UNIVERSITY OF CALIFORNIA SAN DIEGO

Search for anomalous c2v couplings in the VVH production via vector boson scattering

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Physics

by

Yifan Xiang

Committee in charge:

Professor Frank Wuerthwein, Chair
Professor Javier Duarte
Professor Marc Muendler
Professor Yi-zhuang You

2024

The Dissertation of Yifan Xiang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

This dissertation is dedicated to my parents, whose unwavering support has been my anchor throughout the journey of completing my PhD. Your belief in my abilities and endless love have powered me through challenges and inspired me to strive for excellence.

I would also like to extend my deepest gratitude to my friend Jofan, whose generosity and support during one of the most challenging periods of my life were invaluable. Jofan, thank you for providing me a home and comfort when I needed it the most, especially during the sorrowful time of losing a close friend. Your kindness will never be forgotten.

To all of you, thank you for lighting my path with your guidance and love.

EPIGRAPH

There was no answer,
except the general answer life gives to all the most complex and insoluble questions.
That answer is:
one must live in the needs of the day, in other words, become oblivious

Lev Nikolayevich Tolstoy

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Frank Wuerthwein for his support as the chair of my committee. Through multiple drafts and many long nights, his guidance has proved to be invaluable.

This dissertation, in full, is a reprint of the material as it appears in CMS analysis note AN-23-016, Search for anomalous c2v couplings in the VVH production via vector boson scattering. The dissertation author was the primary investigator and author of this analysis note.

VITA

2018 Bachelor of Science in Physics, University of Science and Technology of China

2024 Doctor of Philosophy in Physics, University of California San Diego

PUBLICATIONS

ABSTRACT OF THE DISSERTATION

Search for anomalous c2v couplings in the VVH production via vector boson scattering

by

Yifan Xiang

Doctor of Philosophy in Physics

University of California San Diego, 2024

Professor Frank Wuerthwein, Chair

In this dissertation, we detail the investigation of anomalous C_{2V} couplings in the production process ($pp \to VVH + jj$) with a single-lepton final state. This research utilizes a dataset from proton-proton collisions at ($\sqrt{s} = 13TeV$) acquired by the CMS experiment over the years 2016, 2017 and 2018. The dataset corresponds to an integrated luminosity of $138 fb^{-1}$. The analysis focuses on final states in which the vector bosons and the Higgs boson, decaying hadronically, are identified as single large-radius jets. We explore a semi-leptonic final state, characterizing one of the W boson candidates through the tagging of a high transverse momentum, isolated muon or electron.

Chapter 1

Introduction

1.1 Introduction to CMS

Physicists are captivated by questions that span from the infinite to the infinitesimal: What is the universe made of? How do quarks behave? How does the sun burn?

Physicists from around the world gather at the Large Hadron Collider (LHC) to explore these questions. The LHC is the world's largest and most powerful particle accelerator, situated 100 meters below the Swiss-French border near Geneva. It features a 27 km ring of superconducting magnets that accelerate protons to 7 TeV each, creating head-to-head collisions at 14 TeV. These collisions occur at a staggering rate of 600 million per second, with protons traveling at 99.9999991% the speed of light.

Inside the LHC, two beams of high-energy particles travel in opposite directions and collide, scattering particles in all directions. Massive particle detectors are positioned at the collision points to capture the resulting events. Among the four main experiments at the LHC—ALICE, ATLAS, CMS, and LHCb—CMS is one of the two largest detectors, alongside ATLAS. Although both have similar scientific goals, they employ different technical solutions and magnet system designs to validate each other's results. The Compact Muon Solenoid (CMS) is a general-purpose detector designed to study a wide range of physics topics, including supersymmetry, the Higgs boson, and dark matter.

The CMS collaboration consists of 5,500 physicists, engineers, and students from 241

institutes across 54 countries. Despite its considerable size—15 meters high and 21 meters long—the name ”Compact” reflects its highly integrated design. The detector is optimized to detect muons with great precision. It weighs 14,000 tonnes, has 75 million individual channels, and includes a superconducting cable that generates a 4-Tesla magnetic field, the strongest solenoid magnet ever built.

The CMS detector functions like a high-speed camera, recording collisions 40 million times per second. It comprises several sub-detectors, each with a specific function. The innermost layer, the silicon tracker, traces the paths of charged particles. Next is the Electromagnetic Calorimeter (ECAL), which absorbs electrons and photons. Beyond the ECAL is the Hadron Calorimeter, which measures the energy of hadrons, consisting of quarks and gluons. Then comes the solenoid itself, and finally, the muon detector, which is placed in the outermost layer because muons can pass through several meters of iron without interacting. This powerful and complex machine allows physicists to delve into the fundamental aspects of particle physics, helping to unlock the mysteries of the universe.

1.2 Introduction to Particle Physics

It is well known that matter is made of atoms. In 1897, J.J. Thomson discovered electrons and showed that the atom has a complex structure. In the early 20th century, protons and neutrons were discovered. These three particles constitute the whole world.

In the 1960s, physicists found that protons and neutrons are made of smaller particles called quarks. A proton consists of two up quarks and one down quark, while a neutron consists of two down quarks and one up quark.

However, this is not the end of the story. Up and down quarks, along with electrons and electron neutrinos, are only the first generation of matter. Nature provides us with two additional generations to explain the vast number of discoveries made with accelerators. These 12 particles form the fundamental building blocks of the universe. The last ingredient of the Standard Model

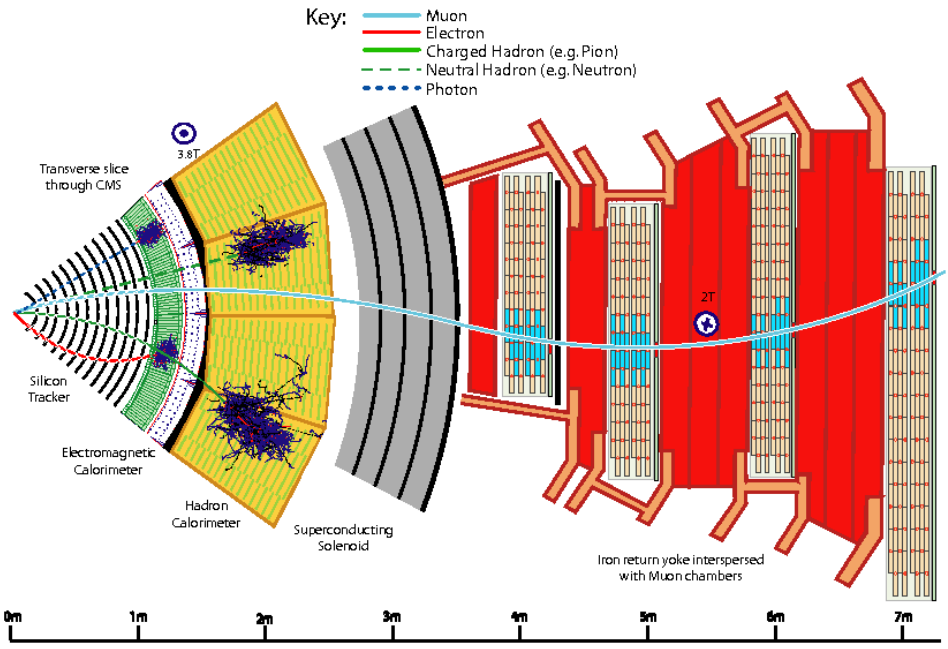


Figure 1.1. The CMS detector at the LHC. This plot shows the various sub-detectors of the CMS detector. The innermost layer is the silicon tracker, followed by the electromagnetic calorimeter, the hadron calorimeter, the solenoid, and the muon detector.

is the force-carrying particles.

The Standard Model describes three fundamental forces. Electromagnetism is mediated by photons, the weak force by W and Z bosons, and the strong force by gluons.

In the 1960s, there were issues with the Standard Model when incorporating the mass of W and Z bosons. In 1964, Peter Higgs proposed a new field, known as the Higgs field, to explain mass. According to his theory, particles acquire mass through interactions with the Higgs field. The particle associated with this field is the Higgs boson. In 2012, the Higgs boson was discovered by the ATLAS and CMS experiments.

1.3 Introduction To The Analysis

The Higgs boson, discovered in 2012 by the ATLAS and CMS collaborations at the CERN LHC [12, 15, 16], has been the subject of extensive studies since its identification. Measurements of the Higgs boson's properties have generally aligned with Standard Model

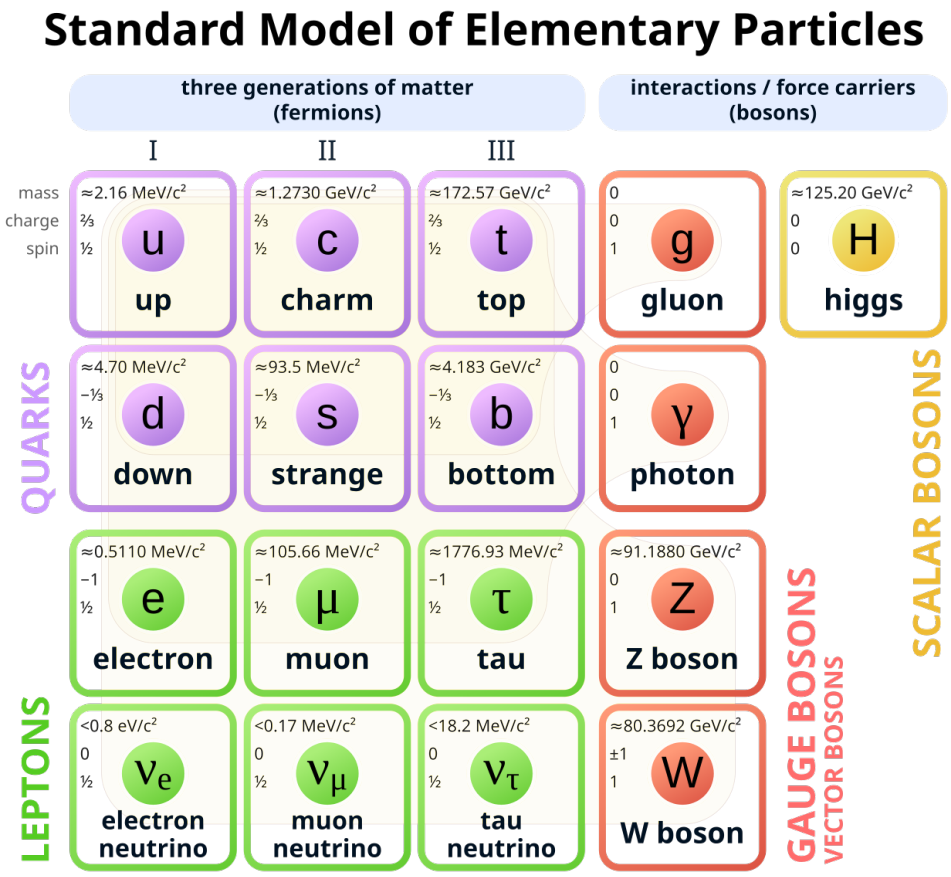


Figure 1.2. The particles of the Standard Model. This plot shows the three generations of matter particles, the force-carrying particles, and the Higgs boson.

(SM) predictions [18, 19]. Following the discovery, research expanded into the production of Higgs boson pairs (HH), a process that enables direct measurements of the Higgs boson trilinear self-coupling and the quartic couplings between two Higgs bosons and two vector bosons. Both ATLAS and CMS have recently published results that constrain these couplings [10, 11], with the CMS results specifying the Higgs-gauge quartic coupling to be within the range [0.62, 1.41] ([0.66, 1.37] expected).

Another significant process, the vector boson scattering (VBS) triboson $W^{\pm}W^{\pm}H$ production, also provides insights into the trilinear Higgs self-coupling c_{λ} and the Higgs-gauge quartic coupling c_{VV} [21, 22]. These couplings, when deviating from the SM predictions, can significantly increase the cross section at high $\hat{s}$, making it possible to probe scenarios beyond

the Standard Model (BSM) [21, 22]. Additionally, the HWW and HZZ (c_V) couplings also contribute to the process.

This analysis focuses primarily on the Higgs-gauge quartic coupling c_{VV}, aiming to set limits that are competitive with the latest publications. The potential modifications from the Standard Model values are expressed through the coupling modifiers κ_λ and κ_{VV}.

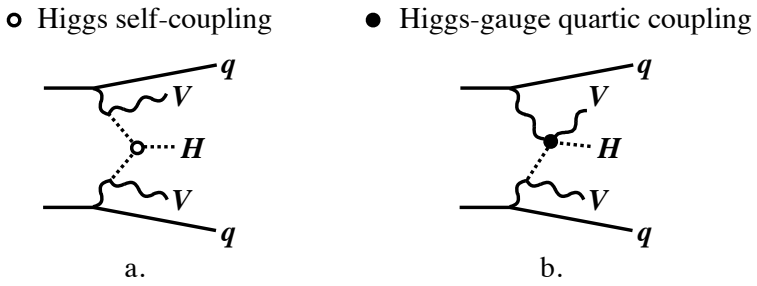


Figure 1.3. Tree-level Feynman diagrams of vector boson scattering multi-boson productions with Higgs boson in the final state, the Higgs self-coupling, and the Higgs-gauge quartic coupling are marked by ○, and ●, respectively.

The cross section increases as a function of the κ_{VV} value in both directions around 1 (SM expectation), as shown in figure 1.4. As the cross section grows, the final state topology also changes. The higher the cross section, corresponding to higher $\hat{s}$ values, the higher is the Lorentz boost of the final state objects and their decay products.

A number of final states can be considered in this analysis. As shown in figure 1.5, the all-hadronic and semi-leptonic final states take the major share of the total VBS VVH cross section and these are the ones covered in this analysis note. Given that we want to probe beyond-SM scenarios, where the final state objects have a large Lorentz boost, our analysis focuses on final state with at least a boson (V or H) reconstructed as a boosted object in a single large cone jet.

Chapter 1, in full, is a reprint of the material as it appears in CMS analysis note AN-23-016, Search for anomalous c2v couplings in the VVH production via vector boson scattering.

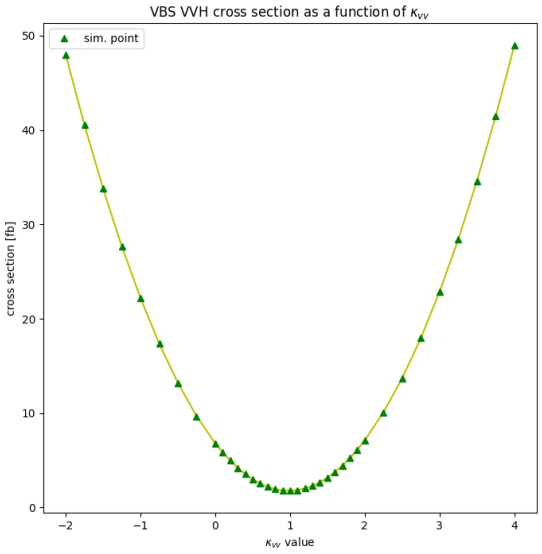


Figure 1.4. Cross section of the VBS VVH production as a function of κ_{VV}.

The dissertation author was the primary investigator and author of this analysis note.

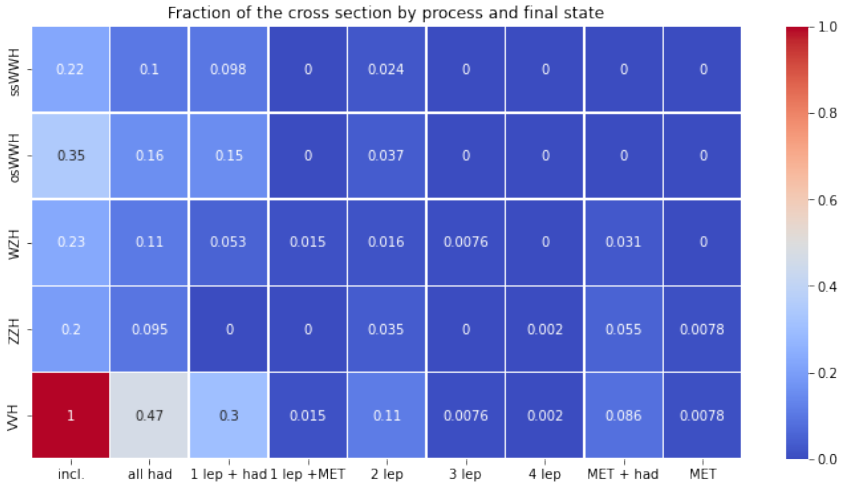


Figure 1.5. Share of the cross-section of the VVH process for all the final states of the VV decays. The dominant final states are the all-hadronic and the semi-leptonic, followed by final states with 2 leptons or MET(the missing energy of the events) +hadrons. NB: In this analysis we consider only the Higgs decays to $b\bar{b}$.

Chapter 2

Samples

In the forthcoming sections of this chapter, we delineate the datasets that form the cornerstone of our analysis, comprising both the empirical data from CMS Run II and the Monte Carlo (MC) simulations that model the expected signal and background processes.

2.1 Collision data

The analysis delineated within this document is predicated upon the proton-proton collision data from Run 2 of the Large Hadron Collider (LHC), as recorded by the CMS experiment. It employs a dataset that embodies an integrated luminosity of 138 fb^{-1}, enabling a robust statistical analysis of the recorded events.

To facilitate this analysis, we utilize several primary datasets. For the single-lepton channel, the SingleMuon and SingleElectron datasets are indispensable, while the JetHT dataset is specifically harnessed for the zero-lepton channel. Each dataset is meticulously curated through the employment of designated triggers, the details of which will be expounded in subsequent sections.

The data tier that serves as the foundation for the analysis is the NanoAOD data tier, specifically version 9, which has been processed under the "Ultra Legacy" (UL, a complete reprocessing of the entire Run 2 dataset with the most updated algorithms and calibrations.) reconstruction campaign. This ensures the utilization of the most refined and comprehensive

dataset available.

An essential aspect of data curation is the implementation of the "golden" JSON selection, (the list of "good" luminosity sections) a rigorous standard applied to all data employed in this analysis. The golden JSON files, along with their respective integrated luminosity contributions for each data-taking period, are systematically cataloged in Table 2.1. These JSON files are paramount in certifying the quality and viability of the collision events included in our study.

Table 2.1. Golden JSON files used to certify the proton-proton collision events for this analysis, along with the corresponding integrated luminosity for each period.

Year	**Golden JSON file**	**Luminosity (fb^{-1})**
2016 (pre-VFP)	Cert 271036-284044 13TeV Legacy2016 Collisions16 JSON.txt	19.52
2016 (post-VFP)	Cert 271036-284044 13TeV Legacy2016 Collisions16 JSON.txt	16.81
2017	Cert 294927-306462 13TeV UL2017 Collisions17 GoldenJSON.txt	41.53
2018	Cert 314472-325175 13TeV Legacy2018 Collisions18 JSON.txt	59.74

2.2 Monte Carlo simulations

In this section, we will delineate the Monte Carlo (MC) samples utilized, which are instrumental in simulating the expected signal and background processes.

2.2.1 Signal

For the comprehensive analysis of the process $pp \to W^{\pm}W^{\pm}H + jj$, Monte Carlo simulations play an indispensable role. These simulations have been generated at leading order (LO) using the `MADGRAPH5 aMC@NLO` generator [13], leveraging a modified version of the Standard Model `MADGRAPH` model with the Higgs-gauge coupling modifier κ_{VV} set to 2. This adjustment is critical as it aligns with the anticipated sensitivity of our analysis.

A suite of four distinct signal samples has been produced to cover each data-taking period, with 2016 being further subdivided into pre- and post-VFP intervals:

- VBSWWH OS Mjj100 4f TuneCP5 [9] 13TeV-madgraph-pythia8
- VBSWWH SS Mjj100 4f TuneCP5 13TeV-madgraph-pythia8
- VBSWZH Mjj100 4f TuneCP5 13TeV-madgraph-pythia8
- VBSZZH Mjj100 4f TuneCP5 13TeV-madgraph-pythia8

These samples are categorized by the vector boson content and charge, each generated with an m_{jj} phase space cut of 100 GeV, where m_{jj} denotes the invariant mass of any two jets. MADGRAPH's simulation output is harmonized with PYTHIA (version 8) for parton showering, employing the CP5 tune and the dipole recoil option for heightened fidelity in the particle shower simulation.

The choice to generate signal samples with $\kappa_{VV} = 2$ is predicated on the expectation that the sensitivity of our analysis will peak around this value. With the applied generator cuts, the cross-section for the process is calculated to be 7.12 fb. Empirical observations have indicated that the kinematic distributions are largely invariant when κ_{VV} is adjusted within a specific range beyond the Standard Model expectations, ensuring the analysis optimized at $\kappa_{VV} = 2$ is robust across a spectrum of kinematic scenarios. It is noteworthy, however, that as κ_{VV} converges towards 1, a discernible variation in event kinematics emerges, resulting in a diminished sensitivity of our approach. Further details on the signal generation and the kinematic distributions as a function of κ_{VV} are available in Appendix A.

Additionally, the signal samples include a reweighting scheme that is applied MADGRAPH. This scheme enables us to survey the κ_{VV} landscape from -2 to 4. The cross-sections associated with these processes demonstrate a quadratic increase as a function of κ_{VV}, aligning with theoretical expectations. The computed cross-sections are systematically tabulated in Table 2.

Table 2.2. The cross sections for WWH, WZH, and ZZH for a spectrum of κ_{VV}. The cross sections for $\kappa_{VV} = 2.0$ are reported by `MADGRAPH5 aMC@NLO`, while the rest are obtained through the reweighting features of `MADGRAPH5 aMC@NLO`.

κ_{VV}	σ [fb]			
	$W^{\pm}W^{\pm}H$	$W^{\pm}W^{\mp}H$	WZH	ZZH
-2.0	8.66	19.66	10.59	9.06
-1.75	7.36	16.6	8.97	7.62
-1.5	6.18	13.8	7.49	6.32
-1.25	5.1	11.28	6.16	5.13
-1.0	4.15	9.02	4.96	4.07
-0.75	3.3	7.03	3.91	3.14
-0.5	2.57	5.3	2.99	2.33
-0.25	1.95	3.85	2.22	1.65
0.0	1.45	2.66	1.59	1.09
0.1	1.28	2.26	1.38	0.9
0.2	1.13	1.91	1.19	0.73
0.3	1.0	1.59	1.02	0.59
0.4	0.88	1.32	0.88	0.46
0.5	0.79	1.09	0.75	0.35
0.6	0.71	0.91	0.66	0.26
0.7	0.65	0.77	0.58	0.19
0.8	0.61	0.67	0.52	0.15
0.9	0.58	0.61	0.49	0.12
1.0	0.58	0.6	0.48	0.11
1.1	0.59	0.63	0.5	0.12
1.2	0.62	0.7	0.53	0.15
1.3	0.67	0.82	0.59	0.21
1.4	0.74	0.98	0.67	0.28
1.5	0.82	1.18	0.78	0.37
1.6	0.93	1.42	0.9	0.48
1.7	1.05	1.71	1.05	0.62
1.8	1.19	2.04	1.23	0.77
1.9	1.35	2.41	1.42	0.94
2.0	1.53	2.83	1.64	1.13
2.25	2.05	4.06	2.28	1.7
2.5	2.68	5.55	3.06	2.39
2.75	3.43	7.32	3.99	3.21
3.0	4.29	9.35	5.05	4.16
3.25	5.27	11.65	6.26	5.23
3.5	6.36	14.22	7.61	6.42
3.75	7.57	17.06	9.1	7.74
4.0	8.88	20.16	10.73	9.18

2.2.2 Background

The Monte Carlo (MC) simulation samples utilized to characterize the background processes in this analysis are cataloged in Tables 2.5 and 2.6. These samples are critical for representing the principal backgrounds encountered in our study, namely top quark, and V+jets production.

A particular focus is given to the treatment of the W+jets background. The cross-sections for W+jets, often a dominant background in studies involving jets and missing transverse momentum, are carefully adjusted to achieve a seamless integration across different energy scales. This is detailed in the following subsection.

LHE H_{T} binned W+jets stitching correction

The differential cross-sections of the W+jets samples categorized by hadronic activity (H_{T}) were not employed in this study (see [6, 7]). Instead, only the inclusive cross-section was utilized, and the yields were adjusted to match the relative proportions observed in the inclusive sample. This approach, however, introduces discontinuities at the LHE H_{T} transition points, as exemplified in Figure 2.1. To achieve a continuous spectrum, a stitching procedure was implemented, mirroring the methodology previously applied to pre-UL samples. In these instances, cross-sections were modified to mitigate recognized discrepancies.

Each sample's LHE H_{T} spectrum was individually fitted using the exponential function $f(x) = e^{ax+b}$. The analysis focused exclusively on either the upper or lower extremity of the spectrum. Continuity between the distributions of higher and lower H_{T}-binned samples was maintained by determining a scale factor based on the intersection point x, described by the equation:

$$\alpha = \frac{f(x) higher H_{\mathrm{T}} - binned}{f(x) lower H_{\mathrm{T}} - binned}. \tag{2.1}$$

For example, in connecting the HT 600-800 GeV sample to the HT 800-1200 GeV sample, the procedure involved:

- Fitting the lower boundary of the W+Jets H_{T} 600-800 sample's LHE H_{T} distribution within the range of $[780, 800]$ using the function $f_1(x) = e^{ax+b}$.
- Similarly, fitting the upper boundary of the W+Jets H_{T} 800-1200 sample's LHE H_{T} distribution within the range of $[800, 840]$ using the function $f_2(x) = e^{ax+b}$.
- The scale factor, α, was then calculated using $\alpha = \frac{f_1(800)}{f_2(800)}$.

The resulting scale factors for all samples are documented in Table 2.3. Additionally, the fitted slopes and the interim weights derived during the LHE H_{T} transitions are tabulated in Table 2.4. The outcomes of the reweighting process are illustrated in Figure 2.2. The adjusted cross-sections for each LHE H_{T} bin following stitching are listed in Table 2.5.

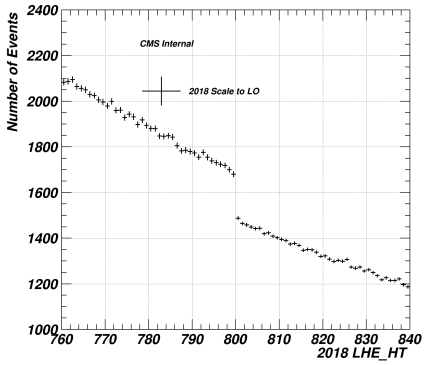


Figure 2.1. Discontinuity at the transition between the LHE H_{T} $[600, 800]$ sample and the LHE H_{T} $[800, 1200]$ sample for the 2018 data taking period.

Table 2.3. WJets global stitching factors by year.

H_{T} bin / year	2016-preVFP	2016-postVFP	2017	2018
LHE $H_{\mathrm{T}} \in [200, 400]$	0.98979	0.9799	0.975082	0.990735
LHE $H_{\mathrm{T}} \in [400, 600]$	1.00073	0.970773	0.980262	1.00713
LHE $H_{\mathrm{T}} \in [600, 800]$	1.02926	0.998517	0.98764	1.04205
LHE $H_{\mathrm{T}} \in [800, 1200]$	1.07868	1.01288	1.02704	1.18941
LHE $H_{\mathrm{T}} \in [1200, 2500]$	1.21157	1.17629	1.22283	1.14644
LHE $H_{\mathrm{T}} \in [2500, \infty]$	1.36417	2.51032	1.0906	1.16974

Table 2.4. W+jets stitching factors for the different H_{T}-binned samples of the 4 data taking periods of the simulation.

2016 preVFP				
end of lower H_{T}-binned sample	fitting func	begin of higher H_{T}-binned sample	fitting func	stitching factor
LHE $H_{\mathrm{T}} \in [100, 200]$	$e^{-0.0149\times x-3.7}$	LHE $H_{\mathrm{T}} \in [200, 400]$	$e^{-0.0147\times x-3.7}$	0.989790
LHE $H_{\mathrm{T}} \in [200, 400]$	$e^{-0.0090\times x-5.4}$	LHE $H_{\mathrm{T}} \in [400, 600]$	$e^{-0.0088\times x-5.5}$	1.011050
LHE $H_{\mathrm{T}} \in [400, 600]$	$e^{-0.0065\times x-6.6}$	LHE $H_{\mathrm{T}} \in [600, 800]$	$e^{-0.0065\times x-6.6}$	1.028510
LHE $H_{\mathrm{T}} \in [600, 800]$	$e^{-0.0055\times x-7.4}$	LHE $H_{\mathrm{T}} \in [800, 1200]$	$e^{-0.0056\times x-7.4}$	1.048010
LHE $H_{\mathrm{T}} \in [800, 1200]$	$e^{-0.0042\times x-8.7}$	LHE $H_{\mathrm{T}} \in [1200, 2500]$	$e^{-0.0041\times x-8.9}$	1.123200
LHE $H_{\mathrm{T}} \in [1200, 2500]$	$e^{-0.0027\times x-11.4}$	LHE $H_{\mathrm{T}} \in [2500, \infty)$	$e^{-0.0026\times x-11.6}$	1.125950
2016 postVFP				
end of lower H_{T}-binned sample	fitting func	begin of higher H_{T}-binned sample	fitting func	stitching factor
LHE $H_{\mathrm{T}} \in [100, 200]$	$e^{-0.0153\times x-3.9}$	LHE $H_{\mathrm{T}} \in [200, 400]$	$e^{-0.0145\times x-4.0}$	0.979900
LHE $H_{\mathrm{T}} \in [200, 400]$	$e^{-0.0096\times x-5.4}$	LHE $H_{\mathrm{T}} \in [400, 600]$	$e^{-0.0095\times x-5.4}$	0.990686
LHE $H_{\mathrm{T}} \in [400, 600]$	$e^{-0.0059\times x-7.2}$	LHE $H_{\mathrm{T}} \in [600, 800]$	$e^{-0.0062\times x-7.1}$	1.028580
LHE $H_{\mathrm{T}} \in [600, 800]$	$e^{-0.0057\times x-7.5}$	LHE $H_{\mathrm{T}} \in [800, 1200]$	$e^{-0.0055\times x-7.7}$	1.014390
LHE $H_{\mathrm{T}} \in [800, 1200]$	$e^{-0.0044\times x-8.7}$	LHE $H_{\mathrm{T}} \in [1200, 2500]$	$e^{-0.0040\times x-9.3}$	1.161330
LHE $H_{\mathrm{T}} \in [1200, 2500]$	$e^{-0.0031\times x-10.6}$	LHE $H_{\mathrm{T}} \in [2500, \infty)$	$e^{-0.0026\times x-12.6}$	2.134100
2017				
end of lower H_{T}-binned sample	fitting func	begin of higher H_{T}-binned sample	fitting func	stitching factor
LHE $H_{\mathrm{T}} \in [100, 200]$	$e^{-0.0153\times x-2.9}$	LHE $H_{\mathrm{T}} \in [200, 400]$	$e^{-0.0146\times x-3.0}$	0.975082
LHE $H_{\mathrm{T}} \in [200, 400]$	$e^{-0.0094\times x-4.5}$	LHE $H_{\mathrm{T}} \in [400, 600]$	$e^{-0.0093\times x-4.5}$	1.005310
LHE $H_{\mathrm{T}} \in [400, 600]$	$e^{-0.0072\times x-5.5}$	LHE $H_{\mathrm{T}} \in [600, 800]$	$e^{-0.0066\times x-5.9}$	1.007530
LHE $H_{\mathrm{T}} \in [600, 800]$	$e^{-0.0060\times x-6.3}$	LHE $H_{\mathrm{T}} \in [800, 1200]$	$e^{-0.0055\times x-6.7}$	1.039890
LHE $H_{\mathrm{T}} \in [800, 1200]$	$e^{-0.0041\times x-8.1}$	LHE $H_{\mathrm{T}} \in [1200, 2500]$	$e^{-0.0040\times x-8.4}$	1.190640
LHE $H_{\mathrm{T}} \in [1200, 2500]$	$e^{-0.0030\times x-9.9}$	LHE $H_{\mathrm{T}} \in [2500, \infty)$	$e^{-0.0026\times x-10.7}$	0.891870
2018				
end of lower H_{T}-binned sample	fitting func	begin of higher H_{T}-binned sample	fitting func	stitching factor
LHE $H_{\mathrm{T}} \in [100, 200]$	$e^{-0.0149\times x-2.4}$	LHE $H_{\mathrm{T}} \in [200, 400]$	$e^{-0.0146\times x-2.5}$	0.990735
LHE $H_{\mathrm{T}} \in [200, 400]$	$e^{-0.0093\times x-4.0}$	LHE $H_{\mathrm{T}} \in [400, 600]$	$e^{-0.0089\times x-4.2}$	1.016540
LHE $H_{\mathrm{T}} \in [400, 600]$	$e^{-0.0065\times x-5.4}$	LHE $H_{\mathrm{T}} \in [600, 800]$	$e^{-0.0067\times x-5.3}$	1.034680
LHE $H_{\mathrm{T}} \in [600, 800]$	$e^{-0.0055\times x-6.2}$	LHE $H_{\mathrm{T}} \in [800, 1200]$	$e^{-0.0054\times x-6.4}$	1.141420
LHE $H_{\mathrm{T}} \in [800, 1200]$	$e^{-0.0041\times x-7.7}$	LHE $H_{\mathrm{T}} \in [1200, 2500]$	$e^{-0.0040\times x-7.7}$	0.963873
LHE $H_{\mathrm{T}} \in [1200, 2500]$	$e^{-0.0026\times x-10.1}$	LHE $H_{\mathrm{T}} \in [2500, \infty)$	$e^{-0.0025\times x-10.4}$	1.020330

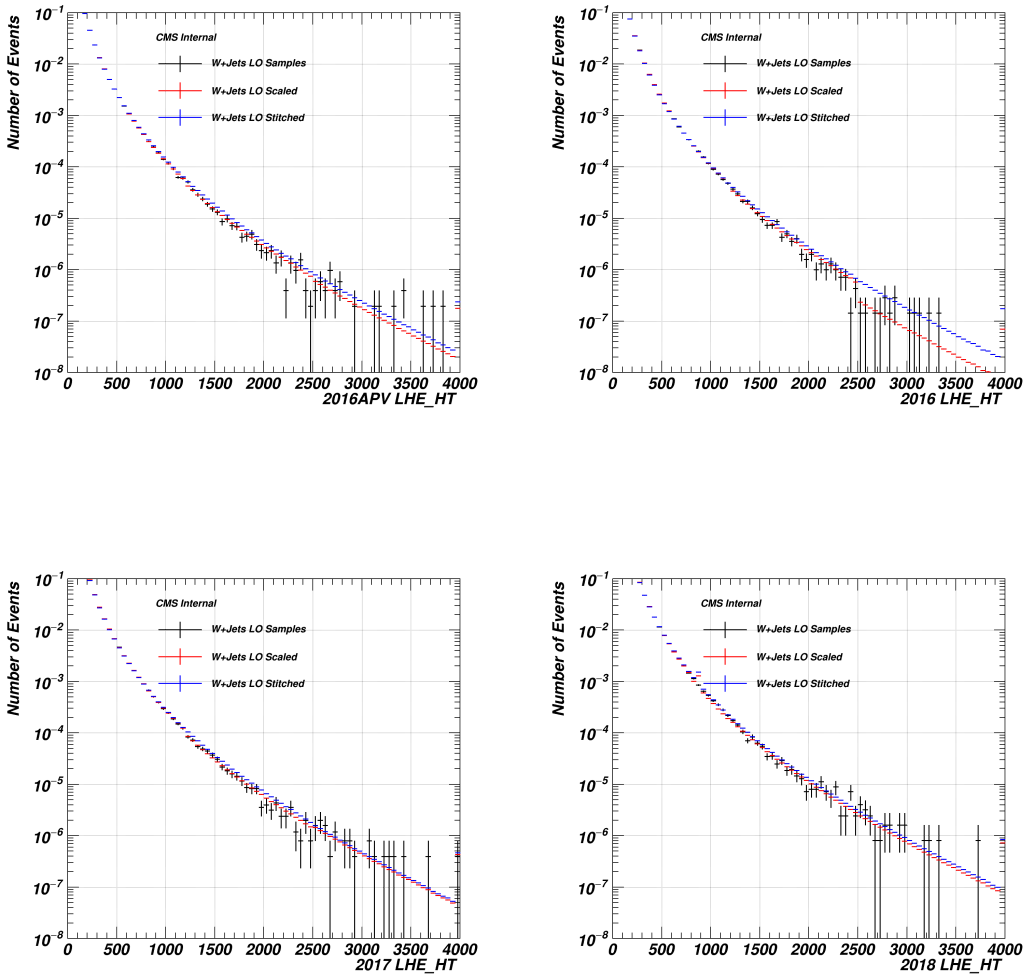


Figure 2.2. W+jets LHE H_{T} spectrum before (red) and after (blue) the stitching and comparison with the inclusive H_{T} sample (black).

Chapter 2, in full, is a reprint of the material as it appears in CMS analysis note AN-23-016, Search for anomalous c2v couplings in the VVH production via vector boson scattering. The dissertation author was the primary investigator and author of this analysis note.

Table 2.5. Background MC samples used in this analysis, corresponding to 2016preVFP, 2016, 2017, 2018 detector conditions (UL), with their respective cross sections in pb. A detailed summary of the stitching the W+jets samples is presented in the following section. 4 cross-sections are reposrted for the W+jets, as the stitching gives slightly different results for each data taking period.

Process	Sample Name	σ [pb]
QCD	/QCD_HT100to200_TuneCP5_13TeV-madgraph-pythia8	27849880
	/QCD_HT200to300_TuneCP5_13TeV-madgraph-pythia8	1716997
	/QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8	351302
	/QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8	31630
	/QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8	6802
	/QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8	1206
	/QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8	98.71
	/QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8	20.2
$t\bar{t}$	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8 [0 3 6 10]	88.29
	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8 [0 3 6 10]	365.34
	TTToHadronic_TuneCP5_13TeV-powheg-pythia8 [0 3 6 10]	377.96
Single Top	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8 [0 3 6 10]	80.95
	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8 [0 3 6 10]	136.02
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 [0 4 7 11]	19.559
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 [0 4 7 11]	19.559
W+jets	/WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8 [0 3 6 10]	1310.78/1283.91 1319.76/1321.16 **
	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8 [0 3 6 10]	1325.9/1303.06 1334.74/1335.7 **
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8 [0 3 6 10]	348.57030/341.046 350.435/351.689 **
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8 [0 3 6 10]	47.308275/45.4362 46.5726/47.1663 **
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8 [0 3 6 10]	11.358487/11.0051 11.1485/11.4196 **
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8 [0 3 8 10]	5.2086934/4.94177 5.02246/5.12389 **
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8 [0 3 6 10]	1.1880809/1.15544 1.183/1.18295 **
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8 [1 4 7 11]	0.024098031/0.0216234 0.0258083/0.0255202 **
Drell-Yan	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8 [0 3 6 10]	20657
	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 [0 3 6 10]	6198

[0] /RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANODSIM
[1] /RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANODSIM
[2] /RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11_ext1-v1/NANODSIM
[3] /RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANODSIM
[4] /RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANODSIM
[5] /RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17_ext1-v1/NANODSIM
[6] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANODSIM
[7] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANODSIM
[8] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v3/NANODSIM
[9] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9_ext1-v2/NANODSIM
[10] /RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANODSIM
[11] /RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANODSIM
[12] /RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1_ext1-v2/NANODSIM
** The W+jets cross sections come from the reweighting to the relative fraction in the inclusive sample and the stitching. They are listed in the order 2016APV, 2016, 2017, 2018. The cross-sections are slightly different due to the stitching.

Table 2.6. Rarer background MC samples used in this analysis, corresponding to 2016preVFP, 2016, 2017, 2018 detector conditions (UL), with their respective cross sections in pb.

Process	Sample Name	σ [pb]
EWK V	EWKWMinus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8 [0 3 7 11]	32.26
	EWKWPlus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8 [0 3 7 11]	39.33
	EWKWminus2Jets_WToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8 [0 3 6 10]	10.67
	EWKWplus2Jets_WToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8 [0 3 6 10]	10.67
	EWKZ2Jets_ZToLL_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8 [0 3 7 11]	6.22
	EWKZ2Jets_ZToNuNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8 [1 4 7 11]	10.72
	EWKZ2Jets_ZToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8 [0 3 6 10]	10.67
$t\bar{t}X$	TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8 [1 3 6 10]	0.2043
	TTWW_TuneCP5_13TeV-madgraph-pythia8 [0 3 6 10]	0.0115
	TTWZ_TuneCP5_13TeV-madgraph-pythia8 [0 3 6 10]	0.003884
	TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8 [0 3 6 10]	0.2529
	TTbb_4f_TTTo2L2Nu_TuneCP5-Powheg-Openloops-Pythia8 [0 3 6 10]	0.04
	TTbb_4f_TTToSemiLeptonic_TuneCP5-Powheg-Openloops-Pythia8 [0 3 6 10]	0.62
	ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8 [1 4 7 11]	0.215
	ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8 [1 4 7 11]	0.1279
Bosons	VBFWH_HToBB_WToLNu_M-125_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8 [0 3 6 10]	0.02656
	VHToNonbb_M125_TuneCP5_13TeV-amcatnloFXFX_madspin_pythia8 [1 4 7 11]	2.207
	WWJJToLNuLNu_EWK_noTop_TuneCP5_13TeV-madgraph-pythia8 [0 3 6 10]	0.284
	WWTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8 [0 3 6 10]	49.997
	WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8 [0 3 7 11]	12.178
	WWTolnulnu_TuneCP5_13TeV-madgraph-pythia8 [0 3 6 10]	12.178
	WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8 [2 5 9 12]	0.2086
	WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8 [2 5 6 12]	0.1651
	WZJJ_EWK_InclusivePolarization_TuneCP5_13TeV_madgraph-madspin-pythia8 [0 3 6 10]	0.01701
	WZTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8 [0 3 6 10]	49.997
	WZTo1L3Nu_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8 [0 3 6 10]	3.05402
	WZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8 [1 4 7 10]	5.6
	WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8 [0 3 7 11]	4.42965
	WZZ_TuneCP5_13TeV-amcatnlo-pythia8 [2 5 9 12]	0.05565
	WminusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8 [0 3 6 10]	0.0490124
	WplusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8 [0 3 6 10]	0.084876
	ZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8 [0 3 6 10]	0.0262749
	ZZJJTo4L_TuneCP5_13TeV-madgraph-pythia8 [1 4 7 11]	0.00884
	ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8 [0 3 6 10]	0.564
	ZZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8 [0 3 6 10]	3.28
	ZZTo4L_M-1toInf_TuneCP5_13TeV_powheg_pythia8 [0 3 6 10]	1.256
	ZZZ_TuneCP5_13TeV-amcatnlo-pythia8 [2 5 9 12]	0.01398
	ggZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8 [0 3 6 10]	0.0024614

[0] /RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANODSIM
[1] /RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANODSIM
[2] /RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11_ext1-v1/NANODSIM
[3] /RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANODSIM
[4] /RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANODSIM
[5] /RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17_ext1-v1/NANODSIM
[6] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANODSIM
[7] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANODSIM
[8] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v3/NANODSIM
[9] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9_ext1-v2/NANODSIM
[10] /RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANODSIM
[11] /RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANODSIM
[12] /RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1_ext1-v2/NANODSIM

Chapter 3

Physics Object

In this section, we delineate the physics objects that are pivotal to our analysis. The primary entities of interest in this study are leptons and jets, each playing crucial roles in the processes under investigation. Specifically, two distinct jet cone sizes are utilized:

1. **AK8 Jets:** These are employed predominantly for the reconstruction of boosted vector bosons that decay into hadrons. The larger cone size of AK8 jets facilitates the capture and analysis of the broader debris patterns characteristic of such high-momentum decay events.

2. **AK4 Jets:** These are essential for identifying the Vector Boson Scattering (VBS) signature. Additionally, AK4 jets are instrumental in the removal of background processes that may obscure or mimic the VBS signal.

The selection and characterization of these jets are tailored to enhance the sensitivity and specificity of the analysis, thereby ensuring robust detection capabilities for the physical phenomena of interest.

3.1 Leptons

In this analysis, we adhere to the ttH Ultra Legacy lepton selection criteria [23] as delineated in CMS internal analysis note (CMS_AN_2019_111). The specific selection parameters

for muons and electrons are detailed in Tables 3.1 and 3.2, respectively. For the signal lepto, we employ the tight identification criterion. Conversely, the veto criterion is utilized to exclude additional leptons.

It is important to note that although τ leptons are included in the final states depicted in Figure 1.5, our focus here is exclusively on isolated electrons and muons. Consequently, hadronic decay modes of τ are expressly excluded from this analysis.

Leptons for this study are extracted from the respective NanoAOD collections. Specifically, electrons in NanoAODv9 have already been corrected for residual energy scale and resolutions. Furthermore, for electron selection, the criterion based on the pseudorapidity, $|\eta|$, is applied to $|\eta_{SC}|$. This is defined in the NanoAOD data-tier branches as |Electron_eta + Electron_etaSC| > 2.5.

Table 3.1. Electron Selectron Criteria for veto and tight working points.

Observable	**veto**	**tight**
p_T	> 7 GeV	> 10 GeV
$\|\eta\|$	< 2.5	< 2.5
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm
|SIP3D| (or d/σ_d)	< 8	< 8
miniPFRelIso_all	< 0.4	< 0.4
Lost hits in the tracker	at most 1	at most 0
mvaFall17V2noIso_WPL (POG ID)	passed	passed
Conversion veto	-	passed
Tight charge	-	$= 2$
Deep Jet of nearby jet	-	<WP-medium
mvaTTHUL [0]	-	> 0.8
H/E	-	< 0.1
1/E - 1/p	-	$>$ -0.04
$\sigma^0_{i\eta i\eta}$	-	$< 0.011/0.030$

[0] $\sigma_{i\eta i\eta}$ is a measure of the spatial distribution of energy deposition in the electromagnetic calorimeter for electrons. The two cut values (0.011/0.030) are for the $|\eta_{SC}|^1 \leq 1.479$ and $1.479 < |\eta_{SC}|^1 < 2.5$ regions respectively.

Table 3.2. Muon selection criteria for veto and tight working points.

Observable	Veto	Tight
p_T	> 5 GeV	> 10 GeV
$\|\eta\|$	< 2.4	< 2.4
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm
sip3d (or d/σ_d)	< 8	< 8
miniPFRelIso_all	≤ 0.40	≤ 0.40
Loose POG ID	true	true
Medium POG ID	-	true
Deep Jet of nearby jet	-	<WP-medium
mvaTTHUL	-	> 0.85

3.1.1 Lepton Scale Factor Treatment

The treatment of lepton scale factors adheres to a factorization prescription as outlined in provided documentation [26, 31]. These scale factors are essential for aligning the lepton efficiencies in Monte Carlo simulations with those observed in data. The scale factors applied include ”reco” and ”loose ID”, which are centrally derived, alongside ”ISO+IP” and ”tight ID”, which are specifically derived to meet the ttH identification criteria. In the 1-lepton channel, single lepton triggers are employed, and a centrally derived scale factor is applied to ensure that the trigger efficiency in the Monte Carlo matches the efficiency measured in actual data. The systematic uncertainties associated with each scale factor and their impact on the analysis are detailed in Chapter 5.

3.2 Ak8 jets

In the study of Vector Boson Scattering involving Higgs and W bosons (VBSVVH), deviations from the Standard Model, particularly in the κ_{VV} coupling, can lead to an increased production rate. Such enhancements often result in the production of highly boosted Higgs and W bosons. When these bosons are produced with substantial momentum, they may decay into

two quarks that are so closely aligned that they appear as a single expansive "fat" jet. For their reconstruction, we employ the anti-k_T jet clustering algorithm with a large radius parameter (R = 0.8), utilizing AK8 jets from the NanoAOD collection where Ultra Legacy (UL) corrections have been applied.

The adoption of AK8 jets for the reconstruction of heavy boosted objects offers several benefits:

- It results in a lower background compared to the resolved case.
- The jet substructure aids in further background discrimination.

The selection criteria for preselecting AK8 jets are detailed in Table 3.3.

Moreover, we employ the ParticleNet graph neural network architecture to classify jets based on the substructure and properties of their constituent particles. This algorithm utilizes the four-momentum vectors of jet constituents, as well as secondary verticles, reconstructed via particle-flow techniques, and processes these through convolutional and fully connected layers to extract features and classify the jets. Compared to traditional jet classification methods, such as those reliant on jet substructure variables or boosted decision trees, ParticleNet has demonstrated significant enhancements in performance.

Specifically, we use the mass-decorrelated version of ParticleNet, termed ParticleNet MD, to minimize mass sculpting effects. In conjunction with the ParticleNet MD taggers, we also employ the ParticleNet regressed mass to capitalize on its superior resolution, notably better than that offered by the softdrop mass. This variable, while not used at the selection stage, is later incorporated as an input to multivariate analyses.

3.2.1 Higgs Tagging

The Higgs tagging score (Hbb Score) is calculated by assessing the likelihood that a fat jet is a Higgs boson decaying to *bb* ($FatJet\_particleNetMD\_Xbb$), relative to its identification as a QCD jet ($FatJet\_particleNetMD\_QCD$). The mathematical formulation for the Higgs score

Table 3.3. Selection criteria for selecting fatjets coming from heavy boosted bosons

Criteria	Cut value
Fat jet p_T	> 250 GeV
Fat jet $\|\eta\|$	< 2.5
Fat jet mass	> 50 GeV
Fat jet softdrop mass	> 40 GeV
Fat Jet jetId(use jet Id only)	Passed
Separation between fat jet and lepton	> 0.8 in ΔR metric

is expressed as follows:

$$\text{Hbb Score} = \frac{ParticleNetMD_{Xbb}}{ParticleNetMD_{Xbb} + ParticleNetMD_{QCD}} \tag{3.1}$$

The criteria for selecting Hbb candidates are thoroughly detailed in Table 3.3. Following the application of all pre-selection criteria, the jet with the highest Hbb Score, as defined above, is selected.

3.2.2 W and Z Tagging

Similarly, the score for W and Z bosons (Vqq Score) is determined by the ratio of the probabilities that a fat jet is identified as a W-jet ($FatJet\_particleNetMD\_Xcc$ and Xqq) to its probability of being classified as a QCD jet ($FatJet\_particleNetMD\_QCD$). The formula for the W score is:

$$\text{Vqq Score} = \frac{ParticleNetMD_{Xcc} + ParticleNetMD_{Xqq}}{ParticleNetMD_{Xcc} + ParticleNetMD_{Xqq} + ParticleNetMD_{QCD}} \tag{3.2}$$

The pre-selection criteria for Vqq fat jets are identical to those for Hbb candidates (see Table 3.3). In the 1-lepton channel, the AK8 jet with the highest Vqq Score is selected as the Vqq candidate after the Hbb jet has been tagged. In scenarios involving the WZH, no specific adjustments are made for Z bosons, given the similarity in appearance of hadronically decaying Z and W bosons.

3.2.3 VBS Jets

VBS jet candidates are initially pre-selected by applying the criteria specified in Table 3.4 to the list of AK4 jets. These jets are sourced from the NanoAOD collection (AK4 CHS) with Ultra Legacy (UL) corrections applied. To refine our selection, a forward-backward jet selection technique is employed.

The highest momentum pre-selected AK4 jet in the $\eta > 0$ region is paired with the highest momentum pre-selected AK4 jet in the $\eta < 0$ region to form the VBS jet candidates. This selection method aligns with the characteristics of VBS events, which typically produce two high-energy jets that are significantly separated in rapidity. This technique not only follows the natural production mechanism of VBS events but also effectively reduces the influence of background processes that generally produce jets more isotropically distributed in rapidity.

In cases where it is not possible to identify distinct VBS jets in both forward and backward regions (i.e., when all jets are observed within the same pseudorapidity sign), the two highest momentum jets are selected as substitutes for the VBS jets. This approach leverages the typical high-energy profile and substantial rapidity separation observed in VBS processes, making it a viable alternative for jet selection. While the forward-backward selection method is preferred for its efficiency in background suppression, choosing the jets with the highest momentum still maintains a reasonable signal-to-background ratio, thus ensuring the integrity of the VBS jet identification process.

3.2.4 Missing Transverse Energy

The missing transverse energy ($p_{T,\text{miss}}$ or MET) is sourced from the NanoAOD collection. In the analysis, Type 1 MET is utilized. Corresponding corrections, as applied to the jets, are also implemented to ensure accuracy and consistency in MET calculations.

Table 3.4. Jet Selection Criteria

Criteria	Cut Value
Jet p_T	> 30 GeV
Jet $\|\eta\|$	< 4.7
Jet $\|\Delta R\|$ from lepton	> 0.4
Jet $\|\Delta R\|$ from Hbb and Vqq AK8 candidate jets	> 0.8
Jet Id [0] for 2016	$>= 1$
Jet Id [0] for 2017 and 2018	$>= 2$
Jet pileup Id (puId) for $p_T >= 50$ jets	> 0

[0] The new criteria used in 2017 and 2018 required jets to pass more stringent requirements on their shape, as well as on their composition of charged and neutral particles, in order to reduce the contamination of jets from pileup and other sources of background events. The Jet_jetId cut value was increased to $>= 2$ in order to ensure that only the highest-quality jets were included in the analysis, while still retaining a high efficiency for signal events.

3.2.5 Treatment of the HEM15/16 Region in 2018 Data

The HEM15/16 issue [8] is addressed following the recommended guidelines. In the Monte Carlo simulations for the 2018 data, adjustments are made to the jet momenta as follows:

- Jets within $-1.57 < \phi < -0.87$ and $-2.5 < \eta < -1.3$ are scaled down by 20%.
- Jets within $-1.57 < \phi < -0.87$ and $-3.0 < \eta < -2.5$ undergo a 35% reduction.

This scaling affects approximately two-thirds of the genuine jets—specifically, those with $p_T > 15$ GeV and passing the tight identification criteria to exclude muons and electrons. The adjustments to jet momenta are also propagated to the MET calculations to maintain consistency in the simulation data.

Chapter 3, in full, is a reprint of the material as it appears in CMS analysis note AN-23-016, Search for anomalous c2v couplings in the VVH production via vector boson scattering. The dissertation author was the primary investigator and author of this analysis note.

Chapter 4

Analysis Techniques

4.1 Preselection

The initial stage of our analysis involves a relatively loose preselection, designed to efficiently filter the dataset while retaining potential signal events. This preselection serves as a critical preliminary step before engaging in the more complex multivariate analysis and is integral to the data-driven background estimation that underpins the reliability of our findings.

The preselection criteria are crafted to exclude data that are clearly non-relevant or of poor quality, thereby enhancing the overall efficiency of subsequent analysis steps. These criteria include basic kinematic filters, data quality checks, and adherence to specific event characteristics that are known to be associated with the signals of interest. Details of these criteria are specified in the following subsections, each tailored to address distinct aspects of the dataset and the experimental setup.

By applying these preselection criteria, we ensure that the data moving forward into the multivariate analysis phase are of high quality and relevance, thus setting a robust foundation for precise and accurate analysis outcomes.

4.1.1 Trigger Selection

For the 1-lepton channel of our analysis, single lepton triggers are employed. These triggers are particularly advantageous within the dense hadronic environment of the Large

Hadron Collider (LHC). They offer a straightforward method to efficiently select events likely to contain the signal processes of interest.

The analysis primarily targets high P_T isolated leptons, thereby ensuring that the trigger thresholds do not pose a limiting factor in event selection. The selection of these triggers is strategically aligned with the characteristics of the targeted leptons, maximizing the efficiency and efficacy of the data acquisition process.

Furthermore, to maintain the integrity of our dataset, a series of data quality filters are applied. These filters, as recommended by the JetMET POG and listed in Table 4.2, are designed to eliminate anomalous high-MET events that may arise due to various reconstruction failures or detector malfunctions. Common issues addressed include noise or dead cells in the ECAL or HCAL, beam halo effects, and event misreconstruction. By applying these event quality filters, the analysis is restricted to using only 'good' events, adhering to the standard practices for CMS analyses.

The specific triggers utilized for each primary dataset and corresponding year of data collection are detailed in Table 4.1. This table provides a comprehensive overview of the trigger configurations, facilitating a clear understanding of the data selection mechanisms implemented in this study.

4.1.2 Lepton Selection

In the process of selecting events for analysis, we focus on events that contain exactly one lepton that meets our tight selection criteria and no veto leptons. This approach is taken to minimize potential background contributions from top quark pair production and to maintain orthogonality with other analysis channels. The specific criteria for tight and veto lepton selection are elaborated in Chapter 3.

Given our interest in Beyond Standard Model (BSM) physics scenarios, particularly those where the Higgs boson's gauge coupling constant, κ_{VV}, deviates from unity, the lepton dynamics are critical. The leptons, originating from the decay of a W boson, are expected to

Table 4.1. HLT triggers used in the analysis.

Year	Sample Name	HLT trigger
2016	Single Muon	IsoMu22 IsoTkMu22 IsoMu24 IsoTkMu24
2016	Single Electron	Ele25_eta2p1_WPTight_Gsf Ele27_eta2p1_WPTight_Gsf
2017	Single Muon	IsoMu24 IsoMu27
2017	Single Electron	Ele35_WPTight_Gsf Ele32_WPTight_Gsf
2018	Single Muon	IsoMu24 IsoMu27
2018	EGamma	Ele35_WPTight_Gsf Ele32_WPTight_Gsf

Table 4.2. Data quality flags used in the analysis.

Flag	Description	Status
Flag_goodVertices	Good vertex requirement	Passed
Flag_globalSuperTightHalo2016Filter	Global super tight halo filter (for data)	Passed
Flag_HBHENoiseFilter	HB/HE noise filter	Passed
Flag_HBHENoiseIsoFilter	HB/HE noise isolation filter	Passed
Flag_EcalDeadCellTriggerPrimitiveFilter	ECAL dead cell trigger primitive filter	Passed
Flag_BadPFMuonFilter	Bad PF muon filter	Passed
Flag_BadPFMuonDzFilter	Bad PF muon Dz filter	Passed
Flag_eeBadScFilter	EE badSC noise filter	Passed
Flag_ecalBadCalibFilter	ECAL bad calibration filter (for 2017 and 2018 data)	Passed

be highly boosted. Consequently, we have implemented a stringent selection criterion on the lepton's transverse momentum P_T, requiring it to exceed 40 GeV.

Details regarding the specific selection criteria applied post-lepton identification are meticulously outlined in Table 4.3. This structured approach ensures that the selected leptons are highly probable to contribute meaningfully to the signal while reducing background noise in the dataset.

4.1.3 AK8 Jets Selection

For the analysis, events must contain at least two AK8 jets that do not overlap with the selected lepton. This requirement ensures that the jets are distinctly associated with the hard

Table 4.3. Lepton selection in 1 lepton analysis

Lepton Selection	Cut Value
ttH lepton tight selection	1
Additional veto ttH lepton	0
lepton p_T	> 40 GeV

scattering event rather than the lepton emission processes. The criteria used to select these fat jets, referred to as AK8 jets, are fully detailed in Section 3.2.

In the selection hierarchy, the $H_{b\bar{b}}$ and $V_{q\bar{q}}$ candidates are prioritized in that order. Each candidate is chosen based on the ParticleNet scores, which assess the likelihood of the jets matching the desired final state characteristics. This scoring system provides a robust method for identifying the jets most likely to represent the $H_{b\bar{b}}$ and $V_{q\bar{q}}$ decay signatures, thereby enhancing the precision of our event classification.

The specific selection process using ParticleNet scores and the related criteria for these candidates are thoroughly explained in Section 3. This approach optimizes the selection of fat jets critical for probing the targeted physical processes within our study.

4.1.4 Top Background Reduction

A predominant source of background in this analysis originates from the semileptonic $t\bar{t}$ (top-antitop) production. To mitigate this, we focus on analyzing the b-jet content of each event.

Extra b-jets

Alongside the selection of VBS jets, extra AK4 jets are identified. These jets are required to not overlap with selected leptons or AK8 jets, ensuring that they are likely associated with the primary hard scattering event rather than secondary processes. Only jets that fall within the tracker acceptance are considered for b-jet identification, optimizing the efficiency and accuracy of b-tagging.

The comprehensive set of pre-selection criteria for these b-jet candidates, including

aspects like jet kinematics and quality filters, is detailed in Table 4.4. This structured selection process is essential for effectively reducing the top background, thereby enhancing the clarity and significance of the analysis results.

Table 4.4. Jet Selection Criteria

Criteria	Cut Value
Jet p_T	> 20 GeV
Jet $\|\eta\|$	< 2.5
Jet $\|\Delta R\|$ from tight ID lepton	> 0.4
Jet $\|\Delta R\|$ from $H_{b\bar{b}}$ and $V_{q\bar{q}}$ candidates	> 0.8
Jet Id [0] for 2016	$>= 1$
Jet Id [0] for 2017 and 2018	$>= 2$
Jet pileup Id (puId) for $p_T >= 50$ jets	> 0

The b-jets within our dataset are tagged using the DeepJet algorithm, a deep learning-based approach that provides high-performance b-jet identification. We adhere to the working points recommended by the BTV (B-tagging and Vertexing) POG (Physics Object Group). These working points define the thresholds for the DeepJet scores that are considered sufficient for a jet to be categorized as a b-jet.

The specific working points employed in this analysis are detailed in Table 4.5. This table lists the criteria that ensure optimal balance between b-tagging efficiency and the misidentification rate, crucial for maintaining the integrity and accuracy of the background reduction strategy.

Table 4.5. DeepJet working points used in the analysis

Working Point by year	Loose	Medium	Tight
2016 preVFP	0.0508	0.2598	0.6502
2016 postVFP	0.0408	0.2489	0.6377
2017	0.0532	0.3040	0.7476
2018	0.0490	0.2783	0.7100

In our effort to further reduce the background from top pair production, a stringent veto

strategy is employed against events containing "tight" b-jets. Specifically, if any of the jets within an event is classified as a "tight" b-jet according to the DeepJet criteria outlined in Table 4.5, the event is excluded from further analysis. This vetoing process effectively minimizes contamination from top quark pair production, which is critical for maintaining the purity of the signal sample.

Conversely, if b-jets only meet the "loose" working point but fail to achieve the "tight" identification standard, the events are not discarded. Instead, these jets are utilized to construct the discriminating variable $m_{\ell b}$. This variable, formed from the invariant mass of the lepton and the loosely identified b-jet, serves as a crucial component in distinguishing signal from background, thereby enhancing the analytical robustness of our study.

Top Background Rejection via $m_{\ell b}$

A crucial technique employed to mitigate the $t\bar{t}$ background involves leveraging the invariant mass of the lepton and the b-quark from the top quark decay. In the semileptonic $t\bar{t}$ decay chain, the top quark invariably decays into a W boson and a b quark. When one of the W bosons (in the $t\bar{t} \rightarrow 1\ell$ sample) decays leptonically into a lepton and a neutrino, the invariant mass $m_{\ell b}$, computed from the lepton and the b-quark, can effectively distinguish events characteristic of the leptonic decay of a top quark.

As the top quark has a well-established mass of approximately 172 GeV, the invariant mass $m_{\ell b}$ should not exceed this threshold. Figure 4.1 illustrates this concept, showing a distinct peak around 172 GeV, consistent with the mass of the top quark. However, due to the neutrino escaping detection and thus carrying away some of the energy and momentum, the $m_{\ell b}$ distribution typically peaks at a lower value, around 120 GeV, declining steadily to near zero by about 150 GeV.

In this analysis, the $m_{\ell b}$ variable is constructed using loosely tagged b-jets and the tightly identified lepton. If two or more such jets are present, the minimum $m_{\ell b}$ value is considered for the event. While $m_{\ell b}$ is not directly used in the selection process, it serves as an important input

to the multivariate analysis, providing a refined measure to further discriminate against the $t\bar{t}$ background.

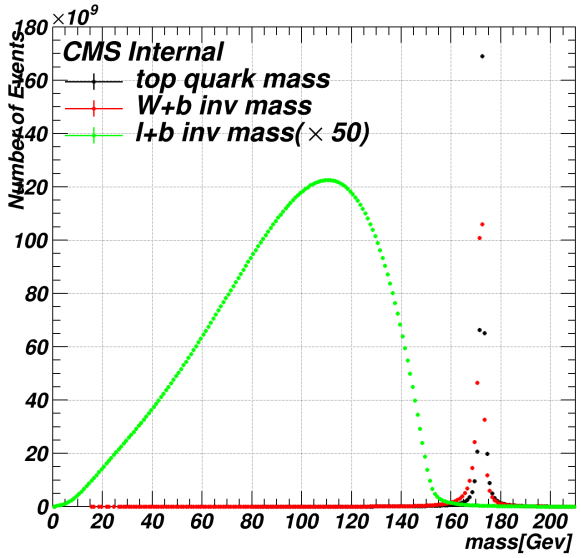


Figure 4.1. This figure illustrates the distribution of the top quark mass at the generator level, the invariant mass of the W+b system originating from the top quark decay, and the invariant mass of the lepton+b system resulting from the top quark decay in the case of leptonic decay of the W boson. All of them are using generator level information. To enhance the visibility of the plot, the green line representing the invariant mass of the lepton and b quark has been scaled up by a factor of 50.

4.1.5 VBS Jets

The selection criteria for VBS jets are comprehensively detailed in Section 3, as these criteria are uniformly applied across our analysis. Events that do not feature at least two additional jets, which can be identified as VBS jet candidates, are excluded from further analysis. This strict requirement ensures that only events with clear jet signatures characteristic of VBS processes are considered, enhancing the purity of the analysis sample.

S_{T} Selection

To further refine our selection prior to the multivariate analysis, we incorporate the S_{T} variable, which is defined as the scalar sum of the transverse momenta of key event components:

$$S_{\mathrm{T}} = p_{T,\ell} + p_{T,H_{b\bar{b}}} + p_{T,V_{q\bar{q}}} + p_{T,\mathrm{miss}} \tag{4.1}$$

for the 1-lepton channel. This variable effectively represents the total transverse energy of the central objects in the event, providing a measure of the overall event activity.

Events are selected for further analysis only if they exhibit an S_{T} greater than 1000 GeV. This threshold is set to ensure a favorable signal-to-background ratio, focusing on events that are highly energetic and thus more likely to be signal rather than background. This criterion is crucial for isolating the most promising events for detailed analysis and is pivotal in maintaining the statistical robustness of our findings.

4.2 Cutflow

The progression of both the background and signal yields through each stage of the selection process is meticulously documented. The cumulative effects of successive selection criteria on the dataset can be observed in Table 4.6, which details the yields after each selection step.

To ensure the accuracy of our Monte Carlo (MC) simulations in representing the actual experimental conditions, several reweighting procedures are implemented:

- **Pileup (PU) Reweighting:** Adjusts the MC to reflect the impact of multiple proton-proton collisions occurring within the same or nearby bunch crossings, which is particularly prevalent in high-luminosity environments like the LHC.
- **Lepton Scale Factors:** These factors correct for discrepancies between the MC simulation and observed data in terms of lepton identification and reconstruction efficiencies.

- **AK4 b-jet Candidate Scale Factors:** Applied to adjust the efficiency of b-jet identification and selection in the MC to match that observed in data.

It is important to note that this reweighting is partial, given that the background is predominantly estimated from actual data rather than MC. Consequently, several corrections and systematic uncertainties are derived and applied only to the signal samples. This approach allows for a more precise adjustment of the signal representation in our analysis, ensuring that it more accurately reflects the experimental observations.

4.3 Kinematic Distribution Comparison between Data and Monte Carlo

To ascertain the consistency and reliability of our Monte Carlo (MC) simulations, it is imperative to validate the data/MC agreement prior to engaging in the training of multivariate discriminators. This validation involves a detailed comparison of kinematic distributions between observed data and MC simulations. Such comparisons are crucial at specific stages of selection, particularly at the pre-selection level, before proceeding to the multivariate analysis phase. The kinematic distributions of several key variables are compared across data and MC to ensure that the simulations accurately reflect the behaviors observed in real data. These variables include distributions for muons, Higgs bosons, W bosons, b-jets, and VBS jets, as well as global event properties. The respective distributions are depicted in Figures 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, and 4.8, with the signal distributions superimposed for clarity. These plots are specifically presented at the selection stage where the two VBS jets have been identified, corresponding to cut C8 in Table 4.6. This stage is critical as it reflects a significant refinement in the event selection process, providing a focused context for evaluating the alignment between the data and MC simulations. This comparison not only verifies the data/MC consistency but also helps in identifying any potential discrepancies that might affect the subsequent analysis stages.

Table 4.6. Semi-leptonic channel cutflow table. Only cross-section reweighting, PU reweighting, lepton scale factors and AK4 b-tagging scale factors are applied.

Cut	Data	Other	TTX	WJets ToLNu	TTbar	Bkg	Signal
C1: Raw [1]	3.0055×10^8	2.4061×10^8	8.2096×10^7	1.0355×10^8	8.0173×10^8	1.2106×10^9	6.8898×10^6
C2: Xsec+PU reweight [2]	$3.0055 \times 10^8 \pm 17336$	$6.2111 \times 10^7 \pm 14425$	$1.1359 \times 10^5 \pm 20.574$	$5.0741 \times 10^7 \pm 6946.6$	$3.5898 \times 10^7 \pm 1300.5$	$1.2654 \times 10^8 \pm 15114$	375.01 ± 0.24877
C3: Trigger [3]	$1.9124 \times 10^8 \pm 13829$	$3.1815 \times 10^7 \pm 8698.8$	66693 ± 15.783	$2.839 \times 10^7 \pm 5149$	$1.9391 \times 10^7 \pm 936.2$	$6.745 \times 10^7 \pm 9317.7$	255.06 ± 0.20595
C4: Lepton p_T [4]	$6.3342 \times 10^7 \pm 7958.9$	$1.2903 \times 10^7 \pm 4526.1$	31861 ± 10.977	$1.7812 \times 10^7 \pm 3892.6$	$1.0948 \times 10^7 \pm 688.68$	$3.4237 \times 10^7 \pm 5176.8$	137.71 ± 0.14998
C5: $H_{b\bar{b}}$ candidate [5]	$1.8116 \times 10^6 \pm 1345.6$	$2.074 \times 10^5 \pm 467.23$	5019 ± 4.3929	$7.5438 \times 10^5 \pm 430.52$	$8.4894 \times 10^5 \pm 194.74$	$1.815 \times 10^6 \pm 663.78$	90.406 ± 0.12149
C6: $V_{q\bar{q}}$ candidate [6]	91474 ± 302.12	11833 ± 109.8	663.27 ± 1.4777	35891 ± 55.939	43190 ± 43.522	91575 ± 130.69	28.473 ± 0.068071
C7: AK4 tight b-jet veto [7]	67820 ± 260.48	10378 ± 105.65	361.56 ± 1.0958	33850 ± 54.551	22034 ± 31.089	66622 ± 122.9	27.261 ± 0.066602
C8: VBS jets [8]	29310 ± 170.79	3319.2 ± 57.775	193.44 ± 0.81773	9332.8 ± 22.241	13092 ± 23.911	25936 ± 66.361	12.126 ± 0.04354
C9: $S_{\mathrm{T}} > 1000$ [9]	12042 ± 108.51	1645.6 ± 40.217	91.368 ± 0.56998	4544.6 ± 11.343	5054.8 ± 14.868	11335 ± 44.342	11.14 ± 0.041716

[1] The number of events that pass a pre-selection stage(1 veto lepton , 4 ak4 jets or 2 ak4 jets and 1 ak8 jet), before any additional cuts or corrections are applied, no xsec weights
[2] Monte Carlo and pileup reweighting applied;
[3] Trigger selection applied;
[4] Exactly one ttH tight lepton with pT $>$ 40 GeV, no additional veto leptons;
[5] one fat jet as Higgs candidate;
[6] one fat jet as W-jet candidate;
[7] No tight AK4 b-tagged jets;
[8] Two additional jets as VBS jets;
[9] Total scalar sum of transverse momenta, ST, greater than 1000 GeV.

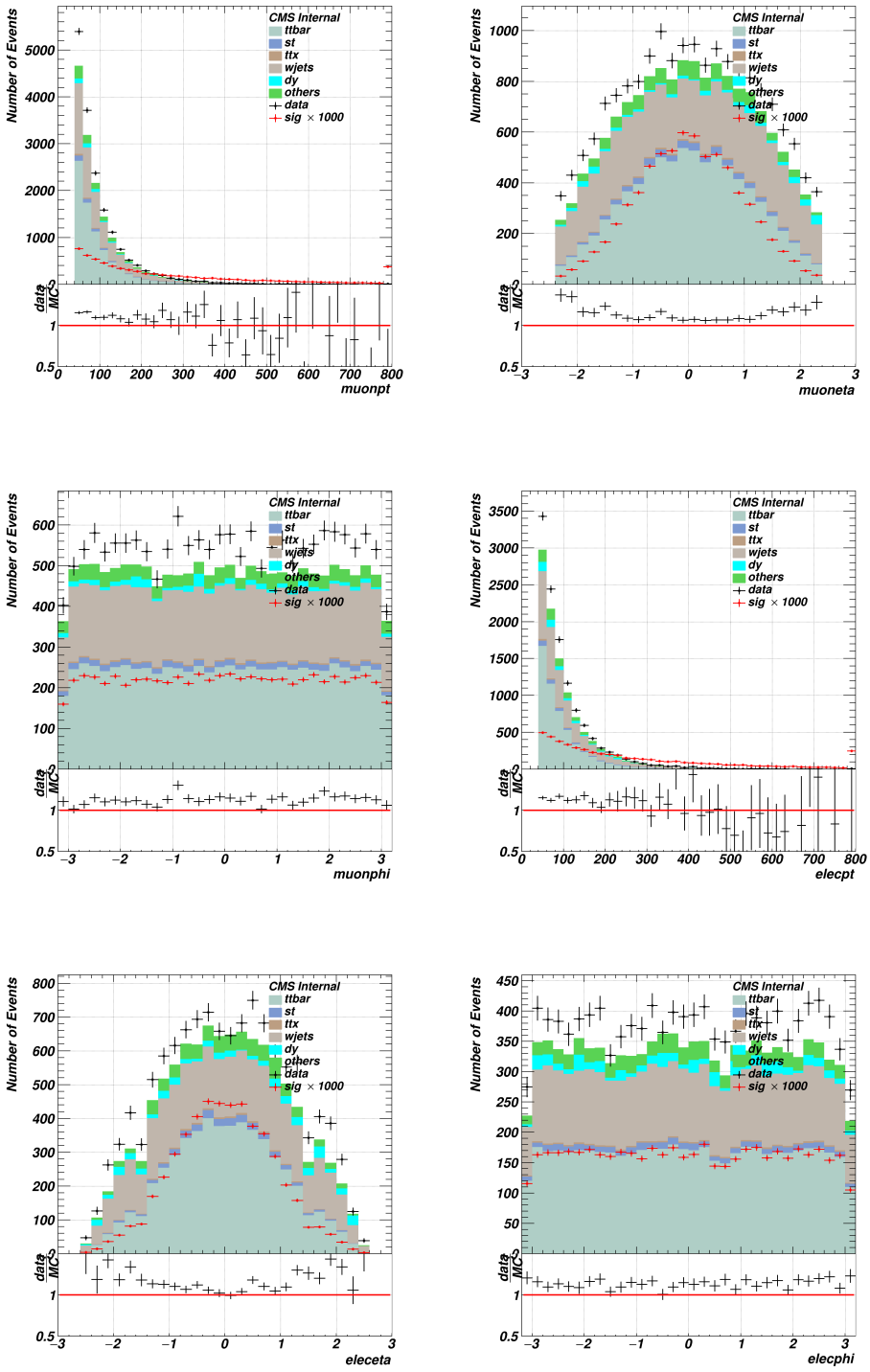


Figure 4.2. Distributions of the and electron muon candidate variables for events passing the VBStag selection criterion in the VBSVVH analysis.

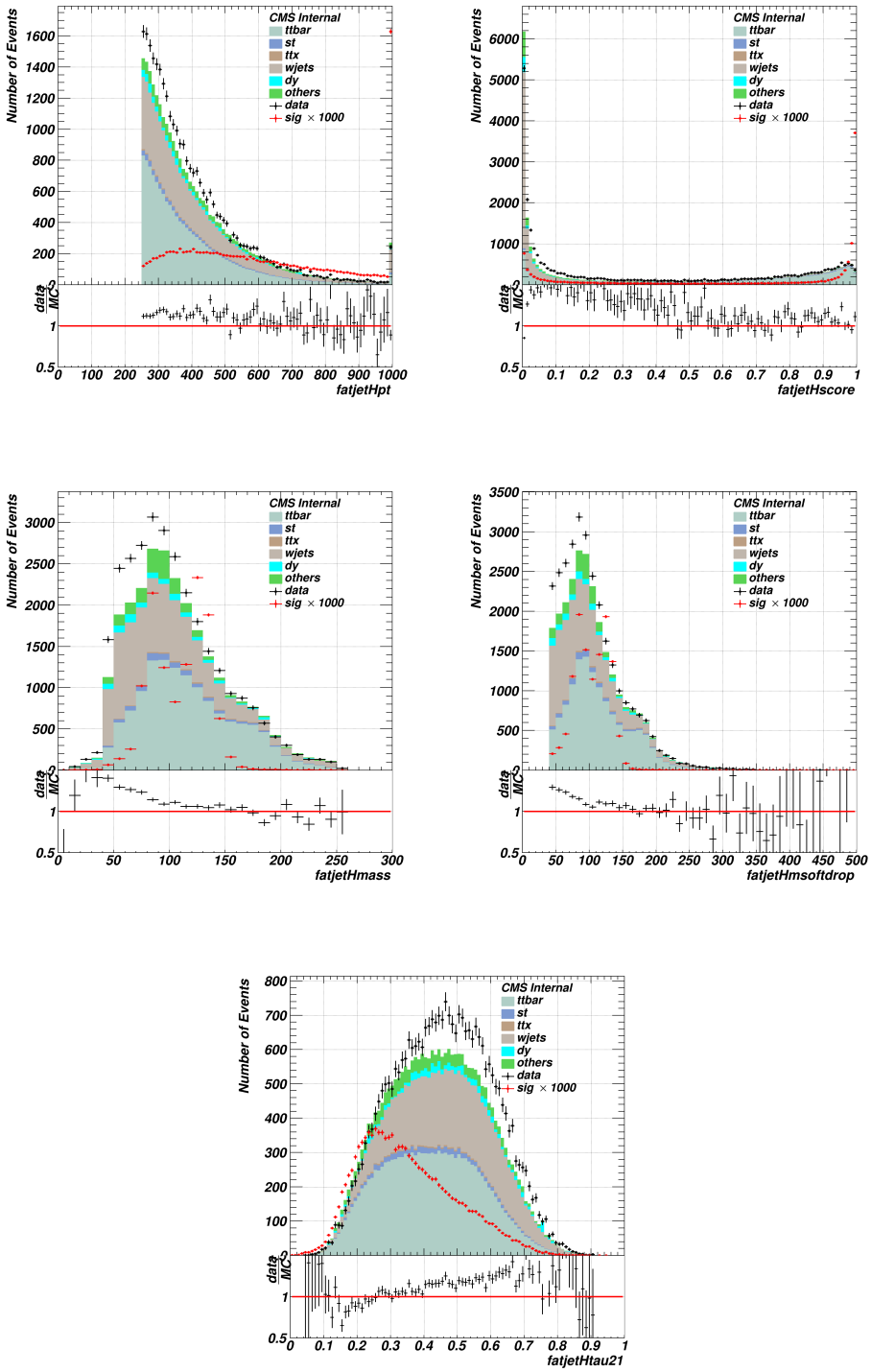


Figure 4.3. Distributions of the H candidate variables for events passing the VBStag selection criterion in the VBSVVH analysis.

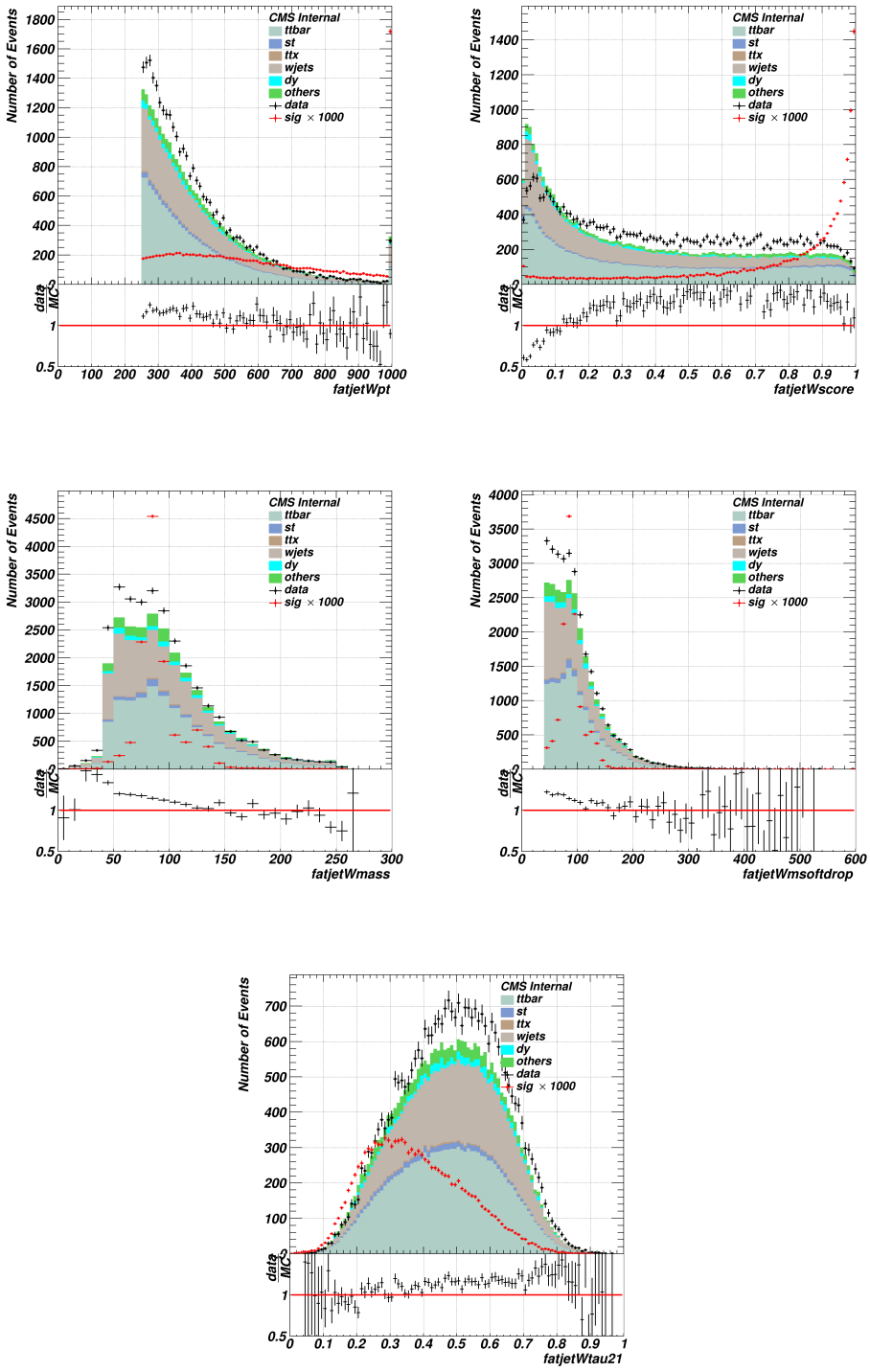


Figure 4.4. Distributions of the W candidate variables for events passing the VBStag selection criterion in the VBSVVH analysis.

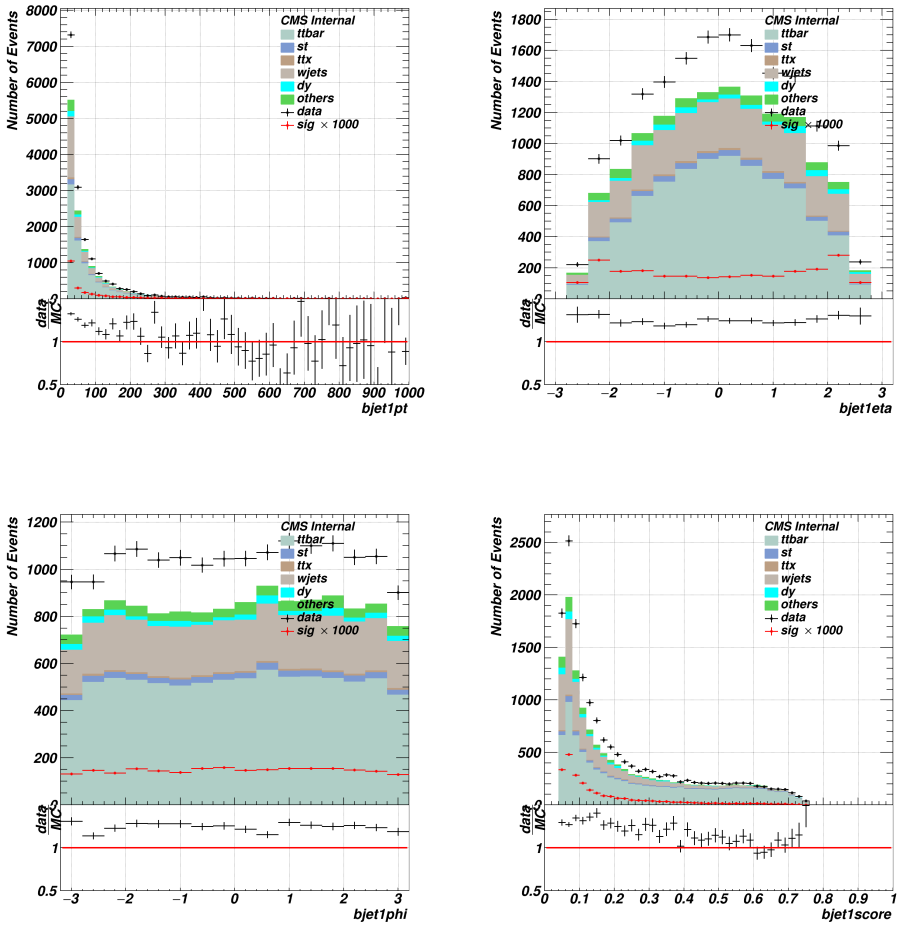


Figure 4.5. Distributions of the 1st b jet candidate variables for events passing the VBStag selection criterion in the VBSVVH analysis.

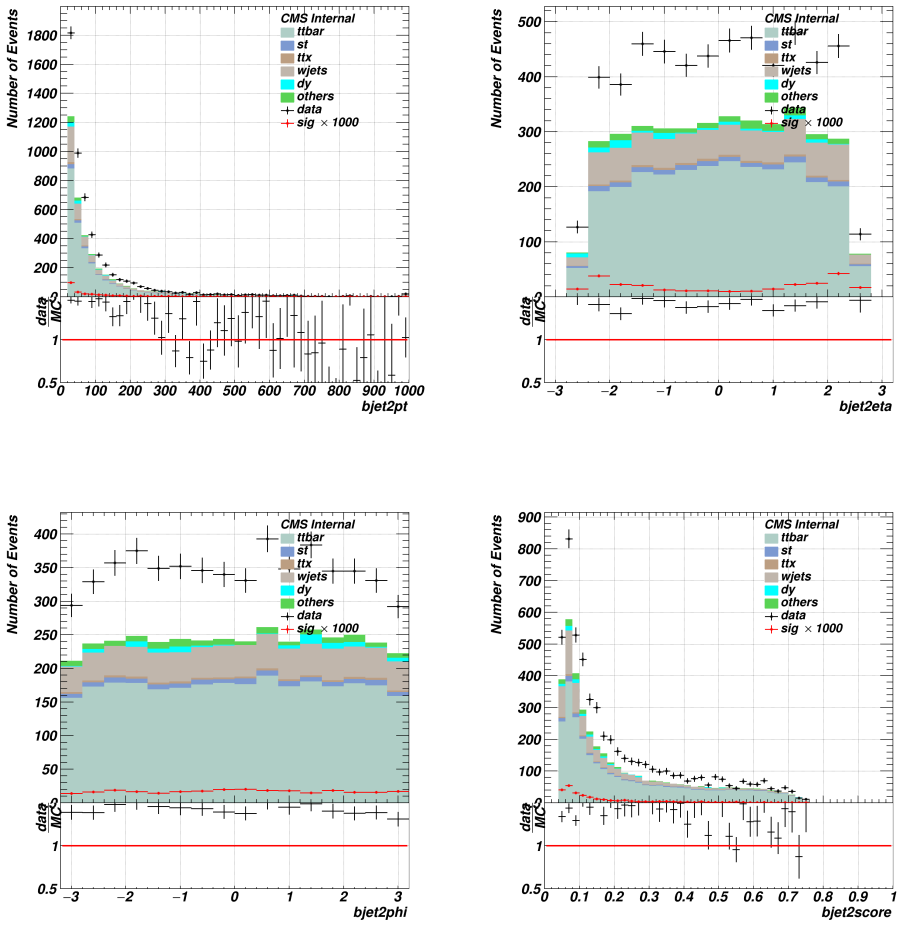


Figure 4.6. Distributions of the 2nd b jet candidate variables for events passing the VBStag selection criterion in the VBSVVH analysis.

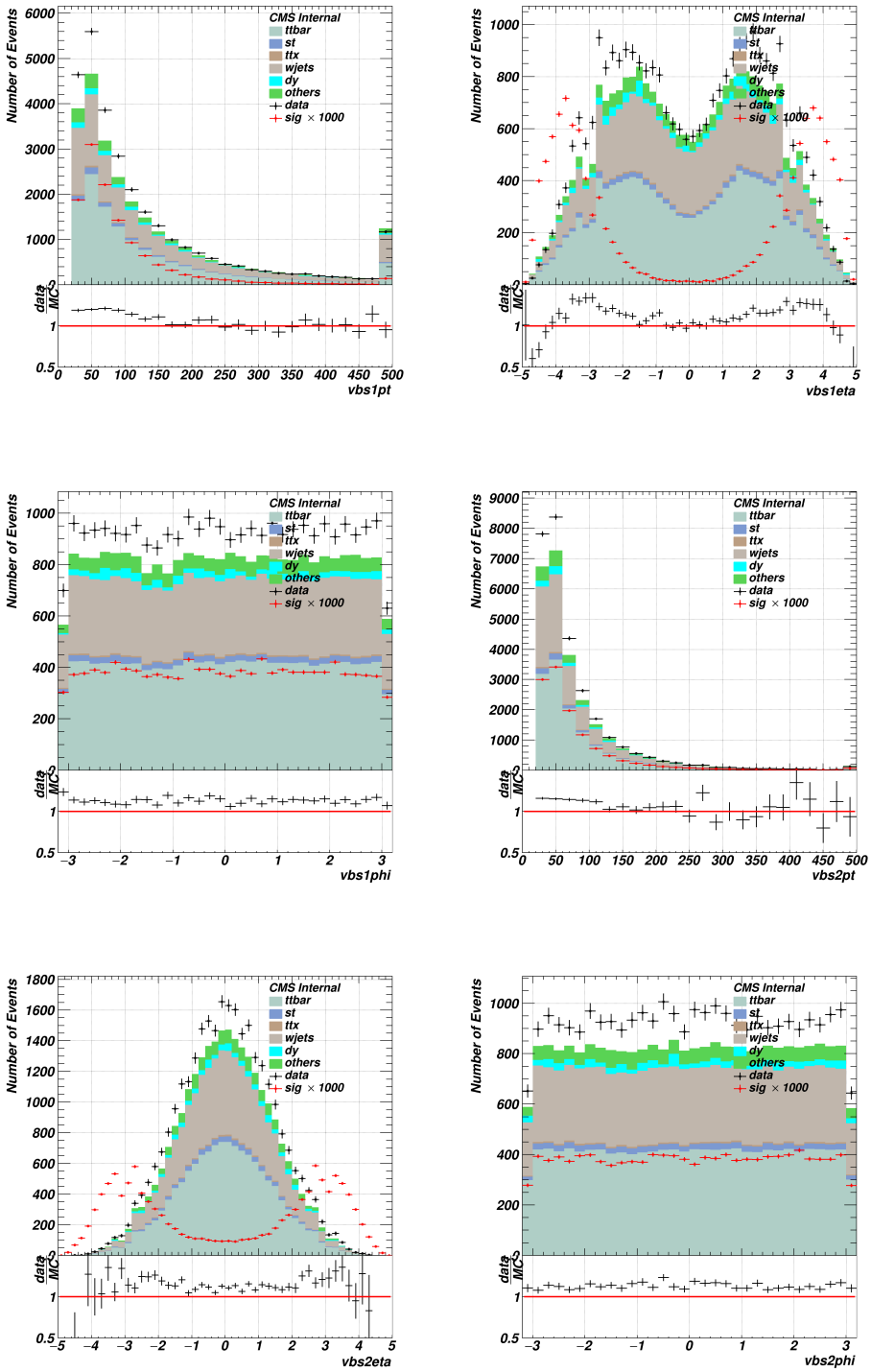


Figure 4.7. Distributions of the vbs jet candidate variables for events passing the VBStag selection criterion in the VBSVVH analysis.

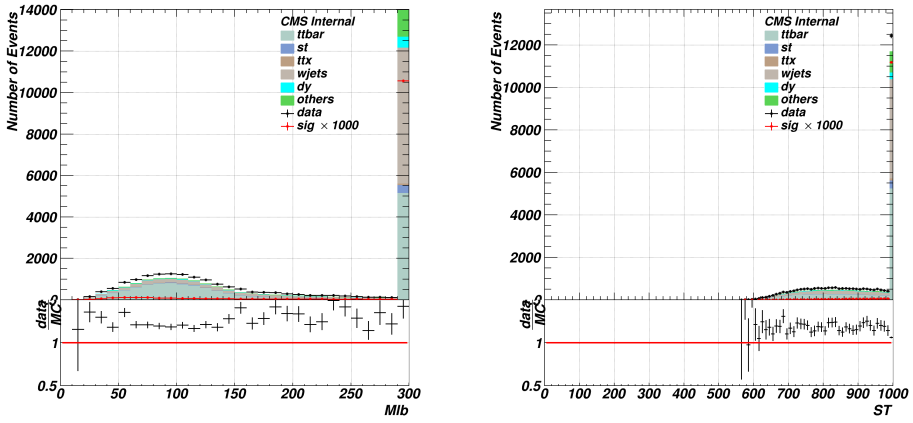


Figure 4.8. Distributions of the $m_{\ell b}$ and S_{T} variables for events passing the VBStag selection criterion in the VBSVVH analysis.

4.4 Multivariate Analysis

Following the preselection phase, our strategy employs a multivariate discriminator to accurately define the optimal signal region. We estimate the background directly from data, which necessitates the use of two approximately orthogonal variables as the ABCD method's arms, one of which is the multivariate discriminator. In the context of our event selection, we differentiate between VBS jets, typically characterized by high pseudo-rapidity, and the central objects located within the rapidity gap.

Given the approximate orthogonality between the features of the central objects and those of the VBS jets, the multivariate discriminator is constructed using only the features of the central objects. Conversely, the VBS jet features are utilized to define the second arm of the ABCD regions. After comprehensive evaluations, a Boosted Decision Tree (BDT) using VBS jet features was developed instead of a single Vector Boson Fusion (VBF) feature. The BDT model, once established, is used to aid the training of a Deep Neural Network (DNN). This DNN is specifically trained to be orthogonal to the VBS BDT, incorporating a loss function with a decorrelation term as cited in [24], allowing the use of these two discriminators as the distinct arms for the ABCD method.

Both the DNN and the BDT are trained within the pre-selection phase space where MC statistics are more abundant, using background simulations against a $\kappa_{VV} = 2$ signal point. Post-training, the signal region optimization involves a thorough scan of the scores from the DNN, the VBS BDT, and the ParticleNetMD scores for the $H_{b\bar{b}}$ and $V_{q\bar{q}}$ candidate fatjets. The background estimation then proceeds by inverting the scores from the DNN and/or the VBS BDT to establish the ABCD anchor regions. Additional model training adjustments include applying Prefire weights, HEM corrections, pileup reweighting, b-tagging and lepton ID uncertainties, puID weights for jets, and jet energy corrections to ensure fidelity and accuracy of the simulated data.

4.4.1 VBS BDT

The Boosted Decision Tree (BDT) dedicated to the VBS jets analysis incorporates several highly discriminative inputs to effectively separate the signal from the background. Central to this analysis are the features $\Delta\eta_{jj}$ and m_{jj}, which represent the pseudorapidity gap and the invariant mass between the two VBS jets, respectively. These features are particularly valuable for their strong correlation with the VBS process characteristics. Additional kinematic features of the jets are also included to enhance the model's predictive accuracy.

The complete list of inputs to the VBS BDT is as follows:

- $\Delta\eta_{jj}$ — the pseudorapidity gap between the VBS jets.
- m_{jj} — the invariant mass of the VBS jets.
- p_T of the first VBS jet.
- η of the first VBS jet.
- p_T of the second VBS jet.
- η of the second VBS jet.

The input distributions for both the signal and the background MC, illustrating these variables, are depicted in Figure 4.9.

For the training of the BDT, the AdaBoost algorithm is employed with the settings "MaxDepth" equal to 2 and "NTrees" equal to 10. This choice reflects a balance between model complexity and overfitting, ensuring robustness and generalizability. The data set is divided evenly between the training and testing phases, with a 50% split, to optimize both the learning and validation of the model's performance.

The BDT scores for signal and background, and the ROC curve are shown in figure 4.10.

4.4.2 DNN

The Deep Neural Network (DNN) is configured to incorporate the Boosted Decision Tree (BDT) output as an additional feature, ensuring that the DNN remains orthogonal to the BDT scores. This design choice enhances the overall robustness and independence of the DNN in the analysis.

The inputs to the DNN include several key variables that are pivotal in characterizing the events:

- $H_{b\bar{b}}$ candidate ParticleNet regressed mass.
- $H_{b\bar{b}}$ candidate p_T.
- $V_{q\bar{q}}$ candidate ParticleNet regressed mass.
- $V_{q\bar{q}}$ candidate p_T.
- Lepton p_T.
- $m_{\ell,b}$ (Invariant mass of the lepton and b-jet).
- MET (or $p_{T,\text{miss}}$).

These inputs are visualized in Figures 4.11 and 4.12. It is noteworthy that the S_{T} variable is not explicitly used, as its components are already included as individual inputs, thus preserving the detailed information they provide.

To ensure effective training, the input features were standardized. The transverse momenta (p_T) of the objects are log-normalized ($p_T \rightarrow \log(p_T)$), and other variables are scaled according to the formula:

$$x \rightarrow \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4.2}$$

This normalization adjusts the range of a variable x so that the selected minimum (x_{min}) and maximum (x_{max}) values are scaled to approximately 0 and 1, respectively.

The architecture of the neural network comprises three hidden layers, each containing 64 nodes. A dual-component loss function is employed: the primary component is the binary cross-entropy for event classification, while the secondary component, weighted at 30% relative importance, ensures decorrelation with the VBS BDT output. Detailed explanations of this loss mechanism are provided in [24] and Appendix 8.4.

Training is conducted over 1000 epochs with a constant learning rate, using large batch sizes that divide the dataset into ten parts. The model state at epoch 500 is selected for use in the analysis as the loss function remains same after this point 4.13. The dataset is split into 80% for training and 20% for testing to balance learning and validation effectively.

The distribution of the DNN output, alongside the loss function and the Receiver Operating Characteristic (ROC) curve, are illustrated in Figure 4.13. These are presented for both the training and testing phases of the model. The figure specifically highlights the results at epoch 500, which has been selected for the final analysis based on its performance metrics.

At the conclusion of the training, the correlation between the VBS BDT and the DNN output is carefully examined. The results indicate that the correlation is effectively managed within the pre-selection phase space, as depicted in Figure 4.14.

4.4.3 Signal Region Optimization

The optimization of the signal region follows the training of the two multivariate discriminators within the preselection phase. Initially, a 1-dimensional scan of the $H_{b\bar{b}}$ and $V_{q\bar{q}}$ candidates' scores is performed to refine their selection thresholds independently. Based on the figure of merit $s/\sqrt{b}$ (signal over the square root of background), the working points are set with $H_{b\bar{b}}$ score greater than 0.5 and $V_{q\bar{q}}$ score greater than 0.7. The detailed cutflow leading up to the region used for the "ABCD" background estimate is documented in Table 4.7.

Following this, a 2-dimensional scan involving the BDT and DNN scores is conducted. This scan ranges from 0.5 to 1 in increments of 0.01. For each point, a simple datacard is produced to compare the signal yield and background predictions, which are made using the

"ABCD" method.

The final selection of the BDT and DNN scores was determined by identifying the combination that provided the highest limit, which ensures an optimal balance between signal sensitivity and systematic error minimization. This approach led to the selection of a BDT score of 0.56 and a DNN score of 0.92 as the cutoff values for the final signal region.

In summary, the criteria selected for defining the signal region are as follows:

- $H_{b\bar{b}}$ score > 0.5
- $V_{q\bar{q}}$ score > 0.7
- DNN > 0.92
- VBS BDT > 0.56.

Table 4.7. Semi-leptonic channel cutflow table from the preselection down to the ABCD region. Only cross-section reweighting, PU reweighting, lepton scale factors and AK4 b-tagging scale factors are applied. The sum of the data yield in the A+B+C+D region is kept blinded. Several reweighting factors are applied to signal MC and background MC, which account for the difference between this table and 4.6

Cut	Other	TTX	WJets	TTbar
$S_{\mathrm{T}} > 1000$	1645.6 ± 40.22	91.37 ± 0.57	4544.6 ± 11.34	5054.8 ± 14.87
$H_{b\bar{b}}$ score > 0.5	436.01 ± 13.40	73.48 ± 0.55	416.4 ± 3.56	3563.7 ± 13.55
$V_{q\bar{q}}$ score > 0.7	158.88 ± 7.38	25.954 ± 0.32	90.67 ± 1.63	710.14 ± 6.07
Cut	data	Bkg	Signal	
$S_{\mathrm{T}} > 1000$	12042 ± 109.74	11335 ± 44.34	11.14 ± 0.04	
$H_{b\bar{b}}$ score > 0.5	4292 ± 65.51	4489.6 ± 19.401	7.86 ± 0.04	
$V_{q\bar{q}}$ score > 0.7	—	985.63 ± 9.70	6.19 ± 0.03	

4.4.4 Background Estimate with the ABCD Method

The estimation of the background in the signal region is conducted entirely from data using the "ABCD" method. This method relies on defining four distinct regions: A, B, C, and D,

as illustrated in Figure 4.15 (left). The regions are strategically established by setting two cuts in the signal region (Region A) and then inversely applying these cuts to define Regions B, C, and D.

The background yield in Region A, designated as A_{pred}, is calculated using the counts from the other three regions with the formula:

$$A_{pred} = B \times \frac{C}{D}. \tag{4.3}$$

The actual data yields and the predicted data in Region A are detailed in Table 4.8. This approach allows for a data-driven estimation of the background, minimizing reliance on model-based assumptions and enhancing the robustness of the analysis.

Additionally, the statistical uncertainty associated with A_{pred} is determined by propagating the statistical uncertainties from the data yields in Regions B, C, and D. The equation for calculating this uncertainty is given by:

$$\varepsilon_{stat} = \sqrt{\left(\frac{\sqrt{B_{data}}}{B_{data}}\right)^2 + \left(\frac{\sqrt{C_{data}}}{C_{data}}\right)^2 + \left(\frac{\sqrt{D_{data}}}{D_{data}}\right)^2} \approx 31\% \tag{4.4}$$

This calculation quantifies the confidence level of the background estimate, ensuring that the statistical validity of the analysis is maintained.

Table 4.8. Data yields and region A prediction for the control region used for the ABCD closure test. The region A yield is kept blind, while A_{pred} is reported.

Region	data yield	prediction
region A	-	0.82
region B	76	
region C	11	
region D	1015	

The data yields for each region are graphically represented in Figure 4.16, where the

yields are plotted as a function of the VBS BDT score on the left and the DNN score on the right. In these plots, the yield from Region A is intentionally kept blind to maintain the integrity of the data analysis process until the final stages. Instead, the estimated yield for Region A (A_{pred}) is shown. This estimated yield is then compared to the corresponding Monte Carlo (MC) simulation to assess the accuracy and reliability of the background estimation. This comparison helps validate the effectiveness of the ABCD method in accurately predicting the background yields from purely data-driven approaches.

4.4.5 Closure Studies

To validate the robustness of our background estimation method, a closure study was performed in a phase space region defined by inverting the $V_{q\bar{q}}$ score cut ($V_{q\bar{q}}$ score < 0.7). To minimize signal contamination and maintain a background composition similar to that in the ABCD regions, the $H_{b\bar{b}}$ score was not inverted but restricted to $H_{b\bar{b}}$ score < 0.95. The ABCD' regions, analogous to the original ABCD method but within this control phase space, are defined using the same BDT and DNN selections applied in the signal region. These regions, labeled A', B', C', and D', are illustrated in Figure 4.15 (right). The data yields and the predictions for region A' are detailed in Table 4.9. The comparison between the predicted data yield (2.41) and the actual data yield (2) indicates a discrepancy, leading to an inferred systematic uncertainty of approximately 20% for the background estimate. This systematic uncertainty is currently being used in our sensitivity analyses. Despite no apparent reason for expecting a significantly different uncertainty value, further closure studies are planned to serve as a cross-check, ensuring the validity and precision of our background estimation approach.

To enhance the validation of our background estimation methodology, an additional cross-check using simulation data was conducted. This simulation-based closure test allows us to verify the consistency of the ABCD method in a controlled environment where the true values and distributions are known. The yields obtained from the simulation, along with the predictions for region A, are presented in Table 4.10. This approach provides an additional

Table 4.9. Data yields and region A prediction for the control region used for the ABCD closure test. The statistical uncertainty on the predicted value is reported in the table.

Region	data yield	prediction
region A'	2	2.20 ± 0.46
region B'	180	
region C'	28	
region D'	2295	

layer of confidence in the robustness of the ABCD method by comparing simulated data outputs directly against the expected predictions.

Table 4.10. MC yields and region A prediction the ABCD closure test. The statistical uncertainties reported in the table.

Region	MC yield	prediction
region A	1.65 ± 0.37	1.30 ± 0.15
region B	71.39 ± 3.59	
region C	16.39 ± 1.60	
region D	896.06 ± 8.86	

To further quantify the reliability of the ABCD method, we assess its closure in Monte Carlo simulations by comparing the actual MC yields, A_{MC}, with the predicted MC yields, A_{MC}^{pred}. The systematic uncertainty derived from this comparison is calculated using the following formula:

$$\varepsilon_{syst} = \frac{2}{A_{MC} + A_{MC}^{pred}} \sqrt{\left(\frac{A_{MC}^{err}}{A_{MC}}\right)^2 + \left(\frac{\varepsilon_{stat}^{MC}}{A_{MC}^{pred}}\right)^2} \approx 17\% \tag{4.5}$$

where A_{MC}^{err} represents the statistical error on the MC yield and ε_{stat}^{MC} denotes the statistical uncertainty on the prediction originating from the B, C, and D regions.

we apply a conservative 20% systematic in the dissertation

This ongoing evaluation process underscores our commitment to ensuring the accuracy and reliability of our background estimation method before its final application in the sensitivity

study. Further updates and final decisions will be thoroughly documented in subsequent revisions of this document.

Chapter 4, in full, is a reprint of the material as it appears in CMS analysis note AN-23-016, Search for anomalous c2v couplings in the VVH production via vector boson scattering. The dissertation author was the primary investigator and author of this analysis note.

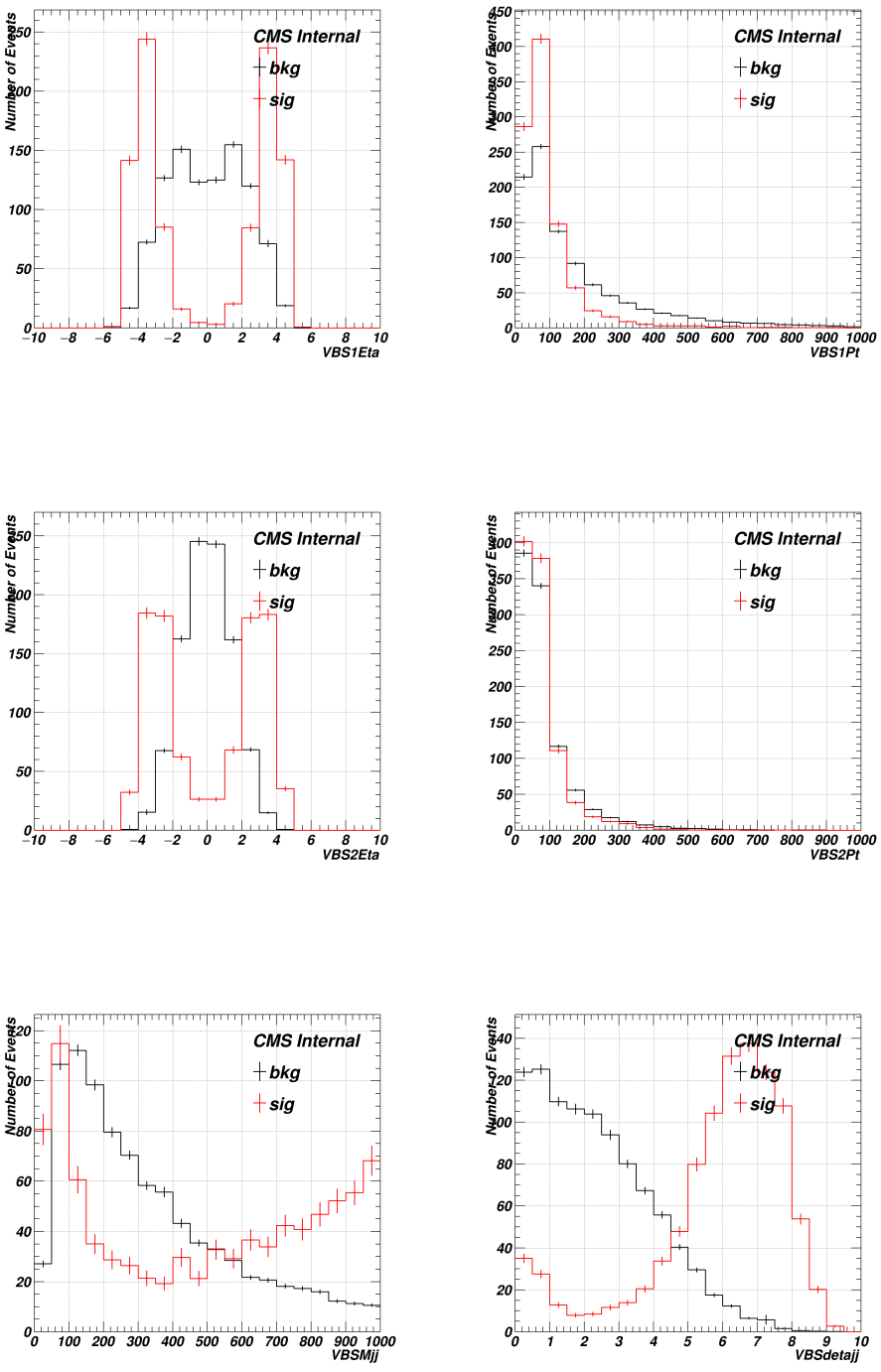


Figure 4.9. Inputs of the VBS BDT. Signal and background are both normalized to 1.

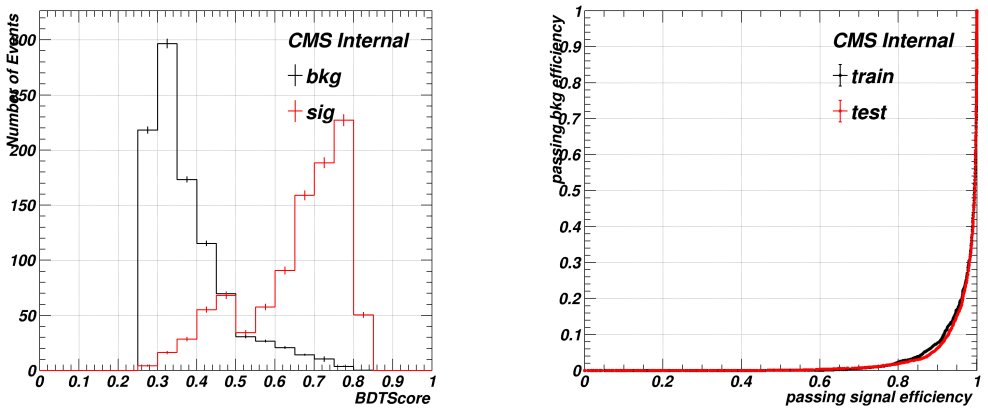


Figure 4.10. VBS BDT output distribution and ROC curve for the test and training sample.

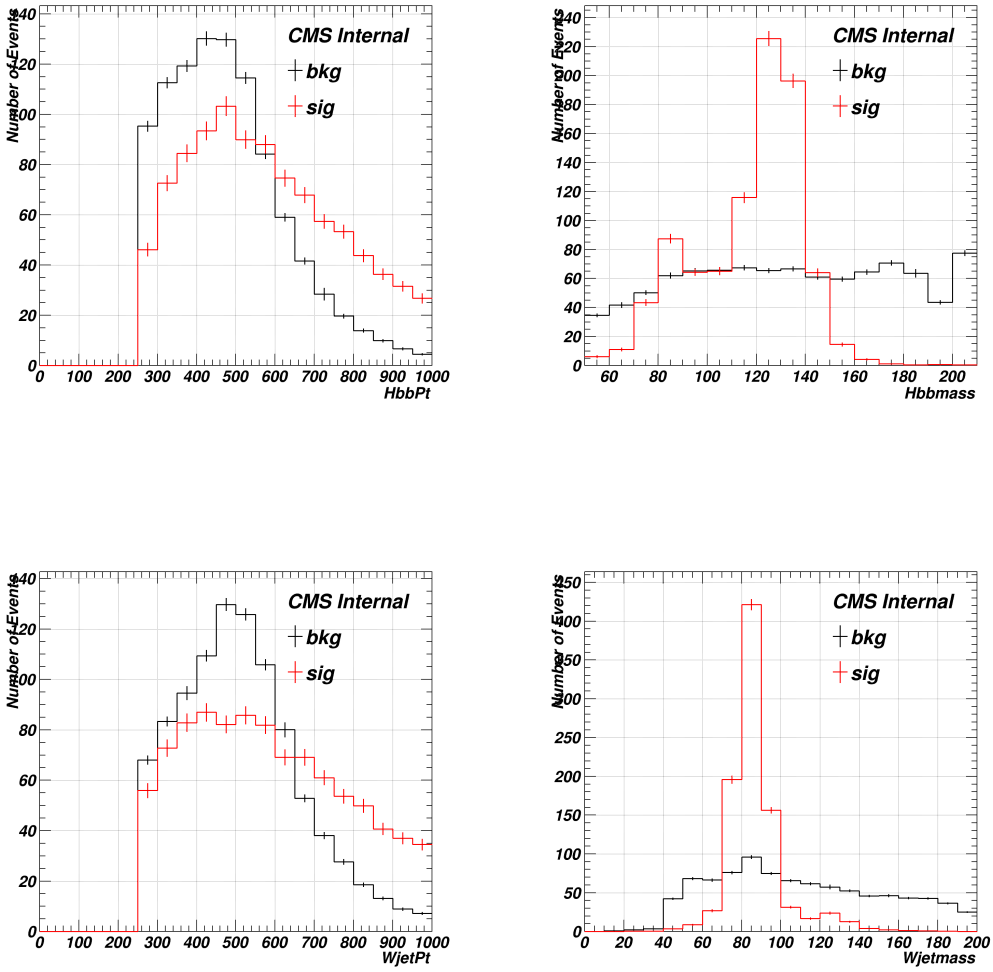


Figure 4.11. Signal and background distributions of the DNN inputs. The signal and background distributions are normalized to 1.

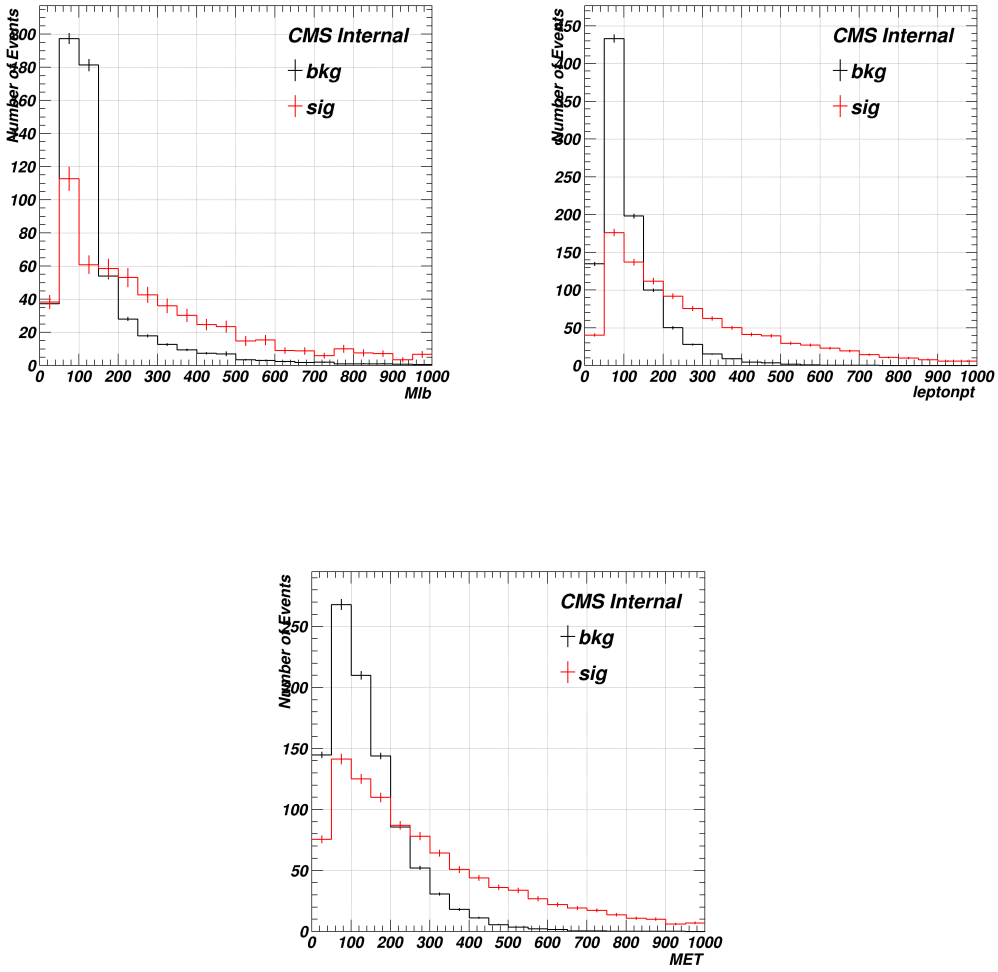


Figure 4.12. Signal and background distributions of the DNN inputs. The signal and background distributions are normalized to 1.

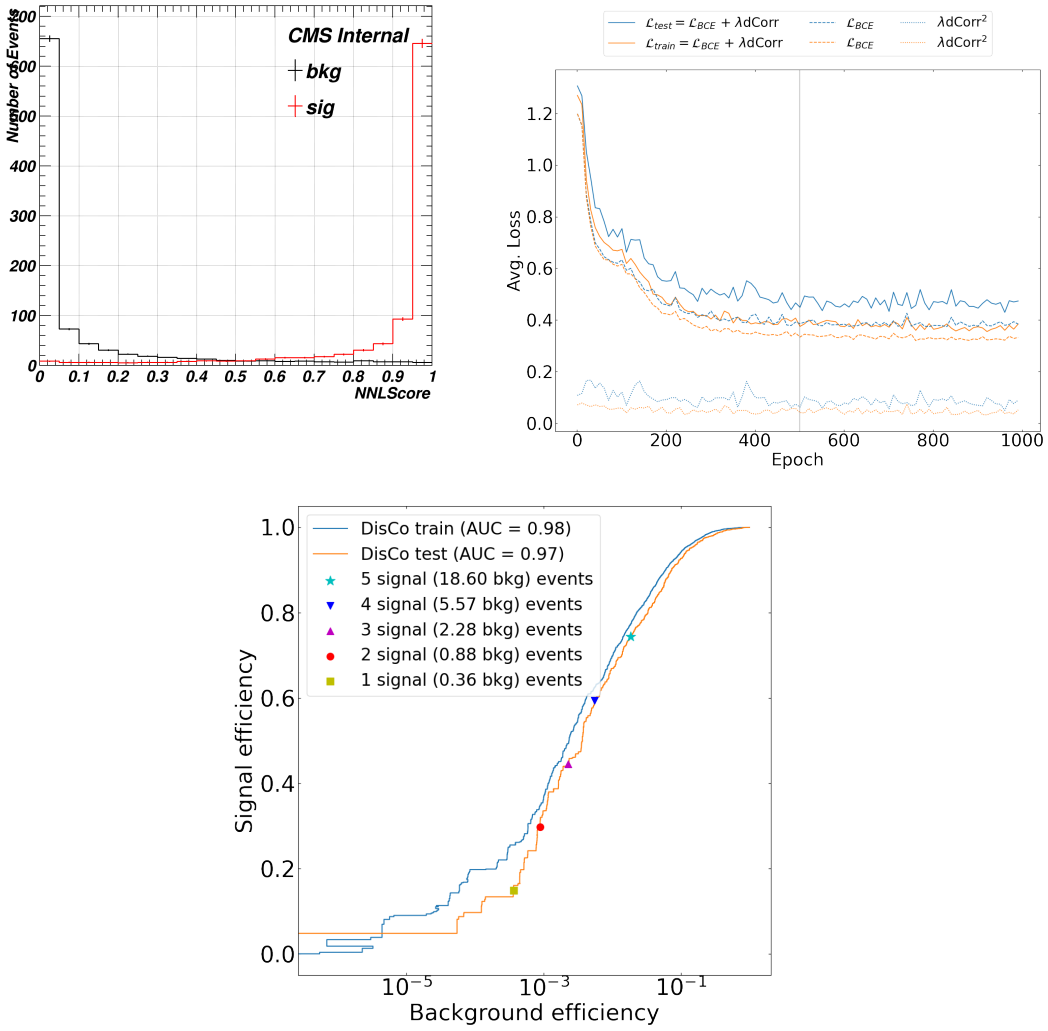


Figure 4.13. DNN output distribution and loss function the test and training sample. ROC curve for the epoch 500 of the training. The epoch 500 was chosen for the analysis.

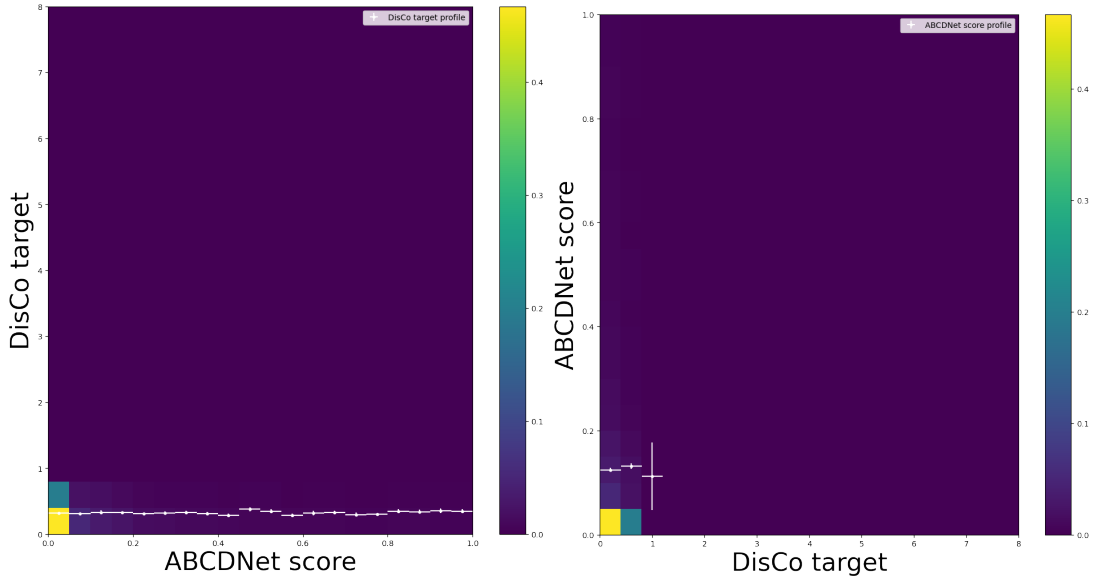


Figure 4.14. Cross-check of the correlation between the VBS BDT score and the DNN output score.

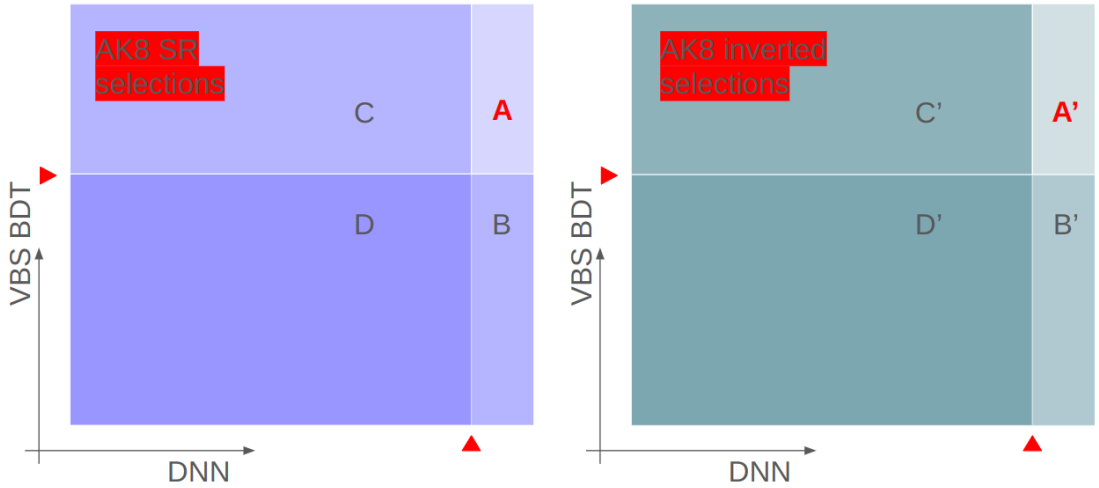


Figure 4.15. schematic representation of the regions used for the ABCD background estimate (left) and the control regions used for the cross-check.

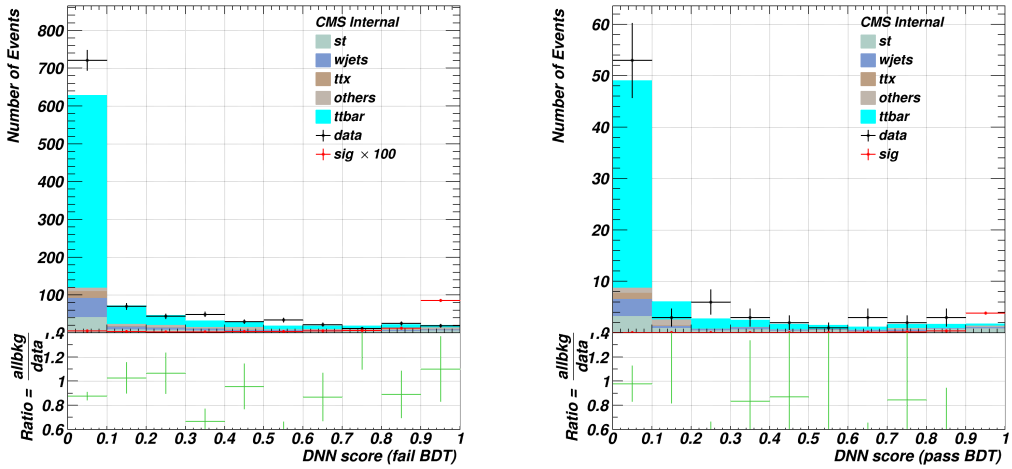


Figure 4.16. Regions B and D are plotted on the left, and regions A and C are plotted on the right.

Chapter 5

Systematic uncertainties

Most sources of systematic uncertainty are evaluated solely for the signal sample, as the background estimate is derived from data. The uncertainties associated with the background yield in the signal region are computed in conjunction with the background estimation process.

Given that our analysis primarily involves a counting experiment, the systematic effects on the shape of signal distributions do not need to be considered. Instead, the variations primarily affect the signal yield and are incorporated either through adjustments in event or object weights, or by recalculating the acceptance of a selection under varied conditions.

Systematic uncertainties are typically quantified by varying individual theoretical scales and experimental corrections (listed in Table 5.5) by one standard deviation. The maximal difference in yield is then taken as the systematic error. These corrections and their associated uncertainties are usually derived centrally to enhance the efficiency of specific selections in the Monte Carlo simulations to match those observed in data.

Corrections are generally applied as an event weight w, whereby the weighted contribution W of each raw Monte Carlo event is the product of the event weights for that event. The yield y in a given signal region, containing N raw Monte Carlo events, is calculated as follows:

$$y = \sum_{i=1}^{N} W_i \tag{5.1}$$

The yield under a systematic variation y_{var} is then determined by adjusting each event weight W_i

to reflect the systematic variation (either up or down):

$$y_{var} = \sum_{i=1}^{N} W_i \times \frac{w_{var}}{w} \tag{5.2}$$

The maximum of the percentage differences, either δ_{up} or δ_{down}, is then adopted as the systematic uncertainty for that particular source:

$$\delta_{var} = \left| 1 - \frac{y_{var}}{y} \right| \tag{5.3}$$

A concise summary of the methods employed to assess each source of systematic uncertainty is provided below:

5.1 PDF Variations

The uncertainty associated with the Parton Distribution Function (PDF) set used in the signal Monte Carlo (MC) generation is assessed using 100 parameter variations from the Hessian PDF set. The PDF uncertainty quantifies the impact of these variations on the model acceptance and is estimated as the standard deviation of the acceptance for each variation [14]. This estimation is formulated as follows:

$$\delta^{PDF} = \sqrt{\sum_{i=1}^{N_{PDF}} \left(1 - \frac{y_i^{var}}{y^{nom}} \right)^2} \tag{5.4}$$

Here, y_i^{var} represents the acceptance under the i-th PDF variation, calculated by adjusting the event weights accordingly:

$$y_i^{var} = \sum_{j=1}^{N_{Events}} w_j \times \frac{w_{i,j}^{PDF}}{R_i} \tag{5.5}$$

The ratio R_i, critical for normalizing the PDF weights against the baseline MadGraph

generator weights for each variation, is defined as:

$$R_i = \frac{\sum_{j=1}^{N_{Events}} w_{i,j}^{PDF}}{\sum_{j=1}^{N_{Events}} w_j^{gen}} \tag{5.6}$$

This method ensures that the PDF variations are systematically and accurately reflected in the uncertainty estimates, providing a robust framework for understanding their impact on the theoretical predictions of the MC simulation.

5.2 muonF Scale and Parton Shower ISR and FSR

The uncertainties associated with the factorization (μ_F) and renormalization (μ_R) scales are derived from the variations specified in NanoAODv9. These scales are varied from 0.5 to 2 times their nominal values, with the following constraint imposed to maintain theoretical consistency:

$$0.5 \leq \frac{\mu_F}{\mu_R} \leq 2$$

For our signal sample, which is generated at Leading Order (LO), the variation in μ_R is determined to be 0.

Additionally, the uncertainties for the initial and final state radiation (ISR and FSR) from the parton shower are assessed using variations provided by Pythia8, also available in NanoAOD. These ISR and FSR scales are independently varied from 0.5 to 2 times their default values.

5.3 Lepton Uncertainties

As detailed in Section 3.1.1, lepton scale factors adhere to a factorization prescription [26,31]. The systematic uncertainties for each scale factor are thoroughly evaluated. The specific repositories for the scale factors and their sources of uncertainties are listed [34, 35].

For electron scale factors concerning reconstruction efficiency and ”Loose ID”, we follow the recommendations of the EGamma group [25], calculating uncertainties accordingly.

Similarly, muon scale factors for reconstruction efficiency and ”Loose ID” are applied as per the guidelines from the Muon POG [28–30], with uncertainties computed in alignment with these standards.

The scale factors for the ttH ID are divided into two categories: ”IP+ISO” and ”Tight ID”. Since our analysis requires that selected leptons pass the tight level lepton ID, both levels of scale factors are applied. The application and uncertainty computation for these factors are conducted following the methodology used in the analysis TOP-22-006.

Lastly, High-Level Trigger (HLT) scale factors are applied. For electrons, these factors are centrally derived, while for muons, they are sourced from SMP-19-002, specifically for the muon HLT paths.

5.4 ParticleNet Scores

The uncertainties in the $H \to b\bar{b}$ score are derived using the gluon splitting proxy method detailed in [1]. This method was originally developed for the $H \to c\bar{c}$ and $H \to b\bar{b}$ analyses. It employs a Boosted Decision Tree (BDT), referred to as the “sfBDT,” to isolate the phase space populated by $H \to b\bar{b}$ jets in the signal, subsequently selecting $g \to b\bar{b}$ jets within that phase space from Monte Carlo simulations of QCD multijet events.

The sfBDT is trained to select suitable $g \to b\bar{b}$ jets to serve as proxies for the $H \to b\bar{b}$ jets in the signal, specifically by vetoing jets with a high gluon contamination rate. Additionally, the sfBDT utilizes input variables involving the basic kinematics of the subjets and the secondary vertices associated with the jet. Once trained, the sfBDT can also be applied to select similar proxy jets from data.

The signal-like jets selected by the sfBDT are then used to measure the efficiency of the ParticleNet Xbb discriminant in QCD Monte Carlo simulations and data. The method is signal-dependent, as it models the gluon splitting proxy, irrespective of the working points.

“Pass” and “fail” regions are defined based on jets passing or failing the ParticleNet

Xbb tagging thresholds. In each region, a fit of the mass of the secondary vertex with the maximum impact parameter d_{xy} significance is performed to distinguish the contributions of b-type ($g \to b\bar{b}$), c-type, and light-type jets. Three scale factors, one for each jet type, are allowed to float in this fit, defined as follows:

$$\mathrm{SF}_i = \frac{\varepsilon_{\mathrm{data},i}}{\varepsilon_{\mathrm{MC},i}} \tag{5.7}$$

where i represents the jet category and ε denotes the efficiency of the ParticleNet tagger.

This measurement is conducted in bins of p_T for a more robust correction and is repeated for each year of UL NanoAOD individually. The bins and working points vary by channel. The scale factor summaries for the 1-lepton channel is reported in Tables 5.1.

Table 5.1. Summary of the ParticleNet Xbb-tagging scale factors for the Xbb > 0.5 working point used in the 1-lepton channel.

Year	P_T range in GeV:				
	[250, 350)	[350, 450)	[450, 550)	[450, ∞)	[450, ∞)
2018	$1.003^{+0.067}_{-0.083}$	$1.172^{+0.180}_{-0.185}$	$1.150^{+0.121}_{-0.095}$	$0.977^{+0.117}_{-0.116}$	$0.998^{+0.168}_{-0.162}$
2017	$1.213^{+0.212}_{-0.201}$	$1.100^{+0.275}_{-0.271}$	$1.216^{+0.192}_{-0.178}$	$1.192^{+0.201}_{-0.188}$	$1.081^{+0.242}_{-0.237}$
2016 (post-VFP)	$1.071^{+0.211}_{-0.174}$	$1.104^{+0.180}_{-0.165}$	$1.163^{+0.155}_{-0.142}$	$1.377^{+0.255}_{-0.240}$	$1.338^{+0.232}_{-0.181}$
2016 (pre-VFP)	$1.131^{+0.160}_{-0.138}$	$1.356^{+0.192}_{-0.148}$	$1.374^{+0.223}_{-0.206}$	$1.376^{+0.242}_{-0.216}$	$1.360^{+0.299}_{-0.257}$

The $\mathrm{V} \to q\bar{q}$ score is calibrated using the method described in [2]. This method employs top pair production events to implement a tag-and-probe method. The tag is a muon (from the semileptonic channel), while the probe is an AK8 jet. AK8 jets are categorized into different templates (merged top, merged W, non-merged top products, others). A fit is performed to the AK8 jet softdrop mass in the "pass" and "fail" regions, defined using the $\mathrm{V} \to q\bar{q}$ tagger working point of interest. Scale factors for each template are derived from the fit, and the merged W scale factor with its uncertainty is used for the analysis to correct the signal.

The results were first reproduced and then recomputed for our working points. The

results for the 0.7 working point, used in the 1-lepton channel, is reported below in Table 5.2.

Table 5.2. Summary of the ParticleNet $V_{q\bar{q}}$-tagging scale factors for the working point used in the analysis.

Year	P_T range in GeV:			
	$[200,\ 300)$	$[300,\ 400)$	$[400, 800)$	$[800\ \infty)$
2016 APV WP 0.7	$0.92^{+0.04}_{-0.04}$	$0.93^{+0.03}_{-0.04}$	$0.94^{+0.06}_{-0.06}$	$0.94^{+0.12}_{-0.12}$
2016 WP 0.7	$0.96^{+0.05}_{-0.05}$	$0.89^{+0.04}_{-0.04}$	$0.85^{+0.07}_{-0.06}$	$0.85^{+0.14}_{-0.12}$
2017 WP 0.7	$0.96^{+0.02}_{-0.02}$	$0.94^{+0.02}_{-0.02}$	$0.93^{+0.04}_{-0.04}$	$0.93^{+0.08}_{-0.08}$
2018 WP 0.7	$0.91^{+0.02}_{-0.02}$	$0.90^{+0.02}_{-0.02}$	$0.85^{+0.04}_{-0.04}$	$0.85^{+0.08}_{-0.08}$

5.5 ParticleNet Mass Regression

The ParticleNet mass regression has been calibrated from the Hcc and Hbb processes [17]. The regression results are consistent with the data in terms of mass scale. A 3% smearing was applied to align the Monte Carlo (MC) resolution with the data. This full smearing is considered as the uncertainty in the measurements.

For this analysis, we followed the same calibration procedure to derive mass scale and resolution corrections, with the associated uncertainties. These corrections were derived for AK8 jets using the method described in [2], where scale and resolution corrections are treated as nuisance parameters. We performed the fit using the regressed mass instead of the soft drop mass and the fit was conducted inclusively across the transverse momentum (p_T) spectrum for AK8 jets. The working point for the "pass" and "fail" determination was set to 0.7. However, when adjusting the working point to 0.8, there's no significant variations.

The calibration results are presented in Tables 5.3 and 5.4. The reported values represent the variations of the nuisance parameters extracted from the fits ($X_{central}$) and a conservative estimate of the uncertainties (X_{up}, X_{down}), rather than the direct outputs of the fit. For the Jet Mass Resolution (JMR), the upward uncertainty was increased to 30%, based on studies performed in

bins of p_T, while the downward uncertainty was set manually as per the previous nominal value. The 2018 Jet Mass Scale (JMS) uncertainty was also increased to 0.1, whereas for other years, the full scale variation was taken as the uncertainty.

The JMS variation must be applied in addition to a 5% factor, as this represents the nuisance parameter used in the fit. The total correction is approximately 1%. The JMR variation is applied on top of a 10% smearing, which is the nominal nuisance parameter. The total effect results in a smearing of approximately 1%, with an upward uncertainty of 3-4% and no smearing as the downward variation.

Table 5.3. JMS Scale Factors

JMS	$X_{central}$	X_{up}	X_{down}
2016 APV	-0.3	0	-0.6
2016	-0.24	0	-0.48
2017	-0.11	0	-0.22
2018	-0.04	0.06	-0.14

Table 5.4. JMR Scale Factors

JMR	$X_{central}$	X_{up}	X_{down}
2016 APV	0.13	0.43	0
2016	0.15	0.45	0
2017	0.09	0.39	0
2018	0.10	0.40	0

5.6 JME Uncertainties

The evaluation of uncertainties related to jet energy corrections (JECs) for both AK4 and AK8 jets follows the protocols and utilizes tools provided by the JETMET group. For the jet energy scale (JES), we employ a reduced set of uncertainties as recommended by [5]. This approach involves decorrelating uncertainties across different years and data-taking periods,

with specific distinctions made for 2016 preVFP and postVFP. Uncertainties recommended for year-specific separation are treated as such, while others remain correlated across all periods.

For AK4 jets, uncertainties are also propagated to the calculation of the Missing Transverse Energy (MET). In addition, we apply Jet Energy Resolution (JER) corrections along with their relative uncertainties specifically for AK4 jets. It is important to note that JER uncertainties are not propagated to the MET and do not apply to AK8 jets.

Furthermore, we incorporate systematic uncertainties on the MET through "unclustered MET" variations, both upwards and downwards. This systematic approach ensures a comprehensive assessment of uncertainties impacting both jet measurements and MET calculations.

5.7 B-tagging Uncertainties

In the 1-lepton analysis, Tight (T) AK4 jets are vetoed, and the Loose (L) working point is utilized to construct the $m_{\ell b}$ variable. We apply the recommended b-tagging scale factors for these fixed working points [3, 4].

For each event passing the selection criteria, a weight is assigned depending on the b-tagging status of AK4 jets within the event. The calculation of this weight takes into account each jet's tagging status during the b-tag veto step and those passing the Loose working point but not the Tight, as follows:

$$w = \frac{\prod_{i=tagged\ T}^{checked-jets} SF_i^T \varepsilon_i^T \prod_{j=tagged\ L,\ not\ T}^{checked-jets} (SF_j^L \varepsilon_j^L - SF_j^T \varepsilon_j^T) \prod_{k=not-tagged\ L}^{checked-jets} (1 - SF_k^L \varepsilon_k^L)}{\prod_{i=tagged\ T}^{checked-jets} \varepsilon_i^T \prod_{j=tagged\ L,\ not\ T}^{checked-jets} (\varepsilon_j^L - \varepsilon_j^T) \prod_{k=not-tagged\ L}^{checked-jets} (1 - \varepsilon_k^L)} \tag{5.8}$$

Here, SF represents the scale factor for each AK4 jet based on p_T, $|\eta|$, and the working point, and ε is the b-tagging efficiency for the sample at the specified working point (L or T), defined as:

$$\varepsilon(p_T, |\eta|, WP) = \frac{N_b^{b-tagged,WP}(p_T, |\eta|)}{N_b^{total}(p_T, |\eta|)} \tag{5.9}$$

Efficiency maps for the main background and signal samples for both Loose and Tight b-tagging working points have been privately calculated and are detailed in Appendix 8.7. Systematic uncertainties are assessed by varying the scale factors up and down as provided by the BTV group. Moreover, scale factors are decorrelated between b/c-jets and light jets to reflect their distinct behaviors.

5.8 Other Experimental Uncertainties (PU, Luminosity, L1 Prefiring)

5.8.1 Luminosity Uncertainty

The uncertainty associated with the luminosity measurement is adopted from the recommendations provided by the Pdmv group [27]. This assessment is crucial as it directly influences the scale of all measured cross sections and thus the overall normalization of signal and background processes.

5.8.2 Pile-Up Uncertainty

The uncertainty on the number of proton-proton (pp) interactions per event, known as pile-up (PU), is calculated by varying the total inelastic pp cross-section by 4.6%. This adjustment follows the guidelines set forth by the Luminosity POG [33], ensuring that the simulation accurately reflects variations observed in actual collision data.

5.8.3 L1 Prefiring Uncertainty

The uncertainties related to the Level 1 (L1) Prefiring weights are determined by varying these weights according to their respective uncertainties [32]. L1 Prefiring typically affects data from earlier runs of the LHC where older detector components had a higher likelihood of

triggering before the actual collision event. Adjusting for this effect is essential for accurate data interpretation, particularly for analyses involving data from the affected periods.

These uncertainties are integral to ensuring the reliability of our results, as they address key aspects of data collection and processing that can significantly impact the final analysis outcomes.

Table 5.5. All corrections applied to the MC used in this analysis for the 1-lepton channel. The source of the correction is also listed, along with whether the correction is used for signal or background MC. Notably, the only correction that is not applied to both signal and background is for the ParticleNet Xbb-tagging efficiency, as the correction used is valid only for real H(bb) fat jets.

Systematic source	1-lepton channel
Pileup reweighting	✓
Pileup jet ID	✓
L1 pre-fire corrections	✓
Electron trigger scale factors	✓
Muon trigger scale factors	✓
Electron reco scale factors	✓
Electron reco-to-Loose scale factors	✓
Electron Loose-to-(IP+ISO) scale factors	✓
Electron (IP+ISO)-to-tight ID scale factors	✓
Muon reco-to-Loose scale factors	✓
Muon Loose-to-(IP+ISO) scale factors	✓
Muon (IP+ISO)-to-tight ID scale factors	✓
ParticleNet Xbb scale factors	✓
ParticleNet W MD scale factors	✓
ParticleNet mass JMS/JMR	✓
DeepJet b-tagging scale factors (b/c jets)	✓
DeepJet b-tagging scale factors (light jets)	✓
MET unclustered	✓
Jet energy scale	✓
Jet energy resolution	✓

Chapter 5, in full, is a reprint of the material as it appears in CMS analysis note AN-23-016, Search for anomalous c2v couplings in the VVH production via vector boson scattering. The dissertation author was the primary investigator and author of this analysis note.

Table 5.6. Values of the systematics for the 1-lepton analysis. The jet energy scale systematics and the sum of multiple systematics (from different sources and data-taking periods) in quadrature. Similarly, the ParticleNet tagger systematics are the sum of the systematics for each data-taking period in quadrature.

Systematic source	1-lepton channel
PDF variations	1.59
μ_F scale	23.95
Parton shower ISR weights	0.64
Parton shower FSR weights	2.97
Pileup reweighting	1.52
Pileup jet ID	0.04
L1 pre-fire corrections	1.02
Electron trigger scale factors	0.56
Muon trigger scale factors	0.23
Electron reco scale factors	0.28
Electron reco-to-Loose scale factors	0.08
Electron Loose-to-(IP+ISO) scale factors	0.12
Electron (IP+ISO)-to-tight ID scale factors	0.44
Muon reco-to-Loose scale factors	0.03
Muon Loose-to-(IP+ISO) scale factors	0.09
Muon (IP+ISO)-to-tight ID scale factors	0.19
ParticleNet Xbb scale factors	2.09 - 6.22
ParticleNet W MD scale factors	1.29 - 2.42
ParticleNet mass JMS/JMR	0.05/0.1
DeepJet b-tagging scale factors (b/c jets)	0.04
DeepJet b-tagging scale factors (light jets)	0.03
MET unclustered	0.11
Jet energy scale	0.02 -3.96
Jet energy resolution	0.99
Luminosity	1.60

Chapter 6

Statistical Analysis

For the 1-lepton channel, we fit the signal and predicted background yield in the signal region to data. The B, C, and D region data yields are included in the fit together with their respective signal contamination.

6.1 One Dimensional Limit Scan

We set upper limits at 95% confidence level on the cross-section of the VBS VVH process for each κ_{VV} generated data point.

The upper limits are calculated using the AsymptoticLimits method in the Combine toolkit [20]. The full list of systematic uncertainties in Table 5.6 are included as nuisance parameters during the fitting procedure.

The result of this fit is shown in Figure 6.1 for the 1-lepton channel. The regions where the expected limit is smaller than the theoretical prediction on the cross section are taken to be excluded values for κ_{VV}.

The impact plots for the 1-lepton channel fit are shown in Figure 6.2.

6.2 Impact Analysis of Systematic Uncertainties

To calculate the impacts of systematic uncertainties on the measurement of the signal strength, we performed the following steps using the Combine toolkit. The impact plots were

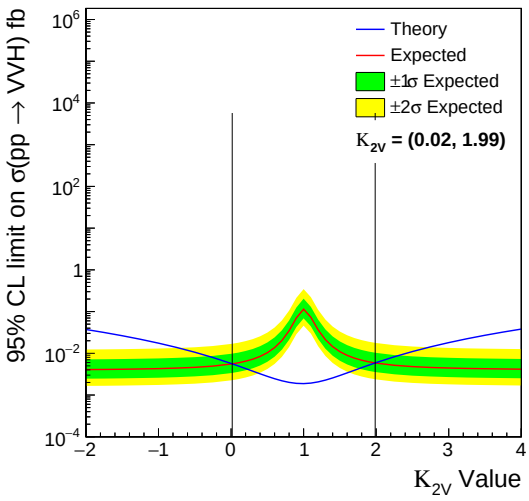


Figure 6.1. 95% confidence level limit as a function of κ_{VV} for the 1-lepton channel.

generated for $C2V = 1, 2$, and 4, both for signal = 1 and signal = 0, utilizing the Asimov dataset $(-t-1)$. Additionally, impact plots were created using the predicted background in region A, determined by $A = \frac{B \times C}{D}$, while keeping the data yield in region A blind.

The Combine results are shown as follows:

Chapter 6, in full, is a reprint of the material as it appears in CMS analysis note AN-23-016, Search for anomalous c2v couplings in the VVH production via vector boson scattering. The dissertation author was the primary investigator and author of this analysis note.

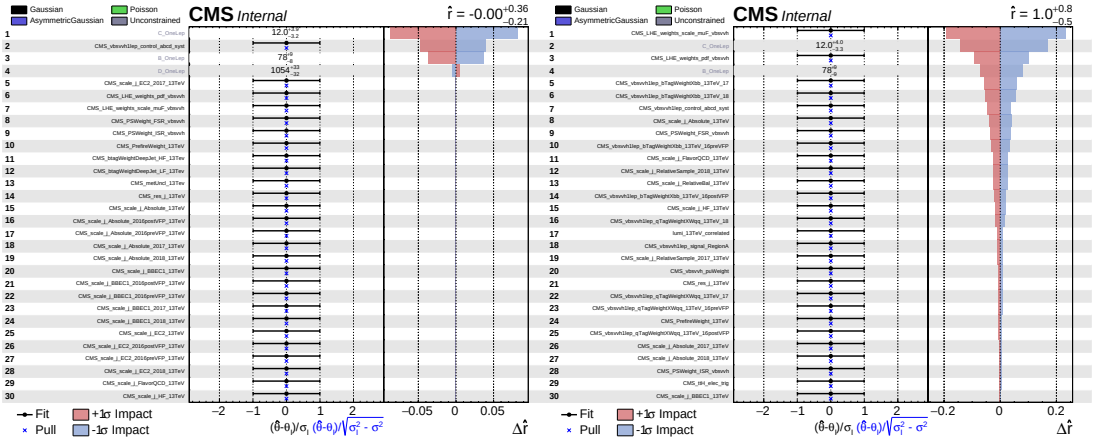


Figure 6.2. Impact plots for the 1-lepton channel fit in the background only hypothesis (left) and in the signal+background hypothesis (right).

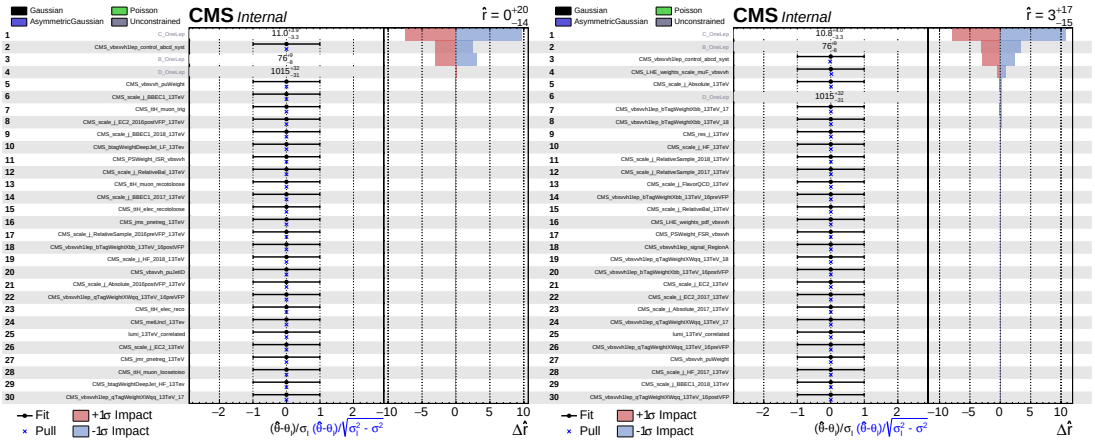


Figure 6.3. Impact plots for $C2V = 1$ with signal strength set to 0. The left plot uses the Asimov dataset ($-t-1$), while the right plot uses the predicted background without the Asimov dataset.

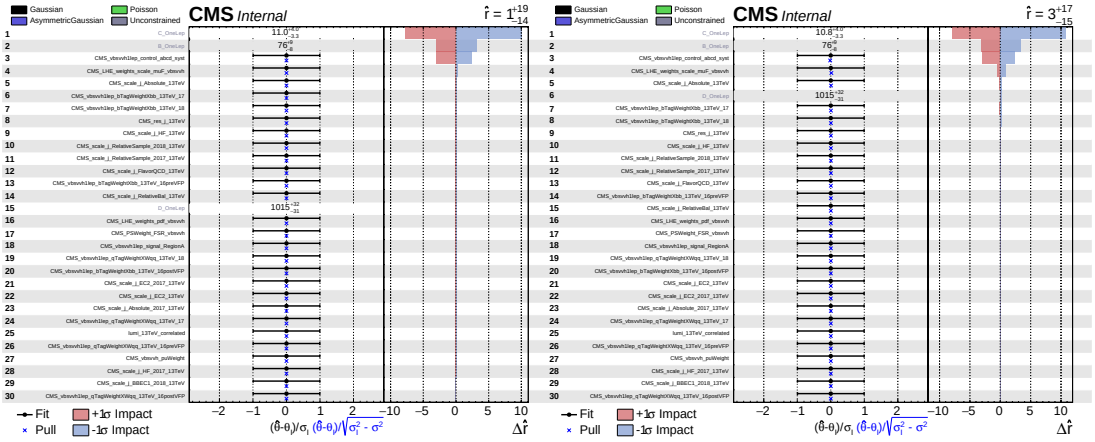


Figure 6.4. Impact plots for $C2V = 1$ with signal strength set to 1. The left plot uses the Asimov dataset ($-t-1$), while the right plot uses the predicted background without the Asimov dataset.

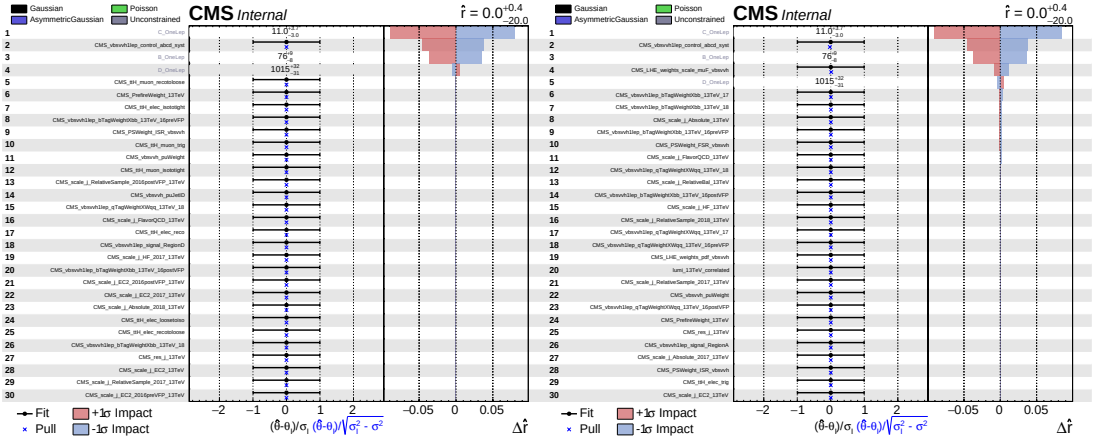


Figure 6.5. Impact plots for $C2V = 2$ with signal strength set to 0. The left plot uses the Asimov dataset ($-t-1$), while the right plot uses the predicted background without the Asimov dataset.

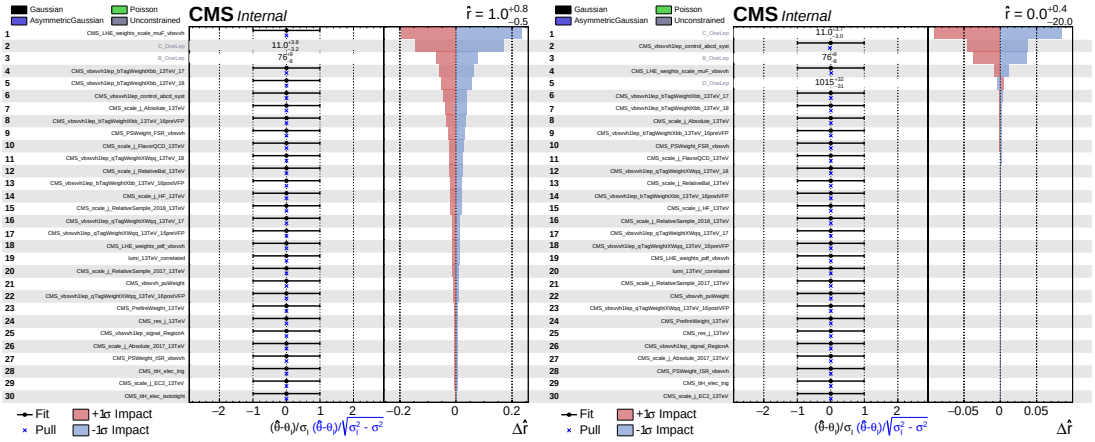


Figure 6.6. Impact plots for $C2V = 2$ with signal strength set to 1. The left plot uses the Asimov dataset ($-t-1$), while the right plot uses the predicted background without the Asimov dataset.

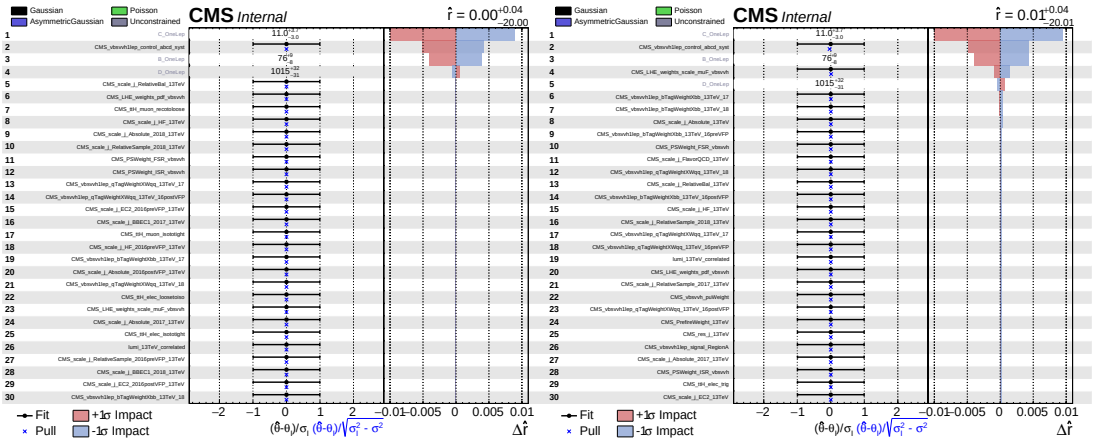


Figure 6.7. Impact plots for $C2V = 4$ with signal strength set to 0. The left plot uses the Asimov dataset ($-t-1$), while the right plot uses the predicted background without the Asimov dataset.

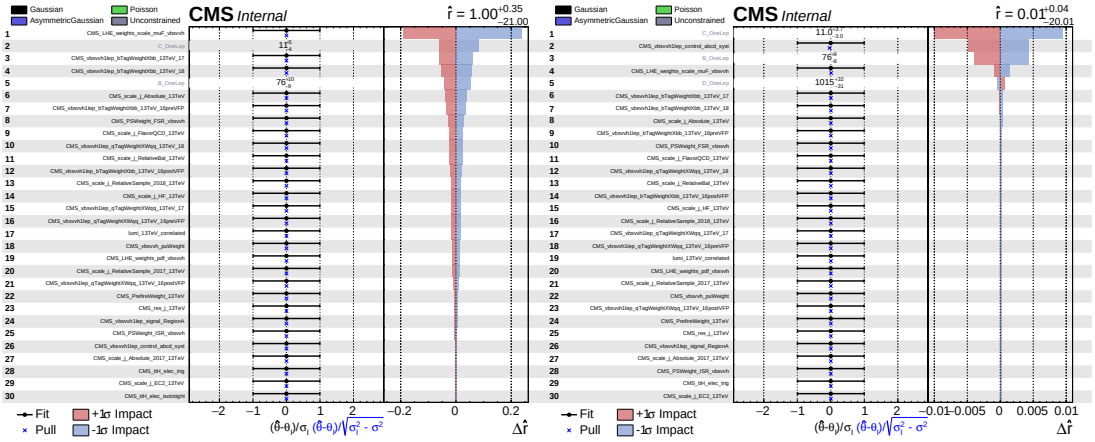


Figure 6.8. Impact plots for $C2V = 4$ with signal strength set to 1. The left plot uses the Asimov dataset ($-t-1$), while the right plot uses the predicted background without the Asimov dataset.

Chapter 7

Result

The below table shows the data yields with unblinded region A in the ABCD method:

Table 7.1. Data yields with unblinded region A.

Region	data yield	prediction
region A	1	0.82
region B	76	
region C	11	
region D	1015	

The final steps for unblinding is to run the Combine with the observed data and compare it with the unblinded ones. The results are shown in the following two plots.

The main difference between the observed and expected results arises from the 0.82 prediction versus 1 data in the A region. This discrepancy has resulted in slightly worse limits in the observed case, but the factor is very small.

The analysis is intended for publication in conjunction with other channels (HIG-24-003). The combined result will be enhanced by a factor of $\sqrt{2}$ through the inclusion of the 0-lepton and other channels, and it will be published by the CMS collaboration.

Chapter 7, in full, is a reprint of the material as it appears in CMS analysis note AN-23-016, Search for anomalous c2v couplings in the VVH production via vector boson scattering. The dissertation author was the primary investigator and author of this analysis note.

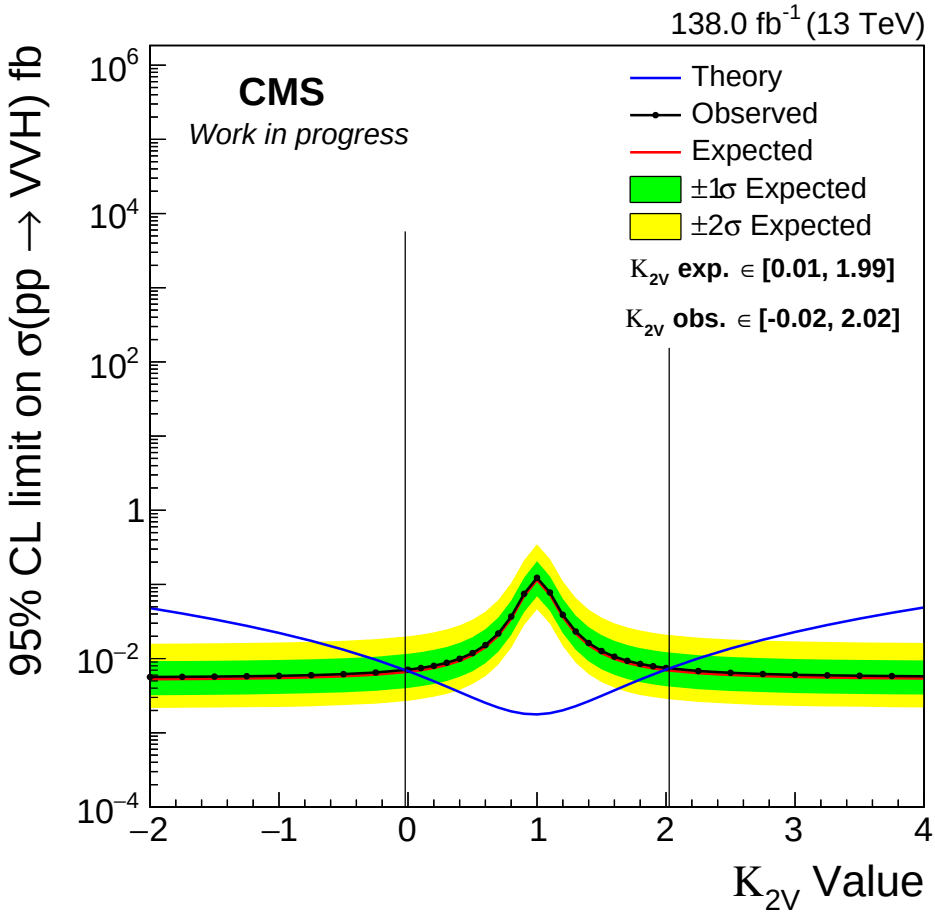


Figure 7.1. Observed and expected limits on the cross-section of the VBS VVH process. This plot includes both the expected limit and the observed limit, run without the `-t -1` option.

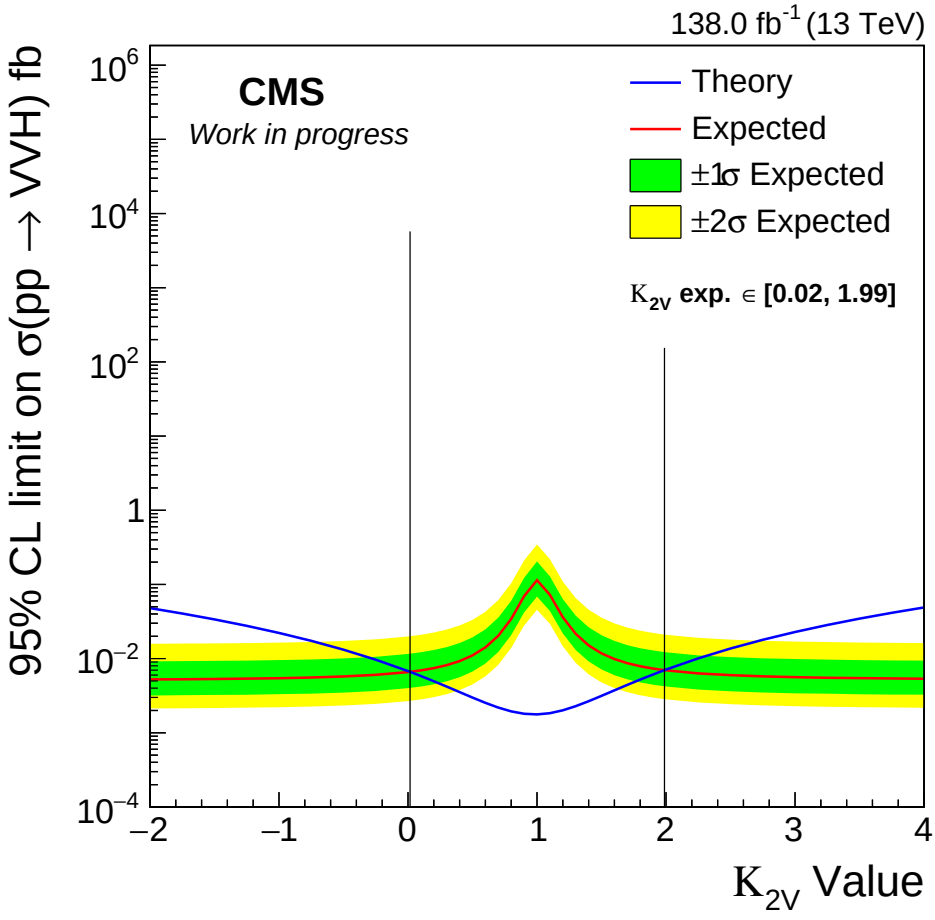


Figure 7.2. Expected limits on the cross-section of the VBS VVH process. This plot includes only the expected limit and was run with the `--run blind` option.

Chapter 8

Appendix

8.1 Signal MadGraph model

The Madgraph model is used to generate MC samples:

```
import model HHVBF_UFO-vbf_hh_4f


generate p p > w+ w- h j j QCD=0


output VBSWWH_OS_VBSCuts_4f_LO -nojpeg
```

Finally, Pythia was used to handle parton showers with the same settings used to generated other CMS MC samples at LO.

8.2 Signal process differential distributions

As the C_{2V} value deviates from the Standard Model (SM) expected value of 1, the anomalous behavior at the high-energy tail becomes more pronounced. The events in these tails exhibit distinct characteristics that can be used to discriminate against other SM background processes, thereby increasing sensitivity towards the search for anomalous coupling behaviors. This section documents the characteristics of the signal process and highlights the features that will be later used to distinguish the signal process from other SM background processes within

the signal region.

The process begins with two quarks from each incoming proton radiating W or Z bosons, which then scatter off while also producing a Higgs boson in the process. The outgoing two quarks, reconstructed as jets, are known as vector boson fusion (VBF) or, in this case, vector boson scattering (VBS) jets. VBS jets are characterized by a large invariant mass (m_{jj}) and pseudorapidity separation ($|\Delta\eta_{jj}|$).

As the C_{2V} coupling strength increases, the incoming quarks carry more energy, causing the VBS jets to become more forward than typical VBS jets. Consequently, both the invariant mass (m_{jj}) and pseudorapidity separation ($|\Delta\eta_{jj}|$) shift to larger values, as illustrated in Figure 8.1. The increase in the individual jet's pseudorapidity ($|\eta|$) can also be observed.

Even more striking differences are seen in the triboson system. As the C_{2V} value deviates from the SM, the W and Z longitudinal polarization mode fraction increases, and the W and Z transverse momenta become larger. Similarly, the Higgs boson also gains a boost. The shape comparison of the kinematic distributions of the bosons can be seen in Figure 8.2. All three bosons gain significant boost as the C_{2V} deviates from the SM to varying degrees.

8.3 Electroweak V Cross-Section Reweighting

During this analysis, a discrepancy emerged between the expected and the Monte Carlo (MC) yield for the EWKWLep. The majority of EWKWLep events in this region had one or both incoming b quarks. Moreover, most EWKWLep events that are VBS W events at the generator level in this region have just one outgoing quark matched to a VBS quark, with the other outgoing quark matched to the $H \to b\bar{b}$ fat jet candidate, which is predominantly a b quark. It became apparent that the kinematics of these b-initiated VBS events were incorrectly simulated.

It was eventually determined that the issue originated at the MADGRAPH level. The following line was originally used to generate the EWKWLep samples in the process card:

```
pp > l vl j j / t t~ h QCD = 0
```

Here, the '/' excludes diagrams containing $H, \mathrm{b}, \bar{\mathrm{b}}$ from generation and ignores any interference from them. However, the process should have been generated using:

```
pp > l vl j j $ t t~ h QCD = 0
```

In this line, the '$' excludes diagrams containing $H, \mathrm{b}, \bar{\mathrm{b}}$ from generation but includes any interference from them. The interference from diagrams containing $\mathrm{b}, \bar{\mathrm{b}}$ is significant. As shown in Fig. 8.3a, the P_T of outgoing b quarks is significantly boosted when using the incorrect MADGRAPH generation line. Essentially, the EWKWLep samples are beyond the Standard Model (BSM) samples where the top quark does not exist, placing them in a familiarly boosted phase space.

As a temporary fix, a scale factor was derived by taking the ratio of the histograms of the outgoing b quark P_T for the correctly generated samples (numerator) and the incorrectly generated samples (denominator). This scale factor, binned in P_T of the outgoing b quark, is sufficient to completely resolve the issue (see Fig. 8.3b). Additionally, the EWKWLep samples have no restriction on the minimum dijet mass, so diboson events are also included in the sample. Since this analysis uses a dedicated diboson sample already, these events are removed to avoid double counting. At the time of writing, the EWKWLep samples have been decommissioned and are being centrally generated with the correct MADGRAPH line.

8.4 ABCDNet

The ABCD method is a widely used technique for data-driven background estimates in high energy physics. The method relies on two statistically independent discriminating variables, used to divide the phase space into four regions. The background in the signal region can be estimated simply using the other three control regions. The variables are typically chosen based on prior physics knowledge.

The paper [24] proposes automating the design of one or both of these variables using machine learning. In our analysis, we only attempt the first method, where a discriminator

is built and kept by construction decorrelated from another variable. In the second case, two discriminators are built and kept independent simultaneously.

In our case, a Deep Neural Network (DNN) is trained to serve as one of the "arms" of a traditional ABCD background estimation, which is described in more detail in the next section. A term dCorr^2 is added to the loss function that trains the DNN to be decorrelated with the other arm:

$$\mathscr{L}[f(\vec{x})] = \mathscr{L}_{\mathrm{BCE}}[f(\vec{x},y)] + \lambda\,\mathrm{dCorr}^2_{y=0}[f(\vec{x}),X_0] \tag{8.1}$$

where $\vec{x}$ is the input vector, y is the truth label (1 for signal, 0 for background), λ is a tunable parameter controlling the size of the decorrelation term, and X_0 is the decorrelation target, which the analyzer is free to choose. Binary Cross Entropy (BCE) is used in our analysis; however, any loss function $\mathscr{L}$ could, in principle, be used in its place.

Moreover, dCorr^2 is the "distance correlation," a statistical quantity that measures the dependence of two variables f and g, based on the "distance covariance" dCov^2 between them:

$$\mathrm{dCov}^2[f,g] = \langle|f-f'|\times|g-g'|\rangle + \langle|f-f'|\rangle\times\langle|g-g'|\rangle - 2\langle|f-f'|\times|g-g''|\rangle \tag{8.2}$$

$$\mathrm{dCorr}^2[f,g] = \frac{\mathrm{dCov}^2[f,g]}{\mathrm{dCov}[f,f]\,\mathrm{dCov}[g,g]} \tag{8.3}$$

For each training, we explored the hyperparameter phase space, particularly tuning the weight of the dCorr^2, λ, and the batch size to ensure good discriminating power and the best possible decorrelation between the DNN output and the second arm used for the ABCD background estimation.

8.5 Background Estimation BDT Plot

8.6 Datacards

8.7 DeepJet b-tagging efficiency maps

8.8 Data/MC for cutflow stages of the 1-lepton analysis

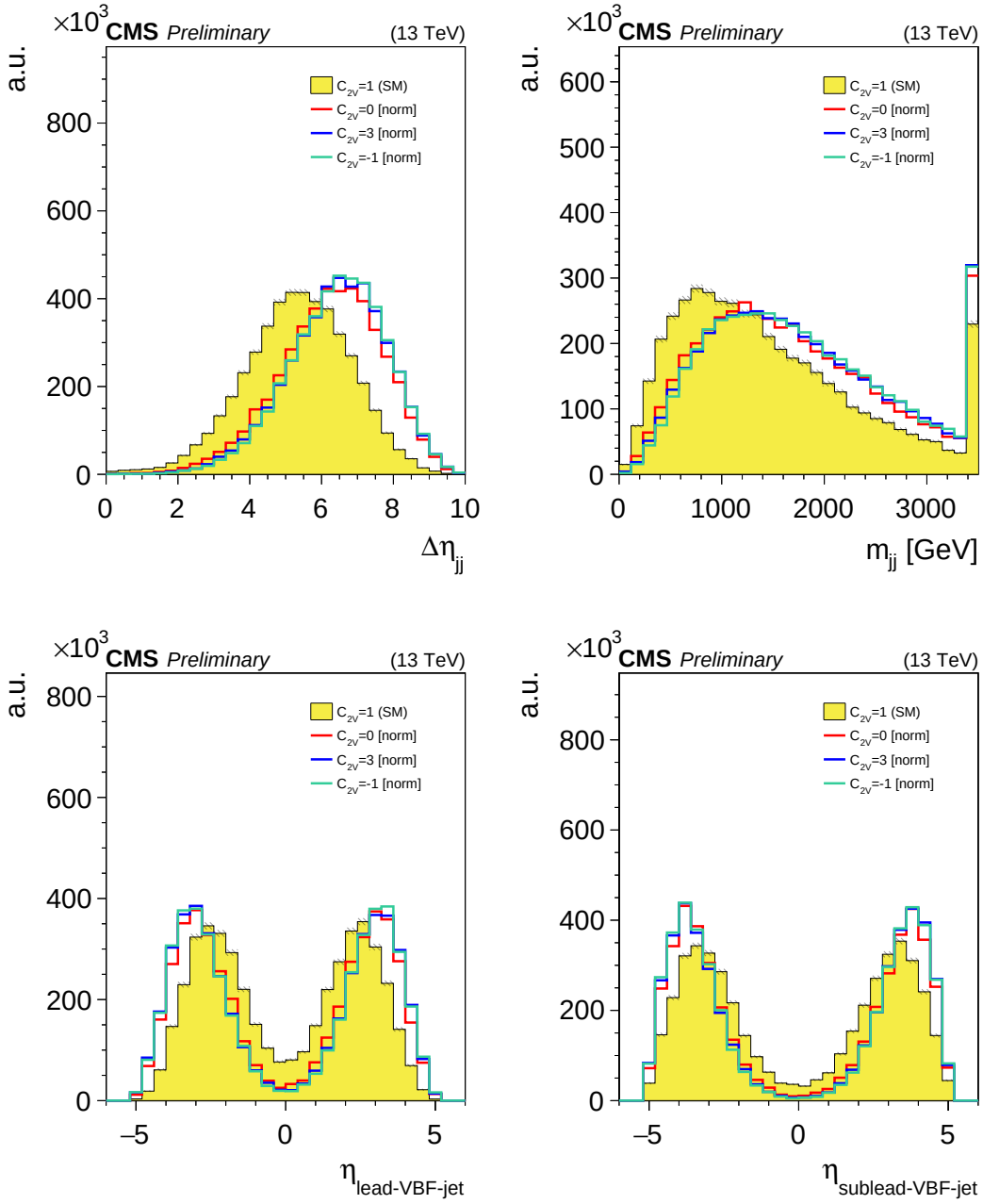


Figure 8.1. Signal process distributions for variables sensitive to the VBS features of the event for various C_{2V} values. The SM expected value of $C_{2V} = 1$ is shown in the yellow histogram, while other C_{2V} values are shown in line histograms with different colors. All histograms are arbitrarily normalized to have the same total integral, thereby comparing only the shape. Only the same-sign WWH production is plotted here.

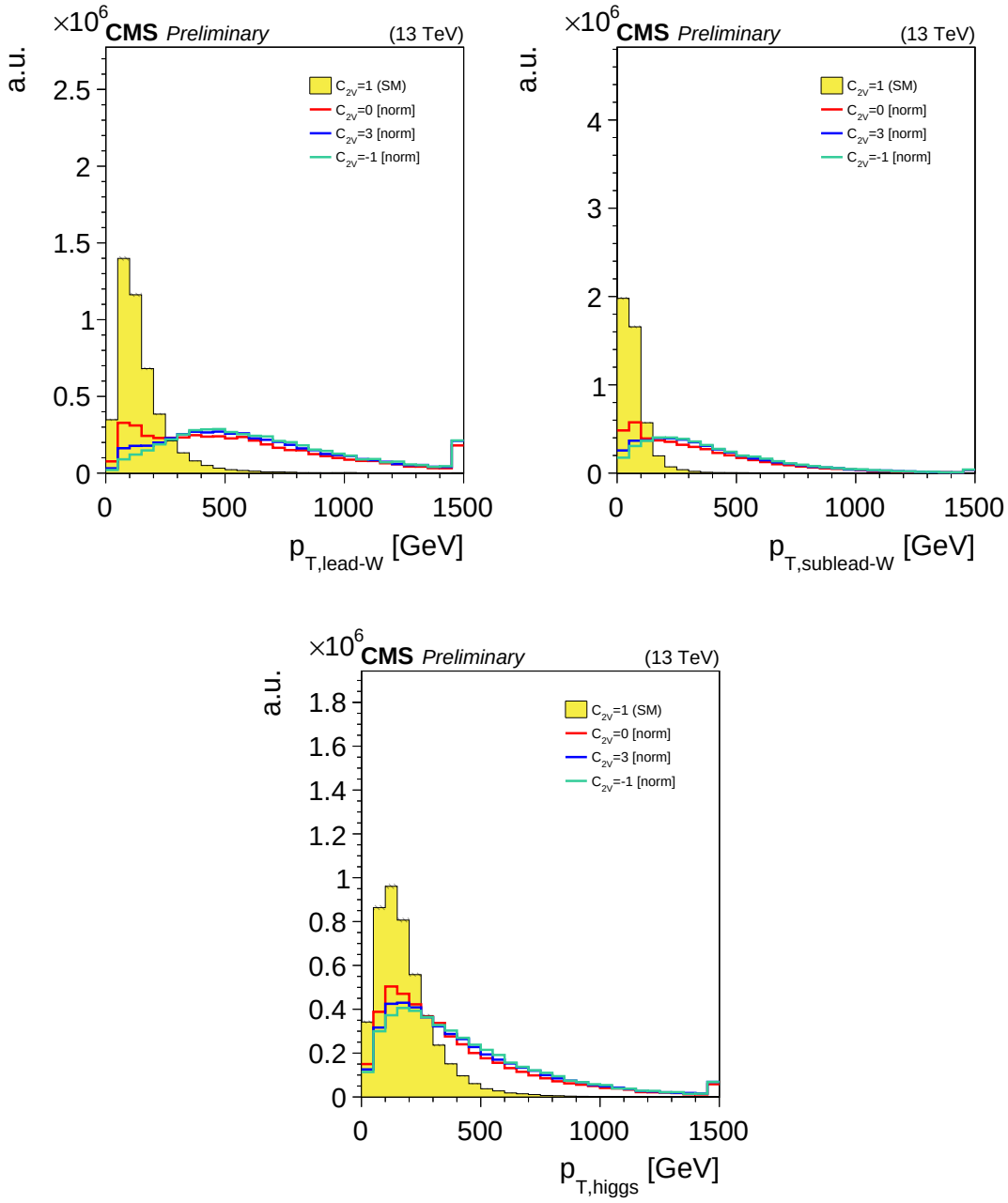


Figure 8.2. Signal process distributions for variables sensitive to the hard scattering energy of the VBS VVH process for various C_{2V} values. The SM expected value of $C_{2V} = 1$ is shown in the yellow histogram, while other C_{2V} values are shown in line histograms with different colors. All histograms are arbitrarily normalized to have the same total integral, thereby comparing only the shape. Only the same-sign WWH production mode is plotted.

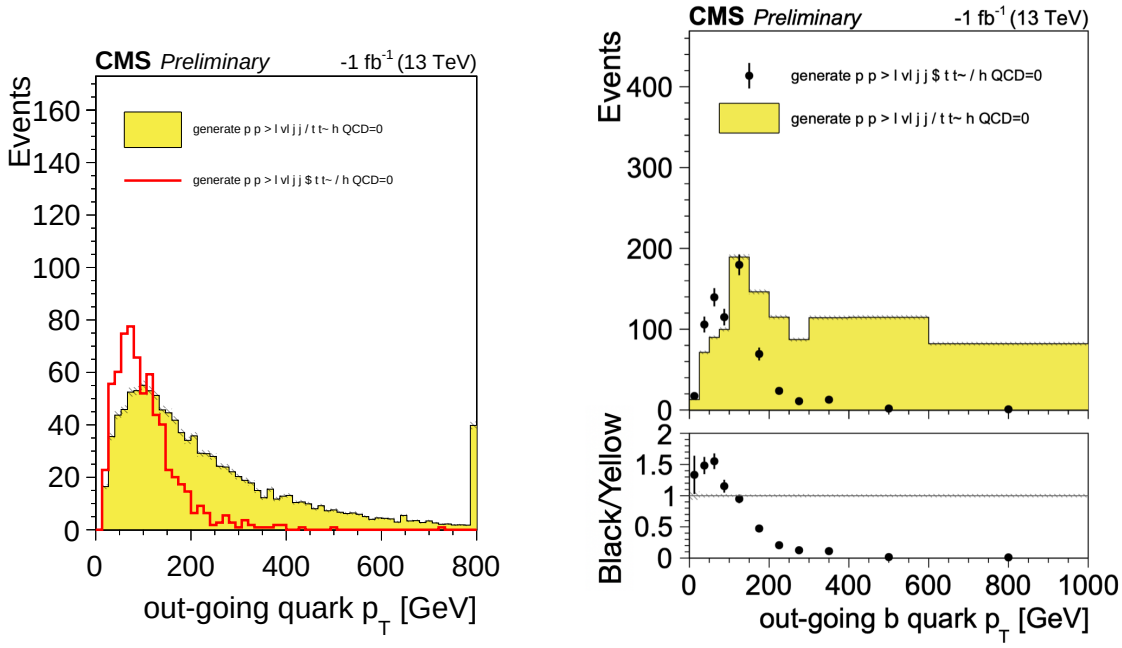


Figure 8.3. The P_T of outgoing b quarks are plotted for the correctly generated samples versus the incorrectly generated samples. The left plot shows the distribution with evenly binned data, while the right plot shows the distribution binned to enable the ratio of correctly to incorrectly generated samples to be taken as a scale factor.

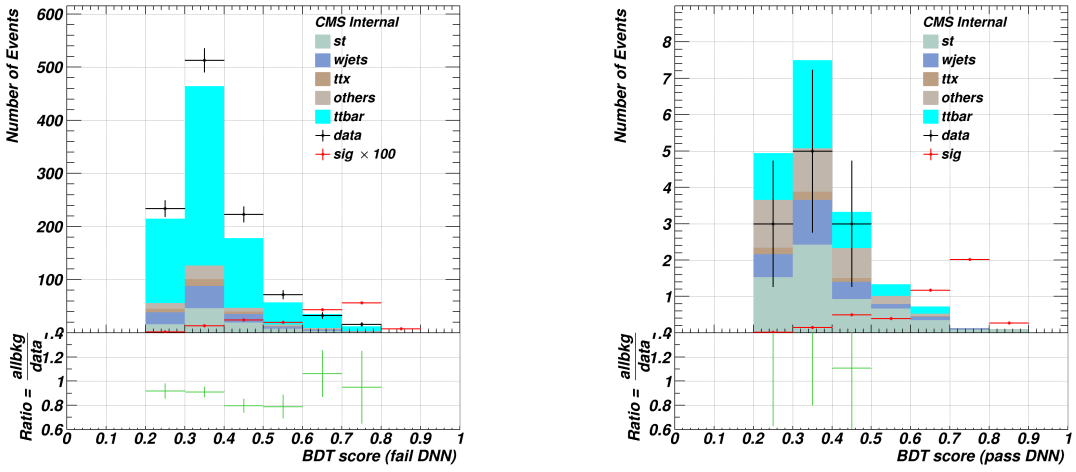
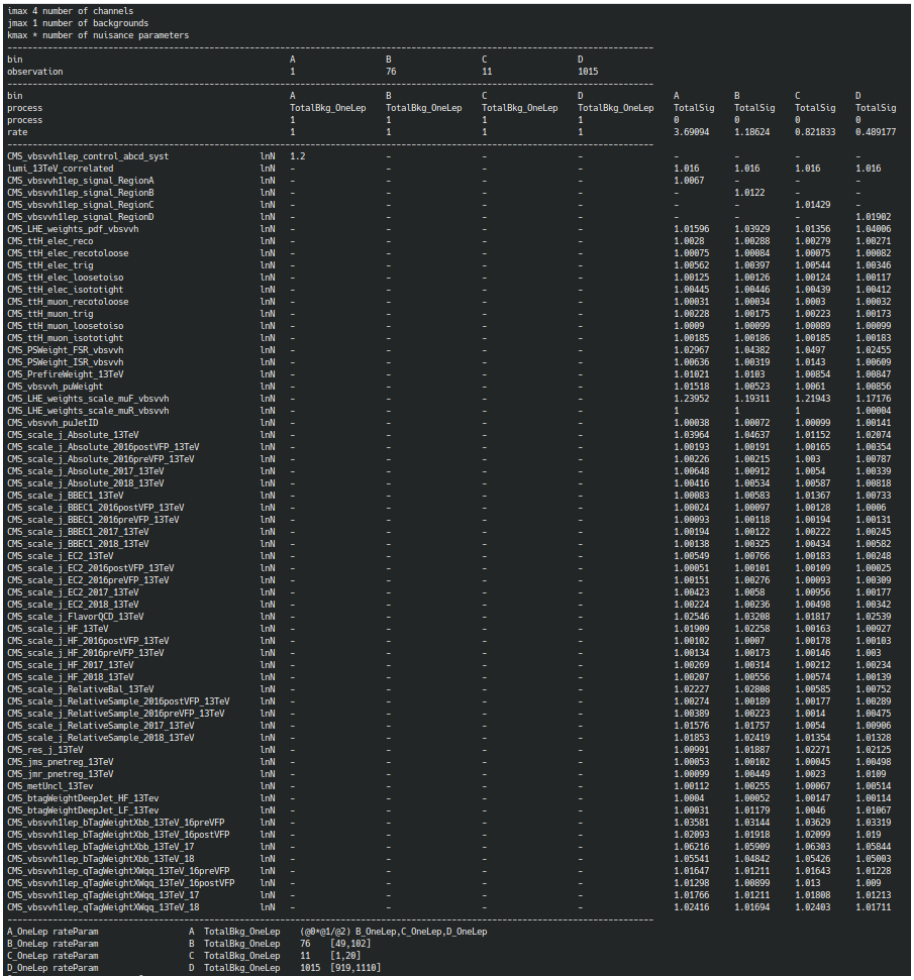


Figure 8.4. Regions C and D are plotted on the left, and regions A and B are plotted on the right.



```
imax 4 number of channels
jmax 1 number of backgrounds
kmax * number of nuisance parameters
--------------------------------------------------------------------------------------------------------------------------------
bin                                                A               B               C               D
observation                                        1               76              11              1015
--------------------------------------------------------------------------------------------------------------------------------
bin                                                A               B               C               D               A          B          C          D
process                                            TotalBkg_OneLep TotalBkg_OneLep TotalBkg_OneLep TotalBkg_OneLep TotalSig   TotalSig   TotalSig   TotalSig
process                                            1               1               1               1               0          0          0          0
rate                                               1               1               1               1               3.69094    1.18624    0.821833   0.489177
--------------------------------------------------------------------------------------------------------------------------------
CMS_vbsvvh1lep_control_abcd_syst                lnN  1.2             -               -               -               -          -          -          -
lumi_13TeV_correlated                           lnN  -               -               -               -               1.016      1.016      1.016      1.016
CMS_vbsvvh1lep_signal_RegionA                   lnN  -               -               -               -               1.0067     -          -          -
CMS_vbsvvh1lep_signal_RegionB                   lnN  -               -               -               -               -          1.0122     -          -
CMS_vbsvvh1lep_signal_RegionC                   lnN  -               -               -               -               -          -          1.01429    -
CMS_vbsvvh1lep_signal_RegionD                   lnN  -               -               -               -               -          -          -          1.01902
CMS_LHE_weights_pdf_vbsvvh                      lnN  -               -               -               -               1.01596    1.03929    1.01356    1.04006
CMS_ttH_elec_reco                               lnN  -               -               -               -               1.0028     1.00288    1.00279    1.00271
CMS_ttH_elec_recotoloose                        lnN  -               -               -               -               1.00075    1.00084    1.00075    1.00082
CMS_ttH_elec_trig                               lnN  -               -               -               -               1.00562    1.00397    1.00544    1.00346
CMS_ttH_elec_loosetoiso                         lnN  -               -               -               -               1.00125    1.00126    1.00124    1.00117
CMS_ttH_elec_isototight                         lnN  -               -               -               -               1.00445    1.00446    1.00439    1.00412
CMS_ttH_muon_recotoloose                        lnN  -               -               -               -               1.00031    1.00034    1.0003     1.00032
CMS_ttH_muon_trig                               lnN  -               -               -               -               1.00228    1.00175    1.00223    1.00173
CMS_ttH_muon_loosetoiso                         lnN  -               -               -               -               1.0009     1.00099    1.00089    1.00099
CMS_ttH_muon_isototight                         lnN  -               -               -               -               1.00185    1.00186    1.00185    1.00183
CMS_PSWeight_FSR_vbsvvh                         lnN  -               -               -               -               1.02967    1.04382    1.0497     1.02455
CMS_PSWeight_ISR_vbsvvh                         lnN  -               -               -               -               1.00636    1.00319    1.0143     1.00609
CMS_PrefireWeight_13TeV                         lnN  -               -               -               -               1.01021    1.0103     1.00854    1.00847
CMS_vbsvvh_puWeight                             lnN  -               -               -               -               1.01518    1.00523    1.0061     1.00856
CMS_LHE_weights_scale_muF_vbsvvh                lnN  -               -               -               -               1.23952    1.19311    1.21943    1.17176
CMS_LHE_weights_scale_muR_vbsvvh                lnN  -               -               -               -               1          1          1          1.00004
CMS_vbsvvh_puJetID                              lnN  -               -               -               -               1.00038    1.00072    1.00099    1.00141
CMS_scale_j_Absolute_13TeV                      lnN  -               -               -               -               1.03964    1.04637    1.01152    1.02074
CMS_scale_j_Absolute_2016postVFP_13TeV          lnN  -               -               -               -               1.00193    1.00191    1.00165    1.00354
CMS_scale_j_Absolute_2016preVFP_13TeV           lnN  -               -               -               -               1.00226    1.00215    1.003      1.00787
CMS_scale_j_Absolute_2017_13TeV                 lnN  -               -               -               -               1.00648    1.00912    1.0054     1.00339
CMS_scale_j_Absolute_2018_13TeV                 lnN  -               -               -               -               1.00416    1.00534    1.00587    1.00818
CMS_scale_j_BBEC1_13TeV                         lnN  -               -               -               -               1.00083    1.00583    1.01367    1.00733
CMS_scale_j_BBEC1_2016postVFP_13TeV             lnN  -               -               -               -               1.00024    1.00097    1.00128    1.0006
CMS_scale_j_BBEC1_2016preVFP_13TeV              lnN  -               -               -               -               1.00093    1.00118    1.00194    1.00131
CMS_scale_j_BBEC1_2017_13TeV                    lnN  -               -               -               -               1.00194    1.00122    1.00222    1.00245
CMS_scale_j_BBEC1_2018_13TeV                    lnN  -               -               -               -               1.00138    1.00225    1.00424    1.00582
CMS_scale_j_EC2_13TeV                           lnN  -               -               -               -               1.00549    1.00766    1.00183    1.00248
CMS_scale_j_EC2_2016postVFP_13TeV               lnN  -               -               -               -               1.00051    1.00101    1.00109    1.00025
CMS_scale_j_EC2_2016preVFP_13TeV                lnN  -               -               -               -               1.00151    1.00276    1.00093    1.00309
CMS_scale_j_EC2_2017_13TeV                      lnN  -               -               -               -               1.00423    1.0058     1.00956    1.00177
CMS_scale_j_EC2_2018_13TeV                      lnN  -               -               -               -               1.00224    1.00236    1.00498    1.00342
CMS_scale_j_FlavorQCD_13TeV                     lnN  -               -               -               -               1.02546    1.03208    1.01817    1.02539
CMS_scale_j_HF_13TeV                            lnN  -               -               -               -               1.01909    1.02258    1.00163    1.00927
CMS_scale_j_HF_2016postVFP_13TeV                lnN  -               -               -               -               1.00102    1.0007     1.00178    1.00183
CMS_scale_j_HF_2016preVFP_13TeV                 lnN  -               -               -               -               1.00134    1.00173    1.00146    1.003
CMS_scale_j_HF_2017_13TeV                       lnN  -               -               -               -               1.00269    1.00314    1.00212    1.00234
CMS_scale_j_HF_2018_13TeV                       lnN  -               -               -               -               1.00207    1.00556    1.00574    1.00139
CMS_scale_j_RelativeBal_13TeV                   lnN  -               -               -               -               1.02227    1.02808    1.00585    1.00752
CMS_scale_j_RelativeSample_2016postVFP_13TeV    lnN  -               -               -               -               1.00274    1.00189    1.00177    1.00289
CMS_scale_j_RelativeSample_2016preVFP_13TeV     lnN  -               -               -               -               1.00389    1.00223    1.0014     1.00475
CMS_scale_j_RelativeSample_2017_13TeV           lnN  -               -               -               -               1.01576    1.01757    1.0054     1.00906
CMS_scale_j_RelativeSample_2018_13TeV           lnN  -               -               -               -               1.01853    1.02419    1.01354    1.01328
CMS_res_j_13TeV                                 lnN  -               -               -               -               1.00991    1.01887    1.02271    1.02125
CMS_jms_pnetreg_13TeV                           lnN  -               -               -               -               1.00053    1.00102    1.00045    1.00498
CMS_jmr_pnetreg_13TeV                           lnN  -               -               -               -               1.00099    1.00449    1.0023     1.0109
CMS_metUncl_13Tev                               lnN  -               -               -               -               1.00112    1.00255    1.00067    1.00514
CMS_btagWeightDeepJet_HF_13Tev                  lnN  -               -               -               -               1.0004     1.00052    1.00147    1.00114
CMS_btagWeightDeepJet_LF_13Tev                  lnN  -               -               -               -               1.00031    1.01179    1.0046     1.01067
CMS_vbsvvh1lep_bTagWeightXbb_13TeV_16preVFP     lnN  -               -               -               -               1.01021    1.03144    1.00629    1.03319
CMS_vbsvvh1lep_bTagWeightXbb_13TeV_16postVFP    lnN  -               -               -               -               1.02093    1.01918    1.02099    1.019
CMS_vbsvvh1lep_bTagWeightXbb_13TeV_17           lnN  -               -               -               -               1.06216    1.05909    1.06303    1.05844
CMS_vbsvvh1lep_bTagWeightXbb_13TeV_18           lnN  -               -               -               -               1.05541    1.04842    1.05426    1.05003
CMS_vbsvvh1lep_qTagWeightXWqq_13TeV_16preVFP    lnN  -               -               -               -               1.01647    1.01211    1.01643    1.01228
CMS_vbsvvh1lep_qTagWeightXWqq_13TeV_16postVFP   lnN  -               -               -               -               1.01298    1.00899    1.013      1.009
CMS_vbsvvh1lep_qTagWeightXWqq_13TeV_17          lnN  -               -               -               -               1.01766    1.01211    1.01808    1.01213
CMS_vbsvvh1lep_qTagWeightXWqq_13TeV_18          lnN  -               -               -               -               1.02416    1.01694    1.02403    1.01711
--------------------------------------------------------------------------------------------------------------------------------
A_OneLep rateParam                A  TotalBkg_OneLep    (@0*@1/@2) B_OneLep,C_OneLep,D_OneLep
B_OneLep rateParam                B  TotalBkg_OneLep    76    [49,102]
C_OneLep rateParam                C  TotalBkg_OneLep    11    [1,20]
D_OneLep rateParam                D  TotalBkg_OneLep    1015  [919,1110]
```

Figure 8.5. Datacard used for the final result in the 1 lepton channel.

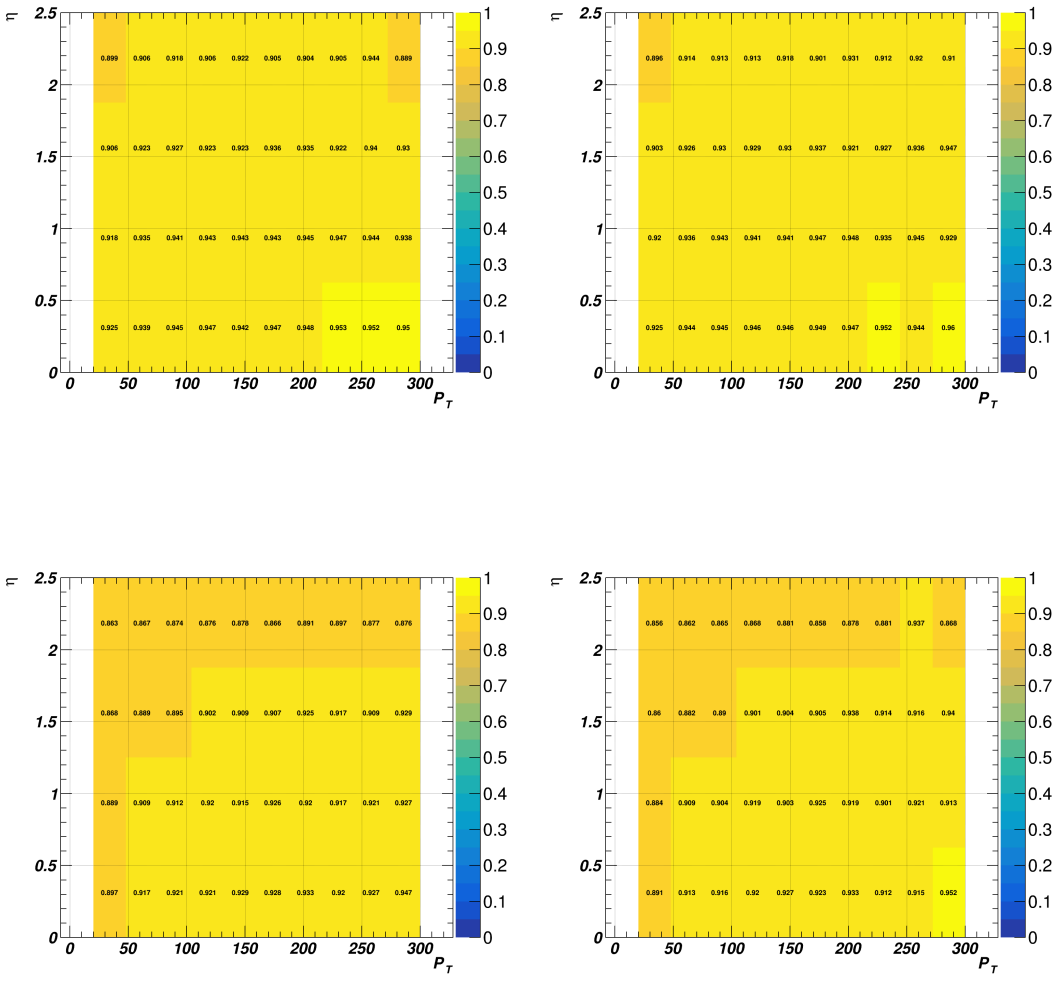


Figure 8.6. The efficiency map, binned in P_T and η, for for b quarks using the Loose DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using $t\bar{t}$ MC in the semi-leptonic final state.

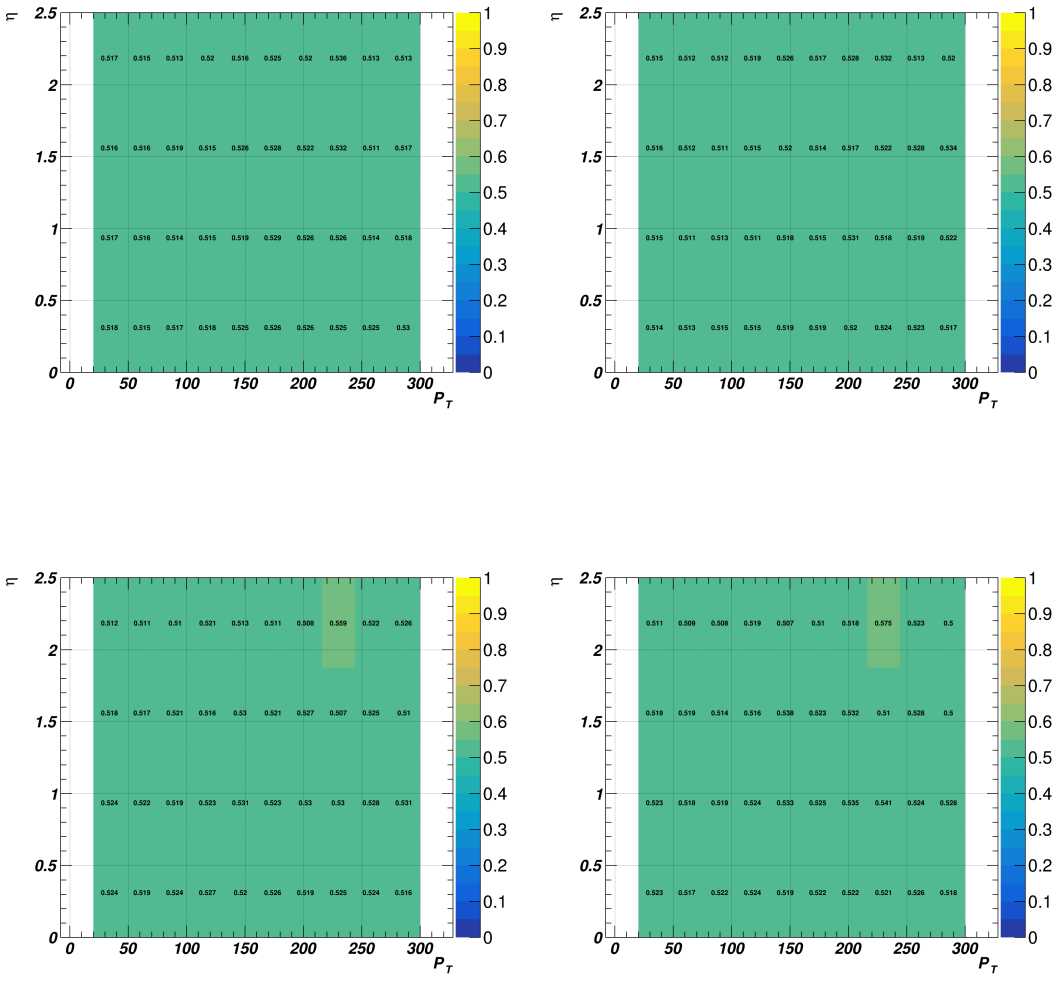


Figure 8.7. The efficiency map, binned in P_T and η, for for c quarks using the Loose DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using $t\bar{t}$ MC in the semi-leptonic final state.

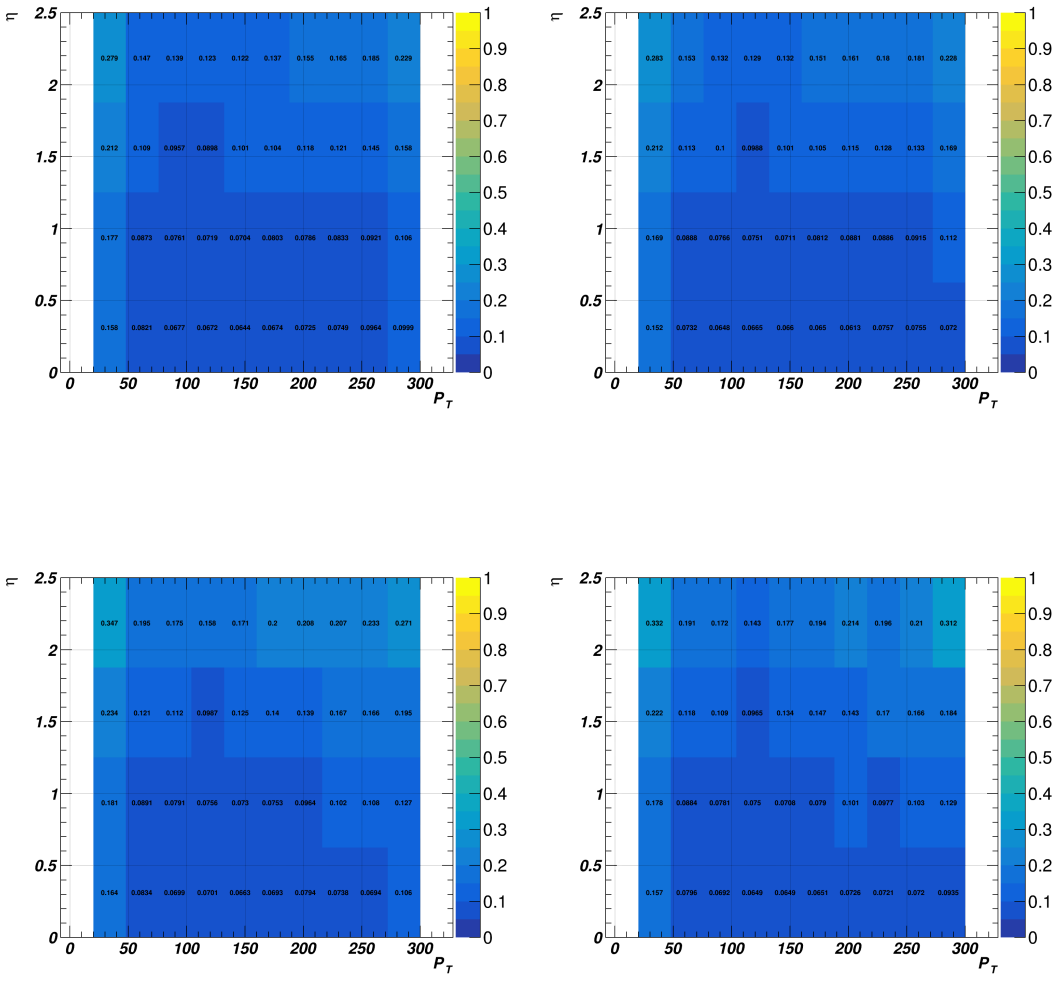


Figure 8.8. The efficiency map, binned in P_T and η, for for light quarks using the Loose DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using $t\bar{t}$ MC in the semi-leptonic final state.

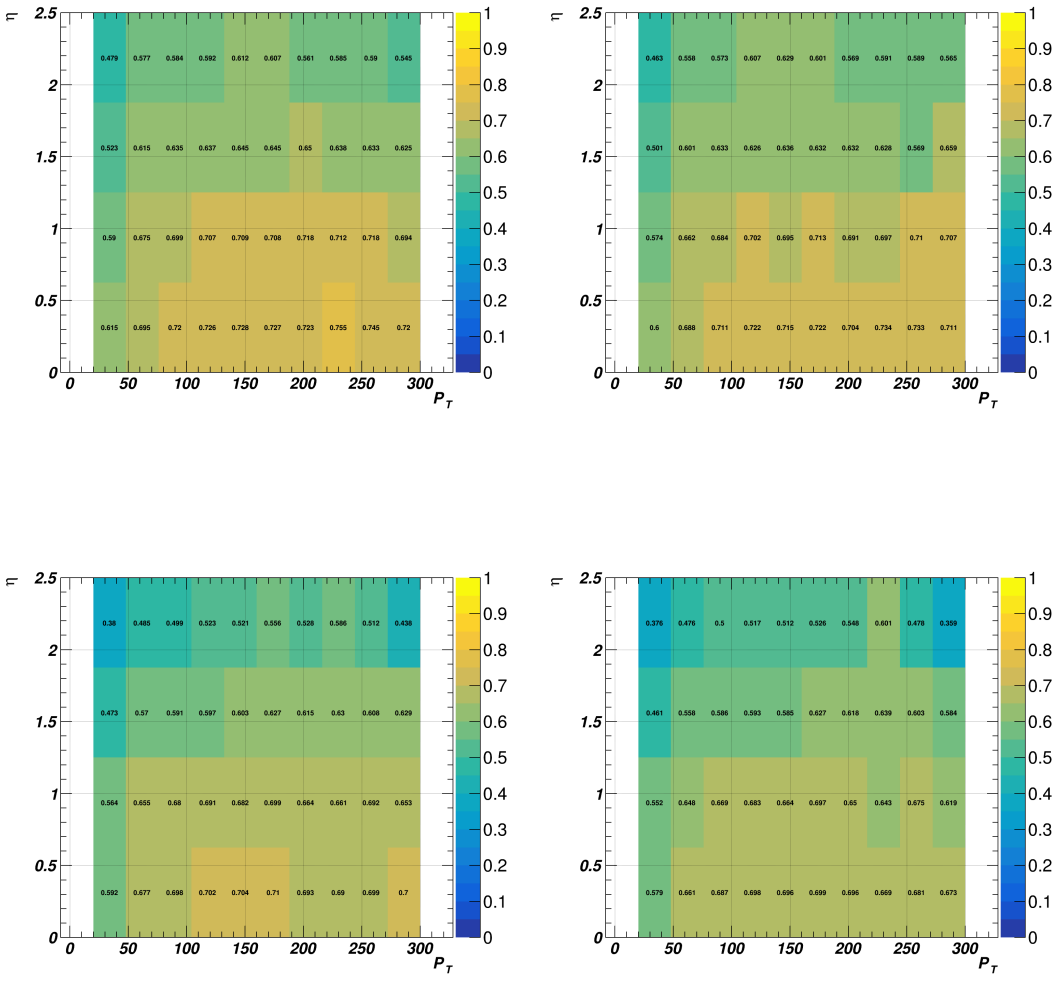


Figure 8.9. The efficiency map, binned in P_T and η, for for b quarks using the Tight DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using $t\bar{t}$ MC in the semi-leptonic final state.

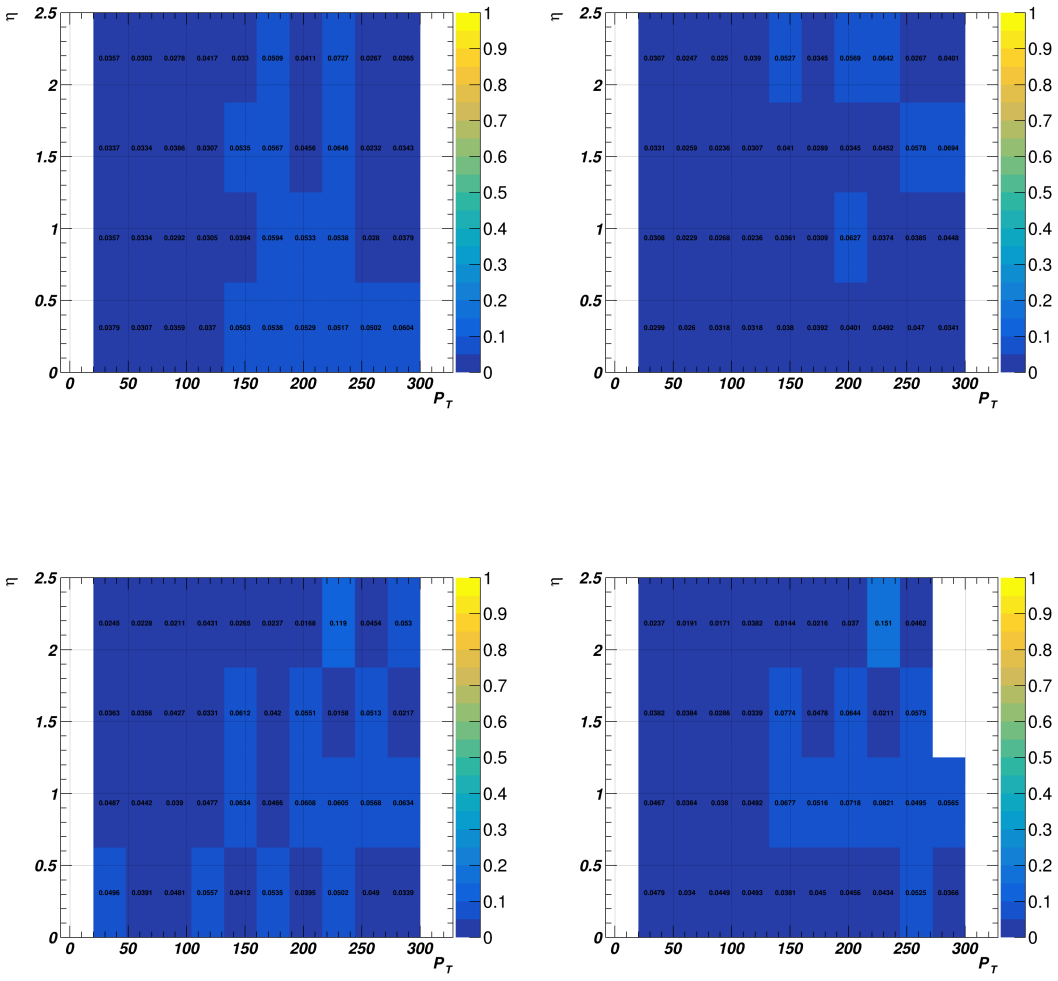


Figure 8.10. The efficiency map, binned in P_T and η, for for c quarks using the Tight DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using $t\bar{t}$ MC in the semi-leptonic final state.

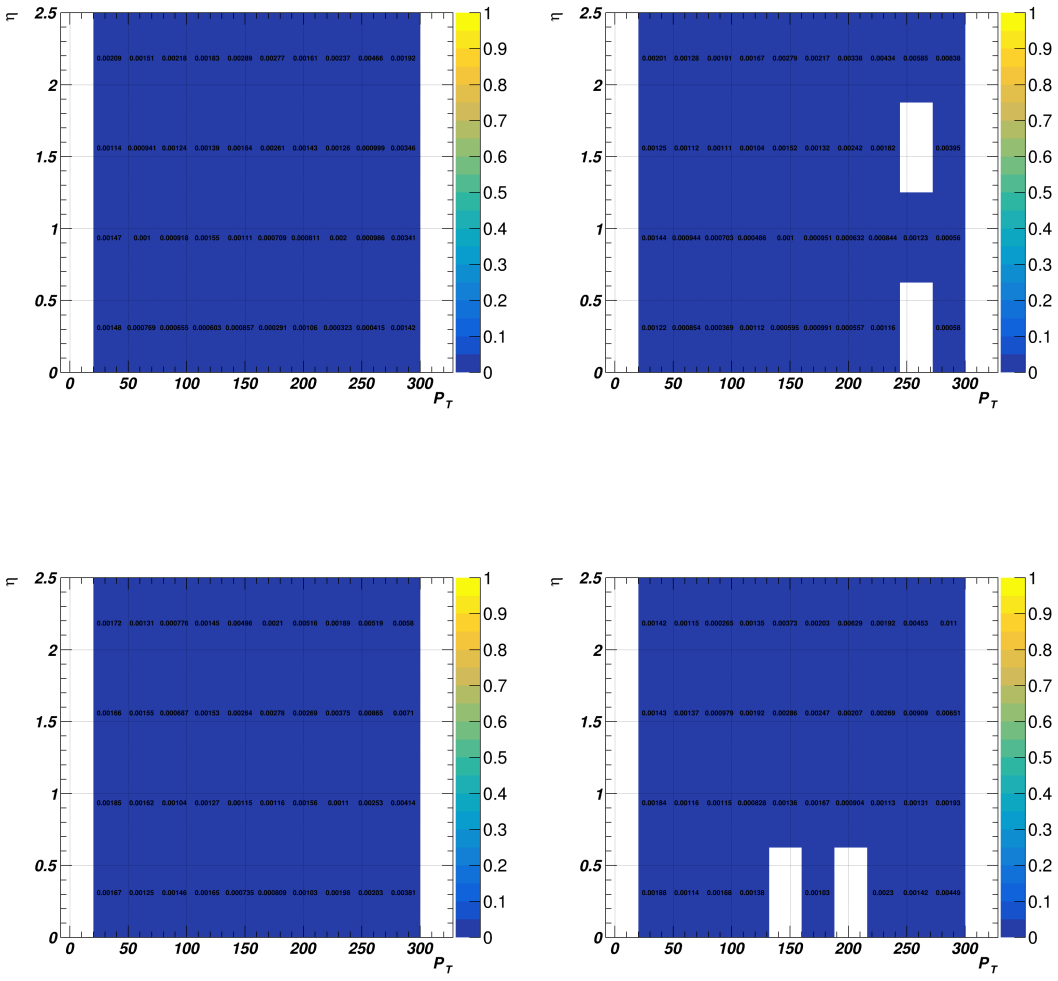


Figure 8.11. The efficiency map, binned in P_T and η, for for light quarks using the Tight DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using $t\bar{t}$ MC in the semi-leptonic final state.

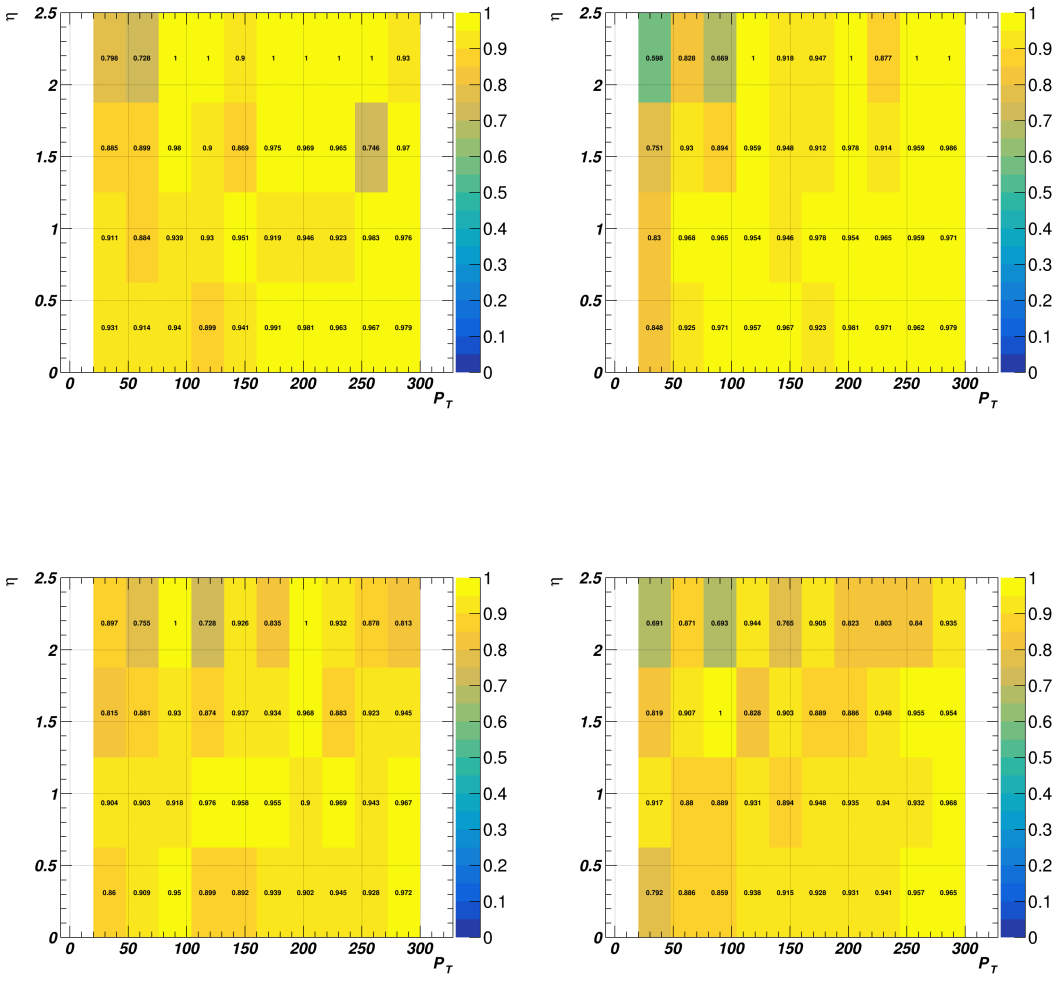


Figure 8.12. The efficiency map, binned in P_T and η, for for b quarks using the Loose DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using the VBS SSWWH signal sample.

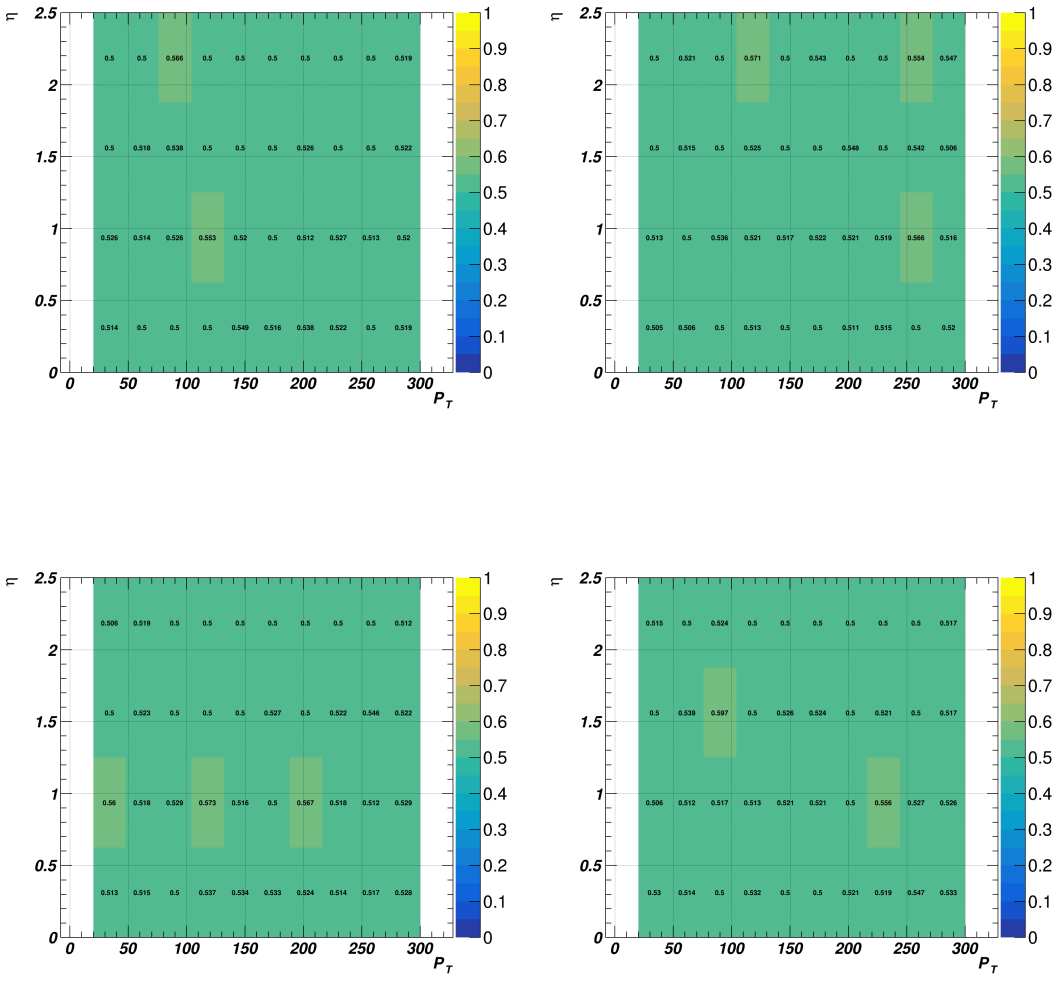


Figure 8.13. The efficiency map, binned in P_T and η, for for c quarks using the Loose DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using the VBS SSWWH signal sample.

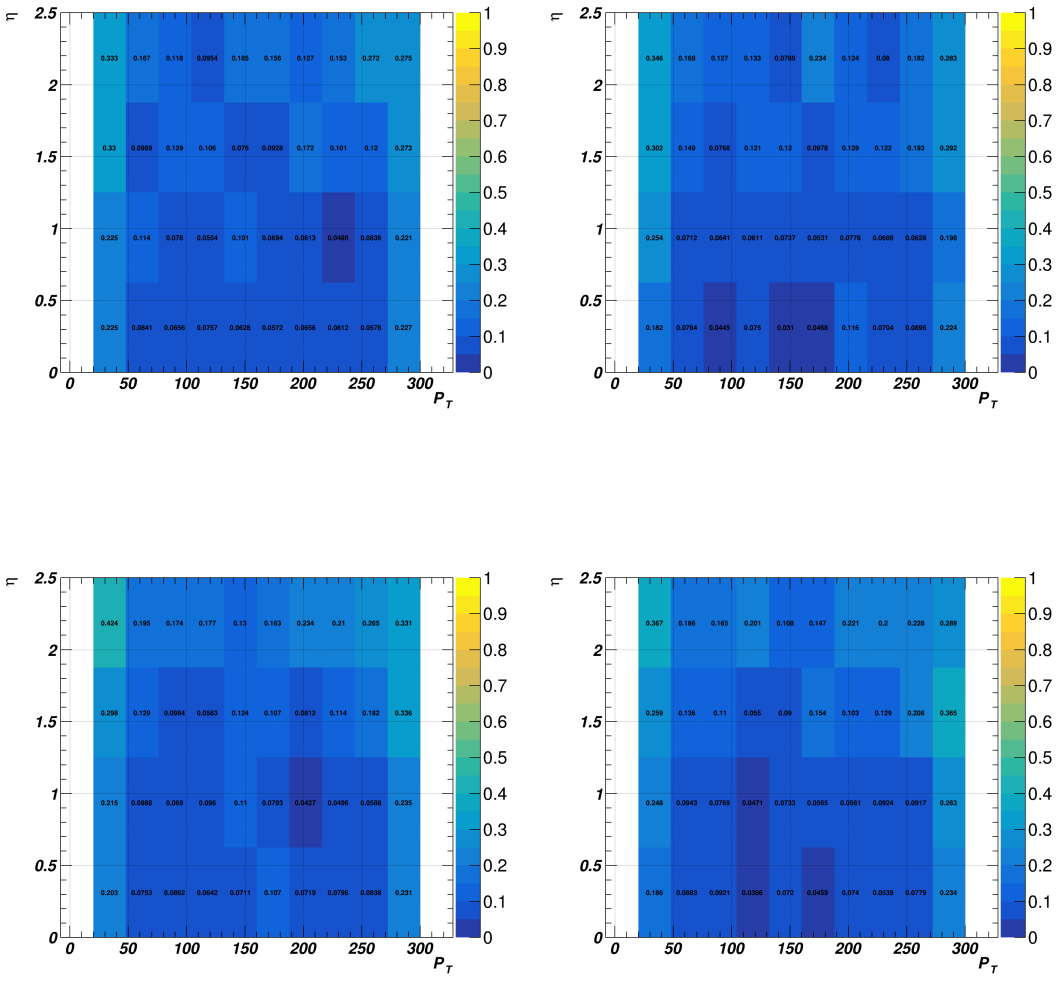


Figure 8.14. The efficiency map, binned in P_T and η, for for light quarks using the Loose DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using the VBS SSWWH signal sample.

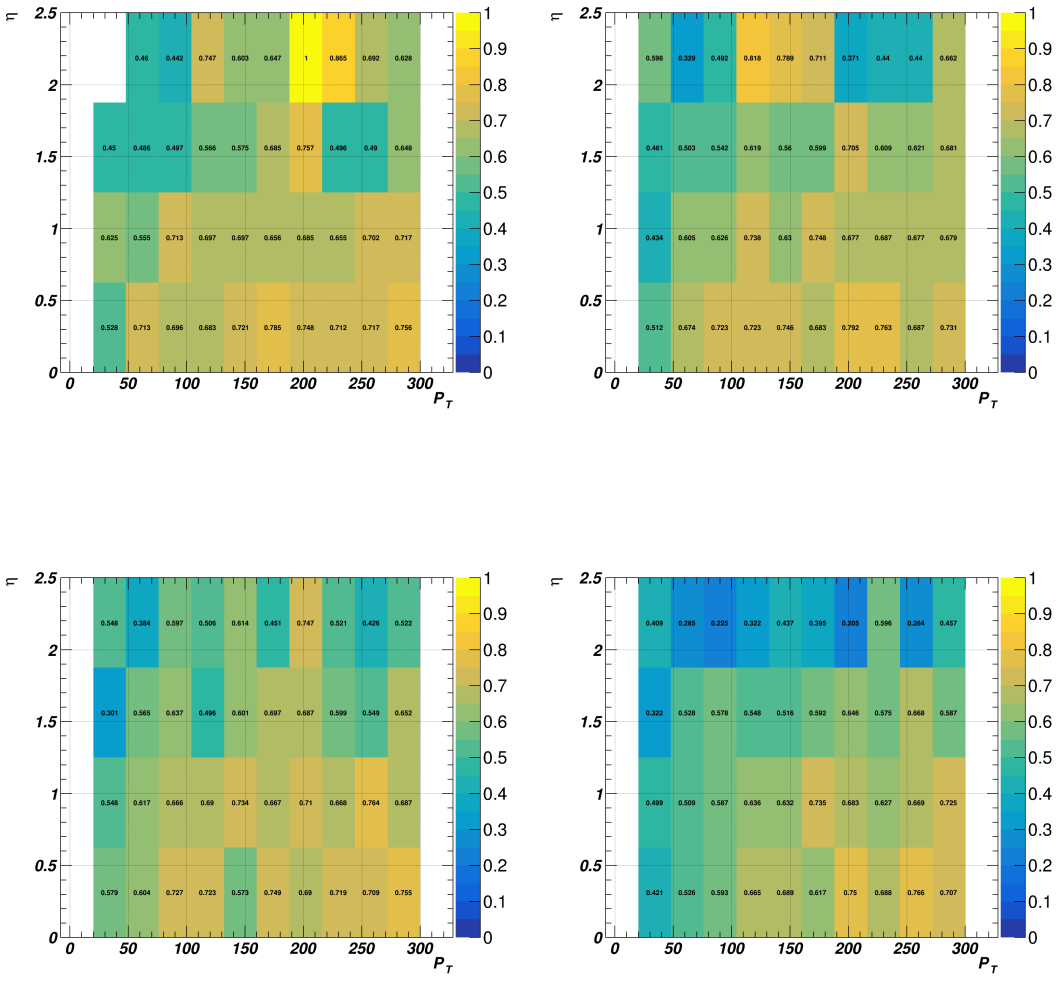


Figure 8.15. The efficiency map, binned in P_T and η, for for b quarks using the Tight DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using the VBS SSWWH signal sample.

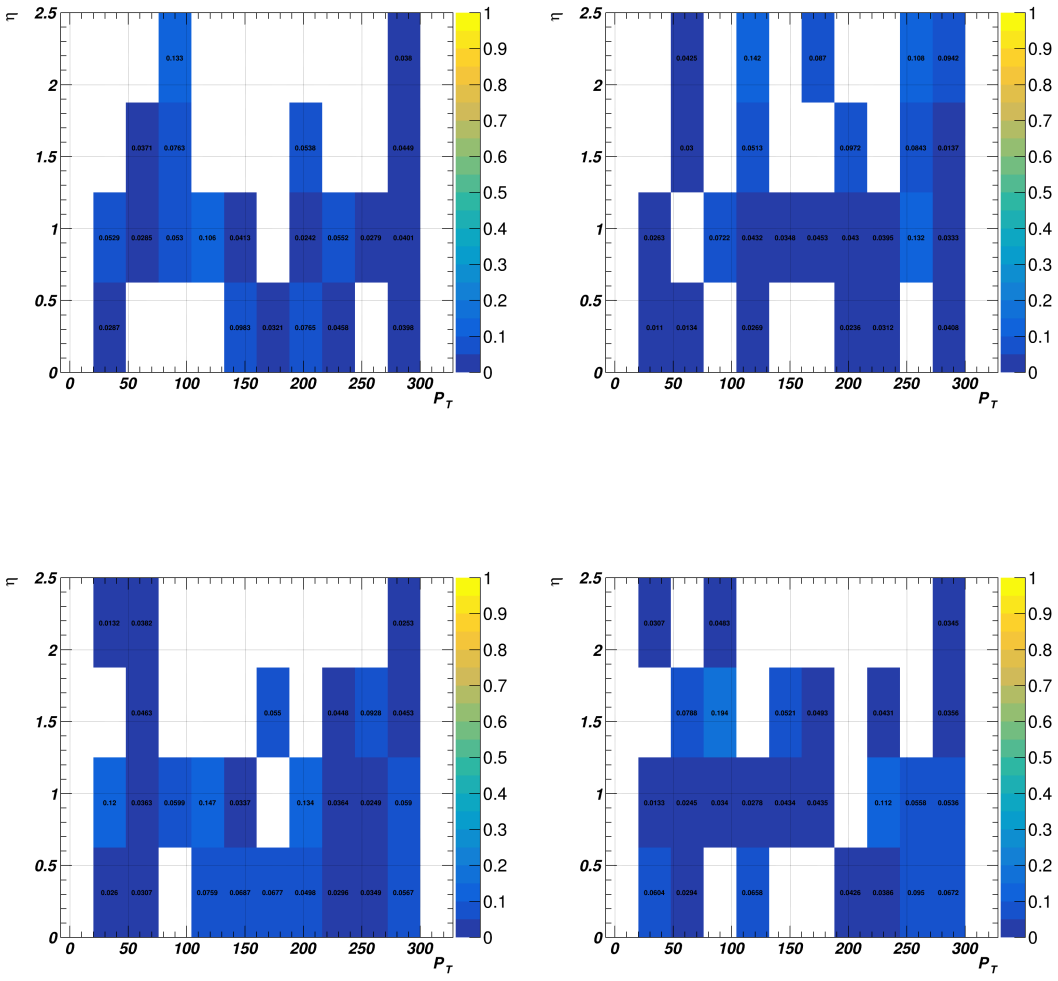


Figure 8.16. The efficiency map, binned in P_T and η, for for c quarks using the Tight DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using the VBS SSWWH signal sample.

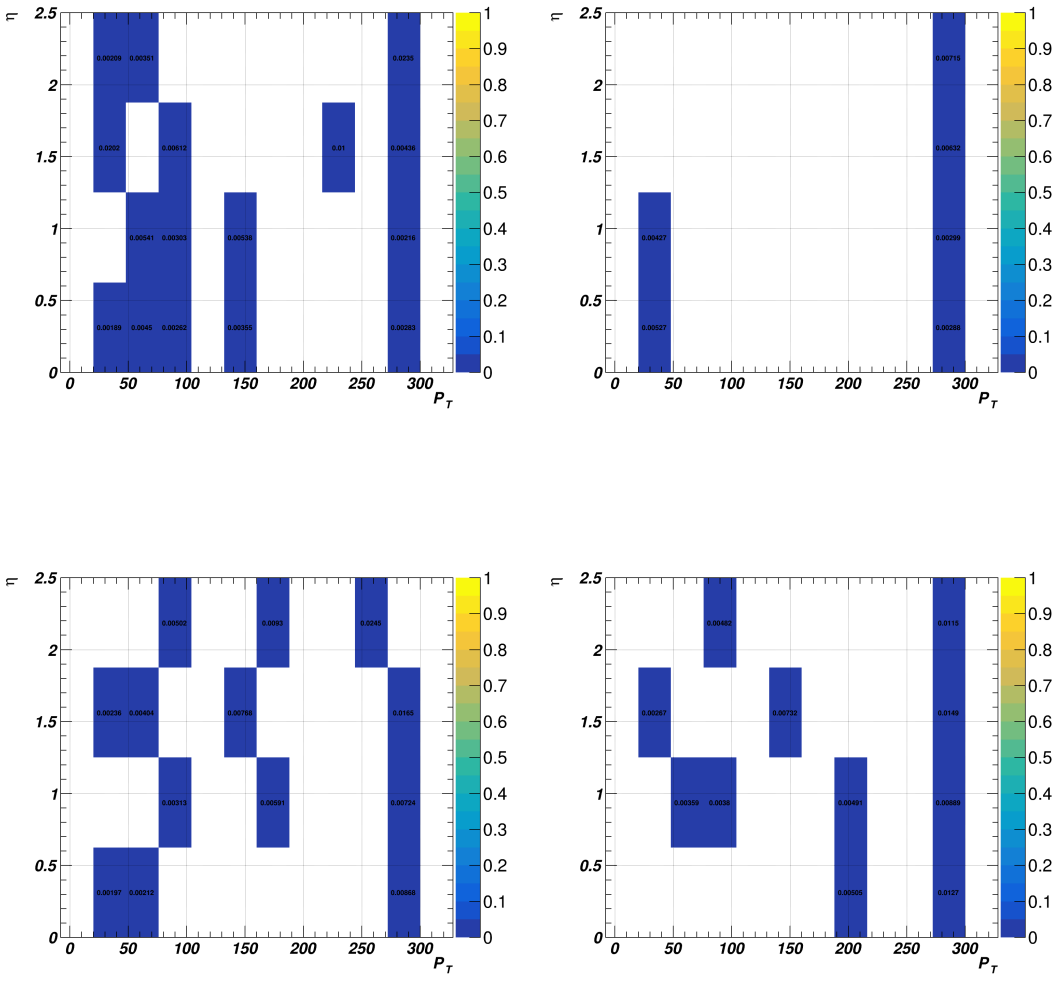


Figure 8.17. The efficiency map, binned in P_T and η, for for light quarks using the Tight DeepJet b-tagging working point is plotted for the 2018 (upper left), 2017 (upper right), 2016 post-VFP (lower left), and 2016 pre-VFP (lower right). The efficiencies are calculated using the VBS SSWWH signal sample.

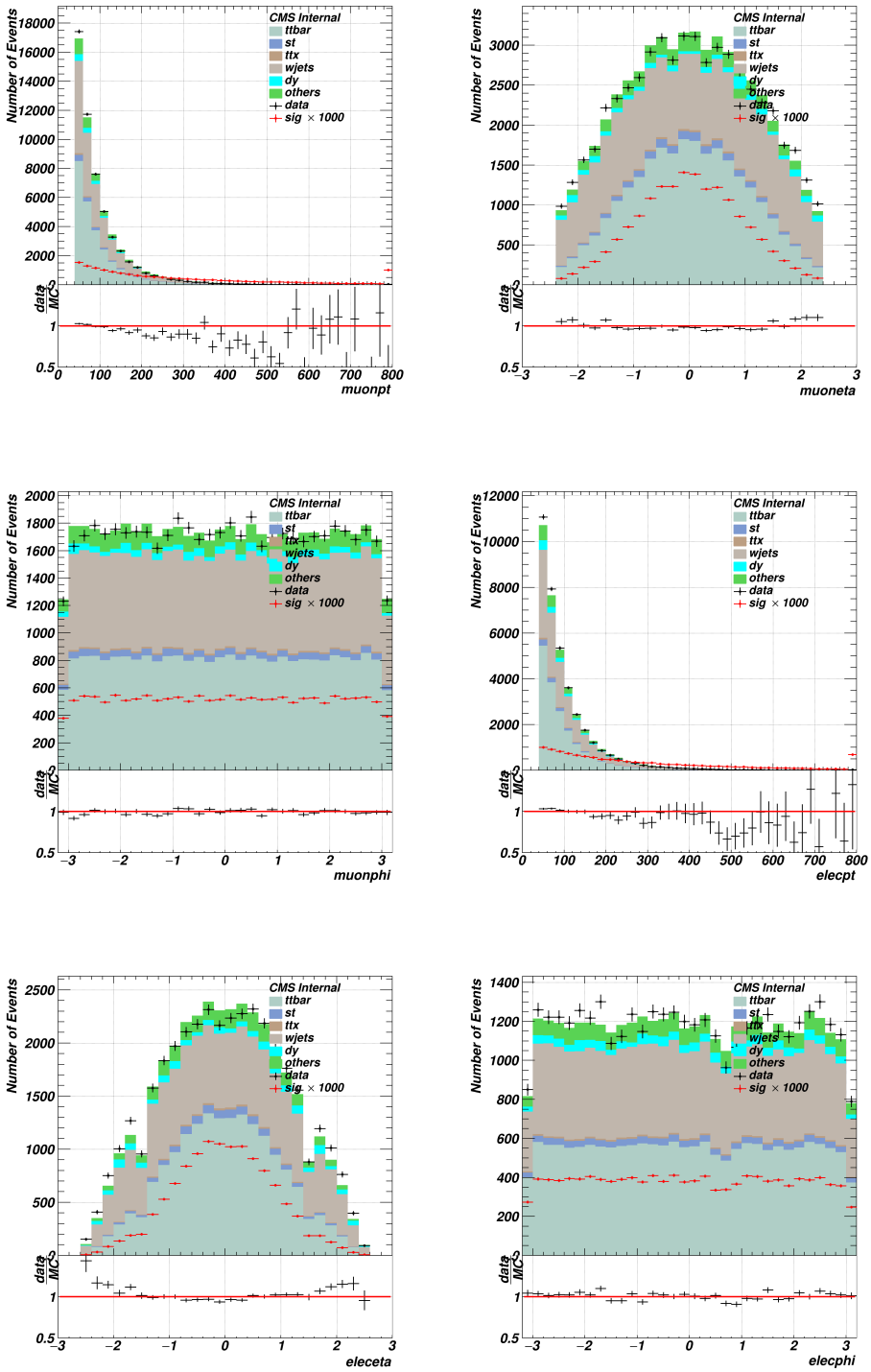


Figure 8.18. Distributions of the and electron muon candidate variables for events passing the fatjettag selection criterion in the VBSVVH analysis.

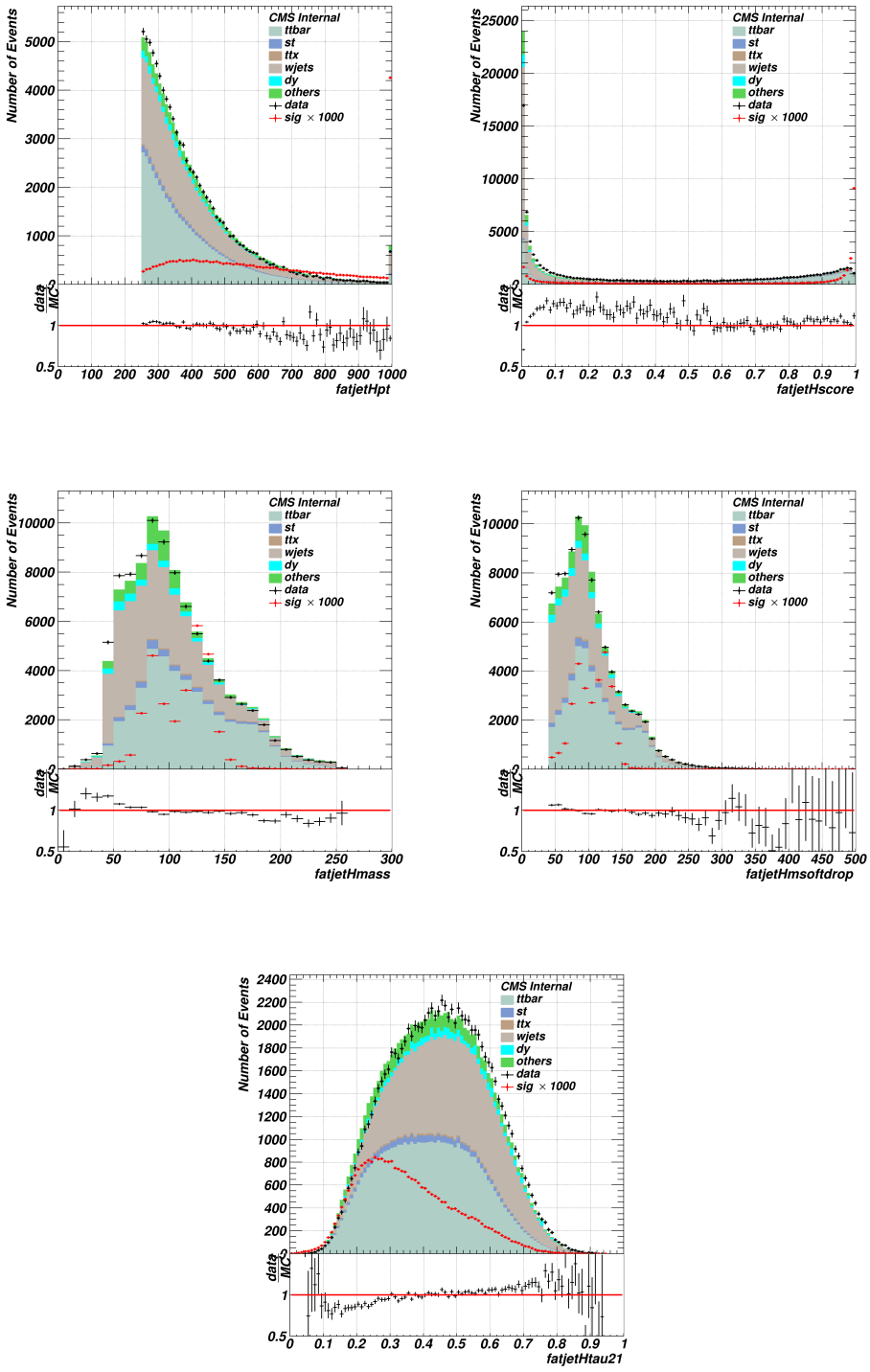


Figure 8.19. Distributions of the H candidate variables for events passing the fatjettag selection criterion in the VBSVVH analysis.

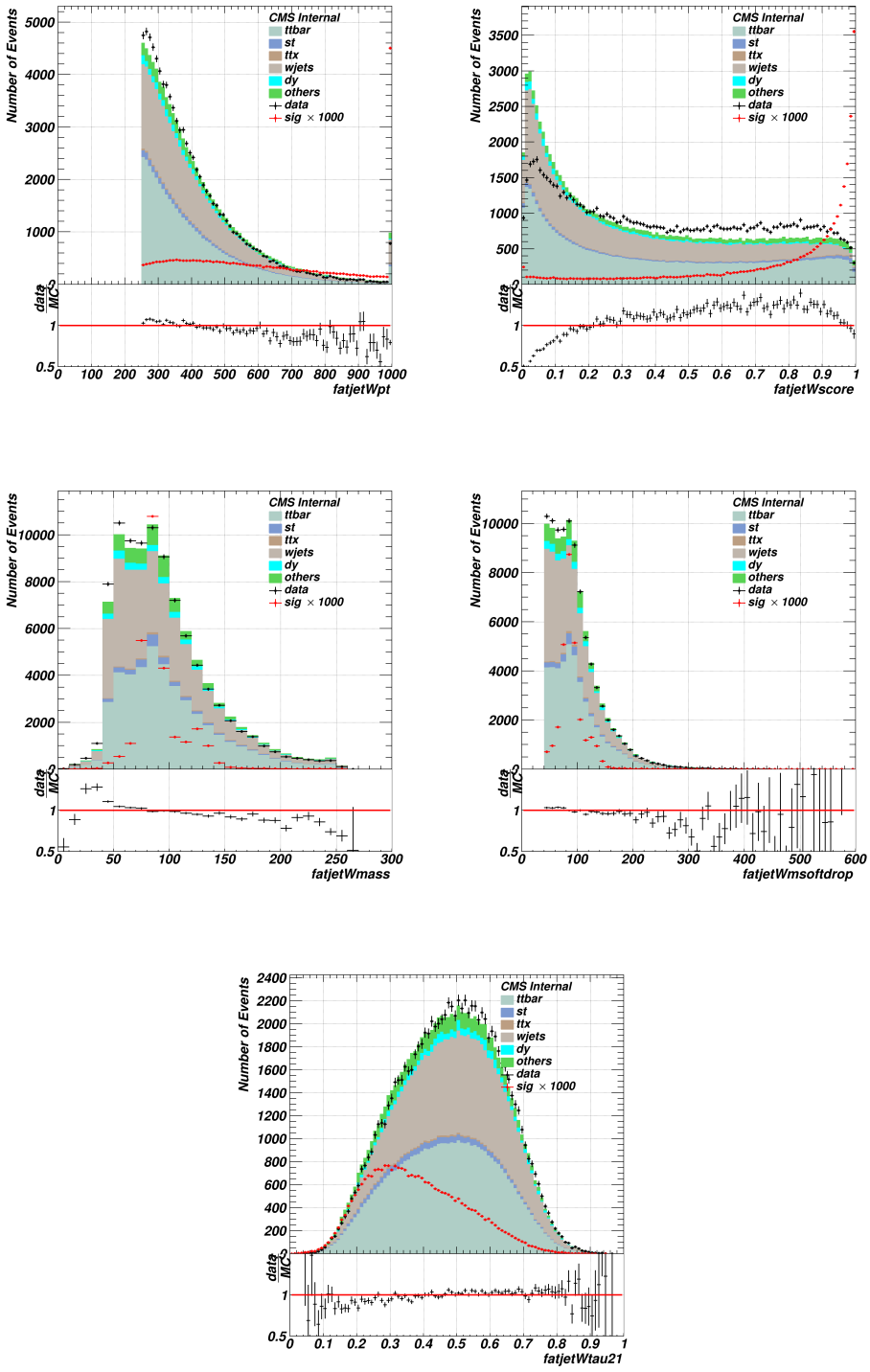


Figure 8.20. Distributions of the W candidate variables for events passing the fatjettag selection criterion in the VBSVVH analysis.

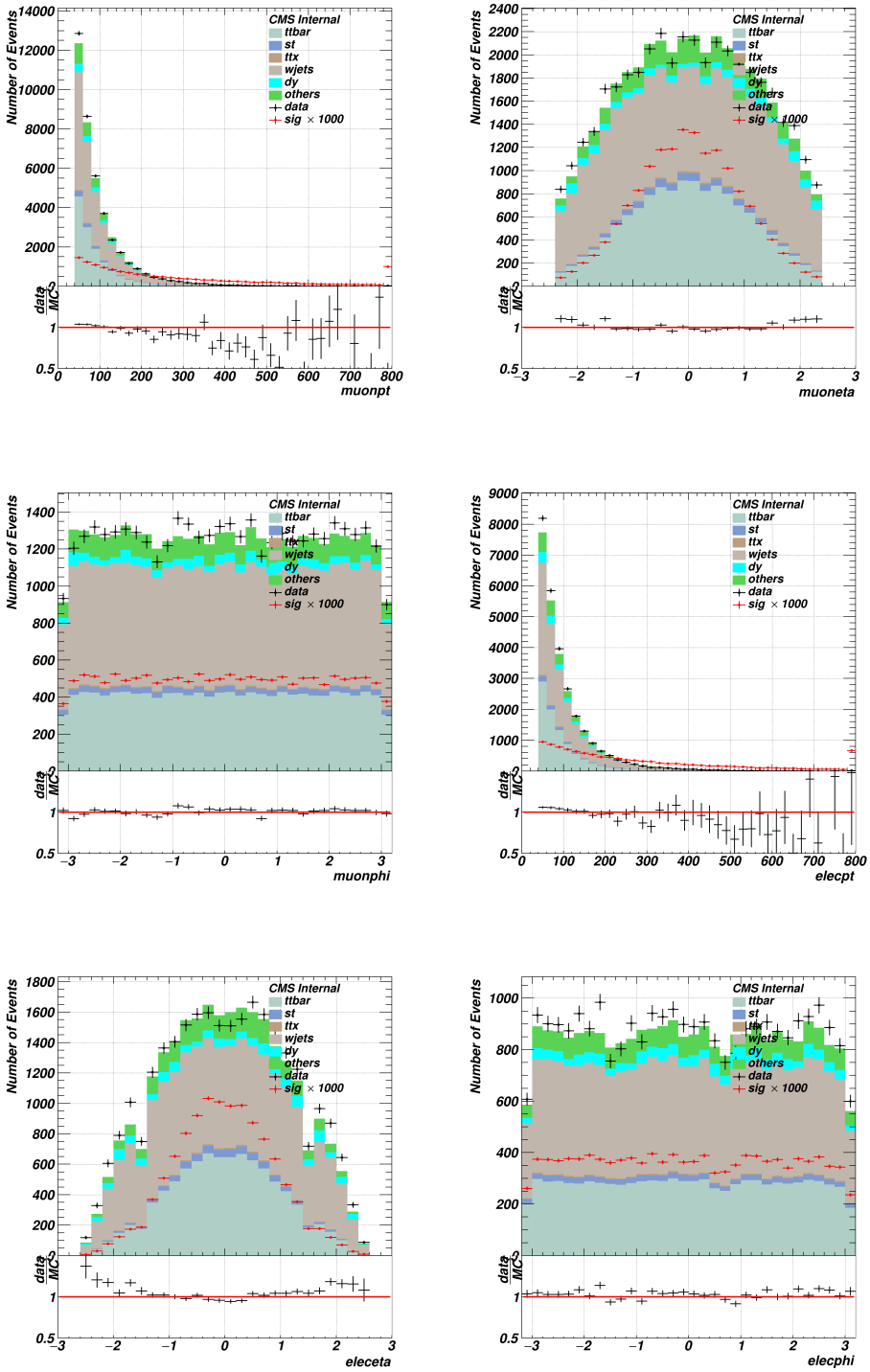


Figure 8.21. Distributions of the and electron muon candidate variables for events passing the tightak4tag selection criterion in the VBSVVH analysis.

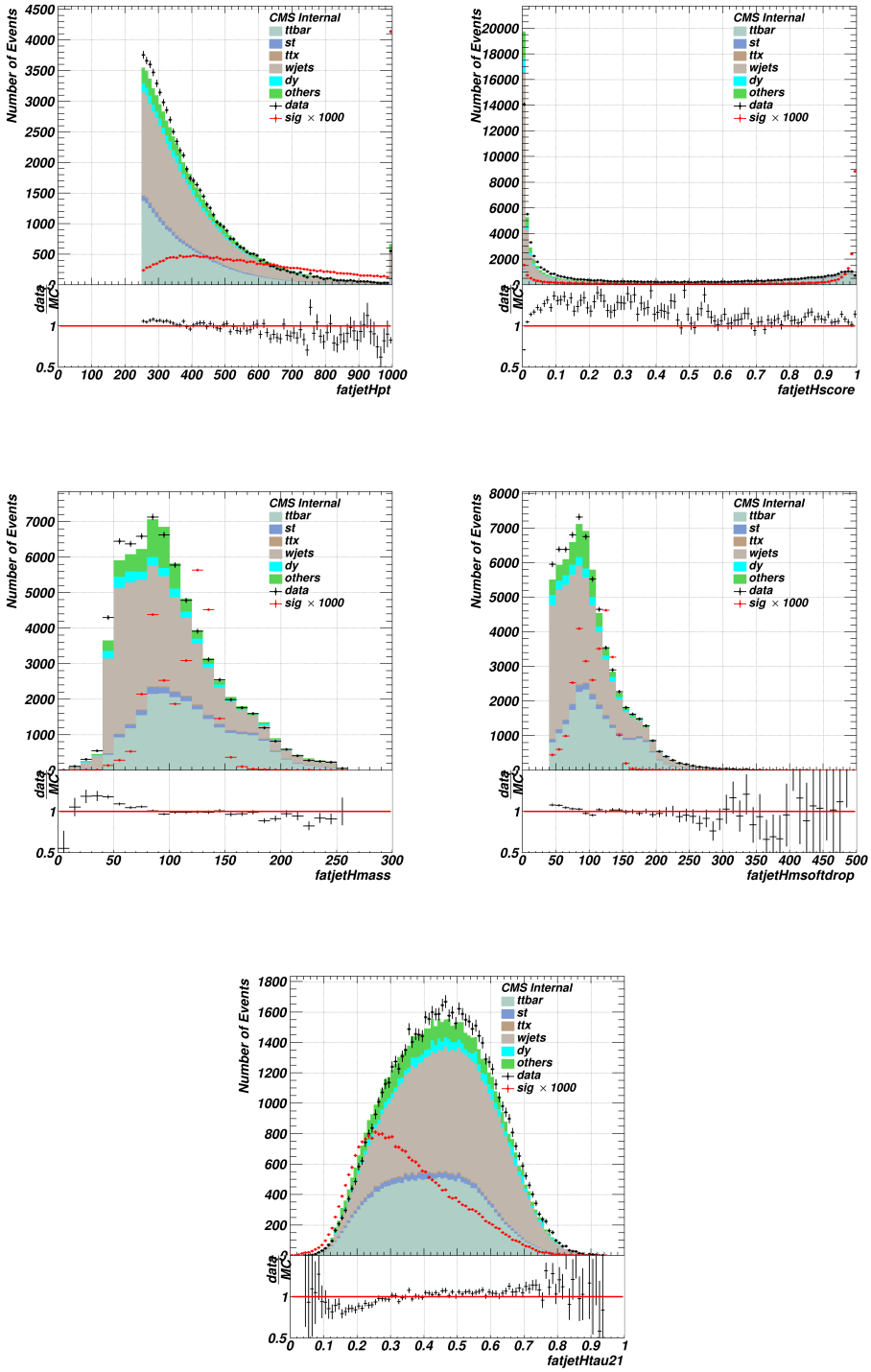


Figure 8.22. Distributions of the H candidate variables for events passing the tightak4tag selection criterion in the VBSVVH analysis.

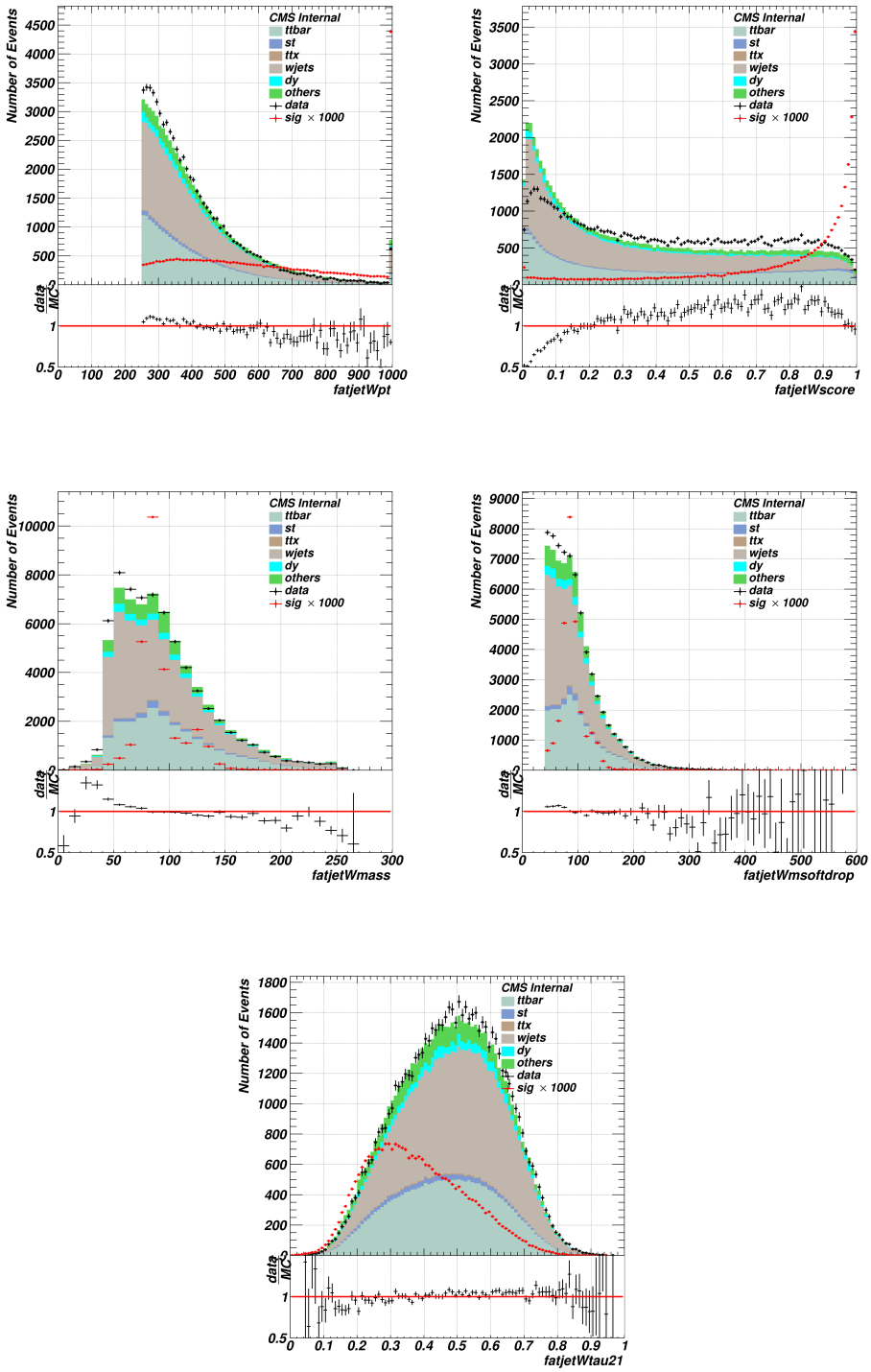


Figure 8.23. Distributions of the W candidate variables for events passing the tightak4tag selection criterion in the VBSVVH analysis.

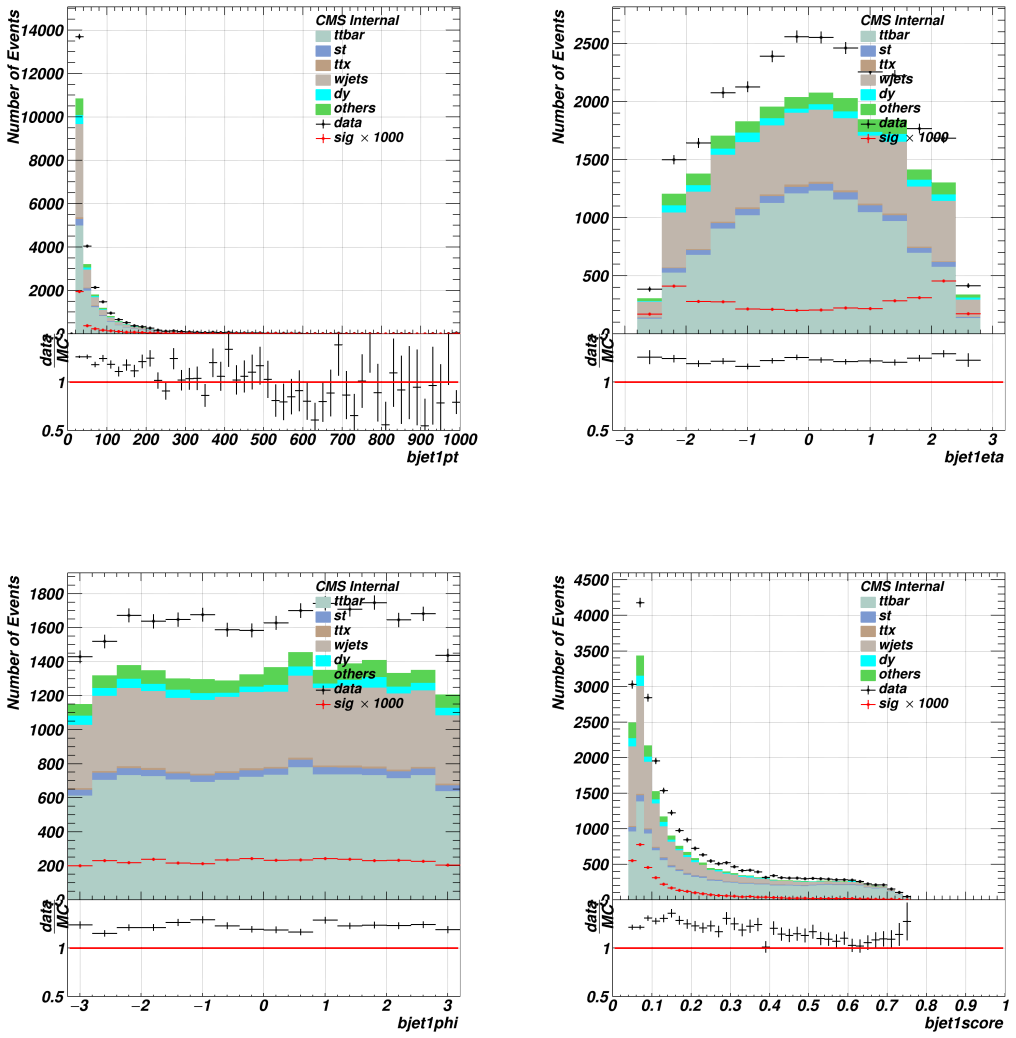


Figure 8.24. Distributions of the 1st b jet candidate variables for events passing the tightak4tag selection criterion in the VBSVVH analysis.

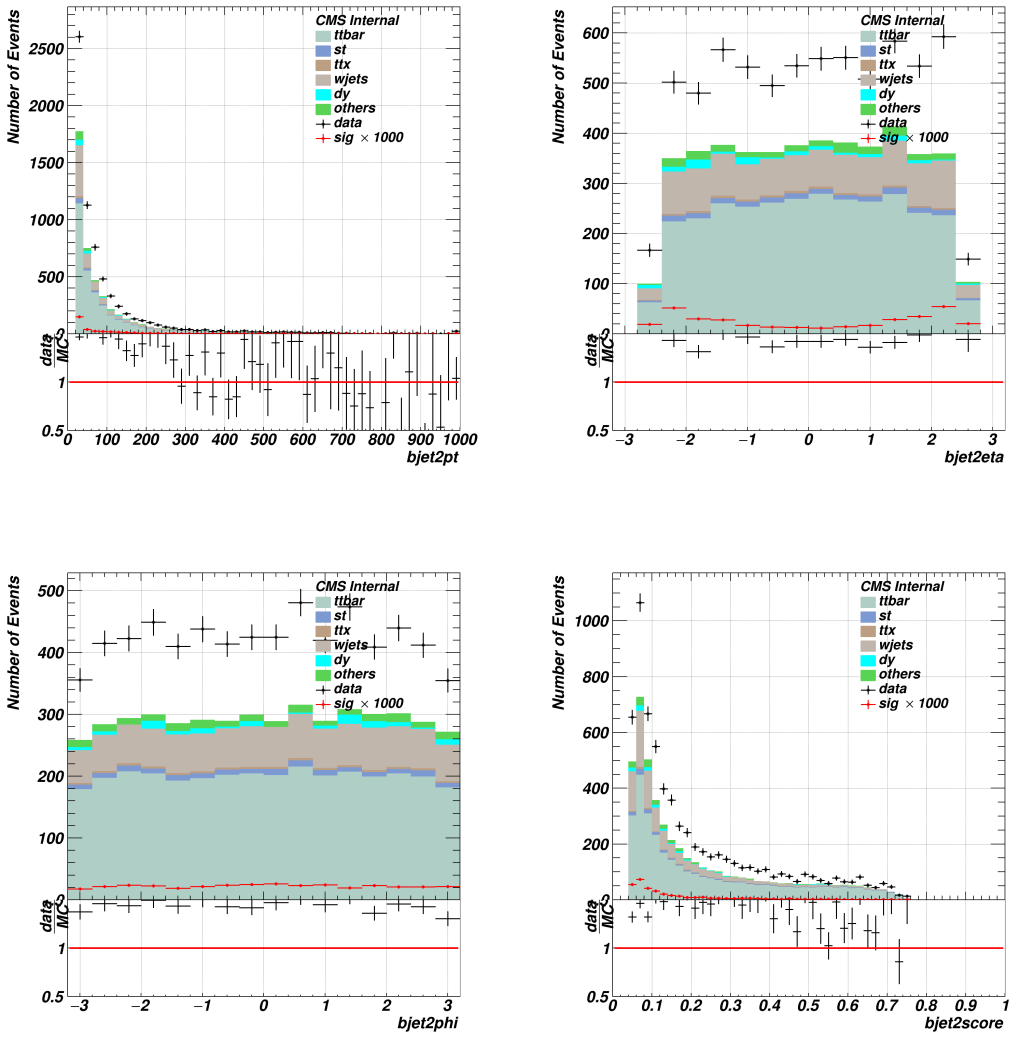


Figure 8.25. Distributions of the 2nd b jet candidate variables for events passing the tightak4tag selection criterion in the VBSVVH analysis.

Bibliography

[1] https://github.com/colizz/boohft-calib.

[2] https://indico.cern.ch/event/1221195/#4-particlenet-sf-for-ul.

[3] Btv scale factor recommendations. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation.

[4] Btv scale factor recommended methods. https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods.

[5] Jme run2 reduced set of uncertainty. https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources#Run_2_reduced_set_of_uncertainty.

[6] Lhe ht w+jets sample issue. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MCKnownIssues#W_Jets_LHEWpt_NLO_Madgraph.

[7] Lhe ht w+jets sample issue (2). https://indico.cern.ch/event/781231/contributions/3263952/subcontributions/274982/attachments/1795600/2926947/190213_mc_discrep_ppd.pdf.

[8] Treatment of the hem15/16 region in 2018 data. https://hypernews.cern.ch/HyperNews/CMS/get/JetMET/2000.html.

[9] Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements. *The European Physical Journal C*, 80(1):4, Jan 2020.

[10] Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks. 5 2022.

[11] Search for nonresonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 1 2023.

[12] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.

[13] 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. *Journal of High Energy Physics*, 2014(7):79, Jul 2014.

[14] Jon Butterworth, Stefano Carrazza, Amanda Cooper-Sarkar, Albert De Roeck, Joël Feltesse, Stefano Forte, Jun Gao, Sasha Glazov, Joey Huston, Zahari Kassabov, Ronan McNulty, Andreas Morsch, Pavel Nadolsky, Voica Radescu, Juan Rojo, and Robert Thorne. PDF4lhc recommendations for LHC run II. *Journal of Physics G: Nuclear and Particle Physics*, 43(2):023001, jan 2016.

[15] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.

[16] 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.

[17] CMS Collaboration. Search for the Higgs boson decaying to charm quarks via associated production with a vector boson using large radius jets with the full Run 2 datase. *CMS Physics Analysis Note*, AN-2021/004, 2021.

[18] ATLAS Collaboration. A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery. *Nature*, 607(7917):52–59, Jul 2022.

[19] CMS Collaboration. A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature*, 607(7917):60–68, Jul 2022.

[20] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), feb 2011.

[21] Christoph Englert, Qiang Li, Michael Spannowsky, Mengmeng Wang, and Lei Wang. VBS $W^{\pm}W^{\pm}H$ production at the HL-LHC and a 100 TeV pp-collider. *Int. J. Mod. Phys. A*, 32(18):1750106, 2017.

[22] Brian Henning, Davide Lombardo, Marc Riembau, and Francesco Riva. Measuring Higgs Couplings without Higgs Bosons. *Phys. Rev. Lett.*, 123(18):181801, 2019.

[23] HIG TTH. Study of Higgs boson production in association with top quarks final states with electrons, muons, and taus using the full Run 2 dataset. *CMS Physics Analysis Note*, AN-2019/111, 2019.

[24] Gregor Kasieczka, Benjamin Nachman, Matthew D. Schwartz, and David Shih. Automating the abcd method with machine learning. *Phys. Rev. D*, 103:035021, Feb 2021.

[25] CMS Collaboration. Egammaul2016to2018. "https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018".

[26] CMS Collaboration. Electronulscalefactors_may2022_update. "https://indico.cern.ch/event/1162757/contributions/4883762/attachments/2448910/4204941/ElectronULScaleFactors_May2022_Update(1).pdf".

[27] CMS Collaboration. Luminosity recommendations for run 2 analyses. https://twiki.cern.ch/twiki/bin/view/CMS/LumiRecommendationsRun2.

[28] CMS Collaboration. Muon recommendations for 2016 ultra legacy data and monte carlo. https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2016.

[29] CMS Collaboration. Muon recommendations for 2017 ultra legacy data and monte carlo. https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2017.

[30] CMS Collaboration. Muon recommendations for 2018 ultra legacy data and monte carlo. https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2018.

[31] CMS Collaboration. Muonulscalefactors_april2022_update. https://indico.cern.ch/event/1154936/contributions/4887489/attachments/2448253/4195313/MuonULScaleFactors_April2022_Update.pdf.

[32] CMS Collaboration. Reweighting recipe to emulate level 1 ecal and muon prefiring. https://twiki.cern.ch/twiki/bin/view/CMS/L1PrefiringWeightRecipe.

[33] CMS Collaboration. Utilities for accessing pileup information for data. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData.

[34] ttH Analysis Group. tth analysis documentation for the legacy analysis. https://github.com/TopEFT/topcoffea/tree/master/topcoffea/data/leptonSF/elec.

[35] ttH Analysis Group. tth analysis documentation for the legacy analysis. https://github.com/TopEFT/topcoffea/tree/master/topcoffea/data/leptonSF/muon.