University of California
Santa Barbara

Search for new physics in fully hadronic final states using the M_{T2} variable at a center-of-mass energy of 13 TeV with the CMS detector

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

Doctor of Philosophy
in
Physics

by

Jason Gran

Committee in charge:

Professor Claudio Campagnari, Chair
Professor Jeffrey D. Richman
Professor Mark Srednicki

June 2016

The Dissertation of Jason Gran is approved.

Professor Jeffrey D. Richman

Professor Mark Srednicki

Professor Claudio Campagnari, Committee Chair

June 2016

Search for new physics in fully hadronic final states using the M_{T2} variable at a center-of-mass energy of 13 TeV with the CMS detector

by

Jason Gran

Acknowledgements

I would like to thank my advisor, Claudio Campagnari, for all of his support and guidance during my five years at UCSB. I must also thank everyone else who has been a part of Claudio's group during my time at UCSB, including Derek Barge, Alex George, Frank Golf, Indara Suarez, Nick Amin, Sicheng Wang, and Bennett Marsh. As a member of the notorious Surf-n-Turf group, I also worked very closely with many people from UCSD and Fermilab. I could not have completed my thesis analysis without all of the work done by members of the MECCA super collaboration, comprised of professors, postdocs, and grad students from UCSB, UCSD, Fermilab, ETH Zurich, and CERN. It was a pleasure working with all of you and I couldn't have asked for a better group of friends and colleagues.

There are many others that have made my experience in Santa Barbara unforgettable. Jeff Richman and Harry Nelson taught me a great deal about particle physics during my first two years at UCSB. I played a huge amount of ping pong with David Jacobson, Curt Nehrkorn, and Ben LaRoque during our first year in the trailer. Speaking of Ben LaRoque, Ben and I shared many hours of coffee breaks over the years, discussing physics and how much better vim is than emacs.

Finally, I would like to thank my family. My parents Jeff and Brenda Gran gave me all the tools and support I needed to get to where I am today. My wife Kasey has been a constant source of love and encouragement during my graduate studies and I am excited about what the future holds for us.

Abstract

Search for new physics in fully hadronic final states using the M_{T2} variable at a center-of-mass energy of 13 TeV with the CMS detector

by

Jason Gran

An inclusive search for new physics in events with fully hadronic final states using the M_{T2} variable is presented. The results are based on a sample of 13 TeV center-of-mass energy proton-proton collisions, corresponding to 2.3 fb^{-1} of integrated luminosity collected with the CMS detector at the LHC in 2015. No excess event yield above the Standard Model background expectation is observed. The results are interpreted as limits on simplified models of supersymmetry, expressed as limits on the masses of potential new colored particles. Assuming a stable LSP with mass less than 500 GeV, gluino masses up to 1550–1750 are excluded at 95% confidence level, depending on the gluino decay mechanism. For models of direct squark pair production, top squarks with masses up to 800 GeV are excluded. Bottom squarks with masses up to 880 GeV are excluded. Light flavor squarks with masses up to 600–1260 GeV, depending on the degeneracy of squark masses, are excluded.

Contents

Chapter 1

Introduction

1.1 The current state of particle physics

The Standard Model (SM) [1–3] of particle physics is a description of the known fundamental particles and their non-gravitational interactions. The particles of the SM are divided into two types, fermions and bosons, based on their intrinsic angular momentum or "spin", where particles with half-integer spin are fermions and particles with integer spin are bosons. Fermions are further divided into two groups called quarks and leptons and are often thought of as the "matter" particles that interact via the exchange of "force mediating" bosons. The SM makes precise testable predictions about the nature of the fundamental particles and accurately explains a very rich variety of observed phenomena.

Despite the tremendous success of the SM, it is not a theory of everything. There are several problems with the SM and observed phenomena that it cannot explain. The SM describes strong and electroweak interactions between particles, but there is currently no consistent quantum theory of gravitational interactions between particles. Many theorists believe that there must be a new theory that takes over at high energy scales (near the Planck scale) where quantum gravitational effects are expected to be important and that

the SM is an effective theory that works well at the electroweak scale. This idea leads to other problems though, as the Planck scale is 10^{17} times larger than the electroweak scale. If there is new physics only at the Planck scale then this new physics should contribute very large terms to quantum loop diagrams, but this is inconsistent with what we observe.

One example of this problem is the Higgs boson mass, which receives loop corrections from all massive particles. If new physics on the order of the Planck scale contributed to these loop corrections then one would expect the Higgs boson mass to be near the Planck scale instead of the 125 GeV that we observe. One way to reconcile this discrepancy is to have extremely precise fine tuning of parameters in the high energy theory such that terms cancel out and the Higgs boson mass remains small. This solution seems unnatural and there are more elegant alternatives. One of the most popular theories that solves this problem is supersymmetry (SUSY) [4], which postulates a symmetry between fermions and bosons. With this symmetry all SM particles have a corresponding superpartner state that differs in spin by 1/2 and the contributions to the loop corrections of the Higgs boson mass would cancel to all orders. The superpartners have not been observed with identical masses to their SM counterparts, so SUSY must be a broken symmetry if it exists. It could be the case then that there exist some particles that we have not yet discovered that belong to some model of SUSY, where the phenomenology of the model depends on how the symmetry is broken.

Another, more concrete, reason to expect that there is more to the particle content of the universe than what is contained in the SM is the existence of dark matter. The existence of dark matter was first proposed in the 1930's [5] as an explanation of the discrepancy between the observed velocity distributions of stars in galaxies and the expected velocity distributions given the amount of observable matter. Since then, a large amount of additional astrophysical evidence for the existence of dark matter has been collected,

such as measurements of gravitational lensing. While the existence of dark matter is not in question, the exact nature of dark matter is unknown. The leading theories postulate that dark matter is composed of some number of new, non-SM, particles that interact very weakly with ordinary matter. There are a number of direct detection experiments that aim to detect dark matter particles as they travel through the Earth, as well as accelerator-based experiments that aim to produce dark matter and infer its presence as missing energy. See Ref. [6] for more information about the astrophysical evidence for dark matter and the experimental efforts to detect dark matter.

The current state of particle physics is that we are looking for clues to help solve problems that the SM cannot solve. The SM can explain a huge amount of the physics we observe, but it seems like there must be something beyond the SM. In the next sections I will briefly describe one of the experimental setups used to look for these clues.

1.2 The Large Hadron Collider

The Large Hadron Collider (LHC) is currently the world's largest and highest energy particle collider. It was designed to look for the Higgs boson and to search for physics beyond the SM at the TeV scale. The LHC is located at CERN, the European Organization for Nuclear Research, near Geneva, Switzerland. The LHC was built over a period of about 10 years starting in 1998 and is housed in the circular tunnel that used to be occupied by the Large Electron Positron (LEP) collider. The LHC is located about 100 m underground with a circumference of 27 km. The LHC is the final stage in the series of linear and circular accelerators that make up the CERN accelerator complex. In this final stage, two beams of protons (or lead ions) circulate in opposite directions around the LHC ring in separate beam pipes. At four locations around the ring the two beams are steered into a collision point, around which an experiment is placed. The corresponding

four experiments are ATLAS, CMS, LHCb, and ALICE. ATLAS and CMS are general purpose experiments capable of detecting many types of particles and have broad physics programs. The LHCb experiment focuses on b-physics and ALICE focuses on heavy ion physics.

The LHC is currently delivering proton-proton collisions with an energy of 6.5 TeV per beam for a total center-of-mass energy of 13 TeV. The protons that make up each beam of the LHC are injected in bunches of $\sim 10^{11}$ protons. The spacing of the bunches is 25 ns, meaning a bunch from one beam crosses through a bunch from the other beam 40 million times per second. In an average bunch crossing there are about 20 proton-proton interactions. This results in more than half a billion collision events every second. Figure 1.1 shows the production cross section and expected number of events per second for several SM processes as a function of the collison energy. More information about the design and performance of the LHC is available at Ref. [7].

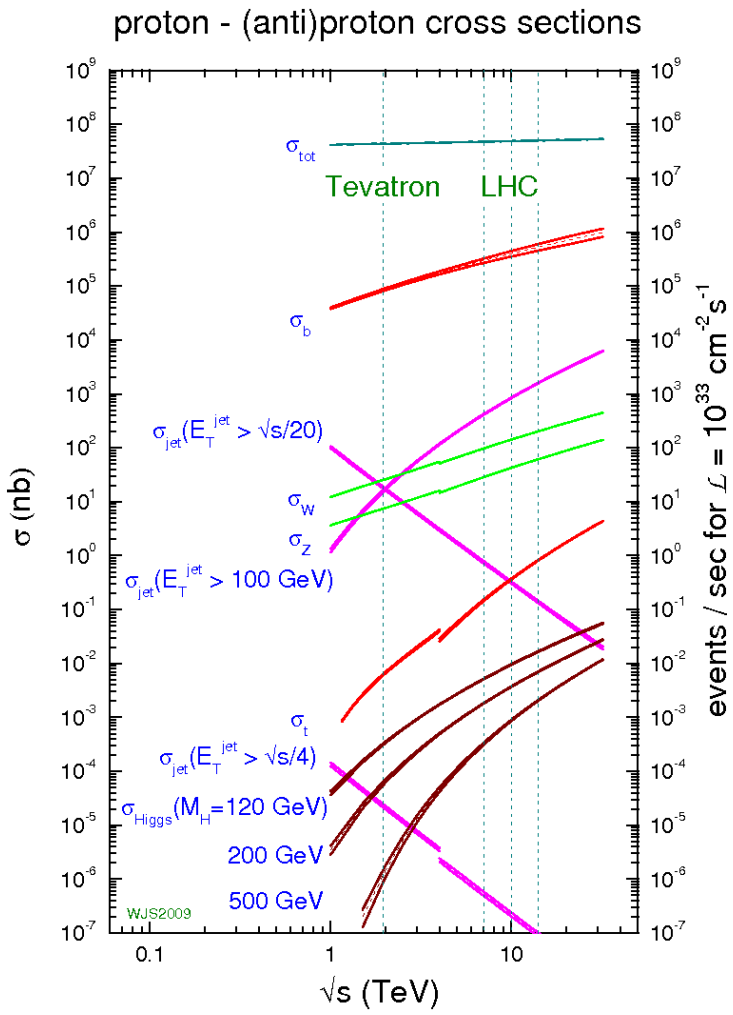

Figure 1.1: The production cross-sections of some SM processes as a function of center-of-mass energy [8].

1.3 The Compact Muon Solenoid experiment

The Compact Muon Solenoid (CMS) experiment is a general purpose experiment designed to analyze collisions produced by the LHC and study whatever TeV-scale physics presents itself. The "Solenoid" in Compact Muon Solenoid is a superconducting solenoid, 13 m in length and 6 m in diameter. This solenoid creates an axial magnetic field of 3.8 T

and causes charged particles to travel in a helical trajectory from which momentum can be calculated. Within the magnetic field volume are several particle detection systems used to measure particle trajectories and energies.

Charged particle trajectories are measures with the tracker, composed of silicon pixels and strips. The tracker covers $0 \leqslant \phi < 2\pi$ in azimuth and $|\eta| < 2.5$ in pseudorapidity, where $\eta \equiv -\ln(\tan(\theta/2))$ and θ is the polar angle of the trajectory of the particle with respect to the counterclockwise beam direction. Outside of the tracker is the electromagnetic calorimeter (ECAL) which is composed of lead-tungstate crystals. Outside of the ECAL is the hadronic calorimeter (HCAL). Together the ECAL and HCAL provide energy measurements of electrons, photons, and hadronic jets in the detector volume corresponding to $|\eta| < 3.0$. Muons within $|\eta| < 2.4$ are identified and measured by gas-ionization detectors embedded in the steel flux-return yoke of the solenoid. Particles that travel at a small angle to the beamline are detected by forward calorimeters at both ends of CMS that cover $3.0 < |\eta| < 5.0$. The coverage provided by these detector components makes CMS nearly hermetic, allowing for a measurement of momentum imbalance in the plane transverse to the beam direction. This momentum imbalance measurement can be used to infer the presence of undetected particles.

A cross-sectional view of the CMS detector is shown in Figure 1.2, with labels for the major subdetectors. A more detailed description of the CMS experiment is available in Ref. [9], including an in-depth description of the detector components, trigger, and data aquisition system.

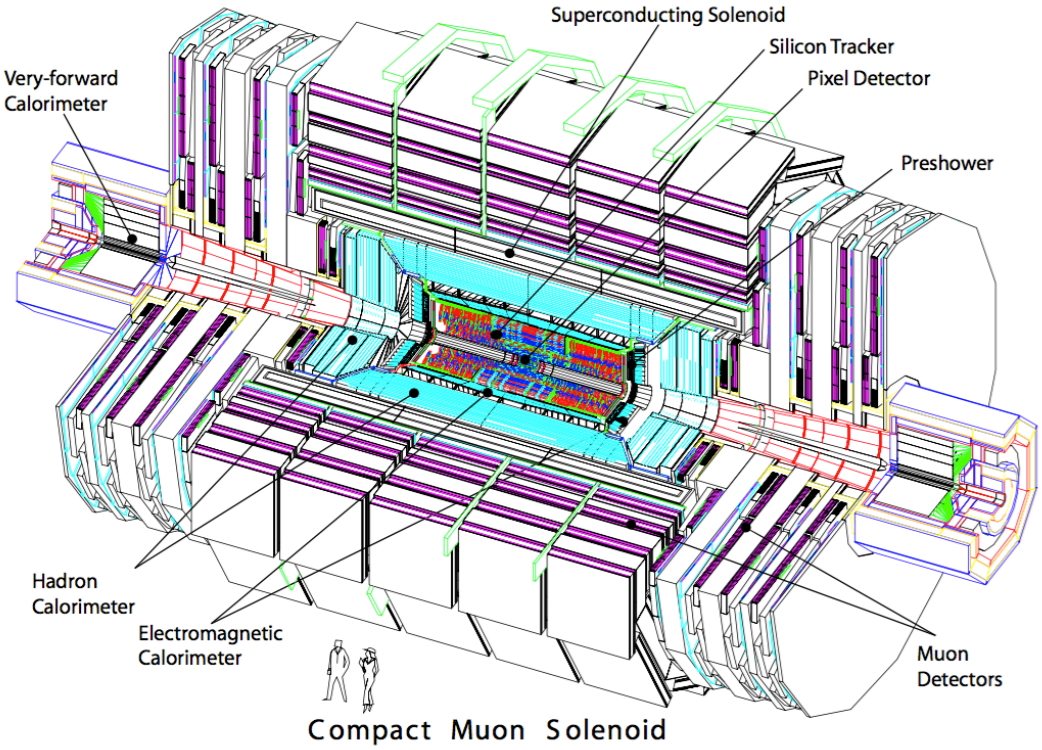

Figure 1.2: A cross-sectional view of the CMS detector [9].

Chapter 2

Searching For New Physics With the M_{T2} Variable

This chapter introduces the M_{T2} variable and describes its use in the analysis. Section 2.1 defines the M_{T2} variable and how it is used. Section 2.2 describes the use of M_{T2} in the case of fully hadronic events. Section 2.3 describes the background suppression provided by the M_{T2} variable.

2.1 Introduction to the M_{T2} variable

To define M_{T2}, we must first define the transverse mass variable, M_{T}. The transverse mass variable is commonly used to measure the mass of a particle that undergoes a two-body decay where one of the decay products is invisible. It was first used to measure the mass of the W boson when it was discovered in 1983 [10]. In the case in which a particle undergoes a two-body decay where one of the decay products is invisible, one does not have enough information to compute the invariant mass of the parent particle. We can measure all components of the momentum of the visible system, but only two of the

three components of the momentum of the invisible particle, since it manifests as missing transverse momentum. Using the information we do have, we define the transverse mass as follows.

$$\begin{aligned}(M_{\mathrm{T}})^2 &= (M_{\mathrm{vis}})^2 + (M_{\chi})^2 + 2\left(E_{\mathrm{T}}^{\mathrm{vis}} E_{\mathrm{T}}^{\chi} - \vec{p}_{\mathrm{T}}^{\,\mathrm{vis}} \cdot \vec{p}_{\mathrm{T}}^{\,\chi}\right) \\ &\approx 2p_{\mathrm{T}}^{\mathrm{vis}} p_{\mathrm{T}}^{\chi}(1 - \cos\Delta\phi)\end{aligned} \tag{2.1}$$

Here the visible system is denoted by "vis" and the invisible system is denoted by "χ". The approximation in the second line of Eq. 2.1 holds in the case that $M_{\mathrm{vis}} \approx M_{\chi} \approx 0$. The quantity $\Delta\phi$ is the azimuthal angle between $\vec{p}_{\mathrm{T}}^{\,\mathrm{vis}}$ and $\vec{p}_{\mathrm{T}}^{\,\chi}$. The transverse mass has a kinematic endpoint at the mass of the mother particle, creating a sharp edge in the distribution, as shown in Figure 2.1 for $\mathrm{W} \to \mu\nu$ events.

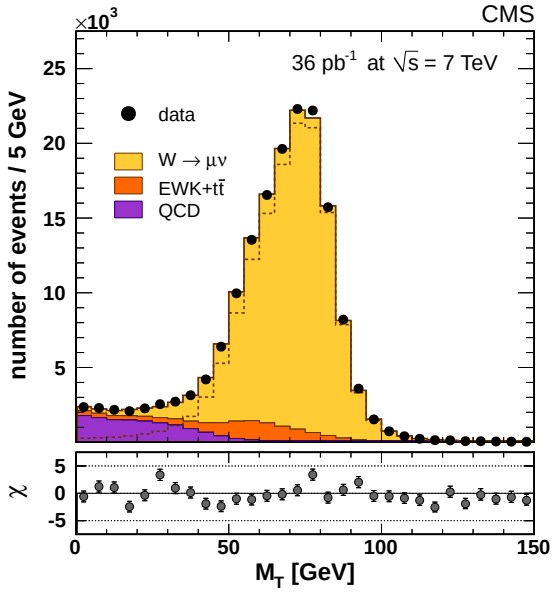

Figure 2.1: Transverse mass distribution for $\mathrm{W} \to \mu\nu$ events. There is a sharp edge near $\mathrm{m_W} \approx 80$ GeV. Taken from Ref. [11].

The M_{T2} variable, sometimes called stransverse mass, is an extension of M_{T} to the

case where there are a pair of particles that each decay semi-invisibly as shown in Figure 2.2. In such a case there are two visible decay systems with known momenta and two invisible decay products with unknown momenta since we can only measure the net missing transverse momentum but not how it is partitioned between the two invisible decay products.

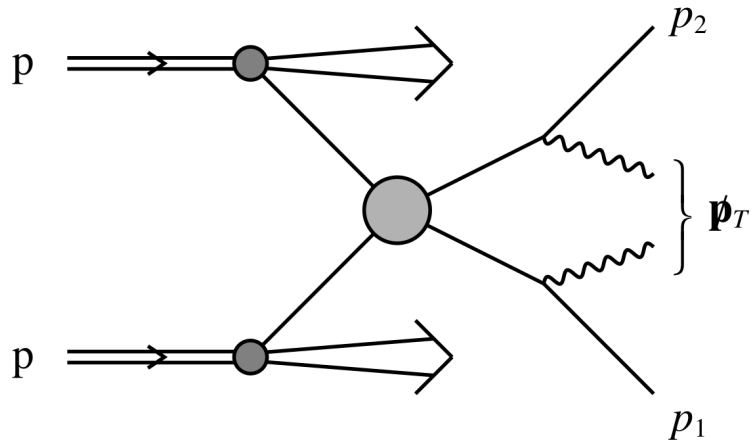

Figure 2.2: Schematic diagram of pair production at a hadron collider where each particle in the pair decays to something visible and something invisible. The momenta of the visible systems can be measured. The sum of the transverse momenta of the invisible decay products is measured as the total missing momentum. Taken from Ref. [12].

Using the available information, we define M_{T2} as follows:

$$M_{\mathrm{T2}}(M_\chi) = \min_{\vec{p}_{\mathrm{T}}^{\,\chi(1)}+\vec{p}_{\mathrm{T}}^{\,\chi(2)}=\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}} \left[\max\left(M_{\mathrm{T}}^{(1)}, M_{\mathrm{T}}^{(2)}\right)\right] \tag{2.2}$$

where $M_{\mathrm{T}}^{(1)}$ and $M_{\mathrm{T}}^{(2)}$ are the transverse masses of each decay chain for some choice of momenta $\vec{p}_{\mathrm{T}}^{\,\chi(1)}$ and $\vec{p}_{\mathrm{T}}^{\,\chi(2)}$ that sum to the total missing momenta. For the true values of $\vec{p}_{\mathrm{T}}^{\,\chi(1)}$ and $\vec{p}_{\mathrm{T}}^{\,\chi(2)}$, $M_{\mathrm{T}}^{(1)}$ and $M_{\mathrm{T}}^{(2)}$ will both be less than or equal to the parent mass. By minimizing $\max(M_{\mathrm{T}}^{(1)}, M_{\mathrm{T}}^{(2)})$ over all possible choices of $\vec{p}_{\mathrm{T}}^{\,\chi(1)}$ and $\vec{p}_{\mathrm{T}}^{\,\chi(2)}$ we ensure that M_{T2} is less than or equal to the parent mass. The mass of the unseen decay product,

M_χ, is used in the calculation of the transverse mass (see equation 2.1) as part of the M_{T2} calculation. This mass cannot be measured and is therefore a free parameter in the M_{T2} definition. This analysis uses $M_\chi = 0$. If the value of M_χ is larger (smaller) than the true value, then the endpoint of the M_{T2} distribution will be above (below) the mass of the pair produced particle.

There are several methods available to compute M_{T2}, of varying speed and precision. The method used in this analysis is called the bisection method. See Ref. [13] for a description of the bisection method and a link to the C++ implementation.

While first introduced [12] as a way to measure the mass of pair-produced particles that may be found at the LHC, M_{T2} is also very useful as a discovery variable itself. As discussed in [14], SM backgrounds to searches for new physics at the LHC are largely suppressed for M_{T2} greater than m_{top}, while the M_{T2} distribution for events with pair production of a new heavy particle will extend to much larger values, providing regions with great discovery potential. Several searches for new physics at the LHC have been performed utilizing M_{T2} as a discovery variable. The strategy used in this analysis is built upon the previous versions of this analysis done at 7 TeV [15] and 8 TeV [16].

2.2 Visible system for multijet events

This analysis uses the M_{T2} variable to search for pair production of a new particle that decays semi-invisibly where the visible decay products contain some number of hadronic jets. In order to calculate M_{T2} for such events we must define two visible systems. The visible systems are obvious for events with only two jets. For events with more than two jets we define the visible systems used in the M_{T2} calculation by clustering all the jets in an event into two hemispheres or "pseudojets" with the intention that each psueudojet contains all the visible particles from one of the parent particles.

To begin the clustering, the two jets with the largest invariant mass are chosen as the seed axes for the pseudojets. Then all other jets are assigned to one of the pseudojets, where jet k is assigned to pseudojet i rather than pseudoejet j if:

$$(E_i - p_i \cos\theta_{ik}) \frac{E_i}{E_i + E_k} \leq (E_j - p_j \cos\theta_{jk}) \frac{E_j}{E_j + E_k} \tag{2.3}$$

where here we are comparing the energy-weighted hemisphere masses, a quantity known as the *Lund distance measure*. The quantity θ_{ik} is the angle between the axis of hemisphere i and jet k. Once all jets are assigned, the pseudojet axes are updated by summing the momenta of all constituent jets. This procedure of assigning jets to one of the pseudojets and then updating the pseudojet axes is repeated until no jets switch from one pseudojet to the other. The resulting pseudojets are used as the visible systems in the M_{T2} calculation.

2.3 $\mathbf{M_{T2}}$ for background processes

One of the main challenges to obtaining sensitivity to new physics signals that decay to hadronic final states is the very large cross section of QCD production of multijet events. Many inclusive searches for new physics require a large amount of missing energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) in order to suppress the QCD multijet background, where the dominant source of $E_{\mathrm{T}}^{\mathrm{miss}}$ is jet mismeasurement. By examining the properties of the M_{T2} variable, we can see that M_{T2} provides excellent suppression of QCD multijet events, often better than using $E_{\mathrm{T}}^{\mathrm{miss}}$ alone.

For a QCD dijet event with large $E_{\mathrm{T}}^{\mathrm{miss}}$, the pseudojets making up the two visible systems will be back-to-back and the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector will be aligned with one of the pseudojets. In this case the global minimum of the transverse masses in the M_{T2} calculation will be

equal to M_{vis}, or the mass of the pseudojet. If we choose to use massless pseudojets (set $E_{\text{T}}^{\text{vis}} = p_{\text{T}}^{\text{vis}}$), then $M_{\text{T2}} = 0$ for perfectly back-to-back dijet events. For this reason, and also so that QCD events with little $E_{\text{T}}^{\text{miss}}$ cannot have large M_{T2} values, we choose to use massless pseudojets in this analysis. M_{T2} will also be close to zero in the case where there are more than two jets but the pseudojets are close to back-to-back. In this sense, the M_{T2} variable provides protection against large missing energy from jet mismeasurement.

In contrast to QCD multijet events, signal events are expected to be asymmetric with large real $E_{\text{T}}^{\text{miss}}$ (i.e., not from jet mismeasurement) and pseudojets that are not back-to-back, causing these events to have large M_{T2} values. The same is true for SM processes that have real $E_{\text{T}}^{\text{miss}}$ from neutrinos, such as $\text{Z} \to \nu\nu$, $\text{W} + \text{jets}$, and $\text{t}\bar{\text{t}}$. It is these processes that make up the majority of the background to this search, with QCD being largely suppressed by requiring a large M_{T2} value.

Figure 2.3 shows the M_{T2} distribution in simulation after the baseline selections described in section 3.3 for QCD, $\text{Z} \to \nu\nu$, $\text{W} + \text{jets}$, $\text{t}\bar{\text{t}}$, and several example SUSY signal models. From this we can see that indeed, QCD falls very quickly with increasing M_{T2}, while events with real $E_{\text{T}}^{\text{miss}}$ extend to large M_{T2}. The strategy for this analysis is to set a minimum M_{T2} threshold to reject a large amount of QCD and then bin in M_{T2} to provide low background regions where a possible signal might be found. The full details of the analysis selection are provided in the next chapter.

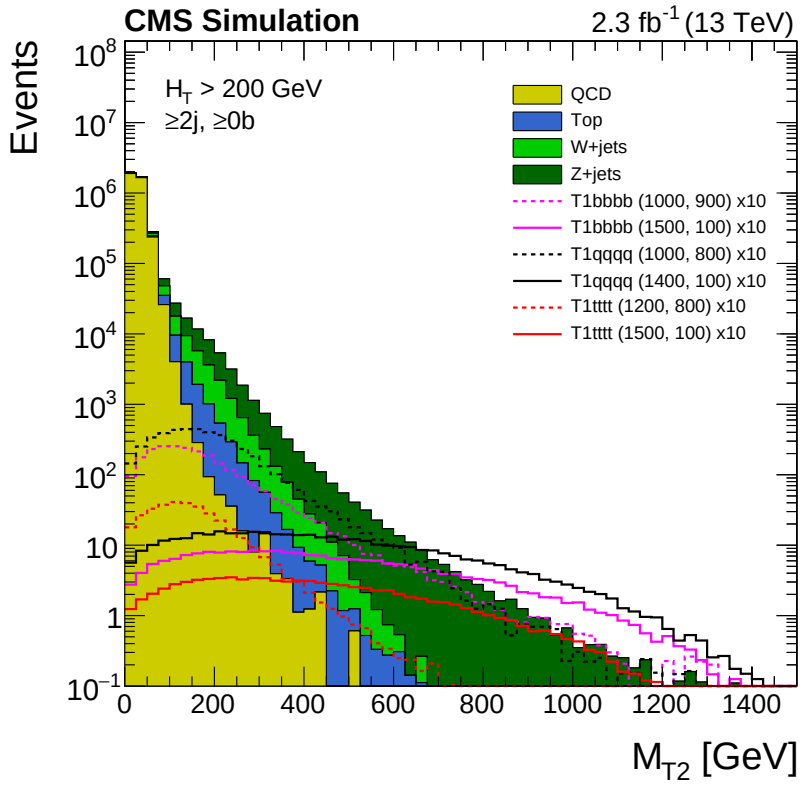

Figure 2.3: Inclusive distribution of M_{T2} after the baseline analysis selection. The SM backgrounds are stacked and example signal points are overlaid with their cross sections scaled up by a factor of 10. Only events with at least two jets are selected. T1bbbb, T1qqqq, and T1tttt represent possible SUSY signal models to be further discussed in chapter 8.

Chapter 3

Event Selection

While this analysis is built around the M_{T2} variable introduced in Chapter 2, additional variables are also used to suppress SM background processes and categorize events. This chapter describes these additional variables and the physics objects (jets, leptons, etc.) from which they are constructed. After defining the relevant objects and variables, I describe how they are used to construct a baseline selection and categorize the remaining events into signal regions.

3.1 Object and variable definitions

Collision events in the CMS detector are reconstructed with using particle-flow (PF) event reconstruction [17]. This reconstruction software aims to identify the particles in a collision by using information from all subdetectors. In the following, the identification criteria are listed for the objects used in this analysis.

3.1.1 Jets

We use ak4CHS jets with the Summer15_25nsV5 jet energy corrections applied. These are jets clustered with the anti-k$_t$ algorithm [18] using a radius of 0.4 where charged particles from non-primary vertices (pileup) are removed before clustering. Jets are required to pass $p_{\mathrm{T}} > 30$ GeV, $|\eta| < 2.5$, and PF jet loose ID [19]. For events with only one jet, we require the jet to satisfy tighter ID requirements:

- PF jet tight ID [19]
- Charged Hadron Fraction > 0.05
- Neutral Hadron Fraction < 0.8
- Neutral EM Fraction < 0.7

These tighter requirements are necessary to reject noise events in data. The requirements on the various energy fractions, for example, remove fake jets arising from spurious energy depositions in a single sub-detector

We identify b-quark jets using the CSVv2IVF algorithm, which is based on the lifetime of b-hadrons. The medium working point is used, corresponding to a cut value of 0.89 [20]. B-tagged jets are required to pass a lower requirement on p_{T}, $p_{\mathrm{T}} > 20$ GeV. This helps adding extra sensitivity to compressed spectrum signals with jets from b-quarks.

3.1.2 MET

The initial momentum of the partons involved in a collision is aligned with the beam axis, with no component transverse to the beam. Using conservation of momentum in the transverse plane we can infer how much missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) there is

in an event, which might correspond to energy carried by undetected particle. $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as:

$$E_{\mathrm{T}}^{\mathrm{miss}} = -\Sigma_i \vec{p}_{T,i} \tag{3.1}$$

where the sum is over the particles in a collision event and $\vec{p}_{T,i}$ is the transverse momentum of the ith particle.

We additionally calculate $H_{\mathrm{T}}^{\mathrm{miss}}$, defined as:

$$H_{\mathrm{T}}^{\mathrm{miss}} = |-\Sigma_j \vec{p}_j| \tag{3.2}$$

where the sum is over the momenta of all jets passing the object requirements above. $H_{\mathrm{T}}^{\mathrm{miss}}$ is different from $E_{\mathrm{T}}^{\mathrm{miss}}$ in that it does not include unclustered energy or energy from soft jets.

3.1.3 MET filters

There are multiple sources of "noise" that may result in an event having a large, incorrect value of $E_{\mathrm{T}}^{\mathrm{miss}}$. We use filters to remove these noisy events that may otherwise enter our search regions. We apply the filters recommended by the JetMET Physics Object Group (POG) [21]:

- primary vertex filter
- CSC beam halo filter
- HBHE noise filter
- HBHE iso noise filter
- ee badSC noise filter

The technical details of these filters are not important, just that they remove undesirable events, so I will not go into further detail here.

We also reject any event containing a jet with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 4.7$ which fails the PF jet loose ID. Such a jet would not be included in the pseudojet clustering, causing an imbalance in the pseudojets, leading to larger values of M_{T2}.

3.1.4 Electrons

The following definition of electron is used in a veto to remove events with leptons from our search regions and to select single lepton events to populate control regions used to esimate the background from events with lost leptons. Electron candidates are required to pass $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.4$. We select electrons passing the POG Veto working point, as defined for Spring15 25ns [22]. The cuts applied are:

- $\sigma_{i\eta i\eta} < 0.0114$ (0.0352 for endcap).

  This variable is calculated using all the RecHits in the 5x5 seed cluster and describes the width of the shower shape in the ECAL.

- $|\Delta\eta_{In}| < 0.0152$ (0.0113 for endcap).

  This is the difference in η between the supercluster and corresponding track.

- $|\Delta\phi_{In}| < 0.216$ (0.237 for endcap).

  This is the difference in ϕ between the supercluster and corresponding track.

- $H/E < 0.181$ (0.116 for endcap).

  This is the ratio of the energy deposited in the HCAL tower just behind the ECAL seed cluster over the energy of the seed cluster. An electron should deposit most of its energy in the ECAL, so an upper limit on this variable is used.

- $|\frac{1}{E} - \frac{1}{p}| < 0.207$ (0.174 for endcap).

  This is a measure of the energy-momentum matching for the supecluster and corresponding track.

- $|d0| < 0.0564$ cm (0.222 for endcap).

  This is the transverse impact parameter with respect to primary vertex.

- $|dz| < 0.472$ cm (0.921 for endcap).

  This is the longitudinal impact parameter with respect to primary vertex.

- number of expected missing inner hits ≤ 2 (3 for endcap).

- conversion veto to reject photon conversions to e^+e^- pairs.

where electrons with $|\eta_{\mathrm{SC}}| > 1.479$ are considered to be in the endcap.

We require electrons to be isolated, cutting on relative mini PF isolation: miniIso/$p_{\mathrm{T}} < 0.1$. Here mini isolation means that the cone size decreases with p_{T}, according to equation 3.3. An effective area pileup correction is applied.

$$\Delta \mathrm{R} = \begin{cases} 0.2 & \text{if } p_{\mathrm{T}} < 50 \text{ GeV} \\ 10 \text{ GeV}/p_{\mathrm{T}} & \text{if } 50 < p_{\mathrm{T}} < 200 \text{ GeV} \\ 0.05 & \text{if } p_{\mathrm{T}} > 200 \text{ GeV} \end{cases} \tag{3.3}$$

3.1.5 Muons

This definition of muon is used in a veto to remove events with leptons from our search regions and to select single lepton events to populate control regions used to esimate the background from events with lost leptons. Muon candidates are required to have $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.4$. We select muons passing the POG Loose working point [23]. The cuts applied are:

- isPFMuon.

  A particle candidate identified as a muon in the CMS Particle Flow Algorithm.

- isGlobalMuon || isTrackerMuon.

  A GlobalMuon candidate is seeded from a track in the muon detectors and then a global fit is used to take into account the muon detector, silicon strip and pixel information. A TrackerMuon is reconstructed by considering all tracker tracks to be potential muon candidates and checking this hypothesis by looking for compatible signatures in the calorimeters and in the muon system.

In addition, we apply a selection on the impact parameter:

- $|d0| < 0.2$ cm, with respect to primary vertex.
- $|dz| < 0.5$ cm, with respect to primary vertex.

We require muons to be isolated, cutting on relative mini PF isolation: miniIso/$p_{\mathrm{T}} < 0.2$, using the definition of cone size from Equation 3.3. An effective area pileup correction is applied.

3.1.6 Isolated tracks

We use "isolated tracks" to identify low p_{T} electrons and muons, as well as tau leptons decaying to leptons or hadrons in order to improve our efficiency for vetoing leptonic events. We select charged PF candidates with different requirements depending on the flavor. For PF electrons and PF muons, we require them to pass $p_{\mathrm{T}} > 5$ GeV, $|dz| < 0.1$ cm, and a track isolation cut of iso/$p_{\mathrm{T}} < 0.2$. The track isolation sum is computed from all charged PF candidates within a cone of $\Delta\mathrm{R} < 0.3$, requiring them to pass $|dz| < 0.1$ cm with respect to the primary vertex. For charged PF hadrons,

we require them to pass $p_{\mathrm{T}} > 10$ GeV, $|dz| < 0.1$ cm, and a track isolation cut of iso$/p_{\mathrm{T}} < 0.1$. The track isolation is computed in the same way as for PF leptons above.

3.1.7 Photons

This definition of photon is used to select $\gamma +$ jets events as part of the estimation of the background to the search from $\mathrm{Z} \to \nu\nu$ events. Photon candidates are required to pass $p_{\mathrm{T}} > 170$ GeV and $|\eta| < 2.5$. We select photons passing the Spring15 POG Loose working point with a customized isolation cut that facilitates the background estimation. For details see Section 4.1. The cuts applied are:

- No pixel seed.

  Photons do not leave hits in the tracker, so it is required that there are no pixel seeds consistent with the position of the ECAL deposits.

- $\sigma_{i\eta i\eta} < 0.0103$ (0.0277 for endcap).

  This variable is calculated using all the RecHits in the 5x5 seed cluster and describes the width of the shower shape in the ECAL.

- $H/E < 0.05$ (0.05 for endcap). This is the ratio of the energy deposited in the HCAL tower just behind the ECAL seed cluster over the energy of the seed cluster. An electron should deposit most of its energy in the ECAL, so an upper limit on this variable is used.

- `chargedHadronIso`$<$ 2.5 GeV, where the `reco::Photon::chargedHadronIso()` variable is used to define the isolation sum.

3.1.8 $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathbf{miss}})$

The variable $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ is defined as the minimum $\Delta\phi$ between $E_{\mathrm{T}}^{\mathrm{miss}}$ and any of the four highest p_{T} jets in the event. For this variable only, we consider jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 4.7$.

3.1.9 Vertices

We consider a reconstructed vertex as good if it satisfies:

- not Fake.

  If no reconstructed vertex is found in an event, a vertex based on the beam-spot is put into the event and is labeled a Fake vertex.

- $N_{dof} > 4$.

  N_{dof} is the number of degrees of freedom in the fit for the position of a vertex based on its associated tracks and is strongly correlated with the number of tracks compatible with arising from the interaction region.

- $|z| < 24$ cm.

  This is the longitudinal distance of the vertex from the nominal interaction point.

- $\rho < 2$ cm.

  This is the distance from the beam axis.

When more than one good reconstructed vertex is found in the event, we use the one with largest $\sum p_{\mathrm{T}}^2$, where the sum runs over all tracks associated to that vertex. This means that all tracks associated to other vertexes are neglected when building jets or isolation variables (known as charged hadron subtraction, or CHS) and the four-momenta of neutral candidates originate from there.

3.2 Signal triggers

This analysis aims to be as inclusive as possible while searching for evidence of new physics processes with jets and no leptons. In the early stages of the development of this analysis it was crucial to develop a trigger strategy that provided access to a wide range of interesting phase space. One way to think about this phase space is to consider the two-dimensional $E_{\mathrm{T}}^{\mathrm{miss}}$– H_{T} plane and how possible new physics events might populate this plane. Events with some compressed new particle spectrum may have few high p_{T} jets and little $E_{\mathrm{T}}^{\mathrm{miss}}$. Events with large splittings in the new particle spectrum could have many high p_{T} jets and large values of $E_{\mathrm{T}}^{\mathrm{miss}}$.

Thinking in this way, we aimed to cover as much as the $E_{\mathrm{T}}^{\mathrm{miss}}$– H_{T} plane as possible when developing the trigger strategy in order to be sensitive to a wide variety of possible new phenomena. There are several limitations that must be taken into account when developing a reasonable trigger, including the trigger rate, how long it takes to make the trigger decision, and politics. The amount of bandwidth a trigger gets is some function of these limitations. For example, it is not reasonable to trigger on every event with at least 10 GeV of $E_{\mathrm{T}}^{\mathrm{miss}}$ because the rate would be much larger than the rate at which the experiment can afford to fully reconstruct and store events. The three signal triggers that were developed and used for this analysis, and their coverage in the $E_{\mathrm{T}}^{\mathrm{miss}}$– H_{T} plane, are shown in Figure 3.1. The names of these triggers and their corresponding selections are given in Table 3.1.

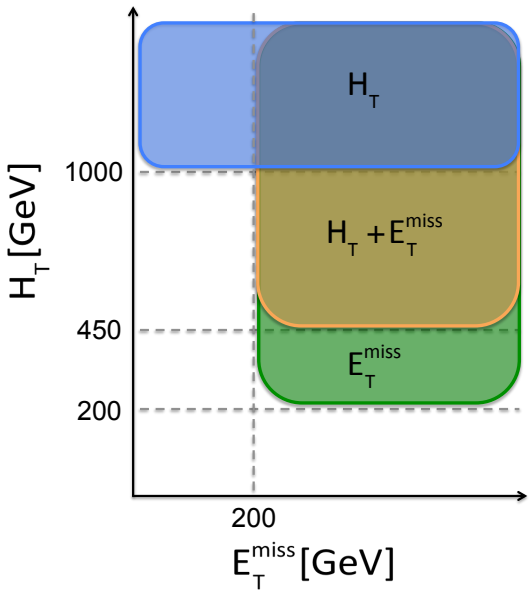

Figure 3.1: Coverage of the H_{T}–$E_{\mathrm{T}}^{\mathrm{miss}}$ plane provided by the signal triggers.

trigger	online selection [GeV]	offline selection [GeV]
`HLT_PFHT800`	$H_{\mathrm{T}} > 800$	$H_{\mathrm{T}} > 1000$
`HLT_PFMETNoMu90_JetIdCleaned_PFMHTNoMu90_IDTight`	$E_{\mathrm{T}}^{\mathrm{miss}} > 90$	$E_{\mathrm{T}}^{\mathrm{miss}} > 200$
`HLT_PFHT350_PFMET100_NoiseCleaned`	$H_{\mathrm{T}} > 350$ $E_{\mathrm{T}}^{\mathrm{miss}} > 100$	$H_{\mathrm{T}} > 450$ $E_{\mathrm{T}}^{\mathrm{miss}} > 200$

Table 3.1: Signal triggers. The online selection is the set of cuts applied on High Level Trigger (HLT) objects before the event is fully reconstructed. The offline selection is the set of analysis cuts used for events coming from these triggers.

3.3 Baseline selection

We apply the following baseline selection to all events used in the analysis signal regions.

- `HLT_PFHT800` OR `HLT_PFHT350_PFMET100_NoiseCleaned`

  OR `HLT_PFMETNoMu90_JetIdCleaned_PFMHTNoMu90_IDTight`: After all baseline cuts,

the efficiency of these triggers is found to be 100% for the H_{T} selection and $> 97\%$ for the $E_{\mathrm{T}}^{\mathrm{miss}}$ selection.

- at least one good vertex
- at least one jet
- $H_{\mathrm{T}} > 1000$ GeV && $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV, or $H_{\mathrm{T}} > 200$ GeV && $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV: these requirements are based on the triggers above. At high H_{T}, a minimal $E_{\mathrm{T}}^{\mathrm{miss}}$ cut is applied as a safety cut for the M_{T2} variable.
- $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}}) > 0.3$: this protects against large $E_{\mathrm{T}}^{\mathrm{miss}}$ from jet mismeasurement in events where the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector is aligned with one of the jets. Overall, the cut rejects a large fraction of QCD events.
- $|\vec{H}_{\mathrm{T}}^{\mathrm{miss}} - \vec{E}_{\mathrm{T}}^{\mathrm{miss}}|/E_{\mathrm{T}}^{\mathrm{miss}} < 0.5$: this protects against a bias in the shape of M_{T2} (especially for QCD events). $E_{\mathrm{T}}^{\mathrm{miss}}$ is sensitive to reconstructed objects with $p_{\mathrm{T}} < 30$ GeV or $|\eta| > 2.4$ whereas these are not used in the pseudojets for M_{T2}, so a large contribution from these objects out of p_{T} or η acceptance can bias the M_{T2} distribution.
- lepton vetoes: to reduce the background from events with a W boson decay, we reject events if they contain:
  - a reco electron or muon candidate as defined in Sections 3.1.4 and 3.1.5
  - a PF electron, muon, or hadron candidate as defined in Section 3.1.6, with the additional requirement of $M_{\mathrm{T}}(\mathrm{cand}, E_{\mathrm{T}}^{\mathrm{miss}}) < 100$ GeV
- $M_{\mathrm{T2}} > 200$ GeV, only for events with at least two jets. This cut provides a large rejection of QCD events (as shown in Figure 3.2).

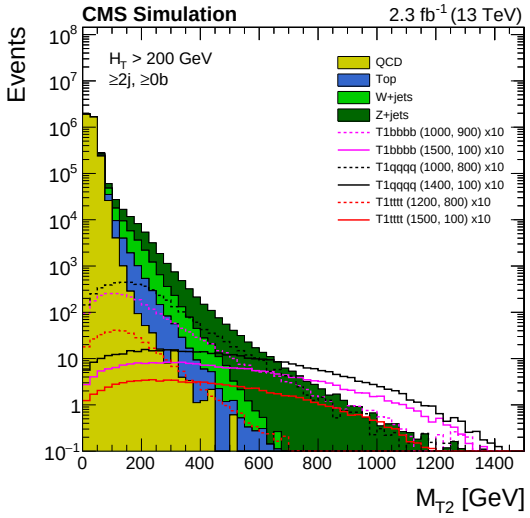

Figure 3.2: Inclusive distribution of M_{T2}. The SM backgrounds are stacked and example signal points are overlaid with their cross sections scaled up by a factor of 10. The signal models are described in Chapter 8. Only events with at least two jets are selected.

3.4 Signal region definitions

Following the baseline selection above, we categorize events in different bins according to the following variables: H_{T}, N_{jets}, $N_{\mathrm{b-tags}}$ and M_{T2}. Figure 3.3 shows distributions of these variables in simulation after the baseline selection of Section 3.3 for SM backgrounds, along with a few overlaid signal models.

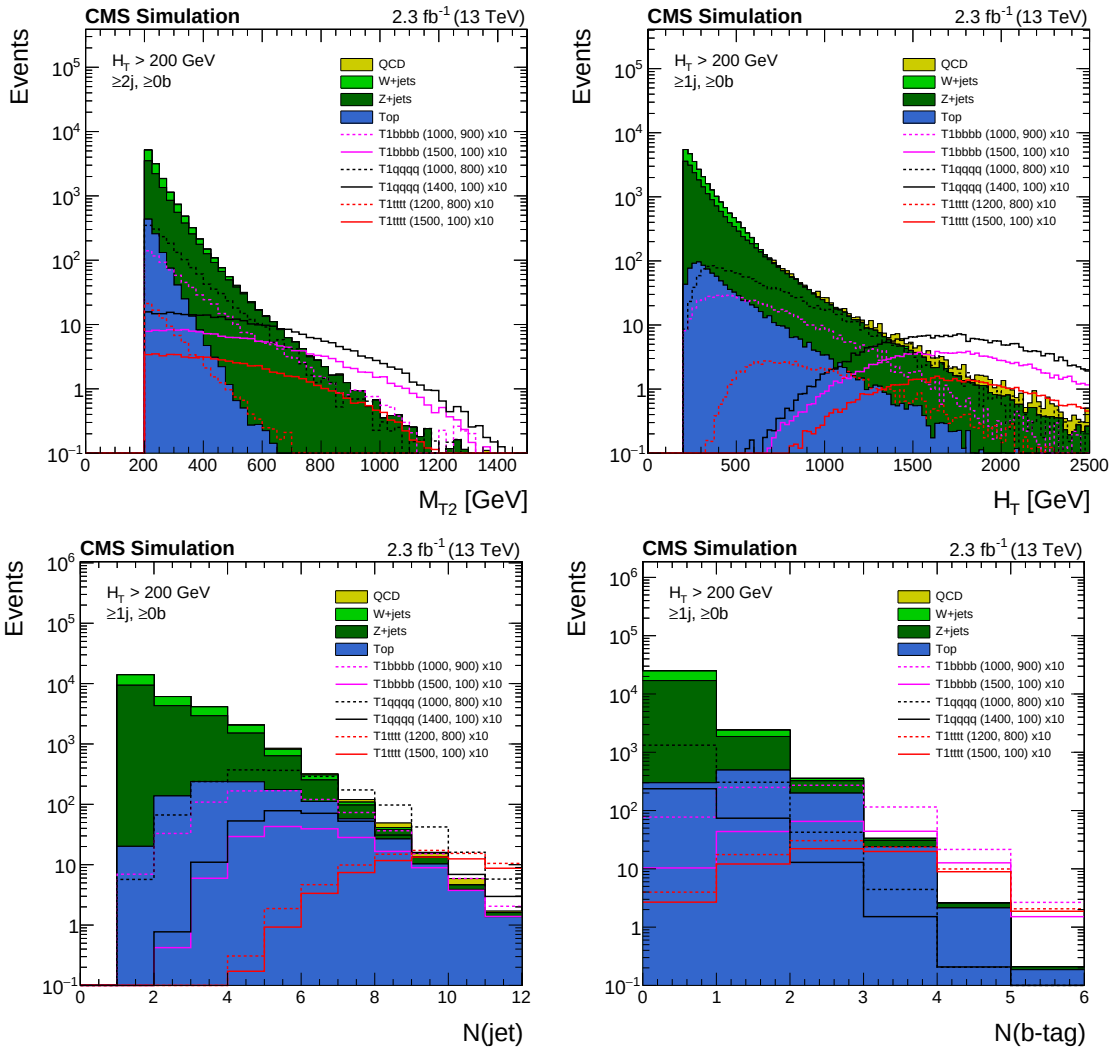

Figure 3.3: Distributions of the variables used for binning the signal regions after the baseline selection. The SM backgrounds are stacked and example signal points are overlaid with their cross sections scaled up by a factor of 10. The signal models are described in Chapter 8. For M_{T2}, only events with at least two jets are selected.

First, we categorize events in H_{T}, N_{jets} and $N_{\mathrm{b-tags}}$:

- 5 bins in H_{T}: [200, 450], [450,575], [575,1000], [1000, 1500], [1500, ∞].

  These bins are also referred to as Very Low H_{T}, Low H_{T}, Medium H_{T}, High H_{T} and Extreme H_{T} regions.

- 11 bins in N_{jets} and $N_{\text{b-tags}}$: 2–3j 0b, 2–3j 1b, 2–3j 2b, 4–6j 0b, 4–6j 1b, 4–6j 2b, ≥7j 0b, ≥7j 1b, ≥7j 2b, 2–6j ≥3b, ≥7j ≥3b

These "topological regions" as defined above in N_{jets} and $N_{\text{b-tags}}$ are shown in Figure 3.4, together with an example of background composition for the Medium H_{T} region.

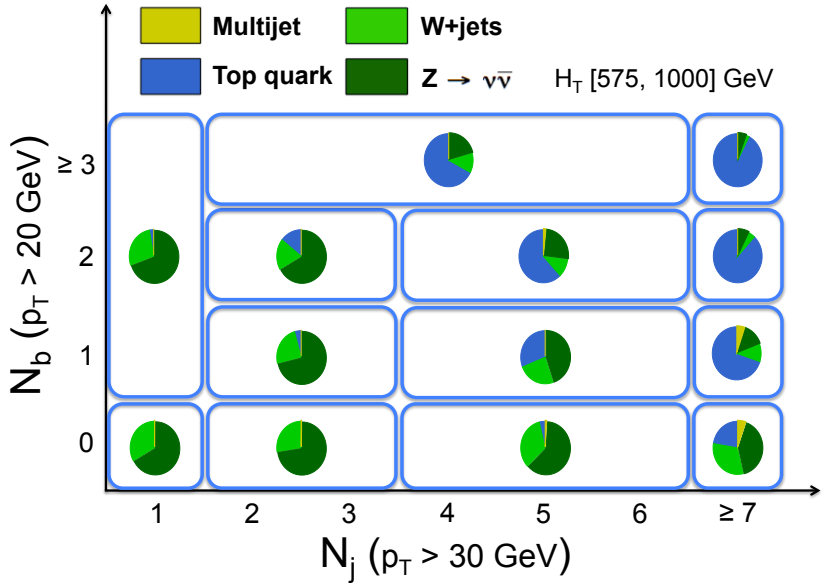

Figure 3.4: Definition of topological regions in number of jets and number of b-tags, with an example of background composition for the Medium H_{T} region. The topological regions are identified by the solid blue lines.

We further subdivide each topological region in bins of M_{T2}. We select the thresholds of such bins using the following criteria:

- The lower edge of the first bin is always $M_{\text{T2}} = 200$ GeV.
- In each H_{T} region, we select the lower threshold of the last M_{T2} bin such as to have one and only one bin expected to contain less than one background event in the most populated N_{jets}-$N_{\text{b-tags}}$ region. Moreover, the upper limit on H_{T} effectively places an upper limit on M_{T2}. Therefore, this lower M_{T2} threshold should not be

larger then the upper limit on H_{T}, in each H_{T} region. This results in the following binning:

  - 3 bins in M_{T2} at Very Low H_{T}: [200,300], [300,400], [400,∞]
  - 4 bins in M_{T2} at Low H_{T}: [200,300], [300,400], [400,500], [500,∞]
  - 5 bins in M_{T2} at Medium H_{T}: [200,300], [300,400], [400,600], [600,800], [800,∞]
  - 5 bins in M_{T2} at High H_{T}: [200,400], [400,600], [600,800], [800, 1000], [1000,∞]
  - 5 bins in M_{T2} at Extreme H_{T}: [200,400], [400,600], [600,800], [800,1000], [1000,∞]

- In each H_{T}-N_{jets}-$N_{\mathrm{b-tags}}$ region, we merge M_{T2} bins that are expected to contain less than one background events.

The resulting bins are summarized in Tables 3.2–3.3 and shown graphically in figure 3.5.

In addition, we also select events with $N_{\mathrm{jets}} = 1$, with the jet required to have p_{T} (jet)$>$ 200 GeV. Monojet events are further categorized in:

- $N_{\mathrm{b-tags}}$: 1j 0b, 1j $\geq$1b
- H_{T}: 7 bins in H_{T}, [200,250], [250,350], [350,450], [450,575], [575,700], [700,1000], [1000,∞]

In each H_{T}-$N_{\mathrm{b-tags}}$ region, we merge H_{T} bins that are expected to contain less than one background event.

Table 3.2: Adopted M_{T2} binning in each topological region of the multi-jet search regions, for the very low, low and medium H_{T} regions.

H_{T} Range [GeV]	Jet Multiplicities	M_{T2} Binning [GeV]
[200, 450]	$2-3$j, 0b	[200, 300, 400, ∞]
	$2-3$j, 1b	[200, 300, 400, ∞]
	$2-3$j, 2b	[200, 300, 400, ∞]
	$4-6$j, 0b	[200, 300, 400, ∞]
	$4-6$j, 1b	[200, 300, 400, ∞]
	$4-6$j, 2b	[200, 300, 400, ∞]
	≥ 7j, 0b	[200, ∞]
	≥ 7j, 1b	[200, ∞]
	≥ 7j, 2b	[200, ∞]
	$2-6$j, ≥ 3b	[200, 300, ∞]
	≥ 7j, ≥ 3b	[200, ∞]
[450, 575]	$2-3$j, 0b	[200, 300, 400, 500, ∞]
	$2-3$j, 1b	[200, 300, 400, 500, ∞]
	$2-3$j, 2b	[200, 300, 400, 500, ∞]
	$4-6$j, 0b	[200, 300, 400, 500, ∞]
	$4-6$j, 1b	[200, 300, 400, 500, ∞]
	$4-6$j, 2b	[200, 300, 400, 500, ∞]
	≥ 7j, 0b	[200, ∞]
	≥ 7j, 1b	[200, 300, ∞]
	≥ 7j, 2b	[200, ∞]
	$2-6$j, ≥ 3b	[200, 300, ∞]
	≥ 7j, ≥ 3b	[200, ∞]
[575, 1000]	$2-3$j, 0b	[200, 300, 400, 600, 800, ∞]
	$2-3$j, 1b	[200, 300, 400, 600, 800, ∞]
	$2-3$j, 2b	[200, 300, 400, 600, ∞]
	$4-6$j, 0b	[200, 300, 400, 600, 800, ∞]
	$4-6$j, 1b	[200, 300, 400, 600, ∞]
	$4-6$j, 2b	[200, 300, 400, 600, ∞]
	≥ 7j, 0b	[200, 300, 400, ∞]
	≥ 7j, 1b	[200, 300, 400, ∞]
	≥ 7j, 2b	[200, 300, 400, ∞]
	$2-6$j, ≥ 3b	[200, 300, 400, ∞]
	≥ 7j, ≥ 3b	[200, 300, 400, ∞]

Table 3.3: Adopted M_{T2} binning in each topological region of the multi-jet search regions, for the high and extreme H_{T} regions.

H_{T} Range [GeV]	Jet Multiplicities	M_{T2} Binning [GeV]
[1000, 1500]	$2-3$j, 0b	[200, 400, 600, 800, 1000, ∞]
	$2-3$j, 1b	[200, 400, 600, 800, ∞]
	$2-3$j, 2b	[200, 400, ∞]
	$4-6$j, 0b	[200, 400, 600, 800, 1000, ∞]
	$4-6$j, 1b	[200, 400, 600, 800, ∞]
	$4-6$j, 2b	[200, 400, 600, ∞]
	≥ 7j, 0b	[200, 400, 600, ∞]
	≥ 7j, 1b	[200, 400, 600, ∞]
	≥ 7j, 2b	[200, 400, ∞]
	$2-6$j, ≥ 3b	[200, 400, ∞]
	≥ 7j, ≥ 3b	[200, 400, ∞]
[1500, ∞]	$2-3$j, 0b	[200, 400, 600, 800, 1000, ∞]
	$2-3$j, 1b	[200, 400, 600, ∞]
	$2-3$j, 2b	[200, ∞]
	$4-6$j, 0b	[200, 400, 600, 800, 1000, ∞]
	$4-6$j, 1b	[200, 400, 600, ∞]
	$4-6$j, 2b	[200, 400, 600, ∞]
	≥ 7j, 0b	[200, 400, ∞]
	≥ 7j, 1b	[200, 400, ∞]
	≥ 7j, 2b	[200, 400, ∞]
	$2-6$j, ≥ 3b	[200, ∞]
	≥ 7j, ≥ 3b	[200, ∞]

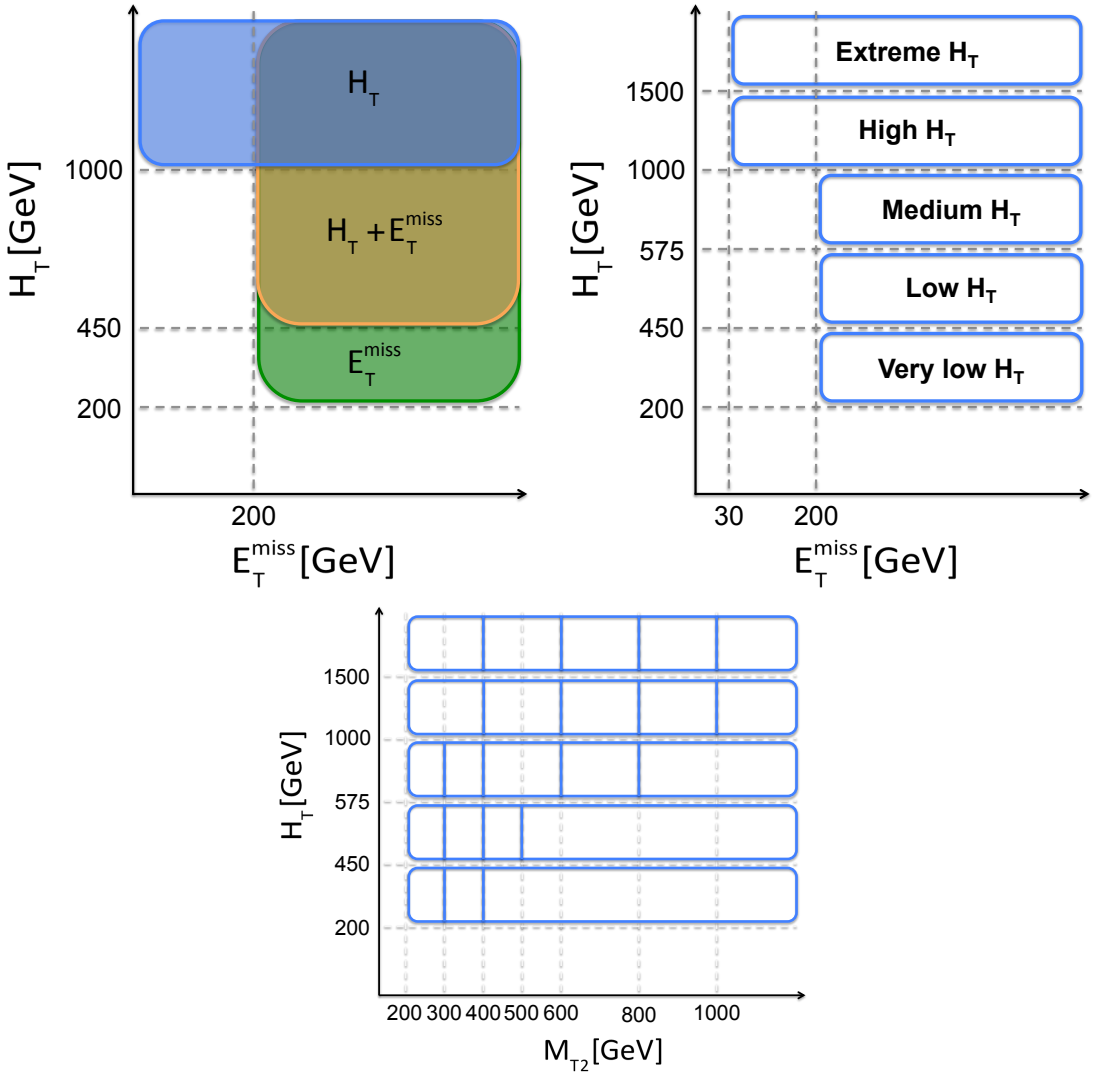

Figure 3.5: The figure in the top left shows the signal trigger coverage in the $E_{\mathrm{T}}^{\mathrm{miss}}$–$H_{\mathrm{T}}$ plane. This plane is divided into several H_{T} regions as shown in the top right figure. The M_{T2} binning in each H_{T} region is shown in the bottom figure.

Chapter 4

Invisible Z Background

Events with Z bosons produced in association with jets, where the Z decays to neutrinos, contain genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ and are the most signal-like of the background processes. There are no handles to reduce the number of $\mathrm{Z} \to \nu\nu$ events entering the search regions that would not also remove a significant amount of signal events, making this background irreducible. This background is estimated primarily from a control sample of $\gamma +$ jets events, relying on the similarities in $\mathrm{Z} +$ jets and $\gamma +$ jets kinematics.

4.1 Prediction using control regions

The background contribution from $\mathrm{Z} \to \nu\nu$ is estimated from $\gamma +$jets control regions. The $\gamma +$jets control regions are populated by events passing the `HLT_Photon165_HE10_v1` trigger with a photon passing the requirements listed in section 3.1.7. An offline cut on the photon p_{T} of 180 GeV is applied, which is in the plateau of the trigger turn on. This photon p_{T} requirement mimics the implicit requirement on the p_{T} of the Z boson due to the $M_{\mathrm{T2}} > 200$ GeV requirement in the baseline selections. The trigger efficiency after the baseline selection for the $\gamma +$jets control region is 97.5%. In order to remove potential

signal contamination, events in the control regions must have $M_{\mathrm{T2}} < 200$ GeV.

The photon is then removed from the event and its transverse momentum is added to the $E_{\mathrm{T}}^{\mathrm{miss}}$ in order to mimic the kinematics of $\mathrm{Z} \to \nu\nu$ events. In order to avoid counting a jet corresponding to the photon, the PF jet closest to the photon, within $\Delta\mathrm{R} < 0.4$, is also removed from the event. With these modifications to the $\gamma+$jets events, all variables involving jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ are recalculated. Using the updated values of these variables, the baseline selections of Section 3.3 are then applied. The $\gamma +$ jets events passing this selection are then divided into the various topological regions defined by H_{T}, N_{jets}, and $N_{\mathrm{b-tags}}$ as described in Section 3.4. Due to limited statistics, the control regions are not further divided into the various M_{T2} bins corresponding to each signal region. Instead, $\mathrm{Z} \to \nu\nu$ monte carlo is used to extrapolate along the M_{T2} dimension to estimate the $\mathrm{Z} \to \nu\nu$ contribution in each signal region. For events with $N_{\mathrm{jets}} = 1$, one control region is defined for each bin of jet p_{T}.

The $\mathrm{Z} \to \nu\nu$ yield in each signal region is then estimated from the corresponding control region as follows:

$$N^{\mathrm{SR}}_{\mathrm{Z}\to\nu\nu}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b}, M_{\mathrm{T2}}) = N^{\mathrm{CR}}_{\gamma}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b}) \times P_{\gamma}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b}) \times f \times R^{\mathrm{Z}/\gamma}_{\mathrm{MC}}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b}) \times k(M_{\mathrm{T2}}) \tag{4.1}$$

where:

- $N^{\mathrm{CR}}_{\gamma}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b})$ is the observed $\gamma+$ jet yield in the control region;
- $P_{\gamma}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b})$ is the prompt-photon purity of the selected control region events (see Section 4.3);
- f is the fraction of prompt photons that are direct (defined as ΔR (parton, photon) > 0.4) over the total number of prompt photons (see Section 4.3);

- $R_{\mathrm{MC}}^{\mathrm{Z}/\gamma}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b})$ is the ratio of $\mathrm{Z} \to \nu\nu$ yields over γ + jets yields taken from MC simulation (where γ refers to direct prompt photons), and corrects for photon acceptance and reconstruction efficiency as well as residual kinematic differences between γ+ jets and $Z(\nu\nu)$+ jets (see Section 4.2);

- $k(M_{\mathrm{T2}})$ is the fraction of events in the corresponding M_{T2} bin in the control region measured in monte carlo.

4.2 Z/γ ratio

A key piece of the transfer factor from control region yields to signal region yields is the Z/gamma ratio, taken from monte carlo separately for each topological region. The leading order diagrams for production of $\mathrm{Z} \to \nu\nu$ and γ + jets events are shown in figure 4.1. There are two things in these diagrams that depend on whether the boson is a Z or a photon: the coupling at the quark-boson vertex and the mass of the boson. The difference in boson mass does not change the kinemtics very much for values of boson p_{T} much larger than the Z mass. Therefore we expect that the ratio of Z production over prompt-photon production to flatten out with increasing boson p_{T} to the ratio of the couplings to quarks. Figure 4.2 shows Z + jets and γ + jets monte carlo yields for events with $H_{\mathrm{T}} > 450$ GeV as a function of boson p_{T} and the ratio of event yields, confirming the flattening of the Z/gamma ratio at large boson p_{T}.

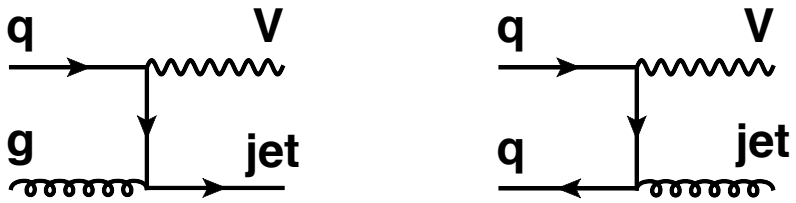

Figure 4.1: Leading-order Feynman diagrams for Z+ jets and γ+ jets processes. In the figure 'V' can be either a Z boson or a photon.

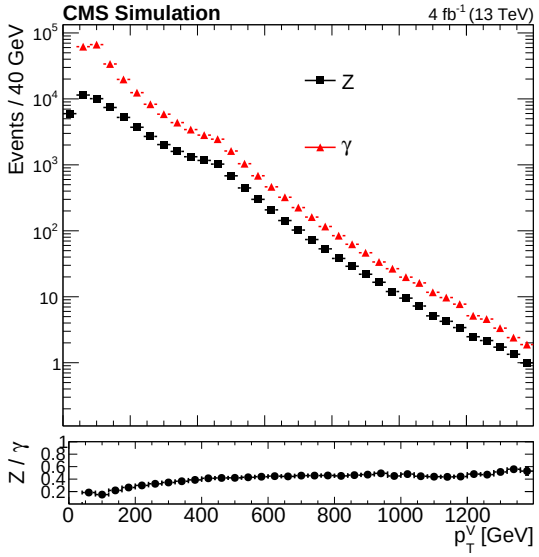

Figure 4.2: Expected number of events with $H_{\mathrm{T}} > 450$ GeV in 4 fb^{-1} of data at 13 TeV, as a function of the transverse momentum of the boson ($p_{\mathrm{T}}(V)$) for γ+ jet (red markers) and $Z(\nu\nu)$+ jets (black markers). The Z/γ ratio as a function of $p_{\mathrm{T}}(V)$ is shown at the bottom.

The Z/gamma ratio is taken from Z + jets and γ + jets monte carlo, computed separately in each topolgical region. Figures 4.3 and 4.4 show the Z/gamma ratio for different H_{T} and $N_{\mathrm{jets}}/N_{\mathrm{b-tags}}$ selections. The value of the Z/gamma ratio is ≈ 0.5 for most regions, so we expect control region yields to be larger than topological region yields by

about a factor of two. The ratio has a mild M_{T2} dependence, but is largely independent of the H_{T}, N_{jets}, and $N_{\mathrm{b-tags}}$ selections.

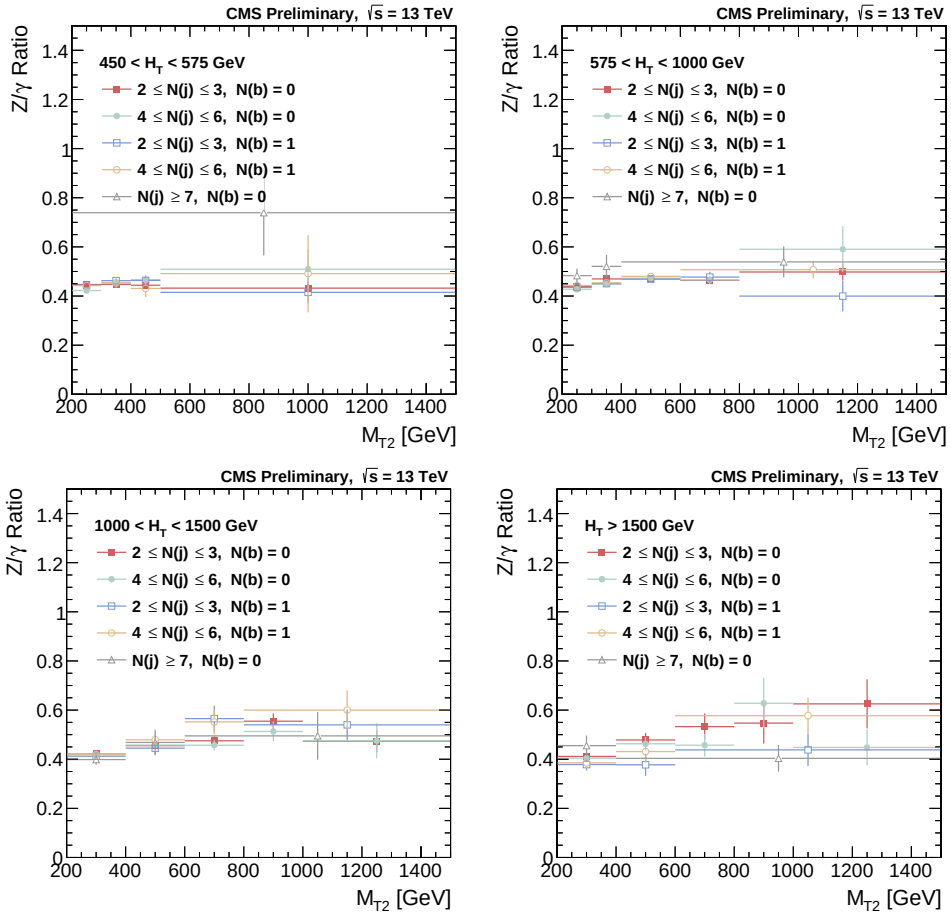

Figure 4.3: Z/γ ratio as a function of M_{T2} for different $N_{\mathrm{jets}}/N_{\mathrm{b-tags}}$ selections in the four analysis H_{T} regions.

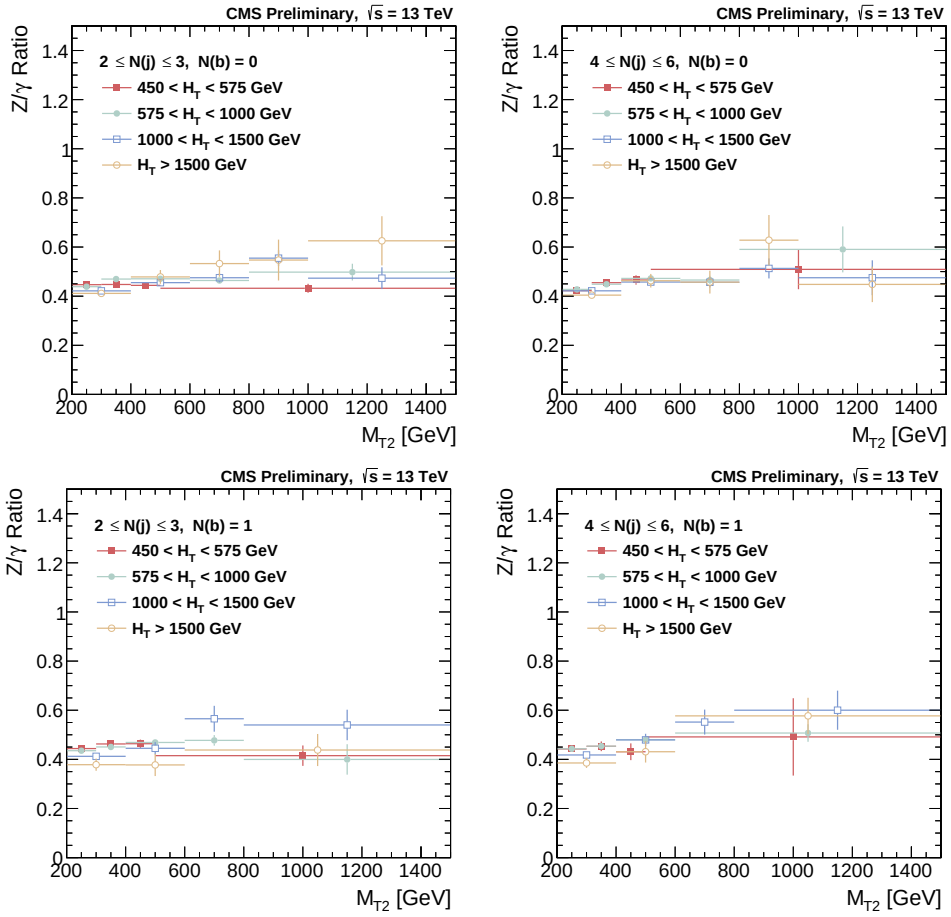

Figure 4.4: Z/γ ratio as a function of M_{T2} for different H_{T} selections for four $N_{\mathrm{jets}}/N_{\mathrm{b-tags}}$ combinations.

4.3 Photon purity

In order to apply the Z/gamma ratio to the γ+jets control region yields we must make sure to only count direct prompt photons. Counting photons from fragmentation at low ΔR(parton, photon) and non-prompt "fake" photons would lead to an overestimation of the Z $\rightarrow \nu\nu$ yield since Z production does not include these processes. Experimentally we cannot distinguish direct prompt photons from fragmentation photons. To estimate the number of direct prompt photons we use a data-driven strategy to estimate the prompt

photon purity,

$$P_\gamma = \frac{\text{prompt}}{\text{prompt + fake}} \tag{4.2}$$

and multiply by the fraction of prompt photons which are direct:

$$f = \frac{\text{direct prompt}}{\text{all prompt}}, \tag{4.3}$$

where f is taken from monte carlo.

To study the identification of prompt photons, as well as the fragmentation and non-prompt backgrounds, we define a prompt photon as a stable particle identified by the MC-truth as a photon, whose mother particle is a photon or a quark. To be considered "matched", a reconstructed photon has to be matched to a prompt photon in direction and momentum. The matching requires $\Delta R < 0.1$, and $0.5 < p_{\mathrm{T}}^{\mathrm{Gen}}/p_{\mathrm{T}}^{\mathrm{reco}} < 2$. A cut of ΔR (parton, photon) > 0.4 is chosen to define direct photons in this analysis. The value of the cut is based on the generator-level cuts present in the Madgraph MC $\gamma+$jets process. A definition based on generator-level isolation (GenIso < 5 GeV), used by other analyses, was found to yield comparable results but a less straightforward combination of the $\gamma+$jets and QCD MC samples. When using the $\gamma+$jets and QCD MC samples to estimate photon yields, events with direct prompt photons are taken only from the $\gamma+$jets MC sample, while events with fragmentation and non-prompt photons are taken from the QCD sample. To avoid double-counting, events with prompt photons satisfying $\Delta R(\text{parton}, \text{photon}) > 0.4$ are removed from the QCD sample.

Figure 4.5, shows the trend of f as a function of H_{T}, M_{T2}, N_{jets} and $N_{\mathrm{b-tags}}$ in the baseline $\gamma+$jets control region. The predictions given by the standard Madgraph QCD sample and the one from a Pythia QCD sample are compared, and while the agreement in shape is acceptable the two predictions differ by a factor of two in yields, predicting

the fragmentation component to be between 5% and 10% of the total prompt photon yield. This fraction is fairly stable as a function of event kinematics, so we use a fixed value of 0.92 to describe it. To cover the observed variations and possible mismodelings of the fragmentation component, we vary the $\Delta R < 0.4$ yields by $\pm 100\%$, yielding $f = 0.92 \pm 0.08$.

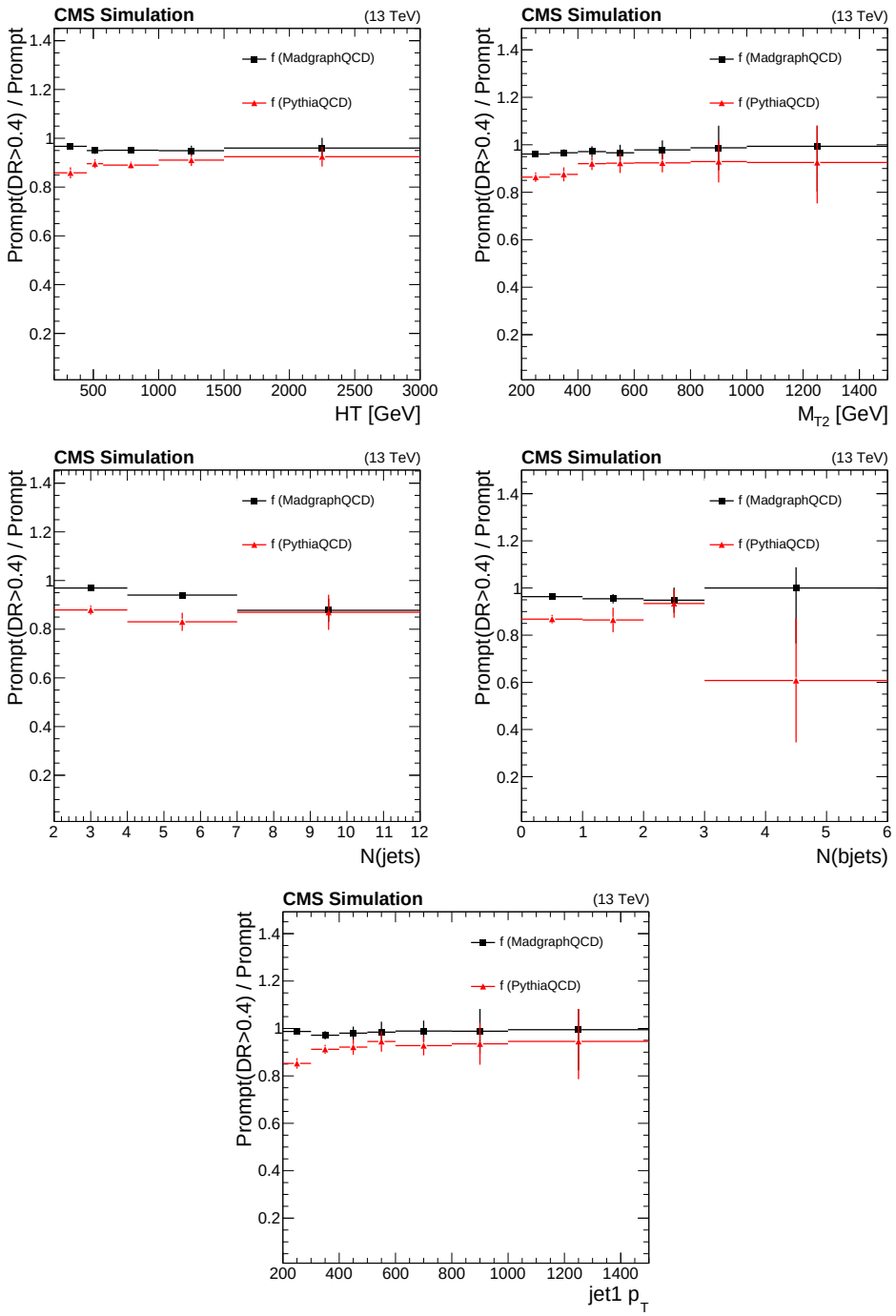

Figure 4.5: Fraction of prompt photons that are direct ($\Delta R(\text{parton}, \text{photon}) > 0.4$), over the total number of prompt photons, as a function of the event H_{T} (top left), M_{T2} (top right), N_{jets} (middle left) and $N_{\text{b-tags}}$ (middle right). The bottom plot show the fraction in the 1-jet region, as a function of H_{T}.

4.3.1 Template fit method

In order to estimate the relative fractions of fake and prompt photons in the control regions we use a template fit method. The aim of the template fit method is to analyze the shape of the photon isolation distribution in data and extract the maximal amount of information regarding the fraction of prompt photons populating it. This is done by defining the expected shapes (templates) of the isolation distributions for prompt and fake photons, and performing, in each region of interest, a templated unbinned maximum-likelihood fit to the data. In this way, the fraction of events which are compatible with prompt photons is extracted.

For prompt photons the isolation template is extracted with a 'random cone' method, taken from [24]. The idea behind the random cone method is that prompt photons are produced at an electromagnetic vertex, hence are expected to be produced without additional hadronic activity. The only hadronic activity that could be present next to prompt photons would come from pile-up or from underlying event activity, and these are assumed to be uncorrelated with the photon direction, and to be diffuse in the detector. Therefore, one assumes that one should see the same amount of activity when opening an isolation cone around a prompt photon or in any other direction in the detector.

The method therefore proceeds as follows: starting from the photon direction, a new direction is identified by rotating ϕ by $+\pi/2$, keeping η constant. This new direction is accepted if:

- no isolated lepton is found within $\Delta R_{\text{iso}} = 0.3$;
- no photon with $p_{\text{T}} > 10$ GeV is found within $2 \cdot \Delta R_{\text{iso}}$;
- no jet with $p_{\text{T}} > 20$ GeV and $|\eta| < 2.5$ is found within $2 \cdot \Delta R_{\text{iso}}$;

If any of these conditions is not fulfilled, the direction is aborted and a second, last, try

is done at $\phi - \pi/2$. If also this fails because of one of the above veto conditions, the event is not used to determine the isolation template. It has been found that the method has a per-event efficiency of about 80%.

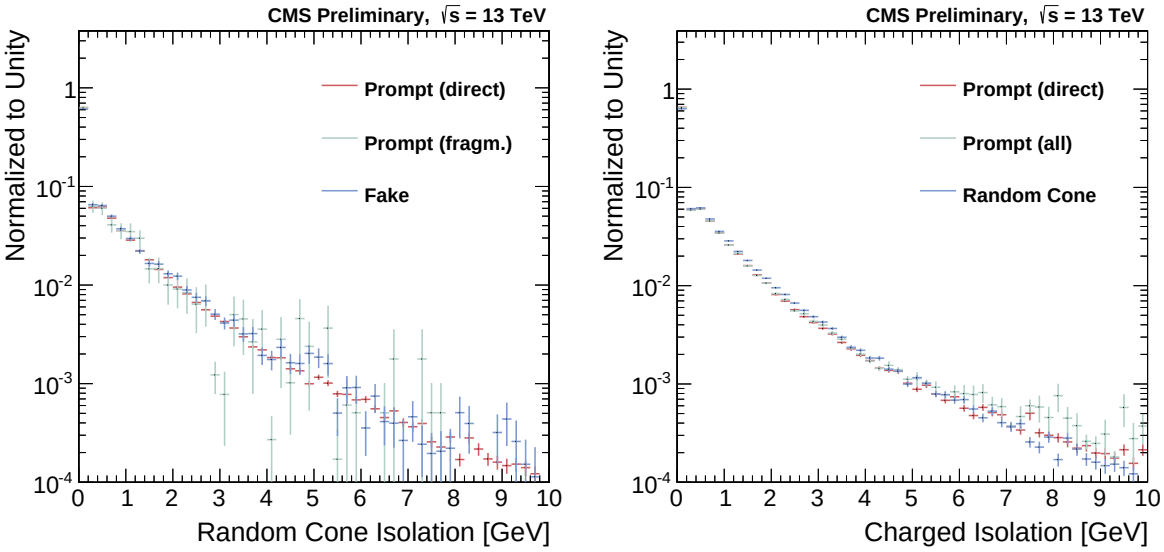

Figure 4.6: Left: random cone isolation template for simulated events in which the initial photon candidate is direct-prompt (red), fragmentation-prompt (green) and fake (blue). Right: isolation of direct-prompt photons (red), of the direct+fragmentation prompt mixture (green) in simulated events, compared to what is obtained with the random cone method (blue). All distributions are normalized to unity.

The obtained isolation templates are shown in Figure 4.6: the left plot shows the template shapes when starting from an initial photon candidate which is direct-prompt (red), fragmentation-prompt (green) and fake (blue). As expected, the random cone isolation distribution does not depend on the nature of the initial seed direction, as it depends only on pile-up and underlying event activity. The right plot instead compares the template obtained with the random cone approach (blue) with the shape obtained on MC-matched direct prompt photons (red) and the mix between direct- and fragmentation-prompt photons (green). As can be seen the random cone describes the expected isolation shape remarkably well, over four orders of magnitude, thus validating the method.

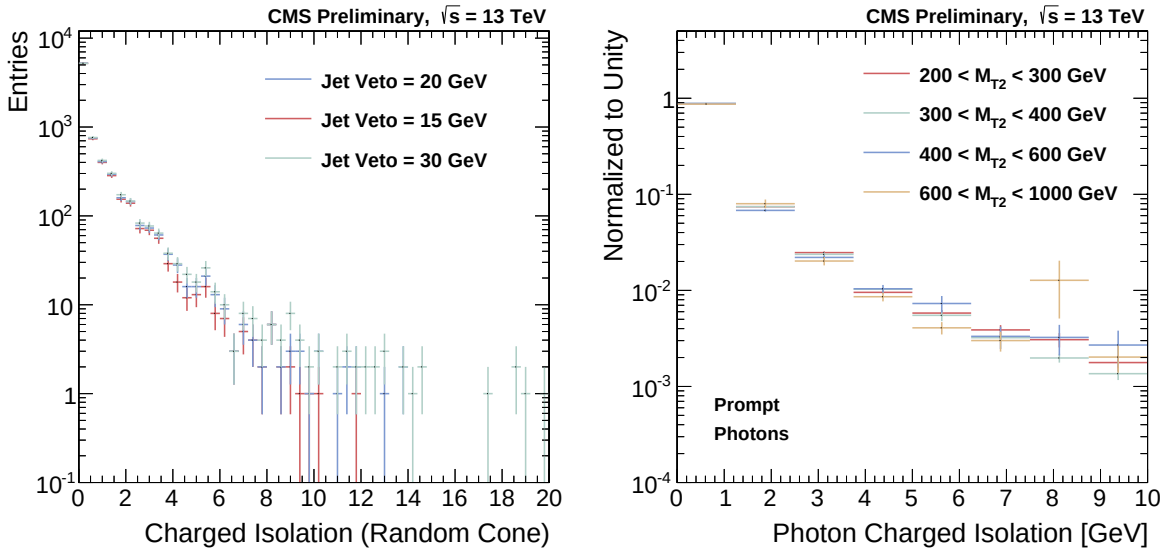

Figure 4.7: Left: Random cone isolation templates obtained in simulated events with the nominal 20 GeV jet-veto threshold (blue), compared to what is obtained by varying this parameter to 15 (red) and 30 GeV (green). Right: prompt photon templates taken from the simulation, for different M_{T2} ranges.

The random cone method was found to be fairly robust for variations of the veto criteria described in the above. In particular, the transverse momentum threshold of jets which have the power of vetoing the two trial directions was varied in order to probe the sensitivity of the template shape on this parameter. This is seen in Figure 4.7 (left), where the template obtained with the nominal 20 GeV threshold (blue) is compared to the ones obtained respectively with 15 (red) and 30 GeV (green). As can be seen, the results are fairly stable. Figure 4.7 (right) shows instead the dependence of the prompt templates as a function of M_{T2}: here templates for photons matched to simulated prompt photons are shown for different M_{T2} ranges (different colors). All templates are normalized to unity. As can be seen the dependence on M_{T2} is found to be negligible.

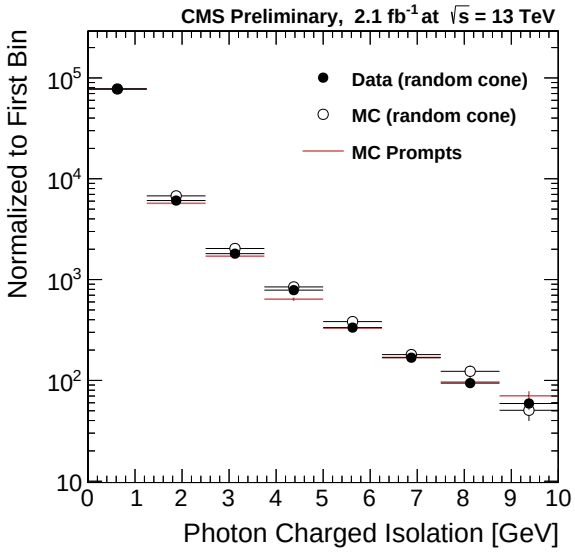

Figure 4.8: Closure of the random cone method to estimate the prompt photon template: black markers show the charged isolation distribution obtained with the random cone method in data; open markers the result of the same method applied to the simulation; the red line instead shows the charged isolation distribution for matched prompt photons in the simulation. Histograms are normalized so as to have the same value as the data in the first bin.

The closure of the random cone method is shown in Figure 4.8: the black markers show the charged isolation distribution obtained with the random cone method in data; open markers the result of the same method applied to the simulation; the red line instead shows the charged isolation distribution for matched prompt photons in the simulation. As can be seen, the three shapes are in good agreement, thus validating the random cone method for obtaining prompt photon isolation templates.

To extract the isolation template for fake photons we use a fake-enriched sample in data obtained by moving to the sidebands the photon shower shape variable $\sigma_{i\eta i\eta}$. The $\sigma_{i\eta i\eta}$ variable assumes lower values for prompt photons (as the transverse containment of an electromagnetic shower is only determined by the Molière radius of the calorimeter), while it assumes larger values for photons produced by meson decays in jets (eg. $\pi^0 \to \gamma\gamma$, $\eta \to \gamma\gamma$) for two main reasons:

- the meson decay produces diphotons with an opening angle (relevant for low-momentum decays);
- the presence of additional nearby photonically-decaying mesons (relevant for high-momentum decays, where the boost makes the opening angle smaller than the ECAL resolution).

Figure 4.9 shows the $\sigma_{i\eta i\eta}$ distribution for photons reconstructed in the ECAL barrel (left) or ECAL endcaps (right): the data is compared to the simulation, in which the prompt (grey), fragmentation (blue) and fake (red) contributions are shown. The simulation is normalized to the integral of the data. For pictorial reasons, these photons have been selected with a very loose isolation requirement (charged isolation < 20 GeV) in order to artificially increase the fake photon yield.

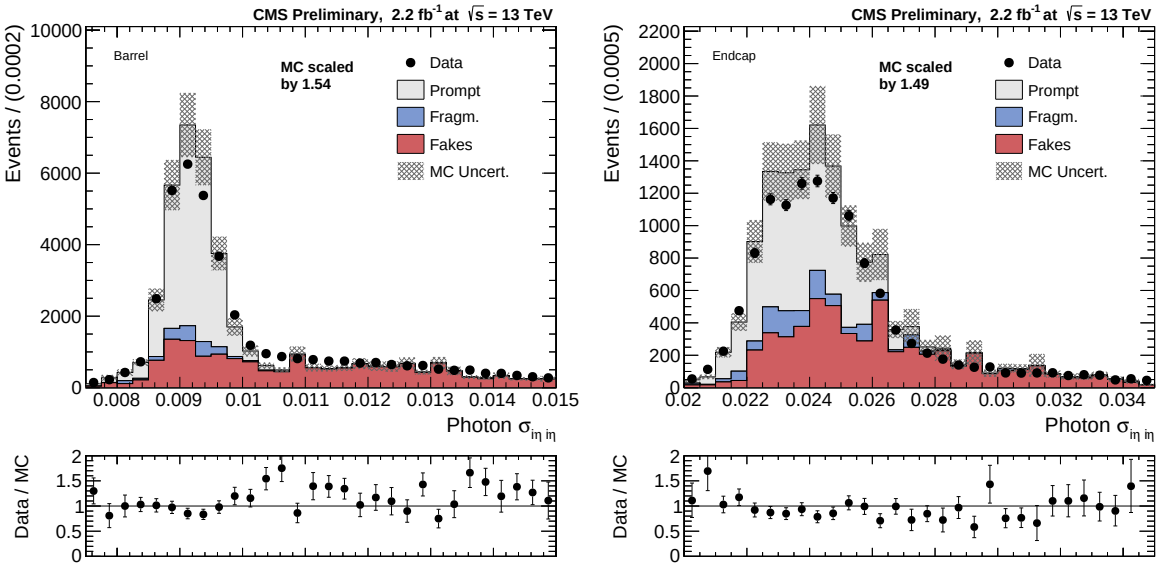

Figure 4.9: Data-MC shape comparison of $\sigma_{i\eta i\eta}$ for photon candidates reconstructed in the ECAL barrel (left) and endcaps (right). Prompt photons are shown in grey, fragmentation photons in blue and fake photons in red. These events have been selected with a very loose isolation requirement, in order to increase the yield of fake photons.

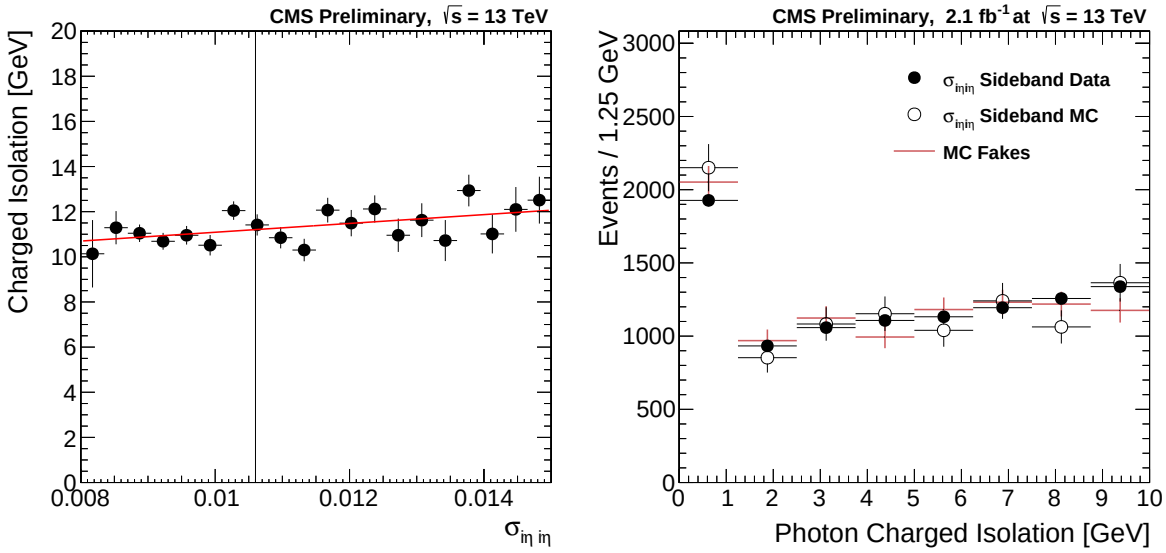

Figure 4.10: Left: average photon charged isolation as a function of $\sigma_{i\eta i\eta}$ in simulated events. The trend is fitted with a linear function, and the fit result is overlayed with a red line. The position of the used photon identification threshold is marked with a vertical black line. Right: comparison between the fake photon isolation template obtained from the $\sigma_{i\eta i\eta}$ sidebands (black markers) and the one obtained by selecting MC-matched fake photons that pass the nominal selection (red crosses).

We define control regions by moving to the sidebands of this variable, defined as $0.011 < \sigma_{i\eta i\eta} < 0.015$ ($0.03 < \sigma_{i\eta i\eta} < 0.035$) for the ECAL barrel (endcaps). As can be seen from the shaded areas in Figure 4.9, these control regions are expected to be dominated by fake photons. So, under the hypothesis that $\sigma_{i\eta i\eta}$ is not correlated to the photon isolation variable we are using, these sidebands can be used to understand the isolation distributions for fake photons. Figure 4.10 (left) shows the trend of the average isolation value as a function of $\sigma_{i\eta i\eta}$, for candidates which pass an isolation preselection cut of 20 GeV. The trend is fitted with a linear function, and the result of the fit is overlayed as a red line. As can be seen, a slight trend is indeed found, but the maximal effect observed over the full $\sigma_{i\eta i\eta}$ range is less than 20%.

Figure 4.10 (right) compares the isolation distribution for photon candidates in the $\sigma_{i\eta i\eta}$ sidebands in data (black markers), to the corresponding events passing the same

$\sigma_{i\eta i\eta}$-sideband selection in the simulation (open markers). The two are then compared to the isolation distribution of MC-matched fake photons which instead pass the nominal $\sigma_{i\eta i\eta}$ requirements (red crosses). The simulation is here normalized to the integral of the data. As can be seen the distributions are in good agreement, within statistical uncertainties, thus validating the $\sigma_{i\eta i\eta}$ sideband method.

Once the templates are defined, the purity is extracted with an unbinned maximum-likelihood fit to the charged isolation distribution of events passing the selections in the photon control region. Photons are required to have a charged isolation value less than 10 GeV ('loose' isolation). The fit is performed in the full $0 - 10$ GeV isolation range, and the templates are obtained from data with the $\sigma_{i\eta i\eta}$ sideband method for fakes, and the random cone method for prompts. The templates are built as histograms with eight 1.25 GeV bins, exactly as in Figures 4.10 (right) and 4.8, respectively for fake and prompt photons. A separate fit is performed in each topological region (integrated over M_{T2}) and all fits use the same templates, which are defined from the inclusive sample.

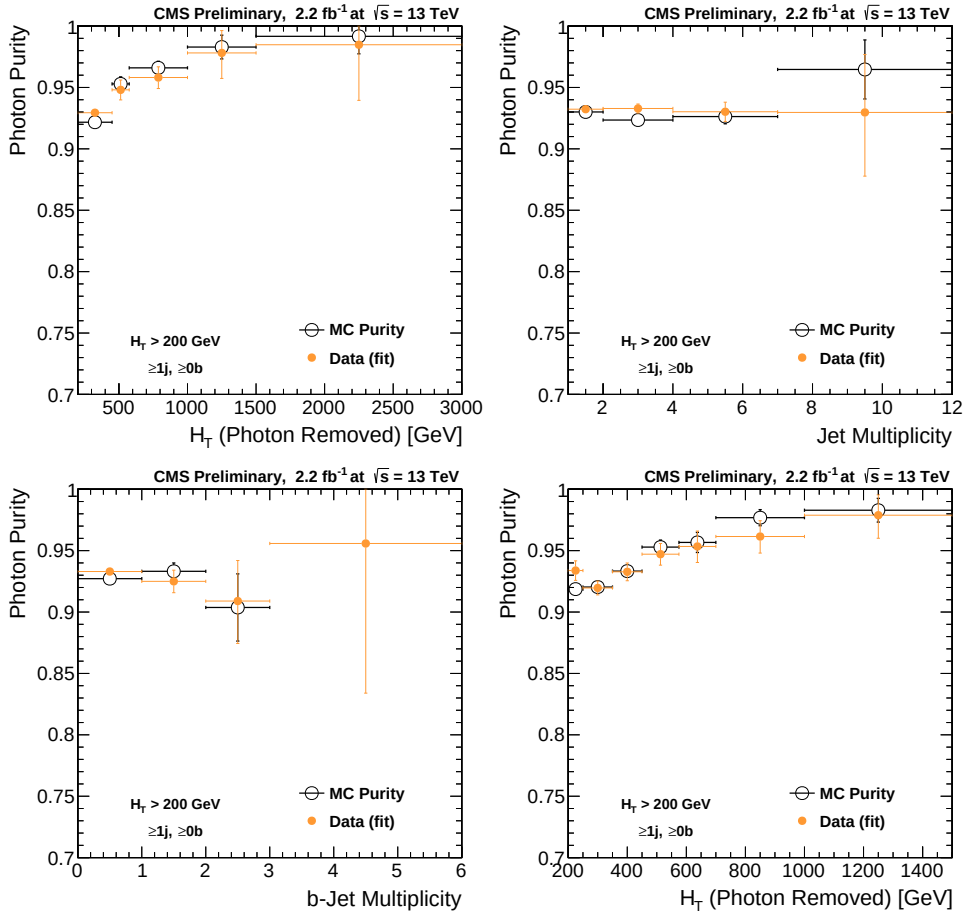

Figure 4.11: Results of the purity template fits in data (yellow markers) compared to the MC truth purity (hollow markers), as a function of H_{T} (top left), jet multiplicity (top right), b-jet multiplicity (bottom left) and finally H_{T} in the monojet binning (bottom left).

From the result of the fit, which gives the purity of photons passing loose isolation requirements, the purity after the nominal 2.5 GeV is extracted. Results for both multijet and monojet events are shown in Figure 4.11, where the purity obtained in data with the template fit (yellow markers) is compared to the one obtained with the MC truth (hollow markers), as a function of H_{T} (top left), jet multiplicity (top right), b-jet multiplicity (bottom left) and finally for H_{T} in the binning used in the monojet search region (bottom right). As can be seen a good level of agreement is achieved between the

data and the simulation. Additionally, one can observe how the fitted purities are very high, typically larger than 95%.

4.3.2 Fake rate method

We can also predict the background from fake-photons using a fake rate (FR) technique. The idea is to count, for each control region bin, the number of events containing a photon which fails the full selection ("Tight") but passes a looser set of requirements ("Loose"). These events, referred to as "LooseNotTight", are for the most part characterized by fake-photons (with a small prompt contribution that must be taken into account). A Fake Rate, defined as the ratio of Tight over Loose fake-photons, is then used to derive the number of Tight fake-photons in each control region bin.

We define the Tight selection as the full photon selection described in Section 3.1.7, while the Loose selection uses a charged isolation cut of 20 GeV (instead of 2.5 GeV). When considering the $\sigma_{i\eta i\eta}$-sideband photons in the baseline γ + jets control region, as will be done in data, the Fake Rate is measured in MC to be 0.10 ± 0.01. This FR value, and its dependence on kinematics, can be compared in MC with the FR of the photons passing the $\sigma_{i\eta i\eta}$ cut, shown in Figure 4.12. The purity difference between using the two Fake Rates, $\sigma_{i\eta i\eta}$-sideband and pass-$\sigma_{i\eta i\eta}$, is considered as a systematic uncertainty.

The LooseNotTight region is composed, in MC, of 85% fake-photons, 8% fragmentation photons and 7% direct photons. While the prompt photon population is predominantly in the Tight region, the fragmentation photon population is evenly distributed (58%/42%) between Tight and LooseNotTight region. To account for normalization uncertainty of the fragmentation photon component, the LooseNotTight fragmentation yield is varied by $\pm$50%, and the resulting change in the purity estimate is considered as a systematic uncertainty.

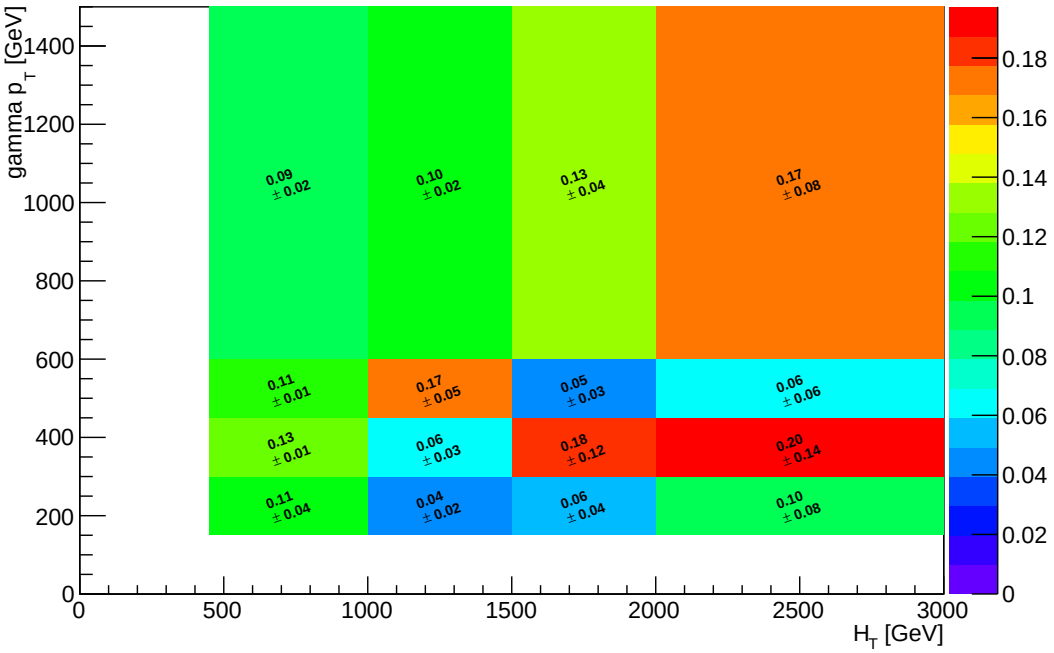

Figure 4.12: Fake Rate (Tight/Loose) for fake-photons in QCD MC as a function of p_{T}^{γ} and H_{T} for photons passing the $\sigma_{i\eta i\eta}$ cut. The difference between this binned FR and the fixed value obtained from the $\sigma_{i\eta i\eta}$-region in MC is considered as a systematic uncertainty.

The purity estimated by the Fake Rate method is shown in Figure 4.13 for the region with $450 < H_{\mathrm{T}} < 575$ GeV, $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV, 2-3j, 0b. In this figure, the black points are based on the nominal method, using a single FR obtained in the $\sigma_{i\eta i\eta}$ sideband and using the nominal yield of LooseNotTight fragmentation photons. Alternative estimates based on the pass-$\sigma_{i\eta i\eta}$ FR and on varied fragmentation yields are shown in red, green and blue. The magenta squares show the purity measured in MC based on the $\gamma+$jets and QCD yields in the Tight region, with statistical uncertainties dominated by the large-weight of QCD events in the Tight region. The different purity estimates are found to agree within 5%, so 5% is taken as an overall systematic uncertainty on the method.

Binomial error propagation is used to estimate the statistical uncertainty of the purity

estimate, taking into account statistical uncertainties in the FR measurement and the LooseNotTight yield. When no LooseNotTight events are found, as in some low-statistics signal regions, the purity is currently estimated to be $100\% \pm 0\%$.

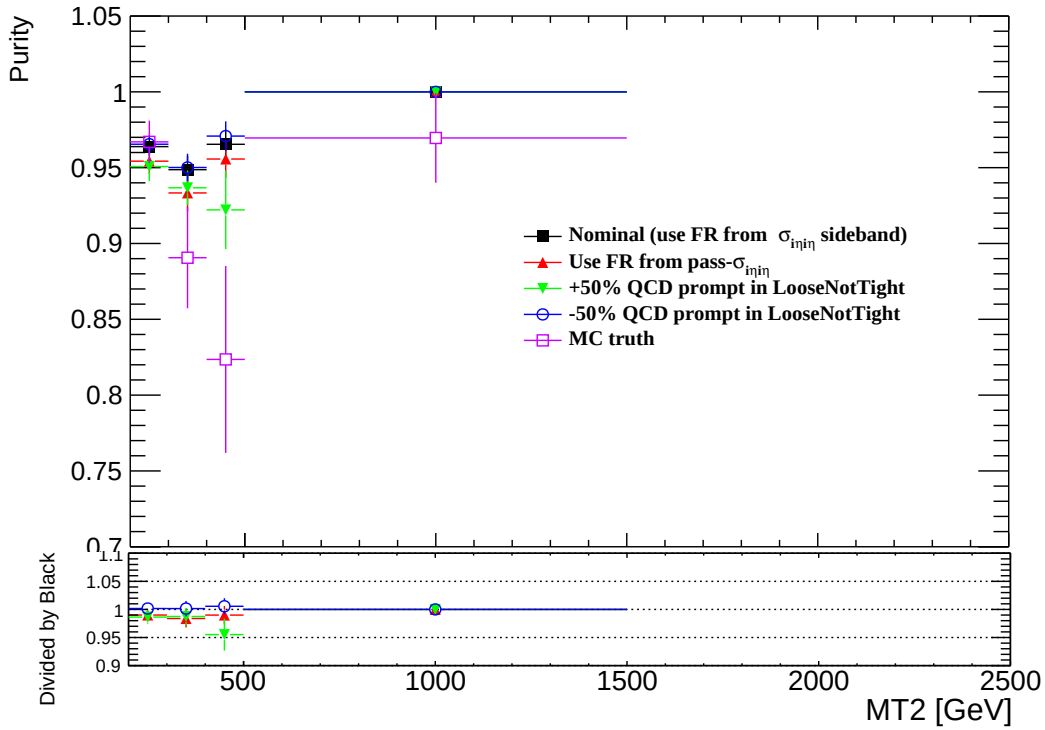

Figure 4.13: Photon purity measured in the region with $450 < H_{\mathrm{T}} < 575$GeV, $E_{\mathrm{T}}^{\mathrm{miss}} > 200$GeV, 2-3j, 0b. The nominal purity is shown in black, while red shows the purity based on the pass-$\sigma_{i\eta i\eta}$ FR, and green/blue show the purity after $\pm 50\%$ variations in LooseNotTight fragmentation photon yields. The magenta squares show the purity measured in MC based on the $\gamma+$jets and QCD yields in the Tight region. The ratios are shown with respect to black.

4.4 Validation using $\mathrm{Z} \to \ell^+\ell^-$ events

We use a $\mathrm{Z} \to \ell^+\ell^-$ control region to validate and correct the Z/γ ratio and to derive uncertainties on the ratio. To define the $\mathrm{Z} \to \ell^+\ell^-$ control region, the baseline selections of Section 3.3 are applied with the exception of the lepton veto. Instead we require exactly

2 electrons or 2 muons passing the reco selections defined in Sections 3.1.4 and 3.1.5. As no specific $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement is made for this region, events are selected using dilepton trigger paths:

- `HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*`
- `HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*`
- `HLT_Ele17_Ele12_CaloId_TrackId_Iso_DZ_v*`

To maintain efficiency at high Z p_{T} in the electron channel despite the isolation requirement at HLT, we additionally use events selected using the same photon trigger used for the $\gamma+$jets control regions, `HLT_Photon165_HE10_v1`. To improve trigger efficiency for the dilepton triggers below, we also require that electrons pass the POG Loose ID (as opposed to the POG Veto), and that the two leptons to have $p_{\mathrm{T}} > 25/20$ GeV. Finally, we require that they have opposite charge and that they form an invariant mass satisfying $|m_{\ell\ell} - m_Z| < 10$ GeV, where m_Z is the nominal Z boson mass.

Similarly to what is done for photons in the $\gamma+$jets control region, the lepton vectors in the transverse plane (p_x, p_y) are added to the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector to simulate a $\mathrm{Z} \to \nu\nu$ event. Since leptons can be reconstructed as (part of) PF jets, we remove the closest jet within $\Delta\mathrm{R} < 0.4$ of each lepton and recompute all variables involving jets and $E_{\mathrm{T}}^{\mathrm{miss}}$.

With the above definition of the $\mathrm{Z} \to \ell^+\ell^-$ control region, we can modify Equation 4.1 by introducing a double-ratio term:

$$N_{Z\nu\nu}^{pred} = N_{\gamma}^{CR} \cdot P_{\gamma} \cdot f \cdot R(Z/\gamma) \cdot \frac{R_{data}(\mathrm{Z}_{\ell\ell}/\gamma)}{R_{MC}(\mathrm{Z}_{\ell\ell}/\gamma).} \tag{4.4}$$

The double ratio term compares the ratio of $\mathrm{Z} \to \ell^+\ell^-$ events to $\gamma+$jets events in data and simulation, and is used to correct the central value of the Z/γ ratio. Since the branching ratio of $\mathrm{Z} \to \ell^+\ell^-$ is smaller than the branching fraction of $\mathrm{Z} \to \nu\nu$, the

$\mathrm{Z} \to \ell^+\ell^-$ control region is not split by H_T, N_jets, and $N_\mathrm{b-tags}$ into the various topological regions. Instead, $R(\mathrm{Z}_{\ell\ell}/\gamma)$ is checked using the full statistics of the baseline selection by plotting $R(\mathrm{Z}_{\ell\ell}/\gamma)$ as a function of H_T, N_jets, and $N_\mathrm{b-tags}$. Figure 4.14 shows $R(\mathrm{Z}_{\ell\ell}/\gamma)$ for data and simulation and the double ratio for H_T, N_jets, and $N_\mathrm{b-tags}$. In order to compare data and simulation, the top background is subtracted from the $\mathrm{Z} \to \ell^+\ell^-$ yields in data and the purity and fragmentation factors and uncertainties are applied to the $\gamma + \mathrm{jets}$ yields in data. Here the top background is taken from MC with a 50% uncertainty. Based on these comparisons, $R(Z/\gamma)$ is corrected by a factor of 0.95 ± 0.11 for each topological region.

The double ratios in figure 4.14 are also used to assign systematic uncertainties on $R(Z/\gamma)$ for each topological region. The uncertainties shown in figure 4.14 include statistical uncertainties and uncertainties from the background subtraction mentioned above. For the multijet regions, the uncertainties on the double ratios in the corresponding H_T, N_jets, and $N_\mathrm{b-tags}$ bins are added in quadrature to determine the total systematic uncertainty. For the monojet regions, the uncertainties on the double ratios in the corresponding H_T and $N_\mathrm{b-tags}$ bin are added in quadrature. The $\mathrm{Z} \to \ell^+\ell^-$ statistics for $N_\mathrm{b-tags} \geq 3$ are insufficient to perform the ratio in this bin, so we take twice the uncertainty of the $N_\mathrm{b-tags} = 2$ bin. The total systematic uncertainty is typically smaller than 20% for the high-background-yield regions, but can be significantly larger in the tails of H_T, N_jets, and $N_\mathrm{b-tags}$.

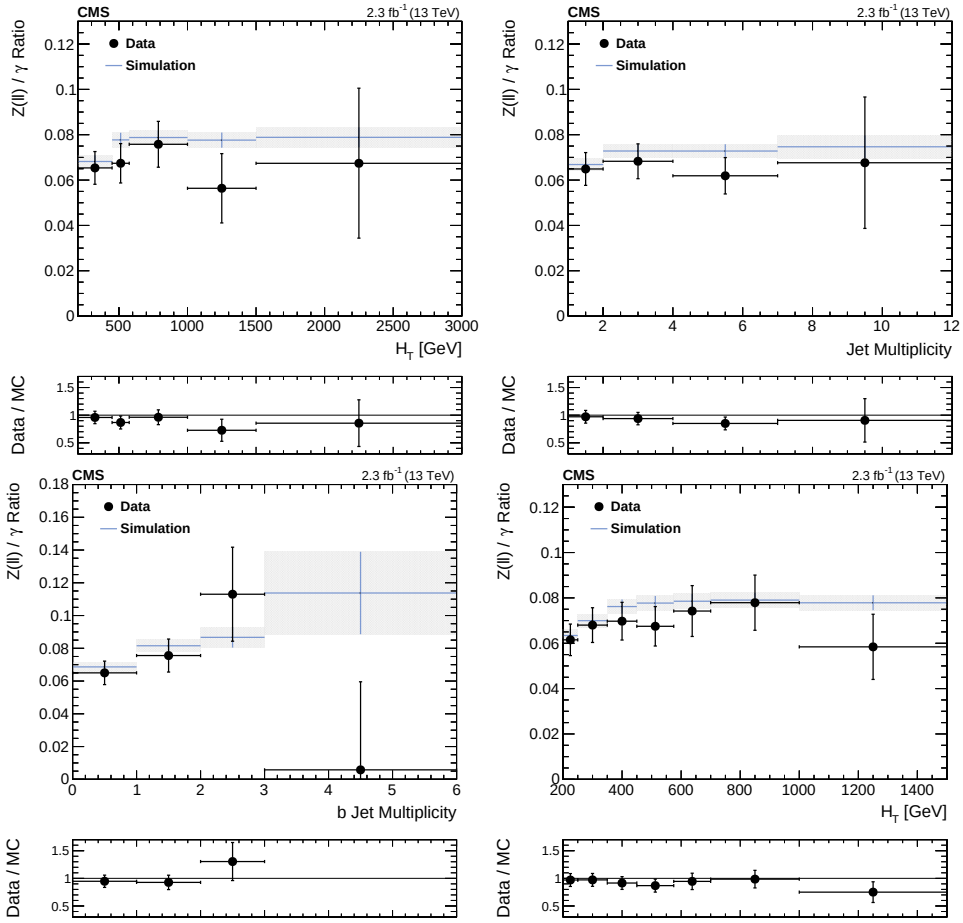

Figure 4.14: $R(\mathrm{Z} \to \ell^+\ell^-/\gamma)$ ratio in data (black markers) and in the simulation (blue lines) as a function of H_T (top left), N_jets (top right), and $N_\mathrm{b-tags}$ in the inclusive region (bottom left) and H_T binned for the monojet region (bottom right).

4.5 M_{T2} extrapolation

The γ+jets control region data is used to predict $\mathrm{Z} \to \nu\nu$ yields in bins of (H_T, $N_\mathrm{b-tags}$, N_jets), where H_T is equivalent to $p_\mathrm{T}^\mathrm{jet1}$ in the monojet regions. The remaining extrapolation in M_{T2} in the multijet regions is performed based on the $\mathrm{Z} \to \nu\nu$ shape predicted by MC in each H_T, $N_\mathrm{b-tags}$, N_jets region. The validity of this approach is tested in data, by comparing the simulated $\mathrm{Z} \to \nu\nu$ shape to the ones obtained from the photon and $\mathrm{W}^\pm \to$

$\ell\nu$ control regions in data. This is shown in Figure 4.15, where the M_{T2} shape obtained from the photon (red markers) and $\mathrm{W}^{\pm} \to \ell\nu$ control regions in data (green markers), is compared to the expected shape of the simulated $\mathrm{Z} \to \nu\nu$ process (black markers), in different H_{T} regions: $200 < H_{\mathrm{T}} < 450$ GeV (top left), $450 < H_{\mathrm{T}} < 575$ GeV (top right), $575 < H_{\mathrm{T}} < 1000$ GeV (center left), $1000 < H_{\mathrm{T}} < 1500$ GeV (center right), and finally $H_{\mathrm{T}} > 1500$ GeV (bottom). All distributions are normalized to have the same area. Events in the $\mathrm{W}^{\pm} \to \ell\nu$ control region explicitly require $N(b) = 0$. As can be seen, no significant discrepancy in the M_{T2} shape is observed between the simulation and the data estimates.

An extra systematic uncertainty is required to account for the MC extrapolation in M_{T2}, based on experimental and theoretical uncertainties. The dominant effects are the generator factorization and renormalization scales, as well as the jet energy scale uncertainties, and together they lead to variations that are at most 20% in the last bin. To take into account potential unknown effects such as NLO electroweak corrections to the $\mathrm{Z} \to \nu\nu$ process, we increase this uncertainty by a factor of two, to 40% in the last bin of each M_{T2} distribution. This uncertainty is also supported by Figure 4.15.

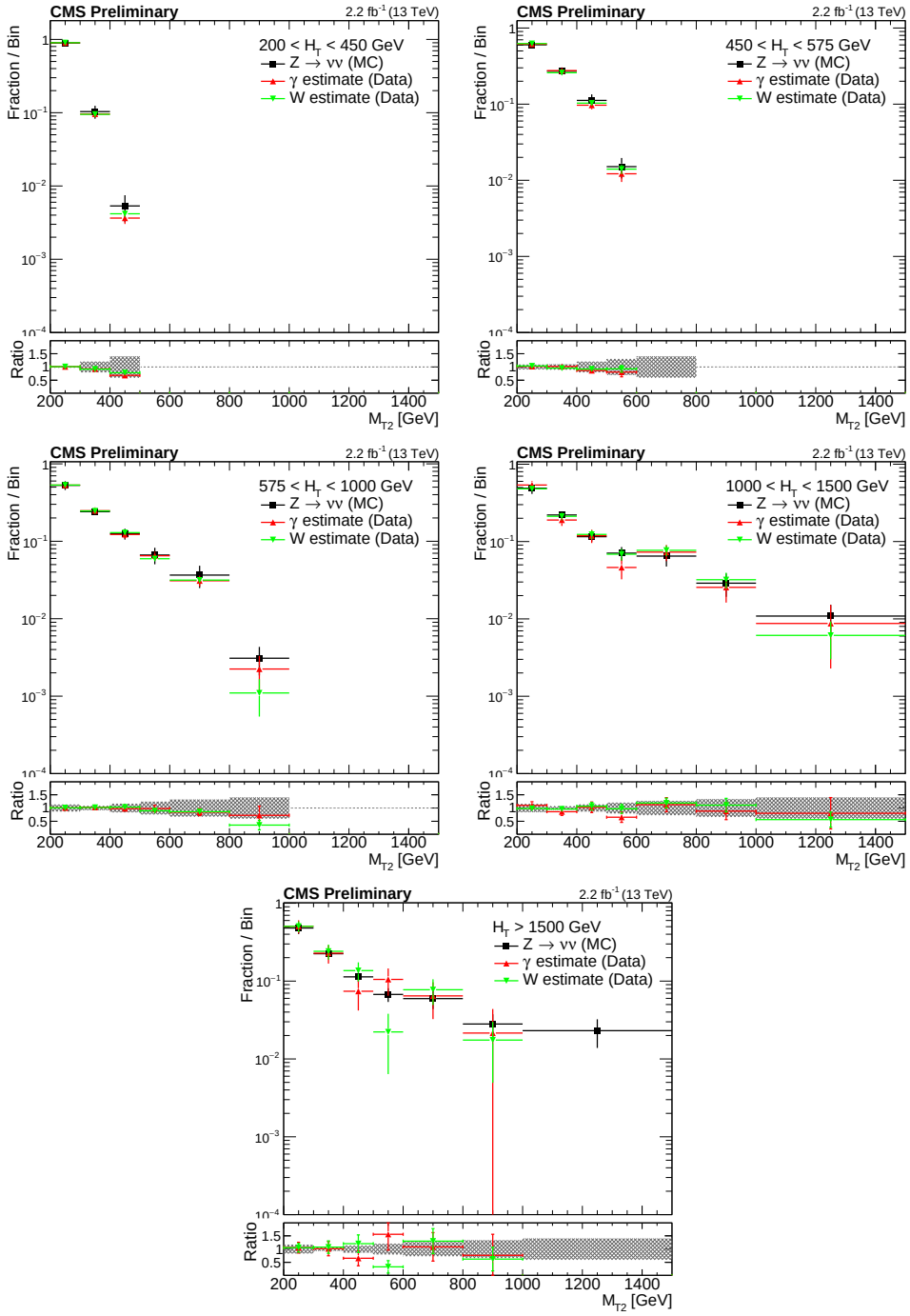

Figure 4.15: Shape comparison of the M_{T2} distribution in the photon (red markers) and $\mathrm{W}^{\pm} \to \ell\nu$ control (green markers) regions in data, compared to the expected shape of the simulated $\mathrm{Z} \to \nu\nu$ process (black markers), in different H_{T} regions: $200 < H_{\mathrm{T}} < 450$ GeV (top left), $450 < H_{\mathrm{T}} < 575$ GeV (top right), $575 < H_{\mathrm{T}} < 1000$ GeV (center left), $1000 < H_{\mathrm{T}} < 1500$ GeV (center right) and finally $H_{\mathrm{T}} > 1500$ GeV (bottom).

4.6 Systematic uncertainty on prediction

The following systematic uncertainties have been assessed on the $Z \to \nu\nu$ prediction:

- control region statistical error: based on the γ + jets control region statistics, uncorrelated across signal bins
- $R(Z/\gamma)$ (stat): based on MC statistics
- $R(Z/\gamma)$ (syst): approximately 20% uncertainty derived from the $R(Z \to \ell^+\ell^-/\gamma)$ measurement as discussed in Section 4.4. This uncertainty includes the $Z \to \ell^+\ell^-$ and γ + jets statistics uncertainty, as well as systematic uncertainty on top subtraction, but no systematic uncertainty on f and purity is included to avoid double counting with the uncertainties below. For each signal region, this uncertainty is composed of three separate nuisance parameters representing the H_T, N_{jets} and N_{b-tags} bins characterizing that region. Each nuisance parameter is correlated across regions sharing the same bin in one of these variables.
- f uncertainty: we take $f = 0.92 \pm 0.08$, correlated across all bins.
- purity (stat): the statistical uncertainty resulting from the template fit (or the Tight/Loose ratio method) is dominated by the yields in the isolation sideband. It is uncorrelated across signal bins.
- purity (syst): includes fake photon template uncertainty (based on the sigmaIetaIeta sideband) and prompt photon template uncertainty (based on random cone method or varying the fragmentation yield) resulting in a 5% uncertainty that is taken as correlated across all bins.
- M_{T2} shape uncertainty: based on MC variations accounting for theoretical (renormalization scale, factorization scale, PDFs) and experimental (JEC, MET) effects.

These effects give at most a 20% variation in the last bin for each topological region. The uncertainty is increased to 40% to account for possible EWK corrections not present in the MC and not tested through the above tests. This uncertainty is implemented as linear morphing with a maximum amplitude of 40% in the last bin of each topological region. Different topological regions are not correlated to each other.

Chapter 5

Lost Lepton Background

The lost lepton background consists primarily of the $t\bar{t}$ and W + jets processes. Smaller contributions come from other SM processes with at least one W boson and jets, such as single top, $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$. Events from these processes typically enter the search regions by having one leptonically-decaying W boson, providing real $E_{\mathrm{T}}^{\mathrm{miss}}$ from the neutrino, where the lepton is either out of acceptance, not isolated, not identified, or not reconstructed as a lepton at all. Lost lepton events make up a large fraction of the total background in nearly all search regions.

The main handle for reducing this background is improving the rejection of events with leptons, as described in Section 5.1. The veto results are incorporated into the analysis selections described in Chapter 3. We predict this background using control regions of single lepton events as described in Section 5.2, including systematic uncertainties on the M_{T2} shape. The uncertainty due to lepton selection efficiency is addressed in Section 5.4. The systematic uncertainties on this background are then summarized in Section 5.5.

5.1 Lepton veto

The most direct way to remove events with a true lepton from the signal regions is to detect and veto on the presence of a charged lepton. This is an optimization problem between maximizing efficiency for finding true leptons (to reduce the background) and minimizing the fake rate of misidentifying jets as leptons (which reduces signal efficiency).

Several objects and selections were investigated for the veto: POG Veto electrons, POG Loose Muons, Hadron Plus Strip (HPS) hadronic taus, PF leptons, and PF hadrons. We varied the p_{T} and isolation requirements, including investigating mini isolation, and also investigated vetoing only candidates with $M_{\mathrm{T}}(\mathrm{cand}, E_{\mathrm{T}}^{\mathrm{miss}}) < 100$ GeV.

For events with a single leptonically-decaying W boson, M_{T} should have an endpoint around the W mass. Meanwhile for signal events, the additional $E_{\mathrm{T}}^{\mathrm{miss}}$ from LSPs allows for higher values of M_{T}, even if the lepton candidate is a fake. Thus an M_{T} cut can effectively protect against inefficiency from fake leptons passing the veto selection in signal events. This can be seen in Figure 5.1. For PF leptons, there is a clear peak at the W boson mass for both $\mathrm{t\bar{t}}$ and $\mathrm{W}+\mathrm{jets}$ while the signal extends to larger values.

Isolated PF hadrons often arise from the charged pion in a one-prong hadronic tau decay, where some of the energy is lost to neutrinos and neutral pions, so the M_{T} distribution peaks even lower than the W mass for $\mathrm{t\bar{t}}$ and $\mathrm{W}+\mathrm{jets}$. QCD jets can also fluctuate to produce isolated PF hadrons, and as these have no relationship to the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event, they can give larger M_{T} values as seen in the high M_{T} tail for $\mathrm{t\bar{t}}$ and $\mathrm{W}+\mathrm{jets}$ in Figure 5.1 (right). Even for PF hadrons, however, an M_{T} cut of 100 GeV captures about 80% of $\mathrm{t\bar{t}}$ and $\mathrm{W}+\mathrm{jets}$ events. The M_{T} distribution for the T1bbbb 1500, 100 signal, which contains no true isolated leptons, extends to large values. We find that vetoing on low M_{T} PF candidates instead of using an explicit tau veto improves the background rejection by 25–30% without sacrificing signal efficiency.

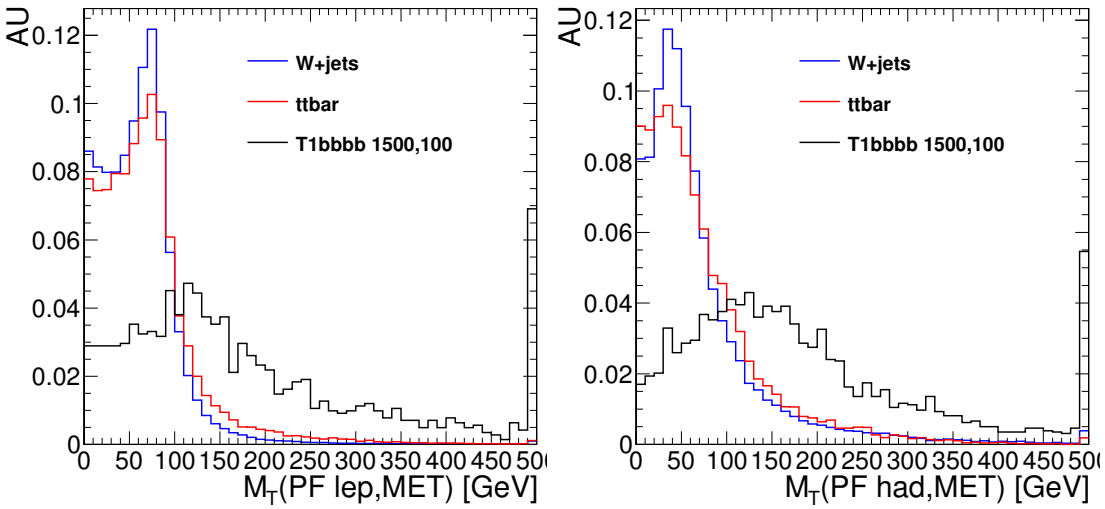

Figure 5.1: Distributions, normalized to unity, of $M_{\mathrm{T}}(\mathrm{cand}, E_{\mathrm{T}}^{\mathrm{miss}})$ for (left) PF leptons and (right) PF hadrons. The PF leptons are required to pass $p_{\mathrm{T}} > 5$ GeV and iso/$p_{\mathrm{T}} < 0.2$, while the PF hadrons are required to pass $p_{\mathrm{T}} > 10$ GeV and iso/$p_{\mathrm{T}} < 0.1$.

After investigating several options for the lepton veto, it was decided to veto an event containing any of the following:

- isolated reco electron with $p_{\mathrm{T}} > 10$ GeV
- isolated reco muon with $p_{\mathrm{T}} > 10$ GeV
- isolated PF lepton with $p_{\mathrm{T}} > 5$ GeV and $M_{\mathrm{T}} < 100$ GeV
- isolated PF hadron with $p_{\mathrm{T}} > 10$ GeV and $M_{\mathrm{T}} < 100$ GeV

where the full details of selections for the veto objects is given in Chapter 3. Table 5.1 shows the efficiency of the lepton veto for Top and W + jets as well as the the efficiency for two signal points.

Table 5.1: Efficiency of the lepton veto for background and signal after all other baseline selection cuts are applied. For T1tttt, only events with no generator level leptonic W decays are considered.

Veto Selection	Top	W + jets	T1tttt 1500,100	T1bbbb 1500,100
electron $p_{\mathrm{T}} > 10$ GeV (mini iso) muon $p_{\mathrm{T}} > 10$ GeV (mini iso) PF lepton $p_{\mathrm{T}} > 5$ GeV, $M_{\mathrm{T}} < 100$ GeV PF hadron $p_{\mathrm{T}} > 10$ GeV, $M_{\mathrm{T}} < 100$ GeV	0.28	0.39	0.92	0.96

5.2 Prediction using control regions

The lost lepton background prediction is made in a data-normalized way using single lepton control regions. To define the single lepton control regions, the baseline selections of Section 3.3 are applied, with the exception of the lepton veto. Instead, we require exactly one candidate passing the reco lepton or PF lepton selections (e or μ only). Often a candidate is found as both a reco lepton and a PF lepton. To avoid double counting, we do not count PF leptons within $\Delta\mathrm{R} < 0.1$ of a reco lepton. We further require $M_{\mathrm{T}}(\mathrm{cand}, E_{\mathrm{T}}^{\mathrm{miss}}) < 100$ GeV in order to reduce potential signal contamination, where a signal event enters a control region and biases the background estimate. These regions use the same triggers as the analysis signal regions.

We further subdivide the events into the categories described in Section 3.4, binning the single lepton control regions in the H_{T}, N_{jets}, and $N_{\mathrm{b-tags}}$ dimensions but not in M_{T2} to preserve statistics. The binning in N_{jets} and $N_{\mathrm{b-tags}}$ is the same as the signal regions except for signal bins with $\geq$7j,$\geq$1b. These are all predicted using control region bins with the same H_{T} selection as the signal bin and $\geq$7j,1–2b. This is motivated by the low control region statistics in bins with $\geq$7j,$\geq$2b as well as potential signal contamination in bins with $\geq$7j,$\geq$3b. In the monojet selection, the single lepton control region is binned in the same $p_{\mathrm{T}}^{\mathrm{jet1}}$ and $N_{\mathrm{b-tags}}$ bins as the signal region, so there is no kinematic extrapolation.

When a true lepton is within detector acceptance, it is usually reconstructed in some

form, even if not found as an isolated lepton candidate. Thus it will likely form a PF jet, if its p_{T} is above the jet threshold. To emulate this effect in the single lepton control region, we remove the closest jet within $\Delta\mathrm{R} < 0.4$ of the lepton and instead count the lepton as a "jet" for the purposes of computing these variables: N_{jets}, $N_{\mathrm{b-tags}}$, H_{T}, $H_{\mathrm{T}}^{\mathrm{miss}}$, $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$, $|\vec{H}_{\mathrm{T}}^{\mathrm{miss}} - \vec{E}_{\mathrm{T}}^{\mathrm{miss}}|/E_{\mathrm{T}}^{\mathrm{miss}}$, the two pseudojets, and M_{T2}.

Signal contamination was checked for several signal benchmark points in all regions. The only signal which showed potential contamination issues was gluino pair production with $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$ (a simplified SUSY model known as "T1tttt"), due to the large branching fraction to leptons and higher jet and bjet multiplicity than SM backgrounds. The contributions from other signals were negligible to our control regions. With the regions defined as above, the contamination from T1tttt is small compared to the expected statistical error on the control region yields, maximally 20% at the signal strength for our expected limits. We therefore consider this effect negligible compared to the uncertainties we assign.

With the above definition of the single lepton control regions, the lost lepton background yield in each signal region is estimated from the corresponding control region as follows:

$$N_{1l}^{\mathrm{SR}}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b}, M_{\mathrm{T2}}) = N_{1l}^{\mathrm{CR}}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b}) \times R_{\mathrm{MC}}^{0l/1l}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b}) \times k_{\mathrm{MC}}(M_{\mathrm{T2}}) \qquad (5.1)$$

where:

- $N_{1l}^{\mathrm{CR}}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b})$ is the observed single lepton yield in the control region;
- $R_{\mathrm{MC}}^{0l/1l}(H_{\mathrm{T}}, \mathrm{N_j}, \mathrm{N_b})$ is the fraction of events for which an isolated lepton candidate is not identified or reconstructed. This factor is obtained from simulation with small corrections to account for differences in lepton efficiency between data and

simulation;

- $k_{\mathrm{MC}}(M_{\mathrm{T2}})$ is the fraction of events in each topological region expected to populate an M_{T2} bin and is obtained from simulation.

The normalization to data accounts for the kinematic modeling of all variables except M_{T2} (and $N_{\mathrm{b-tags}}$ in the case of $\geq$7j,$\geq$1b). The systematic uncertainty on the extrapolation in the remaining variables comes from variations in MC of experimental and theoretical uncertainties, including factorization and renormalization scales, PDFs, jet energy scale uncertainties (propagated to the $E_{\mathrm{T}}^{\mathrm{miss}}$), and b-tagging scale factor uncertainties. The largest variations seen are around 15% in the highest M_{T2} bins from theoretical uncertainties, and variations of up to 40% are seen in low statistics bins from jet energy scale variations. The single lepton control region has typically 1–2 times as many expected events as the signal selection in a given H_{T}, N_{jets}, and $N_{\mathrm{b-tags}}$ bin, and the control region statistical error is propagated to the uncertainty on the prediction.

Figures 5.2–5.4 show distributions of kinematic variables in the single lepton baseline control region, with at least 2 jets and $H_{\mathrm{T}} > 200$ GeV, in data compared with the expectations from MC. Figure 5.2 shows the nominal control region including both electrons and muons, while Figs. 5.3 and 5.4 show electrons and muons separately. The shapes and normalization agree well between electrons and muons. In general, compared to data, MC has fewer events at low jet multiplicity and an H_{T} spectrum that falls more slowly compared to data. After applying b-tagging scale factors, the MC distribution of $N_{\mathrm{b-tags}}$ agrees fairly well with data. The modeling of M_{T2} also looks consistent within statistical errors.

To validate the modeling of W + jets and $\mathrm{t\bar{t}}$ kinematics, we select on $N_{\mathrm{b-tags}} = 0$ and $N_{\mathrm{b-tags}} \geq 2$ to create regions of high purity for each process. The distributions are shown in Figure 5.5 for the W+jets enhanced region and in Figure 5.6 for the $\mathrm{t\bar{t}}$ enhanced

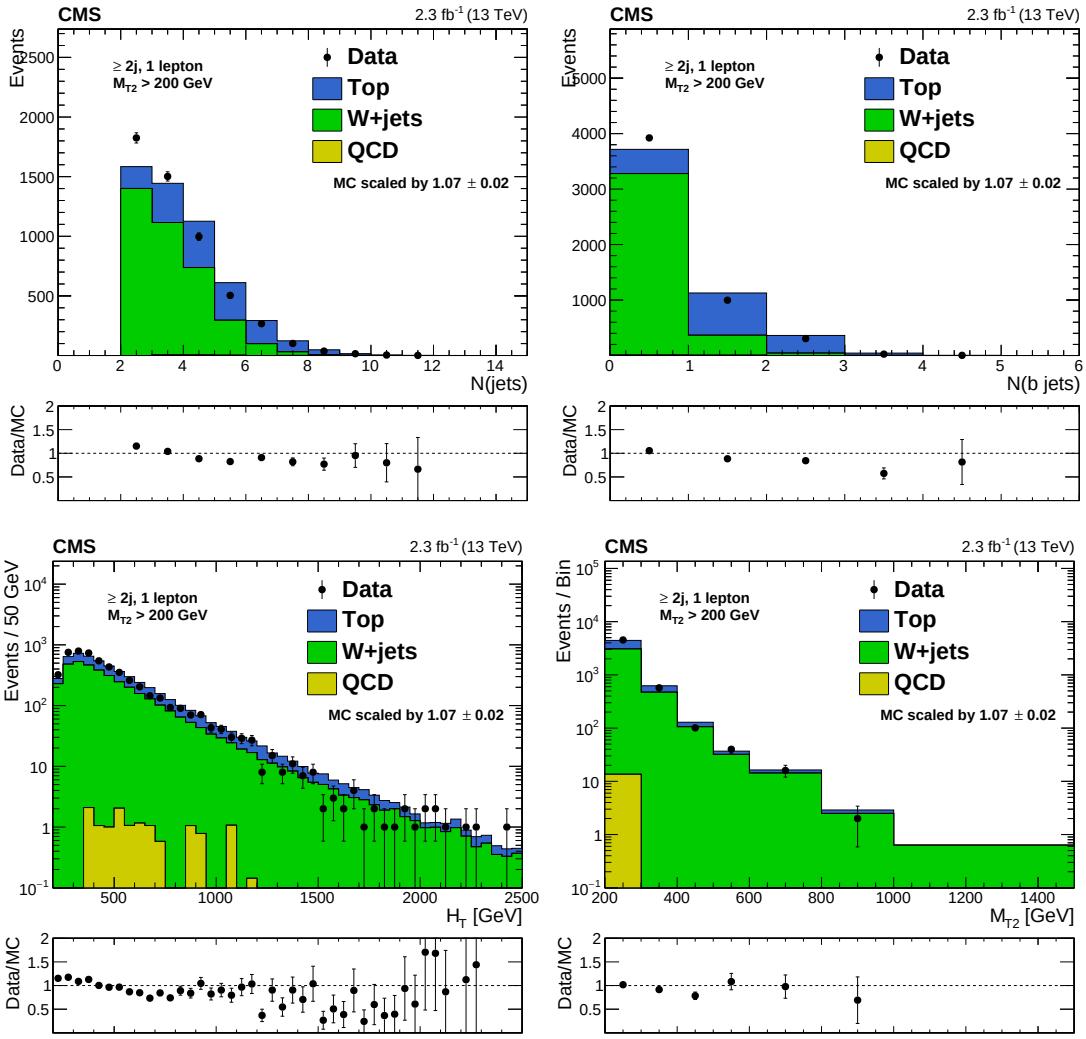

Figure 5.2: Distributions of data and MC predictions for the baseline single lepton control region selection with at least 2 jets and $H_{\mathrm{T}} > 200$ GeV. Shown are (top left) N_{jets}, (top right) $N_{\mathrm{b-tags}}$, (bottom left) H_{T}, and (bottom right) M_{T2}. MC is normalized to data with the scaling given in the figures.

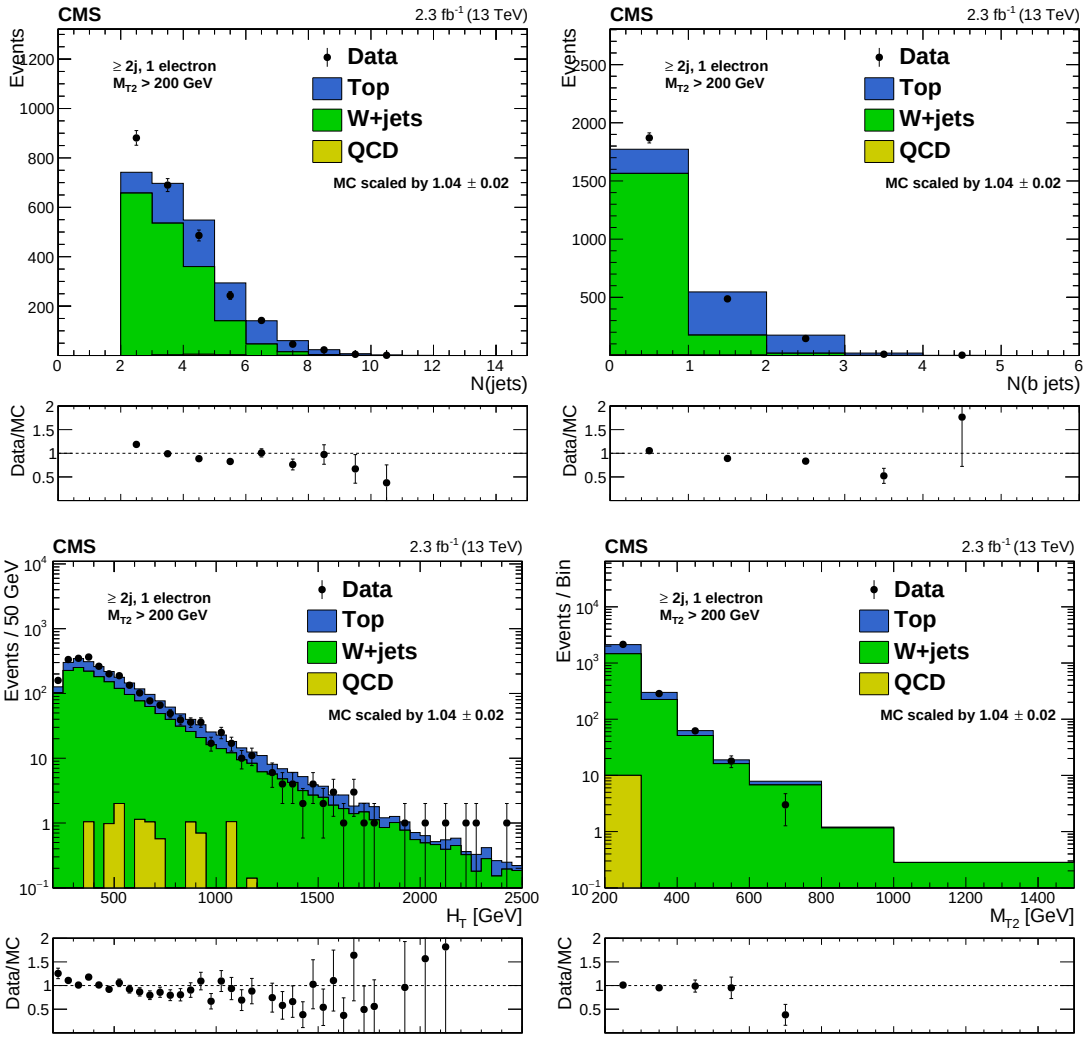

Figure 5.3: Distributions of data and MC predictions for the baseline single lepton control region selection with an electron, at least 2 jets, and $H_{\mathrm{T}} > 200$ GeV. Shown are (top left) N_{jets}, (top right) $N_{\mathrm{b-tags}}$, (bottom left) H_{T}, and (bottom right) M_{T2}. MC is normalized to data with the scaling given in the figures.

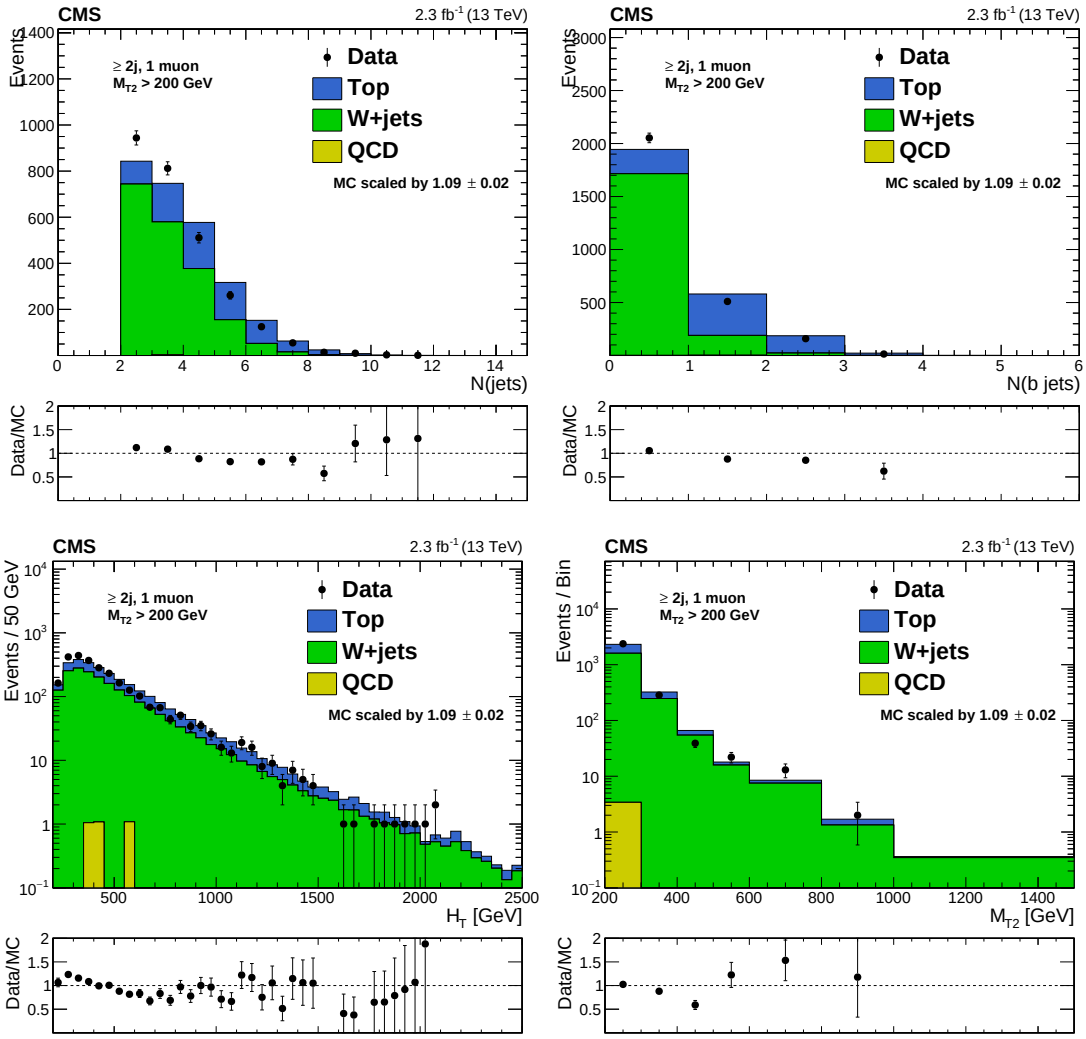

Figure 5.4: Distributions of data and MC predictions for the baseline single lepton control region selection with a muon, at least 2 jets, and $H_{\mathrm{T}} > 200$ GeV. Shown are (top left) N_{jets}, (top right) $N_{\mathrm{b-tags}}$, (bottom left) H_{T}, and (bottom right) M_{T2}. MC is normalized to data with the scaling given in the figures.

region. The trends seen in the inclusive region persist in these.

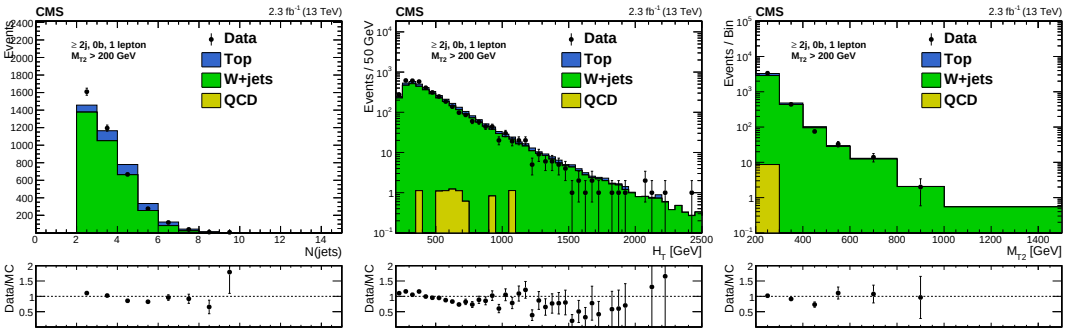

Figure 5.5: Distributions of data and MC predictions for the single lepton control region selection enhanced in W + jets: at least 2 jets, no b-tagged jets, and $H_{\mathrm{T}} > 200$ GeV. Shown are (left) N_{jets}, (center) H_{T}, and (right) M_{T2}. MC is normalized to data with the scaling given in the figures.

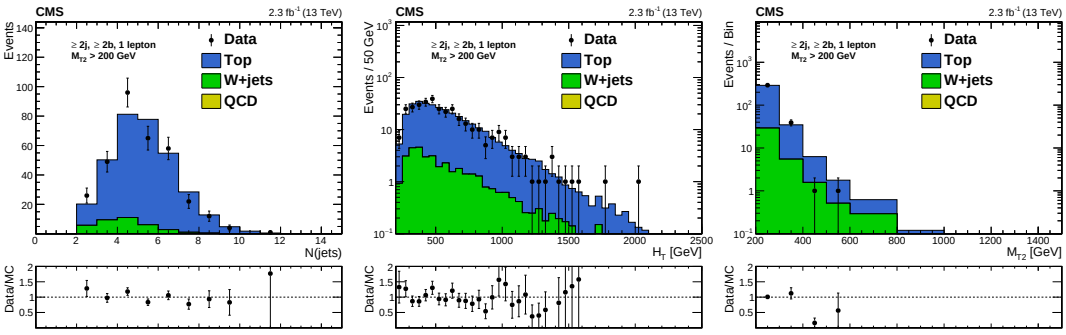

Figure 5.6: Distributions of data and MC predictions for the single lepton control region selection enhanced in $\mathrm{t\bar{t}}$: at least 2 jets, at least 2 b-tagged jets, and $H_{\mathrm{T}} > 200$ GeV. Shown are (left) N_{jets}, (center) H_{T}, and (right) M_{T2}. MC is normalized to data with the scaling given in the figures.

Since the M_{T2} shape is taken from MC for the background prediction method, we further check the M_{T2} shape in bins of H_{T} as well as the $p_{\mathrm{T}}^{\mathrm{jet1}}$ shape in the monojet region. The results are shown in Fig. 5.7. Generally good agreement is seen, with a couple discrepant bins in the distribution for $1000 < H_{\mathrm{T}} < 1500$ GeV. The background estimate procedure essentially normalizes MC to data in bins of H_{T}, N_{jets}, and $N_{\mathrm{b-tags}}$. We show the M_{T2} shape in Fig. 5.8 after this is done, comparing data to the sum of

lost lepton backgrounds, for different b-tag selections. Again good agreement is seen. Finally, as another way of looking at the M_{T2} shape modeling, we compare our nominal prediction, which extrapolates using MC in M_{T2}, to a "bin-by-bin" prediction which takes these shapes from data instead. From these studies we conclude that an M_{T2} shape uncertainty with 40% change in the maximal M_{T2} bin is sufficient to cover any mismodeling by MC.

5.3 Signal contamination

Despite the selections intended to reduce contributions from signal to the single lepton control regions, signal contamination can be non-negligible in some regions of phase space where the signal is kinematically similar to the background. A contribution from signal to the control region would result in an overestimation of the lost lepton background.

To account for this in our interpretations, we treat the amount by which the lost lepton background would be overestimated as a reduction in signal efficiency. Specifically, in each analysis bin, we define:

$$N_{sig}^{SR\prime} = N_{sig}^{SR} - TF \cdot N_{sig}^{CR} \tag{5.2}$$

where N_{sig}^{SR} and N_{sig}^{CR} are the predicted signal in the signal region and control region bins, respectively, and TF is the transfer factor from control to signal region used in the lost lepton estimate. Then the quantity $N_{sig}^{SR\prime} \leq N_{sig}^{SR}$ is used in calculating the limit on the signal cross section.

This treatment has been used in several other SUSY analyses (e.g. [25]) and has the useful property that $N_{sig}^{SR\prime}$ depends linearly on the signal cross section. This can be seen

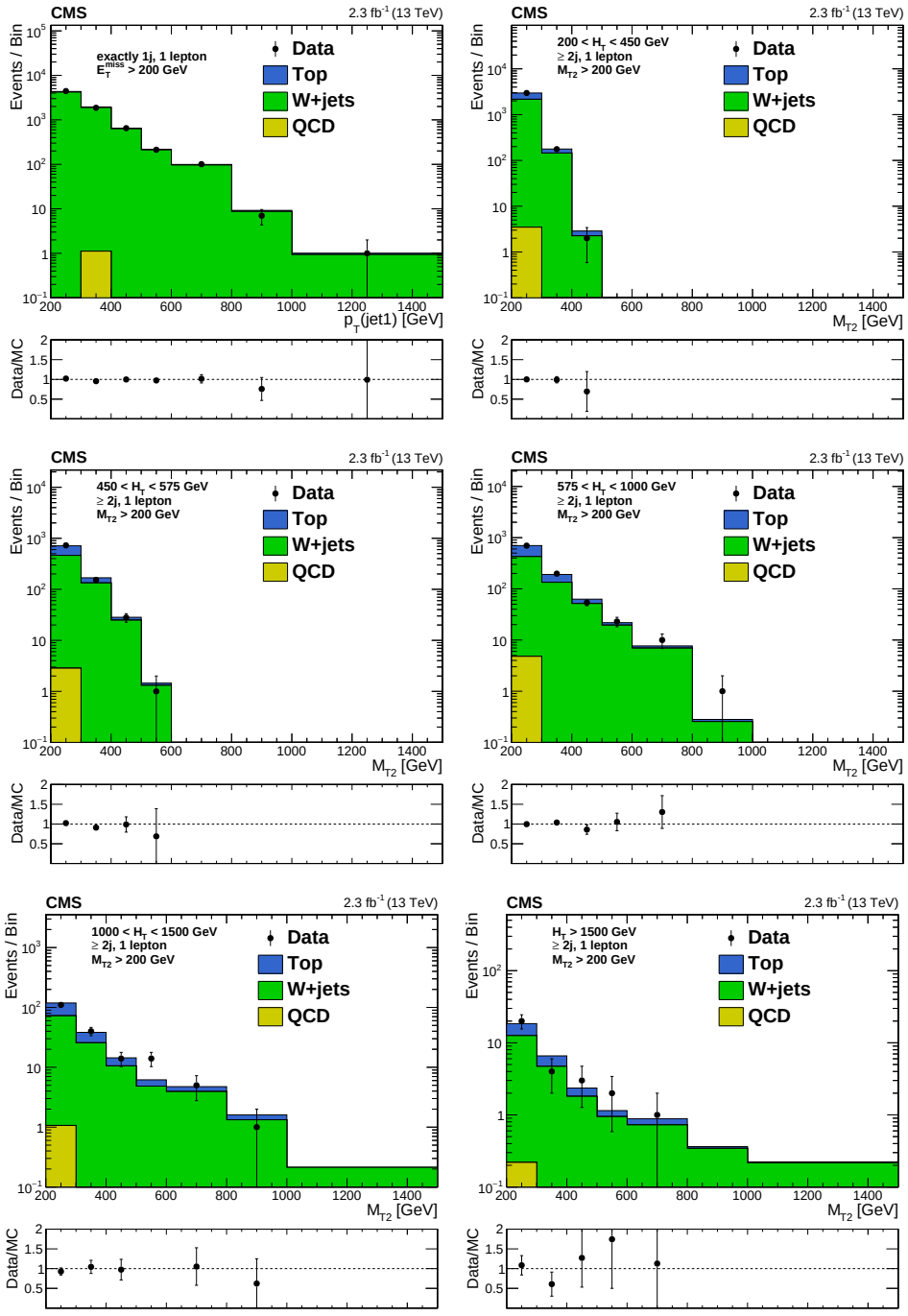

Figure 5.7: Distributions of $p_{\mathrm{T}}^{\mathrm{jet1}}$ or M_{T2} data and MC predictions for the single lepton control regions in the monojet region (top left) or in the analysis H_{T} bins for regions with at least 2 jets. MC is normalized to data in each plot to compare the shapes.

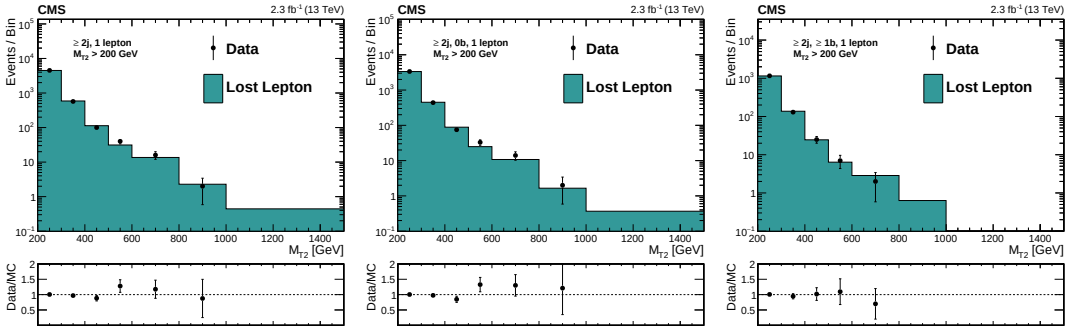

Figure 5.8: Distributions of data and MC predictions for the single lepton control region selection, after MC is normalized to data in the CR bins of H_{T}, N_{jets}, and $N_{\mathrm{b-tags}}$, for an inclusive selection (left), events with no b-tags (center), and events with at least one b-tag (right).

by rewriting as below:

$$N_{sig}^{SR\prime} = \sigma_{sig} \cdot \mathcal{L} \cdot (\epsilon_{sig}^{SR} - TF \cdot \epsilon_{sig}^{CR}) \tag{5.3}$$

where σ_{sig} is the signal cross section, $\mathcal{L}$ is the luminosity, and ϵ_{sig}^{SR} and ϵ_{sig}^{CR} are the efficiencies for the signal to populate the signal region and control region, respectively.

5.4 Systematic uncertainties on lepton efficiency

Tag and probe with $\mathrm{Z} \to \ell^+\ell^-$ events is used to evaluate the lepton (e, μ) efficiencies in data. We use the results of the official SUSY PAG tag and probe group, which have been approved by the respective POGs [26]. Identification efficiencies are computed in bins of lepton p_{T} and η, while isolation efficiencies are computed in bins of p_{T} and nearby activity. The scale factors are compatible with unity with uncertainties on the efficiency ranging from one to a few percent.

The uncertainty on lepton selection efficiency affects both the number of events we expect in our single lepton control region as well as the number of events in our lepton-

vetoed signal region. For example, if the lepton efficiency is varied up, we expect to see more events in our control region and fewer events in our signal region. We evaluate the effect of the lepton efficiency uncertainty on the transfer factor from the control region to the signal region in each bin of H_{T}, N_{jets}, $N_{\mathrm{b-tags}}$ in Fig. 5.9. The maximal effect seen is at the level of 3%, and we take this as the uncertainty on the background estimate.

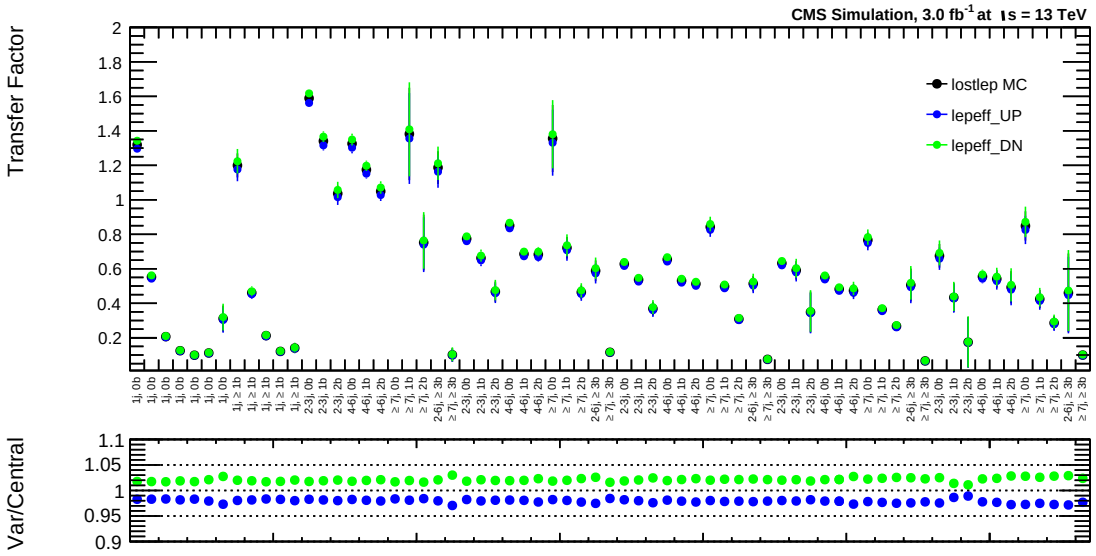

Figure 5.9: Effect of varying the lepton (e, μ) efficiency by its uncertainty on the transfer factor between the control and signal region, for each bin of H_{T}, N_{jets}, $N_{\mathrm{b-tags}}$. From left to right are the monojet regions then the multijet regions in order of increasing H_{T}.

The efficiency for hadronically decaying tau leptons has also been checked in MC, since getting a pure sample of hadronic taus by selecting tracks is not possible in data. The isolation efficiency is seen as a function of PF candidate p_{T} for the different lepton flavors in Fig. 5.10. For this plot, the cut applied to all flavors is trkiso/$p_{\mathrm{T}} < 0.1$, the nominal cut for PF hadrons, to compare the different flavors on a common baseline. The 1-prong hadronic taus are also shown divided into those with no π^0 mesons in the decay compared to those with at least one π^0. The isolation efficiency is consistent between electrons, muons, and 1-prong taus with no π^0 mesons, while 1-prong taus with at least

one π^0 have lower efficiency. This is due to conversions of the photons from the π^0 decay, and most 1-prong taus have at least one π^0 in the decay. Since this effect is well-understood, we take half of the difference in isolation efficiency between 1-prong taus (the black curve) and muons as an uncertainty on our tau selection, which ranges between 4–7% on the tau efficiency depending on p_{T}. We take 7% and find that it also covers the differences in total veto efficiency as a function of the main kinematic variables (Fig. 5.11). For 3-prong taus, our PF hadron veto is very inefficient and most of them fail isolation. The typical selection efficiency is around 6%, as seen in Fig. 5.11, and we take a 100% relative uncertainty on this.

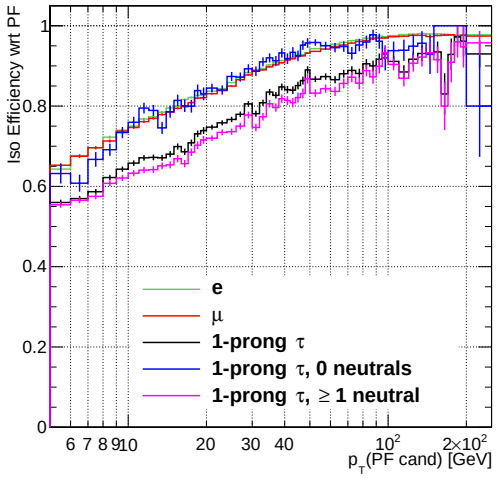

Figure 5.10: Isolation efficiency for a cut of trkiso/$p_{\mathrm{T}} < 0.1$ for PF candidates as a function of candidate p_{T}, broken down by lepton flavor. For hadronic taus, the efficiency is also shown for those with no π^0 mesons in the decay compared to those which have at least one π^0.

The uncertainty on tau selection efficiency is propagated to the full lost lepton background prediction in MC, considering only taus which are in p_{T} acceptance (the leading charged daughter at generator level must have $p_{\mathrm{T}} > 10$ GeV). The result on the prediction can be seen in Fig. 5.12. For 1-prong taus, it is always less than 3% for H_{T} bins

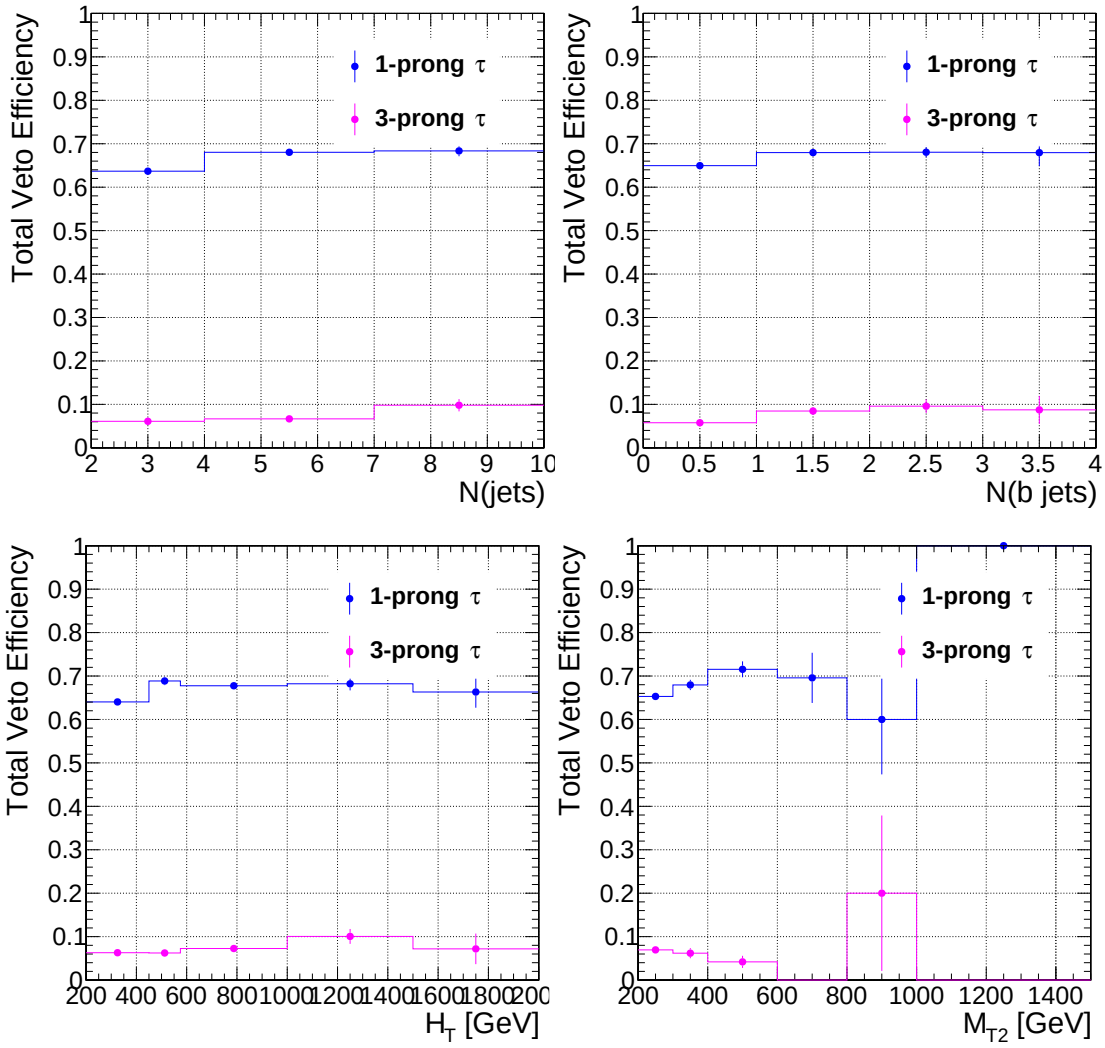

Figure 5.11: Total veto efficiency for tau leptons as a function of the main kinematic variables, defined as the fraction of events with a generated hadronic tau in p_{T} acceptance (p_{T} of leading daughter greater than 10 GeV) which are rejected by any component of our lepton veto.

below 1500 GeV and within 6% for $H_{\mathrm{T}} > 1500$ GeV. For 3-prong taus, it is always less than 1% in all bins. Although not shown on this plot, the monojet regions were also checked and found to be within 3% for 1-prong taus and 1% for 3-prong taus.

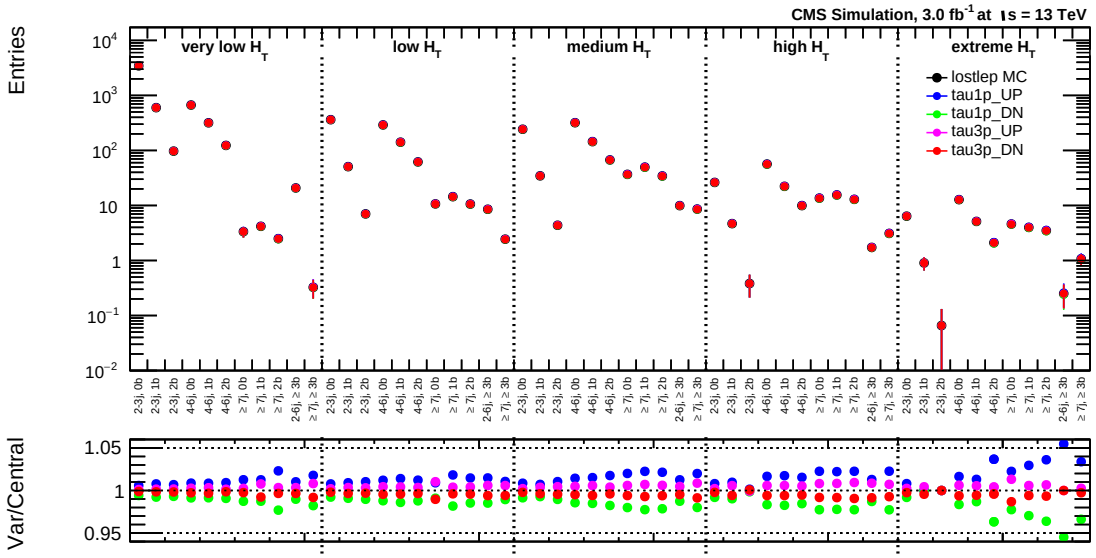

Figure 5.12: Effect of varying the tau selection efficiency on the total lost lepton background prediction for each topological region.

We also depend on the MC modeling of the lepton veto and M_{T} cut efficiency to extrapolate from the control regions to the signal regions. We define a control region using $\mathrm{Z} \to \ell^+\ell^-$ events with one lepton treated as "missing" to mimic a leptonic W event and validate this modeling in data. From data/MC comparisons in the control region, we assign a systematic uncertainty of 3% on the M_{T} cut efficiency modeling in MC.

5.5 Systematic uncertainties on prediction

The following systematic errors have been assessed on the lost lepton prediction:

- control region statistical error: the poisson error on the observed data count in each control region bin is used. This error is taken as correlated among all signal bins

that share the same control region bin.

- signal region MC statistical error: this is accounting for the fact that MC will be used to extrapolate to the expected yields in the signal regions. It uses the available MC statistics in a signal bin and ranges between 1–50%, depending on the bin. This error is taken as uncorrelated in all signal bins.

- lepton selection efficiency: 7% based on lepton efficiency studies. This covers the $\leq 3\%$ effect of the lepton (e, μ) selection efficiency, the 3% effect for the M_{T} cut efficiency, the $\leq 6\%$ effect for 1-prong taus and the 1% effect for 3-prong taus all described in the previous section. This error is taken as correlated among all signal bins that share the same control region bin.

- transfer factor from control to signal region: 10–40% based on MC variation studies. This includes the effect of MC theoretical variations on lepton acceptance, jet energy scale uncertainties, and also the impact of b-tagging scale factor uncertainties on the $\geq$7j,$\geq$1b bins. The largest uncertainty, 40%, is assigned to bins with $200 < H_{\mathrm{T}} < 450$ GeV and $N_{\mathrm{jets}} \geq 7$ due to sensitivity to (and low statistics for) jet energy scale variations. This error is taken as correlated among all signal bins that share the same control region bin.

- M_{T2} shape uncertainty: for multijet bins only, based on MC variation studies and validated with comparisons of MC to data in the single lepton control region. This uncertainty is implemented as linear morphing with a maximum amplitude of 40% in the last bin of each topological region. Different topological regions are not correlated to each other.

Chapter 6

QCD Background

QCD multijet events contain no genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ and are strongly suppressed by the baseline selection, especially the $M_{\mathrm{T2}} > 200$ GeV requirement. Any residual background comes from jet mismeasurement resulting in large experimental $E_{\mathrm{T}}^{\mathrm{miss}}$. The QCD background in signal regions with at least two jets is estimated from data control regions obtained by inverting the $\Delta\phi_{\mathrm{min}}$ selection as described in Section 6.1. An alternative estimation method to cross check these results is described in Chapter 7. The QCD background in the monojet signal regions is estimated from dijet control regions in data with one hard jet and one soft jet as described in Section 6.2.

6.1 Multijet prediction using the $\Delta\phi$-ratio method

The multijet background consists predominantly of light-flavor and gluon multijet events. The baseline cut of $M_{\mathrm{T2}} > 200$ GeV rejects much of this background and the remaining yields are estimated using data control samples. For events with at least two jets, a multijet enriched control sample for each H_{T} bin is obtained by inverting the requirement $\Delta\phi_{\mathrm{min}} = \Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$, described in section 3.1.8. The control regions are

then used to predict the multijet yield in the signal regions by extrapolating from low to high $\Delta\phi_{\text{min}}$. For the bins with $H_{\text{T}} > 1$ TeV, the control region events are selected using the signal triggers. The events in the control regions for bins with $H_{\text{T}} < 1$ TeV are selected using prescaled H_{T} triggers, as the low H_{T} unprescaled signal triggers also have a large online $E_{\text{T}}^{\text{miss}}$ requirement.

To extrapolate from low to high $\Delta\phi_{\text{min}}$, we use the ratio

$$r_{\phi}(M_{\text{T2}}) = N(\Delta\phi_{\text{min}} > 0.3)/N(\Delta\phi_{\text{min}} < 0.3), \tag{6.1}$$

defined as the ratio of the number of events with $\Delta\phi_{\text{min}} > 0.3$ to the number of events with $\Delta\phi_{\text{min}} < 0.3$. From simulation, the functional form of this ratio as a function of M_{T2} is found to be well described by a power law function:

$$r_{\phi}(M_{\text{T2}}) = \frac{N(\Delta\phi_{\text{min}} > 0.3)}{N(\Delta\phi_{\text{min}} < 0.3)} = a \cdot M_{\text{T2}}^{b}, \tag{6.2}$$

for $M_{\text{T2}} > 50$ GeV (see Figure 6.1). The functional form is parameterized by two values a and b. These parameters are obtained from a fit to data in an M_{T2} sideband.

For regions with $H_{\text{T}} < 1000$ GeV, a fit window of $60 < M_{\text{T2}} < 100$ GeV is used. For the high and extreme H_{T} regions, statistics allow for a slightly smaller fit window of $70 < M_{\text{T2}} < 100$ GeV which improves the fit. The choice of the upper threshold of the fit window is chosen such that the contribution of events from electroweak and top-quark production is maintained to be small with respect to QCD multijet. With the current integrated luminosity, it is not possible to perform a reliable fit to the ratio in each topological region. Instead, the $\Delta\phi_{\text{min}}$ ratio is fit in each of the five H_{T} regions inclusively in N_{jets} and $N_{\text{b-tags}}$, and factors measured in data are used to obtain the yield in each N_{jets} and $N_{\text{b-tags}}$ bin. Figure 6.1 shows the expected ratio from simulation in

each of the five H_{T} regions.

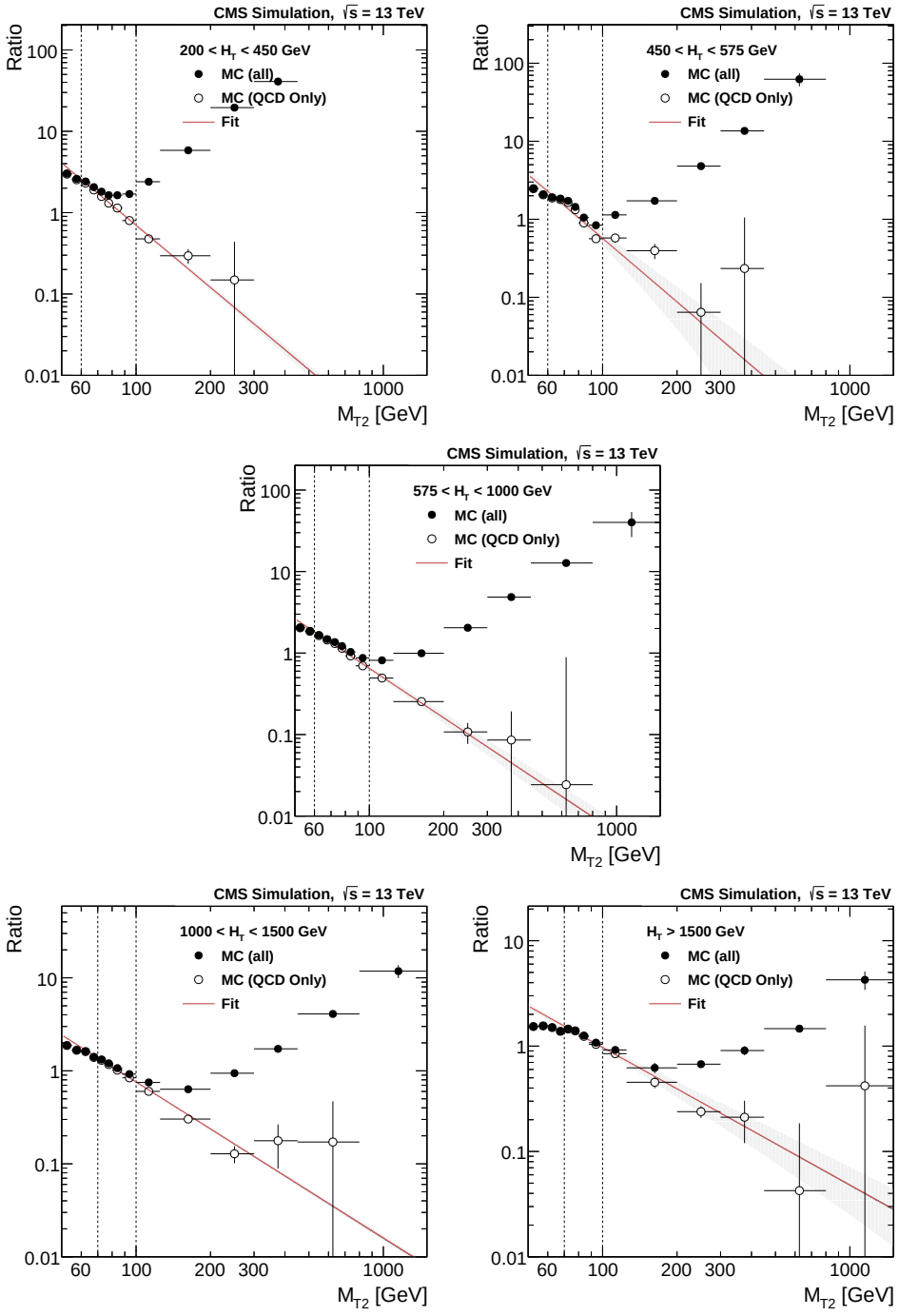

Figure 6.1: Expected distributions from simulation of the ratio $r_\phi(M_{\mathrm{T2}}) = N(\Delta\phi_{\mathrm{min}} > 0.3)/N(\Delta\phi_{\mathrm{min}} < 0.3)$ as a function of M_{T2}, for the low (top left), medium (top right), high (medium left), and extreme (medium right) H_{T} regions. The full points represent the ratio as predicted from simulation using all background components, while the hollow points represent the expected ratio from QCD multijet. The errors are MC statistics. The red line and the band around it show the fit to a power law function perfomed in the fit window, with its associated fit uncertainties.

A systematic uncertainty is assigned to cover variations in the r_ϕ value by choosing different fit ranges. Two alternative fit ranges are chosen by shifting the lower boundary one bin to the right or to the left. In order to recover statistics in the right variation, also the higher boundary is shifted one bin to the right. The systematic uncertainty is taken as the maximal deviation among the variations with respect to the default fit window. For $M_{\mathrm{T2}} > 200$ GeV this systematic is found to range between 60–138% (low H_{T}), 18–42% (medium H_{T}), 16–44% (high H_{T}) and 60–200% (extreme H_{T}), increasing with M_{T2}.

Figure 6.2 shows the ratio from data with an integrated luminosity of 2.3 fb^{-1}. The full markers in these plots show the ratio for all the data in each inverted $\Delta\phi$ control region, including non-QCD events. The hollow markers show the ratio after non-QCD events are subtracted, where the non-QCD yield is taken from MC. Due to high trigger prescales, the statistics of the low and medium H_{T} regions are small, resulting in large fit uncertainties that increase with M_{T2}. For the very low H_{T} region there is currently no trigger that allows access to the low M_{T2} region used to perform the fit. For this reason the fit parameters obtained from the low H_{T} region are also used to predict the QCD yields in the very low H_{T} region. This choice is justified by simulation and validated in the closure tests discussed later.

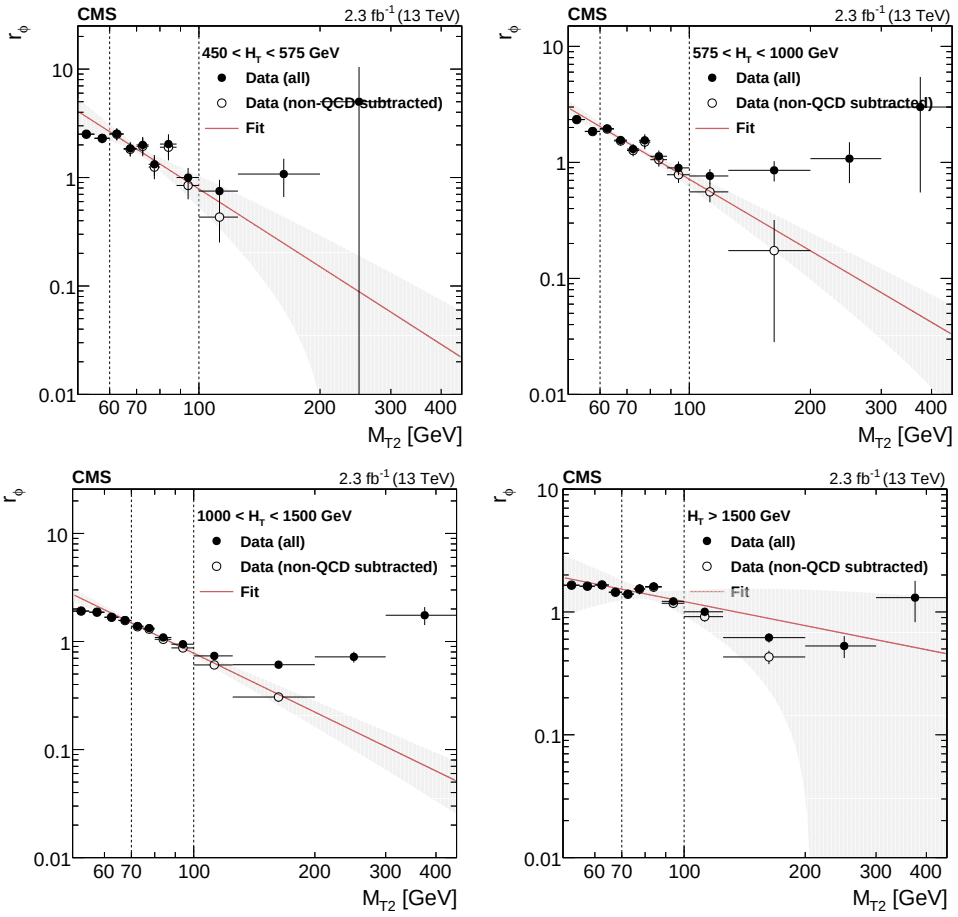

Figure 6.2: Distributions from data of the ratio $r_\phi(M_{\mathrm{T2}}) = N(\Delta\phi_{\mathrm{min}} > 0.3)/N(\Delta\phi_{\mathrm{min}} < 0.3)$ as a function of M_{T2}, for the low (top left), medium (top right), high (bottom left), extreme (bottom right) H_{T} regions. The full points represent the ratio from data before subtracting the non-QCD component, while the hollow points represent the data after the non-QCD contribution has been subtracted. The data in the high and extreme H_{T} regions has been collected with the unprescaled HLT_PFHT800 trigger, while for the low and the medium H_{T} regions the prescaled HLT_PFHT350 (prescale 180) and HLT_PFHT475 (prescale 60) triggers have been used. The red line and the band around it show the fit to a power law function perfomed in the fit window, with its associated fit uncertainties.

From the inverted $\Delta\phi_{\text{min}}$ control regions for each H_{T} region and the corresponding $\Delta\phi_{\text{min}}$ ratio fits, we can obtain an estimate for the QCD yields in each H_{T} – M_{T2} region, inclusive in N_{jets} and $N_{\text{b-tags}}$:

$$N_{\text{inc}}^{SR}(H_{\text{T}}, M_{\text{T2}}) = N_{\text{inc}}^{CR}(H_{\text{T}}) \cdot r_{\phi}(M_{\text{T2}}). \tag{6.3}$$

In order to estimate the QCD yields in each signal region, specified by some H_{T} – M_{T2} – N_{jets} – $N_{\text{b-tags}}$ combination, we multiply equation 6.3 by two factors, f_j and r_b. f_j is the fraction of QCD events falling in bin N_{jets}, and r_b is the ratio of events with $N_{\text{b-tags}}$ b-tagged jets over all events in each N_{jets} bin.

From QCD simulation we find that f_j and r_b are compatible at high ($\Delta\phi > 0.3$) and low ($\Delta\phi < 0.3$) $\Delta\phi$ values (figures 6.3 and 6.4). Similarly, we find that f_j and r_b do not depend on M_{T2} (Figures 6.5 and 6.6). Figure 6.7 shows that r_b does not depend on H_{T}. Unsurprisingly, Figures 6.3 and 6.4 also show that f_j has a dependency on H_{T} and r_b has a dependency on N_{jets}, respectively. Given these observations, we estimate the QCD yield in each signal region as follows:

$$N_{j,b}^{SR}(M_{\text{T2}}) = N_{\text{inc}}^{CR}(H_{\text{T}}) \cdot r_{\phi}(M_{\text{T2}}) \cdot f_j(H_{\text{T}}) \cdot r_b(N_{\text{jets}}). \tag{6.4}$$

We assign systematic uncertainties on f_j and r_b for the assumption of $\Delta\phi$ and M_{T2} (also H_T for r_b) invariance. The values of the systematics are determined by calculating the RMS of the variations shown in figures 6.3 – 6.7, choosing the maximum from the different H_{T} (N_{jets}) regions for each f_j (r_b). These systematics are summarized in table 6.1.

Observable	f_{23}	f_{46}	f_{7+}	r_0	r_1	r_2	r_{3+}
Syst. Error	25%	7%	20%	8%	20%	35%	70%

Table 6.1: Systematic relative uncertainties for f_j and r_b associated to the assumptions of invariance with respect to $\Delta\phi$, M_{T2} and H_T (H_T only for r_b). The uncertainties on f_j (r_b) are applicable to all H_T (N_j) regions. For the very low H_T regions the uncertainties for f_j are doubled as described in the text.

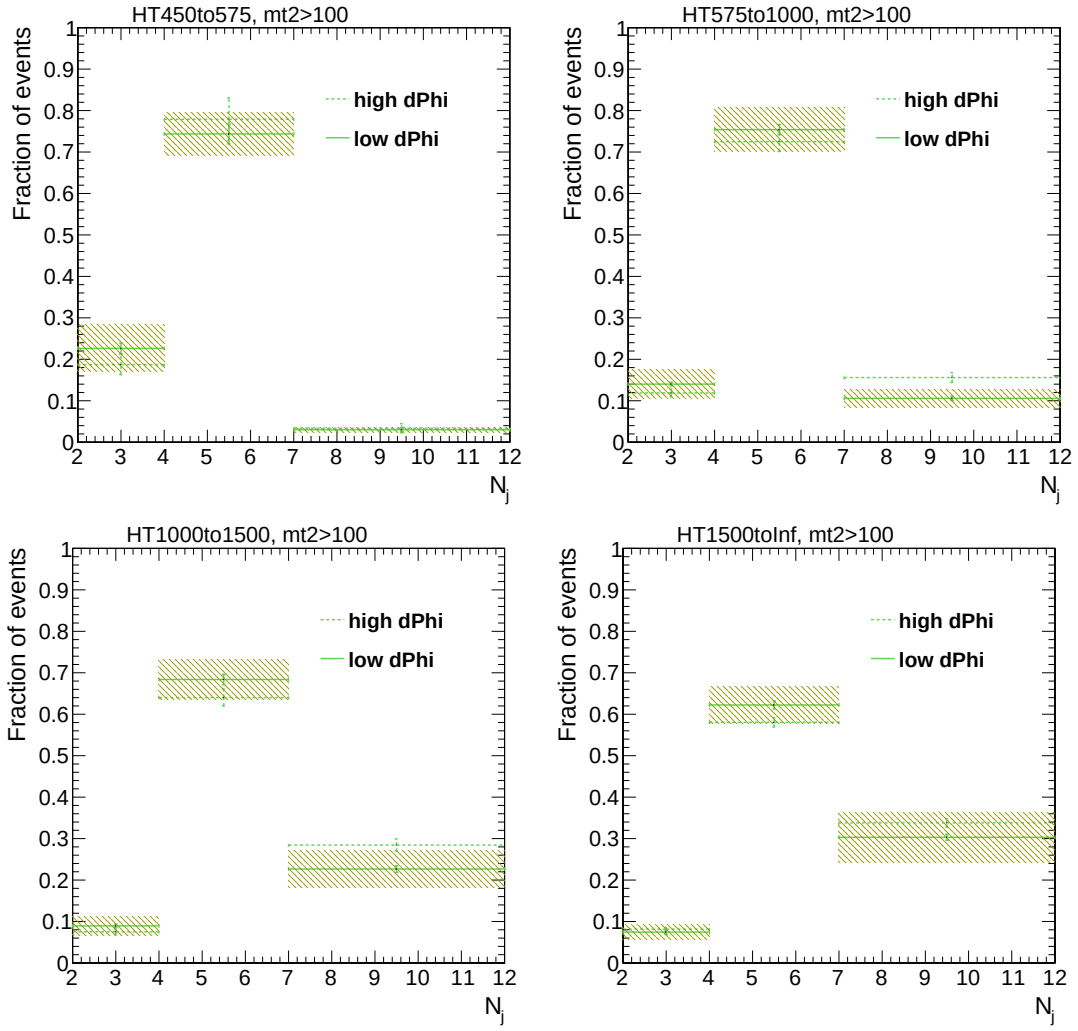

Figure 6.3: Values of f_j from QCD simulation for $\Delta\phi > 0.3$ (dashed lines) and $\Delta\phi < 0.3$ (solid lines) events, in the different H_T regions (indicated on the top of each pad). The uncertainties on the points are MC stats, and the band correspond to the systematic uncertainty tabulated in Table 6.1 and centered in the nominal value (low $\Delta\phi$, $M_{\mathrm{T2}} > 100$ GeV).

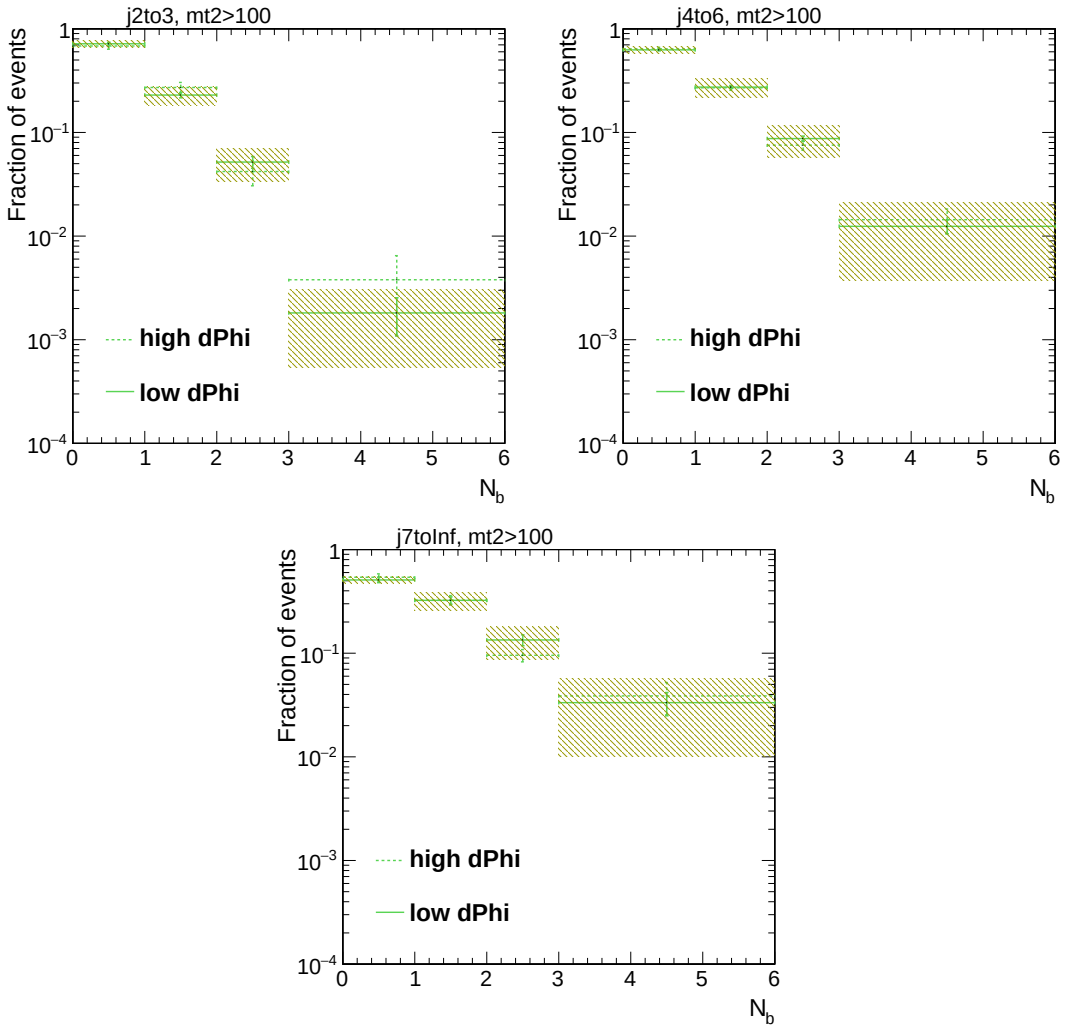

Figure 6.4: Values of r_b from QCD simulation for $\Delta\phi > 0.3$ (dashed lines) and $\Delta\phi < 0.3$ (solid lines) events, in the different N_j regions (indicated on the top of each pad). The uncertainties on the points are MC stats, and the band correspond to the systematic uncertainty tabulated in Table 6.1 and centered in the nominal value (low $\Delta\phi$, $M_{\mathrm{T2}} > 100$ GeV).

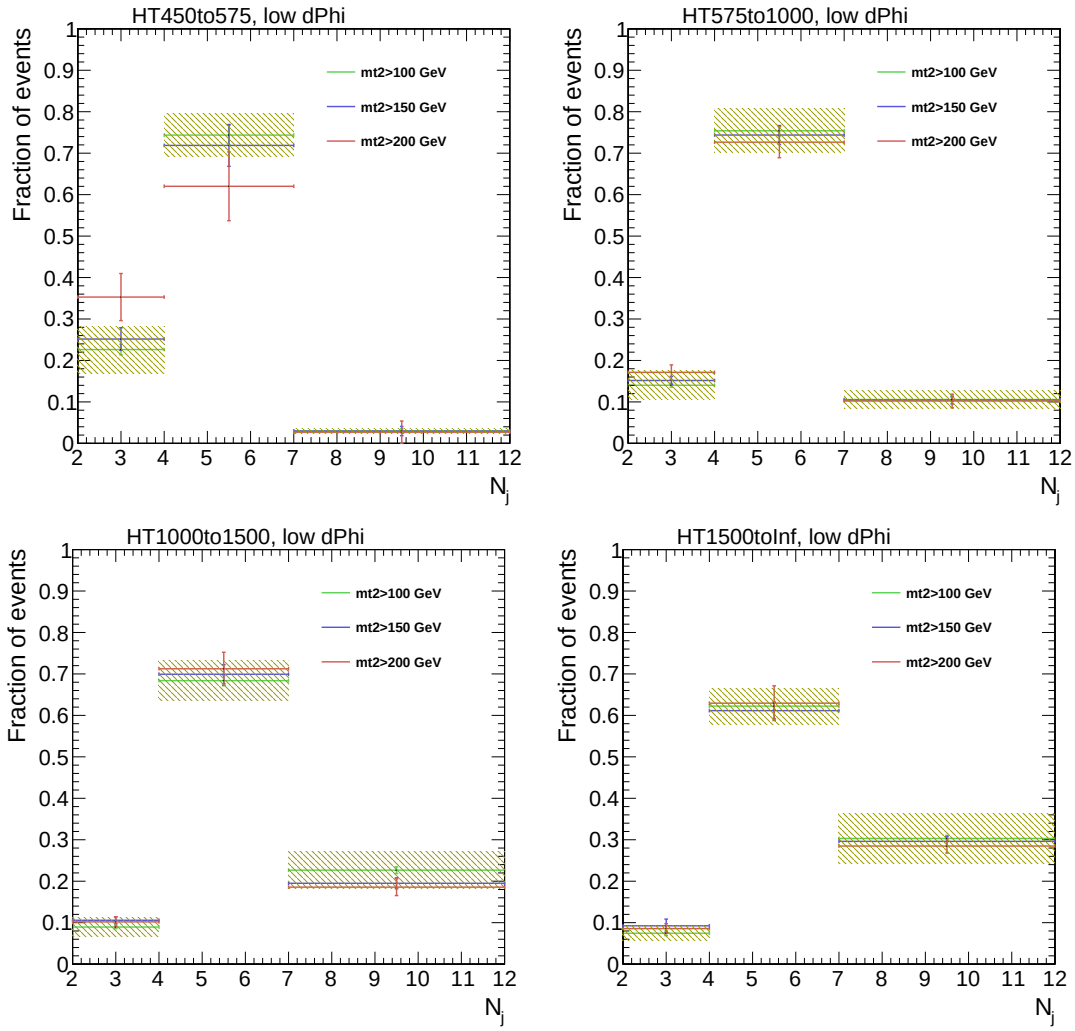

Figure 6.5: Values of f_j from QCD simulation for different M_{T2} thresholds (different colors), in the different H_T regions (indicated on the top of each pad). The uncertainties on the points are MC stats, and the band correspond to the systematic uncertainty tabulated in Table 6.1 and centered in the nominal value (low $\Delta\phi$, $M_{\mathrm{T2}} > 100$ GeV).

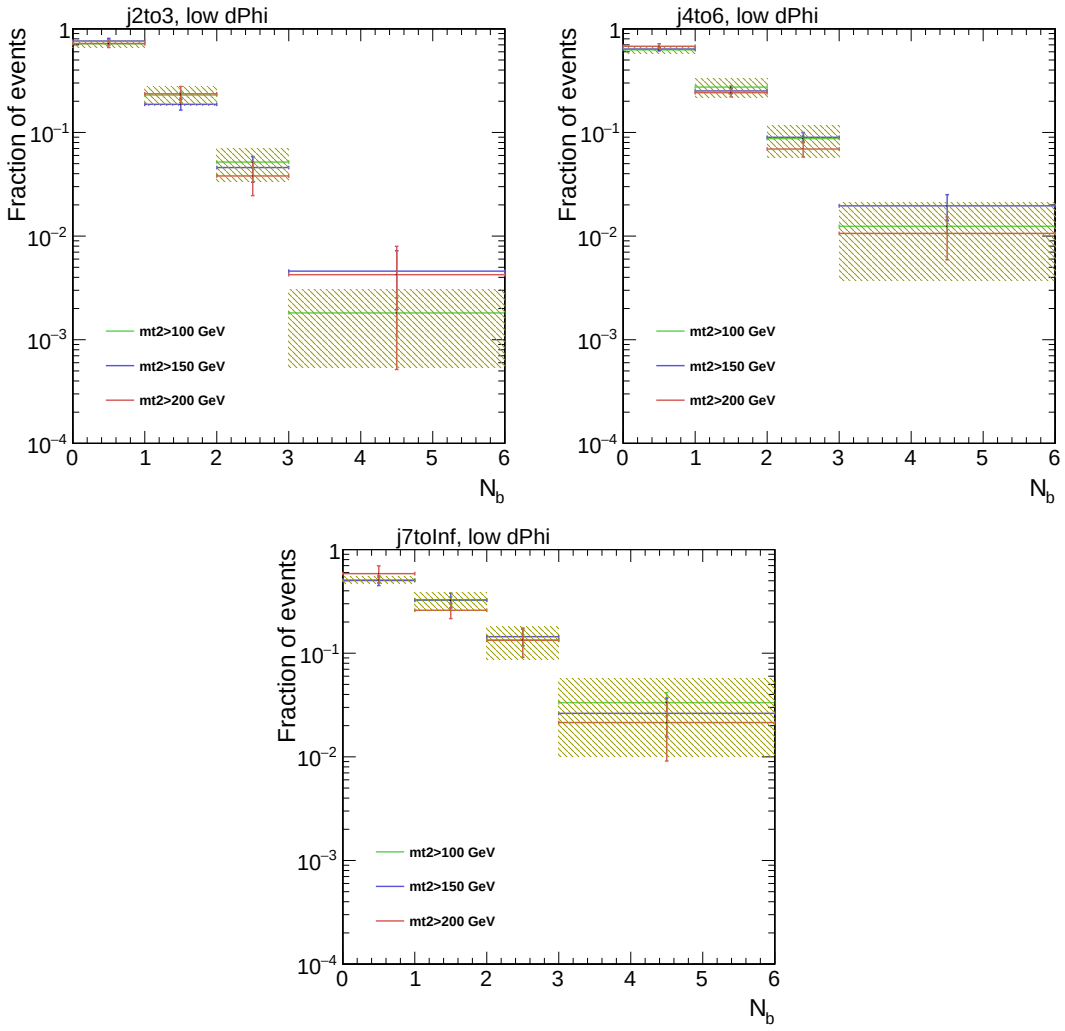

Figure 6.6: Values of r_b from QCD simulation for different M_{T2} thresholds (different colors), in the different N_j regions (indicated on the top of each pad). The uncertainties on the points are MC stats, and the band correspond to the systematic uncertainty tabulated in Table 6.1 and centered in the nominal value (low $\Delta\phi$, $M_{\mathrm{T2}} > 100$ GeV).

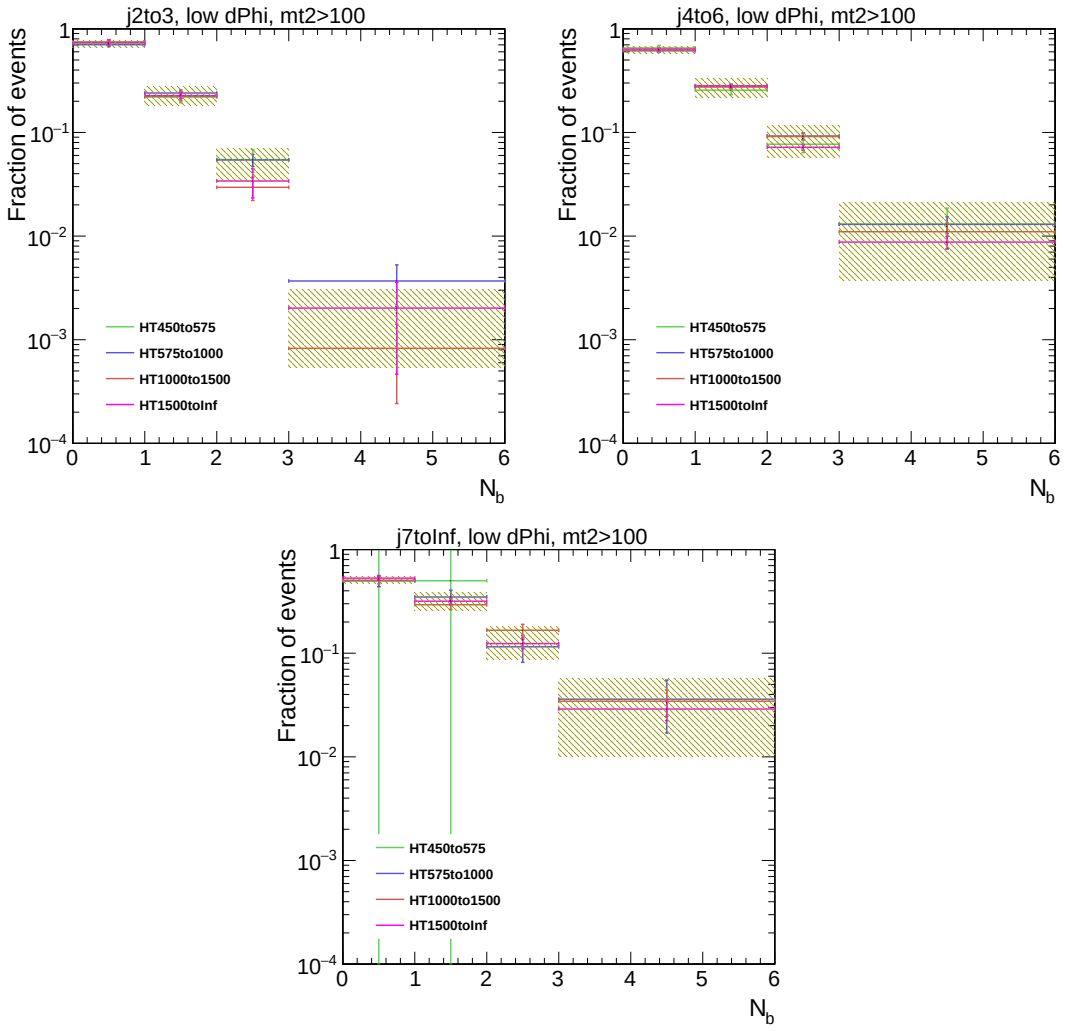

Figure 6.7: Values of r_b from QCD simulation for the different H_T regions (different colors), in the different N_j regions (indicated on the top of each pad). The uncertainties on the points are MC stats, and the band correspond to the systematic uncertainty tabulated in Table 6.1 and centered in the nominal value (low $\Delta\phi$, $M_{\mathrm{T2}} > 100$ GeV).

We measure the values of f_j and r_b directly from data using the QCD-enriched regions of $\Delta\phi < 0.3$ and $100 < M_{\mathrm{T2}} < 200$ GeV, per H_T region and per N_{jets} region respectively for f_j and r_b. Figures 6.8 and 6.9 show these measurements and their comparison to the predictions from QCD MC, where good agreement is found. The measurements of r_b are performed in data using the unprescaled HLT_PFHT800 trigger. For the measurements of f_j in the high and extreme H_{T} regions, the unprescaled HLT_PFHT800 trigger is used while for the medium and low H_{T} regions the prescaled HLT_PFHT475 and HLT_PFHT350 triggers are used. The small contribution (2–9 %) from non-QCD processess is subtracted using MC.

For the very low H_{T} regions, for which there is no existing prescaled trigger covering the H_{T} phase space, f_j is measured using the unprescaled HLT_PFMET90_PFMHT90 trigger, that covers the full H_{T} phase space although at $100 < M_{\mathrm{T2}} < 200$ GeV one lives in the turnon of the trigger. This is fine because we have shown that f_j only depends on H_{T} and not on M_{T2}, so we are not biasing the result, although the unknown amount of non-QCD contribution cannot be subtracted (unless the turnon is modeled). It is expected that the amount of non-QCD events in this region is much smaller that the number of QCD events, so the non-QCD subtraction will make very little difference. In any case the systematic uncertainties for the very low H_{T} regions are conservatively doubled to account for this.

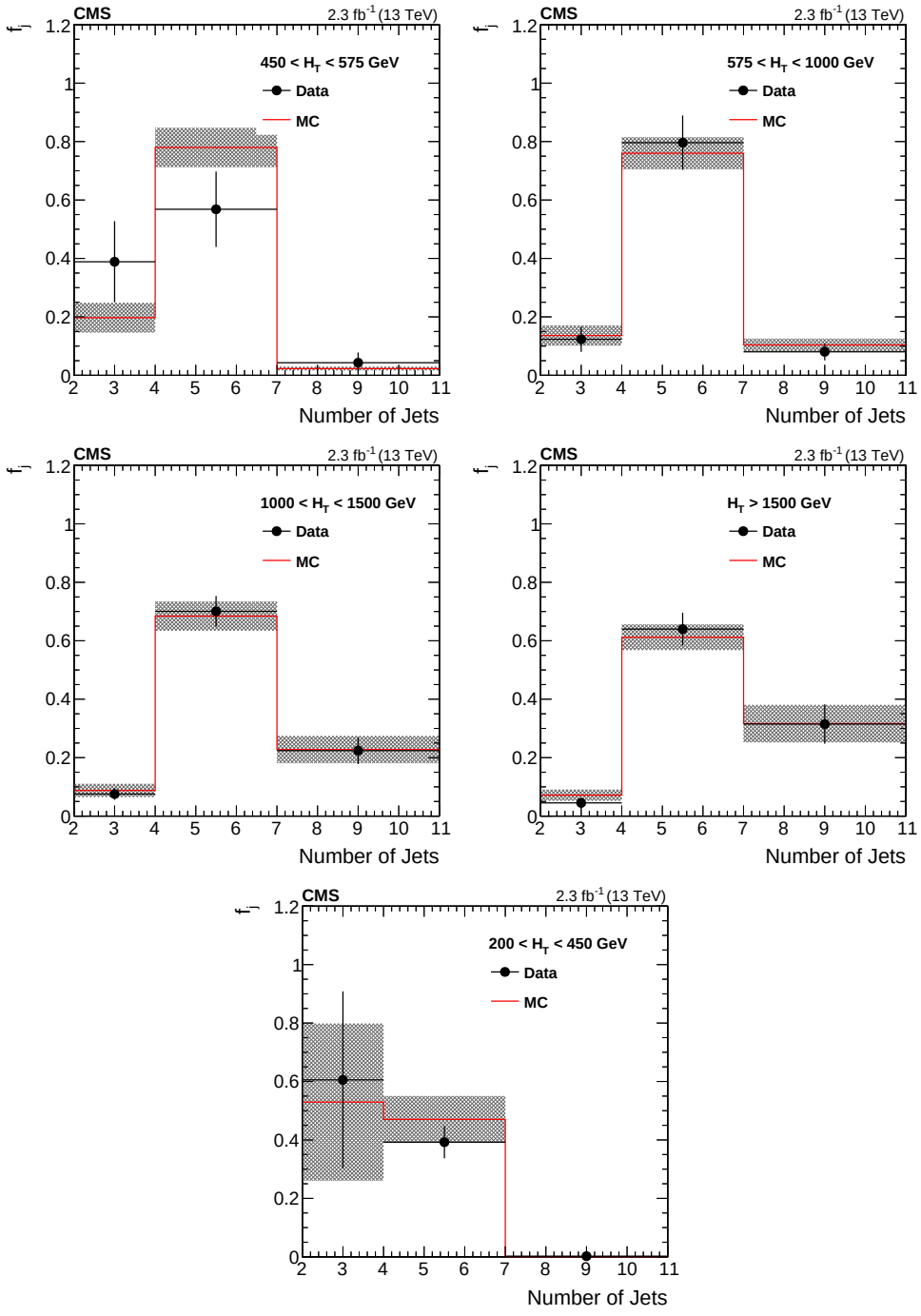

Figure 6.8: Values of f_j measured in data using $\Delta\phi < 0.3$ and $100 < M_{\mathrm{T2}} < 200$ GeV, compared to the values predicted by QCD simulation. The uncertainties include the statistical uncertainy and the systematic uncertainties in table 6.1.

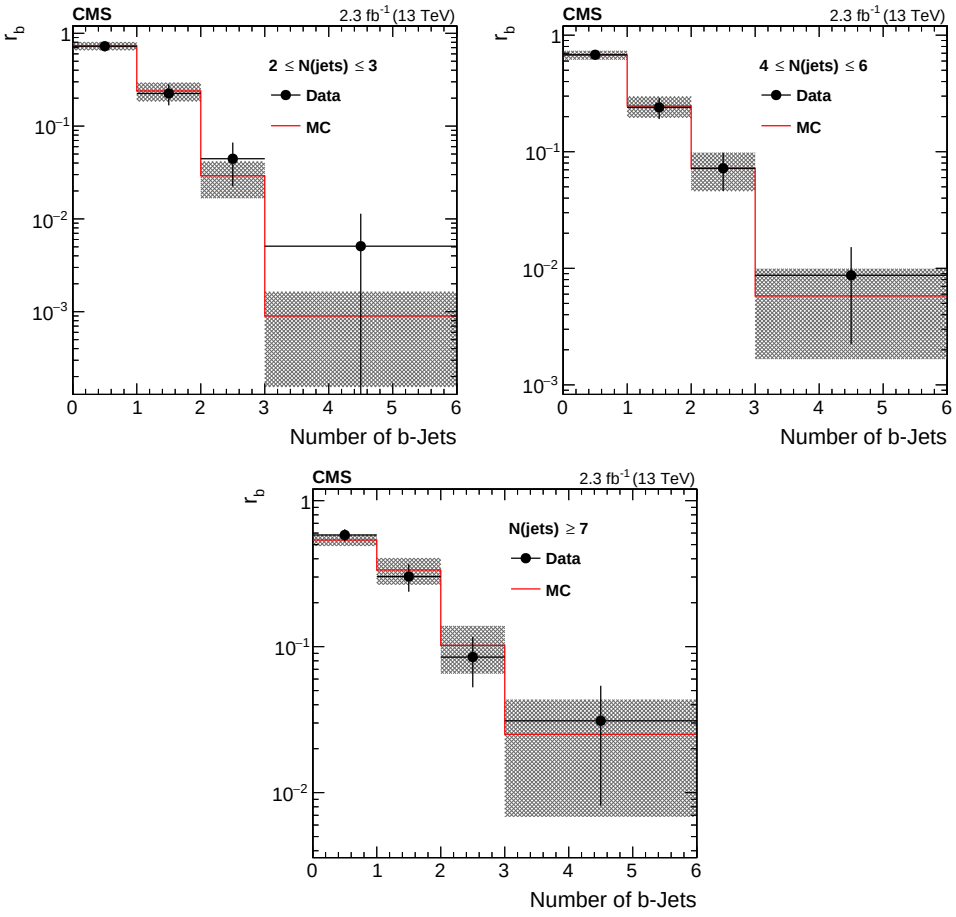

Figure 6.9: Values of $\hat{r}_b$ measured in data using $\Delta\phi < 0.3$ and $100 < M_{\mathrm{T2}} < 200$ GeV, compared to the values predicted by QCD simulation. The uncertainties include the statistical uncertainy and the systematic uncertainties in table 6.1.

Figure 6.10 shows a closure test performed on MC putting all the ingredients together. For the very low H_{T} region MC prediction, the r_ϕ fit parameters have been replaced by the parameters obtained from the low H_{T} fit, as is done for data. In these plots the data-driven QCD estimate performed using MC simulation is compared with the pure MC prediction, in the signal regions ($M_{\mathrm{T2}} > 200$ GeV) and also in validation regions ($100 < M_{\mathrm{T2}} < 200$ GeV) which we use to check the QCD background estimation in data.

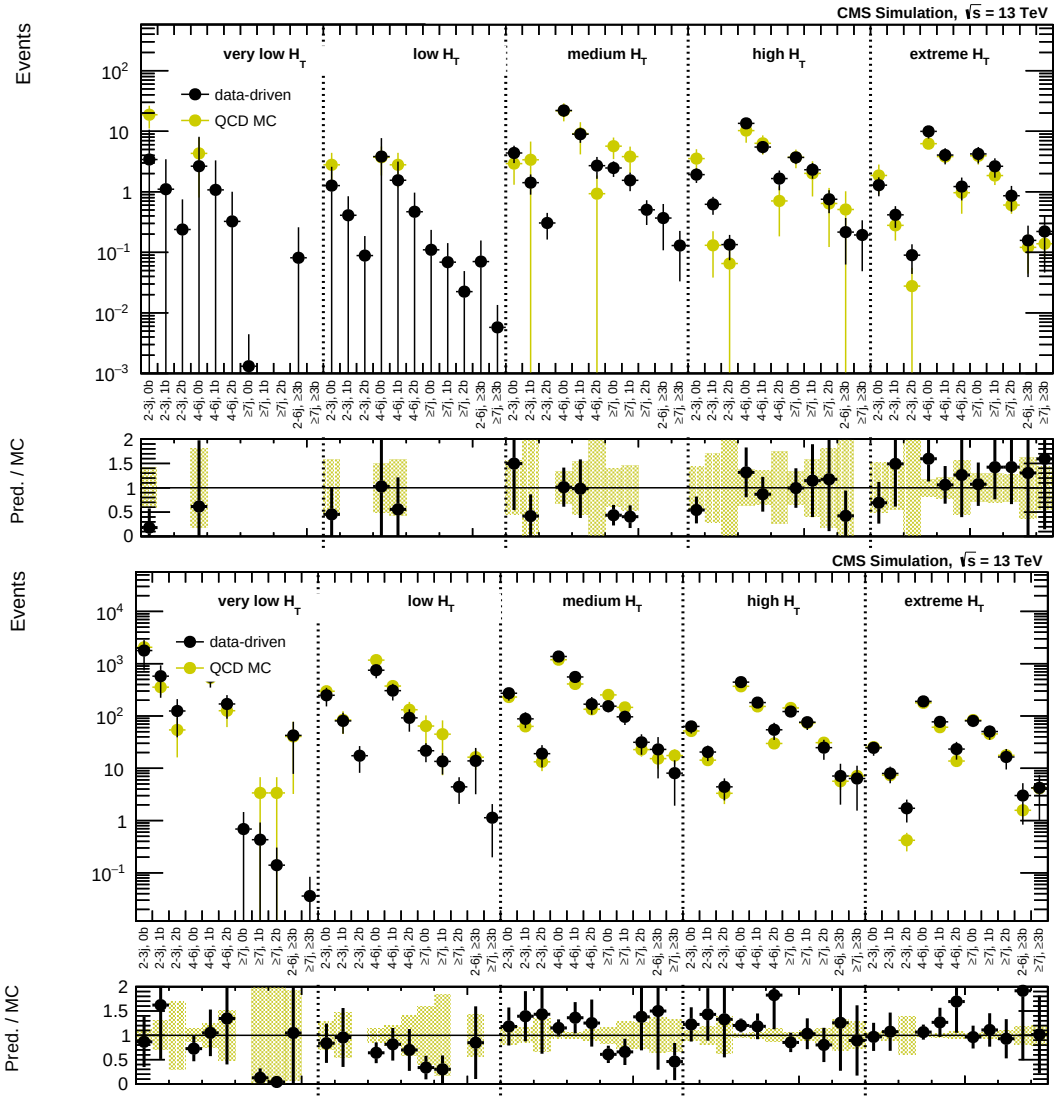

Figure 6.10: MC closure test comparison of the data-driven QCD estimate using MC with the pure MC prediction in the signal region $M_{\mathrm{T2}} > 200$ GeV (top) and in the validation region $100 < M_{\mathrm{T2}} < 200$ GeV (bottom).

Figure 6.11 shows a validation of the method performed in 2.3 fb^{-1} of data in the region $100 < M_{\mathrm{T2}} < 200$ GeV. The high-$\Delta\phi$ data is compared with the background prediction where QCD has been estimated using the data-driven method on 2.3 fb^{-1} data, and the non-QCD contribution is taken from simulation.

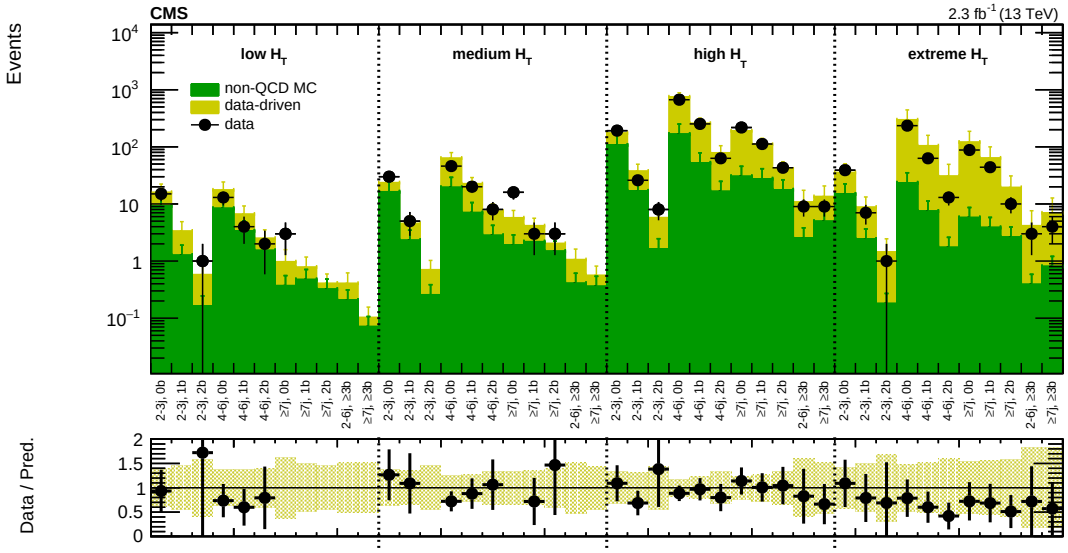

Figure 6.11: Validation of the data-driven QCD estimate in the region $100 < M_{\mathrm{T2}} < 200$ GeV. The points are data with $\Delta\phi > 0.3$, triggered by the H_{T}-only triggers (prescaled for low and medium H_{T}). The green histogram is the non-QCD contribution as expected from simulation, while the yellow histogram is the QCD estimation using the data-driven method on 2.3 fb^{-1} of data.

6.2 Monojet prediction

In monojet events $H_{\mathrm{T}} \approx E_{\mathrm{T}}^{\mathrm{miss}}$, and the missing energy is pointing opposite to the jet. Therefore, the $\Delta\phi$-ratio method cannot be employed. Instead, an alternative approach has been devised, which makes use of unbalanced dijet events, and extrapolates to low subleading jet transverse momentum.

The control region for this method is defined as events with:

- exactly 2 jets
- leading jet $p_{\mathrm{T}} > 200$ GeV
- $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV
- $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}}) < 0.3$

- (no cut on M_{T2})

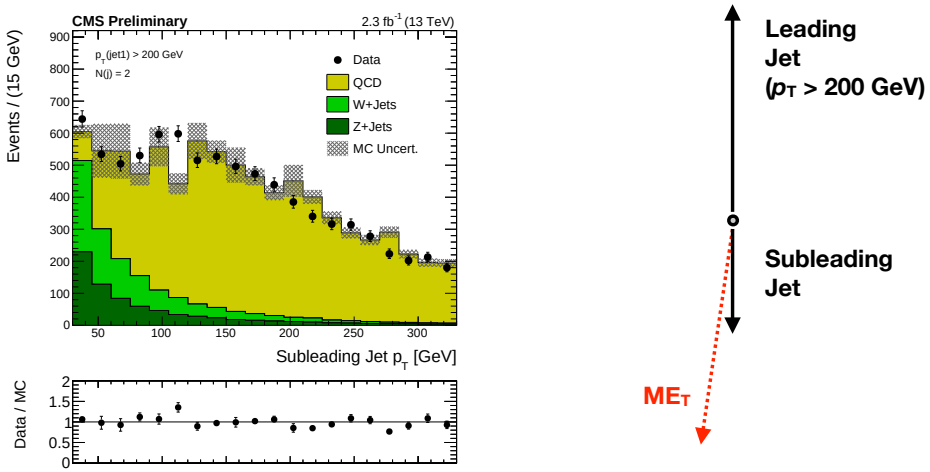

Figure 6.12: Left: Subleading jet transverse momentum for dijet events with $p_{\mathrm{T}}^{\mathrm{jet1}} > 200$ GeV. The simulation is normalized to the data yield. Right: cartoon depicting unbalanced dijet events.

Note that the $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ cut makes this region orthogonal to the multijet signal region, whereas the dijet requirement makes it orthogonal to the monojet signal region. The transverse momentum of the subleading jet for these events is shown in Figure 6.12 (left), where the simulation is normalized to the data yield. As can be seen the plot has a lower bound at 30 GeV, as that is the jet counting threshold: events with subleading jet with $0 < p_{\mathrm{T}}^{\mathrm{jet2}} < 30$ GeV will be classified as monojet events. As $p_{\mathrm{T}}^{\mathrm{jet2}}$ gets smaller, these events become more unbalanced, with a hard leading jet and a soft subleading jet, with the $E_{\mathrm{T}}^{\mathrm{miss}}$ pointing close to the direction of the subleading jet, as shown schematically in Figure 6.12 (right).

We use events with $30 < p_{\mathrm{T}}^{\mathrm{jet2}} < 60$ GeV to put an upper limit on the amount of QCD in the $0 < p_{\mathrm{T}}^{\mathrm{jet2}} < 30$ GeV range (the monojet signal regions). This is done using the fact that the number of events with $0 < p_{\mathrm{T}}^{\mathrm{jet2}} < 30$ GeV is expected to be smaller than the number of events with $30 < p_{\mathrm{T}}^{\mathrm{jet2}} < 60$ GeV as the former requires larger larger jet

mismeasurement than the latter. The QCD estimate $N_{QCD}(p_{\mathrm{T}}^{\mathrm{jet1}})$, for a given monojet $p_{\mathrm{T}}^{\mathrm{jet1}}$ bin, is obtained as:

$$N_{QCD}(p_{\mathrm{T}}^{\mathrm{jet1}}) = N_{data}(30-60, p_{\mathrm{T}}^{\mathrm{jet1}}) \cdot f_{QCD}(30-60, p_{\mathrm{T}}^{\mathrm{jet1}})$$

where $N_{data}(30-60, p_{\mathrm{T}}^{\mathrm{jet1}})$ is the number of data dijet events with with $30 < p_{\mathrm{T}}^{\mathrm{jet2}} < 60$ GeV and leading jet in a given $p_{\mathrm{T}}^{\mathrm{jet1}}$ bin, and $f_{QCD}(30-60, p_{\mathrm{T}}^{\mathrm{jet1}})$ is the fraction of events in simulation passing the same selection that come from QCD (in order to remove the contribution of electroweak processes). We assign a 50% uncertainty on $f_{QCD}(30-60, p_{\mathrm{T}}^{\mathrm{jet1}})$.

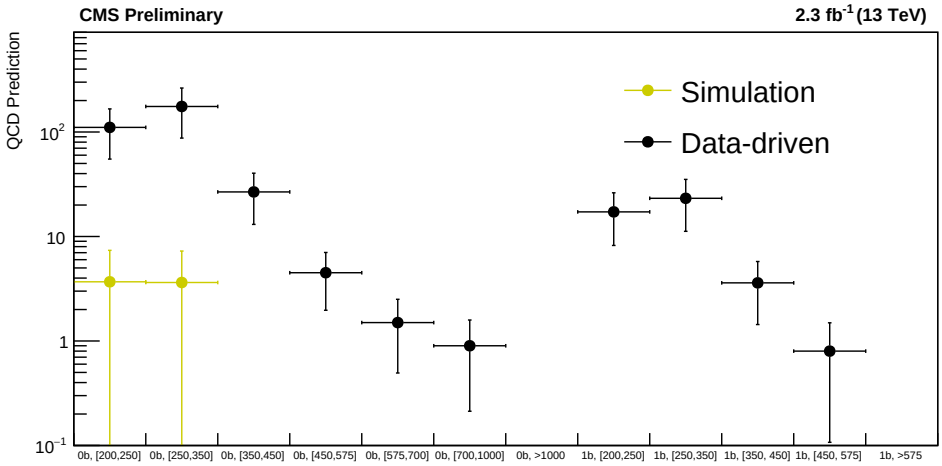

Figure 6.13: MC-closure of the QCD monojet background estimation: the estimation computed on the simulation (black markers) is compared to the MC-truth QCD background in each monojet search bin.

The closure of this method is shown in Figure 6.13, where the results of the data-driven (black markers) is compared to the MC-truth QCD background (yellow markers) in each monojet search bin. As can be seen the estimate seems to provide a reasonable upper bound for this background process, hence validating the method.

Chapter 7

Rebalance and Smear Method

The rebalance and smear (R&S) method provides a cross check of the $\Delta\phi$-ratio method described in Chapter 6 that is used to estimate the qcd multijet background. The predictions from the R&S method serve only to strengthen our confidence in the predictions from the $\Delta\phi$-ratio method, and are not used outside of this comparison.

The R&S estimation is performed in two distinct steps. In the first step, multijet events are "rebalanced" by adjusting the jet p_{T}'s such that the resulting $E_{\mathrm{T}}^{\mathrm{miss}}$ is close to zero. The rebalancing is done through a likelihood maximization that takes into account the jet resolution. The result of the rebalancing step is an inclusive sample of "true" QCD events that is used as seed events for the second step, the smearing. In the smearing step, the p_{T} of the jets in each rebalanced seed event are smeared according to the jet response in order to simulate the instrumental effects that lead to nonzero $E_{\mathrm{T}}^{\mathrm{miss}}$. The smearing can be done many times for each rebalanced event in order to accumulate statistics in the tails of kinematic distributions such as $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{T2}.

7.1 Event selections

For the rebalance and smear prediction in data, the input data sample comes from the unprescaled HLT_PFHT800 signal trigger as well as several prescaled pure H_{T} triggers with online H_{T} requirements as low as 200 GeV. In order to use an event from one of these triggers, the offline H_{T} must be at least 100 GeV greater than the online H_{T} threshold, in the efficiency plateau. If an event passes more than one of these triggers then the lowest prescale value is used. Due to lack of appropriate triggers, the R&S estimate cannot be done for the very low H_{T} region, and is therefore only used to predict the multijet background in signal regions with $H_{\mathrm{T}} > 450$ GeV. In addition to the trigger selections, events must contain at least one good vertex, two jets with $p_{\mathrm{T}} > 10$ GeV and pass the standard event cleaning filters in order to be used in the rebalancing step. No other selections are applied.

7.2 Jet response templates

A key ingredient in the rebalance and smear estimation is the jet response templates. These templates are distributions of $p_{T,jet}^{\mathrm{reco}}/p_{T,jet}^{\mathrm{gen}}$ which are derived in MadGraph QCD Monte Carlo. The templates are binned in gen level jet p_{T} and η with the following bin edges:

PtBinEdges = 0, 20, 30, 50, 80, 120, 170, 230, 300, 380, 470, 570, 680, 800, 1000, 1300, 1700, 2200, 2800, 3500, 4300, 5200, 6500 GeV

EtaBinEdges = 0, 0.3, 0.5, 0.8, 1.1, 1.4, 1.7, 2.3, 2.8, 3.2, 4.1, 5.0

There are separate templates derived for b-tagged jets, as decays of b-jets often contain larger real $E_{\mathrm{T}}^{\mathrm{miss}}$ from neutrinos than light flavor decays. For medium b-tagged jets the b-jet specific templates are used. Templates that are inclusive in jet flavor are used

for non b-tagged jets.

7.3 Rebalancing procedure

The rebalancing procedure adjusts the p_{T} of jets in an event with the aim of reproducing the true hard scatter event which has no $E_{\mathrm{T}}^{\mathrm{miss}}$. Note that only the jet p_{T} is modified while the jet direction remains unchanged. All jets above 10 GeV are used in the rebalancing, except those that both fail the pileup jet ID and have $p_{\mathrm{T}} < 100$ GeV. The pileup jet ID is used in an attempt to avoid rebalancing jets from the hard scatter against pileup jets. Jets not considered in the rebalancing are added back to the rebalanced event in order to retain effects due to pileup.

In rebalancing an event with n reco jets that satisfy the requirements above, we attempt to find the most likely configuration of n "true" jets that satisfy the rebalancing condition

$$\sum_{i=1}^{n} \vec{p}_{T,i}^{\,\mathrm{true}} + \vec{p}_{T,\mathrm{soft}}^{\,\mathrm{true}} = 0 \tag{7.1}$$

where $\vec{p}_{T,\mathrm{soft}}^{\,\mathrm{true}}$ is the transverse momentum in the true event due to jets below the 10 GeV threshold and unclustered energy that is needed to balance the event.

In order to be able to find a solution to Eq. 7.1 we use $\vec{p}_{T,\mathrm{soft}}^{\,\mathrm{reco}}$ as an estimate for $\vec{p}_{T,\mathrm{soft}}^{\,\mathrm{true}}$. Here $\vec{p}_{T,\mathrm{soft}}^{\,\mathrm{reco}}$ is the imbalance in the reco event due to jets below the 10 GeV threshold and unclustered energy. Enforcing this balancing condition exactly may not always result in the best representation of the true event since the angle and magnitude of the measured $\vec{p}_{T,\mathrm{soft}}^{\,\mathrm{reco}}$ can be affected by soft pileup and mismeasurement of soft activity from the hard scatter. Therefore we loosen the rebalancing condition to

$$\sum_{i=1}^{n} \vec{p}_{T,i}^{\,\mathrm{true}} + \vec{p}_{T,\mathrm{soft}}^{\,\mathrm{reco}} \approx 0. \tag{7.2}$$

How the loosening of this condition is done is made more precise in the description of the likelihood maximization below.

To rebalance an event with n reco jets and find the most likely configuration of n true jets we maximize the likelihood L, where

$$L = \prod_{i=1}^{n} \mathrm{Prob}(p_{T,i}^{\mathrm{true}} | p_{T,i}^{\mathrm{reco}}) \times G(\frac{p_{x,\mathrm{soft}}^{\mathrm{reco}} - p_{x,\mathrm{soft}}^{\mathrm{true}}}{\sigma_T^{\mathrm{soft}}}) \times G(\frac{p_{y,\mathrm{soft}}^{\mathrm{reco}} - p_{y,\mathrm{soft}}^{\mathrm{true}}}{\sigma_T^{\mathrm{soft}}}) \tag{7.3}$$

with

$$G(x) \equiv exp(-x^2/2). \tag{7.4}$$

The term $\mathrm{Prob}(p_{T,i}^{\mathrm{true}} | p_{T,i}^{\mathrm{reco}})$ in equation 7.3 is the probability that the true p_{T} of jet$_i$ is equal to $p_{T,i}^{\mathrm{true}}$ when the measured p_{T} is $p_{T,i}^{\mathrm{reco}}$. This probability is taken from the gaussian core of the jet response templates. The two G(x) terms in equation 7.3 enforce the approximate balancing condition (equation 7.2). The definition of the soft terms in equation 7.3 are given in equations 7.5 and 7.6. The balancing in the x and y directions is done to within some width σ_T^{soft}. A value of 20 GeV is chosen for σ_T^{soft} which is approximately the width of the x and y components of $E_{\mathrm{T}}^{\mathrm{miss}}$ from MinBias events.

$$\vec{p}_{T,\mathrm{soft}}^{\mathrm{reco}} \equiv -\vec{E}_{\mathrm{T}}^{\mathrm{miss}} - \sum_{\text{jets pt>10 GeV}} \vec{p}_{T,i}^{\mathrm{reco}} \tag{7.5}$$

$$\vec{p}_{T,\mathrm{soft}}^{\mathrm{true}} \equiv -\sum_{i=1}^{n} \vec{p}_{T,i}^{\mathrm{true}} \tag{7.6}$$

In practice the likelihood maximization is done by minimizing -log(L) using minuit [27]. The minimization is done by finding the n parameters c_1, ..., c_n such that $p_{T,i}^{\mathrm{true}} \equiv \frac{1}{c_i} p_{T,i}^{\mathrm{reco}}$ minimize the -logL. To calculate $\mathrm{Prob}(p_{T,i}^{\mathrm{true}} | p_{T,i}^{\mathrm{reco}})$ we look at the response template for jets with $p_{\mathrm{T}} = \frac{1}{c_i} p_{T,i}^{\mathrm{reco}}$ and find the probability of c_i. The rebalanced event will have jets with p_{T} scaled by $\frac{1}{c_i}$ for the corresponding c_i.

7.4 Smearing procedure

Once a sample of rebalanced events has been obtained, the next step is to smear the jets in these events many times. Each rebalanced event is smeared (100×prescale) number of times. For each smearing, the p_{T} of each jet in the rebalanced event is scaled by a random factor drawn from the corresponding jet response template. If an event contains jets that were not considered in the rebalancing procedure (i.e. they failed the pileup jet ID) then those jets remain in the event without any smearing. After the smearing has been done, all jet-related quantities are recalculated and the analysis selections are applied. Histograms are filled for each smeared event that passes the analysis selections with a weight of 0.01 events. The rebalance and smear predictions for kinematic distributions and event yields are taken from these histograms.

7.5 Performance in Monte Carlo

Figures 7.1 - 7.4 show kinematic distributions from Monte Carlo and R&S based on Monte Carlo after a very loose selection of $H_{\mathrm{T}} > 1000$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV. Figures 7.1 - 7.4 show the same distributions after a selection of 450 GeV $< H_{\mathrm{T}} <$ 1000 GeV GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV. The rebalance and smear method models the shapes of these distributions quite well. There is an overall normalization difference of about 10% introduced by the $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV cut due to differences in the modeling of very low $E_{\mathrm{T}}^{\mathrm{miss}}$ events. Figures 7.9 shows Monte Carlo closure in the topological regions after the baseline selection.

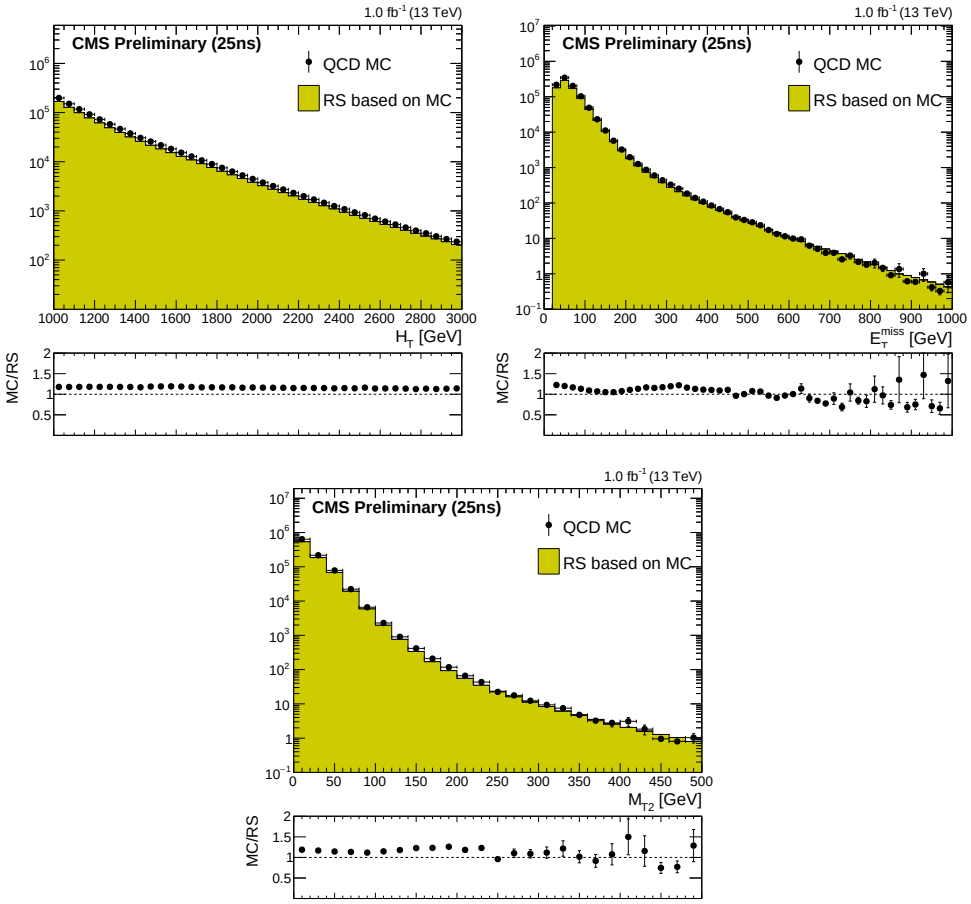

Figure 7.1: H_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$, and M_{T2} distributions for Monte Carlo and R&S based on MC. The selection is $H_{\mathrm{T}} > 1000$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV.

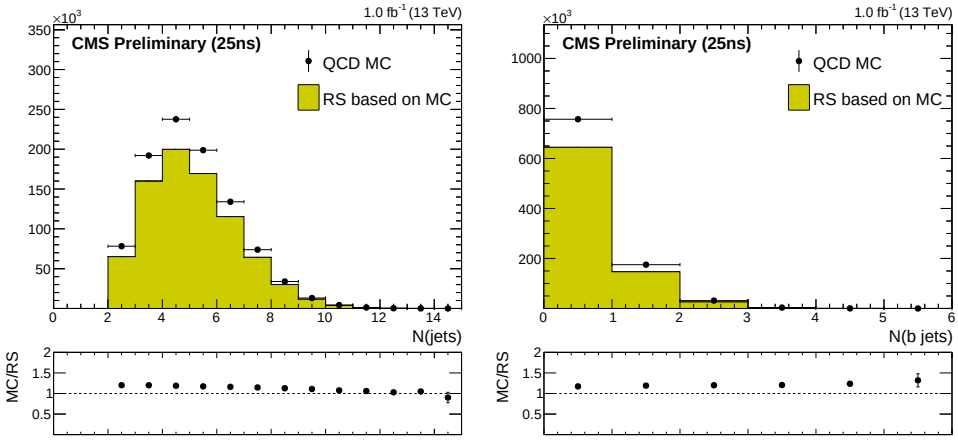

Figure 7.2: N_{jets} and $N_{\text{b}-\text{tags}}$ distributions for Monte Carlo and R&S based on MC. The selection is $H_{\text{T}} > 1000$ GeV and $E_{\text{T}}^{\text{miss}} > 30$ GeV.

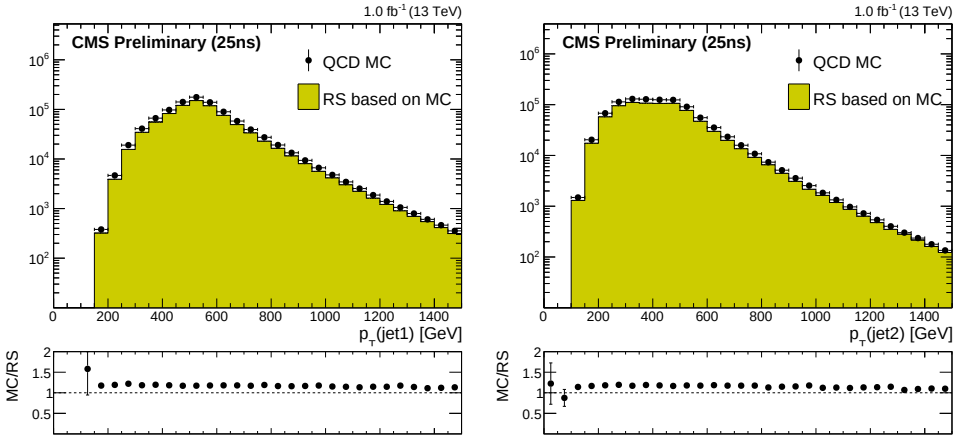

Figure 7.3: Leading and subleading jet p_{T} distributions for Monte Carlo and R&S based on MC. The selection is $H_{\text{T}} > 1000$ GeV and $E_{\text{T}}^{\text{miss}} > 30$ GeV.

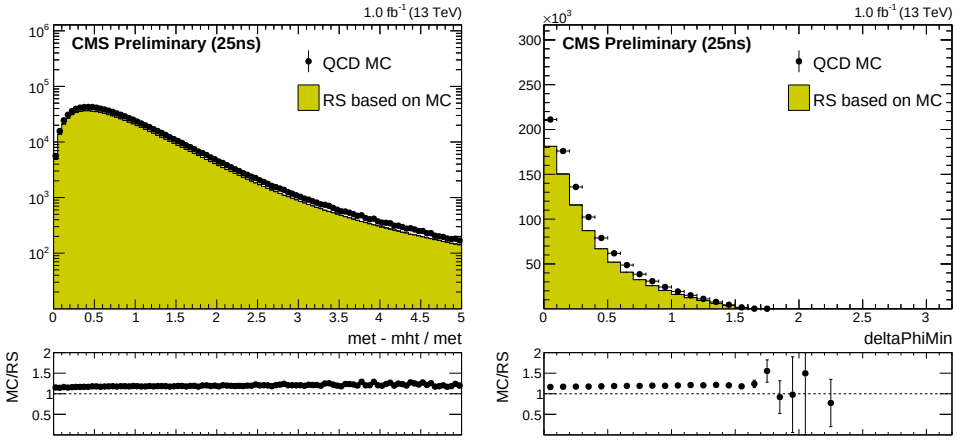

Figure 7.4: $|\vec{H}_{\mathrm{T}}^{\mathrm{miss}} - \vec{E}_{\mathrm{T}}^{\mathrm{miss}}|/E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ distributions for Monte Carlo and R&S based on MC. The selection is $H_{\mathrm{T}} > 1000$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV.

7.6 Electroweak contamination

The input to the rebalancing step in data comes from pure H_{T} triggers with no attempt to remove any possible contamination from non-QCD processes. Most electroweak events are rebalanced to $E_{\mathrm{T}}^{\mathrm{miss}}$ close to zero just like actual QCD events and contribute an extremely small amount to the final prediction since the cross section for electroweak processes is much smaller than the QCD cross section. However some configurations of electroweak events prove difficult to rebalance, such as events with $E_{\mathrm{T}}^{\mathrm{miss}}$ in one hemisphere and all jets in the other hemisphere. An example of one such Monte Carlo event is shown in Figure 7.10 The $E_{\mathrm{T}}^{\mathrm{miss}}$ in these events is reduced in the rebalancing step but can still be rather large. When the $E_{\mathrm{T}}^{\mathrm{miss}}$ after rebalancing is large basically every smeared event will also have large $E_{\mathrm{T}}^{\mathrm{miss}}$ and will therefore contribute to the final prediction much more than if the smeared $E_{\mathrm{T}}^{\mathrm{miss}}$ was actually a product of sampling the tails of the jet response templates.

In order to remove contamination to the R&S prediction from electroweak events that are difficult to rebalance we require the $E_{\mathrm{T}}^{\mathrm{miss}}$ after rebalancing to be less than 100 GeV.

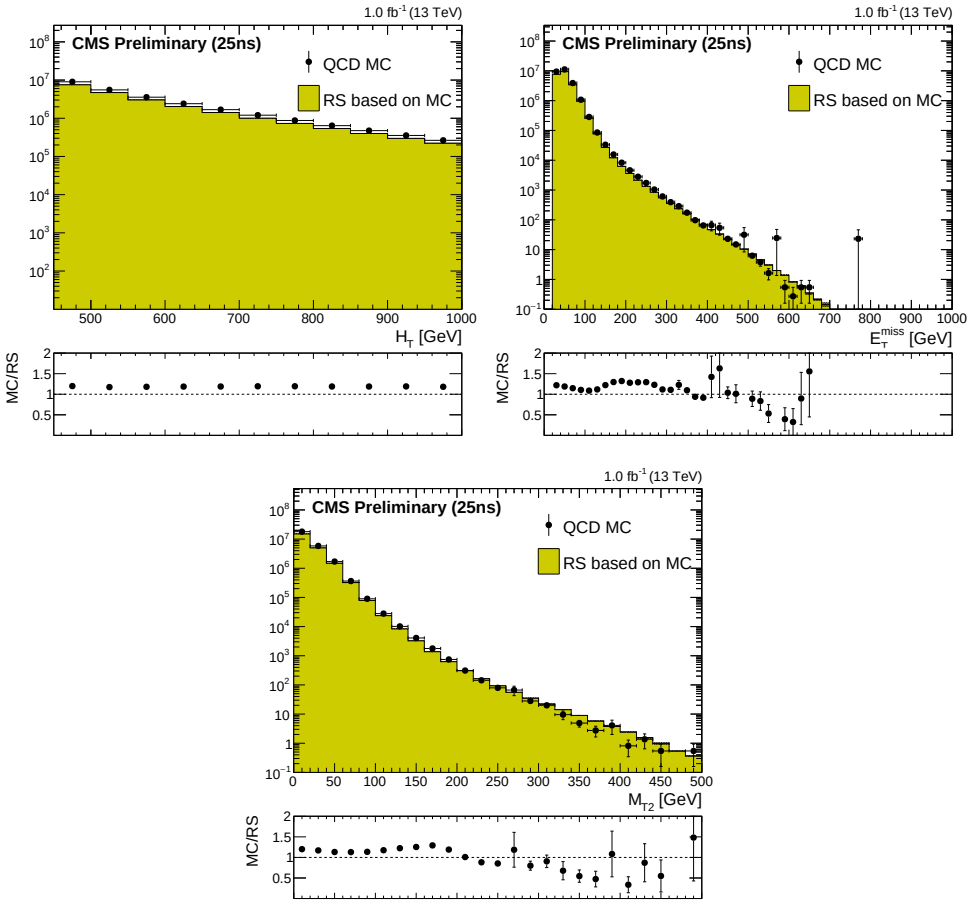

Figure 7.5: H_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$, and M_{T2} distributions for Monte Carlo and R&S based on MC. The selection is 450 GeV $< H_{\mathrm{T}} <$ 1000 GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV.

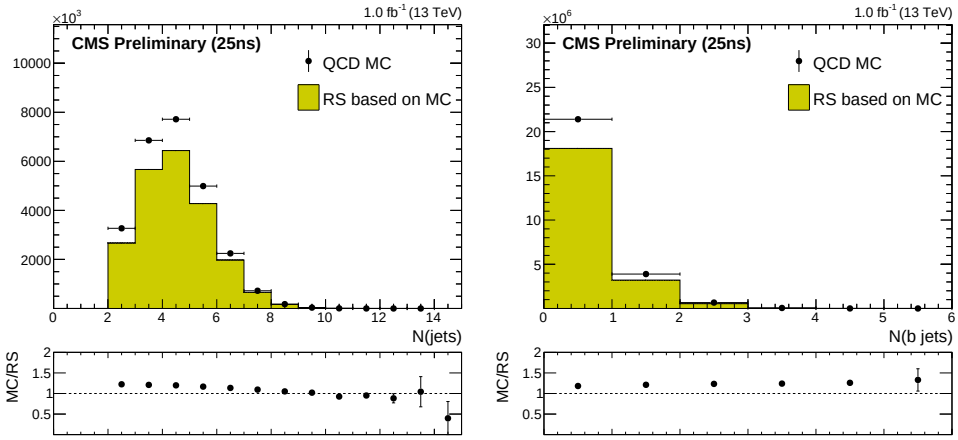

Figure 7.6: N_{jets} and $N_{\text{b-tags}}$ distributions for Monte Carlo and R&S based on MC. The selection is 450 GeV $< H_{\text{T}} <$ 1000 GeV and $E_{\text{T}}^{\text{miss}} > 30$ GeV.

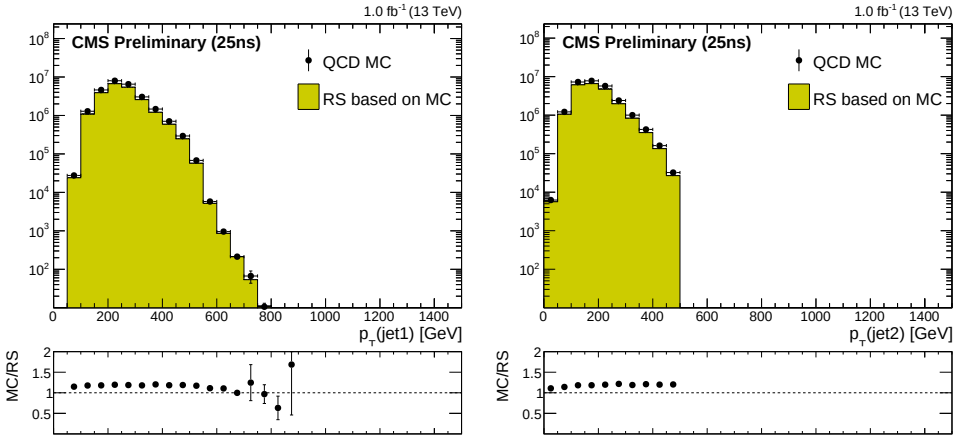

Figure 7.7: Leading and subleading jet p_{T} distributions for Monte Carlo and R&S based on MC. The selection is 450 GeV $< H_{\text{T}} <$ 1000 GeV and $E_{\text{T}}^{\text{miss}} > 30$ GeV.

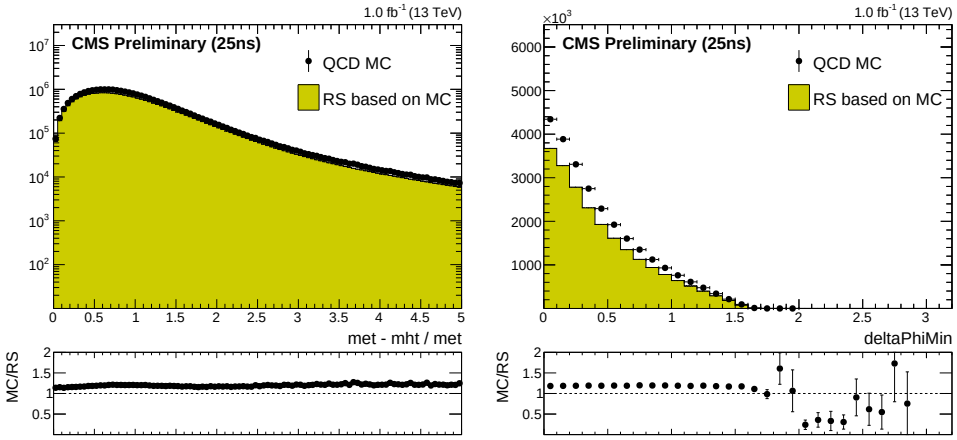

Figure 7.8: $|\vec{H}_{\mathrm{T}}^{\mathrm{miss}} - \vec{E}_{\mathrm{T}}^{\mathrm{miss}}|/E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ distributions for Monte Carlo and R&S based on MC. The selection is 450 GeV $< H_{\mathrm{T}} <$ 1000 GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV.

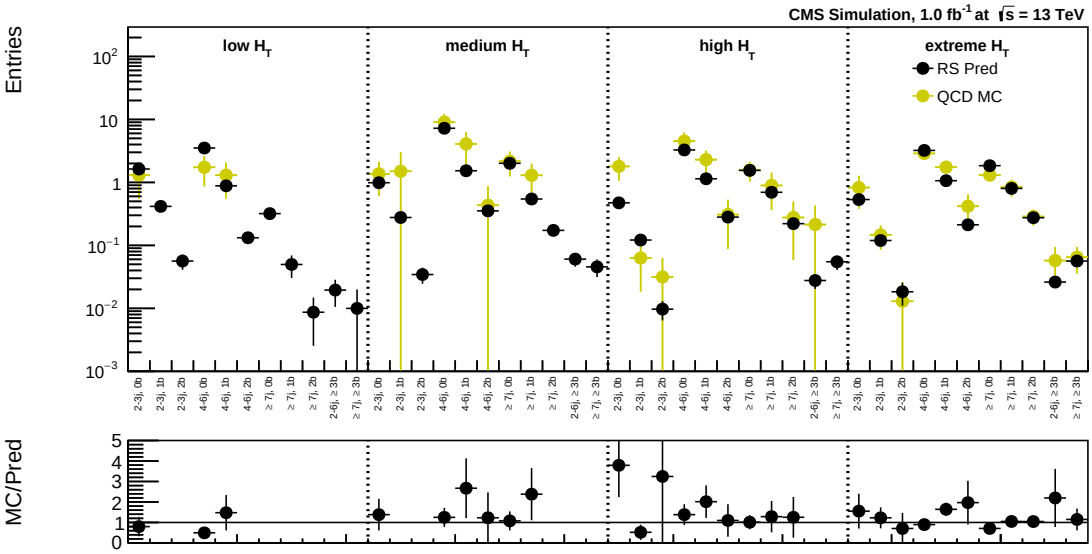

Figure 7.9: R&S Monte Carlo closure in topological regions after the baseline selection. The bottom histogram shows the ratio of yields in each region.

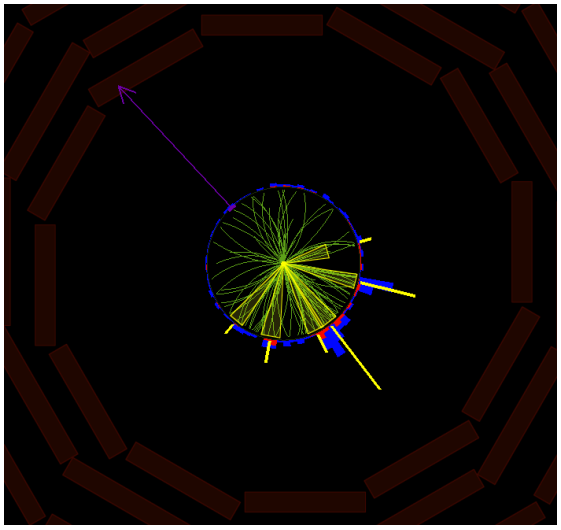

Figure 7.10: Example of an Invisible Z event in Monte Carlo with a configuration that leads to large $E_{\mathrm{T}}^{\mathrm{miss}}$ after rebalancing. This event has 517 GeV of $E_{\mathrm{T}}^{\mathrm{miss}}$ before rebalancing and 311 GeV of $E_{\mathrm{T}}^{\mathrm{miss}}$ after rebalancing.

Figures 7.11 and 7.12 show the rebalanced $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for smeared QCD and electroweak MC events that enter the signal regions. The electroweak events are the sum of events from $\mathrm{Z} \to \nu\nu$, $\mathrm{W} + \mathrm{jets}$, and ttbar Monte Carlo. From these distributions we can determine the effect of requiring $E_{\mathrm{T}}^{\mathrm{miss}} < 100$ GeV after rebalancing. We compare the QCD yield integrated over rebalanced $E_{\mathrm{T}}^{\mathrm{miss}}$ to the sum of the QCD and EWK yields with rebalanced $E_{\mathrm{T}}^{\mathrm{miss}} < 100$ GeV and take a scale factor to correct for the difference. The uncertainty on this scale factor is chosen to be the difference from 1.0.

Table 7.1: Scale factors to correct for loss of QCD events due to rebalanced $E_{\mathrm{T}}^{\mathrm{miss}} < 100$ GeV requirement.

	low H_{T}	med H_{T}	high H_{T}	ext H_{T}
QCD total yield	7.07	12.45	7.40	7.63
QCD + EWK reb $E_{\mathrm{T}}^{\mathrm{miss}} < 100$ GeV	6.89	11.63	7.38	7.63
Correction Factor	1.03 ± 0.03	1.07 ± 0.07	1.0	1.0

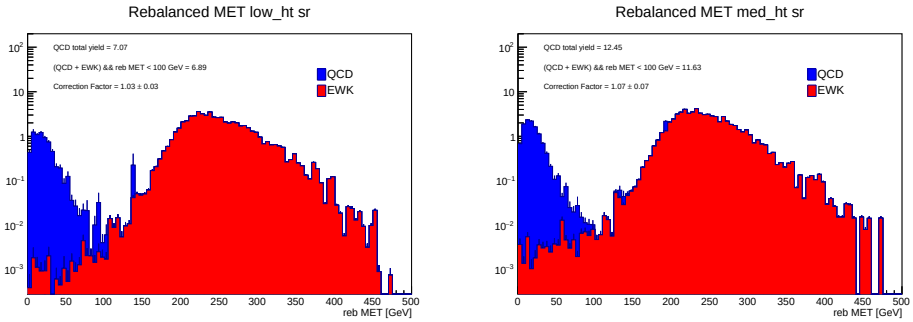

Figure 7.11: Rebalanced $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for QCD and EWK smeared events in the low HT and med H_{T} regions after the baseline selection.

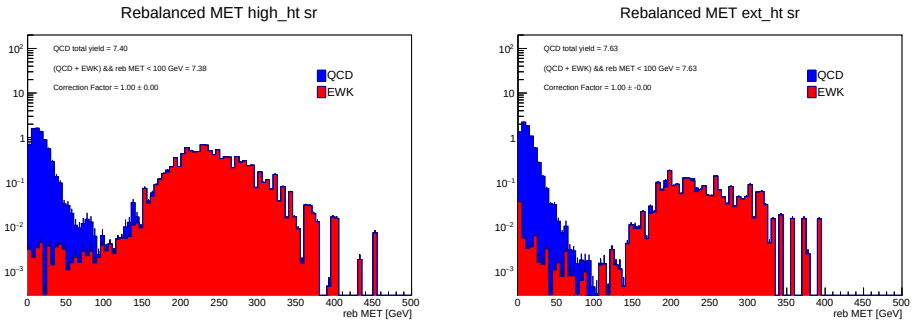

Figure 7.12: Rebalanced $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for QCD and EWK smeared events in the high HT and ext H_{T} regions after the baseline selection.

7.7 Performance in data control regions

In order to gauge the performance of the rebalance and smear method in data we define three control regions that are orthogonal to the search regions and enriched in QCD. The first control region is obtained from the baseline selection by inverting the $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ cut, requiring $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}}) < 0.3$. The second control region is the M_{T2} sideband 100 GeV $< M_{\mathrm{T2}} <$ 200 GeV. The third control region is defined by both inverting the $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ selection and selecting the M_{T2} sideband. Figures 7.13 - 7.17 show several kinematic distributions in the control regions for 450 GeV $< H_{\mathrm{T}} <$ 1000 GeV and $H_{\mathrm{T}} >$ 1000 GeV separately. Non-QCD background contributions in these control regions are taken from Monte Carlo. Selecting the M_{T2} sideband for 450 GeV $< H_{\mathrm{T}} <$ 1000 GeV is not enough to make QCD a significant fraction of the total background in this control region, so it is not shown.

7.8 Systematic uncertainties

From the Monte Carlo closure results in the signal regions (Figure 7.9) we take scale factors to correct for non-closure and assign an uncertainty on these scale factors as the difference from 1.0. We take one scale factor for the low and medium H_{T} regions and a separate scale factor for the high and extreme H_{T} regions. Each scale factor is the ratio of QCD yield to R&S yield in the corresponding regions. Table 7.2 gives the correction factors and Figure 7.18 shows the Monte Carlo closure in the signal regions after applying the correction factors.

Table 7.2: Scale factors to correct Monte Carlo non-closure in the signal regions

	low+med H_{T}	high+ext H_{T}
Correction Factor	1.35 ± 0.35	1.20 ± 0.20

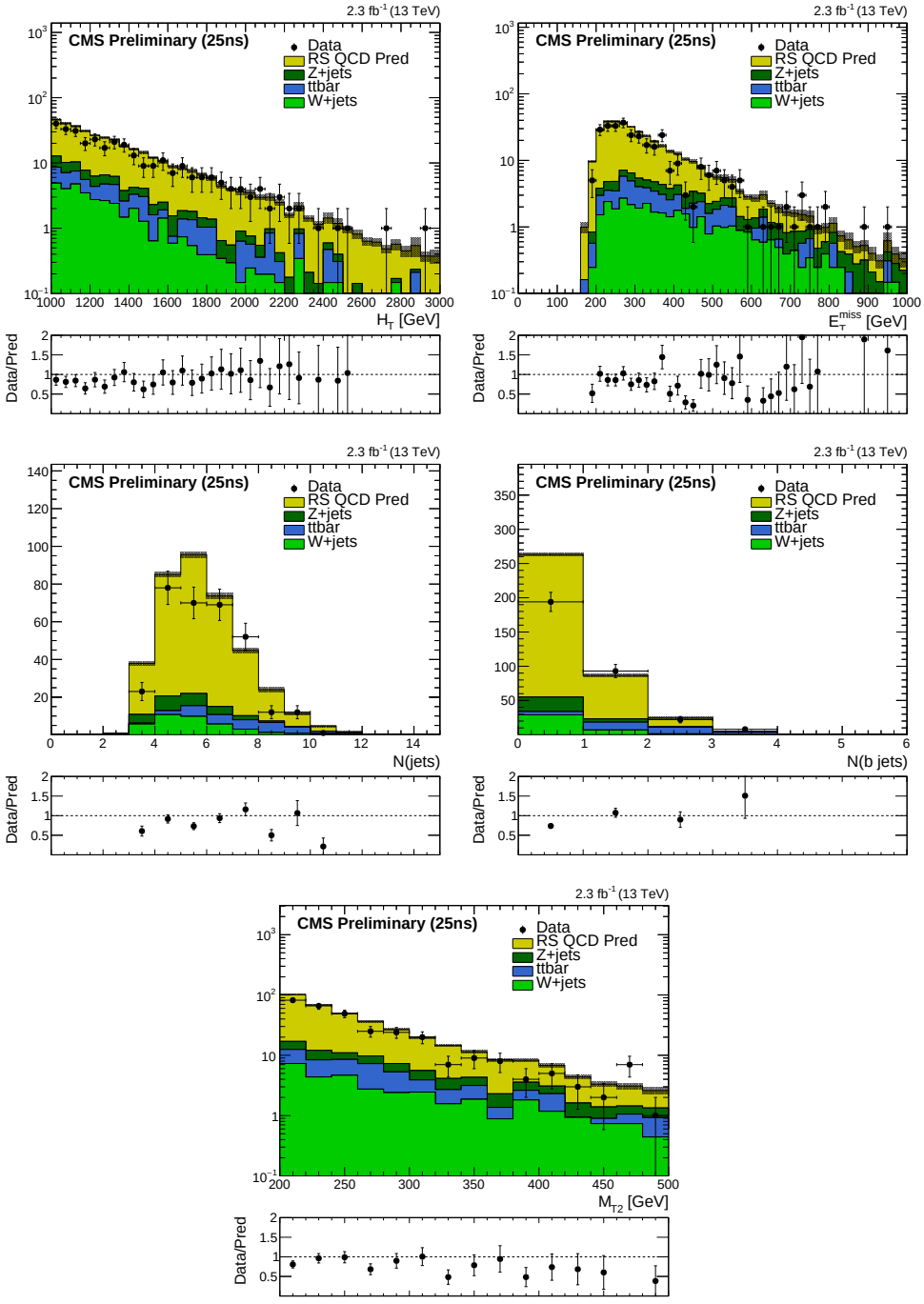

Figure 7.13: Comparison of kinematic distributions for data and background in the inverted $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ control region for $H_{\mathrm{T}} > 1000$ GeV. The QCD background is from the rebalance and smear data-driven prediction. Non-QCD backgrounds are taken from Monte Carlo.

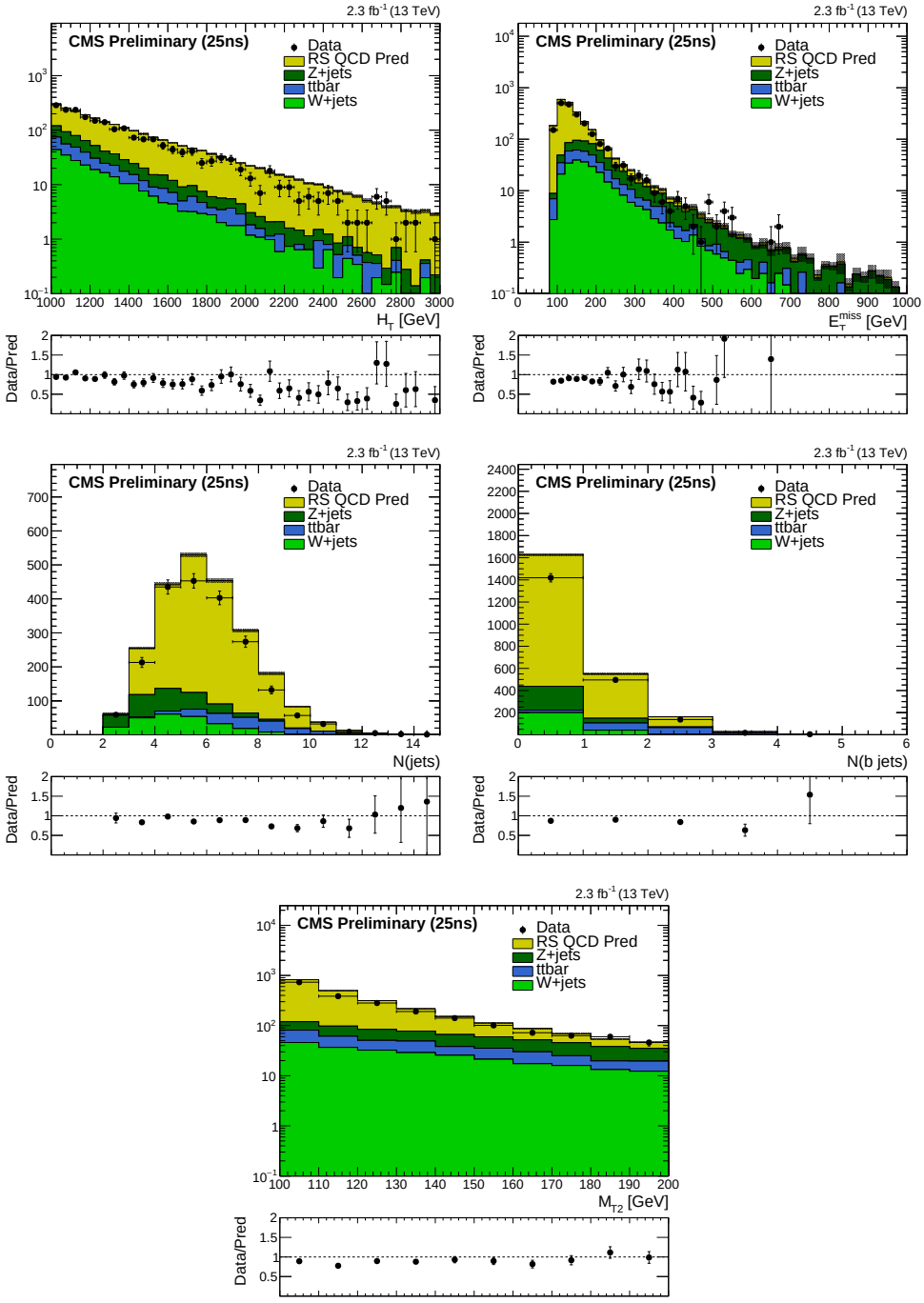

Figure 7.14: Comparison of kinematic distributions for data and background in the M_{T2} sideband control region (100 GeV $< M_{\mathrm{T2}} <$ 200 GeV) for $H_{\mathrm{T}} >$ 1000 GeV. The QCD background is from the rebalance and smear data-driven prediction. Non-QCD backgrounds are taken from Monte Carlo.

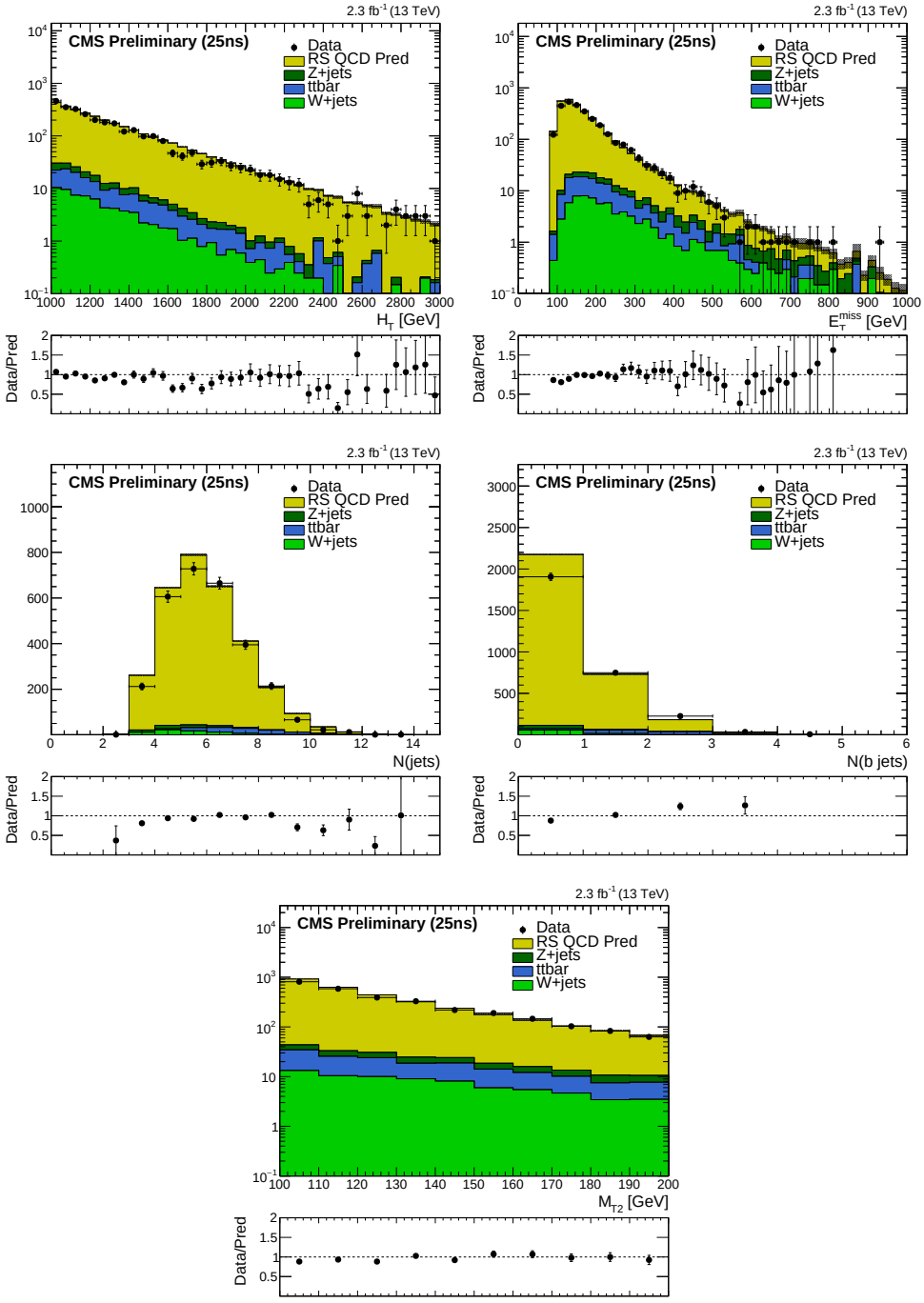

Figure 7.15: Comparison of kinematic distributions for data and background in the M_{T2} sideband + inverted $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ control region for $H_{\mathrm{T}} > 1000$ GeV. The QCD background is from the rebalance and smear data-driven prediction. Non-QCD backgrounds are taken from Monte Carlo.

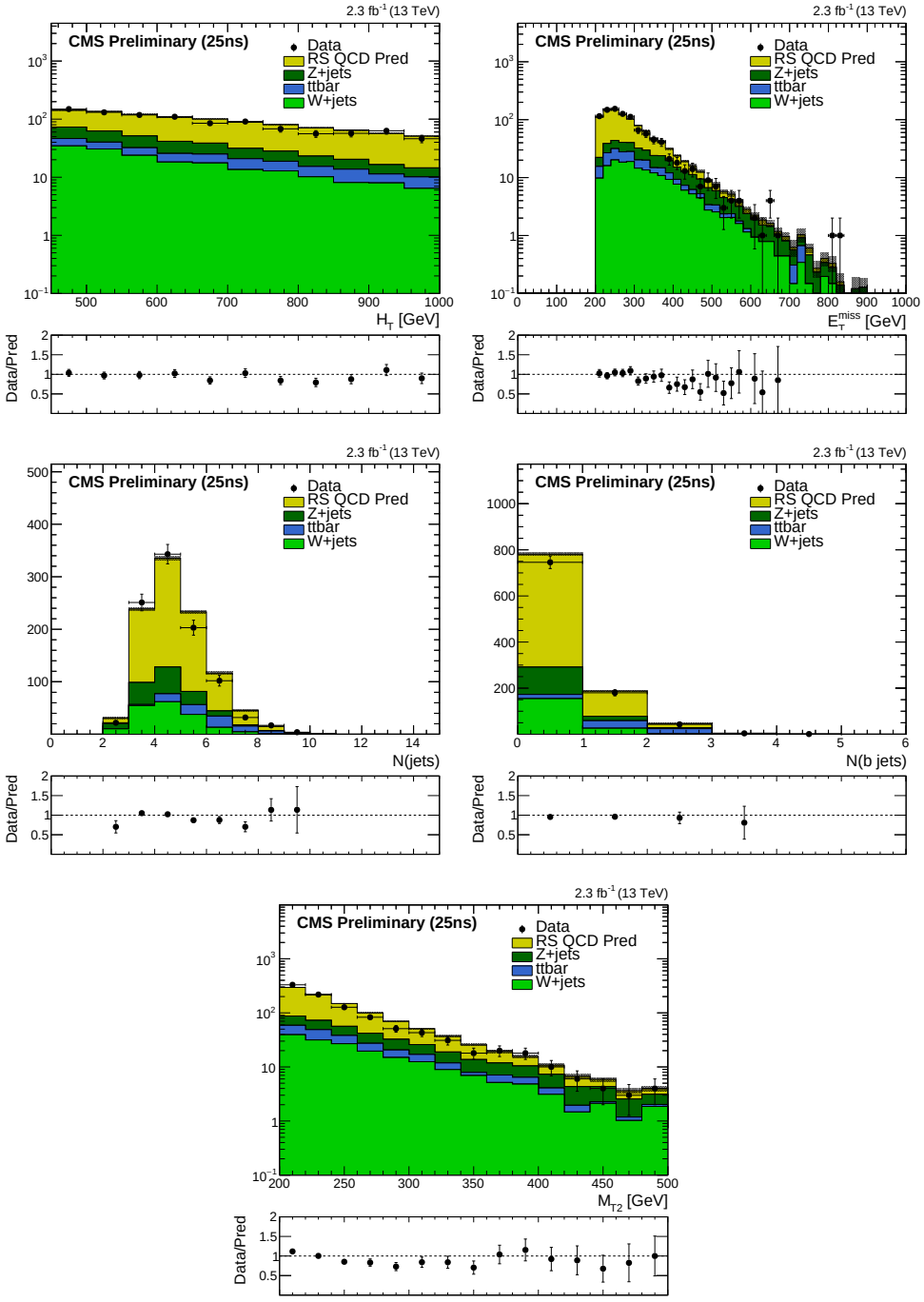

Figure 7.16: Comparison of kinematic distributions for data and background in the inverted $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ control region for 450 GeV $< H_{\mathrm{T}} <$ 1000 GeV. The QCD background is from the rebalance and smear data-driven prediction. Non-QCD backgrounds are taken from Monte Carlo.

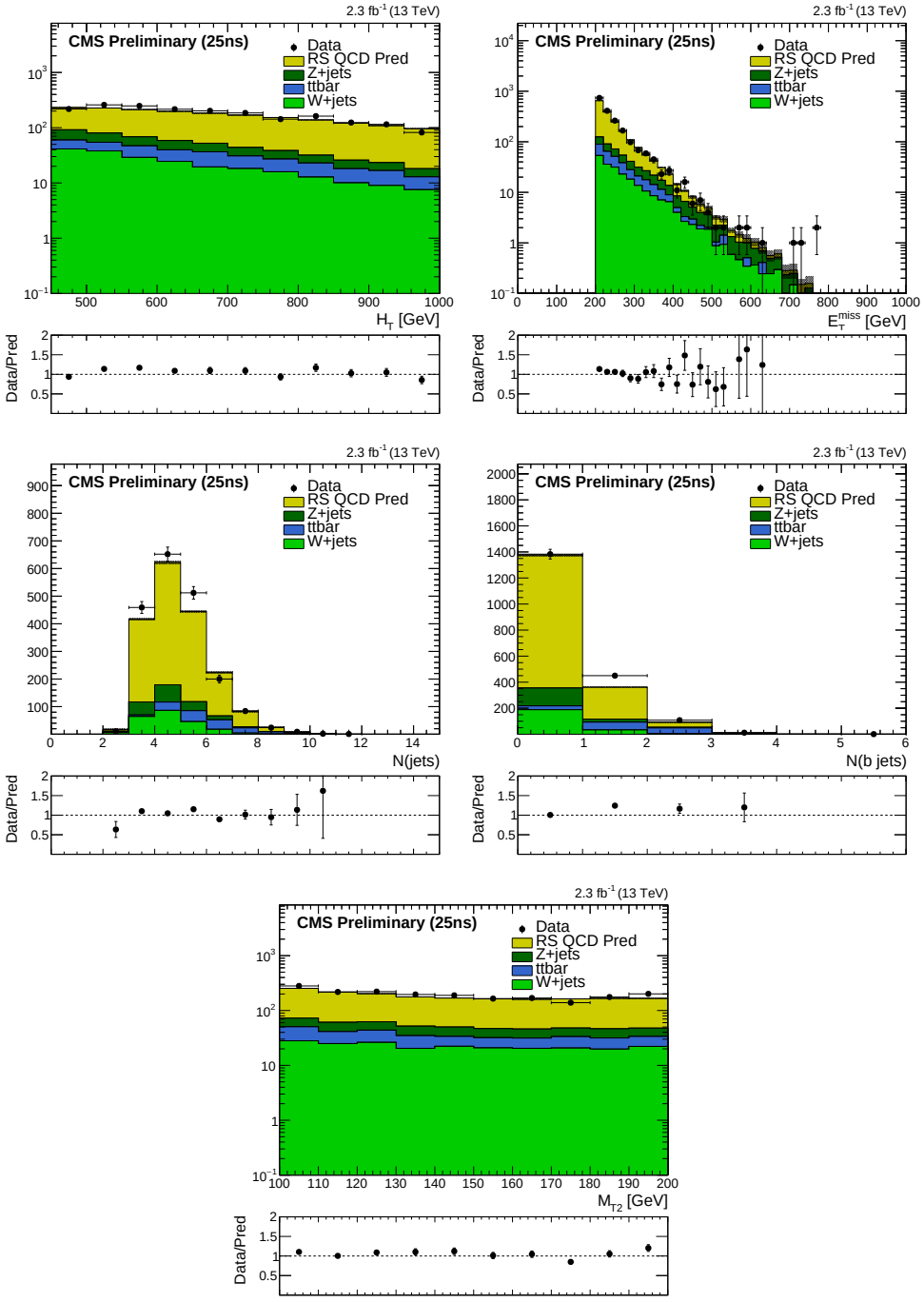

Figure 7.17: Comparison of kinematic distributions for data and background in the M_{T2} sideband + inverted $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ control region for 450 GeV $< H_{\mathrm{T}} <$ 1000 GeV. The QCD background is from the rebalance and smear data-driven prediction. Non-QCD backgrounds are taken from Monte Carlo.

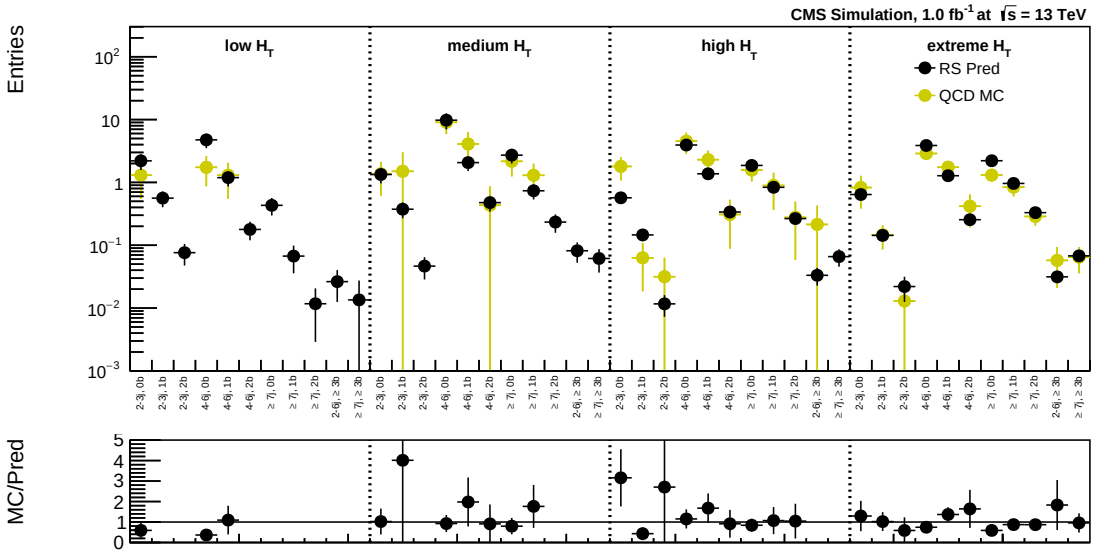

Figure 7.18: R&S Monte Carlo closure in topological regions after the baseline selection with mc closure correction factors applied.

After applying the correction factors derived from Monte Carlo closure and the correction factors to account for the effects of the cut on rebalanced $E_{\mathrm{T}}^{\mathrm{miss}}$ from Section 7.6 we derive an additional uncertainty from the non-closure in the data control regions defined in Section 7.7. Figures 7.19–7.21 show the closure in these data control regions. Table 7.3 summarizes the level of non-closure in each of these control regions. We take the largest discrepancy in Table 7.3 of 34% as a systematic for all regions. Table 7.4 summarizes all of the correct factors and uncertainties.

Table 7.3: Non closure in data control regions for low+med H_{T} and high+ext H_{T}. The non-closure is defined as (data-pred)/data. There is no entry for the M_{T2} sideband low+med H_{T} region because QCD is a negligible background here. Uncertainties are statistical.

	low+med H_{T}	high+ext H_{T}
inverted $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$	31% ± 3%	34% ± 6%
M_{T2} sideband	NA	30% ± 3%
M_{T2} sideband + inverted $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$	24% ± 2%	25% ± 2%

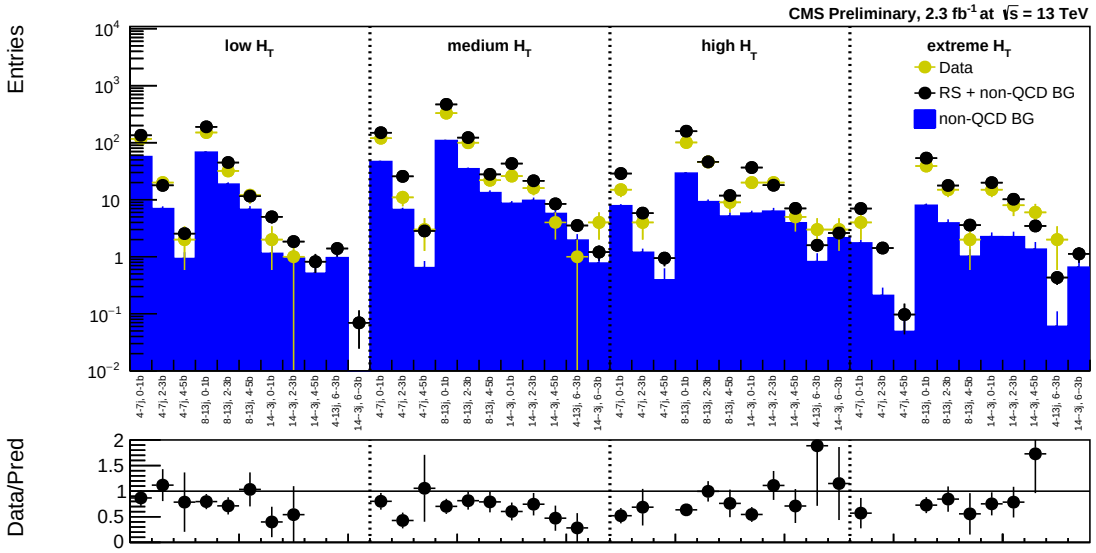

Figure 7.19: R&S closure in the inverted $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ data control region. Non-QCD backgrounds are taken from Monte Carlo.

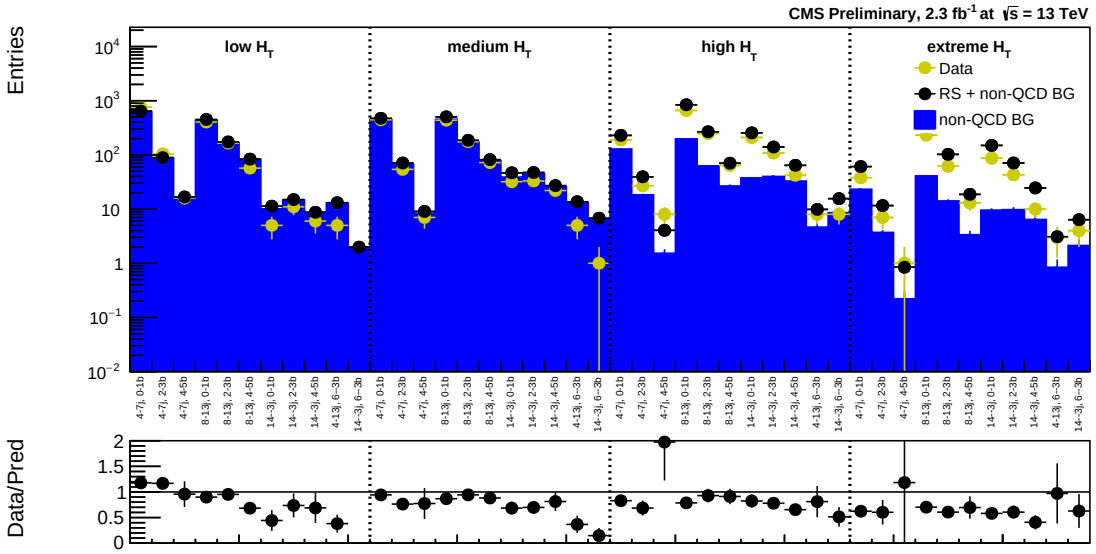

Figure 7.20: R&S closure in the M_{T2} sideband (100 GeV $< M_{\mathrm{T2}} <$ 200 GeV) data control region. Non-QCD backgrounds are taken from Monte Carlo.

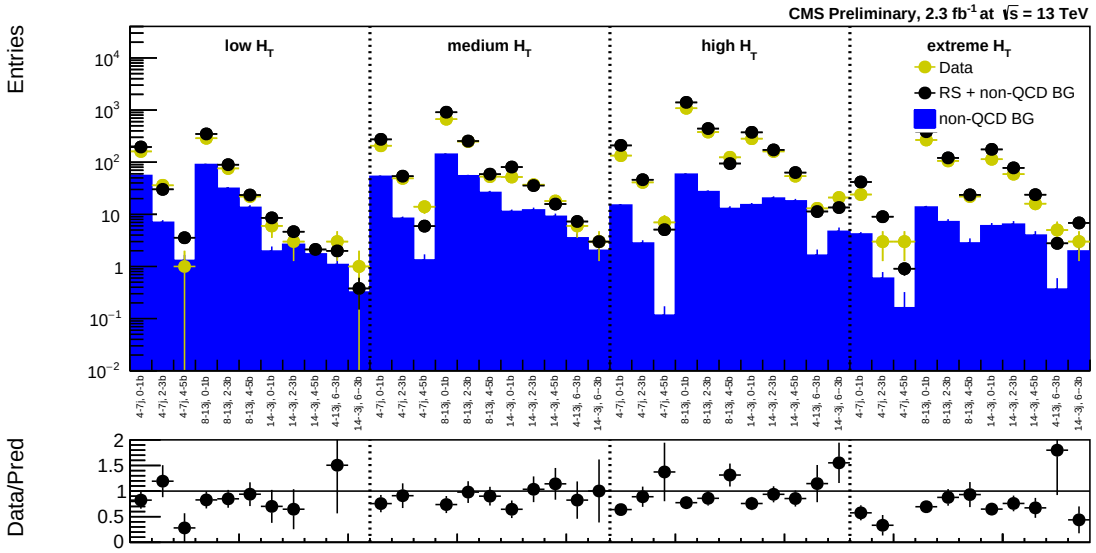

Figure 7.21: R&S closure in the M_{T2} sideband + inverted $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ data control region. Non-QCD backgrounds are taken from Monte Carlo.

Table 7.4: Summary of correction factors and uncertainties.

	low H_{T}	med H_{T}	high H_{T}	ext H_{T}
Rebalance $E_{\mathrm{T}}^{\mathrm{miss}}$ Correction	1.03 ± 0.03	1.07 ± 0.07	NA	NA
MC Closure Correction	1.35 ± 0.35	1.35 ± 0.35	1.20 ± 0.20	1.20 ± 0.20
Data CR Closure Uncertainty	34%	34%	34%	34%

7.9 Comparison with $\Delta\phi$-ratio method

Figure 7.22 shows a comparison of the data-driven predictions from the R&S method and the $\Delta\phi$-ratio method for topological regions with at least two jets and $H_{\mathrm{T}} > 450$ GeV. All correction factors and uncertainties are included in these estimates. The R&S predictions are larger than the $\Delta\phi$-ratio predictions for most regions, but the two methods agree in every topological region within the errors on the predictions, increasing our confidence in the multijet background estimation.

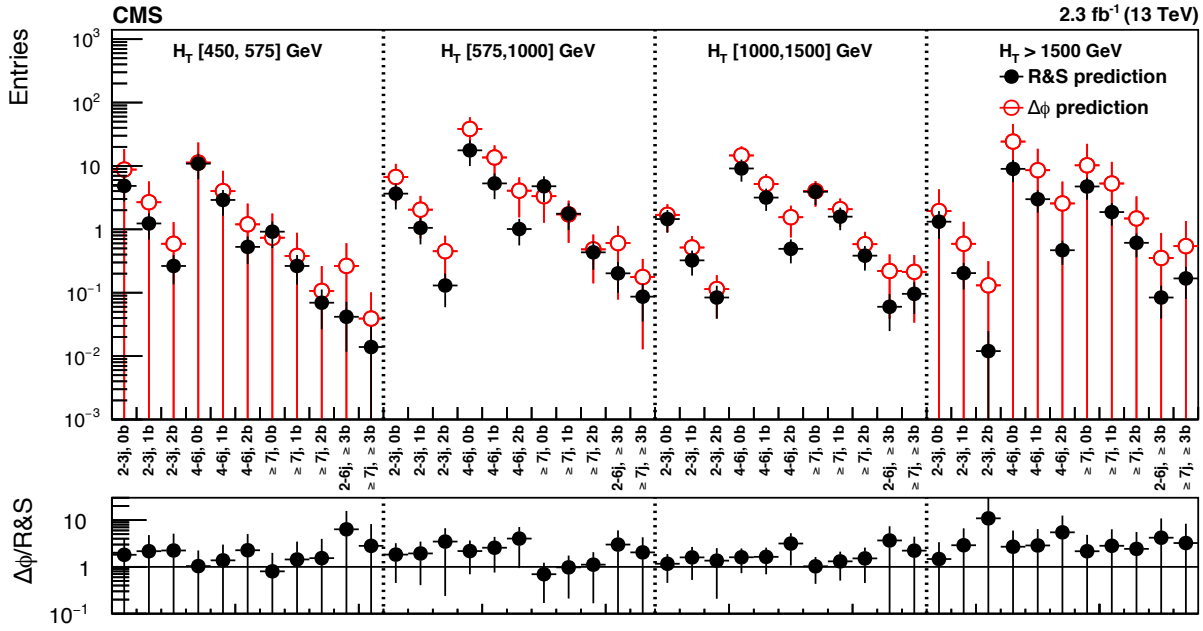

Figure 7.22: Comparison of the data-driven predictions in the signal regions from the R&S method and the $\Delta\phi$-ratio method. The uncertainties are statistical and systematic.

Chapter 8

Results and Interpretation

Figure 8.1 shows the expected and observed yields for each topological region (integrated over M_{T2}). Figures 8.2–8.7 show the expected and observed yields in every signal region. The background yields in these figures are taken from the data-driven estimates described in Chapters 4–6 and the uncertainty bands represent the full uncertainty on the data-driven estimates. As these figures show, there is no statistically significant deviation in the observed yields from the Standard Model expectation, providing no evidence for new physics. This may sound boring, but I assure you it is not! We can use these search results to constrain models of what new physics might look like. Setting limits on these models is useful because it tells theorists what isn't there, and can help guide future model-building. This is also useful for experimentalists, because then we can compare limits and discuss why my limits are the best.

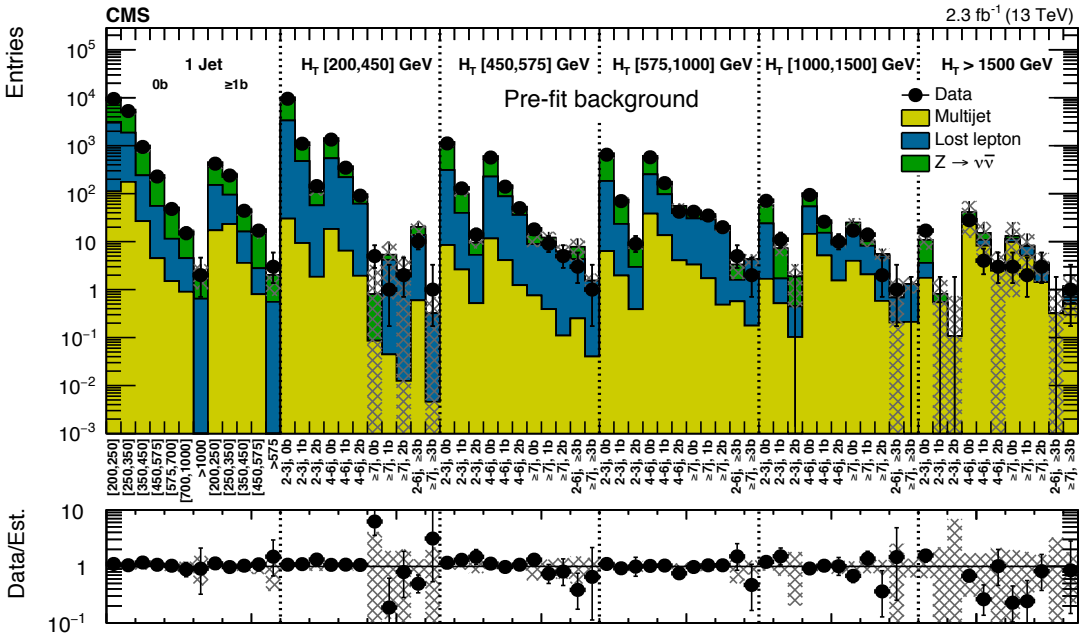

Figure 8.1: Expected (pre-fit) and observed yields in the analysis binning, for all regions and integrated over M_{T2}. For the monojet, on the x-axis the H_{T} binning is shown (in GeV).

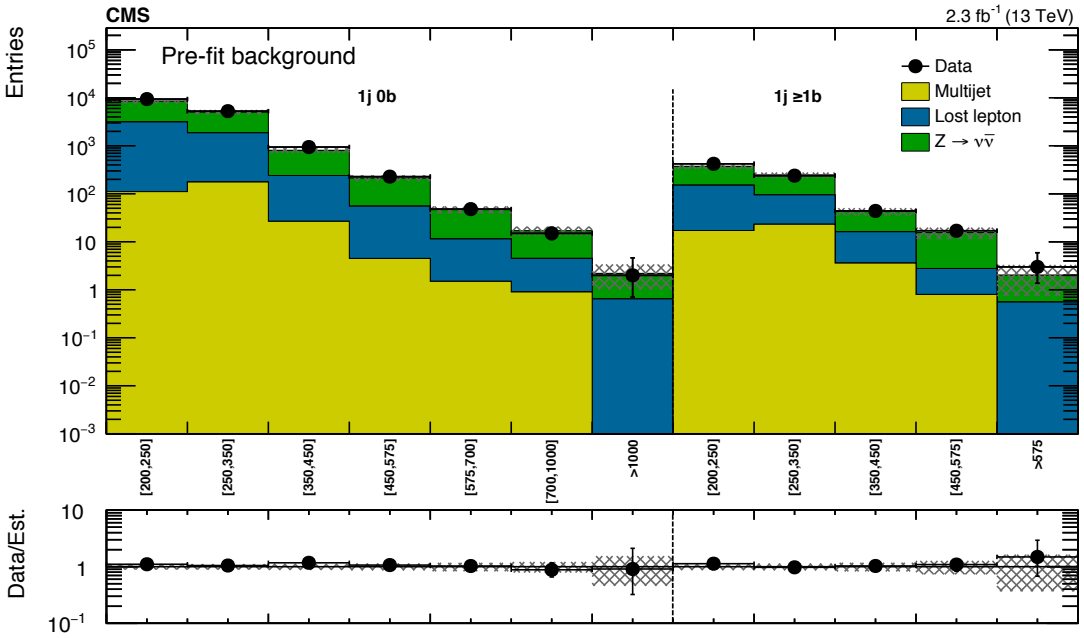

Figure 8.2: Expected (pre-fit) and observed yields in the analysis binning, for the monojet region, with p_T(jet)$>$ 200 GeV, in each of the $N_{\mathrm{b-tags}}$ bins. On the x-axis, the H_{T} binning is shown (in GeV).

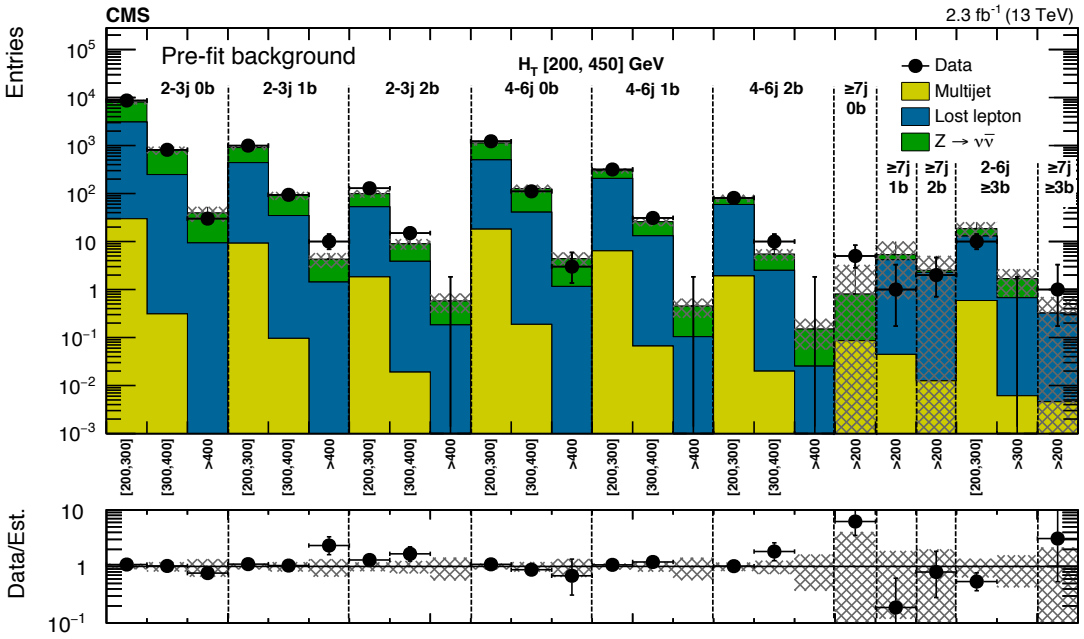

Figure 8.3: Expected (pre-fit) and observed yields in the analysis binning, for the H_{T} region [200,450] and each of the N_{jets} and $N_{\mathrm{b-tags}}$ bins. On the x-axis, the M_{T2} binning is shown (in GeV).

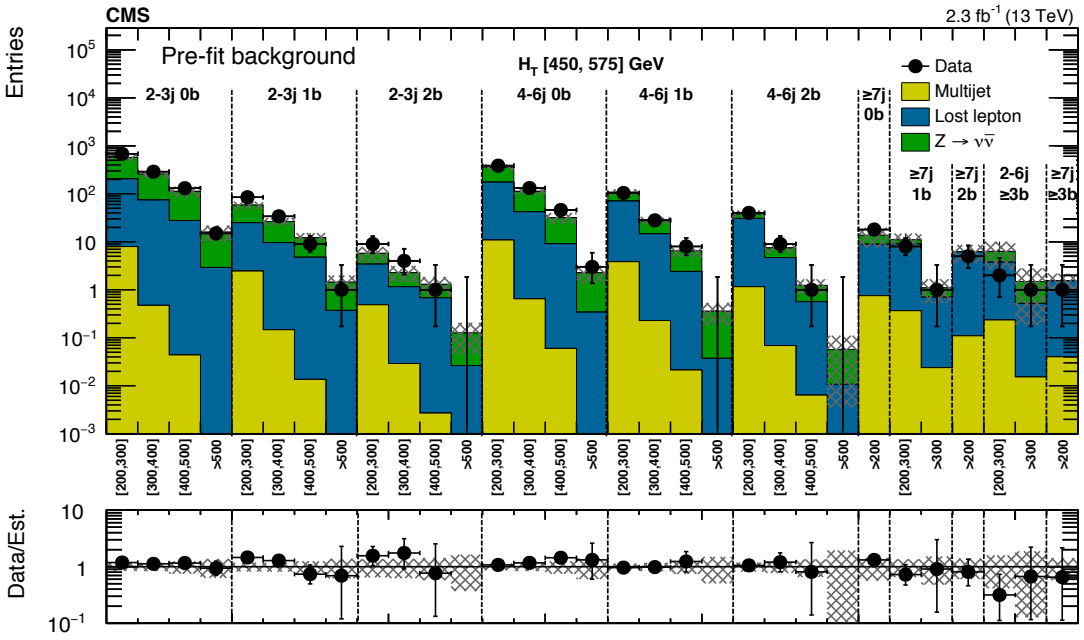

Figure 8.4: Expected (pre-fit) and observed yields in the analysis binning, for the H_{T} region [450,575] and each of the N_{jets} and $N_{\mathrm{b-tags}}$ bins. On the x-axis, the M_{T2} binning is shown (in GeV).

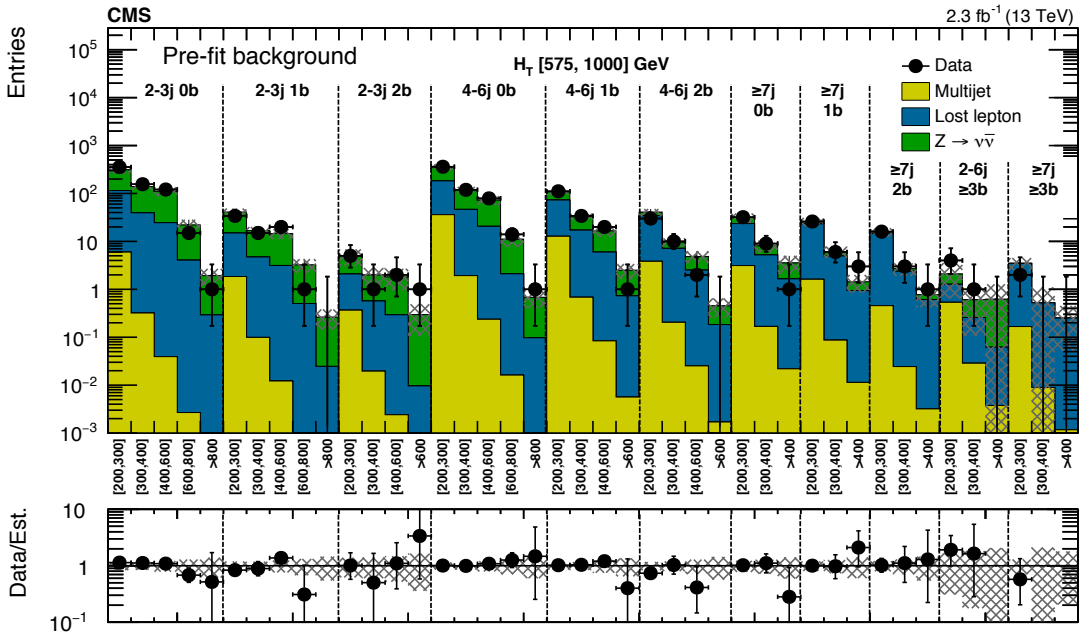

Figure 8.5: Expected (pre-fit) and observed yields in the analysis binning, for the H_{T} region [575,1000] and each of the N_{jets} and $N_{\mathrm{b-tags}}$ bins. On the x-axis, the M_{T2} binning is shown (in GeV).

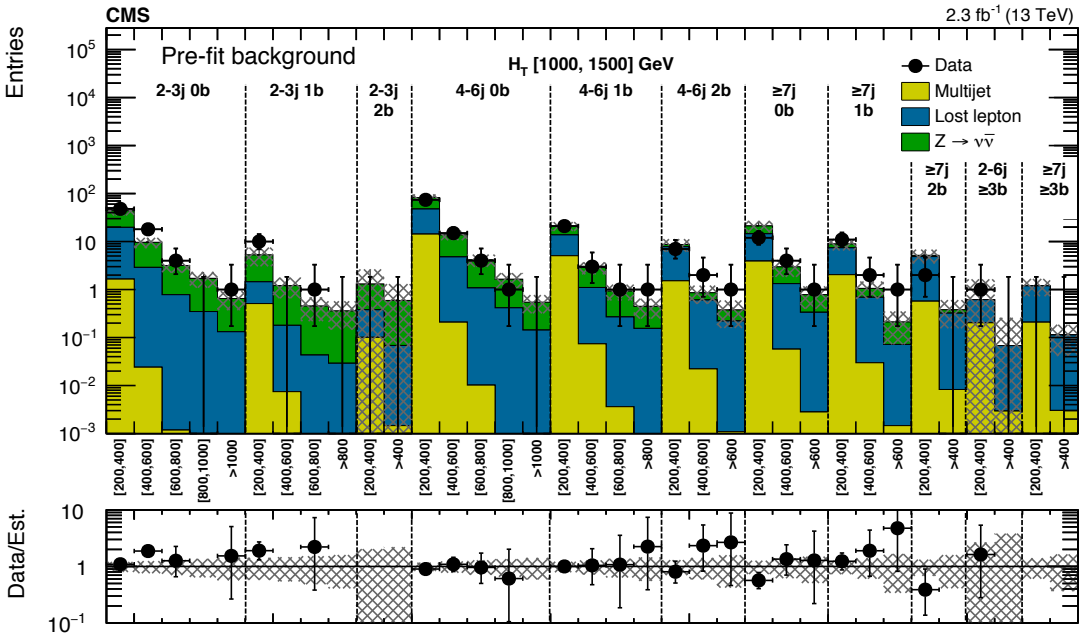

Figure 8.6: Expected (pre-fit) and observed yields in the analysis binning, for the H_{T} region [1000,1500] and each of the N_{jets} and $N_{\mathrm{b-tags}}$ bins. On the x-axis, the M_{T2} binning is shown (in GeV).

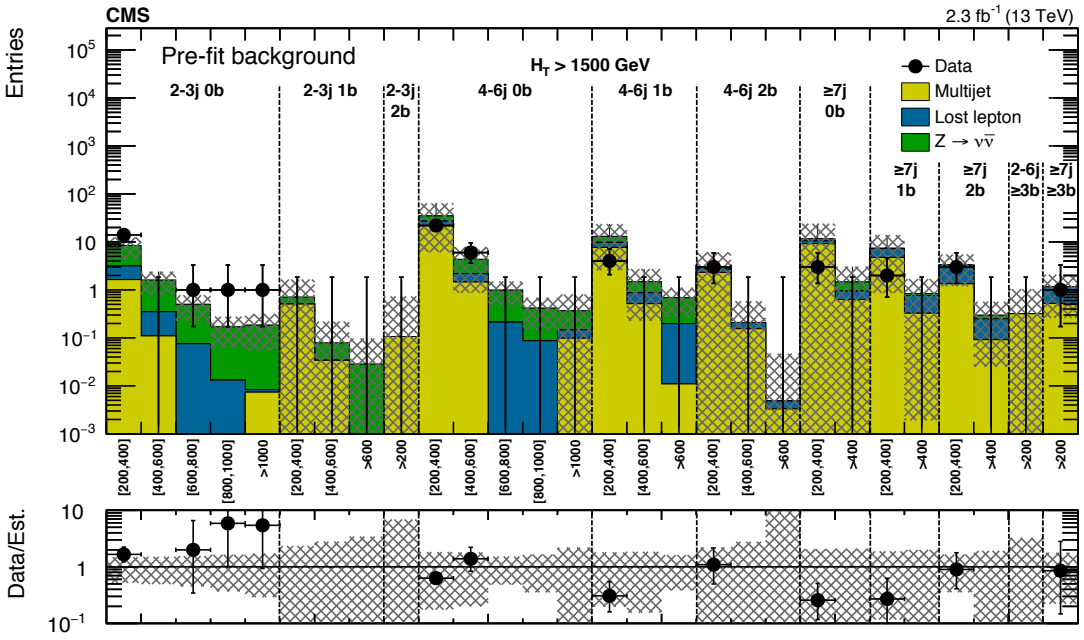

Figure 8.7: Expected (pre-fit) and observed yields in the analysis binning, for the H_{T} region [1500,∞] and each of the N_{jets} and $N_{\mathrm{b-tags}}$ bins. On the x-axis, the M_{T2} binning is shown (in GeV).

We set limits on several simplified SUSY models [28] of gluino and squark pair production. Figure 8.8 shows the diagrams for the models considered here. For gluino (squark) pair production, the simplified models assume that all supersymmetric particles other than the gluino (squark) and the lightest neutralino are too massive to be produced directly, and that the gluino (squark) undergoes a prompt decay. In the three models with gluino pair production, each gluino decays with 100% branching fraction into the lightest supersymmetric particle (LSP) and either a pair of b quarks ($\tilde{g} \to b\bar{b}\tilde{\chi}_1^0$), top quarks ($\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$), or light-flavor quarks ($\tilde{g} \to q\bar{q}\tilde{\chi}_1^0$), where the decay proceeds through an off-shell squark of the same flavor.

The constraints on a given model are set by determining the 95% confidence level upper limits on the production cross section for points in the two dimensional plane of gluino (or squark) mass and LSP mass. If the upper limit on the cross section for a

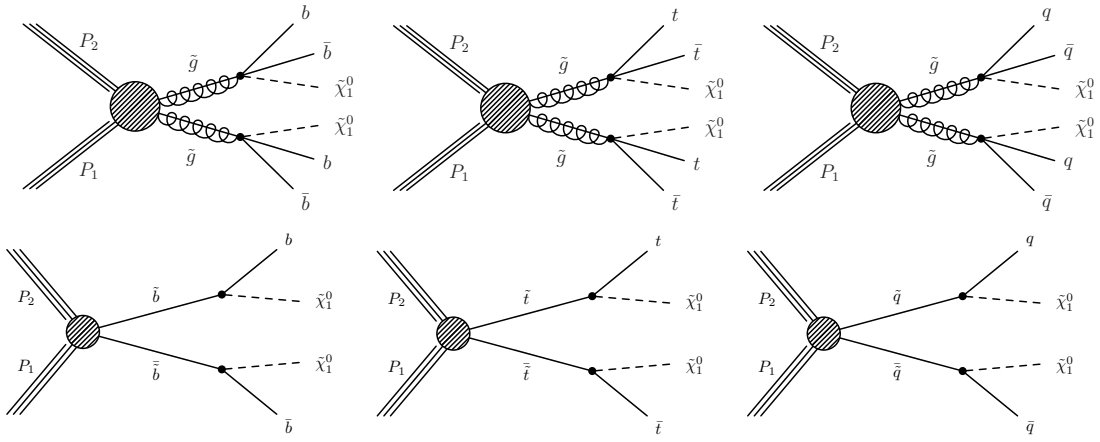

Figure 8.8: (Above) Diagrams for the three considered scenarios of gluino pair production. The depicted three-body decays are assumed to proceed through off-shell squarks. (Below) Diagrams for the three considered scenarios of direct bottom squark, top squark, and light flavor squark pair production.

signal point is less than the cross section for that signal point, then it is excluded. The signal cross sections used here are calculated at NLO+NLL order in α_s [29–33]. The cross section limit for a given point is extracted using the frequentist-motivated "CL_s" technique, which uses the likelihood ratio of the signal + background hypothesis and the background-only hypothesis [34–36]. The limits shown here make use of the asymptotic approximation [37].

When computing limits, the backgrounds in the signal regions are estimated by performing a maximum-likelihood fit to the data using the CMS Higgs Combine tool [38] in both the signal + background and background-only hypotheses. For each signal point, only the signal region bins with a nonzero signal yield are included in the fit. For signal models with leptonic decays, potential signal contamination in the single lepton control regions is taken into account as described in Section 5.3.

The uncertainties considered on signal yields are listed in Table 8.1 along with their typical values and whether they are taken as correlated across signal bins. The largest uncertainties come from MC statistics, b-tagging efficiency, and for model points with

small mass splittings, the recoil "ISR" uncertainty. The uncertainty on acceptance due to renormalization and factorization scales was found to be relatively small and a flat value of 5% is taken. The uncertainties due to PDFs and jet energy scale were found to be compatible with statistical fluctuations for bins with lower MC statistics. A flat value of 5% is taken for the jet energy scale motivated by high statistics bins.

Table 8.1: Signal systematic uncertainties with their typical values in individual signal bins. Also indicated is whether the uncertainty is taken as correlated across signal bins or not.

Source	Typical Values [%]	Correlated?
Luminosity	5	✓
MC statistics	1–100	-
Renormalization and factorization scales	5	-
"ISR" recoil	0–30	✓
B-tagging efficiency, heavy flavor	0–40	✓
B-tagging efficiency, light flavor	0–20	✓
Lepton efficiency (models with leptons only)	0–20	✓
Jet energy scale	5	-

Figure 8.9 shows exclusion limits on gluino-mediated squark production models and Figure 8.10 show exclusion limits on squark pair production. A summary of gluino production exclusion limits and squark production exclusion limits is shown in Figure 8.11. Table 8.2 summarizes the best observed limits on sparticle and LSP masses for each interpretation. The white diagonal band in the top right plot of Figure 8.10, showing limits on top squark production, corresponds to the region $|m_{\tilde{t}} - m_t - m_{LSP}| < 25$ GeV. In this compressed region, the selection efficiency of top squark events depends strongly on the difference in mass between the top squark and the LSP. Due to the finite granularity of the signal MC samples in this region, a precise determination of the cross section upper limit is not possible.

In the limit of a massless LSP, the limits on most squark and gluino production models are extended by 200–300 GeV with respect to the 8 TeV results from Run1 of the

LHC. We didn't discover anything, but at least we extended the exclusion boundaries!

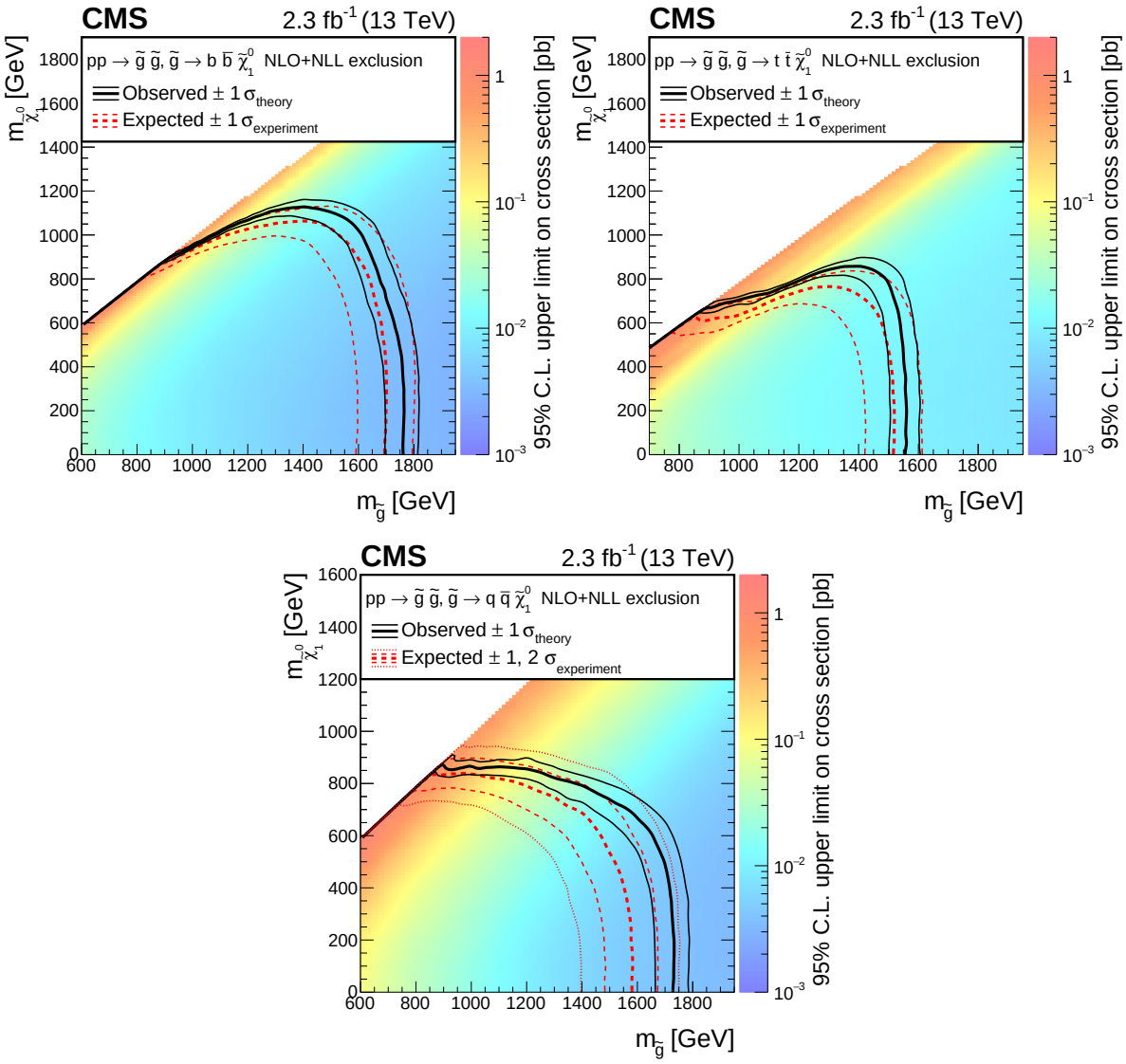

Figure 8.9: Expected and observed exclusion limits at 95% CL on gluon-mediated bottom squark production (top left), gluon-mediated top squark production (top right), and gluon-mediated light-flavor squark production (bottom).

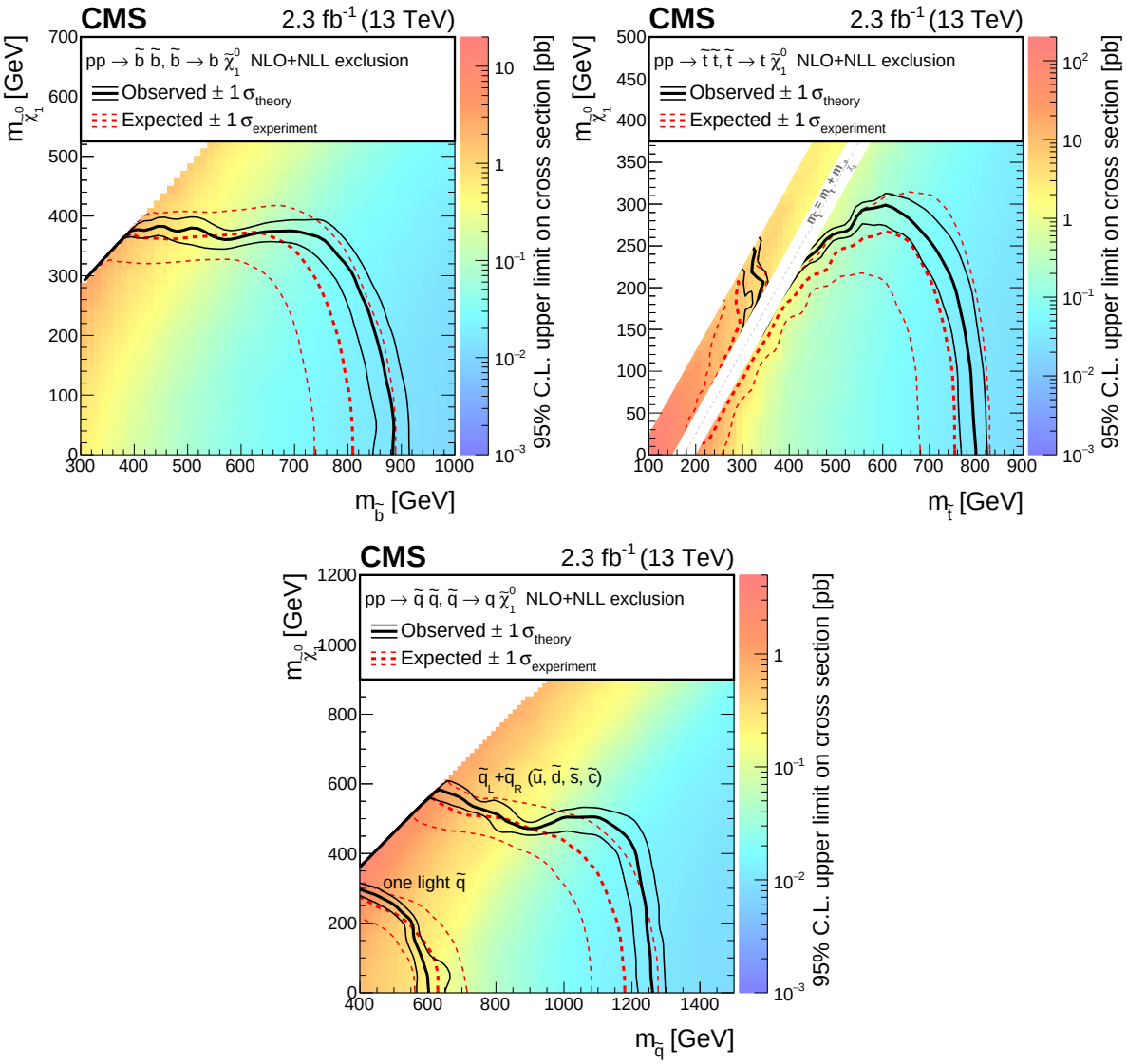

Figure 8.10: Expected and observed exclusion limits at 95% CL on bottom squark pair production (top left), top squark pair production (top right) and light-flavor squark pair production (bottom).

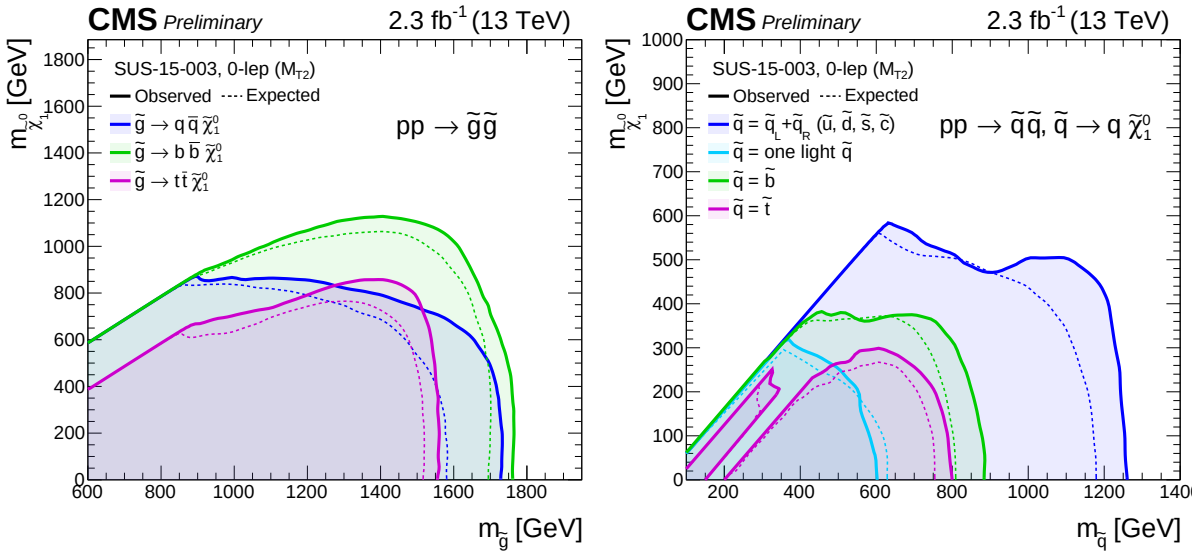

Figure 8.11: Summary of exclusion limits for gluino production (left) and squark production (right).

Table 8.2: Summary of 95% CL observed exclusion limits for different SUSY simplified model scenarios. The limit on the mass of the produced sparticle is quoted for a massless LSP, while for the lightest neutralino the best limit on its mass is quoted.

Simplified model	Limit on produced sparticle mass [GeV] at $m_{\tilde{\chi}_1^0} = 0$GeV	Best limit on LSP mass [GeV]
Direct squark production		
Bottom squark	880	380
Top squark	800	300
Single light squark	600	300
8 degenerate light squarks	1260	580
Gluino mediated production		
$\tilde{g} \to b\bar{b}\tilde{\chi}_1^0$	1750	1125
$\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$	1550	825
$\tilde{g} \to q\bar{q}\tilde{\chi}_1^0$	1725	850

Bibliography

[1] S. L. Glashow, *Partial-symmetries of weak interactions*, *Nuclear Physics* **22** (1961), no. 4 579 – 588.

[2] S. Weinberg, *A model of leptons*, *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266.

[3] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc.* **C680519** (1968) 367–377.

[4] S. P. Martin, *A Supersymmetry primer*, hep-ph/9709356. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[5] F. Zwicky, *Spectral displacement of extra galactic nebulae*, *Helv. Phys. Acta* **6** (1933) 110–127.

[6] K. Garrett and G. Duda, *Dark Matter: A Primer*, *Adv. Astron.* **2011** (2011) 968283, [arXiv:1006.2483].

[7] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation* **3** (2008), no. 08 S08001.

[8] `https://mstwpdf.hepforge.org/plots/crosssections2009_v2.png`.

[9] **CMS** Collaboration, S. Chatrchyan *et. al.*, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[10] **UA1** Collaboration, *Experimental observation of isolated large transverse energy electrons with associated missing energy at sqrt(s) = 540 gev*, *Physics Letters B* **122** (1983), no. 1 103 – 116.

[11] **CMS** Collaboration, *Measurements of inclusive w and z cross sections in pp collisions at sqrt(s) = 7 tev*, *JHEP* **10** (2011), no. 132.

[12] C. Lester and D. Summers, *Measuring masses of semi-invisibly decaying particle pairs produced at hadron colliders*, *Physics Letters B* **463** (1999), no. 1 99 – 103.

[13] H.-C. Cheng and Z. Han, *Minimal kinematic constraints and mt2*, *Journal of High Energy Physics* **2008** (2008), no. 12 063.

[14] A. J. Barr and C. Gwenlan, *The race for supersymmetry: Using m_{T2} for discovery*, *Phys. Rev. D* **80** (Oct, 2009) 074007.

[15] **CMS** Collaboration, *Search for supersymmetry in hadronic final states using mt2 in pp collisions at sqrt(s) = 7 tev*, *JHEP* **10** (2012), no. 18.

[16] **CMS** Collaboration, *Searches for supersymmetry using the mt2 variable in hadronic events produced in pp collisions at 8 tev*, *JHEP* **05** (2015), no. 78.

[17] **CMS Collaboration** Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*, Tech. Rep. CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr, 2009.

[18] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **04** (2008) 063, [arXiv:0802.1189].

[19] `https://twiki.cern.ch/twiki/bin/view/CMS/JetID`.

[20] `https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation74X50ns`.

[21] `https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2`.

[22] `https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2`.

[23] `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId`.

[24] **CMS** Collaboration, *Measurement of differential cross sections for the production of a pair of isolated photons in pp collisions at sqrt(s)=7 tev*, *The European Physical Journal C* **74** (2014), no. 11.

[25] "Search for direct stop pair production in the semi-leptonic channel at 8 tev." AN-2014/067.

[26] `https://twiki.cern.ch/twiki/bin/view/CMS/SUSLeptonSF`.

[27] F. James and M. Roos, *MINUIT-a system for function minimization and analysis of the parameter errors and correlations*, *Comput. Phys. Commun.* **10** (1975), no. 6 343–67.

[28] **CMS** Collaboration, *Interpretation of searches for supersymmetry with simplified models*, *Phys. Rev. D* **88** (Sep, 2013) 052017.

[29] W. Beenakker, R. Höpker, M. Spira, and P. M. Zerwas, *Squark and gluino production at hadron colliders*, *Nucl. Phys. B* **492** (1997) 51.

[30] A. Kulesza and L. Motyka, *Threshold resummation for squark-antisquark and gluino-pair production at the LHC*, *Phys. Rev. Lett.* **102** (2009) 111802.

[31] A. Kulesza and L. Motyka, *Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC*, *Phys. Rev. D* **80** (2009) 095004.

[32] W. Beenakker, S. Brensing, M. Krämer, A. Kulesza, E. Laenen, and I. Niessen, *Soft-gluon resummation for squark and gluino hadroproduction*, *JHEP* **12** (2009) 041.

[33] W. Beenakker, S. Brensing, M. Krämer, A. Kulesza, E. Laenen, L. Motyka, and I. Niessen, *Squark and gluino hadroproduction*, *Int. J. Mod. Phys. A* **26** (2011) 2637.

[34] A. L. Read, *Presentation of search results: The CL_s technique*, *J. Phys. G* **28** (2002) 2693.

[35] A. L. Read, *Modified frequentist analysis of search results (the CL_s method)*, *CERN-OPEN* **205** (2000).

[36] ATLAS and CMS Collaborations, *Procedure for the LHC Higgs boson search combination in summer 2011*, CMS NOTE/ATL-PHYS-PUB ATL-PHYS-PUB-2011-011, CMS-NOTE-2011-005, 2011.

[37] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554, [arXiv:1007.1727].

[38] "Higgs combine tool." `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit`.