

CMS Draft Analysis Note

The content of this note is intended for CMS internal use and distribution only

2015/02/06

Head Id: 276410

Archive Id: 276495

Archive Date: 2015/02/05

Archive Tag: trunk

Search for supersymmetry in the multijet and missing transverse momentum channel in pp collisions at 13 TeV

A. Borzou¹, K. Hatakeyama¹, H. Liu¹, J. W. Gary², P. Jandir², F. Lacroix², O. Long², H. Wei², J. Bradmiller-Feld³, J. Richman³, M. Franco Sevilla³, A. Ovcharova³, W. T. Ford⁴, F. O. Jensen⁴, T. Mulholland⁴, K. Stenson⁴, V. D. Elvira⁵, J. Hirschauer⁵, N. Tran⁵, A. Whitbeck⁵, A. Dräger⁶, K. Heine⁶, S. Kurz⁶, M. Niedziela⁶, C. Sander⁶, S. Sharma⁷, B. Mahakud⁸, K. A. Ulmer⁹, and R. Patel⁹

¹ Baylor University

² University of California, Riverside

³ University of California, Santa Barbara

⁴ University of Colorado at Boulder

⁵ Fermi National Accelerator Laboratory

⁶ University of Hamburg

⁷ Indian Institute of Science Education and Research, Pune

⁸ Tata Institute of Fundamental Research

⁹ Texas A&M University

Abstract

An inclusive search for supersymmetry in all-hadronic channels, based on the H_T , H_T^{miss} , N_{jet} , and $N_{\text{b-jet}}$ variables, is described. This note is intended to document the analysis proposed by the RA2/b group for an early CMS publication utilizing the 13 TeV data to be collected in 2015, and was prepared as part of the PHYS14 exercise. The proposed analysis techniques, expected sources of systematic uncertainty, and anticipated sensitivity are discussed.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor:	The RA2/b team
PDFTitle:	Search for supersymmetry in the multijet and missing transverse momentum channel in pp collisions at 13 TeV
PDFSubject:	CMS
PDFKeywords:	CMS, physics, software, computing

Please also verify that the abstract does not use any user defined symbols

DRAFT

1 Introduction

This note describes a search for supersymmetry (SUSY) in all-hadronic final states. Many models for physics beyond the standard model (SM) predict final states with jets and transverse momentum imbalance \vec{p}_T^{miss} . The momentum imbalance arises from weakly interacting new-physics particles that escape without detection. Among new-physics models, SUSY has received much attention because it may predict unification of the strong, weak, and electromagnetic forces, and provide a dark matter candidate. In particular, “natural” SUSY models minimize the fine tuning associated with the value of the Higgs boson mass and its radiative corrections. In natural SUSY, the top squark, bottom squark, gluino, and higgsino are required to be “light”, i.e., to have masses at the electroweak energy scale, making them potentially accessible at the LHC.

Amongst possible SUSY production processes, gluino pair production has the largest cross section, and – in the natural SUSY framework – can be expected to lead to top- and bottom-squark production. Gluino pair production leads to final states containing a large number of jets.

We present a general search for gluino pair production leading to final states with large \vec{p}_T^{miss} and large jet multiplicity. The data are examined in bins of N_{jet} and $N_{\text{b-jet}}$, where N_{jet} is the number of jets and $N_{\text{b-jet}}$ is the number of jets tagged as arising from a bottom-quarks (b jet). Bottom- and top-squark production will naturally lead to final states with tagged b jets. Events with top-squark production can be identified in a complementary way by examining events with large values of N_{jet} . The search is thus sensitive to natural SUSY scenarios. The search variables are H_T , H_T^{miss} , N_{jet} , and $N_{\text{b-jet}}$, where H_T is the scalar sum of the transverse momentum values (p_T) of the jets, while H_T^{miss} is the magnitude of the vector sum of the jet momenta: H_T characterizes the level of hadronic activity in an event, which is typically larger in new-physics scenarios than in the SM, while H_T^{miss} is a measure of \vec{p}_T^{miss} . The search is performed in exclusive bins of these four observables.

This study combines and extends search strategies developed using the CMS 8 TeV data, specifically the study in Ref. [1] (SUS-12-024), which examined data in bins of $N_{\text{b-jet}}$ but not N_{jet} , and which proved sensitive to SUSY scenarios with gluino decays to bottom squarks, and the study in Ref. [2] (SUS-13-012), which examined data in bins of N_{jet} but not $N_{\text{b-jet}}$, and proved sensitive to scenarios with gluino decay to top squarks. The two approaches are combined in a unified framework to yield improved sensitivity.

2 Trigger

The trigger for the search, single-lepton, and QCD control samples will be HLT_PFHT350_PFMET120_NoiseCleaned_v1, which is currently seeded at Level 1 by L1_HTT150 or L1_HTT175, though this may be changed to L1_HTT200. To reduce the rate into the online particle flow, this HLT path requires online calo-based $H_T > 280$ GeV and calo-based $E_T^{\text{miss}} > 70$ GeV. Figure 1 shows the trigger turn-on for the Level 1, online calo-based, and final HLT particle flow requirements as a function of the offline particle flow E_T^{miss} and H_T . The curves were made using CMSSW_7.1.7 and a privately produced SUSY MC sample used for trigger development.

This trigger may need to be changed to the L1_HTT200 seed when the machine reaches its full instantaneous luminosity. The final HLT particle flow H_T threshold of 350 GeV was chosen with this in mind, since the turn-on for lower PFHT thresholds would become limited by the

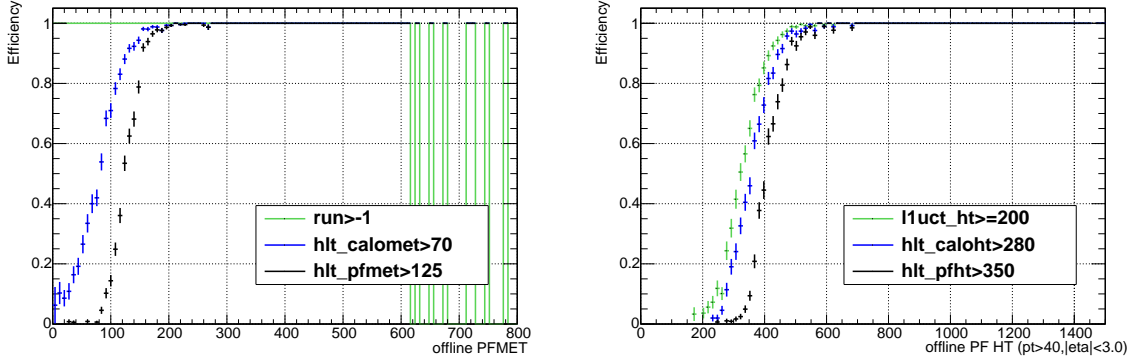


Figure 1: Turn-on curves for the Level 1 (green), online calo-based (blue), and HLT particle flow (black) requirements as a function of offline PFMET (left) and offline PFHT (right) for the HLT_PFHT350_PFMET120_NoiseCleaned_v1 trigger.

L1_HTT200 requirement, as can be seen in Fig. 1. The HLT particle flow E_T^{miss} threshold of 120 GeV was selected in order to keep the HLT rate at full instantaneous luminosity below around 40 Hz.

Our baseline selection requires $H_T > 500$ GeV and $H_T^{\text{miss}} > 200$ GeV. Assuming E_T^{miss} and H_T^{miss} to be roughly interchangeable for values of either variable above 150 GeV, we expect the trigger to be $> 95\%$ efficient for events that satisfy the baseline selection. We will confirm this with newer Monte Carlo samples and adjust the baseline selection if necessary in order to maintain $> 95\%$ trigger efficiency.

3 Event samples

3.1 Standard model MC samples

Monte Carlo (MC) samples reconstructed with CMMSW release 7.2 (Phys14DR) are used throughout. The SM samples are listed in Table 1. The cross sections listed are NLO unless otherwise noted. All samples except signal scans use the PU20bx25 pileup scenario, which simulates a pileup distribution with an average of 20 interactions per bunch crossing and a 25 ns interval between bunches.

3.2 Signal models

T1qqqq, T1tttt, T1bbbb and mass points selected for the study, with motivation.

4 Event selection

The search region for the analysis contains events with large H_T^{miss} , H_T , multiple jets, and no leptons. The analysis targets three signal models differing primarily in the number of jets and b-tagged jets produced in the gluino decay. Each of these models can be further divided into separate topologies with different mass splitting between the gluino and LSP. To maximize sensitivity to this diverse collection of topologies, a loose baseline selection is initially applied in H_T^{miss} , H_T , and N_{jet} . As indicated in Table 4, this baseline selection preserves about two-thirds of the non-compressed T1bbbb and T1qqqq events (which have no leptons), about one-third of the non-compressed T1tttt events (which may have leptons), and 10-20% of the compressed

Table 1: Standard model MC samples used in the analysis.

Dataset	σ (pb)	$\int \mathcal{L} dt$ (fb ⁻¹)
QCD_HT-250To500_13TeV-madgraph	6.705×10^5	4×10^{-3}
QCD_HT-500To1000_13TeV-madgraph	2.674×10^4	0.1
QCD_HT-1000ToInf_13TeV-madgraph	769.7	1.9
TTJets_MSDecaysCKM.central_Tune4C_13TeV-madgraph-tauola	806.1	31.6
TTbarH_M-125_13TeV_amcatnlo-pythia8-tauola	0.509	398
TTWJets_Tune4C_13TeV-madgraph-tauola	1.152	214
TTZJets_Tune4C_13TeV-madgraph-tauola	2.232	111.7
TToLeptons_s-channel-CSA14_Tune4C_13TeV-aMCatNLO-tauola	2	250
TBarToLeptons_s-channel-CSA14_Tune4C_13TeV-aMCatNLO-tauola	1	250
TToLeptons_t-channel-CSA14_Tune4C_13TeV-aMCatNLO-tauola	103.4	38.6
TBarToLeptons_t-channel_Tune4C_CSA14_13TeV-aMCatNLO-tauola	61.6	32.5
T_tW-channel-DR_Tune4C_13TeV-CSA14-powheg-tauola	35	28.2
Tbar_tW-channel-DR_Tune4C_13TeV-CSA14-powheg-tauola	35	27.8
WJetsToLNu_HT-100to200_Tune4C_13TeV-madgraph-tauola	1817	2.9
WJetsToLNu_HT-200to400_Tune4C_13TeV-madgraph-tauola	471.6	10.5
WJetsToLNu_HT-400to600_Tune4C_13TeV-madgraph-tauola	55.61	83.4
WJetsToLNu_HT-600toInf_Tune4C_13TeV-madgraph-tauola	18.81	243.6
ZJetsToNuNu_HT-100to200_Tune4C_13TeV-madgraph-tauola	372.6	13.4
ZJetsToNuNu_HT-200to400_Tune4C_13TeV-madgraph-tauola	100.8	45.1
ZJetsToNuNu_HT-400to600_Tune4C_13TeV-madgraph-tauola	11.99	369.8
ZJetsToNuNu_HT-600toInf_Tune4C_13TeV-madgraph-tauola	4.113	1085
WH_HToBB_WToLNu_M-125_13TeV_powheg-herwigpp	0.0853	2345
ZH_HToBB_WToLNu_M-125_13TeV_powheg-herwigpp	0.099	1022
Samples for control sample studies		
DYJetsToLL_M-50_HT-100to200_Tune4C_13TeV-madgraph-tauola	194.3	20.9
DYJetsToLL_M-50_HT-200to400_Tune4C_13TeV-madgraph-tauola	52.24	89.3
DYJetsToLL_M-50_HT-400to600_Tune4C_13TeV-madgraph-tauola	6.546	753.3
DYJetsToLL_M-50_HT-600toInf_Tune4C_13TeV-madgraph-tauola	2.179	2062
GJets_HT-100to200_Tune4C_13TeV-madgraph-tauola	1534	3.1
GJets_HT-200to400_Tune4C_13TeV-madgraph-tauola	489.9	9.3
GJets_HT-400to600_Tune4C_13TeV-madgraph-tauola	62.05	70
GJets_HT-600toInf_Tune4C_13TeV-madgraph-tauola	20.87	208

Table 2: MC samples for signal SMS model points. Note that all samples are generated with FullSim.

Dataset	σ (pb)	$\int \mathcal{L} dt$ (fb ⁻¹)
SMS-T1tttt_2J_mG1-1500_mLSP-100_Tune4C_13TeV-madgraph-tauola	0.014	7447
SMS-T1tttt_2J_mG1-1200_mLSP-800_Tune4C_13TeV-madgraph-tauola	0.086	1171
SMS-T1bbbb_2J_mG1-1500_mLSP-100_Tune4C_13TeV-madgraph-tauola	0.014	7410
SMS-T1bbbb_2J_mG1-1000_mLSP-900_Tune4C_13TeV-madgraph-tauola	0.325	298.5
SMS-T1qqqq_2J_mG1-1400_mLSP-100_Tune4C_13TeV-madgraph-tauola	0.025	4067
SMS-T1qqqq_2J_mG1-1000_mLSP-800_Tune4C_13TeV-madgraph-tauola	0.325	297.1

events. The events passing this selection are then binned in H_T^{miss} , H_T , N_{jet} , and $N_{\text{b-jet}}$. $N - 1$ plots of these variables with respect to the baseline selection are presented in Figure 2. In addition, $N - 1$ plots of these variables in the most sensitive regions for each model point are shown in Appendix A.

The following cuts define the baseline selection:

- $N_{\text{jet}} \geq 4$:

Since every pair of gluinos decays to four quarks, all events must contain at least four “Good” jets, defined by

- $p_T > 30 \text{ GeV}$,
- $|\eta| < 2.4$,
- Passing “loose” jet ID criteria for PF jets [3]:
 - Neutral hadron fraction < 0.99 ,
 - Neutral EM fraction < 0.99 ,
 - Number of constituents > 1 ,
 - Charged hadron fraction > 0 ,
 - Charged multiplicity > 0 ,
 - Charged EM fraction < 0.99 .

The loose p_T cut on the jets is chosen to increase sensitivity to models with small mass splitting between the gluino and LSP and thus softer H_T^{miss} and H_T spectra.

- $H_T > 500 \text{ GeV}$, where $H_T = \sum_{\text{jets}} p_T$. The jets must meet the criteria listed above.
- $H_T^{\text{miss}} > 200 \text{ GeV}$, where $H_T^{\text{miss}} = \left| \sum_{\text{jets}} \vec{p}_T \right|$. The jets included in this sum must satisfy $p_T > 30 \text{ GeV}$, $|\eta| < 5$, and the “loose” jet ID listed above.

- Muon veto:

Muon candidates are selected using the POG-recommended “Tight Muon” selection [4]. Muon candidates are required to have $p_T > 10 \text{ GeV}$ and $|\eta| < 2.4$. The recommended PF relative-isolation criterion with $\delta\beta$ -based corrections and $\Delta R = 0.4$ is used with a “loose” cut of < 0.20 to eliminate events with an isolated muon.

- Electron veto:

Electron candidates are selected using the POG-recommended “Cut Based VETO” selection for the Phys14 PU20bx25 pileup scenario [5]. Electron candidates are required to have $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$. The recommended PF relative-isolation criterion with $\delta\beta$ -based corrections and $\Delta R = 0.3$ is used with a requirement of < 0.33 (< 0.38) in the barrel (endcap) region to eliminate events with an isolated electron.

- Angular cut $\Delta\hat{\phi}_{\text{min}} > 4.0$:

We use the resolution-normalized azimuthal separation between jets and the E_T^{miss} vector, denoted $\Delta\hat{\phi}$ [6, 7], to reject QCD events. This variable is less correlated with E_T^{miss} than the separation before normalization, $\Delta\phi_{\text{min}}$, making it useful for performing a data-driven estimate of the QCD background (see Section 6.2). For jet j , we define:

$$\Delta\hat{\phi}_j \equiv \Delta\phi(j, E_T^{\text{miss}}) / \sin^{-1}(\Delta_{Tj} / E_T^{\text{miss}}). \quad (1)$$

Δ_{Tj} is defined by:

$$\Delta_{Tj} = 0.1 \frac{\sqrt{\sum_{i \neq j} [p_x^j p_y^i - p_y^j p_x^i]^2}}{p_T^j}, \quad (2)$$

where the sum runs over all jets i passing the “good” criteria listed above. We calculate $\Delta\hat{\phi}_j$ for the leading three jets and call the minimum value $\Delta\hat{\phi}_{\min}$.

- Isolated track veto: Following the baseline event selection and standard muon and electron event vetoes described above, the dominant source of background is $t\bar{t}$, single top, and W +jets events with one $W \rightarrow l\nu$ decay. To further suppress these lost lepton backgrounds, we reject events that have one or more tracks with $p_T > 15$ GeV, $|\eta| < 2.4$, and relative isolation less than 0.1 within a cone of 0.3. Isolation is calculated from charged PF candidates consistent with the primary vertex ($|dz(PV)| < 0.05$ cm). To retain more signal, and thus improve signal-to-background event discrimination, events with one “isolated track,” as defined above, are only rejected if they satisfy

$$m_T(tk, E_T^{\text{miss}}) = \sqrt{2p_T^{tk} E_T^{\text{miss}} (1 - \cos \Delta\phi)} < 100 \text{ GeV} \quad (3)$$

Where p_T^{tk} is the transverse momentum of the track and $\Delta\phi$ is the azimuthal separation between the track and E_T^{miss} vector. As illustrated in Section 3, a large fraction of the tracks from events with one-and-only-one source of E_T^{miss} , i.e. one $W \rightarrow l\nu$ decay, satisfy this requirement. Placing this m_T restriction on the definition of the track saves up to 30% of $T1tttt$ events while increasing the lost lepton background by less than 10%.

A detailed study of the sources of these lost lepton events, before and after the cut on isolated tracks, is presented in section 6.1.

Table 3: Cutflow and expected yields at 4 fb^{-1} for SM backgrounds after baseline selection. The MC are scaled according to the cross sections listed in Table 1. Note that the large amount of QCD after baseline selection is due in part to one event with a weight of 1005.

Cut	$t\bar{t}$	QCD	Z+jets	W+jets	Single top	Other
Start	3249990	2792038930	1958012	9452080	952000	377
$N_{\text{jet}} \geq 4$	2263404	539767728	67043	761941	197620	22
$H_T > 500 \text{ GeV}$	434376	36841661	10268	92531	29673	3
$H_T^{\text{miss}} > 200 \text{ GeV}$	16202	23871	2724	9622	2399	1
Muon veto	11872	23850	2723	7066	1755	1
Electron veto	7226	23332	2681	4185	1052	1
$\Delta\hat{\phi}_{\min} > 4$	5174	5144	2292	3033	673	0
Iso. track veto	3890 ± 22	4982 ± 1038	2260 ± 4	2219 ± 8	492 ± 8	0 ± 0

5 Search variables and binning

5.1 Nominal analysis: H_T , H_T^{miss} , N_{jet} , and $N_{\text{b-jet}}$

Includes definition of the search and control regions, N-1 plots demonstrating the benefit of considering all four dimensions. Includes presentation of background and signal in the search and control regions (from MC).

Nominal binning to use is as follows:

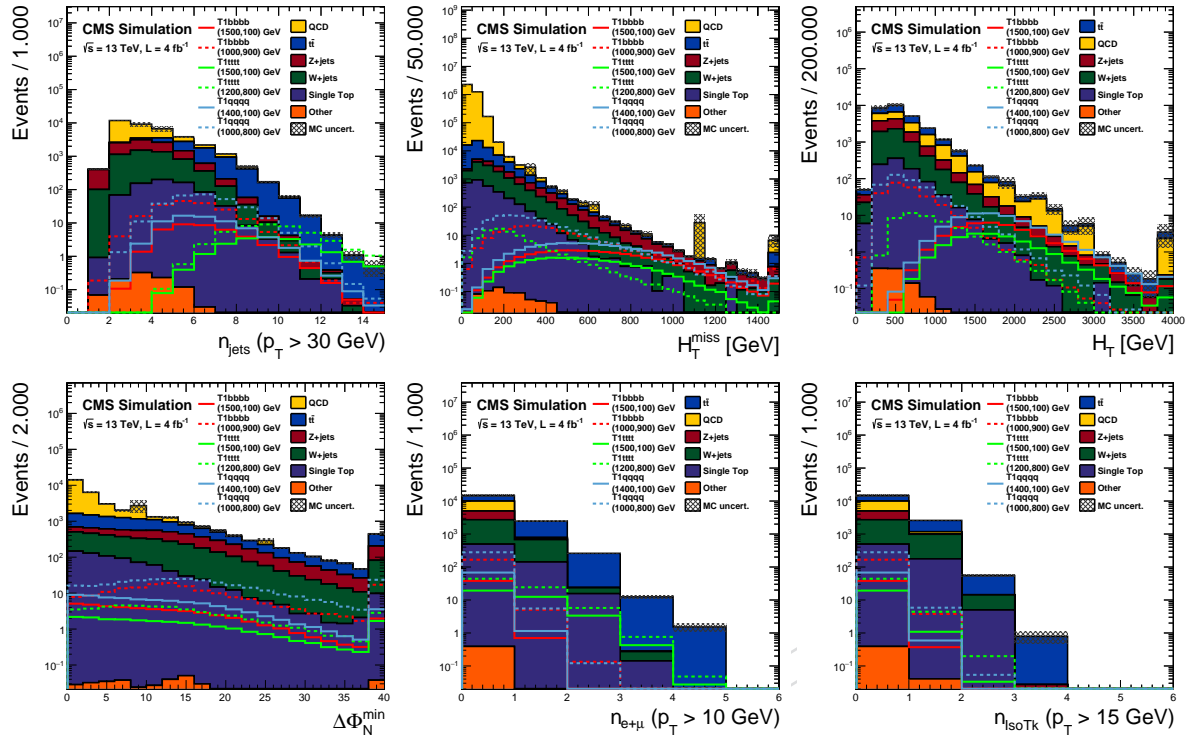


Figure 2: $N - 1$ plots of the variables included in the baseline selection, described above. The SM and signal MC samples are scaled to 4 fb^{-1} .

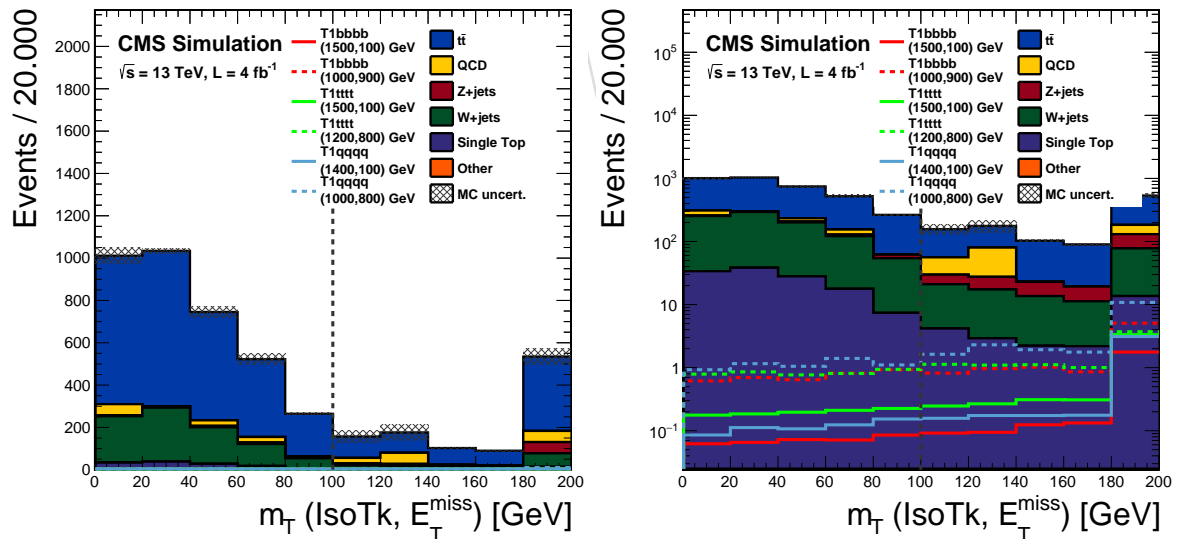


Figure 3: $m_T(tk, E_T^{\text{miss}})$ for signal and background for events passing the baseline selection and containing one isolated track (with no m_T restriction) on a linear (left) and log (right) scale. Tracks with $m_T > 100$ GeV, to the right of the gray dashed line, are not counted as “veto tracks,” so these events would enter our search region. The SM and signal MC samples are scaled to 4 fb^{-1} .

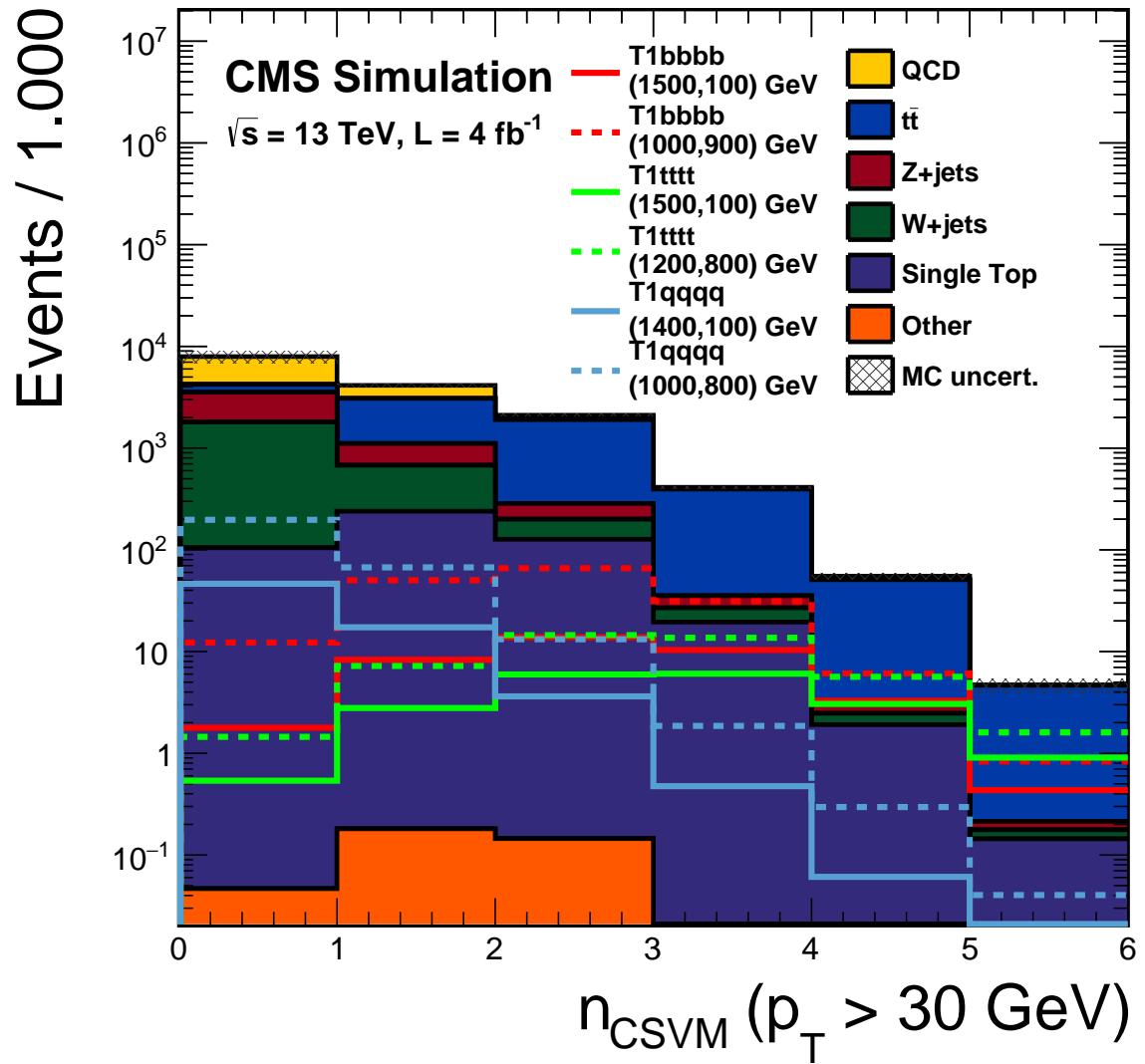


Figure 4: The number of b-tagged jets after the baseline selection. There is no minimum $N_{\text{b-jet}}$ requirement in the baseline selection.

Table 4: Cutflow and expected yields at 4 fb^{-1} for SMS model points after baseline selection. The MC are scaled according to the cross sections listed in Table 2.

Cut	T1tttt (1500, 100)	T1tttt (1200, 800)	T1bbbb (1500, 100)	T1bbbb (1000, 900)	T1qqqq (1400, 100)	T1qqqq (1000, 800)
Start	56.8	342.6	56.8	1301.6	101.2	1301.6
$N_{\text{jet}} \geq 4$	56.7	341.6	54.8	618.4	97.8	997.8
$H_T > 500 \text{ GeV}$	56.7	262.3	54.8	236.4	97.7	503.8
$H_T^{\text{miss}} > 200 \text{ GeV}$	49.9	112.5	49.3	193.5	87.8	325.5
Muon veto	37.6	84.2	49.2	192.4	87.7	325.3
Electron veto	25.1	56.7	48.4	187.2	86.3	318.9
$\Delta\phi_{\text{min}} > 4$	20.4	48.7	38.4	170.4	68.5	285.8
Iso. track veto	19.3	44.1	38.0	166.6	67.9	279.9
Baseline efficiency	0.34	0.13	0.67	0.13	0.67	0.22

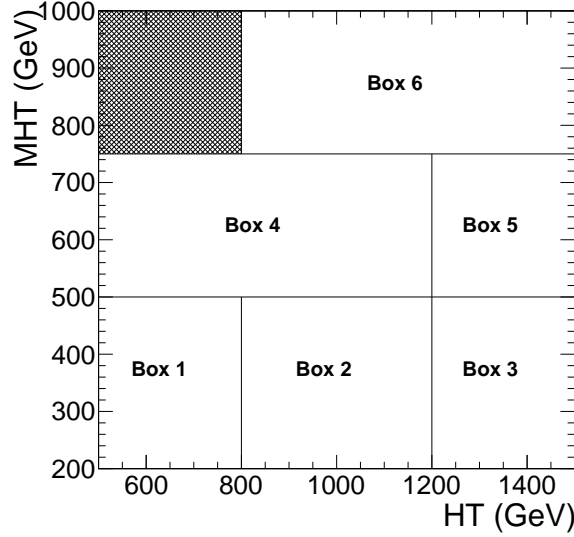


Figure 5: Two-dimensional plane in H_T and H_T^{miss} showing the signal bins. The same H_T and H_T^{miss} regions are used for each N_{jet} and $N_{\text{b-jet}}$ count.

- N_{jet} : 4–6, 7–8, ≥ 9
- $N_{\text{b-jet}}$: 0, 1, 2, ≥ 3
- H_T : 500–800, 800–1200, $\geq 1200 \text{ GeV}$
- H_T^{miss} : 200–500, 500–750, $\geq 750 \text{ GeV}$

The binning is rectangular as above, except the 12 bins with $H_T < 800$ and $H_T^{\text{miss}} > 750$ are removed. Additionally, for $500 < H_T^{\text{miss}} < 750$, a single range with $500 < H_T < 1200$ is used, and for $H_T^{\text{miss}} > 750$, a single range with $H_T > 800$ is used. The merging of these bins further reduces the bin count by 24, for a total of 72 bins. The 6 bins in the H_T and H_T^{miss} plane are shown visually in Fig. 5.

5.2 Alternative search variables: $\sum m_{\text{jet}}$

We are also investigating the possibility of re-clustering jets into larger-radius ($R = 1.2$) jets and using the properties of those jets to discriminate between signal and background. In particular,

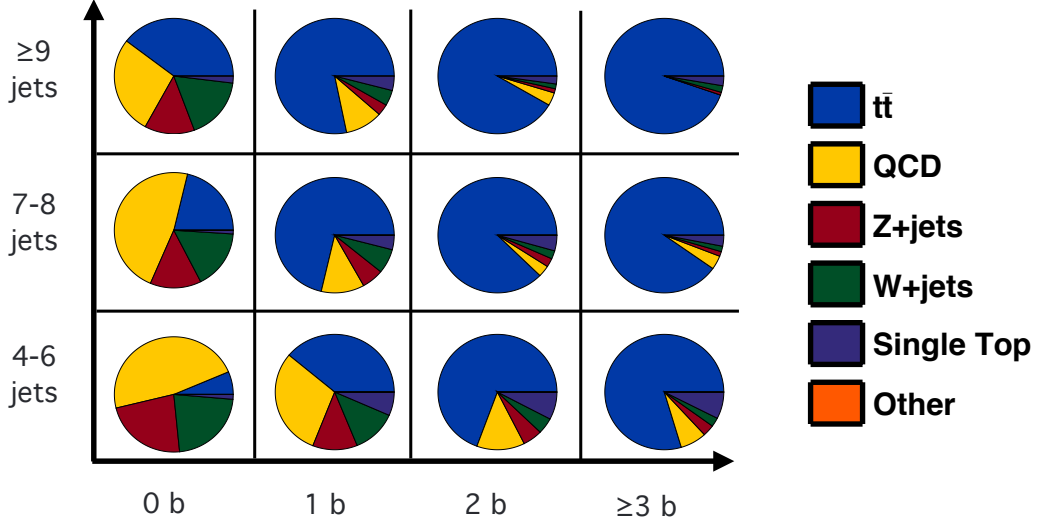


Figure 6: Relative expected contribution from each SM background in bins of N_{jet} and $N_{b\text{-jet}}$, integrated over all $H_T^{\text{miss}} > 200$ GeV and $H_T > 500$ GeV bins.

the scalar sum of the masses of these large radius jets, $\sum m_{\text{jet}}$, has been proposed as an alternative search variable to H_T in searches for high-multiplicity signals [8]. A detailed study of this variable and other topics relevant to large-radius jets is documented in a separate Analysis Note [cite later].

6 Background evaluation

Discussion of data-driven techniques, including the impact of uncertainties on the sensitivity, closure tests, etc., for all the following.

6.1 Top quark and W+jets events

Following the baseline kinematic selection and the rejection of events with reconstructed and identified electrons and muons (described in Chapter 4), about 50% of the expected SM background (integrated over all search bins) comes from $t\bar{t}$ and W+jets events with a lepton that is “lost,” or not identified. This fraction grows significantly (up to 90%) in bins of higher N_{jet} and $N_{b\text{-jet}}$ (Figure 6).

Of the lost-lepton events remaining after the electron and muon event vetoes but before the cut on isolated tracks, about 45% contain hadronically-decaying taus, as illustrated in Figure 7. The isolated track veto cut rejects about 30% of the hadronic tau background and a smaller but significant fraction of the other lost leptons (see Table 5).

Work is ongoing to improve the isolated track definition to treat charged hadron tracks separately from charged lepton tracks, which may help reject additional lost lepton events.

To estimate the amount of events involving electrons and muons which are not rejected by the explicit lepton veto two method are under study. Both take advantage of the amount of isolated electrons and muons.

Table 5: Expected number of $t\bar{t}$ events at 4 fb^{-1} before and after isolated track veto. The events are categorized by the number of type of leptonic W decays.

Cut	$1 W \rightarrow \tau \rightarrow \text{had}$	$1 W \rightarrow \tau \rightarrow e/\mu$	$1 W \rightarrow e/\mu$	$2 W \rightarrow e/\mu/\tau$	hadronic
Before veto	2233	756	1418	446	50
After veto	1573	640	1084	315	47
Efficiency	0.70	0.85	0.76	0.71	0.93

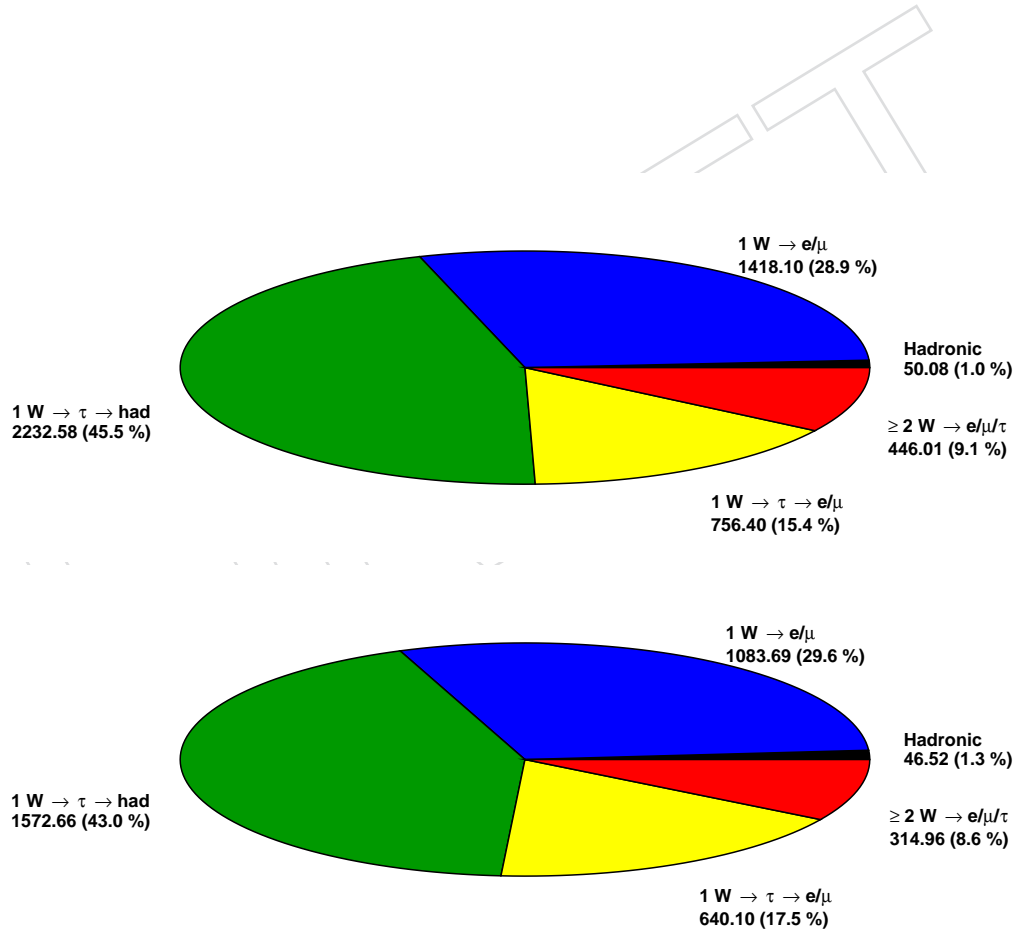


Figure 7: Leptonic composition of $t\bar{t}$ background at truth level, before (above) and after (below) isolated track veto.

6.1.1 Classical Lost-lepton (e/μ) estimation method

Arne

Events including muons or electrons either directly from a W decay or via the intermediate

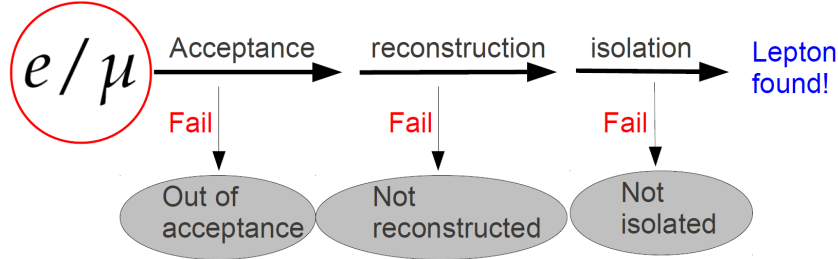


Figure 8: Sketch of the criteria electrons and muons from W decays must meet in order to be rejected by the explicit lepton veto.

stat of a tau may pass the event selection if either of the following steps is not met (see Fig: 8):

- Kinematic acceptance cuts
- Reconstruction criteria
- Isolation definition

The basic concept of this data-driven method is to estimate the amount of electrons and muons failing either of these steps by selecting a control-sample of exactly one well isolated muon or electron and weighting these events according to the lepton efficiencies of each of these steps separately.

In the following the method is discussed starting with the efficiencies determination followed by defining the single lepton control-sample and the computation of the prediction and coverage for the minor contribution of di-leptonic $t\bar{t}$ decays.

Determining the lepton efficiencies

The main crucial input to this method is the choice and the binning of the parametrization variables of the lepton efficiency maps. Extensive studies have been performed and are ongoing in finding and optimizing the variables. Each of the steps (see Fig. 8) has been investigated independently.

The efficiencies have been determined from the $t\bar{t}$ and W+jets MC samples applying the baseline selection without any lepton criteria.

The following parametrization variables have been investigated:

- Search variables: H_T , H_T^{miss} , $N(\text{jets})$ & $B(\text{tags})$
- Lepton p_T
- Activity around the lepton $A_{\mu/e}$ defined as Eq: 4,5

$$A_{\mu} = \sum_{\text{Jet}}^{\Delta R < 1.0} p_T(\text{Jet}) * (\text{ChargedEMFraction} + \text{ChargedHadFraction}) \quad (4)$$

$$A_e = \sum_{\text{Jet}}^{\Delta R < 1.0} p_T(\text{Jet}) * (\text{ChargedHadFraction}) \quad (5)$$

Currently the following efficiency parametrization has been used for electrons and muons:

- Acceptance: H_T^{miss} & $B(\text{tags})$ (2D binning) (see. Fig. 9)

- Reconstruction: p_T and activity around the lepton A (2D binning) (see. Fig. 10)
- Isolation: p_T and activity around the lepton A (2D binning) (see. Fig. 11)
- Purity of the lepton control sample (electron CS only): H_T^{miss} & $N(\text{jets})$ (2D binning) (see Fig. 12)
- m_T -cut efficiency: p_T and activity around the lepton A (2D binning) (see. Fig. 13)
- Dil-leptonic contribution to single lepton control sample: $N(\text{jets})$ (1D binning) (see. Fig. 14)
- Dil-leptonic selection efficiency : $N(\text{jets})$ (1D binning) (see. Fig. 15)

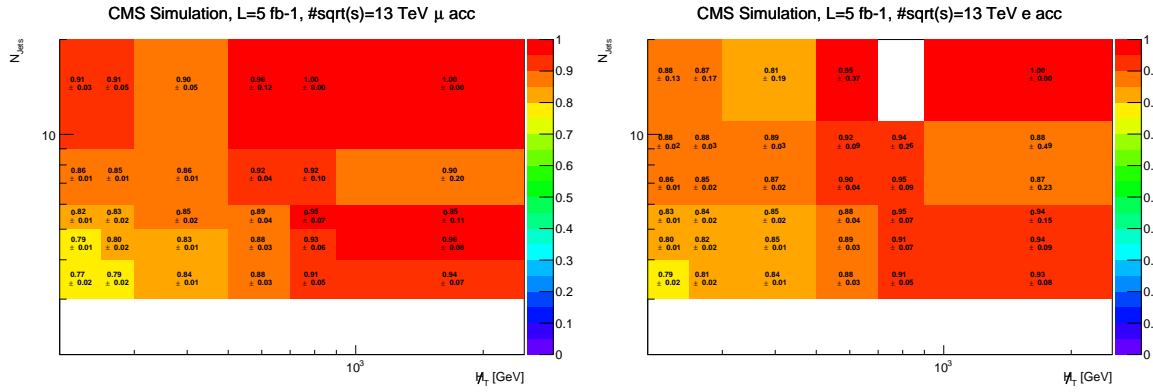


Figure 9: Muon and electron acceptance efficiency maps.

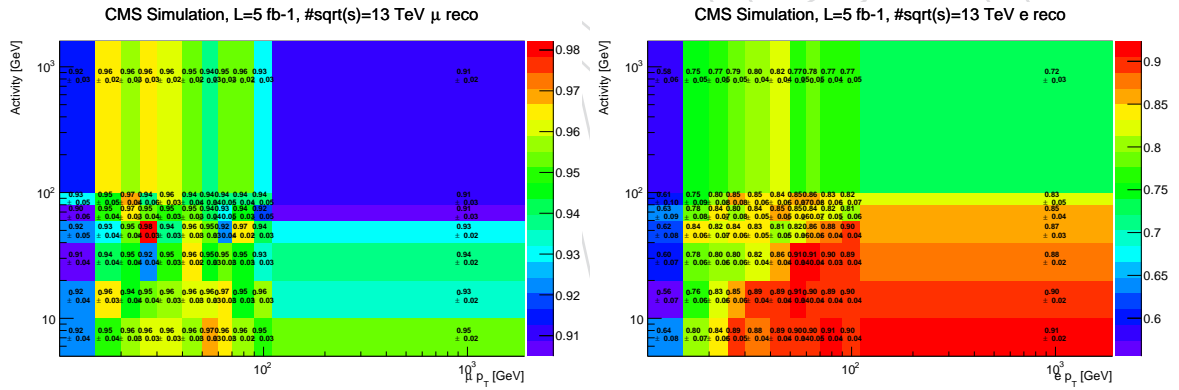


Figure 10: Muon and electron reconstruction efficiency maps.

Selecting the single control-sample

The single electron and muon control sample is selected by requiring exactly one well isolated electron and no isolated muon (or one isolated muon and no isolated electrons) and no isolated tracks. These events originate from a W to lepton and neutrino decay which opens the opportunity to computing the transverse projected mass (Eq. 6) of the W as can be seen in Fig. 16. A cut at 100 GeV rejecting events resulting in a higher m_T value reduces possible signal contamination of the control sample (while preserving about 90 % of single prompt W +jets and $t\bar{t}$ events. Table 6 shows the amount of expected events vs the amount of single μ and e control sample for the search bins.

In addition purity studies have been performed showing for the muonic control sample a purity of 99 % thus no correction for impurity while the single electron control sample shows a significant contamination of non-prompt electrons leading to a purity around 90% for which the control sample is corrected for.

The last correction accounts for di-leptonic events in which one of the prompt leptons is lost.

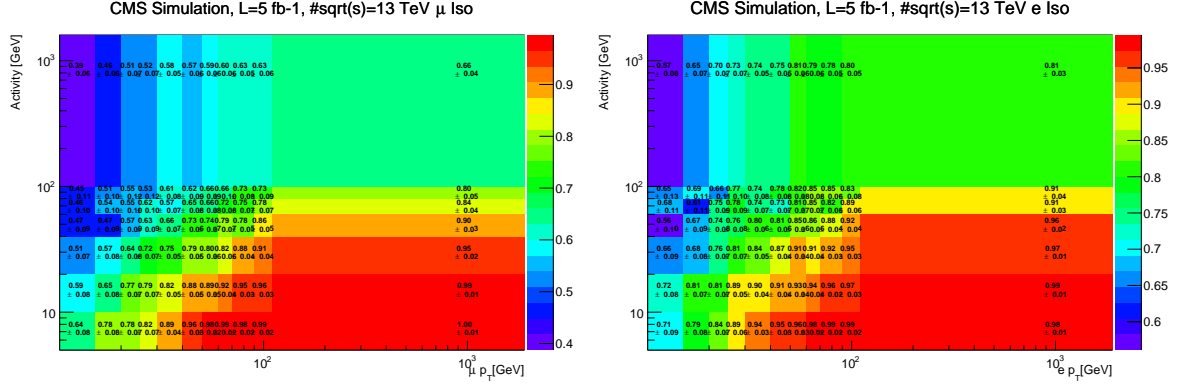


Figure 11: Muon and electron isolation efficiency maps.

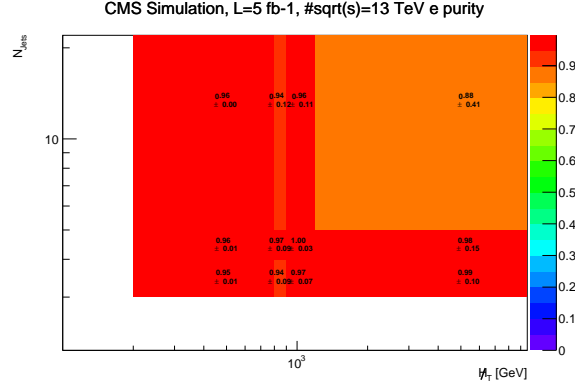


Figure 12: Purity of single electron control sample.

The amount of these events is already significantly reduced by applying $m_T < 100 \text{ GeV}$ but still about 3% of the total single lepton control sample originates from di-leptonic $t\bar{t}$ decays (see Fig. 14). These events need to be evaluated separately since it is by far more likely to lose one than two leptons.

$$m_T = \sqrt{2p_T(\mu)E_T^{\text{miss}}(1 - \cos(\Delta\Phi))} \quad (6)$$

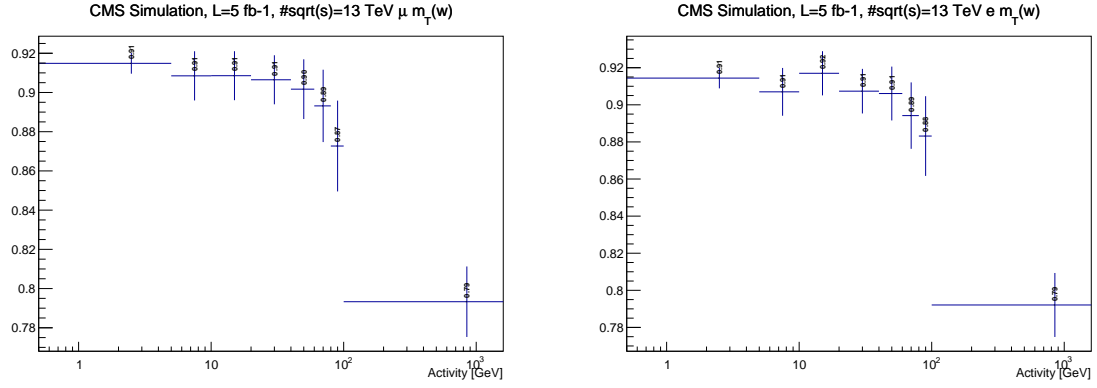
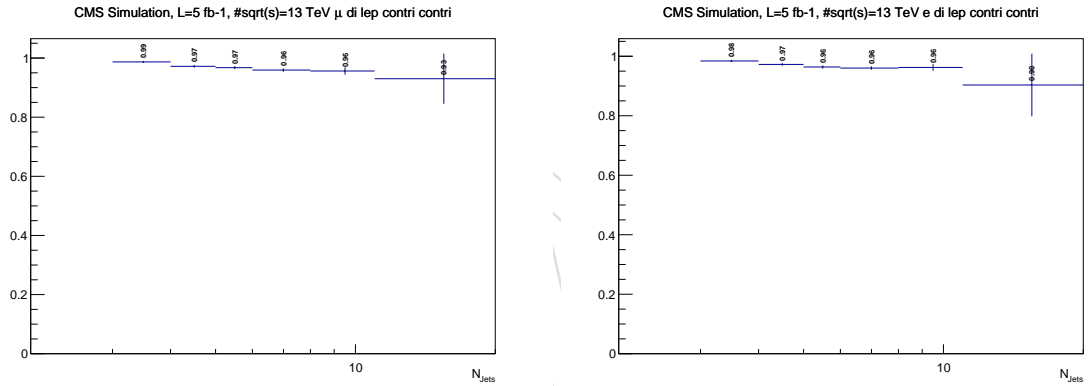
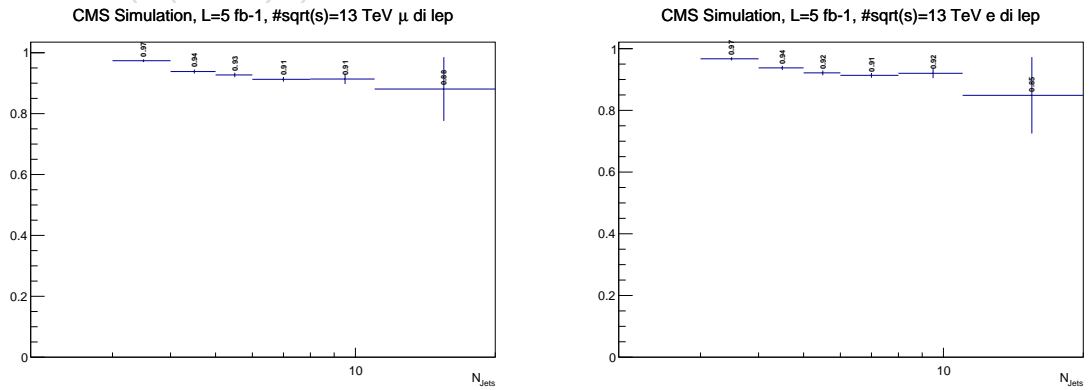
Figure 13: Muon and electron m_T -cut efficiency maps.

Figure 14: Muon and electron di-leptonic contribution to single lepton control sample.

Figure 15: Muon and electron di-leptonic $t\bar{t}$ event rejection efficiency.

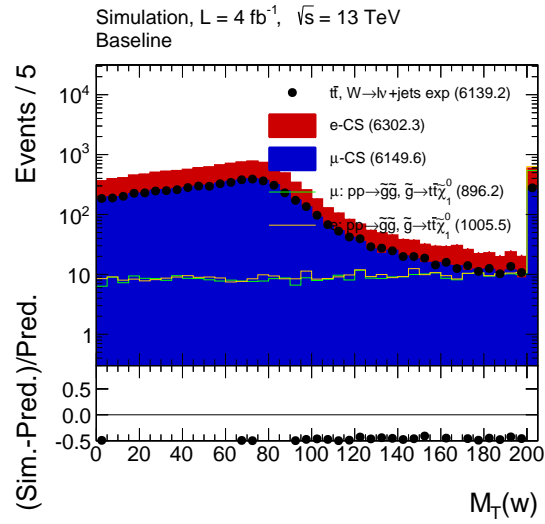


Figure 16: m_T distribution of the single electron and muon control sample selected from W+jets and $t\bar{t}$ events together with a typical signal distribution (here T1tttt increased to 400 fb for visibility purpose).

Selection	MCEX	ElecCS	MuCS	Total Control-Sample
Baseline	6139.2 ± 22.5	5451.4 ± 20.8	5365.3 ± 20.4	10816.7 ± 50.0
B=0, NJet=4-Inf, HT=500-800, MHT=200-500	1375.2 ± 9.0	1358.0 ± 9.1	1372.3 ± 9.1	2730.4 ± 21.9
B=1, NJet=4-6, HT=500-800, MHT=200-500	989.8 ± 10.3	846.6 ± 9.5	836.1 ± 9.4	1682.7 ± 22.9
B ≥, NJet=4-6, HT=500-800, MHT=200-500	97.0 ± 3.5	63.2 ± 2.8	56.9 ± 2.7	120.0 ± 6.7
B=1, NJet=6-Inf, HT=500-800, MHT=200-500	173.6 ± 4.6	176.6 ± 4.7	158.3 ± 4.4	334.8 ± 11.0
B=2, NJet=6-Inf, HT=500-800, MHT=200-500	145.9 ± 4.3	133.6 ± 4.1	135.2 ± 4.1	268.8 ± 9.9
B ≥, NJet=6-Inf, HT=500-800, MHT=200-500	41.2 ± 2.3	34.6 ± 2.1	31.2 ± 2.0	65.8 ± 5.0
B=0, NJet=4-Inf, HT=800-1200, MHT=200-500	346.9 ± 4.0	328.9 ± 3.8	321.5 ± 3.7	650.4 ± 9.1
B=1, NJet=4-6, HT=800-1200, MHT=200-500	241.5 ± 4.9	180.6 ± 4.2	166.1 ± 4.0	346.7 ± 9.9
B ≥, NJet=4-6, HT=800-1200, MHT=200-500	27.0 ± 1.8	15.8 ± 1.4	15.1 ± 1.4	30.9 ± 3.3
B=1, NJet=6-Inf, HT=800-1200, MHT=200-500	105.0 ± 3.5	95.7 ± 3.3	86.7 ± 3.2	182.5 ± 7.9
B=2, NJet=6-Inf, HT=800-1200, MHT=200-500	93.7 ± 3.4	80.7 ± 3.2	76.2 ± 3.1	156.9 ± 7.6
B ≥, NJet=6-Inf, HT=800-1200, MHT=200-500	31.2 ± 2.0	25.2 ± 1.8	20.6 ± 1.6	45.8 ± 4.2
B=0, NJet=4-Inf, HT=1200-Inf, MHT=200-500	77.8 ± 1.8	61.4 ± 1.5	57.8 ± 1.4	119.2 ± 3.6
B=1, NJet=4-6, HT=1200-Inf, MHT=200-500	46.4 ± 2.1	26.1 ± 1.5	22.9 ± 1.4	49.1 ± 3.5
B ≥, NJet=4-6, HT=1200-Inf, MHT=200-500	5.2 ± 0.8	2.4 ± 0.5	1.1 ± 0.3	3.5 ± 1.2
B=1, NJet=6-Inf, HT=1200-Inf, MHT=200-500	29.1 ± 1.8	21.3 ± 1.5	17.6 ± 1.4	38.9 ± 3.6
B=2, NJet=6-Inf, HT=1200-Inf, MHT=200-500	26.4 ± 1.8	16.9 ± 1.4	17.6 ± 1.5	34.5 ± 3.5
B ≥, NJet=6-Inf, HT=1200-Inf, MHT=200-500	7.9 ± 1.0	6.4 ± 0.9	4.7 ± 0.8	11.0 ± 2.1
B=0, NJet=4-Inf, HT=500-800, MHT=500-750	28.7 ± 1.1	40.5 ± 1.0	41.7 ± 1.0	82.2 ± 2.5
B=1, NJet=4-6, HT=500-800, MHT=500-750	10.2 ± 0.9	12.3 ± 0.9	14.3 ± 1.0	26.6 ± 2.3
B ≥, NJet=4-6, HT=500-800, MHT=500-750	0.2 ± 0.1	0.6 ± 0.3	0.2 ± 0.1	0.8 ± 0.5
B=1, NJet=6-Inf, HT=500-800, MHT=500-750	0.6 ± 0.3	0.6 ± 0.3	0.4 ± 0.2	1.0 ± 0.6
B=2, NJet=6-Inf, HT=500-800, MHT=500-750	0.2 ± 0.1	0.4 ± 0.2	0.4 ± 0.2	0.8 ± 0.5
B ≥, NJet=6-Inf, HT=500-800, MHT=500-750	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
B=0, NJet=4-Inf, HT=800-1200, MHT=500-750	28.8 ± 0.9	49.7 ± 1.1	51.3 ± 1.1	101.0 ± 2.6
B=1, NJet=4-6, HT=800-1200, MHT=500-750	15.1 ± 1.2	16.8 ± 1.1	14.1 ± 0.9	30.8 ± 2.5
B ≥, NJet=4-6, HT=800-1200, MHT=500-750	0.7 ± 0.3	0.5 ± 0.2	0.7 ± 0.3	1.1 ± 0.6
B=1, NJet=6-Inf, HT=800-1200, MHT=500-750	2.3 ± 0.5	2.4 ± 0.4	2.1 ± 0.4	4.5 ± 1.1
B=2, NJet=6-Inf, HT=800-1200, MHT=500-750	1.7 ± 0.5	1.7 ± 0.4	2.1 ± 0.5	3.8 ± 1.1
B ≥, NJet=6-Inf, HT=800-1200, MHT=500-750	0.3 ± 0.2	0.4 ± 0.2	0.2 ± 0.1	0.6 ± 0.5
B=0, NJet=4-Inf, HT=1200-Inf, MHT=500-750	13.0 ± 0.8	14.0 ± 0.6	13.4 ± 0.6	27.4 ± 1.4
B=1, NJet=4-6, HT=1200-Inf, MHT=500-750	7.1 ± 0.8	5.4 ± 0.6	4.8 ± 0.6	10.3 ± 1.5
B ≥, NJet=4-6, HT=1200-Inf, MHT=500-750	0.6 ± 0.3	0.3 ± 0.2	0.0 ± 0.0	0.3 ± 0.4
B=1, NJet=6-Inf, HT=1200-Inf, MHT=500-750	3.4 ± 0.6	3.1 ± 0.6	2.7 ± 0.5	5.9 ± 1.3
B=2, NJet=6-Inf, HT=1200-Inf, MHT=500-750	2.0 ± 0.5	1.8 ± 0.5	1.8 ± 0.5	3.6 ± 1.1
B ≥, NJet=6-Inf, HT=1200-Inf, MHT=500-750	1.2 ± 0.4	0.4 ± 0.2	0.3 ± 0.2	0.6 ± 0.5
B=0, NJet=4-Inf, HT=0800-Inf, MHT=750-Inf	8.1 ± 0.4	17.4 ± 0.6	17.8 ± 0.6	35.2 ± 1.4
B=1, NJet=4-Inf, HT=0800-Inf, MHT=750-Inf	4.1 ± 0.5	4.8 ± 0.4	4.2 ± 0.3	9.0 ± 0.9
B=2, NJet=4-Inf, HT=0800-Inf, MHT=750-Inf	1.2 ± 0.3	1.3 ± 0.3	1.1 ± 0.3	2.4 ± 0.7
B ≥, NJet=4-Inf, HT=0800-Inf, MHT=750-Inf	0.3 ± 0.2	0.0 ± 0.0	0.4 ± 0.2	0.4 ± 0.2

Table 6: Expected events vs amount of expected single electron and muon control sample for the exclusive search regions.

Predicting the single and di-leptonic $t\bar{t}$ and W+jets background

After selecting and correcting the single electron and muon control sample weights are computed according to the efficiencies event by event .

To model the not isolated muons the muon control sample is weighted according to:

$$!ISO^\mu = CS \cdot \frac{1 - \epsilon_{ISO}^\mu}{\epsilon_{ISO}^\mu} \quad (7)$$

with $!ISO^{\text{lepton}}$ being the applied weight and $\epsilon_X^{\mu/e}$ being the efficiency of the criteria X for the lepton.

The next step is to model the not reconstructed muons which is done by taking the muon isolation and reconstruction (in)efficiency into account. The weight is calculated according to the following equation:

$$!Reco^\mu = CS \cdot \frac{1}{\epsilon_{ISO}^\mu} \cdot \frac{1 - \epsilon_{Reco}^\mu}{\epsilon_{Reco}^\mu} \quad (8)$$

The last step in modeling the lost muons is to take also the muons into account which fall out of the detector acceptance. The weight has to include the isolation and reconstruction

(in)efficiencies together with the out of acceptance efficiency:

$$!Acc^\mu = CS \cdot \frac{1}{\epsilon_{ISO}^\mu} \cdot \frac{1}{\epsilon_{Reco}^\mu} \cdot \frac{1 - \epsilon_{Acc}^\mu}{\epsilon_{Acc}^\mu} \quad (9)$$

The electrons are modeled using the same muon CS. This is valid since the decay of a W to a μ or e has according to the lepton universality the same probability. All muon (in)efficiencies need to be taken into account before modeling the lost electrons resulting in more complex equations. The following equations are used to model the not isolated (Eq. 10), not reconstructed (Eq. 11) and the out of acceptance electrons (Eq. 12).

$$!ISO^e = CS \cdot \frac{1 - \epsilon_{ISO}^e}{\epsilon_{ISO}^\mu} \cdot \frac{\epsilon_{Reco}^e}{\epsilon_{Reco}^\mu} \cdot \frac{\epsilon_{Acc}^e}{\epsilon_{Acc}^\mu} \quad (10)$$

$$!Reco^e = CS \cdot \frac{1}{\epsilon_{ISO}^\mu} \cdot \frac{1 - \epsilon_{Reco}^e}{\epsilon_{Reco}^\mu} \cdot \frac{\epsilon_{Acc}^e}{\epsilon_{Acc}^\mu} \quad (11)$$

$$!Acc^e = CS \cdot \frac{1}{\epsilon_{ISO}^\mu} \cdot \frac{1}{\epsilon_{Reco}^\mu} \cdot \frac{1 - \epsilon_{Acc}^e}{\epsilon_{Acc}^\mu} \quad (12)$$

The final equation to predict all lost leptons is:

$$\text{Total Lost Leptons} = C \cdot \sum_{i=e,\mu} (!Iso^i + !Reco^i + !Acc^i) \quad (13)$$

with C being a correction factor consisting of the correction factor for the m_T cut efficiency and a small factor for di-leptonic events of the order of 1% of the total prediction (including for the electron control sample the purity correction).

The same is done for the single electron control sample.

Finally the two independent predictions are combined to predict the total amount of lost electrons and muons. A closure test of the full method can be seen in Fig. 17 for the baseline selection. Overall good agreement can be observed for all search variables. In addition a closure test for each of the search bins can be seen in Tab. 7.

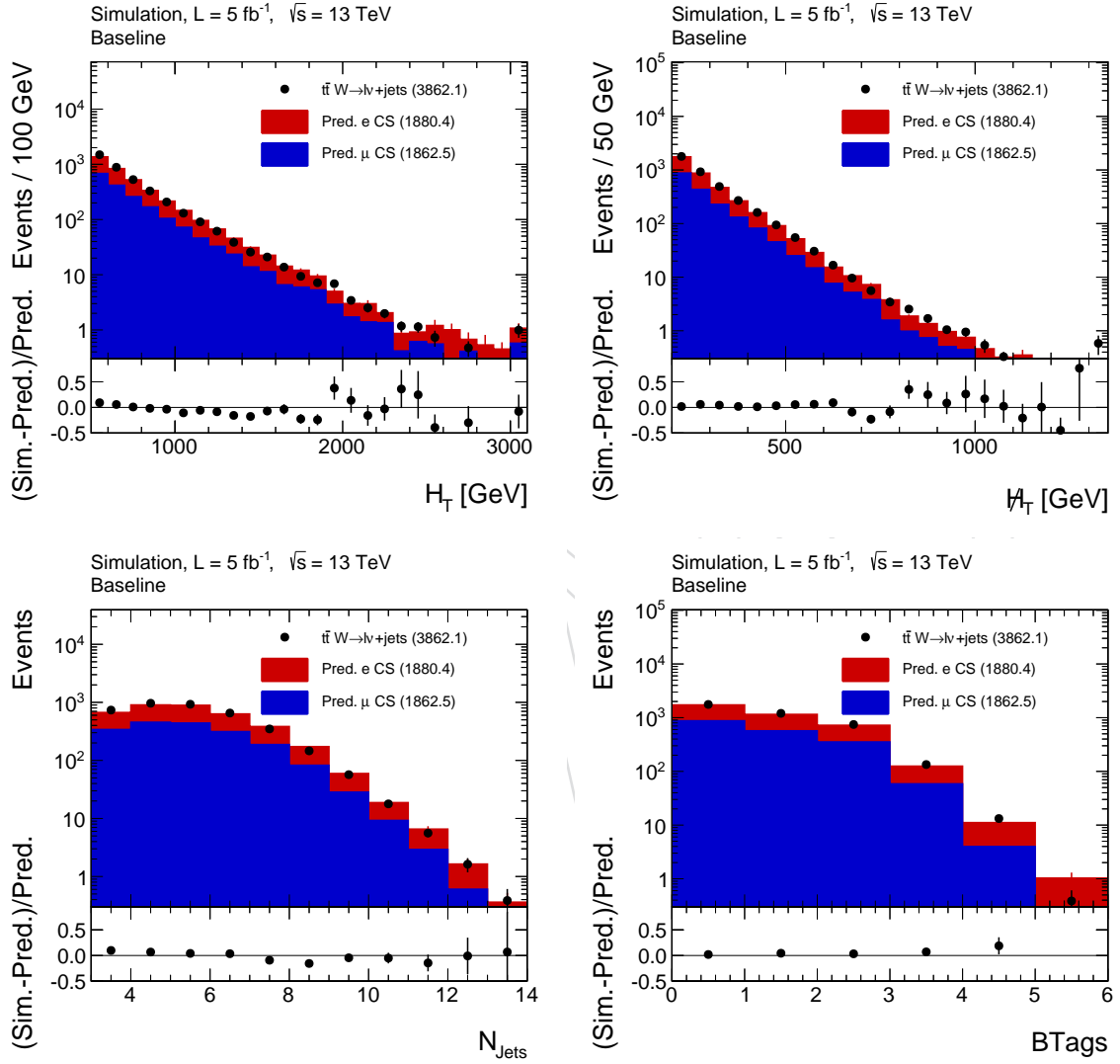


Figure 17: Comparison of the expected and predicted lost electrons and muons using both the single electron and muon control sample. Overall good agreement can be observed for all search variables.

Selection	MCEX	ElecPr	MuPr	Total MC prediction
Baseline	6139.2 \pm 22.5	3036.0 \pm 14.8	2997.3 \pm 14.8	6033.3 \pm 35.8
B=0, NJet=4-Inf, HT=500-800, MHT=200-500	1375.2 \pm 9.0	667.7 \pm 5.4	673.8 \pm 5.3	1341.5 \pm 13.0
B=1, NJet=4-6, HT=500-800, MHT=200-500	989.8 \pm 10.3	485.4 \pm 6.7	481.2 \pm 6.7	966.6 \pm 16.2
B \geq , NJet=4-6, HT=500-800, MHT=200-500	97.0 \pm 3.5	41.4 \pm 2.2	39.8 \pm 2.2	81.1 \pm 5.3
B=1, NJet=6-Inf, HT=500-800, MHT=200-500	173.6 \pm 4.6	93.0 \pm 2.9	83.7 \pm 2.8	176.8 \pm 6.9
B=2, NJet=6-Inf, HT=500-800, MHT=200-500	145.9 \pm 4.3	73.5 \pm 2.6	76.7 \pm 2.8	150.3 \pm 6.5
B \geq , NJet=6-Inf, HT=500-800, MHT=200-500	41.2 \pm 2.3	19.9 \pm 1.4	17.7 \pm 1.3	37.7 \pm 3.3
B=0, NJet=4-Inf, HT=800-1200, MHT=200-500	346.9 \pm 4.0	195.6 \pm 3.0	190.5 \pm 2.9	386.1 \pm 7.2
B=1, NJet=4-6, HT=800-1200, MHT=200-500	241.5 \pm 4.9	132.0 \pm 3.8	122.3 \pm 3.7	254.2 \pm 9.1
B \geq , NJet=4-6, HT=800-1200, MHT=200-500	27.0 \pm 1.8	13.1 \pm 1.3	12.8 \pm 1.4	25.9 \pm 3.2
B=1, NJet=6-Inf, HT=800-1200, MHT=200-500	105.0 \pm 3.5	62.1 \pm 2.6	56.7 \pm 2.5	118.9 \pm 6.1
B=2, NJet=6-Inf, HT=800-1200, MHT=200-500	93.7 \pm 3.4	53.9 \pm 2.4	52.7 \pm 2.5	106.7 \pm 5.9
B \geq , NJet=6-Inf, HT=800-1200, MHT=200-500	31.2 \pm 2.0	16.7 \pm 1.4	14.1 \pm 1.3	30.8 \pm 3.2
B=0, NJet=4-Inf, HT=1200-Inf, MHT=200-500	77.8 \pm 1.8	42.5 \pm 1.4	39.6 \pm 1.2	82.1 \pm 3.2
B=1, NJet=4-6, HT=1200-Inf, MHT=200-500	46.4 \pm 2.1	21.5 \pm 1.5	20.3 \pm 1.5	41.8 \pm 3.7
B \geq , NJet=4-6, HT=1200-Inf, MHT=200-500	5.2 \pm 0.8	2.3 \pm 0.6	0.8 \pm 0.3	3.1 \pm 1.3
B=1, NJet=6-Inf, HT=1200-Inf, MHT=200-500	29.1 \pm 1.8	15.4 \pm 1.3	14.2 \pm 1.3	29.6 \pm 3.2
B=2, NJet=6-Inf, HT=1200-Inf, MHT=200-500	26.4 \pm 1.8	13.7 \pm 1.3	14.0 \pm 1.3	27.8 \pm 3.2
B \geq , NJet=6-Inf, HT=1200-Inf, MHT=200-500	7.9 \pm 1.0	5.4 \pm 0.9	3.9 \pm 0.7	9.3 \pm 2.0
B=0, NJet=4-Inf, HT=500-800, MHT=500-750	28.7 \pm 1.1	13.0 \pm 0.5	12.7 \pm 0.4	25.6 \pm 1.1
B=1, NJet=4-6, HT=500-800, MHT=500-750	10.2 \pm 0.9	4.6 \pm 0.5	5.7 \pm 0.6	10.3 \pm 1.2
B \geq , NJet=4-6, HT=500-800, MHT=500-750	0.2 \pm 0.1	0.3 \pm 0.2	0.1 \pm 0.0	0.4 \pm 0.4
B=1, NJet=6-Inf, HT=500-800, MHT=500-750	0.6 \pm 0.3	0.2 \pm 0.1	0.1 \pm 0.0	0.3 \pm 0.2
B=2, NJet=6-Inf, HT=500-800, MHT=500-750	0.2 \pm 0.1	0.2 \pm 0.1	0.1 \pm 0.1	0.3 \pm 0.2
B \geq , NJet=6-Inf, HT=500-800, MHT=500-750	0.0 \pm 0.0	0.0 \pm 0.0	0.0 \pm 0.0	0.0 \pm 0.0
B=0, NJet=4-Inf, HT=800-1200, MHT=500-750	28.8 \pm 0.9	16.9 \pm 0.5	17.6 \pm 0.5	34.5 \pm 1.2
B=1, NJet=4-6, HT=800-1200, MHT=500-750	15.1 \pm 1.2	6.8 \pm 0.6	6.2 \pm 0.6	13.0 \pm 1.4
B \geq , NJet=4-6, HT=800-1200, MHT=500-750	0.7 \pm 0.3	0.3 \pm 0.1	0.4 \pm 0.2	0.7 \pm 0.4
B=1, NJet=6-Inf, HT=800-1200, MHT=500-750	2.3 \pm 0.5	1.0 \pm 0.2	0.9 \pm 0.2	1.9 \pm 0.6
B=2, NJet=6-Inf, HT=800-1200, MHT=500-750	1.7 \pm 0.5	0.6 \pm 0.2	1.1 \pm 0.3	1.7 \pm 0.5
B \geq , NJet=6-Inf, HT=800-1200, MHT=500-750	0.3 \pm 0.2	0.3 \pm 0.1	0.1 \pm 0.0	0.3 \pm 0.3
B=0, NJet=4-Inf, HT=1200-Inf, MHT=500-750	13.0 \pm 0.8	7.0 \pm 0.5	7.0 \pm 0.5	14.1 \pm 1.1
B=1, NJet=4-6, HT=1200-Inf, MHT=500-750	7.1 \pm 0.8	3.9 \pm 0.6	3.4 \pm 0.6	7.3 \pm 1.5
B \geq , NJet=4-6, HT=1200-Inf, MHT=500-750	0.6 \pm 0.3	0.3 \pm 0.2	0.0 \pm 0.0	0.3 \pm 0.4
B=1, NJet=6-Inf, HT=1200-Inf, MHT=500-750	3.4 \pm 0.6	2.5 \pm 0.6	2.0 \pm 0.5	4.5 \pm 1.3
B=2, NJet=6-Inf, HT=1200-Inf, MHT=500-750	2.0 \pm 0.5	1.6 \pm 0.5	1.1 \pm 0.3	2.7 \pm 1.1
B \geq , NJet=6-Inf, HT=1200-Inf, MHT=500-750	1.2 \pm 0.4	0.5 \pm 0.3	0.2 \pm 0.1	0.7 \pm 0.6
B=0, NJet=4-Inf, HT=0800-Inf, MHT=750-Inf	8.1 \pm 0.4	4.7 \pm 0.2	4.7 \pm 0.2	9.4 \pm 0.5
B=1, NJet=4-Inf, HT=0800-Inf, MHT=750-Inf	4.1 \pm 0.5	1.4 \pm 0.2	1.1 \pm 0.1	2.5 \pm 0.4
B=2, NJet=4-Inf, HT=0800-Inf, MHT=750-Inf	1.2 \pm 0.3	0.6 \pm 0.2	0.4 \pm 0.2	1.0 \pm 0.5
B \geq , NJet=4-Inf, HT=0800-Inf, MHT=750-Inf	0.3 \pm 0.2	0.0 \pm 0.0	0.3 \pm 0.2	0.3 \pm 0.2

Table 7: Comparison of prediction from single electron and muon control sample vs expected lost-leptons in exclusive search regions. Overall reasonable agreement between expectation and combined prediction can be observed.

6.1.2 Hadronically decaying τ estimation

Ahmad

6.1.3 Extrapolating background estimates from medium to high H_T^{miss}

Owen

The lost-lepton method described in Sec.6.1.1 relies on a control sample to search sample ratio of about 1 to 1. For our most sensitive search bins, with background below a few events, and relatively good signal efficiency, the statistical uncertainty on the lost-lepton background estimate can be substantial and sometimes the dominant background uncertainty. The expected exclusion sensitivity (i.e. the expected upper limit on the signal cross section if the signal does not exist) does not strongly depend on the statistical uncertainty on the background in these very low background bins. However, reducing the statistical uncertainty on the lost-lepton background estimate can substantially improve the discovery sensitivity (i.e. the significance of an excess due to a real signal), since a smaller background uncertainty reduces the likelihood of the null hypothesis if an excess is observed. This is the motivation for the method described in this subsection.

The lost-lepton H_T^{miss} spectrum depends on the underlying p_T spectrum of the leptonically decaying W and the angular distribution of the W decay. The W p_T spectrum depends on the top quark p_T spectrum, for which the Monte Carlo is less reliable than the angular distribution of the W decay, which is well understood (see Sec. 7 of [7] and [9]). The H_T^{miss} and W p_T are related by

$$H_T^{\text{miss}} \approx p_T^W (1 - \cos \Delta\theta_T) / 2 \quad (14)$$

where $\Delta\theta_T$ is the angle between the direction of the W in the lab frame and the direction of the charged lepton in the W rest frame, all in the transverse plane. The W p_T can be reconstructed in single lepton control sample events, which provides a data control sample for the W p_T spectrum. We use Monte Carlo distributions of $(1 - \cos \Delta\theta_T) / 2$ for lost leptons to translate the observed W p_T spectrum into a lost lepton H_T^{miss} spectrum. This lost lepton H_T^{miss} spectrum is then used to estimate a high-to-medium H_T^{miss} ratio R^{MHT} , which can be used to extrapolate a lost-lepton background estimate at medium H_T^{miss} , from the method described in Sec.6.1.1, to high H_T^{miss} , thus reducing the statistical uncertainty of the lost-lepton background estimate in the most sensitive search bins.

Figure 18 shows an example single lepton control sample H_T^{miss} vs p_T^W distribution. The statistical uncertainty on the H_T^{miss} extrapolation ratio R^{MHT} is mostly determined by the number of single lepton control sample events with $p_T^W > H_{T^{\text{miss}}}^{\text{low}}$, where $H_{T^{\text{miss}}}^{\text{low}}$ is the low edge of the high- H_T^{miss} search bin (see App.C for details). For example, the extrapolation ratio R^{MHT} going from H_T^{miss} bin 1 to H_T^{miss} bin 2, which is from 500 to 750 GeV, uses all single lepton control sample events with $p_T^W > 500$ GeV, while the standard lost-lepton method only uses single-lepton events with H_T^{miss} between 500 to 750 GeV to predict the background for H_T^{miss} bin 2. Figure 18 illustrates the source of the improvement in control sample statistics, which is typically more than a factor of four.

Figure 19 shows the $(1 - \cos \Delta\theta_T) / 2$ distribution of lost lepton events, including the separate contributions from $W \rightarrow e, \mu$, $W \rightarrow \tau \rightarrow e, \mu$, and $W \rightarrow \tau \rightarrow \text{hadrons}$. When the lepton is emitted in the direction opposite the W p_T , $\Delta\theta_T$ is close to π , the lepton p_T is soft, and $(1 - \cos \Delta\theta_T) / 2$ is close to 1. Figure 19 also shows the ratio of lost lepton events to single lepton control sample events R^{LC} as a function of the reconstructed W p_T . This isn't flat, indicating that lost lepton and control sample events have different W p_T spectra, so this must be taken into account. Both the $(1 - \cos \Delta\theta_T) / 2$ and R^{LC} distributions have some dependence on N_{jet} and H_T . This is taken into account by using separate distributions for each N_{jet} and H_T bin combination, as shown in Appendix C.

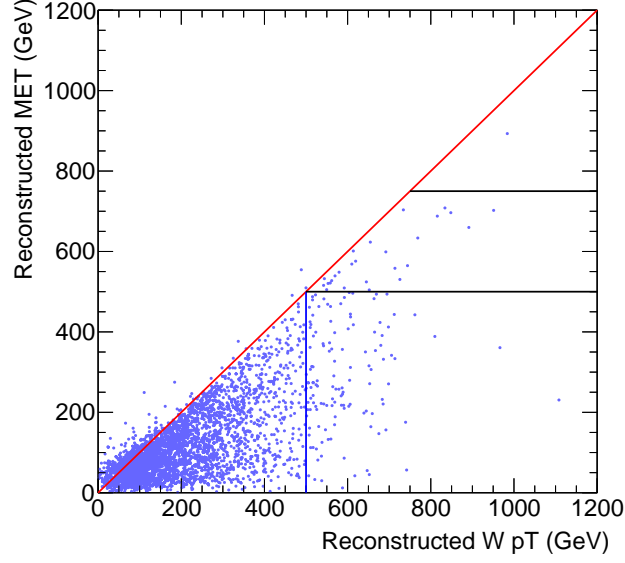


Figure 18: Example H_T^{miss} vs W p_T distribution for single-lepton control sample events from the $t\bar{t}$ Monte Carlo with $H_T > 1200$ GeV, $N_{\text{jet}} \geq 7$, and ≥ 3 . The horizontal lines define H_T^{miss} bin 2 (500 to 750). The H_T^{miss} extrapolation method uses all events above the vertical line with $p_T^W > 500$ GeV in the estimation of the H_T^{miss} spectrum above 500 GeV.

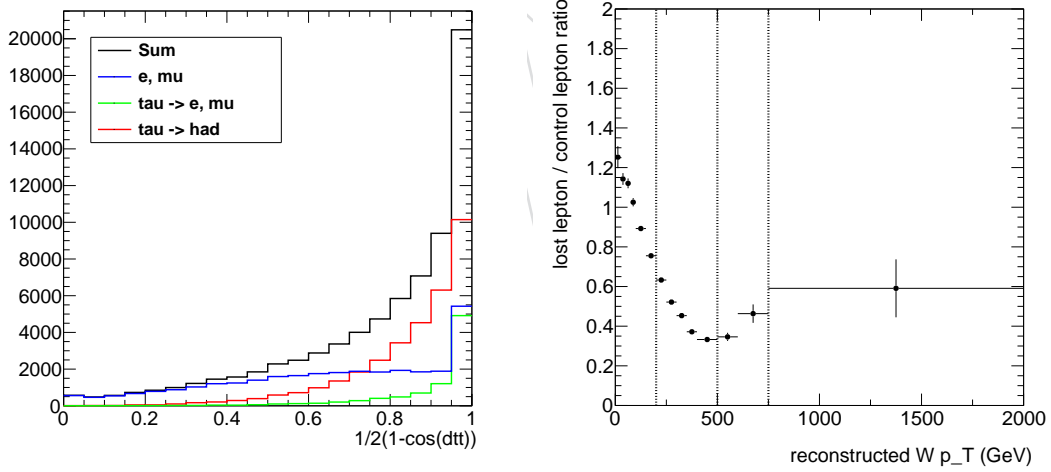


Figure 19: Distribution of $(1 - \cos \Delta\theta_T) / 2$ (left) for lost leptons and the ratio of lost lepton events / control lepton events as a function of p_T^W from the $t\bar{t}$ Monte Carlo. The left figure is summed over all search bins and shows the contributions from $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ (blue), $W \rightarrow \tau\nu$ with $\tau \rightarrow e$ or μ (green), and $W \rightarrow \tau\nu$ with $\tau \rightarrow$ hadrons (red). The right plot is an example distribution corresponding to events in N_{jet} bin 1, H_T bin 2. The vertical dashed lines in the right plot show the location of the H_T^{miss} bin boundaries. The points below 200 GeV are not used in the analysis.

242 The high / medium H_T^{miss} ratios for each N_{jet} , H_T bin combination are derived from H_T^{miss} dis-
 243 tributions constructed from the single lepton control sample, where each event makes a con-
 244 tribution from the corresponding $(1 - \cos \Delta\theta_T) / 2$ distribution weighted by the lost / control
 245 ratio R^{LC} . The resulting H_T^{miss} distributions are compared with the Monte Carlo lost lepton

246 H_T^{miss} distributions in Fig.20. The agreement is reasonable.

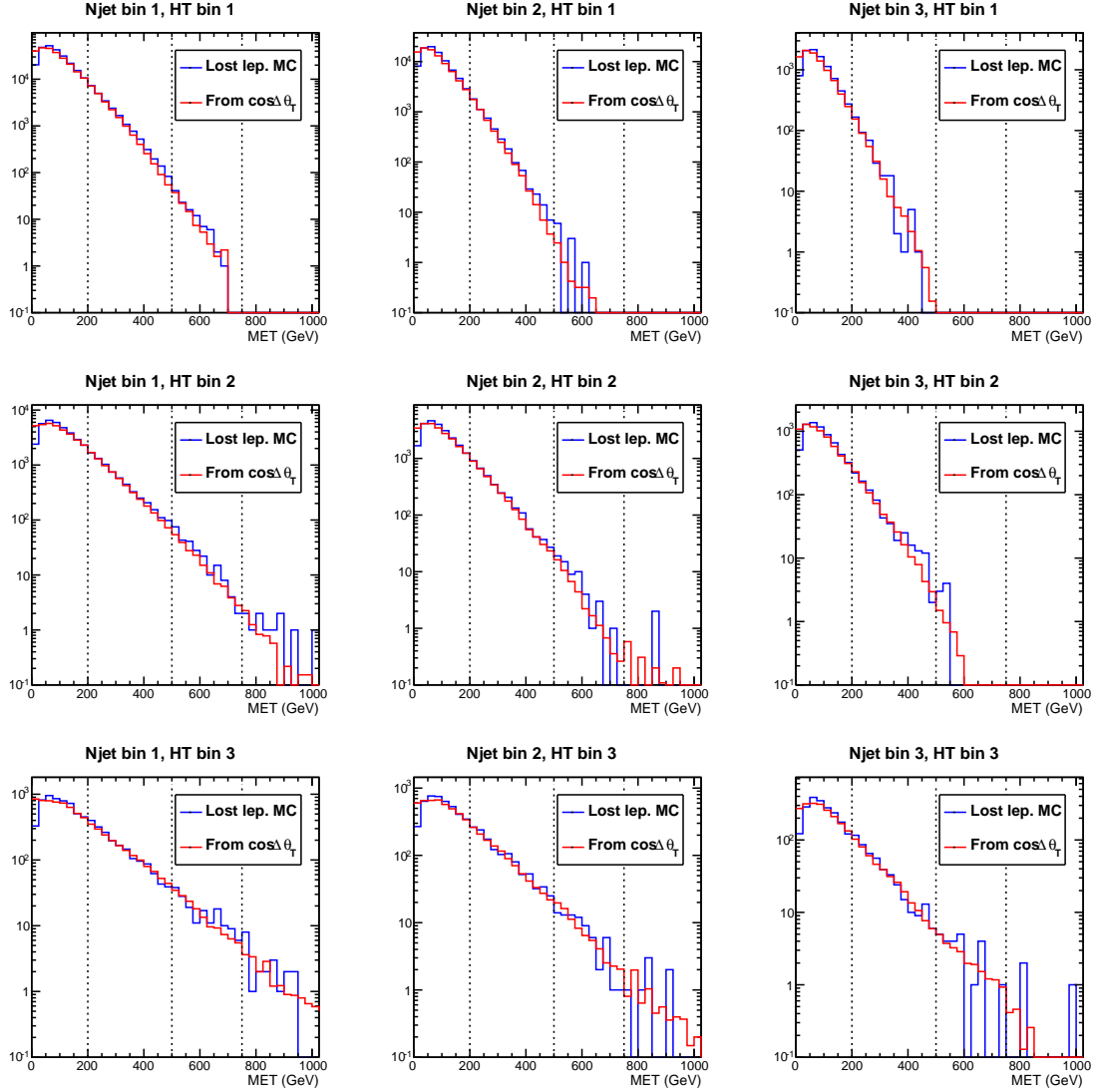


Figure 20: Distributions of H_T^{miss} for the 9 N_{jet} , H_T analysis bin combinations for the $t\bar{t}$ Monte Carlo sample. The blue distributions are from the lost lepton Monte Carlo, while the red distributions are built from the single lepton control sample, applying the method described in this section.

247 Figure 21 compares the predicted number of $t\bar{t}$ lost lepton background events with the observed
 248 number for all of the search bins for H_T^{miss} bins 2 and 3, where the extrapolation was done from
 249 H_T^{miss} bin 1. The agreement is reasonable. The difference in error bars is an indication of the
 250 improvement in the statistical uncertainty of the lost lepton background estimate. More details
 251 on the H_T^{miss} extrapolation method, including how the statistical errors from the single lepton
 252 control sample are propagated to the BG predictions, are given in Appendix C.

253 This method includes both events with a single $W \rightarrow e, \mu$ decay and events with a $W \rightarrow \tau$
 254 decay, including hadronic tau decays. The hadronic tau events contribute to both the lost lepton
 255 $(1 - \cos \Delta\theta_T) / 2$ distribution (Fig.19, left) and the numerator of the lost / control ratio R^{LC}
 256 (Fig.19, right) but are essentially absent from the single lepton control sample.

257 The $(1 - \cos \Delta\theta_T) / 2$ distributions and lost / control ratio R^{LC} as a function of p_T^W are taken

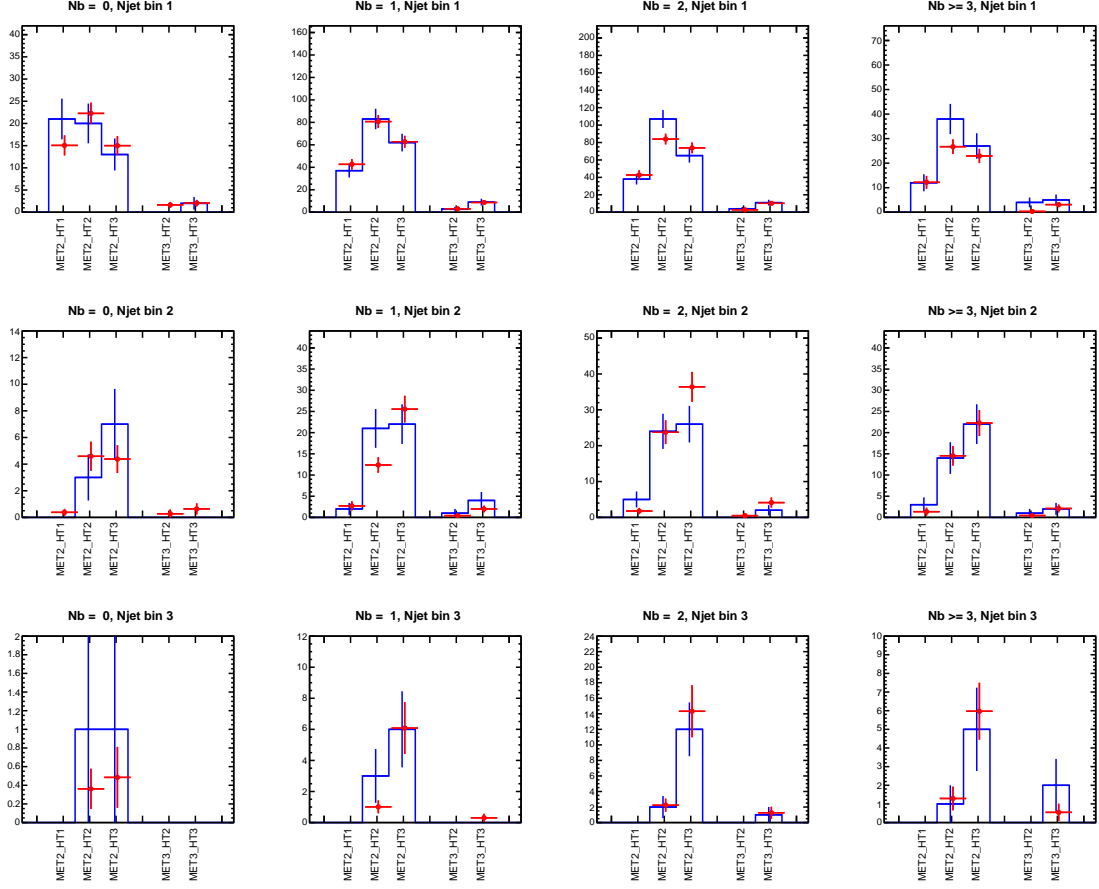


Figure 21: Number of $t\bar{t}$ lost lepton background events in the analysis search bins from the Monte Carlo (blue) and the H_T^{miss} extrapolation method (red). The equivalent integrated luminosity of the $t\bar{t}$ Monte Carlo sample is about 31 fb^{-1} .

from Monte Carlo and will depend on the lepton acceptance, identification, and isolation efficiency. These will be corrected, based on data calibrations, and varied in the evaluation of the systematic uncertainties. We do not anticipate these systematic uncertainties to be larger than the statistical uncertainties in the most sensitive search bins. The material shown in this section is based on the $t\bar{t}$ Monte Carlo, though it will be extended to include the other lost lepton sources: W +jets and single top. No trigger requirement was imposed for the material shown in this section. We expect to use the single lepton events passing the main HT-MET cross trigger with a $H_T^{\text{miss}} > 200 \text{ GeV}$ requirement. This will lower the control sample statistics, but by a relatively small amount for the high p_T^W region which matters most. The studies in this section were done using E_T^{miss} instead of H_T^{miss} , though there is a good agreement between the two for events with $p_T^W > 200 \text{ GeV}$. We will soon switch over to H_T^{miss} .

6.2 QCD multijet events

Owen, Pawandee, Jeff, Jack

We are considering two techniques for estimating the QCD background: using a low- $\Delta\hat{\phi}_{\min}$ control sample as in [1, 10] and using the rebalance-and-smear technique as in [2, 11]. The low- $\Delta\hat{\phi}_{\min}$ method is simpler to implement, so we plan to use it in this first version of the analysis. We will not discuss the rebalance-and-smear method in this note.

The H_T^{miss} in QCD multijet events is almost always due to a single mismeasured jet in the event, thus the H_T^{miss} direction is usually close to a jet. The $\Delta\hat{\phi}_{\min}$ variable is the minimum ϕ difference between H_T^{miss} and one of the three highest p_T jets, normalized to take into account the scaling of the $\Delta\hat{\phi}_{\min}$ resolution with H_T^{miss} , as described in Sec.4.

The low $\Delta\hat{\phi}_{\min}$ region is significantly enriched in QCD events. The sample of events with the $\Delta\hat{\phi}_{\min}$ requirement inverted (i.e. $\Delta\hat{\phi}_{\min} < 4$) serves as the QCD control sample. The QCD background for each search bin is estimated from the corresponding bin in the QCD control sample. After subtracting the other backgrounds (top, $W + \text{jets}$, invisible Z) and any signal contamination from the control sample yield, the background estimate is the control sample yield times a transfer factor that is the ratio of events passing / failing the $\Delta\hat{\phi}_{\min} > 4$ requirement for QCD events. This ratio R^{QCD} has some dependence on the variables that define the search bins: H_T^{miss} , H_T , $N_{\text{b-jet}}$, and N_{jet} , though by using $\Delta\hat{\phi}_{\min}$, the H_T^{miss} dependence is strongly reduced.

We model the H_T^{miss} , H_T , $N_{\text{b-jet}}$, and N_{jet} dependence of R^{QCD} by assuming that it factorizes. That is, we assume the H_T dependence does not depend on H_T^{miss} , $N_{\text{b-jet}}$, or N_{jet} and similarly for H_T^{miss} , $N_{\text{b-jet}}$, and N_{jet} . To be explicit, we use

$$R_{i,j,k,l}^{\text{QCD}} = K_{\text{MHT},i}^{\text{QCD}} \cdot K_{\text{HT},j}^{\text{QCD}} \cdot K_{\text{nb},k}^{\text{QCD}} \cdot K_{\text{nj},l}^{\text{QCD}}, \quad (15)$$

where i, j, k, l are the H_T^{miss} , H_T , $N_{\text{b-jet}}$, and N_{jet} bin indices. This model assumption is tested by directly fitting QCD Monte Carlo $R_{i,j,k,l}^{\text{QCD}}$ values for the K^{QCD} parameters and seeing if the model is able to describe all $R_{i,j,k,l}^{\text{QCD}}$ values simultaneously. The results of this test are shown in Fig.22. Many of the bins in Fig.22 have no black point, meaning that there weren't enough QCD MC events in that bin to define R^{QCD} . We have tested the factorization hypothesis of the QCD model with looser binning and have found reasonable agreement (may add this to an appendix).

To remove ambiguity in the K^{QCD} factors, we fix $K_{\text{MHT},1}^{\text{QCD}}$, $K_{\text{nb},1}^{\text{QCD}}$, and $K_{\text{nj},1}^{\text{QCD}}$ to one. In the 8 TeV RA2b analysis[1, 10], we were able to determine many of the K^{QCD} factors directly from the data. This was possible because for each of the analysis dimensions, the lowest bin in each of the other n-1 dimensions had enough QCD to constrain the K^{QCD} factors. We will attempt this in the 13 TeV analysis, though it's likely that some will need to be constrained by the QCD MC.

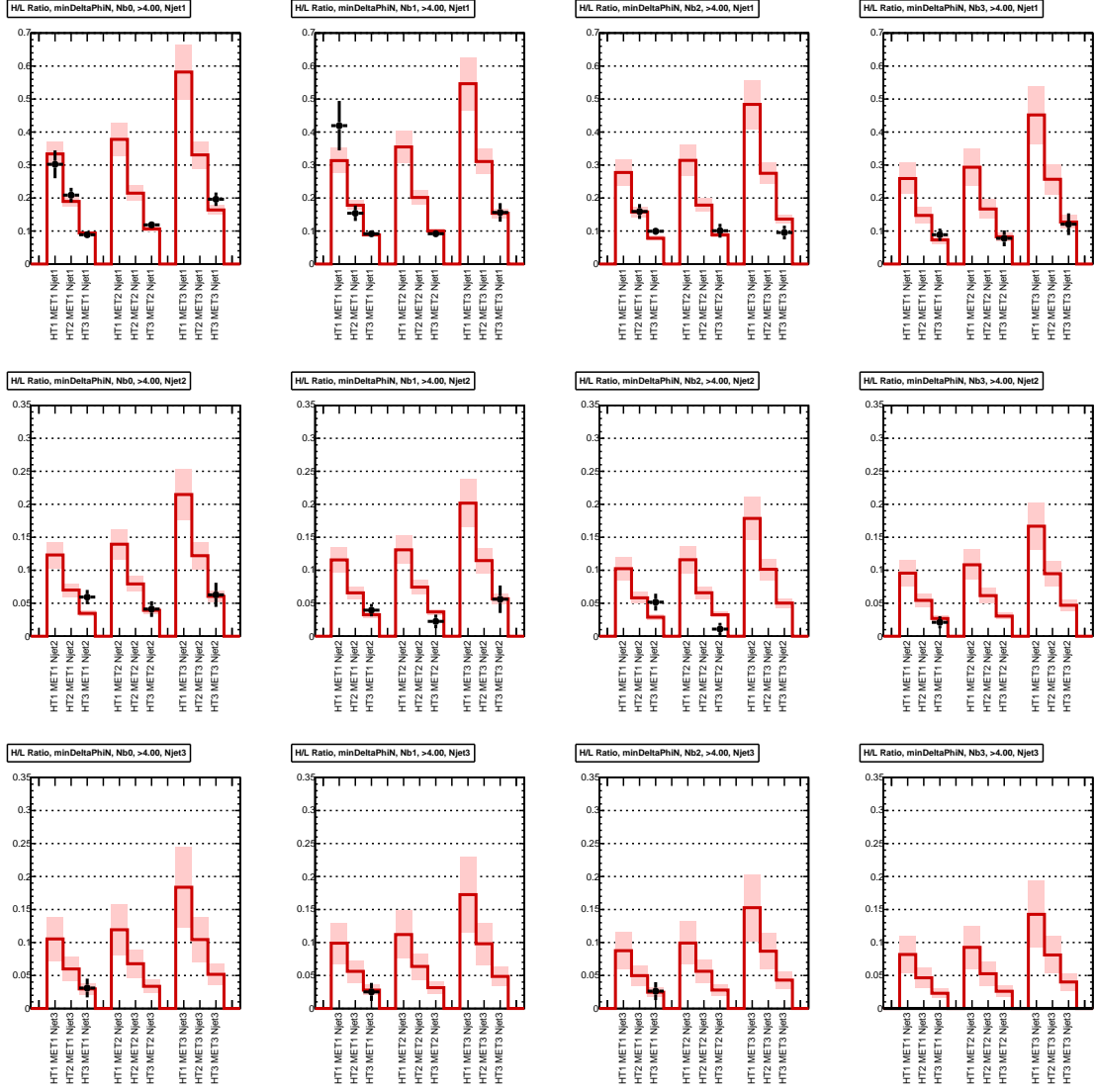


Figure 22: Chisquare fit of the QCD Monte Carlo $R_{ij,k,l}^{QCD}$ values (black points with error bars). The fit result for the QCD model is given by the red histogram, where the shaded region is the uncertainty from the chisquare fit. Bins without a black point do not have enough QCD MC events contributing to define the $R_{ij,k,l}^{QCD}$ ratio. The rows are different N_{jet} bins and the columns are different N_{b-jet} bins.

6.3 Z+jets with $Z \rightarrow \nu\bar{\nu}$

Events that produce a Z boson in association with jets represent an irreducible background for our search when the Z decays to two neutrinos. We are pursuing two techniques to measure this background. In Section 6.3.1, we describe our method to measure $Z \rightarrow \nu\bar{\nu}$ by reconstructing $Z \rightarrow \ell^+\ell^-$ decays, where $\ell = e$ or μ , and removing the leptons from the event. Section 6.3.2 describes another method we are using to measure this background by selecting a γ +jets sample.

6.3.1 $Z \rightarrow \ell^+\ell^-$ with $\ell = e, \mu$

Troy, Bill F.

We select $Z \rightarrow \ell^+\ell^-$ candidates in the Phys14 H_T binned Drell-Yan (DY) MC sample listed in Table 1. Unlike what we expect from a Double Lepton triggered data sample, the reconstructed Z purity in the DY sample is very near 100%. To more closely model the data, we perform our data driven estimate on a MC dataset which includes DY and $t\bar{t}$ MC.

Leptons are identified with the same criteria as lepton vetoes described in Section 4, with the additional requirement of lepton $p_T \geq 17$ GeV. Oppositely signed pairs of selected leptons are combined to form Z candidates in the invariant mass range 91.188 ± 15 GeV. The lepton p_T is added to the E_T^{miss} in the event to simulate $Z \rightarrow \nu\bar{\nu}$. We call this new quantity modifiedMET. Additionally, $\Delta\phi_{\text{min}}$ is recomputed with modifiedMET. The same cuts are applied to this sample as given in Section 4.

We observe very few events over parts of the kinematic plane. However, in both $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ we find that the kinematic shapes in each of the N_{jet} and $N_{\text{b-jet}}$ bins agree with the kinematic shapes inclusive in N_{jet} and $N_{\text{b-jet}}$ bins. To determine $Z \rightarrow \ell^+\ell^-$ events in any of our 72 bins, we count the events in the H_T, H_T^{miss} plane and scale the observed events by the extrapolation factor $\mathcal{F}_{N_{\text{jet}}, N_{\text{b-jet}}} = \frac{(Z \rightarrow \ell^+\ell^-)_{N_{\text{jet}}, N_{\text{b-jet}}}}{(Z \rightarrow \ell^+\ell^-)_{\text{Inclusive}}}$, where $\mathcal{F}_{N_{\text{jet}}, N_{\text{b-jet}}}$ is integrated in H_T^{miss} and H_T . To compute the $Z \rightarrow \nu\bar{\nu}$ background, we use:

$$N(Z \rightarrow \nu\bar{\nu})_{N_{\text{jet}}, N_{\text{b-jet}}}^{H_T, H_T^{\text{miss}}} = N(Z \rightarrow \ell^+\ell^-)_{\text{Inclusive}}^{H_T, H_T^{\text{miss}}} \cdot \mathcal{F}_{N_{\text{jet}}, N_{\text{b-jet}}} \cdot \mathcal{P} \cdot \mathcal{R} / (\mathcal{A} \cdot \epsilon), \quad (16)$$

where \mathcal{P} is the purity of the Z sample, $\mathcal{R} = 5.95$ (need to add reference to PDG) is the ratio of branching fractions, \mathcal{A} is the lepton acceptance, and ϵ is the efficiency.

6.3.1.1 Lepton acceptance The acceptance region is defined as Z decays where both muons have $p_T \geq 17$ GeV and $|\eta| < 2.4$, or both electrons have $p_T > 17$ and $|\eta| < 2.5$, and where the Z invariant mass lies within 15 GeV of the nominal Z mass. The efficiency \mathcal{A} for both leptons to lie within the acceptance is calculated from MC truth information from Z decays in the DY MC sample.

6.3.1.2 Lepton selection efficiency Our lepton selection efficiency is measured in Drell-Yan using the full MC statistics. A scale factor is then applied using the ratio of our Double Lepton sample efficiency, measured using Tag and Probe, over Drell-Yan MC efficiency. The Tag and Probe efficiency is measured inclusively in the H_T^{miss}, H_T plane. Results from the fits for passing and failing probes are shown in Figure 23. The Tag and Probe efficiency is obtained by dividing passing signal by the sum of passing and failing signal. The efficiency to reconstruct a Z is found by squaring the lepton selection efficiency.

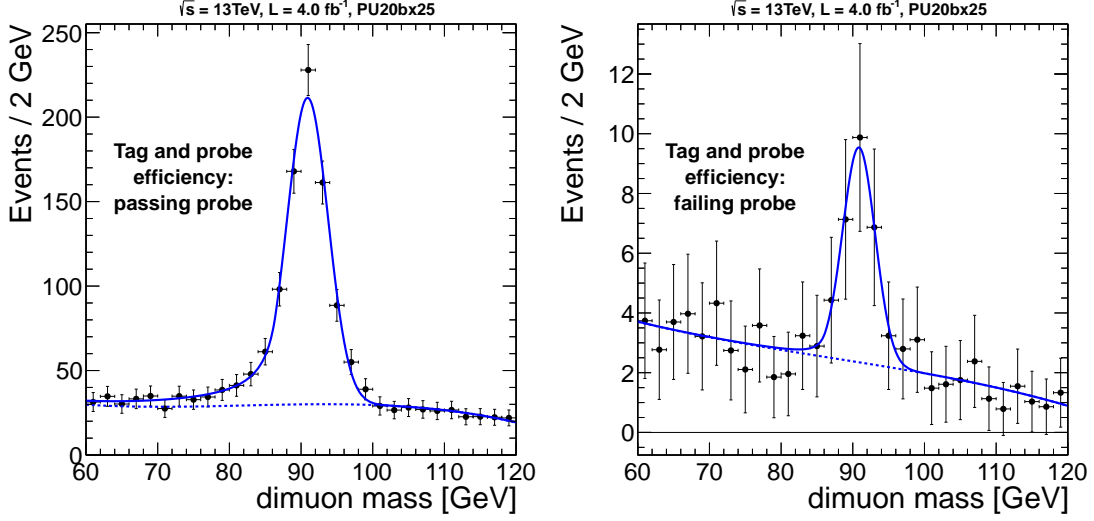


Figure 23: Fits to reconstructed Z candidates to obtain the lepton selection efficiency from the tag and probe study. The $\mu^+\mu^-$ final state is shown for passing (left) and failing (right) probe. Will add corresponding $Z \rightarrow e^+e^-$ plots to Appendix.

6.3.1.3 Z sample purity

The purity is obtained from the fits shown in Figure 24, as the area of the peak function divided by the total area within the Z mass window. Templates for the signal and background shapes are obtained from fitting the full $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ distributions in the Double Lepton sample without any event selection. These template shapes are fixed when fitting for the fraction of signal events in the H_T^{miss}, H_T bin.

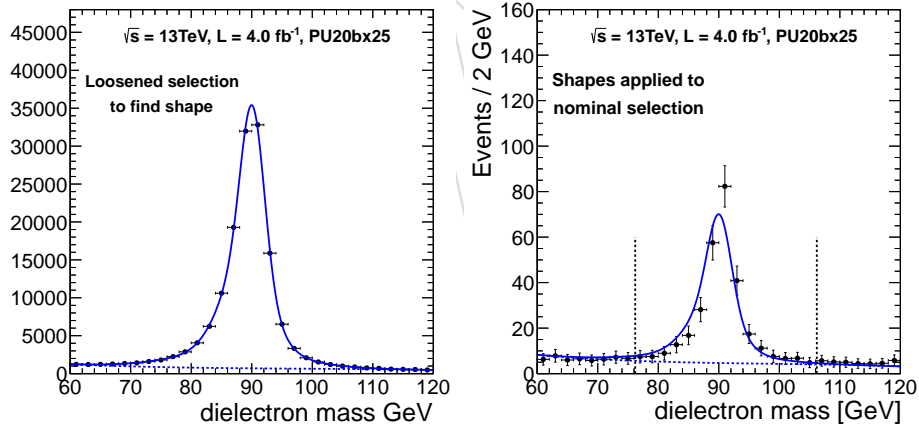


Figure 24: Fits to reconstructed Z candidates in e^+e^- final states with no selection requirements (left) and with the nominal selection requirements. The parameters of the peak shape found in the left plot are fixed for the fit to the distribution in the right plot. Will include $Z \rightarrow \mu^+\mu^-$ in Appendix. Above right plot actually has the additional requirement ≥ 1 CSVL b-tag, these plots will be remade without this requirement. Will also fix axes in left plot.

6.3.1.4 Extrapolation factors

We find good H_T and H_T^{miss} shape agreement between the 1 b-jet sample and inclusive sample, as shown in Figure 25. Also, within statistical error, we find that the shapes remain relatively constant across the N_{jet} bins. Further studies on the relationship between our kinematic bins and our $N_{\text{jet}}, N_{\text{b-jet}}$ multiplicity bins are being pursued. The values for the extrapolation factors are determined from event yields in Table 8, where $\mathcal{F}_{N_{\text{jet}}, N_{\text{b-jet}}}$ is found by dividing the number of events in a particular N_{jet} and $N_{\text{b-jet}}$ bin by the total number of events integrated in these two dimensions. We see very little events in the

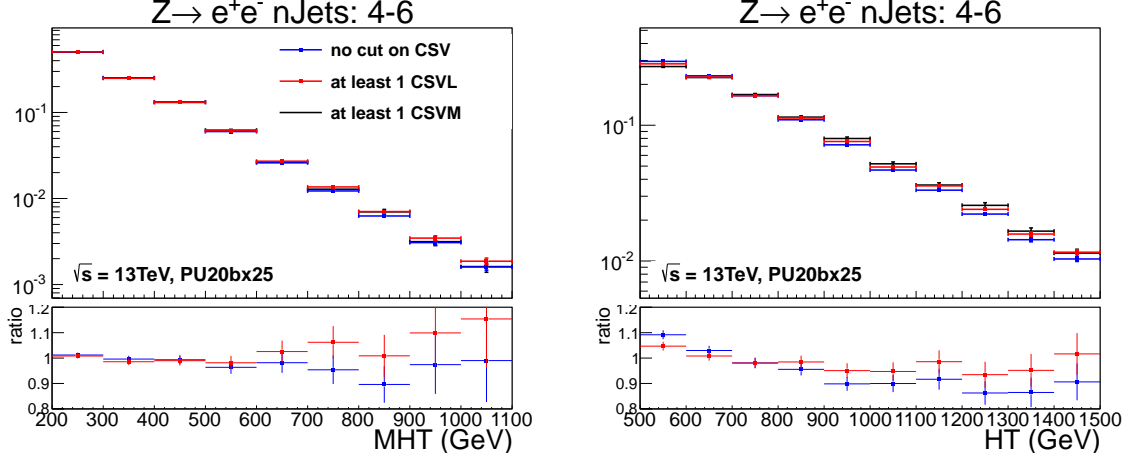


Figure 25: H_T^{miss} and H_T shapes of the $Z \rightarrow e^+e^-$ sample in the 4-6 N_{jet} bin inclusive in b-tags (blue), with at least 1 CSVM b-tag (black), and shown for comparison with at least 1 CSVL b-tag (red). Plots are normalized to unit area. Will put the $Z \rightarrow \mu^+\mu^-$ and the rest of the N_{jet} binned distributions in the Appendix.

largest N_{jet} bins. To rectify our lack of statistics in data, we are considering auxiliary methods of estimating the population in these bins. Alternative methods we are considering include finding a data control sample that matches the N_{jet} and $N_{\text{b-jet}}$ content of the $Z \rightarrow \ell^+\ell^-$ sample, or by using MC for these bins that is verified to be consistent with data throughout the N_{jet} and $N_{\text{b-jet}}$ plane. Table 9 shows the performance of the extrapolation procedure for two bins.

Table 8: Number of $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ events expected for 4 fb^{-1} in the N_{jet} and $N_{\text{b-jet}}$ bins. Uncertainties reflect full MC statistics. In the large N_{jet} bins, we occasionally see few or zero $t\bar{t}$ events, resulting in irregular statistical uncertainties. These numbers are used to compute $\mathcal{F}_{N_{\text{jet}}, N_{\text{b-jet}}}$.

	$Z \rightarrow e^+e^-$			$Z \rightarrow \mu^+\mu^-$		
	4-6 Jets	7-8 Jets	9+ Jets	4-6 Jets	7-8 Jets	9+ Jets
0 b-tagged Jets	176.7 ± 0.9	5.93 ± 0.17	0.32 ± 0.03	167.3 ± 0.8	5.88 ± 0.17	0.32 ± 0.03
1 b-tagged Jet	90.2 ± 1.5	5.8 ± 0.5	0.59 ± 0.18	81.4 ± 1.2	4.01 ± 0.20	0.50 ± 0.18
2 b-tagged Jets	33.7 ± 1.4	4.2 ± 0.5	0.10 ± 0.02	32.5 ± 1.3	2.88 ± 0.41	0.13 ± 0.02
3+ b-tagged Jets	6.2 ± 0.6	0.912 ± 0.22	0.36 ± 0.18	5.3 ± 0.6	1.50 ± 0.34	0.20 ± 0.13

6.3.1.5 Systematic uncertainties The systematic uncertainties for the factors entering into the $Z \rightarrow \nu\bar{\nu}$ background calculation from $Z \rightarrow \ell^+\ell^-$ are summarized in Table 10.

6.3.1.6 Event yields and closure test Tables 8 and 11 show the $Z \rightarrow \ell^+\ell^-$ yields expected for 4 fb^{-1} in the N_{jet} , $N_{\text{b-jet}}$ and H_T , H_T^{miss} planes respectively. We are considering auxiliary methods for determining the extrapolation factors in the most statistically limited N_{jet} and $N_{\text{b-jet}}$ bins.

Table 9: Predicted and observed $Z \rightarrow e^+e^-$ events for 4 fb^{-1} from $N_{\text{jet}} N_{\text{b-jet}}$ extrapolation procedure in each of the $H_T^{\text{miss}} H_T$ bins as depicted in Figure 5 for $N_{\text{jet}} 4-6$; $N_{\text{b-jet}}=1$ and $N_{\text{jet}} 7-8$; $N_{\text{b-jet}}=1$ bins. Uncertainties represent MC statistics. To obtain the full $Z \rightarrow \nu\bar{\nu}$ prediction in these bins we apply the remaining components of Equation 16 found in Table 11.

Bin	Box 1	Box 2	Box 3	Box 4	Box 5	Box 6
$N(Z \rightarrow e^+e^-)$	$N_{\text{jet}} 4-6; N_{\text{b-jet}}=1; (\mathcal{F} = 0.278 \pm 0.005)$					
predicted	63.8 ± 0.6	14.2 ± 0.3	2.82 ± 0.08	7.08 ± 0.13	0.78 ± 0.05	1.52 ± 0.07
observed	63.3 ± 1.4	14.3 ± 0.6	2.95 ± 0.21	7.20 ± 0.33	0.87 ± 0.15	1.54 ± 0.15
$N(Z \rightarrow e^+e^-)$	$N_{\text{jet}} 7-8; N_{\text{b-jet}}=1; (\mathcal{F} = 0.0178 \pm 0.0014)$					
predicted	4.10 ± 0.33	0.91 ± 0.07	0.182 ± 0.015	0.455 ± 0.037	0.051 ± 0.005	0.098 ± 0.009
observed	3.08 ± 0.42	1.89 ± 0.28	0.396 ± 0.130	0.243 ± 0.130	0.137 ± 0.130	0.096 ± 0.130

Table 10: Systematic uncertainties related to the extraction of the $Z \rightarrow \nu\bar{\nu}$ background estimate.

Contribution	size (%)	
	$Z \rightarrow \mu^+\mu^-$	$Z \rightarrow e^+e^-$
Purity	—	—
Acceptance	—	—
Lepton selection efficiency	—	—
MC closure	—	—
Extrapolation (0 b-jets)	—	—
Extrapolation (1 b-jet)	—	—
Extrapolation (2 b-jets)	—	—
Extrapolation (≥ 3 b-jets)	—	—
Extrapolation (4-6 jets)	—	—
Extrapolation (7-8 jets)	—	—
Extrapolation (≥ 9 jets)	—	—

Table 11: Number of $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ events expected for 4 fb^{-1} in the H_T and H_T^{miss} bins as depicted in Figure 5. The uncertainties in these events represent MC statistics. These numbers are scaled by $\mathcal{F}_{N_{\text{jet}}, N_{\text{b-jet}}}$ in Equation 16 to obtain the $Z \rightarrow \nu\bar{\nu}$ prediction in all 4 analysis dimensions.

Bin	Box 1	Box 2	Box 3	Box 4	Box 5	Box 6
$N(Z \rightarrow e^+e^-)$	229.8 ± 2.3	51.1 ± 1.1	10.17 ± 0.28	25.5 ± 0.5	2.83 ± 0.16	5.47 ± 0.26
$N(Z \rightarrow \mu^+\mu^-)$	211.7 ± 2.0	49.6 ± 1.0	10.43 ± 0.34	23.7 ± 0.5	2.43 ± 0.08	4.06 ± 0.17
$\mathcal{A}(Z \rightarrow e^+e^-)$	0.851 ± 0.004	0.849 ± 0.008	0.850 ± 0.017	0.919 ± 0.012	0.915 ± 0.036	0.945 ± 0.027
$\mathcal{A}(Z \rightarrow \mu^+\mu^-)$	0.775 ± 0.004	0.786 ± 0.008	0.797 ± 0.017	0.906 ± 0.013	0.907 ± 0.037	0.925 ± 0.029
$\epsilon_{\text{sel}}(Z \rightarrow e^+e^-)$	0.928 ± 0.020	0.930 ± 0.021	0.926 ± 0.024	0.934 ± 0.022	0.936 ± 0.034	0.936 ± 0.028
$\epsilon_{\text{sel}}(Z \rightarrow \mu^+\mu^-)$	0.947 ± 0.025	0.949 ± 0.026	0.945 ± 0.029	0.950 ± 0.027	0.920 ± 0.038	0.911 ± 0.033
$\mathcal{P}(Z \rightarrow e^+e^-)$	0.831 ± 0.010	0.840 ± 0.020	0.955 ± 0.033	0.941 ± 0.022	0.946 ± 0.065	0.917 ± 0.052
$\mathcal{P}(Z \rightarrow \mu^+\mu^-)$	0.869 ± 0.009	0.850 ± 0.020	0.927 ± 0.037	0.942 ± 0.023	0.995 ± 0.047	0.963 ± 0.050
$Z \rightarrow \nu\bar{\nu}$ ($Z \rightarrow e^+e^-$)	1550 ± 54	348 ± 16	79.3 ± 4.8	178 ± 8	19.9 ± 2.2	36.0 ± 3.2
$Z \rightarrow \nu\bar{\nu}$ ($Z \rightarrow \mu^+\mu^-$)	1575 ± 63	354 ± 18	80.8 ± 5.7	162 ± 9	18.8 ± 1.7	30.3 ± 2.7
$Z \rightarrow \nu\bar{\nu}$ (MC)	1550	399	93.6	167	22.3	33.2

6.3.2 Estimating $(Z \rightarrow \nu\nu) + \text{jets}$ background with $\gamma + \text{jets}$

Andrew, Jim

The missing transverse momentum in $Z \rightarrow \nu\nu + \text{jets}$ events is dominated by the invisible, high- p_T Z boson. A single photon control sample with H_T^{miss} recalculated while neglecting the photon can be used to predict the $Z \rightarrow \nu\nu$ background because the two processes have similar kinematics for high p_T bosons and because differing kinematics for lower p_T bosons can be corrected with a combination of simulation and data.

As shown in Eqs. 17 and 18, we estimate the number of $Z \rightarrow \nu\nu + \text{jets}$ events contributing to a single analysis bin ($N_{Z \rightarrow \nu\nu}^{\text{exp}}$) using our determination of the number of $\gamma + \text{jets}$ produced before detector effects ($N_{\gamma}^{\text{theory}}$), which we obtain from the number of reconstructed $\gamma + \text{jets}$ events passing the event selection (N_{γ}^{obs}) by applying corrections for misreconstructed photons (β_{purity}), photon acceptance (ϵ_{acc}), photon identification efficiency (ϵ_{ID}), and photon isolation efficiency (ϵ_{ISO}). These efficiency parameters are measured in simulation with corrections determined from data ($C_{\text{DATA/MC}}$). Next, this estimate of the true number of $\gamma + \text{jets}$ is multiplied by the phenomenological ratio of $Z \rightarrow \nu\nu + \text{jets}$ events and $\gamma + \text{jets}$ events, which is obtained from simulation with leading order MADGRAPH+PYTHIA ($R_{Z(\nu\nu)/\gamma}^{\text{MC}}$) to account for differences in production between the two processes.

We expect that $R_{Z(\nu\nu)/\gamma}^{\text{MC}}$ will exhibit H_T and H_T^{miss} dependence only at low boson p_T , and that $R_{Z(\nu\nu)/\gamma}^{\text{MC}}$ will have little dependence on jet multiplicity because the hadronic recoil is mostly independent of the produced boson. In the analysis of 8 TeV data, these expectations were confirmed for events with no b-tagged jets. In the current analysis, we are working to understand how the method performs in events with at least one b-tagged jet.

We use NLO simulation and $Z \rightarrow \mu\mu$ data to crosscheck this method and to estimate systematic uncertainties related to the assumption that the simulation correctly models $R_{Z(\nu\nu)/\gamma}^{\text{MC}}$. The $Z \rightarrow \mu\mu$ data is especially important to validate the method at high jet and b-jet multiplicity where NLO calculations are still not available.

Before 13 TeV data arrives, we study this method using simulation such that $C_{\text{DATA/MC}}$ are taken to be unity and N_{γ}^{obs} is taken from the simulation.

$$N_{Z \rightarrow \nu\nu}^{\text{exp}} = R_{Z(\nu\nu)/\gamma}^{\text{MC}} \cdot N_{\gamma}^{\text{theory}} \quad (17)$$

$$N_{Z \rightarrow \nu\nu}^{\text{exp}} = R_{Z(\nu\nu)/\gamma}^{\text{MC}} \cdot \frac{\beta_{\text{purity}}}{\epsilon_{\text{acc}} \cdot \epsilon_{\text{ID}} \cdot \epsilon_{\text{ISO}} \cdot C_{\text{DATA/MC}}} \cdot N_{\gamma}^{\text{obs}}. \quad (18)$$

6.3.2.1 Datasets The simulated samples used to study this background estimation method include sets of $\gamma + \text{jet}$, $Z \rightarrow \nu\nu$, and $Z \rightarrow \mu\mu$ events generated with MADGRAPH+PYTHIA programs in bins of H_T . All samples are from the Phys14DR-PU20bx25_PHYS14_25_V1-v1 production campaign. These samples along with their cross sections and number of events simulated are shown in Table 12.

6.3.2.2 Photon reconstruction For the single photon control sample, we require at least one well identified and isolated photon candidate having at least 100 GeV of transverse momentum and pseudorapidity $|\eta| < 2.5$, excluding the barrel/endcap transition, $1.4442 < |\eta| < 1.566$. Identification and isolation requirements intended to reject electrons and pions misreconstructed as photons are adopted from the Egamma POG's recommendations for 8 TeV data

Dataset	Events Simulated	σ [pb]
GJets_HT-400to600_Tune4C_13TeV-madgraph-tauola	4802346	62.05
GJets_HT-600toInf_Tune4C_13TeV-madgraph-tauola	4341179	20.87
ZJetsToNuNu_HT-400to600_Tune4C_13TeV-madgraph-tauola	4158549	11.99
ZJetsToNuNu_HT-600toInf_Tune4C_13TeV-madgraph-tauola	4201502	4.11

Table 12: Datasets used in γ +jets method.

and 13 TeV data/simulation, where available. The identification criteria, which include requirements for low hadronic activity (H/E), a shower shape ($\sigma_{i\eta i\eta}$) consistent with a photon, and an associated pixel seed veto, are shown in Table 13 for events with a photon in the barrel or endcap.

The isolation requirements restrict the energy sum from particle flow (PF) candidates within a cone of $\Delta R < 0.3$ around the momentum vector of the photon candidate. Specifically, we require that the energy from charged hadrons ($\text{Iso}_{\text{pfCh.}}$), neutral hadrons ($\text{Iso}_{\text{pfNu.}}$), and electromagnetic particles $\text{Iso}_{\text{pfGa.}}$ not exceed the p_T -dependent thresholds shown in Table 13. The isolation energy from each of the three particle species is corrected for pileup with the per-event average pileup energy density (ρ) and per-photon η -dependent effective areas shown in Table 14.

Variable	Barrel	Endcap
Pixel Seed Veto	Yes	Yes
H/E	< 0.05	< 0.05
$\sigma_{i\eta i\eta}$	< 0.011	< 0.031
$\max(\text{Iso}_{\text{pfCh.}} - \text{EA} \cdot \rho, 0)$	< 0.7	< 0.5
$\max(\text{Iso}_{\text{pfNu.}} - \text{EA} \cdot \rho, 0)$	$< 0.4 + 0.04 \cdot p_T$	$< 1.5 + 0.04 \cdot p_T$
$\max(\text{Iso}_{\text{pfGa.}} - \text{EA} \cdot \rho, 0)$	$< 0.5 + 0.005 \cdot p_T$	$< 1.0 + 0.005 \cdot p_T$

Table 13: Photon identification and isolation requirements for events with photons in the barrel and endcap.

$ \eta $	pfCh.	pfNu.	pfGa.
0.0 - 1.0	0.012	0.030	0.148
1.0 - 1.479	0.010	0.057	0.130
1.479 - 2.0	0.014	0.039	0.112
2.0 - 2.2	0.012	0.015	0.216
2.2 - 2.3	0.016	0.024	0.262
2.3 - 2.4	0.020	0.039	0.260
> 2.4	0.012	0.072	0.266

Table 14: Effective areas used in pileup correction as a function of photon pseudorapidity.

6.3.2.3 Photon reconstruction efficiencies and purity

As shown in Eqs. 18, we correct the observed number of γ +jets events (N_{γ}^{obs}) for the reconstruction efficiencies (ϵ_{ID} , ϵ_{ISO}) and the purity of the sample (β_{purity}).

Simulation-based efficiencies for the ID (ϵ_{ID}) and isolation (ϵ_{ISO}) requirements are shown in Figs. 26 and 27 as functions of the variables used for analysis binning: H_T , H_T^{miss} , N_{jets} , and $N_{\text{b-tags}}$. The ID efficiency is computed relative to events passing generator-level acceptance,

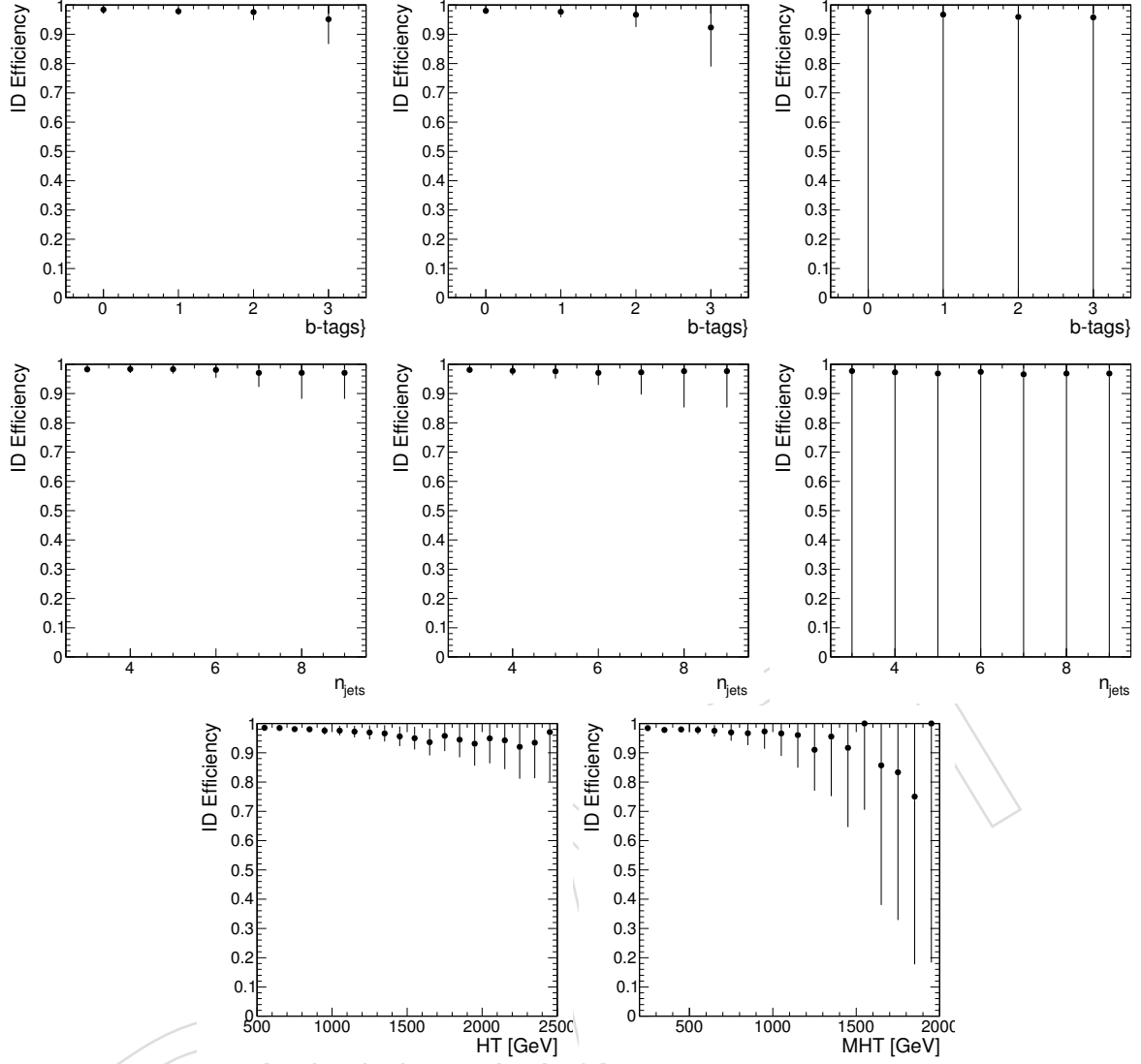


Figure 26: Efficiency of photon ID requirements versus kinematic variables. The left, middle and right plots of the first and second rows show events with low ($200 < H_T^{\text{miss}} < 350$ GeV), medium ($350 < H_T^{\text{miss}} < 500$ GeV), and high ($H_T^{\text{miss}} > 500$ GeV) missing transverse momentum.

$\epsilon_{\text{ID}} = N_{\text{ID}}/N_{\text{acc}}$, while the isolation efficiency is computed relative to the ID efficiency, $\epsilon_{\text{iso}} = N_{\text{iso}}/N_{\text{ID}}$. In both cases, the expected general trend is observed: efficiencies increase with H_T^{miss} (because it is effectively the photon p_T in this method), increase with H_T because of its strong correlation with photon p_T , and decreases with jet multiplicity because of the higher probability for photon-jet overlap. The $C_{\text{data/MC}}$ factors that correct these simulation-based efficiencies (ϵ_{ID} and ϵ_{iso}), which will be measured in 13 TeV data using tag-and-probe methods, are currently assumed to be unity.

The β_{purity} correction for misreconstructed photons is measured in data with a two-component fit of the $\sigma_{i\eta j\eta}$ distribution. The prompt-photon PDF for this fit comes from simulation, the non-prompt-photon PDF comes from data with a photon candidate failing exactly one of the isolation criteria.

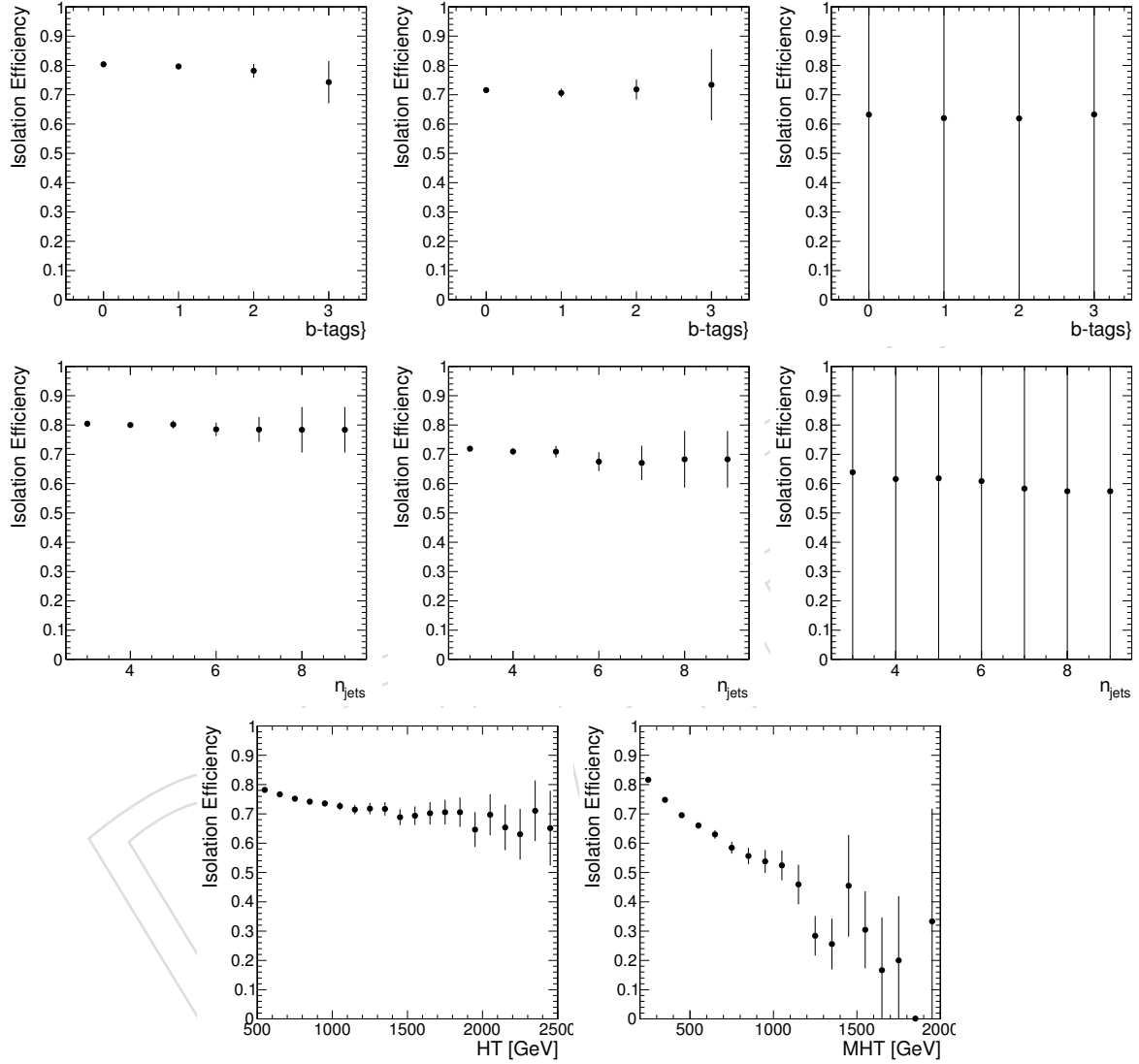


Figure 27: Efficiency of isolation requirements versus kinematic variables. The left, middle and right plots of the first and second rows show events with low ($200 < H_T^{\text{miss}} < 350$ GeV), medium ($350 < H_T^{\text{miss}} < 500$ GeV), and high ($H_T^{\text{miss}} > 500$ GeV) missing transverse momentum.

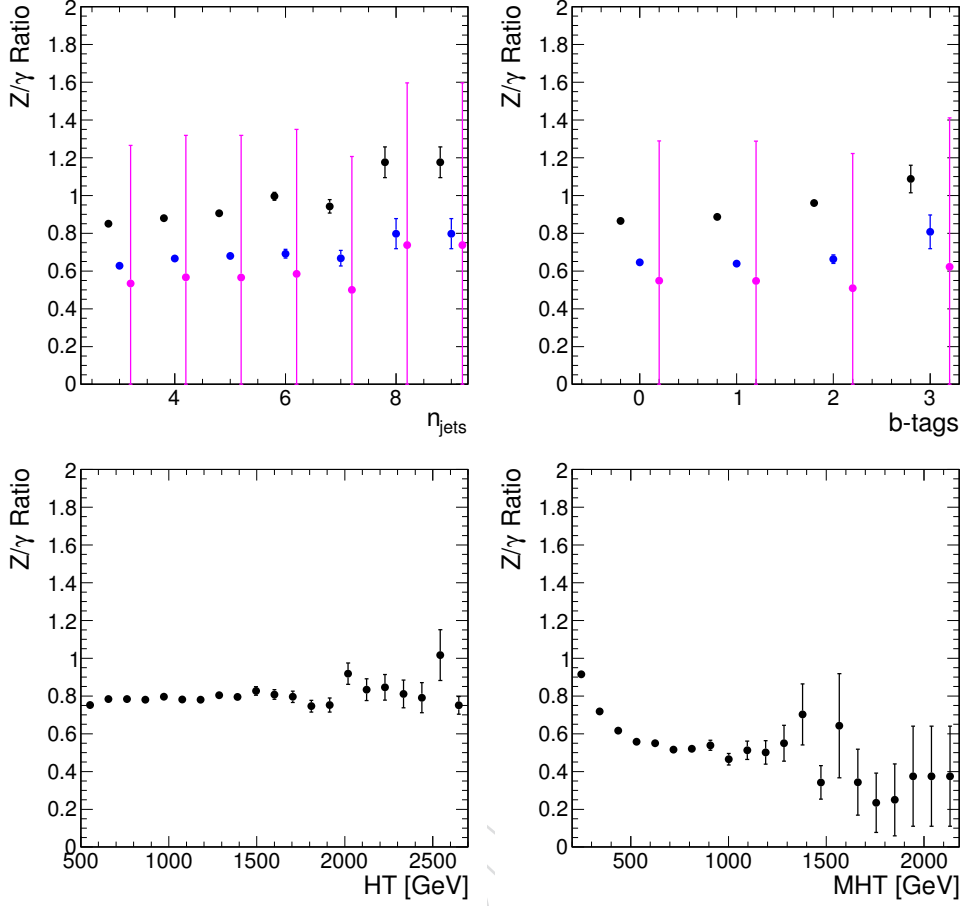


Figure 28: Distributions of Z/γ ratio ($R_{Z(\nu\bar{\nu})/\gamma}^{\text{MC}}$) versus various kinematic variables.

6.3.2.4 Z/γ phenomenological ratio To account for the difference in cross sections and small differences in kinematic distributions of Z and γ events, we apply a $R_{Z(\nu\bar{\nu})/\gamma}^{\text{MC}}$ scale factor which is the ratio of the number of $Z \rightarrow \nu\bar{\nu} + \text{jets}$ and $\gamma + \text{jets}$ events computed with leading order generator-level simulation (MADGRAPH+PYTHIA) in each analysis bin. Figure 28 shows the Z/γ ratio computed as a function of the analysis variables. The ratio is independent of H_T (H_T^{miss}) for $H_T > 500$ GeV ($H_T^{\text{miss}} > 500$ GeV). We observe approximately 30% dependence on both N_{jets} and $N_{\text{b-tags}}$. This dependence was not observed in the 8 TeV analysis, and we are investigating the effects of pileup as a potential cause of the discrepancy.

6.3.2.5 Systematic uncertainty on $\gamma + \text{jets}$ method The dominant systematic uncertainty for this method comes from our understanding of the accuracy of the leading-order modeling of the $R_{Z(\nu\bar{\nu})/\gamma}^{\text{MC}}$ phenomenological ratio. We estimate the systematic uncertainty on this ratio in two ways. First, we adopt the difference between NLO and LO calculations of the ratio as a systematic uncertainty. However, NLO calculations are only available for $N_{\text{jets}} < 4$, so a separate study is needed to understand the performance of the method at high jet multiplicity. To this end, we use the double ratio defined in Eq. 19, as was done for the 8 TeV RA2 analysis, to correct for potential bias coming from mismodeling of the high jet and high b-tag multiplicity bins.

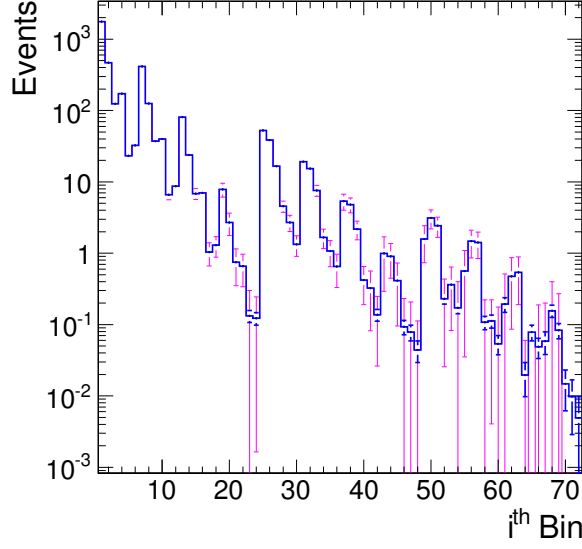


Figure 29: Closure test results.

$$N_{Z \rightarrow \nu\nu}^{\text{exp}} = R_{Z(\nu\bar{\nu})/\gamma}^{\text{MC}} \cdot \frac{R_{Z(\mu\mu)/\gamma}^{\text{data}}}{R_{Z(\mu\mu)/\gamma}^{\text{MC}}} \cdot \frac{\beta_{\text{purity}}}{\epsilon_{\text{acc}} \cdot \epsilon_{\text{ID}} \cdot \epsilon_{\text{ISO}} \cdot C_{\text{DATA/MC}}} \cdot N_{\gamma}^{\text{obs}}. \quad (19)$$

This double ratio is expected to be essentially independent of all analysis variables, but any residual dependence will be adopted as a systematic uncertainty. The use of this double ratio revives some dependence on the small $Z \rightarrow \mu\mu$ control sample, which we are specifically attempting to avoid with this γ +jets events; however, because the double ratio will be shown to be flat, the ratio will be coarsely binned thereby minimizing statistical uncertainty.

6.3.2.6 Closure test for photon method We perform a closure test of the method using MC in place of data. As demonstrated by Equation 20, the closure test trivially gives the correct central value for the $Z \rightarrow \nu\nu$ prediction because the efficiencies and $R_{Z(\nu\bar{\nu})/\gamma}^{\text{MC}}$ are computed in each of the analysis bins. Nevertheless, this simple test provides a benchmark for the expected statistical uncertainty of the method (as well as a sanity check of the method).

$$N_{Z \rightarrow \nu\nu}^{\text{MC}} = R_{Z(\nu\bar{\nu})/\gamma}^{\text{MC}} \cdot \frac{N_{\gamma}^{\text{MC}}}{\epsilon_{\text{acc}} \cdot \epsilon_{\text{ID}} \cdot \epsilon_{\text{ISO}}} = \frac{N_{Z \rightarrow \nu\nu}^{\text{MC}}}{N_{\gamma}^{\text{gen}}} \cdot N_{\gamma}^{\text{ISO}} \cdot \frac{N_{\gamma}^{\text{ID}}}{N_{\gamma}^{\text{ISO}}} \cdot \frac{N_{\gamma}^{\text{acc}}}{N_{\gamma}^{\text{ID}}} \cdot \frac{N_{\gamma}^{\text{gen}}}{N_{\gamma}^{\text{acc}}} \quad (20)$$

Figure 29 shows the expected background from the $Z \rightarrow \nu\nu$ simulation compared to the prediction from this γ +jets method. As expected, the two prediction are identical for this trivial MC test. The uncertainties for the γ +jets method are the uncertainties expected for 5/fb of 13 TeV data.

The computed efficiencies (ϵ_{acc} , ϵ_{ID} , ϵ_{iso}) and $R_{Z(\nu\bar{\nu})/\gamma}^{\text{MC}}$ are shown in Tables 15, 16, and 17. In Tables 18, 19, and 20, we show the MC prediction for the number of events in the γ +jets control region, the MC prediction for the number of $Z \rightarrow \nu\nu$ +jets in the analysis search bins, and the $Z \rightarrow \nu\nu$ +jets prediction based on this γ +jets method.

Bin	ϵ_{acc}	ϵ_{ID}	ϵ_{iso}	$R_{Z/\gamma}^{\text{MC}}$
1	0.971 ± 0.0062	0.987 ± 0.0063	0.782 ± 0.0054	0.78 ± 0.0039
2	0.969 ± 0.0095	0.981 ± 0.0097	0.776 ± 0.0083	0.839 ± 0.0064
3	0.97 ± 0.019	0.96 ± 0.019	0.77 ± 0.016	0.87 ± 0.013
4	0.97 ± 0.013	0.98 ± 0.014	0.66 ± 0.01	0.539 ± 0.006
5	0.97 ± 0.039	0.97 ± 0.039	0.63 ± 0.029	0.69 ± 0.022
6	0.98 ± 0.028	0.97 ± 0.028	0.54 ± 0.019	0.51 ± 0.012
7	0.97 ± 0.013	0.98 ± 0.013	0.78 ± 0.011	0.793 ± 0.0081
8	0.97 ± 0.018	0.98 ± 0.019	0.76 ± 0.016	0.85 ± 0.012
9	0.97 ± 0.034	0.95 ± 0.034	0.75 ± 0.029	0.84 ± 0.023
10	0.96 ± 0.027	0.97 ± 0.028	0.64 ± 0.021	0.55 ± 0.013
11	0.96 ± 0.067	0.96 ± 0.068	0.63 ± 0.051	0.6 ± 0.033
12	0.98 ± 0.055	0.96 ± 0.055	0.54 ± 0.037	0.51 ± 0.024
13	0.97 ± 0.03	0.98 ± 0.031	0.77 ± 0.026	0.86 ± 0.021
14	0.96 ± 0.043	0.97 ± 0.044	0.77 ± 0.038	0.89 ± 0.031
15	0.98 ± 0.08	0.93 ± 0.078	0.76 ± 0.069	0.86 ± 0.055
16	0.97 ± 0.063	0.96 ± 0.064	0.67 ± 0.05	0.51 ± 0.027
17	1.0 ± 0.16	1.0 ± 0.16	0.6 ± 0.11	0.51 ± 0.067
18	1.0 ± 0.14	0.9 ± 0.13	0.48 ± 0.083	0.46 ± 0.053
19	0.96 ± 0.099	1.0 ± 0.1	0.76 ± 0.088	0.96 ± 0.076
20	1.0 ± 0.16	0.9 ± 0.15	0.7 ± 0.14	1.3 ± 0.16
21	1.0 ± 0.24	0.8 ± 0.21	0.7 ± 0.21	0.8 ± 0.15
22	1.0 ± 0.21	1.0 ± 0.22	0.6 ± 0.15	0.6 ± 0.11
23	1.0 ± 0.53	0.9 ± 0.48	0.5 ± 0.35	0.7 ± 0.31
24	1.0 ± 0.45	1.0 ± 0.45	0.6 ± 0.31	0.5 ± 0.18

Table 15: Acceptance, identification, and isolation efficiencies and the Z/γ phenomenological ratio computed in various analysis bins of the low jet multiplicity region ($4 \leq n_{\text{jets}} \leq 6$). Errors account for statistical errors of the MC only. Errors are always report to two significant figures and central values are quote to the most significant figure of the error.

i^{th} Bin	ϵ_{acc}	ϵ_{ID}	ϵ_{iso}	Z/γ Ratio
25	0.96 ± 0.039	0.99 ± 0.041	0.79 ± 0.035	0.9 ± 0.028
26	0.96 ± 0.034	0.97 ± 0.035	0.74 ± 0.029	0.91 ± 0.025
27	0.95 ± 0.052	0.96 ± 0.054	0.73 ± 0.045	0.93 ± 0.039
28	0.96 ± 0.077	0.97 ± 0.08	0.6 ± 0.057	0.53 ± 0.034
29	1.0 ± 0.11	1.0 ± 0.11	0.58 ± 0.079	0.68 ± 0.063
30	1.0 ± 0.14	1.0 ± 0.14	0.49 ± 0.087	0.5 ± 0.058
31	0.95 ± 0.064	0.98 ± 0.067	0.76 ± 0.056	0.87 ± 0.045
32	0.96 ± 0.055	0.97 ± 0.056	0.75 ± 0.047	0.9 ± 0.04
33	0.96 ± 0.079	0.96 ± 0.081	0.76 ± 0.07	0.98 ± 0.062
34	1.0 ± 0.14	1.0 ± 0.14	0.8 ± 0.12	0.55 ± 0.063
35	1.0 ± 0.17	1.0 ± 0.17	0.5 ± 0.11	0.62 ± 0.087
36	1.0 ± 0.22	0.9 ± 0.2	0.6 ± 0.16	0.6 ± 0.1
37	1.0 ± 0.12	1.0 ± 0.12	0.8 ± 0.11	0.81 ± 0.076
38	1.0 ± 0.11	1.0 ± 0.11	0.73 ± 0.091	1.16 ± 0.099
39	0.9 ± 0.13	0.9 ± 0.14	0.8 ± 0.12	0.77 ± 0.083
40	0.9 ± 0.23	1.0 ± 0.26	0.7 ± 0.2	0.5 ± 0.1
41	0.9 ± 0.31	0.9 ± 0.32	0.5 ± 0.2	0.7 ± 0.17
42	1.0 ± 0.33	1.0 ± 0.33	0.4 ± 0.17	0.29 ± 0.088
43	0.9 ± 0.3	1.0 ± 0.32	0.5 ± 0.19	1.0 ± 0.25
44	0.9 ± 0.23	0.9 ± 0.23	0.8 ± 0.22	1.0 ± 0.18
45	1.0 ± 0.37	0.9 ± 0.35	0.9 ± 0.34	1.0 ± 0.29
46	1.0 ± 0.71	1.0 ± 0.71	0.8 ± 0.57	0.9 ± 0.49
47	1.0 ± 0.82	1.0 ± 0.82	1.0 ± 0.82	1.0 ± 0.63
48	1.0 ± 0.71	0.8 ± 0.57	1.0 ± 0.82	0.4 ± 0.25

Table 16: Acceptance, identification, and isolation efficiencies and the Z/γ phenomenological ratio computed in various analysis bins of the medium jet multiplicity region ($7 \leq n_{jets} \leq 8$). Errors account for statistical errors of the MC only. Errors are always report to two significant figures and central values are quote to the most significant figure of the error.

i^{th} Bin	ϵ_{acc}	ϵ_{ID}	ϵ_{iso}	Z/γ Ratio
49	1.0 ± 0.26	1.0 ± 0.26	0.9 ± 0.24	1.2 ± 0.23
50	1.0 ± 0.14	1.0 ± 0.14	0.7 ± 0.11	1.1 ± 0.12
51	1.0 ± 0.15	1.0 ± 0.15	0.8 ± 0.13	1.0 ± 0.12
52	1.0 ± 0.41	0.9 ± 0.38	0.7 ± 0.34	0.7 ± 0.24
53	0.9 ± 0.33	0.9 ± 0.34	0.5 ± 0.23	0.8 ± 0.22
54	1.0 ± 0.5	1.0 ± 0.5	0.3 ± 0.2	0.8 ± 0.32
55	1.0 ± 0.43	1.0 ± 0.47	0.8 ± 0.39	1.0 ± 0.36
56	0.9 ± 0.19	1.0 ± 0.21	0.7 ± 0.17	1.1 ± 0.17
57	0.9 ± 0.17	0.9 ± 0.18	0.8 ± 0.17	0.9 ± 0.14
58	1.0 ± 0.47	1.0 ± 0.47	0.6 ± 0.31	0.5 ± 0.18
59	0.9 ± 0.41	1.0 ± 0.47	0.8 ± 0.39	0.4 ± 0.16
60	1.0 ± 1.0	1.0 ± 1.0	0.5 ± 0.61	1.0 ± 0.79
61	1.0 ± 0.81	1.0 ± 0.81	1.0 ± 0.81	1.3 ± 0.78
62	1.0 ± 0.38	1.0 ± 0.38	0.7 ± 0.31	1.0 ± 0.29
63	1.0 ± 0.29	1.0 ± 0.29	0.6 ± 0.22	0.9 ± 0.21
64	1.0 ± 1.0	1.0 ± 1.0	1.0 ± 1.0	0.4 ± 0.32
65	1.0 ± 1.4	1.0 ± 1.4	0.0 ± 0.0	3.0 ± 3.1
66	1.0 ± 1.4	1.0 ± 1.4	1.0 ± 1.4	2.0 ± 2.0
67	1.0 ± 1.1	1.0 ± 1.1	0.3 ± 0.33	0.6 ± 0.5
68	0.8 ± 0.57	1.0 ± 0.82	1.0 ± 0.82	1.5 ± 0.8
69	1.0 ± 1.0	1.0 ± 1.0	0.5 ± 0.61	2.0 ± 1.2
70	1.0 ± 1.4	1.0 ± 1.4	0.0 ± 0.0	0.6 ± 0.65
71	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	$100000000.0 \pm 69000000.0$
72	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	$50000000.0 \pm 49000000.0$

Table 17: Acceptance, identification, and isolation efficiencies and the Z/γ phenomenological ratio computed in various analysis bins of the high jet multiplicity region ($9 \leq n_{jets}$). Errors account for statistical errors of the MC only. Errors are always report to two significant figures and central values are quote to the most significant figure of the error.

i^{th} Bin	γ +Jets	$Z \rightarrow (\nu\nu)$	Data-driven $Z \rightarrow (\nu\nu)$ Est.
1	1698.0 \pm 8.8	1767.0 \pm 3.8	1770.0 \pm 24.0
2	411.0 \pm 3.3	467.0 \pm 1.5	467.0 \pm 9.7
3	103.0 \pm 1.6	124.7 \pm 0.78	125.0 \pm 5.1
4	201.0 \pm 2.5	172.1 \pm 0.96	172.0 \pm 5.2
5	19.8 \pm 0.72	23.2 \pm 0.34	23.0 \pm 2.0
6	32.7 \pm 0.93	32.3 \pm 0.4	32.0 \pm 2.1
7	387.0 \pm 4.1	414.0 \pm 1.8	410.0 \pm 11.0
8	106.0 \pm 1.7	125.6 \pm 0.79	126.0 \pm 5.1
9	30.4 \pm 0.89	37.3 \pm 0.43	37.0 \pm 2.8
10	44.0 \pm 1.1	39.9 \pm 0.46	40.0 \pm 2.5
11	6.4 \pm 0.41	6.6 \pm 0.18	7.0 \pm 1.0
12	8.6 \pm 0.47	8.7 \pm 0.21	9.0 \pm 1.1
13	68.0 \pm 1.7	80.7 \pm 0.81	81.0 \pm 5.3
14	19.1 \pm 0.71	23.8 \pm 0.34	24.0 \pm 2.3
15	5.5 \pm 0.38	6.8 \pm 0.18	7.0 \pm 1.2
16	8.5 \pm 0.49	7.0 \pm 0.19	7.0 \pm 1.0
17	1.1 \pm 0.17	1.03 \pm 0.071	1.0 \pm 0.37
18	1.3 \pm 0.18	1.29 \pm 0.08	1.3 \pm 0.41
19	5.7 \pm 0.5	7.8 \pm 0.25	8.0 \pm 1.7
20	1.4 \pm 0.19	2.7 \pm 0.11	2.7 \pm 0.94
21	0.5 \pm 0.12	0.75 \pm 0.061	0.7 \pm 0.4
22	0.6 \pm 0.13	0.66 \pm 0.059	0.7 \pm 0.32
23	0.08 \pm 0.045	0.13 \pm 0.025	0.1 \pm 0.17
24	0.16 \pm 0.064	0.12 \pm 0.024	0.1 \pm 0.12

Table 18: Estimated event yields in both signal and the signal photon control region from both MC and the γ +jets data driven background estimation in various bins of the low jet multiplicity region ($4 \leq n_{jets} \leq 6$). Errors correspond to statistical errors expected from 5/fb for the data-driven estimation and the single photons control sample yields. Errors are always report to two significant figures and central values are quote to the most significant figure of the error.

i^{th} Bin	γ +Jets	$Z \rightarrow (\nu\nu)$	Data-driven $Z \rightarrow (\nu\nu)$ Est.
25	44.0 ± 1.5	52.7 ± 0.7	53.0 ± 4.5
26	29.7 ± 0.89	38.7 ± 0.44	39.0 ± 2.9
27	11.9 ± 0.56	16.6 ± 0.29	17.0 ± 2.0
28	4.9 ± 0.36	4.5 ± 0.15	4.5 ± 0.81
29	2.2 ± 0.24	2.7 ± 0.11	2.7 ± 0.69
30	1.3 ± 0.18	1.33 ± 0.081	1.3 ± 0.43
31	15.7 ± 0.87	19.2 ± 0.42	19.0 ± 2.8
32	11.9 ± 0.57	15.3 ± 0.28	15.0 ± 1.9
33	5.4 ± 0.38	7.6 ± 0.19	8.0 ± 1.3
34	2.3 ± 0.27	1.66 ± 0.093	1.7 ± 0.51
35	0.8 ± 0.15	1.07 ± 0.072	1.1 ± 0.43
36	0.6 ± 0.13	0.65 ± 0.056	0.7 ± 0.32
37	4.9 ± 0.49	5.3 ± 0.22	5.0 ± 1.4
38	2.8 ± 0.27	4.8 ± 0.16	5.0 ± 1.2
39	1.9 ± 0.22	2.2 ± 0.1	2.2 ± 0.65
40	0.5 ± 0.11	0.42 ± 0.045	0.4 ± 0.24
41	0.21 ± 0.074	0.32 ± 0.04	0.3 ± 0.25
42	0.18 ± 0.069	0.14 ± 0.026	0.1 ± 0.11
43	0.4 ± 0.14	0.99 ± 0.095	1.0 ± 0.71
44	0.7 ± 0.13	0.91 ± 0.067	0.9 ± 0.47
45	0.31 ± 0.091	0.41 ± 0.045	0.4 ± 0.32
46	0.08 ± 0.045	0.09 ± 0.021	0.1 ± 0.14
47	0.08 ± 0.045	0.08 ± 0.02	0.1 ± 0.13
48	0.08 ± 0.045	0.04 ± 0.015	0.04 ± 0.069

Table 19: Estimated event yields in both signal and the signal photon control region from both MC and the γ +jets data driven background estimation in various bins of the med jet multiplicity region ($7 \leq n_{jets} \leq 8$). Errors correspond to statistical errors expected from 5/fb for the data-driven estimation and the single photons control sample yields. Errors are always report to two significant figures and central values are quote to the most significant figure of the error.

i^{th} Bin	γ +Jets	$Z \rightarrow (\nu\nu)$	Data-driven $Z \rightarrow (\nu\nu)$ Est.
49	1.2 ± 0.24	1.6 ± 0.12	1.6 ± 0.84
50	2.0 ± 0.23	3.1 ± 0.13	3.1 ± 0.93
51	1.7 ± 0.21	2.4 ± 0.11	2.4 ± 0.78
52	0.21 ± 0.074	0.23 ± 0.036	0.2 ± 0.2
53	0.21 ± 0.074	0.36 ± 0.042	0.4 ± 0.29
54	0.05 ± 0.037	0.17 ± 0.029	0.2 ± 0.23
55	0.4 ± 0.15	0.56 ± 0.07	0.6 ± 0.54
56	0.9 ± 0.15	1.48 ± 0.09	1.5 ± 0.64
57	0.9 ± 0.16	1.41 ± 0.084	1.4 ± 0.59
58	0.13 ± 0.058	0.11 ± 0.023	0.1 ± 0.11
59	0.18 ± 0.069	0.11 ± 0.023	0.1 ± 0.11
60	0.03 ± 0.026	0.05 ± 0.016	0.1 ± 0.12
61	0.15 ± 0.088	0.19 ± 0.045	0.2 ± 0.32
62	0.3 ± 0.11	0.47 ± 0.05	0.5 ± 0.39
63	0.34 ± 0.094	0.54 ± 0.051	0.5 ± 0.36
64	0.05 ± 0.037	0.02 ± 0.0098	0.02 ± 0.04
65	0.0 ± 0.0	0.08 ± 0.02	0.0 ± 0.0
66	0.03 ± 0.026	0.05 ± 0.015	0.0 ± 0.14
67	0.03 ± 0.026	0.06 ± 0.021	0.1 ± 0.14
68	0.08 ± 0.045	0.16 ± 0.03	0.2 ± 0.25
69	0.03 ± 0.026	0.08 ± 0.02	0.1 ± 0.19
70	0.0 ± 0.0	0.015 ± 0.0085	0.0 ± 0.0
71	0.0 ± 0.0	0.01 ± 0.0069	0.0 ± 0.0
72	0.0 ± 0.0	0.005 ± 0.0049	0.0 ± 0.0

Table 20: Estimated event yields in both signal and the signal photon control region from both MC and the γ +jets data driven background estimation in various bins of the high jet multiplicity region ($9 \leq n_{jets}$). Errors correspond to statistical errors expected from 5/fb for the data-driven estimation and the single photons control sample yields. Errors are always report to two significant figures and central values are quote to the most significant figure of the error.

456 **7 Statistical treatment**

457 Owen, Nhan

458 **8 Sensitivity studies**

459 **9 Summary**

460 Bill G.

DRAFT

References

- [1] CMS Collaboration, “Search for gluino mediated bottom- and top-squark production in multijet final states in pp collisions at 8 TeV”, *Phys. Lett. B* **725** (2013) 243–270, doi:10.1016/j.physletb.2013.06.058.
- [2] CMS Collaboration, “Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **06** (2014) 055, doi:10.1007/JHEP06(2014)055.
- [3] Harel, A. and Schieferdecker, P., “Jet Identification”, (2010). <https://twiki.cern.ch/twiki/bin/view/CMS/JetID>.
- [4] CMS Muon POG, “Baseline muon selections for 2012 data (CMSSW 52X and above)”, (2012). <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId#The2012Data>.
- [5] CMS E/gamma POG, “Cut Based Electron ID for Run 2”, (2014). <https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2>.
- [6] CMS Collaboration, “Search for supersymmetry in events with b-quark jets and missing transverse energy in pp collisions at 7 TeV”, *Phys. Rev. D* **86** (2012) 072010, doi:10.1103/PhysRevD.86.072010.
- [7] The RA2b Group, “Search for New Physics in Events with b-Jets and Missing Transverse Energy with the full 2011 Data Sample”, *CMS Analysis Note* **AN-11-409** (2011).
- [8] A. Hook, E. Izaguirre, M. Lisanti, and J. G. Wacker, “High multiplicity searches at the LHC using jet masses”, *Phys. Rev. D* **85** (Mar, 2012) 055029, doi:10.1103/PhysRevD.85.055029.
- [9] A. Czarnecki, J. Korner, and J. Piclum, “Helicity fractions of W bosons from top quark decays at NNLO in QCD”, *Phys. Rev. D* **81** (2010) 111503, doi:10.1103/PhysRevD.81.111503.
- [10] The RA2b Group, “Search for supersymmetry in pp collisions at 8 TeV using the shape of H_T , E_T^{miss} , and b-quark jet multiplicity distributions”, *CMS Analysis Note* **AN-12-081** (2012).
- [11] The RA2 Group, “Search for new physics in multijets and missing momentum final state at 8 TeV”, *CMS Analysis Note* **AN-12-350** (2012).
- [12] R. D. Cousins, J. T. Linnemann, and J. Tucker, “Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **595** (2008), no. 2, 480 – 501, doi:http://dx.doi.org/10.1016/j.nima.2008.07.086.
- [13] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **1407** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [14] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, arXiv:1410.3012.

- 501 [15] Giovanni Zevi Della Porta, “ECAL and HCAL reconstruction for 25ns bunch spacing”,
502 (2014). [https://indico.cern.ch/event/339666/session/2/contribution/](https://indico.cern.ch/event/339666/session/2/contribution/20/material/slides/0.pdf)
503 [20/material/slides/0.pdf](https://indico.cern.ch/event/339666/session/2/contribution/20/material/slides/0.pdf).

DRAFT

A N-1 plots for most sensitive regions

A rough optimization is performed over the four search variables (H_T , H_T^{miss} , N_{jet} , and $N_{\text{b-jet}}$) to identify regions of high sensitivity (“superbins”) at 4 fb^{-1} for each model point. This study motivates the choice of binning (Chapter 5) and also offers insight into which variables are most powerful at discriminating signal from background. The results are reported in this section in the form of $N - 1$ plots (Figures 30-35), as well as the expected signal and background yields and the Z_{Bi} figure of merit for each region (Table 21). The Z_{Bi} figure of merit approximates the expected significance for n_s signal events and n_b background events assuming a poisson-distributed background measurement performed in a control sample with $\tau = \mu_{\text{off}}/\mu_{\text{on}}$ times the amount of statistics as the search region [12]. For this exercise, we use $\tau = 1$, which is approximately the ratio of the number of events in the single lepton control sample to the number of events in the zero lepton search region.

Table 21: Expected signal and background yields and significance (Z_{Bi} at 4 fb^{-1} after tighter selection in search variables. Each table corresponds to a “superbin” (region of high sensitivity) for a particular signal point. The baseline selection is defined in Chapter 4. This “ $N - 1$ ” selection is illustrated in Figures 30-35.

Cut	T1tttt ($m_{\tilde{g}} = 1500 \text{ GeV}$, $m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$)	SM	Z_{Bi}
Baseline	19.3 ± 0.1	13659.6 ± 1038.0	0.1
$H_T > 1000 \text{ GeV}$	18.4 ± 0.1	2043.1 ± 92.4	0.3
$N_{\text{b-jet}} \geq 2$	15.3 ± 0.1	322.1 ± 13.4	0.6
$N_{\text{jet}} \geq 9$	9.9 ± 0.1	36.2 ± 2.8	1.0
$H_T^{\text{miss}} > 500 \text{ GeV}$	5.8 ± 0.1	1.9 ± 0.4	1.6
Cut	T1tttt ($m_{\tilde{g}} = 1200 \text{ GeV}$, $m_{\tilde{\chi}_1^0} = 800 \text{ GeV}$)	SM	Z_{Bi}
Baseline	44.1 ± 0.4	13659.6 ± 1038.0	0.3
$N_{\text{b-jet}} \geq 2$	35.4 ± 0.3	2143.1 ± 50.9	0.5
$N_{\text{jet}} \geq 9$	23.2 ± 0.3	81.0 ± 3.6	1.6
Cut	T1bbbb ($m_{\tilde{g}} = 1500 \text{ GeV}$, $m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$)	SM	Z_{Bi}
Baseline	38.0 ± 0.1	13659.6 ± 1038.0	0.2
$N_{\text{b-jet}} \geq 2$	27.9 ± 0.1	2143.1 ± 50.9	0.4
$H_T > 1200 \text{ GeV}$	24.6 ± 0.1	150.0 ± 9.5	1.3
$H_T^{\text{miss}} > 500 \text{ GeV}$	16.6 ± 0.1	16.2 ± 1.2	2.2
Cut	T1bbbb ($m_{\tilde{g}} = 1000 \text{ GeV}$, $m_{\tilde{\chi}_1^0} = 900 \text{ GeV}$)	SM	Z_{Bi}
Baseline	166.6 ± 1.5	13659.6 ± 1038.0	1.0
$N_{\text{b-jet}} \geq 2$	104.1 ± 1.2	2143.1 ± 50.9	1.6
$H_T^{\text{miss}} > 500 \text{ GeV}$	33.7 ± 0.7	50.7 ± 2.1	2.8
Cut	T1qqqq ($m_{\tilde{g}} = 1400 \text{ GeV}$, $m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$)	SM	Z_{Bi}
Baseline	67.9 ± 0.3	13659.6 ± 1038.0	0.4
$H_T > 1400 \text{ GeV}$	51.5 ± 0.2	466.2 ± 33.8	1.6
$H_T^{\text{miss}} > 400 \text{ GeV}$	41.6 ± 0.2	150.6 ± 28.0	2.2
Cut	T1qqqq ($m_{\tilde{g}} = 1000 \text{ GeV}$, $m_{\tilde{\chi}_1^0} = 800 \text{ GeV}$)	SM	Z_{Bi}
Baseline	279.9 ± 1.9	13659.6 ± 1038.0	1.7
$H_T^{\text{miss}} > 300 \text{ GeV}$	179.0 ± 1.6	4899.2 ± 1011.5	1.8
$N_{\text{jet}} \geq 6$	114.3 ± 1.2	1001.5 ± 60.6	2.5

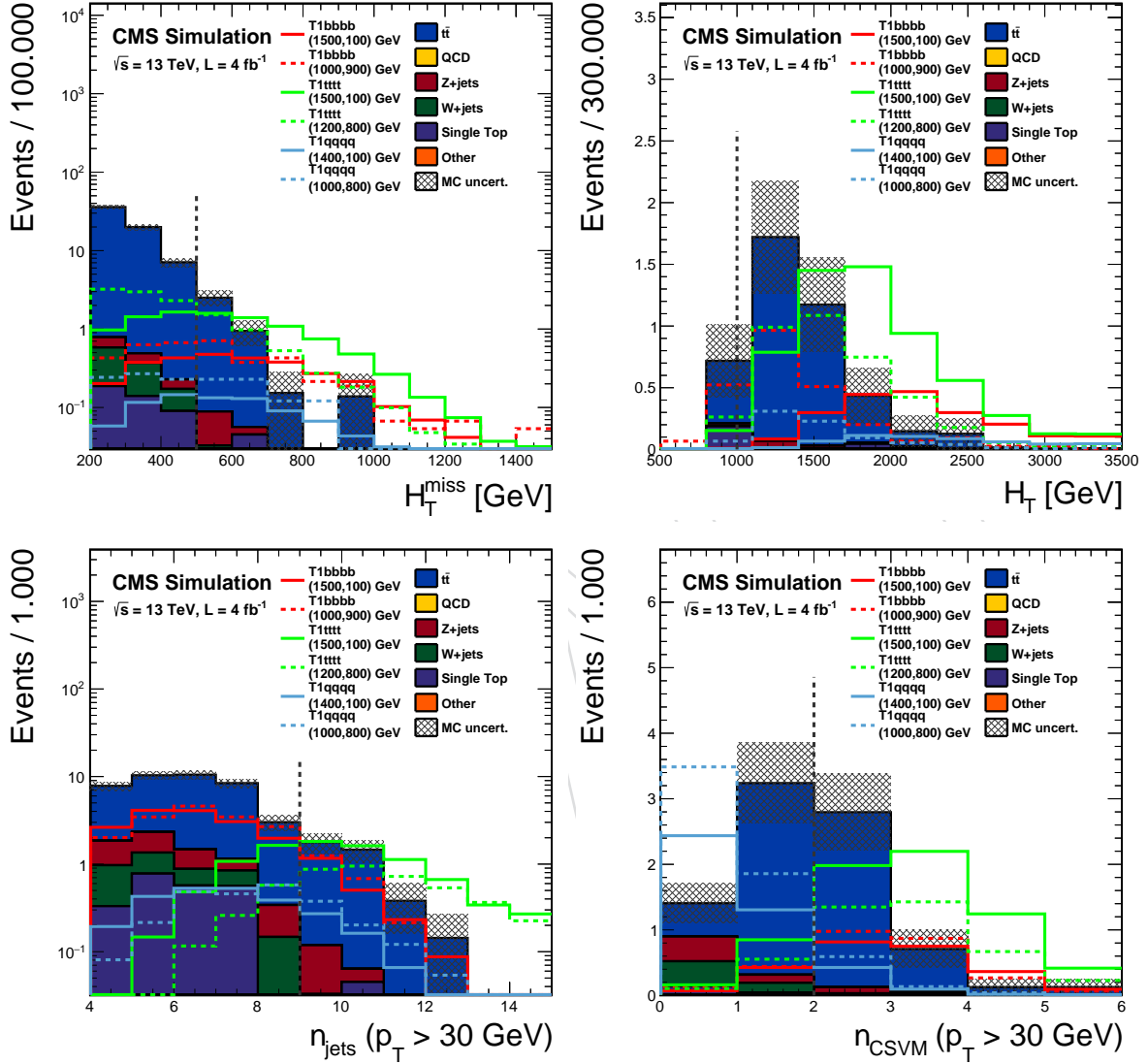


Figure 30: $N-1$ plots of the binning variables for the most sensitive region for the non-compressed $T1tttt$ point, $m_{\tilde{g}} = 1500 \text{ GeV}$, $m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$ (solid green line). The boundaries for this region, marked in the plots by dashed gray vertical lines, are $H_T > 1000 \text{ GeV}$, $H_T^{\text{miss}} > 500 \text{ GeV}$, $N_{b\text{-jet}} \geq 2$, and $N_{\text{jet}} \geq 9$.

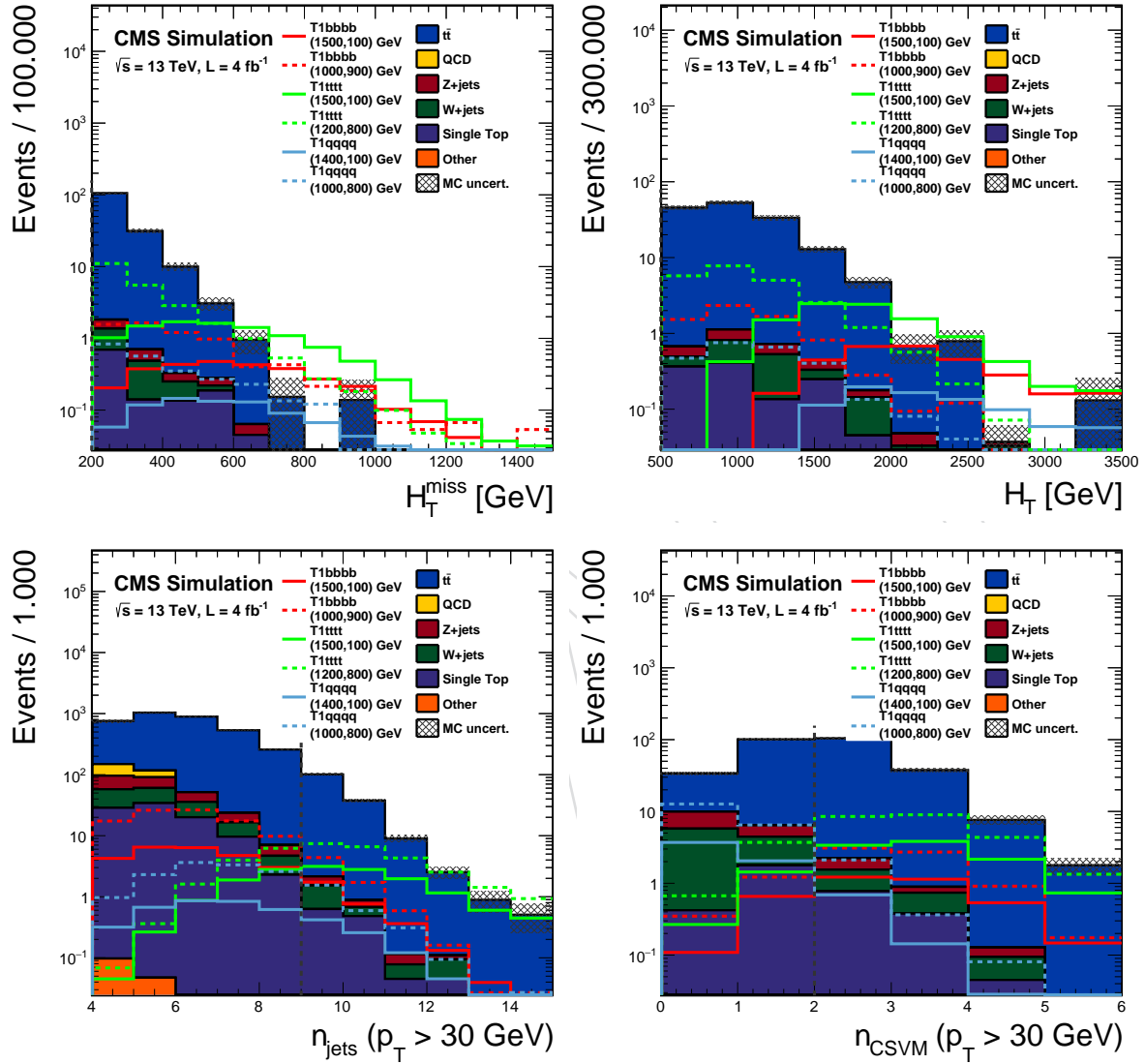


Figure 31: $N - 1$ plots of the binning variables for the most sensitive region for the compressed $T1tttt$ point, $m_{\tilde{g}} = 1200$ GeV, $m_{\tilde{\chi}_1^0} = 800$ GeV (dashed green line). The boundaries for this region, marked in the plots by dashed gray vertical lines, are $H_T > 500$ GeV, $H_T^{\text{miss}} > 200$ GeV, $N_{b\text{-jet}} \geq 2$, and $N_{\text{jet}} \geq 9$.

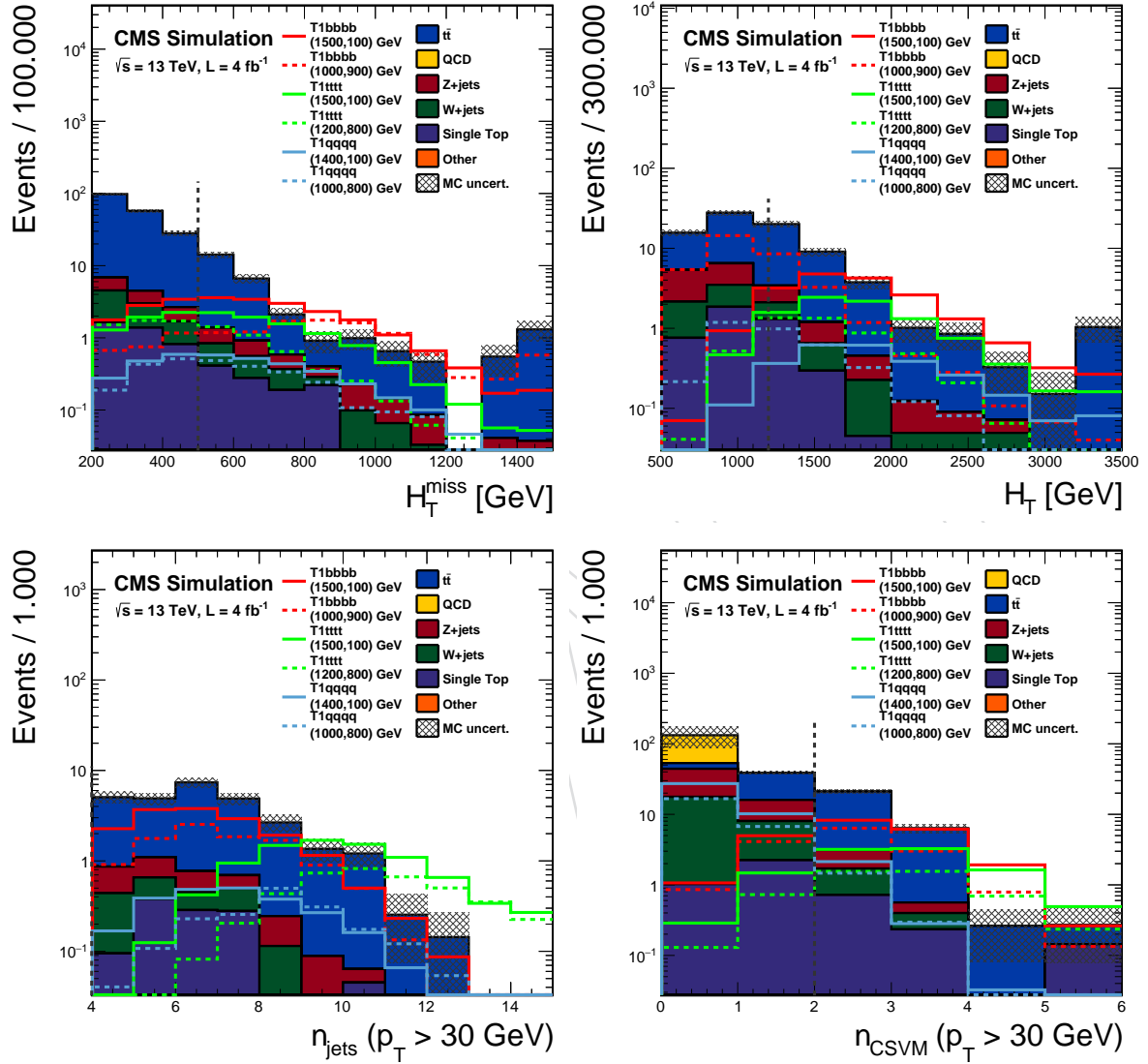


Figure 32: $N-1$ plots of the binning variables for the most sensitive region for the non-compressed T1bbbb point, $m_{\tilde{g}} = 1500$ GeV, $m_{\tilde{\chi}_1^0} = 100$ GeV (solid red line). The boundaries for this region, marked in the plots by dashed gray vertical lines, are $H_T > 1200$ GeV, $H_T^{\text{miss}} > 500$ GeV, $N_{\text{b-jet}} \geq 2$, and $N_{\text{jet}} \geq 4$.

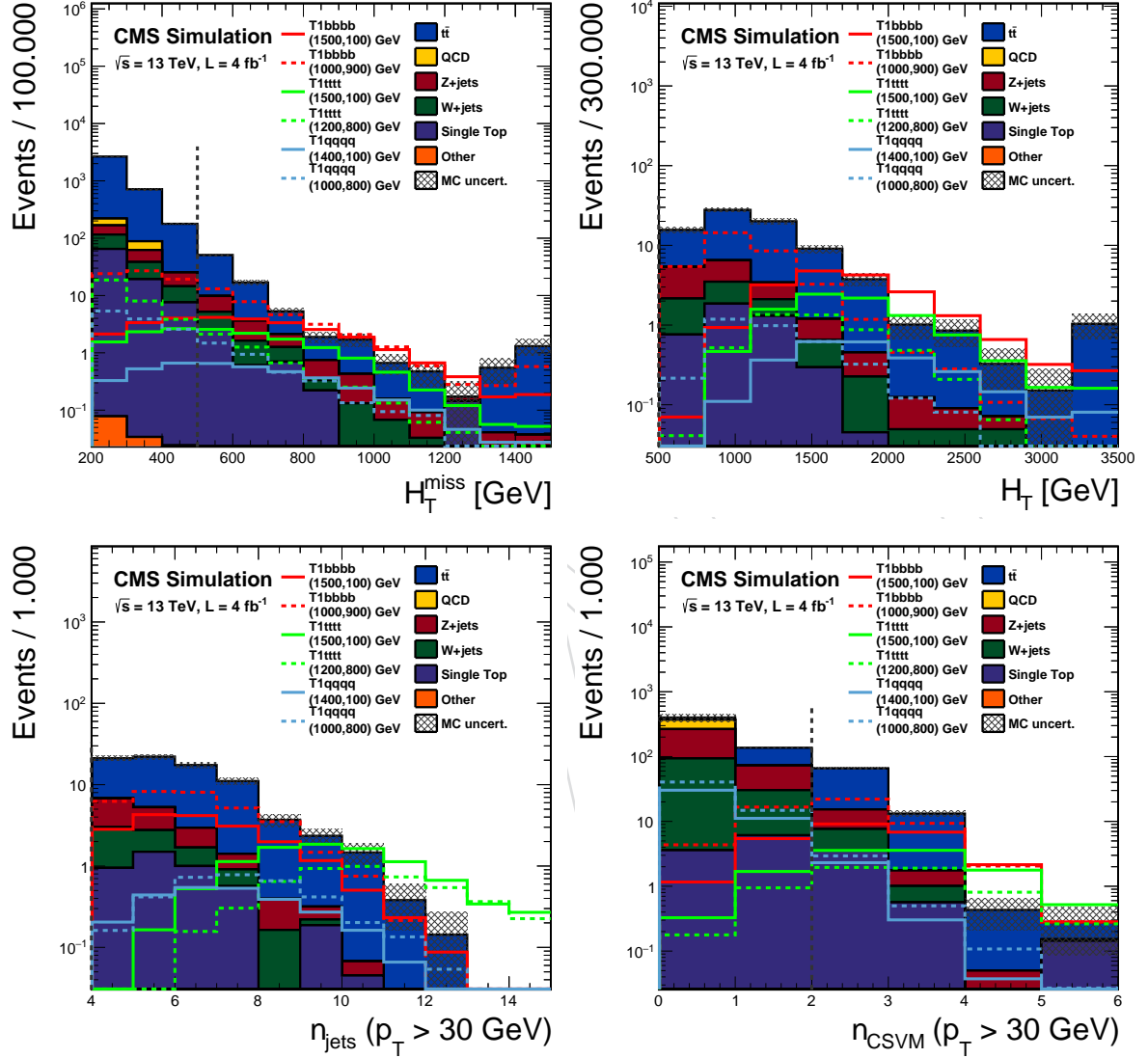


Figure 33: $N-1$ plots of the binning variables for the most sensitive region for the non-compressed T1bbbb point, $m_{\tilde{g}} = 1000$ GeV, $m_{\tilde{\chi}_1^0} = 900$ GeV (dashed red line). The boundaries for this region, marked in the plots by dashed gray vertical lines, are $H_T > 500$ GeV, $H_T^{\text{miss}} > 500$ GeV, $N_{\text{b-jet}} \geq 2$, and $N_{\text{jet}} \geq 4$.

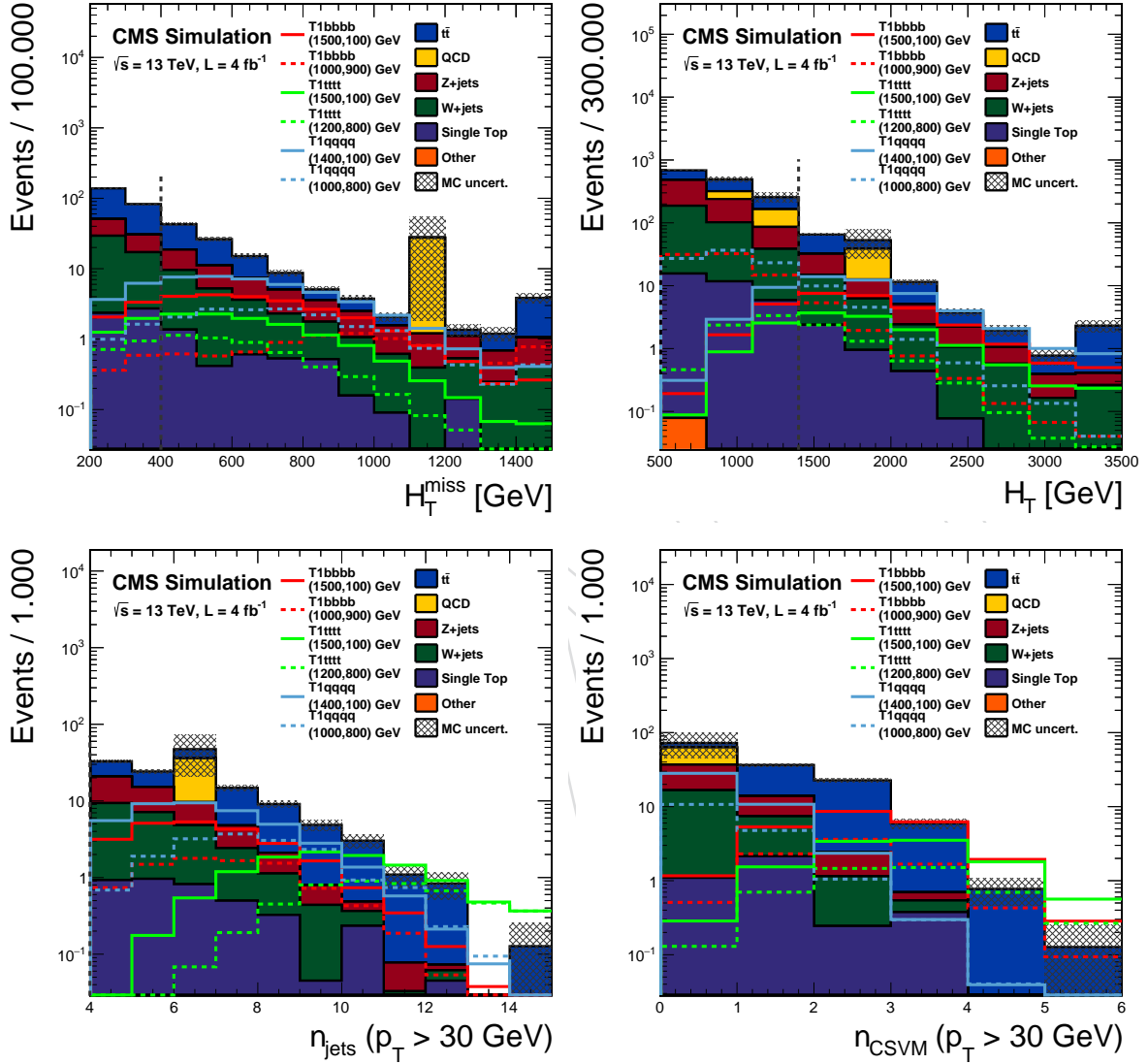


Figure 34: $N-1$ plots of the binning variables for the most sensitive region for the non-compressed $T1qqqq$ point, $m_{\tilde{g}} = 1400 \text{ GeV}$, $m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$ (solid blue line). The boundaries for this region, marked in the plots by dashed gray vertical lines, are $H_T > 1400 \text{ GeV}$ and $H_T^{\text{miss}} > 400 \text{ GeV}$, and $N_{\text{jet}} \geq 4$.

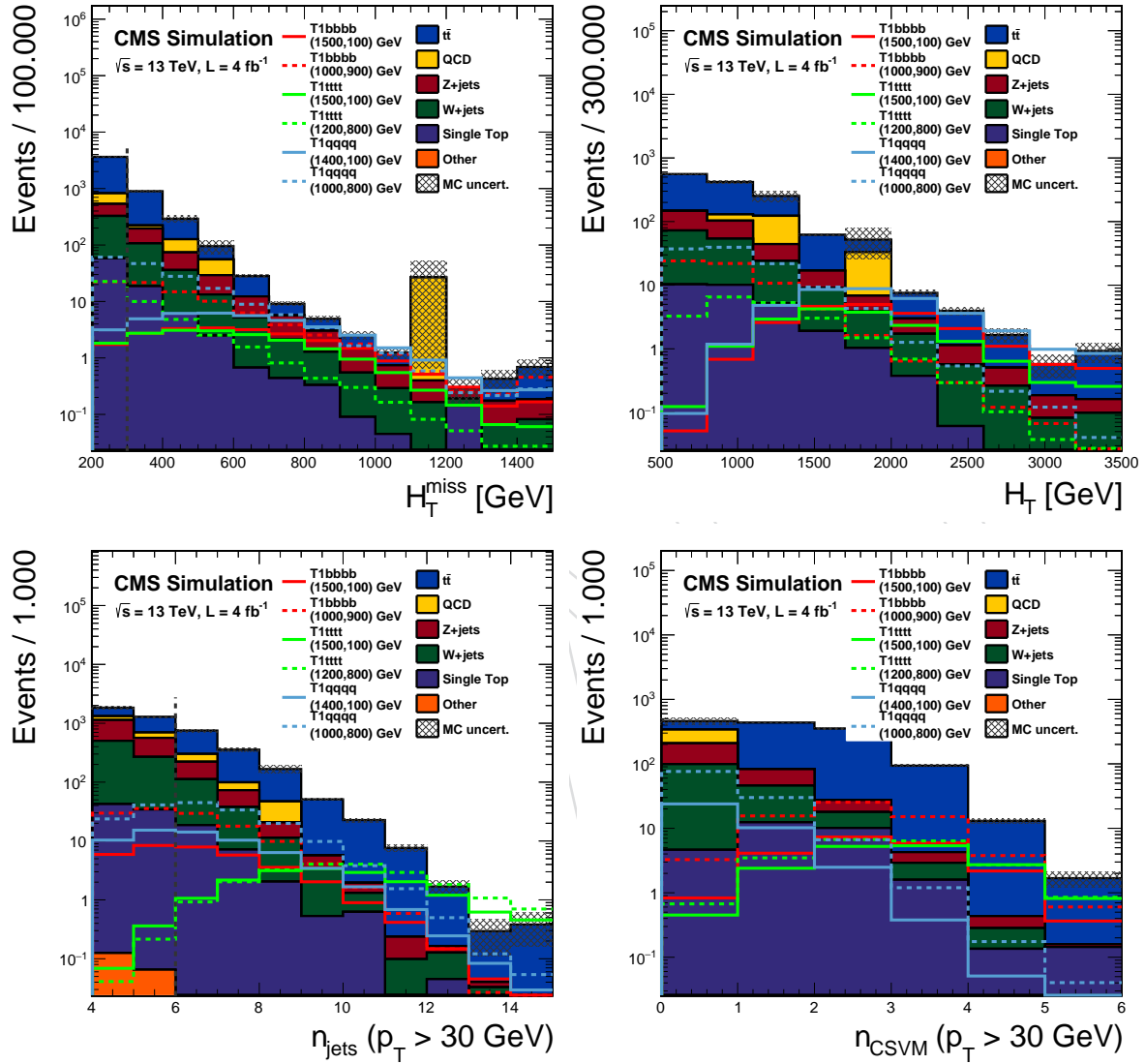


Figure 35: $N - 1$ plots of the binning variables for the most sensitive region for the compressed $T1qqqq$ point, $m_{\tilde{g}} = 1000$ GeV, $m_{\tilde{\chi}_1^0} = 800$ GeV (dashed blue line). The boundaries for this region, marked in the plots by dashed gray vertical lines, are $H_T > 500$ GeV, $H_T^{\text{miss}} > 300$ GeV, and $N_{\text{jet}} \geq 6$.

B Comparison of QCD samples

As part of the Phys14 MC production campaign, two sets of QCD samples were produced:

1. Samples binned in generator-level H_T , generated with Madgraph [13]
2. Samples binned in generator-level \hat{p}_T , generated with Pythia 8 [14]

The Phys14 Madgraph samples more closely resemble the QCD MC that will be used throughout the upcoming run, but they have insufficient statistics for studying the closure of the QCD background estimation methods described in Section 6.2. Thus we use the Phys14 Pythia 8 samples in those studies. Here we present a few basic comparisons between the Phys14 Madgraph and Pythia 8 samples and their CSA14 (CMSSW_7_0_6_patch1) versions.

Table 22 includes a baseline event selection cutflow comparison of the four sets of samples. After baseline selection, in particular the QCD-suppressing cut on $\Delta\hat{\phi}_{\min}$, the expected yields from the two sets of Phys14 samples agree within statistical uncertainty. The expected yields from the CSA14 samples are consistently larger than the corresponding Phys14 yields. These disagreements may be explained by known differences in reconstruction between CMSSW_7_0_6_patch1 and CMSSW_7_2_0_patch1, including the improved mitigation of out-of-time pileup effects in CMSSW_7_2_0_patch1 [15].

Table 22: Comparison of cutflow and expected yields at 4 fb^{-1} for QCD samples after baseline selection. Note that the CSA14 samples are reconstructed with CMSSW_7_0_6_patch1, while the Phys14 samples are reconstructed with CMSSW_7_2_0_patch1.

Cut	CSA14 Pythia	CSA14 Madgraph	Phys14 Pythia	Phys14 Madgraph
Start	33065553 ± 18251	2792038900 ± 3315130	33065553 ± 21286	2792038930 ± 1642779
$H_T > 500 \text{ GeV}$	30667176 ± 17512	94910354 ± 282016	30196895 ± 20254	65239789 ± 71885
$N_{\text{jet}} \geq 4$	15725313 ± 12459	60902673 ± 249387	13472302 ± 13420	36841661 ± 62033
$H_T^{\text{miss}} > 200 \text{ GeV}$	51028 ± 598	30294 ± 5897	36371 ± 532	23871 ± 1123
Muon veto	50843 ± 597	30276 ± 5897	36322 ± 532	23850 ± 1123
Electron veto	34988 ± 466	29946 ± 5896	35479 ± 525	23332 ± 1119
$\Delta\hat{\phi}_{\min} > 4$	6018 ± 209	7600 ± 4126	4892 ± 216	5144 ± 1039

C Details of the H_T^{miss} extrapolation method

The distributions of $(1 - \cos \Delta\theta_T)/2$ and the lost lepton / control sample lepton ratio as a function of p_T^W in the analysis bins of N_{jet} and H_T are shown in Figures 36 and 37.

The details of how the lost lepton H_T^{miss} distribution is constructed from the single lepton control sample are as follows. For each single lepton control sample event, the $(1 - \cos \Delta\theta_T)/2$ PDF is translated into a H_T^{miss} PDF using the relation $H_T^{\text{miss}} \approx p_T^W (1 - \cos \Delta\theta_T)/2$. The integral of the H_T^{miss} pdf for the high (medium) H_T^{miss} bin for event i is I_{ijk}^H (I_{ijk}^M), where event i falls in N_{jet} , H_T bin combination j and p_T^W bin k . The contribution of the event is weighted by the lost / control ratio R_{jk}^{LC} . Summing over all single lepton control sample events gives

$$S_j^H = \sum_{i,k} R_{jk}^{LC} I_{ijk}^H \quad \text{and} \quad S_j^M = \sum_{i,k} R_{jk}^{LC} I_{ijk}^M \quad (21)$$

and the H_T^{miss} ratio $R_j^{\text{MHT}} = S_j^H / S_j^M$.

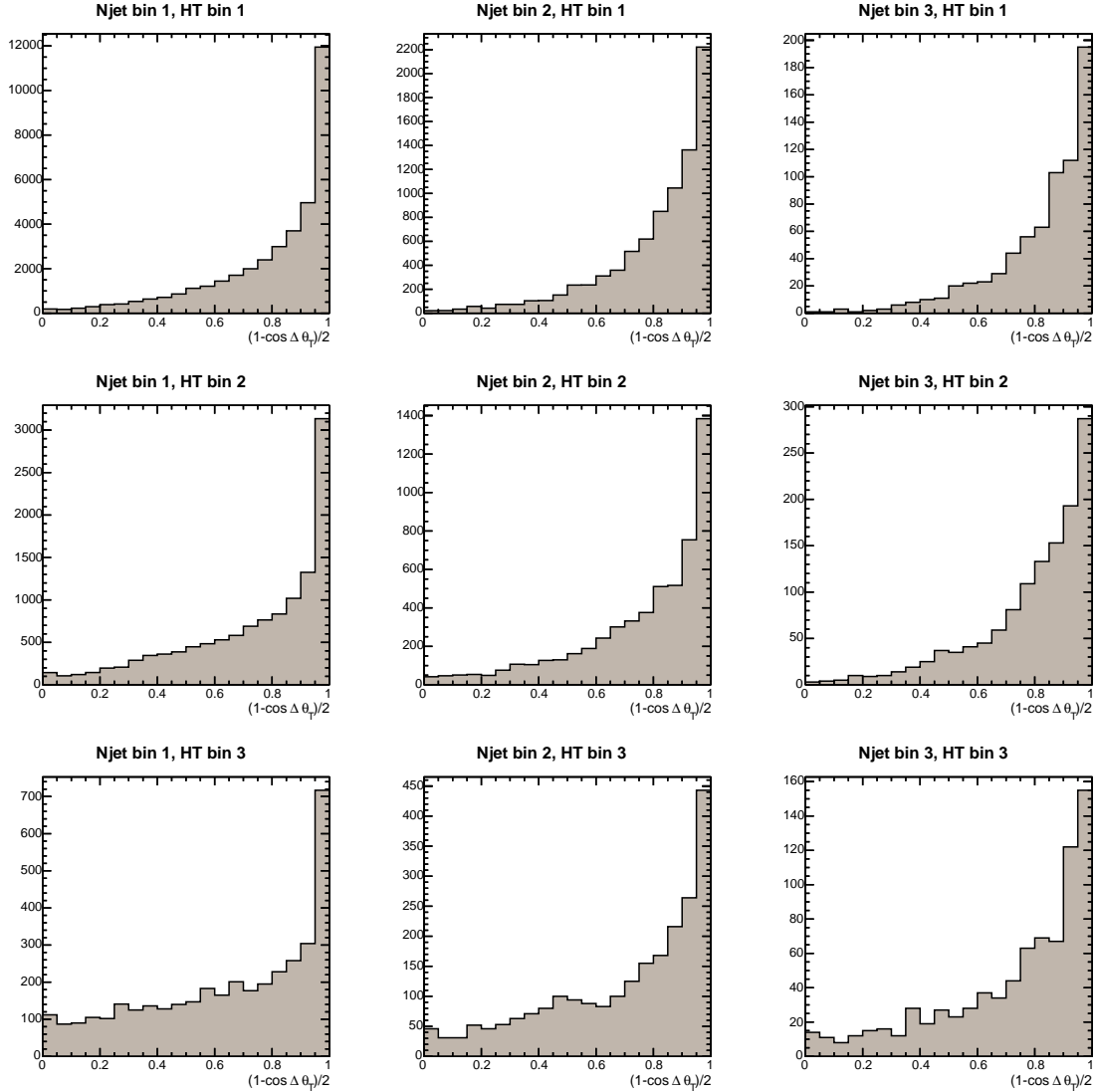


Figure 36: Distribution of $(1 - \cos \Delta\theta_T)/2 \approx H_T^{\text{miss}} / p_T^W$ for lost leptons in the analysis bins of N_{jet} and H_T for the $t\bar{t}$ Monte Carlo sample.

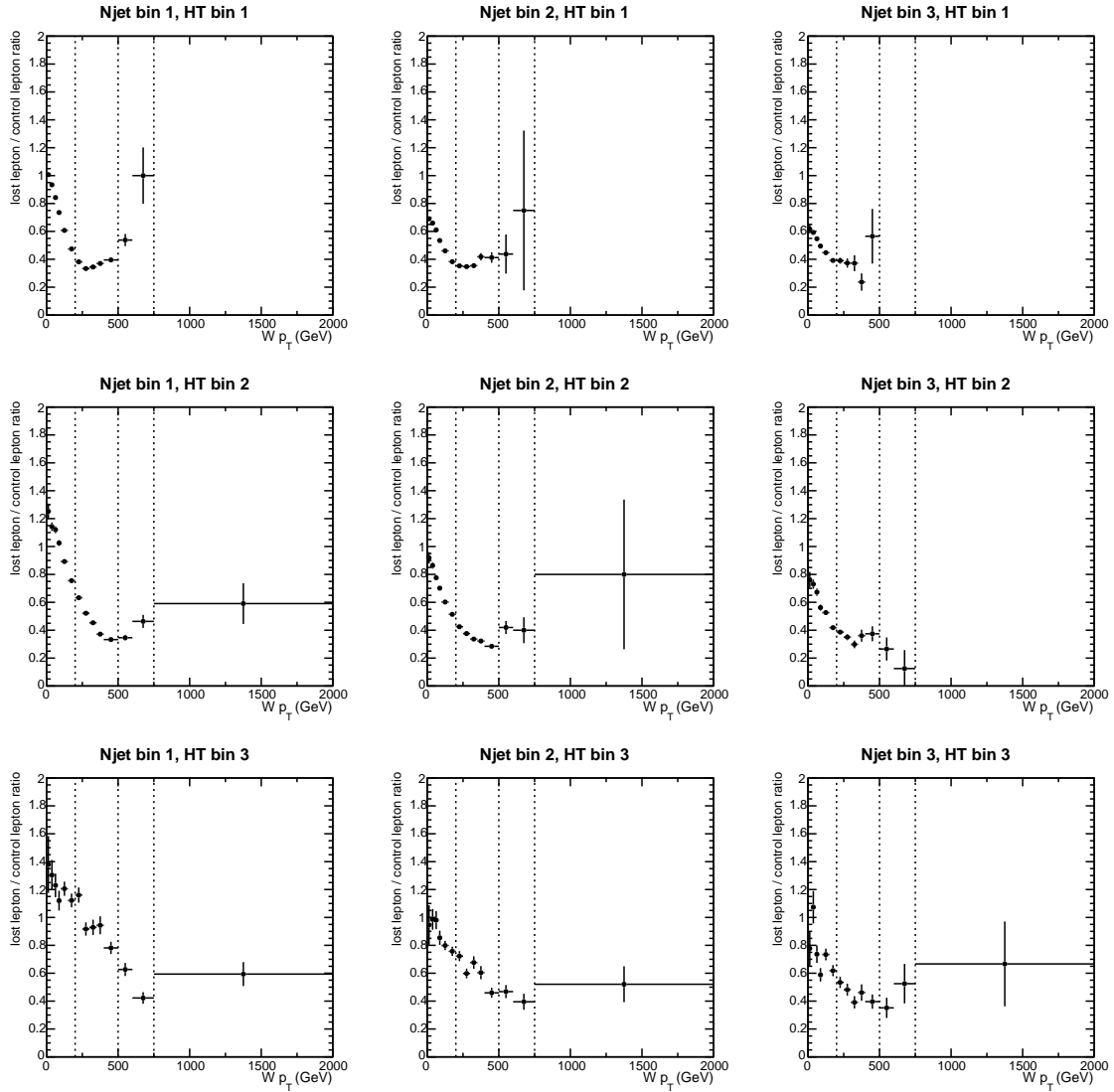


Figure 37: The lost lepton / control sample lepton ratio as a function of p_T^W in the analysis bins of N_{jet} and H_T for the $t\bar{t}$ Monte Carlo sample. The vertical dashed lines are at the boundaries of the analysis H_T^{miss} bins. The points below 200 GeV are not used in the analysis.

The uncertainty on R_j^{MHT} due to the single lepton control sample statistics and the uncertainty on the R_{jk}^{LC} is given by

$$\sigma_{R_j^{\text{MHT}}}^2 = \sum_{i,k} \left(\frac{R_{jk}^{\text{LC}} I_{ijk}^H}{S_j^M} - \frac{S_j^H}{(S_j^M)^2} R_{jk}^{\text{LC}} I_{ijk}^M \right)^2 + \sum_k \left(\frac{\sum_i I_{ijk}^H}{S_j^M} - \frac{S_j^H}{(S_j^M)^2} \sum_i I_{ijk}^M \right)^2 \sigma_{R_{jk}^{\text{LC}}}^2 \quad (22)$$

which is approximately

$$\sigma_{R_j^{\text{MHT}}}^2 \approx (R_j^{\text{MHT}})^2 \frac{1}{N_j^H} + \sum_k \left(\frac{\sum_i I_{ijk}^H}{S_j^M} - \frac{S_j^H}{(S_j^M)^2} \sum_i I_{ijk}^M \right)^2 \sigma_{R_{jk}^{\text{LC}}}^2, \quad (23)$$

536 where N_j^H is the number of single lepton control sample events with a non-zero contribution to
 537 the high H_T^{miss} bin sum S_j^H , which is the number of events with $p_T^W > H_T^{\text{miss}}_{\text{low}}$, where $H_T^{\text{miss}}_{\text{low}}$
 538 is the low edge of the high H_T^{miss} bin. The first term dominates for bins with relatively small
 539 N_j^H , giving the somewhat expected result of $\sigma_{R_j^{\text{MHT}}} / R_j^{\text{MHT}} \approx 1 / \sqrt{N_j^H}$. In the analysis, the full
 540 expression given in Eqn.22 will be used.