CMS Physics Analysis Summary

Contact: cms-pag-conveners-susy@cern.ch 2016/08/04

Search for *R*-parity-violating SUSY in final states with zero or one lepton and large multiplicity of jets and b-tagged jets

The CMS Collaboration

Abstract

Preliminary results are reported from a search for physics beyond the standard model in proton-proton collisions at $\sqrt{s} = 13$ TeV, focusing on the signature of large multiplicity of jets and b-tagged jets, in a final state with zero or one reconstructed lepton. The data sample corresponds to an integrated luminosity of 2.7 fb^{-1}, recorded by the CMS experiment at the Large Hadron Collider. The results are interpreted in terms of limits on the parameter space for the *R*-parity-violating supersymmetric extension of the standard model in a benchmark model of gluino pair production where each gluino decays via $\widetilde{\text{g}} \to \text{tbs}$. Assuming the gluino decays solely to tbs, gluino masses smaller than 1360 GeV are excluded at a 95% confidence level.

1 Introduction

Searches for physics beyond the Standard Model (SM) are motivated by several considerations, ranging from strong astrophysical evidence for dark matter [1] to theoretical problems associated with explaining the observed particle masses and with maintaining the mass hierarchies in the presence of quantum corrections [2, 3]. Thus, in spite of its success in describing a vast range of phenomena, the SM is almost certainly incomplete as a description of fundamental particles and their interactions. Supersymmetry (SUSY) can provide a solution to some of these problems [4–11].

We search for new physics in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, focusing on the signature of at least four energetic jets, at least one of which must be b-tagged. The analysis is performed on a sample of proton-proton collisions with an integrated luminosity of $2.7\,\text{fb}^{-1}$, collected by the CMS experiment during 2015 at the Large Hadron Collider (LHC).

In many SUSY models, a multiplicatively conserved quantum number, R-parity, is introduced [12]. The R-parity of a particle is defined as $(-1)^{2s+3(B-L)}$, where s is the spin of the particle and B and L are its baryon and lepton number, respectively. If R-parity is conserved, SUSY particles are produced in pairs, and SUSY particle decay chains end at the lightest supersymmetric particle (LSP), typically a neutralino. As the LSP is stable in these models, and interacts only via the weak interaction, the presence of two neutralinos in the final state generally leads to a signature of large missing transverse energy $E_{\text{T}}^{\text{miss}}$. The absence of any evidence for such signatures provides an experimental motivation for a search for R-parity-violating (RPV) SUSY, as does the desire for complete coverage of the model space. A review of R-parity-violating SUSY can be found in Ref. [13].

There is no purely theoretical reason that R-parity should be conserved. If R-parity-violation is permitted, additional terms in the superpotential are allowed. The R-parity-violating superpotential is

$$W = \frac{1}{2}\lambda^{ijk} L_i L_j \bar{\text{e}}_k + \lambda^{'ijk} L_i Q_j \bar{\text{d}}_k + \mu^{'i} L_i H_u + \frac{1}{2}\lambda^{''ijk} \bar{\text{u}}_i \bar{\text{d}}_j \bar{\text{d}}_k. \tag{1}$$

Here L_i, Q_j, and H_u are $SU(2)$ doublets corresponding to leptons, quarks, and the Higgs boson, respectively. The fields $\bar{\text{e}}_k$, $\bar{\text{u}}_i$, and $\bar{\text{d}}_j$ are the charged lepton, up-type quark and down-type quark $SU(2)$ singlets. Color indices are suppressed, and Latin letters denote family indices. Other discrete symmetries that violate only baryon number or lepton number share the same phenomenological virtues as R-parity, and lead to non-zero terms in Eq. 1, even if R-parity conservation is not imposed [14].

In this analysis, we focus on a particular model of R-parity-violation, minimal flavor-violating (MFV) SUSY [15]. This model assumes that R-parity-violating couplings arise from the SM Yukawa couplings, Y_{d} and Y_{u} (but not $Y_{\text{d}}^{\dagger}$ and $Y_{\text{u}}^{\dagger}$). This assumption results in a baryon-number-violating term in the superpotential, the last term in Eq. 1, taking the form:

$$W_{\text{BNV}} = \frac{1}{2} w'' \left(Y_{\text{u}} \bar{\text{u}}\right) \left(Y_{\text{d}} \bar{\text{d}}\right) \left(Y_{\text{d}} \bar{\text{d}}\right) \tag{2}$$

where $Y_{\text{u}} = (1/v_{\text{u}}) V_{\text{CKM}}^{\dagger} \text{diag}(m_{\text{u}}, m_{\text{c}}, m_{\text{t}})$ and $Y_{\text{d}} = (1/v_{\text{d}}) \text{diag}(m_{\text{d}}, m_{\text{s}}, m_{\text{b}})$. The vacuum expectation values of the up- and down-type Higgs are denoted v_{u} and v_{d}, respectively.

As the RPV couplings are proportional to the Yukawa couplings in this model, heavier generations have larger couplings. In these models, the absence of any observations of baryon

and lepton number violation involving the lightest two generations is a result of the small mass of those states. In the MFV scenario considered here, the gluino decays primarily via $\widetilde{g} \to \bar{t}\widetilde{t} \to \bar{t}\bar{b}\bar{s}$ (charge conjugate reactions are implied throughout this document). Color conservation constrains the coupling λ''^{ijk} to be anti-symmetric in the last two indices, excluding the possibility of gluino decays to tbb. The coupling of a top squark to a bottom and a strange quark therefore provides the largest allowed coupling. Pair production of gluinos that decay in this way is the primary focus of this analysis.

The simplified model that is used for interpretation of these results makes several assumptions about the SUSY mass spectrum. It assumes that squarks other than the top squark are much heavier than the gluino so that their effect is negligible. The top squark is assumed to be heavier than the gluino and therefore is off-shell in the gluino decay. As the off-shell decay of the top squark in this channel results in a three-body gluino decay, constraints from searches for dijet resonances (i.e. $\widetilde{t} \to \bar{b}\bar{s}$) are not applicable to this scenario. However, large jet and b-jet multiplicity is a very common feature of many models of physics beyond the standard model [16, 17], and this inclusive search is generically sensitive to those scenarios.

There have been several searches [18–21] at CMS for R-parity-violating SUSY. In Ref. [21], a strategy similar to that of this search was used separately for the all-hadronic and one-lepton final states. Gluinos decaying to tbs were excluded at 95% confidence level for masses less than 0.98 TeV and 1.03 TeV in the all-hadronic and one-lepton analyses, respectively. In this document, we study the all-hadronic and one-lepton modes in a combined analysis. A combination of the two channels allows for an improved constraint on background normalizations, and provides improved sensitivity with the integrated luminosity delivered in 2015.

This analysis searches for an excess of events with a large number of b-tagged jets, N_b, in signal regions determined as a function of the jet multiplicity, N_{jet}, and the sum of the masses of large radius jets, M_J. The quantity M_J was proposed in phenomenological studies [22–24] and was used by ATLAS for searches in the all-hadronic signature [25, 26] and by CMS for searches in the single-lepton signature [27]. While the M_J and N_{jet} distributions may not be well modeled by simulation, we expect that the distribution of the number of b-tagged jets, N_b, can be reliably modeled within a given M_J and N_{jet} region. The reason for that expectation is that the leading mechanisms for producing b quarks, QCD pair production and top quark decay, are well modeled in simulation. Other production modes are single b quark production in flavor-excitation modes and gluon splitting to $b\bar{b}$, which both generate only small contributions. Mistagging of light parton jets generates most of the b-tagged jets in the $N_b > 2$ region, with a rate that can be constrained from the data.

The background normalizations are determined in each bin of M_J and N_{jet} in a fit to the N_b distribution. The normalizations are determined separately in these bins for two reasons. The shape of the N_b distribution changes significantly as a function of the event kinematics, and is particularly sensitive to the number of jets in the event, which may not be modeled perfectly. Also, by binning the analysis in a large number of different kinematic regions, it is possible to demonstrate that the dependence on jet p_T of the b-tagging efficiency for true b partons, and the mistagging rate for light partons, is modeled correctly. As a result, the shape of the N_b distribution becomes limited by experimental uncertainties, mostly those related to b-tagging efficiency, rather than by theoretical uncertainties.

2 Event selection

The key issue in the data analysis is the suppression and estimation of the dominant SM backgrounds, which arise primarily from the production of $\mathrm{t\bar{t}}$ pairs and QCD multijet events (QCD). Backgrounds from $(\mathrm{W,Z})$+jets production, from single-top quark production, and from $\mathrm{t\bar{t}}(\mathrm{W,Z,H,t\bar{t}})$ production are also considered but these contributions are very small, less than a few percent.

Background suppression is achieved through the application of multiplicity requirements on the jets reconstructed in the event, together with threshold requirements. Four jets with transverse momentum $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ are required, and H_{T}, the scalar sum of all jet p_{T}, is required to be greater than 1.5 TeV (1.2 TeV) for the fully hadronic (one-lepton) final state. We further suppress the standard model backgrounds with a baseline selection on the sum of the masses of large radius jets, $M_{\mathrm{J}} > 500\,\mathrm{GeV}$, as described in detail later in this section.

After these requirements, QCD multijet events emerge as the largest contribution to the background in the fully hadronic event sample, while $\mathrm{t\bar{t}}$ events are the dominant background in the one-lepton sample and are nonnegligible in the large M_{J} and N_{jet} regions of the fully hadronic sample.

2.1 Data samples and triggers

The data sample is collected with a trigger requiring at least 800 GeV of H_{T}. The trigger efficiency is determined using prescaled triggers with looser H_{T} requirements, and is cross-checked with an independent trigger requiring a single muon. The trigger efficiency is consistent with 100% for all kinematic regions used in the analysis. The data sample corresponds to an integrated luminosity of $2.7\,\mathrm{fb}^{-1}$.

The modeling of backgrounds is done with Monte Carlo (MC) generated event samples that are processed with a GEANT4-based simulation [28] of the CMS detector and reconstructed with the same algorithms as the data. The background predictions use these MC samples, with corrections derived from data. The dominant backgrounds arise from QCD and $\mathrm{t\bar{t}}$ events, and are generated at leading order with MADGRAPH5_AMC@NLO 2.2.2 [29] (hereafter referred to as MADGRAPH). Smaller backgrounds from leptonic and hadronic decays of W and Z bosons are also generated with MADGRAPH, while the single top backgrounds are generated with POWHEG [30] or MADGRAPH at next-to-leading order. These sub-dominant backgrounds are normalized according to the most precise cross section calculation available. Parton showering, hadronization, and the effects of the underlying event are simulated with PYTHIA 8.205 [31].

Representative RPV SUSY signal models are generated with MADGRAPH for gluino masses ranging from 0.75 to 1.5 TeV, and they are processed with a parametric fast simulation [32]. In these signal samples, we assume that the gluino is the LSP and that the masses of all other particles are large enough that they have no effect. However, the top squark mass is assumed to be small enough that the gluino decays promptly.

2.2 Event selection

The offline pre-selection requirements are designed to be simple and robust. Reconstructed vertices must satisfy a set of quality requirements including $|z| < 24\,\mathrm{cm}$ and $\rho < 2\,\mathrm{cm}$, where z and ρ are the longitudinal and transverse distances of the beam spot with respect to the nominal CMS interaction point. The vertex with the largest Σp_{T}^2, summed over its constituent charged particle tracks, is chosen as the primary vertex.

Jets are reconstructed using the particle-flow (PF) algorithm [33, 34] which combines information from all components of the detector. The clustering is performed using the anti-k_t clustering algorithm [35, 36] with a size parameter R of 0.4. The effect of multiple proton-proton interactions in a single bunch crossing, referred to as pileup, is mitigated by considering only those charged particle tracks that originate from the primary vertex, and reweighting the number of primary interactions per bunch crossing to match the data. Jet candidates are required to satisfy quality criteria that suppress noise and spurious energy deposits and to satisfy $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$. These jets, as well as any identified electrons or muons, are then further clustered with an anti-k_t algorithm with $R = 1.2$ using the FASTJET package [36]. The sum of the masses of these large-radius jets is called M_J. We use here the same definition of M_J as that of Ref. [27].

The b tagging selection used is the medium working point of the Combined Secondary Vertex v2 discriminator with inclusive vertex finding (CSV) [37]. The medium working point corresponds to a b tagging efficiency (light-jet mistag rate) of $\approx .60$ (≈ 0.01) for low p_T central jets; the efficiency is reduced at high p_T and $|\eta|$.

Muons and electrons are identified with the following selections. Muons must have $p_T > 20\,\text{GeV}$, $|\eta| < 2.4$, and a distance of closest approach to the beamline in the direction transverse to the beam of $|d_0| < 0.2\,\text{cm}$, and along the beam of $|d_z| < 0.5\,\text{cm}$. Electrons in the barrel region ($|\eta| \leq 1.479$) are required to have $p_T > 20\,\text{GeV}$, $|d_0| < 0.0118\,\text{cm}$, $|d_z| < 0.373\,\text{cm}$, and the ratio of the energy in the HCAL to that in the ECAL less than 0.0876. Electrons in the endcap ($1.479 < |\eta| < 2.5$) are required to have $p_T > 20\,\text{GeV}$, $|d_0| < 0.0739\,\text{cm}$, $|d_z| < 0.602\,\text{cm}$, and the ratio of the energy in the HCAL to that in the ECAL less than 0.0678.

A measure of lepton isolation is determined by calculating the sum of the p_T of PF candidates, excluding the lepton itself, within an $\eta - \phi$ region around the lepton. In order to reduce the rate of accidental overlaps between jets and a prompt lepton in high-multiplicity or highly boosted events, the radius of this region varies with lepton p_T as follows:

$$R = \begin{cases} 0.2, & p_T^{\text{lep}} \leq 50\ \text{GeV} \\ \frac{10\ \text{GeV}}{p_T^{\text{lep}}}, & 50 < p_T^{\text{lep}} < 200\ \text{GeV} \\ 0.05, & p_T^{\text{lep}} \geq 200\ \text{GeV}, \end{cases} \tag{3}$$

The resulting p_T sum, divided by the p_T of the lepton, is required to be less than 0.1 (0.2) for muons (electrons).

Events are required to have at least one good primary proton-proton vertex, at least four jets satisfying $p_T \geq 30\,\text{GeV}$ and $|\eta| \leq 2.4$, and exactly zero or one isolated muons or electrons. At least one b-tagged jet is required in order to suppress W+jets events.

The acceptance times efficiency for the signal to enter the kinematic regions used in the analysis ranges from 71.0% for a gluino mass of 1.0 TeV to 88.5% for a gluino mass of 1.4 TeV.

3 Analysis overview

The analysis is performed in bins of N_{jet}, M_J, and N_{lep}, as shown in Table 1. The N_{jet} bins are 4–5, 6–7, 8–9 and ≥ 10; the M_J bins are $500 < M_J \leq 800\,\text{GeV}$ and $M_J > 800\,\text{GeV}$; and N_{lep} is either zero or one. Signal events in the fully hadronic (one-lepton) channel have ten (eight) quarks.

Table 1: A representation of the bins of the analysis after the baseline selection. The bins labeled "CR" are background dominated and serve as a control region for the fit, while the bins labeled "SR" correspond to signal regions. Even within the high N_{jet} and M_J bins, events with $N_b \leq 2$ are background dominated and act as control regions for the fit.

N_{lep}	M_J [GeV]	N_{jet}			
		4–5	6–7	8–9	≥ 10
0	500–800	CR	CR	SR	SR
	> 800	CR	CR	SR	SR
1	500–800	CR	SR	SR	
	> 800	CR	SR	SR	

Therefore the low N_{jet} bins (4–7 in the fully hadronic case and 4–5 in the one-lepton case) have only small contributions from signal events and are used to study the data prior to examining the regions with larger N_{jet}. In the $N_{lep} = 1$ bin, the highest two N_{jet} bins are merged, as $t\bar{t}$ events in which one W boson decays to leptons typically have two fewer jets than those where both W bosons decay hadronically. The H_T selection is $H_T > 1.5\,\text{TeV}$ ($H_T > 1.2\,\text{TeV}$) in the hadronic (single-lepton) signal regions. This H_T requirement is much higher than in analyses that consider E_T^{miss} because in the scenario considered here there is no neutralino to carry away energy.

The dominant background to the fully hadronic signal regions arises from QCD multijet events with smaller backgrounds mainly from $t\bar{t}$ production; therefore, the modeling of the QCD component is crucial. Corrections to the QCD background prediction are derived from data, after subtracting non-QCD background, to predict the spectrum of b tagged jets, N_b. There are three main concerns for this procedure: the simulation of the N_{jet} and M_J distributions, the flavor composition, and the b quark production mechanism.

The N_{jet} and M_J distributions are difficult to model properly in MC. This difficulty is avoided by binning the sample in N_{jet} and M_J and then fitting the N_b distribution in data. This determines the background normalizations in each (N_{jet}, M_J) bin because the $N_b \leq 2$ bins are background dominated. Once these normalizations are taken from data, the remaining corrections are small. The shape of the N_b distribution can be predicted reliably because the b quark production mechanisms and the b tagging efficiencies and mistagging rates can be understood with the data control samples. The fit across N_{jet}, M_J, and N_b includes these corrections and can further reduce the systematic uncertainties on the corrections because the low (N_{jet}, M_J) bins are background dominated even at high N_b. The signal sensitivity comes from the $N_b \geq 3$ bins at high (N_{jet}, M_J), as illustrated in Table 1.

The primary source of additional b quarks, beyond the two b quarks from direct $b\bar{b}$ production or decays of pairs of top quarks, arises from gluon splitting, $g \to b\bar{b}$. The rate for this process may not be simulated properly, so we constrain it in control samples with two nearby b-tagged jets. Events with low $\Delta R_{b\bar{b}}$, where $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is measured between the two b jets, are used as a probe of gluon splitting events, while the region of high $\Delta R_{b\bar{b}}$ is used to normalize the control samples [38].

Another challenge for modeling b quark production arises from uncertainties in the flavor composition of the QCD multijet background from a mismodeling of either the flavor composition at the generator level or b tagging efficiencies. The b tagging efficiencies are corrected as a function of $|\eta|$ and p_T [37], and the flavor composition is corrected with the results of a fit of the b tagging discriminant distribution in a QCD-dominated low N_{jet} sideband.

With this procedure, it is possible to make a data-driven estimation of the background in the variables N_{jet}, M_{J}, and N_{b}. The signal yield is extracted in a fit of the N_{b} distribution in the signal regions.

4 Flavor composition correction in the fully hadronic sample

In contrast to $\mathrm{t\bar{t}}$ events, in which the flavor composition of the jets is largely determined by the well known top quark and W boson branching fractions, in QCD multijet events the flavor composition is not strongly constrained. To ensure that the QCD simulation has the appropriate flavor composition, events are reweighted to match the flavor composition seen in data. This reweighting is performed by fitting the distribution of the CSV discriminant for all jets in the 0.89–1.00 range, where the lower bound of this range corresponds to the tagging threshold for the medium working point of the discriminator.

The fit is performed including the statistical uncertainty on the MC prediction as nuisance parameters in the fit according to the Barlow-Beeston prescription [39]. To reduce signal contamination, only the low N_{jet} = 6–7 events are used.

Templates are generated according to the number of jets passing the event selection that are matched using generator information to charm or bottom partons. There are three types of templates for the QCD multijet background:

- Events with at least one generator-level b quark
- Events with no generator-level b quarks but at least one generator-level c quark
- Events with no generator-level b or c quarks

The overall normalization of the QCD component is first fixed to the data yield minus the yield of non-QCD events. The fraction of the total number of events that have at least one true b jet passing the jet selection is denoted f_{b}. The fraction of the total number of events that have no true b jets but at least one true c jet passing the jet selection are denoted f_{c}. The fraction of the remainder is denoted f_{light}. The fit provides new fractions f'_{i} where the fit fractions are defined as

$$f'_{\text{i}} = \frac{n_{\text{i}}}{n_{\text{b}} + n_{\text{c}} + n_{\text{light}} + n_{\text{non-QCD}}} \tag{4}$$

where n_{b} and n_{c} are the fitted yields of the bottom and charm components, while n_{light} and $n_{\text{non-QCD}}$ are the fixed yields of light parton and non-QCD (mainly $\mathrm{t\bar{t}}$) events. The index i corresponds to b, c, light parton, and non-QCD events. By construction, $\sum_i f_i = \sum_i f'_i = 1$.

As a consequence of the small light parton yield, and the similarity between the light parton and charm quark shapes, we fix the light parton contribution in the fit and only allow the charm contribution to float. This procedure results in a conservative estimate of the uncertainty because charm quarks have a larger mistag rate and therefore assuming that all of the difference is due to the charm component generates a larger change to the N_{b} distribution than if the light parton fraction is varied by the same amount. Fits have been performed with 100% variations of the light parton component to determine that no additional uncertainty is necessary as a result of fixing this component.

We obtain the f'_{i} factors of Eq. 4 from a fit of the CSV distribution in a control sample defined by $6 \leq N_{\text{jet}} \leq 7$, and $N_{\text{b}} \geq 2$; the result is shown in Fig. 1. The chi-square per degree of freedom, χ^2/ndof=18.2/20, indicates that the fit models the data well. To check for dependence

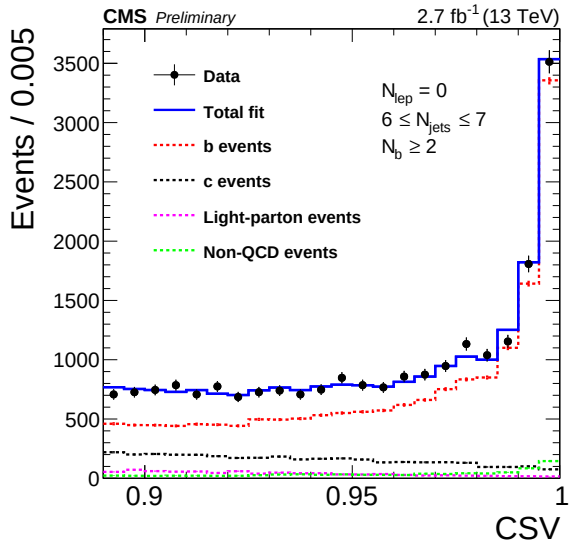

Figure 1: Fit to the CSV distribution in data in the CSV > 0.89 region for $H_T > 1.5$ TeV, $N_b \geq 2$, and $6 \leq N_{jet} \leq 7$. Error bars indicate the statistical uncertainties in the simulated samples (most of which are smaller than the line width) and the data.

of the fit on the number of jets, the event kinematics, or the discriminant shape modeling, we repeat the fit with different numbers of jets ($N_{jet} \geq 8$), alternate selections on M_J, and after removing the four highest or lowest CSV bins, respectively. The resulting variations provide systematic uncertainties on the central value of the fit parameters. Based on these fits, a per-event weight $w_{b-jet} = f'_{flavor}/f_{flavor}$ is used to reweight the QCD samples. These weights are $0.98 \pm 0.01 \pm 0.07$ and $1.10 \pm 0.06 \pm 0.34$ for b and c events, respectively.

5 Gluon splitting systematic uncertainty

Jets containing b quarks are produced in three different ways in QCD multijet events: pair production ($q\bar{q} \to b\bar{b}$), flavor excitation ($bq \to bq$), and gluon splitting ($g \to b\bar{b}$). The first two processes are important primarily in the initial hard scatter while the second is suppressed by the small intrinsic b quark content of the proton. The rate of gluon splitting in the collinear region is difficult to model [38], making an estimation of the systematic uncertainty from data crucial. This uncertainty is determined using the $\Delta R_{b\bar{b}}$ distribution, where ΔR is computed between the two b-tagged jets in events with $N_b = 2$. Events with $N_b > 2$ are not used in this procedure to avoid possible effects of signal contamination.

The measured $\Delta R_{b\bar{b}}$ distribution is compared to simulation for the all-hadronic sample in Fig. 2. This distribution is normalized in the high $\Delta R_{b\bar{b}}$ region, $\Delta R_{b\bar{b}} \geq 2.4$, which is dominated by pair-production, for which the shape is known to be modeled well [38], although the H_T and N_{jet} dependence is not expected to agree without this normalization. The difference between data and simulation in the $\Delta R_{b\bar{b}} \leq 1.6$ region is then used as a measure of gluon splitting mis-modeling.

The deviation from unity of the data/simulation ratio in the low $\Delta R_{b\bar{b}}$ region, combined with the uncertainty of the test, is used to determine a systematic uncertainty for gluon splitting, f_{gs}. The data/simulation ratio is consistent with unity except for an 18% difference in the $N_{jet} = 4$–5 region. The $\pm 1\sigma$ variations of the N_b distribution are formed by weighting events containing a

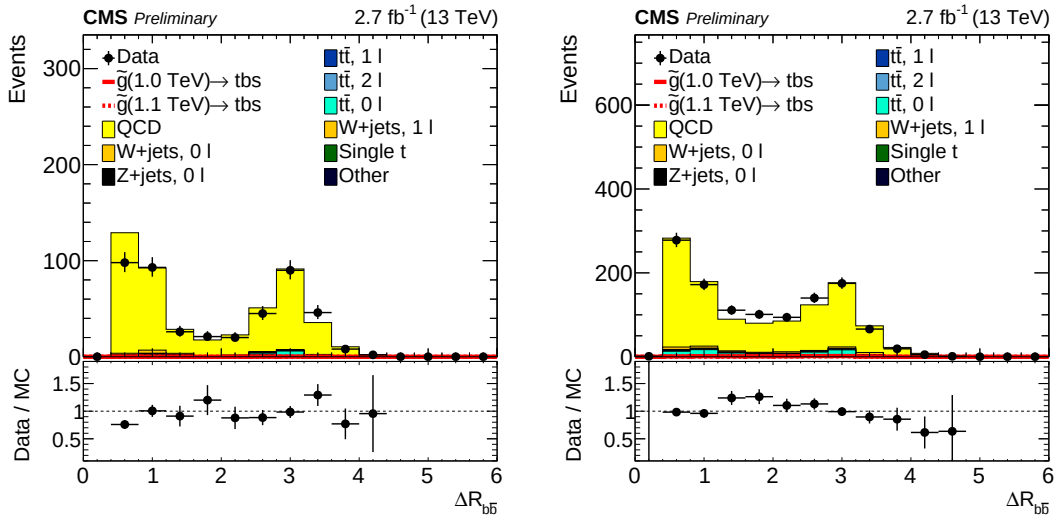

Figure 2: Distributions of $\Delta R_{b\bar{b}}$ after normalization in the $\Delta R_{b\bar{b}} \geq 2.4$ region for selections of N_{jet} = 4–5 (left) and N_{jet} = 6–7 (right). Events are required to have zero leptons, with $M_J > 500\,GeV$ and $N_b = 2$. As shown by the red lines, the contribution from signal events is negligible for the gluino masses considered here.

gluon splitting to $b\bar{b}$ or $c\bar{c}$ by a factor $1 \pm f_{gs}$. Assuming that the deviation in the $\Delta R_{b\bar{b}}$ distribution is entirely due to gluon splitting to heavy flavor, as opposed to mistagged light parton jets, is conservative because such events generate the largest change in the N_b distribution. As will be shown in the next section, the final effect of the gluon splitting systematic uncertainty is numerically small both because large mismodelings are not evident and because the contribution to gluon splitting in the high N_b bins is small compared to the contribution from mistagged light parton jets.

6 Systematic uncertainties

6.1 Background systematic uncertainty

Systematic uncertainties arise primarily from the scale factors used to correct the b tagging efficiencies. The b-tagging efficiency and its uncertainty are derived from QCD and $t\bar{t}$ events in less extreme regions of phase space than those considered in this analysis. This leads to larger uncertainties, for both signal and background predictions, than the uncertainties arising from the QCD flavor composition and gluon splitting described in Section 4 and Section 5, respectively.

Because the N_b distributions are determined from simulated events, it is important to fully explore the possible effect of parton distribution function (PDF) uncertainties on the templates, though numerically the effect is found to be small. To evaluate PDF uncertainties, events are weighted by $w_1 w_2$, where w_i is the ratio

$$f^{\text{replica}}(x, Q^2) / f^{\text{nominal}}(x, Q^2)$$

for each incident proton, and replica is the new PDF set. For the QCD and $t\bar{t}$ samples, the normalization is taken from the fit that will be described in Section 7, and so the weights are renormalized to preserve the overall QCD and $t\bar{t}$ cross section, respectively; this renormalization is purely a change of variables with no physical content. Similarly, since the gluino pair

production cross section is related to the signal strength extracted in the fit, the PDF weights for the signal sample are renormalized as well. However, for all other backgrounds the weights are not renormalized in order to take into account the background cross section uncertainty. In the NNPDF scheme, each of the $N_{\text{var}} = 100$ PDF variations is the result of a best fit to a resampling of the probability distribution modeling the data [40, 41]. To take into account all of these variations separately while maintaining the overall uncertainty, the size of each variation is scaled down by a factor of $1/\sqrt{N_{\text{var}}} = 0.1$, and all are considered simultaneously.

Uncertainties arising from variations of the renormalization and factorization scales by a factor of two are taken into account for all backgrounds. Separate uncertainties are assigned for independent variations of these scales as well as a simultaneous variation.

As the modeling of $\mathrm{t\bar{t}}$ events is well understood, the effect of its uncertainties are small, motivating a treatment based on theoretical uncertainties. Event weights corresponding to variations of the renormalization and factorization scales are used to determine templates for systematic variations. The top quark p_{T} spectrum is reweighted to match the data. To check that the assigned systematic uncertainties cover generator based uncertainties, the N_{b} spectrum from the MADGRAPH generator at leading order is compared to that predicted by the POWHEG generator and MADGRAPH generator at NLO precision. This comparison finds that the three samples are consistent. There is a small uncertainty arising from the charged lepton identification efficiency, which is nearly negligible because the $\mathrm{t\bar{t}}$ normalizations are determined from data.

Uncertainties in the cross sections of backgrounds other than QCD or $\mathrm{t\bar{t}}$ are estimated by varying the renormalization and factorization scales, as well as the PDFs. The integrated luminosity is varied according to its uncertainty of 2.7% [42], and is relevant only for these subleading backgrounds, as the other backgrounds are normalized to data. For the subleading backgrounds, the luminosity uncertainty is small compared to cross section uncertainties.

Systematic uncertainties associated with jet energy scale, jet energy resolution [43], and pileup are evaluated with the same procedure for all background components. The pileup uncertainty is evaluated by reweighting the pileup distribution after varying the nominal minimum bias cross section of $\sigma = 69$ mb by its uncertainty of 5%. Jet energy scale uncertainties are calculated by varying the p_{T} of $R = 0.4$ jets as a function of p_{T} and η. The systematic uncertainty arising from jet energy resolution is determined by applying an $|\eta|$-dependent factor to the jet p_{T} to match the jet energy resolution observed in data. The full effect of this procedure is taken as an uncertainty.

All systematic uncertainties are modeled with a template-morphing technique [44], with the exception of the luminosity uncertainty, which is applied as a lognormal constraint for each $(N_{\text{jet}}, M_{\text{J}})$ bin. The systematic uncertainties on the background yield in the fit regions with the best signal sensitivity are summarized in Fig. 3. The values of the uncertainties given in the figure are their pre-fit values, though the normalization of the dominant background (QCD multijet events in the $N_{\text{lep}} = 0$ bin and $\mathrm{t\bar{t}}$ events in the $N_{\text{lep}} = 1$ bin) is adjusted to match the data normalization.

6.2 Signal systematic uncertainty

Several of the systematic uncertainties affecting the signal yield are evaluated in the same way as the background yield. These are the jet energy scale, jet energy resolution, pileup, and the scale factors for the b tagging efficiency. Theoretical uncertainties associated with PDFs and scale variations are evaluated in a similar manner for signal as for background; the only difference is that these variations are constrained to not change the overall normalization. As the

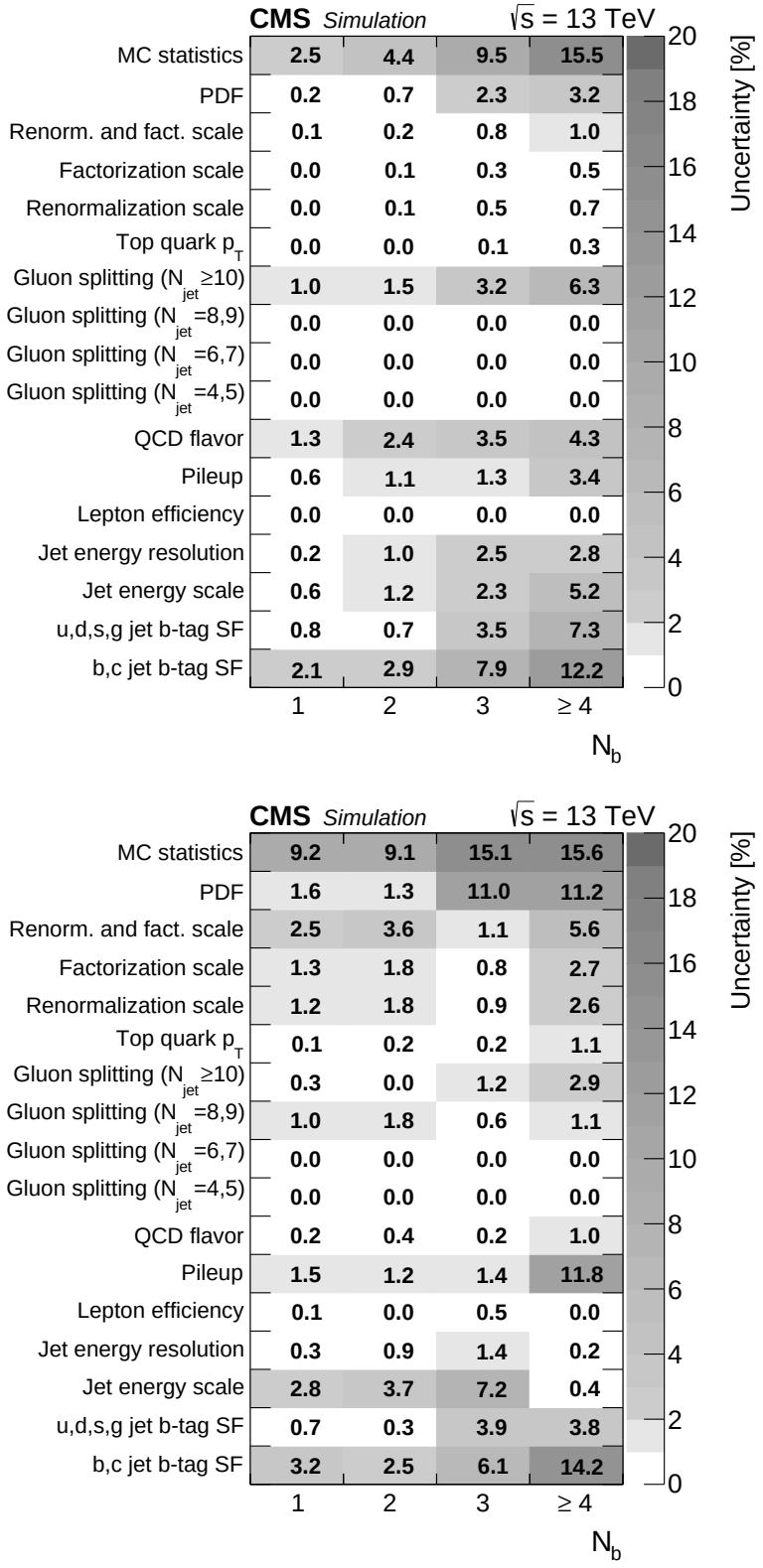

Figure 3: Background systematic uncertainties in the fit regions with the best signal sensitivity. The selection used in the top plot is $N_{\text{lep}} = 0$, $N_{\text{jet}} \geq 10$, $H_{\text{T}} > 1.5\,\text{TeV}$, and $M_{\text{J}} > 800\,\text{GeV}$ and the selection used in the bottom plot is $N_{\text{lep}} = 1$, $N_{\text{jet}} \geq 8$, $H_{\text{T}} > 1.2\,\text{TeV}$, and $M_{\text{J}} > 800\,\text{GeV}$.

trigger efficiency is consistent with 100 percent to within 0.1%, no uncertainty is assigned for trigger efficiency. Additional uncertainties arise from differences in the modeling of efficiencies between the full simulation and the fast simulation. These uncertainties are applied separately for the reconstruction and selection efficiency of leptons as well as for the b tagging efficiency.

To account for discrepancies in the modeling of initial state radiation, the $p_{\mathrm{T,sum}} = p_{\mathrm{T},\tilde{\mathrm{g}}1} + p_{\mathrm{T},\tilde{\mathrm{g}}2}$ of the gluino-gluino system is reweighted based on results of Z-recoil studies. A systematic uncertainty of 6%, 10%, and 15% is assigned for $p_{\mathrm{T,sum}}$ in the range $[40, 100)$, $[100, 150)$, and $[150, \infty)$ GeV, respectively.

The systematic uncertainties in the $N_{\mathrm{lep}} = 0$ and $N_{\mathrm{lep}} = 1$ bins with the best signal sensitivity are shown in Fig. 4 for the SUSY signal with $m_{\tilde{\mathrm{g}}} = 1200\,\mathrm{GeV}$. The dominant signal systematic uncertainties arise from PDFs, the jet energy scale, and the scale factors associated with b tagging efficiency.

7 Fit validation

Validation regions at low N_{jet} (N_{jet} = 4–7 for $N_{\mathrm{lep}} = 0$ and N_{jet} = 4–5 for $N_{\mathrm{lep}} = 1$) are studied before examining the signal region; fitting the N_{b} spectrum allows one to validate the fit and its modeling of uncertainties via nuisance parameters. A binned maximum likelihood fit is performed simultaneously in the M_{J}, N_{jet}, and N_{lep} regions. As the M_{J} and N_{jet} dependence of the QCD and $\mathrm{t\bar{t}}$ contributions may be modeled imperfectly by the MC, separate normalizations for these contributions are allowed in each bin of M_{J} and N_{jet}. Systematic uncertainties are incorporated by interpolating between templates in which the systematic uncertainties are varied by $\pm 1\sigma$. Nuisance parameters governing the change in the N_{b} distribution are allowed to vary in the fit subject to a Gaussian constraint. The constraints are normalized so that $\mu = 0$ corresponds to the unmodified systematic uncertainty and $\sigma = 1$ corresponds to a 1 standard deviation variation. The ability of the fit to model the data can be judged by determining the extent to which the nuisance parameters that govern the mean and width of the systematic uncertainties are changed from 0 and 1, respectively. The MC statistical uncertainties are taken into account by variations in which each single bin is varied according to its statistical uncertainty.

The likelihood used in the fit is

$$\mathcal{L} = \prod_{\substack{i \in N_{\mathrm{lep}} \\ j \in N_{\mathrm{jet}} \\ k \in N_{\mathrm{b}}}} P(N_{ijk} | \mu_{\mathrm{signal}} \nu_{ijk,\mathrm{signal}} + \mu_{ijk,\mathrm{QCD}} \nu_{ijk,\mathrm{QCD}} + \mu_{\mathrm{t\bar{t}}} \nu_{ijk,\mathrm{t\bar{t}}} + \nu_{ijk,\mathrm{W+jets}} + \nu_{ijk,\mathrm{other}}) \prod_{\mathrm{m} \in \mathrm{syst}} P(N_{ijk} | \theta_m) \tag{5}$$

where the N_{jet} bins are 4–5, 6–7, 8–9, and $\geq$10 for $N_{\mathrm{lep}} = 0$ and 4–5, 6–7, and $\geq$8 for $N_{\mathrm{lep}} = 1$. Only the lower N_{jet} regions are used in the validation region fits. The fitted signal strength is μ_{signal}. The $\mu_{ij,\mathrm{QCD}}$ and $\mu_{ij,\mathrm{t\bar{t}}}$ are the M_{J} and N_{jet}-dependent normalization constants. The signal, QCD, $\mathrm{t\bar{t}}$, W+jets and backgrounds other than QCD and $\mathrm{t\bar{t}}$ are relative to nominal values specified by $\nu_{ijk,\mathrm{signal}}$, $\nu_{ijk,\mathrm{QCD}}$, $\nu_{ijk,\mathrm{t\bar{t}}}$, $\nu_{ijk,\mathrm{W+jets}}$, $\nu_{ijk,\mathrm{other}}$, respectively. In other words, there is one signal yield and fixed minor background yields. The QCD ($\mathrm{t\bar{t}}$) normalizations are allowed to float in each $(M_{\mathrm{J}}, N_{\mathrm{jet}})$ bin for $N_{\mathrm{lep}} = 0$ ($N_{\mathrm{lep}} = 1$). In the $N_{\mathrm{lep}} = 0$ bins, the $\mathrm{t\bar{t}}$ yields are taken from the corresponding $N_{\mathrm{lep}} = 1$ bin, except with N_{jet} differing by two. This procedure is motivated by the fact that events with fully hadronic $\mathrm{t\bar{t}}$ decays have two more jets than those in which the $\mathrm{t\bar{t}}$ pair decays semileptonically. Similarly, in the $N_{\mathrm{lep}} = 1$ bins the QCD yields are taken from the corresponding $N_{\mathrm{lep}} = 0$ bin. Each systematic uncertainty is parameterized by a single nuisance parameter, θ_m where m denotes a systematic uncertainty, common in all $(M_{\mathrm{J}}, N_{\mathrm{jet}}, N_{\mathrm{lep}})$ bins;

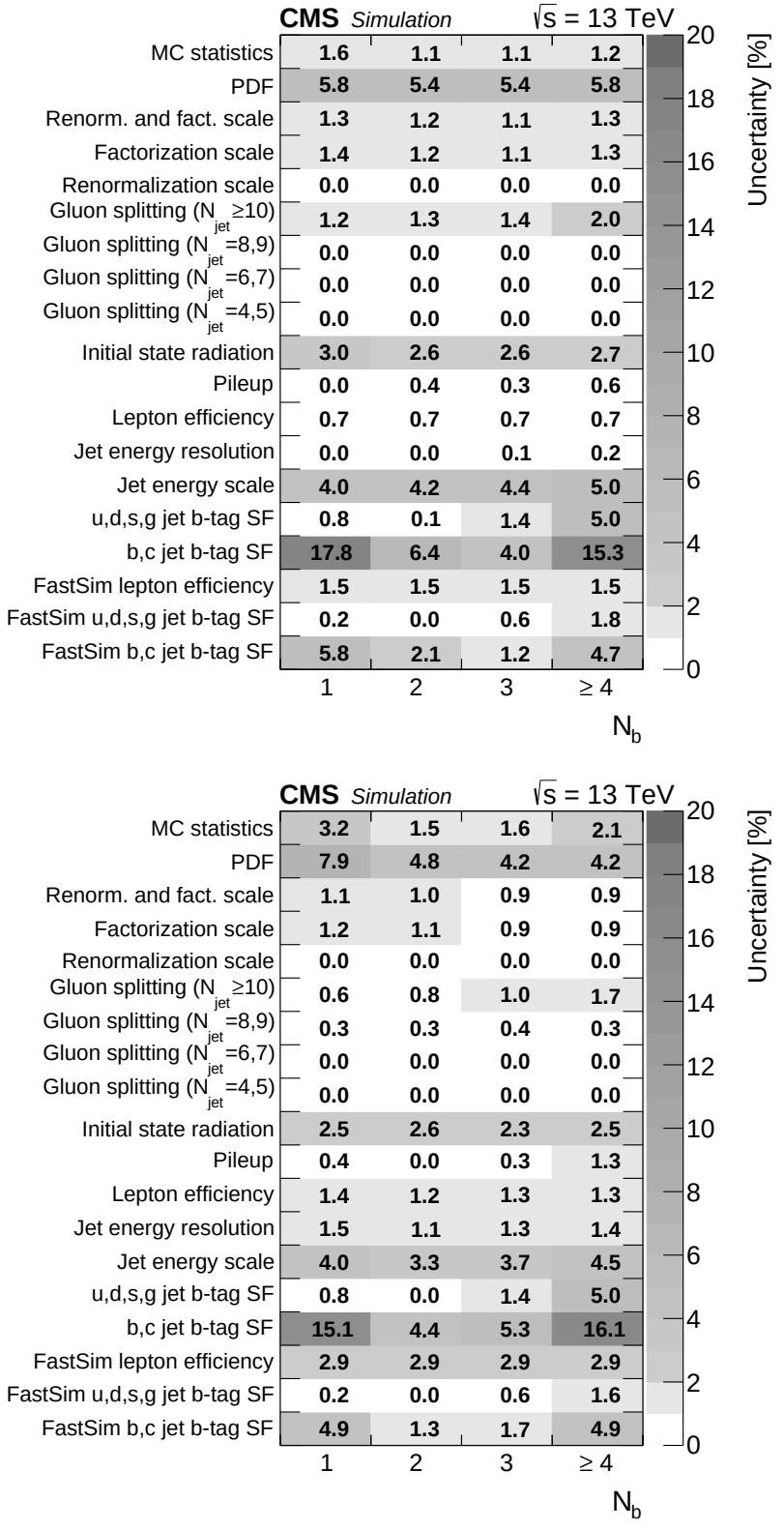

Figure 4: Signal systematic uncertainties for $m_{\tilde{g}} = 1200\,\text{GeV}$ in the signal regions with the best signal sensitivity. The selection used in the top plot is $N_{\text{lep}} = 0$, $N_{\text{jet}} \geq 10$, $H_{\text{T}} > 1500\,\text{GeV}$, and $M_{\text{J}} > 800\,\text{GeV}$ and the selection used in the bottom plot is $N_{\text{lep}} = 1$, $N_{\text{jet}} \geq 8$, $H_{\text{T}} > 1200\,\text{GeV}$, and $M_{\text{J}} > 800\,\text{GeV}$.

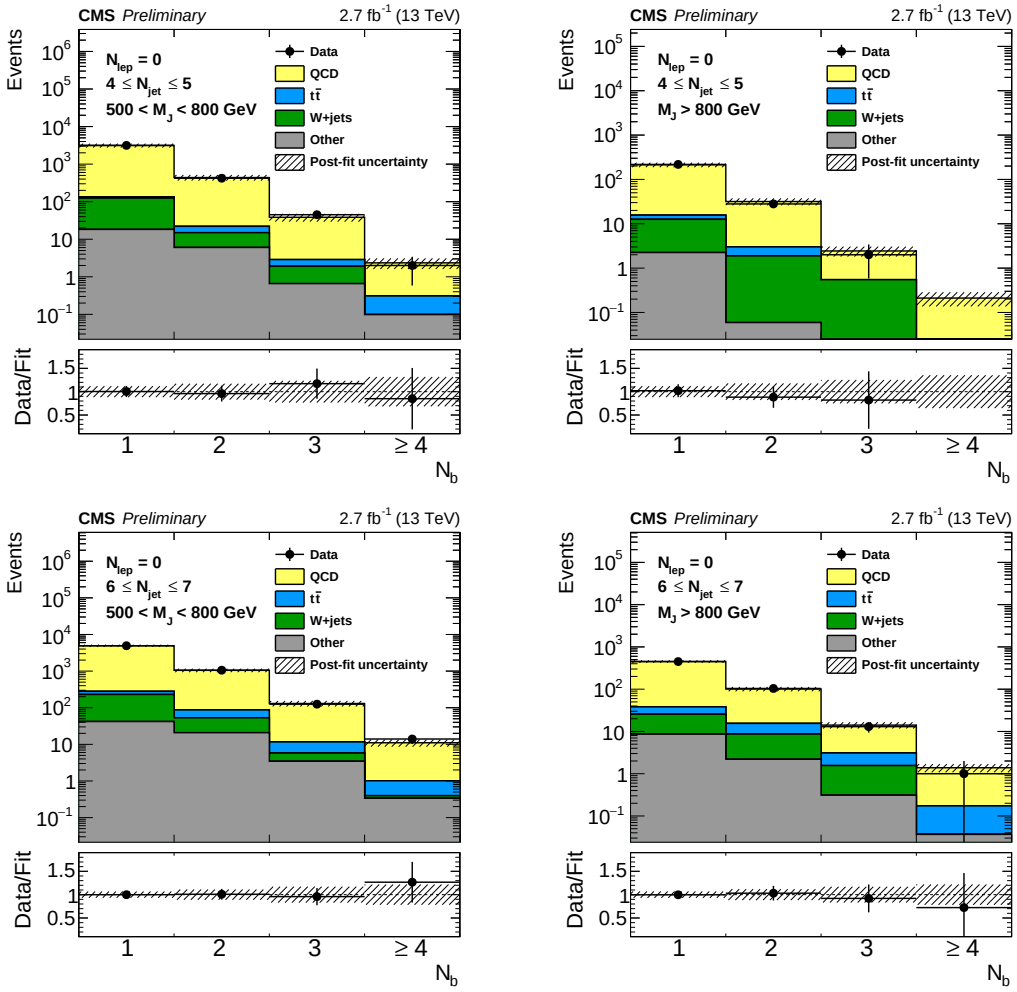

Figure 5: Fit to data in the validation regions with $N_{\text{lep}} = 0$. Plots in the left column require $500 < M_{\text{J}} < 800\,\text{GeV}$, while $M_{\text{J}} > 800\,\text{GeV}$ is required in the right column. The top row shows the $4 \leq N_{\text{jet}} \leq 5$ region, and the bottom row shows the $6 \leq N_{\text{jet}} \leq 7$ region. The uncertainty displayed in the ratio plot as error bars reflects data statistics only.

the nuisance parameters are also common between signal and background. Exceptions are the systematic uncertainties associated with gluon splitting, which is assigned separate nuisance parameters for different N_{jet} selections, and those associated with MC statistics, which have separate nuisance parameters for each $(M_{\text{J}}, N_{\text{jet}}, N_{\text{lep}}, N_{\text{b}})$ bin. In the control sample fit, the signal yield is set to zero, and bins with $N_{\text{jet}} \geq 8$ ($N_{\text{jet}} \geq 6$) bins are excluded for the hadronic (one-lepton) sample. The resulting fits to data are shown in Figures 5–6.

All of the nuisance parameters are consistent with their pre-fit values within the uncertainties, providing confidence in the modeling of systematic uncertainties. In particular, the theoretical uncertainties due to scale variations and PDF uncertainties have not been shifted, consistent with the assumption that the N_{b} distribution has little theoretical uncertainty. The nuisance parameters that are most constrained in the fit are those associated with b-tagging or jet energy scale. However, these are the nuisance parameters that have the largest effect on the b-tag multiplicity distribution and so it is expected that they will be constrained by the large statistics in the control regions. The nuisance parameters for PDF uncertainties (MC statistics) are pulled negligibly, with a χ^2 per degree of freedom of $0.022/100$ ($0.28/64$), indicating that these

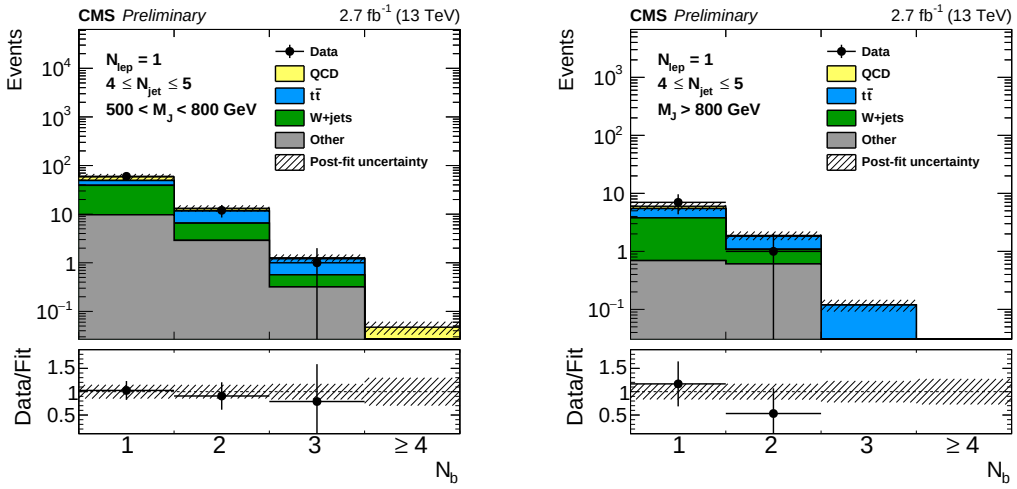

Figure 6: Fit to data in the validation regions with $N_{\text{lep}} = 1$. Plot on the left requires $500 < M_{\text{J}} < 800\,\text{GeV}$, while $M_{\text{J}} > 800\,\text{GeV}$ is required on the right plot. The uncertainty displayed in the ratio plot as error bars reflects data statistics only.

uncertainties do not have a significant impact on the fit.

Signal injection studies have been performed to validate the fit procedure prior to examining the signal region. An ensemble of data sets were generated from the simulated data, normalized assuming the integrated luminosity of the data, to study biases in the fit. For each gluino mass, a signal with the expected yield is injected to determine that the signal yield can be extracted from the fit without bias. The mean fitted signal strength is consistent with the expected value of 1.0 for all tested gluino masses.

8 Results

This section discusses the result of the fit performed simultaneously in all $(M_{\text{J}}, N_{\text{jet}}, N_{\text{lep}})$ bins. This fit differs from the fit discussed previously in Section 7 only by the additional of the signal regions at large values of N_{jet}, and the inclusion of a signal component. The pulls of the nuisance parameters after the fit are examined as a check of the validity of the fit. The fitted N_{b} distributions are found to be consistent with the null hypothesis, and we set upper limits on the cross section for gluino pair production.

The behavior of the fit is inspected by examining the pull of nuisance parameters in both signal plus background and background-only fits. All of the nuisance parameters are consistent with their pre-fit uncertainties, providing a confidence in the modeling of systematic uncertainties in the full fit.

The post-fit N_{b} distributions are shown in Figure 7 and 8 for the fully hadronic and 1-lepton signal regions, respectively. As the normalization of the dominant background component in each (M_{J}, N_{jet}, N_{lep}) region is taken from data and the nuisance parameters (including per-N_{b}-bin systematic uncertainties) are profiled in the fit, each post-fit background distribution is expected to match more closely to the data than a comparison of four independent bins would. Tables 2 and 3 compare the observed data yields in each N_{b} bin to the post-fit normalization of backgrounds, and the yields expected for a gluino with $m_{\tilde{\text{g}}} = 1200\,\text{GeV}$. The best-fit signal strength, in units of the gluino pair production cross section, is at the physical boundary of 0.0 in the fits at each assumed gluino mass.

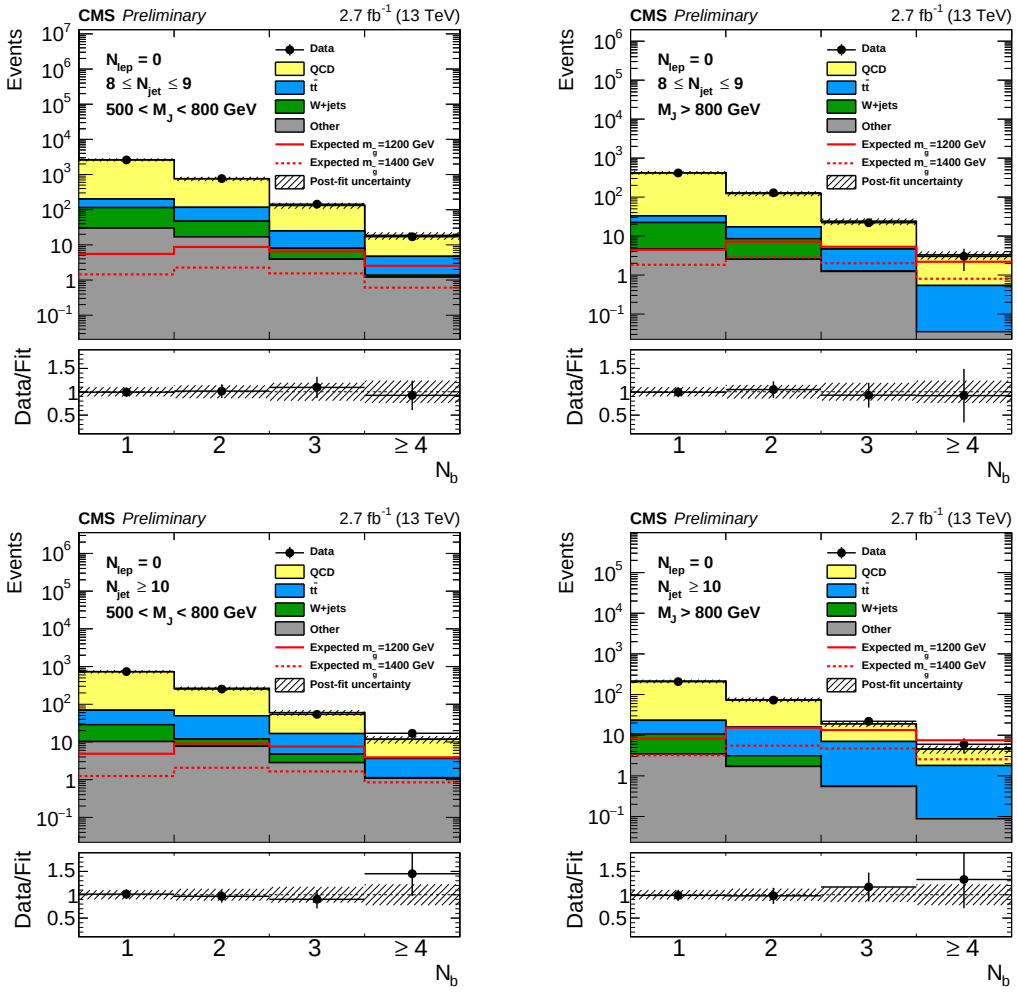

Figure 7: The post-fit N_b distributions in the 0-lepton signal region. Plots in the left column require $500 < M_J < 800\,\text{GeV}$, while $M_J > 800\,\text{GeV}$ is required in the right column. The top row shows the $8 \leq N_{jet} \leq 9$ signal region, and the bottom row shows the $N_{jet} \geq 10$ signal region. The uncertainty displayed in the ratio plot as error bars reflects data statistics only.

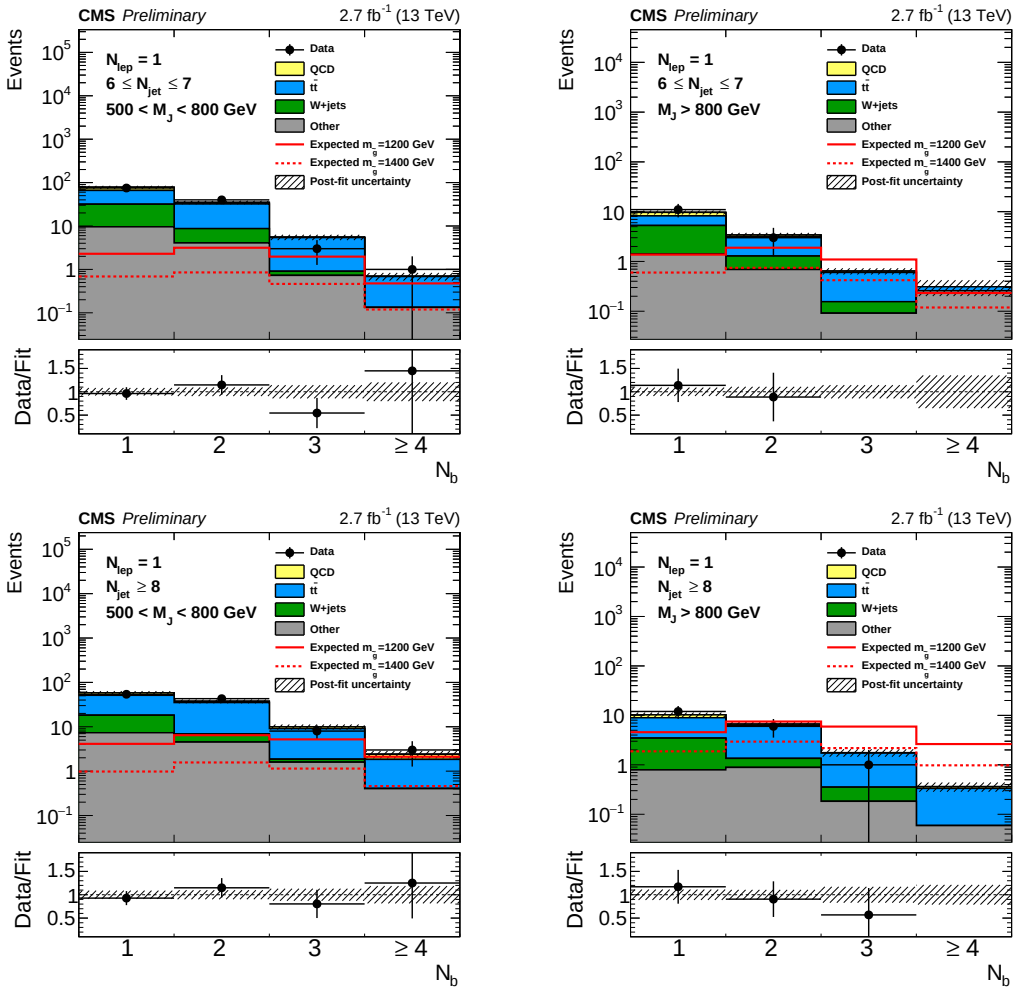

Figure 8: The post-fit N_b distributions in the 1-lepton signal region. Plots in the left column require $500 < M_J < 800\,\text{GeV}$, while $M_J > 800\,\text{GeV}$ is required in the right column. The top row shows the $6 \leq N_{jet} \leq 7$ signal region, and the bottom row shows the $N_{jet} \geq 8$ signal region. The uncertainty displayed in the ratio plot as error bars reflects data statistics only.

Table 2: Table of post-fit normalization of backgrounds, data, and expected yields for a gluino with $m_{\tilde{g}} = 1200\,\text{GeV}$ in events with $N_{\text{lep}} = 0$ and $H_T > 1500\,\text{GeV}$. Systematic uncertainties are included.

N_b	QCD	$t\bar{t}$	W+jets	Other	All bkg.	Data	$m_{\tilde{g}} = 1200\,\text{GeV}$
$8 \leq N_{\text{jet}} \leq 9$, $500 < M_J < 800\,\text{GeV}$							
1	2424.5 ± 245.1	85.5 ± 6.5	84.7 ± 20.0	30.1 ± 4.7	2624.7 ± 261.2	2603	5.5
2	639.8 ± 93.3	69.5 ± 7.6	30.0 ± 7.3	16.9 ± 2.9	756.2 ± 104.9	767	8.7
3	106.9 ± 21.6	16.8 ± 2.4	4.1 ± 1.3	3.9 ± 0.7	131.7 ± 25.3	144	6.5
≥ 4	13.7 ± 3.3	3.3 ± 0.9	0.2 ± 0.1	1.2 ± 0.2	18.4 ± 4.4	17	2.5
$8 \leq N_{\text{jet}} \leq 9$, $M_J > 800\,\text{GeV}$							
1	386.9 ± 39.3	10.7 ± 1.1	17.5 ± 3.9	4.7 ± 0.6	419.8 ± 41.9	416	4.5
2	106.0 ± 15.9	8.6 ± 1.1	6.0 ± 1.6	2.5 ± 0.7	123.2 ± 17.9	129	7.2
3	19.1 ± 3.8	3.4 ± 0.6	0.0 ± 0.0	1.2 ± 0.3	23.7 ± 4.5	22	5.3
≥ 4	2.7 ± 0.7	0.5 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	3.3 ± 0.8	3	2.2
$N_{\text{jet}} \geq 10$, $500 < M_J < 800\,\text{GeV}$							
1	657.3 ± 66.2	40.6 ± 4.7	18.5 ± 4.0	10.3 ± 1.5	726.7 ± 71.8	736	4.9
2	212.2 ± 29.0	37.5 ± 4.0	4.3 ± 1.5	7.7 ± 1.2	261.7 ± 32.9	253	9.1
3	43.2 ± 7.4	11.9 ± 2.3	1.9 ± 0.6	2.8 ± 0.6	59.9 ± 10.1	54	7.6
≥ 4	8.1 ± 1.9	2.5 ± 0.6	0.0 ± 0.0	1.1 ± 0.2	11.7 ± 2.6	17	3.9
$N_{\text{jet}} \geq 10$, $M_J > 800\,\text{GeV}$							
1	187.7 ± 18.8	12.5 ± 2.1	7.2 ± 1.8	3.5 ± 0.7	210.9 ± 21.2	208	8.3
2	58.7 ± 8.2	12.9 ± 1.4	1.4 ± 0.8	1.7 ± 0.3	74.7 ± 9.8	73	15.3
3	11.9 ± 2.1	6.4 ± 1.1	0.0 ± 0.0	0.5 ± 0.1	18.9 ± 2.9	22	13.4
≥ 4	2.7 ± 0.6	1.7 ± 0.5	0.0 ± 0.0	0.1 ± 0.0	4.5 ± 1.0	6	7.5

Table 3: Table of post-fit normalization of backgrounds, data, and expected yields for a gluino with $m_{\tilde{g}} = 1200\,\text{GeV}$ in events with $N_{\text{lep}} = 1$ and $H_T > 1200\,\text{GeV}$. Systematic uncertainties are included.

N_b	QCD	$t\bar{t}$	W+jets	Other	All bkg.	Data	$m_{\tilde{g}} = 1200\,\text{GeV}$
$6 \leq N_{\text{jet}} \leq 7$, $500 < M_J < 800\,\text{GeV}$							
1	12.0 ± 0.8	34.2 ± 1.9	22.1 ± 4.9	9.6 ± 2.6	77.9 ± 6.3	75	2.3
2	2.7 ± 0.3	23.5 ± 1.9	4.6 ± 1.0	4.1 ± 1.0	34.9 ± 3.2	40	3.2
3	0.1 ± 0.0	4.5 ± 0.6	0.2 ± 0.2	0.7 ± 0.1	5.5 ± 0.8	3	2.0
≥ 4	0.0 ± 0.0	0.6 ± 0.1	0.0 ± 0.0	0.1 ± 0.0	0.7 ± 0.1	1	0.5
$6 \leq N_{\text{jet}} \leq 7$, $M_J > 800\,\text{GeV}$							
1	1.4 ± 0.1	3.0 ± 0.2	3.9 ± 0.9	1.4 ± 0.3	9.7 ± 0.9	11	1.4
2	0.3 ± 0.0	1.8 ± 0.2	0.6 ± 0.1	0.7 ± 0.2	3.4 ± 0.4	3	1.9
3	0.0 ± 0.0	0.4 ± 0.1	0.1 ± 0.0	0.1 ± 0.0	0.6 ± 0.1	0	1.1
≥ 4	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.3 ± 0.1	0.3 ± 0.1	0	0.2
$N_{\text{jet}} \geq 8$, $500 < M_J < 800\,\text{GeV}$							
1	7.1 ± 0.9	32.9 ± 2.5	10.8 ± 2.7	7.3 ± 1.6	58.2 ± 5.1	54	4.1
2	2.4 ± 0.4	28.2 ± 2.9	2.4 ± 0.7	4.5 ± 0.8	37.4 ± 3.6	43	6.5
3	0.9 ± 0.2	7.2 ± 1.0	0.3 ± 0.1	1.6 ± 0.2	10.0 ± 1.3	8	5.2
≥ 4	0.6 ± 0.2	1.4 ± 0.3	0.0 ± 0.0	0.4 ± 0.1	2.4 ± 0.4	3	2.1
$N_{\text{jet}} \geq 8$, $M_J > 800\,\text{GeV}$							
1	1.3 ± 0.2	5.5 ± 0.6	2.7 ± 0.6	0.8 ± 0.3	10.3 ± 1.1	12	4.6
2	0.4 ± 0.1	4.9 ± 0.5	0.5 ± 0.1	0.9 ± 0.3	6.6 ± 0.7	6	7.5
3	0.0 ± 0.0	1.4 ± 0.2	0.2 ± 0.2	0.2 ± 0.0	1.8 ± 0.3	1	5.9
≥ 4	0.0 ± 0.0	0.3 ± 0.1	0.0 ± 0.0	0.1 ± 0.0	0.4 ± 0.1	0	2.6

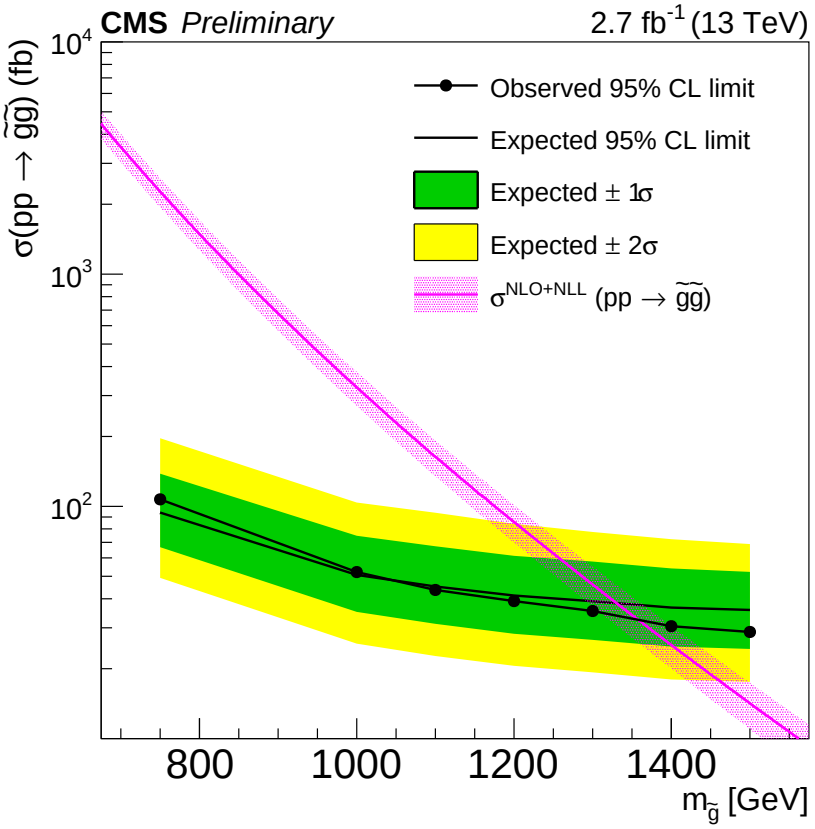

Figure 9: Cross section upper limits at 95% CL compared to the gluino pair production cross section (magenta). The expected limits (black solid line) and their ± 1 standard deviation (green) and ± 2 standard deviation (yellow) limits are shown. The observed limit is displayed as a black solid line with dots.

No significant excesses of events above the standard model predictions are observed, and we set upper limits on the gluino pair production cross section, calculated in the asymptotic approximation [45]. Figure 9 shows the expected and observed cross-section upper limit at 95% CL. Comparing that limit to the gluino pair production cross section [46], we expect to exclude $m_{\tilde{g}} < 1340\,\text{GeV}$. Since no signal is extracted, the observed limit is comparable, excluding gluino masses with $m_{\tilde{g}} < 1360\,\text{GeV}$.

9 Summary

A search has been performed for new physics in high multiplicity final states with zero or one reconstructed lepton. The data used in this analysis were collected by the CMS experiment at $\sqrt{s} = 13\,\text{TeV}$ and correspond to an integrated luminosity of $2.7\,\text{fb}^{-1}$. The results of the search are interpreted in the context of R-parity-violating supersymmetry with a model in which the gluino decays exclusively to tbs. No significant excesses are observed, and cross section limits have been set at 95% confidence level. These limits correspond to the exclusion of gluino masses of $m_{\tilde{g}} < 1360\,\text{GeV}$ within this scenario.

References

[1] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints", *Phys. Rept.* **405** (2005) 279, doi:10.1016/j.physrep.2004.08.031, arXiv:hep-ph/0404175.

[2] E. Witten, "Dynamical Breaking of Supersymmetry", *Nucl. Phys. B* **188** (1981) 513, doi:10.1016/0550-3213(81)90006-7.

[3] S. Dimopoulos and H. Georgi, "Softly Broken Supersymmetry and SU(5)", *Nucl. Phys. B* **193** (1981) 150, doi:10.1016/0550-3213(81)90522-8.

[4] P. Ramond, "Dual theory for free fermions", *Phys. Rev. D* **3** (1971) 2415, doi:10.1103/PhysRevD.3.2415.

[5] Y. A. Golfand and E. P. Likhtman, "Extension of the algebra of Poincaré group generators and violation of P invariance", *JETP Lett.* **13** (1971) 323.

[6] A. Neveu and J. H. Schwarz, "Factorizable dual model of pions", *Nucl. Phys. B* **31** (1971) 86, doi:10.1016/0550-3213(71)90448-2.

[7] D. V. Volkov and V. P. Akulov, "Possible universal neutrino interaction", *JETP Lett.* **16** (1972) 438.

[8] J. Wess and B. Zumino, "A Lagrangian model invariant under supergauge transformations", *Phys. Lett. B* **49** (1974) 52, doi:10.1016/0370-2693(74)90578-4.

[9] J. Wess and B. Zumino, "Supergauge transformations in four dimensions", *Nucl. Phys. B* **70** (1974) 39, doi:10.1016/0550-3213(74)90355-1.

[10] P. Fayet, "Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino", *Nucl. Phys. B* **90** (1975) 104, doi:10.1016/0550-3213(75)90636-7.

[11] H. P. Nilles, "Supersymmetry, supergravity and particle physics", *Phys. Rep.* **110** (1984) 1, doi:10.1016/0370-1573(84)90008-5.

[12] G. R. Farrar and P. Fayet, "Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry", *Phys. Lett. B* **76** (1978) 575, doi:10.1016/0370-2693(78)90858-4.

[13] R. Barbier et al., "R-parity violating supersymmetry", *Phys. Rept.* **420** (2005) 1, doi:10.1016/j.physrep.2005.08.006, arXiv:hep-ph/0406039.

[14] L. E. Ibanez and G. G. Ross, "Discrete gauge symmetries and the origin of baryon and lepton number conservation in supersymmetric versions of the standard model", *Nucl. Phys. B* **368** (1992) 3, doi:10.1016/0550-3213(92)90195-H.

[15] C. Csáki, Y. Grossman, and B. Heidenreich, "Minimal flavor violation supersymmetry: A natural theory for R-parity violation", *Phys. Rev. D* **85** (2012) 095009, doi:10.1103/PhysRevD.85.095009, arXiv:1111.1239.

[16] J. A. Evans, Y. Kats, D. Shih, and M. J. Strassler, "Toward Full LHC Coverage of Natural Supersymmetry", *JHEP* **07** (2014) 101, doi:10.1007/JHEP07(2014)101, arXiv:1310.5758.

[17] J. A. Evans, "A Swarm of Bs", *JHEP* **08** (2014) 073, doi:10.1007/JHEP08(2014)073, arXiv:1402.4481.

[18] CMS Collaboration, "Search for three-jet resonances in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett. B* **718** (2012) 329, doi:10.1016/j.physletb.2012.10.048, arXiv:1208.2931.

[19] CMS Collaboration, "Search for Three-Jet Resonances in pp Collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **107** (2011) 101801, doi:10.1103/PhysRevLett.107.101801, arXiv:1107.3084.

[20] CMS Collaboration, "Searches for light- and heavy-flavour three-jet resonances in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Lett. B* **730** (2014) 193, doi:10.1016/j.physletb.2014.01.049, arXiv:1311.1799.

[21] CMS Collaboration, "Searches for R-parity-violating supersymmetry in pp collisions at $\sqrt{s}$ = 8 TeV in final states with 0–4 leptons", (2016). arXiv:1606.08076. Submitted to Phys. Rev. D.

[22] A. Hook, E. Izaguirre, M. Lisanti, and J. G. Wacker, "High Multiplicity Searches at the LHC Using Jet Masses", *Phys. Rev. D* **85** (2012) 055029, doi:10.1103/PhysRevD.85.055029, arXiv:1202.0558.

[23] T. Cohen, E. Izaguirre, M. Lisanti, and H. K. Lou, "Jet Substructure by Accident", *JHEP* **03** (2013) 161, doi:10.1007/JHEP03(2013)161, arXiv:1212.1456.

[24] S. El Hedri, A. Hook, M. Jankowiak, and J. G. Wacker, "Learning How to Count: A High Multiplicity Search for the LHC", *JHEP* **08** (2013) 136, doi:10.1007/JHEP08(2013)136, arXiv:1302.1870.

[25] ATLAS Collaboration, "Search for massive supersymmetric particles decaying to many jets using the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev. D* **91** (2015) 112016, doi:10.1103/PhysRevD.91.112016, arXiv:1502.05686.

[26] ATLAS Collaboration, "Search for new phenomena in final states with large jet multiplicities and missing transverse momentum at $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS experiment", *JHEP* **10** (2013) 130, doi:10.1007/JHEP10(2013)130, 10.1007/JHEP01(2014)109, arXiv:1308.1841. [Erratum: *JHEP* **01** (2014) 109].

[27] CMS Collaboration, "Search for supersymmetry in pp collisions at $\sqrt{s}$ = 13 TeV in the single-lepton final state using the sum of masses of large-radius jets", (2016). arXiv:1605.04608.

[28] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[29] 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* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[30] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[31] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[32] CMS Collaboration, "The fast simulation of the CMS detector at LHC", *J. Phys. Conf. Ser.* **331** (2011) 032049, doi:10.1088/1742-6596/331/3/032049.

[33] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and Missing E_T", technical report, CERN, 2009.

[34] CMS Collaboration, "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", technical report, CERN, 2010.

[35] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[36] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[37] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016.

[38] CMS Collaboration, "Measurement of $B\overline{B}$ angular correlations based on secondary vertex reconstruction at $\sqrt{s} = 7$ TeV", *JHEP* **03** (2011) 1103, doi:10.1007/JHEP03(2011)136, arXiv:1102.3194.

[39] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* **77** (1993) 219, doi:10.1016/0010-4655(93)90005-W.

[40] NNPDF Collaboration, "A determination of parton distributions with faithful uncertainty estimation", *Nucl. Phys. B* **809** (2009) 1, doi:10.1016/j.nuclphysb.2008.09.037, 10.1016/j.nuclphysb.2009.02.027, arXiv:0808.1231. [Erratum: Nucl. Phys.B816,293(2009)].

[41] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[42] CMS Collaboration, "CMS Luminosity Measurement for the 2015 Data Taking Period", Technical Report CMS-PAS-LUM-15-001, CERN, Geneva, 2016.

[43] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", Technical Report CERN-PH-EP-2015-305. CMS-JME-13-004-003. arXiv:1607.03663, CERN, Geneva, Jul, 2016. Submitted to J. Instrum.

[44] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies, "Handling uncertainties in background shapes", *JINST* **10** (2015) P04015, doi:10.1088/1748-0221/10/04/P04015, arXiv:1408.6865.

[45] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)].

[46] M. Kramer et al., "Supersymmetry production cross sections in pp collisions at $\sqrt{s} = 7$ TeV", (2012). arXiv:1206.2892.