PHYSICAL REVIEW D **81,** 035005 (2010)

Modeling missing transverse energy in V + jets at CERN LHC

Victor Pavlunin

Department of Physics, University of California, Santa Barbara, California, USA, 93106-9530

(Received 26 June 2009; revised manuscript received 2 December 2009; published 3 February 2010)

I discuss a method to model the instrumental response of the CMS and ATLAS detectors at high missing transverse energy to dominant standard model V + jets backgrounds, where V is a Z, γ or W, using multijet QCD events. The method is developed for new physics searches in early data at the LHC with minimal recourse to simulation.

DOI: 10.1103/PhysRevD.81.035005 PACS numbers: 12.60.−i, 13.85.Qk, 14.70.Fm, 14.70.Hp

I. INTRODUCTION

The LHC enters a new energy regime to explore the origin of the electroweak symmetry breaking, search for and study physics beyond the standard model (SM). At its design center-of-mass energy, new physics production cross sections may be significant so that a data sample of modest integrated luminosity, 100 pb^{-1} or less, may contain a large number of new particles. The challenge is to distinguish events with new particles from those, many orders of magnitude more copious, attributed to the SM with limited understanding of the SM production rates and detector performance in early LHC data.

Missing transverse energy, $\not\!\!E_T$, has discriminating power to reveal new particles interacting weakly with ordinary matter produced via high energy parton collisions in laboratory conditions at the LHC. These weakly interacting particles may comprise the dark matter of our Universe [1]. They are expected in new physics models, such as R-parity conserving supersymmetry [2] and many others [3]. Missing transverse energy allows to perform a broad search sensitive to the presence of such particles in collision data and is an observable that may lead to an early discovery at the LHC [4]. At the same time, missing transverse energy is one of the most difficult observables to measure precisely and simulate accurately [5] because it is measured by multiple detector subsystems and subject to mismeasurements and backgrounds in any of them.

In this paper, I discuss a new method [6] to predict backgrounds at high $\not\!\!E_T$ for new physics searches in signatures consistent with SM V + jets and $t\bar{t}$ + jets [7], where V is a Z, γ or W. I assume that new particles are heavy and decay to SM particles emitting multiple jets so that high sensitivity is expected at high $\not\!\!E_T$ and a large number of jets. Since the main sources of artificial $\not\!\!E_T$ come from the system of jets, the detector and noncollision effects, I model the instrumental response to the system of jets in V + jets and other effects at high $\not\!\!E_T$ *in situ* using multijet QCD events. This method complements and extends the work of Ref. [8], where events with high rapidity objects are used to model SM V + jets and multijet backgrounds in new physics searches without heavy reliance on $\not\!\!E_T$. The emphasis of this work, as that of Ref. [8], is on robustness against imperfections of background modeling required for new physics searches in early LHC data.

II. OVERVIEW

Monte Carlo (MC) simulation capable of modeling the detector response to SM processes is a great asset in new physics searches. However, there are two challenges in searches of early LHC data based on MC simulation. First, the SM V + jets production rates are difficult to predict from first principles. MC techniques are unreliable in predicting backgrounds with a large number of jets and need to be tuned with high $\sqrt{s}$ data. Theory calculations at sufficiently high order in many cases do not exist [9]. The structure functions have significant uncertainties in the small x range accessible at the LHC [10]. Second, significant uncertainties in the calibration of the experimental apparatus are expected in early data taking. Missing transverse energy is an observable that is particularly difficult to measure precisely and simulate accurately, since large jet energy fluctuations, detector artifacts, collision related and noncollision effects can produce non-Gaussian high $\not\!\!E_T$ tails. These artificial $\not\!\!E_T$ tails may resemble a signature of a new weakly interacting particle.

To introduce the method, let us consider an event with a Z boson reconstructed in the dimuon channel and four jets. The four-momentum of the Z is well measured so that the system of the four jets and other effects unrelated to the dimuon system are the main source of $\not\!\!E_T$ in this event. In order to develop a search in $\not\!\!E_T$ based on MC simulation, one would need to identify, understand and simulate the detector response to each of these effects. Instead, I model these effects *in situ* using multijet QCD events as follows: A sample of QCD events with four jets that have approximately the same configuration as the four jet system of the Z + jets event is selected. A $\not\!\!E_T$ prediction, or a template, for this Z + jets event is obtained using the $\not\!\!E_T$ distribution measured in the selected QCD sample and normalized to unity. This procedure is repeated for other Z + jets events with four jets. The $\not\!\!E_T$ templates are summed to obtain a SM $\not\!\!E_T$ prediction for all Z + jets events with four jets.

The photon momentum in γ + jets is also well measured so that the same procedure applies to obtain a SM $\not\!\!E_T$

prediction in the γ + jets sample. In W + jets and $t\bar{t}$ + jets with one of the two top quarks decaying semileptonically, the l + jets + $\not\!\!E_T$ signature, there is genuine $\not\!\!E_T$ from the undetected neutrino produced in W decays. To avoid reliance on MC and theory, I model the neutrino $|\vec{p}_T|$ spectra using the charged lepton $|\vec{p}_T|$ spectra. If the W bosons are not polarized in the transverse plane, the two spectra should be the same. Event selection and polarized W bosons produced in top quark decays lead to differences in the charged lepton and neutrino spectra. However, these differences are small and can be accounted for by corrections. A prediction for artificial $\not\!\!E_T$ in W + jets and $t\bar{t}$ + jets is derived in the same manner as for Z + jets and combined with a modeled neutrino $|\vec{p}_T|$ spectrum to obtain a SM $\not\!\!E_T$ prediction in the l + jets + $\not\!\!E_T$ final state.

In each channel, the search is made in $\not\!\!E_T$ distributions of events with the same number of jets, or events with at least a certain number jets. Since higher sensitivity to new physics is expected in events with a large number of jets, the focus is to model the high $\not\!\!E_T$ region in V + jets events with 3 or more jets. Events with 2 jets are valuable as a validation and calibration sample. This method is developed for searches in early LHC data. It will work best if the LHC start-up is quick, new particles are strongly produced and not very heavy, e.g., such as squarks and gluinos in the low mass minimal supergravity (mSUGRA) CMS and ATLAS benchmarks [6]. With this in mind, a prediction of SM backgrounds in high $\not\!\!E_T$ tails to about 20% may be sufficient to reveal new physics. For this reason, an accuracy benchmark for this method is to predict SM backgrounds in V + jets at high $\not\!\!E_T$ and a large number of jets (3 jets or more) to about 20% or better in a data-driven manner.

III. EXPERIMENTAL ASPECTS

The CMS and ATLAS experiments use multipurpose detectors at the European Organization for Nuclear Research (CERN). Detailed descriptions of the detectors can be found in Ref. [11]. The detectors are capable of reconstructing electrons and muons with high efficiencies and low fake rates for lepton $|\vec{p}_T| > 20$ GeV in the $|\eta| < 2.5$ range [12]. (In this paper the symbol l is used to denote an e or μ, but not τ. Charge-conjugate modes are implied.) In both detectors, photons and jets can be reconstructed reliably within $|\eta| < 2.5$ and $|\eta| < 3.0$, respectively.

To study the method, mock data samples are generated for the following SM processes: Z + jets (5.0 fb^{-1}, up to 5 partons, $Z \to l^+l^-$), W + jets (1.0 fb^{-1}, up to 5 partons, $W \to l\nu_l$), $t\bar{t}$ + jets (1.0 fb^{-1}, up to 4 partons, $t\bar{t} \to l\nu_l bbjj$), γ + jets (400.0 pb^{-1}, up to 5 partons), and QCD jets (1.0 pb^{-1}, up to 5 partons) [13]. (The same samples are used in Ref. [8].) The integrated luminosity listed in parentheses is used everywhere in tests in this paper but Sec. XI. These samples are generated with ALPGEN [14] at the parton level. PYTHIA [15] is used for parton showering, hadronization, simulation of the underlying event, and jet reconstruction. To model features of a new physics signal in search distributions, mock signal data samples for mSUGRA benchmark points LM1 and LM4 [2,16,17] are generated with PYTHIA.

Electrons and muons are required to have $|\vec{p}_T|$ of at least 20 GeV in the $|\eta| < 2.5$ range. Photons are reconstructed above the $|\vec{p}_T|$ threshold of 30 GeV in the $|\eta| < 2.5$ range. Jets are reconstructed using the PYCELL algorithm for $R = 0.5$ [15] and required to be within $|\eta| < 3.0$. A low jet $|\vec{p}_T|$ threshold of 20 GeV is used in the $\not\!\!E_T$ measurements in order to collect the energy deposited in the calorimeters to a fuller extent. Higher jet $|\vec{p}_T|$ thresholds, 50 GeV or more, are used to measure other observables in a robust manner as indicated below. I require that the leading jet and $\vec{\not\!\!E}_T$ be not aligned in the transverse plane within 0.15 radians: $0.15 < |\Delta\phi^{\text{lead jet-}\not\!\!E_T}| < (\pi - 0.15)$. (The jet with the highest $|\vec{p}_T|$ in an event is the leading jet of this event. Any other jet in this event is a nonleading jet.) It is assumed that the triggering and event reconstruction efficiency in each channel is 50%.

The vector of missing transverse energy, $\vec{\not\!\!E}_T$, is calculated as the vector opposite to the sum of $\vec{p}_T$ measurements of charged leptons, photons, and jets in each event. Since electron, muon and photon momenta are measured with high precision compared to jets, their contribution to artificial $\not\!\!E_T$ in events with a large number of jets is negligible. To emulate detector resolution effects for jets, jet energies measured by PYTHIA are smeared. The jet smearing function has three components: a) a Gaussian with

$$\sigma(|\vec{p}_T|)/|\vec{p}_T| = \sqrt{(7.0/|\vec{p}_T|)^2 + (1.2/\sqrt{|\vec{p}_T|})^2 + (0.04)^2},$$

where $|\vec{p}_T|$ is measured in GeV, b) an exponential low-side tail that stretches from $[1.0 - 2\sigma(|\vec{p}_T|)/|\vec{p}_T|]$ to 0.0 added to the Gaussian component 3% of the time, and c) similarly, an exponential high-side tail from $[1.0 + 2\sigma(|\vec{p}_T|)/|\vec{p}_T|]$ to infinity added to the Gaussian 1% of the time. Figure 1 shows the $|\vec{p}_T|$ dependence of the Gaussian smearing and the full smearing function with the non-Gaussian tails for a few fixed jet $|\vec{p}_T|$ values. This jet smearing function is constructed based on studies of the CMS and ATLAS detectors [6,11] to represent the expected jet response characteristic of the two detectors. Since the jet system tends to be the dominant source of artificial $\not\!\!E_T$, all three components of the jet smearing function are varied in the studies of robustness and limitations of the method discussed below.

The selection criteria used in the paper are not optimized to any new physics model. Instead, they are chosen to ensure robust detector performance and maintain sensitivity to a wide range of new physics models at high $\not\!\!E_T$ and a large number of jets. The selection criteria can be modified without producing a significant effect on the method's performance. The mSUGRA benchmarks listed above are

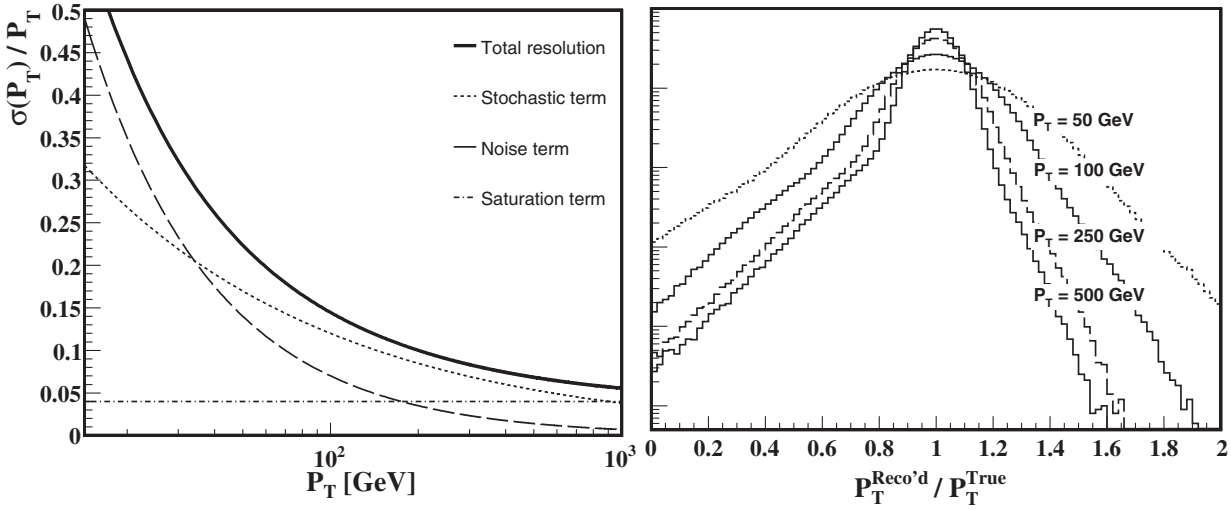

FIG. 1. Left: the Gaussian contribution to the jet energy resolution as a function of true jet $|\vec{p}_T|$ is shown in the solid line. The noise, stochastic, and saturation contributions to the jet resolution function are shown separately. Right: the jet energy smearing functions for 500, 250, 100, and 50 GeV $|\vec{p}_T|$ jets are shown in the solid, dashed, dotted-dashed, and dotted lines.

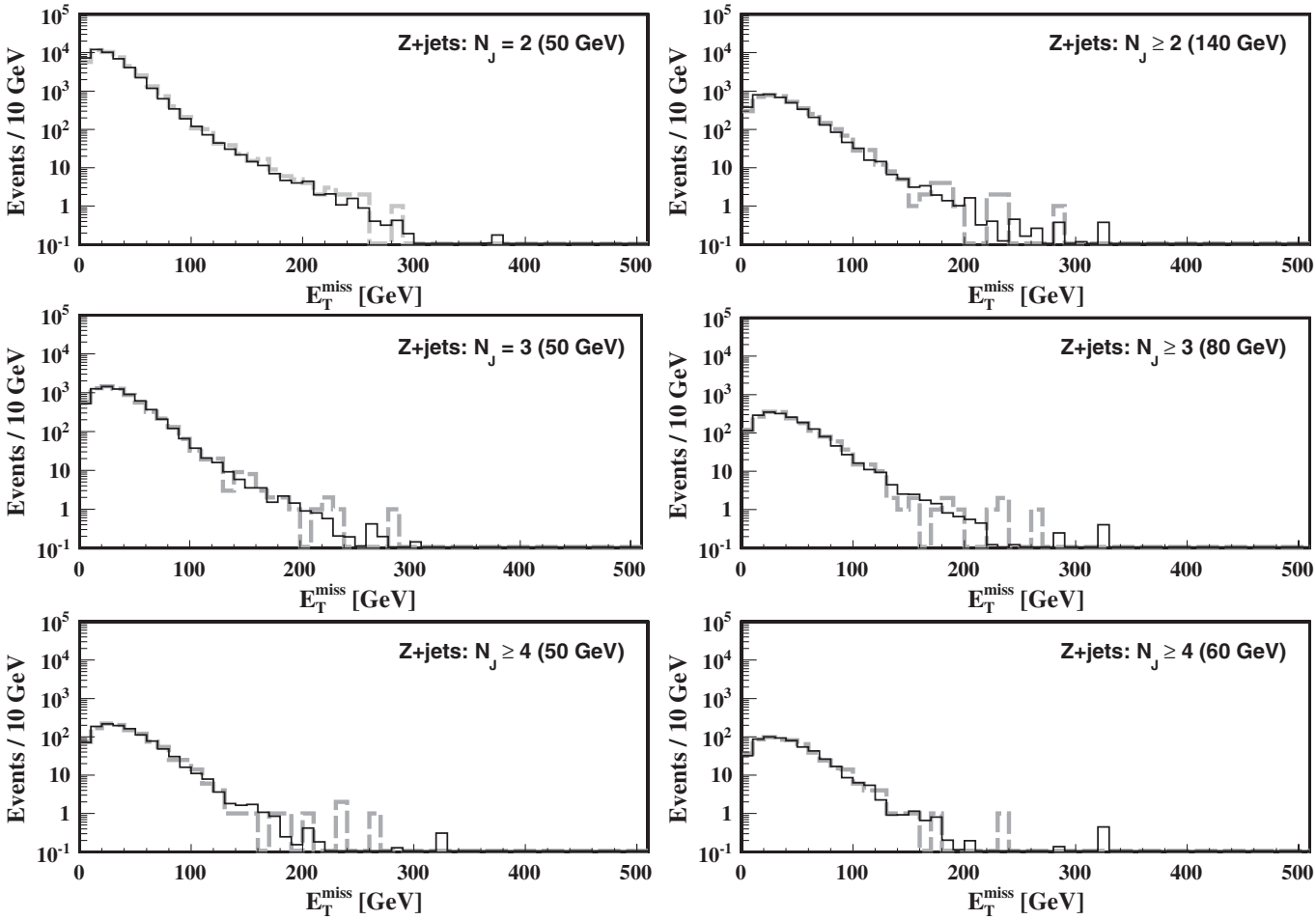

FIG. 2. Algorithm performance in Z + jets for N_J of, or at least of, 2, 3, and 4 in the first, middle, and third rows, respectively. The first (second) column is for the 50 GeV (high) jet $|\vec{p}_T|$ thresholds for N_J. The observed $\not\!\!E_T$ distributions are shown in the dashed lines, their predictions obtained using multijet QCD events are the solid lines.

used only to illustrate generic features of a new physics contribution. The goal of this paper is to demonstrate the scope of the method and its performance rather than to attain high sensitivity to a specific model or to quantify it.

IV. ALGORITHM

I will describe in detail the algorithm in this section and present results of its tests in the next section. Let us assume that the V momentum in a V + jets event is known. The resolution and other effects producing artificial missing transverse energy for this event are modeled using multijet QCD events with a kinematic configuration of jets similar to that in the V + jets event.

A prominent difference between the jet systems in V + jets and QCD is that in V + jets events the jet system recoils against the V while in QCD events it is at rest in the transverse plane. A key is to select multijet QCD events to predict $\not\!\!E_T$ in a manner that captures effects generating artificial $\not\!\!E_T$ *in situ* but allows for the difference stemming from the boost of the jet system in the transverse plane in V + jets.

It is not necessary to model accurately every degree of freedom in the jet system of V + jets on an event-by-event basis by QCD for two reasons. First, each V + jets event is modeled by a large sample of QCD events so that mismodeled correlations are averaged out over this sample of QCD events. Second, $\not\!\!E_T$ is measured for a large V + jets sample so that mismodeled correlations are averaged out over V + jets events as well. These averaging effects allow to develop a simple algorithm.

Multijet QCD events are selected using two variables: (1) N_J, number of jets above a high $|\vec{p}_T|$ threshold (50 GeV or higher), and (2) $J_T \equiv \sum |\vec{p}_T^{\,\text{jet}}|$ for jets above a low 20 GeV $|\vec{p}_T|$ threshold (the same jet threshold is used for $\not\!\!E_T$ measurements) [18]. A QCD sample is selected for each pair (N_J, J_T bin), the width of J_T bins is 10 GeV (100 GeV) below (above) 1 TeV. A $\not\!\!E_T$ template is obtained for each of these samples as a $\not\!\!E_T$ distribution in that sample normal-

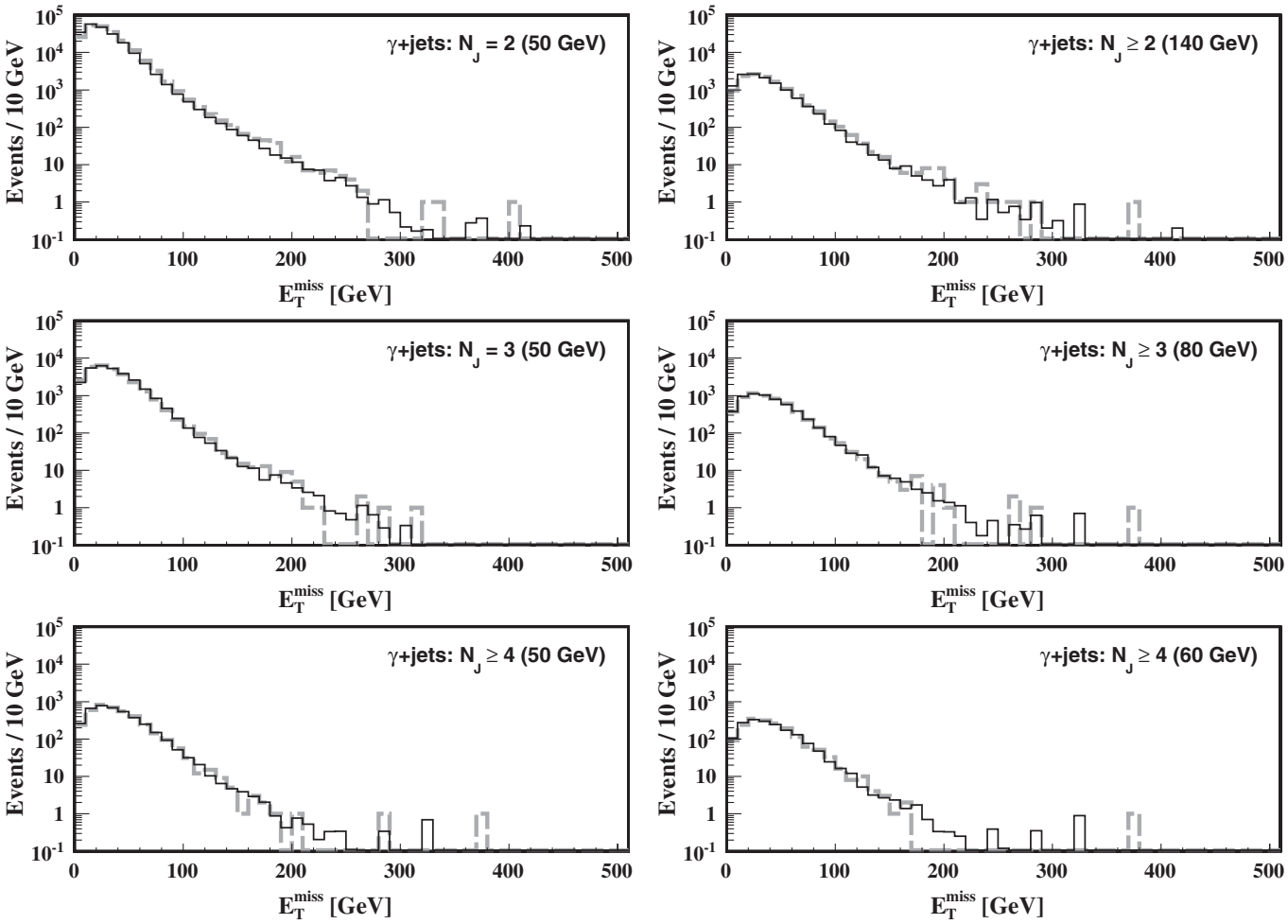

FIG. 3. Algorithm performance in γ + jets for N_J of, or at least of, 2, 3, and 4 in the first, middle, and third rows. The first (second) column shows results for the 50 GeV (high) jet $|\vec{p}_T|$ threshold(s) for N_J. The observed $\not\!\!E_T$ distributions are shown in the dashed lines, their predictions are the solid lines.

ized to unity. For each V + jets event, N_J, and J_T are measured and used to select the $\not\!\!E_T$ template with the same N_J in the corresponding J_T bin, which represents an artificial $\not\!\!E_T$ prediction for this V + jets event. These templates can be summed over, for example, all V + jets events to obtain a $\not\!\!E_T$ prediction for the entire V + jets sample.

Two sets of jet $|\vec{p}_T|$ thresholds are used to measure N_J. In the first set, the jet $|\vec{p}_T|$ threshold for N_J is 50 GeV. QCD events for this jet threshold can be collected using prescaled low $|\vec{p}_T|$ jet triggers. In the second set, the jet thresholds are equal to (or higher than) the jet $|\vec{p}_T|$ thresholds that can be used in unprescaled multijet triggers. For the second set, I use 140 GeV for $N_J = 2$, 80 GeV for $N_J = 3$, and 60 GeV for $N_J \geq 4$. These jet $|\vec{p}_T|$ thresholds can be changed depending on the trigger rates in data without significant effect.

The V momentum in Z + jets and γ + jets is well measured so that the application of the algorithm is straightforward in these channels. A comparison of predicted and observed yields is shown in Figs. 2 and 3 for Z + jets and γ + jets, respectively, where the left (right) column shows results for the 50 GeV (high) jet $|\vec{p}_T|$ thresholds for N_J. It is seen that the prediction (solid line) is very good for $N_J = 3$ (50 GeV threshold), $N_J \geq 3$ (high threshold), and $N_J \geq 4$ (50 GeV and high thresholds). For events with $N_J = 2$ (50 GeV threshold) or $N_J \geq 2$ (high threshold), the measurement (dashed line) is about 20% below the prediction at $\not\!\!E_T \leq 20$ GeV. The mechanism responsible for this bias is discussed in Sec. VI. Since new physics is not expected to contribute at small $\not\!\!E_T$, to remove this bias, the prediction is normalized to the measurement in the $\not\!\!E_T \in [50, 100]$ GeV interval. This is done for the $\not\!\!E_T$ predictions in Z + jets and γ + jets events with $N_J = 2$ (50 GeV threshold) or $N_J \geq 2$ (high threshold) everywhere in the rest of this paper.

In W + jets and $t\bar{t}$ + jets events reconstructed in the l + jets + $\not\!\!E_T$ channel, there is genuine missing transverse energy from undetected neutrinos produced in W decays. Initially, to study only the effect of $\not\!\!E_T$ mismeasurements, I

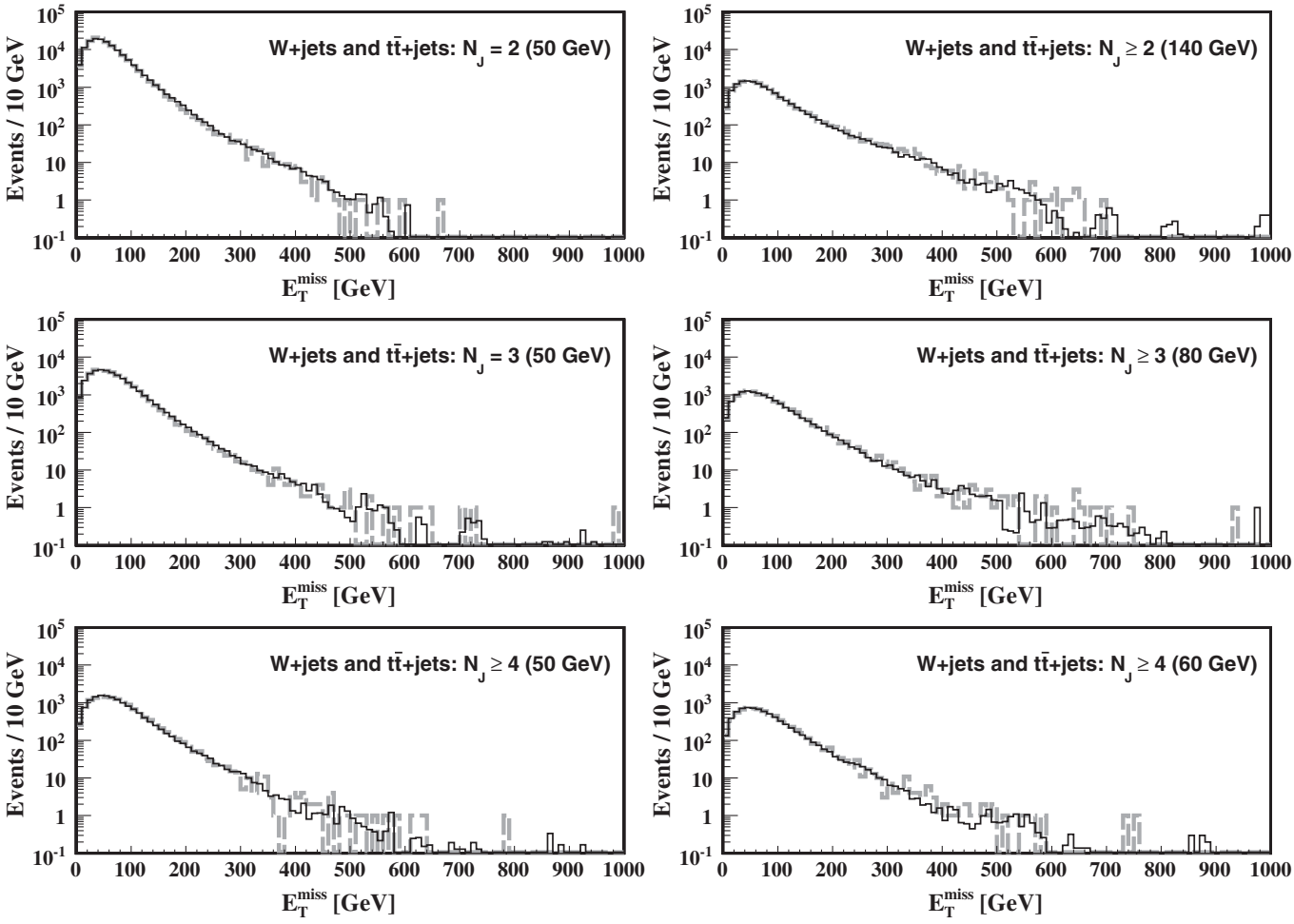

FIG. 4. Algorithm performance in l + jets + $\not\!\!E_T$ for N_J of, or at least of, 2, 3, and 4 in the first, middle and third rows. The first (second) column shows results for the 50 GeV (high) jet $|\vec{p}_T|$ threshold(s) for N_J. The observed $\not\!\!E_T$ distributions are shown in the dashed lines, their predictions are the solid lines.

consider the dominant $W \rightarrow l\nu_l$ and $t\bar{t} \rightarrow l\nu_l b\bar{b}jj$ contributions and assume that the neutrino $|\vec{p}_T|$ spectra are known until Sec. IX. To model $\not\!\!E_T$ resolution effects, the neutrino $|\vec{p}_T|$'s are smeared with the artificial $\not\!\!E_T$ predictions obtained from multijet QCD events. This is done on an event-by-event basis assuming that the neutrino $\vec{p}_T$ and the artificial $\vec{\not\!\!E}_T$ interfere at a random angle ϕ distributed uniformly from 0 to π in the transverse plane. Figure 4 shows how well the method works for the 50 GeV (first column) and high (second column) jet $|\vec{p}_T|$ thresholds for N_J in the l + jets + $\not\!\!E_T$ final state, where both W + jets and $t\bar{t}$ + jets are included according to their expected production cross sections.

In Sec. IX, it is demonstrated that one can approximate the neutrino $|\vec{p}_T|$ spectra by the charged lepton $|\vec{p}_T|$ spectra. The contribution from $W \rightarrow \tau\nu_\tau$ in W + jets and $t\bar{t}$ + jets is also considered in Sec. IX. With these extensions, the method can be used to predict the $\not\!\!E_T$ distribution in the l + jets + $\not\!\!E_T$ final state, which has high sensitivity to a variety of new physics models with new weakly interacting particles in early data.

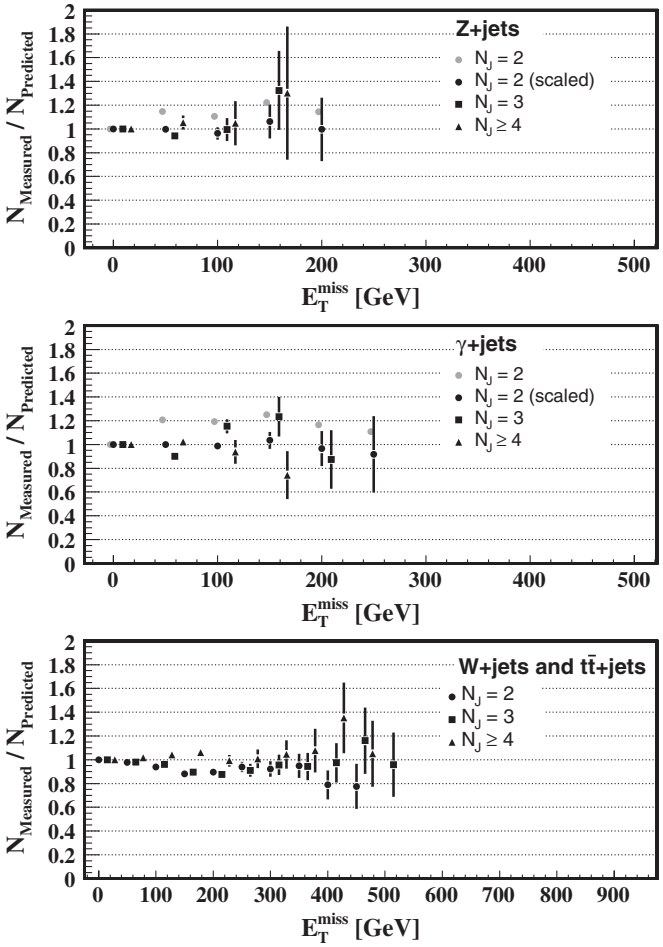

FIG. 5. Ratios of observed and estimated integrated yields for Z + jets (top), γ + jets (middle), l + jets + $\not\!\!E_T$ (bottom) obtained for the 50 GeV jet $|\vec{p}_T|$ threshold for N_J. In each plot three types of markers are shown for $N_J = 2$ (circles), 3 (squares) and ≥ 4 (triangles). The shaded markers for Z + jets and γ + jets show the ratios before the predictions are normalized at low $\not\!\!E_T$ as described in the text. Note, the ratios are correlated since yields are integrated upwards.

For brevity, in the rest of the paper, I present results of studies for the 50 GeV jet threshold used to measure N_J. They have higher statistical precision than those for higher jet $|\vec{p}_T|$ thresholds for N_J. Ratios of observed and predicted yields, $N_{\text{Observed}}/N_{\text{Predicted}}$, are shown in Fig. 5, where the yields in each $\not\!\!E_T$ bin are integrals of the distributions shown in the first columns of Figs. 2–4 from that bin's $\not\!\!E_T$ value to infinity. The algorithm performs at least as well when the set of higher jet $|\vec{p}_T|$ thresholds for N_J is used.

Since the QCD production cross section is very large at the LHC, only a small QCD sample is needed for this method to work, e.g., 1 pb^{-1} of QCD is used to model $\not\!\!E_T$ distributions in 5 fb^{-1} of Z + jets in this paper. Again, the QCD sample for templates can be collected via prescaled small $|\vec{p}_T|$ jet triggers and unprescaled high $|\vec{p}_T|$ multijet triggers. Because of the large QCD production cross section, the relative contribution from electroweak processes with genuine $\not\!\!E_T$ from neutrinos in this sample is negligible for searches in early data.

V. ROBUSTNESS

The goal of this method is to capture effects generating high artificial $\not\!\!E_T$ *in situ* using multijet QCD events. To demonstrate how well the method works, I present a set of tests in which increased jet misreconstruction is introduced. In each test, an identical change to the mock data samples for V + jets and the QCD sample is made and the analysis procedure is repeated. Figure 6 shows how drastic the effect of these changes on the $\not\!\!E_T$ distribution in γ + jets can be. (The γ + jets channel in Fig. 6 is used since there is no genuine $\not\!\!E_T$ in this final state and it has a larger yield than Z + jets.) Test results are presented in Fig. 7 for $N_J = 3$ as ratios of observed and estimated integrated yields. For brevity, test results for $N_J = 2$ and ≥ 4 are not shown but discussed and compared to those in Fig. 7.

Jet reconstruction efficiencies are not equal to unity. To test if the method models effects due to undetected jets accurately, an identical source of jet inefficiency is introduced in V + jets and QCD events. I remove jets that fall in a veto cone of $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.8$ at $(\eta, \phi) = (0.0, 0.0)$ [19], where ϕ is the azimuthal angle. Since softer jets are more likely to be lost, only nonleading jets are removed in the veto cone in this test. The effect of this inefficiency on the $\not\!\!E_T$ distribution in γ + jets for $N_J = 3$ can be assessed by comparing the solid line with the dashed grey line in Fig. 6. Test results for Z + jets, γ + jets and l + jets + $\not\!\!E_T$, W + jets and $t\bar{t}$ + jets, with $N_J = 3$ are shown in Fig. 7 in circular markers. The increased artificial $\not\!\!E_T$ tail due to lost jets is modeled accurately by the method. The same conclusion holds for $N_J = 2$ and ≥ 4.

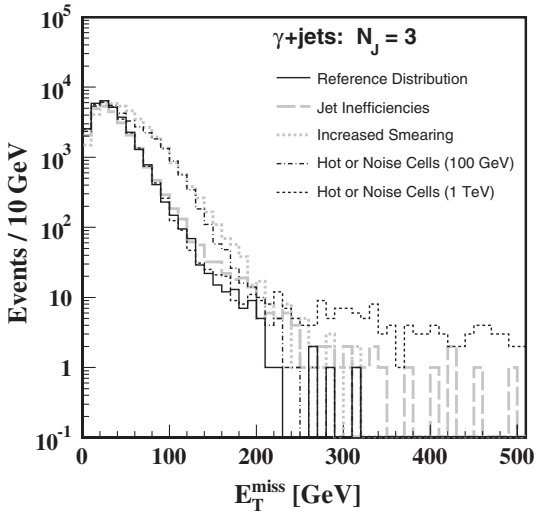

FIG. 6. Illustration of effects associated with jet misreconstruction on artificial $\not\!\!E_T$ in γ + jets in the $N_J = 3$ bin for the jet $|\vec{p}_T|$ threshold of 50 GeV. The black line is a reference $\not\!\!E_T$ distribution from Fig. 3. Jet reconstruction inefficiencies (dashed grey), increased jet energy smearing (dotted grey) and extraneous energy (dotted-dashed black and dotted black) from the tests in Secs. V and VI significantly increase artificial $\not\!\!E_T$.

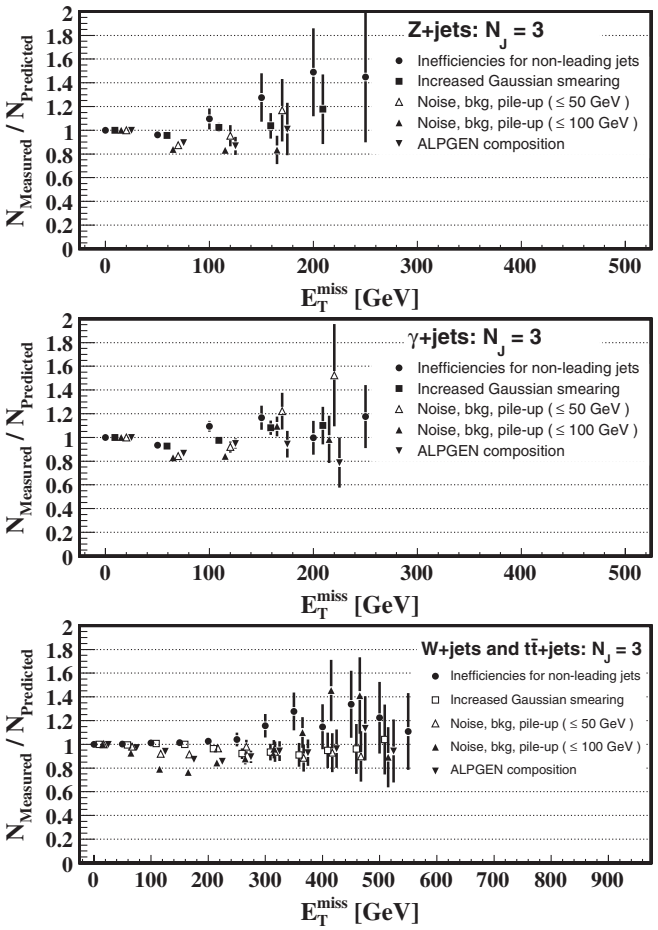

FIG. 7. Ratios of observed and estimated integrated yields in Z + jets (top), γ + jets (middle), W + jets, and $t\bar{t}$ + jets (bottom) all for $N_J = 3$ and the 50 GeV jet $|\vec{p}_T|$ threshold for N_J from robustness tests in Sec. V. Circles, squares, triangles-up, triangles-down are for tests with increased inefficiencies for nonleading jets, increased Gaussian jet energy smearing, extraneous energy contributions and a modified N_J composition of the QCD sample. Note, ratios in the plots are correlated as tests are made using events drawn from the same mock data samples, and yields are integrated upwards.

In the next test, the jet energy smearing is increased. Two tests are made: (a) the Gaussian $\sigma(|\vec{p}_T|)$ is doubled with the area and shape of the non-Gaussian tails unchanged, and (b) the area of the low and high-side non-Gaussian tails is doubled with the Gaussian component kept unchanged. The effect of the additional jet energy smearing in test (a) on the $\not\!\!E_T$ distribution in γ + jets for $N_J = 3$ is shown in Fig. 6 in the dotted grey line. Ratios of observed and estimated yields from test (a) are in square markers in Fig. 7 for $N_J = 3$. Similar results are observed in the other two N_J bins, and in test (b). Again, templates constructed from multijet QCD events capture effects from additional jet smearing *in situ* so that the level of background at high $\not\!\!E_T$ in V + jets is predicted accurately.

Hot cells or noise in the calorimeters, backgrounds from the proton beams, cosmic rays, underlying event or additional pp interactions in the same bunch crossing contribute extra energy and jets erroneously attributed to those produced in V + jets and QCD processes. Since additional jets have a higher probability to be soft, I test the method's ability to model such effects by adding extra jets with a soft uniform $|\vec{p}_T|$ spectrum from 0 to 50 GeV with a 20% probability to each V + jets or QCD event. These extra jets change J_T and $\not\!\!E_T$, but do not change N_J. The predictions are good in this test as seen in Fig. 7 in open triangular (up) markers.

I repeat the previous test with a uniform $|\vec{p}_T|$ spectrum of additional energy contributions covering the range from 0 to 100 GeV added with a 10% probability to V + jets and QCD events. This produces a strong effect on the $\not\!\!E_T$ distribution shown for γ + jets with $N_J = 3$ in the dotted-dashed line in Fig. 6. Ratios of observed and estimated yields for the three V + jets processes in the $N_J = 3$ bin are in solid triangular (up) markers in Fig. 7. I find that the prediction is consistent with the measurement to about 20% or better in the $N_J = 3$ and ≥ 4 bins. In the $N_J = 2$ bin in Z + jets and γ + jets, a bias is observed. The origin of this bias stems from differences in N_J and J_T distributions between V + jets and QCD. I discuss it and more stringent tests with extraneous energy contributions in the next section.

The cross section ratios for V + jets and QCD processes: $\sigma^{V+\text{jets}}(n\text{jets})/\sigma^{\text{QCD}}(n\text{jets})$, $\sigma^{V+\text{jets}}(n\text{jets})/\sigma^{V+\text{jets}}(n+1\text{jets})$ or $\sigma^{\text{QCD}}(n\text{jets})/\sigma^{\text{QCD}}(n+1\text{jets})$, where n is equal to 2 or more, in LHC data are likely to differ from that of ALPGEN used in this study. There may also be

differences in other differential distributions in the jet system of V + jets or multijet QCD events between LHC data and ALPGEN. To test how sensitive the method is to such differences, I vary the ALPGEN ratios $\sigma^{\mathrm{QCD}}(n\,\mathrm{jets})/\sigma^{\mathrm{QCD}}(n+1\,\mathrm{jets})$, $n \geq 2$, by a factor of 1.5 up or down. Test results with reduced ratios for $N_J = 3$ are shown in Fig. 7 in triangular (down) markers. The $\not\!\!E_T$ predictions are good in this test because they are made on an event-by-event basis using QCD events with the same N_J and J_T. QCD events with other values of N_J and J_T are included only if they are misreconstructed, which is a second order effect, but it can become significant in regimes where distributions fall or rise steeply. Test results for $N_J = 3$ and ≥ 4 are all good. For $N_J = 2$ in Z + jets and γ + jets, when the $\sigma^{\mathrm{QCD}}(n\,\mathrm{jets})/\sigma^{\mathrm{QCD}}(n+1\,\mathrm{jets})$ ratios are reduced the prediction improves; when the ratios are increased the prediction becomes biased. The origin of this bias is the same as in the previous test and is discussed in the next section.

In conclusion of this section, the quality of the $\not\!\!E_T$ prediction improves at larger N_J. The $\not\!\!E_T$ prediction is robust for $N_J = 3$ and ≥ 4 in all tests. Events with $N_J = 2$ are more susceptible to biases for two reasons. First, there are significant differences in the differential distributions describing jets in QCD and V + jets: in QCD dijets, the jets come mainly from leading order parton interactions, while in V + jets, the jets are from higher order processes. Second, the averaging effects discussed in Sec. IV are not as strong when the number of jets is small. Nevertheless, only two tests for $N_J = 2$ are biased in this section. Any other effect that generates artificial $\not\!\!E_T$ in the same manner in the jet system of V + jets and multijet QCD events should be modeled *in situ* by the method. I next discuss the method's limitations revealed in more stringent tests.

VI. LIMITATIONS

I increase the degree of jet misreconstruction up to a point where the method becomes biased to explore the boundaries of the domain where the method works. This allows to understand in greater detail mechanisms that may lead to a bias. At the end of this section, I discuss how to avoid regimes where the method is biased.

The test with the jet veto cone introduced in the previous section is repeated with a modification such that leading jets falling into the veto cone are removed. This is a stringent test since leading jets are less likely to be undetected. Test results are shown in Fig. 8 in circular markers, for brevity, only for γ + jets in the $N_J = 3$ bin. While the prediction partly takes into account the effect of undetected leading jets, it underestimates the background at high $\not\!\!E_T$ in that N_J bin in Z + jets and γ + jets. The prediction is biased because in QCD events $\not\!\!E_T$ is always less than J_T, by the definition of $\not\!\!E_T$ in Sec. III. In V + jets, V is a Z or γ here, $\not\!\!E_T$ can be greater than J_T when the leading jet recoiling against an energetic V boson in the transverse plane is lost. The V + jets events with $\not\!\!E_T$ larger than J_T can not be modeled by the algorithm in Sec. IV. This bias is larger for $N_J = 2$, while in the $N_J \geq 4$ bin, the prediction is good for both Z + jets and γ + jets. In l + jets + $\not\!\!E_T$, W + jets and $t\bar{t}$ + jets combined, due to a genuine $\not\!\!E_T$ contribution from neutrinos to the full $\not\!\!E_T$, this bias does not appear.

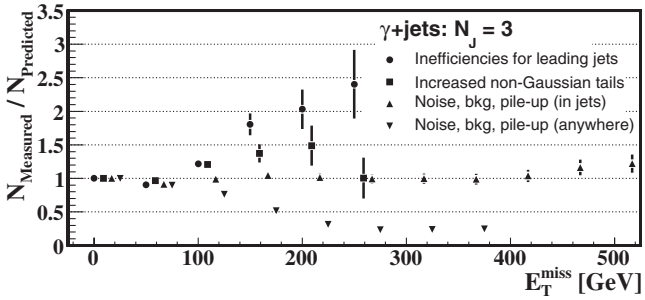

FIG. 8. Ratios of observed and estimated integrated yields in γ + jets for $N_J = 3$ and the 50 GeV jet $|\vec{p}_T|$ threshold for N_J from tests in Sec. VI. Circles, squares, triangles-up and triangles-down are for tests with increased inefficiencies for leading jets, increased non-Gaussian jet energy fluctuations and extraneous energy contributing in jets or anywhere in the calorimeters. Note, these ratios are correlated as tests are made using events drawn from the same mock data samples, and yields are integrated upwards.

I repeat the test with increased jet energy mismeasurements after tripling the area of the lower non-Gaussian tail in the jet response function and reducing the magnitude of its slope on the logarithmic scale of the lower plot in Fig. 1 by a factor of 2. The prediction is biased in the $N_J = 2$ bin for both Z + jets and γ + jets. The quality of the prediction improves in the $N_J = 3$ bin, shown in Fig. 8 in square markers for γ + jets, and it is good for $N_J \geq 4$ in Z + jets and γ + jets. One should expect a bias for large low-side tails in the jet response function appearing via the same mechanism operating in the previous test. The V + jets events containing jets fluctuated down in $|\vec{p}_T|$ can have J_T that is less than $\not\!\!E_T$. Such events can not be modeled by the algorithm of Sec. IV. In the l + jets + $\not\!\!E_T$ final state, the prediction of the full $\not\!\!E_T$, which includes the neutrino momentum, is good in all N_J bins. Note, for a large low-side tail in the jet response function, the jet energy scale may become biased. Effects due to a jet energy scale offset are discussed in Sec. VIII.

I make two tests with a harder spectrum of additional energy contributions unrelated to γ + jets and QCD events. In these tests, the spectrum of additional energy is uniform in E_T from 0 to 1 TeV added with a 1% probability to both V + jets and QCD. Since there is no genuine $\not\!\!E_T$ in γ + jets and QCD, the requirement on $|\Delta\phi^{\text{lead jet}\text{-}\not\!\!E_T}|$ (Sec. III) that $\vec{\not\!\!E}_T$ and the leading jet be not aligned in the transverse plane removes a fraction of events with high E_T extraneous contributions. In the first test,

additional energy depositions contribute only to jets that are above the $|\vec{p}_T|$ threshold for N_J. The $\not\!\!E_T$ distribution in γ + jets, $N_J = 3$, is shown in the dotted black line in Fig. 6 with a large artificial high $\not\!\!E_T$ tail. Ratios of observed and estimated yields are in triangular (up) markers in Fig. 8. In γ + jets, the prediction is good for $N_J = 3$ and ≥ 4, and it is biased in the $N_J = 2$ bin for the following reason. The J_T spectrum in QCD events tends to be softer than that in V + jets events with the same N_J. (This effect is most pronounced for $N_J = 2$.) The fraction of soft QCD multijet events promoted to higher J_T by extraneous energy depositions tends to be larger than that fraction in V + jets. Since such events have larger $\not\!\!E_T$ due to the extraneous energy depositions unbalanced in the transverse plane, the level of background at high $\not\!\!E_T$ is overestimated.

In the second test, extraneous energy contributions are added randomly in the $\eta - \phi$ plane so that N_J also tends to increase. Ratios of observed and estimated yields for γ + jets, $N_J = 3$, are in triangular (down) markers in Fig. 8. The prediction overestimates the background in all N_J bins. This happens because $\sigma_{\rm QCD}(n\,{\rm jets})/\sigma_{\rm QCD}(n+1\,{\rm jets})$, $n \geq 2$, is higher than $\sigma_{V+{\rm jets}}(n\,{\rm jets})/\sigma_{V+{\rm jets}}(n+1\,{\rm jets})$ in the mock data samples so that the fraction of events with $N_J = n$ reconstructed erroneously in the $N_J = (n+1)$ bin due to an extra energy deposition is higher in QCD compared to that fraction in V + jets. (Again, this effect is most pronounced for $N_J = 2$.) Since these misreconstructed events have larger $\not\!\!E_T$, the prediction overestimates the background. The mechanisms leading to a bias described in this and the previous paragraphs are also responsible for biases noted in the previous sections.

Test results with extraneous energy contributions for Z + jets are qualitatively similar to those for γ + jets. In l + jets + $\not\!\!E_T$, the biasing effects discussed above are intertwined with additional effects due to the presence of a neutrino in the final state and the $t\bar{t}$ + jets contribution. The genuine $\not\!\!E_T$ from the neutrino makes the requirement on $|\Delta\phi^{\text{lead jet-}\not\!\!E_T}|$ less efficient in suppressing high E_T extraneous contributions. The $t\bar{t}$ + jets events contribute to further differences in N_J and J_T spectra between W + jets and QCD. I find that in l + jets + $\not\!\!E_T$ the prediction tends to overestimate the background in the tests with extraneous energy depositions, and the quality of the prediction improves with N_J.

Regimes with severely misreconstructed events where the method may become biased need to be avoided. By imposing event quality criteria or improving the jet reconstruction, e.g., using the tracking systems, [20] one can reduce the number of such events. Moderately misreconstructed events are modeled *in situ* by the method. The V + jets sample with $N_J = 2$ is the most challenging for this method. This makes two jet events a good sample with which to validate the algorithm in data. The method performs better at higher J_T and N_J, where the sensitivity to new physics is higher. There are several reasons for that a) there are fewer differences between the hadronic systems in V + jets and QCD, b) the averaging effects over V + jets and QCD events are stronger, and c) the jet reconstruction performs better at higher jet $|\vec{p}_T|$.

VII. $t\bar{t}$ + jets

SM $t\bar{t}$ + jets events, where $t\bar{t} \rightarrow l\nu_l b\bar{b}q\bar{q}$, constitute a dominant background in the l + jets + $\not\!\!E_T$ signature for $N_J \geq 3$. The shapes of N_J and J_T spectra in these events differ from those in QCD events collected for templates and from those in V + jets. The calorimeter response to b jets in $t\bar{t}$ + jets differs from that of light quark and gluon jets [21]. These effects lead to a bias in the prediction of artificial $\not\!\!E_T$ in $t\bar{t}$ + jets. To demonstrate this bias clearly, Fig. 9 (top) shows the artificial $\not\!\!E_T$ in $t\bar{t}$ + jets for $N_J \geq 4$ (dashed line), where the neutrino four-momentum is assumed to be measured so that it is included in the $\not\!\!E_T$ calculation, and its prediction using QCD templates (solid line). Note, at large $\not\!\!E_T$, this bias is an order of magnitude smaller compared to the genuine $\not\!\!E_T$ from the neutrino in the final state having the $|\vec{p}_T|$ spectrum shown in the dotted-dashed line in the same figure. When the neutrino $|\vec{p}_T|$ spectrum is combined with the artificial $\not\!\!E_T$ in the full $\not\!\!E_T$ prediction, the bias becomes insignificant as seen in Fig. 9 (bottom).

The artificial $\not\!\!E_T$ is a dominant contributor in events with small genuine $\not\!\!E_T$. Figure 9 (bottom) shows that the accuracy of its prediction is sufficient to model the full $\not\!\!E_T$

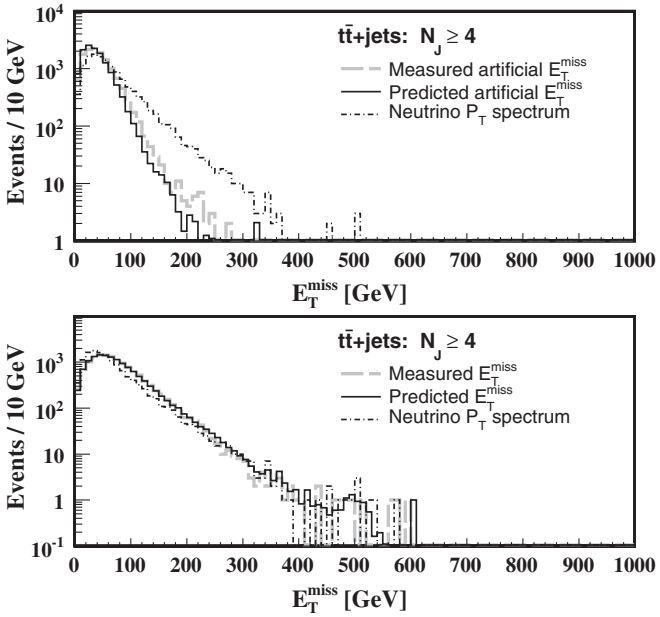

FIG. 9. Top: artificial $\not\!\!E_T$ in $t\bar{t}$ + jets (dashed) and its prediction (solid) for $N_J \geq 4$ and the 50 GeV jet $|\vec{p}_T|$ threshold for N_J. Bottom: full $\not\!\!E_T$ (dashed) and its prediction (solid) combining both the neutrino $|\vec{p}_T|$ spectrum and the artificial $\not\!\!E_T$ prediction for the same N_J and jet $|\vec{p}_T|$ selection. In both plots the neutrino spectrum is shown as the dotted-dashed line.

distribution at small $\not\!\!E_T$. At high $\not\!\!E_T$, the missing momentum from the neutrino dominates over artificial $\not\!\!E_T$ so that the accuracy of the full $\not\!\!E_T$ prediction is highly dependent on how well the neutrino spectrum is modeled. The modeling of neutrino $|\vec{p}_T|$ spectra is discussed in Sec. IX.

Despite the fact that the bias in the artificial $\not\!\!E_T$ prediction for $t\bar{t}$ + jets is insignificant in the full $\not\!\!E_T$ prediction in l + jets + $\not\!\!E_T$, it is instructive to examine how it behaves when selection criteria or the algorithm of Sec. IV are modified. Two observations can be made. First, the bias becomes smaller when the jet $|\vec{p}_T|$ threshold for $\not\!\!E_T$ and J_T is reduced or the η coverage for jets is increased since the total energy is collected to a fuller extent with more inclusive requirements. Optimal requirements on these variables can only be determined using data because at smaller $|\vec{p}_T|$ and larger $|\eta|$ more noise and backgrounds are expected. Second, in $t\bar{t}$ + jets, there tends to be more jets included in the $\not\!\!E_T$ and J_T calculations that are below the jet $|\vec{p}_T|$ threshold for N_J. Since the jet resolution improves as the jet $|\vec{p}_T|$ grows, the prediction can be improved by making $\not\!\!E_T$ templates in coarse bins of $R(J_T) = J_T^{\text{high}}/J_T$, where J_T^{high} is a scalar sum of jet $|\vec{p}_T|$'s for jets above the $|\vec{p}_T|$ threshold for N_J. Alternatively, the same effect can be achieved by modifying the composition of the QCD sample used for templates. Finally, the modeling of b jets in $t\bar{t}$ + jets can be improved by removing a fraction of jet $|\vec{p}_T|$ measurements in QCD events that is expected to be carried by muons and neutrinos in semileptonic decays of beauty and charm quarks in b jets.

VIII. JET ENERGY SCALE

Jet energy measurements could be systematically biased in early data. Let us consider a case when jet energies are under-measured uniformly in jet $|\vec{p}_T|$. Such mismeasurements cancel to first order in $\not\!\!E_T$ measurements in QCD events. In V + jets, since the jet system recoils against the V, the jet energy mismeasurements add up coherently along the V direction in the transverse plane. To avoid a bias due to this difference, the jet energy scale needs to be calibrated. Since the method is capable to model large tails in the jet response function, a precise calibration of the jet energy scale as a function of η and ϕ (azimuthal angle) is not required. The jet energy scale can be calibrated with sufficient accuracy using standard techniques based on γ + jets and Z + jets ($N_J \leq 2$) processes [20] in very early data.

Figure 10 gives a comparison of $\not\!\!E_T$ distributions in Z + jets (top) and l + jets + $\not\!\!E_T$ (bottom) for $N_J = 3$ without (solid) and with a 10% (dashed) and 20% (dotted) jet energy scale offset downwards uniform in jet $|\vec{p}_T|$. The prediction becomes good for a 10% or smaller offset in Z + jets and γ + jets. One may reduce the effect from a residual jet energy scale offset on $\not\!\!E_T$ in Z + jets and γ + jets by normalizing the predicted $\not\!\!E_T$ shape to the observed

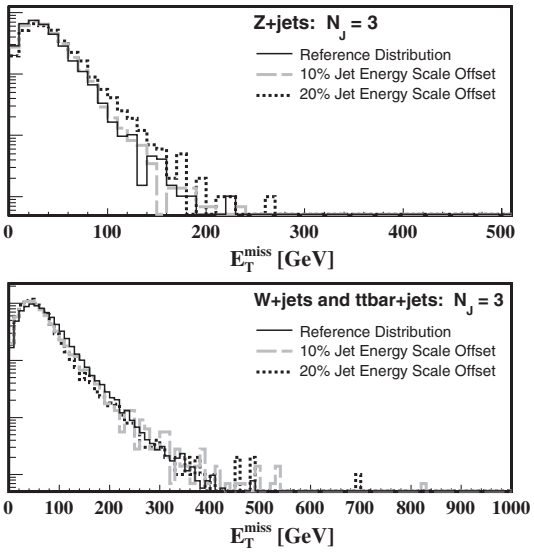

FIG. 10. The $\not\!\!E_T$ distributions without (solid) and with 20% (dotted) or 10% (dashed) jet energy scale offset in Z + jets (top) and l + jets + $\not\!\!E_T$ (bottom) for $N_J = 3$ with the 50 GeV jet threshold for N_J. The distributions in each plot are normalized to the same area.

distribution in the small $\not\!\!E_T$ region, for example, for $\not\!\!E_T \in [50, 100]$ GeV. Demands on the precision of the jet energy calibration in l + jets + $\not\!\!E_T$ are higher. Finally, even before the jet energy scale is calibrated, one can make a search in the projection of $\vec{\not\!\!E}_T$ on the axis perpendicular to the V direction (the l direction in l + jets + $\not\!\!E_T$) in the transverse plane, $\not\!\!E_{TT}$. Searches in $\not\!\!E_{TT}$ are less sensitive to effects associated with the jet energy scale offset since those lead to a bias along the V direction.

IX. NEUTRINO SPECTRA IN W DECAYS

In the l + jets + $\not\!\!E_T$ signature, dominated by W + jets and $t\bar{t}$ + jets, there are one or more undetected neutrinos in the final state. To model $\not\!\!E_T$ in these events, one needs a prediction or a measurement of the neutrino $|\vec{p}_T|$ spectra, which can be combined with $\not\!\!E_T$ resolution predictions from QCD templates. The neutrino $|\vec{p}_T|$ spectra could be obtained from MC simulation. Or, the neutrino $|\vec{p}_T|$ spectra can be modeled in a data-driven manner using charged lepton $|\vec{p}_T|$ spectra as described in this section.

A. $W \to l\nu_l$

The solid and dashed lines in plot (a) of Fig. 11 are the neutrino and charged lepton $|\vec{p}_T|$ spectra in W + jets for $N_J = 3$, $W \to l\nu_l$, passing all selection of Sec. III but the requirement on the charged lepton $|\vec{p}_T|$ of at least 20 GeV. It is seen that the two $|\vec{p}_T|$ spectra have consistent shapes so that the charged lepton $|\vec{p}_T|$ spectra can be used to model the neutrino $|\vec{p}_T|$ spectra in W + jets. Note, the W bosons in W + jets tend to be produced in the transverse-minus

helicity state (left-handed) rather than in the transverse-plus helicity state (right-handed). These polarization effects are present even in the transverse plane for $N_J \geq 2$ so that W^+ (W^-) bosons in W + jets tend to produce charged leptons with a $|\vec{p}_T|$ spectrum that is softer (harder) compared to the neutrino $|\vec{p}_T|$ spectrum. However, in the entire W + jets sample, W^+ and W^- combined, these polarization effects are averaged and largely disappear [22] so that the charged lepton and neutrino $|\vec{p}_T|$ spectra have very similar shapes seen in plot (a) of Fig. 11. The application of a $|\vec{p}_T|$ threshold on the charged lepton makes its spectrum harder, while the neutrino spectrum becomes softer, as seen in plot (b) of the same Figure. To model the neutrino spectrum using the reconstructed charged lepton spectrum in W + jets, the effect of the charged lepton $|\vec{p}_T|$ threshold and the residual differences between charged lepton and neutrino $|\vec{p}_T|$ spectra need to be corrected. The corrections can be obtained from MC simulation.

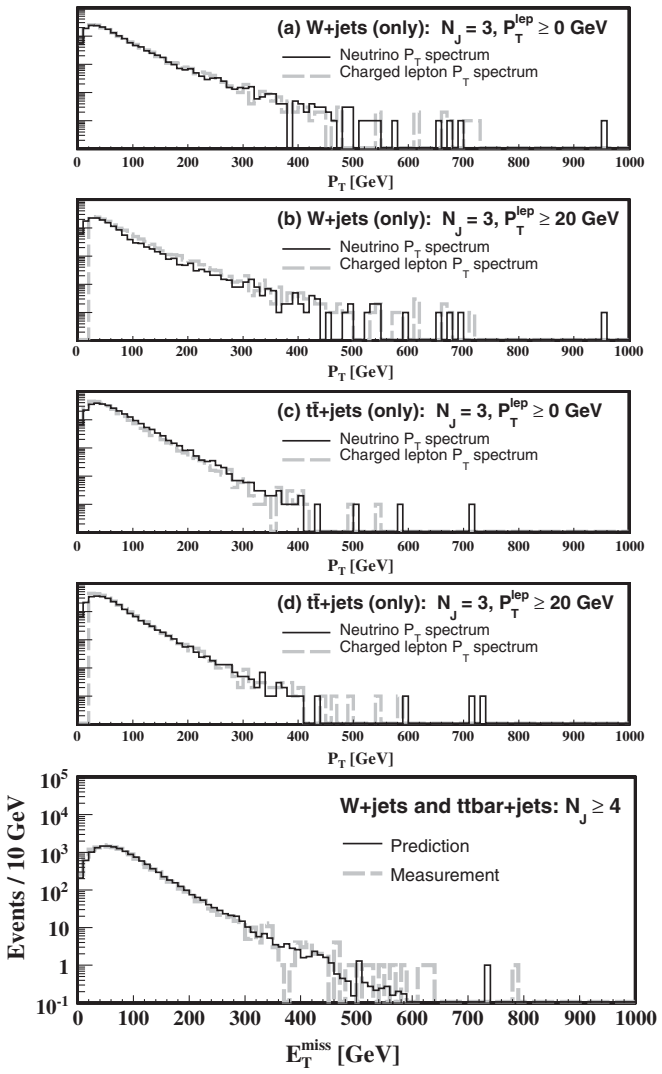

FIG. 11. Comparisons of charged lepton and neutrino spectra in W + jets (a, b) and $t\bar{t}$ + jets (c, d) without (a, c) and with (b, d) charged lepton $|\vec{p}_T|$ thresholds all for $N_J = 3$. The lower plot shows a $\not\!\!E_T$ prediction based on the charged lepton spectrum in l + jets + $\not\!\!E_T$, $N_J \geq 4$, without any corrections. In all plots, the threshold for N_J is 50 GeV. (Different sequences of random numbers are used to smear jet energies in these plots.)

The solid and dashed lines in plot (c) of Fig. 11 are the neutrino and charged lepton $|\vec{p}_T|$ spectra in $t\bar{t}$ + jets for $N_J = 3$, $t\bar{t} \rightarrow l\nu_l b\bar{b}q\bar{q}$, after the selection of Sec. III but without a threshold requirement on the charged lepton $|\vec{p}_T|$. In the SM, 30% of W^+ (W^-) bosons in t ($\bar{t}$) quark decays are left-handed (right-handed) and the rest are longitudinally polarized [23]. Left-handed W^+ and right-handed W^- bosons tend to produce charged leptons with a $|\vec{p}_T|$ spectrum that is softer compared to the neutrino $|\vec{p}_T|$ spectrum as seen in plot (c). Since the two spectra have similar shapes, it is possible to use the charged lepton $|\vec{p}_T|$ spectrum to model the neutrino spectrum $|\vec{p}_T|$ in $t\bar{t}$ + jets. Again, when a charged lepton $|\vec{p}_T|$ threshold is applied, the charged lepton spectrum becomes harder while the neutrino spectrum becomes softer, which leads to a higher consistency between the two spectra seen in plot (d). Nevertheless, the effects of the W polarization in top decays and the event selection, mainly due to the charged lepton $|\vec{p}_T|$ threshold, in $t\bar{t}$ + jets, in general, need to be corrected.

In order to determine corrections to the charged lepton spectra for W + jets and $t\bar{t}$ + jets from MC simulation, one needs to measure the shape of the $|\vec{p}_T|$ dependence of lepton reconstruction efficiencies and the relative fractions of W + jets and $t\bar{t}$ + jets in the data sample. The former can be readily done via a standard technique based on $Z \rightarrow l^+l^-$ decays [24]. The latter should come from an independent measurement. With these two ingredients, corrections can be determined from MC simulation.

Since corrections to the charged lepton spectra are small, the reliance on details of MC simulation to determine the neutrino $|\vec{p}_T|$ spectra is minimal. For a 20 or 15 GeV threshold on charged lepton $|\vec{p}_T|$, no corrections are required to predict the $\not\!\!E_T$ distributions in $t\bar{t}$ + jets in all N_J bins to 20% or better in the mock data samples. Corrections are needed for W + jets. The lower plot in Fig. 11 shows the $\not\!\!E_T$ distribution and its prediction in l + jets + $\not\!\!E_T$, W + jets and $t\bar{t}$ + jets combined, for $N_J \geq 4$ based on the charged lepton spectrum without corrections. Since $t\bar{t}$ + jets dominates over W + jets in the $N_J \geq 4$ ($N_J = 3$) bin, the prediction is good to 15% (25%) at high $\not\!\!E_T$ without corrections. The $N_J \geq 4$ bin, where the prediction is the most robust, is likely to have the highest sensitivity to a new physics contribution compared to lower jet multiplicity events.

B. $W \rightarrow \tau\nu_\tau$

In the l + jets + $\not\!\!E_T$ signature, there is background from tauonic W decays in W + jets and $t\bar{t}$ + jets. Tauonic W

decays produce at least one additional neutrino that is a source of differences between the charged lepton and neutrino $|\vec{p}_T|$ spectra.

There are two types of tauonic W decays that contribute significant background: (1) W + jets and $t\bar{t}$ + jets, where $W^- \to \tau\bar{\nu}_\tau$ with $\tau \to l\bar{\nu}_l\nu_\tau$, and (2) $t\bar{t}$ events, where $W^- \to l\bar{\nu}_l$ and $W^+ \to \bar{\tau}\nu_\tau$ with $\bar{\tau} \to$ (hadrons $\bar{\nu}_\tau$). The contribution from tauonic W decays is an order of magnitude smaller compared to that from $W \to l\nu_l$ decays. (The tauonic background of type 2 can be suppressed by vetoing events with isolated single hadronic tracks.) The τ branching fractions are well known. Therefore, the effects from $W \to \tau\nu_\tau$ on $\not\!\!E_T$ predictions can be well modeled by an additional smooth correction to the charged lepton $|\vec{p}_T|$ spectra that can be determined by MC simulation. Since the contribution from tauonic W decays is smaller compared to that from $W \to l\nu_l$ decays in l + jets + $\not\!\!E_T$, corrections for them are not discussed further in this paper.

X. SYSTEMATIC UNCERTAINTIES

Systematic uncertainties need to account for the statistical precision and biases of the method's background predictions at high $\not\!\!E_T$. Mechanisms by which biases may appear are discussed in Sec. VI. The method's susceptibility to them can be studied in both data and MC.

A sample of events with $N_J = 2$ for a small jet $|\vec{p}_T|$ threshold, such as 50 GeV, is more sensitive to biases. The relative contribution from new physics can not be large in this sample. Therefore, these events can be used to validate the method's performance and place an upper bound on its biases at higher N_J and jet $|\vec{p}_T|$. Similarly, the application of event quality criteria are expected to reduce the number of severely misreconstructed events that may lead to biases in the prediction of artificial $\not\!\!E_T$. By varying the event quality selection criteria, one can determine if the method is subject to such biases or estimate their size. An excess due to a new physics contribution should be stable under variations of these criteria.

In the l + jets + $\not\!\!E_T$ channel, lepton $|\vec{p}_T|$ spectra are used to model neutrino $|\vec{p}_T|$ spectra. There are several sources of the systematic uncertainty associated with this modeling as MC is used to obtain corrections to the charged lepton $|\vec{p}_T|$ spectra. Because these corrections are small, uncertainties due to MC used to extract them enter only at second order. They can be estimated by varying the composition of the MC samples used to measure them and the reconstruction efficiencies of leptons and jets within their uncertainties. The uncertainties in the composition of the MC samples should come from an independent measurement of the relative W + jets and $t\bar{t}$ + jets cross sections for different N_J. Note, in Sec. IX A, it is demonstrated that these corrections may be negligible for $N_J \geq 4$ to obtain a prediction at high $\not\!\!E_T$ to 20% or better in early data.

The QCD background to signal events with one or more fake leptons or photons, cross-feeds among V + jets processes and other secondary backgrounds are not considered in this paper. These backgrounds as well as dileptons from $t\bar{t}$ + jets could be accounted in the $\not\!\!E_T$ distributions and their predictions by the following procedure. For each significant secondary background contribution, one can obtain a control sample in data and estimate the number of events from this background contribution in the entire search sample [6,24]. Next, one can measure the $\not\!\!E_T$ distribution and make its prediction in that control sample, and normalize both to the expected number of events for this background contribution. The $\not\!\!E_T$ distribution and its prediction from the control sample could then be subtracted from the $\not\!\!E_T$ distribution and its prediction for the entire sample, respectively.

A large new physics contamination to QCD at large J_T, in general, may bias the prediction at large $\not\!\!E_T$ and hide a new physics contribution to V + jets. I find that even under the most optimistic scenarios for new physics cross sections such a contamination does not lead to a significant bias.

Even though the reliance on MC is much reduced in this method, MC can be used to validate the method and constrain its systematic biases as is done in this paper. Nevertheless, a study of control data samples is needed to develop, optimize and validate the final algorithm and to quantify its systematic uncertainties.

XI. PREDICTIONS WITH SIGNAL

The algorithm's performance with a new physics contribution is illustrated in Fig. 12 in the Z + jets and l + jets + $\not\!\!E_T$ channels in events with $N_J = 3$ (left) and ≥ 4 (right) for the 50 GeV threshold. Figure 13 shows the corresponding distributions in events with $N_J \geq 3$ (left) and ≥ 4 (right) for the high $|\vec{p}_T|$ jet thresholds. The integrated luminosity of the mock data samples in these figures is 200 pb^{-1} for $\sqrt{s} = 14$ TeV. New physics contributions in the figures are similar to those from mSUGRA benchmarks LM4 and LM1 [17] for Z + jets and l + jets + $\not\!\!E_T$, respectively. The plots show SM backgrounds with new physics contributions (dashed) and their $\not\!\!E_T$ predictions (solid) from QCD templates. The dotted-dashed lines represent SM backgrounds only to ease comparisons.

New physics events tend to have large J_T and $\not\!\!E_T$. It is seen that the addition of a signal contribution with large J_T does not bias the prediction significantly at high $\not\!\!E_T$. An excess of signal events above the background prediction stands out clearly in both channels. Since in l + jets + $\not\!\!E_T$ the neutrino spectrum is modeled based on the charged lepton spectrum in each N_J bin, the method works best in this signature when the charged lepton spectrum in new physics events is soft compared to the $|\vec{p}_T|$ spectrum produced by new weakly interacting particles [25].

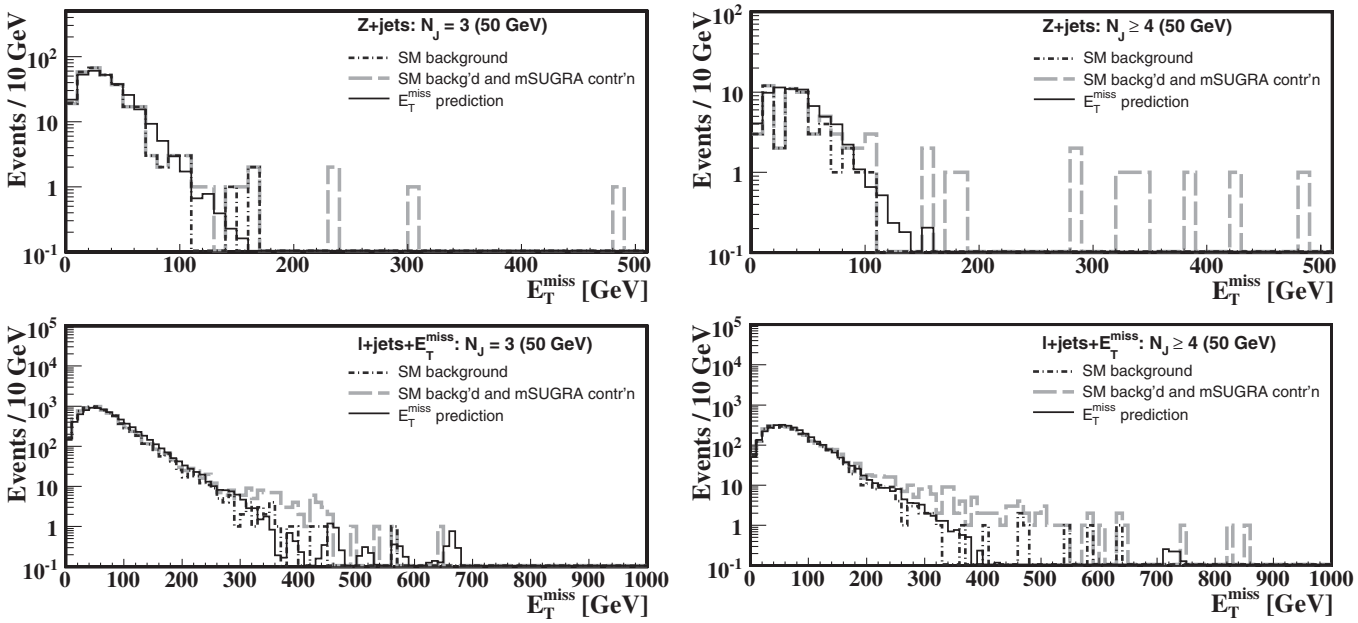

FIG. 12. Observed (dashed) and predicted (solid) SM Z + jets (top) and l + jets + $\not\!\!E_T$ (bottom) for $N_J = 3$ (left) and $N_J \geq 4$ (right) with new physics contributions from mSUGRA benchmarks [17]. The dotted-dashed lines highlight the SM contributions. The $|\vec{p}_T|$ threshold for N_J is 50 GeV. The plots correspond to 200 pb^{-1} at $\sqrt{s} = 14$ TeV.

XII. CONCLUSION

I have presented a new method to predict SM backgrounds at high $\not\!\!E_T$ and a large number of jets, N_J, within a context of a search for new phenomena in final states consistent with SM V + jets: Z + jets, γ + jets and W + jets. The artificial $\not\!\!E_T$ in each V + jets event is modeled *in situ* using multijet QCD events with a configuration of jets similar to that in the V + jets event. The genuine $\not\!\!E_T$ contribution from neutrinos in the l + jets + $\not\!\!E_T$ channel, dominated by W + jets and $t\bar{t}$ + jets, is modeled based on the charged lepton $|\vec{p}_T|$ spectrum.

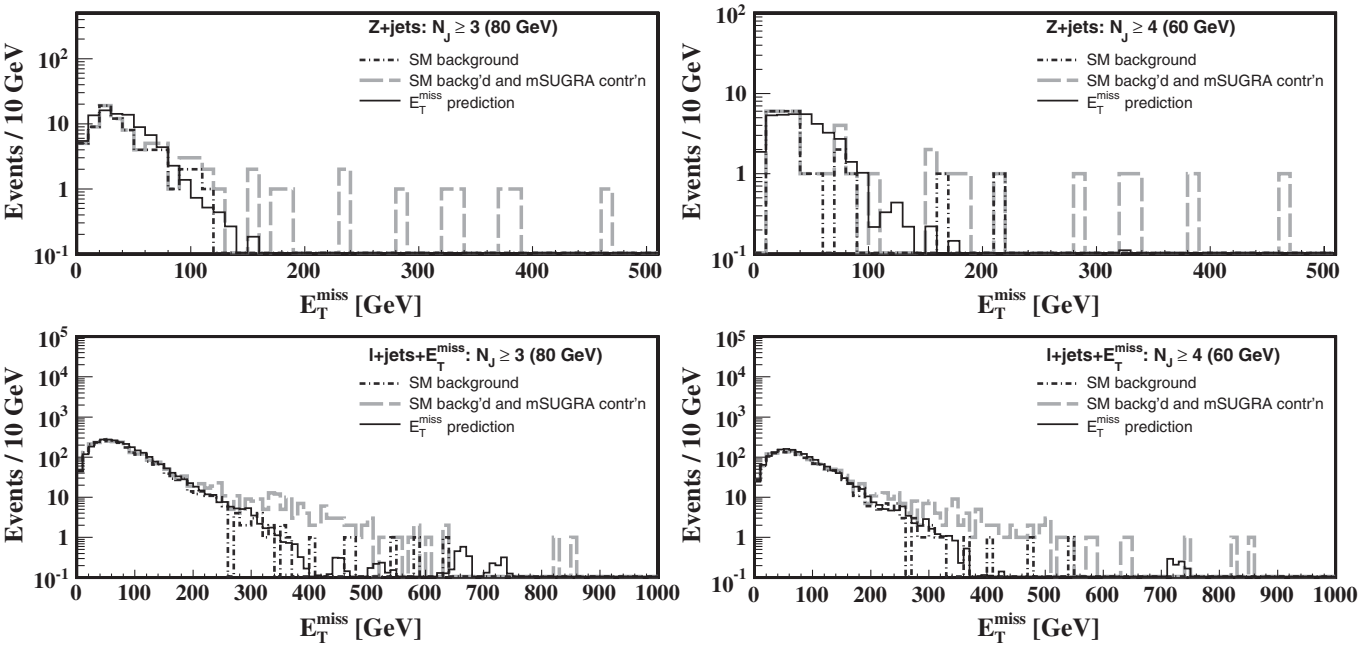

FIG. 13. Observed (dashed) and predicted (solid) SM Z + jets (top) and l + jets + $\not\!\!E_T$ (bottom) for $N_J \geq 3$ (left) and $N_J \geq 4$ (right) with new physics contributions from mSUGRA benchmarks [17]. The dotted-dashed lines highlight the SM contributions. The plots are made for the high $|\vec{p}_T|$ thresholds for N_J and correspond to 200 pb^{-1} at $\sqrt{s} = 14$ TeV.

The method performs reasonably well in robustness tests. I have identified mechanisms by which it may become biased, discussed systematic uncertainties in its background predictions and procedures to estimate them. A new physics contamination of the QCD sample does not lead to a significant bias. The method has discriminating power to reveal a new physics contribution at high $\not\!\!E_T$ and a large number of jets. It can be applied to data with minimal recourse to MC simulation in early LHC running when robust data-driven SM background predictions play a key role in searches for new phenomena.

[1] J. Dunkley *et al.*, Astrophys. J. Suppl. Ser. **180**, 306 (2009); G. Bertone, D. Hooper, and J. Silk, Phys. Rep. **405**, 279 (2005), and references therein.

[2] J. Wess and B. Zumino, Nucl. Phys. **B70**, 39 (1974); Y. A. Golfand and E. P. Likhtman, JETP Lett. **13**, 323 (1971); D. V. Volkov and V. P. Akulov, JETP Lett. **16**, 438 (1972).

[3] For example, models of little Higgs with T-parity: N. Arkani-Hamed, A. G. Cohen, and H. Georgi, Phys. Lett. B **513**, 232 (2001); H. C. Cheng and I. Low, J. High Energy Phys. 09 (2003) 051.

[4] Models with extra dimensions also lead to signatures with large $\not\!\!E_T$ at the LHC: N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, Phys. Lett. B **429**, 263 (1998); L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999); T. Appelquist, H.-C. Cheng, and B. A. Dobrescu, Phys. Rev. D **64**, 035002 (2001).

[5] A. Abachi *et al.* (D0 Collaboration), Phys. Rev. Lett. **75**, 618 (1995); B. Abbott *et al.* (D0 Collaboration), Phys. Rev. Lett. **82**, 29 (1999); T. Affolder *et al.* (CDF Collaboration), Phys. Rev. Lett. **85**, 1378 (2000); **88**, 041801 (2002).

[6] G. L. Bayatian *et al.* (CMS Collaboration), Report Nos. CMS TDR 8.2 and CERN/LHCC-2006-021; Report Nos. CMS TDR 8.1 and CERN/LHCC-2006-001; G. Aad *et al.* (ATLAS Collaboration) arXiv:0901.0512.

[7] The W + jets and $t\bar{t}$ + jets processes are studied in the l + jets + $\not\!\!E_T$ signature and treated in the same manner in this paper. In a few places, where differences between the two processes are important, they are discussed separately.

[8] V. Pavlunin and D. Stuart, Phys. Rev. D **78**, 035012 (2008).

[9] For a recent review see J. M. Campbell, J. W. Huston, and W. J. Stirling, Rep. Prog. Phys. **70**, 89 (2007); See also C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, Phys. Rev. D **69**, 094008 (2004); C. F. Berger *et al.*, Phys. Rev. Lett. **102**, 222001 (2009).

[10] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Eur. Phys. J. C **14**, 133 (2000).

[11] S. Chatrchyan *et al.* (CMS Collaboration) JINST **3**, S08004 (2008); G. Aad *et al.* (ATLAS Collaboration), JINST **3**, S08003 (2008).

[12] Pseudorapidity is $\eta = -\ln[\tan(\theta/2)]$, where θ is the particle's polar angle to the beam line.

[13] The start-up $\sqrt{s}$ of pp collisions at LHC is expected to be lower now. For consistency with the previous study [8], I use the same ALPGEN samples for $\sqrt{s}$ of 14 TeV that were used in Ref. [8]. The method has also been tested for lower center-of-mass energies.

[14] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, J. High Energy Phys. 07 (2003) 001; S. Mrenna and P. Richardson, J. High Energy Phys. 05 (2004) 040.

[15] T. Sjöstrand, S. Mrenna, and P. Skands, J. High Energy Phys. 05 (2006) 026.

[16] A. H. Chamseddine, R. Arnowitt, and P. Nath, Phys. Rev. Lett. **49**, 970 (1982).

[17] The minimal supergravity is a restricted model of supersymmetry characterized by only five free parameters defined at the grand unification scale: m_0, $m_{1/2}$, A_0, $\tan\beta$, sign(μ). For LM4 (LM1) benchmarks they are set to 210 (60) GeV, 285 (250) GeV, 0 (0), 10 (10), + (+). The total LM4 cross section is 25.1 pb (NLO), with $\sigma[Z(l^+l^-)$ + jets$] \sim 0.6$ pb. The total LM1 cross section is 54.9 pb (NLO); about a third of it includes one or more leptons in the final state. For more information see G. L. Bayatian *et al.* (CMS Collaboration), Report Nos. CMS TDR 8.2 and CERN/LHCC-2006-021.

[18] Energy depositions in the calorimeters from electrons [$Z(ee)$ + jets and e + jets + $\not\!\!E_T$] and photons (γ + jets) are reconstructed as jets. Such jets are well measured since most of their energy is collected by the electromagnetic calorimeter. In order to avoid double counting the energy depositions from electrons and photons in $\not\!\!E_T$, only jets are used in $\not\!\!E_T$ for final states with electrons or photons. In J_T and N_J measurements, jets from electrons and photons are not used.

[19] This is a simple model of jet losses that allows to study how the method performs if there are other sources of jet inefficiency, for example, more uniform in η and ϕ.

[20] For a recent review of jet reconstruction and calibration see S. D. Ellis *et al.*, Prog. Part. Nucl. Phys. **60**, 484 (2008).

[21] The contribution from b jets to the bias in artificial $\not\!\!E_T$ prediction is smaller than that due to differences in N_J and J_T spectra between $t\bar{t}$ + jets and QCD. Resolution effects specific to higher track multiplicity b jets are not simulated in this study.

[22] The effect of W polarization in W + jets on the charged lepton and neutrino $|\vec{p}_T|$ spectra are more pronounced in the forward regions at high $|\eta_{\text{charged lepton}}|$, especially outside the fiducial η coverage for charged leptons. Therefore, they can be enhanced or reduced by varying the $\eta_{\text{charged lepton}}$ selection.

[23] G. L. Kane, G. A. Ladinsky, and C. P. Yuan, Phys. Rev. D **45**, 124 (1992); M. Fischer, S. Groote, J. G. Korner, and M. C. Mauser, Phys. Rev. D **63**, 031501(R) (2001); V. M.

Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **100**, 062004 (2008); T. Aaltonen *et al.* (CDF Collaboration), Phys. Lett. B **674**, 160 (2009).

[24] A. Abulencia *et al.* (CDF Collaboration), J. Phys. G **34**, 2457 (2007); D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. **94**, 091803 (2005).

[25] More generally, in $l + \text{jets} + \not{E}_T$, one could develop a search for new physics by studying the consistency of the charged lepton $|\vec{p}_T|$ spectrum with that of the neutrino.