CMS Physics Analysis Summary

Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector

The CMS Collaboration

Abstract

The performance of missing transverse momentum reconstruction algorithms is presented using 13 TeV pp collision data collected with the CMS detector in 2016, corresponding to an integrated luminosity of 12.9 fb^{-1}. Events with anomalous missing transverse momentum are studied, and the performance of algorithms used to identify and remove those events is presented. The scale and resolution for missing transverse momentum and the performance of advanced missing transverse momentum algorithms are measured using events with an identified Z boson or isolated photon.

1 Introduction

The presence of neutrinos and hypothetical neutral weakly interacting particles produced in the proton-proton (pp) collisions provided by the LHC at CERN can only be inferred through the momentum imbalance in the plane perpendicular to the beam direction since such particles do not leave a signal in the CMS detector. A precise measurement of all detectable and reconstructed physics objects (e, μ, τ, γ, jets, etc) can be used to determine the missing transverse momentum ($\vec{E}_\mathrm{T}^\mathrm{miss}$). The magnitude of $\vec{E}_\mathrm{T}^\mathrm{miss}$ is referred to as the missing transverse energy and is denoted by $E_\mathrm{T}^\mathrm{miss}$.

The $E_\mathrm{T}^\mathrm{miss}$ is one of the most important observables in many searches for physics beyond the standard model, such as supersymmetry and extra dimensions, as well as for collider-based dark matter searches. It also played a crucial role in the discovery of the Higgs boson, in particular in analysis with neutrinos in the final state, such as WW. In addition, the precise measurement of $E_\mathrm{T}^\mathrm{miss}$ is critical for measurements of standard model physics involving W bosons and top quarks.

In addition to imperfect resolution of all detectable and reconstructed physics objects, the $E_\mathrm{T}^\mathrm{miss}$ measurement is also sensitive to overlapping detector signals from additional interactions in the previous, the same, or subsequent bunch crossings (commonly referred to as in-time and out-of-time pileup), particle misidentification, as well as detector malfunctions. A detailed understanding of these effects is important for maintaining an optimal $E_\mathrm{T}^\mathrm{miss}$ performance. It is, therefore, essential to study $E_\mathrm{T}^\mathrm{miss}$ reconstruction in detail with data.

This document describes the $E_\mathrm{T}^\mathrm{miss}$ reconstruction algorithms and associated corrections, together with performance studies conducted in 13 TeV pp data with 25 ns bunch-spacing, recorded with the CMS detector in 2016. Previous studies of the missing transverse energy reconstruction in 7 and 8 TeV data were presented in [1, 2].

In this document, a brief description of the CMS detector is presented in Section 2. A description of the physics object reconstruction and the event selection along with the data and Monte Carlo (MC) simulation samples used is provided in Section 3. In Section 4, the reconstruction and associated correction of $E_\mathrm{T}^\mathrm{miss}$ are shown. In Section 5, sources of anomalous $E_\mathrm{T}^\mathrm{miss}$ measurements from known detector induced backgrounds, artifacts and methods for identifying them are described. In Section 6, the $E_\mathrm{T}^\mathrm{miss}$ scale and resolution studies are shown based on the measurements made with event samples containing isolated photon or Z boson candidates. Section 7 reports the performance of $E_\mathrm{T}^\mathrm{miss}$ reconstruction algorithms developed to mitigate the effects of a large number of pileup interactions. The algorithm that provides an estimate of the $E_\mathrm{T}^\mathrm{miss}$ significance is described and its performance presented in Section 8. A summary is given in Section 9.

2 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the superconducting solenoid volume are a silicon pixel and strip tracker, a lead tungsten crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity [3] coverage provided by the barrel and endcap detectors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid.

In the barrel section of the ECAL, an energy resolution of about 1% is achieved for unconverted

or late-converting photons in the tens of GeV energy range. The remaining barrel photons have a resolution of about 1.3% up to a pseudorapidity of $|\eta| = 1$, rising to about 2.5% at $|\eta| = 1.4$. In the endcaps, the resolution of unconverted or late-converting photons is about 2.5%, while the remaining endcap photons have a resolution between 3 and 4% [4]. When combining information from the entire detector, the jet energy resolution amounts typically to 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV, to be compared to about 40%, 12%, and 5% obtained when the ECAL and HCAL calorimeters alone are used.

In the region $|\eta| < 1.74$, the HCAL cells have widths of 0.087 in pseudorapidity and 0.087 in azimuth (ϕ). In the η-ϕ plane, and for $|\eta| < 1.48$, the HCAL cells map on to 5×5 ECAL crystals arrays to form calorimeter towers projecting radially outwards from close to the nominal inter-action point. At larger values of $|\eta|$, the size of the towers increases and the matching ECAL arrays contain fewer crystals. Within each tower, the energy deposits in ECAL and HCAL cells are summed to define the calorimeter tower energies, subsequently used to provide the energies and directions of hadronic jets.

Muons are measured in the pseudorapidity range $|\eta| < 2.4$, with detection planes made using three technologies: drift tubes, cathode strip chambers, and resistive plate chambers. Matching muons to tracks measured in the silicon tracker results in a relative transverse momentum resolution for muons with $20 < p_T < 100$ GeV of 1.3–2.0% in the barrel and better than 6% in the endcaps, The p_T resolution in the barrel is better than 10% for muons with p_T up to 1 TeV [5].

The silicon tracker measures charged particles within the pseudorapidity range $|\eta| < 2.5$. It consists of 1440 silicon pixel and 15 148 silicon strip detector modules and is located in the 3.8 T field of the superconducting solenoid. For non isolated particles of $1 < p_T < 10$ GeV and $|\eta| < 1.4$, the track resolutions are typically 1.5% in p_T and 25–90 (45–150) μm in the transverse (longitudinal) impact parameter [6]

The first level (L1) of the CMS trigger system, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select the most interesting events in a fixed time interval of less than 4 μs. The high-level trigger (HLT) processor farm further decreases the event rate from around 100 kHz to less than 1 kHz, before data storage.

A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [3].

3 Data sets and event selection

The reported studies are performed using the data set recorded with the CMS detector in 2016 corresponding to an integrated luminosity of 12.9 fb^{-1} of proton-proton collisions with 25 ns bunch-spacing at $\sqrt{s} = 13$ TeV.

Event reconstruction at the CMS experiment is performed using a particle-flow (PF) algorithm [7, 8] which reconstructs and identifies individual particles using an optimized combination of all sub-detector information. Photon candidates are identified as energy depositions in the ECAL called superclusters. The energy of charged hadrons is estimated using track momentum and the corresponding ECAL and HCAL energy depositions. These energies are calibrated for the nonlinear response of the calorimeters and zero suppression. The energy of neutral hadrons is obtained from the corresponding calibrated ECAL and HCAL energy.

Jets used in the studies are reconstructed from the clustering of the PF candidates, using the anti-k_T algorithm [9]. The anti-k_T jet clustering is performed with a distance parameter of 0.4.

The jet energy is corrected to account for the impact of additional proton-proton interactions on the basis of the event energy density scaled proportionally to the jet area. Jets are further corrected to account for non-uniform and non-linear detector response. Residual data vs. simulation calibrations based on control samples are then applied to correct for data vs. simulation differences in the jet energy scale [10].

Candidate photons used in the reported studies are reconstructed within the geometrical acceptance defined by $|\eta| < 1.4442$ and for transverse momentum $p_T > 50\,\text{GeV}$, with a cluster reconstruction algorithm based on energy deposited in the ECAL. This cluster reconstruction allows for almost complete recovery of the energy of the photons that convert due to the material in front of the ECAL. For the selected photon candidate identification criteria based on the cluster width, shower shape and isolation based on PF candidates are applied. Lastly to distinguish photons from electrons, the candidate photons are required to not have any associated track seeds in the pixel detector.

Candidate electrons used in the reported studies are reconstructed within the geometrical acceptance defined by $|\eta| < 2.5$ and for a transverse momentum $p_T > 20$ GeV, with an algorithm that combines information from the ECAL and the tracker. Energy clusters in the ECAL are built with a topological algorithm that allows for energy sharing between the crystals. The measurement of the electron momentum relies on a multivariate regression that combines a regression-based supercluster energy measurement in the ECAL and the momentum estimate from the tracker [11]. For the selected electron candidate, identification criteria based on shower shapes, and isolation based on PF candidates are applied. Finally, to reject electrons produced by photon conversion, the electrons associated to a track with a missing hit in any of the innermost three layers of the pixel tracker are rejected.

Candidate muons used in the reported studies are reconstructed within the geometrical acceptance $|\eta| < 2.4$ and for $p_T > 20$ GeV, by combining information from the silicon tracker and the muon system. The matching between the inner and outer tracks proceeds either outside-in, starting from a track in the muon system, or inside-out, starting from a track in the silicon tracker. For the muon candidates, the pileup corrected isolation sum is required to be smaller than 15% of the muon transverse momentum.

Simulated standard model (SM) background samples are produced for the Z, W, WZZ, WWZ, and $t\bar{t}$ using MADGRAPH5_AMC@NLO [12] interfaced with PYTHIA 8.2 [13] and using the CUETP8M1 tune for hadronization and fragmentation, where jets from the matrix element calculations are matched to the parton shower following the FxFx matching prescription [14] for next-to-leading-order (NLO) processes and the MLM prescription for leading-order (LO) processes [15]. The QCD and γ + jets samples are produced using MADGRAPH in LO mode interfaced with PYTHIA 8.2 using the CUETP8M1 tune for hadronization and fragmentation. The single top-quark background sample is produced with POWHEG V2 [16], and a set of diboson samples are produced with PYTHIA 8.2. The NNPDF 3.0 [17] parton distribution functions (PDF) are used in all generated samples. Interactions of all final-state particles with the CMS detector are simulated with GEANT4 [18]. All Monte Carlo samples are reweighted to be consistent with the reconstructed primary vertex distribution as measured in the data. This is demonstrated in Fig 1, which shows agreement in the reconstructed vertex multiplicity for $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ candidate events respectively.

A dilepton sample corresponding to final states containing at least two leptons is used for the E_T^{miss} scale and resolution measurement presented in Section 6. The $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ candidate events were collected using a combination of dilepton trigger with thresholds of 17/8 GeV for muons and 23/12 GeV for electrons and single lepton triggers with higher thresholds
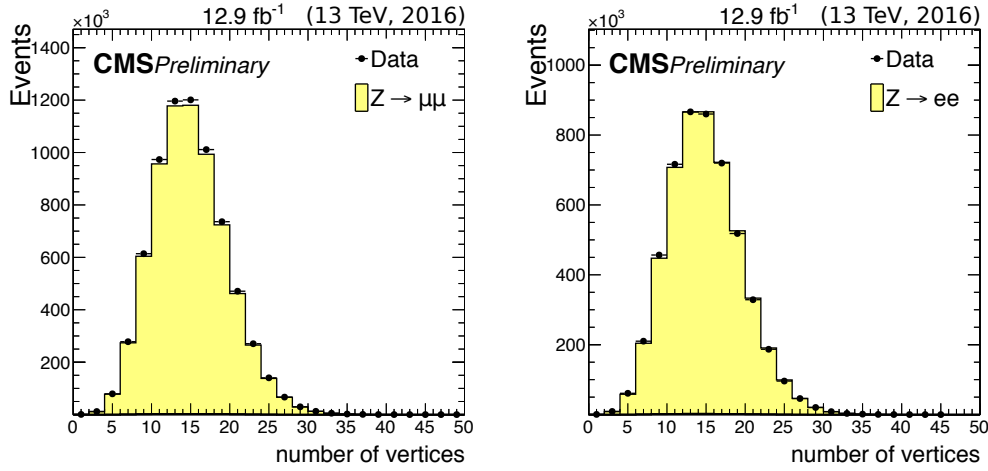
Figure 1: Multiplicity of reconstructed vertices for $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ candidate events.

of 24 GeV and 27 GeV for muons and electrons respectively. Events with an invariant mass of the dimuon or dielectron system outside of the Z-boson mass (M_{ll}) window 81 GeV $< M_{ll} <$ 101 GeV are rejected. The $t\bar{t}$ and single-top (top) processes as well as dibosons (VV) processes are the dominant backgrounds in both the $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ samples and are referred to as EWK in the figures.

A direct-photon sample corresponding to final states containing at least one photon and at least one jet is used for the measurements of $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution. Photon events were collected with a set of triggers based on the measured p_{T} of the hardest reconstructed photon candidate in the event. The p_{T} thresholds of the triggers range from 30 GeV to 165 GeV. Offline selected photons are required to have $p_{\mathrm{T}} > 50$ GeV passing the selection criteria based on cluster shape and isolation. Events with leptons are rejected. The background processes that are considered for the direct photon analysis are QCD multijet events as well as production of single photons produced in association with the Wor Z boson (refereed to as EWK).

The boson transverse momentum spectra are shown in Fig. 2 for Z and isolated γ candidate events.

The dijet data sample is used in the studies of anomalous high-$E_{\mathrm{T}}^{\mathrm{miss}}$ events presented in Section 5 was collected with a hadronic energy sum trigger, which requires total hadronic energy in the event to be larger than 800 GeV. Dijet events are selected by requiring a leading jet with $p_{\mathrm{T}} > 400$ GeV, and events are required to have at least another jet with $p_{\mathrm{T}} > 200$ GeV. Events with an identified photon, electron and muon are rejected.

4   The $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction

The $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ can be defined as the imbalance in the transverse momentum of all particles that interact with the detectors. Because of the momentum conservation, $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ correspond to the transverse momentum that is carried by weakly interacting particles, such as neutrinos. The $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is computed as the negative of the vectorial sum of transverse momenta of all PF particles and is also referred to as PF $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$. This quantity is used in the majority of CMS analyses because of its high performance. Its peformance is discussed in Section 6 in detail.

Minimum energy thresholds in the calorimeters, inefficiencies in the tracker, nonlinearity of the response of the calorimeter for hadronic particles could lead to overestimated or underes-
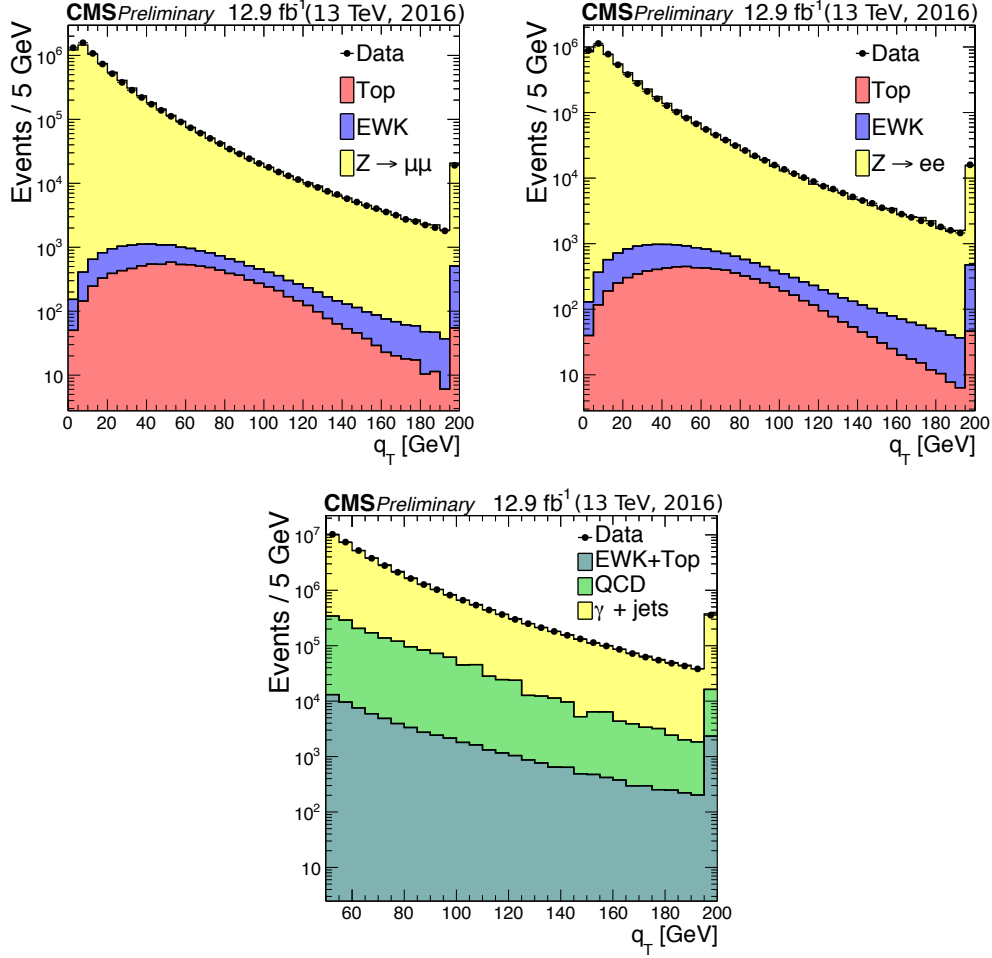
Figure 2: Distributions of Z/γ transverse momentum in $Z \to \mu^+\mu^-$ (left), $Z \to e^+e^-$ (right), and isolated photon events (bottom). The last bin contains overflow content. The top contribution corresponds to the top pair and single top production processes. The EWK contribution corresponds to the diboson, $Z\gamma$ and $W\gamma$ production processes.

timated values of E_T^{miss}. The bias on the E_T^{miss} measurement is reduced by correcting the p_T of the jets to the particle level jet p_T using the jet energy corrections [10] and propagating the correction into the $\vec{E}_T^{\mathrm{miss}}$ according to

$$\vec{E}_T^{\mathrm{misscorr}} = \vec{E}_T^{\mathrm{miss}} - \sum_{\mathrm{jets}} (\vec{p}_{\mathrm{T,jet}}^{\,\mathrm{corr}} - \vec{p}_{\mathrm{T,jet}}), \qquad (1)$$

where the superscript "corr" refers to the corrected values.

This "type-I" correction for $\vec{E}_T^{\mathrm{miss}}$ uses jet energy scale corrections for all corrected jets with $p_T > 15\,\mathrm{GeV}$ that have less than 0.9 of their energy deposited in ECAL. Furthermore, if a muon is found in a jet, its 4-momentum is subtracted from the 4-momentum of the jet when performing the correction and is added back to a corrected object.

A similar correction strategy has been also adopted by the HLT level E_T^{miss}. At the HLT, $\vec{E}_T^{\mathrm{miss}}$ is calculated as the negative of the vectorial sum over transverse momenta of all jets while the "type-I" correction for E_T^{miss} uses jet energy scale corrections for all jets with $p_T > 35\,\mathrm{GeV}$ that have less than 0.9 of electromagnetic energy fraction. Similarly to the offline reconstruction case, if a muon is found in a jet, its 4-momentum is subtracted from the 4-momentum of the

jet when performing the correction and is added back to a corrected object. The improved efficiency curves of corrected E_T^{miss} can be seen in Fig. 3. Because of the improved efficiency of the triggers, the online thresholds can be moved to higher values yielding up to 10% reduction in rate for the same offline performance.
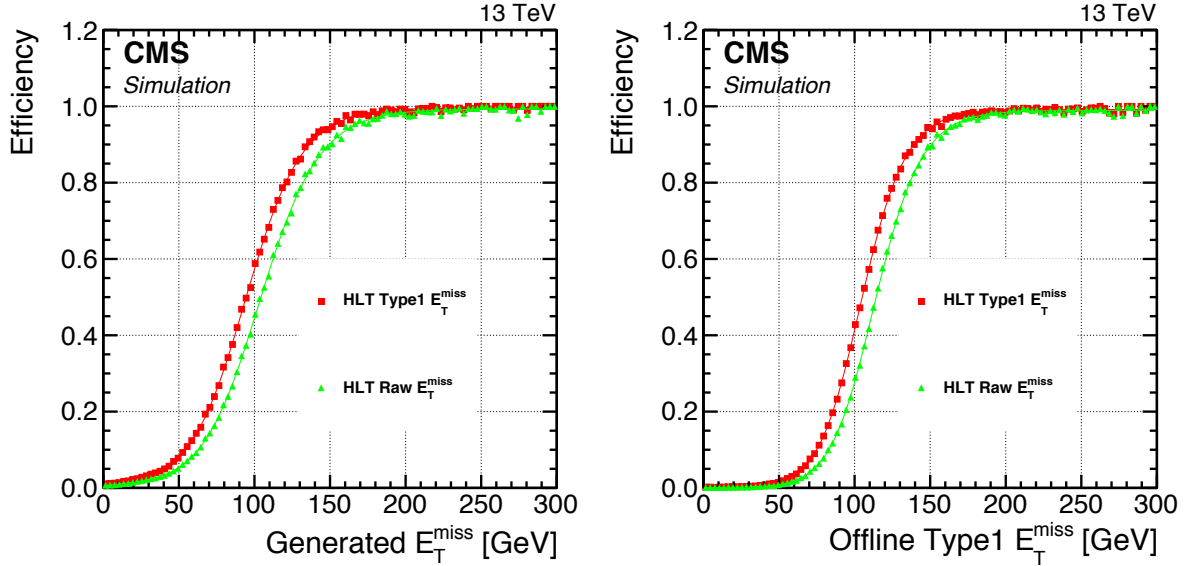

Figure 3: The HLT uncorrected and jet energy scale corrected E_T^{miss} trigger efficiencies as a function of generated E_T^{miss} (left) and the offline corrected E_T^{miss} (right).

The uncertainty of E_T^{miss} has strong dependence on the exact topology of the final state of a given study. Events where the transverse momenta of the jets in the final state are small compared to the p_T of a lepton (often true, for example, for leptonic W decays) will have very different uncertainties than those dominated by several high p_T jets. To estimate the uncertainty on the E_T^{miss}, one can factorize it into its components of jets, electrons, muons, taus, photons and jets of $p_T < 10\,\text{GeV}$ and PF candidates not within jets (defined as the unclustered energy) and vary each object within its scale and resolution uncertainties. For the unclustered energy uncertainty the low energy PF particles are further divided into groups to account for the resolution effect of different detector components: charged PF hadrons (dominated by the tracker resolution), neutral PF hadrons (dominated by the HCAL resolution), photon PF candidates (dominated by the ECAL resolution) and PF candidates associated with the forward hadronic (HF) calorimeter (dominated by the forward HCAL resolution). The largest contributions to the unclustered energy uncertainty are due to the neutral PF hadrons and HF PF candidates. The magnitude of the uncertainties due to jet energy scale correction and unclustered energy scale variations are found to be 5% and 30% respectively in an inclusive $t\bar{t}$ simulation.

The reported studies in this document uses "type-I" corrected E_T^{miss} and the reported uncertainties are computed by varying jet energy scale correction and unclustered energy scale per event.

5 Anomalous E_T^{miss}

Anomalous high-E_T^{miss} events can be due to various phenomena. In the ECAL, spurious deposits may appear due to particles striking sensors in the ECAL photodetectors, or from real showers with non-collision origins such as those caused by beam halo particles. ECAL dead cells can cause real energy to have been missed, again leading to a spurious imbalance. In the

HCAL, spurious energy can arise due to noise in the hybrid photodiode (HPD) and readout box (RBX) electronics, as well as direct particle interactions with the light guides and photo-multiplier tubes of the forward calorimeter. These sources have been studied extensively in the data collected in the 2010 to 2012 LHC running period [1, 2]. Details of the various algorithms (filters) developed to identify and suppress such anomalous high E_T^{miss} events are given below.

The geometrical patterns of hybrid photodiode or readout box channels as well as the pulse shape and timing information are utilized by various HCAL barrel and endcap (HBHE) algorithms to identify and eliminate the noise signals originating from HCAL. These noise algorithms operate both in noise-cleaning and noise-rejecting modes such that the noisy energy deposits are removed from the event reconstruction in the cleaning mode, and the entire event is removed from the physics data sets in the rejecting mode. The isolation-based noise filter utilizes a topological algorithm where energy deposits in HCAL and ECAL as well as the tracking measurements are combined and compared in order to identify isolated instances of noise activity in HBHE. The negative energy filter (NEF) is a pulse-shape filter that acts at the cluster reconstruction level targeting spike-like noise signals in HBHE HPDs and is used in the noise-cleaning mode. It relies on the known pulse-shapes of HPDs (similar to the RBX pulse-shape filters), but explicitly allows for the presence of in-time and out-of-time pileup when testing for anomalous pulse-shapes.

In the ECAL, much of the electronics noise and spurious signals from particle interactions with the photodetectors is removed at the reconstruction based on ECAL information only (local reconstruction) with topological and timing based selection. The remaining effects that lead to high-E_T^{miss} signatures such as anomalously high energy supercrystals (cluster of clusters extended in ϕ to recover the spread energy), and the lack of information for channels due to not properly functioning precision readout are removed through dedicated event filters.

In 2016, five ECAL endcap (EE) supercrystals have been identified and masked which give anomalously high energies (TeV regime) due to producing high amplitude anomalous pulses, in several channels at once. Furthermore in a limited number of ECAL towers, the crystal by crystal precision readout is not available. For the cases where the precision readout is not available, the trigger primitive (TP) information is available, and can be used to estimate the magnitude of the visible energy "lost". However, the TP information saturates above 127.5 GeV. If the TP value is found to be close to the saturation value for the "lost" channels, the event is filtered out.

Machine-induced backgrounds, especially the production of muons when beam protons undergo collisions upstream of the detector also called as beam halo, can cause high-E_T^{miss} signatures that are unrelated to the hard scattering in the collision event. These beam halo particles will travel nearly parallel to the collision axis and will leave a calorimeter deposit along a line at constant ϕ in the calorimeter. In addition, interactions in the cathode strip chambers (CSCs), a subdetector with good reconstruction performance for both collision and non-collision muons, will often line up with the deposit. As a result, filters that are used to identify events contaminated by the beam halo muons rely on various information related to the geometric quantities, energy deposits, and timing signatures. The filters used in 2016 are found to reject 85% in a halo-enriched sample, whereas per-event mistag probability determined from a simulation of inclusive pp interactions for the filter is found to be $< 0.01\%$. An example event display for a beam halo event is shown in Fig. 4 where collinear hits in the CSCs are visible.

Another source of anomalous high-E_T^{miss} comes from the low-quality muon during the muon-tracking iteration which leads to a high-p_T track that survives while the muon is not considered good by the PF algorithm. These low-quality tracks however are still considered charged
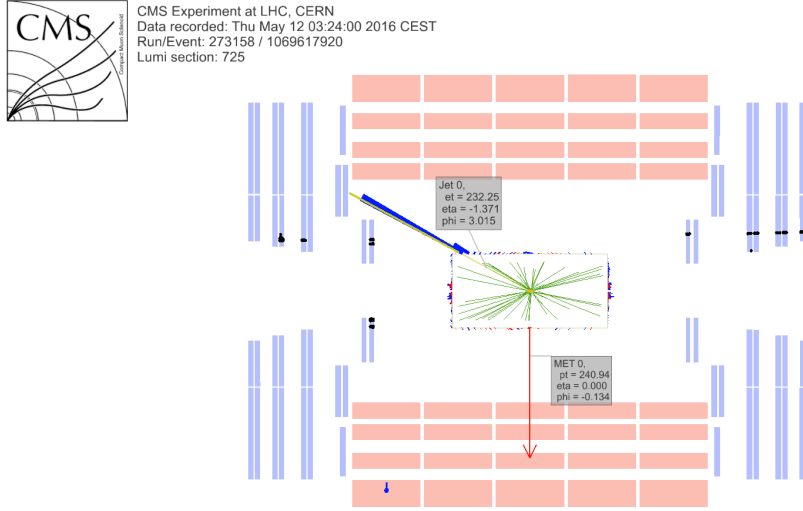
Figure 4: Event display for a beam halo event with collinear hits in the CSC (black), E_T^{miss} of 250 GeV and a jet of 232 GeV. The hadronic deposit is spread in η but it is narrow in ϕ.

hadrons therefore is used in the E_T^{miss} computation, creating a large anomalous E_T^{miss}. The affected events are removed based on the purity of the reconstructed tracks and the relative p_T error of the muons.

The effect of the above mentioned cleaning algorithms is shown in Fig. 5 for events in the dijet topology. The jet identification criterion is presented in Ref. [19] and limits the maximum neutral hadron energy fraction to 0.9. The E_T^{miss} distribution for events passing this criterion and all cleaning algorithms shows a substantial reduction of the high E_T^{miss} tail and agrees well with the simulated distributions.

6   The E_T^{miss} scale and resolution measurement

The performance (scale and resolution) of E_T^{miss} can be studied in events with an identified Z boson or an isolated photon. Momenta of muons originating from Z boson decays are reconstructed with resolutions of 1-6% [5], and electron and photon energies are reconstructed with resolutions of 1-4% [20]. Momenta of jets are reconstructed with less precision, with typical resolutions of 5-15% [10]. As a consequence, the E_T^{miss} resolution in Z or γ + jets events is dominated by the hadronic activity in the event. E_T^{miss} distributions in $Z \to \mu^+\mu^-$, $Z \to e^+e^-$, and photon events are presented in Fig. 6. The data distributions are modeled well by the simulation. The increase in the uncertainty band in Fig. 6 around 60 GeV is due to the large impact of the unclustered energy uncertainties.

The performance of E_T^{miss} is measured through comparing the momenta of the vector boson to the hadronic recoil system. In Figure 7 the vector boson momentum in the transverse plane is shown as $\vec{q}_T$, and transverse momentum of the hadronic recoil, defined as the vectorial sum of the transverse momenta of all particles except the vector boson (or its decay products, in the case of Z bosons), is shown as $\vec{u}_T$. Momentum conservation in the transverse plane dictates that $\vec{q}_T + \vec{u}_T + \vec{\not{E}}_T = 0$.

A well measured Z/γ provides an unique event axis along the unit vector $\hat{q}_T$ and a precise
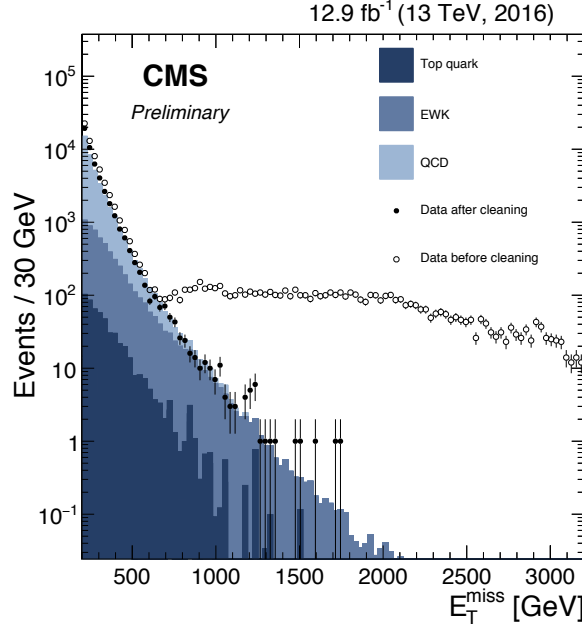
Figure 5: The $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions for events passing the dijet selection with the 2016 cleaning algorithms applied including the one based on jet identification requirements (filled markers), without the 2016 cleaning algorithms applied (open markers), and from simulation (filled histograms). The top quark contribution corresponds to the top pair and single-top production processes. The EWK contribution corresponds to $Z \rightarrow ll$, $Z \rightarrow \nu\nu$, $W \rightarrow l\nu$ and diboson processes.

momentum scale where $q_{\mathrm{T}} \equiv |\vec{q}_{\mathrm{T}}|$. Two signed components of the hadronic recoil, parallel ($u_{\|}$) and perpendicular ($u_{\perp}$), can then be computed by projecting the hadronic recoil on to this axis. The decomposition of the hadronic recoil into $u_{\|}$ and $u_{\perp}$ components is used to evaluate $E_{\mathrm{T}}^{\mathrm{miss}}$ characteristics.

In the following we refer to $-\langle u_{\|}\rangle / \langle q_{\mathrm{T}}\rangle$ as the $\vec{E}_{\mathrm{T}}$ response which is closely related to jet energy scale corrections and jet parton flavour. The response and resolution of $E_{\mathrm{T}}^{\mathrm{miss}}$ are studied with a parametrization of the $u_{\|} + q_{\mathrm{T}}$ and $u_{\perp}$ distributions by a Voigtian function, defined by the convolution of a Breit-Wigner (BW) distribution and a Gaussian (G) distribution:

$$\mathrm{V}(x; \sigma, \gamma) = \int \mathrm{G}(y, \sigma) \mathrm{BW}(x - y, \gamma) dy. \tag{2}$$

This parametrization well describes the observed $u_{\|} + q_{\mathrm{T}}$ and $u_{\perp}$ distributions. Full width at half maximum (FWHM) of the Voigtian form, divided by $2\sqrt{2\ln 2}$, is used to describe the resolution in $u_{\|}$ and $u_{\perp}$, denoted by $\sigma(u_{\|})$ and $\sigma(u_{\perp})$. As with the response, we will examine the resolution as a function of q_{T}.

Distributions of $u_{\|}$ and $u_{\perp}$ in $Z \rightarrow \mu^+ \mu^-$ and $Z \rightarrow e^+ e^-$, and photon events are shown in Fig. 8. The back-to-back nature of events implies that $u_{\|}$ is balanced with q_{T}, thus centering $u_{\|} + q_{\mathrm{T}}$ around zero. The $u_{\perp}$ distribution is found to be symmetric because of the isotropic nature of energy fluctuations due to detector noise and underlying event.

Figure 9 shows the response curves as a function of q_{T}, extracted from data and simulation in $Z \rightarrow \mu^+ \mu^-$, $Z \rightarrow e^+ e^-$, and photon events. Good agreement is observed between the data and simulation in all three different final states. The curves show that the $E_{\mathrm{T}}^{\mathrm{miss}}$ is able to fully recover the hadronic recoil activity corresponding to a Z boson q_{T} of 50 GeV, whereas below that,
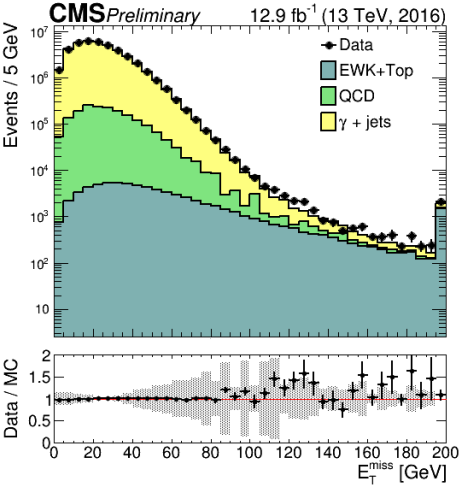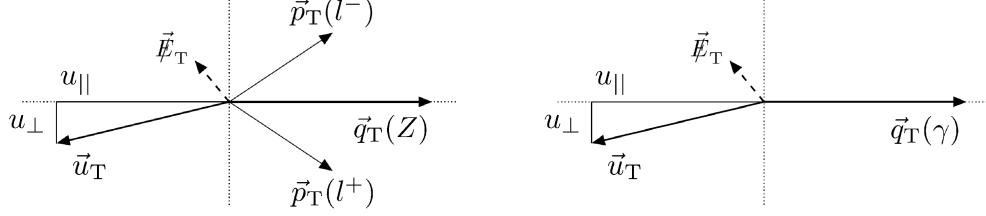
Figure 7: Illustration of $Z \to \ell^+ \ell^-$ (left) and photon (right) event kinematics in the transverse plane. The vector $\vec{u}_T$ denotes the vectorial sum of all particles reconstructed in the event except for the two leptons from the Z decay (left) or the photon (right).

7 PUPPI E_T^{miss}

New techniques have been developed to further reduce the pileup dependence of jets and E_T^{miss}. One of these techniques is pileup per particle identification (PUPPI) [21] and attempts to use local shape information, event pileup properties and tracking information to reduce the pileup dependence on various jet and E_T^{miss} variables.

This method relies on the computation of a local α variable which contrasts the collinear structure of QCD with the soft diffuse radiation coming from pileup. A weight for each particle is computed on an event-by-event basis based on the local shape for a charged pileup. The weights then are used to rescale the 4-momenta of the particles depending on their pileup likeliness. In CMS, a configuration with a different α definition is adopted for particles in the central and forward regions.

In the central region, the shape variable for a given particle i is defined as

$$\alpha_i = \log \sum_{\substack{j \in Ch, PV \\ j \neq i}} \left(\frac{p_{Tj}}{\Delta R_{ij}} \right)^2 \Theta(R_0 - \Delta R_{ij}) \tag{3}$$

where Θ is the step function, i refers to the particle in question and j to the neighboring charged particles from the primary vertex within a cone of radius R_0. The charged particles are considered to be coming from the primary vertex of the event or, if unassociated, to be within $d_z < 0.3$ cm, where d_z is the distance along the z axis with respect to the leading vertex. In the forward region and in the absence of tracking coverage, the summation index j in Eq. 3 extends over all neighbouring particles with a cone of radius R_0.

A χ^2 approximation

$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})^2}{RMS_{PU}^2}, \tag{4}$$

where $\bar{\alpha}_{PU}$ is the median value of the α_i distribution for pileup particles in the event and RMS_{PU} is the corresponding RMS, is used to determine the likeliness of a particle to be from pileup. In the tracker region, $\bar{\alpha}_{PU}$ and RMS_{PU} are calculated using all charged pileup particles (i.e. all charged particles associated to the pileup vertices) while in the forward region they are calculated using all the particles in the event.
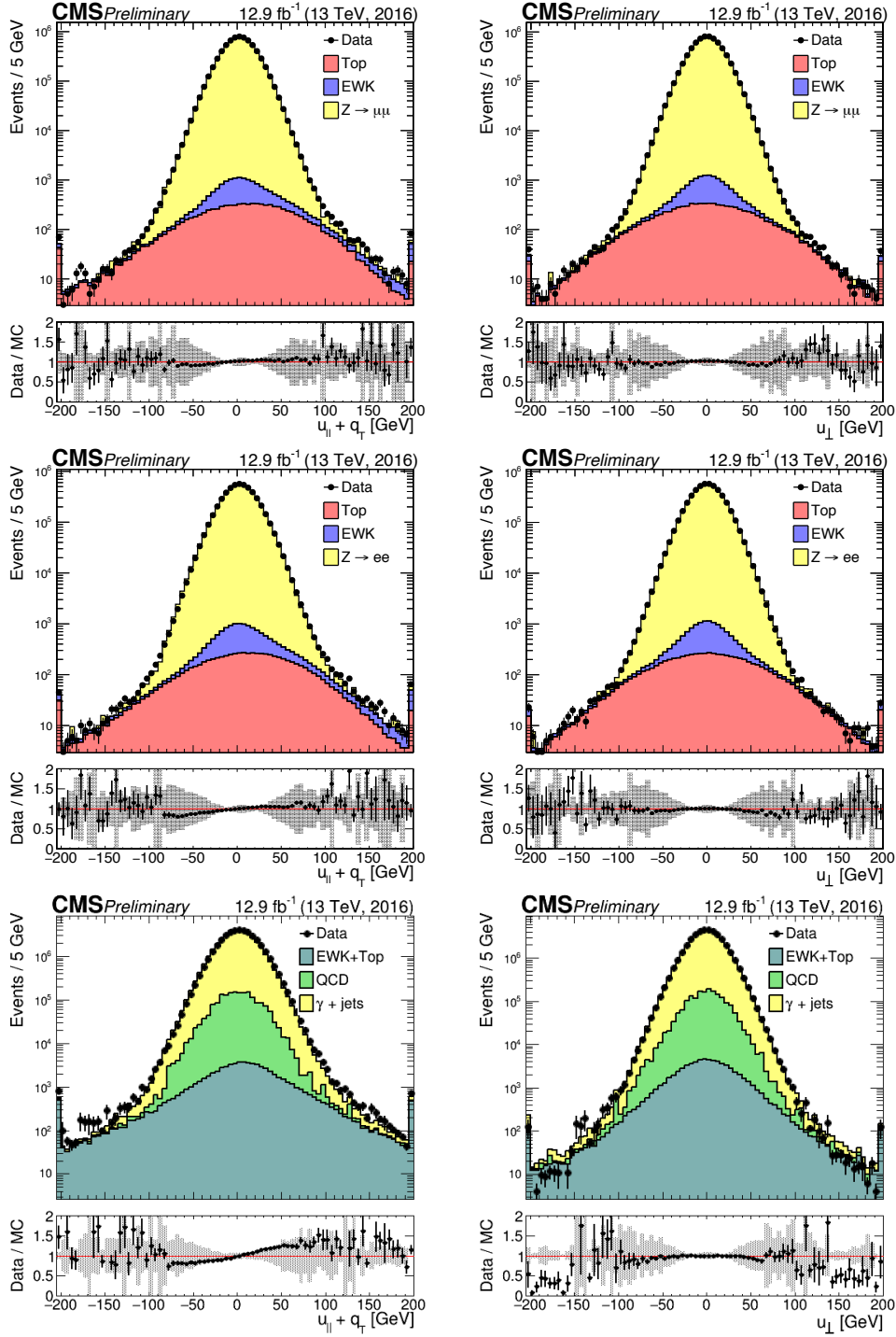
Figure 8: Distributions of the parallel and perpendicular components of the hadronic recoil, in data (filled markers) and simulation for $Z \to \mu^+\mu^-$(top), $Z \to e^+e^-$(middle) and γ (bottom) events. The points in the lower panel of each plot show the data/MC ratio, including the statistical uncertainties of both data and simulation. The systematic uncertainty due to the jet energy corrections and the systematic uncertainty in the unclustered energy is also displayed on the ratio. The last bin contains overflow content. The top contribution corresponds to the top pair and single top production processes. The EWK contribution corresponds to the diboson, $Z\gamma$ and $W\gamma$ production processes.
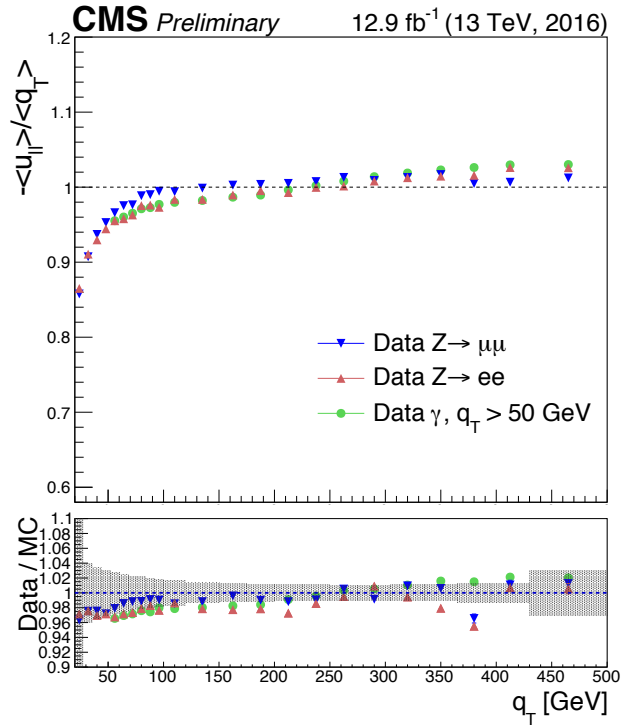
Figure 9: Response is shown for $Z \rightarrow \mu^+\mu^-$ events, $Z \rightarrow e^+e^-$ events and γ events. The upper frame shows the response in data; the lower frame shows the ratio of data to simulation with the error band displaying the systematic uncertainty of the simulation, estimated as the $Z \rightarrow e^+e^-$ channel systematic uncertainty.
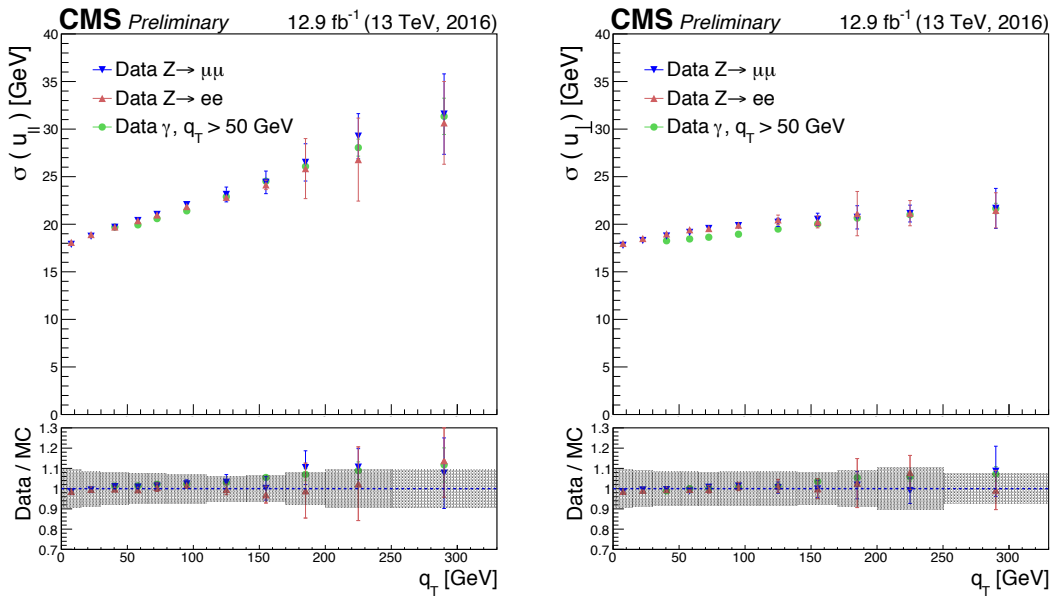

Figure 10: Resolution is shown for $Z \rightarrow \mu^+\mu^-$ events, $Z \rightarrow e^+e^-$ events and γ events. The upper frame shows the resolution in data; the lower frame shows the ratio of data to simulation with the error band displaying the systematic uncertainty of the simulation, estimated as the $Z \rightarrow e^+e^-$ channel systematic uncertainty.

The chosen algorithm parameters are similar to the recommendation from Ref. [21] where the cone radius is set to 0.4. The particle weights (w_i) are required to be larger than 0.01 and the minimum scaled p_T of the neutral particles is required to be $w_i \cdot (p_{Ti} > (A + B \cdot n_{PV}))$. In this equation, A and B are tunable parameters which are tuned as a function of pseudorapidity and n_{PV} is the reconstructed vertex multiplicity. The optimization of the tunable parameters are performed to achieve the best jet p_T and E_T^{miss} resolutions in pseudorapidity regions $0 < |\eta| < 2.5$, $2.5 < |\eta| < 3$, and $|\eta| > 3$. No additional pileup corrections are applied to jets clustered from these weighted inputs.

Once a weight per PF particle is determined, the E_T^{miss} can be computed using the 4-vector weighted by the PUPPI weight. PUPPI E_T^{miss} distributions for $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow e^+e^-$ events are presented in Fig. 11. The data distributions are modeled reasonably well by the simulation.
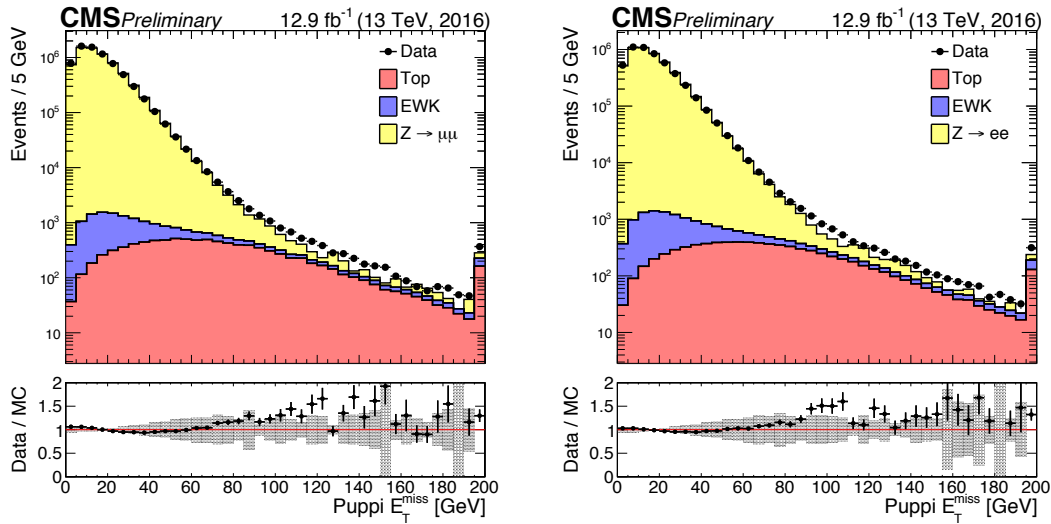

Figure 11: PUPPI E_T^{miss} in $Z \rightarrow \mu^+\mu^-$(left), $Z \rightarrow e^+e^-$(right), and isolated photon events (bottom). The points in the lower panel of each plot show the data/MC ratio, including the statistical and systematic uncertainties due to jet energy corrections and unclustered energy. The last bin contains overflow content. The top contribution corresponds to the top pair and single top production processes. The EWK contribution corresponds to the diboson, $Z\gamma$ and $W\gamma$ production processes.

The performance (resolution) of PUPPI E_T^{miss} then can be studied in events with an identified Z boson or an isolated photon similar to the case of E_T^{miss} as described in Section 6. Distributions of $u_{\parallel}$ and $u_{\perp}$ for $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow e^+e^-$ events are shown in Fig. 12.

Figure 13 shows the response curves as a function of q_T, extracted from data and simulation in $Z \rightarrow \mu^+\mu^-$ events compared to E_T^{miss} response. Difference in response is observed between PUPPI E_T^{miss} and particle flow E_T^{miss} that is on the order of 4% at the low q_T. The difference in response is due to over-removal of the pileup candidates from the event. For the cases for clustered jets, this over-removal is corrected for. However, the unclustered energy is not corrected for this effect. At low q_T the the uncorrected unclustered energy contributions are dominating compared to the corrected energy of the recoiling jets.

Figure 14 shows the response corrected resolution curves as a function of the number of vertices for $u_{\parallel}$ and $u_{\perp}$, extracted from data and simulation in $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow e^+e^-$ events for E_T^{miss} and PUPPI E_T^{miss}. The resolutions measured for PUPPI E_T^{miss} show less degradation in resolution in higher pileup scenarios compared to E_T^{miss}.
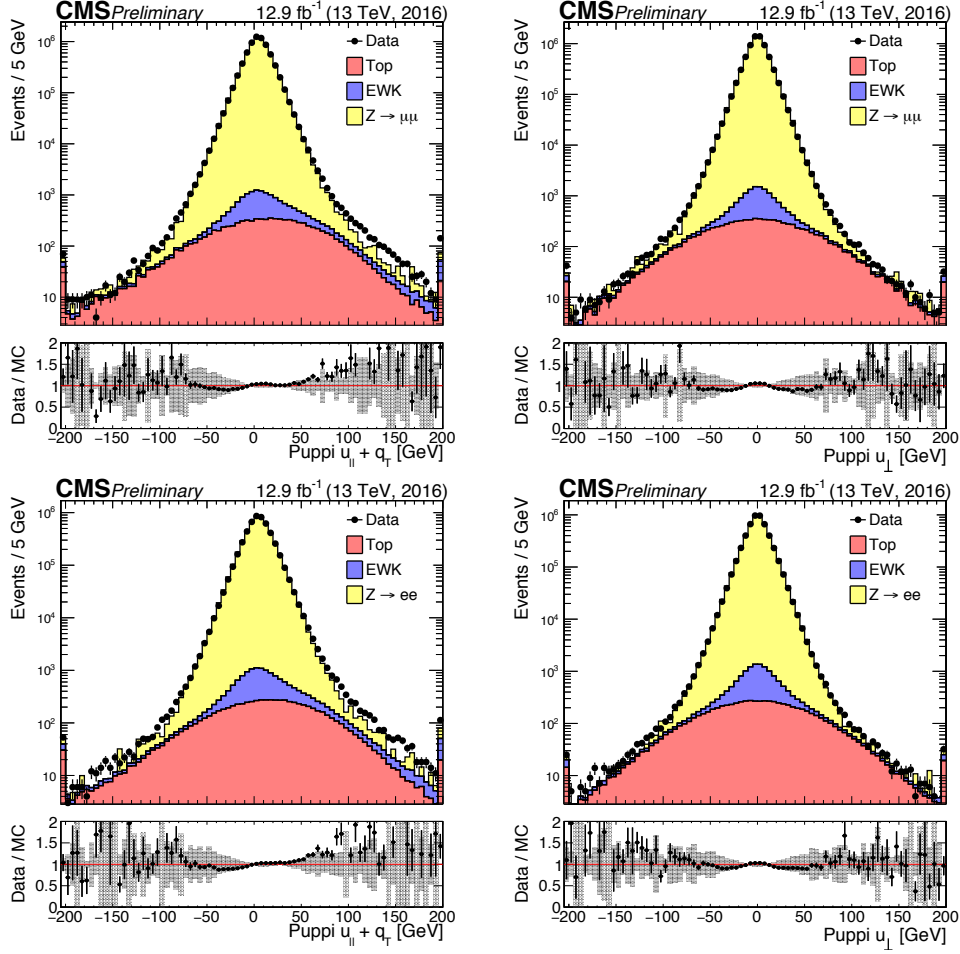
Figure 12: Distributions of the parallel and perpendicular components of the hadronic recoil for PUPPI E_T^{miss}, in data (filled markers) and simulation for $Z \to \mu^+\mu^-$ (top) and $Z \to e^+e^-$ (bottom) events. The points in the lower panel of each plot show the data/MC ratio, including the statistical uncertainties (green band) of both data and simulation. The red band displays the systematic uncertainty due to the jet energy corrections and the blue band displays the systematic uncertainty in the unclustered energy. The last bin contains overflow content. The top contribution corresponds to the top pair and single top production processes. The EWK contribution corresponds to the $Z\gamma$ and $W\gamma$ production processes.

8 The E_T^{miss} significance

In many analysis using E_T^{miss}, it is useful to distinguish between events with genuine E_T^{miss} and those with spurious E_T^{miss}. The E_T^{miss} significance variable, denoted $\mathcal{S}$, helps to make this distinction on an event-by-event basis. The significance variable is computed with the reconstructed E_T^{miss}, as well as the jets, leptons, and low-p_T unclustered hadronic activity in each event. This approach allows to construct a significance variable that applicable to a variety of event topologies at CMS. The variable is described in detail in Refs. [1, 2]. Here we give an overview of updates, tuning, and performance studies conducted using the CMS 13 TeV data set.

The significance is defined as the log-likelihood ratio

$$\mathcal{S} \equiv 2 \ln \left(\frac{\mathcal{L}(\vec{\varepsilon} = \sum \vec{\varepsilon}_i)}{\mathcal{L}(\vec{\varepsilon} = 0)} \right), \tag{5}$$
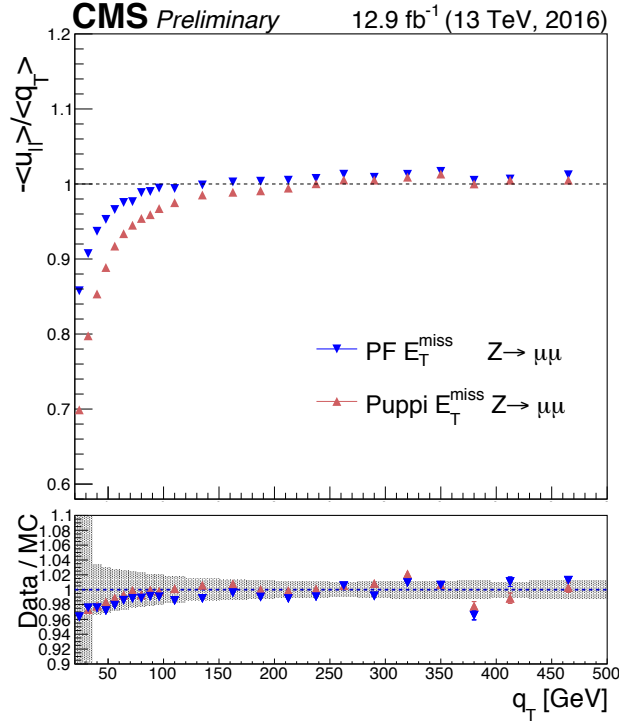
Figure 13: Response is shown for $Z \to \mu^+\mu^-$ events for E_T^{miss} and PUPPI E_T^{miss}. The lower frame shows the ratio of data to simulation with the error band displaying the systematic uncertainty of the simulation
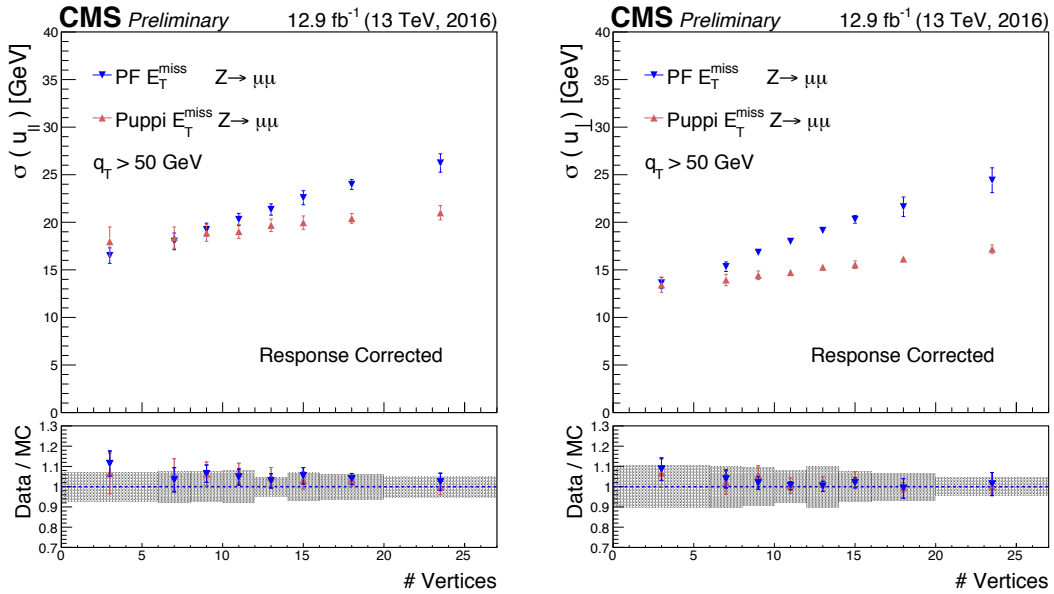

Figure 14: Resolution is shown for $Z \to \mu^+\mu^-$ events. The upper frame shows the resolution in data for E_T^{miss} (red triangle) and PUPPI E_T^{miss} (blue triangle); the lower frame shows the ratio of data to simulation with the error band displaying the systematic uncertainty of the simulation.

where the $\vec{\varepsilon}$ is the *true* E_T^{miss} and $\sum \vec{\varepsilon}_i$ is the *observed* E_T^{miss}. In the numerator, we evaluate the likelihood that the true value of E_T^{miss} equals the observed value, while the denominator corresponds to the *null hypothesis* (that the true E_T^{miss} is zero). In most cases, with a very good

approximation the likelihood $\mathcal{L}(\vec{\varepsilon})$ has the form of a Gaussian distribution. Then, the significance can be written,

$$\mathcal{S} = \left(\sum \vec{\varepsilon}_i \right)^\dagger \mathbf{V}^{-1} \left(\sum \vec{\varepsilon}_i \right), \tag{6}$$

where $\mathbf{V}$ is the 2×2 $E_{\mathrm{T}}^{\mathrm{miss}}$ covariance matrix. In this formulation, $\mathcal{S}$ is conveniently a χ^2 variable with two degrees of freedom (one degree of freedom for each component of $E_{\mathrm{T}}^{\mathrm{miss}}$) for events with zero true $E_{\mathrm{T}}^{\mathrm{miss}}$.

The covariance matrix $\mathbf{V}$ in Eq. 6 models the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution smearing in each event. It is constructed by propagating the individual resolutions of the objects entering the $E_{\mathrm{T}}^{\mathrm{miss}}$ sum. In most cases, the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution captured in covariance $\mathbf{V}$ is primarily determined by the hadronic components of the event. For the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance variable, this includes jets with $p_{\mathrm{T}} > 15$ GeV, and objects with $p_{\mathrm{T}} < 15$ GeV including low-p_{T} jets and unclustered hadronic activity. Jets in the high-p_{T} component enter into the total covariance $\mathbf{V}$ with an individual covariance of the form:

$$\mathbf{U} = \begin{pmatrix} \sigma_{p_{\mathrm{T}}}^2 & 0 \\ 0 & p_{\mathrm{T}}^2 \sigma_\phi^2 \end{pmatrix}. \tag{7}$$

Here, the quantities $\sigma_{p_{\mathrm{T}}}$ and σ_ϕ are measured and then retuned with a combination of simulation and data-based techniques.

All objects momenta falling below the 15 GeV threshold, including low-p_{T} jets and unclustered hadronic activity, are summed vectorially, and the summed momentum is assigned into a single object with $\vec{p}_{\mathrm{T}} = \sum_i \vec{p}_{\mathrm{T}i}$. The resolution of this object is parametrized by the scalar p_{T} sum of its constituents:

$$\sigma_{uc}^2 = \sigma_0^2 + \sigma_s^2 \sum_{i=1}^{n} |\vec{p}_{\mathrm{T}i}|, \tag{8}$$

where the values of σ_0^2 and σ_s^2 are determined using a data driven technique as explained in Ref [2]. The resolution of this object is assumed to be isotropic in the transverse place of the detector.

Electrons and muons are assumed to have perfect resolutions when compared to the hadronic component of the event, and make no contribution to the covariance $\mathbf{V}$.

The performance of the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance variable in the $\mathrm{Z} \to \mu\mu$ and $\mathrm{Z} \to ee$ events which nominally have zero *true* $E_{\mathrm{T}}^{\mathrm{miss}}$ is shown in Fig. 15 for one jet requirement respectively. The core of the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance spectrum follows an ideal χ^2 distribution, but begins to slightly deviate from a perfect χ^2 at large values of $\mathcal{S}$, where the $t\bar{t}$ background becomes significant. In both channels good agreement between data and simulation is observed.

9   Summary

The performance of missing transverse momentum reconstruction algorithms is studied and presented using 13 TeV pp collision data collected with the CMS detector in 2016, corresponding to an integrated luminosity up to 12.9 fb^{-1}. The performance of algorithms used to identify and remove events with anomalous missing transverse momentum have been studied using dijet events. The scale and resolution of the missing transverse momentum determination has been measured using events with an identified Z boson or isolated photon. The measured scale and resolution in data are found to be in agreement with the expectations from the simulation. The performance of an advanced $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithm specifically developed
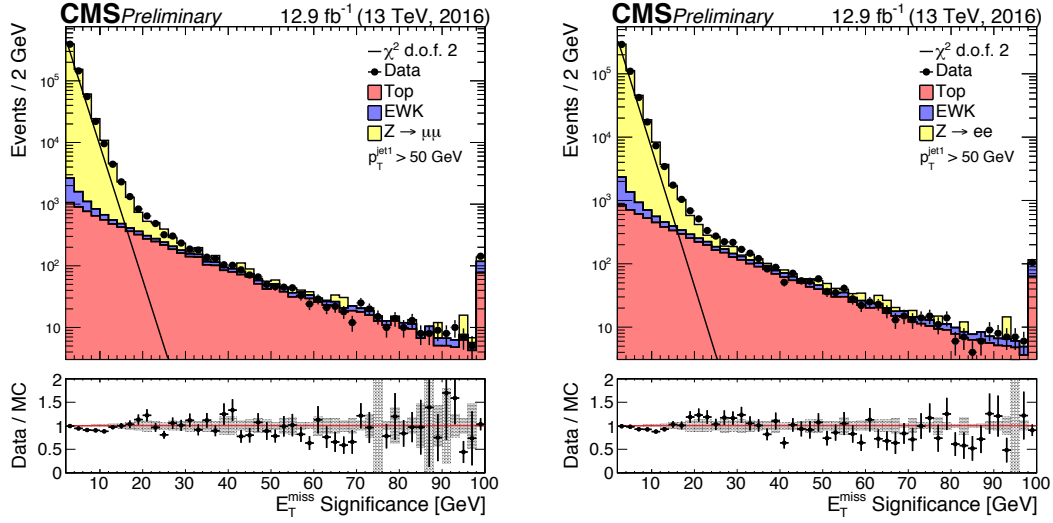
Figure 15: Distribution of E_T^{miss} significance in the $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ channel for one jet selection respectively. The black straight line corresponds to a χ^2 distribution of 2 degrees of freedom. The gray band in the lower pad displays the statistical uncertainty of the simulation.

to cope with a large number of pileup interactions has also been presented. This algorithm shows a significantly reduced dependence of the E_T^{miss} resolution on pileup interactions in $Z \rightarrow \mu^+\mu^-$ events. The studies presented in this note provide a solid foundation for all CMS measurements with E_T^{miss} in the final state, including measurements of the Higgs boson, W boson and top quark and searches for new neutral weakly interacting particles.

References

[1] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011) P09001, `doi:10.1088/1748-0221/6/09/P09001`, `arXiv:1106.5048`.

[2] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 02, P02006, `doi:10.1088/1748-0221/10/02/P02006`, `arXiv:1411.0511`.

[3] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, `doi:10.1088/1748-0221/3/08/S08004`.

[4] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P08010, `doi:10.1088/1748-0221/10/08/P08010`, `arXiv:1502.02702`.

[5] CMS Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, `doi:10.1088/1748-0221/7/10/P10002`, `arXiv:1206.4071`.

[6] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, `doi:10.1088/1748-0221/9/10/P10009`, `arXiv:1405.6569`.

[7] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr, 2009.

[8] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", Technical Report CMS-PAS-PFT-10-001, CERN, 2010, 2015.

[9] 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`.

[10] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", `arXiv:1607.03663`.

[11] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at s = 8 TeV", *JINST* **10** (2015), no. 06, P06005, `doi:10.1088/1748-0221/10/06/P06005`, `arXiv:1502.02701`.

[12] 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`.

[13] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, `doi:10.1016/j.cpc.2015.01.024`, `arXiv:1410.3012`.

[14] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, `doi:10.1007/JHEP12(2012)061`, `arXiv:1209.6215`.

[15] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **01** (2007) 013, `doi:10.1088/1126-6708/2007/01/013`, `arXiv:hep-ph/0611129`.

[16] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **09** (2009) 111, `doi:10.1007/JHEP02(2010)011,10.1088/1126-6708/2009/09/111`, `arXiv:0907.4076`. [Erratum: JHEP02,011(2010)].

[17] R. D. Ball et al., "Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology", *Nucl. Phys.* **B849** (2011) 296–363, `doi:10.1016/j.nuclphysb.2011.03.021`, `arXiv:1101.1300`.

[18] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250, `doi:10.1016/S0168-9002(03)01368-8`.

[19] CMS Collaboration, "Search for dark matter production in association with jets, or hadronically decaying W or Z boson at $\sqrt{s}$=13 TeV", CMS Physics Analysis Summary CMS-PAS-EXO-16-013, CERN, 2016.

[20] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV", *JINST* **10** (2015), no. 08, P08010, `doi:10.1088/1748-0221/10/08/P08010`, `arXiv:1502.02702`.

[21] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup Per Particle Identification", *JHEP* **10** (2014) 059, `doi:10.1007/JHEP10(2014)059`, `arXiv:1407.6013`.