
CMS Physics Analysis Summary

Contact: cms-pog-conveners-jetmet@cern.ch

2013/04/23

Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV with the CMS Detector

The CMS Collaboration

Abstract

The performance of missing transverse momentum reconstruction algorithms is presented using 8 TeV pp collision data collected with the CMS detector, corresponding to an integrated luminosity up to $12.2 \pm 0.5 \text{ fb}^{-1}$. 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, including the effects of multiple pp interactions in the same, earlier, and later bunch crossings (pileup interactions), are measured using events with an identified Z boson or isolated photon. They are in general well described by the simulation. Advanced missing transverse momentum reconstruction algorithms are also developed specifically to mitigate the effects of large numbers of pileup interactions on missing transverse momentum resolution. Using these advanced algorithms, the dependence of the missing transverse momentum resolution on pileup interactions is reduced significantly.

1 Introduction

The Compact Muon Solenoid (CMS) detector [1] can detect most species of stable particles produced in proton-proton collisions delivered by the Large Hadron Collider (LHC). Notable exceptions are neutrinos and hypothetical neutral weakly interacting particles that have yet to be discovered. Although these particles do not leave a signal in the detector, their presence can be inferred from the momentum imbalance in the plane perpendicular to the beam direction, known as missing transverse momentum and denoted by \vec{E}_T . Its magnitude is denoted by E_T and is referred to as missing transverse energy.

Missing transverse momentum plays a critical role in many physics analyses at the LHC. It is a key variable in many searches for physics beyond the standard model, such as supersymmetry and extra dimensions as well as for collider dark matter searches. It also played an important role in studies contributing to the discovery of a new boson at a mass of around 125 GeV, in particular in channels with the WW and $\tau\tau$ final states [2]. In addition, the precise measurement of \vec{E}_T is critical for some measurements of standard model physics, particularly those involving W bosons and top quarks.

The \vec{E}_T reconstruction is sensitive to detector malfunctions and various reconstruction effects resulting in particle momentum mismeasurements and particle misidentification. Precise calibration of all physics objects is crucial for the \vec{E}_T reconstruction, and \vec{E}_T is particularly sensitive to multiple proton-proton interactions in the same, earlier, and later bunch crossings (pileup interactions). Thus, it is essential to study \vec{E}_T reconstruction in detail with data. This note describes \vec{E}_T reconstruction algorithms and corrections developed by the CMS Collaboration and their performance evaluated in 8 TeV proton-proton collision data corresponding to an integrated luminosity up to $12.2 \pm 0.5 \text{ fb}^{-1}$. The average number of interactions in a proton-proton bunch crossing is approximately 20. The studies of the missing transverse momentum reconstruction in 7 TeV data are presented in Ref. [3].

This note is organized as follows. In Section 2, the data and Monte Carlo simulation samples used for these studies, together with the event selection criteria, are described. In Section 3, the different algorithms for reconstructing \vec{E}_T are presented. In Section 4, sources of anomalous \vec{E}_T measurements from known detector artifacts and methods for identifying them are described. In Section 5, the \vec{E}_T scale and resolution are reported based on measurements using events containing Z boson or photon candidates. Studies presented in Section 5 include a detailed evaluation of \vec{E}_T resolution degradation due to pileup interactions. Section 6 reports the performance of advanced \vec{E}_T reconstruction algorithms developed to cope with large numbers of pileup interactions. Conclusions are given in Section 7.

2 Data Samples and Event Selection

The data samples used for studies presented in this note were collected from February through October 2012 in proton-proton collisions at a centre-of-mass energy $\sqrt{s} = 8 \text{ TeV}$, corresponding to an integrated luminosity up to $12.2 \pm 0.5 \text{ fb}^{-1}$. For all studies we require at least one well-identified event vertex whose z position is less than 24 cm away from the nominal centre of the detector and whose transverse distance from the z -axis is less than 2 cm. The vertex with the largest sum of the transverse momentum squared of all tracks associated to that vertex is considered to be the primary vertex that corresponds to the origin of the hard-scattering process.

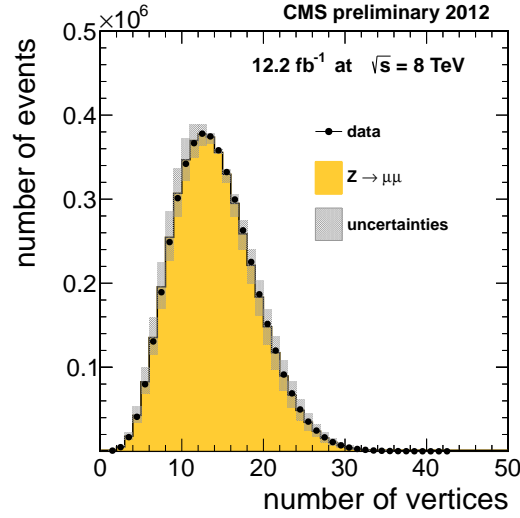


Figure 1: Vertex multiplicity distribution for $Z \rightarrow \mu^+\mu^-$ candidate events in the 2012 data and MC samples, as defined in Section 2.2. The quoted uncertainty is dominated by the uncertainty on the total inelastic pp scattering cross section measurement [10, 11].

CMS uses global event reconstruction, also known as particle-flow (PF) event reconstruction [4, 5], which consists of reconstructing and identifying each particle with an optimized combination of all subdetector information. In this process, particles emerging from all vertices (charged hadrons, neutral hadrons, photons, muons, and electrons) are individually reconstructed. These particles, referred to as PF particles, are then used to reconstruct the \vec{E}_T , jets, and hadronic τ decays, and to quantify the isolation of leptons and photons.

The collision data are compared to samples of simulated events that are generated either using PYTHIA 6 [6], with a parameter set referred to as tune Z2 [7] for the QCD, $\gamma\gamma$, and $\gamma + \text{jets}$ processes, or with MADGRAPH 5 [8] interfaced with PYTHIA 6 for top ($t\bar{t}$ and single-top), $Z + \text{jets}$, $W + \text{jets}$, and diboson processes. The generated events are passed through the CMS detector simulation, which is based on GEANT4 [9]. The detector geometry description includes realistic subsystem conditions such as the simulation of non-functioning channels. The simulated events are reweighted so that the distribution of the number of pileup interactions in the simulation matches that expected for the distribution of instantaneous luminosities recorded for the data. This is demonstrated in Fig. 1, which shows agreement in the vertex multiplicity (N_{vtx}) distribution between data and MC simulation samples. The total uncertainty on the N_{vtx} distribution is dominated by the uncertainty on the total inelastic pp scattering cross section measurement [10, 11] that affects the pileup profile in the simulation sample.

The event selection criteria used to study these data and simulation samples are discussed below.

2.1 Dijet Event Selection

The dijet data sample used in the studies of anomalous high- \vec{E}_T events presented in Section 4 was collected with a single jet trigger, which requires at least one jet in the event with transverse momentum $p_T > 320$ GeV. Some of the anomalous event cleaning algorithms similar to those described in Ref. [3] were applied at the trigger level. The integrated luminosity collected by this trigger corresponds to $11.5 \pm 0.5 \text{ fb}^{-1}$, while all the other data samples used in this note correspond to $12.2 \pm 0.5 \text{ fb}^{-1}$. Offline, hadronic jets are clustered from reconstructed particles

with the infrared- and collinear-safe anti- k_t algorithm, operated with size parameter 0.5. The jet momentum is determined as the vectorial sum of all particle momenta in this jet, and is found in the simulation to be within 2% to 10% of the true momentum over the whole detector acceptance for jets with p_T above 20 GeV. Jet energy corrections are derived from the simulation, and are confirmed with in-situ measurements using the energy balance of dijet and $\gamma + \text{jets}$ events [12].

Dijet events are selected by requiring a leading jet with $p_T > 400$ GeV, which ensures that events are in the region where the trigger is fully efficient, and events are required to have at least another jet with $p_T > 200$ GeV.

2.2 $Z \rightarrow \ell^+\ell^-$ Event Selection

The $Z \rightarrow \ell^+\ell^-$ events, where ℓ is either a muon or an electron, are used in the \vec{E}_T scale and resolution studies presented in Sections 5 and 6. The $Z \rightarrow \mu^+\mu^-$ events were collected using a trigger that requires the presence of two muons with p_T thresholds of 17 and 8 GeV, respectively. We require two muon candidates that pass the following criteria. Muon candidates are required to be reconstructed in the tracker and in the muon chambers, have a reconstructed transverse momentum (p_T^μ) greater than 20 GeV, and be reconstructed in the pseudorapidity range $|\eta| < 2.1$. Muon candidates are required to be isolated in order to reject muons produced by B or D hadron decays. A muon is considered to be isolated if the following condition is satisfied:

$$R_{\text{Iso}}^\ell \equiv \left[\sum_{\text{HS-charged}} p_T + \max \left(\sum_{\text{neutrals}} p_T + \sum_{\text{photons}} p_T - 0.5 \times \sum_{\text{PU-charged}} p_T, 0 \right) \right] / p_T^\mu < 0.1, \quad (1)$$

where $\sum_{\text{HS-charged}} p_T$, $\sum_{\text{neutrals}} p_T$, and $\sum_{\text{photons}} p_T$ are respectively the scalar sums of the transverse momenta of charged particles from the primary hard-scatter (HS) vertex, neutral hadrons, and photons within a distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.4$ in η - ϕ space of the muon. The $0.5 \times \sum_{\text{PU-charged}} p_T$ term is associated with charged hadrons from pileup (PU) vertices and provides a correction that takes into account the additional energy produced by pileup photons and neutral hadrons. Muon reconstruction and identification are described in detail in Ref. [13].

The $Z \rightarrow e^+e^-$ events were collected using a double-electron trigger with p_T thresholds of 17 and 8 GeV. The events are required to have two electron candidates that pass the following selection criteria. Electron candidates are reconstructed by associating a cluster reconstructed in the electromagnetic calorimeter (ECAL) to a track reconstructed by a Gaussian-sum filter (GSF) algorithm [14]. Such electron candidates are required to be in the ECAL fiducial volume defined by $|\eta| < 1.44$ and $1.57 < |\eta| < 2.5$. The electron candidates are required to be isolated based on Eq. (1) but with $\Delta R < 0.3$. Finally, to reject electrons produced by photon conversion, the electrons associated to a GSF track with a missing hit in any of the innermost three layers of the pixel tracker are rejected.

Events with an invariant mass of the dimuon or dielectron system outside of the Z mass window $60 < M < 120$ GeV are rejected. The invariant mass spectrum is presented in Fig. 2 and the distribution of the transverse momentum of the $Z \rightarrow \ell^+\ell^-$ candidate, q_T , is shown in Fig. 3. The data distributions are well modeled by the simulation.

2.3 Direct Photon Event Selection

Photon events were collected with a set of triggers based on the measured p_T of the leading photon candidate in the event. The p_T boundaries of the triggers were: 50, 75, 90, 135, and

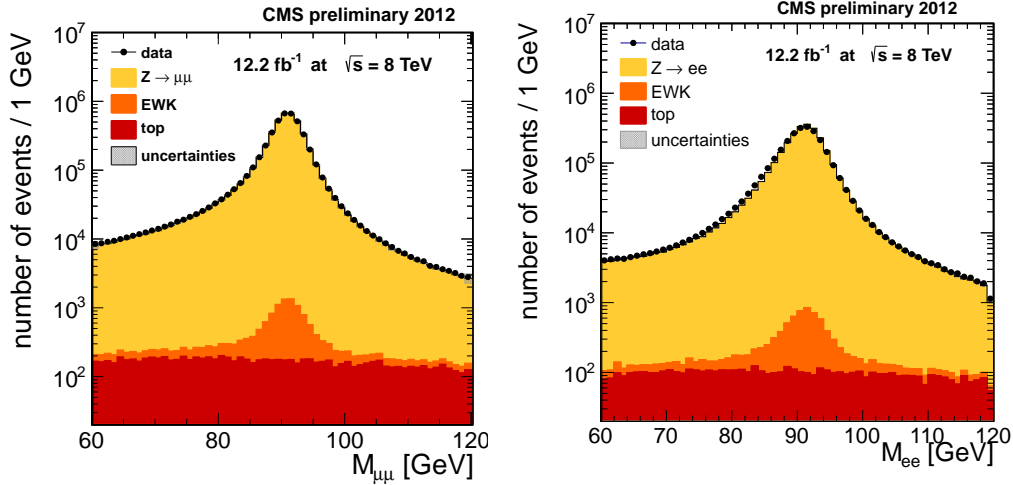


Figure 2: Dilepton invariant mass distributions in events passing the $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) selections.

150 GeV. The rates of the four lower p_T triggers were randomly reduced (prescaled) because of limited data acquisition system bandwidth. The approximate effective values of these prescaling factors are 900, 150, 71, and 1.33, respectively. The photon events are selected offline by requiring each event to contain exactly one reconstructed photon candidate.

Photon candidates are selected from clusters of energy in the ECAL within the pseudorapidity coverage $|\eta| < 1.44$. Various identification criteria are applied in order to correctly identify photons with high efficiency and suppress the misidentification of electrons, jets, or spurious ECAL signals as photons [15]. To suppress the misreconstruction of hadronic jets as photons, photon candidates are required to be isolated. Isolation from other activity in the event is determined through a combination of independent isolation sums for different types of PF particles, corrected for pileup effects and for the contributions from the photon itself, in a cone of $\Delta R = 0.3$ around the photon trajectory. In addition, to prevent the misidentification of electrons as photons, the photon candidate clusters must not match any track reconstructed in the pixel detector that is associated with the primary vertex.

The background processes that are considered are QCD multijet events, diphoton production, production of single W bosons, and single photons produced in association with a Z or a W boson. Although the majority of QCD multijet events fail the photon selection, QCD multijet events constitute a dominant background due to their overwhelming production cross section. QCD jets that pass the photon selection are typically enriched in $\pi^0 \rightarrow \gamma\gamma$ and contain little hadronic activity; therefore, the detector response to these jets is similar to that of single photons. Since the simulated QCD multijet events passing the photon selection tend to suffer from low statistics, the $\gamma + \text{jets}$ MC sample is used to simulate the QCD multijet background in this study. Because of different quark/gluon-initiated jet fractions in the recoil system, the energy scale of the hadronic recoil is $\sim 3\%$ lower in QCD multijet events than in single photon events [12]; thus, the hadronic recoil is scaled down when the $\gamma + \text{jets}$ MC sample is used to model the QCD multijet background events. The events are reweighted so that the simulated QCD multijet background contribution reproduces the observed q_T spectrum in data after subtracting the expected prompt photon, diphoton, and electroweak background contributions. The photon q_T distributions in data and simulated samples are presented in Fig. 3.

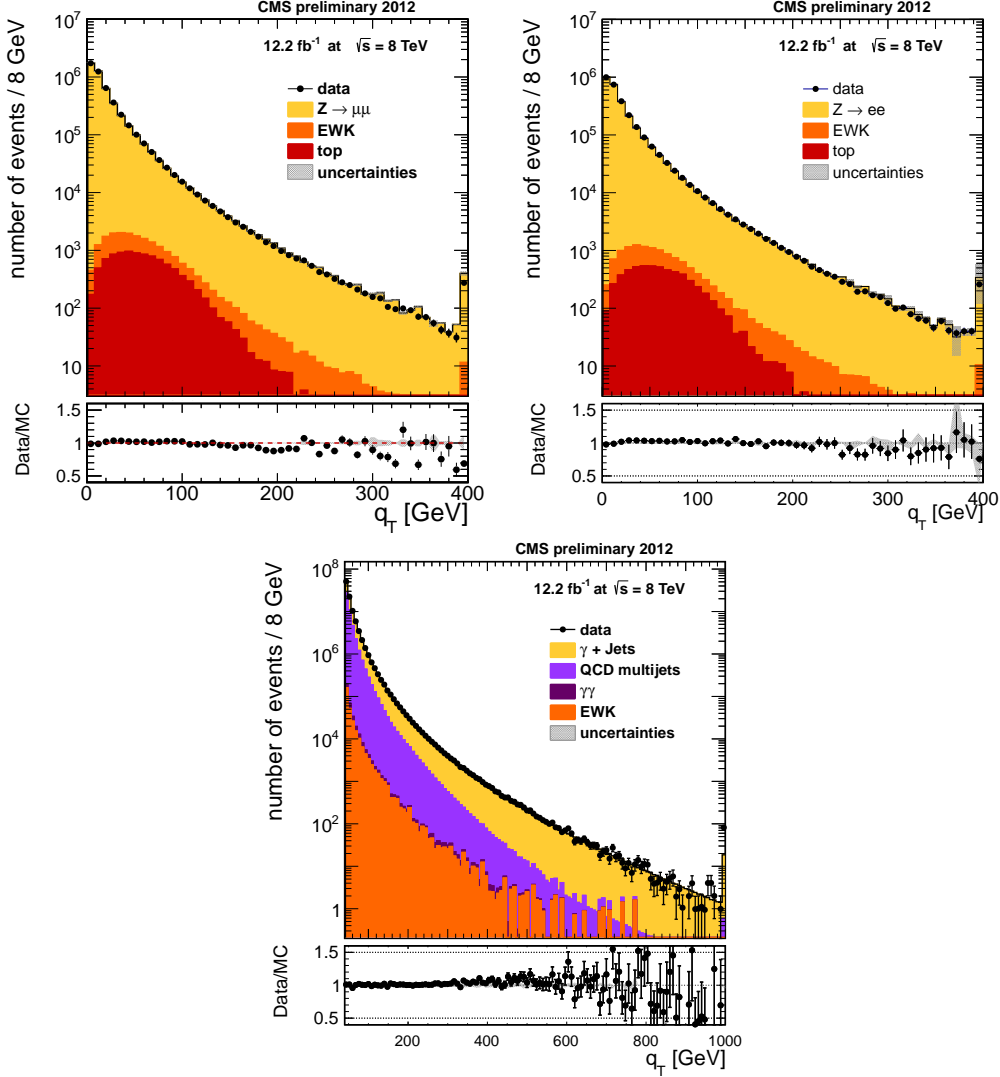


Figure 3: q_T distributions in events with $Z \rightarrow \mu^+\mu^-$ (left), $Z \rightarrow e^+e^-$ (right), and photon events (bottom). The points in the lower section of the plots contain the statistical uncertainties of data and simulation, and the grey error band shows the systematic uncertainty on the simulation.

3 Reconstruction of \vec{E}_T

We define \vec{E}_T as the imbalance in the transverse momentum of all visible particles that interact via the electromagnetic or strong forces in the final state of proton-proton collisions. Because momentum is conserved in each direction, \vec{E}_T is the transverse momentum that is carried by weakly interacting particles, such as neutrinos. CMS has developed several distinct and complementary algorithms to reconstruct \vec{E}_T [3]. The \vec{E}_T reconstructed using a particle-flow technique is used in the majority of CMS analyses due to its high performance; thus, it is discussed most frequently in this note. Additionally, we present a comparison of the performance of PF \vec{E}_T and Calo \vec{E}_T . Two advanced \vec{E}_T reconstruction algorithms specifically developed to mitigate effects from large numbers of pileup interactions are discussed in Section 6.

The negative of the vectorial sum over all PF particle transverse momenta is PF \vec{E}_T . Calo \vec{E}_T is calculated using the energies contained in calorimeter towers and their directions relative to the centre of the detector.

The magnitude of the \vec{E}_T can be underestimated for a variety of reasons, including minimum energy thresholds in the calorimeters, p_T thresholds and inefficiencies in the tracker, neutrinos from leptonic decays of particles, and the nonlinearity of the response of the calorimeter for hadronic particles due to its non-compensating nature. This bias is significantly reduced by correcting the p_T of the jets to the true (particle level) p_T using the jet energy corrections [12] and recomputing \vec{E}_T , as shown in Eq. (2):

$$\vec{E}_T^{\text{corr}} = \vec{E}_T - \sum_{\text{jets}} (\vec{p}_{T,\text{jet}}^{\text{corr}} - \vec{p}_{T,\text{jet}}), \quad (2)$$

where the superscript “corr” refers to the corrected values.

This “type-1” correction for \vec{E}_T uses jet energy scale corrections for all jets that have less than 0.9 of their energy in the ECAL and corrected $p_T > 10$ GeV for PF \vec{E}_T , and for jets with $p_T > 20$ GeV for Calo \vec{E}_T .

In addition to the correction discussed above, further corrections are introduced in 2012 data analyses in order to improve the performance of the \vec{E}_T reconstruction in events with large numbers of pileup interactions. Each pileup interaction is a minimum bias pp interaction. The true \vec{E}_T is close to zero in minimum bias pp interactions, and the vectorial \vec{p}_T sum of charged particles is expected to be well balanced by that of neutral particles. However, due to the calorimeter nonlinearity and minimum energy thresholds in the calorimeters used for the PF reconstruction, \vec{E}_T points on average in the direction of the vectorial \vec{p}_T sum of neutral particles, which induces \vec{E}_T . The “type-0” correction accounts for this effect by using the the vectorial \vec{p}_T sum of charged particles associated to pileup vertices as an estimator of the induced \vec{E}_T .

The correction is parametrized by $f(v) = c_1 \cdot v \cdot (1.0 + \text{erf}(-c_2 \cdot v^{c_3}))$ where $\vec{v} = \sum_{\text{charged}} \vec{p}_T$ is the vectorial \vec{p}_T sum of charged particles associated with a given pileup vertex and $v \equiv |\vec{v}|$. The coefficients $c_1 = -0.70$, $c_2 = 0.030$, and $c_3 = 0.91$ are obtained by fitting the \vec{E}_T component parallel to the \vec{v} direction as a function of v in simulated single minimum bias events. When the type-0 correction is applied to the collision data and simulation samples with pileup interactions, the factor $f(v) \cdot \vec{v}/v$, which gives the expected total \vec{E}_T for each pileup interaction, is summed over all pileup vertices and is subtracted from the reconstructed \vec{E}_T :

$$\vec{E}_T^{\text{corr}} = \vec{E}_T - \vec{\Delta}_{\text{PU}} = \vec{E}_T - \sum_{\text{PU-vertices}} f(v) \frac{\vec{v}}{v}. \quad (3)$$

Although particles are produced uniformly in ϕ , some ϕ -asymmetry is observed in the \vec{p}_T sums of calorimeter energy deposits, tracks, and particles reconstructed by the particle-flow algorithm, leading to a ϕ -asymmetry in \vec{E}_T . The cause for this asymmetry has not yet been identified unambiguously; however, it may be attributed to ϕ -dependence of the detector response, imperfect alignment of different detector subsystems, or a ~ 4 mm shift between the centre of the detector and beamline [16].

The ϕ -asymmetry is present not only in the collision data but also in Monte Carlo simulated events. The observed \vec{E}_T ϕ -asymmetry is due to a shift in the E_x and E_y components, which increases approximately linearly with the number of reconstructed vertices. This correlation is utilized for a correction procedure. The \vec{E}_T ϕ -asymmetry corrections are determined separately for data and Monte Carlo simulated events. Linear functions are fitted to the correlation of E_x

and \cancel{E}_y to N_{vtx} , the number of reconstructed vertices:

$$\begin{aligned}\langle \cancel{E}_x \rangle &= c_{x_0} + c_{x_s} \cdot N_{\text{vtx}}, \\ \langle \cancel{E}_y \rangle &= c_{y_0} + c_{y_s} \cdot N_{\text{vtx}}.\end{aligned}\tag{4}$$

The linear dependence of $\langle \cancel{E}_x \rangle$ and $\langle \cancel{E}_y \rangle$ on N_{vtx} is used to correct $\vec{\cancel{E}}_T$ on an event-by-event basis as:

$$\begin{aligned}\cancel{E}_x^{\text{corr}} &= \cancel{E}_x - \langle \cancel{E}_x \rangle = \cancel{E}_x - (c_{x_0} + c_{x_s} \cdot N_{\text{vtx}}), \\ \cancel{E}_y^{\text{corr}} &= \cancel{E}_y - \langle \cancel{E}_y \rangle = \cancel{E}_y - (c_{y_0} + c_{y_s} \cdot N_{\text{vtx}}).\end{aligned}\tag{5}$$

The coefficients c_{x_0} , c_{x_s} , c_{y_0} , and c_{y_s} are determined separately from $Z \rightarrow \mu^+ \mu^-$ candidate events in data and simulation samples, and they are validated in event samples with different final states. These coefficients for the PF $\vec{\cancel{E}}_T$ are shown in Table 1.

Table 1: The parameters for the $\vec{\cancel{E}}_T$ ϕ -asymmetry corrections for PF $\vec{\cancel{E}}_T$ for data and MC simulation.

	c_{x_0} (GeV)	c_{x_s} (GeV)	c_{y_0} (GeV)	c_{y_s} (GeV)
Data	0.27	0.32	-0.23	-0.17
Simulation	0.12	0.02	0.28	-0.13

In this note, the type-1 correction is applied to both PF and Calo $\vec{\cancel{E}}_T$ and the type-0 correction is also applied to PF $\vec{\cancel{E}}_T$. Additionally, all the \cancel{E}_T distributions are corrected for the ϕ -asymmetry. Furthermore, in the Monte Carlo simulated events, jet momenta are smeared in order to account for the jet resolution differences between data and simulation [12], and the $\vec{\cancel{E}}_T$ is recomputed based on the smeared jet momenta.

4 Large \cancel{E}_T due to Misreconstruction

In Ref. [3] we showed the results of studies of anomalous high- \cancel{E}_T events in the data collected during 2010 LHC running, associated with particles striking sensors in the ECAL barrel detector, as well as those caused by beam-halo particles and ECAL dead cells. Studies of anomalous $\vec{\cancel{E}}_T$ events caused by (1) HCAL hybrid photodiode and readout box electronics noise and (2) direct particle interactions with the light guides and photomultiplier tubes of the forward calorimeter were discussed in Ref. [17].

In the 2012 data, we have identified several new types of anomalous events populating the high \cancel{E}_T tail. There are a few channels in the ECAL endcaps that occasionally produce high amplitude anomalous pulses. The affected events are removed based on the total energy and the number of low-quality hits in the channels. A misfire of the HCAL laser calibration system in the HCAL barrel (HB), endcap (HE), or forward (HF) regions can produce false signals in almost all channels in a sub-detector. If this misfire overlaps with a bunch crossing and it results in a trigger, the event can be contaminated and a large fake $\vec{\cancel{E}}_T$ is induced. The affected events are removed based on the hit occupancies in the channels that are used for signal and calibration readout.

Another source of fake $\vec{\cancel{E}}_T$ comes from the track reconstruction. The silicon strip tracker can be affected by coherent noise, which causes the presence of a few ten-thousands of clusters widely distributed in the silicon detectors. A significant fraction of these events are vetoed at

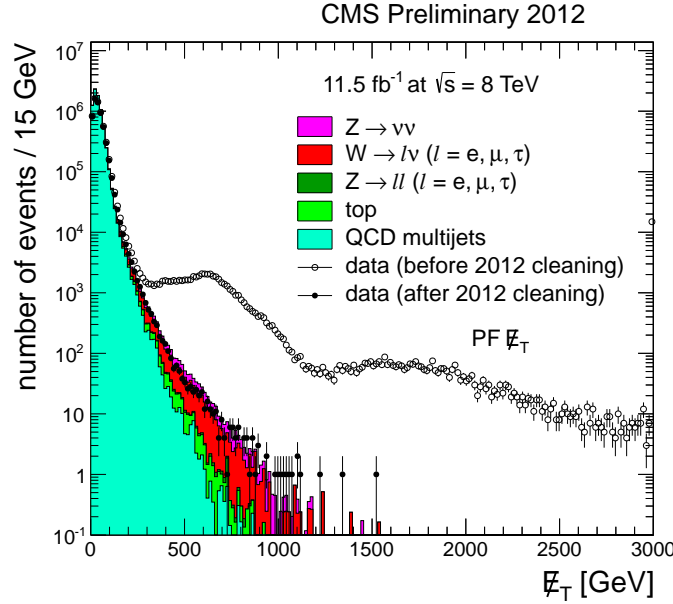


Figure 4: PF E_T distributions for events passing the dijet selection without the 2012 cleaning algorithms applied (open markers), with the 2012 cleaning algorithms applied including the one based on jet identification requirements (filled markers), and events from the simulation (filled histograms).

early stages of the online trigger selection; however, the veto is not fully efficient and some of these events are read out and reconstructed. In such events, there are several fake tracks reconstructed with p_T of hundreds or thousands of GeV. These fake tracks can mimic charged particles, which are then clustered in jets with very high p_T , thus creating a large fake E_T . The affected events are removed based on the number of clusters in the silicon strip and pixel detectors.

Although the rejection of anomalous high- E_T events due to noise in HB and HE were studied in Ref. [3], further developments were necessary to cope with different LHC running conditions, including high luminosities and a shorter bunch crossing interval of 50 ns. A set of algorithms was developed to exploit the differences between noise and signal pulse shapes in order to reject noise hits from event reconstruction. The CMS hadron calorimeter signals are digitized in time intervals of 25 ns, and signals in neighboring time intervals are used to define the pulse shape. In these algorithms, a comparison of the measured and the ideal signal pulse shape is made, and a compatibility test to a signal hypothesis is performed. A log of the ratio of the χ^2 of the two hypotheses is used to identify a channel as noisy. An energy dependent cut is chosen for these algorithms using events collected during runs with no circulating beams in the LHC.

Figure 4 shows the PF E_T distribution for dijet events. The anomalous events with PF E_T around 600 GeV are mainly due to misfires of the HCAL laser calibration system and the anomalous events with PF E_T around 2 TeV are mainly caused by the electronics noise in HB and HE. Even after all the anomalous event cleaning algorithms developed for the 2012 data are applied, we still find some residual anomalous E_T events in the tail of the PF E_T distribution. These events are confirmed to be removed by a jet identification requirement (in which the neutral hadron energy fraction of the jet is less than 0.90 and the photon energy fraction is

less than 0.95), which is used by many physics analyses in CMS (e.g. Ref. [18]). The event is rejected if any jet fails the jet identification criteria. The PF \cancel{E}_T distribution for events passing all cleaning algorithms and jet identification requirements shows a substantial reduction of the high PF \cancel{E}_T tail, and agrees well with the distribution for the simulation.

5 Missing Transverse Momentum Scale and Resolution

In this section, we present studies of the performance of $\vec{\cancel{E}}_T$ using $Z \rightarrow \mu^+\mu^-$, $Z \rightarrow e^+e^-$, and photon events. Momenta of muons originating from Z decays are reconstructed with resolutions of $\sigma_{p_T}/p_T \sim 1\text{--}6\%$ [13], and electron and photon energies are reconstructed with resolutions of $\sigma_E/E \sim 1\text{--}4\%$ [19]. Momenta of jets are reconstructed with less precision, with resolutions of typically $\sigma_{p_T}/p_T \sim 10\text{--}15\%$ [12]. As a consequence, the $\vec{\cancel{E}}_T$ resolution in Z or γ + jets events is dominated by the resolution with which the hadronic activity in the event is reconstructed.

The PF \cancel{E}_T distributions in $Z \rightarrow \mu^+\mu^-$, $Z \rightarrow e^+e^-$, and photon events are presented in Fig. 5. Good agreement of data and simulation is observed in all distributions. Uncertainty bands in the distributions for $Z \rightarrow \mu^+\mu^-$, $Z \rightarrow e^+e^-$, and photon events include uncertainties on the lepton and photon energy scales (0.2% for muons, 0.6% for barrel electrons and photons, and 1.5% for endcap electrons), jet energy scale (2–10%), jet energy resolution (6–15%), and the energy scale of low energy particles, defined as the unclustered energy (10%). The bulge in the uncertainty band in Fig. 5 around PF $\cancel{E}_T = 70$ GeV is due to a maximized impact of jet energy resolution fluctuations in events in this region. The uncertainty becomes smaller at higher PF \cancel{E}_T values where events are dominantly coming from the top or electroweak background processes and the PF \cancel{E}_T distributions have shallower slopes.

While there is no genuine $\vec{\cancel{E}}_T$ in $Z \rightarrow \ell^+\ell^-$ and photon events, it can be induced by removing the well measured vector boson from the event reconstruction. We use this induced $\vec{\cancel{E}}_T$ to study the $\vec{\cancel{E}}_T$ scale and resolution in the following sections.

5.1 Definitions of \cancel{E}_T Scale and Resolution

In $Z \rightarrow \ell^+\ell^-$ and photon events, we can probe the detector response to the global hadronic system and measure the scale and resolution of $\vec{\cancel{E}}_T$ by comparing the transverse momentum of the well measured and well understood vector boson to that of the hadronic recoil system. Many physics and experimental issues contribute to the measured induced $\vec{\cancel{E}}_T$ in these events, such as effects due to jet energy resolution, fluctuating jet composition, underlying event activity, pileup, detector noise, and finite detector acceptance.

The following notation is used: the vector boson momentum in the transverse plane is \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 \vec{u}_T . Momentum conservation in the transverse plane requires $\vec{q}_T + \vec{u}_T + \vec{\cancel{E}}_T = 0$. Figure 6 summarizes these kinematic definitions.

The presence of a well measured Z or γ provides both a momentum scale, $q_T \equiv |\vec{q}_T|$, and a unique event axis, along the unit vector \hat{q}_T . The hadronic recoil can be projected onto this axis, yielding two signed components, parallel (u_{\parallel}) and perpendicular (u_{\perp}) to the event axis. Since $u_{\parallel} \equiv \vec{u}_T \cdot \hat{q}_T$, and the observed hadronic system is usually in the opposite hemisphere from the boson, u_{\parallel} is typically negative. In the following we refer to $-\langle u_{\parallel} \rangle / q_T$ as the $\vec{\cancel{E}}_T$ response, and

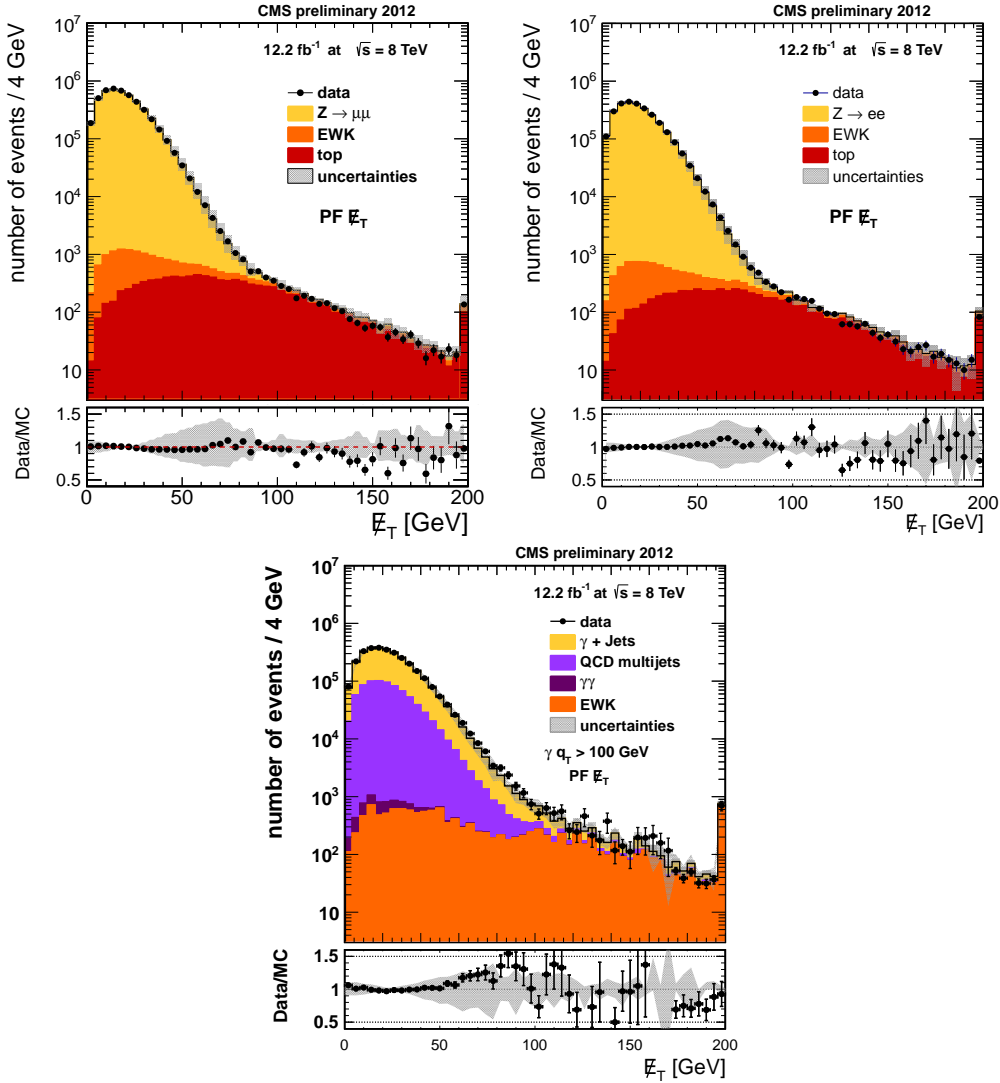


Figure 5: PF E_T distribution in events with $Z \rightarrow \mu^+\mu^-$ (left), $Z \rightarrow e^+e^-$ (right), and photon events (bottom). The points of the lower section of the plots contain the statistical uncertainties of data and simulation, and the grey error band is the systematic uncertainty on the simulation.

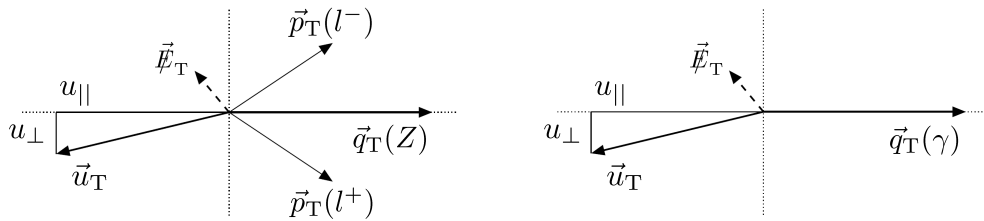


Figure 6: Illustration of $Z \rightarrow \ell^+\ell^-$ (left) and photon (right) event kinematics in the transverse plane. The vector \vec{u}_T denotes the vectorial sum of all particles reconstructed in the event except for the two leptons from the Z decay (left) or the photon (right).

denote a distribution of $-\langle u_{\parallel} \rangle / q_T$ versus q_T as a response curve.

The resolution is assessed with a parametrization of the $u_{\parallel} + q_T$ and u_{\perp} distributions by a Voigtian function, defined by the convolution of a Breit-Wigner (BW) distribution and a Gaus-

sian (G) distribution:

$$V(x; \sigma, \gamma) = \int G(y, \sigma) BW(x - y, \gamma) dy,$$

as it is found to describe the observed $u_{\parallel} + q_T$ and u_{\perp} distributions very well. The resolution in u_{\parallel} and u_{\perp} , denoted by $\sigma(u_{\parallel})$ and $\sigma(u_{\perp})$, is given by the full width at half maximum (FWHM) of the Voigtian form, divided by $2\sqrt{2 \ln 2}$. As with the response, we will examine the resolution as a function of q_T .

5.2 Measurement of PF \vec{E}_T Scale and Resolution

As mentioned earlier, we decompose the recoil with respect to the boson (Z or γ) direction in the transverse plane in order to study the \vec{E}_T scale and resolution. Distributions of the u_{\perp} and $u_{\parallel} + q_T$ components of the recoil calculated from PF \vec{E}_T are shown in Fig. 7 for $Z \rightarrow \mu^+ \mu^-$, $Z \rightarrow e^+ e^-$, and photon events. The back-to-back nature of events implies that u_{\parallel} is balanced with q_T , thus making $u_{\parallel} + q_T$ centred around zero and approximately symmetric. The data/MC ratio versus the boson q_T shows a rising slope, indicating some differences in the \vec{E}_T scale between data and MC at the level of the quoted systematic uncertainty. Due to the nature of the perpendicular recoil component energy fluctuations (detector noise and underlying event), the distribution of u_{\perp} is symmetric. The discrepancy between data and MC simulation observed in the photon u_{\perp} spectrum is due to the limitations of the leading-order event generator PYTHIA used for simulating photon events. The uncertainty on the $u_{\parallel} + q_T$ and u_{\perp} distributions increases around ± 70 GeV due to the jet energy resolution uncertainty.

The response curves extracted from data, $-\langle u_{\parallel} \rangle / q_T$ versus q_T , are shown in Fig. 8 for $Z \rightarrow \mu^+ \mu^-$, $Z \rightarrow e^+ e^-$, and photon events. The agreement between data and simulation is reasonable for each channel. The Z curves indicate that the PF \vec{E}_T is able to fully recover the hadronic recoil activity corresponding to a boosted Z of $q_T \sim 40$ GeV. Below 40 GeV, the uncorrected unclustered energy contribution starts to be significant compared to the corrected energy of the recoiling jets, leading to an under-estimation of the response. The γ + jets curves are 2–3% lower than the Z curves at $q_T < 100$ GeV, which is mainly attributed to the higher fraction of gluon jets in the hadronic recoil in photon events because of a sizable amount of QCD multi-jets background. The resolution curves, $\sigma(u_{\parallel})$ and $\sigma(u_{\perp})$ versus q_T , are shown in Fig. 9. The resolution increases with increasing q_T , and the data and simulation curves are in reasonable agreement for each channel.

Since the q_T spectrum is not the same between the Z and γ + jets channels, the comparison of resolution curves between the Z and γ + jets channels may be affected due to their dependence on the q_T spectrum. Thus, for the remaining resolution curves shown in this section, both Z and γ + jets events are required to satisfy $q_T > 100$ GeV and a reweighting of both Z data and simulation is applied on an event-by-event basis to make a q_T spectrum similar to that of γ + jets data. Figure 10 shows the resolution of the PF \vec{E}_T projections along the x - and y -axes as a function of $\sum E_T$. The $\sum E_T$ is the scalar sum of E_T of all the particles reconstructed by the particle-flow reconstruction except for dileptons from Z's or photons. Resolution curves are found to be in agreement between different channels and the measured curves in data are well described by the simulation. The resolution curves can be parametrized by:

$$\sigma(E_x, E_y) = \sigma_0 + \sigma_s \sqrt{\sum E_T}, \quad (6)$$

where σ_0 is the intrinsic detector noise resolution and σ_s is the \vec{E}_T resolution stochastic term. The σ_0 parameter is not well constrained in fits and has sizable uncertainties. The stochastic term is $\sigma_s \sim 0.6$ and is compatible for different channels as shown in Table 2.

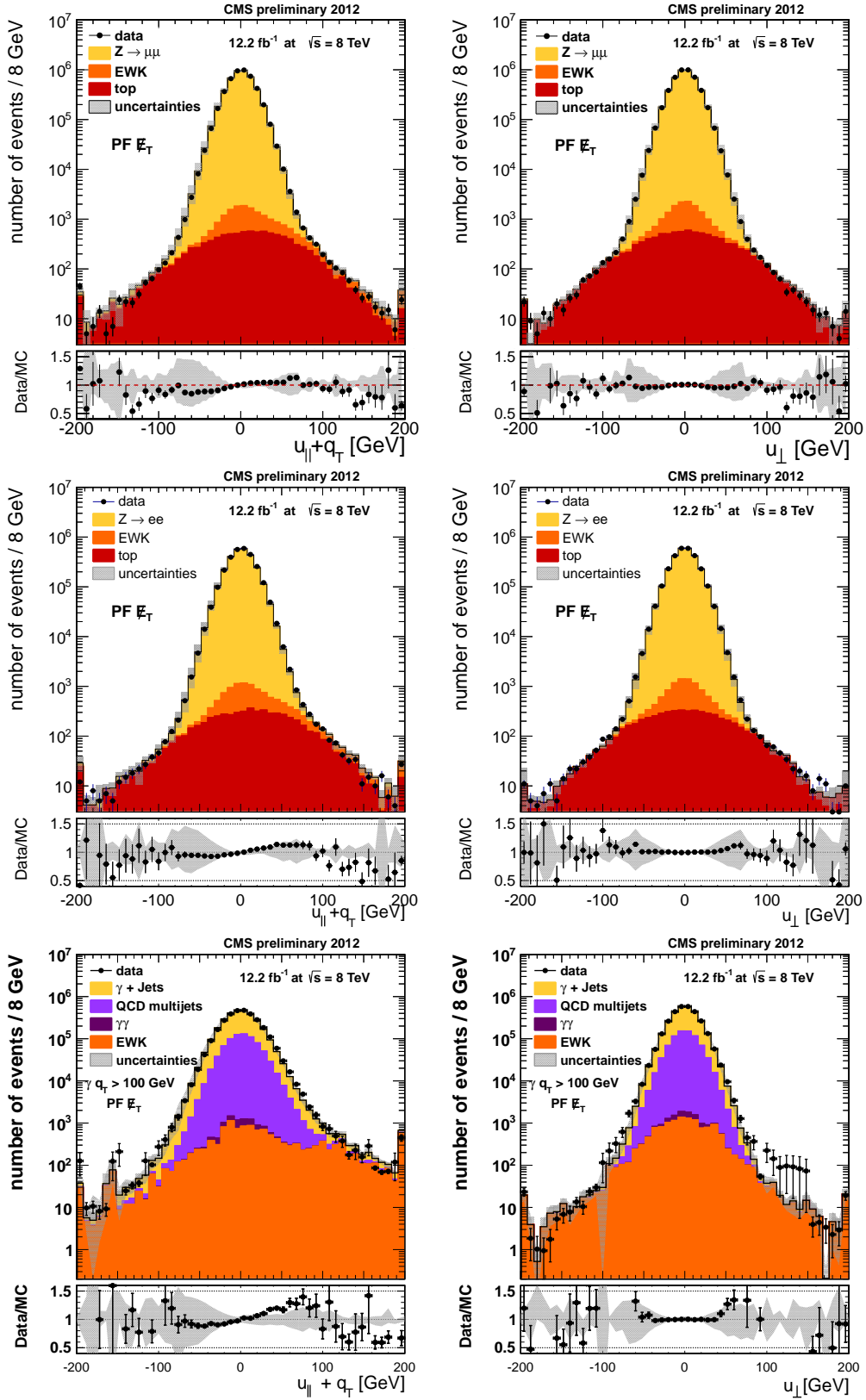


Figure 7: $u_{\parallel} + q_T$ distributions for PF \vec{E}_T for $Z \rightarrow \mu^+\mu^-$ (top, left), $Z \rightarrow e^+e^-$ (middle, left), events (bottom, left); u_{\perp} distributions for PF \vec{E}_T for $Z \rightarrow \mu^+\mu^-$ (top, right), $Z \rightarrow e^+e^-$ (middle, right), and photon events (bottom, right). Statistical uncertainties and systematic uncertainties are shown as grey bands on the prediction from the simulation.

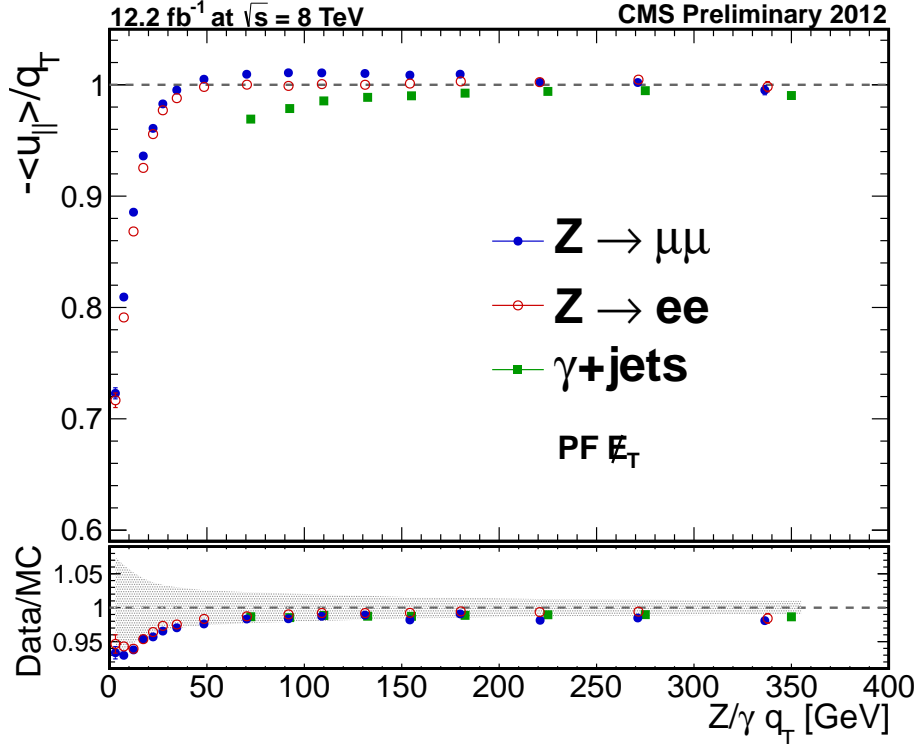


Figure 8: Response curves for PF \vec{E}_T for events with Z and γ . Results are shown for $Z \rightarrow \mu^+\mu^-$ events (full blue circles), $Z \rightarrow e^+e^-$ events (open red circles), and photon events (full green squares). The upper frame shows the response in data; the lower frame shows the ratio of data to simulation. The q_T value for each point is determined based on the average q_T value in data contributing to each point.

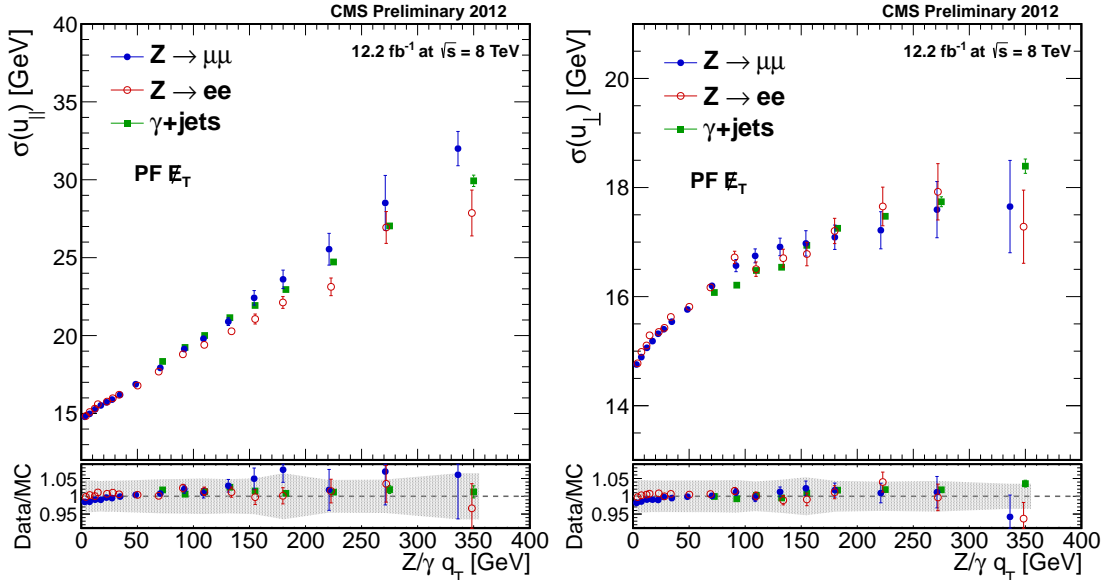


Figure 9: Resolution curves of the parallel recoil component (left) and perpendicular recoil component (right) versus $Z/\gamma q_T$ for PF \vec{E}_T for events with Z and γ . Results are shown for $Z \rightarrow \mu^+\mu^-$ events (full blue circles), $Z \rightarrow e^+e^-$ events (open red circles), and photon events (full green squares). The upper frame of each figure shows the response in data; the lower frame shows the ratio of data to simulation. The q_T value for each point is determined based on the average q_T value in data contributing to each point.

Table 2: Parametrization results of the resolution curves for the components of \vec{E}_T , as functions of $\sum E_T$. The parameter values σ_0 and σ_s are obtained from collision data. For each parameter, we also present R , the ratio of values obtained in data and simulation. For the ratios, the first uncertainty is from the fit, and the second uncertainty corresponds to the propagation of jet energy scale, jet energy resolution, lepton/photon energy scale, and unclustered energy scale uncertainties in the parametrization.

channel	\vec{E}_x component			
	σ_0 (GeV)	$R = \sigma_0(\text{data})/\sigma_0(\text{MC})$	σ_s (GeV ^{1/2})	$R = \sigma_s(\text{data})/\sigma_s(\text{MC})$
$\gamma + \text{jets}$	0.37 ± 0.42	$0.12 \pm 0.14 \pm 0.19$	0.61 ± 0.01	$1.15 \pm 0.03 \pm 0.15$
$Z \rightarrow e^+e^-$	0.05 ± 0.59	$0.05 \pm 0.59 \pm 0.05$	0.63 ± 0.02	$1.07 \pm 0.05 \pm 0.11$
$Z \rightarrow \mu^+\mu^-$	0.87 ± 0.36	$0.40 \pm 0.20 \pm 1.24$	0.62 ± 0.01	$1.10 \pm 0.03 \pm 0.14$
	\vec{E}_y component			
	σ_0 (GeV)	$R = \sigma_0(\text{data})/\sigma_0(\text{MC})$	σ_s (GeV ^{1/2})	$R = \sigma_s(\text{data})/\sigma_s(\text{MC})$
$\gamma + \text{jets}$	0.17 ± 0.37	$0.05 \pm 0.11 \pm 0.13$	0.62 ± 0.01	$1.17 \pm 0.03 \pm 0.16$
$Z \rightarrow e^+e^-$	0.90 ± 0.57	$0.45 \pm 0.31 \pm 0.30$	0.59 ± 0.02	$1.07 \pm 0.05 \pm 0.12$
$Z \rightarrow \mu^+\mu^-$	1.42 ± 0.41	$1.02 \pm 0.42 \pm 3.61$	0.60 ± 0.01	$1.02 \pm 0.04 \pm 0.03$

Figure 11 shows the resolution curves $\sigma(u_{\parallel})$ and $\sigma(u_{\perp})$ versus the number of primary vertices N_{vtx} , for both Z channels and the $\gamma + \text{jets}$ channel. The offset of the curve is related to the resolution in Z or $\gamma + \text{jets}$ events without pileup and the dependence with respect to N_{vtx} indicates how much the pileup degrades the \vec{E}_T resolution. Since the hard-scatter interaction and each additional collision are uncorrelated, these resolution curves can be parametrized by the function:

$$f(N_{\text{vtx}}) = \sqrt{\sigma_c^2 + \frac{N_{\text{vtx}}}{0.7} \times \sigma_{\text{PU}}^2}, \quad (7)$$

where σ_c is the resolution term induced by the hard-scatter interaction and σ_{PU} is the resolution term induced on average by one additional pileup collision. The factor 0.7 accounts for the fact that only 70% of pp interactions produce a reconstructed vertex. Results of the parametrizations are given in Table 3. From Table 3 one can see that different channels are compatible with each other, and that the simulation describes well the performance obtained in data. One may also notice that for one additional pileup interaction the PF \vec{E}_T resolution is degraded by 3.3–3.7 GeV in quadrature. Figure 12 demonstrates the resolution improvements in PF \vec{E}_T with respect to Calo \vec{E}_T .

6 Pileup-mitigated \vec{E}_T and its Performance

Since the vast majority of pileup interactions do not have significant \vec{E}_T and the average value of \vec{E}_T projected in any direction is zero, the effect on \vec{E}_T response from pileup interactions is small. However, as shown in Section 5, pileup interactions have a considerable effect on the \vec{E}_T resolution. Table 3 shows that each pileup interaction adds an additional 3.3–3.7 GeV of smearing to the \vec{E}_T resolution in quadrature in both the transverse and parallel components with respect to the boson \vec{q}_T in $Z \rightarrow \mu^+\mu^-$, $Z \rightarrow e^+e^-$, and photon events. In events where the recoil p_T is small and the instantaneous luminosity is around the mean value of 2012 data taking, which corresponds to approximately 20 pileup interactions, the contribution to the \vec{E}_T resolution from pileup interactions is larger than the contribution from the true recoil of the pp interaction producing Z bosons or photons.

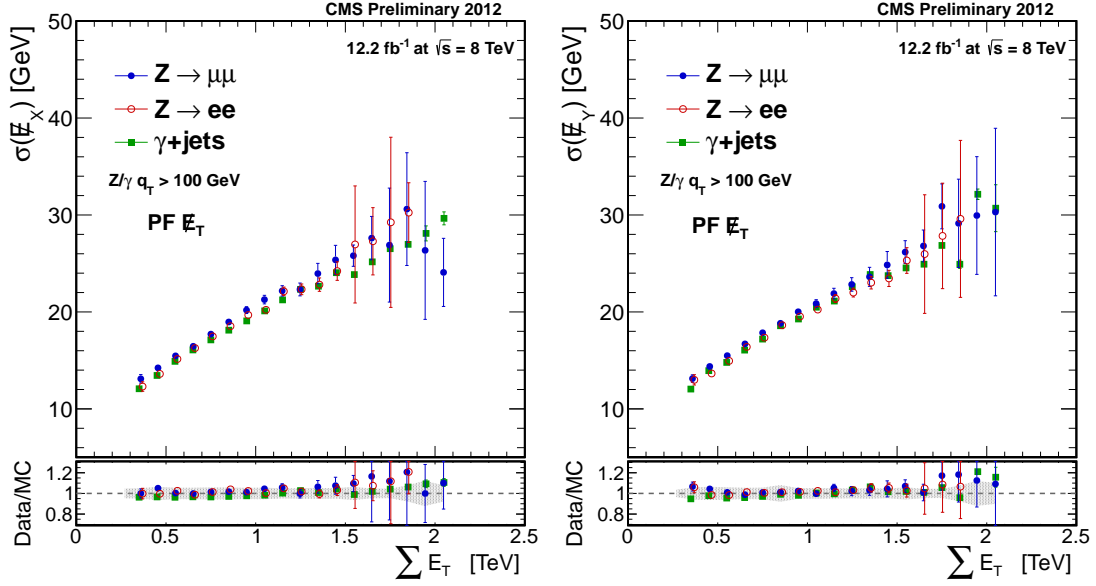


Figure 10: Resolution of the PF \vec{E}_T projection along the x-axis (left) and the y-axis (right) as a function of ΣE_T for events with Z and γ . Results are shown for $Z \rightarrow \mu^+\mu^-$ events (full blue circles), $Z \rightarrow e^+e^-$ events (open red circles), and photon events (full green squares). The upper frame of each figure shows the response in data; the lower frame shows the ratio of data to simulation.

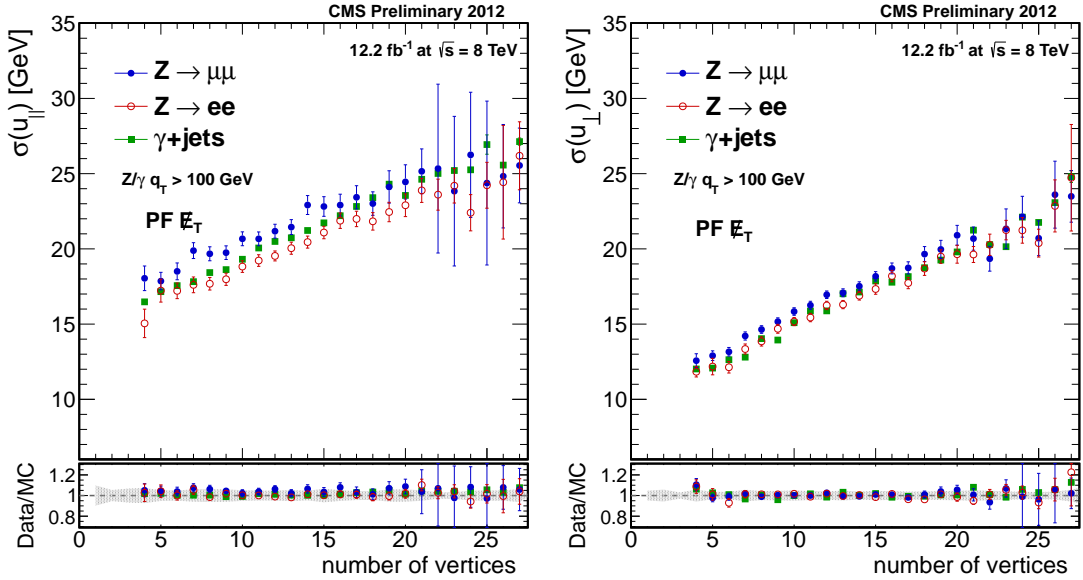


Figure 11: Parallel recoil component (left) and perpendicular recoil component (right) resolution curves versus the number of reconstructed vertices for PF \vec{E}_T for events with Z and γ . Results are shown for $Z \rightarrow \mu^+\mu^-$ events (full blue circles), $Z \rightarrow e^+e^-$ events (open red circles), and photon events (full green squares). The upper frame of each figure shows the response in data; the lower frame shows the ratio of data to simulation.

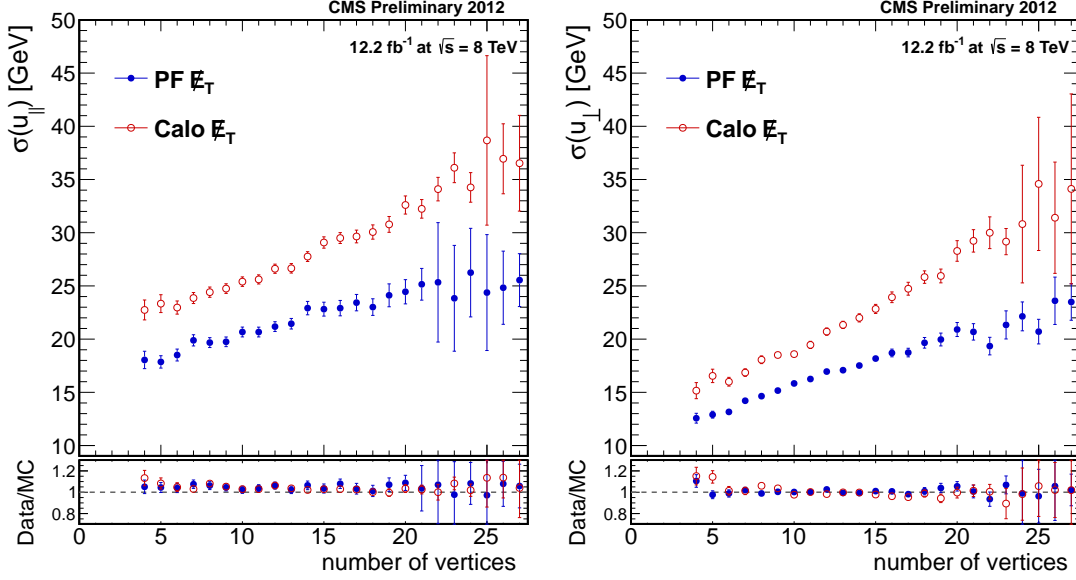


Figure 12: Parallel recoil component (left) and perpendicular recoil component (right) resolution curves for the PF \cancel{E}_T and Calo \cancel{E}_T algorithms versus the number of reconstructed vertices for Calo \cancel{E}_T for $Z \rightarrow \mu^+ \mu^-$ events. The upper frame of each figure shows the response in data; the lower frame shows the ratio of data to simulation.

Table 3: Parametrization results of the resolution curves for the u_{\parallel} and u_{\perp} components as functions of N_{vtx} . The parameter values σ_c and σ_{PU} are obtained from collision data. For each parameter, we also present R , the ratio of values obtained in data and simulation. For the ratios, the first uncertainty is from the fit, and the second uncertainty corresponds to the propagation of jet energy scale, jet energy resolution, lepton/photon energy scale, and unclustered energy scale uncertainties in the parametrization.

channel	u_{\parallel} component			
	σ_c (GeV)	$R = \sigma_c(\text{data})/\sigma_c(\text{MC})$	σ_{PU} (GeV)	$R = \sigma_{\text{PU}}(\text{data})/\sigma_{\text{PU}}(\text{MC})$
$\gamma + \text{jets}$	13.48 ± 0.15	$0.95 \pm 0.01 \pm 0.06$	3.73 ± 0.03	$1.06 \pm 0.01 \pm 0.06$
$Z \rightarrow e^+ e^-$	13.18 ± 0.45	$0.97 \pm 0.05 \pm 0.08$	3.52 ± 0.09	$1.03 \pm 0.04 \pm 0.08$
$Z \rightarrow \mu^+ \mu^-$	15.74 ± 0.28	$1.06 \pm 0.03 \pm 0.06$	3.46 ± 0.07	$1.02 \pm 0.03 \pm 0.04$
	u_{\perp} component			
	σ_c (GeV)	$R = \sigma_c(\text{data})/\sigma_c(\text{MC})$	σ_{PU} (GeV)	$R = \sigma_{\text{PU}}(\text{data})/\sigma_{\text{PU}}(\text{MC})$
$\gamma + \text{jets}$	7.53 ± 0.08	$0.92 \pm 0.01 \pm 0.10$	3.43 ± 0.01	$1.03 \pm 0.00 \pm 0.06$
$Z \rightarrow e^+ e^-$	8.39 ± 0.41	$1.08 \pm 0.08 \pm 0.14$	3.29 ± 0.06	$0.97 \pm 0.02 \pm 0.07$
$Z \rightarrow \mu^+ \mu^-$	9.55 ± 0.23	$1.04 \pm 0.04 \pm 0.06$	3.33 ± 0.04	$1.00 \pm 0.02 \pm 0.05$

In this section we discuss two algorithms developed to reduce the effect of pileup interactions on \vec{E}_T reconstruction, hereafter referred to as the “No-PU PF \vec{E}_T ” and “MVA PF \vec{E}_T ” algorithms. These algorithms are based on identifying two distinct classes of PF particles in each event: PF particles that are likely to originate from the primary hard-scattering pp interaction (HS particles) and PF particles that are likely to originate from pileup interactions (PU particles). Both algorithms attempt to minimize the influence of pileup interactions on \vec{E}_T resolution at the cost of occasionally over-subtracting the contributions of PU particles to \vec{E}_T and thus leading to a \vec{E}_T response reduced by a few GeV.

Separation of charged PF particles originating from the primary hard-scattering pp interaction and those from pileup interactions is best performed by matching them to the primary vertex or pileup vertices. Jets originating primarily from pileup interactions (pileup jets) can also be identified by associating charged PF particles in jets to the primary and pileup vertices and also by using the jet shape variables. Pileup jets often appear as an agglomeration of lower p_T sub-jets. We use a multivariate boosted decision tree (BDT) algorithm in order to identify pileup jets and it is referred to as “MVA-based pileup jet identification discriminator” (MVA pileup jet ID) [2]. Both No-PU and MVA PF \vec{E}_T algorithms utilize the MVA pileup jet ID.

The MVA pileup jet ID incorporates variables designed to identify pileup jets by their jet shapes or through their vertex quality properties. The jet shape information used in the algorithm consists of two sets of variables, which are defined as:

$$\langle \Delta R \rangle = \frac{1}{p_T^{\text{jet}}} \sum_{\text{all particles}} p_T \Delta R, \quad (8)$$

$$p_T^{\Delta R}(X) = \frac{1}{p_T^{\text{jet}}} \sum_{\text{particles with } X \leq \Delta R < X+0.1} p_T, \quad (9)$$

where p_T and p_T^{jet} are the transverse momentum of each PF particle and the jet, respectively, and ΔR is the distance between each PF particle and the jet axis in η - ϕ space. The summations are performed over all PF particles associated to the jet, and $p_T^{\Delta R}(X)$ is calculated for five values of X from 0.0 to 0.4 in steps of 0.1. Two vertex-specific variables are also added to the MVA training:

$$\beta = \frac{\sum_{\text{HS-charged}} p_T}{\sum_{\text{all charged}} p_T}, \quad \beta^* = \frac{\sum_{\text{PU-charged}} p_T}{\sum_{\text{all charged}} p_T}, \quad (10)$$

where $\sum_{\text{HS-charged}} p_T$ and $\sum_{\text{PU-charged}} p_T$ are the scalar p_T sums of charged PF particles in the jet and associated with the primary hard-scatter vertex and pileup vertices, respectively. The $\sum_{\text{all charged}} p_T$ is the scalar p_T sum of all charged PF particles in the jet. The number of charged PF particles and the number of neutral PF particles within a jet are also added as inputs to the MVA pileup jet ID. To enhance the discrimination power at low p_T , the jet p_T , η , and ϕ are included in the MVA training. The MVA is then trained in 16 separate categories, that are defined with four jet p_T and η intervals, using a sample of simulated $Z \rightarrow \mu^+ \mu^-$ events. During the MVA training, jets are identified as originating from the hard scatter by matching the reconstructed jets to true hard-scatter jets in simulated events. We determine a separate working point in each of the 16 categories, which corresponds to a 95% identification efficiency for true hard-scatter jets with $p_T > 25$ GeV.

More details of the No-PU and MVA PF \vec{E}_T algorithms and their performance in $Z \rightarrow \mu^+ \mu^-$ and $Z \rightarrow e^+ e^-$ events are presented in the following sections. These algorithms provide a crucial improvement to physics analyses sensitive to low to moderate \vec{E}_T , such as Higgs boson searches in τ final states [20].

6.1 No-PU PF \vec{E}_T

The first step in the No-PU PF \vec{E}_T algorithm is to divide the particles reconstructed by the PF algorithm into two distinct categories: particles that are likely to originate from the primary hard-scattering pp interaction and particles that are likely to originate from pileup interactions. The \vec{E}_T is then computed by separately weighting contributions from various objects in the event.

The particles that are considered likely to originate from the primary hard-scattering pp interaction are:

- “leptons” (electrons/photons, muons, and hadronic tau decays),
- particles within jets of $p_T > 30$ GeV that pass the MVA pileup jet ID (“HS-jets”),
- charged hadrons not clustered in jets with $p_T > 30$ GeV and associated to the hard-scatter vertex (“HS-charged”).

The selection of electrons/photons, muons, and hadronic tau decays is analysis-specific. Jets, neutral particles, and charged particles overlapping with any selected lepton within $\Delta R < 0.3$ are excluded from the \vec{E}_T reconstruction in order to avoid double-counting.

Particles that are considered likely to originate from pileup interactions are:

- charged hadrons that are neither within jets of $p_T > 30$ GeV nor associated to the hard-scatter vertex (“PU-charged”),
- neutral particles not within jets of $p_T > 30$ GeV (“neutrals”),
- particles within jets of $p_T > 30$ GeV that fail the MVA pileup jet ID (“PU-jets”).

Particles within the first category enter the \vec{E}_T in the usual way. The contribution of particles in the second category is reweighted to reduce its contribution. This reweighting factor is based on the ratio of transverse momenta carried by tracks that are not within jets of $p_T > 30$ GeV and are either associated to the primary vertex or not:

$$S_F = \frac{\sum_{\text{HS-charged}} p_T}{\sum_{\text{HS-charged}} p_T + \sum_{\text{PU-charged}} p_T}. \quad (11)$$

Based on this scale factor, the No-PU PF \vec{E}_T is then computed as:

$$\begin{aligned} \vec{E}_T = & - \left[\sum_{\text{leptons}} \vec{p}_T + \sum_{\text{HS-jets}} \vec{p}_T + \sum_{\text{HS-charged}} \vec{p}_T \right. \\ & \left. + S_F \cdot \left(\alpha \cdot \sum_{\text{PU-charged}} \vec{p}_T + \beta \cdot \sum_{\text{neutrals}} \vec{p}_T + \gamma \cdot \sum_{\text{PU-jets}} \vec{p}_T + \delta \cdot \vec{\Delta}_{\text{PU}} \right) \right]. \quad (12) \end{aligned}$$

The $\vec{\Delta}_{\text{PU}}$ term corresponds to the type-0 correction applied to the PF \vec{E}_T (cf. Eq. (3)) and improves the No-PU PF \vec{E}_T resolution. The parameters α , β , γ , and δ have been determined by numerical optimization of the \vec{E}_T resolution using a sample of simulated $Z \rightarrow \mu^+ \mu^-$ events. The optimal values found by this procedure are given in Table 4.

6.2 MVA PF \vec{E}_T

The MVA PF \vec{E}_T algorithm is based on a set of multivariate regressions, which provide an improved measurement of the \vec{E}_T in the presence of a high number of pileup interactions. The

Table 4: Values of No-PU PF \vec{E}_T algorithm parameters (cf. Eq. (12)).

Parameter	value
α	1.0
β	0.6
γ	1.0
δ	1.0

MVA PF \vec{E}_T is computed as a correction to the hadronic recoil \vec{u}_T reconstructed from PF particles. The correction is obtained in two steps. First we compute a correction to the azimuthal angle of \vec{u}_T by training a BDT to match the azimuthal angle of $-\vec{q}_T$, which is considered as the true hadronic recoil direction, in simulated $Z \rightarrow \mu^+\mu^-$ events. In the second step, we train another BDT to predict the magnitude of the true \vec{u}_T on a dataset where we have already corrected the direction of the \vec{u}_T using the regression function from the first step. The corrected \vec{u}_T is then added to \vec{q}_T to obtain the negative MVA PF \vec{E}_T . Both regressions are trained on a sample of simulated $Z \rightarrow \mu^+\mu^-$ events and validated in a sample of $Z \rightarrow \mu^+\mu^-$ events from data which cover q_T values up to about 350 GeV. These BDT regressions are not optimized for higher q_T events.

To construct the MVA PF \vec{E}_T , we compute five \vec{E}_T variables calculated from PF particles. These \vec{E}_T variables are listed below:

- the negative vectorial sum of all PF particles in the transverse plane (PF \vec{E}_T),
- the negative vectorial sum of all charged PF particles that have been associated to the selected hard-scatter vertex,
- the negative vectorial sum of all charged PF particles that have been associated to the hard-scatter vertex and all neutral PF particles within jets that have passed the MVA pileup jet ID,
- the negative vectorial sum of all charged PF particles that have not been associated to the hard-scatter vertex and all neutral PF particles within jets that have failed the MVA pileup jet ID,
- the negative vectorial sum of all charged PF particles that have been associated to the hard-scatter vertex and all neutral PF particles (also those that have not been clustered into jets) plus the positive vectorial sum of all neutral PF particles within jets that have failed the MVA pileup jet ID.

For each of the \vec{E}_T variables, the vector \vec{u}_T is computed. The magnitude and azimuthal angle of all five types of \vec{u}_T , scalar sum of transverse momenta of all PF particles of each respective \vec{E}_T variable, the momentum vectors of the two highest p_T jets in the event, and the number of primary vertices are used as inputs to the BDT regression.

6.3 Recoil Correction for No-PU and MVA PF \vec{E}_T

Both the No-PU PF \vec{E}_T and MVA PF \vec{E}_T show reduced dependence on pileup interactions compared to PF \vec{E}_T as is shown in Section 6.4. However, they show larger differences in both scale and resolution between data and MC simulation. In order to compensate for this data-simulation difference, a “recoil” correction is derived from a $Z \rightarrow \mu^+\mu^-$ event sample. In the recoil correction, the u_\perp and $u_\parallel + q_T$ components of the Z boson recoil are parametrized as a

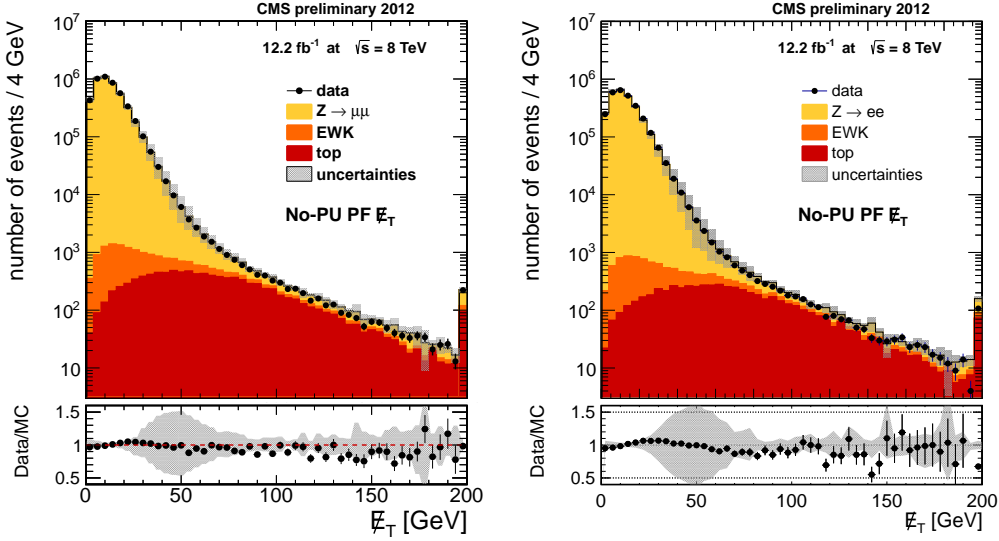


Figure 13: No-PU PF \vec{E}_T distributions in $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) events. The recoil correction is applied to simulated events.

function of the Z boson q_T separately for events having 0, 1, and ≥ 2 jets with $p_T > 30$ GeV. A third-order polynomial and a first-order polynomial, both functions of the boson q_T , are used to describe the resolution and scale of the recoil, respectively. A separate fit is performed on the data and the MC simulation. The parametrization of the recoil as determined from data is used to correct \vec{E}_T in simulated events by rescaling the response and smearing by the difference in quadrature of the resolution. This correction is performed separately as a function of the q_T of the reconstructed Z boson and the jet multiplicity. Finally, the \vec{E}_T of each simulated event is re-computed by replacing the original recoil with the corrected recoil and adding back the \vec{q}_T of the event. The recoil correction allows the No-PU and MVA PF \vec{E}_T algorithms to be used in event topologies where the u_{\parallel} and u_{\perp} can be clearly defined. The recoil correction improves the data-simulation agreement in the performance of the No-PU and MVA PF \vec{E}_T algorithms by typically decreasing the \vec{E}_T response by $\sim 4\%$ and degrading the \vec{E}_T resolution by $\sim 6\%$ in simulated events with q_T around 10–15 GeV.

The No-PU and MVA PF \vec{E}_T distributions are shown in Figs. 13–14 for $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow e^+e^-$ events. Distributions for data and simulation are in agreement within the uncertainties. The increase of the systematic uncertainty in the No-PU and MVA PF \vec{E}_T distributions at $E_T \sim 50$ GeV is dominated by the jet energy resolution uncertainty.

6.4 Measurement No-PU and MVA PF \vec{E}_T Scale and Resolution

Figures 15–16 show the response of the No-PU PF \vec{E}_T and MVA PF \vec{E}_T in $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow e^+e^-$ events, respectively. The data and MC simulation distributions show a fair agreement. The No-PU PF \vec{E}_T response approaches unity slower than the standard PF \vec{E}_T whose response versus q_T is shown in Fig. 8. This is because, when the No-PU PF \vec{E}_T algorithm reweights the transverse momenta of particles from pileup interactions, some particles from the primary pp interaction are also reweighted and it reduces the No-PU PF \vec{E}_T scale. The MVA PF \vec{E}_T response is around 0.9 even at high q_T which is because the BDT training for the MVA PF \vec{E}_T used in this study is optimized for the improved resolution rather than for the unity response.

The u_{\perp} and u_{\parallel} resolutions are shown in Figs. 17–18 versus Z boson q_T , and in Fig. 19 the resolutions of PF \vec{E}_T , No-PU PF \vec{E}_T , and MVA PF \vec{E}_T are compared as a function of N_{vtx} . The

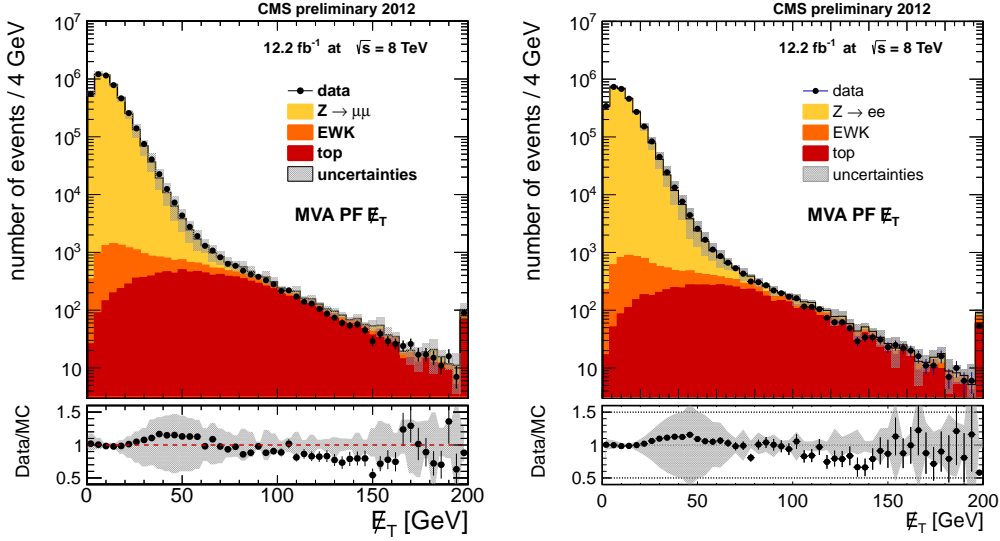


Figure 14: MVA PF \vec{E}_T distributions in $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) events. The recoil correction is applied to simulated events.

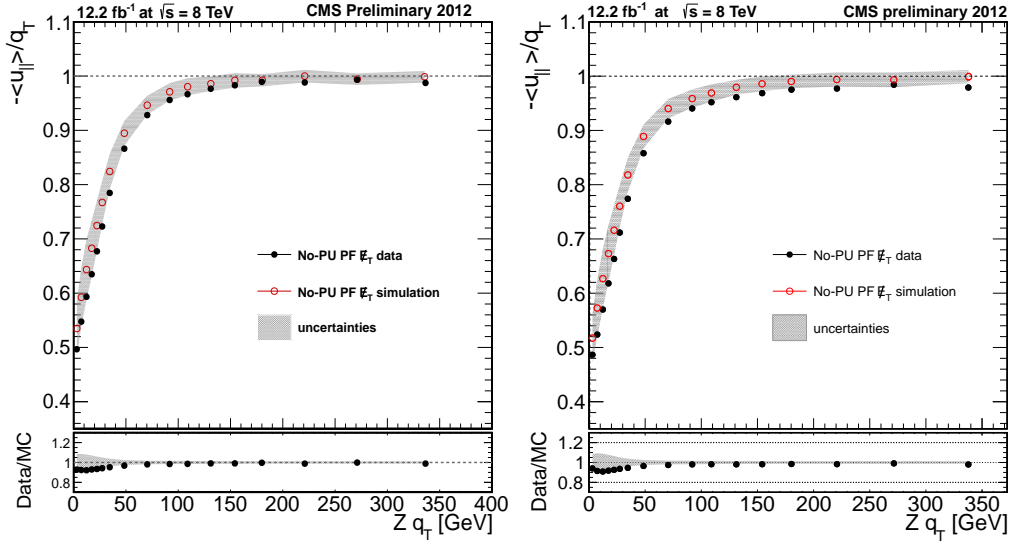


Figure 15: Response as a function of q_T for No-PU PF \vec{E}_T in $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) events.

PF \vec{E}_T used here includes all the corrections described in Section 3, i.e. the type-0, type-1, and ϕ -asymmetry corrections for both data and simulation and additionally jet smearing for simulation, unlike uncorrected PF \vec{E}_T used in comparison with MVA PF \vec{E}_T in Ref. [20]. The No-PU PF \vec{E}_T and particularly MVA PF \vec{E}_T show significantly reduced dependence of the resolution on pileup interactions in both data and MC simulation.

7 Conclusions

The performance of \vec{E}_T reconstruction algorithms has been studied using data collected in 8 TeV pp collisions with the CMS detector at the LHC. The data samples used in this note were collected from February through October 2012 and correspond to an integrated luminosity up to $12.2 \pm 0.5 \text{ fb}^{-1}$. The \vec{E}_T reconstruction algorithms and corrections are described with

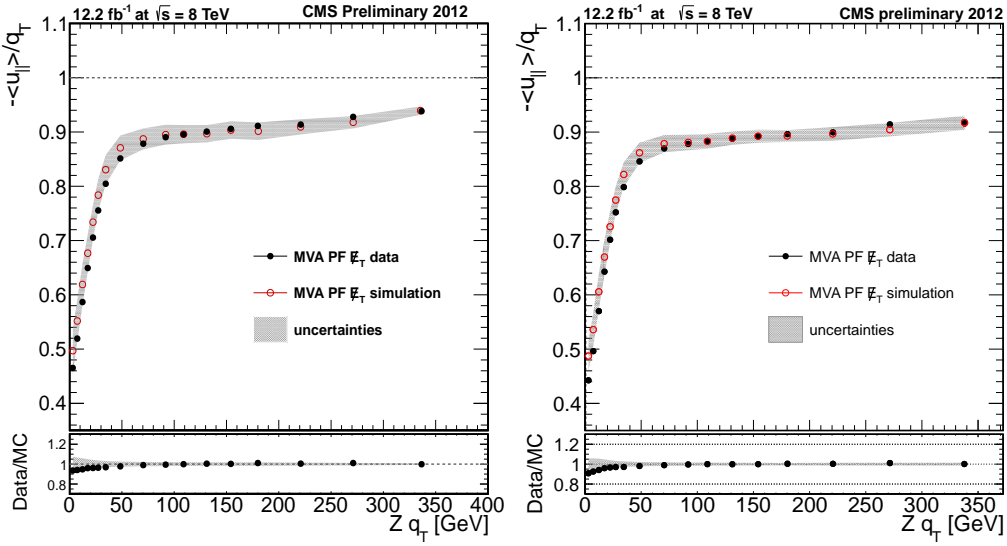


Figure 16: Response as a function of q_T for MVA PF \vec{E}_T in $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) events.

an emphasis on changes compared to those used with the 7 TeV pp collision data collected in 2010 [3]. Events with artificially high \vec{E}_T in a dijet event sample are examined, and it is found that a majority of such events can be identified and removed by the anomalous event cleaning algorithms.

The scale and resolution of PF \vec{E}_T as well as the degradation of the PF \vec{E}_T due to pileup interactions have been measured in $Z \rightarrow \mu^+\mu^-$, $Z \rightarrow e^+e^-$, and photon events. The measured PF \vec{E}_T scale and resolution in data agree with the expectations from the simulation after correcting for the jet energy scale and resolution differences between data and MC simulation. Pileup interactions are found to degrade the PF \vec{E}_T resolution by 3.3–3.7 GeV in quadrature per additional pileup interaction.

The performance of two advanced \vec{E}_T reconstruction algorithms specifically developed to cope with large numbers of pileup interactions has also been studied. They show a significantly reduced dependence of the \vec{E}_T resolution on pileup interactions in $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow e^+e^-$ events.

The studies presented in this note provide a solid foundation for all CMS measurements with \vec{E}_T in the final state, including measurements of the W boson and top quark, searches for new neutral weakly interacting particles, and studies of the properties of the Higgs-like boson.

References

- [1] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [2] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [3] CMS Collaboration, “Missing transverse energy performance of the CMS detector”, *JINST* **6** (2011) P09001, doi:10.1088/1748-0221/6/09/P09001,

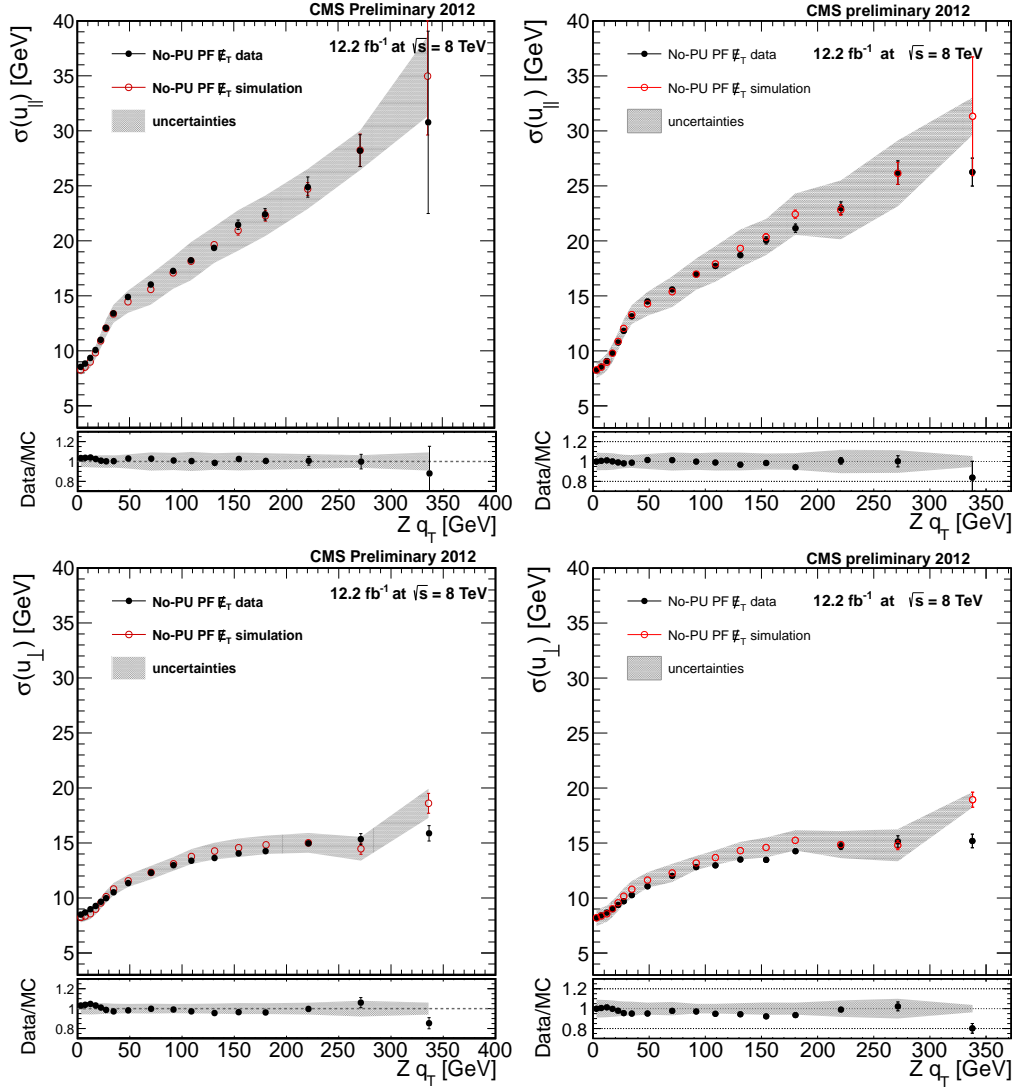


Figure 17: Parallel (top) and perpendicular (bottom) resolution as a function of q_T for No-PU PF \vec{E}_T in $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) events.

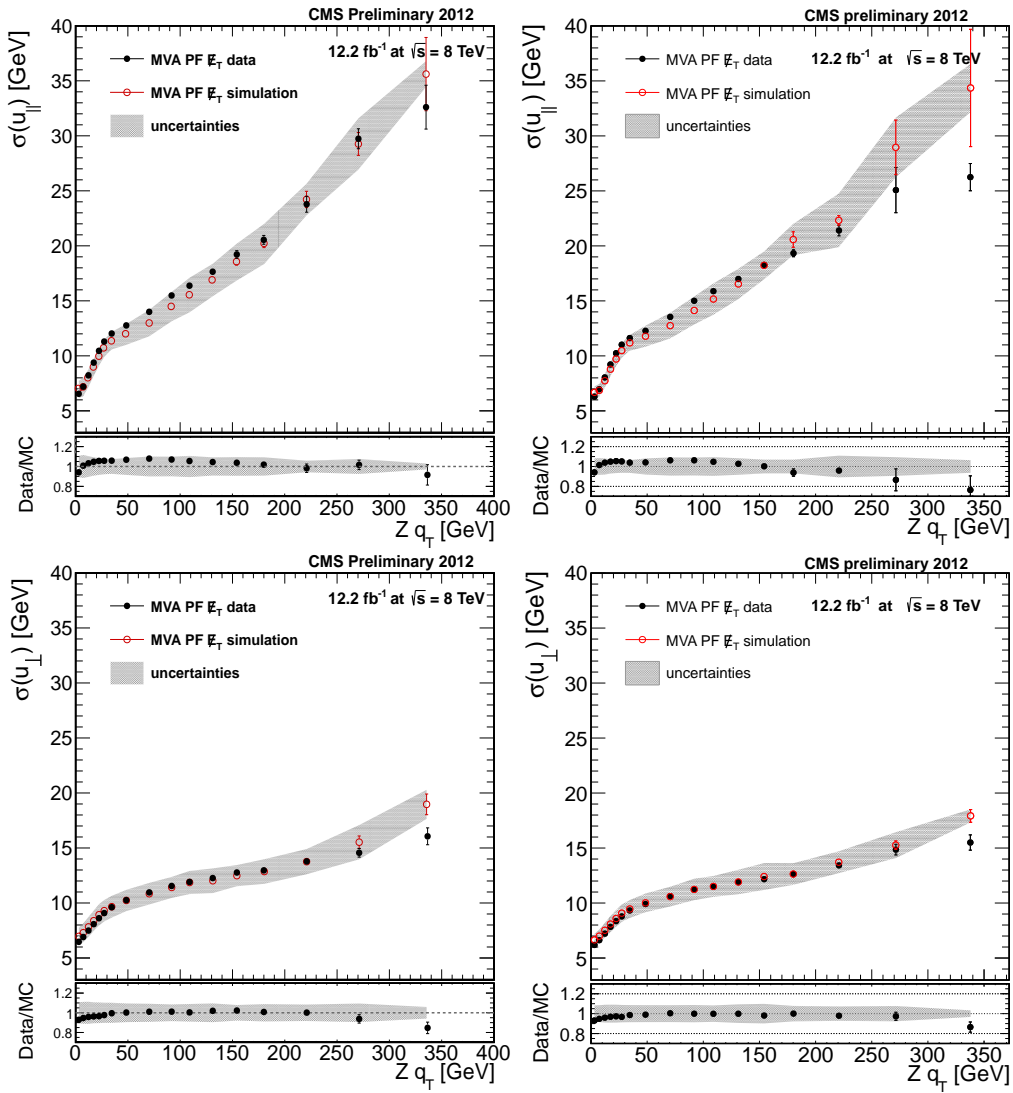


Figure 18: Parallel (top) and perpendicular (bottom) resolution as a function of q_T for MVA PF \vec{E}_T in $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) events.

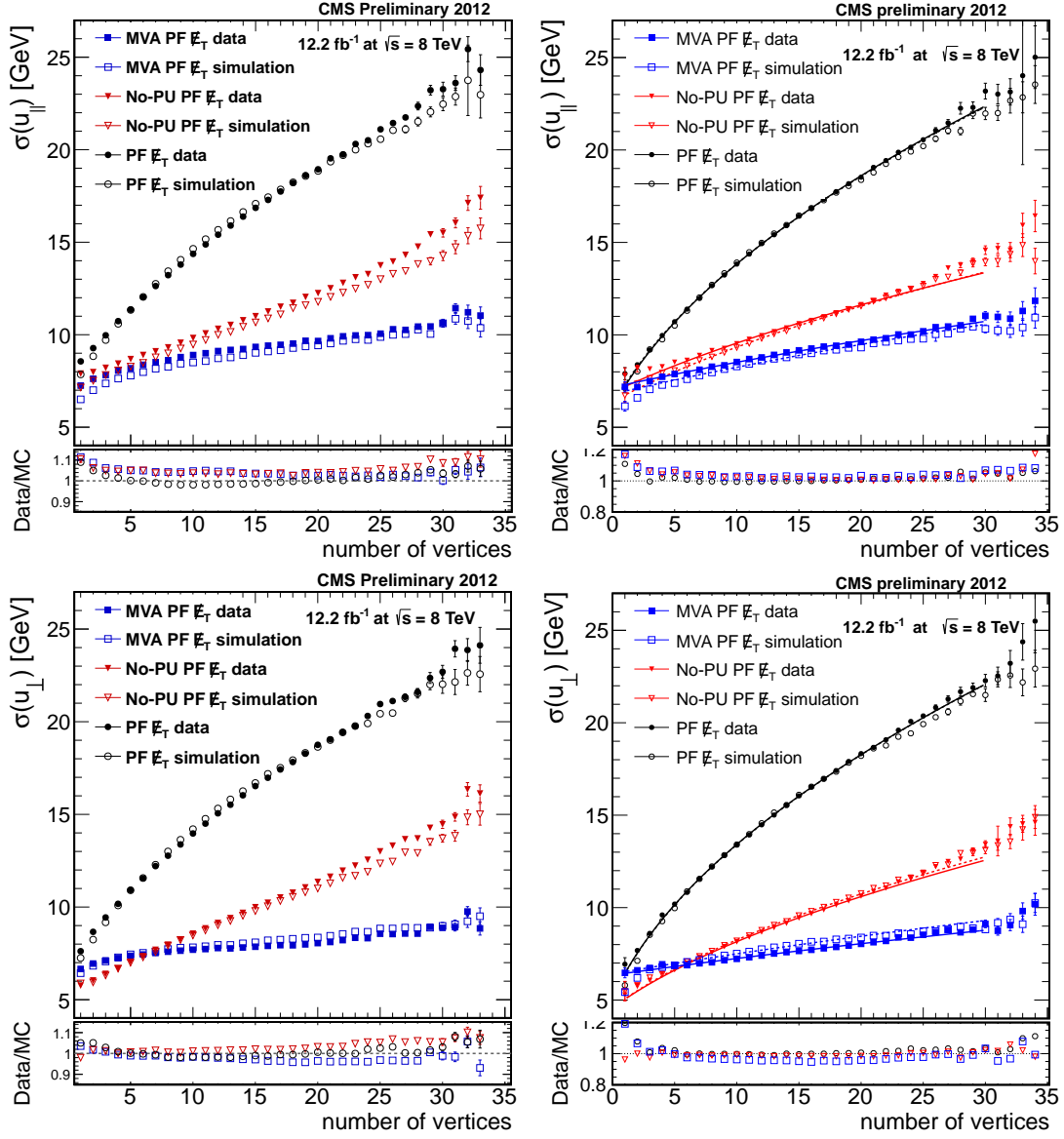


Figure 19: Parallel (top) and perpendicular (bottom) resolution as a function of the number of reconstructed vertices for PF \vec{E}_T , No-PU PF \vec{E}_T , and MVA PF \vec{E}_T in $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) events.

- arXiv:1106.5048.
- [4] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T ”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, (2009).
 - [5] CMS Collaboration, “Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector”, CMS Physics Analysis Summary CMS-PAS-PFT-10-001, (2010).
 - [6] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 physics and manual”, *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.
 - [7] R. Field, “Early LHC Underlying Event Data – Findings and Surprises”, (2010). arXiv:1010.3558.
 - [8] J. Alwall *et al.*, “MadGraph/MadEvent v4: The New Web Generation”, *JHEP* **09** (2007) 028, doi:10.1088/1126-6708/2007/09/028, arXiv:0706.2334.
 - [9] Geant4 Collaboration, “Geant4: A simulation toolkit”, *Nucl. Instrum. and Methods A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
 - [10] CMS Collaboration, “Inelastic pp cross section at 7 TeV”, CMS Physics Analysis Summary CMS-PAS-FWD-11-001, (2011).
 - [11] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting – Summer 2012 Update”, CMS Physics Analysis Summary CMS-PAS-LUM-12-001, (2012).
 - [12] CMS Collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS”, *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.
 - [13] 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.
 - [14] CMS Collaboration, “Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC”, CMS Note CMS-2005/001, (2005).
 - [15] CMS Collaboration, “Photon reconstruction and identification at $\sqrt{s} = 7$ TeV”, CMS Physics Analysis Summary CMS-PAS-EGM-10-005, (2010).
 - [16] CMS Collaboration, “Studies of Tracker Material”, CMS Physics Analysis Summary CMS-PAS-TRK-10-003, (2010).
 - [17] A. Apresyan, “Identification and mitigation of anomalous signals in CMS HCAL”, CMS Conference Report CERN-CMS-CR-2012-238, (2012).
 - [18] CMS Collaboration, “Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 7$ TeV”, *Phys. Rev. Lett.* **109** (2012) 171803, doi:10.1103/PhysRevLett.109.171803, arXiv:1207.1898.
 - [19] CMS Collaboration, “Electron performance with 19.6 fb⁻¹ of data collected at $\sqrt{s} = 8$ TeV with the CMS detector”, CMS Performance Note CMS-DP-2013-003, (2013).
 - [20] CMS Collaboration, “Search for the Standard-Model Higgs boson decaying to tau pairs in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV”, CMS Physics Analysis Summary CMS-PAS-HIG-13-004, (2013).