

# Chapter 1

## Event Selection

In keeping with the phenomenology described in Sec. ??, the candidate GGM events selected in this search consist of two high- $E_T$  photons and a significant momentum imbalance transverse to the beam, indicating the production of an escaping gravitino. This momentum imbalance is usually referred to as *missing transverse energy* and is denoted by the symbol  $\cancel{E}_T$ .

However, in order to use real CMS data (as opposed to simulation) to derive predictions for the backgrounds to the search, *control samples* distinct from the *candidate* two-photon sample must be collected. These samples consist of different numerical combinations of photons, electrons, and jets, and are explained in more detail in Chapter ?. Since this search is performed in the high- $\cancel{E}_T$  tail of the  $\cancel{E}_T$  distribution, where adequate detector simulation is very difficult, it is advantageous to use *data-driven* background estimates, which capture the true detector response, over numbers derived from simulation.

In the following sections, the reconstruction of photons, electrons, jets, and  $\cancel{E}_T$  is explained. Sec. 1.1 begins with an explanation of the high level reconstruction. It is followed by Sec. 1.2, which describes the triggers used to collect the candidate and control samples. Finally, the chapter concludes with a measurement of the photon

identification efficiency in Sec. 1.4.

## 1.1 Object Reconstruction

This section describes the *offline* object reconstruction, i.e. the reconstruction of particle objects from events that have already been triggered and written to permanent storage, as opposed to the building of trigger objects explained in Secs. ?? and 1.2.

### 1.1.1 Photons

#### Uncalibrated EB/EE Hits

Photon reconstruction begins with the ADC count value for each of the 10 recorded time samples per ECAL crystal per trigger. To construct an *uncalibrated hit*, the gain (1, 6, or 12; see Sec. ??) of each sample is determined and the ADC count value scaled appropriately. The pedestal is estimated from the average of the first three samples, which, for a properly time in hit, should contain no signal. This pedestal value is subtracted from the rest of the samples. Finally, the amplitude of the pulse is reconstructed using a predetermined weight for each sample [?]. The weights correspond to the pulse shape expected from the MGPA and shaping circuit response. The time of the hit is also reconstructed using the ratios between neighboring time samples [?]. A typical ECAL channel pulse shape is shown in Figure 1.1.

#### Calibrated EB/EE Hits

In the next phase of the photon reconstruction, calibrations are applied to the uncalibrated hits to form *calibrated hits* with energy measured in GeV. Channels are excluded from seeding calibrated hits if

- they are excessively noisy,

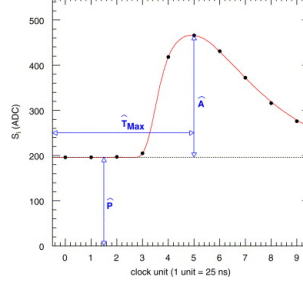


Figure 1.1: Typical ECAL channel pulse shape.  $\hat{P}$  is the pedestal value,  $\hat{A}$  is the pulse amplitude, and  $\hat{T}_{\max}$  is the hit time. The red line is the assumed pulse shape from which the weights are derived. Reprinted from ref. [?].

- they are stuck in fixed gain,
- they are totally dead,
- they have one or more neighboring dead channels, or
- they do not have good trigger primitives (i.e. trigger primitive is missing, saturated, or spike-like).

In addition, no uncalibrated hits that are spike-like are eligible for calibration. The calibrations applied are crystal transparency loss corrections measured continuously by the laser/LED system, energy intercalibrations (relative energy calibration between crystals), absolute scale calibrations between ADC counts and GeV,<sup>1</sup> and time intercalibrations (relative time calibration between crystals).

The ECAL crystals were pre-calibrated before installation in CMS using laboratory light yield and photodetector gain measurements [?]. In addition, some EB and EE crystals were intercalibrated using test beams [?], and all EB crystals were intercalibrated with cosmic ray muons [?]. EE precalibrations were validated with LHC *splash events* in 2009 [?, ?], in which the beam was dumped onto a collimator approximately 150 meters upstream of CMS, causing a spray of muons to enter CMS at one

<sup>1</sup>The ADC-GeV scale factors (one for EB and one for EE) are defined such that the sum of fully calibrated and scaled hits in a particular  $5 \times$  cluster of crystals (plus the associated energy deposited in ES) is 50 GeV for a 50 GeV incident unconverted photon [?].

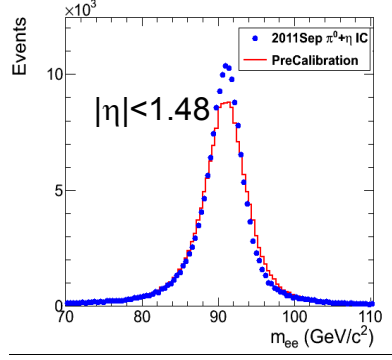


Figure 1.2:  $Z$  peak reconstructed using pre-LHC calibration constants (red) or September 2011  $\pi^0/\eta$ -derived intercalibration constants (blue). Reprinted from ref. [?].

endcap and exit at the other. Splash events were also used to derive time intercalibration constants. Before colliding beam operations commenced, the intercalibration precision was estimated to be 0.5%-2.2% in EB and 1%-5% in EE [?].

Three calibration methods were employed once colliding beam operations began:

- $\phi$  symmetry relative calibration between crystals, exploiting the azimuthal symmetry of CMS
- $\pi^0$  and  $\eta$  relative calibration between crystals, using the diphoton decays of these particles
- $E/p$  absolute calibration, comparing the momentum measured in the tracker  $p$  to the energy measured in the ECAL  $E$  of a sample of electrons from  $Z$  decay

By September 2011, the intercalibration precision in EB was measured to be between 0.3% and 1.1% using the  $\pi^0/\eta$  method [?]. Figure 1.2 shows the improvement in  $Z$  reconstruction from pre-LHC calibration constants to the latest  $\pi^0/\eta$ -derived constants.

## Calibrated ES Hits

ES calibrated hits are formed from the three samples read out per sensor. Just as in the case of EB/EE crystals, ES uncalibrated hits gain-adjusted, pedestal-subtracted, and reconstructed using weights. To make a calibrated ES hit, intercalibration constants, angle correction constants, and a MIP-GeV absolute scale factor are applied.

## Clustering

After calibrated ECAL hits are formed, they must be clustered into shapes that represent the energy deposit from a single particle. *Basic clusters* are formed around seed hits, defined as a hit that

- has calibrated  $E_T > 1(0.18)$  GeV in EB(EE),
- does not originate from a dead channel or one with faulty hardware,
- is not poorly calibrated,
- was reconstructed with the standard algorithm (i.e. not a special recovery algorithm for channels with subpar data integrity),
- is not saturated,
- is not spike-like (**define spike rejection variables here**), and
- is in time (EB).

EB basic clusters are formed around the seeds via the *hybrid* algorithm, while EE basic clusters are formed with the *multi5x5* algorithm [?]. In addition to non-radiating electrons and unconverted photons, both algorithms are designed to also recover all of the energy associated with electron bremsstrahlung deposits and photon conversions. The geometry of the CMS magnetic field means that bremsstrahlung and conversions will tend to spread the shower out in  $\phi$ , not  $\eta$ . Both algorithms work by forming

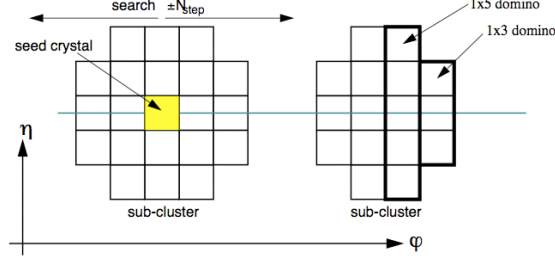


Figure 1.3: Hybrid algorithm in EB. The shower extent is essentially constant in  $\eta$ , but spreads out in  $\phi$  as the two sub-clusters (or basic clusters) are grouped into the same supercluster. Reprinted from ref. [?].

basic clusters around seeds, then combining the basic clusters into *superclusters* (SC) by searching in a window extended in the  $\phi$  direction for all basic clusters consistent with bremsstrahlung radiation from the primary electron, or with a photon conversion. Figure 1.3 illustrates the hybrid algorithm in EB. In EE, the energy deposited in ES must also be added into the total clustered energy sum.

Figure 1.4 shows the effect of superclustering on  $Z \rightarrow ee$  reconstruction.

### Supercluster Corrections

The total clustered ECAL energy is defined as

$$E = F \times \sum_{i=1}^{n_{\text{crystal}}} G \times c_i \times A_i \quad (1.1)$$

where  $G$  is the ADC-GeV or MIP-GeV scale factor,  $c_i$  are the intercalibration constants,  $A_i$  is the uncalibrated hit amplitude in ADC counts, and  $F$  is SC correction factor.  $G$  and  $c_i$  were explained in Sec. 1.1.1.  $F$  is a product of three factors for hybrid SCs (two for multi5x5 SCs) [?]:

1.  $C_{EB}(\eta)$ , which compensates for lateral energy leakage due to the crystal off-pointing in EB. These corrections are taken from MC simulation [?] and were confirmed in test beams [?].

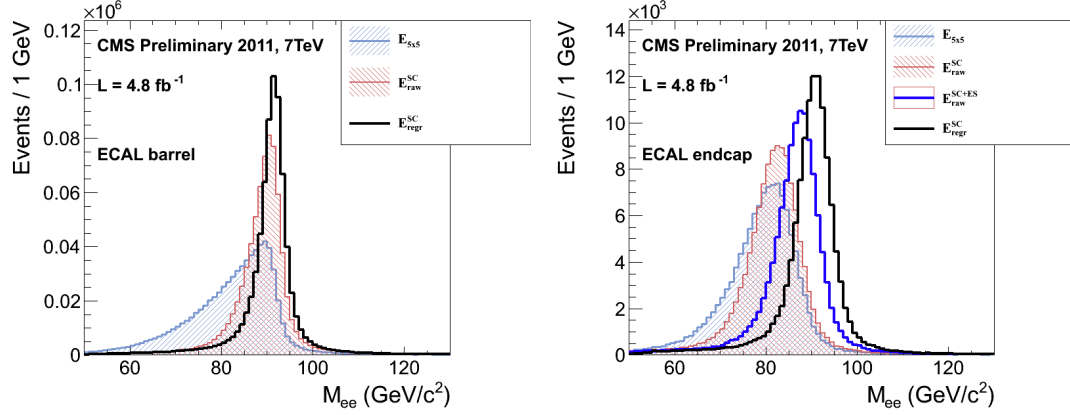


Figure 1.4:  $Z$  peak reconstructed in the dielectron channel for different kinds of clustering. The left plot is for EB and the right plot for EE. The constituent hits were calibrated with the best available intercalibrations and laser calibrations as of December 2011. The light blue histogram shows the reconstruction using a  $5 \times 5$  energy sum, the red histogram shows the reconstruction using the SC energy for crystals only (the dark blue histogram on the right-hand side adds in the energy from ES), and the black histogram shows the reconstruction after the SCs are corrected using a multivariate method [?]. Reprinted from Fig. 30 of ref. [?].

2.  $f(\text{brem})$ , which corrects for the biases in the clustering algorithms for showers characterized by differing amounts of bremsstrahlung. These corrections are taken from MC simulation [?].
3. Residual correction  $f(E_T, \eta)$ , due to the variation in  $\eta$  of detector material traversed by a primary electron or photon, and to any residual  $E_T$  dependence of the reconstruction. These corrections are determined from  $Z \rightarrow ee$  data samples.

As a benchmark of ECAL calibration performance, the extra energy smearing in MC needed to achieve data/MC agreement in the  $Z$  width was between  $\sim 0.9\%$  (in the central part of EB for electrons with little bremsstrahlung) and  $\sim 3.3\%$  (in the outer part of EE for heavily radiating electrons) [?].

## From Supercluster to Photon

The CMS photon object is any SC with  $E_T > 10$  GeV and  $H/E < 0.5$ , unless the SC  $E_T > 100$  GeV, in which case the  $H/E$  requirement is dropped.  $H/E$  is defined as the ratio of energy in the HCAL in a  $0.15$  cone around the SC centroid, directly behind the SC, to the SC energy. SCs with  $R9 > 0.94(0.95)$  in EB(EE), where  $R9$  is defined as the ratio of the energy in the central  $3 \times 3$  cluster of crystals divided by the SC energy  $E_{3 \times 3}/E_{SC}$ , are the best calibrated and most accurate type of electromagnetic shower. Therefore, for these objects, the photon energy is defined as the energy sum of the fully calibrated hits in the central  $5 \times 5$  cluster around the seed (with  $C_{EB}(\eta)$  applied for EB photons). For all other SCs, the photon energy is equal to the fully corrected SC energy (cf. Sec. 1.1.1).

In this search, candidate photons and *fake photons* ( $f$ , “fakes”) are further selected according to the criteria listed in Table 1.1. Fakes are used in the determination of the QCD background, as explained in Chapter ?? .  $I_{\text{comb}}$  is defined as

$$I_{\text{comb}} = I_{\text{ECAL}} - 0.0792\rho + I_{\text{HCAL}} - 0.0252\rho + I_{\text{track}} \quad (1.2)$$

where  $I_{\text{ECAL}}$ ,  $I_{\text{HCAL}}$ , and  $I_{\text{track}}$  are  $E_T$  sums in the annular regions defined in Figure 1.5 and  $\rho$  is the average pileup energy density in the calorimeters (per unit  $\eta \cdot \phi$ ) as measured with the Fastjet algorithm [?, ?]. Note that the ECAL and track isolation veto strips at constant  $\eta$  ensure that the isolation cuts are similarly efficient for converted photons, radiating electrons, and unconverted photons.

Figure 1.6 shows the  $\rho$  distribution for a sample of events with at least two 25 GeV EM objects passing the  $|\eta|$ ,  $H/E$ , and  $R9$  requirements in Table 1.1, and passing the trigger requirements in Table 1.3, representing the full 2011 dataset. Since average  $\rho$  is  $\sim 5$  GeV, and there is a long tail above this average value, it is necessary to



Table 1.1: Selection criteria for photons and fakes.

Variable	Cut ( $\gamma$ )	Cut ( $f$ )
SC $ \eta $	$< 1.4442$	$< 1.4442$
$H/E$	$< 0.05$	$< 0.05$
$R9$	$< 1$	$< 1$
Has pixel seed	No	No
$I_{\text{comb}}, \sigma_{i\eta i\eta}$	$< 6 \text{ GeV} \ \&\& \ < 0.011$	$(\geq 6 \text{ and } < 20 \text{ GeV}) \    \ \geq 0.011$

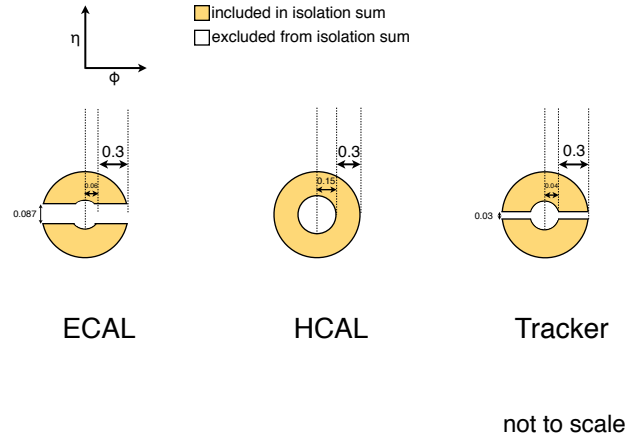


Figure 1.5: ECAL, HCAL, and track Isolation cones.

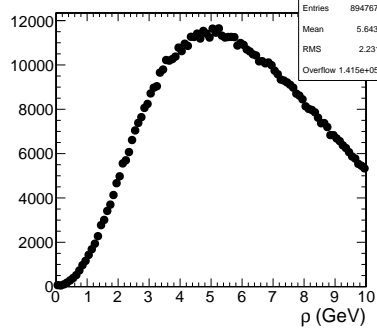


Figure 1.6:  $\rho$  distribution for a sample of events with at least two 25 GeV EM objects passing the  $|\eta|$ ,  $H/E$ , and  $R9$  requirements in Table 1.1, and passing the trigger requirements in Table 1.3. This sample covers the full 2011 dataset.

subtract pileup energy from the ECAL and HCAL isolation cones to recover otherwise clean photons in events with large pileup. The ECAL and HCAL *effective areas* of 0.0792 and 0.0252 (**Change these to the final values used in the analysis.**), respectively, are calculated by fitting the average ECAL or HCAL isolation energy vs.  $\rho$  in a sample of  $Z \rightarrow ee$  events to a straight line. The slope of the line—which has the units of  $\eta \cdot \phi$ , or area—is the effective area.

**Summarize the combined isolation lower bound optimization here.**

The upper bound on fake photon combined isolation guarantees that poorly isolated dijet events, with  $\cancel{E}_T$  resolution dissimilar to the candidate diphoton events, do not enter the  $\cancel{f}\bar{f}$  sample. The exact value of 20 GeV (cf. Table 1.1) arises from a low- $\cancel{E}_T$   $\cancel{f}\bar{f}/\gamma\gamma$   $\chi^2$  optimization procedure [?]. Figure 1.7 shows the value of the Neyman’s  $\chi^2$  between the  $\cancel{f}\bar{f}$  and  $\gamma\gamma$   $\cancel{E}_T$  distributions, truncated at either 25 or 50 GeV, vs. upper bound on fake combined isolation. As shown in the figure, 20 GeV very nearly minimizes the  $\chi^2$ , while also being large enough that a sufficient number of  $\cancel{f}\bar{f}$  events may be collected.

**Summarize the effective area calculation here?**

A “pixel seed” is defined as a hit in the pixel detector consistent with a track extrapolated from the position of the ECAL SC back to the primary vertex. Real photons, having no charge and therefore no bending in the magnetic field, should not

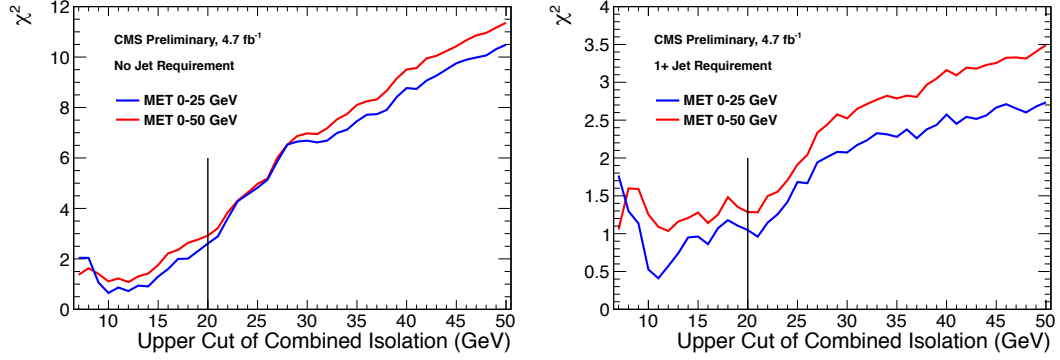


Figure 1.7: Neyman’s  $\chi^2$  between the  $ff$  and  $\gamma\gamma \cancel{E}_T$  distributions, truncated at either 25 (red) or 50 (blue) GeV, vs. upper bound on fake combined isolation. The left plot includes all events; the right plot is for events with  $\geq 1$  jet defined as in Table 1.2, but with the  $\Delta R$  cleaning criteria applied to the two primary EM objects and all additional electrons, photons, and fake photons. The full reweighting and normalization procedure is employed in the  $\cancel{E}_T$  calculation (see Sec. ??). Error bars include statistical, reweighting, and normalization error (see Sec. ??). Reprinted from Fig. 9 of ref. [?].

have a pixel seed.

### 1.1.2 Electrons

Electrons are reconstructed identically to photons, except that in the electron case the presence of a pixel seed is enforced, rather than vetoed.<sup>2</sup> Photons and electrons are defined by very similar criteria so that  $Z \rightarrow ee$  events can be used to model the QCD background in the two-photon sample without introducing any bias in the electron energy measurement (cf. Sec. ??).

### 1.1.3 Jets and Missing Transverse Energy

#### Particle Flow

In this analysis, jets and  $\cancel{E}_T$  are formed from *particle flow* (PF) candidates. The particle flow algorithm [?, ?] uses information from all CMS subdetectors to reconstruct

<sup>2</sup>In many CMS analyses, electrons are reconstructed very differently from photons. In particular, a special tracking algorithm [?] is used to best follow a radiating electron. However, in this analysis, the electron tracking is not used.

as accurately as possible the positions and momenta of all visible jet constituents, exploiting the fine granularity of the tracker and ECAL to achieve a greatly improved momentum resolution over calorimeter-only jets [?]. The PF algorithm is summarized below [?].

1. Reconstruct the fundamental detector objects via iterative procedures
  - Tracks in the inner silicon layers
    - High efficiency and low fake rate for charged hadrons in jets
    - Relaxed primary vertex constraint allows photon conversions, particles originating from nuclear interactions in the silicon, and long-lived particles to be reconstructed
  - Calorimeter clusters
  - Muon tracks in the outer muon layers
2. Create a “block” of linked fundamental objects
  - Link silicon tracks to calorimeter clusters via  $\Delta R_{\text{track-cluster}}$  (account for electron bremsstrahlung)
  - Link clusters in one calorimeter layer to clusters in a separate layer via  $\Delta R_{\text{cluster-cluster}}$
  - Link silicon tracks to muon tracks via global track  $\chi^2$
3. ID the particles in the block
  - If global (silicon + muon layers) muon  $p_T$  is compatible with silicon track  $p_T$ , ID as a muon and remove corresponding tracks from block
  - ID electron tracks via special algorithm and removed all corresponding tracks and cluster from block

- Remove fake tracks from the block
- Remove excess track-cluster links via  $\Delta R$  track – cluster minimization (but allow multiple tracks to be associated to one cluster)
- If the cluster energy is significantly larger than the energy of the linked track, ID as a PF photon or PF neutral hadron and remove corresponding clusters from the block
- If the cluster is not linked to a track, ID as a PF photon or PF neutral hadron and remove corresponding clusters from the block
- Remaining track-cluster links are PF charged hadrons

## Jets

PF candidates are clustered into jets by means of the anti- $k_T$  algorithm with  $R = 0.5$  [?]. In this algorithm, all possible pairs of PF candidates  $i, j$  are looped over, and the momenta of the pair that minimize the distance variable

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2 \max(k_{Ti}^2, k_{Tj}^2)} \quad (1.3)$$

are combined, where  $k_{Ti}$  is the transverse momentum of “combined” PF candidate  $i$ . The constituent PF candidates are clustered together. The process is repeated until  $d_{ij} > 1/k_{Ti}^2$  for all pairs of clustered PF momenta [?]. An illustration is given in Figure 1.8. The anti- $k_T$  algorithm is infrared and collinear safe, leading to well-behaved theoretical predictions and ease of comparison between data and MC simulation. It also tends to form circular jets, making it easy for experimental effects such as expected out-of-cone energy and fiducial acceptance to be measured or simulated. For these reasons, the anti- $k_T$  jet clustering algorithm was chosen for this analysis.

Once jets are clustered, they must be corrected for biases in the energy mea-

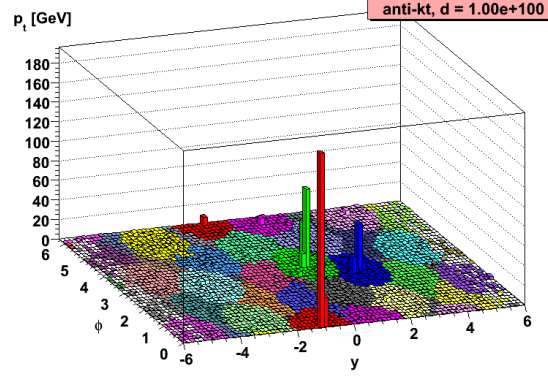


Figure 1.8: Example event display showing jets clustered via the anti- $k_T$  algorithm.  $y$  is pseudorapidity.

surement due to non-compensation [?], invisible energy (lost to overcoming nuclear binding energy, in neutrinos, or in unclustered muons, for example) [?], detector geometry and cracks [?], zero suppression and trigger inefficiencies [?], pileup, and effects of the clustering algorithm [?]. Four multiplicative correction factors are applied to the raw jet four-momentum  $p_\mu^{\text{raw}}$  [?]:

- $C_{\text{offset}}(p_T^{\text{raw}})$ , which accounts for extra energy due to noise, pileup, and the underlying event;
- $C_{\text{MC}}(C_{\text{offset}}p_T^{\text{raw}}, \eta)$ , which is derived from MC and accounts for most of the  $p_T$  and  $\eta$  dependence;
- $C_{\text{rel}}(\eta)$ , which accounts for the remaining differences in uniformity over the entire calorimeter between data and MC; and
- $C_{\text{abs}}(C_{\text{rel}}C_{\text{MC}}C_{\text{offset}}p_T^{\text{raw}})$ , which accounts for the remaining differences in linearity over the full  $p_T$  range between data and MC.

Figure 1.9 shows the total jet energy correction factor  $C_{\text{offset}}C_{\text{MC}}C_{\text{rel}}C_{\text{abs}}$  vs.  $\eta$  for jets reconstructed with the anti- $k_T$  algorithm,  $R = 0.5$ . The PF jet corrections are more uniform across  $\eta$  than those of CALO jets (composed of simple calorimeter towers) or JPT jets (Jet Plus Tracks; composed of calorimeter energies replaced,

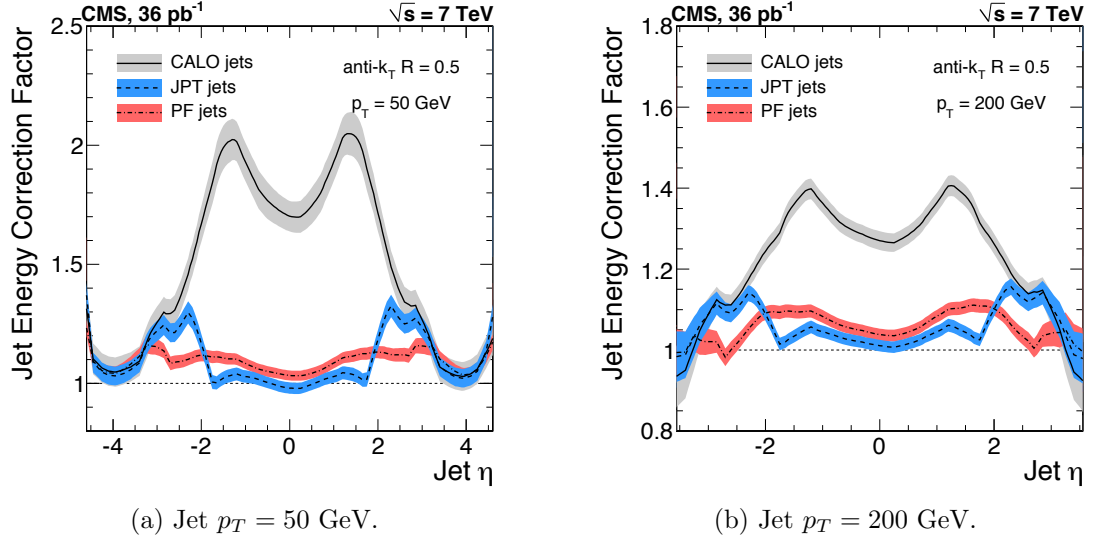


Figure 1.9: Total jet energy correction factor  $C_{\text{offset}}C_{\text{MC}}C_{\text{rel}}C_{\text{abs}}$  vs.  $\eta$ , including uncertainty band, for jets reconstructed with the anti- $k_T$  algorithm,  $R = 0.5$ . Reprinted from Fig. 26 of ref. [?].

where possible, with matching track  $p_T$ ) [?]. In addition, for  $p_T$  in the range 30-200 GeV and  $|\eta|$  up to 2.0, the PF jet energy correction uncertainty is lower than that of the other two types of jets, and never exceeds  $\sim 3\%$  [?]. The superior performance of PF jets motivates their use in this search.

In this analysis, candidate and QCD control events are binned by number of jets satisfying the criteria in Table 1.2.

### Missing Transverse Energy

To be consistent with the jet reconstruction,  $\cancel{E}_T$  in this analysis is also reconstructed from PF candidates. Raw  $\cancel{E}_T$  is defined as

$$\cancel{E}_{T\text{raw}} = \left| - \sum_{i=1}^{n_{\text{PF}}} \vec{p}_{Ti} \right| \quad (1.4)$$

where  $n_{\text{PF}}$  is the number of PF candidates in the event.  $\cancel{E}_{T\text{raw}}$  must be corrected for the same effects that necessitate jet corrections, since  $\cancel{E}_{T\text{raw}}$  is usually the result of jet mis-measurement (except, of course, in electroweak physics processes that include an energetic neutrino, or SUSY production). Therefore, CMS *Type-I*  $\cancel{E}_T$  corrections simply involve replacing the PF jets with their corrected energies (cf. Sec 1.1.3) and recalculating  $\cancel{E}_T$ . Only jets with *electromagnetic fraction* (EMF) below 90% and  $p_T > 20$  GeV are replaced. This ensures that very electromagnetic jets (as well as isolated leptons, which also receive no correction), which consist chiefly of neutral pions and are measured accurately by the ECAL, do not receive a correction derived for jets with a large fraction of their energy in charged hadrons. In addition, the  $p_T$  cut guarantees that jet corrections are only applied where they are known to within a few percent. Since this search looks for excesses in the  $\cancel{E}_T$  distribution typically above 100 GeV of  $\cancel{E}_T$ , it is not very sensitive to the mis-measurement of low- $p_T$  jets.

**Check this statement and add illustrative plots.**

Figure 1.10 shows the  $\sigma$  of a Gaussian fit to the x- and y-components of calibrated  $\cancel{E}_T$  vs. the calibrated PF  $E_T$  sum in a sample of events containing at least two jets with  $p_T > 25$  GeV. Again, PF  $\cancel{E}_T$  outperforms  $\cancel{E}_T$  constructed of calorimeter towers or track-corrected calorimeter deposits.

## 1.2 HLT

From the objects described in Sec. 1.1, four samples of events are formed:

- $\gamma\gamma$  candidate sample, in which the two highest  $E_T$  objects are photons,
- $e\gamma$  control sample, in which the two highest  $E_T$  objects are one electron and one photon,
- $ee$  control sample, in which the two highest  $E_T$  objects are electrons, and



Table 1.2: Definition of HB/HE hadronic jets. **Add a footnote describing the PF electron and PF muon definitions, with references.**

Variable	Cut
Algorithm	L1FastL2L3Residual corrected PF
$p_T$	$> 30$ GeV
$ \eta $	$< 2.6$
Neutral hadronic energy fraction	$< 0.99$
Neutral electromagnetic energy fraction	$< 0.99$
Number of constituents	$> 1$
Charged hadronic energy	$> 0.0$ GeV if $ \eta  < 2.4$
Number of charged hadrons	$> 0$ if $ \eta  < 2.4$
Charged electromagnetic energy fraction	$< 0.99$ if $ \eta  < 2.4$
$\Delta R$ to nearest electron, muon, or one of the two primary EM objects	$> 0.5$

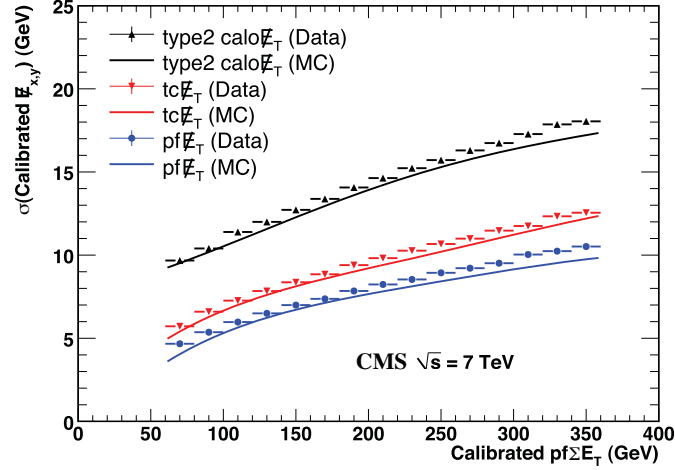


Figure 1.10:  $\sigma$  of a Gaussian fit to the x- and y-components of calibrated  $\cancel{E}_T$  vs. the calibrated PF  $E_T$  sum in a sample of events containing at least two jets with  $p_T > 25$  GeV.  $\sigma$  is calibrated such that the  $\cancel{E}_T$  scale is equal for all three algorithms. PF  $\sum E_T$  is corrected, on average, to the particle level using a Pythia v8 simulation [?]. The blue markers (data) and line (MC) refer to PF jets. Reprinted from Fig. 13 of ref. [?].

- $f\bar{f}$  control sample, in which the two highest  $E_T$  objects are fakes.

In all samples, the leading EM object is required to have  $E_T > 40$  GeV, while the trailing EM object is required to have  $E_T > 25$  GeV. The high level triggers used to select the four samples, by run range, are listed in Table 1.3. No trigger is prescaled.

Table 1.3: HLT paths triggered by the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $f\bar{f}$  samples, by run range. No triggers are prescaled.

Run range	$\gamma\gamma$	$e\gamma$	$ee$	$f\bar{f}$
160404-163261	Photon26_ IsoVL_ Photon18	Photon26_ IsoVL_ Photon18	Photon26_ IsoVL_ Photon18	Photon26_ IsoVL_ Photon18
161216-166967	Photon36_ CaloIdL_ Photon22_ CaloIdL	Photon36_ CaloIdL_ Photon22_ CaloIdL	Photon36_ CaloIdL_ Photon22_ CaloIdL	Photon36_ CaloIdL_ Photon22_ CaloIdL
166347-180252	Photon36_ CaloIdL_ IsoVL_ Photon22_ CaloIdL_ IsoVL	Photon36_ CaloIdL_ IsoVL_ Photon22_ CaloIdL_ IsoVL	Photon36_ CaloIdL_ IsoVL_ Photon22_ CaloIdL_ IsoVL	Photon36_ CaloIdL_ IsoVL_ Photon22_ CaloIdL_ IsoVL  Photon36_ CaloIdL_ IsoVL_ Photon22_ R9Id  Photon36_ R9Id_ Photon22_ CaloIdL_ IsoVL  Photon36_ R9Id_ Photon22_ R9Id

Each piece of the HLT path name is defined as follows.

- “Photon”: Energy deposit in the ECAL that fired an L1 trigger (cf. Sec. ??). For Photon26\_IsoVL\_Photon18, the L1 seed  $E_T$  threshold is 12 GeV, while for all other triggers in Table 1.3 it is 20 GeV.
- Integer following the word “Photon”:  $E_T$  threshold in GeV for offline reconstructed photon, using the full photon reconstruction of Sec. 1.1.1 minus the laser calibrations and assuming the primary vertex at (0, 0, 0).
- “CaloIdL”: For EB photons,  $H/E < 0.15$  and  $\sigma_{i\eta i\eta} < 0.014$ .
- “IsoVL”:  $I_{\text{ECAL}} < 0.012E_T + 6$  GeV,  $I_{\text{HCAL}} < 0.005E_T + 4$  GeV, and  $I_{\text{track}} < 0.002E_T + 4$  GeV.
- “R9Id”:  $R9 > 0.8$ .

In addition, the versions of HLT\_Photon26\_IsoVL\_Photon18 and Photon36\_CaloIdL\_Photon22\_CaloIdL that were active during runs 160404-163268 included a cut  $E_{\text{max}}/E_{5\times 5} < 0.98$  for spike rejection.  $E_{\text{max}}$  is the energy in the highest energy crystal of the EM cluster and  $E_{5\times 5}$  is the energy in the  $5\times 5$  crystal matrix around the seed crystal. For runs after 163268, Swiss cross spike rejection of individual crystals from HLT quantities was performed (cf. Sec. 1.1.1). All information about the evolution of the CMS HLT settings can be found in the HLT configuration browser at <http://j2eeps.cern.ch/cms-project-confdb-hltdev/browser/>.

As an example of the naming convention just described, the HLT path Photon36\_CaloIdL\_IsoVL\_Photon22\_R9Id is fired if one photon is found with  $E_T > 36$  GeV passing the CaloIdL and IsoVL requirements, and another is found with  $E_T > 22$  GeV passing the R9Id requirement.

### 1.3 Event Quality

To suppress instrumental backgrounds, a set of event quality cuts are applied to the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $ff$  samples. First, all events are required to pass a good run selection, as determined by the CMS Physics Validation Team (<https://twiki.cern.ch/twiki/bin/view/CMS/PVTMain>, CERN computing ID needed). The good run selection excludes luminosity sections during which a sufficient part of the CMS detector was unpowered or malfunctioning. Such conditions could occur if, for example, a high voltage supply trips off in the middle of a run, or a DAQ error corrupts data quality but is not spotted until after the data have been collected. The severity of a detector problem is judged by its effect on a wide range of analyses and reconstruction algorithms. Of the  $\text{N fb}^{-1}$  of integrated luminosity delivered by the LHC in 2011,  $4.68 \text{ fb}^{-1}$  passed the good run selection. This analysis is performed on the entire 2011 certified dataset.

Second, all events must contain at least one good interaction vertex. The criteria for a good vertex are:

- $\chi^2 \neq 0 \parallel \text{ndof} \neq 0 \parallel N_{\text{tracks}} \neq 0$ , where  $\chi^2$  and ndof are calculated for the track fit to the vertex, and  $N_{\text{tracks}}$  is the number of tracks in the vertex fit
- $\text{ndof} > 4$
- $|z| < 24 \text{ cm}$ , where  $z$  is the  $z$ -coordinate of the vertex position
- $|\rho| < 2 \text{ cm}$ , where  $\rho$  is the transverse displacement of the vertex position from the beam line

The good vertex requirement eliminates non-collision backgrounds such as beam scraping, beam halo, cosmic muon interactions, and instrumental effects.

Third, the two electromagnetic objects in the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $ff$  events must be separated in  $\phi$  by at least 0.05. This requirement protects against beam halo

bremsstrahlung, in which a halo muon traveling parallel to the beam line radiates an energetic photon while itself depositing a large amount of energy in the ECAL. In this case, the two ECAL hits would likely be at the same  $\phi$  (and  $\rho$ ). An example cartoon is shown in Figure ?? **Insert figure.**

Fourth, the two EM objects must be separated in  $R$  by at least 0.6. Since the isolation cone size used is 0.3, this ensures that the isolation energy of one EM object cannot be in the veto strip (Fig. 1.5) of the other.

Finally, the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $ff$  events must pass an HCAL noise filter and ECAL dead channel filter. The HCAL noise filter guarantees that all HCAL reconstructed hits are inconsistent with any noise source. Noise sources [?] include:

- Ion feedback in the HPDs absent any true incident photons, in which a thermal electron ionizes a molecule in the HPD acceleration gap, faking a real signal
- HPD discharge affecting nearly all channels in the same HPD [?], partially explained by the effect of the 4 T CMS magnetic field on the flashover voltage of the dielectric [?]
- Concurrent signals in nearly all 72 channels of a single RBX, as yet unexplained
- HF PMT window hits (as opposed to the usual quartz fiber hits)
- ADC saturation

Since HCAL noise may induce fake jets or  $\cancel{E}_T$ , events are rejected if any of the following criteria are true:

- Any HPD has  $> 17$  hits
- A single HPD has  $> 10$  hits, but every other HPD has zero hits
- An RBX has  $> 10$  zero-ADC-count hits

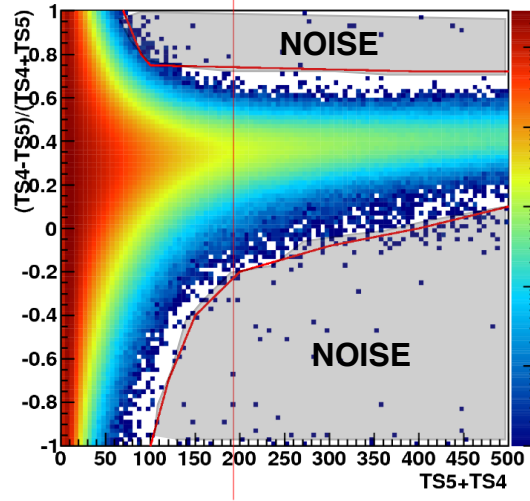


Figure 1.11:  $(TS4 - TS5)/(TS4 + TS5)$  vs.  $TS4 + TS5$  for a minimum bias sample. HB/HE hits are considered noisy if they lie in the sparsely populated gray region labeled "NOISE" defined by the curved red lines. Adapted from ref. [?].

- Any HB/HE reconstructed hit corresponding to an RBX with  $> 50$  GeV of energy fails a two-dimensional cut defined by the variables  $(TS4 - TS5)/(TS4 + TS5)$  vs.  $TS4 + TS5$ , where  $TS4(TS5)$  is the hit amplitude in the fourth(fifth) time sample read out for that hit. The cut is defined in Fig. 1.11.

The ECAL dead channel filter is designed to flag events in which significant EM energy was deposited in a masked region of the ECAL by using the trigger primitive information for the corresponding trigger tower. Energy deposited in a masked region of ECAL can cause fake  $\cancel{E}_T$ . Events are rejected if the trigger primitive  $E_T$  exceeds the maximum value of 63.75 GeV in any trigger tower that is masked in the readout.

## 1.4 Photon Identification Efficiency

In order to determine the cross section (or cross section upper limit) for a GGM signal, the photon identification efficiency is needed. Since no suitably large sample of  $Z \rightarrow \mu\mu\gamma$  events in CMS exists yet, the efficiency calculation relies on the similarity between detector response to electrons and photons. A scale factor to correct the MC

photon ID efficiency to the real photon efficiency for the data is obtained from the ratio of the electron efficiency from the data to the electron efficiency from MC.

The different types of photon ID variables—calorimeter and track isolation, ratio of hadronic to electromagnetic energy of the shower, and transverse shower shape—are chosen so that their distributions for isolated electrons and photons are similar.<sup>3</sup> Figure ?? shows distributions of photon ID variables for MC electrons and photons, where a MC electron(photon) is a `reco::Photon` object matched to a generated electron(photon) originating from a  $Z$  decay(the hard interaction) within  $\Delta R = 0.3$ . MC electrons come from the `DYJetsoLL` sample, while MC photons come from the `GJet` sample (see Appendix ??). **Make these plots.** The shapes of the distributions agree well. **Also include data/MC comparisons?**

The photon selection efficiency is

$$\epsilon_\gamma = \epsilon_\gamma^{\text{MC}} \times \frac{\epsilon_e^{\text{data}}}{\epsilon_e^{\text{MC}}} \quad (1.5)$$

where

- $\epsilon_\gamma$  is the photon ID efficiency in data,
- $\epsilon_\gamma^{\text{MC}}$  is the photon ID efficiency in MC,
- $\epsilon_e^{\text{data}}$  is the electron ID efficiency obtained using  $Z \rightarrow ee$  electrons in the data that satisfy the photon ID cuts, and
- $\epsilon_e^{\text{data}}$  is the electron ID efficiency obtained using  $Z \rightarrow ee$  electrons in MC that satisfy the photon ID cuts.

The ratio  $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}}$  is defined as the scale factor by which the GGM signal MC photon ID efficiency must be multiplied to give an estimate of the photon ID efficiency

---

<sup>3</sup> $R9$  differs between photons and radiating electrons, but the requirement  $R9 < 1$  is loose enough not to introduce problems with the use of electrons to measure the photon ID efficiency.

in data. The photon ID requirements of Table 1.1 plus the IsoVL HLT requirement described in Sec. 1.2 and Table 1.3 are repeated in Table 1.4.

Table 1.4: Candidate photon ID requirements.

Variable	Cut
$I_{\text{ECAL}}$	$< 0.012E_T + 6 \text{ GeV}$
$I_{\text{HCAL}}$	$< 0.005E_T + 4 \text{ GeV}$
$I_{\text{track}}$	$< 0.002E_T + 4 \text{ GeV}$
$H/E$	$< 0.05$
$\sigma_{i\eta i\eta}$	$< 0.011$
$I_{\text{ECAL}} - 0.0792\rho + I_{\text{HCAL}} - 0.0252\rho + I_{\text{track}}$	$< 6 \text{ GeV}$
$R9$	$< 1$

#### 1.4.1 Tag and Probe Method

A *Z tag and probe* method is utilized to measure the efficiency of the photon ID cuts in Table 1.1. The tag is a well-identified electron. The probe, by contrast, is as loosely identified as possible, and all tags must pass the probe criteria in addition to the stringent tag criteria. The tag and probe criteria used in this study are shown in Table 1.5.

The invariant mass of the tag and probe are required to be within a narrow window around  $Z$  mass. Assuming that the probabilities of the tag and probe legs of the  $Z$  decay to pass the photon ID cuts are uncorrelated, the efficiency can be estimated as

$$\epsilon = \frac{N_{\text{tag-pass}}}{N_{\text{tag-pass}} + N_{\text{tag-fail}}} \quad (1.6)$$

where  $N_{\text{tag-pass}}$  is the number of tag-probe pairs in which the probe leg passes the photon ID cuts under study and  $N_{\text{tag-fail}}$  is the number of tag-probe pairs in which the probe leg fails the cuts. Implicit in these definitions is a double counting of pairs in which both electrons pass the tag and probe criteria [?]. In addition, in the rare



Table 1.5: Tag and probe criteria. The superscript 0.4 indicates that the isolation variable was calculated in a cone of  $\Delta R = 0.4$  around the photon candidate. The isolations without superscripts use the standard  $\Delta R = 0.3$  cones.

Variable	Cut	
	Tag	Probe
RECO object	photon	photon
HLT	HLT_Ele17_CaloIdVT_CaloIsoVT_TrkIdT_TrkIsoVT_SC8_Mass30_v* (must have fired the 17 GeV leg)	—
$H/E$	$< 0.05$	$< 0.15$
$I_{\text{ECAL}}^{0.4}$	$< 0.006E_T + 4.2 \text{ GeV}$	—
$I_{\text{HCAL}}^{0.4}$	$< 0.0025E_T + 2.2 \text{ GeV}$	—
$I_{\text{track}}^{0.4}$	$< 0.001E_T + 2.0 \text{ GeV}$	—
$E_T$	$> 25 \text{ GeV}$	—
SC $E_T$	—	$> 15 \text{ GeV}$
SC $ \eta $	$< 1.4442$	$< 1.4442$
$\sigma_{i\eta i\eta}$	$< 0.009$	—
Has pixel seed?	Yes	—
Track match type	General track	—
Track match $\Delta R$	$< 0.04$	—
Track match $p_T$	$> 15 \text{ GeV}$	—
Track match $ \eta $	$< 1.479$	—

circumstance (less than 1% in MC [?]) that two or more probes may be matched to one tag, the pair with invariant mass closest to the  $Z$  mass is chosen.

In practice,  $N_{\text{tag-pass}}$  and  $N_{\text{tag-fail}}$  are returned by a simultaneous unbinned maximum likelihood fit to the invariant mass distributions of tag-pass and tag-fail events, with appropriate signal and background PDF assumptions. The fit form used is

$$\begin{aligned} f_{\text{tag-pass}}(m_{\text{tag-pass}}) &= \epsilon N_S f_S^{\text{pass}}(m_{\text{tag-pass}}) + N_B^{\text{pass}} f_B^{\text{pass}}(m_{\text{tag-pass}}) \\ f_{\text{tag-fail}}(m_{\text{tag-fail}}) &= (1 - \epsilon) N_S f_S^{\text{fail}}(m_{\text{tag-fail}}) + N_B^{\text{fail}} f_B^{\text{fail}}(m_{\text{tag-fail}}) \end{aligned} \quad (1.7)$$

where  $f_{\text{tag-pass}}(m_{\text{tag-pass}})$  and  $f_{\text{tag-fail}}(m_{\text{tag-fail}})$  are the tag-pass and tag-fail PDFs, respectively;  $\epsilon$  is the efficiency;  $N_S$  is the total number of  $Z$  signal events summed over both samples;  $f_S^{\text{pass}}(m_{\text{tag-pass}})$  and  $f_S^{\text{fail}}(m_{\text{tag-fail}})$  are the tag-pass and tag-fail signal PDFs, respectively;  $N_B^{\text{pass}}$  and  $N_B^{\text{fail}}$  are the numbers of background events in the tag-pass and tag-fail samples, respectively; and  $f_B^{\text{pass}}(m_{\text{tag-pass}})$  and  $f_B^{\text{fail}}(m_{\text{tag-fail}})$  are the tag-pass and tag-fail background PDFs, respectively. This particular implementation of the tag and probe methodology is based on tag CMSSW\_4\_2\_5 of the CMSSW package `PhysicsTools/TagAndProbe`, and uses the MINUIT2 [?] library, as coded in RooFit [?], for the likelihood maximization. For this study, CMSSWv4.2.8 was used.

For both samples, the signal shape is assumed to be a Crystal Ball function [?] convoluted with the  $Z$  generated lineshape, while the background shape is a PDF that describes the falling background as well as the kinematic turn-on at low invariant mass. The background PDF, called “RooCMSShape” [?], is given by

$$\text{RooCMSShape}(x) = \begin{cases} 1e20 & \text{for } (x - \mu)\gamma < -70 \\ 0 & \text{for } (x - \mu)\gamma > 70 \\ \text{erfc}((\alpha - x)\beta) \exp(-(x - \mu)\gamma) & \text{otherwise} \end{cases} \quad (1.8)$$

where  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\mu$  are parameters of the fit, most of which are held fixed. In the three lowest  $E_T$  bins, all parameters of the tag-pass and tag-fail background PDFs are left floating, because the effect of the relaxed  $E_T$  cuts has a significant effect on the background shape. More details of the signal and background PDFs are given in Table 1.6. The fixed signal and background parameter values were determined by fitting a small sample ( $0.0 \leq \eta < 0.25$ ) of Fall11 MC signal (DYJetsToLL) and background (QCD\_Pt-20to30\_BCtoE, QCD\_Pt-30to80\_BCtoE, QCD\_Pt-80to170\_BCtoE, GJet\_Pt-20\_doubleEMEnriched, WJetsToLNu, TTJets) with parameters left floating.

Some fit examples are shown in Figures ?? and ?. In Fig. ??, which shows fits to data and MC for  $15 \text{ GeV} \leq \text{probe } E_T \leq 20 \text{ GeV}$ , the kinematic turn-on is below the invariant mass range covered by the plot. The exponentially falling background is easily seen underneath the signal, and is especially pronounced in the background-dominated tag-fail sample. **Include these plots.**

#### 1.4.2 Photon Efficiency Scale Factor $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}}$

Figure 1.12 shows the dependence of the photon ID efficiency scale factor  $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}}$  on  $E_T$ ,  $\eta$ ,  $\Delta R_{\gamma\text{jet}}$ , and  $N_{\text{jet}}$ , where jets are defined as in Sec. 3.5.2 (**insert reference to “Jet Selection” section here**).  $\Delta R_{\gamma\text{jet}}$  **plot will be added during second round**. Errors are statistical only. For the plot of the scale factor vs.  $\Delta R_{\gamma\text{jet}}$ , only events with  $\geq 1$  jets are included. There no significant dependence of the scale factor on these variables, so only one scale factor is computed from the entire dataset.

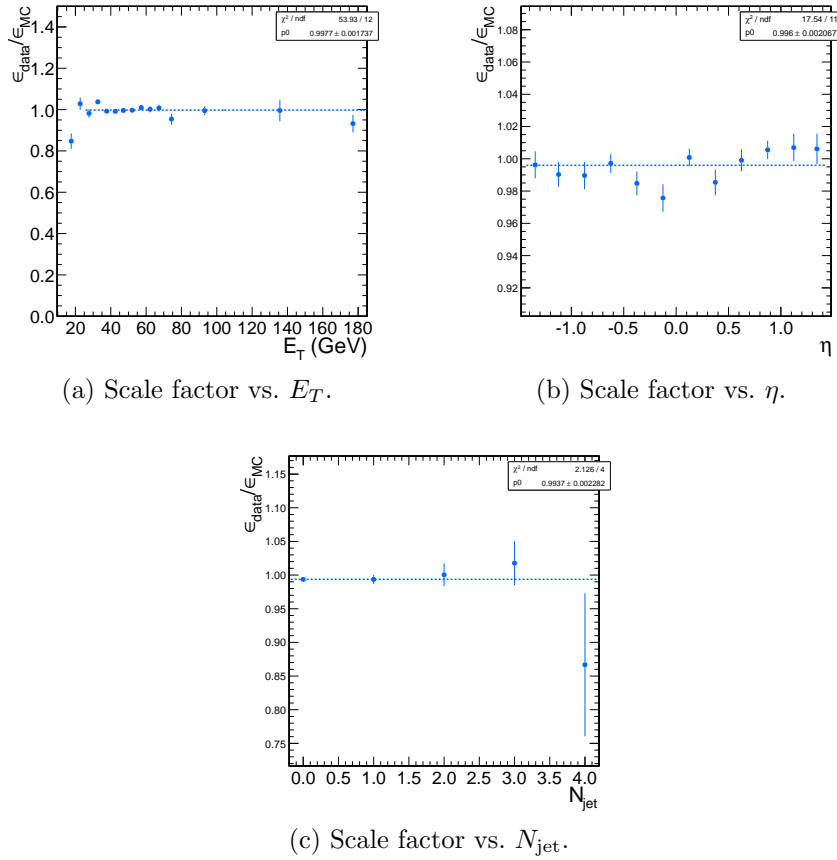


Figure 1.12: Dependence of the photon ID efficiency scale factor on some kinematic variables. Errors are statistical only.

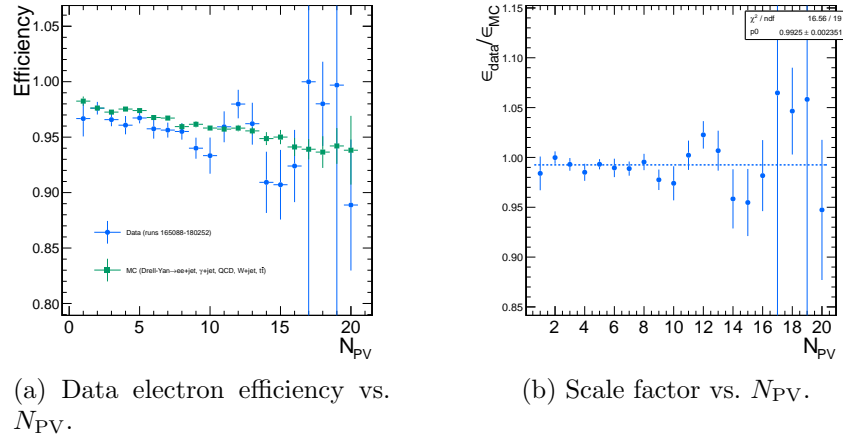


Figure 1.13: Dependence of the photon ID efficiency scale factor on the number of primary vertices per event. Errors are statistical only.

The effect of pileup is studied by comparing the efficiencies  $\epsilon_e^{\text{data}}$  and  $\epsilon_e^{\text{MC}}$  vs. the number of primary vertices ( $N_{PV}$ ) in the event. The efficiency drops for events with large  $N_{PV}$ , even after using pileup-corrected isolation cuts, as can be seen in Figure 1.13a. However, the MC tracks the data, and the scale factor is flat vs.  $N_{PV}$ , as shown in Fig. 1.13b.

The scale factor is measured to be  $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}} = 0.994 \pm 0.002(\text{stat.}) \pm ?(\text{syst.})$ . Four main sources of systematic error, in addition to the statistical error of 0.2%, were studied.

#### Different behavior of electrons and photons in MC

Even though the photon ID cuts are designed to be similarly efficient for both electrons and photons, there might be a small difference in the performance between the two kinds of particles, e.g. because of electron bremsstrahlung. To check this effect, the MC electron ID efficiency was calculated using a  $Z \rightarrow ee$  sample and the MC photon ID efficiency was calculated using a  $\gamma + \text{jets}$  sample. Both samples were reconstructed in CMSSWv3.6. Half the difference between these two results, 0.5%, was taken as an error on the scale factor. **This result is now out of date because it used an older selection. Should this check be redone?**

**Pileup** To account for the possibility that the MC simulation may not adequately reproduce the data in a high pileup environment, the data/MC scale factor for events with 1-4 good reconstructed primary vertices was calculated, along with the same for events with  $\geq 7$  good reconstructed primary vertices. This particular division of the data was chosen because the pileup distribution in data peaks at  $\sim 7$  primary vertices per event. The difference between the scale factors from both samples,  $?$ , was taken as an error on the scale factor from pileup.

**Signal fit over/underestimation** It was found that the signal fit slightly underestimates the data in the tag-pass sample, and slightly overestimates it in the tag-fail sample. To cover this effect with a systematic error, the efficiencies in data and MC, and then the scale factor, were recalculated using the background (from fit) subtracted integrals of the tag-pass and tag-fail distributions, rather than the fitted signal yields in those distributions. The difference between the scale factor found in this way and the nominal scale factor,  $?$ , was taken as an error on the scale factor. **Not sure this is a problem anymore.**

**Signal and background shape assumption** The largest source of systematic error comes from the signal and background shape assumptions. To assess the magnitude of this error, the tag-pass and tag-fail tail parameters (Crystal Ball  $n$  and  $n$ ) were varied by  $\pm 1\sigma$ , and the background shape was varied between “RooCMSShape, exponential, power law, and quadratic. All possible combinations of varied parameters were generated, and the data and MC were refit and new scale factors generated according to those combinations. The error was taken as the largest deviation of the scale factor from nominal,  $?$ %. **Also, in many bins, there is zero background fitted under the tag-pass peak. Background is a small effect here anyway, but we should get a sys-**

tematic for this somehow.

Table 1.6: Parameter values for the signal and background PDFs for the different samples. The background PDF applies to all efficiency bins except the four lowest  $E_T$  bins, which use a floating RooCMSShape background. When a bracketed range is given, the parameter is allowed to float within that range. When a constant is given, the parameter is fixed to that constant.

PDF	Crystal Ball fit parameters				RooCMSShape fit parameters			
	$\mu$	$\sigma$	$\alpha$	n	$\mu$	$\alpha$	$\beta$	$\gamma$
Tag-pass signal	[-1.0, 1.0]	[1.0, 3.0]	0.87	97.0	N/A	N/A	N/A	N/A
Tag-fail signal	[-1.0, 1.0]	[1.0, 3.0]	0.73	134.9	N/A	N/A	N/A	N/A
Tag-pass back-ground	N/A	N/A	N/A	N/A	65.0	61.949	0.04750	0.01908
Tag-fail back-ground	N/A	N/A	N/A	N/A	$\alpha$	[50.0, 100.0]	0.065	0.048