Shining Light on Dark Matter,
One Photon at a Time

by

Brandon Leigh Allen

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctorate of Science in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Author . . . .
Department of Physics
May 18, 2019

Certified by. . . . .
Christoph E.M. Paus
Professor
Thesis Supervisor

Accepted by . . . .
Nergis Mavalvala
Associate Department Head for Education

Shining Light on Dark Matter,

One Photon at a Time

by

Brandon Leigh Allen

Submitted to the Department of Physics
on May 18, 2019, in partial fulfillment of the
requirements for the degree of
Doctorate of Science in Physics

Abstract

A search is conducted for new physics in final states containing a photon and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. The data collected by the CMS experiment at the CERN LHC correspond to an integrated luminosity of 35.9 inverse femtobarns. No deviations from the predictions of the standard model are observed. The results are interpreted in the context of dark matter production and limits on new physics parameters are calculated at 95% confidence level. For the two simplified dark matter production models considered, the observed (expected) lower limits on the mediator masses are both $950\,(1150)$ GeV for 1 GeV dark matter mass.

Thesis Supervisor: Christoph E.M. Paus
Title: Professor

Acknowledgments

This is the acknowledgements section. You should replace this with your own acknowledgements.

Contents

Chapter 1

Reconstruction

How do we turn electrical signals into physics.

1.1 Tracks

1.2 Primary Vertex

1.3 Particle Flow

1.4 Photons

We select high-E_T photons from the ECAL Barrel. First, we apply the following cuts to select high-E_T photon candidates:

- Super cluster $E_\mathrm{T} > 175\,\mathrm{GeV}$

- Super cluster $|\eta| < 1.4442$.

We use the uncorrected E_T of the supercluster as photon E_T. The use of supercluster raw E_T is motivated by an observation that this energy correction causes photon objects with large cluster shower width to exhibit unphysical energies. Figure 1-1 is a profile of the magnitude of the energy correction in bins of $\sigma_{i\eta i\eta}$. As an illustration, an unphysically large correction is causing the transverse momentum of the photon

object in the event shown in Fig. 1-2 to be nearly twice as large as the $E_{\mathrm{T}}^{\mathrm{miss}}$, which is supposed to balance the visible, i.e., photon momentum. Photon objects with wide showers are used to estimate the hadron-to-photon misidentification background, while the photon energy resolution has an insignificant effect. Therefore, unbiased supercluster energy was chosen over the corrected photon energy.

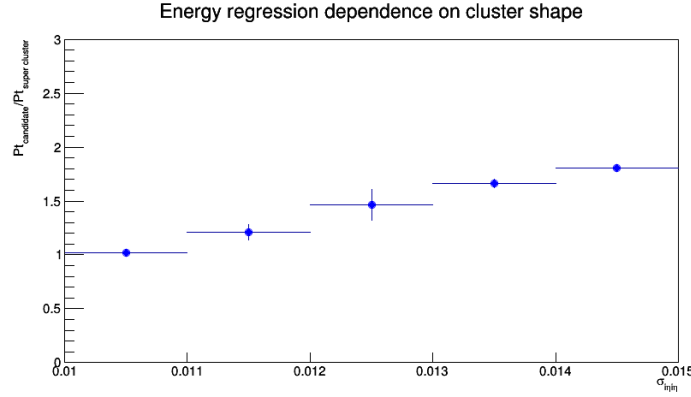Figure 1-1: Magnitude of the energy correction on the photon object in bins of $\sigma_{i\eta i\eta}$.
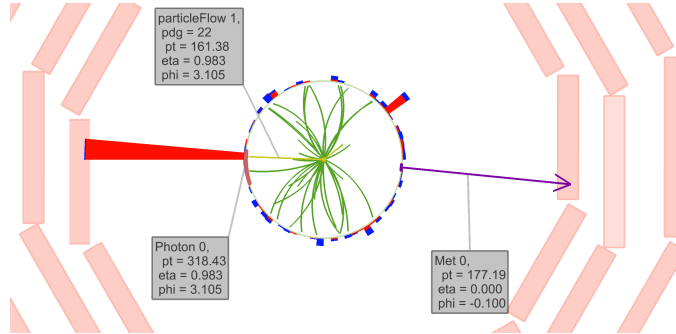

Figure 1-2: An example event where a photon with a wide shower receives a large energy correction.

To reduce hadron-to-photon misidentification rate, we apply the following collection of isolation and shower shape cuts, which will hereby be referred to as the ID:

- $H/E < 0.0260$

- $\sigma_{i\eta i\eta} < 0.01040$

10

- ρ-corrected maximum PF charged hadron isolation $I_{\mathrm{CH}}^{\max} < 1.146\,\mathrm{GeV}$

- ρ-corrected PF neutral hadron isolation $I_{\mathrm{NH}} < 2.792\,\mathrm{GeV} + 0.0112 \times E_{\mathrm{T}}^{\gamma} + 0.000028 \times (E_{\mathrm{T}}^{\gamma})^2 / \mathrm{GeV}$

- ρ-corrected PF photon isolation $I_{\gamma} < 2.176\,\mathrm{GeV} + 0.0043 \times E_{\mathrm{T}}^{\gamma}$

In the identification criteria, the maximum PF charged hadron isolation $I_{\mathrm{CH}}^{\max}$ is the maximum of the standard PF charged hadron isolation computed for all reconstructed vertices. Standard PF charged hadron isolation is computed with respect to the primary vertex. Since the object with the highest p_{T} in the selected events is typically a photon, which has no intrinsic association to a vertex, it is possible that the identified primary vertex does not correspond to the interaction from which the photon object originate. In such cases, the photon object can be surrounded by charged hadrons, i.e., a part of a jet, and still appear isolated under the standard charged hadron isolation. The use of maximum isolation is a conservative measure to address such misidentification.

Effective areas for isolations are also recomputed to maintain flat pileup dependence as given in Table 1.1

Table 1.1: Effective areas for isolations.

Isolation	$|\eta| < 1.0$	$1.0 < |\eta| < 1.479$
maximum PF charged hadron isolation	0.01064	0.1026
PF neutral hadron isolation	0.0597	0.0807
PF photon isolation	0.1210	0.1107

To reject electrons from the candidate sample, we require that no electron track seeds can be associated to the super cluster. This is known as the pixel seed veto.

To clean the candidate sample from photon objects originating from non-collision sources. we apply the following collection of cuts, which combined with the pixel seed veto constitutes the γ ID:

- Beam halo tagger $E_{\mathrm{MIP}} < 4.9\,\mathrm{GeV}$

- $\sigma_{i\eta i\eta} > 0.001$

11

- $\sigma_{i\phi i\phi} > 0.001$

- Cluster seed time $|t_{\text{seed}}| < 3$.

Beam halo tagger is the total energy deposited in ECAL by a hypothetical beam halo muon that passes through the photon cluster. Lower bounds for $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$ as well as the requirement on the cluster seed time are employed to reject spurious photon objects arising from "ECAL spikes", or anomalous electronic signal induced at the ECAL Barrel photodetectors by nuclear or ionizing interactions in the photocathode.

1.5   Electrons

1.6   Muons

1.7   Hadrons and jets

1.7.1   Hadronic taus

1.8   Missing Transverse Energy

Chapter 2

Calibration

How good is the reconstruction.

2.1 Trigger Efficiency

2.2 Photon Identification Efficiency

When measuring the photon efficiency scale factor, we split the photon ID described
in Section 1.4 into two parts, which we call the e/γ portion and the γ portion. The
e/γ portion of the ID consists of the H/E, $\sigma_{i\eta i\eta}$, and PF isolation cuts and is measured
using the "tag-and-probe" (TP) method as these variables have similar efficiencies for
physical electrons and photons. The γ portion of the ID consists of the pixel seed
veto. We measure the efficiency of γ portion on a sample of physical photons using a
$\sigma_{i\eta i\eta}$ template fit method.

We perform both efficiency estimates as a function of p_T with the binning [175,200],
[200,250], [250,300], [300,350], [350,400] and [400,∞). This binning was chosen based
on the number of available events in data for the failing probes fit in the TP method
and the background template for the $\sigma_{i\eta i\eta}$ fits.

2.2.1 e/γ ID Efficiency

The efficiency corresponding to the e/γ part of the photon ID is estimated by exploiting the Z boson decay into an e^+e^- pair. In this "tag-and-probe" (TP) method, a high-quality electron object (tag) is identified in a single photon data sample, and the accompanying electron is sought for in the pool of elecromagnetic objects (probes) in the event. The area of the peak in the mass distribution of the tag-probe system around the Z boson mass (between $81\,\mathrm{GeV}$ and $101\,\mathrm{GeV}$) is then measured once applying the $e\gamma$ ID requirements on the probe and once inverting the requirements. Denoting the two areas N_{pass} and N_{fail}, respectively, the resulting effiency $\epsilon_{e/\gamma}$ is given by

$$\epsilon_{e/\gamma} = \frac{N_{\mathrm{pass}}}{N_{\mathrm{pass}} + N_{\mathrm{fail}}}. \tag{2.1}$$

The TP measurement is performed on a subset of the single photon triggered events where there is an electron object (tag) passing the "tight" identification criteria in addition to the triggering photon (probe). All possible tag-probe combinations are considered; if the tag object can also serve as a probe and the probe object as a tag, which is a common occurrence in the case when the probe is electron-like (passes the $e\gamma$ ID), then the two combinations are considered independently to avoid the bias caused by somehow preferring one object over another to use as the probe.

The tag-probe mass distributions are then fit to extract N_{pass} and N_{fail}. The fit model is composed of two templates, where one template describes a pure $Z \rightarrow ee$ line shape and the other describes the background contributions. The backgrounds to the fits include W+jets, diboson, and $t\bar{t}$ productions, which are all negligible and estimated to contribute by less than 1%. Minor contribution from processes that do not involve true electrons, such as diphoton production with a strongly asymmetric conversion on one of the photons and misidentification of a QCD jet as an electron, are predicted to be negligible from MC studies.

The $Z \rightarrow ee$ template is given by an analytic shape of Breit-Wigner distribution convoluted with the Crystal Ball function. The mass and width parameters of Breit-Wigner distribution are fixed to PDG values. Crystal Ball parameters are allowed

to float in the fit. It is well known that Breit-Wigner distribution usually does not describe the mass distribution well when the tag or the probe is under kinematically exclusive selections. However, at this high probe p_T scale, selected events are mostly of Z+jets topology with a boosted Z boson, which makes the selection rather inclusive in terms of the tag-probe invariant mass.

The background template is taken from events collected by the single photon trigger where an additional muon object is present, making use of the fact that the most of the background processes in both fits are symmetric in lepton flavor. In order to mitigate the statistical fluctuation in the background sample, the actual template is constructed by a Gaussian kernel estimation of the mass distribution of this muon-probe sample.

The floating parameters of the fits are therefore the normalizations of the $Z \rightarrow ee$ and backgroud templates and the Crystal Ball smearing parameters. Selected example fits are shown in Figure 2-1.

The statistical uncertainty of the fits is estimated by generating toy data from the nominal fit result with the same number of entries as the fit target distribution. The mass distribution of the toy data is then fit with the same model with the parameters floating. This procedure is repeated 100 times to obtain a distribution of the $Z \rightarrow ee$ event yields, and its standard deviation is taken as the statistical uncertainty of the fit. Relative statistical uncertainty on the efficiency is 10%.

To estimate the effect of potential mismodeling in the fits, alternative fits varying the background and signal templates are performed first. In the alternative-background fit, a simple linear function is tested. In the alternative-signal fit, no Crystal Ball convolution is performed to the signal template and the mass and width of the Breit-Wigner function are allowed to vary. Resulting best-fit distributions of these alternative models are then used to generate a large number of toy distributions, which are fit by the nominal model. The average shift of the fit result from the nominal value is then taken as the uncertainty. The relative uncertainty on the efficiency varies from 2 to 4% depending on the probe p_T bin.

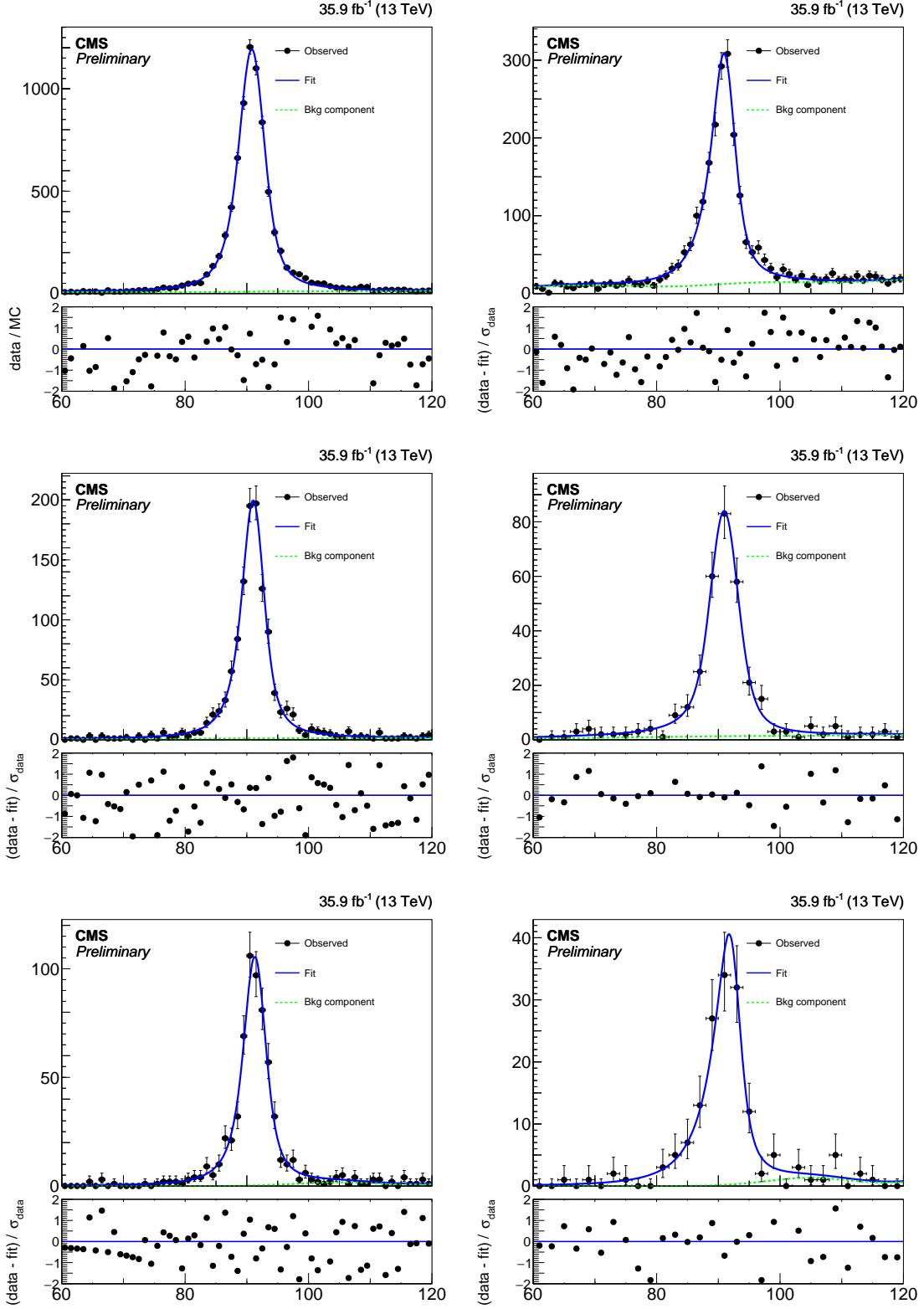The MC efficiency is taken from counting the number truth-matched electrons

Figure 2-1: Fits to the mass distributions for pass (left) and fail (right) selections, in bins of probe p_T: $175 < p_T < 200\,\mathrm{GeV}$ (top), $300 < p_T < 350\,\mathrm{GeV}$ (middle), $p_T > 400\,\mathrm{GeV}$ (bottom). The blue solid line represents the full fit model, and the green dashed line its background component.

16

passing and failing the e/γ part of the ID from a $Z \to ee$ sample. Additionally, the MC efficiency is computed using the same procedure as in data as a cross-check. The two methods are consistent within their uncertainties.

The data efficiencies, MC efficiencies, and resulting scale factors as a function of p_T are shown in Figure 2-2. The scalefactors are consistent with unity within the uncertainties. The numerical values are given in Table 2.1. We use the bin by bin scale factor corresponding to the truth values in the analysis.
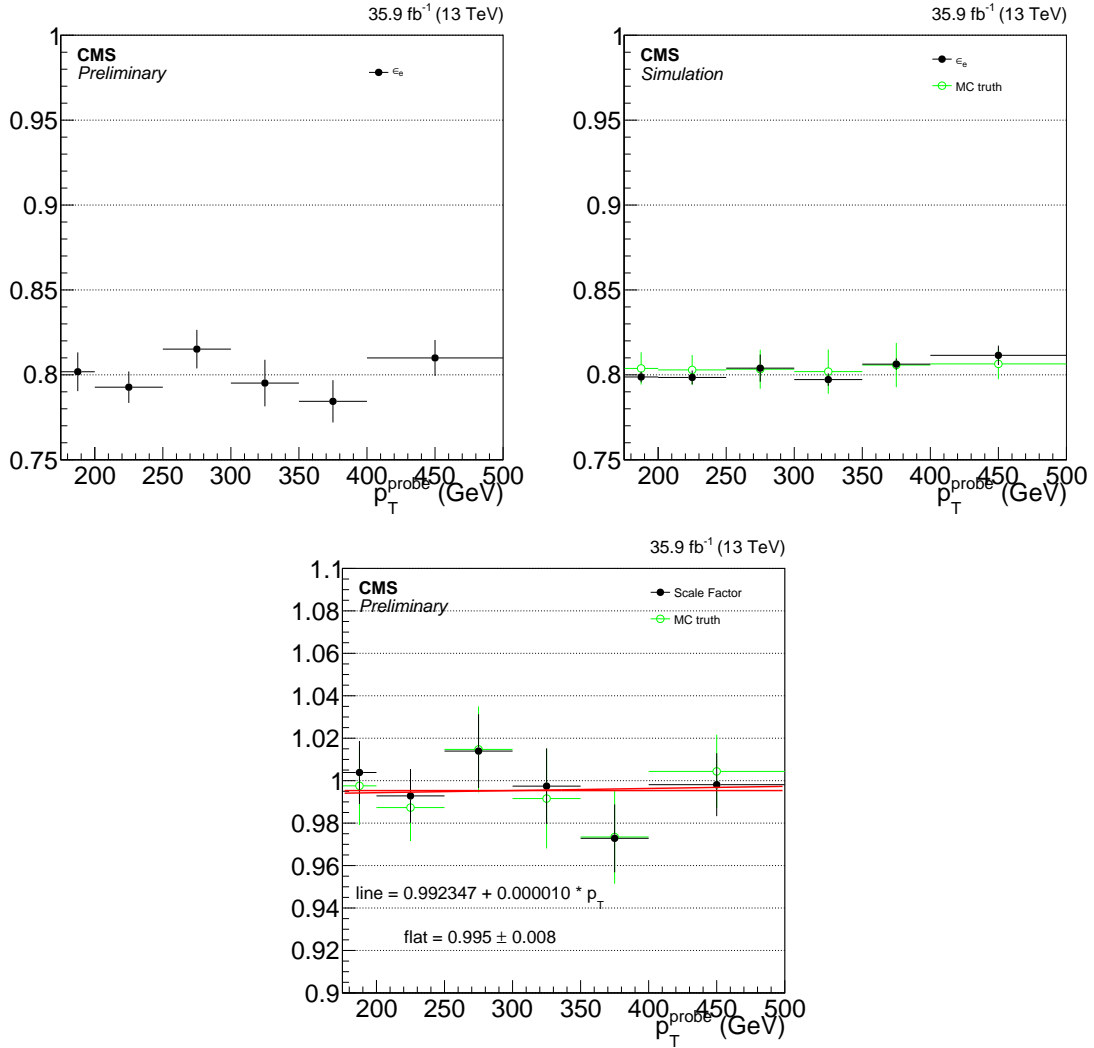

Figure 2-2: e/γ component of the photon identification efficiency for data (top-left) and MC (top-right) and corresponding scale factor (bottom) as a function of photon p_T.

Table 2.1: e/γ scale factors as a function of photon p_{T}.

$p_{\mathrm{T}}^{\mathrm{probe}}$ (GeV)	MC Fit	Truth
(175, 200)	1.014 ± 0.008	1.009 ± 0.016
(200, 250)	1.003 ± 0.008	0.999 ± 0.014
(250, 300)	1.014 ± 0.010	1.016 ± 0.019
(300, 350)	1.002 ± 0.014	0.997 ± 0.022
(350, 400)	0.986 ± 0.012	0.987 ± 0.022
(400, 6500)	0.988 ± 0.011	0.999 ± 0.016

2.2.2 γ-specific ID Efficiency

To measure the efficiency of the γ-specific component of the photon ID, we use a $\sigma_{i\eta i\eta}$ template fit to extract the number of true photons from a pool of photon objects passing the e/γ ID.

The measurement is performed over a EM object+jet control sample where we require one jet passing with $p_{\mathrm{T}} > 100\,\mathrm{GeV}$ and $|\eta| < 2.5$ which passes the loose jet ID. The EM object passes the e/γ ID with the execption of the following relaxed cuts:

- $\sigma_{i\eta i\eta} < 0.015$

- $I_{\mathrm{CH}} < 11.0\,\mathrm{GeV}$.

Additionally, we apply a $E_{\mathrm{T}}^{\mathrm{miss}} < 60\,\mathrm{GeV}$ cut to make this region orthogonal to the signal region of the analysis.

We then fit the $\sigma_{i\eta i\eta}$ distribution of the EM object with a template describing the $\sigma_{i\eta i\eta}$ shape of true photons and another describing the hadronic background. The real photon template is taken from γ+jets MC requiring the photon to pass the e/γ ID except for the $\sigma_{i\eta i\eta}$ requirement. The fake photon template is taken from the same data control sample, requiring $5\,\mathrm{GeV} < I_{\mathrm{CH}} < 7\,\mathrm{GeV}$. The integral of the post-fit real photon template below $\sigma_{i\eta i\eta} = 0.0104$ is the number of true photons in the target sample.

The fit is performed once for all EM objects and then once for EM objects passing the γ-specific ID criteria. The ratio of the numbers of true photons obtained in the two fits is the efficiency.

The $\sigma_{i\eta i\eta}$ template fit method in its simplest form fits the observed distribution with the following fit function:

$$P(f; \sigma_{i\eta i\eta}) = f \cdot h_s(\sigma_{i\eta i\eta}) + (1 - f) \times h_b(\sigma_{i\eta i\eta}), \tag{2.2}$$

where h_s is the signal template, h_b is the background template, and f is the fraction of true photons in the target sample. Both the target template and the fit function are normalized to unity, removing the number of photon candidates in the target sample N as a fit parameter and leaving f as the only free parameter.

However, the hadronic background template, taken from the data control sample, has contributions from real photons. The amount of this "photon contamination" depends on the sideband choice, but is finite even for a sideband with very large I_{CH}. As described below, we perform additional fits with the background templates from alternative sidebands $3.5\,\text{GeV} < I_{CH} < 5\,\text{GeV}$ ("near") and $7.5\,\text{GeV} < I_{CH} < 9\,\text{GeV}$ ("far") to assess the systematic uncertainty. The photon contamination of the nominal and far sideband is 10-15%, and in the near sideband, it can go up to approximately 20% (see Figure 2-3).

To remove the photon contamination from the background templates, we modify the method and create a new background template h_b^{sub} from the original background template h_b by subtracting the true photon shape in the sideband $h_{S'}$. After normalization to unity, we obtain the expression

$$h_b^{\text{sub}}(\sigma_{i\eta i\eta}) = \frac{h_b(\sigma_{i\eta i\eta}) - S'/B \cdot h_{s'}(\sigma_{i\eta i\eta})}{1 - S'/B}, \tag{2.3}$$

where B is the number of photon candidates in the sideband and S' is the number of true photons in the sideband.

To determine S', we start with the number of true photons in the target sample, $f \cdot N$. We then scale this by the ratio of the relative fractions of true MC photons in
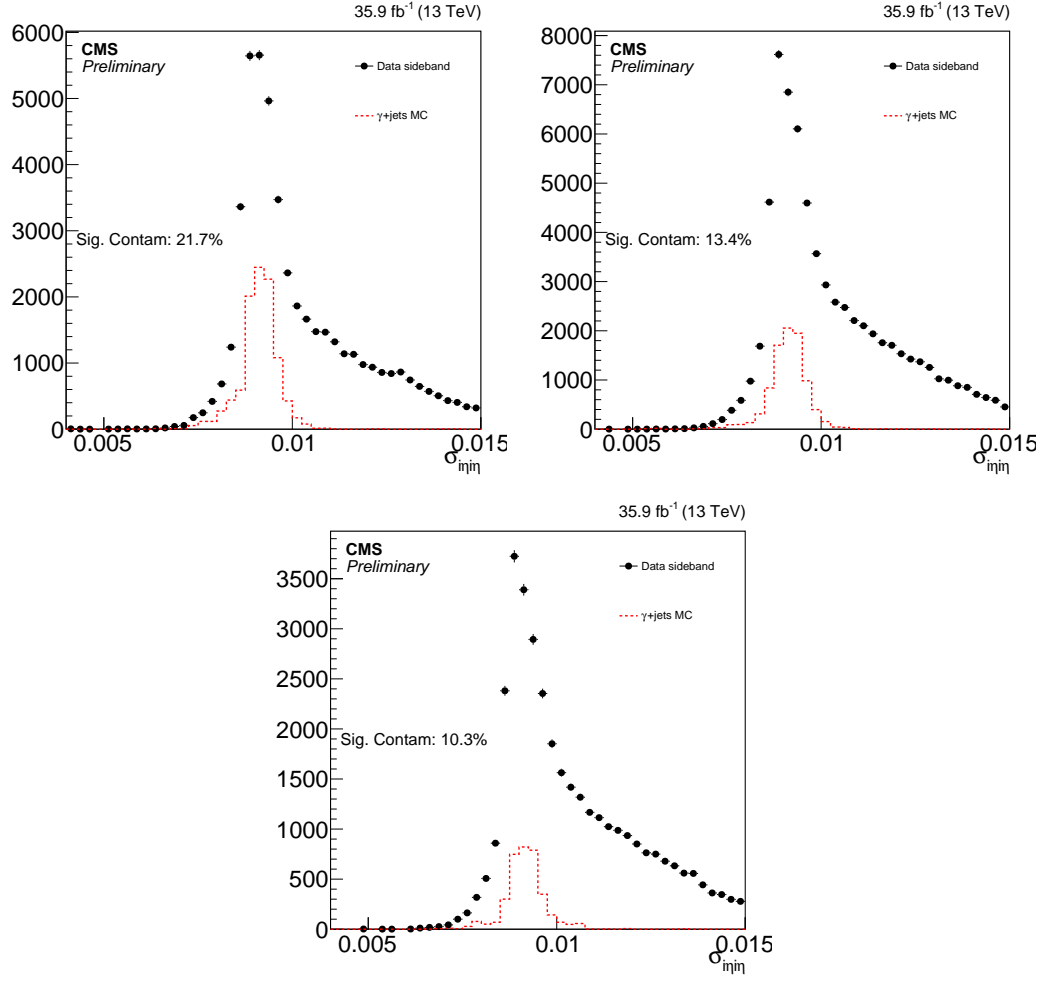
Figure 2-3: Signal contamination in the [3.5,5.0] (left), [5.0,7.5] (middle), and [7.5,9.0] (right) isolation sidebands.

the I_{CH} sideband r_{sb} and in the signal region r_{sig}, giving us the expression

$$S' = f \cdot \frac{r_{\mathrm{sb}}}{r_{\mathrm{sig}}} \cdot N. \tag{2.4}$$

Going back to our original fit function and replacing h_b with h_b^{sub} gives us

$$P(f; \sigma_{i\eta i\eta}) = f \cdot h_s(\sigma_{i\eta i\eta}) + (1 - f) \times \frac{h_b(\sigma_{i\eta i\eta}) - S'(f)/B \cdot h_{s'}(\sigma_{i\eta i\eta})}{1 - S'(f)/B}, \tag{2.5}$$

which converges to the original fit function if $S' = 0$, i.e., if there is no photon contamination in the sideband. Note that f is still the only free parameter for this new function as S' only depends on f and $r_{\mathrm{sb}}/r_{\mathrm{sig}}$ is set constant in the fit (see discussion of systematics for more detail).

There are four main sources of systematic uncertainty for this measurement. The first comes from the sideband choice, as the relative rates of different types of fake photons varies with I_{CH}. The second comes from the true photon I_{CH} shape, as this is used to determine the normalization of true photons in the sideband. Currently, this shape is taken from MC and thus there is the potential to mismodel the effects of the underlying event and pile-up. The third comes from the true photon $\sigma_{i\eta i\eta}$ distribution. As we take this from MC as well, we can mismodel the signal template shape. Finally, at high p_{T}, we suffer from low yields in our I_{CH} sidebands, which can lead to fluctutations that unduly influence the fit.

The uncertainty due to sideband choice is simply the larger of the differences of the purities measured using the near and far sidebands versus the nominal sideband. Figure 2-4 shows fits using the three sidebands for the $[400,\infty)$ p_{T} bin.

To measure the uncertainty due to the I_{CH} shape, we look at the I_{CH} for electrons in $Z \to ee$ events in both data and MC. Using these distributions, we obtain a data/MC scale factor which we apply to the MC true photon I_{CH} distribution to obtain a scaled MC distribution. This process is shown in Figure 2-5. Then, we recount the photons using this new distribution and take the difference in the values obtained using the raw MC and scaled MC distributions as a systematic uncertainty.

To measure the uncertainty due to the signal template $\sigma_{i\eta i\eta}$ shape, we look at the
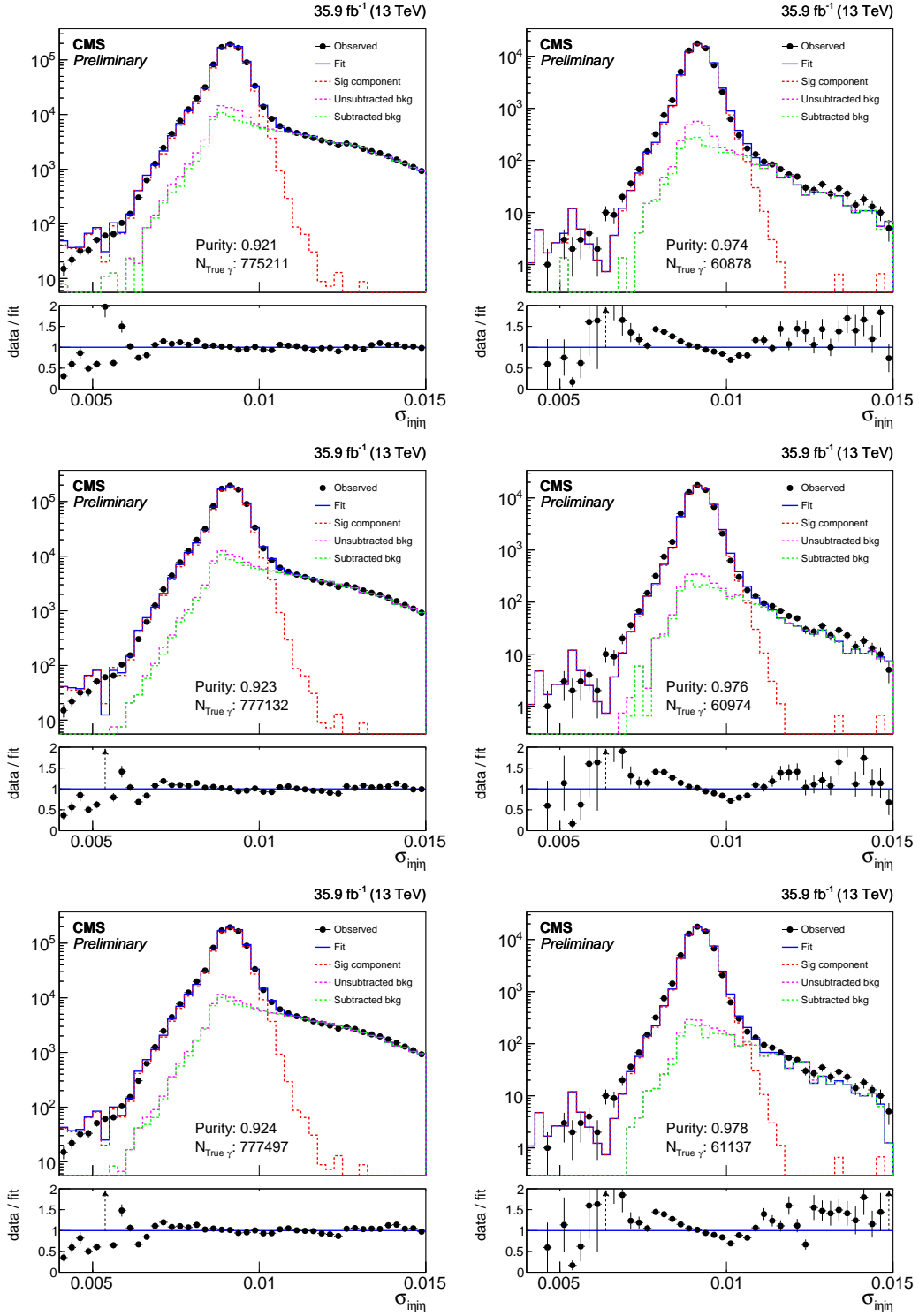
Figure 2-4: Fits to the $\sigma_{i\eta i\eta}$ distributions for the [175, 200] (left) and [400,∞) (right) p_{T} bins using the [3.5,5.0] (top), [5.0,7.5] (middle), and [7.5,9.0] (bottom) isolation sidebands. The blue solid line represents the full fit model, the red dashed line its signal component, and the green dashed line its background component.
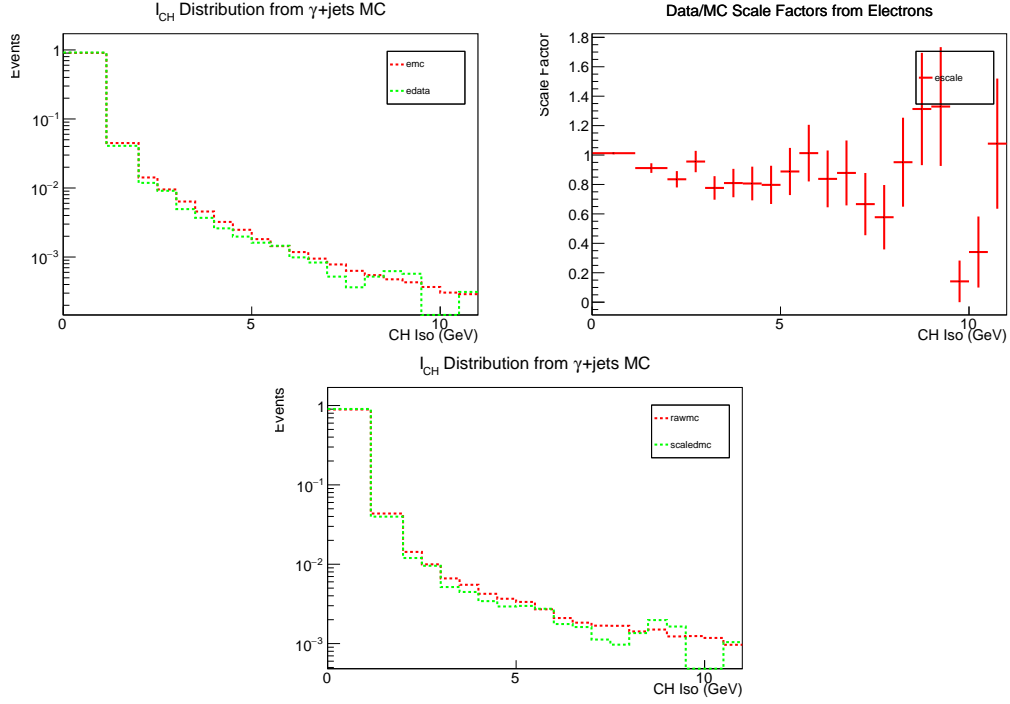
Figure 2-5: Top left: I_{CH} distributions of electrons in data and MC in $Z \to ee$ events. Top right: data/MC scale factor obtained from the electron I_{CH} distributions. Bottom: I_{CH} distributions of the MC photon objects used to estimate the amount of photon contamination in the background template, before and after applying the data/MC scale factor.

$\sigma_{i\eta i\eta}$ distributions for electrons in both data and MC. Again we compare $Z \to ee$ events in data and MC. From the $\sigma_{i\eta i\eta}$ distributions of high-purity electron samples, obtain a data/MC scale factor which we apply to the MC true photon $\sigma_{i\eta i\eta}$ distribution to obtain a scaled MC distribution. Then, we recount the photons using this new distribution and take the difference in the values obtained using the raw MC and scaled MC distributions as a systematic uncertainty. The difference between fits with and without the $\sigma_{i\eta i\eta}$ scaling are shown in Figure 2-6.
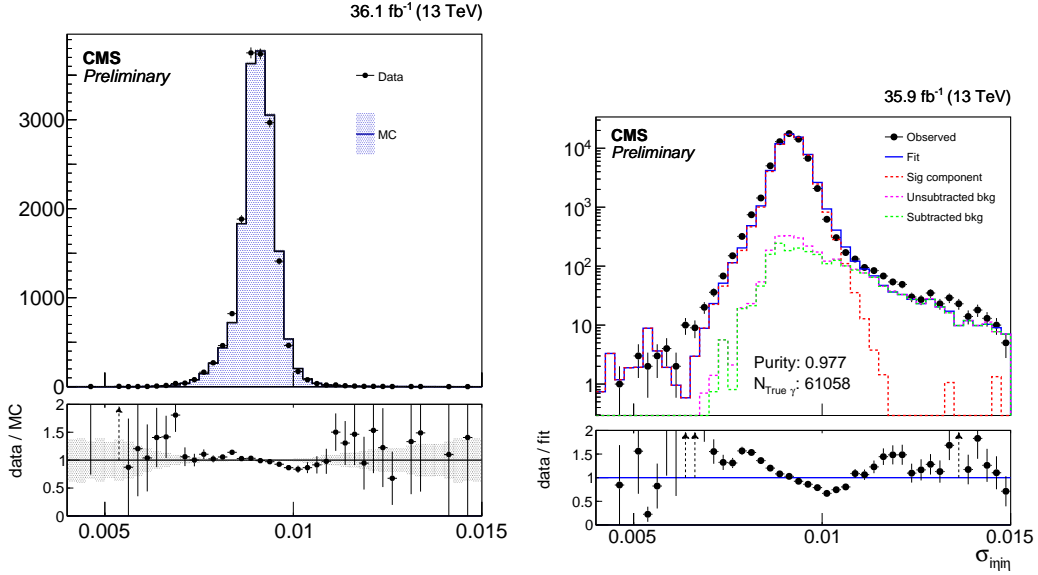

Figure 2-6: Left: Comparison of $\sigma_{i\eta i\eta}$ distributions between data and MC in $Z \to ee$ events. Lower panel shows the data/MC $\sigma_{i\eta i\eta}$ scale factor. Right: Result of the fit with true-photon template with the data/MC $\sigma_{i\eta i\eta}$ scale factor applied to the true-photon template.

To estimate the uncertainty due to statistical fluctuations in our background templates, we generate toys from the background template from data. We then repeat the fit with each of these toys and plot the distribution of the difference between the purity value obtained from the toy templates versus the nominal template. We take the standard deviation of this distribution, shown in Figure 2-7, as a systematic uncertainty.

The values obtained for each systematic uncertainty on the true photon count of the denominator are shown in Table 2.2 in bins of p_T. The relative uncertainties on the numerator are similar, and in the efficiency, each uncertainty source is considered
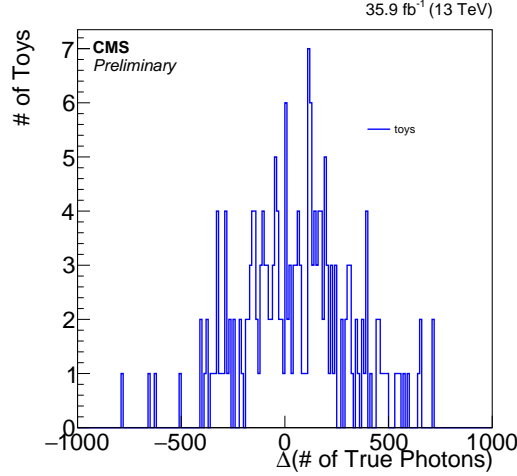
Figure 2-7: Shift in true-photon yields, extracted from alternative fits varying the background template within its statistical uncertainty. Nominal photon count in this specific E_T^γ bin is 6.64×10^5.

as fully correlated.

Table 2.2: Relative uncertainties on the estimated number of true photons in the denominator sample.

p_T Range (GeV)	Sources of Systematic Uncertainty			
	Sideband	I_{CH} Shape	Signal Shape	Bgkd. Stats
(175, 200)	0.09	0.18	0.05	0.04
(200, 250)	0.01	0.16	0.06	0.03
(250, 300)	0.14	0.16	0.06	0.05
(300, 350)	0.12	0.16	0.07	0.08
(350, 400)	0.23	0.11	0.05	0.10
(400, ∞)	0.27	0.09	0.05	0.05

The MC effiency of the γ-specific ID is determined by counting the number of truth-matched photons passing e/γ part of the ID and the full ID. However, there is a complication, the γ+jets region in data has approximately 5% contamination from electrons before applying the pixel veto, as shown in Figure 2-8. Thus, we combine appropriately cross-section weighted γ+jets, W+jets, and $t\bar{t}$ samples and truth match to both electrons and photons. Additionally, we apply a 14% uncertainty on the W+jets and $t\bar{t}$ yields to account for the NLO cross-section ratio uncertainties with respect to γ+jets at this p_T range that is uncorrelated between the numerator and

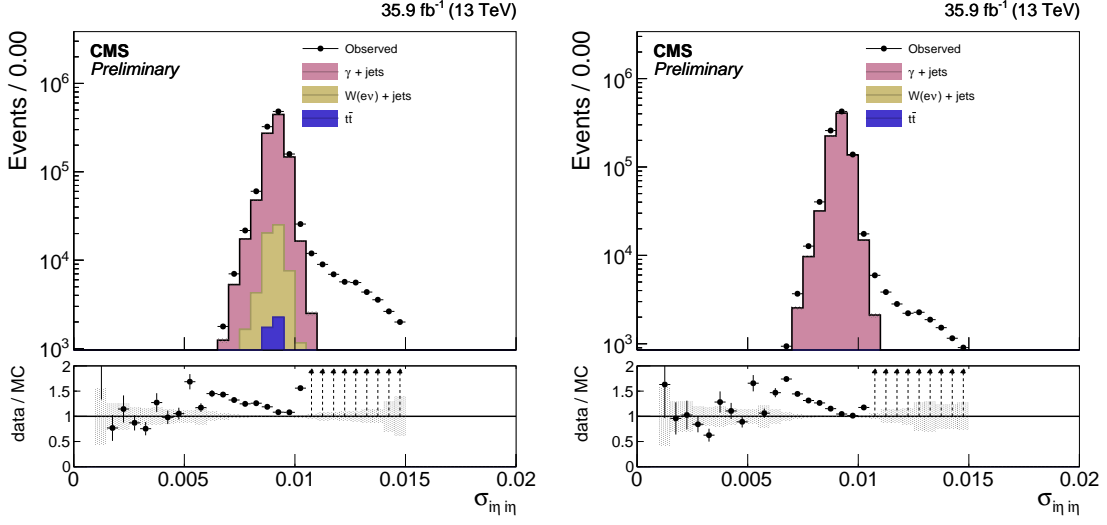denominator as a negligible amount of electron events survive the pixel veto.

Figure 2-8: Electron contamination in γ+jets region before (left) and after (right) applying the pixel seed veto.

The data efficiency, MC efficiency, and the scale factor for the γ-specific ID as a function of p_T are shown in Figure 2-9. As there is no significant trend in the scale factor as a function of p_T we apply a flat scale factor of 0.984 ± 0.009 for all of the MC-based background and signal models in the analysis.
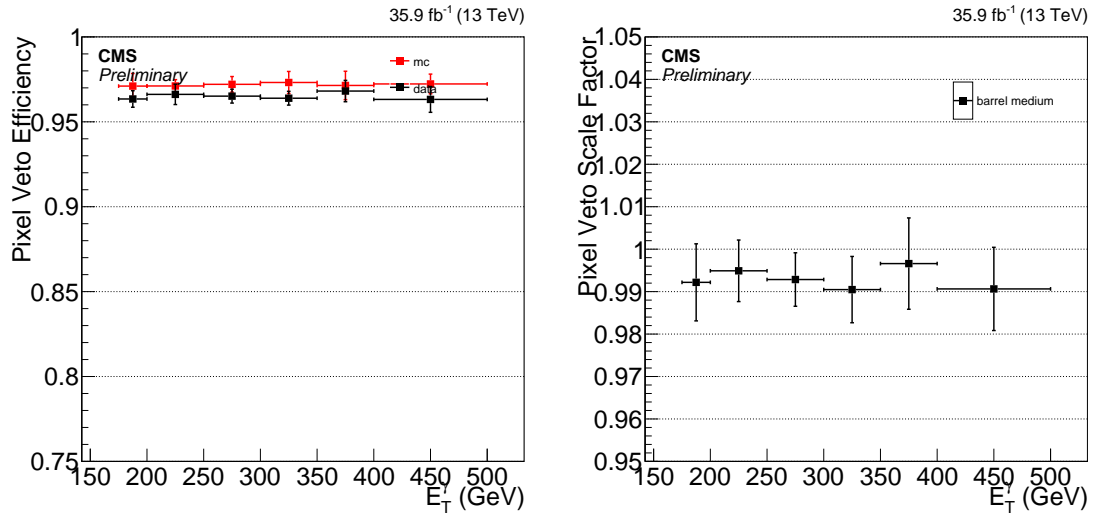
2.3 Lepton Scale Factors

2.4 Jet Energy Scale

Figure 2-9: Photon pixel veto efficiencies (left) and corresponding scale factor (right) as a function of photon p_T.

Chapter 3

The Monophoton Analysis

The main event.

3.1 Event Selection

3.2 Irreducible backgrounds
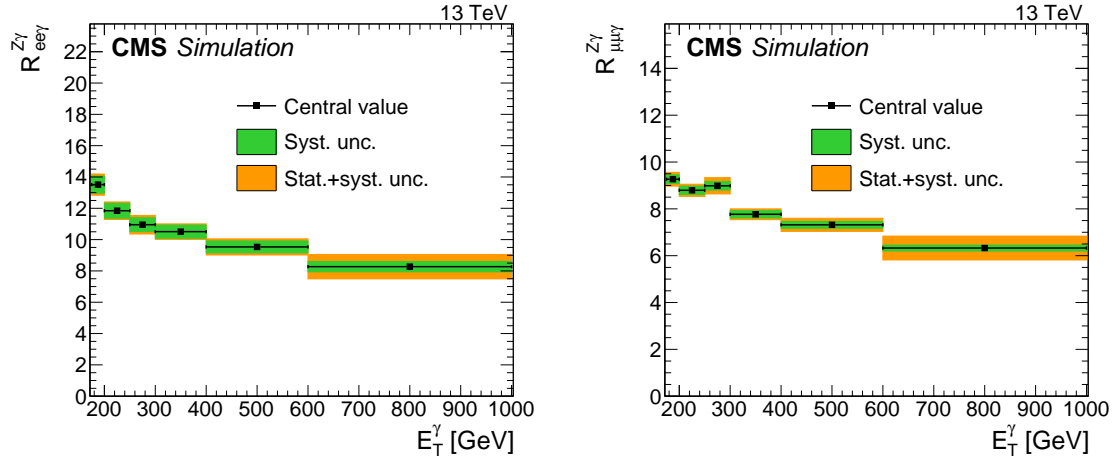

Figure 3-1:   Transfer factors $R^{Z\gamma}_{ee\gamma}$ (left) and $R^{Z\gamma}_{\mu\mu\gamma}$ (right). The uncertainty bands in green (inner) and orange (outer) show the systematic uncertainty, and the combination of systematic and statistical uncertainty arising from limited MC sample size, respectively. The systematic uncertainties considered are the uncertainties in the data-to-simulation correction factors ρ for the lepton identification efficiencies.

Using the transfer factor $R_{\ell\ell\gamma}^{Z\gamma}$, the total estimated event yield $T_{\ell\ell\gamma}$ in each dilepton control region in the ith bin of the E_{T}^{γ} distribution can be expressed as

$$T_{\ell\ell\gamma,i} = \frac{N_i^{Z\gamma}}{R_{\ell\ell\gamma,i}^{Z\gamma}} + b_{\ell\ell\gamma,i},$$

$$(3.1)$$

where $N^{Z\gamma}$ is the number of $Z(\to \nu\bar{\nu})+\gamma$ events in the combined signal regions and $b_{\ell\ell\gamma}$ is the predicted contribution from other background sources in the dilepton control region, namely $t\bar{t}\gamma$, $VV\gamma$, and misidentified hadrons. The subscript i indicates that the quantities are evaluated in bin i of the E_{T}^{γ} distribution.
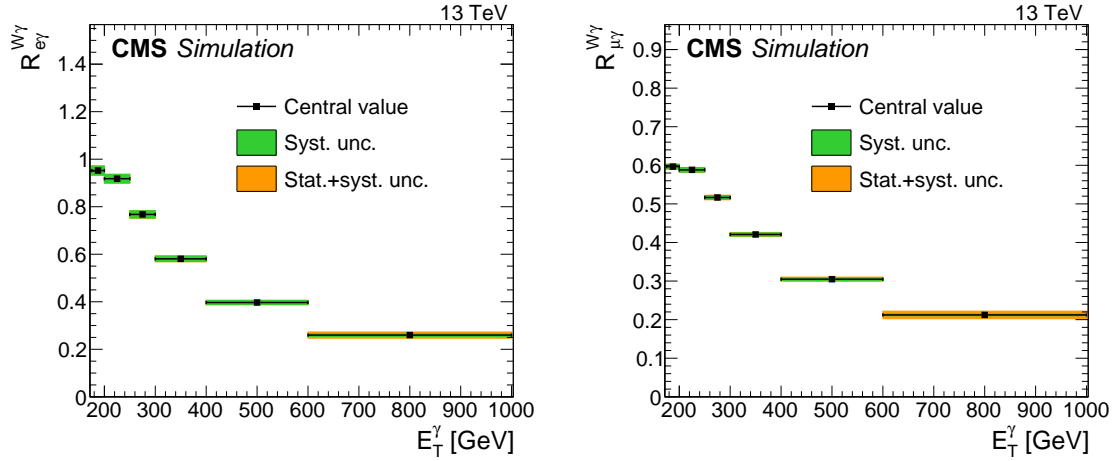

Figure 3-2: Transfer factors $R_{e\gamma}^{W\gamma}$ (left) and $R_{\mu\gamma}^{W\gamma}$ (right). The uncertainty bands in green (inner) and orange (outer) show the systematic uncertainty, and the combination of systematic and statistical uncertainty arising from limited MC sample size, respectively. The systematic uncertainties considered are the uncertainties in the data-to-simulation correction factors ρ for the lepton identification efficiencies.

Using $R_{\ell\gamma}^{W\gamma}$ and $f_{W\gamma}^{Z\gamma}$, the total estimated event yield $T_{\ell\gamma}$ in each single-lepton control region in the ith bin of the E_{T}^{γ} distribution can be expressed as

$$T_{\ell\gamma,i} = \frac{N_i^{Z\gamma}}{R_{\ell\gamma,i}^{W\gamma} f_{W\gamma,i}^{Z\gamma}} + b_{\ell\gamma,i},$$

$$(3.2)$$

where $b_{\ell\gamma}$ is the predicted contribution from other background sources in the single-lepton regions, namely misidentified electrons and hadrons and other minor SM processes.
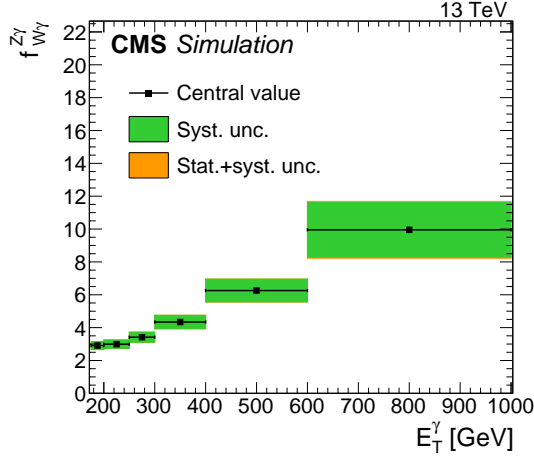
Figure 3-3: Transfer factor $f_{W\gamma}^{Z\gamma}$. The uncertainty bands in green (inner) and orange (outer) show the systematic uncertainty, and the combination of systematic and statistical uncertainty arising from limited MC sample size, respectively. The systematic uncertainties considered are the uncertainties from higher-order theoretical corrections.

3.2.1 Higher-order corrections to V+γ differential cross sections

We apply the correction factors shown in Fig. 3-4, which are combinations of Sudakov suppression factors and photon-induced enhancements, and are provided by the authors of Ref. [?] in addition to the NNLO QCD correction.

Figure 3-5 shows the effect of systematic uncertainty in the ratio between the $Z(\to \nu\bar{\nu})+\gamma$ and $W(\to \ell\nu)+\gamma$ processes with respect to nominal value for $Z\gamma$ and $W\gamma$ respectively.

3.3 Misidentified electrons

An electron can be misidentified as a photon if the association of tracks or track seeds to the supercluster in ECAL fails in the reconstruction step. The production of a single W boson decaying to an electron and a neutrino is a high-rate process, and it mimics the photon plus E_T^{miss} signature if the electron is misidentified.

The rate at which this misidentification occurs is proportional to the inefficiency
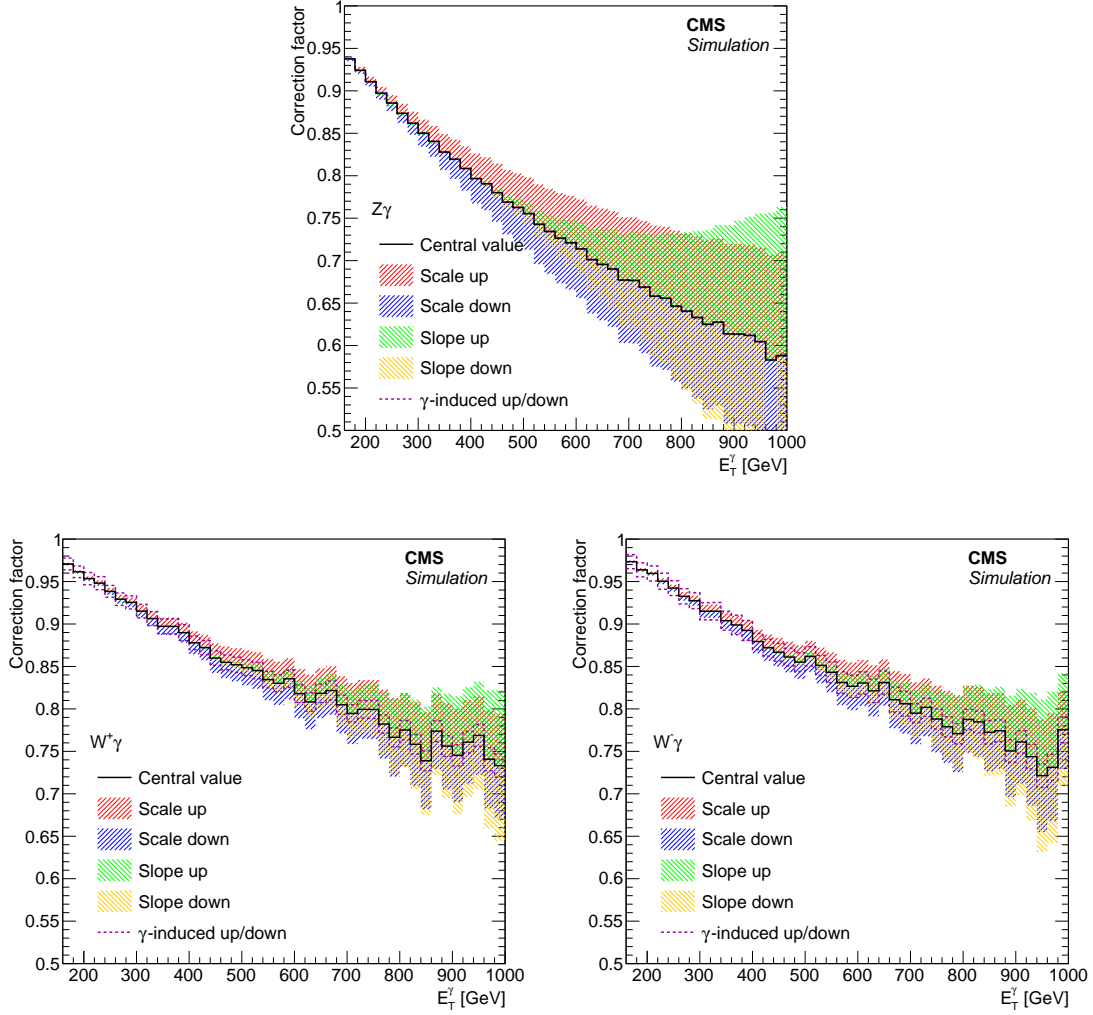
31

Figure 3-4: Electroweak NLO cross section corrections as a function of photon p_{T} for $Z(\to \nu\bar{\nu})+\gamma$ (top), $W^+ + \gamma$ (bottom left), and $W^- + \gamma$ (bottom right) processes, overlaid with uncertainty bands. See text for descriptions of the individual components of the uncertainty. The uncertainty due to γ-induced production is negligible in $Z(\to \nu\bar{\nu})+\gamma$ production.
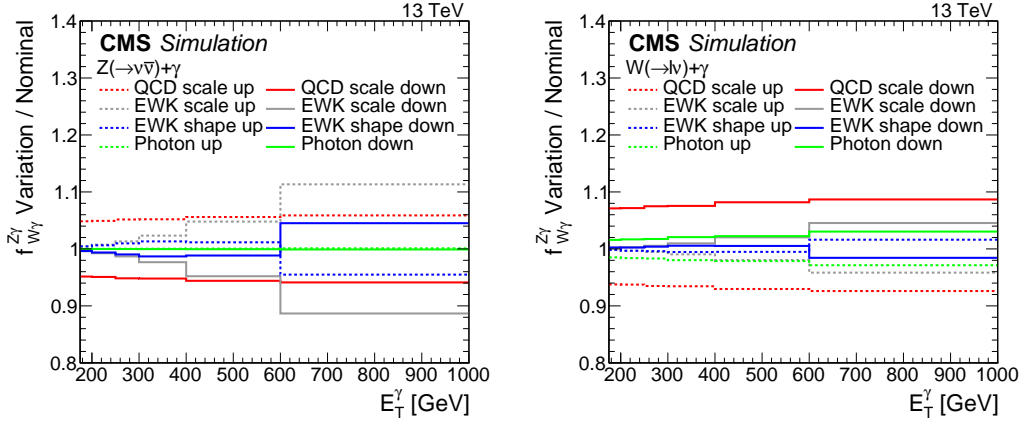
Figure 3-5: Systematic uncertainty in the transfer factors for $Z(\to \nu\bar{\nu})+\gamma$ (left) and $W(\to \ell\nu)+\gamma$ (right). The last bin includes all events with $E_T^\gamma > 1000\,\text{GeV}$.

$1 - \epsilon_e^{\text{track}}$ of the tracking, defined over the electrons passing the photon identification criteria described in Sec. 1.4 except the electron veto. This partial identification is denoted as $e\gamma$ ID in the following. If one assumes that the kinematic and other critical properties of the electron plus E_T^{miss} events are unaffected by the electron misidentification, it is possible to model the electron misidentification background by taking a proxy sample with well-identified electrons and scaling this sample by $R_e = (1 - \epsilon_e^{\text{track}})/\epsilon_e^{\text{track}}$.

The "tag-and-probe" method described in Section 2.2.1 with appropriate changes is used to measure the efficiency corresponding to the factor R_e in data.

The first such change is that the sample is split into $e\gamma$ and ee categories depending on whether the probe passes or fails the electron veto requirement. Probes in both categories must also pass the $e\gamma$ ID. Denoting the area of the peak in each cateogry $N_{e\gamma}$ and N_{ee}, respectively, the ratio $N_{e\gamma}/N_{ee}$ is equal to R_e up to minor systematic corrections.

The second such change is in the background model used in the TP fits. The backgrounds to the $e\gamma$ fit consist of processes with actual electron and photon in the final state, such as $W\gamma$ and $Z \to ee$ with a hard radiation off one of the electrons. Because of this, we scale the mass distrubition of the $\mu + \gamma$ sample by the ratio of electron-probe to muon-probe events taken from MC to account for the different rates

of FSR and bremsstrahlung between muons and electrons. As an alternative template to assess the systematic effect introduced by the choice of the background template, the unscaled mass distribution is also tested.

Figure 3-6 shows the six fits performed on ee and $e\gamma$ in bins of probe p_T, from which the R_e factor used for the estimation of the electron misidentification background is derived.

The proxy sample for the background estimation is obtained by identical event selection as that described in Sec. **??**, but with the pixel-seed veto inverted on the photon candidate object.

Figure 3-7 shows the derived R_e factor as a function of E_T^γ. The electron proxy sample is reweighted by R_e depending on the p_T of the electron object.

3.4   Misidentified hadrons

The estimation of hadron misidentification background proceeds in multiple steps. First, the fraction of hadronic objects within a pool of photon candidate objects in the photon plus jet control region is measured. This measurement is described in detail in Section 2.2.2. Figure 3-8 and Table 3.1 show the final impurity and associated uncertainties as a function of p_T.

p_T (GeV)	Nominal	Sources of Systematic Uncertainty			
		Sideband	CH Iso Shape	Signal Shape	Bgkd. Stats
(175, 200)	4.31 ± 0.21	0.09	0.18	0.05	0.04
(200, 250)	3.39 ± 0.17	0.01	0.16	0.06	0.03
(250, 300)	2.44 ± 0.22	0.14	0.16	0.06	0.05
(300, 350)	1.99 ± 0.23	0.12	0.16	0.07	0.08
(350, 400)	1.43 ± 0.28	0.23	0.11	0.05	0.10
(400, ∞)	0.63 ± 0.30	0.27	0.09	0.05	0.05

Table 3.1: Impurities for photons as a function of p_T.

Following this measurement, another control sample is formed where the photon in the photon plus jet sample is replaced by a hadronic proxy object. The hadronic proxy object is a reconstructed photon object which pass the photon ID described in Section 1.4 with the execption of failing at least one of the following cuts:
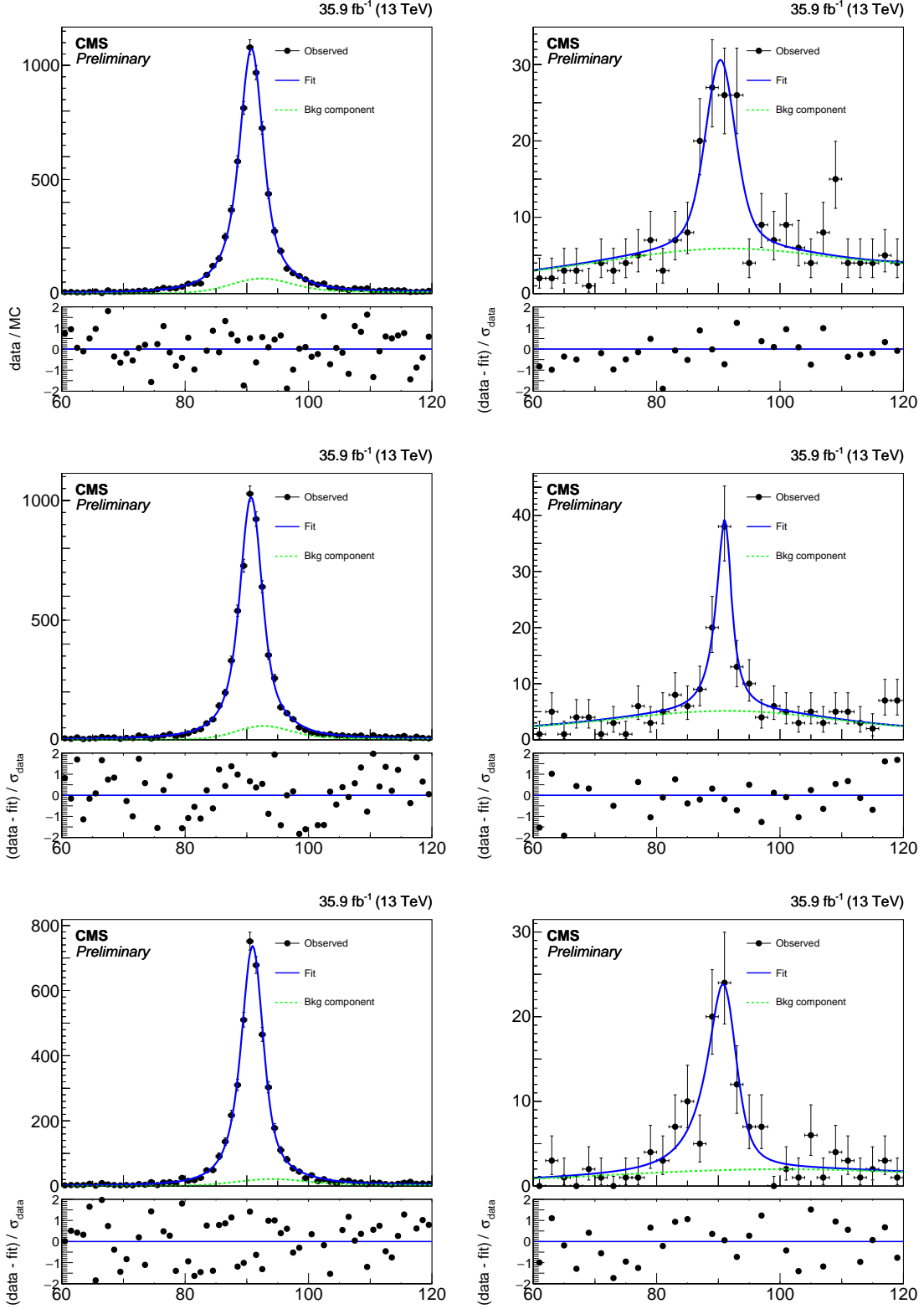
Figure 3-6: Fits to the mass distributions for ee (left) and $e\gamma$ (right) selections, in bins of probe p_T: $175 < p_\mathrm{T} < 200\,\mathrm{GeV}$ (top), $200 < p_\mathrm{T} < 250\,\mathrm{GeV}$ (middle), $p_\mathrm{T} > 250\,\mathrm{GeV}$ (bottom). The blue solid line represents the full fit model, and the green dashed line its background component.
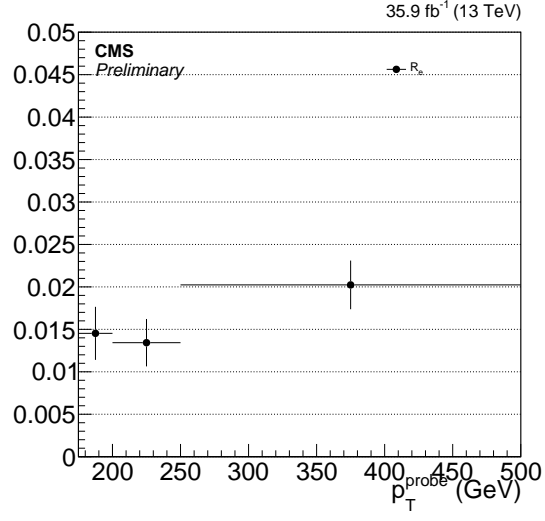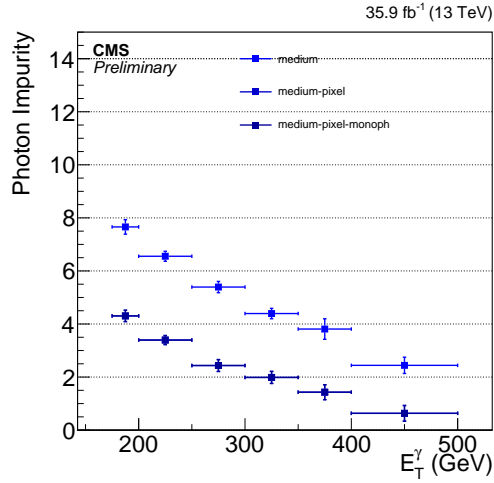
Figure 3-7: Electron to photon fake rate R_e.

Figure 3-8: Impurities for photons as a function of p_T. The different bands show the effects of adding different stages of the full ID, starting with the baseline ID and isolation and successively adding the pixel seed veto.

- $\sigma_{i\eta i\eta} < 0.01022$

- PF Charged Hadron isolation < 0.441 GeV .

Additionally, we apply a $E_T^{miss} < 60$ GeV cut to make this region orthogonal to the signal region of the analysis.

The hadronic transfer factor R_h, which measures the rate at which hadronic proxy objects result in hadrons that are misidentified as candidate photons, is obtained by dividing the estimated number of misidentified hadrons in the photon plus jet sample by the number of events in the hadron proxy + jet control region as a function of p_T. Figure 3-9 shows the transfer factor R_h along with the various distributions used for its derivation.

Finally, a third control sample of events with a hadronic proxy object and $E_T^{miss} > 170$ GeV is prepared. Under the assumption that the R_h stays constant regardless of whether the event has a high-p_T jet or E_T^{miss}, this proxy plus E_T^{miss} sample is then weighted by R_h to arrive at an estimate of the misidentified hadron plus E_T^{miss} background of this analysis.

To estimate the uncertainty on this background, we repeat the above method using tighter and looser definitions of the hadron proxy object. The tighter definition differs from the nominal by the following cuts:

- ρ-corrected PF Neutral Hadron isolation $< 0.264 + 0.014 \times p_T^\gamma + 0.000019 \times (p_T^\gamma)^2$.

- ρ-corrected PF Photon isolation $< 2.362 + 0.0053 \times p_T^\gamma$,

and the looser definition differs from the nominal by the following cuts:

- ρ-corrected PF Neutral Hadron isolation $< 10.910 + 0.014 \times p_T^\gamma + 0.000019 \times (p_T^\gamma)^2$.

- ρ-corrected PF Photon isolation $< 3.630 + 0.0053 \times p_T^\gamma$.

The different distributions from the nominal, tight, and loose selections are shown in Figure 3-10. The tight and loose shapes are taken as the one sigma band around the nominal estimate. Additionally, there is an uncertainty coming from the estimation of the photon purity. Figure 3-11 shows the resulting shapes from moving the shapes generated by a one sigma shift in the purity.
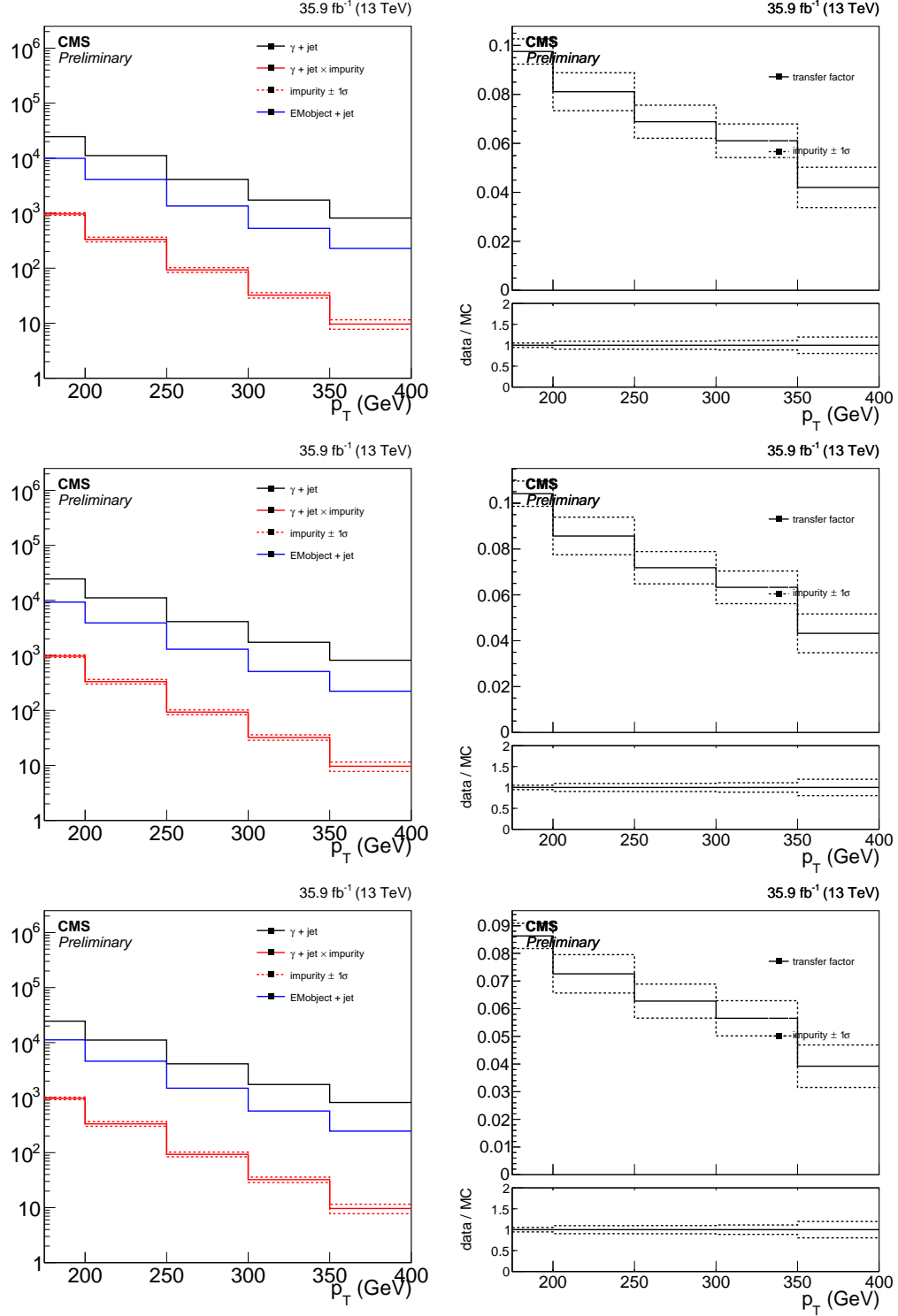
Figure 3-9: Left: The p_T distribution of the candidate photon object in the photon + jet control sample (black), the result of scaling it with the impurity (red), and the p_T distribution of the hadronic proxy object in the proxy + jet control sample (blue). Right: Hadronic transfer factor R_h, which is the ratio of the red and blue distributions in the left plot. Top: Nominal hadron proxy object. Middle: Tighter hadron proxy object. Bottom: Looser hadron proxy object.
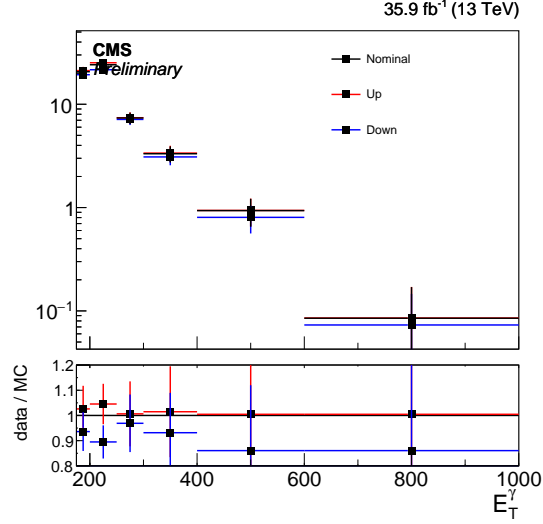
Figure 3-10: The p_T distribution of the estimated contribution from hadronic fakes in the signal region. The distribution labeled Up (Down) comes from the tighter (looser) selection. The systematic uncertainty resulting from this variation is around 5% at the low end of our p_T range and increases to 15% after $p_T > 400$ GeV.
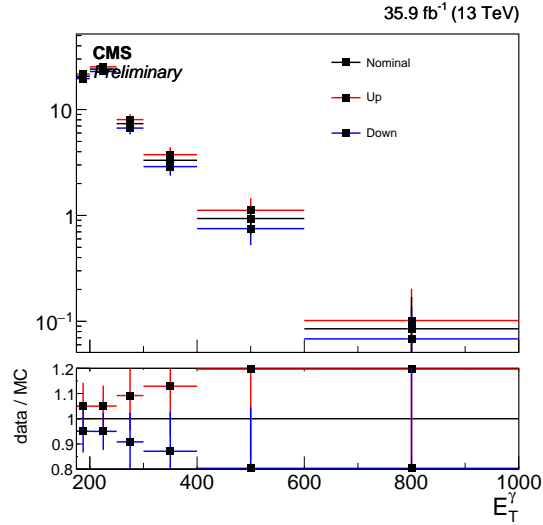

Figure 3-11: The p_T distribution of the estimated contribution from hadronic fakes in the signal region. The distribution labeled Up (Down) comes from varying the purity one sigma up (down). The systematic uncertainty resulting from this variation is around 5% at the low end of the p_T range and increases to 20% after $p_T > 400$ GeV.
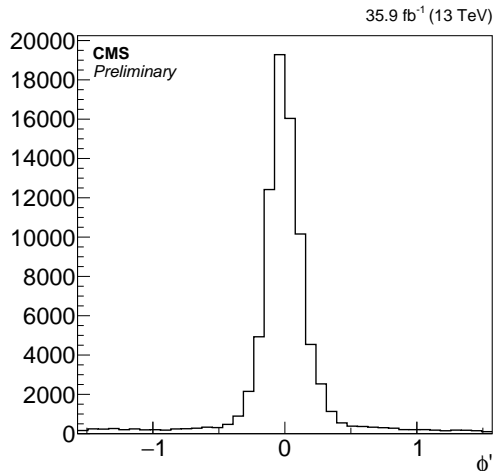
Figure 3-12: Folded ϕ' distribution of the halo sample.

3.5 Spikes

3.6 Beam halo

Figure 3-12 shows the ϕ distribution of the halo showers obtained from the single photon data set, requiring $E_{\mathrm{T}}^{\gamma} > 175\,\mathrm{GeV}$ and $E_{\mathrm{T}}^{\mathrm{miss}} > 170\,\mathrm{GeV}$. Here, halo showers are defined as photon objects that fail the MIP total energy cut in events where beam halo MET filter (one component of the "MET filters" mentioned in Section **??**). The distribution is shifted and folded to make the peaking behavior clear. The resulting variable on the horizontal axis is named ϕ'.

The splitting of the signal region can be thought of as a two-bin fit. Collision processes occupy the relative fractions of phase space in the horizontal (H) and vertical (V) signal regions, $C_H = 1/\pi$ and $C_V = (\pi - 1)/\pi$, respectively. The corresponding fractions for beam halo events are determined by selecting a halo-enriched sample where the halo identification is inverted. Thus, a fit of the two signal regions provides an estimate of the overall normalization of the beam halo background, denoted h. The E_{T}^{γ} dependence of the halo background is encoded in $n_{K,i}^{\mathrm{halo}}$, the unit-normalized beam halo prediction in the ith bin of the signal region $K \in \{H, V\}$. Using the notation introduced in Section 3.2, the total estimated background T_K in the two signal

regions are

$$T_{K,i} = C_K(N_i^{Z\gamma} + N_i^{W\gamma}) + hn_{K,i}^{\text{halo}} + C_K b_{K,i}$$
$$= C_K(1 + f_{W\gamma i}^{Z\gamma}{}^{-1})N_i^{Z\gamma} + hn_{K,i}^{\text{halo}} + C_K b_{K,i},$$

(3.3)

where $b_{K,i}$ is the total contribution to bin i of region K from electron and hadron misidentification, ECAL spikes, and other minor SM background processes.

3.7 Other minor SM background processes

The SM $t\bar{t}\gamma$, VVγ, $Z(\to \ell\bar{\ell})+\gamma$, $W \to \ell\nu$, and γ+jets processes are minor ($\sim$10%) background processes in the signal region. Although $Z(\to \ell\bar{\ell})+\gamma$ and γ+jets do not involve high-p_T invisible particles, the former can exhibit large E_T^{miss} when the leptons fail to be reconstructed, and the latter when jet energy is severely mismeasured. The estimates for all five processes are taken from MADGRAPH5_aMC@NLO simulations at LO in QCD and can be found in Tables 3.2 and 3.3.

3.8 Statistical Interpretation

Free parameters of the fit are the yield of $Z(\to \nu\bar{\nu})+\gamma$ background in each bin of the signal regions $(N_i^{Z\gamma})$ and the overall normalization of the beam halo background (h). Bin-by-bin yields of $W(\to \ell\nu)+\gamma$ and $Z(\to \ell\bar{\ell})+\gamma$ samples in all regions are related to the yield of $Z(\to \nu\bar{\nu})+\gamma$ through the MC prediction through the transfer factors defined in Section 3.2. The transfer factors are allowed to shift within the aforementioned theoretical and experimental uncertainties.

The background-only likelihood that is maximized in the fit is

$$\mathcal{L} = \prod_i \left\{ \mathcal{L}_{\text{signal}} \times \mathcal{L}_{\text{single-lepton}} \times \mathcal{L}_{\text{dilepton}} \right\} \times \mathcal{L}_{\text{nuisances}}$$

$$= \prod_i \left\{ \prod_{K=H,V} \mathcal{P}\left(d_{K,i} \Big| T_{K,i}(\vec{\theta})\right) \times \prod_{\ell=e,\mu} \mathcal{P}\left(d_{\ell\gamma,i} \Big| T_{\ell\gamma,i}(\vec{\theta})\right) \times \prod_{\ell=e,\mu} \mathcal{P}\left(d_{\ell\ell\gamma,i} \Big| T_{\ell\ell\gamma,i}(\vec{\theta})\right) \right\} \times \prod_j \mathcal{N}(\theta_j)$$

$$= \prod_i \left\{ \begin{array}{c} \displaystyle\prod_{K=H,V} \mathcal{P}\left(d_{K,i} \Big| \left(1 + f_{W\gamma,i}^{Z\gamma}{}^{-1}(\vec{\theta})\right) C_K N_i^{Z\gamma} + hn_{K,i}^{\text{halo}}(\vec{\theta}) + C_K b_{K,i}(\vec{\theta})\right) \\[1.5em] \displaystyle\times \prod_{\ell=e,\mu} \mathcal{P}\left(d_{\ell\gamma,i} \Big| \frac{N_i^{Z\gamma}}{R_{\ell\gamma,i}^{W\gamma}(\vec{\theta}) f_{W\gamma,i}^{Z\gamma}(\vec{\theta})} + b_{\ell\gamma,i}(\vec{\theta})\right) \\[1.5em] \displaystyle\times \prod_{\ell=e,\mu} \mathcal{P}\left(d_{\ell\ell\gamma,i} \Big| \frac{N_i^{Z\gamma}}{R_{\ell\ell\gamma,i}^{Z\gamma}(\vec{\theta})} + b_{\ell\ell\gamma,i}(\vec{\theta})\right) \end{array} \right\} \times \prod_j \mathcal{N}(\theta_j),$$

$$(3.4)$$

following the notation introduced in Section 3.2, and where $\mathcal{P}(n|\lambda)$ is the Poisson probability of n for mean λ, $\mathcal{N}$ denotes the unit normal distribution, and $d_{X,i}$ is the observed number of events in bin i of region X. Systematic uncertainties are treated as nuisance parameters in the fit and are represented by $\vec{\theta}$. Each quantity Q_j with a nominal value $\overline{Q}_j$ and a standard deviation of the systematic uncertainty σ_j appears in the likelihood function as $\overline{Q}_j \exp(\sigma_j \theta_j)$.

3.9   Results

3.9.1   Pre-fit and post-fit distributions

Figure 3-13 shows the observed E_{T}^{γ} distributionsin the four control regions compared with the results from simulations before and after performing the simultaneous fit across all the control samples and signal region, and assuming absence of any signal. Figure 3-14 shows the observed E_{T}^{γ} distributions in the horizontal and vertical signal regions compared with the results from simulations before and after performing a combined fit to the data in all the control samples and the signal region. The observed distributions are in agreement with the prediction from SM and noncollision backgrounds.
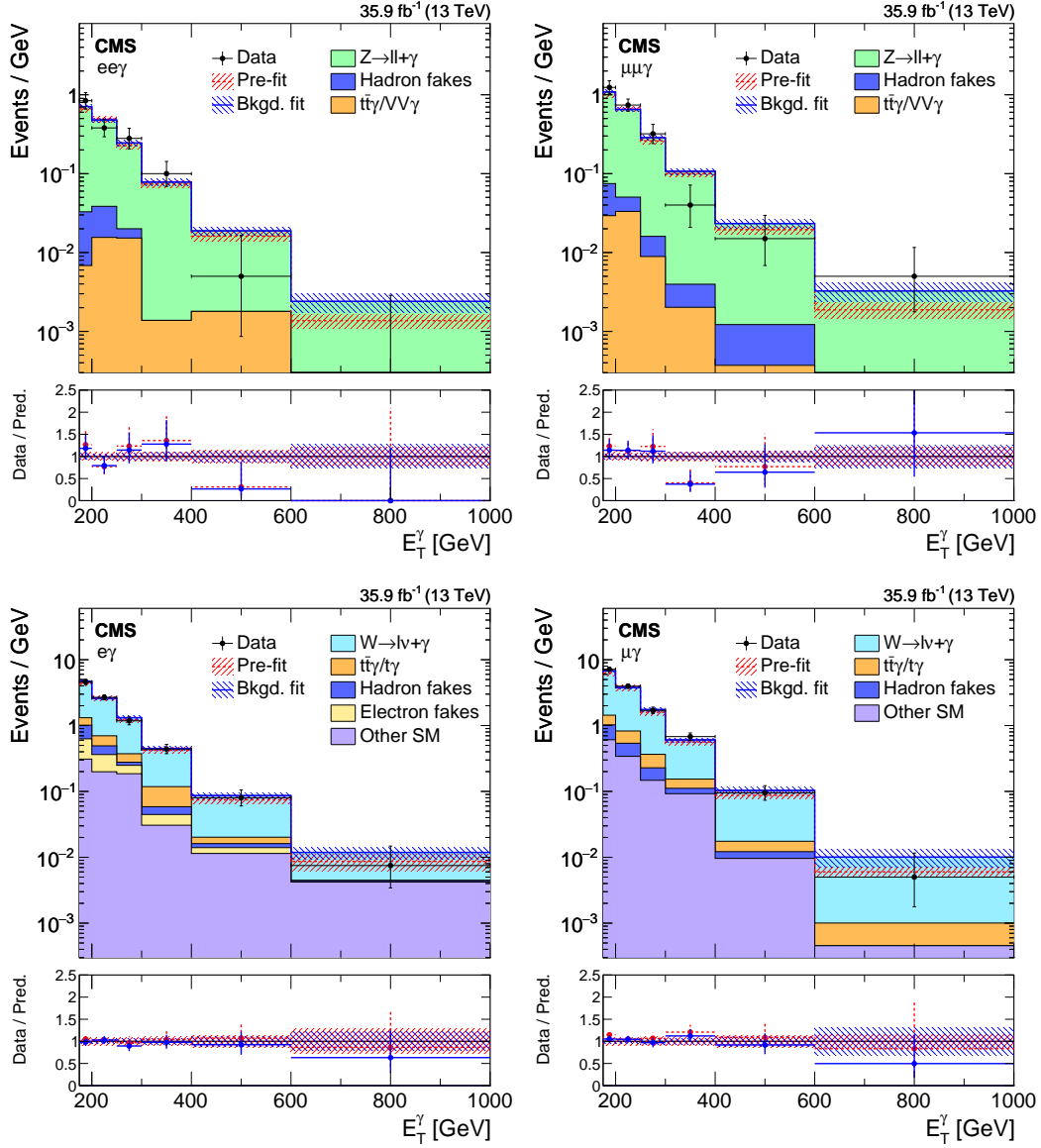
Figure 3-13: Comparison between data and MC simulation in the four control regions: $ee\gamma$ (upper left), $\mu\mu\gamma$ (upper right), $e\gamma$ (lower left), $\mu\gamma$ (lower right) before and after performing the simultaneous fit across all the control samples and signal region, and assuming absence of any signal. The last bin of the distribution includes all events with $E_T^\gamma > 1000$ GeV. The ratios of data with the pre-fit background prediction (red dashed) and post-fit background prediction (blue solid) are shown in the lower panels. The bands in the lower panels show the post-fit uncertainty after combining all the systematic uncertainties.
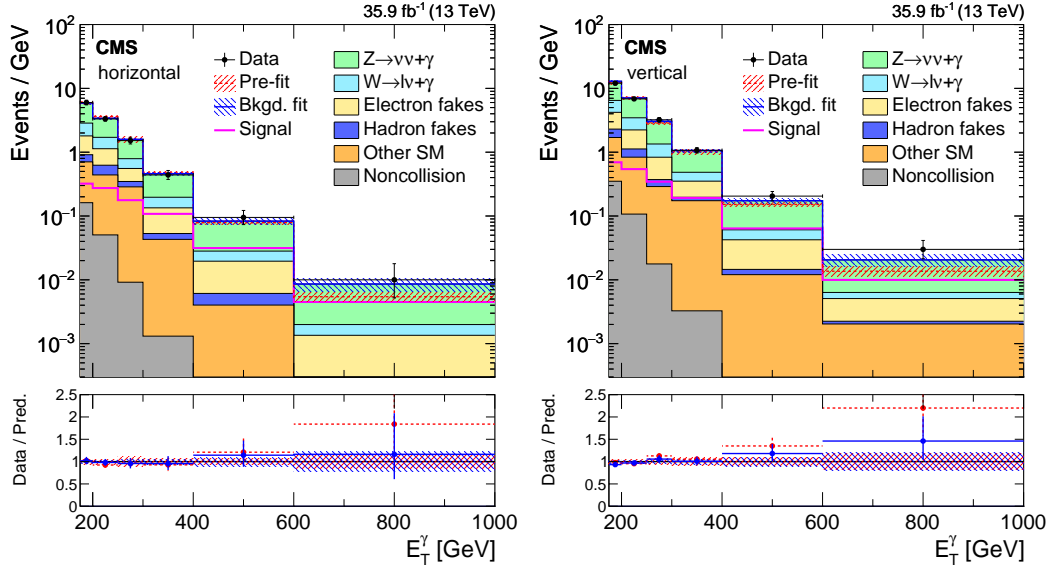
Figure 3-14: Observed E_T^γ distributions in the horizontal (left) and vertical (right) signal regions compared with the post-fit background expectations for various SM processes. The last bin of the distribution includes all events with $E_T^\gamma > 1000$ GeV. The expected background distributions are evaluated after performing a combined fit to the data in all the control samples and the signal region. The ratios of data with the pre-fit background prediction (red dashed) and post-fit background prediction (blue solid) are shown in the lower panels. The bands in the lower panels show the post-fit uncertainty after combining all the systematic uncertainties. The expected signal distribution from a 1 TeV vector mediator decaying to 1 GeV DM particles is overlaid.

Table 3.2: Expected event yields in each E_T^γ bin for various background processes in the horizontal signal region. The background yields and the corresponding uncertainties are obtained after performing a combined fit to data in all the control samples, excluding data in the signal region. The observed event yields in the horizontal signal region are also reported.

E_T^γ [GeV]	[175, 200]	[200, 250]	[250, 300]	[300, 400]	[400, 600]	[600, 1000]
$Z\gamma$	81.2 ± 8.0	88.2 ± 8.4	38.8 ± 4.8	26.8 ± 3.7	8.8 ± 1.9	1.4 ± 0.7
$W\gamma$	27.9 ± 3.7	29.9 ± 3.9	11.4 ± 1.7	6.3 ± 1.2	1.4 ± 0.4	0.1 ± 0.1
Misid. electrons	22.5 ± 2.7	25.7 ± 2.7	10.5 ± 1.0	8.2 ± 0.7	2.7 ± 0.2	0.5 ± 0.0
Misid. hadrons	5.2 ± 2.2	9.3 ± 1.8	3.1 ± 0.7	1.0 ± 0.3	0.4 ± 0.1	0.0 ± 0.0
Other SM	13.6 ± 2.0	19.6 ± 1.3	13.9 ± 0.4	4.2 ± 0.2	0.8 ± 0.0	0.1 ± 0.0
ECAL spikes	4.3 ± 1.3	2.7 ± 0.8	0.5 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Total prediction	154.6 ± 8.3	175.4 ± 8.8	78.2 ± 5.3	46.6 ± 4.0	14.1 ± 2.1	2.1 ± 0.8
Observed	150 ± 12	166 ± 13	76.0 ± 8.7	44.0 ± 6.6	19.0 ± 4.4	4.0 ± 2.0

Table 3.3: Expected event yields in each E_T^γ bin for various background processes in the vertical signal region. The background yields and the corresponding uncertainties are obtained after performing a combined fit to data in all the control samples, excluding data in the signal regions. The observed event yields in the vertical signal region are also reported.

E_T^γ [GeV]	[175, 200]	[200, 250]	[250, 300]	[300, 400]	[400, 600]	[600, 1000]
$Z\gamma$	172 ± 17	190 ± 18	83 ± 10	58.6 ± 7.9	18.0 ± 3.9	3.1 ± 1.6
$W\gamma$	59.9 ± 7.8	63.6 ± 7.8	24.6 ± 3.5	13.4 ± 2.4	3.0 ± 0.8	0.3 ± 0.2
Misid. electrons	48.4 ± 5.6	56.2 ± 5.1	23.4 ± 1.8	15.7 ± 1.4	5.6 ± 0.4	1.2 ± 0.1
Misid. hadrons	15.1 ± 4.4	14.5 ± 3.1	4.2 ± 0.8	2.3 ± 0.8	0.5 ± 0.1	0.1 ± 0.1
Other SM	33.8 ± 4.1	36.6 ± 2.7	13.6 ± 0.5	17.1 ± 0.6	2.4 ± 0.1	0.8 ± 0.0
ECAL spikes	9.3 ± 2.8	5.7 ± 1.7	0.9 ± 0.3	0.3 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
Total prediction	339 ± 18	366 ± 19	150 ± 11	107.5 ± 8.7	29.6 ± 4.3	5.4 ± 1.7
Observed	301 ± 17	342 ± 19	161 ± 13	107 ± 10	41.0 ± 6.4	12.0 ± 3.5

The expected yields in each bin of E_T^γ for all backgrounds in the horizontal and vertical signal regions after performing a combined fit to data in all the control samples, excluding data in the signal regions, are given in Tables 3.2 and 3.3, respectively. The covariances between the predicted background yields across all the E_T^γ bins in the two signal regions are shown in Fig. 3-15. The expected yields together with the covariances can be used with the simplified likelihood approach detailed in Ref. [?] to reinterpret the results for models not studied in this thesis

.

3.9.2 Limits

Figure 3-16 shows the 95% CL upper cross section limits with respect to the corresponding theoretical cross section ($\mu_{95} = \sigma_{95\%}/\sigma_\mathrm{theory}$) for the vector and axial-vector mediator scenarios, in the M_med–m_DM plane. The solid black (dashed red) curves are the observed (expected) contours of $\mu_{95} = 1$. The σ_theory hypothesis is excluded at 95% CL or above in the region with $\mu_{95} < 1$. The uncertainty in the expected upper limit includes the experimental uncertainties. For the simplified DM LO models considered, mediator masses up to $950\,\mathrm{GeV}$ are excluded for values of m_DM less than $1\,\mathrm{GeV}$.
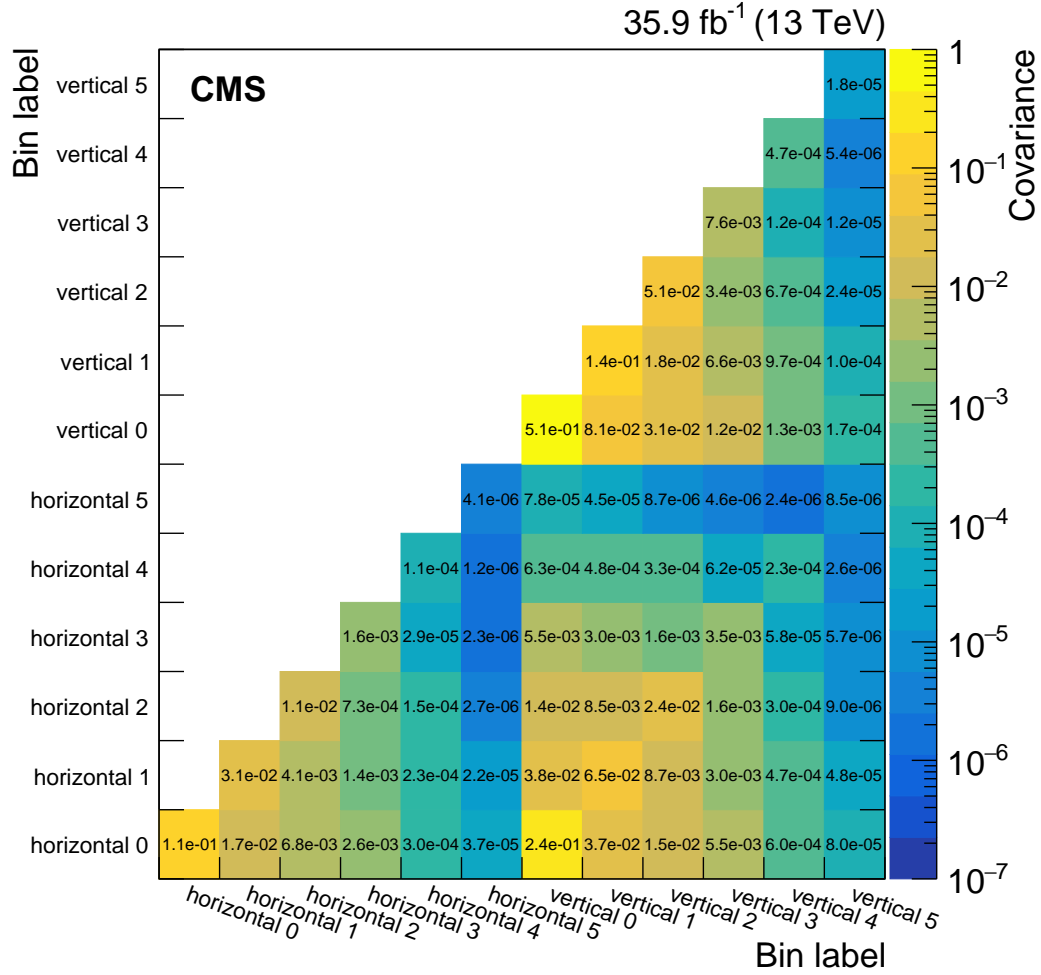
Figure 3-15: Covariances between the predicted background yields in all the E_{T}^{γ} bins of the horizontal and vertical signal regions. The bin labels specify which signal region the bin belongs to and what number bin it is for that region.
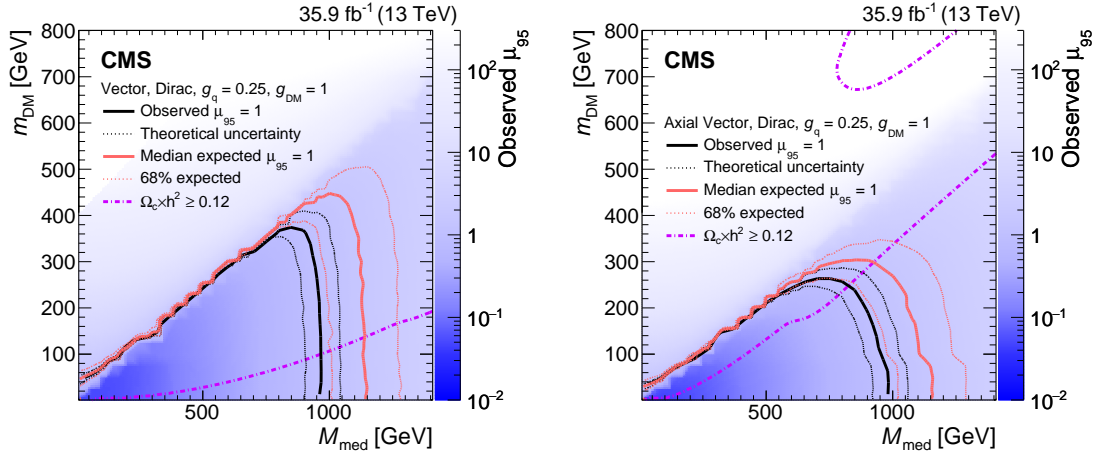
Figure 3-16: The ratio of 95% CL upper cross section limits to the theoretical cross section (μ_{95}), for DM simplified models with vector (left) and axial-vector (right) mediators, assuming $g_q = 0.25$ and $g_{DM} = 1$. Expected $\mu_{95} = 1$ contours are overlaid in red. The region under the observed contour is excluded. For DM simplified model parameters in the region below the lower violet dot–dash contour, and also above the corresponding upper contour in the right hand plot, cosmological DM abundance exceeds the density observed by the Planck satellite experiment.

Chapter 4

Comparison with Other Results

We're not doing this in a vacuum.

4.1 Monophoton

4.2 Monojet / Mono-Z

4.3 Direct Detection

We show the results in Fig. 4-1.

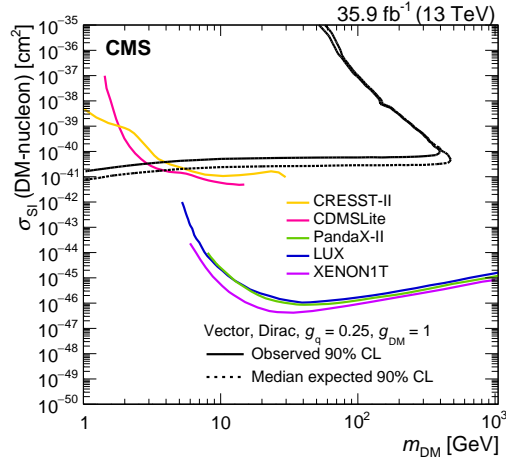## 4.4 Indirect Detection

We show the results in Fig. 4-2.

Figure 4-1: The 90% CL exclusion limits on the χ–nucleon spin-independent scattering cross sections involving the vector operator as a function of the m_{DM}. Simplified model DM parameters of $g_q = 0.25$ and $g_{\mathrm{DM}} = 1$ are assumed. The region to the upper left of the contour is excluded. On the plots, the median expected 90% CL curve overlaps the observed 90% CL curve. Also shown are corresponding exclusion contours, where regions above the curves are excluded, from the recent results by the CDMSLite [?], LUX [?], PandaX-II [?], XENON1T [?], and CRESST-II [?].
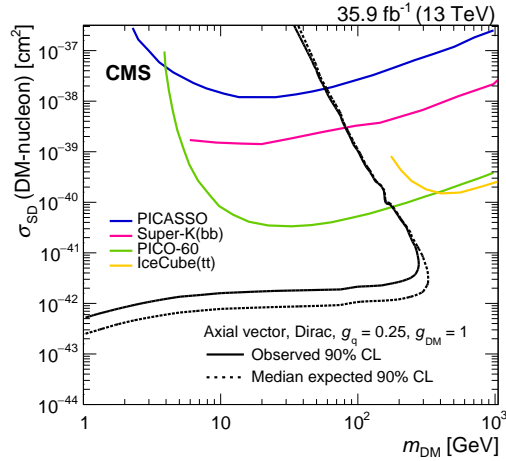

Figure 4-2: The 90% CL exclusion limits on the χ–nucleon spin-dependent scattering cross sections involving the axial-vector operator as a function of the m_{DM}. Simplified model DM parameters of $g_q = 0.25$ and $g_{\mathrm{DM}} = 1$ are assumed. The region to the upper left of the contour is excluded. On the plots, the median expected 90% CL curve overlaps the observed 90% CL curve. Also shown are corresponding exclusion contours, where regions above the curves are excluded, from the recent results by the PICO-60 [?], IceCube [?], PICASSO [?] and Super-Kamiokande [?] Collaborations.