Studies on the background estimation for the measurement of $Z(\to \ell^+\ell^-)\gamma$ differential cross-sections

DESY SUMMER SCHOOL PROJECT
September 4, 2019

Joschka Birk, University of Freiburg

supervised by

Kurt Brendlinger, Beate Heinemann, Yee Chinn Yap

Abstract

The dominant background in the ATLAS $Z(\to \ell\ell)\gamma$ cross-section measurement at $\sqrt{s} = 13\,\text{TeV}$ is the Z+jets background. The method used to estimate this background is the 2D sideband method which is based on the discrimination between jets and photons using the photon identification variables and the photon isolation variable. The studies presented in this report are focused on the behaviour of the correlation between these discriminating variables as a function of both the transverse momentum p_{T}^{γ} and the pseudorapidity η_{γ} of the photon candidate. Photon candidates with high p_{T}^{γ}, high η_{γ} and a matching conversion vertex are found to be most problematic for the background estimation. Efforts to mitigate the issues related to this method, and reduce the corresponding systematic uncertainty, are discussed.

Contents

1 Introduction

The measurement of the production of a Z boson in association with a prompt photon performed by the ATLAS experiment [1] is a precision measurement that is compared with theoretical predictions. The measurement is performed using the data recorded by the ATLAS Detector [2] at the LHC between 2015 and 2018 (Run 2). During that time, the LHC provided proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, and the integrated luminosity of the full dataset used in the measurement is $139\,\text{fb}^{-1}$.

The measurement is carried out with a sample dominated by events where the photon is emitted from a quark leg in the initial state as shown in Figure 1a, which is achieved by requiring the invariant mass of the two leptons $m(\ell\ell)$ to be larger than 40 GeV and the sum $m(\ell\ell)+m(\ell\ell\gamma)$ to be larger than 182 GeV where the second term in the sum is the invariant mass of the $\ell^+\ell^-\gamma$ system. This selection removes most of the events with photons radiated from one of the leptons in the final state, which is illustrated in Figure 1b. A graphical demonstration of the cut on the sum $m(\ell\ell) + m(\ell\ell\gamma)$ can be seen in Figure 2, which shows a scatter plot of $m(\ell\ell)$ and $m(\ell\ell\gamma)$ for $Z\gamma$ events with two muons. The scatter plot shows that there are two dominant populations, one located at $m(\mu\mu) \approx m_Z$, representing the initial state radiation events and one located at $m(\mu\mu\gamma) \approx m_Z$, which corresponds to the events from the final state radiation.

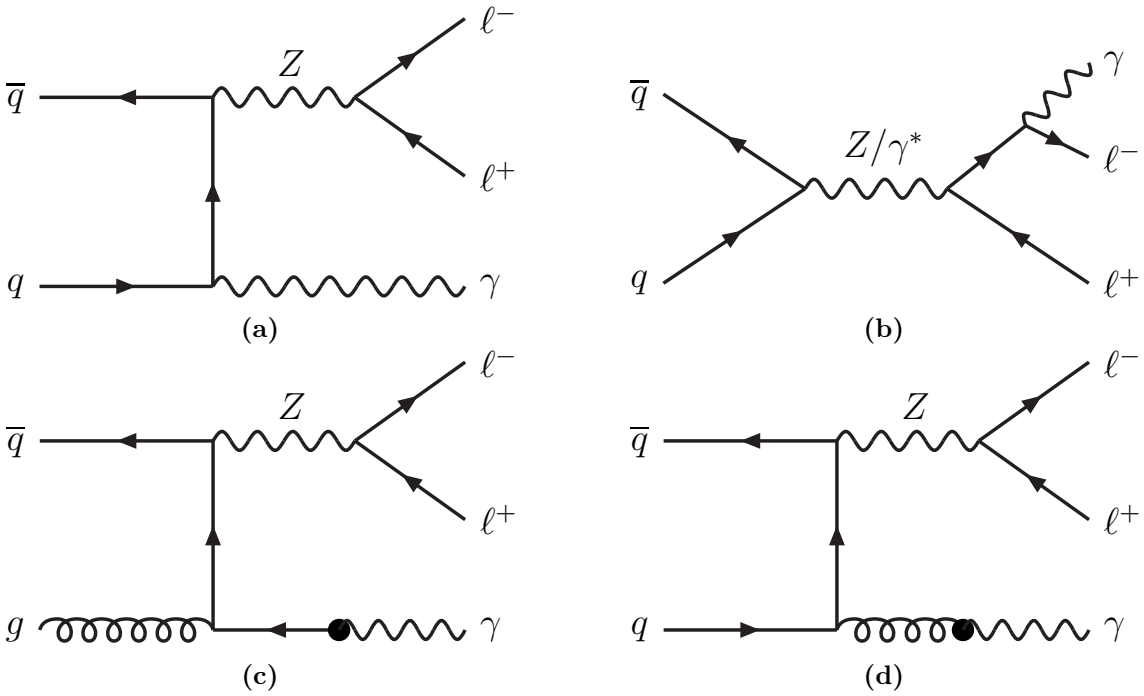


Figure 1: Feynman diagrams of the different processes for $\ell^+\ell^-\gamma$ production: **(a)** $q\bar{q}$ annihilation with photon radiation at a quark leg; **(b)** $q\bar{q}$ annihilation with photon radiation from a final state lepton; **(c)** and **(d)** Z + $q(g)$ processes in which a photon is produced from the fragmentation of the quark or gluon [1].

The events that originate from processes in which the photon is produced from fragmentation of either a quark or a gluon, shown in Figure 1c and Figure 1d, are removed by requiring the photons to be isolated, which means that they are not accompanied by much nearby QCD activity.
The events used for the analysis are required to contain a high-energy photon together

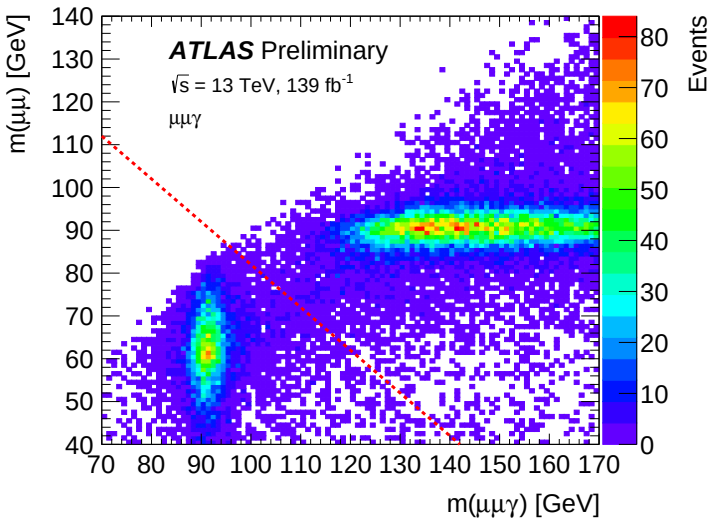


Figure 2: A scatter plot of the two variables $m(\ell\ell)$ and $m(\ell\ell\gamma)$ for the muons which satisfy all remaining selection criteria. The plot shows that there are two dominant populations, one located at $m(\mu\mu) \approx m_Z$, representing the initial state radiation events and one located at $m(\mu\mu\gamma) \approx m_Z$, which corresponds to the events from the final state radiation. The cut $m(\ell\ell) + m(\ell\ell\gamma) > 182\,\text{GeV}$ is shown as a dashed red line [1].

with an opposite-sign, same-flavour lepton (electron or muon) pair. The corresponding photon and electron candidates have to satisfy several isolation and identification (ID) criteria which are not discussed in detail here. Furthermore, there are several cuts concerning the kinematics of the photon and lepton candidates. The resulting event selection is summarised in Table 1.

However, even though the measurement considers only events with $p_{\text{T}}^{\gamma} > 30\,\text{GeV}$, the studies presented in this report are performed using a lower cut of $p_{\text{T}}^{\gamma} > 15\,\text{GeV}$ to also investigate the events with low photon transverse momentum, in order to investigate the possibility of including this region of phase space in future measurements.
In order to perform a precision measurement of the cross section

$$\sigma_{pp\to\ell\ell\gamma} = \frac{N_{\text{sig}}}{C \cdot \int \mathcal{L}\, dt}\,, \tag{1}$$

where C represents a factor which takes detector inefficiencies and resolution into account, the number of signal events N_{sig} has to be determined with a very high precision. To obtain the number of signal events in the signal region, the number of background events N_{bkg} that pass the event selection has to be subtracted from the total number N_{obs} of selected events in the signal region

$$N_{\text{sig}} = N_{\text{obs}} - N_{\text{bkg}}\,. \tag{2}$$

Therefore, the precision of the number of signal events and consequently the precision of the measured cross section is strongly influenced by the uncertainty of the background estimation. In this measurement, the dominant background is the Z + jets background, which are processes including the production of a Z boson in which a jet is misidentified as a photon. This background is estimated using a data-driven method that uses the photon ID and photon isolation as discriminating variables between jets and photons. In

Table 1: Definition of the $\ell^+\ell^-\gamma$ signal region. The photon and lepton candidates are required to satisfy cuts on the kinematics as well as criteria concerning the identification and isolation of the candidate. The event selection itself is just requiring the invariant mass of the two leptons as well as the sum of that mass and $m(\ell\ell\gamma)$ to be larger than a given threshold.

	Photons	Electrons	Muons
Kinematics	$p_{\mathrm{T}} > 30\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.37$ excl. $1.37 < \lvert\eta\rvert < 1.52$	$p_{\mathrm{T}} > 30,25\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.37$ excl. $1.37 < \lvert\eta\rvert < 1.52$	$p_{\mathrm{T}} > 30,25\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.5$
Identification Isolation	Tight FixedCutLoose [3] $\Delta R(\ell,\gamma) > 0.4$	Medium FCLoose [4] $\Delta R(\mu, e) > 0.2$	Medium FCLoose_FixedRad [5]
Event selection	$m(\ell\ell) > 40\,\mathrm{GeV}$, $m(\ell\ell) + m(\ell\ell\gamma) > 182\,\mathrm{GeV}$		

this report, the term background refers to the Z+jets background contribution while other background contributions are specified if mentioned. The method used for the background estimation, called the 2D sideband method, uses the assumption that the photon ID and the photon isolation are uncorrelated variables. However, even if these two variables have a small correlation, this method can still be applied taking the corresponding corrections due to the correlation into account.

This report presents studies on this background estimation with a focus on the correlation between the photon ID and the photon isolation. The general concept of photon identification and photon isolation is briefly explained in Section 2, followed by a short explanation of the 2D sideband method in Section 3. The results of the studies are presented in Section 4.

2 Photon identification and isolation

Photon identification is performed via cuts on different variables describing the photon shower development in the calorimeter, or "shower shapes". Additionally, photon candidates in the signal region need to pass an isolation criteria describing the hadronic activity around the photon within a cone [1] of specific size. This chapter gives a brief introduction to the photon identification performed in ATLAS based on Reference [3].

2.1 Photon ID

The identification of photon candidates is carried out using different variables providing a good separation between prompt photons and non-prompt photons or hadronic jets. A schematic representation of all the used variables is shown in Figure 3. While prompt photons are expected to have pretty narrow shower shapes, non-prompt photons originating from the decay of neutral hadrons, for example pions, are expected to show broad shower shapes. There are two reference sets of cuts, called *loose'* and *tight* although in this report

[1] The coordinate system used in ATLAS is a right-handed coordinate system with the origin at the centre of the detector and the z-axis along the beam line. Points in the transverse plane (perpendicular to the beamline) are described by cylindrical coordinates (r,ϕ), with ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$, with θ being the polar angle measured from the z-axis. Angular distances are measured in units of ΔR, which is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

the loose' set of cuts is also referred to as *non-tight*. The tight photon identification requires the photon candidate to pass all the cuts on the shower shape variables as described in [3]. The loose' or non-tight photon ID corresponds to photons which do not pass the tight ID but still satisfy the cuts of a "looser" identification, which means that some variables required in the tight ID are relaxed.

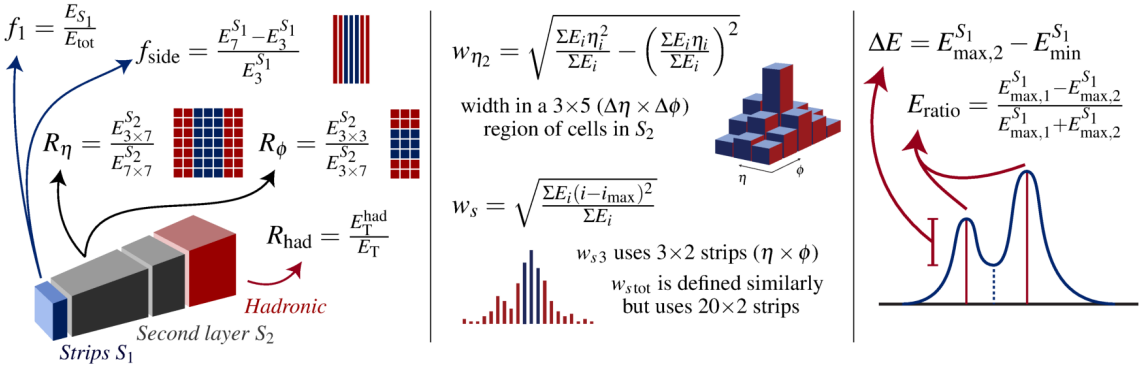


Figure 3: Schematic representation of the variables used in the photon ID from Reference [6]. The variables describe the shower shape of the photon candidate in the electromagnetic calorimeter. $E_X^{S_N}$ is the electromagnetic energy collected in the N-th longitudinal layer of the electromagnetic calorimeter in a cluster of properties X.

2.2 Isolation

The isolation of photon candidates is defined via a *track isolation* and a *calorimeter isolation*.

The track isolation is defined using the sum $p_{\text{T}}^{\text{iso}}$ of the scalar transverse momenta of tracks with $p_{\text{T}} > 1\,\text{GeV}$ measured within a cone of $\Delta R = 0.2$ around the photon direction. The momentum of tracks associated to the photon itself (e.g. in the case of a converted photon, see Section 2.3) is not included in the calculation of the sum. To pass the track isolation, the fraction of this sum of transverse momenta and the transverse momentum of the photon itself has to be smaller than five percent

$$\frac{p_{\text{T}}^{\text{iso}}}{p_{\text{T}}^{\gamma}} < 0.05\,. \tag{3}$$

The calorimeter isolation of a photon is defined using the variable $E_{\text{T}}^{\text{cone20}}$, which is the sum of transverse energy of topological clusters measured in the same cone as $p_{\text{T}}^{\text{iso}}$. However, as for the track isolation, a specific volume closer to the photon is not included in this sum since this variable is intended to measure the hadronic activity nearby the photon but should not include contributions originating from the photon itself, which is schematically shown in Figure 4. Prompt photons are not expected to be accompanied by a lot of nearby activity in the electromagnetic calorimeter while non-prompt photons are expected to show a lot of activity in terms of measured energy in the calorimeter within the cone of $\Delta R = 0.2$. Therefore, the calorimeter isolation is defined via the criteria

$$\frac{E_{\text{T}}^{\text{cone20}}}{p_{\text{T}}^{\gamma}} < 0.065\,. \tag{4}$$

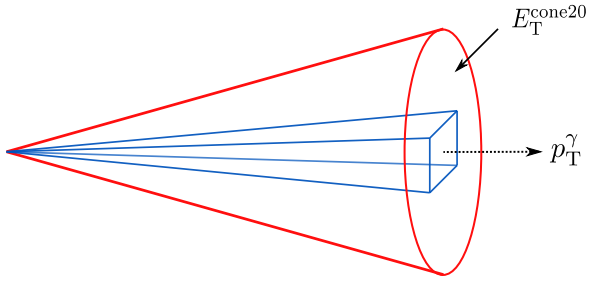


Figure 4: Schematic representation of the isolation cone.

2.3 Photon conversion

Photons measured in the ATLAS Detector are reconstructed from energy clusters in the electromagnetic calorimeter (ECAL). Since photons can convert into electron-positron pairs by interacting with the detector material before entering the ECAL, photons are classified as either *converted* or *unconverted* depending on if there is a matching reconstructed conversion vertex or a matching track which could originate from a photon conversion. This is schematically shown in Figure 5. The distance ΔR of the electron-positron-pair is typically very small, which means that the converted photon will be reconstructed as a single topological energy cluster [7] in the ECAL.

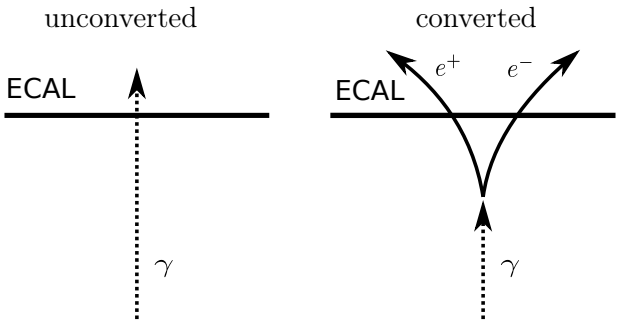


Figure 5: Schematic representation of a unconverted photon (left) and a converted photon (right) entering the electromagnetic calorimeter. The sketch represents the transverse plane (perpendicular to the beam axis) of the ATLAS detector.

3 Background estimation using the 2D sideband method

The 2D sideband method is a method commonly used to perform a data-driven background estimation. The selected events are distributed in a plane generated by the photon ID variable and the photon isolation variable. The isolation variable is the calorimeter isolation and the event selection includes the requirement that all events satisfy the track isolation criteria. This plane is then divided into four different regions: the signal region A which is dominated by signal events and three control regions B, C and D. A schematic representation of these regions is shown in Figure 6 as well as the corresponding normalised distributions of the Z+jets background in these regions. The general idea of the method is then to estimate the number of background events in the signal region (A) using the number of background events observed in data in the control regions (B, C and D).
The four regions are defined by a cut on the isolation variable $E_{\mathrm{T}}^{\mathrm{cone20}} - 0.065\, p_{\mathrm{T}}^{\gamma}$ and the (binary) variable of the photon ID. The latter variable splits the sample into events with

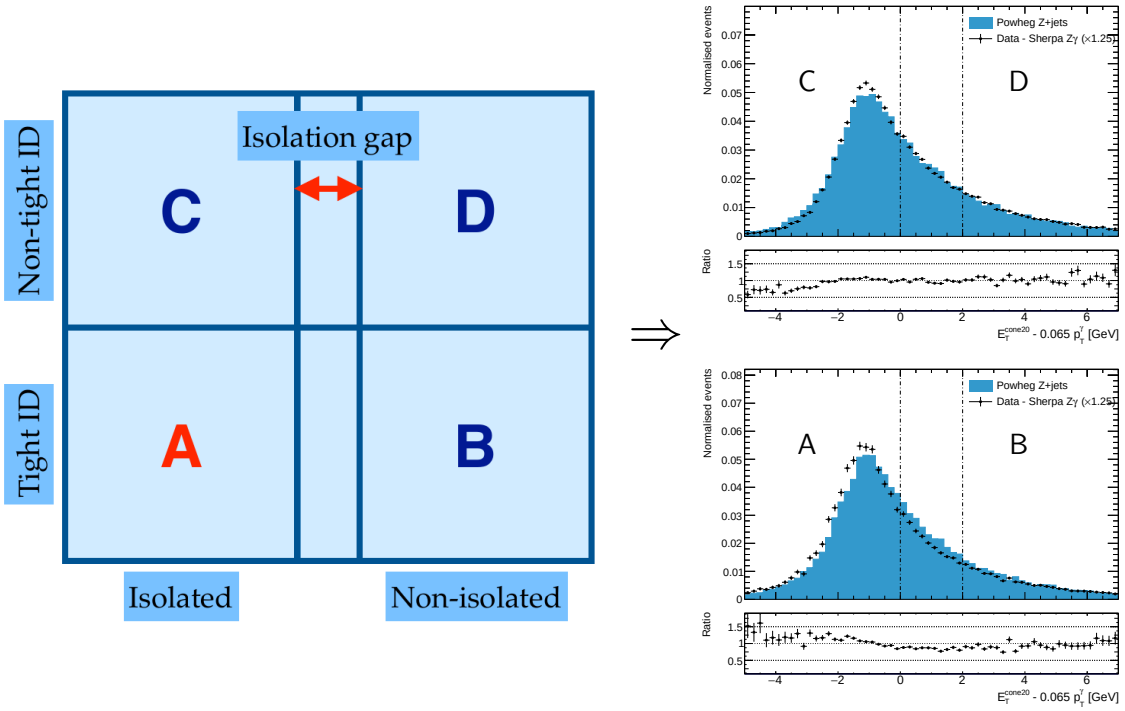


Figure 6: Left: Schematic representation of the 2D plane used for the background estimation. The signal region A is expected to be dominated by signal while the control regions B, C and D are dominated by background events. This allows to use the number of background events observed in the control regions B, C and D to calculate an estimate for the number of background events in the signal region A. Right: Normalised isolation distributions for tight and non-tight ID for Z+jets events. The blue distribution shows the Z+jets MC simulation and the black points show the normalised distribution of data after subtracting the $Z\gamma$ MC simulated sample.

photons satisfying the tight photon ID and the events with the photons corresponding to the non-tight (loose') photon ID. The cut on the isolation variable is performed by cutting on $E_{\mathrm{T}}^{\mathrm{cone20}} - 0.065\, p_{\mathrm{T}}^{\gamma} < 0$ to obtain isolated photons, which is equivalent to the isolation criteria in Equation 4, and cutting on $E_{\mathrm{T}}^{\mathrm{cone20}} - 0.065\, p_{\mathrm{T}}^{\gamma} > 2\,\mathrm{GeV}$ for the non-isolated photons. The isolation gap of $E_{\mathrm{gap}} = 2\,\mathrm{GeV}$ is used to avoid signal leakage into the control regions B and D since these are supposed to be background dominated.
The resulting regions are:

- Signal region A: The photon candidate satisfies the isolation criteria and the tight identification criteria.
- Control region B: The photon candidate is not isolated but passes the tight identification criteria.
- Control region C: The photon candidate is isolated but fails the tight identification (but passes the loose' identification).
- Control region D: The photon candidate is not isolated and fails the tight identification (but passes the loose' identification).

The general assumption in the 2D sideband method is that the two discriminating variables are uncorrelated for the Z+jets background events, which means

$$\frac{N_A^{Z+\text{jets}}}{N_B^{Z+\text{jets}}} = \frac{N_C^{Z+\text{jets}}}{N_D^{Z+\text{jets}}} \tag{5}$$

with $N_X^{Z+\text{jets}}$ being the number of Z+jets events in region X with $X = A, B, C, D$. The accuracy of this assumption is is measured with the correlation R defined as

$$R = \frac{N_A^{Z+\text{jets}} \times N_D^{Z+\text{jets}}}{N_B^{Z+\text{jets}} \times N_C^{Z+\text{jets}}}\,, \tag{6}$$

which is equal to 1 if Equation 5 is correct, i.e. if there is no correlation between the photon ID and the photon identification for the Z+jets background events. To account for the signal leakage in the control regions, signal leakage factors are defined by relating the amount of signal in each control region relative to the amount observed in the signal region:

$$c_B = \frac{N_B^{\text{sig}}}{N_A^{\text{sig}}}\,, \quad c_C = \frac{N_C^{\text{sig}}}{N_A^{\text{sig}}} \quad \text{and} \quad c_D = \frac{N_D^{\text{sig}}}{N_A^{\text{sig}}}\,. \tag{7}$$

Using these signal leakage factors and the number N_X^{bkg} of non-Z+jets background events in the control regions, the number of signal events in region A can be calculated as

$$\begin{aligned} N_A^{\text{sig}} &= N_A - N_A^{\text{bkg}} - N_A^{Z+\text{jets}} \\ &= N_A - N_A^{\text{bkg}} - R\,\frac{\left(N_B - N_B^{\text{bkg}} - c_B\,N_A^{\text{sig}}\right)\cdot\left(N_C - N_C^{\text{bkg}} - c_C\,N_A^{\text{sig}}\right)}{N_D - N_D^{\text{bkg}} - c_D\,N_A^{\text{sig}}}\,. \end{aligned} \tag{8}$$

This equation contains N_A^{sig} but expanding the equation and rearranging of the terms leads to a quadratic formula which can be solved for N_A^{sig}. The resulting formula then contains the correlation R, the signal leakage factors c_B, c_C, c_D, the total number of events in the four regions N_X and the (MC-estimated) numbers N_X^{bkg} of non-Z+jets background events in the four regions.

The strategy is then to use the MC sample of the Z+jets background (generated with Powheg [8]) to obtain the correlation R as well as using the MC sample of the $Z\gamma$ signal (generated with Sherpa [9]) to calculate the signal leakage factors. Afterwards, the numbers of total events in the four regions used for the calculation of N_A^{sig} is the number of observed data events in these regions, which leads to a data-driven background estimate. The MC samples of both signal and background are normalised to the corresponding cross section of the predicted sample.

The calculation of the background estimate in the signal region is applied separately within each bin of a given observable for the measurement of the corresponding differential cross section for that observable. Nevertheless, for all the bins the same value is used for the correlation which is calculated inclusively. This means that all phase space regions are assumed to show the same correlation between isolation and identification.
The uncertainty of the Z+jets estimate is calculated using the statistical uncertainties of the numbers of events in the regions A, B, C, D in data as well as the statistical and systematic uncertainties assigned to the signal leakage factors and the correlation R. The systematic uncertainty assigned to R is quite large in order to cover the fluctuations seen when calculating R in bins of p_{T}^{γ} and η_{γ}. These fluctuations are discussed in more detail in Section 4.1. However, the current implementation of the background estimation shows

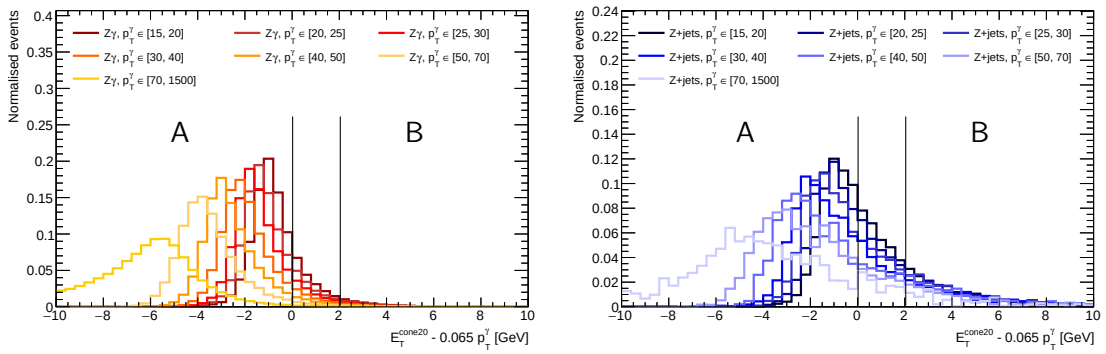
a large uncertainty of the purity, which is defined as the fraction of estimated signal events and overall measured events in the signal region A, for events with very low-transverse-momentum photon candidates. Therefore, the events used in the measurement are required to satisfy the criteria $p_{\mathrm{T}}^{\gamma} > 30\,\mathrm{GeV}$.

4 Results

The studies presented in the following section show investigations of different variables used in the Z+jets background estimation, with a focus on the correlation R between the photon ID and the photon isolation. The overall goal of these studies is to improve the accuracy of the 2D sideband method and to reduce the systematic uncertainty originating from this method.

4.1 Studies of the isolation distribution and signal leakage

As the distribution of $E_{\mathrm{T}}^{\mathrm{cone20}}$ changes with the transverse momentum of the photon candidate and because p_{T}^{γ} itself is part of the isolation variable, the isolation distribution changes its shape for different p_{T}^{γ} ranges, which is shown in Figure 7 for the $Z\gamma$ MC and Z+jets MC. The isolation gap is also shown in the plot to visualise the amount of signal leakage which is rejected by the gap. The signal leakage is supposed to be limited since it is also contributing to the uncertainty of the final background estimation. On the other hand, the isolation gap should not be chosen too large as this would lead to small numbers of Z+jets events in the control regions B and D and thereby increase the uncertainty induced by the statistics in these regions. Looking at the shape of the $Z\gamma$ distributions in Figure 7a, it



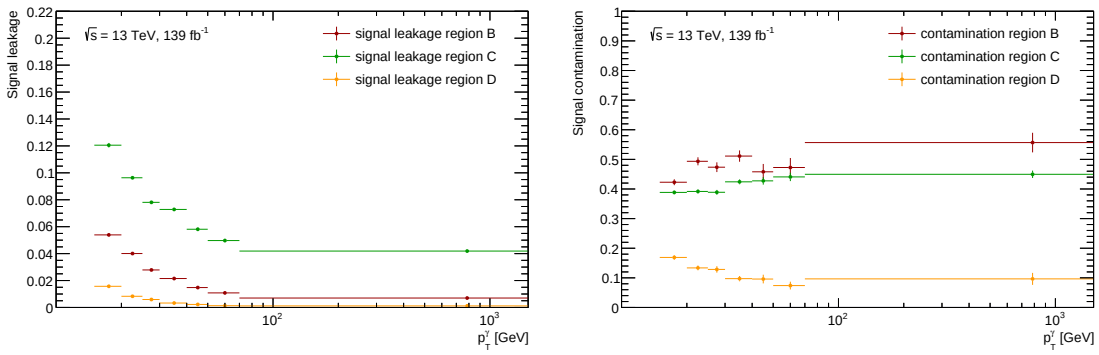
(a) **(b)**

Figure 7: Normalised distributions of the isolation variable for the tight photons (regions A and B). The different colors show the distribution for different ranges in p_{T}^{γ} with lighter colors representing higher p_{T}^{γ}.

seems that the signal leakage becomes less of a problem for high transverse momentum of the photon. However, this is only true when comparing the amount of signal in the given control region to the corresponding number of signal events in region A, which is the definition of the signal leakage factors in Equation 7. When considering the number of signal events in a control region with the number of $Z\gamma$ events and Z+jets events in that region, called signal contamination

$$\text{sig. contamination} = \frac{N_X^{\mathrm{sig}}}{N_X^{\mathrm{sig}} + N_X^{Z+\mathrm{jets}}}\,, \tag{9}$$

the leakage of signal events appears to be very relevant at high p_{T}^{γ} since the signal events make up around half of the events in control region B for example. This can be seen in Figure 8, which shows both the signal leakage and the signal contamination in the control regions as a function of the transverse momentum of the photon. The calculation is performed in bins of p_{T}^{γ} and the uncertainty shown in the plots represents the statistical uncertainty.



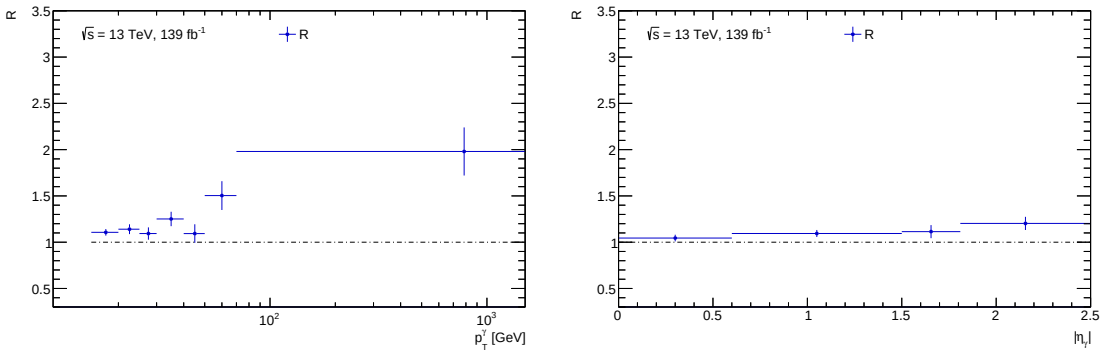
(a) **(b)**

Figure 8: Correlation R as a function **(a)** of the photon transverse momentum and **(b)** of the pseudorapidity of the photon. Only statistical uncertainties are included in the error bars.

4.2 Studies on the correlation R

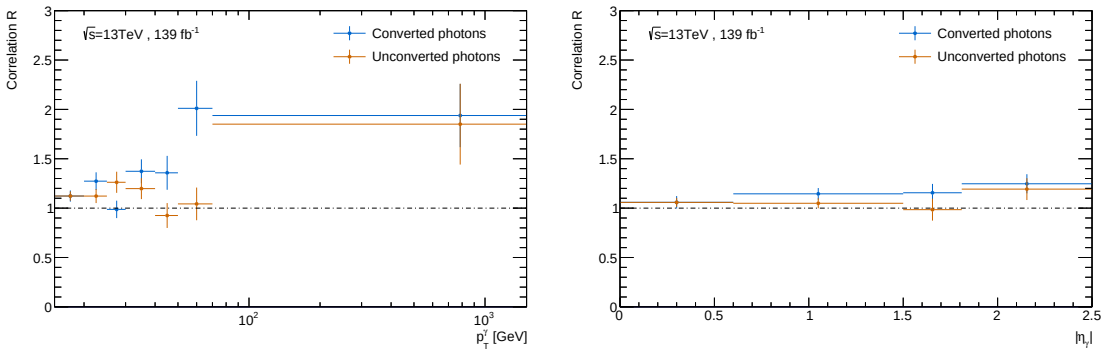
In order to investigate the correlation R of the photon ID and the isolation, R is calculated as a function of both the transverse momentum and the pseudorapidity of the photon. As can be seen in Figure 9a, the correlation seems to increase for higher p_{T}^{γ}, especially the last bin shows a quite large correlation close to 2. However, the correlation seems to be relatively stable with respect to the pseudorapidity of the photon, which is shown in Figure 9b. Note that the stability of the R value as a function of pseudorapidity is mainly a reflection of the behavior at low p_{T}^{γ}, because the sample overwhelmingly contains events with low-p_{T}^{γ} photon candidates. The p_{T}^{γ} distributions of both data and MC are shown in Section 4.4.

In order to see if the p_{T}^{γ} or η_{γ} dependence of R is different for unconverted and converted photons, the samples are split into subsamples of unconverted and converted photons. Afterwards, the calculations of the correlation as a function of p_{T}^{γ} and η_{γ} are repeated for both subsamples. The corresponding plots are shown in Figure 10. Comparing the p_{T}^{γ} dependence for unconverted and converted photons in Figure 10a, the correlation tends to depend more on the photon transverse momentum for converted photons. While the correlation is distributed around a value of 1.0 for all bins except the last bin for unconverted photons, the correlation seems to be increasing with p_{T}^{γ} for the subsample containing the converted photon candidates. This indicates that the correlation seems to be more of a problem for photon candidates that convert before entering the ECAL.
The η_{γ} dependence in Figure 10b of the correlation, on the other hand, seems to be unaffected by this split of the sample. For both unconverted and converted photons the correlation still seems to be quite unaffected by the pseudorapidity of the photon.



(a) **(b)**

Figure 9: Correlation R as a function **(a)** of the photon transverse momentum and **(b)** of the pseudorapidity of the candidate photon. The uncertainties shown in the plot are only the statistical uncertainties.



(a) **(b)**

Figure 10: Correlation R as a function **(a)** of the transverse momentum and **(b)** of the pseudorapidity of the photon candidate for unconverted (orange) and converted (blue) candidates.

However, since the correlation shows some interesting behaviour for high p_{T}^{γ} and because the measurement of the $Z\gamma$ cross section in Reference [1] only uses events with $p_{\mathrm{T}}^{\gamma} > 30\,\mathrm{GeV}$, the η_{γ} dependence is also calculated separately for unconverted and converted photons in events with this p_{T}^{γ} cut. The resulting graphs of the correlation, which are shown in Figure 11, are quite different from the previous plots of the η_{γ} dependence. The correlation as a function of $|\eta_{\gamma}|$ starts to fluctuate more for the unconverted photons, although this can be explained by much smaller statistics and R still is quite constant and close to 1.0 within the statistical uncertainties. The corresponding plot for the converted photons, however, shows a slight trend of increasing correlation for higher η_{γ}. Therefore, converted, high-transverse-momentum and high-η_{γ} photon candidates seem to be the most problematic events in terms of the correlation in the 2D sideband method and thereby also in terms of the background estimation itself.

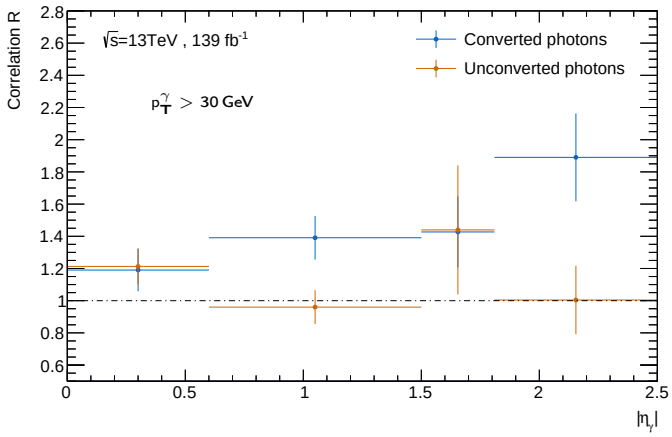


Figure 11: Correlation R as a function of the pseudorapidity η_γ of photon candidates which carry a large transverse momentum ($p_{\mathrm{T}}^{\gamma} > 30\,\mathrm{GeV}$). The orange graph shows the relation for the unconverted and the blue graph for the converted photon candidates.

4.3 Loose isolation and gap in photon ID

In the same way that the photon ID has its non-tight definition which means that the photon candidates need to pass some looser requirements on the shower shape variables, a loose isolation requirement is tested for the isolation. This is intended to remove very non-isolated photon candidates, which may have a stronger correlation with the photon ID variables than the bulk of the isolation distribution. Two different definitions for a loose isolation requirement are tested, namely $E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}}^{\gamma} < 0.2$ and $E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}}^{\gamma} < 0.6$. The resulting graphs of R as a function of p_{T}^{γ} are shown in Figure 12 together with the graph corresponding to having no loose isolation requirement. The values of the correlation do not change much when comparing the three different graphs in the plot. Therefore, introducing the loose isolation criteria does not seem to improve the p_{T}^{γ} dependence and thus removing these very unisolated events at the expense of statistics does not seem to be useful.

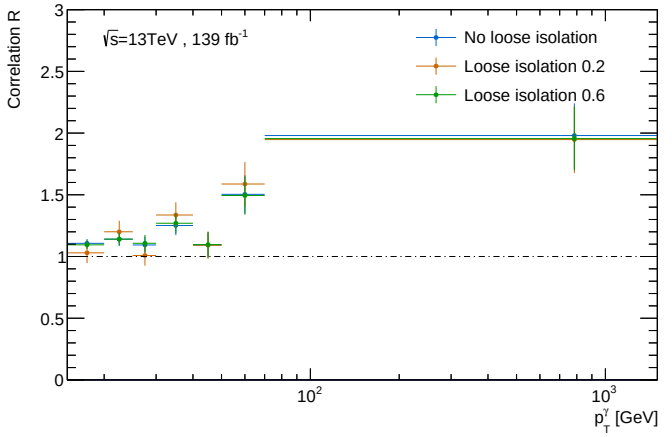
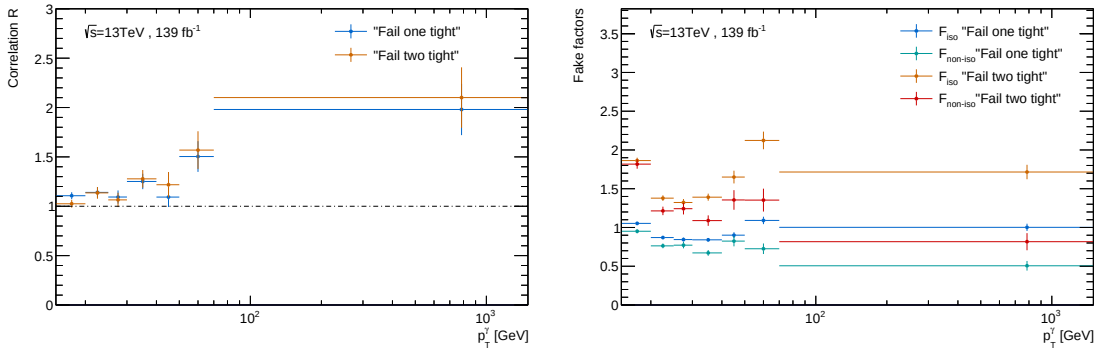
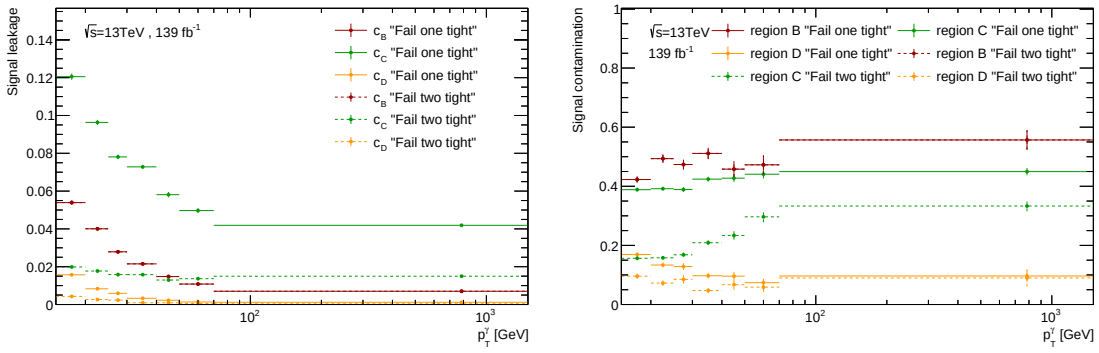


Figure 12: The p_{T}^{γ} dependence of the correlation without a loose isolation definition (blue), with the loose isolation requirement being $E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}}^{\gamma} < 0.2$ (orange) and $E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}}^{\gamma} < 0.6$ (green).



(a) **(b)**

Figure 13: p_{T}^{γ} dependence **(a)** of the correlation R and **(b)** of the fake factors for both the normal case using the LoosePrime4 i.e. requiring the photon candidates to fail only one of the tight cuts but pass all the cuts in LoosePrime4 (bluish colours) and the case with the additional requirement that non-tight events have to fail at least two of the tight cuts to be classified as non-tight (shown in the reddish colours).



(a) **(b)**

Figure 14: p_{T}^{γ} dependence **(a)** of the signal leakage and **(b)** of the signal contamination for both the normal case using the LoosePrime4 i.e. requiring the photon candidates to fail only one of the tight cuts (but pass all the cuts in LoosePrime4), and the case with the additional requirement that non-tight events have to fail at least two of the tight cuts to be classified as non-tight. The solid lines show the normal case and the dotted lines show the case with requiring candidates to fail two or more tight cuts.

In addition, to see if introducing some kind of gap between the tight photons and the non-tight photons has an effect on the correlation, the definition of the non-tight regions are slightly modified: instead of requiring these photon candidates to fail one or more of the tight cuts, they are required to fail at least two of the tight cuts. In this way, events that fail only one of the tight cuts are removed from the regions C and D. The resulting graph of the correlation as a function of p_{T}^{γ} is shown in Figure 13a. The changes of the R values are again only very small. However, this removes many of the events in the regions C and D, which has a large effect on the so-called fake factors F_{iso} and $F_{\text{non-iso}}$, defined as

$$F_{\mathrm{iso}} = \frac{N_A^{Z+\mathrm{jets}}}{N_C^{Z+\mathrm{jets}}} \quad \text{and} \quad F_{\text{non-iso}} = \frac{N_B^{Z+\mathrm{jets}}}{N_D^{Z+\mathrm{jets}}} \ . \tag{10}$$

In some bins of p_{T}^{γ} these fake factors change by a factor of approximately 2, which can be seen in Figure 13b. This is a very significant change, especially because these values are usually preferred to be smaller than 1. However, at the same time the signal leakage and the signal contamination improve significantly by introducing this additional requirement for the non-tight photons, which can be seen in Figure 14. Using the modified definition of the non-tight regions would lead to smaller signal leakage-induced uncertainties in the final background estimate but at the same time could lead to larger uncertainty contributions induced by the increased statistical uncertainty of the correlation.

4.4 p_{T}^{γ} distributions

Since the studies of the correlation show a p_{T}^{γ} dependence of R, the distribution of p_{T}^{γ} itself is investigated in more detail as a wrong modelling of p_{T}^{γ} in MC could bias the background estimation. The distributions are shown in Figure 15 for all four regions used in the 2D sideband method for both MC and data.

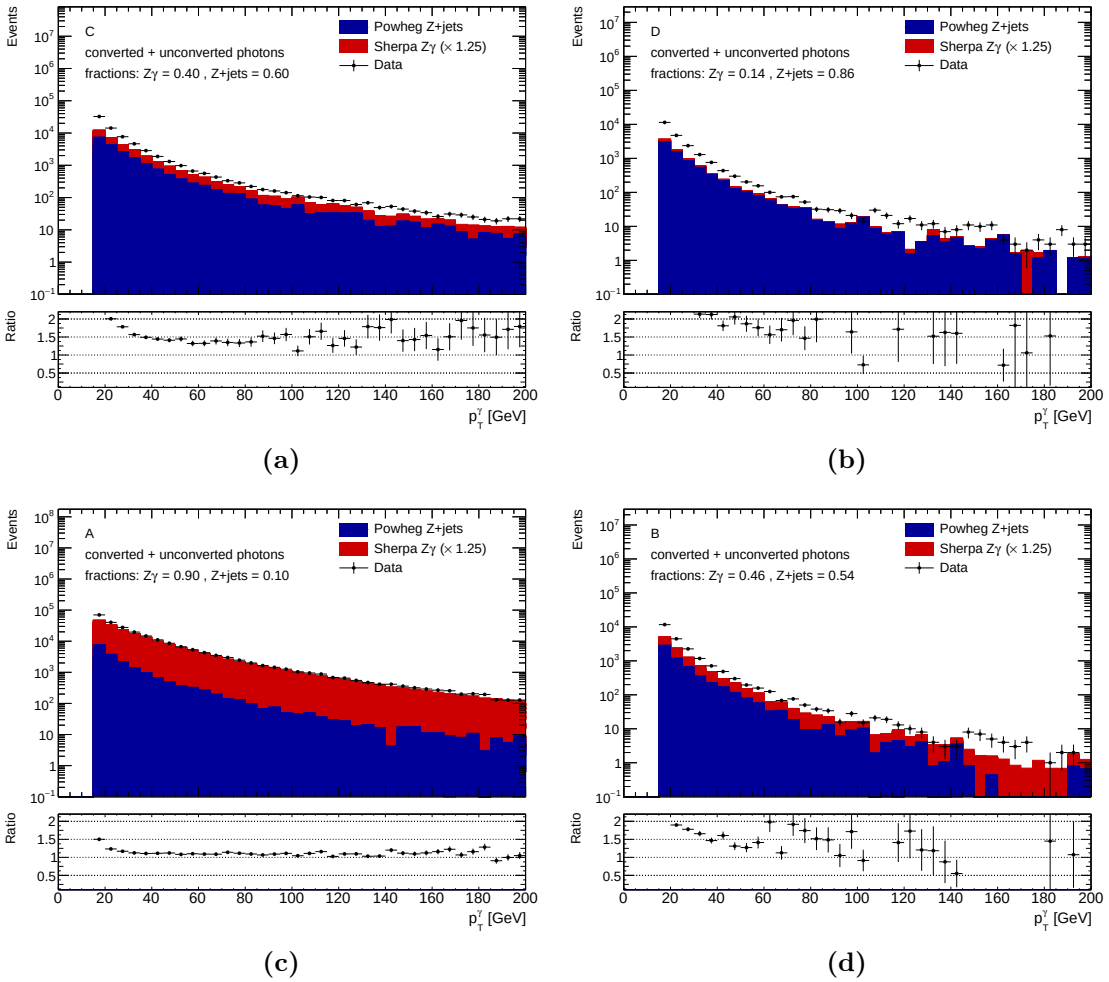


Figure 15: p_{T}^{γ} distribution in the four different regions of the 2D sideband method. The red and blue stacked histograms show the MC prediction of $Z\gamma$ signal and Z+jets background and the black points show the p_{T}^{γ} distribution observed in data. The ratio graphs in the bottom panel of all four plots show the ratio between data and the (stacked) MC.

Even though there are some contributions not included in the MC, e.g. $t\bar{t}\gamma$ and $Z \to \ell\ell$ events with a photon candidate from a pile-up event, the MC seems to underestimate the data - especially at low p_{T}^{γ}. As a rough estimate for the number of Z+jets events obtained by the data driven method, the background estimation is performed unbinned, with an inclusive $R = 1.14$ and neglecting the non-Z+jets background contributions. Comparing the resulting number of events in A obtained by this calculation with the number of events in region A in the Powheg Z+jets sample, the data driven background estimate leads to a much larger number of events, corresponding to a factor of 2.6. This large difference between MC and data-driven estimate indicates that the Powheg Z+jets sample underestimates the number of events seen in data. As long as the shape of the distribution is modelled correctly, this is not a problem because the MC estimation of the Z+jets background is only used for the estimation of the correlation R, which is not affected by a constant factor. However, the current implementation of the method seems to not predict the p_{T}^{γ} distribution correctly in the control regions B, C, and D even when this constant factor of 2.6 is applied, which means that an inclusive calculation might be problematic.

5 Conclusion

The studies presented in this report show investigations of the Z+jets background estimation for the $Z\gamma$ cross section measurement by ATLAS [1]. The calculation of the background estimate in this measurement is performed using the 2D sideband method, which is providing a data-driven background estimate. An important parameter in these calculations is the correlation R, which represents the correlation between the photon ID and the photon isolation, the two variables used to discriminate jets from photons.

The investigations of this correlation showed that this parameter changes as a function of the transverse momentum of the photon candidate but stays quite constant as a function of the pseudorapidity of the photon candidate when using the full MC Z+jets sample. To understand better where in phase space the problematic events are, the sample was split into converted photons and unconverted photons. This showed that the p_{T}^{γ} dependence gets even larger and the correlation seems to also depend on the pseudorapidity of the photon when looking only at converted photons.
The behaviour of the correlation as a function of p_{T}^{γ} did not change much when introducing either a loose cut on the isolation variable and thereby removing very non-isolated photon candidates or requiring the non-tight photon candidates to fail at least two of the cuts included in the tight requirements. However, modifying the non-tight definition reduced the amount of signal leakage significantly. Nevertheless, to decide if any different definitions of the regions should be used, the resulting effect on the total uncertainty of the background estimate should be calculated.

As these studies show that the correlation R between the photon ID and the photon isolation is not the same for different areas in phase space, a possible improvement of the background estimation would be to perform the estimation using a correlation which is binned in p_{T}^{γ} and η_{γ}. Even if this does not significantly impact the central value of the Z+jet estimate, a comparison between the estimation with the inclusive R value and an estimation using a binned version of R would be interesting to see. Of course a binned background estimation should only be considered if the statistics are still sufficient in the binned application.
Since the dependence on p_{T}^{γ} is larger than the η_{γ} dependence, a binned application with

respect to the transverse momentum of the photon candidate would probably be the choice if the two dimensional application would suffer too hard from the lower statistics. Using a binned R value could potentially improve the R-value induced uncertainty in the Z+jets background estimation. The way the method is currently implemented, a large systematic uncertainty is assigned to the R value in order to cover the spread of the R values when calculated in bins of p_{T}^{γ}. This uncertainty would not be required if the R value would be made p_{T}^{γ} dependent and especially low p_{T}^{γ} events would have a lower systematic uncertainty. Additionally, in this way, the correlation-induced uncertainty of the background estimation might not be as large as it is so far for low-p_{T}^{γ} events. As a result of that, events with $15\,\mathrm{GeV} < p_{\mathrm{T}}^{\gamma} < 30\,\mathrm{GeV}$ could potentially be also included in the measurement.

6 References

[1] *Measurement of $Z\gamma \to \ell^+\ell^-\gamma$ differential cross-sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2019-034. Geneva: CERN, July 2019. URL: http://cds.cern.ch/record/2682846.

[2] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08003.

[3] The ATLAS Collaboration. "Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016". In: *Eur. Phys. J. C* 79 (Oct. 2018), 205. 55 p. DOI: 10.1140/epjc/s10052-019-6650-6. arXiv: 1810.05087. URL: https://cds.cern.ch/record/2643391.

[4] ATLAS Collaboration. "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} =$ 13 TeV". In: arXiv:1902.04655 (Feb. 2019), arXiv:1902.04655. arXiv: 1902.04655 [physics.ins-det].

[5] ATLAS Collaboration. "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 TeV". In: arXiv:1603.05598 (Mar. 2016), arXiv:1603.05598. arXiv: 1603.05598 [hep-ex].

[6] James Saxon. "Discovery of the Higgs Boson, Measurements of its Production, and a Search for Higgs Boson Pair Production". Presented 13 06 2014. July 2014. URL: https://cds.cern.ch/record/1746004.

[7] The ATLAS Collaboration. "Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}$=7 TeV". In: *The European Physical Journal C* 73.3 (2013), p. 2304. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-013-2304-2. URL: https://doi.org/10.1140/epjc/s10052-013-2304-2.

[8] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *Journal of High Energy Physics* 2010.6 (June 2010), p. 43. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2010)043. URL: https://doi.org/10.1007/JHEP06(2010)043.

[9] M. Schönherr and F. Krauss. "Soft photon radiation in particle decays in SHERPA". In: *Journal of High Energy Physics* 2008.12, 018 (Dec. 2008), p. 018. DOI: 10.1088/1126-6708/2008/12/018. arXiv: 0810.5071 [hep-ph].