Moscow International School of Physics 2022

Likelihood fit based estimation of the background induced by the misidentification of a jet as a photon at *pp* collider experiment

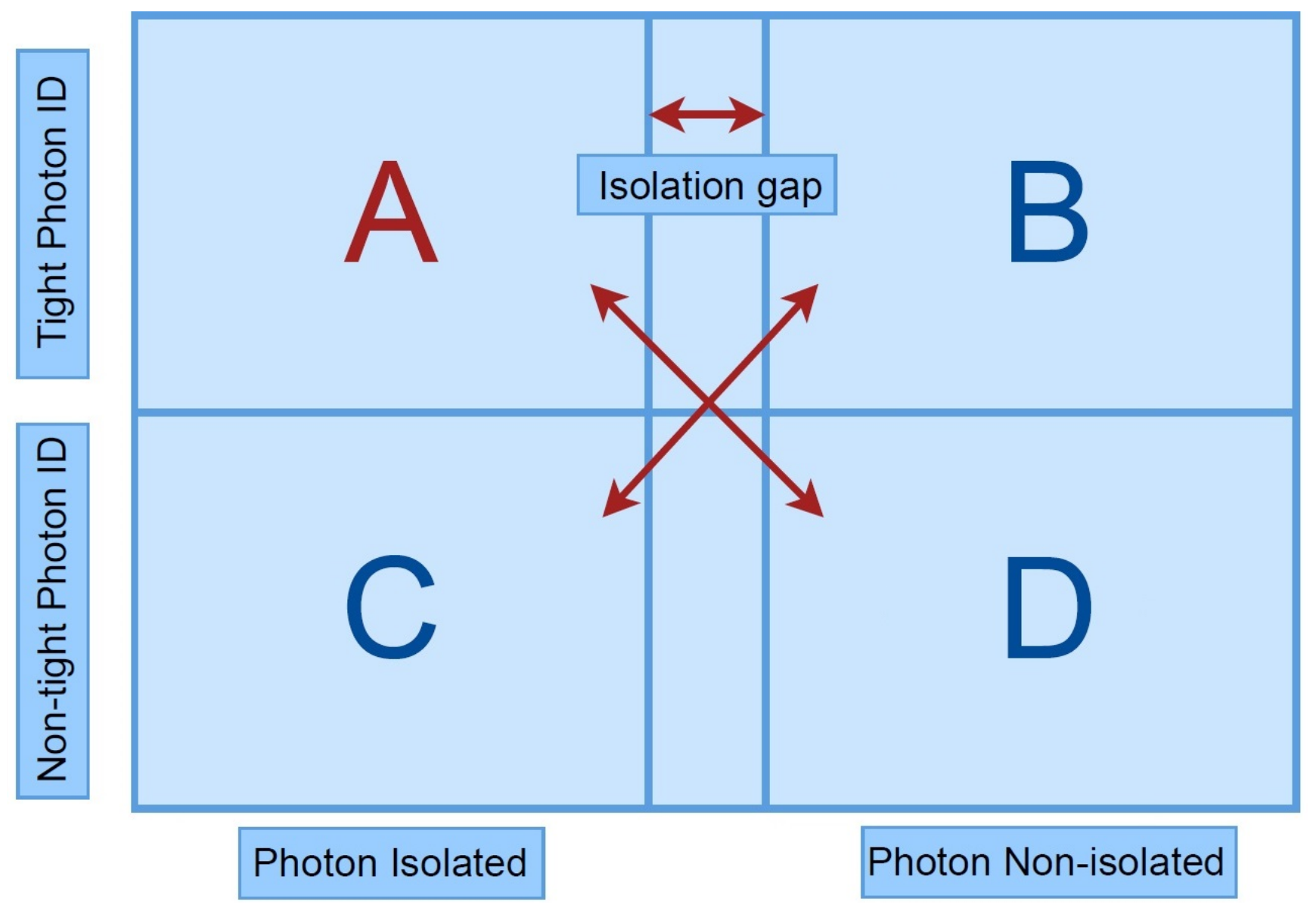
K. Kazakova, E. Soldatov, D. Pyatiizbyantseva, K. Savelyev

Introduction

Background processes emerging from object misidentification are not well-modeled in Monte-Carlo (MC). Data-driven methods are usually used to obtain more accurate estimates for such backgrounds. One of these methods — two-dimensional sideband method (ABCD-method) is used by many physics analyses at the LHC. However, it is rather laborious. In this study the alternative maximum likelihood based method of the estimation of the background from object misidentification was developed.

Methodology

In the estimation of the background arising from the misidentification of jet as a photon ($jet \to \gamma$) ABCD-method splits the phase space basing on identification (*tight*[1] and *loose*) and isolation (*isolated*[2] and *non-isolated*) criteria for photons. Signal region (SR) A that should mainly consist of signal events and control regions (CR) B, C and D that should consist of background events are illustrated below.



The main assumption of the ABCD-method is the absence of the correlation, that corresponds to $R = \frac{N_A N_D}{N_B N_C} = 1$. The developed likelihood-based approach uses more general relationships between regions and takes into account the shape of distributions in each region [1]. The essence of the method is to perform a fit [2] of the likelihood function, that is defined as:

$$L(N_{ji}|f_{F_{ji}}, f_{N_j}) = \prod_{j=A}^{B,C,D} \prod_{i=1}^{N_{bins}} \mathrm{Pois}(N_{ji}|\nu_{b_{ji}} + \nu_{\gamma_{ji}} f_{F_{ji}} + \nu_{s_{ji}} f_{N_j}),$$

where model parameters are defined as:

- f_{N_j} – varying parameter for signal in each region;
- $f_{F_{ji}}$ – varying parameter for estimated background in each region and bin;
- $\nu_{b_{ji}}$ – number of events in MC backgrounds (exc. $jet \to \gamma$);
- $\nu_{s_{ji}}$ – number of signal events;
- $\nu_{\gamma_{ji}}$ – number of estimated background ($jet \to \gamma$) events.

It is constructed with the assumption that $R = 1$ for each bin in the distribution for $jet \to \gamma$ background:

$$1 = \frac{\nu_{\gamma_{Ai}} f_{F_{Ai}} \cdot \nu_{\gamma_{Di}} f_{F_{Di}}}{\nu_{\gamma_{Bi}} f_{F_{Bi}} \cdot \nu_{\gamma_{Ci}} f_{F_{Ci}}}.$$

To avoid the redundancy of the model the following limitations are applied: $f_{F_{Bi}} = f_{F_{Di}}$. In this way the number of $jet \to \gamma$ events in SR: $N_A^{jet\to\gamma} = \nu_{\gamma_{Ai}} f_{F_{Ai}}$.

The proposed method significantly reduces the number of steps to be done to obtain the estimate compared to ABCD-method.

The results of the fit

The likelihood-based approach was applied to associated $Z\gamma$ production with Z-boson decaying into neutrinos ($Z \to \nu\bar{\nu}$). One of the backgrounds comes from γ+j events, when large $E_{\mathrm{T}}^{\mathrm{miss}}$ is created by a combination of mismeasured jet energy and real $E_{\mathrm{T}}^{\mathrm{miss}}$ from heavy quarks decays. Zj events come from $jet \to \gamma$ misidentification. The phase space of SR used in the analysis is defined in the table below.

Selection	Cut value
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 130 GeV
E_{T}^{γ}	> 150 GeV
Number of tight photons	$N_\gamma = 1$
Lepton veto	$N_e = 0$, $N_\mu = 0$

The processes considered in the analysis were generated in MadGraph5_aMC [3] MC[3] event generator using pp-collisions with $\sqrt{s} = 13$ TeV and the integrated luminosity of 139 fb^{-1}. Pythia8 [4] and Delphes [5] were used for parton showering and hadronization and detector simulation respectively. The fit was performed for ϕ_γ and η_γ. Table on the right represents the estimates depending on N_{bins}. The final estimate is chosen based on the $\chi^2/N_{d.o.f.}$ value in SR. The estimate of $jet \to \gamma$ events in SR obtained by likelihood method is $N_A^{jet\to\gamma} = 3179^{+117}_{-108} \pm 69$ events for ϕ_γ and $N_A^{jet\to\gamma} = 3243^{+126}_{-122} \pm 48$ events for η_γ. The MC prediction is $N_A^{jet\to\gamma} = 3093 \pm 178$ events.

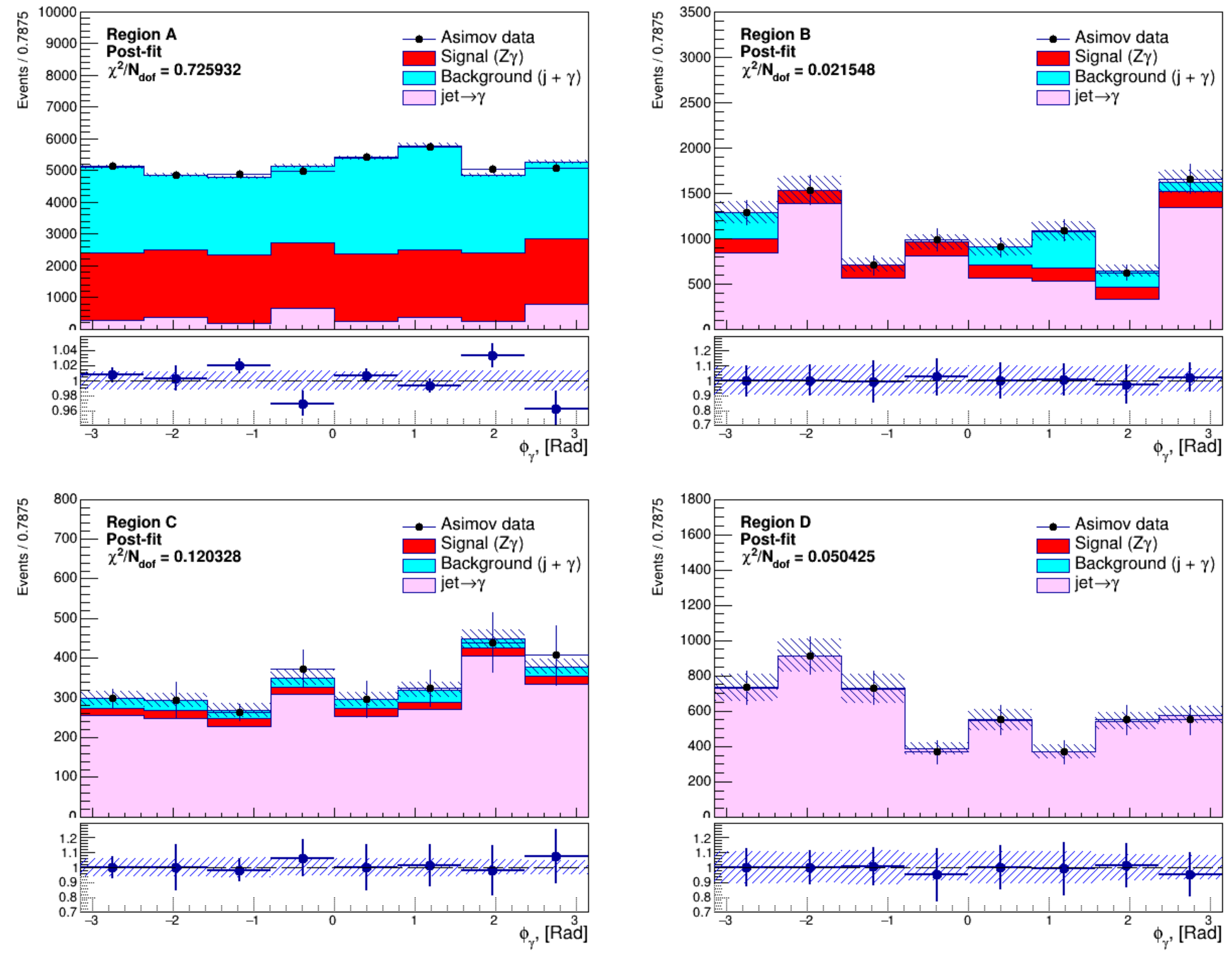


Figure 1: The result of the likelihood fit for ϕ_γ

N_{bins}	ϕ_γ			η_γ		
	Estimate	R-factor	$\chi^2/N_{d.o.f.}$	Estimate	R-factor	$\chi^2/N_{d.o.f.}$
6	3255^{+111}_{-106}	1.04 ± 0.03	0.45	3238^{+129}_{-125}	1.03 ± 0.03	0.39
7	2906^{+110}_{-108}	0.94 ± 0.03	0.73	3243^{+126}_{-122}	1.04 ± 0.02	0.55
8	3179^{+117}_{-108}	1.04 ± 0.03	0.73	3276^{+141}_{-137}	1.04 ± 0.02	0.26
9	3119^{+130}_{-127}	1.01 ± 0.03	0.62	3251^{+133}_{-130}	1.05 ± 0.02	0.50

Conclusions

The alternative likelihood-based method of estimation of $jet \to \gamma$ events was developed. It uses the information about the shape of the distributions in the regions and provides the much simpler way to obtain the estimate of the number of background events. The final estimates for different variables and MC prediction coincide within the error.

References

1. Buttinger W. // Background Estimation with the ABCD Method. 2018. P.8-16.
2. W. Verkerke, D. Kirkby. // RooFit Users Manual V2.91. 2019. Sept. P.112.
3. Allwall J. et. al. // JHEP. 2014. V.2014. №07. P.079.
4. Sjostrand T. et. al. // Comput. Phys. Commun. 2015. V.191. P.159-177.
5. de Favereau J. et. al. // JHEP. 2014. V.2014. №2. P.057.

Contact

- katerina.kazakova@cern.ch

---

[1] Tight photon satisfies all criteria of the EM shower shape; [2] Isolated photon satisfies $E_{\mathrm{T}}^{cone40} - 0.022 p_{\mathrm{T}} < 2.45$ GeV; [3] Asimov data is a sum of all MC processes.