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

K. Kazakova[1], E. Soldatov[1], D. Pyatiizbyantseva[1], K. Savelyev[2]

[1] NRNU Moscow Engineering Physics Institute, Moscow, Russia

[2] Moscow State University, Moscow, Russia

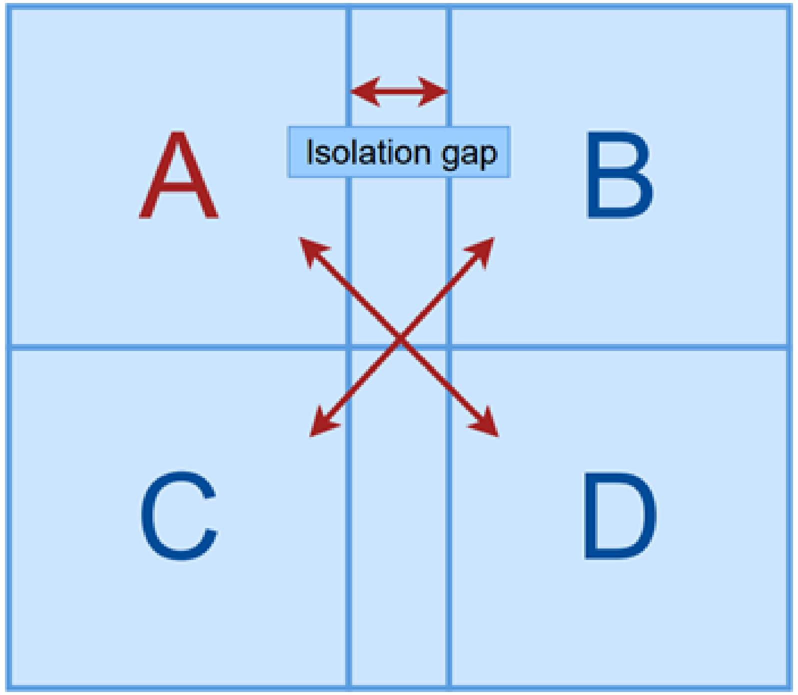
The XXVI International Scientific Conference of Young Scientists and Specialists, 24 – 28 October 2022

Motivation and estimation techniques I

- Background processes emerging from object misidentification are not well-modeled in Monte-Carlo. One of the most often used data-driven method is two-dimensional sideband method (ABCD–method). However, it is rather laborious.

Briefly about ABCD-method:

- splits the phase space into 4 regions based on 2 discriminating variables;
- the main assumption is $R = \frac{N_A N_D}{N_B N_C} = 1$



MC

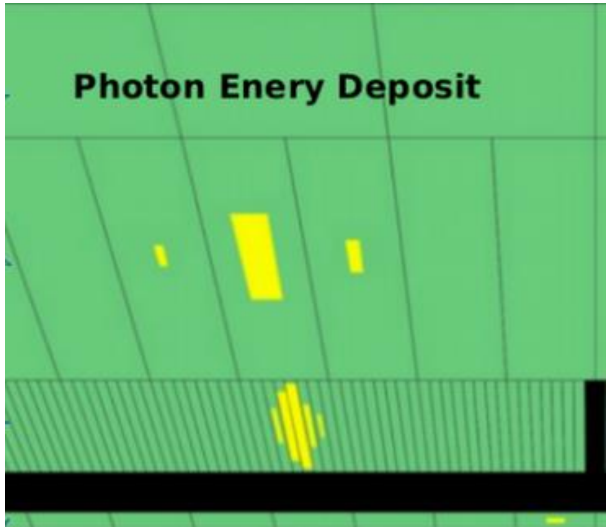
Disadvantages:

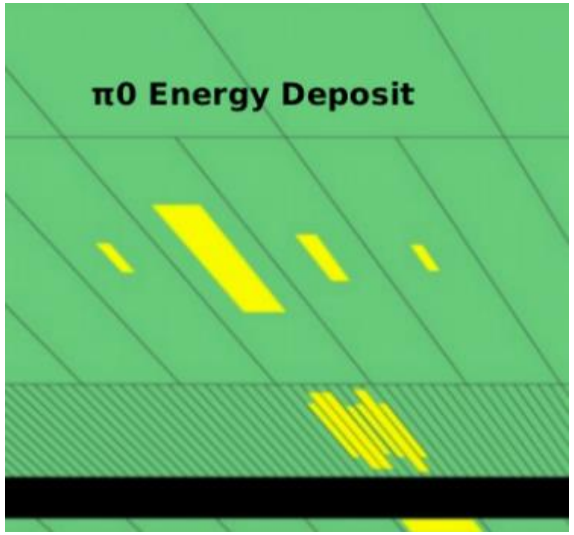
- Extremely long R factor estimation process;
- It doesn't use the information about the shape of the distributions in the regions.

Motivation: to develop an alternative estimation method which greatly simplifies the estimation

Motivation and estimation techniques II

The background induced by the misidentification of a jet as a photon is studied in this analysis.



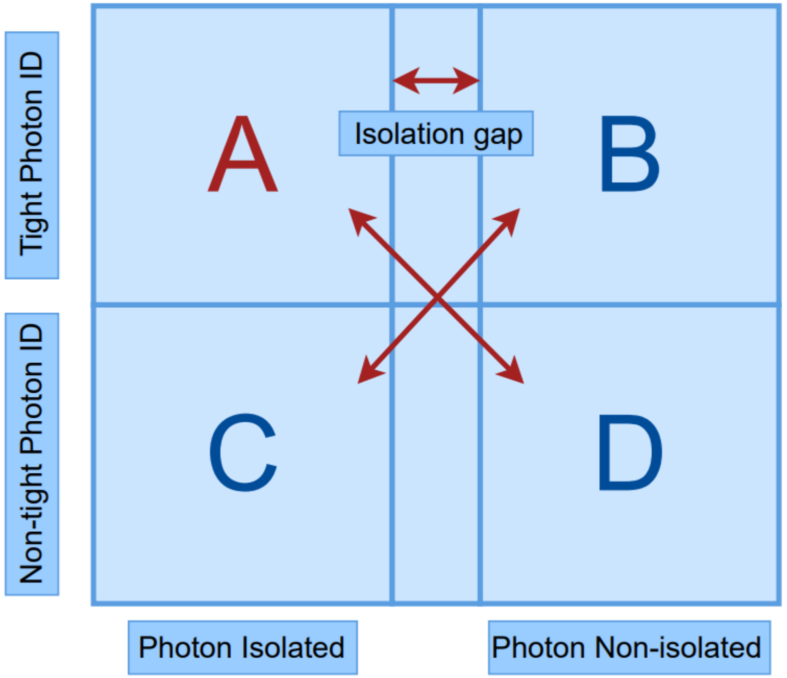


Hadronic jets in which a π^0 carries a significant fraction of the energy may be misidentified as isolated photons

⇒ the SR will be contaminated with jet $\rightarrow \gamma$

ABCD-method for jet $\rightarrow \gamma$:

- the phase space is splitted into 4 regions based on the identification (*tight* or *loose'*) and isolation (*isolated* or *non-isolated*) criteria for photons;
- the main assumption is the absence of correlation between identification and isolation criteria.



Likelihood-based approach I

- The main idea: to fit signal and other backgrounds distributions except jet → γ to data in all ABCD regions

The essence of the method is to perform a fit of the likelihood function, which 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 (excl. jet → γ);
- $\nu_{s_{ji}}$ – number of signal events;
- $\nu_{\gamma_{ji}}$ – number of estimated background (jet → γ) events.

Likelihood-based approach II

- Likelihood based approach is constructed with the assumption that R = 1 for each bin in the distribution for jet → γ 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 limitation is applied: $f_{F_{Bi}} = f_{F_{Di}}$

The search of minimum of likelihood function is performed with RooFit toolkit:

$$\frac{\partial L}{\partial f_{F_{ji}}} = 0, \quad \frac{\partial L}{\partial f_{N_j}} = 0$$

This way the number of jet → γ 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

MC samples

- The likelihood-based approach was applied to associated Zγ production with Z-boson decaying into neutrinos ($Z \to \nu\nu$). One of the backgrounds comes from γ+j events. Zj events come from jet → γ misidentification

- The processes considered in the analysis were generated in MadGraph5 MC event generator using pp collisions with $\sqrt{s}$ = 13 TeV and the integrated luminosity of 139 fb^{-1}

- Pythia8 and Delphes were used for parton showering and hadronization and detector simulation respectively.

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

Event selection criteria for Zγ candidate events

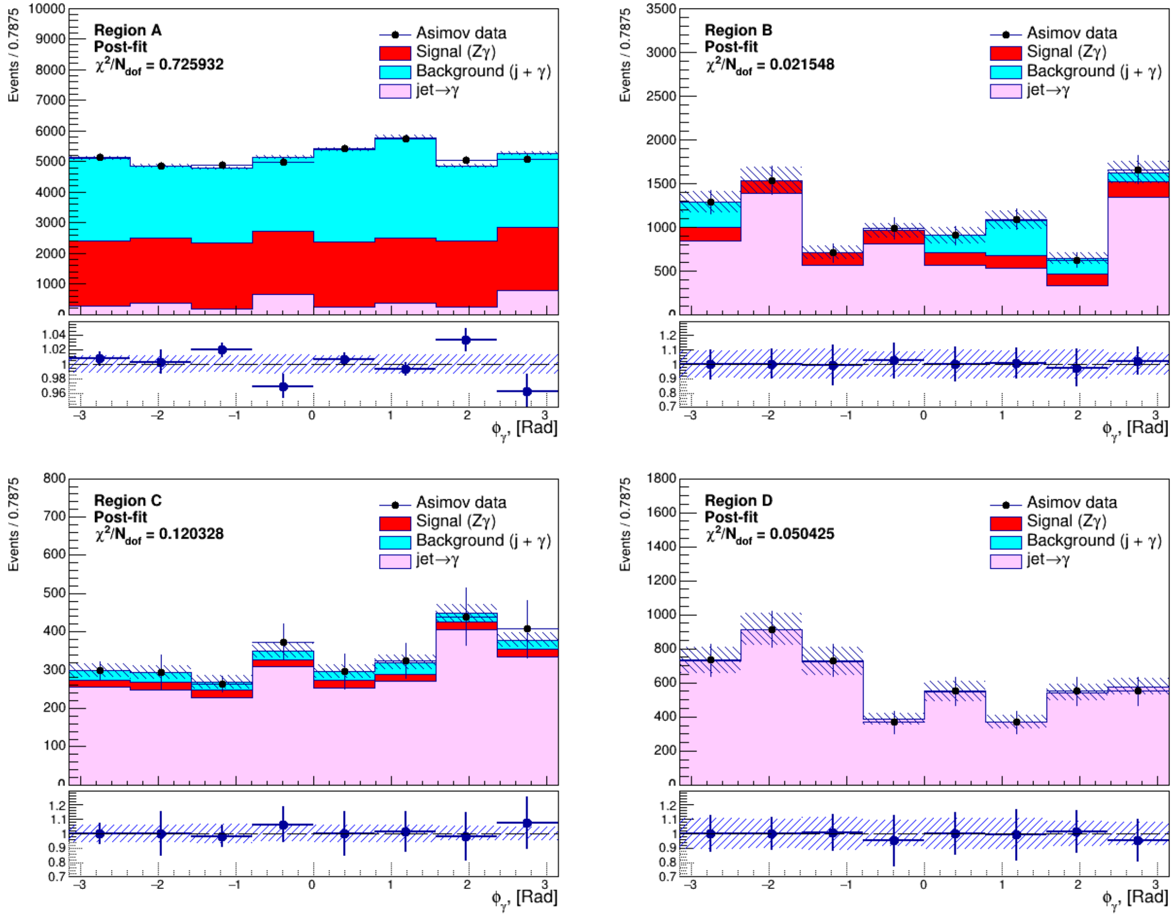
Thus the study uses Asimov data which is not real data but the sum of MC generated processes, the likelihood-based estimate of jet → γ background and MC prediction should coincide. It is so-called «closure test».

The results of the fit

The fit was performed for ϕ_γ and η_γ :

The final estimate is chosen based on the $\chi^2/N_{d.o.f.}$ value in SR and R-factor.

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



- The systematic uncertainties were derived by variating the value of isolation gap by ±σ in non-isolated control regions.
- The estimate of jet → γ events in SR obtained by likelihood method is $N_A^{jet\to\gamma} = 3179^{+117}_{-108} \pm 69$ for ϕ_γ and $N_A^{jet\to\gamma} = 3243^{+126}_{-122} \pm 48$ for η_γ
- The MC prediction is $N_A^{jet\to\gamma} = 3093 \pm 178$ events

Summary

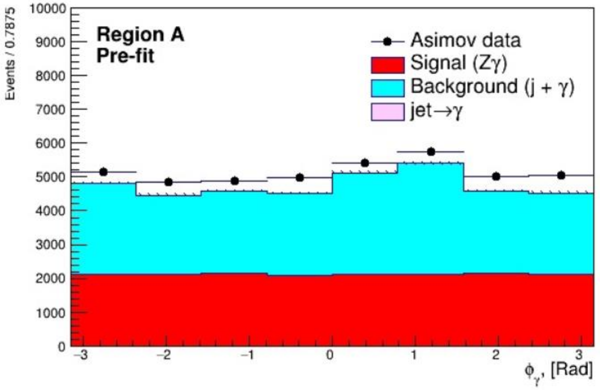
- The alternative likelihood-based method of estimation of jet $\to \gamma$ events was developed
- The method uses the information about the shape of the distributions in the regions and provides a much simpler way to obtain the estimate of the number of background events.
- The final estimates for different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

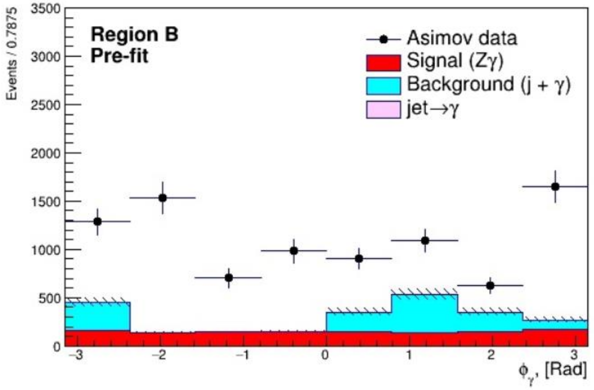
Thanks for your attention!

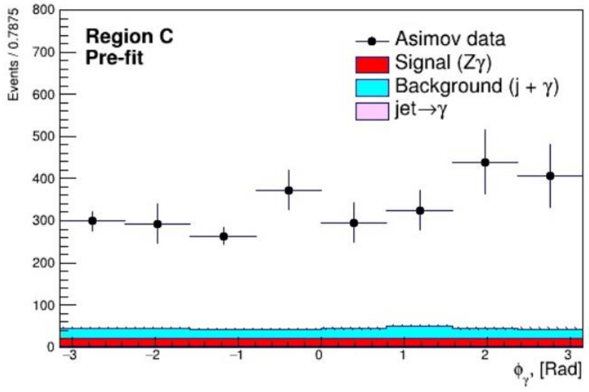
BACK-UP

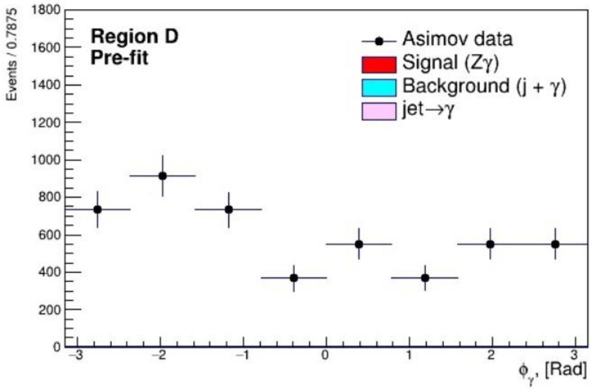
The results of the fit

Pre-fit for ϕ_γ

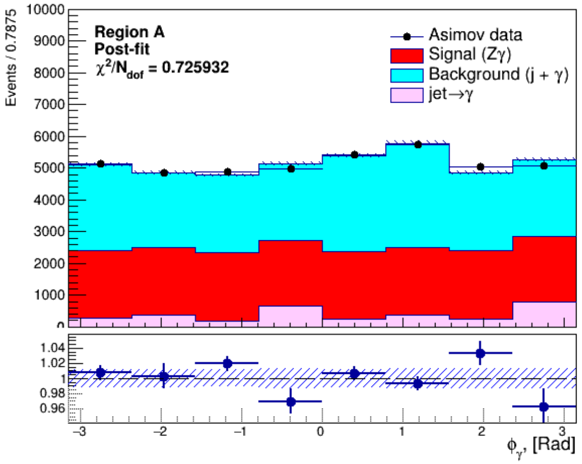


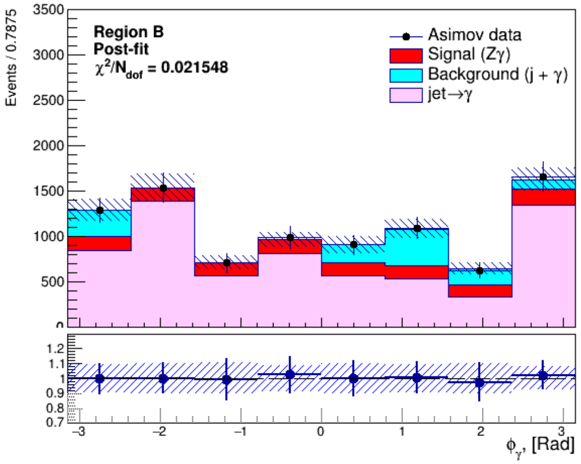


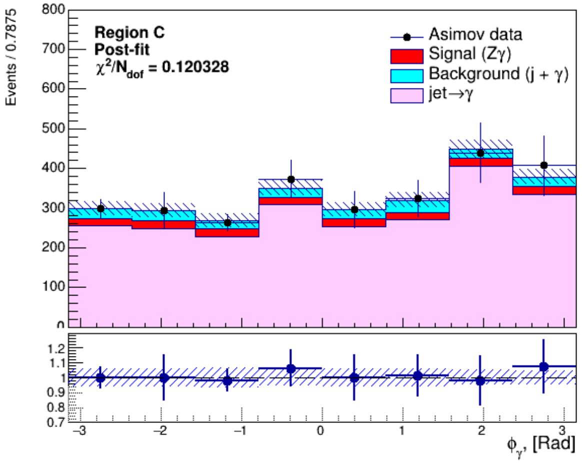


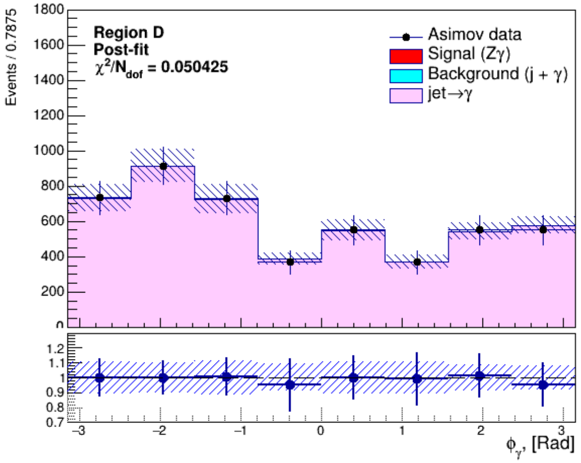


Post-fit for ϕ_γ



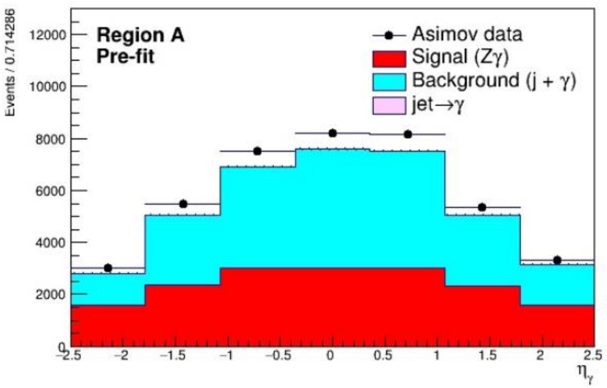


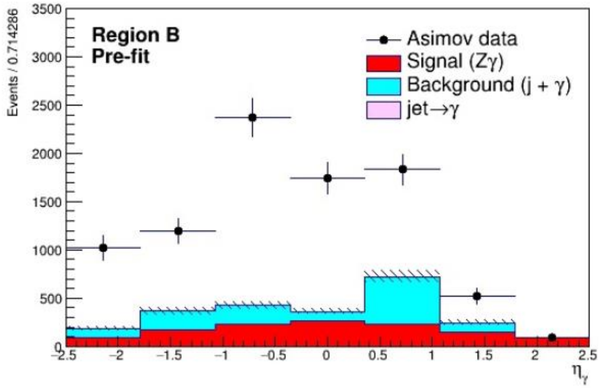


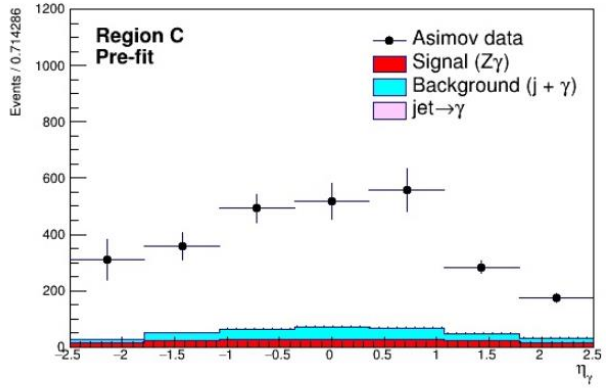


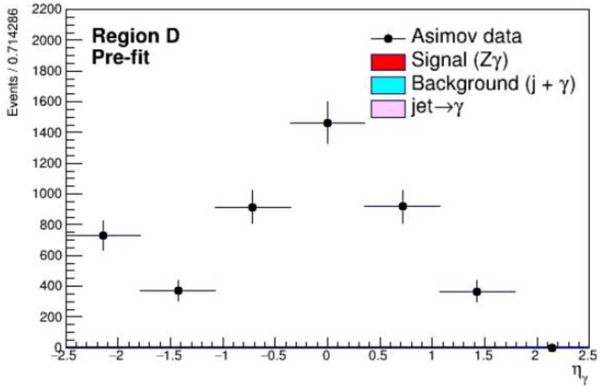
The results of the fit

Pre-fit for η_γ









Post-fit for η_γ