$B^0 \to \pi^0\pi^0$ analysis

Sebastiano

Overview

BF and A_{CP} of $B^0 \to \pi^0\pi^0$ decays: fundamental measurements at Belle II.

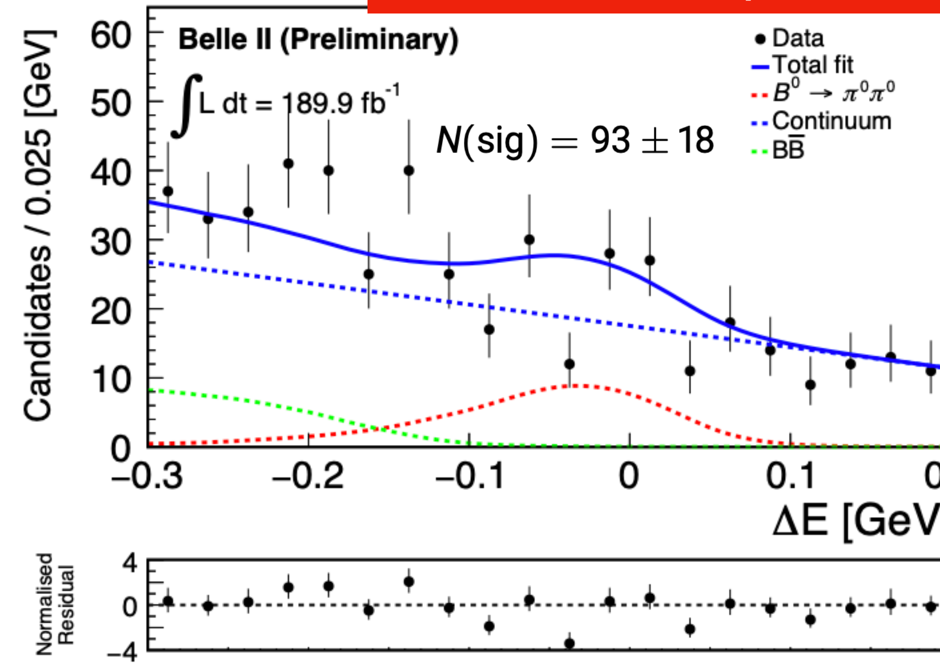
Results (@189.9fb^{-1}) by Francis shown at ICHEP2022.

Now: prepare new analysis for pre-LS1 dataset.

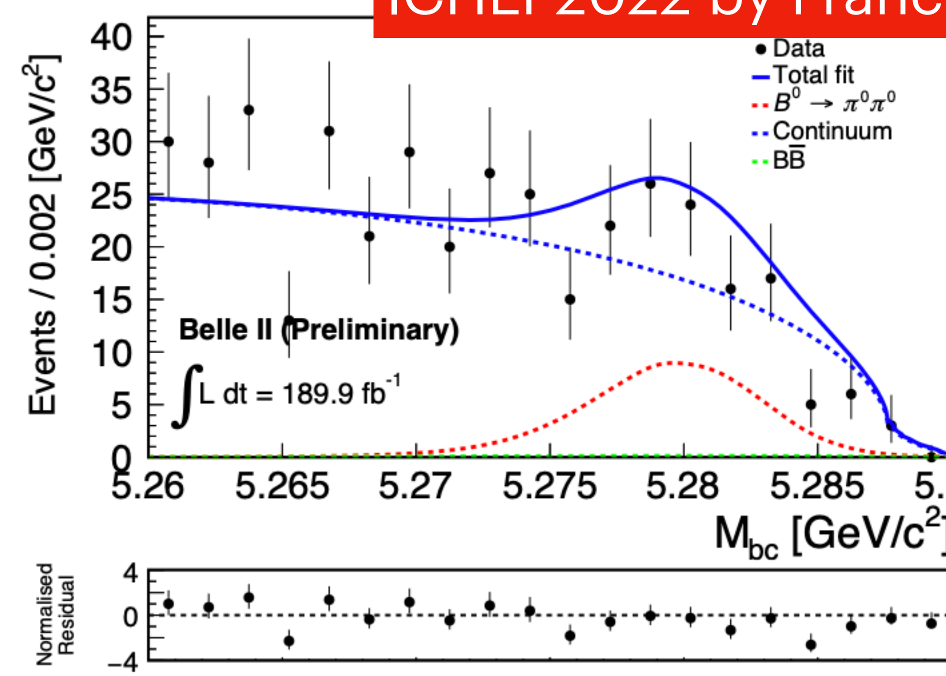
Plan:

- revisit photonMVA looking at variables with good data/MC agreement
- revisit CSBDT adding BTag variables to suppress even more $e^+e^- \to q\bar{q}$
- Introduce specific BDT trained against continuum ρ's

ICHEP2022 by Francis



ICHEP2022 by Francis



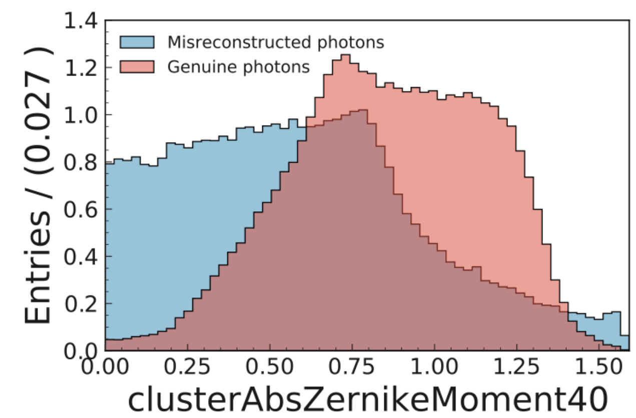
$\mathcal{A}^{CP} = -0.14 \pm 0.46$ (stat.) ± 0.07 (syst.)
$\mathcal{B} = (1.27 \pm 0.25$ (stat.) ± 0.17 (syst.)$) \cdot 10^{-6}$

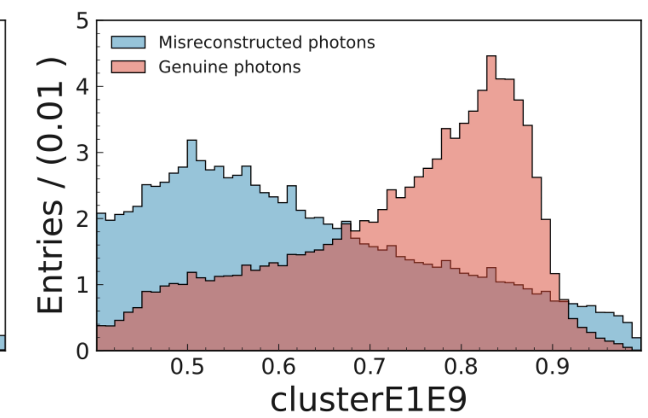
Photon MVA

Photon MVA

Distinguish between real photons and “false” photons: beam backgrounds, other particles, energy releases from other particles (split-offs)….

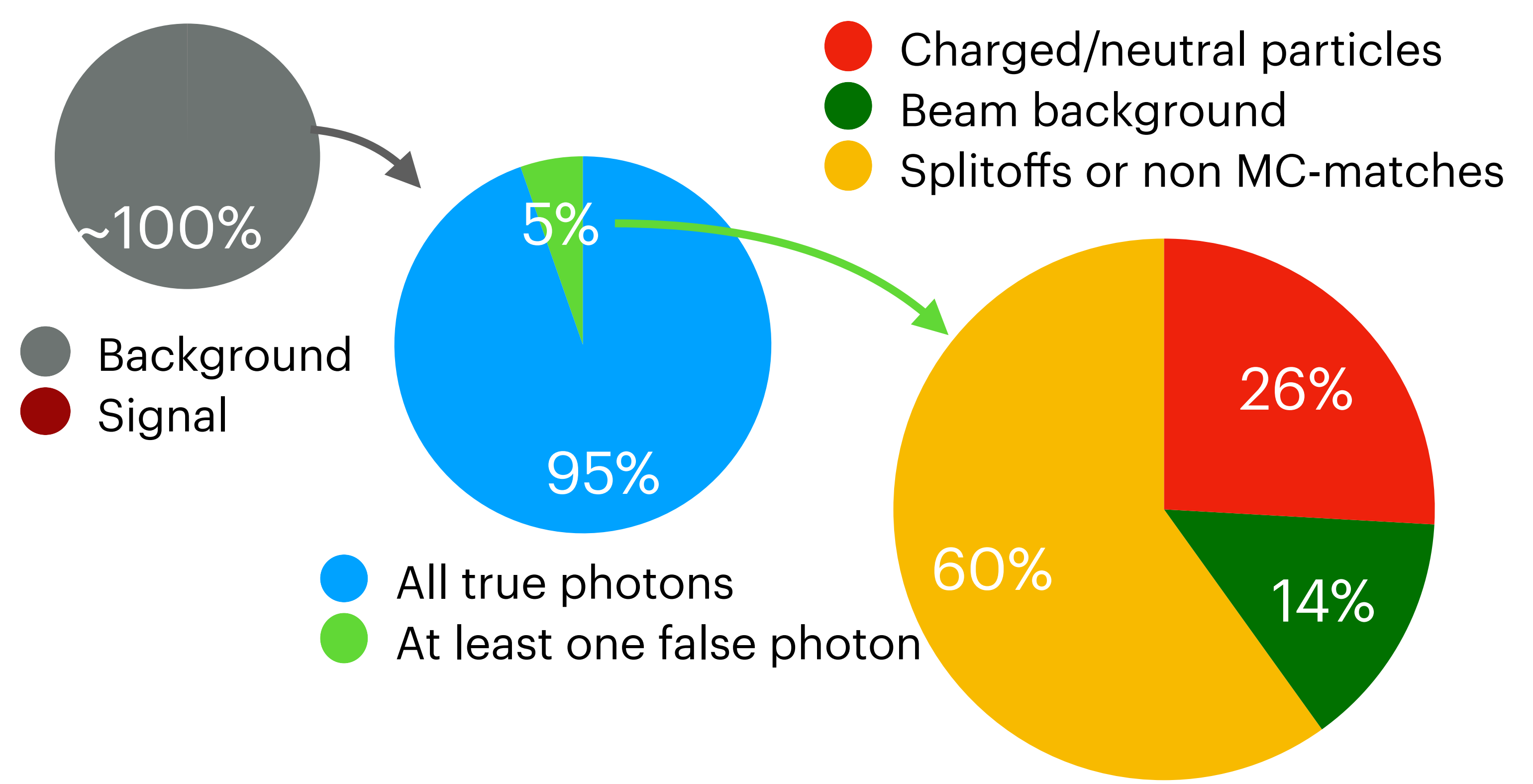
Combine highly-discriminant cluster- and photon-variables in a MVA.





False photons have usually low energies, while $B^0 \to \pi^0\pi^0$ photons high-energy.

After the default selection on photons and π^0’s, the residual bkg is mainly composed by true combinatorial π^0’s.



Photon MVA: inputs validation

Ideally we need a sample of true photons and a sample of false photons (difficult to obtain).

Use inclusive sample of photons from $D^* \to D^0(K\pi\pi^0)\pi$ decays: apply same π^0 selections of my analysis $\to$ same π^0 kinematic distributions.

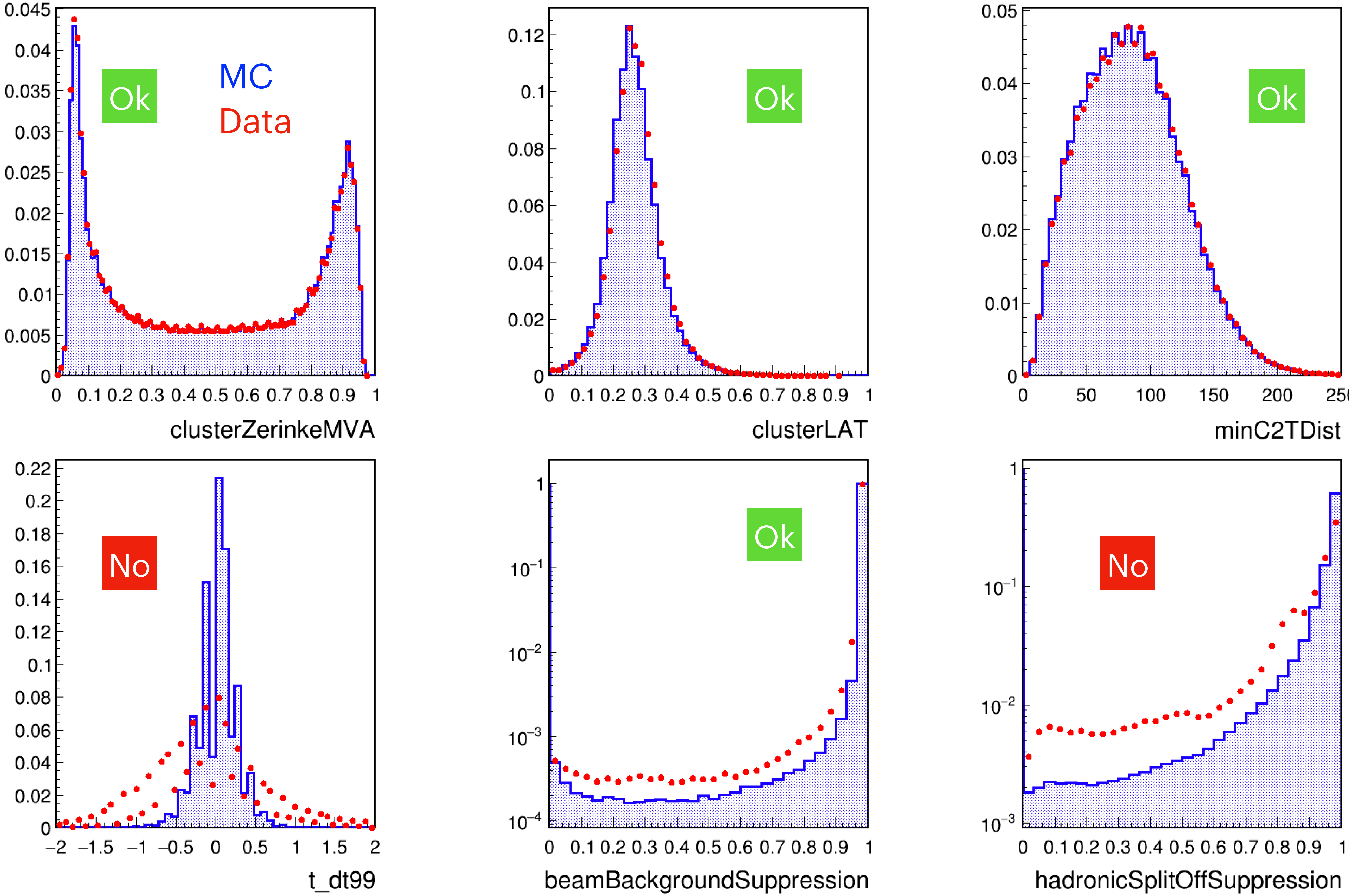
Sample is signal dominated $\to$ ~all true photons (as in $B^0 \to \pi^0\pi^0$).

Compare input distributions using MC14rd (1 ab^{-1})/Proc12+AllBuckets(189 fb^{-1}) and MC15ri (200 fb^{-1})/Proc13c1(8 fb^{-1}).

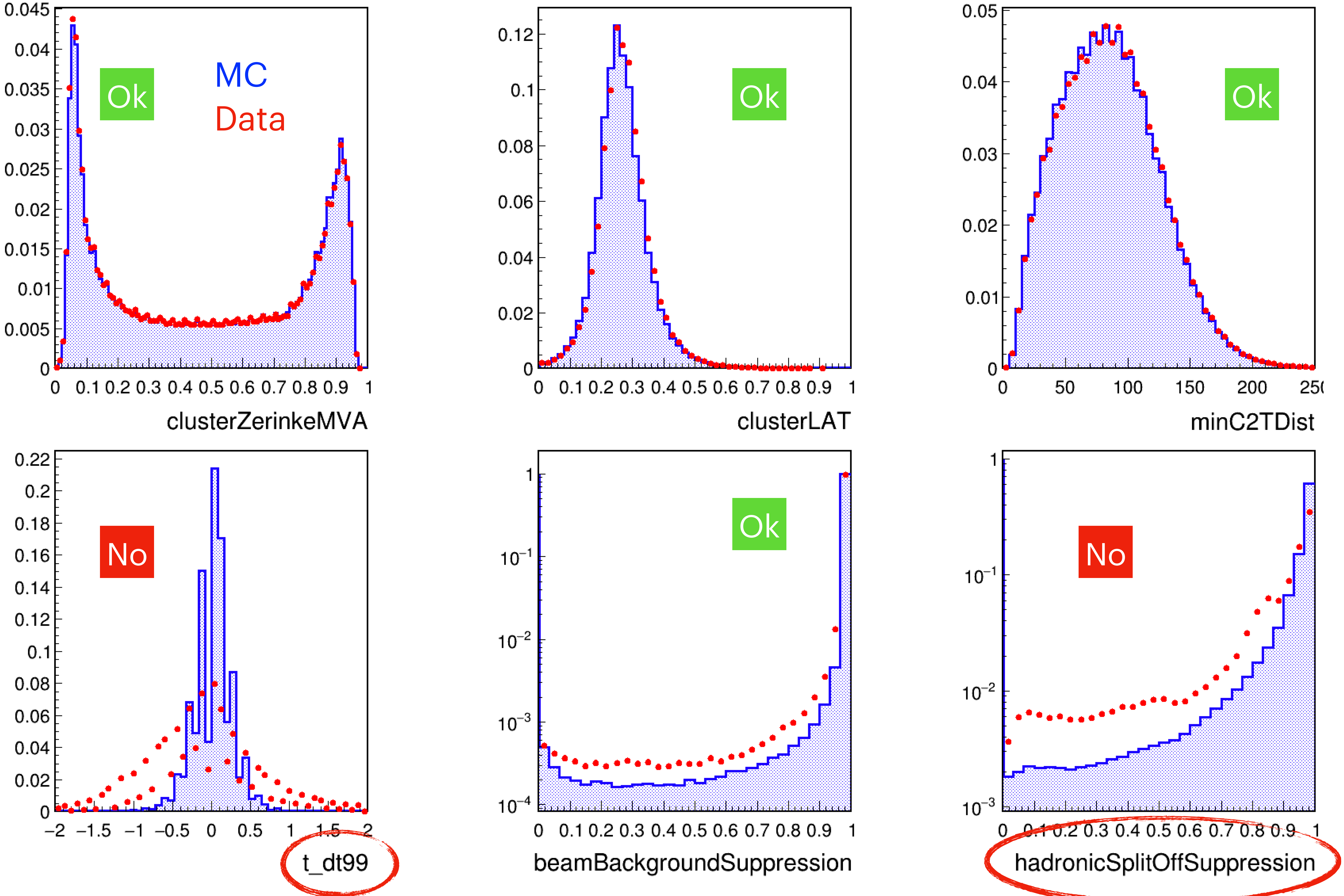
MC14 vs Proc12+AllBuckets

Release-05

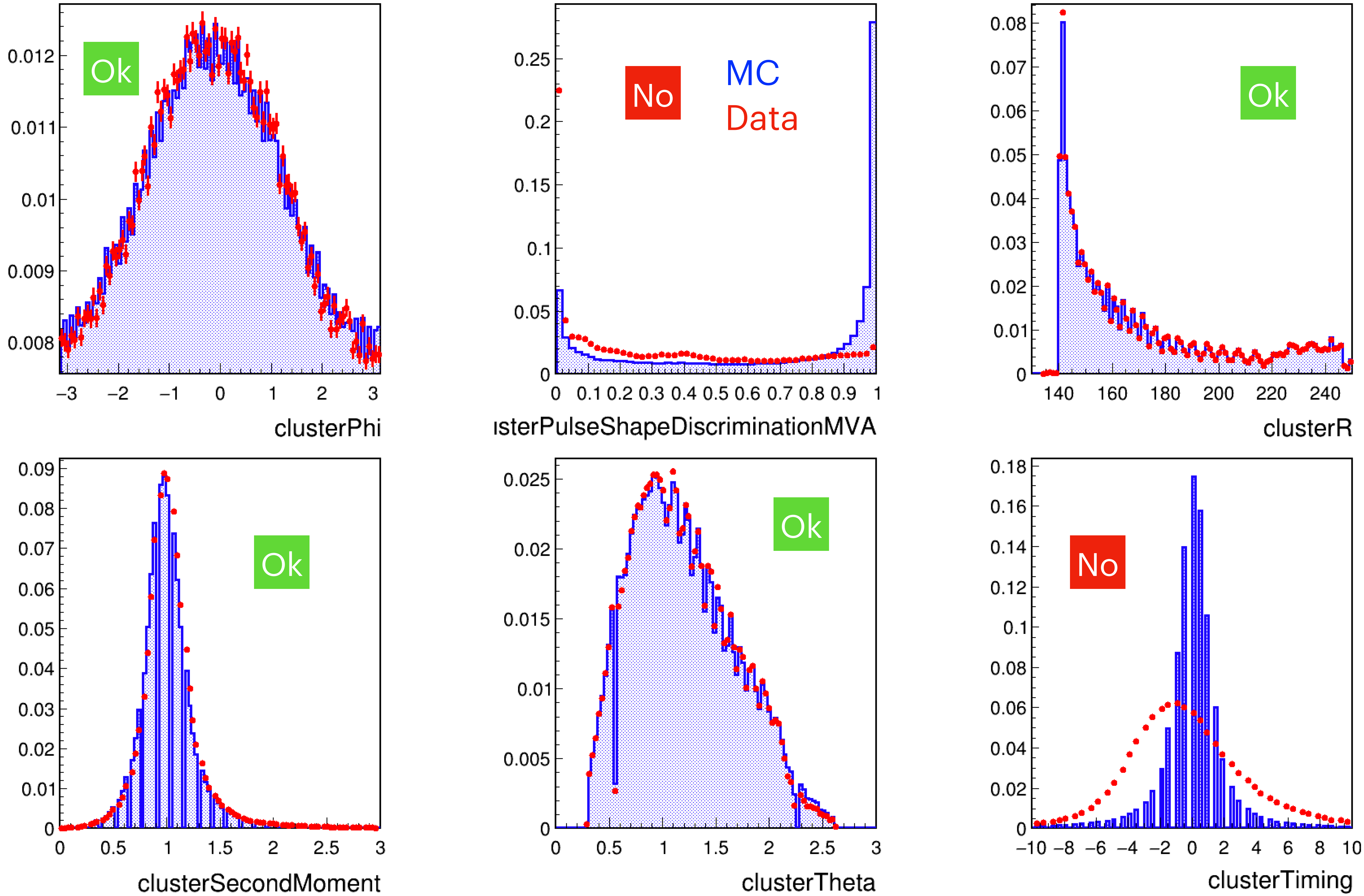
Photon MVA: inputs validation (rel-05)



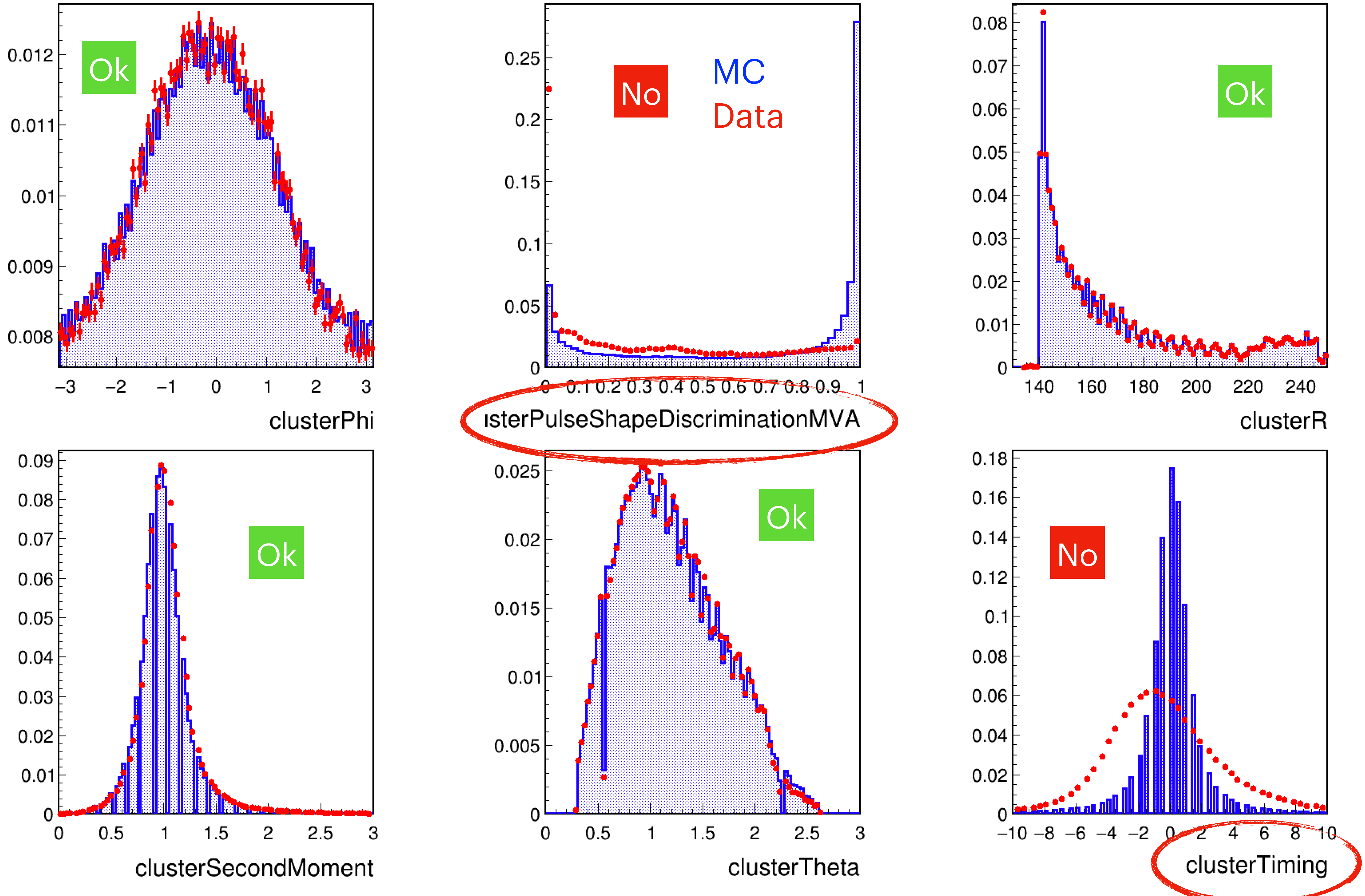
Photon MVA: inputs validation (rel-05)



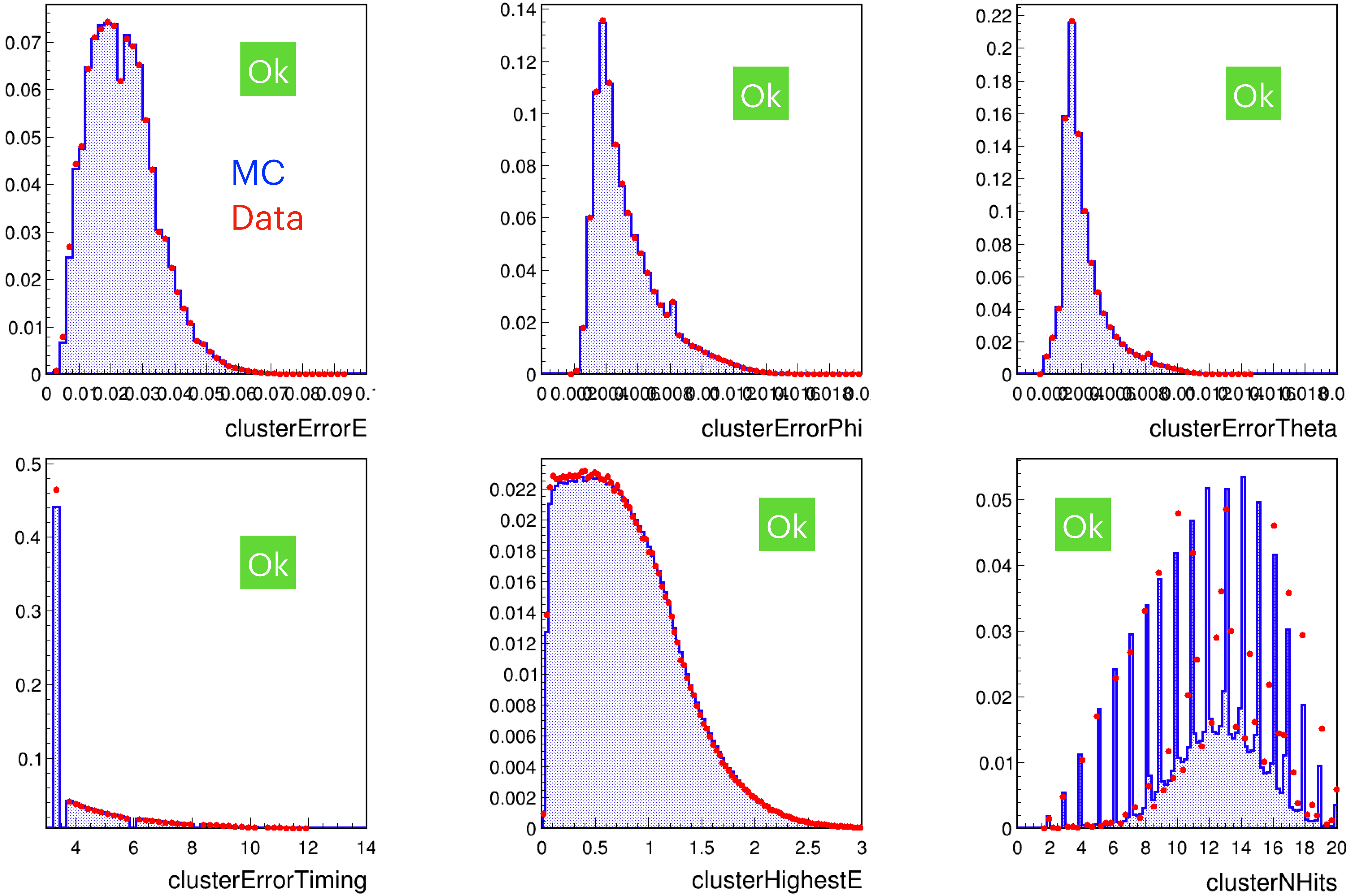
Photon MVA: inputs validation (rel-05)



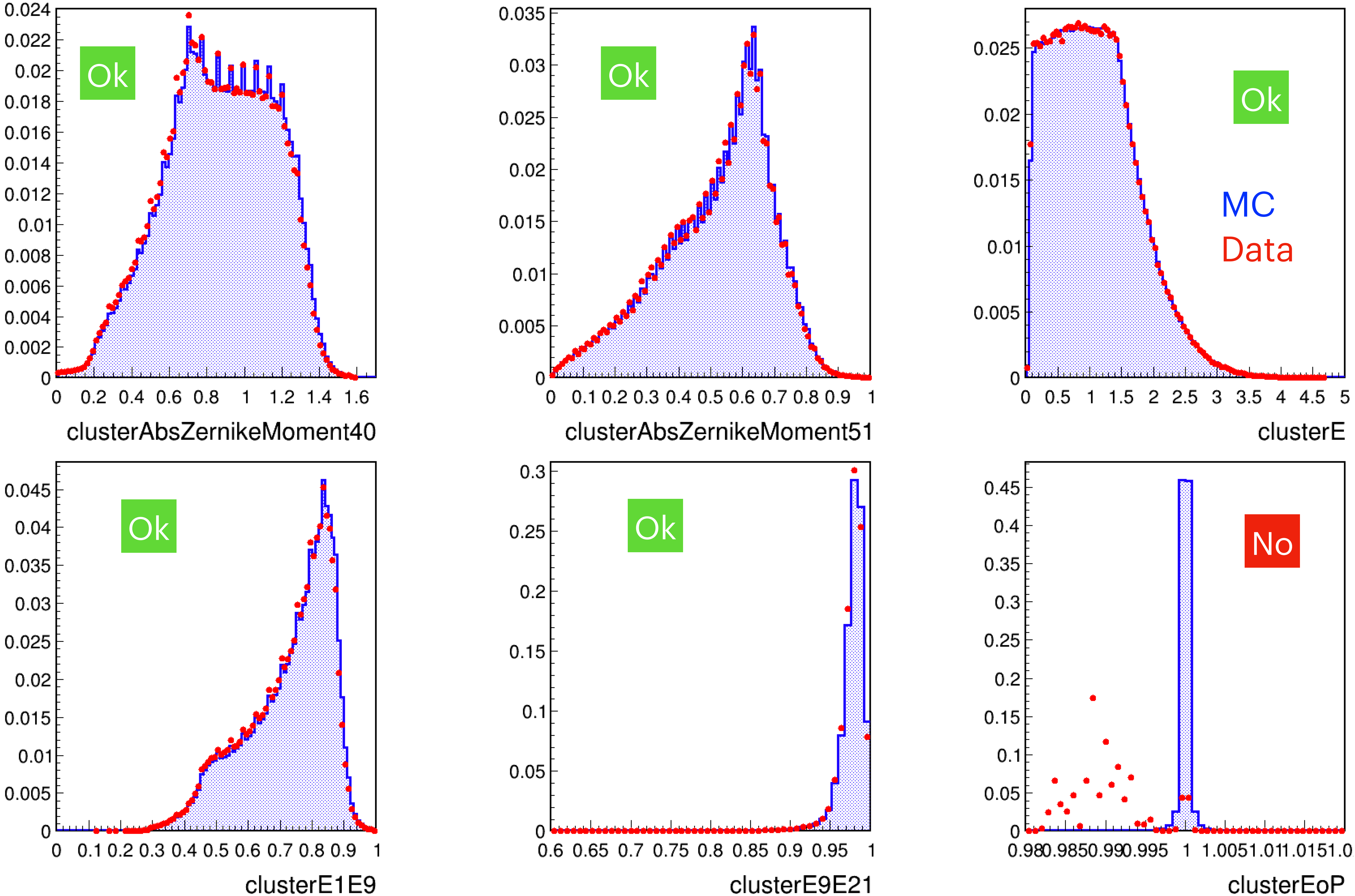
Photon MVA: inputs validation (rel-05)



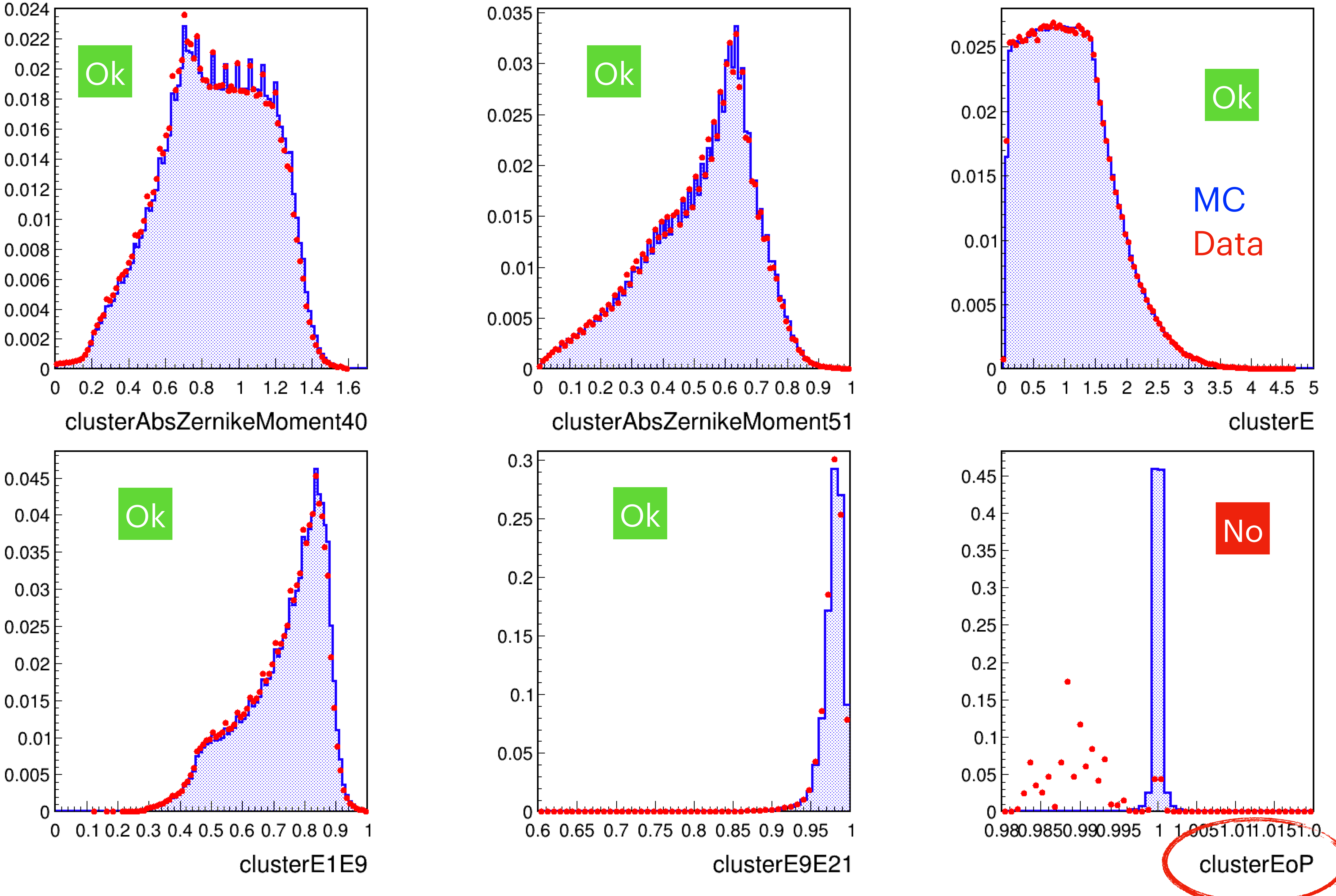
Photon MVA: inputs validation (rel-05)



Photon MVA: inputs validation (rel-05)



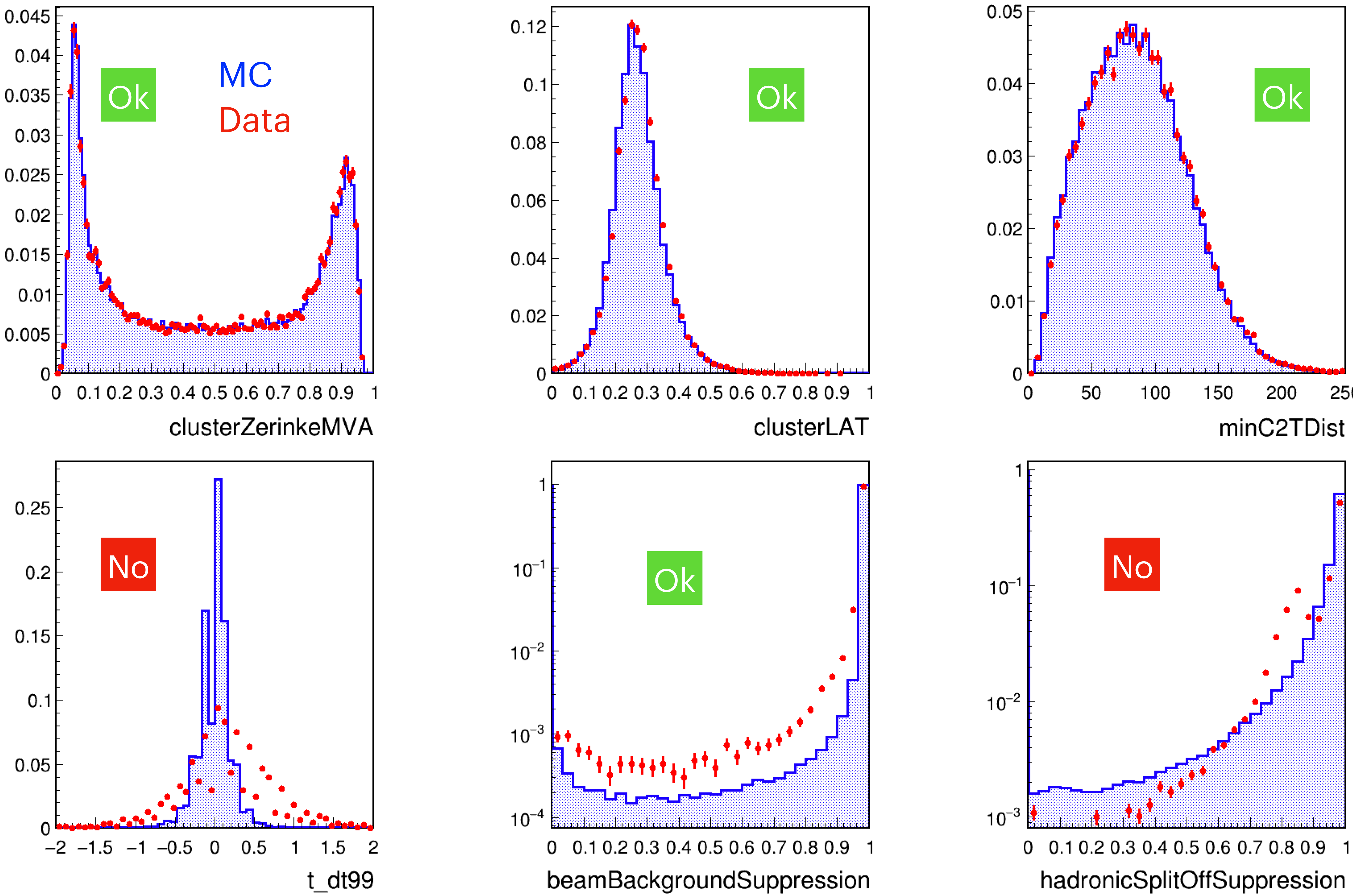
Photon MVA: inputs validation (rel-05)



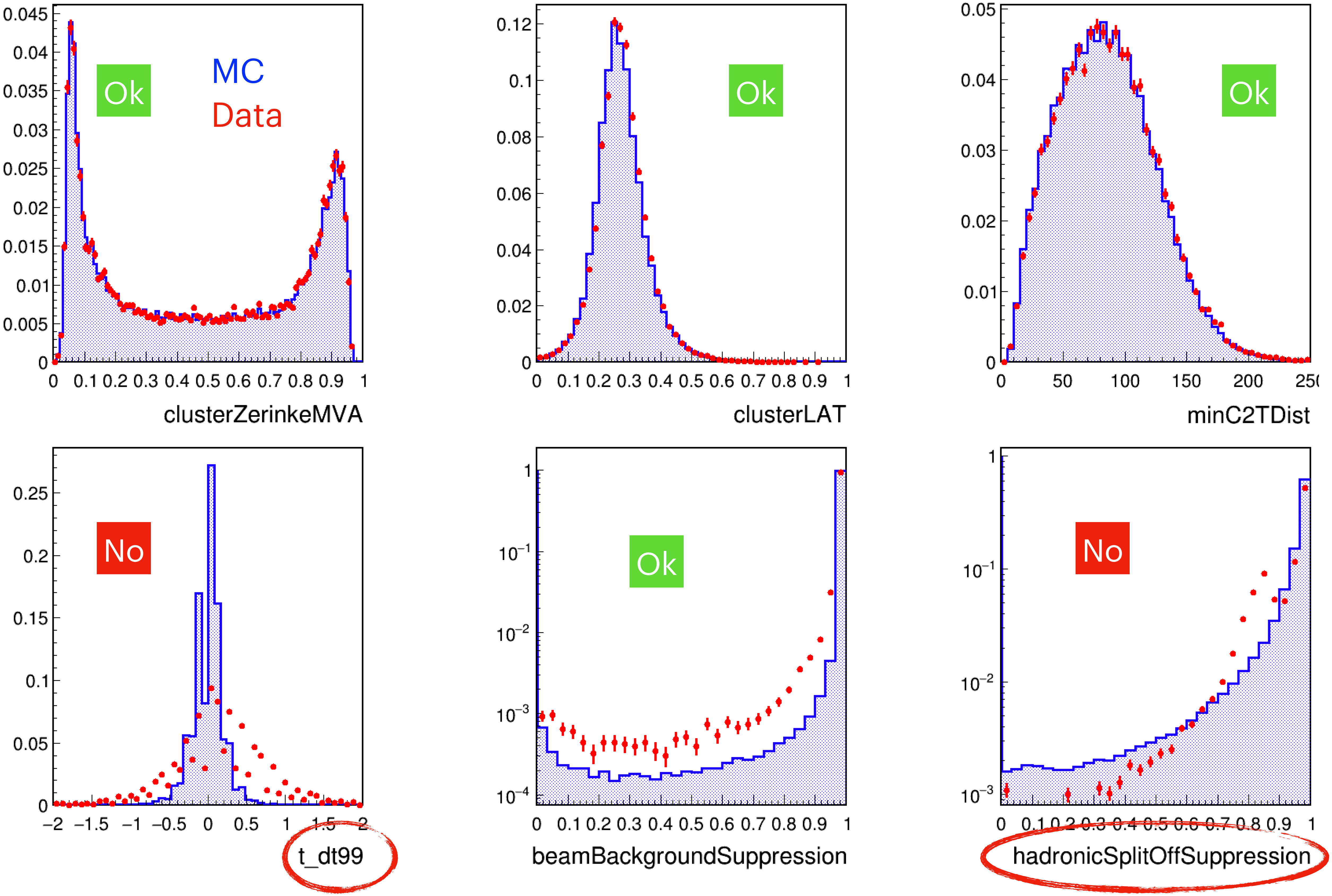
MC15 vs Proc13

Release-06

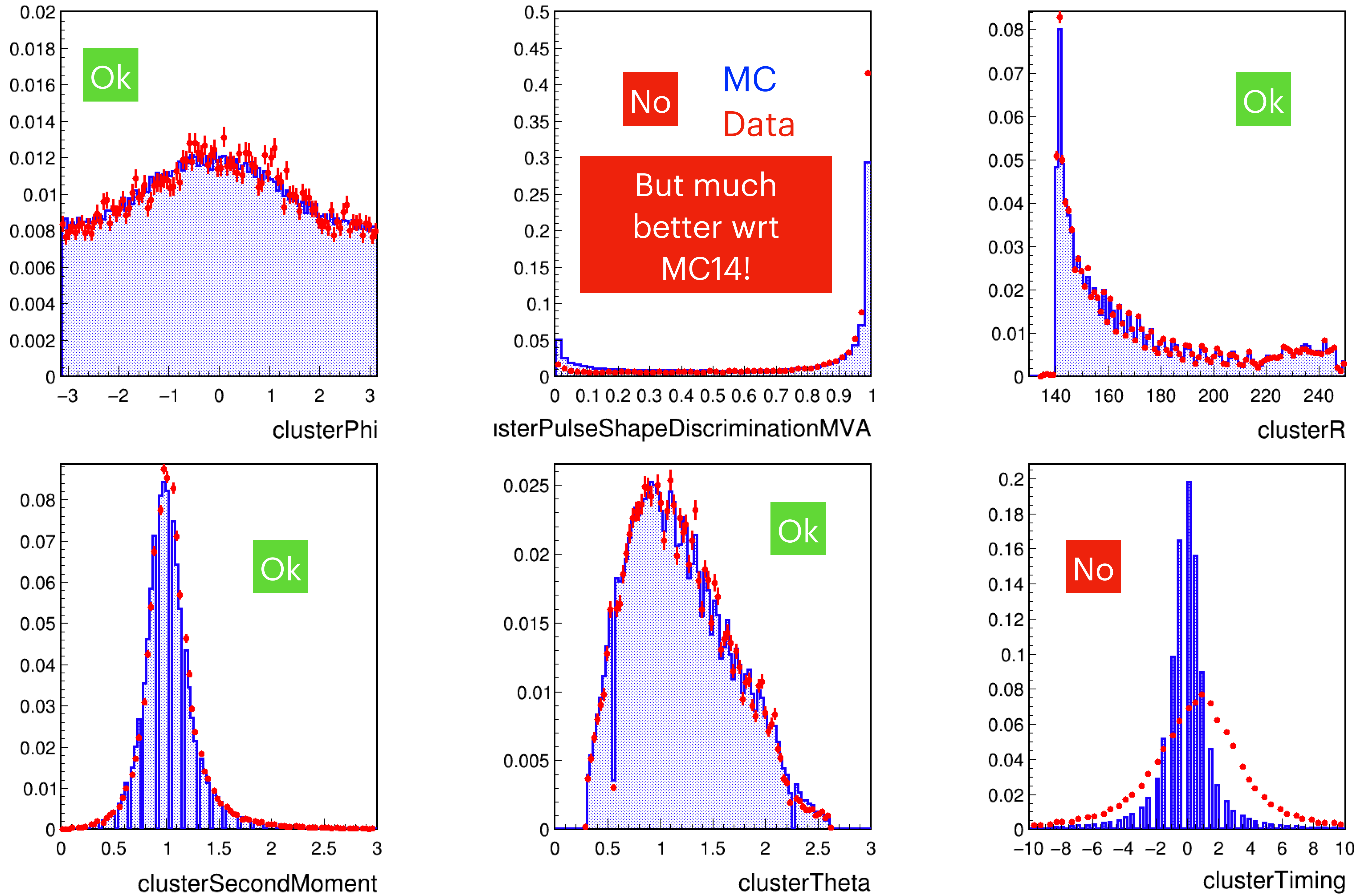
Photon MVA: inputs validation (rel-06)



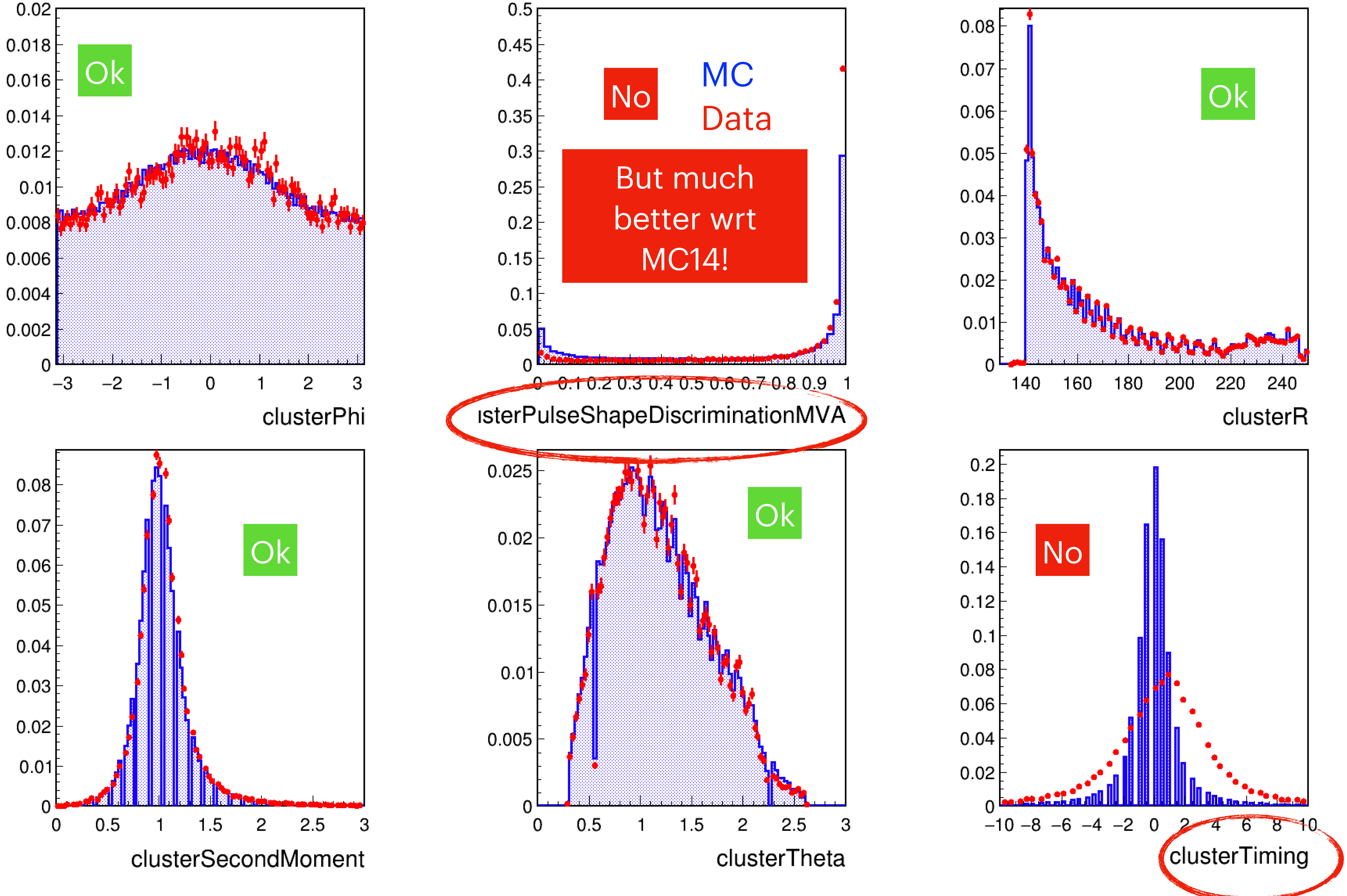
Photon MVA: inputs validation (rel-06)



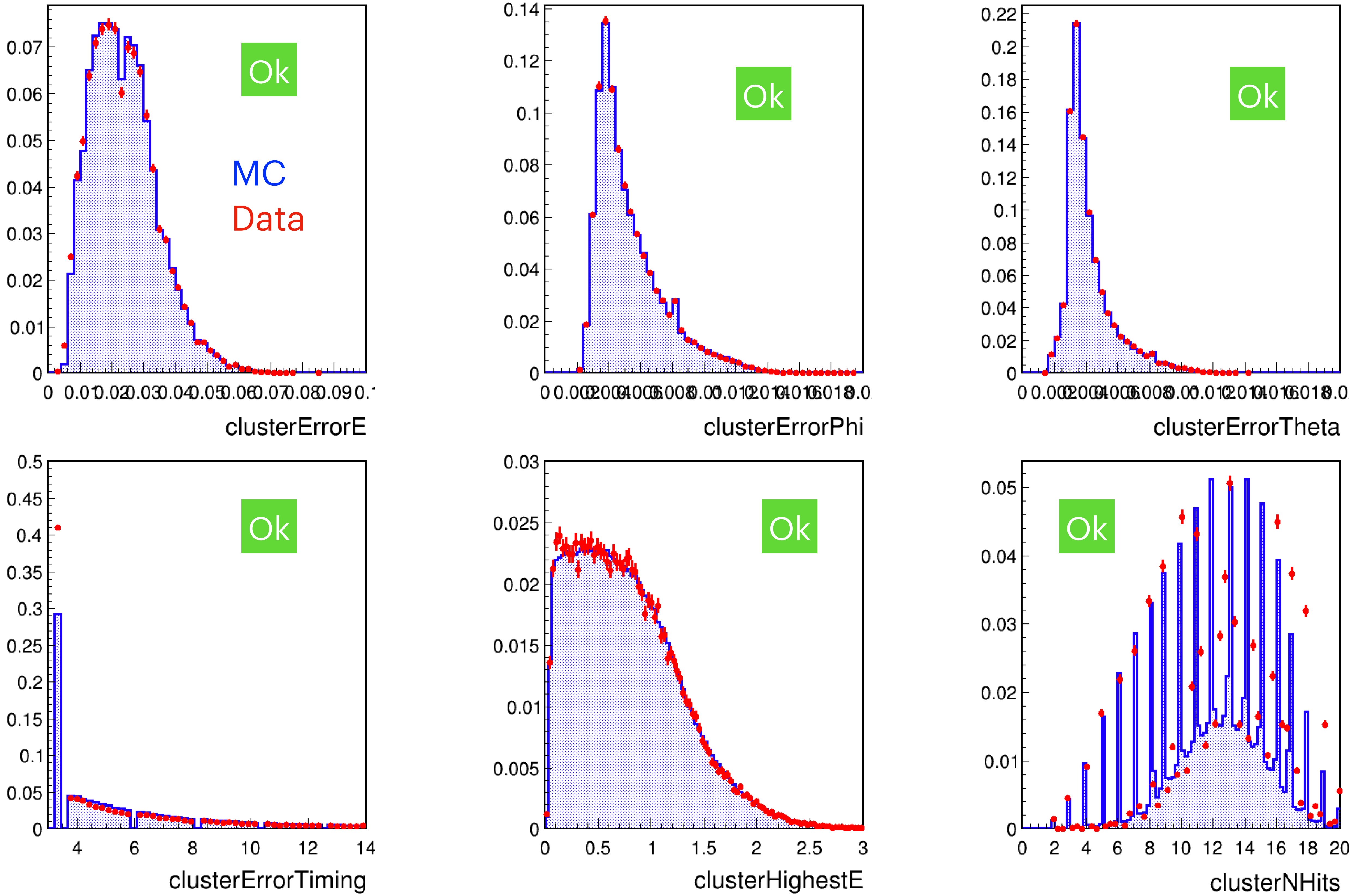
Photon MVA: inputs validation (rel-06)



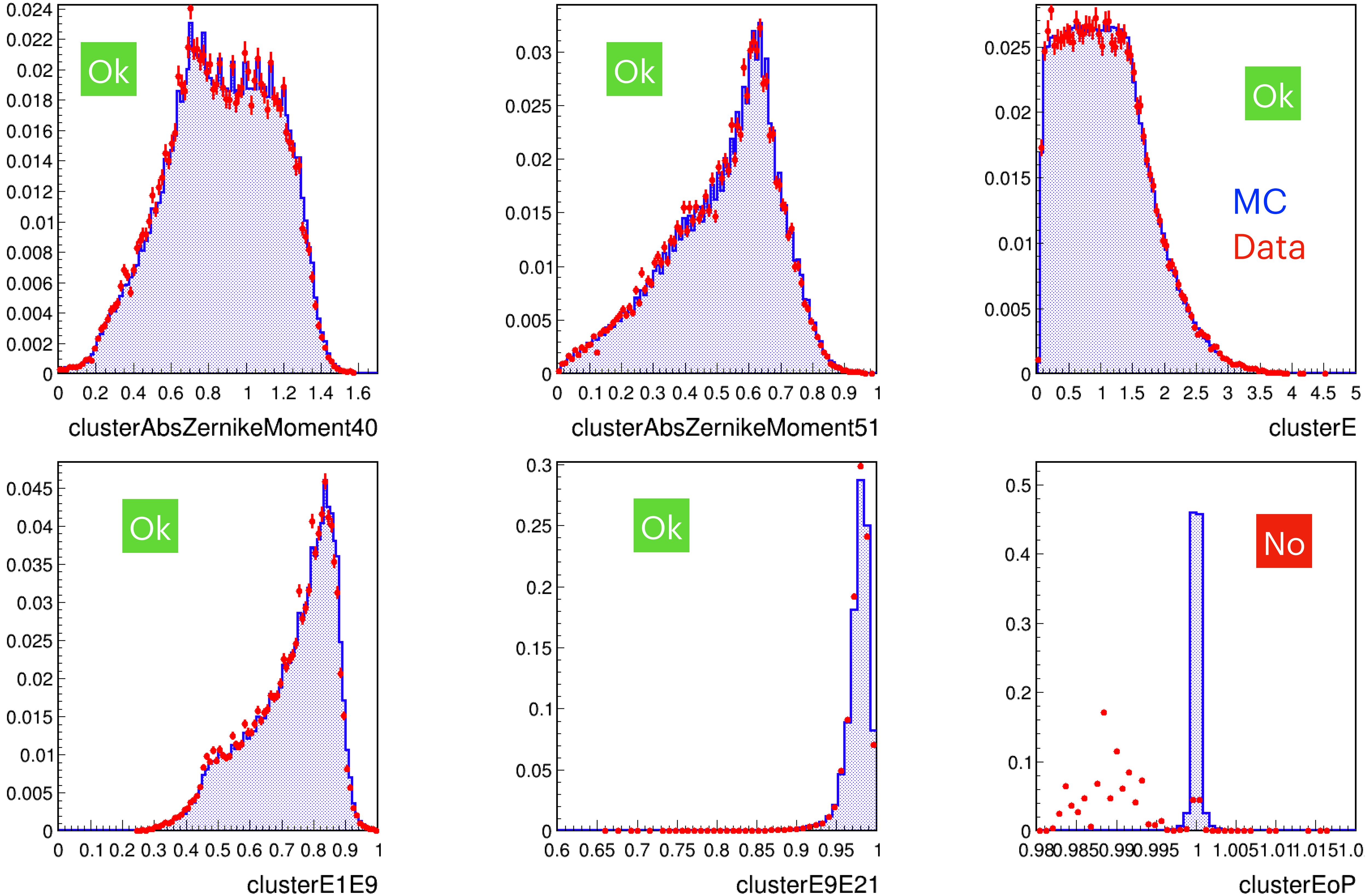
Photon MVA: inputs validation (rel-06)



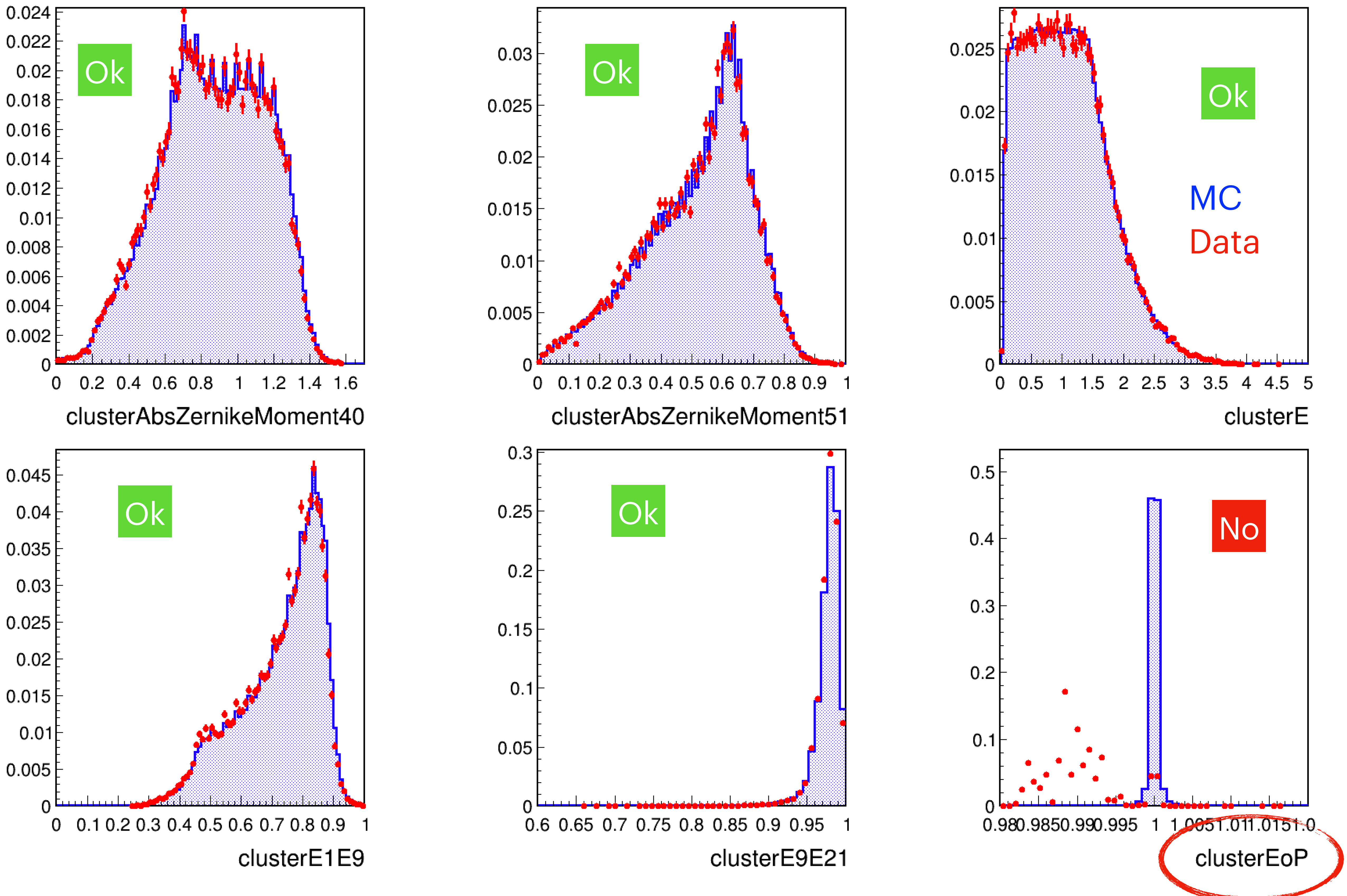
Photon MVA: inputs validation (rel-06)



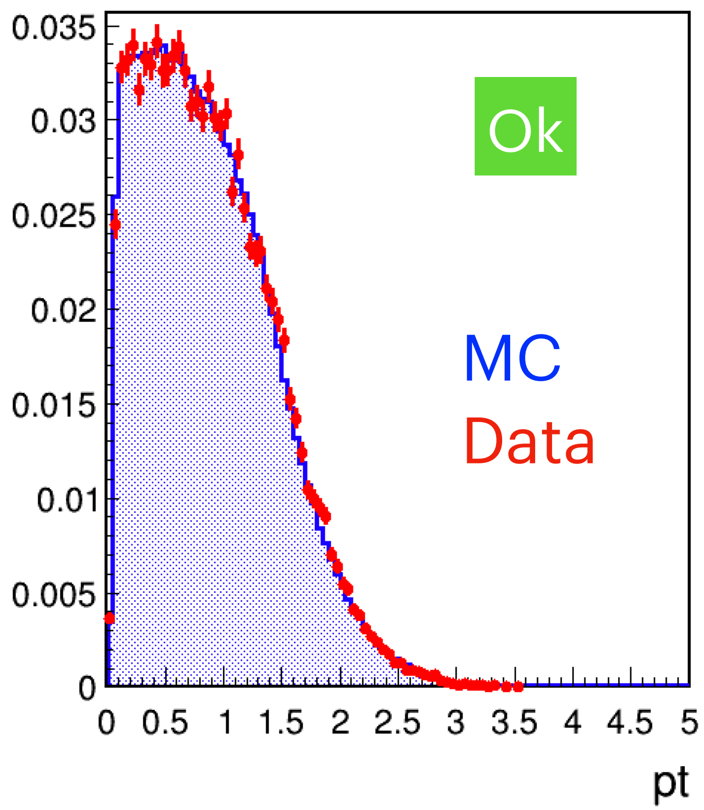
Photon MVA: inputs validation (rel-06)

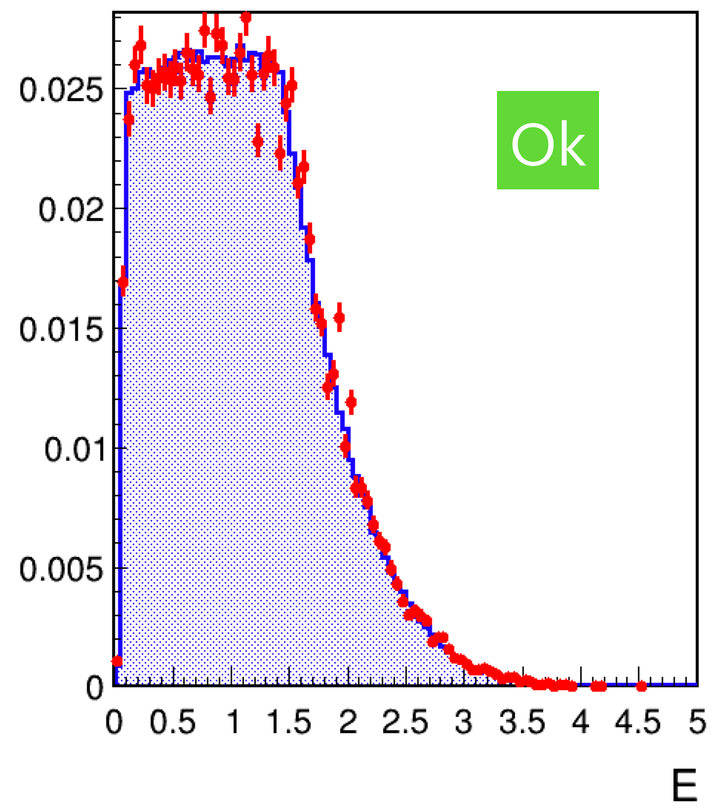


Photon MVA: inputs validation (rel-06)



Photon MVA: inputs validation (rel-06)

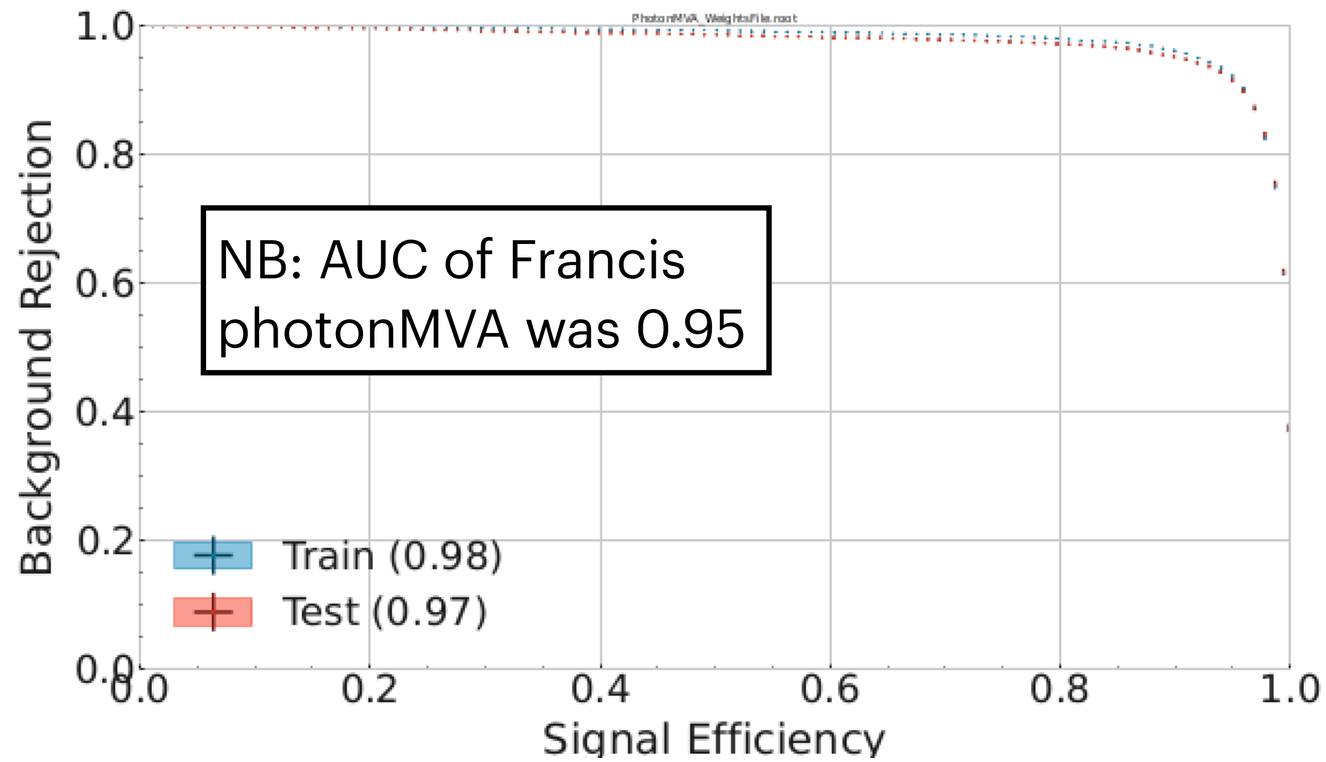


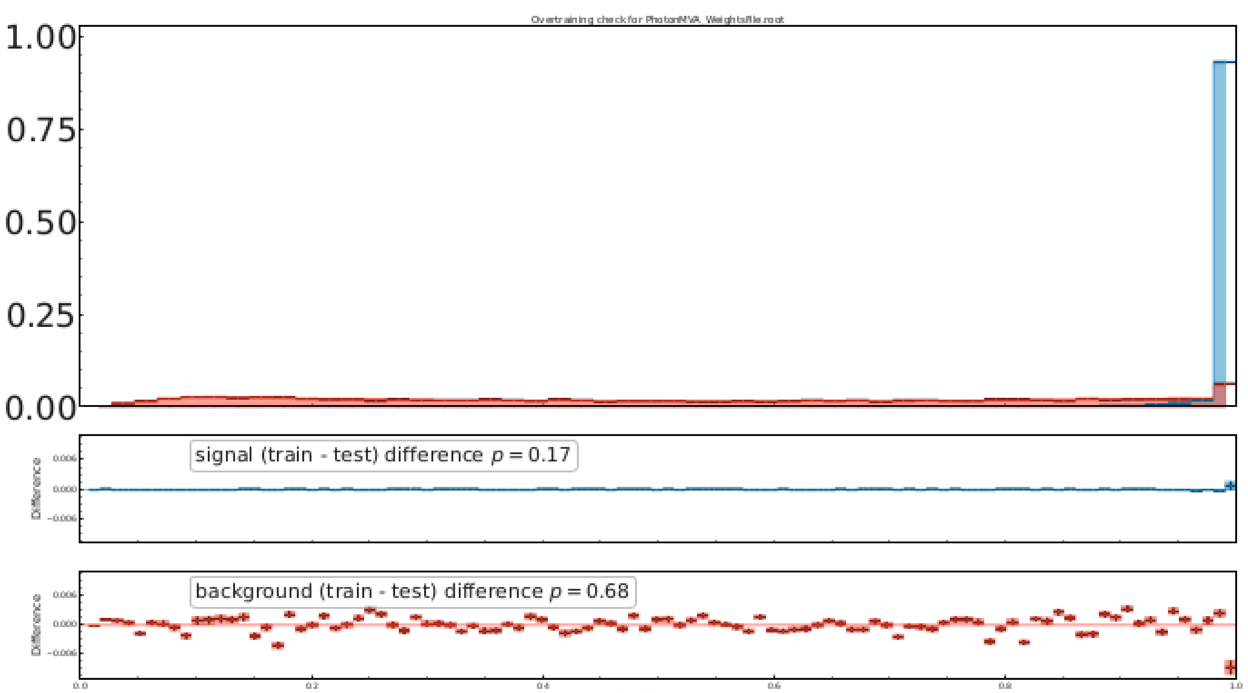


Photon MVA results using release-06

Train on MC sample after applying all π^0 selections.

Inputs (after pruning)
pt
clusterE1E9
clusterErrorPhi
clusterHighestE
clusterSecondMoment
clusterZernikeMVA
minC2TDist
clusterLAT
clusterNHits
clusterTheta
beamBackgroundSuppression

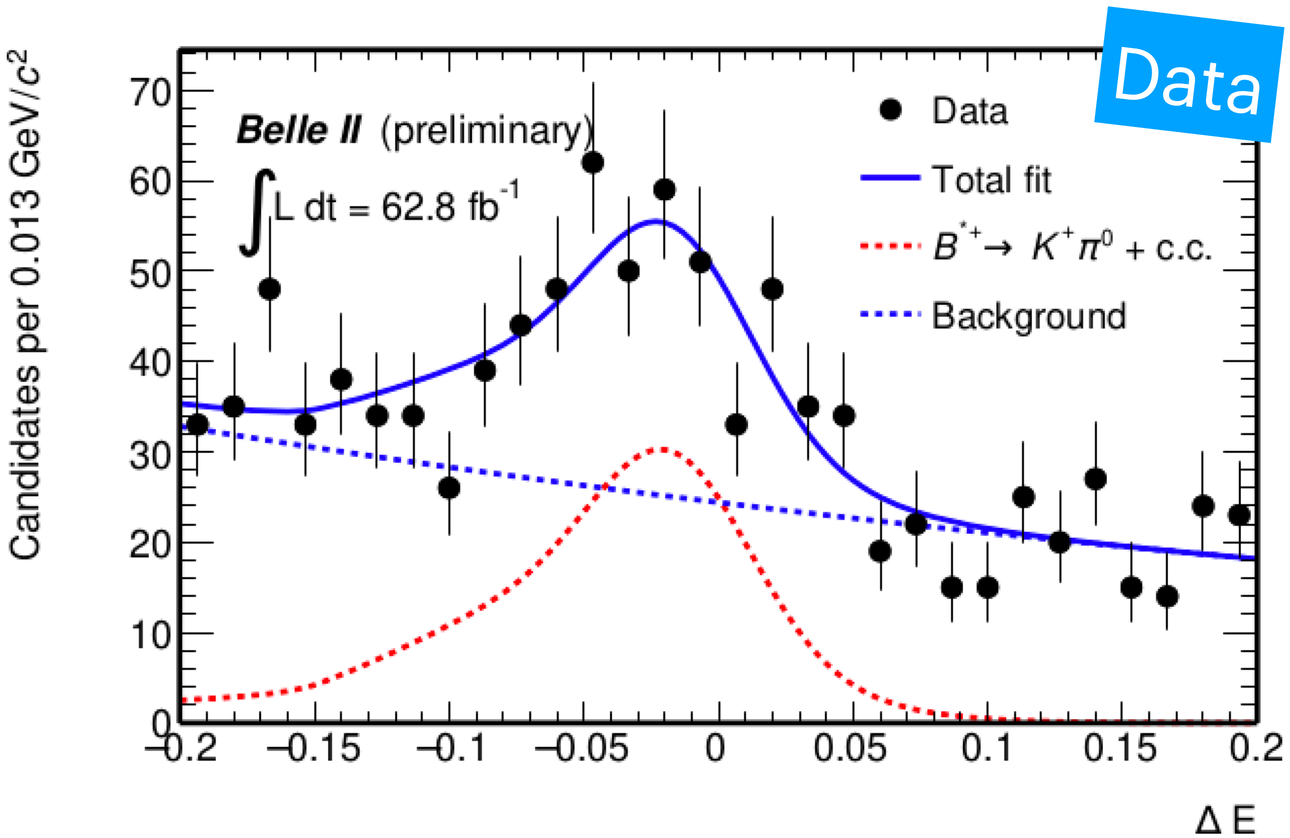




Photon MVA validation

Apply photonMVA to $B^+ \to K^+\pi^0$ proc13 sample (chunk1+chunk2 — 62fb-1).

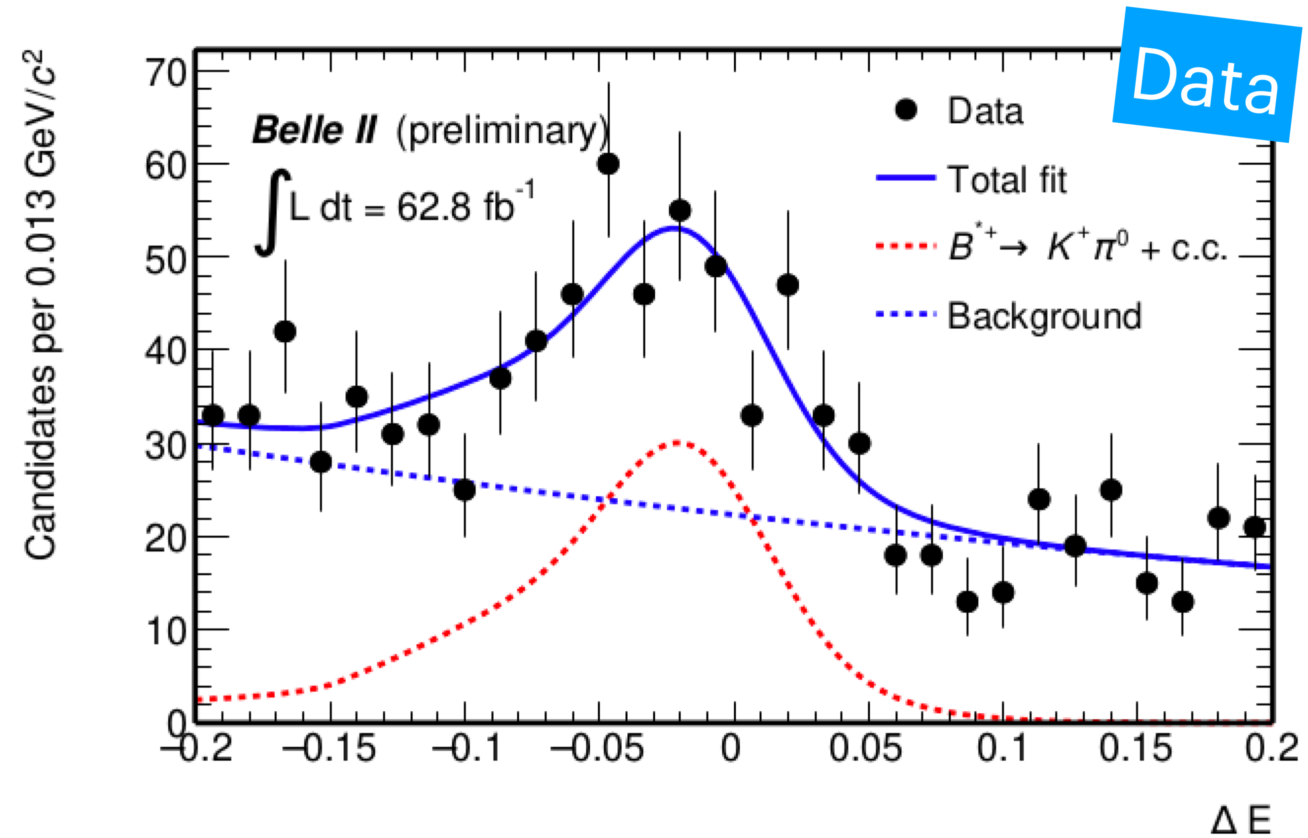
No photonMVA



Background: 742.64 ± 40.1

Signal: 260.35 ± 33.6

PhotonMVA>0.2



Background: 679.23 ± 38.6 (-8,5%)

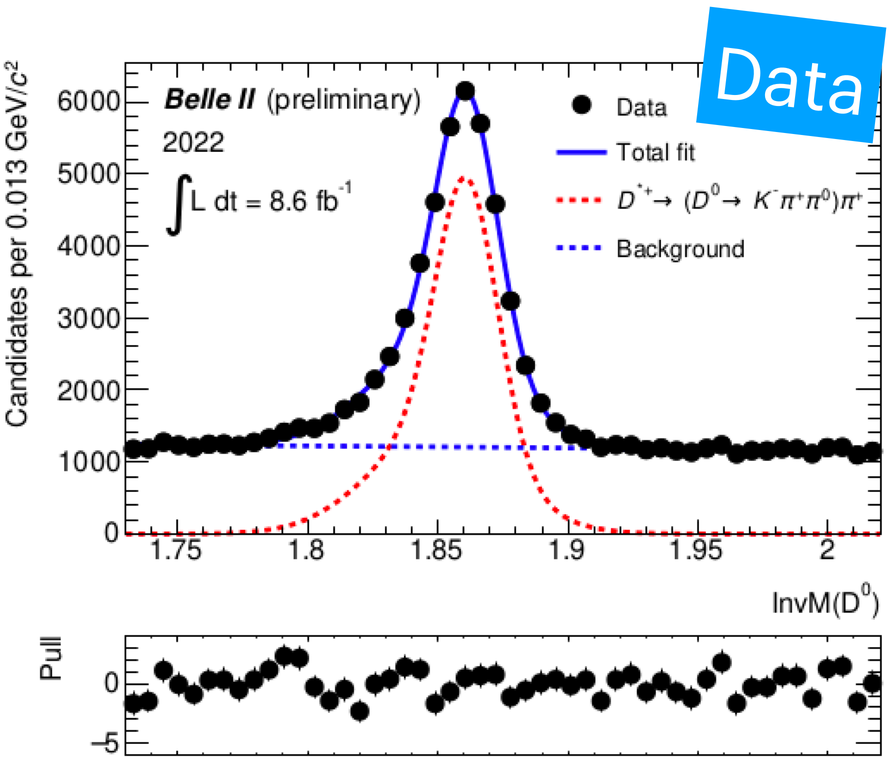
Signal: 258.76 ± 32.6 (-0.6%)

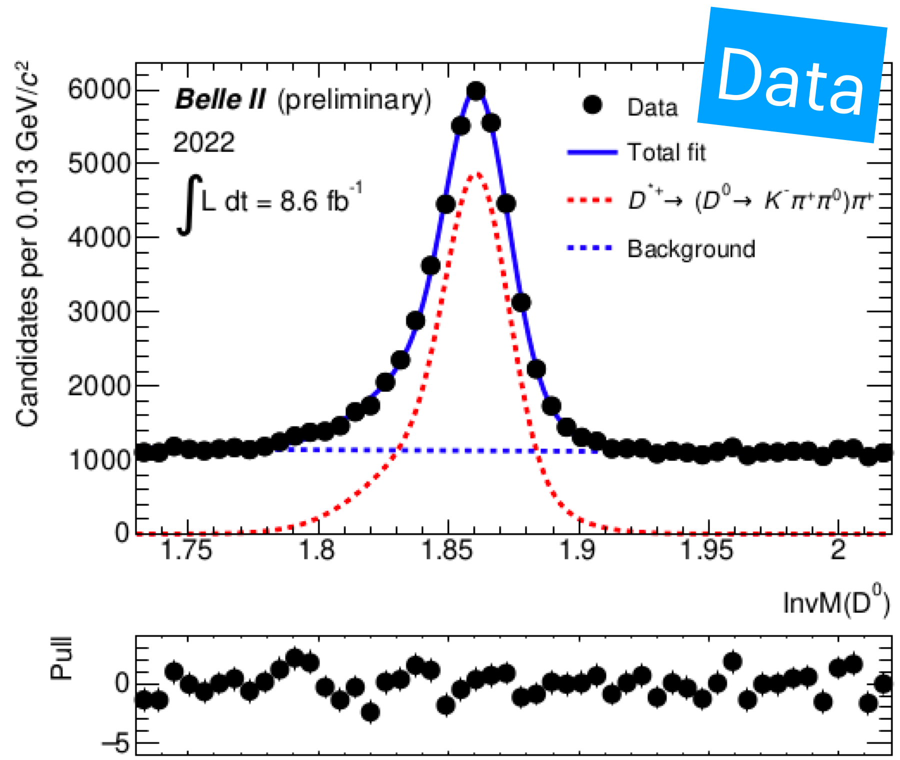
PhotonMVA works well. Modest impact, but still useful.

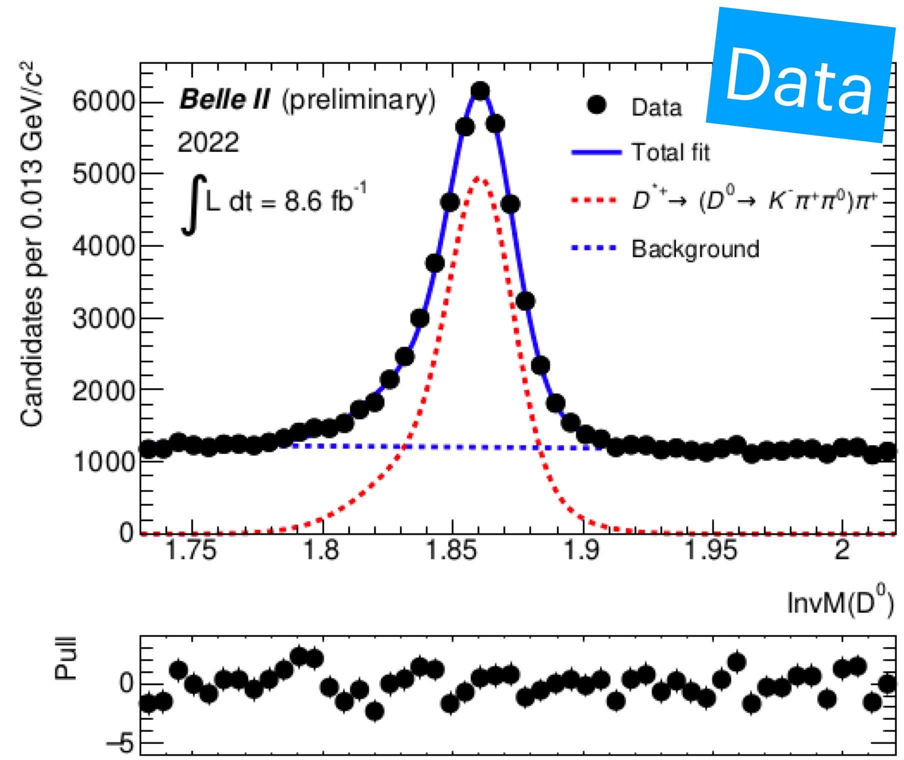
Photon MVA validation

Apply photonMVA to $D^* \rightarrow D^0(K\pi\pi^0)\pi$ proc13 sample (chunk1 — 8.6fb^{-1}).

No photonMVA | My PhotonMVA>0.05 (optimised) | Francis PhotonMVA>0.05







Background: 59902 ± 74
Signal: 33944 ± 69
Significance: 110.804

Background: 56066 ± 490 (-6.4%)
Signal: 33422 ± 528 (-1.5%)
Significance: 111.724

Background: 57410 ± 70 (-4.1%)
Signal: 33519 ± 61 (-1.3%)
Significance: 111.158

PhotonMVA works well. Modest impact, but still useful.

CSBDT

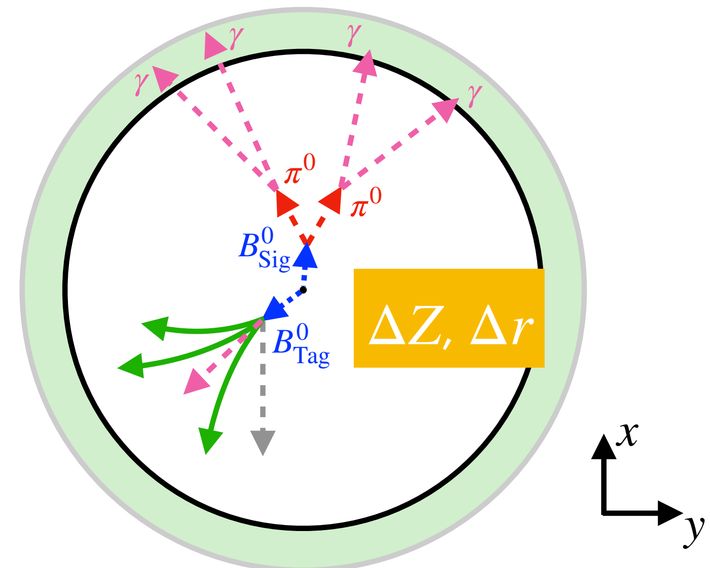
CSBDT summary

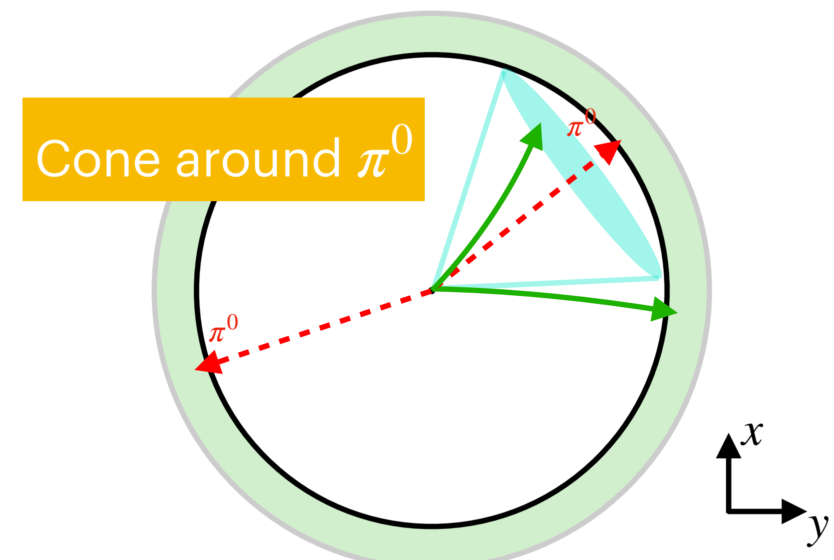
Create continuum-suppression BDT using event-shape variables and B_{Tag} variables, avoiding large correlations (<10% — was 5% for Francis) and/or sculpting.

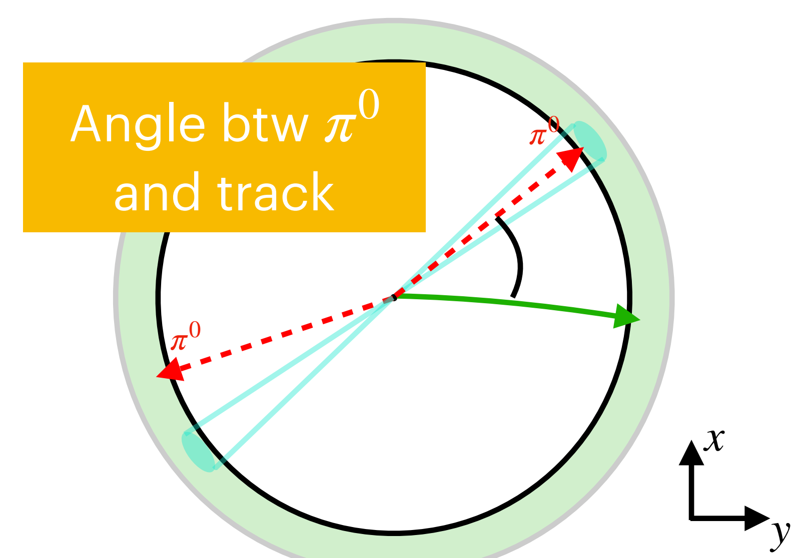
Must check if the use of B_{Tag} variables sculpts or introduces large correlations in the flavour tagger variables.

Note: 6.7% of the signal events doesn’t have a B_{Tag} vertex → remove these events (bkg: -9.4%).

New possible inputs:



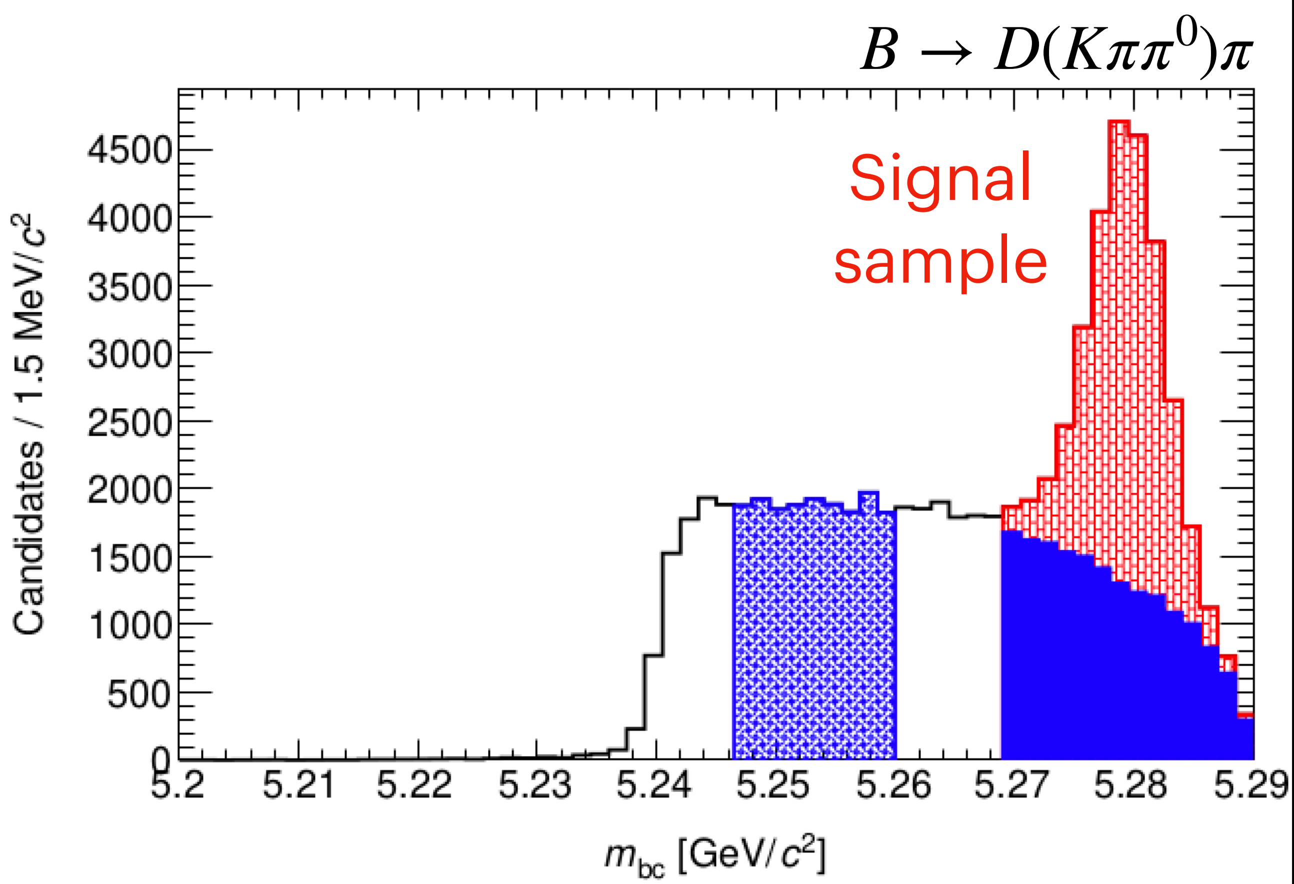




CSBDT: inputs validation

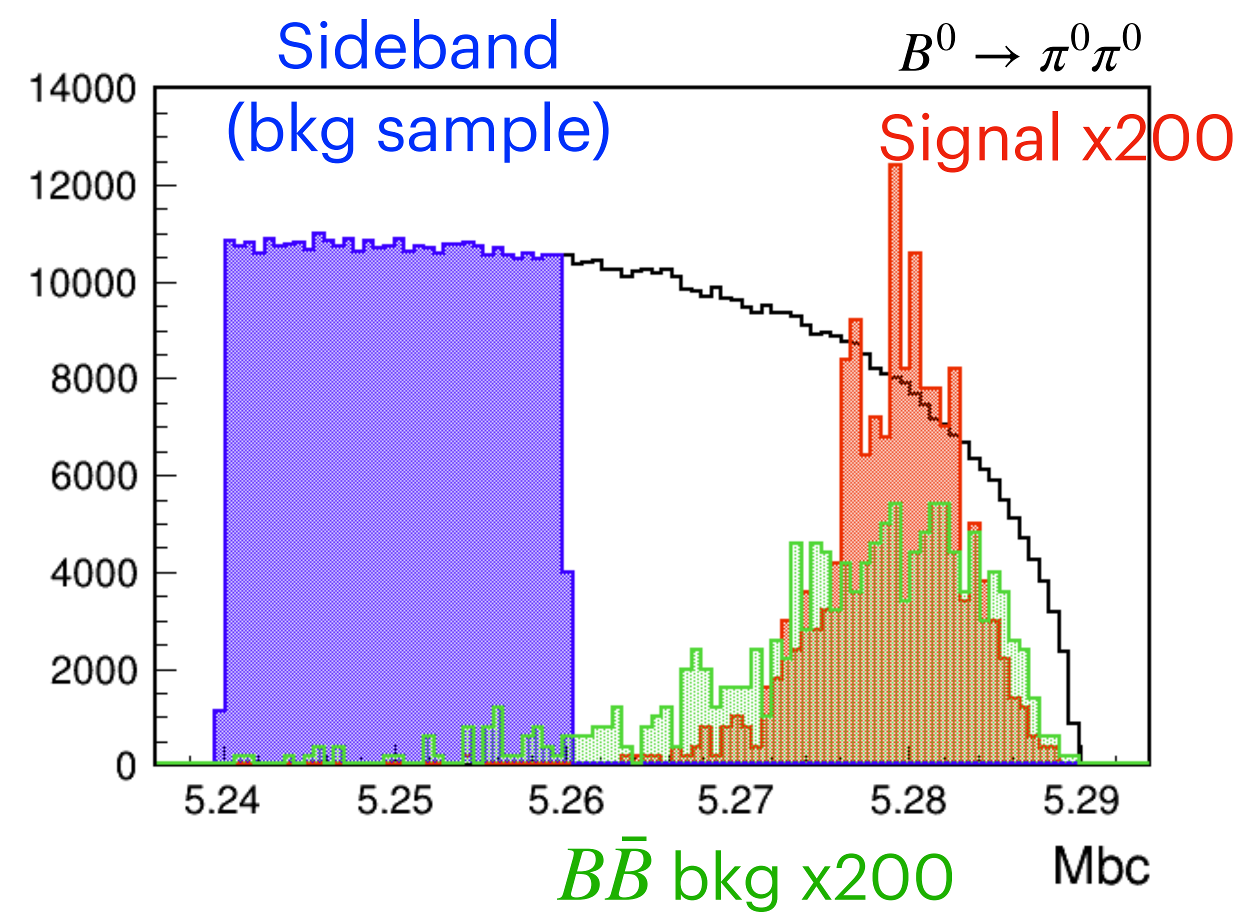
Signal: use $B \to D(K\pi\pi^0)\pi$ sideband-subtracted data (proc13) and sideband-subtracted $B \to D(K\pi\pi^0)\pi$ MC15

Do not use $B \to D(K\pi\pi^0)\pi$ for bkg because of the different compositions



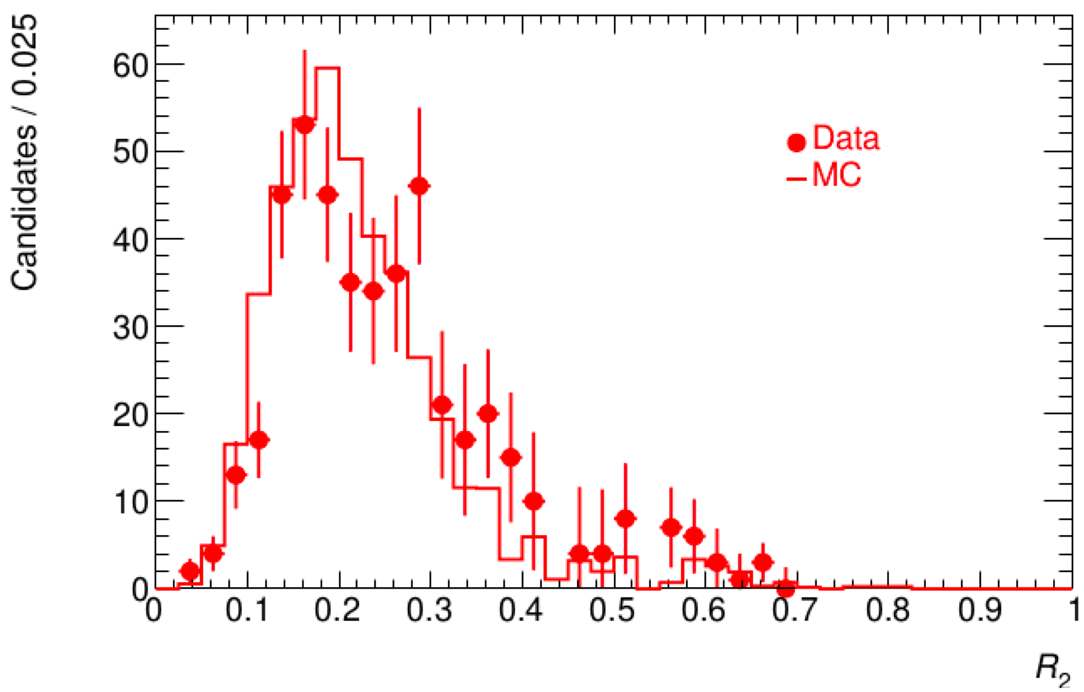
Background: use $B^0 \to \pi^0\pi^0$ sideband data (proc13) and $B^0 \to \pi^0\pi^0$ sideband MC15

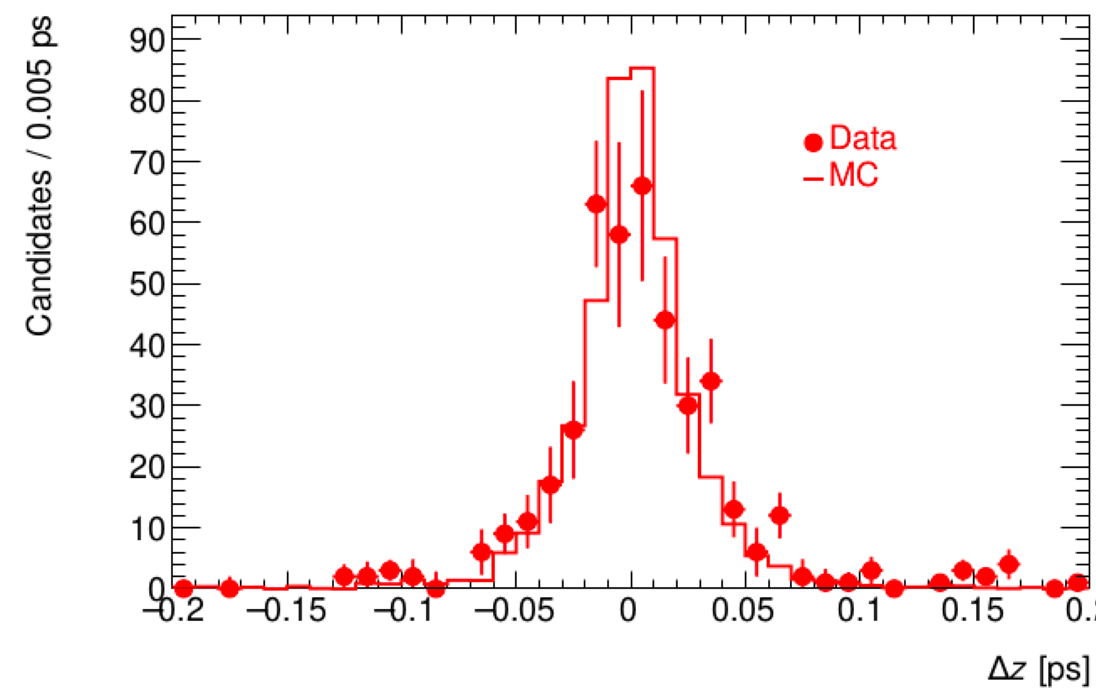
Need to check if bkg composition is the same in sideband and signal region

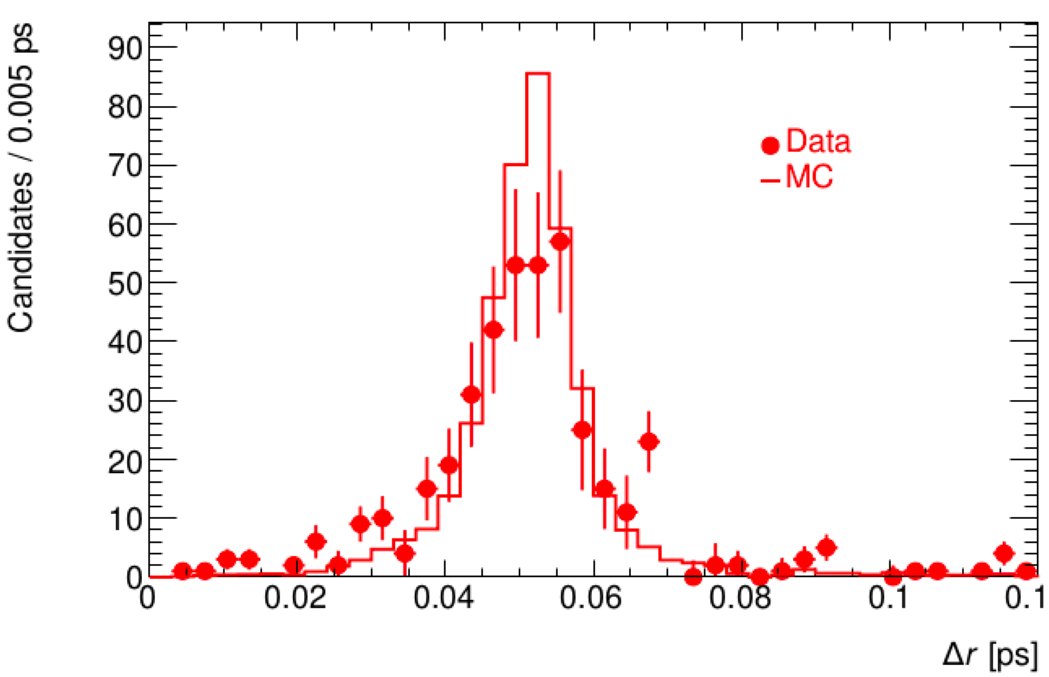


Inputs validation — Signal only

Use $B \to D(K\pi\pi^0)\pi$ sideband-subtracted data (proc13) and sideband-subtracted $B \to D(K\pi\pi^0)\pi$ MC15.



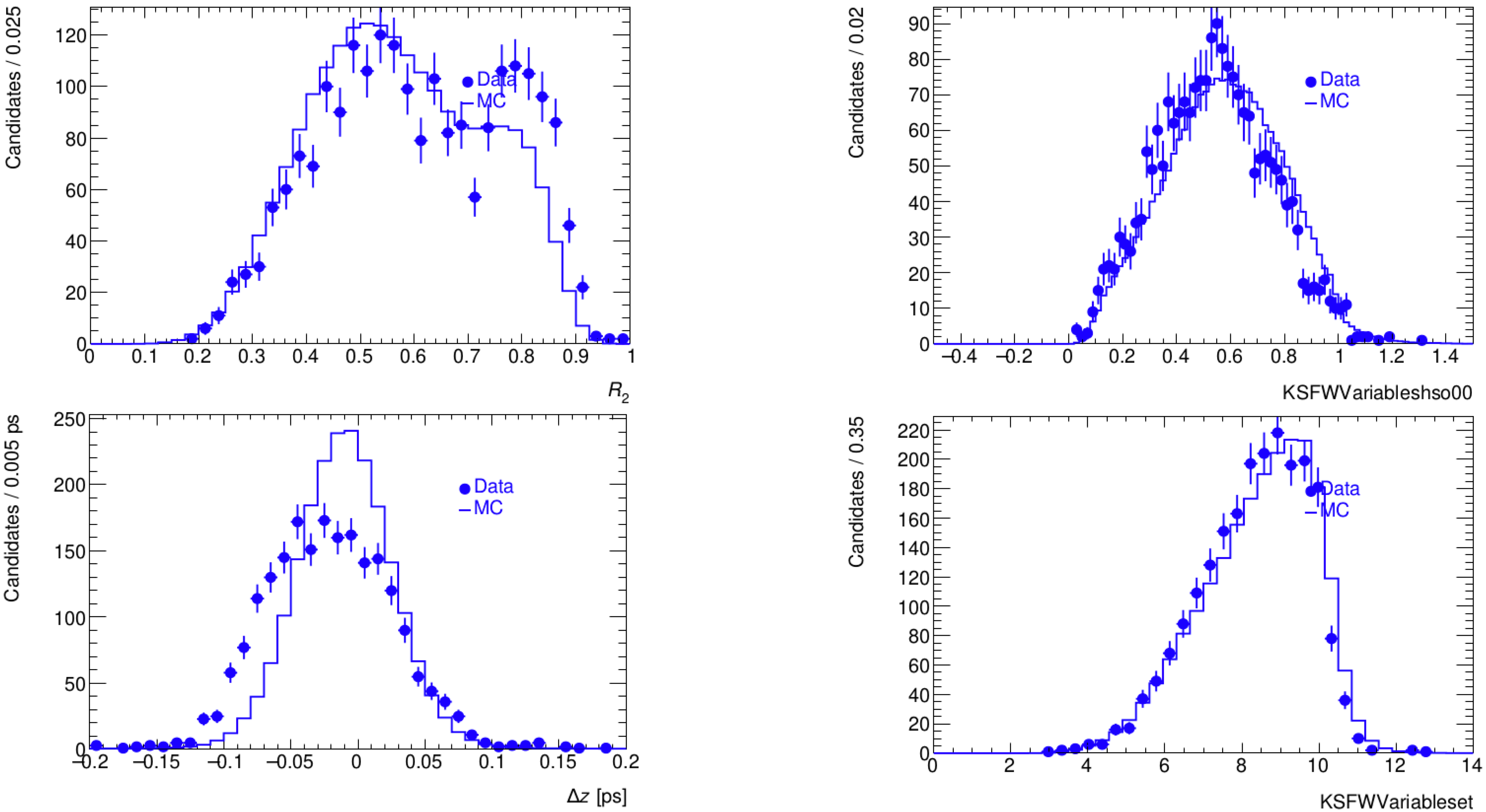




Sample has poor statistics, but do not observe any large discrepancy.

Inputs validation — Background only

Use $B^0 \to \pi^0\pi^0$ sideband data (proc13) and $B^0 \to \pi^0\pi^0$ sideband MC15

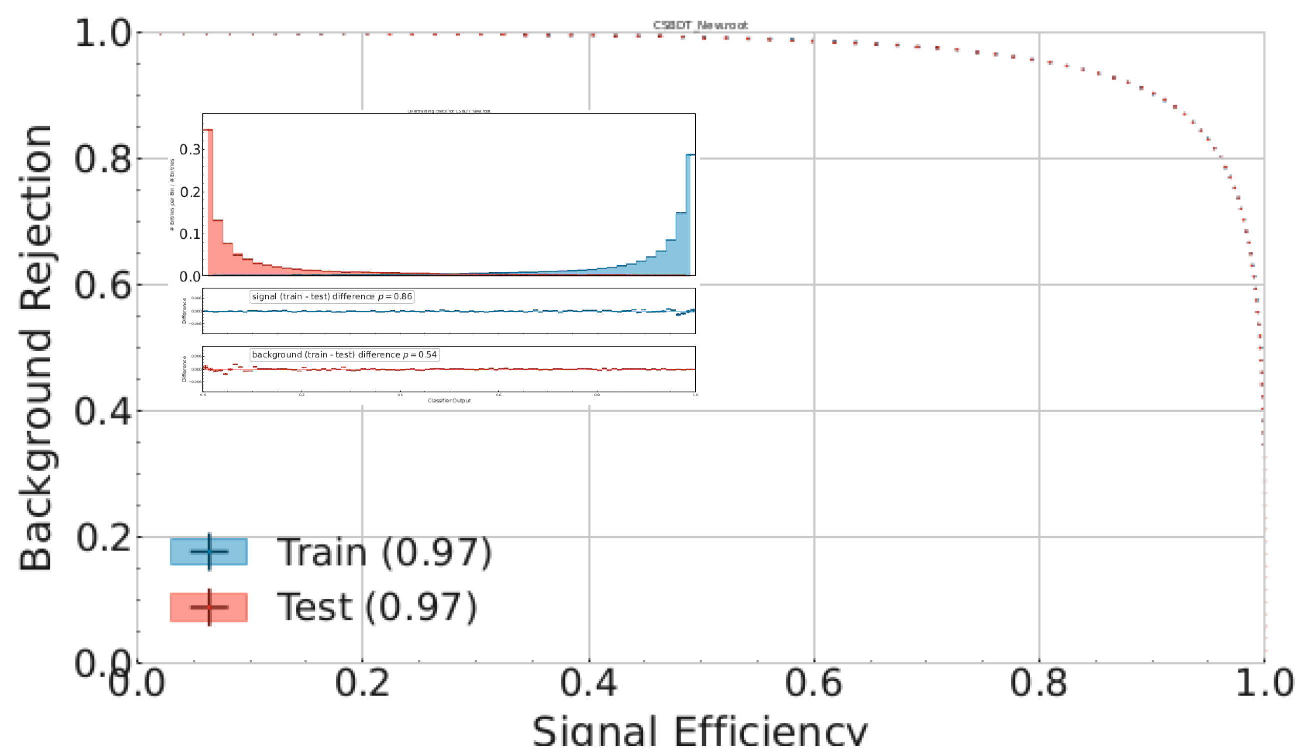


Observe variables with some discrepancies.
Better to use directly sideband data to train the CSBDT

CSMVA (preliminary) result

Train on MC sample after applying all π^0 selections.

Inputs (after pruning)
7 Kakuno-Super-Fox-Wolfram moments
cosTBTO
1 CleoCone
cosTheta*
R2
thrustOm
ΔZ (BTag)
Δr (BTag)
thrustAxisCosTheta
angle between π^0's
cosHelicityAngle
KSFWVariableset
KSFWVariablesmm2



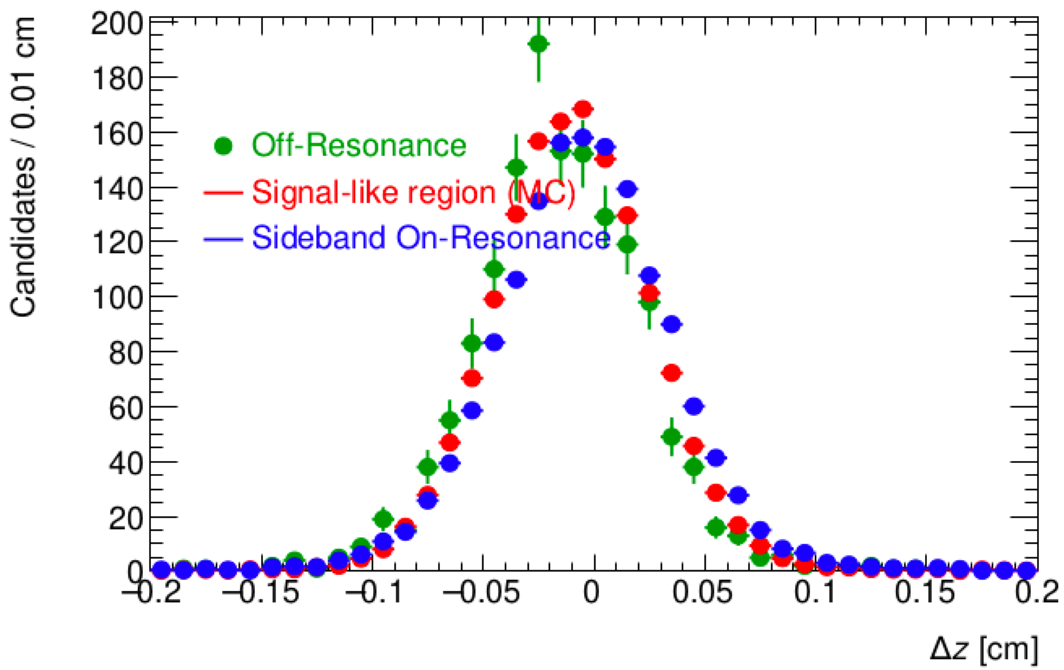
Better performance wrt old BDT (AUC=0.95).

Will repeat this using off-res data for the bkg training.

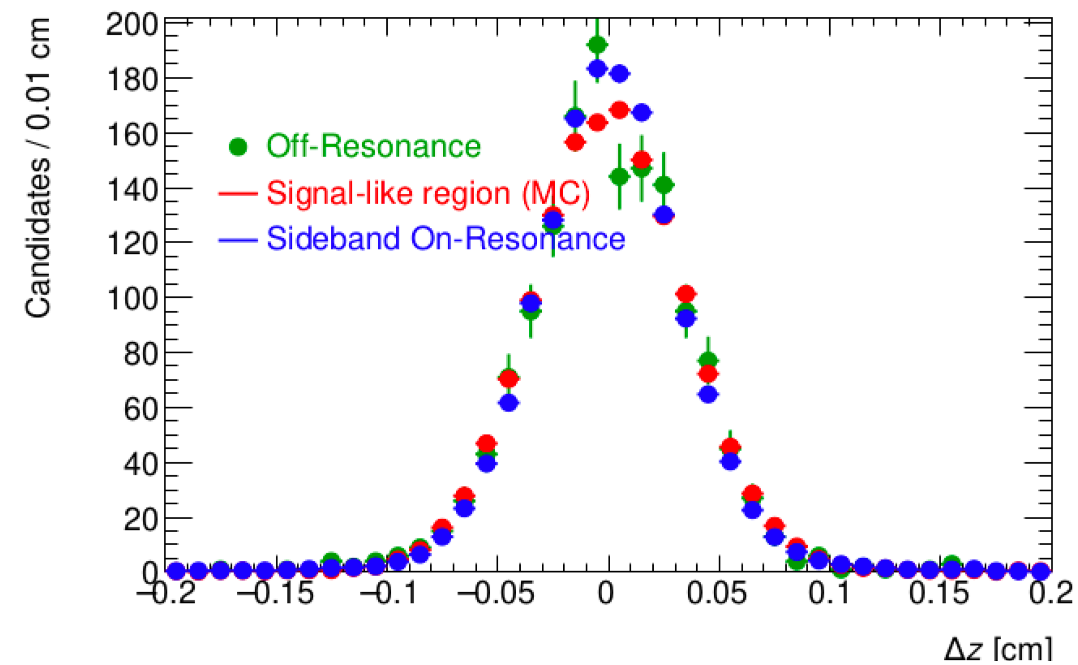
Which are the correct Δr and ΔZ distributions?

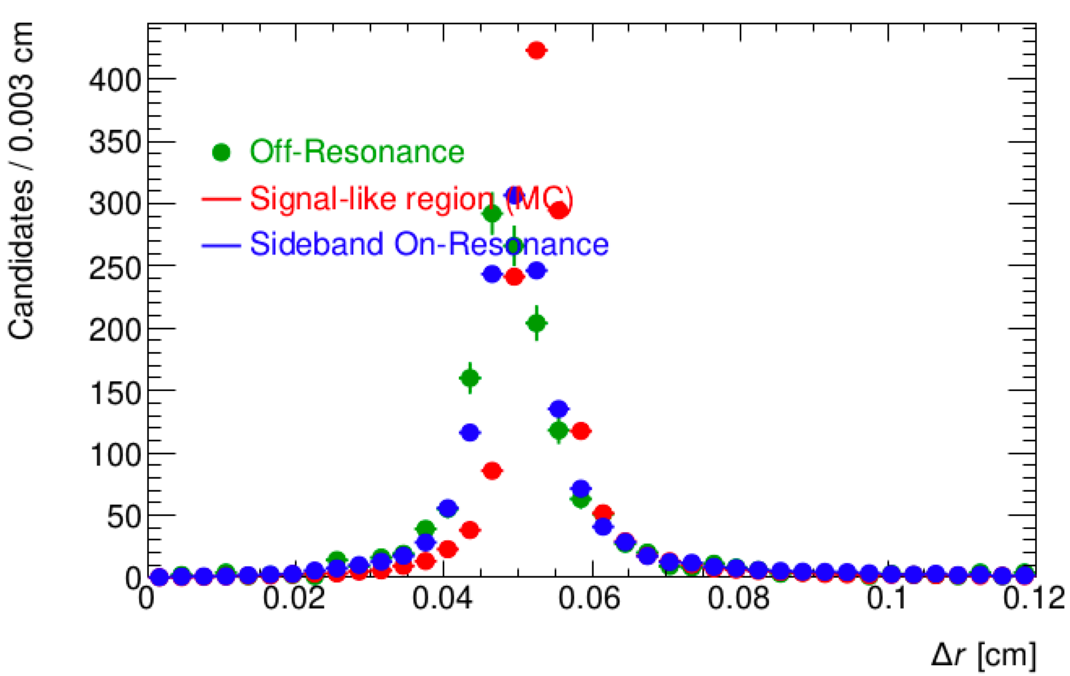
Sideband data and off-resonance data have different Δr and ΔZ distributions.
Use Δr and ΔZ distributions with respect to the IP (not the lab origin).

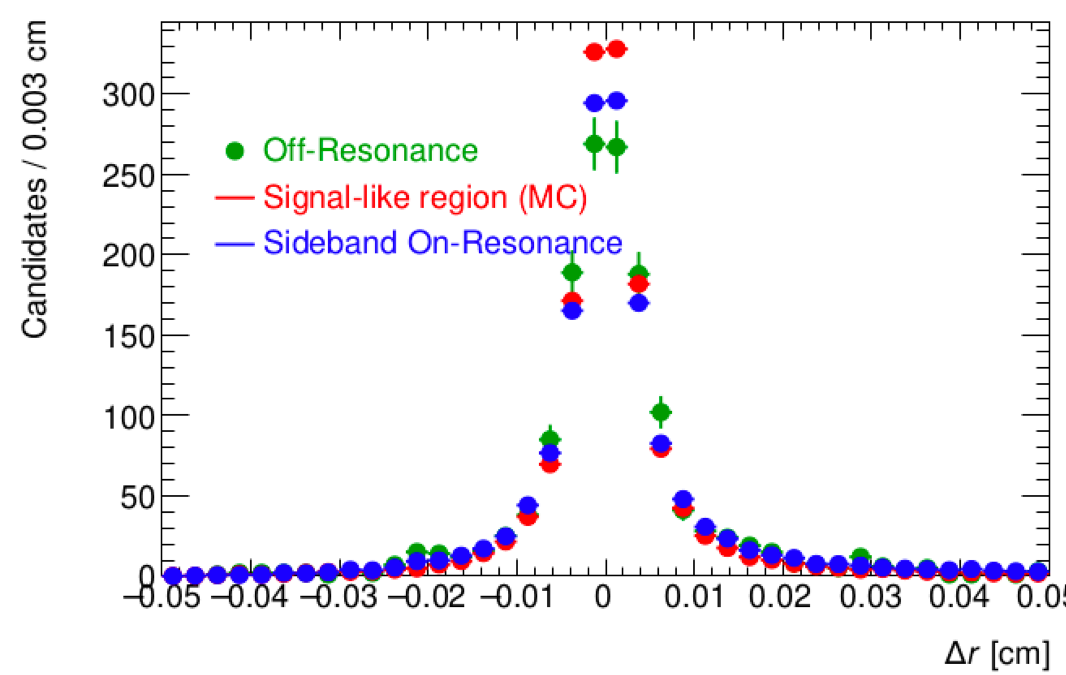
with respect to the origin



with respect to the IP



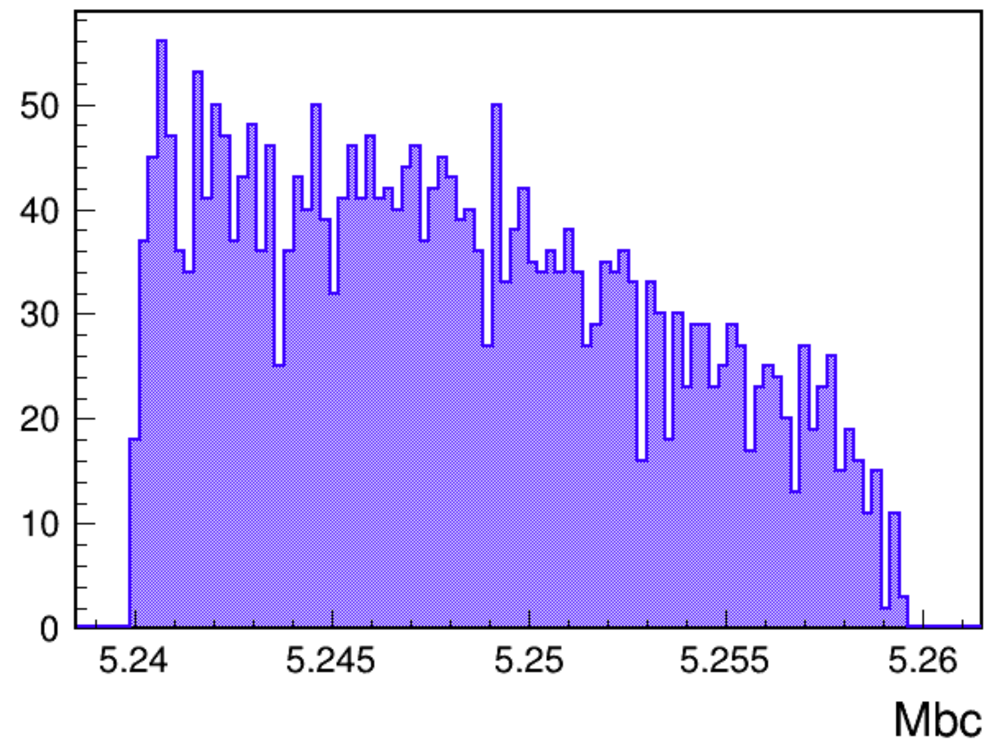




Improved situation

CSMVA using data to train bkg

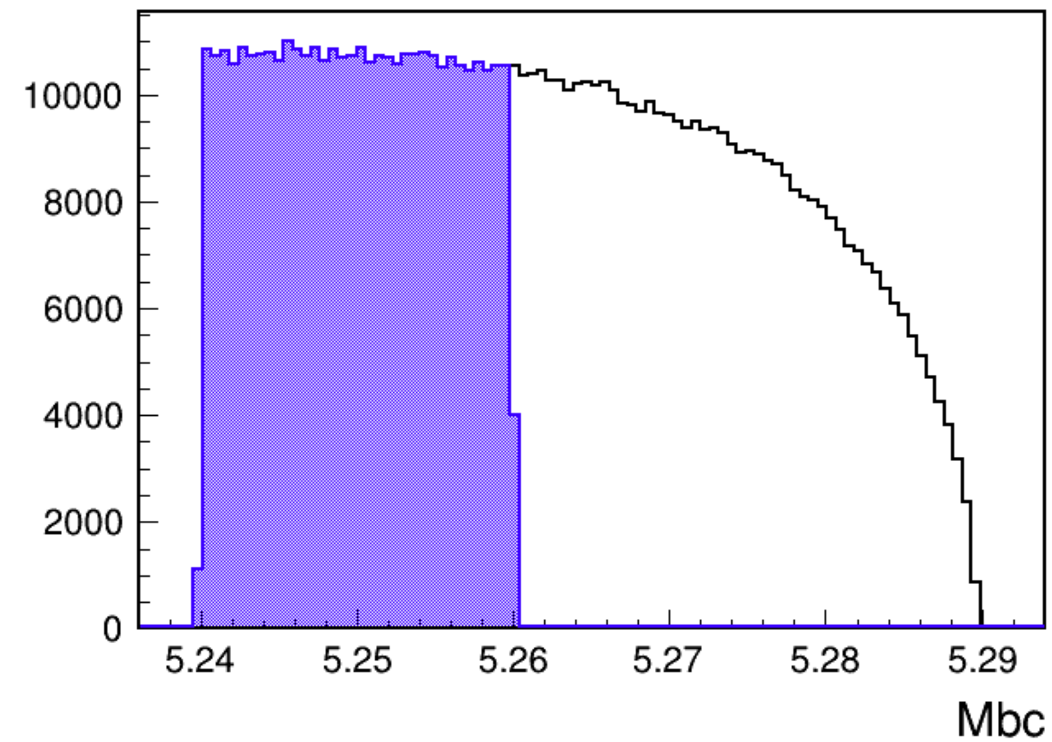
Off-resonance data



Pro: describes well the background in the signal-region in all the variables

Cons: very small amount of data

Sideband on-resonance data

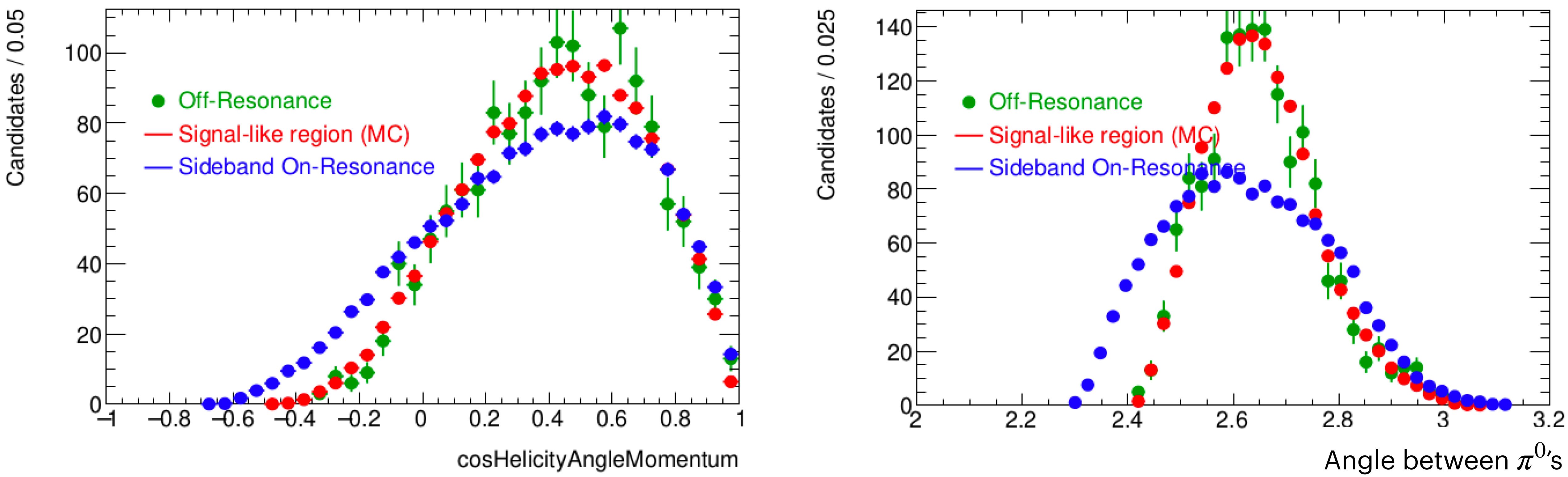


Pro: very large amount of data

Cons: doesn't describe well bkg in the signal region in two distributions (angles between pions)

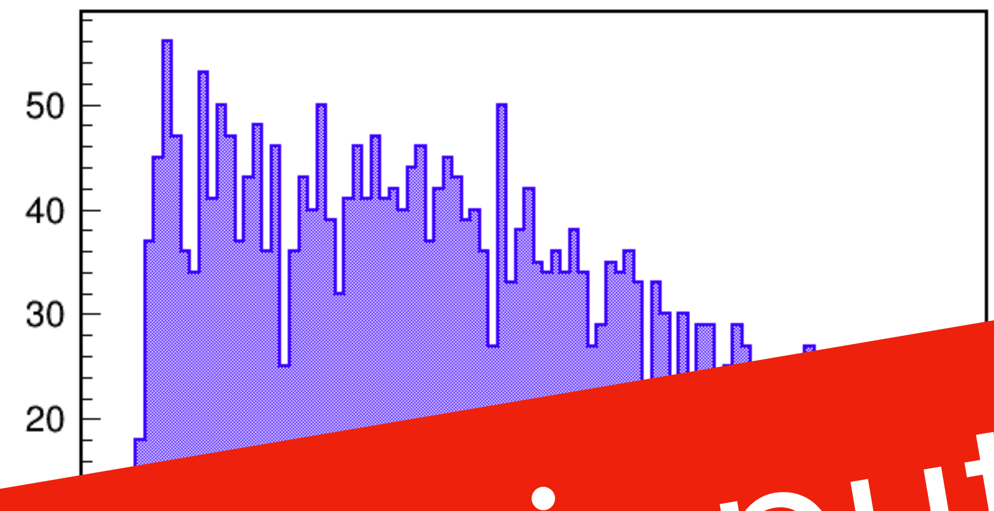
Sideband on-resonance data

Only two distributions not describing well bkg in signal region:



CSMVA using data to train bkg

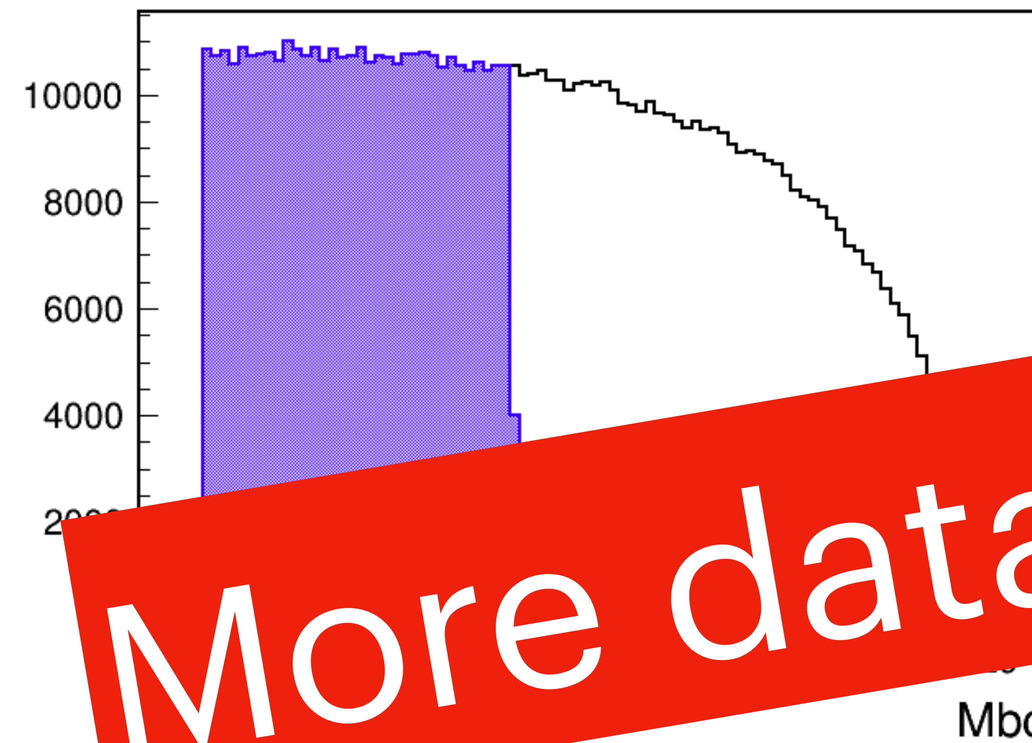
Off-resonance data



More input variables

Cons: very small amount of data

Sideband on-resonance data



More data

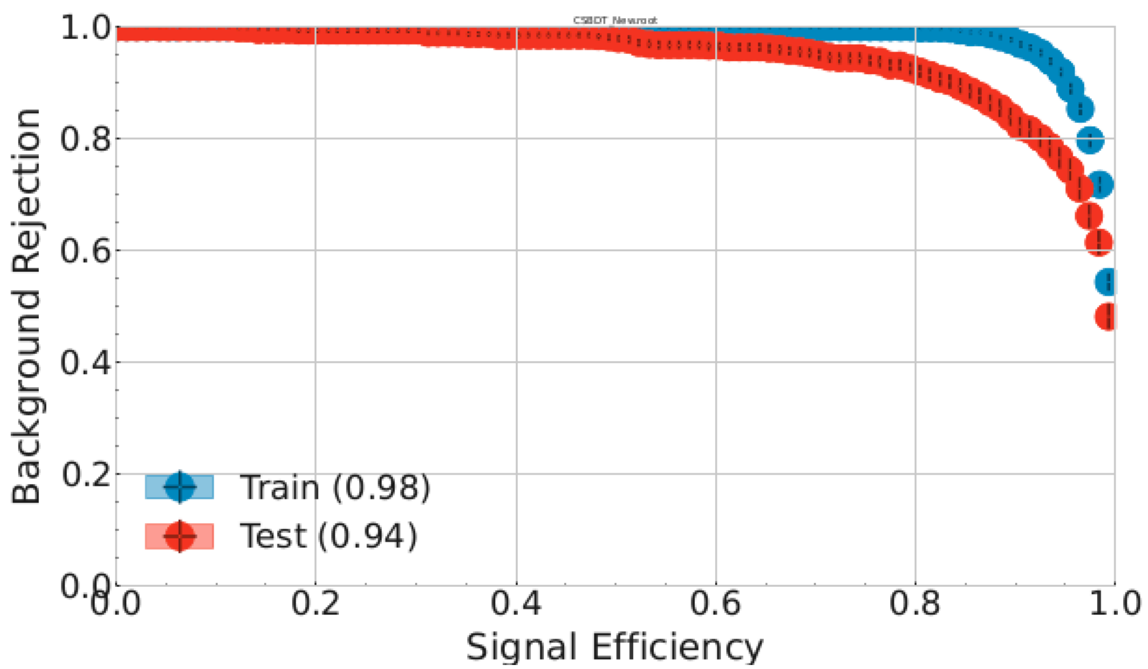
Pro: very large amount of data

Cons: doesn't describe well bkg in the signal region in two distributions (angles between pions)

CSMVA using data to train bkg

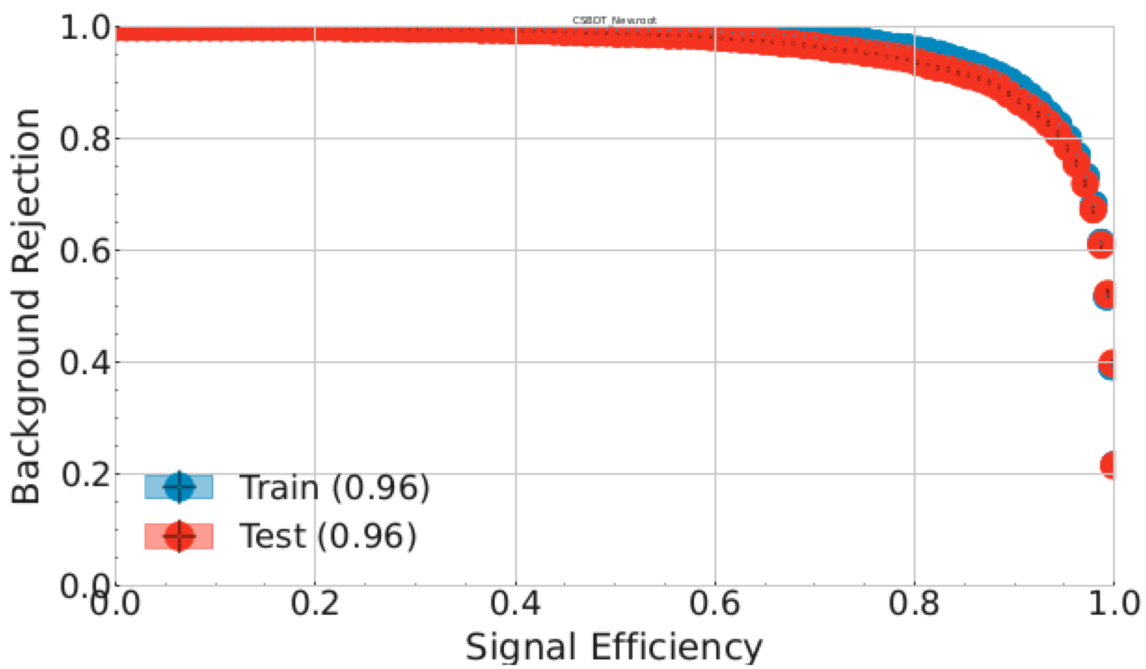
Off-resonance data

Do not exclude the 2 variables from the BDT.



Sideband on-resonance data

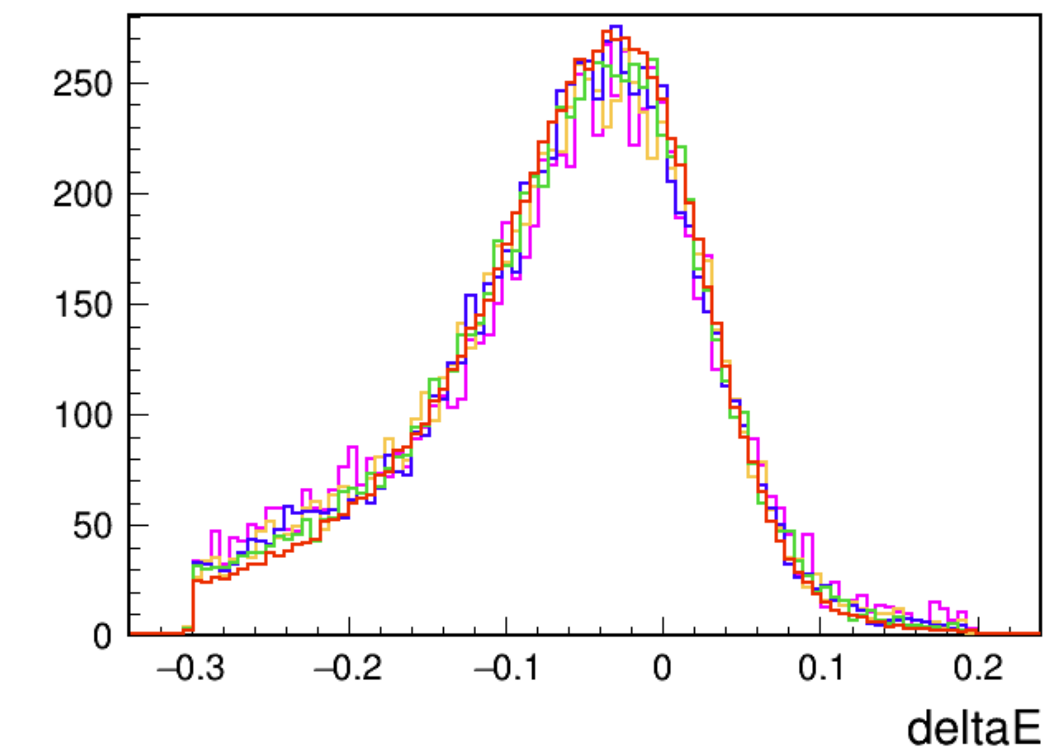
Exclude the 2 variables from the BDT.

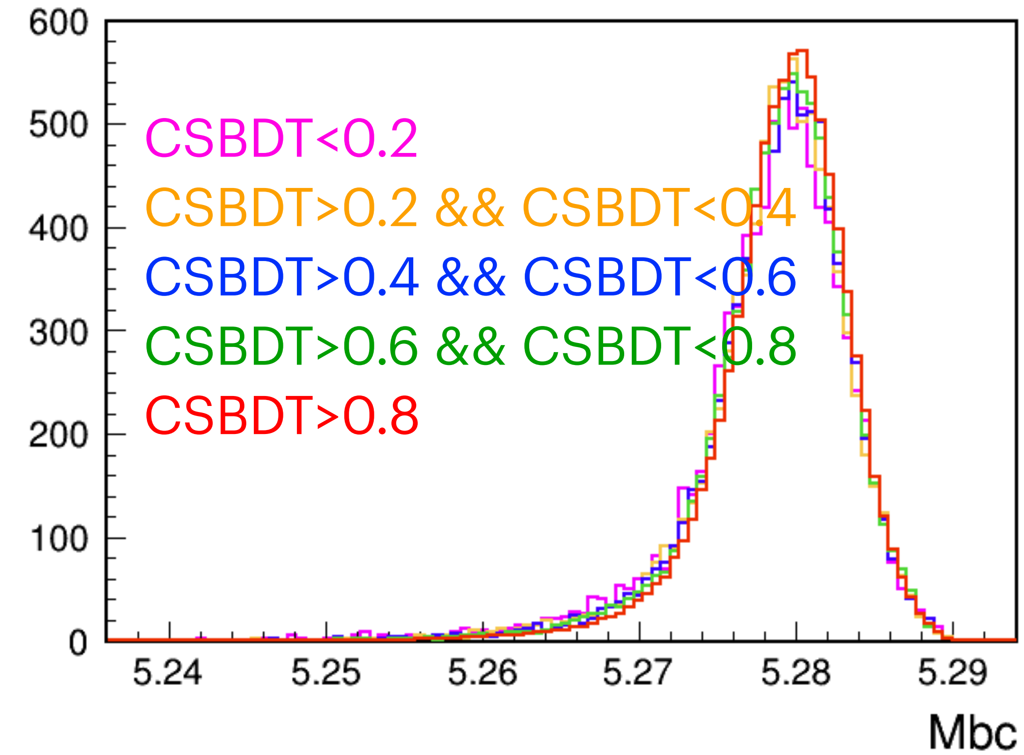


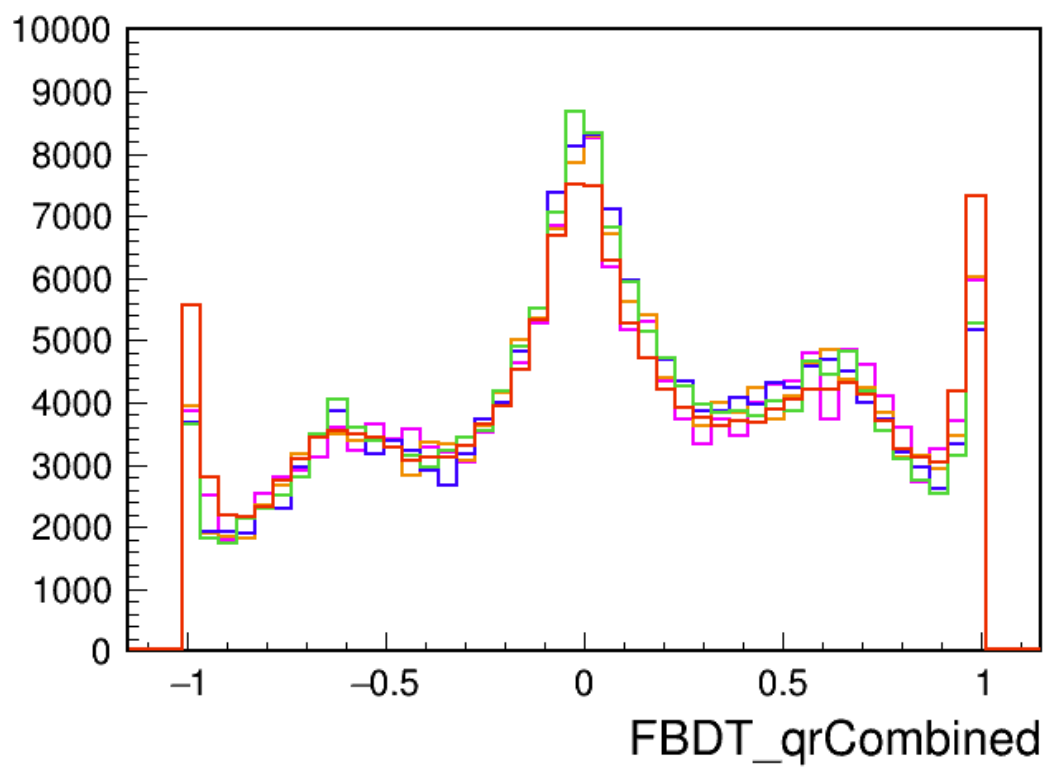
Small off-resonance data sample makes CSBDT not very reliable, while sideband data gives -2% in AUC.

CSBDT dependences with fit variables

Draw fit variables in slices of CSBDT (signalMC only).







CSBDT>0.5 && CSBDT<0.6
CSBDT>0.6 && CSBDT<0.7
CSBDT>0.7 && CSBDT<0.8
CSBDT>0.8 && CSBDT<0.9
CSBDT>0.9

Some sculpting in qr

qr variables on SignalMC

$B^0 \to \pi^0\pi^0$

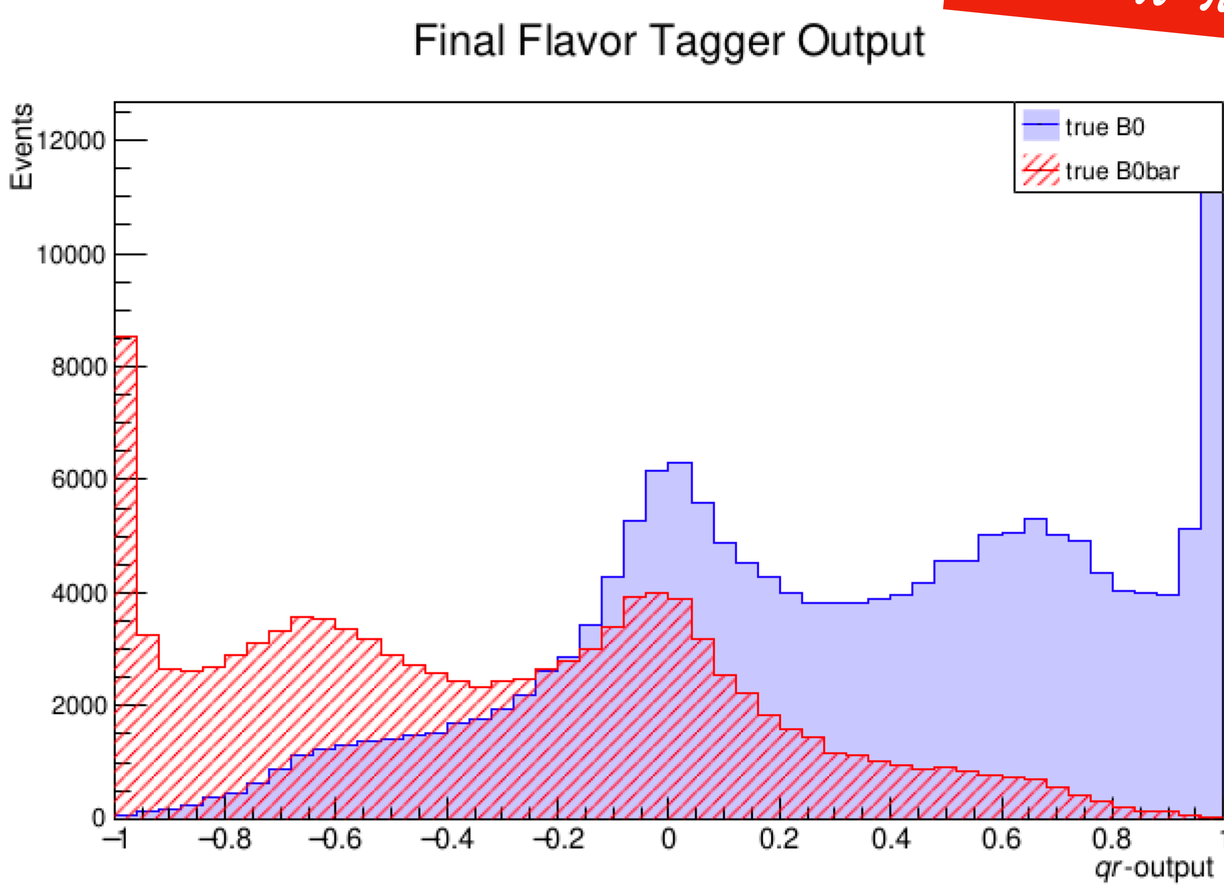
r- Interval	ε_i	$\Delta\varepsilon_i$	$w_i \pm \delta w_i$	$\Delta w_i \pm \delta\Delta w_i$	$\varepsilon_{eff,i} \pm \delta\varepsilon_{eff,i}$	$\Delta\varepsilon_{eff,i} \pm \delta\Delta\varepsilon_{eff,i}$
$0.000 - 0.100$	16.9	0.17	47.51 ± 0.24	2.47 ± 0.47	0.0420 ± 0.0080	-0.0829 ± 0.0193
$0.100 - 0.250$	16.6	0.12	41.01 ± 0.24	0.85 ± 0.47	0.5356 ± 0.0283	-0.0981 ± 0.0571
$0.250 - 0.500$	21.1	0.75	29.90 ± 0.20	0.64 ± 0.39	3.4112 ± 0.0675	-0.0948 ± 0.1350
$0.500 - 0.625$	11.7	-0.26	20.87 ± 0.23	1.87 ± 0.46	3.9781 ± 0.0667	-0.5973 ± 0.1346
$0.625 - 0.750$	11.5	-0.03	15.09 ± 0.21	0.81 ± 0.41	5.5976 ± 0.0728	-0.2756 ± 0.1460
$0.750 - 0.875$	8.7	0.06	8.27 ± 0.18	0.56 ± 0.37	6.0356 ± 0.0653	-0.1225 ± 0.1307
$0.875 - 1.000$	13.5	-0.81	1.74 ± 0.07	0.21 ± 0.14	12.5839 ± 0.0721	-0.8637 ± 0.1445
Total			$\varepsilon_{eff} = \sum_i \varepsilon_i \cdot \langle 1 - 2w_i \rangle^2 = 32.18 \pm 0.16$		$\Delta\varepsilon_{eff} = -2.13 \pm 0.32$	

$B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0$

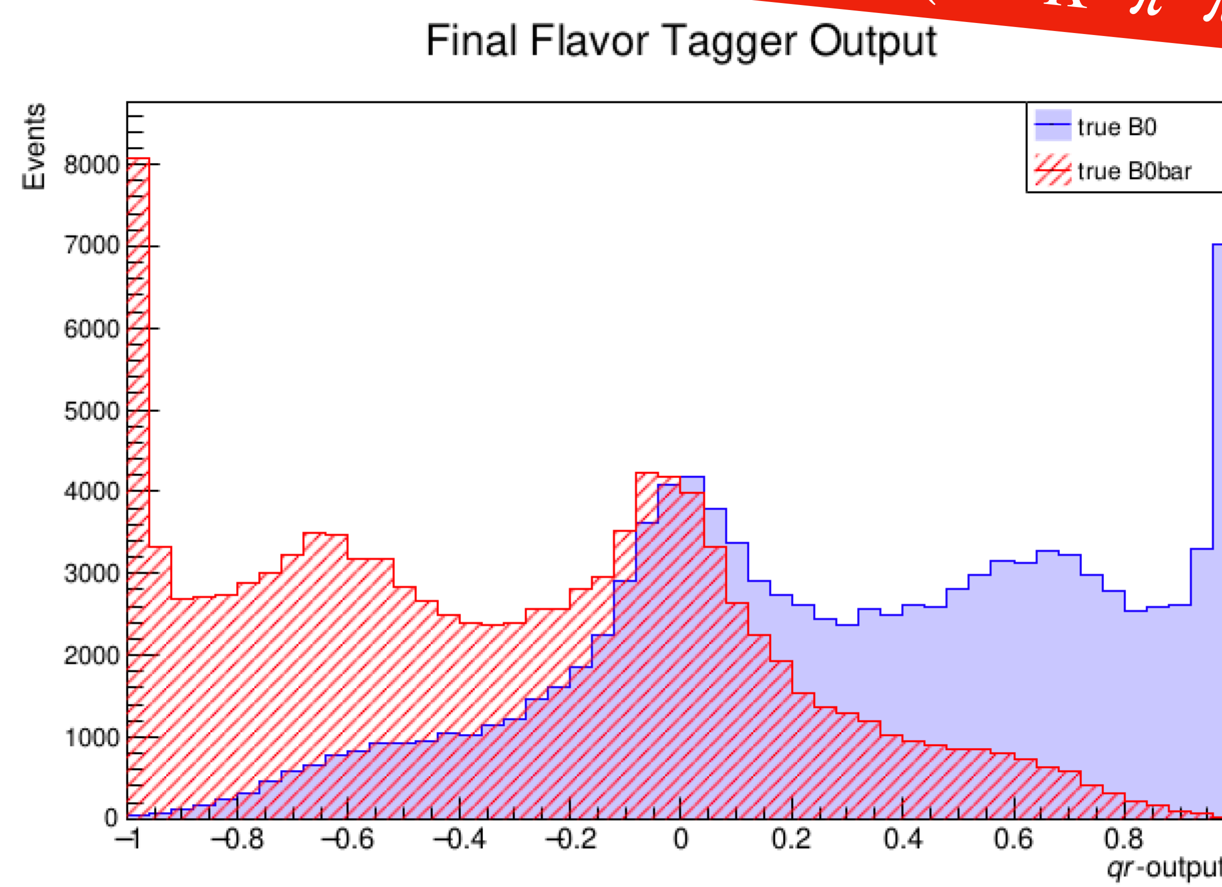
r- Interval	ε_i	$\Delta\varepsilon_i$	$w_i \pm \delta w_i$	$\Delta w_i \pm \delta\Delta w_i$	$\varepsilon_{eff,i} \pm \delta\varepsilon_{eff,i}$	$\Delta\varepsilon_{eff,i} \pm \delta\Delta\varepsilon_{eff,i}$
$0.000 - 0.100$	17.7	0.27	47.27 ± 0.26	2.58 ± 0.51	0.0526 ± 0.0099	-0.0986 ± 0.0217
$0.100 - 0.250$	16.8	0.26	41.09 ± 0.26	1.07 ± 0.52	0.5322 ± 0.0311	-0.1189 ± 0.0623
$0.250 - 0.500$	21.1	0.60	30.37 ± 0.22	0.24 ± 0.43	3.2461 ± 0.0728	0.0138 ± 0.1457
$0.500 - 0.625$	11.5	0.06	21.40 ± 0.26	1.66 ± 0.52	3.7607 ± 0.0723	-0.4175 ± 0.1445
$0.625 - 0.750$	11.2	-0.13	15.23 ± 0.23	0.97 ± 0.46	5.4192 ± 0.0794	-0.3649 ± 0.1588
$0.750 - 0.875$	8.6	-0.22	8.46 ± 0.20	0.46 ± 0.41	5.9675 ± 0.0721	-0.2828 ± 0.1442
$0.875 - 1.000$	13.1	-0.85	1.62 ± 0.08	0.47 ± 0.15	12.2983 ± 0.0783	-1.0309 ± 0.1565
Total			$\varepsilon_{eff} = \sum_i \varepsilon_i \cdot \langle 1 - 2w_i \rangle^2 = 31.28 \pm 0.17$		$\Delta\varepsilon_{eff} = -2.30 \pm 0.34$	

qr variables on SignalMC

$B^0 \to \pi^0\pi^0$



$B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0$




```
TOTAL NUMBER OF TAGGED EVENTS  =  273256
TOTAL AVERAGE EFFICIENCY  (q=+-1)=  100.00 +- 0.00 %
TOTAL AVERAGE EFFECTIVE EFFICIENCY  (q=+-1)=  32.184039 +- 0.156912 %
TOTAL AVERAGE EFFECTIVE EFFICIENCY ASYMMETRY (q=+-1)=  -2.134811 +- 0.315022 %
B0-TAGGER  TOTAL EFFECTIVE EFFICIENCIES: 31.14 +- 0.20 % (q=+1)  33.27 +- 0.25 % (q=-1)
FLAVOR PERCENTAGE (MC):                  60.52 % (q=+1)  39.48 % (q=-1)  Diff=21.03 %
```

```
TOTAL NUMBER OF TAGGED EVENTS  =  214462
TOTAL AVERAGE EFFICIENCY  (q=+-1)=  100.00 +- 0.00 %
TOTAL AVERAGE EFFECTIVE EFFICIENCY  (q=+-1)=  31.276453 +- 0.171015 %
TOTAL AVERAGE EFFECTIVE EFFICIENCY ASYMMETRY (q=+-1)=  -2.299739 +- 0.341994 %
B0-TAGGER  TOTAL EFFECTIVE EFFICIENCIES: 30.15 +- 0.24 % (q=+1)  32.45 +- 0.24 % (q=-1)
FLAVOR PERCENTAGE (MC):                  49.59 % (q=+1)  50.41 % (q=-1)  Diff=-0.82 %
```

qr variables on SignalMC15

Check after CS selection (>0.7).

	Default CS	Default CS + Δr and ΔZ	Default CS + Δr, ΔZ, and ROE tracks	Sato-san parameters
Total effective efficiency (q=+-1)	33.48 ± 0.20 %	33.16 ± 0.19 %	34.81 ± 0.19 %	33.73 ± 0.03 %
Total effective efficiency asymmetry	-2.84 ± 0.40 %	-2.58 ± 0.39 %	-2.50 ± 0.39 %	−0.09 ± 0.06 %
B^0 effective efficiency	32.10 ± 0.24 %	31.90 ± 0.24 %	33.59 ± 0.24 %	33,69 ± ? %
B^0bar effective efficiency	34.93 ± 0.31 %	34.48 ± 0.30 %	36.09 ± 0.30 %	33,78 ± ? %

Flavor tagging validation: to do

Sato-san parameters

	Signal channel	Control channel
Data	~~$B^0 \to \pi^0\pi^0$ Data~~	$B^0 \to D^-(K^+\pi^-\pi^-)\pi^+$ Data Samples in production
MC	$B^0 \to \pi^0\pi^0$ MC Done	$B^0 \to D^-(K^+\pi^-\pi^-)\pi^+$ MC Samples in production

Flavor tagging: $B^0 \to D^-(K^+\pi^-\pi^-)\pi^+$ in MC

Using ~100fb^-1 of MC with no selections at all, I obtain $\Delta\varepsilon = -3 \pm 1.3\,\%$.

Expect not very good precision with the current data.

ρ MVA

ρ MVA

Beyond the CS: identify the principal bkg components.

	Events that have at least a π^0 from ...
$\rho(770)^+$	47.1%
Z^0 (direct from e+e-)	75.0%

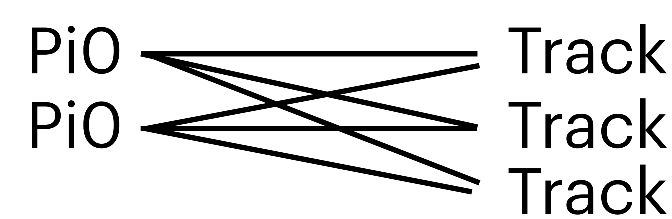
Large number of continuum π^0's come from a ρ $\rightarrow$ develop a specific BDT (in addition to the default CS BDT).

Combine each track in the event with each π^0.

Use kinematic and angular variables to distinguish between ρ's and other particles.

Example:

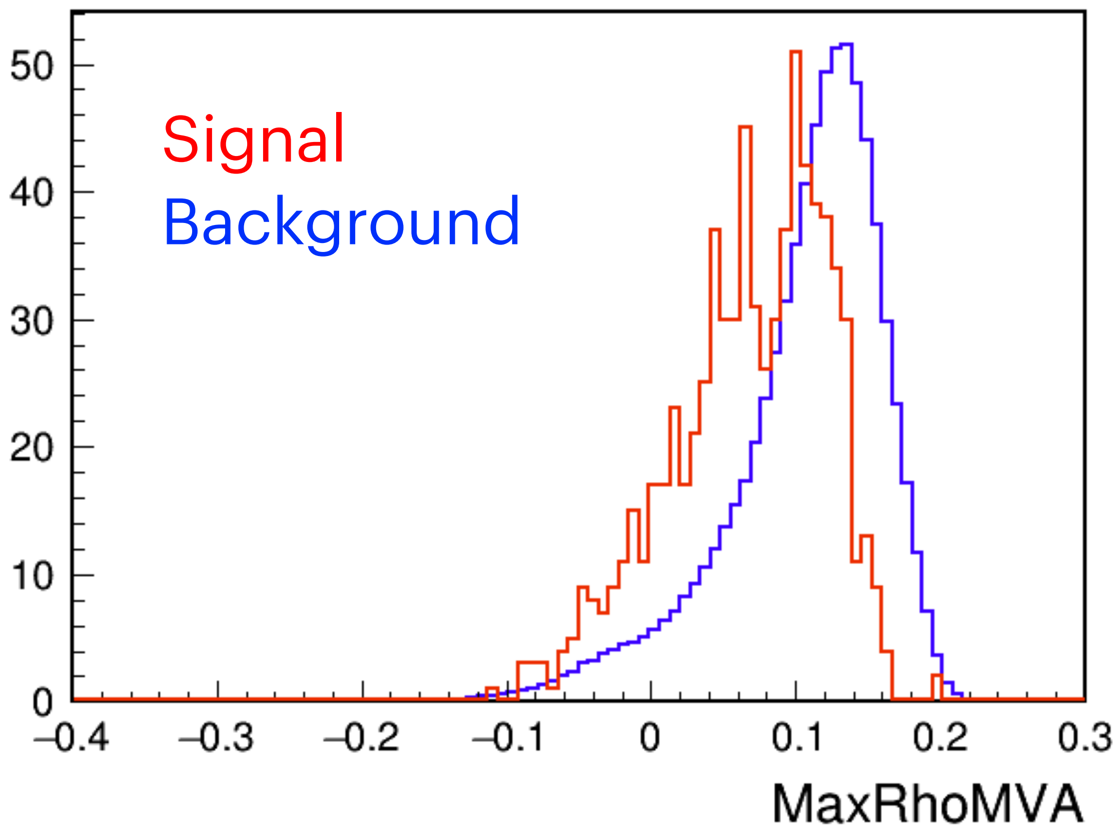
1 candidate



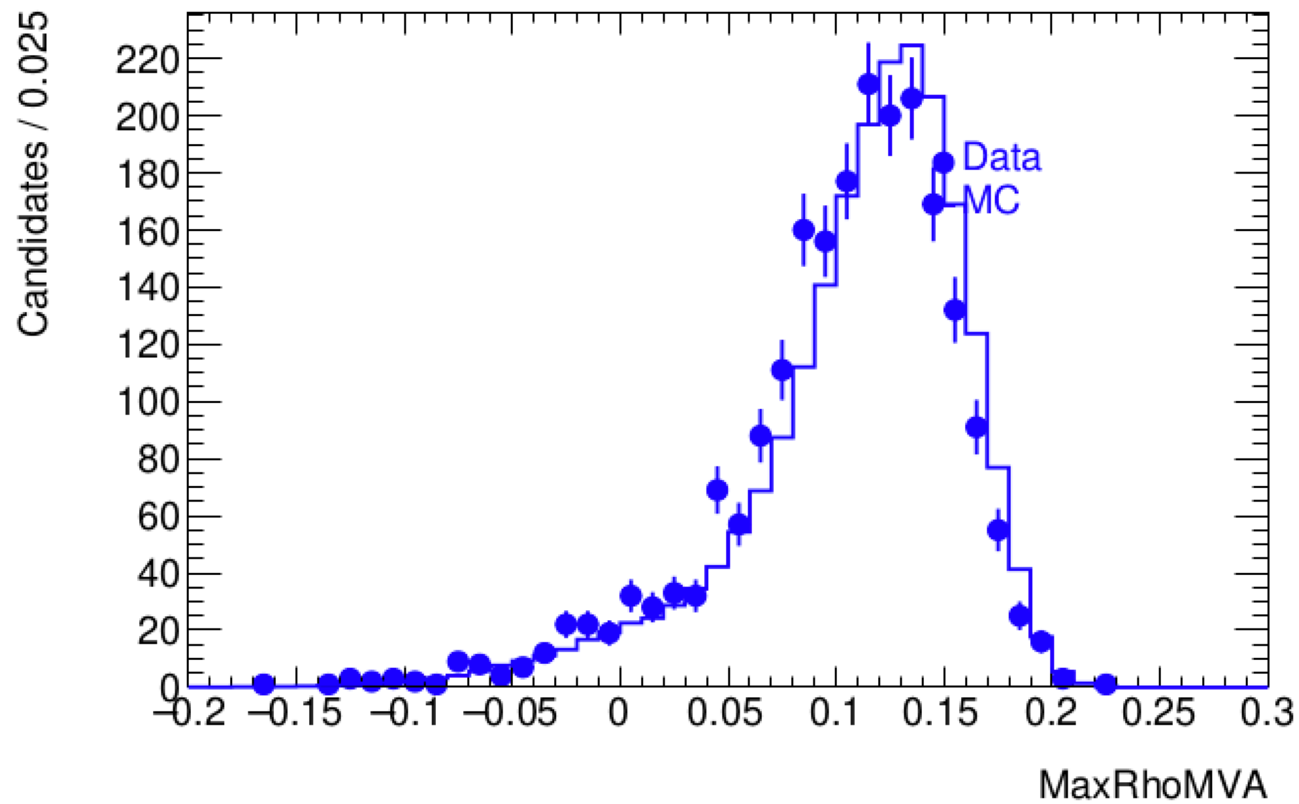
-> 6 rho sub-candidates

Max ρMVA distribution

Each candidate has for example 20 ρ sub-candidates. Take the one with largest rhoMVA (the one more similar to a ρ).



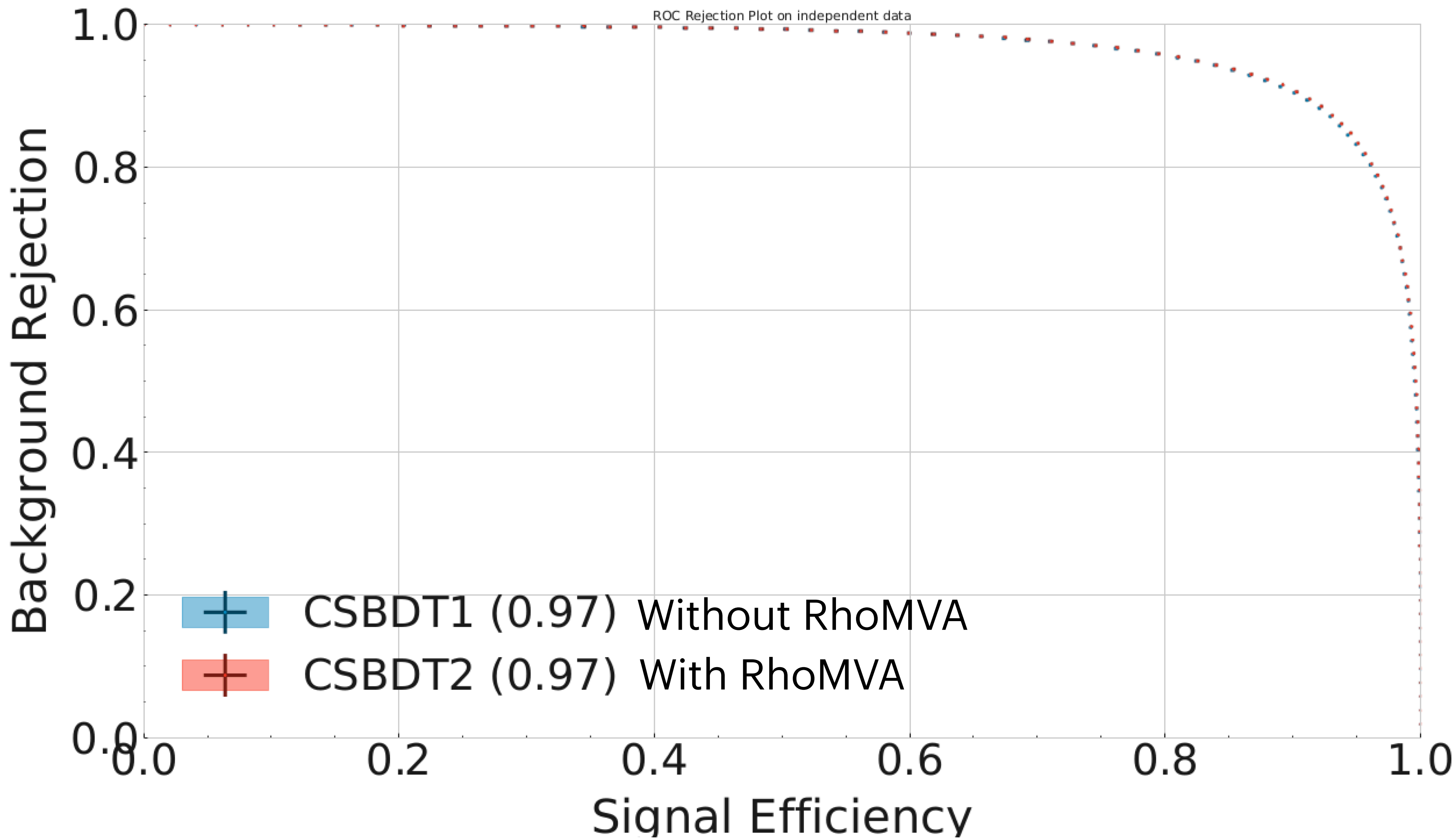
Validation: use $B^0 \to \pi^0\pi^0$ sideband (inclusive sample of true and false ρ).



Variable gives separation, and discrepancy is acceptable

Total candidates	Candidates with at least one rho	Candidates where the rho has been correctly identified
788473	285585	158393

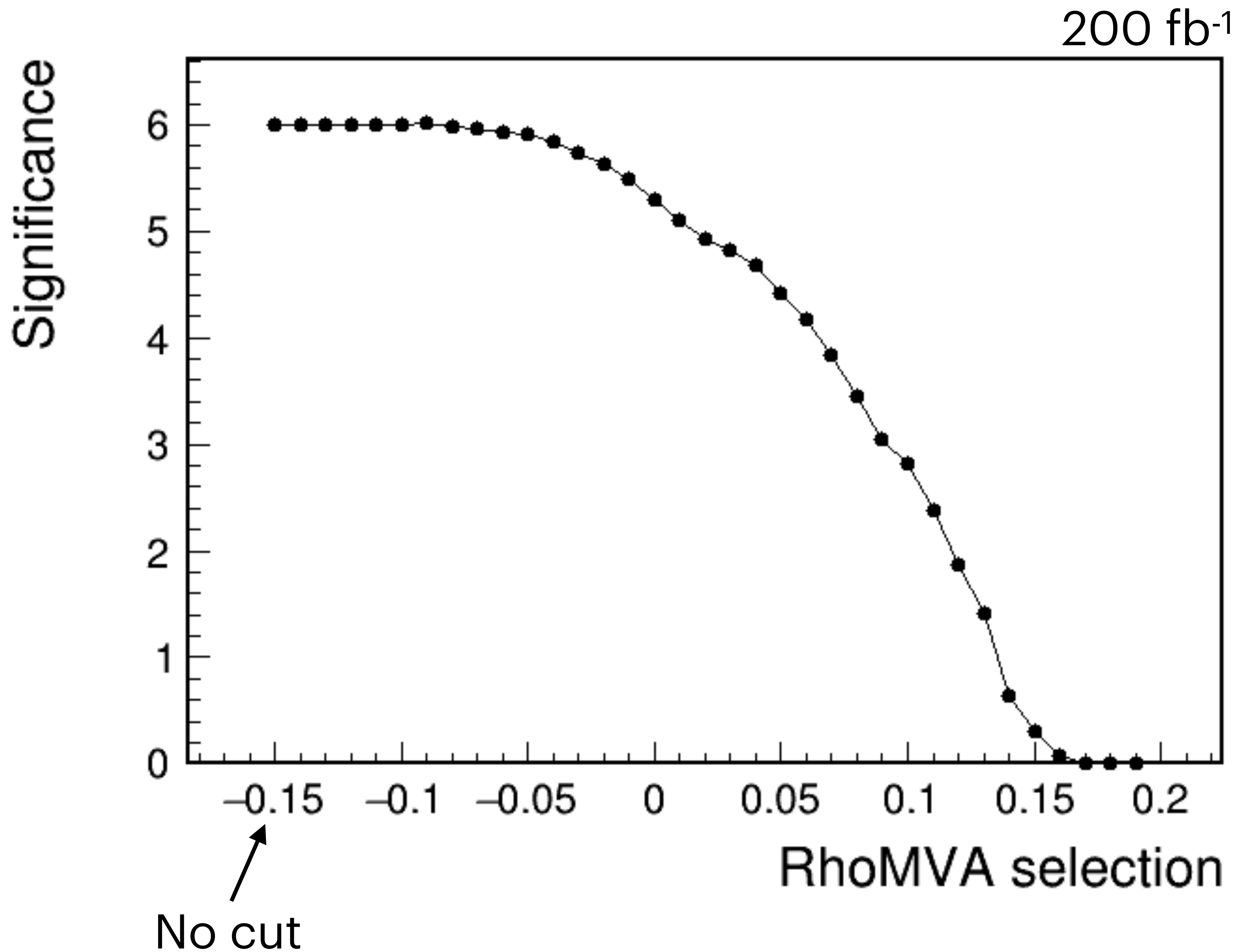
Use ρMVA as input of the CSBDT



Inclusion of ρMVA gives no improvement

Other possibility: ρMVA after the CSBDT

Apply first the selection on the CSBDT (>0.8), -0.2<ΔE<0.1 and Mbc>5.27, then various selections on ρMVA and calculate significance $S/\sqrt{S+B}$.



No gain in significance after selection on ρMVA.

Summary

Prepare $B^0 \to \pi^0\pi^0$ analysis for pre-LS1 dataset.

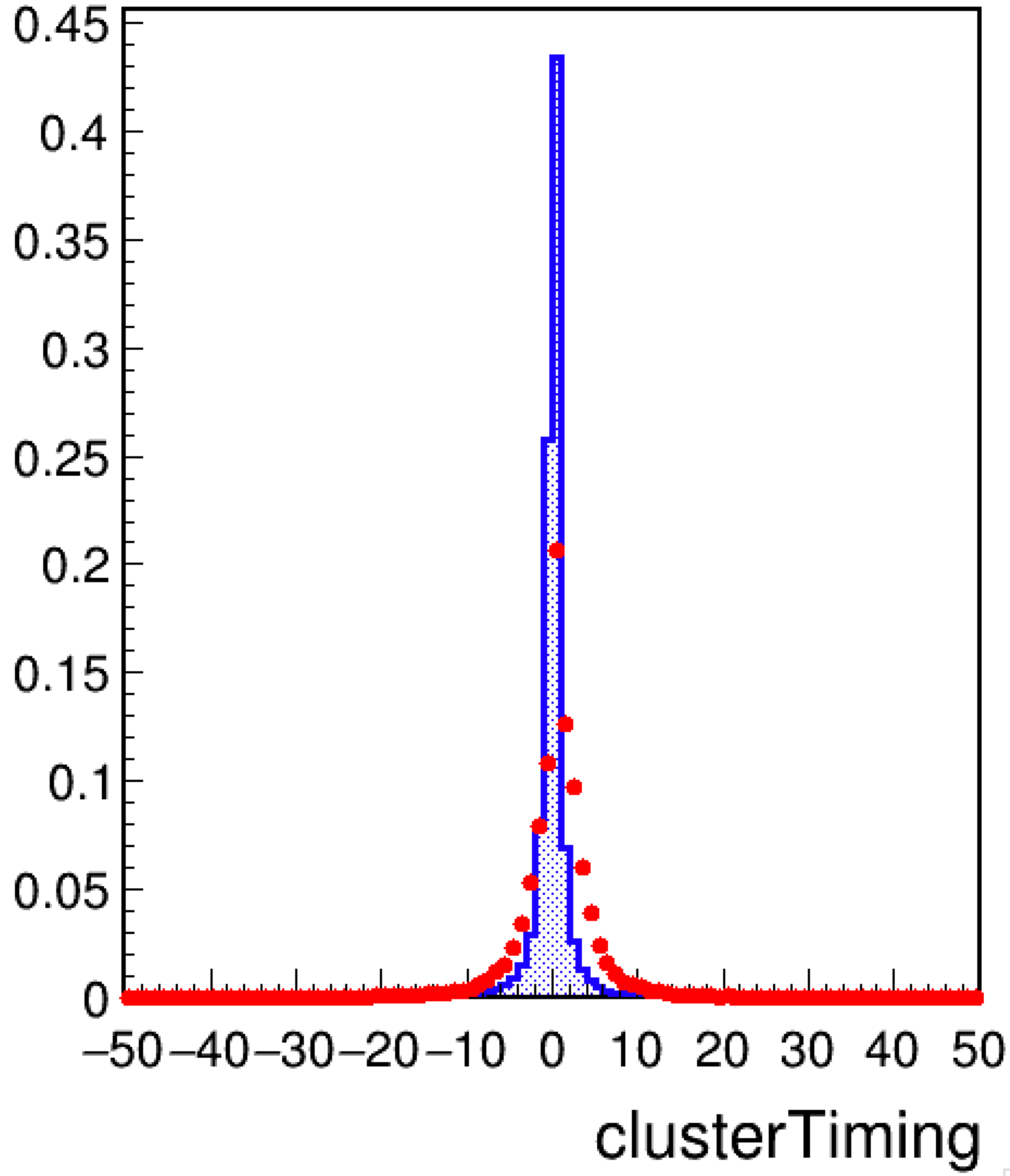
Revisited photonMVA: use new variables with good data/MC agreement.
Already validated on data.

Revisited CSBDT: add B_{Tag} variables to suppress even more continuum.
Variables are ready, but need to repeat training using off-res data (is it enough?). Check how the use of B_{Tag} variables impacts the flavour tagger.

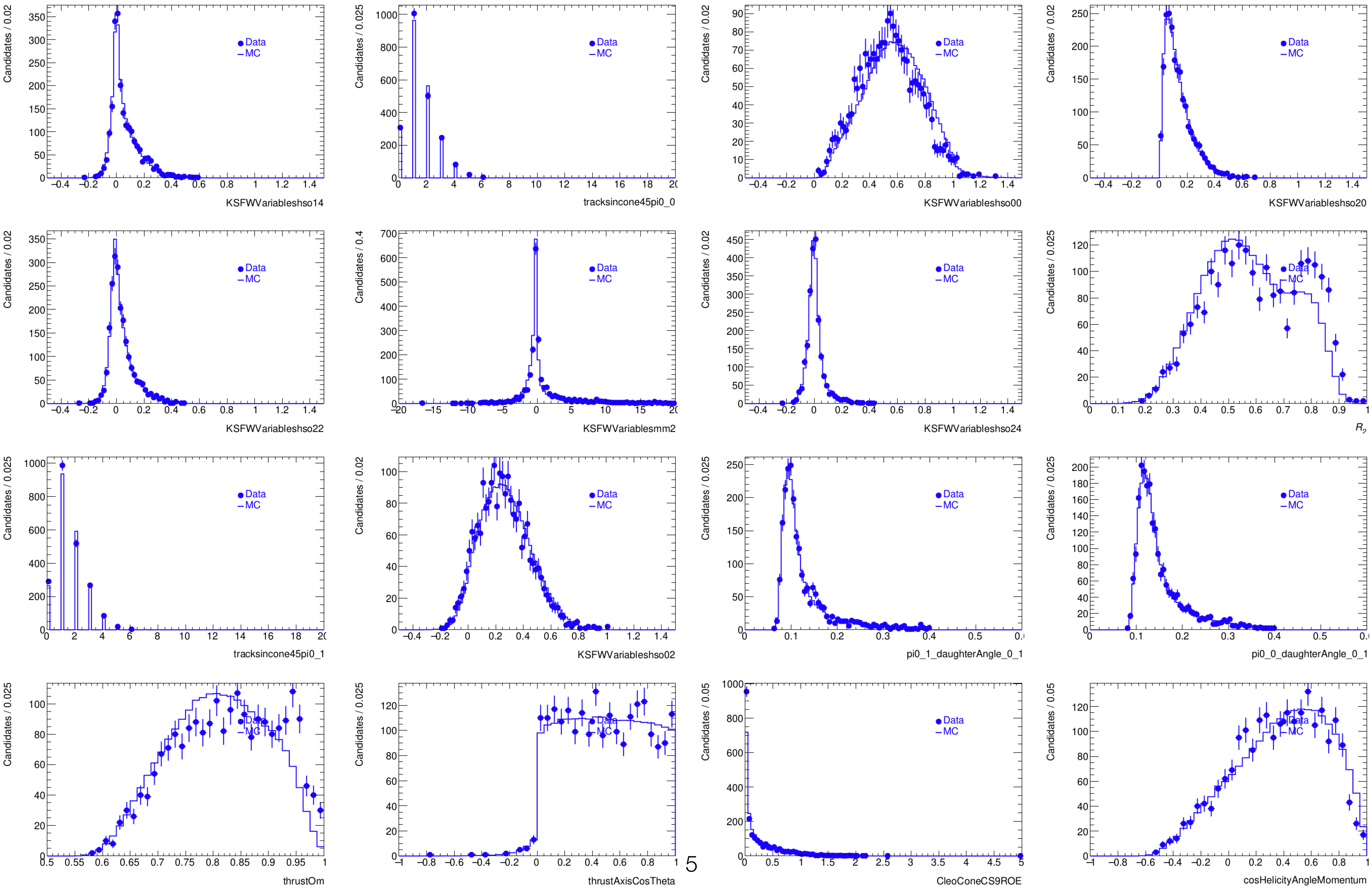
Introduced ρBDT: improvement is negligible, maybe not useful to add it in the analysis.

Backup

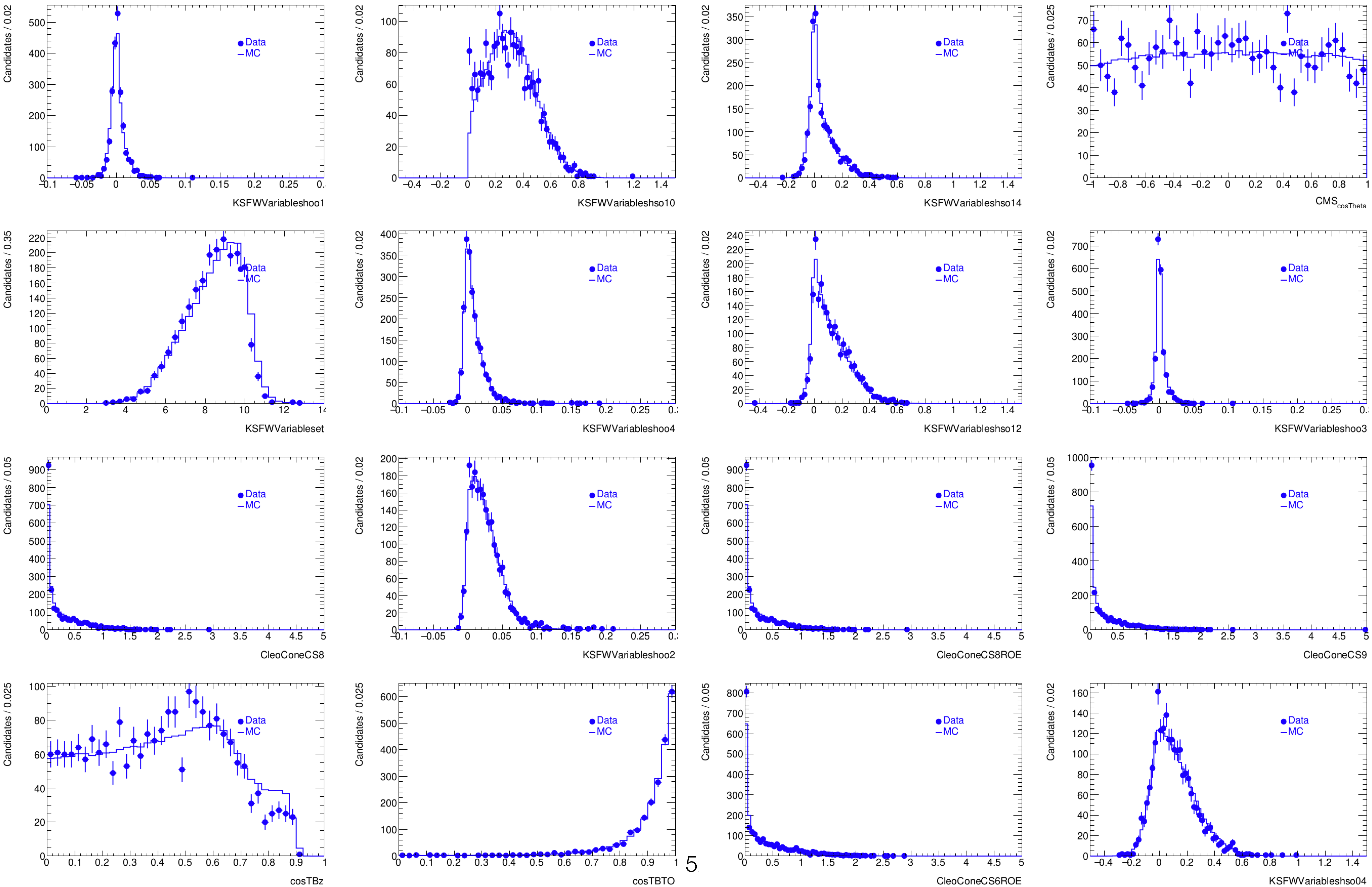
ClusterTiming (rel-06)



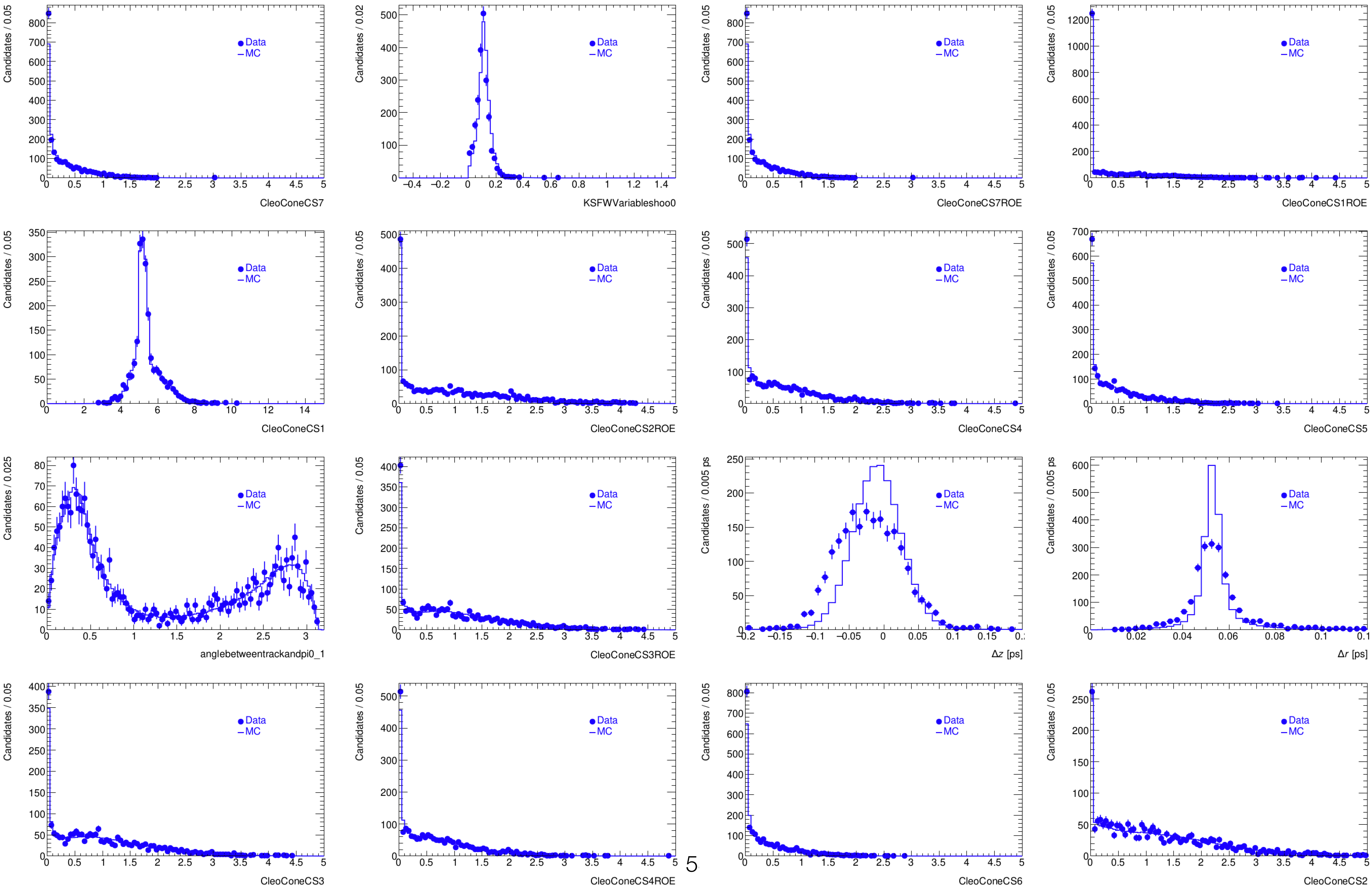
Inputs validation — Background only



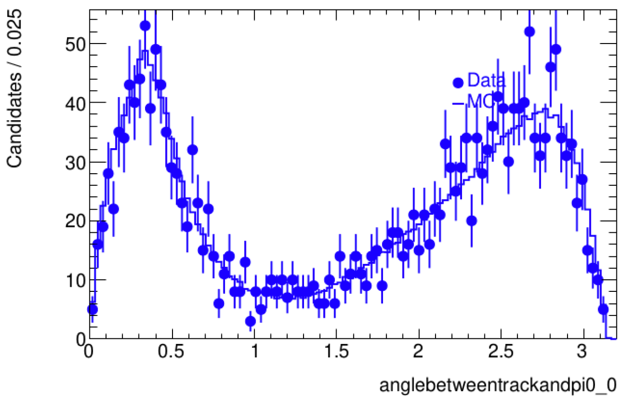
Inputs validation — Background only



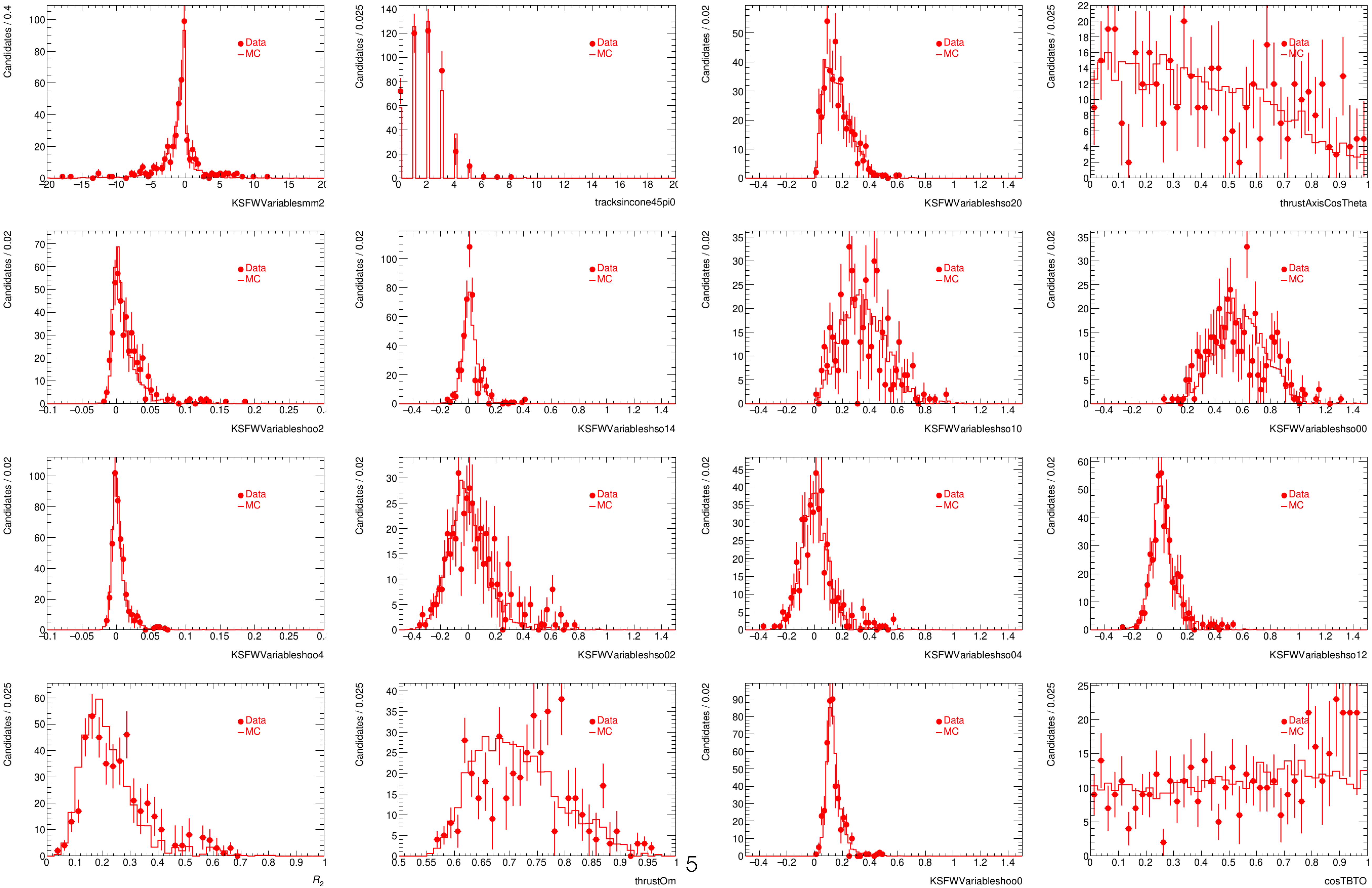
Inputs validation — Background only



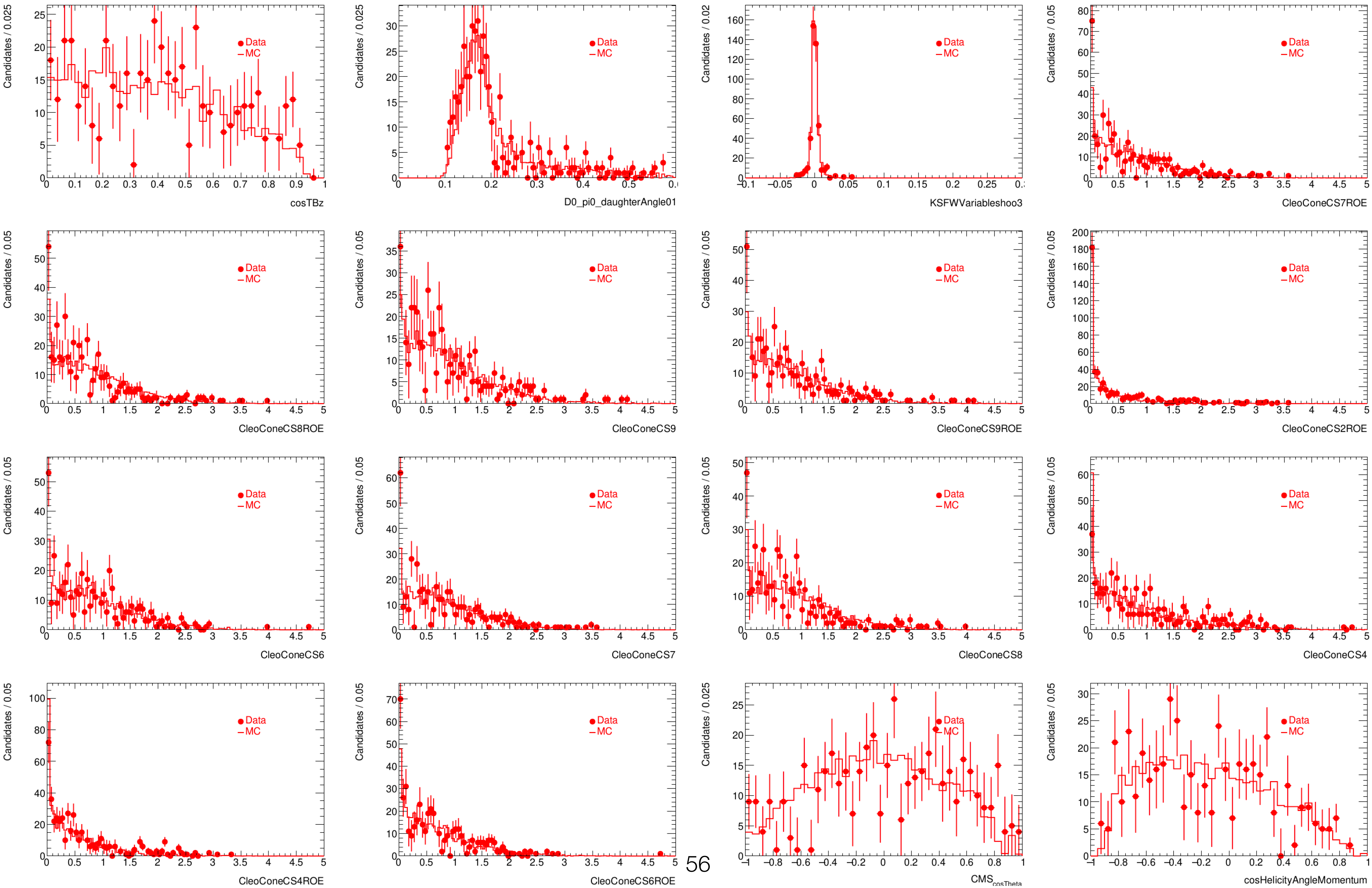
Inputs validation — Background only



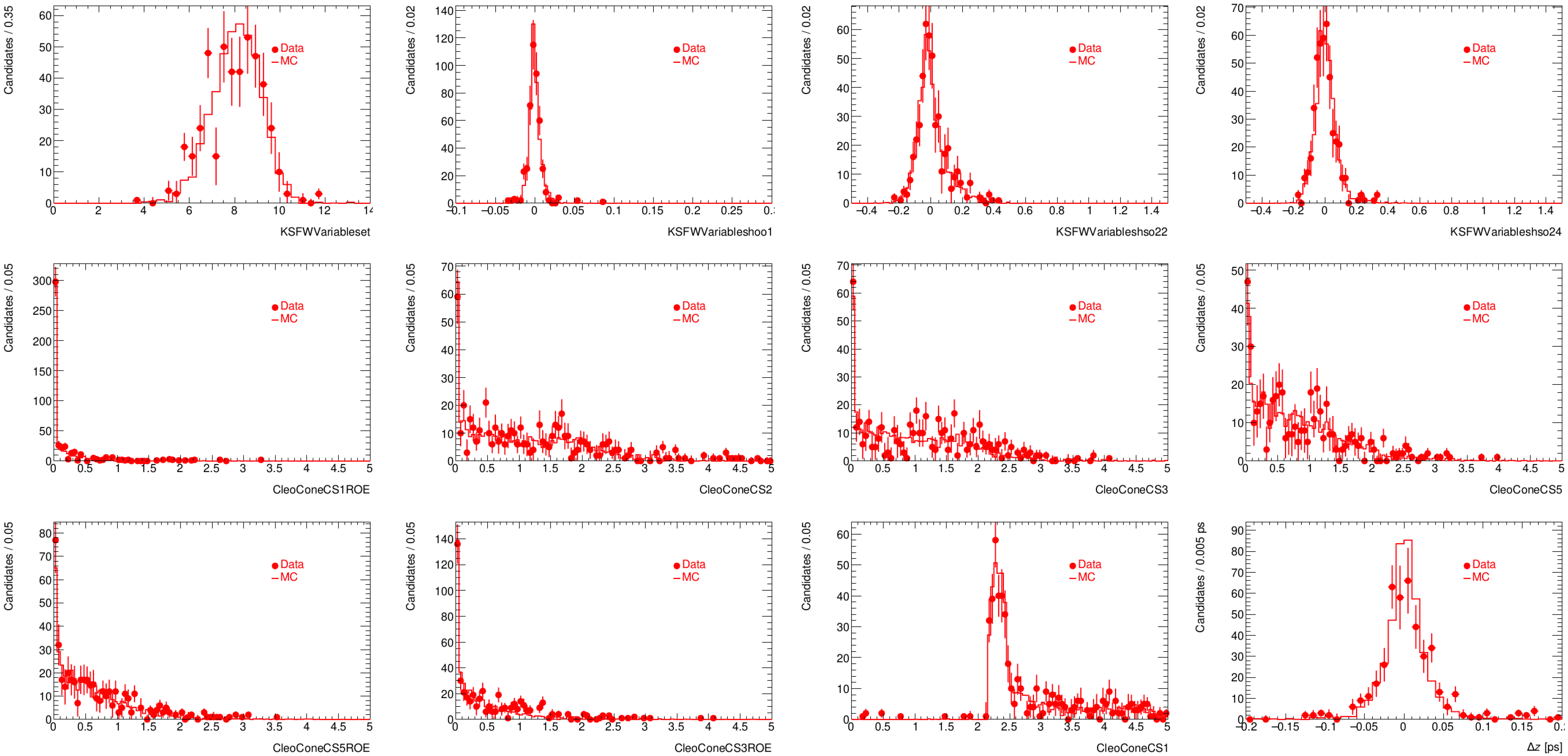
Inputs validation — Signal only

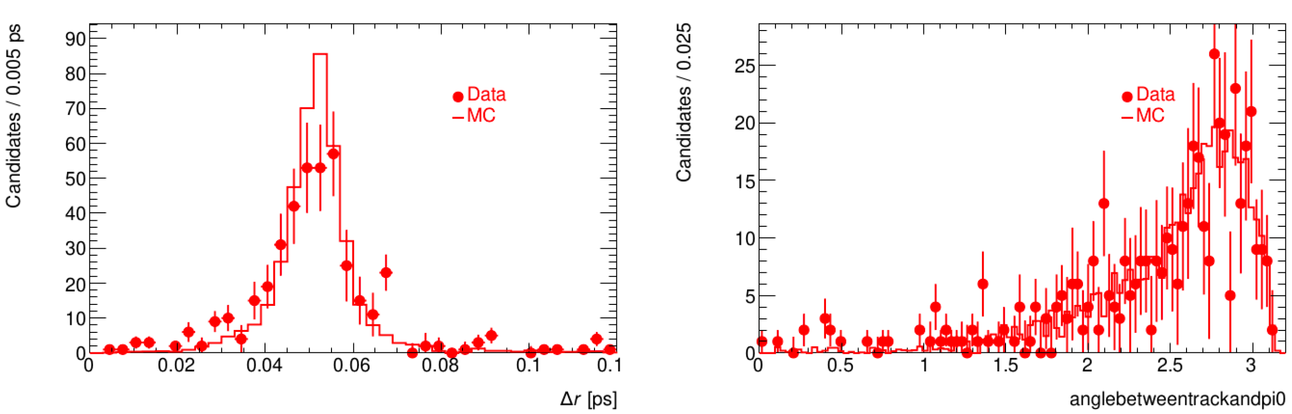


Inputs validation — Signal only



Inputs validation — Signal only

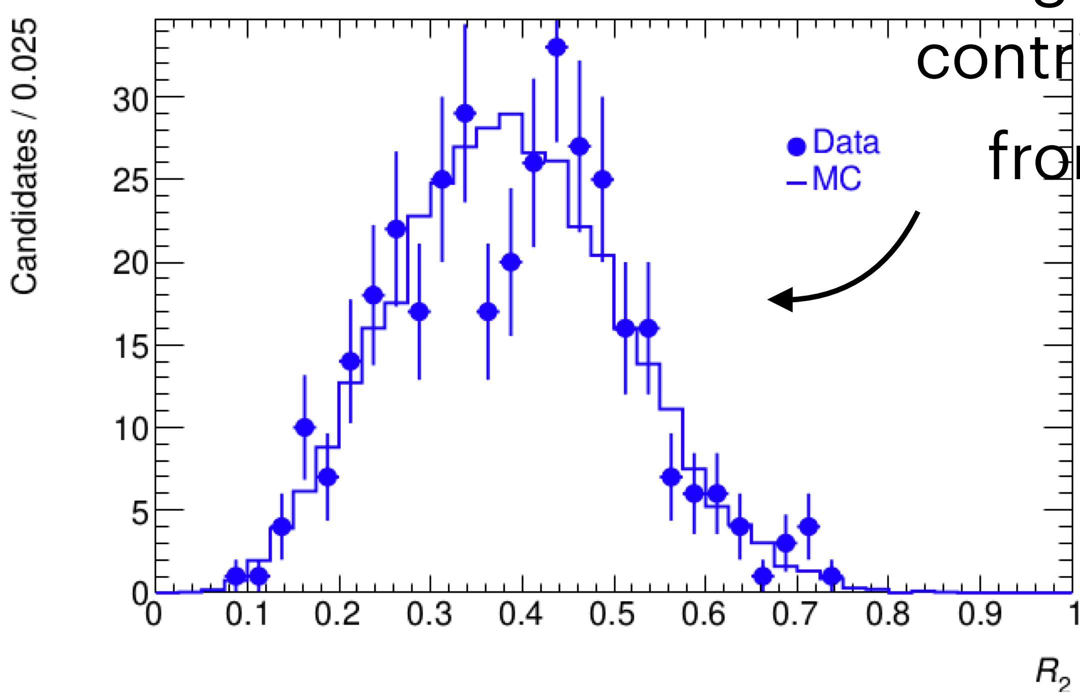


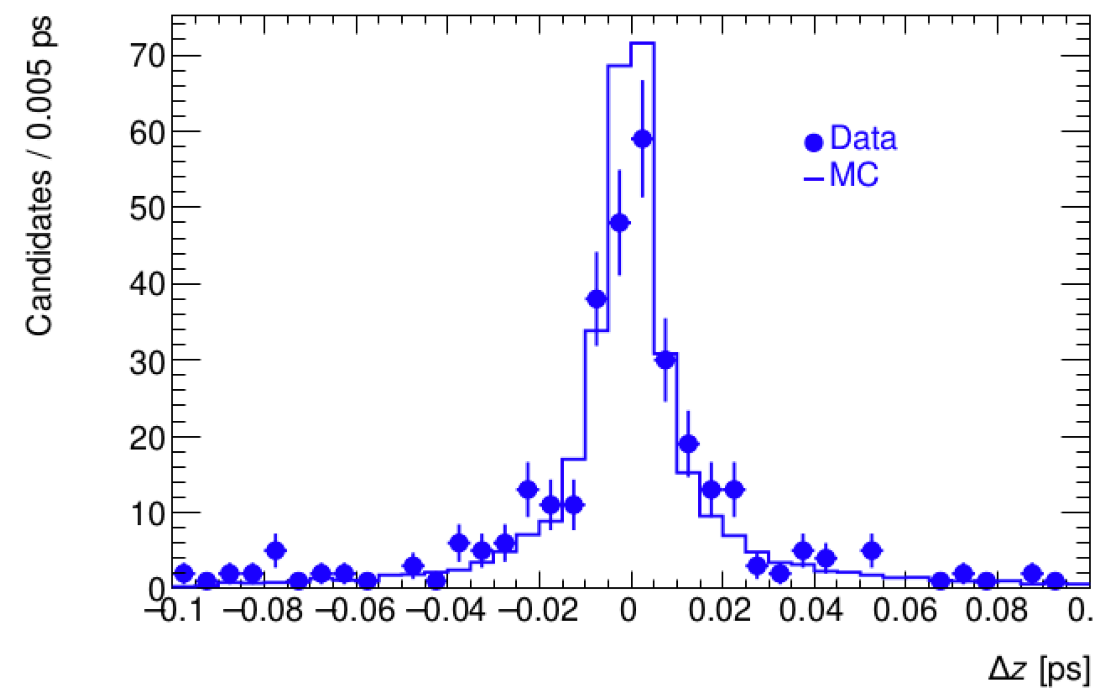


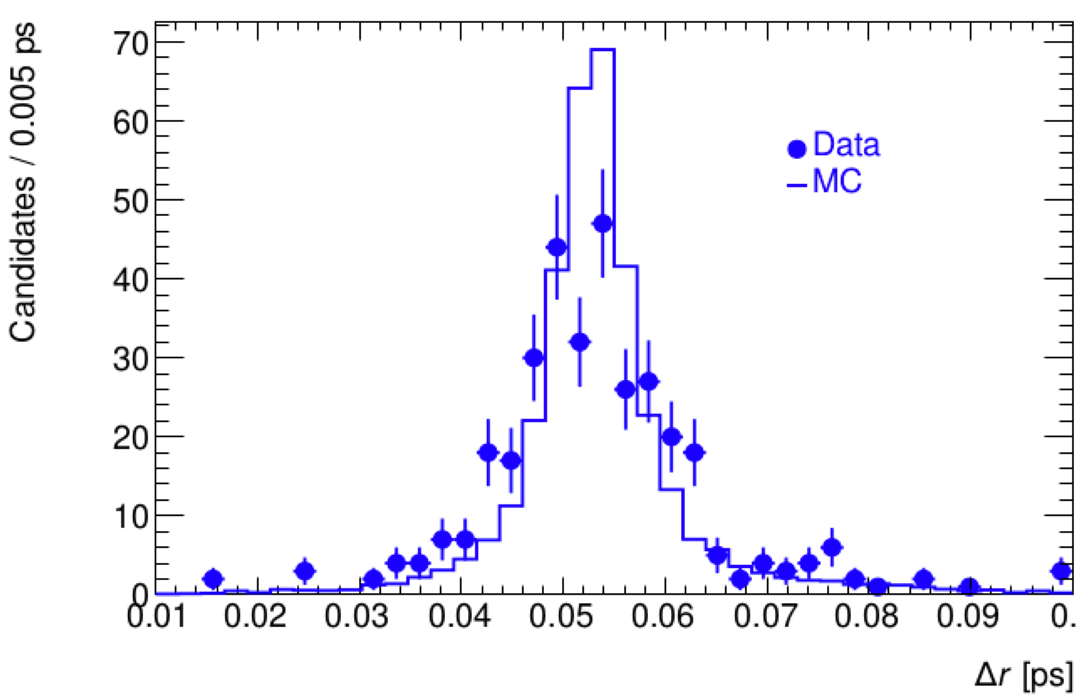
Need more statistics, but no large discrepancy observed

Check — Background only using $B \to D(K\pi\pi^0)\pi$ sideband

Negligible contribution from $\tau\tau$





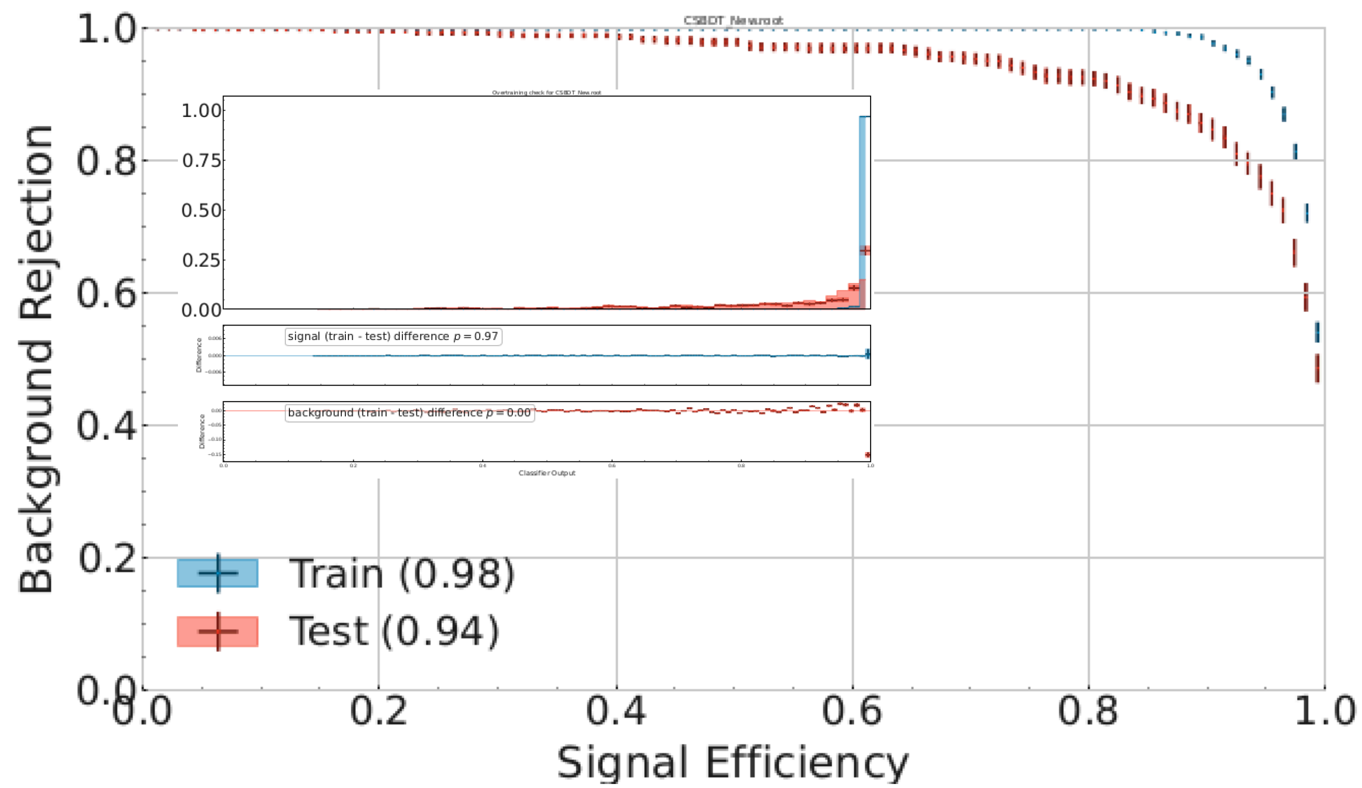


Need more statistics — but observe smaller discrepancies in Δr and ΔZ wrt $B^0 \to \pi^0\pi^0$. Why?

CSMVA using off-res data for the bkg

Train on off-res data and signalMC after applying all π^0, ΔE, and M_{bc} selections.

Inputs (after pruning)
7 Kakuno-Super-Fox-Wolfram moments
cosTBTO
1 CleoCone
cosTheta*
R2
thrustOm
ΔZ (BTag)
Δr (BTag)
thrustAxisCosTheta
angle between π^0's
cosHelicityAngle
KSFWVariableset
KSFWVariablesmm2



Train bkg sample (from offres): 1000 events
Train sig sample (from MC): 180000 events
Test bkg sample (from offres): 500 events
Test sig sample (from MC): 90000 events

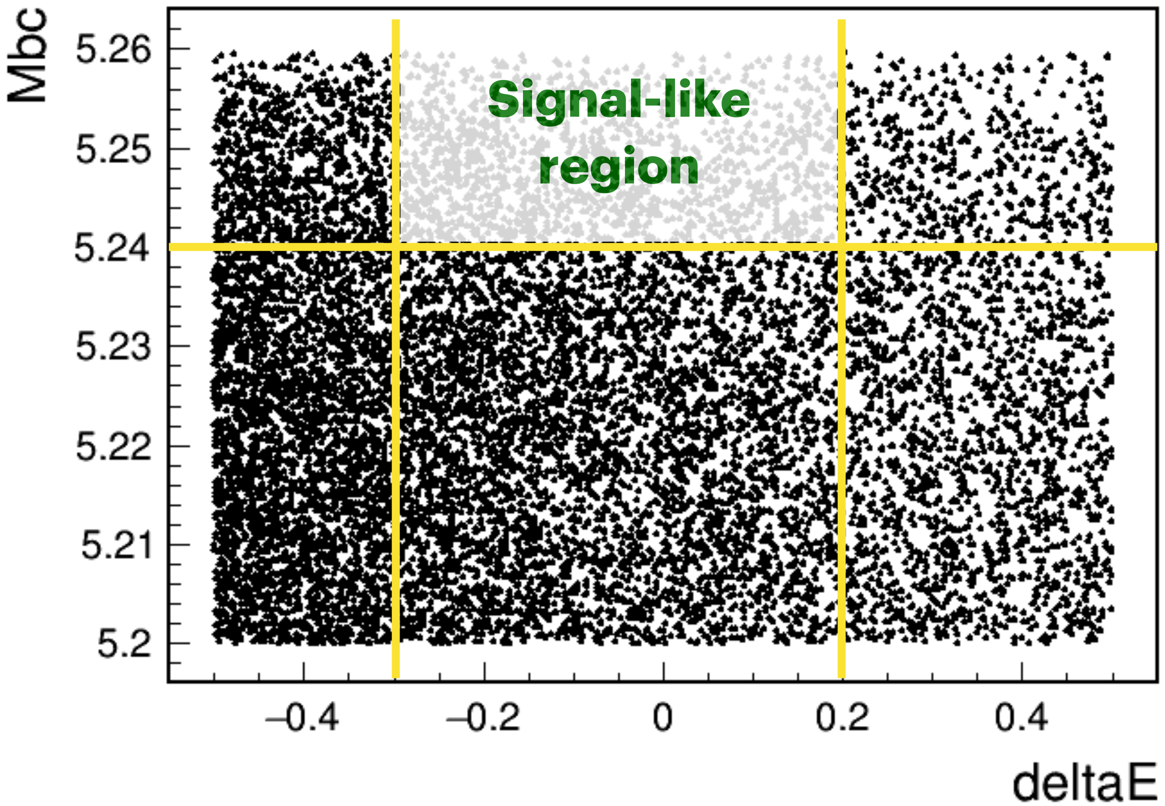
Very small off-res sample ($9 fb^{-1}$)
→ poor BDT (total off-res sample will be $18 fb^{-1}$)

CSMVA using off-res data for the bkg

But what off-resonance data can I use?

In previous result I was using only the signal-like region

Off-resonance data:



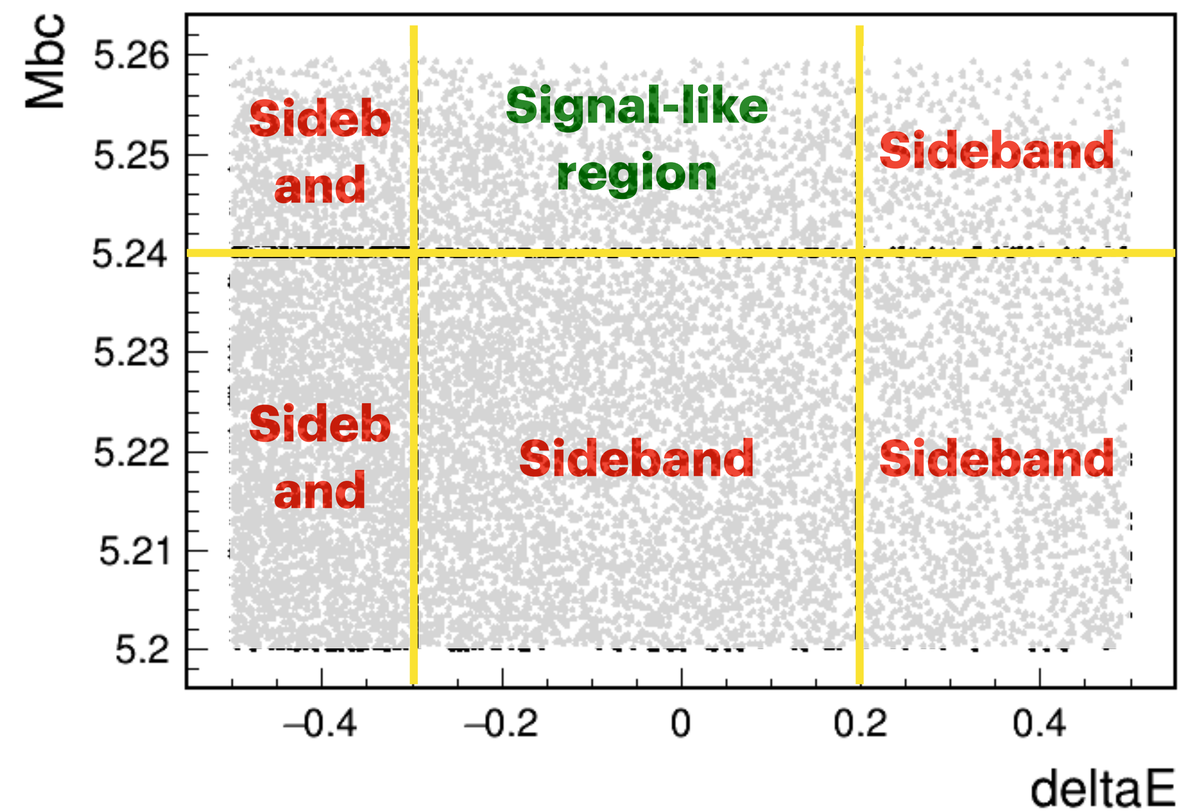
CSMVA using off-res data for the bkg

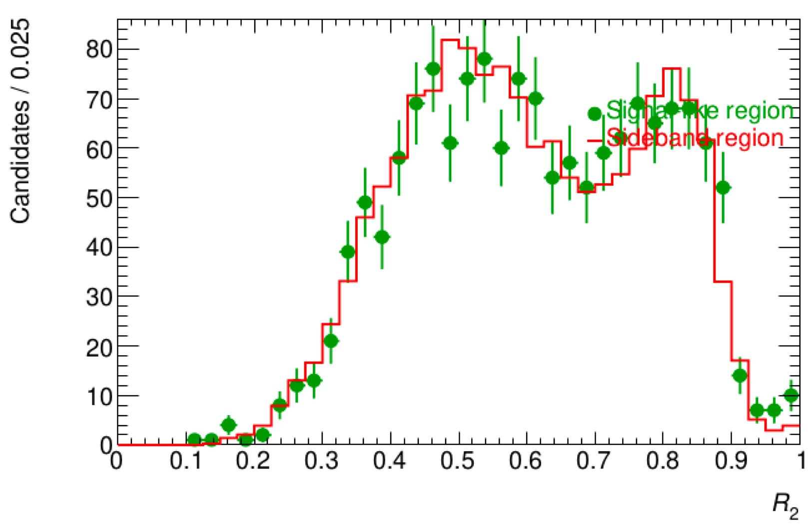
But what off-resonance data can I use?

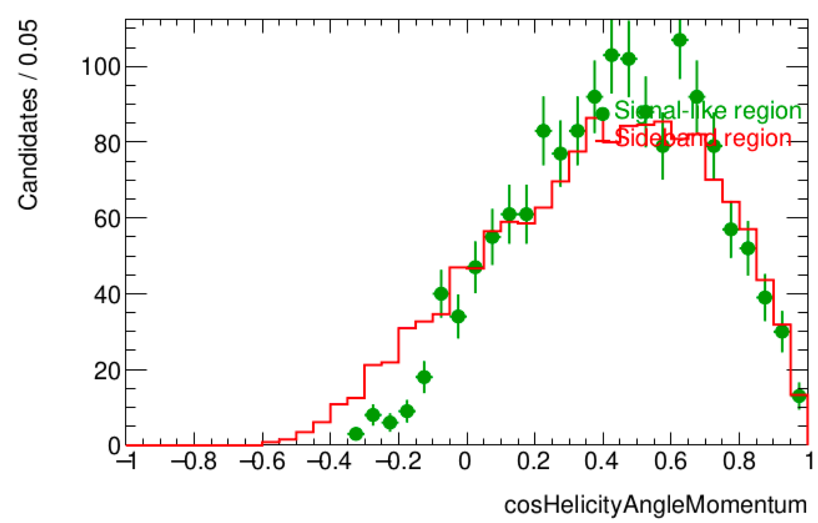
In previous result I was using only the signal-like region

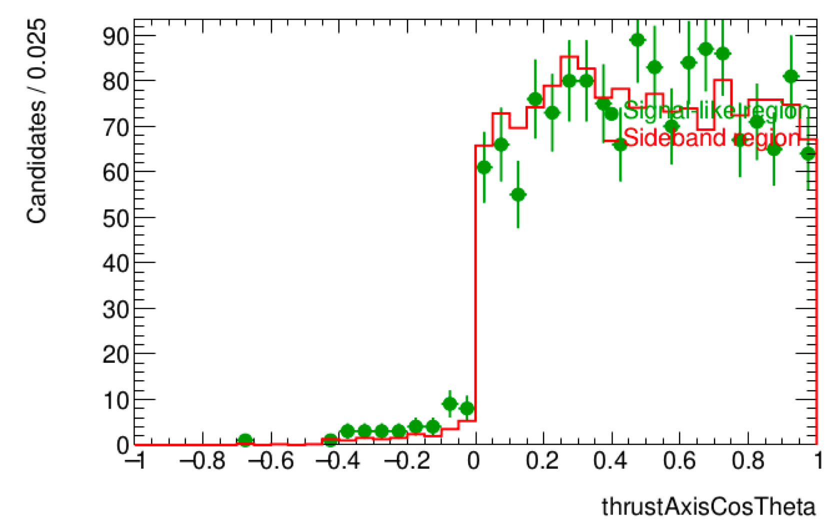
Compare CS input distributions in signal-like and sideband regions (in off-resonance data):

Off-resonance data:







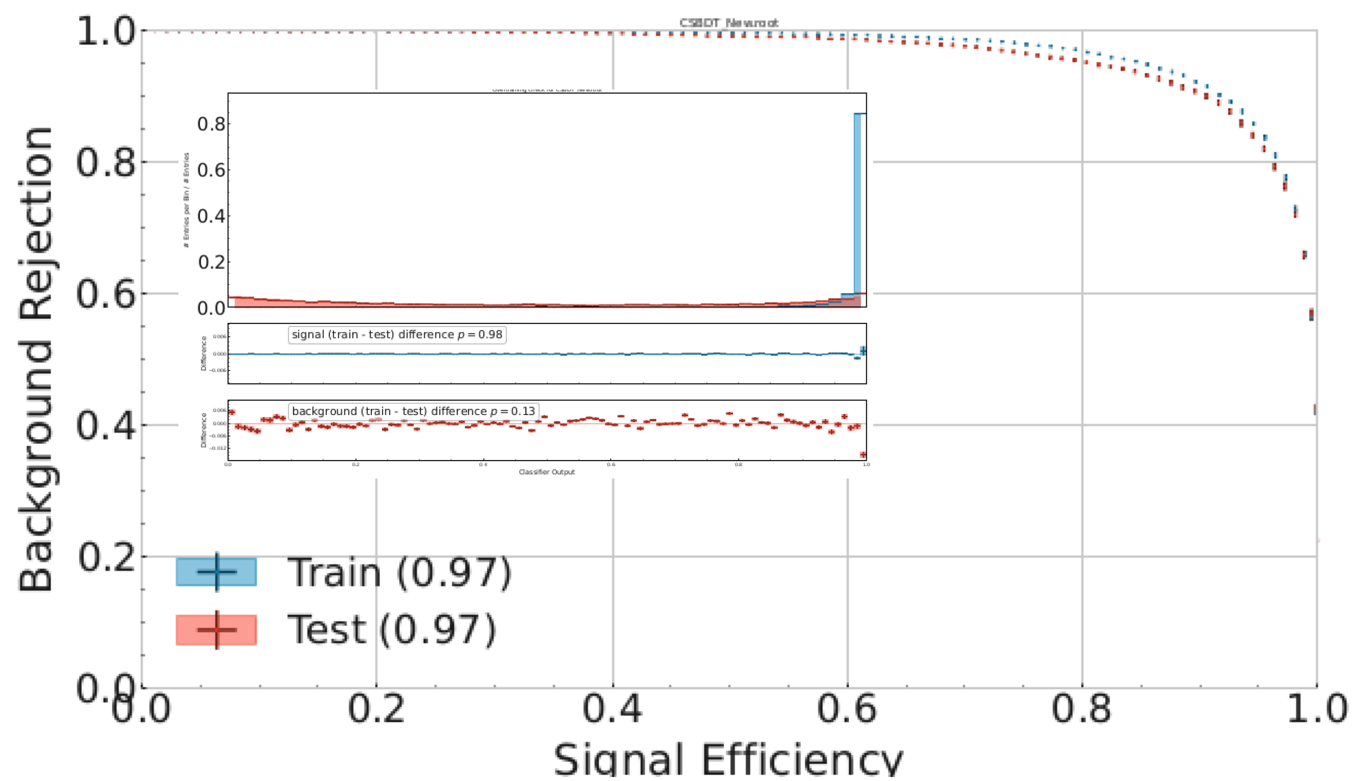


All variables that have discrepancies are not CS inputs anymore (after pruning), except cosHelAngleMomentum.

CSMVA using off-res data for the bkg

Train on off-res data and signalMC after applying all π^0 selections.
Use all off-resonance data (including sidebands). Exclude cosHelicityAngle.

Inputs (after pruning)
7 Kakuno-Super-Fox-Wolfram moments
cosTBTO
1 CleoCone
cosTheta*
R2
thrustOm
ΔZ (BTag)
Δr (BTag)
thrustAxisCosTheta
angle between π^0's
~~cosHelicityAngle~~
KSFWVariableset
KSFWVariablesmm2

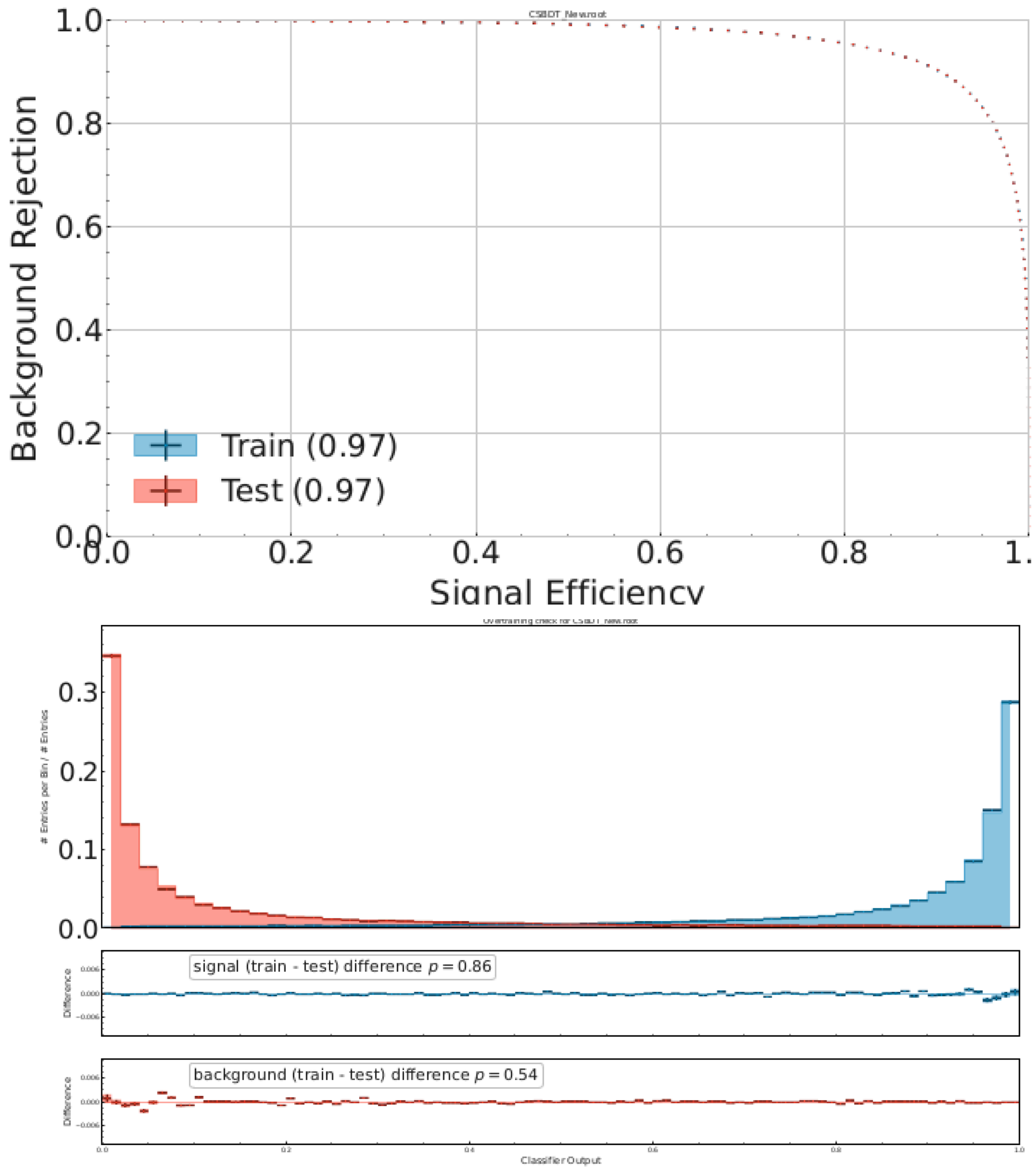


Train bkg sample (from offres): 8000 events
Train sig sample (from MC): 180000 events
Test bkg sample (from offres): 4000 events
Test sig sample (from MC): 90000 events

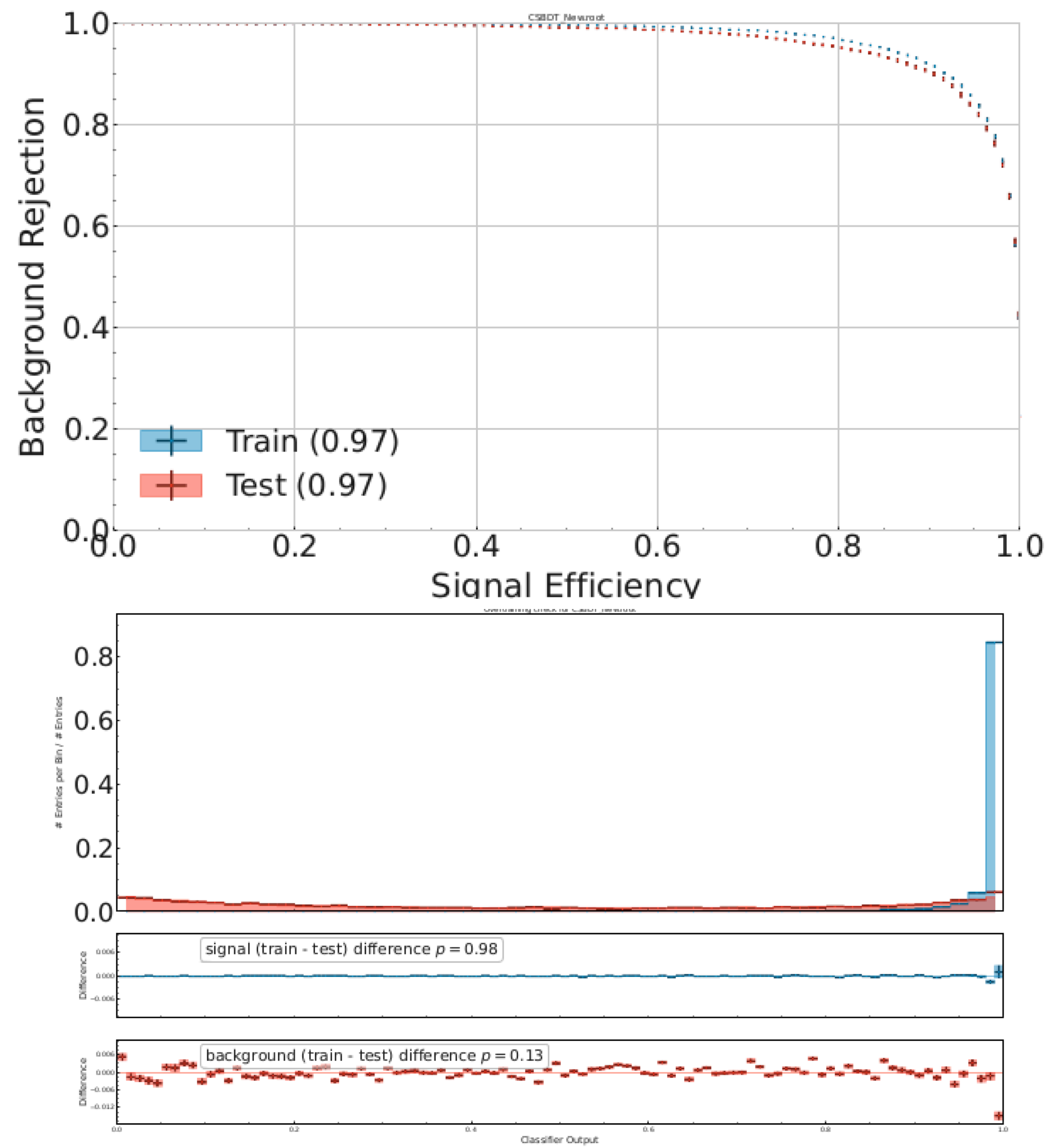
Better result wrt previous one

CSMVAs comparison

All MC



Off-res data + signalMC



ROC curves are the same, but distributions are quite different

Sideband in off-res data is fine.

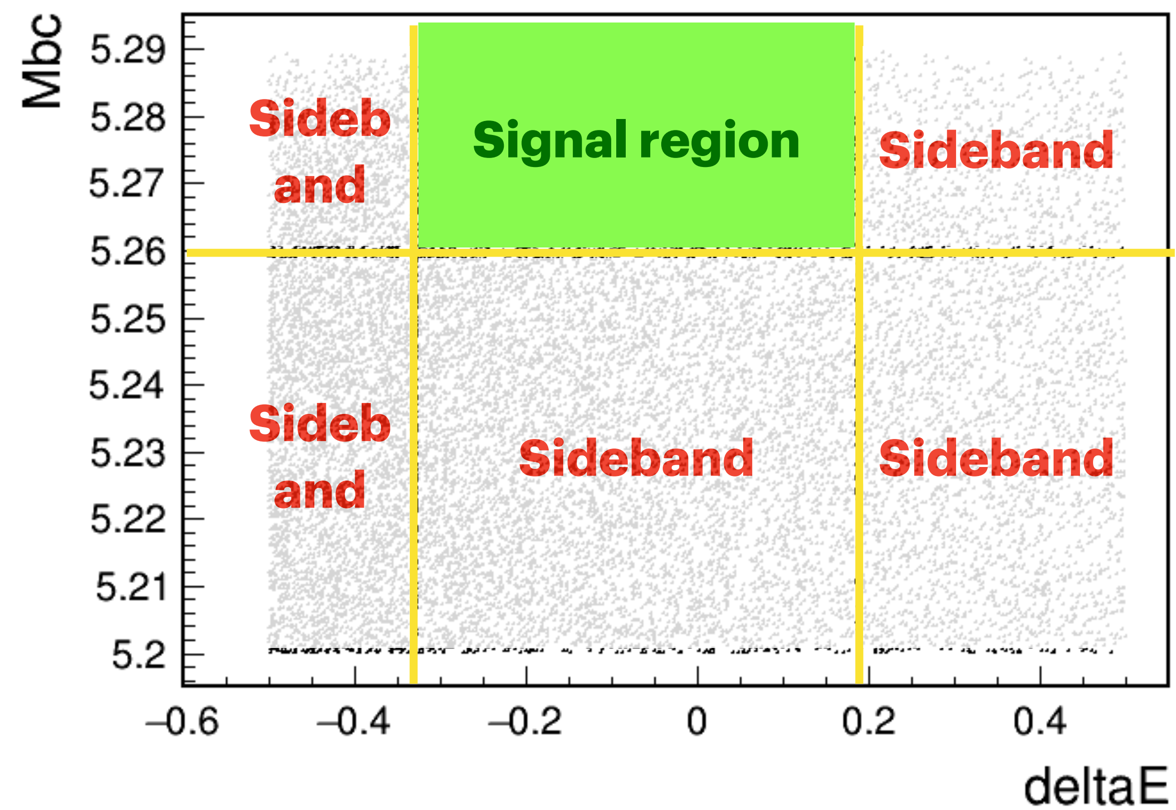
But then what about the
on-resonance sideband data?

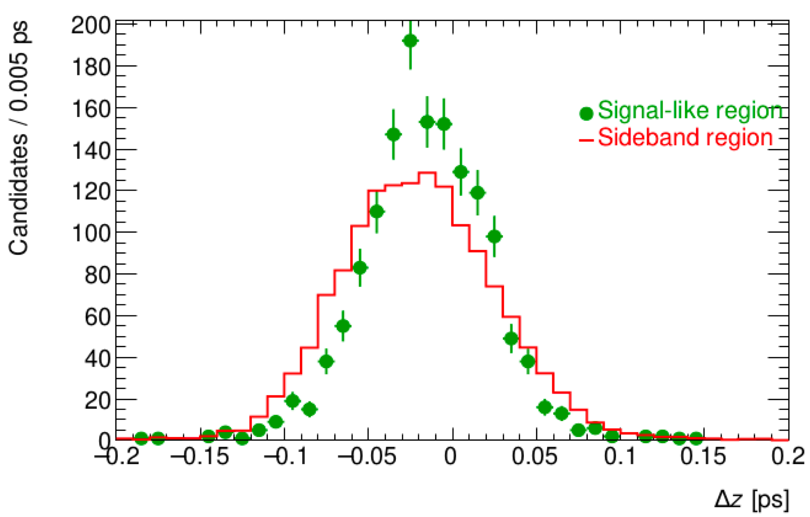
Sideband data vs off-res data

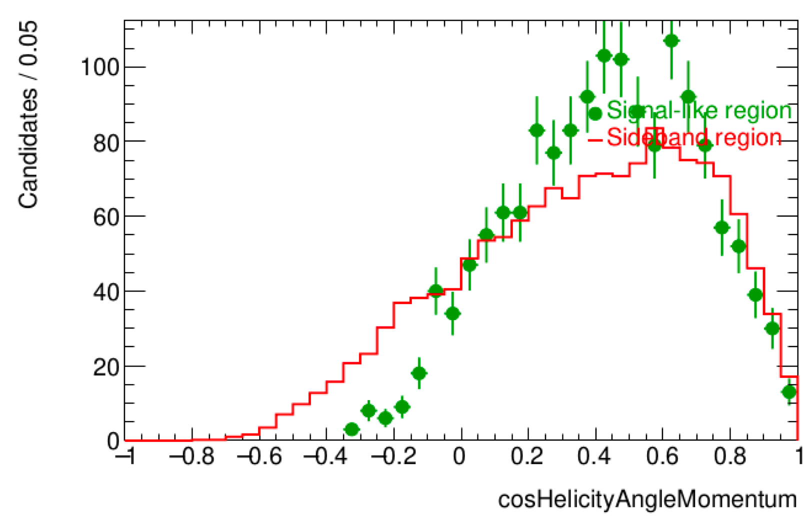
Sideband distributions seem good in off-res.

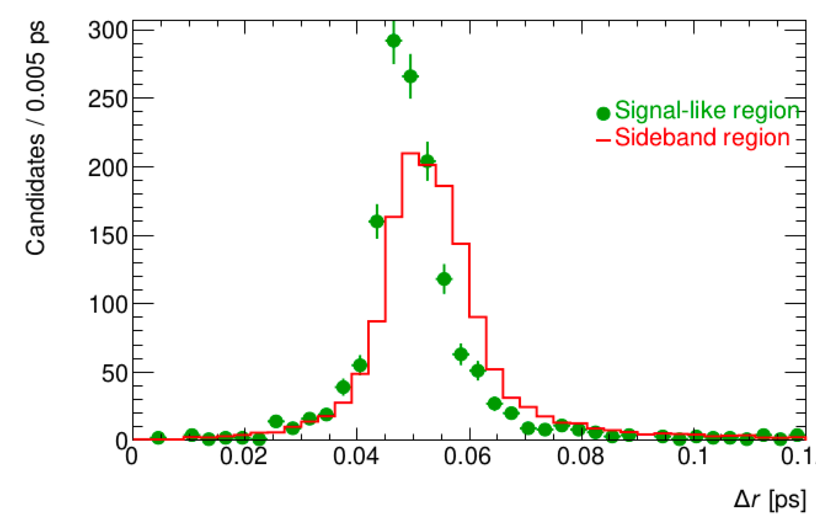
Compare CS inputs distributions in signal-like region (in off-resonance data) and sideband regions (in on-resonance data):

Data:





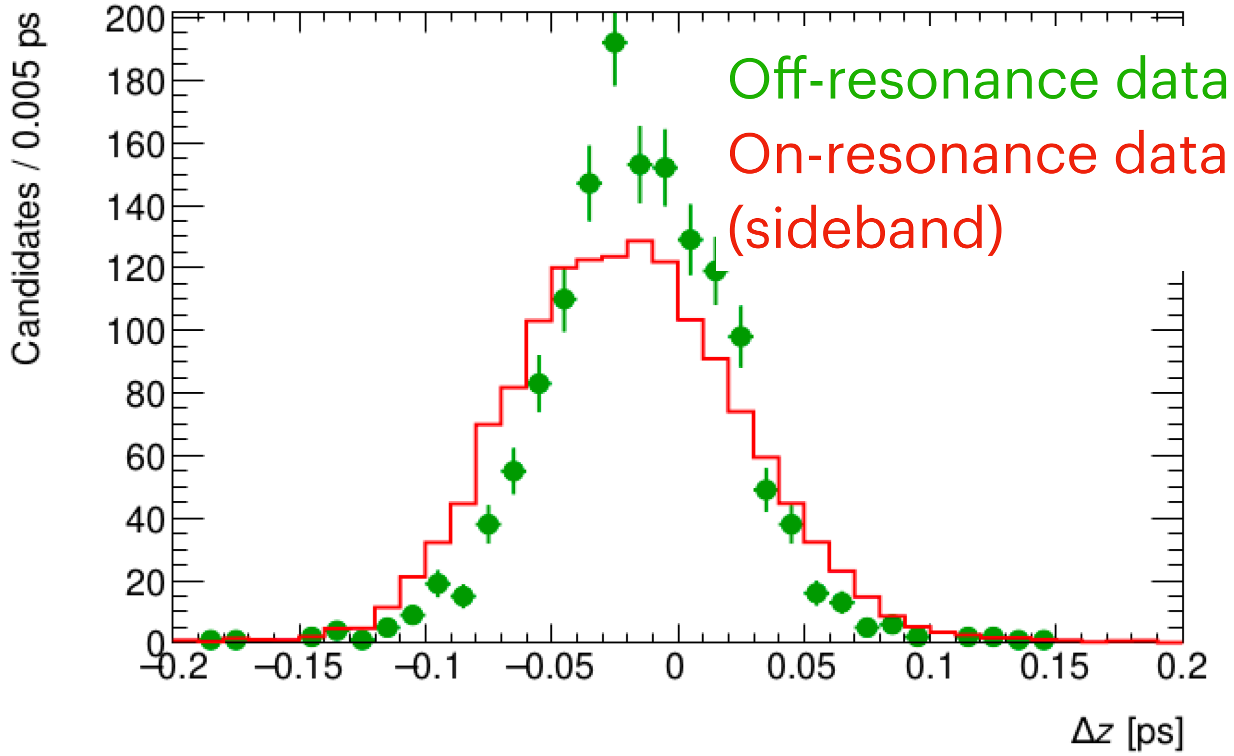


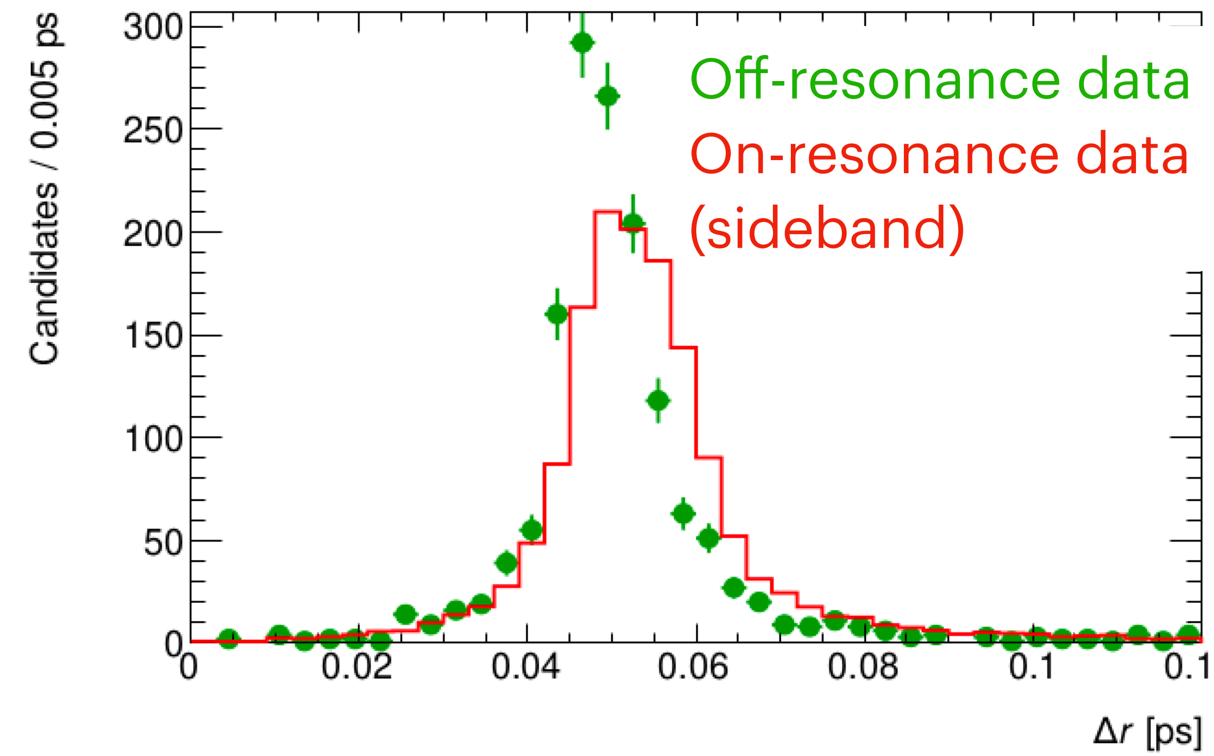


Observe large discrepancies also in Δr and ΔZ.

Which are the correct Δr and ΔZ distributions?

Sideband data and off-resonance data have different r and ΔZ distributions.





I need to understand which one reproduces correctly the signal region:

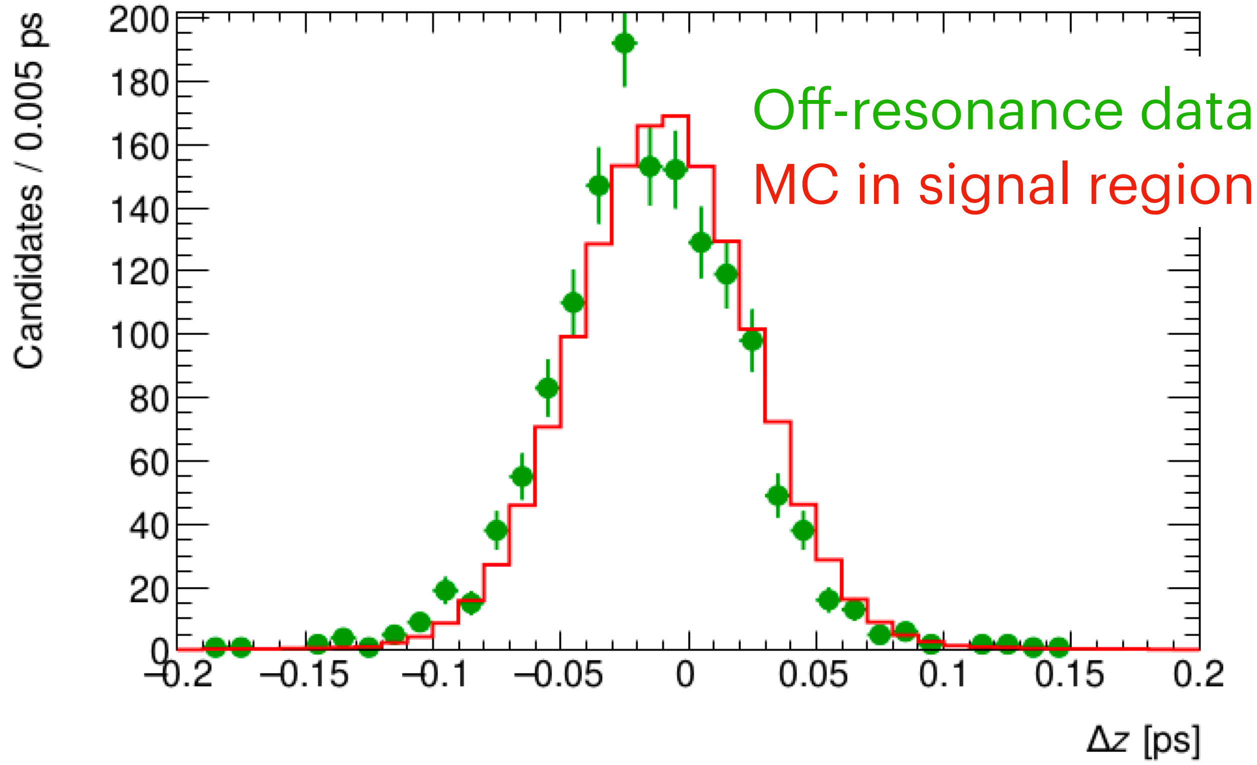
Red is correct → use sideband data (and exclude cosHelAngle from the inputs)

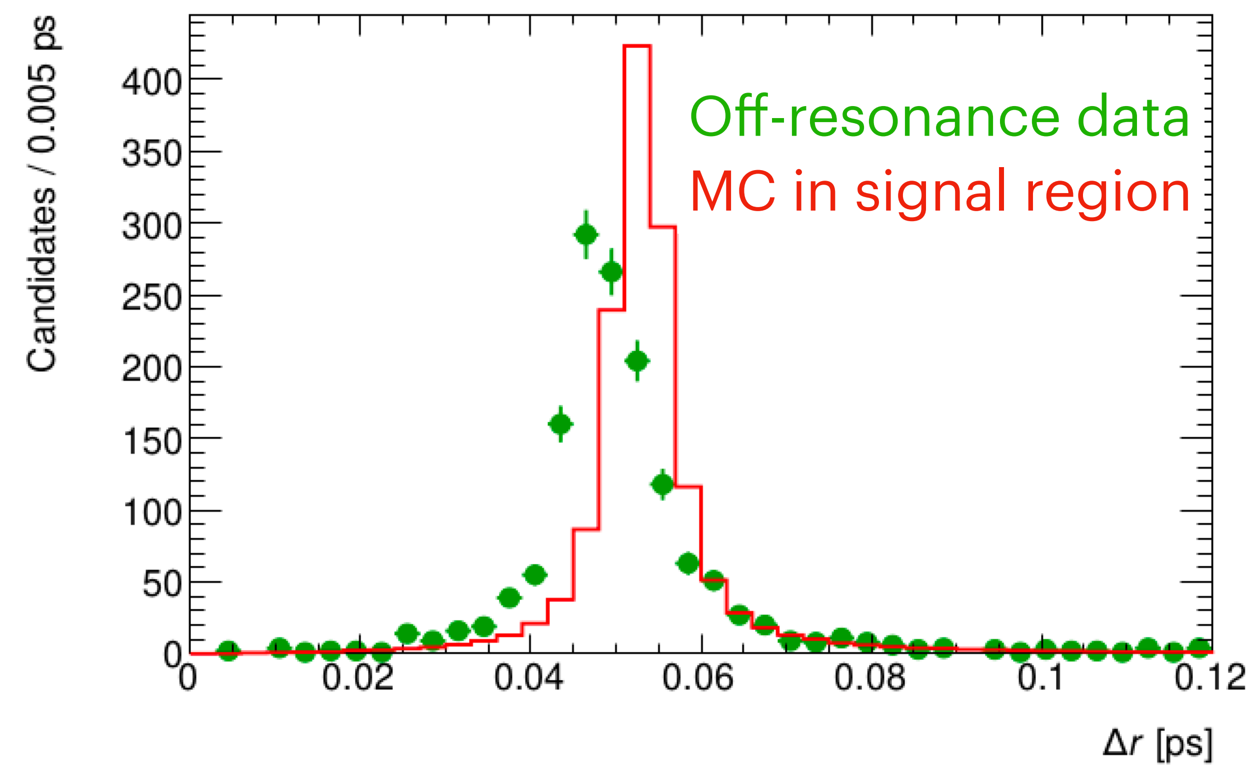
Green is correct → use off-res data (and exclude cosHelAngle from the inputs)

Stiil thinking how to do this.

Which are the correct Δr and ΔZ distributions?

Sideband data and off-resonance data have different r and ΔZ distributions.



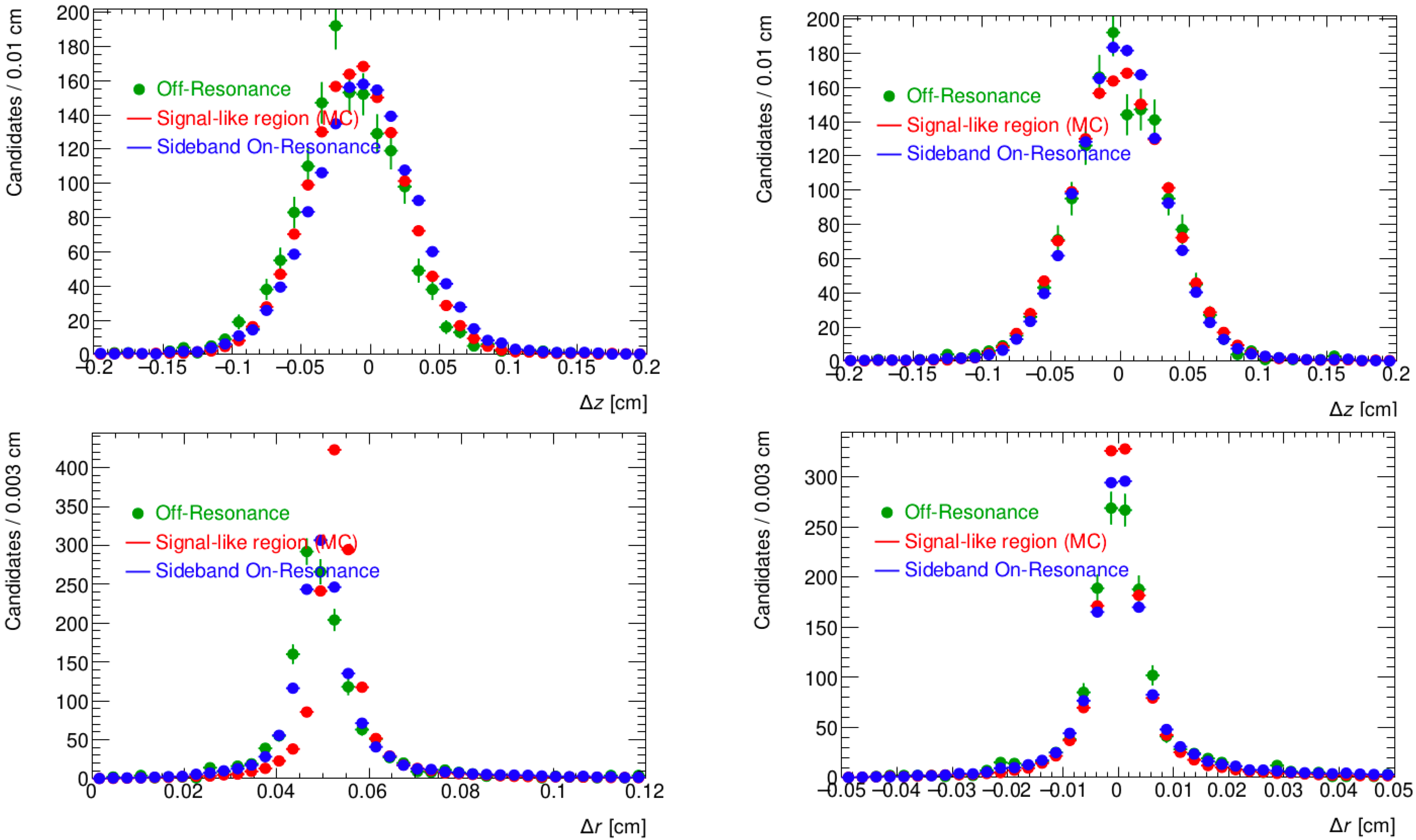


Metti tutti e tre insieme dai

Which are the correct Δr and ΔZ distributions?

Sideband data and off-resonance data have different Δr and ΔZ distributions.
Use Δr and ΔZ distributions with respect to the IP (not the lab origin).

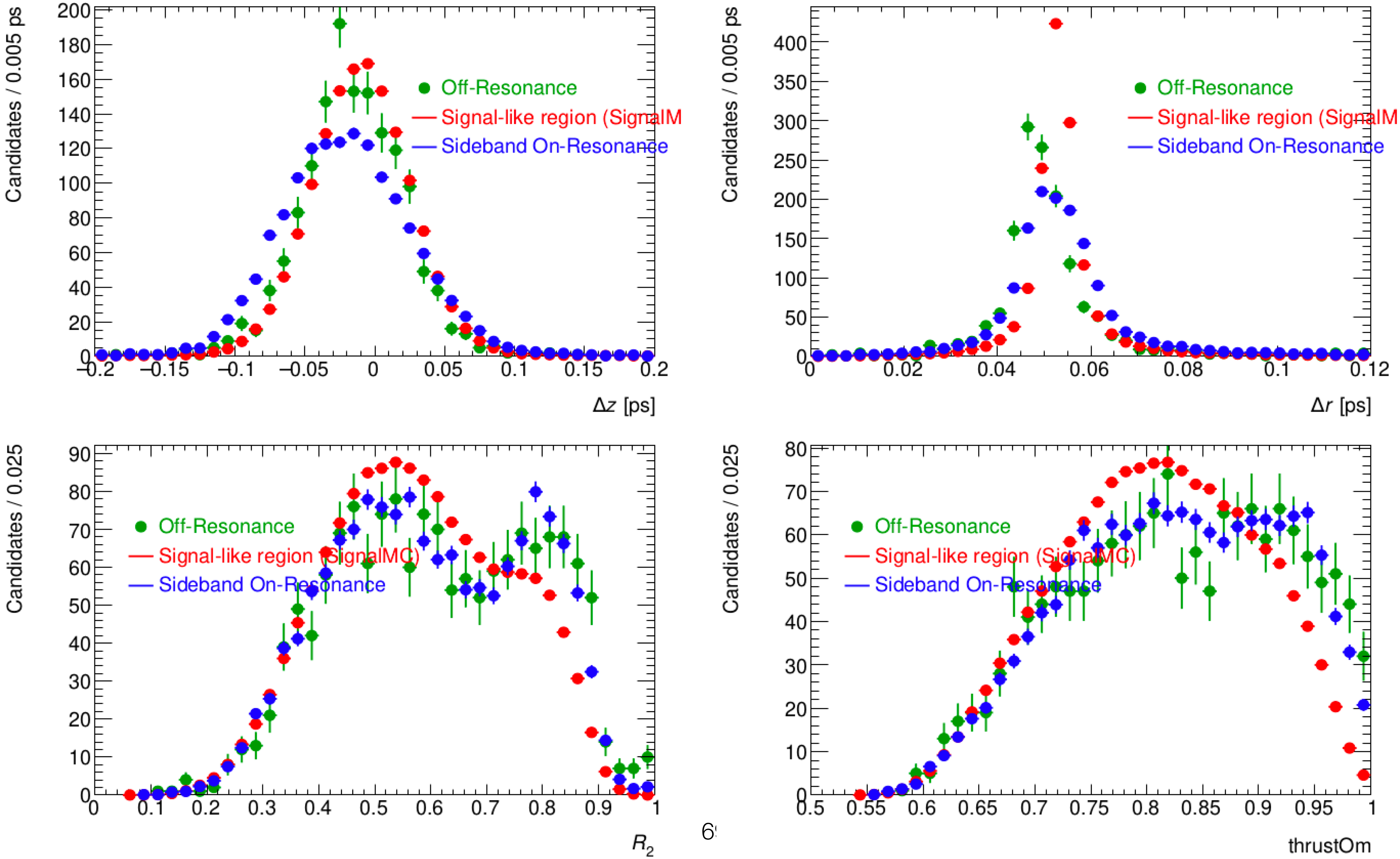
with respect to the origin | with respect to the IP



Improved situation

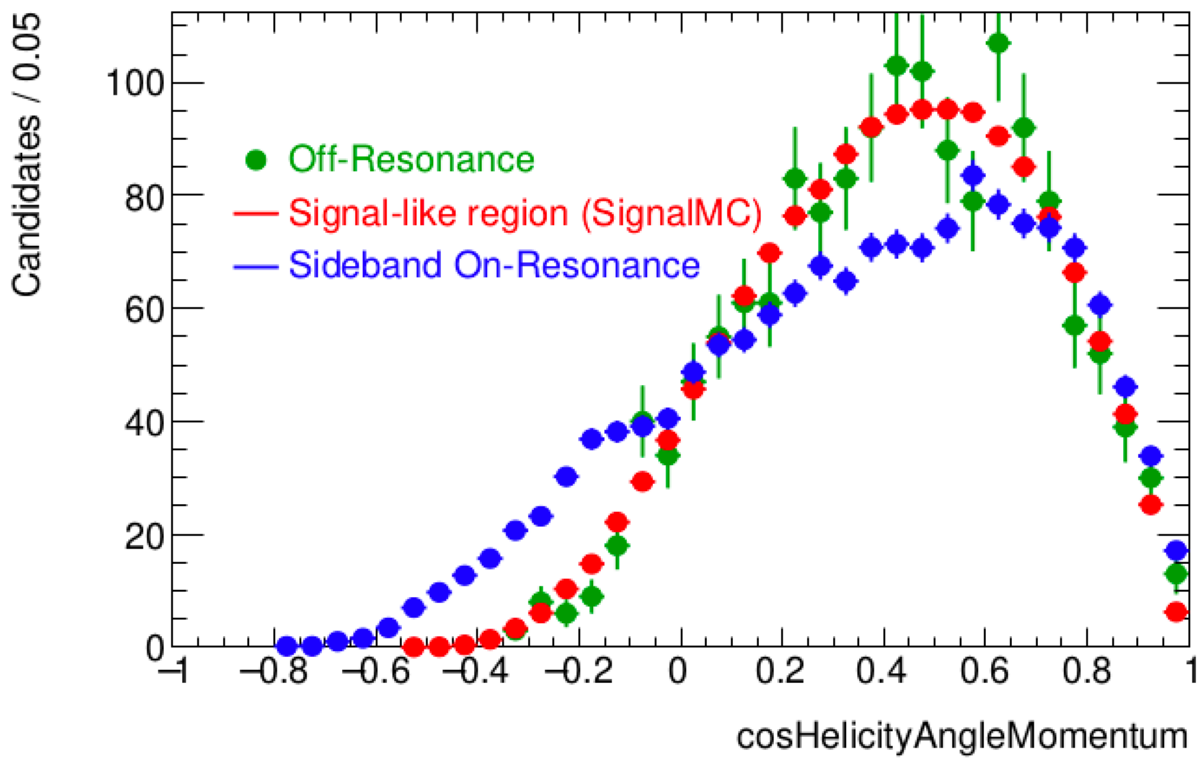
Which are the correct Δr and ΔZ distributions?

Sideband data and off-resonance data have different r and ΔZ distributions.

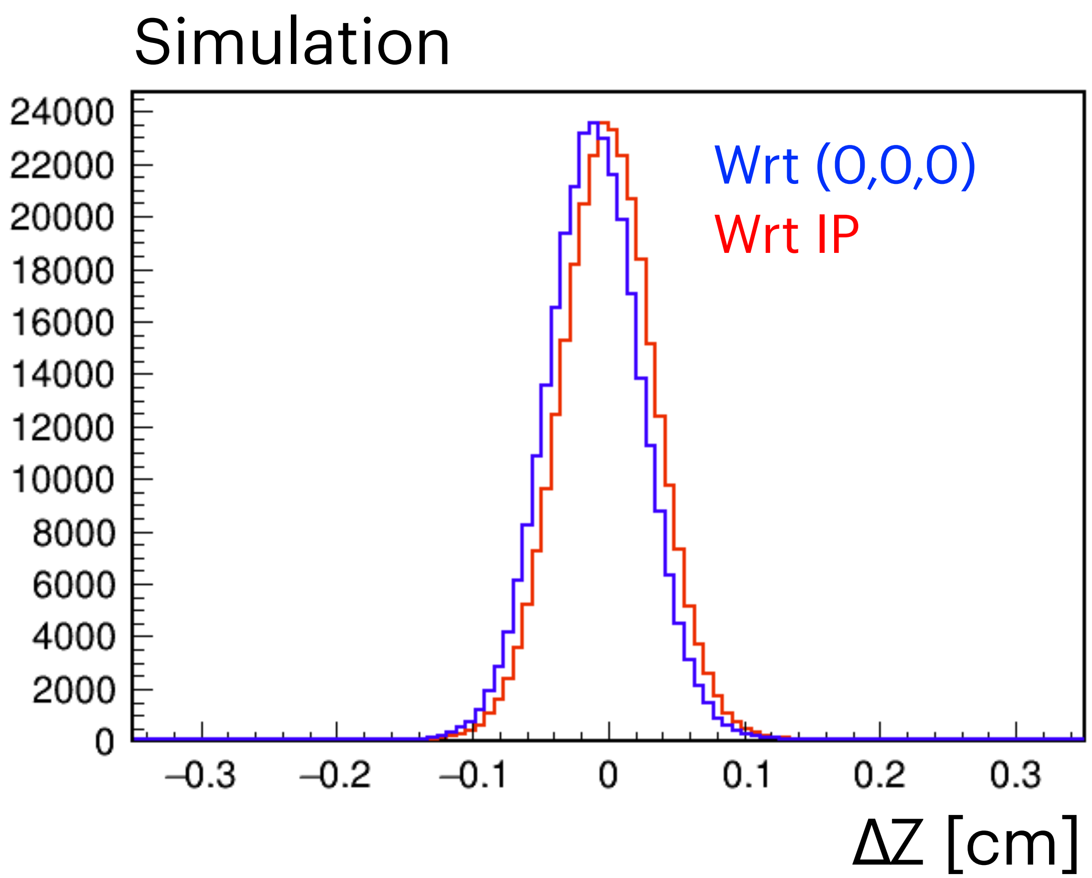


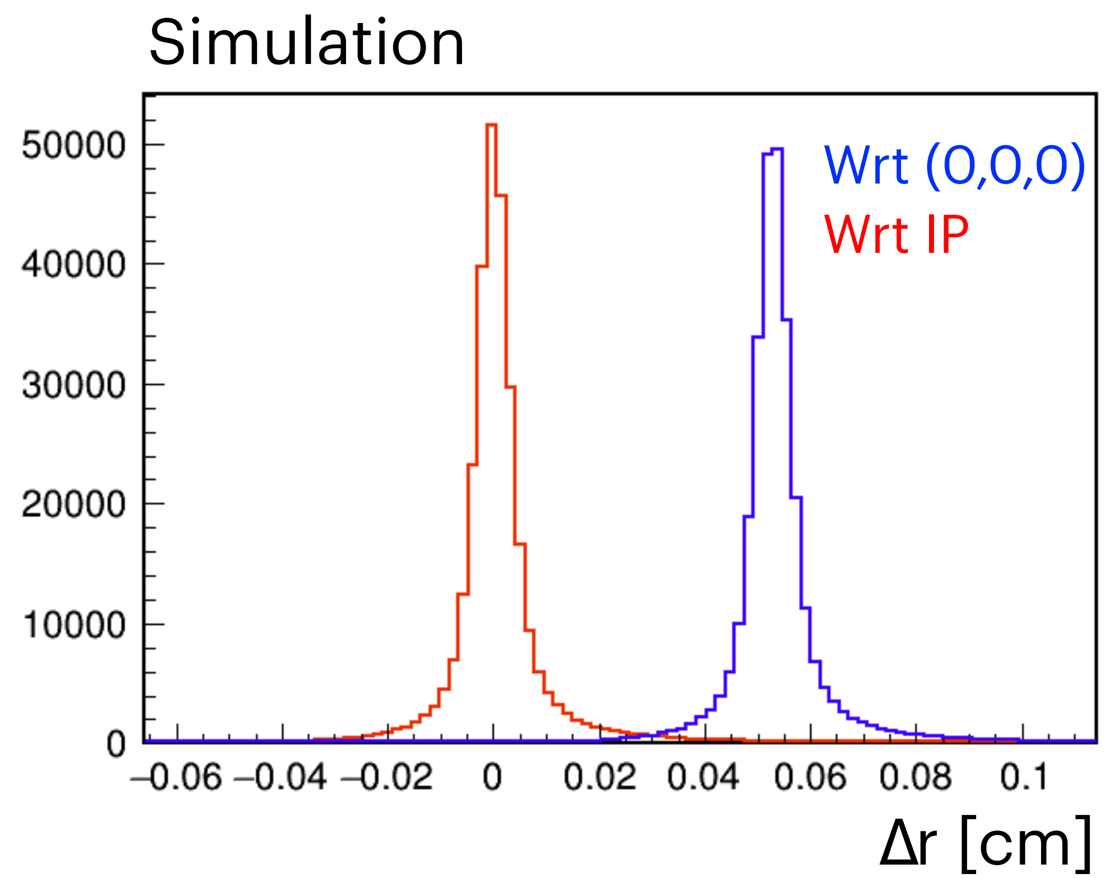
Which are the correct Δr and ΔZ distributions?

Sideband data and off-resonance data have different r and ΔZ distributions.



Which are the correct Δr and ΔZ distributions?





Now we have the expected distributions!