Efficiency studies for a new BDT-based Photon Id WP in the search for $HH \rightarrow bb\gamma\gamma$

Leonardo Carminati, Giulia Maineri, Stefano Manzoni, Elena Mazzeo, Davide Mungo, Ruggero Turra

University of Milan

Borsa di studio INFN

- borsa di studio trimestrale "La fisica delle particelle per esplorare l'Universo"
- in un grande laboratorio di ricerca: CERN, KEK, BEPC, PSI o Fermilab
- per laureandi triennali e magistrali




link ai bandi disponibili, link al bando e ai progetti dell'anno scorso

My project

Project:

Photon identification optimization for the HH->bbyy search with the ATLAS detector at the LHC

This project aims to improve the identification performance of photons with the ATLAS detector using advanced ML techniques. Currently, a photon is defined at the analysis level by a series of rectangular selections on calorimeter-based discriminant variables. Three different working points in terms of signal efficiency/background rejection are defined for these requirements. In this project, we aim to improve the current photon identification selections using advanced ML techniques and compare the performance with existing menus. In addition, the ML approach will provide an opportunity to overcome the limitations of the fixed working point of the selection by providing a (pseudo)continuous identification menu. The main challenge of the project will be to measure in situ the efficiency and scaling factors for the new identification algorithm using photons from Z->ee/mumu radiative decays data events. The main physical target of this project is to improve the sensitivity on the production of di-higgs in the final yybb state with Run data 3 by increasing the efficiency of signal selection. One possibility that we would like to explore is the inclusion of the photon identification score directly into the optimization of the analysis, so that the analysis can exploit photons of different qualities depending on the specific kinematic features of a given event category.

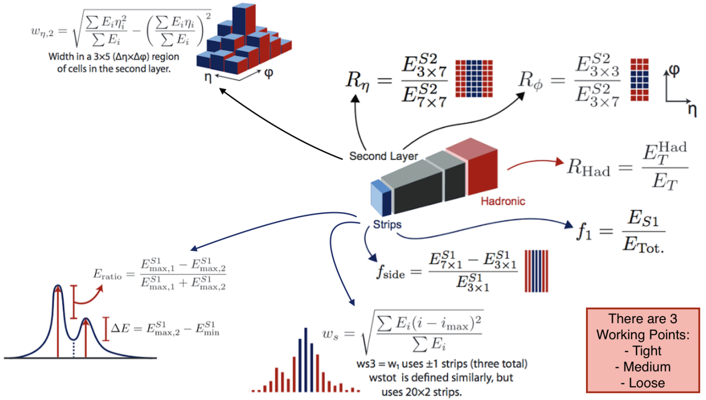
Laboratory:

Contact persons:

Leonardo Carminati, Stefano Manzoni, Elena Mazzeo, Ruggero Turra, Davide Mungo

A new BDT-based WP

The current ATLAS photon ID algorithm is based on a set of **rectangular selections** on calorimetric **shower shape** variables



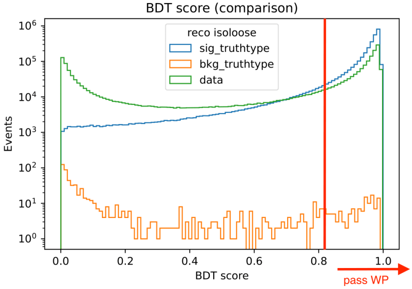
There are 3 Working Points:
- Tight
- Medium
- Loose

Typically, photon-related analyses use the Tight WP.

A new BDT-based WP

To improve the sensitivity of the $HH \rightarrow bb\gamma\gamma$ analysis, a new photon ID based on a **Boosted Decision Trees** algorithm combining the information of the same discriminating variables of the cut-based WPs has been developed.

A **new Working Point** has been tuned, cutting the BDT score output distribution such to have the same background rejection as the tight WP.

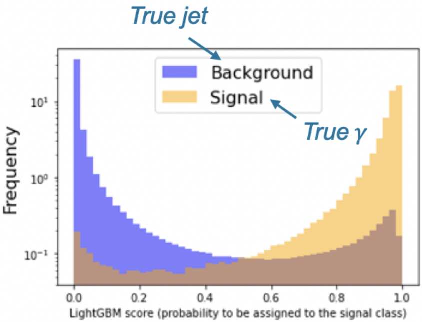


BDT optimization

The BDT was trained by Marta on **mc16** samples, using the **LightGBM** package.

Training samples:

- Signal → photons from Pythia γ+jets MC sample;
- Background → jets from Pythia multijets MC sample.



Input variables:

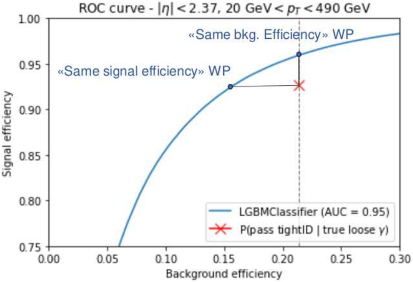
- photon kinematics: η, p_T;
- shower shapes: R_η, R_ϕ, w_η^1, w_η^2, $w_{s,tot}$, R_{had}, R_{had1}, f_1, ΔE, E_{ratio};
- photon conversion: convRadius, convType.

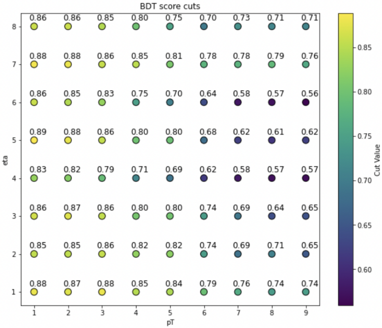
Tommaso is currently studying the WP on MC20 samples.[a]

[a] https://indico.cern.ch/event/1466239/contributions/6173967/attachments/2945101/5175802/phID_BDT.pdf

A new candidate WP, based on the BDT score distribution, has been defined by choosing the BDT score cut that guarantees the **same background rejection** of the standard Tight WP. The WP-tuning is done in η, p_T bins.

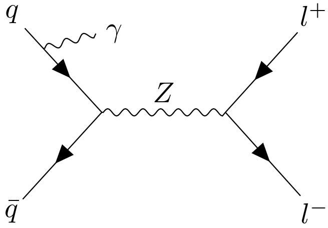
$$\varepsilon_{sig(bkg)} = \text{P (pass ID} \mid \text{loose } \gamma \text{ matched to true(fake) } \gamma \text{)}$$

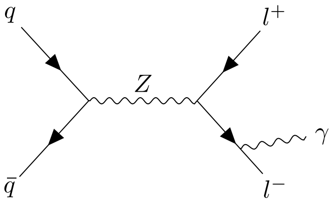




Efficiency of the WP

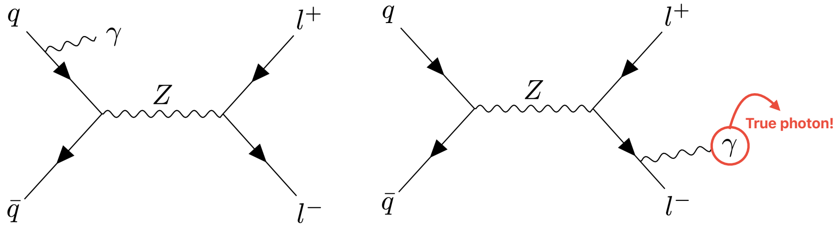
We want to compute the efficiency of this WP in data and Monte Carlo and then calculate corrective factors for MC.
$\Rightarrow$ we need a pure sample of photons. How to extract true photons from data sample?
$\Rightarrow Z(\to ll\gamma)$





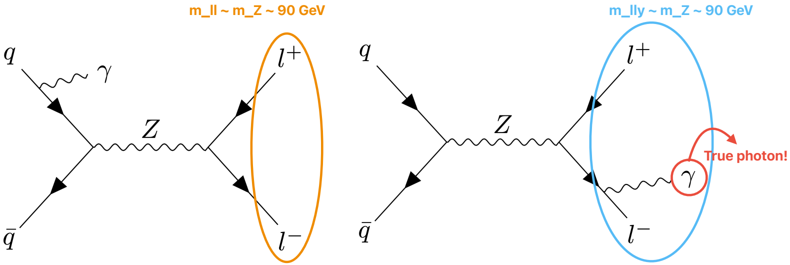
Efficiency of the WP

We want to compute the efficiency of this WP in data and Monte Carlo and then calculate corrective factors for MC.
$\Rightarrow$ we need a pure sample of photons. How to extract true photons from data sample?
$\Rightarrow Z(\rightarrow ll\gamma)$



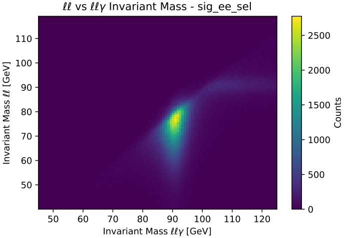
Efficiency of the WP

We want to compute the efficiency of this WP in data and Monte Carlo and then calculate corrective factors for MC.
$\Rightarrow$ we need a pure sample of photons. How to extract true photons from data sample?
$\Rightarrow Z(\rightarrow ll\gamma)$

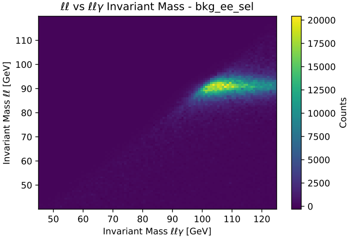


m_{ll} vs $m_{ll\gamma}$

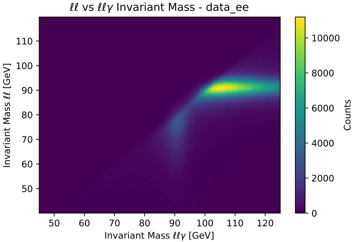
Electronic channel



Signal $Z\gamma$

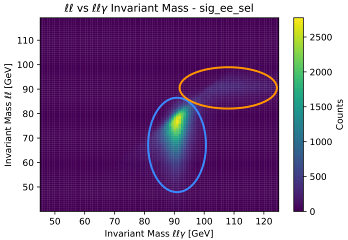


Background Z+jets

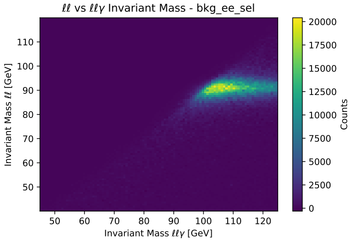


Data

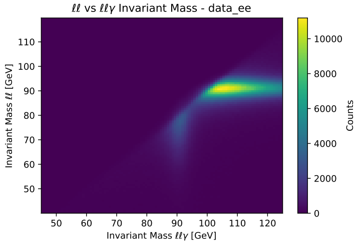
Electronic channel



Signal $Z\gamma$



Background Z+jets

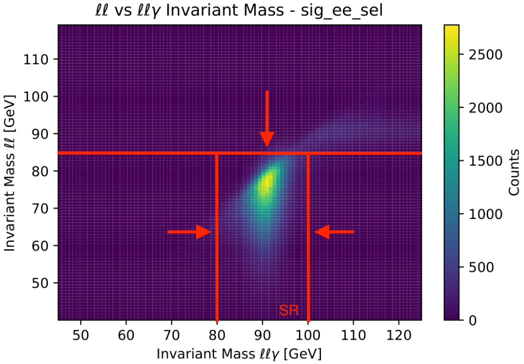


Data

Signal Region definition

Define a Signal Region:

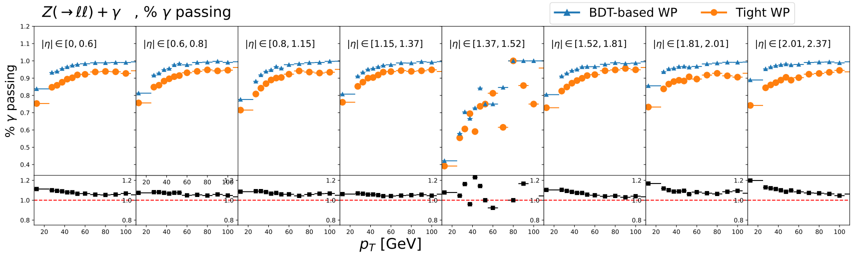
- $m_{ll} < 85$ GeV
- 80 GeV $< m_{ll\gamma} < 100$ GeV
- ph.isoloose = True



Application of the BDT-based WP to $Z\gamma$ MC

Quick check of the efficiency of the BDT-based WP optimized to have the same background rejection as the Tight WP:

- no requirements on $m_{ll\gamma}$, m_{ll}, photon isolation
- counting the fraction of truth photons passing the WP, no fit performed
- no separation for different convType and leptonic channels



→ gain of efficiency!

Efficiency and Scale Factors

We want to calculate the efficiency of this WP:

$$\varepsilon_s(\eta, p_T, convType) = \frac{N_s^{pass}(\eta, p_T, convType)}{N_s^{all}(\eta, p_T, convType)}$$

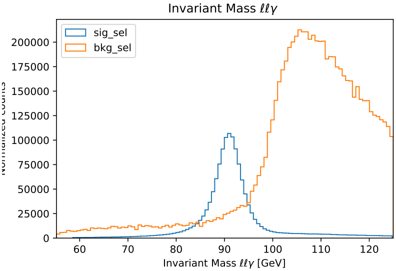
...and correct the efficiency in Monte Carlo on data, with Scale Factors (SFs):

$$SF(\eta, p_T, convType) = \frac{\varepsilon_s^{data}(\eta, p_T, convType)}{\varepsilon_s^{MC}(\eta, p_T, convType)}$$

BDT-based WP efficiency on data

Let's start with the efficiency on data ε_s^{data}.

$$\varepsilon_s^{data}(\eta, p_T, convType) = \left[\frac{N_s^{pass}(\eta, p_T, convType)}{N_s^{all}(\eta, p_T, convType)} \right]^{data}$$



Data sample contains both signal and background
⇒ fit $m_{ll\gamma}$ distribution with a s+b template
⇒ estimate signal yields (pass and all) integrating PDF_s

BDT-based WP efficiency on data

Let's start with the efficiency on data ε_s^{data}.

$$\varepsilon_s^{data}(\eta, p_T, convType) = \left[\frac{N_s^{pass}(\eta, p_T, convType)}{N_s^{all}(\eta, p_T, convType)}\right]^{data}$$

In order to compute N_s^{pass}, N_s^{all} we perform a fit on the $m_{ll\gamma}$ distribution:

- for $p_T < 20\,\text{GeV}$: fit data with $n_s \cdot \text{PDF}_s + n_b \cdot \text{PDF}_b$;
- for $p_T > 20\,\text{GeV}$: fit data with $n_s \cdot \text{PDF}_s$.

where:

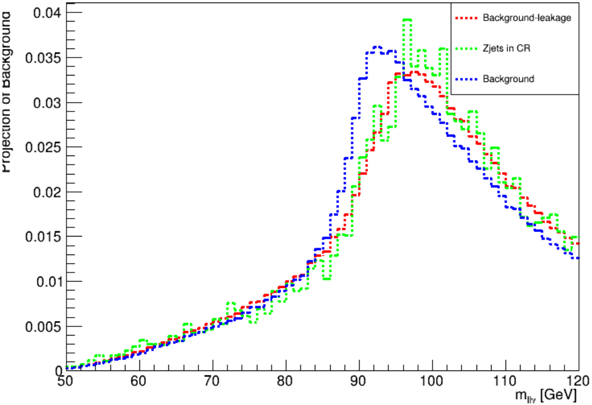
- PDF_s is a template based on the $Z\gamma$ MC sample;
- PDF_b is a template extracted from data in a Control Region (anti-isol & anti-loose), subtracting the signal leakage from $Z\gamma$ MC

BDT-based WP efficiency on data: background template

PDF_b is a template extracted from data in a Control Region (anti-isol & anti-loose), subtracting the signal leakage from $Z\gamma$ MC

$\rightarrow$We chose this template due to lack of statistic in Zjets causing fit instability

Check $m_{ll\gamma}$ shape:



PDF_b = [data - Zγ]$_{CR}$

Control Region:

- ph.isoloose!=1
- ph.loose_id!=1

BDT-based WP efficiency on data

$$\varepsilon_s^{data}(\eta, p_T, convType) = \left[\frac{N_s^{pass}(\eta, p_T, convType)}{N_s^{all}(\eta, p_T, convType)}\right]^{data}$$

In order to compute N_s^{pass}, N_s^{all} we perform a fit on the $m_{ll\gamma}$ distribution:

- for $p_T < 20$ GeV: fit data with $n_s \cdot \mathrm{PDF}_s + n_b \cdot \mathrm{PDF}_b$;
- for $p_T > 20$ GeV: fit data with $n_s \cdot \mathrm{PDF}_s$.

We then compute N_s^{pass}, N_s^{all} as:

$$N_s^{pass} = n_s^{pass} \left(\int_{80\mathrm{GeV}}^{100\mathrm{GeV}} PDF_s^{pass} \, d(m_{ll\gamma}) \right)$$

$$N_s^{all} = n_s^{all} \left(\int_{80\mathrm{GeV}}^{100\mathrm{GeV}} PDF_s^{all} \, d(m_{ll\gamma}) \right)$$

BDT-based WP efficiency on data: binned

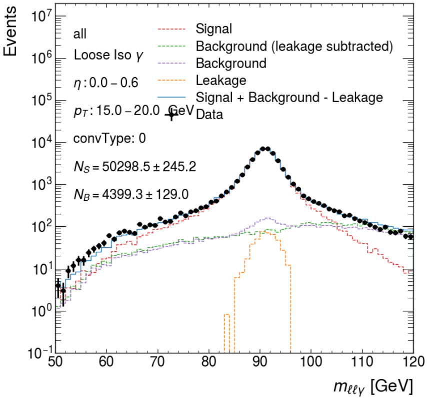
Let's then perform the fit in **bins** of η, p_T and convType and separately for the two **leptonic channels**.

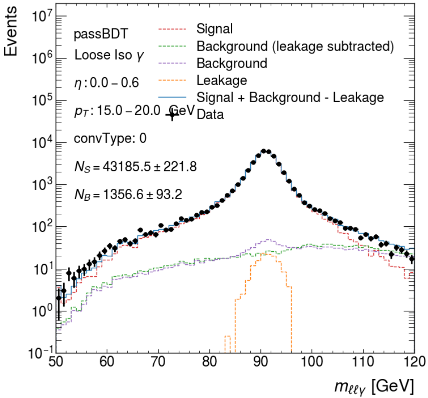
- channel $\in$ ["ee", "$\mu\mu$"]
- $\eta \in [(0, 0.6), (0.6, 1.37), (1.52, 1.81), (1.81, 2.37)]$
- $p_T \in [10, 15, 20, 25, 30, 35, 40, 50, 70, 150]$GeV
- convType $\in [0, > 0]$

For each bin the fit is performed both for the "pass" and "all" configurations.

BDT-based WP efficiency on data: binned

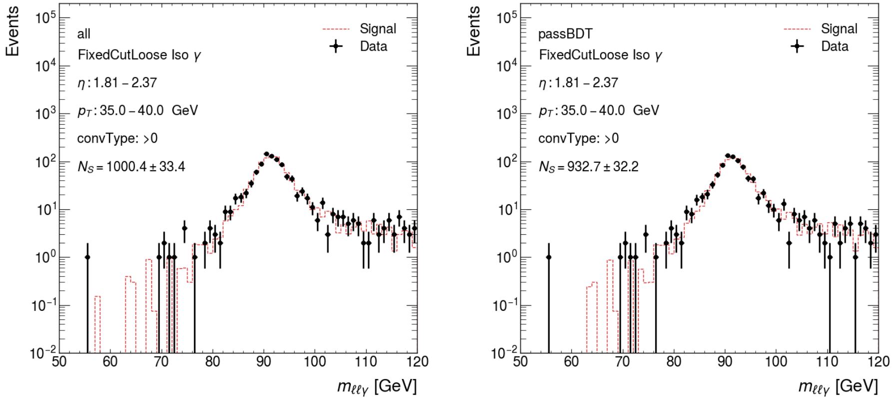
Example: Electronic channel, $\eta \in [0, 0.6]$, $p_T \in [15, 20]$GeV, unconverted





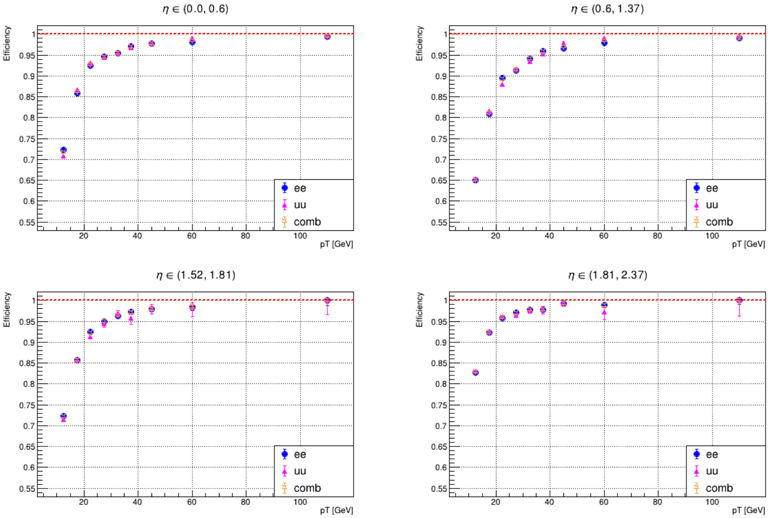
BDT-based WP efficiency on data: binned

Example: Electronic channel, $\eta \in [1.81, 2.37]$, $p_T \in [35, 40]$GeV, converted

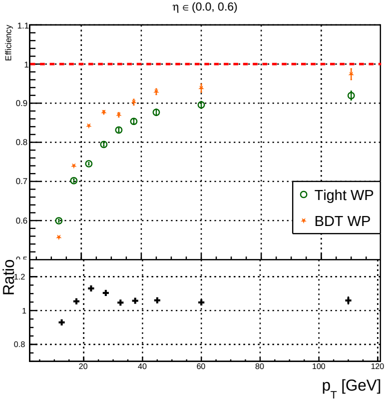


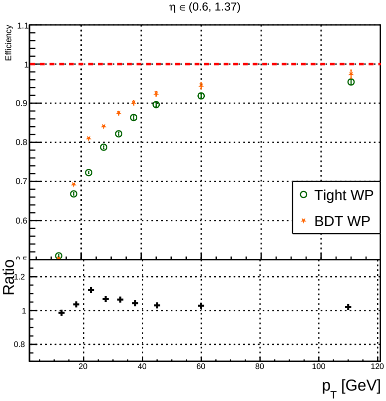
BDT-based WP efficiency on data: binned

Example: Combined channel, unconverted *only statistic errors in these plots*



Example: Efficiency on data, combined channel, converted





BDT-based WP efficiency on MC: calculation

Then we need the efficiency on MC: ε_s^{MC}.

$$\varepsilon_s^{MC}(\eta, p_T, convType) = \left[\frac{N_s^{pass}(\eta, p_T, convType)}{N_s^{all}(\eta, p_T, convType)} \right]^{MC}$$

BDT-based WP efficiency on MC: calculation

Then we need the efficiency on MC: ε_s^{MC}.

$$\varepsilon_s^{MC}(\eta, p_T, convType) = \left[\frac{N_s^{pass}(\eta, p_T, convType)}{N_s^{all}(\eta, p_T, convType)} \right]^{MC}$$

In order to compute N_s^{pass}, N_s^{all} we perform again a fit on the $m_{ll\gamma}$ distribution:

- fit $Z\gamma$ MC sample with $n_s \cdot \mathrm{PDF}_s$

where PDF_s is a template based on $Z\gamma$ MC samples.

BDT-based WP efficiency on MC: calculation

Then we need the efficiency on MC: ε_s^{MC}.

$$\varepsilon_s^{MC}(\eta, p_T, convType) = \left[\frac{N_s^{pass}(\eta, p_T, convType)}{N_s^{all}(\eta, p_T, convType)} \right]^{MC}$$

In order to compute N_s^{pass}, N_s^{all} we perform again a fit on the $m_{ll\gamma}$ distribution:

- fit $Z\gamma$ MC sample with n_s·PDF_s

where PDF_s is a template based on $Z\gamma$ MC samples.

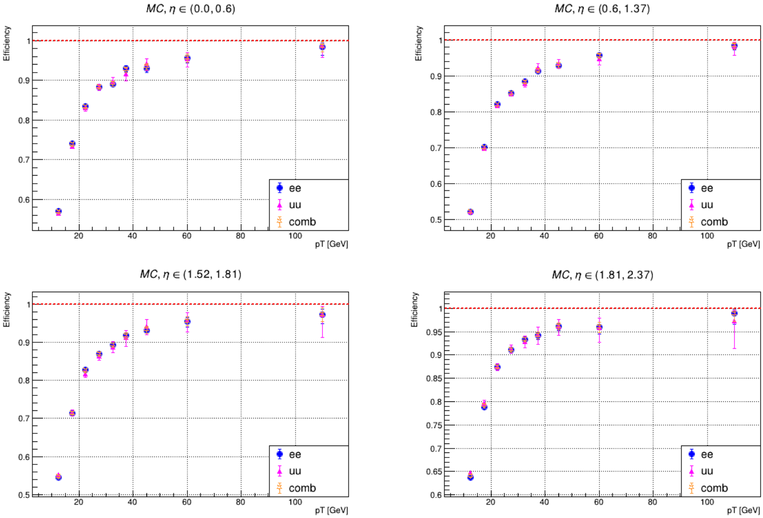
We then compute N_s^{pass}, N_s^{all} as:

$$N_s^{pass} = n_s^{pass} \left(\int_{80\text{GeV}}^{100\text{GeV}} PDF_s^{pass} \, d(m_{ll\gamma}) \right)$$

$$N_s^{all} = n_s^{all} \left(\int_{80\text{GeV}}^{100\text{GeV}} PDF_s^{all} \, d(m_{ll\gamma}) \right)$$

BDT-based WP efficiency on MC: binned

Example: Combined channels, converted

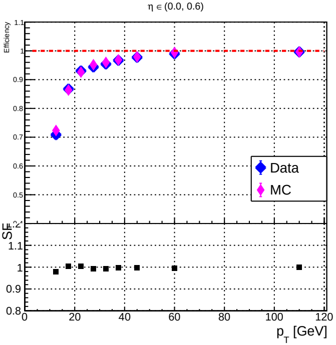


We are now ready to calculate Scale Factors (SFs).

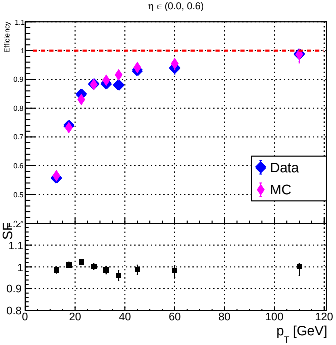
$$SF(\eta, p_T, convType) = \frac{\varepsilon_s^{data}(\eta, p_T, convType)}{\varepsilon_s^{MC}(\eta, p_T, convType)}$$

Example: Muonic channel, $\eta \in [0, 0.6]$

convType = 0



convType > 0



Systematic uncertainties on SFs

We need to introduce **systematic** uncertainties on SFs:

- uncertainties on the **signal Monte Carlo model** σ_{MCgen}
  $\rightarrow$ re-calculate the SFs using samples generated with different MC generators.

Systematic uncertainties on SFs

We need to introduce **systematic** uncertainties on SFs:

- uncertainties on the **signal Monte Carlo model** σ_{MCgen}
  $\rightarrow$ re-calculate the SFs using samples generated with different MC generators.
- uncertainties on the **fitting procedure** σ_{fit}
  $\rightarrow$ re-calculate the SFs varying the fit range.

Systematic uncertainties on SFs

We need to introduce **systematic** uncertainties on SFs:

- uncertainties on the **signal Monte Carlo model** σ_{MCgen}
  $\rightarrow$ re-calculate the SFs using samples generated with different MC generators.
- uncertainties on the **fitting procedure** σ_{fit}
  $\rightarrow$ re-calculate the SFs varying the fit range.
- uncertainties on the **background model** σ_{bkg}
  $\rightarrow$ re-calculate the SFs varying the background template.

Systematic uncertainties on SFs

We need to introduce **systematic** uncertainties on SFs:

- uncertainties on the **signal Monte Carlo model** σ_{MCgen}
  $\rightarrow$ re-calculate the SFs using samples generated with different MC generators.
- uncertainties on the **fitting procedure** σ_{fit}
  $\rightarrow$ re-calculate the SFs varying the fit range.
- uncertainties on the **background model** σ_{bkg}
  $\rightarrow$ re-calculate the SFs varying the background template.

We will then assign to SFs a total uncertainty:

$$\sigma_{tot} = \sqrt{\sigma^2_{stat} + \sigma^2_{MCgen} + \sigma^2_{fit} + \sigma^2_{bkg}}$$

where

$$\sigma_{MCgen} = SF_{(\text{varied MC gen})} - SF_{nom} \qquad \sigma_{fit} = SF_{(\text{varied fit-range})} - SF_{nom}$$

$$\sigma_{bkg} = SF_{(\text{varied bkg model})} - SF_{nom}$$

Systematic uncertainties on SFs: combination

Let's combine the systematic contributions with the statistic uncertainty:

$$\sigma_{tot} = \sqrt{\sigma_{stat}^2 + \sigma_{MCgen}^2 + \sigma_{fit}^2 + \sigma_{bkg}^2}$$

Example

comb, unconverted, $\eta \in (1.81, 2.37)$

pT [GeV]	SF	Stat.Error	Gen.Error	Fit.Error	Bkg.Error	Total.Error
12.5	0.9917	0.0032	0.0001	0.0010	0.0236	0.0238
17.5	0.9994	0.0025	0.0005	0.0001	0.0062	0.0067
22.5	1.0005	0.0024	0.0089	0.0006	0.0000	0.0092
27.5	0.9971	0.0025	0.0001	0.0002	0.0000	0.0026
32.5	0.9965	0.0030	0.0009	0.0005	0.0000	0.0032
37.5	0.9939	0.0044	0.0008	0.0005	0.0000	0.0045
45.0	1.0048	0.0042	0.0006	0.0009	0.0000	0.0044
60.0	1.0009	0.0072	0.0009	0.0003	0.0000	0.0073
110.0	1.0050	0.0082	0.0000	0.0009	0.0000	0.0083

Conclusions and next steps

Summary:

- The **efficiency** of a new BDT-based WP for photons has been computed, in different η, p_T, convType bins and different leptonic channels, using $Z\gamma$ sample.
- For the calculation of the efficiency, a **fit** of the 3-bodies invariant mass $m_{ll\gamma}$ has been used. For $p_T < 20\,\text{GeV}$, a **background** template from data in a proper CR has been used, subtracting the signal leakage.
  $\rightarrow$ the new WP allows a gain in efficiency
- The efficiency has been computed in both data and Monte Carlo and **Scale Factors** have been extracted as a ratio between data and Monte Carlo.
  $\rightarrow$ SFs are close to 1 as expected
  $\rightarrow$ Scale factors and their systematic uncertainties for BDT-based WP are of the same order as the ones of the Tight WP
- **Systematic uncertainties** on the SFs have been computed considering contributions from signal modelling, fitting procedure and background modelling.

Next steps:

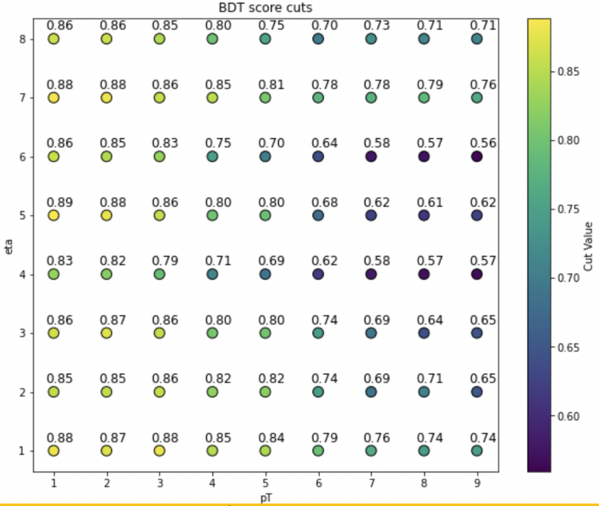
- apply same efficiency and scale factors calculation procedure using the new BDT-based WP optimized on **MC20** and **MC23**

Backup

Tuning a new WP

$|\eta|$ bins: $[0, 0.6, 0.8, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37]$

p_T bins: $[20, 25, 30, 35, 40, 45, 50, 55, 65, 490]$GeV.



Preliminary studies: used ntuples

Let's apply the BDT-based WP to $Z\gamma$, Zjets MC samples, full Run2 data sample.

Used ntuples can be found in:
dir = /eos/atlas/atlascerngroupdisk/perf-egamma/photonID/NTUP_ZLLG/
(thanks to Nadezda).

Monte Carlo[2]: $Z\gamma$, Z+jets

- MC16a: dir + mc16a_13TeV/00-03-01
- MC16d: dir + mc16d_13TeV/00-03-01
- MC16e: dir + mc16e_13TeV/00-03-01

Signal $Z\gamma$: mc16*.Sh_224_NN30NNLO_eegamma_LO_pty_7_15.*root - mc16*.Sh_224_NN30NNLO_mumugamma_LO_pty_140_E_CMS.*.root
Background Z+jets: mc16*.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV0_70*.root - mc16*.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV1000_E_CMS.*.root

Data[2]

- 2015: dir + data15_13TeV/00-03-01, $L = 3219{,}56\,\mathrm{pb}^{-1}$
- 2016: dir + data16_13TeV/00-03-01, $L = 32\,988{,}1\,\mathrm{pb}^{-1}$
- 2017: dir + data17_13TeV/00-03-01, $L = 43\,587{,}3\,\mathrm{pb}^{-1}$
- 2018: dir + data18_13TeV/00-03-01, $L = 58\,450{,}1\,\mathrm{pb}^{-1}$

[0] https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ZllgAnalysis

Preliminary studies: used ntuples

Pre-selections: cuts applied at ntuple dumper level [1].

- Photon:
  - $p_T^{min} = 5\,\text{GeV}$;
  - $\eta^{max} = 2.37$
  - crack region: $\eta \in [1.37, 1.52]$.
- Electrons (Muons):
  - $p_T^{min} = 10\,\text{GeV}$;
  - $\eta^{max} = 2.47(2.5)$;
  - crack region: $\eta \in [1.37, 1.52]$;
  - isolation: loose;
  - $sig_{d_0}^{max} = 5(3)$, max transverse impact parameter significance;
  - ${z_0}^{max} = 0.5$, max longitudinal impact parameter .
- $ll\gamma$ candidates:
  - opposite charged leptons;
  - trigger match;
  - $m_{ll} \in [40, 120]\text{GeV}$;
  - $m_{ll\gamma} \in [45, 125]\text{GeV}$;
  - $\Delta R(p_T^{\mu}, p_T^{\gamma}) > 0.01$;
  - $\Delta R(p_T^{e}, p_T^{\gamma}) > 0.2$.

[1] https://gitlab.cern.ch/ATLAS-EGamma/Software/PhotonID/RadiativeZ/-/blob/master/ZllgAnalysis/Root/AnalysisConfig.cxx

Further cuts and selections

True γ signal from $Z\gamma$: $13 \leq$ ph.truth_type ≤ 15

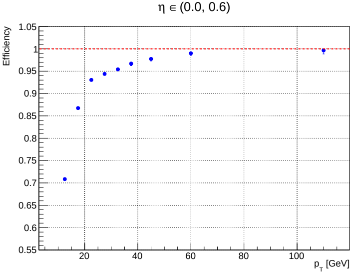
Fake γ background from Z+jets: ph.truth_type ≤ 12 or ph.truth_type ≥ 16

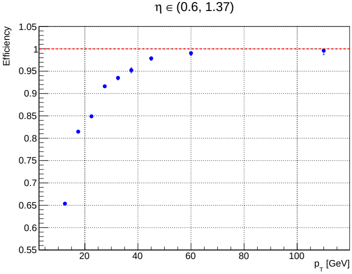
where

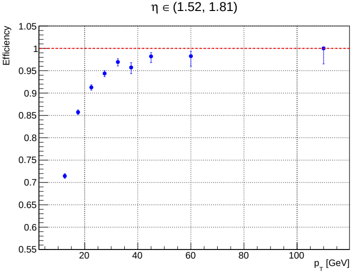
- UnknownPhoton $= 13$;
- IsoPhoton $= 14$;
- NonIsoPhoton $= 15$.

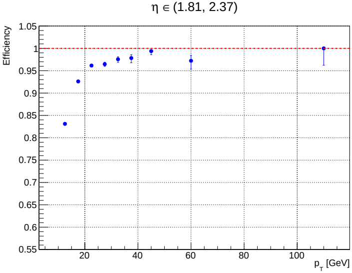
BDT-based WP efficiency on data: binned

Example: Muonic channel, unconverted *only statistic errors in these plots*



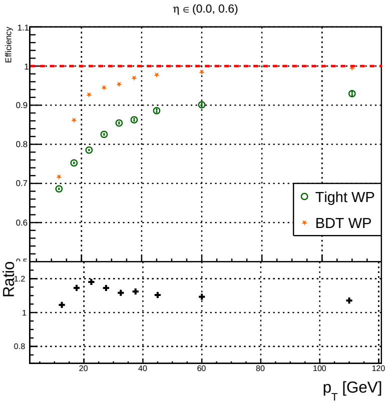


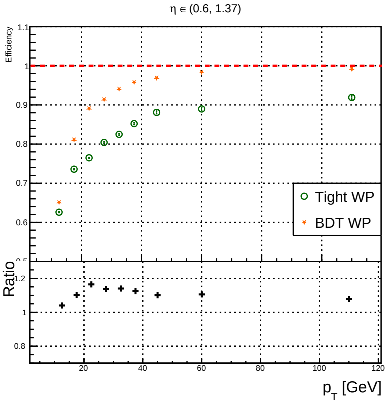




Comparison with the Tight WP

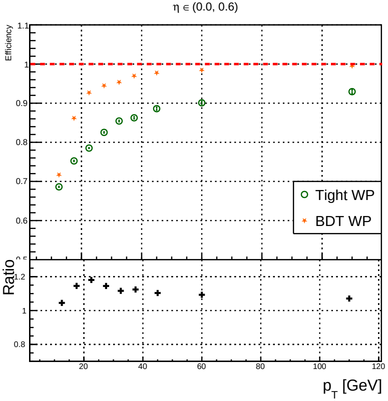
Example: Efficiency on data, combined channel, unconverted

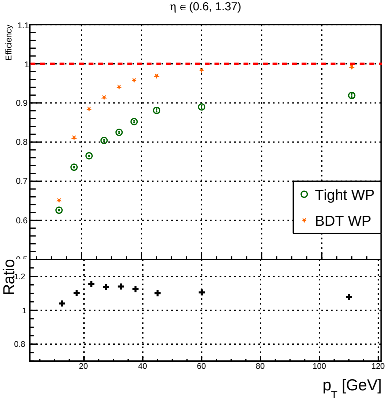




Comparison with the Tight WP

Example: Efficiency on MC, combined channel, unconverted





BDT optimization

Selection on the training samples

- Loose selection on shower shapes variables: $f_1 > 0.005$ and $E_{277} > 0{,}1\,\text{MeV}$.
- A selection on the photon transverse momentum: $20\,\text{GeV} < p_T < 1\,\text{TeV}$ has been applied (the $p_T > 20\,\text{GeV}$ threshold is already present in the p_T profiles of the two MC samples).

+

Signal:

- pdgID = 22
- mother pdgID < 100
- truth type = 14 (only direct photons)

Background:

- pdgID != 22
  or
- pdgID = 22 + mother pdgID < 100

Shower shape variables

Category	Description	Name
Hadronic leakage	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster (used over the ranges $\|\eta\| < 0.8$ and $\|\eta\|.1.37$)	R_{had1}
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \|\eta\| < 1.37$)	R_{had}
EM second layer	Ratio of the sum of the energy of the cells contained in a $3 \times 7 \eta \times \phi$ rectangle(measured in cell units) to the sum of the cell energies in a 7×7 rectangle, both centred around the most energetic cell	R_η
	Lateral shower width, $\sqrt{((\Sigma E_i \eta_i^2)(\Sigma E_i) - ((\Sigma E_i \eta_i)(\Sigma E_i))^2)}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated with a window of 3×5 cells	$w_{\eta 2}$
	Ratio of the sum of the energies of the cells contained in a $3 \times 3 \eta \times \phi$ rectangle(measured in cell units) to the sum of the cell energies in a $3 \times 7 \eta \times \phi$ rectangle, both centred around the most energetic cell	R_ϕ
EM first layer	Total lateral shower width, $\sqrt{(\Sigma E_i (i - i_{max})^2)(\Sigma E_i)}$, where i runs over all cells in a window of $\Delta\eta \approx 0.0625$ and i_{max} is the index of the highest-energy cell	w_{stot}
	Lateral shower width, $\sqrt{(\Sigma E_i (i - i_{max})^2)(\Sigma E_i)}$ where i runs over all cells in a window of 3 cells around the highest-energy cell	$w_{s3}/w_{\eta 1}$
	Energy fraction outside core of three central cells, within seven cells	f_{side}
	Difference between the energy of the cell associated with the second maximum, and the energy reconstructed in the cell with the smallest value found between the first and second maxima	ΔE_s
	Ratio of the energy difference between the maximum energy deposi and the energy deposit in a secondary maximum in the cluster to the sum of these energies	E_{ratio}
	Ratio of the energy measured in the first layer of the electromagnetic calorimeter to the total energy of the EM cluster	f_1

https://indico.ihep.ac.cn/event/17584/contributions/115526/attachments/65947/77808/atlas-note.pdf

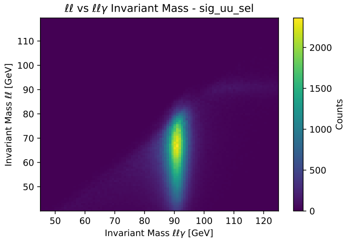
ConvRadius

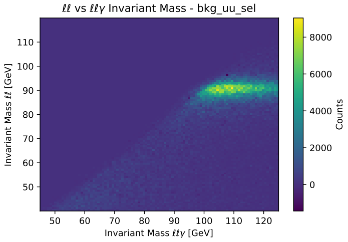
convRadius
= radius of the conversion vertex computed as $\sqrt{x^2 + y^2}$

Same bkg efficiency WP

Since $HH \rightarrow \gamma\gamma bb$ is statistically limited, we have privileged a higher signal efficiency while retaining a strong rejection for fake γ.

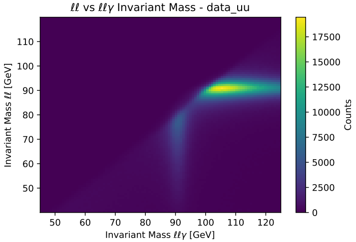
We have therefore opted for the "Same bkg. efficiency" WP.





Signal $Z\gamma$

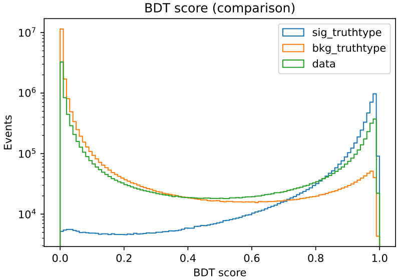
Background Z+jets



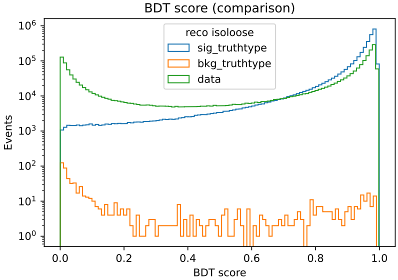
Data

Application of the BDT

Isoloose* not-applied



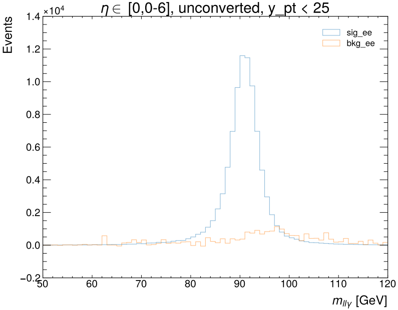
Isoloose* applied

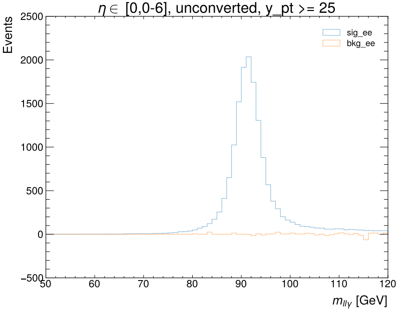


$*ph.isoloose = True$

Why neglecting background for $p_T > 25$ GeV

Question: Are we sure we can neglect background for $p_T > 25$ GeV?



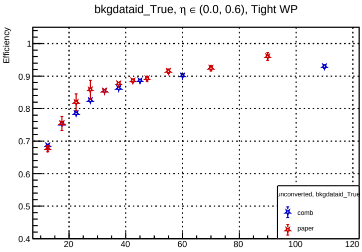


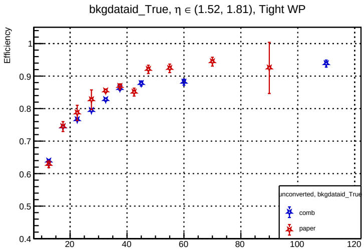
BDT-based WP efficiency: validation of the fit procedure

In order to validate the fitting procedure, we compared our results regarding **Tight WP** with the one in the Full Run-2 ATLAS Photon Identification Note[2].

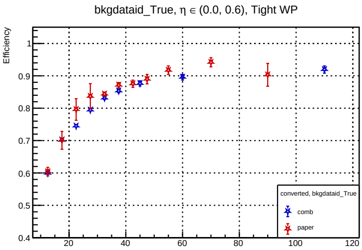
[2] https://cds.cern.ch/record/2711624/files/ATL-COM-PHYS-2020-139.pdf

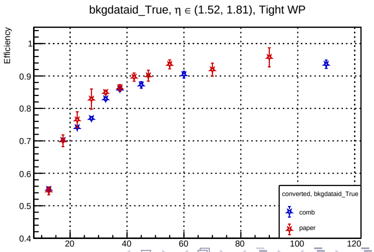
Unconverted





Converted





BDT-based WP efficiency: validation of the fit procedure

Note: In these plots, paper **uncertainties** comprehend both statistic and systematic uncertainties, while our results only include statistic uncertainties. (paper = Full Run-2 ATLAS Photon Identification Note[3])

Once the fitting procedure seems to be strong and stable, we can apply it to the **new BDT-based WP** to calculate efficiencies.

[3] https://cds.cern.ch/record/2711624/files/ATL-COM-PHYS-2020-139.pdf

Scale Factors in bins of convType

Why is it important to calculate SFs separately for converted and unconverted photons?

In MC there is a certain fraction of converted and unconverted photons, that depends on η (converted photons are from 10% to 60%).

BDT takes convType in input, using it to separate true and fake photons.

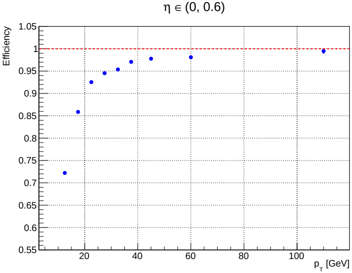
So the optimized cut on the BDT score, is the optimal one for a sample with the MC fractions of converted/unconverted, that will be different from the optimal one for data.

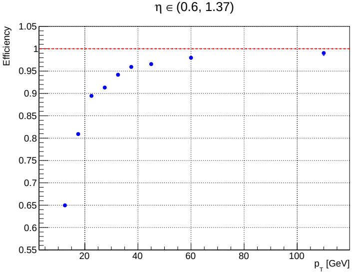
We cannot simply apply this cut on data, because in data the fractions converted/unconverted are different. We need to separate converted and unconverted photons.

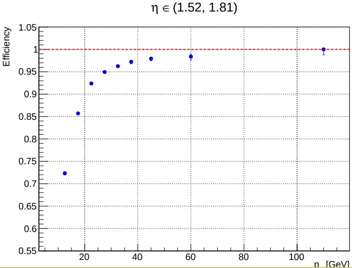
BDT-based WP efficiency on data: binned

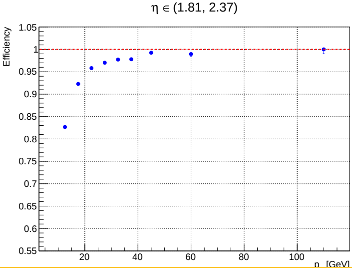
Example: Electronic channel, unconverted

only statistic errors in these plots



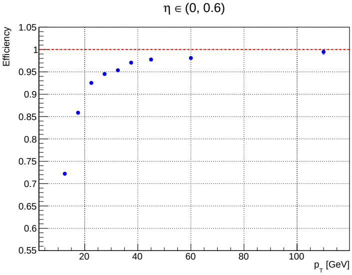


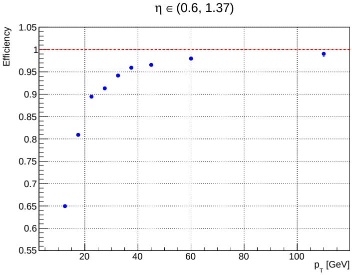


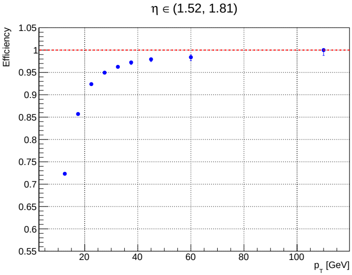


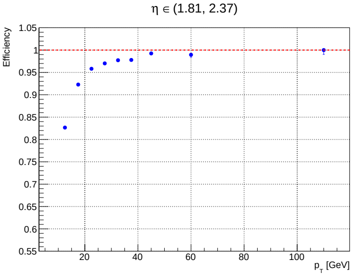
Efficiency on data, Electrons

Unconverted

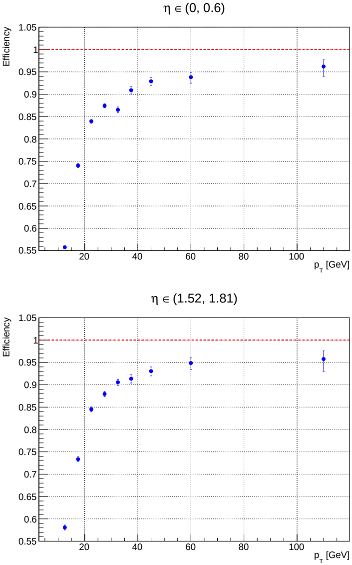






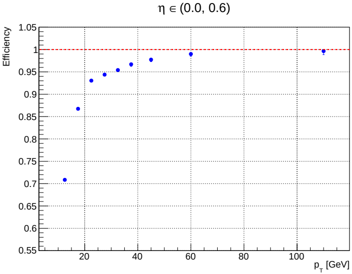


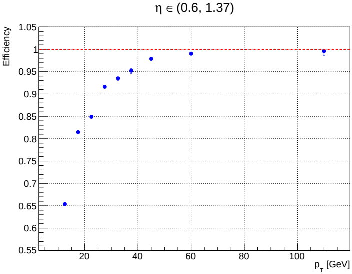
Converted

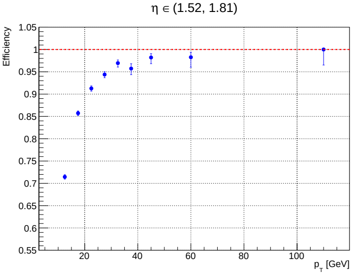


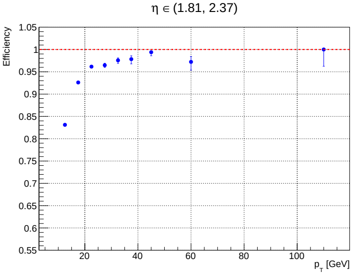
Efficiency on data, Muons

Unconverted



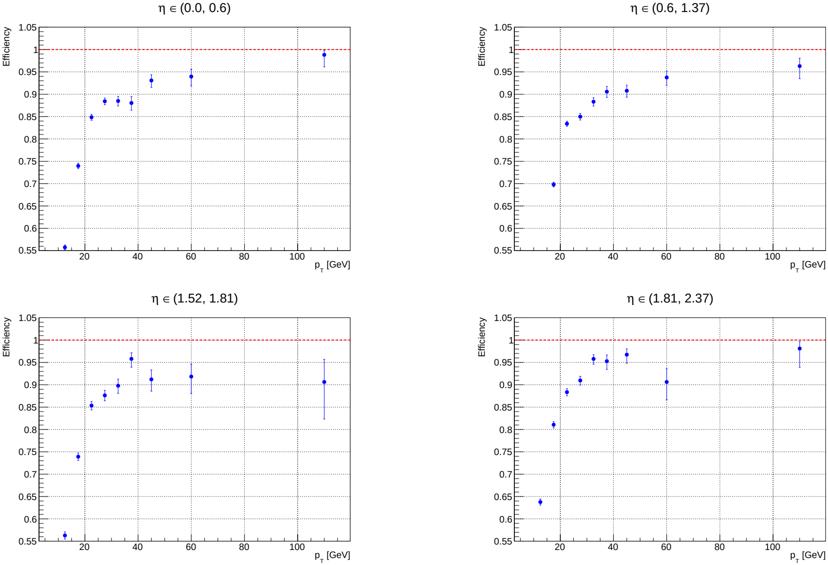




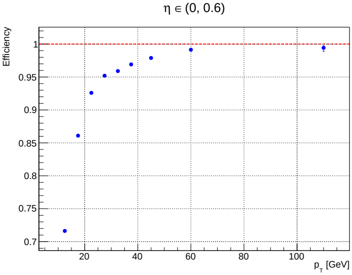


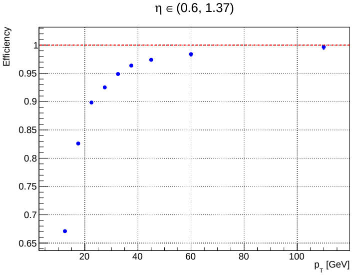
Efficiency on data, Muons

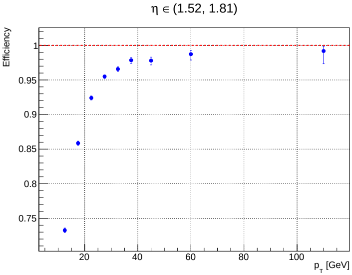
Converted

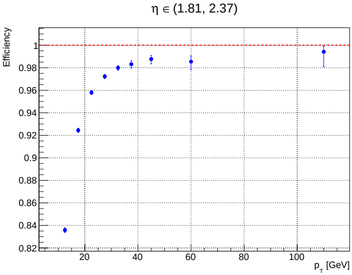


Unconverted



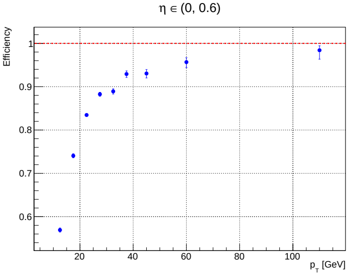


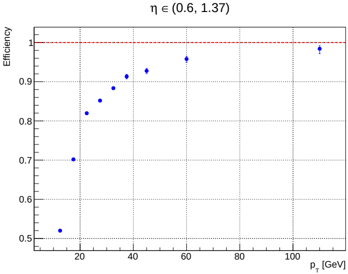


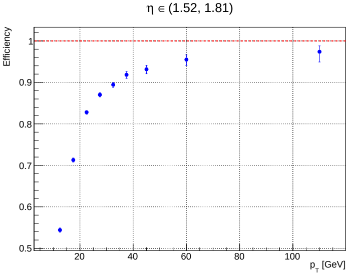


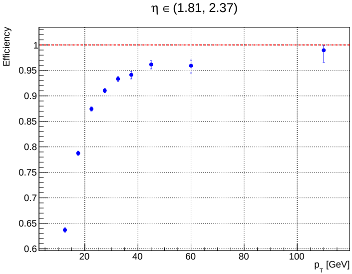
Efficiency on MC, Electrons

Converted



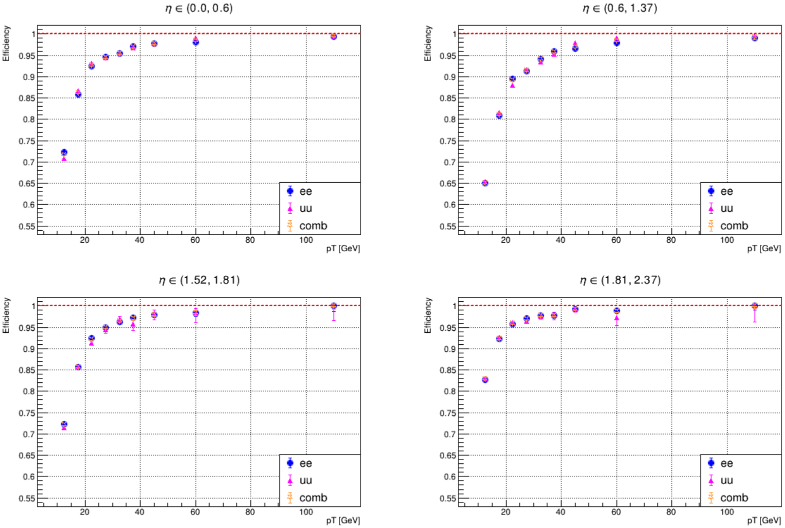






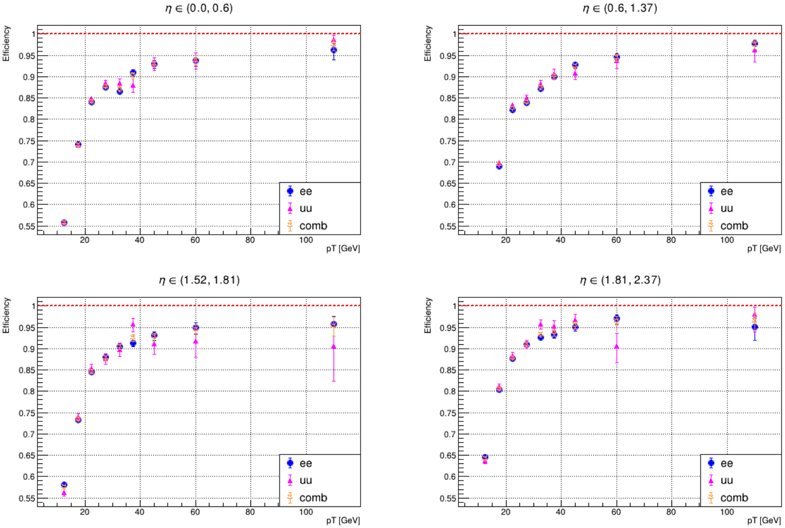
Efficiency on data, combined

Unconverted



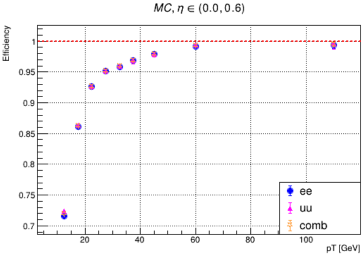
Efficiency on data, combined

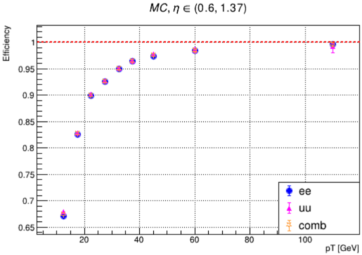
Converted

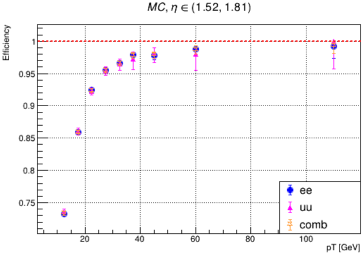


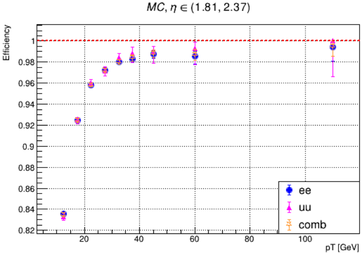
Efficiency on MC, combined

Unconverted



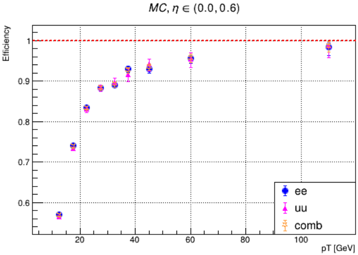


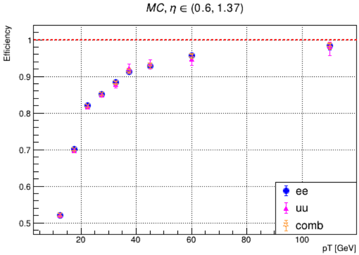


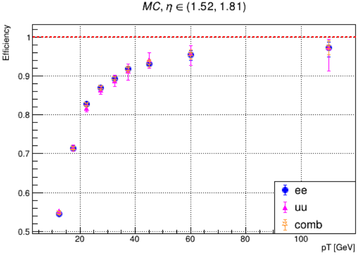


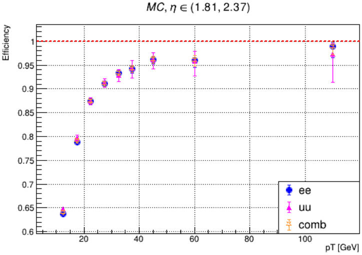
Efficiency on MC, combined

Converted



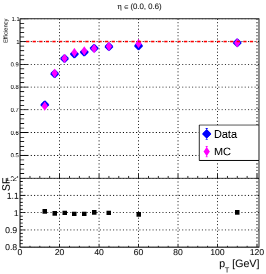


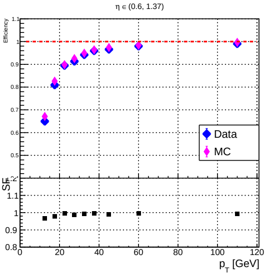


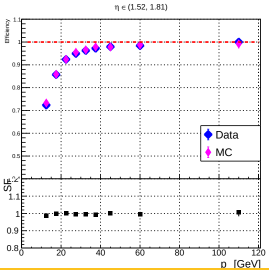


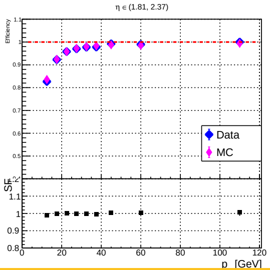
Efficiency with SF, Electrons

Unconverted



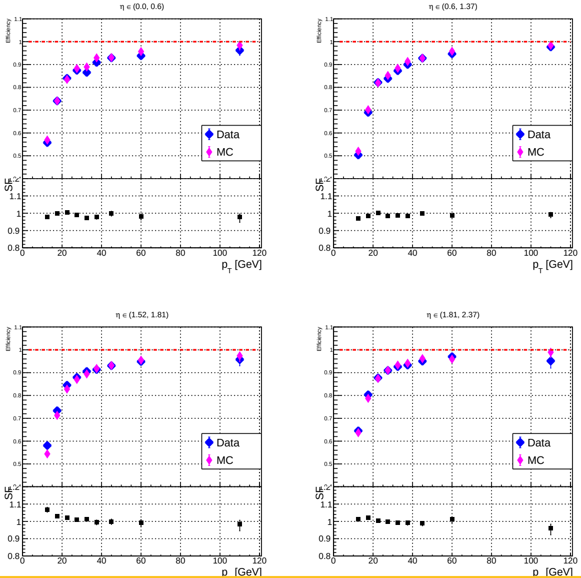




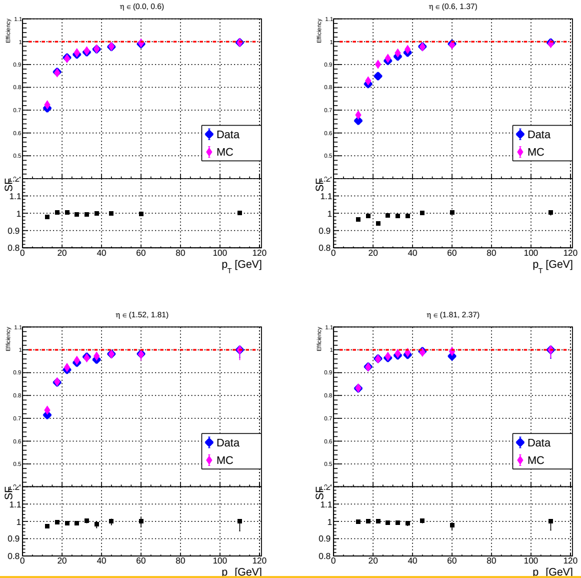


Efficiency with SF, Electrons

Converted

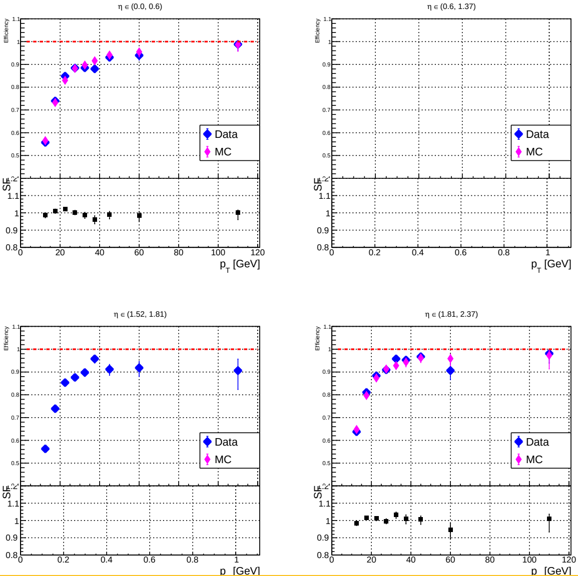


Unconverted



Efficiency with SF, Muons

Converted



Unconverted

combined, unconverted, $\eta \in$ (0,0.6)

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9974	0.0027	0.0027
17.5	1.0000	0.0021	0.0021
22.5	1.0010	0.0018	0.0018
27.5	0.9926	0.0019	0.0019
32.5	0.9944	0.0025	0.0025
37.5	1.0008	0.0032	0.0032
45.0	0.9986	0.0032	0.0032
60.0	0.9922	0.0034	0.0034
110.0	1.0004	0.0051	0.0051

combined, unconverted, $\eta \in$ (0.6,1.37)

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9663	0.0027	0.0027
17.5	0.9808	0.0022	0.0022
22.5	0.9835	0.0021	0.0021
27.5	0.9874	0.0022	0.0022
32.5	0.9909	0.0026	0.0026
37.5	0.9936	0.0033	0.0033
45.0	0.9945	0.0033	0.0033
60.0	0.9990	0.0039	0.0039
110.0	0.9963	0.0052	0.0052

combined, unconverted, $\eta \in$ (1.52,1.81)

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9823	0.0053	0.0053
17.5	0.9978	0.0044	0.0044
22.5	0.9976	0.0039	0.0039
27.5	0.9934	0.0039	0.0039
32.5	0.9978	0.0045	0.0045
37.5	0.9925	0.0059	0.0059
45.0	1.0012	0.0069	0.0069
60.0	0.9972	0.0093	0.0093
110.0	1.0067	0.0068	0.0151

combined, unconverted, $\eta \in$ (1.81,2.37)

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9917	0.0032	0.0032
17.5	0.9994	0.0025	0.0025
22.5	1.0005	0.0024	0.0024
27.5	0.9971	0.0025	0.0025
32.5	0.9965	0.0030	0.0030
37.5	0.9939	0.0044	0.0044
45.0	1.0048	0.0042	0.0042
60.0	1.0009	0.0072	0.0072
110.0	1.0050	0.0051	0.0114

Scale Factors, combined

Converted

combined, converted, $\eta \in$ (0,0.6)

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9820	0.0088	0.0088
17.5	1.0030	0.0072	0.0072
22.5	1.0109	0.0064	0.0064
27.5	0.9937	0.0064	0.0064
32.5	0.9770	0.0085	0.0085
37.5	0.9742	0.0106	0.0106
45.0	0.9954	0.0116	0.0116
60.0	0.9816	0.0148	0.0148
110.0	0.9888	0.0192	0.0192

combined, converted, $\eta \in$ (0.6,1.37)

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9671	0.0075	0.0075
17.5	0.9891	0.0058	0.0058
22.5	1.0077	0.0052	0.0052
27.5	0.9879	0.0055	0.0055
32.5	0.9909	0.0066	0.0066
37.5	0.9847	0.0084	0.0084
45.0	0.9943	0.0089	0.0089
60.0	0.9880	0.0107	0.0107
110.0	0.9906	0.0140	0.0140

combined, converted, $\eta \in$ (1.52,1.81)

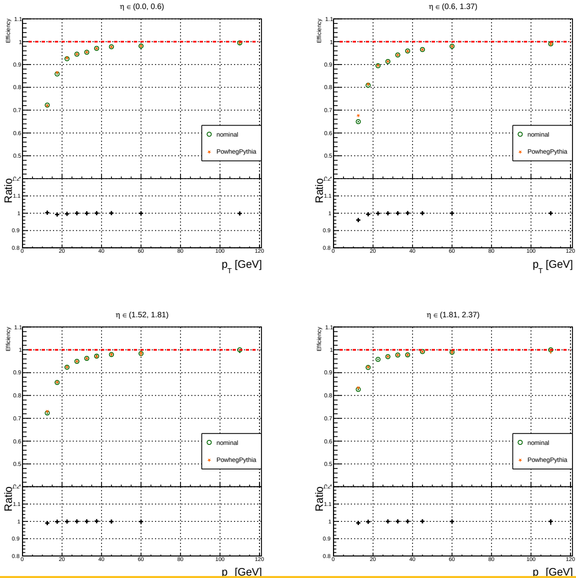
pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	1.0521	0.0114	0.0114
17.5	1.0306	0.0087	0.0087
22.5	1.0262	0.0076	0.0076
27.5	1.0112	0.0079	0.0079
32.5	1.0123	0.0094	0.0094
37.5	1.0073	0.0122	0.0122
45.0	0.9935	0.0138	0.0138
60.0	0.9884	0.0174	0.0174
110.0	0.9779	0.0293	0.0293

combined, converted, $\eta \in$ (1.81,2.37)

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	1.0034	0.0081	0.0081
17.5	1.0190	0.0067	0.0067
22.5	1.0054	0.0061	0.0061
27.5	0.9986	0.0063	0.0063
32.5	1.0008	0.0074	0.0074
37.5	0.9953	0.0109	0.0109
45.0	0.9923	0.0111	0.0111
60.0	1.0056	0.0162	0.0162
110.0	0.9762	0.0242	0.0263

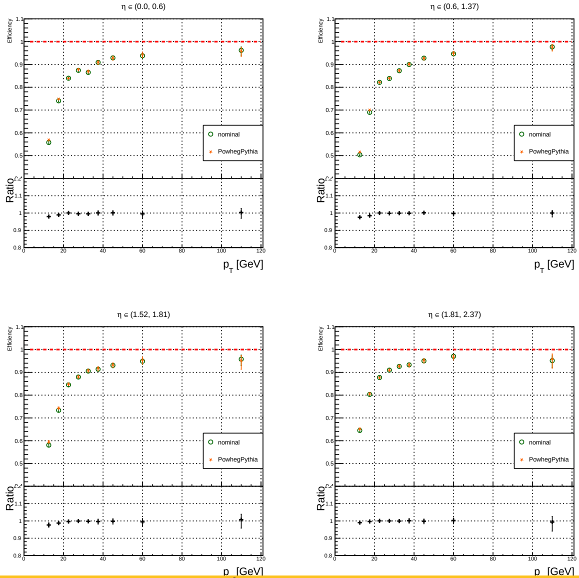
Systematic, Electrons

Unconverted



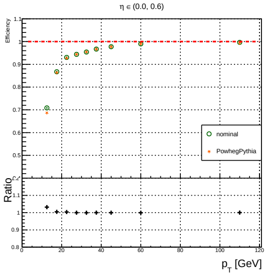
Systematic, Electrons

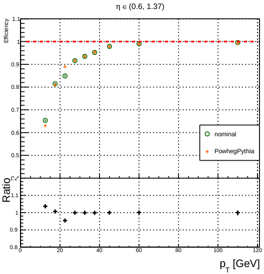
Converted

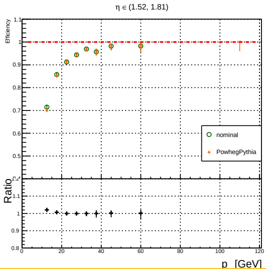


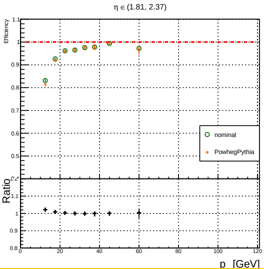
Systematic, Muons

Unconverted



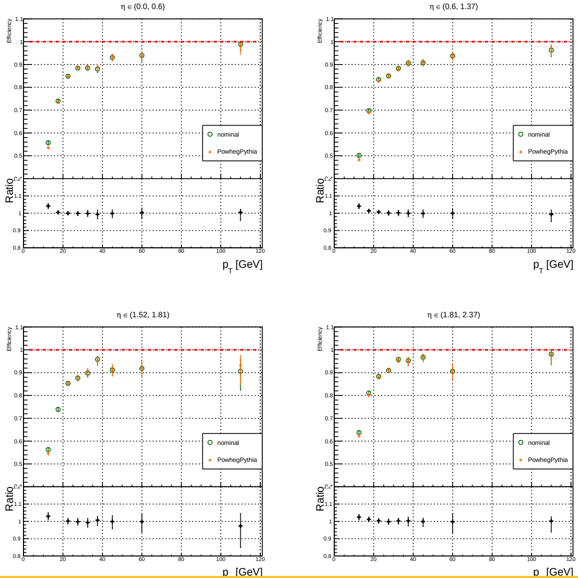






Systematic, Muons

Converted



Scale Factors, combined, with systematic

Unconverted

comb, unconverted, $\eta \in$ (0.0, 0.6)

pT [GeV]	SF	Stat.Error	Gen.Error	Fit.Error	Bkg.Error	Total.Error
12.5	0.9974	0.0027	0.0134	0.0029	0.0090	0.0166
17.5	1.0000	0.0021	0.0030	0.0006	0.0254	0.0257
22.5	1.0010	0.0018	0.0013	0.0024	0.0000	0.0033
27.5	0.9926	0.0019	0.0001	0.0007	0.0000	0.0021
32.5	0.9944	0.0025	0.0004	0.0002	0.0000	0.0025
37.5	1.0008	0.0032	0.0006	0.0275	0.0000	0.0276
45.0	0.9986	0.0032	0.0012	0.0004	0.0000	0.0034
60.0	0.9922	0.0034	0.0011	0.0023	0.0000	0.0043
110.0	1.0004	0.0051	0.0000	0.0005	0.0000	0.0051

comb, unconverted, $\eta \in$ (0.6, 1.37)

pT [GeV]	SF	Stat.Error	Gen.Error	Fit.Error	Bkg.Error	Total.Error
12.5	0.9663	0.0027	0.0076	0.0153	0.0046	0.0179
17.5	0.9808	0.0022	0.0018	0.0080	0.0200	0.0217
22.5	0.9835	0.0021	0.0107	0.0122	0.0000	0.0164
27.5	0.9874	0.0022	0.0003	0.0012	0.0000	0.0025
32.5	0.9909	0.0026	0.0007	0.0019	0.0000	0.0033
37.5	0.9936	0.0033	0.0007	0.0017	0.0000	0.0038
45.0	0.9945	0.0033	0.0002	0.0003	0.0000	0.0033
60.0	0.9990	0.0039	0.0009	0.0018	0.0000	0.0044
110.0	0.9963	0.0052	0.0010	0.0004	0.0000	0.0053

comb, unconverted, $\eta \in$ (1.52, 1.81)

pT [GeV]	SF	Stat.Error	Gen.Error	Fit.Error	Bkg.Error	Total.Error
12.5	0.9823	0.0053	0.0005	0.0094	0.0326	0.0343
17.5	0.9978	0.0044	0.0005	0.0041	0.0077	0.0097
22.5	0.9976	0.0039	0.0003	0.0043	0.0000	0.0058
27.5	0.9934	0.0039	0.0000	0.0503	0.0000	0.0504
32.5	0.9978	0.0045	0.0005	0.0022	0.0000	0.0051
37.5	0.9925	0.0059	0.0024	0.0025	0.0000	0.0068
45.0	1.0012	0.0069	0.0035	0.0347	0.0000	0.0356
60.0	0.9972	0.0093	0.0004	0.0003	0.0000	0.0093
110.0	1.0067	0.0109	0.0000	0.0012	0.0000	0.0110

comb, unconverted, $\eta \in$ (1.81, 2.37)

pT [GeV]	SF	Stat.Error	Gen.Error	Fit.Error	Bkg.Error	Total.Error
12.5	0.9917	0.0032	0.0001	0.0010	0.0236	0.0238
17.5	0.9994	0.0025	0.0005	0.0001	0.0062	0.0067
22.5	1.0005	0.0024	0.0089	0.0006	0.0000	0.0092
27.5	0.9971	0.0025	0.0001	0.0002	0.0000	0.0026
32.5	0.9965	0.0030	0.0009	0.0005	0.0000	0.0032
37.5	0.9939	0.0044	0.0008	0.0005	0.0000	0.0045
45.0	1.0048	0.0042	0.0006	0.0009	0.0000	0.0044
60.0	1.0009	0.0072	0.0009	0.0003	0.0000	0.0073
110.0	1.0050	0.0082	0.0000	0.0009	0.0000	0.0083

Scale Factors, combined, with systematic

Converted

comb, converted, $\eta \in$ (0.0, 0.6)

pT [GeV]	SF	Stat.Error	Gen.Error	Fit.Error	Bkg.Error	Total.Error
12.5	0.5576	0.0034	0.0023	0.0060	0.0361	0.0368
17.5	0.7401	0.0036	0.0040	0.0034	0.0750	0.0753
22.5	0.8424	0.0037	0.0002	0.0052	0.0000	0.0064
27.5	0.8769	0.0040	0.0014	0.0009	0.0000	0.0043
32.5	0.8704	0.0055	0.0015	0.0022	0.0000	0.0062
37.5	0.9027	0.0072	0.0014	0.0072	0.0000	0.0103
45.0	0.9295	0.0073	0.0003	0.0017	0.0000	0.0075
60.0	0.9386	0.0103	0.0044	0.0083	0.0000	0.0139
110.0	0.9741	0.0138	0.0044	0.0019	0.0000	0.0146

comb, converted, $\eta \in$ (0.6, 1.37)

pT [GeV]	SF	Stat.Error	Gen.Error	Fit.Error	Bkg.Error	Total.Error
12.5	0.5029	0.0028	0.0002	0.0051	0.0136	0.0148
17.5	0.6925	0.0030	0.0040	0.0032	0.0045	0.0075
22.5	0.8250	0.0031	0.0018	0.0060	0.0000	0.0070
27.5	0.8410	0.0035	0.0019	0.0016	0.0000	0.0043
32.5	0.8747	0.0043	0.0028	0.0034	0.0000	0.0061
37.5	0.9011	0.0057	0.0005	0.0025	0.0000	0.0062
45.0	0.9240	0.0058	0.0006	0.0022	0.0000	0.0062
60.0	0.9445	0.0076	0.0020	0.0001	0.0000	0.0078
110.0	0.9742	0.0104	0.0029	0.0049	0.0000	0.0118

comb, converted, $\eta \in$ (1.52, 1.81)

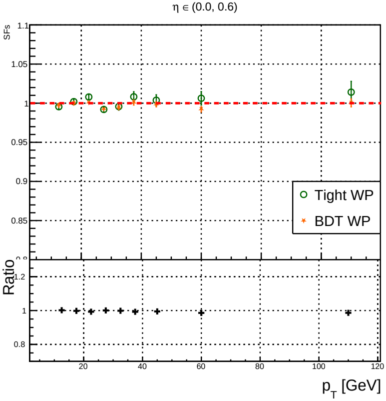
pT [GeV]	SF	Stat.Error	Gen.Error	Fit.Error	Bkg.Error	Total.Error
12.5	0.5752	0.0045	0.0040	0.0061	0.0095	0.0128
17.5	0.7352	0.0045	0.0073	0.0147	0.0150	0.0227
22.5	0.8471	0.0046	0.0005	0.0139	0.0000	0.0147
27.5	0.8789	0.0050	0.0001	0.0056	0.0000	0.0075
32.5	0.9045	0.0061	0.0016	0.0066	0.0000	0.0091
37.5	0.9243	0.0080	0.0027	0.0108	0.0000	0.0137
45.0	0.9277	0.0092	0.0035	0.0068	0.0000	0.0120
60.0	0.9447	0.0120	0.0038	0.0047	0.0000	0.0135
110.0	0.9523	0.0221	0.0013	0.0092	0.0000	0.0240

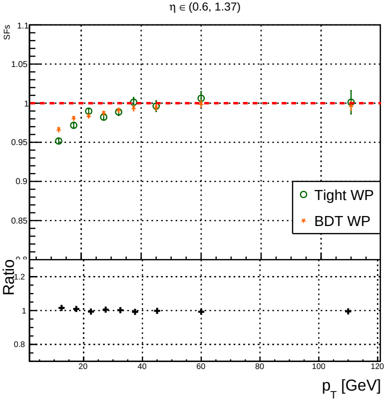
comb, converted, $\eta \in$ (1.81, 2.37)

pT [GeV]	SF	Stat.Error	Gen.Error	Fit.Error	Bkg.Error	Total.Error
12.5	0.6427	0.0039	0.0008	0.0004	0.0116	0.0190
17.5	0.8054	0.0039	0.0010	0.0054	0.0084	0.0133
22.5	0.8789	0.0040	0.0077	0.0012	0.0000	0.0099
27.5	0.9097	0.0042	0.0012	0.0015	0.0000	0.0051
32.5	0.9333	0.0051	0.0015	0.0064	0.0000	0.0089
37.5	0.9372	0.0075	0.0012	0.0047	0.0000	0.0095
45.0	0.9542	0.0078	0.0016	0.0014	0.0000	0.0085
60.0	0.9646	0.0104	0.0016	0.0084	0.0000	0.0140
110.0	0.9638	0.0204	0.0066	0.0091	0.0000	0.0242

Comparison with the Tight WP

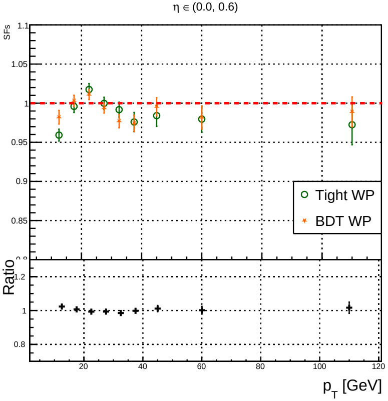
Example: combined channel, unconverted

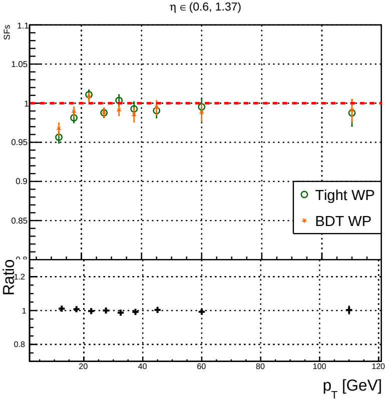




Comparison with the Tight WP

Example: combined channel, converted

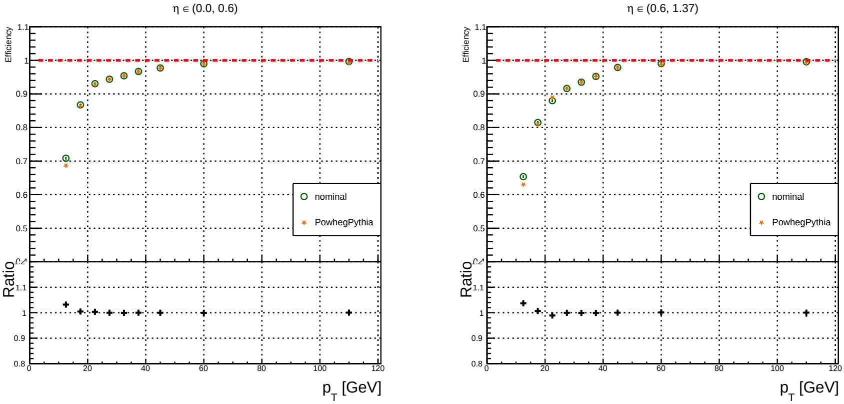




Monte Carlo generators systematic
→ re-calculate the SFs on samples generated with different generators:

- mc16*.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee.*.root
- mc16*.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zmumu.*.root

Example: Muonic channel, unconverted



Scale Factors, combined

Unconverted

combined, unconverted, $\eta \in (0,0.6)$

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9974	0.0027	0.0027
17.5	1.0000	0.0021	0.0021
22.5	1.0010	0.0018	0.0018
27.5	0.9926	0.0019	0.0019
32.5	0.9944	0.0025	0.0025
37.5	1.0008	0.0032	0.0032
45.0	0.9986	0.0032	0.0032
60.0	0.9922	0.0034	0.0034
110.0	1.0004	0.0051	0.0051

combined, unconverted, $\eta \in (0.6,1.37)$

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9663	0.0027	0.0027
17.5	0.9808	0.0022	0.0022
22.5	0.9835	0.0021	0.0021
27.5	0.9874	0.0022	0.0022
32.5	0.9909	0.0026	0.0026
37.5	0.9936	0.0033	0.0033
45.0	0.9945	0.0033	0.0033
60.0	0.9990	0.0039	0.0039
110.0	0.9963	0.0052	0.0052

combined, unconverted, $\eta \in (1.52,1.81)$

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9823	0.0053	0.0053
17.5	0.9978	0.0044	0.0044
22.5	0.9976	0.0039	0.0039
27.5	0.9934	0.0039	0.0039
32.5	0.9978	0.0045	0.0045
37.5	0.9925	0.0059	0.0059
45.0	1.0012	0.0069	0.0069
60.0	0.9972	0.0093	0.0093
110.0	1.0067	0.0068	0.0151

combined, unconverted, $\eta \in (1.81,2.37)$

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9917	0.0032	0.0032
17.5	0.9994	0.0025	0.0025
22.5	1.0005	0.0024	0.0024
27.5	0.9971	0.0025	0.0025
32.5	0.9965	0.0030	0.0030
37.5	0.9939	0.0044	0.0044
45.0	1.0048	0.0042	0.0042
60.0	1.0009	0.0072	0.0072
110.0	1.0050	0.0051	0.0114

Scale Factors, combined

Converted

combined, converted, $\eta \in$ (0,0.6)

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9820	0.0088	0.0088
17.5	1.0030	0.0072	0.0072
22.5	1.0109	0.0064	0.0064
27.5	0.9937	0.0064	0.0064
32.5	0.9770	0.0085	0.0085
37.5	0.9742	0.0106	0.0106
45.0	0.9954	0.0116	0.0116
60.0	0.9816	0.0148	0.0148
110.0	0.9888	0.0192	0.0192

combined, converted, $\eta \in$ (0.6,1.37)

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	0.9671	0.0075	0.0075
17.5	0.9891	0.0058	0.0058
22.5	1.0077	0.0052	0.0052
27.5	0.9879	0.0055	0.0055
32.5	0.9909	0.0066	0.0066
37.5	0.9847	0.0084	0.0084
45.0	0.9943	0.0089	0.0089
60.0	0.9880	0.0107	0.0107
110.0	0.9906	0.0140	0.0140

combined, converted, $\eta \in$ (1.52,1.81)

pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	1.0521	0.0114	0.0114
17.5	1.0306	0.0087	0.0087
22.5	1.0262	0.0076	0.0076
27.5	1.0112	0.0079	0.0079
32.5	1.0123	0.0094	0.0094
37.5	1.0073	0.0122	0.0122
45.0	0.9935	0.0138	0.0138
60.0	0.9884	0.0174	0.0174
110.0	0.9779	0.0293	0.0293

combined, converted, $\eta \in$ (1.81,2.37)

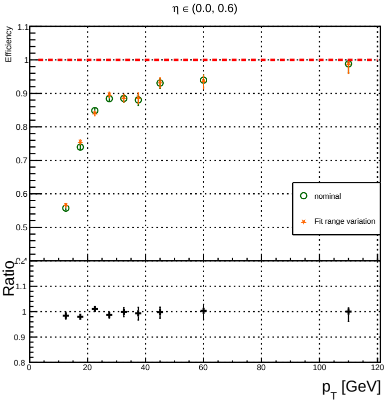
pT [GeV]	SF	Stat.Error up	Stat.Error down
12.5	1.0034	0.0081	0.0081
17.5	1.0190	0.0067	0.0067
22.5	1.0054	0.0061	0.0061
27.5	0.9986	0.0063	0.0063
32.5	1.0008	0.0074	0.0074
37.5	0.9953	0.0109	0.0109
45.0	0.9923	0.0111	0.0111
60.0	1.0056	0.0162	0.0162
110.0	0.9762	0.0242	0.0263

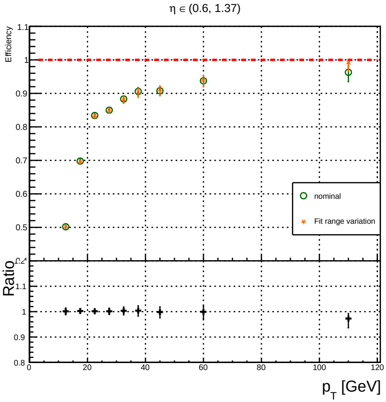
Systematic uncertainties on SFs: fitting procedure

fitting procedure systematic

→ re-calculate the SFs varying the fit-range: $[55, 105]$GeV → [65,105]GeV

Example: Muonic channel, converted





background modelling systematic
→ re-calculate the SFs varying the background template: Zjets instead of data in CR

Example: Muonic channel, unconverted