egamma performance and photons

Marcello Fanti
University of Milano and INFN

- Energy calibration for electrons and photons
- Evaluation of photon purity
- What we expect/plan from early data?

Some preliminary remarks...

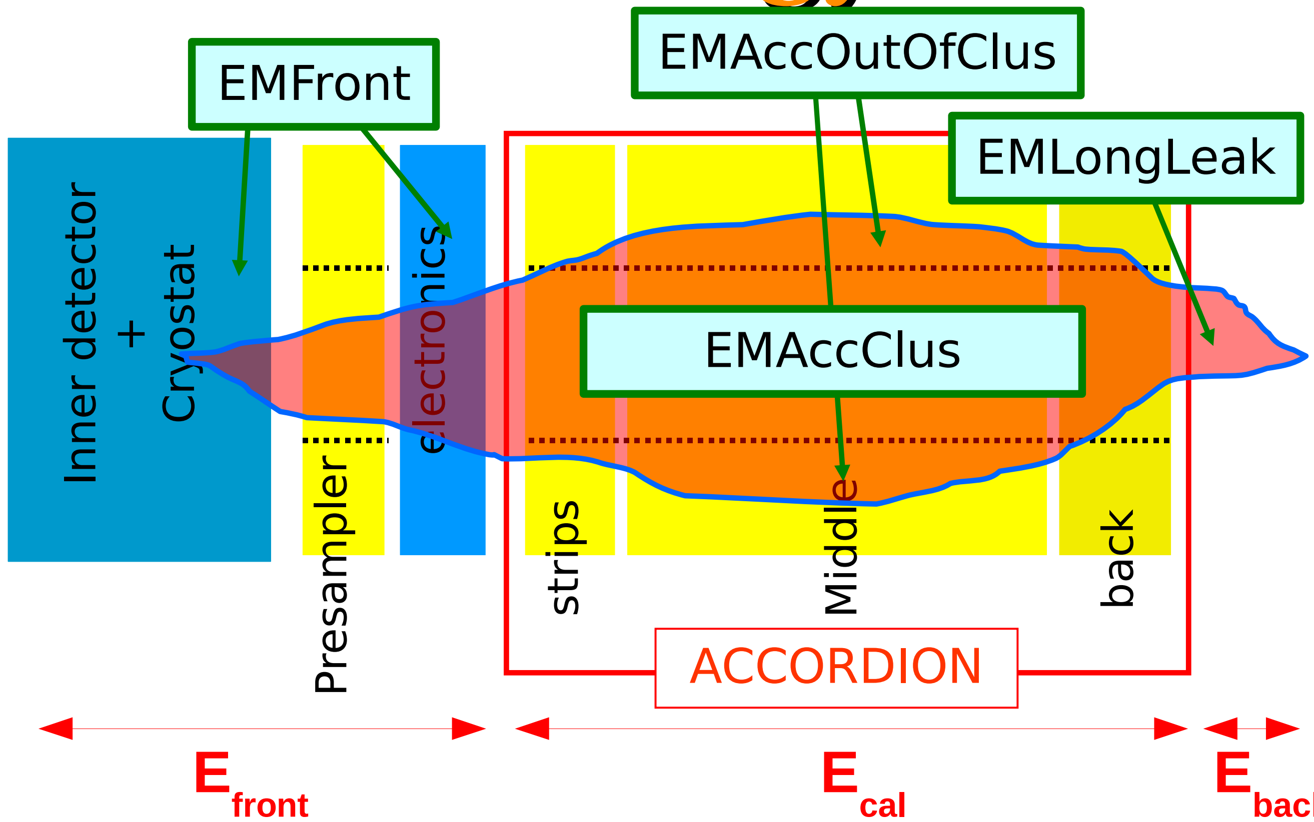
- Focus on photon physics (inclusive, di-photons) and Higgs search in $H\to\gamma\gamma$ decay channel (but for the latter, no discovery/exclusion can be assessed with ~ 200 pb^{-1} ... please refer to talk by Daniela Rebuzzi)
- Some QCD photon physics can be carried out with first year of data...
  - Need to evaluate the purity ***P*** and the efficiency ε of the sample for cross-section evaluation: $\sigma = (N * P) / (\varepsilon * \int L\, dt)$.
  - Purity discussed in ***this talk***;
  - Some info about efficiency later on...
- Also need energy calibration! (also in ***this talk***)
  - Related to photon physics, but of much more general interest
- CAVEAT: We don't have “candles”: $\pi^0\to\gamma\gamma$ and $\eta\to\gamma\gamma$ have low pt and require special photon reconstruction (topo-cluster based), cannot use tag&probe etc ...

e/gamma calibration

D. Banfi
L. Carminati
L. Mandelli

(related to photon physics, but of course of much more general interest)

Energy calibration



Calibration recipe

$$E_{corr} = E_{front} + E_{cal} + E_{back}$$

$$E_{cal} = C_{cal}(X,\eta)\,[1+f_{out}(X,\eta)]\,E_{acc}$$

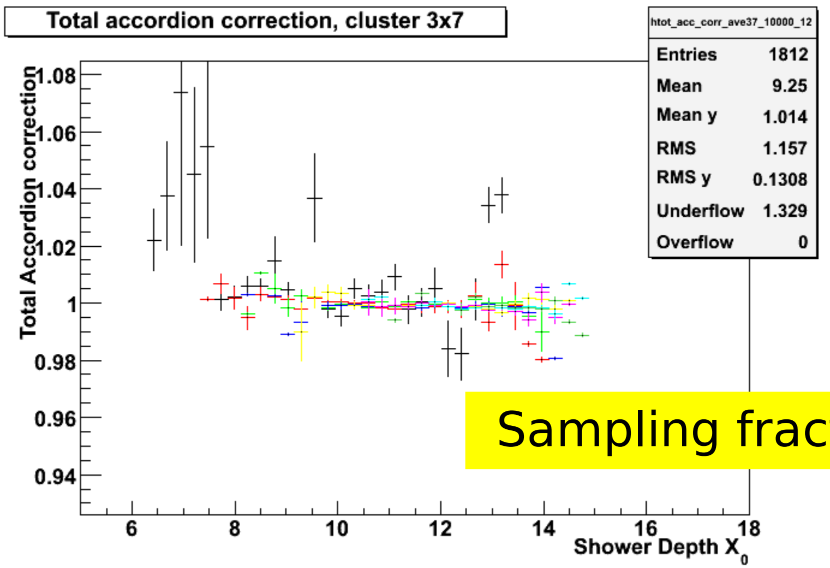
$$f_{leak} = E_{back} / E_{cal}$$
$$= f^{0}_{leak}(\eta)\,X + f^{1}_{leak}(\eta)\,e^{X}$$

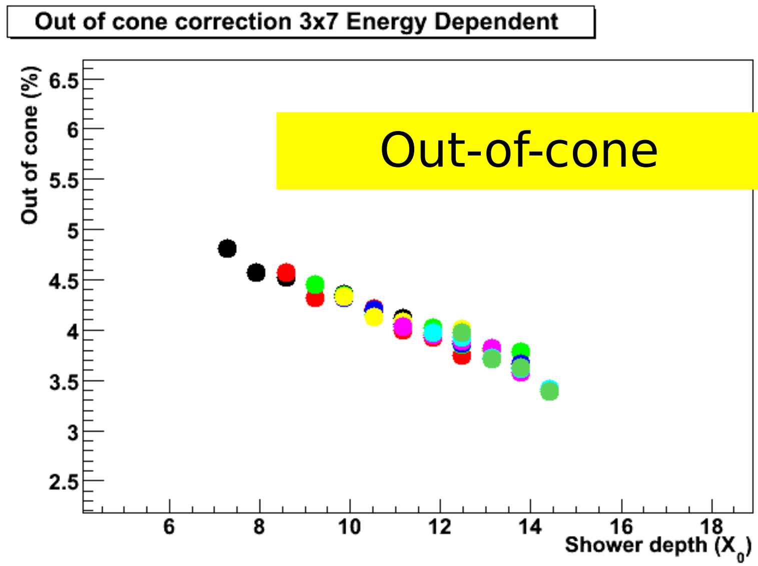
$$E_{front} = a(E_{cal},\eta) + b(E_{cal},\eta)\,E_{PS} + c(E_{cal},\eta)\,E_{PS}^{2}$$

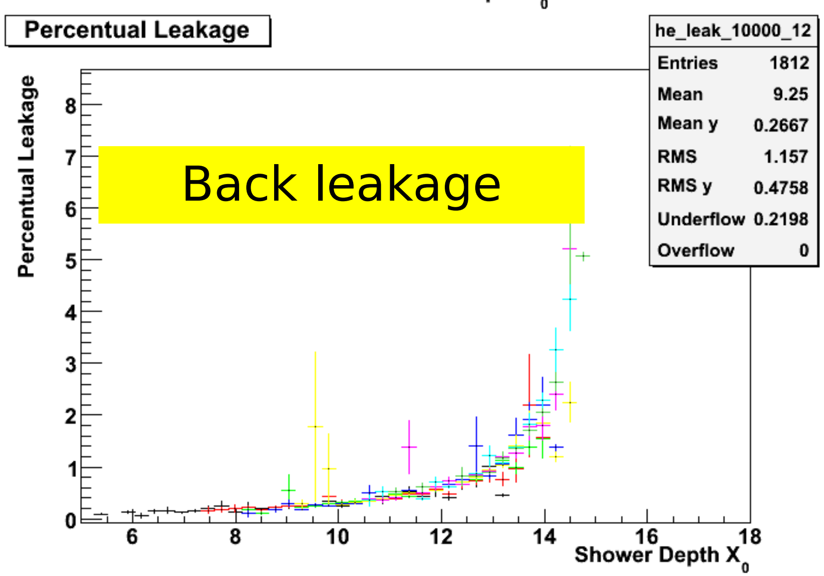
- MonteCarlo-based: use “calibration hits” (special runs where energy hits in dead/inactive material are saved)
- All corrections are factorized, and correlated to observable deposits in active layers and to material description
- Corrections in/beyond accordion are energy-independent if parametrized vs ***shower depth X*** (longitudinal barycentre from PS and 3 accordion layers)

Some examples of corrections...

Colors mark different energies

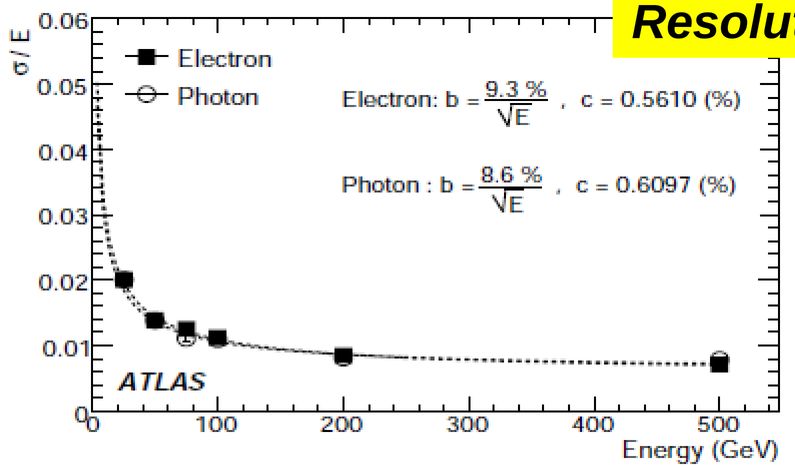




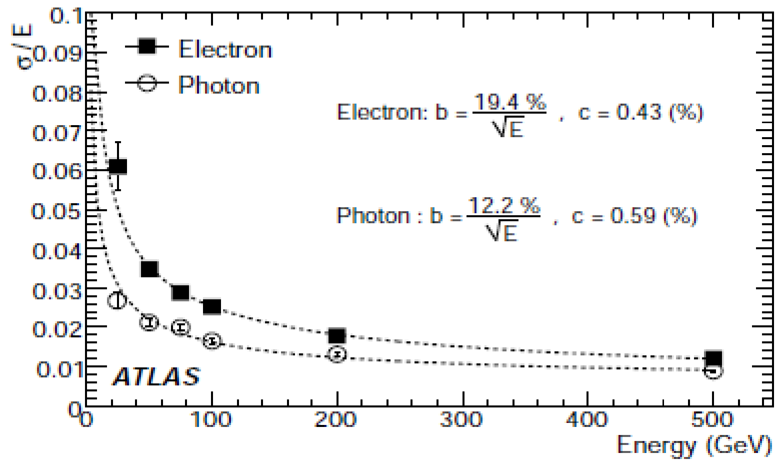


Performance

Resolution

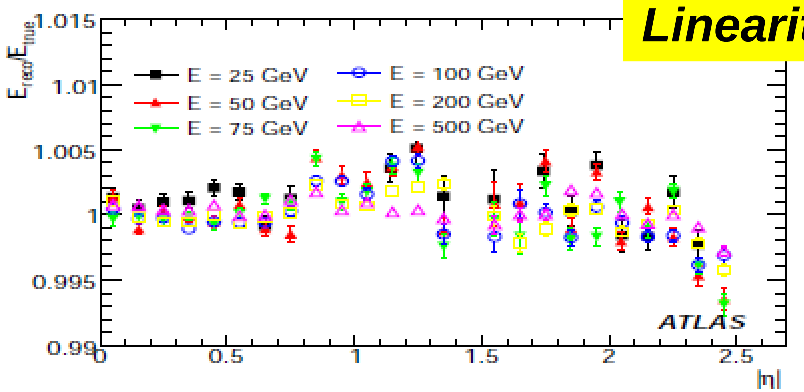


(a) $|\eta| = 0.3$.

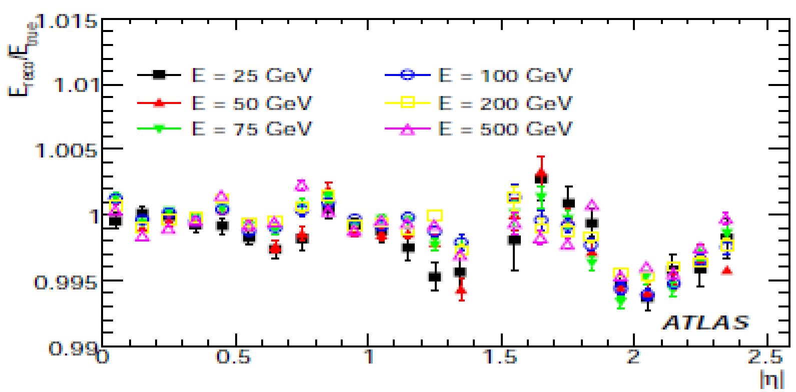


(b) $|\eta| = 1.65$.

Linearity



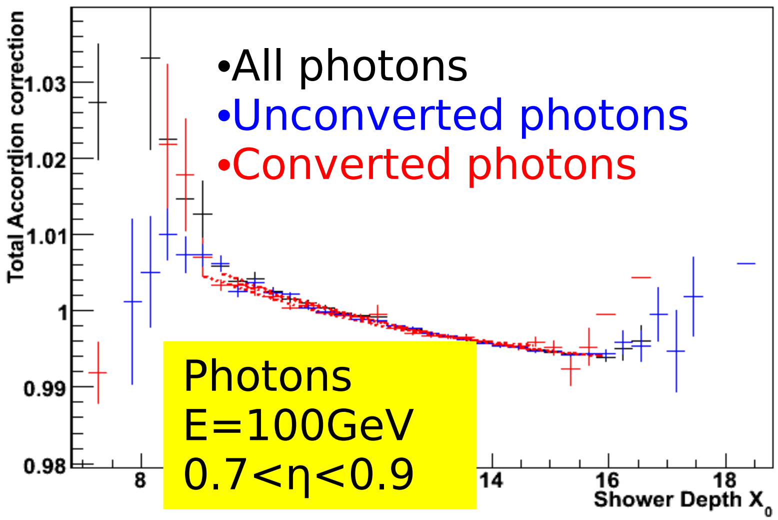
(a) Electrons.



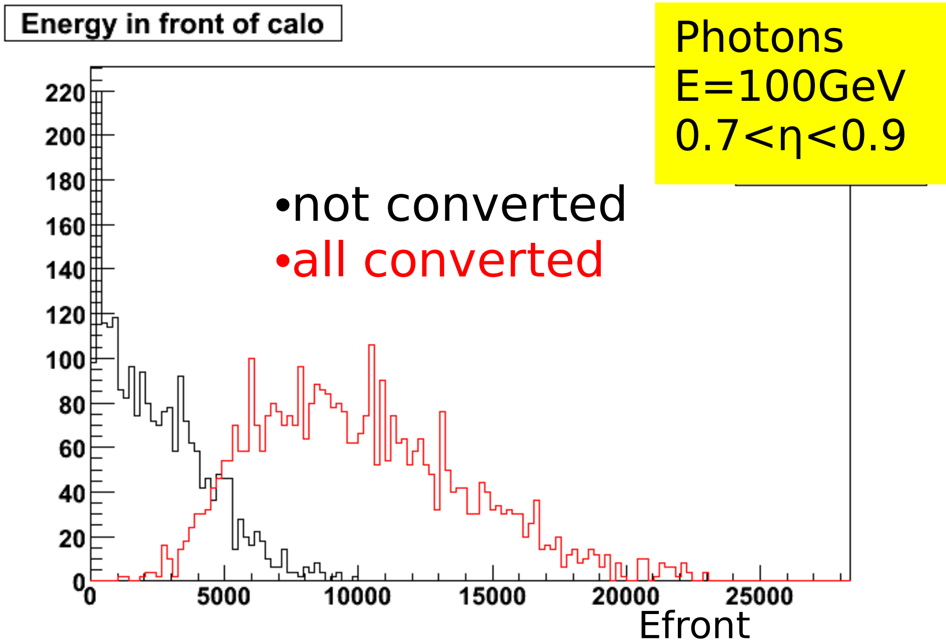
(b) Photons.

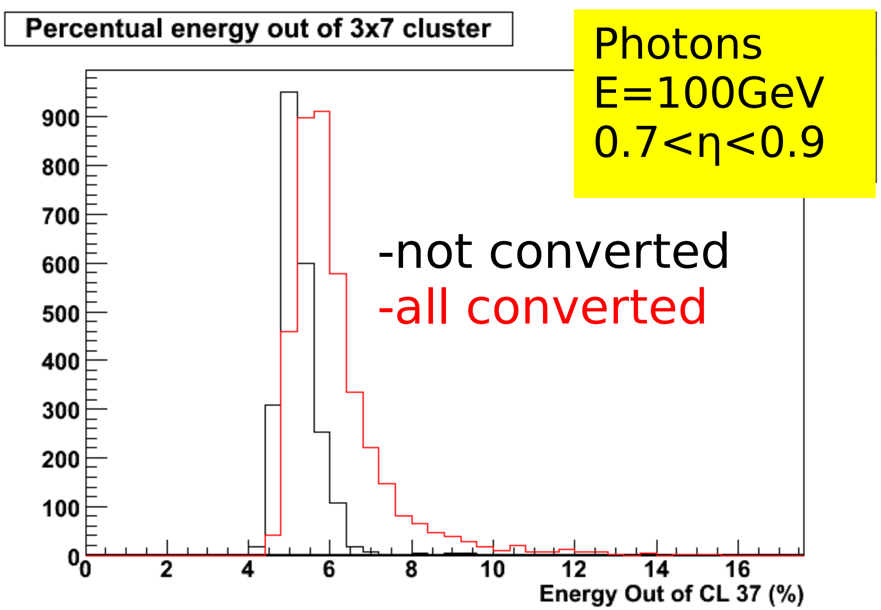
Converted photons

- Large probability (~30%) of photon conversion in ID cavity
- No effect on sampling fraction in accordion...

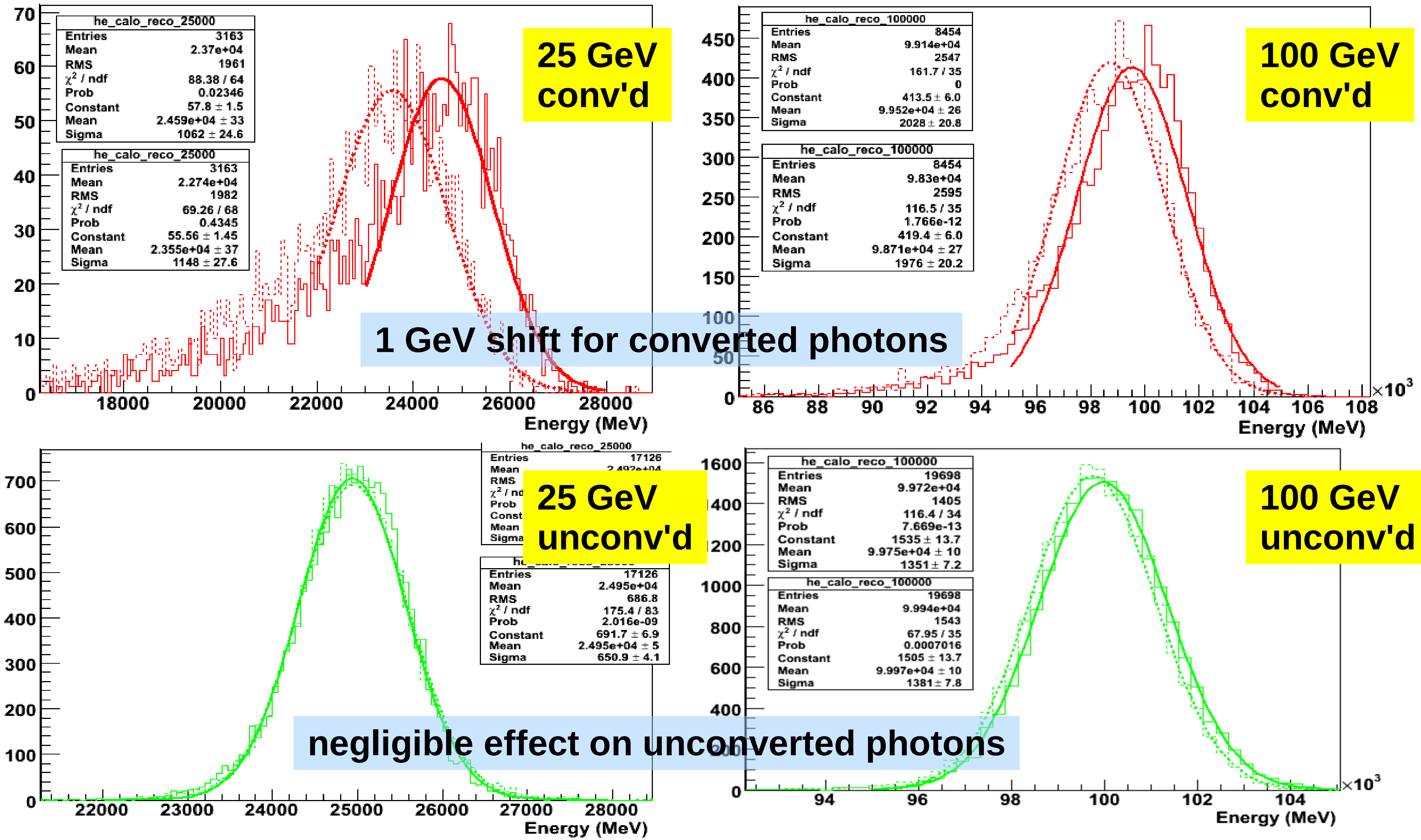


- ... but larger energy deposited in front of EmCalo and out-of-cluster
  - Need dedicated calibration, derived for truly-converted photons and applied to photons with detected conversion

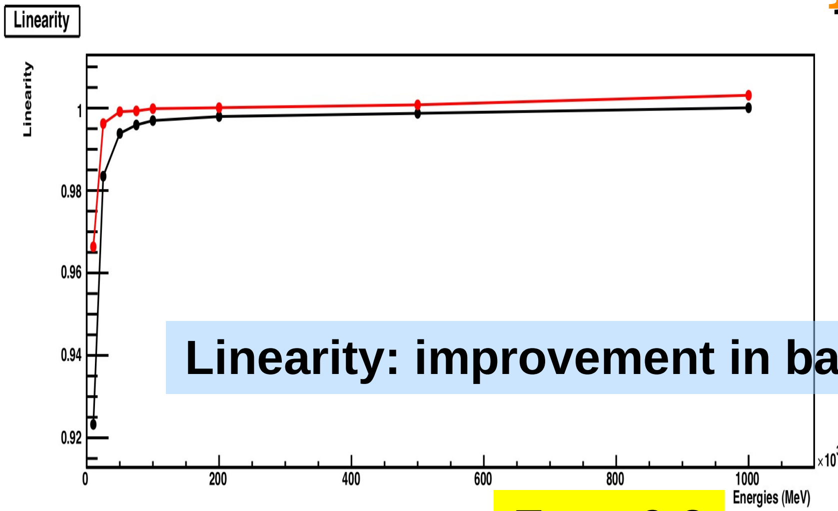


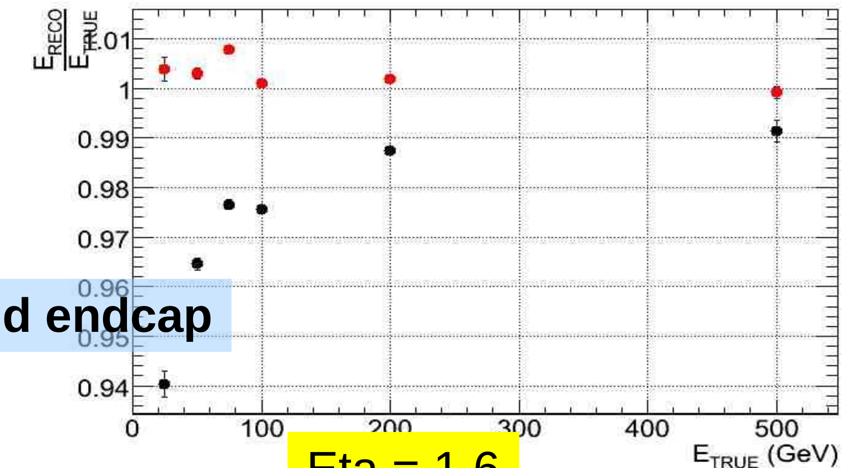


Converted photons (cont'd)



Converted photons (cont'd)

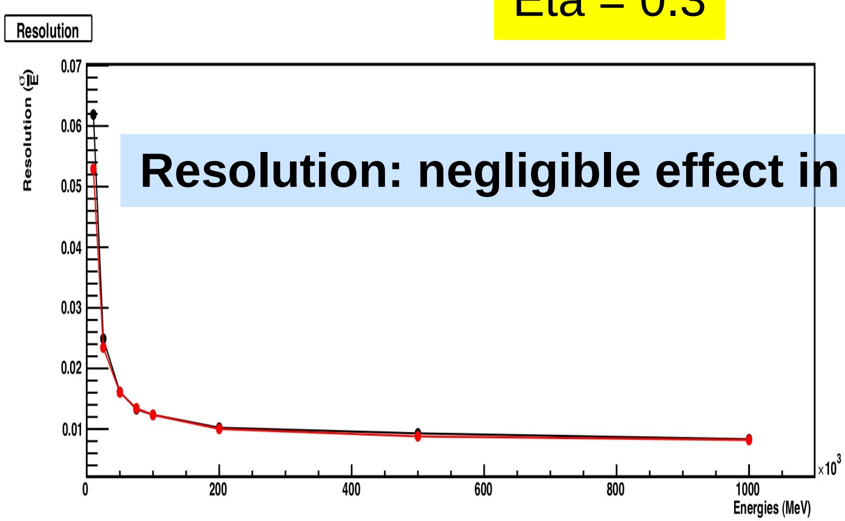


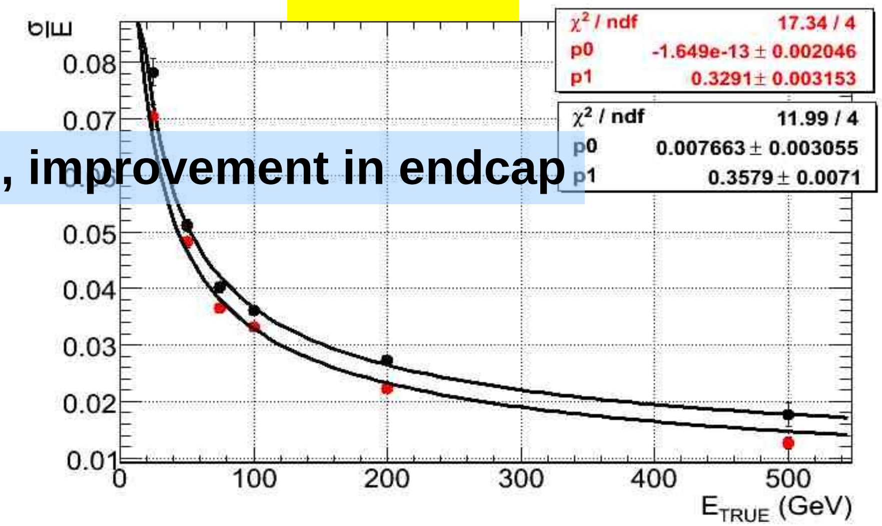


Linearity: improvement in barrel and endcap

Eta = 0.3

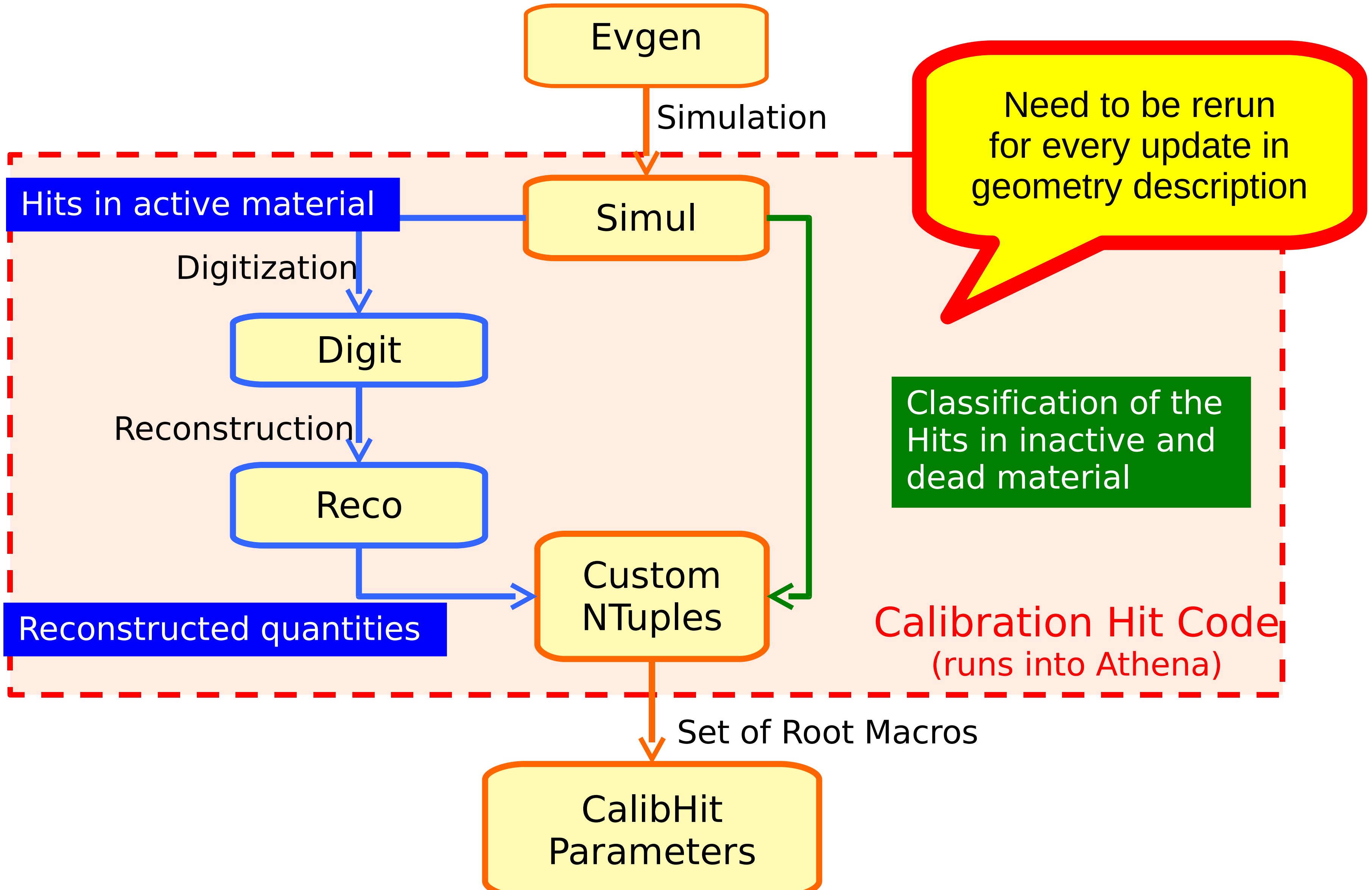
Eta = 1.6





Resolution: negligible effect in barrel, improvement in endcap

Calibration flow chart



Understanding material

- Not our “direct responsibility”, but strictly related to energy calibration

ID based methods

Number and radius of γ-conversions and bremsstrahlung

tools to X-ray material in 3D

... but efficient for small radii only

ID & calo

E/p of high Pt electrons

EM calo based methods
Thanks to its high lateral and longitudinal granularity

Energy flow of minimum bias events

&

Electron shower shape

Complementary to ID-based:
less detailed in x,y,z but more sensitive to outer radii

- First months of data taking will tell us about material distributions: if too far from hypothesized, new full-simulations with calibration hits will be necessary.

Photon purity

D. Banfi
L. Carminati
C. Costa
M. Fanti
L. Mandelli
F. Tartarelli

Definitions

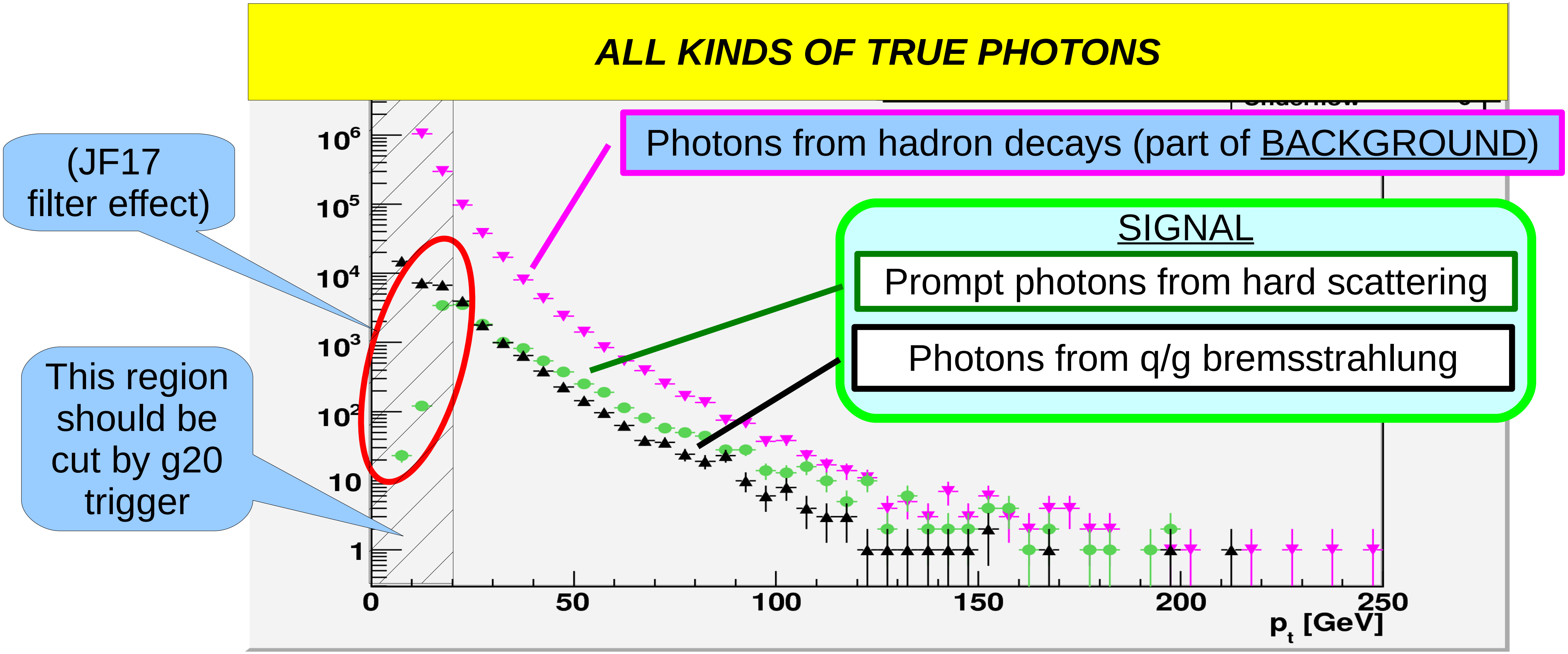
- Signal/background definitions:
  - ***Signal***: photons from **hard scattering** ("prompt") and from **fragmentation** (quark/gluon bremstrahlung) [distinction being actually generator-dependent, thus unphysical!]
  - ***Background***: the rest (secondary photons from hadron decays and other calorimetric clusters not coming from photons)
- Signal/background truth-based classification
  - achieved through cone matching (ΔR) between reconstructed candidates and truth particles
- Purity can be defined as:
  - $\mathbf{P = (\varepsilon_{\gamma} * \sigma_{\gamma}) / (\varepsilon_{\gamma} * \sigma_{\gamma} + \varepsilon_{Bkg} * \sigma_{Bkg})}$ ...
  - ... but cannot be directly worked out from MC truth, unless you blindly trust efficiencies and cross-section ratios: critical because QCD is heavily involved (perturbative order in pQCD, fragmentation models...)

Main issues

- All methods exploit measurable variables that behave differently for signal and background (*"discriminating variables"*)
  - Need to know how signal/background behave wrt discriminating variables
  - But... can we trust simulation? See later on...
- First data will have lots of things to understand...
  - Need more independent methods, use many variables, cross-validate methods with each other
- Performance may change a lot with analysed data set
  - Use a MC data set which realistically represents what we expect from first data (correct mixtures of processes, to have a signal/background ratio near to the real one...)

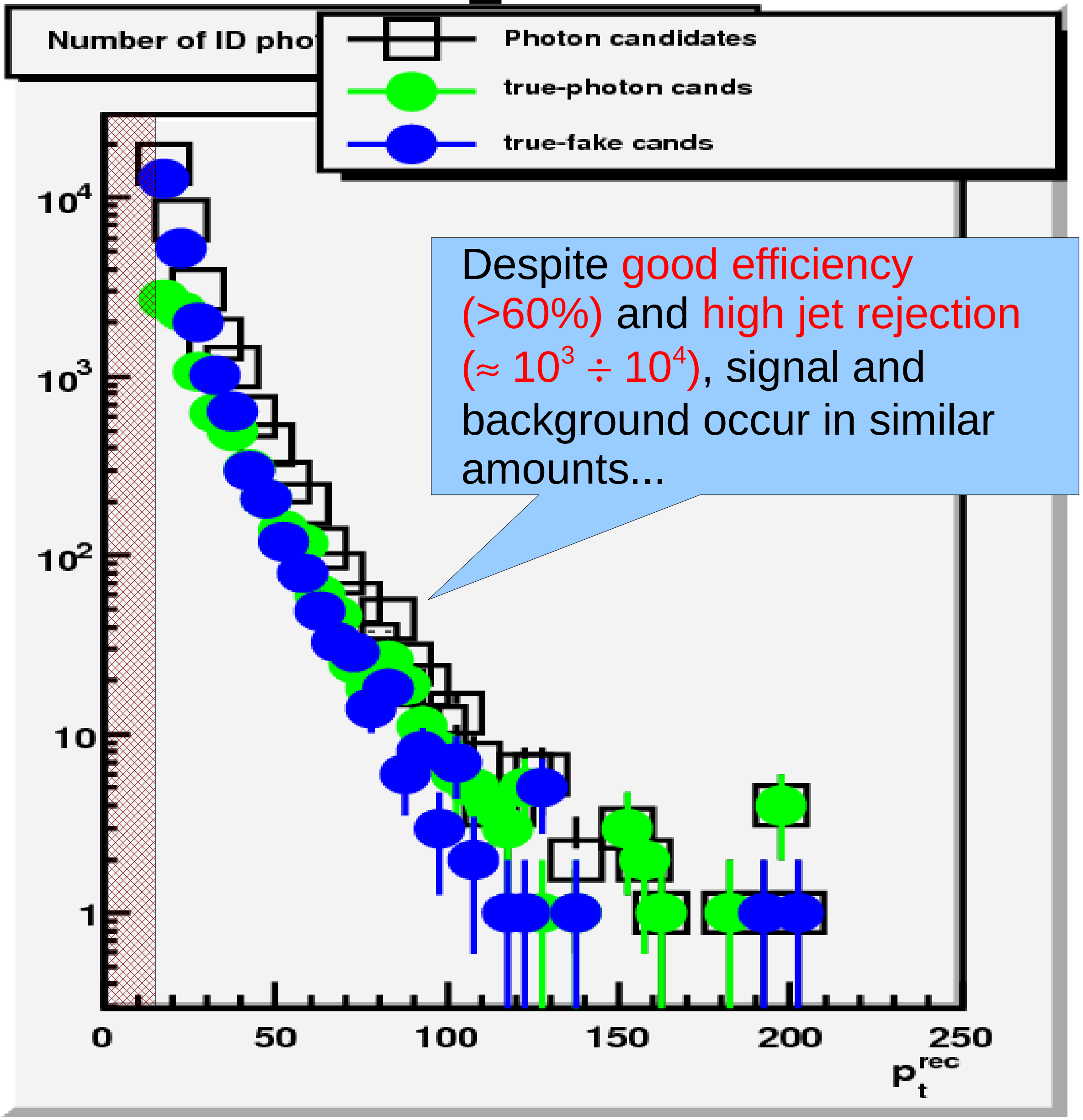
Data set

- Data set: **pythia JF17 + JF35** (mc08.105802 + mc08.105807) :
  - ~ 9.2 M (~0.1 pb^{-1}) fully-simulated events with ***di-jet-like*** topology, different processes are balanced according to their cross-sections
  - Mostly di-jets, plus ~12600 gamma-jet, and ~1900 other processes
  - Dominating processes at LHC ⇒ **realistic sample**



Candidates' sample

- Photon Identification:
  - **PhotonAODCollection** + conversion recovery from ElectronAODCollection
  - **IsEM(PhotonTight)**
    - Little had leakage
    - Narrow shower width in Middle and Strips
    - No 2nd maximum in strips (γ / π^0 separation)
  - **Track isolation**
    - Σ_{tracks} p_t < 4 GeV (inside cone dR<0.3)
- Fiducial region:
  - $|\eta| < 1.37$, $p_t > 15$ GeV



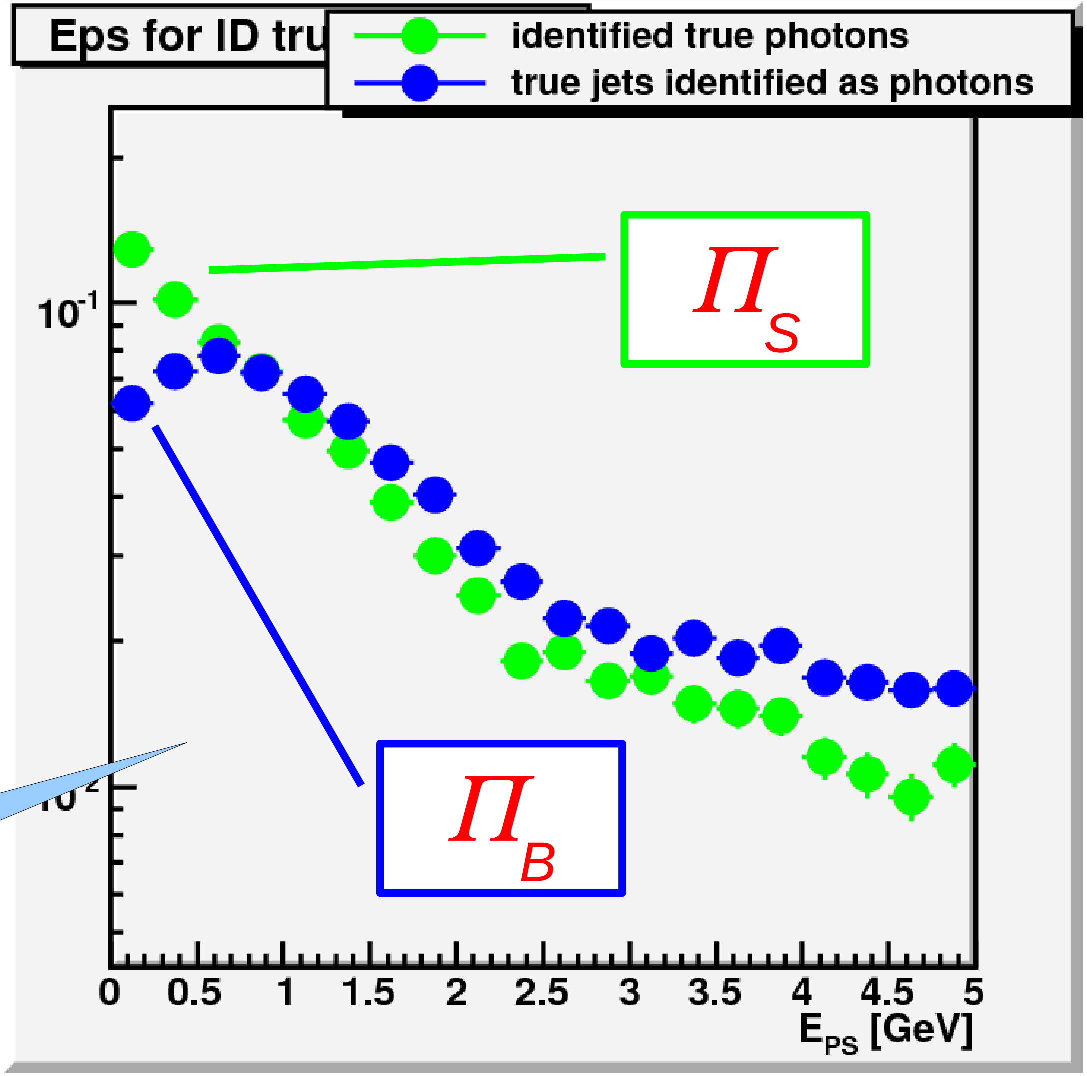
Despite good efficiency (>60%) and high jet rejection ($\approx 10^3 \div 10^4$), signal and background occur in similar amounts...

Purity evaluation (basics)

- Estimate the photon purity *P* in the candidates' sample:
  - Choose a **<u>separation variable</u>** ξ having pdf's sensibly different for true signal photons and background: $\Pi_S(\xi)$ and $\Pi_B(\xi)$;
  - Get the pdf from observed candidates sample: $\Pi_{obs}(\xi)$;
  - Find value of P that best fits:

$$\Pi_{obs}(\xi) = P\,\Pi_S(\xi) + (1-P)\,\Pi_B(\xi)\ .$$

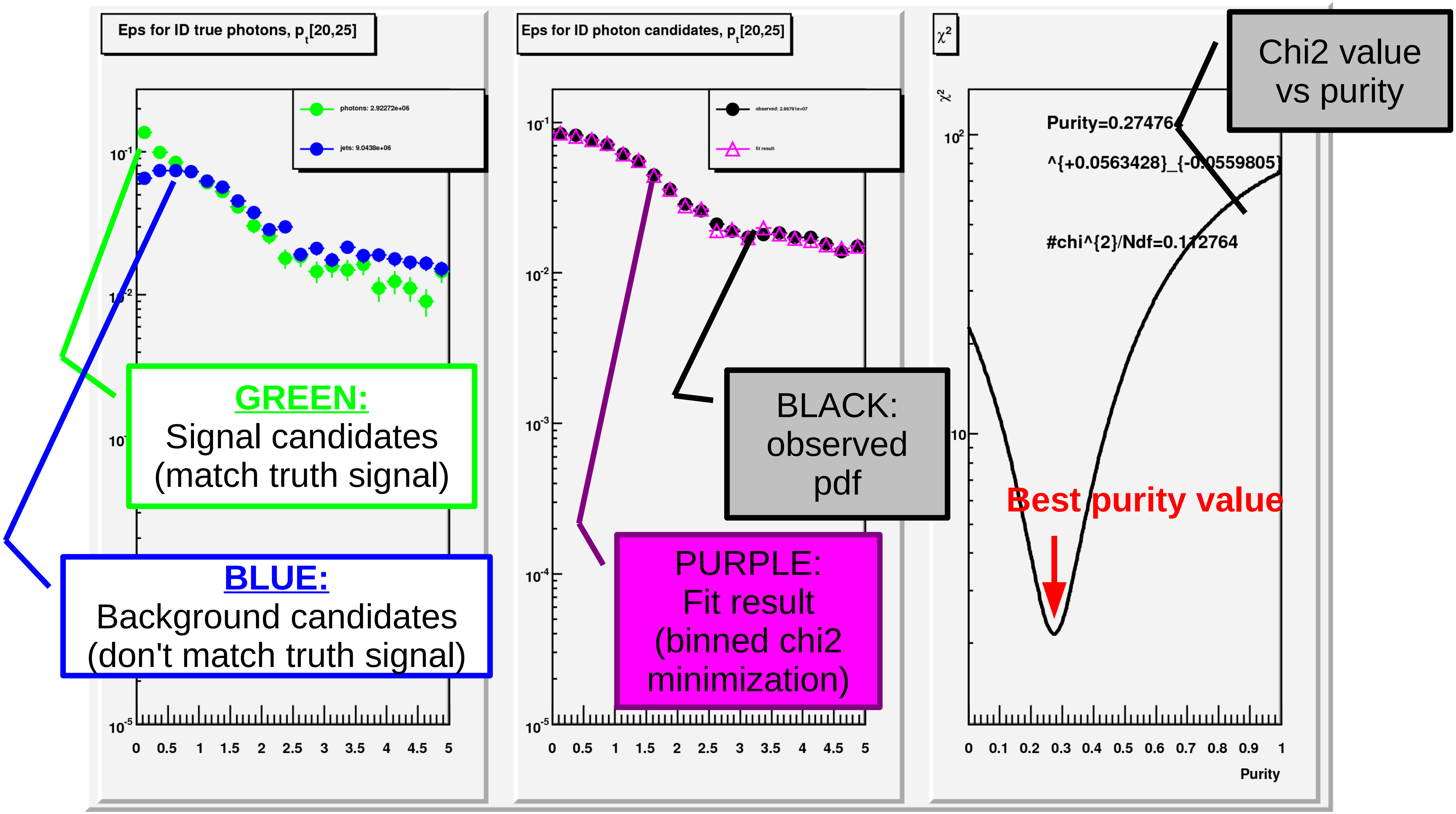
- Choose $\xi \equiv E_{PS}$:



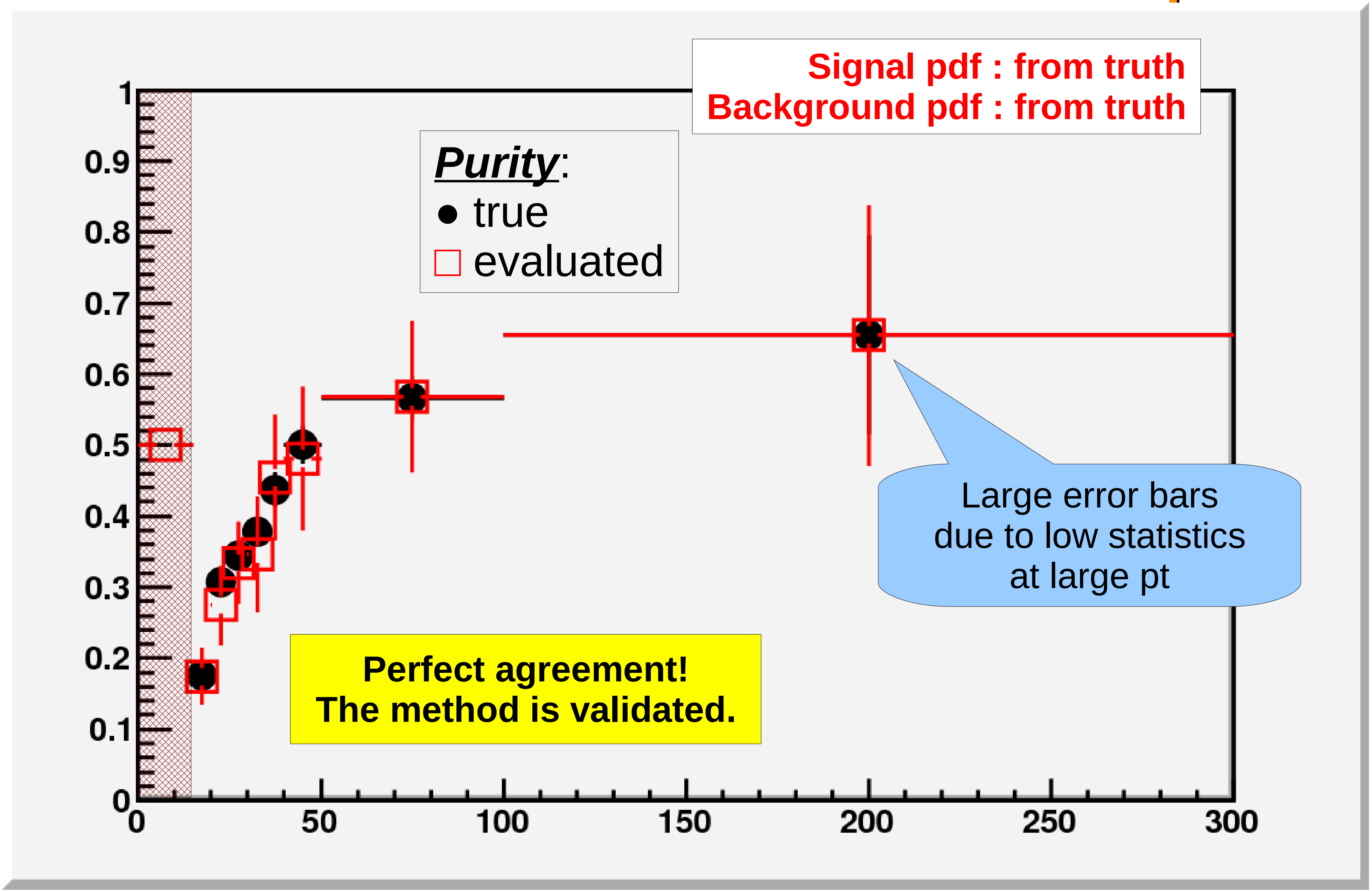
Larger E_{PS} in background:
- in jets (more track activity)
- In $h^0 \to 2\gamma$ (more conversions)

Example: purity fit (20<pt<25)

- Find value of P that best fits: $\Pi_{obs}(\xi) = P\, \Pi_S(\xi) + (1-P)\, \Pi_B(\xi)$.



Purity evaluation using E_{ps}

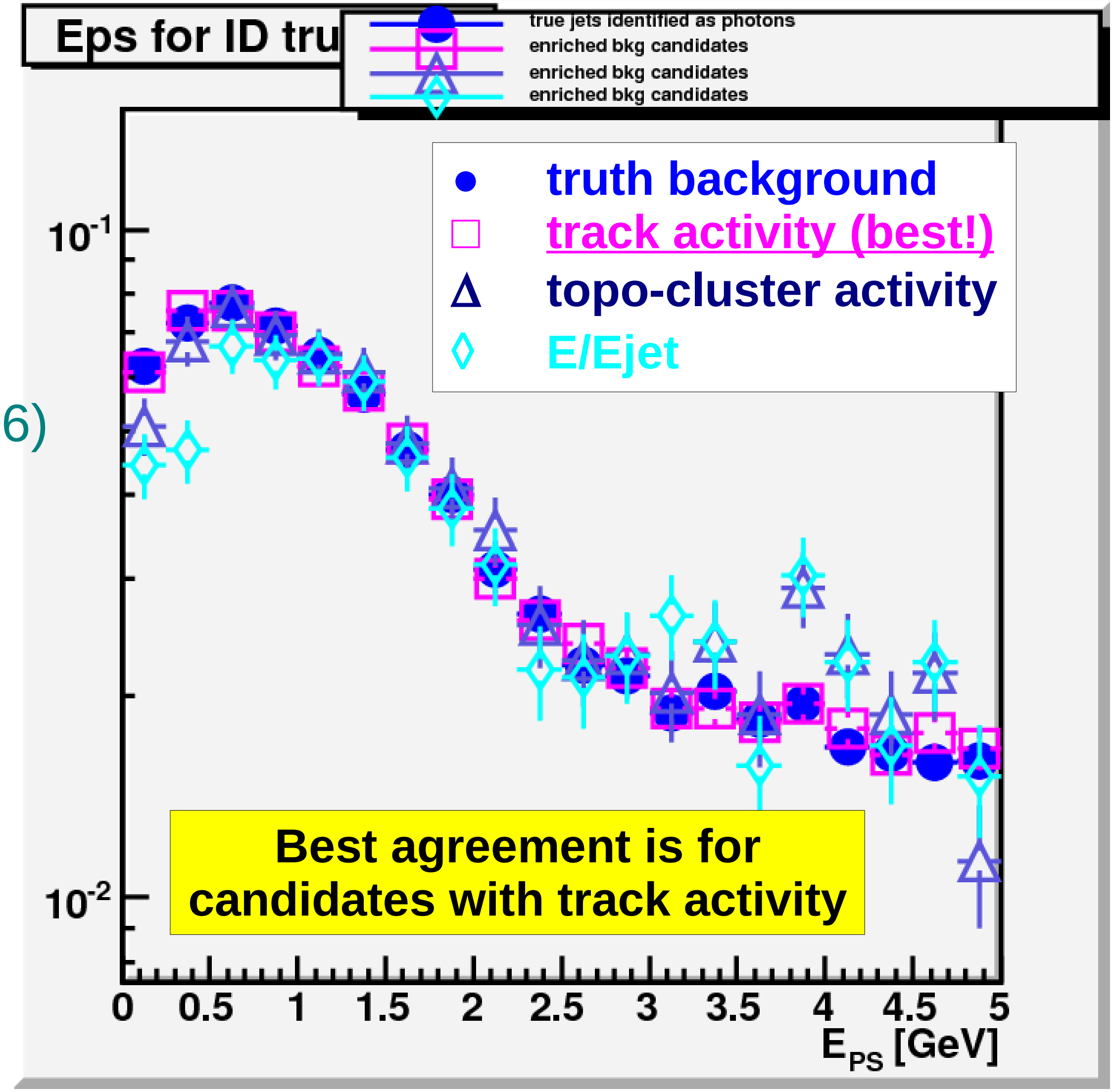


Do we trust simulation ?

- Background:
  - Depends much on fragmentation/hadronization model
  - Try to extract information from real data
  - Build background-enriched sample as control sample
- Signal:
  - Probably ok for prompt photons, more model-dependent for fragmentation (q/g bremsstrahlung)
  - Try to build signal-enriched sample from our candidate sample (but more difficult than for backgrounds, too many jets!)
  - Possible systematics study: use only prompt photons instead of signal (=prompt+brem) ?
- Therefore, use MC truth mostly as a tool to validate our assumptions and check final results

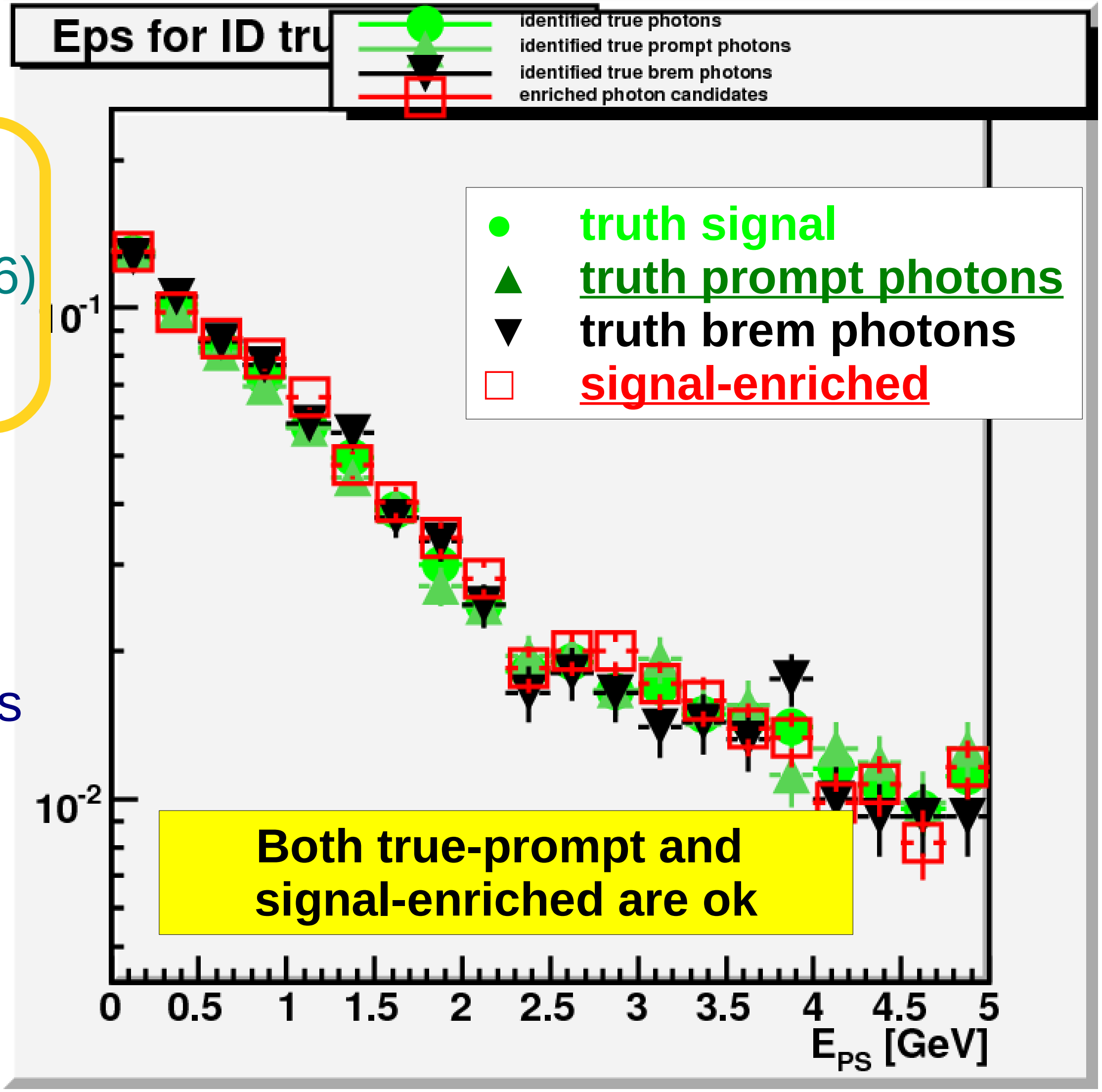
Background pdf's

- Build *background-enriched* sample: several recipes tried:
  - **Track activity (best!)**
    - $\Sigma_{tracks}\ p_t > 1$ GeV (dR<0.3)
  - Topo-cluster activity
    - $p_t / (\Sigma_{topo\text{-}clus}\ p_t) < 0.5$ (dR<0.6)
  - Cluster/jet energy ratio
    - $E / E_{jet} < 0.5$
- Require tracks activity around the photon candidate:
  - Signal contamination < 10%
  - Obtained pdf agrees with "true" one

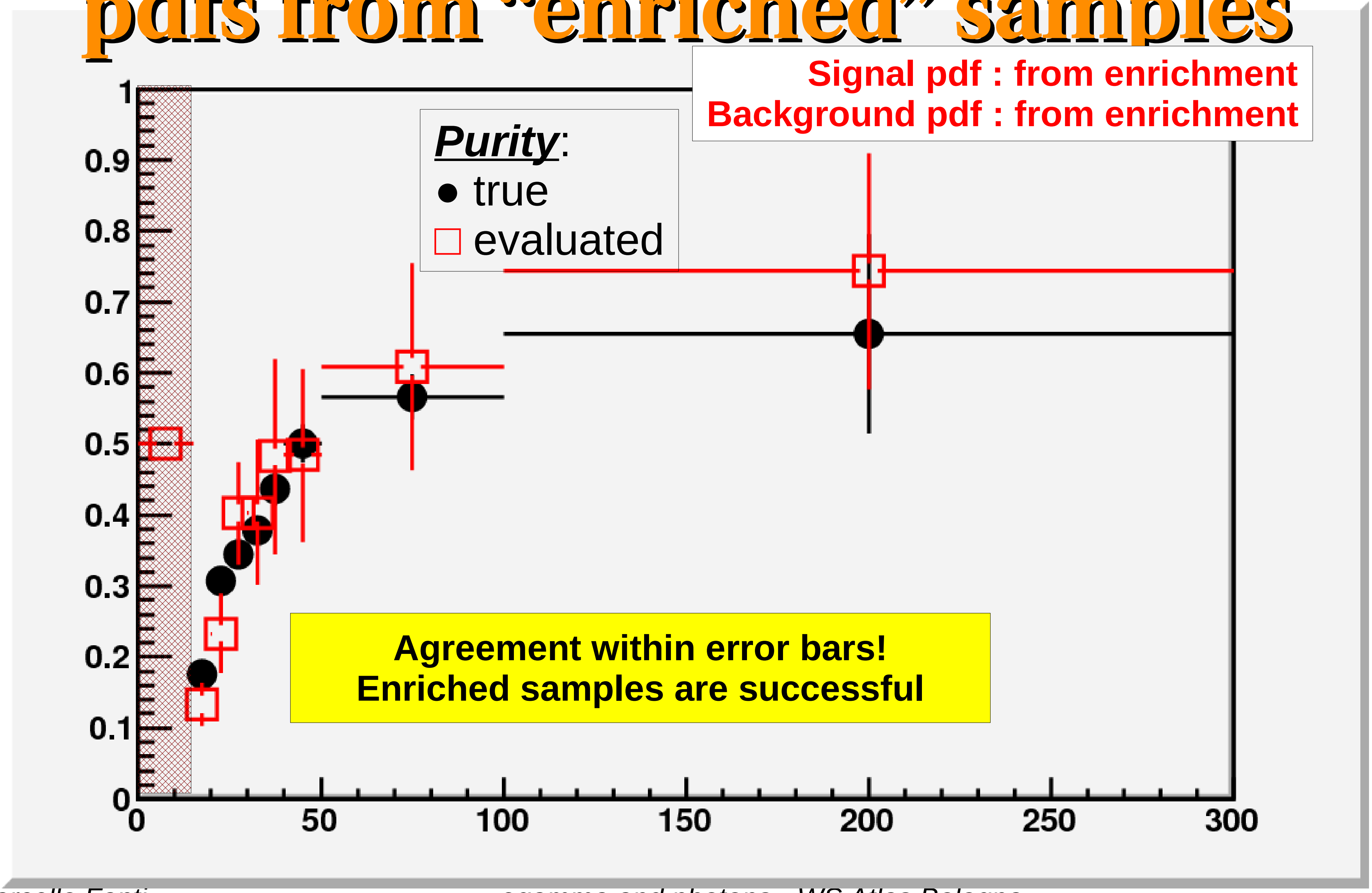


Signal pdf's

- Try to build *signal-enriched* sample
  - Cuts:
    - $\Sigma_{tracks}\ p_t < 1$ GeV (dR<0.3)
    - $p_t / (\Sigma_{topo\text{-}clus}\ p_t) > 0.9$ (dR<0.6)
    - $E / E_{jet} > 0.8$
  - Pdf agrees with "true" one
- ... But signal at ~ 70%, not so "rich", less robust!
- Use pdf from *true prompt photons* as cross-check

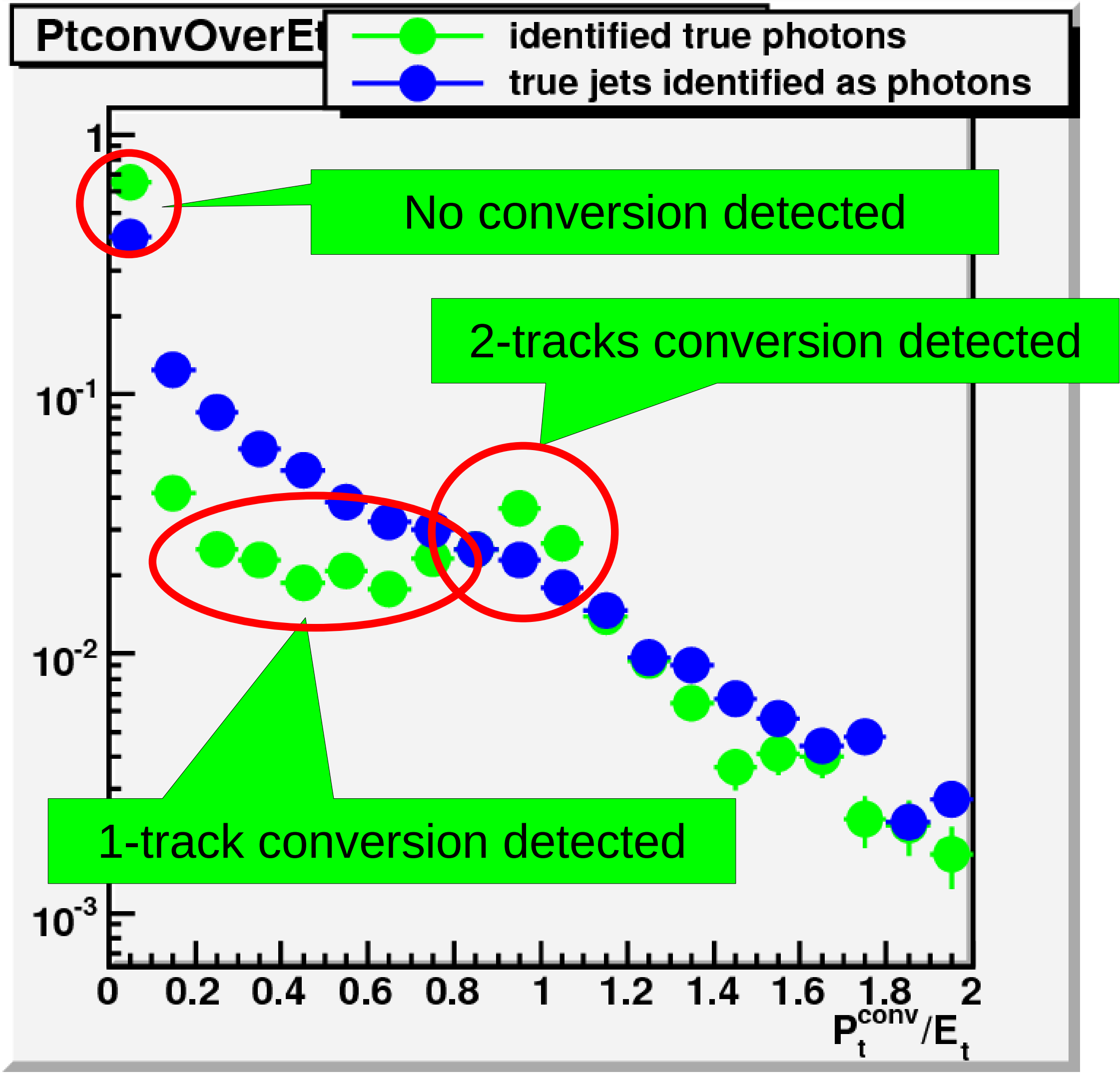


Purity evaluation using E_{ps} : pdfs from “enriched” samples

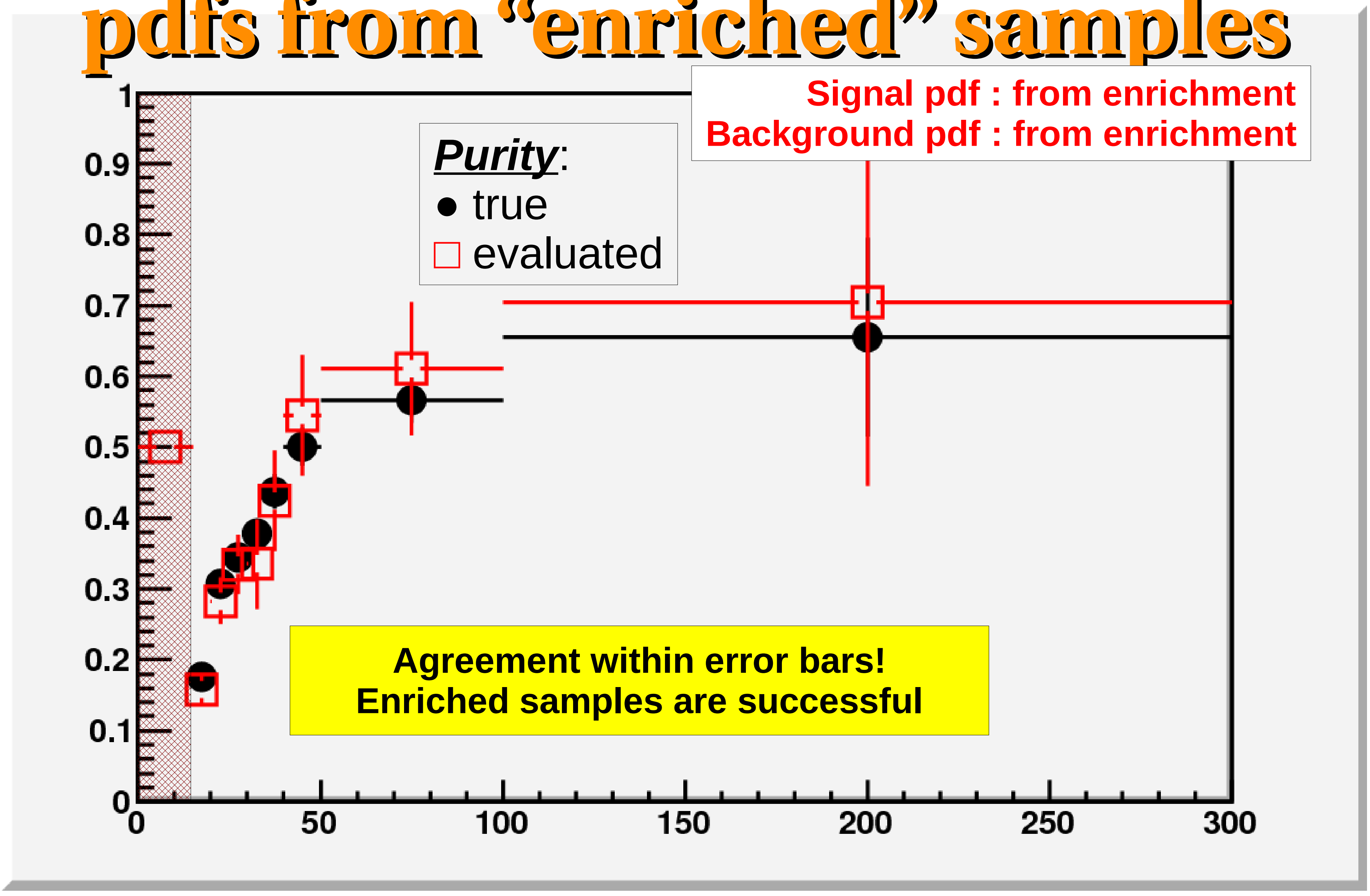


Using P_t of detected conversions

- Detected photon conversions:
  - **2-tracks conversion**: 2-tracks vertex, displaced from primary vertex, with at least a track pointing to em cluster
  - **1-track conversion**: em cluster with an associated track, without hits in B-layer
- Define P_t^{conv} as:
  - $P_t^{track1}+P_t^{track2}$ for 2-tracks conv
  - P_t^{track1} for 1-track conv
- Use $\xi \equiv P_t^{conv}/E_T$.



Purity evaluation using P_t^{conv}/E_t : pdfs from “enriched” samples



Photon physics with early data

Photon efficiency

- We are not directly involved at the moment, but this topic is strictly connected to our needs
- Photon efficiency will be MonteCarlo-based, and depend on the sample (e.g. photons from Higgs are more isolated than photons from QCD fragmentation...) but data-driven techniques provide handles to validate and/or correct MC efficiency in some cases
- Tag&probe with $Z \to ee$ and translate to photons
  - Apply photon ID cuts to “probe electron”, get efficiency, compare to “true” efficiency (photon cuts on true electron), and derive corrections if needed, to be applied to any truth-based efficiency --- could be feasible with 200 pb^{-1} collected data
- FSR in $Z \to \mu\mu\gamma$ events
  - After 200 pb^{-1} , ~5000 candidates are expected after analysis cuts; using 4 eta-bins and 8 pt-bins, efficiency could be estimated with ~ 5% uncertainty ...
- In the beginning, g20 trigger will be used; estimated efficiency 100%

Purity

- In principle we are ready:
  - Pdf method with more independent variables.
  - Also an independent method (2-D sidebands, not shown here) which gives comparable results.
  - If all agree, we are “confortable”
  - If they don't agree, or if extracted purity is too far from expected one, comparison could help in finding where the problems are...
  - ... But need to test methods on a different fragmentation model (e.g. Herwig)
- With 0.1 pb^{-1} we “are able” to get purity with ~5% uncertainty, with 8 pt-bins pt>15 GeV (no η-binning, need to believe in MC η distribution)
- 10 pb^{-1} statistics could allow η-binning (e.g. 4 bins, being thus more model-independent!) and smaller error bars, thus allowing systematic studies across different methods

Photon analysis

- Photon physics with 10 pb^{-1} :
  - expected $\approx 10^6$ gamma-jets + $\approx 10^3$ di-gammas
  - Purity from data, efficiency from MC
  - Inclusive photon physics
- Photon physics with 200 pb^{-1} :
  - $\approx 2 \cdot 10^7$ gamma-jets + $\approx 2 \cdot 10^4$ di-gammas
  - Data-driven purity and efficiency
  - Inclusive photon and di-photon physics

Software for photon physics

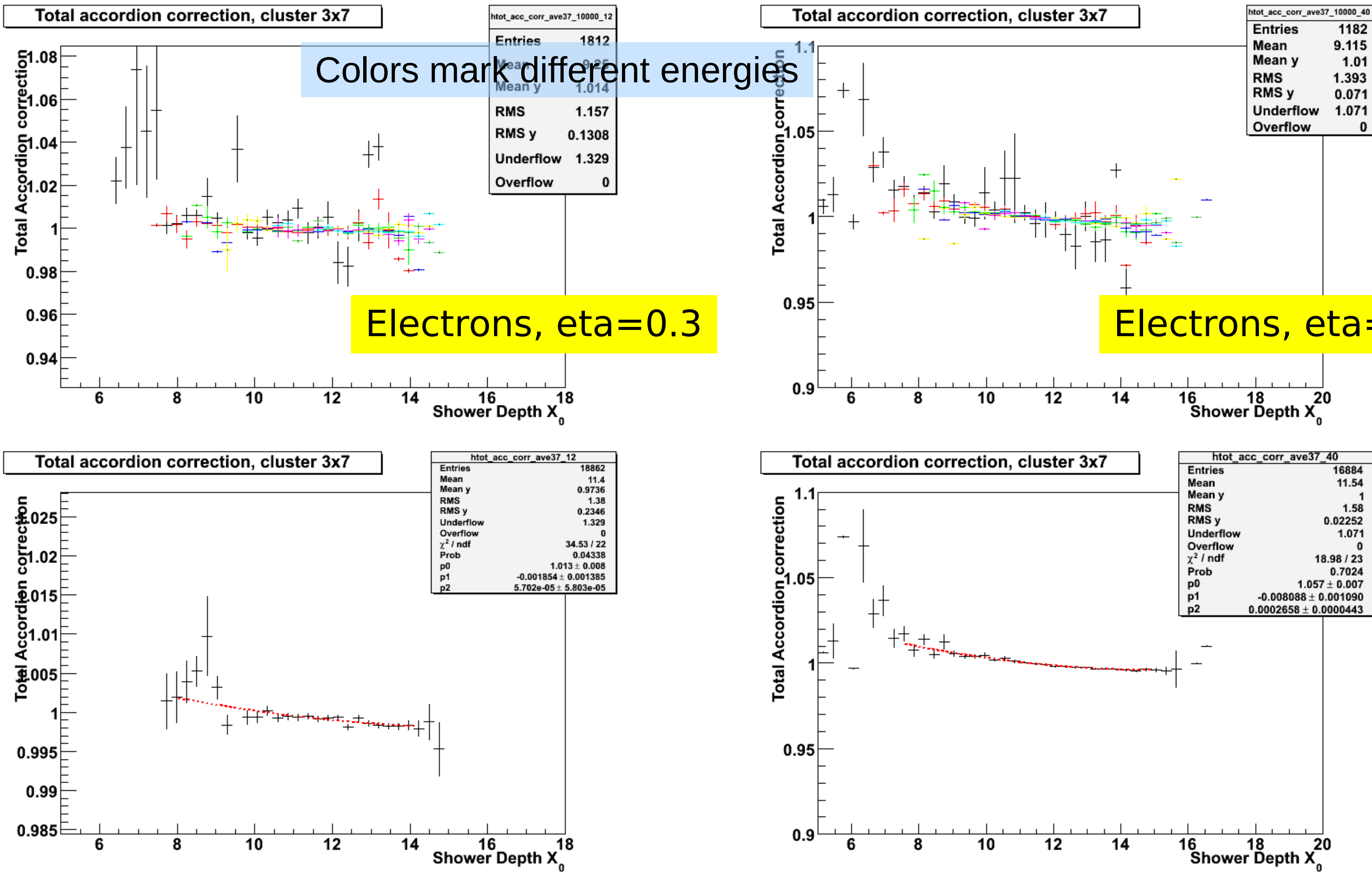
L. Carminati
M. Fanti

Software

- Using athena package HiggsAnalysisUtils (originaly written for H$\rightarrow\gamma\gamma$, then extended to generic photon physics)
  - https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HiggsAnalysisUtils
  - Runs on athena, on AOD and D1PD
  - Uses dedicated tools and egamma tools
  - Produces ntuples (“D3PD”) containing mainly info for reconstructed photons, jets, and associated truth info, and trigger blocks
  - Publicly available on CVS
- Soon to become PhotonAnalysisUtils ☺
- Already working efficiently, need to be interfaced with InSituPerformance for storing results to DB, and access luminosity blocks info

BACKUP SLIDES

Accordion correction



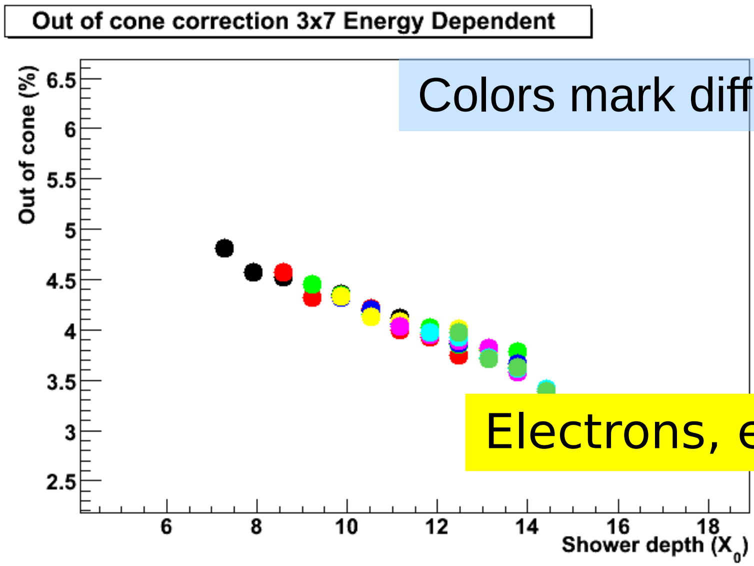
Colors mark different energies

Electrons, eta=0.3

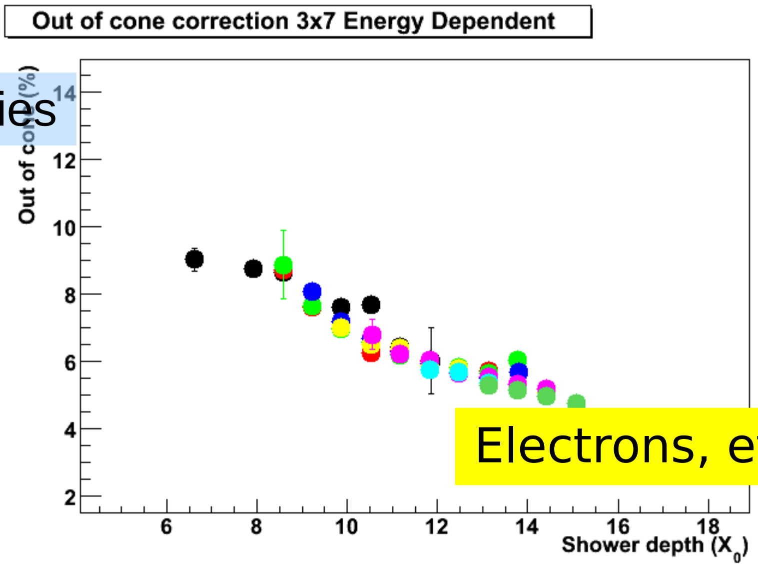
Electrons, eta=1.2

Out-of-cluster correction

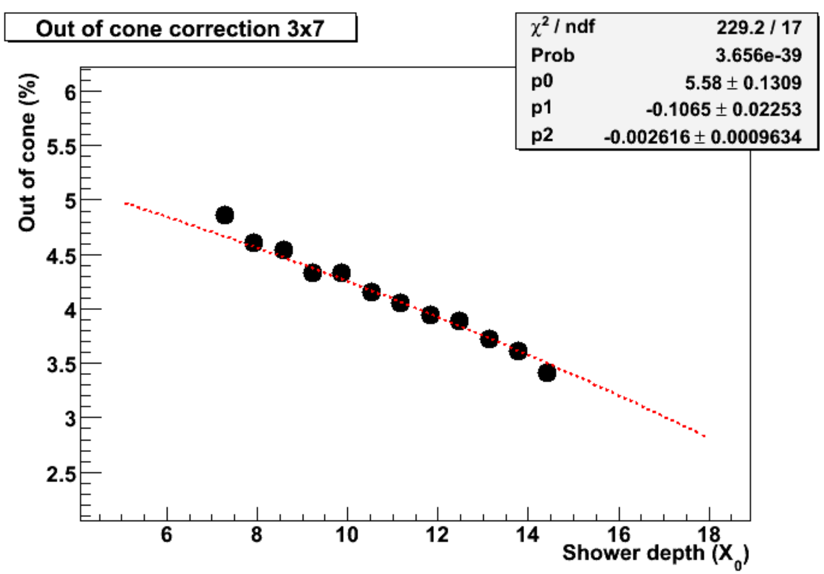
Colors mark different energies

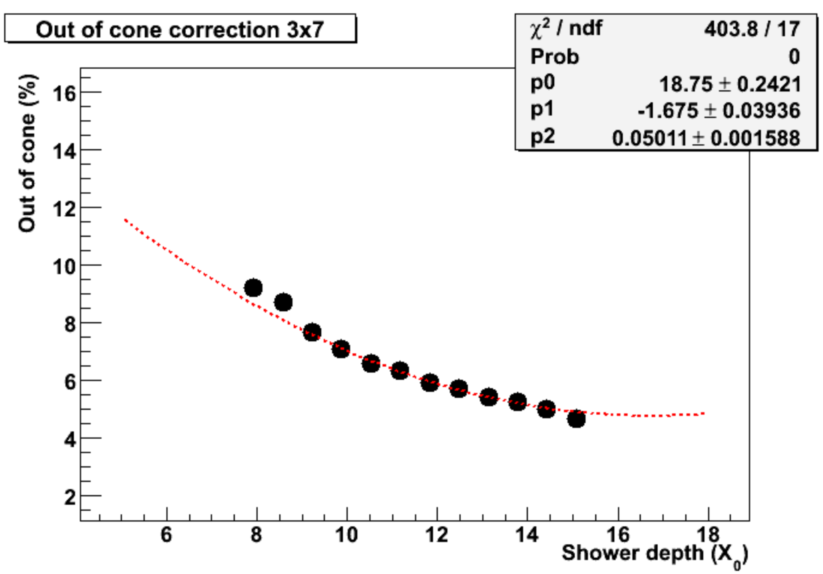


Electrons, eta=0.3



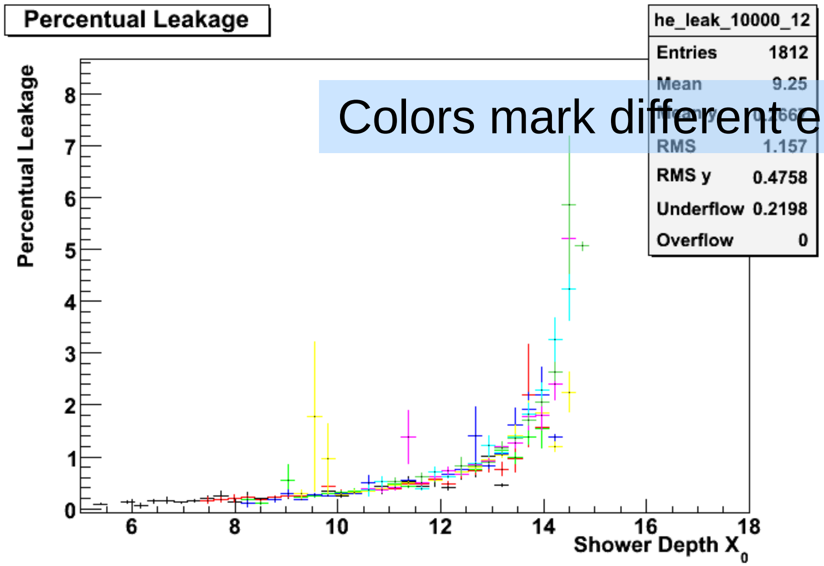
Electrons, eta=1.2

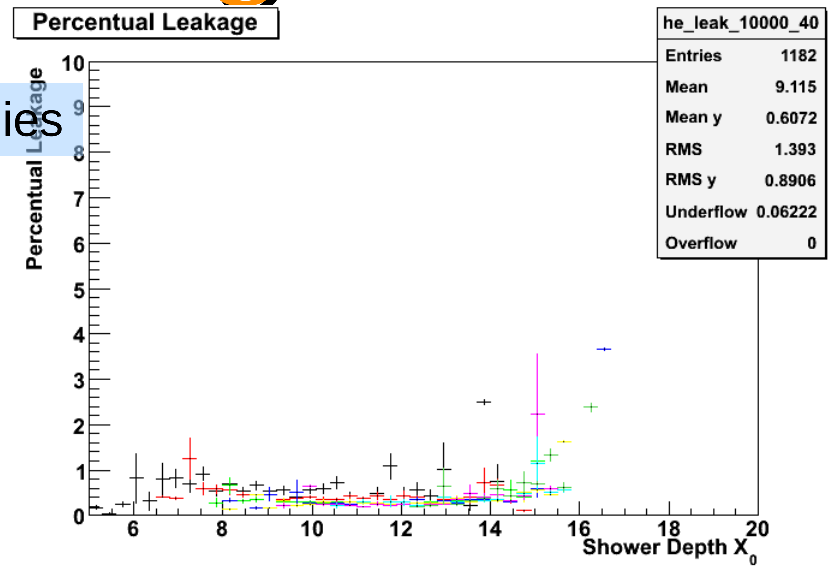


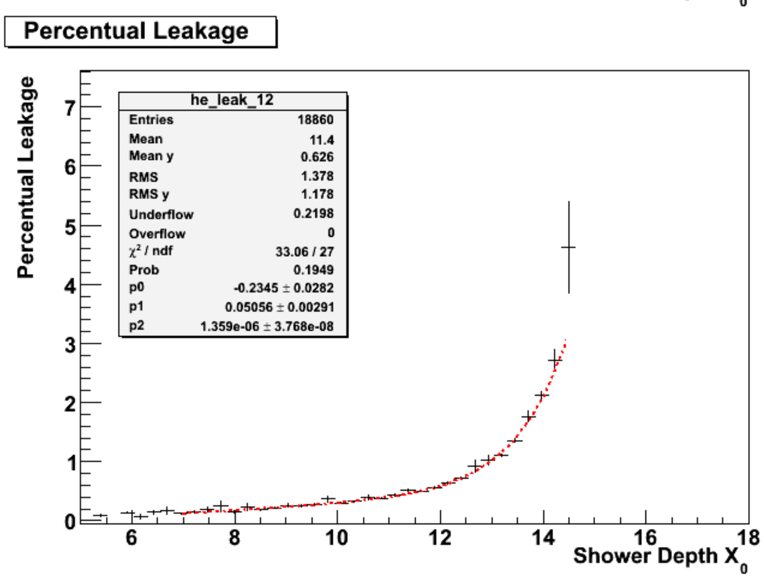


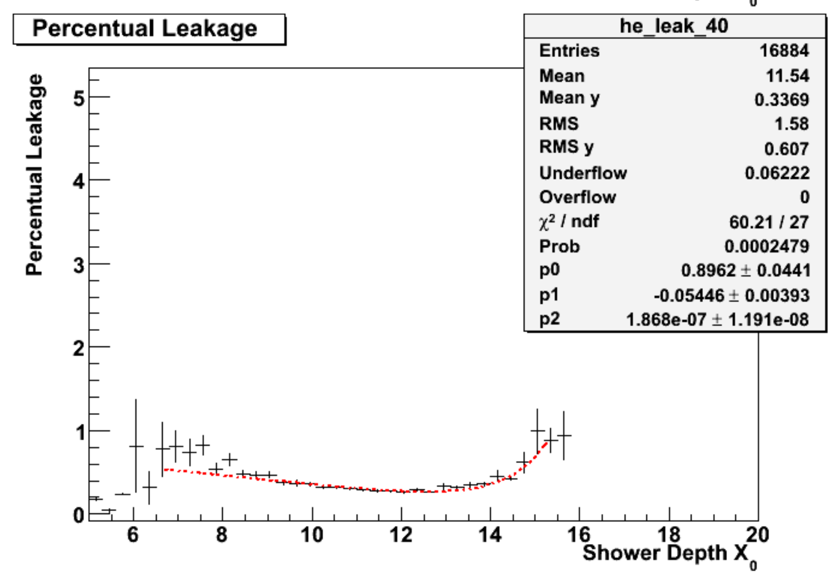
Back leakage

Colors mark different energies

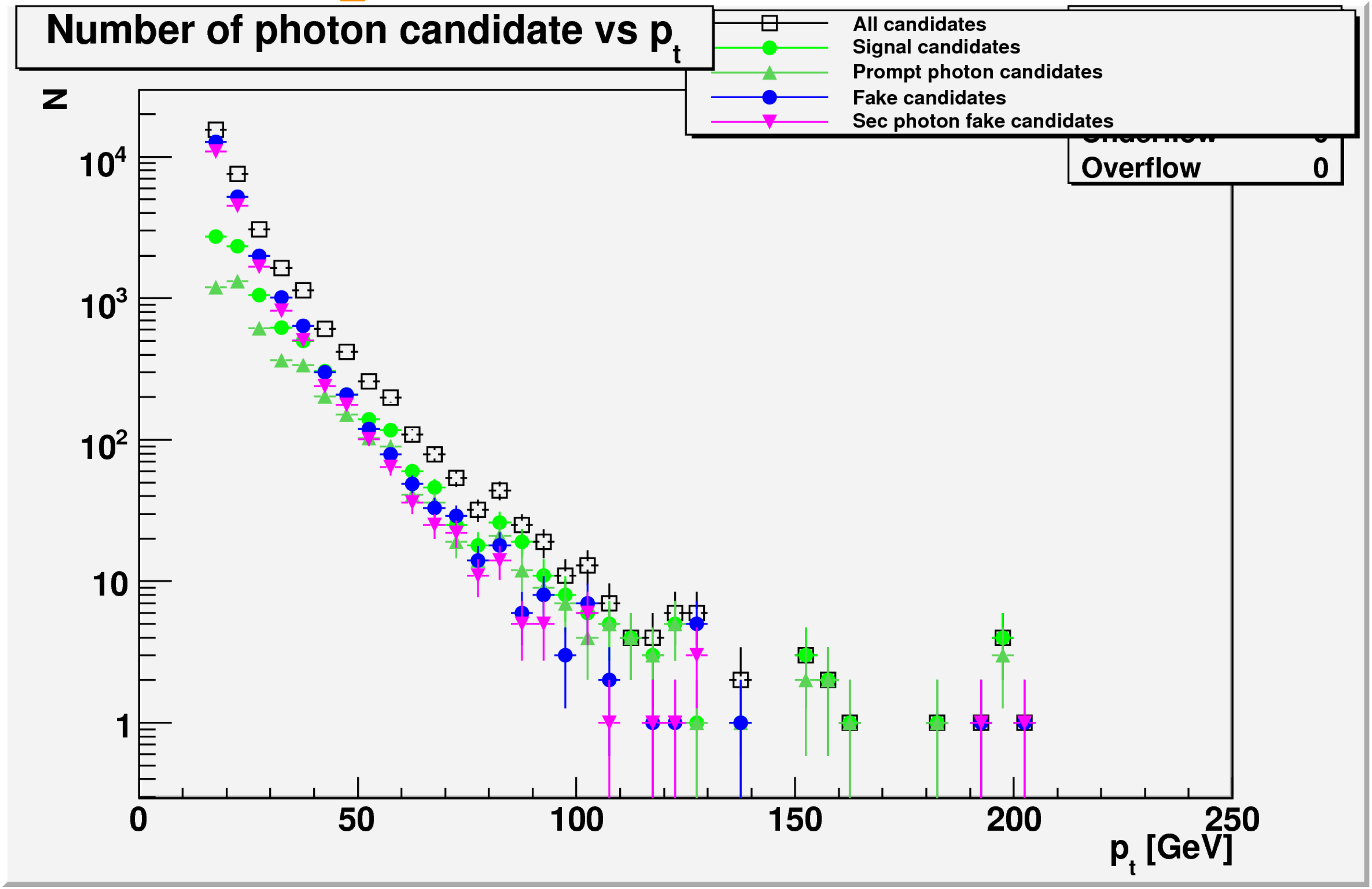




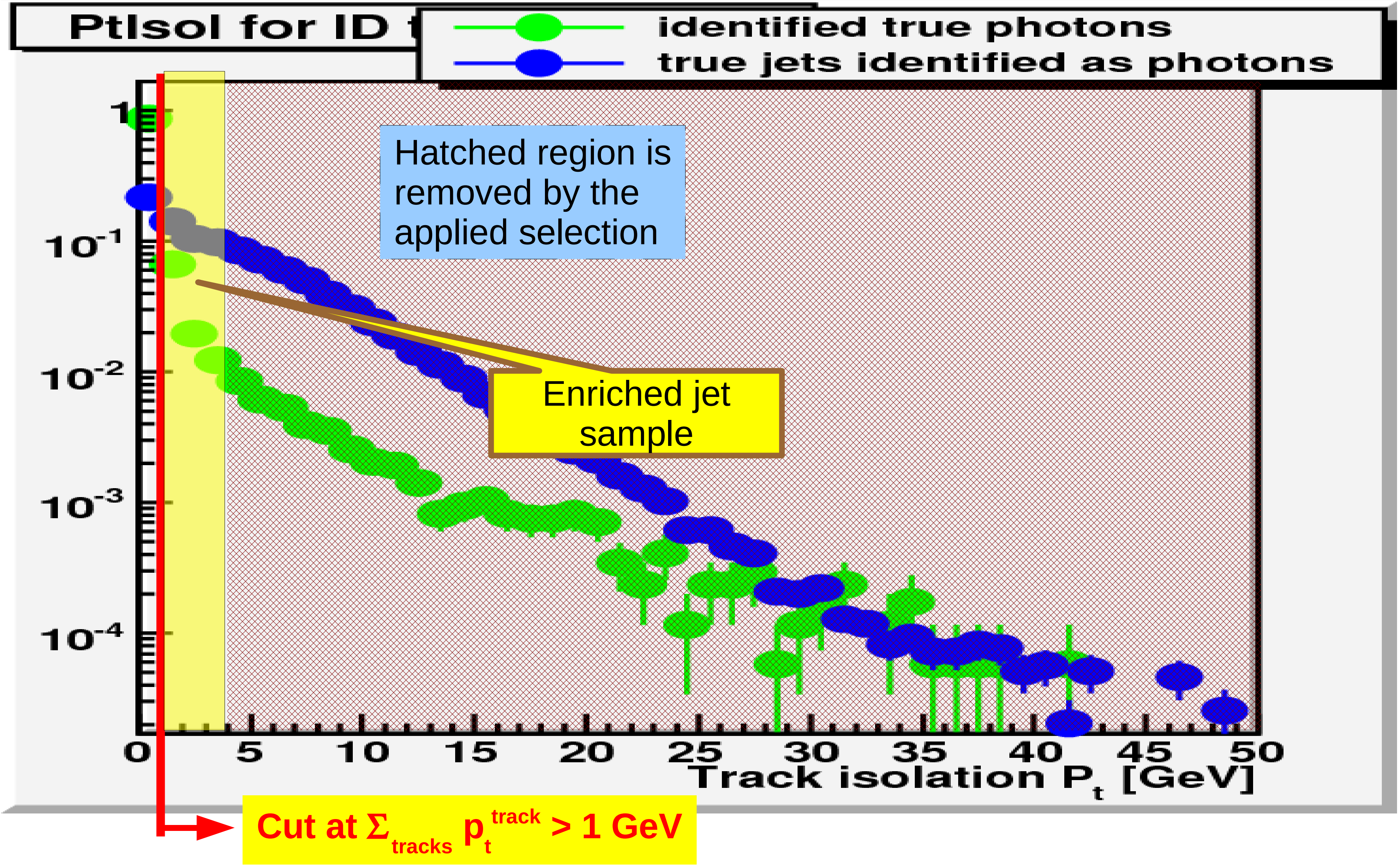




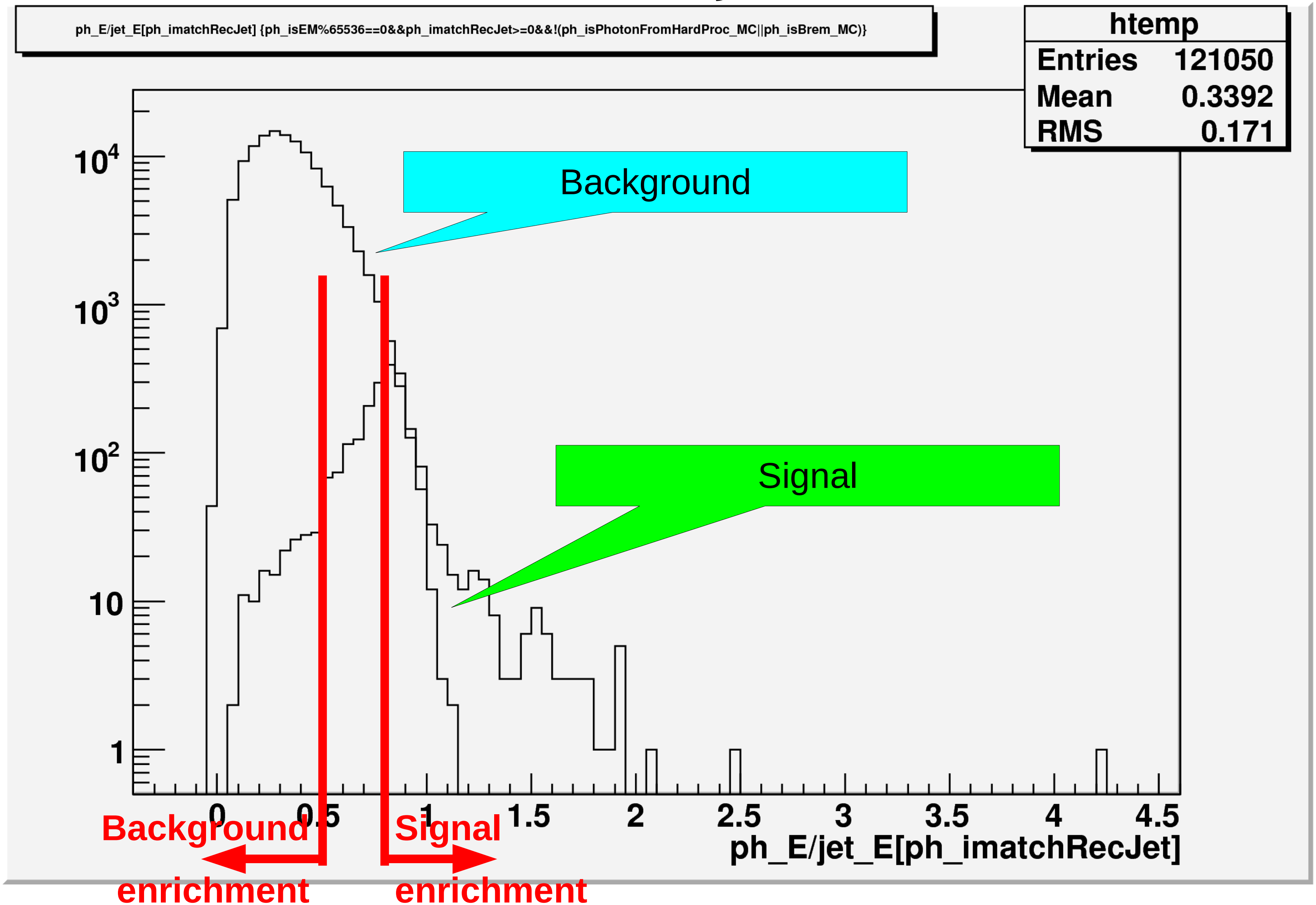
Composition of candidates



Track isolation

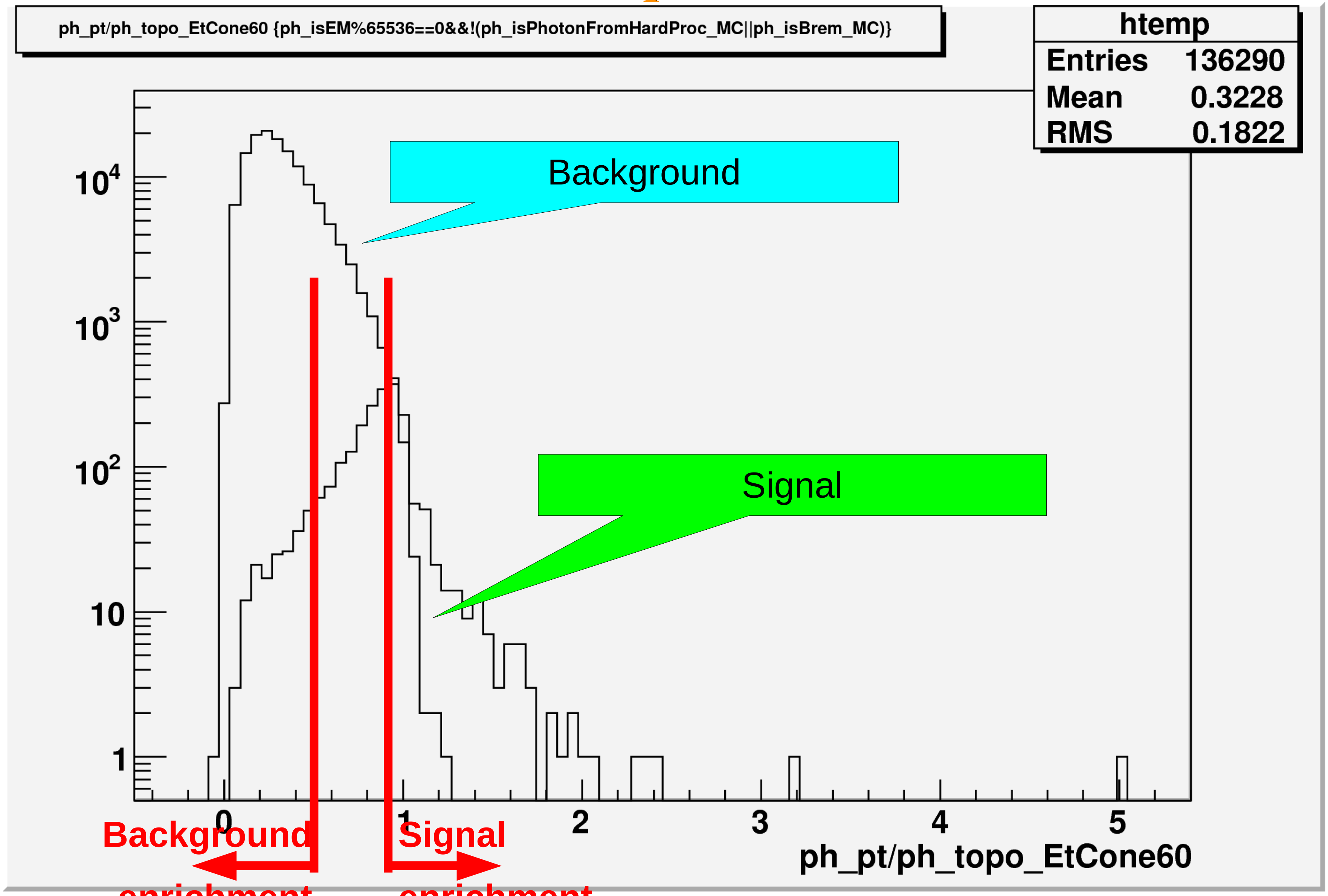


E / E_{jet}



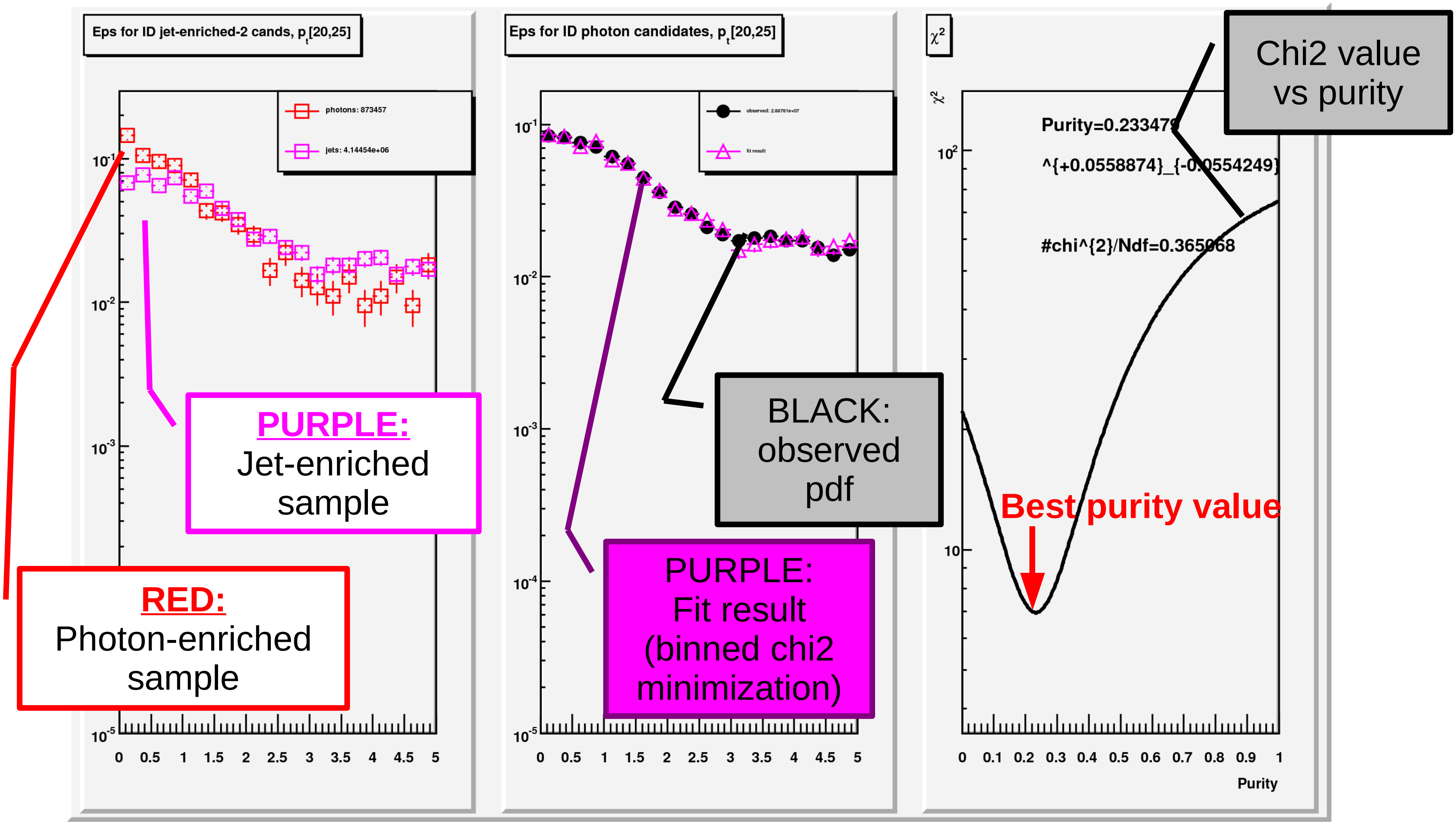
ph_E/jet_E[ph_imatchRecJet] {ph_isEM%65536==0&&ph_imatchRecJet>=0&&!(ph_isPhotonFromHardProc_MC||ph_isBrem_MC)}
htemp
Entries 121050
Mean 0.3392
RMS 0.171
Background
Signal
ph_E/jet_E[ph_imatchRecJet]
Background enrichment
Signal enrichment

$p_t / (\Sigma_{topo\text{-}clus}\, p_t)$

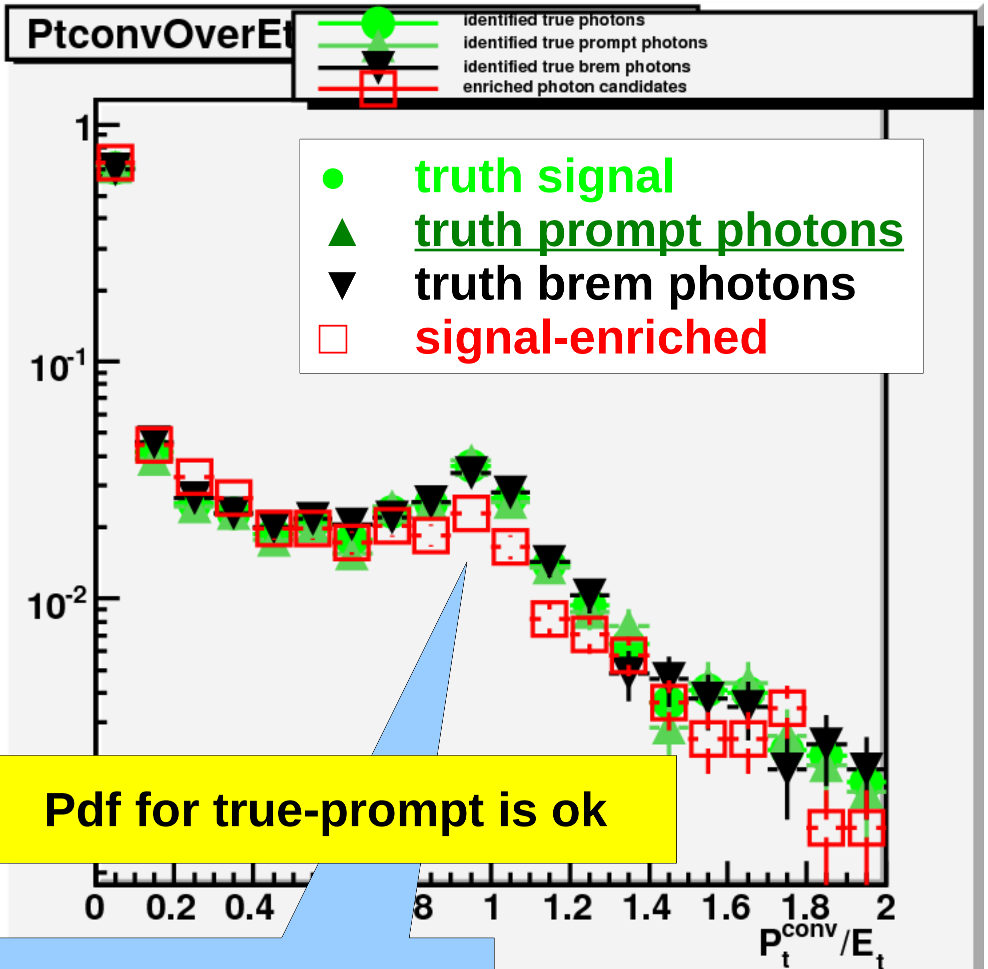


Example: purity fit (20<pt<25)

- Find value of P that best fits: $\Pi_{obs}(\xi) = P\,\Pi_S(\xi) + (1-P)\,\Pi_B(\xi)$.

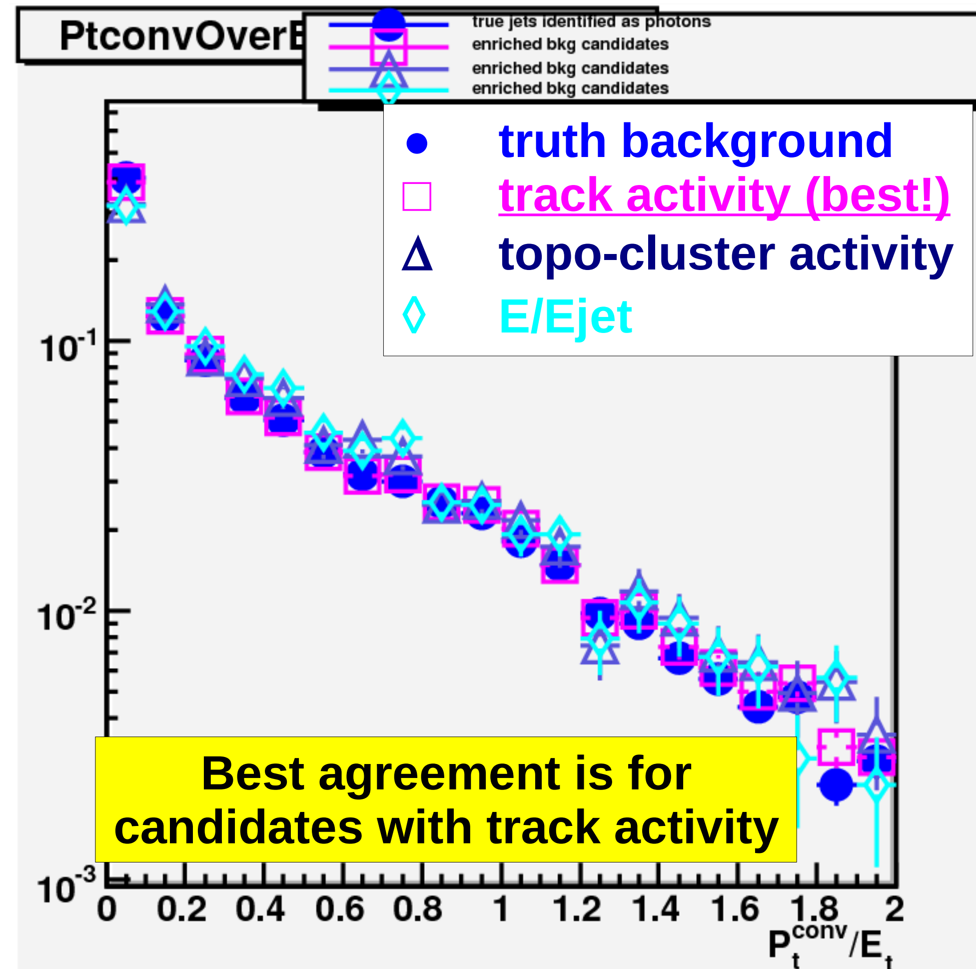


Enriched samples



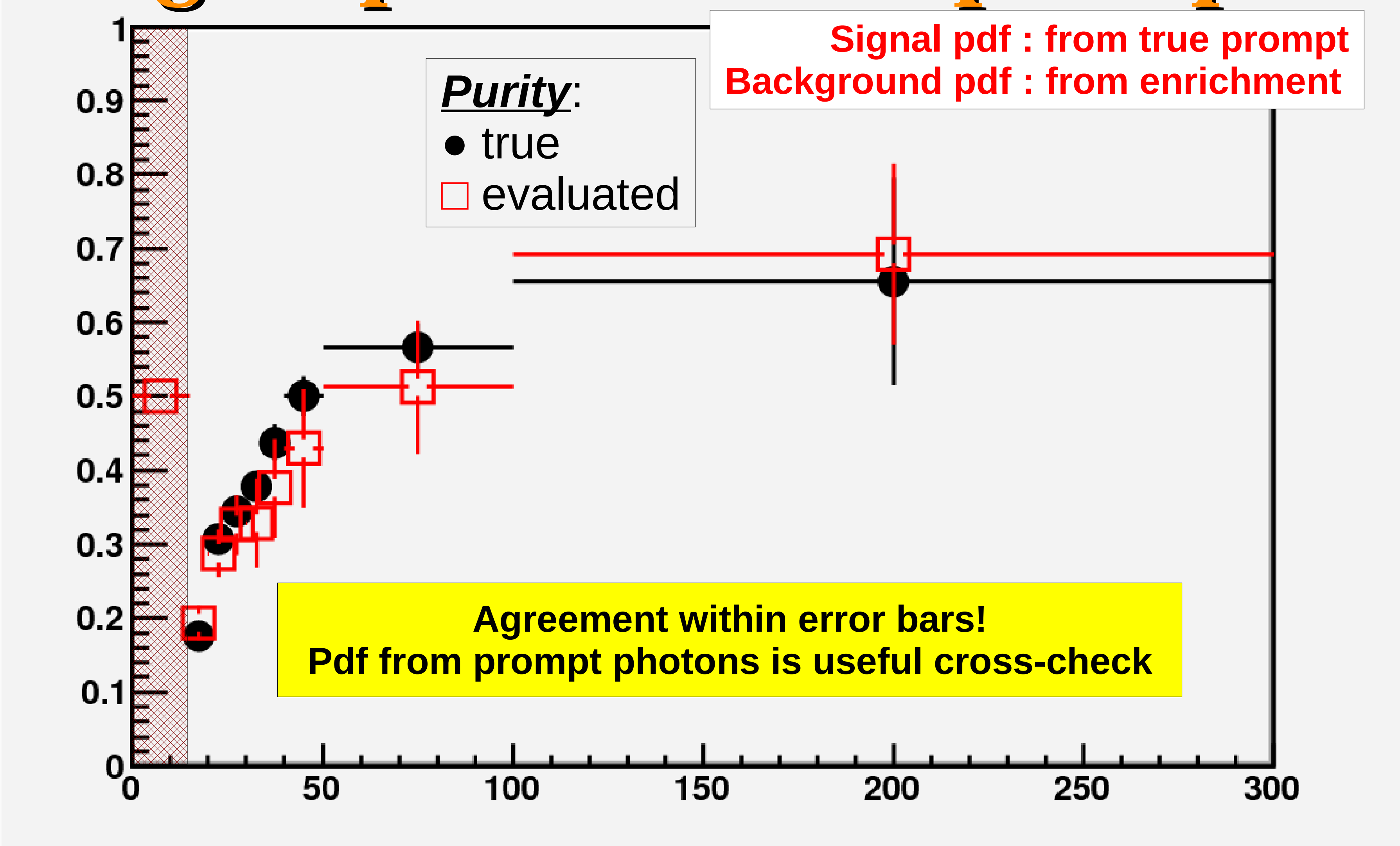
Pdf for true-prompt is ok

Peak at 1
not well reproduced
in enriched sample.
Will it be a problem?



Best agreement is for
candidates with track activity

Purity evaluation using P_t^{conv}/E_t : signal pdf from true prompts

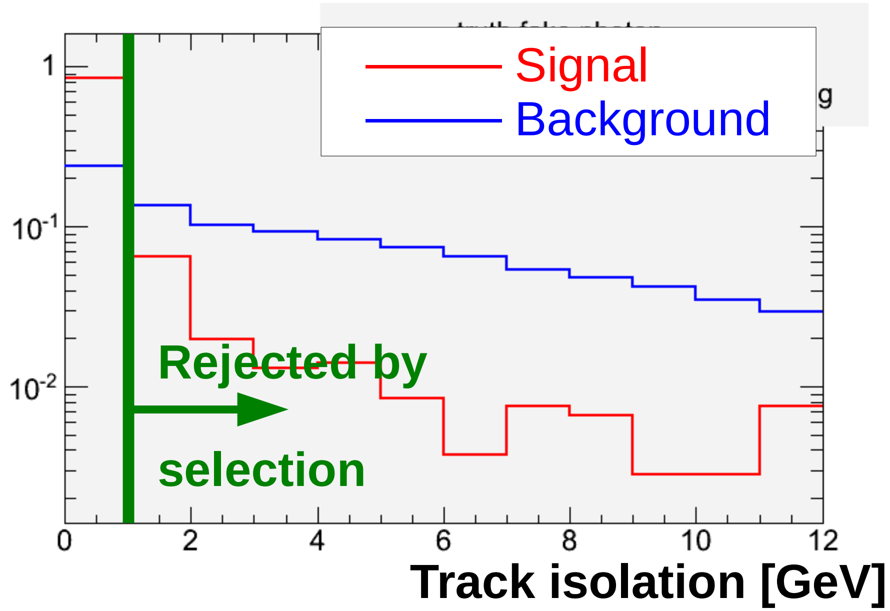


2-dimensional side bands

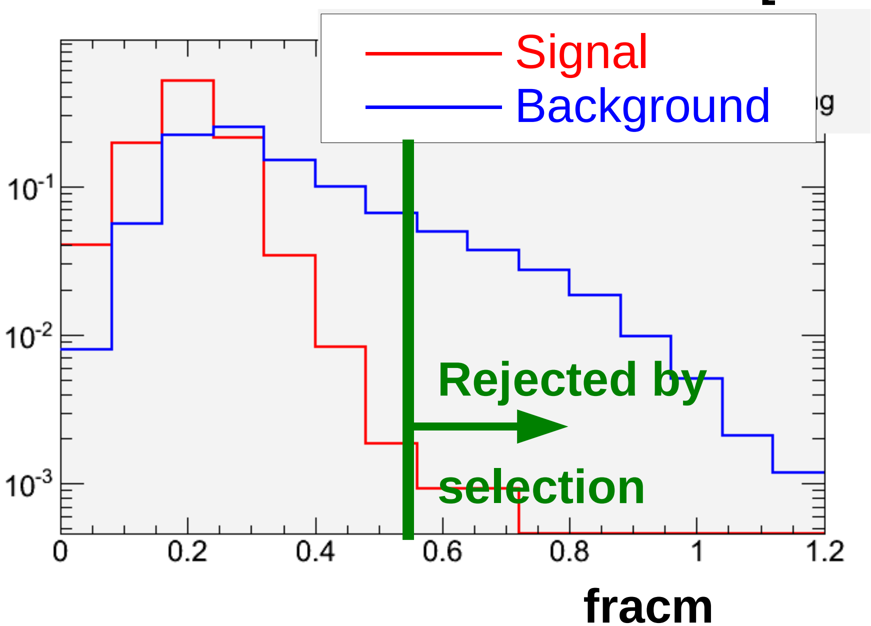
(Zhijun Liang)

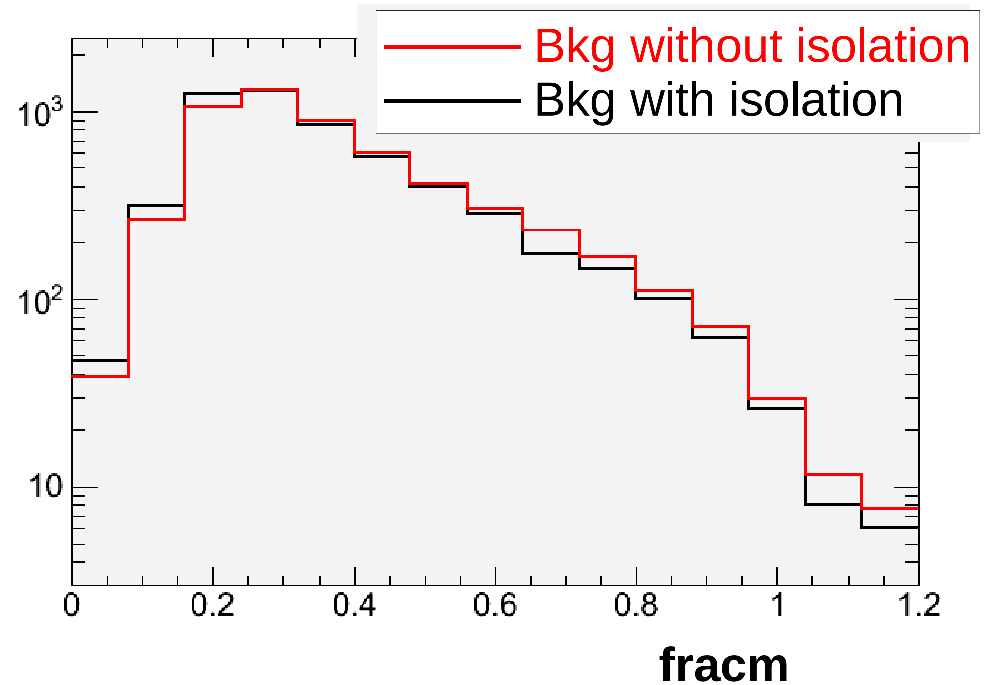
(let's remind that having another independent method could be very useful, when looking at 1^{st} data)

The method

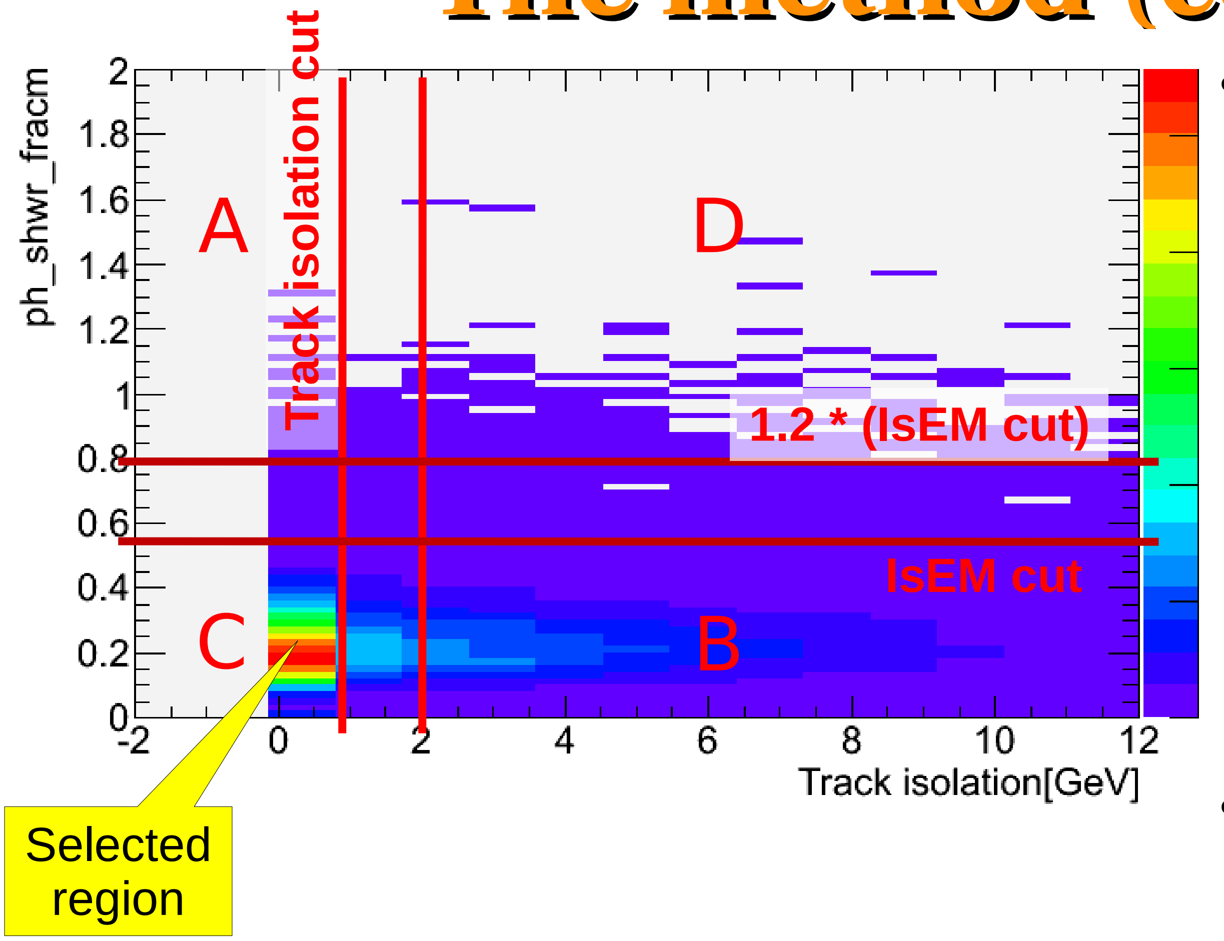


- 2 discriminating variables:
  - track isolation Σ_{tracks} p_t (inside cone dR=0.3);
  - "fracm" = ($E_{\pm 7}$ - $E_{\pm 3}$) / $E_{\pm 7}$ (out-of-core energy fraction in strips)
- No correlation for background:





The method (cont'd)



- Assumptions:
  - regions A,B,D dominated by background (checked against truth: signal < 6%)

  $N_{A,B,D} \cong N^{(fake)}_{A,B,D}$

  - Fracm and track isolation independent (see previous slide)

  $N^{(fake)}_{C} / N^{(fake)}_{A} = N^{(fake)}_{B} / N^{(fake)}_{D}$

- From candidates' counts in A,B,D predict fake counts in C:

$$N^{(fake)}_{C} = (N^{(fake)}_{A} N^{(fake)}_{B}) / N^{(fake)}_{D} \cong (N_{A} N_{B}) / N_{D}$$

$$P = 1 - N^{(fake)}_{C} / N_{C} \cong 1 - (N_{A} N_{B}) / (N_{c} N_{D})$$

Example of calculation

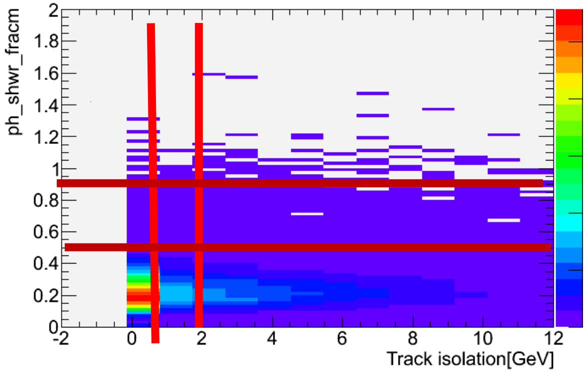
Table for photon candidate Pt~[30GeV,35GeV]

	Total events	truth Photons (N’’)	Truth fakes (N’)
Control region A	184	7 (NA’’)	177 (NA’)
Control region B	1202	67 (NB’’)	1135 (NB’)
Signal region C	1211	542 (NC’’)	669 (NC’)
Control region D	338	0 (ND’’)	338 (ND’)

- The truth information shows:
- Background events in region A B D is at least 20 times larger than signal events
- NA’>>NA’’ and NB’>>NB’’
- Formulae for background events NA’/NC’=ND’/NB’ are validated

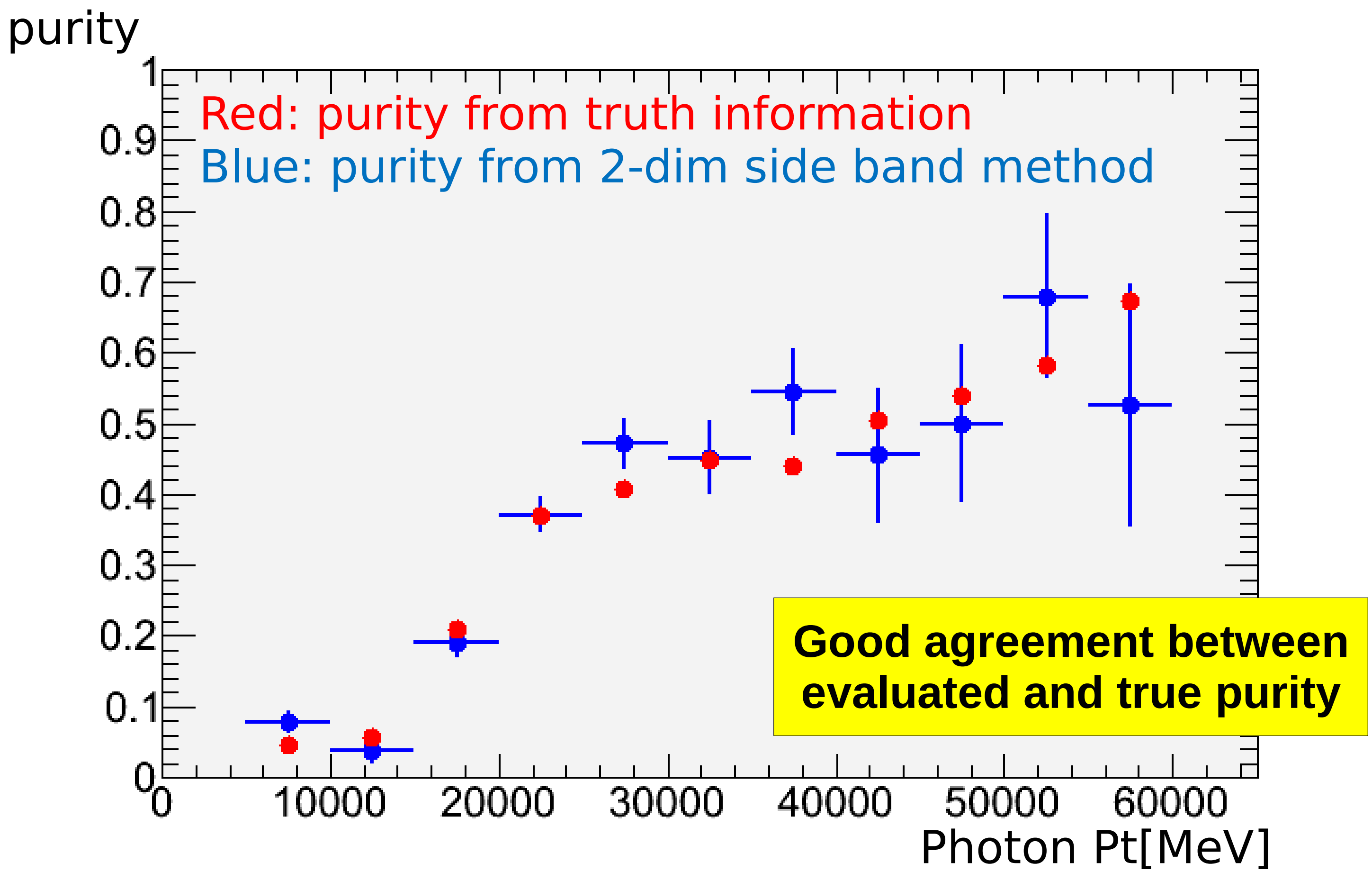
$P_{calc} = 1 - (184*1202)/(1211*338) = 0.46$

$P_{true} = 542/1211 = 0.45$



46

Result of 2-dim side bands method



Caveat: due to different track isolation in candidates' definition, these values are not directly comparable with those from pdf's method)