1 Why do we need purity

For the γ-hadron analysis we are interested in correlated direct photons (trigger particles), i.e. coming directly from a hard process, to the opposite side hadrons. In this purpose, in a whole sample of EMCal clusters, we need to be able to identifie direct photons clusters and reject background clusters coming from $\pi^0$and η for the most of them.

To do this selection, one can first apply some basic cuts to remove the main part of clusters in which we are not interested in :

— Cut on particle momentum : The photon coming from a hard process is more energetic than the hadrons coming from parton emitted back to this photon and than other particles coming from less hard processes. Hence one ask for the analyzed cluster to be associated to the most energetic particle in the event.

— Cut on the shape of the cluster : Figure 1 shows that photons' clusters have a low λ_0^2 value, i.e they are circular cluster. We clearly see on this figure that applying a cut at low λ_0^2 ($\lambda_0^2 < 0.3$ for this analysis) allows to keep almost all clusters coming from photons but reject most of the clusters coming from $\pi^0$and η , which have high λ_0^2 clusters.

— Cut on isolation : Photons of interest are emitted alone in their propagation direction such as there is a minor hadronic activity around them. One can define a cone around the photon candidate with a radius $R^2 = d\phi^2 + d\eta^2$ in which we measure the hadronic activity. We define an photon candidate as an isolated one if ADD THE CUT FOR ANTI ISOLATION, to be discussed if we keep ptmax or change to sum pT

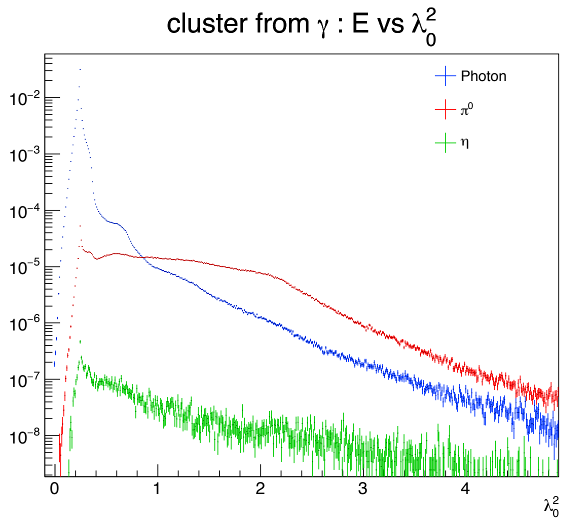

FIGURE 1 – λ_0^2 distribution for for photons, $\pi^0$and η . The second peak in photon distribution is due to the decay of π^0into two photons in an elliptic cluster.

Even after one applies these cuts on the initial sample of clusters, a non negligible amount of background clusters are still remaining in our sample of isolated and circular clusters. ADD PLOT OF CONTRIBUTIONS PROPORTIONS Hence it is needed to estimate what is the proportion of clusters that indeed comes from direct photons in this sample. This proportion is called purity and the next sections will explain the methods developed to estimate this purity.

2 Notations

First, let's present the notation that will be used in the next sections. The phase space in isolation and λ_0^2 is represented in figure 2. The zone A is the one in which we want to estimate the purity. The B, C and D zones allow us to estimate the background contribution in zone A as we will explain later (see section 3).

- The amount of particles is expressed as :
  - S : amount of direct photons
  - B : amount of background
    (π^0, η , gamma decays (π^0, η), ...)
  - N : S + B total number of particles
- The isolation criteria can be defined as CHECK THIS PART :
  - isol : no particle in isolation cone with $p_T > 0.5$ GeV/c (A, B)
  - $\neq$ isol : at least 1 particle in isolation cone with $p_T > p_T^{thres}$ (C, D)
- The circularity of the clusters required is pictured as TO BE DISCUSSED RESEND EMAIL :
  - $<$: round shape cluster $\lambda_0^2 < 0.3$ (A, C)
  - $>$: elliptic cluster $\lambda_0^2 > 0.3$ (B, D)
- We will split the background contributions and name them as follow :
  - π^0 : 2 photons coming from a $\pi^0$and merged into one single cluster (can be elliptic at low energy and circular at energy)
  - $\gamma_{\pi^0}^{paired}$: 1 photon coming from a π^0into one cluster which has it's partner in the isolation cone
  - $\gamma_{\pi^0}^{single}$: 1 photon coming from a π^0into one cluster which has lost it's partner (the partner is not in the isolation cone)

  The same notation will be used for the η

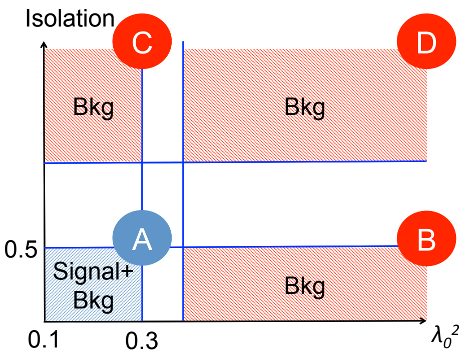

FIGURE 2 – The phase space in isolation and λ_0^2

3 Purity estimation

Now the notations have been presented. The purity definition can be given in terms of particles amount. The purity is defined as follow :

$$p = \frac{\text{Nb of direct photons' clusters}}{\text{Nb of isolated and circular clusters}} = \frac{S_<^{\text{isol}}}{N_<^{\text{isol}}} = 1 - \frac{B_<^{\text{isol}}}{N_<^{\text{isol}}} \quad (1)$$

The purity can't be estimated directly, hence one as to estimate the amount of background in zone A using the knowledge of background contributions in zones B, C and D.

3.1 First method

This method was the first to be developed (ref to Nicolas' thesis ?). In this method, one assumes there is a proportionality between the background contributions in the four zones. This assumption leads to this equation :

$$\frac{B_<^{\text{isol}}}{B_<^{\neq \text{ isol}}} = \frac{B_>^{\text{isol}}}{B_>^{\neq \text{ isol}}} \tag{2}$$

Replacing $B_<^{\text{isol}}$ in equation 1 we obtain :

$$p_1 = 1 - \frac{B_<^{\neq \text{ isol}}/N_<^{\text{isol}}}{B_>^{\neq \text{ isol}}/B_>^{\text{isol}}} \tag{3}$$

Unfortunately the assumption leading to this estimation canot be satisfied. For instance some background contributions exists at low λ_0^2 and not high λ_0^2 and vice versa. The $\pi^0$and η decay into two photons that can lead either in one elliptic cluster where the two decay photons are merged or two separated circular clusters. The second configuration implies that the photon candidate's cluster, around which we measure the hadronic activity, is less isolated as in the first configuration as the second decay photon participates in the hadronic activity around the candidate. This leads to a modification of the background ratio at low λ_0^2 in 2 in comparison to the ratio at high λ_0^2 . On figure **??** we show that the decay photons contribution to the background is not negligible and the modification of the ratio at low λ_0^2 can't be disregarded. On another hand several particles can be very closed to the photon and could formed with the latter a single elongated cluster (at high λ_0^2), called multi-contributions cluster (MCC) ref to alexis. Such a cluster is more isolated as if the particles surrounding the photon will have given separated clusters. In this case the background ratio at high λ_0^2 is modified compared to the one at low λ_0^2 . ADD PLOT OF MCC ?

We now know that this method can't be used anymore. Hence one has to find another way to estimate the photons purity.

3.2 Second method

A first thought is to use the simulation (MC) to correct the biases coming from the gamma decays and MCC contributions to the background. In order to do this one assumes the following :

$$\left(\frac{B_<^{isol}/N_<^{\neq isol}}{N_>^{isol}/N_>^{\neq isol}}\right)_{data} = \left(\frac{B_<^{isol}/N_<^{\neq isol}}{N_>^{isol}/N_>^{\neq isol}}\right)_{MC-JJ} \Leftrightarrow \left(\frac{B_<^{isol}/B_<^{\neq isol}}{B_>^{isol}/B_>^{\neq isol}}\right)_{data} = \left(\frac{B_<^{isol}/B_<^{\neq isol}}{B_>^{isol}/B_>^{\neq isol}}\right)_{MC-JJ} \tag{4}$$

This leads to a purity defined as :

$$p_2 = 1 - \left(\frac{N_<^{\neq \text{ isol}}/N_<^{\text{isol}}}{N_>^{\neq \text{ isol}}/N_>^{\text{isol}}}\right)_{data} \times \left(\frac{N_<^{isol}/N_<^{\neq isol}}{N_>^{isol}/N_>^{\neq isol}}\right)_{MC(JJ)} \tag{5}$$

This assumption means that there is no signal in B, C and D zones which is obvious and the isolation of the background is well reproduced by the simulation in the four zones. ADD PLOT fisol data vs MC – how do we deal with this raise at low M02 ? see figure 3

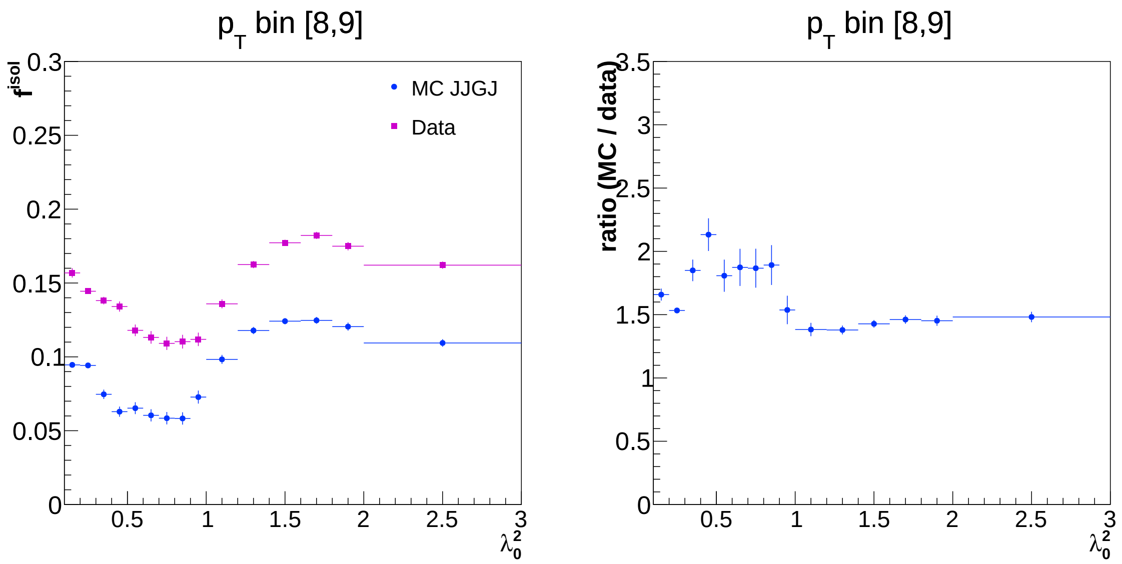

FIGURE 3 – Isolation fraction in MC and data (left panel) and the ratio between the two (right panel).

3.2.1 Closure test and signal contamination

To first test the validity of the method one can perform a closure test. The closure test consist in replacing the data term in equation 5 by a GJ + JJ mixing. If the closure test succeed the purity found with p_2 (p_{output}) should be the same as the true purity (p_{input}), as defined in equation 1, using GJ + JJ simulation.

As a reminder the original cut for anti-isolation is $p_T > 1.0$ GeV/c CHECK and the cut in λ_0^2 is 0.3. Using these cuts we obtain the closure test presented in figure 4. We clearly see

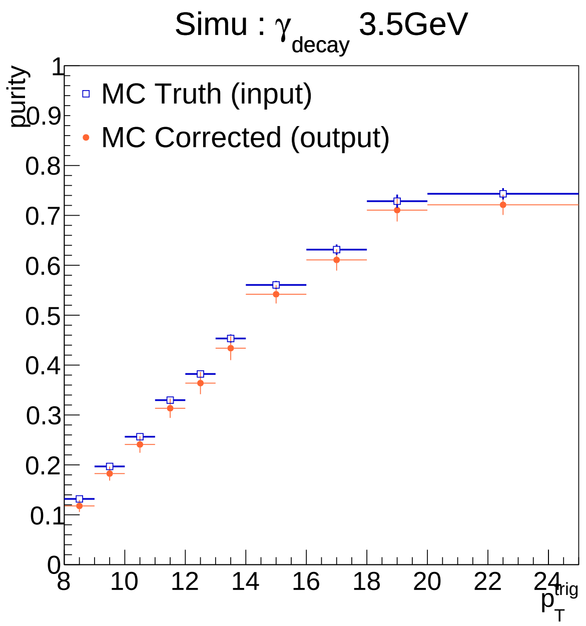

FIGURE 4 – Comparison of the true purity (blue) and the corrected purity using p_2 method applied on simulation (orange). The JJ simulation used for the correction term is the gamma decay triggered with the threshold of 3.5 GeV/c. The closure test fails here as the two curve are not compatible.

the closure test fails. One can assesses that the difference between p_{input} and p_{output} is due to the presence of the signal in the background regions (B, C and D)dire que ça va dans le bon sens. Figure 5 shows that indeed there is signal (black points) in background regions. The contamination is not negligible in B and C zones where the signal proportion grows up to 10% in zone B and 5% in zone C. The aim now is to find cuts that allow to get rid off signal in background regions to estimate the latter properly.

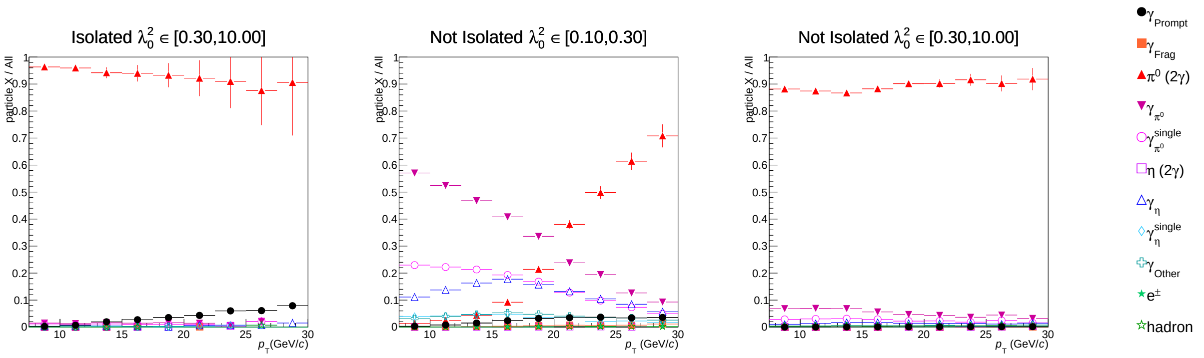

FIGURE 5 – Proportion of particles in B (left), C (center) and D (right) zones. The signal is represented by the black points (prompt photons).

A first approach is to implement the banana shape from ADD REF TO GUSTAVO'S AN to remove the major part of signal. Figure 6 shows the shape of such a function. One can see this function allows to remove the signal λ_0^2 distribution tail after 0.3. Even with this change a small signal contamination remains and the low threshold of the banana shape as be 0.6 and change the anti isolation cut to 3.0 GeV/c to get rid off any signal contamination in the background regions (see figure 7).

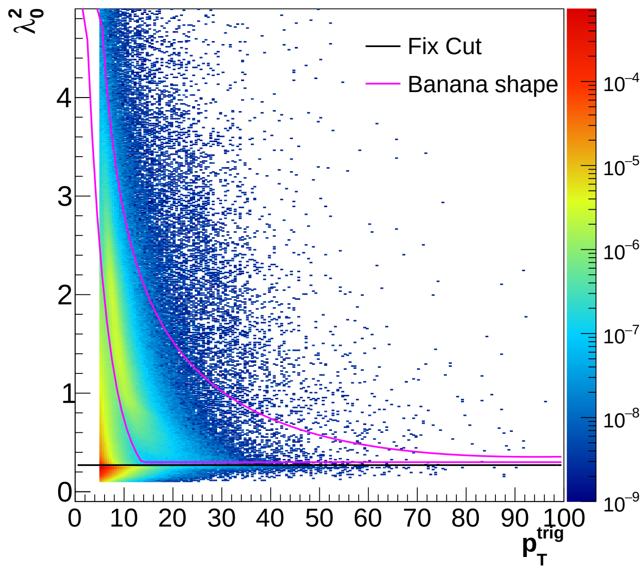

FIGURE 6

One can redo the closure test with the new tight cuts. The result is shown on figure 8 where we see a good agreement between p_{input} and p_{output}.

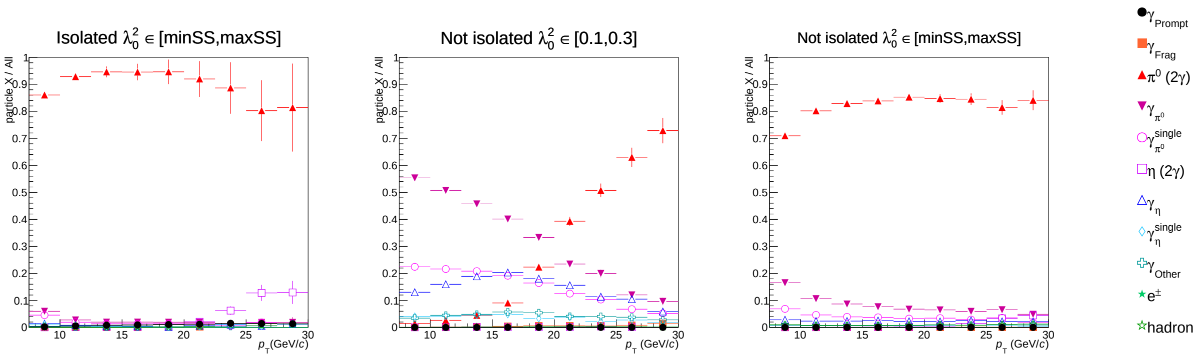

FIGURE 7 – Proportion of particles in B (left), C (center) and D (right) zones. The signal contamination is now negligible in the three background regions

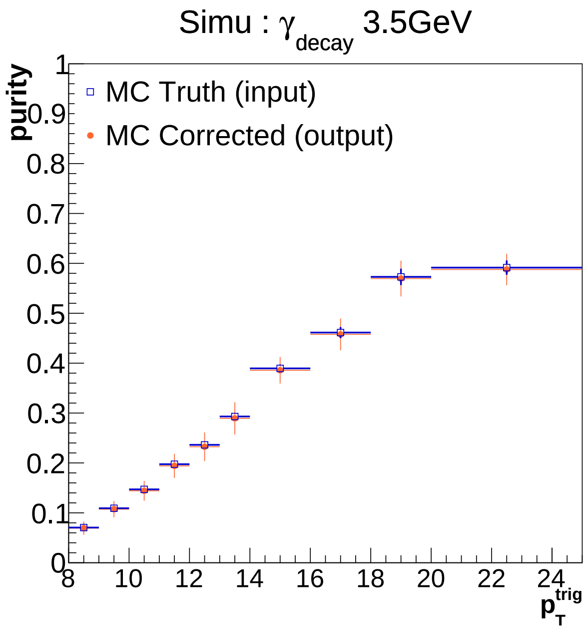

FIGURE 8 – Comparison of the true purity (blue) and the corrected purity using p_2 method applied on simulation (orange). The JJ simulation used for the correction term is the gamma decay triggered with the threshold of 3.5 GeV/c. The closure test succeed when using tight cuts for the anti-isolation and λ_0^2 limit.

One can also look at the relative difference between the two purities to see that the cuts chosen are the best ones. Looking at figure 9 we see that for a large range in p_T the chosen cuts allow to have a very small relative difference while not killing all the statistic of the sample.

These results prove that the p_2 method has to be used with tight cuts to obtain a non bias result.

3.2.2 Results for data

One applies now the p_2 method on data to obtain the purity. As several JJ simulations exist we have to first checked that they give compatible results in their validity area : do we take pi0 trig ?

- γ_{decay} 3.5 GeV/c : valid in all analysis p_T range
- γ_{decay} 7 GeV/c : unbiased only after 16 GeV/c add plot from Julien ?

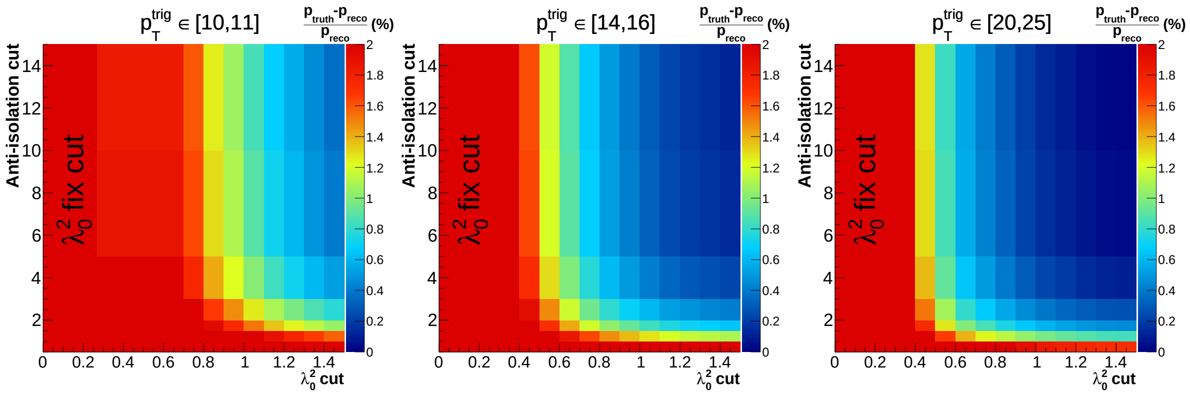

FIGURE 9 – Relative difference between p_{input} and p_{output} for p_2 method. The relative difference is shown for three p_T bins : [10,11] (left), [14,16] (center) and [20,25] (right).

Figure 10 shows the purity obtain with a correction from these two JJ simulation. We can see that the results are compatible after 16 GeV/c. [h !] As the two simulation are compatible we

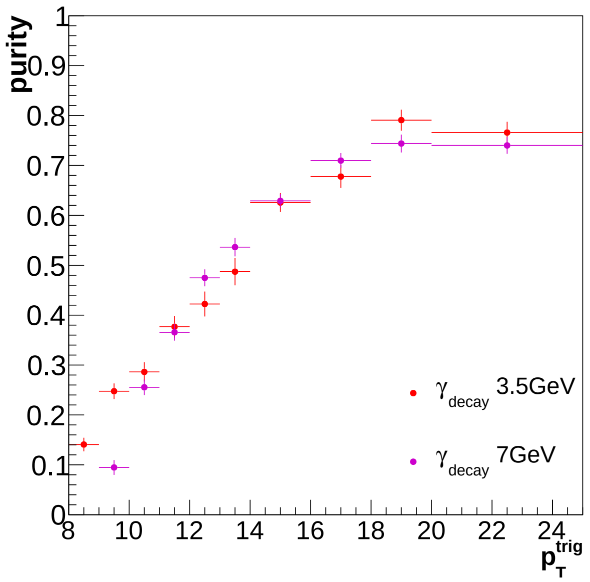

FIGURE 10 – Purities for a correction factor using γ_{decay} 3.5 GeV/c simulation (red) and γ_{decay} 7 GeV/c simulation (magenta).

used, in order to reduce statistical error, the low threshold simulation up to 16 GeV/c and the high threshold simulation after this value. The final result for p_2 method applied on data is given on figure 11.

3.3 Third method

This method also relies on a correction coming from simulation to correct biases coming from gamme decays and MCC, but here the correction term is extracted from GJ + JJ mixing simulation. Here one assumes :

$$\left(\frac{N_{<}^{\neq isol}/B_{<}^{isol}}{N_{>}^{\neq isol}/N_{>}^{isol}}\right)_{data} = \left(\frac{N_{<}^{\neq isol}/B_{<}^{isol}}{N_{>}^{\neq isol}/N_{>}^{isol}}\right)_{MC-GJ+JJ} \tag{6}$$

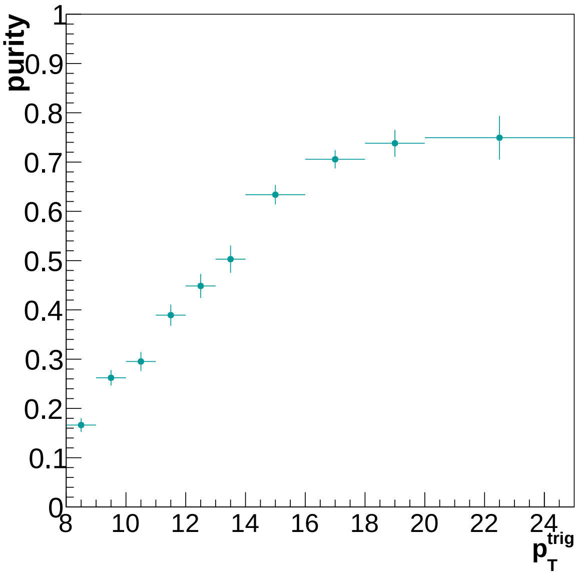

FIGURE 11

This leads to purity expressed as follow :

$$p_3 = 1 - \left(\frac{N_<^{\neq \text{ isol}} / N_<^{\text{isol}}}{N_>^{\neq \text{ isol}} / N_>^{\text{isol}}} \right)_{data} \times (1 - p_{\text{MC}}^{truth}) \left(\frac{N_<^{isol} / N_<^{\neq isol}}{N_>^{isol} / N_>^{\neq isol}} \right)_{MC(GJ+JJ)} \quad (7)$$

With this method 2 main assumptions are made :

- The signal contamination in the background regions is the same in data and simulation (GJ+JJ)
- The isolation fractions ($\frac{N^{isol}}{N^{noisol}+N^{isol}}$) are the same at low and high λ_0^2 regions

The validity of the assumptions will be discussed later on. Note that one can't perform a closure test on this method because one will go back to p_{input} by construction.

3.3.1 Results for data

For this method the standard cuts will be used (no need to care about signal contamination if one trust first assumption), i.e. anti isolation cut at 1.0 GeV/c and λ_0^2 limit at 0.3 without banana shape function. As for p_2 one will use the two JJ simulation (see section 3.2.2) to construct GJ+JJ simulation, γ_{decay} 3.5 GeV/c below 16 GeV/c and γ_{decay} 7 GeV/c above this value. The result obtain for p_3 is shown on figure 12.

3.3.2 Hypothesis validity

To test the validity of the first assumption one can compare the results from p_3 with standard cuts with p_2 with tight cuts which, we know, is free of signal contamination. If the signal contamination is well reproduce in MC compared to data, then using GJ+JJ mixing should correct from signal contamination and the two methods with these specific cuts should be compatible. On figure 13 we see that the results are very close but this is not a perfect match as one could expect if the signal contamination is corrected with GJ+JJ simulation. This shows that a GJ+JJ mixing simulation may not be enough to correct signal contamination in data in background regions.

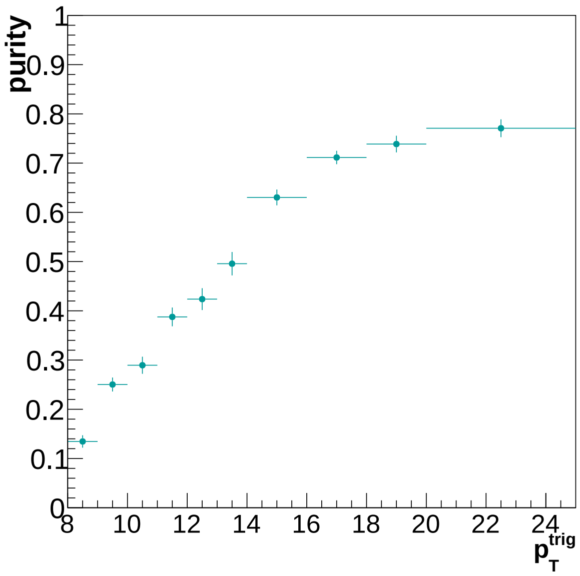

FIGURE 12

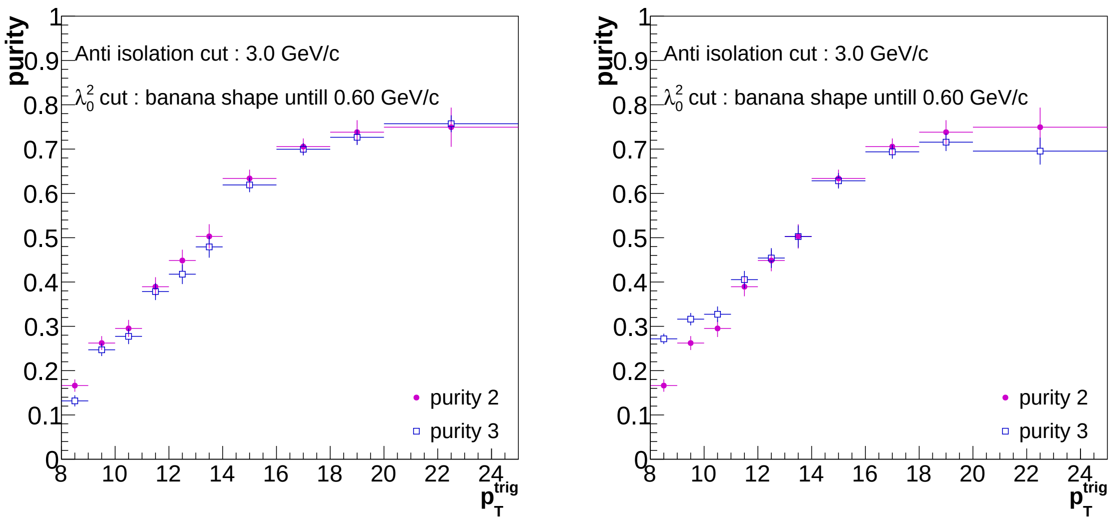

FIGURE 13

3.4 Fourth method

3.5 Comparison of the methods

4