Qualification Task: AFT-487

Martino Centonze

May 7, 2020

Acceptance of the track association algorithm

The idea is to investigate how much the combined processes of track selection and track association are able to detect the decay products of b hadrons and quantify how much the overall efficiency of these algorithms affect the performance of our taggers.

Tracks to jets association: Shrinking Cone Association.

$$\Delta R_{cone}(p_T) = P_1 + \exp(P_2 + P_3 * p_T) \tag{1}$$

where $P_1 = 0.239$, $P_2 = -1.220$, $P_3 = -1.64 * 10^{-5}$ and p_T (in [MeV]) is the transverse momentum of the jet. From MonteCarlo truth: decay products of b hadrons –> "children".

The goal: for each child, look for a matching track within the sample of tracks associated to the b jet.

The algorithm

1. select the b tagged jets from MC truth,
2. use the pdg-Id of tracks and children, and those of the respective parents, in order to find all possible matchings according to the following criteria:

$$pdg\_Id(track) = pdg\_Id(child) \tag{2}$$

$$pdg\_Id(parent\_track) = pdg\_Id(parent\_child) \tag{3}$$

3. in case of multiple matchings to the same child or to the same track (*i.e.* one child is associated to multiple tracks or viceversa), select the *best* one, by minimizing (at the same) time the following quantities

$$|\Delta p_T| \equiv |p_{T,child} - p_{T,track}| \tag{4}$$

$$\Delta R \equiv \sqrt{(\Delta\phi_{track,child})^2 + (\Delta\eta_{track,child})^2} \tag{5}$$

Here, I stress the fact that there are no cuts applied in this algorithm (spoiler: I'll introduce cuts later on, motivating them).

Results

number of children: 804481
number of matched children: 661624
number of children without match: 142857
average efficiency: 0.82242
number of children with one single match: 373042 (a relative fraction of 0.56383 of the number of total matches)
number of double matches: 243837 (relative fraction of 0.36854)
number of triple matches: 31532 (relative fraction of 0.04766)

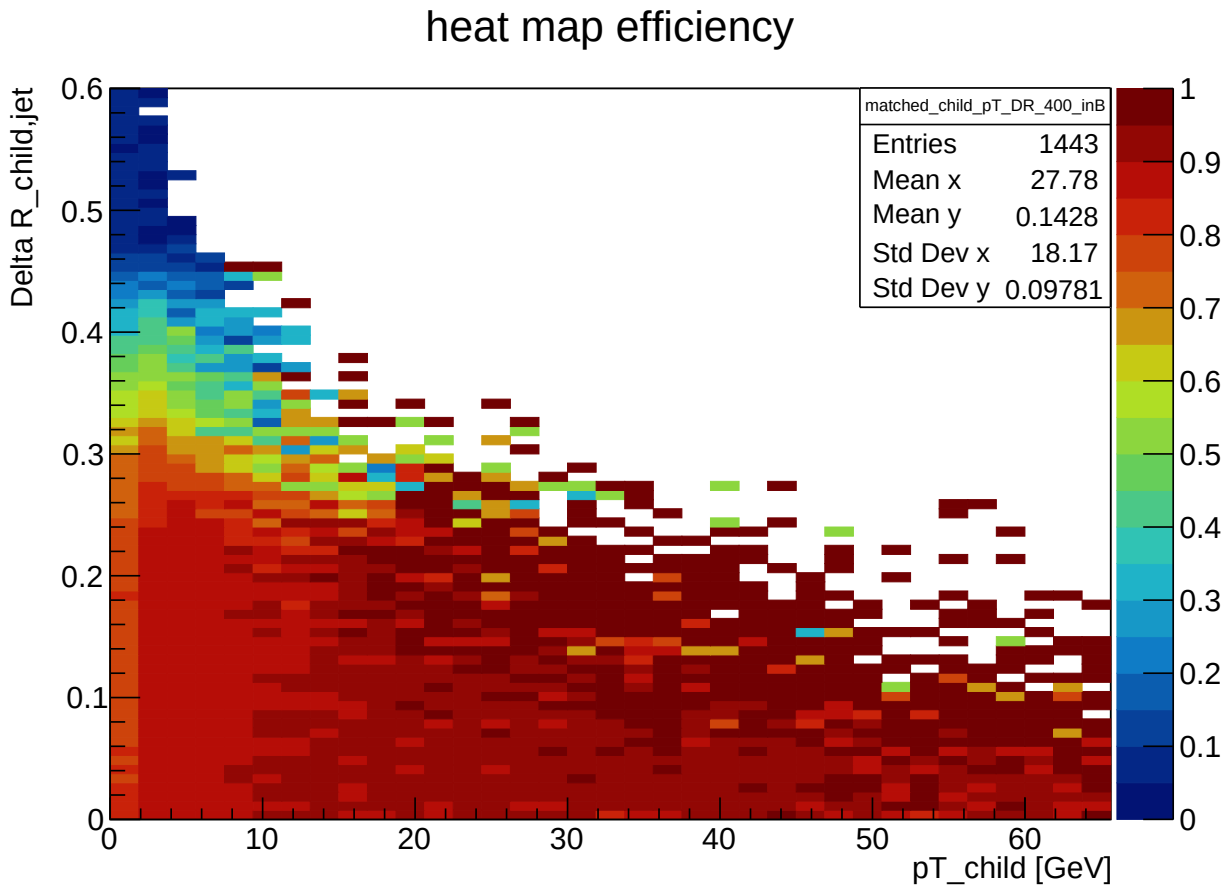
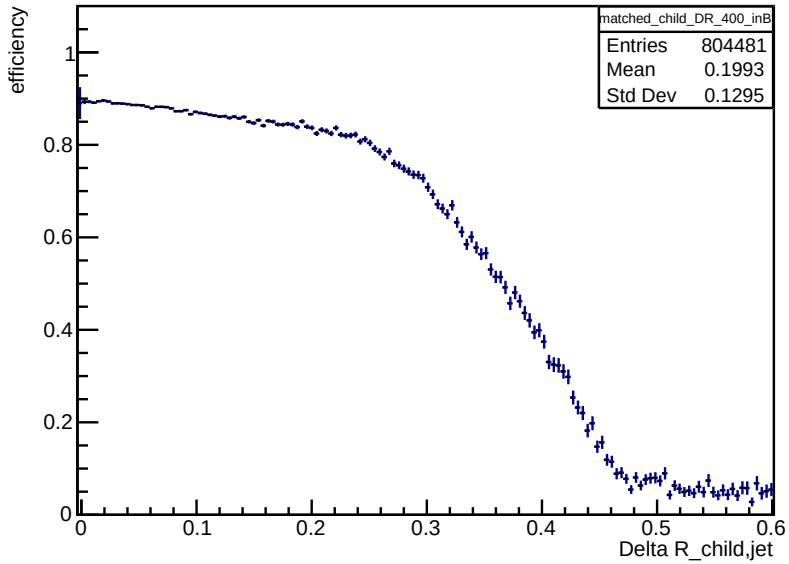
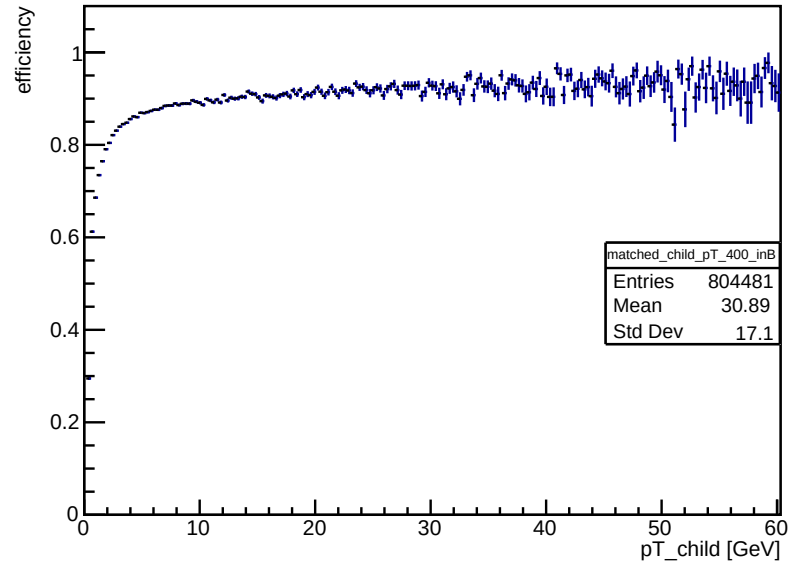


Figure: On the right of the plot, the colour scale relative to the values of the efficiency ϵ. The efficiency is estimated by dividing the set of matched children by the set of all children, for each b jet.

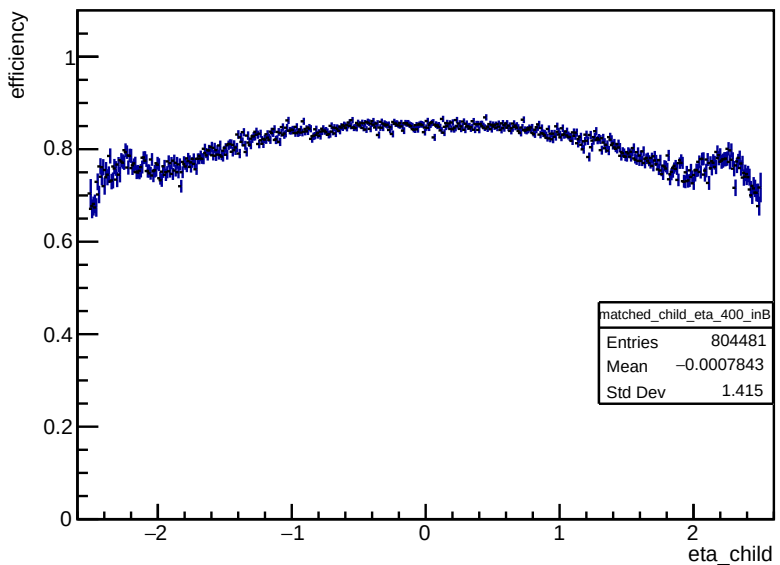
efficiency_vs_DR



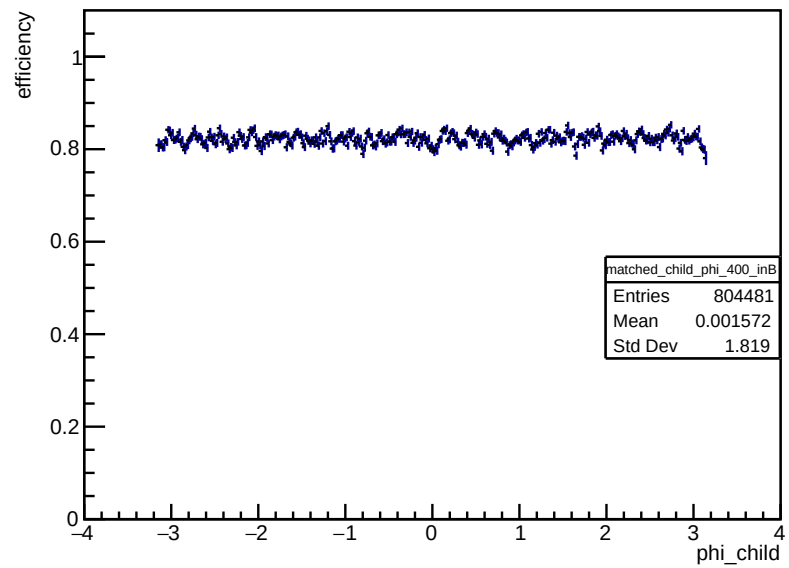
efficiency_vs_pT



efficiency_vs_eta



efficiency_vs_phi

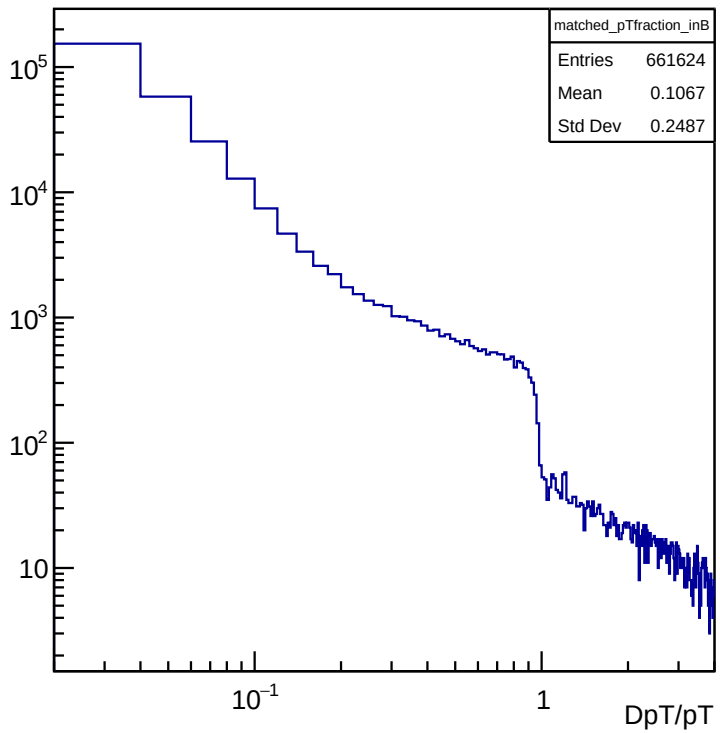


The results show the presence of "bad" correspondences, namely for:

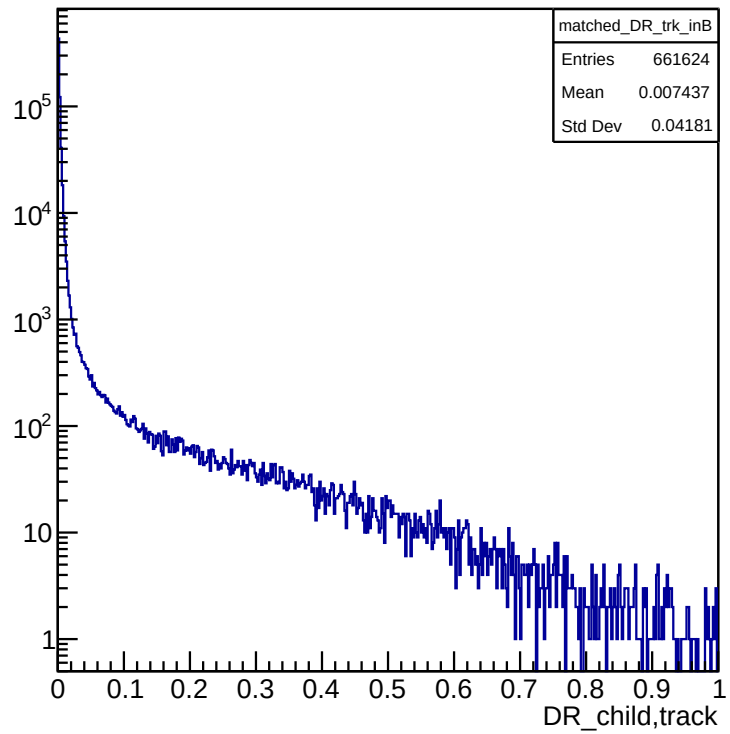
$$\frac{|\Delta p_T|}{p_{T,child}} > 1$$

$$\Delta R_{child,jet} > 0.5$$

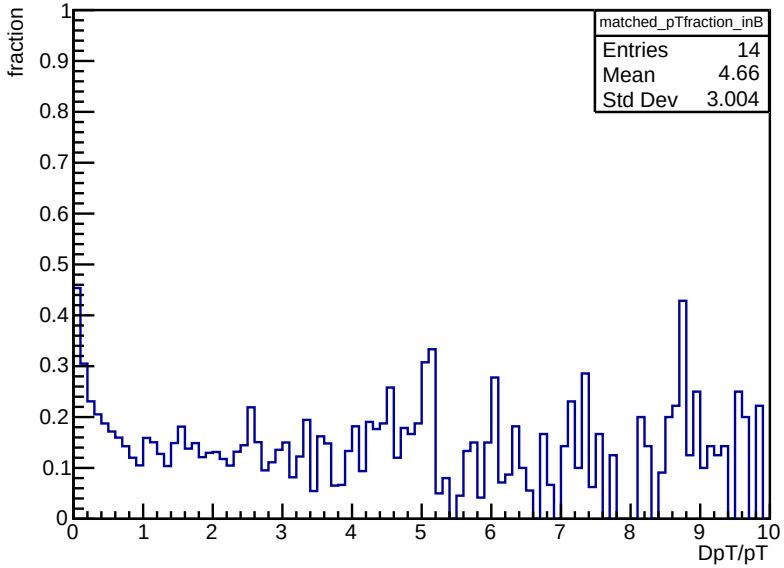
matched_DpT/pT_child_inB



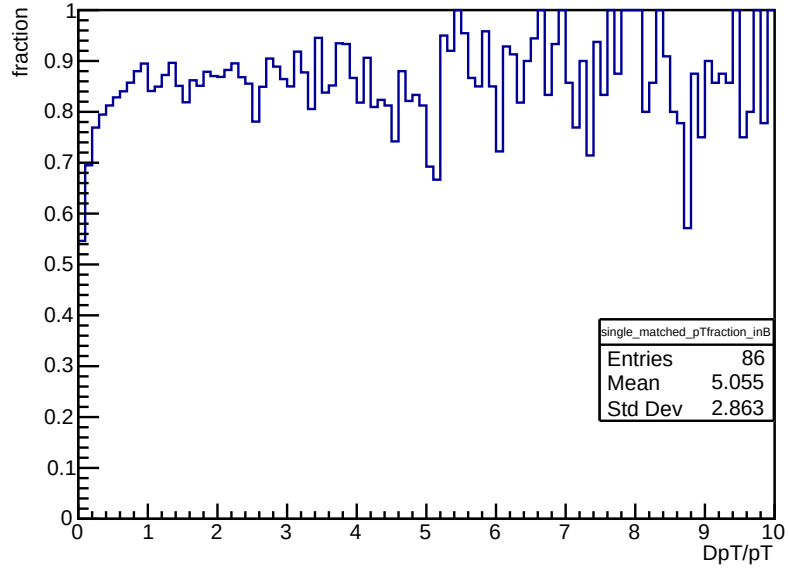
DR



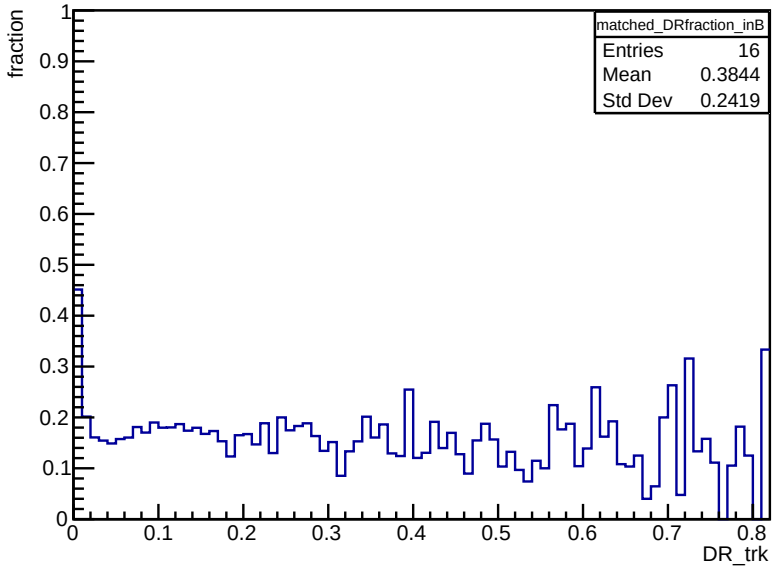
(matched-singlematched)/matched_vs_DpT/pT



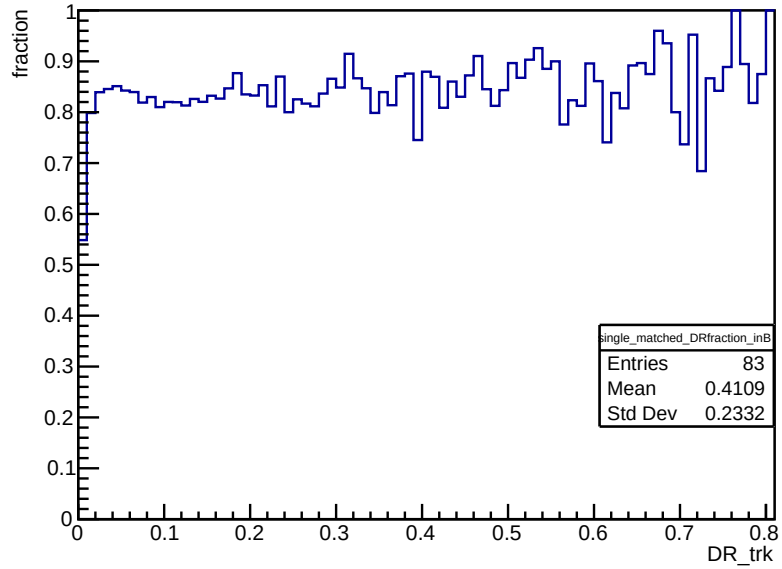
singlematched/matched_vs_DpT/pT



(matched-singlematched)/matched_vs_DR_trk



singlematched/matched_vs_DR_trk



About the 85% of these "bad" matchings come from single matches, which have no input control on $|\Delta p_T|/p_{T,child}$ and ΔR.

Proposal:

Apply a cut on $|\Delta p_T|/p_{T,child}$ and ΔR:

$$\frac{|\Delta p_T|}{p_{T,child}} < 1 \tag{6}$$

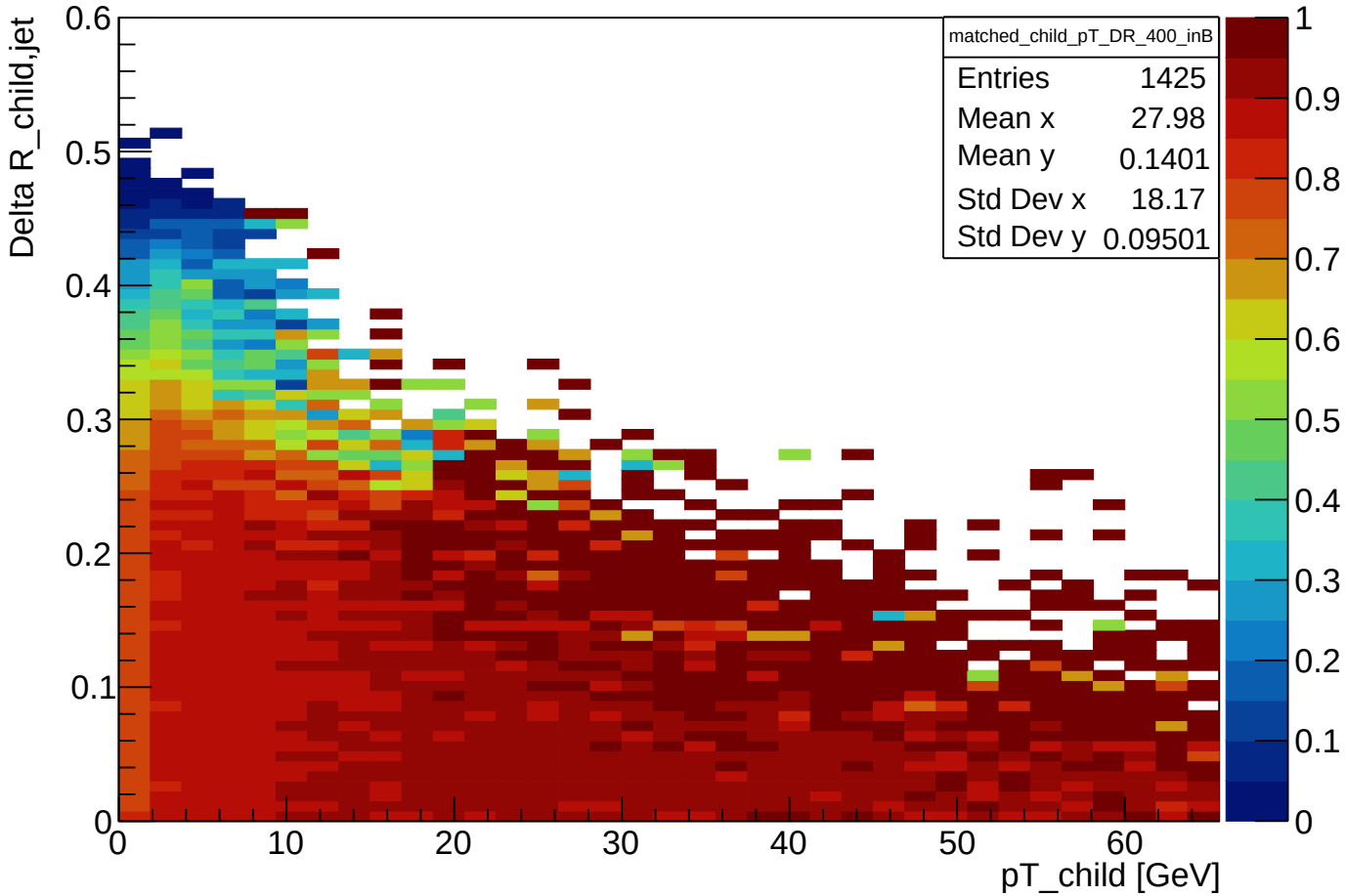
$$\Delta R < 0.1 \tag{7}$$

Here I also modify the algorithm, by changing one of the two minimization criteria, that is $|\Delta p_T|/p_{T,child}$ instead of $|\Delta p_T|$ of eq. 4. Results:

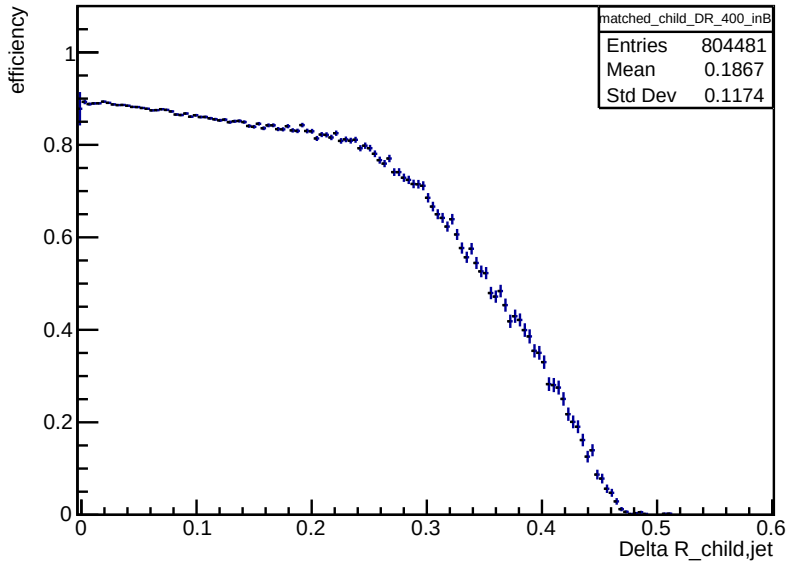
number of children: 804481
number of matched children: 653602
number of children without match: 150879
average efficiency: 0.81245
number of children with one single match: 366818, (a relative fraction of 0.56123 of the number of total matches)
number of double matches: 241245, (a relative fraction of 0.36910 of the number of total matches)
number of triple matches: 32156, (a relative fraction of 0.04920 of the number of total matches).

The average efficiency is stable: ~ 0.81 vs ~ 0.82.

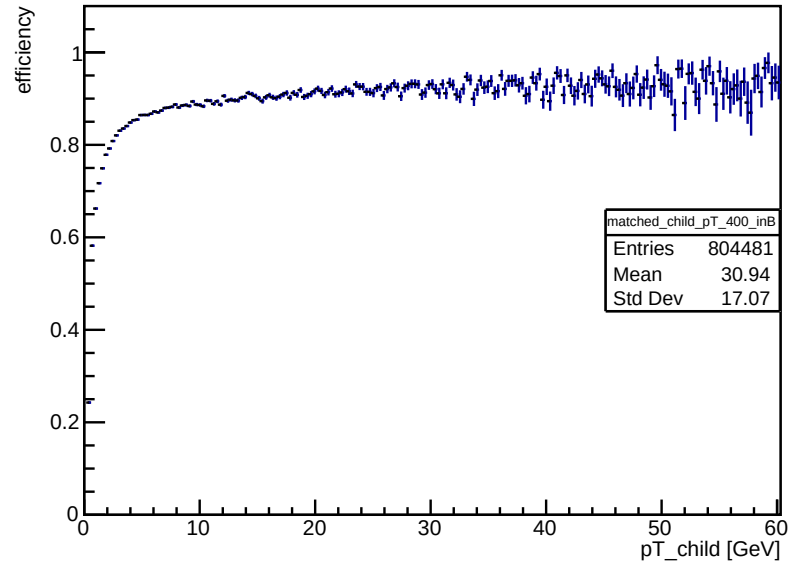
heat map efficiency



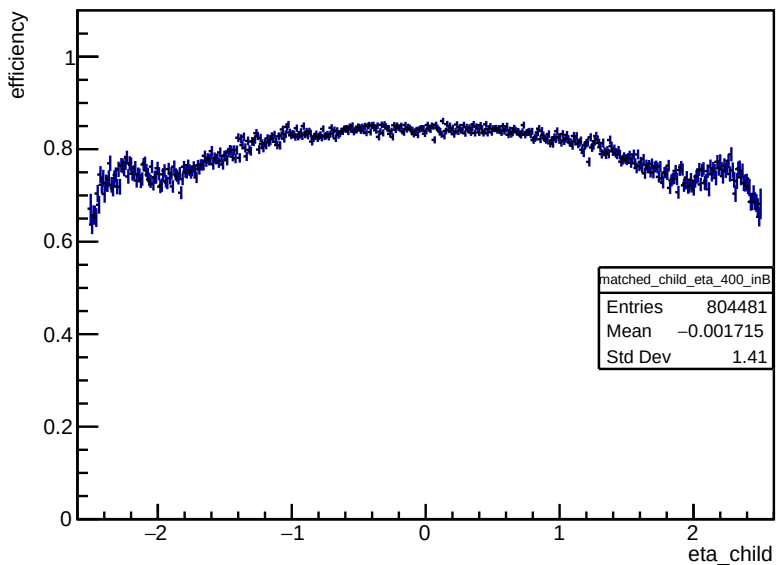
efficiency_vs_DR



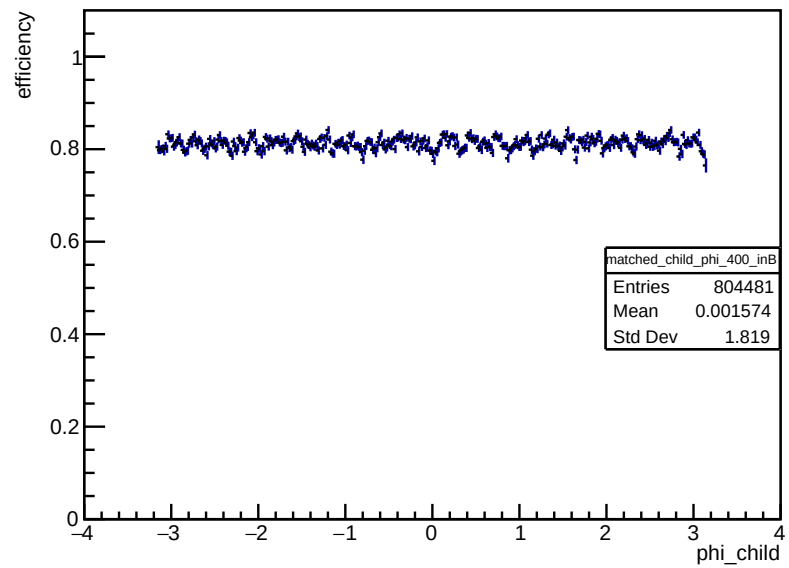
efficiency_vs_pT



efficiency_vs_eta

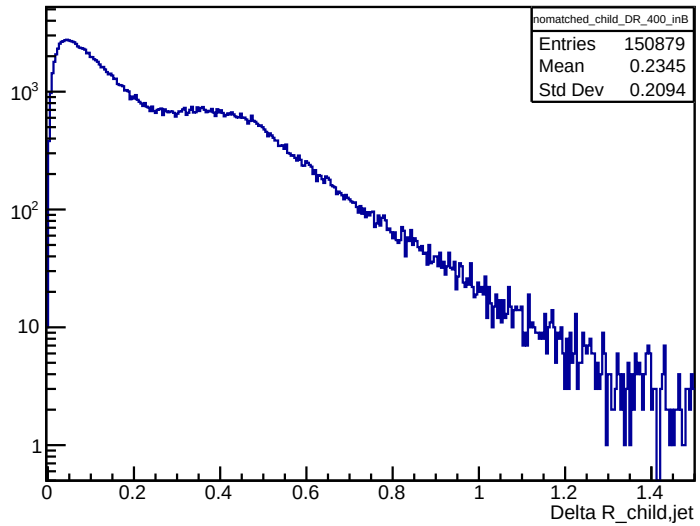


efficiency_vs_phi

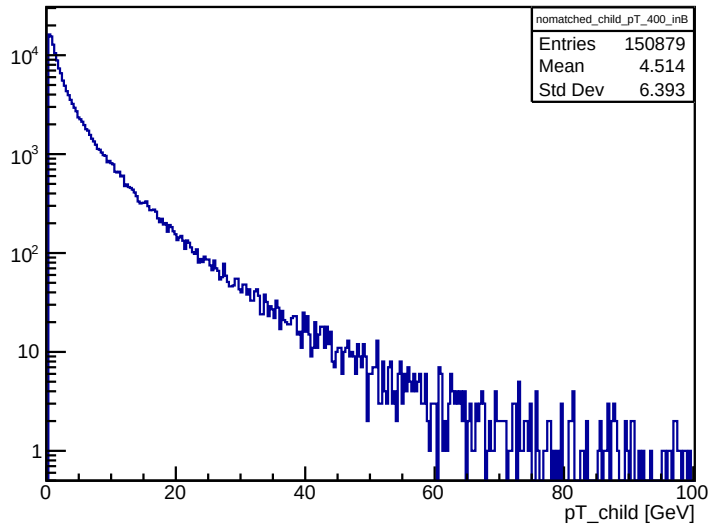


A look at the unmatched children

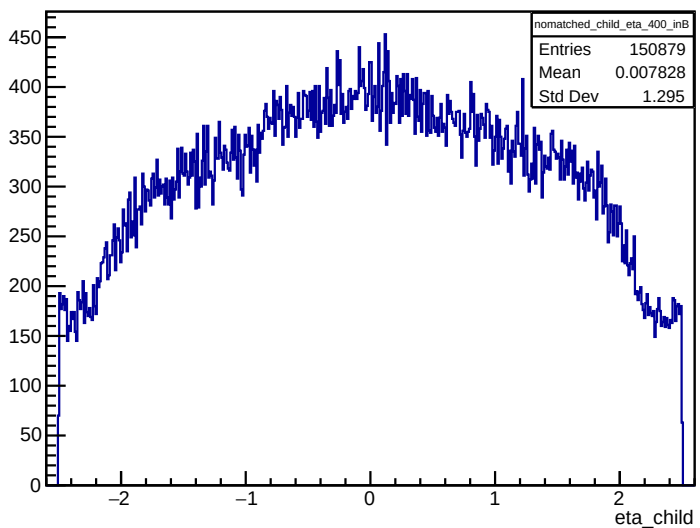
nomatched_child_DR_400_inB



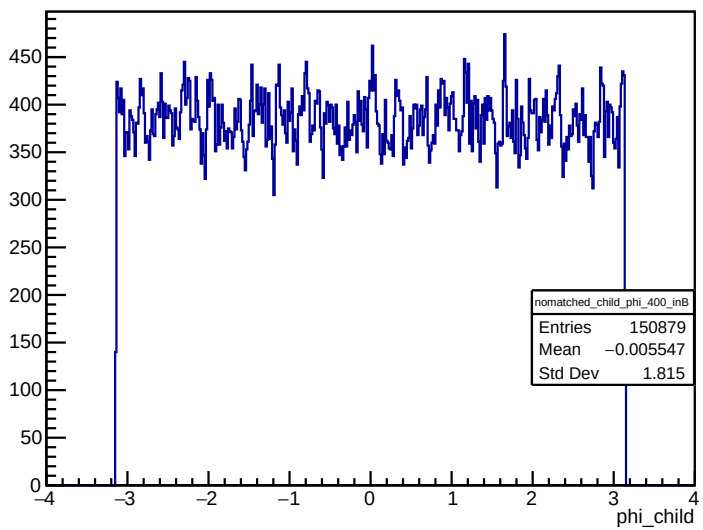
nomatched_child_pT_400_inB



nomatched_child_eta_400_inB



nomatched_child_phi_400_inB



A look at the unmatched children

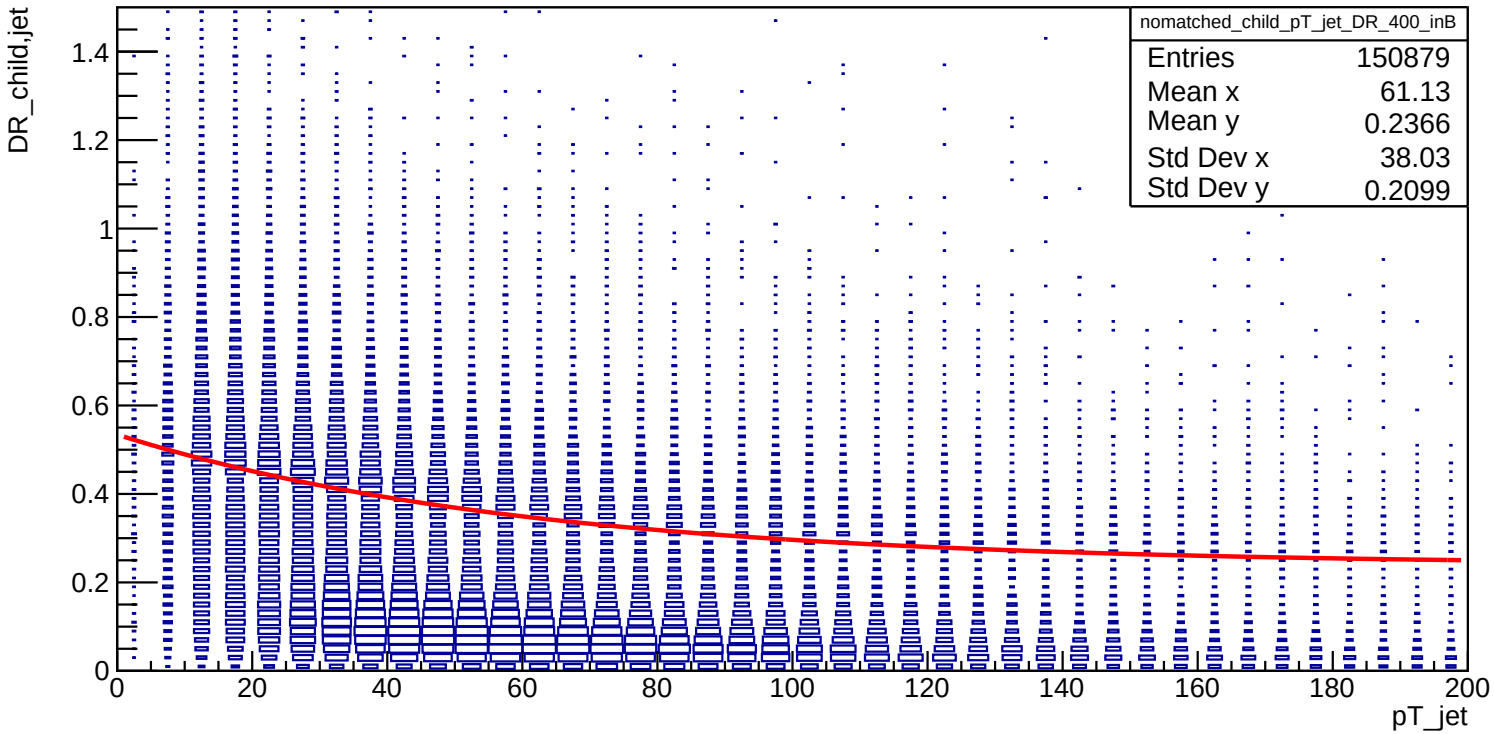


Figure: Here the number of unmatched children is plotted against the pT of the jet and DR, while in red is plotted the cone size (eq. 1). Clearly visible the pT dependent DR cut, but most of the the lost children are at small DR. An effect of track selection?

Conclusions

What is left to do:

1. select a stable criterion for performing the track-child association, in order to standardize the measure of the efficiency ϵ,
2. use the algorithm in order to measure the efficiency of different track-jet association schemes, analysing eventual differences in ϵ,
3. allow comparison of the performance of track association algorithms.

My code is available in my public repository on GitHub $\rightarrow$ `https://github.com/martinosal/root_selector`

Thank you!