Plans

- Taggers: IPxD, SVF, JetFitter
  - IPxD: re-tune with inclusive dijet samples:
    - Weights added.
    - Centrality dependence added.
    - Setup LOCALGROUPDISK space for inclusive n-tuples.
    - templates and evaluation using different pT cuts .
  - SVF & JetFitter Performance: physics variables:
    - Added comparison with light jet
    - Simplified cuts being compared
    - ROC curve for vertexing efficiency of different tagger & centrality

B-Tagging Workflow

Track to Jet Association:
ParticleJetTool

ReTag

Jet Fitter:
Select Tracks (pT, quality, etc.)
Fit for common SV

Vkal SVF:
Select Tracks (pT, quality, etc.)
Fit for common SV
Find V0 tracks

IPxD:
Select Tracks (exclude V0 tracks)
Assign probablity

Physics Variables

DL1

-track pass **TrkSelectionTool**
-associate each track to closest jet
-if dR < threshold, track is associated
Shrinking Cone: higher jet pT, smaller the threshold dR
Fixed Cone: threshold dR = 0.4 for all jet pR

Pre-Tagging Selection

Default Setup:
Shrinking Cone, min track pT > 0.5 GeV

MC Samples

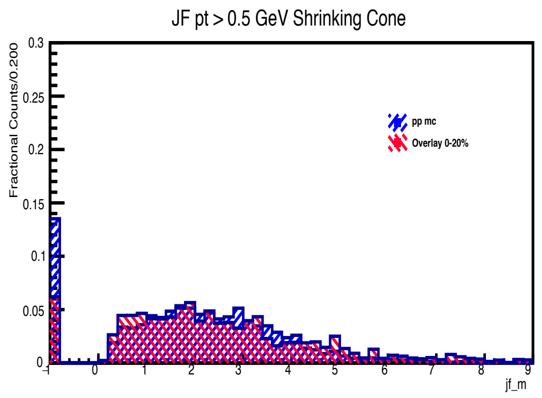
- pp MC and MC overlay (JetFitter and SV1 plots):
  - pp MC: 50k events (12.5k each for JZ1-JZ4) of pythia dijets events at 5.02 TeV, applied with bbar filter Selection on Jets.
  - Configuration file: https://gitlab.cern.ch/atlas-physics/pmg/infrastructure/mc15joboptions/blob/master/share/DSID420xxx/MC15.420271.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ1_bbfilter.py
  - Overlay: pp MC + 2018 minBias data to simulate underlying events.
- pp Inclusive dijets samples (IP3D):
  - https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HIJetMCSamples#Pythia8_dijets_8M_per_sample_in
  - Pythia8 dijets - 8M per sample in 21.0.93

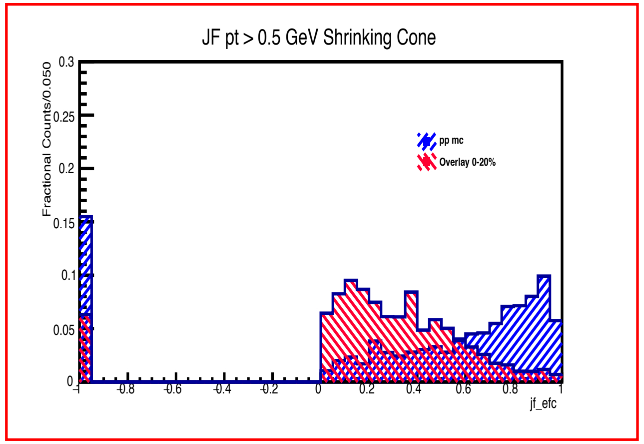
- Selection on Jets:
  - Reco jets with ΔR(truth-reco) < 0.3
  - $p_T^{truth\ jet} > 50$ GeV or 100 GeV (see back up plots)

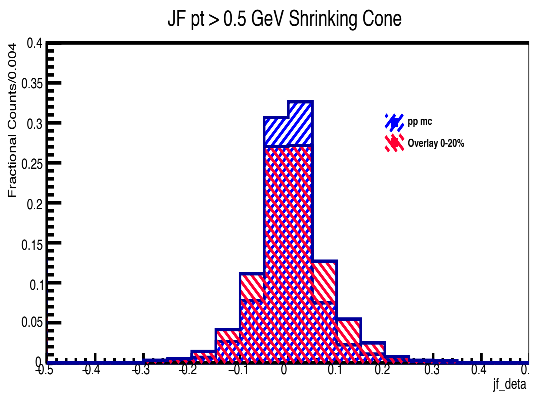
- B-Jets: jets with a truth B hadron associated with it. Similarly for C-jets
  - $p_T^B > 5$ GeV
  - ΔR(jet-BHadron) < 0.3

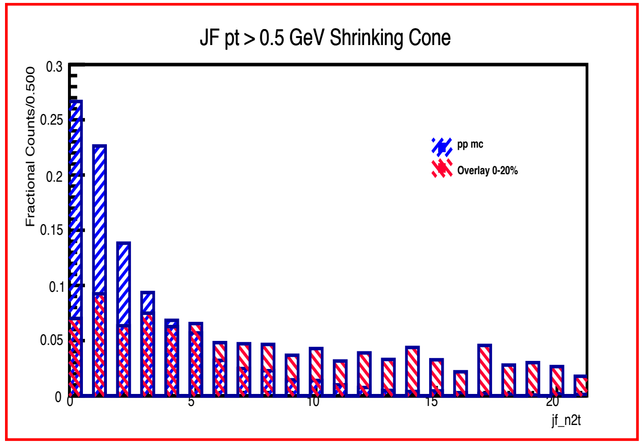
- Tool: https://gitlab.cern.ch/stapiaar/tagging_framework_hi/tree/master/
  - The most updated modified version is at https://gitlab.cern.ch/xiaoning/hiretagging_framework

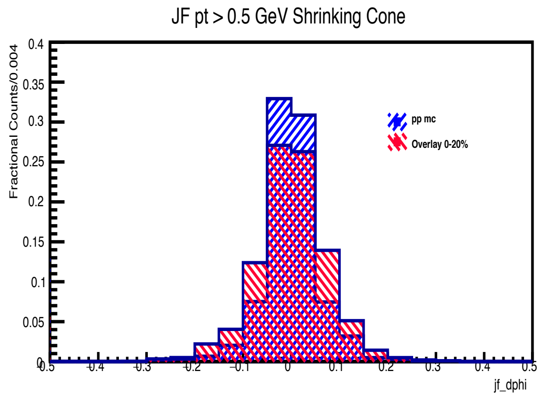
JF Variables at Default Setup

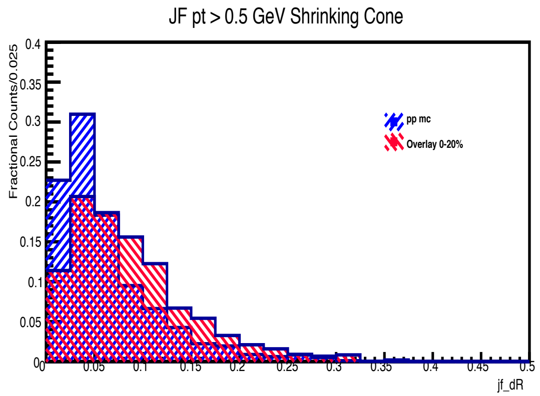


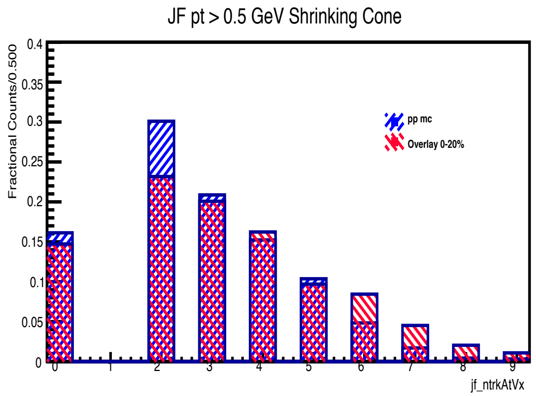




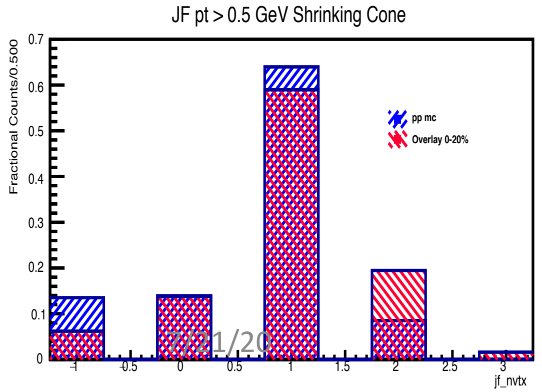


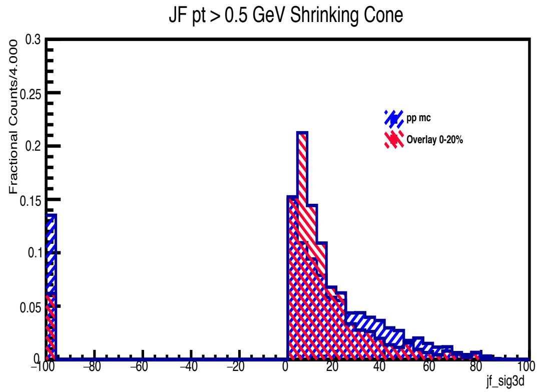


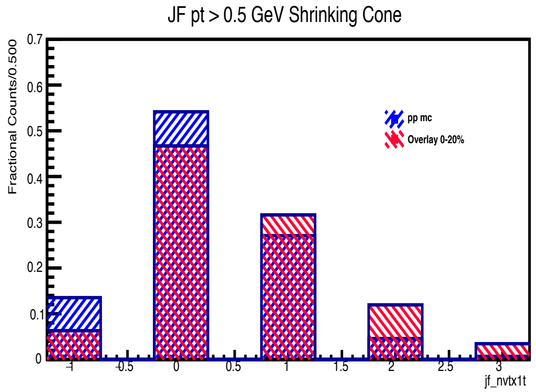




- Energy fraction and number of 2-track vertices are heavily modified.



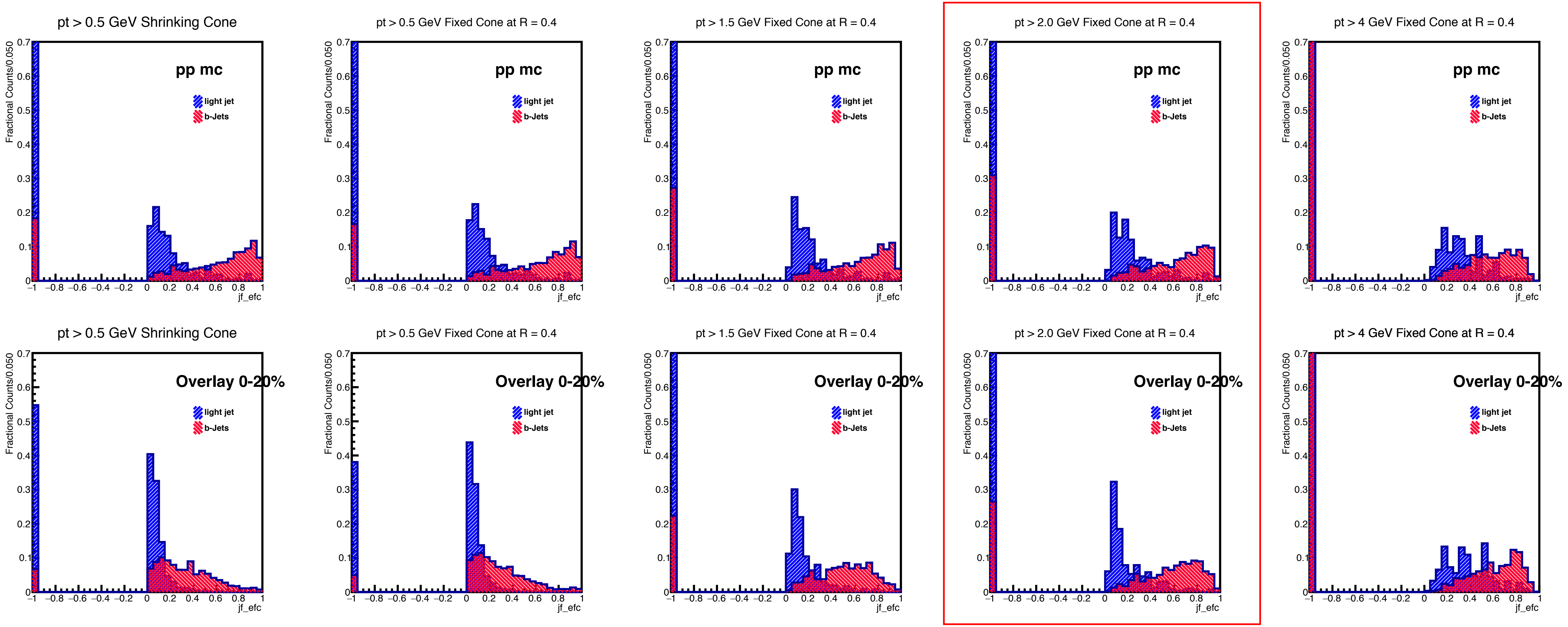




JF Energy Fraction

↓ Visually the two distributions are the most distinct at 2 GeV

↓ over cutting causes light jet to right-shift as well



Top Row: pp mc
Bottom Row: Overlay 0-20%

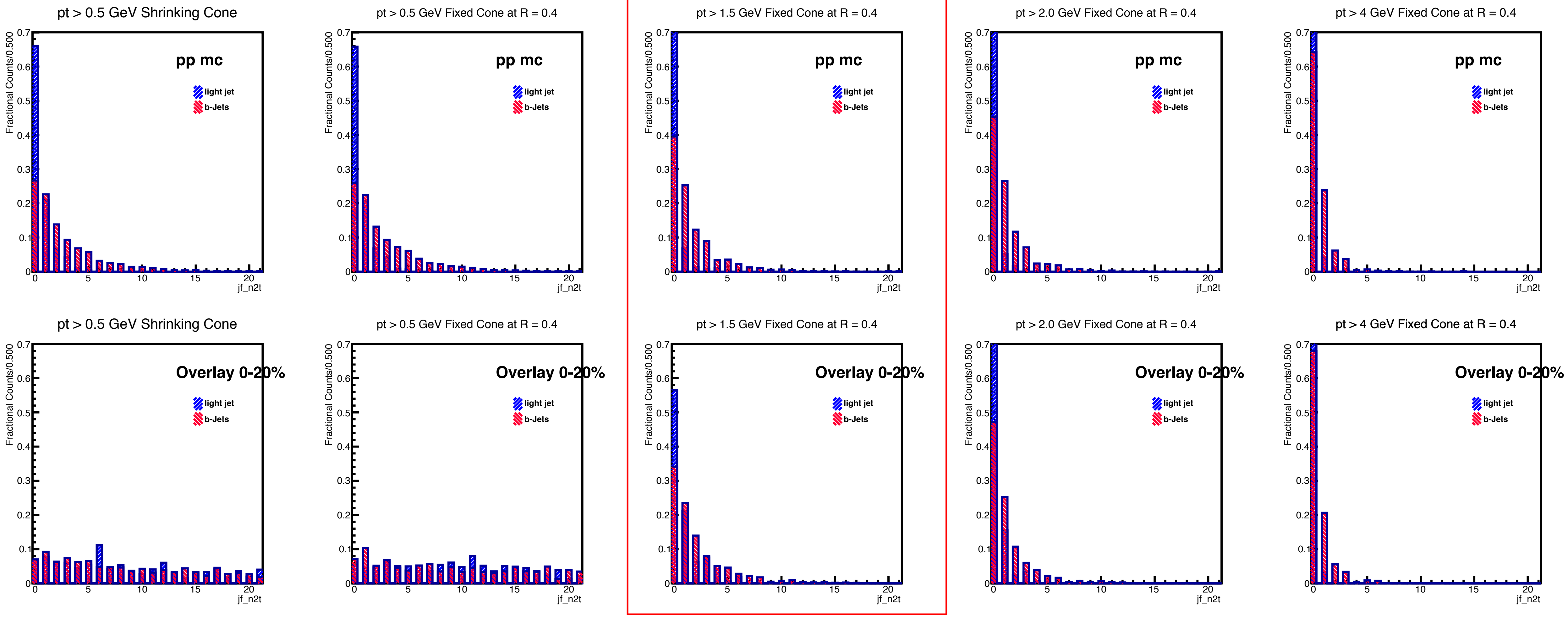
From left to right:

1. No pT Shrinking Cone
2. No pT Fixed Cone
3. Min pt = 1.5 GeV Fixed Cone
4. Min pt = 2.0 GeV Fixed Cone
5. Min pT = 4.0 GeV Fixed Cone

Red: b-jet
Blue: light jet

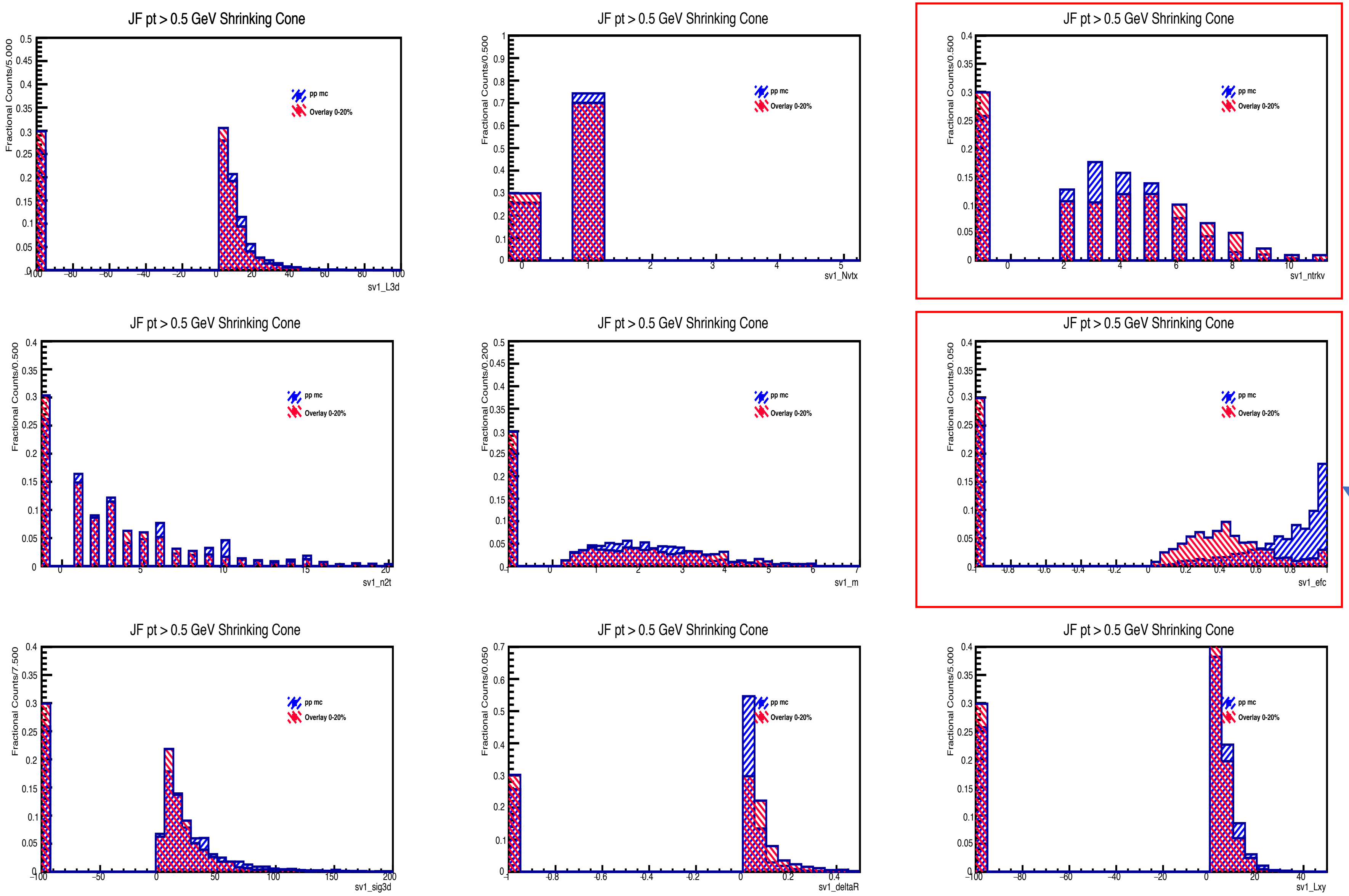
JF n2t (2-trk vertices candidates)

↓ over cutting leaves too few bins



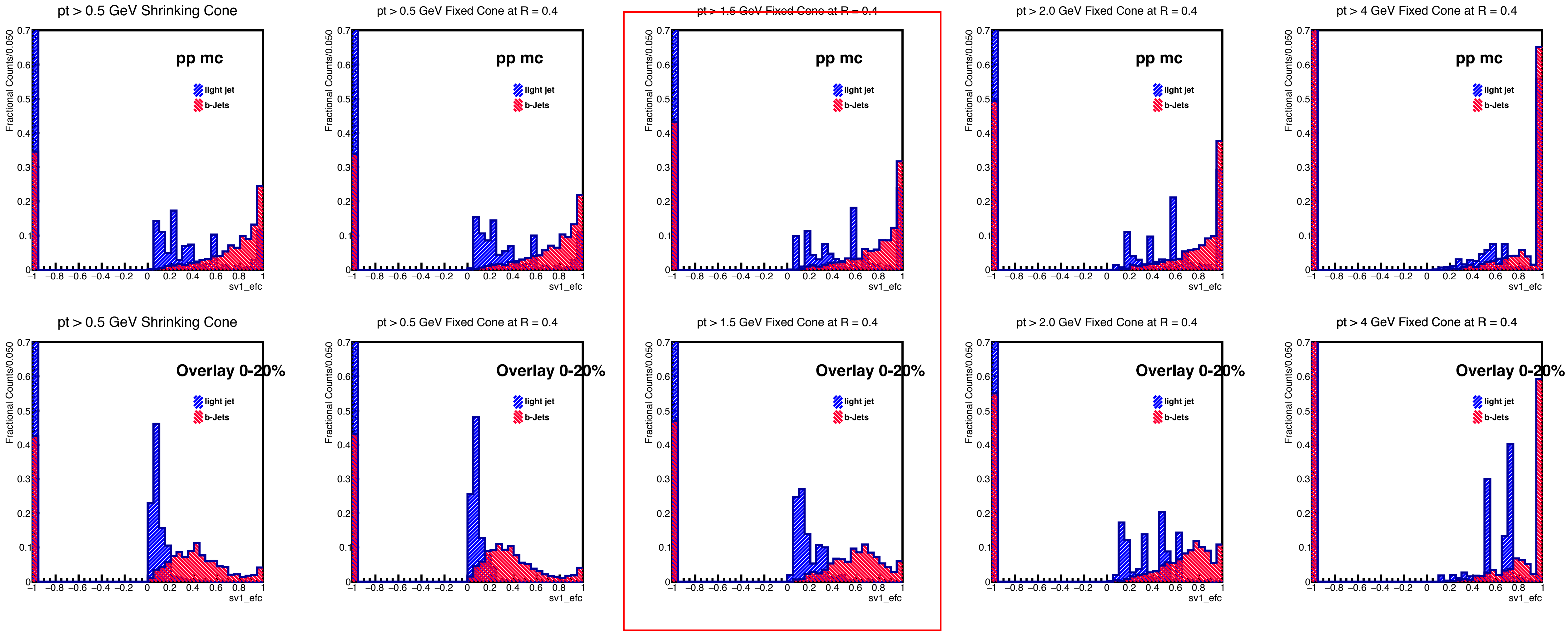
- Before applying cuts in pt, distribution is even, possibly for combinatorics of UE tracks.
- Starting at 1.5 GeV or above, overlay have a similar distribution as pp.

SV Variables at Default Setup



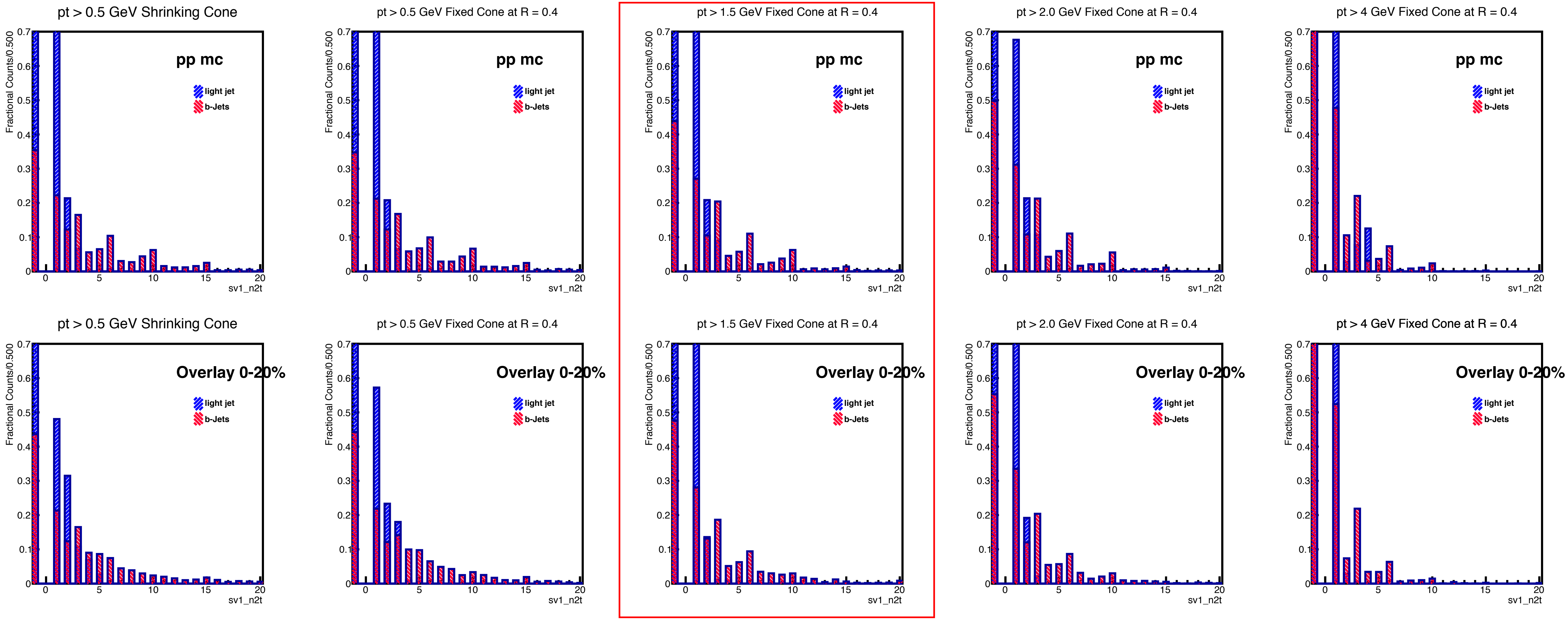
- Energy fraction and number of 2-track vertices are heavily modified.
- Peak at 1 is due to missing tracks. See back up slide 18.

SV Energy Fraction



- Visually speaking, light and b-jets distributions are the most different/separated when cutting at 1.5 GeV
- Over cutting or under cutting right-shift/left-shift both distributions.

SV n2t (2-trk vertices candidates)

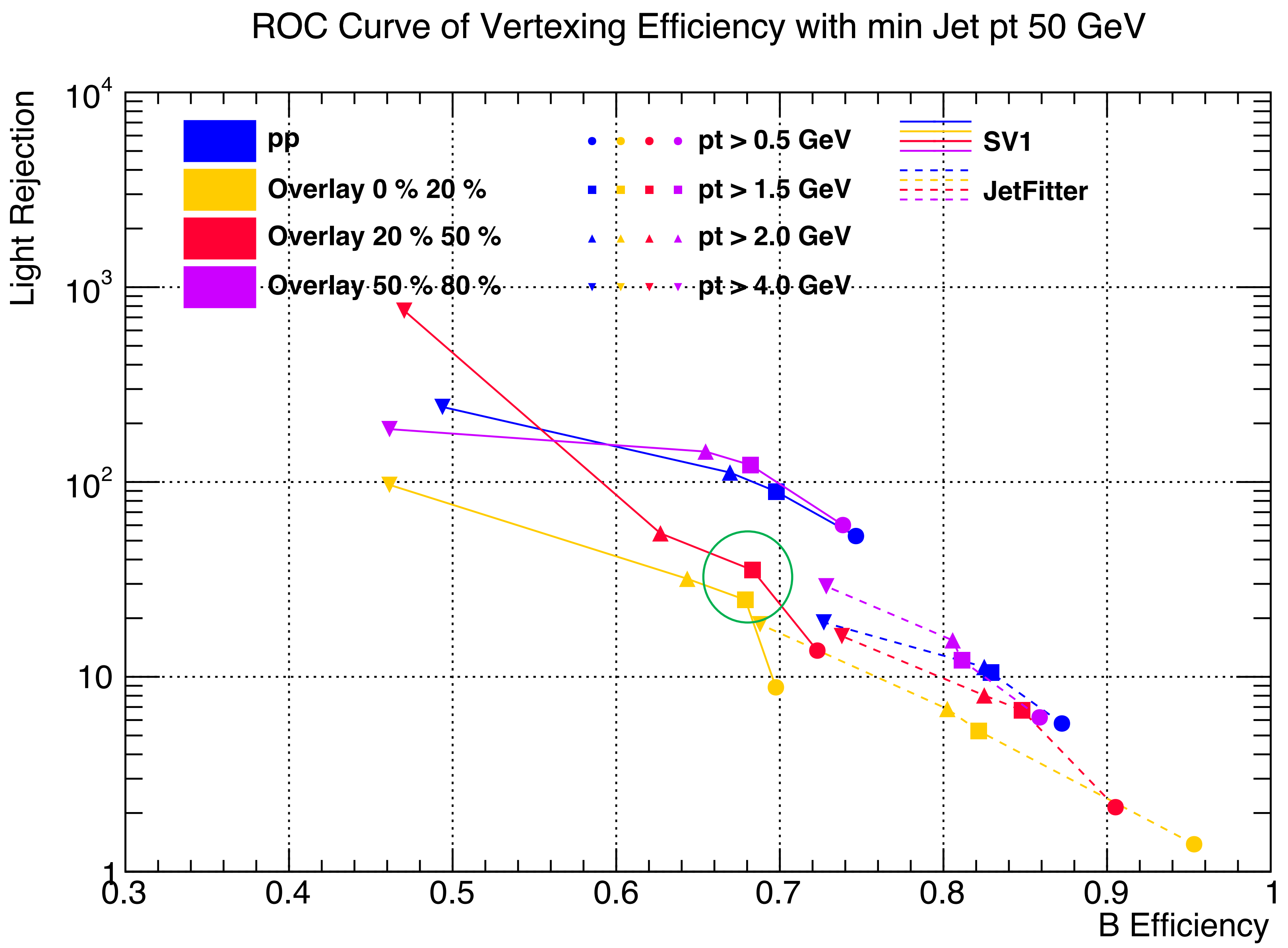


- 1.5 GeV or above cuts reduces light jet candidates in overlay to 1 or 0

↑ overcutting removes too many b-jets candidates as well

Secondary Vertexing Performance for JetFitter and SVF

ROC curve for JF and SV1 Vertexing Efficiency



- For JetFitter, cutting at pT = 1.5 GeV or pT = 2.0 GeV gives the most similar performance of different centrality.
- In most central events in SV1, cutting at 1.5 GeV have a uprise in performance

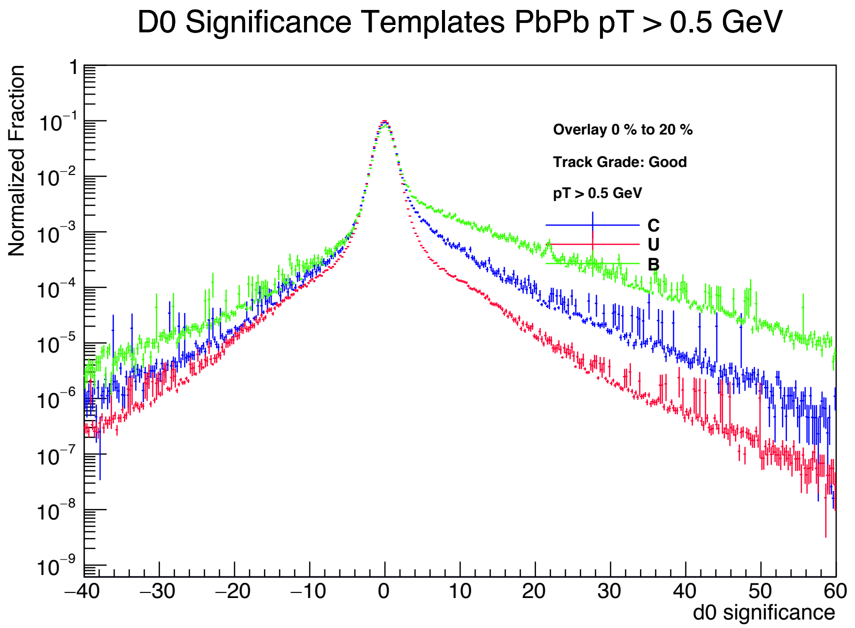
- What else can we tell from the plot?
  - JetFitter is better in efficiency and worse in rejection than SV
  - Ideas on how to choose the best cut?

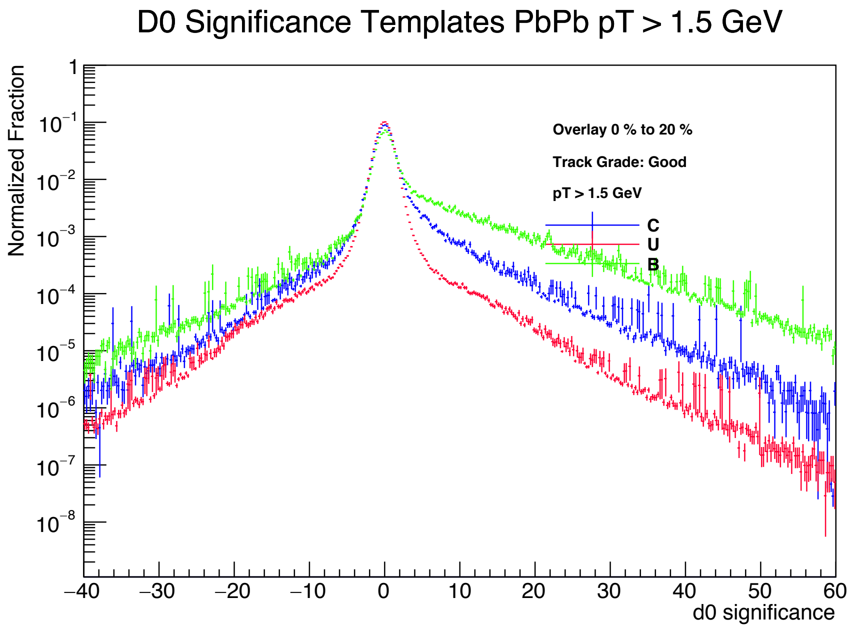
IP3D Templates Remaking

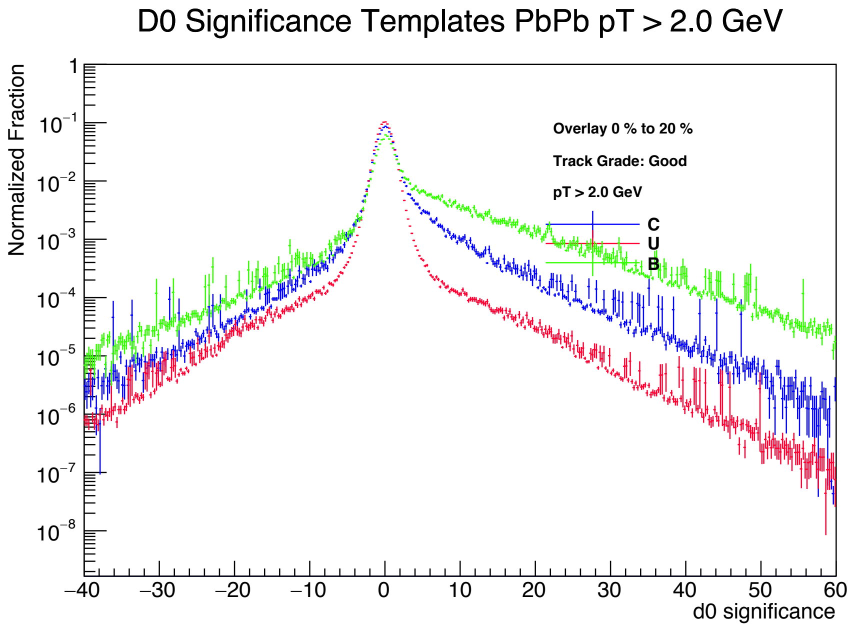
- Made templates with 8M inclusive dijet samples
- Jets selection:
  - pT > 50 GeV; 100 GeV
  - Truth matched
  - Rapidity < 2.1
  - Default JVT (Jet Vertex Tagger)* score related requirements
    - JVT is used to suppress pile-up interactions, will shut this down in future templates making
  - Flavour Labelling in templates:
    - jet_LabDr_HadF branch in the ntuple
    - based on dR < 0.3, requiring min hadron pT > 5 GeV, implemented in ParticleJetTools.
  - Retraining: make new templates with 8M sample and use these templates for evaluating the same MC.

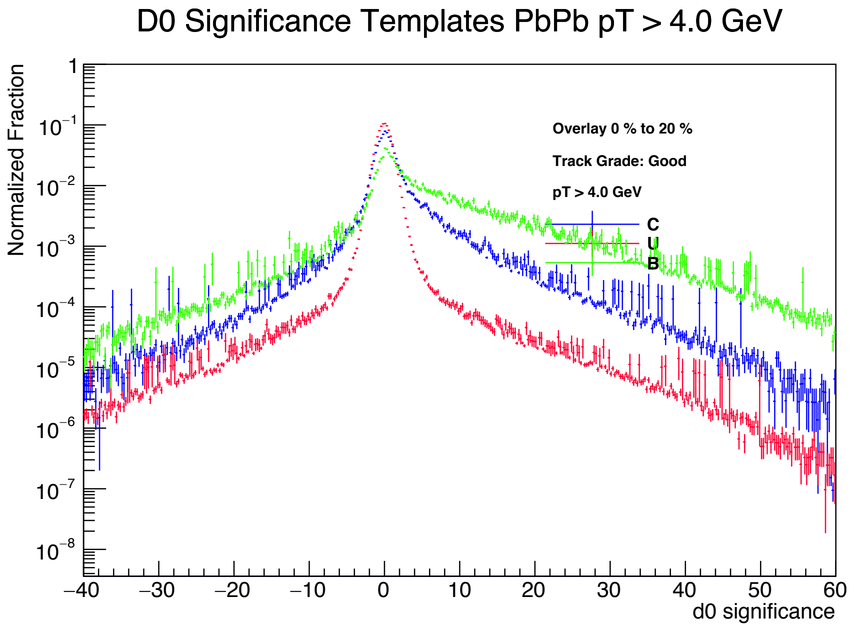
*:ATLAS Collaboration, Tagging and suppression of pileup jets with the ATLAS detector, ATLAS-CONF-2014-018, url: https://cds.cern.ch/record/1700870.

New Templates Making

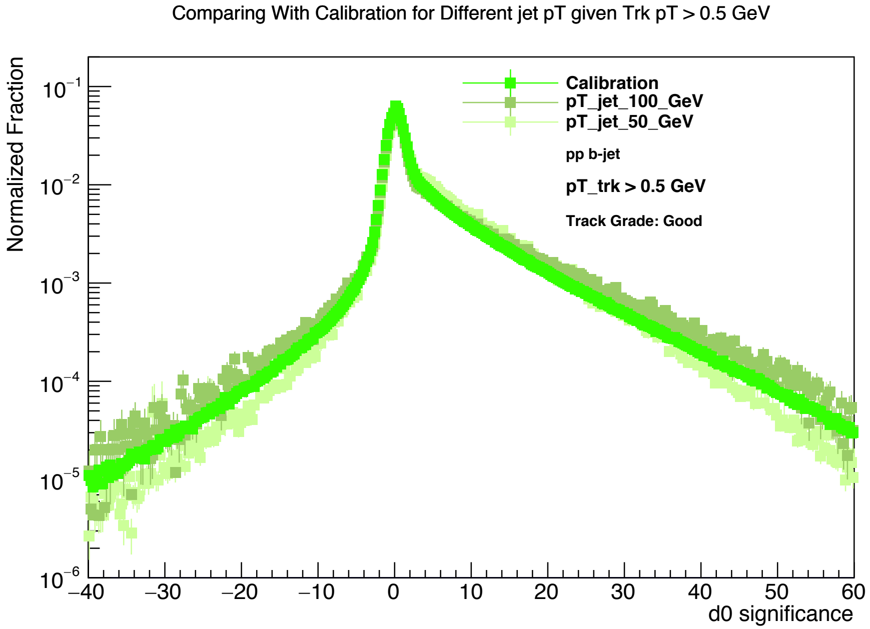


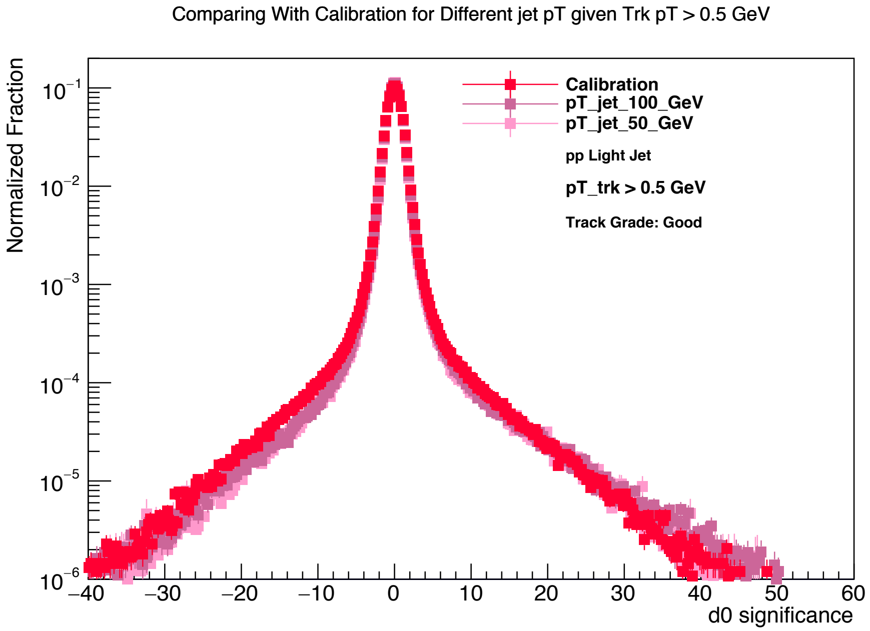


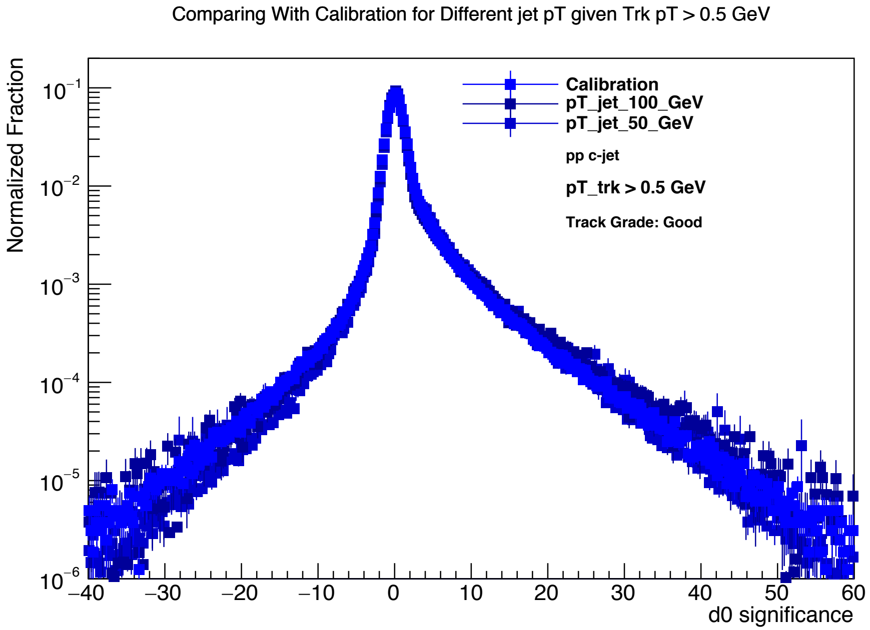




Comparison to calibration templates 2D

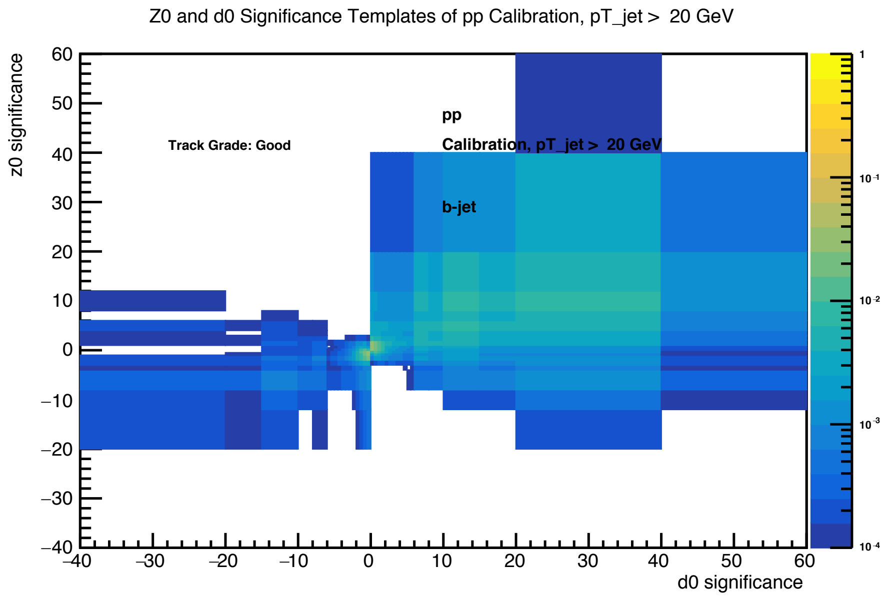


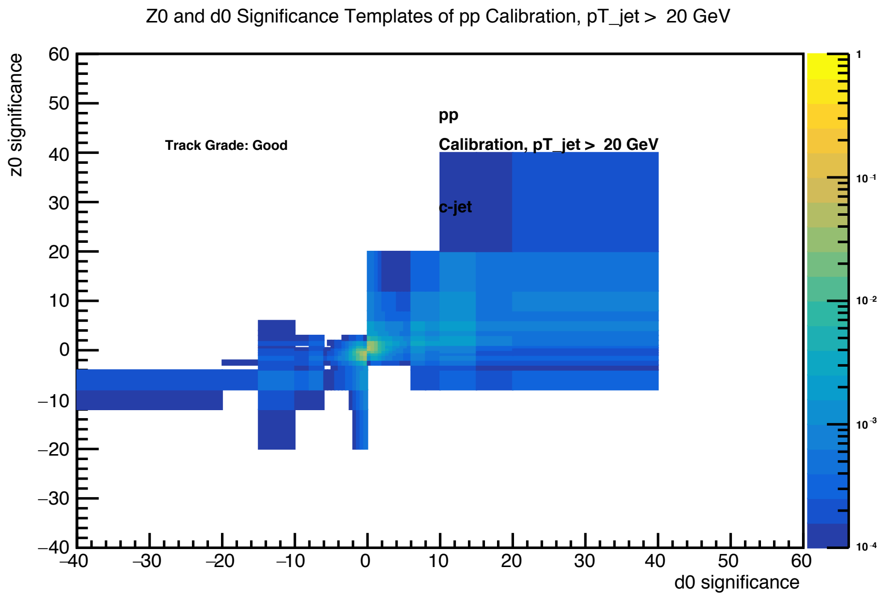


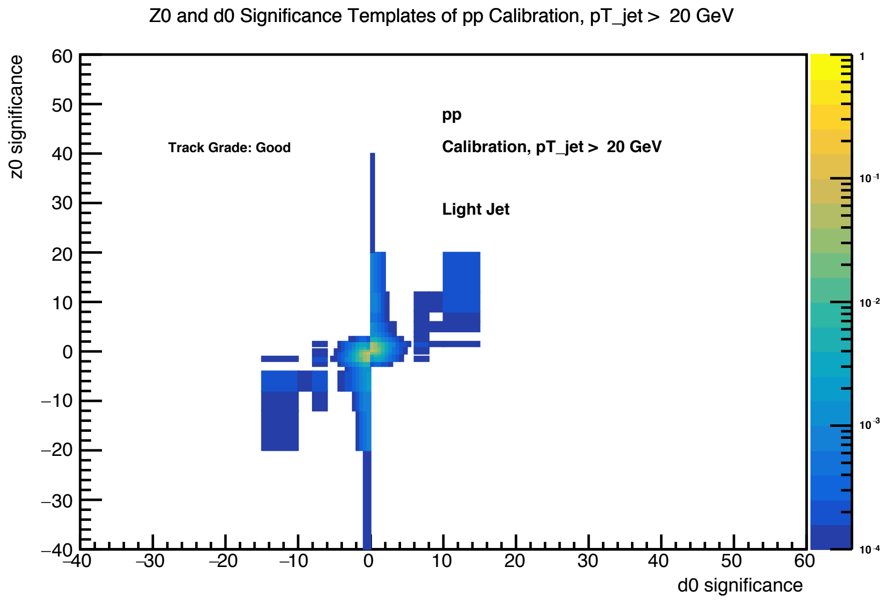


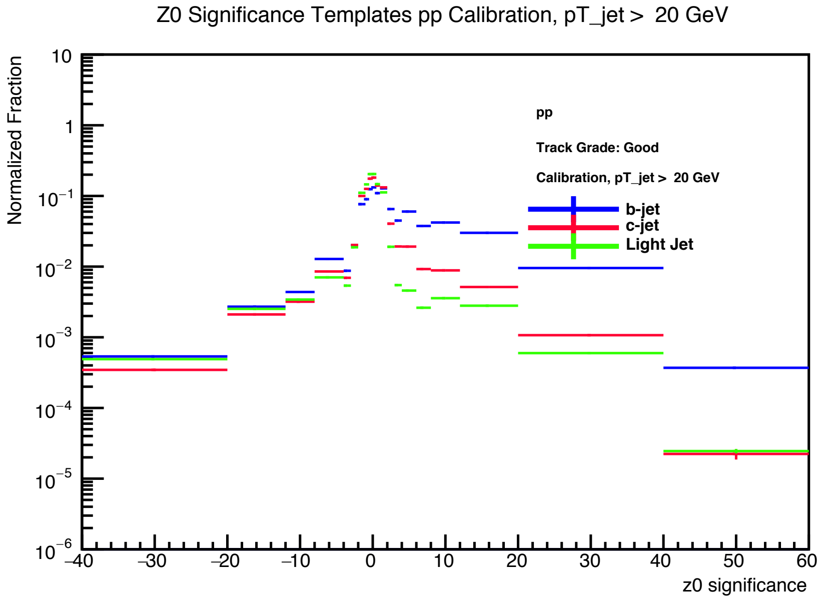
pp templates look similar qualitatively.
To do: ratio plots of templates.

Comparison to calibration templates 3D: calibration

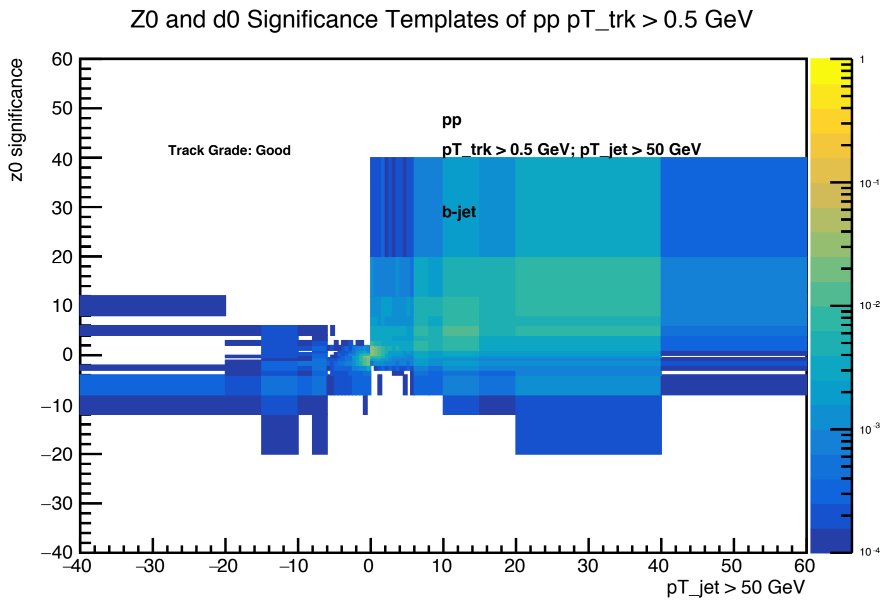


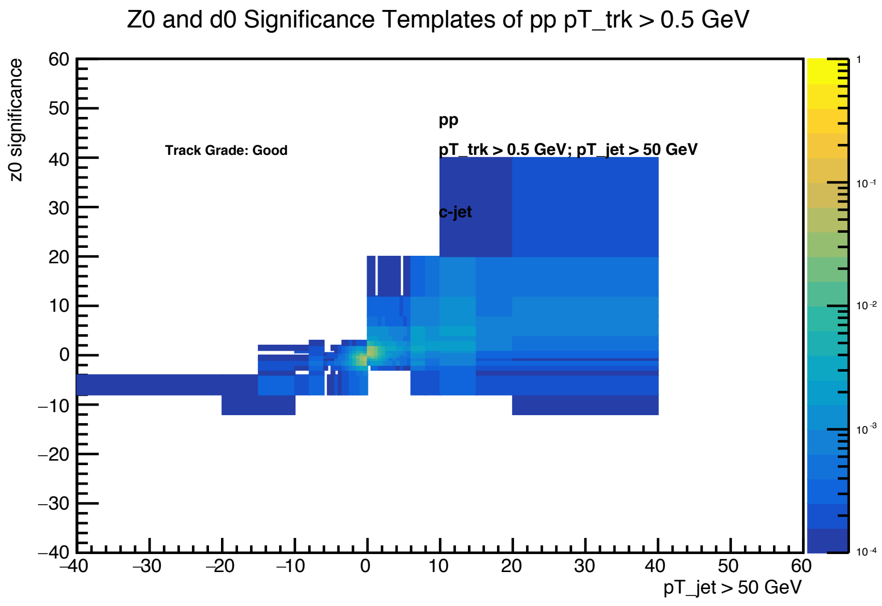


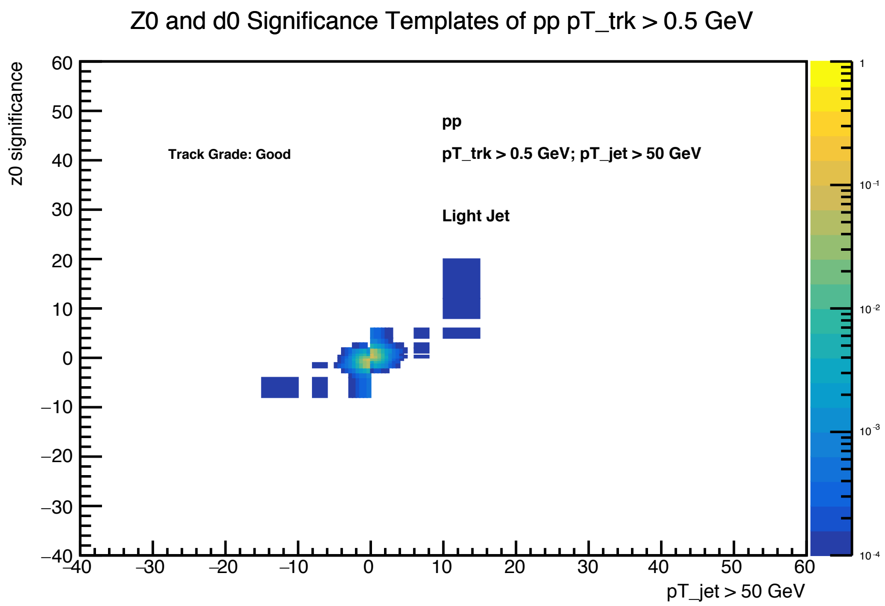


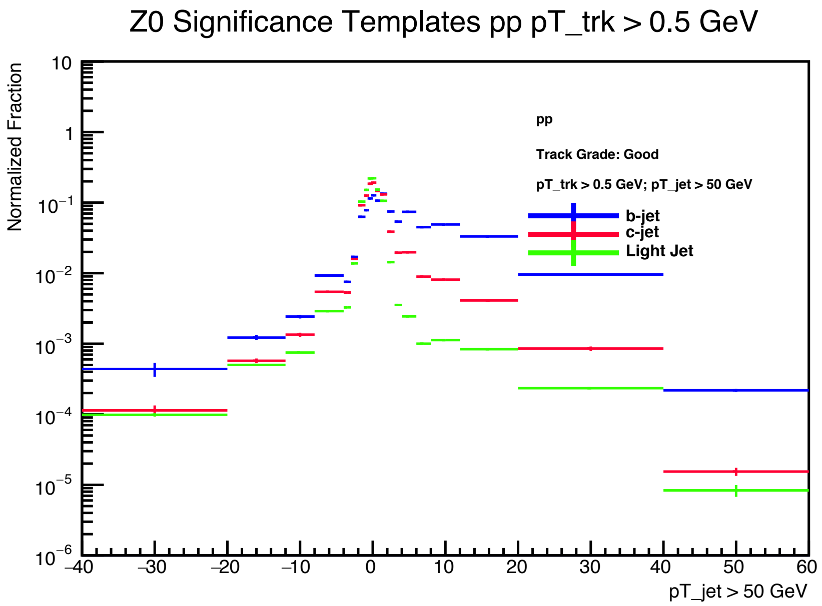


Comparison to calibration templates 3D: inclusive pp









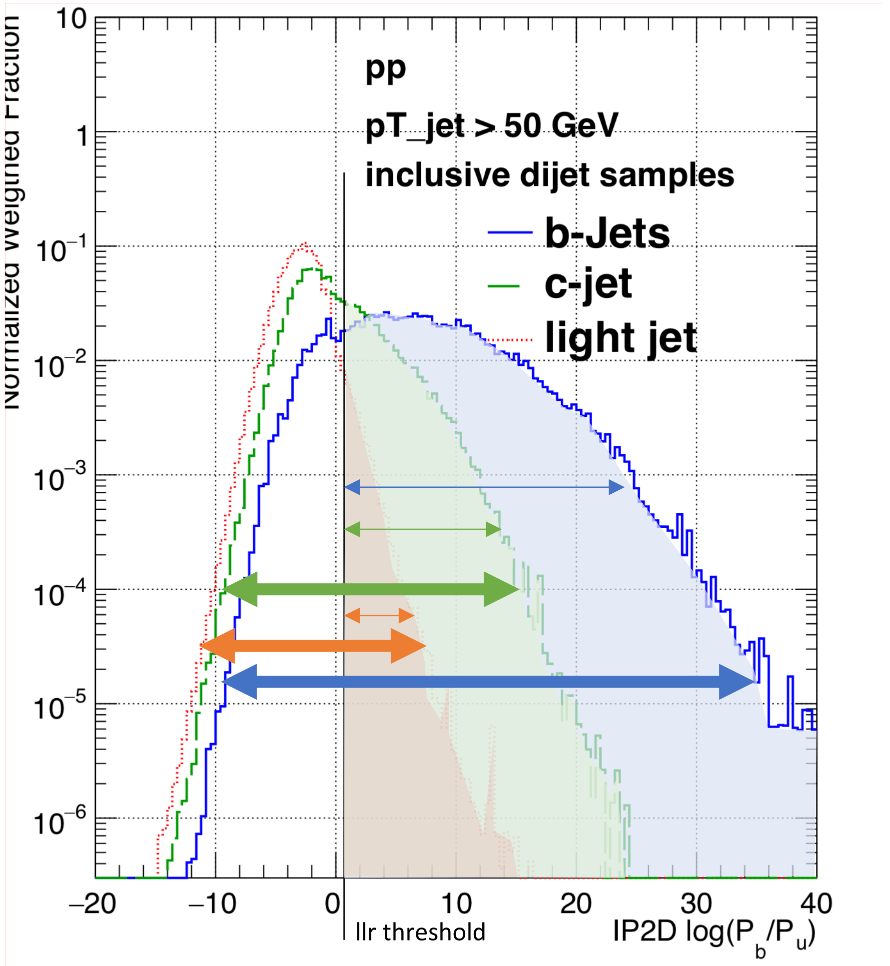
Less well populated in high Z0 bins, for all three flavors.

Plots are normalized to 1 since calibration was made with data.

- What data? (tag just says "run2")

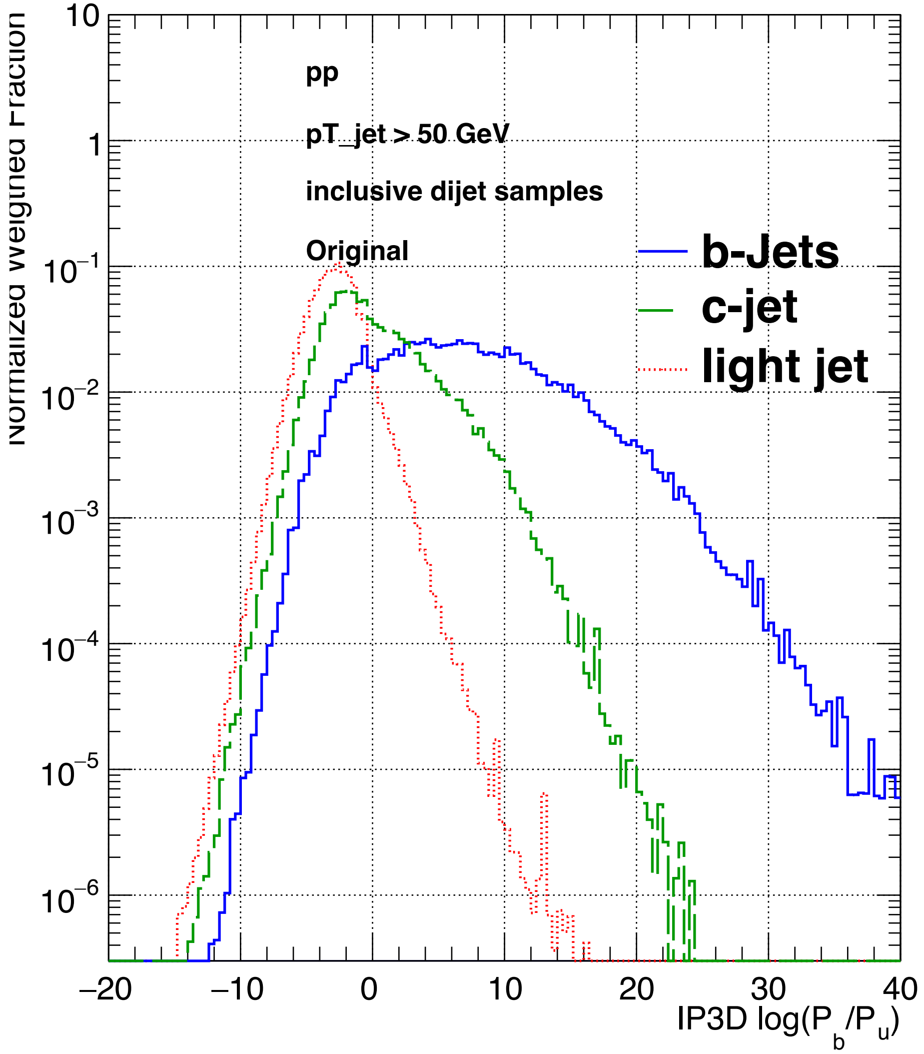
LogLikelihood Plot for (p_b/p_u)

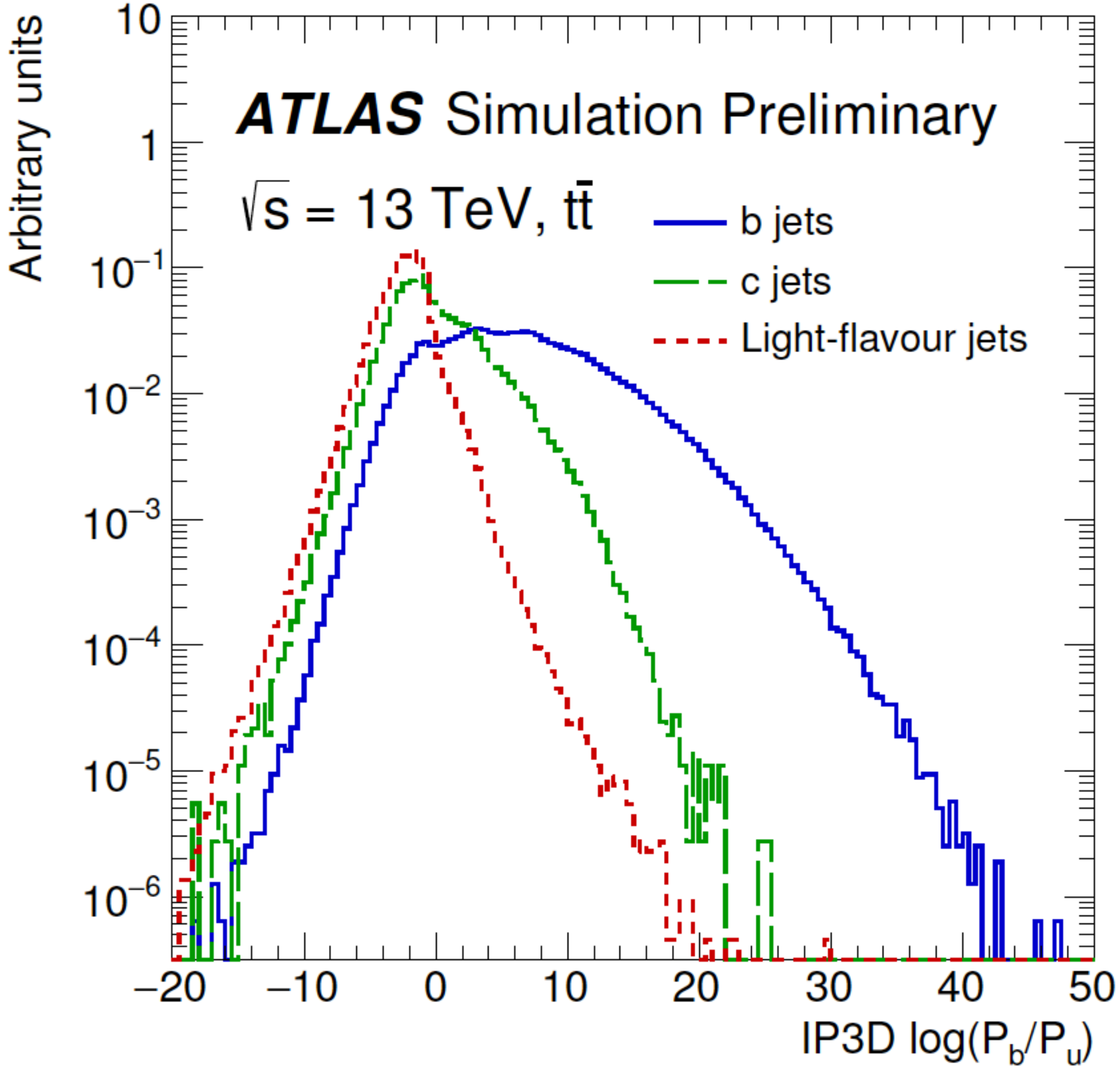
- $log^{jet}_{(\frac{P_b}{P_u})} = \sum log^{trk}_{(\frac{P_b}{P_u})}$ for qualified tracks, P_b and P_u are extracted from corresponding templates based on track quality.
- P_{flav} = template bin content/template integral.
- To make each point of the ROC curve:
  - Set a threshold, integrate from threshold rightward.
  - Divide the integral/total integration for each flavor
  - Efficiency = ratio of b-jet
  - Purity = 1/ratio of light-jet



Ratio = thin arrow integral/thick arrow integral

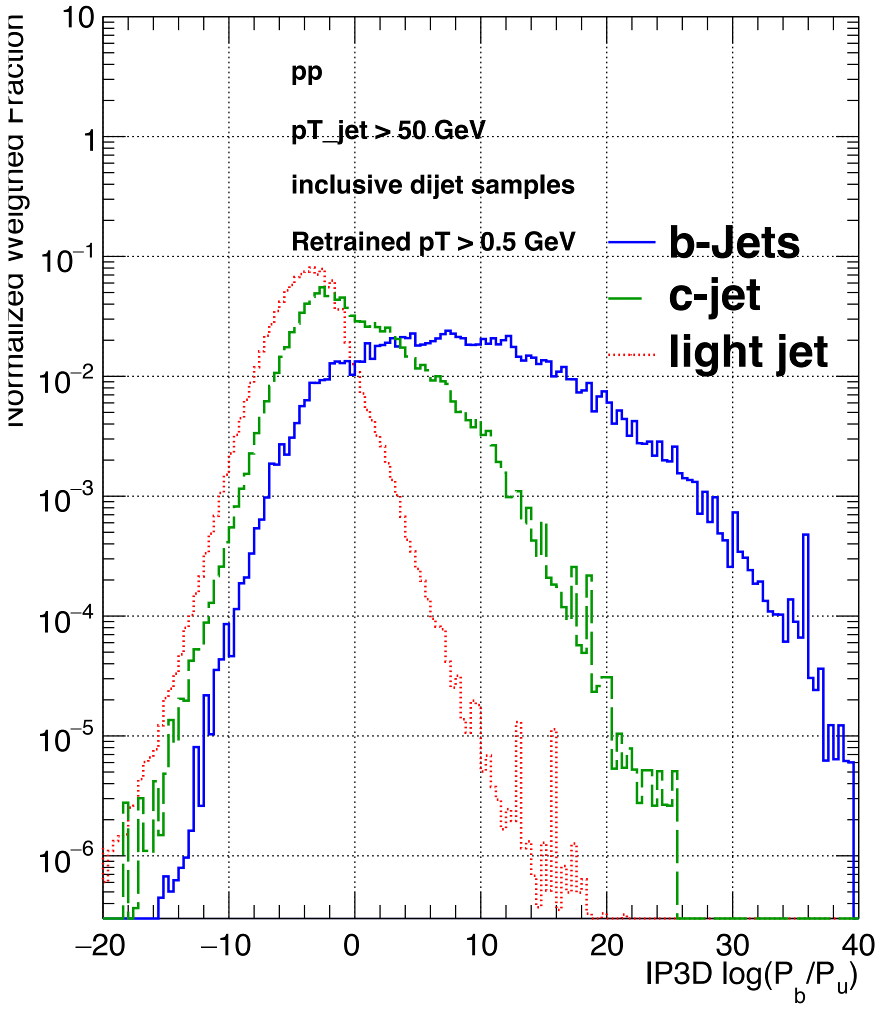
Comparison with ttbar samples

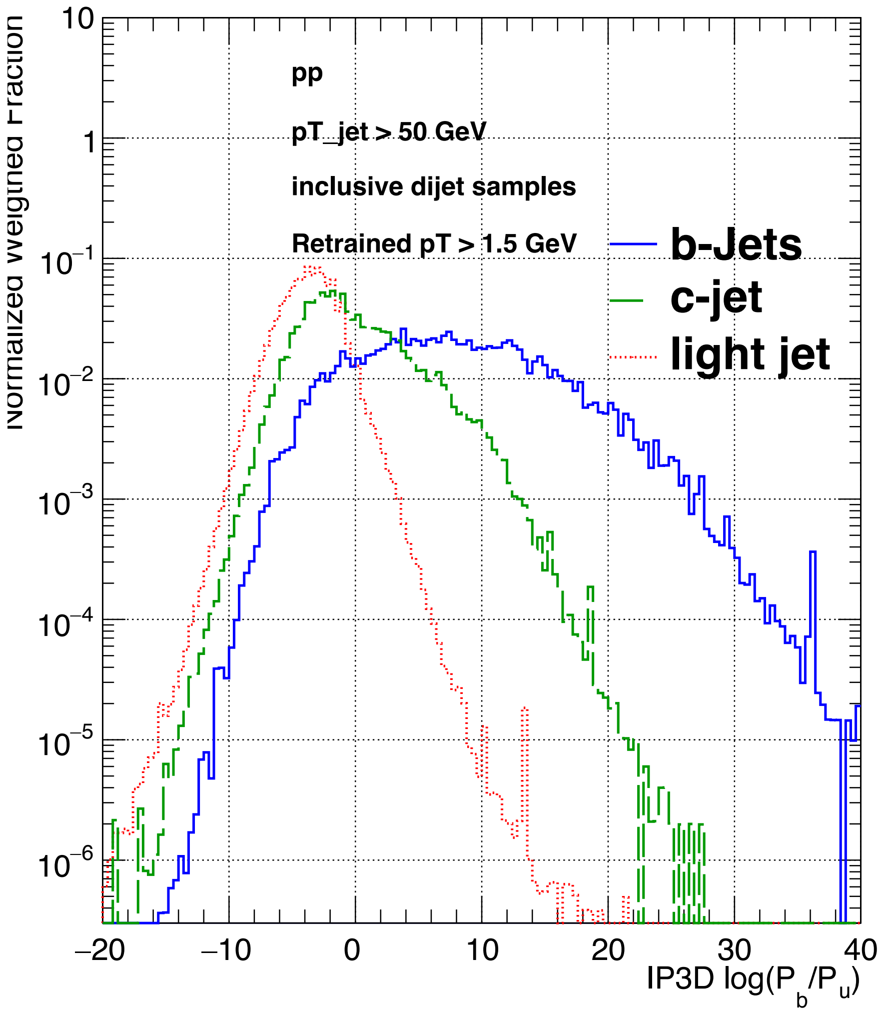




Qualitatively similar.
Peaks are slightly more smeared.

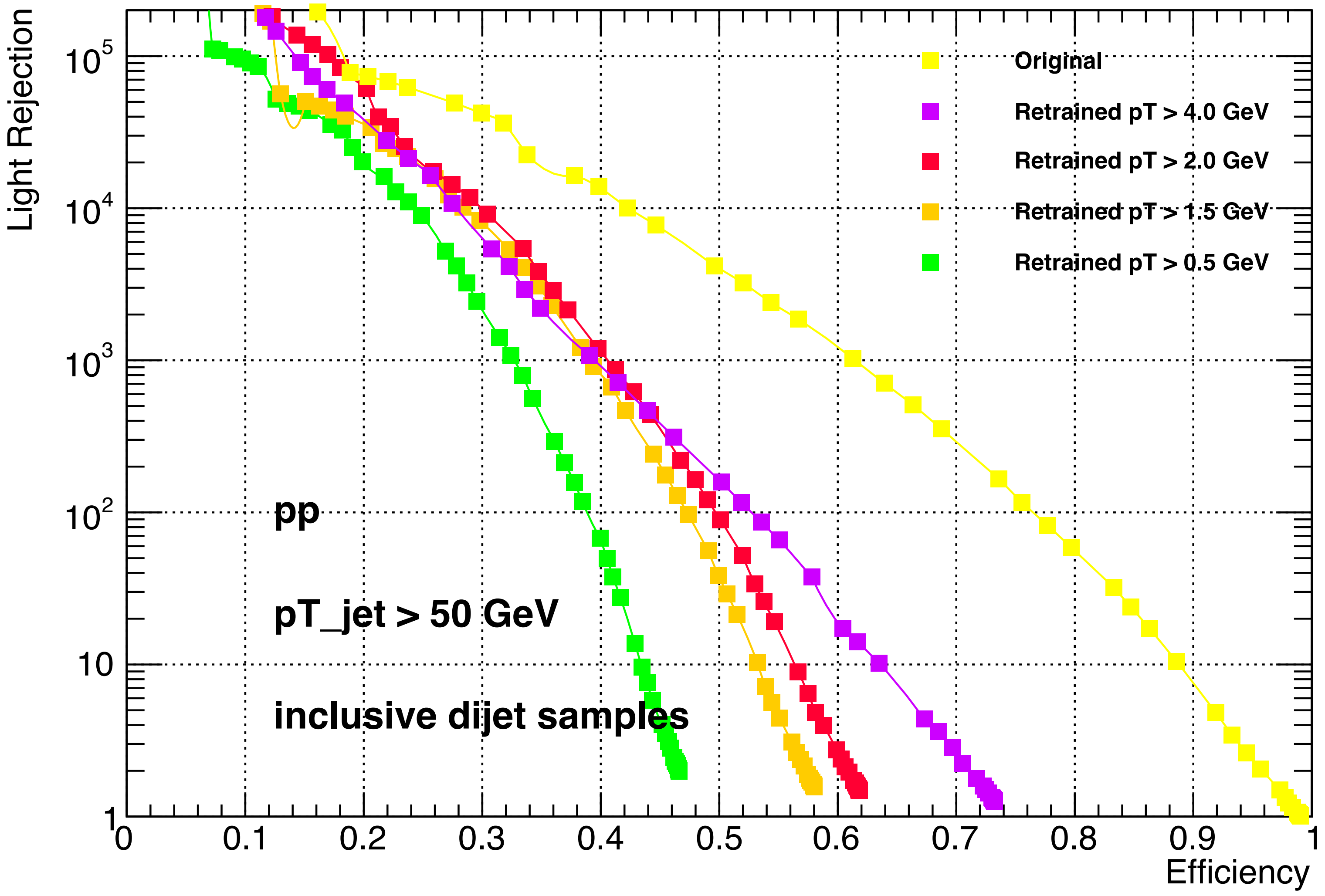
Retrained llr plots



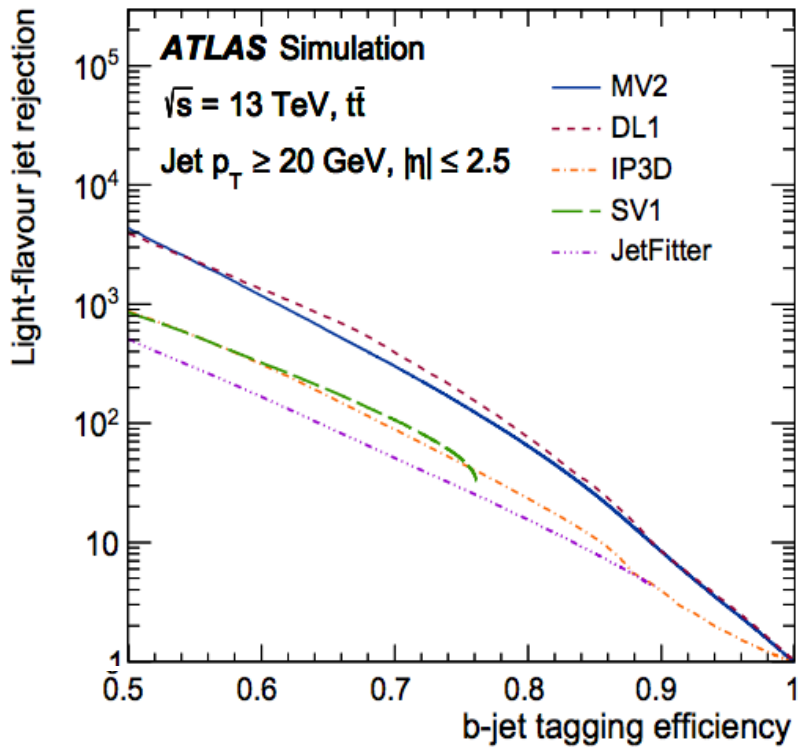


- Distribution even more smeared than using original templates.

IP3D ROC curve with pp

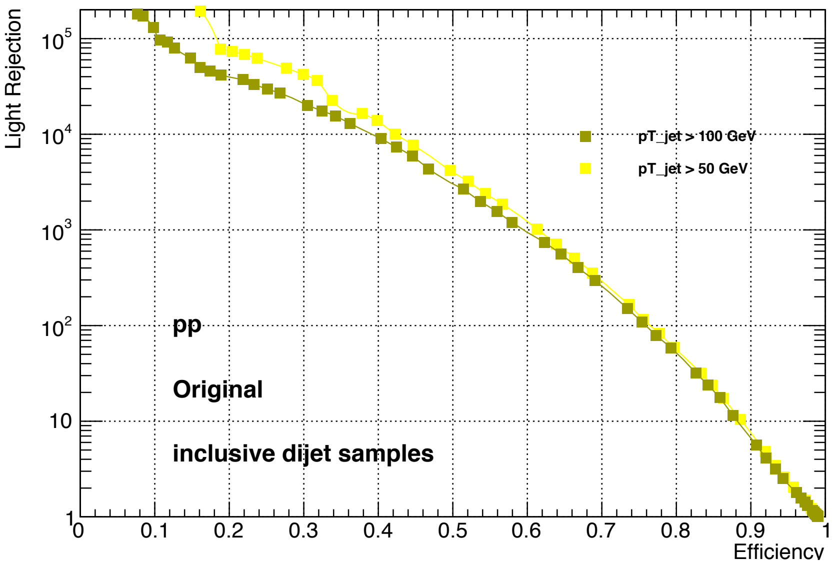


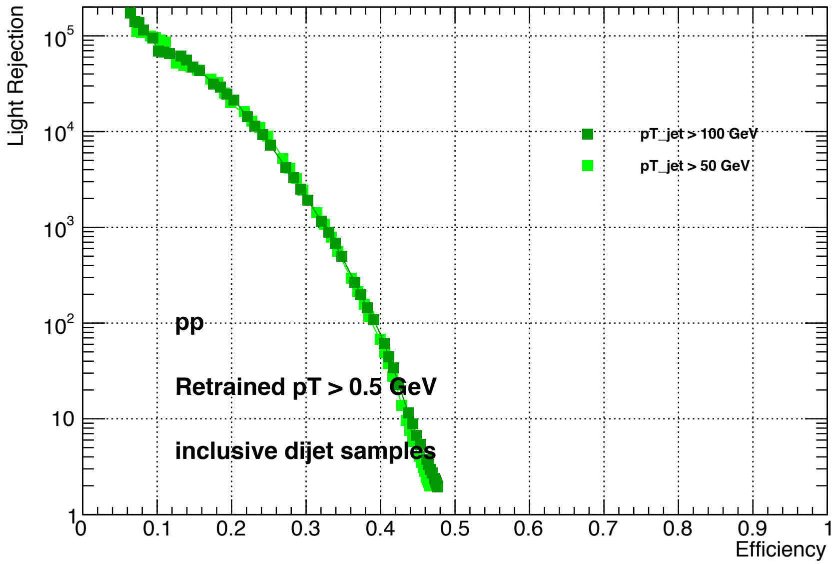
- Original templates have the best performance.
- With increased cuts on pT track, the performance is improved.

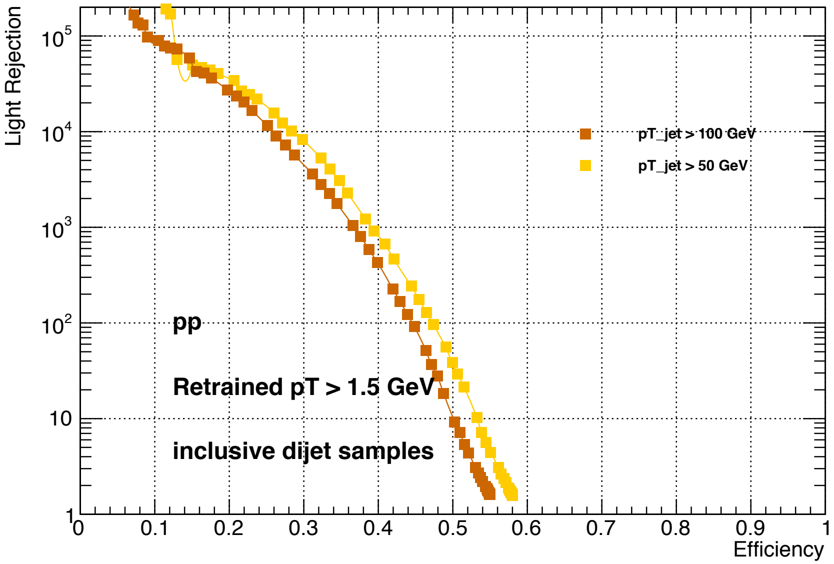


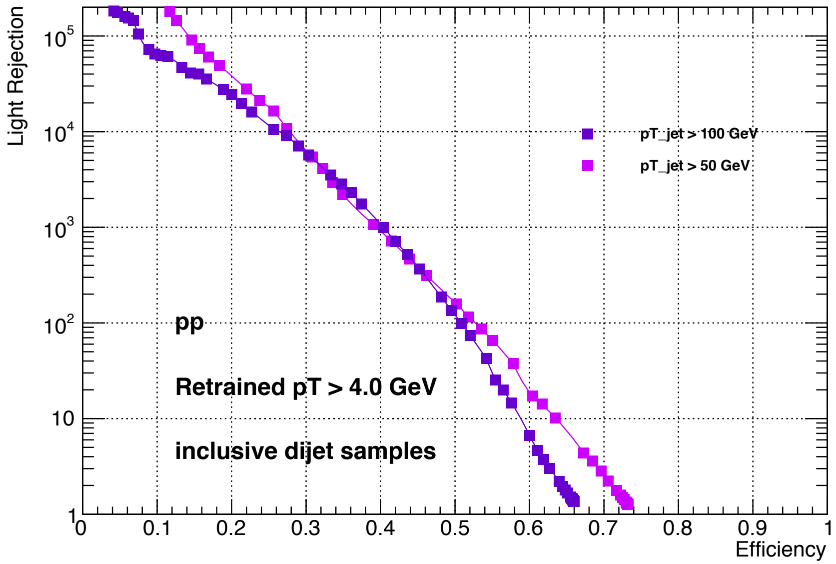
https://arxiv.org/pdf/1907.05120.pdf

Comparing Different jet pT









For the pp samples concerned,

- Without cuts on track pT, cutting jet pT at 100 GeV is performing marginally better
  - Possibly due to underlying effects in low pT jets
- With cuts on track pT, cutting at jet pT 50 GeV is performing slightly better
  - pT dependent performance? pT dependent ROC curve?

Summary

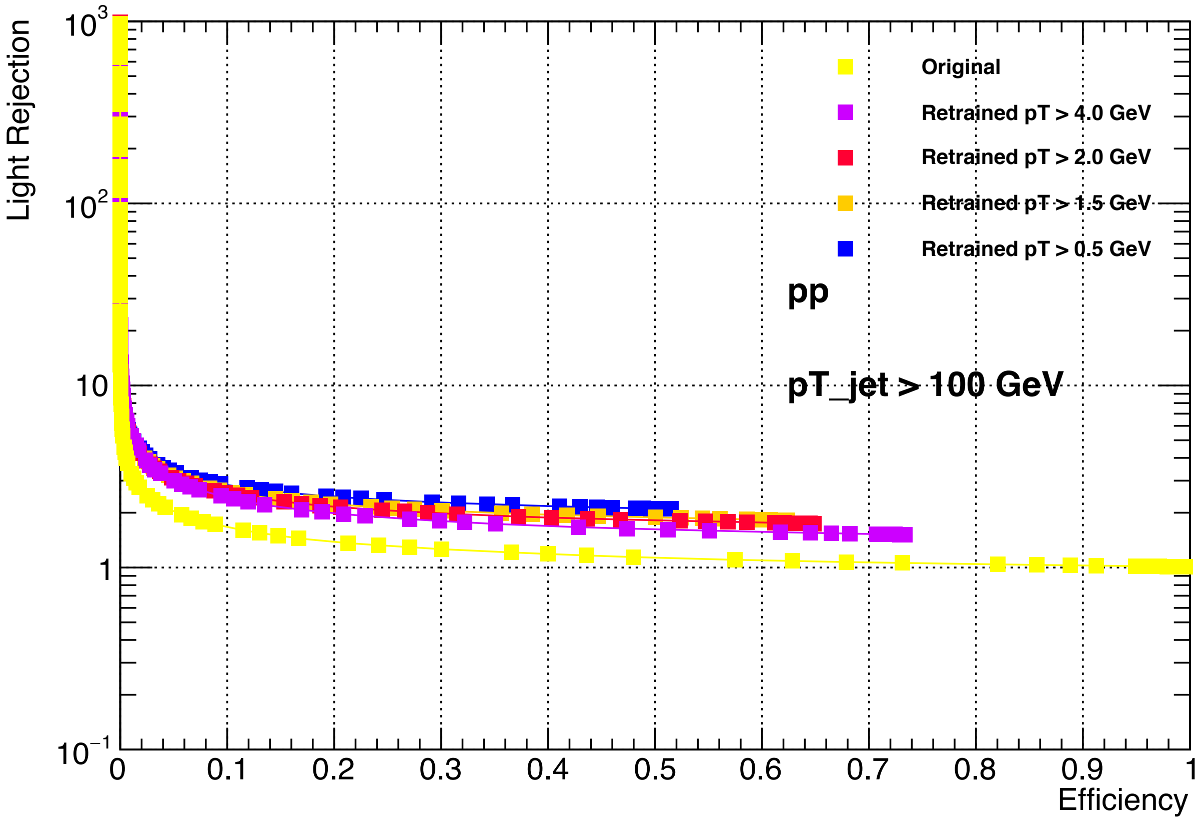
- Comparing some typical physics variables, cutting pT at 1.5 GeV optimized most modified variables.
  - some variables are better optimized at 2.0 GeV cut, should combine results of IPxD for choosing the best cut.
  - To better evaluate everything together, consider writing a simple probabilistic classification for each tagger.
- IP3D for pp shows worse performance with re-training.
  - Statistics?
  - Try plotting for PbPb.

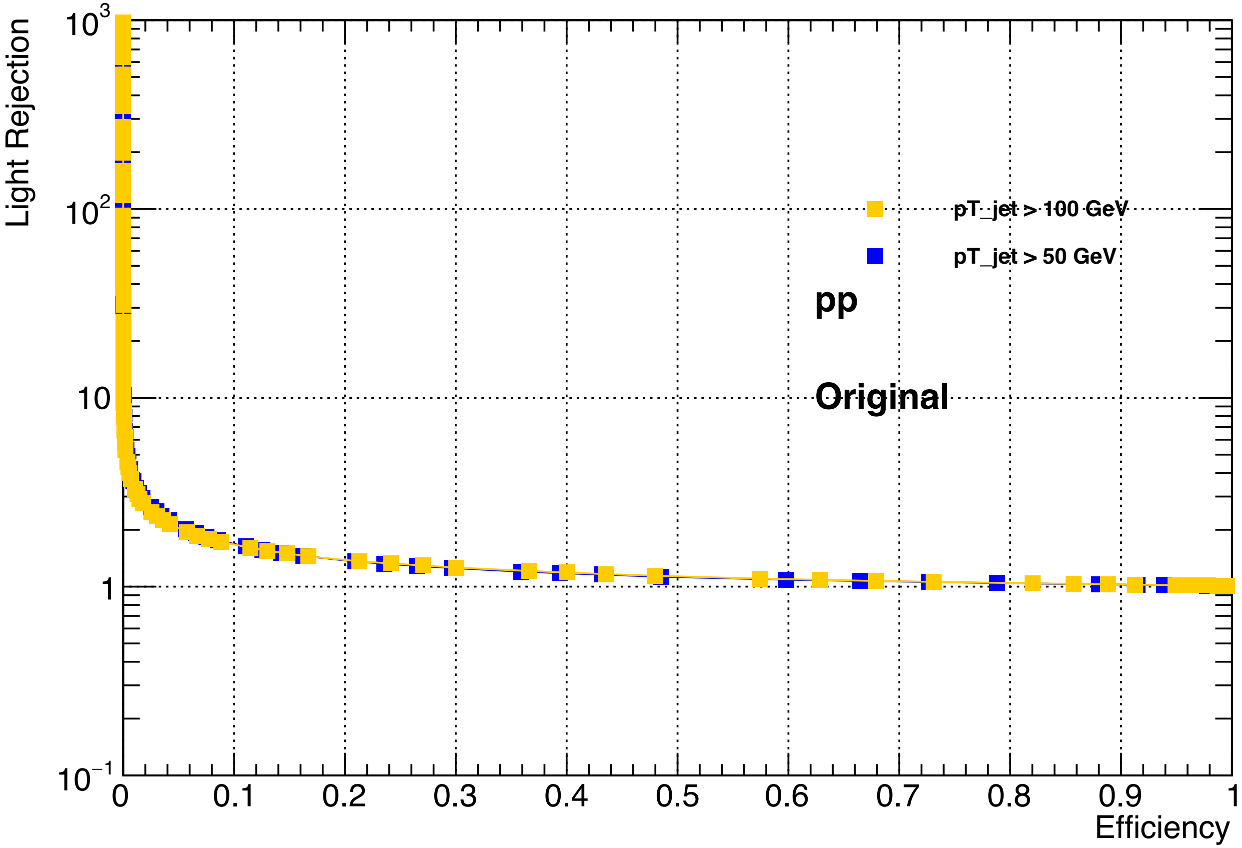
Back up (I lied in the title clearly)

Checklist

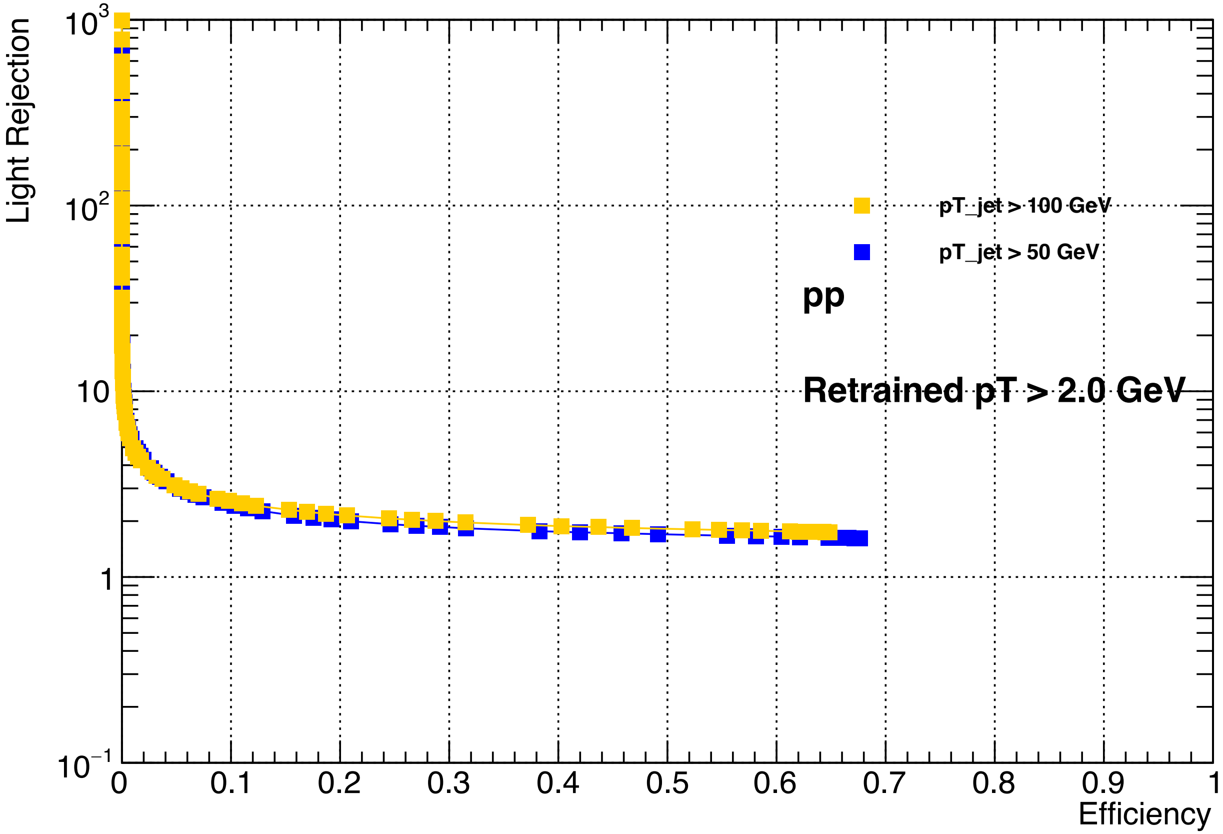
- Jet pT? (Slide 16 -17)
  - Jet pT > 50 GeV
  - Jet pT > 100 GeV
- Centrality? (Slide 18)
  - Is it just the pp samples? Compared most central at jet pT > 100 GeV with pp
- Samples? (Slide 18)
  - Compared small 50k samples to 8M inclusive samples
- Track quality? Other ideas?
- Look for single IPxD performance plots. (There're none in the btagging performance papers, have it always been this bad??)

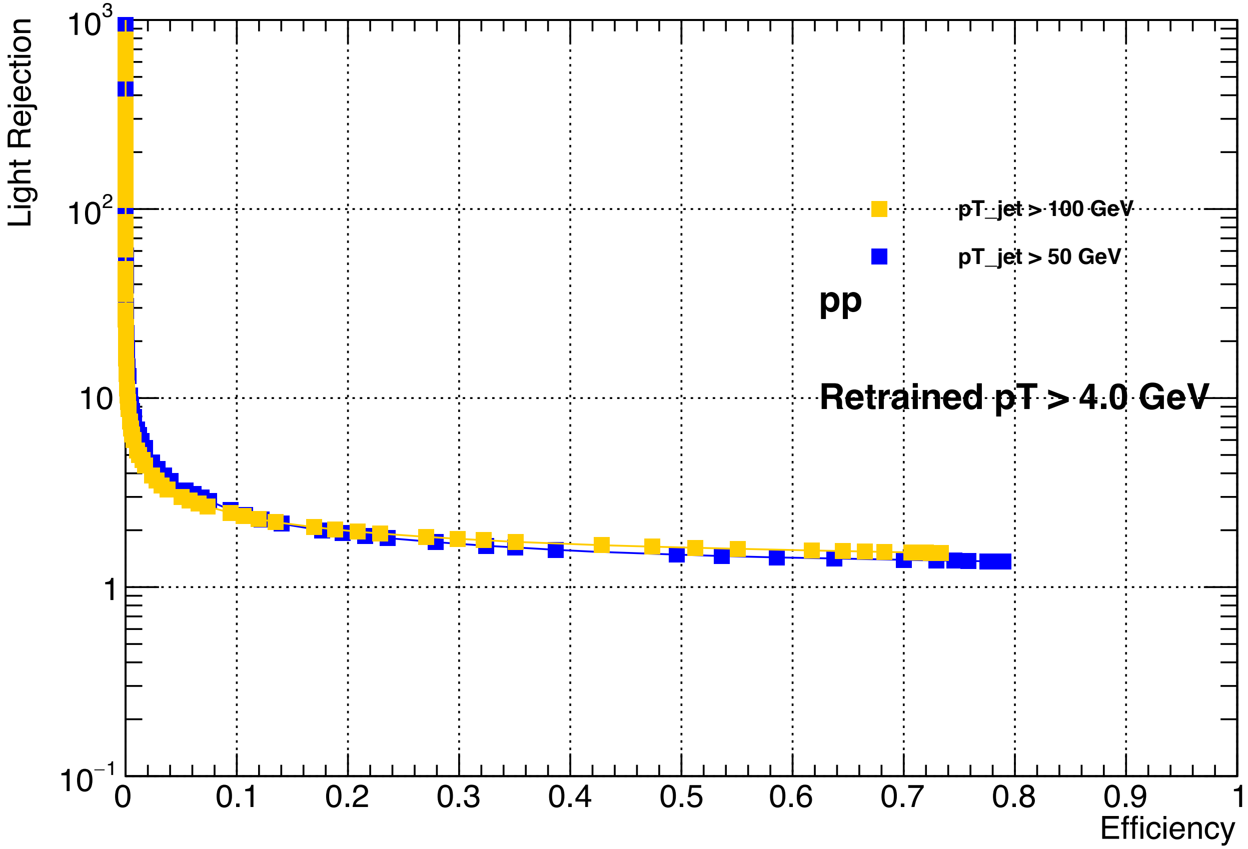
Different Jet pT comparisons





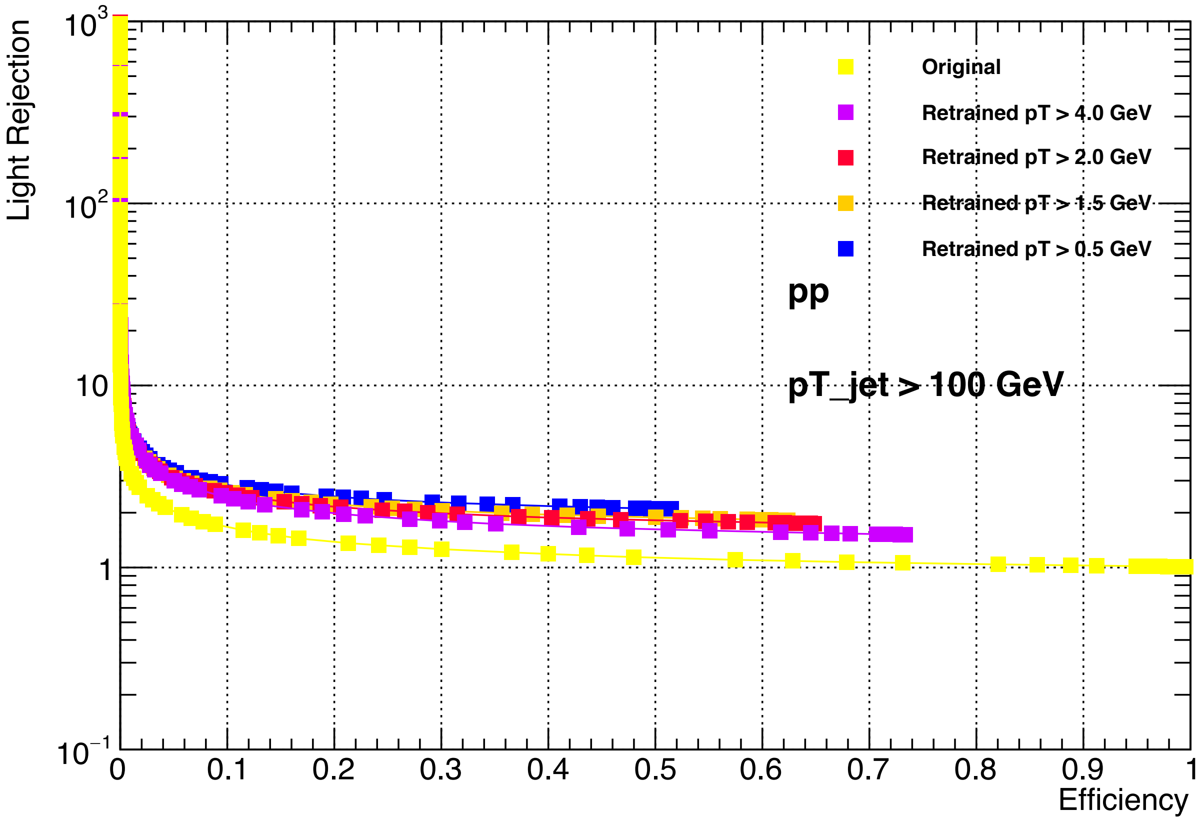
Original: evaluations from using the default calibration templates. (BTagCalibRUN2Onl-08-40.root, 2017 pp templates)
Retrained pT > xx GeV: evaluations from using inclusive samples' templates with the cut track pT > xx GeV and jet pT > xx GeV

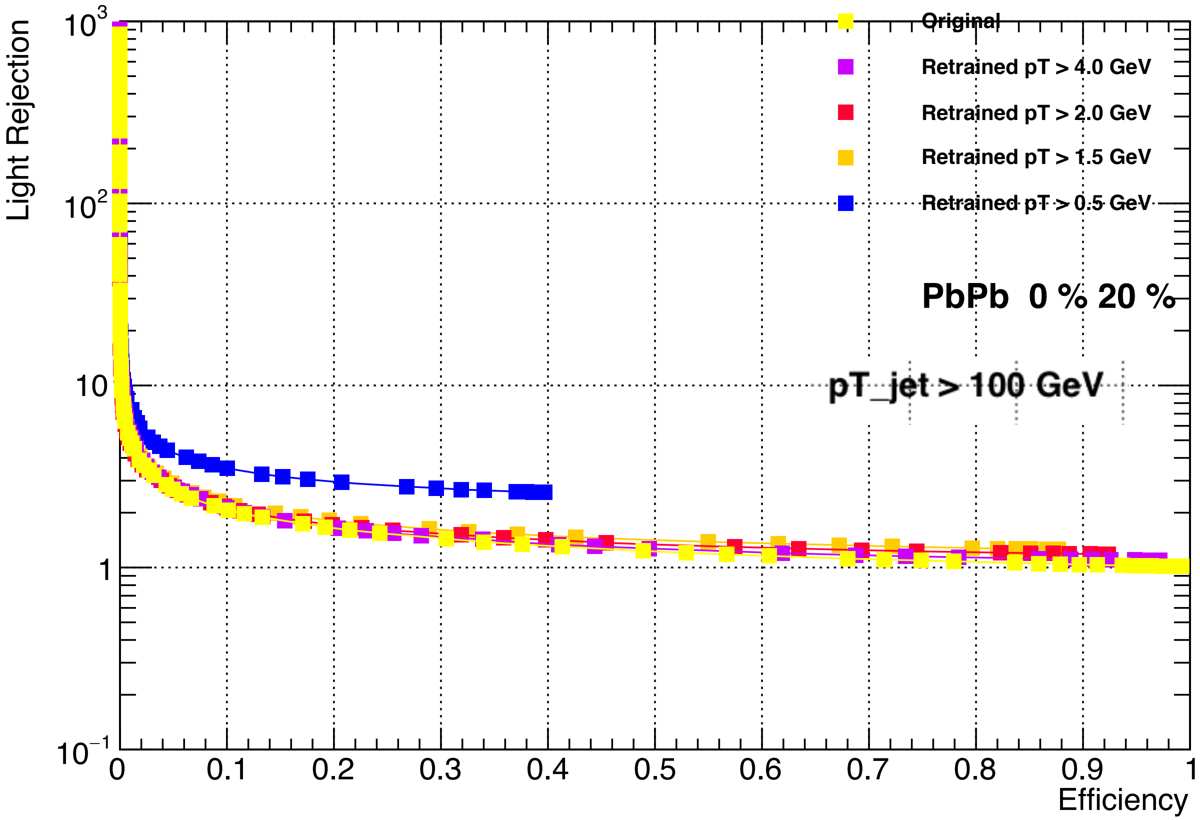




pT_jet > 100 GeV with the same llr cut threshold is better in purity but worse in efficiency, as expected
In general the performance is similarly bad.

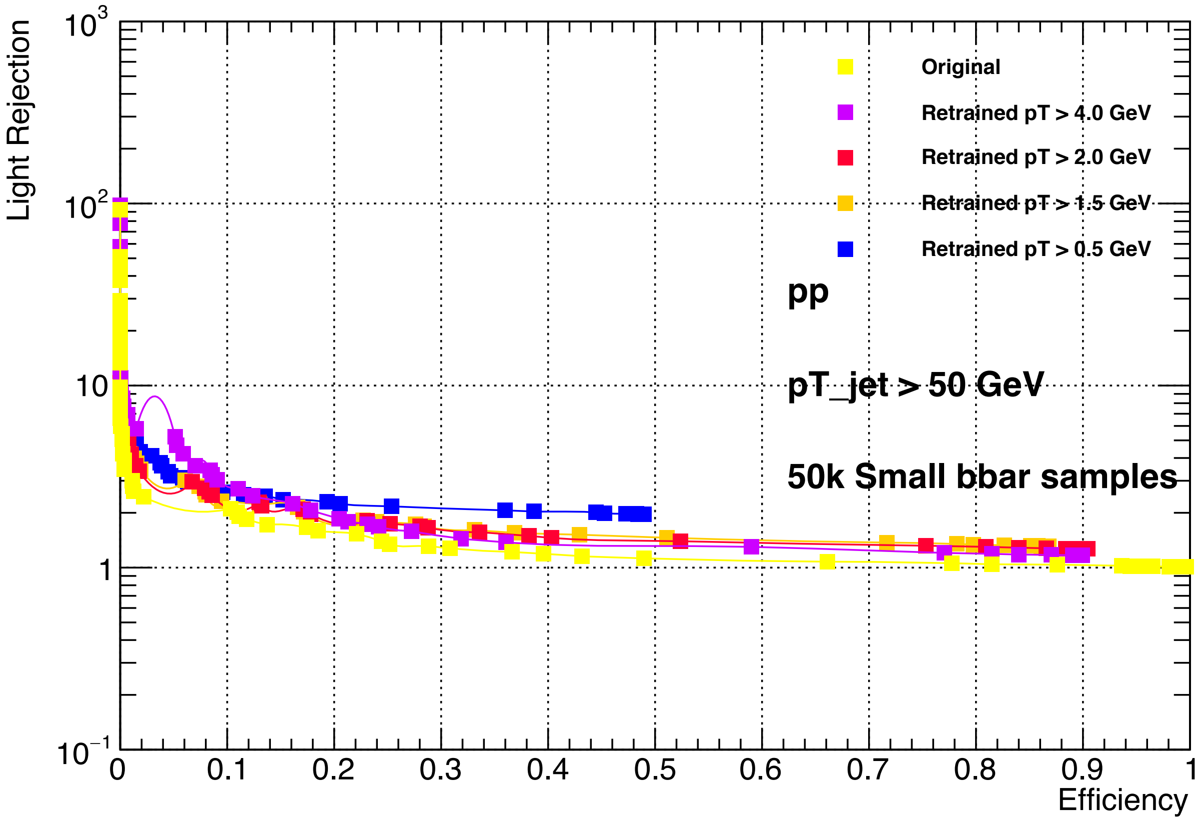
Comparison with PbPb (?)

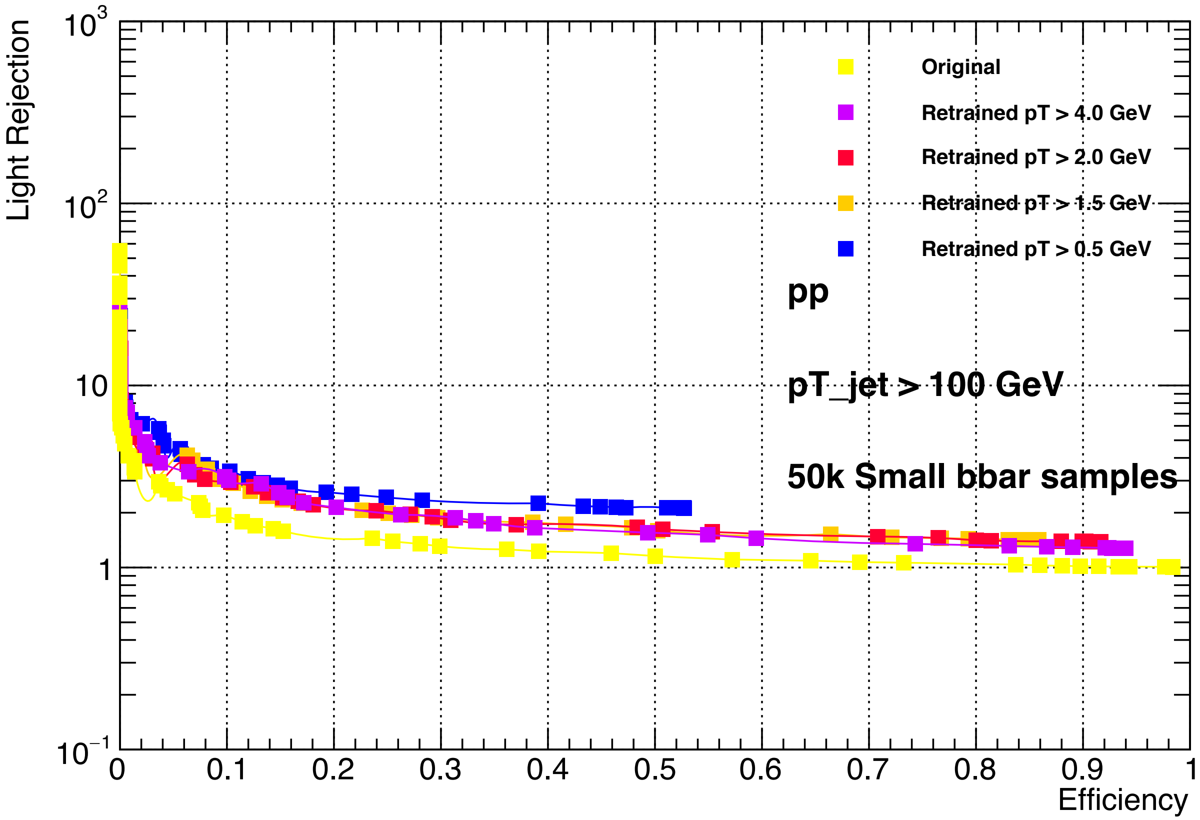




Most central events are better in light rejection but worse in efficiency if we don’t cut on track pT. (??)
Others better in efficiency but worse in light rejection.

Comparison with Small samples





Small bbar samples: https://its.cern.ch/jira/browse/ATLHI-240
Performance is qualitatively similar to inclusive dijet samples.