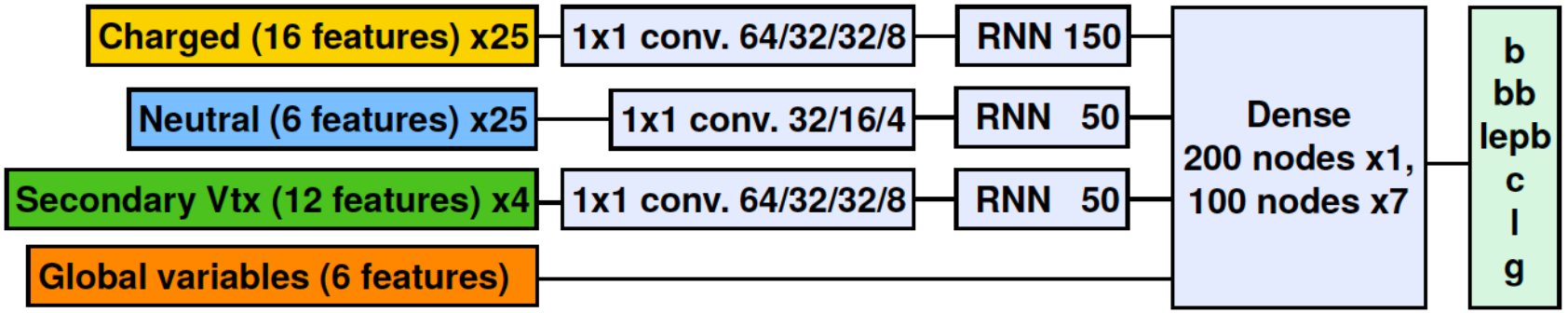
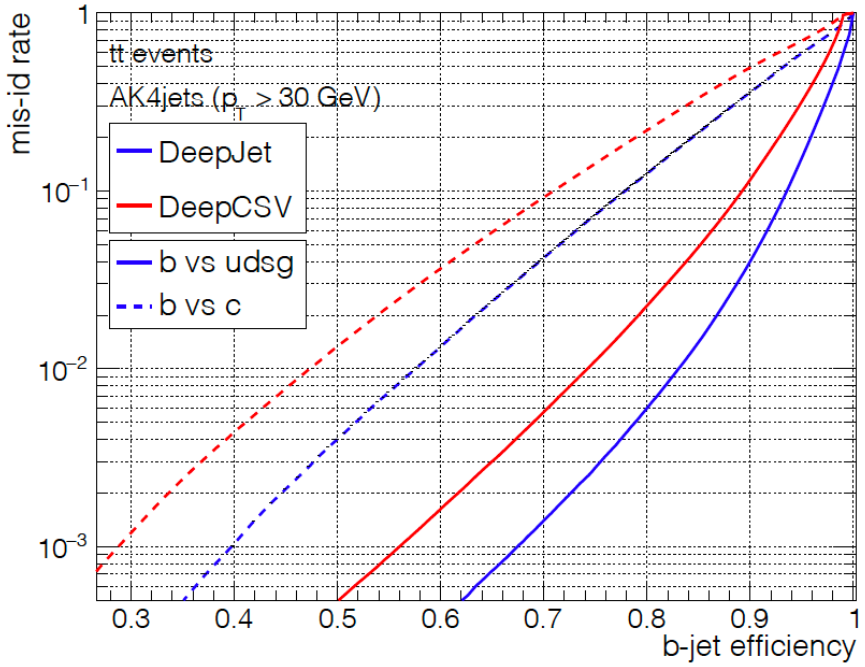
CMS DeepJet

„Jet Flavour Classification Using DeepJet“, arXiv:2008.10519



- considers information of all jet constituents (neutral and charged), secondary vertices and global event variables —> 650 input variables
- features of the first 25 charged/neutral candidates w.r.t. displacement significance / shortest angular distance to secondary vertex (p_T ordering) considered
- secondary vertices ordered by flight distance significance



Input files & first steps

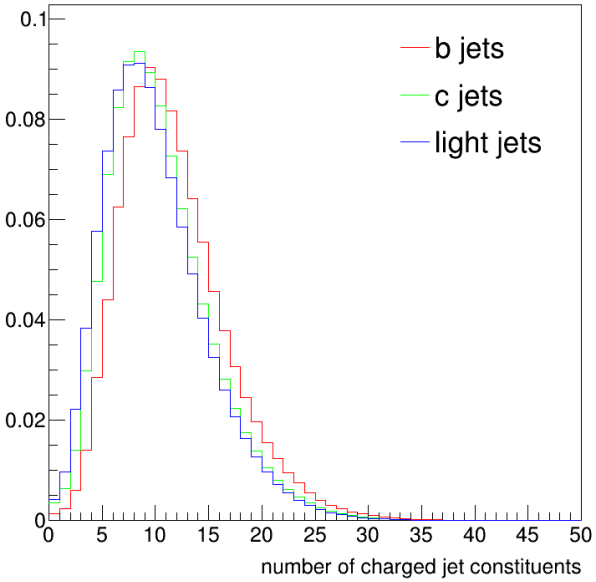
- **here**: classify jets into three different classes: b jets, c jets and light jets
- input files: **events with 6 jets** (b,c,u,d,s)

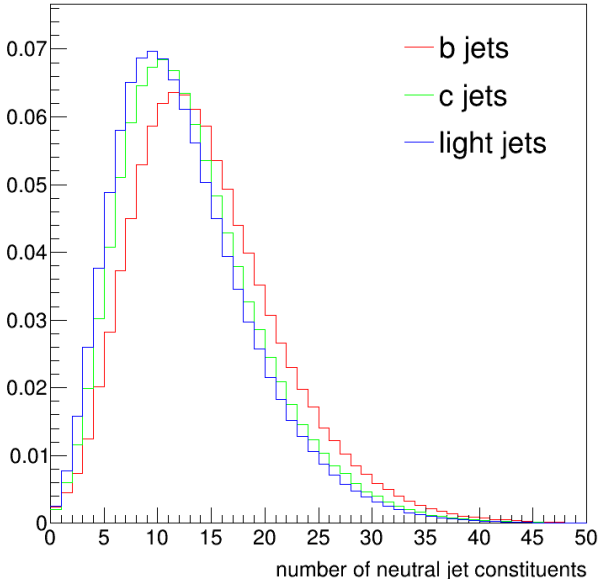
  /pnfs/desy.de/ilc/prod/ilc/mc-opt-3/ild/dst-merged/500-TDR_ws/flavortag/ILD_l5_o1_v02/v02-00-01/

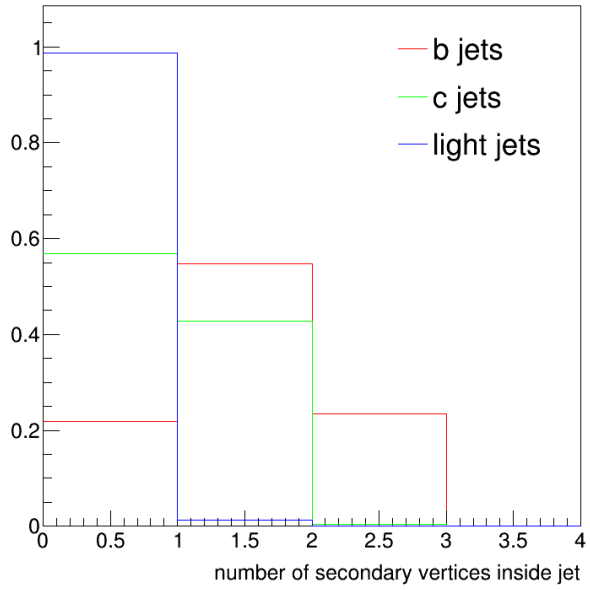
- **first steps** : run PV & SV finder, jet clustering and vertex refinement of LCFIPlus

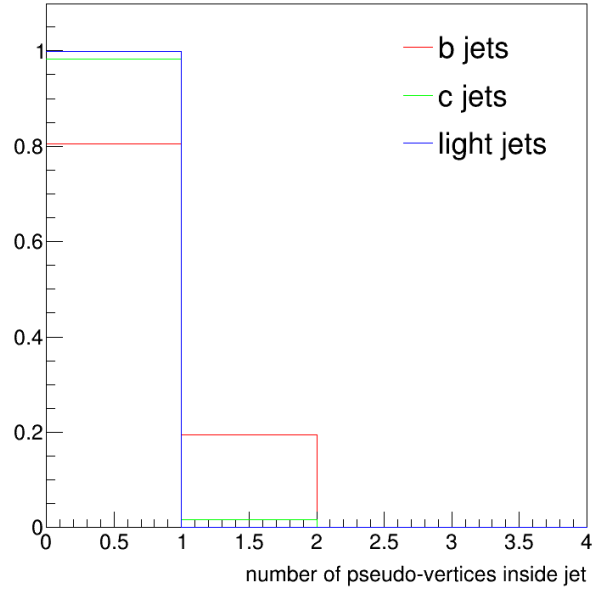
- note: SV collections contains secondary vertices and pseudo-vertices

(Possible) Input features - global variables









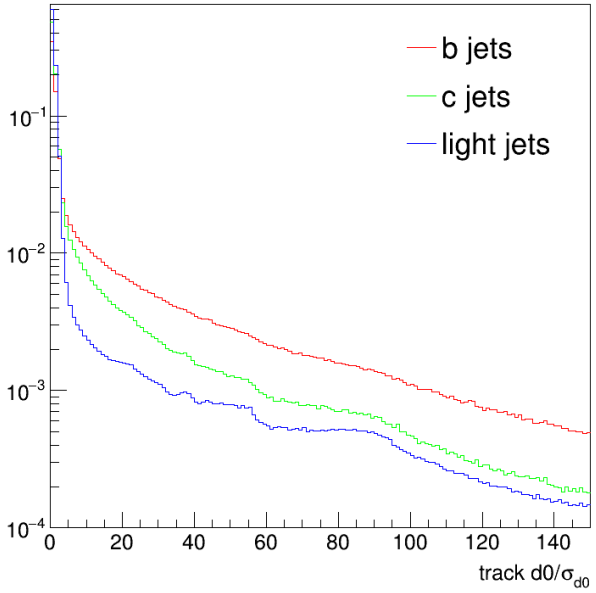
- jet p, p_T, energy, θ, φ & rapidity

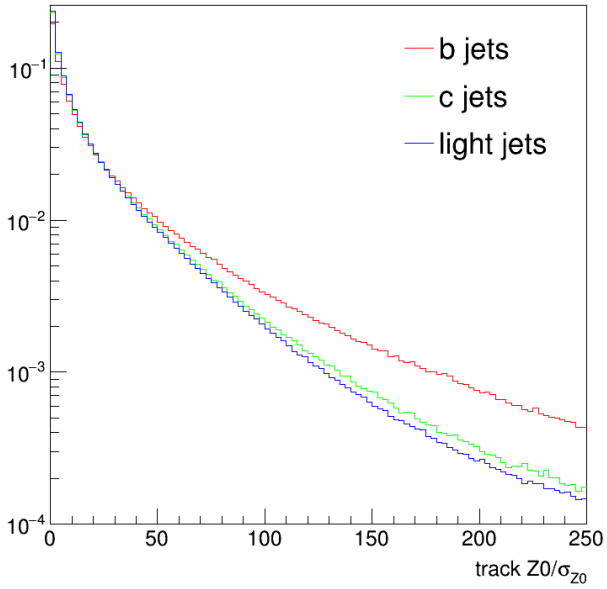
  plots in /afs/desy.de/user/m/mameyer/public/figures/deepjet

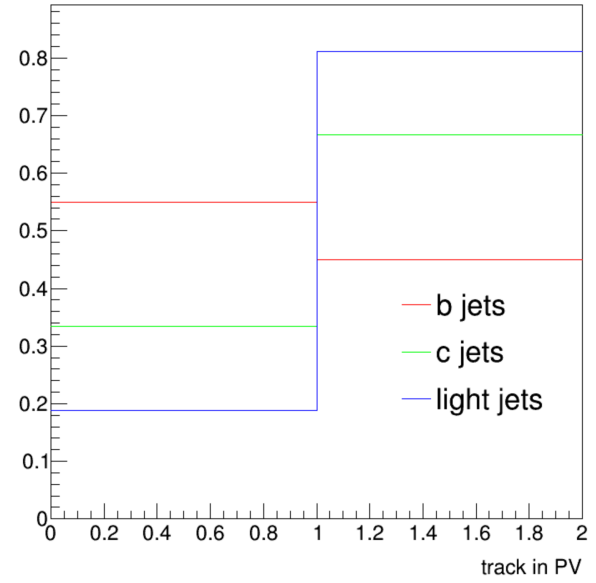
note:

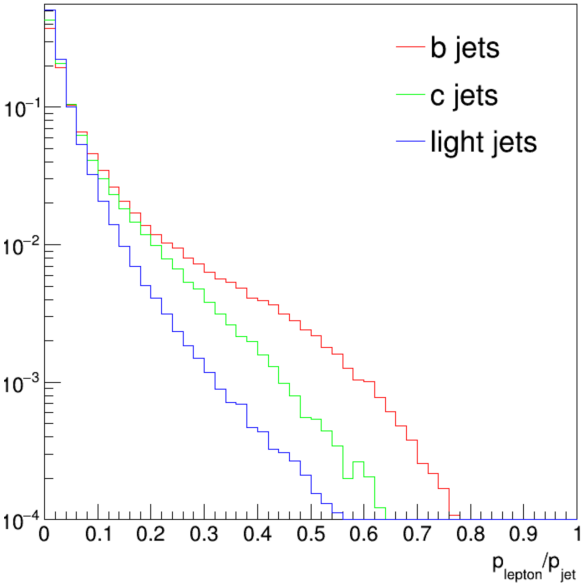
- jet-vertex assignment implemented via common tracks of jet & vertex
- in some cases vertex contains tracks of more than one jet —> take jet-vertex pair with most tracks in common (**to do**: **check**)

Input features - charged jet constituents



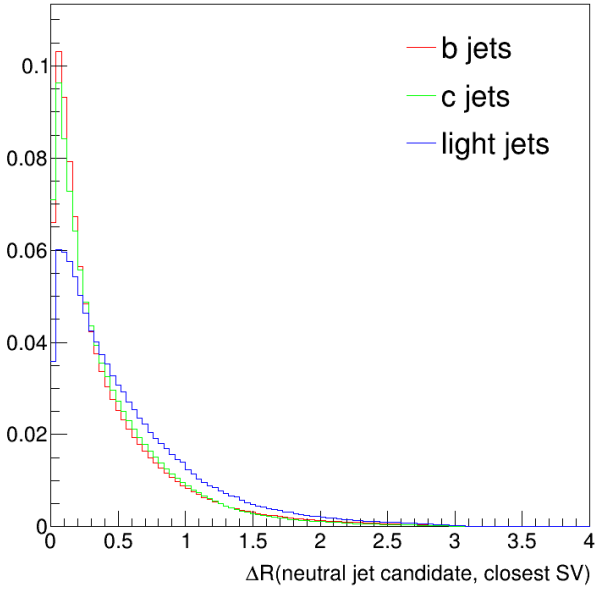


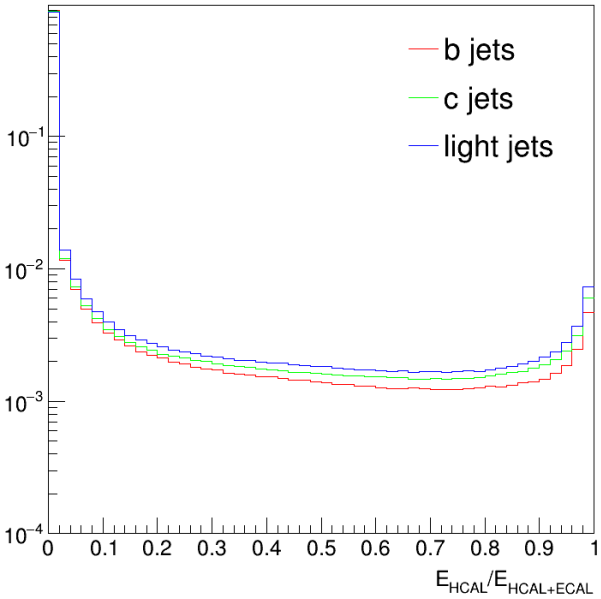




- momentum of charged jet constituent, track θ & p_T relative to jet axis, dot product of the jet and track momentum (divided by jet momentum)
- ΔR between jet axis and track
- d0, Z0, 3D IP and their significances
- fraction of the jet momentum & energy carried by track
- ΔR (track & closest SV inside jet)
- track used in PV fit?, χ2/ndf & ndf
- is electron?, is muon?
- momentum fraction of electrons/muons w.r.t. jets

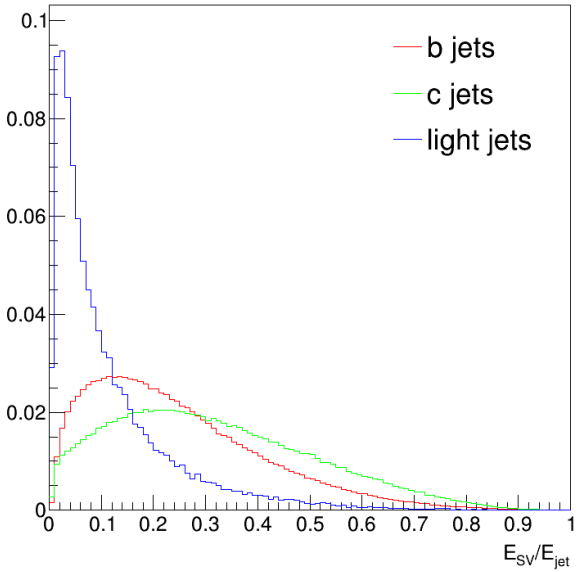
Input features - neutral jet constituents

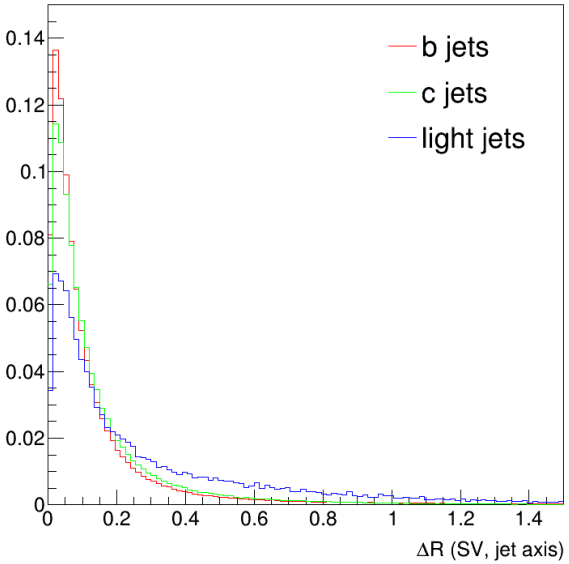


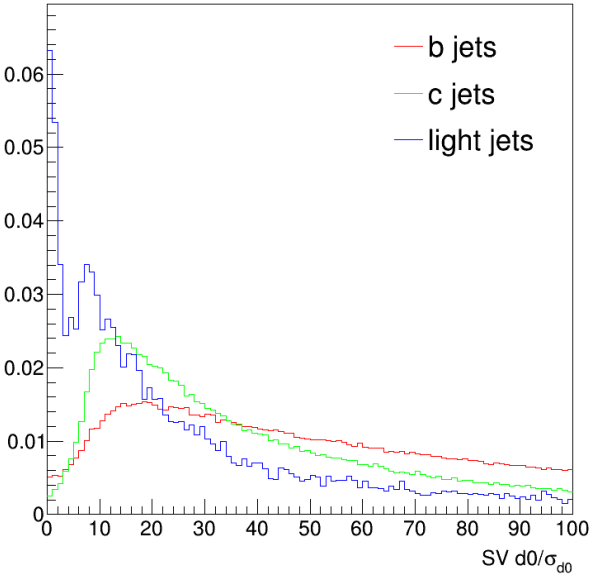


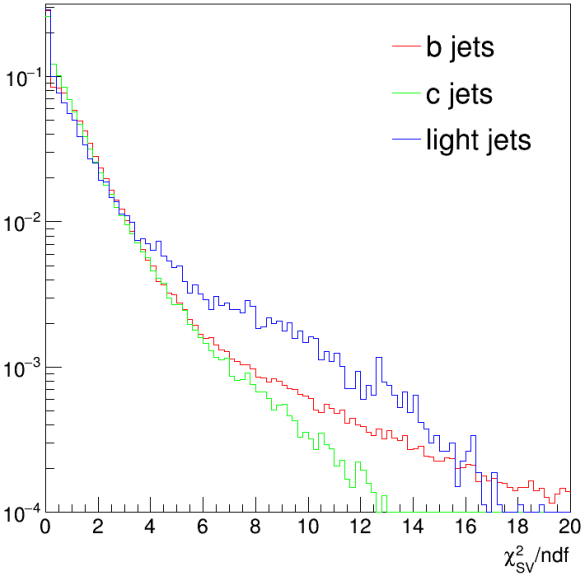
- fraction of the jet momentum & energy carried by neutral jet constituent
- ΔR(jet axis, neutral candidate),
- is photon?
- fraction of neutral candidate energy deposited in the hadronic calorimeter
- momentum of neutral jet constituent

Input features - secondary vertices





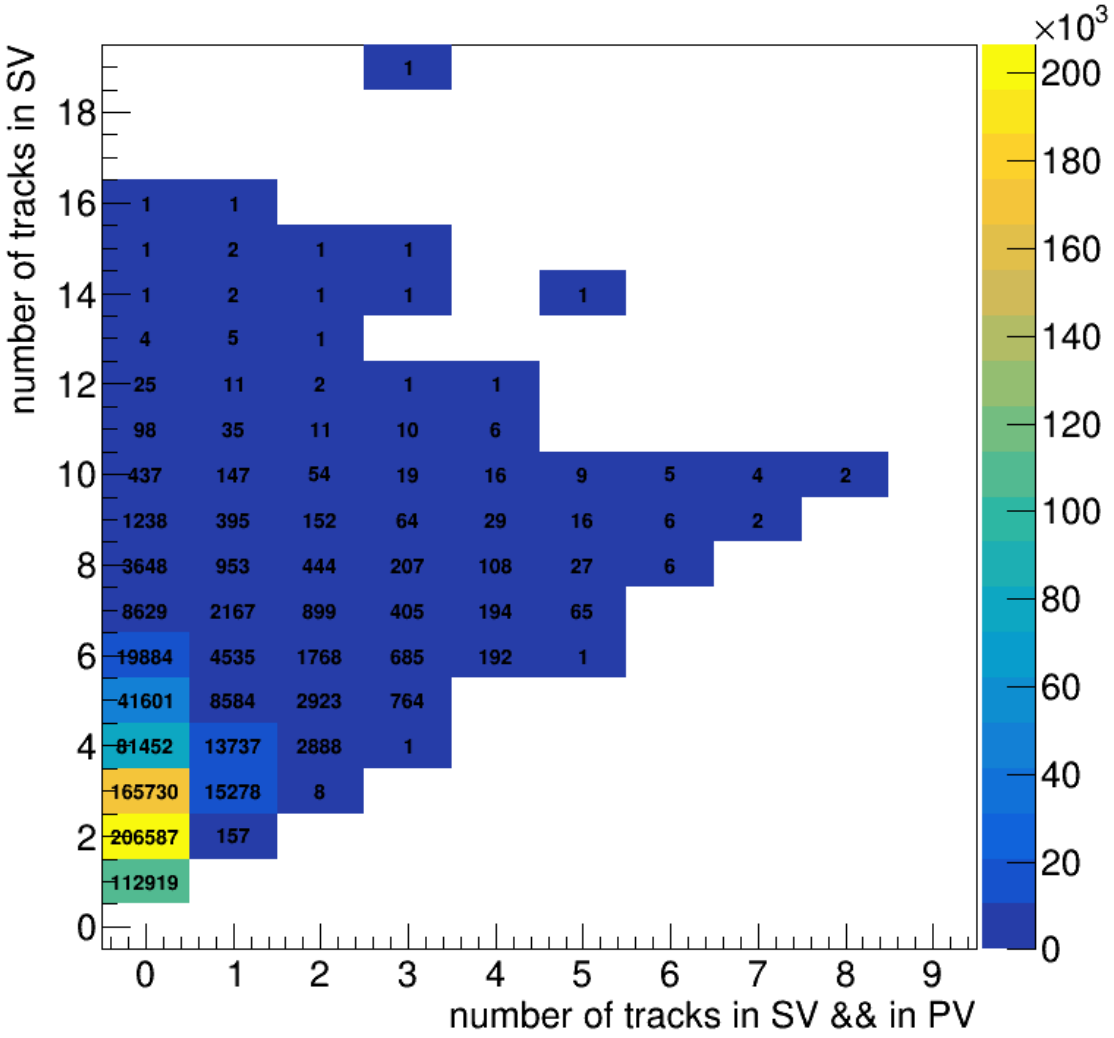




- vertex p & E,
- ΔR(closest jet, SV) & ΔR(SV, jet)
- secondary vertex mass
- number of tracks
- X2 , χ2/ndf & ndf
- d0, Z0, 3DIP and their significances
- cosine of the angle between the secondary vertex flight direction and the direction of the secodary vertex momentum
- ratio of the secondary vertex energy to the jet energy

Tracks in PV && in SV

- ~ 8% of the SVs (collection RefinedVertex) contain tracks that have also been used to reconstruct the primary vertex
- in LCFIPlus: first PV, then SV finding; tracks assigned to SV because of lower χ2; not removed from PV, no re-fitting of PV
- affects properties of PV:
  - for IP & IP significances: PV covariance matrix, PV position
  - track used in PV : fixed



Other remarks

- bug in LCFIPlus: double counting of jet constituents (reported in https://indico.cern.ch/event/817641/contributions/3413383/attachments/1836363/3009210/190430_LCFIPlus_SoftwareMeeting.pdf)

➡ checked, no double counting found

- charged reconstructed particles can have multiple tracks attached to them (kinks)
- 0.1% of all charged jet constituents

➡ choose track with highest momentum, fixed double counting of tracks

Preprocessing of features

- if a value of a features is not available, the value has to be replaced with an appropriatly chosen value
  - ΔR(SV, neutral/charged jet constituent) set to -10 if no SV found inside jet
  - neutral HCAL fraction set to -10 if no deposits inside ECAL, HCAL
- **NaNs** in significances
  - NaNs in covariance matrix of PV
- **inf** in track χ2/ndf
  - ndf = 0

} removed jets

- ordering of particle objects by (as applied by CMS)
  - impact parameter significance for charged jet constituents
  - shortest angular distance to a secondary vertex (by momentum if there is no secondary vertex) for neutral jet constituents
  - flight distance significance for secondary vertices

Other news & to do

- asked if CMS architecture public somewhere
- started to implement NN model and training

to do:

- check why vertices with tracks from more than one jet
- flavor Id variables, signed impact parameters
- particle net
- contact Taikan?

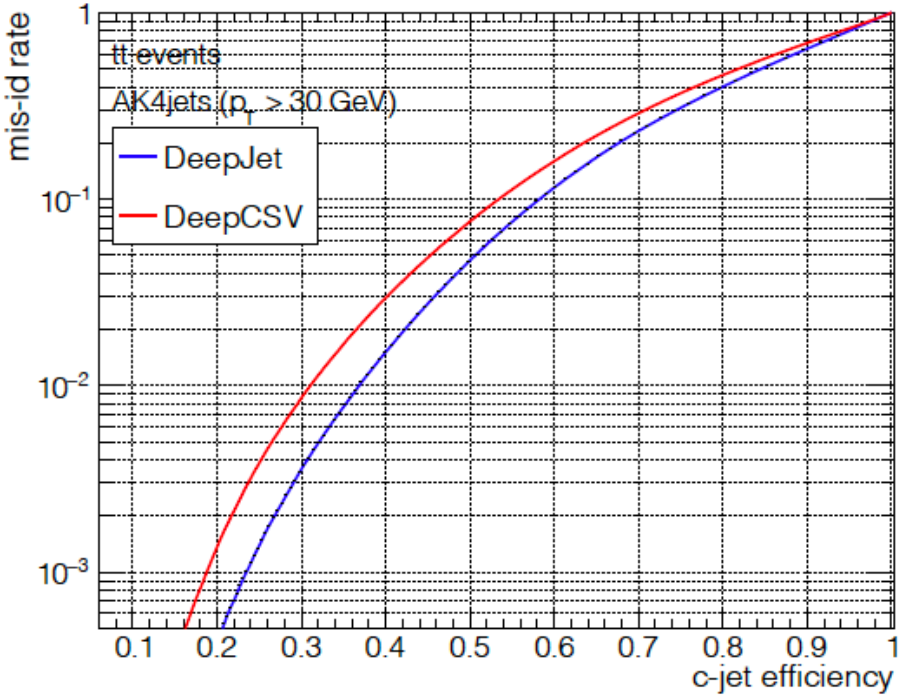
Variables used by LCFIPlus

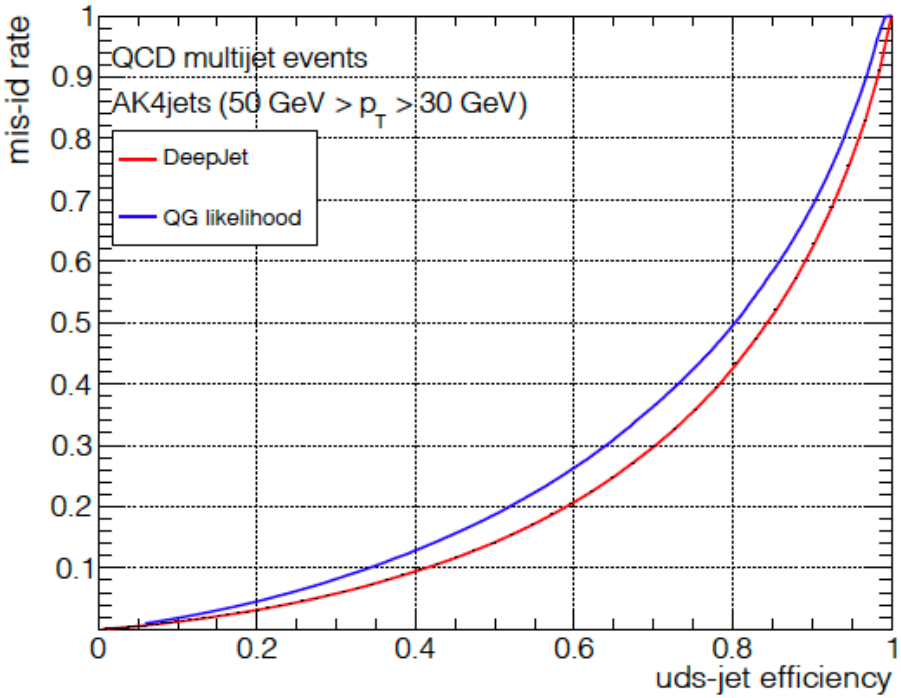
Name	Description	Normalization factor	Used by category
trk1d0sig	d0 significance of track with highest d0 significance	1	A, B, C, D
trk2d0sig	d0 significance of track with second highest d0 significance	1	A, B, C, D
trk1z0sig	z0 significance of track with highest d0 significance	1	A, B, C, D
trk2z0sig	z0 significance of track with second highest d0 significance	1	A, B, C, D
trk1pt	transverse momentum of track with highest d0 significance	$1/E_{\mathrm{jet}}$	A, B, C, D
trk2pt	transverse momentum of track with second highest d0 significance	$1/E_{\mathrm{jet}}$	A, B, C, D
jprobr	joint probability in the r-phi plane using all tracks	1	A, B, C, D
jprobr5sigma	joint probability in the r-phi plane using all tracks having impact parameter significance exceeding 5 sigma	1	A, B, C, D
jprobz	joint probability in the z projection using all tracks	1	A, B, C, D
jprobz5sigma	joint probability in the z projection using all tracks having impact parameter significance exceeding 5 sigma	1	A, B, C, D
d0bprob	product of b-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
d0cprob	product of c-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
d0qprob	product of q-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
z0bprob	product of b-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
z0cprob	product of c-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
z0qprob	product of q-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
nmuon	number of identified muons	1	A, B, C, D
nelectron	number of identified electrons	1	A, B, C, D
trkmass	mass of all tracks exceeding 5 sigma significance in d0/z0 values	1	A, B, C, D

Variables used by LCFIPlus

Name	Description	Normalization factor	Used by category
1vtxprob	vertex probability with all tracks associated in vertices combined	1	B, C, D
vtxlen1	decay length of the first vertex in the jet (zero if no vertex is found)	$1/E_{\mathrm{jet}}$	B, C, D
vtxlen2	decay length of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{\mathrm{jet}}$	D
vtxlen12	distance between the first and second vertex (zero if number of vertex is less than two)	$1/E_{\mathrm{jet}}$	D
vtxsig1	decay length significance of the first vertex in the jet (zero if no vertex is found)	$1/E_{\mathrm{jet}}$	B, C, D
vtxsig2	decay length significance of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{\mathrm{jet}}$	D
vtxsig12	vtxlen12 divided by its error as computed from the sum of the covariance matrix of the first and second vertices, projected along the line connecting the two vertices	$1/E_{\mathrm{jet}}$	D
vtxdirang1	the angle between the momentum (computed as a vector sum of track momenta) and the displacement of the first vertex	E_{jet}	B, C, D
vtxdirang2	the angle between the momentum (computed as a vector sum of track momenta) and the displacement of the second vertex	E_{jet}	D
vtxmult1	number of tracks included in the first vertex (zero if no vertex is found)	1	B, C, D
vtxmult2	number of tracks included in the second vertex (zero if number of vertex is less than two)	1	D
vtxmult	number of tracks which are used to form secondary vertices (summed for all vertices)	1	D
vtxmom1	magnitude of the vector sum of the momenta of all tracks combined into the first vertex	$1/E_{\mathrm{jet}}$	B, C, D
vtxmom2	magnitude of the vector sum of the momenta of all tracks combined into the second vertex	$1/E_{\mathrm{jet}}$	D
vtxmass1	mass of the first vertex computed from the sum of track four-momenta	1	B, C, D
vtxmass2	mass of the second vertex computed from the sum of track four-momenta	1	D
vtxmass	vertex mass as computed from the sum of four momenta of all tracks forming secondary vertices	1	B, C, D
vtxmasspc	mass of the vertex with minimum pt correction allowed by the error matrices of the primary and secondary vertices	1	B, C, D
vtxprob	vertex probability; for multiple vertices, the probability P is computed as 1-P = (1-P1)(1-P2)...(1-PN)	1	B, C, D

DeepJet - performance





Input collections

- jet collection: RefinedJets
- primary vertex collection: PrimaryVertex_NewVtx
- secondary vertex collection: RefinedVertex (includes secondary vertices and pseudo-vertices)

Input features - global variables

CMS variable	What I implemented differently	Note
Jet p_T	jet **p,** jet **p_T** & jet **energy**	
Jet η	jet **θ**, jet **φ,** jet **rapidity**	
The number of charged particle flow candidates in the jet		
The number of neutral particle flow candidates in the jet		
The number of secondary vertices in the jet		SVs & pseudo-vertices have to be assigned to a jet, implemented via common tracks of jet and vertex, in some cases a vertex contains tracks of more than one jet —> take jet-vertex pair with most tracks in common
The number of primary vertices in the event	replaced with **number of pseudo vertices** in the jet	

Input features - charged jet constituents

CMS variable	What I implemented differently	Note
Charged track η relative to the jet axis	charged track **θ** relative to the jet axis	
Charged track p_T relative to the jet axis		
Dot product of the jet and track momentum		
Dot product of the jet and track momentum divided by the magnitude of the jet momentum		
ΔR between the jet axis and the track	ΔR(η,φ) -> **ΔR(θ,φ), ΔR(y,φ)**	
The track 2D impact parameter value	added additionally **Z0**	
The track 2D impact parameter significance	added additionally **Z0 significance**	
The track 3D impact parameter value		not calculated via closest approach of jet in 3D, but with sqrt($d0^2$+dZ^2)

Input features - charged jet constituents

CMS variable	What I implemented differently	Note
The track 3D impact parameter significance		
The track distance to the jet axis	Closest approach of track to jet axis, used HelixClass from MarlinUtil, approximate jet as straight helix	has to be checked
Fraction of the jet momentum carried by the track	fraction of the **momentum & energy** carried by the track	
ΔR between the track and the closest secondary vertex	ΔR(η,φ) -> **ΔR(θ,φ)**	consider only vertices inside the jet
An integer flag that indicate whether the track was used in the primary vertex fit		
The charged candidate' s PUPPI weight	not applicable —> not implemented	
χ2 of the charged track fit	stored **χ2/ndf & ndf**	
An integer flag which indicate the quality of the fitted track, based on number of detector hits used for the reconstruction as well as the overall χ2 of the charged track fit	values? not known, idea: check track quality criteria used in MarlinPandora or take number of hits in different sub-systems	not implemented yet

Input features - charged jet constituents

- additional variables not used by CMS:
  - charge of charged jet constituents
  - integer flag indicating whether charged candidate is an electron
  - integer flag indicating whether charged candidate is a muon
  - momentum of charged jet constituent

Input features - neutral jet constituents

CMS variable	What I implemented differently	Note
Fraction of the jet momentum carried by the neutral candidate	Fraction of the **jet momentum** and fraction of the **jet energy**	
ΔR between the jet axis and the neutral candidate	ΔR(η,φ) -> **ΔR(θ,φ) & ΔR(y,φ)**	
An integer flag indicating whether the neutral candidate is a photon		
Fraction of the neutral candidate energy deposited in the hadronic calorimeter		
ΔR between the neutral candidate and the closest secondary vertex	ΔR(η,φ) -> **ΔR(θ,φ)**	consider only vertices inside the jet
The neutral candidates PUPPI weight	not applicable —> not implemented	

- additional variables not used by CMS:
  - momentum of neutral jet constituent

Input features - secondary vertices

CMS variable	What I implemented differently	Note
Secondary vertex p_T	vertex **momentum & energy**	
ΔR between the jet axis and the secondary vertex	ΔR(η,φ) -> **ΔR(θ,φ), find closest jet in ΔR & ΔR(θ,φ) between SV and jet SV is assigned to**	jet closest in ΔR is not always the jet the vertex is assigned to
Secondary vertex mass		
Number of tracks in the secondary vertex		
X2 of the secondary vertex fit		
Reduced X2 of the secondary vertex fit	stored **χ2/ndf & ndf**	
The secondary vertex 2D impact parameter value	added additionally **Z0**	
The secondary vertex 2D impact parameter significance	added additionally **Z0 significance**	

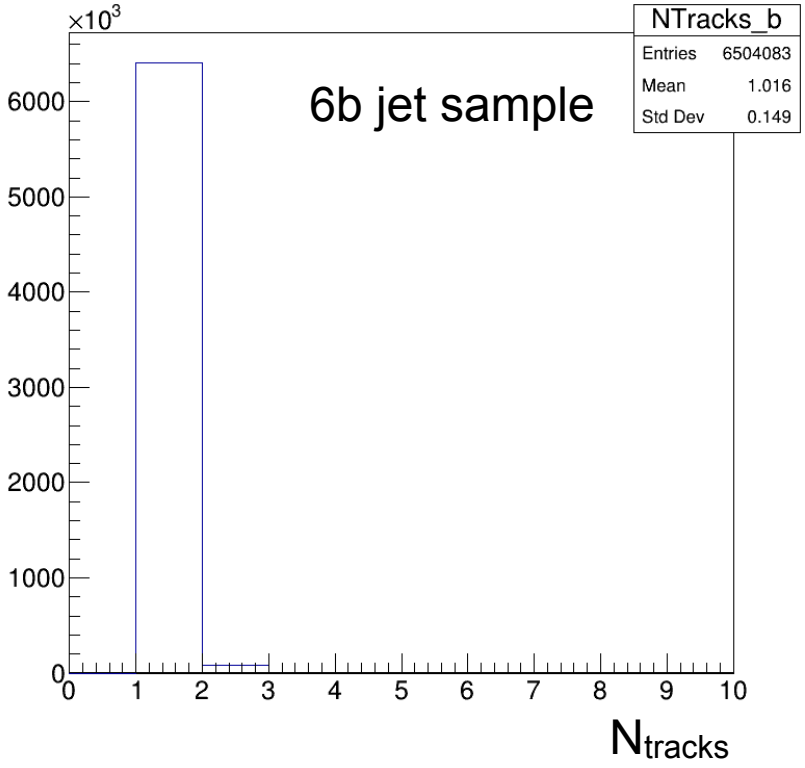
Input features - secondary vertices

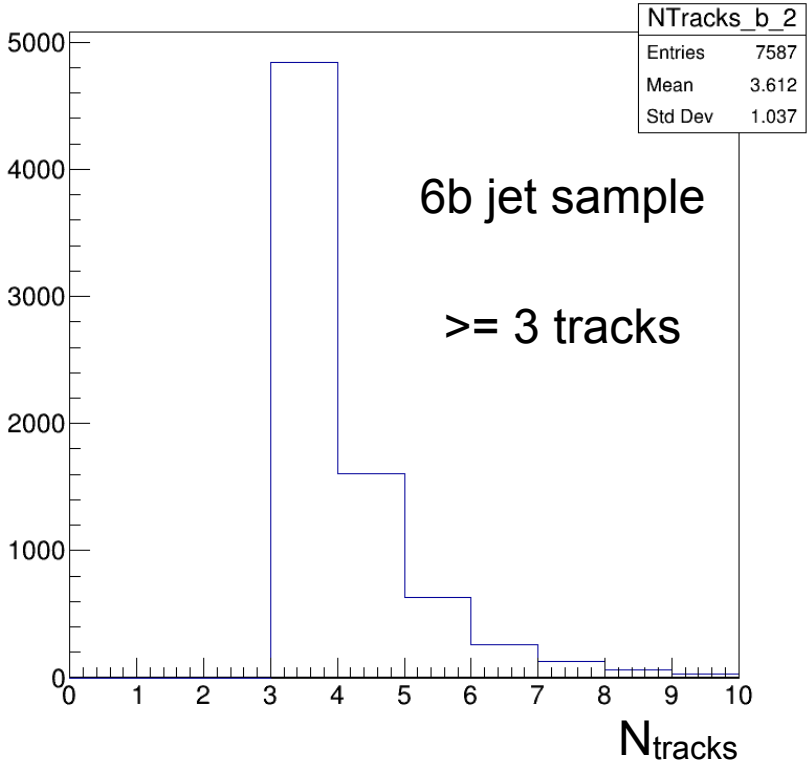
CMS variable	What I implemented differently	Note
The secondary vertex 3D impact parameter value		sqrt($d0^2+dZ^2$)
The secondary vertex 3D impact parameter significance		
Cosine of the angle between the secondary vertex flight direction and the direction of the secodary vertex momentum		
Ratio of the secondary vertex energy to the jet energy		take jet assigned to SV via number of common tracks

Charged jet candidates with multiple tracks

- charged reconstructed particles can have multiple tracks attached to them (kinks)
- 0.1% of all charged jet constituents

➡ choose track with highest momentum, fixed double counting of tracks





- consistent with Yasser's studies (https://indico.desy.de/event/29952/contributions/103222/attachments/65812/81377/Yasser_21_04_27.pdf)