Preparation for Week 1

Laurie McClymont

UCL ATLAS Meeting
29/05/15

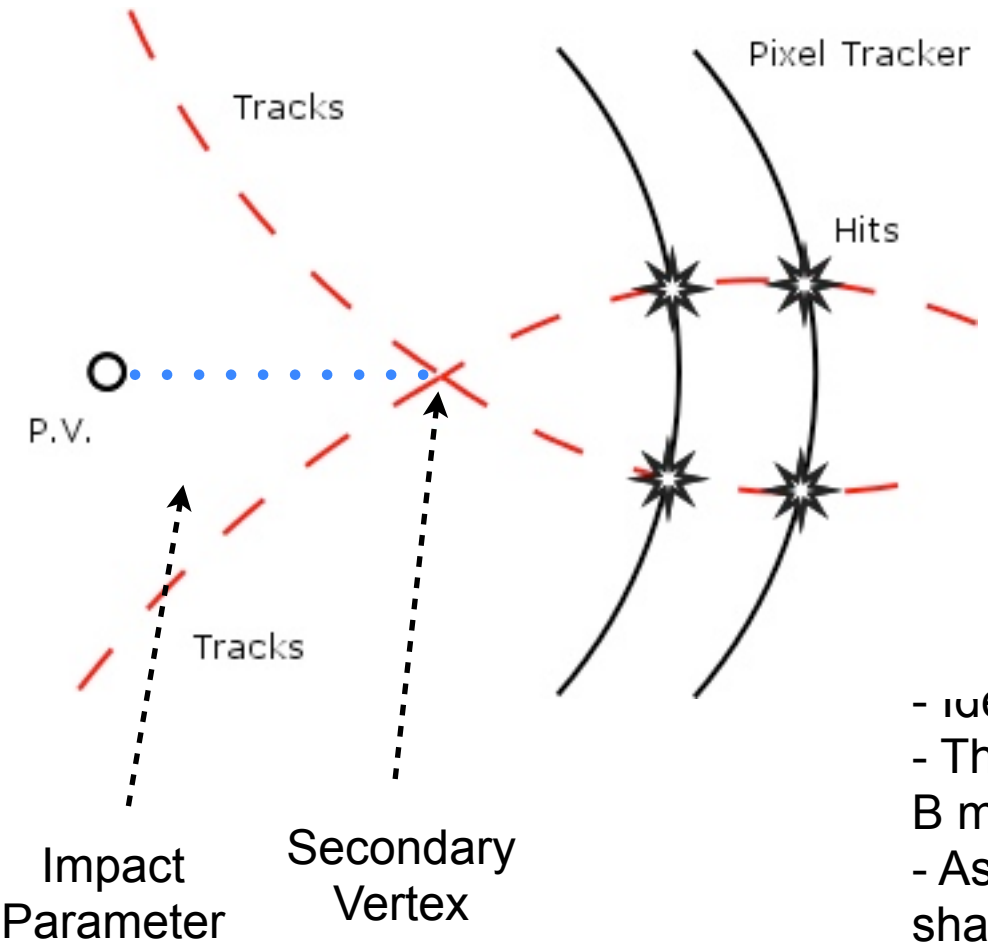
Aims

- First data available in Week 1 (Next Week!)
- We want to look at this data and see if the b-Tagging variables are performing as expected.

Process

- Valerio Produces NTuples using Run2BtagOptimisationFramework.
- I have code that reads the NTuples and fills histograms with quantities from the NTuples.
- Code takes ~10 mins to run over NTuples.

B-tagging - IP3D, SV1, JF



IP3D

- Look for tracks with a large impact parameter significance = (I.P. / σ)

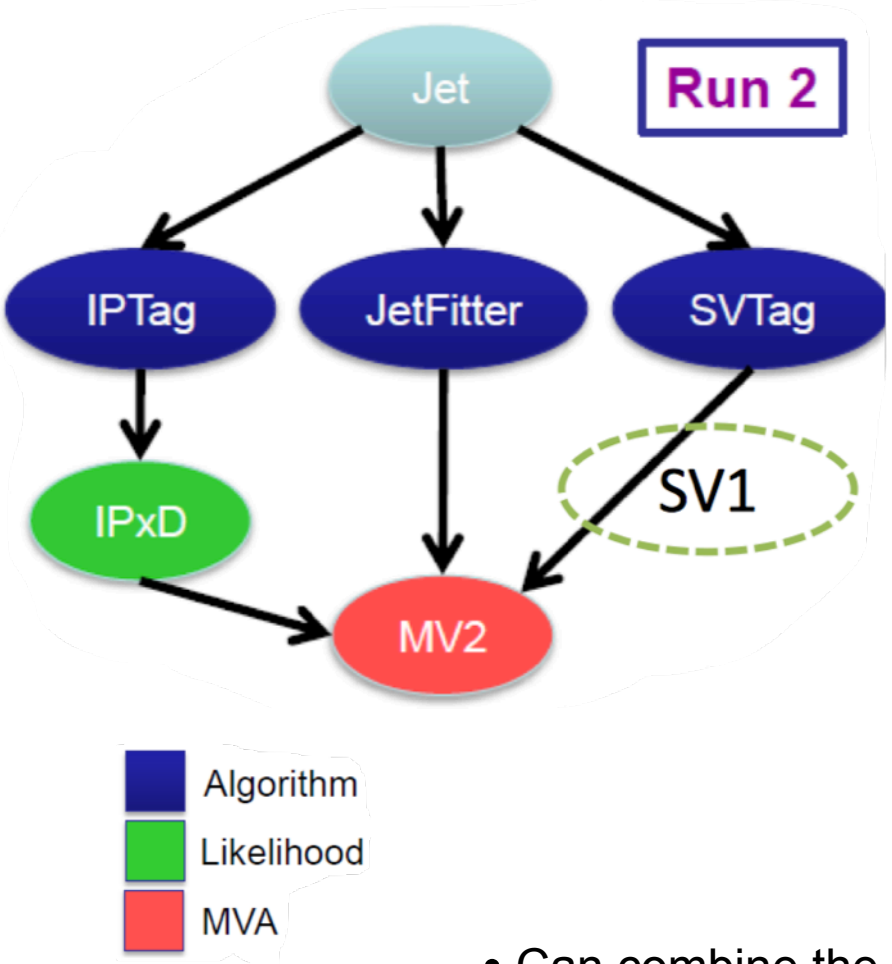
SV1

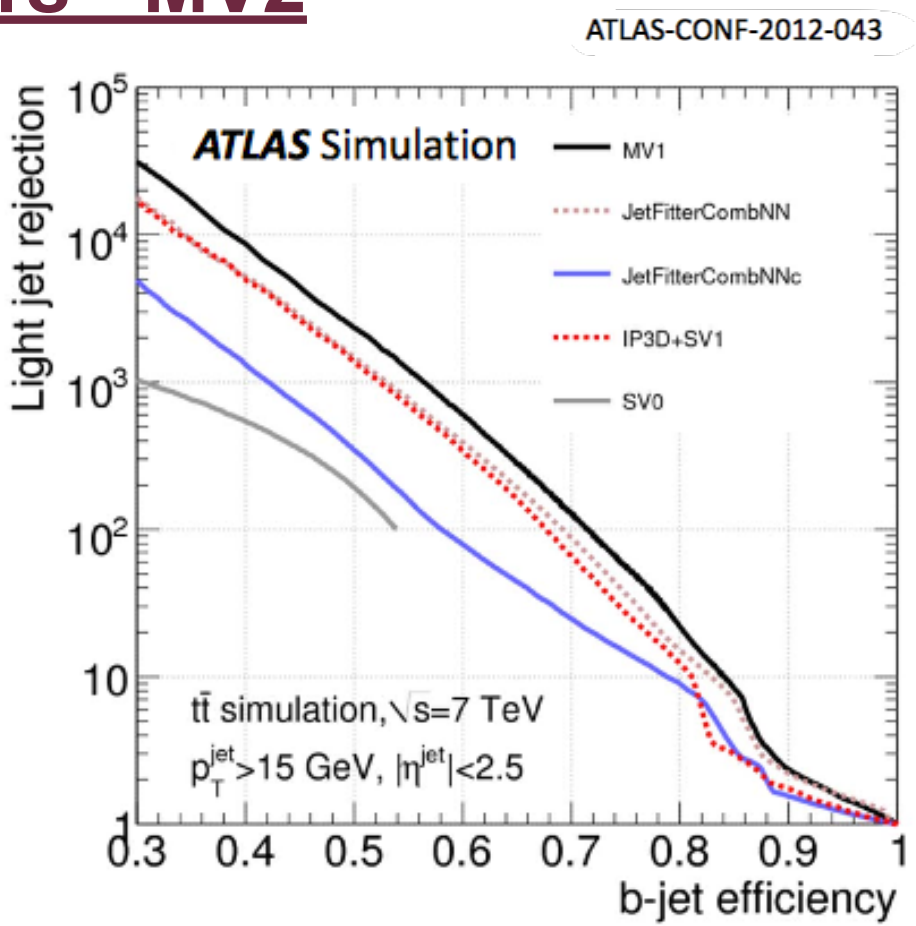
- Identify a secondary vertex
- Look for large flight path significance = (F.P. / σ)

JF

- Identify Second and Tertiary Vertices.
- These vertices correspond to decay of B meson and decay of D meson.
- Assumes these vertices lie on an shared B flight axis

Multi-Variate B Taggers - MV2





- Can combine the basic algorithms in a neural network.
- This leads to improved performance.
- MV2c20 - Trained on a sample containing 20% charm.

Samples and Cuts

Samples

```
user.vdao.mc15_13TeV.*.Pythia8EvtGen_jetjet_JZ*W
.merge.AOD.*.BTAGNTUP_OrigV5slim_BTAGSTREAM/
```

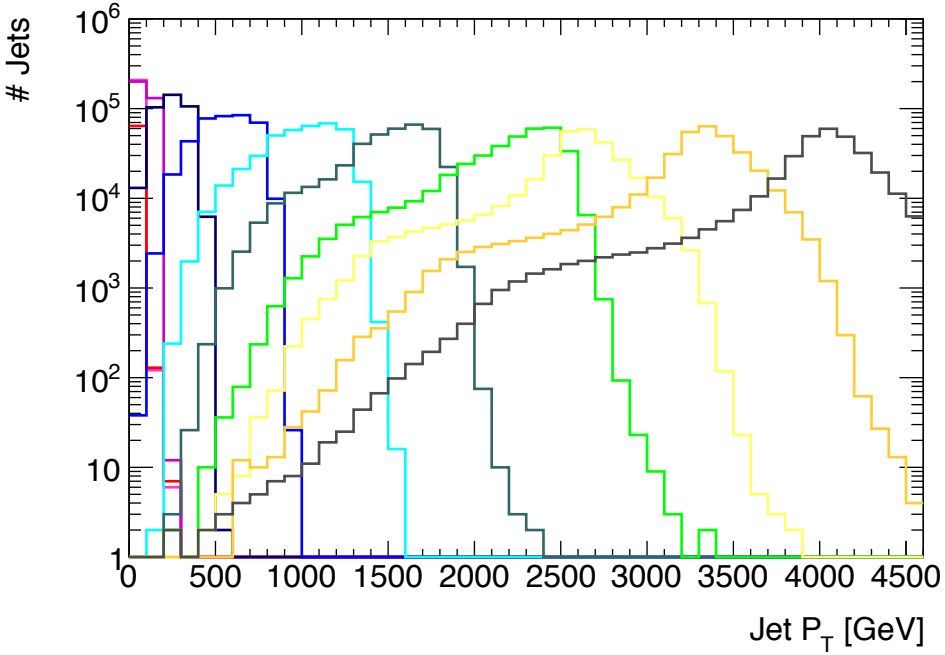
- A week 1 dijet sample for comparison to data.
- 2,161,636 split into 10 JZ slices each containing ~200,000 events.
- JZ slices must be re-weighted to get smooth jet-P_T spectrum.

Details/Cuts

- $20 < P_T < 4600$ GeV
- Leading and Sub-Leading Jet Only
- |eta| < 2.5
- JVT > 0.941
- njets ≥ 2

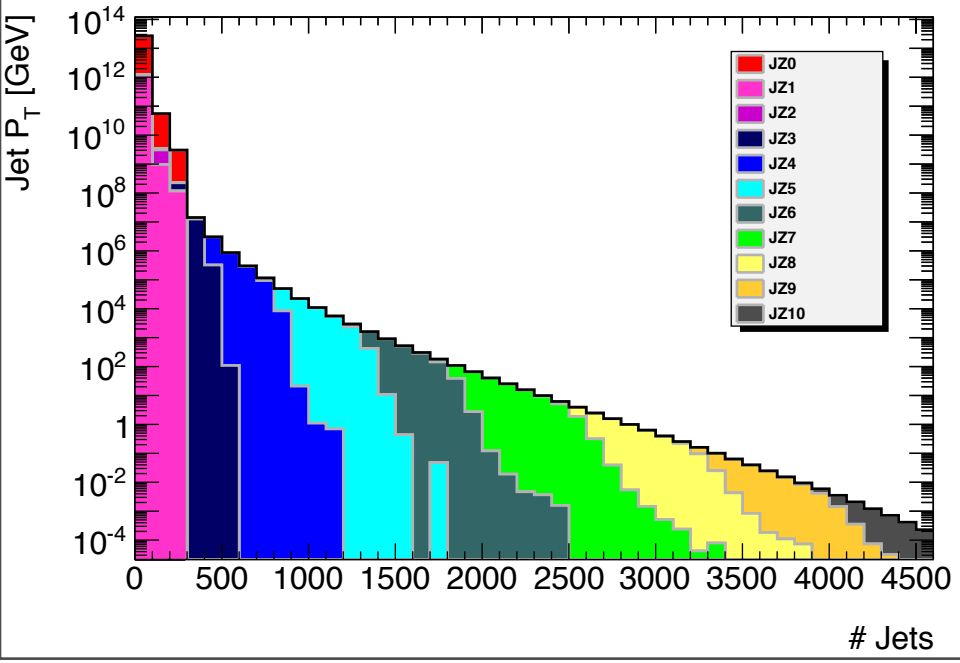
- LabDr_HadF truth matching.
- AntiKt4EMTopoJets

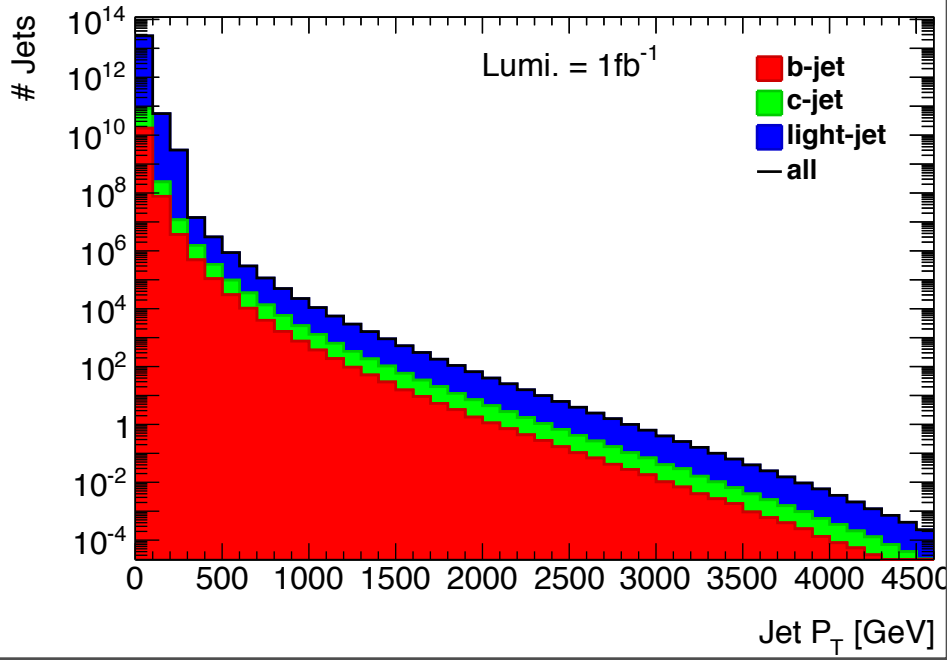
Di-jet sample re-weighting



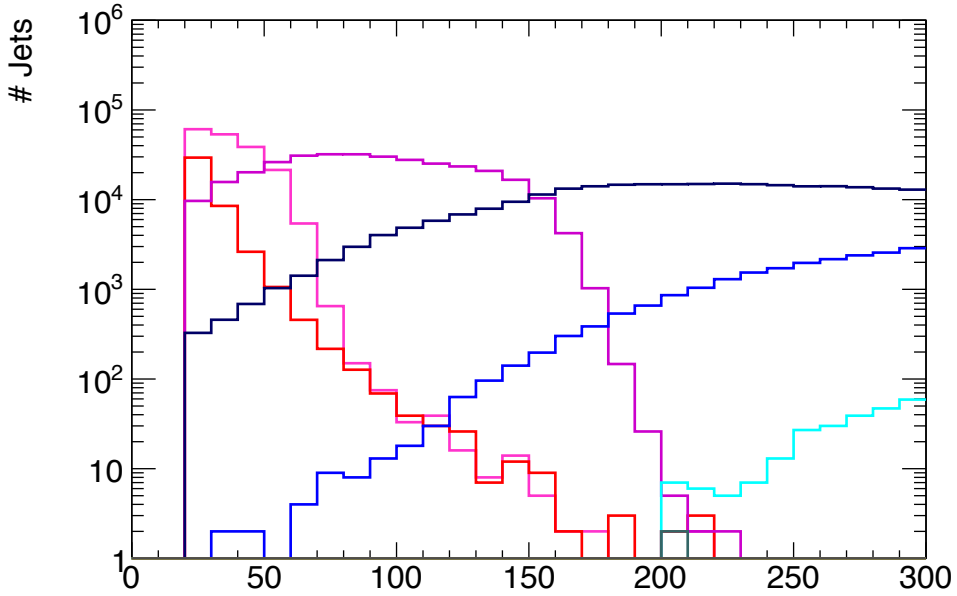
Total Weight = *mcwg*(Filter Eff.)*(CS[fb])*(Lumi[fb^{-1}])* / *(# Events)*

Xs(fb)	Eff.	Slice and Energy
7.8420E+13	1.0240E+00	#JZ0W 0-20 GeV
7.8420E+13	6.7198E-04	#JZ1W 20-60 GeV
2.4334E+12	3.3264E-04	#JZ2W 60-160 GeV
2.6454E+10	3.1953E-04	#JZ3W 160-400 GeV
2.5464E+08	5.3009E-04	#JZ4W 400-800 GeV
4.5536E+06	9.2325E-04	#JZ5W 800-1300 GeV
2.5752E+05	9.4016E-04	#JZ6W 1300-1800 GeV
1.6214E+04	3.9282E-04	#JZ7W 1800-2500 GeV
6.2505E+02	1.0162E-02	#JZ8W 2500-3200 GeV
1.9640E+01	1.2054E-02	#JZ9W 3200-3900 GeV
1.1961E+00	5.8935E-03	#JZ10W 3900-4600 GeV



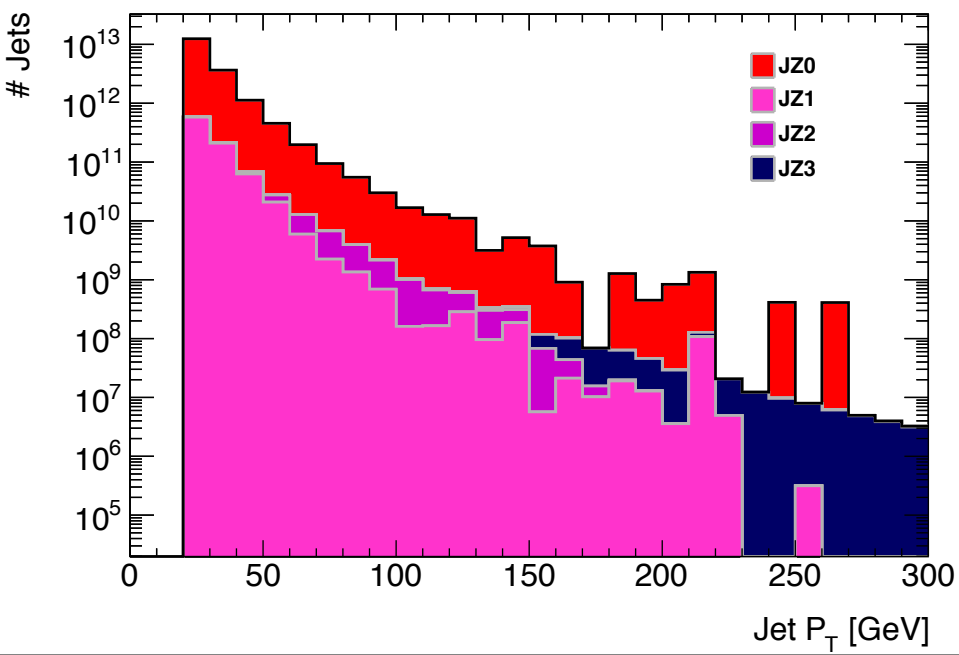


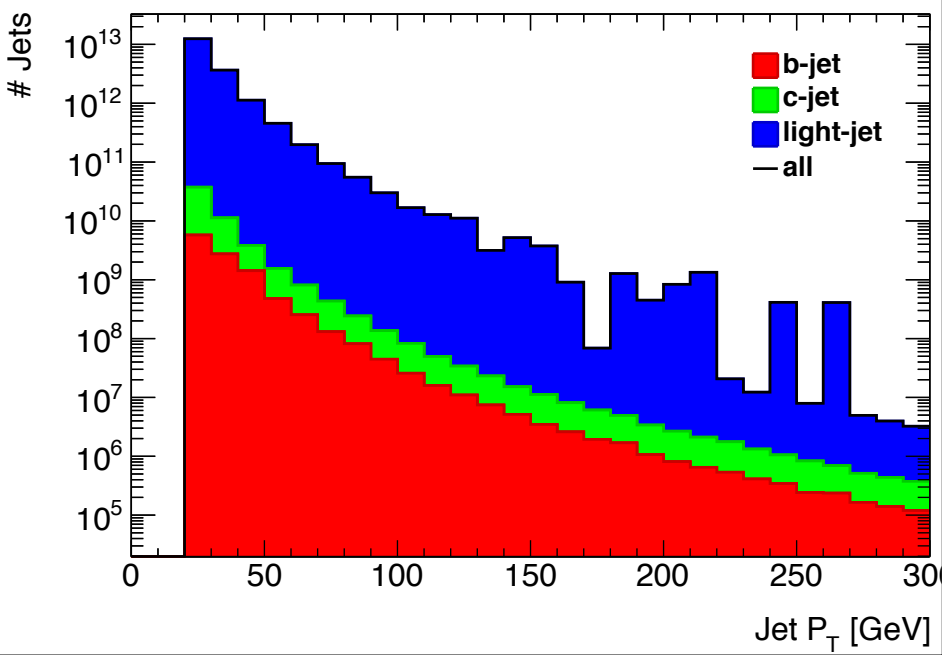
Di-jet sample re-weighting - Zoom In



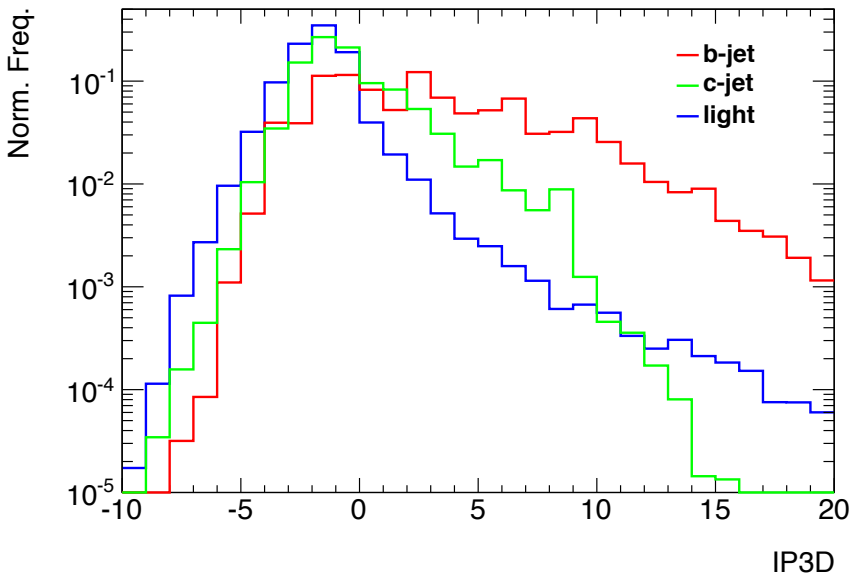
Jet P_T = 20-300 GeV

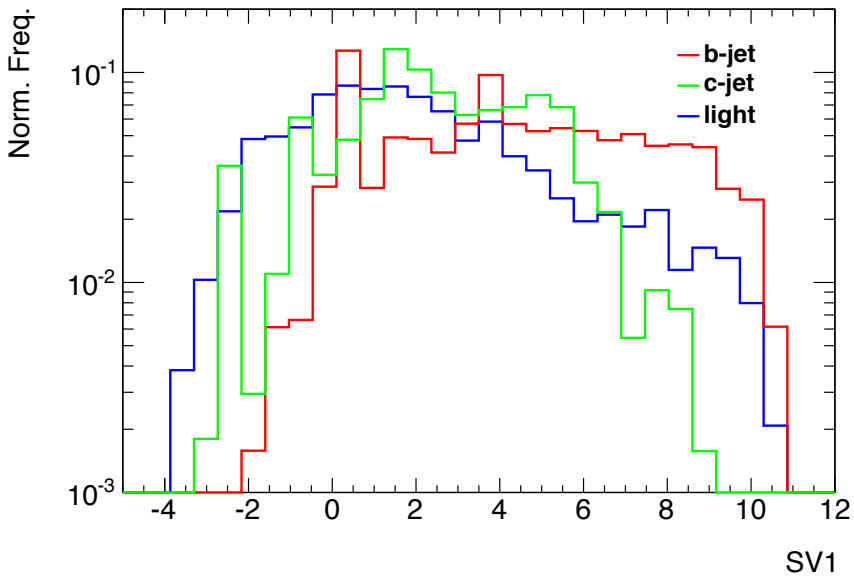
Xs (fb)	Eff.	Slice and Energy
7.8420E+13	1.0240E+00	#JZ0W 0-20 GeV
7.8420E+13	6.7198E-04	#JZ1W 20-60 GeV
2.4334E+12	3.3264E-04	#JZ2W 60-160 GeV
2.6454E+10	3.1953E-04	#JZ3W 160-400 GeV
2.5464E+08	5.3009E-04	#JZ4W 400-800 GeV

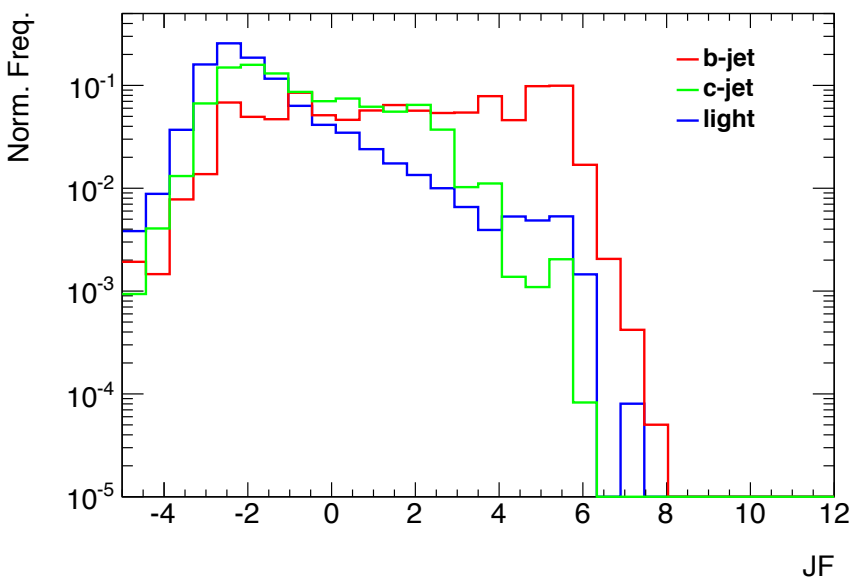


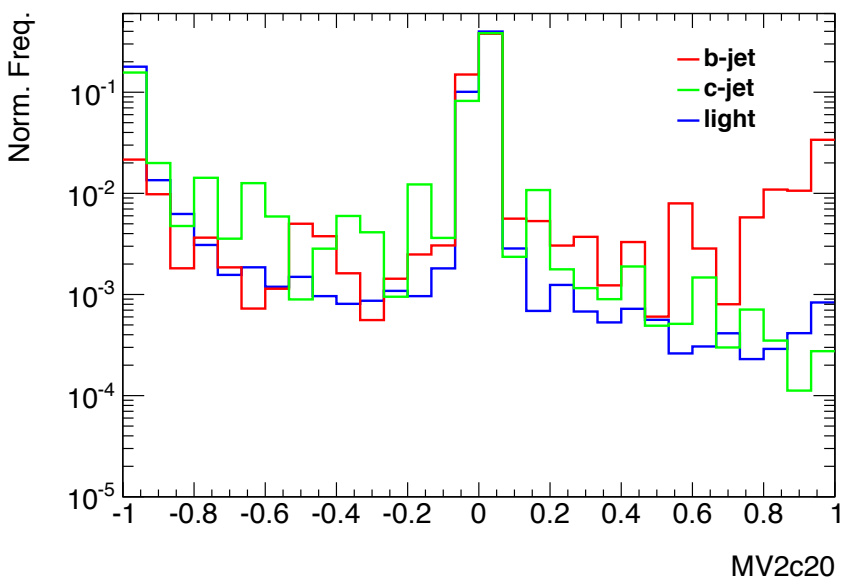


Normalised Discriminants - Full P_T Range



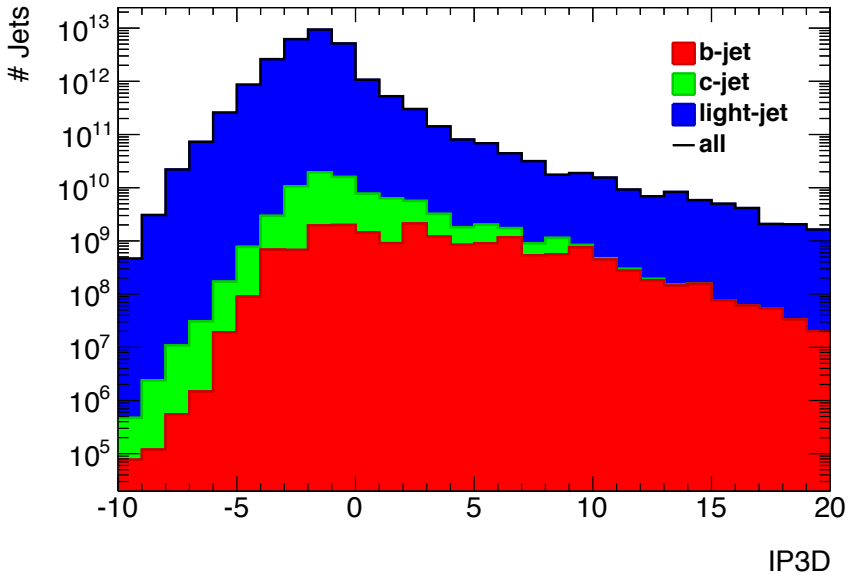


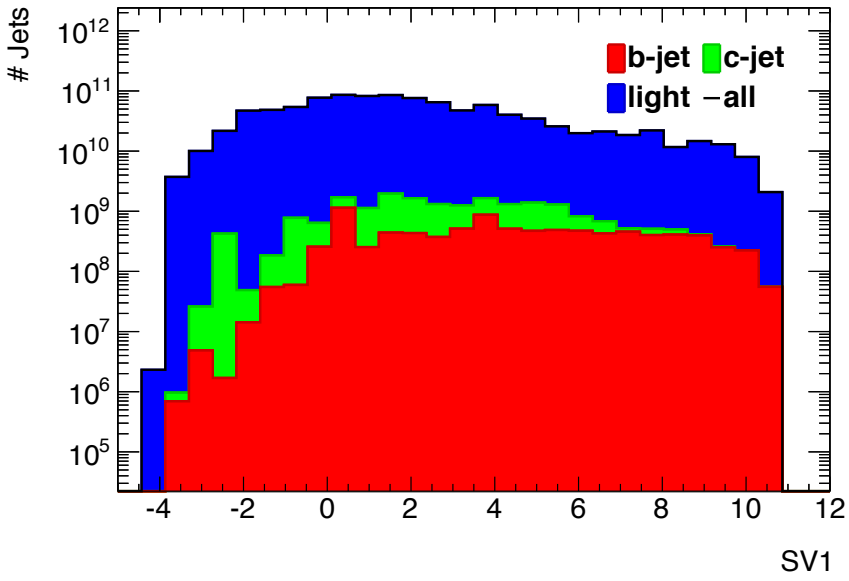


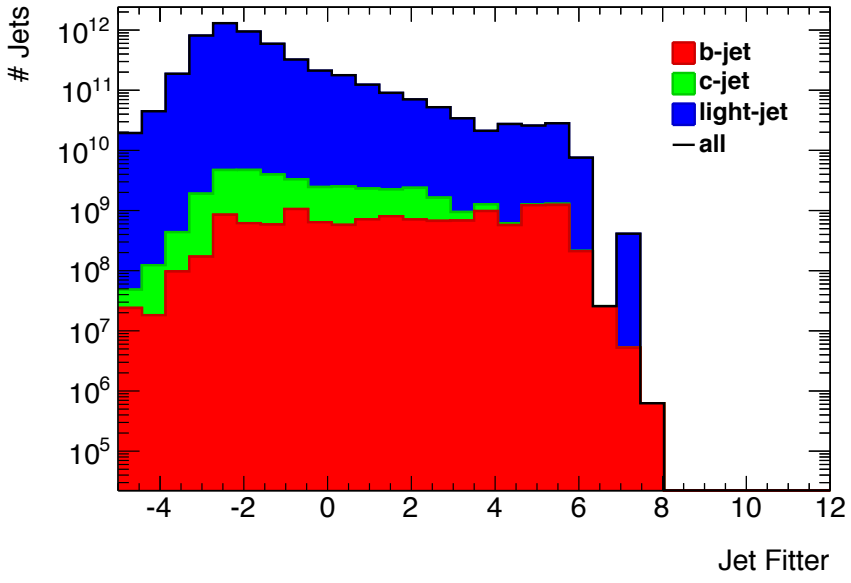


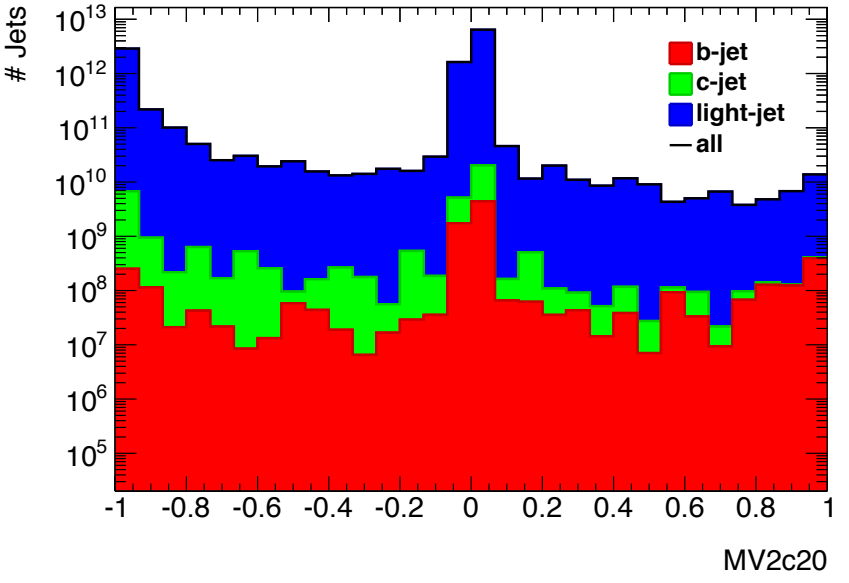
Discriminant Stack - Full P_T Range

P_T Range = 20-4600 GeV



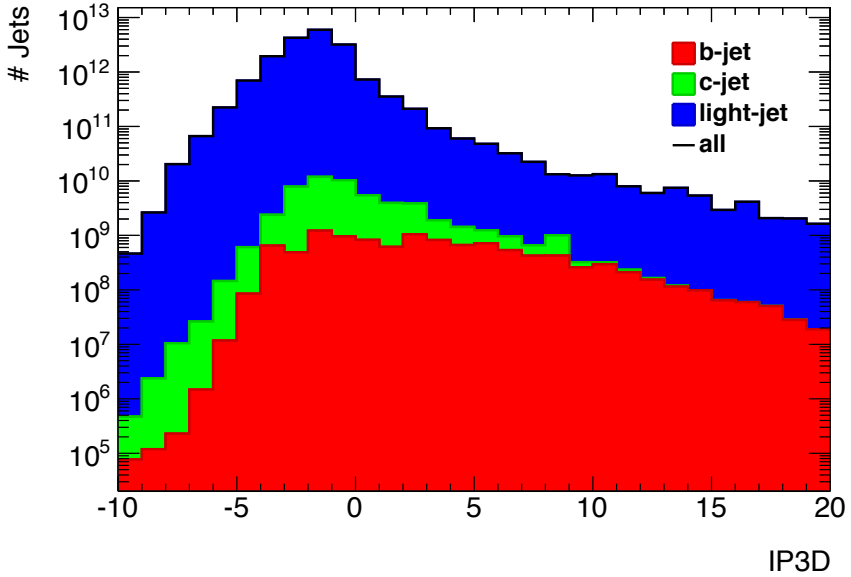


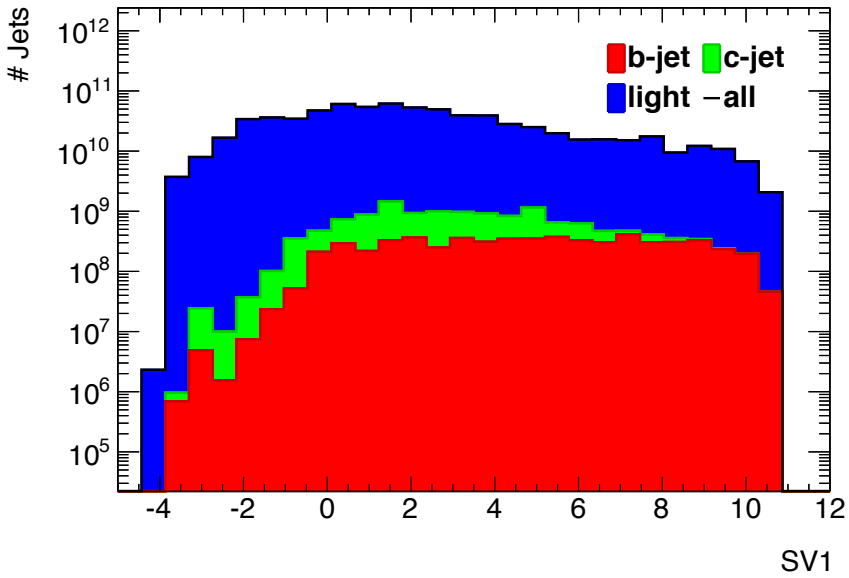


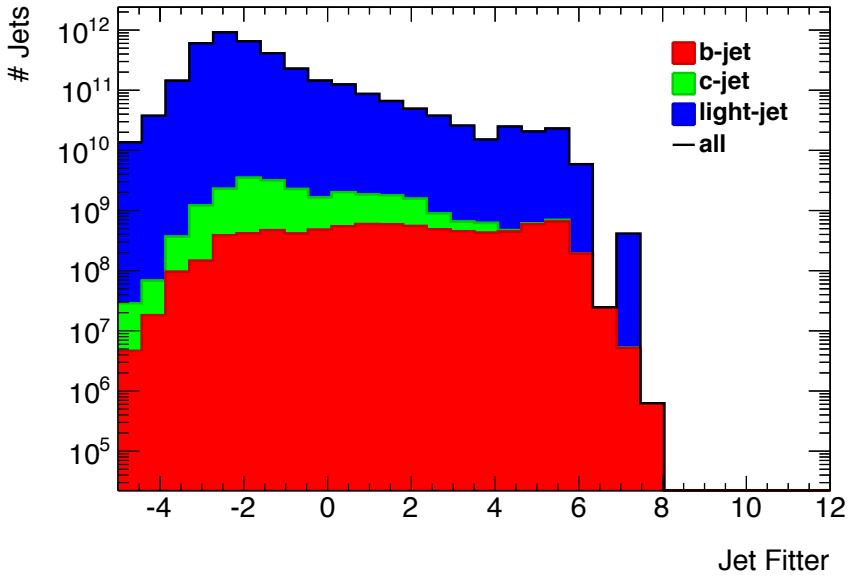


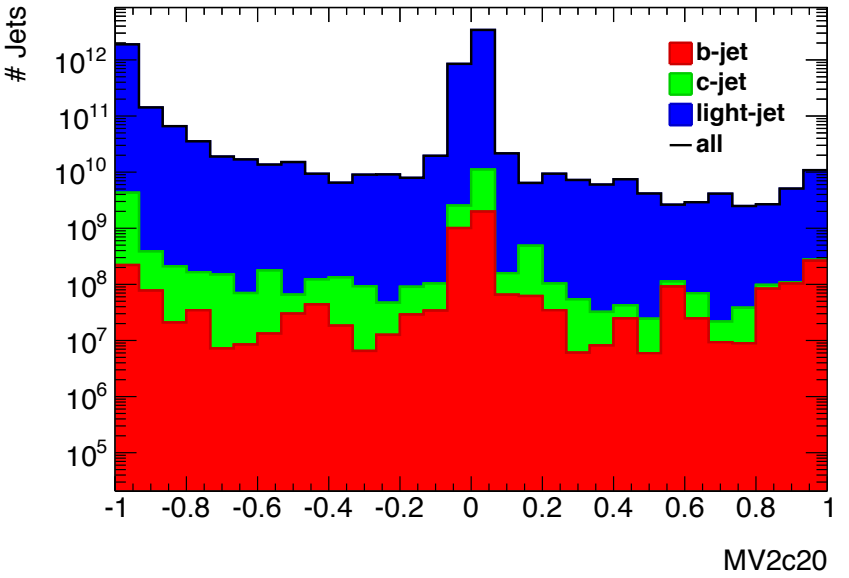
Discriminant Stack - 20 < P_T < 300 GeV

P_T Range = 20-300 GeV



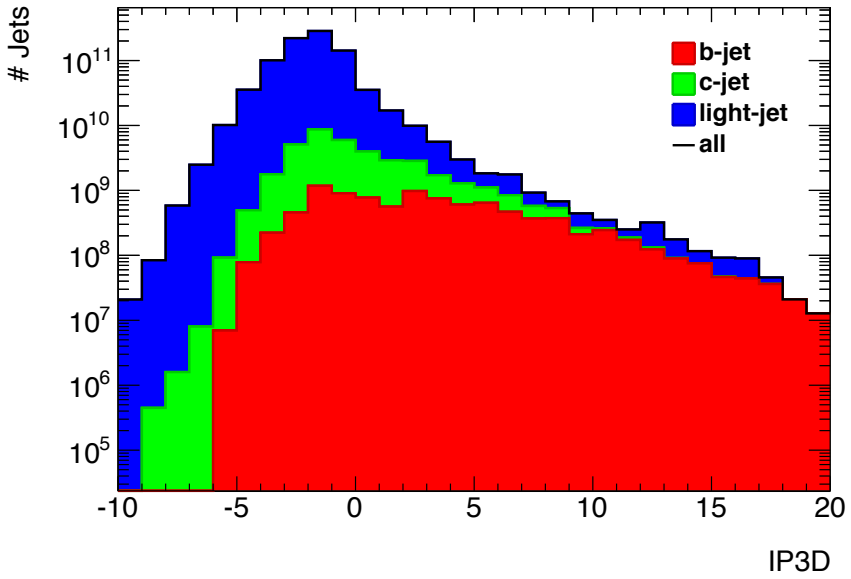


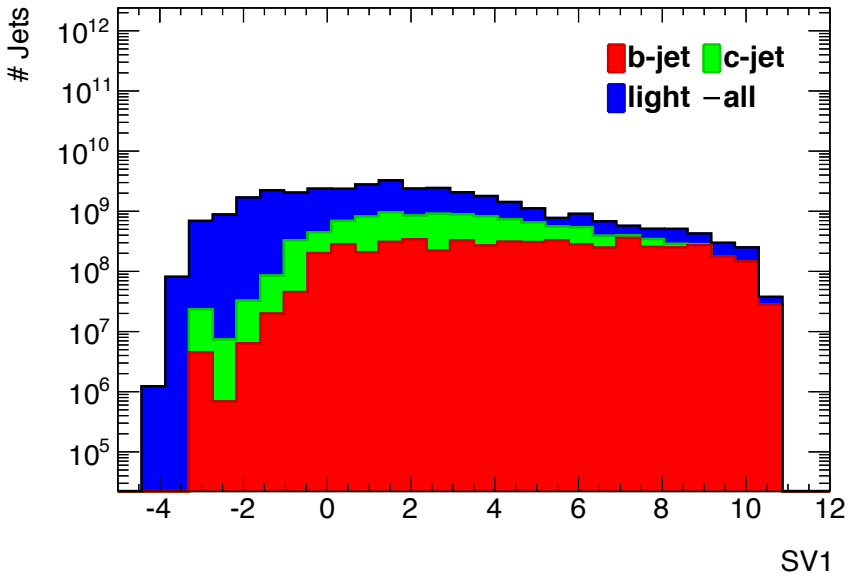


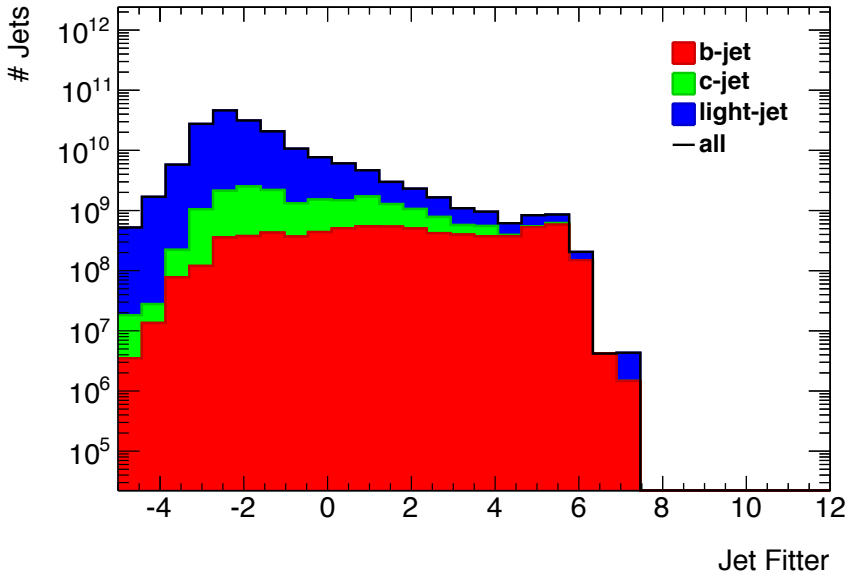


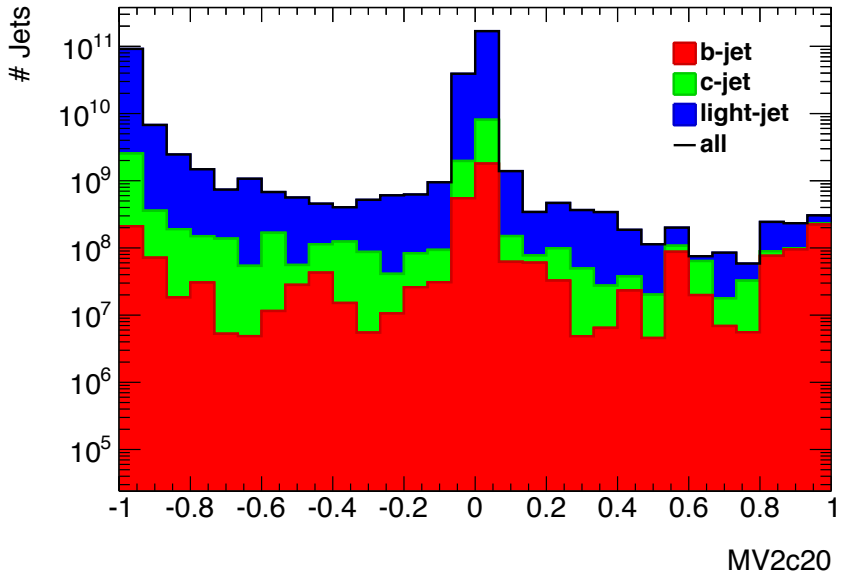
Discriminant Stack - JZ1W

JZ1W - 20-60 GeV



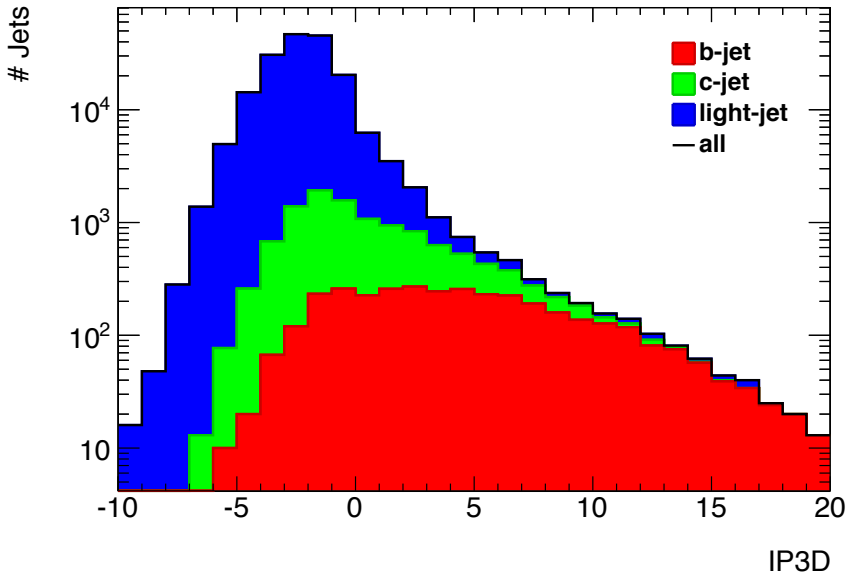


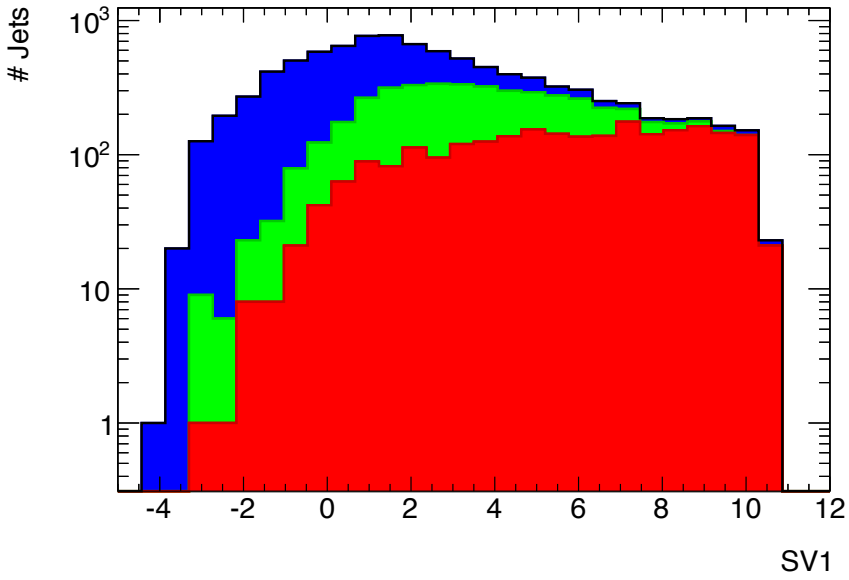


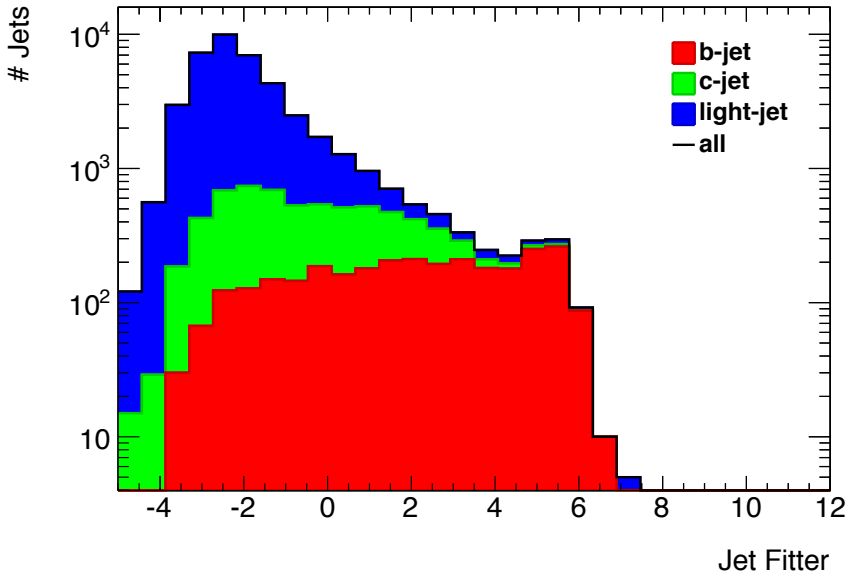


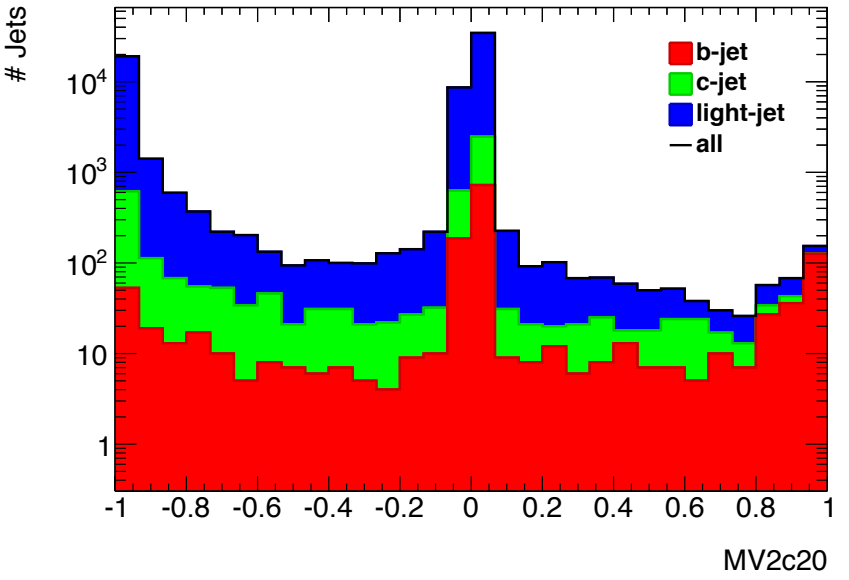
Discriminant Stack - JZ1W - no Weights applied.

JZ1W - 20-60 GeV Weight = 1









To Do

- Still work to do!

- We want to better understand shape of the discriminant.
  -Why is there a peak at 0 in the mv2c20?

  - Plot Leading Jet Only
  - Plot P_T slices
  - mv2c00/mv2c10
  - Show the under-fill bin - algorithms have no output.
  - Check weighting

- Understand Sample

  - PV Distributions.

UCL

Track Selection Optimisation for high-P_T b-Tagging update

Track Cuts and Some Definitions

Track Cuts

From Talk by R. Zaidan at Flav Tag Workshop 2015

In this study I have applied these cuts manually in my analysis code.

	IP3D	SV1	JFit
$p_T \geq$	1000	700	769.2
$\lvert\eta\rvert \leq$	2.5	2.5	2.5
$N_{SI} \geq$	7	7	7
$N_{SCT} \geq$	-	4	4
$N_{PIX} \geq$	2	1	1
$N_{IBL} \geq$	1	-	-
$N_{IBL} + N_{BL} \geq$	-	-	-
$N^{SH}_{PIX} + \frac{N^{SH}_{SCT}}{2} \leq$	-	-	1
$N^{HOLE}_{SI} \leq$	-	-	-
$N^{HOLE}_{PIX} \leq$	-	-	-
$d_0 \leq$	1	5	3.5
$z_0 * \sin(\theta) \leq$	1.5	25	5
$\sigma(d_0) \leq$	-	1	0.35
$\sigma(z_0) \leq$	-	5	2.5
$\chi 2/NDF \leq$	-	-	3.5

Definitions

From B = Any track associated to the decay of the B or C Hadron

From PuFake = Any track without a well matched truth particle.

From Geant = Any track created by a GEANT interaction.

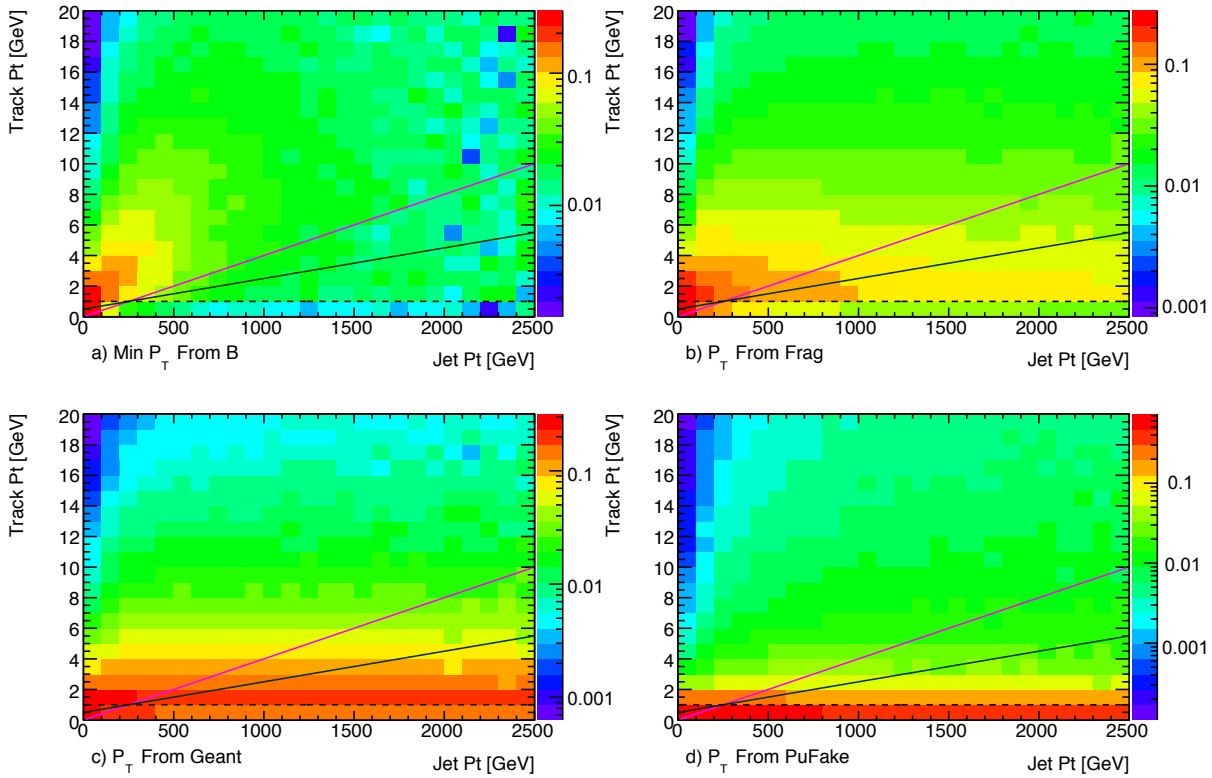
From Frag = Any track not From B or From Geant

From Other = Any track not From B

Selection Efficiency = $\frac{\text{\# Tracks FromX Selected By a Cut}}{\text{\# Truth Tracks From X}}$

Fraction of Tracks = $\frac{\text{\# Tracks FromX Selected By a Cut}}{\text{Total \# Tracks Selected by a Cut}}$

Track Optimisation for High-P_T b-Tagging



Z' bb sample

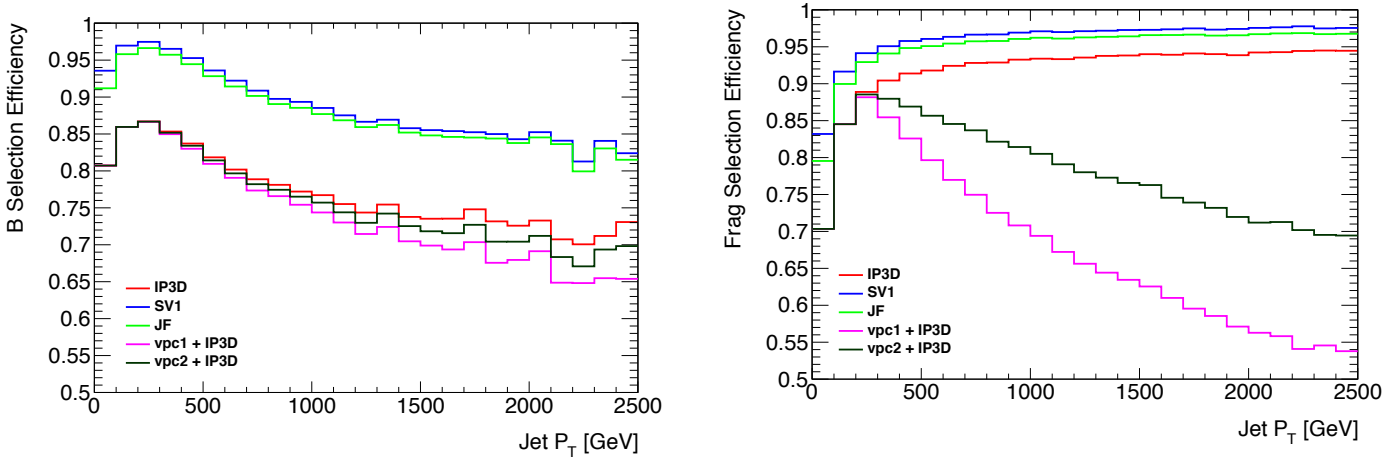
- We have studied the distribution of track-P_T against increasing jet-P_T.

- Identified that tracks from B are more dependant than tracks from other origins.

- Implemented two jet-P_T dependant cuts on track-P_T.

- These cuts can be seen on the left by the magenta and dark green lines.

- These cuts show much promise



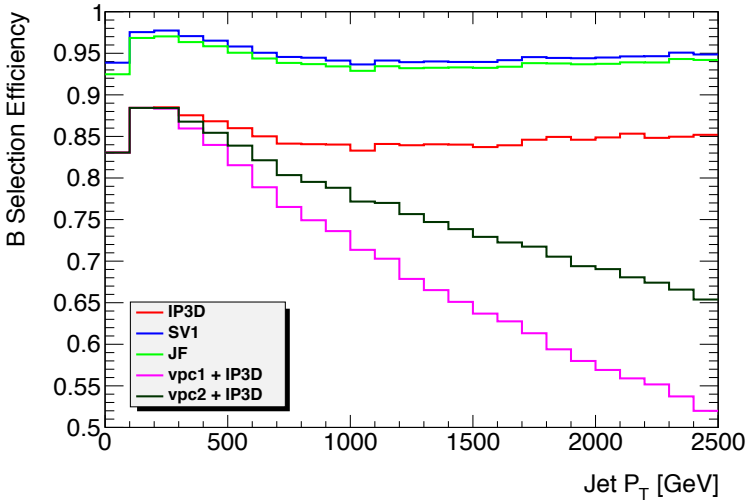
- Little effect on track selection efficiency of tracks from B.

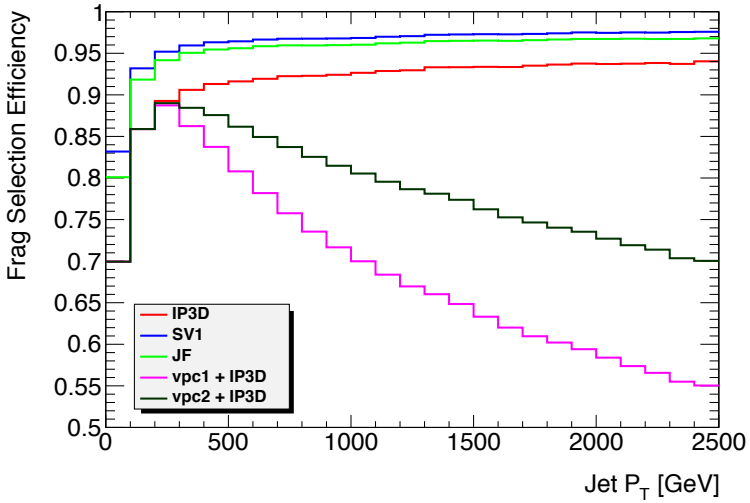
- Larger effect on tracks from other sources (particularly fragmentation) which should reduce fake rates.

See Full Talk: https://indico.cern.ch/event/393645/contribution/13/material/slides/0.pdf

Further Optimisation Studies

QCD Sample





- Applying the same cuts to a QCD sample shows larger cut on tracks from B.
- This is consistent with the removal of some B-tracks that correspond to gluon splitting
- Reduces QCD background for exotic resonances.

Future Aims

• Examine other track selection cuts, such as d0 and z0, to see if there are any optimisations that can be done for high-P_T.
- Study is underway and first results will be very soon.

• Produce ROC curves to demonstrate how the jet-P_T dependant cuts effect b-Tagging performance.
- This should be done soon.

• Next; investigate, understand and optimise the b-tagging algorithms themselves.
• Integrate b-Tagging findings into di-jet framework.