



Flavour Tagging Commissioning with Data

Andreas Korn & Laurie McClymont,
Flavour Tagging Group

Joint CTIDE and Flav. Tag Meet
16/09/15



Aims

- Pub note for Data Commissioning for Flavour Tagging in Run2 Data
- Comparing data to MC in dijet events to test our understanding of flavour tagging.
- A note has been produced commissioning with top events.
- This talk will introduce key notable discrepancies, with a particular interest in tracking

Progress on Note

- First set of plots produced, framework in place.
- Note is written: <https://cds.cern.ch/record/2032461>
- First reading has occurred and first set of comments have now been addressed.
- Delaying for better understanding of distributions.



3 Samples

- MC Sample:

- **Full xAOD**

- 50ns dijet MC sample data
 - Split into 4 slices and the re-weighted (see backup) JZ1W-JZ4W - No JZ0W used.
 - ~ 8M Events.

`"mc15_13TeV.361021.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ1W.merge.AOD.e3569_s2576_s2132_r6630_r6264/"
"mc15_13TeV.361022.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2W.merge.AOD.e3668_s2576_s2132_r6630_r6264/"
"mc15_13TeV.361023.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3W.merge.AOD.e3668_s2576_s2132_r6630_r6264/"
"mc15_13TeV.361024.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4W.merge.AOD.e3668_s2576_s2132_r6630_r6264/"`

- Data Sample:

- 50ns data from stable beam collisions.
 - **FTAG** derivation
 - ~4M Events from 6 Runs: 270806, 270953, 271048, 271421, 271516 and 271595

`"data15_13TeV.00270806.physics_Main.merge.DAOD_FTAK1.f611_m1463_p2375/"
"data15_13TeV.00270953.physics_Main.merge.DAOD_FTAK1.f611_m1463_p2375/"
"data15_13TeV.00271048.physics_Main.merge.DAOD_FTAK1.f611_m1463_p2375/"
"data15_13TeV.00271421.physics_Main.merge.DAOD_FTAK1.f611_m1463_p2375/"
"data15_13TeV.00271516.physics_Main.merge.DAOD_FTAK1.f611_m1463_p2375/"
"data15_13TeV.00271595.physics_Main.merge.DAOD_FTAK1.f611_m1463_p2375/"`

- We are using NTuples created using Run2BTagOptimisationFramework



4 Details and Cuts

- 20.1.5.3 with all tags recommended by CP group
- Running xAOD fix on full xAOD

• **HLT_j60 Trigger for MC and Data with Leading Jet $P_T > 70$ GeV.**

- AntiKt4EMTopoJets
- Run1LooseBadCuts and “ugly” jet removal.
- Jet Calibration:

```
- calibfile  ="JES_MC15Prerecommendation_April2015.config"
- calSeg      ="JetArea_Residual_Origin_EtaJES_GSC" (_Insitu for data)
```

• GRL = *“data15_13TeV.periodAllYear_DetStatus-v63-pro18-01_DQDefects-00-01-02_PHYS_StandardGRL_All_Good.xml”*

Select event if:

- $n_{\text{jets}} \geq 2$ with $P_T > 35$ GeV, $|\eta| < 2.4$ and $\text{JVT} > 0.641$ if $P_T < 50$ GeV
- Leading jet, $P_T > 70$ GeV, $|\eta| < 2.4$ and $\text{JVT} > 0.641$ if $P_T < 50$ GeV

Then plot subleading if subleading jet has:

- $P_T > 35$ GeV
- $|\eta| < 2.5$
- $\text{JVT} > 0.641$ if $P_T < 50$ GeV

Monte Carlo Cuts

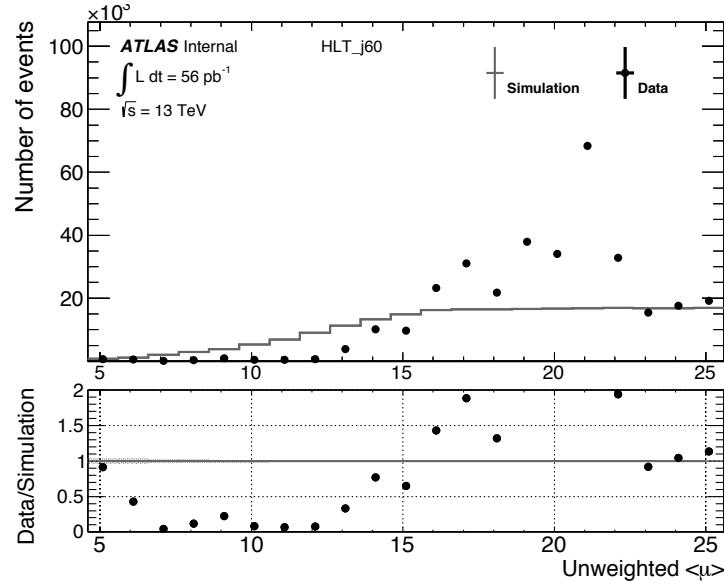
Truth Dijet Test applied to MC to clean sample
- $(\text{Lead } P_T + \text{Sublead } P_T)/2 < 1.4 * \text{Truth Lead } P_T$

- LabDr_HadF truth matching.

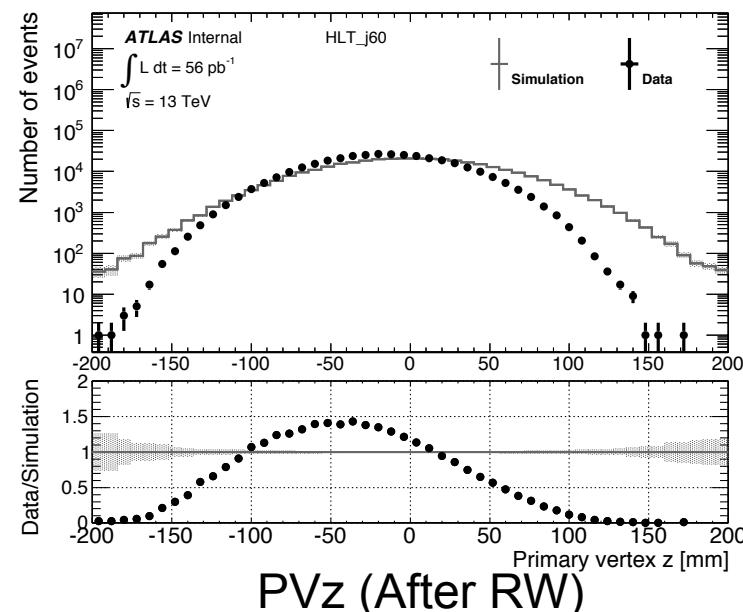
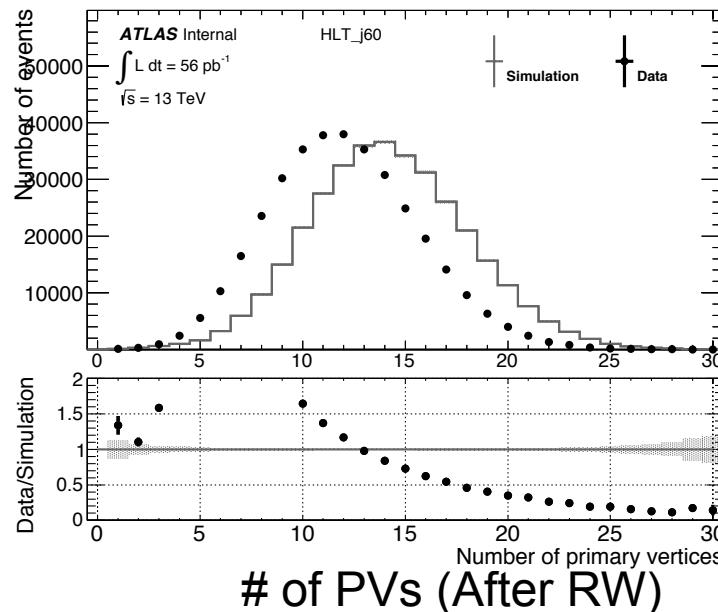
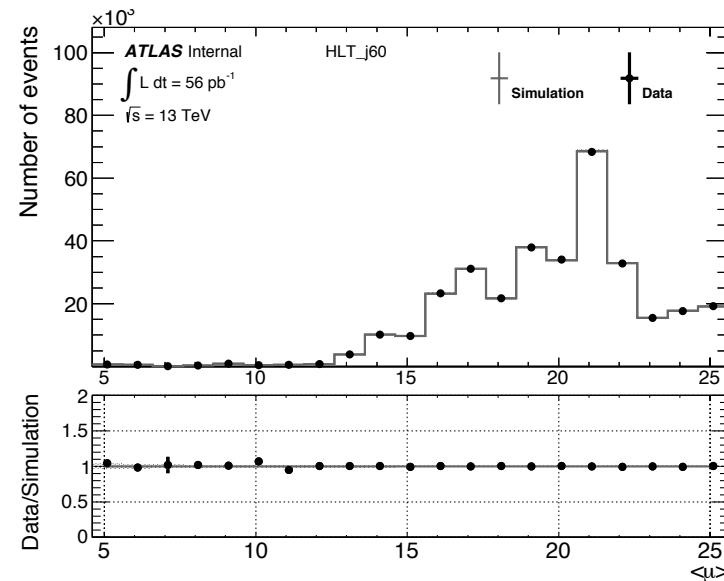


5 $\langle\mu\rangle$ Reweighting

Before:



After:

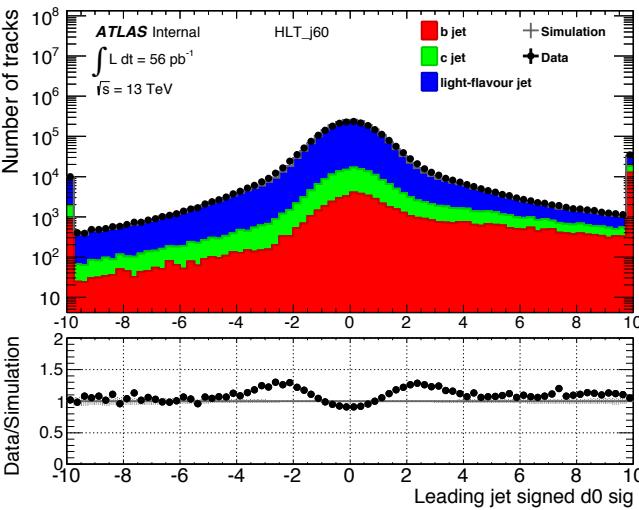




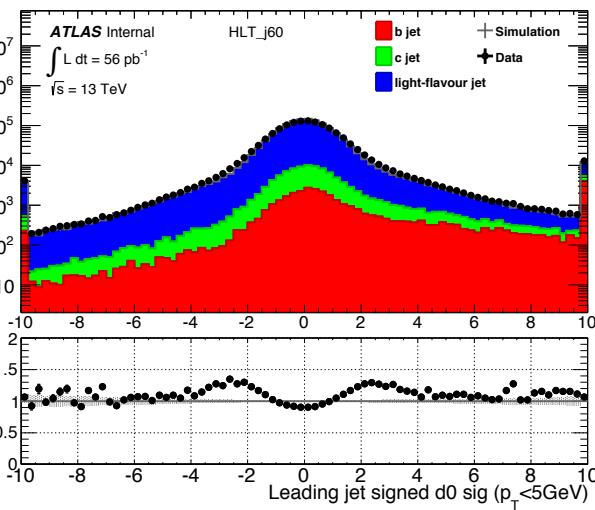
6 Signed d0/z0 Significance - Varying Track pT



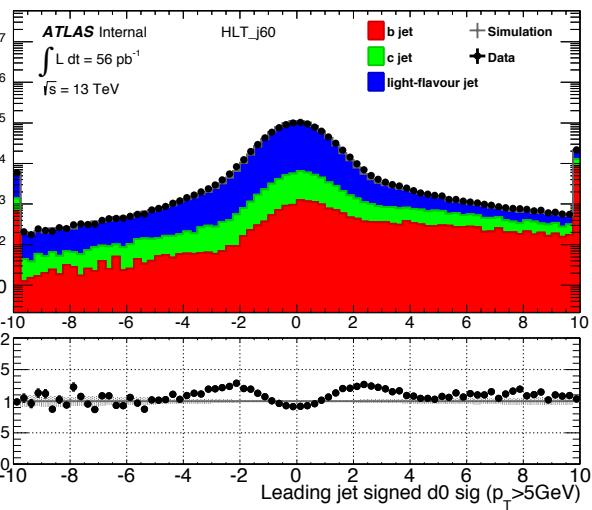
d0 Significance:



All ($p_T > 1 GeV$)

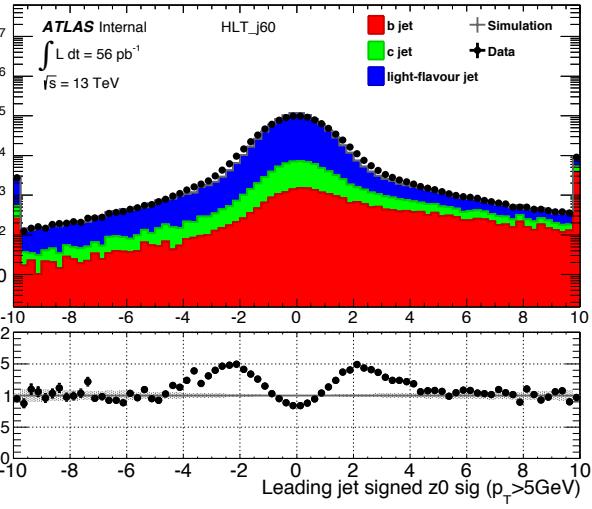
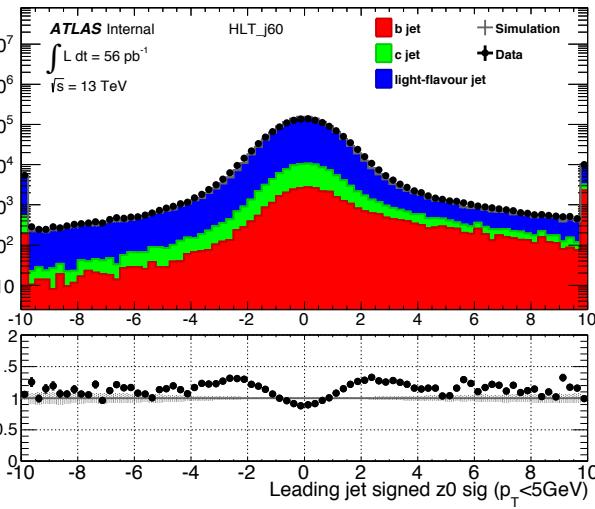
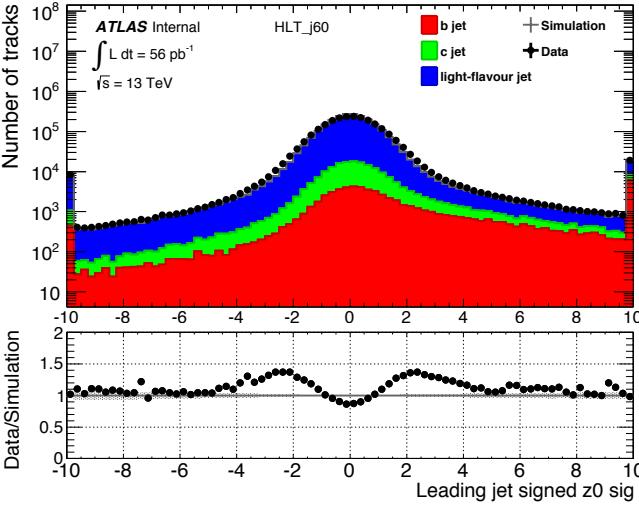


1 GeV < track pT < 5 GeV



track pT > 5 GeV

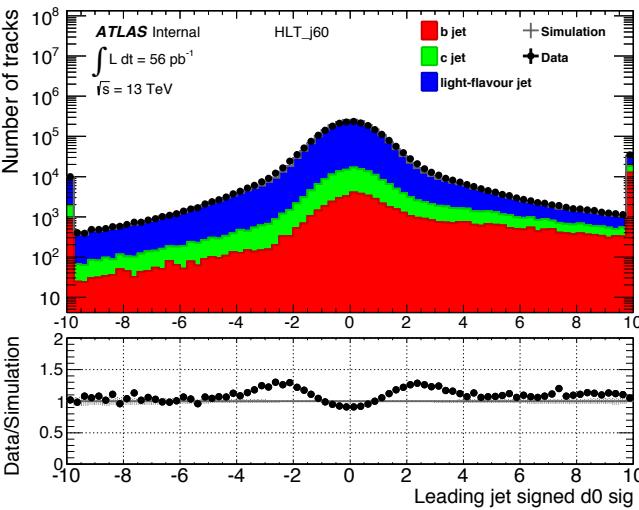
z0 Significance:



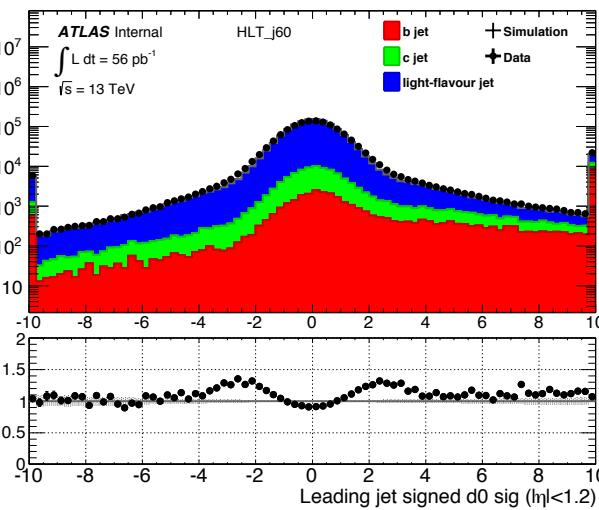


7 Signed d0/z0 Significance - Varying n

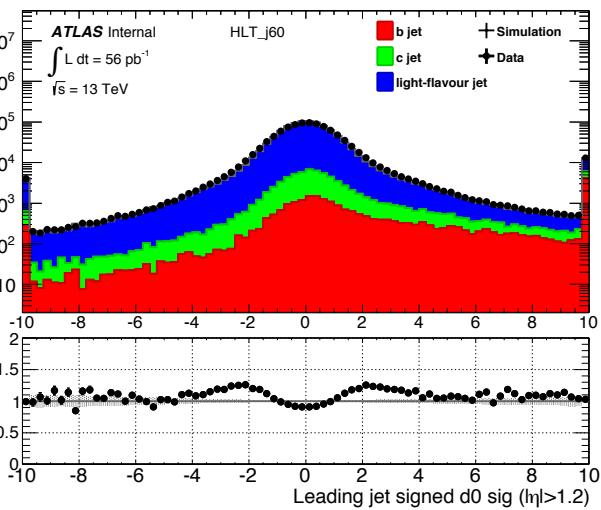
d0 Significance:



All

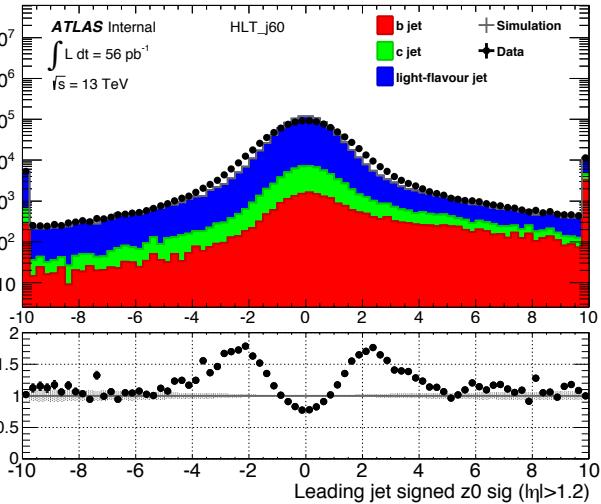
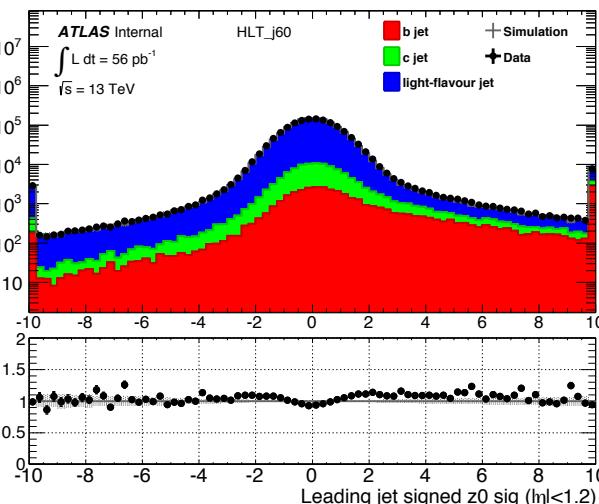
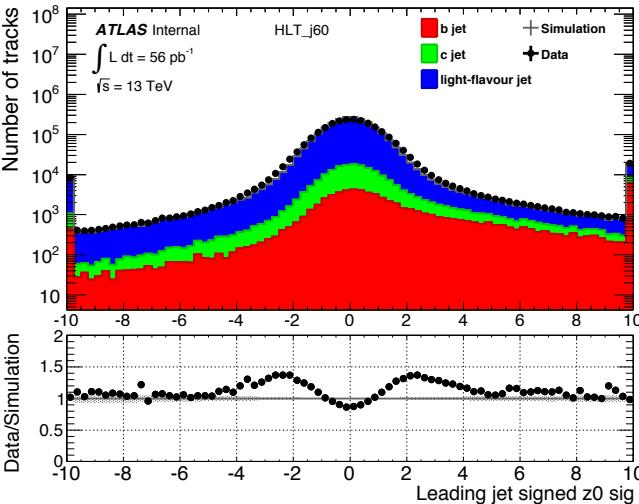


$|\eta| < 1.2$



$|\eta| > 1.2$

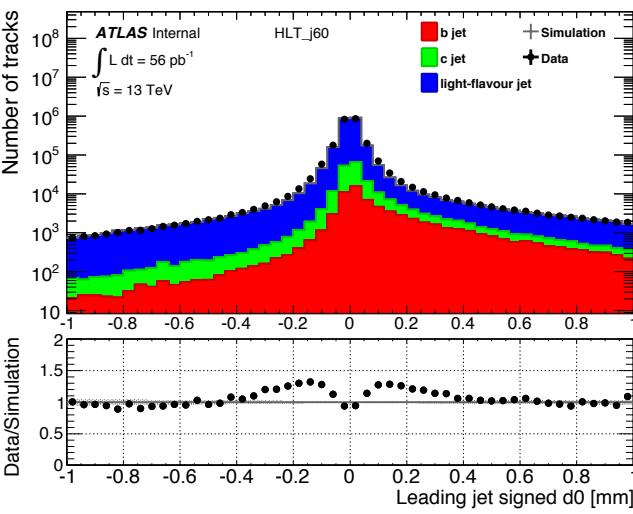
z0 Significance:



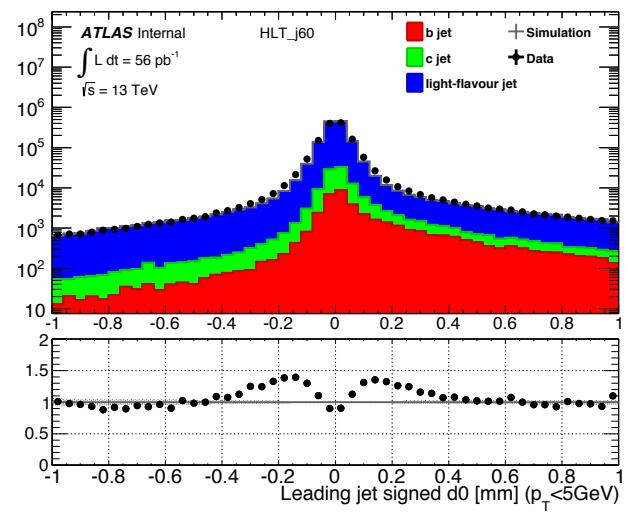


8 Signed d0/z0 - Varying Track pT

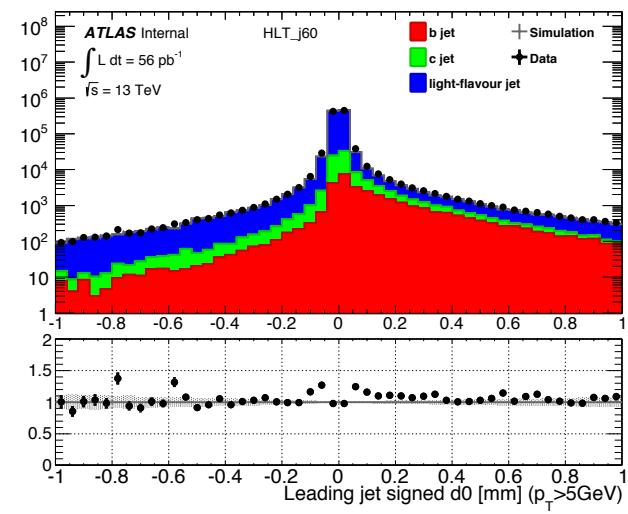
d0:



All ($p_T > 1 \text{ GeV}$)

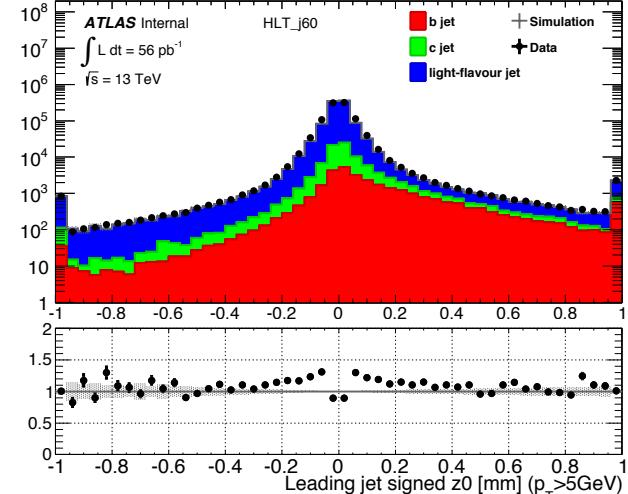
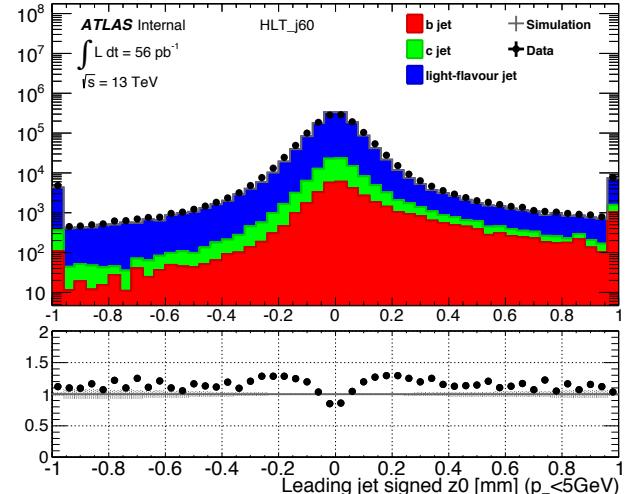
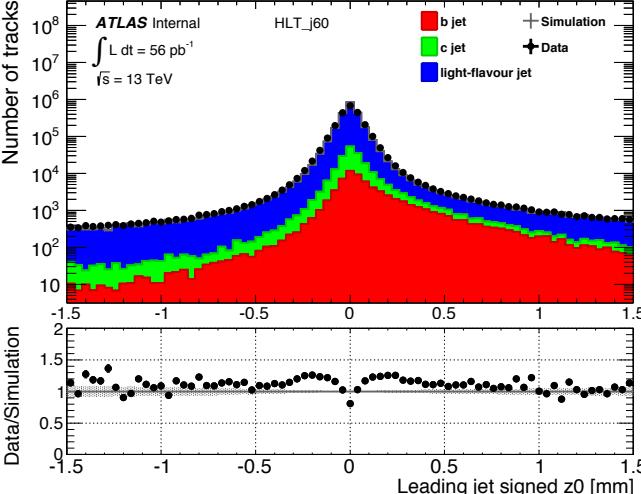


1 GeV < track pT < 5 GeV



track pT > 5 GeV

z0:

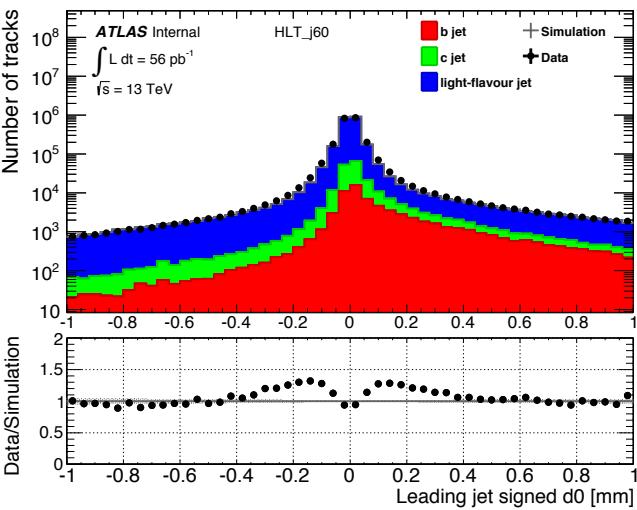




9 Signed d0/z0 - Varying η



d0:

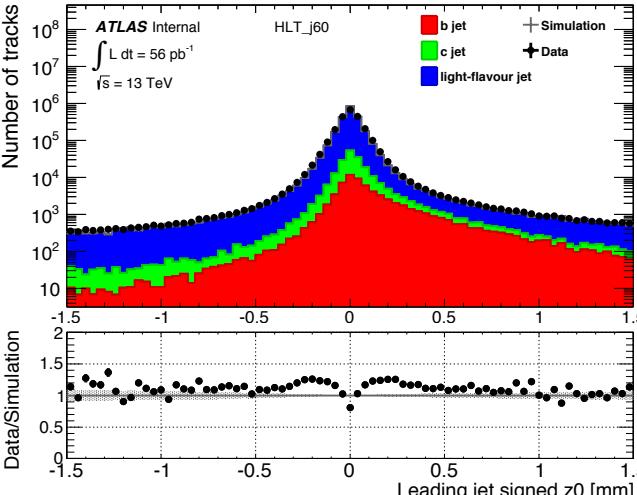


All

$|\eta| < 1.2$

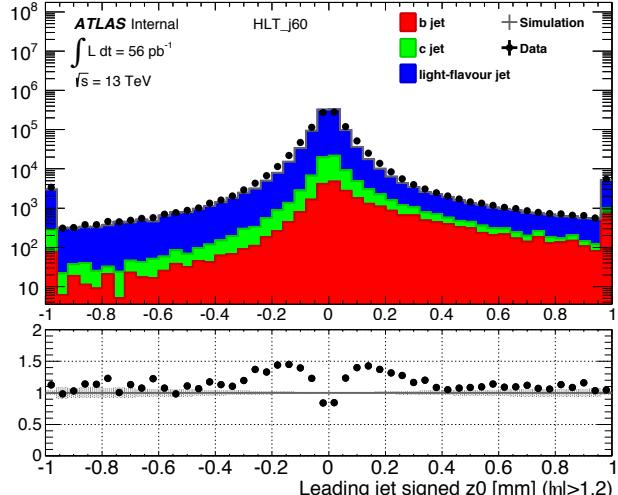
$|\eta| > 1.2$

z_0 :



$|\eta| < 1.2$

$|\eta| > 1.2$

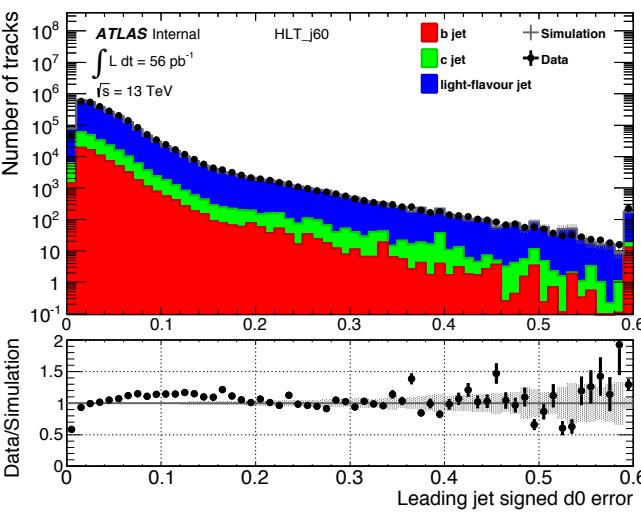




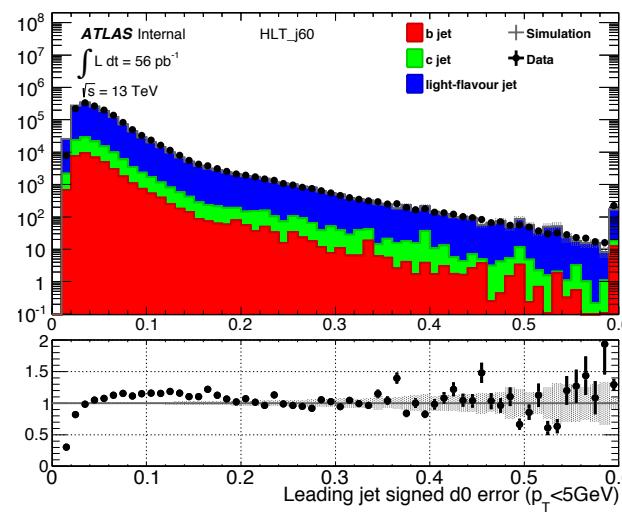
10 Signed d0/z0 Error - Varying Track pT



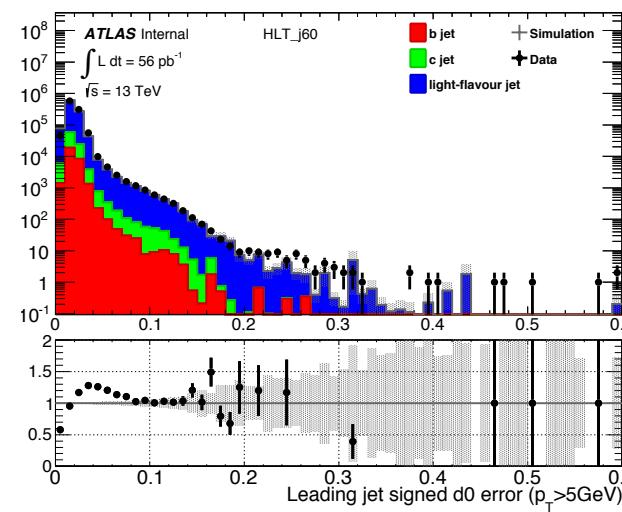
d0 Error:



All ($p_T > 1$ GeV)

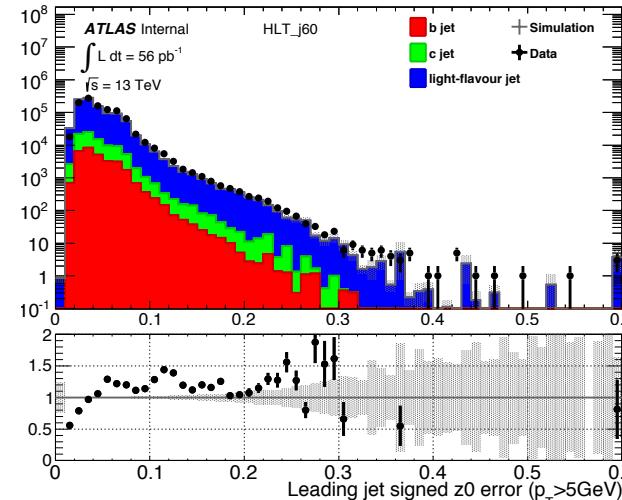
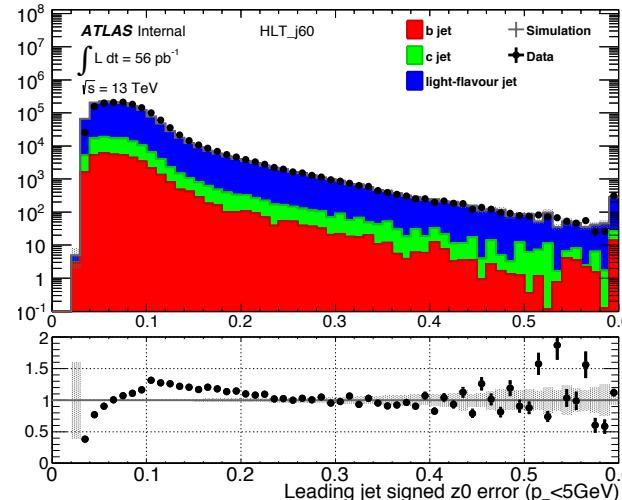
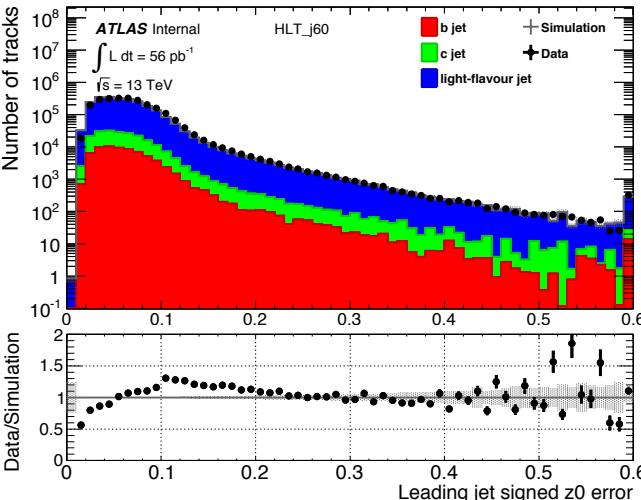


1 GeV < track pT < 5 GeV



track pT > 5 GeV

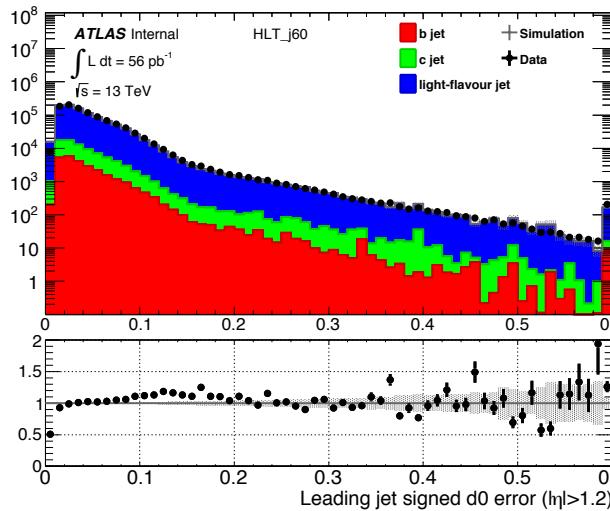
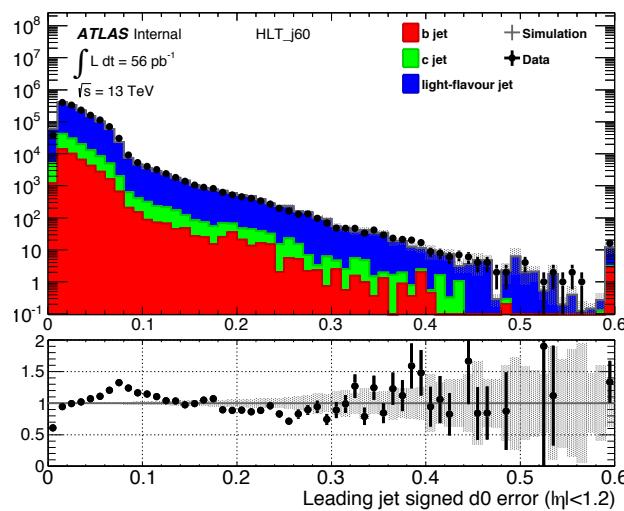
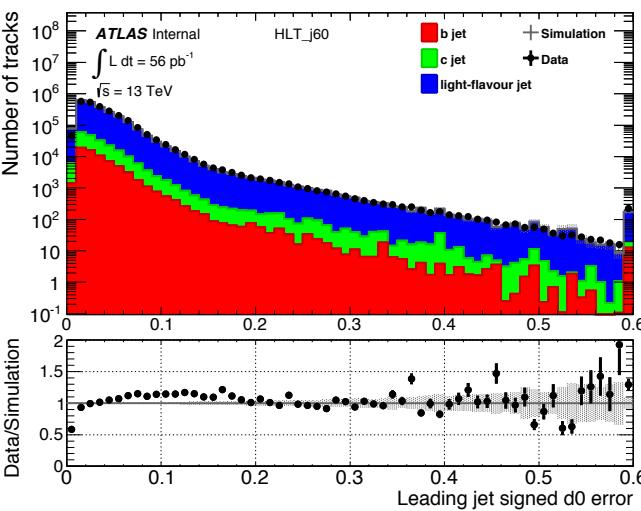
z0 Error:





11 Signed d0/z0 Error - Varying η

d0 Error:

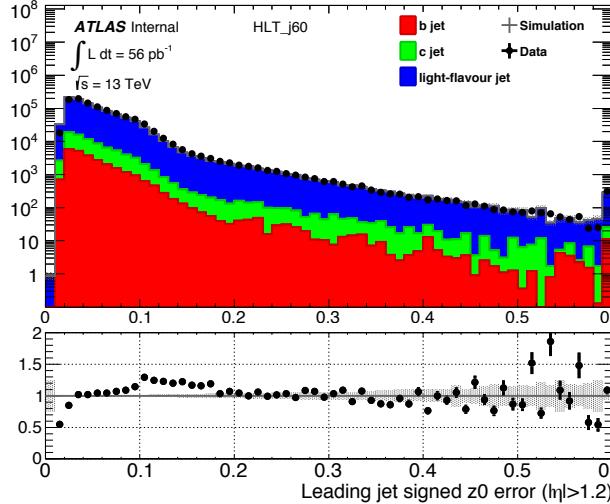
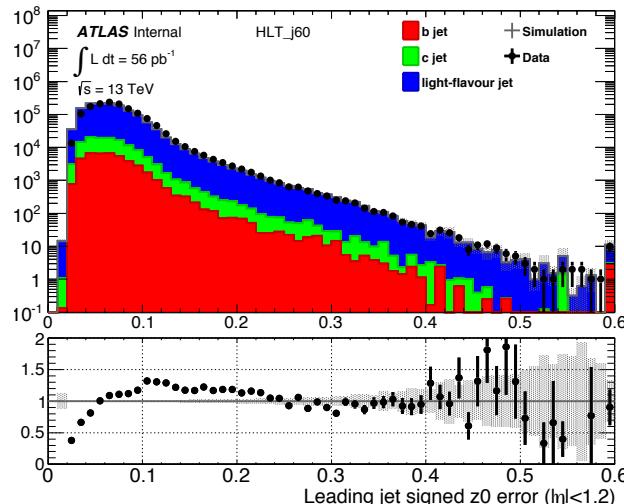
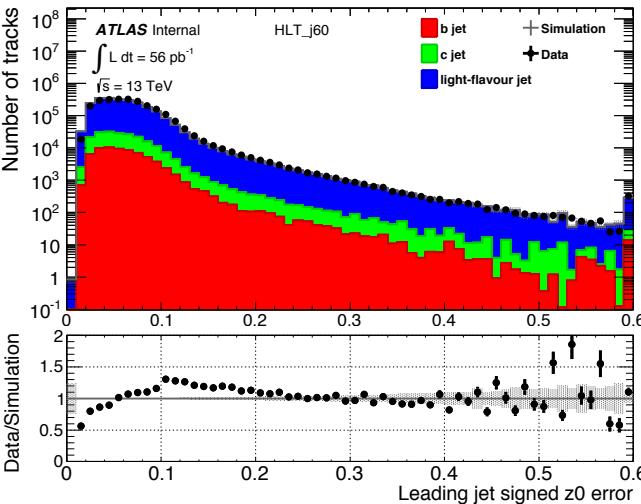


All

$|\eta| < 1.2$

$|\eta| > 1.2$

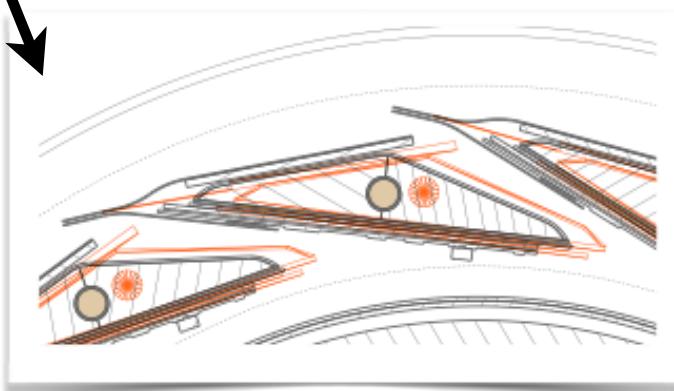
z0 Error:





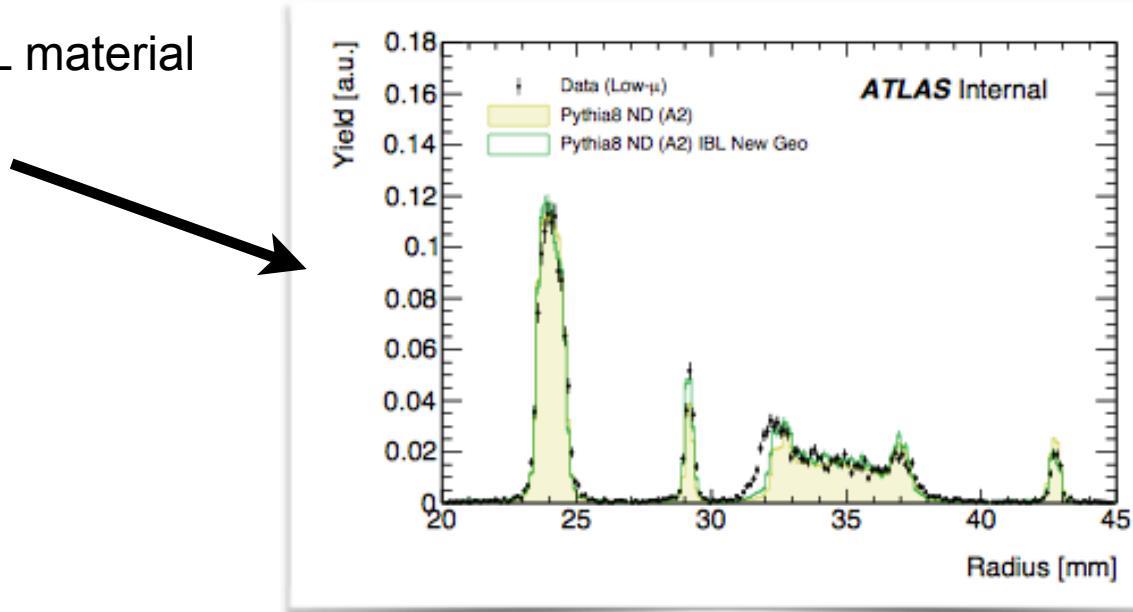
12 IP Update from Tracking Group

- Geometry tag missing 23% IBL material
 - New Geo. Tag Produced
 - In validation.
- Angle of overlap issue for IBL
 - Unlikely to be a large effect.



GEO Model
IBL

Simone Pagan Griso, Heather Gray
[https://indico.cern.ch/event/433839/
contribution/6/attachments/
1128840/1612854/PC_20jul.pdf](https://indico.cern.ch/event/433839/contribution/6/attachments/1128840/1612854/PC_20jul.pdf)

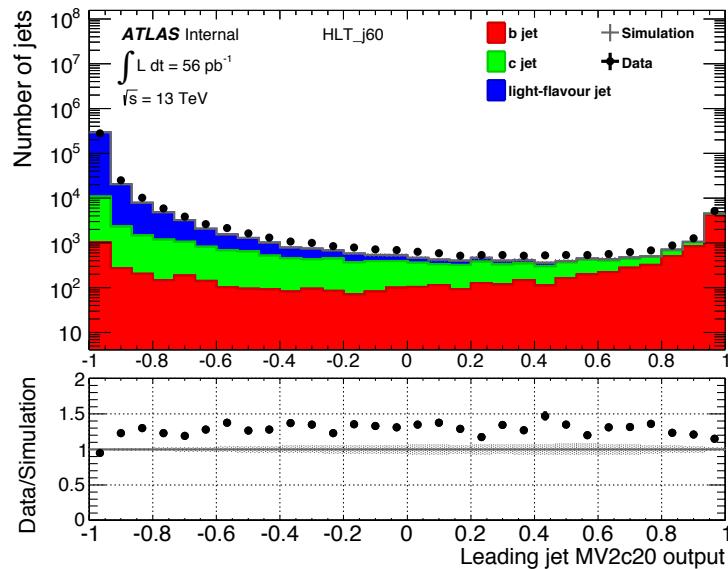


- These issues affect the error estimation and hence the d_0/z_0 significance distribution
- Also the d_0 and z_0 resolution is also effected
 - The d_0/z_0 distributions can be matched again using a smearing distribution.

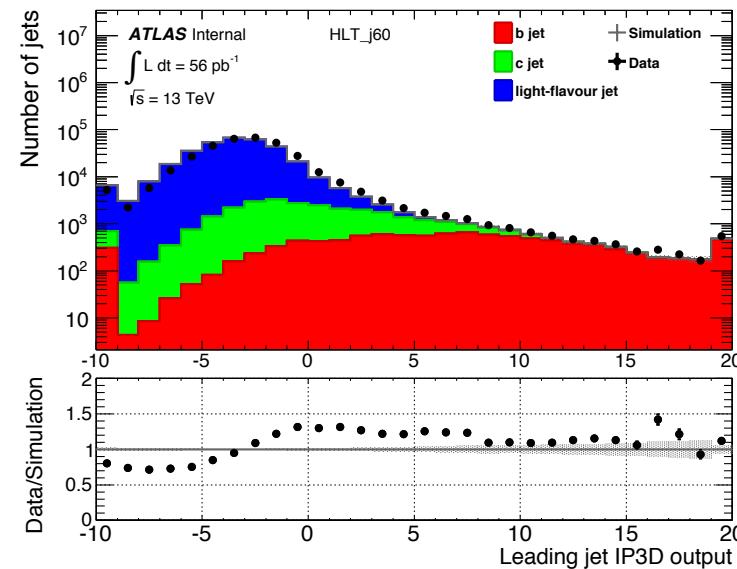


13 Taggers Leading Jet

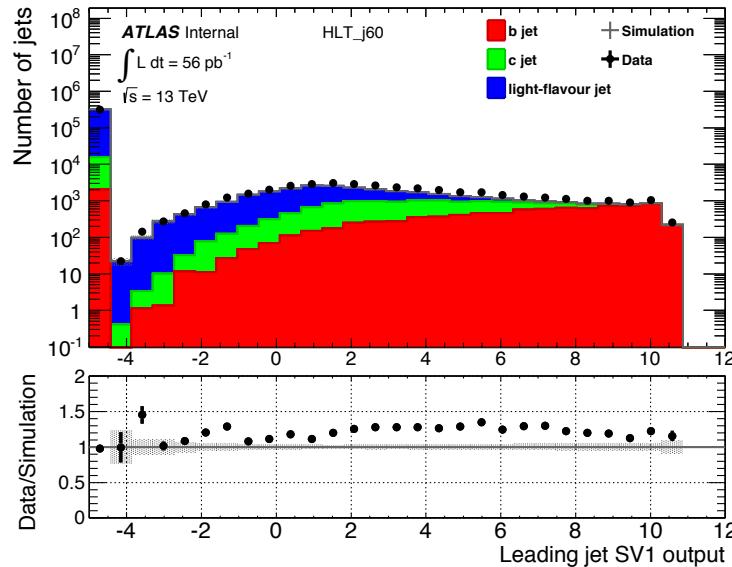
MV2c20:



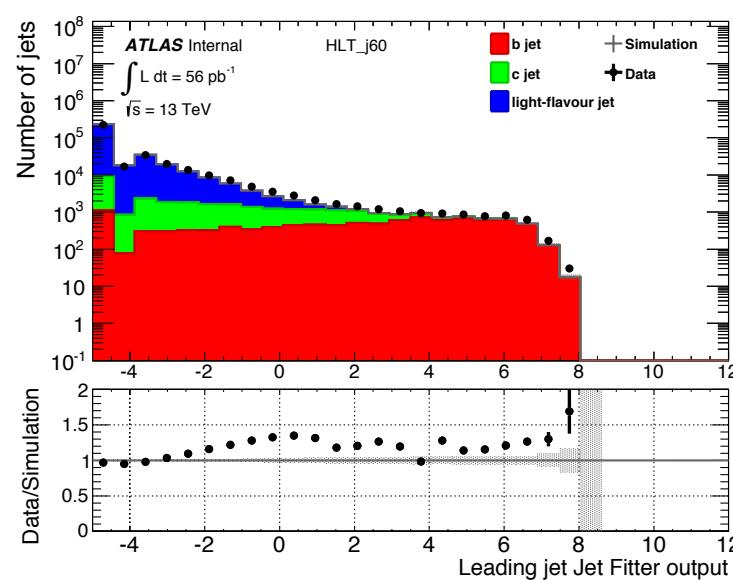
IP3D:



SV1:



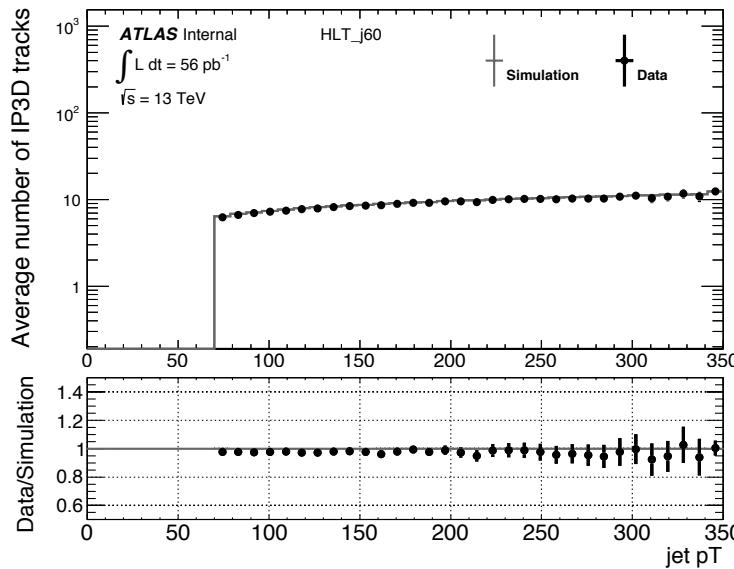
Jet Fitter:



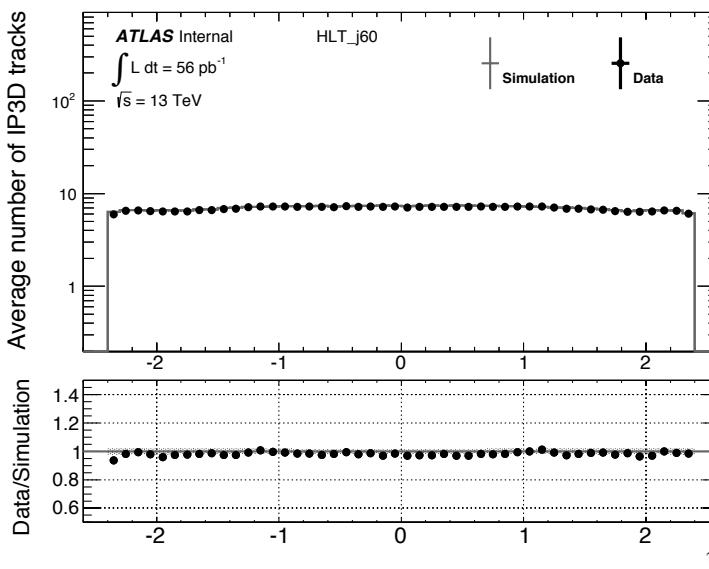


14 # IP3D Tracks / # Jets

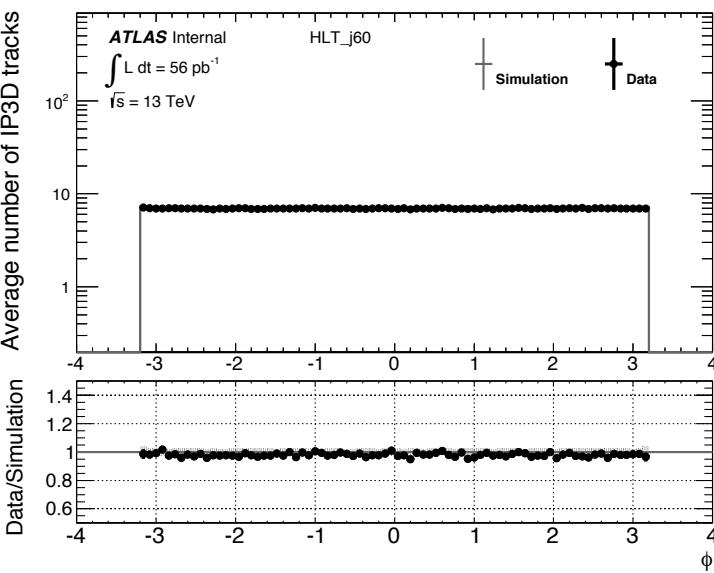
Vs. Jet pT:



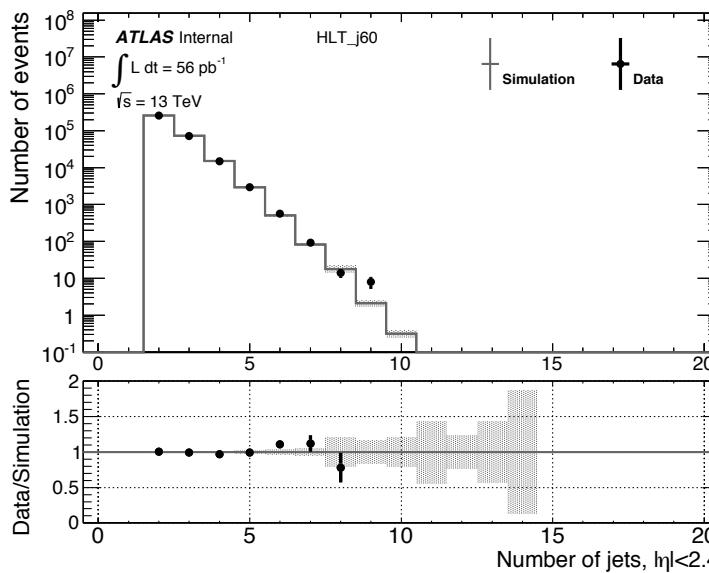
Vs. Eta:



Vs. Phi:



Jet Multiplicity:

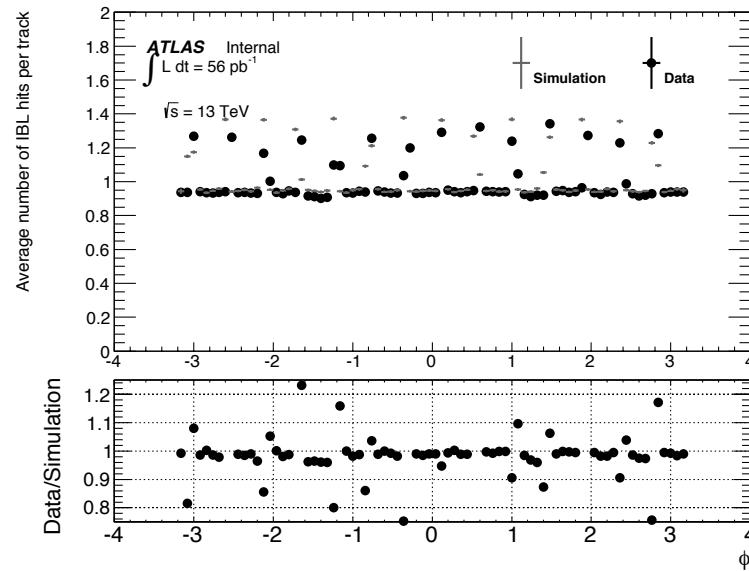




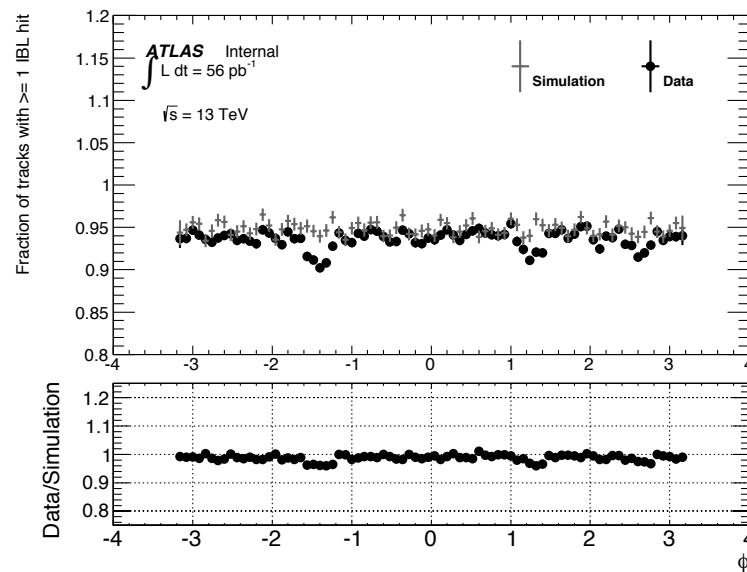
15 Track Studies - Average # IBL Track Hits



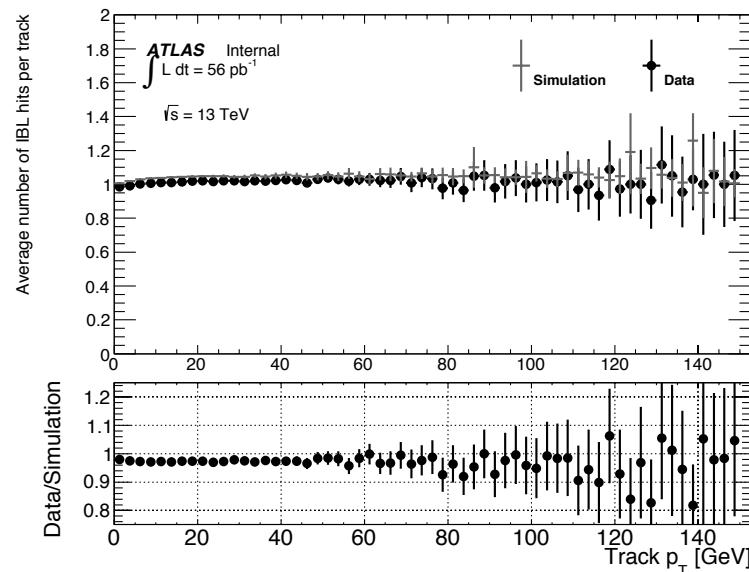
vs. Phi



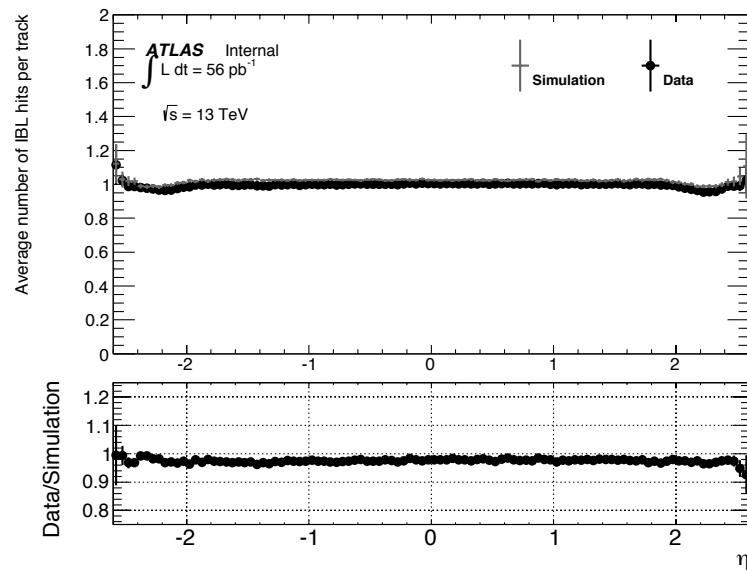
Hits ≥ 1 vs. Phi



vs. Track pT

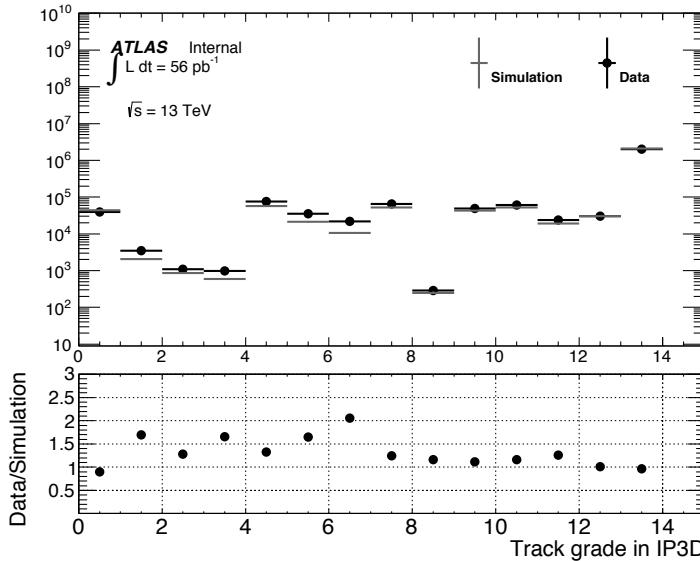


vs. Eta





16 IP3D Track Categories



IP3D Grades Tracks into 14 categories - shown below

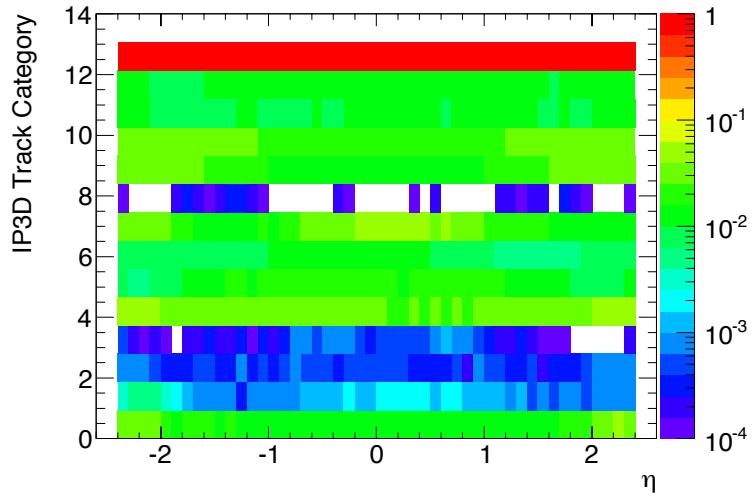
Next Slides:
Categories as a function of
 $|\eta|$ and jet- p_T

- 0 - 0In0NIn_EInENIn: no hit in first 2 layer and both expected
- 1 - 0In0NIn_EIn: no hit in first 2 layer and expected in the first
- 2 - 0In0NIn_ENIn: no hit in first 2 layer and expected in the second
- 3 - 0In0NIn: no hit in first 2 layer and not expected
- 4 - 0In1NIn_EIn: no hit in first layer and expected
- 5 - 0In1NIn: no hit in first layer and not expected
- 6 - 1In0NIn_ENIn: no hit in second layer and expected
- 7 - 1In0NIn_ENIn: no hit in second layer and expected
- 8 - First2LayerShared: shared in both first and second layer
- 9 - PixelShared: at least 1 share in first 2 layers or shared in other layers
- 10 - SCTShared: >1 shared in SCT (noPixShared)
- 11 - First2LayerSplit: at least a split hit in one of the first 2 layer
- 12 - PixelSplit: split hits in other pixel layers
- 13 - Good: split hits in other pixel layers



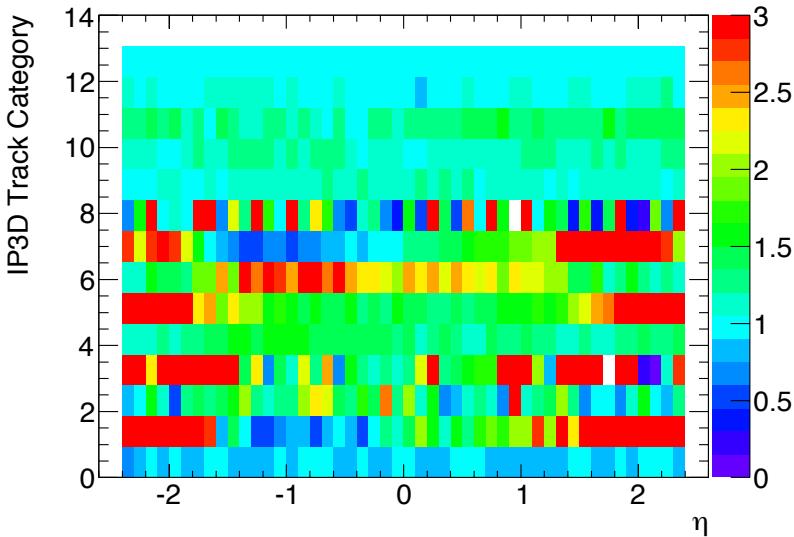
17 vs. Eta

Data

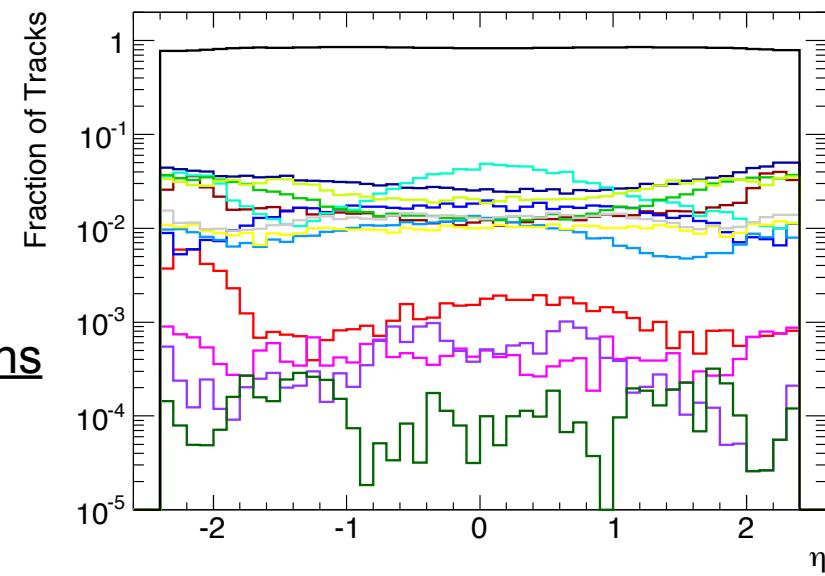


- 0 - 0In0NIn_EInENIn: no hit in first 2 layer and both expected
- 1 - 0In0NIn_EIn: no hit in first 2 layer and expected in the first
- 2 - 0In0NIn_ENIn: no hit in first 2 layer and expected in the second
- 3 - 0In0NIn: no hit in first 2 layer and not expected
- 4 - 0In1NIn_EIn: no hit in first layer and expected
- 5 - 0In1NIn: no hit in first layer and not expected
- 6 - 1In0NIn_ENIn: no hit in second layer and expected
- 7 - 1In0NIn_ENIn: no hit in second layer and expected
- 8 - First2LayerShared: shared in both first and second layer
- 9 - PixelShared: at least 1 share in first 2 layers or shared in other layers
- 10 - SCTShared: >1 shared in SCT (noPixShared)
- 11 -First2LayerSplit: at least a split hit in one of the first 2 layer
- 12 -PixelSplit: split hits in other pixel layers
- 13 -Good: split hits in other pixel layers

Data to MC Ratio



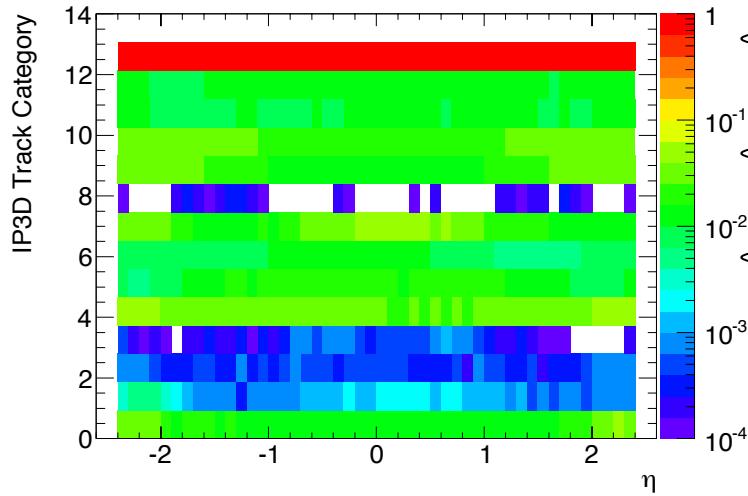
Data
Projections





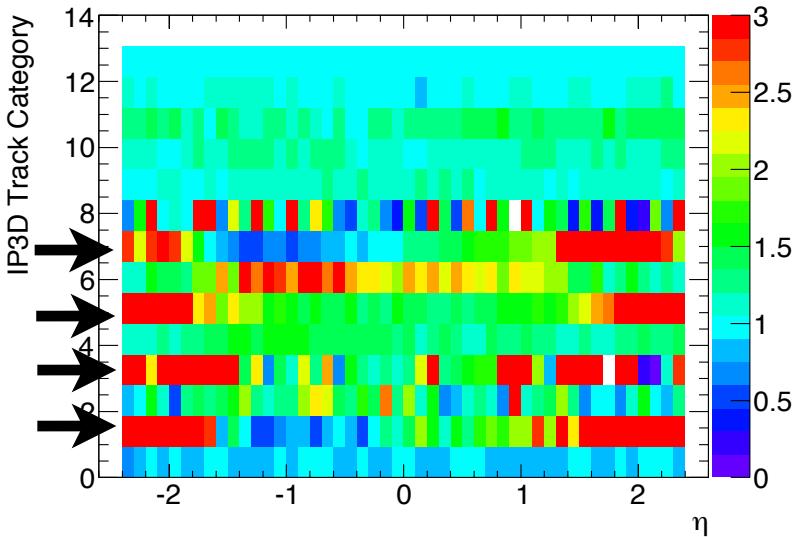
18 vs. Eta

Data

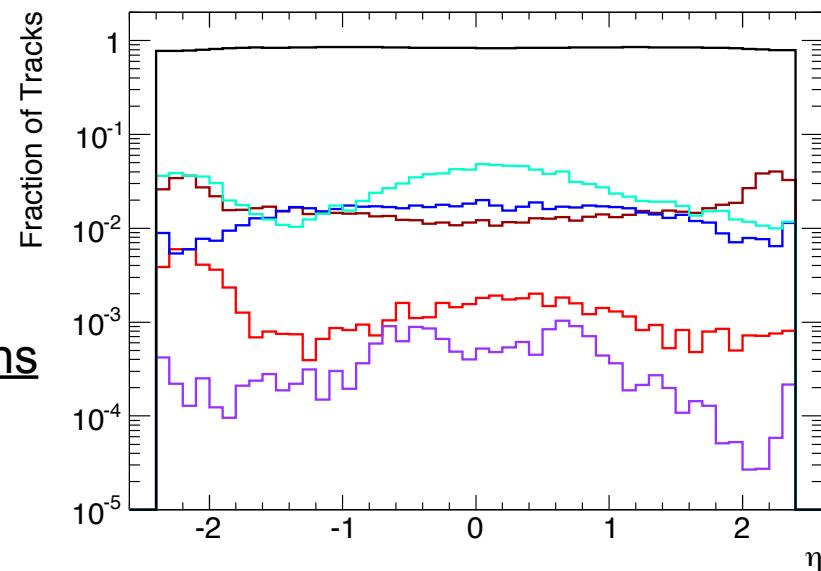


- 0 - 0In0NIn_EInENIn: no hit in first 2 layer and both expected
- 1 - 0In0NIn_EIn: no hit in first 2 layer and expected in the first
- 2 - 0In0NIn_ENIn: no hit in first 2 layer and expected in the second
- 3 - 0In0NIn: no hit in first 2 layer and not expected
- 4 - 0In1NIn_EIn: no hit in first layer and expected
- 5 - 0In1NIn: no hit in first layer and not expected
- 6 - 1In0NIn_ENIn: no hit in second layer and expected
- 7 - 1In0NIn_ENIn: no hit in second layer and expected
- 8 - First2LayerShared: shared in both first and second layer
- 9 - PixelShared: at least 1 share in first 2 layers or shared in other layers
- 10 - SCTShared: >1 shared in SCT (noPixShared)
- 11 -First2LayerSplit: at least a split hit in one of the first 2 layer
- 12 -PixelSplit: split hits in other pixel layers
- 13 -Good: split hits in other pixel layers

Data to MC Ratio



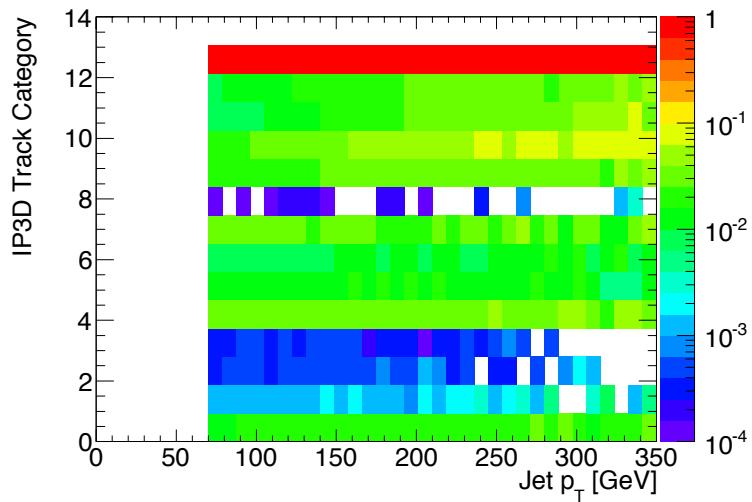
Data
Projections





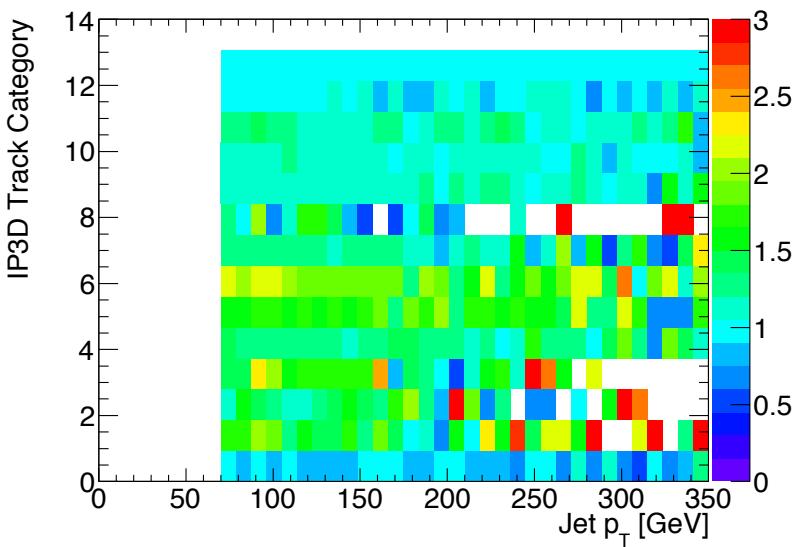
19 vs. Jet p_T

Data

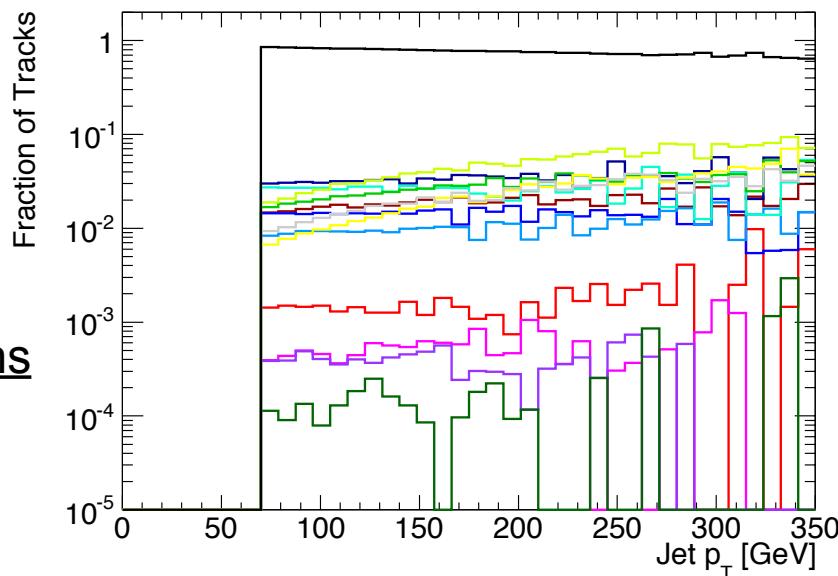


- 0 - 0In0NIn_EInENIn: no hit in first 2 layer and both expected
- 1 - 0In0NIn_EIn: no hit in first 2 layer and expected in the first
- 2 - 0In0NIn_ENIn: no hit in first 2 layer and expected in the second
- 3 - 0In0NIn: no hit in first 2 layer and not expected
- 4 - 0In1NIn_EIn: no hit in first layer and expected
- 5 - 0In1NIn: no hit in first layer and not expected
- 6 - 1In0NIn_ENIn: no hit in second layer and expected
- 7 - 1In0NIn_ENIn: no hit in second layer and expected
- 8 - First2LayerShared: shared in both first and second layer
- 9 - PixelShared: at least 1 share in first 2 layers or shared in other layers
- 10 - SCTShared: >1 shared in SCT (noPixShared)
- 11 -First2LayerSplit: at least a split hit in one of the first 2 layer
- 12 -PixelSplit: split hits in other pixel layers
- 13 -Good: split hits in other pixel layers

Data to MC Ratio



Data Projections





20 Unbiased b-Enhanced Sample

- As b-jets are most interesting to us we want to test our modelling in a sample that has an increased fraction of b-jets.
- However using b-tagging tools to increase fraction b-jets biases the sample
 - This is because we would only choose jets with “good” properties
 - i.e. A reconstructed SV.
- Hence what we do is tag the leading jet, and study the subleading jet.
 - Leading jet tagged at 70% efficiency point ($MV2c20 > -0.0436$)
 - Flavour correlation in QCD then means that the subleading jet forms an unbiased b-enhanced sample

Working Point	Leading(L) or Subleading(SL) Jet?	Number of Jets	Fraction of b-jets	Fraction of c-jets	Fraction of Light Jets
100	L	6.50E+07	0.02962 +/- 0.0007	0.06323 +/- 0.0005	0.9072 +/- 0.0001
	SL	6.00E+07	0.03198 +/- 0.0007	0.06122 +/- 0.0005	0.9068 +/- 0.0001
70	L	2.20E+06	0.6508 +/- 0.0008	0.2760 +/- 0.0013	0.0732 +/- 0.0024
	SL	2.00E+06	0.1369 +/- 0.0019	0.0687 +/- 0.0027	0.7944 +/- 0.0007

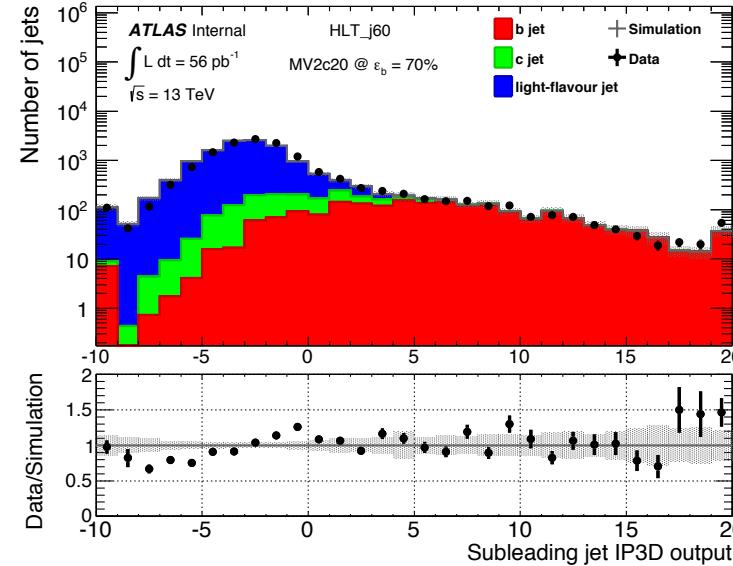
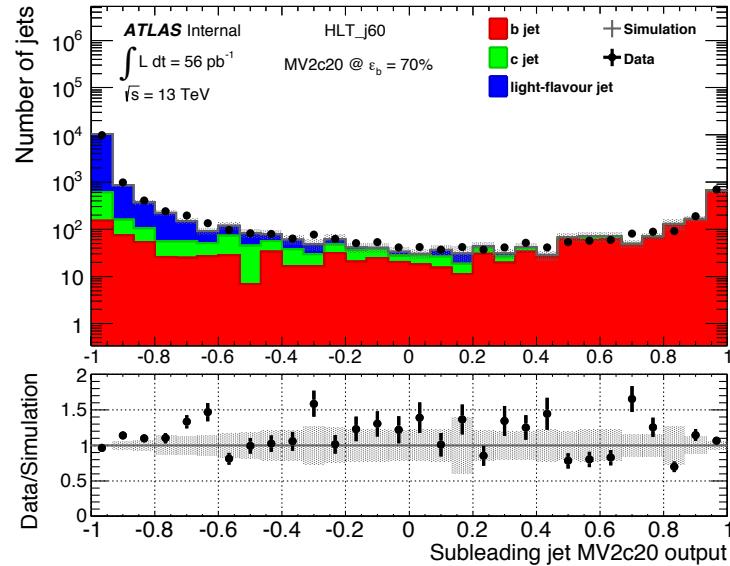


21 b Enhanced Sample - Subleading Jet

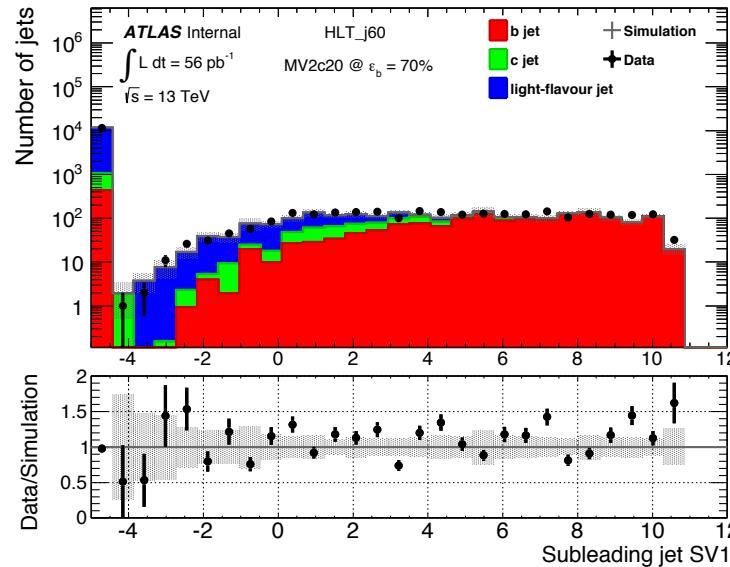
MV2c20:

Leading MV2c20 > -0.0436 which is 70% b-efficiency

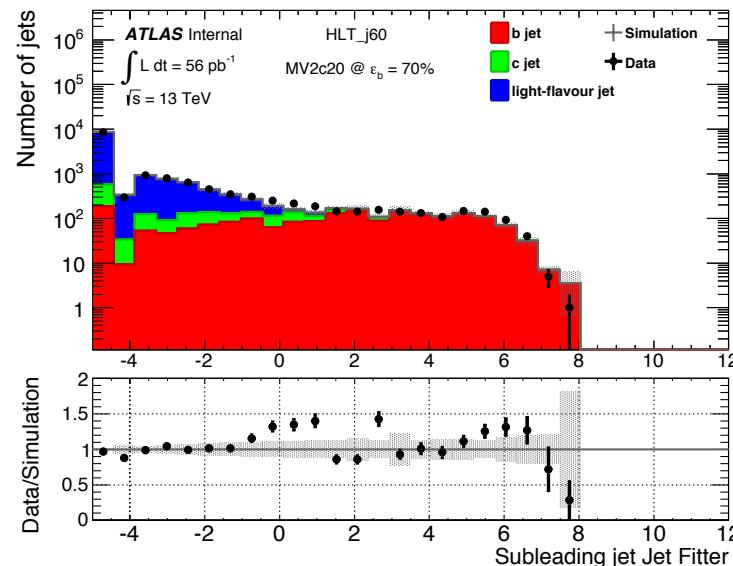
IP3D:



SV1:



Jet Fitter:



Conclusions

IP resolution discrepancies are observed in data.

- The knock on effect can be seen in the IP3D distribution

Track Hits and Number of Tracks typically well understood

IPxD categories

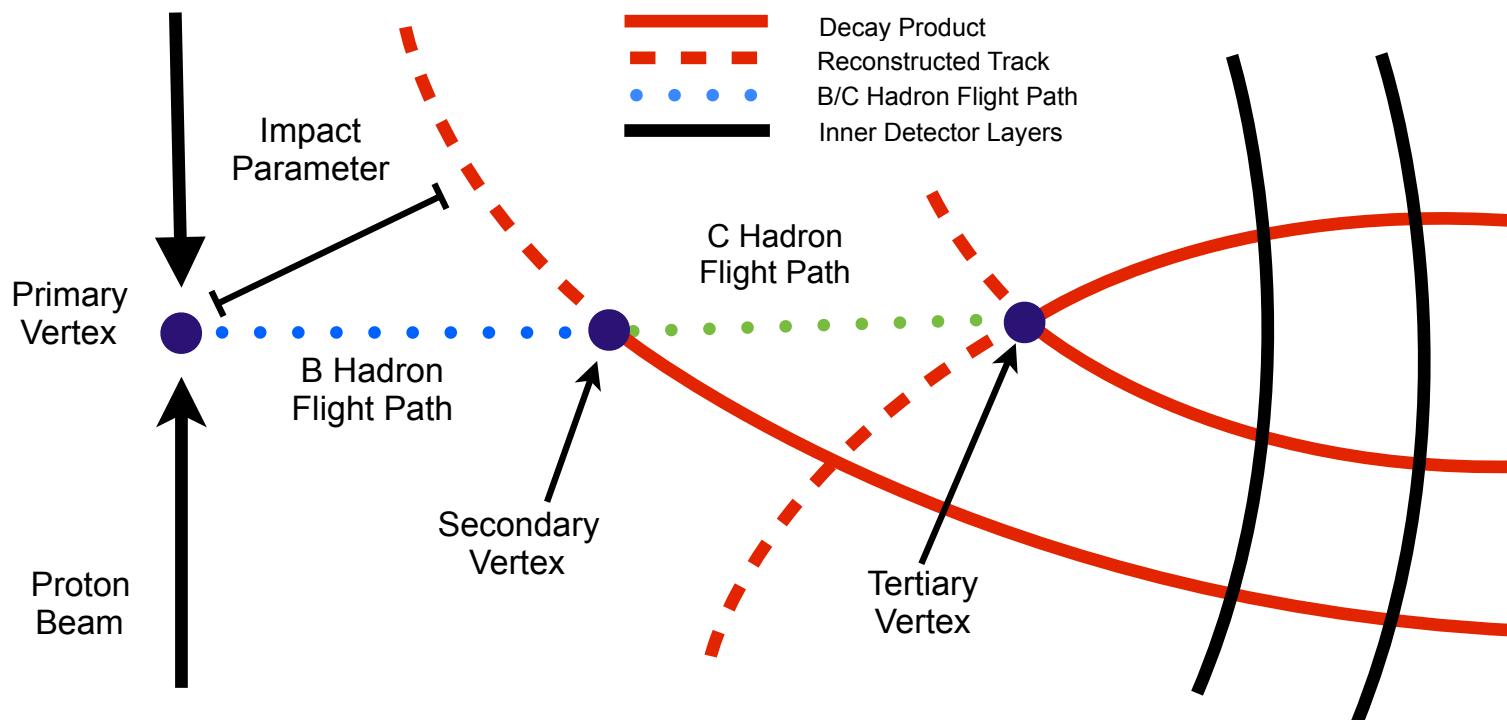
- In general agreement is reasonable
- However some categories show strong $|\eta|$ dependance.

MV2c20 not well modelled in Leading Jet sample

- Much better Modelled in b-Enhanced sample
- Why the modelling improves is not fully understood.



Backup

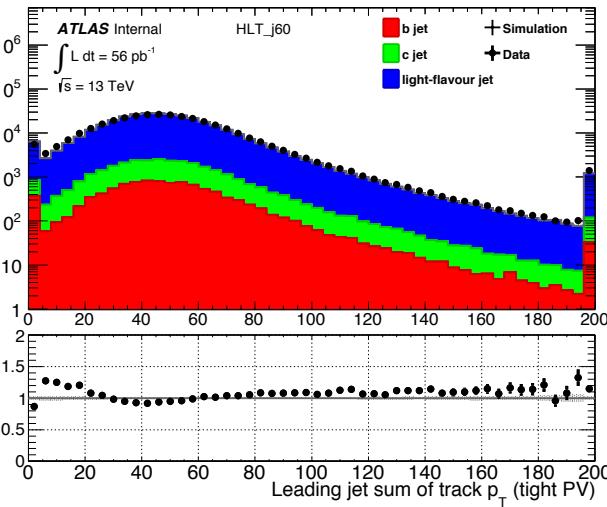
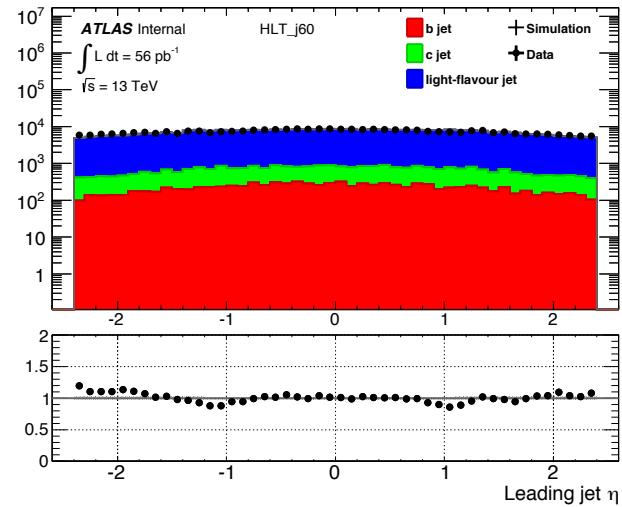
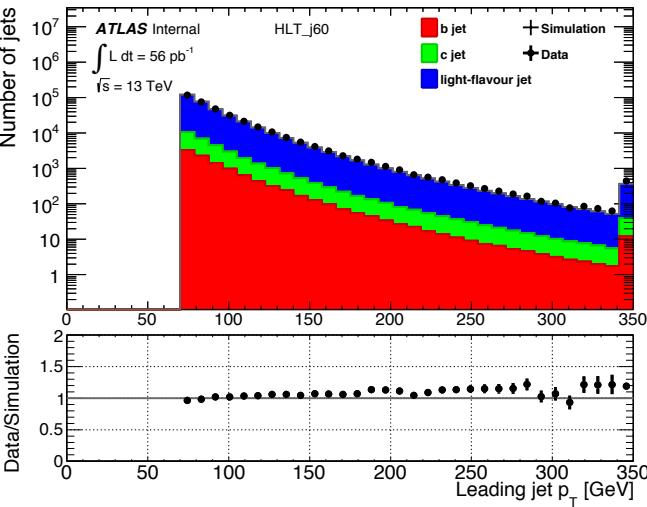


- IP2D and IP3D: Use impact parameter distributions to discriminate between flavours
- SV1: Search for Secondary Vertex from crossing of tracks
- Jet Fitter: Reconstructs full decay chain by searching for many vertices along a shared jet flight axis
- MV2c20: Base algorithms are combined in a BDT to give optimal performance.



25 Jet Kinematic Distributions

Leading Jet:

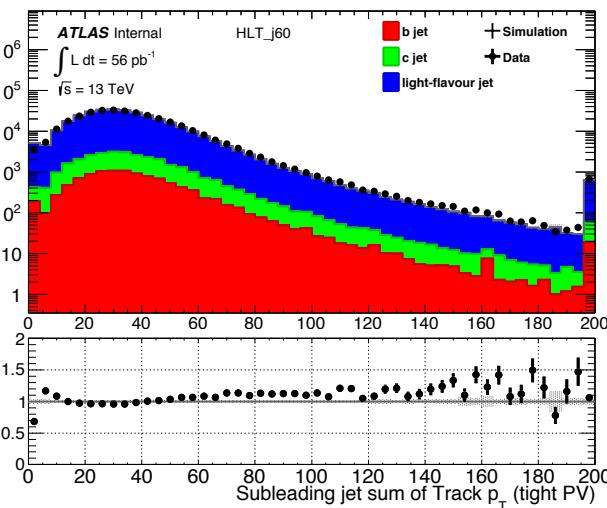
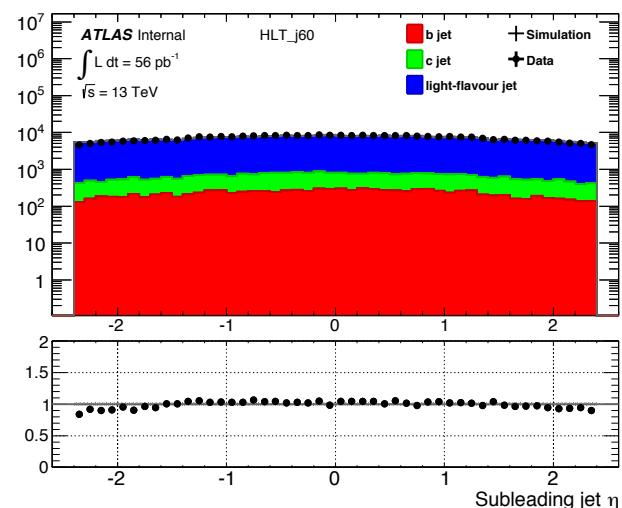
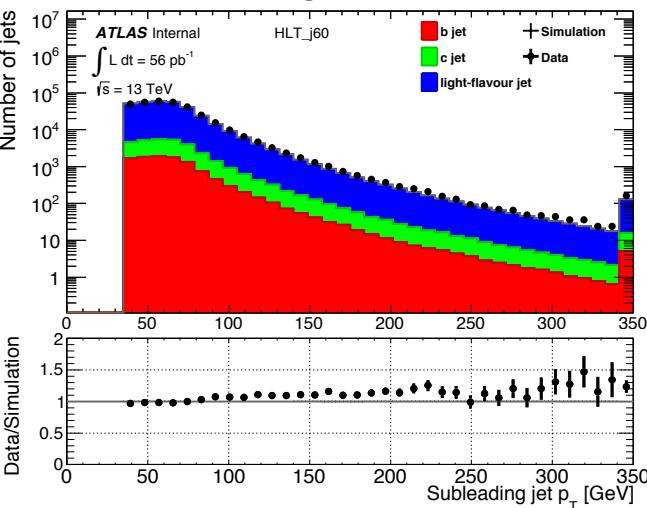


Jet P_T

Eta

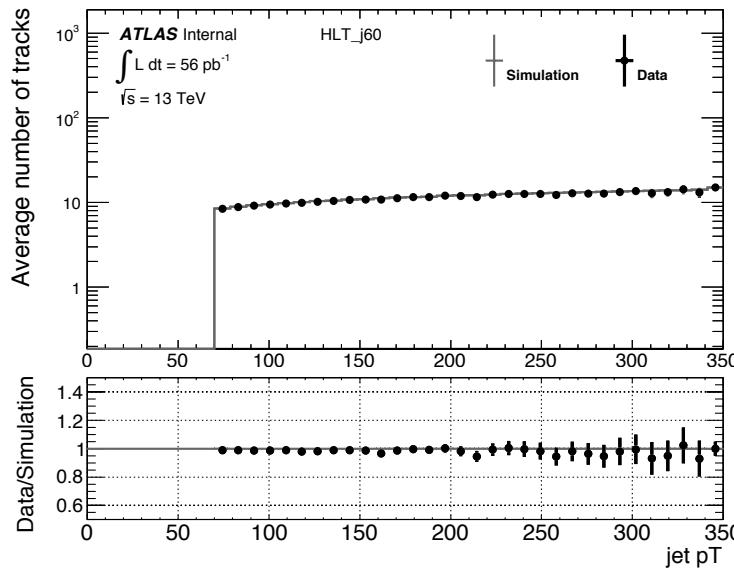
Sum of Track P_T

Sub-Leading Jet:

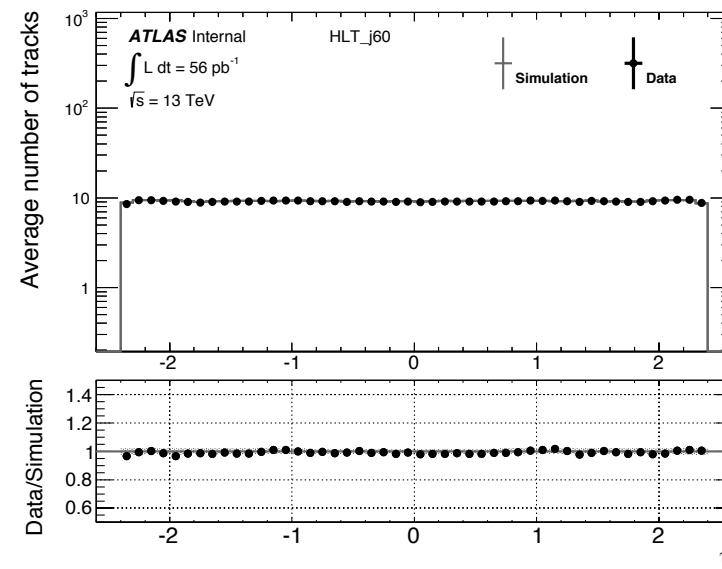




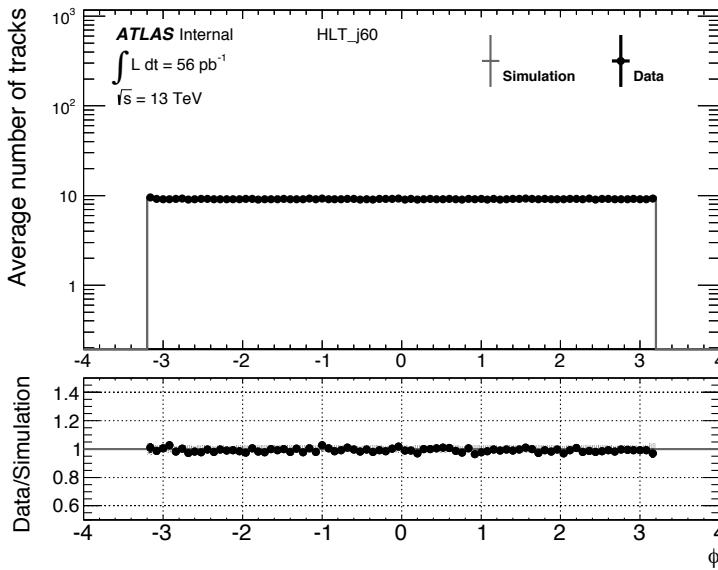
Vs. Jet pT:

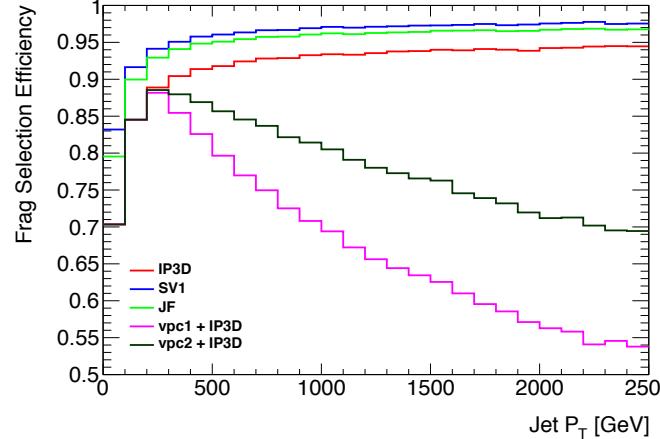
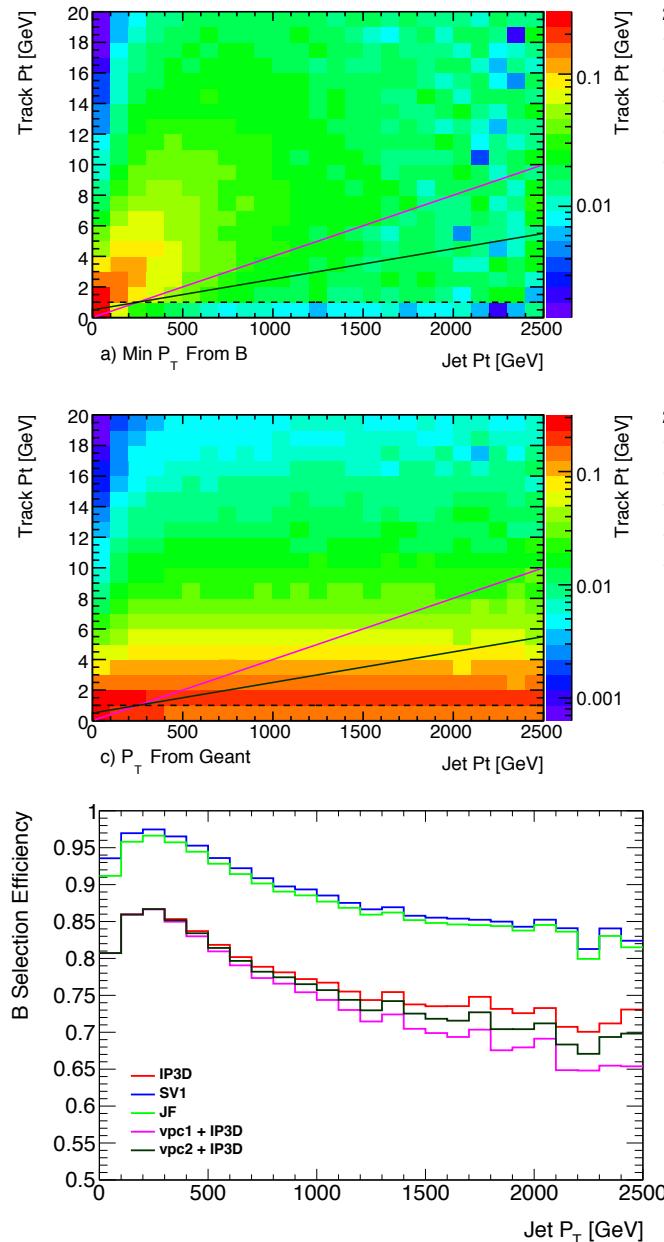
Tight PV

Vs. Eta:

Tight PV

Vs. Phi:

Tight PV



$Z' bb$ 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.