Comparing PYTHIA8 and Herwig7 MC Generators using both $t\bar{t}$ and Z' events in GN2 algorithm for B-tagging
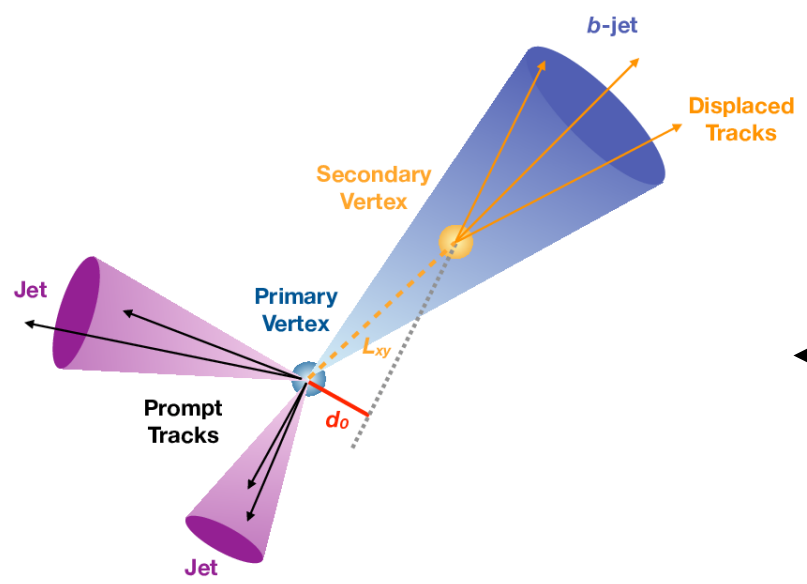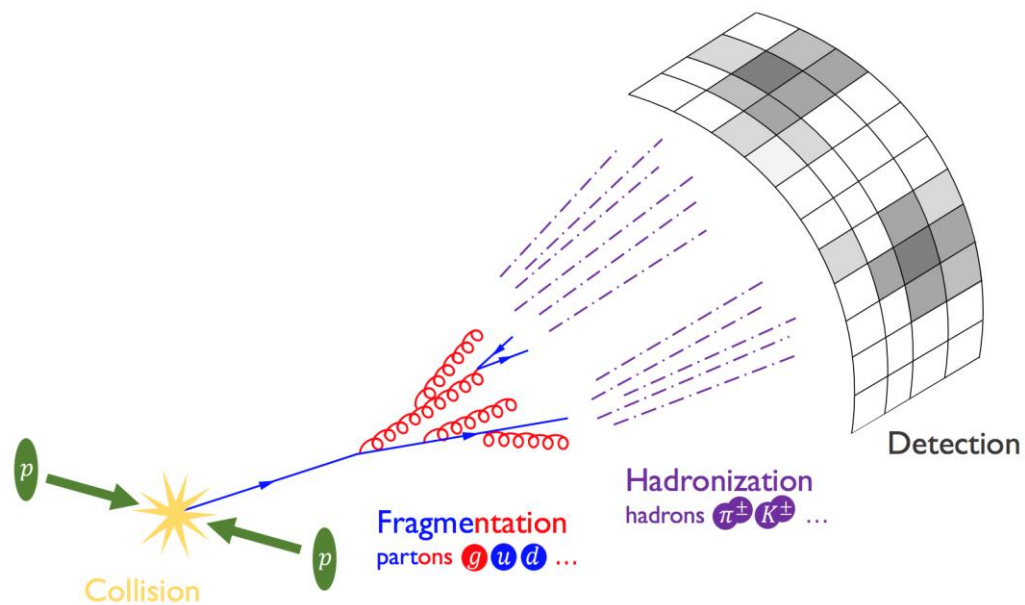
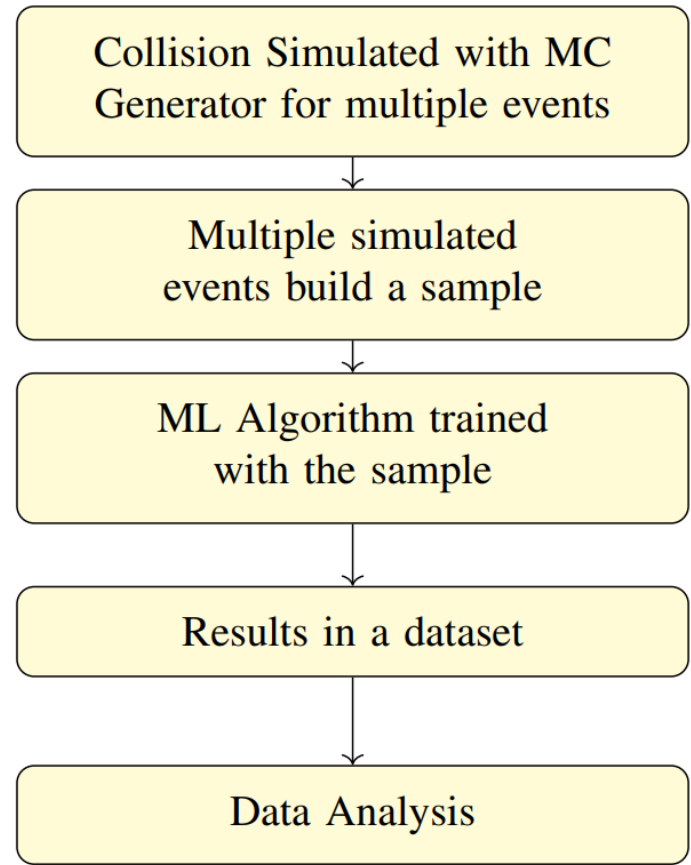Salvador Torpes

Helena Santos

Collision

p p

Fragmentation
partons g u d ...

Hadronization
hadrons $\pi^{\pm}$ $K^{\pm}$ ...

Detection

b-jet

Displaced Tracks

Secondary Vertex

Primary Vertex

Jet

L_{xy}

d_0

Prompt Tracks

Jet

B-Tagging

Collision Simulated with MC Generator for multiple events

Multiple simulated events build a sample

ML Algorithm trained with the sample

Results in a dataset

Data Analysis

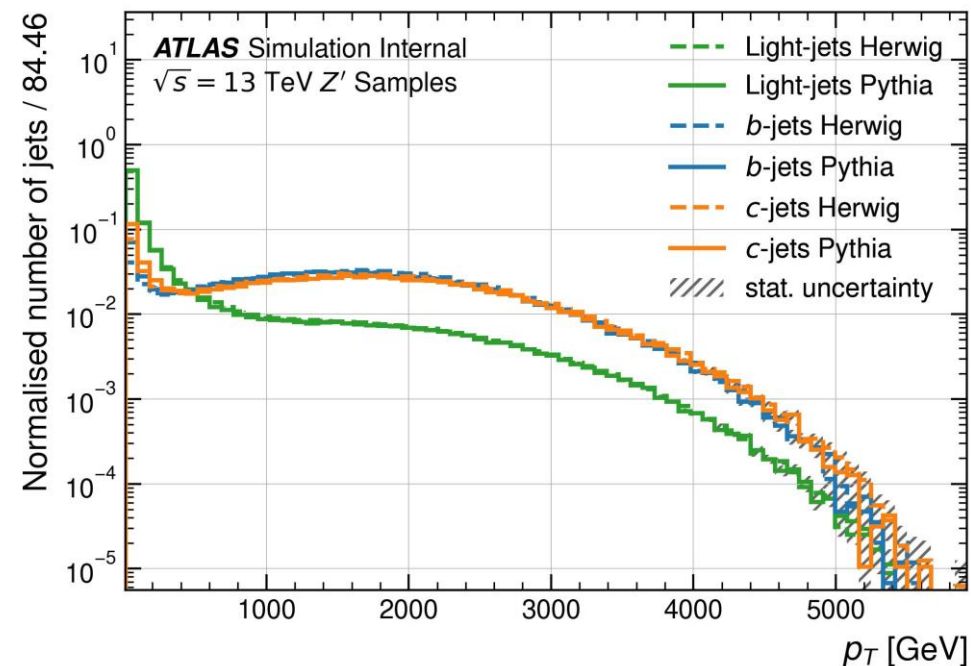| ML Tagger	MC Generator (Simulates the Sample)	Events on the Sample
GN2	Pythia	Zprime
		$t\bar{t}$
	Herwig	Zprime
		$t\bar{t}$

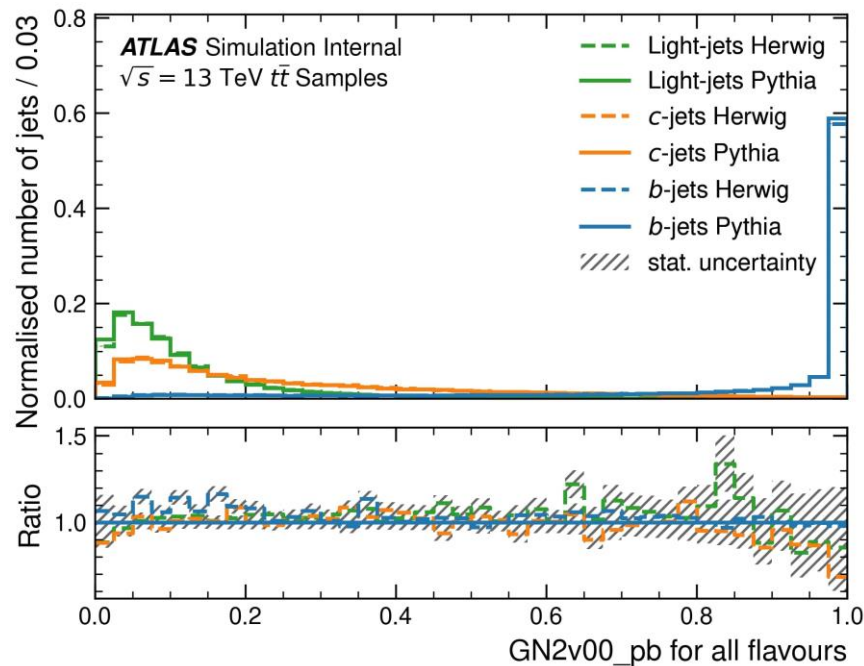# Jet Variables

Transverse Momentum pT

$t\bar{t}$ Sample
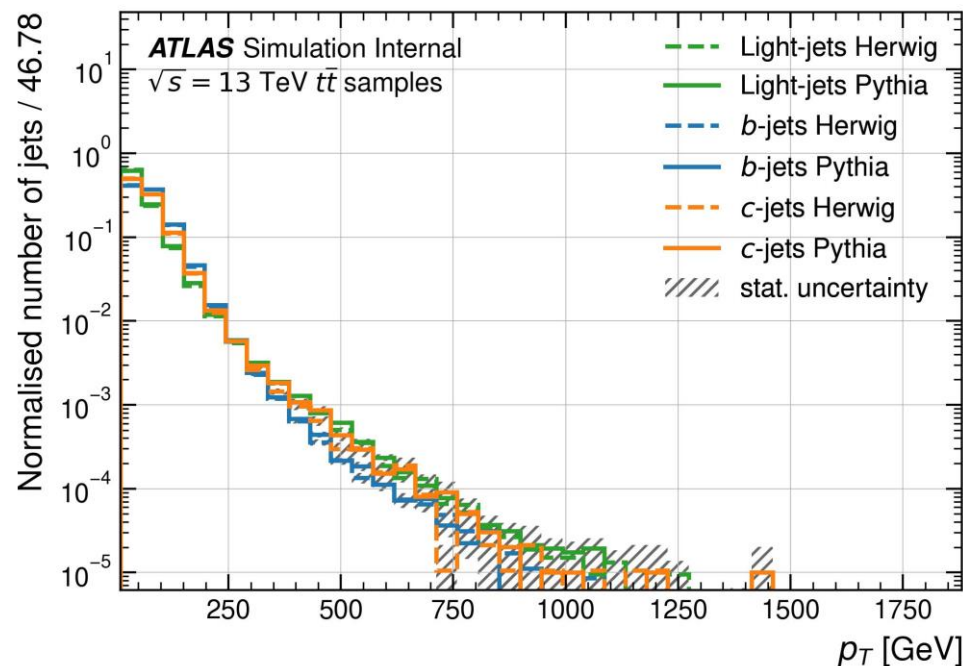
Z' Sample
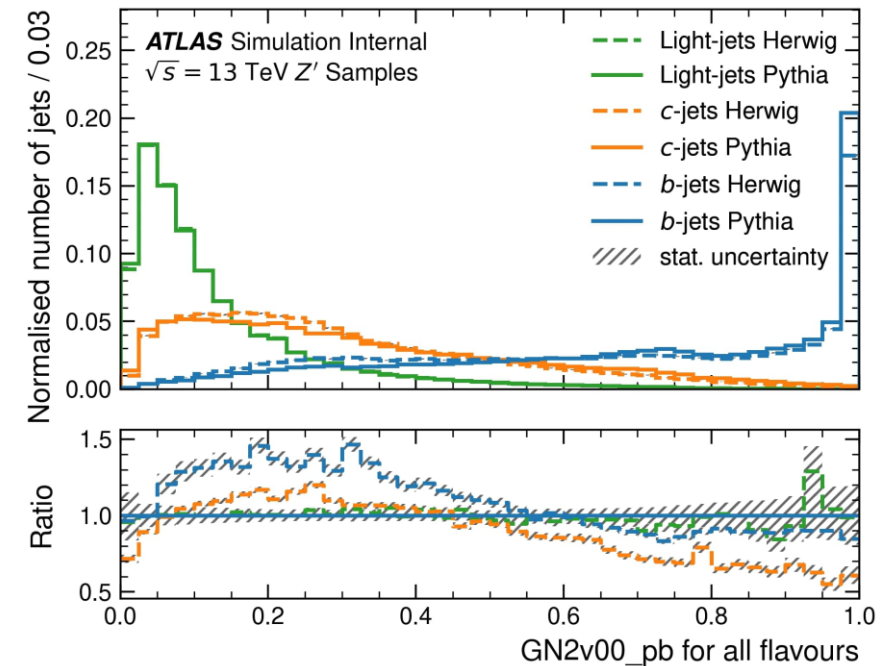

- No differences between Herwig and Pythia;
- Zprime samples have higher pT than ttbar ones;

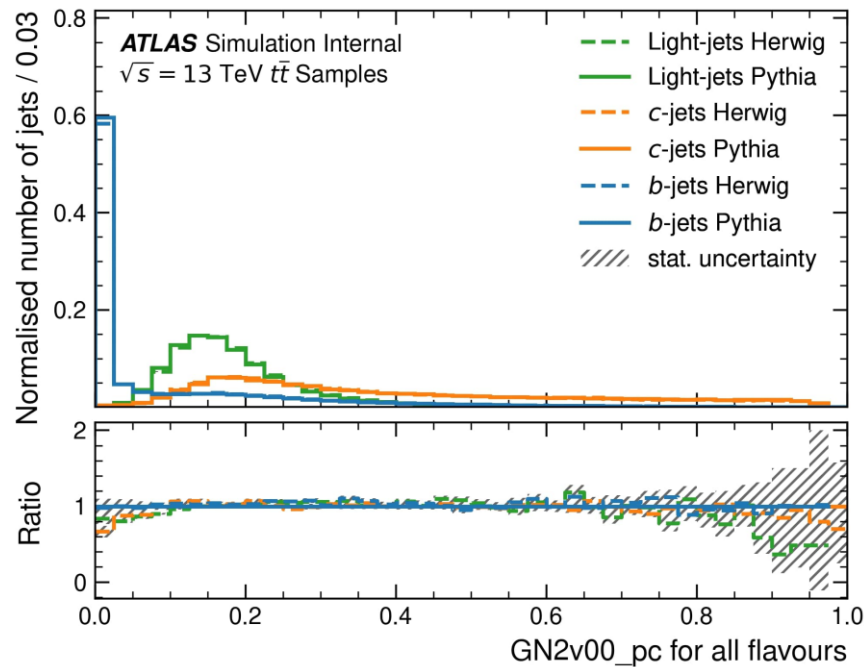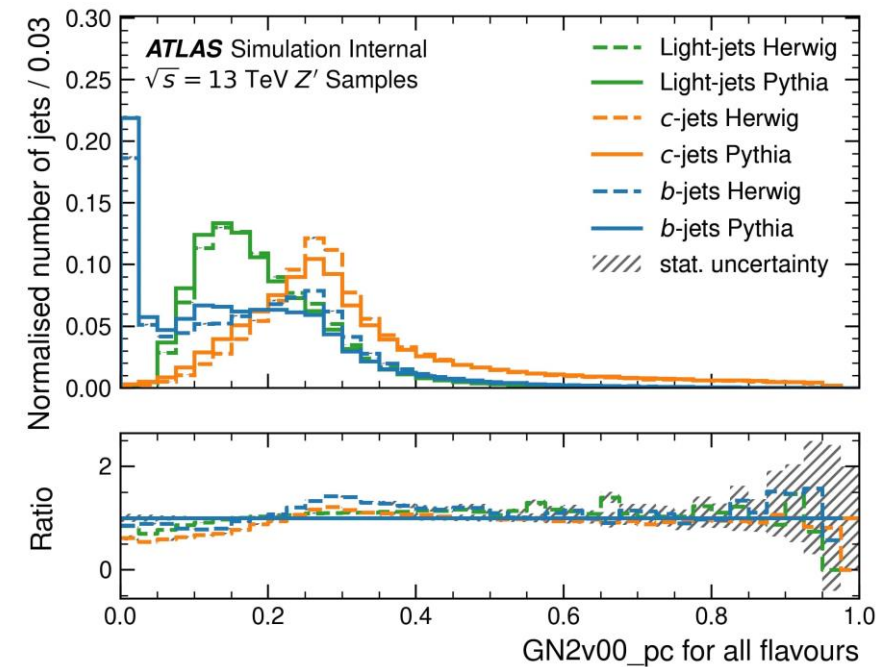# Score Variables from GN2 – Score GN2v00_pb

$t\bar{t}$ Sample

Z' Sample

- Score b peaks at 1 as expected for b jets and around 0 for other jets;
- Differences between Herwig and Pythia can be noticed;
- Peaks of Zprime samples are less intense because this they have higher pT and B Tagging algorithms are not as exact for high pT as they are for lower pT (ttbar) – therefore we have more dispersion in Zprime;

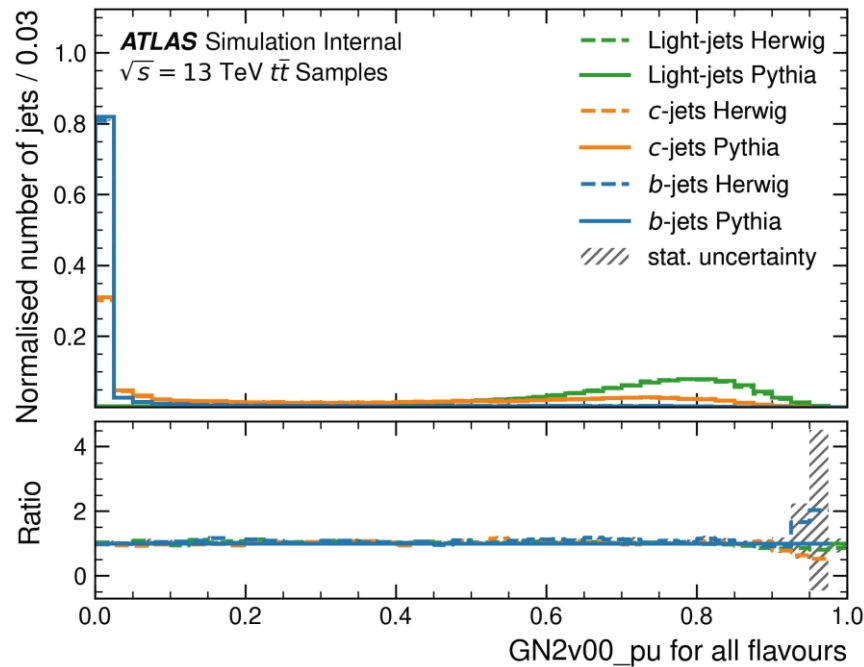# Score Variables from GN2 – Score GN2v00_pc

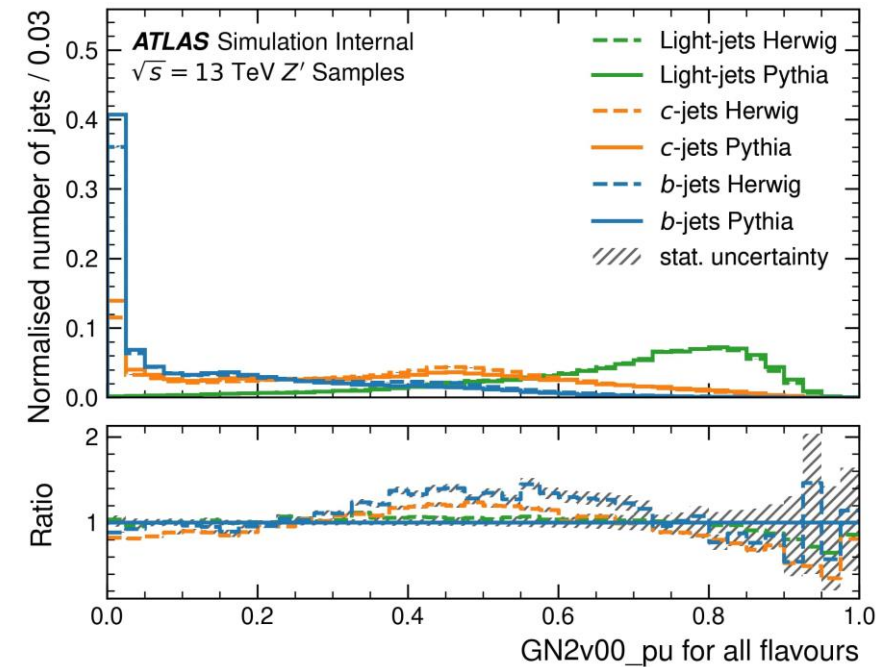$t\bar{t}$ Sample

Z' Sample

- Score c peaks around 0 for all flavours

Score Variables from GN2 – Score GN2v00_pu

$t\bar{t}$ Sample

Z' Sample

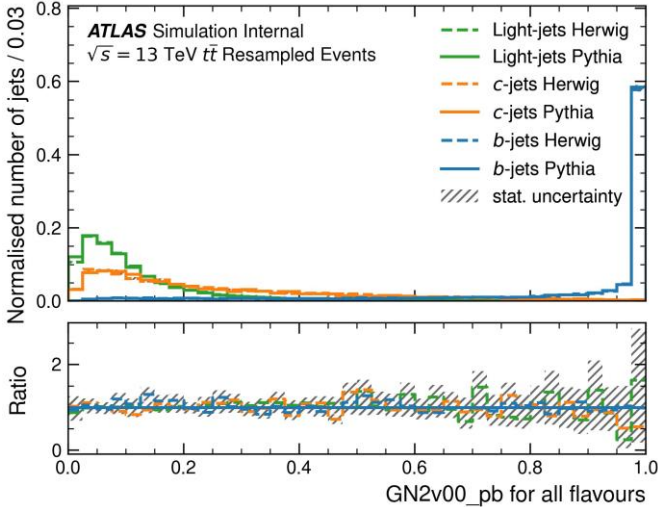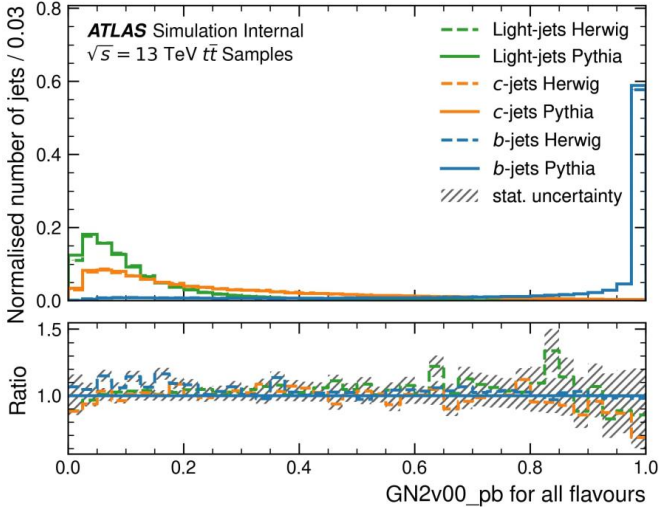- Score u peaks around 0 for all flavours

Resampled Events

Score b for all flavours separated

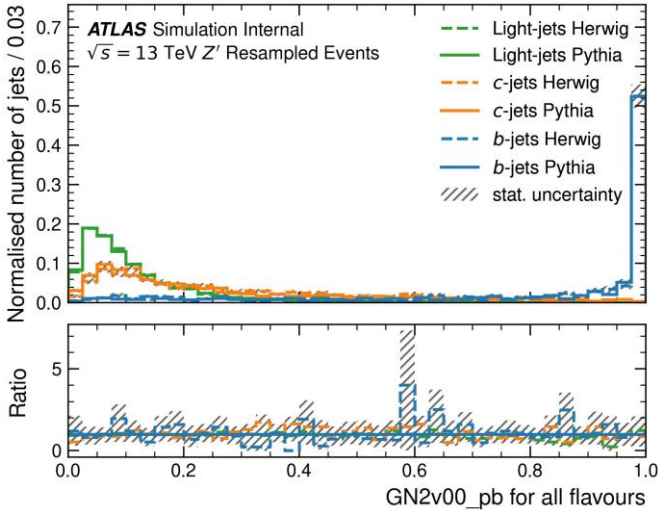|	Non Resampled Events	Resampled Events
$t\bar{t}$ Herwig and Pythia		

Score b for all flavours separated

	Non Resampled Events	Resampled Events
Z' Herwig and Pythia		
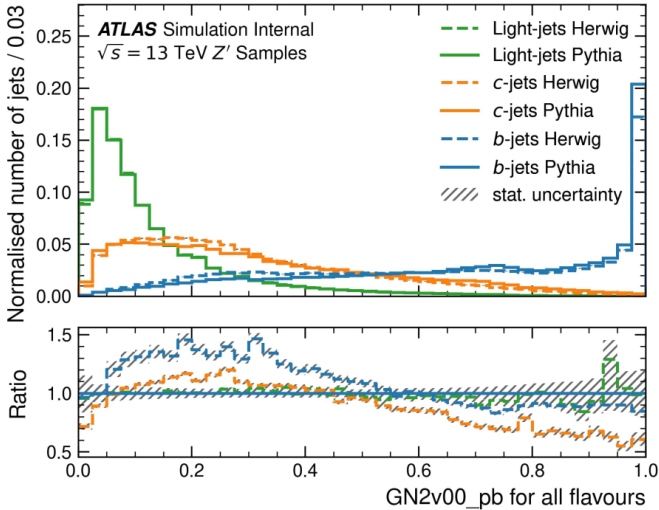
Tendential Lines for Different GN2 Scores

$t\bar{t}$ events - GN2 Herwig

Fit for GN2v00_pb
Fit for GN2v00_pc
Weighted Mean B Jets
Weighted Mean C Jets

scores

GN2v00_pu
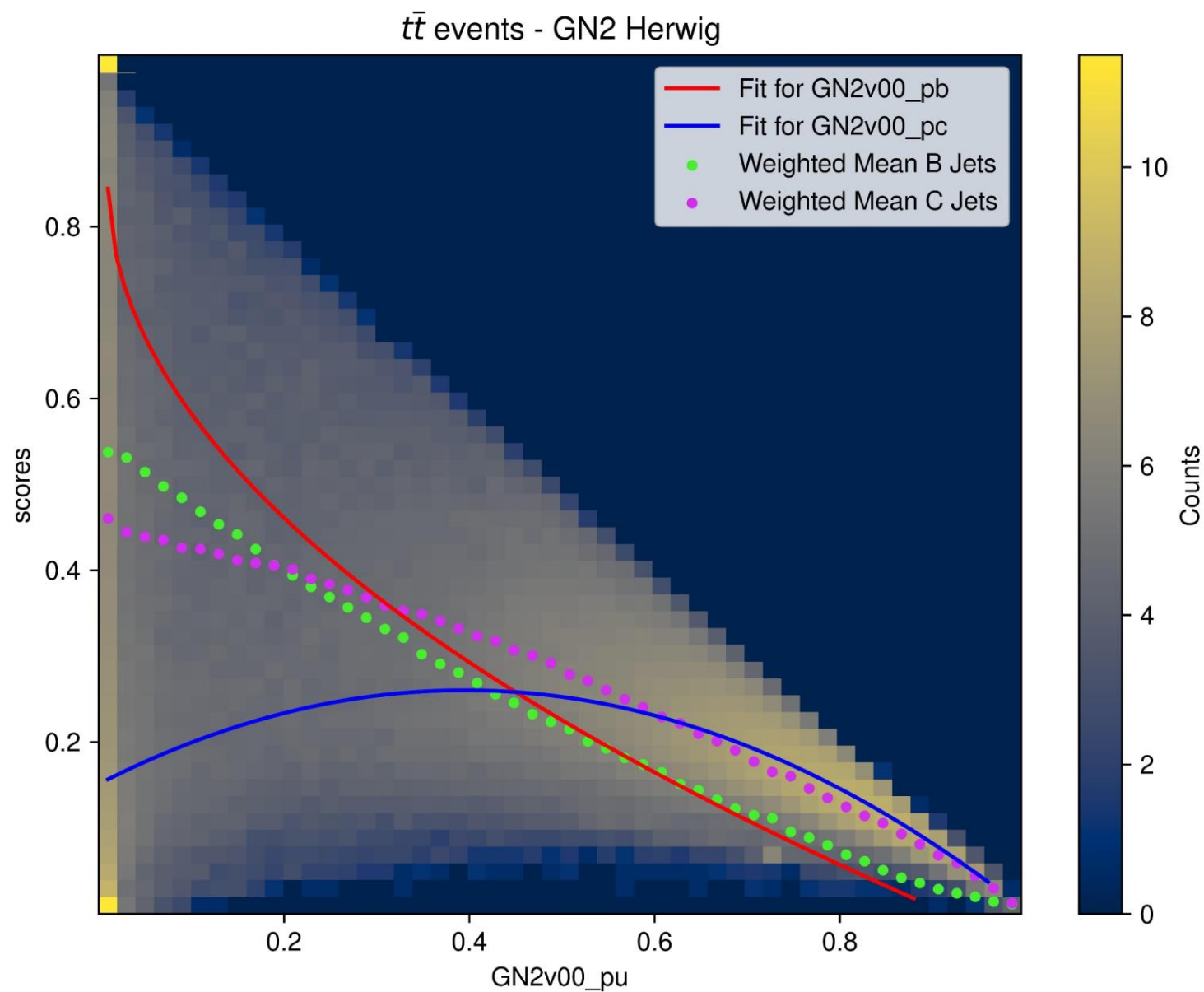
Counts

In order to observe the behavior, we intended to study, two types of techniques were used:

1. 'Tendential Lines': These are curves fitted to the points of the histogram with the highest counts. They have no physical meaning and they are present only to help us understand the behavior of the scores.

2. Weighted Mean: For each x bin of the histogram, we calculated the weighted mean of the y bins according to the number of counts of each bin. This allows us to observe the region where the scores have the highest counts.

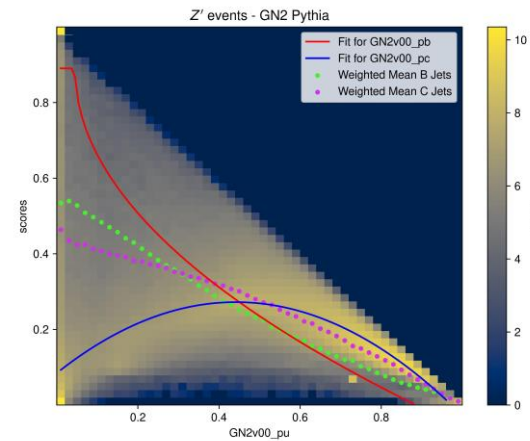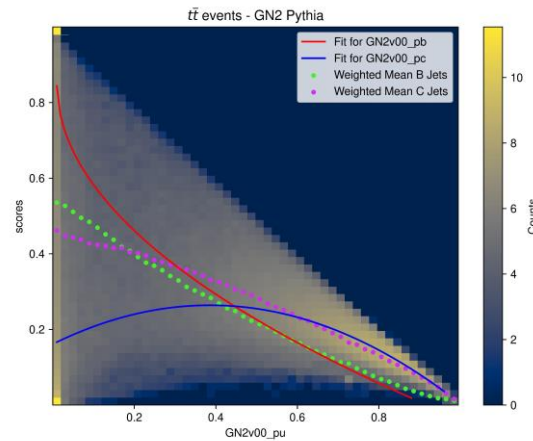# Score u and c+b GN2 with weighted means for both c and b
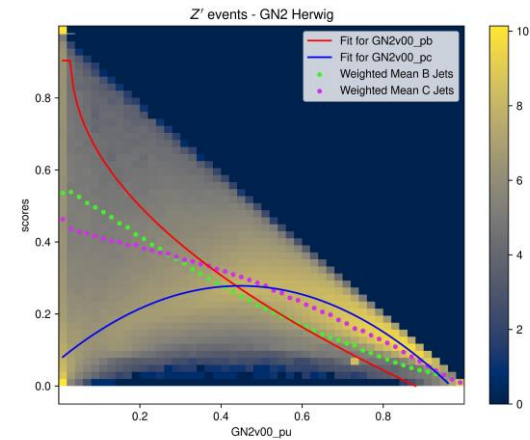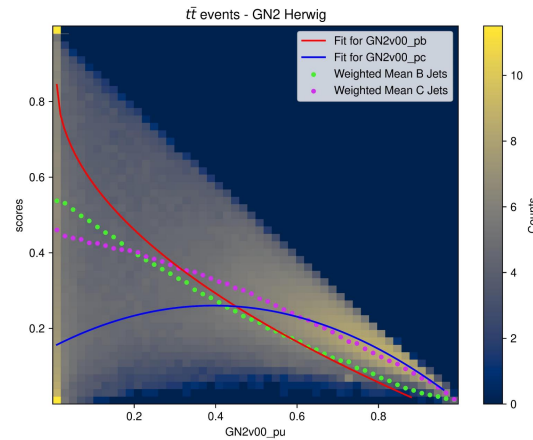
|	$t\bar{t}$ Sample	Z' Sample
Herwig		
Pythia		

We can see that from both the weighted mean and the fits there is a common behaviour for all four cases: when the jet has low u score, it's b score is bigger than it´s score and when it has a high u score the opposite verify. We have confirmed the same tendency verifies for RNNIP scores.
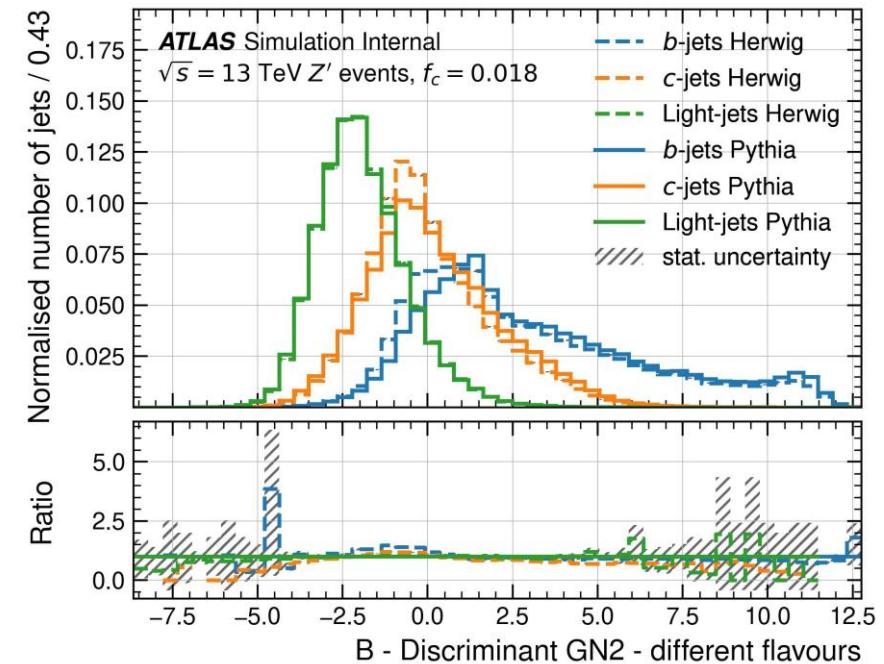
B – Discriminant

B Discriminant for all flavours separated
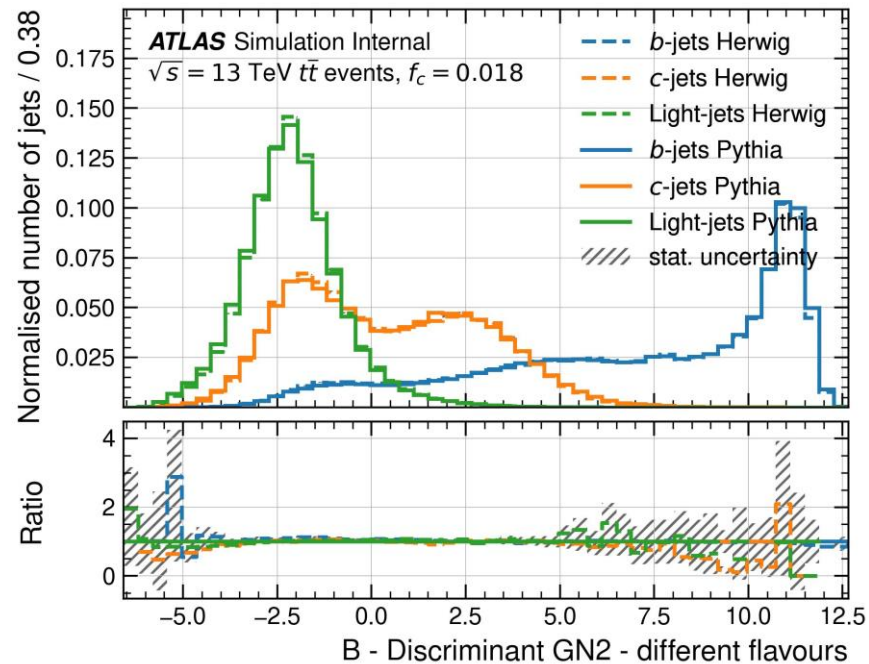
$t\bar{t}$ Sample

Z' Sample

$$D_b = \log\left(\frac{P_b}{f_c \times P_c + f_u \times P_u}\right)$$

End

Back Up

MC Generators

- For MC Generators PYTHIA8 and Herwig7, we considered both $t\bar{t}$ and Z' samples with 1 000 000 jets each for track and jet variables;

- Z' Herwig: user.pgadow.500567.e7954_s3681_r13144_p5627.tdd.EMPFlow.24_2_2.23-04-28_pflow_april23_output.h5

- Z' Pythia: user.pgadow.500568.e7954_s3126_r13144_p5627.tdd.EMPFlow.24_2_2.23-04-28_pflow_april23_output.h5

- $t\bar{t}$ Herwig: user.pgadow.601414.e8472_s3873_r13829_p5627.tdd.EMPFlowTruth.24_2_4.23-06-27_partonshower_output.h5

- $t\bar{t}$ Pythia: user.pgadow.601229.e8453_s3873_r13829_p5627.tdd.EMPFlowTruth.24_2_4.23-06-27_partonshower_output.h5
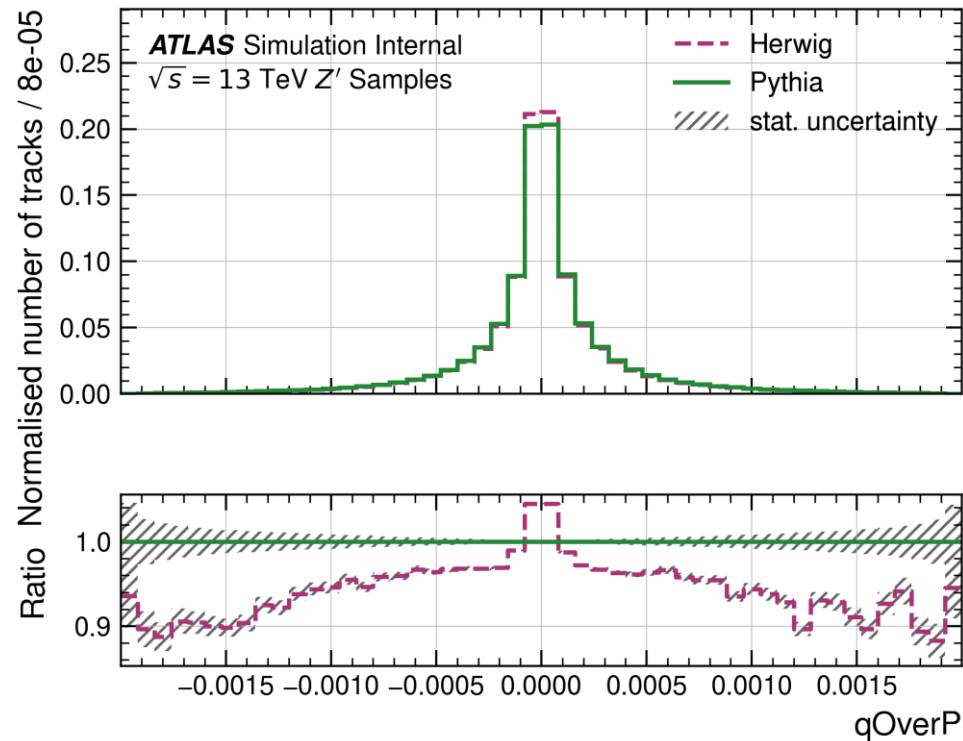
Track Variables

qOverP
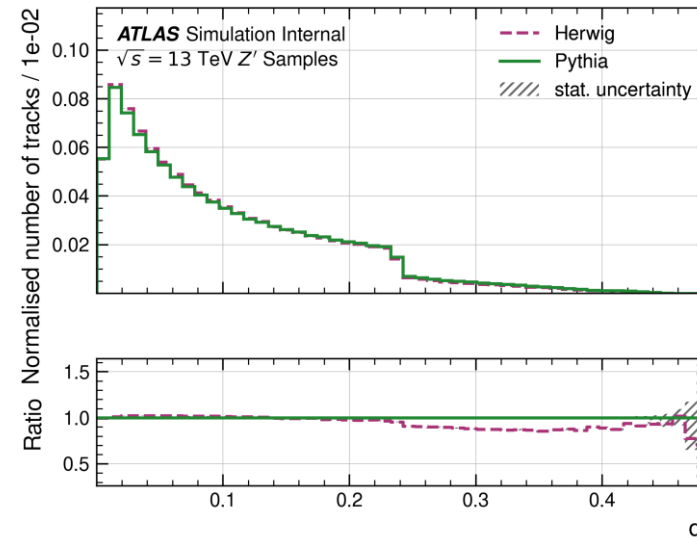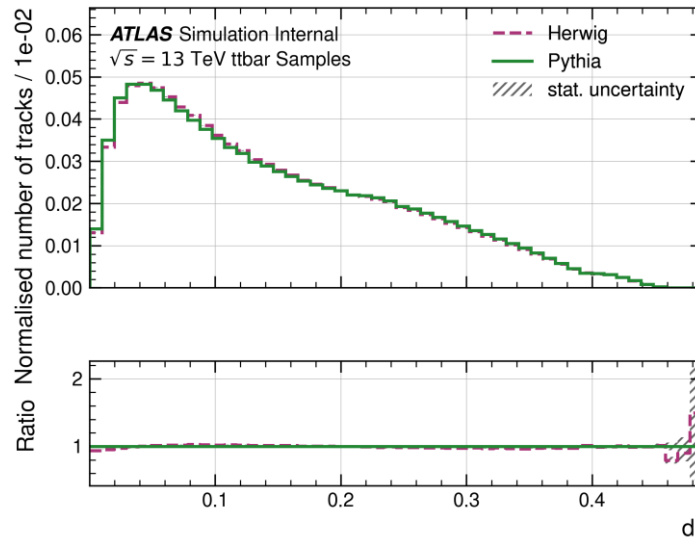

$t\bar{t}$ Sample
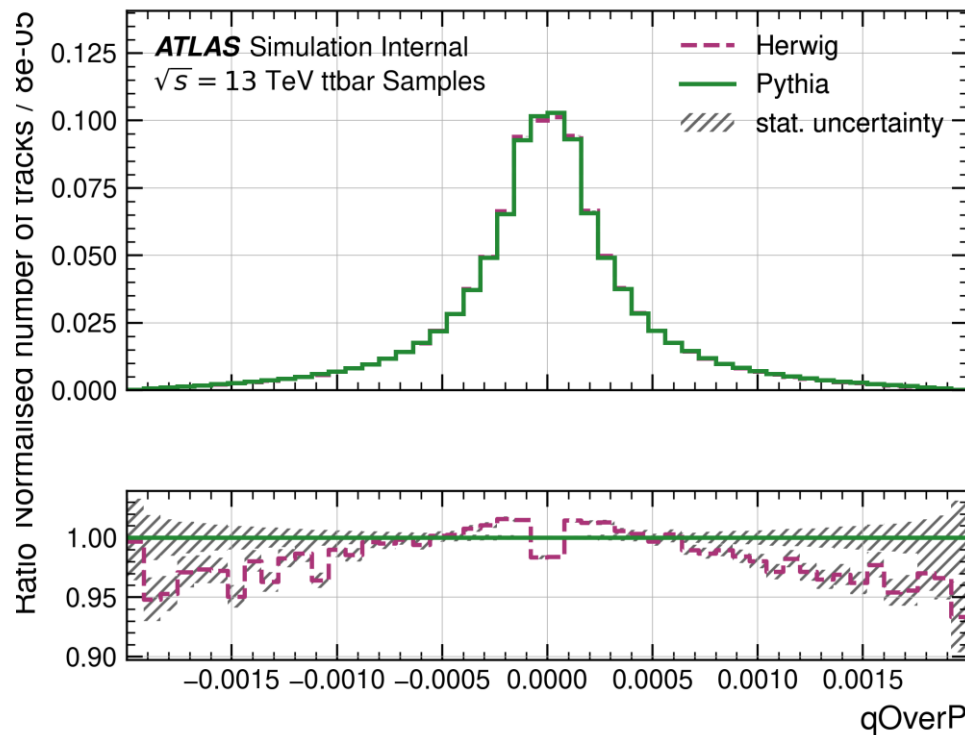
Z' Sample

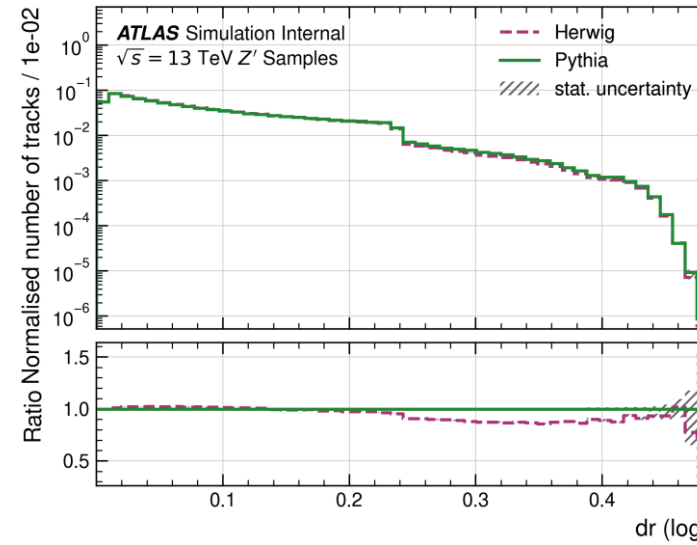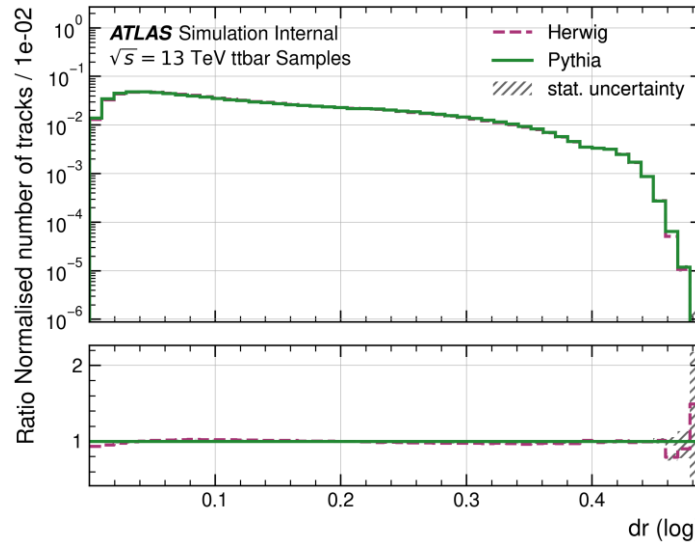- As expected, both distributions are symmetrical because it is equally likely for a charged particle to have charge q or –q.
- Differences between Herwig and Pythia can be observed – they are more significant in the Z' sample.

Delta r

$t\bar{t}$ Sample

Z' Sample

dr with Log scale

dr with linear scale

- Displacement around 0,24 only present in the Z' Sample;
- As expected, the amount of jets decreases as the delta r increases – the further away from the centre of the jet, the least tracks you find;
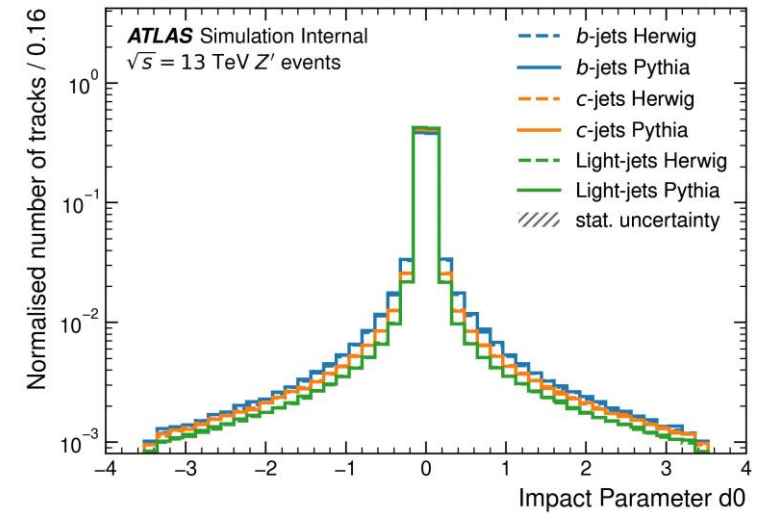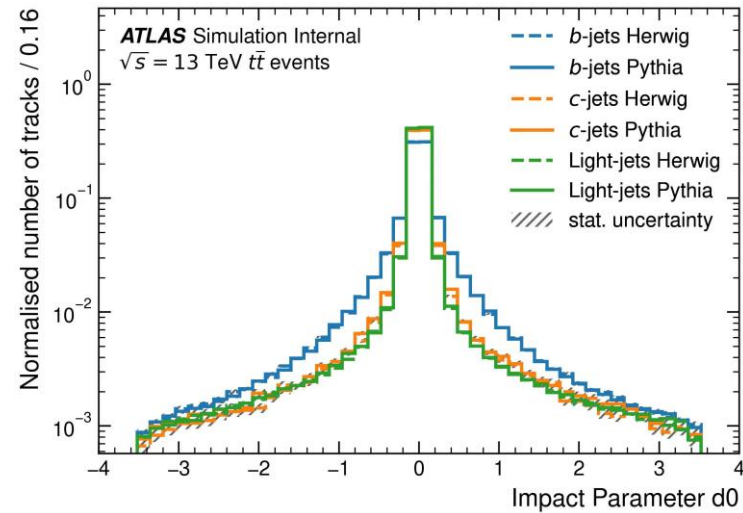- Pythia and Herwig values differ on the Z' sample starting at dr = 0,24;

d0 Plot

d0 all flavours separated

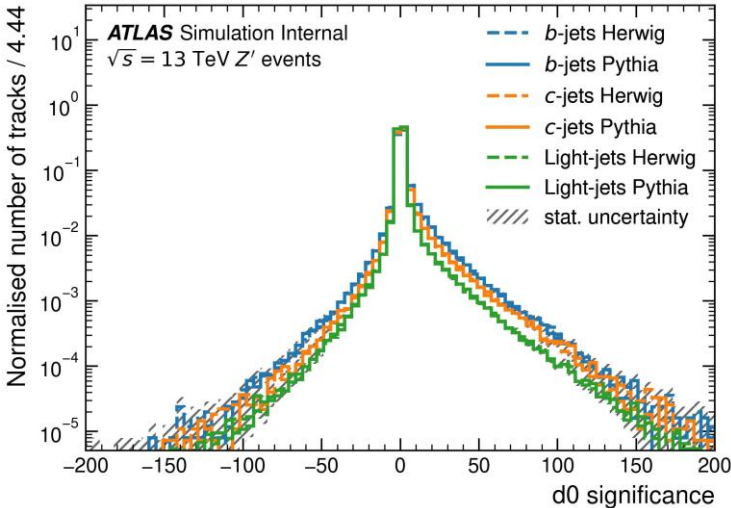|	$t\bar{t}$ Sample	Z' Sample
Herwig and Pythia		

d0 significance all flavours separated

	$t\bar{t}$ Sample	Z' Sample
Herwig and Pythia		
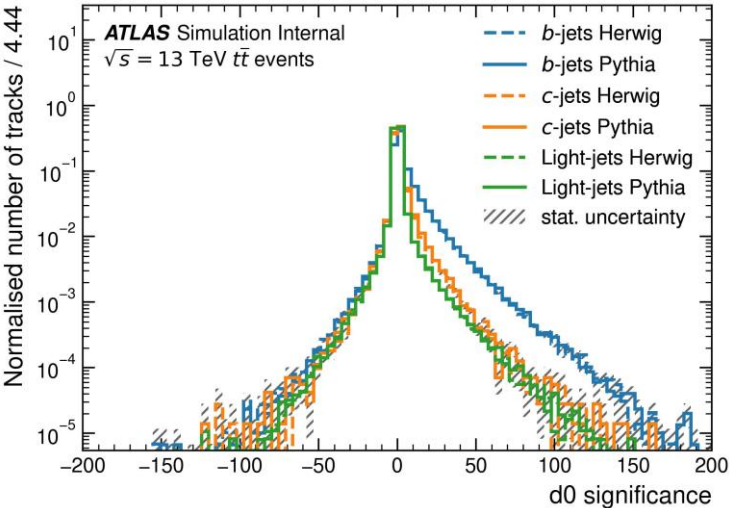
Number of Tracks per Jet

Number of Tracks all flavours separated

	$t\bar{t}$ Sample	Z' Sample
Herwig and Pythia		