August 12, 2025

ADP MEETING
SUMMER STUDENT PRESENTATION

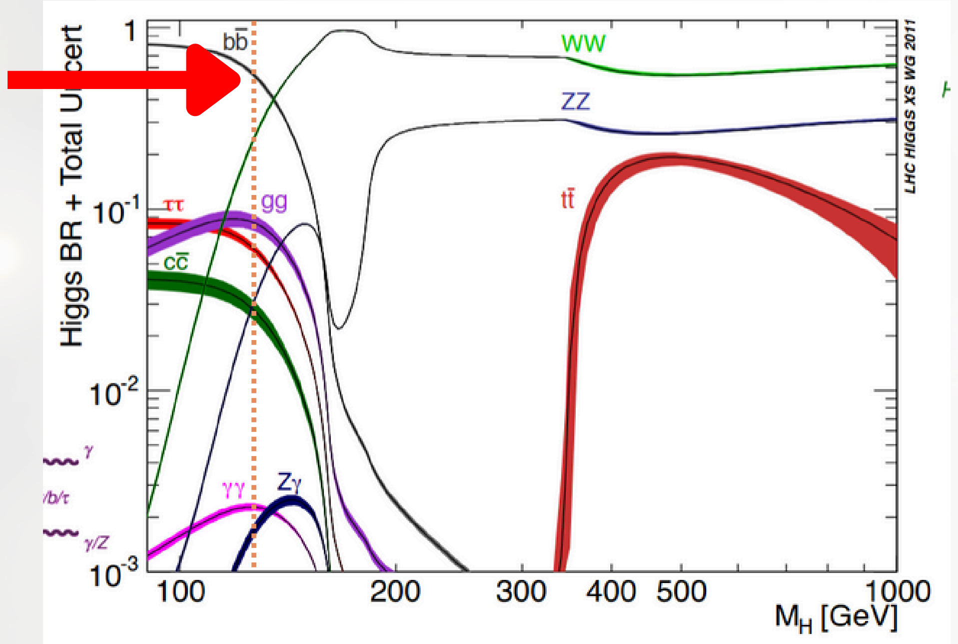
Francis Lance T. Jumawan
University of the Philippines Diliman

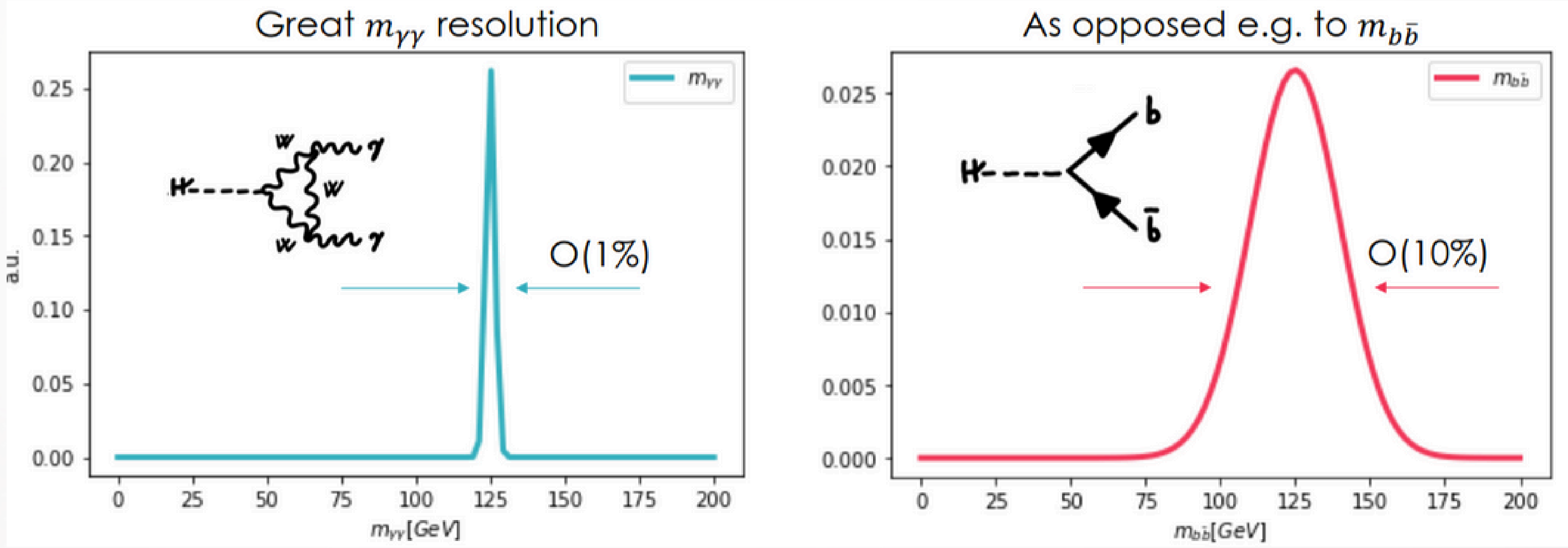
Supervisors:
Dr. Valentina Cairo and Dr. Lorenzo Santi
CERN (HH Analysis Group)

THE PROJECT

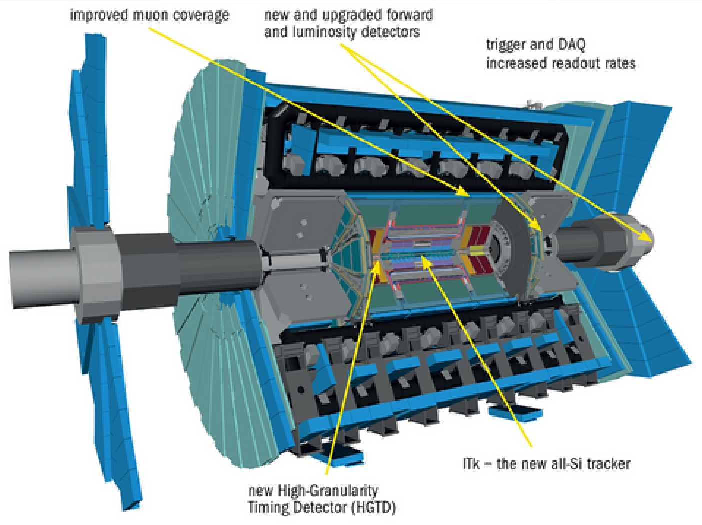
Leveraging Forward Flavour Tagging for bbH(H → γγ) Studies in ATLAS at the HL-LHC

The Higgs-bottom Yukawa coupling y_b governs the interaction between the Higgs field and the b quark

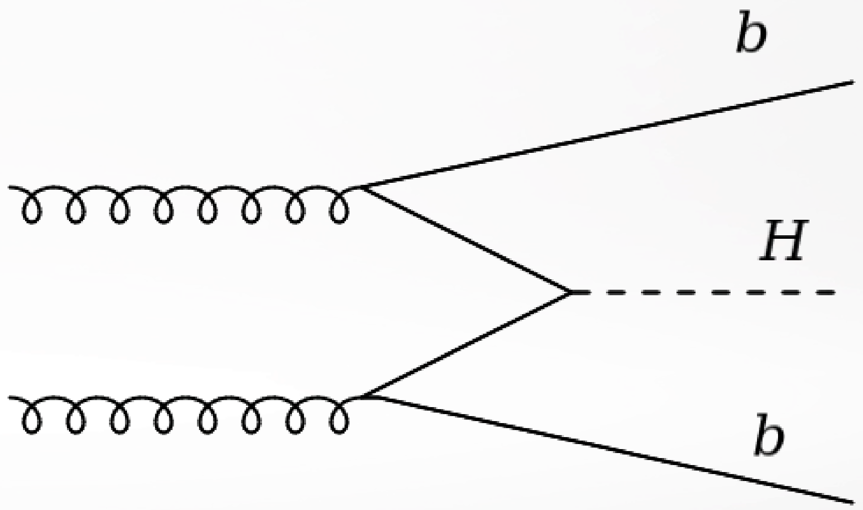




In bbH production, bjets are usually forward and thus have a soft pT



ATLAS EXPERIMENT

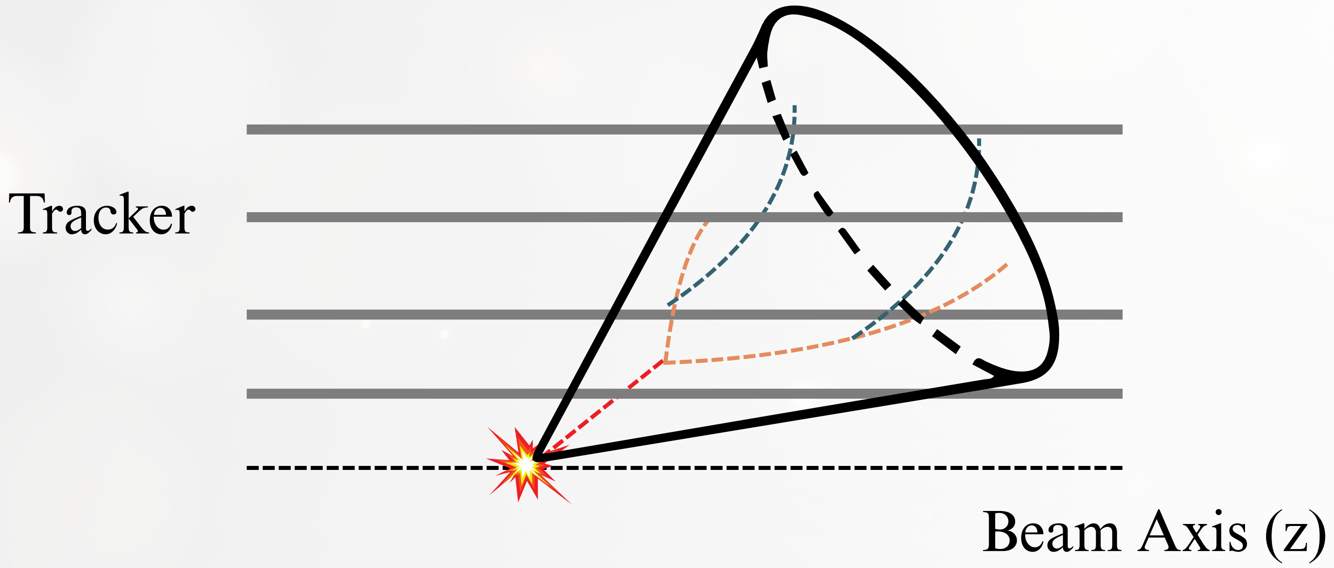


Probes y_b directly at tree level

REPORT
TABLE OF CONTENTS

INTRODUCTION
WHAT IS FLAVOUR TAGGING?

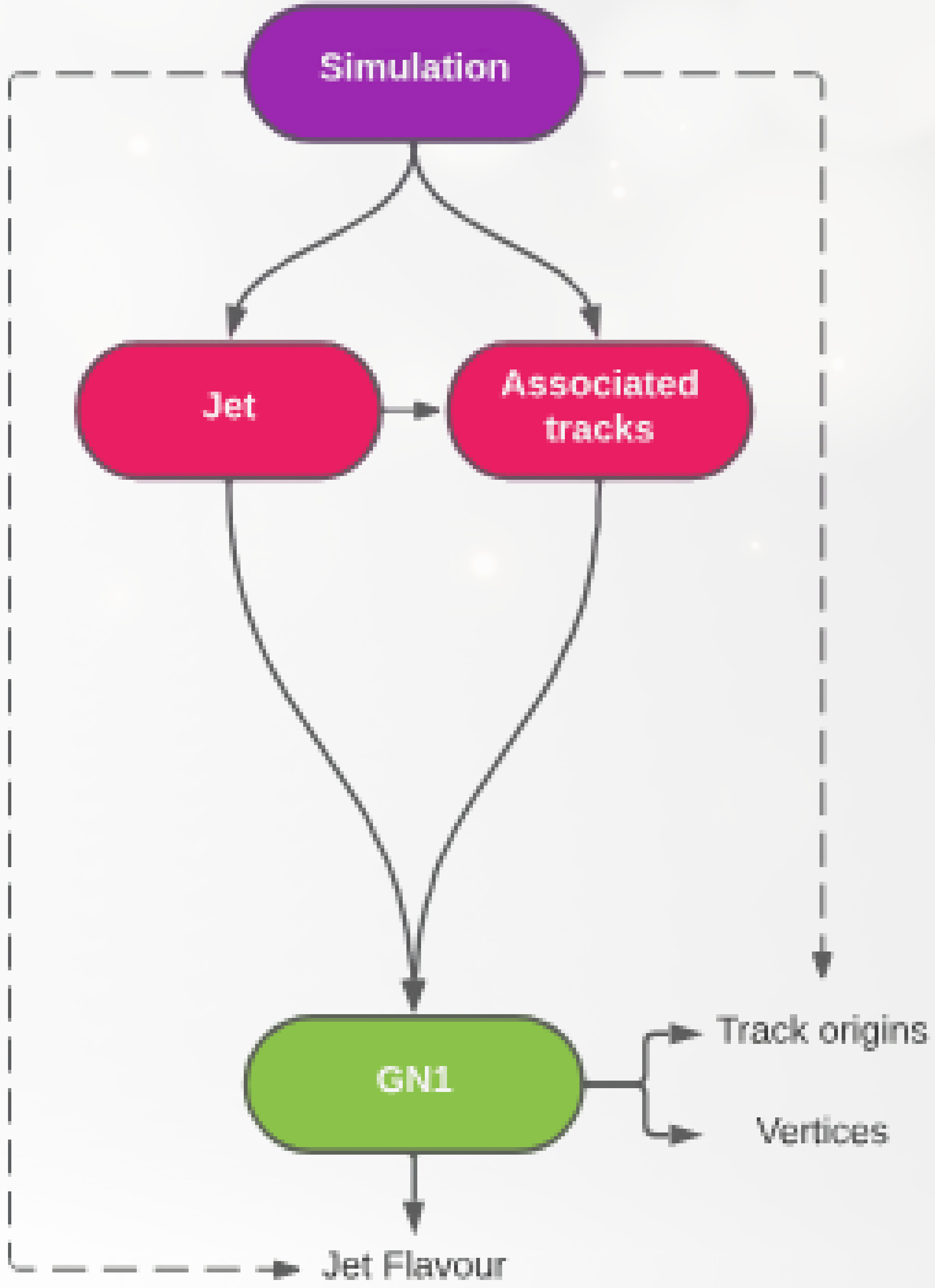
- Identify the flavour of the parton from which a jet originates.



GN2

A tag from 1 to 6 (b-jet like)

WP 77%: 4, 5, 6
WP 85%: 3, 4, 5, 6



Reality and Limitations

- No HL-LHC samples currently available. (**Run 2 and Run 3** samples)
  - Which still has eta acceptance of 2.5 and below
  - The sample versions I am using is from the $HH \to b\bar{b}\gamma\gamma$

Probing the Higgs b-quark coupling (Y_b) : bbH Analysis Feasibility study

ABOUT THE SAMPLES

SIGNAL

Higgs produced in association with b-quarks (bbH)

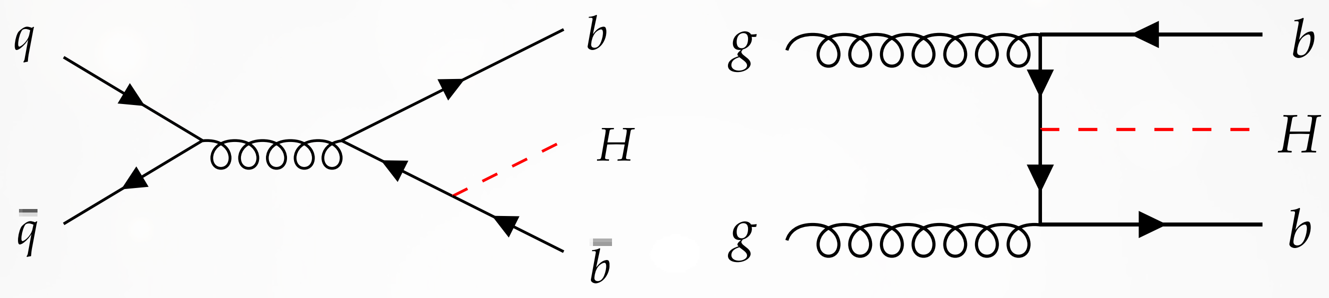
Final States: bb(H → yy)

- 2 b-jets and 2 photons.

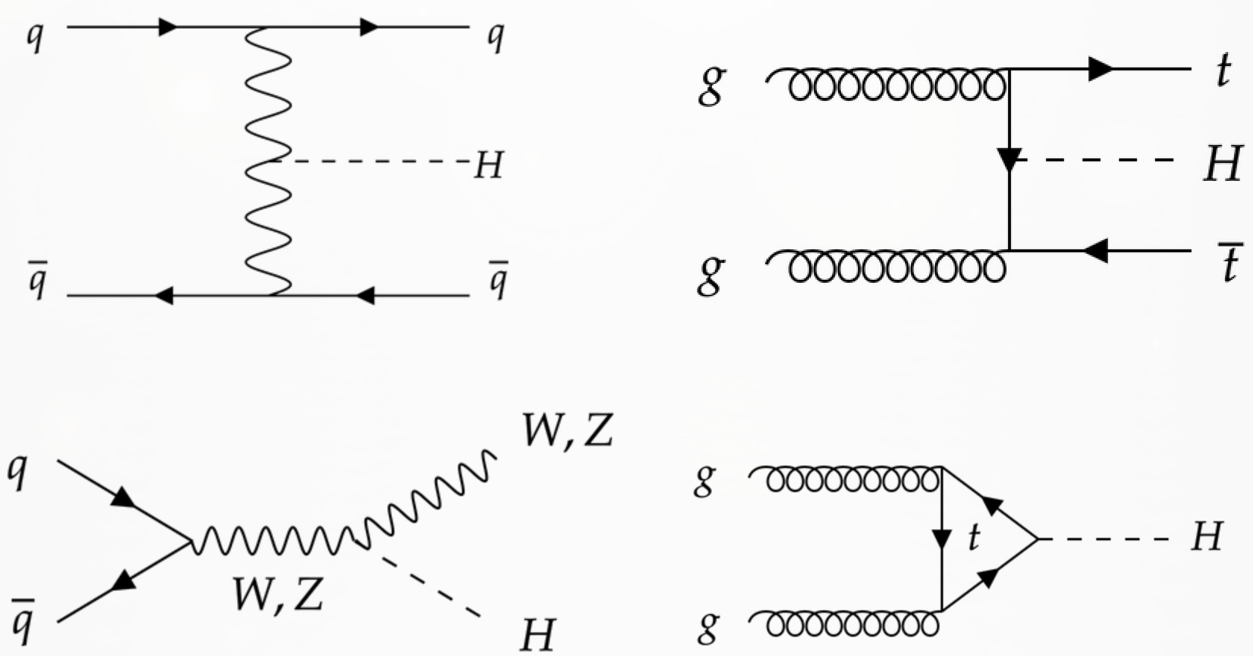
Cross-section: ~0.5 pb at 13 TeV

Feynman Diagrams

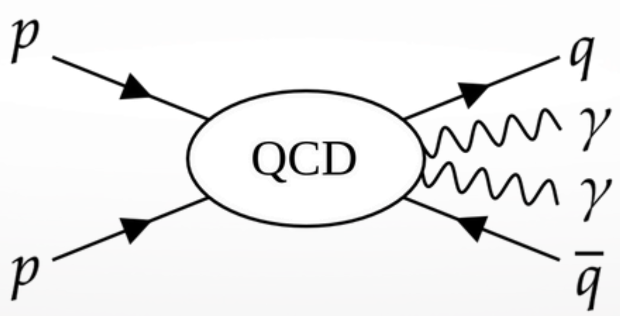
Signal



Resonant Backgrounds



Non-resonant Backgrounds



BACKGROUND

Processes with the same finals states

Final States: 2 jets and 2 photons.

QCD + yy

Vector Boson Fusion

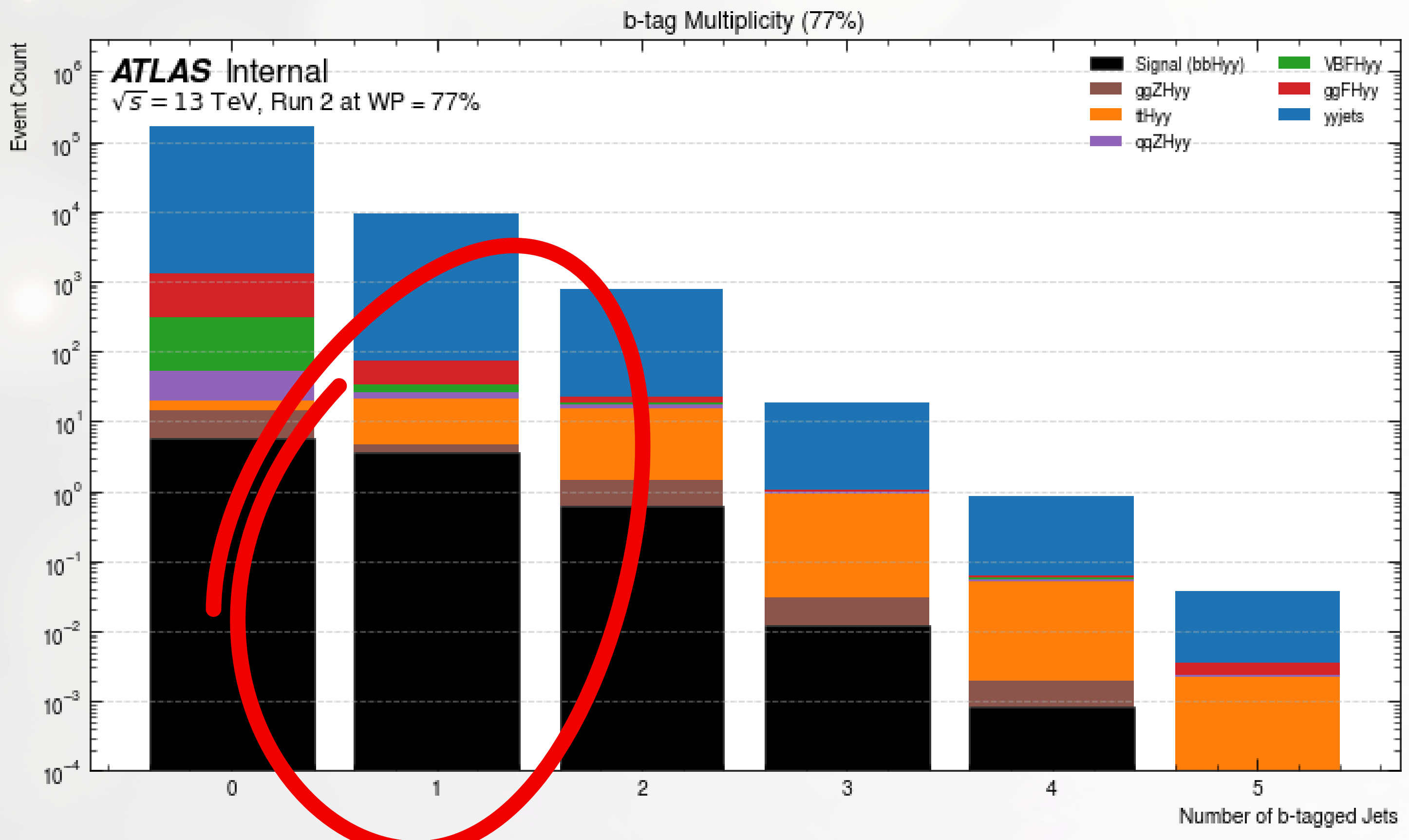
gluon-gluon Fusion

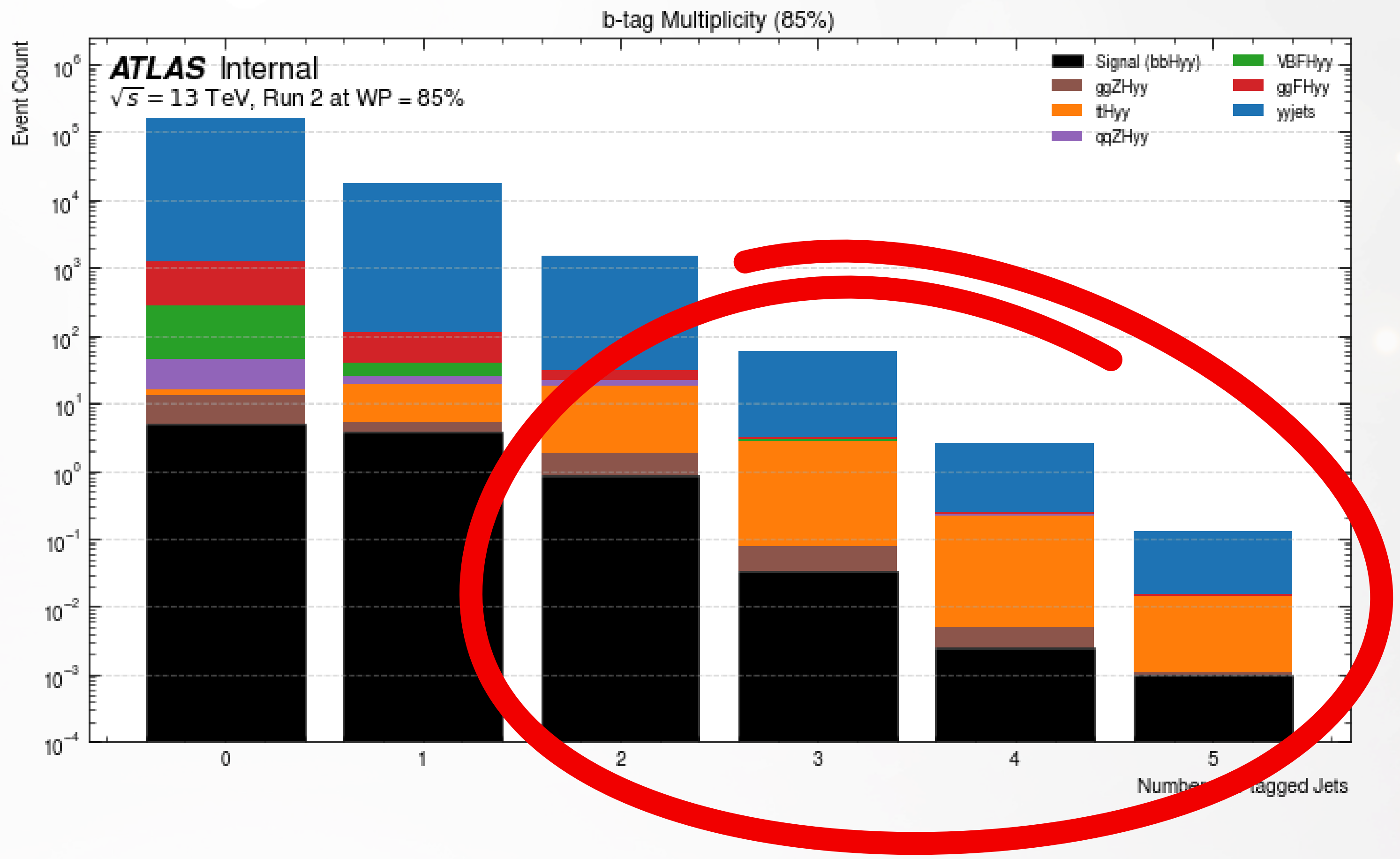
Higgstrahlung

Higgs production with top quarks (ttH)

ANALYSIS
FLOWCHART

- KINEMATIC ANALYSIS
- SIGNAL REGION
- MULTIVARIATE ANALYSIS
- BINNED SIGNIFICANCE
- STATISTICAL FIT





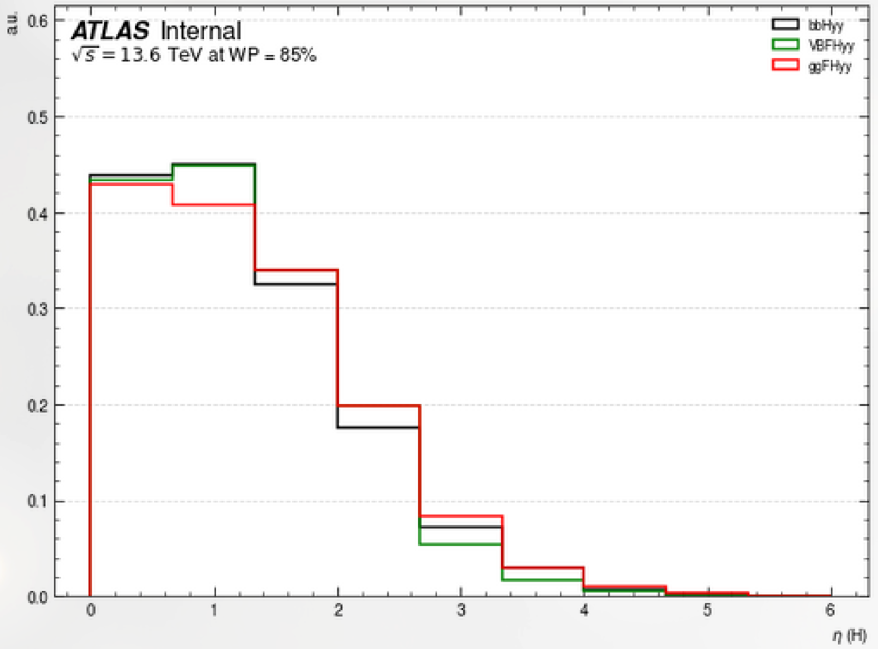
KINEMATIC ANALYSIS

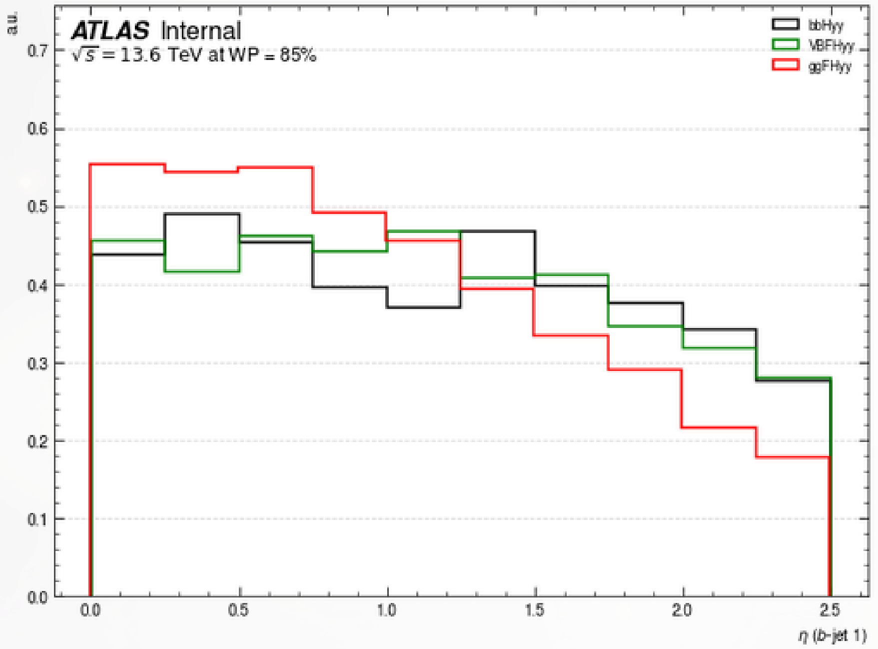
1

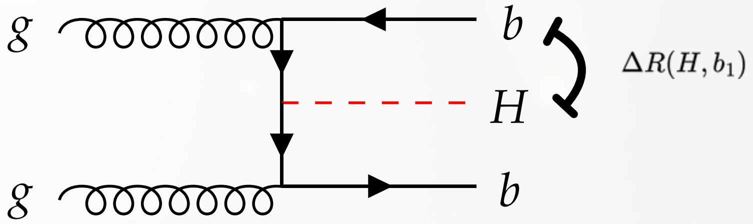
2

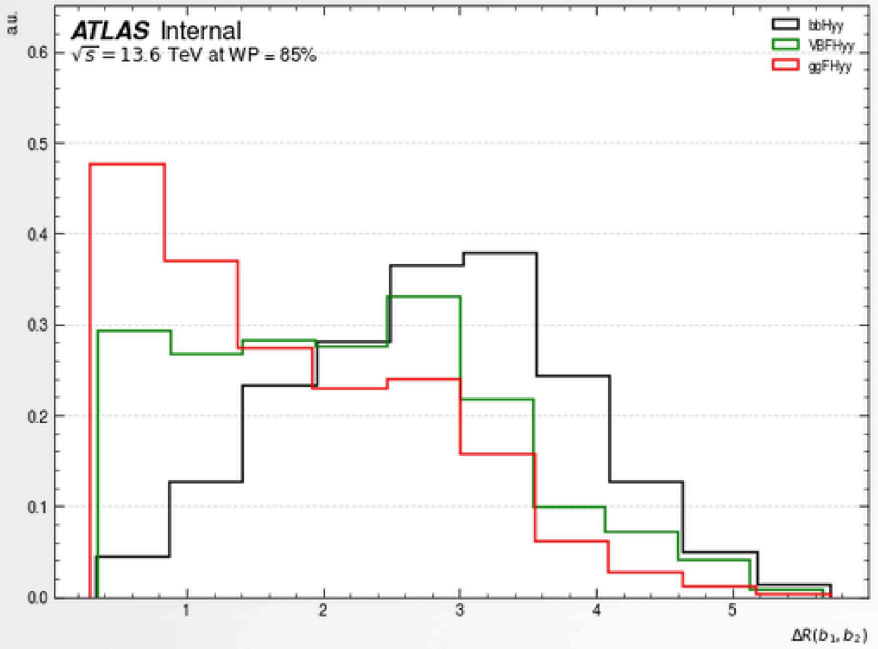
KINEMATIC ANALYSIS

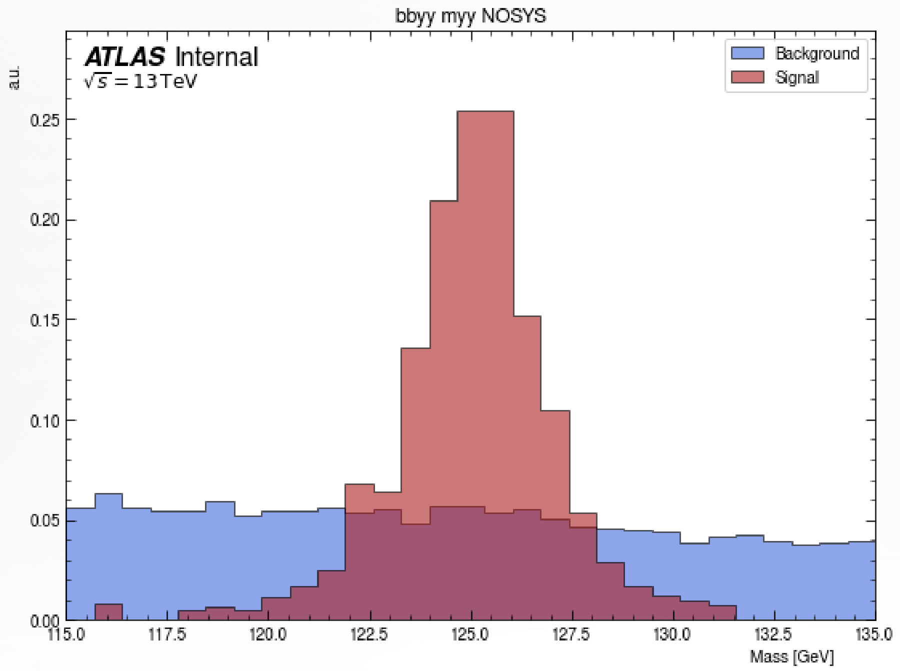
1

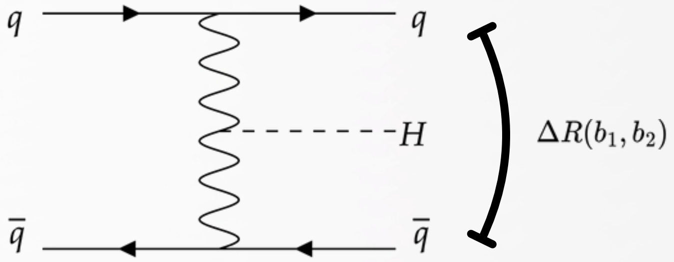










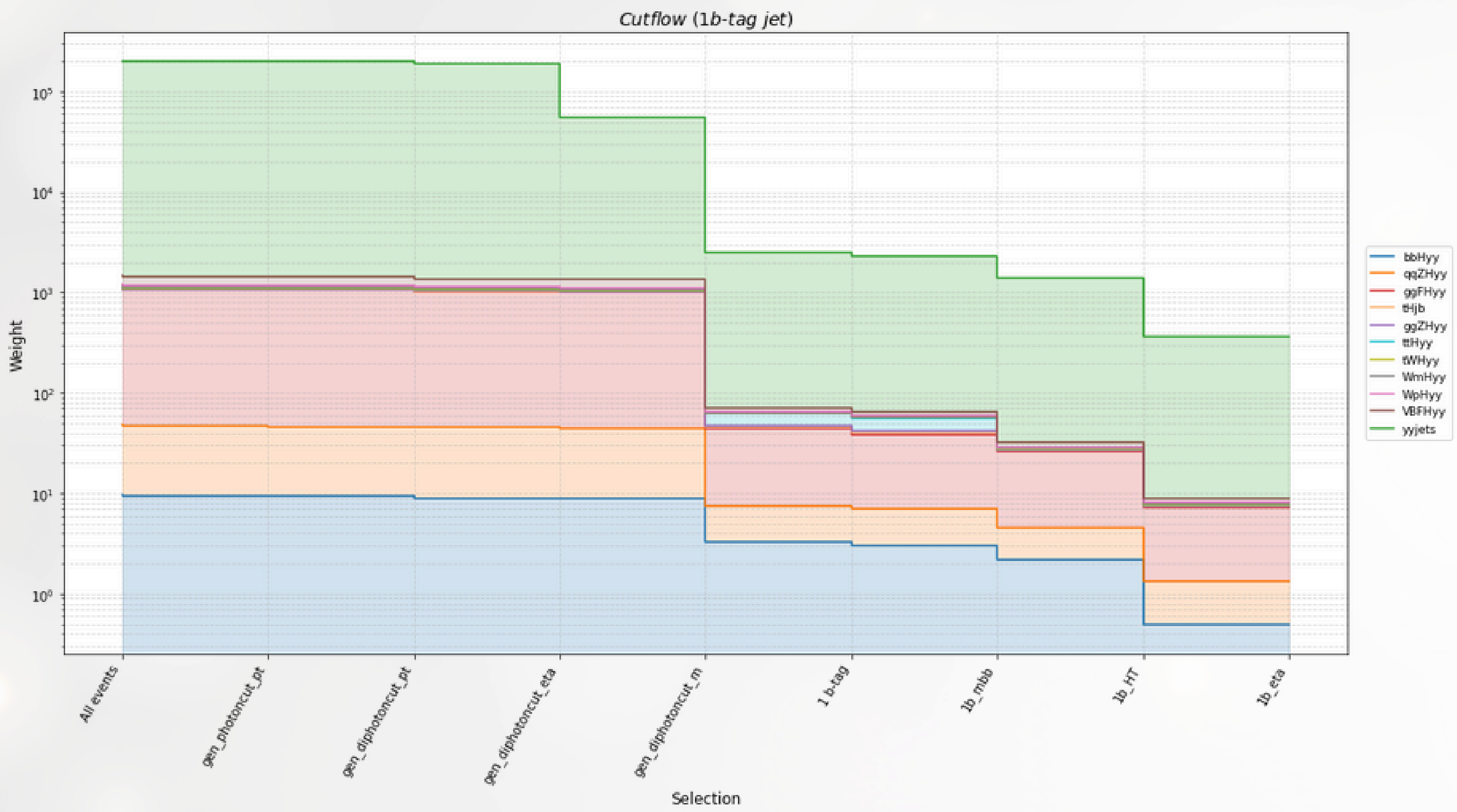


2

CUT-FLOW ANALYSIS

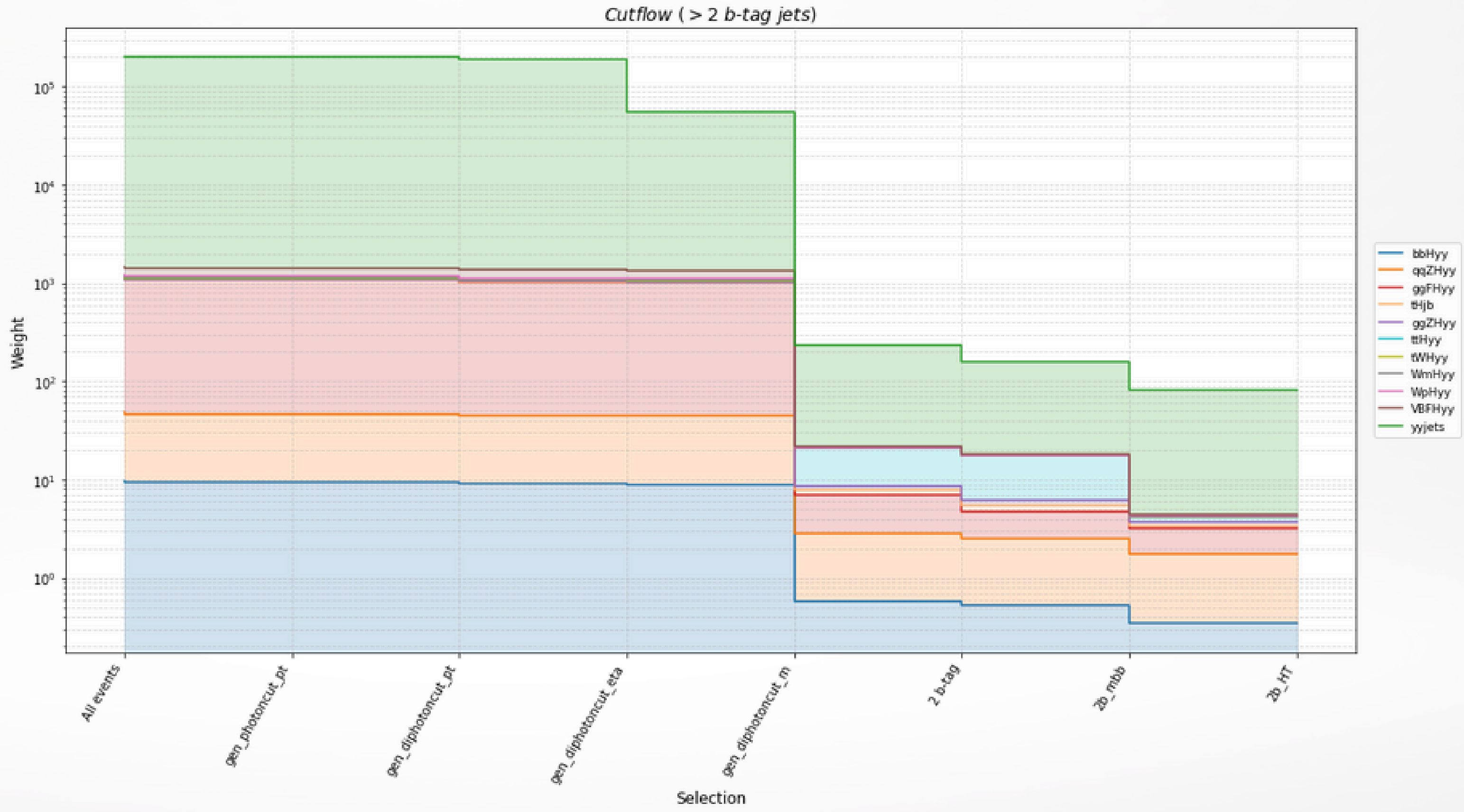
(ATTEMPT 1)

2



$$Z_{\text{total}} = 0.026$$

$$Z_{\text{total}} = 0.038$$

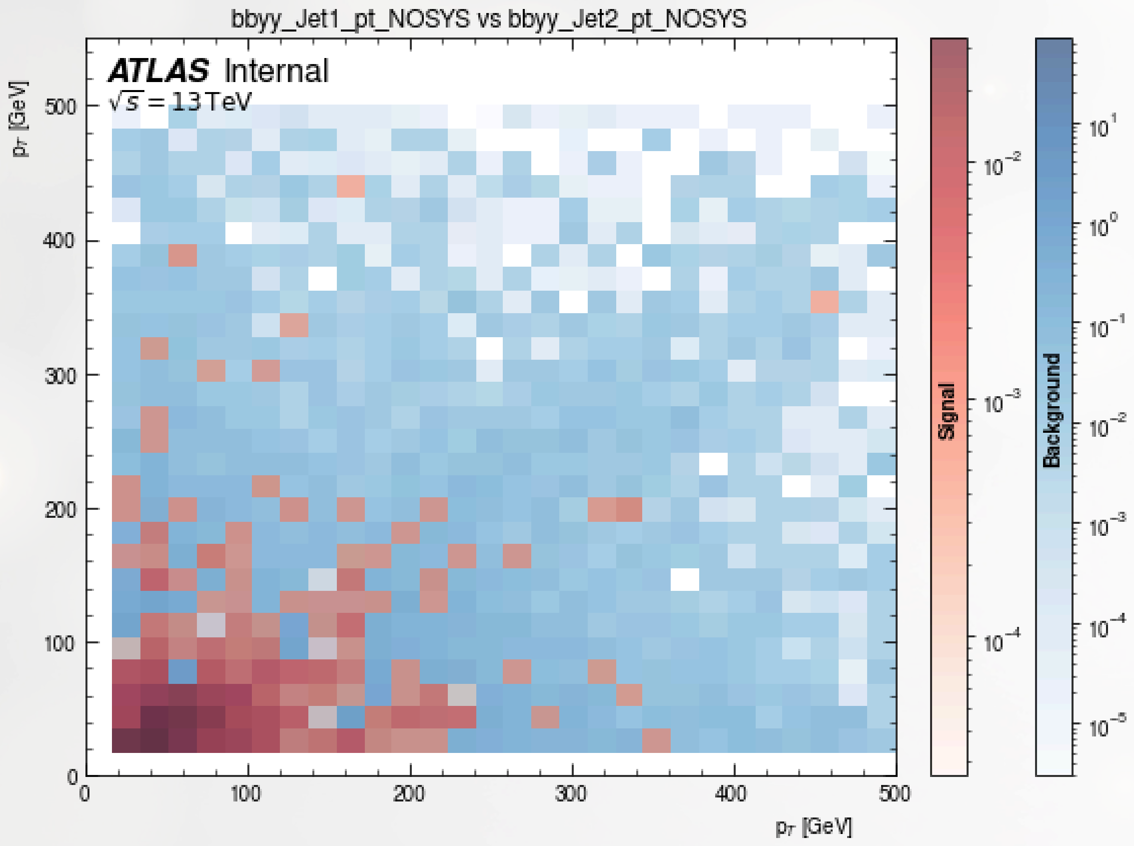


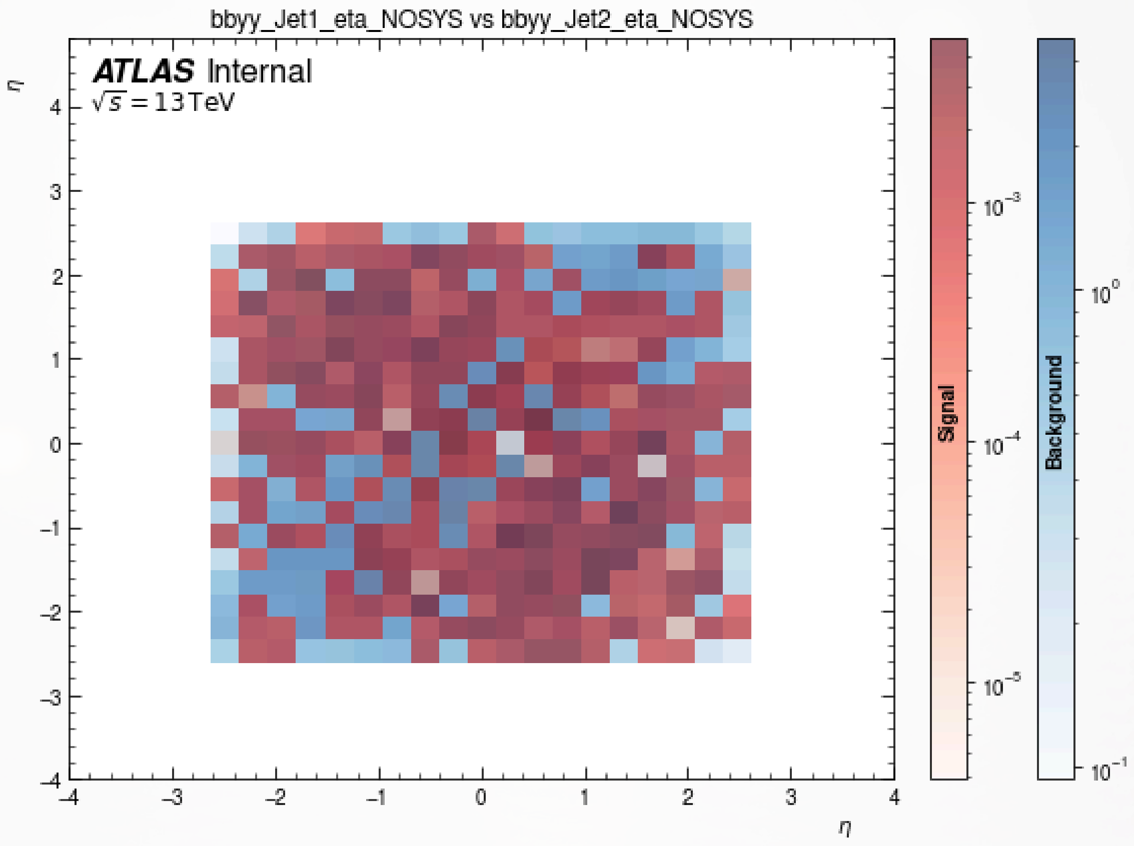
3

3

FEATURES

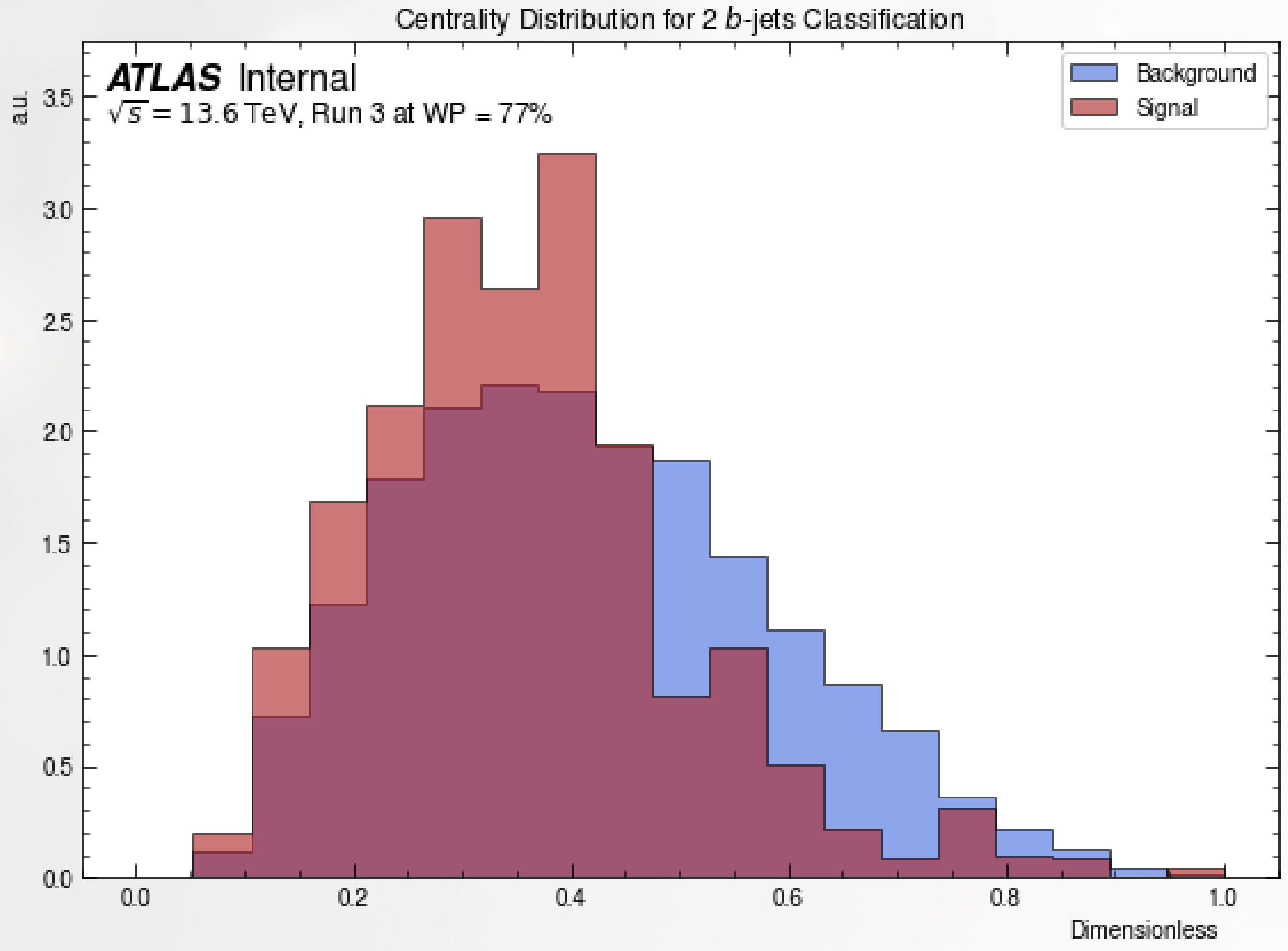
leading jets (b-jet candidates)



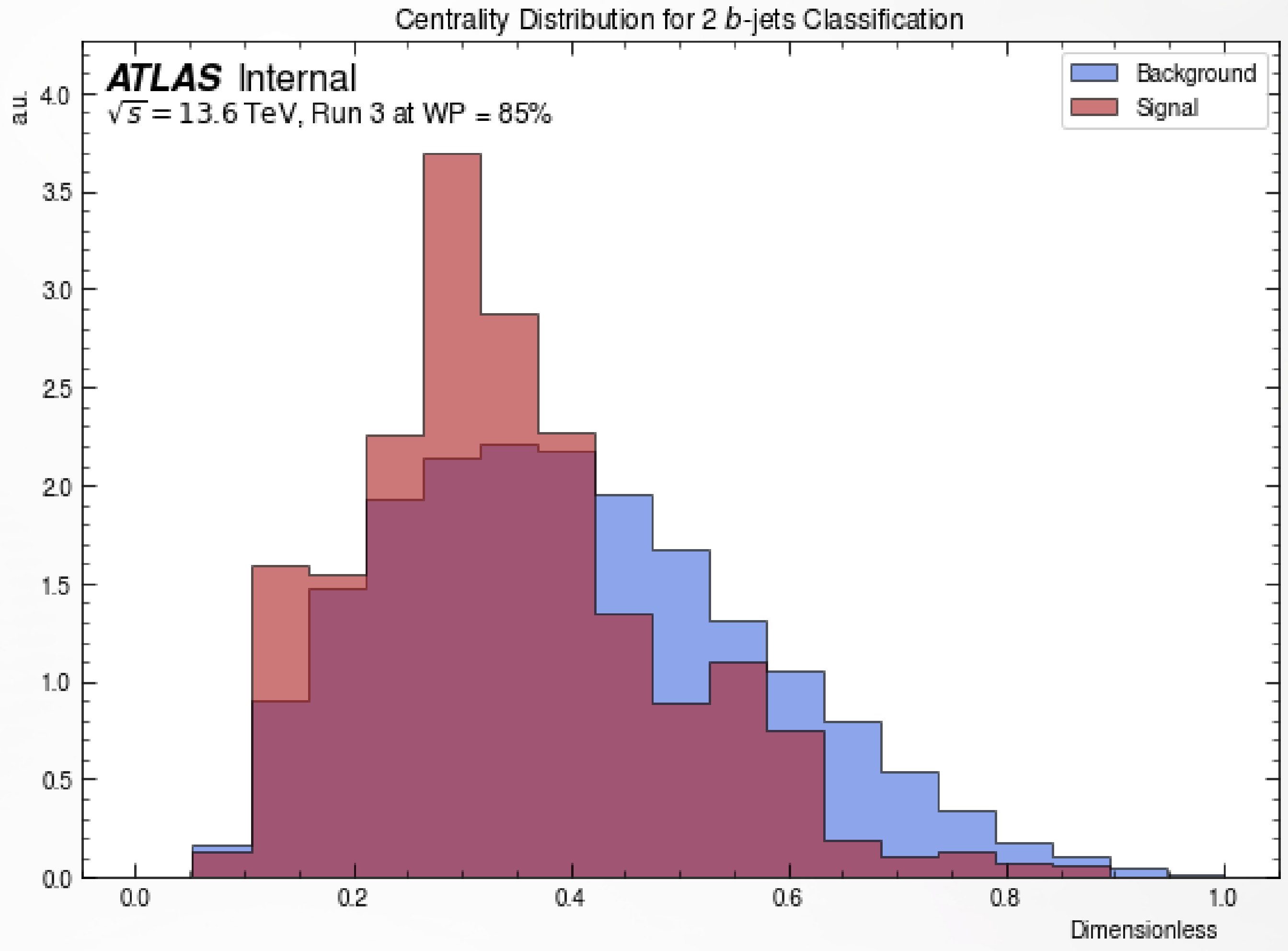


$$\text{Central}_{p_T} = \frac{\sum_{i=1}^{4} (\text{Jet } \{\text{i}\}) p_T \exp(-|\eta_i|)}{\sum_{i=1}^{4} (\text{Jet } \{\text{i}\}) p_T}$$

4



Centrality Distribution for 2 b-jets Classification
ATLAS Internal
√s = 13.6 TeV, Run 3 at WP = 77%
Background
Signal
a.u.
Dimensionless



Centrality Distribution for 2 b-jets Classification
ATLAS Internal
√s = 13.6 TeV, Run 3 at WP = 85%
Background
Signal
a.u.
Dimensionless

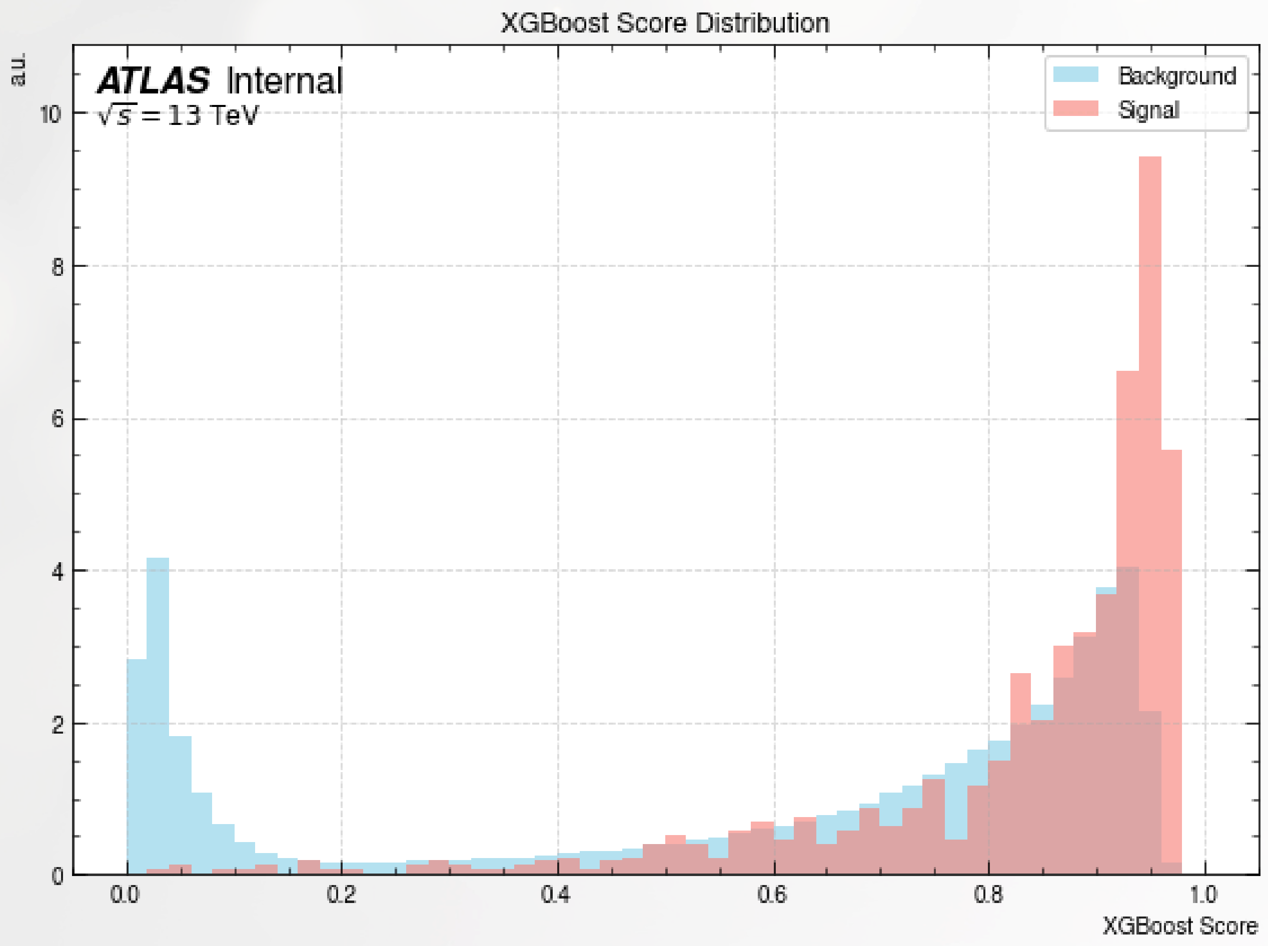
3

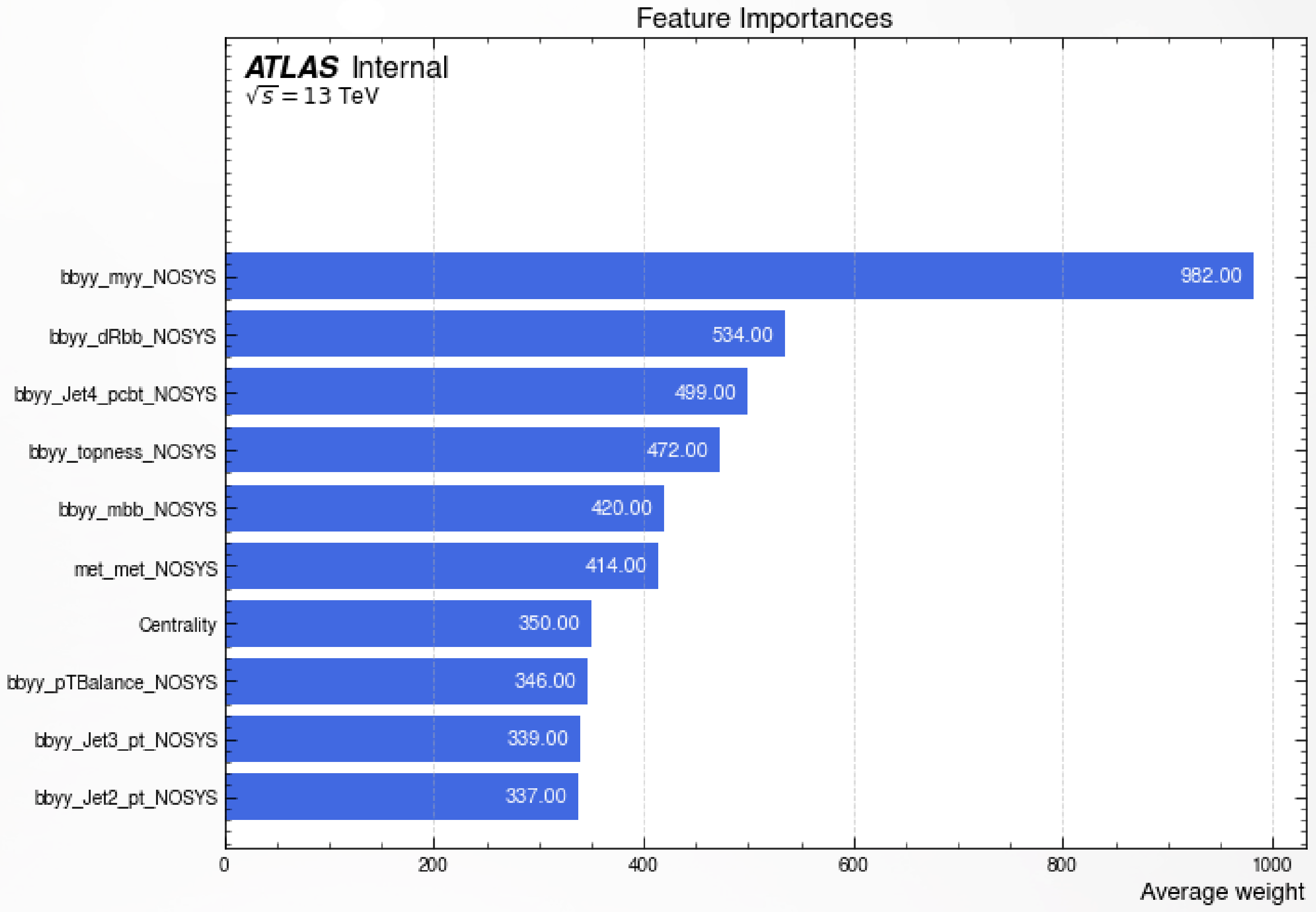
FEATURES

4

3

BOOSTED DECISSION TREE

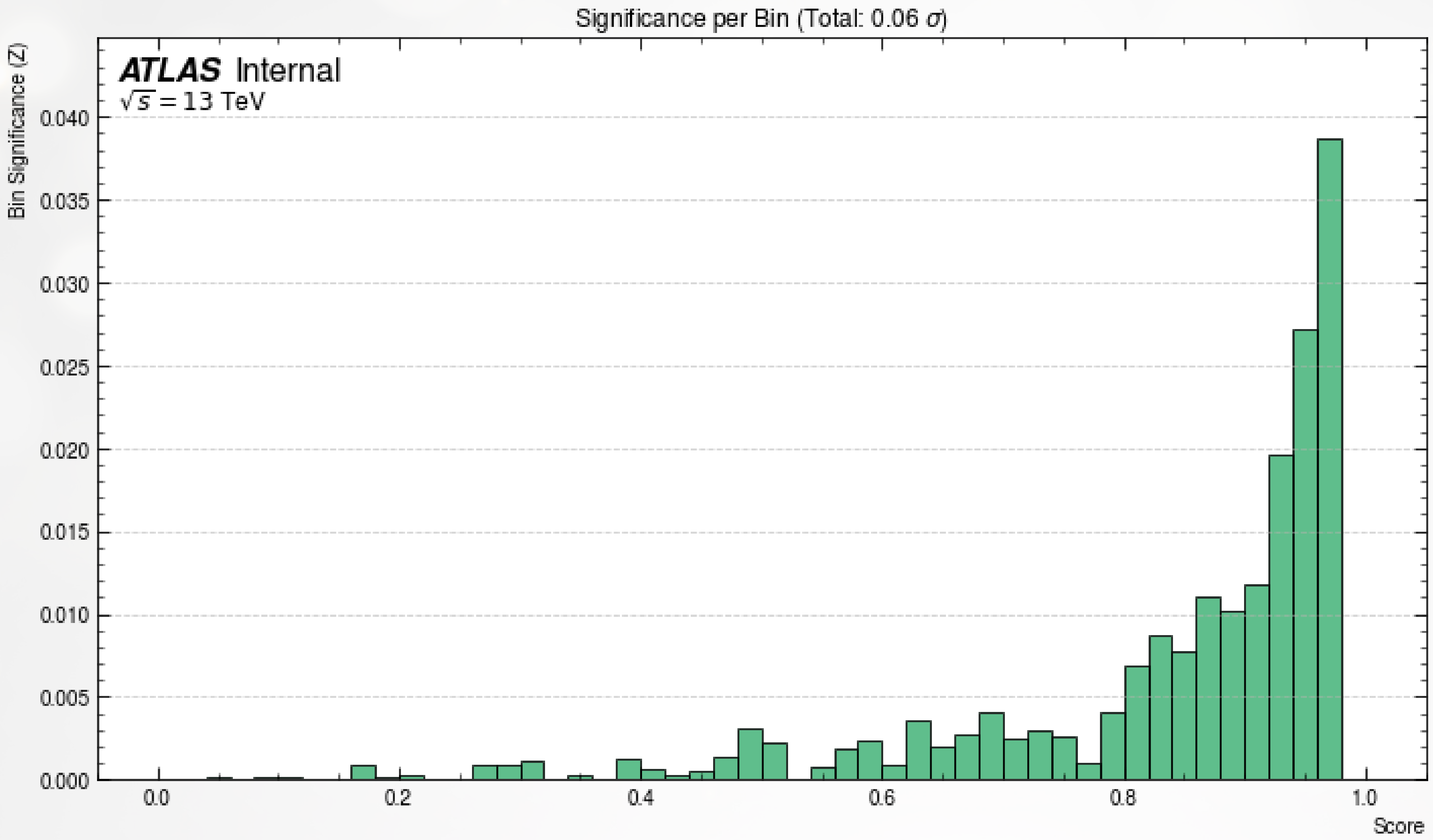




4

4

BINNED SIGNIFICANCE



$$Z_{total} = \sqrt{\sum_i Z_i^2} \text{ , where } Z = \sqrt{2\left((n_s + n_b)\log\left(1 + \frac{n_s}{n_b}\right) - n_s\right)}$$

5

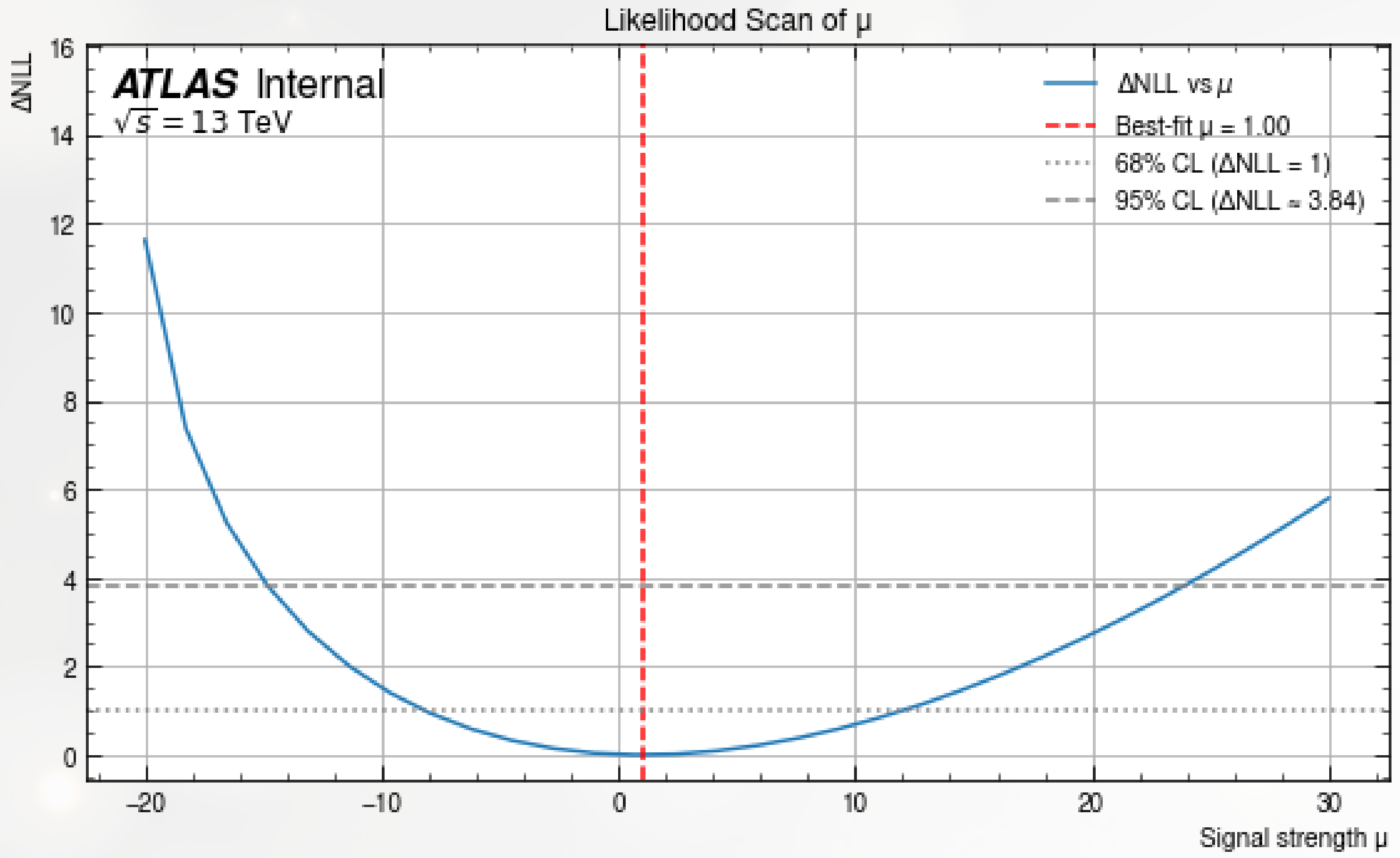
Run 2 Significance	WP = 77%	WP = 85%
1 b-tag	0.09	0.08
2 b-tag	0.05 (0.04)	0.05 (0.06)

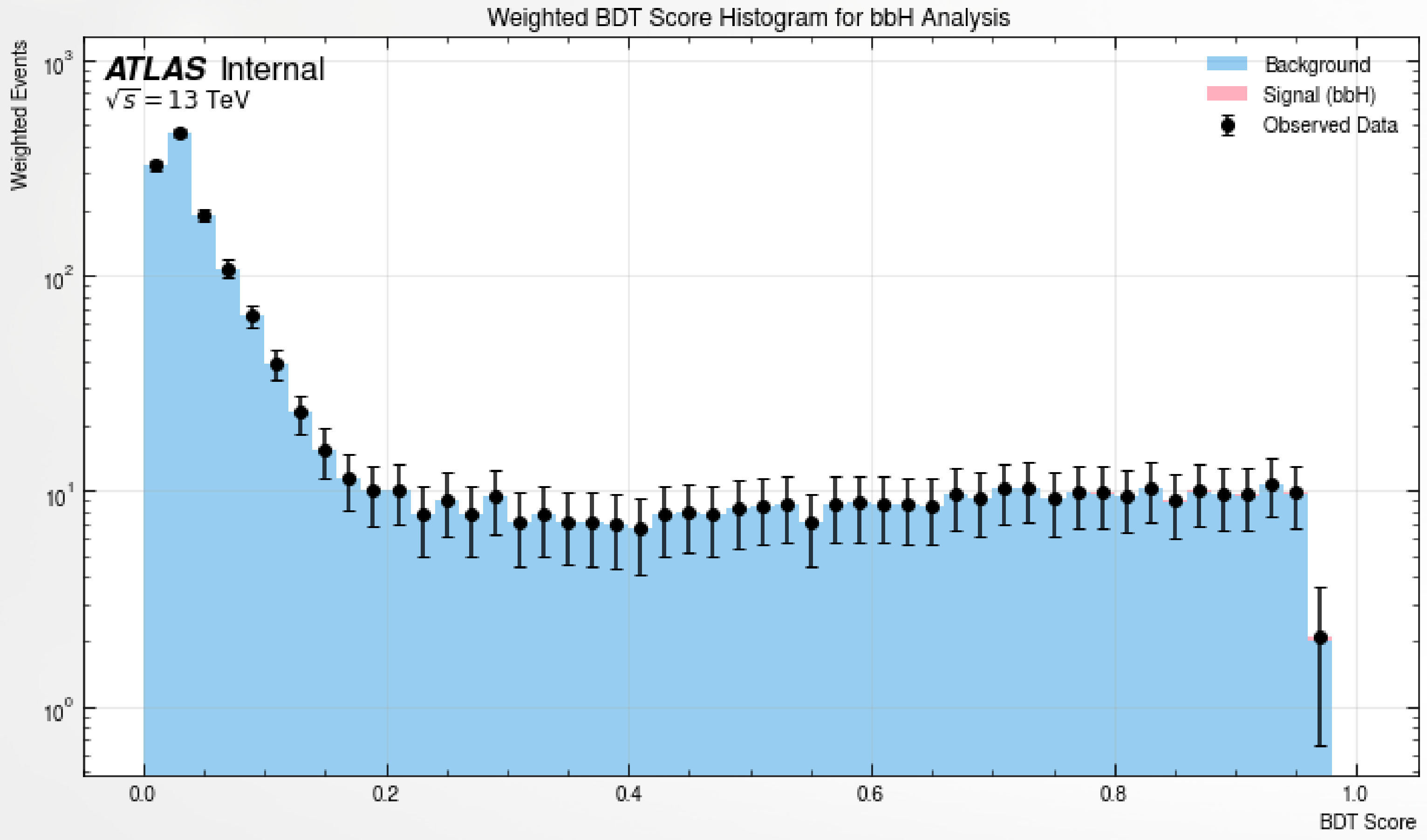
Run 3 Significance	WP = 77%	WP = 85%
1 b-tag	0.08	0.07
2 b-tag	0.02 (0.08)	0.03 (0.06)

4

BINNED SIGNIFICANCE

5





In finding the signal strength (μ), the negative log likehood is minimized

$$NLL = -\log(\mathcal{L})$$

where the likelihood is defined as,

$$\mathcal{L}(\text{data}, \mu) = \prod_i \mathcal{P}(n_i | \mu \cdot s_i + b_i)$$

Test statistic

$$q(\mu) = -2 \ln \left(\frac{\mathcal{L}(\mu)}{\mathcal{L}(\mu_{best})} \right)$$

The 95% region is defined as $q(\mu) \leq 3.84$

Result:

Signal strength (μ): 1.000
95% CL interval: μ ∈ [-14.83, 24.83]

5

STATISTICAL FIT

CONCLUSION

Our current feasibility study shows that with the Run 2/3 acceptance and statistics, the sensitivity is very limited. The main challenges are the small number of signal events, significant kinematic overlap with resonant backgrounds, and detector geometry

Overall, while the current results show the difficulty of the measurement with present datasets, they also point to clear directions for improvement at the HL-LHC

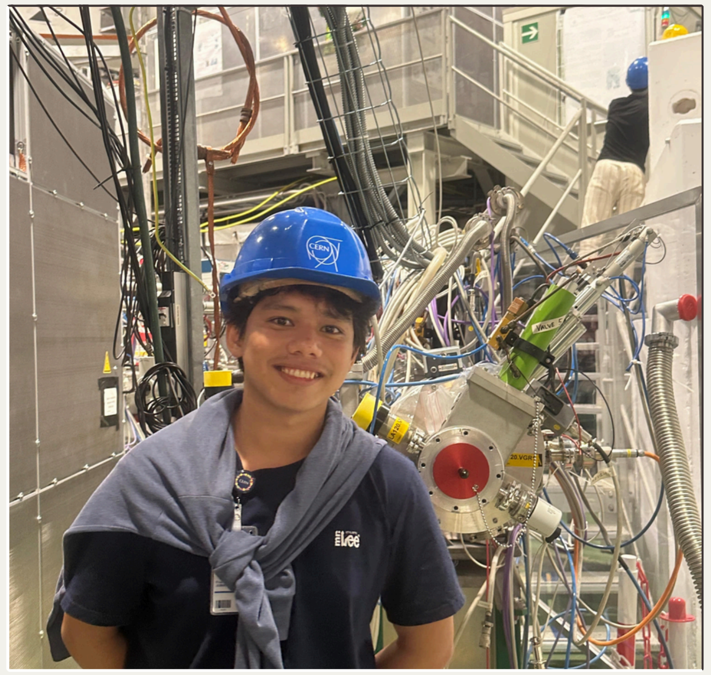
NEXT STEPS

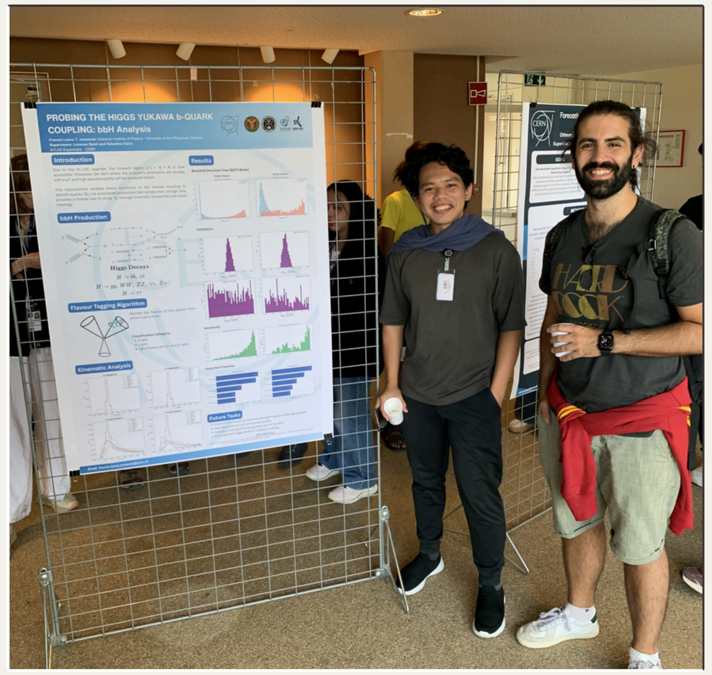
- HL-LHC conditions will give more statistics and extended η coverage
- Better signal–background separation using additional discriminating features and more advanced machine learning
- Improvements in forward b-tagging performance will also help
- A statistical fit on on data will allow us to set meaningful upper limits on the signal strength μ

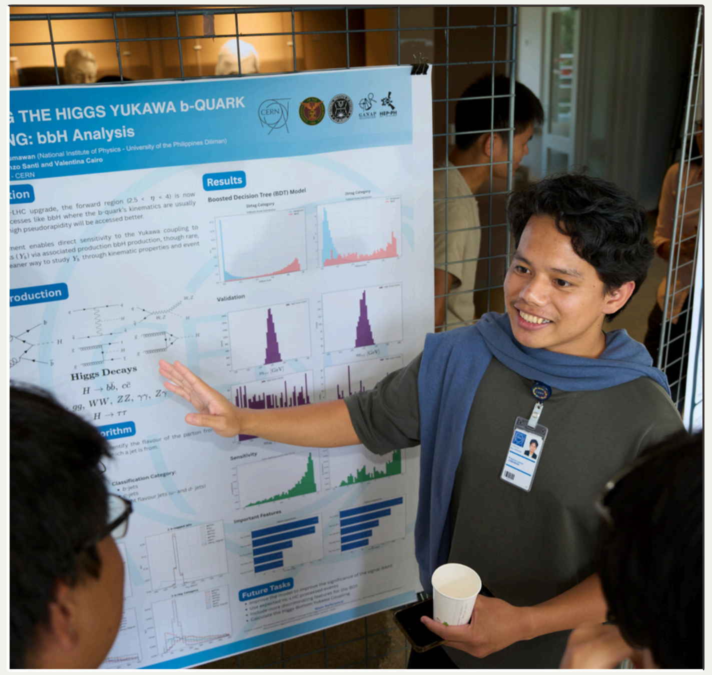
THANK YOU

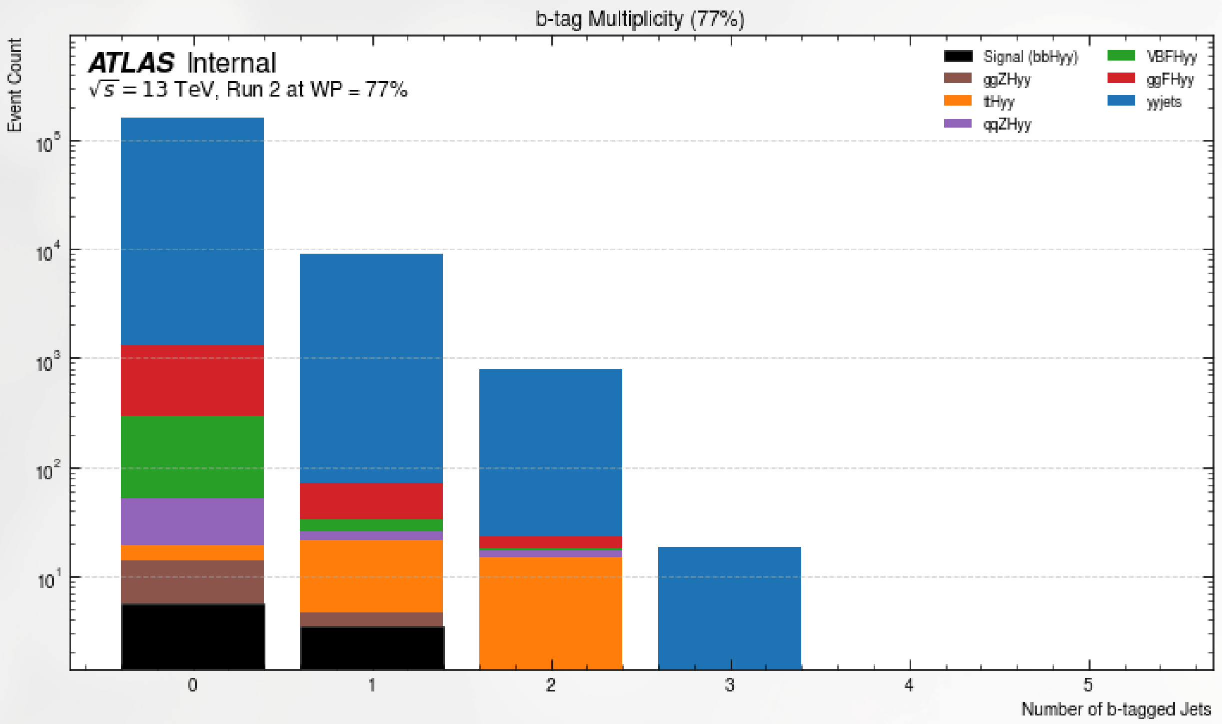
THANK YOU

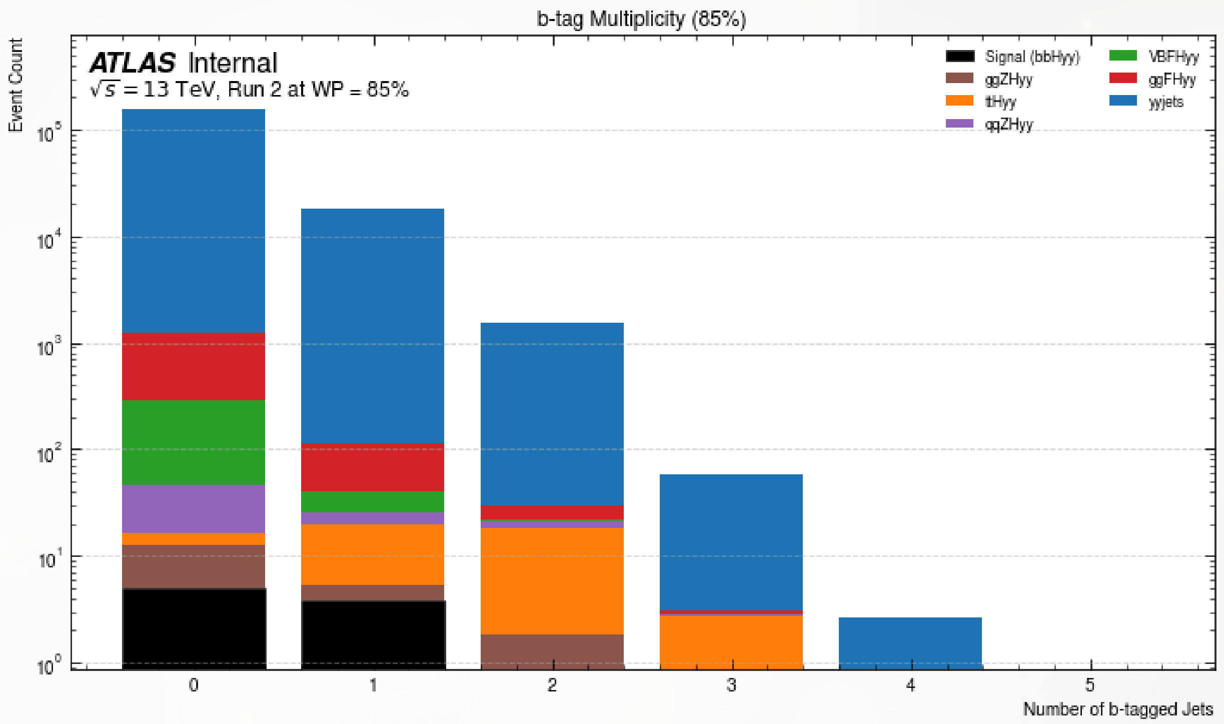
SUMMER STUDENT PROGRAM 2025

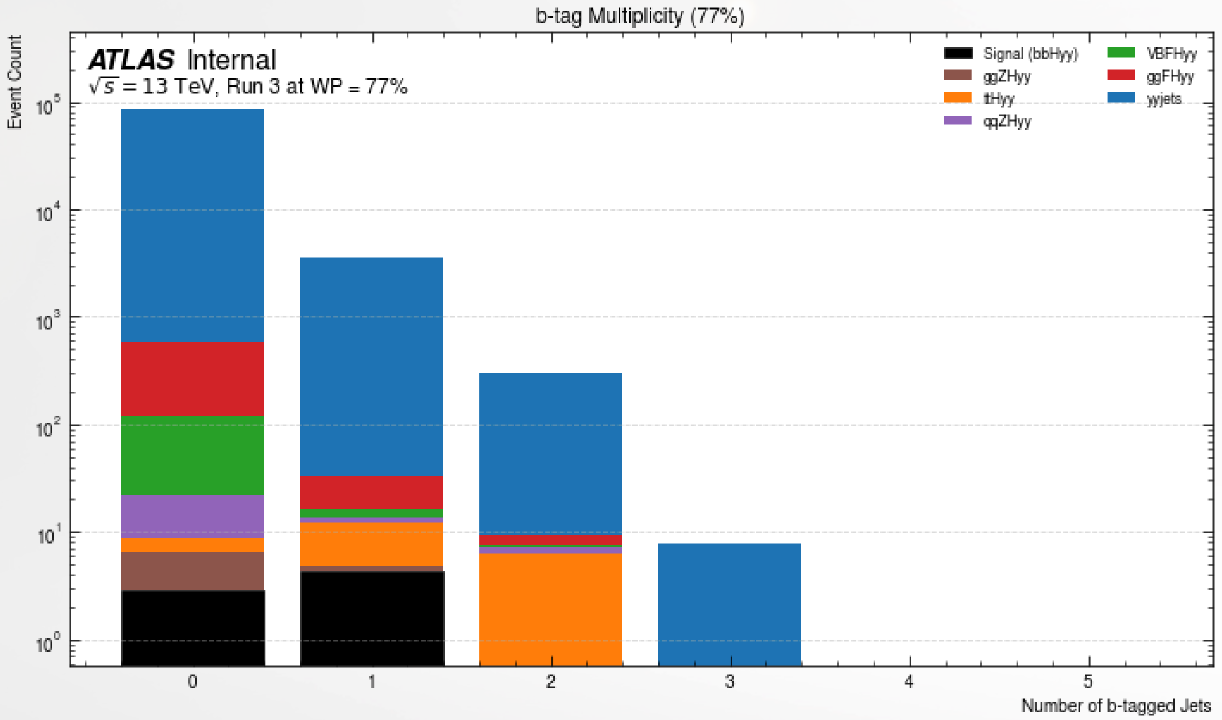


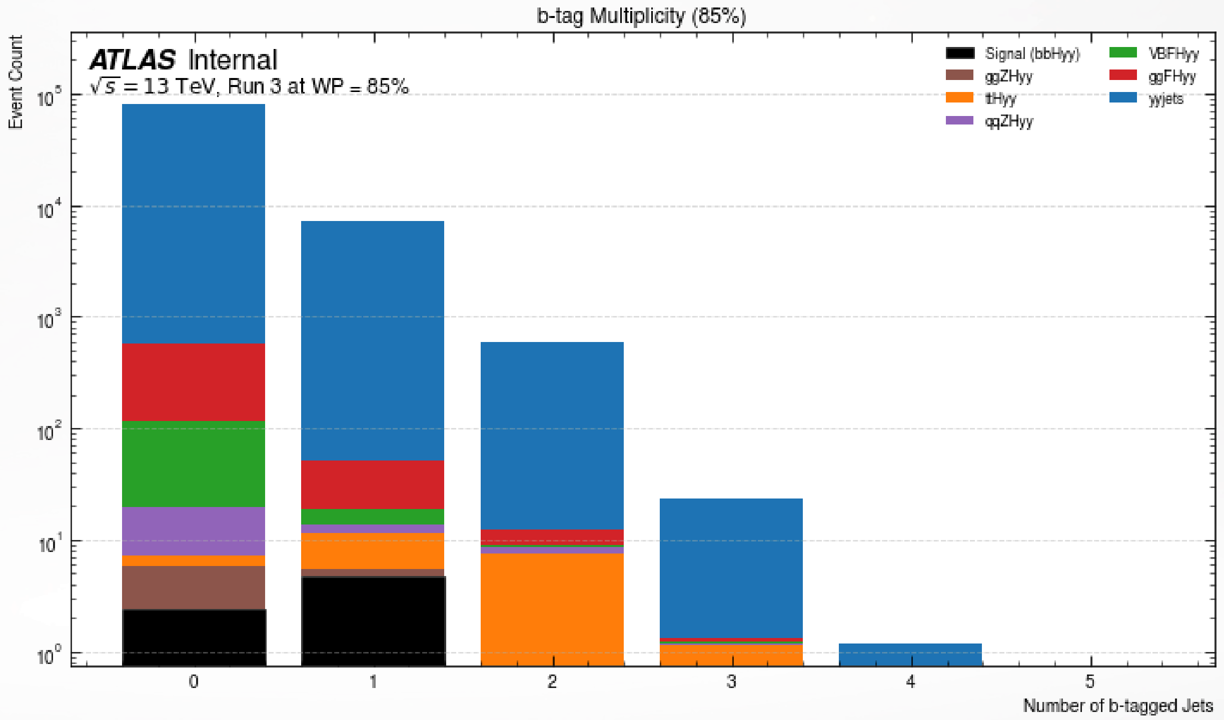












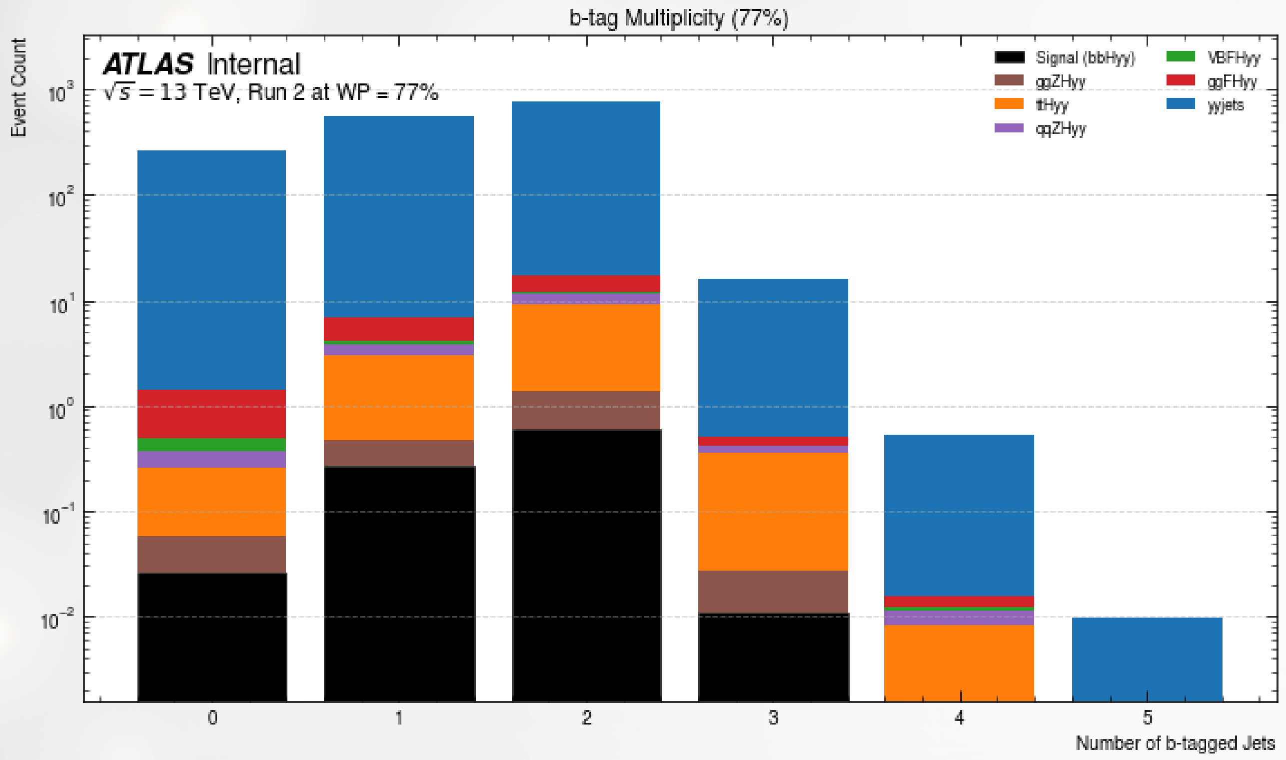
1

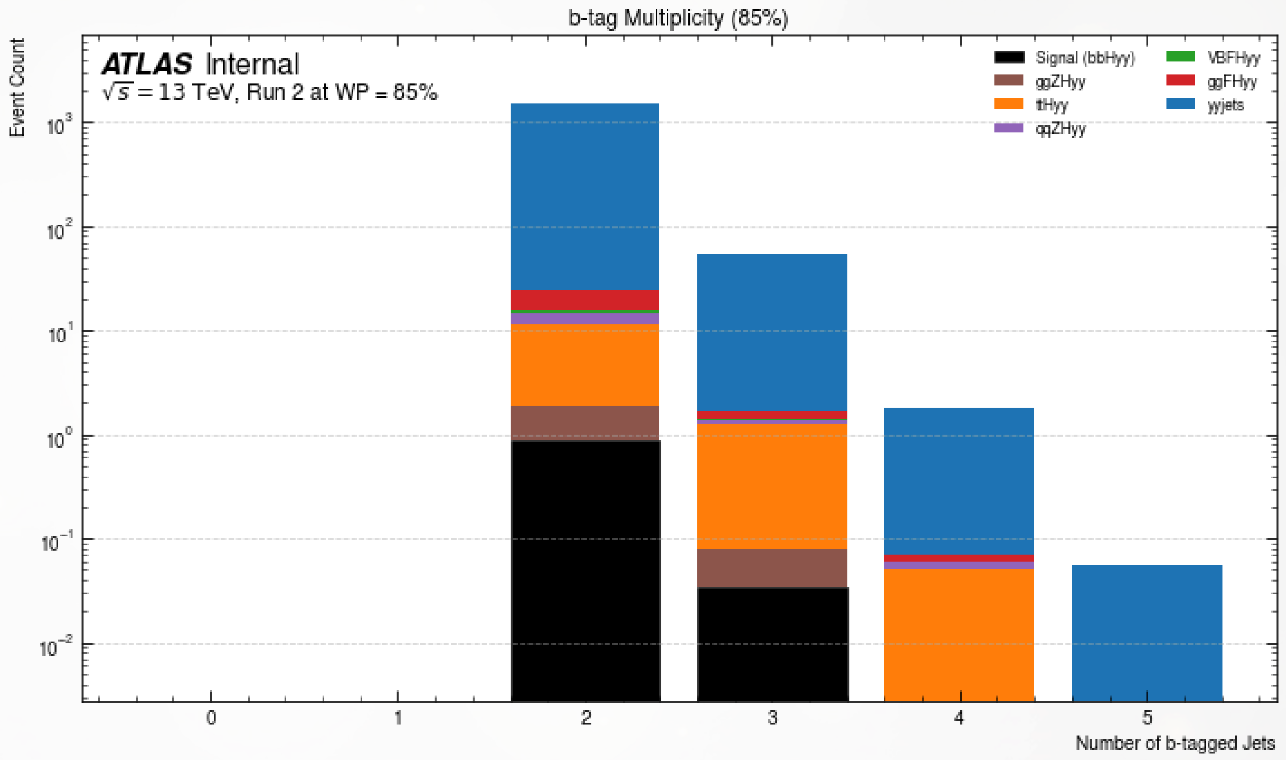
KINEMATIC ANALYSIS

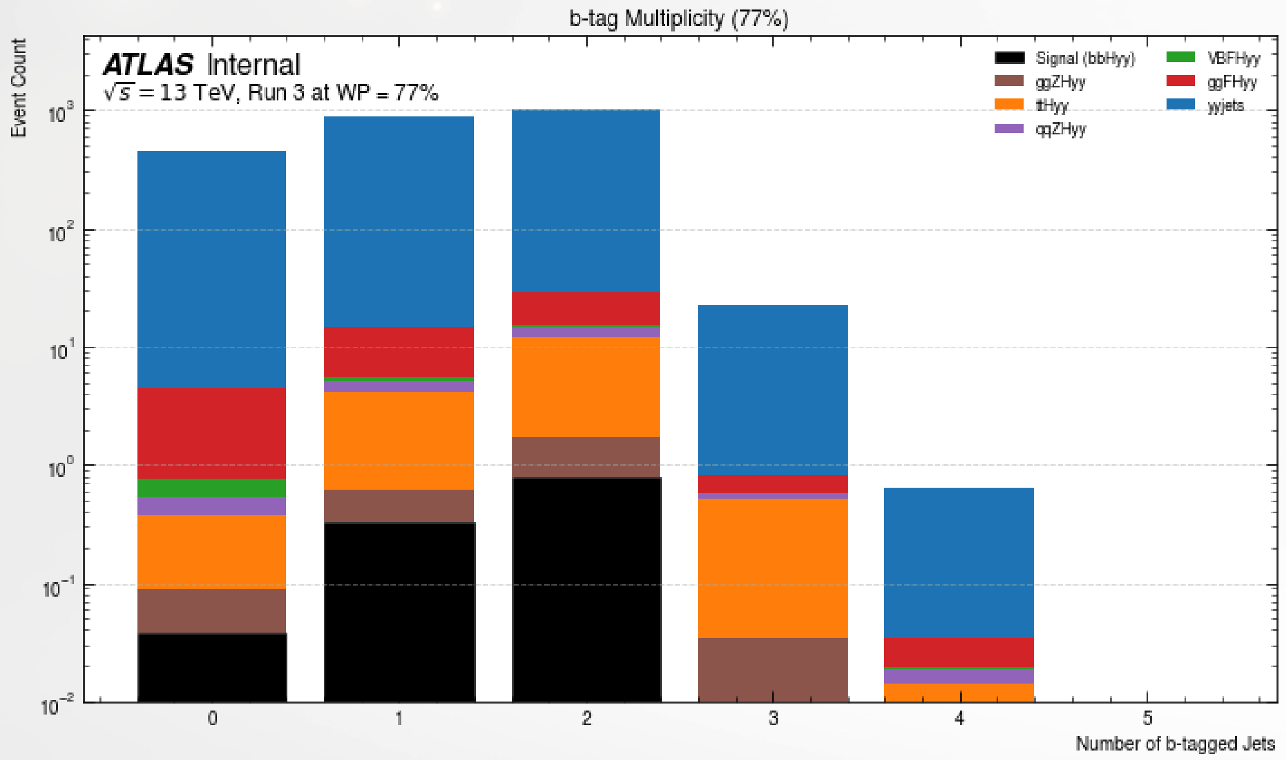
2

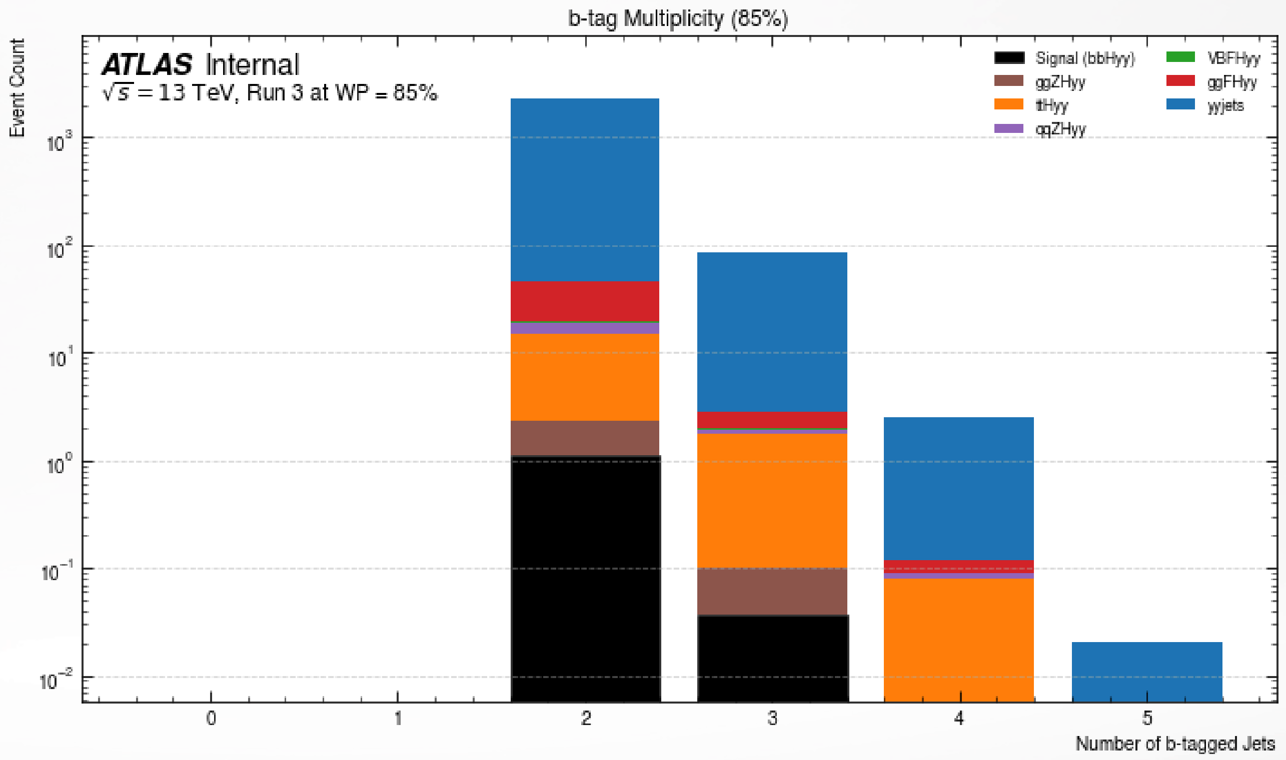
KINEMATIC ANALYSIS

1



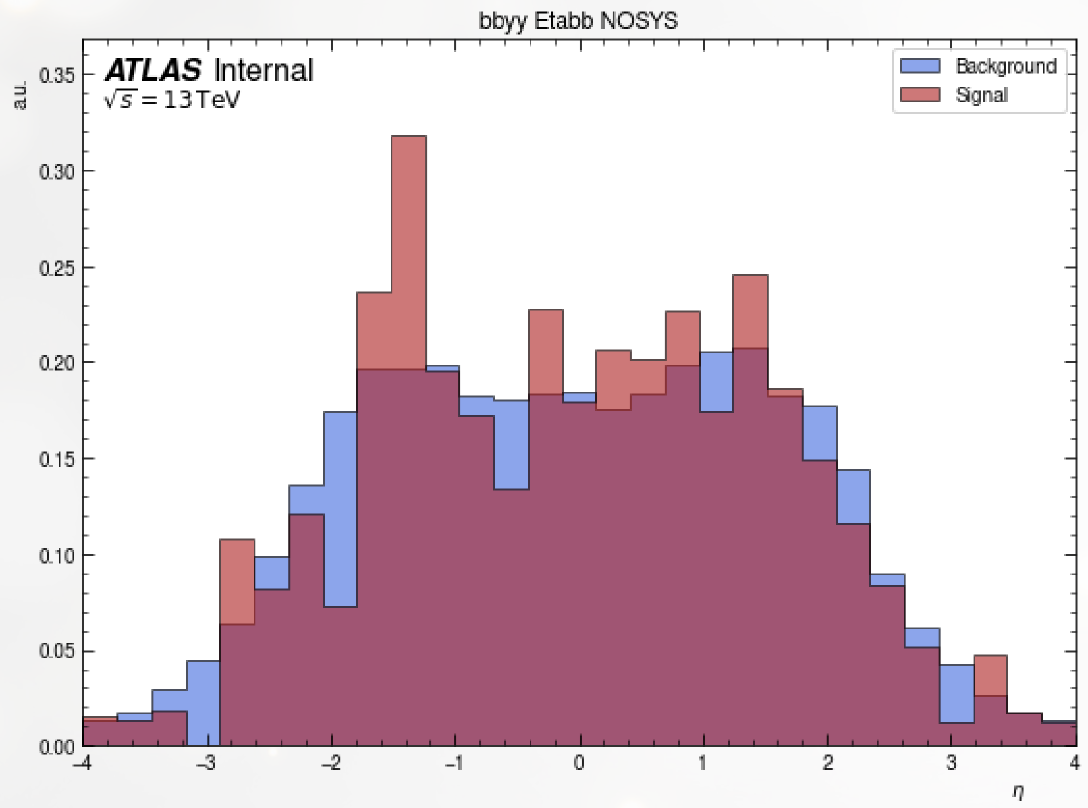


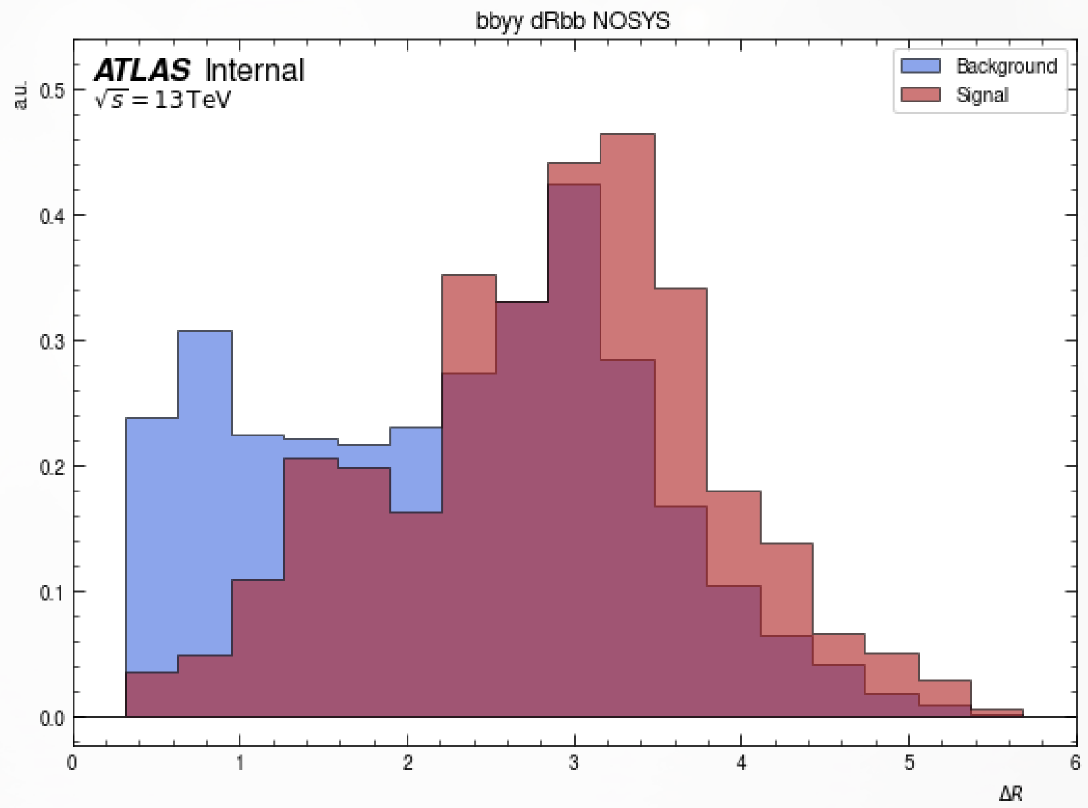


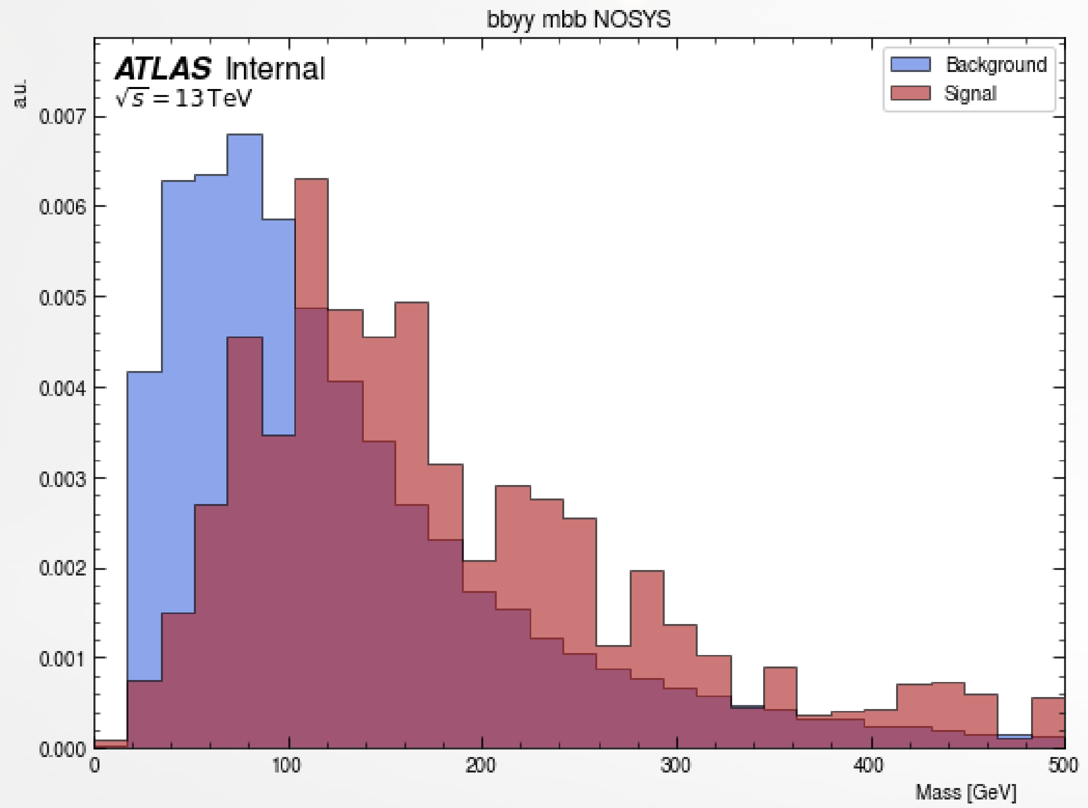


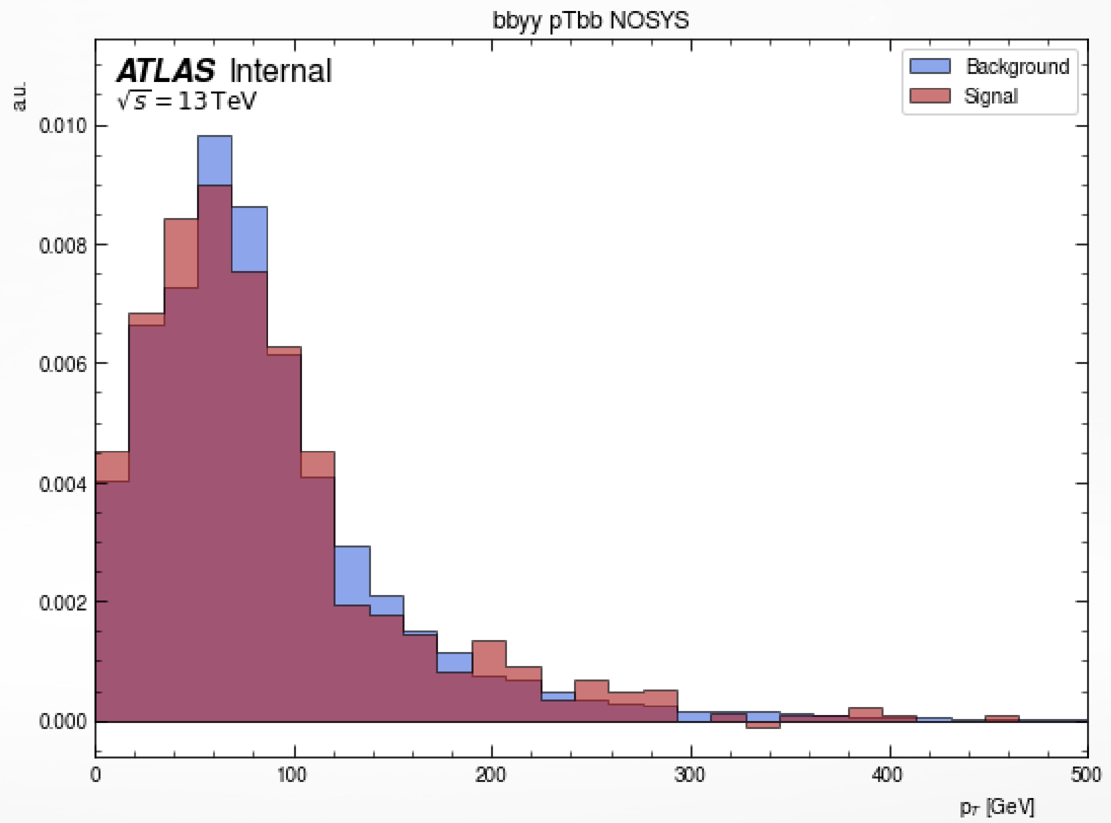
2

bb-System







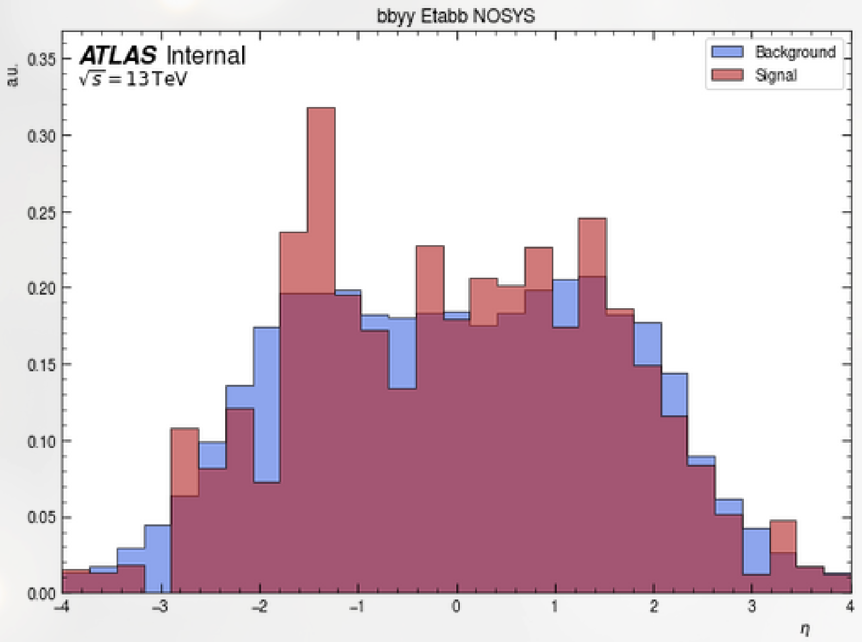


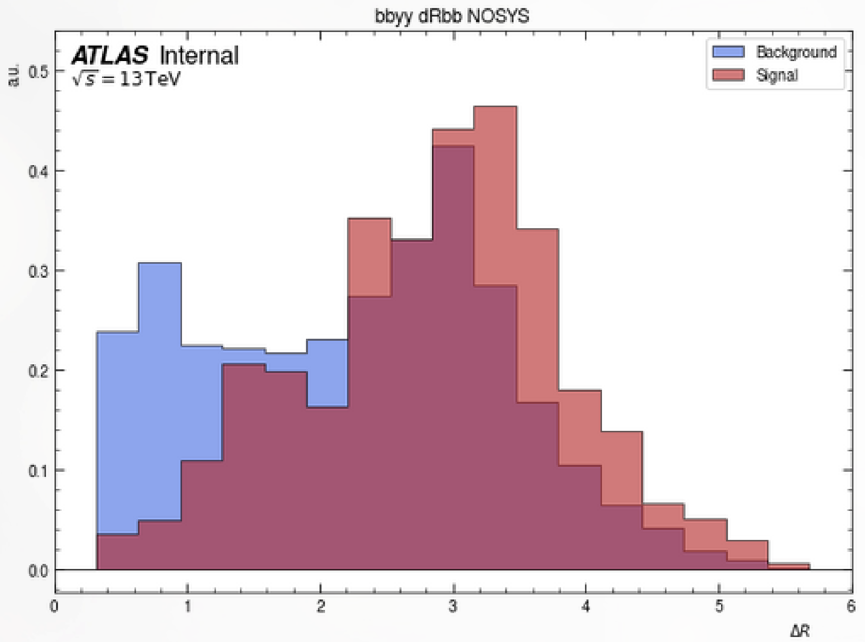
1

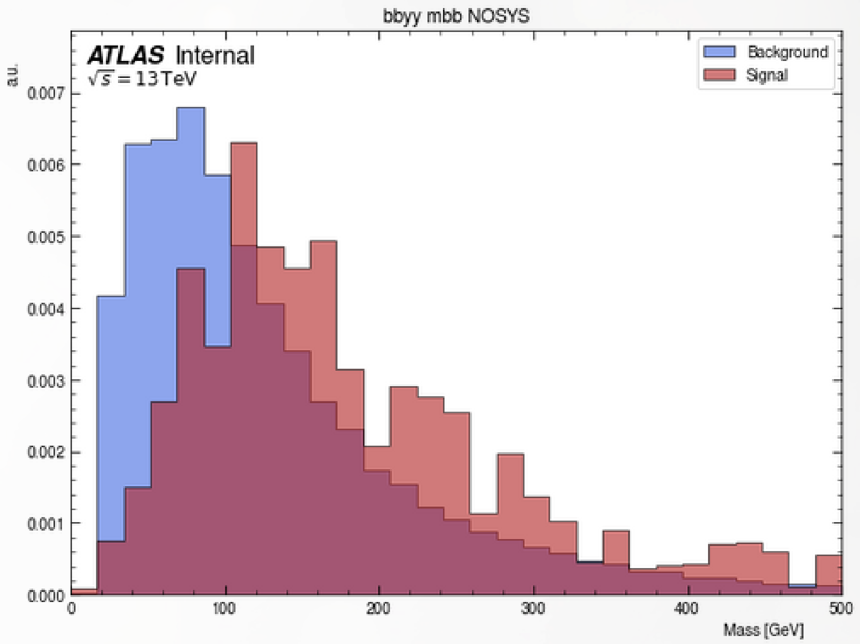
KINEMATIC ANALYSIS

2

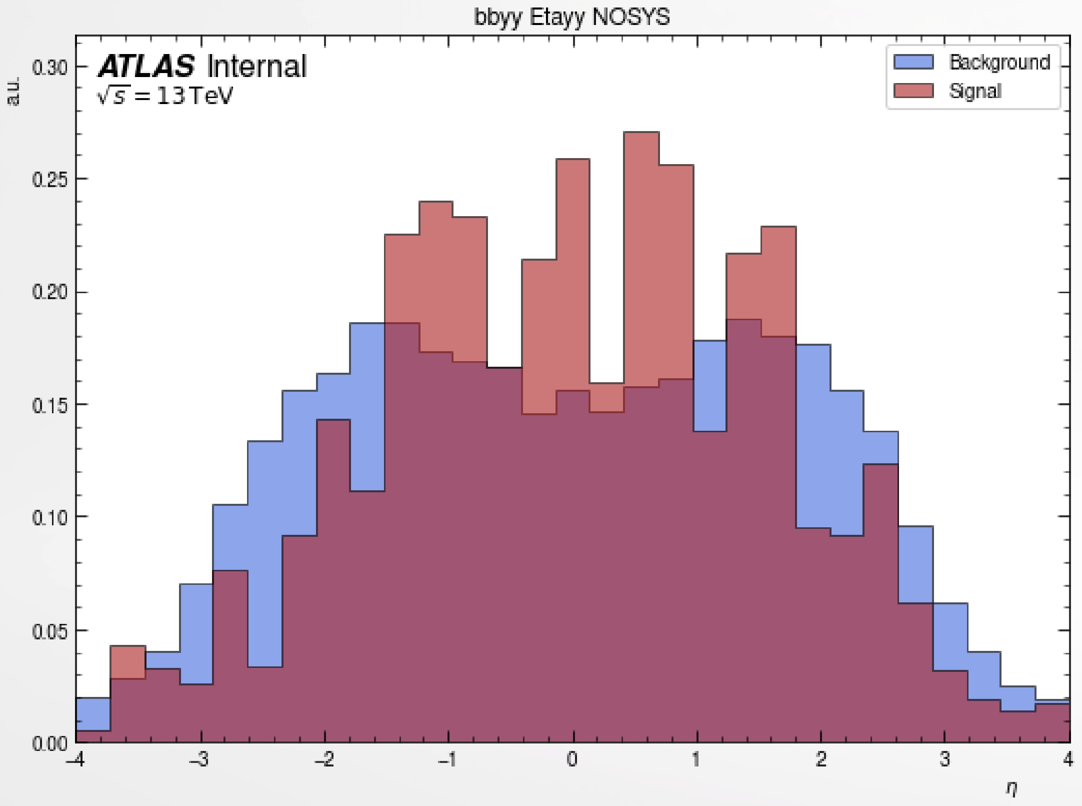
bb-System

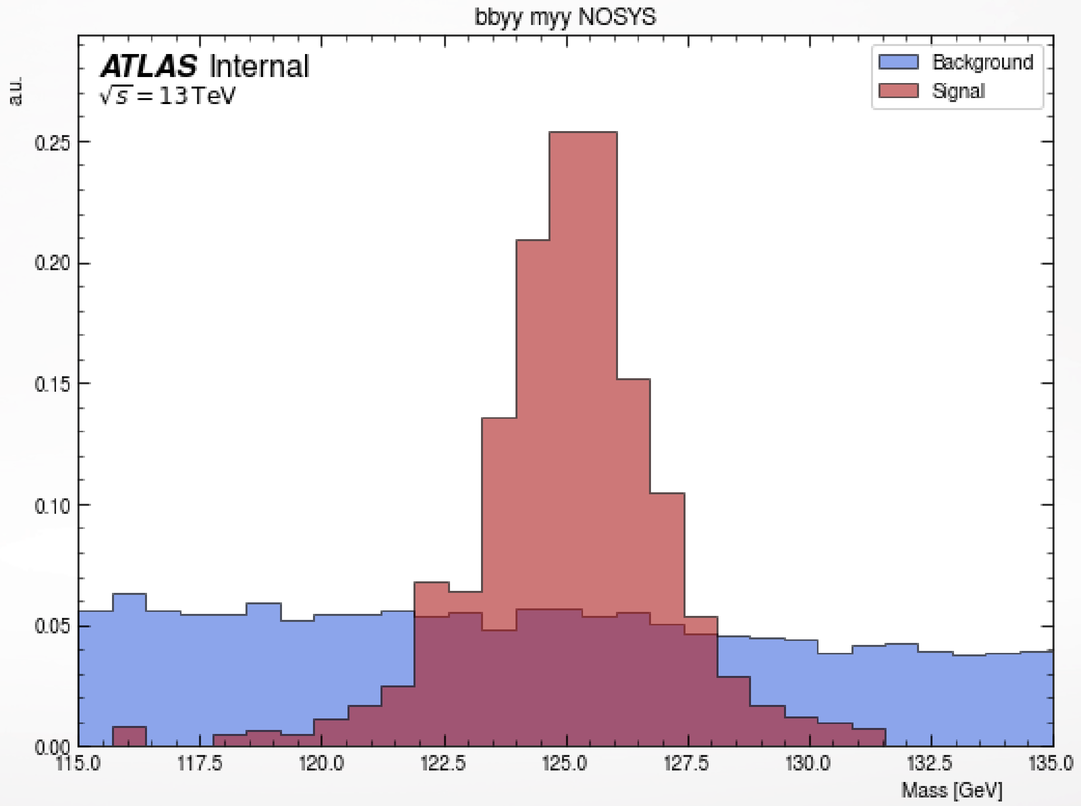






yy-System



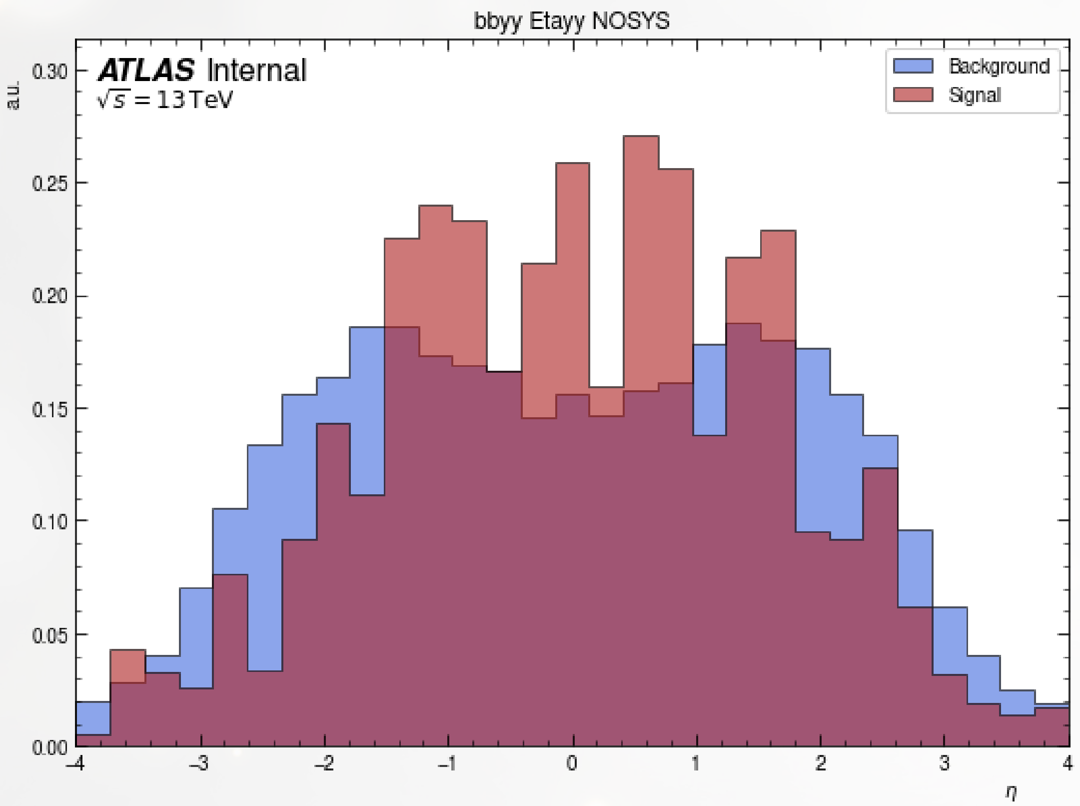


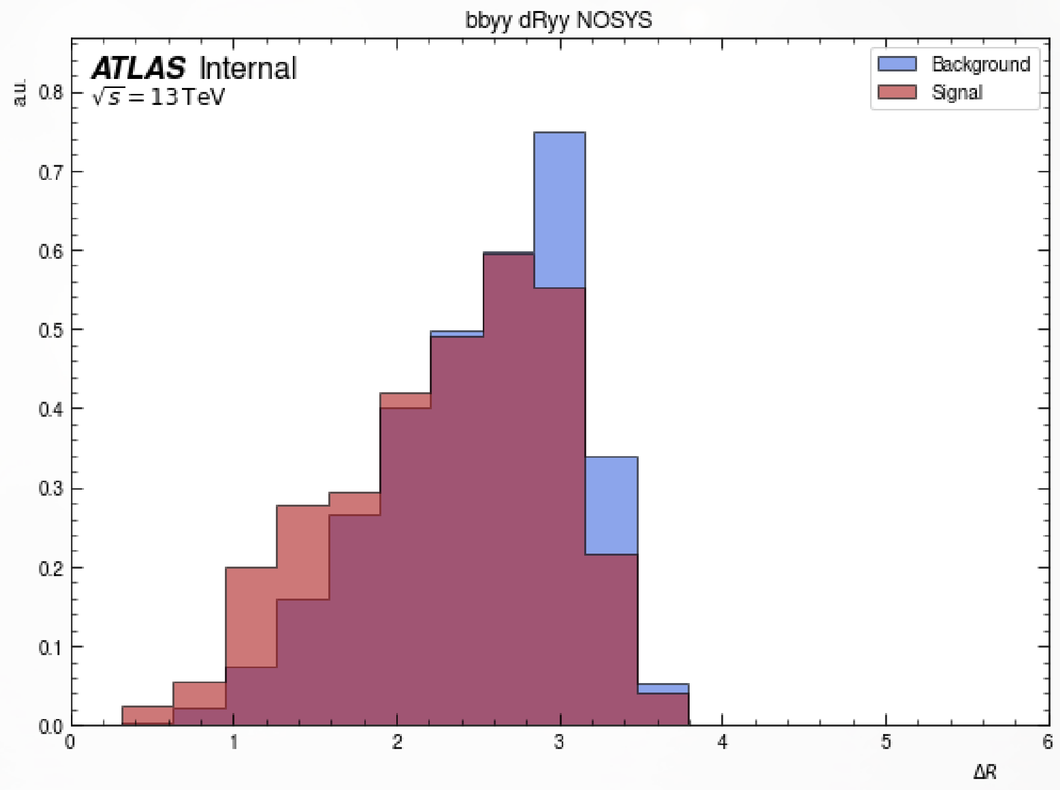
1

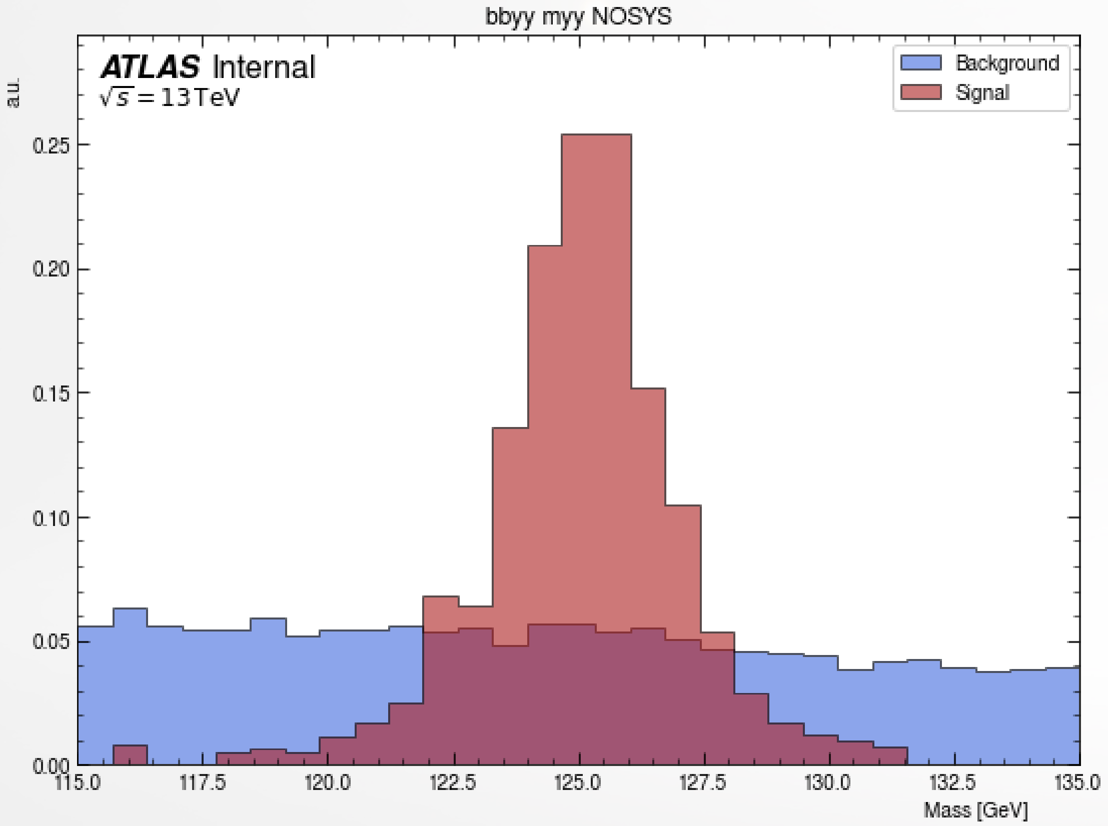
KINEMATIC ANALYSIS

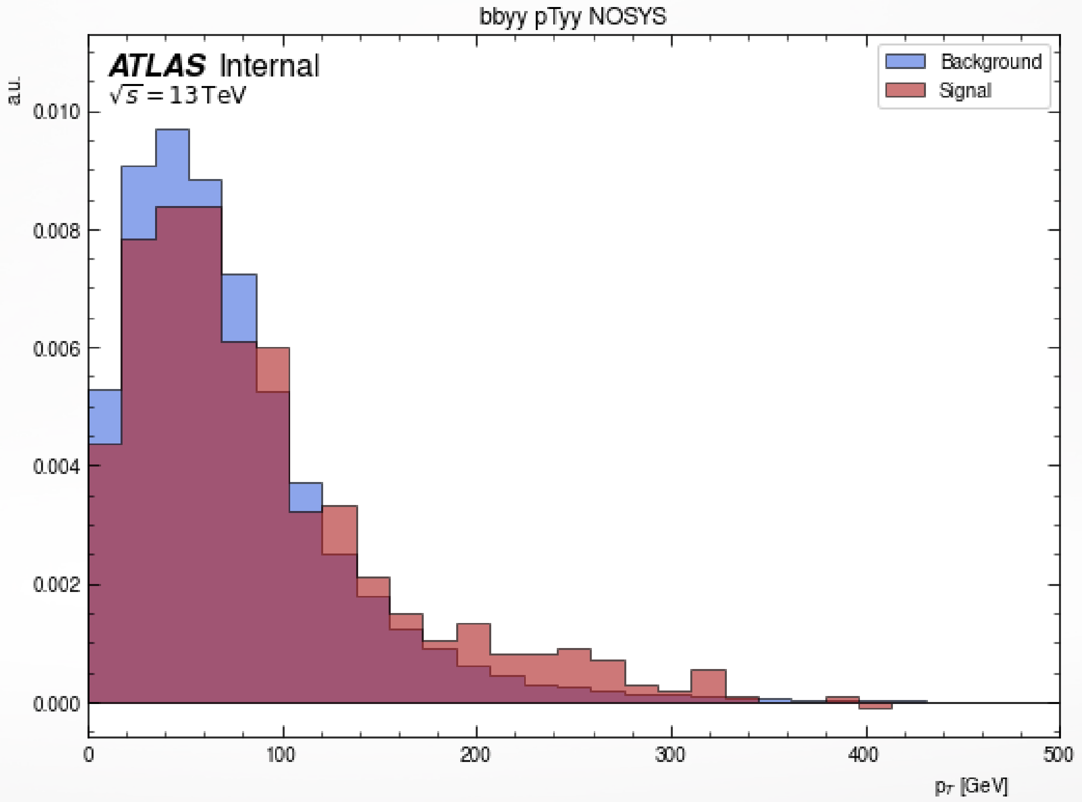
2

yy-System







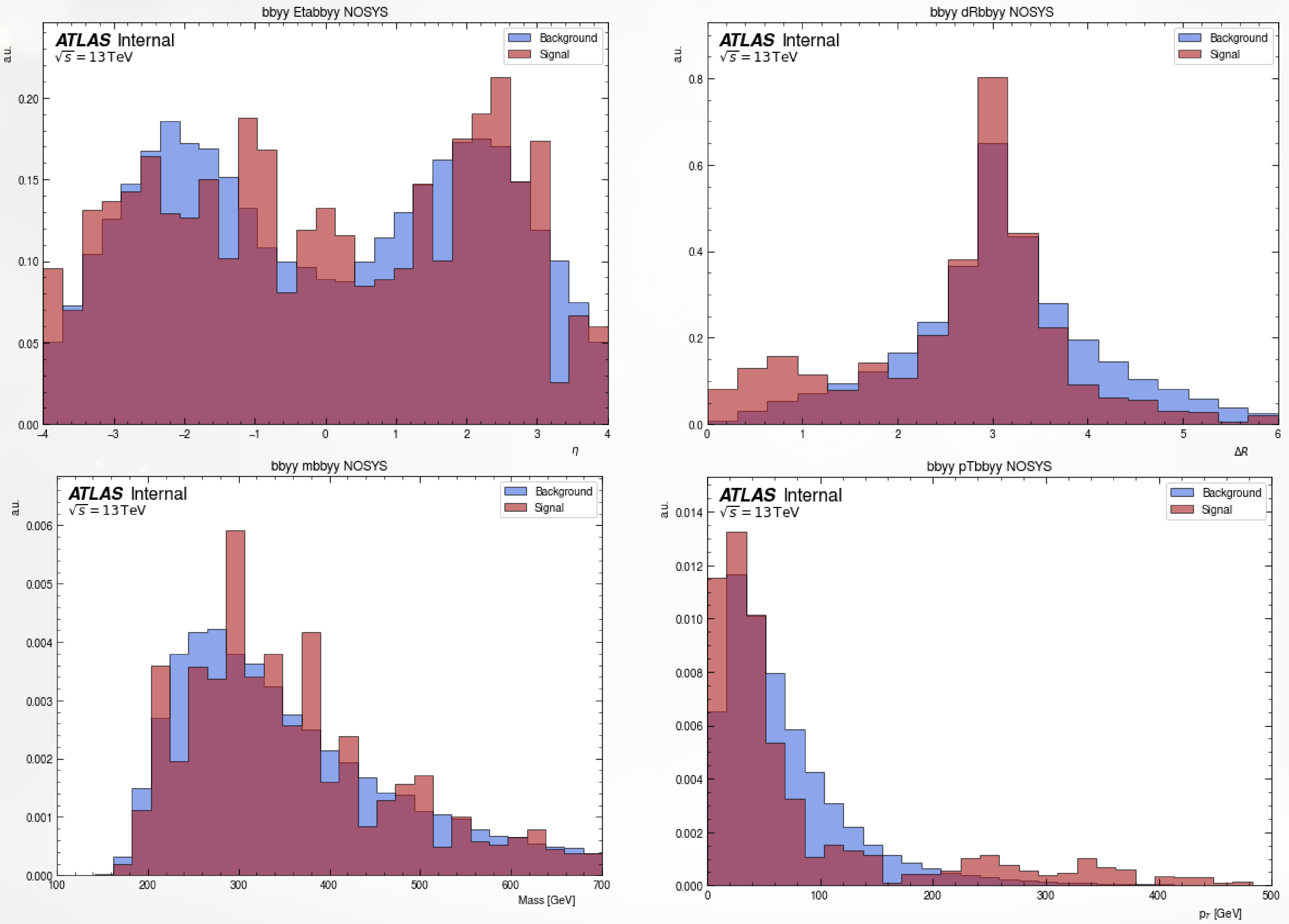


1

KINEMATIC ANALYSIS

2

bbyy-System

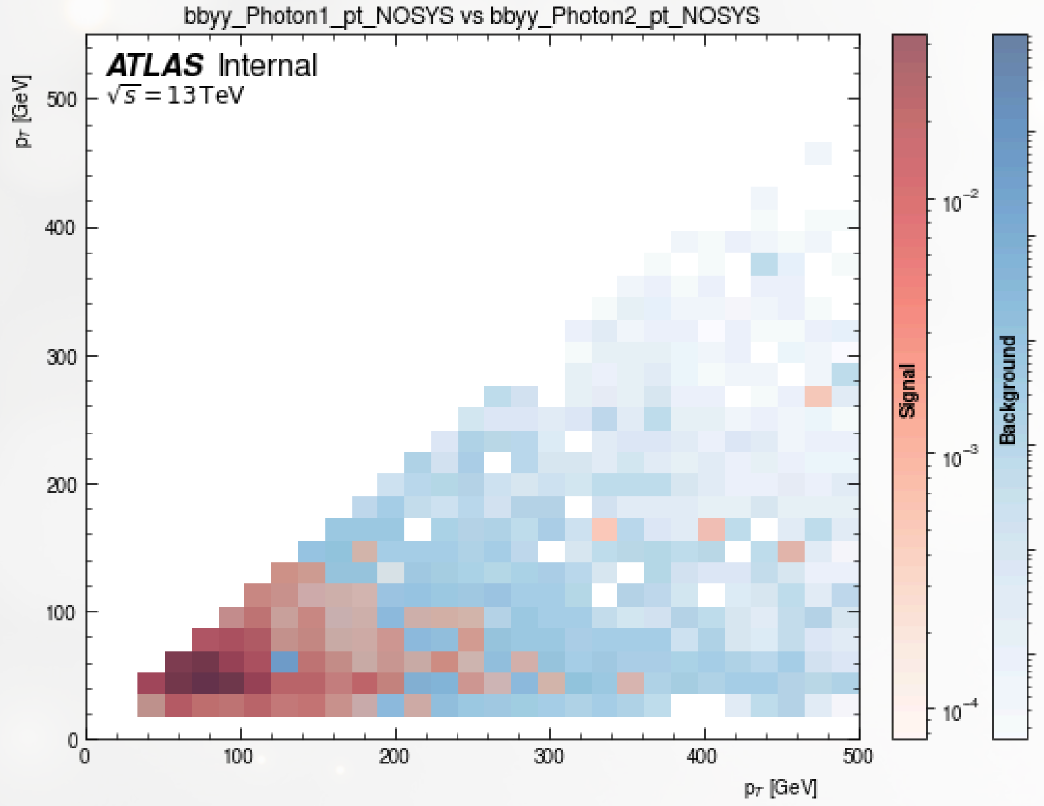


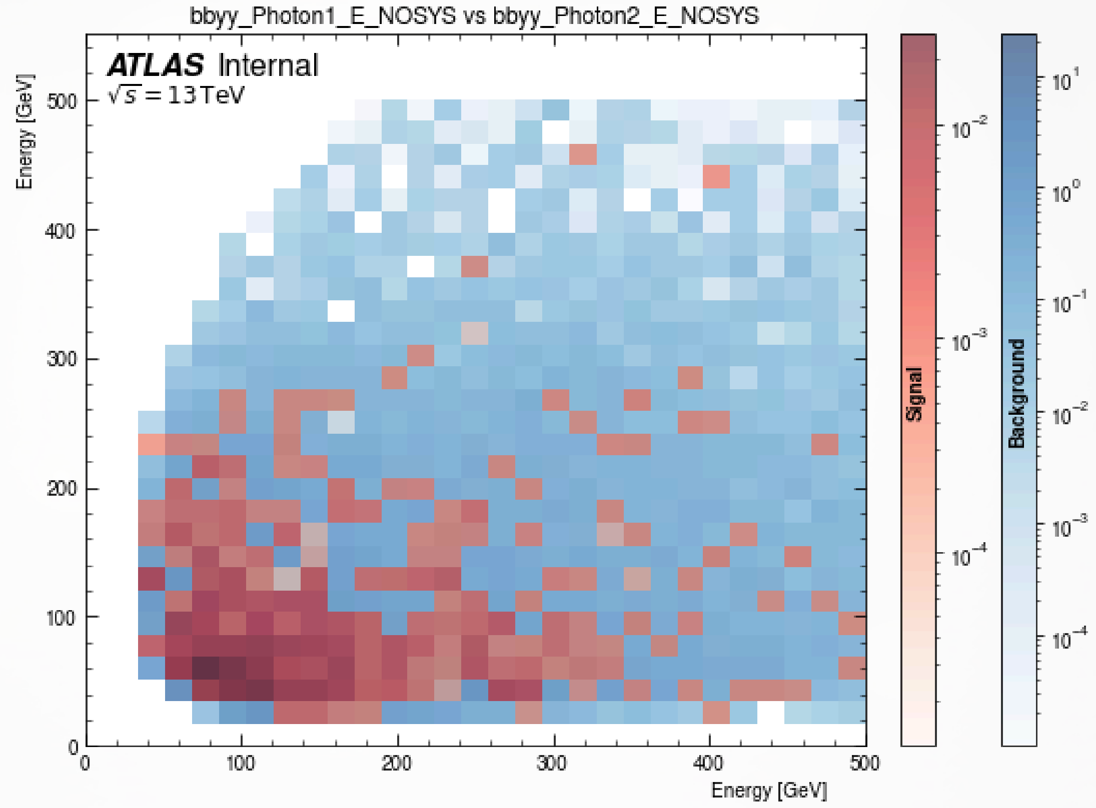
1

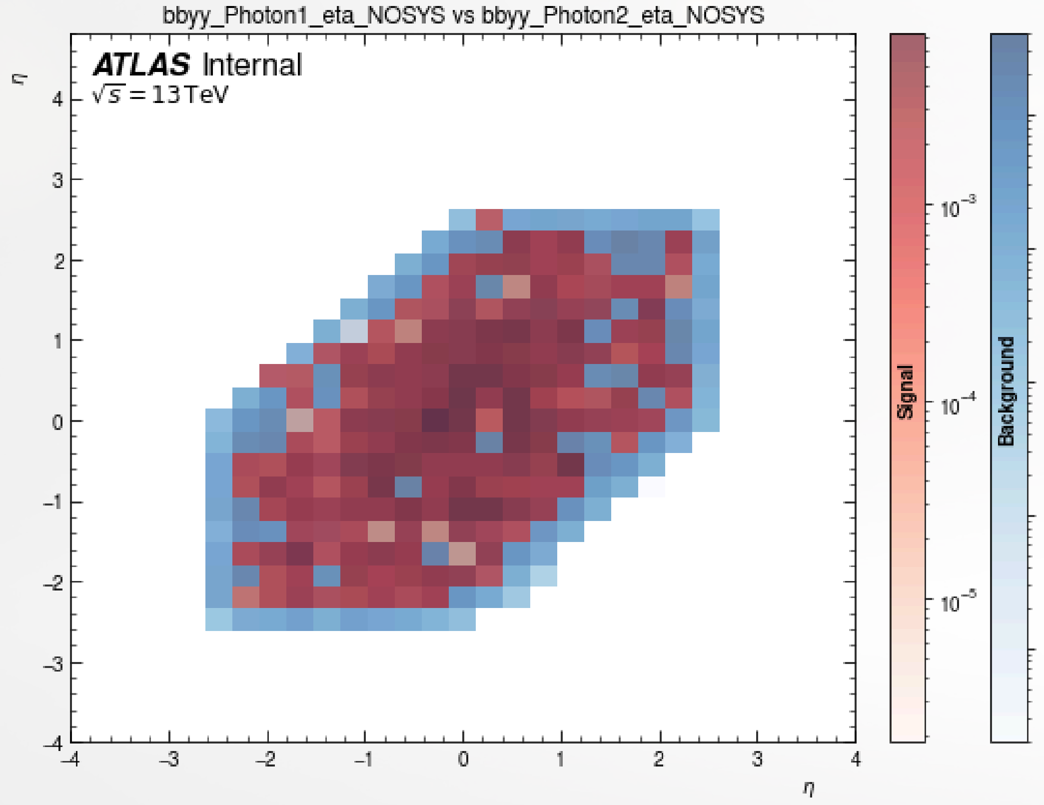
KINEMATIC ANALYSIS

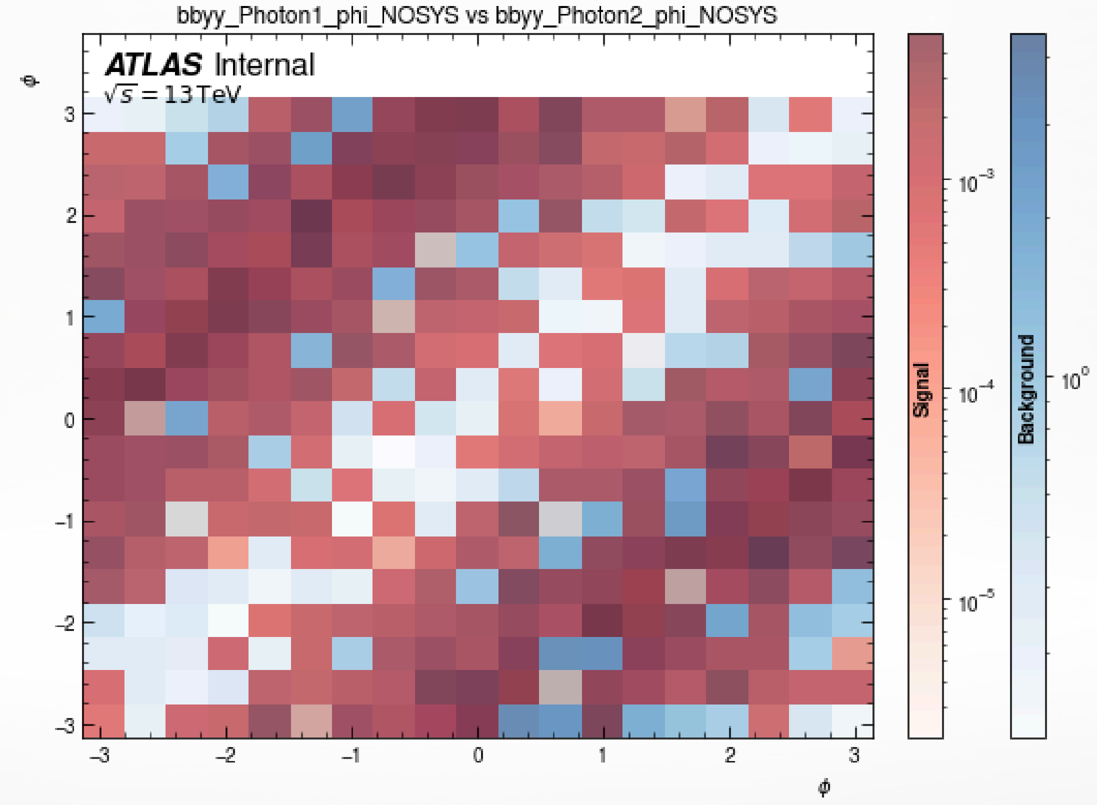
2

photon candidates







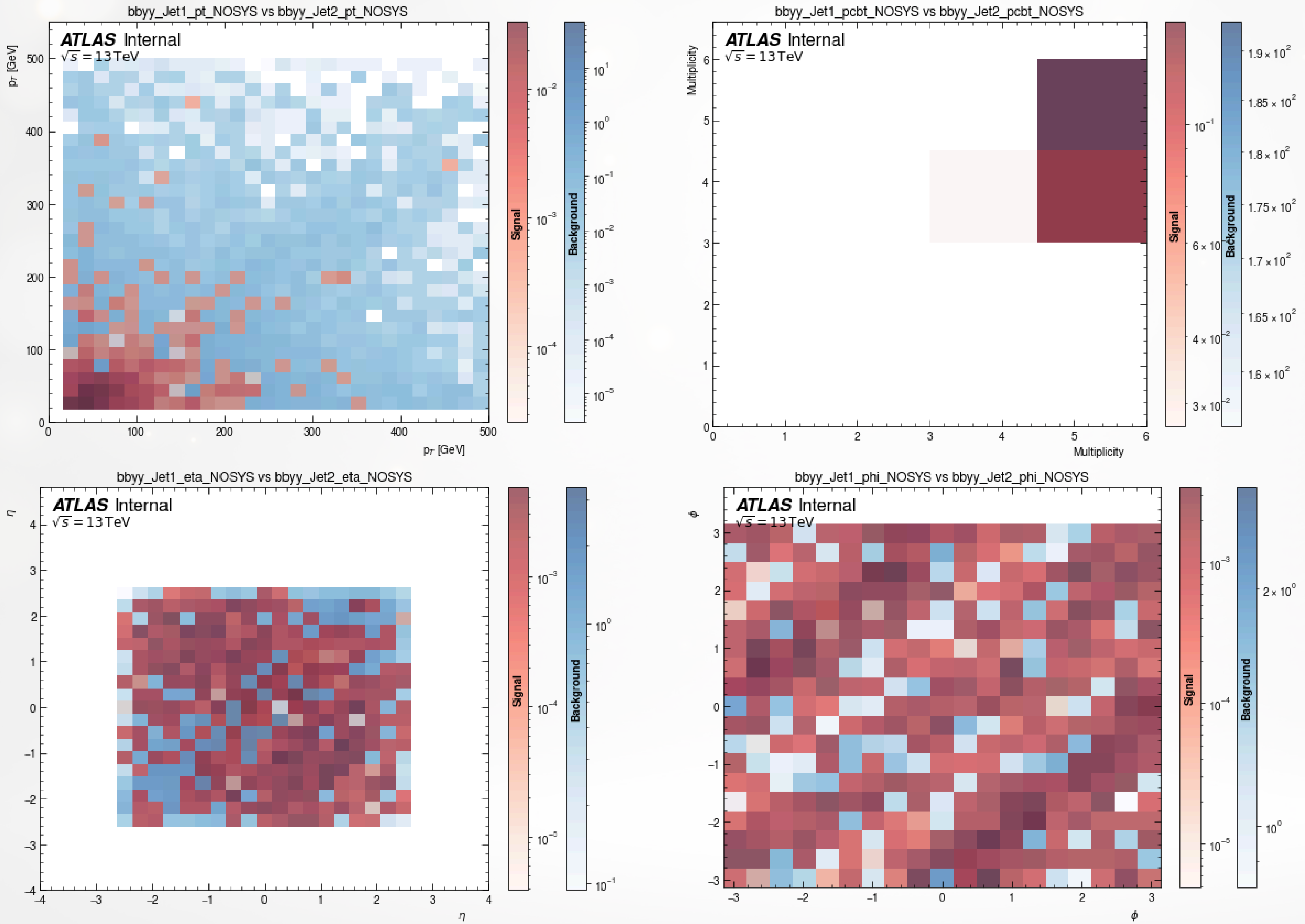


1

KINEMATIC ANALYSIS

2

leading jets (b-jet candidates)

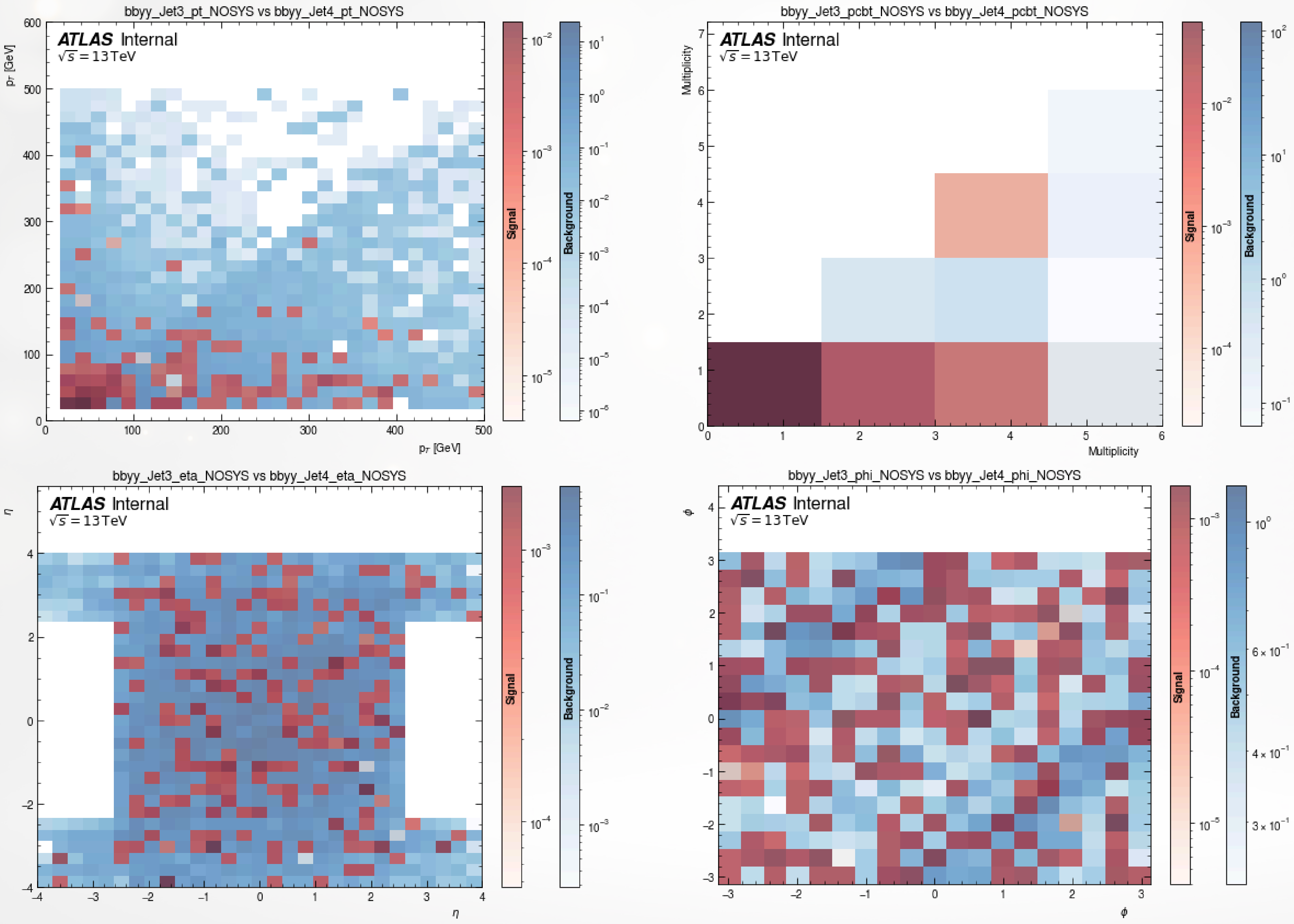


1

KINEMATIC ANALYSIS

2

subleading jets (Jet 3 and Jet 4)

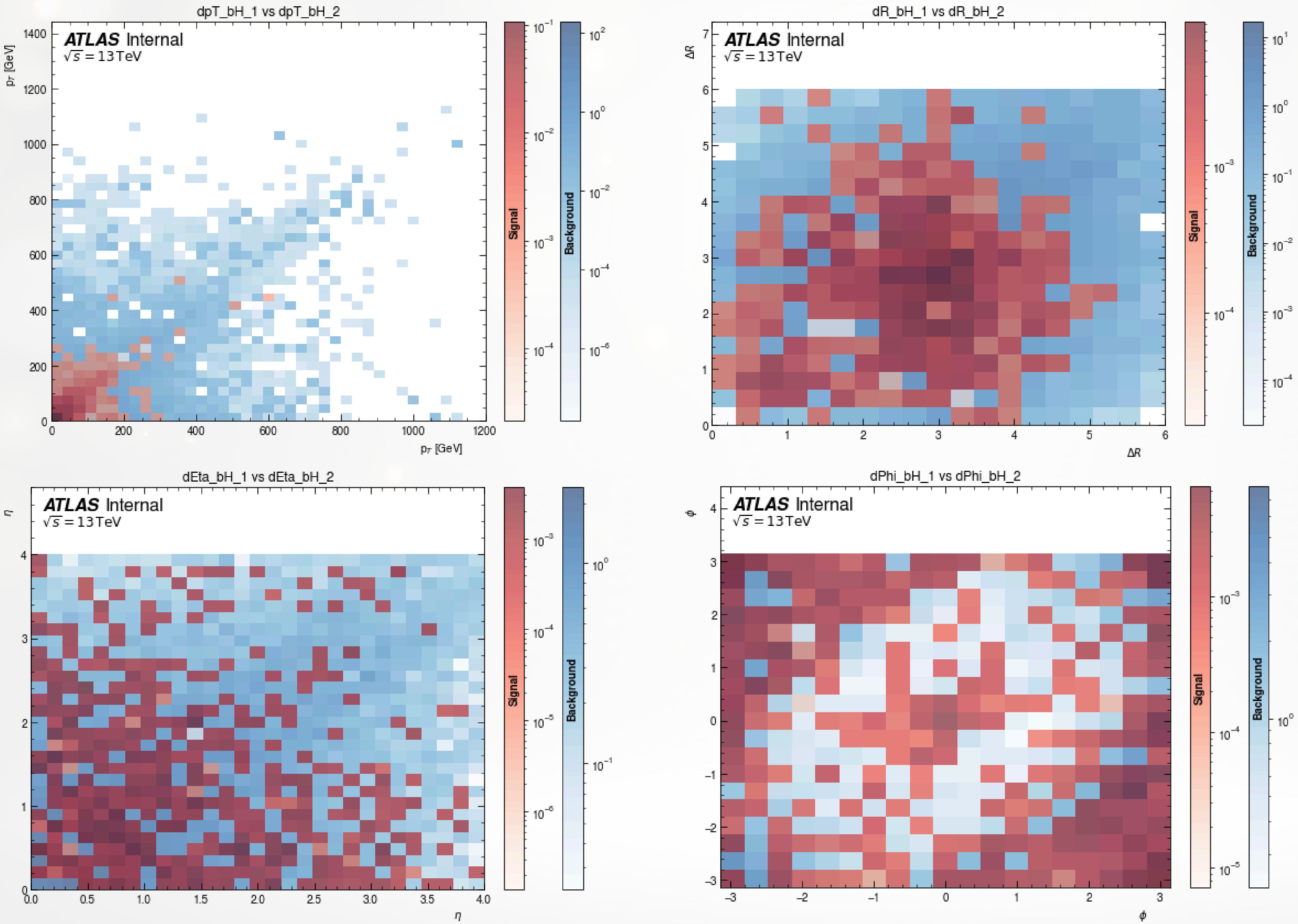


1

KINEMATIC ANALYSIS

2

bH System

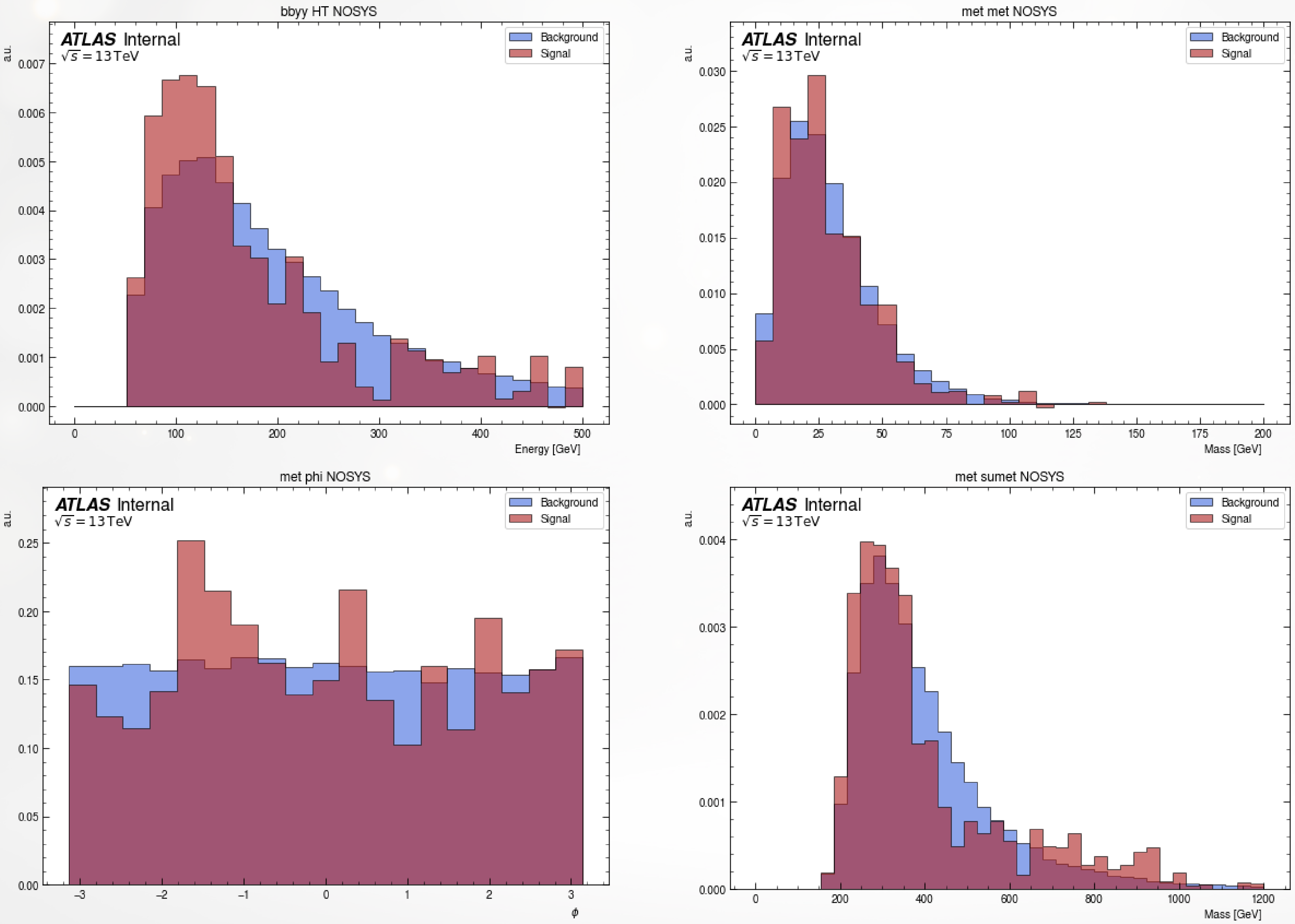


1

KINEMATIC ANALYSIS

2

Other variables

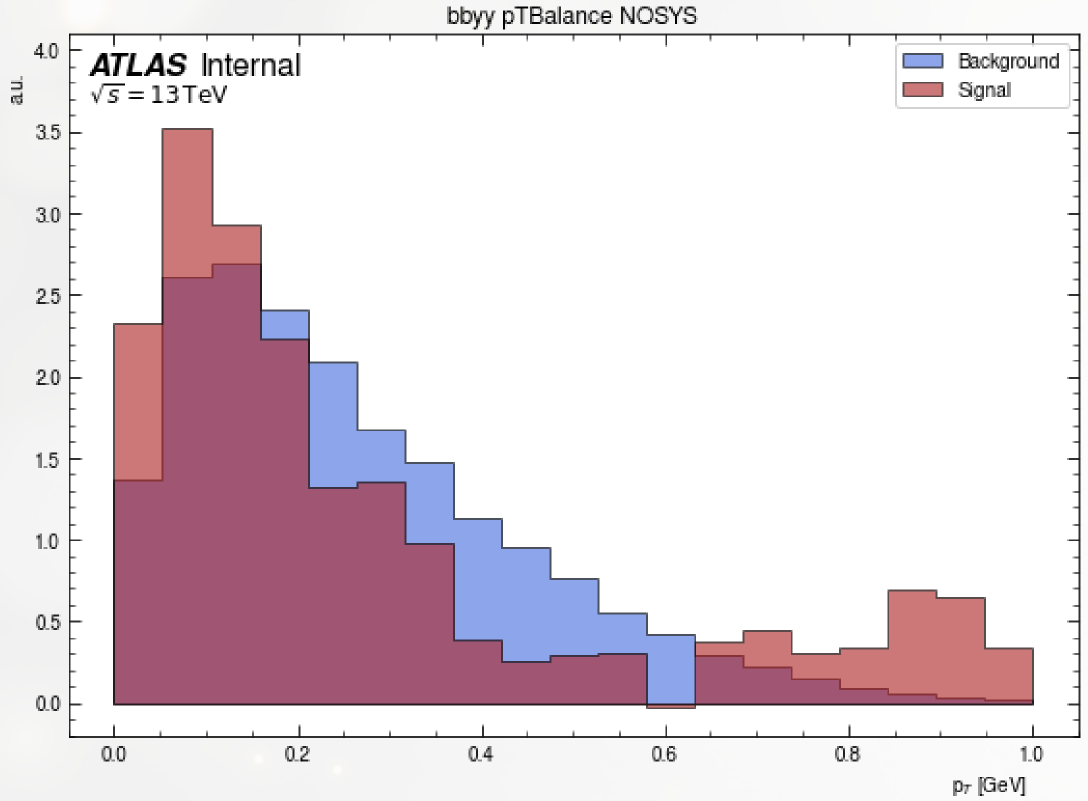


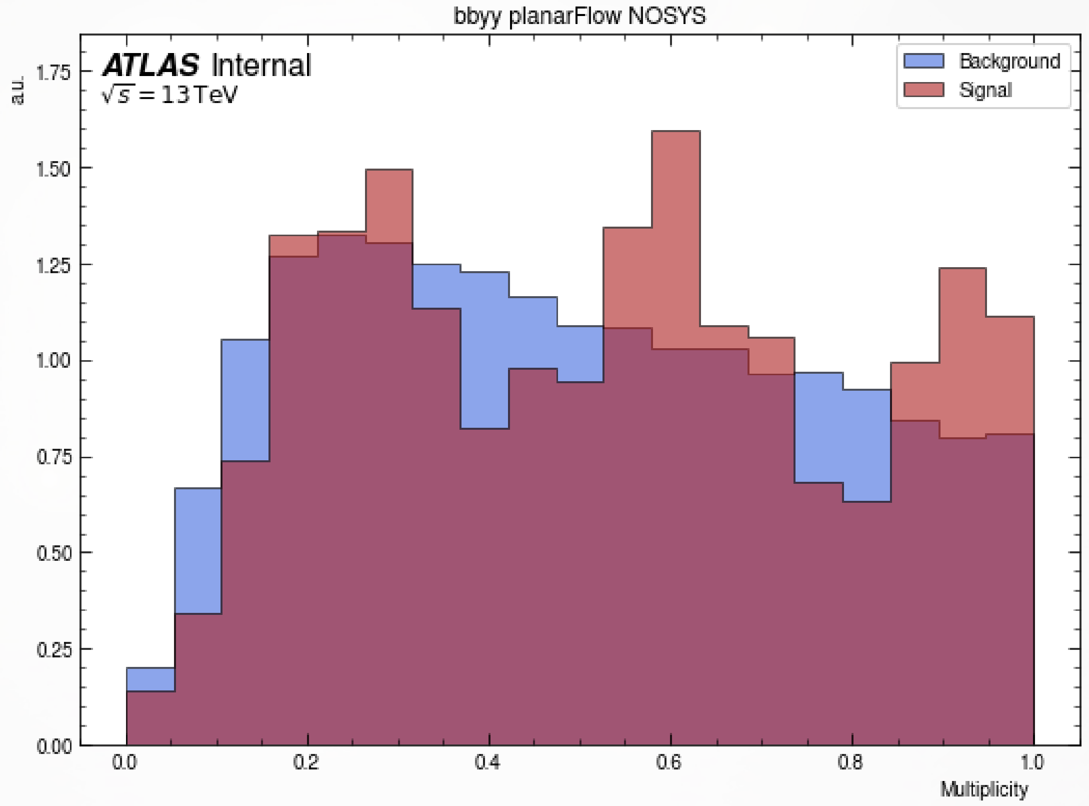
1

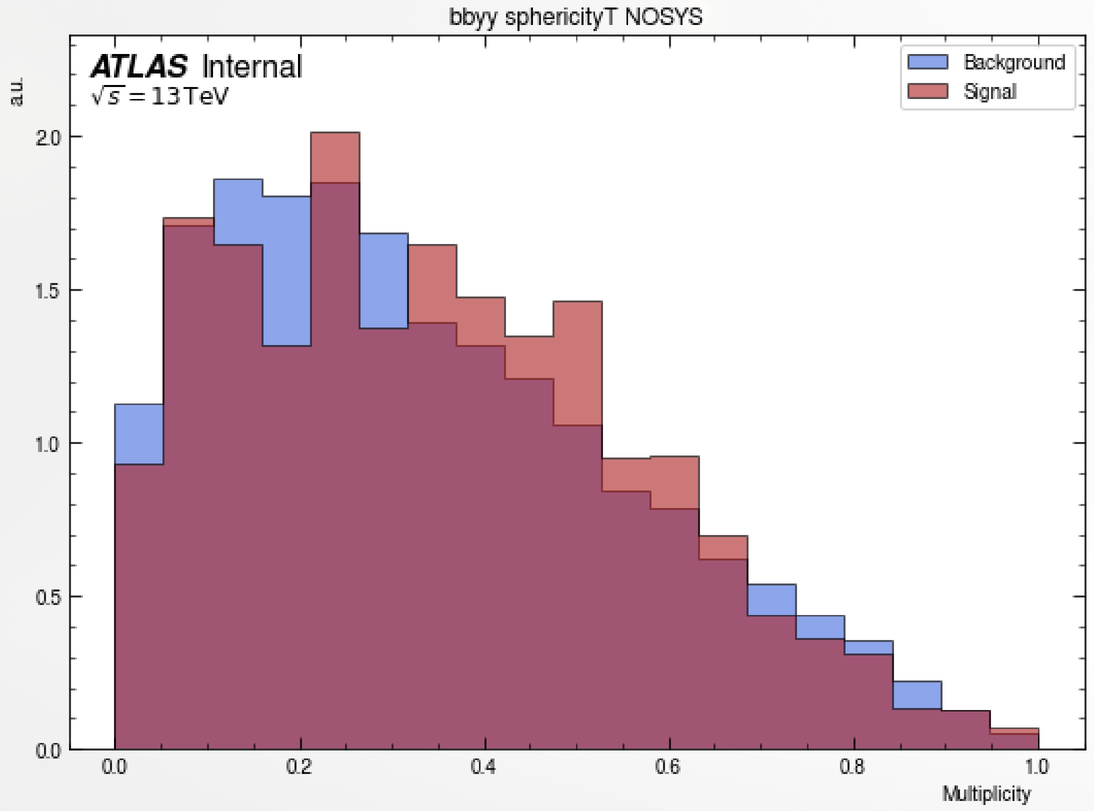
KINEMATIC ANALYSIS

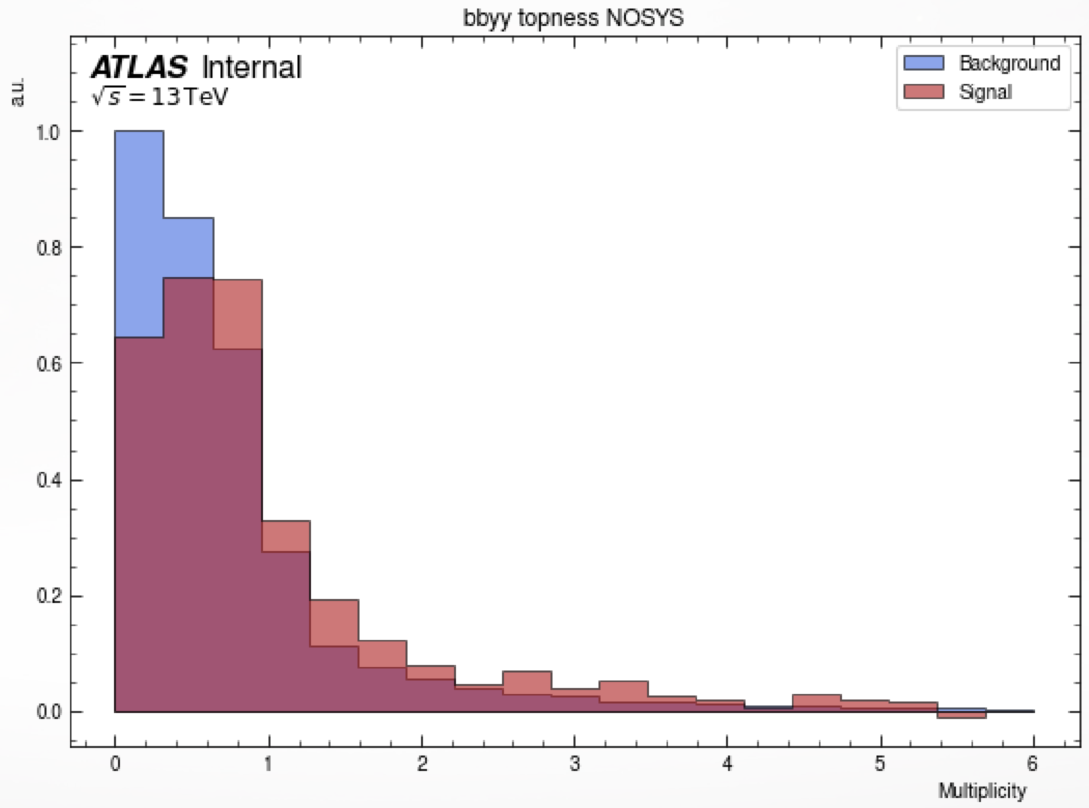
2

Other variables







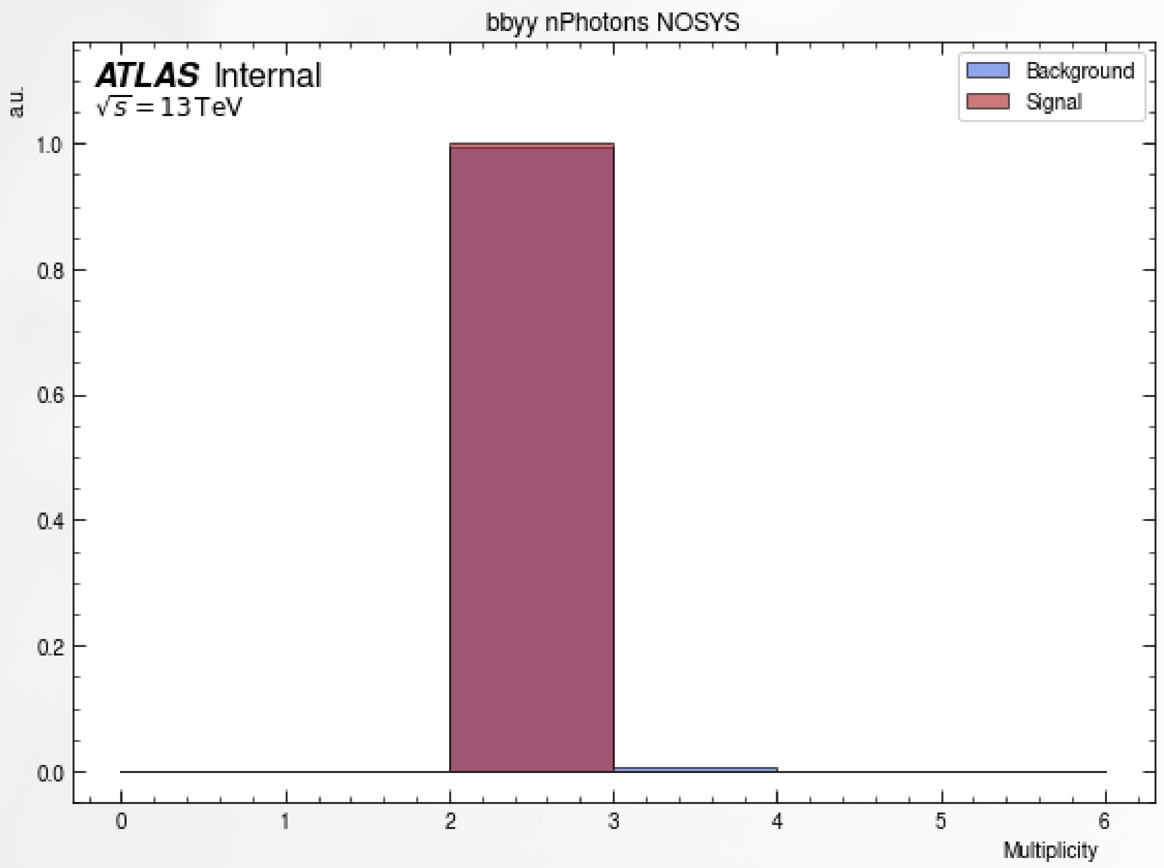


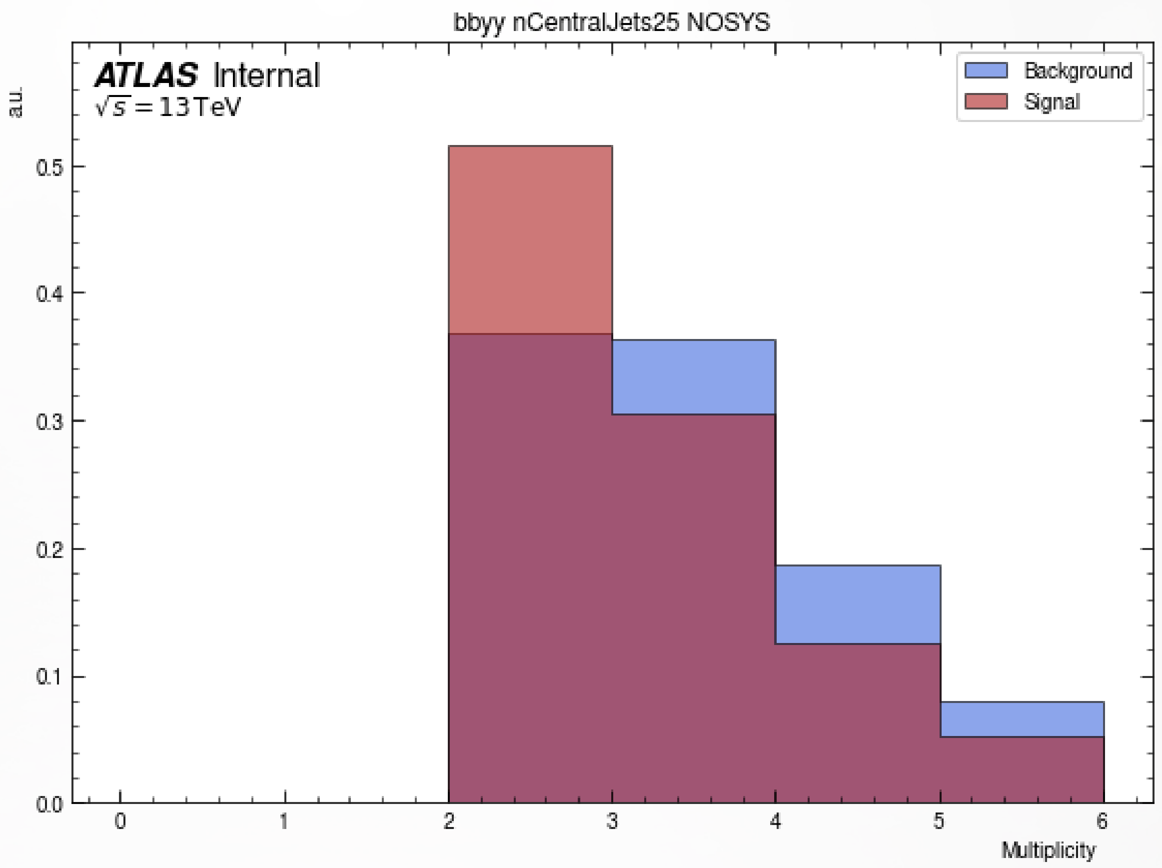
1

KINEMATIC ANALYSIS

2

Other variables





1

KINEMATIC ANALYSIS

2

2

FEATURES

b-tagging ratio

$$b\text{-tag ratio} = \frac{\text{pcbt}_{\text{jet 1}} + \text{pcbt}_{\text{jet 2}}}{\text{pcbt}_{\text{jet 3}} + \text{pcbt}_{\text{jet 4}} + 0.001}$$

Central pT

$$\text{Central}_{p_T} = \frac{\sum_{i=1}^{4} (\text{Jet \{i\}})p_T \exp(-|\eta_i|)}{\sum_{i=1}^{4} (\text{Jet \{i\}})p_T}$$

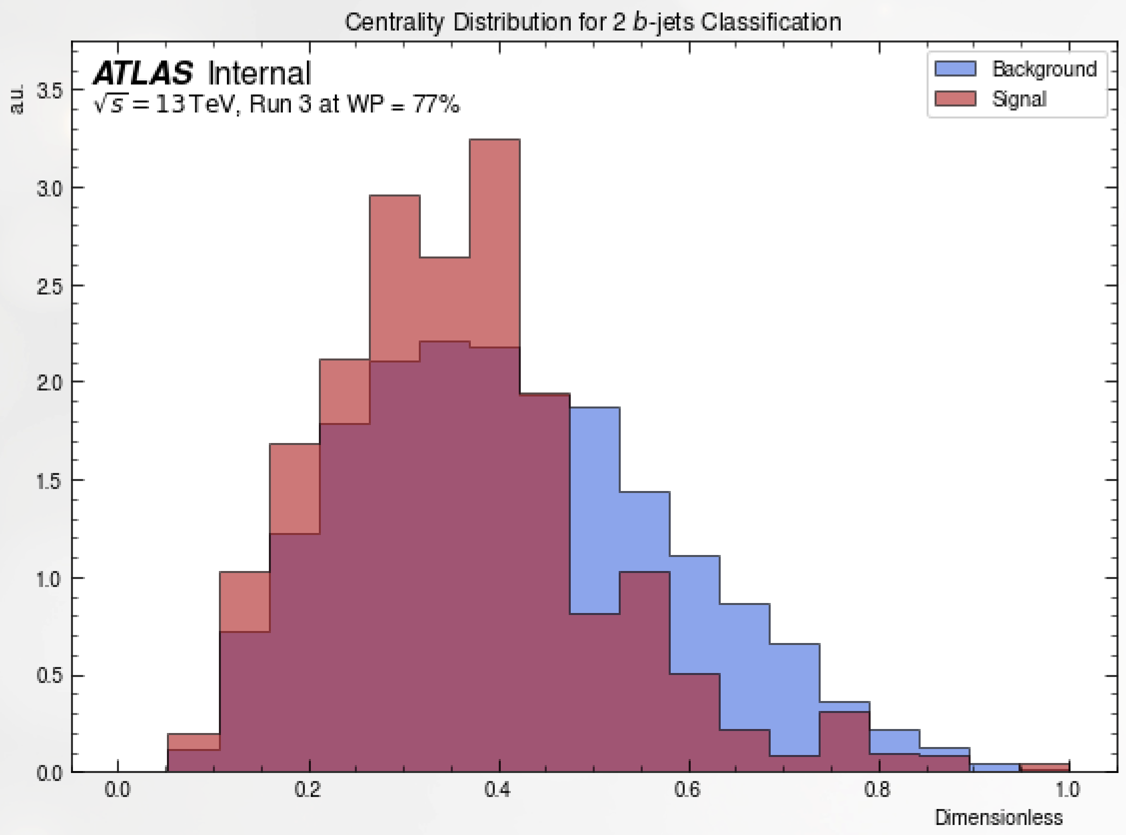
Vector pT Sum

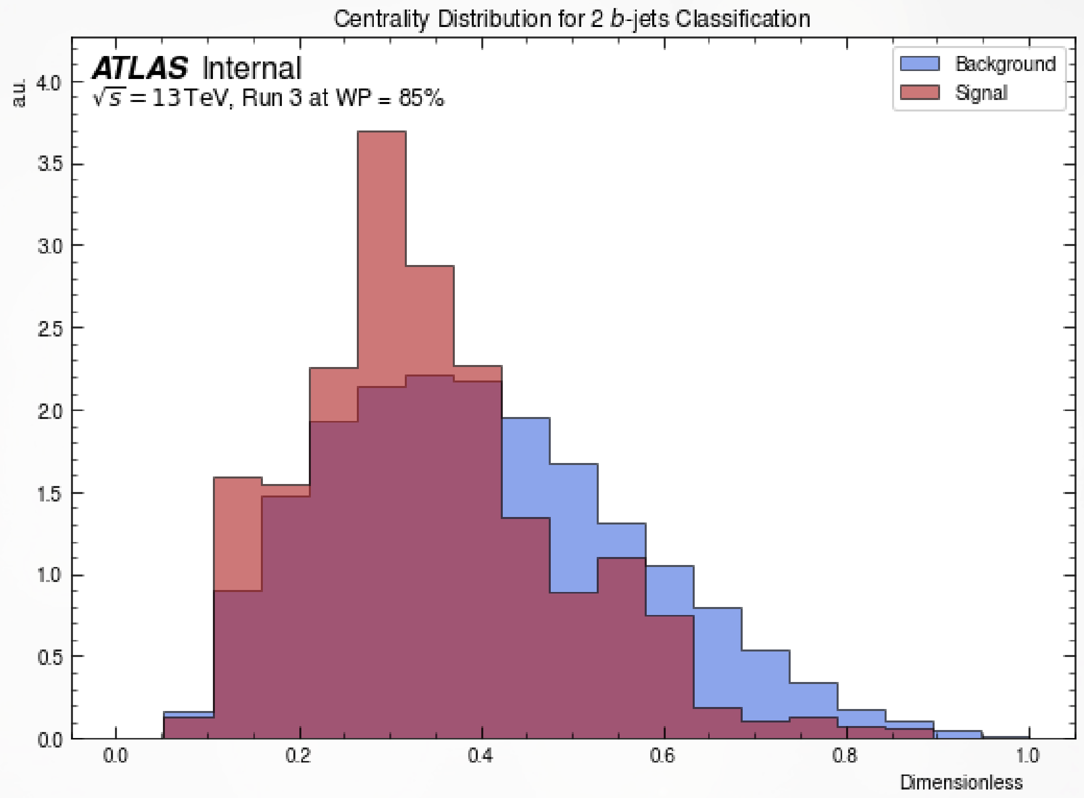
$$\text{Jet\_ph}_{\vec{p_T}} = \frac{\left|\sum_{i=1}^{4} \text{Jet } \vec{p_{T,i}} + \sum_{i=1}^{2} \gamma\, \vec{p_{T,i}}\right|}{\sum_{i=1}^{4} \text{Jet } p_{T,i} + \sum_{i=1}^{2} \gamma\, p_{T,i}}$$

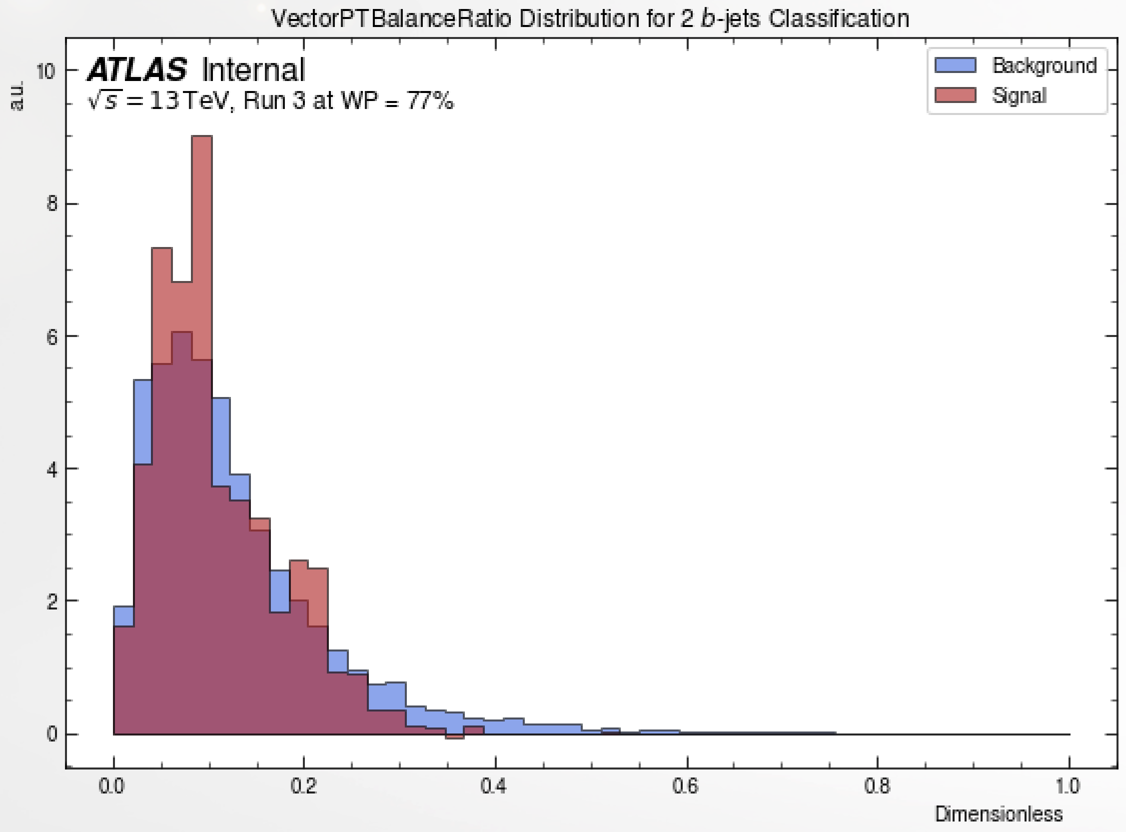
Higgs pT

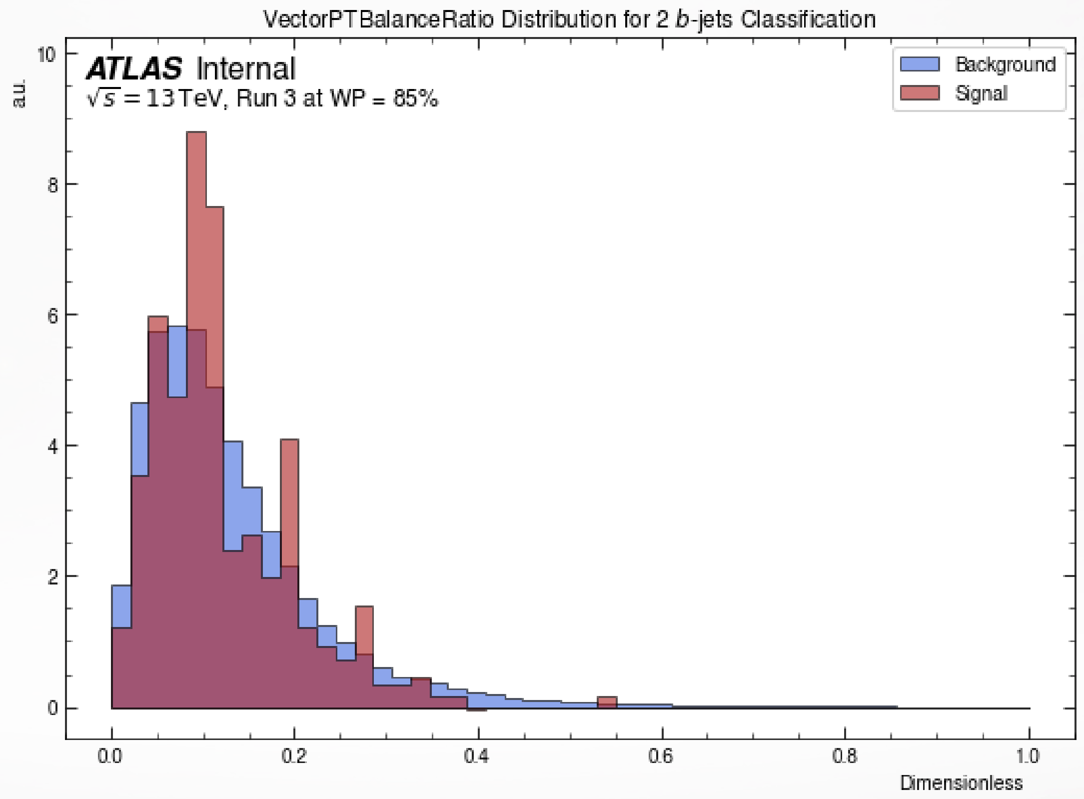
$$\text{Higgs}_{p_T \text{ Symmetry}} = \frac{pT_{\gamma\gamma} - pT_{bb}}{pT_{\gamma\gamma} + pT_{bb}}$$

3


Centrality Distribution for 2 b-jets Classification
ATLAS Internal
√s = 13 TeV, Run 3 at WP = 77%
Background
Signal
a.u.
Dimensionless


Centrality Distribution for 2 b-jets Classification
ATLAS Internal
√s = 13 TeV, Run 3 at WP = 85%
Background
Signal
a.u.
Dimensionless


VectorPTBalanceRatio Distribution for 2 b-jets Classification
ATLAS Internal
√s = 13 TeV, Run 3 at WP = 77%
Background
Signal
a.u.
Dimensionless


VectorPTBalanceRatio Distribution for 2 b-jets Classification
ATLAS Internal
√s = 13 TeV, Run 3 at WP = 85%
Background
Signal
a.u.
Dimensionless

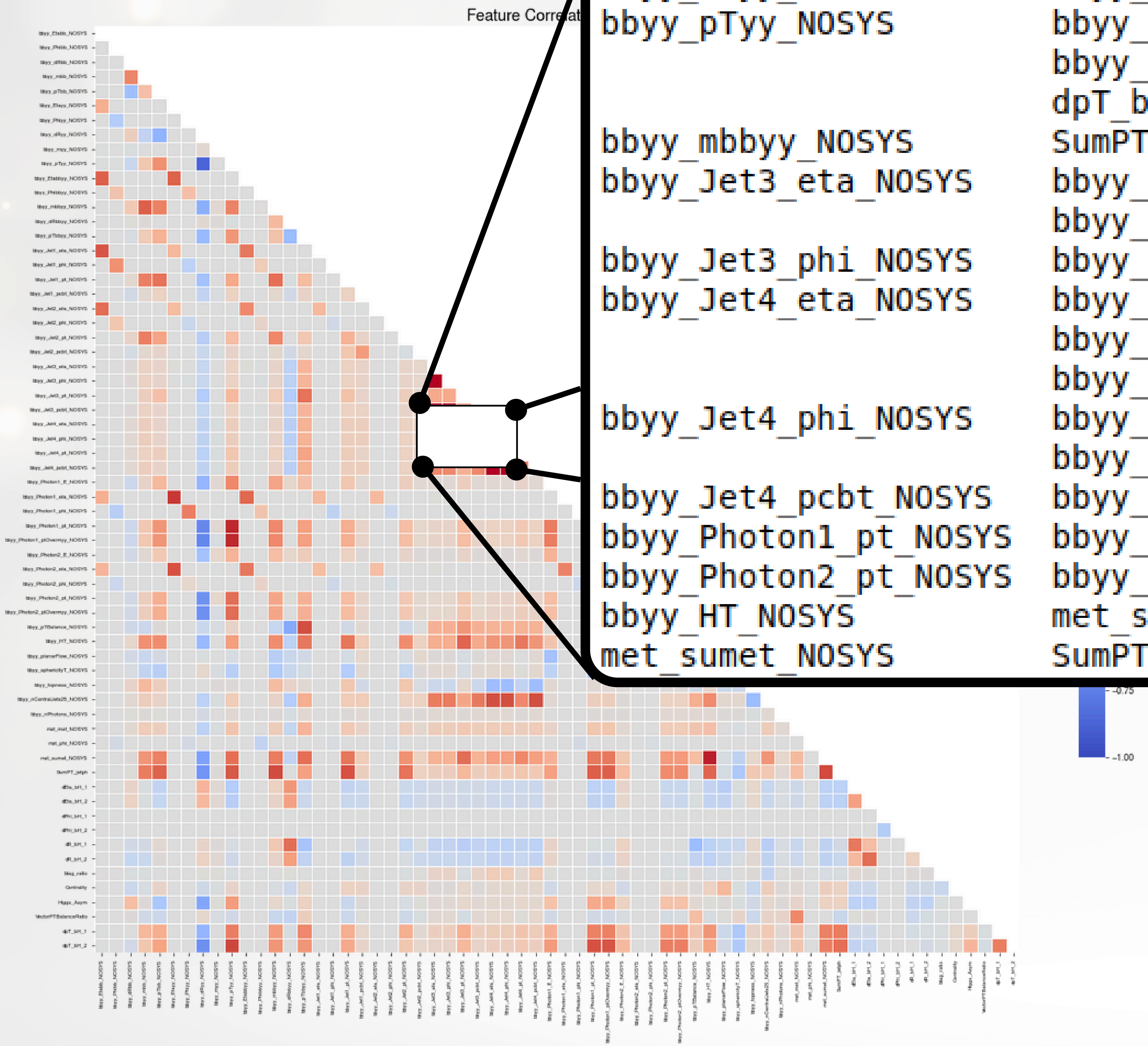
2

FEATURES

3

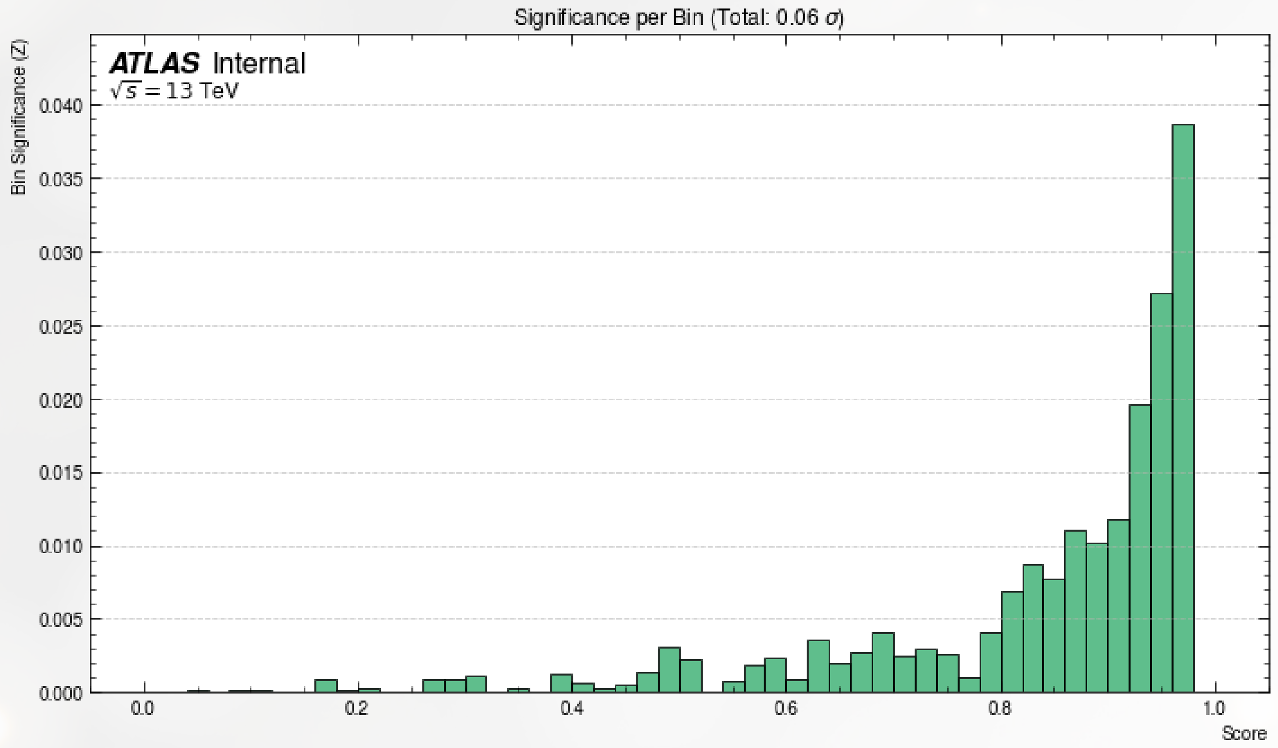
2

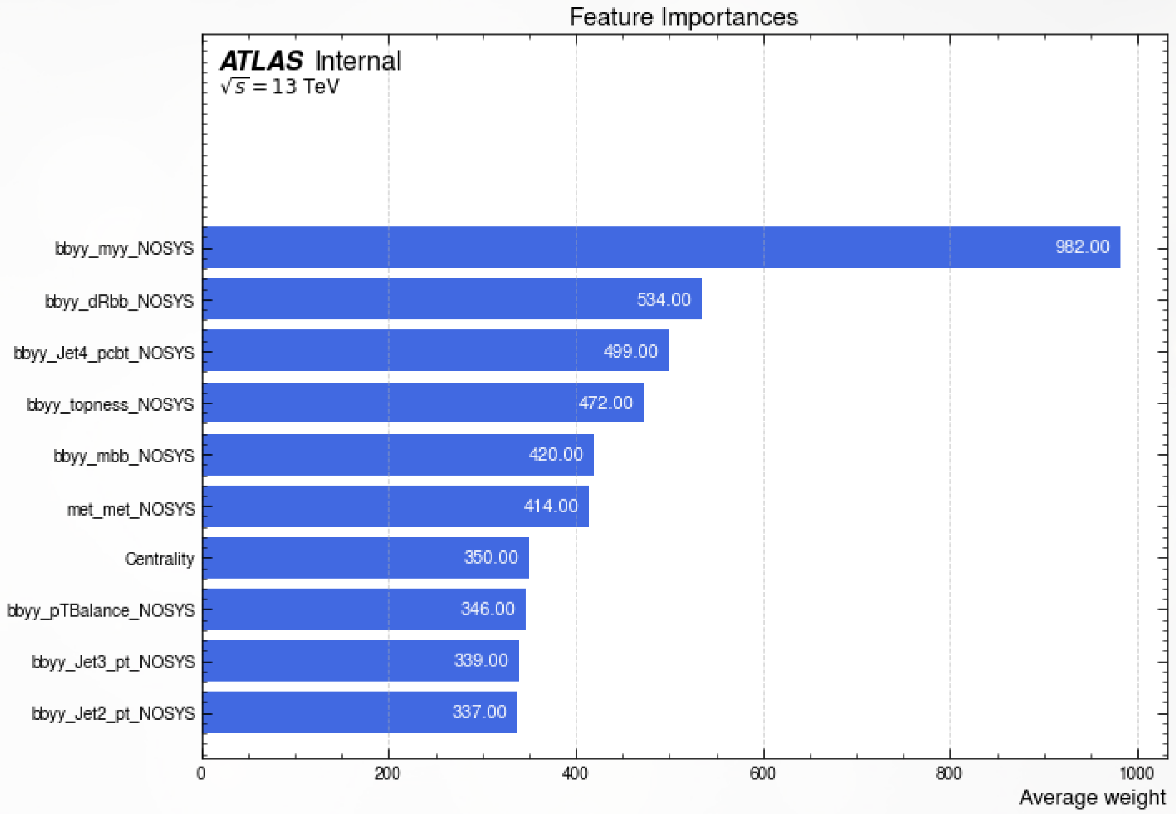
FEATURES

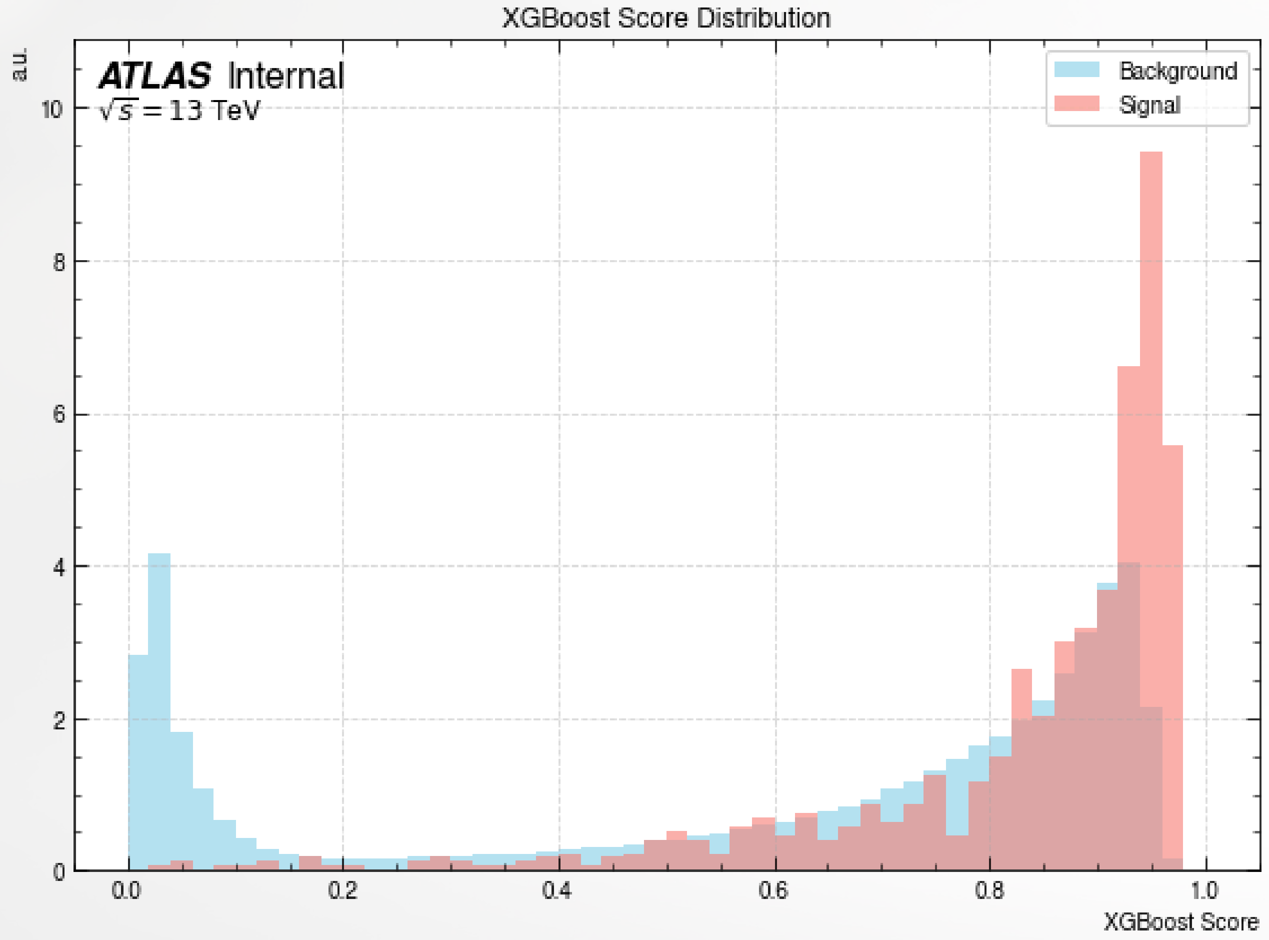


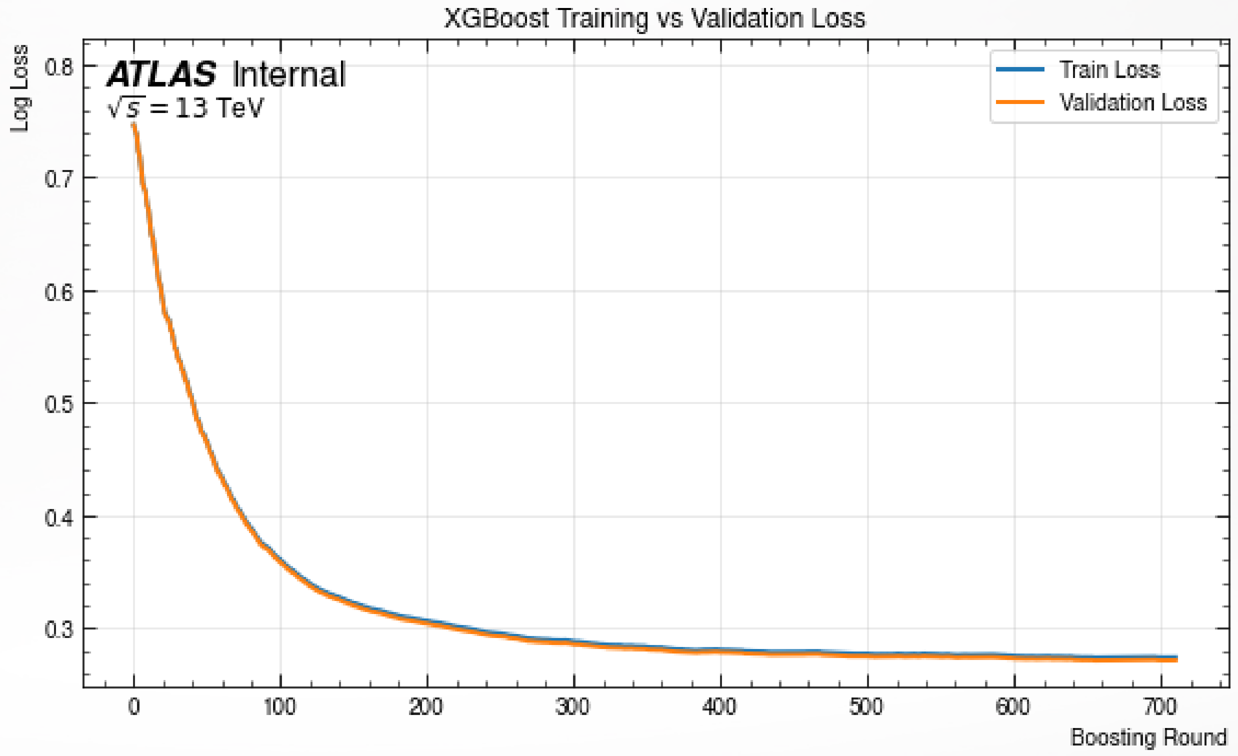
bbyy_Etayy_NOSYS	bbyy_Photon1_eta_NOSYS	0.876659
bbyy_dRyy_NOSYS	bbyy_pTyy_NOSYS	0.871529
bbyy_pTyy_NOSYS	bbyy_Photon1_pt_NOSYS	0.923363
	bbyy_Photon1_ptOvermyy_NOSYS	0.923016
	dpT_bH_2	0.833532
bbyy_mbbyy_NOSYS	SumPT_jetph	0.817284
bbyy_Jet3_eta_NOSYS	bbyy_Jet3_phi_NOSYS	0.998100
	bbyy_Jet3_pcbt_NOSYS	0.998990
bbyy_Jet3_phi_NOSYS	bbyy_Jet3_pcbt_NOSYS	0.998482
bbyy_Jet4_eta_NOSYS	bbyy_Jet4_phi_NOSYS	0.999119
	bbyy_Jet4_pcbt_NOSYS	0.999545
	bbyy_nCentralJets25_NOSYS	0.800250
bbyy_Jet4_phi_NOSYS	bbyy_Jet4_pcbt_NOSYS	0.999429
	bbyy_nCentralJets25_NOSYS	0.800260
bbyy_Jet4_pcbt_NOSYS	bbyy_nCentralJets25_NOSYS	0.805193
bbyy_Photon1_pt_NOSYS	bbyy_Photon1_ptOvermyy_NOSYS	0.994959
bbyy_Photon2_pt_NOSYS	bbyy_Photon2_ptOvermyy_NOSYS	0.992702
bbyy_HT_NOSYS	met_sumet_NOSYS	0.945356
met_sumet_NOSYS	SumPT_jetph	0.827633

3





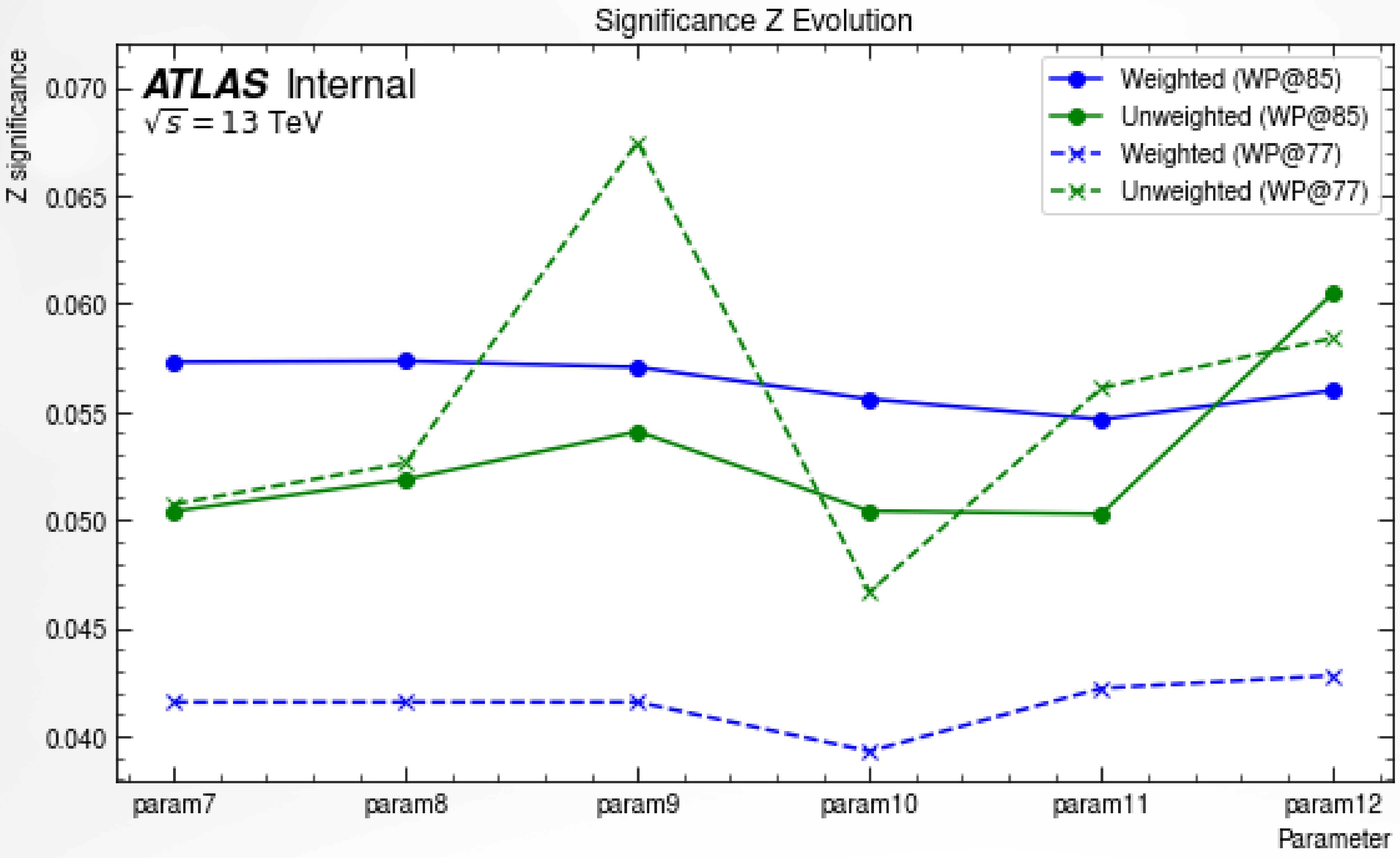




3

BOOSTED DECISSION TREE

4



3

BOOSTED DECISSION TREE

4

Data: Version 6

Run 2 Significance	WP = 77%	WP = 85%
1 b-tag	0.09	0.08
2 b-tag	0.05	0.05

Run 3 Significance	WP = 77%	WP = 85%
1 b-tag	0.08	0.07
2 b-tag	0.02	0.03

3

BOOSTED DECISION TREE

4

Data: Version 9 (Has a required 2 b-tagged at WP = 85)

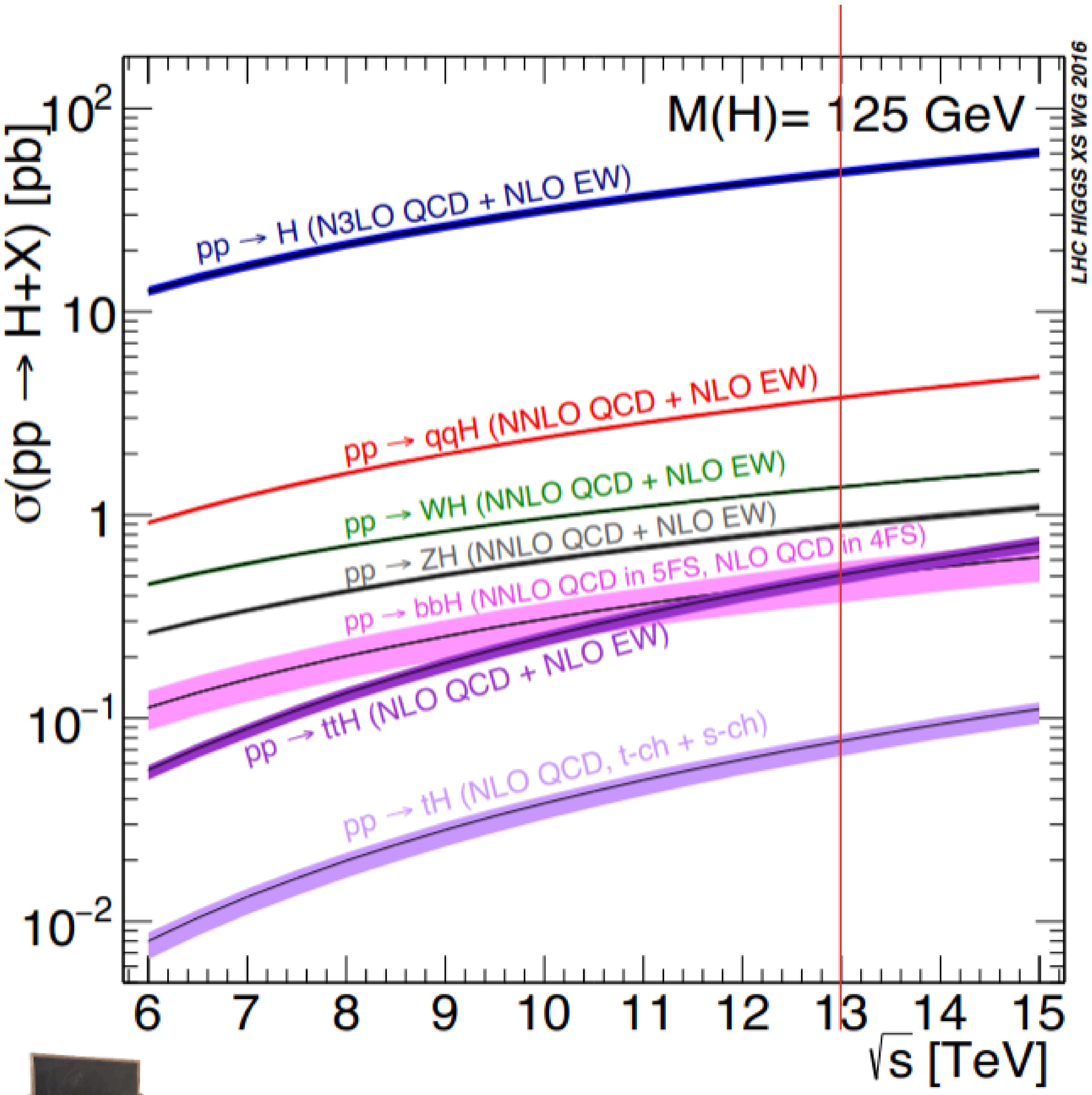
Run 2 Significance	**WP = 77%**	**WP = 85%**
2 b-tag	0.04	0.06

Run 3 Significance	**WP = 77%**	**WP = 85%**
2 b-tag	0.08	0.06

3

BOOSTED DECISION TREE

4


M(H)= 125 GeV
LHC HIGGS XS WG 2016
σ(pp → H+X) [pb]
pp → H (N3LO QCD + NLO EW)
pp → qqH (NNLO QCD + NLO EW)
pp → WH (NNLO QCD + NLO EW)
pp → ZH (NNLO QCD + NLO EW)
pp → bbH (NNLO QCD in 5FS, NLO QCD in 4FS)
pp → ttH (NLO QCD + NLO EW)
pp → tH (NLO QCD, t-ch + s-ch)
√s [TeV]

5

STATISTICAL FIT

In finding the signal strength (μ), the negative log likehood is minimized

$$NLL = -\log(\mathcal{L})$$

where the likelihood is defined as,

$$\mathcal{L}(\text{data}, \mu) = \prod_i \mathcal{P}(n_i|\mu \cdot s_i + b_i)$$

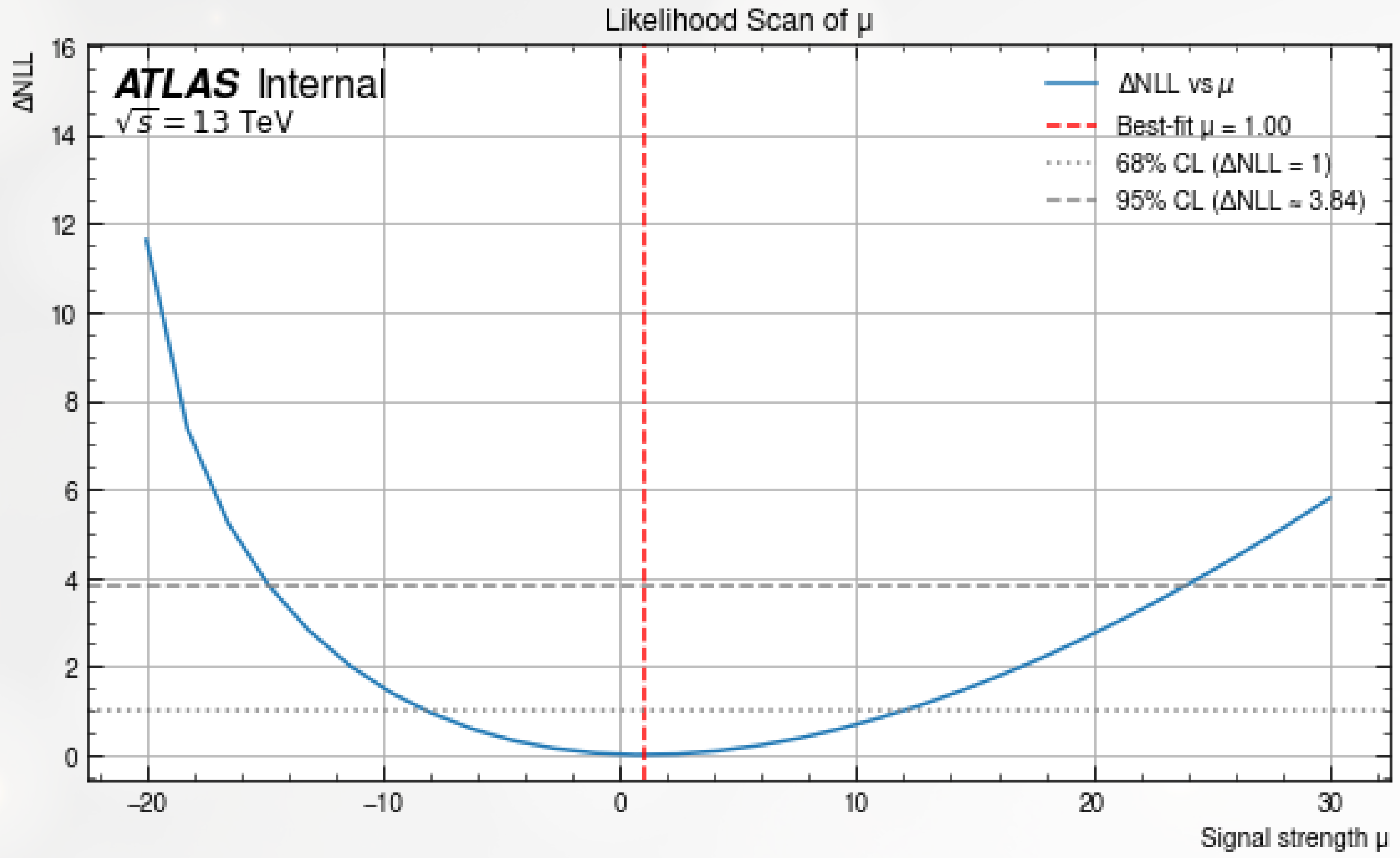
and P is a Poisson Distribution defined as,

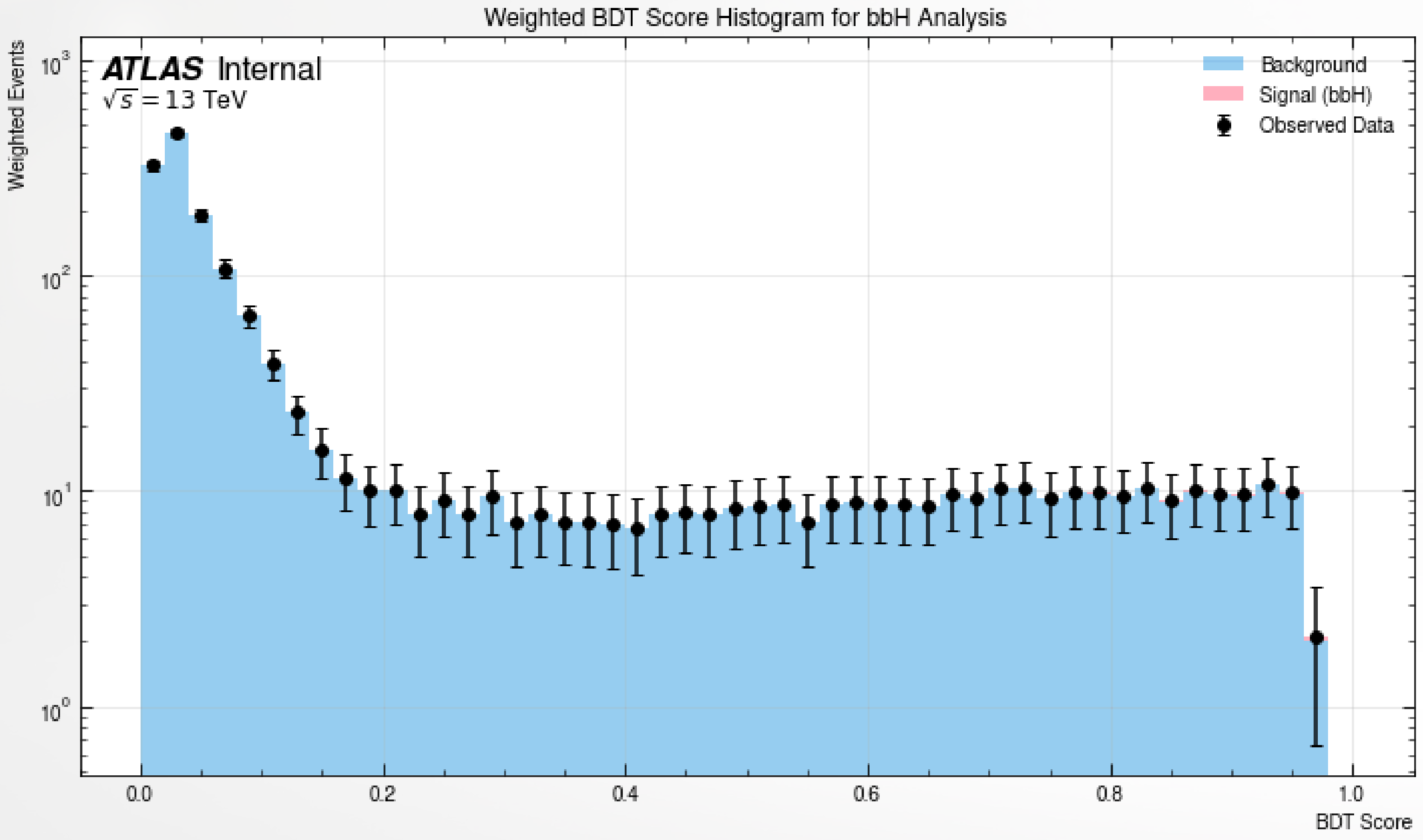
$$\mathcal{P}(n_i|\mu \cdot s_i + b_i) = \frac{(\mu \cdot s_i + b_i)^n \exp(-(\mu \cdot s_i + b_i))}{n!}$$

In getting the range of μ that can still explain the the data found with a certain confidence level

$$q(\mu) = -2\ln\left(\frac{\mathcal{L}(\mu)}{\mathcal{L}(\mu_{best})}\right)$$

The 95% region is defined as $q(\mu) \leq 3.84$





Signal strength (μ): 1.000

95% CL interval: μ ∈ [-14.83, 24.83]

5

STATISTICAL FIT