



Spurious Signal Tests

Laurie McClymont, Andreas Korn

Di-bjet Meet
30 June 2016



- **Confirm that fits discrepancies are not significantly occurring**
 - Fit discrepancy = A difference in shape between fitting function and background shape
 - Fit discrepancy may hide true signal or create fake signal
- **Test fit function by performing fits to background only data-set**
 - Use MC for representative background only data-set
 - Create data-like distributions by applying poisson fluctuations
 - Study fit quality - BH p-value, Chi2 p-value, Deficit Hunter p-value
 - Search for evidence of spurious signal

Event Selection

Pythia8EvtGen MC Di-Jet Sample

- **HLT_j380**
- 2016 MC
- di-b-jet Ntuple production

Scale to 10ifb

- Will update for final lumi

• **Standard Dijet Resonance Cuts**

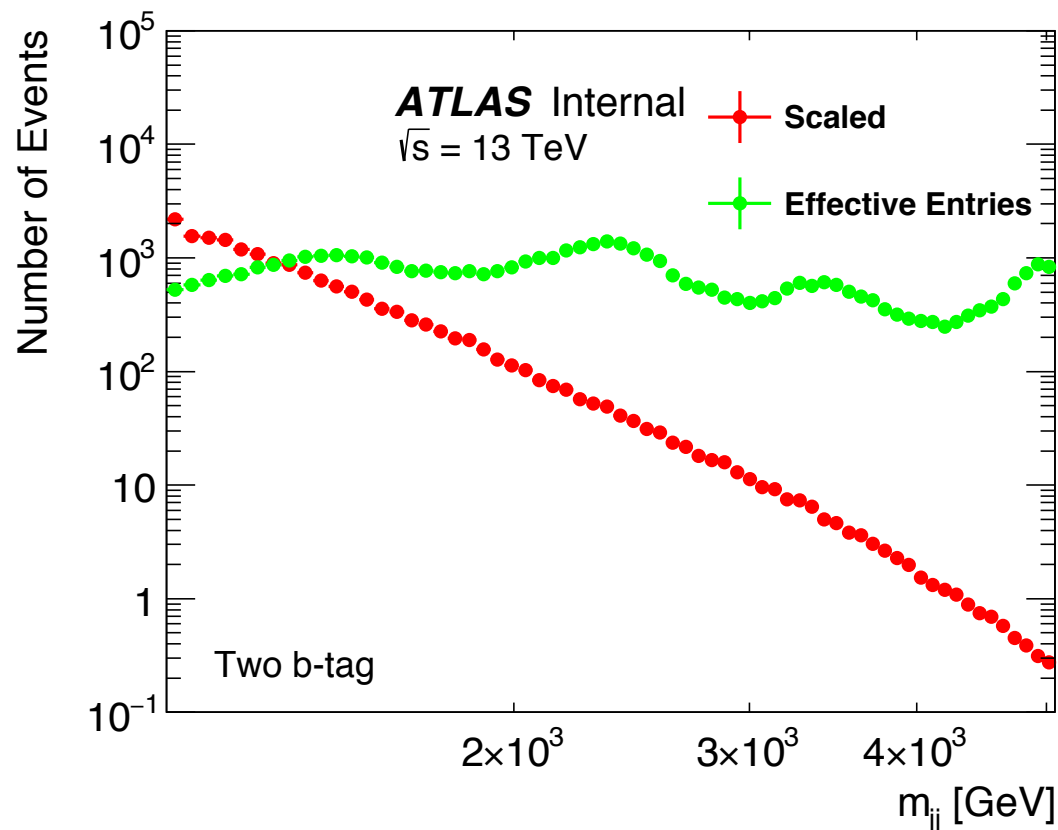
- Leading Jet $p_T > 440$ GeV
- Sublead Jet $p_T > 60$ GeV
- $|y^*| < 0.6$
- $m_{jj} > 1100$ GeV

• **MV2c10**

- Using fixed cut 85% for both jets
- mbb_fix_8585
- mbj_inc_fix_8585



We want data-like distributions, poisson fluctuations applied to precise background estimate



Where Effective Entries > Scaled: We have enough stats for fit tests

Scaled > Eff Ent. : Errors driven by MC, can't make data-like distributions

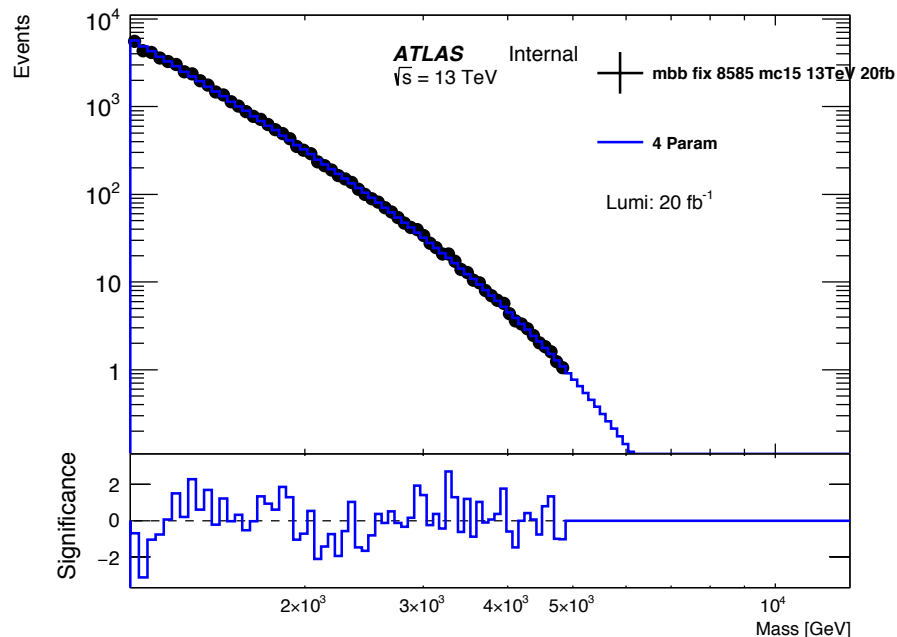
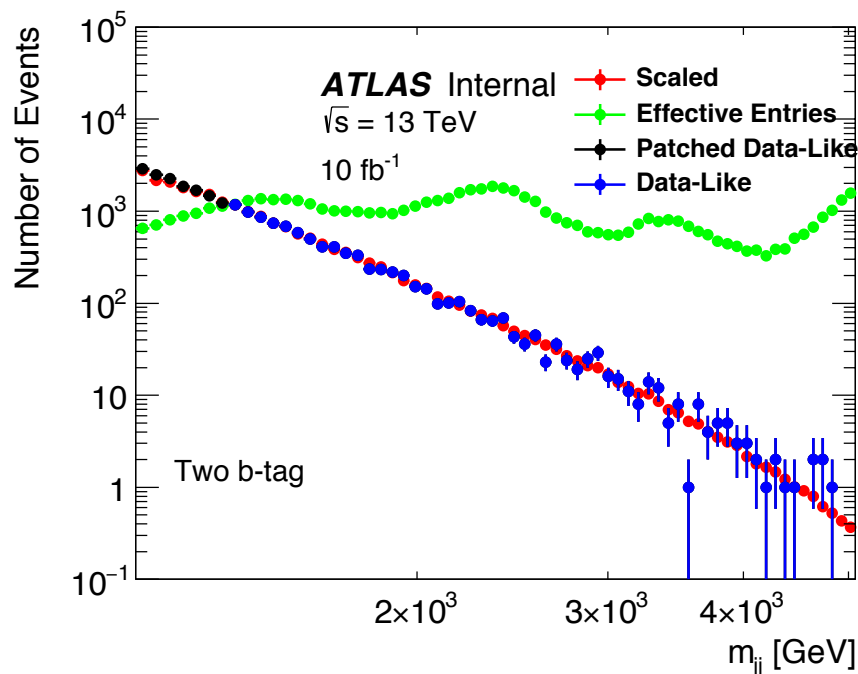
- We do not have a high precision prediction for background here
- In this case, there are 7 bins where this is true
- Mass range: 1100 - 1341 GeV



Patch 6 bins using fit to full spectrum

Procedure:

- Scale to 20ifb => Large range
- **Cut off** => Consider up to 1
=> True MC errors
- Fit using **4 para fit function**
- Use this fit for 7 discrepant bins



$\chi^2: 93.585$ $NdF: 115.000$

Make Data Like Distribution

Procedure:

- Use Scaled distribution from MC
- (Use patched fit for those 7 bins)
- Apply poisson fluctuations
- This can be done for many seeds



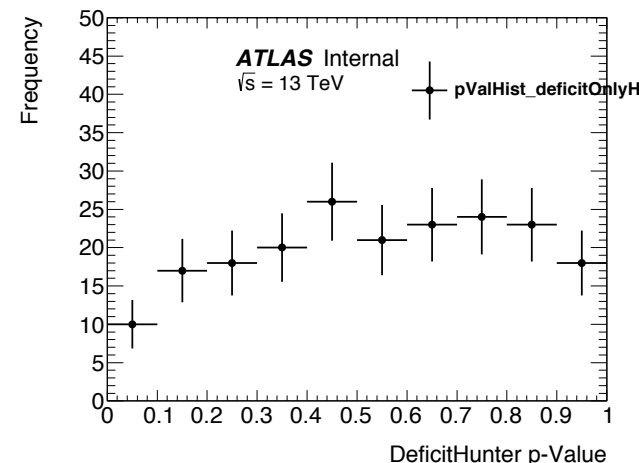
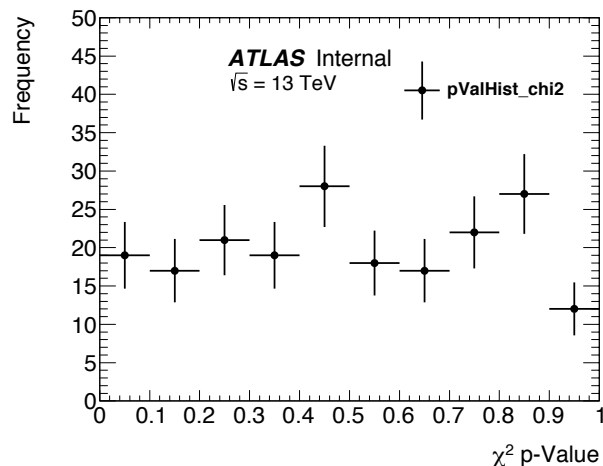
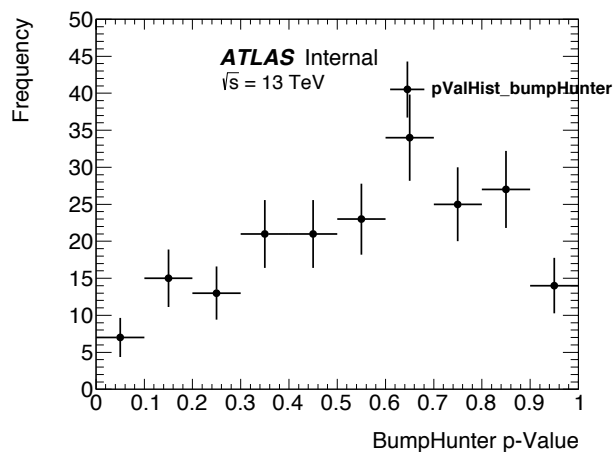
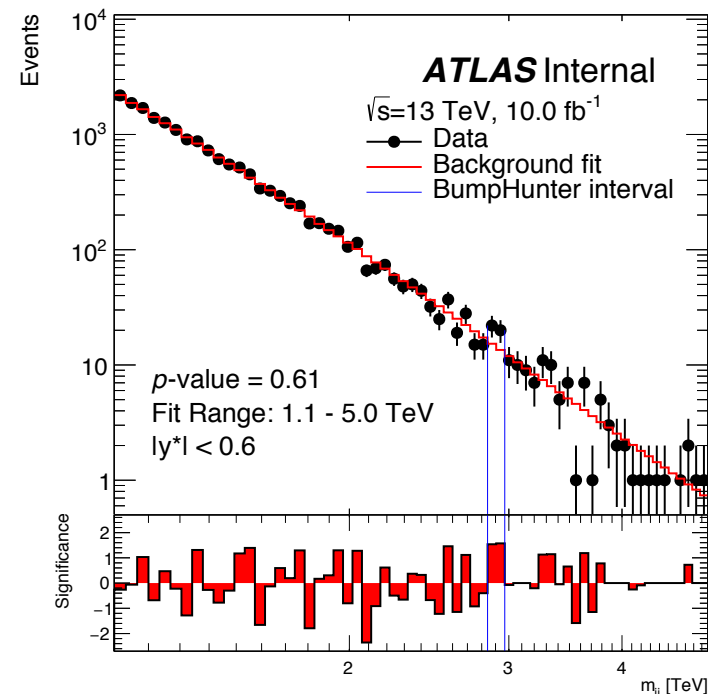
5 Fit Tests - mbb

Fit to data-like distribution

- Using 4 parameter fit function
 - *Can change according to Wilks' procedure run on full data-set.*
- Carry out bump hunter procedure
- Also run on deficit hunter (*Allow deficit only*)
- Also look at Chi2 p-value

Run over many seeds, get global distribution

- 200 seeds
- Look for any biases towards p-values

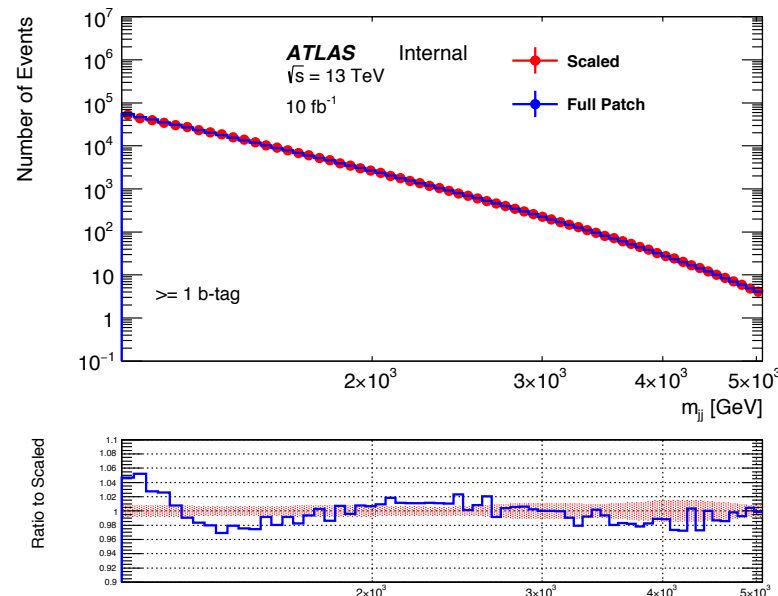
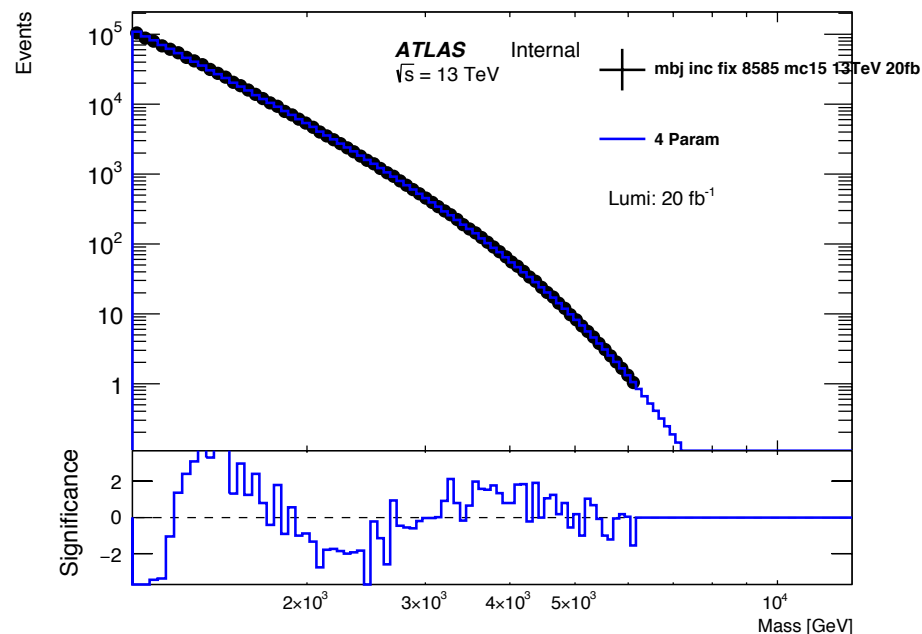
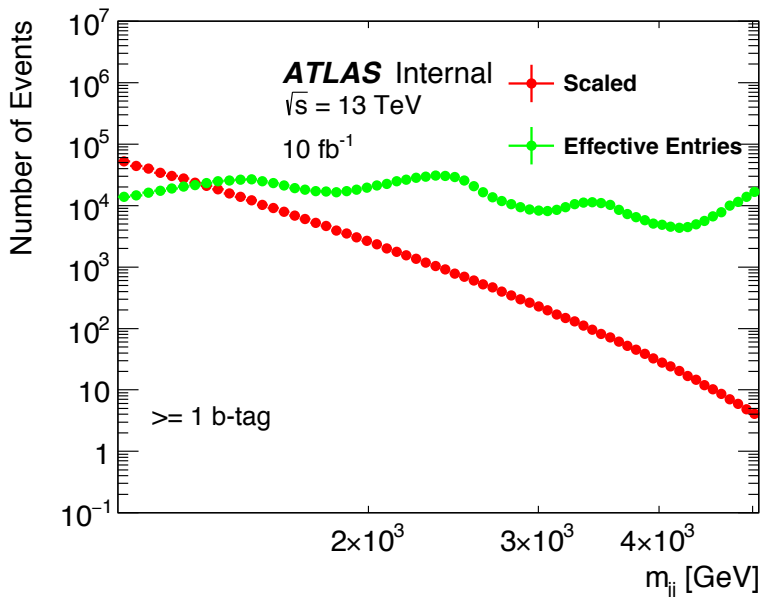




Patch 6 bins using fit to full spectrum

Procedure:

- Scale to 20ifb => Large range
- **Cut off** => **Consider up to 1**
=> **True MC errors**
- Fit using **3 para fit function**
- Use this fit for 7 discrepant bins



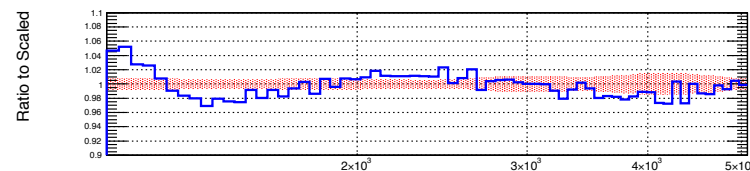
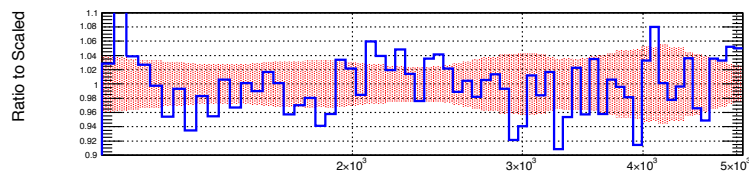
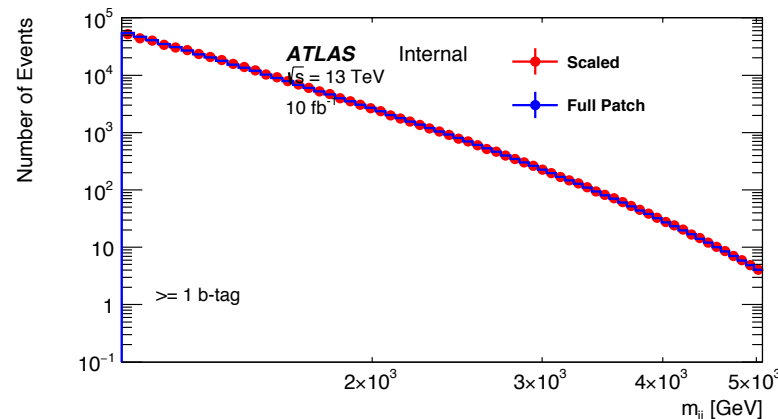
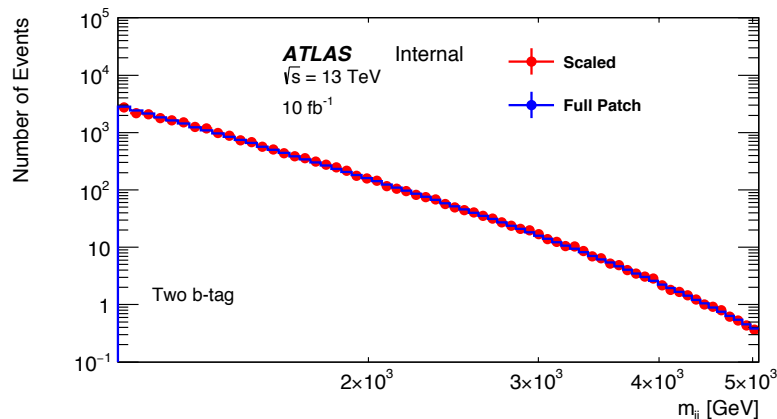
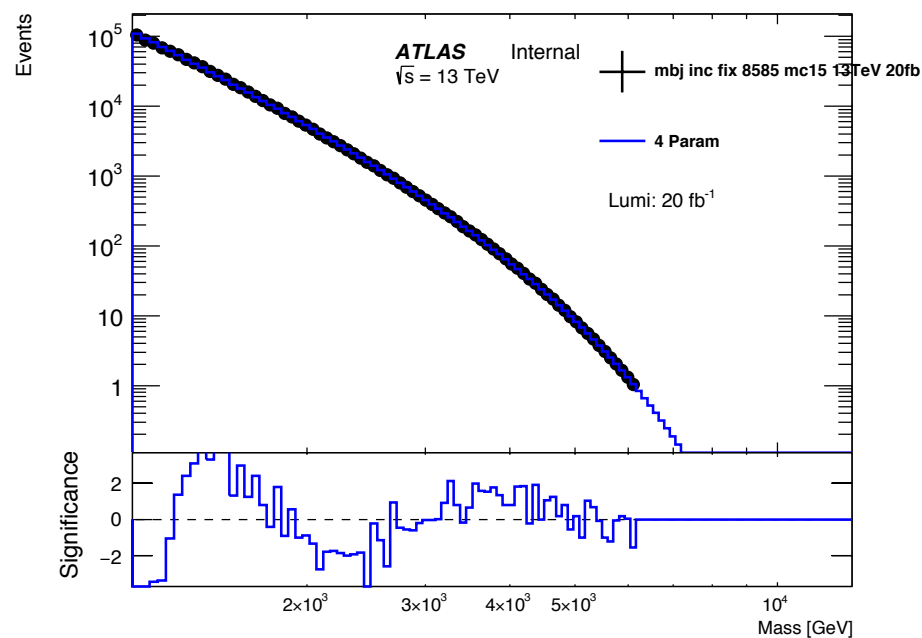


7 mbj Patch

Patch 6 bins using fit to full spectrum

Procedure:

- Scale to 20ifb => Large range
- **Cut off => Consider up to 1**
=> True MC errors
- Fit using **3 para fit function**
- Use this fit for 7 discrepant bins





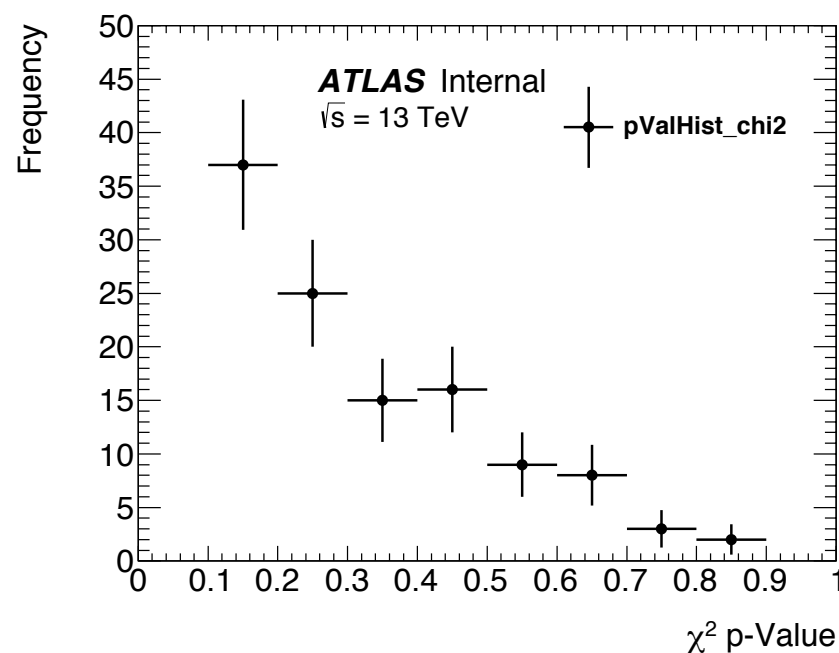
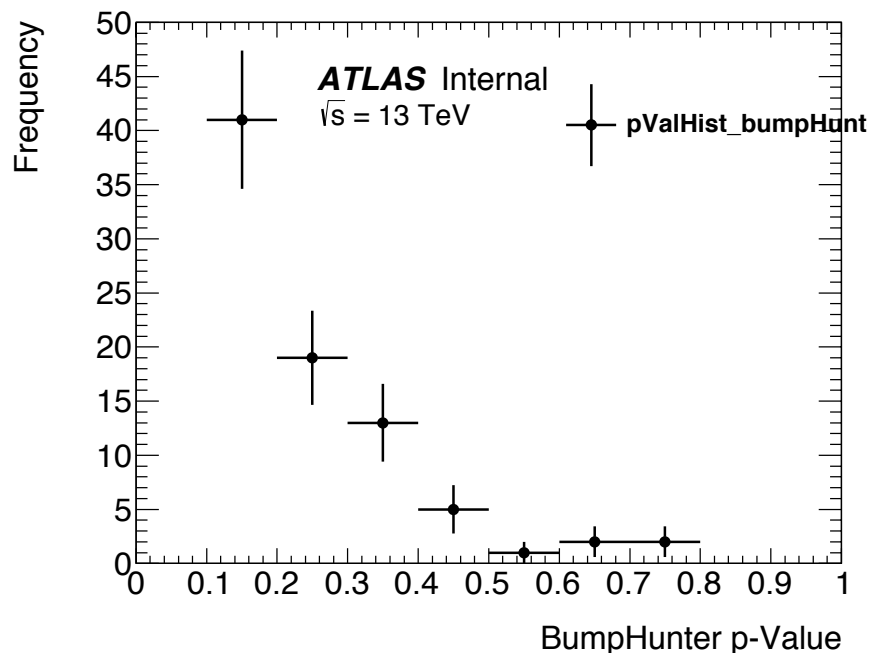
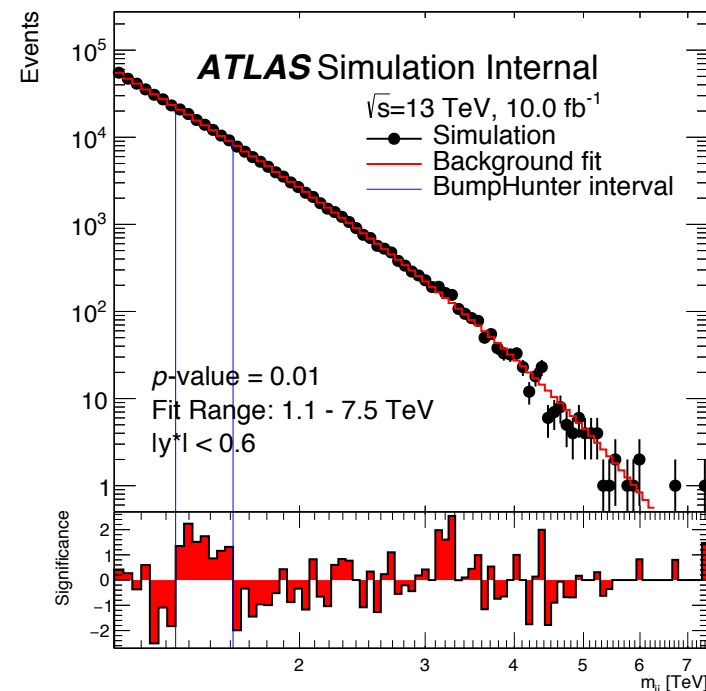
8 Fit Tests - mbb

Fit to data-like distribution

- Using 4 parameter fit function
 - *Can change according to Wilks' procedure run on full data-set.*
- Carry out bump hunter procedure
- Also run on deficit hunter (*Allow deficit*)
- Also look at Chi2 p-value

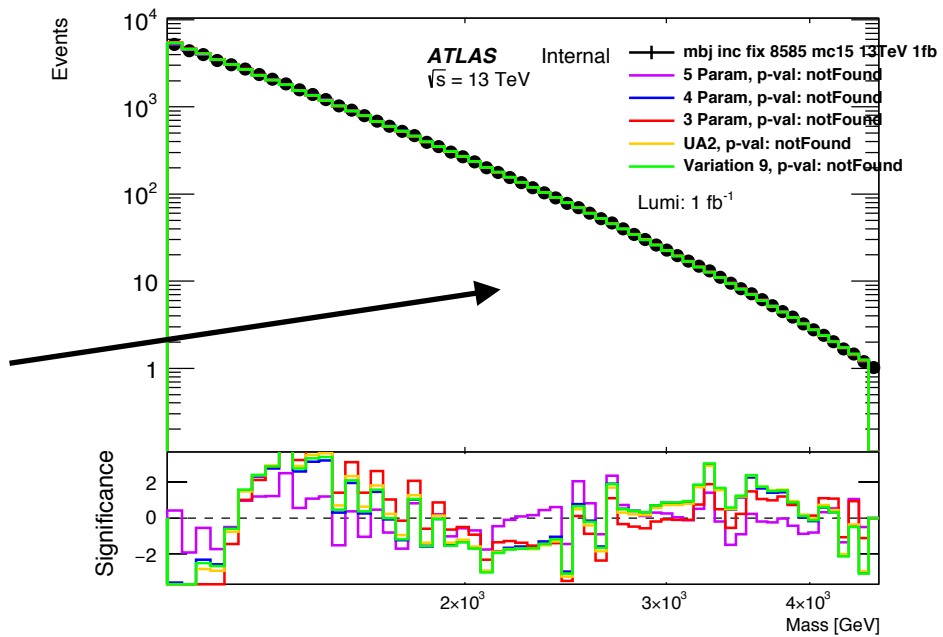
Run over many seeds, get global distribution

- 200 seeds
- Look for any biases towards p-values

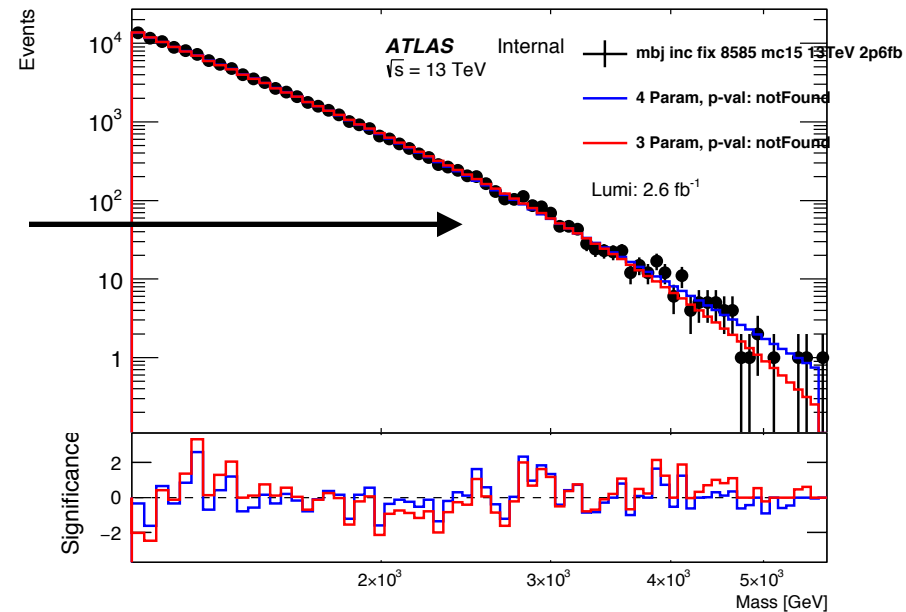


Possible solutions

1) Fix the patch



2) Fit to data-like only.



3) TLA sample with truth tagging

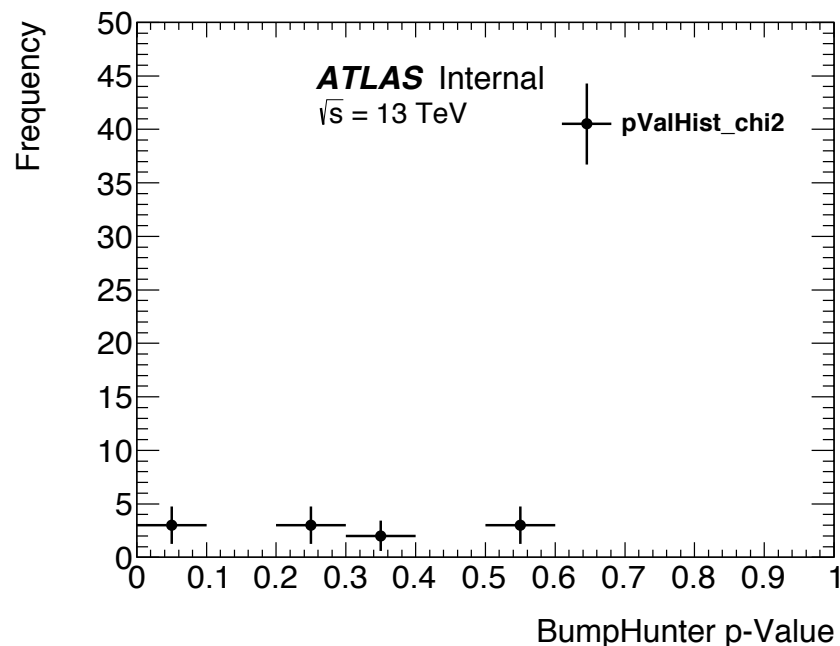
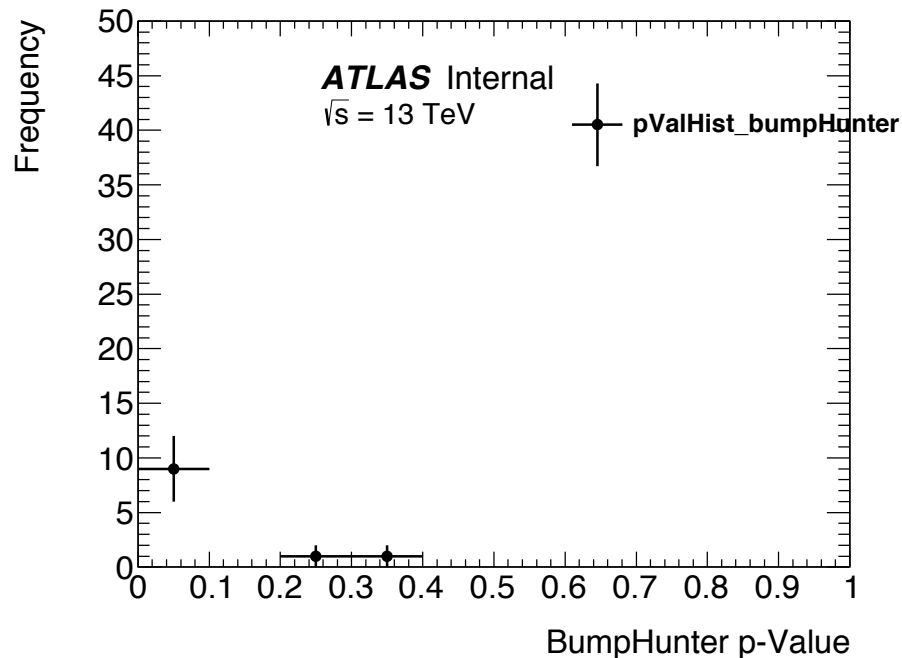
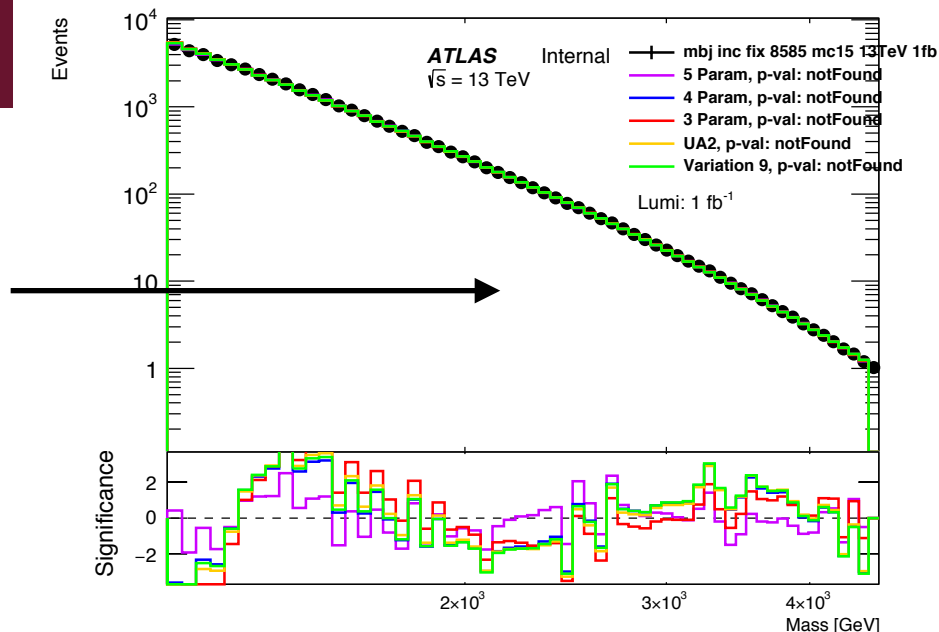


10 Possible Solutions - (1)

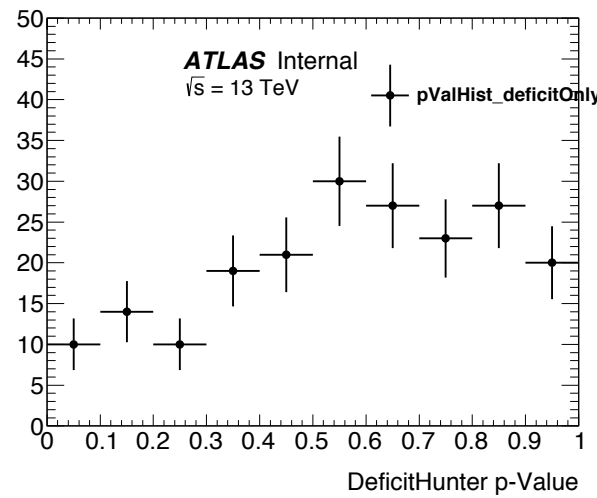
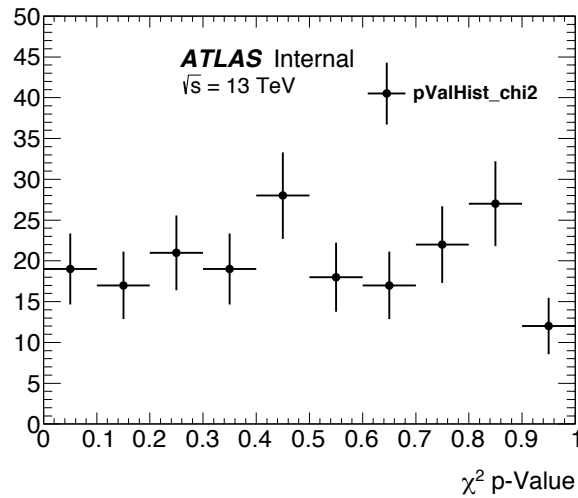
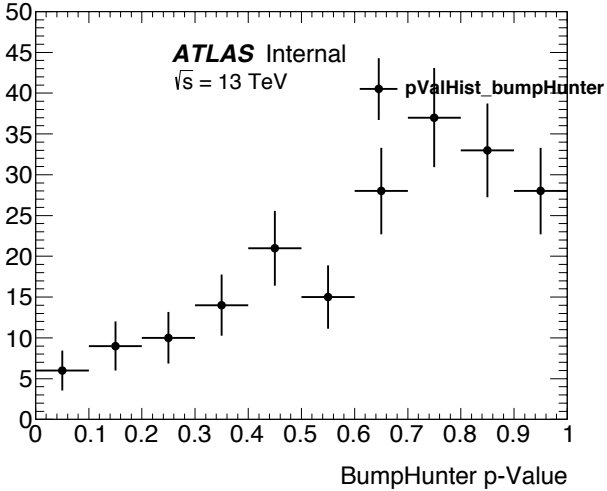
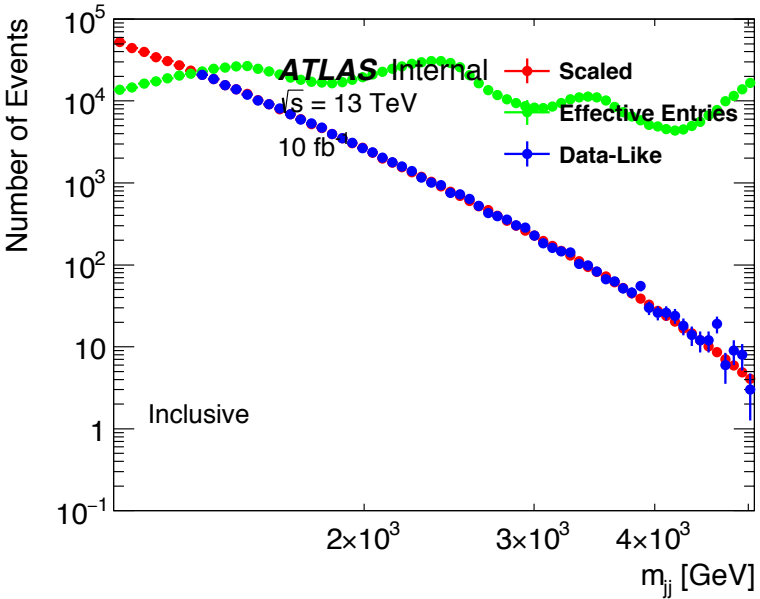
1) Fix the patch

- Possible that changing to 5 para and fitting to a different mass range might fix the problem

Didn't Work



- 2) Fit to data-like only.
- Perform tests to region where we have precision (no patch)
 - Doing now!
 - 200 seeds done, looks good much better
 - Why the bias towards BH high-values?





3) TLA sample with truth tagging

- TLA used large truth MC sample to perform tests
- Can we use this with emulation of b-tagging
- Does such a tool exist?

Probably best option long term!

Unofficial tool exists, see [here!](#)

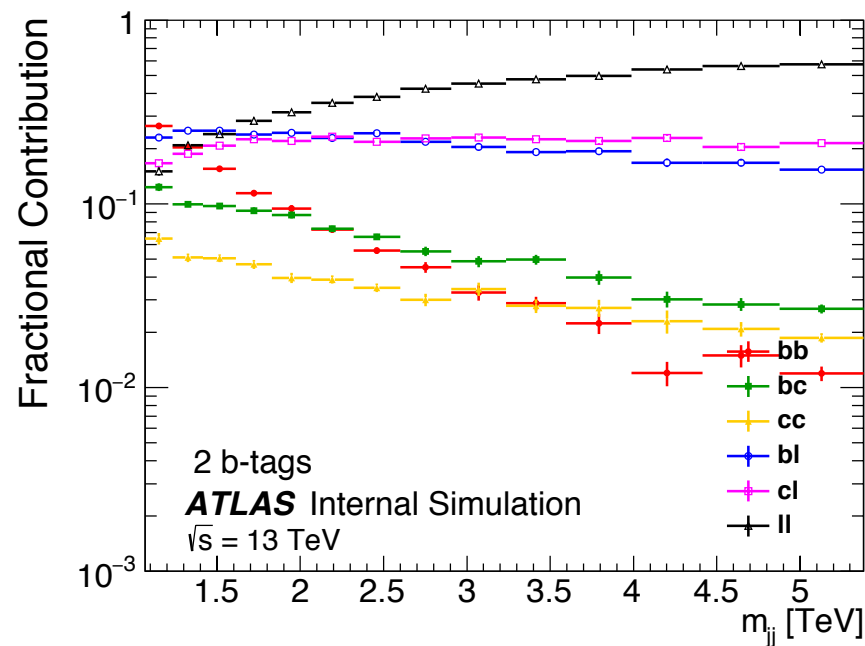
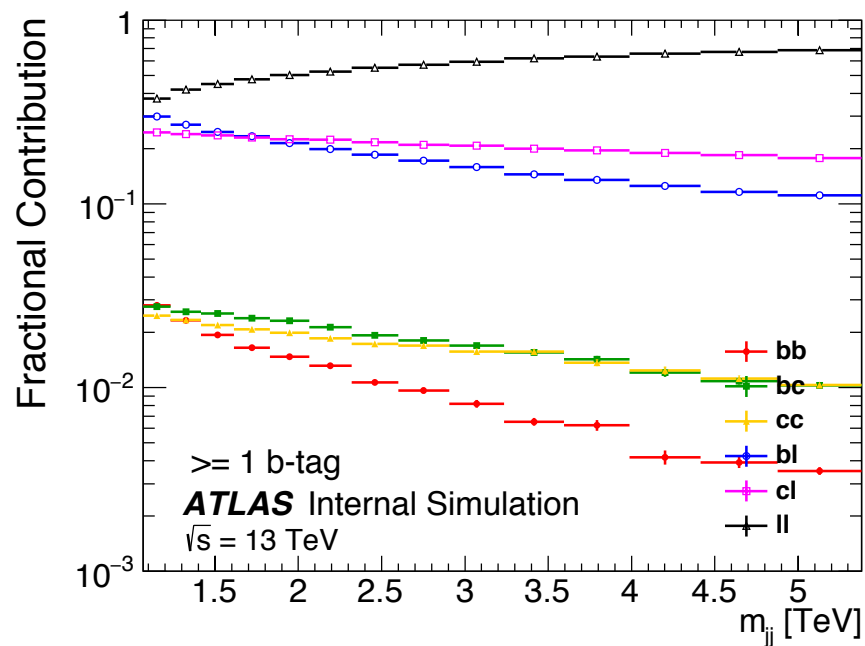
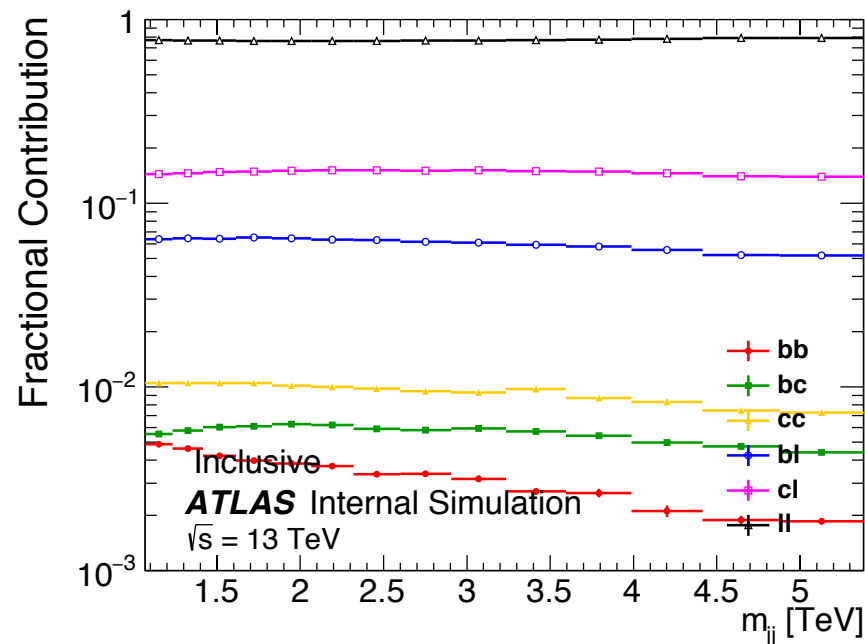
- Needs to be implemented
- MC validation

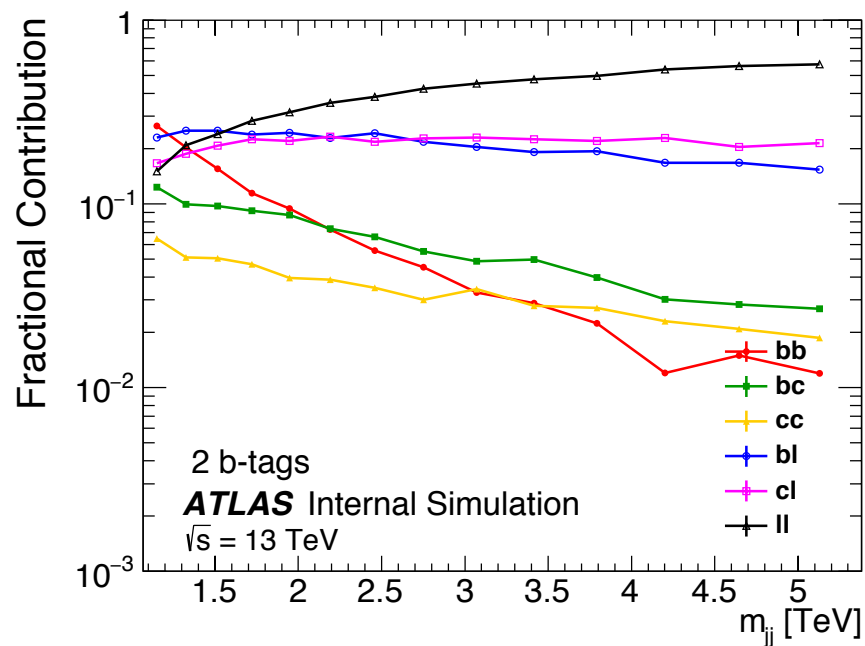
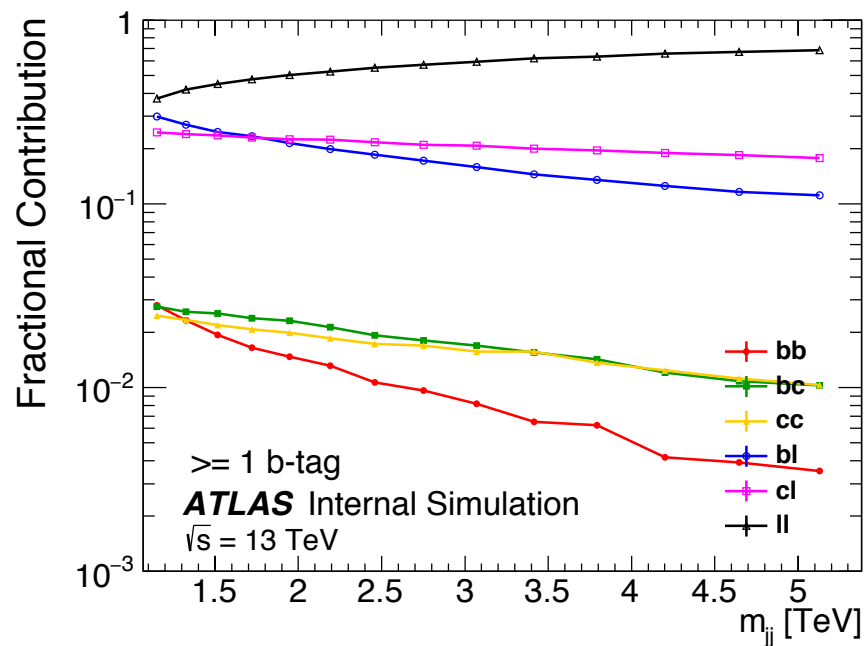
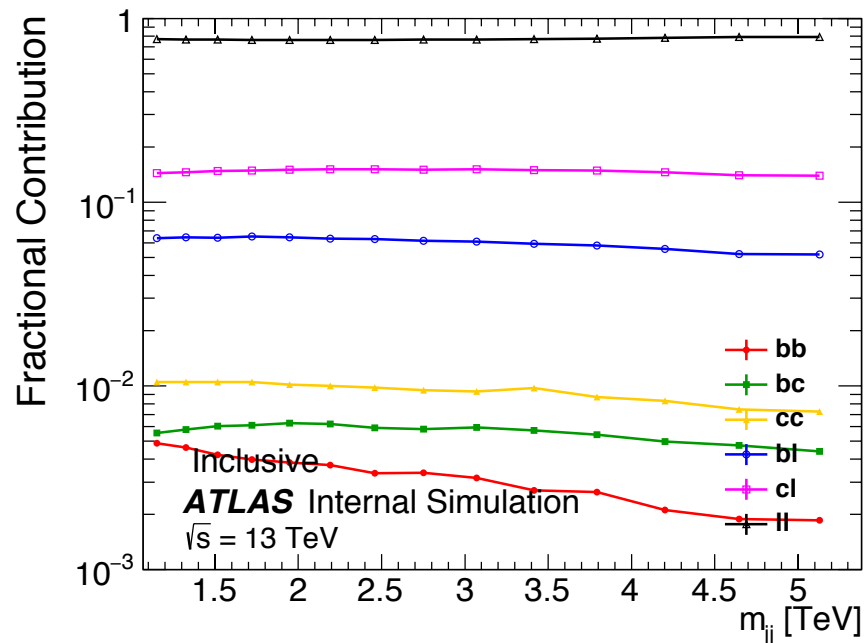
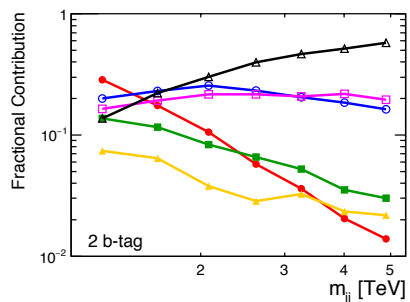
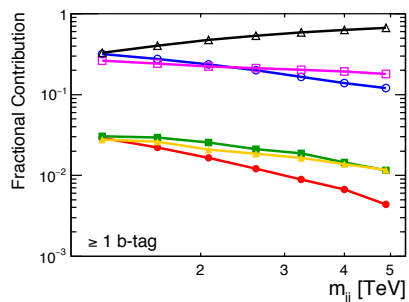
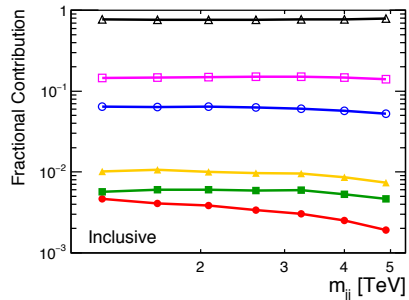


13 Flavour Fractions

Smooth

II dominates at high masses







- **Tested 4 parameter function in data-like distributions at 10 fib**
 - 2-tag => Framework set up to create data-like distributions using a patch (mbb)
 - ≥ 1 tag => Framework set up to create truncated data-like distributions.
 - We can then fit to get p-values (BH, DH, Chi2)
- **Global p-value distributions**
 - 200 seeds
 - No evidence of spurious signal
 - No discrepant bumps or deficits
 - Chi2 p-value suggests that overall fit quality is good
- **Flavour fractions**
 - Dominated by II at high mass
 - Smoothly changing flavour fractions
 - Will put it on one canvas, like in Moriond.

To Do

- **Update note**