Flavour Fit Studies

Laurie McClymont, Andreas Korn

With thanks to Jeff and Lydia!

Di-bjet Meet

3 February 2016

Getting the Flavour Fractions

- We want to understand how varying the flavour composition will affect the fitting function.
  => Are the fitting functions robust to changes flavour composition?
  => Vary the amount that different flavour combinations contribute and fit.

- Comment from JDM Approval
  => Further tests requested.
  => Are we introducing a spurious signal in changing flavour fraction?

Details

Pythia8EvtGen MC Di-Jet Sample
- di-b-jet Ntuple production

Standard Dijet Resonance Cuts
- Leading Jet pT > 410 GeV
- Sublead Jet pT > 50 GeV
- $|y^*| < 0.6$
- mjj > 1100 GeV

Using fixed cut 85% for both jets.
- mbb_fix_8585

Cone matching truth flavour
- jetHadronConeExclTruthLabelID

Work Flow

phys-exotics/jdm/dijet/inputs/Btag/MC15_DiJet_20151104

Packages used

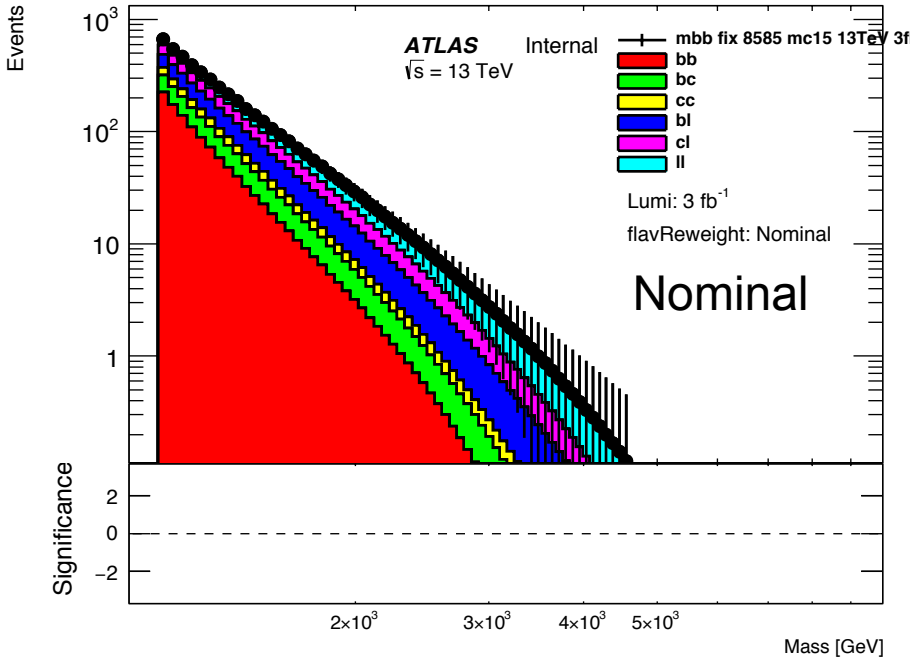
- DijetHelpersPackage:
=> Create scaled distributions.
=> Vary flavour fractions.
=> Create p-values of fit.

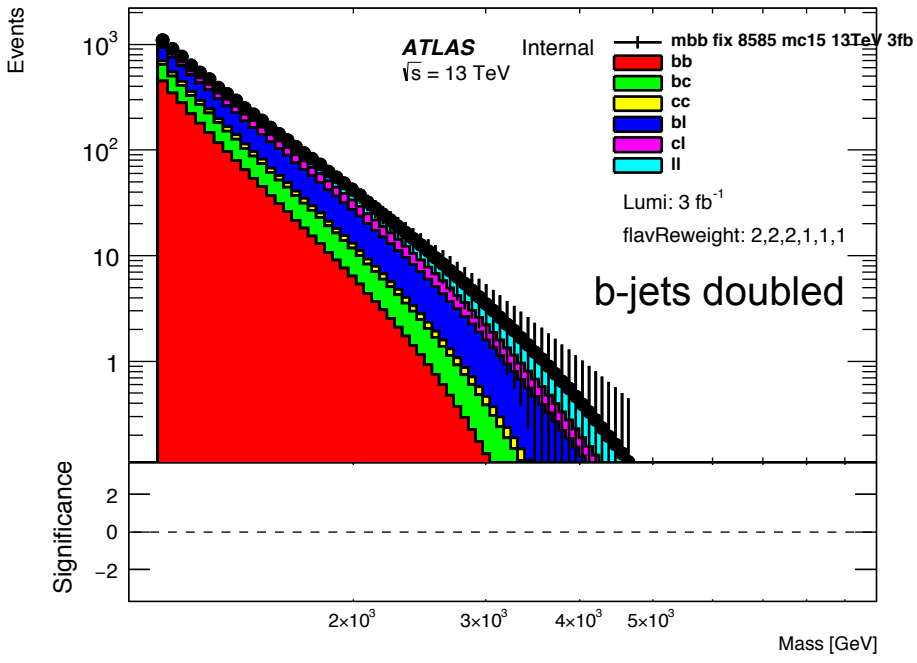
- Dijet Statistical Packages:
=> Using search phase from this package
=> Bumphunter to search for discrepant regions
=> Spurious signal check

Stacking the Flavour Fractions

- Create flavour fractions
  => Creating histograms from fits to 20 fb^{-1} scaled to 3 fb^{-1}
  => Adding the fractions in different ways to produce various spectra

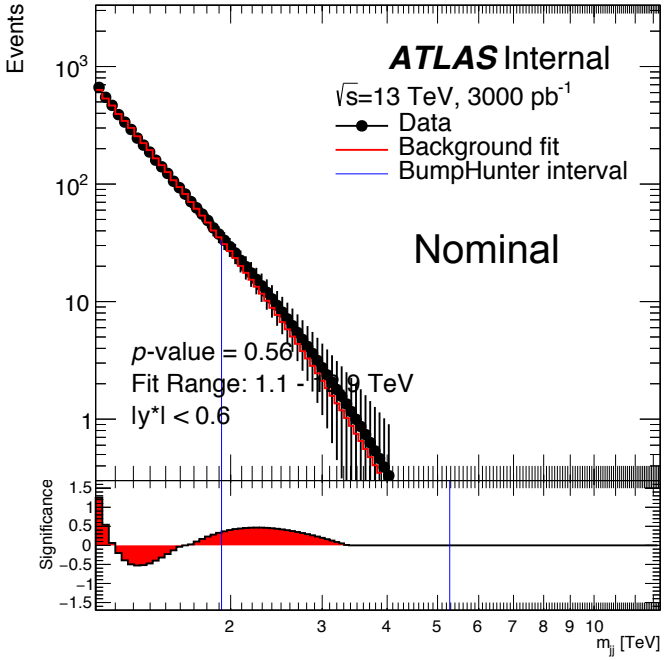
- This creates new scaled like distributions.

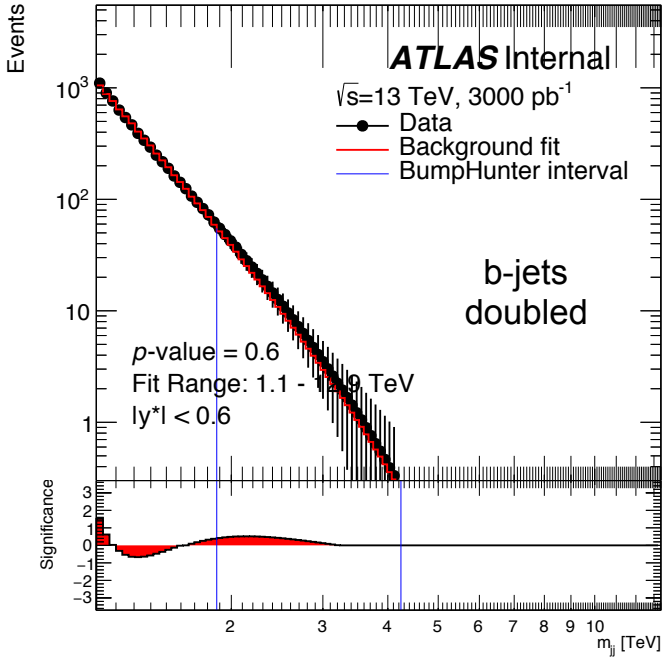


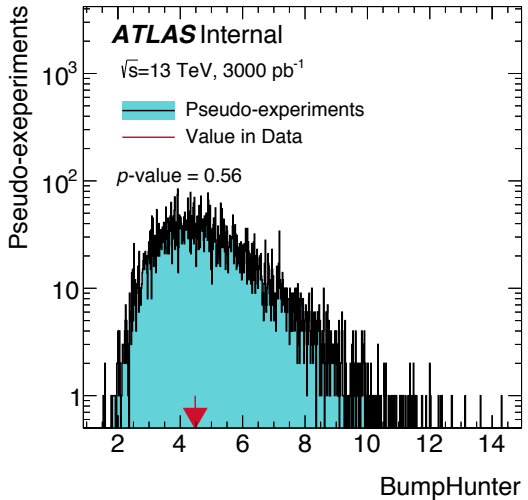


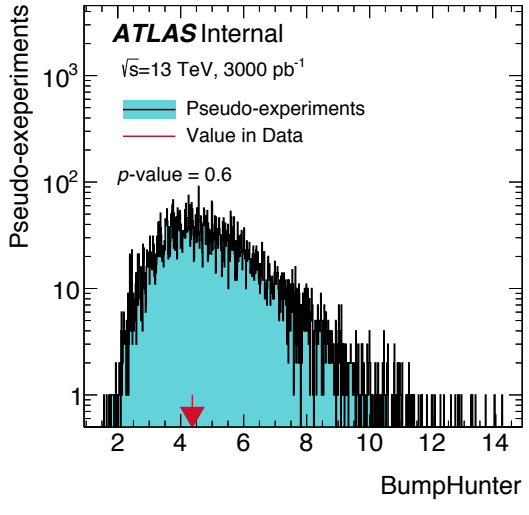
Looking for a spurious signal

- We then fit to this scaled using 3-parameter fit function
- Run bumphunter to look for significant deviations



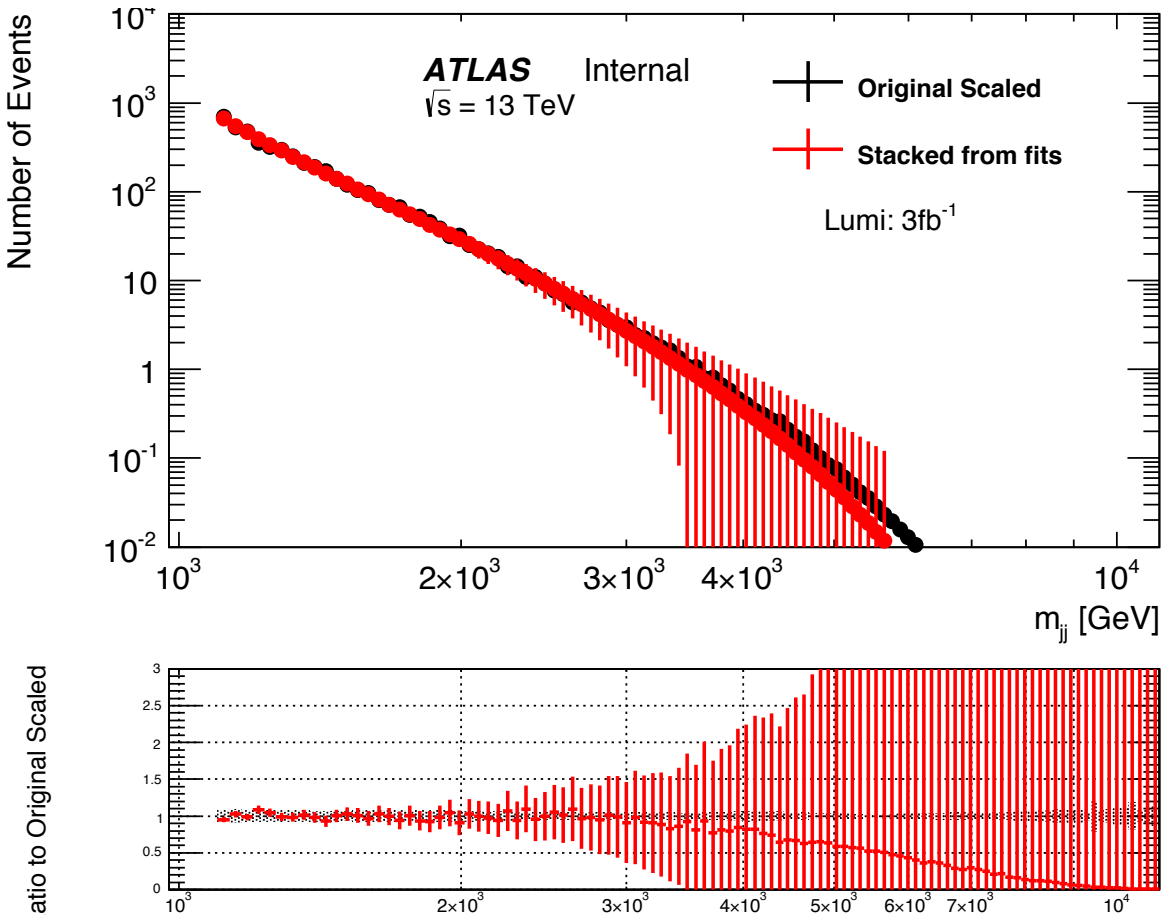






Looking for a spurious signal

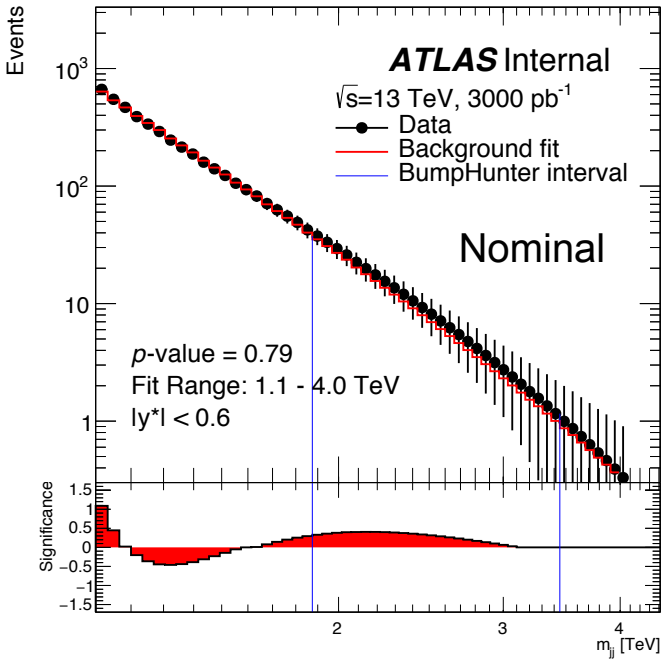
- However I know, that stacking doesn't match MC above 4 TeV

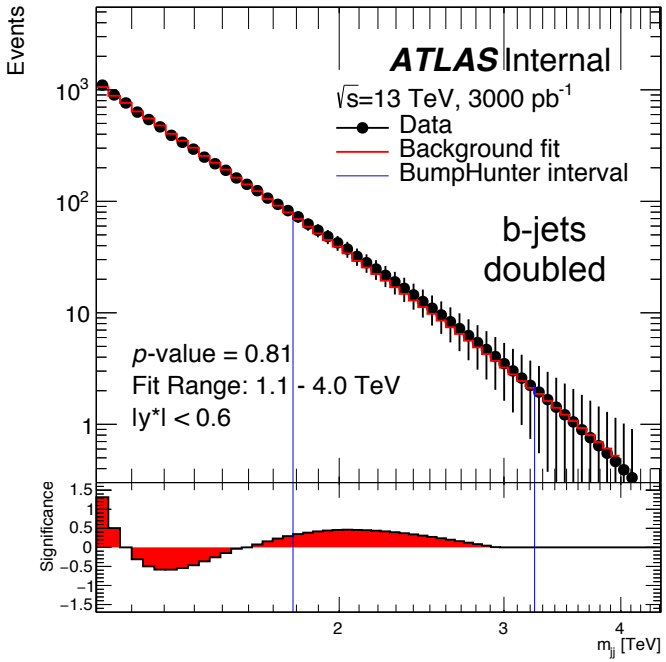


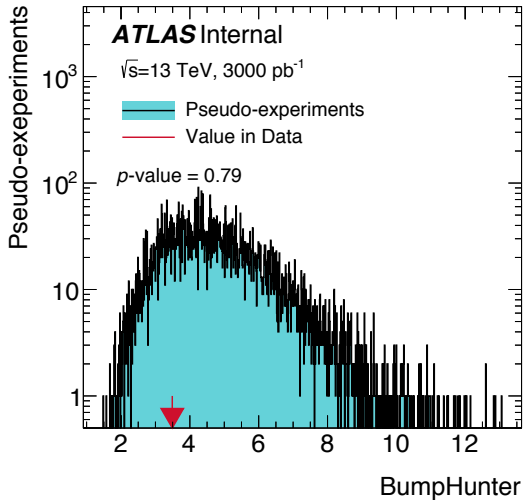
- So I will also try cutting off at 4 TeV (and 5 TeV)

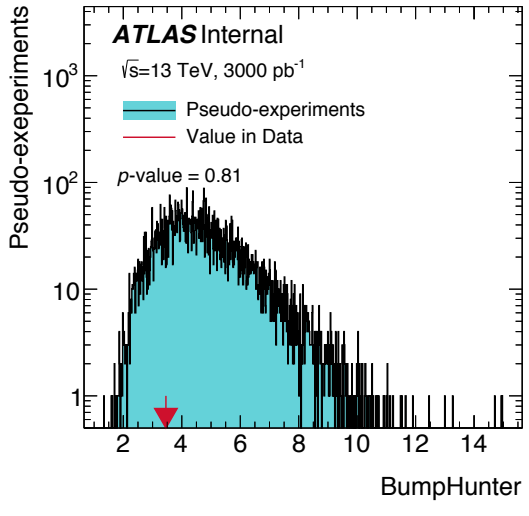
Looking for a spurious signal

- Cut off at 4 TeV









Looking for a spurious signal

- Cut off at 5 TeV == To be added!

Nominal

b-jets
doubled

Conclusions

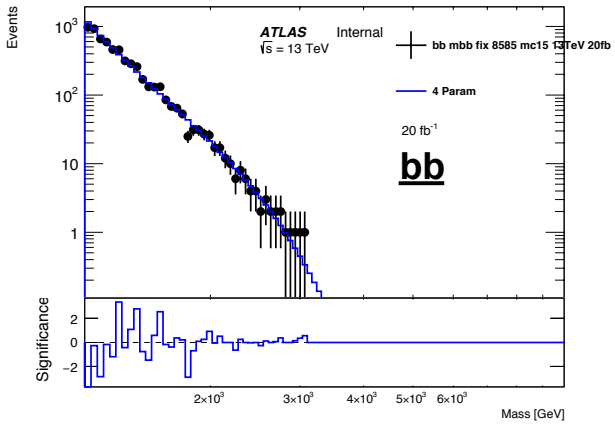
- **Spurious Signal Tests**
  - Check scaled dijet mass spectrum for any large deviations.
  - Cutting scaled spectra off at 4 TeV
  - Bumphunter values:
    - Nominal = 0.79
    - b-Jet Content = 0.81
  - No obvious problem here.
  - This is documented and ready to go in note…

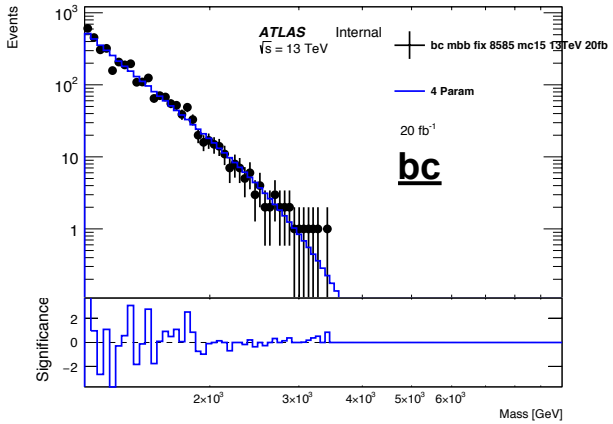
- **Also done last week: p-Value of fitting function**
  - Fitted to ‘data-like’ distributions.
  - We see no drop in performance (p-value) in the case where b-jet content is doubled.
  - Evidence that fit is robust to flavour fraction.
  - Systematic from fit parameters and fit function choice are enough.
  - This is documented and in note already

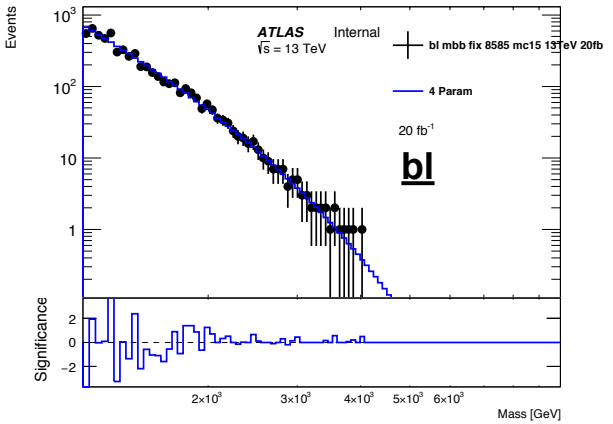
Backup:
p-Value of Fits Study

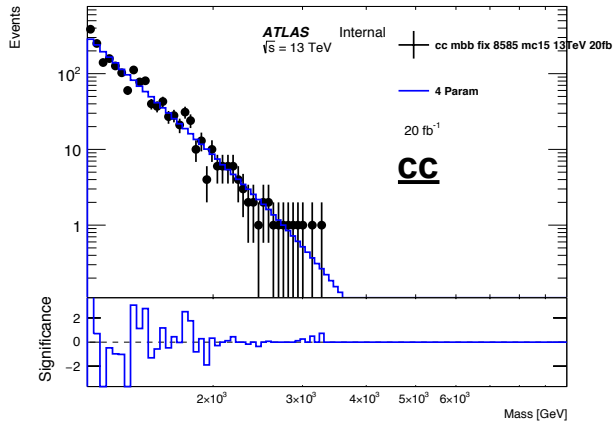
Getting the Flavour Fractions

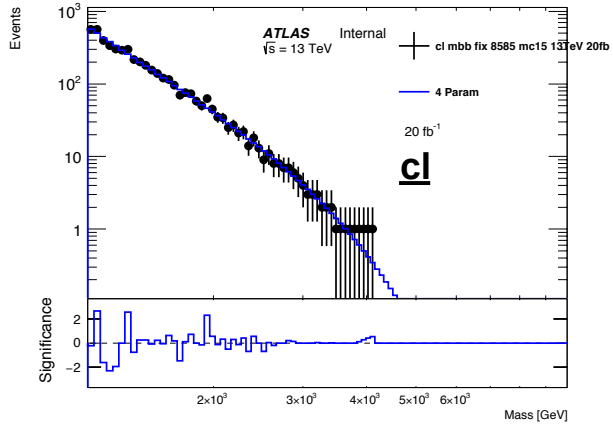
- Flavour fractions are extracted from MC using truth information.
- The dijet mass spectrums for these flavour fractions are then scaled to 20fb^{-1}
- The dijet mass spectrums are fitted to using the 4-parameter fit function.

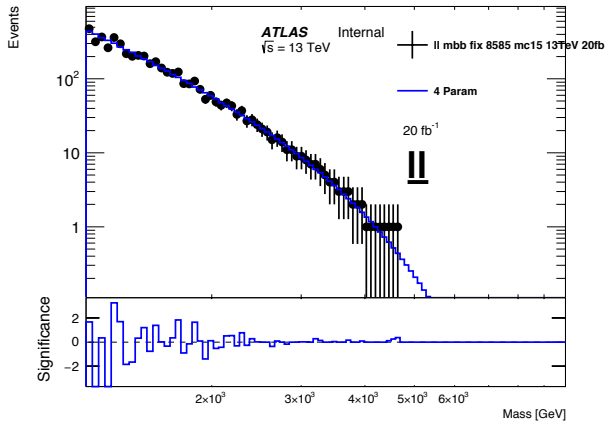








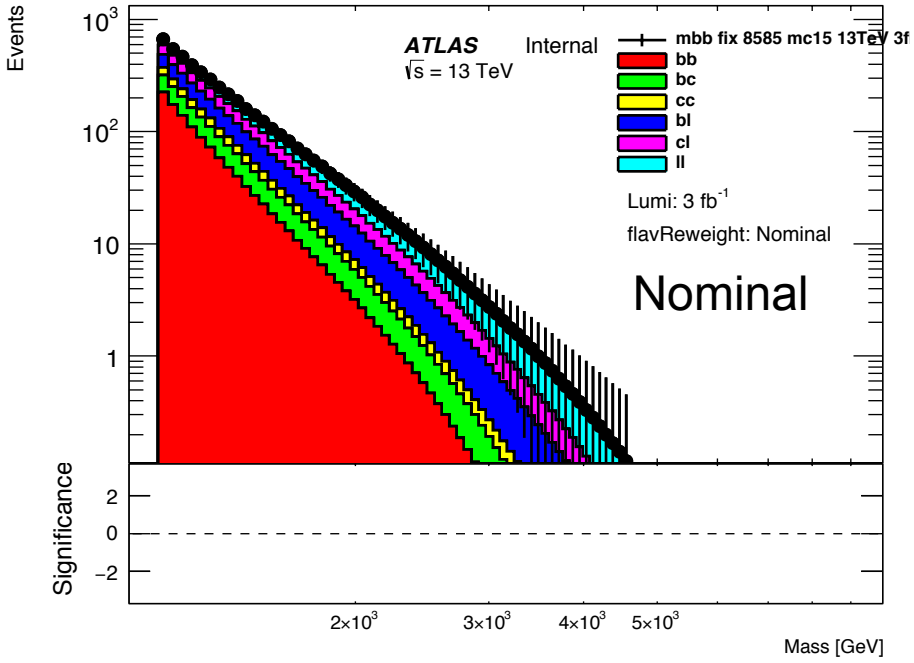


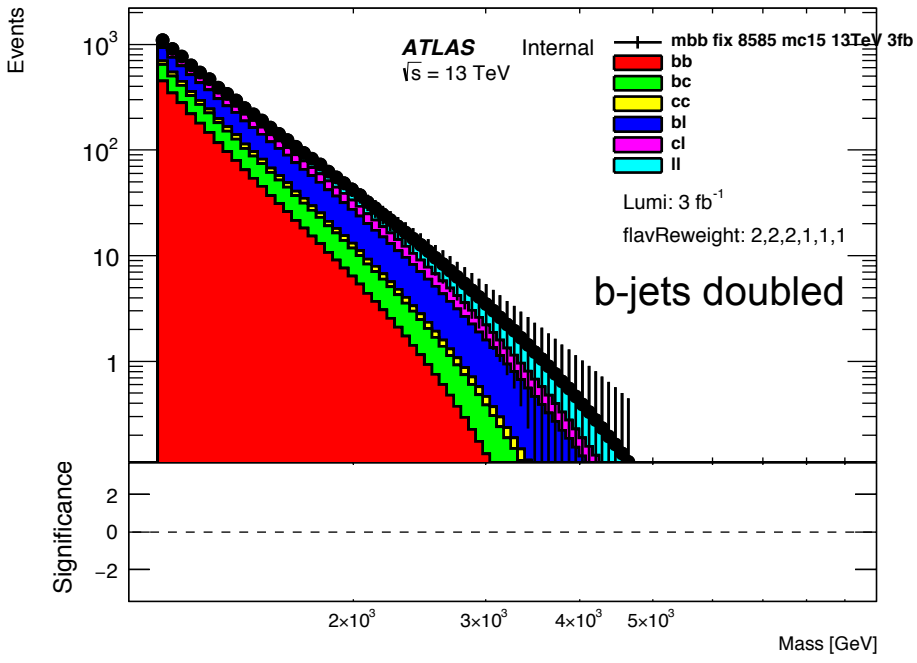


Stacking the Flavour Fractions

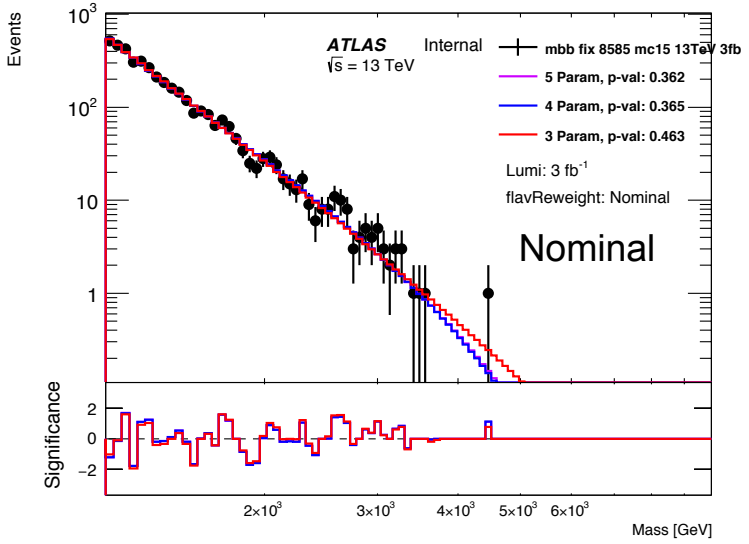
- Create flavour fractions
  => Creating histograms from fits to 20 fb^{-1} scaled to 3 fb^{-1}
  => Adding the fractions in different ways to produce various spectra

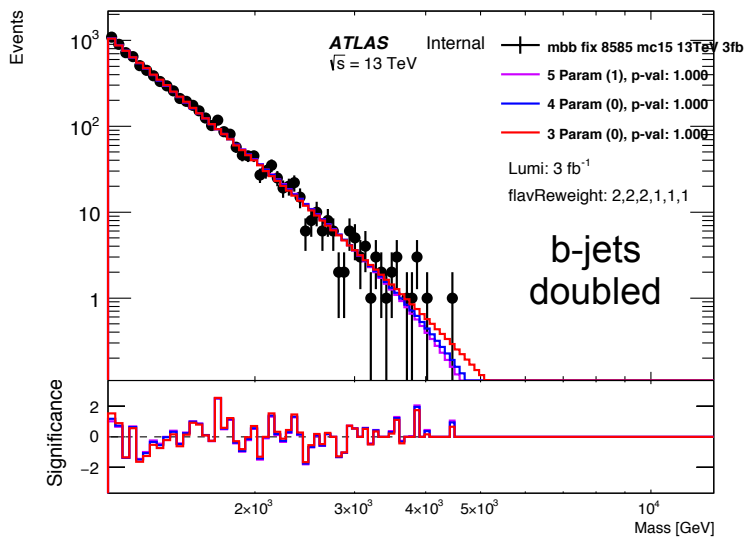
- This creates new scaled like distributions.





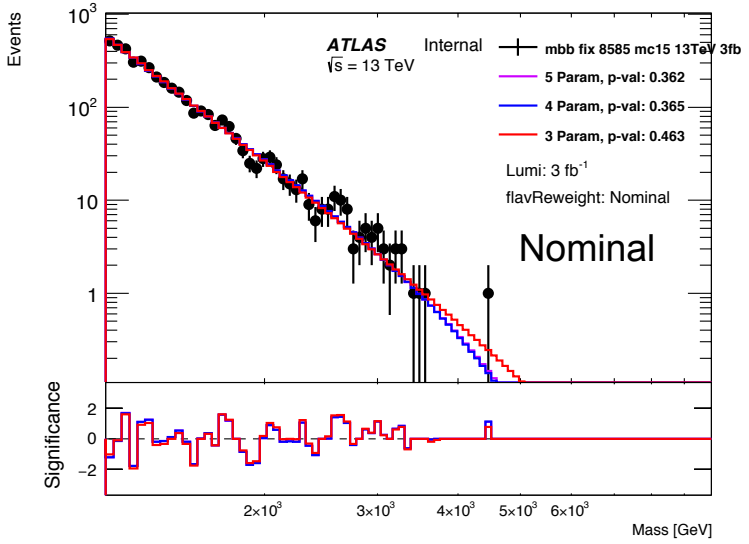
- By applying poisson fluctuations we can create ‘data-like’ distribution
- These are fitted using the 3, 4 and 5 parameter fit function

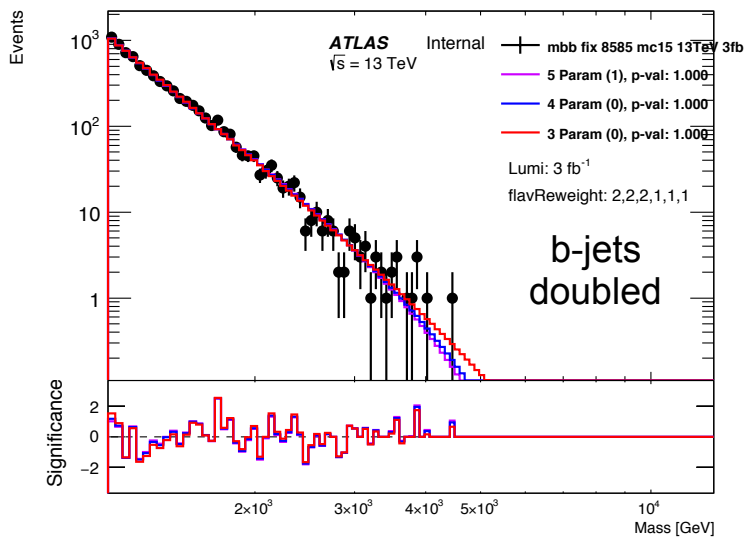




Making Data-Like

- By applying poisson fluctuations we can create ‘data-like’ distribution
- These are fitted using the 3, 4 and 5 parameter fit function

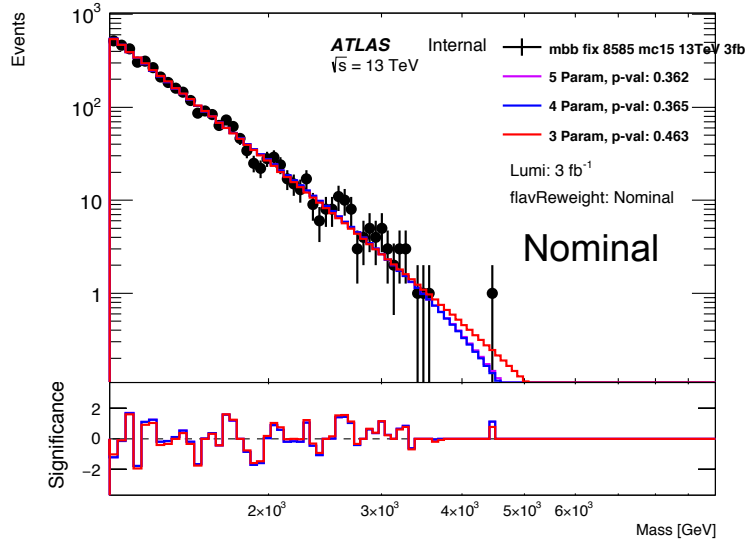


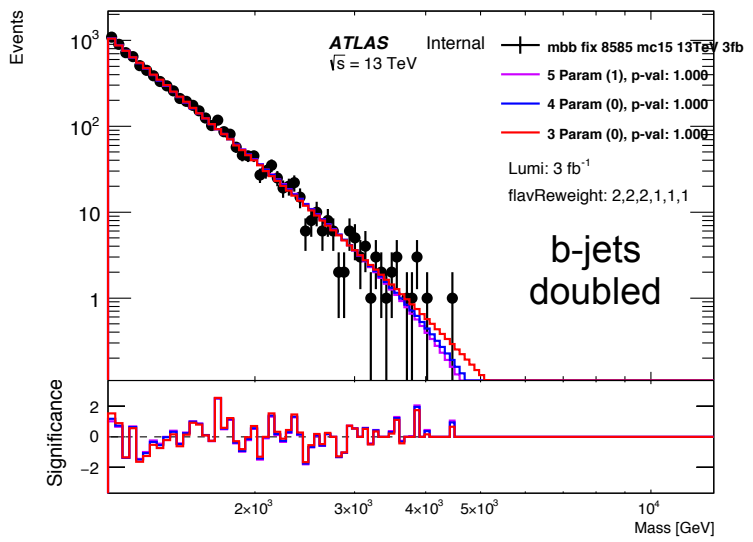


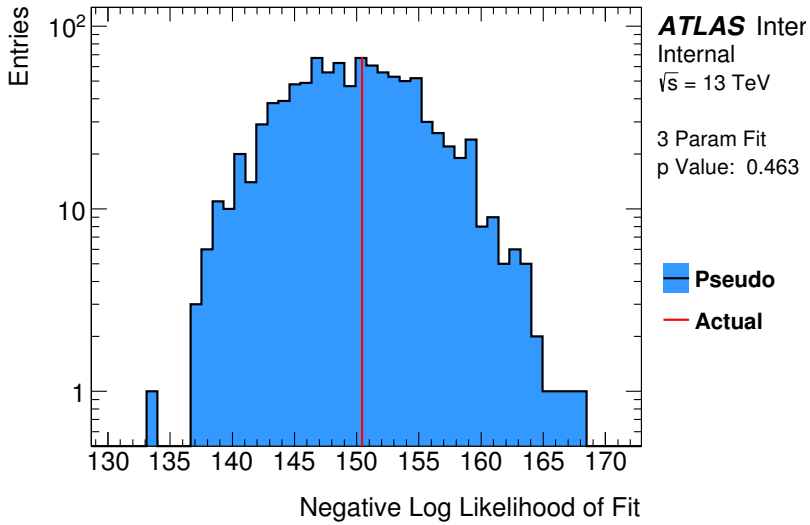
To calculate p-value of a fit:

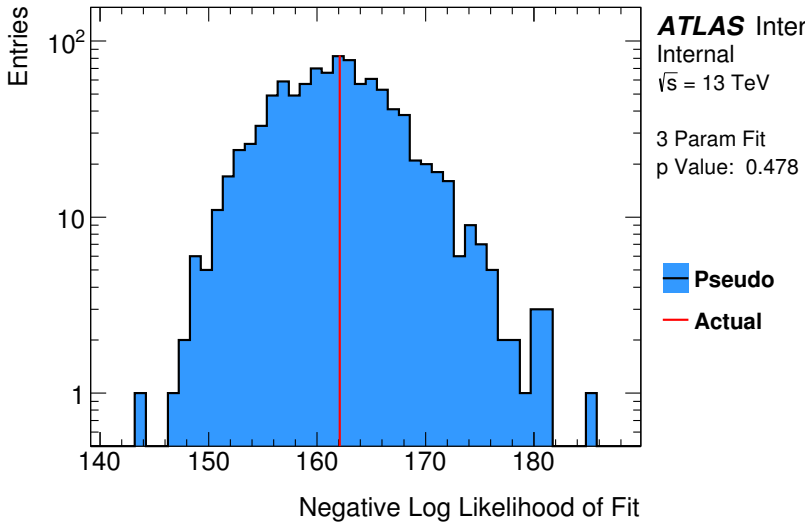
1. Take the fit function and apply poisson fluctuations. (Pseudo-experiment)
2. Re-fit to the pseudo-data using the same fit function.
3. Compare quality of fit to pseudo-experiment to that of the original fit.
   - For a measure quality of fit I use negative log likelihood
4. Repeat 1000 times and count fraction of pseudo-experiments that have a worse quality of fit than the original fit.

- By applying poisson fluctuations we can create ‘data-like’ distribution
- These are fitted using the 3, 4 and 5 parameter fit function



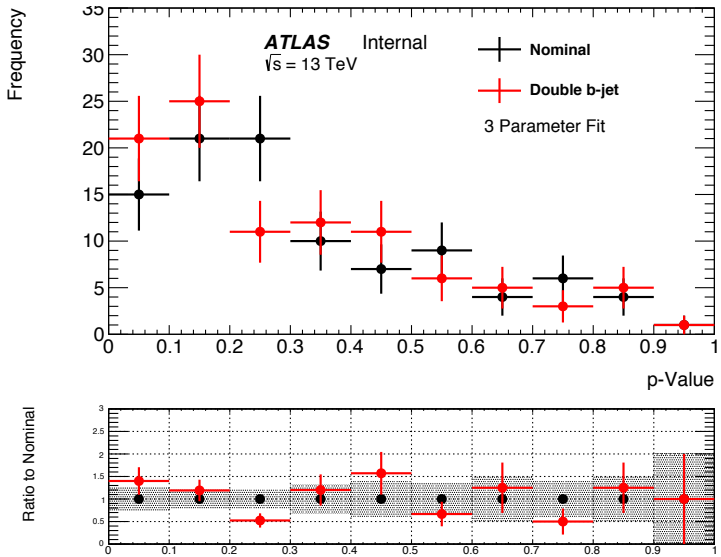


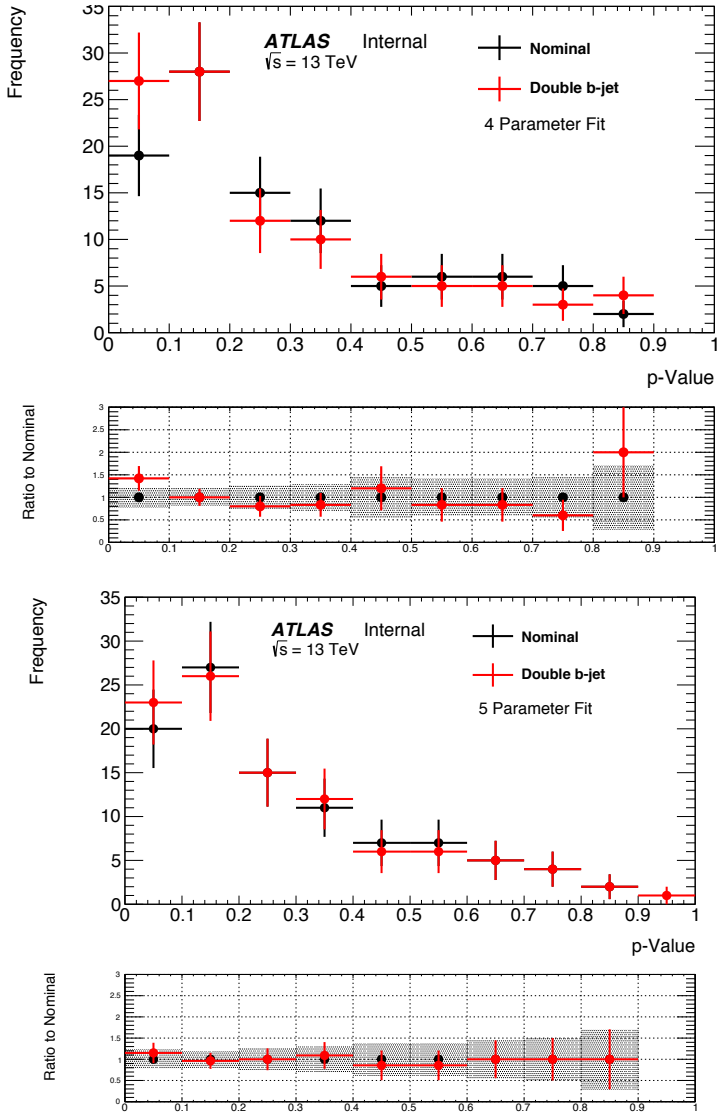




Many More Data-Like

- Different sets of poisson fluctuations means a different ‘data-like’ spectrum
- Each ‘data-like’ dist. can be fitted to, giving a different p-value for each fit variation.
- 100 different data-like distributions have been studied





Mean p-values

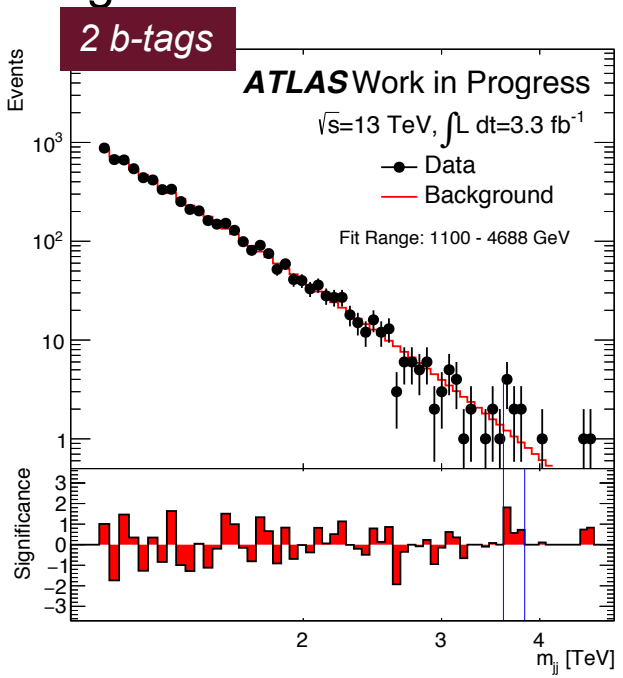
	3-Para. Fit	4-Para. Fit	5-Para. Fit
Nominal	0.325 +/- 0.024	0.280 +/- 0.023	0.283 +/- 0.022
b-jet Doubled	0.308 +/- 0.024	0.267 +/- 0.022	0.276 +/- 0.022

Backup:
b-Tagged Dijet Analysis

- **Fitting Function**
  - We use a smoothly falling fitting function to fit to background

$$f(x) = p_1(1-x)^{p_2}(x)^{p_3+p_4\ln x+p_5\ln x^2}$$

where, $x = m_{jj}/\sqrt{s}$



  - This comes in a 3, 4 or 5 parameter versions
    - Setting p_4 and p_5 to zero

- **Varying Flavour Composition**
  - It is known that the fitting function can fit to the Monte-Carlo
  - However MC is not a perfect prediction of reality
  - What if there are more *b*-jets in the data than in the MC
    - Can we still fit to data in this case? Is our fitting function robust.