Flavour Composition Studies:
Quality of Fit Discussion
Spurious Signal Discussion

Laurie McClymont,
di-b-jet Analysis Team

Di-b-jet Meeting

17 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.

- **A) Response to Fabrizio's comment @ Ed Board Meeting 10/02/2016**
  => Currently NLL is used as a measure of QoF in calculations of p-value = Not Valid
  => Our proposal, still fit by minimising NLL (Our fit procedure is the same)
  => Use Chi2 as quality of fit, still calculate p-values using toys.

- **B) Response to Klaus' comment @ Ed Board Meeting 10/02/2016**
  => Fitting to scaled distribution is not enough to show no spurious signal.
  => We have injected poisson noise, and then searched for spurious signal.
  => No discrepant number of low bumpHunter p-values.

Details

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

Standard Dijet Resonance Cuts

Using fixed cut 85% for both jets = mbb_fix_8585

Cone matching truth flavour
- jetHadronConeExclTruthLabelID

Work Flow

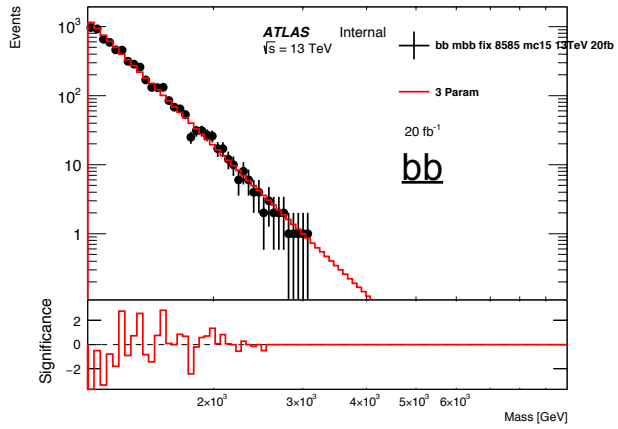
- DijetHelpersPackage:
=> Create scaled and data-like distributions.
=> Vary flavour fractions.
=> Create p-values of fit.

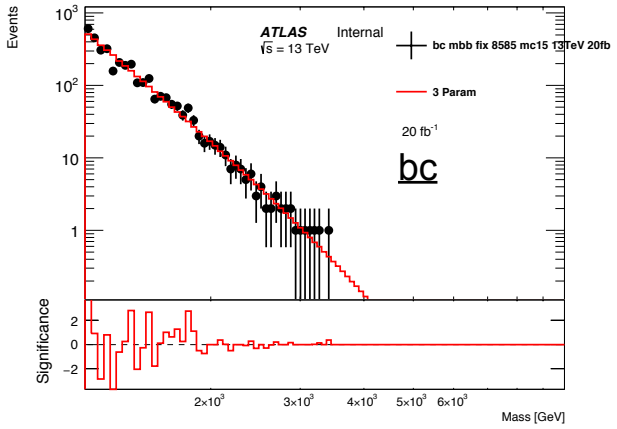
- Dijet Statistical Packages:
=> Using search phase from this package
=> Bumphunter to search for discrepant regions
=> Spurious signal check

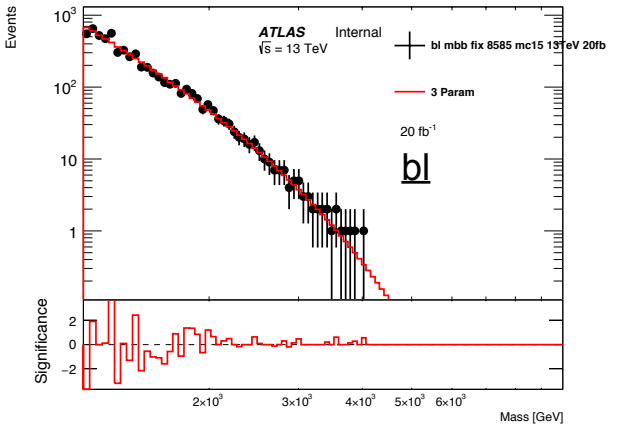
A) Chi2 used for QoF

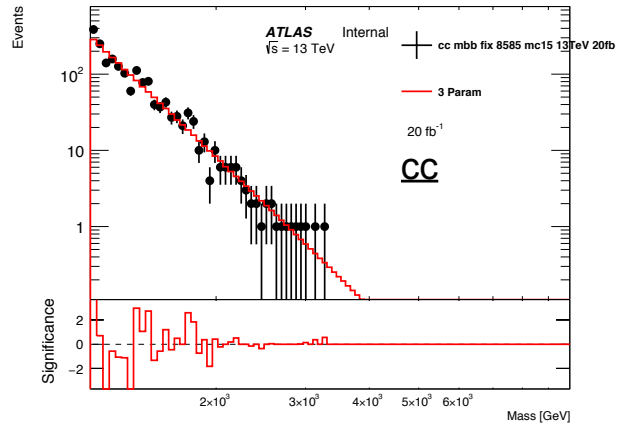
Getting the Flavour Fractions

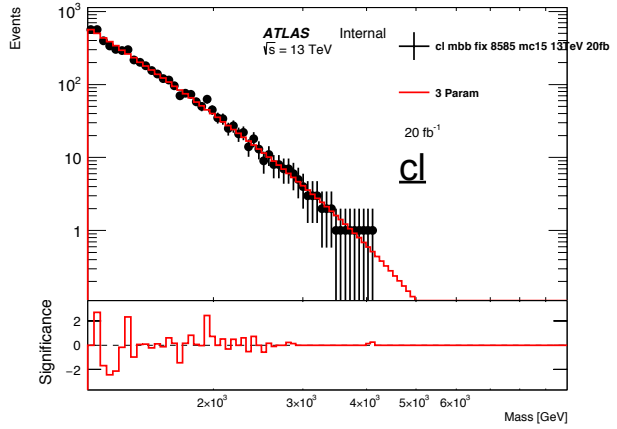
- Flavour fractions are extracted from MC using truth information
- The dijet mass spectrums for these flavour fractions are then scaled to 20fb^{-1}
- The dijet mass spectrums are fitted to using the **3-parameter** fit function.

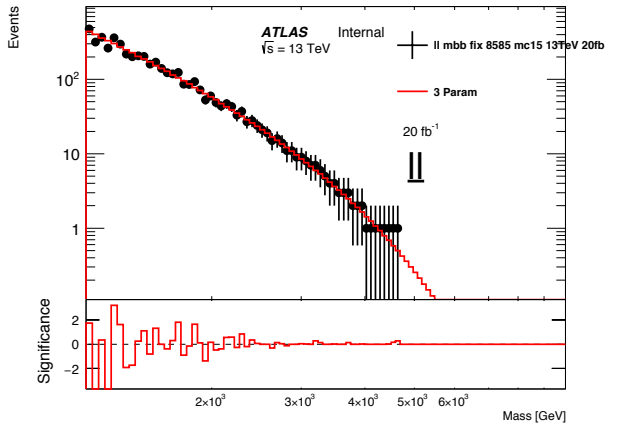






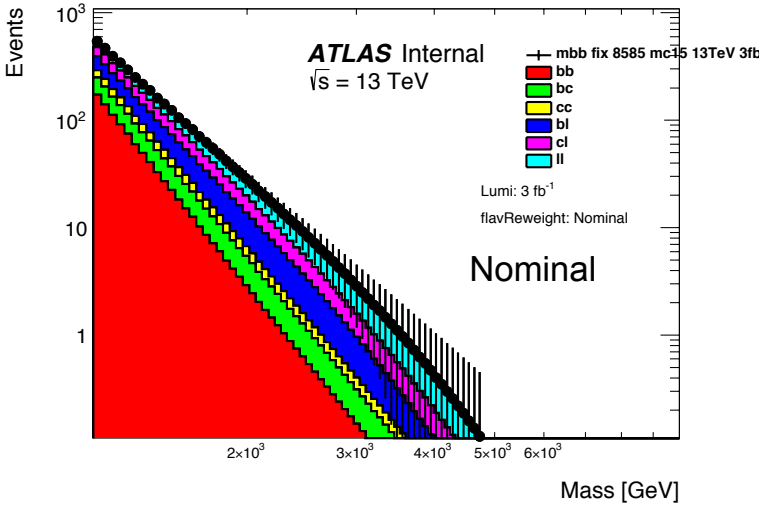


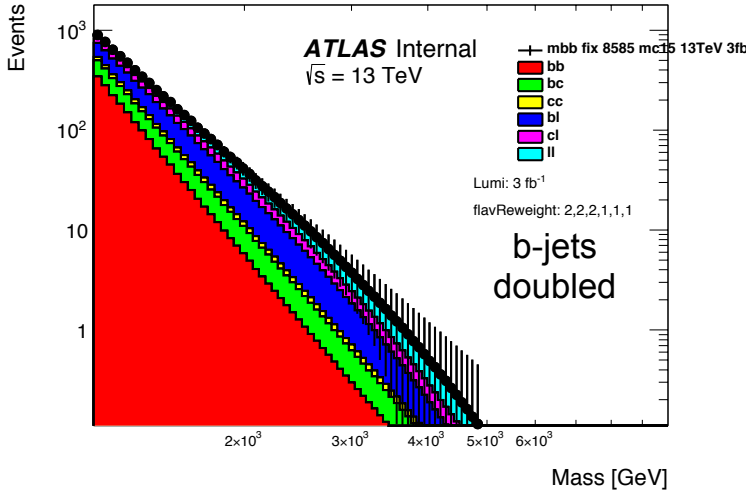


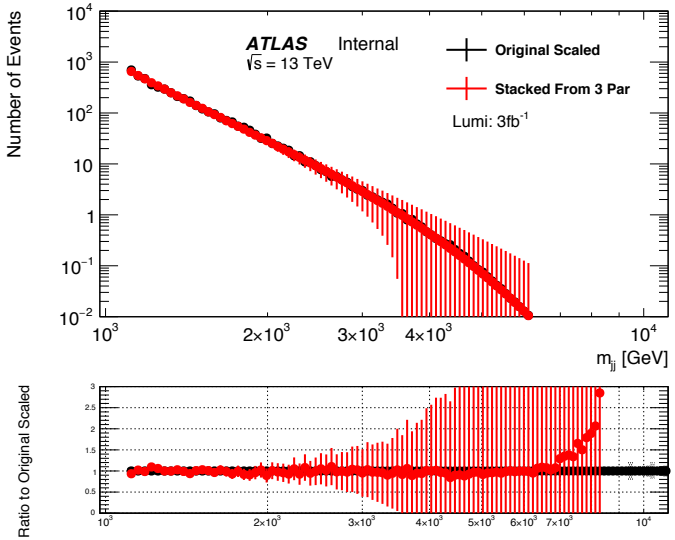


Stacking the Flavour Fractions

- Creates new scaled like distributions.
  => Adding templates from fits to 20 fb^{-1} scaled to 3 fb^{-1}
  => Adding the fractions in different ways to produce various spectra



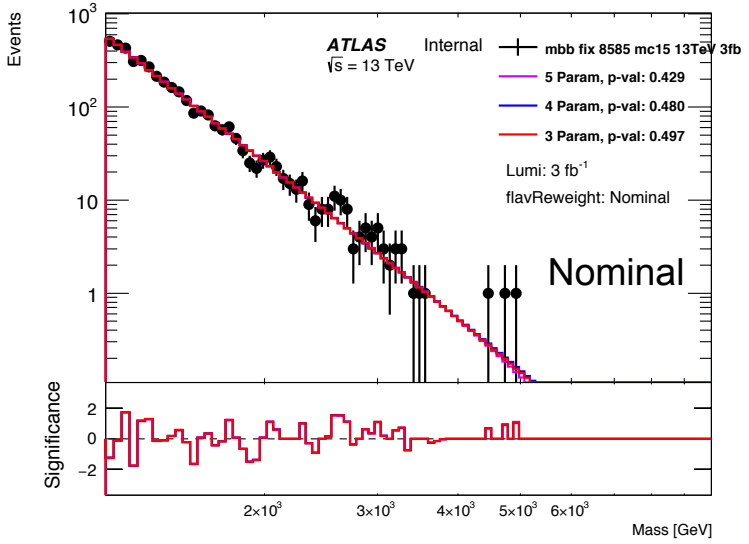


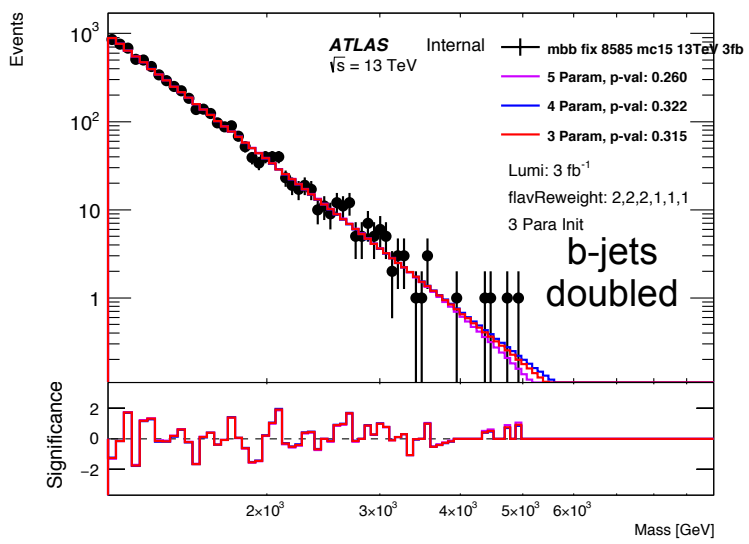


- Able to reproduce MC up to large masses.
- We conclude that the procedure of fitting to flavour fractions then stacking is appropriate.

Making Data-Like - Negative Log Likelihood

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



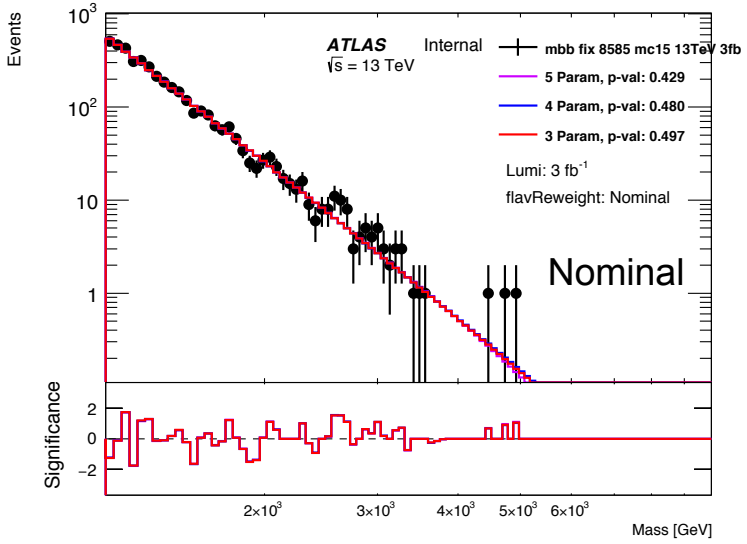


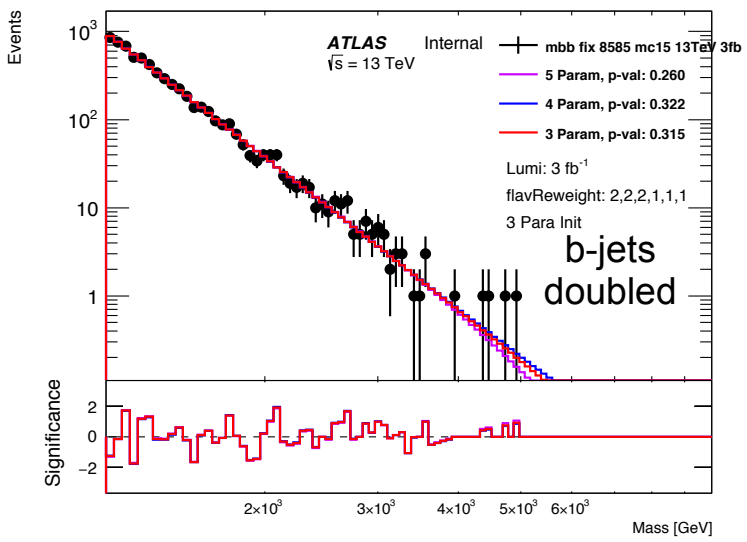
To calculate p-value of a fit:

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 we use negative log likelihood** **(Not Valid)**
4. Repeat 1000 times and count fraction of pseudo-experiments that have a worse quality of fit than the original fit.

Making Data-Like - Now using Chi²

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



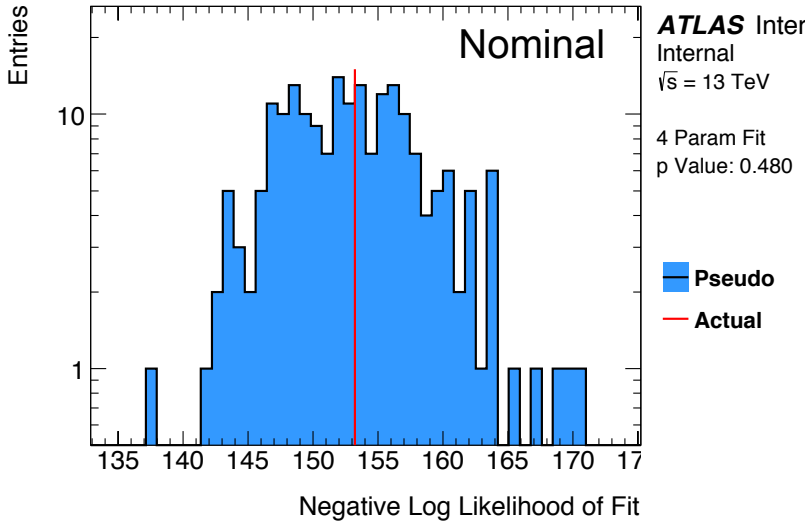


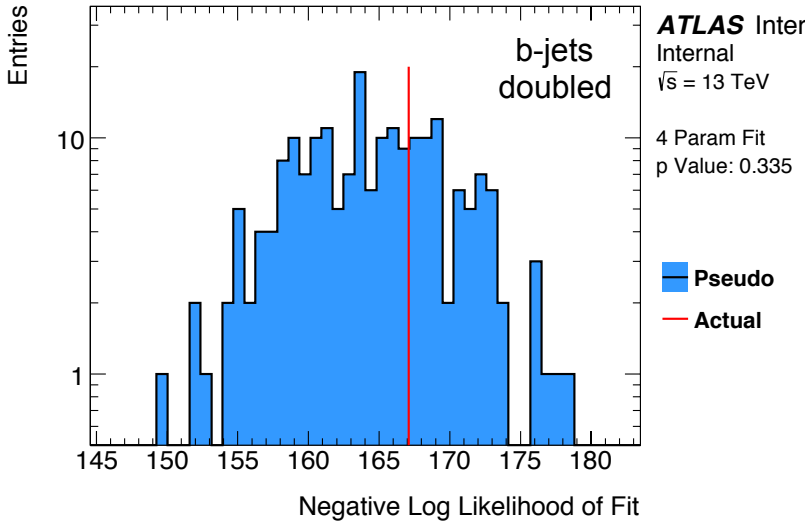
To calculate p-value of a fit:

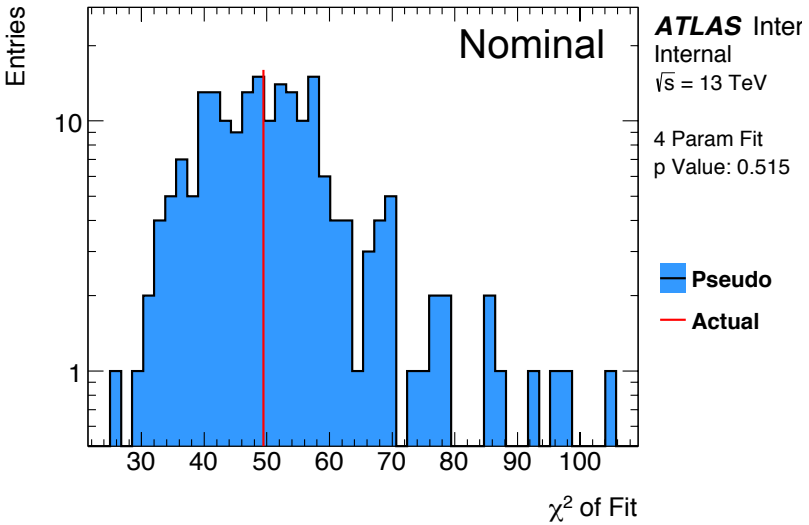
1. Take the fit function and apply poisson fluctuations. (Pseudo-experiment)
2. Re-fit to the pseudo-data using the same fit function.
3. Compare quality of fit to pseudo-experiment to that of the original fit.
   - **For a measure quality of fit we now use Chi²**
4. Repeat 1000 times and count fraction of pseudo-experiments that have a worse quality of fit than the original fit.

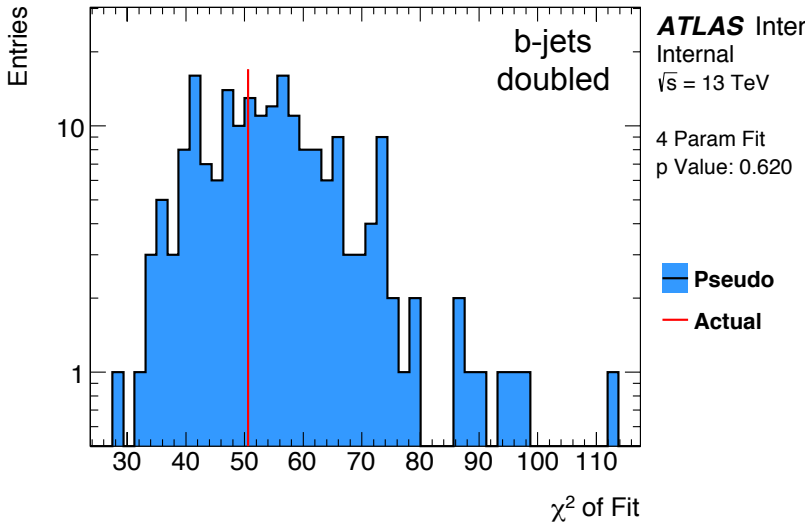
Data-Like p-Values

- By applying poisson fluctuations we can create ‘data-like’ distribution
- Fitted to using the 3 parameter by minimising NLL, this fixes parameters.
- We then use 200 toys and Chi^2 as QoF test. (200 allows us to repeat this more...)



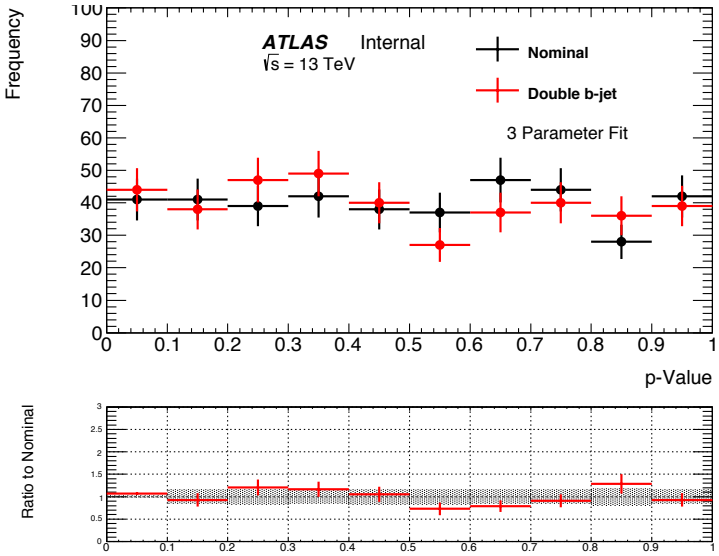


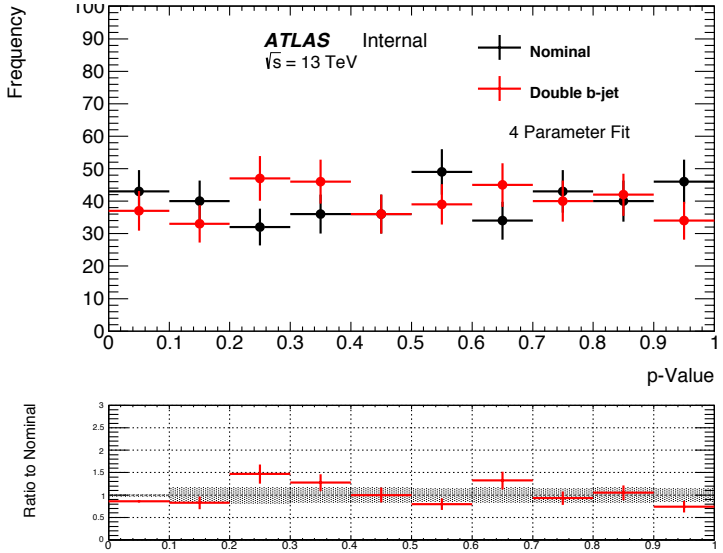


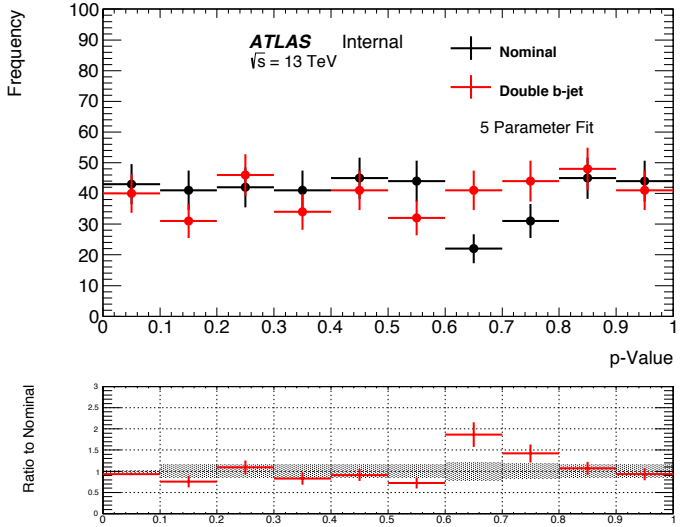


p-Value Studies

- 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.
- 400 different data-like distributions have been studied





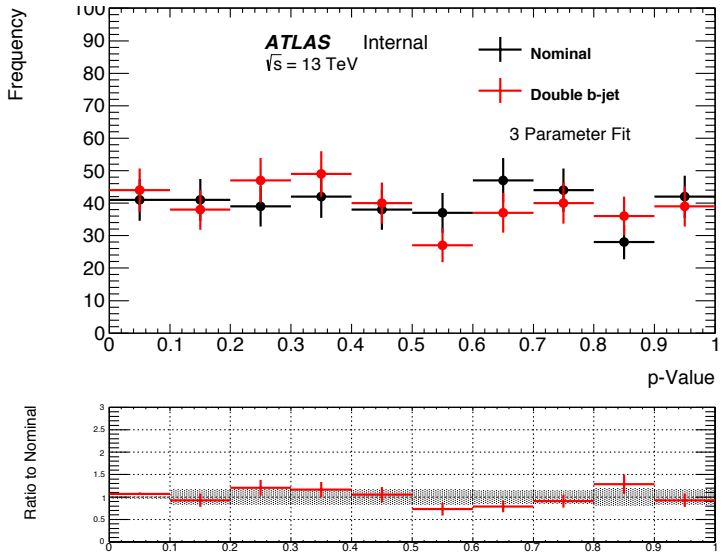


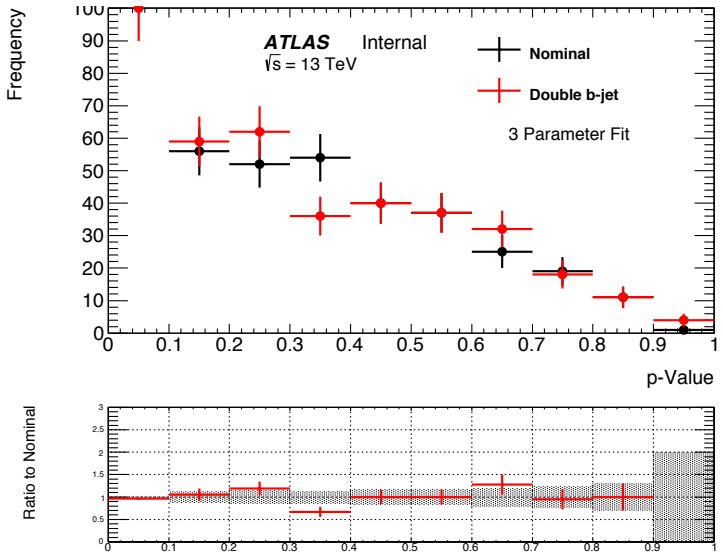
Mean p-values

	3-Para. Fit	4-Para. Fit	5-Para. Fit
Nominal	0.492 +/- 0.014	0.508 +/- 0.015	0.488 +/- 0.015
b-jet Doubled	0.478 +/- 0.014	0.495 +/- 0.014	0.514 +/- 0.015

p-Value Studies - Comparison with NLL

- 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.
- 400 different data-like distributions have been studied





Chi2

	3-Para. Fit
Nominal	0.492 +/- 0.014
b-jet Doubled	0.478 +/- 0.014

NLL

	3-Para. Fit
Nominal	0.305 +/- 0.012
b-jet Doubled	0.311 +/- 0.012

Conclusions

- **Change to Chi2 for QoF**
  - NLL is not a good measure of QoF, so no longer used to calculate p-values.
  - Minimising NLL can be used to choose and fix fitting parameters.
  - We now use Chi^2 as a measure of QoF.
  - p-Values can be calculated by throwing toys and comparing Chi^2 values.

- **Changes to Statistics**
  - We increased the number of 'data-like' distributions studied (100 => 400)
  - This allows us to compare distributions of p-values better.
  - To allow this, reduce the number of toys from 1000 => 200

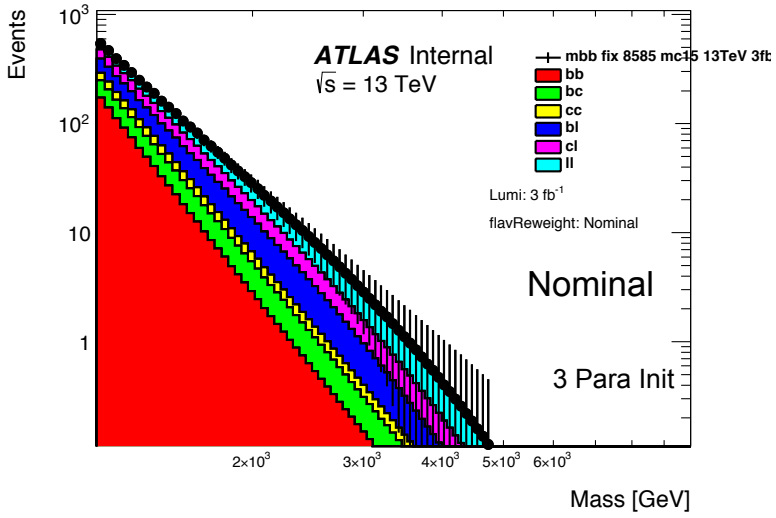
- **Results**
  - We see that the Chi^2 produces flat distributions of p-values (slide 8)
  - There is large discrepancy of mean p-values between Nominal and b-Jet Content doubled
  - Conclusions of study should not be changed
    - Fit is robust to flavour fraction changes
    - No need for additional

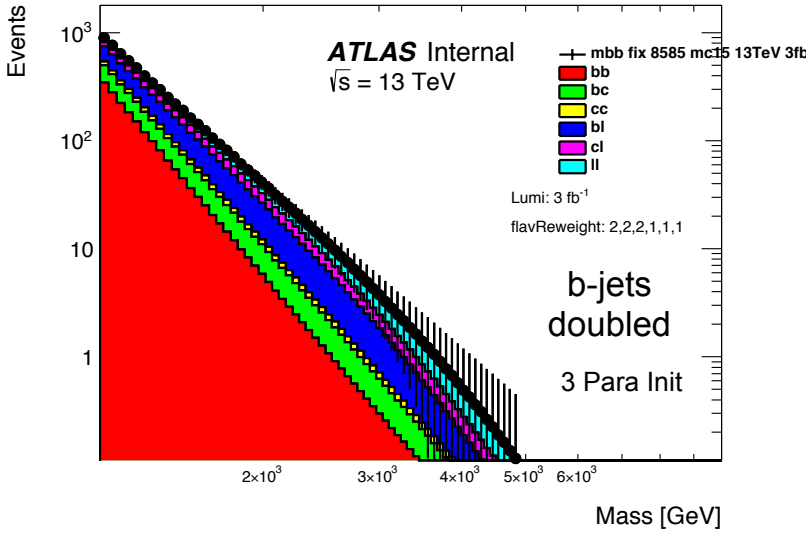
- **Documented in SVN**

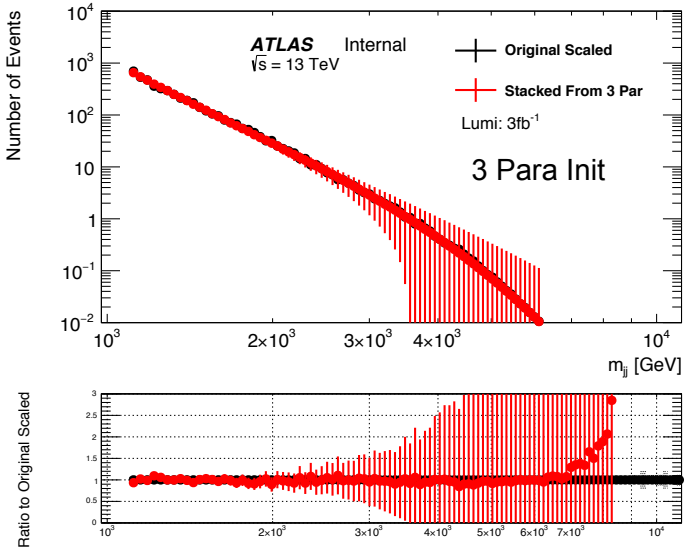
B) Spurious Signal Check

Previous Test for Spurious Signal - Intro

- New for v1.5



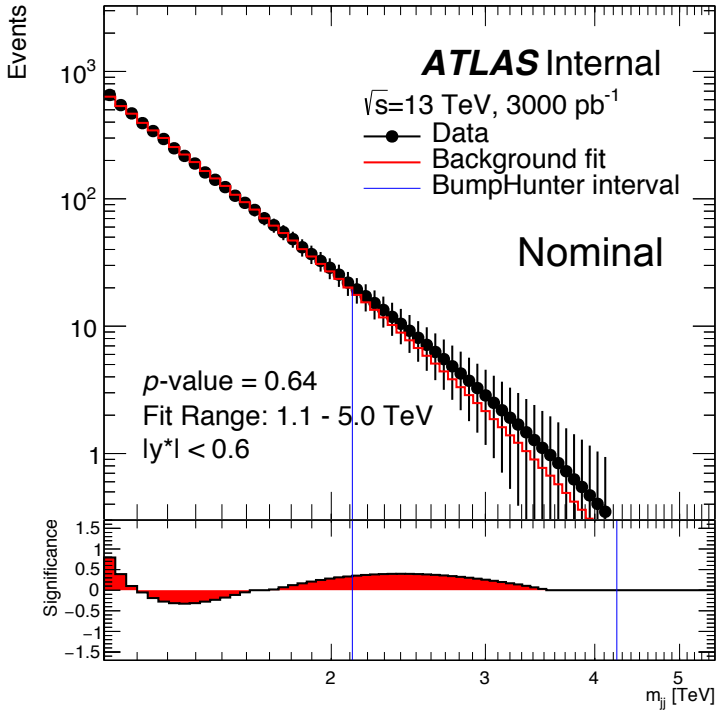


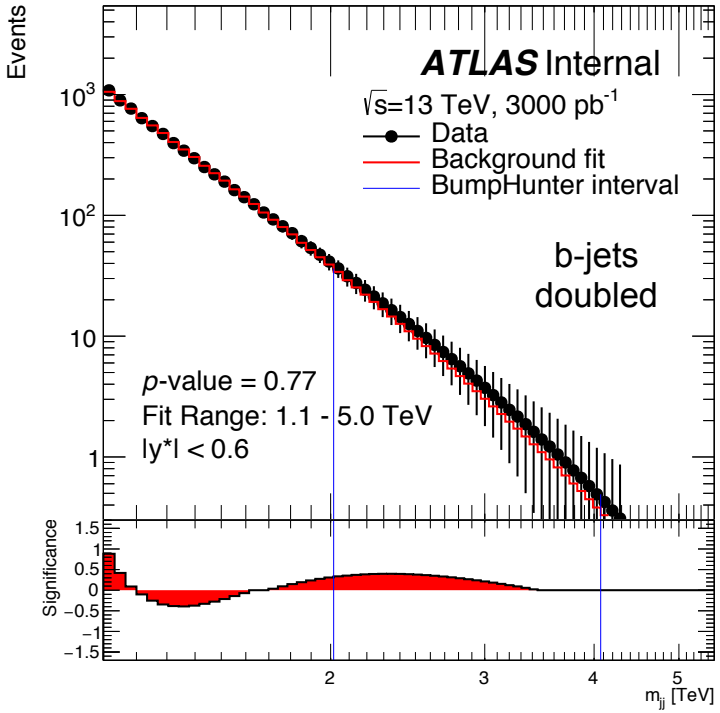


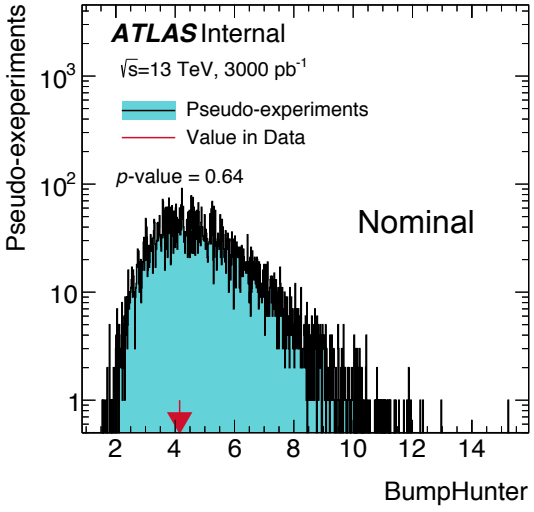
- Test for spurious signal
  => Use scaled spectra before Poisson noise
  => Fit to this spectra using 3 par. function
  => BumpHunter will identify discrepant region
  => BH then can calculate p-value

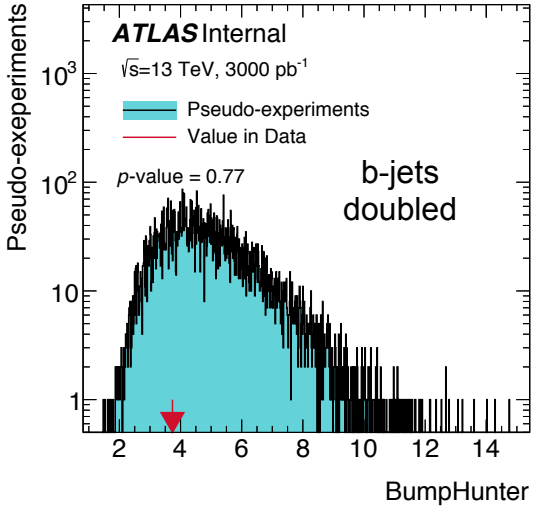
- Mass Range of Fit
  => 1.1 - 5 TeV
  => Larger than mass range in data.

Previous Test for Spurious Signal - Results









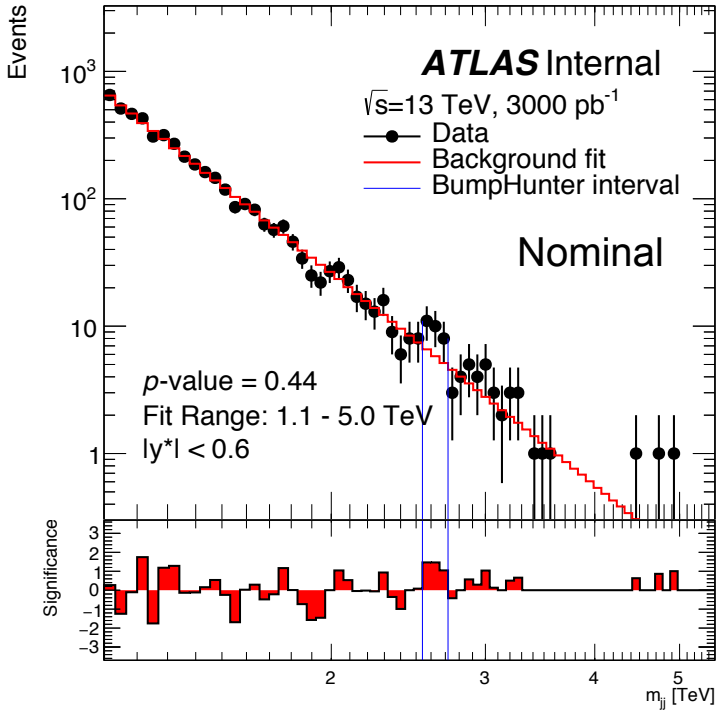
- No significant spurious signal found.
- Consistent p-Value in both flav. composition cases
- Wide discrepant region => Unlike benchmark models

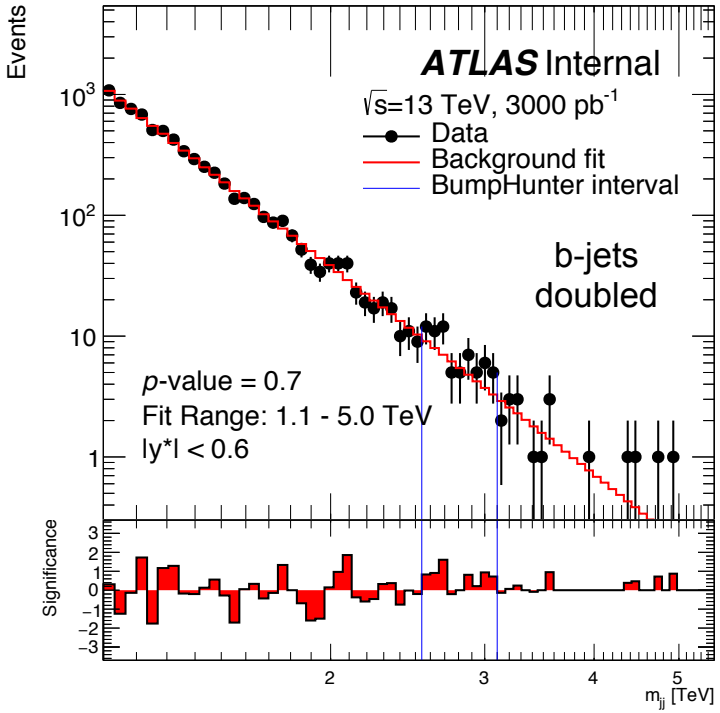
Updated Test for Spurious Signal - Intro

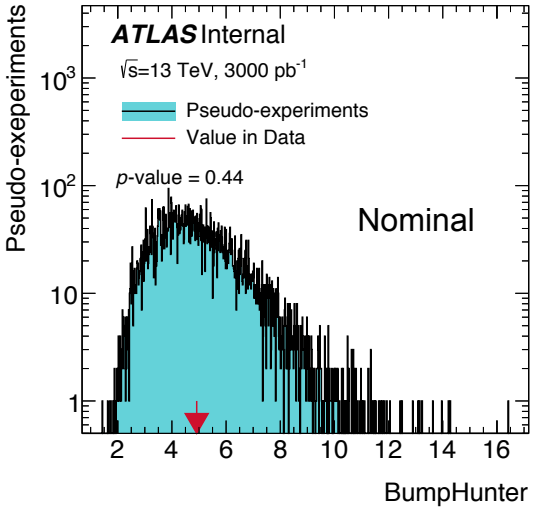
- **Use ‘data-like’ distributions**
  - Poisson fluctuate scaled distribution to create ‘data-like’ distribution
  - Then run BumpHunter on the ‘data-like’ distribution
  - Calculate a p-value of most discrepant region.

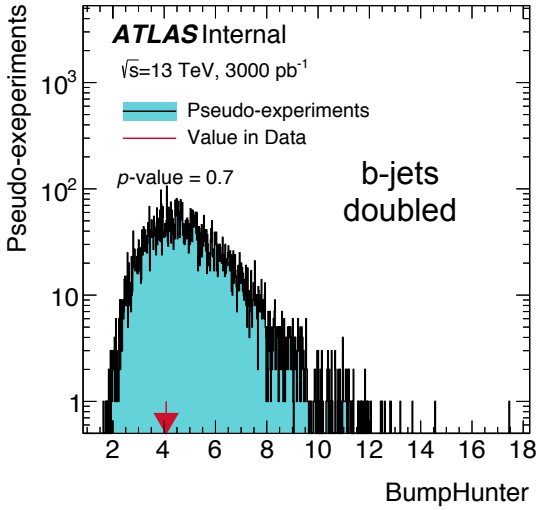
- **Look for trends in p-values**
  - Repeat process above 200 times
  - Study distribution of p-values in Nominal case and double b-jet case
  - Look for any trends in BumpHunter p-value
    - Are deviations between fit and truth causing a bias at low p-values?

Update Test for Spurious Signal - Single Seed





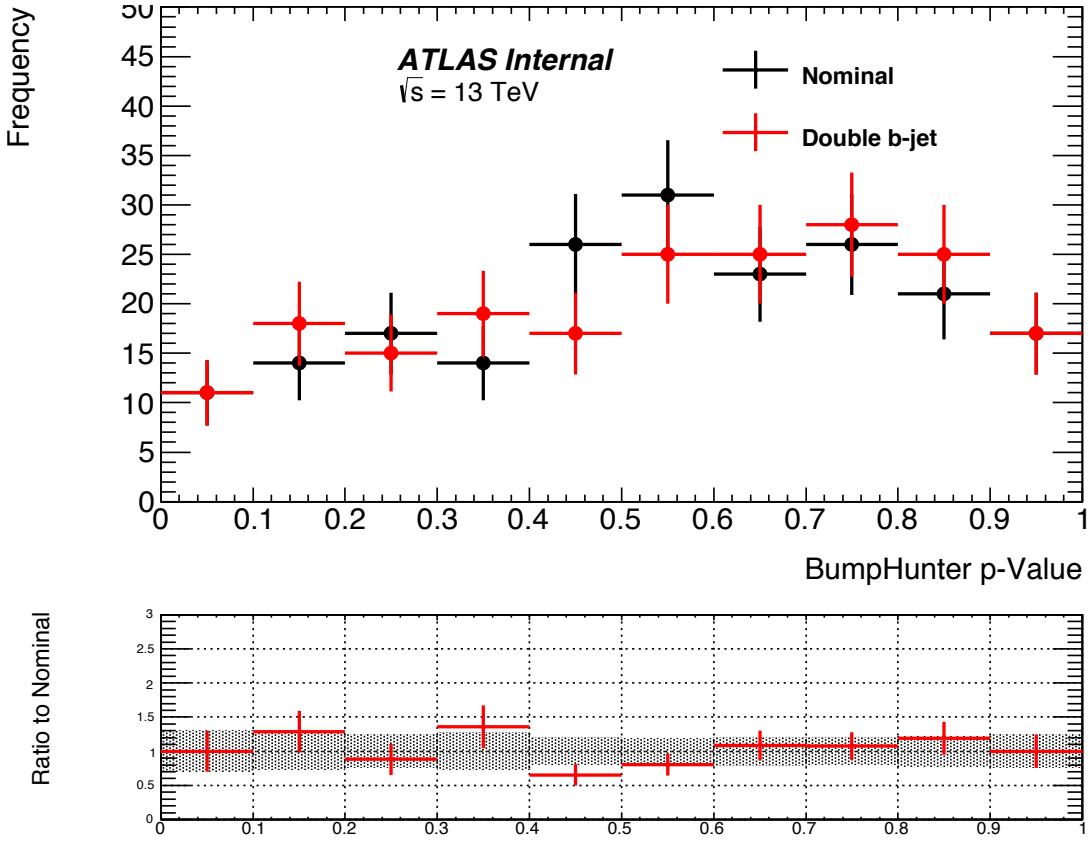




- Poisson fluctuate to create ‘data-like’ distribution
- This slide shows just one seed
  - But no large p-value here
- Repeat 200 times

p-Value Studies

- Different sets of poisson fluctuations means a different ‘data-like’ spectrum
- Each ‘data-like’ dist. checked from bumps which gives a bumpHunter p-value.
- 200 different data-like distributions have been studied



Nominal	0.542 +/- 0.018
b-jet Doubled	0.547 +/- 0.018

- No clear bias in p-value
- Mean p-values and distributions consistent between both flavour composition cases.

Conclusions

- **Use ‘data-like’ distributions**
  - Poisson fluctuate scaled distribution to create ‘data-like’ distribution
  - Then run BumpHunter on the ‘data-like’ distribution
  - Calculate a p-value of most discrepant region.

- **Look for trends in p-values**
  - Repeat process above 200 times
  - Study distribution of p-values in Nominal case and double b-jet case
  - Look for any trends in BumpHunter p-value
    - Are deviations between fit and truth causing a bias at low p-values?

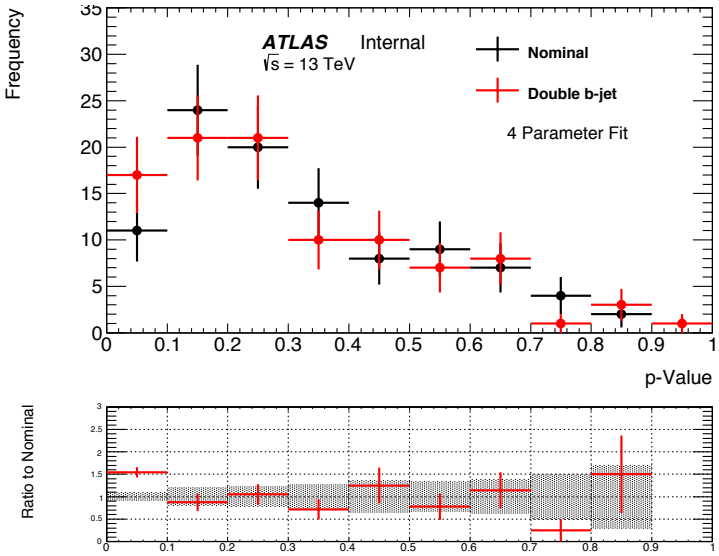
- **Results**
  - No clear bias in p-value distributions
  - No large number of ‘false alarms’
  - Mean p-values and distributions consistent between both flavour composition cases.

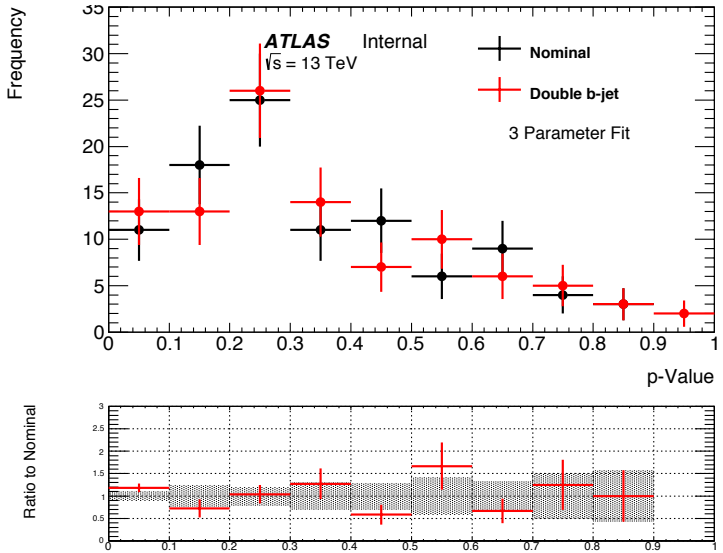
- **Documented in SVN**

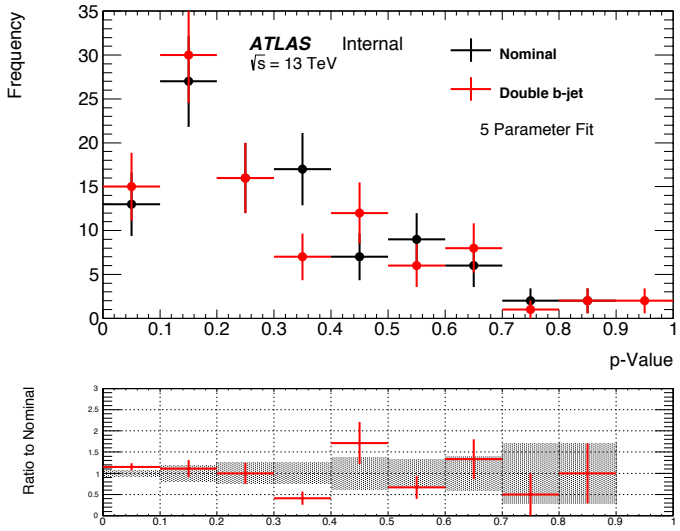
Backup:

Previous p-Value studies

- 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





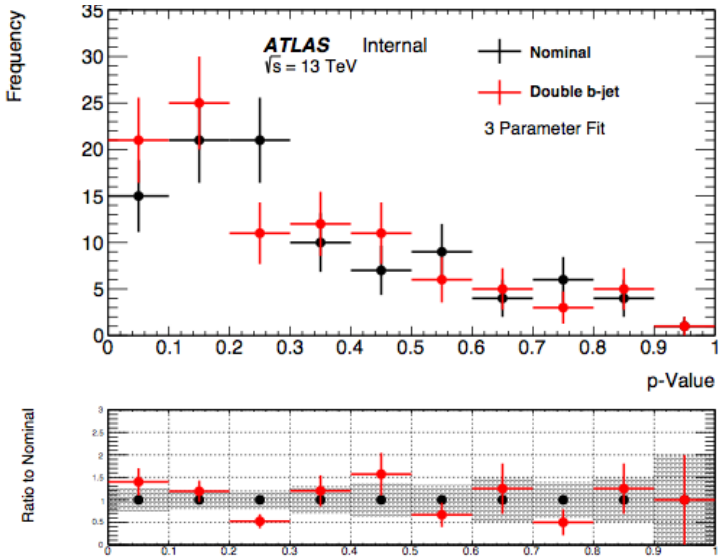


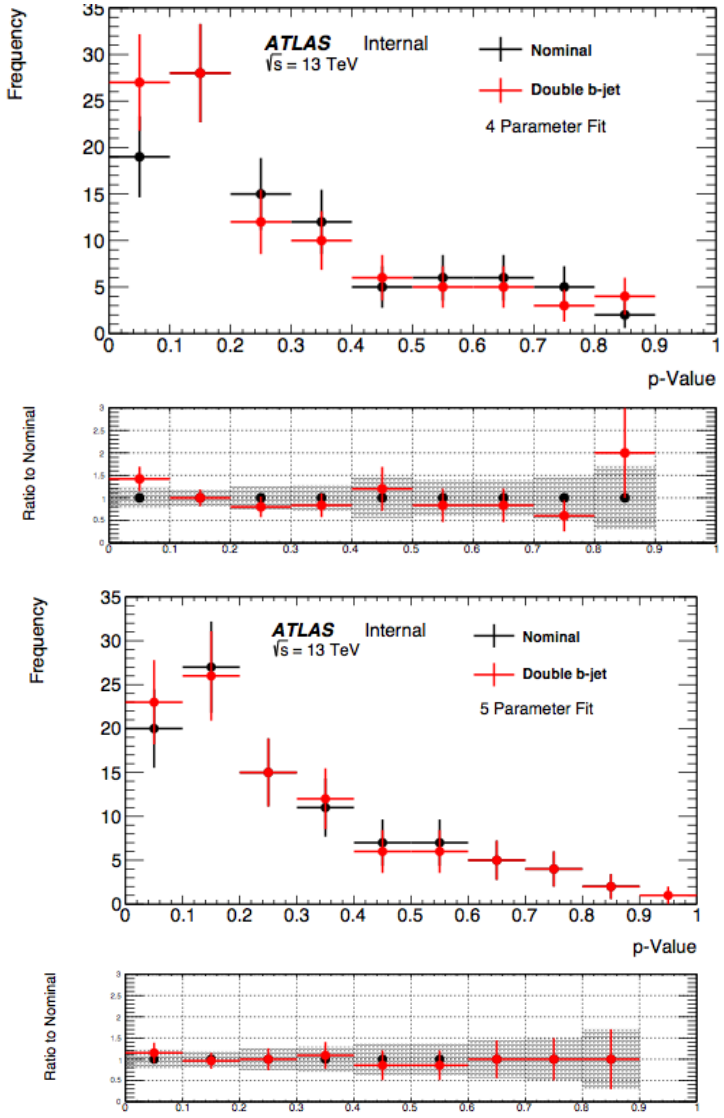
Mean p-values

	3-Para. Fit	4-Para. Fit	5-Para. Fit
Nominal	0.336 +/- 0.021	0.311 +/- 0.022	0.296 +/- 0.021
b-jet Doubled	0.347 +/- 0.023	0.307 +/- 0.022	0.297 +/- 0.022

p-Values in the Old Case: 4 Para Init.

- 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