UCL

Flavour Composition Studies

Laurie McClymont, Andreas Korn

13 January 2016

Introduction

- 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 on CDS
  => A lot of p-values are 1, is this plausible
  => This is what I want to address

Figure 17: Same comment about the fit legend text—why do they all have a p-value of 1? Also, does doubling the b-content just happen to improve the fits such that each of the 3,4,5 parameter fits are identical or are the higher order contributions just close to 0 for these fits?

Details

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

Standard Dijet Resonance Cuts
- Leading Jet p_T > 410 GeV
- Sublead Jet p_T > 50 GeV
- $|y^*| < 0.6$
- m_{jj} > 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`

Use DijetHelpersPackage:

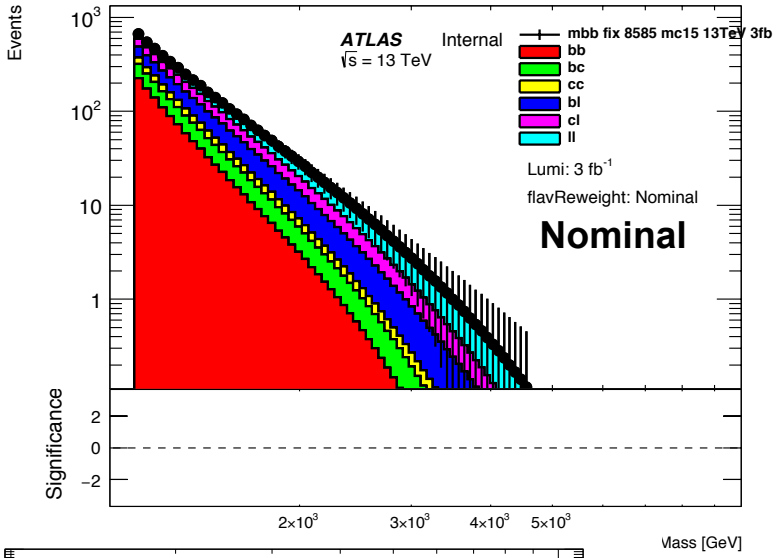
Create Histograms and merge slices
- *makeStandardHistograms.py*
- *plotStandardPlots.py*

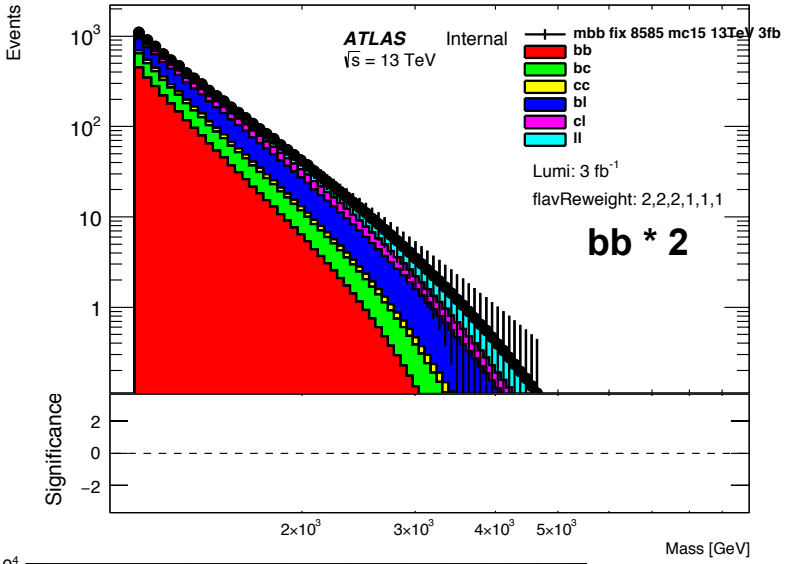
- Fit 20ifb smooth histos using *singleFit.py*
- Create data-like using from these fits

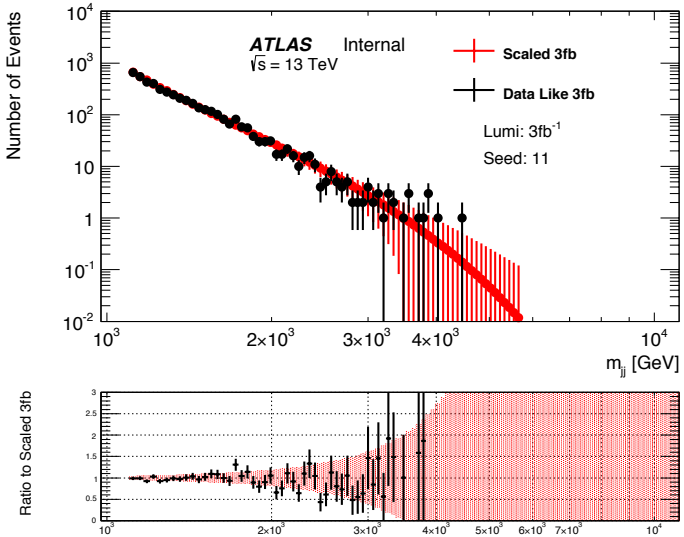
Fit spectrums and make some plots
- *singleFit.py*
- *plotSingleFit.py*

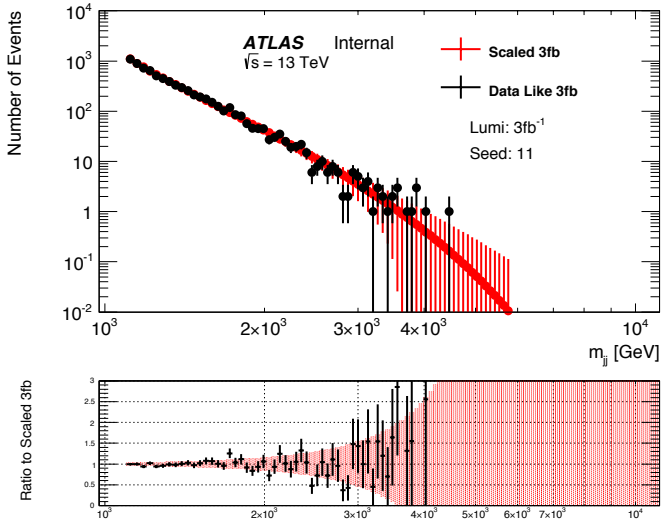
Combining Flavour Composition

- Create flavour fractions
  - => Creating histograms from fits to 20ifb scaled to 3ifb.
  - => Adding the fractions in different ways to produce various spectra.
- This creates new scaled distributions from which we can create data-like MC.



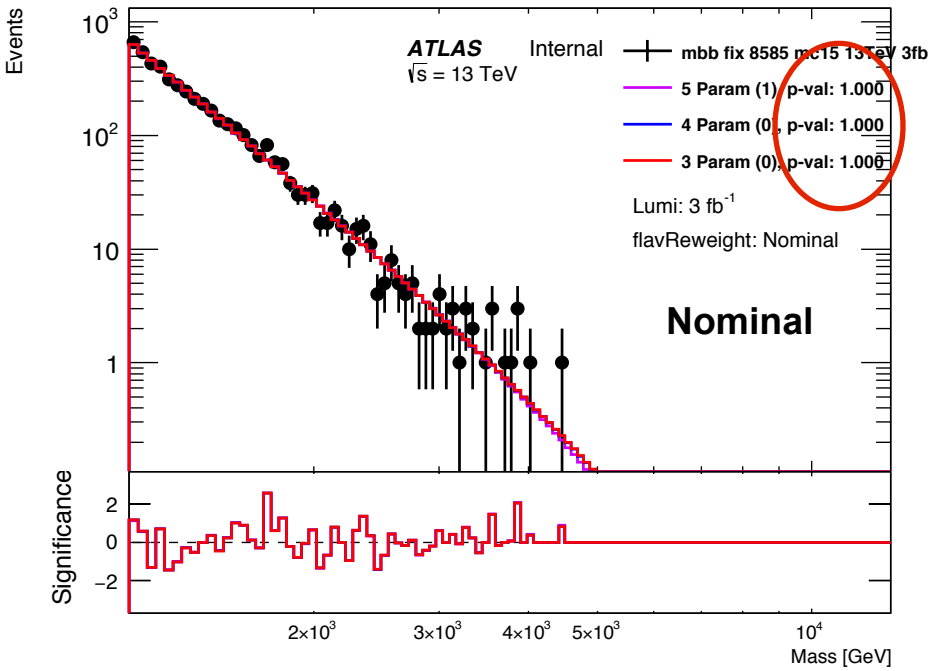


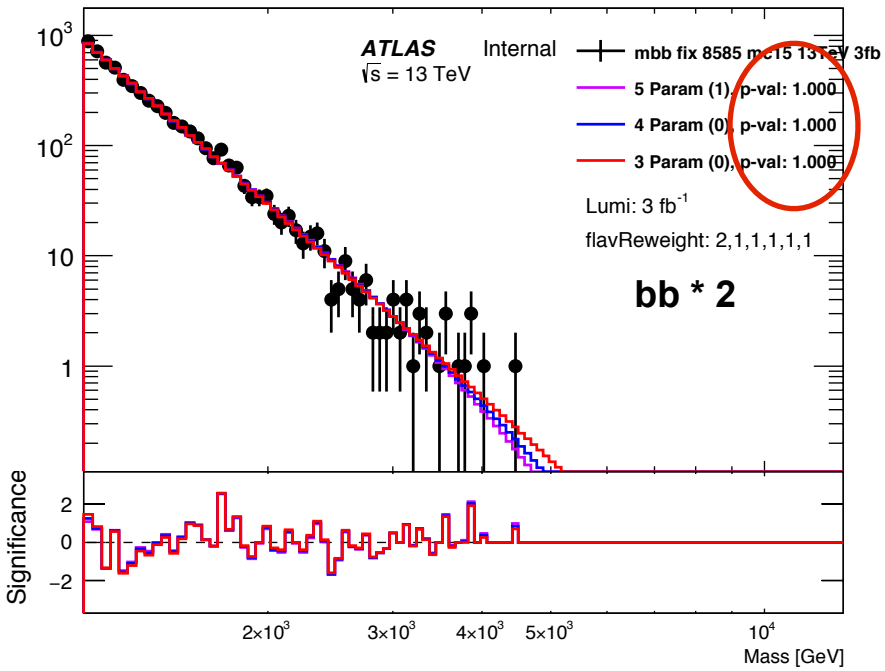




Fitting Procedure

- We then fit to data-like distributions

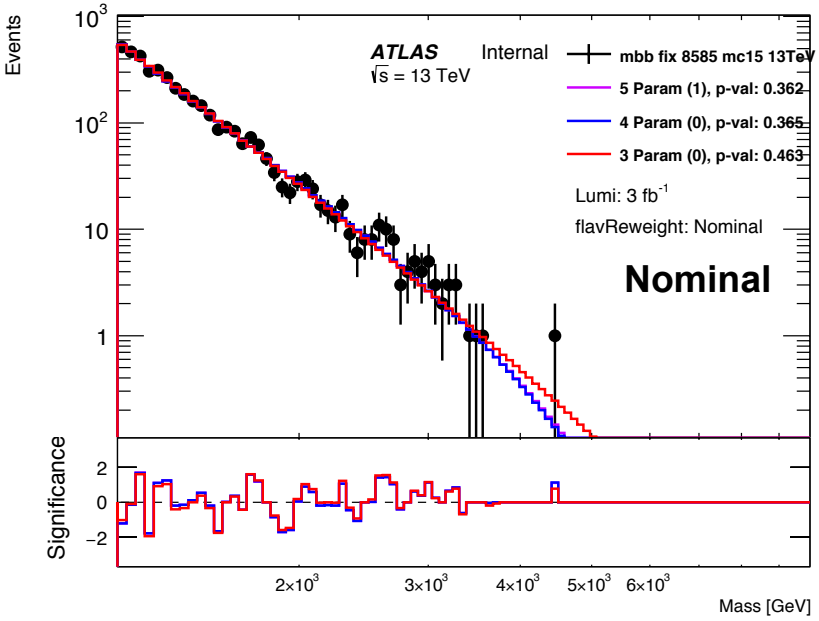


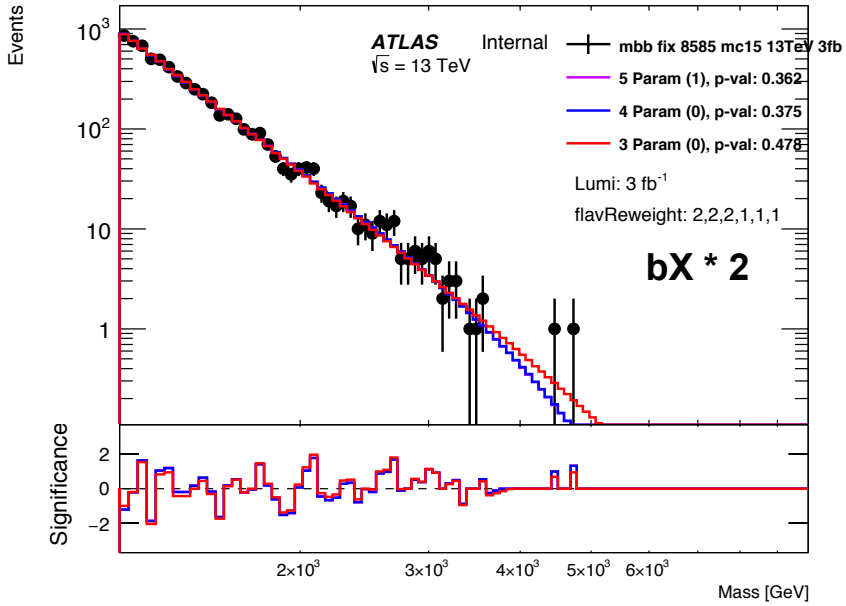


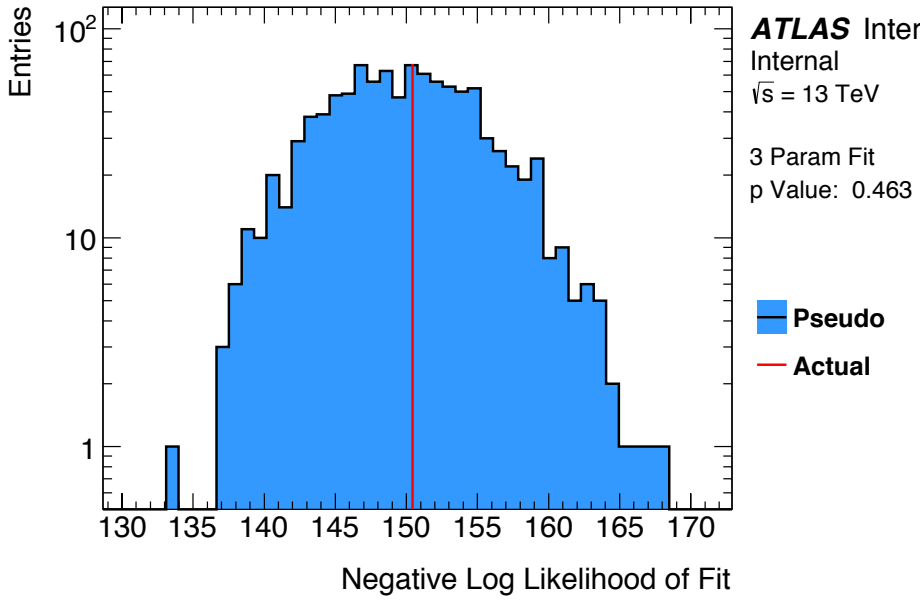
- Very high p-value (p ~ 1.000, really??)
  => These values taken from Chi^2/Ndof
  => Question on this on cds

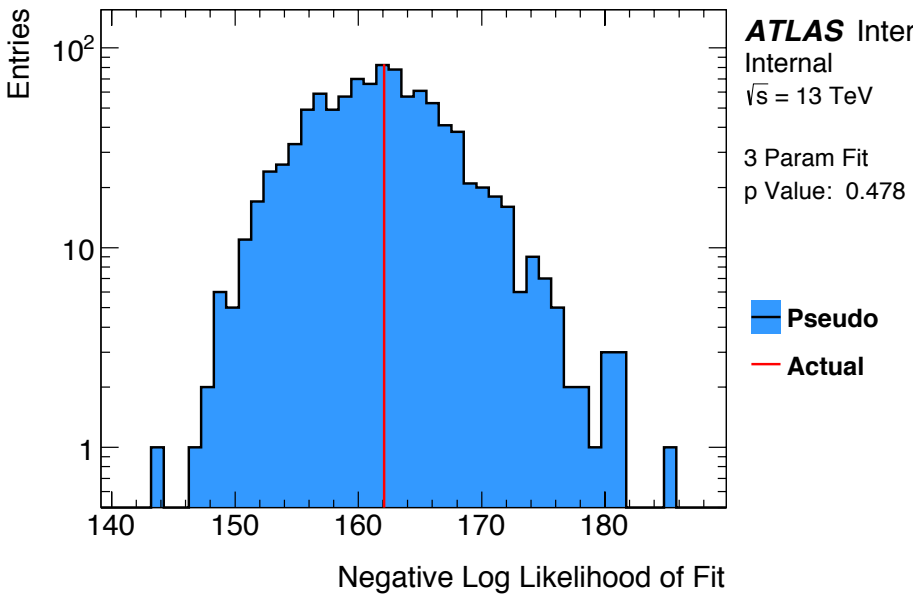
- Try using toys to create p-Values (Advised by Jeff)
  => Produce pseudo-data distributions based on the fit function parameters.
  => Fit to these distributions.
  => Ask number of times that fit to the pseudo is better than the original fit
  => I use 1000 toys here.

Fitting Procedure with toys for p-value





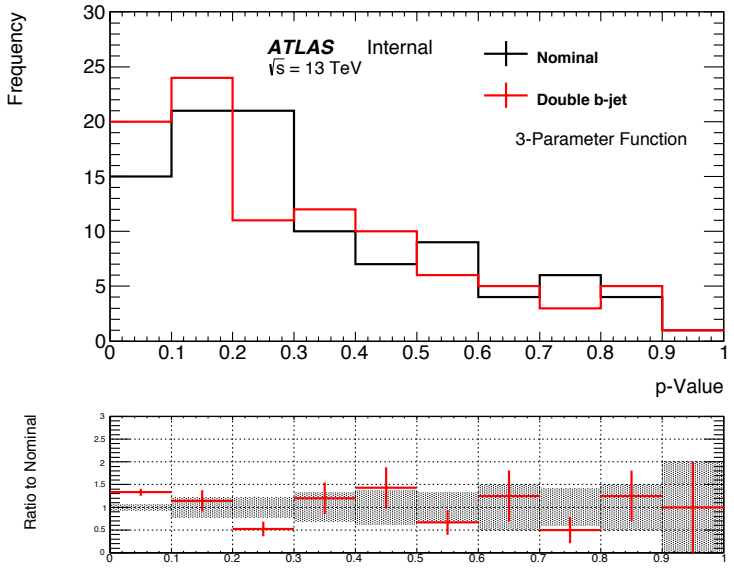


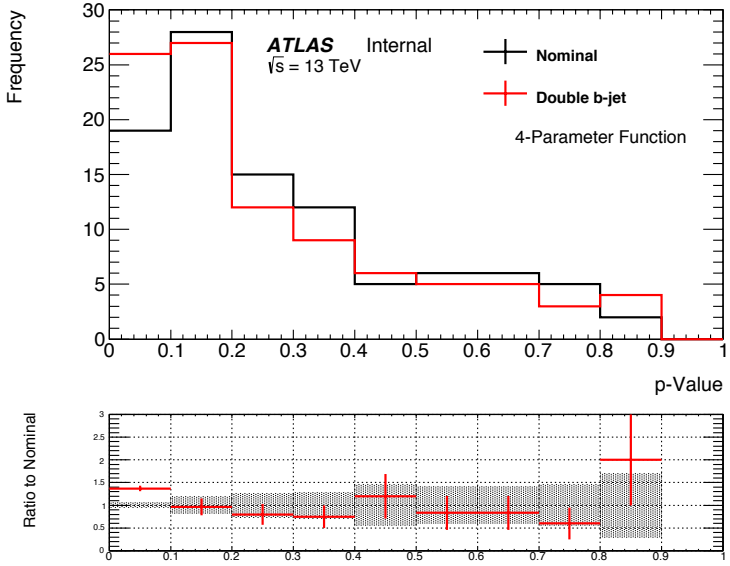


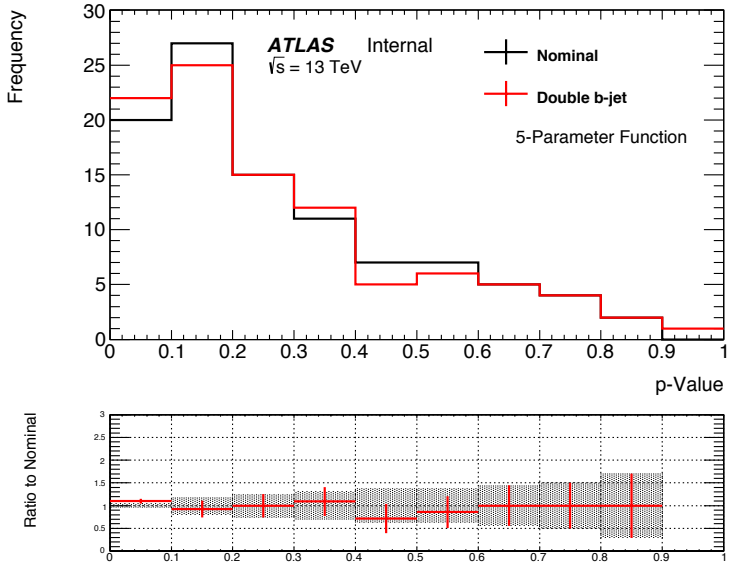
p-values for different data-like seeds

- Each seed represents a different initial fluctuation, a different data-like distribution
  => For each seed a p-value can be created
  => 100 Seeds tried.

	# Parameter		
Mean	3	4	5
Nominal	0.325	0.280	0.283
bX*2	0.311	0.270	0.280







Conclusions

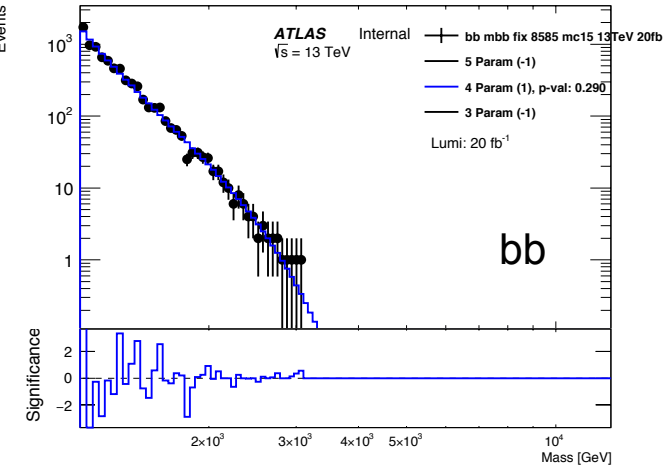
- Creating p-values by using pseudo-experiments.

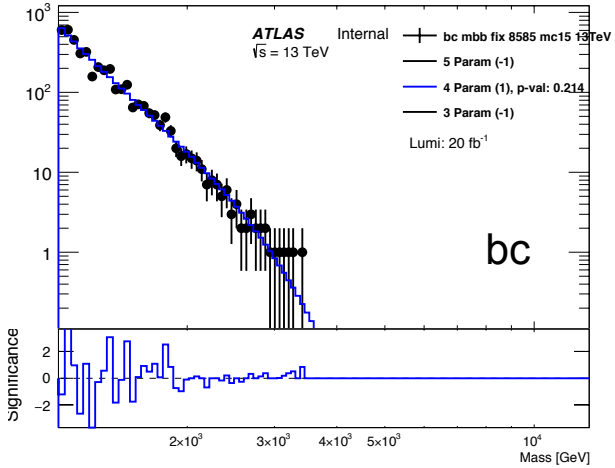
- Low p-values when using toys.
=> Really want to try and understand why...

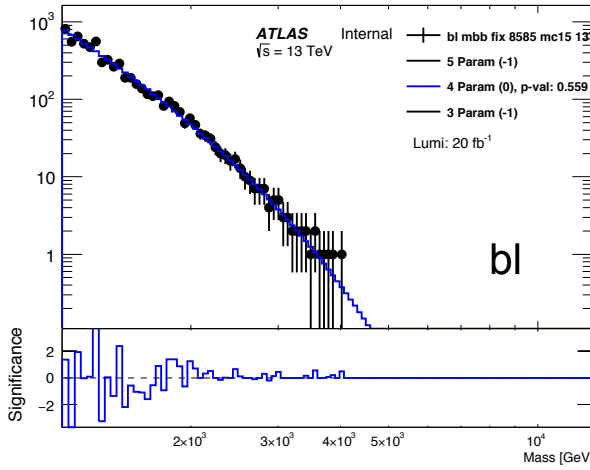
- What is the best way of displaying results
=> Choose one typical example
=> Show a grid of results
=> An average of several p-values

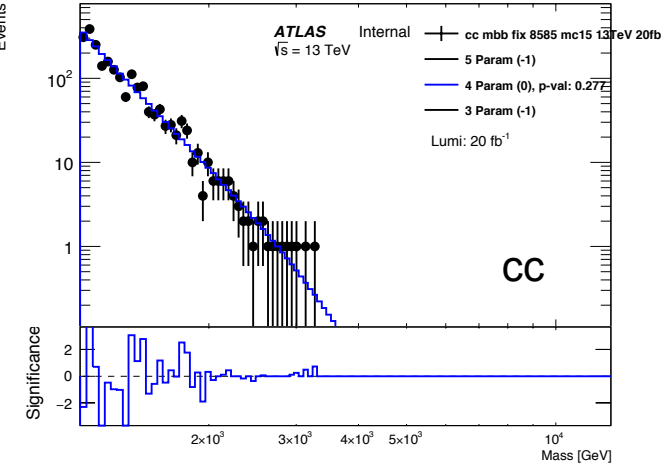
Fit to Flavour Fractions

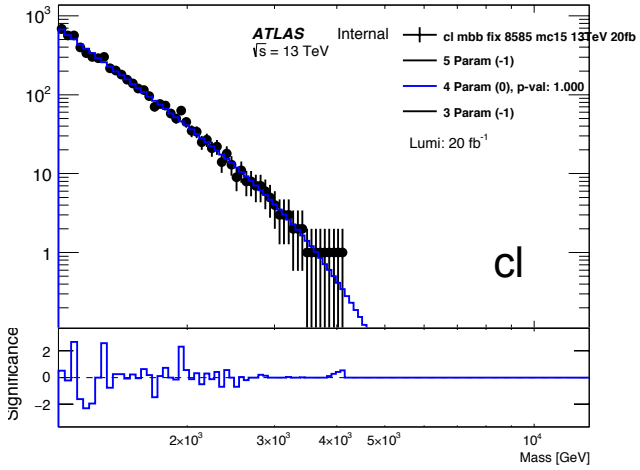
- Extract flavour fractions from 20 ifb of MC using truth information.
- Fit to smooth distributions (rounded to integer) using 4 parameter fitting function.
- Use these fits as templates to model the shape of the flavour fractions.

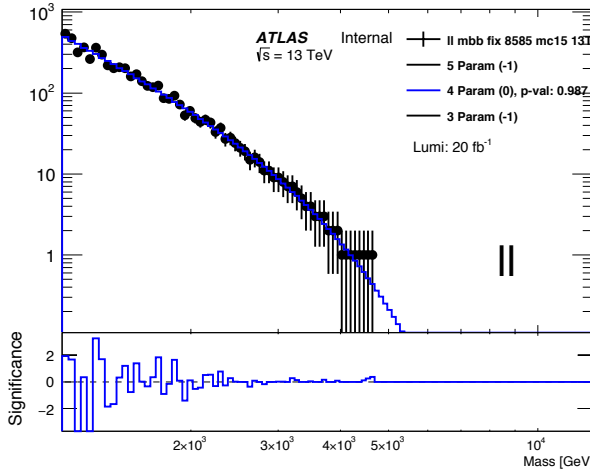








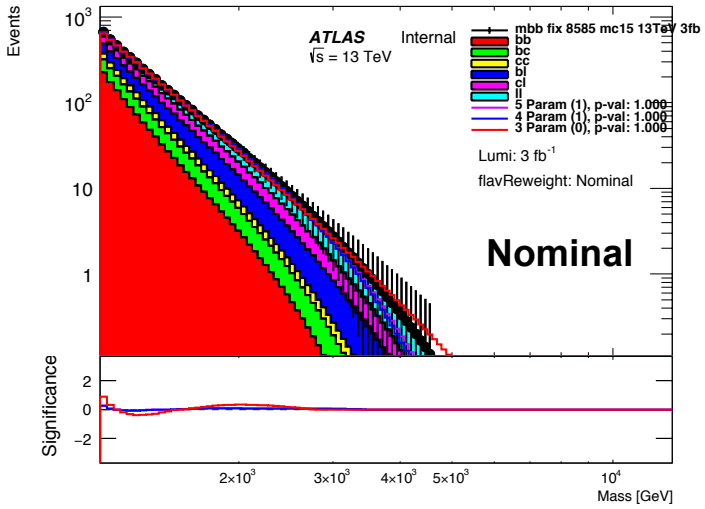


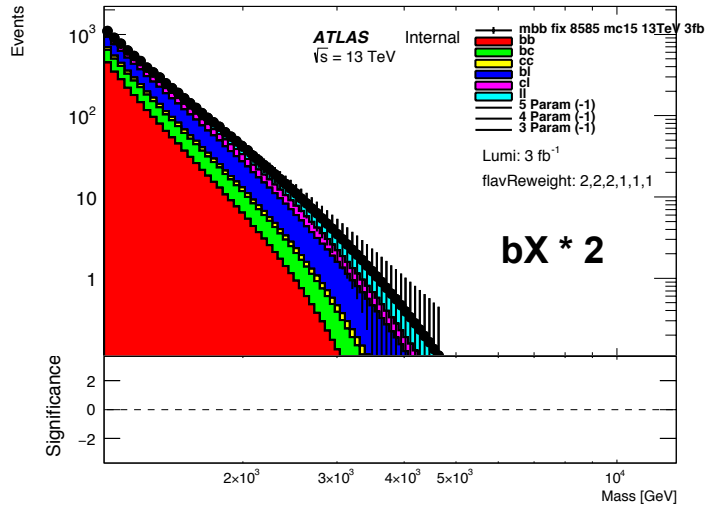


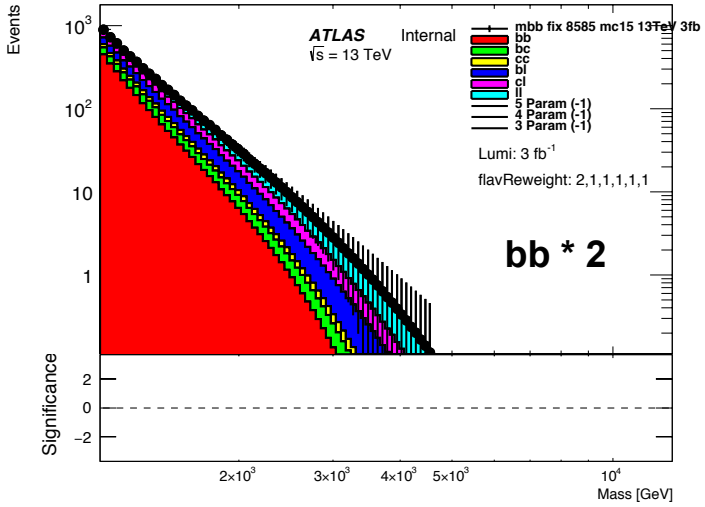
Need to look at errors, here they are poisson. => Maybe I should fit scaled that is truncated at 0.5

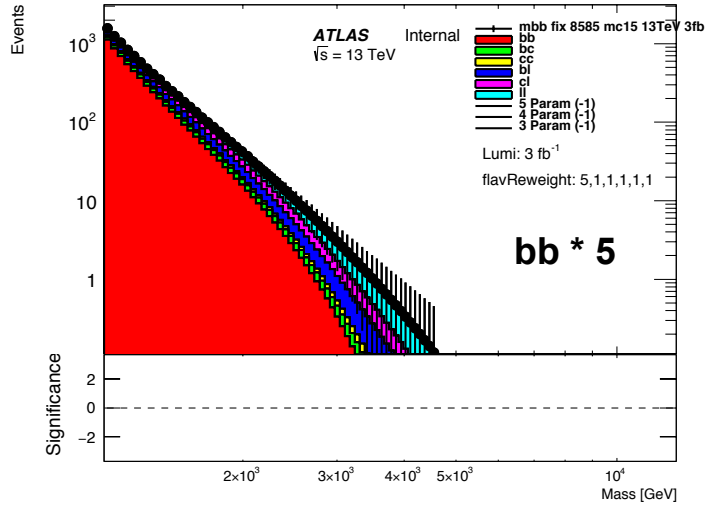
Combining Flavour Composition

- Then combine histograms by,
  => Creating histograms from fits scaled to 3ifb.
  => Adding the fractions in different ways to produce various spectra.
- This creates new scaled distributions from which we can create data-like MC.



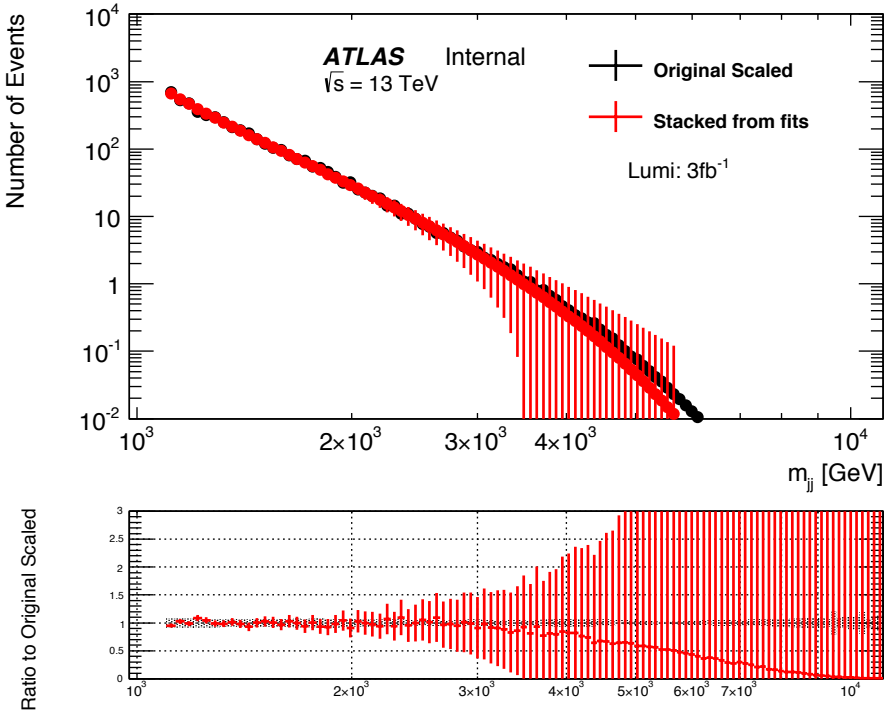






Comparing Nominal to Original Spectrum

- Compare nominal produced from stacking to original MC sample
  => To validate that our procedure for producing distributions is valid.



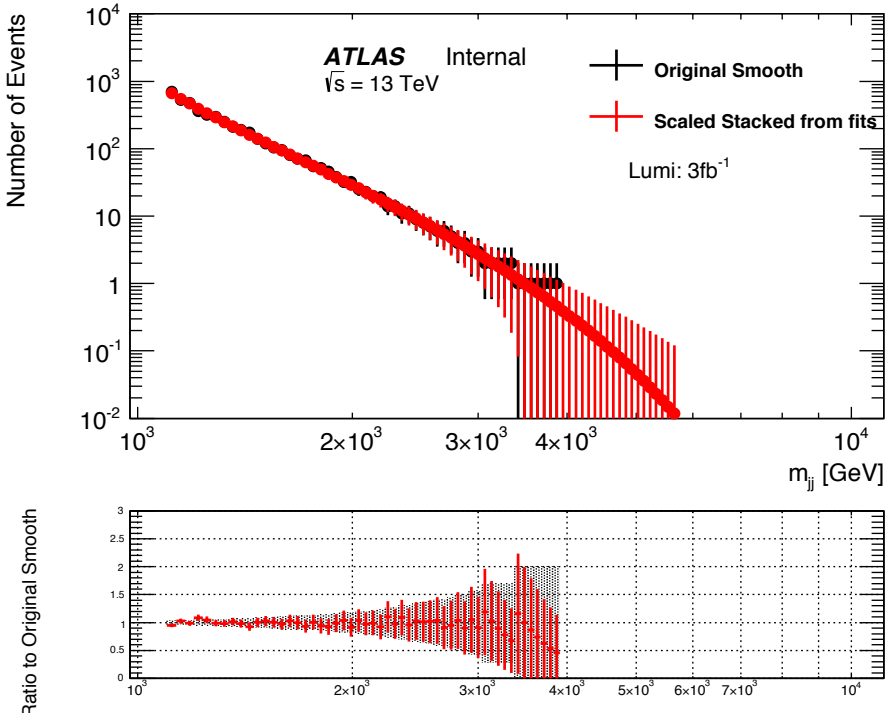
Smooth means
- rounded to one
- poisson errors

Smooth distributions was what was originally fitted to
=> Comparison is good.

Matches well in region where nEvents > 1

Some differences in tail of distribution.

=> All within errors.



Comparing the shapes

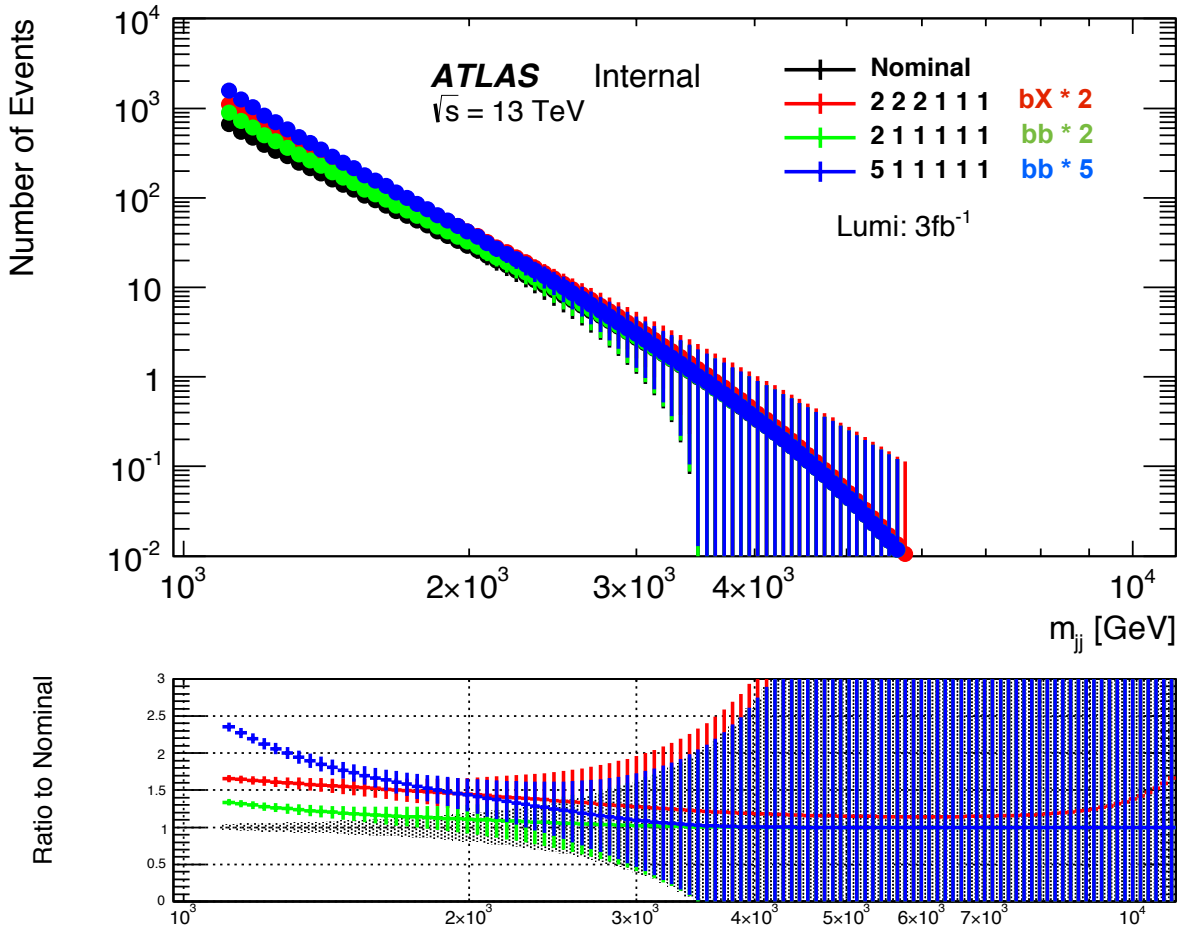
- Compare how different these spectra actually are...

=> 1 1 1 1 1 1 = Nominal, no reweighing.

=> 2 2 2 1 1 1 = double contribution of bb, bc, bl

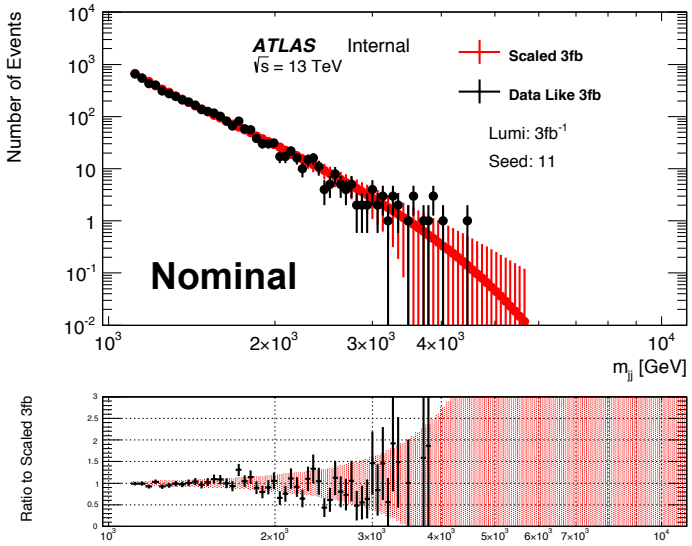
=> 2 1 1 1 1 1 = double contribution of bb

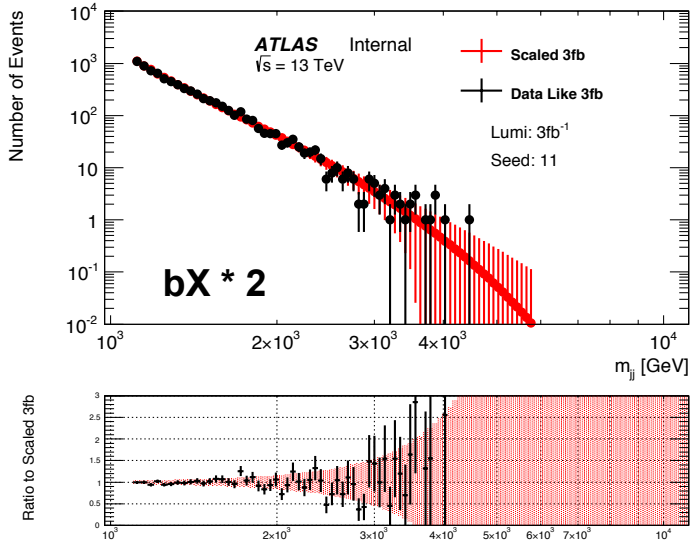
=> 5 1 1 1 1 1 = quintuple contribution of bb

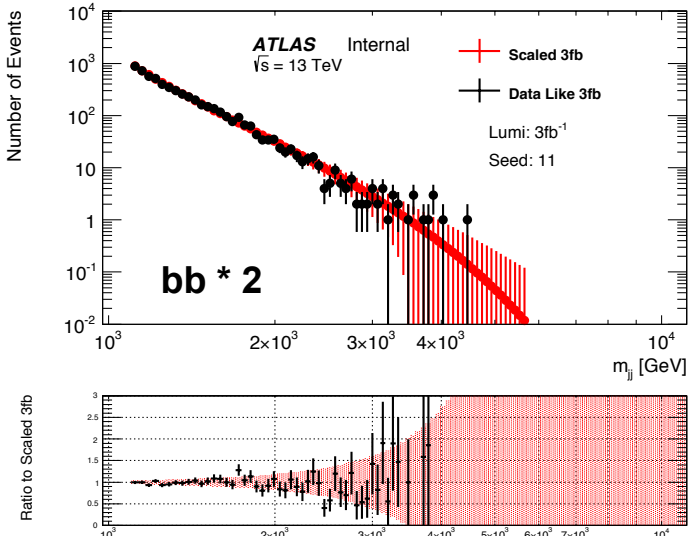


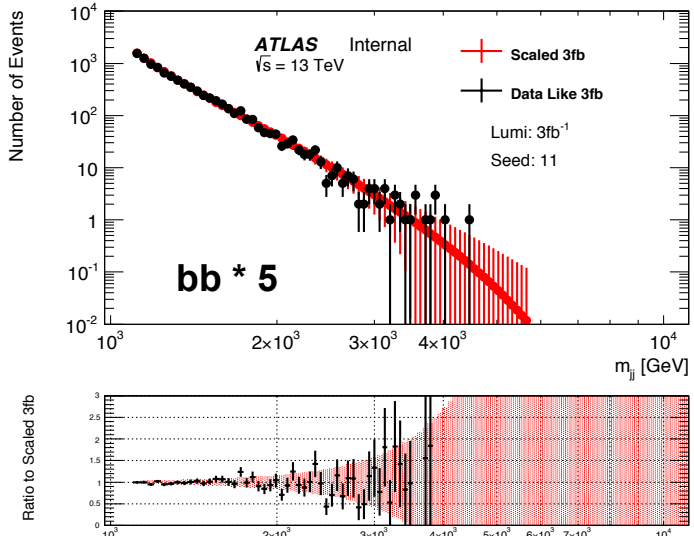
Making Data-Like Monte Carlo

- Make data-like
  => Poisson fluctuating using scaled distribution as mean entry.
- This creates new scaled distributions from which we can create data-like MC.



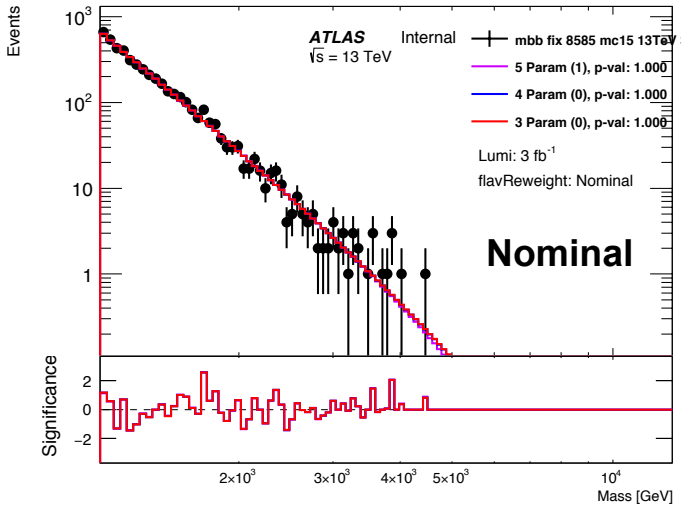


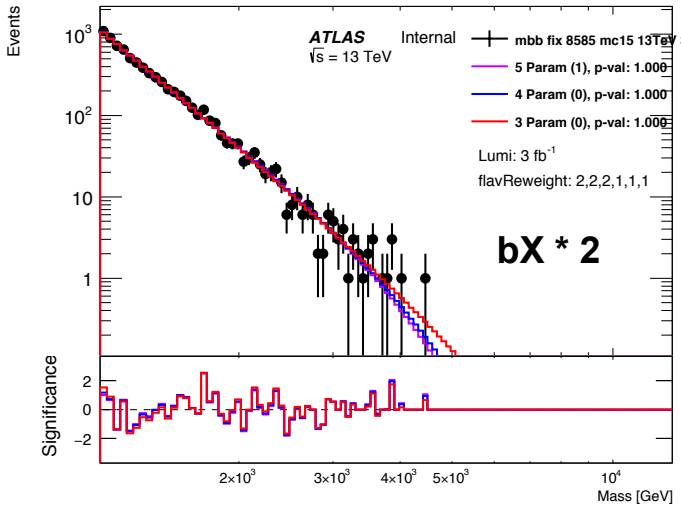


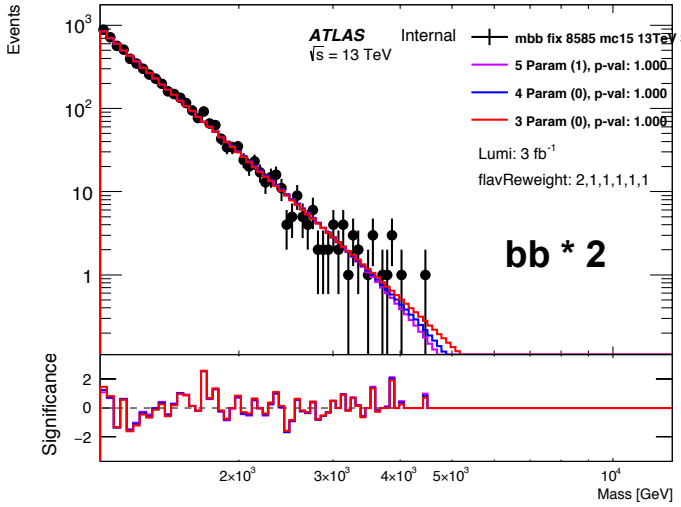


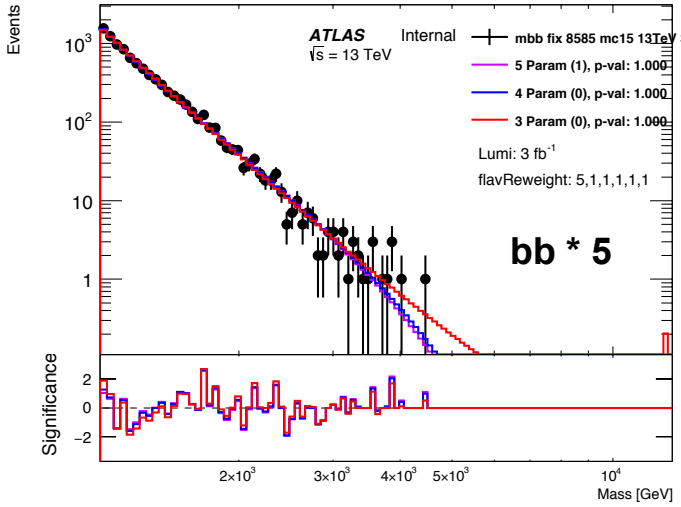
Fitting Procedure

- We then fit to each of our different combinations using the all our fitting functions
- Strong performance in all cases
  => However 3 parameter fit splits from others as weighting increases.

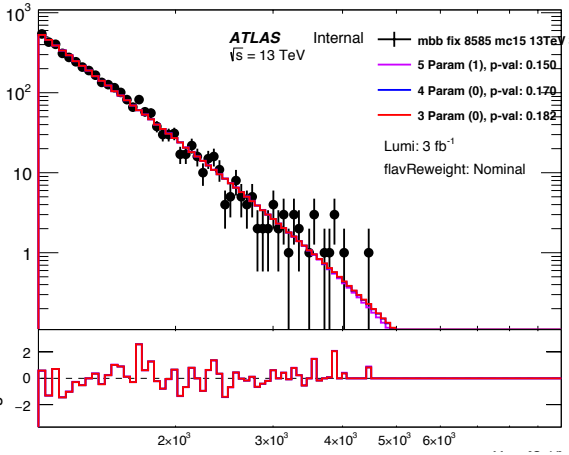


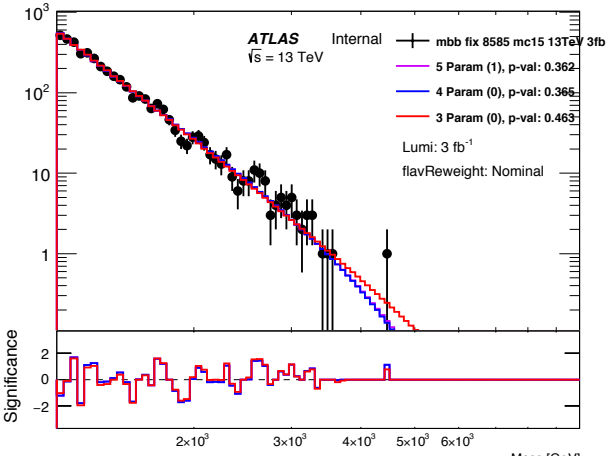


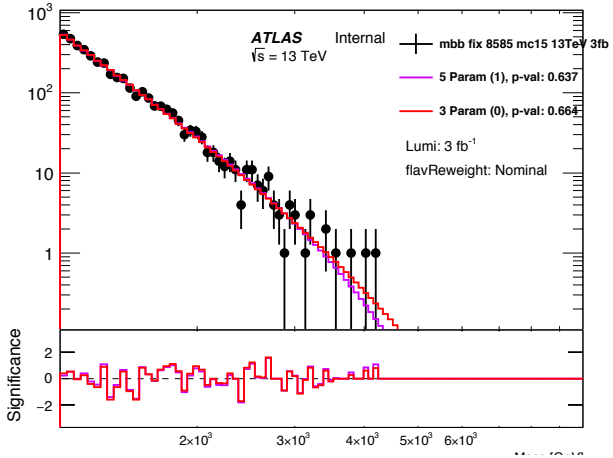


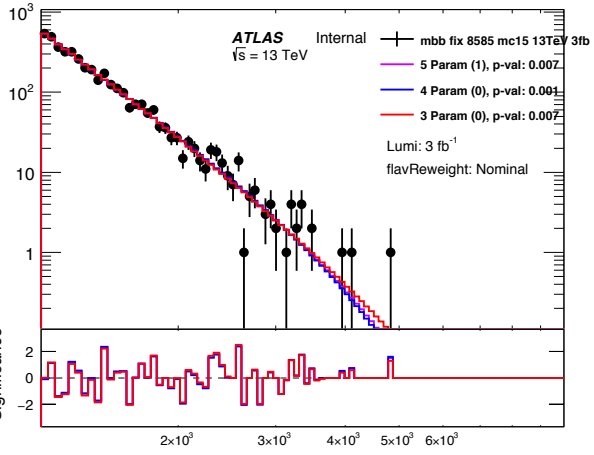


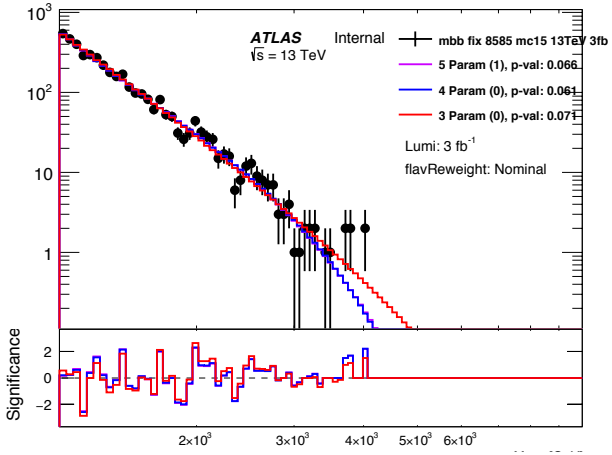
Nominal, All Seeds

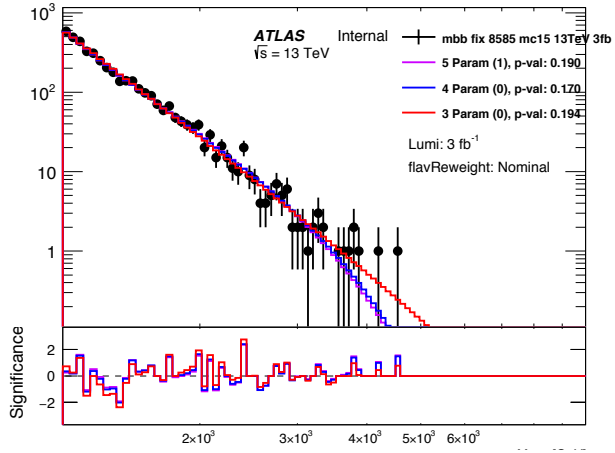


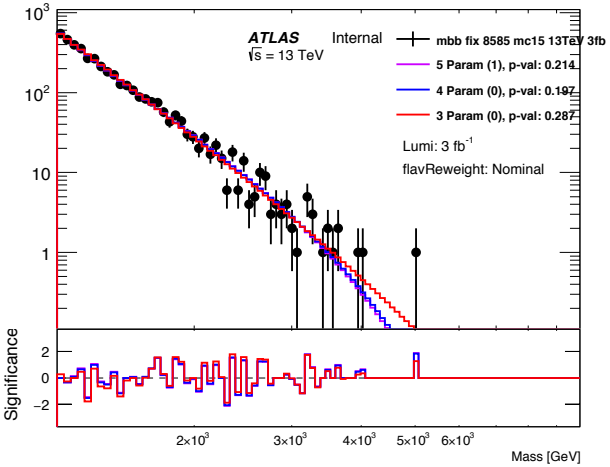


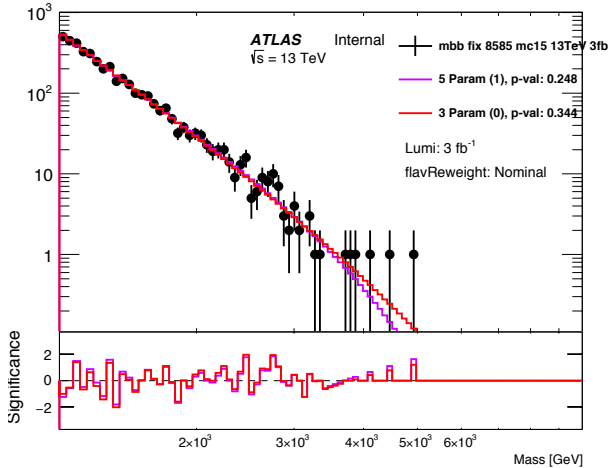


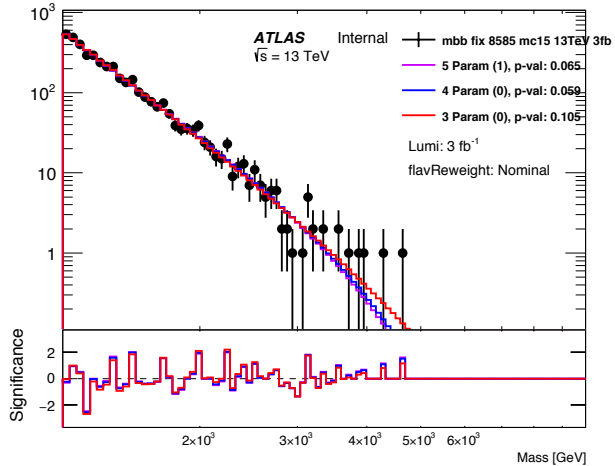
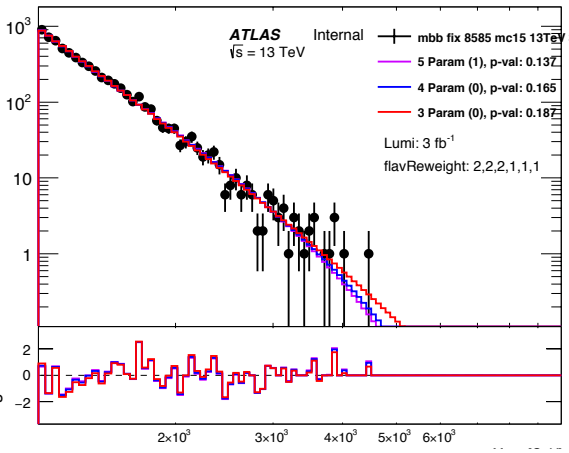
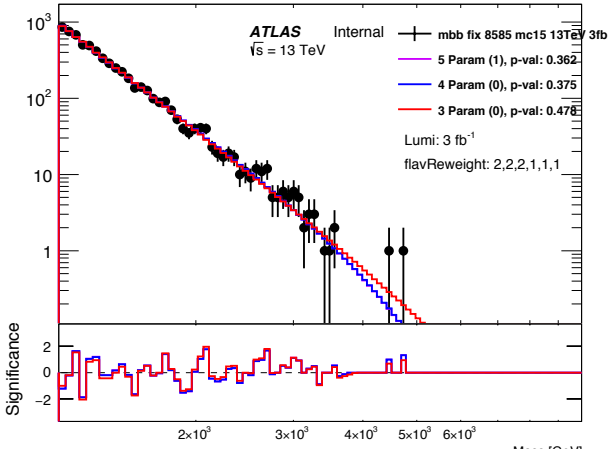
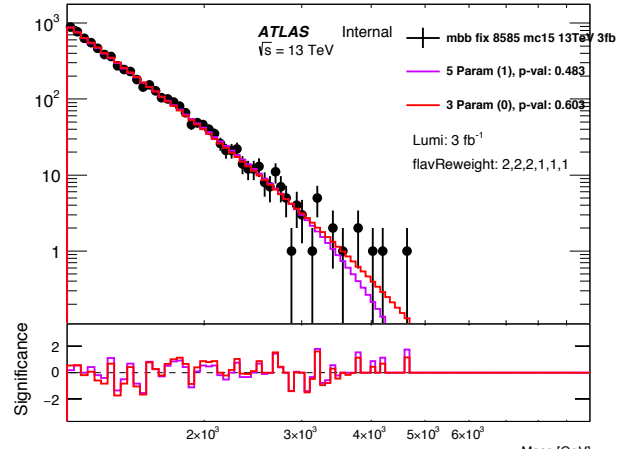
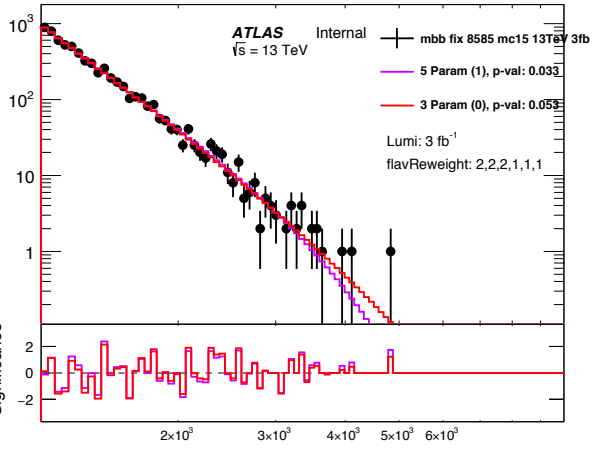
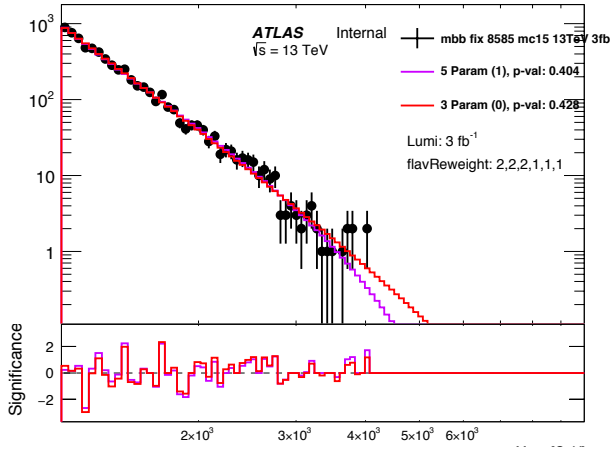
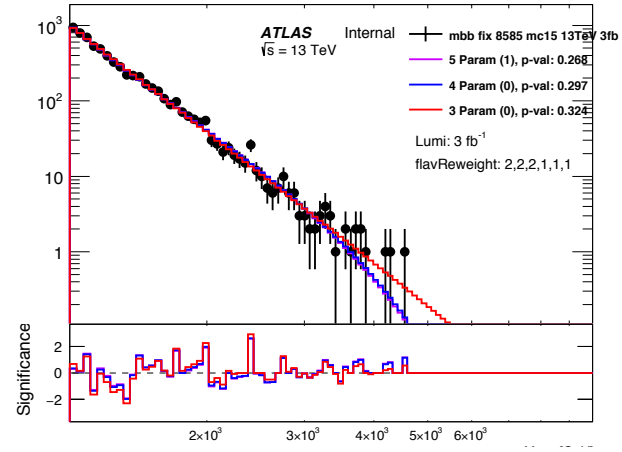
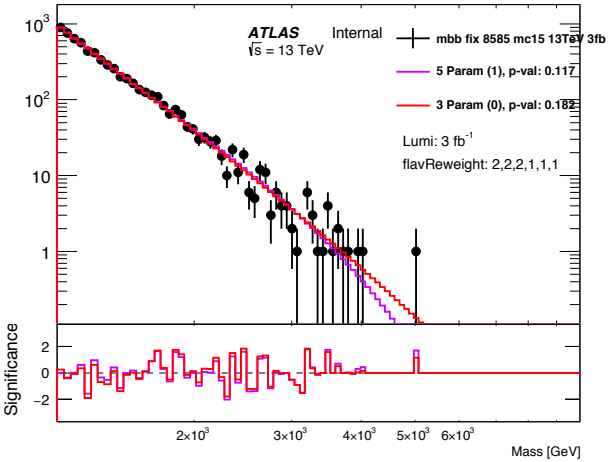
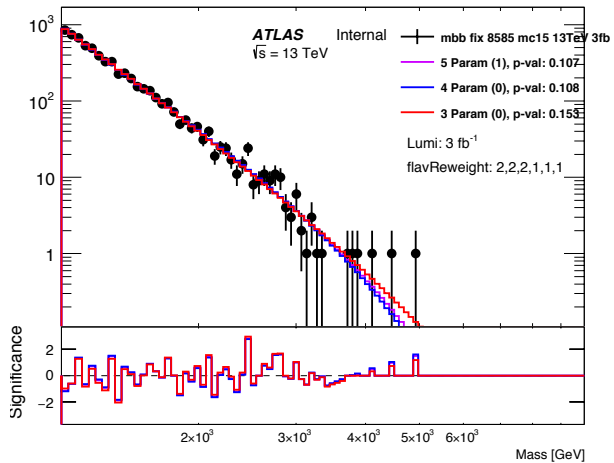
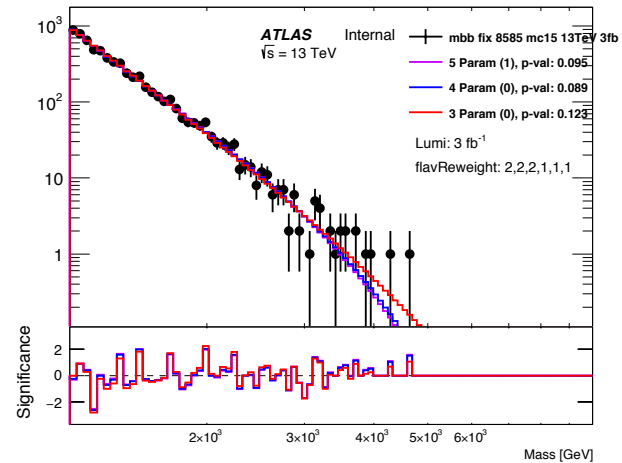












bX*2 All Seeds