bJES calibration for precision top physics

(cut out of this discussion)

Miguel Jiménez & Juan Ramírez

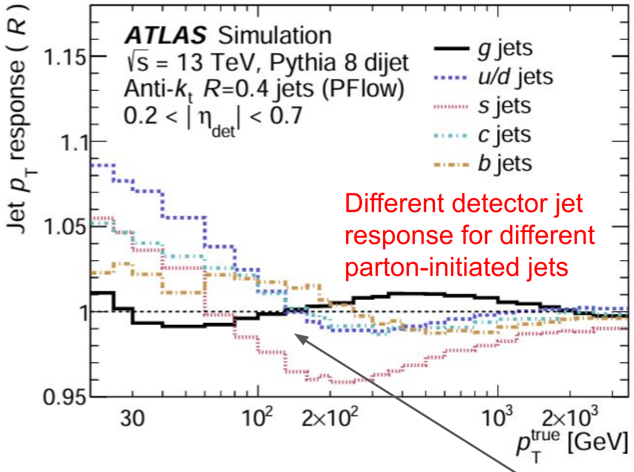
gitlab repo

10-12 November

Top-quark mass workshop, Valencia

Supervisors: Marcel Vos and Davide Melini

Just a word on jet flavour



Jet Flavour Matters!

GN2 Working Point	c-quark	light-quark	b-quark	gluon
Inclusive	11.7%	68.5%	1.9%	17.8 %
90%	44.6%	39.9%	2.8%	12.7%
85%+77%	44.3%	32.4%	10.1%	13.08%
70%+65%	8.5%	10.3%	63.8%	17.4%

Table 25: Flavour composition 150-1000 GeV. Generator: Pythia8

Classic calibration based on an inclusive sample of light-quarks initiated jets

General summary of the analysis: (Nominal method → Method 1)

It will be documented in this int note (with other techniques): ANA-JETM-2025-01

Goal: Measure bJES and cJES in situ responses with a precision below per mil level

1. Create in situ response histograms in a pT photon basis. Do this for both data and MC. The main observable is the direct balance ratio → pT leading jet / (pT photon * cos(DeltaPhi(ph,jet)))
2. We extract gaussian / poisson fits per bin to get the mean of the response distribution.
3. Perform the double ratio R_DB_MC / R_DB_data. The nominal sample is usually PowHeg+Pythia8, but this can be changed if we see there is a better non-closure for other generator. With this the nominal JES factor is gotten.
4. We do 1-3 for an inclusive sample (light quarks) and for three different tagged samples (65%-70% WP: mainly b-quarks), (77%-85% WP: mainly c-quark) and (90% WP: mainly c and light quarks).
5. We extract the nominal bJES / JES factors making the ratio of the tagged over the inclusive sample. Then we do a constant fit to the histogram.
6. The uncertainties are bootstrapped using the bootstrap package. The modelling uncertainties are just the rough comparison of the bJES / JES factors that one generator gives respect to the nominal Py8. As alternative we are using Herwig7. The statistical uncertainty is assumed to not be highly correlated so it is directly propagated (b-jet sample much higher than inclusive).
7. About alternative method, more in the backup.

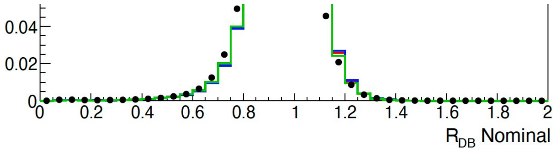
2

1. Same logic as Nominal method: create in situ response histos of R_DB for MC and data, fit them, perform the ratio R_DB_MC/R_DB_data to get the (b)JES; then extract bJES/JES by doing the double ratio of b-tagged over inclusive; BUT
2. The alternative method combines all the pT bins between 150 GeV and 1 TeV to get a single R_DB distribution to fit.
3. Statistical uncertainties are obtained propagating the uncertainties on the gaussian fit mean for data and MC.
4. Modelling uncertainties are obtained as in the Nominal method.
5. Systematic uncertainties are obtained comparing the final bJES/JES between the nominal and the variated samples (with some subtleties, check backup) → NO BOOTSTRAPPING.

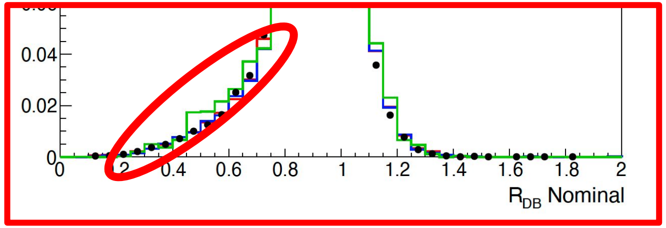
Nominal calibration

Fit Range Studies

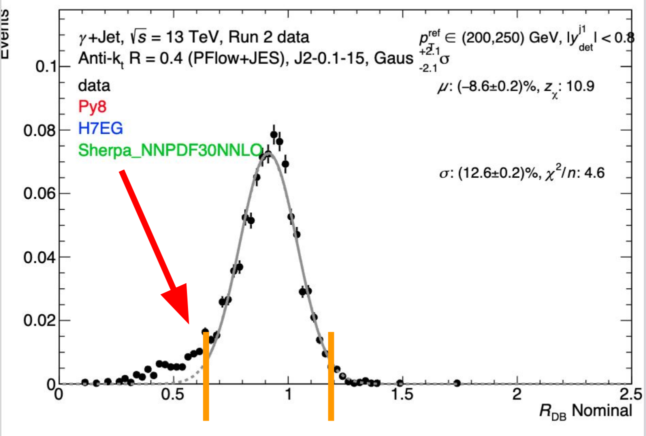
Inclusive (Sl.6)

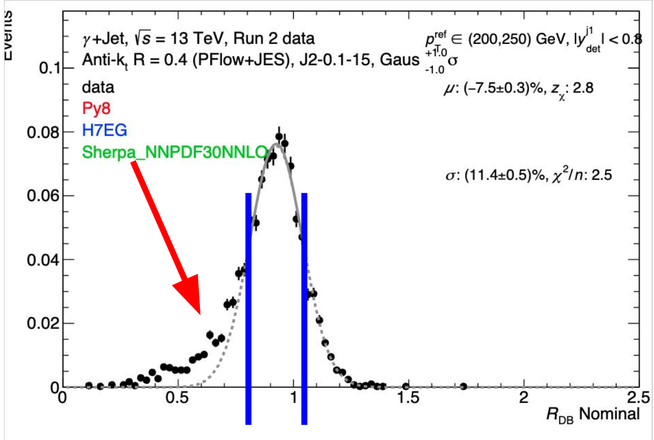


WP70+65: bottom (Sl. 6)



- Nominal strategy for fitting R_DB distributions → gaussian fit in in symmetric range μ +/- 2.1σ, defined by the full distribution's mean and RMS, μ and σ.
- This works for the inclusive sample (symmetric as well), but for b-tagged samples (especially tight b-tag) there is a neutrino tail at low R_DB that **biases the fits towards lower R_DB values** when included in the range.





NEW New studies on this → crystal ball fits, fake subtraction fits, other ideas that can come up

Method (1), data, pT bin (200,250) GeV. Nominal +/-2.1σ fit vs +/-1.0σ fit.

Fit Range Studies. Nominal MC

WP	(c)bJES/JES
70+65	0.9978
85+77	0.9967
90	0.9991

Table 16: Pythia8, nominal $\pm 2.1\,\sigma$ fit

WP	(c)bJES/JES
70+65	1.0028
85+77	0.9970
90	0.9998

Table 17: Pythia8, alternative $\pm 1.2\,\sigma$ fit

WP	(c)bJES/JES
70+65	1.0048
85+77	0.9986
90	0.9989

Table 18: Pythia8, alternative $\pm 1.0\,\sigma$ fit

Decreasing the range of the gaussian fit rises bJES/JES, particularly at WP70+65, and gives better chi2/red (backup)

WP	(c)bJES/JES
70+65	1.0007
85+77	0.9973
90	1.0003

Table 19: Pythia8, alternative FF method

FF has a higher bJES/JES than the nominal as well, but is still smaller than the fits with shorter ranges for WP70+65 (comparison of whole distributions → neutrino tail effect?)

Systematic effects

Summary of Central values & Uncertainties

Method 1: Nominal → fit to the bJES / JES value. Bootstrapped uncertainties

Quantity	bJES/JES (65% ‖ 70% WP)	cJES/JES (77% ‖ 85% WP)	cJES/JES (90% WP)
central value Pythia8	1.0015	0.9984	1.0004
central value Herwig7	1.0008	0.9966	0.9995
central value SHERPA2.X	0.9853	0.9970	0.9974
stat. uncertainty	0.0018	0.0012	0.0009
MC stat. unc.	0.0004	0.0002	0.0002
photon scale unc.	0.0015	0.0027	0.0002
photon res. unc.	0.0015	0.0023	0.0001
$\Delta\phi$ unc.	0.0008	0.0003	0.0003
veto cut unc.	0.006	0.0009	0.0002
Fit range unc (2.1σ - 1.5σ)			
Fit range unc Py8	0.005	0.002	0.0007
Fit range unc Herwig	0.003	0.00007	0.0006
Fit range unc Sherpa	0.009	0.0034	0.001
total uncertainty	0.008	0.0027	0.0014

Method 2: Direct extraction from response distributions

Quantity	bJES/JES (65% ‖ 70% WP)	cJES/JES (77% ‖ 85% WP)	cJES/JES (90% WP)
central value Pythia8	0.9978 **- 0.004**	0.997 - 0.0014	0.9991 - 0.0013
central value Herwig7	0.9970 **- 0.004**	0.996 - <0.001	0.9982 - 0.0013
central value Sherpa	0.9807 **- 0.005**	0.996 - 0.001	0.9980 + <0.001
stat. uncertainty	0.0009	0.0007	0.0006
MC stat. unc. (Py8)	0.0004	0.00023	0.00019
photon scale unc.	0.0006	0.00023	0.0014
photon res. unc.	0.0004	0.0007	0.0004
$\Delta\phi$ cut unc.	0.0024	0.0007	0.0012
veto cut unc.	0.0015	0.004	0.0009
fit range unc. ($2.1\sigma - 1.5\sigma$, Py8)	0.0010	0.0023	0.0015
total uncertainty	0.0032	0.005	0.0027

Central values for **WP90 and WP85+77 agree** +/-0.001 between both methods, but **this fails for WP70+65 (x4/5 times larger)**!

Systematic uncertainties show fluctuations between the 2 methods at the per mille level
E.g. for WP70+65, leading syst for (1) is Veto cut, but for (2) is Δϕ, which is x4 larger than in (1)!

The **selection of the fit range** turns out to be one of the **main uncertainties**, if not the leading! (more on this in next slides)

Breakdown of uncertainties. What is ok and what is not

1. The uncertainty on some systematic effects is reduced respect to inclusive calibration, i.e GES goes from (approx) 0.5% to 0.1% → Expected → it is OK
   Cancellation of systematic effects as well in GER, topology cuts
2. Topology cuts → a bit messy (distribution not gaussian at low pT) → differ between two methods → bad, for the first method we rely on stat significance based on 100 bootstrap (>2sigma) → This is not OK
3. In principle the fit should not affect that much, but yet does → Improve of fitting strategy → Just work in progress
4. Differences between generators → can be fixed with dedicated MC2MC corrections (already implemented in the general code).

Bottlenecks (work in progress)

Combination of Z+jets and ɣ+jets results

- Performed following the nominal (1) method and including the Z+jets samples in the central ($|y\_j1^det|$ < 0.8) region (huge thank you to Magda for lending us your results! Check her HCW2025 contribution for more details).
- Extraction of bJES/JES is performed individually for ɣ+jets, Z(→ee)+jets and Z(→μμ)+jets. After validating with the plots that the results are compatible, we perform a fit of all the points from the 3 analyses.
- Caveat: the MC generators' version aren't exactly the same between the analyses (for more details, check backup).
  - ɣ+jets uses Pythia8 samples and Sherpa 2.2.2
  - Z+jets uses Powheg+Pythia8 and Sherpa 2.2.11
  - Herwig and FxFx are exclusive of ɣ+jets and Z+jets, respectively.

Conclusions

- Preliminary results are shown for Nominal (1) and alternative (2) methods
- First look at combination of Gamma+jets and Z+jets: complementary coverage (gamma+jets: 150-1000 GeV, Z+jets: 20-300 GeV); good agreement between 3 sets of results
- The method still requires a fine-tuning to improve both fit quality and significance to see a real effect of the systematic uncertainties.

Discussion & Outlook

- Once the b-jet regression is available, compare old and new results.
- Converge on a standard fit procedure for R_DB fits.
- Forward Folding technique looks promising, but requires further development
- Produce a preliminary result for the paper. Any deadline for this?