Inclusive Jet Analysis

The toy RM, step by step
(Part 1)

Patrick L.S. Connor

Deutsches Elektronen-Synchrotron
Hamburg

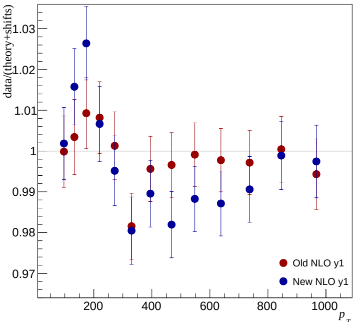
7 May 2020

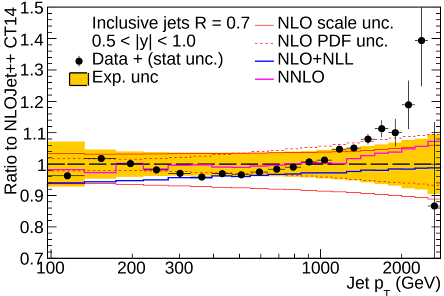
Introduction.

Introduction

Status

- AN and paper draft shared with L2 & L3 conveners
  $\longrightarrow$ no feedback...
- However, still trying to improve understanding of systematic effects
  $\longrightarrow$ remove (or at least reduce) additional bin-to-bin systematic uncertainties.
- Suspecting issue(s) somewhere in unfolding because of tensions in QCD fits:
  - tension between three first rapidity bins and forward region;
  - instabilities close to the edge of the phase space.





Introduction

Generalities

Traditionally, two methods exist to construct the RM:

- **sample** obtained directly from matching jets in MC samples
- **toy** constructed by hand from "a" gen-level spectrum and the smearing obtained from the MC samples

⟶ pros and cons ⟶ *see appendix*

Present analysis

- Avoid additional statistical uncertainties from MC samples (also a natural way to avoid problem of regularisation).
- Reduce model dependence as much as possible.
- Correctly account for correlations among jets in a single event.
- Disentangle all possible sources of systematic uncertainties to improve control on pulls in QCD fits.

⟶ despite the (potential) need for multi-dimensional unfolding, seems better to go for a toy RM.

Introduction

Current implementation (corresponding to the AN)

- Toy 1D unfolding.
- Background, inefficiencies & migrations among rapidity bins applied as correction factors.
- Transverse momentum is Gaussian-smeared (*i.e.* no correction for non-Gaussian deviations).

Revisited (and improved) procedure

1. First reproduce effect of detector to go from gen to rec level $\longrightarrow$ topic of the day.
2. Then construct iteratively a gen-level spectrum with which we will construct the RM to unfold the rec-level $\longrightarrow$ topic of a next presentation.

Rec level.

Background
Inefficiencies
Resolution
Migrations among rapidity bins
Toy / rec

Rec level

Effects

- Gaussian width $\longrightarrow$ resolution
- Gaussian mean $\longrightarrow$ energy scale
- Background $\longrightarrow$ fake jets
- Inefficiencies $\longrightarrow$ miss jets
- *Cross-rapidity migrations*
- *Corrections to Gaussian Ansatz*

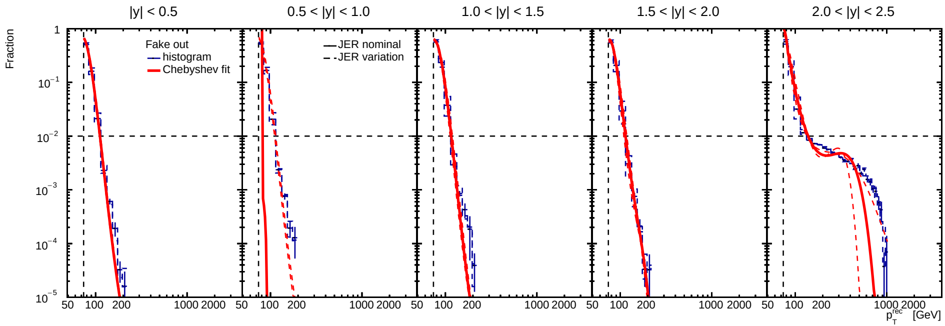
$\longrightarrow$ for the two last ones: the migrations among rapidity bins are usually treated as miss/fake entries in 1D unfolding; corrections to Gaussian Ansatz are usually ignored (although sometimes a Crystal-Ball function is used...)

Goal

Get a smooth function describing each of these effects separately.

Background

Migrations out of phase space

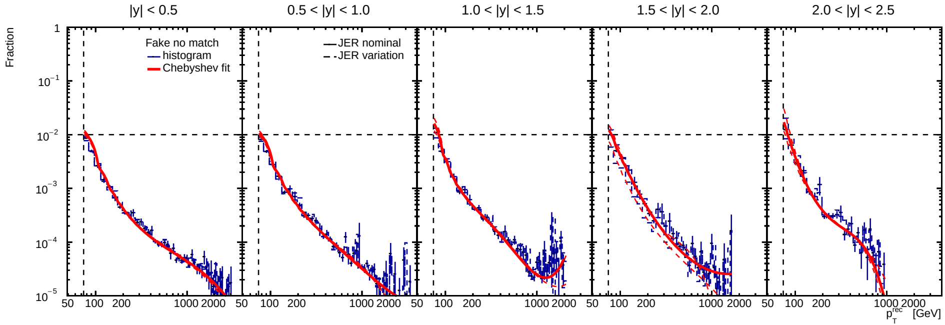


Figure

- Steeply falling spectrum.
- Applied as a correction factor after the convolution.
- Significantly large values may have large impact.
- Result in two first rec-level bins is expected to be unstable.
- Migrations to $|y| > 2.5$ are visible in the last rapidity bin.

Background

Real background

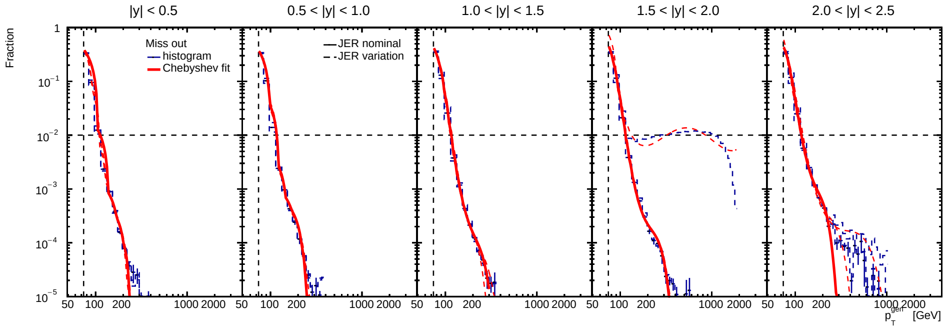


Figure

- Most likely corresponding to pile-up jets and detector noise.
- Applied as a correction factor after the convolution.
- No significantly large values.
- Not expected to cause troubles.

Inefficiencies

Migrations out of phase space

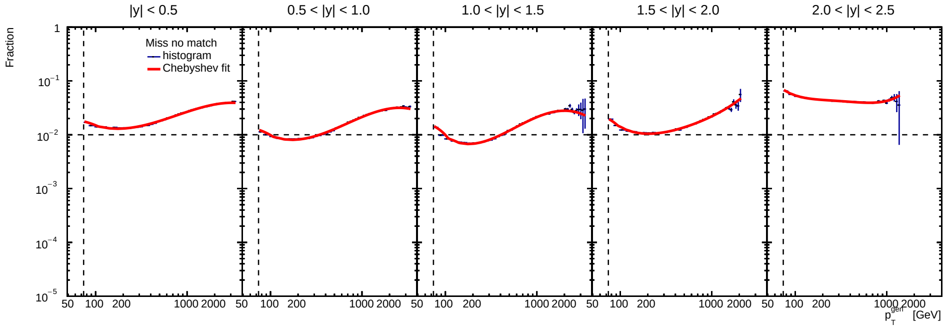


Figure

- Steeply falling spectrum.
- Applied as a correction factor in the convolution.
- Significantly large values may have large impact, especially if extracted from fit instead of histogram.
- Result in two first rec-level bins is expected to be unstable.

Inefficiencies

Real inefficiencies



Figure

- Most likely corresponding to MET filters (high p_T), hot zones (global effect), prefiring issue (large η).
- Applied as a correction factor in the convolution.
- No significantly large values.
- Not expected to cause troubles.

Resolution

Definition

$$\Delta = \frac{p_T^{\text{rec}} - p_T^{\text{gen}}}{p_T^{\text{gen}}} \qquad (1)$$

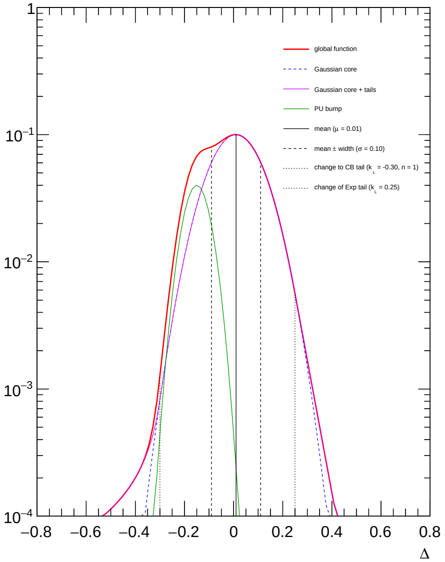
$\longrightarrow$ on the LHS (RHS), migrations to lower (larger) p_T

Shape

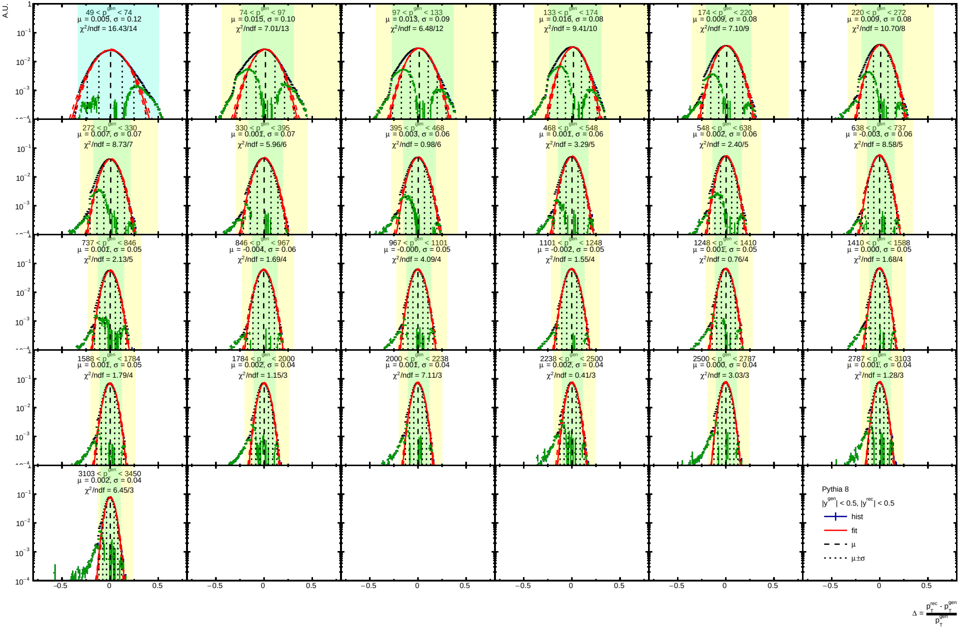
- Only RHS of the core is found to be Gaussian in all phase space!
- Deviation on LHS is due to matching with pile-up (see back-up).
- Additional tails, especially at low p_T and large y.

Next slides

Normalised resolution curves (histogram, Gaussian fit, and difference)



Resolution


A.U.
$49 < p_T^{gen} < 74$
μ = 0.005, σ = 0.12
χ^2/ndf = 16.43/14
$74 < p_T^{gen} < 97$
μ = 0.015, σ = 0.10
χ^2/ndf = 7.01/13
$97 < p_T^{gen} < 133$
μ = 0.013, σ = 0.09
χ^2/ndf = 6.48/12
$133 < p_T^{gen} < 174$
μ = 0.016, σ = 0.08
χ^2/ndf = 9.41/10
$174 < p_T^{gen} < 220$
μ = 0.009, σ = 0.08
χ^2/ndf = 7.10/9
$220 < p_T^{gen} < 272$
μ = 0.009, σ = 0.08
χ^2/ndf = 10.70/8
$272 < p_T^{gen} < 330$
μ = 0.007, σ = 0.07
χ^2/ndf = 8.73/7
$330 < p_T^{gen} < 395$
μ = 0.001, σ = 0.07
χ^2/ndf = 5.96/6
$395 < p_T^{gen} < 468$
μ = 0.003, σ = 0.06
χ^2/ndf = 0.98/6
$468 < p_T^{gen} < 548$
μ = 0.001, σ = 0.06
χ^2/ndf = 3.29/5
$548 < p_T^{gen} < 638$
μ = 0.002, σ = 0.06
χ^2/ndf = 2.40/5
$638 < p_T^{gen} < 737$
μ = -0.003, σ = 0.06
χ^2/ndf = 8.58/5
$737 < p_T^{gen} < 846$
μ = 0.001, σ = 0.05
χ^2/ndf = 2.13/5
$846 < p_T^{gen} < 967$
μ = -0.004, σ = 0.06
χ^2/ndf = 1.69/4
$967 < p_T^{gen} < 1101$
μ = -0.000, σ = 0.05
χ^2/ndf = 4.09/4
$1101 < p_T^{gen} < 1248$
μ = -0.002, σ = 0.05
χ^2/ndf = 1.55/4
$1248 < p_T^{gen} < 1410$
μ = 0.001, σ = 0.05
χ^2/ndf = 0.76/4
$1410 < p_T^{gen} < 1588$
μ = 0.000, σ = 0.05
χ^2/ndf = 1.68/4
$1588 < p_T^{gen} < 1784$
μ = 0.001, σ = 0.05
χ^2/ndf = 1.79/4
$1784 < p_T^{gen} < 2000$
μ = 0.002, σ = 0.04
χ^2/ndf = 1.15/3
$2000 < p_T^{gen} < 2238$
μ = 0.001, σ = 0.04
χ^2/ndf = 7.11/3
$2238 < p_T^{gen} < 2500$
μ = 0.002, σ = 0.04
χ^2/ndf = 0.41/3
$2500 < p_T^{gen} < 2787$
μ = 0.000, σ = 0.04
χ^2/ndf = 3.03/3
$2787 < p_T^{gen} < 3103$
μ = 0.001, σ = 0.04
χ^2/ndf = 1.28/3
$3103 < p_T^{gen} < 3450$
μ = 0.002, σ = 0.04
χ^2/ndf = 6.45/3
Pythia 8
$|y^{gen}| < 0.5$, $|y^{rec}| < 0.5$
hist
fit
μ
$\mu\pm\sigma$
$\Delta = \frac{p_T^{rec} - p_T^{gen}}{p_T^{gen}}$

Resolution

Pile-up bump?

Matching?

- Scan rec-level and look for a match a gen-level or *vice-versa*? $\longrightarrow$ does not change
- Within a cone, we take the highest p_T jet
  $\longrightarrow$ but don't take the closest (too sensitive to pile-up & screws up the shape of the distribution).

Ideas?

- Include condition based on PU jet discriminator?
  $\longrightarrow$ limited to low p_T^{rec}
- Look at jet constituents?
  $\longrightarrow$ potentially very long and tedious task...
- ...?

Resolution

Pile-up bump?

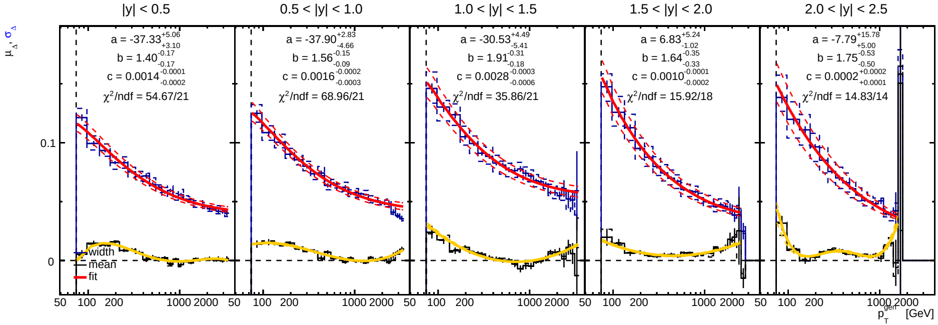
Is it relevant?

- The bump itself is only a small (but sizeable, unlike the tails) effect ($\sim$ statistical uncertainties).
- At the moment, it is just ignored, and some small test show that the effect is indeed minor.
- At the end, we should account for it...
- ... but more important is that it spoils the fit of the core of the resolution curve...

$\longrightarrow$ **especially the mean will be less precise**

Resolution

Resolution width



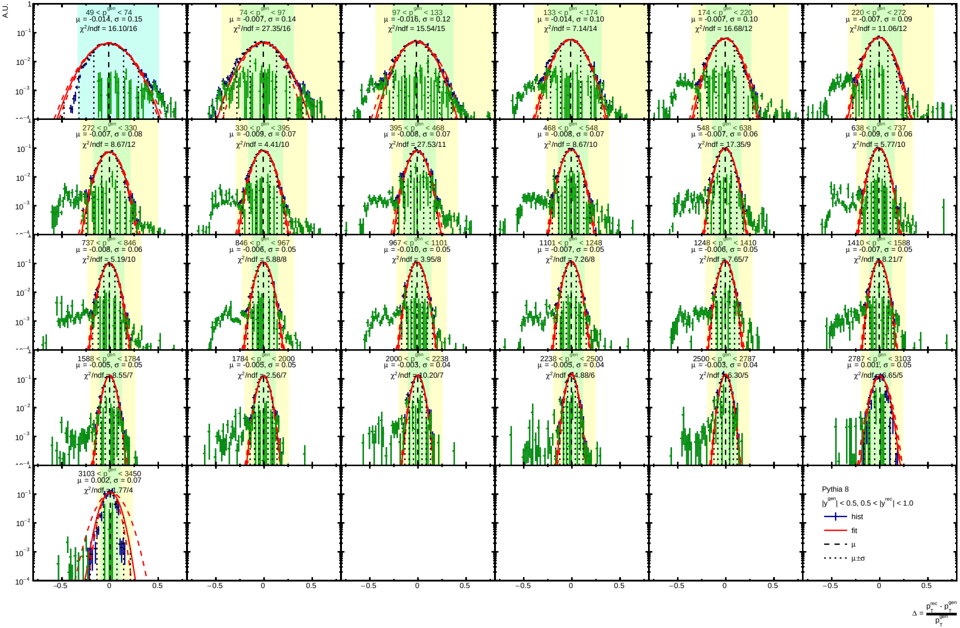
Fit function for resolution width

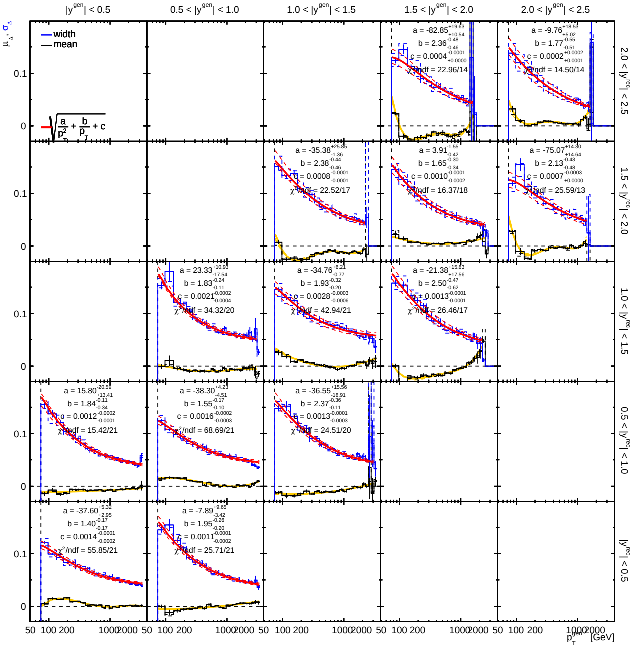
$$\frac{\Delta}{p_T^{\text{gen}}} = \sqrt{\frac{a}{p_T^{\text{gen}2}} + \frac{b}{p_T^{\text{gen}}} + c} \qquad (2)$$

Figure

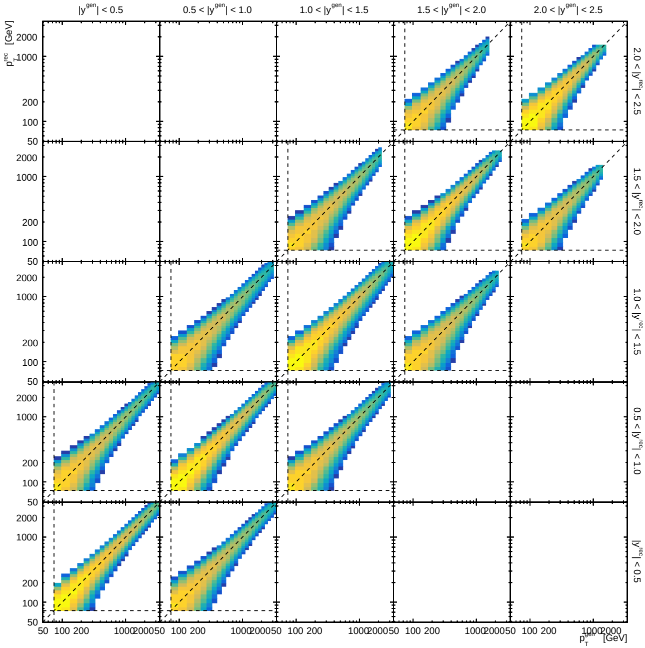
- Shown here for 1D (see later for 2D).
- **Only the RHS of the core of the resolution curve is found to be Gaussian.**
- Mean is shown in black, and fitted with a Chebyshev polynomial in orange (note: not exactly used as such later in the presentation).

Migrations among rapidity bins





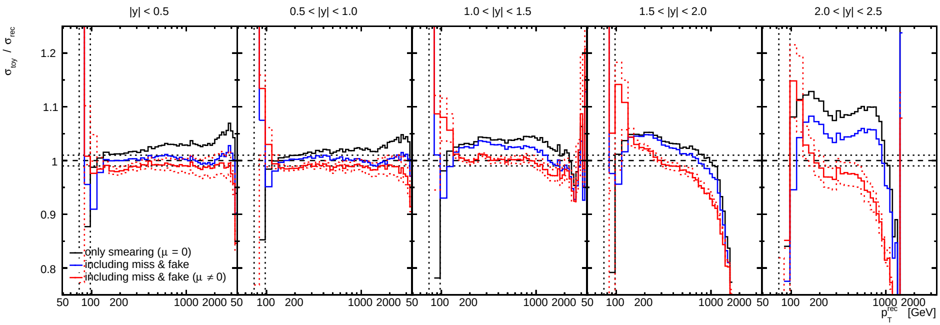
|y^gen| < 0.5
0.5 < |y^gen| < 1.0
1.0 < |y^gen| < 1.5
1.5 < |y^gen| < 2.0
2.0 < |y^gen| < 2.5
2.0 < |y^rec| < 2.5
1.5 < |y^rec| < 2.0
1.0 < |y^rec| < 1.5
0.5 < |y^rec| < 1.0
|y^rec| < 0.5
μ_Δ, σ_Δ
width
mean
$\sqrt{\frac{a}{p_T^2} + \frac{b}{p_T} + c}$
a = -82.85
b = 2.36
c = 0.0004
χ²/ndf = 22.96/14
a = -9.76
b = 1.77
c = 0.0002
χ²/ndf = 14.50/14
a = -35.38
b = 2.38
c = 0.0008
χ²/ndf = 22.52/17
a = 3.91
b = 1.65
c = 0.0010
χ²/ndf = 16.37/18
a = -75.07
b = 2.13
c = 0.0007
χ²/ndf = 25.59/13
a = 23.33
b = 1.83
c = 0.0021
χ²/ndf = 34.32/20
a = -34.76
b = 1.93
c = 0.0028
χ²/ndf = 42.94/21
a = -21.38
b = 2.50
c = 0.0013
χ²/ndf = 26.46/17
a = 15.80
b = 1.84
c = 0.0012
χ²/ndf = 15.42/21
a = -38.30
b = 1.55
c = 0.0016
χ²/ndf = 68.69/21
a = -36.55
b = 2.37
c = 0.0013
χ²/ndf = 24.51/20
a = -37.60
b = 1.40
c = 0.0014
χ²/ndf = 55.85/21
a = -7.89
b = 1.95
c = 0.0011
χ²/ndf = 25.71/21
p_T^gen [GeV]


|y^gen| < 0.5
0.5 < |y^gen| < 1.0
1.0 < |y^gen| < 1.5
1.5 < |y^gen| < 2.0
2.0 < |y^gen| < 2.5
p_T^rec [GeV]
p_T^gen [GeV]
2.0 < |y^rec| < 2.5
1.5 < |y^rec| < 2.0
1.0 < |y^rec| < 1.5
0.5 < |y^rec| < 1.0
|y^rec| < 0.5
50
100
200
1000
2000

Figure

Gen level is smoothed and compared to the rec level of Pythia after all usual corrections for $p_T > 74$ GeV.

Toy / rec



Comments

- The effect of the mean and width of the resolution are by far the dominant effects.
- Background and inefficiencies are also sizeable effects.
- However, the deviations from the Gaussian shape themselves are minor effect (but the pile-up bump screws up the fit of the resolution curves).

$\longrightarrow$ actually, the mean used here is not extracted from the fit, but extracted from the histogram (better, but still suboptimal).

Summary & Conclusions.

Summary & Conclusions

- The goal of the exercise is to define a set of analytical functions able to reproduce the effect of the detector in order to construct a toy RM.
- Only missing piece is a robust fit that gives a good value for the mean.
- Once this is done, proceed to next step and reconstruct the gen level of Pythia based on analytical functions, and then go to unfold real data (AK4 + AK7).

Thank you for your attention!

Back-up.

Sample vs toy

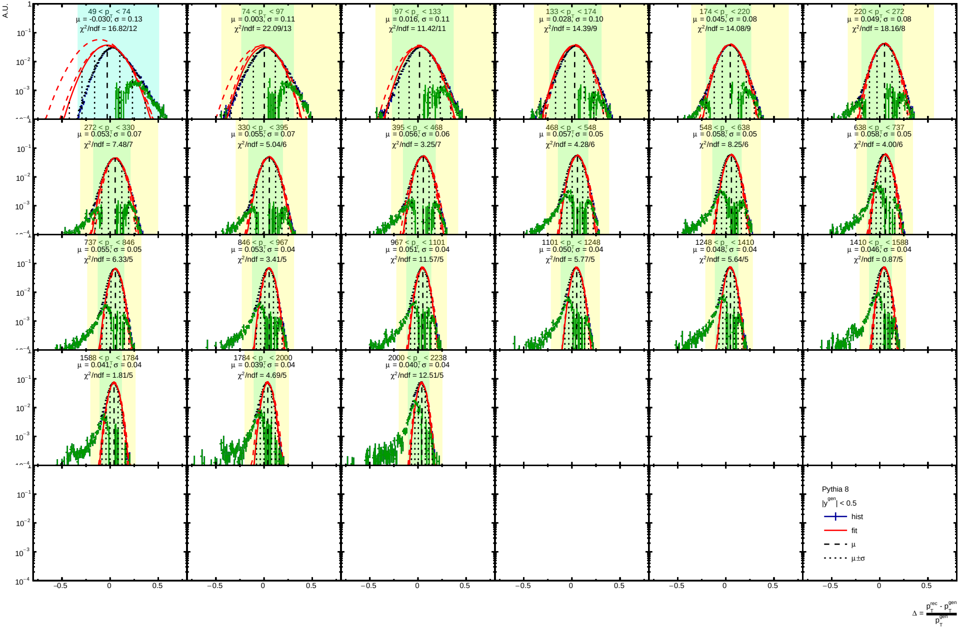
Sample

1. Easier to implement, even for $d > 1$.
2. Limited by statistics of the sample, potentially implying the need for regularisation.
3. Model dependence not straightforward to estimate.
4. Should include all effects from the detector at a time.

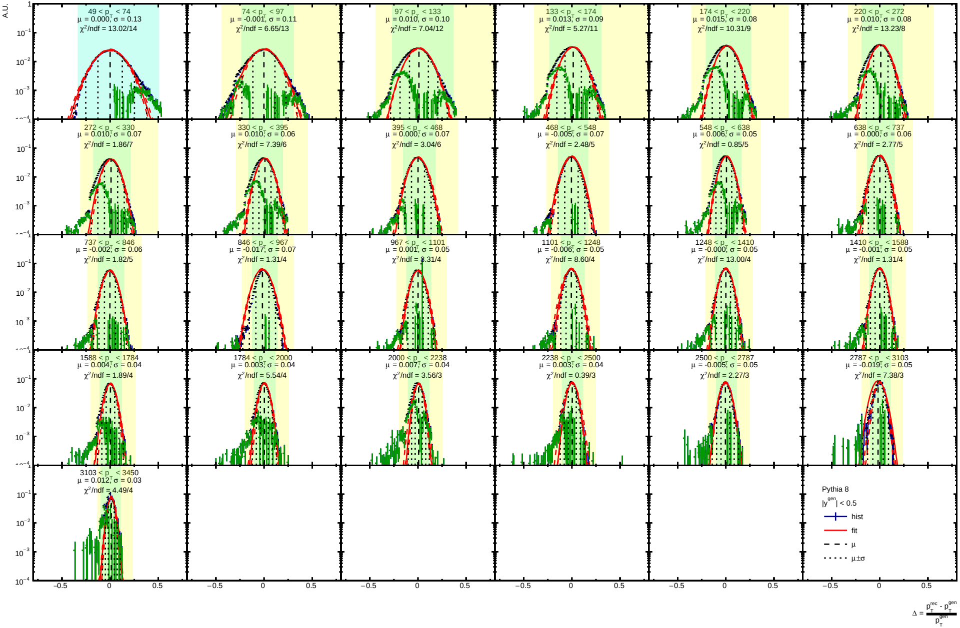
Toy

1. Less easy to implement, especially for $d > 1$.
2. Take "some theory curve" and smear it with resolution obtained from the MC samples (after JES corrections & JER smearing).
3. Not limited by statistics, avoid additional regularisation.
4. Better control on model dependence by appropriate choice of gen-level distribution.

Pythia Flat w/o PU



Pythia Flat w/ PU



Glossary I

AK4 anti k_T algorithm ($R = 0.4$). 22

AK7 anti k_T algorithm ($R = 0.7$). 22

AN Analysis Note. 3, 5

JER Jet Energy Resolution. 24

JES Jet Energy Scale. 24

MC Monte Carlo. 4, 24

PU pile-up. 14

QCD Quantum Chromodynamics. 3, 4

RM Response Matrix. 4, 5, 22

References I