Inclusive jet measurement

Toy, episode 2

Patrick L.S. Connor

Deutsches Elektronen-Synchrotron
Hamburg

4 June 2020

Introduction.

Outline
Remaining challenges
Reminder

Outline

Today

- Reviewing remaining challenges before proceeding to QCD fits (by Katerina & Toni).
- Then focusing on recent investigations to improve the toy approach.
- Last time, I reported on unsuccessful attempts to include non-Gaussian deviations in the fit: this is now under better control.
- Work in (good) progress.

Warning

This is a quite technical presentation, but it potentially concerns all analyses using a toy approach to perform unfolding.

Introduction I

Jet Energy Resolution

- Mean of resolution matters in smearing procedure, and cannot be assumed to be zero (especially for AK7).
- Tails of resolution seem to matter as well.

$\longrightarrow$ impact on construction of toy RM.

Unfolding

- Suspicions on background estimation.
- Currently observing a non-physical tension between low- and high-rapidity bins.
- Missing proper estimation of correlations among rapidity bins.

$\longrightarrow$ currently working on improving construction of toy RM, technical details are discussed in this presentation.

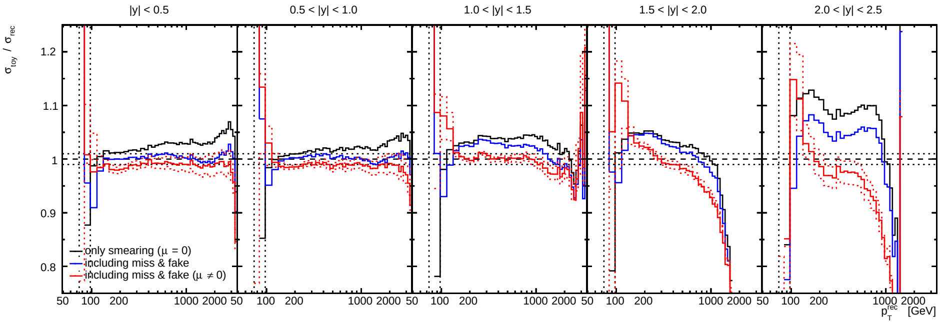
Introduction II

Bin-to-bin non-statistical deviations

- A few sources already identified and solved (JES, effective lumi instead of prescales, inconsistency in JSON file, ...).
- Seems to be present also in other analyses.
- Unclear whether there is anything left at detector level (Chebyshev fits indicate it is the case).
- There might be a contribution from the unfolding to the issue.

$\longrightarrow$ once the unfolding is improved, we can go back to these questions.

Reminder



From a few weeks ago

Gen level is smoothed and smeared, then compared to the rec level of Pythia after all usual corrections for $p_T > 74\ \mathrm{GeV}$:

- Expecting agreement at one.
- Using here a **purely Gaussian smearing**, also accounting for miss and fakes.
- Closure is failing for $|y| > 1.0$: can it come from the tails?

$\longrightarrow$ work in progress, few insights on next slides...

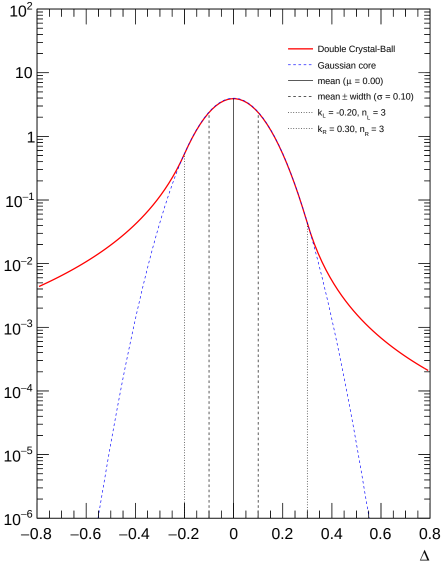
Fitting with Crystal-Ball.

Fitting with Crystal-Ball

Recent progresses

- Setting up automated algorithm to fit the resolution curves with double-sided Crystal-Ball function (shown on next slide) to get the non-Gaussian effects in the tail.
- Non-Gaussian effects in the core should in principle be covered by the JER uncertainties (see figure & back-up).
- Main difficulty is to find an algorithm to fit so many parameters.

$\longrightarrow$ now we want to investigate deviations from Gaussian behaviour with Crystal-Ball tail



(Double) Crystal Ball

Definition

$$f(x) = N \cdot \begin{cases} A_2(B_2 + z)^{-n+1} & \text{for} \quad z \geq \alpha_2 \\ \exp \frac{-1}{2} z^2 & \text{for} - \alpha_1 < z < \alpha_2 \\ A_1(B_1 - z)^{-n+1} & \text{for} \quad z \leq -\alpha_1 \end{cases}$$

where

$$z = \frac{x - \mu}{\sigma}$$

$$A_i = \left(\frac{n_i}{|\alpha_i|}\right)^n \exp \frac{-1}{2} |\alpha_i|^2$$

$$B_i = \frac{n_i}{|\alpha|} - |\alpha_i|$$

$$C_i = \frac{n_i}{\alpha_i} \frac{1}{n_i - 1} \exp \frac{-1}{2} |\alpha_i|^2$$

$$D = \sqrt{\frac{\pi}{2}} \left(\text{erf} \frac{|\alpha_2|}{\sqrt{2}} + \text{erf} \frac{|\alpha_1|}{\sqrt{2}} \right)$$

$$N = \frac{1}{\sigma(C_1 + C_2 + D)}$$

Find the transition points

Problem

We need to automate the fit of many many parameters in ~ 100 bins, including the transitions points
$\longrightarrow$ brutal force method with releasing all parameters just does not work: one needs to provide realistic starting values and ranges.

Trick to find the Gaussian core

In log scale, a Gaussian is just a parabola
$\longrightarrow$ first (second) derivative of the logarithm of the Gaussian is just a line with a nonzero slope (constant):

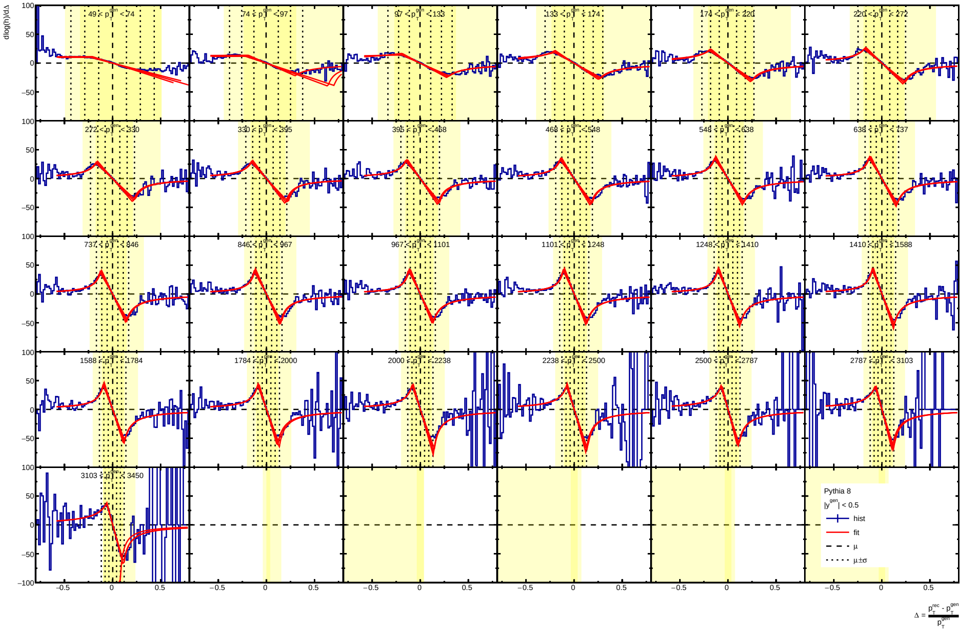
$$\log f(x) = \log N - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big(\log f(x)\Big) = -\frac{x-\mu}{\sigma^2}$$

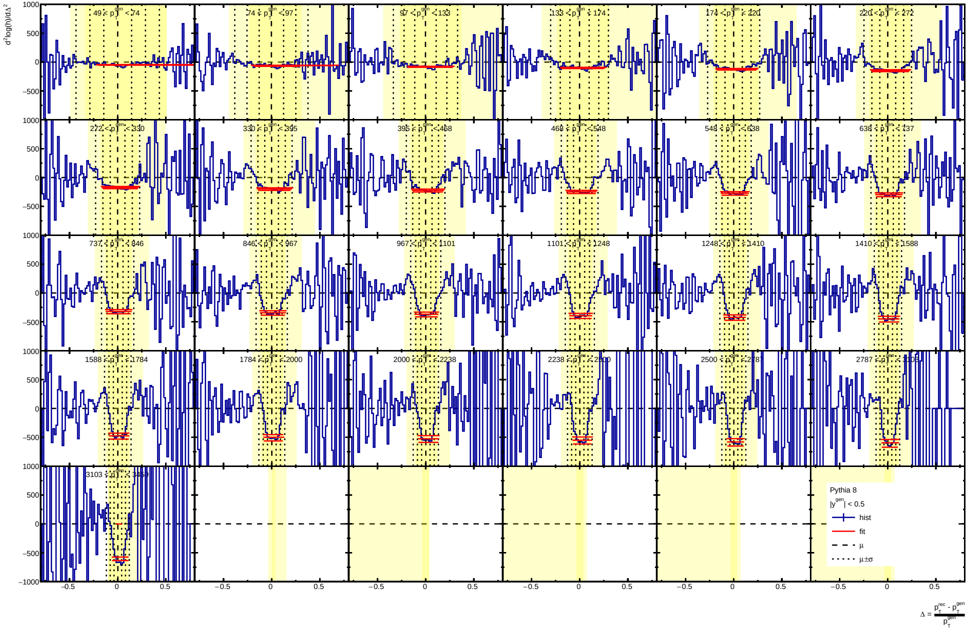
$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Big(\log f(x)\Big) = -\frac{1}{\sigma^2}$$

$\longrightarrow$ easy way to determine transition region between core and tails.

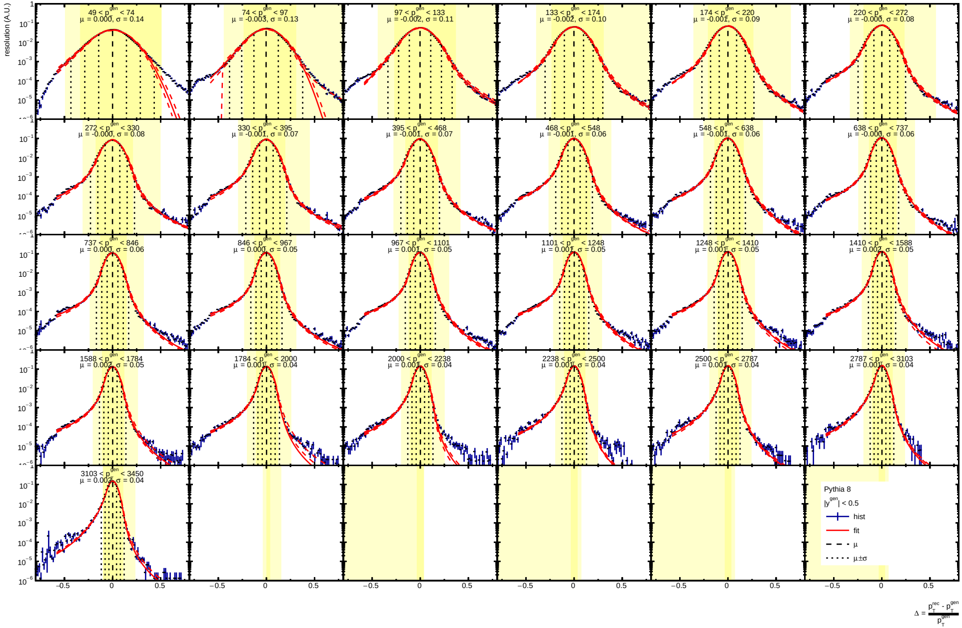
First derivative of log



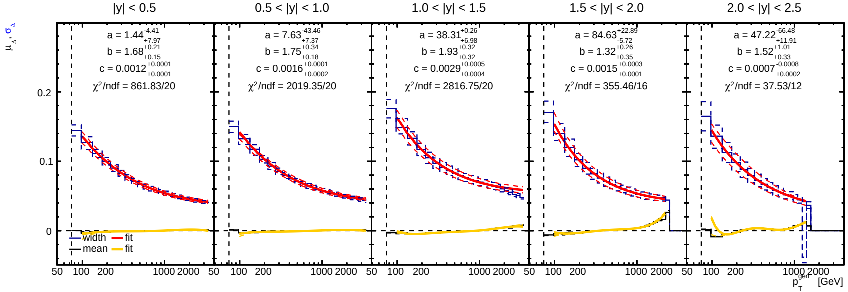
Second derivative of log



Fit in log scale



Gaussian-core parameters



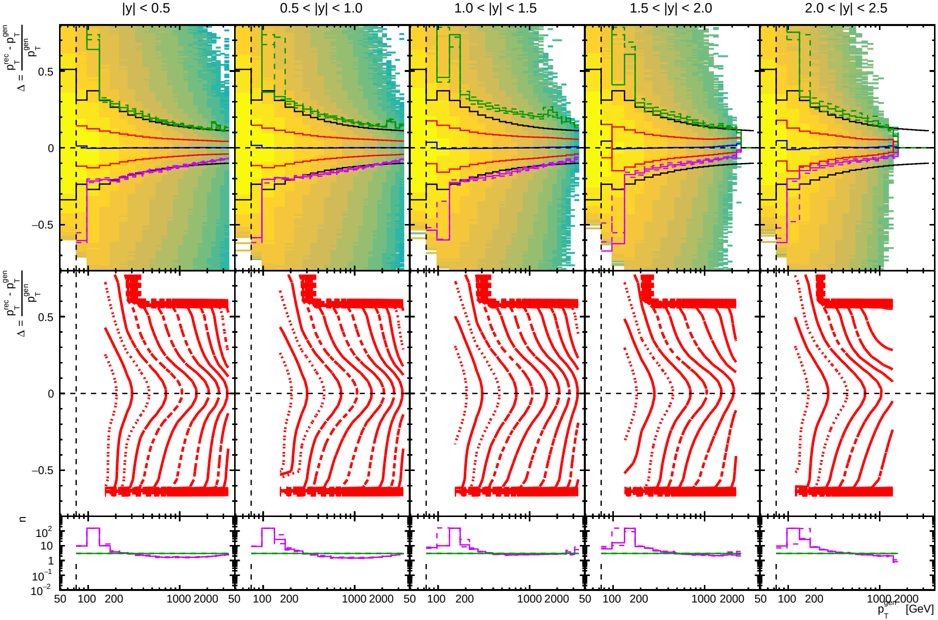
Figure

- Fits seem reasonable, event at low p_T (before, there were big jumps).
- Currently going down to 97, but it seems that one could go down to 74 $\longrightarrow$ this is new, since I reran the whole analysis with lower p_T thresholds to allow matching with lower p_T.

Next slide

We show about core & tail parameters (μ, σ, $k_{L,R}$, $n_{L,R}$) as well as 2D resolution function.

All parameters

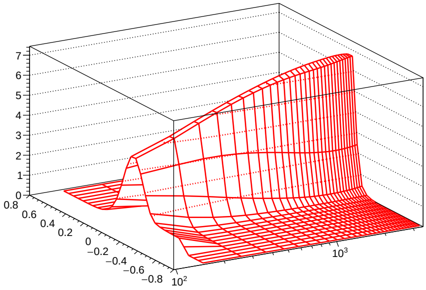


Back to the toy.

Short recap & plans

Fit with double Crystal-Ball function

- Fit seems to work nicely in each $p_T, y)$ bin separately.
- Now we want a continuous function over p_T.
- We can also fit all various parameters quite easily to get a continuous behaviours as a function of p_T, using Chebyshev polynomials.



Reminder

- The distribution at detector level is a convolution of the cross section and of the resolution function.
- To construct the RM, we need to integrate in two dimensions in each region of the phase space corresponding to a bit.
- Then we can fold the cross section to compare to the reco level distribution from Pythia (see first figure of the presentation).

Back to the toy

Problem

Integrating numerically takes way too long...

Solution

Integrate analytically over the resolution!

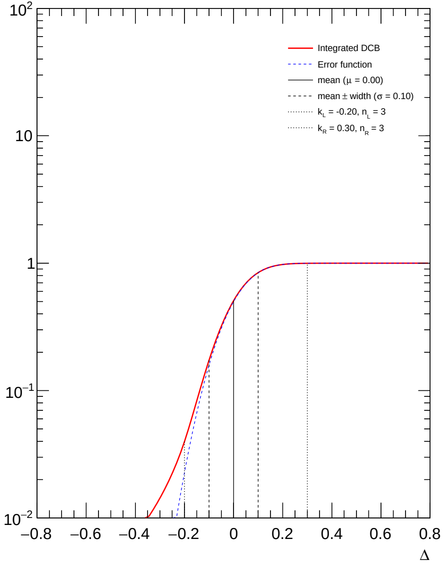
Gaussian core

The error function is precisely defined as the integral of a Gaussian:

$$\int \exp \frac{-1}{2} z^2 \, \mathrm{d}z = \sqrt{\frac{\pi}{2}} \operatorname{erf} \frac{z}{\sqrt{2}}$$

Crystal Ball tail

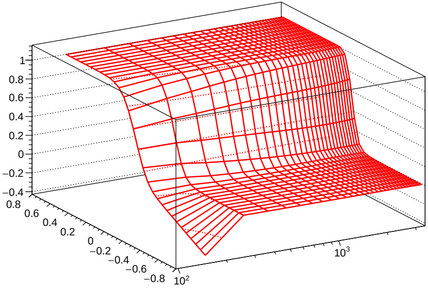
$$\int NA \cdot (B \pm z)^{-n} \, \mathrm{d}z = \frac{\pm NA}{n-1} (B \pm z)^{1-n}$$



Recap & prospects

In practice

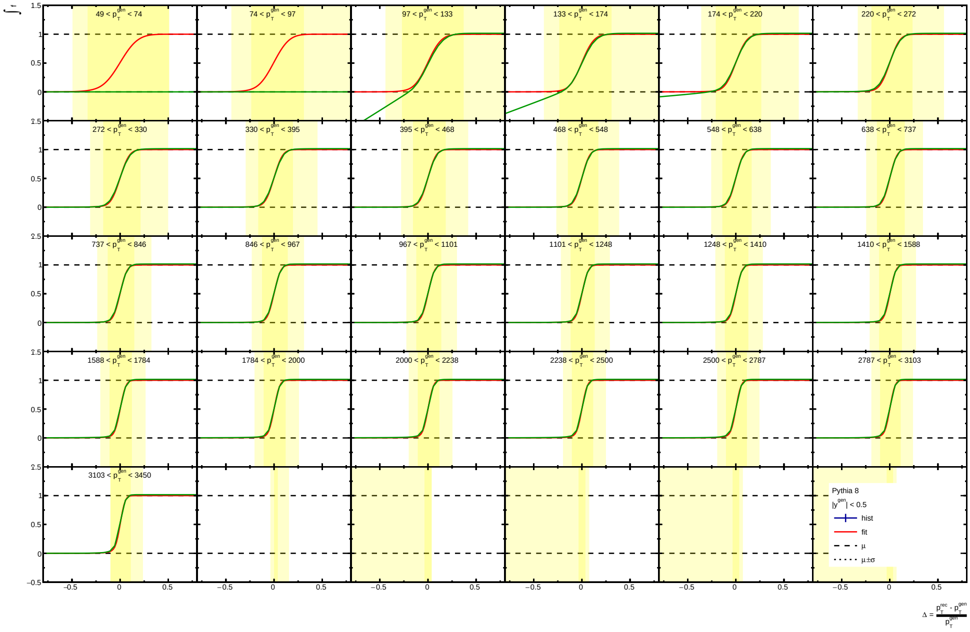
- Take all the parameters from the fit.
- Just plug them in the formula for the semi-analytical integral.
- The integral over the resolution is shown in the figure.
- (Then, in principle, one just needs to integrate over p_T.)



Next slide

We look at the integral, either in each p_T bin (red) or in the smoothed 2D function (green).

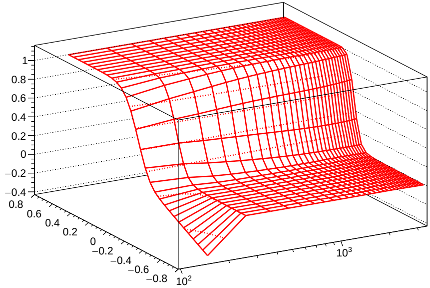
Integral


49 < p_T^gen < 74
74 < p_T^gen < 97
97 < p_T^gen < 133
133 < p_T^gen < 174
174 < p_T^gen < 220
220 < p_T^gen < 272
272 < p_T^gen < 330
330 < p_T^gen < 395
395 < p_T^gen < 468
468 < p_T^gen < 548
548 < p_T^gen < 638
638 < p_T^gen < 737
737 < p_T^gen < 846
846 < p_T^gen < 967
967 < p_T^gen < 1101
1101 < p_T^gen < 1248
1248 < p_T^gen < 1410
1410 < p_T^gen < 1588
1588 < p_T^gen < 1784
1784 < p_T^gen < 2000
2000 < p_T^gen < 2238
2238 < p_T^gen < 2500
2500 < p_T^gen < 2787
2787 < p_T^gen < 3103
3103 < p_T^gen < 3450
Pythia 8
|y^gen| < 0.5
hist
fit
μ
μ±σ
Δ = (p_T^rec - p_T^gen) / p_T^gen

Recap & prospects

Recap

- The CB fit and the integral themselves seem to work reasonably well (red curve on former slide).
- However, suspecting the issue to come from the use of inconsistent fits for the different parameters as a function of p_T (green curve).



Prospects

- Once this is fixed, we will be able to test the impact of the tails on the toy.
- If closure is successful, we can go back to unfolding (first still in simulation, then in data).

Summary & Conclusions.

Summary & Conclusions

- Significant progresses to fit resolution with Crystal-Ball function.
- Trying to use the non-Gaussian deviations in the toy: now having some difficulties with the smooth interpolation.
- As soon as closure is successful, we can go back to unfolding.

Thank you for your attention!

Back-up.

Acronyms I

AK7 anti k_T algorithm $(R = 0.7)$. 4

JER Jet Energy Resolution. 8

JES Jet Energy Scale. 5

QCD Quantum Chromodynamics. 3

RM Response Matrix. 4, 17

References I