SPONSORED BY THE

Federal Ministry
of Education
and Research

Data Driven Determination of Jet Smearing Function

C. Sander

Susy Group meeting - 28th April 09

QCD Background from MET Tails

ABCD Method: If two well separating and uncorrelated variables are present, number of bg events in signal region can be estimated.

- Challenges:
  - For RA2 Correlations in the separating variables are present (→ extended ABCD method with cross check by more than one variable pair)
  - Signal contamination results in overestimation of bg events and shape of MHT distribution

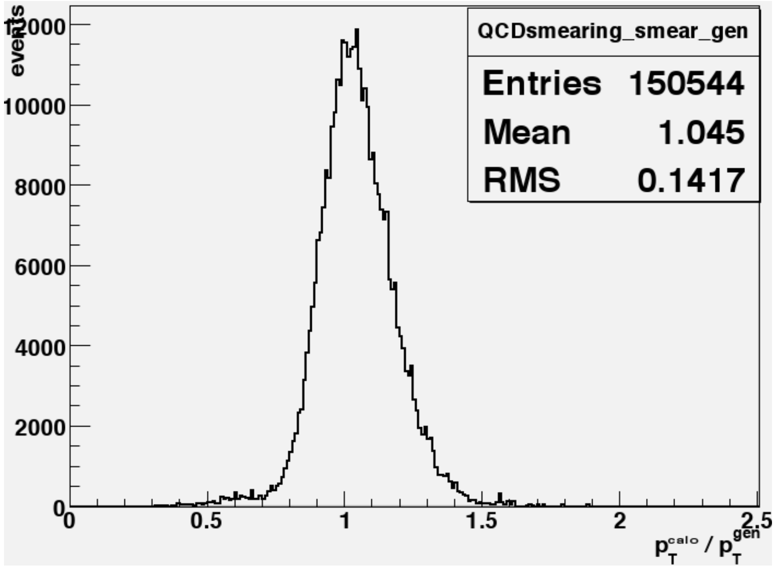
Smearing Method: Smear "well measured" QCD events according to assumed jet resolution function → "new" sample → perform cuts and look at distributions as for a MC background sample

- Challenges:
  - Normalization of smeared events
  - Data driven determination of smearing function
  - Acceptance effects (e.g. if HF neglected)

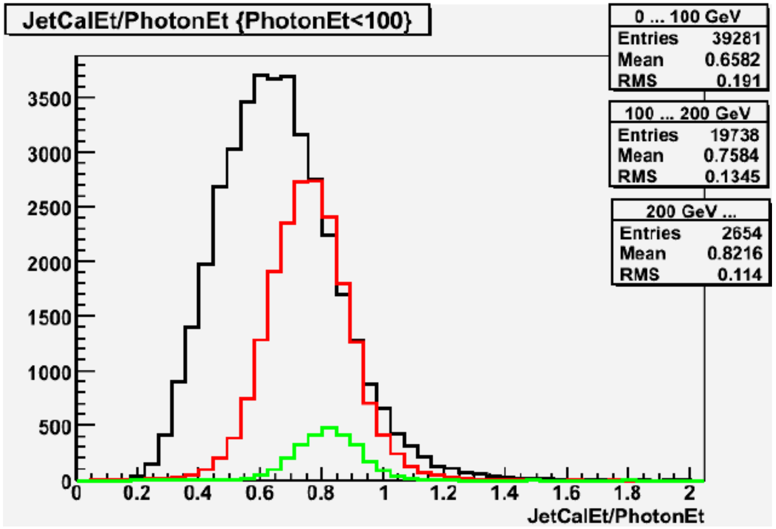
- Use MC truth: match genJets with corrected caloJets and histogram jet resolution for different p_T and η bins
- Use γ-jet or Z-jet events where truth is “known” due to precise ECal and tracker information
- ATLAS or MET projection method: select N-jet events with MET parallel or anti-parallel to one of the leading jets ... but in reality all jets are smeared
- New idea: use N-jet events but assume all jets to be potentially smeared

General advantage of N-jet events:

- Higher statistics at larger energies
- Similar flavour content of used sample and wanted bg

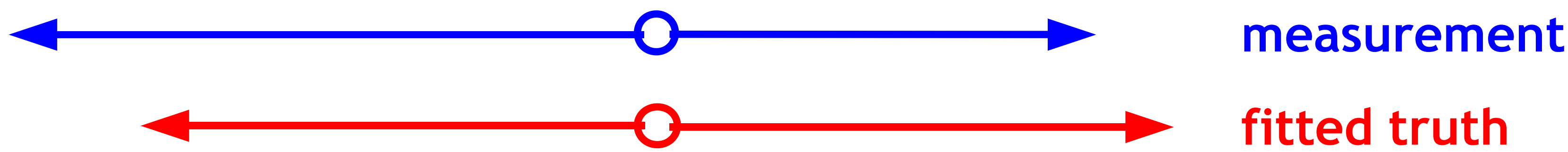


smearing func from MC truth



smearing func from γ-jet

New Idea: DiJet Balance



1) Start with first assumption of jet resolution function $r(t, m)$, e.g. from MC truth or Gaussian resolution

2) For each event: find best truth value t (same for both jets) to maximize product of $r(t, m_1)$ and $r(t, m_2)$

3) With fitted t and measured values m_i new approximation of $r'(t, m)$

4) Iterate (back to 2) until convergence

Algorithm can not work since most probable truth is determined event by event. An event with one well and one heavily miss measured jet will choose the truth somewhere in between both measurements to maximize the probability → tails will not be modelled

Newest Idea: DiJet Balance

Fit jet response function for a whole sample of events simultaneously

- Calculate probability of a DiJet event for given response function r and prior truth probability desity function f (e.g. $= 1$):

$$p_{j_1,j_2} = \frac{\int_{\text{truth}} r(j_1/t) \cdot r(j_2/t) f(t) dt}{\int \int \int_{\text{truth}} r(j_1/t) \cdot r(j_2/t) f(t) dt \; dj_1 \; dj_2}$$

- Evaluation of normalization integral: First substitute $s_1 = j_1/t$ and $s_2 = j_2/t$ and integrate over s_1 and s_2 (normalized $= 1$)

$$\int_{\text{truth}} t^2 dt = \frac{1}{3}(T^3_{\text{max}} - T^3_{\text{min}})$$

- Overall probability has to be maximized: $p_{\text{total}} = \prod_{i=1}^{N} p^i_{j_1,j_2}$
- Minimize negative log Likelihood:

$$L = -\sum_{i=1}^{N} \log(p^i_{j_1,j_2})$$

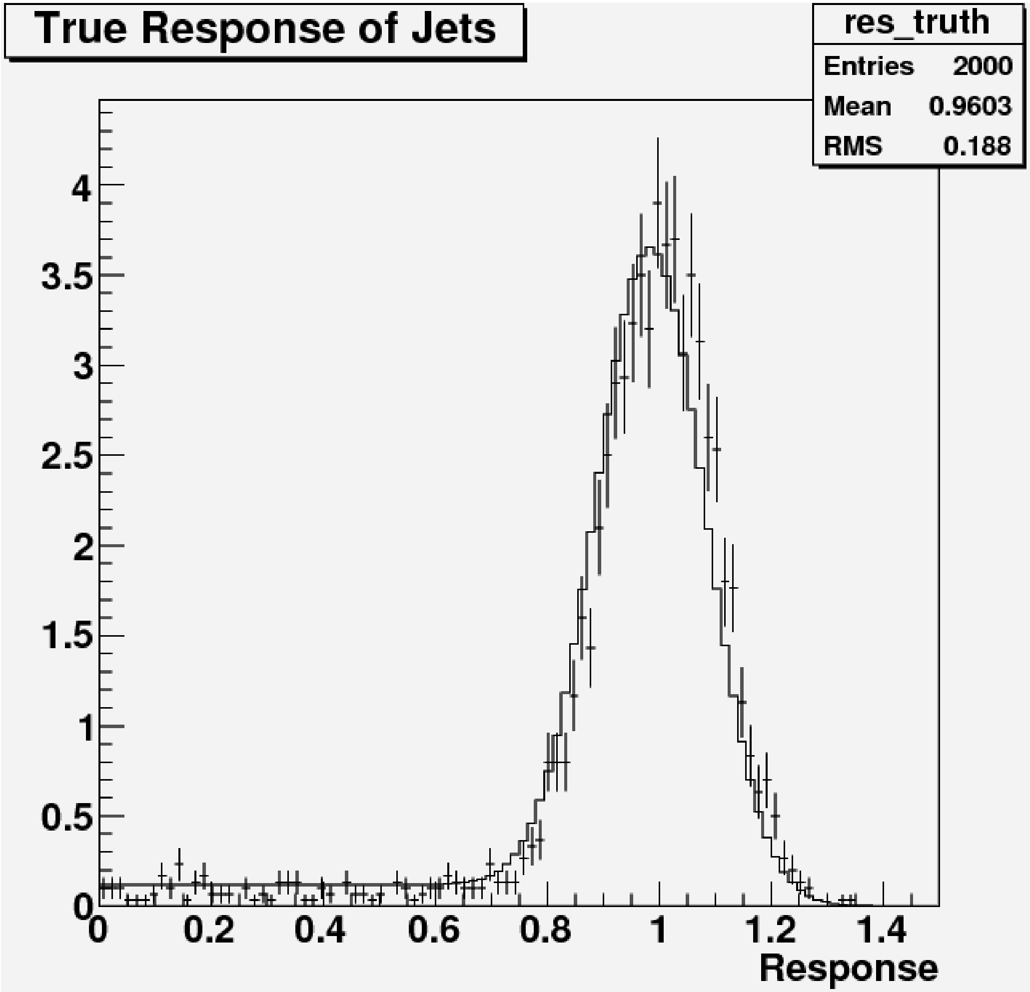
Proof of Principle: ToyMC

- Dice DiJet truth flat between $T_{\max}$ and $T_{\min}$ (here 100 and 1000)
- Smear both jets normal distributed around peak response (here 1.0)
- For both jets there is the small probability (here 0.1) that only a fraction x (here flat distributed from 0 to 1) from the total energy is measured
- Choose a parametrization for the response function (this is the most difficult part). Some possibilities:
  - Histogram (advantage: easily normalizable; disadvantage: many free parameters, discontinuities)
  - Spline (advantage: smooth behavior; disadvantage: difficult to normalize, negative values ???)
  - Functional expression (advantage: small number of free parameters, sometimes simply normalizable; disadvantage: might not describe the data correctly)
- Here (normal distribution + Fermi distribution):

$$r(x) = c \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} + (1-c) \cdot \frac{1}{T \log(1 + \exp \mu/T)} \cdot \frac{1}{\exp{(x-\mu)/T} + 1}$$

- Set peak response $\mu = 1$
- Start with pure normal distribution with $\sigma = 0.15$
- Result (using LVMINI):
  - $c = 0.88(01)$
  - $\sigma = 0.098(2)$
  - $T = 0.04901(6)$

This looks fine ☺



- But many open questions:
  - Systematic shift?
  - How to generalize to NJet events? (in principal possible but integration over more than one dimension needs many resources)
  - Best parametrization?