QCD background smearing function studies

CMS Hamburg / DESY Meeting

Christian Autermann, Jula Draeger, Ulla Gebbert, Christian Sander, Matthias Schröder, Peter Schleper, Torben Schum, Hartmut Stadie, Jan Thomsen

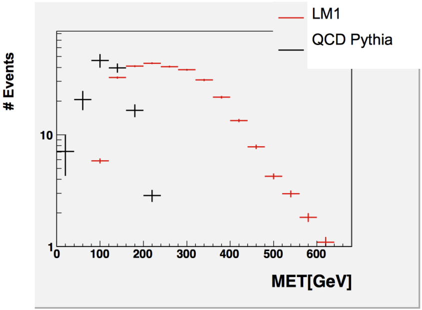
July 15th, 2009

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

Motivation: Understanding the QCD background to $\not\!\!E_T$

- Missing transverse energy, $\not\!\!E_T$, is an important signature in the search for new physics
- Large $\not\!\!E_T$-background expected from QCD events in the all-hadronic channel:
  - Particles invisible to the calorimeter e.g. μ or ν
  - Mismeasurement due to intrinsic calorimeter resolution
  - Mismeasurement due to detector acceptance
  - ...



Understanding of QCD contribution to $\not\!\!E_T$ important

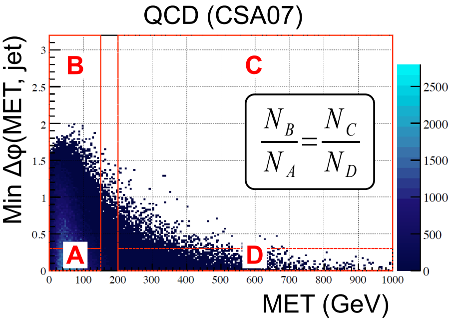
Outline

Outline

Different concepts to estimate the QCD $\not{E}_T$ contribution

- Estimation from MC simulation
- ABCD method[1]
  - Two uncorrelated variables
  - Definition of background (**A**, **B**, **D**) and signal (**C**) dominated regions by cuts on these variables
  - Prediction of **number** of QCD events in signal region

$$\boxed{N_C = N_D \frac{N_B}{N_A}}$$

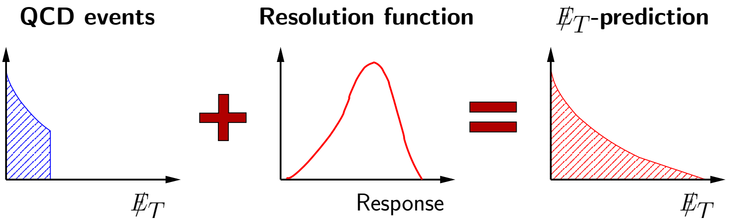


- Jet smearing method
  - This talk

[1] Talk by T. Schum http://indico.cern.ch/conferenceOtherViews.py?view=cms\&confId=37672

The jet smearing method

1. Selection of well measured QCD events
2. Smearing with resolution function $p_T^{\text{meas}}/p_T^{\text{true}}$

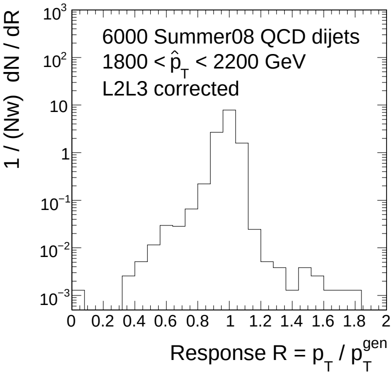


- Application of resolution function to QCD events[2] (UCSB)
  - Selection criteria
  - Normalisation of smeared events
  - Effect of double smearing
- Determination of resolution function $\longrightarrow$ This talk

[2] Talk by M. Rydenfelt http://indico.cern.ch/conferenceDisplay.py?confId=61021

Determination of the resolution function

- True jet p_T from MC simulation
  - Depends completely on simulation
- γ-jet or Z-jet events
  - $p_T^{\gamma,Z}$ relatively well measured
  - Low statistics
- MET projection method[3] (Iowa)
  - N-jets with $\not{E}_T$ parallel to one jet
  - Other jets assumed to be measured correctly
  - Neglects mismeasurement of other jets



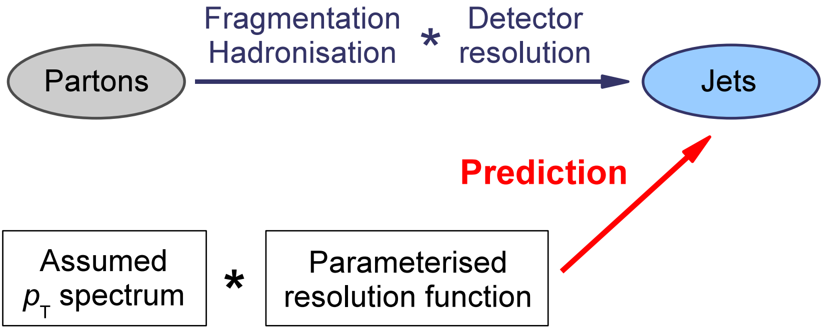
Resolution fit ⟶ this talk

- N-jet events
- All jets assumed to be mismeasured by resolution function

[3] Talk by E. Albayrak `http://indico.cern.ch/conferenceDisplay.py?confId=62099`

Outline

Concept of the resolution fit method



Details for dijet events

- In each event i, probability for a given dijet configuration p_T^{j1}, p_T^{j2} is

$$\mathcal{P}^i_{1,2} = \frac{1}{\text{norm}} \int_0^\infty \mathrm{d}p_T^{\text{true}}\, f(p_T^{\text{true}}) \cdot r_b(p_T^{j1}/p_T^{\text{true}}) \cdot r_b(p_T^{j2}/p_T^{\text{true}})$$

  - f is the probability density function (pdf) of p_T^{true}
  - r_b is the parameterised resolution function i.e. the response pdf
- Likelihood $\tilde{\mathcal{L}} = \prod_{i=0}^{N_{\text{evt}}} \mathcal{P}^i_{1,2}$ maximal for correct resolution function r_b
- Minimisation w.r.t. b of negative log-likelihood function

$$\mathcal{L} = -\sum_{i=0}^{N_{\text{evt}}} \ln(\mathcal{P}^i_{1,2})$$

- Appropriate definition of probability allows to include other data types e.g. γ-jet events

Possible parameterisations of the resolution function

Analytic function

- Smooth behaviour
- Small number of parameters
- Normalisation simple in some cases
- Functional form difficult to determine

Step function

- Normalisation simple
- Describes any distribution if appropriately binned
- Many parameters
- Discontinuities

Spline

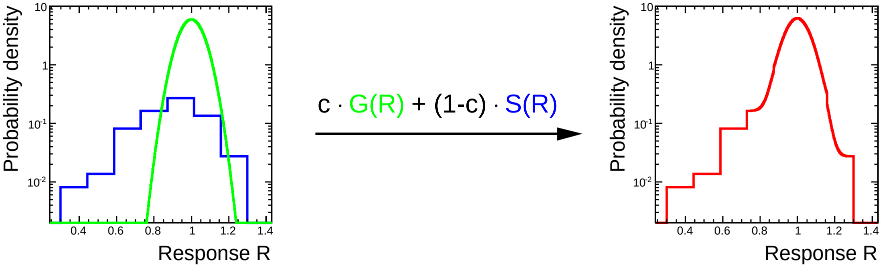
- Smooth behaviour
- Normalisation difficult
- Technically complex

The used resolution parameterisation

Superposition of Gaussian and interpolated step function

$$r_b(R) = c \cdot G(R; 1, \sigma) + (1 - c) \cdot S_N(R; \mathbf{b})$$

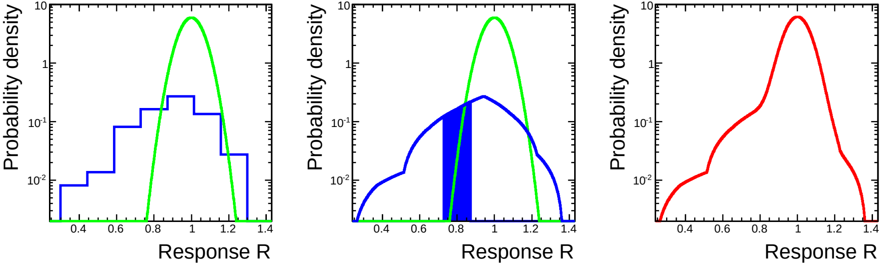
- Normalisation c
- Central Gaussian G around 1 (assume calibrated jets)
- Step function S with N parameters $\mathbf{b}$ to describe tails



Details of the resolution parameterisation

Step function S with N parameters $\mathbf{b}$ to describe tails

- Bin content $b_i \geq 0$ by construction — fit guided by penalty terms
- Actual $S_N(R)$ is linear interpolation of adjacent bin contents
- Only $N-1$ parameters are fitted $\longrightarrow$ fixed scale



The fitting framework

- Extension of the *Kalibri* tool from the University of Hamburg
- Originally developed for jet calibration via an **unbinned fit**
- Utilises `LVMINI` by V. Blobel
  - Limited memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm
  - Support for $\mathcal{O}(10^4)$ parameters
- Highly configurable object-oriented framework
- Support for **different parameterisations** and data sets
- Support for **parameter limits**
- **Multiple threads** to exploit use of multi-core processors
- Automated production of **control plots**
- Validation via **Toy Monte Carlo**
  - Generation of jet 4-momenta according to specified p_T spectrum
  - Distribution of jet energy on towers according to specified jet shape
  - Simulation of tower measurement according to specified resolution model

More information at `http://wwwiexp.desy.de/groups/cms/kalibri/api/index.html`

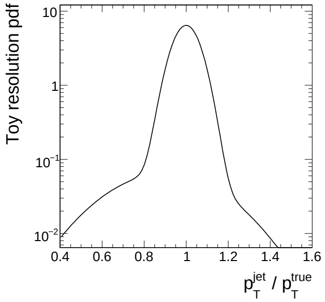
Outline

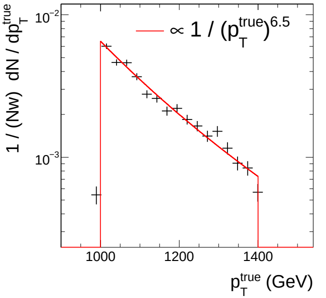
Input sample from ToyMC

- Generation of 2000 dijet events
- Simulated resolution is sum of two Gaussians

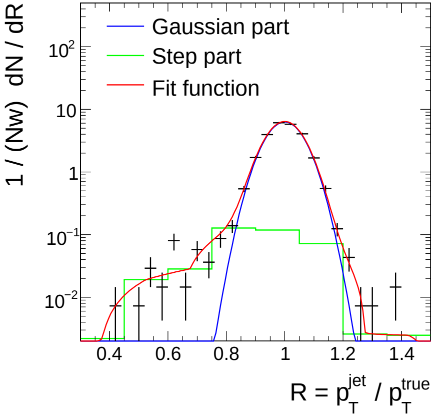
$$c \cdot G_0(1, \sigma_0) + (1 - c) \cdot G_1(\mu_1, \sigma_1)$$

  where
  - $c = 0.96$
  - $\sigma_0 = 0.06$
  - $\mu_1 = 0.9$
  - $\sigma_1 = 0.25$
- Resolution is energy independent
- p_T^{true}-spectrum falling as $1/x^{6.5}$





Results from Toy MC



- Step function with 8 bins from $0.3 < R < 1.5$
- Correct spectrum $\propto 1/(p_T^{true})^{6.5}$ assumed during fit
- Fit reproduces resolution well (numbers in the backup)

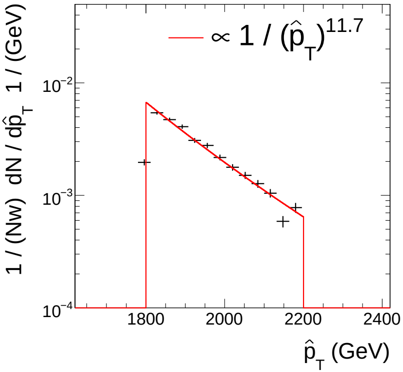
Outline

Application on Summer08 QCD dijets

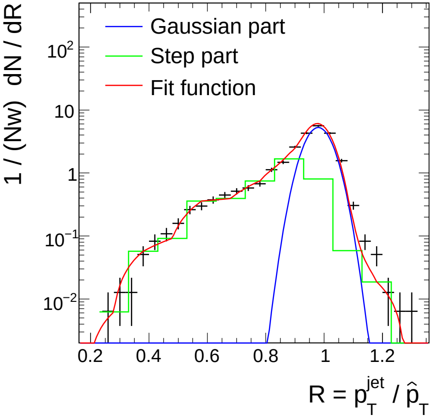
- Example $\hat{p}_T$ bin
  - $1800 < \hat{p}_T < 2200$ GeV
- Selected 1963 events out of 2000
  - Two jets leading in p_T^{gen}
  - `Jet - GenJet` matching criteria $\Delta R < 0.25$
  - $|\eta| < 1.3$ for both jets
  - $0.05 <$ hadronic fraction < 0.95
- No weights applied
- p_T^{true}-spectrum from fit on $\hat{p}_T$

$$\frac{\mathrm{d}N}{\mathrm{d}\hat{p}_T} \propto \frac{1}{\hat{p}_T^n}, \qquad n = 11.7$$

- L2L3 jet energy corrections applied

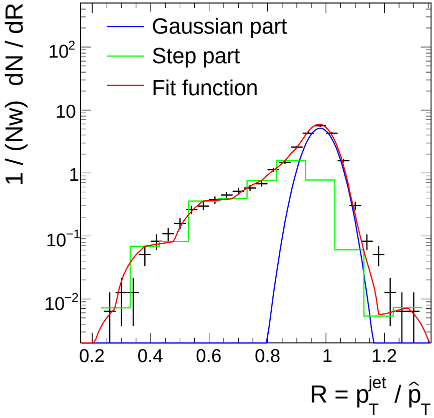


Results for Summer08 QCD dijets



- Step function with 11 bins from $0.23 < R < 1.33$
- Correct $\hat{p}_T$ spectrum $\propto 1/\hat{p}_T^{11.7}$ assumed during fit
- Fit reproduces resolution reasonably well (numbers in the backup)

Results for a fitted $\hat{p}_T$ spectrum



- Exponent n in $\hat{p}_T$ power law spectrum determined by fit
- Result $n = 11.8 \pm 0.5$ compatible with correct value 11.7
- Method can unfold spectrum and resolution

Summary

- The jet smearing method allows an estimation of the $\not{E}_T$ contribution from QCD events
- A data driven approach for the determination of the resolution function has been presented
- The method fits a parameterised resolution function using dijet events
- The resolution fit has been demonstrated to work for events generated by a Toy MC simulation
- First promising results from the Summer08 dijet production have been shown

Outlook

- Determination of resolution function for different $\hat{p}_T$ and η bins
  - Interpolation possible?
  - (How to) use knowledge of E-dependence of detector resolution?
- In data binning in p_T^{jet} and η
  - Cuts for selection of clean dijet signatures ($\Delta\phi$, E_T^{Jet3})?
  - Bias due to falling spectrum and cut-off effects?
- Knowledge about heavy flavour jet contribution
  - (How to) use in resolution parameterisation?

Backup

Mathematical details on the resolution fit method

In each event i, probability for a given dijet configuration p_T^{j1}, p_T^{j2} is

$$\mathcal{P}_{1,2}^i = \frac{1}{\text{norm}} \int_0^\infty \mathrm{d}p_T^{\text{true}}\, f(p_T^{\text{true}}) \cdot r_b(p_T^{j1}/p_T^{\text{true}}) \cdot r_b(p_T^{j2}/p_T^{\text{true}})$$

- f is the probability density function (pdf) of p_T^{true}
- r_b is the parameterised resolution function
- The normalisation is

$$\text{norm} = \int\int\int \mathrm{d}p_T^{\text{true}} \mathrm{d}p_T^{j1} \mathrm{d}p_T^{j1}\, f(p_T^{\text{true}}) \cdot r_b(p_T^{j1}/p_T^{\text{true}}) \cdot r_b(p_T^{j2}/p_T^{\text{true}})$$

Fitted parameters for Toy example

Parameter	Scale	Start value	Fitted value
c	1	0.96	0.944033
μ	1	1	1
σ	0.1	0.6	0.599104
0	0.01	1	0.928335
1	0.1	1	0.806321
2	0.1	1	1.19273
3	0.5	1	1.07337
4	0.5	1	1
5	0.5	1	0.604311
6	0.1	1	0.10949
7	0.01	1	1.04362

Fitted parameters for Summer08 dijets — fixed spectrum

Index	Scale	Start value	Fitted value
c	1	0.7	0.578505
μ	1	0.98	0.98
σ	0.1	0.5	0.436586
0	0.01	1	0.834385
1	0.1	1	0.76398
2	0.1	1	1.22063
3	0.5	1	0.954766
4	0.5	1	1.05236
5	1	1	1
6	2	1	1.11558
7	1	1	1.07393
8	0.1	1	0.782341
9	0.05	1	0.496435
10	0.01	1	0.0246184

Fitted parameters for Summer08 dijets — fitted spectrum

Index	Scale	Start value	Fitted value
c	1	0.7	0.590911
μ	1	0.98	0.98
σ	0.1	0.5	0.4589
0	0.01	1	0.943574
1	0.1	1	0.896644
2	0.1	1	1.06795
3	0.5	1	0.934932
4	0.5	1	1.02109
5	1	1	1
6	2	1	1.02793
7	1	1	1.01331
8	0.1	1	0.78706
9	0.05	1	0.139978
10	0.01	1	0.951093