Progress on QCD background estimation methods for the hadronic SUSY search

CMS SUSY meeting

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

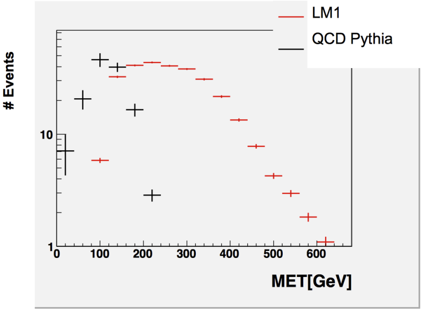
RA2 working group

October 6th, 2009

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

Motivation: Understanding the QCD background to $\not{E}_T$

- RA2 signature: ≥ 3 jets $+ \not{E}_T$
- Large $\not{E}_T$-background expected from QCD events
  - 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

1. Data driven QCD background determination methods
   - ABCD method
   - Jet smearing method
2. The response fit method
   - Technique
   - Parameterisation of the response function
   - Parameterisation of the p_T^{true} spectrum
   - Proof of principle for one p_T^{dijet} bin

Outline

1. Data driven QCD background determination methods
   - ABCD method
   - Jet smearing method
2. The response fit method
   - Technique
   - Parameterisation of the response function
   - Parameterisation of the p_T^{true} spectrum
   - Proof of principle for one p_T^{dijet} bin

Concept of the ABCD method (T. Schum [1])

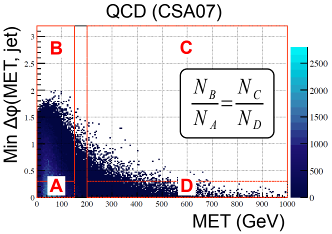
- Two uncorrelated variables
- Cuts: regions dominated by
  - background (**A**, **B**, **D**)
  - signal (**C**)
- Prediction of **number** of QCD events in signal region

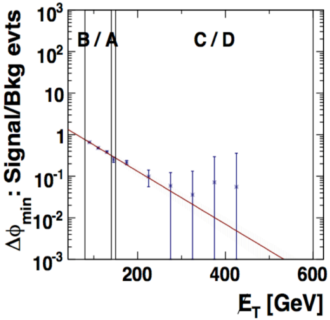
$$N_C = N_D \frac{N_B}{N_A}$$

- Extended ABCD method

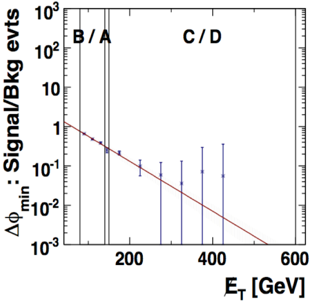
$$N_C = \int_{\not{E}_T} \mathrm{d}N_D \frac{\mathrm{d}N_B}{\mathrm{d}N_A}$$

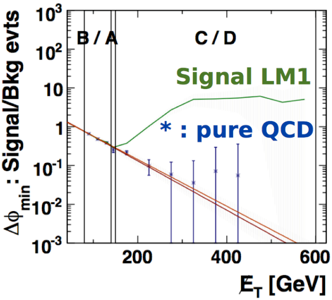
- E.g. $\min\Delta\phi(\not{E}_T, \text{jet})$ vs $\not{E}_T$: exponential fit on $\frac{\mathrm{d}N_B}{\mathrm{d}N_A}$





Background estimation from ABCD method (T. Schum)

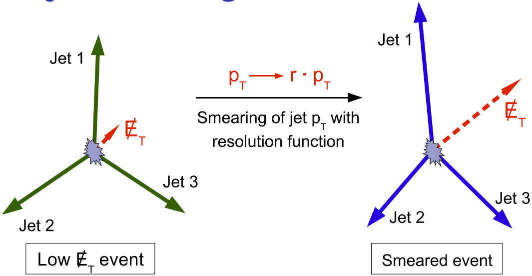




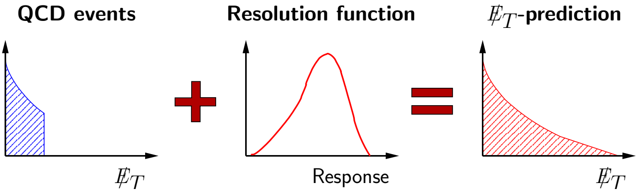
	#signal	#true QCD	#estimated QCD	stat. error
QCD	0	91	92	±21
QCD+LM1	318	91	102	±23

- Good estimation of the number of QCD events
- Fit relatively robust (≈ 10%) w.r.t. signal contamination
- Studies done for other variables with different correlations

Concept of the jet smearing method

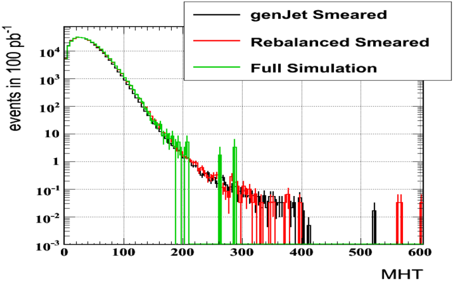


1. Selection of well measured QCD events
2. Smearing with response function



Application of the response function (M. Rydenfelt [2])

- Smearing of **measured** p_T with response function
- Bias in $\not{E}_T$ prediction
  - Double smearing
  - Selection cut on $\not{E}_T$



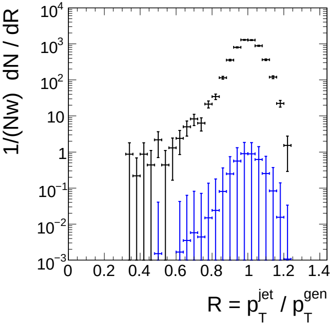
Rebalancing of QCD events

$$p_{T,1} \rightarrow p'_{T,1} = -\sum_{i=2} p_{T,i}$$

- Smearing of artifical, perfectly balanced QCD events

Good reproduction of the $\not{E}_T$ spectrum

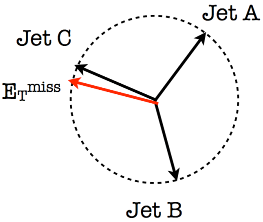
Determination of the response function

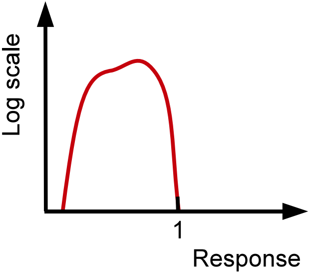


- Summer08 QCD dijets
- $470 < \hat{p}_T < 800$ GeV
- Weighted for $100\,\text{pb}^{-1}$
- Dijet selection applied (see slide 17)

- Direct measurement of response from γ-jet events
  - p_T^γ well measured
  - Low statistics: $\sigma(\gamma\text{-jet}) = \mathcal{O}(2 \cdot 10^{-3})\sigma(\text{dijet})$
- Determination of response from N-jet events
  - MPF method (Iowa)
  - Response fit method (Hamburg)

$\not\!\!E_T$ projection fraction (MPF) method (Iowa [3])

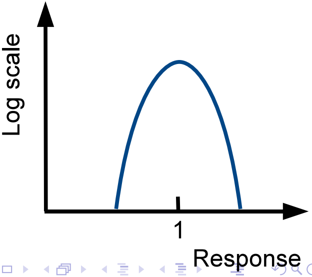




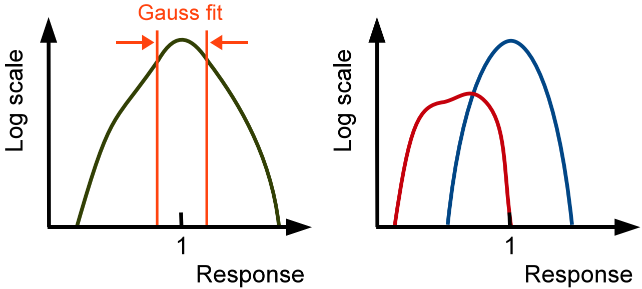
- 3-jet events with $\not\!\!E_T$ parallel to jet C

$$R = 1 - \frac{\vec{\not\!\!E}_T \cdot \vec{p}^{\,true}_{T,c}}{|\vec{p}^{\,true}_{T,c}|^2}, \quad \vec{p}^{\,true}_{T,c} \approx -(\vec{p}^{\,true}_{T,a} + \vec{p}^{\,true}_{T,b})$$

- Selection bias → tail of response
- Gaussian bulk of response from γ-jet events



Combination of bulk and tail from dijet events (Iowa)



- $\not{E}_{\mathrm{T}}$ projection fraction for both jets in dijet events

$$R_a = 1 + \frac{\vec{\not{E}}_{\mathrm{T}} \cdot \vec{p}_{\mathrm{T},b}}{|\vec{p}_{\mathrm{T},b}|^2}$$

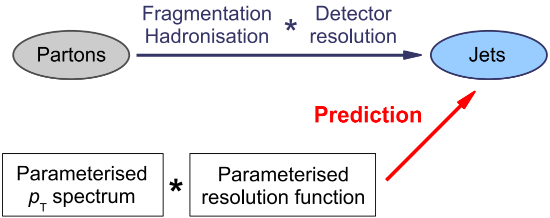
- Peak fitted with Gaussian
- Normalisation of tail to bulk contribution in response from ratio of integral of fit function to number of events in left tail

Outline

1. Data driven QCD background determination methods
   - ABCD method
   - Jet smearing method

2. The response fit method
   - Technique
   - Parameterisation of the response function
   - Parameterisation of the p_T^{true} spectrum
   - Proof of principle for one p_T^{dijet} bin

Concept of the response fit method



- In each event i, probability for a given dijet configuration p_T^1, p_T^2 is

$$\mathcal{P}_i^{1,2} \propto \int_0^\infty \mathrm{d}p_T^{\mathrm{true}}\, f_{\mathbf{b}}(p_T^{\mathrm{true}}) \cdot r_{\mathbf{b}}(p_T^1/p_T^{\mathrm{true}}) \cdot r_{\mathbf{b}}(p_T^2/p_T^{\mathrm{true}})$$

- $f_{\mathbf{b}}$ is the probability density function (pdf) of p_T^{true}
- $r_{\mathbf{b}}$ is the response pdf

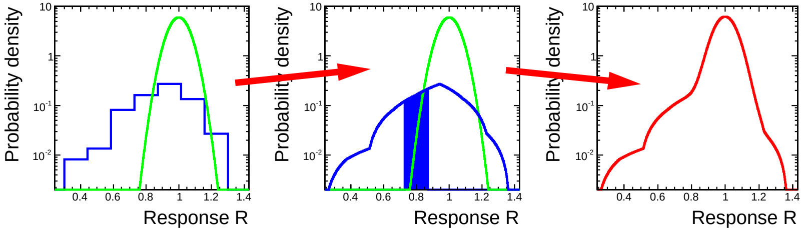
- Likelihood $\tilde{\mathcal{L}}(\mathbf{b}) = \prod_{i=0}^{N_{\mathrm{evt}}} \mathcal{P}_i^{1,2}$ maximal for best parameter values $\mathbf{b}$
- Inclusion of other data types by appropriate definition of probability

Parameterisation of the response function

Superposition of Gaussian and interpolated step function

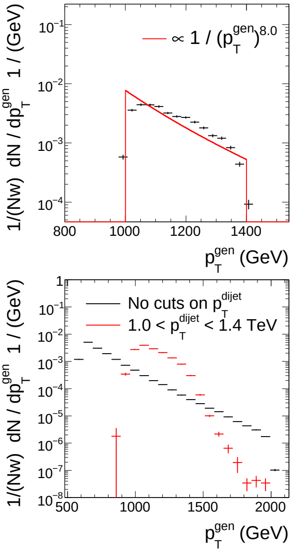
$$r_{\mathbf{b}}(R) = c \cdot G(R; \mathbf{b}) + (1 - c) \cdot S_N(R; \mathbf{b}) \qquad R = p_T / p_T^{\text{true}}$$

- Central Gaussian G: $\frac{\sigma(R)}{R} = \frac{a_1}{p_T^{\text{true}}} \oplus \frac{a_2}{\sqrt{p_T^{\text{true}}}} \oplus a_3$
- Step function S with N parameters
- Actual $S_N(R)$ is linear interpolation of adjacent bin contents



- Separate fits in different $p_T^{\text{dijet}} = \frac{1}{2}(p_T^1 + p_T^2)$ bins → interpolation

Parameterisation of the p_T^{true} spectrum



- p_T^{true} spectrum parameterised with powerlaw

$$f(p_T^{true}) \propto \frac{1}{(p_T^{true})^n}$$

- Fair description of p_T^{gen} spectrum
- **Note**: $p_T^{gen} \neq p_T^{true}$, but used for validation purposes

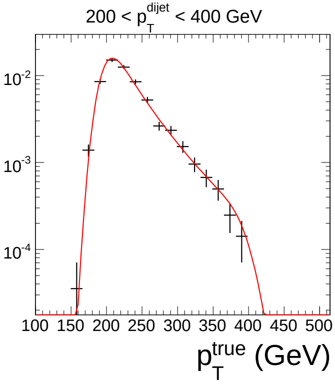
Binning in p_T^{dijet}

Migration due to resolution and steeply falling spectrum

Description of p_T^{true} spectrum including cuts on p_T^{dijet}

Inclusion of p_T^{dijet} cuts into pdf of p_T^{true}

$$f(p_T^{true}) \propto f_0(p_T^{true}) \int_{p_{T,min}^{dijet}}^{p_{T,max}^{dijet}} dx \, \frac{r(x/p_T^{true})p_T^{true}}{\sqrt{2}}$$



- Bin: $p_{T,min}^{dijet} < p_T^{dijet} < p_{T,max}^{dijet}$
- Response pdf r
- Pure spectrum $f_0 \propto 1/(p_T^{true})^n$
- Demonstration with `ToyMC` sample using Gaussian response

Description of p_T^{true} spectrum

Event selection

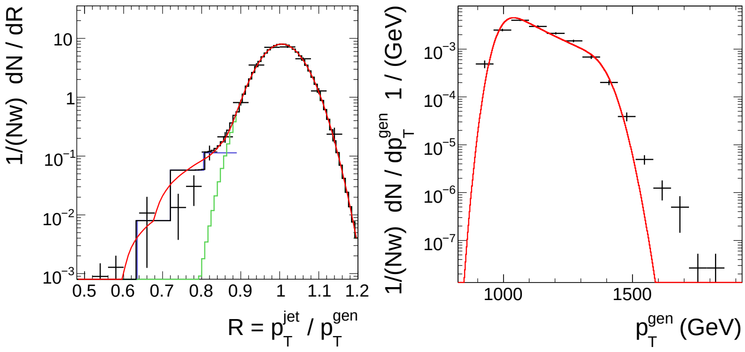
- L2L3 corrected Summer08 QCDDijets
- $600 < \hat{p}_T < 2200$ GeV
- Dijet selection similar to `CMS AN-2008/031`[1]

Dijet selection

- 2 jets leading in p_T with $|\eta| < 0.8$
- $p_T^3/p_T^{dijet} < 0.1$ or $p_T^3 < 2$ GeV
- $|\Delta\phi^{1,2}| > 2.7$
- $0.07 < p_T^{had}/p_T^{1,2} < 0.95$
- $\Delta R(\mathtt{jet}, \mathtt{genjet}) < 0.1$ (important for validation)

[1] Determination of the Relative Jet Energy Scale at CMS from Dijet Balance

Results for one bin $1000 < p_T^{dijet} < 1400$ GeV



- Step function with 4 bins from $0.5 < R < 0.9$
- Fixed Gaussian response from fit on L2L3 corrected jets assumed in p_T^{true} spectrum ($\rightarrow$ deviations at large p_T^{gen})
- Reasonable description of response

Summary & outlook

Summary

- Redundant methods for a data driven estimation of the $\not{E}_T$ contribution from QCD events are being developed
  - ABCD method
  - Jet smearing method
- Data driven approaches for the determination of the jet response function have been presented
  - MPF method
  - Response fit method
- The method determines the response and spectrum by maximising the probability of the measured dijet configuration

Outlook

- Response function in different p_T^{dijet} and η bins
- Evaluation of systematic uncertainties of the method
- Combination of background estimations from different methods

Backup

References

- ABCD method
  - Talk by T. Schum:
    http://indico.cern.ch/conferenceDisplay.py?confId=68147
  - Talk by C. Auterman:
    https://indico.desy.de/conferenceOtherViews.py?view=standard&confId=2175
- Jet smearing by rebalancing
  - Talk by M. Rydenfeldt
    http://indico.cern.ch/conferenceDisplay.py?confId=61021
- MPF method
  - Talk by T. Yetkin
    http://indico.cern.ch/conferenceDisplay.py?confId=68147
  - Talk by E. Albayrak
    http://indico.cern.ch/conferenceOtherViews.py?view=cms&confId=65582

Backup on ABCD method (T. Schum)

- Studied variables (correlation)
  - min$\Delta\phi(\not{E}_{\mathrm{T}}, \mathrm{jet})$ vs $\not{E}_{\mathrm{T}}$ (exponential)
  - min$\Delta\phi(\not{E}_{\mathrm{T}}, \mathrm{MPT})$ vs $\not{E}_{\mathrm{T}}$ (constant)
- SUSYPATV5 samples
  - QCD madgraph
  - SUSY LM1
- Event selection from RA2 default baseline cut scenario
  - $p_{\mathrm{T}}^1 > 180$ GeV, $p_{\mathrm{T}}^2 > 110$ GeV, $p_{\mathrm{T}}^3 > 50$ GeV
  - Direct lepton veto
  - $\not{E}_{\mathrm{T}} > 150$ GeV
  - min$\Delta\phi > 0.3$

Backup on MPF method (Iowa)