Update on QCD background estimation

UHH CMS SUSY Meeting

Matthias Schröder

November 19th, 2009

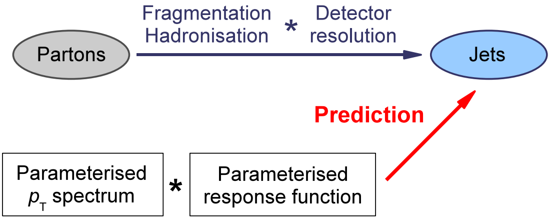
GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Outline

1. Reminder
   - Technique of the response fit method
   - p_{T} dependent response
2. Technical details of the minimisation
3. Results for one $p_{\mathrm{T}}^{\mathrm{dijet}}$ bin
4. Results for p_{T} range
5. Conclusions

Concept of the response fit method



- In each event i, probability density of the 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

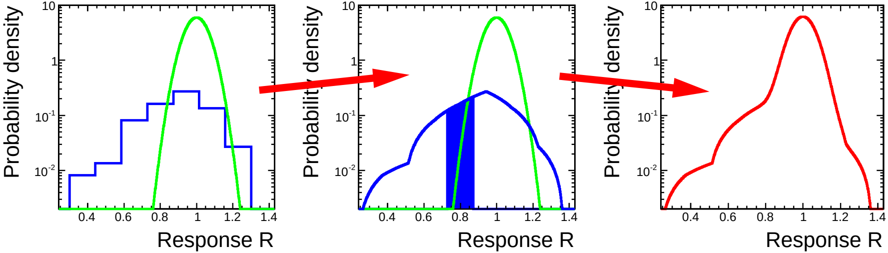
- Likelihood $\tilde{\mathcal{L}}(\mathbf{b}) = \prod_{i=0}^{N_{\mathrm{evt}}} \mathcal{P}_i^{1,2}$ maximal for correct parameter values $\mathbf{b}$

Parameterisation of the response function

Superposition of Gaussian and interpolated step function

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

- Central Gaussian G around 1 (assume calibrated jets)
- Step function S with N parameters
- Actual $S_N(R)$ is linear interpolation of adjacent bin contents



p_T dependent response parameterisation

- Width of central Gaussian $\frac{\sigma(p_T/p_T^{true})}{\langle p_T/p_T^{true}\rangle} = \frac{a_1}{p_T^{true}} \oplus \frac{a_2}{\sqrt{p_T^{true}}} \oplus a_3$
- No p_T^{true} dependence of step function
- Separate fits in different p_T^{gen} bins
- Interpolation of parameters
- **Data driven**: binning in p_T^{dijet}
- Migration due to resolution and steeply falling spectrum

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

$$f_{\mathbf{b}}(p_T^{true}) \propto \hat{f}_{\mathbf{b}}(p_T^{true}) \int_{p_{T,min}^{dijet}}^{p_{T,max}^{dijet}} \mathrm{d}x \, \frac{r_{\mathbf{b}}(x/p_T^{true})p_T^{true}}{\sqrt{2}}$$

- Bin: $p_{T,min}^{dijet} < p_T^{dijet} < p_{T,max}^{dijet}$
- Response pdf $r_{\mathbf{b}}$, pure spectrum $\hat{f}_{\mathbf{b}} \propto 1/(p_T^{true})^n$

Outline

Technical details of the minimisation

- Calculation of likelihood computationally intensive due to integrations

$$
\begin{aligned}
\mathcal{P}_i^{1,2} &= \frac{1}{\mathcal{N}} \int_0^\infty \mathrm{d}p_\mathrm{T}^\mathrm{true} \int_{p_\mathrm{T,min}^\mathrm{dijet}}^{p_\mathrm{T,max}^\mathrm{dijet}} \mathrm{d}x\, \frac{r_\mathbf{b}(x/p_\mathrm{T}^\mathrm{true})p_\mathrm{T}^\mathrm{true}}{\sqrt{2}} \cdot (p_\mathrm{T}^\mathrm{true})^{-n} \cdot r_\mathbf{b}(p_\mathrm{T}^1/p_\mathrm{T}^\mathrm{true}) \cdot r_\mathbf{b}(p_\mathrm{T}^2/p_\mathrm{T}^\mathrm{true}) \\
\mathcal{N} &= \int_0^\infty \mathrm{d}p_\mathrm{T}^\mathrm{true} \underbrace{\int_{p_\mathrm{T,min}^\mathrm{dijet}}^{p_\mathrm{T,max}^\mathrm{dijet}} \mathrm{d}x\, \frac{r_\mathbf{b}(x/p_\mathrm{T}^\mathrm{true})p_\mathrm{T}^\mathrm{true}}{\sqrt{2}}}_{\mathcal{C}} \cdot (p_\mathrm{T}^\mathrm{true})^{2-n}
\end{aligned}
$$

- Integral $\mathcal{C}$ describes effects on $p_\mathrm{T}^\mathrm{true}$ by cuts on $p_\mathrm{T}^\mathrm{dijet}$
- Calculated only once per iteration to save CPU time
  1. Storing $\mathcal{C}$ for different values of $p_\mathrm{T}^\mathrm{true}$ using current parameter values $\mathbf{b}$
  2. Minimising $\mathcal{L}$ by varying $\mathbf{b}$ and n
  3. Starting again at 1
- Implementation of faster integration using Simpson's 3/8 rule

Outline

1. Reminder
   - Technique of the response fit method
   - p_T dependent response
2. Technical details of the minimisation
3. Results for one p_T^{dijet} bin
4. Results for p_T range
5. Conclusions

Event selection

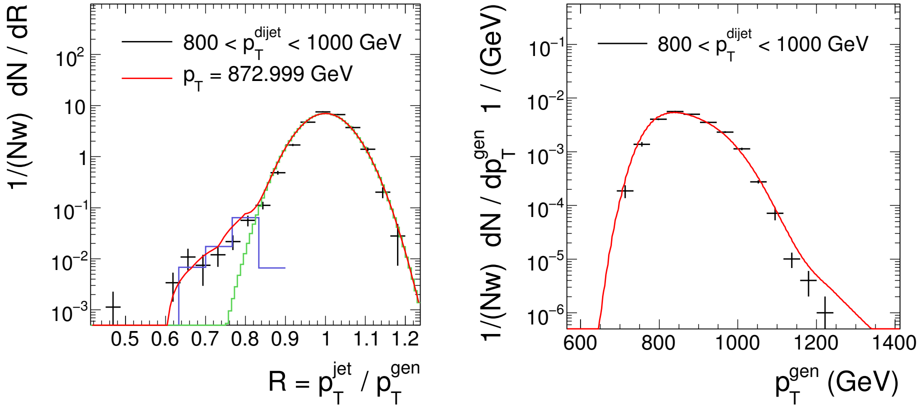
- L2L3 corrected Summer08 QCDDijets
- $800 < p_T^{\text{dijet}} < 1000$ GeV
- Dijet selection similar to `CMS AN-2008/031`[1]

Dijet selection

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

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

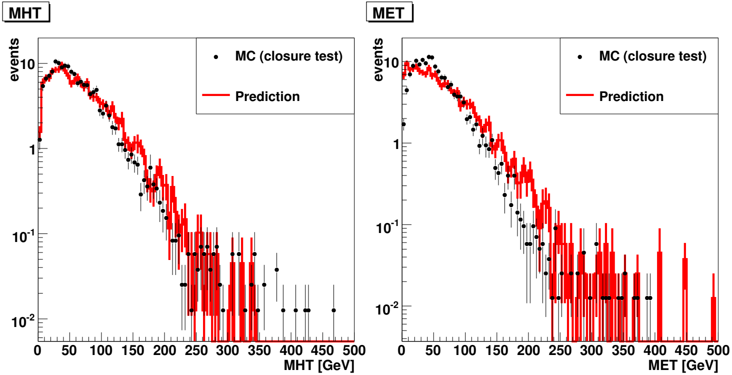
Fit results for one bin $800 < p_T^{\mathrm{dijet}} < 1000$ GeV



- Step function with 6 bins from $0.5 < R < 0.9$
- Fitted pdf shown for mean p_T of p_T^{gen} distribution
- Good description of response and truth spectrum

Consistency check: predicted MHT and $\not{E}_T$ spectra

- Selection of same events as used to fit response function

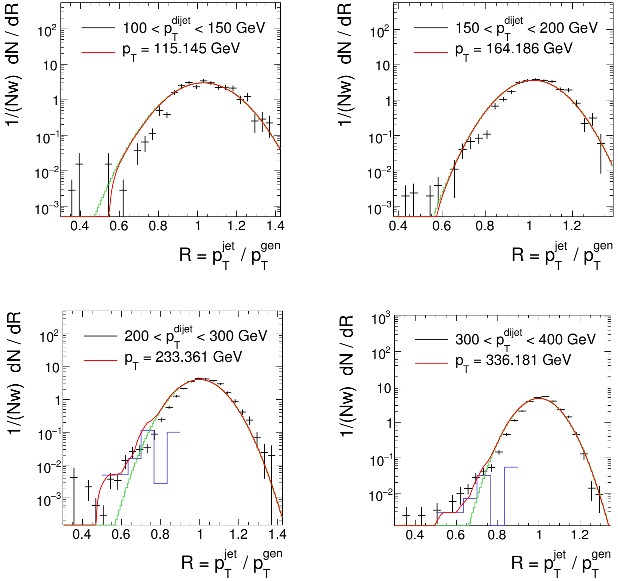


- Data points (MC) from L2L3 corrected `sisCone5CaloJets`
- Prediction from smeared genjets

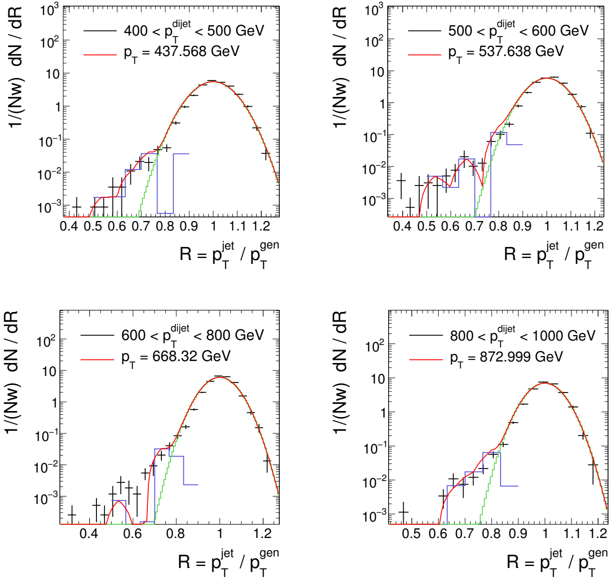
- Data points (MC) from `metNoHFHO`
- Prediction from smeared genjets

Outline

Fit results for p_T range $80 < \hat{p}_T < 1000$ GeV

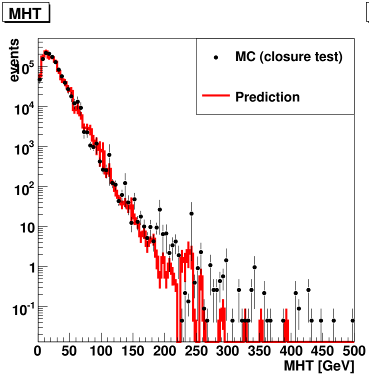


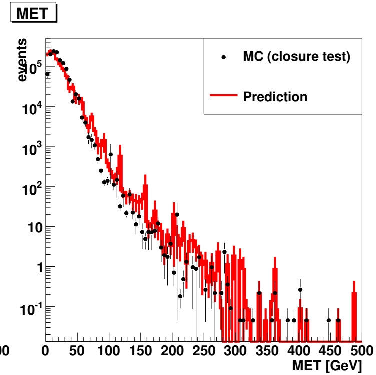
Fit results for p_T range $80 < \hat{p}_T < 1000$ GeV



Consistency check: $80 < \hat{p}_T < 1000$ GeV

- Dijet event selection applied, weighted for 100 pb^{-1}
- Smearing with response pdf corresponding to p_T of smeared genjet





- Data points (MC) from L2L3 corrected `sisCone5CaloJets`
- Prediction from smeared genjets

- Data points (MC) from `metNoHFHO`
- Prediction from smeared genjets

Outline

1. Reminder
   - Technique of the response fit method
   - p_{T} dependent response
2. Technical details of the minimisation
3. Results for one $p_{\mathrm{T}}^{\mathrm{dijet}}$ bin
4. Results for p_{T} range
5. Conclusions

Conclusions

- Underestimation of MHT spectrum
- Problem with fit of response function in low p_T bins: one-sided step function reaches far into Gaussian part
    - Usage of step function covering whole R range
    - Adjustment of step function range by width of Gaussian (faster!)
- Systematic uncertainties of method
    - Parameterisation of spectrum
    - Binning of step function
    - Binning of response functions in p_T^{dijet}
    - QCD selection cuts