Estimation of QCD Multijet Background for Top Antitop Events from Data

Lukas Kreczko, Sebastian Naumann, Peter Schleper, Georg Steinbrück , Roger Wolf

Hamburg CMS Meeting, 25.02.2009

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

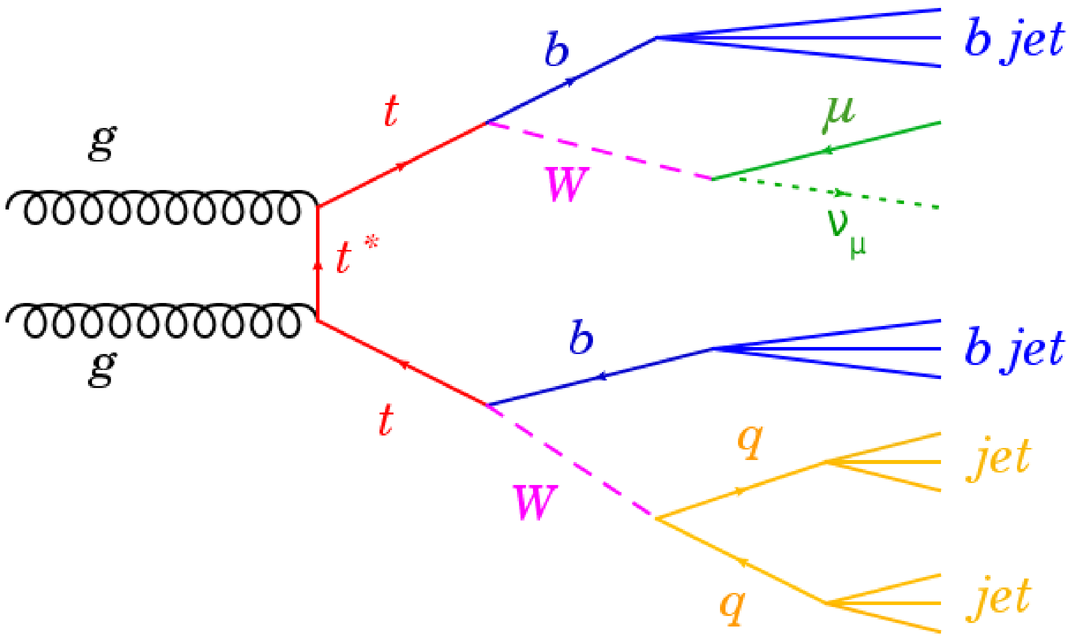
Universität Hamburg

Overview

- Samples and Signal
- Method for background estimation from data
- Robustness of the method
- Summary and Outlook

Top-Quark

Top-Signal: semileptonic with muon in final state



Important backgrounds:
- QCD Multijet Events (this analysis)
- W + n Jets
- tt̄ full hadronic
- tt̄ dileptonic muon

Used Samples

- incl $t\bar{t}$ + n jets (CSA07)
- W + n jets (CSA07)
- QCD μ enriched: μ are selected on generator level with minimum pT of 15GeV. No decay in flight.

Selection to ensure high QCD statistics:

- four jets with $E_T > 20$ GeV and $|\eta| < 2.4$
- one muon with $p_T > 20$ GeV and $|\eta| < 2.1$

Matrix Method

$$N_{loose} = N_{QCD} + N_{sig}$$

Number of events before a cut

$$N_{tight} = \epsilon_{QCD} \cdot N_{QCD} + \epsilon_{sig} \cdot N_{sig}$$

Number of events after a cut

ε_{SIG} and ε_{QCD} :the efficiencies to pass the cut

If the efficiencies are known, the number of signal and QCD events can be calculated:

$$N_{sig} = \frac{N_{tight} - \epsilon_{QCD} \cdot N_{loose}}{\epsilon_{sig} - \epsilon_{QCD}}$$

$$N_{QCD} = \frac{\epsilon_{sig} \cdot N_{loose} - N_{tight}}{\epsilon_{sig} - \epsilon_{QCD}}$$

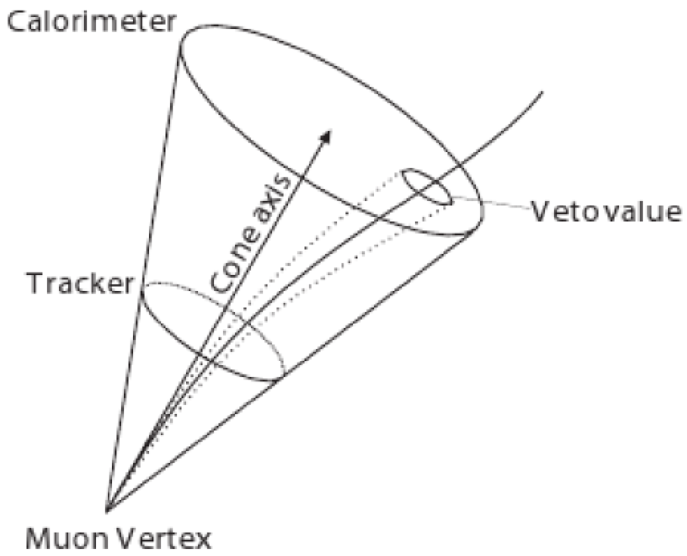
Cut-Variable: Muon Isolation

Muon isolation describes the activity in a cone of 0.3 in ΔR around a muon

Tracker Isolation: Sum of all pT(tracks) in the cone < **3 GeV**.

Calorimeter Isolation: Sum of all ET in the cone < **5 GeV**.

Minimal distance to closest Jet < 0.5 in ΔR



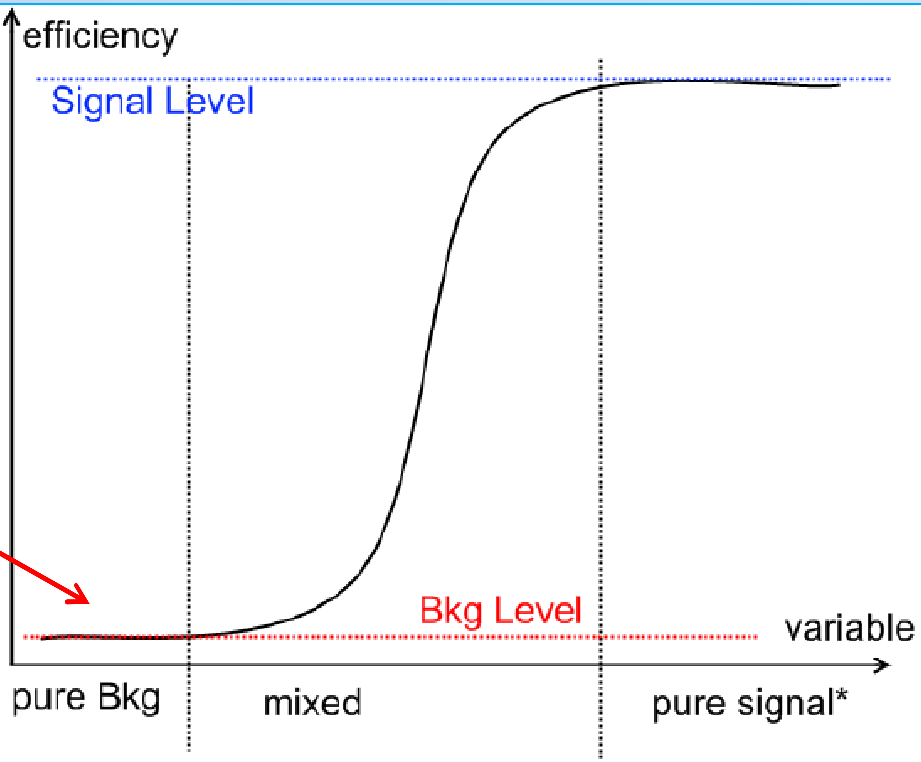
Top-signal as well as W+ jets contain isolated muons => both as signal

Estimate of efficiencies

ε_{SIG} can be gained from simulation, ε_{QCD} from data

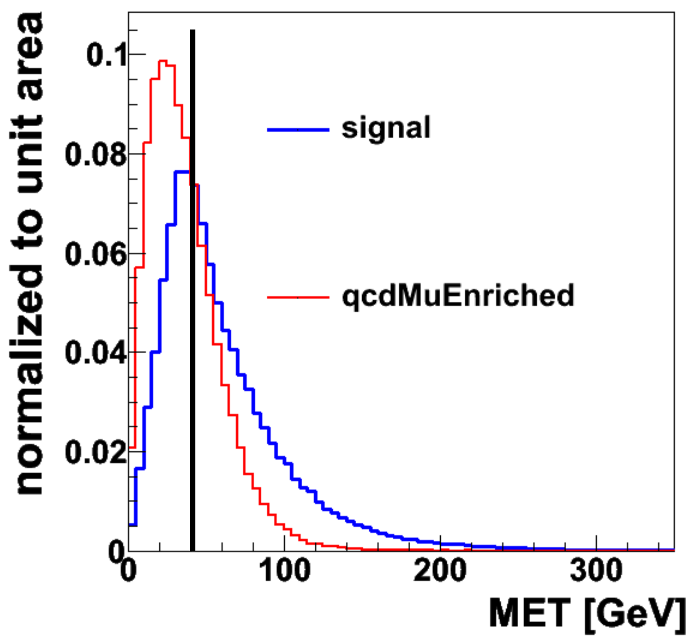
a control variable is neccessary destinguish between background and signal

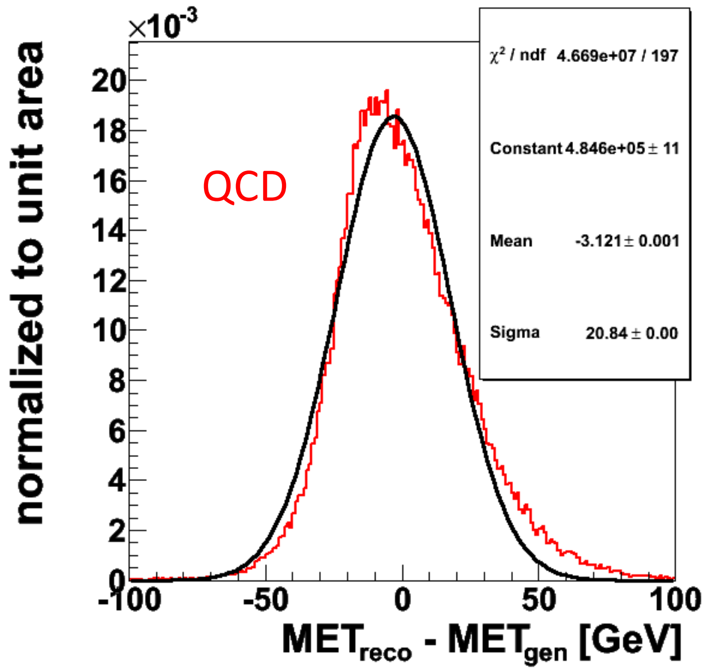
ε_{QCD} from here



For perfect variable ε_{SIG} from here

Missing E_T as control variable





Displayed errors are statistical

Estimation of events before isolation cut:

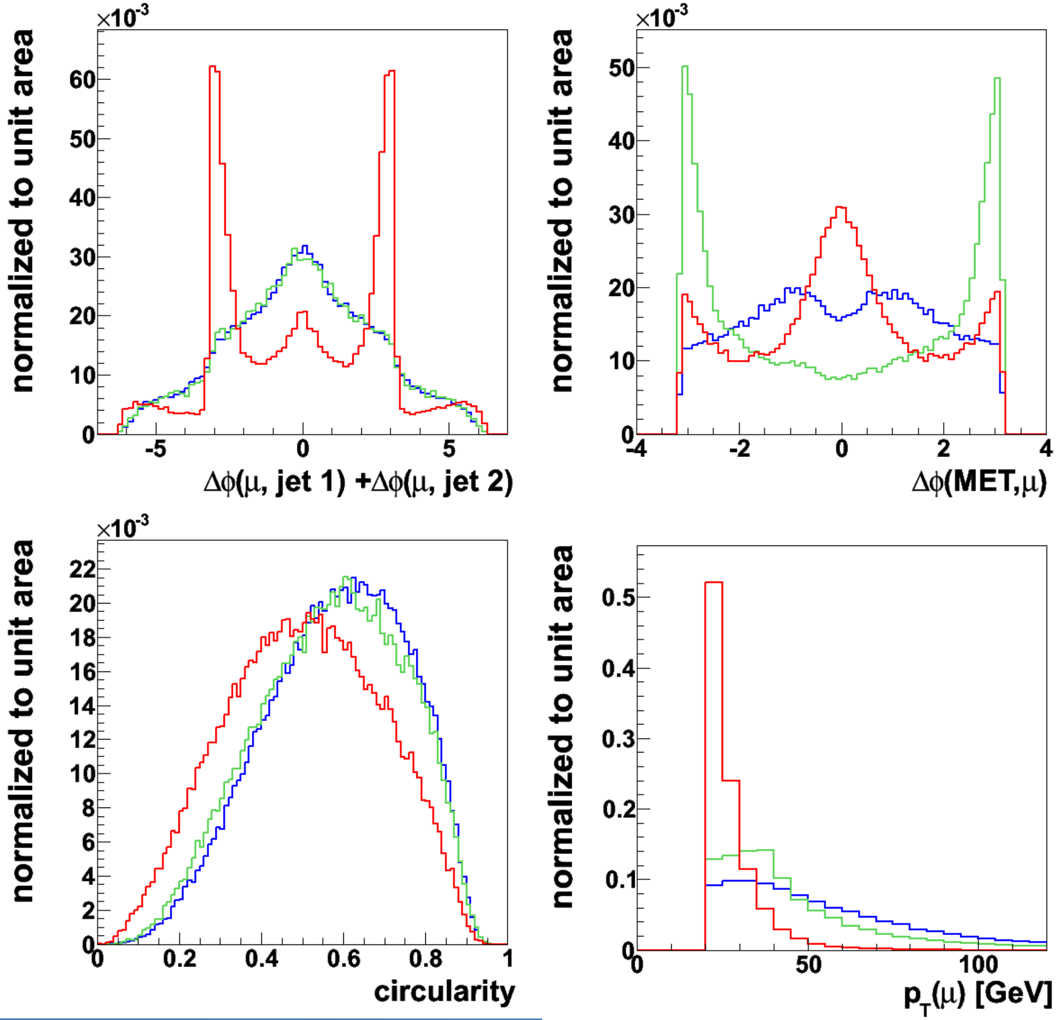
- QCD: 469860 ± 1300, true number: 465171
- Signal: 646 ± 394, true number : 5336

Estimation of events after isolation cut:

- QCD: 11734± 245, true number 8598
- Signal: 432 ± 268, true number 3568

40% difference from the true value for QCD and up to 90% for signal

New control variable



W+ n Jets

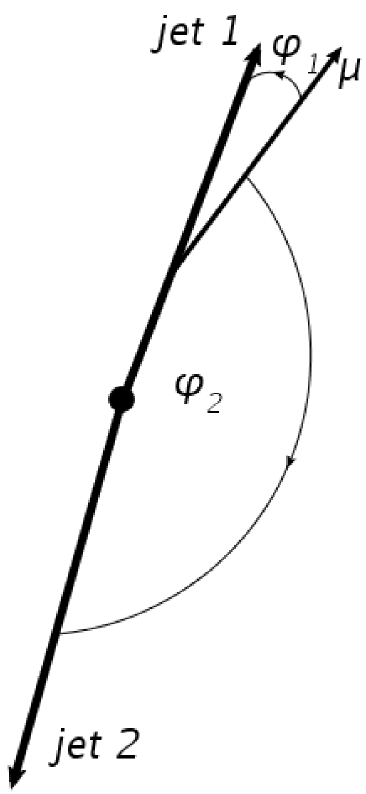
QCDMuEnriched

Top-Signal

- clear separation of QCD and signal
- stable isolation efficiency over the complete region
- as far independent from detector calibration as possible
- not correlated with the selection variable

$\Delta\Phi(\mu, jet1) + \Delta\Phi(\mu, jet2)$

QCD

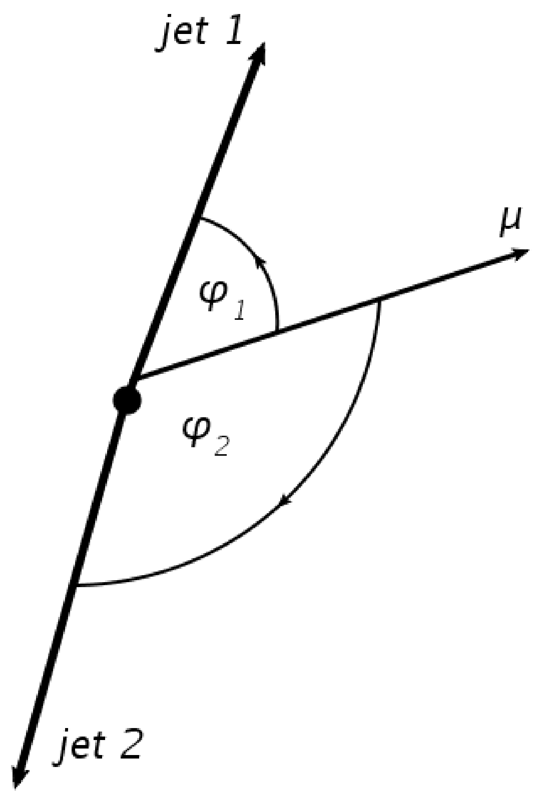


QCD: μs emerging from jets, Di-jet structure leads to values of π

signal: μs from real W-bosons, higher angle between μ and jets thus low values

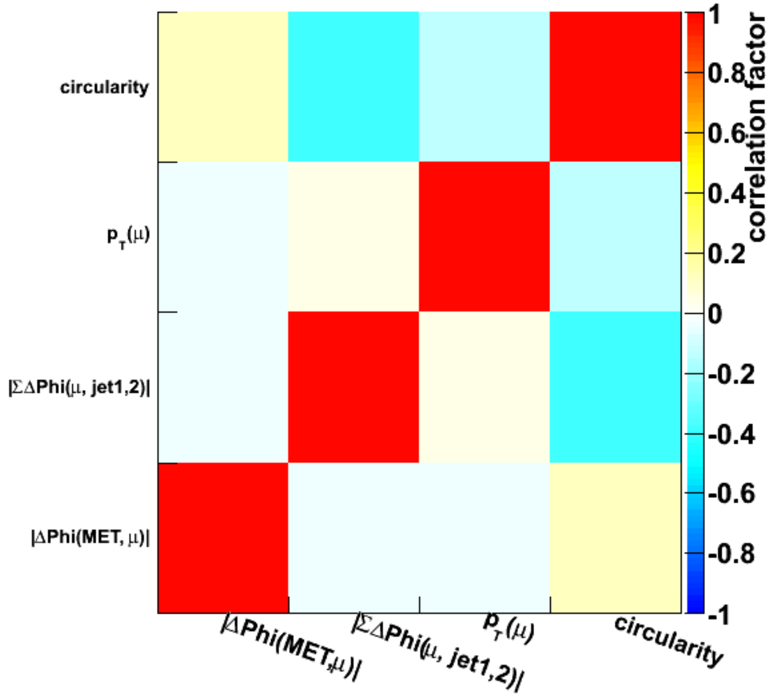
two leading jet on the same side
=> values above π

signal



Multivariate Ansatz

Correlation between variables

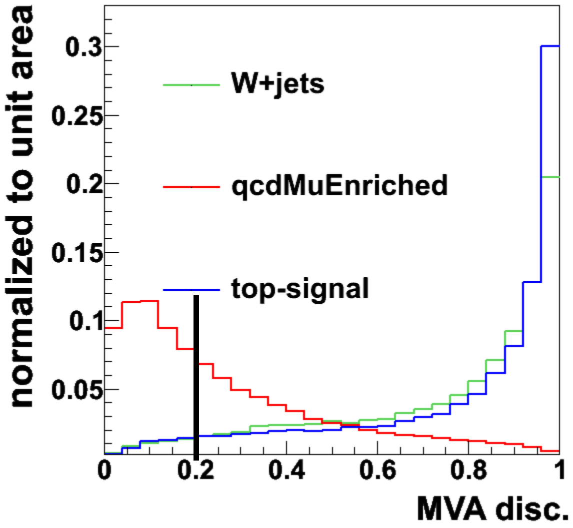


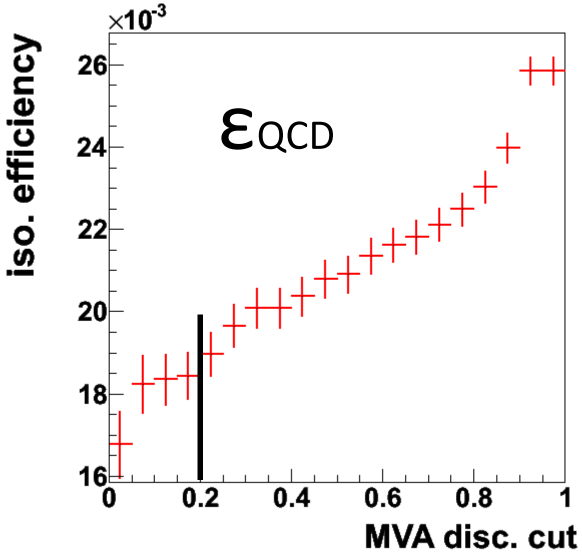
single variable matches all criteria
=> combination via a multivariate method

Only ΔΦ(μ, jet1) + ΔΦ(μ, jet2) and circulation have a bigger correlation of -0.4

Before creating the likelihood discriminator the variables are decorrelated

Multivariate Ansatz





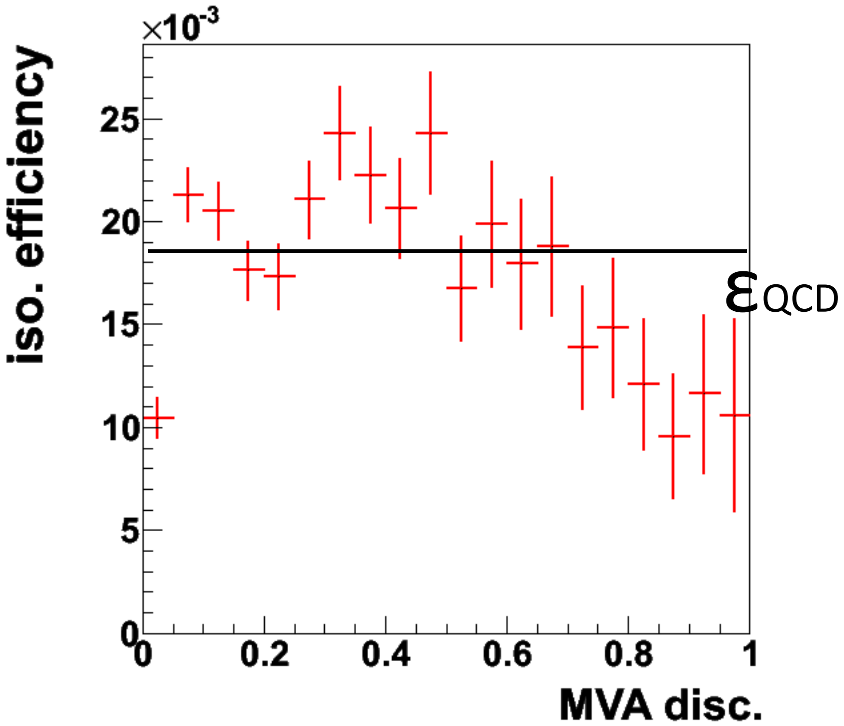
Control region with discriminator < 0.2 : contains 50% of total QCD events and has 5% signal contamination

Estimation of events before isolation cut :
- QCD: 465138 ± 1200, true number: 465171
- Signal: 5369 ± 436 , true number : 5336

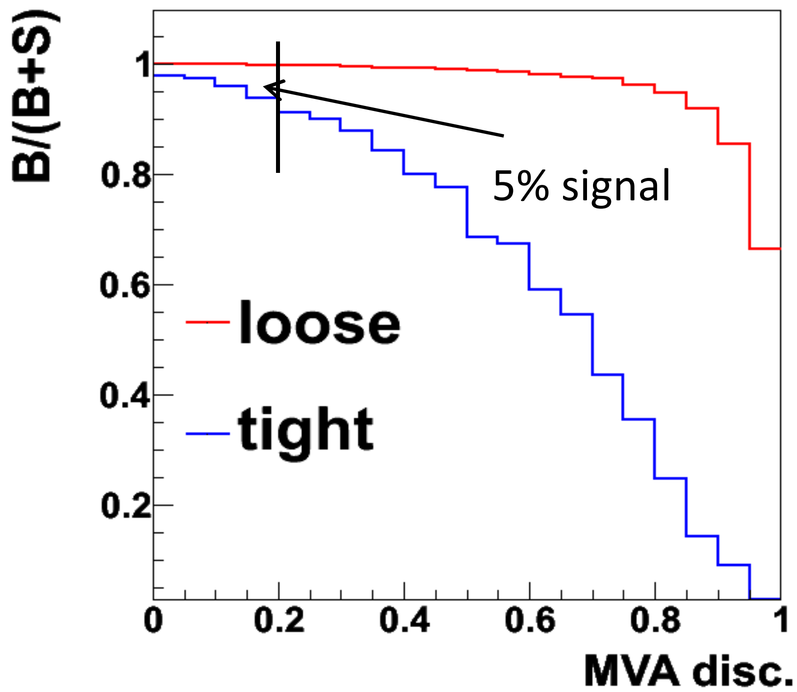
Estimation of events after isolation cut :
- QCD : 8575 ± 270, true number : 8598
- Signal: 3591 ± 292, true number : 3568

Two effects on ε_{QCD}

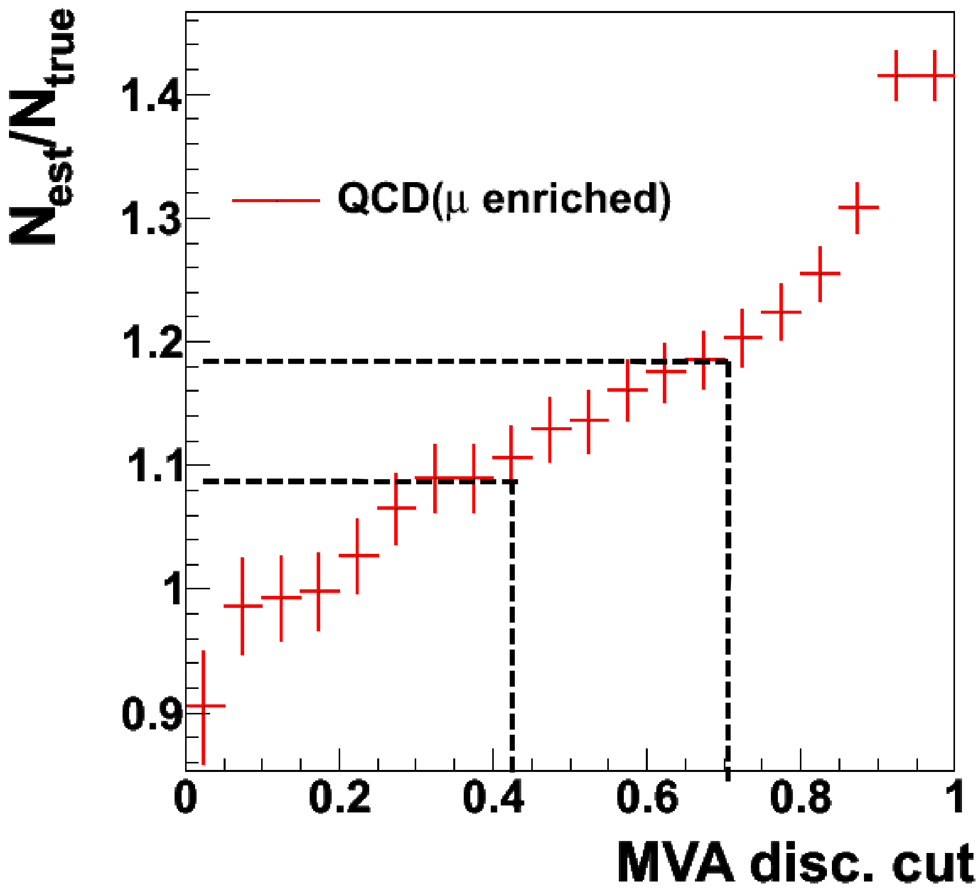
ε_{QCD} varies as a function of MVA discriminator



Signal contamination raises with MVA discriminator



Change of the control region



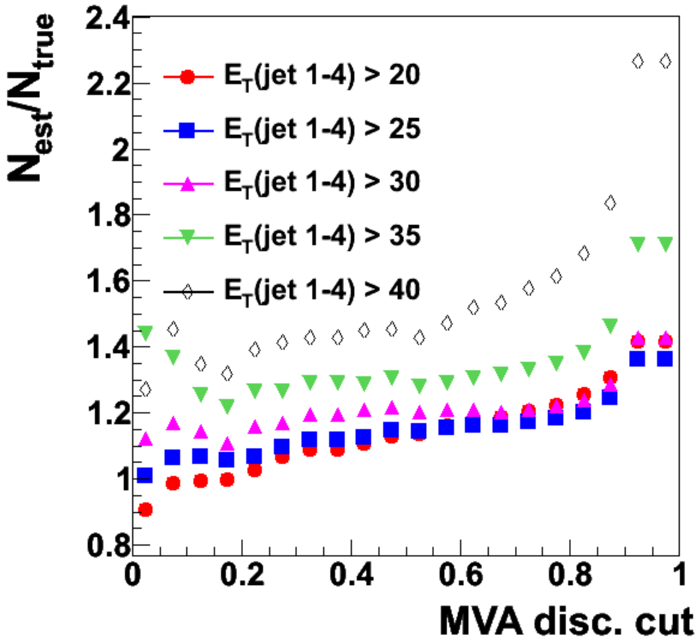
Depending on the error allowance on the QCD estimation after cut, the control region can be changed.

10% uncertainty allows cut < 0.4 with 20% signal contamination

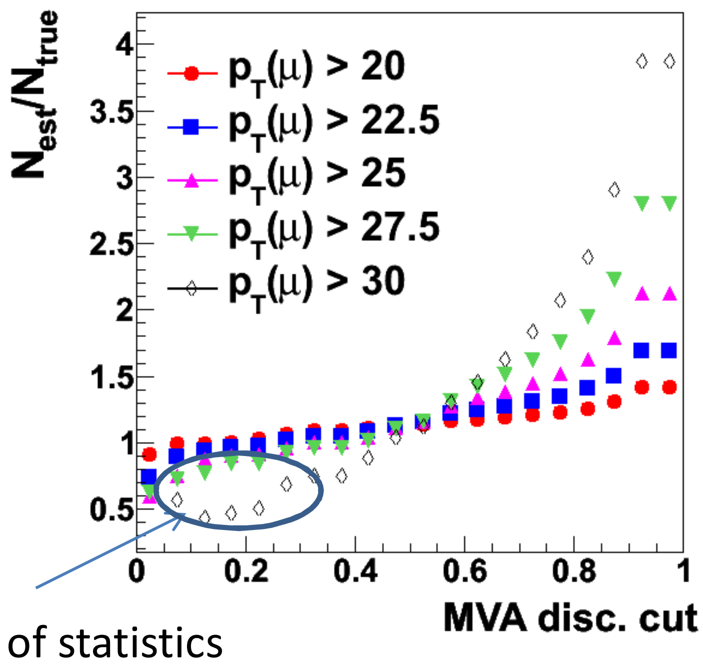
20% uncertainty allows cut < 0.7 with 55% signal contamination

Harder selection cuts

Cut on E_T of the four leading jets



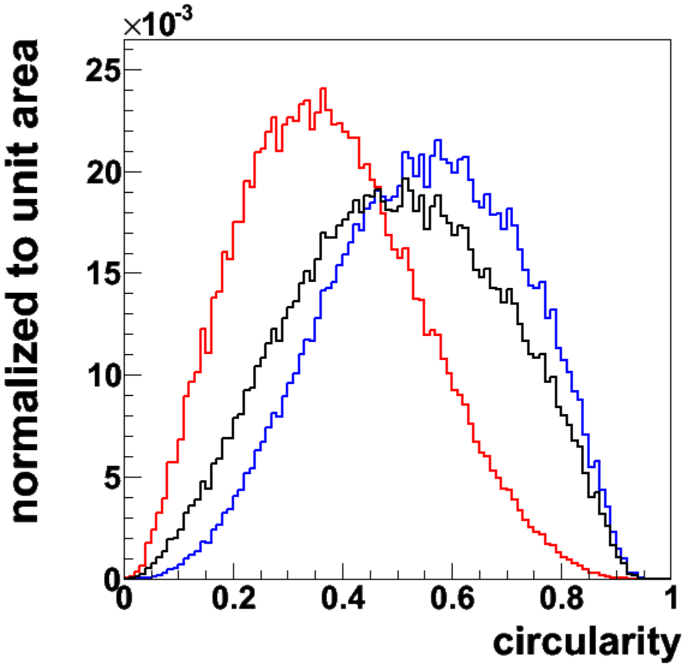
Cut on p_T of the leading muon



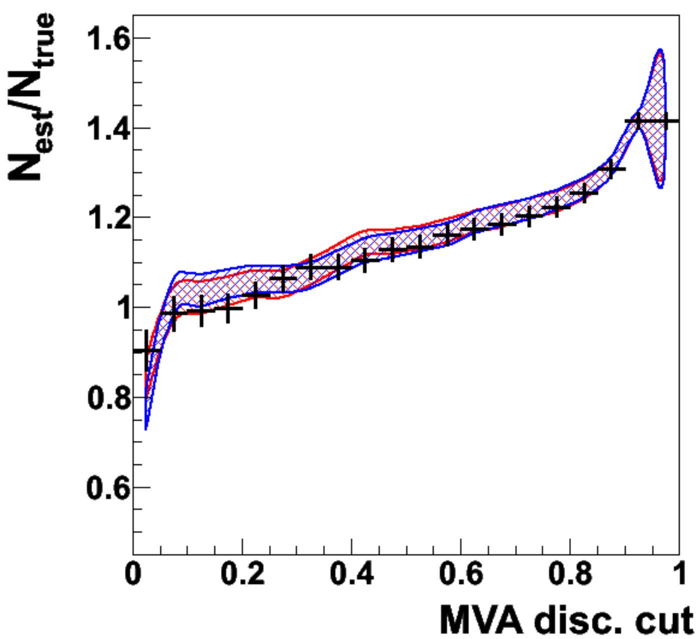
Harder cuts increase signal contamination => up to 40% error

Universität Hamburg

Change of Variable shape



Shape was changed by reweighting of QCD events towards background (red) and signal (blue)



Change of shapes leads to an overestimation of QCD events up to 10%

Universität Hamburg

Change of normalization

Estimation of events after isolation cut for **2 x N(QCD)**:

- QCD : 16764 ± 394, true number : 17196
- Signal: 4000 ± 419, true number : 3568

Estimation of events after isolation cut for **2 x N(W+jets)**:

- QCD : 8908 ± 265, true number : 8598
- Signal: 6306 ± 292, true number : 6616

Effects on changing Top-signal are small, since W+jets dominate the signal by factor of 6

Effect from the mismeasurement of ε_{QCD} are ≈3%, signal influence makes 5%

Universität Hamburg

Summary

- MET is not suitable at CMS for the Matrix Method, but MVA techniques can provide a good control variable
- Variables independent from detector calibration such as angle between objects and event shape variables
- Accepting estimation errors of 10-20% the Method is robust to normalization and shape errors

Universität Hamburg

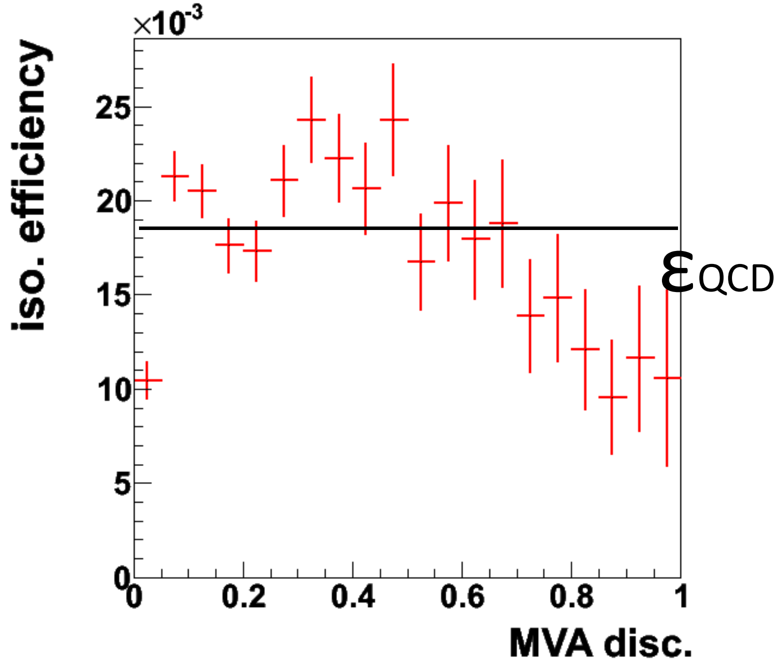
Outlook

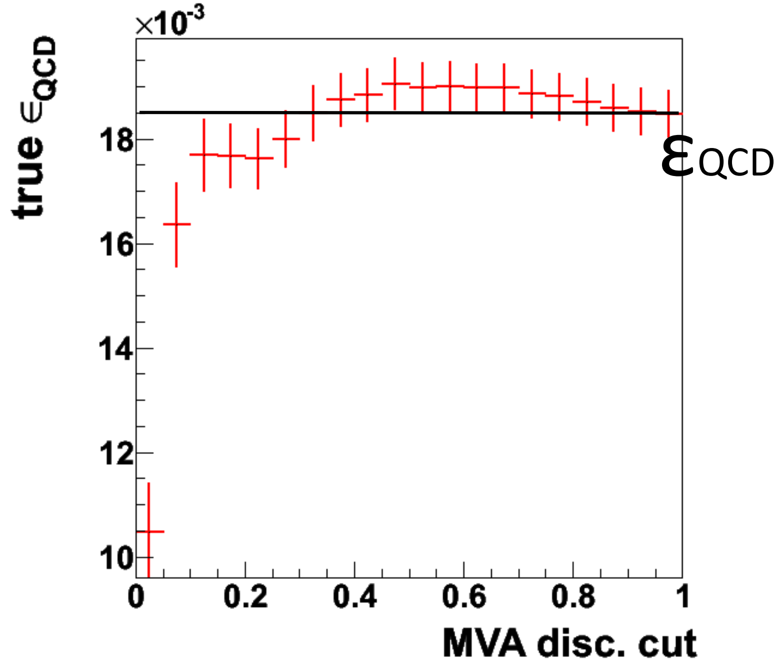
- Since $p_T(\mu)$ is used for cuts it should be replaced as MVA input.
- Verification of method robustness with other simulated samples
- If more statistic becomes available: study method with harder isolation criteria
- Study the dependency of input variables on ε_{QCD}

Universität Hamburg

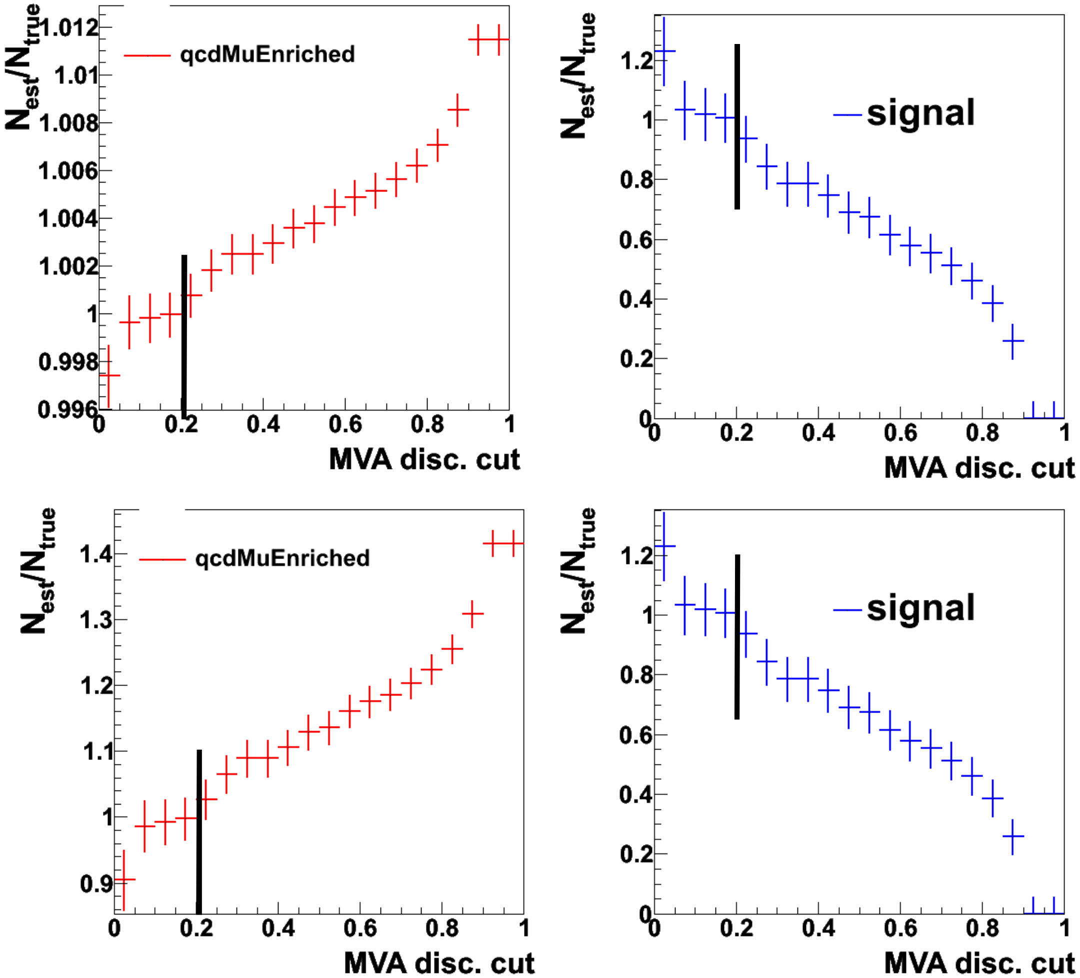
Backup Slides

Backup slides

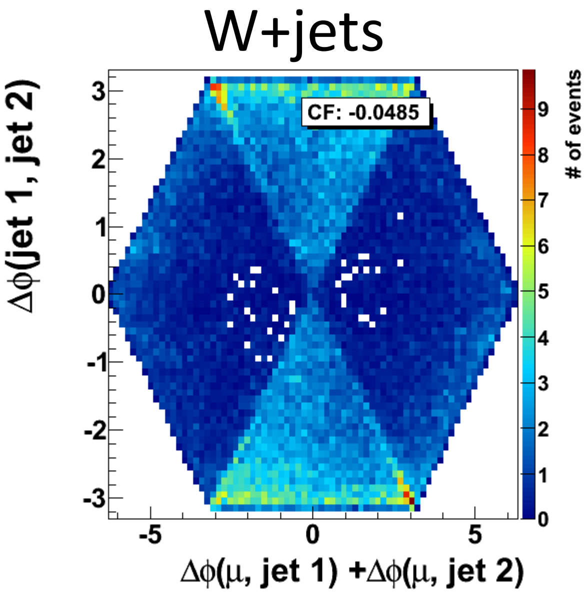


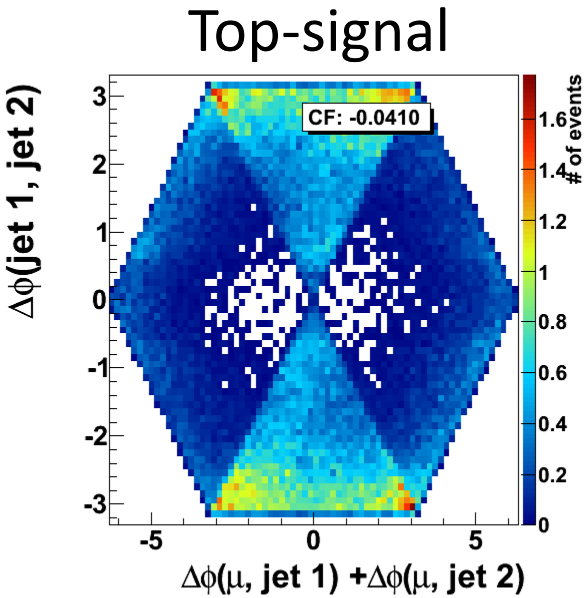


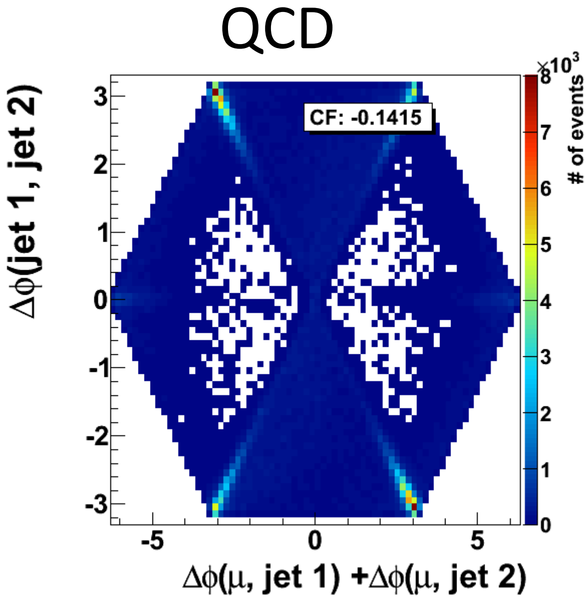
Estimation quality



Backup slides

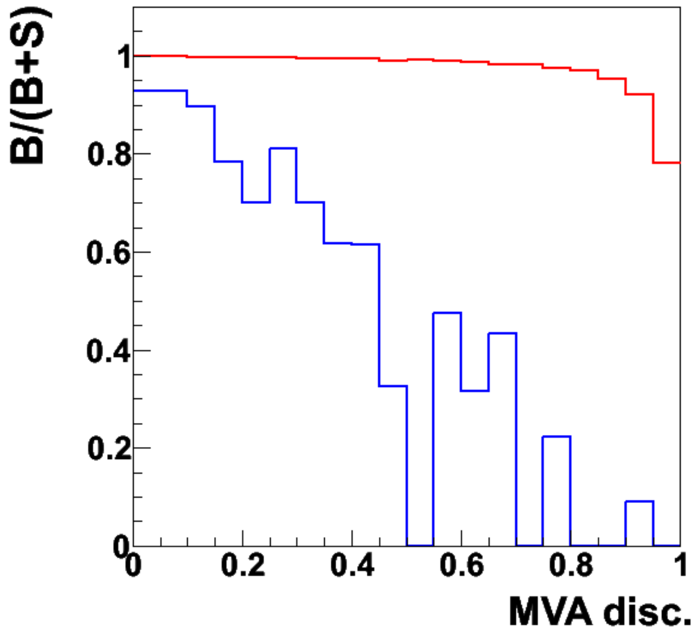






Backup slides

jet cuts



muon cuts