Unfolding at CMS

Matthias Weber

On behalf of the CMS collaboration

ETH Zurich

Outline

- Example of Matrix inversion:
  - Measurements of the charge ratio of atmospheric muons
- Example of Iterative bayesian unfolding (D'Agostini)
  - Charge particle multiplicities in pp interactions
- Example of Singular Value Decomposition
  - First measurement of hadronic event shapes in pp collisions
- Example of Tikhonov regularization
  - Measurement of the inclusive jet pt spectrum in pp collisions

Unfolding procedure:

Matrix inversion

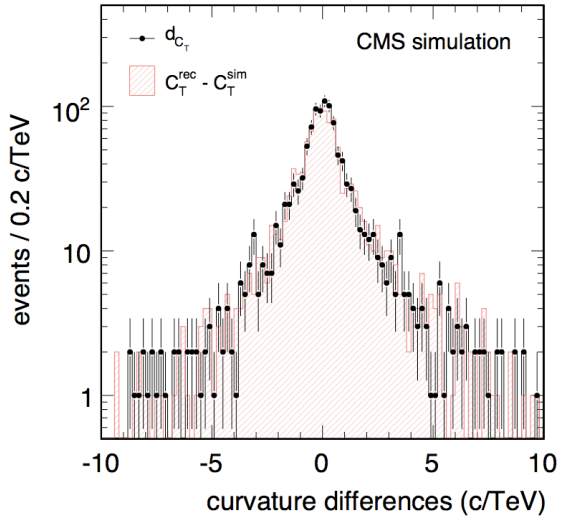
Measurements of the charge ratio of atmospheric muons *(Physics Letters **B 692 (2010)**)*

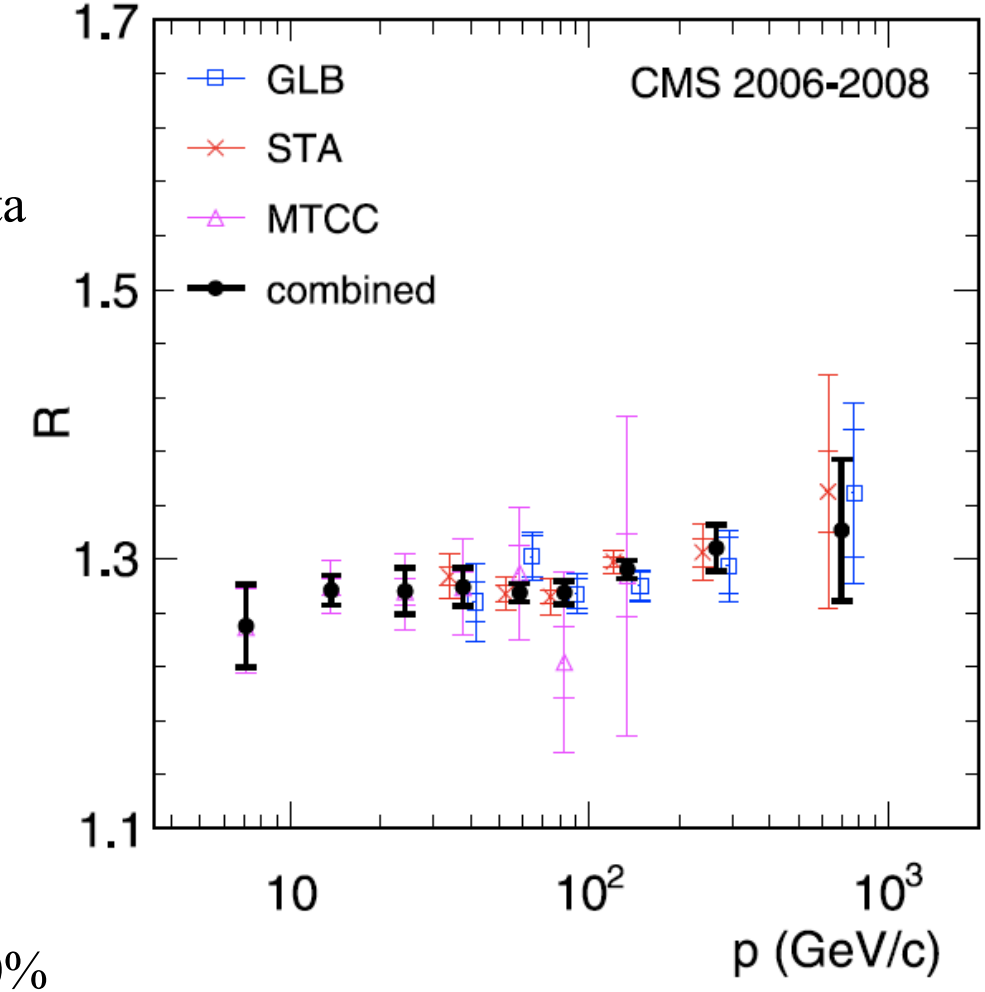
Matrix Inversion Example

Muon charge ratio R = number of positive to negative charged atmospheric muons

- Determine matrix from data, using momentum resolution function and charge misalignment measured in data

  -> reproduce MC predictions





- Unfold using matrix inversion
  - Sum of offdiagonal elements <20%

Unfolding procedure:

Iterative "Bayesian" Unfolding

Charged particle multiplicities

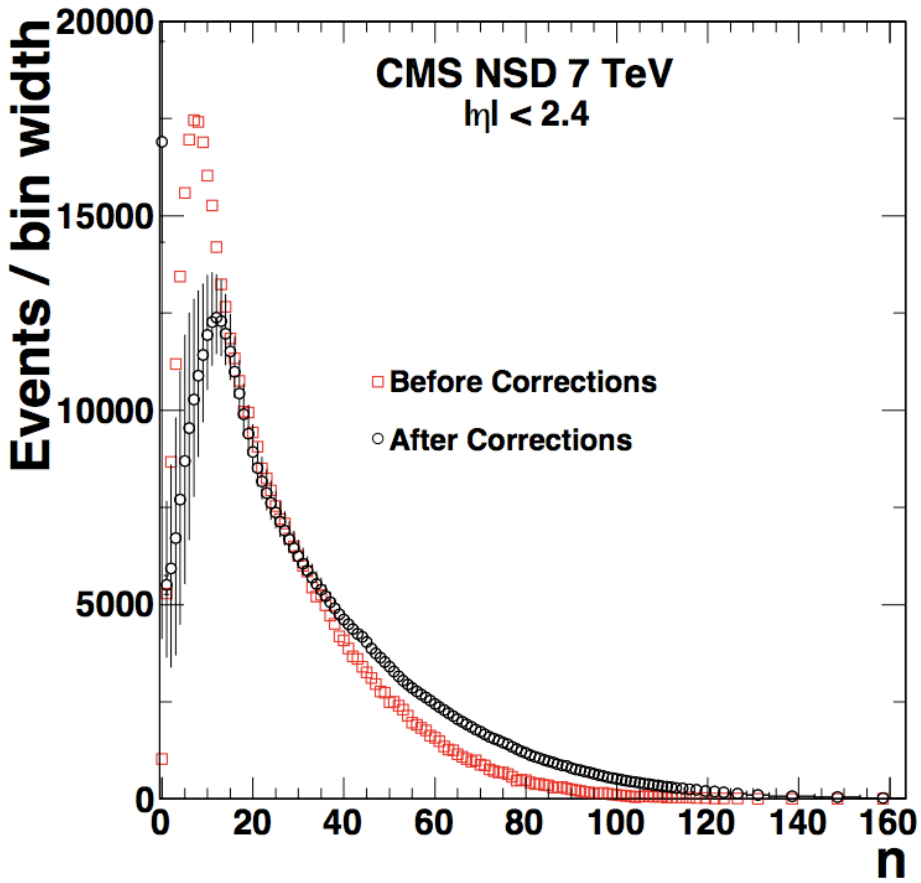
(arXiv:1011.5531v1)

Unfolding corrections

- Unfolding due to inefficiencies in track reconstruction and acceptance, secondary particles, decay products of long-lived hadrons

- Use bayesian unfolding procedure

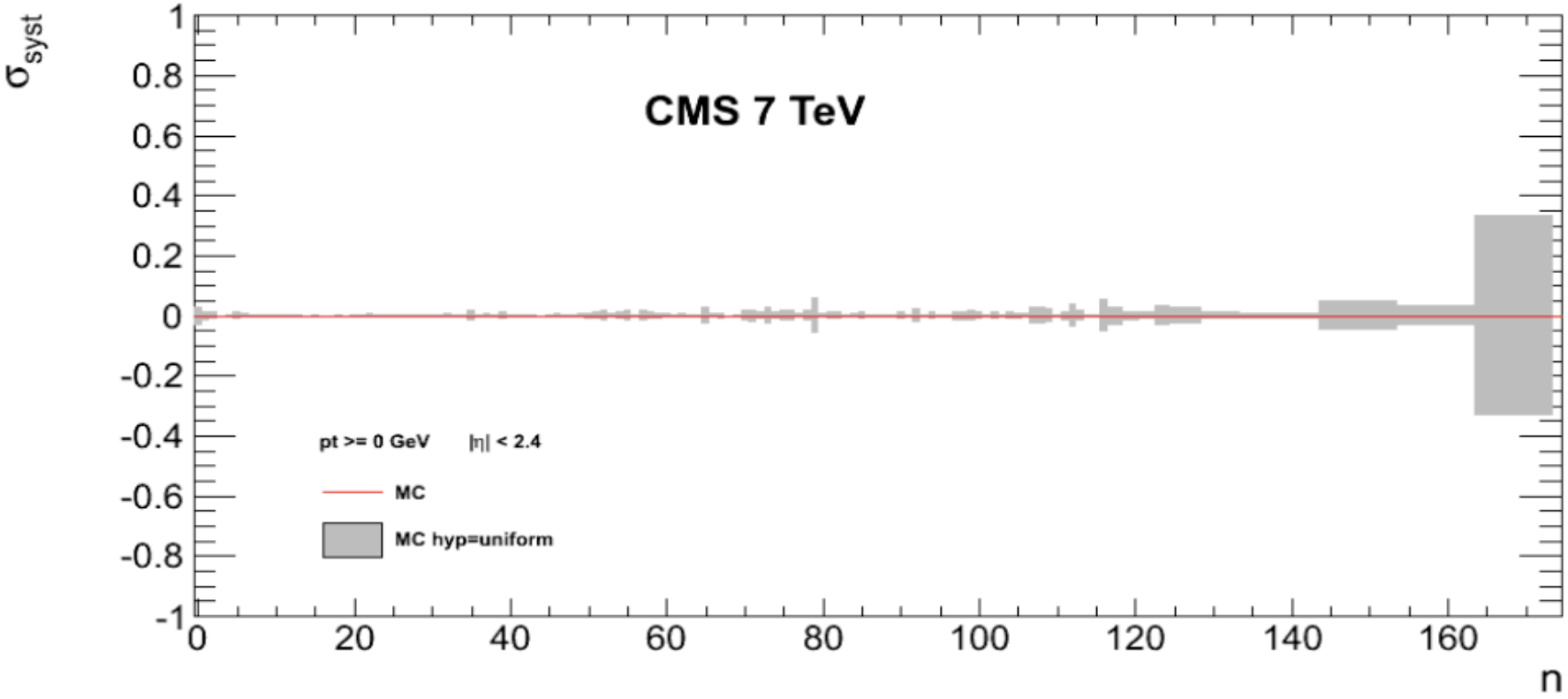
  G. D'Agostini, Nucl. Inst. Meth. A362 (1995) 487



Dependence on prior

Hypothesis 1: Uniform
Hypothesis 2: MC shape

Tests on Monte-Carlo:
Need to determine the #iterations for each hypothesis !

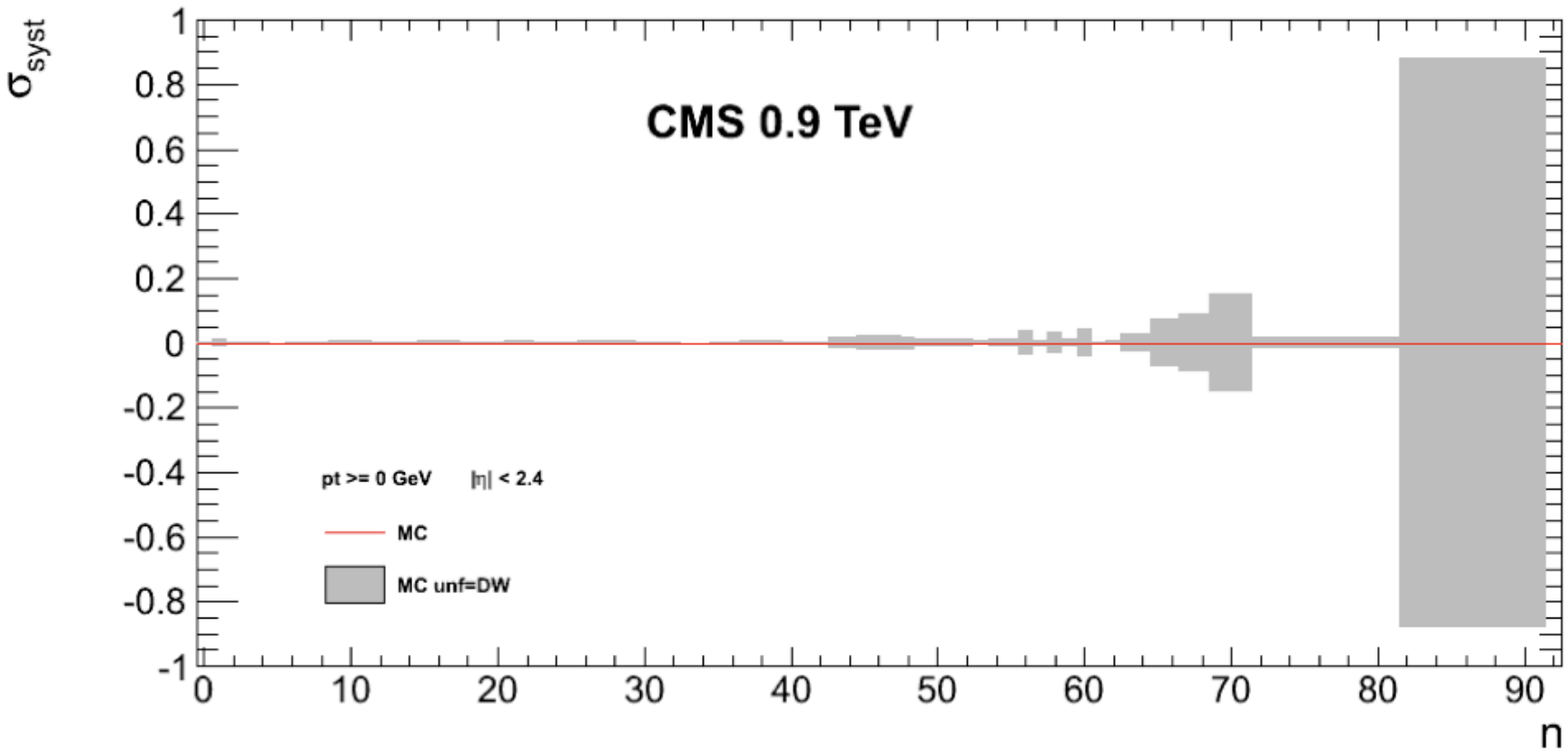


→ No dependence in Prior observed (after few iterations), tested at all energies and for all bins.

MC dependency

MC 1: PYTHIA D6T @ 900 GeV
MC 2: PYTHIA DW @ 900 GeV

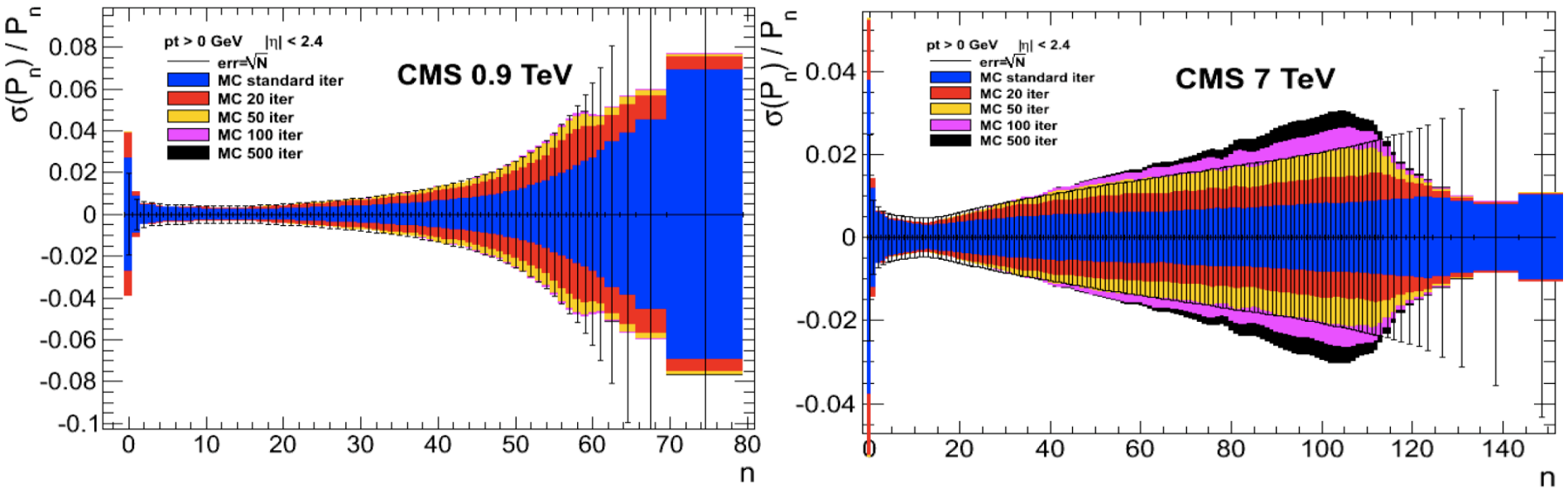
Tests on Monte-Carlo !



→ No effect beyond statistical errors / fluctuations !

Statistical error test

- Performed on MC with high statistics and no fluctuations in input reco data
- increasing the #iterations "fill" sqrt(N) expected errors

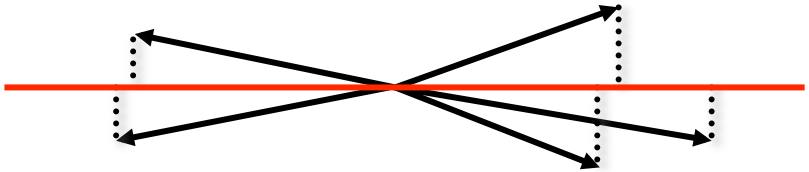


Take central value from small number of iterations (chi2 ~1) and the errors from large number of iterations

Unfolding procedure:
Singular Value Decomposition
Event Shape Analysis

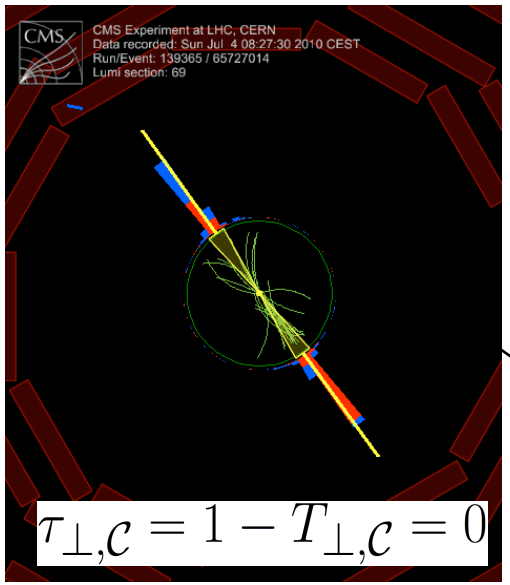
Illustration: central transverse thrust

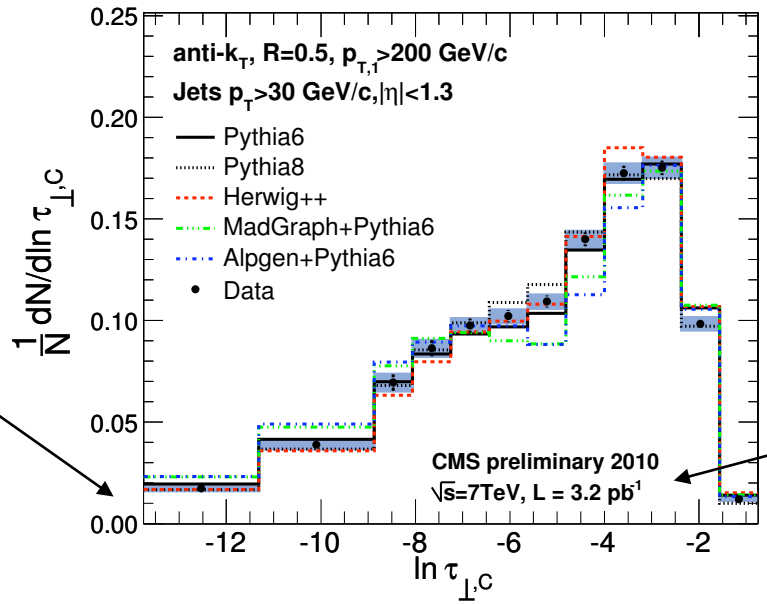
Defined in terms of four momenta in the transverse plane, Particle flow jet momenta are used as input for the event shape calculation

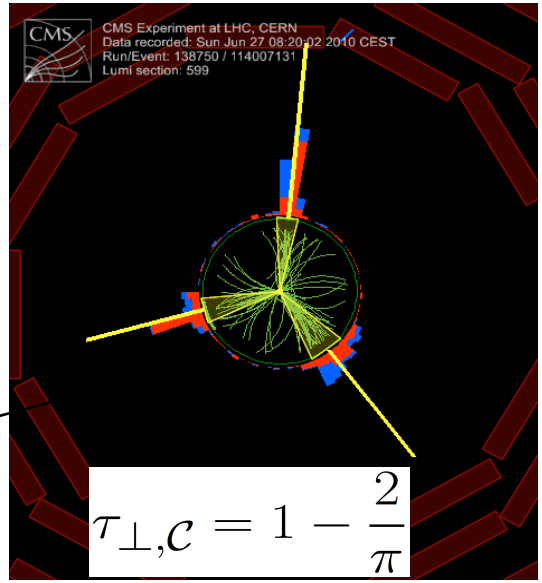


$$T_{\perp,C} \equiv \max_{\vec{n}_T} \frac{\sum_{i\in C} \left|\vec{p}_{\perp,i} \cdot \vec{n}_T\right|}{\sum_{i\in C} p_{\perp,i}}$$

Divide phasespace in bins of the leading jet p_T: 90 GeV/c<$p_{T,1}$<125 GeV/c (low), 125 GeV/c<$p_{T,1}$<200 GeV/c (medium), $p_{T,1}$>200 GeV/c (high)







SVD Unfolding

- The SVD method is based on the singular value decomposition of the response matrix.

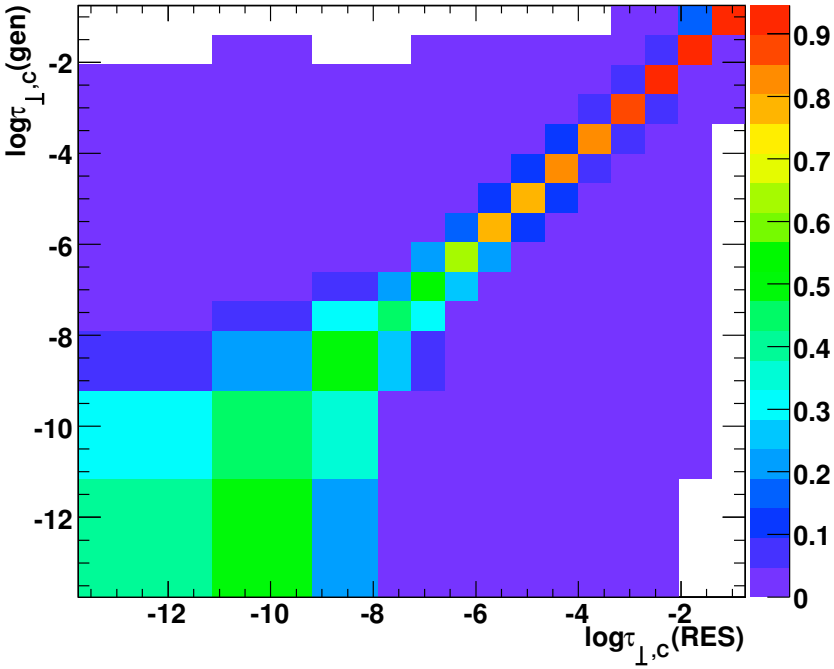
  Hoecker, Kartvelishvili, Nucl.Instrum.Methods A 372 (1996)469

- Smear jets in energy, η, φ. Calculate transverse thrust out of these smeared jets

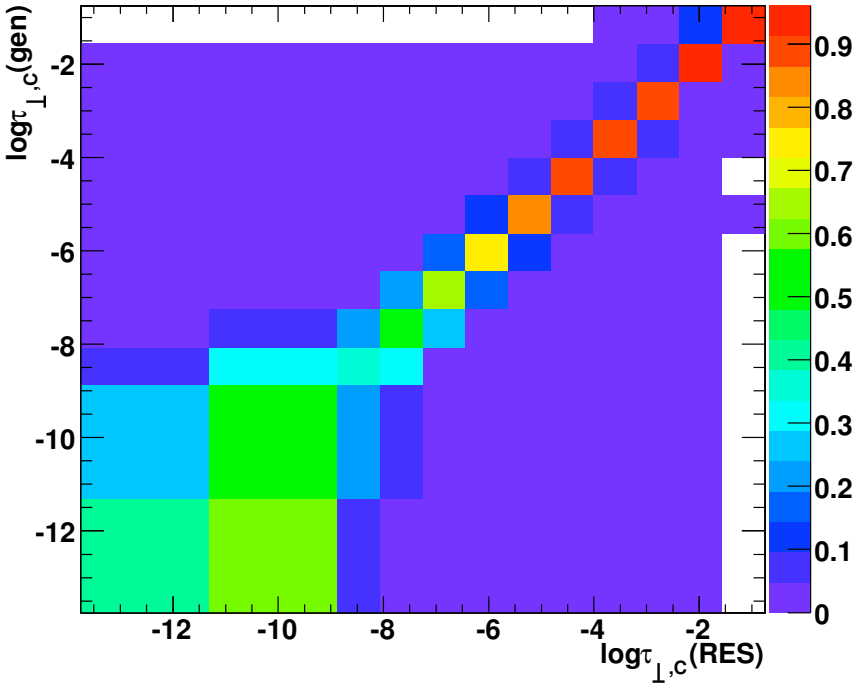
  -> pseudodata distribution
- Use the response matrix derived from Pythia6-QCD samples (Tune D6T).
- Check the closure between the unfolded pseudo-data distribution and the generator hadron level distribution.
- Choose the regularization parameter such that the chi2-value between the unfolded and the generator level distributions is minimal.

Response matrix

- Response matrix of central transverse thrust for the medium and high bin, derived from Pythia6 QCD MC



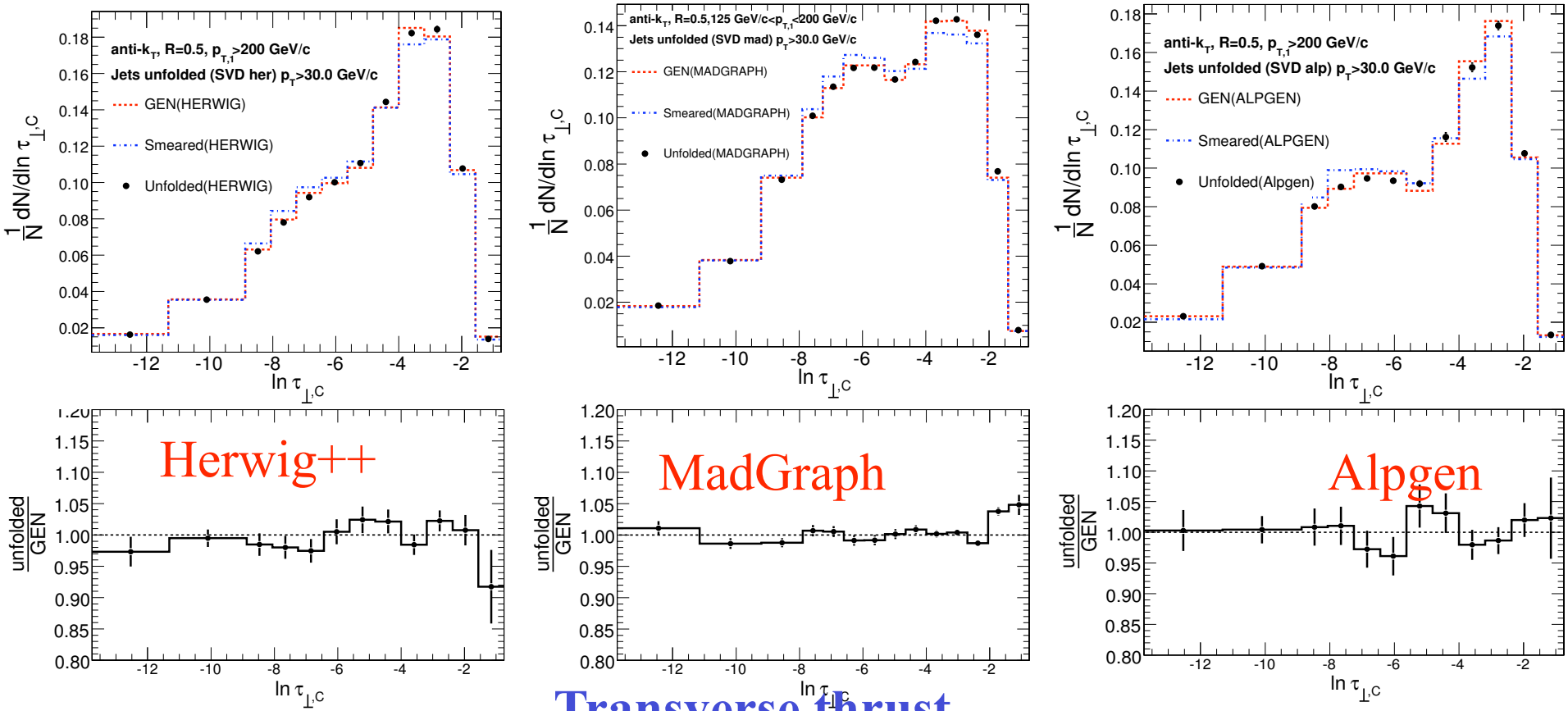
medium



high

Significant off-diagonal elements, especially for low event shape values (>30 %), elements > 0.90 only for highest event-shape values

Cross check for other generators: Unfolded (SVD) vs Gen

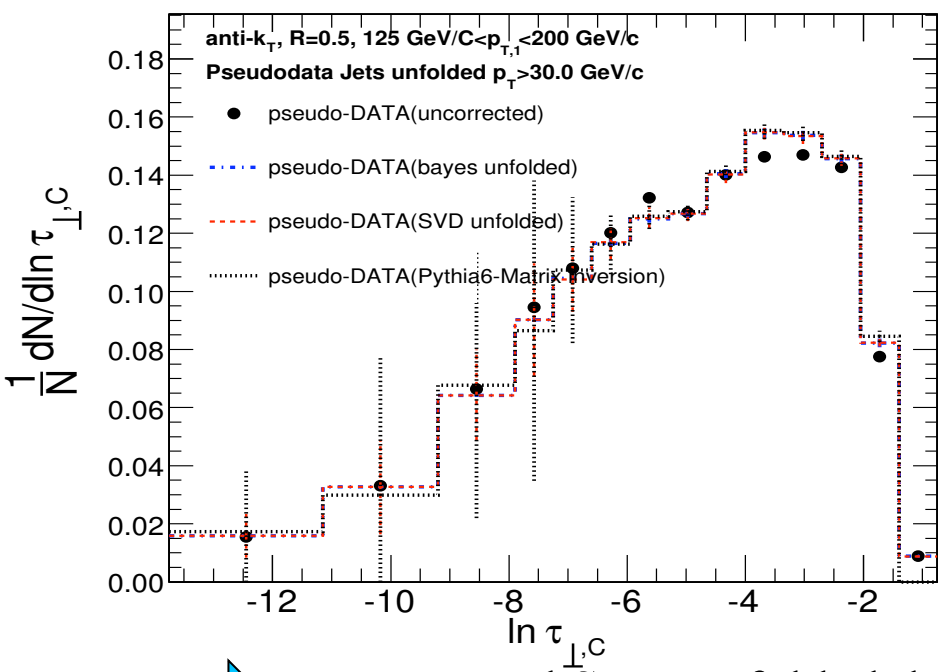


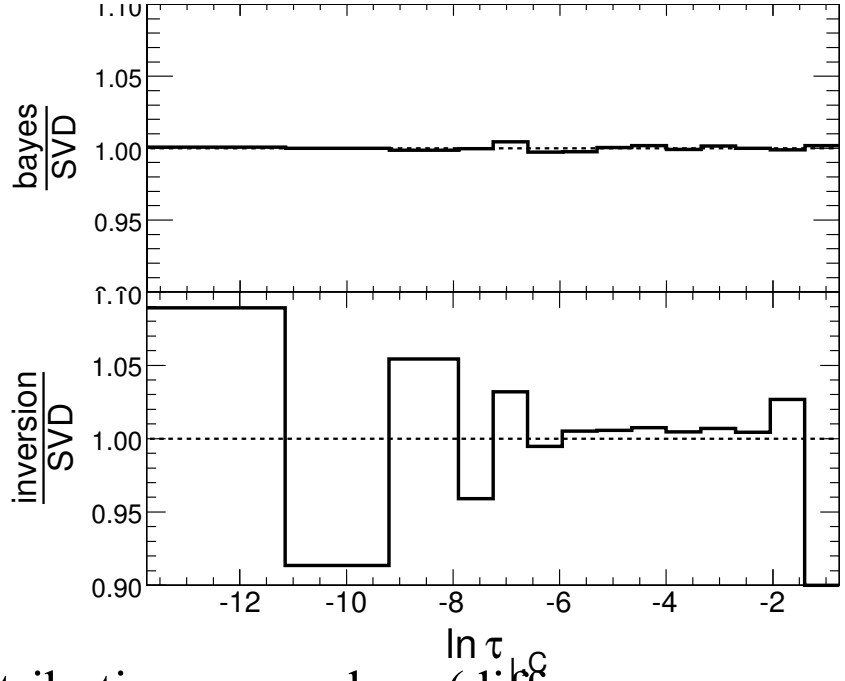
Transverse thrust

Good closure for other MC generators can be observed, using the response matrix from Pythia6

Cross check: compare different unfolding procedures

Compare different unfolding procedures:
SVD unfolding, iterative bayesian unfolding and the *inversion* of the response matrix

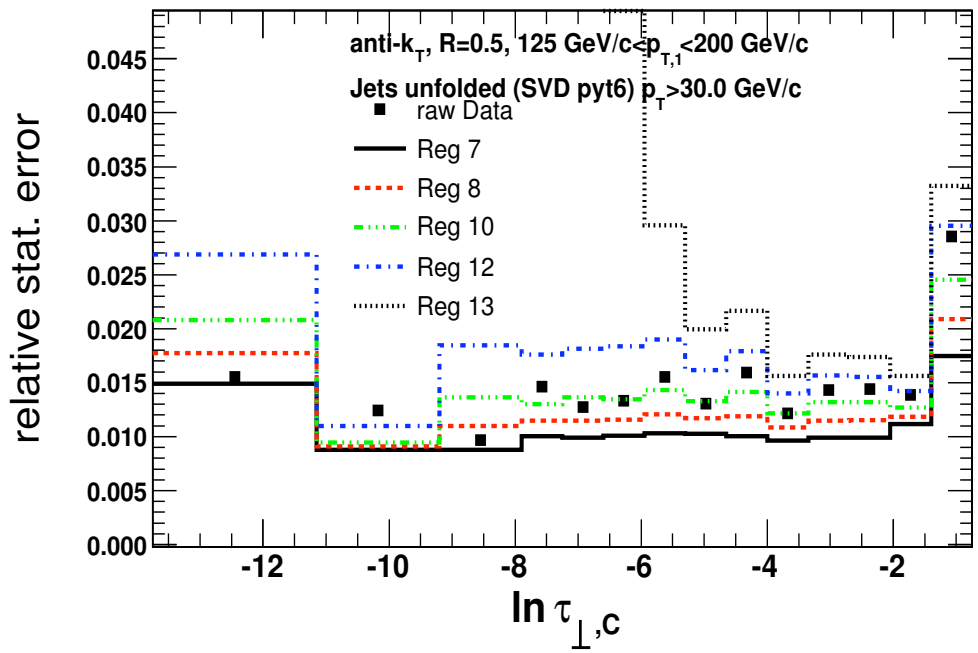


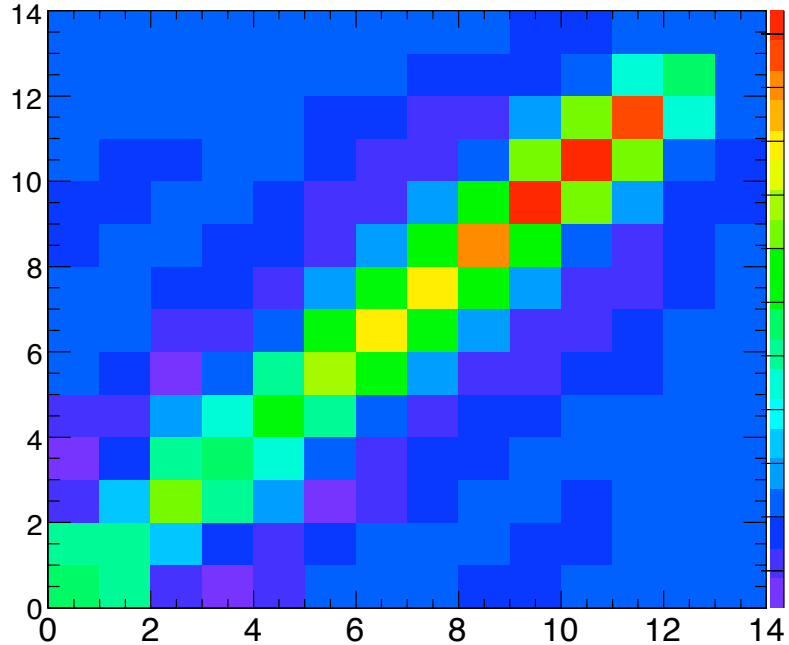


Bayes and SVD unfolded data distributions very close (differences <1 %). For the medium and high bin the matrix inversion also leads to a satisfactory result (<2% differences for higher event shape values).

SVD error propagation

- Compare relative statistical errors in pre-unfolded data and unfolded data, using certain regularization parameters





For certain regularization choices the relative stat. errors of all bins after unfolding < errors before unfolding and sqrt(N)

Chi2-tests

Check if a bias towards a particular generator is introduced by the unfolding:

1. Compare to chi2 between MC hadron level distribution and unfolded "data" distribution
2. smear generator level by the Pythia6 response matrix, calculate chi2 between smeared MC and pre-unfolded data distribution

central transverse thrust, medium bin

MC Generator	χ^2 values between data and smeared mc	χ^2 values between unfolded data and Gen mc
PYTHIA6	421	398
HERWIG++	211	200
MADGRAPH	2590	2570
ALPGEN	3860	3860

→ Chi2 order is the same before and after unfolding, values are similar

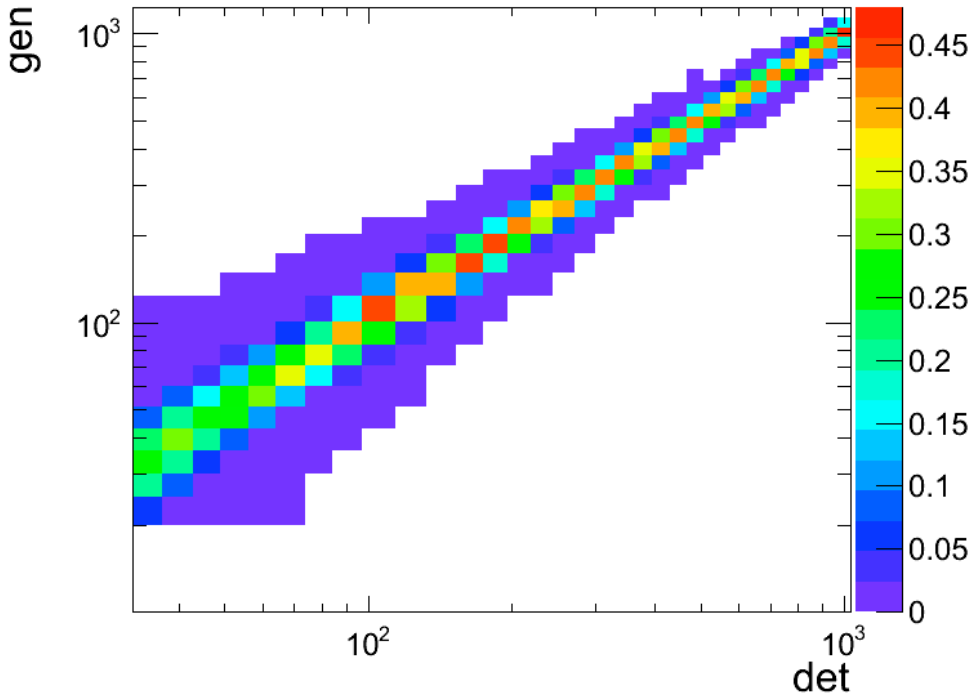
Tikhonov regularization

Measurement of the inclusive jet pt spectrum and jet shapes in pp collisions

Matrix Definition

- Available MC generators fail to produce the high pt spectrum

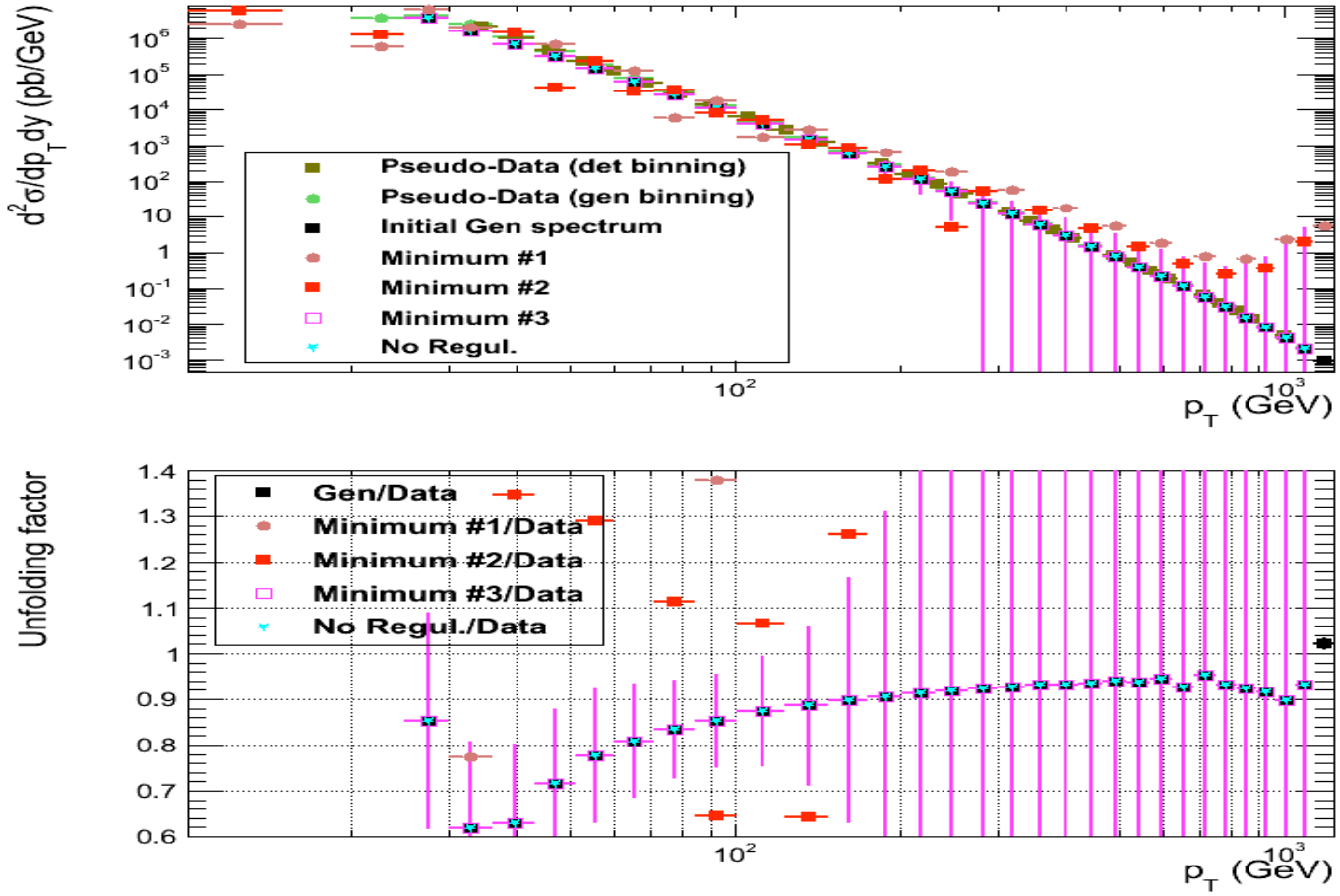
- Determine the response matrix by smearing a theory curve by jet resolution functions



➡ Suitable choice of number of bins, number of data n vs theory μ bins important for solution of the regularization method

First closure tests and errors

- No square response matrix R; system to solve and regularize: $R^T n = R^T R \mu$
- Errors propagation of regularized solution by matrix inversion leads to big errors



Summary

- Several unfolding methods are used in CMS analyses
  - Matrix Inversion, bayesian iterative unfolding, SVD, Tikhonov

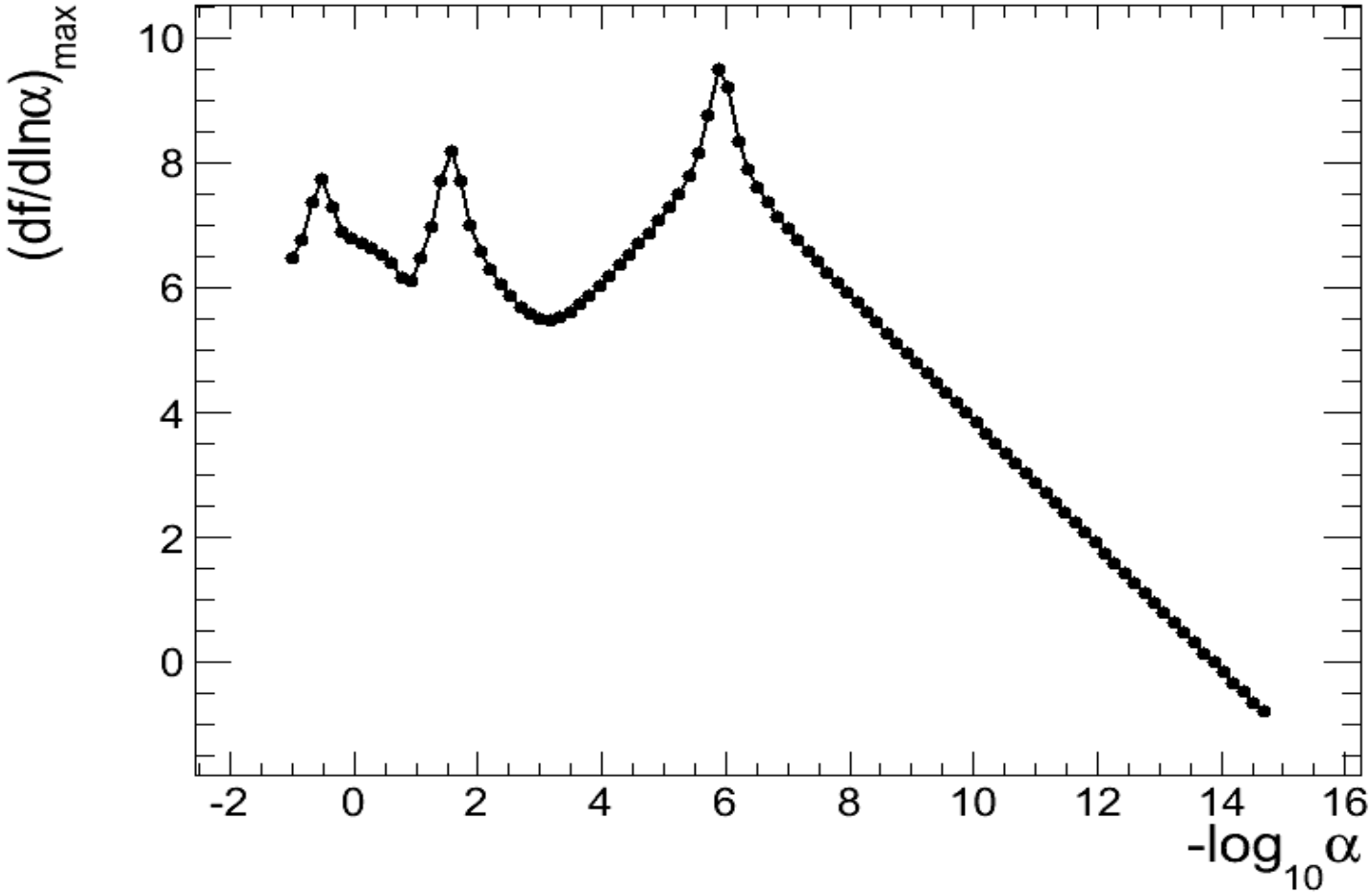
- Test for unfolding involve model dependence, closures of the unfolding, pull distributions

- Issues in error propagation
  - Errors sometimes smaller as sqrt(N)
  - Use errors of matrix inversion or errors of least regularized solution

BACKUP

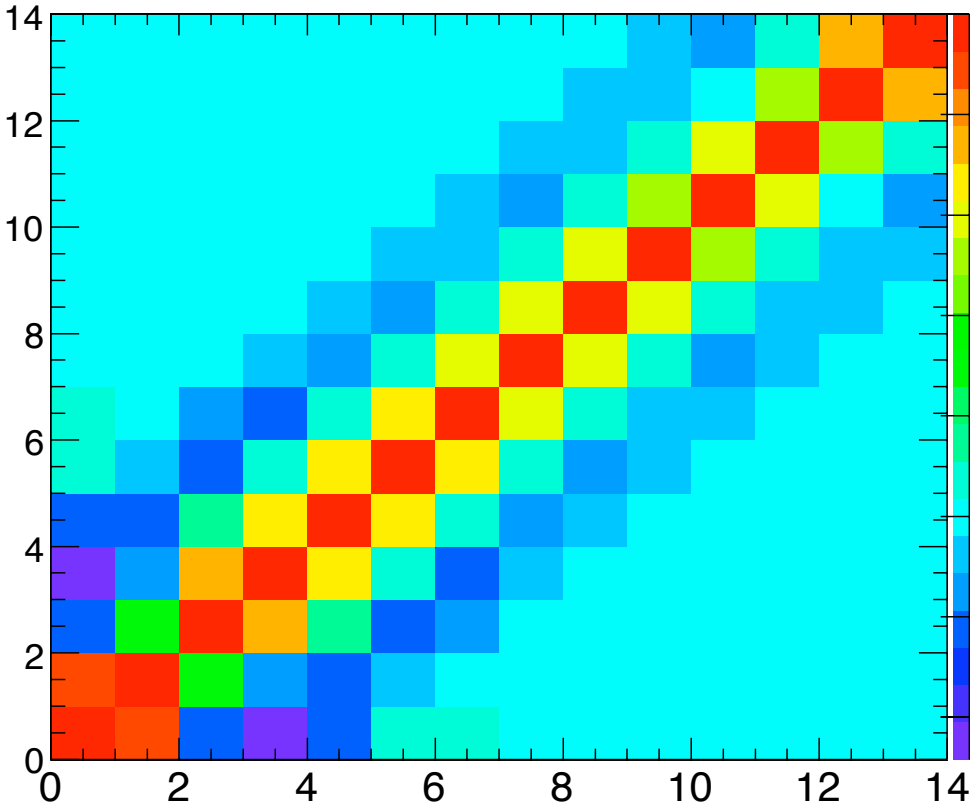
Thikonov regularization function

Minima correspond to possible stable solution. First minima depends on initial conditions. Deepest and smoothest minima should be best solution



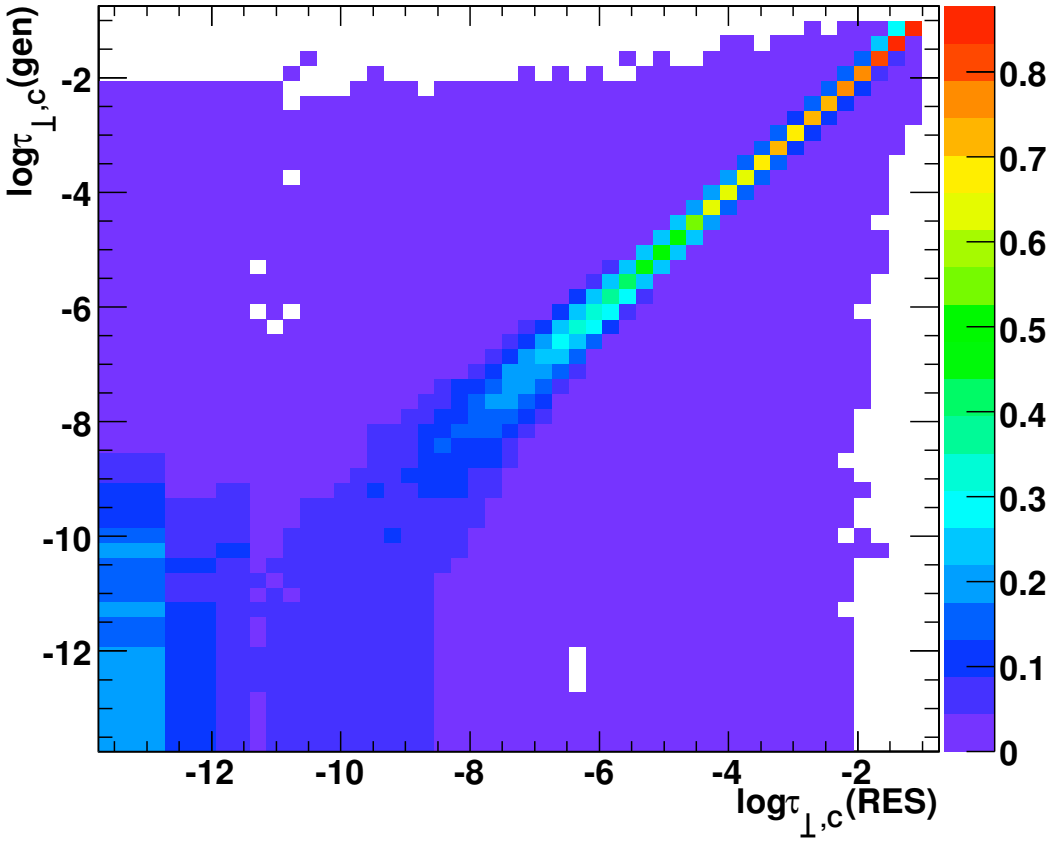
Event Shape Correlation Matrix

correlation matrix: pythia6 D6T is used as model, pseudodata is also pythia6

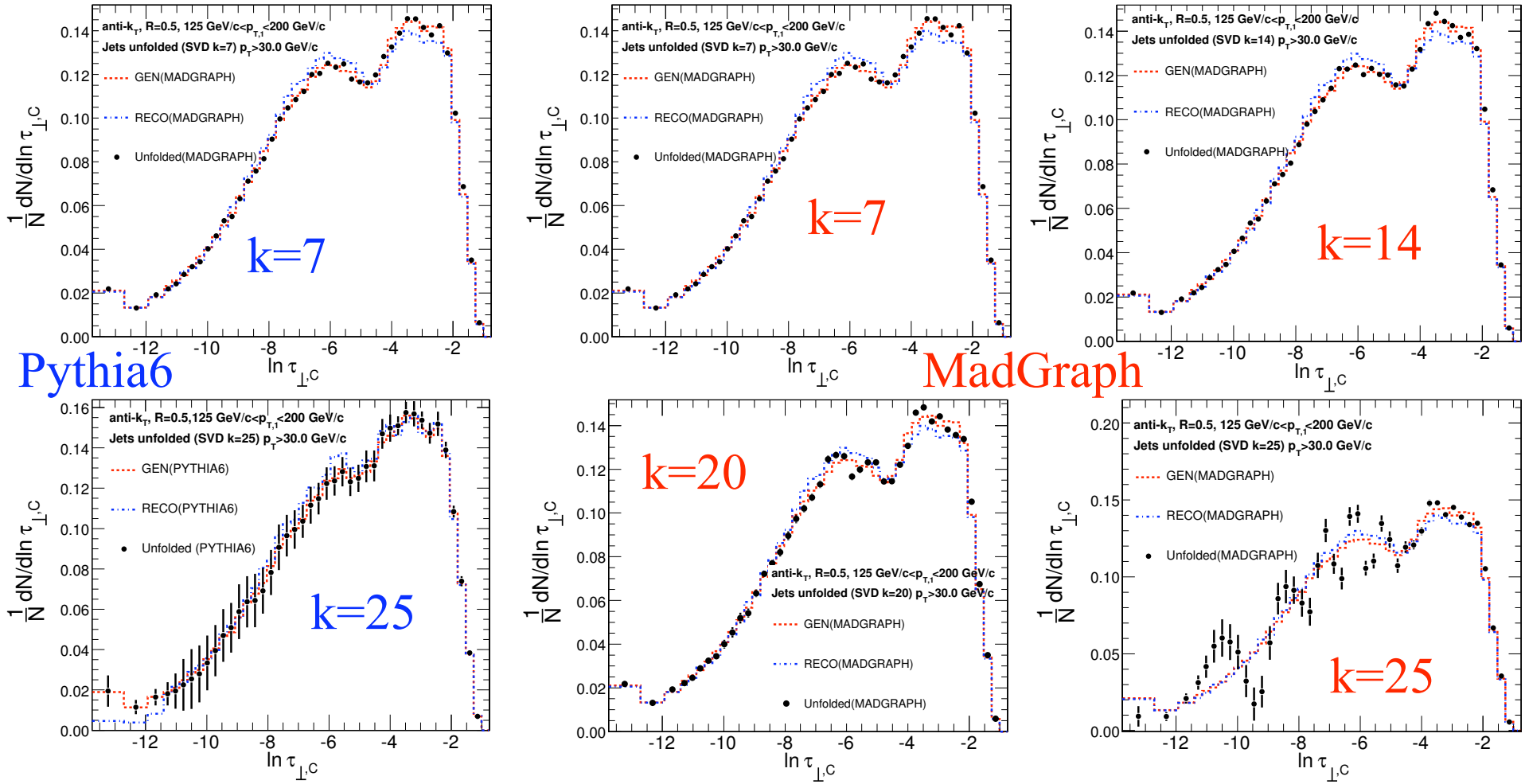


Event Shape: finer binning in response matrix

Pythia6 D6T is used as to determine the response matrix, the pseudodata values are derived using smeared particle jets as input



Event Shape: varying the regularization parameter



Big oscillations in case of madgraph for k>25, best chi2 for k=7