JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

Experiences with GURU

Sebastian Eckweiler - Universität Mainz

GURU

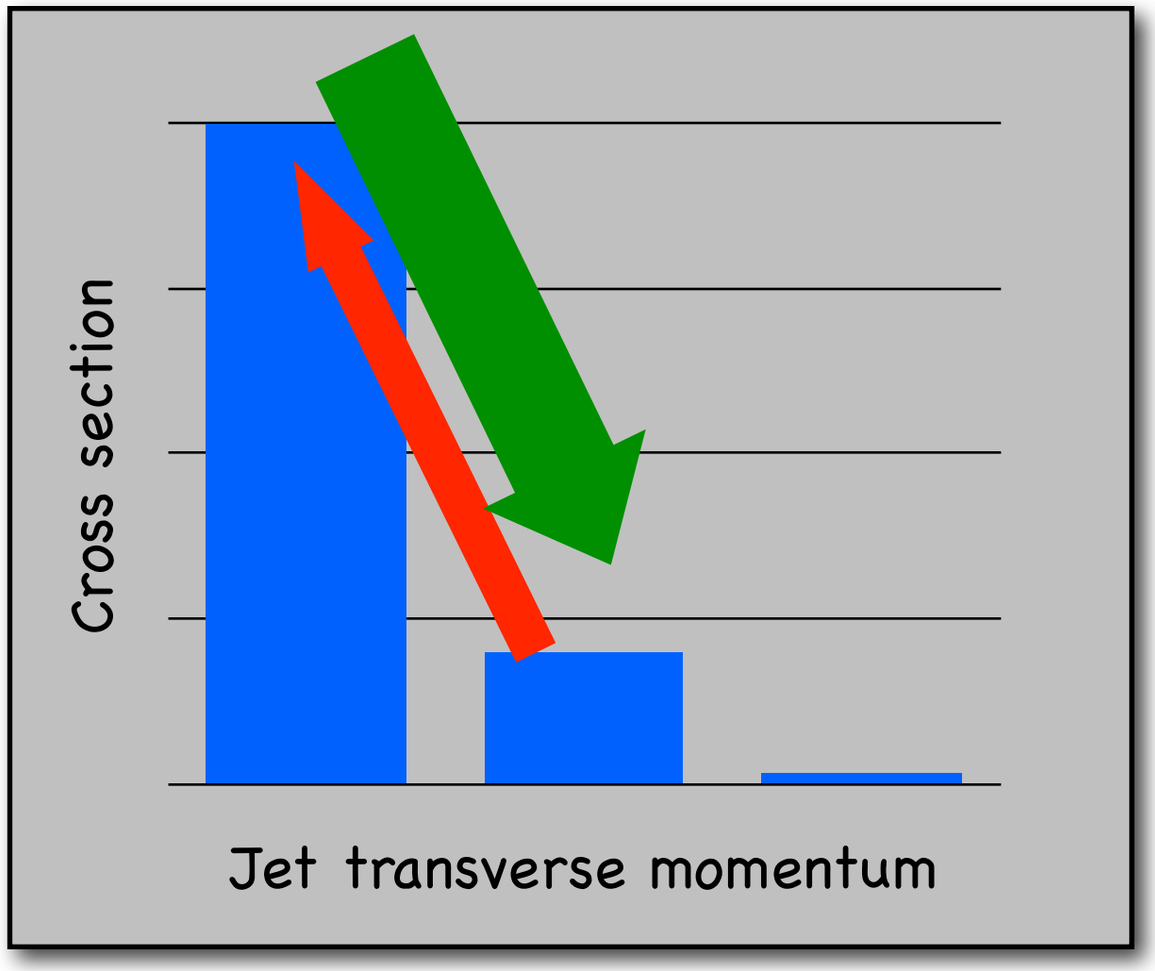
- Tool for regularized unfolding:
  ‚*SVD Approach to Data Unfolding*‘: http://arxiv.org/abs/hep-ph/9509307v2

- c++ wrapper available on the web:
  http://www-d0.fnal.gov/~ghesketh/unfolding/

- How does it work?
  - Unfolding problem treated as sys. of. equations, e.g.: t * R = b
  - With:
    b = measured quantities (‚*data*‘)
    t = true quantities, ideally the result of the unfolding
    R = response matrix, usually derived from Monte Carlo simulations
  - GURU employs singular value decomposition to solve this system of eq
  - Rotate such that R becomes diagonal - solution(s) becomes trivial

Application of GURU

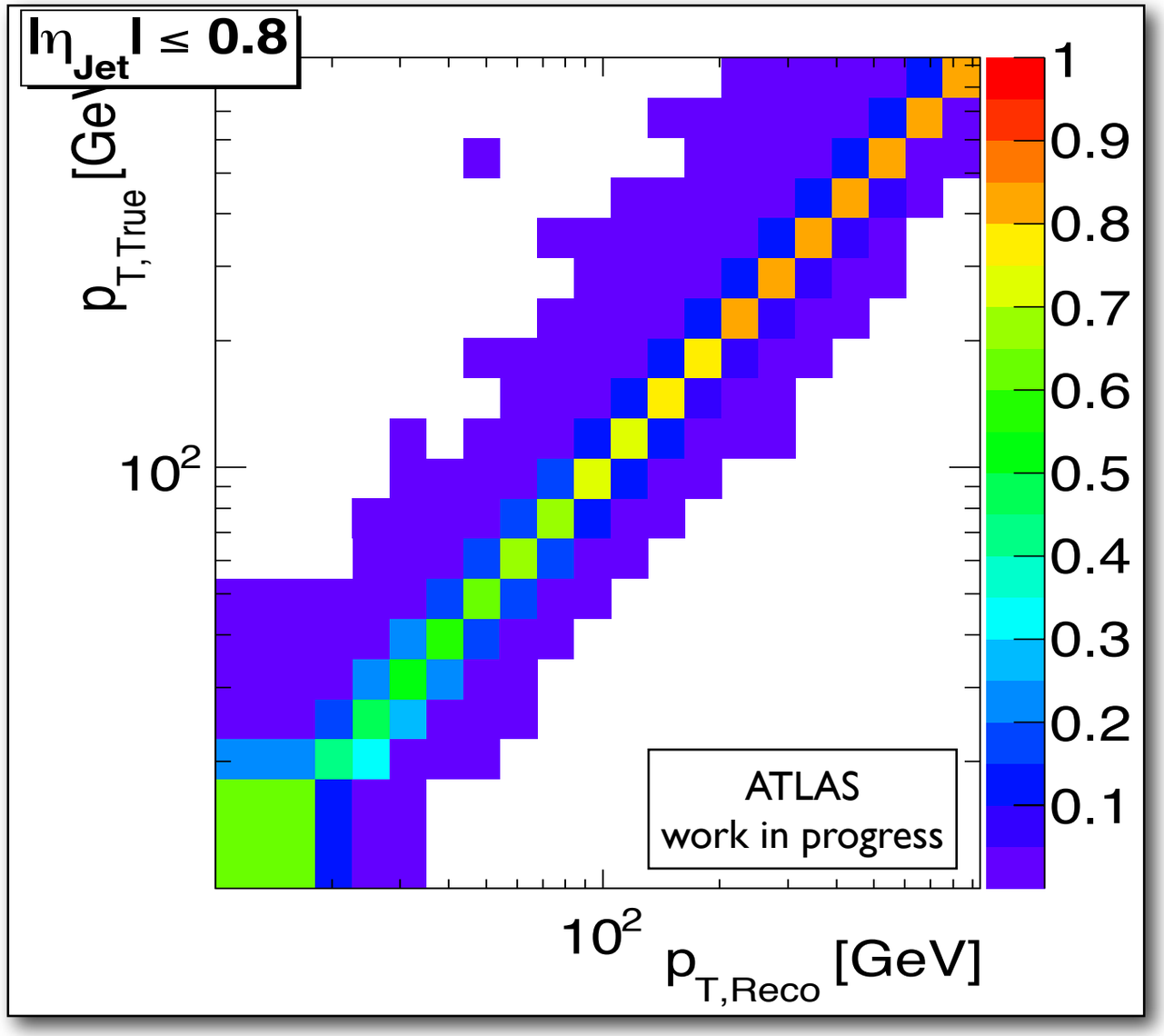
- Use case: unfolding of incl. jet cross section
- Unfortunately measured quantities aren‘t identical to the observable we‘d like to measure, but (hopefully) closely related to them
- Concerning jets: energy reconstructed in calorimeter != energy of stable particles, due to:
  - Energy-scale: assumed to be 1
  - Energy-resolution:
    can‘t do too much about that,
    once the calorimeter is built
- Impact on cross section:
  - Migrations between bins distort measured shape
  - Net effect in general:
    increase of measured cross section



Unfolding with GURU

- How to use GURU?
- In principle only 2 inputs needed (not completely true for this case):
  - b = data as a TH1
  - R = Response Matrix as TH2

- In our case:
- R(i,j) =
  probability that reconstructed jet with $p_{T,Reco}$ in bin i has an associated truth jet with $p_{T,True}$ in bin j



- Matrix mainly diagonal:
  $\langle p_{T,Reco}/p_{T,True} \rangle \approx 1$
  (means our calibration makes sense)
- Migration primarily to neighboring bins
- Probability decreasing by ~1/10 with every bin

Experience with GURU

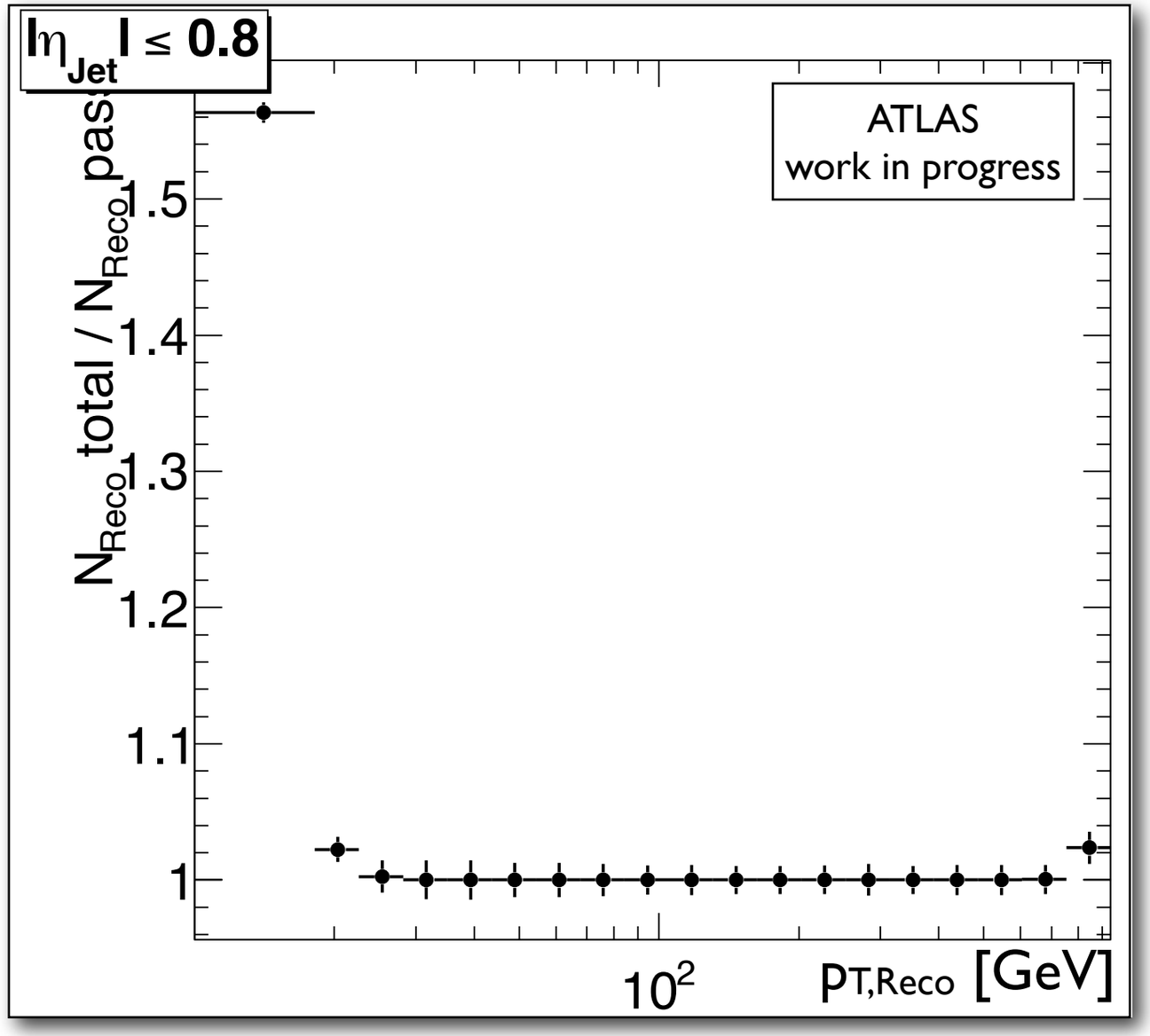
- The response matrix only handles truth jets between $p_{T,True,Min}$ and $p_{T,True,Max}$ (here [50, ~900] GeV)

- But:
  Reco-Jets with 50 GeV may have nearest truth jets with only 45 GeV.

- Those losses/gains at the boundaries are not visible in the response matrix
  (esp. as long as truth and reco bins are identical)

- To correct for that, include the probability that:

  Jet in reco bin i has truth jet between $p_{T,True,Min}$ & $p_{T,True,Max}$



- Apply as an acceptance correction before the unfolding

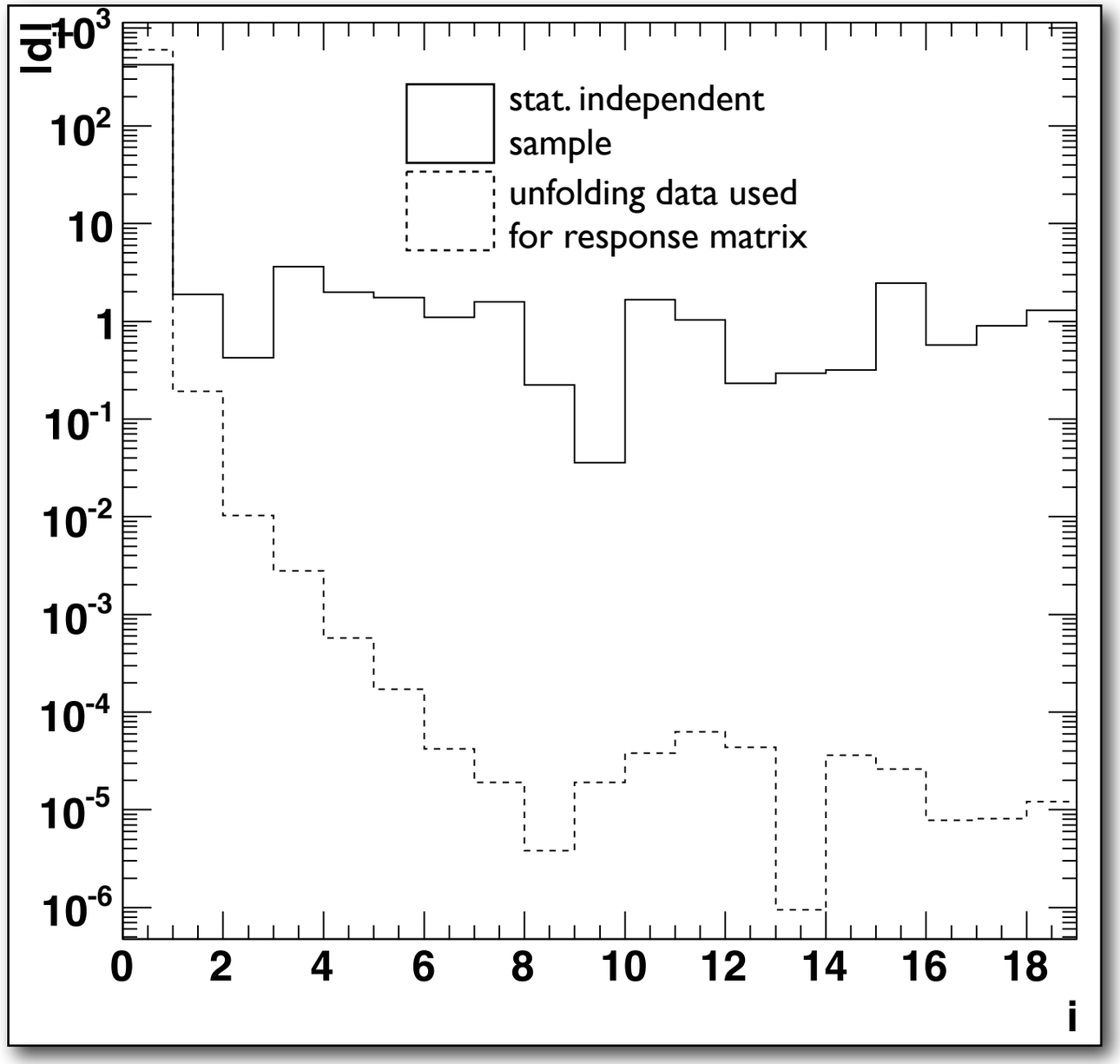
What GURU does

- Starting off with: $R*t = b$
- R is diagonalized and system is rotated accordingly: $S*z = d$
- Equations are scaled such that: $\sigma(d_i) = 1$
- Since S is diagonal the solution is straight forward:
- Still, this is not regularized... $z_i = d_i / s_i$
- Adding the regularization this becomes: $z_i = s_i * d_i / (s_i^2 + \tau)$

- With τ being the regularization parameter one has to choose
- GURU chooses N values $\tau = s_i^2$ and gives one result per τ
- This is where GURU stops and one has to employ some own thinking

Choice of τ

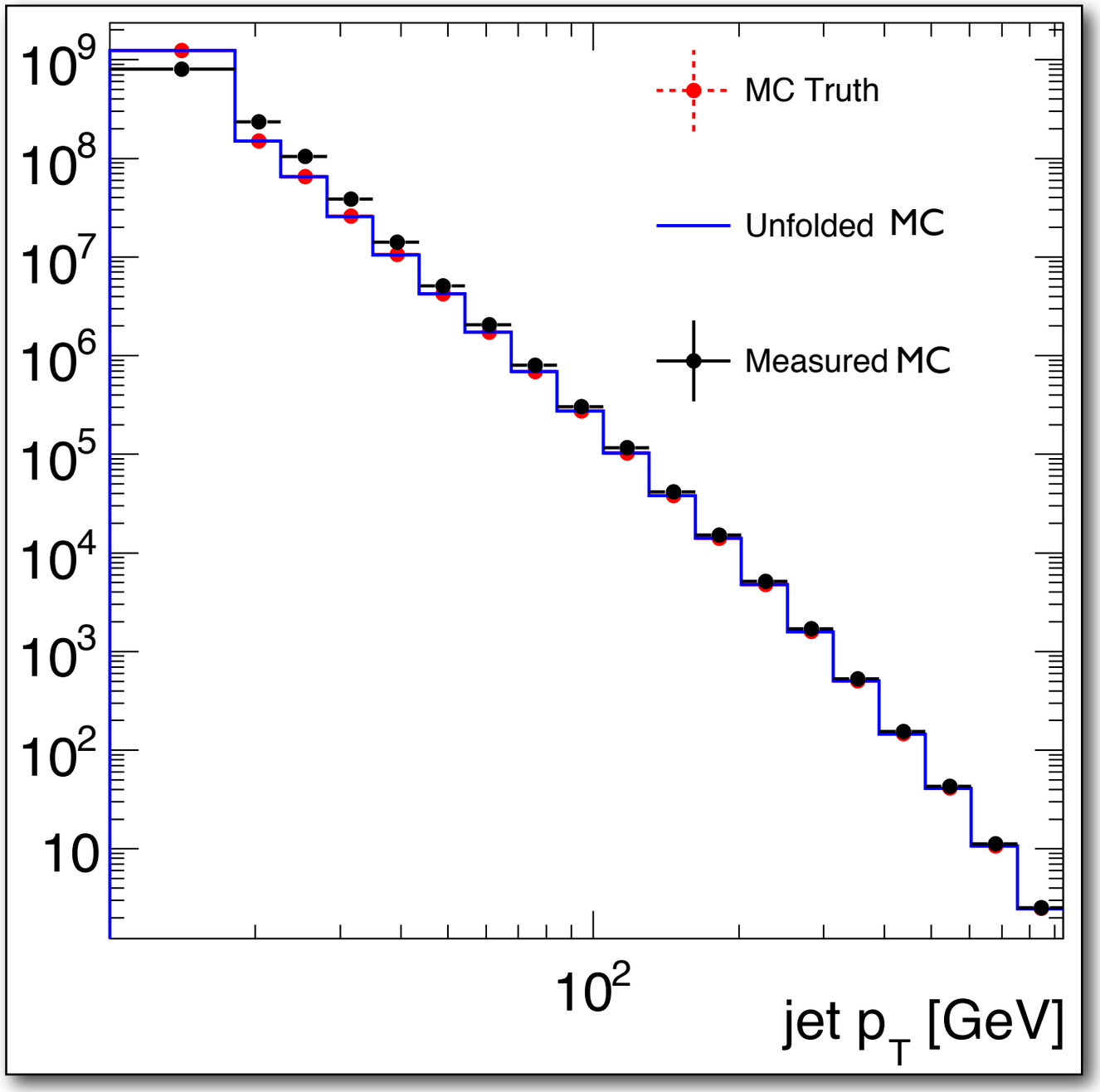
- There's ‚dedicated' section on this in the paper
- To choose a proper regularization, the study of the d_i is crucial
- d_i are (usually?) exponentially decreasing
- Since $\sigma(d_i) = 1$, everything close to one hints at a statistically insignificant component
- Find k where $d_i(i>k) \approx 1$
- Here, choose:
  $\tau = s_1^2$ or s_2^2



Results

- In these closure tests GURU basically gives perfect results

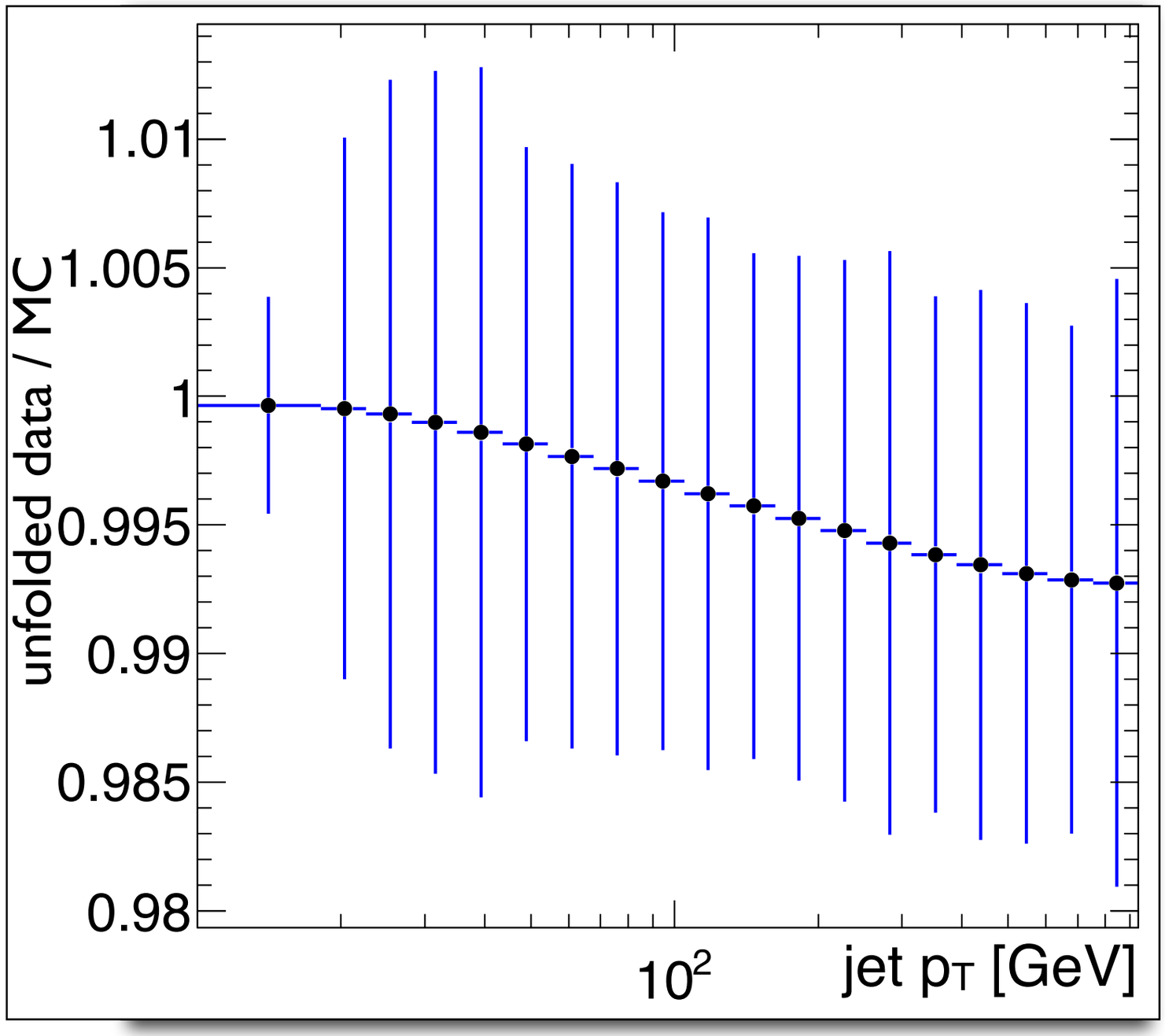
- No surprise, since response matrix is perfectly well known



Results

- In these closure tests GURU basically gives perfect results

- No surprise, since response matrix is perfectly well known



Summary

- Started studying the use of GURU for the unfolding in the measurement of a incl. jet cross section
- Personally biased judgement:
- Positive:
  - It works
  - Some tripwires in its use exist - e.g. the response matrix needs careful definition
  - It‘s readily available on the web and assuming some ROOT experience easy to use
  - Last but not least: the author replies to emails ;)
- Negative:
  - It certainly has some black-box character -
    rotate&scale matrices are part of the output, but hardly intuitive
  - Can be unstable
    - output is very limited and tough to interpret