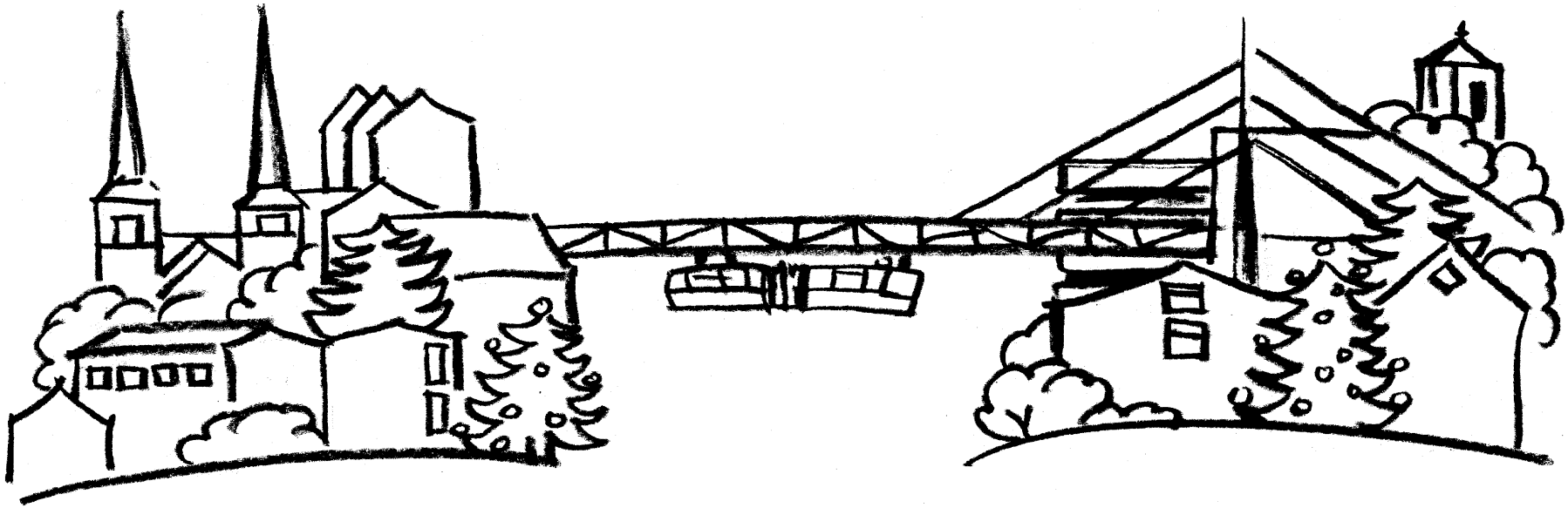
Experience with Bayesian Unfolding (after method proposed by D´Agostini)



Marisa Sandhoff

Bergische Universität Wuppertal

Contents

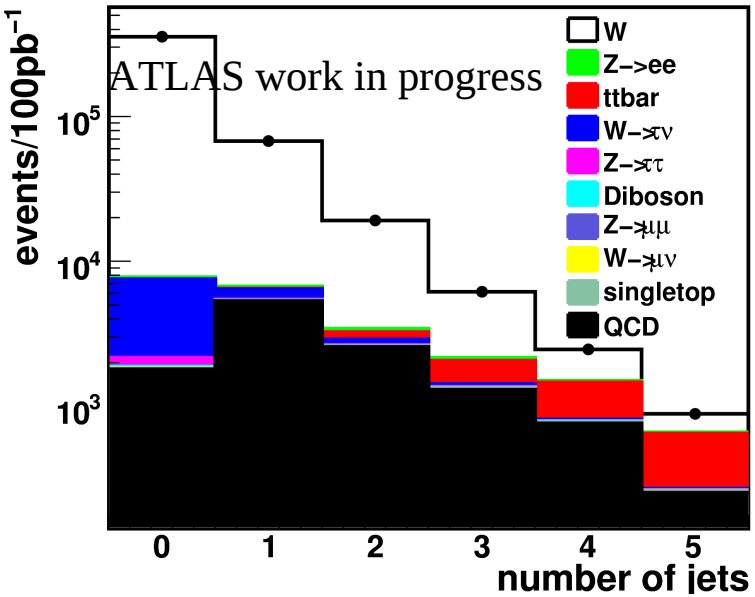
- Motivation
- Strategy of unfolding an experimental distribution
  - W <-> Z migrations
  - Model dependency of migration matrix
- Estimation of statistical errors
- Summary

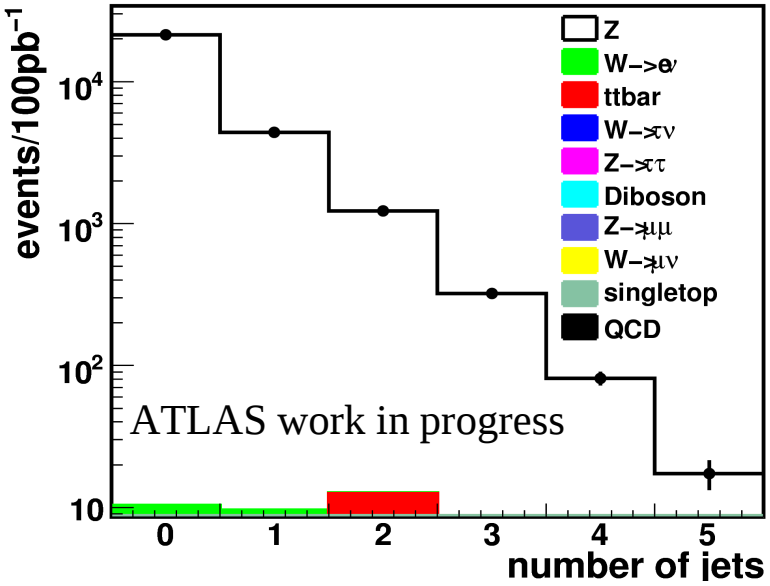
Motivation

- Production of W and Z bosons + jets
  central process at LHC
  -> important background processes
  -> test QCD predictions, test and tune MC generators
- Aim of analysis:
  measure **ratio W+njets/Z+njets**
- **Unfold W+jets and Z+jets** then calculate ratio
  -> compare directly to QCD and MC generator predictions, without detector and reconstruction effects

Analysis with unfolding -> as I learned by doing..

- Always first step: Select W and Z candidates



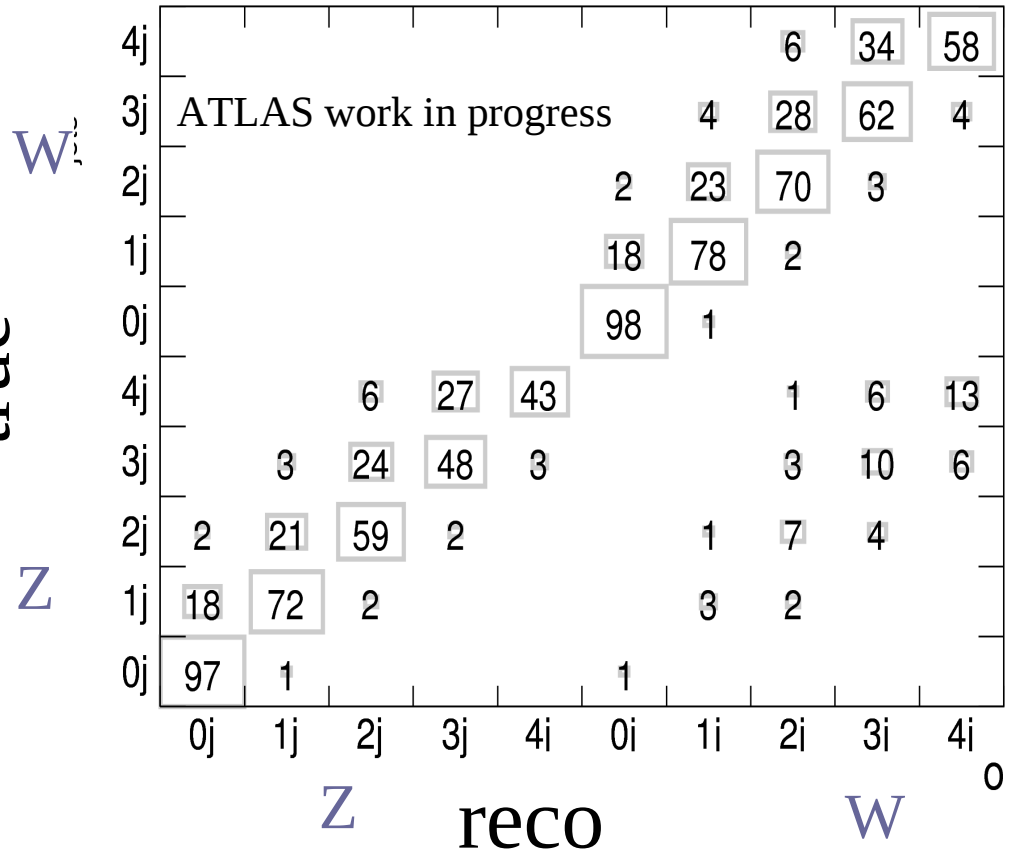


- First subtract background -> then unfold only signal (tried to handle bg during unfolding -> large off-diagonal elements, does not converge!)

Next “problem”: W <-> Z Migrations!

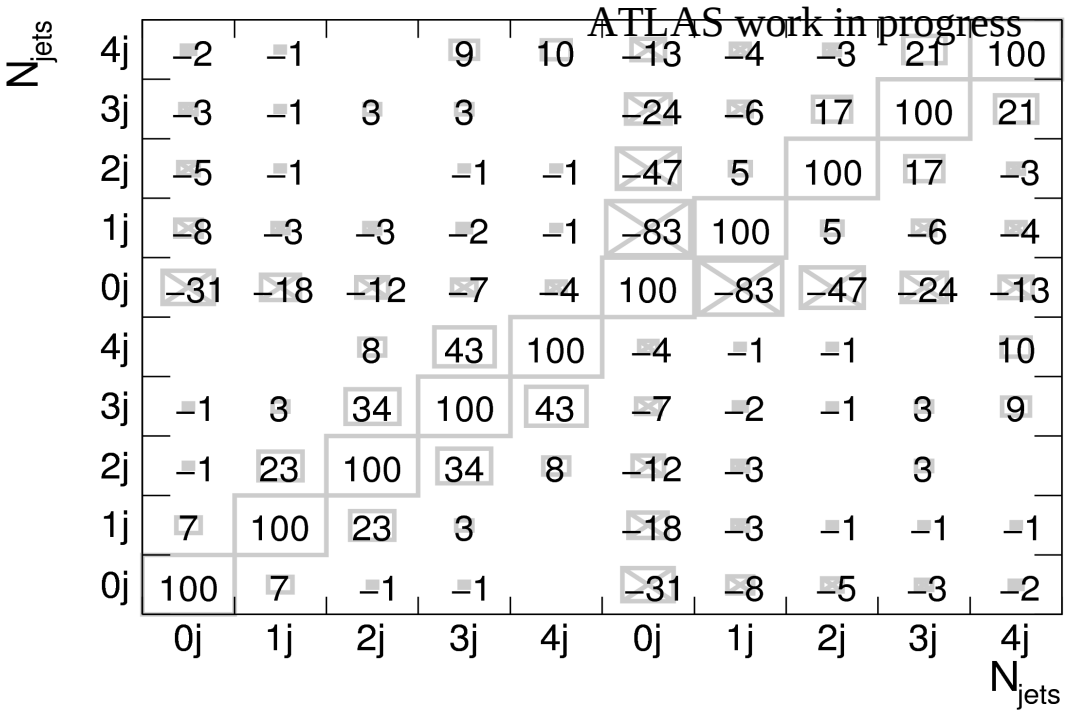
- W <-> Z migrations exist
  - Many Z events reconstructed as W -> how to handle?
- Correct during unfolding?
- Migration matrix with large off-diagonal elements! (typical problem for unfolding)
- 10 times more W than Z

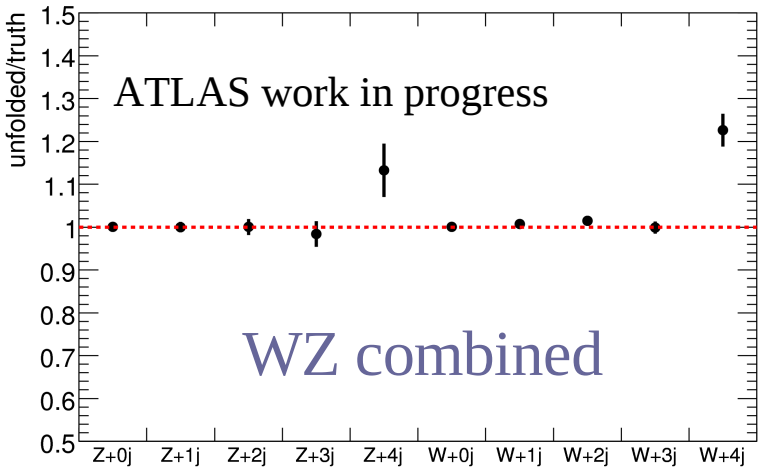
Migration prob. in %

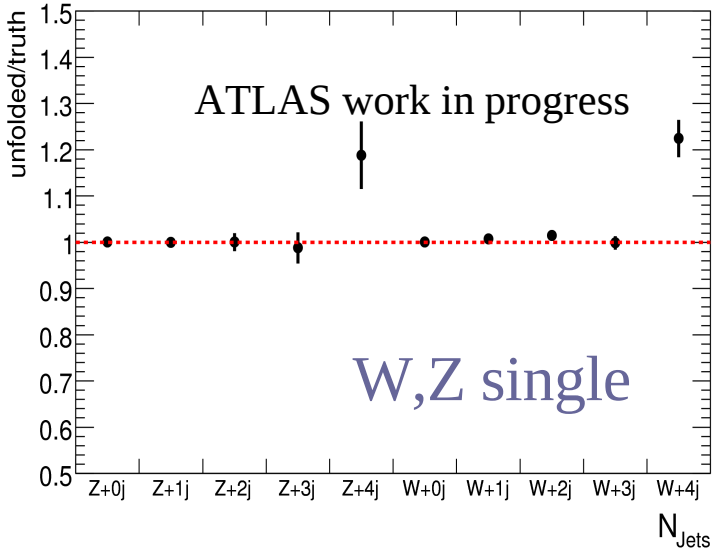


W <-> Z Migrations II

- ... what happens during unfolding?
- -> works, but.... (different results!!!)
- -> errors do not change??? Strange! Strong correlations! Also strange...





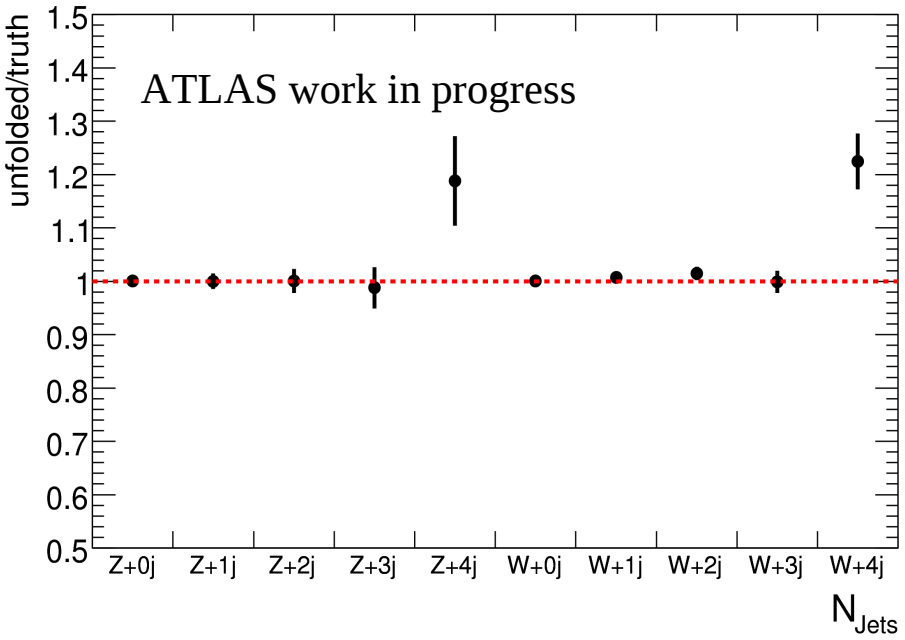


Lessons learned

- Factorize:
  - Event selection
  - Background subtraction
  - **Unfolding**
    - Better unfold **W and Z separately**!
  - Unfolding = re-sorting of events!
    -> Calculate migration matrix with selected events, **same cuts** like for W and Z
- After unfolding:
  Signal efficiency: from “all” -> “selected”
  no migrations here! Only 1/ε!

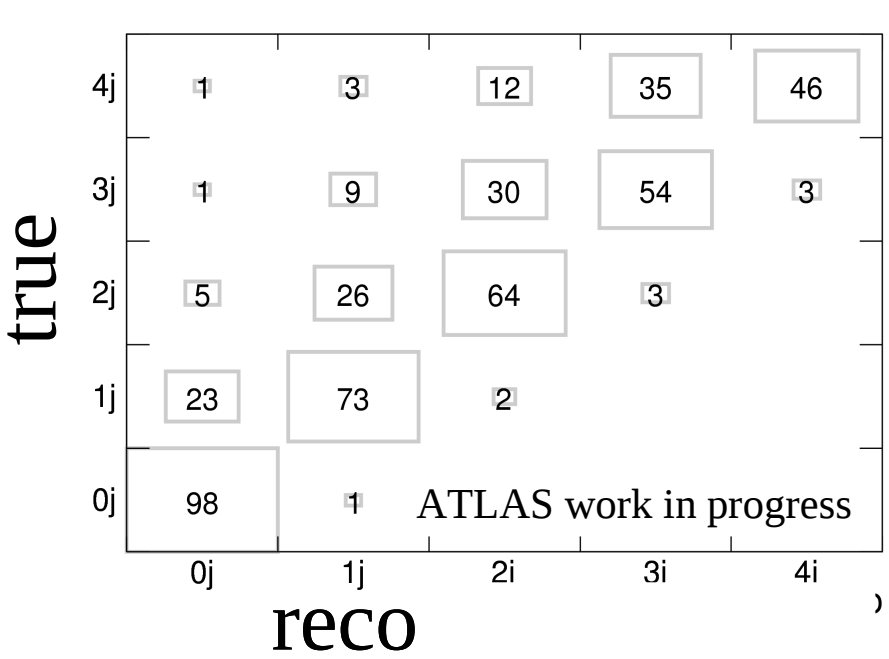
One “problem”: Need a migration matrix!

- Migration matrix from simulated events
  - Two “models”: **event generator** and **detector simulation**
- Closure test inside one model -> works fine! (binary bin effects: no out of acceptance correction)
- But cannot calculate mig matrix with data.. have to **rely on Monte Carlo**..

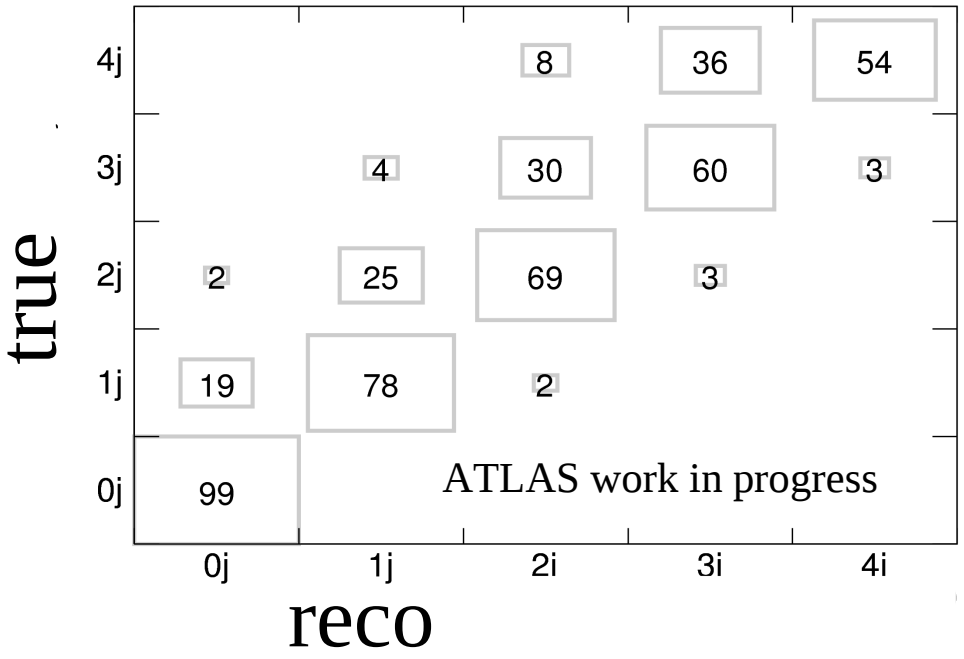


Comparison of Migration Matrices

- Alpgen



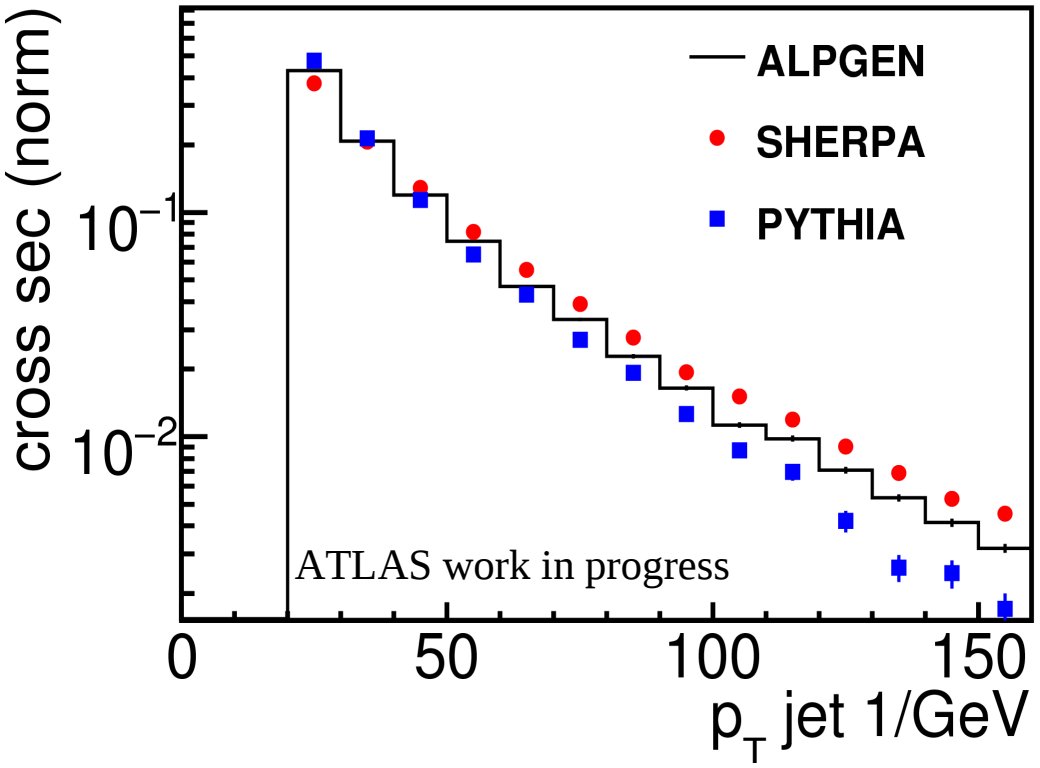
Sherpa



- Less migrations for Sherpa! Why?

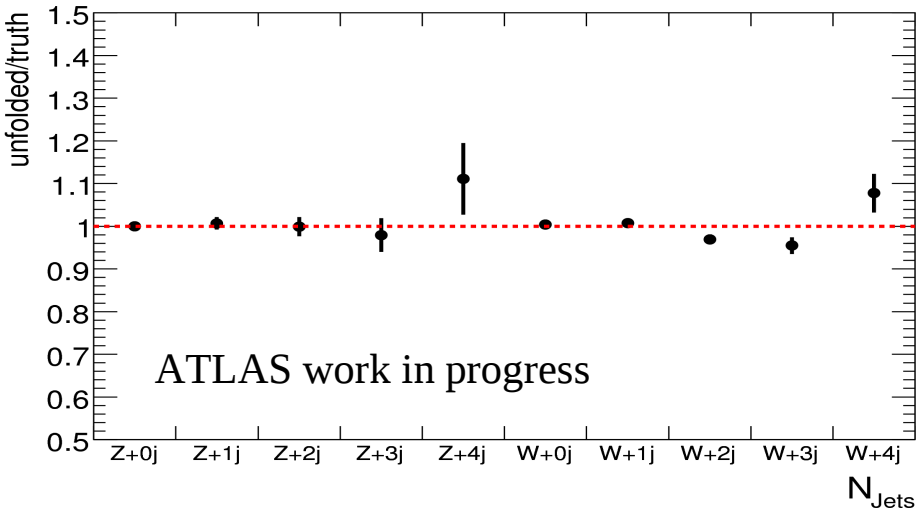
Jet pT: ALPGEN and SHERPA

- PT Sherpa larger than pT Alpgen-> differences understood!
- “hidden” physic effects reveal as detector effects -> take care!
- -> important to **compare data and MC** in jet pT distributions
- migration matrix has to reflect data

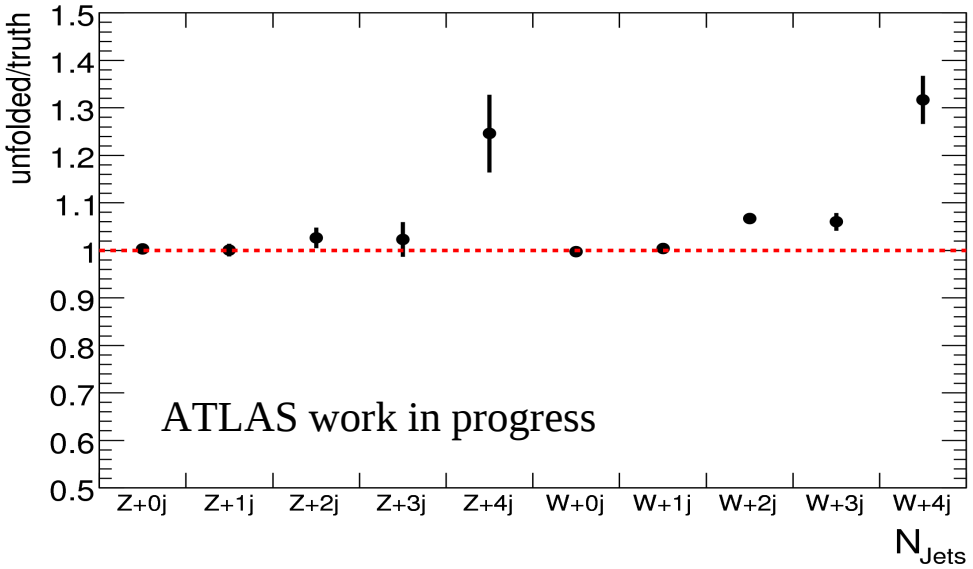


Consequences ALPGEN/SHERPA

- Alpgen with Sherpa



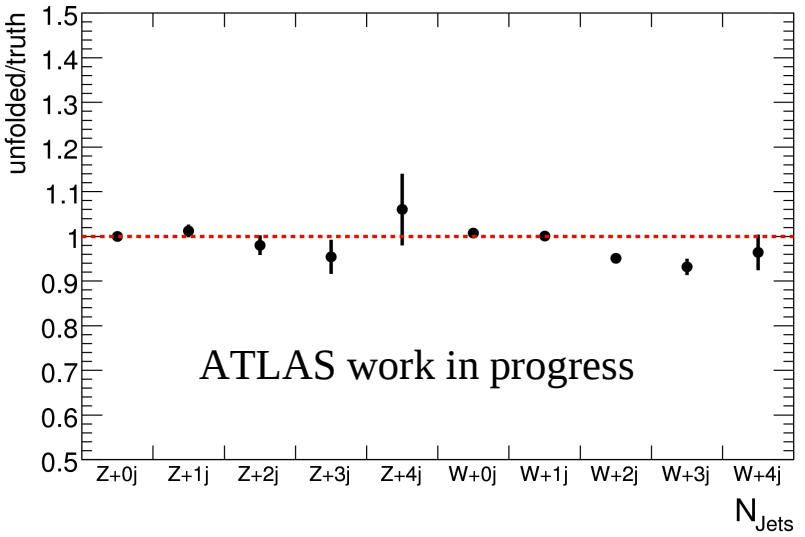
Sherpa with Alpgen



- Sherpa: less migrations -> Alpgen unfolded with Sherpa: too few events back to high jet multiplicities (and vice versa) -> but only **small effects**!!!!

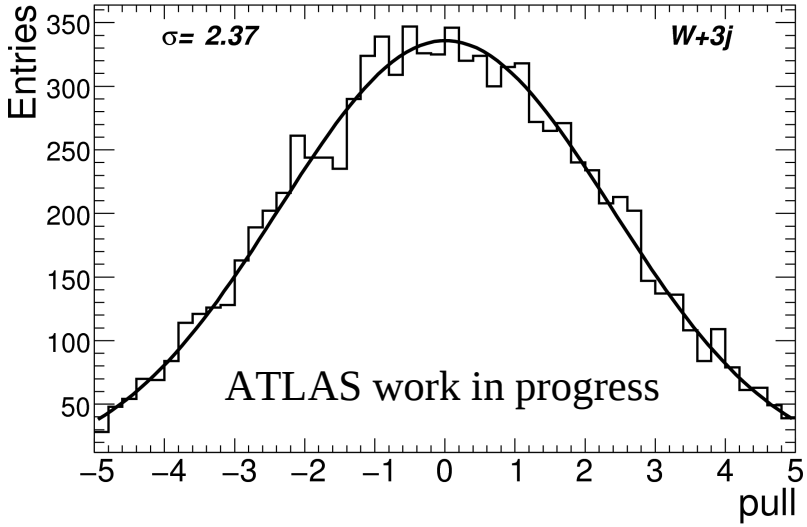
My Analysis

- No data available (at time of analysis ;-))
- Migration matrix as independent as possible:
  - Alpgen for “pseudo data” (with full detector simulation)
  - Sherpa for migration matrix (with fast detector simulation)
    -> both “models” different
- Very good agreement despite these differences!!!
- (implemented D´Agostini´s formulae in C++ routine -> cross checked with original FORTRAN code)

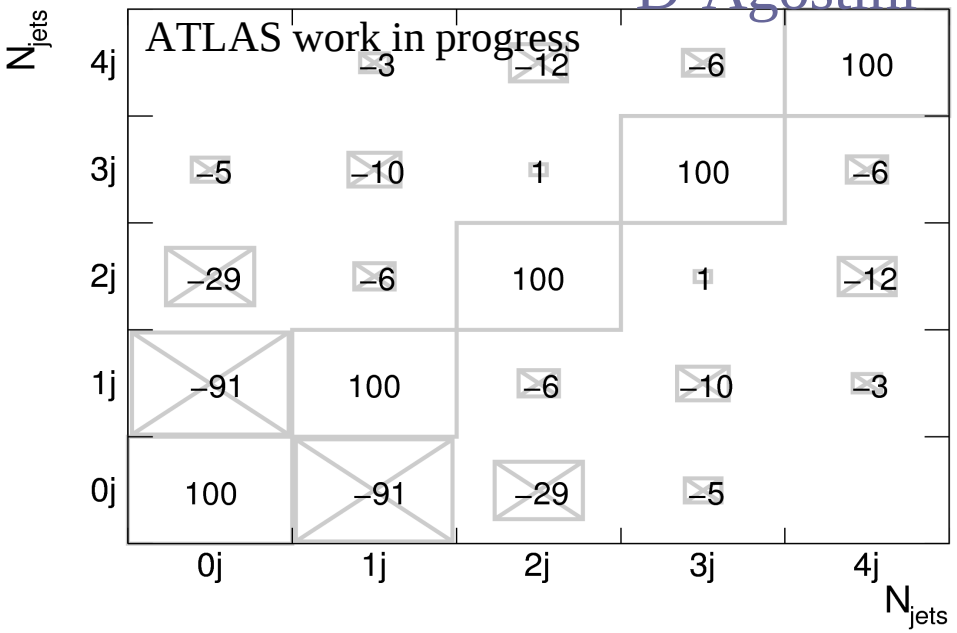


Error estimation -> always difficult ...

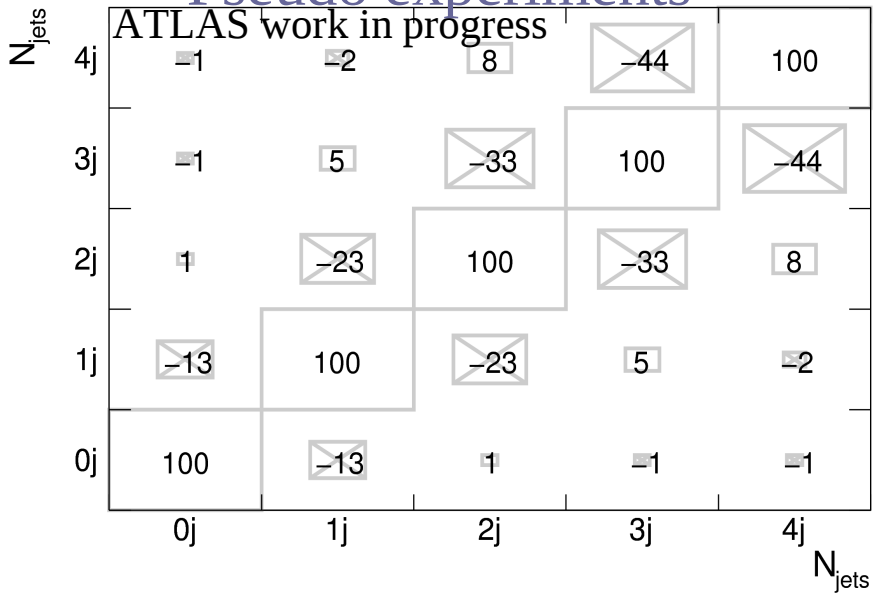
- Correlations, repeated cross talk between bins
- Pseudo experiments to test errors
  -> too small! Problem not gaussian?
  -> took errors from pseudo exp.



D´Agostini



Pseudo experiments



Summary

- Unfolding of W+jets and Z+jets distributions
- Factorize: event selection, background subtraction, unfolding, signal efficiency
- Unfold = only resort events of signal distribution
- Unfold W and Z separately
- Cannot avoid model dependency of migration matrix
  - Control plots MC <-> data important
  - “hidden” physics effects are revealed in mig matrix
- Took errors from pseudo experiments