Unfolding the method of “Unfolding” in HEP

Medha Chakraborty

Motivation

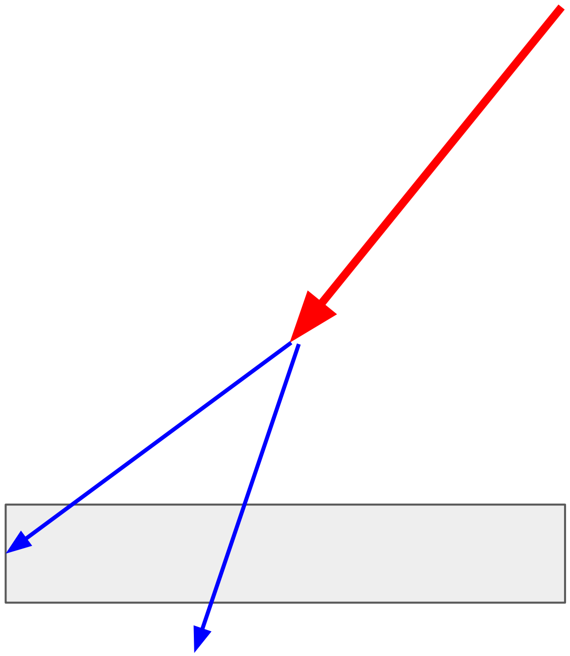
1. Particles passing through the detector, deposit charge/energy in them.
2. The energy or momentum of primary particles are reconstructed using the above measured quantities of secondary particles.
3. i) Detector resolution

   ii) Statistical fluctuations

   iii) Background

   iv) Efficiency

   Lead to bias in reconstructed energy or momentum of the primary

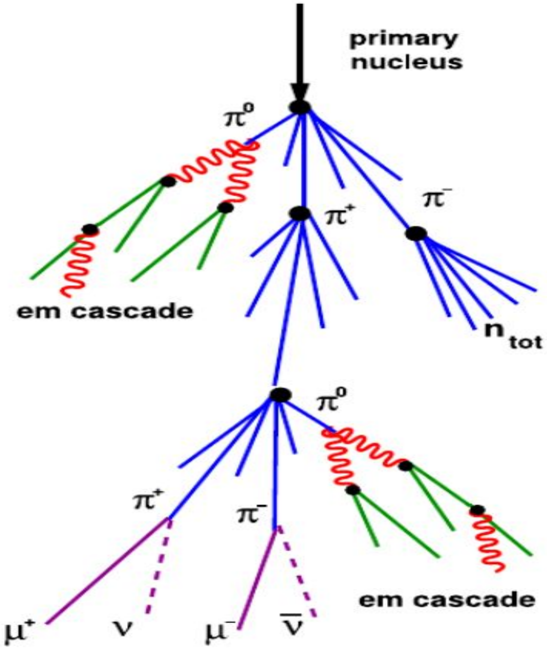


Examples

Accelerator

$$t\bar{t} \rightarrow W^+bW^-\bar{b} \rightarrow qq'bl\nu\bar{b}$$

Astroparticle



Reconstruction

Generate primaries in Monte Carlo (target quantities: Energy/ momentum)

↓

Secondaries produced through governing physics

↓

Passed through the detector response (get measurable quantities)

↓

Relate the measured quantities to the primary unknowns

We know the true and the reconstructed value of target quantities. The distributions of true and reconstructed quantities do not match!

Example

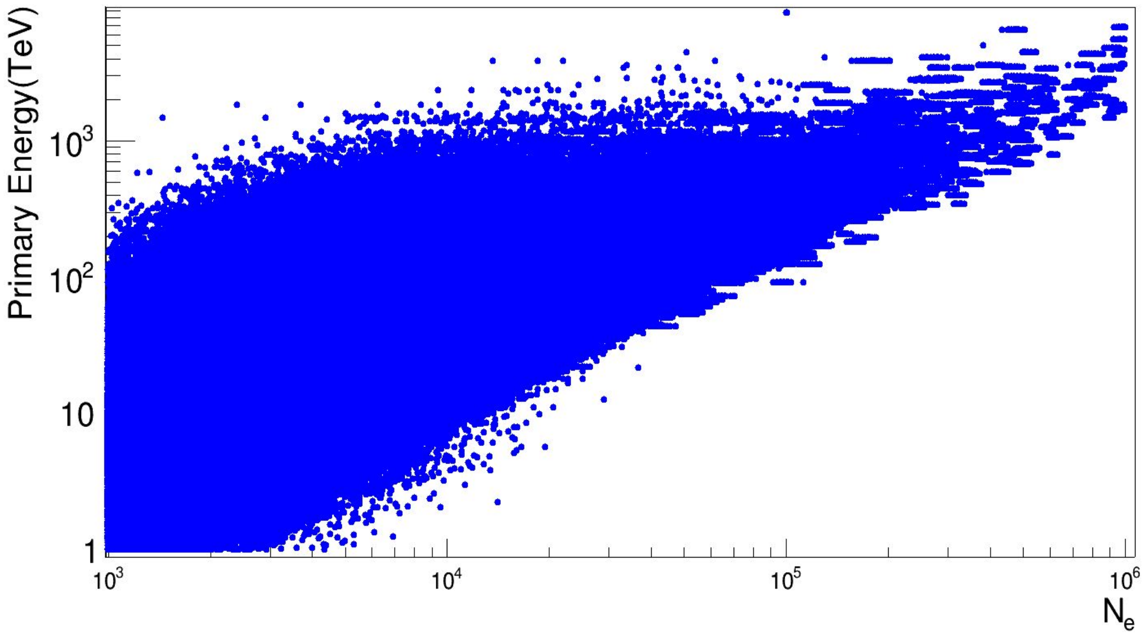
The energy of primary, reconstructed

$\log(E_T) = A \log(N_e) + c$, E_T: True energy, N_e: Number of secondaries

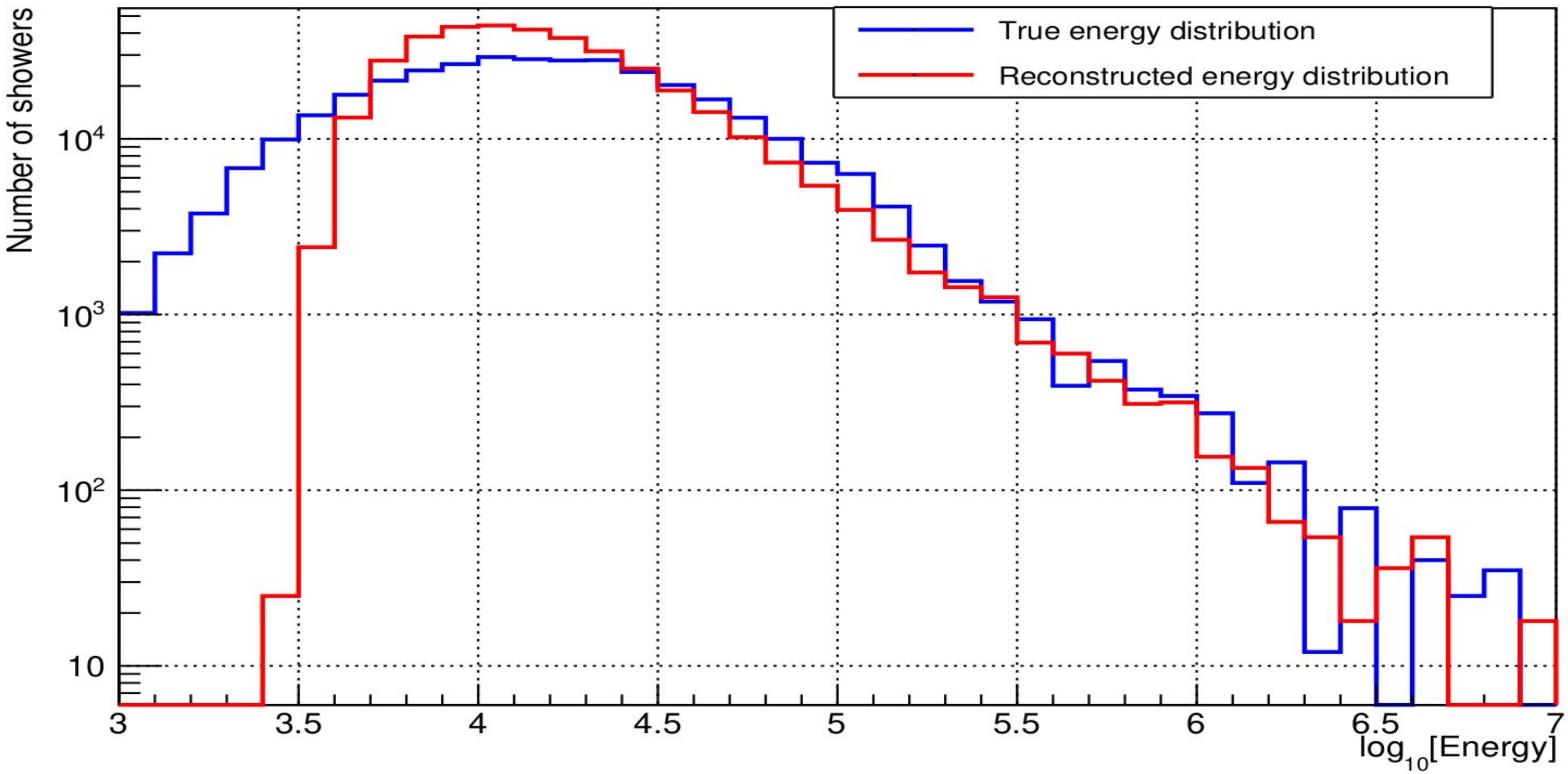
Fluctuation in number of events recorded.

N_e due to shower formation.

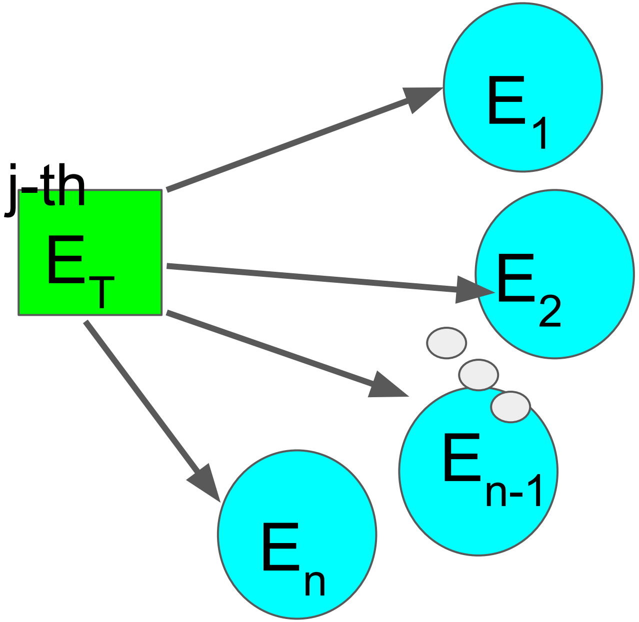
Detector resolution in measuring N_e.



Example (continued…) Mismatch!!! Migration!!!



Response matrix



Probability of true energy E_T in the j-th bin migrating to E_R in the i-th bin.

Smearing probability, calculated in the response matrix.

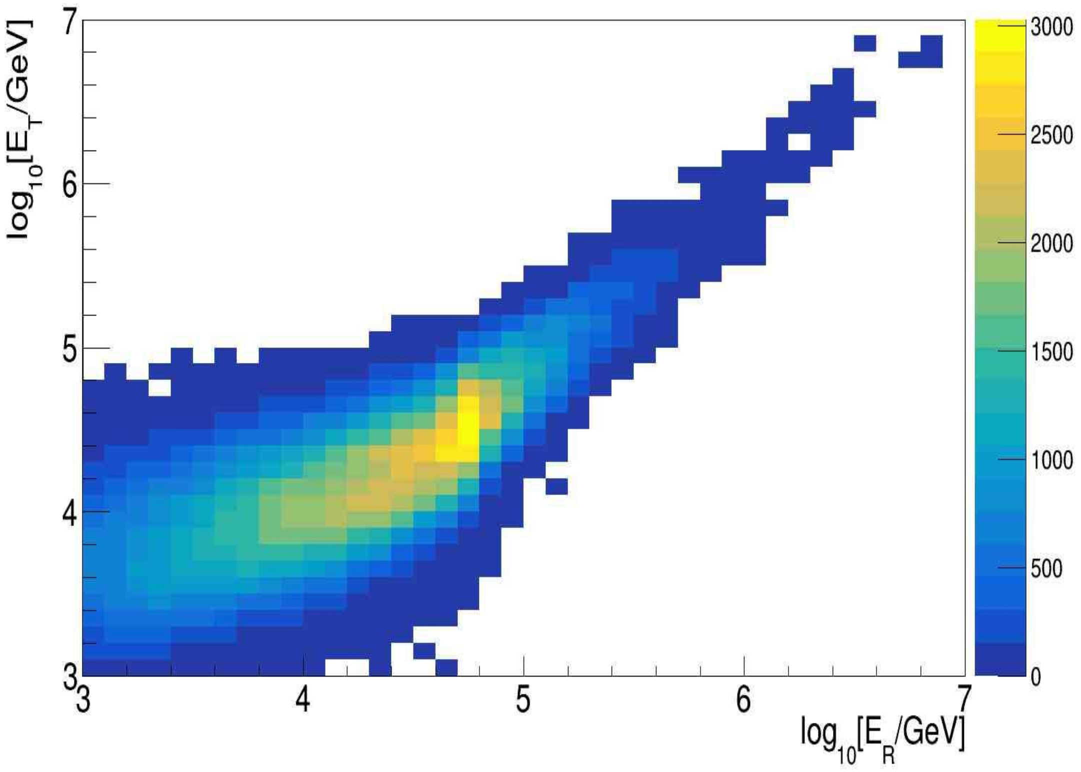
$$P_{ij}(E_R|E_T) = n_{ij}/\Sigma_i n_{ij}$$

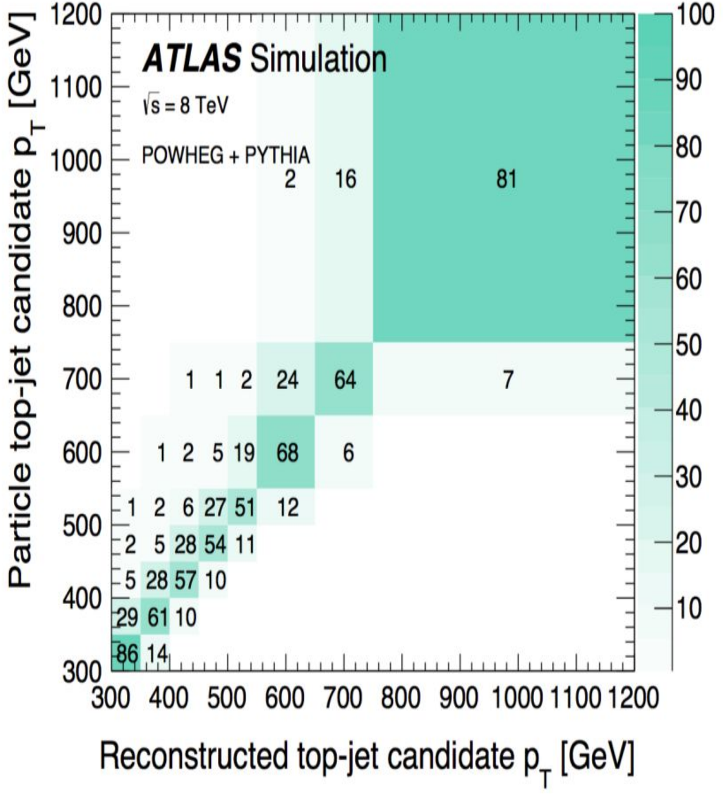
This give the response matrix

Here, $\Sigma_i n_{ij}$ is the number of events in j-th bin of true energy distribution

$$\mathbf{E_R} = P_{ij}(E_R|E_T)\mathbf{E_T}$$

Response matrix





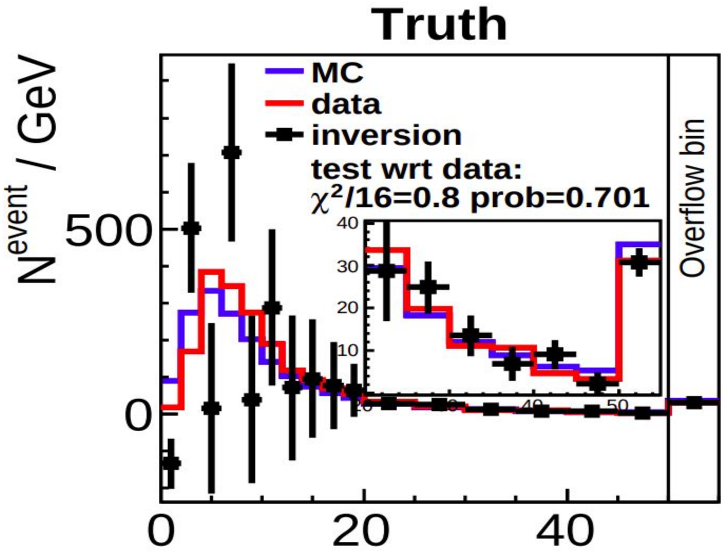
Unfolding

Response P_{ij} calculated.

$$\mathbf{E_R} = P_{ij}(E_R|E_T)\mathbf{E_T}$$

Inversion unfolding :

$$\mathbf{E_T} = P_{ij}^{-1}(E_R|E_T)\mathbf{E_R}$$



1. Inversion unfolding: Unable to handle large statistical fluctuations. Singularity of matrix and unstable results. Inconsistency in many cases
2. Bin-to-bin unfolding: Efficiency of each bin is calculated from simulation. Can’t handle migration effects accurately.

Shift to regularized methods.

Regularized unfolding

Iterative process.

Have a probabilistic approach.

Several methods:

1. Bayesian iterative method
2. Single value decomposition etc

Bayesian iterative method will be discussed further.

Bayesian unfolding

1. Assume an initial vector (Normalized true vector) : $\mathbf{p_0}$
2. The response or smearing matrix is calculated using simulated data.

   $P(E_R|E_T)$ is calculated. (Train)

3. We have to calculate the Unfolding matrix $P(E_T|E_R)$

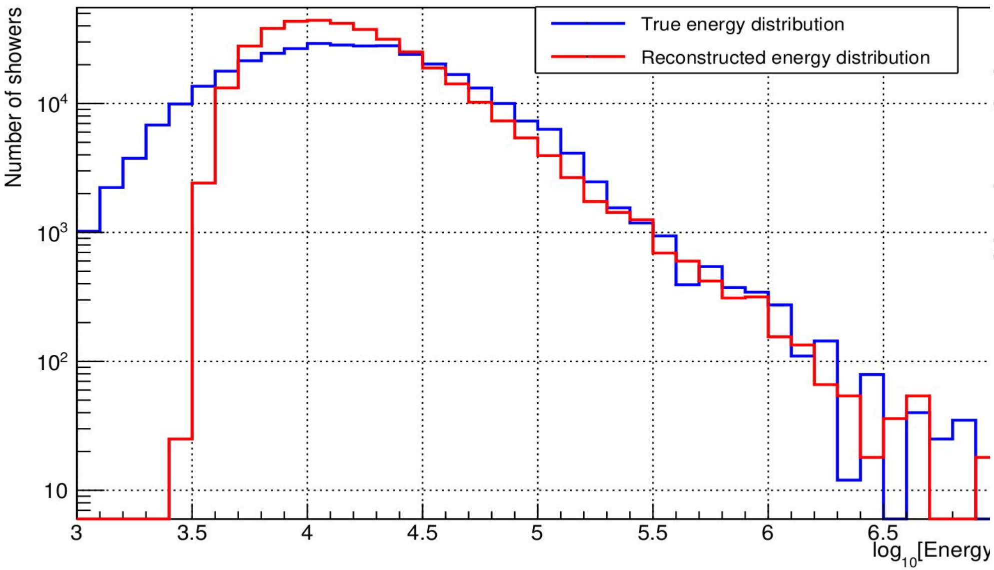
$$P(E_T|E_R) = \frac{\mathrm{P}(\mathrm{E}_R|\mathrm{E}_T).\mathrm{p}_0(\mathrm{E}_T)}{\Sigma_{E'_T}\mathrm{P}(\mathrm{E}_R|\mathrm{E}'_T).\mathrm{p}_0(\mathrm{E}'_T)}$$

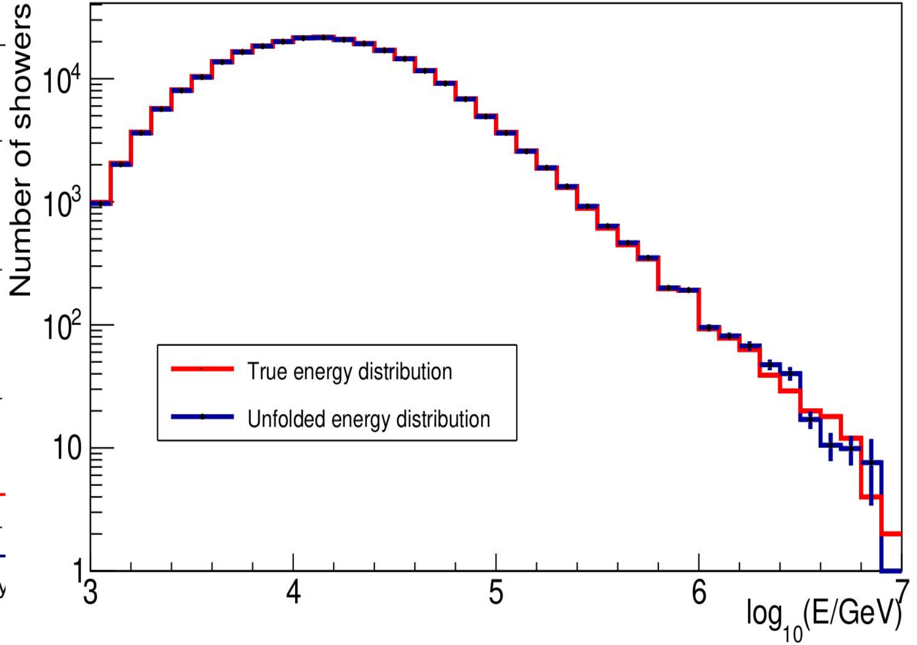
4. Using this, we find the true distribution corresponding to the given reconstructed distribution

$$\mathrm{E}_{T1} = \Sigma_{\mathrm{E}'_R}\mathrm{P}(\mathrm{E_T}|\mathrm{E_R})\mathrm{E}'_\mathrm{R}$$

5. Using the above, $\mathbf{p_0}$ is calculated again and the iterations go on till convergent solutions are obtained for unfolded distribution.

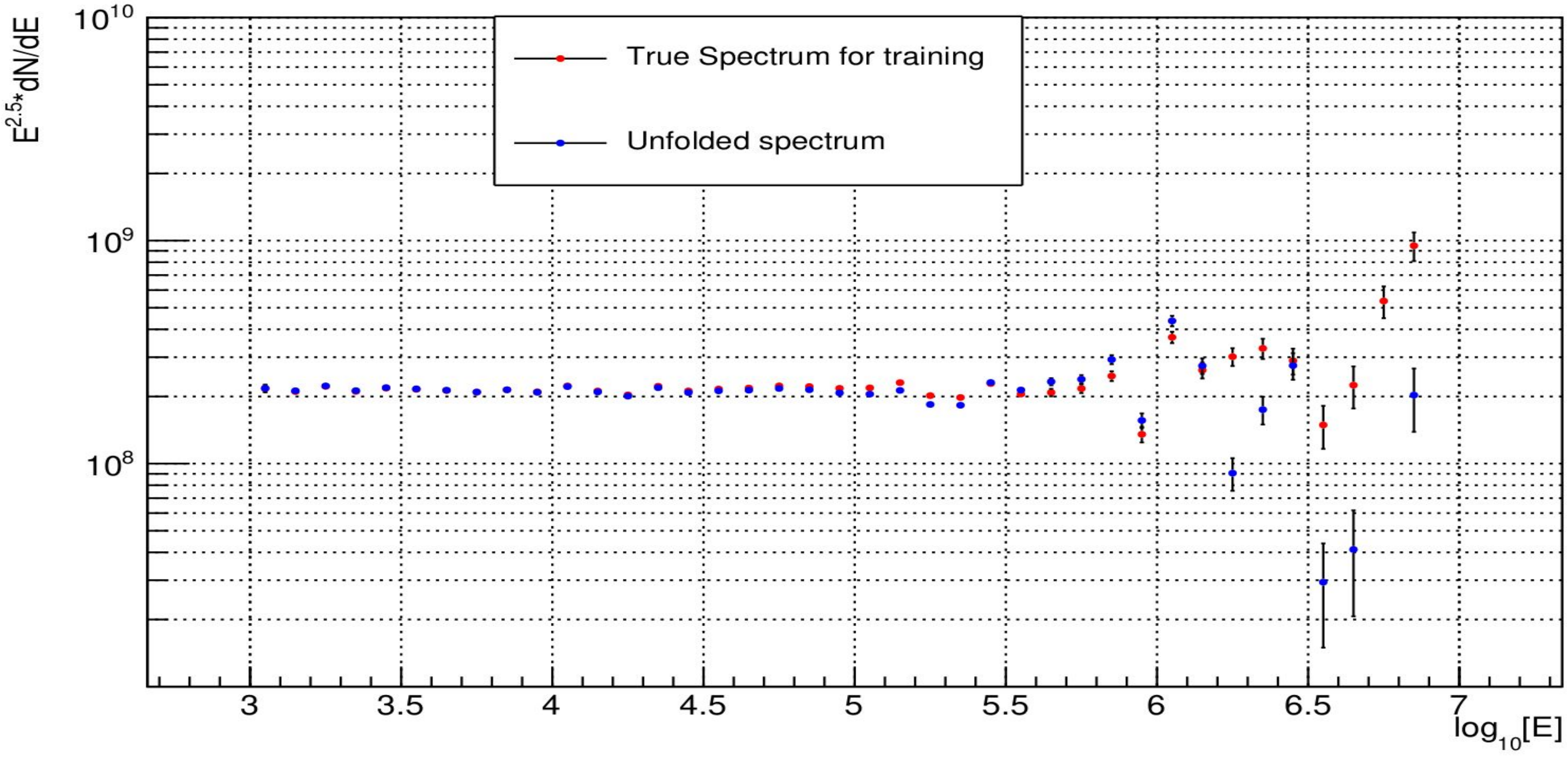
Example 1: (RooUnfold package was used for unfolding)





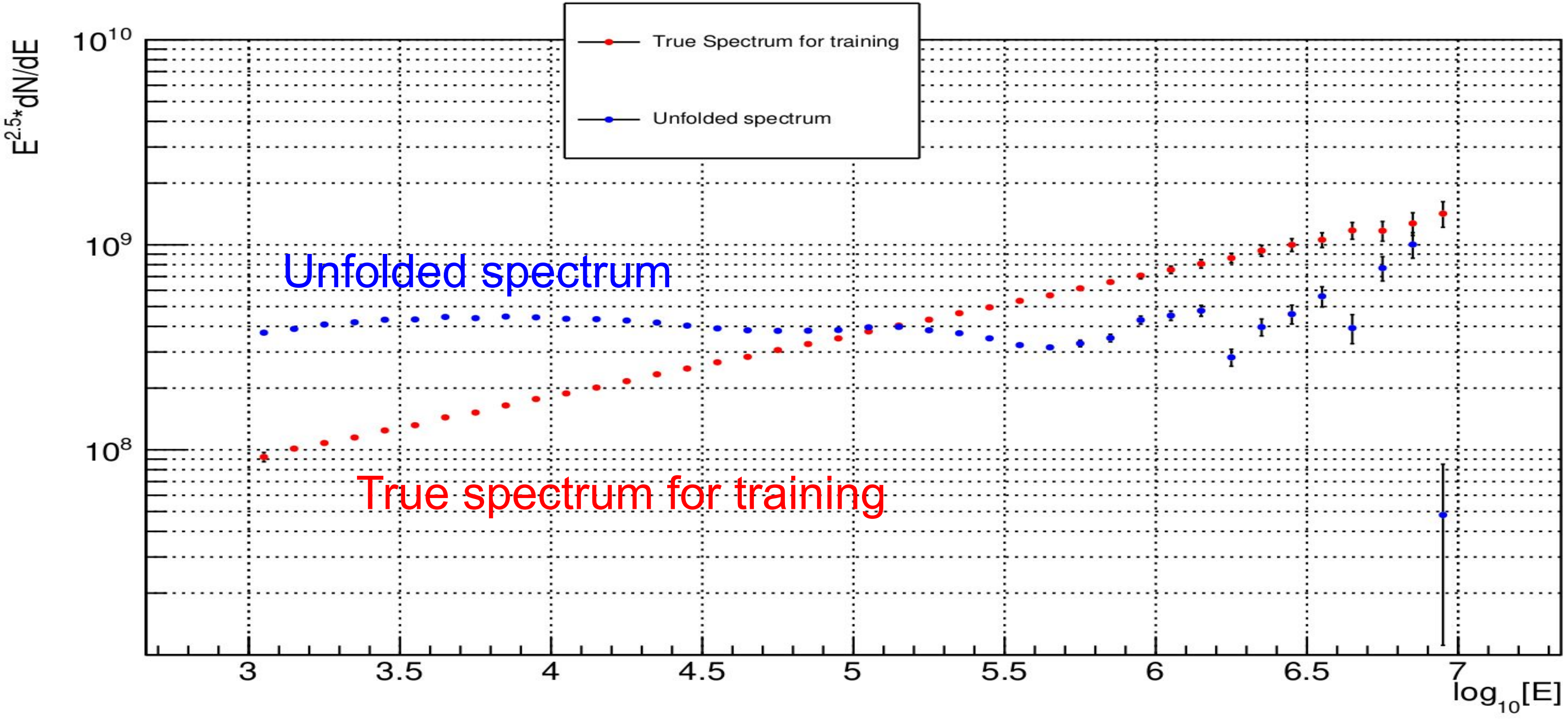
The simulated data set was divided into two parts: One for training (response), the other was used for testing (unfolding and comparing with what is expected)

Example 1: Energy spectrum

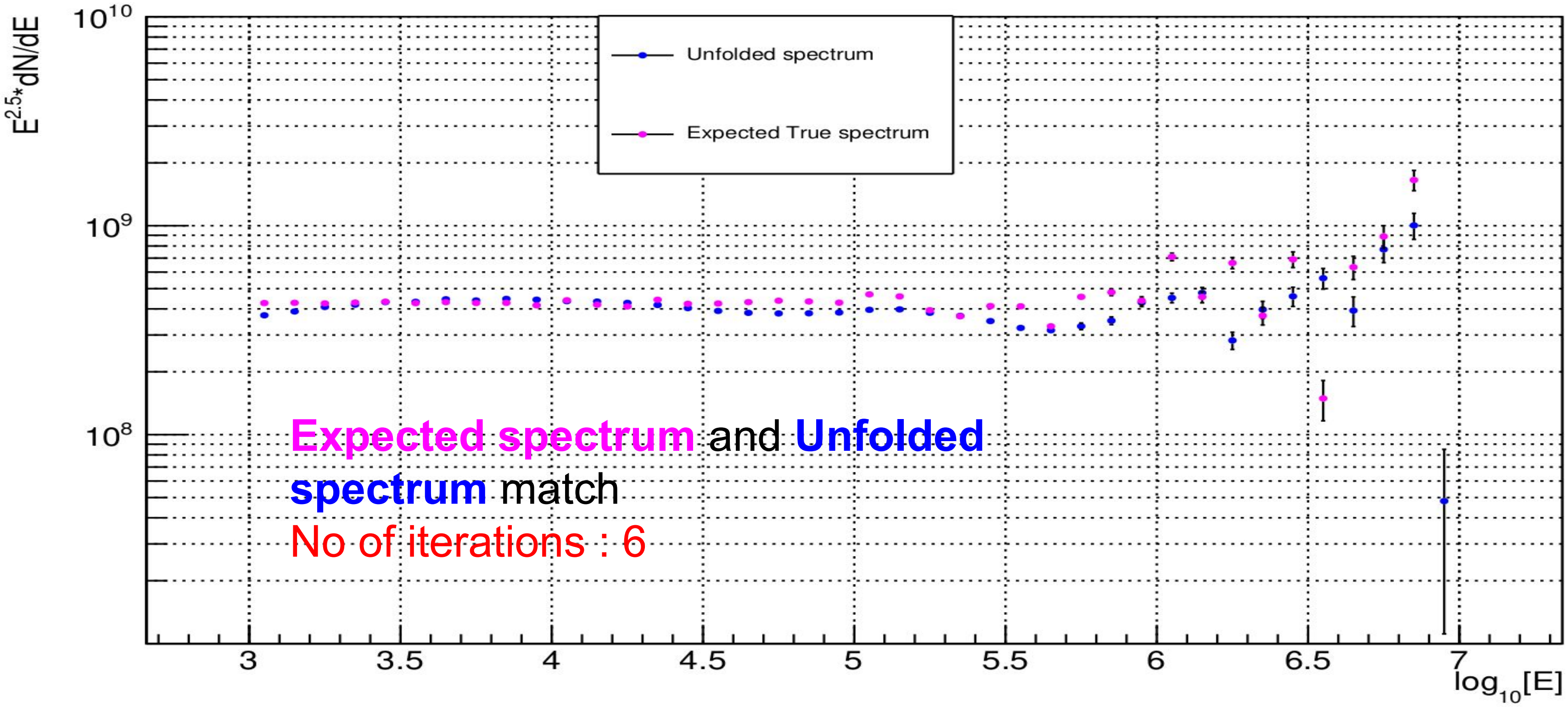


No of iterations: 2

Example 2: In the previous data set, the training and testing distributions are similar, here we demonstrate with dissimilar distributions



Example 2: In the previous data set, the training and testing distributions are similar, here we demonstrate with dissimilar distributions



Expected spectrum and Unfolded spectrum match

No of iterations : 6

Setting the number of iterations

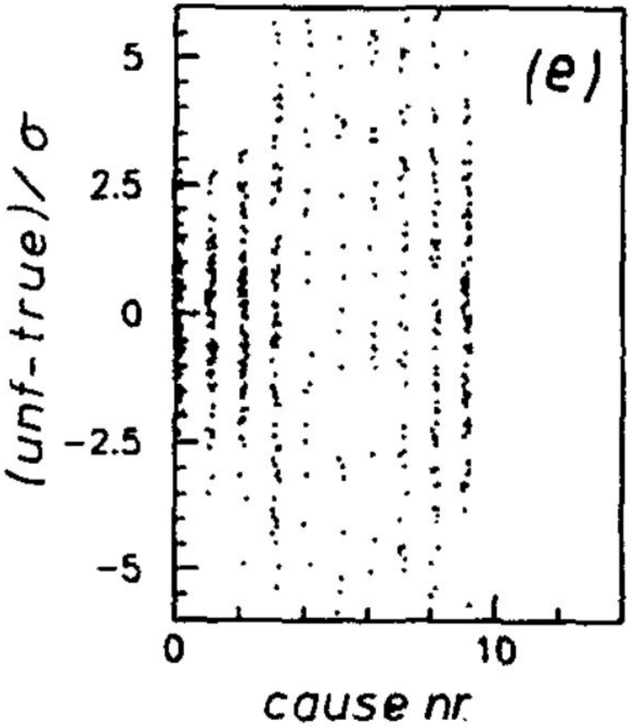
Regularization parameter

The best possible initial guess guided by physics. In RooUnfold, the initial guess is taken as training distribution.

$\Delta\chi^2$ in successive iterations is less than a certain value (0.01, to be determined by user).

A convergence criteria is to be set from tests on simulation.

Usually converges in few iterations (~4).



Uncertainities due to unfolding

1. Initial guess of $\mathbf{p_0}$
2. Error in response matrix: Systematic: Determination of response matrix from simulation. Statistical : Less no of events

$$\mathrm{n}(\mathrm{C_i}) = \Sigma_{\mathrm{j}=1}^{\mathrm{n_E}} M_{ij} \mathrm{n}(\mathrm{E_j})$$

Error propagation matrix calculated as,

$$\frac{\partial \hat{n}(\mathrm{C}_i)}{\partial n(\mathrm{E}_j)} = M_{ij} + \frac{\hat{n}(\mathrm{C}_i)}{n_0(\mathrm{C}_i)} \frac{\partial n_0(\mathrm{C}_i)}{\partial n(\mathrm{E}_j)} - \sum_{k=1}^{n_\mathrm{E}} \sum_{l=1}^{n_\mathrm{C}} \frac{n(\mathrm{E}_k)\epsilon_l}{n_0(\mathrm{C}_l)} M_{ik} M_{lk} \frac{\partial n_0(\mathrm{C}_l)}{\partial n(\mathrm{E}_j)}$$

$$V(\hat{n}(\mathrm{C}_k), \hat{n}(\mathrm{C}_l)) = \sum_{j,s=1}^{n_\mathrm{E}} \sum_{i,r=1}^{n_\mathrm{C}} \frac{\partial \hat{n}(\mathrm{C}_k)}{\partial P(\mathrm{E}_j|\mathrm{C}_i)} V(P(\mathrm{E}_j|\mathrm{C}_i), P(\mathrm{E}_s|\mathrm{C}_r)) \frac{\partial \hat{n}(\mathrm{C}_l)}{\partial P(\mathrm{E}_s|\mathrm{C}_r)}$$

Efficiency and background

Efficiency : True E satisfies the quality cut but reconstructed E does not

Background : True E does not satisfy the quality cut but reconstructed E does.

THANK YOU

References

1. G. D’Agostini, A multidimensional unfolding method based on Bayes’ theorem, Nucl. Instrum. 256 Meth. A362 (1995) 487 (Primary)
2. https://cds.cern.ch/record/2229001/files/ATL-PHYS-PROC-2016-189.pdf
3. Arxiv 1611.01927v2
4. RooUnfold package docs: http://hepunx.rl.ac.uk/adye/software/unfold/RooUnfold.html . (errors in unfolding)
5. Statistical methods for data analysis by Luca Lista