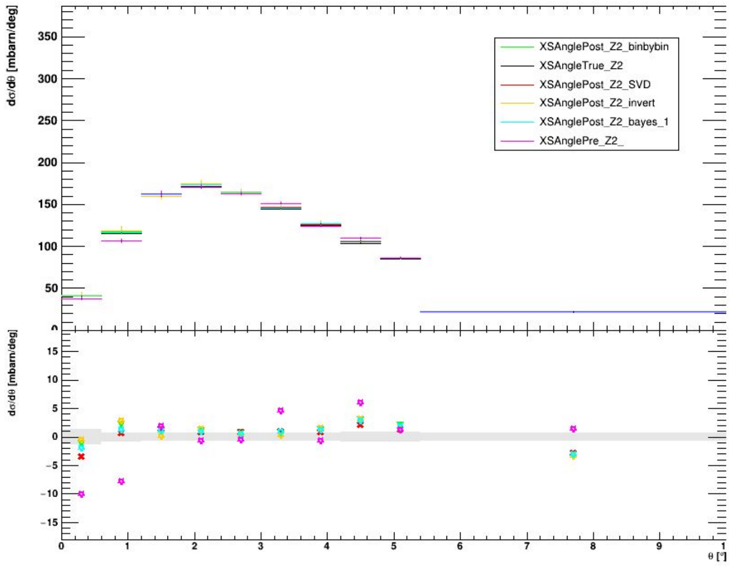
Unfolding Validation Tests

FOOT physics meeting 21/05/2024

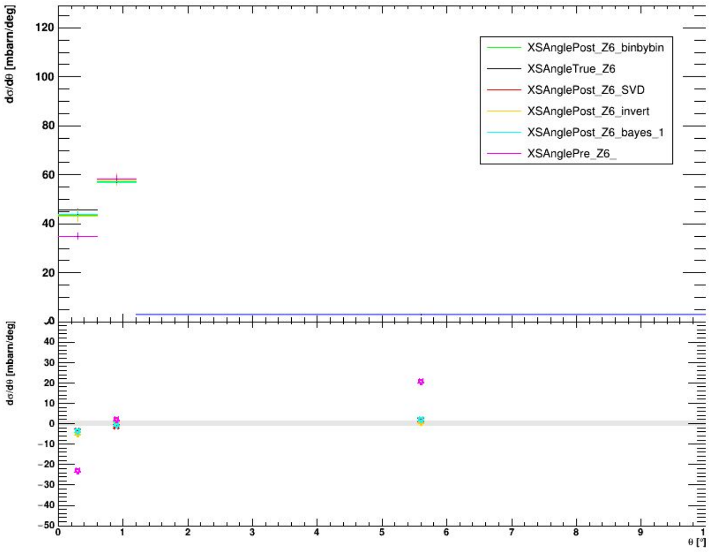
Alberto Mengarelli INFN BO
Riccardo Ridolfi UniBO & INFN

Unfolding Method Comparison

Z=2



Z=6

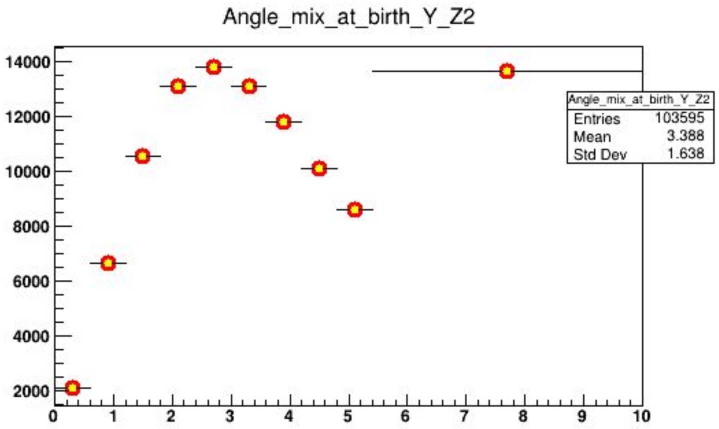


Different methods behave the same way → Bayes Iterative method chosen

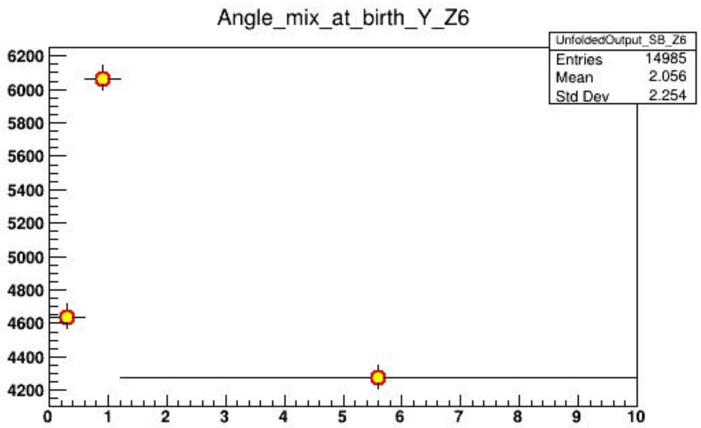
Closure test

First trivial check to perform as starting point: unfolding the same input distribution used to construct the response matrix etc..This must give a perfect closure.

Z=2



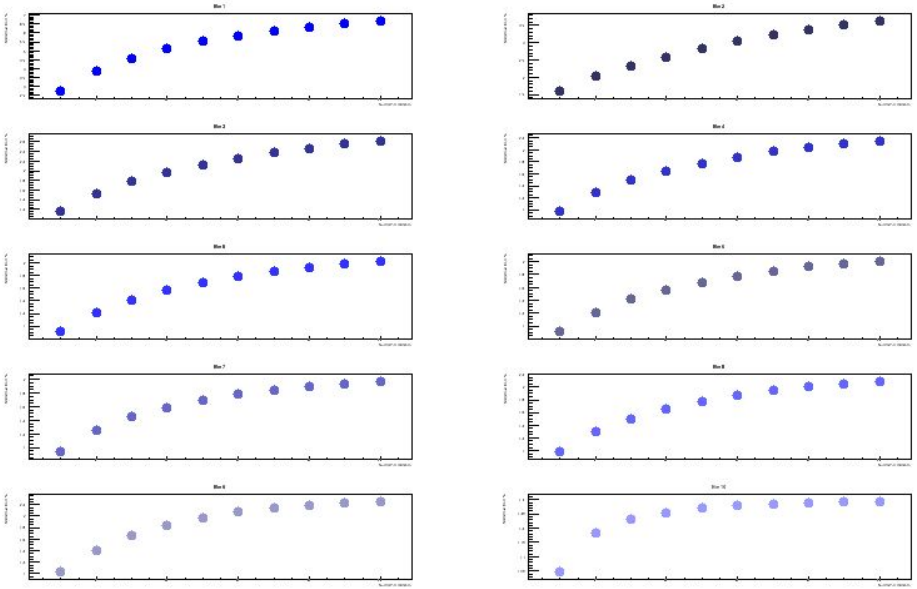
Z=6



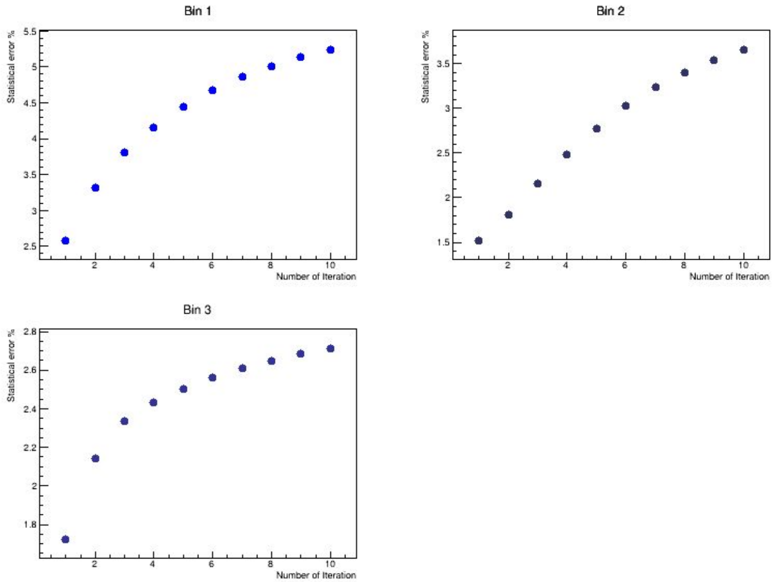
Studies on the number of iterations: Statistical uncertainty

The degree of convergence of the iterative procedure is checked by comparing each iteration to the previous one. In particular, the parameters taken into account are the bin-wise statistical error (it increases with the number of iterations, reaching a sort of plateau)

Z=2



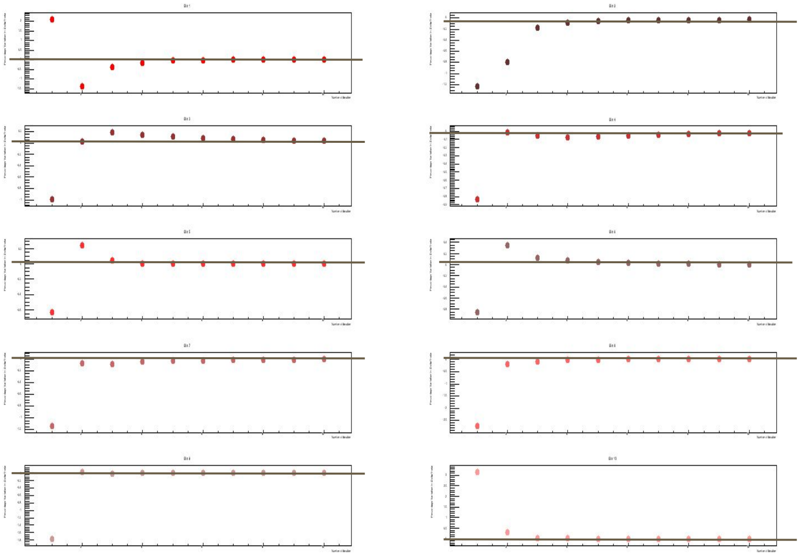
Z=6



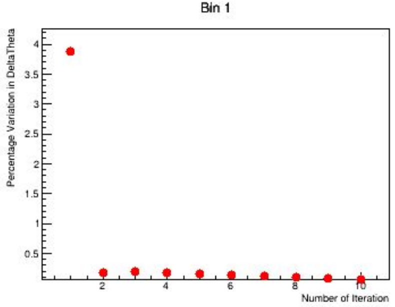
Studies on the number of iterations: bin residuals

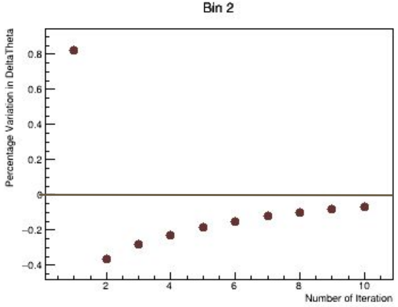
bin-wise residuals (the bin-by-bin difference between the unfolded distribution at the i-th iteration and the unfolded distribution at the $i-1$-th iteration, which must converge to 0).

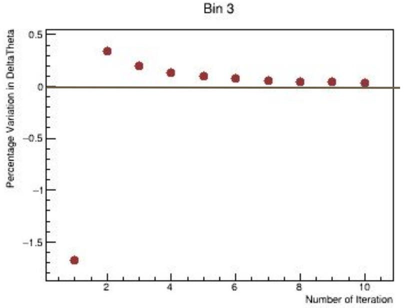
Z=2



Z=6







Studies on the number of iterations: Average correlation

An optimal choice of the regularization parameter is the one that minimizes the average correlation factor:

$$\rho_{\mathrm{avg}} = \frac{1}{M_x}\sum_{j=1}^{M_x}\rho_j \,.$$

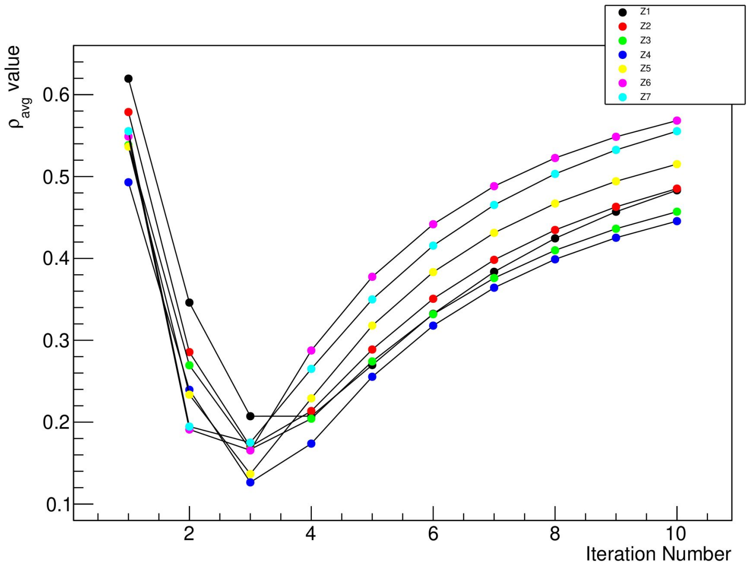
where ρ_j global correlation coefficient of bin j is defined as

$$\rho_j = \sqrt{1-\left((\boldsymbol{V_{xx}})_{jj}\left(\boldsymbol{V_{xx}}^{-1}\right)_{jj}\right)^{-1}} \,.$$

M_x : ndof ; V_{xx} : Cov. matrix

$$\rho_{\mathrm{avg}} = \frac{1}{N_{\mathrm{bins}}}\sum_{i=1}^{N_{\mathrm{bins}}}\rho_i \qquad \rho_i = \sqrt{1-\left[C_{ii}\left(C^{-1}\right)_{ii}\right]^{-1}} \,.$$

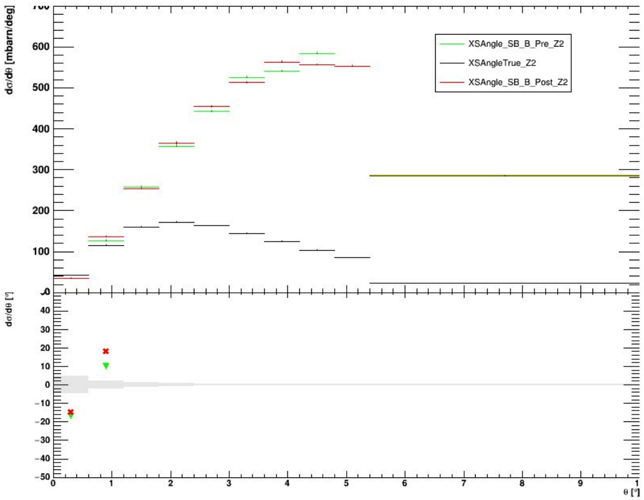
S. Schmitt, Data Unfolding Methods in High Energy Physics, EPJ Web Conf. 137 (2017) 11008, ed. by Y. Foka, N. Brambilla and V. Kovalenko, arXiv: 1611.01927



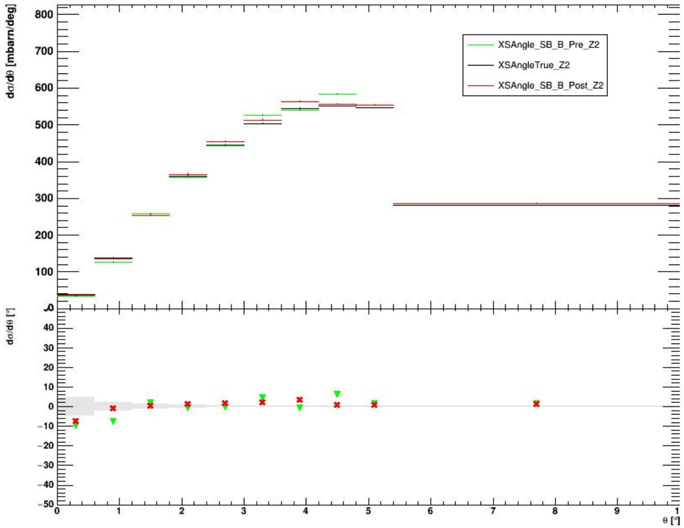
Stress tests

Due to the specific choice of the Monte Carlo sample for the training of the unfolding, it is necessary to check whether this choice could introduce a bias via the unfolding. To do this check the MC reweighting is required, in order to change the shapes of the distributions and get a varied distribution used as pseudo-data.

Only pseudo-data reweighted



Both pseudo-data and truth reweighted



Summary & next steps

- Standard sanity check on the unfolding machinery has been presented
- No unexpected behavior of the machinery has been found
- Performing checks on the dependence of different figures of merit on the number of iterations, 3 iterations could be chosen as the optimal value for **iterative bayesian unfolding**

...About systematics of the unfolding method

- “External”: difference compared to an alternative method
- “Internal”: difference between using 3 iteration and 4

expected to be at subpercent level

Table 1. Data tests performed on various unfolding methods.

	Y_{unf}	$\chi^2_A/N_{d.f.}$	$\mathrm{Prob}(\chi^2_A, N_{d.f.})$
expectation	4584	$N_{d.f.}$ / $N_{d.f.}$	0.5
bin-by-bin	4521	34.7 / 0	n.a.
matrix inversion	4584	0 / 0	n.a.
template fit	4572	12.0 / 16	0.74
constrained template fit	4584	12.0 / 16	0.74
Tikhonov τ = 0.0068	4584	13.4 / 16	0.64
Tikhonov τ = 0.012	4584	15.0 / 16	0.52
EM method n = 0	4537	5069 / 0	n.a.
EM method n = 20	4585	5.9 / 0	n.a.
EM method n = 100	4584	4.2 / 0	n.a.
EM method n = 1000	4584	3.9 / 0	n.a.
IDS n = 1	4584	76.8 / 0	n.a.
IDS n = 3	4584	26.1 / 0	n.a.
IDS n = 10	4584	8.0 / 0	n.a.
IDS n = 30	4584	4.9 / 0	n.a.

backup