Tunfold Studies for Blindtest
Status Report

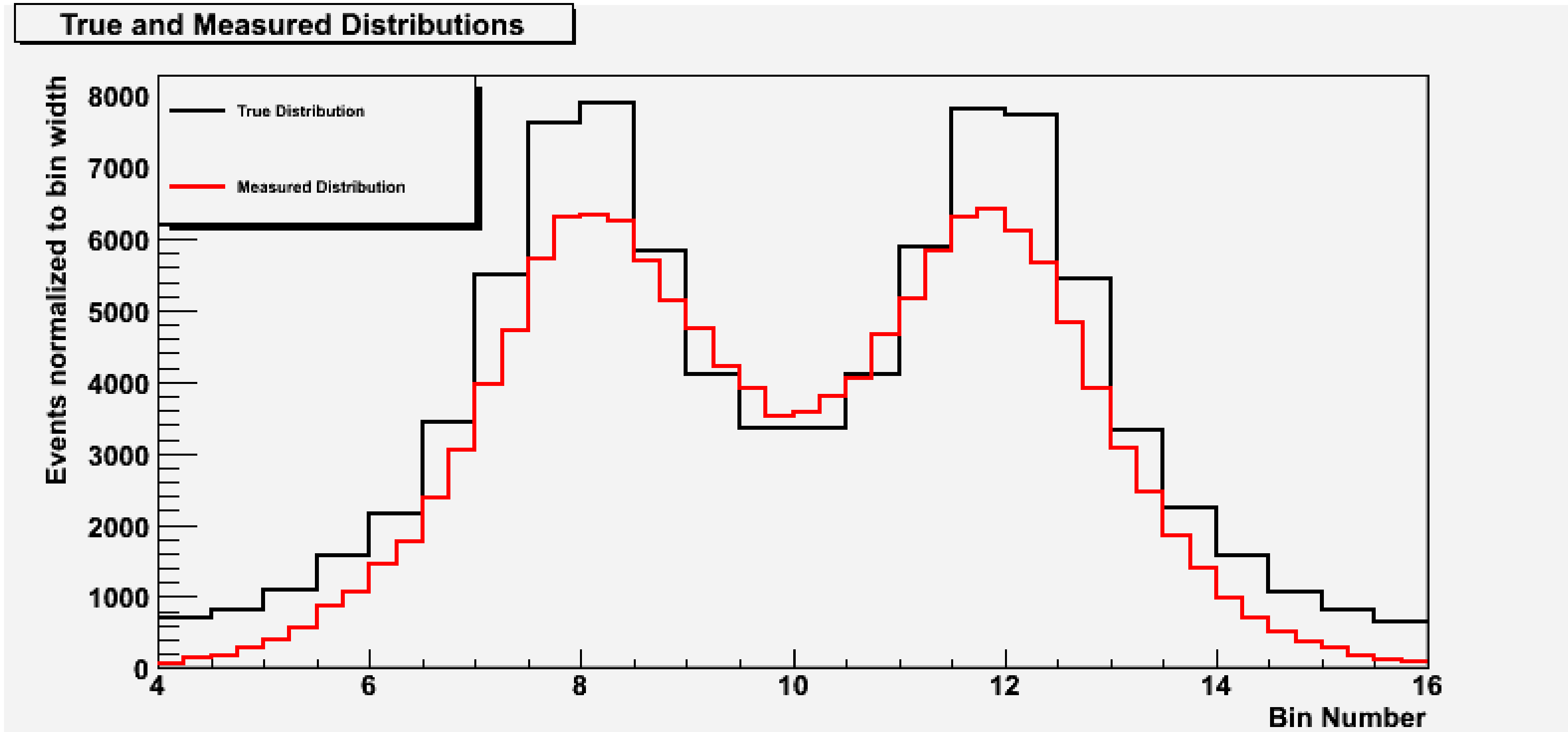
Cipriano, Pedro; Flucke, Gero

Unfolding meeting
9 June 2011

Outline

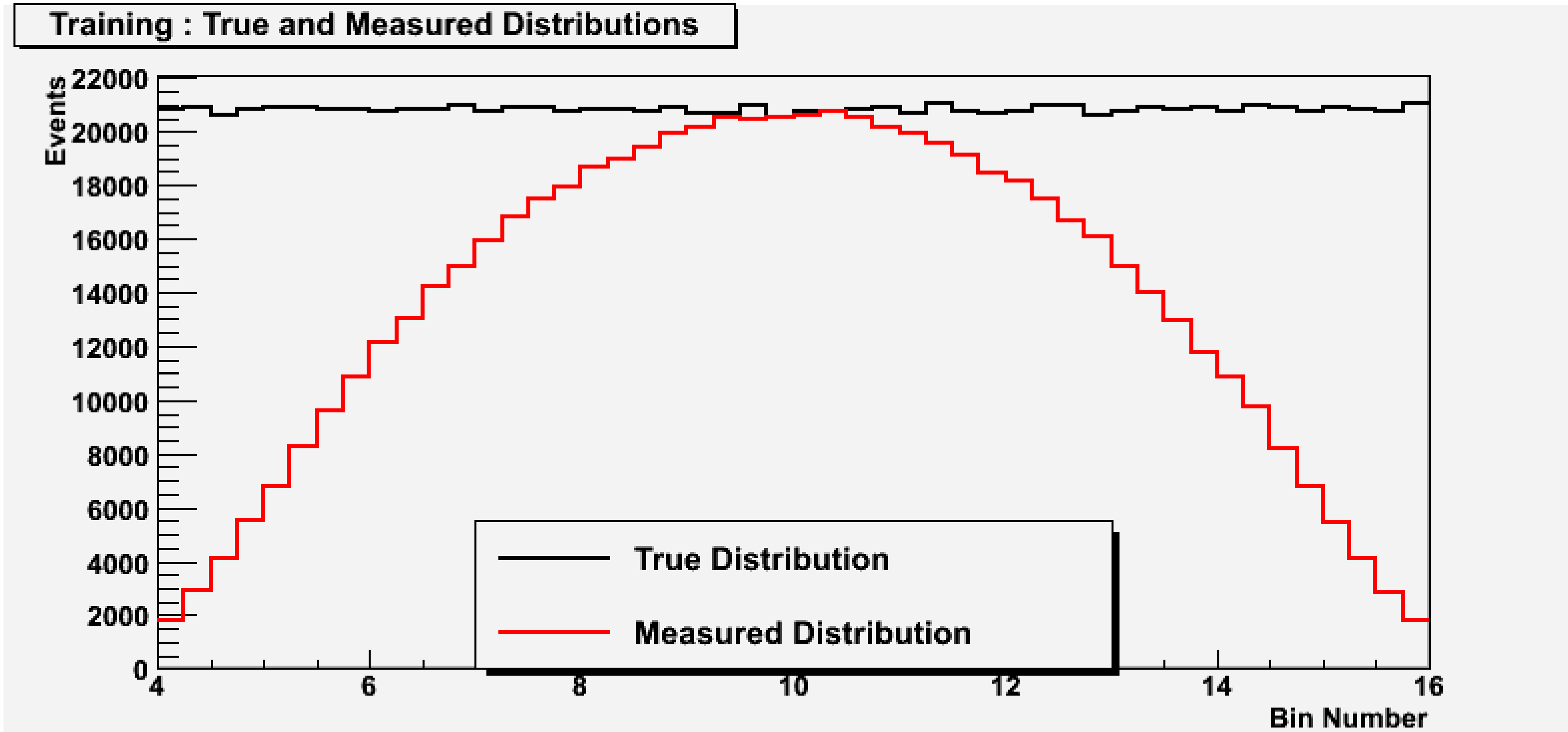
- Unfolding Setup
  - True and Measured distribution
  - Response matrix
  - Unfolding with Tunfold for smearing = bin width
  - Correlations
- Methods
- TUnfold
  - Brief explanation on the different Modes
- Quality of the result
  - Correlations
  - Pull
  - Chi^2
- Comparing Different Methods
- Conclusion
- Plans

True and measured



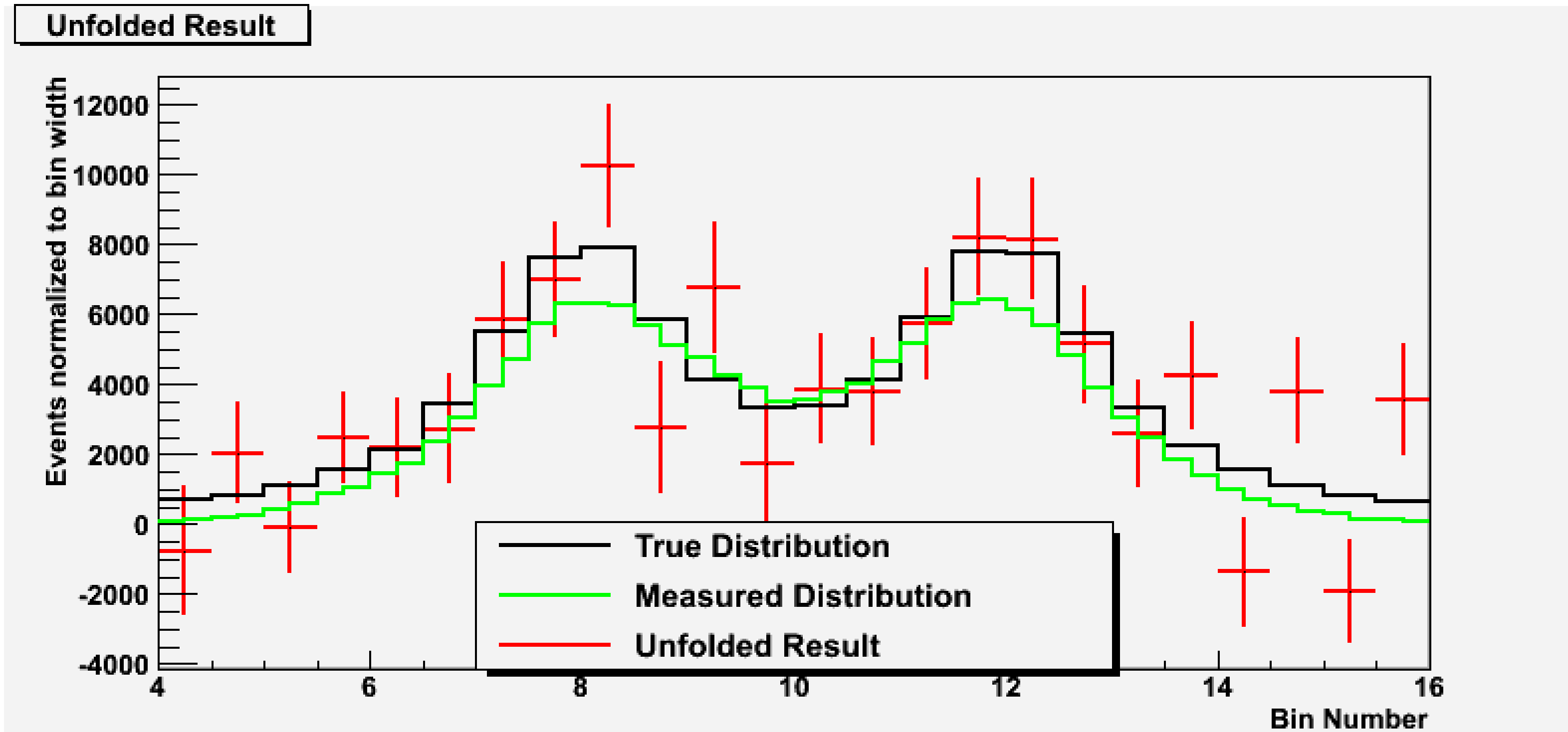
- True obtained from 10^5 dices of a Double BreitWigner function
- Measured obtained from true by applying to each event a Gaussian smearing with a width = 0.5, no bias and an acceptance = $1 - ((x - 10)^2 / 36)$
- Number of Measured bins = 48 (bin width = 0.25)
- Number of True bins = 24 → bin width = 0.5 (Nikolai proposed bin width = 0.4)

Response Matrix



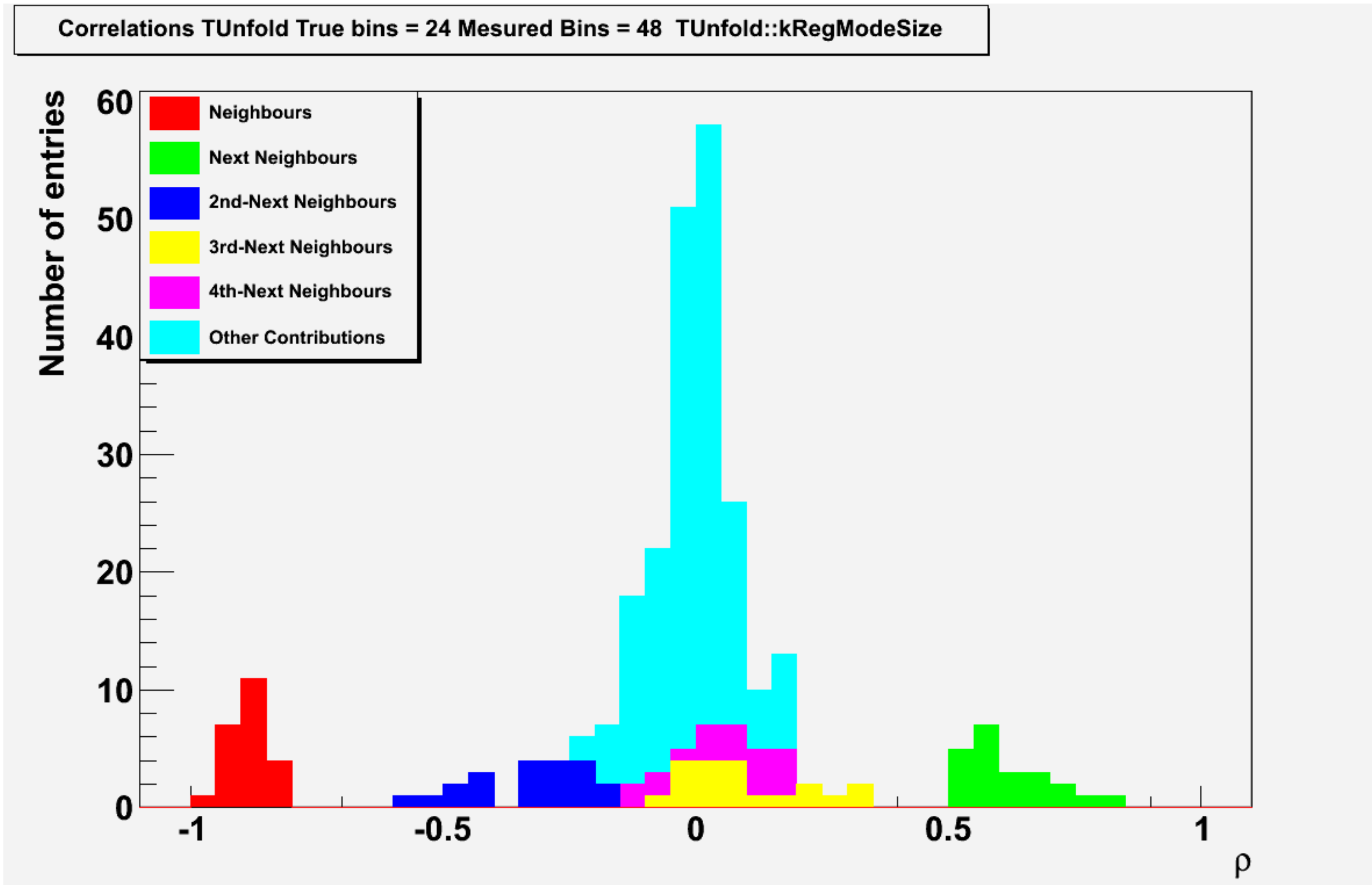
- True obtained from 10^6 dices of a flat function
- Smearing, bias, acceptance and binning as set for the measurement

Unfolding with Tunfold for smearing = bin width



- Software: RooUnfold 1.0.3
- Unfolding Algorithm: TUnfold
- Regularization mode : KregModeSize

Bin Correlations



- Neighbour bins almost 100% correlated
- Sign of R alternates for further neighbours
- The absolute value of R decreases

Methods Summary

- Testing:
  - Other smearing values
  - Different regularization modes

- Software: RooUnfold 1.0.3
- Unfolding Algorithm: TUnfold
- Regularization modes : KregModeSize, KregModeDerivative and KregCurvature
- Number of unfolding tries = 10^4
- Number of measured events = 10^5
- Number of training events = 10^6
- Number of Measured bins = 48
- Number of True bins = 24
- Smearing values = {0.3, 0.5, 0,7}
- Displacement = 0.0
- Acceptance : $1 – ((x – 10)^2 / 36)$

Tunfold: Methods

- TUnfold solves the inverse problem:
- $chi^2 = (y-Ax)^T Vyx^{-1} (y-Ax) + r^2 (L(x-x0))^T L(x-x0) + \lambda \, sum\_i(y(i) - Ax(i))$
  - y: vector of measured values
  - Vyx: covariance matrix for y
  - A: response matrix
  - x: unknown underlying distribution
  - r: parameter, defining the regularisation strength
  - L: matrix of regularisation conditions
  - x0: bias distribution
  - λ: lagrangian multiplier
  - y(i): one component of the vector y
  - Ax(i): one component of the vector Ax

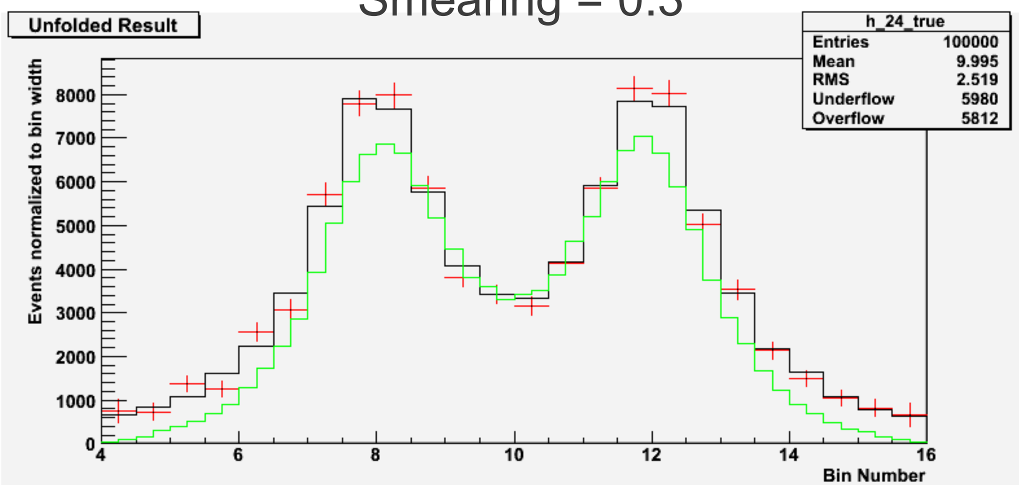
  - The λ term was not used in this study

The different TUnfold regularization modes

- 4 different regularization modes are available on Tunfold:
  - KRegModeNone → No regularization
  - KRegModeSize
    - minimize the size of (x-x0)
    - usally is the one present in the literature
    - the bias x0 usually is not present
  - KregModeDerivative
    - minimize the 1st derivative of (x-x0)
    - Create positive correlations between bins
  - KregModeCurvature
    - minimize the 2nd derivative of (x-x0)
    - Create positive correlations between bins

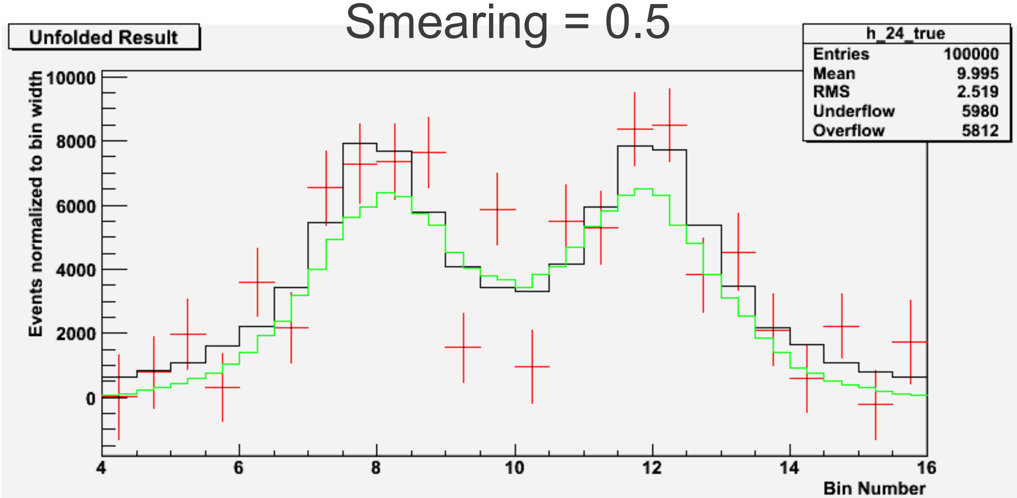
KregModeSize Unfolded Result

Smearing = 0.3

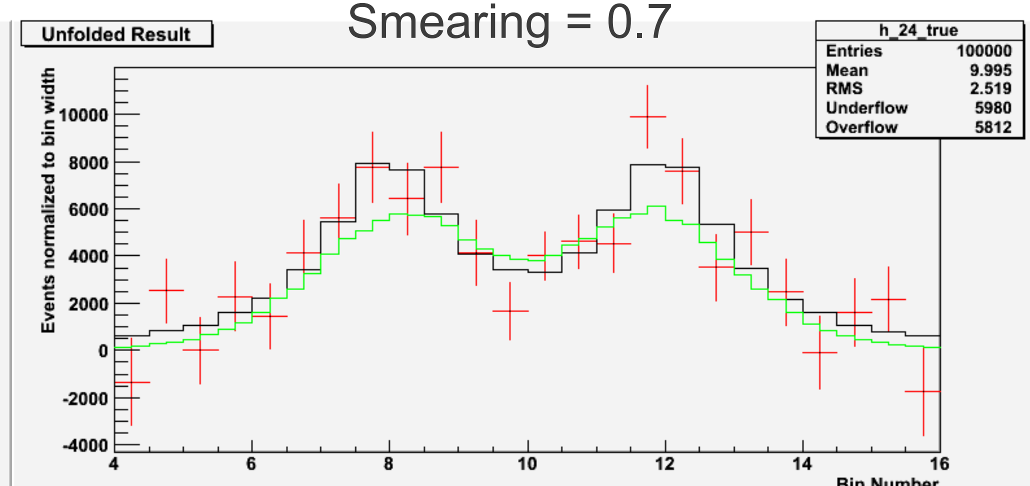


- Bin width = 0.5
- Unfolding with a smearing smaller than bin width seems to give decent results
- For bigger values large errors appear and there are a significant deviation from the True values

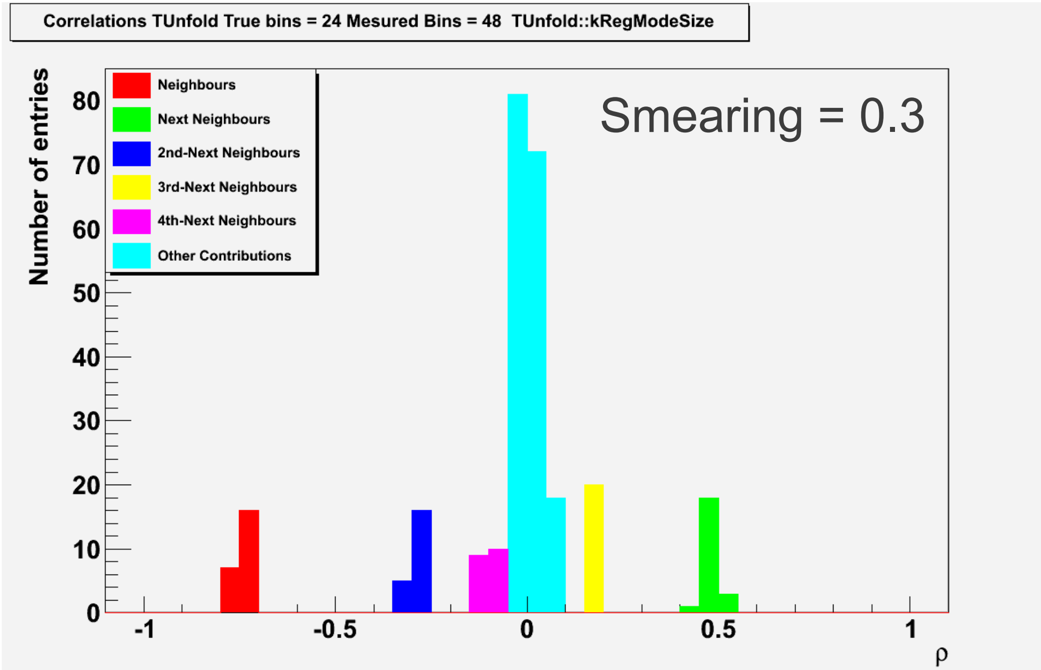
Smearing = 0.5



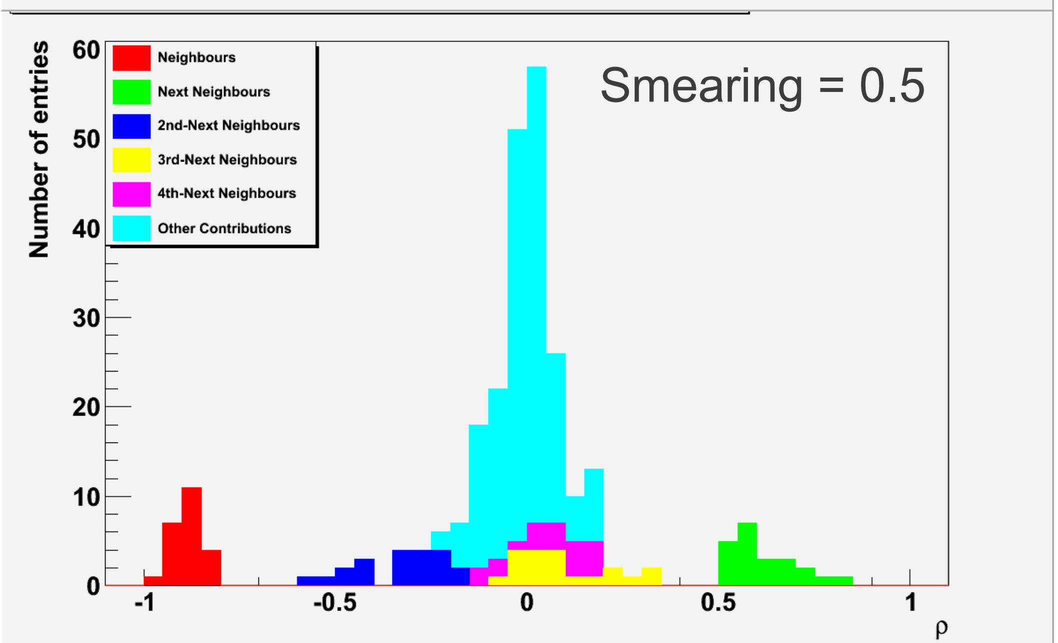
Smearing = 0.7

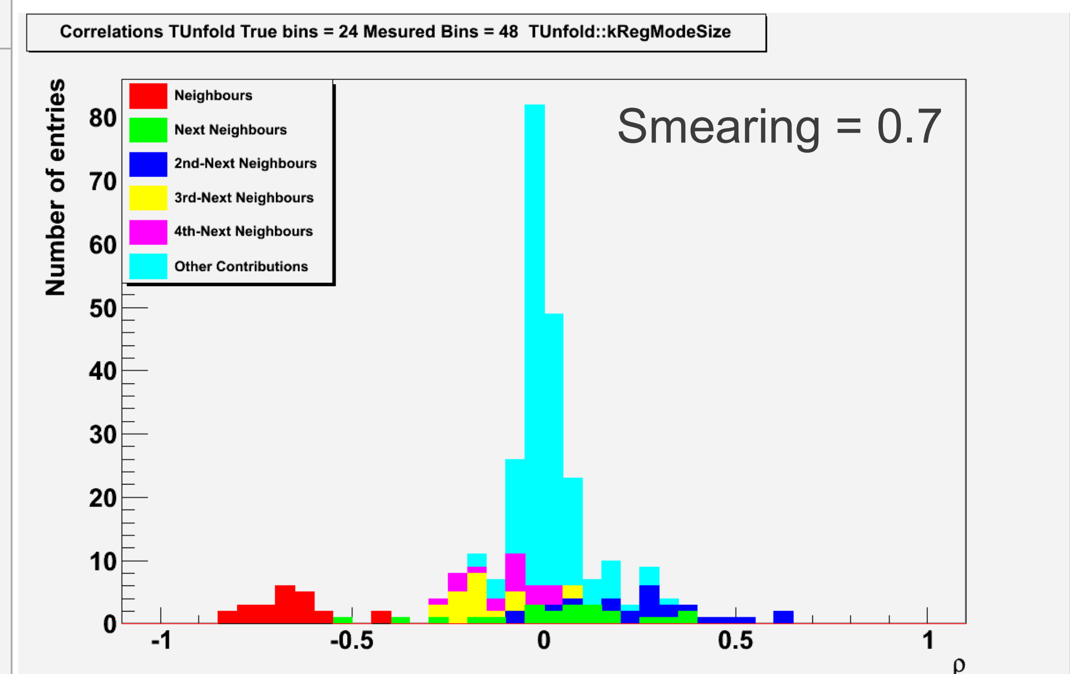


Bin Correlations

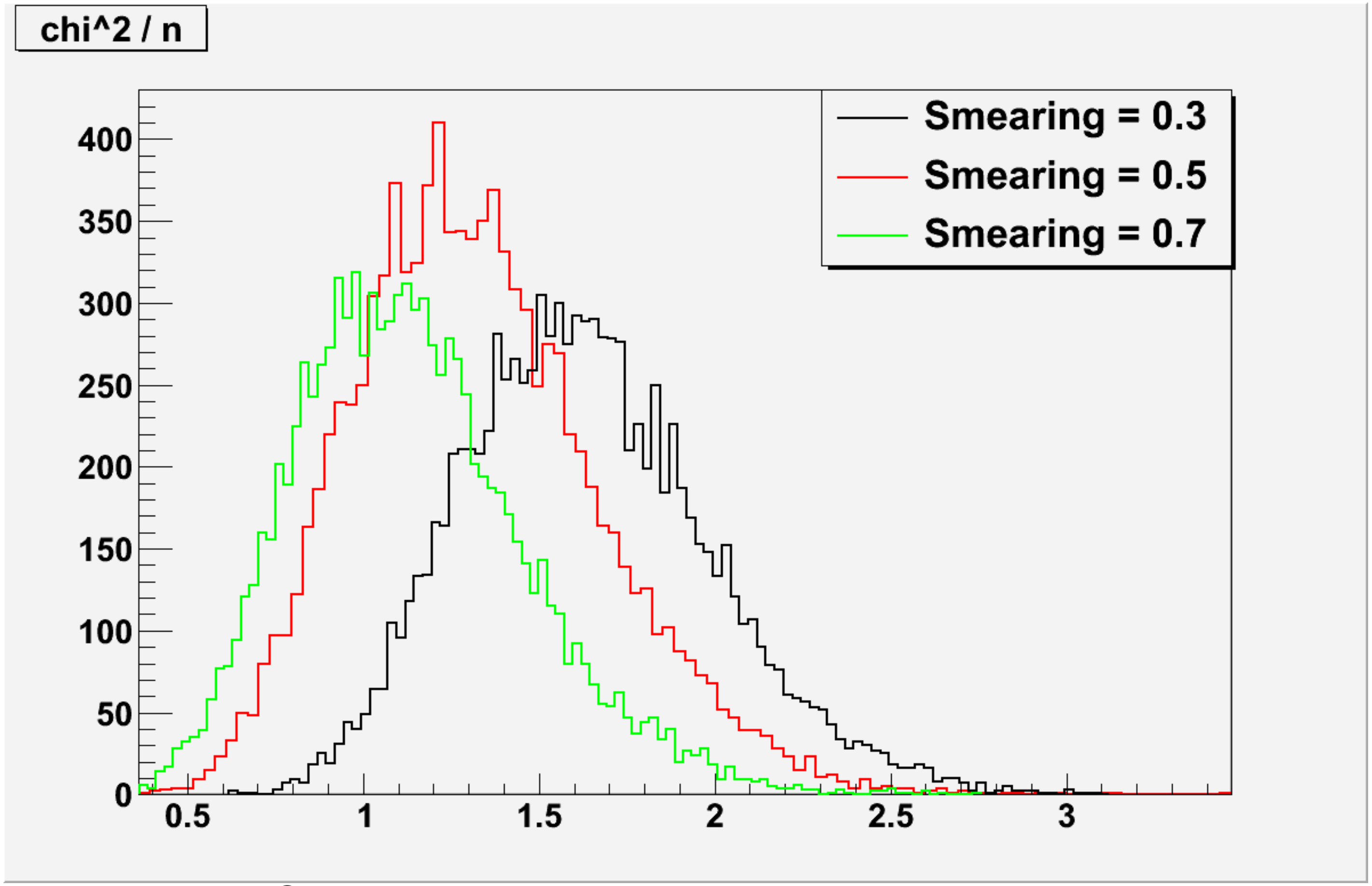


- Correlations smear out with the increase of the smearing value



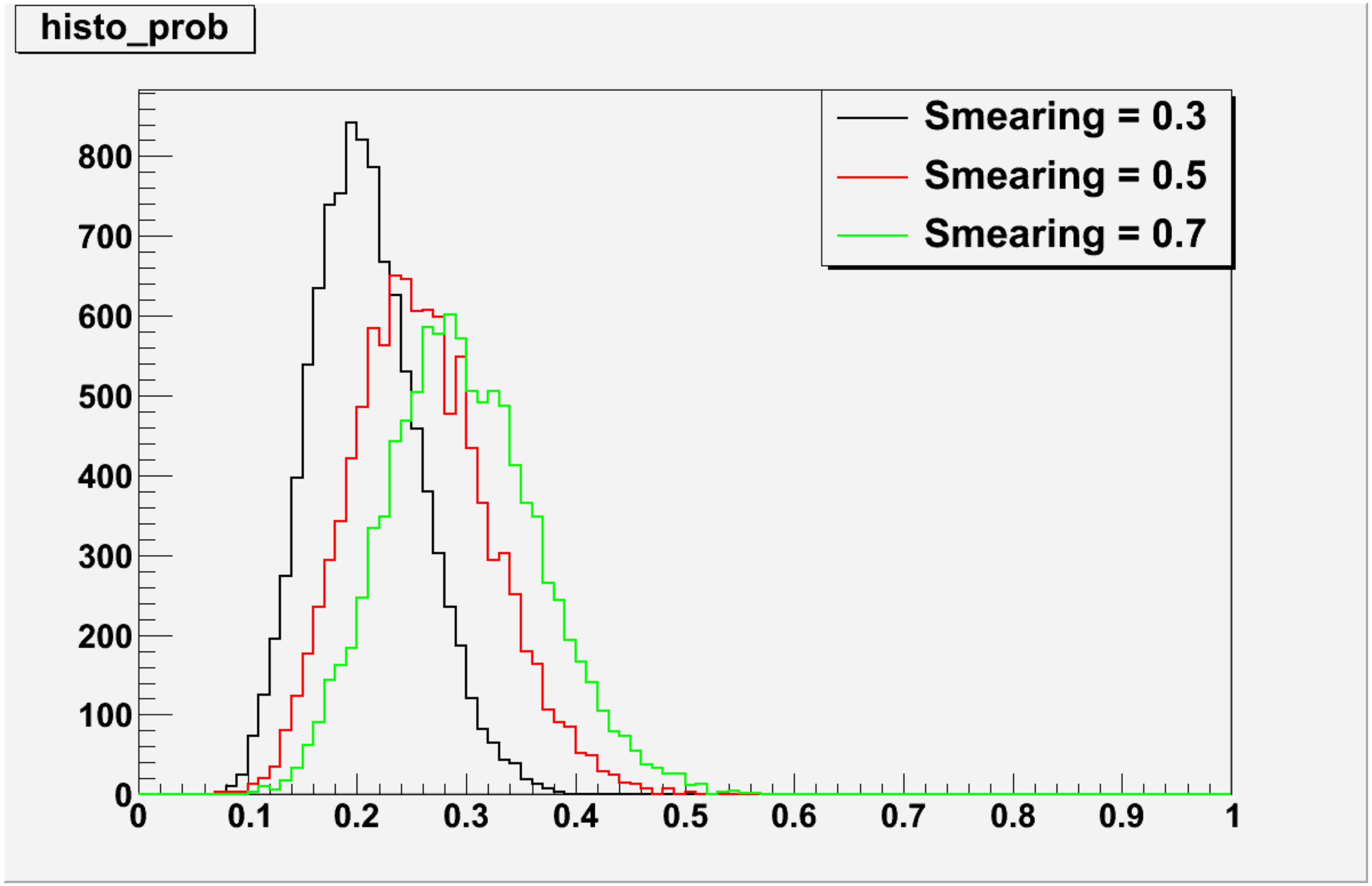


Chi^2



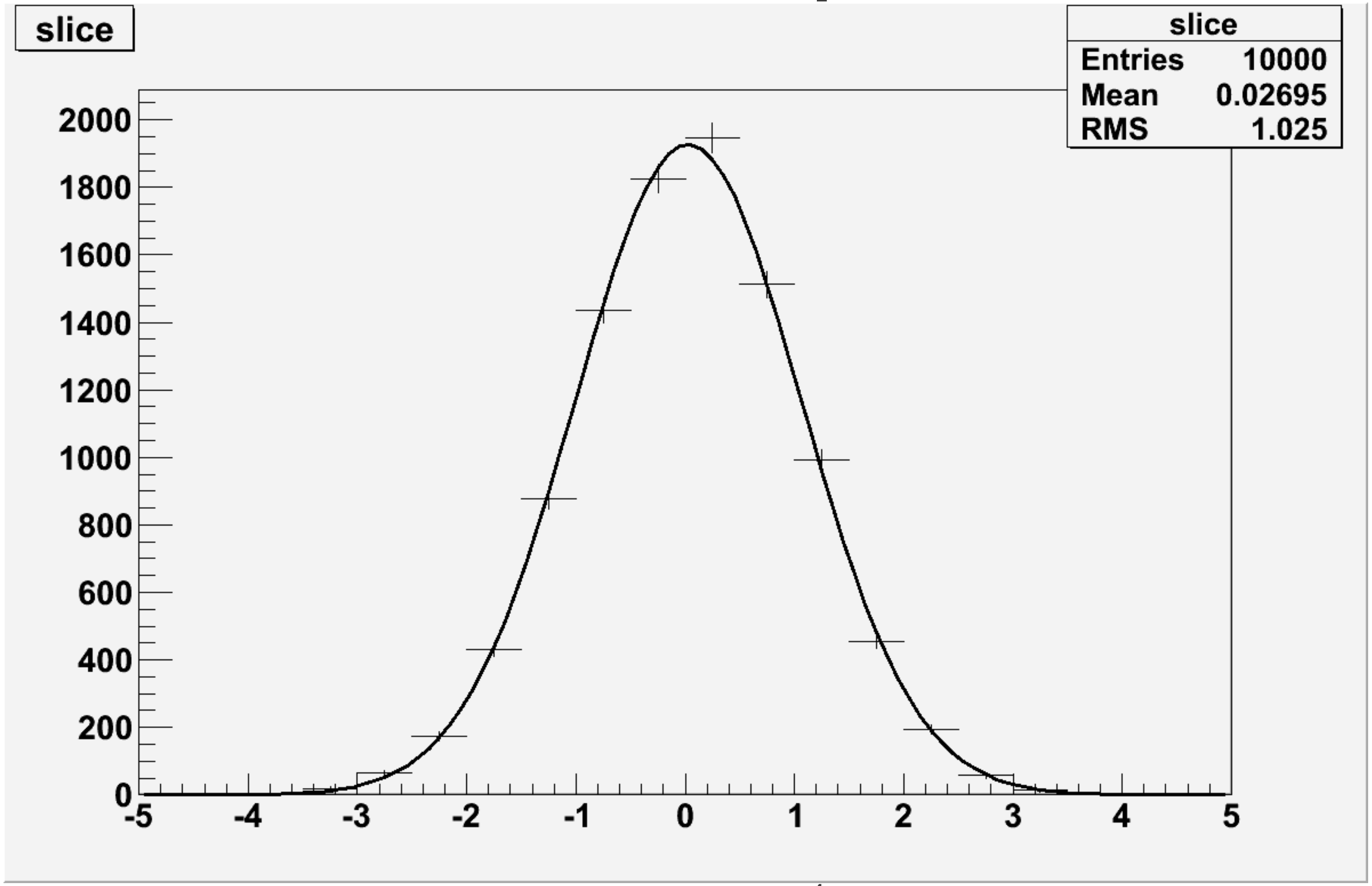
- Chi^2 divided by degrees of freedom
  - Expected mean value : 1
- Small smearing values tend to have a higher value of chi^2

Chi^2 Probability



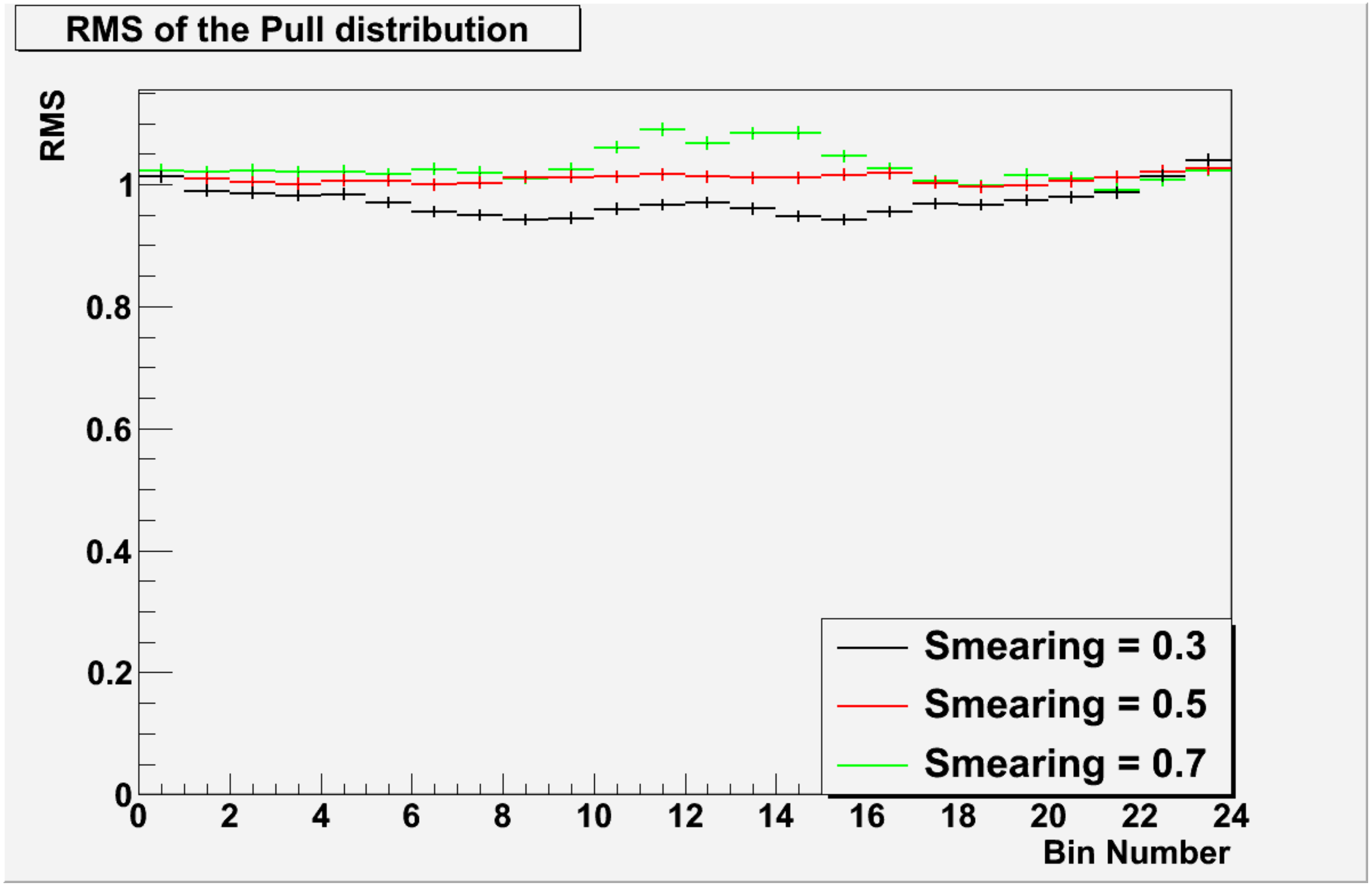
- Flat distribution expected
- Peak under 0.5 → the fit is too bad

Pull example



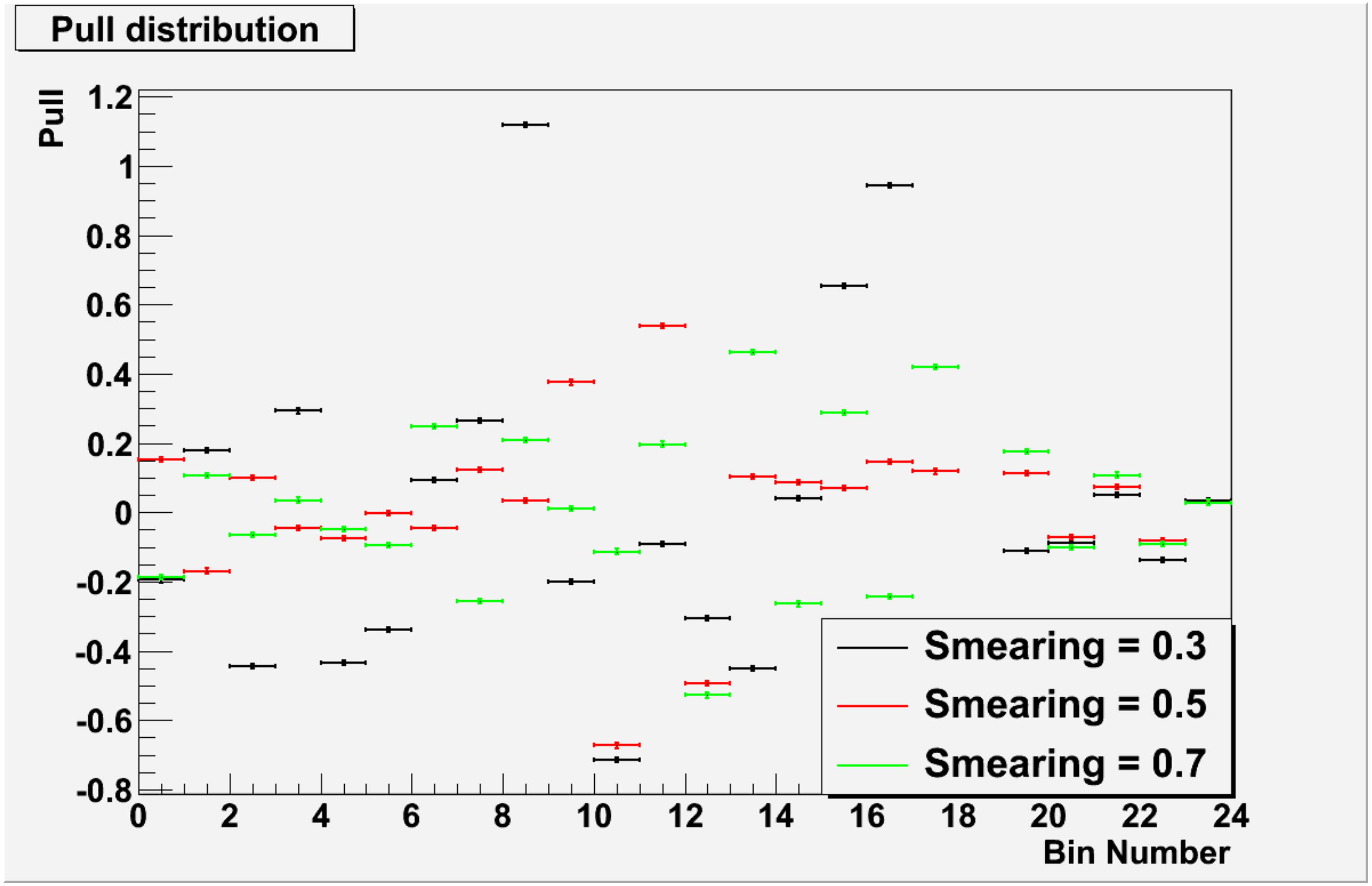
- Distribution obtained with 10^4 independent runs
  - Pull = (Unfold – True)/Sigma
    - Bin Number: 24
- Pull distribution for each bin is Gaussian shaped

RMS of the Pull



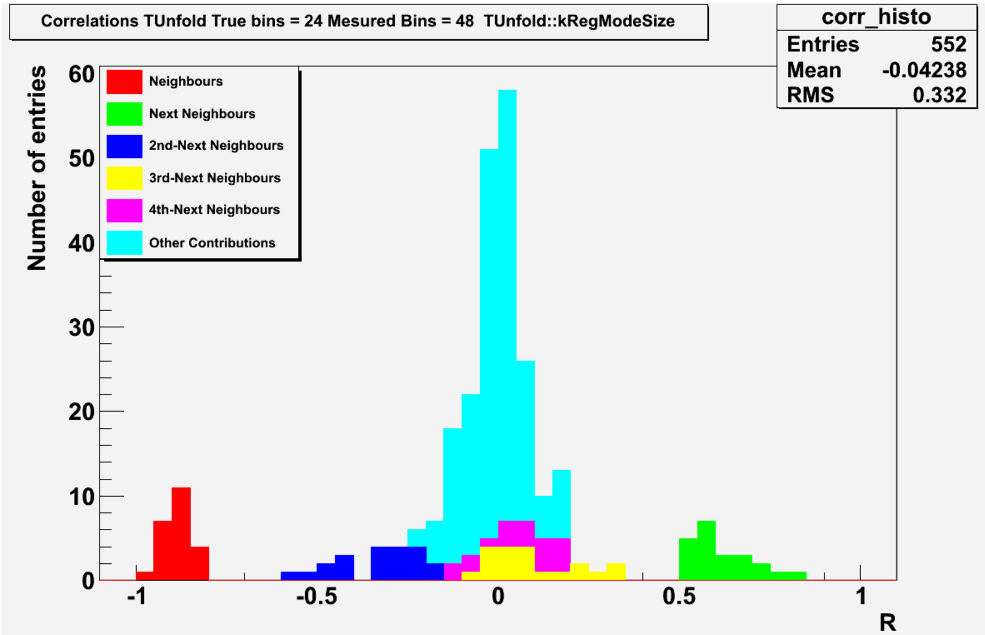
- Expected flat = 1 distribution
- No significant deviation found

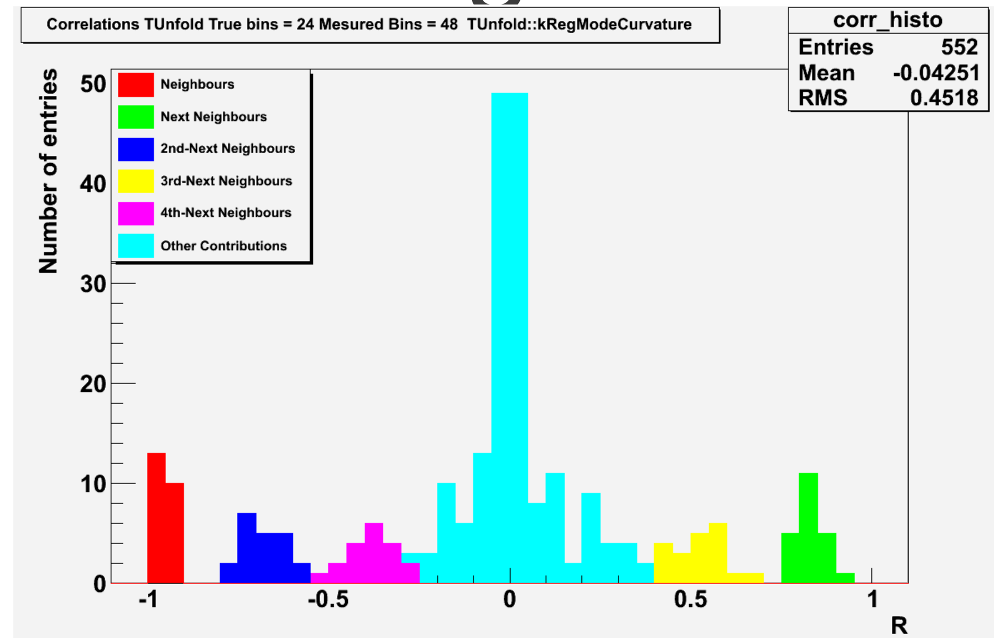
Pull Distribution

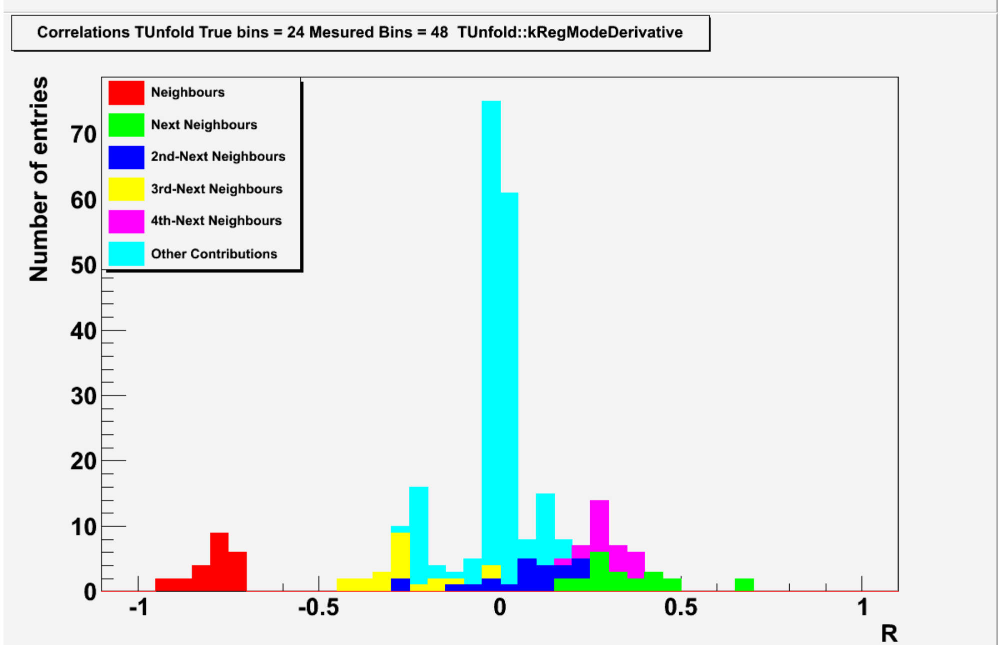


- Biases up to 1sigma
- Alternating sign between neighbors
- Bigger smearing → less pull

Bin Correlations, Smearing = 0.5

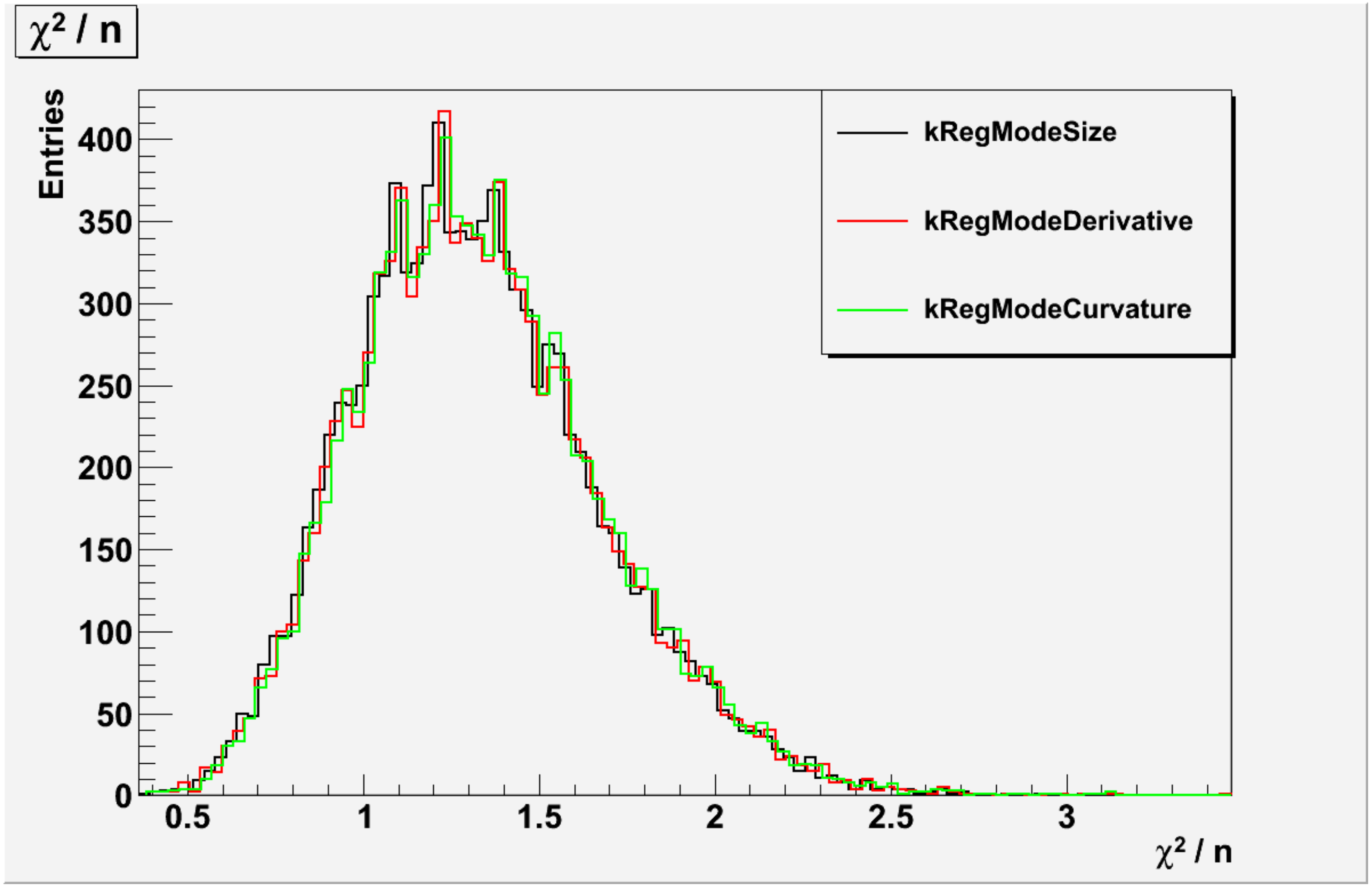






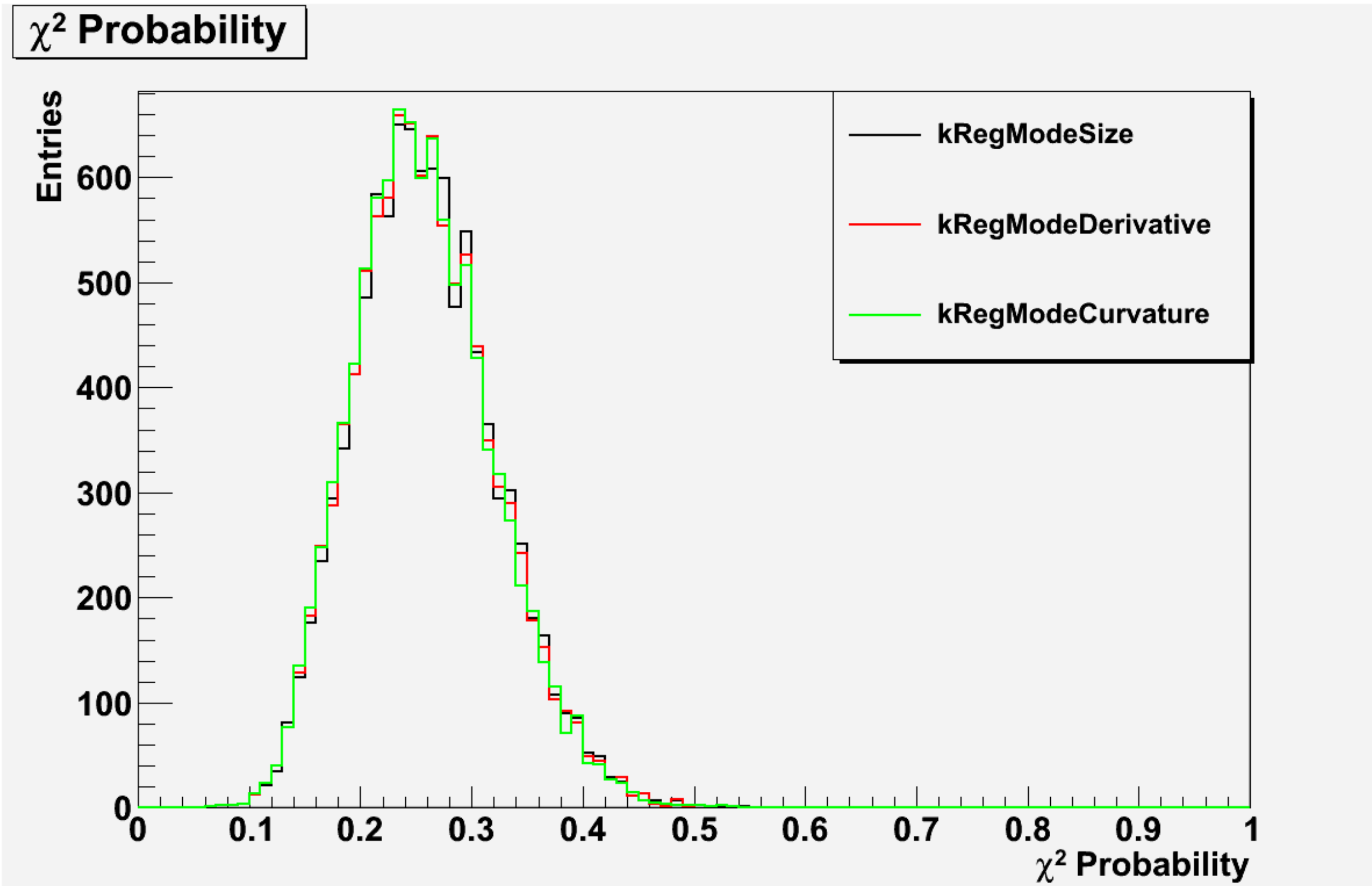
- Different regularization modes tested
- Patterns change with the mode used

Chi^2 distribution



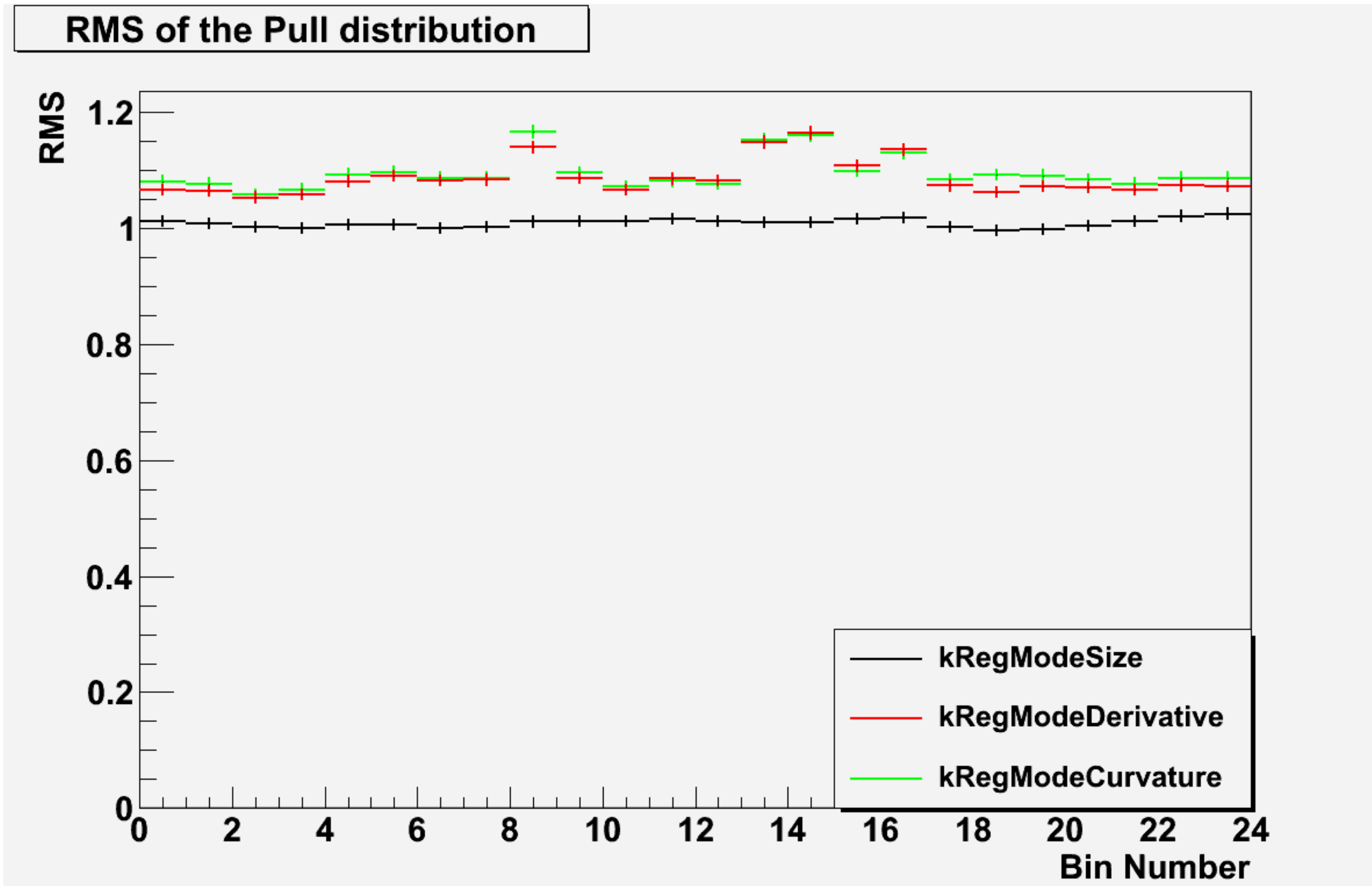
- Chi^2 divided by degrees of freedom
  - Expected mean value : 1
- No significant change due to the regularization mode

Chi^2 probability



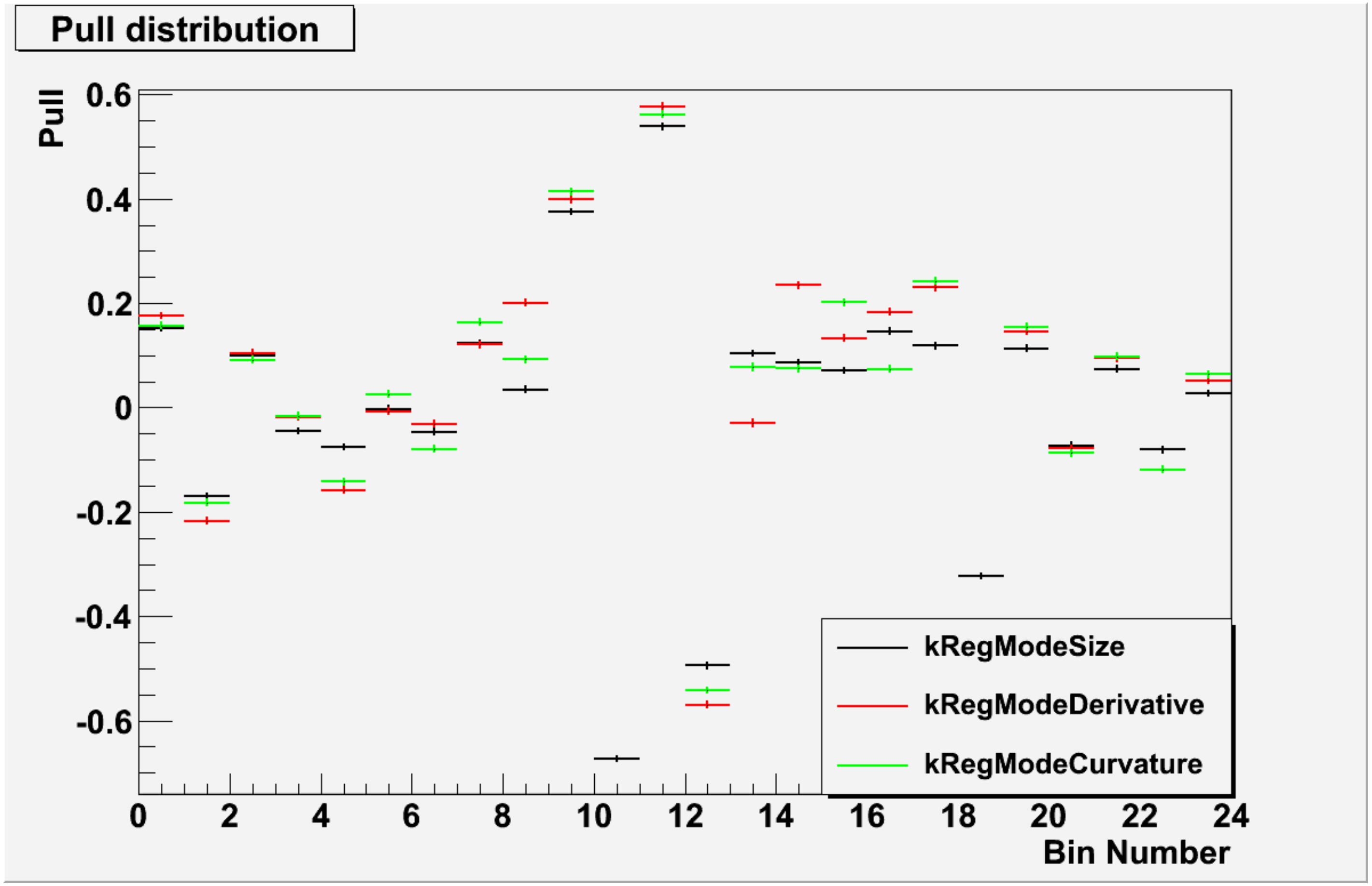
- Flat distribution expected
- Peak under 0.5 → the fit is too bad
- No significant differences between the methods

RMS of the Pull



- Expected flat = 1 distribution
- Derivative and Curvature modes are not so flat as the Size

Pull distribution



- Biases up to -0.6sigma
- Alternating sign between neighbors
- No significant change between regularization modes

Summary

- Several smearing values had been tested
  - The mean value of Chi2 tends to the expected values as the smearing decreases. Why?
- All the regularization modes of TUnfold had been compares for smearing = 0.5
  - Correlations change between modes
  - Chi2 and Pull do not show significant differences between the modes, except for Pull RMS
- Clearly the aim for a bin width smaller than Gaussian smearing is too ambitious

Plans

- Discuss this results with Stefan Schmidt (TUnfold creator)
- Look at Γ curve
- Try to understand/use λ-term
- In principle ready for the blind test