

# **Statistical analysis of Data/MC comparisons**

**Serguei Bitioukov**

**Institute for High Energy Physics  
named by A.A.Logunov of NRC «Kurchatov Institute»**

17/10/2018 GEANT4 Hadronic Working Group Meeting

# Introduction

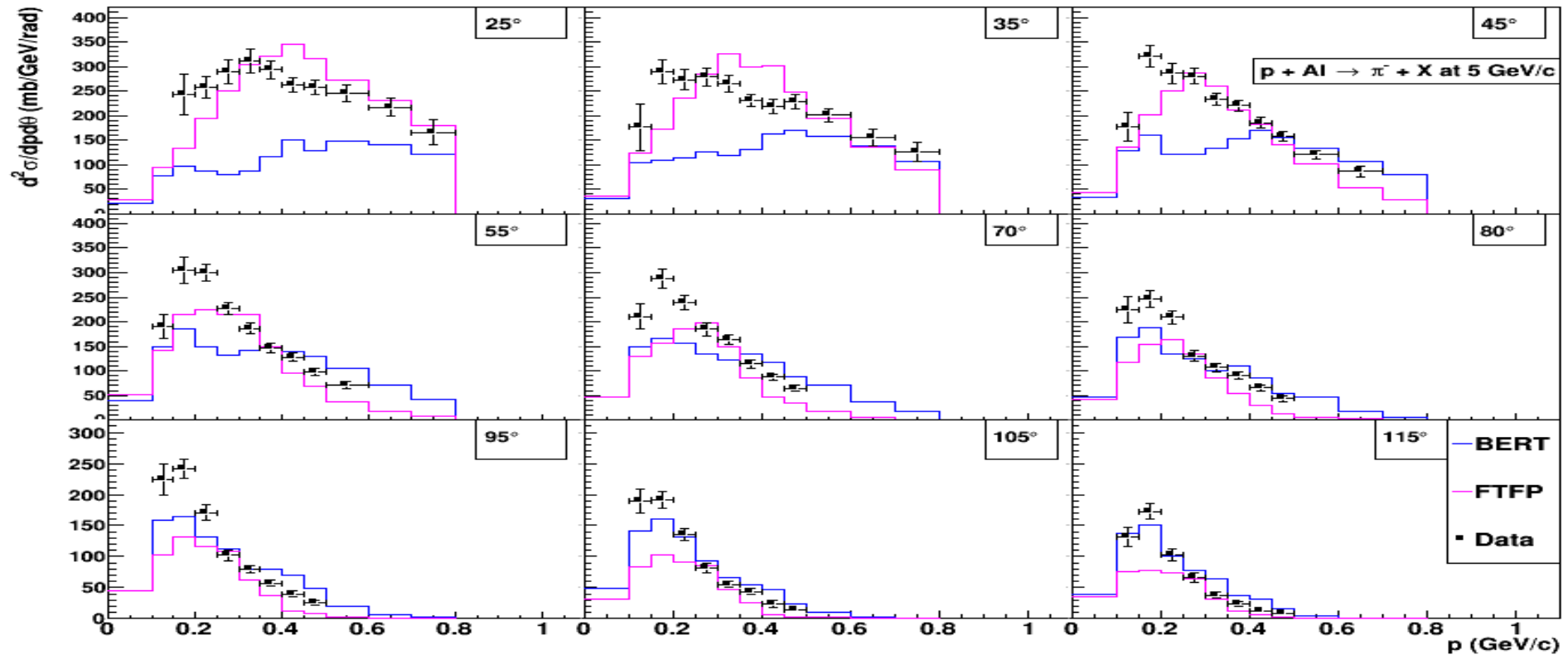
There is a problem to classify comparisons between simulation and data in the case when data consists of many different results for different parts of phase space.

In the report are considered the possibilities of new method [arXiv:1302.2651,2013; EPJ+, 128 (2013) 143] for such type comparisons.

As example are taken the experimental data of experiment HARP. The author would like to thank Prof. V. Ivantchenko for help with access to these data and for very useful discussions.

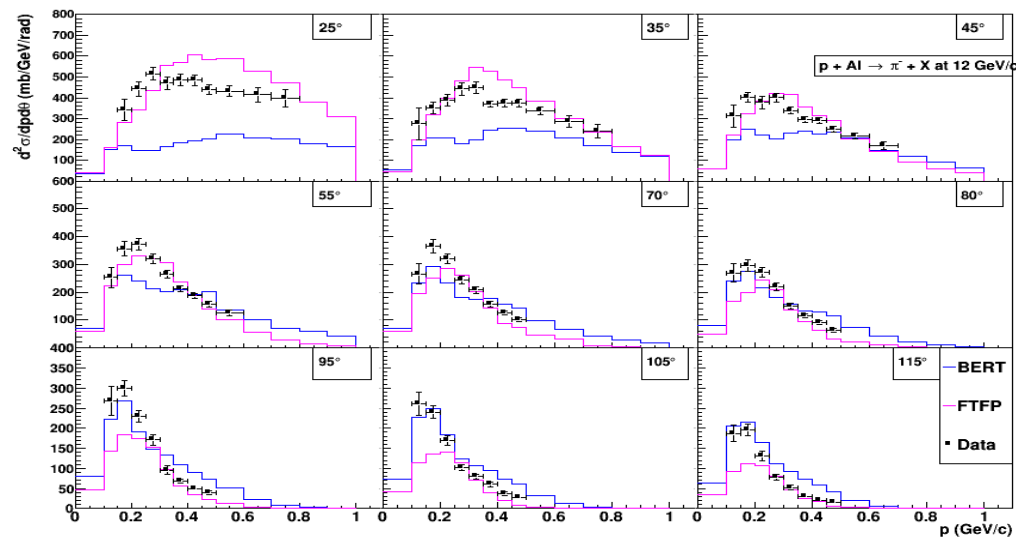
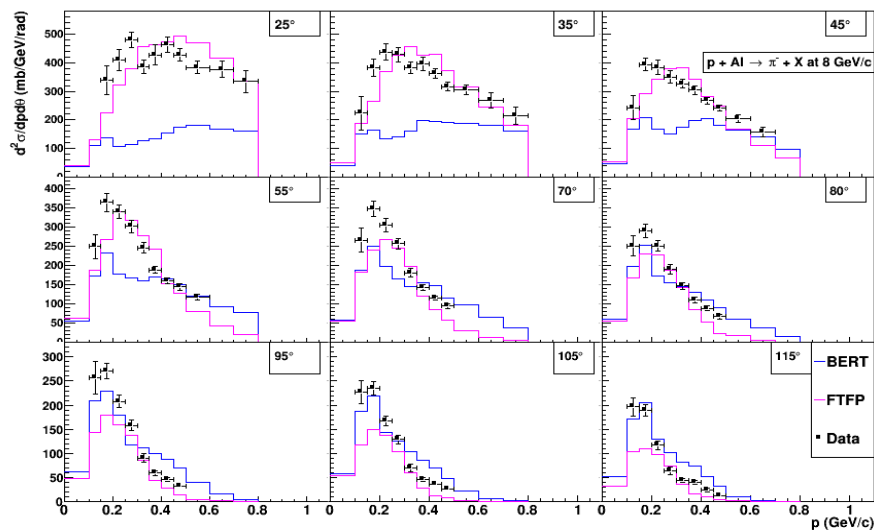
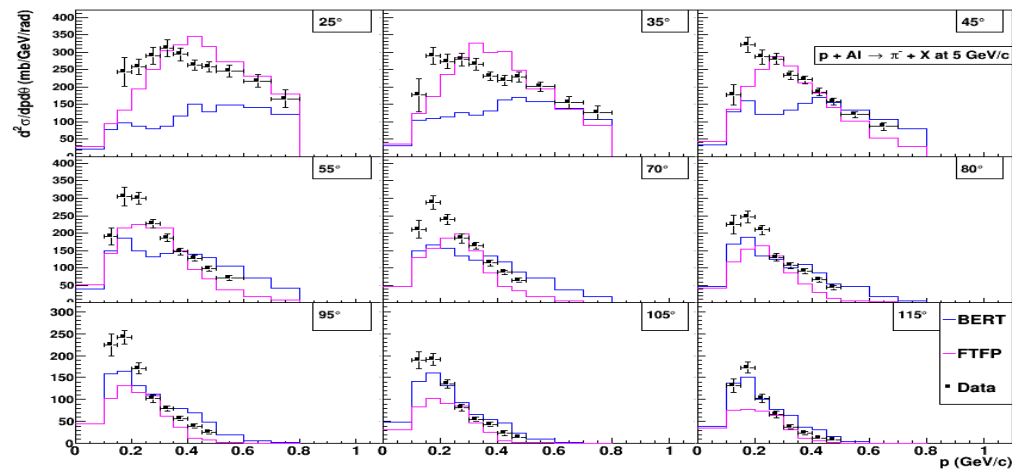
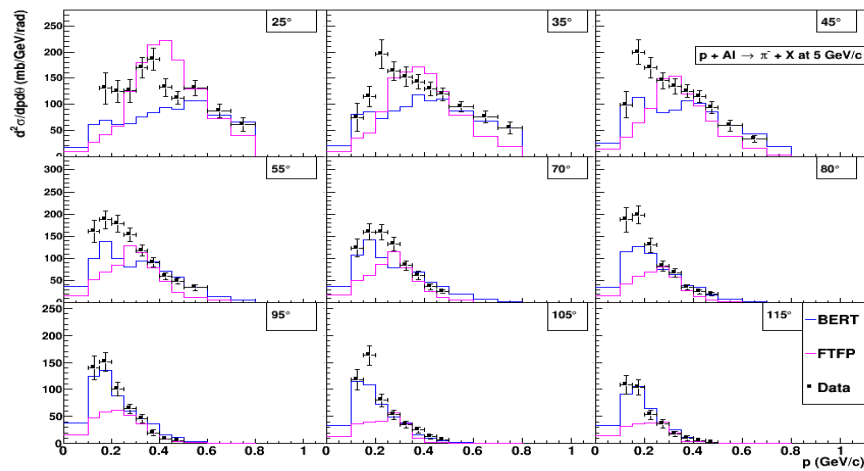
# Goal: compare models with data I

Experiment:  $p \text{ Al} \rightarrow \pi^- + X$ ,  $p \text{ Ta} \rightarrow \pi^- + X$ . 9 ranges of angles.  
GEANT4 (10.4ref07) -- 3 models: FTFP (Fritiof string model in G4), BERT (Bertini cascade), BIC (binary cascade)



# Goal: compare models with data II

4 energies of incoming protons: 3 GeV, 5 GeV, 8 GeV, 12 GeV.



# Method of comparison

Two dependencies:  $x_{11} \pm \sigma_{11}, x_{21} \pm \sigma_{21}, \dots, x_{n1} \pm \sigma_{n1}$  and  
 $x_{12} \pm \sigma_{12}, x_{22} \pm \sigma_{22}, \dots, x_{n2} \pm \sigma_{n2}$ .

Introduce the “significance of difference” for point #  $i$ :  $s_i = \frac{x_{i1} - x_{i2}}{\sqrt{\sigma_{i1}^2 + \sigma_{i2}^2}}$ .

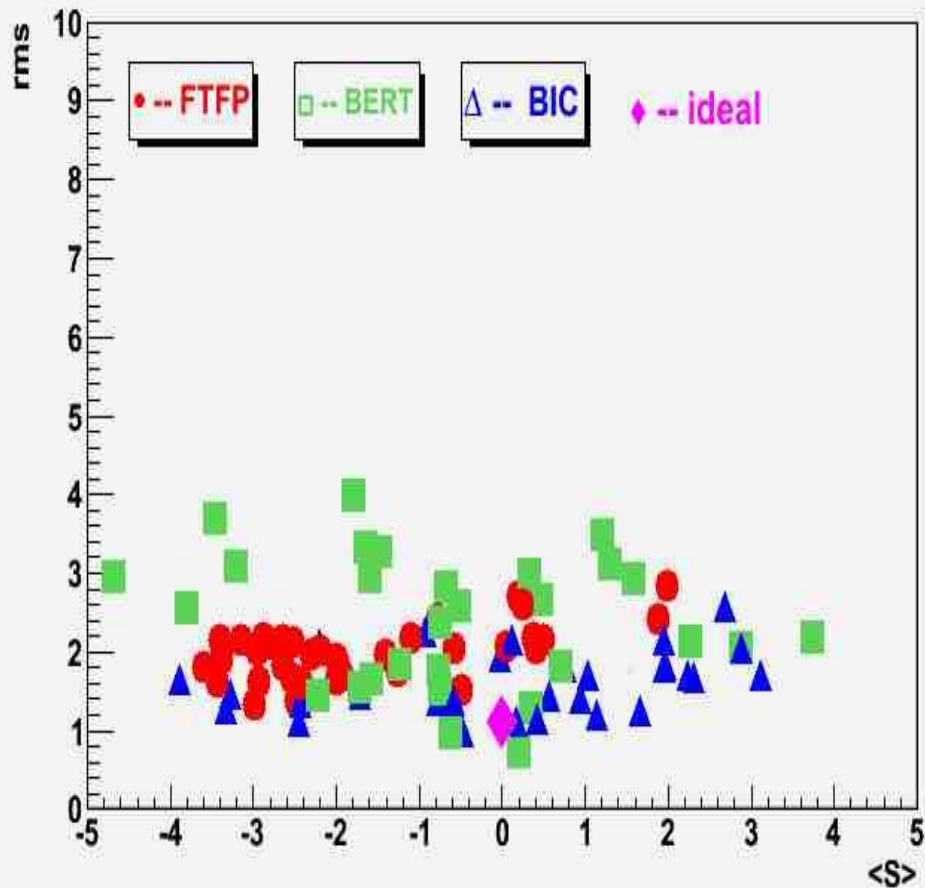
Often is used  $\chi^2 = \sum_{i=1}^n s_i^2$  as a test statistics for comparison.

As known,  $\frac{\chi^2}{n} = \bar{S}^2 + rms^2$ , where  $\bar{S} = \frac{\sum_{i=1}^n s_i}{n}$ ,  $rms = \sqrt{\frac{\sum_{i=1}^n (s_i - \bar{S})^2}{n}}$ ,  
i.e.  $\chi^2$  is nonlinearity sum of two independent test statistics.

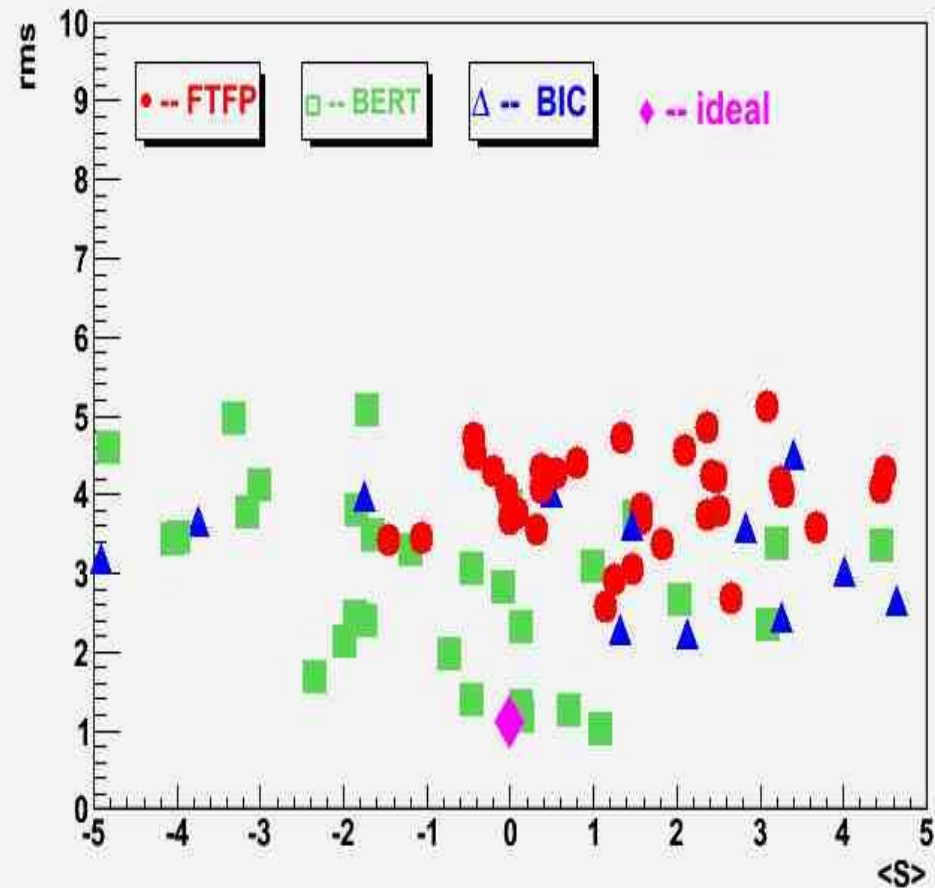
Introduce the new bidimensional test statistics  $Srms = (\bar{S}, rms)$ .

# Two reactions: p Al and p Ta

p Al [3, 5, 8, 12] GeV pi- large angle (10.4ref07)

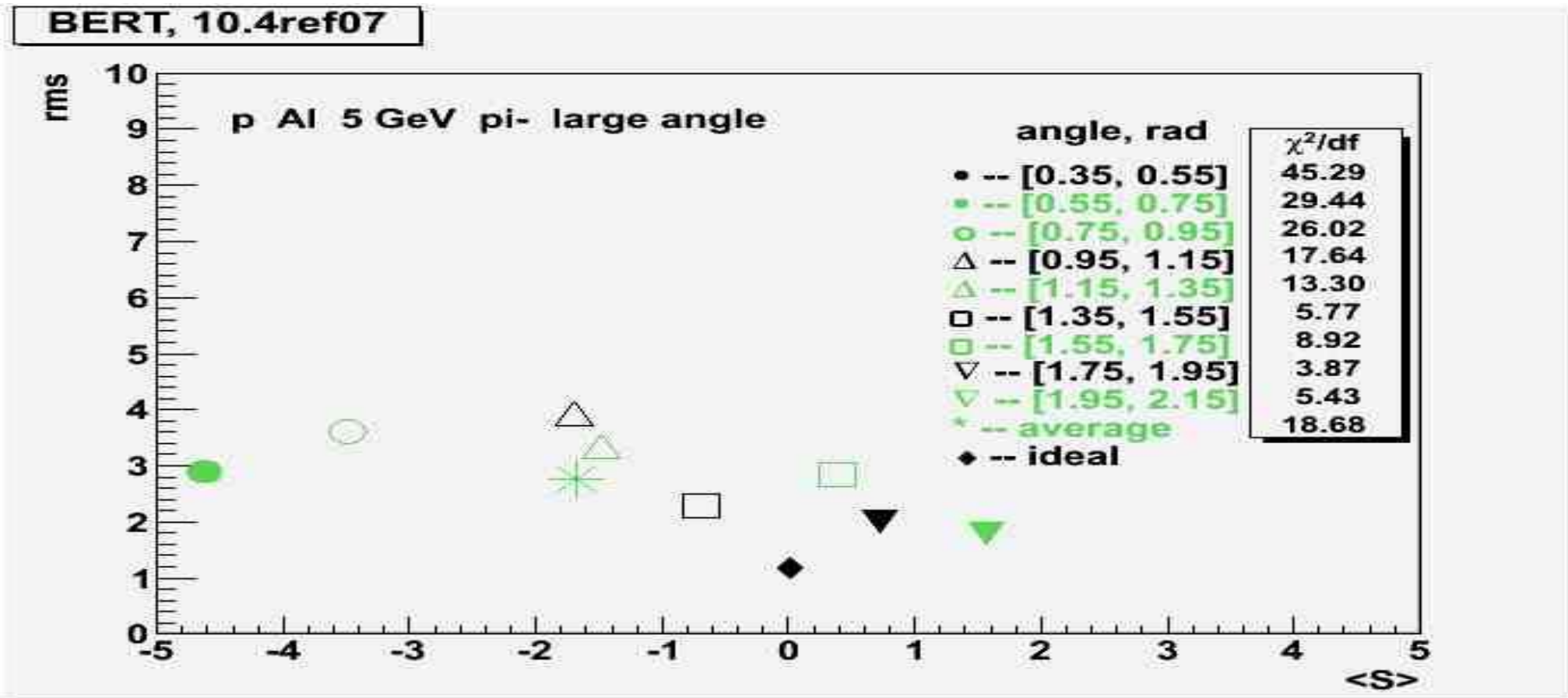


p Ta [3, 5, 8, 12] GeV pi- large angle (10.4ref07)



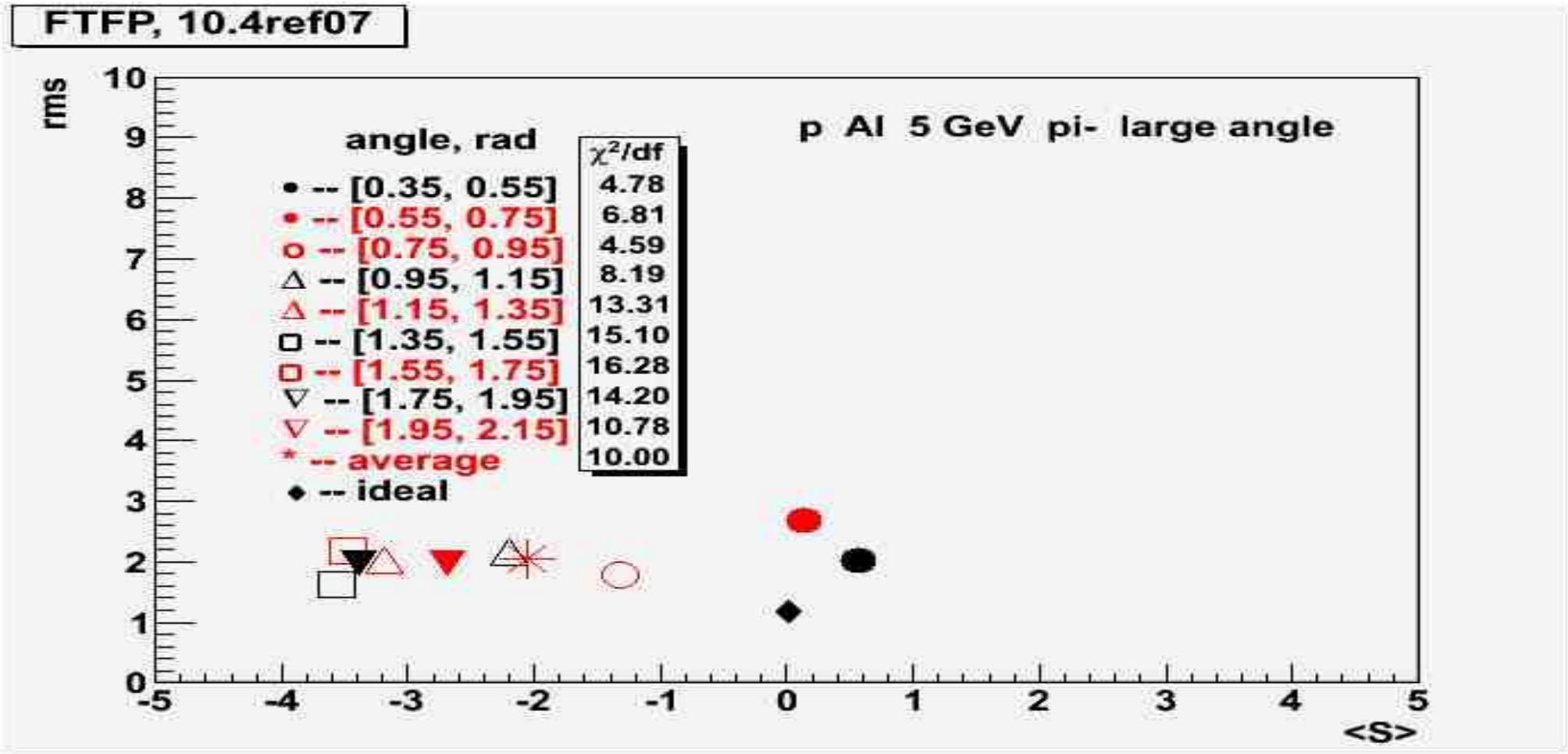
# Advantage of bidimensional test statistics

As seen, the angular dependent migration of mean “significance” from small angles to large angles exists. It points the way for possible correction of model.



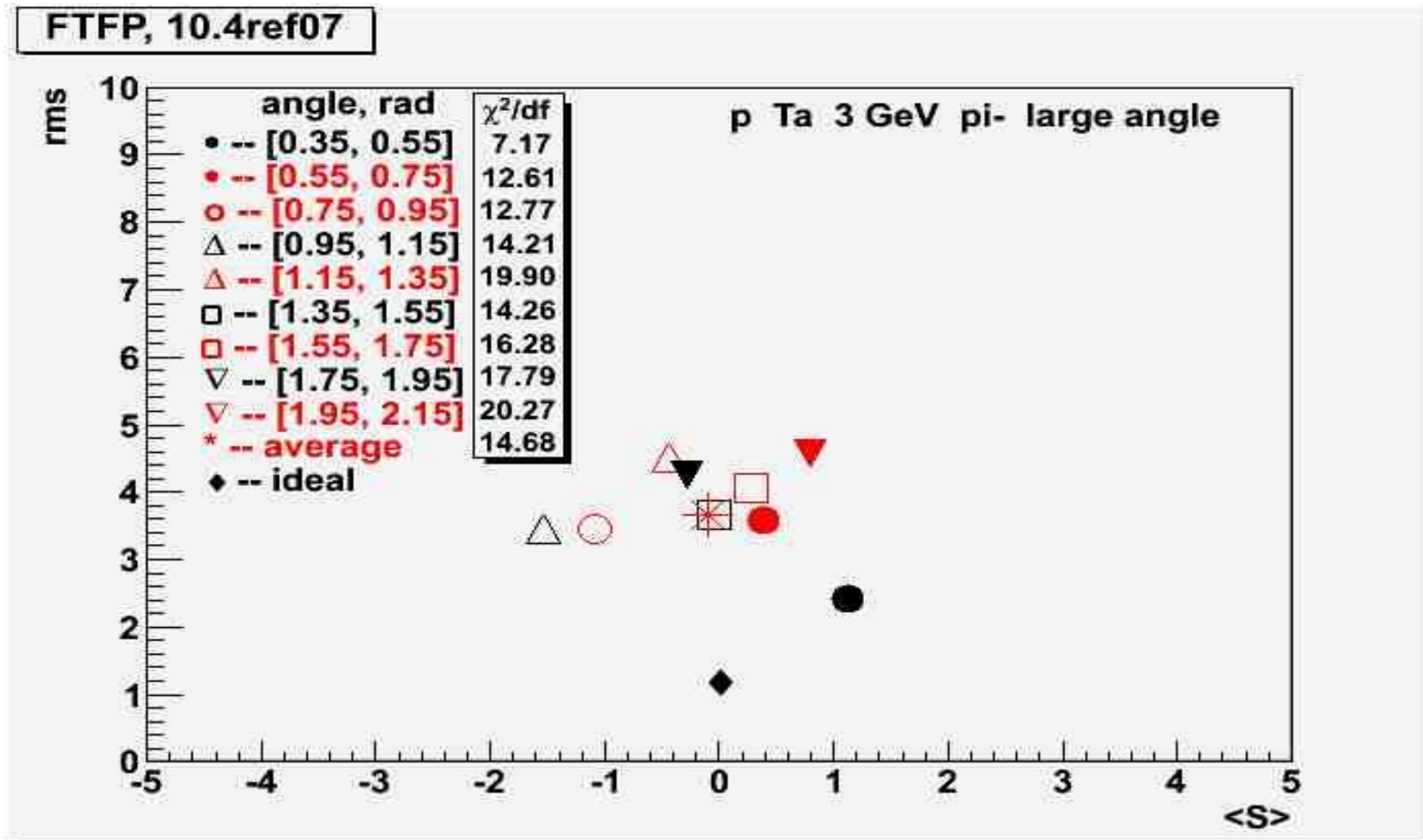
# Advantage of bidimensional test statistics

In case of FTFP model, we have reverse dependence. The correction of model is also needed.





# No angular dependent migration



# Conclusion

The described method provides more information than  $\chi^2$  analysis.

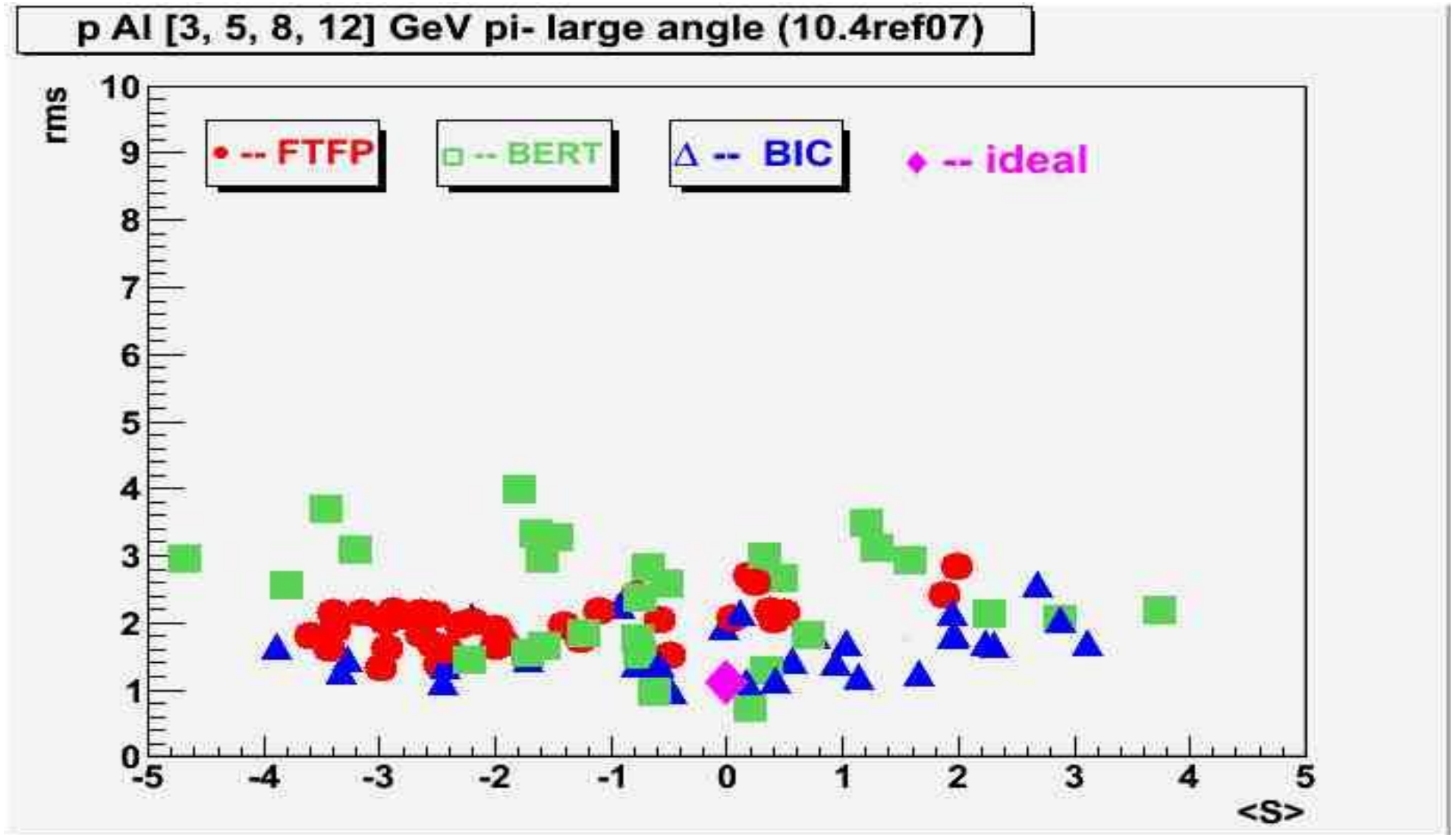
This approach for comparison of data allows also :

- to compare multidimensional dependences,
- to compare two sets of several dependences,
- to compare the dependences produced by different experiments.

This approach may be used for characterization of data/MC agreement during tuning of model parameters in the frame of Geant4 project and for control of model evolution between versions.

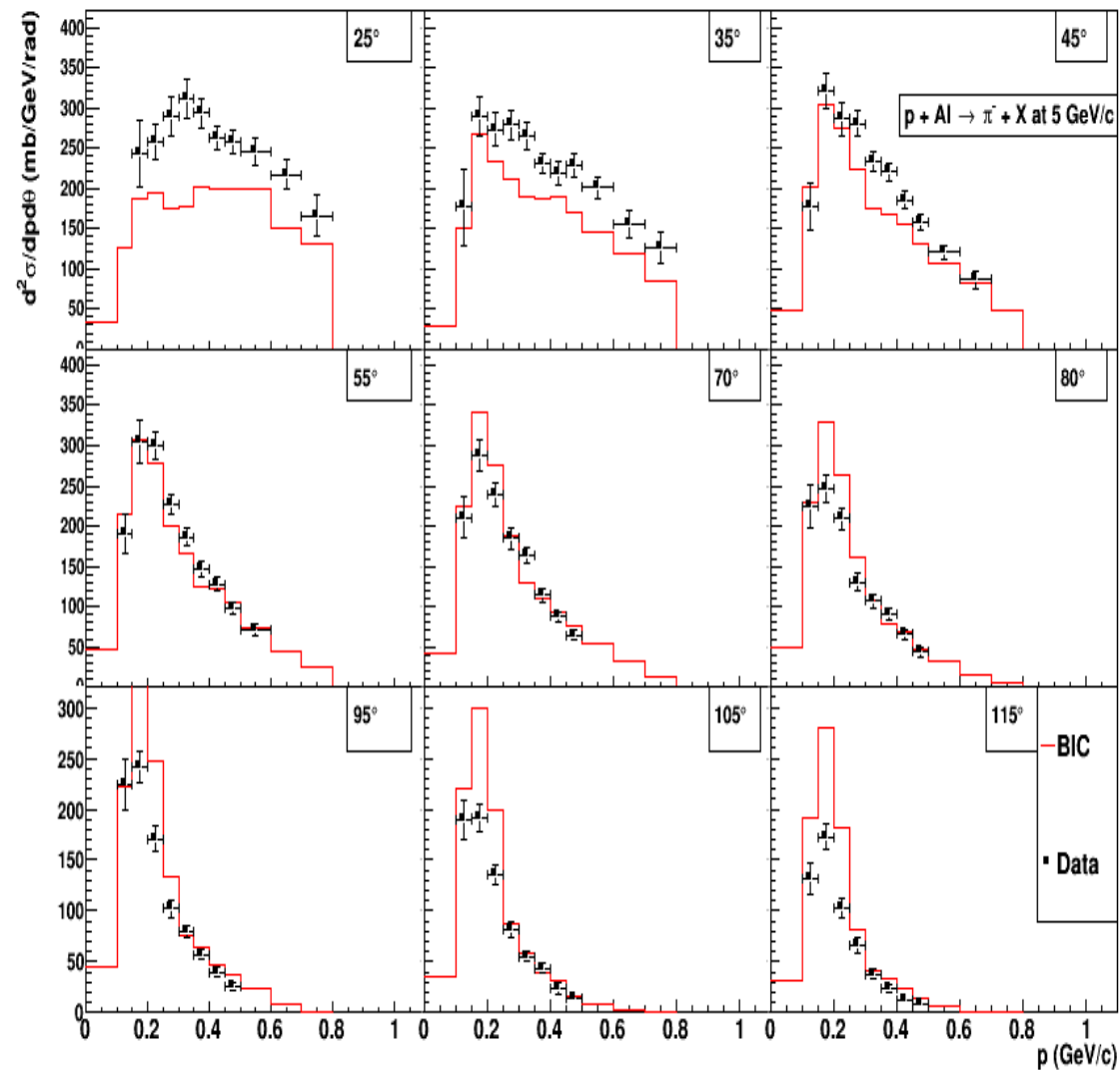
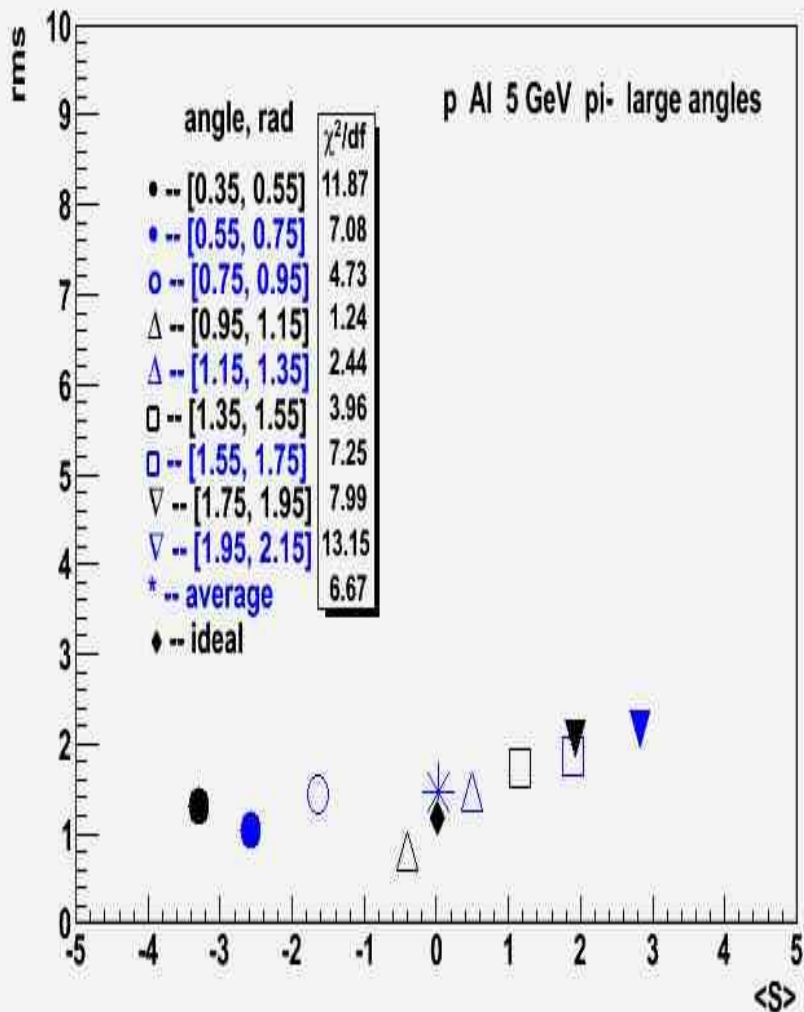
# Back up

# Slide 4 (new coordinates)



# BIC model versus data

BIC, 10.4ref07



# Ta

