

# Status of pseudo-experiments generation for PDF correlation studies

S. Amoroso, S. Glazov,  
S. Mikhalkov, V. Novik

# Introduction

It's important to understand correlated or not PDF measurements from different groups to calculate uncertainty.

PDFs determined by different groups (ABMP, CTEQ-TEA, MMHT, NNPDF, ...) are expected to be correlated due to:

- Common data samples used (e.g. HERA combined data)
- Similar theory predictions (NNLO DGLAP, NNLO coefficient functions, often identical APPLgrids)

## **Tools for the correlation measurement**

- Use xFitter as a baseline tool for toy generation
- Prepare scripts for toys generation and validation. Store the toys in a common repository

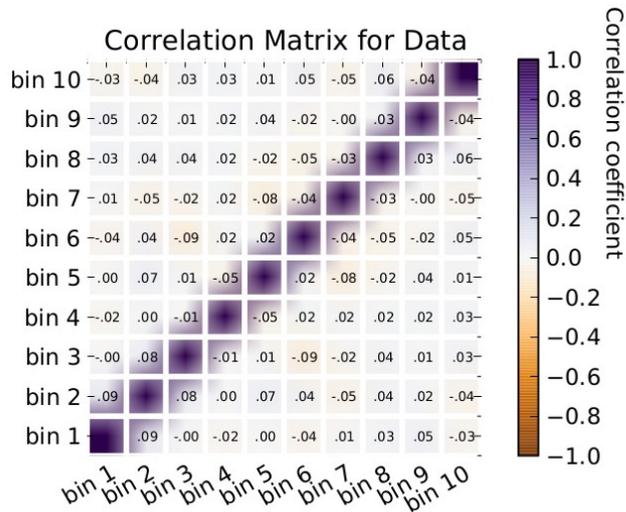
# Status of toys production

- Code to produce/validate toys in xFitter format established
  - Available at <https://gitlab.cern.ch/smikhalc/test>
- Tool functionality:
  - generation of toys,
  - covariance matrix plotting,
  - add toys bins in initial .dat file
- We were done tests with several types of datasets.
  - HERA inclusive cross sections (arXiv1506.06042)
  - ATLAS 7 TeV W, and Z double differential data. (arXiv1612.03016)
  - CMS 7 TeV jet data (arXiv1212.6660)
  - E866 Drell Yan ratio data (arXiv0103030)
- 1000 Toys for most of discussed datasets already exist.
- Result of toy production was changed from 1 file with all toys to 1 file for the each toy.
- Script for conversion in CTEQ input format was created.

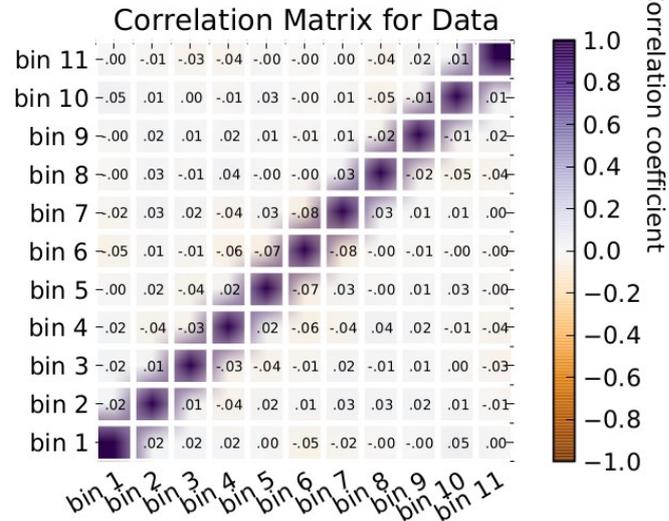
# Matrix definition and examples

$$C_{i,j} = \text{pearsonr}(\text{toys}_i, \text{toys}_j)$$

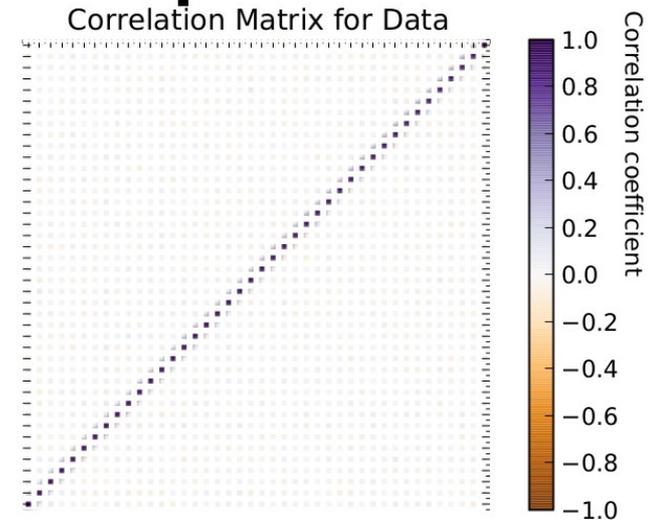
where  $\text{toys}_i$  array of  $n$  toys of bin  $i$ ;



Tevatron e866 dy arXiv0103030



LHC ATLAS wzProduction  
arXiv1612.03016



HERA inclusiveDis  
arXiv1506.06042

# Next steps

- Generate 1000 toys for each discussed data set
- CMS data sets have a large number of bins.
  - It is a difficulty for matrices plotting.
  - We need change our plotting code
- Another difficulty is input datasets differences.
  - We have to change our script to universal application for each input data set type.
- Simplicity and multipurpose script development

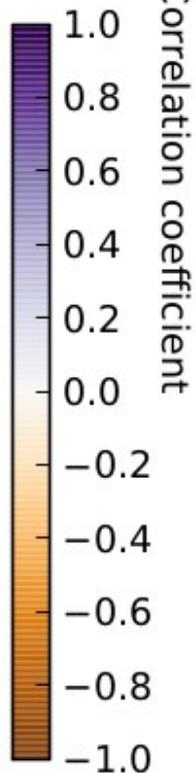
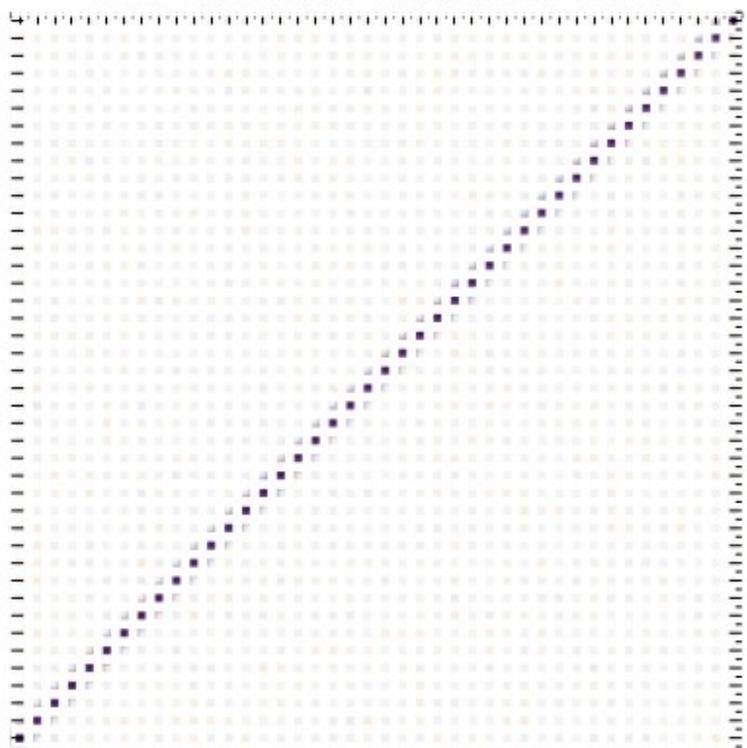
**Thank you for your attention!**

# Backup

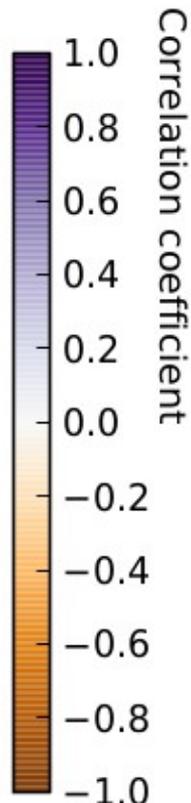
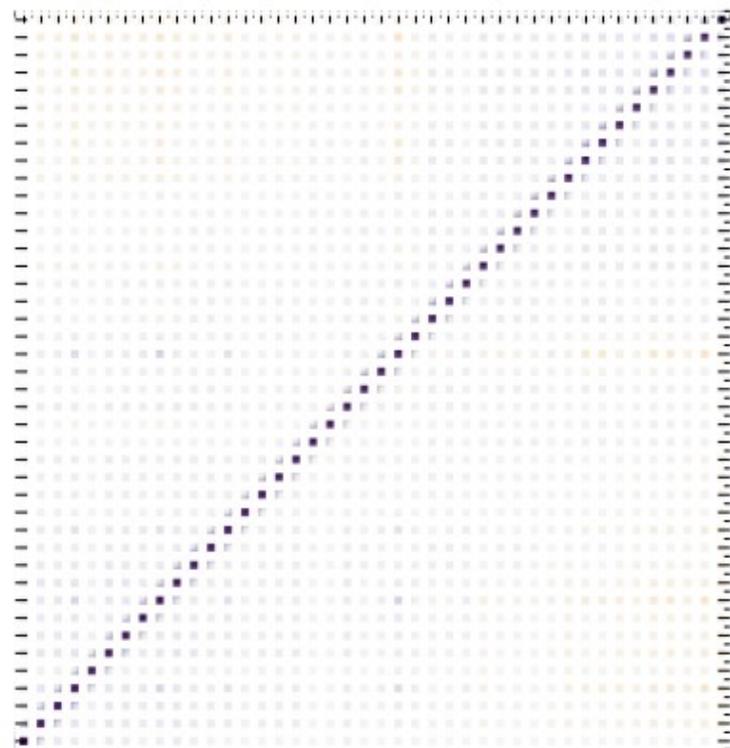
# Covariance matrices

HERA1+2\_CEm

Correlation Matrix for Data



Correlation Matrix for xFitter



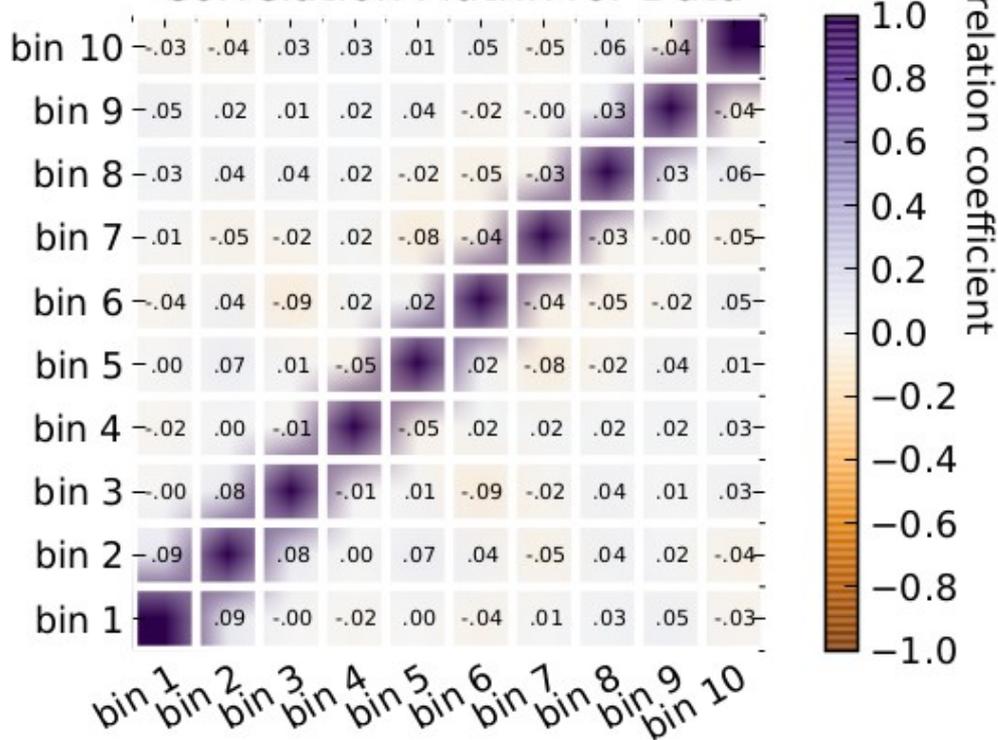
Matrix from the toys

Matrix from the xFitter method

# Covariance matrices

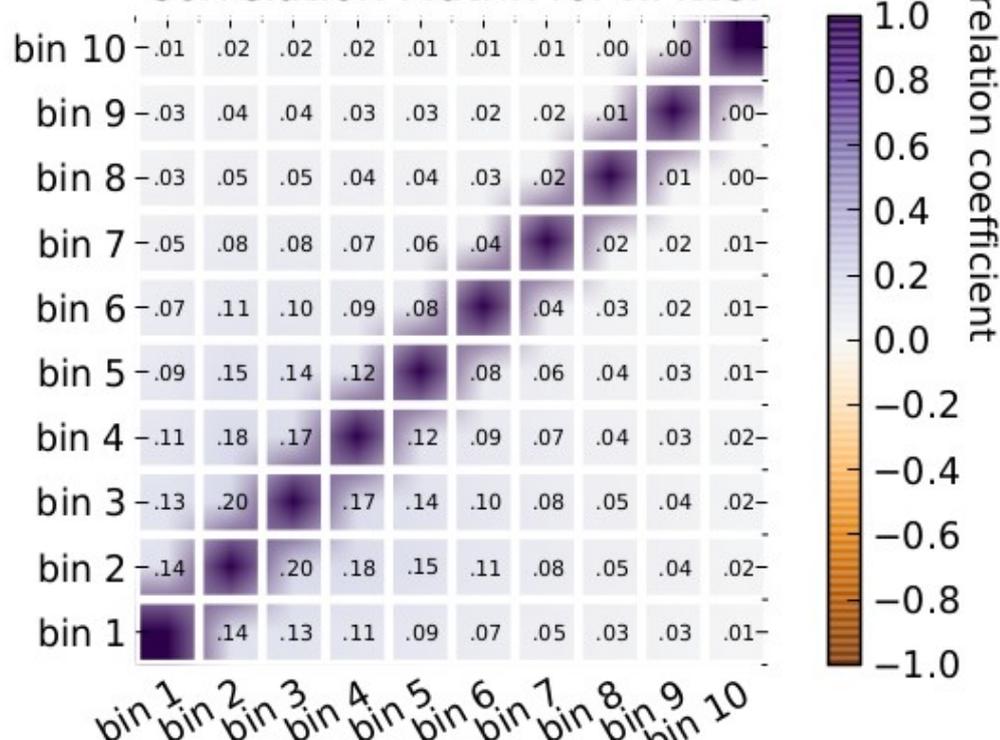
Tevatron/e866/x2-hi-mass

### Correlation Matrix for Data



Matrix from the toys

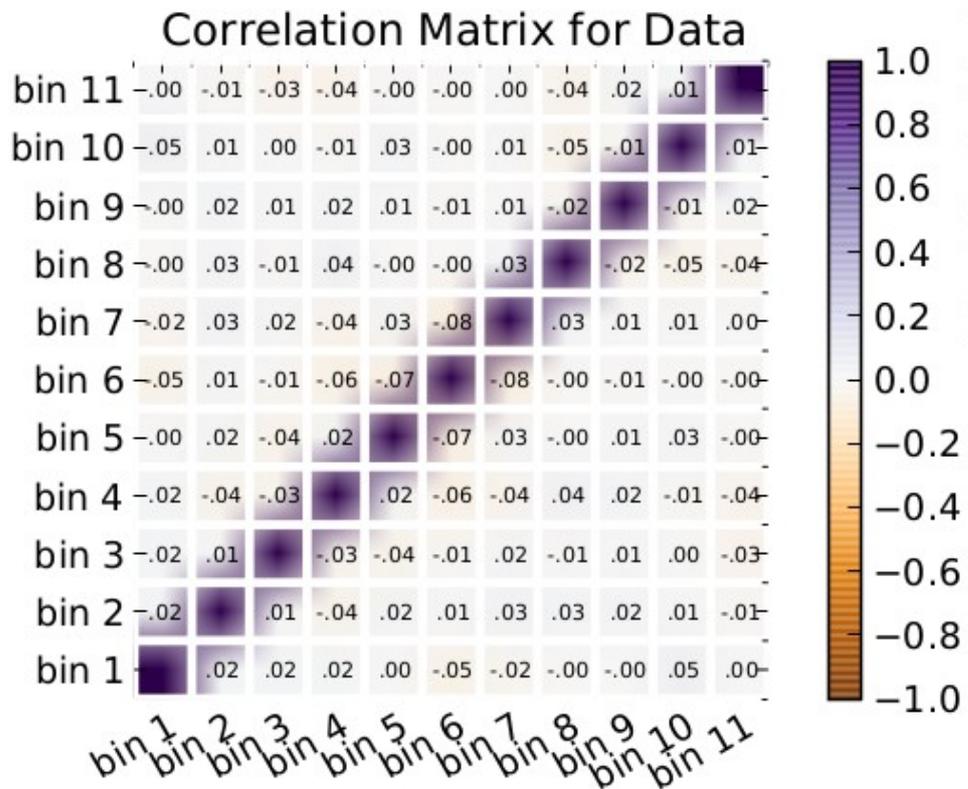
### Correlation Matrix for xFitter



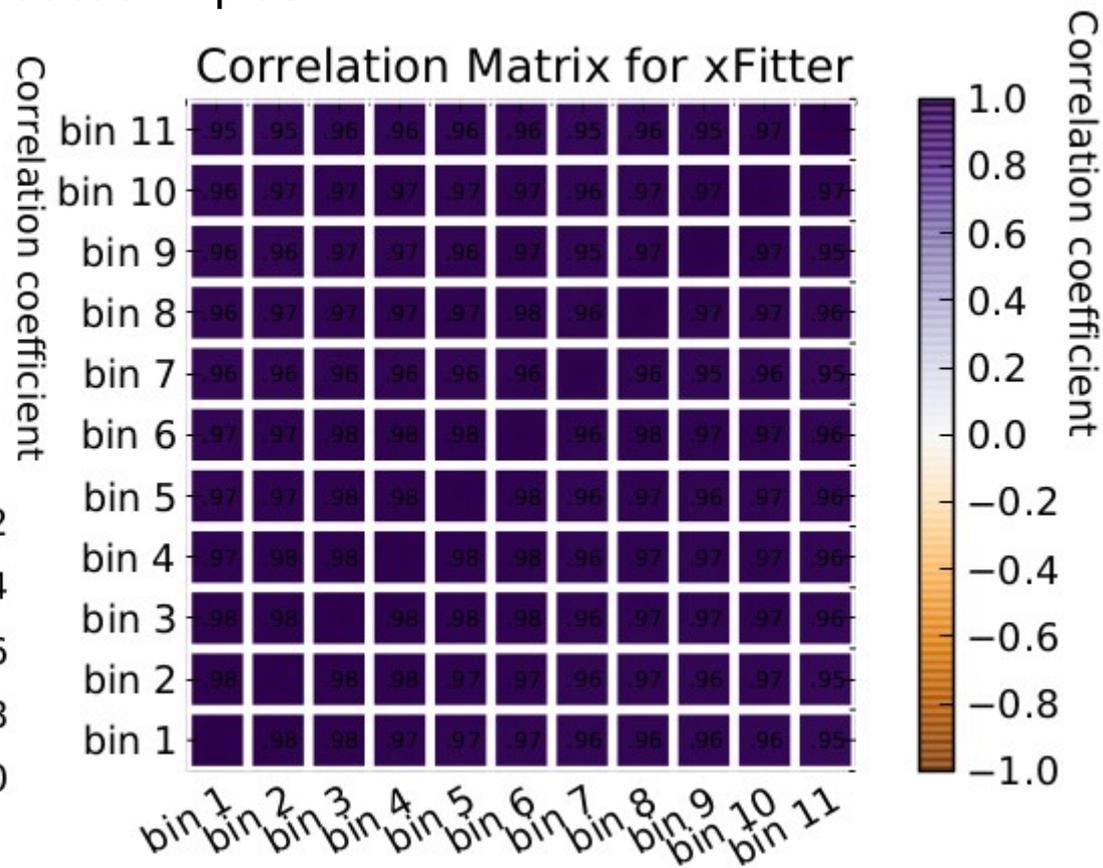
Matrix from the xFitter method

# Covariance matrices

ATLAS/wzProduction/wplus



Matrix from the toys



Matrix from the xFitter method