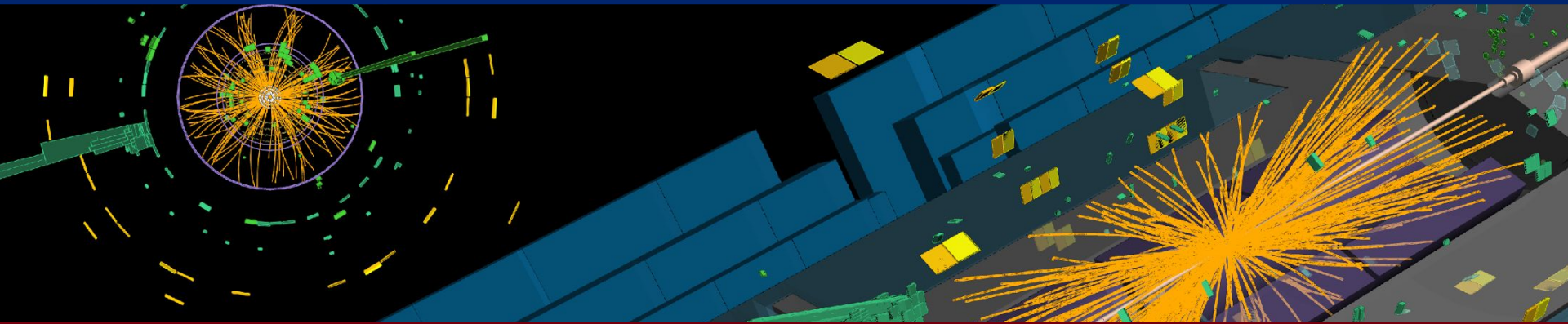


# Comparing unfolding methods



DIANA-HEP Topical Meeting

13 Nov 2017

Chris Meyer



# Introduction

- Many unfolding methods available and used in analyses
  - Overview of common methods in RooUnfold by Tim
  - Bin-by-bin, D'Agostini iterative, SVD, IDS, matrix method, ...
- Different methods have different benefits
  - Bin-by-bin: simple, although often avoided due to larger bias
  - Matrix inversion: minimally biased, but result often oscillates
  - Regularized: SVD, D'Agostini iterative, IDS
    - Avoids oscillating result while keeping minimal bias
- Important to compare when choosing for a given analysis
- In this talk:
  - A few things to think about when making the comparison

# Possible sources of unfolding biases

- MC simulation doesn't perfectly describe collected data
  - If it did, we wouldn't need to measure anything
- Regularization procedure
  - E.g. Makes constraints on smoothness of spectra
  - Overconstraints can lead to biases
- Large bin migrations can amplify biases
  - From either of the first points above

# General shape bias

- To compare methods, define a bias metric
- Simple to do with MC simulation
  - Unfold a reco spectra, compare with known truth spectra
  - Avoid identical MC simulation used for migration matrix...
    - Assuming no bugs → expect minimal bias (trivial closure)
- For robust test, consider different input/output pairs

# Data-driven method

- Described in article by [B. Malaescu](#), [used in ATLAS](#)
- Change the shape of the input MC to better match data
  - Change the truth shape using a smooth function
  - Propagate this change to the reconstructed spectra
    - Continue iterating until reconstructed spectra looks like data
  - Result: **data-like reconstructed spectra, with matching truth**
- Unfold data-like reconstructed spectra
  - Use the nominal migration matrix
- Compare **unfolded data-like spectra** with **expected truth**
- Resulting difference is the bias

# Other shape differences

- Can also consider different MC reweightings
  - Reweighting of different kinematic variables (such as [Higgs  \$p\_T\$](#) )
    - Different than the one being unfolded
  - Often choose one that's known to be mismodelled
- Other shape differences relevant to analysis?

## Choose the shape difference appropriate for your analysis

- Precision standard model measurement?
  - Perhaps in-situ / alternate variable shape reweighting
- Spectra that might include new physics?
  - Likely other shape differences that could be important

**The idea is to consider a range of possibilities, then determine which unfolding method performs best**

# Statistical uncertainty

- The amount of regularization affects the statistical uncertainty
  - Smoothness constraint
  - Iterative procedures
- Consider realistic statistical uncertainty in bias tests
  - Throw psuedo-spectrum from data-like reconstructed spectra
    - Consider expected Poisson uncertainty, not MC uncertainty
  - Unfold and compare to true spectra
  - Throwing many psuedo-spectra provides the covariance matrix

# What to do with the numbers?

- The previous slides provide a bias and stat unc
- Not obvious how to combine this information...
  - And very much depends on the situation
- A few general thoughts on method choice:
  - Often prefer bias < stat unc in most bins
  - Statistical uncertainty doesn't increase/decrease too much
    - Both individual bins, and sum of bins (see next point)
  - Statistical correlations need to be considered
    - Migration methods introduce correlations between bins
    - Simple quadrature sum of each bin no longer correct



# Summary

- Important to compare different unfolding methods
  - Choice is highly analysis dependant
  - Also compare different regularization for the same method
- Things to consider when choosing:
  - Size of unfolding bias (for reasonable shape difference)
  - Change in statistical error
  - Resulting correlations between bins