Tools for Unbinned Unfolding

Ryan Milton

AI4EIC 2025

In collaboration with: V. Mikuni, T. Lee, M. Arratia, T. Wamorkar, B. Nachman

UC RIVERSIDE

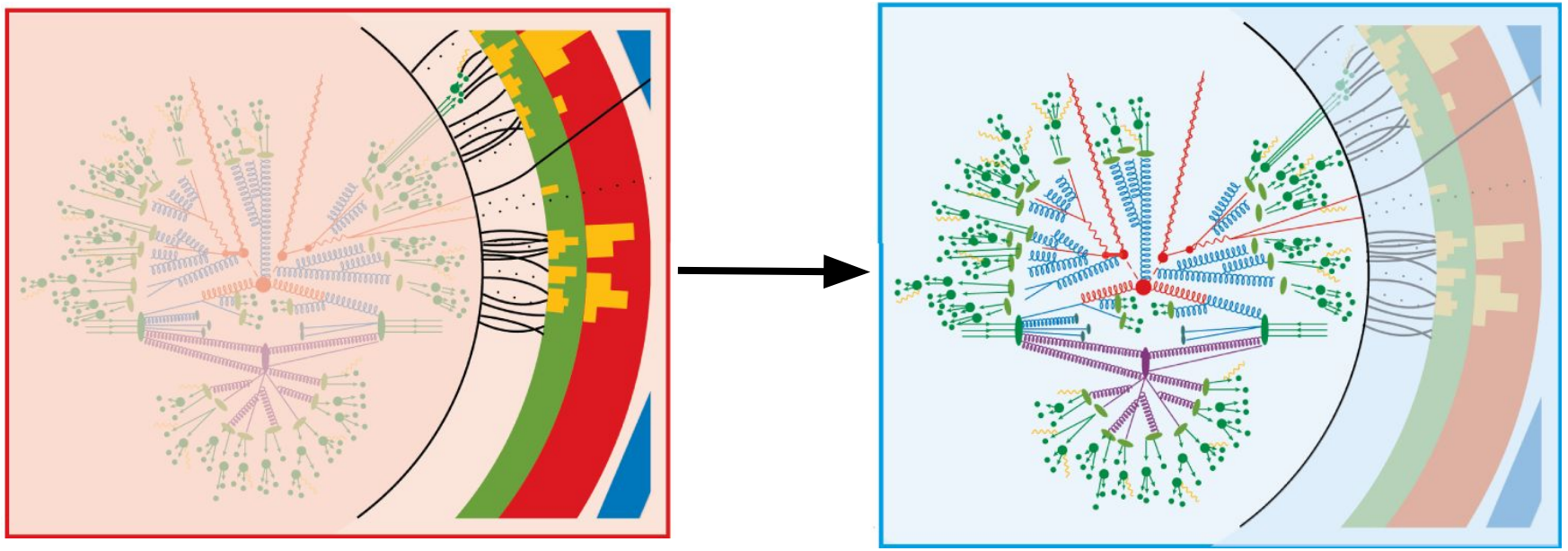
BERKELEY LAB

SLAC NATIONAL ACCELERATOR LABORATORY

NSF CSSI 2311666

Overview of unfolding

- Objective: Remove detector distortions from experimental data



Experimental data

Physics information

Binned unfolding

$$m = Rt$$

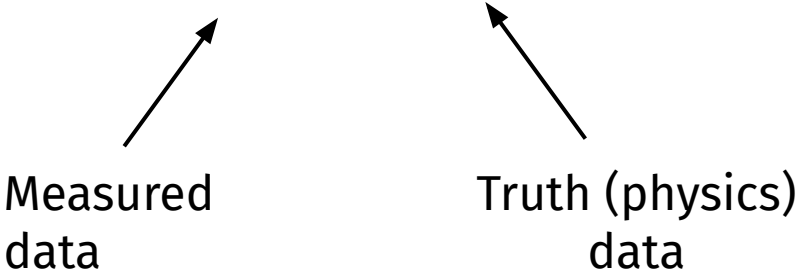
Binned unfolding

$$m = Rt$$

Measured data

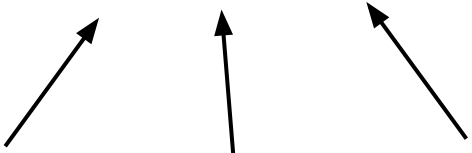
Binned unfolding

$$m = Rt$$



Binned unfolding

$$m = Rt$$



Measured data

Truth (physics) data

Response matrix

R_{ij} = Pr(measure in bin i | truth in bin j)

Made with simulation data

Binned unfolding

$$m = Rt$$

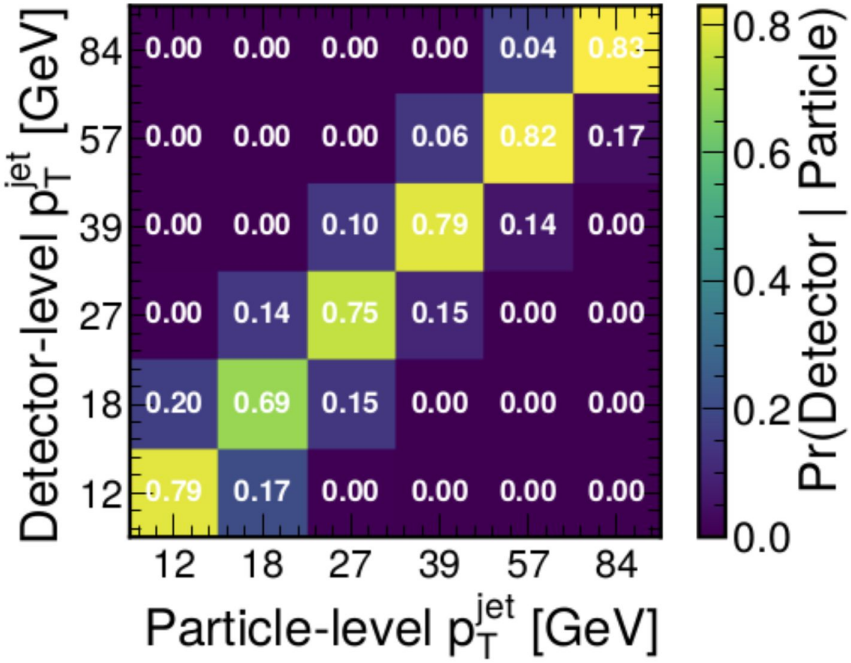
Measured data

Truth (physics) data

Response matrix

R_{ij} = Pr(measure in bin i | truth in bin j)

Made with simulation data

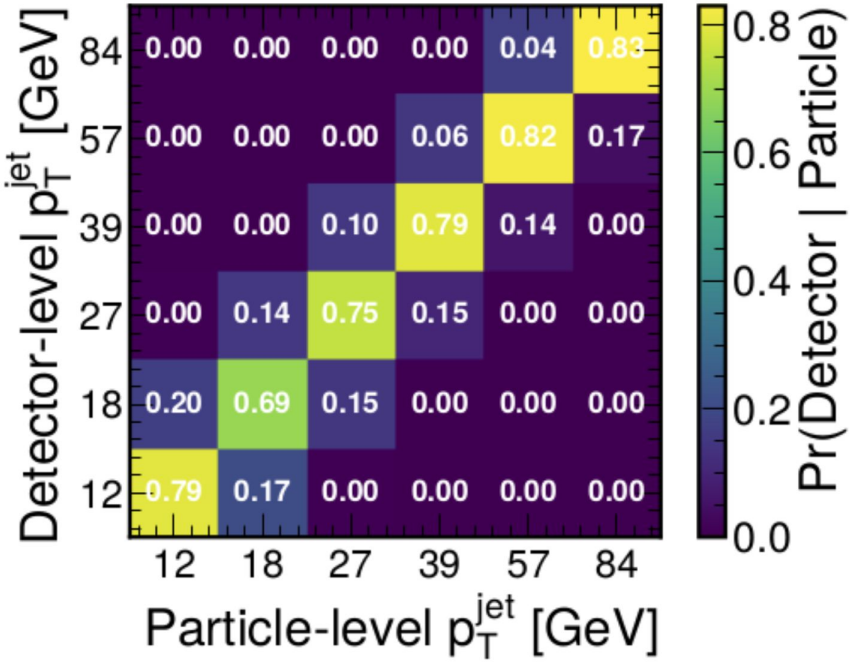


Example of response matrix

Binned unfolding

$$m = Rt$$

$$t = R^{-1}m$$



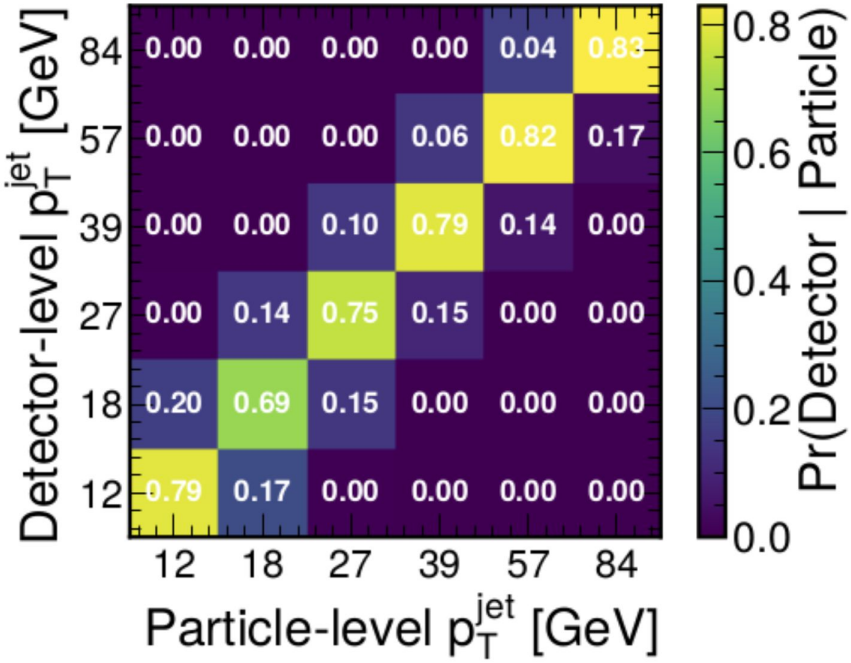
Example of response matrix

Binned unfolding

$$m = Rt$$

$$t = R^{-1}m$$

Main idea of binned unfolding is to ~invert the response matrix!



Example of response matrix

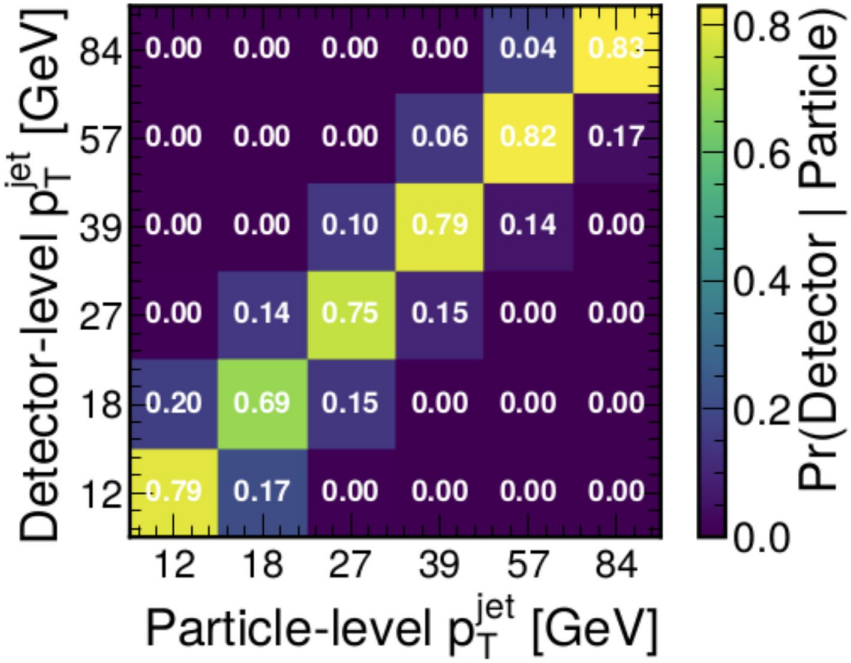
Binned unfolding

$$m = Rt$$

Examples:
Singular value decomposition,
Iterative Bayesian unfolding

$$t = R^{-1}m$$

Main idea of binned unfolding is to ~invert the response matrix!



Example of response matrix

Drawbacks of binned unfolding

Drawbacks of binned unfolding

- How to define optimal binning?
  - Must be chosen before unfolding procedure
  - Binning choices makes it difficult to compare between experiments and to publish data

Drawbacks of binned unfolding

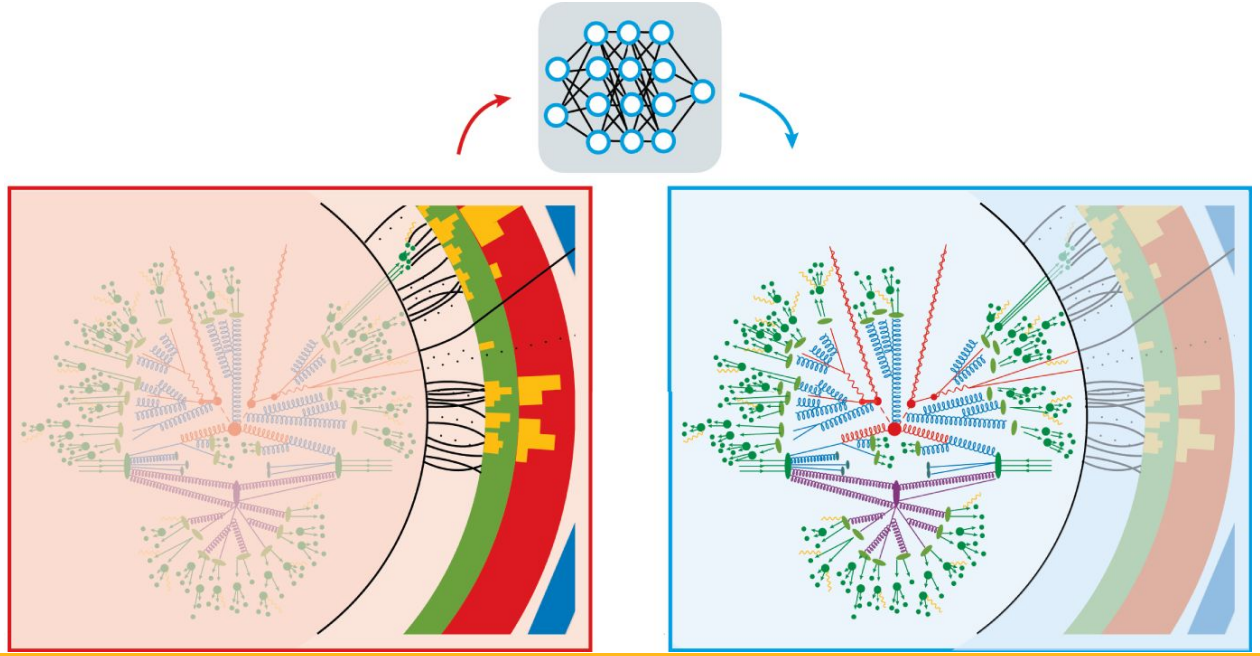
- How to define optimal binning?
  - Must be chosen before unfolding procedure
  - Binning choices makes it difficult to compare between experiments and to publish data
- Difficult to scale histograms to multiple dimensions by including multiple distributions

Drawbacks of binned unfolding

- How to define optimal binning?
  - Must be chosen before unfolding procedure
  - Binning choices makes it difficult to compare between experiments and to publish data
- Difficult to scale histograms to multiple dimensions by including multiple distributions
- Integrate over quantities thought to be irrelevant for a specific analysis
  - For other observables, like those motivated by future theoretical insight, may have to repeat analysis from scratch

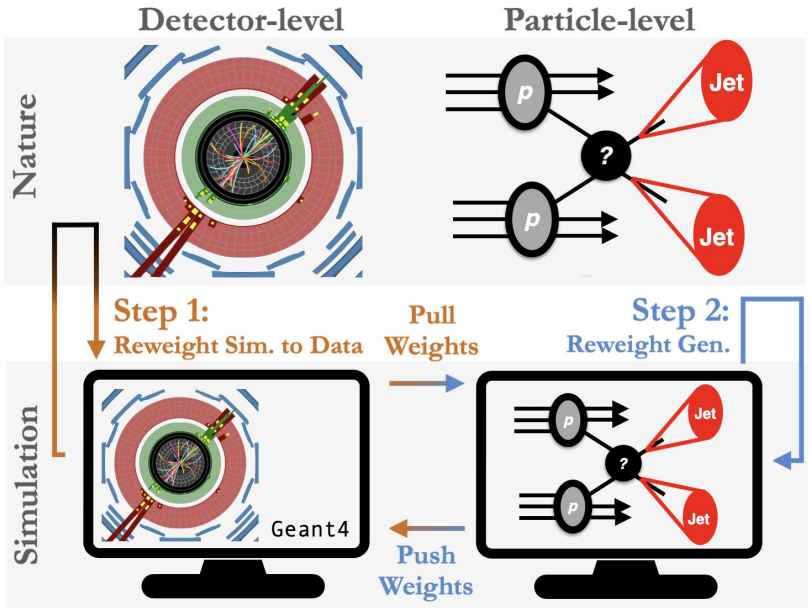
Unbinned unfolding motivation

- Motivates an unbinned unfolding method using machine learning
- Naturally unbinned and can handle high dimensions



Unbinned unfolding with OmniFold

- Iterative Bayesian Unfolding (IBU) is a popular binned unfolding algorithm
- Each iteration can be broken down into two steps
- In each step, a likelihood ratio is approximated
- Can estimate these two ratios using classifiers instead
- Use these likelihood ratios as unfolding weights!



Adapted from A. Andreassen, et al., Phys. Rev. Lett. **124**, 182001 (2020)

Applications of OmniFold

- OmniFold has been applied to many experimental analyses!

PHYSICAL REVIEW LETTERS **128,** 132002 (2022)

Measurement of Lepton-Jet Correlation in Deep-Inelastic Scattering with the H1 Detector Using Machine Learning for Unfolding

Measurement of event shapes in minimum bias events from pp collisions at 13 TeV

The CMS Collaboration

PHYSICAL REVIEW LETTERS **133,** 261803 (2024)

Simultaneous Unbinned Differential Cross-Section Measurement of Twenty-Four Z + jets Kinematic Observables with the ATLAS Detector

G. Aad *et al.**
(ATLAS Collaboration)

PHYSICAL REVIEW D **108,** L031103 (2023)

Multidifferential study of identified charged hadron distributions in Z-tagged jets in proton-proton collisions at $\sqrt{s} = 13$ TeV

R. Aaij *et al.**
(LHCb Collaboration)

Measurement of CollinearDrop jet mass and its correlation with SoftDrop groomed jet substructure observables in $\sqrt{s} = 200$ GeV pp collisions by STAR

Youqi Song (Wright Laboratory, Yale University)

on behalf of the STAR Collaboration

Status of unbinned unfolding tools

- PyPi omnifold:
  - Architectures: Multilayer perceptron (MLP) , point-edge transformer (PET)
  - Any number of dimensions
  - May need GPUs to train
- RooUnfold inspired omnifold (also has a Python interface):
  - Architecture: Boosted decision tree (BDT)
  - Low dimensional unfolding
  - Simple set-up: No data preprocessing, no GPUs needed


PAPER

Tools for unbinned unfolding

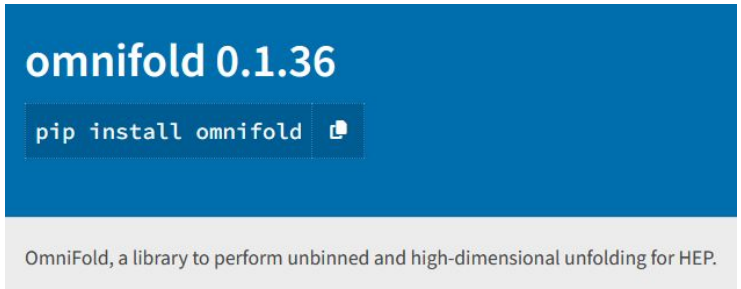
Ryan Milton, Vinicius Mikuni, Trevin Lee, Miguel Arratia, Tanvi Wamorkar and Benjamin Nachman

Published 23 May 2025 • © 2025 IOP Publishing Ltd and Sissa Medialab. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Journal of Instrumentation, Volume 20, May 2025

Citation Ryan Milton *et al* 2025 *JINST* **20** P05034

DOI 10.1088/1748-0221/20/05/P05034



Application to jet measurements

- Test unfolding quality using closure test
  - Use two Monte Carlo simulations so we know what the truth looks like!
- Use Pythia as Monte Carlo (Gen/Sim), Herwig as pseudodata (“Truth”/“Data”) to get unfolded jet observables
- Using Delphes fast simulation of CMS detector

Application to jet measurements

- Test unfolding quality using closure test
  - Use two Monte Carlo simulations so we know what the truth looks like!
- Use Pythia as Monte Carlo (Gen/Sim), Herwig as pseudodata (“Truth”/“Data”) to get unfolded jet observables
- Using Delphes fast simulation of CMS detector

Method	**ML type**	**Input**	**Unfolding type**
OmniFold	PET	All particles	Unbinned

Application to jet measurements

- Test unfolding quality using closure test
  - Use two Monte Carlo simulations so we know what the truth looks like!
- Use Pythia as Monte Carlo (Gen/Sim), Herwig as pseudodata (“Truth”/“Data”) to get unfolded jet observables
- Using Delphes fast simulation of CMS detector

Method	**ML type**	**Input**	**Unfolding type**
OmniFold	PET	All particles	Unbinned
MultiFold	BDT, MLP (DNN)	All jet observables	Unbinned

Application to jet measurements

- Test unfolding quality using closure test
  - Use two Monte Carlo simulations so we know what the truth looks like!
- Use Pythia as Monte Carlo (Gen/Sim), Herwig as pseudodata (“Truth”/“Data”) to get unfolded jet observables
- Using Delphes fast simulation of CMS detector

Method	**ML type**	**Input**	**Unfolding type**
OmniFold	PET	All particles	Unbinned
MultiFold	BDT, MLP (DNN)	All jet observables	Unbinned
UniFold	BDT	Individual jet observables	Unbinned

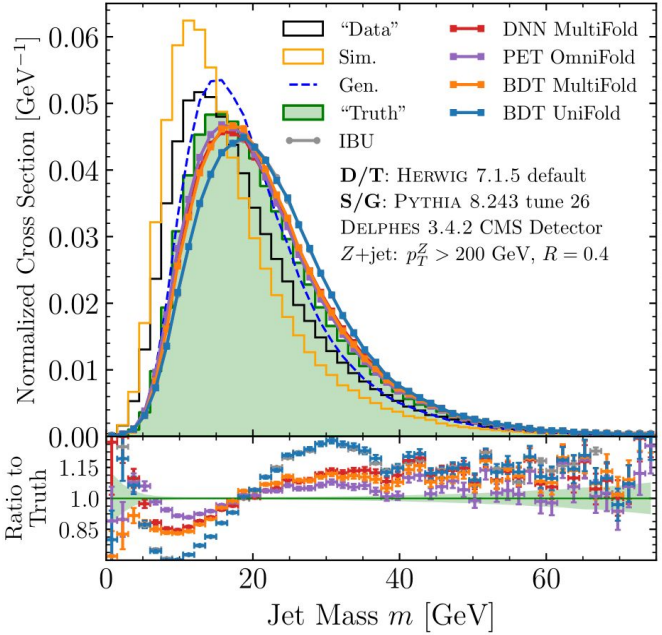
Application to jet measurements

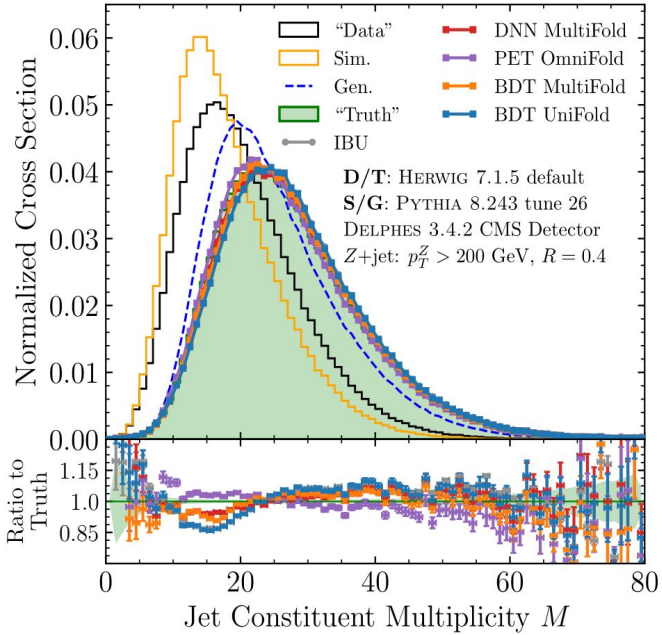
- Test unfolding quality using closure test
  - Use two Monte Carlo simulations so we know what the truth looks like!
- Use Pythia as Monte Carlo (Gen/Sim), Herwig as pseudodata (“Truth”/“Data”) to get unfolded jet observables
- Using Delphes fast simulation of CMS detector

Method	**ML type**	**Input**	**Unfolding type**
OmniFold	PET	All particles	Unbinned
MultiFold	BDT, MLP (DNN)	All jet observables	Unbinned
UniFold	BDT	Individual jet observables	Unbinned
IBU	N/A	Individual jet observables	Binned

Performance on jet observables

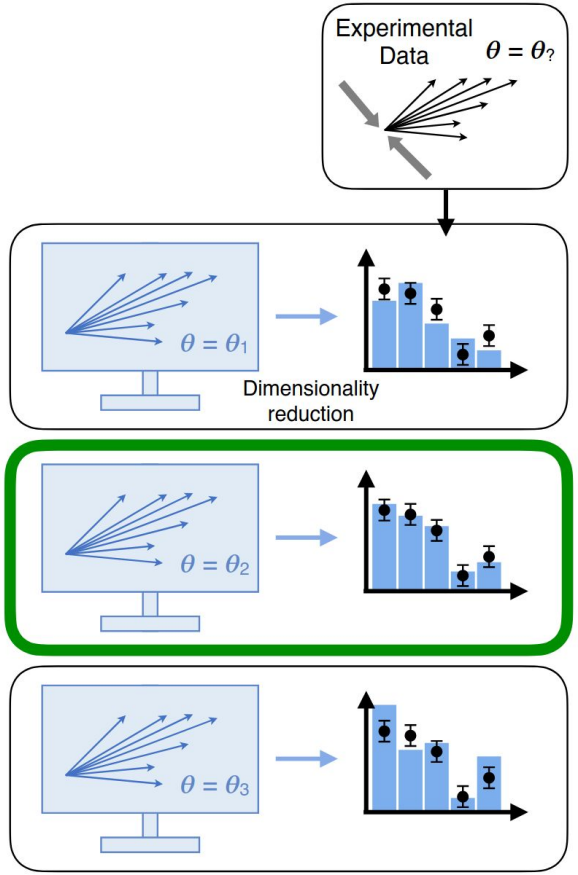
- Trying to match the “Truth” distributions by applying weights to Gen
- Other jet observables in paper: Width, Soft Drop mass, N-subjettiness ratio, groomed jet momentum fraction





Binned inference

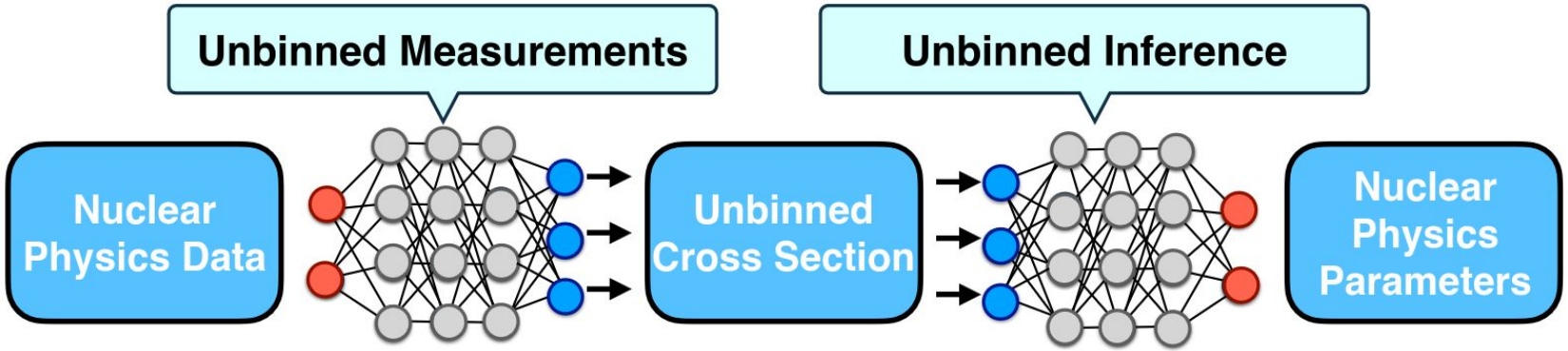
- To go from measurements to matching with theory, need to do inference
- Typically, a functional form is assumed and parameters are fit
  - Computationally expensive and depends on binning



Conventional inference

ML-based inference

- New approach: Train a model to output probability of data given parameters of PDFs and correlations
- Use this model with unbinned unfolded data!

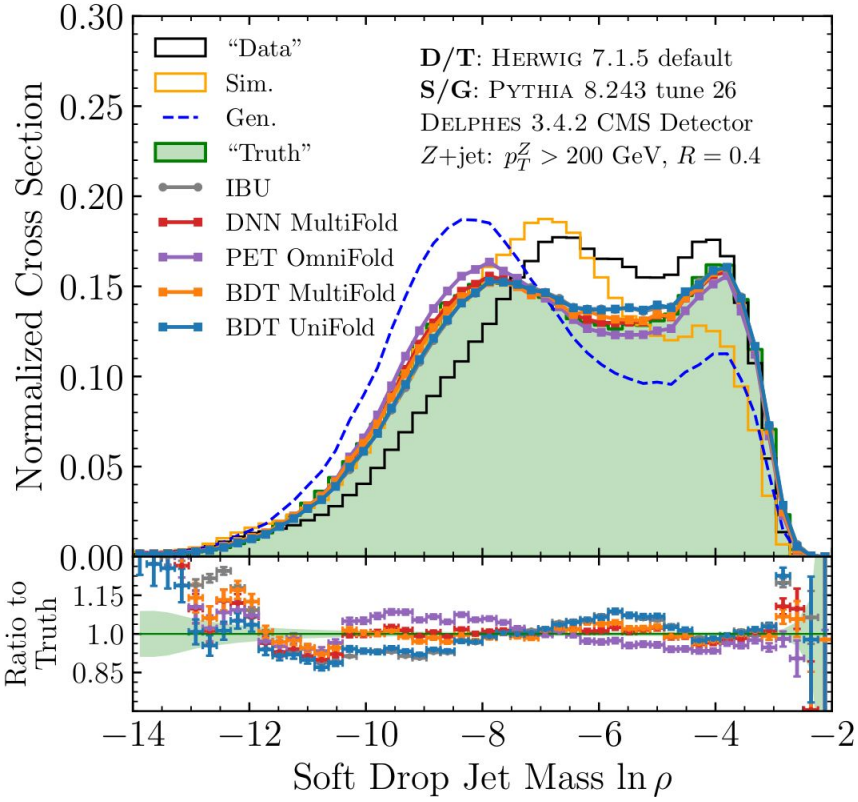


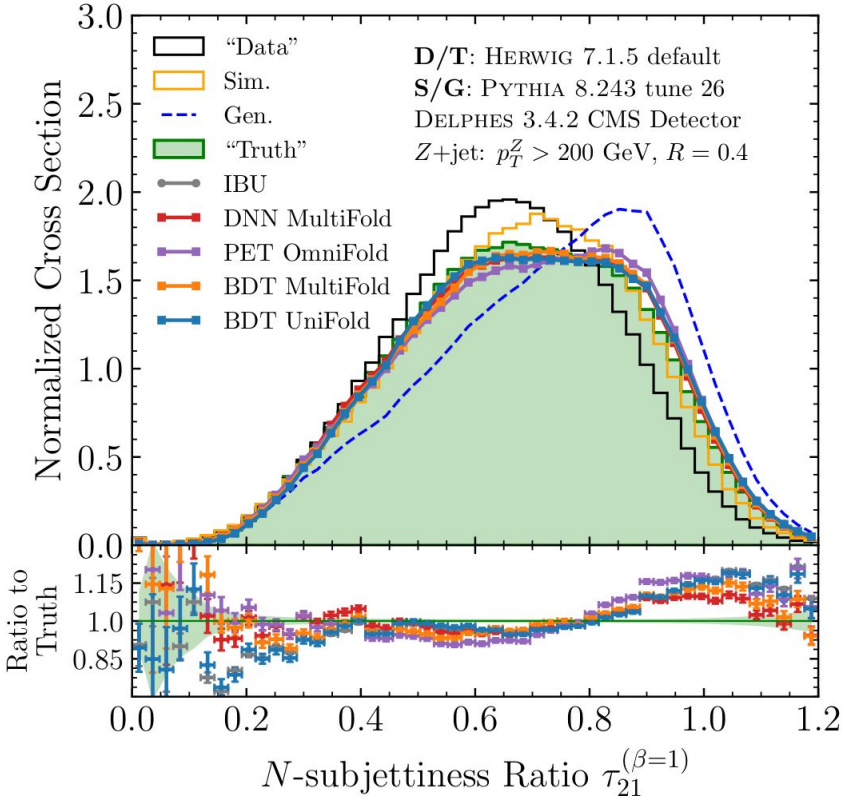
Summary and outlook

- Binned unfolding is limited, motivating an ML unbinned unfolding approach
  - With ML: Maintain correlations, high dimensional unfolding, no reliance on bin definitions. Also better for data preservation.
- We have developed powerful and easy to use tools to do unbinned unfolding
- These tools prove effective in unfolding jet observables in comparison to traditional binned methods
- Can create a fully unbinned pipeline with unbinned inference
- Incorporation of BDT OmniFold into RooUnfold is ongoing

Thank you!

Backup: Other jet observables





Backup: Other jet observables