University of Zurich[UZH]

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

ML workshop in CHIPP 2024 Annual meeting

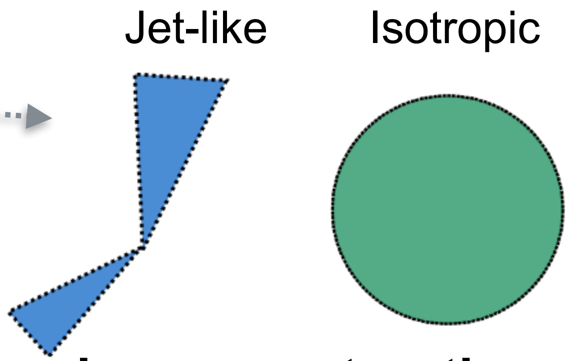
Weijie Jin, Kyle Cormier, Florencia Canelli

Event shape observables in proton-proton collisions

Event shape observables:

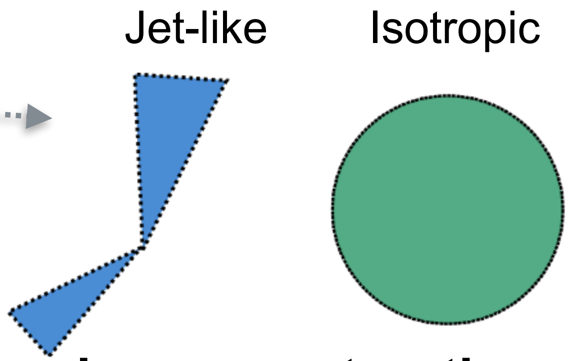
Variables describing the "**shapes** " of the events

→ Functions of the momentum of the final state particles



We focus on **charged particles** ← **precise reconstruction of tracks**

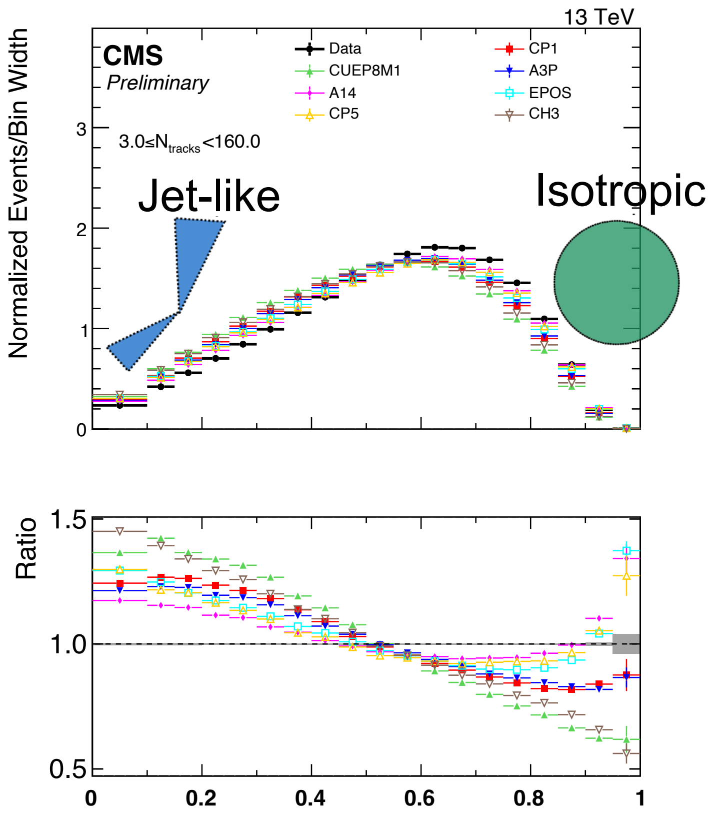
Event shape observables in proton-proton collisions



Event shape observables:

Variables describing the “**shapes** ” of the events

→ Functions of the momentum of the final state particles

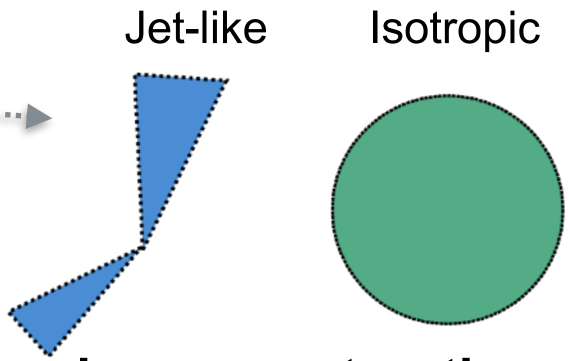


S_T^{tracks} (high purity, pt>0.5 GeV)

We focus on **charged particles** ← **precise reconstruction of tracks**

← An example: transverse sphericity

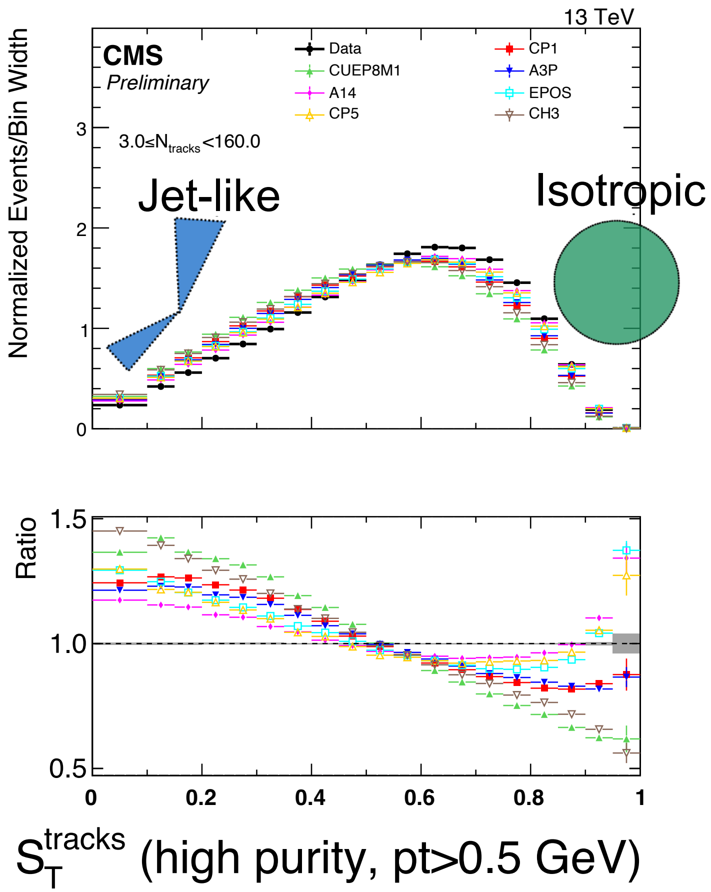
Event shape observables in proton-proton collisions



Event shape observables:

Variables describing the "**shapes** " of the events

→ Functions of the momentum of the final state particles



We focus on **charged particles** ← **precise reconstruction of tracks**

← An example: transverse sphericity

Observables for measurement:

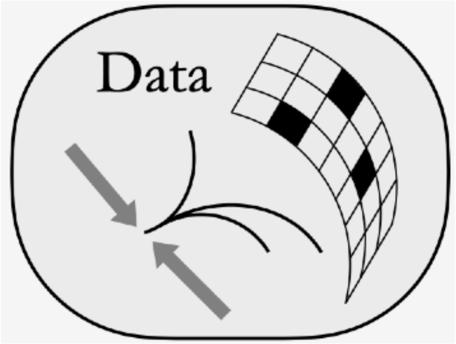
From charged particles p_T > 0.5 GeV, $|\eta|$ < 2.4

- Charged particle multiplicity
- Invariant mass of charged particles

- Sphericity (+ transverse)
- Thrust (+transverse)
- Broadening
- Isotropy

Machine-learning-based unfolding

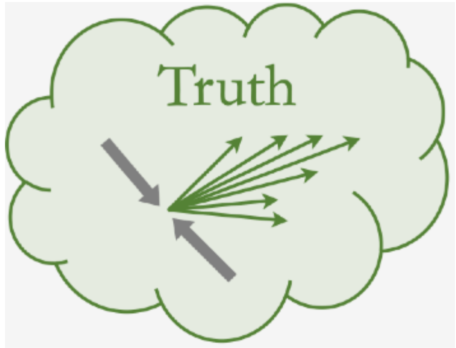
Unfold with a **machine-learning-based** algorithm: **Multifold***



Event shapes
of detector-level objects

arrays of multiplicity, sphericity, thrust …
of **tracks**

Multifold



Event shapes of particles

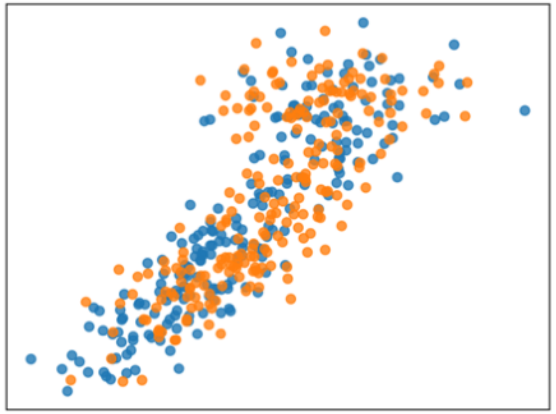
arrays of multiplicity, sphericity, thrust …
of **charged particles**

simultaneously unfold of 8 observables

→ theoretical interpretation, generator tuning …

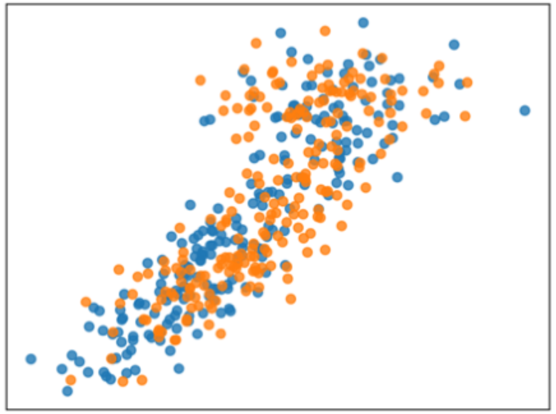
* https://arxiv.org/abs/1911.09107, https://arxiv.org/abs/2105.04448

Unbinned multi-dimensional unfolding and uncertainty estimation



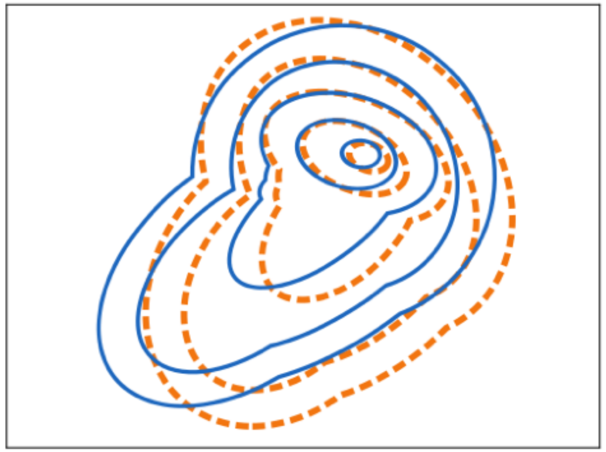
← A typical binary classifier to distinguish two sets

Unbinned multi-dimensional unfolding and uncertainty estimation

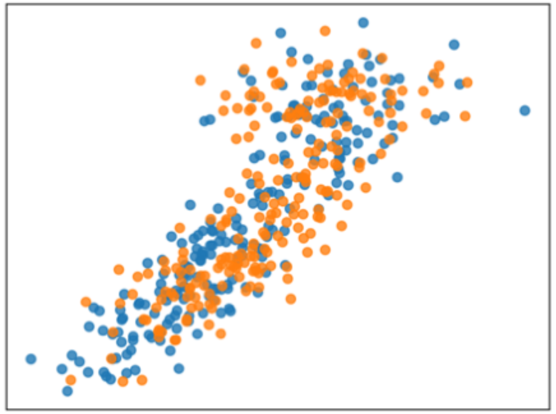


← A typical binary classifier to distinguish two sets

What it actually did: learn the differences in the distributions →
(likelihood ratio)

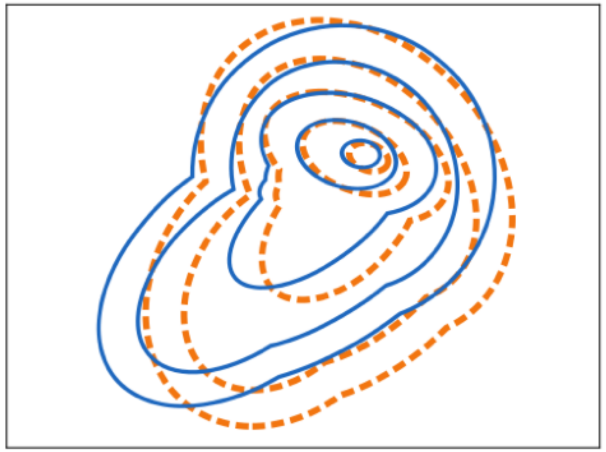


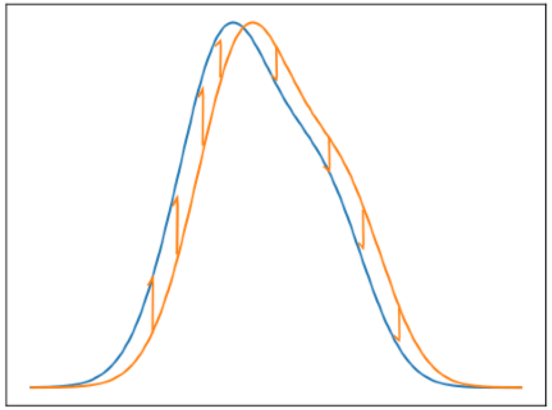
Unbinned multi-dimensional unfolding and uncertainty estimation



← A typical binary classifier to distinguish two sets

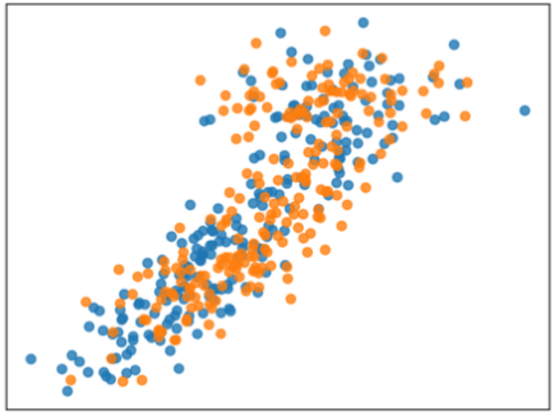
What it actually did: learn the differences in the distributions →
(likelihood ratio)





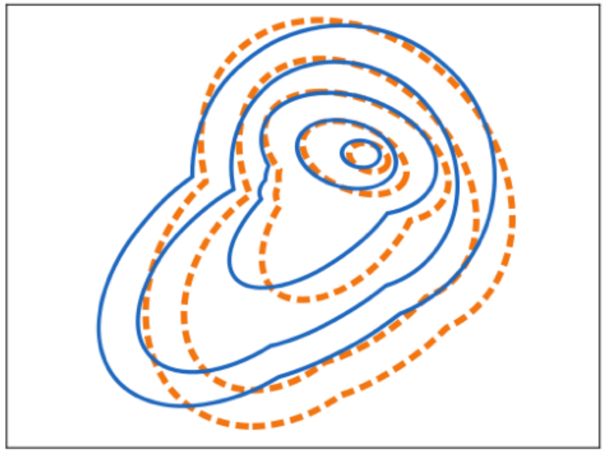
← We can use the classification scores to weight **MC** to **data**, and **nominal sample** to **systematic variations**

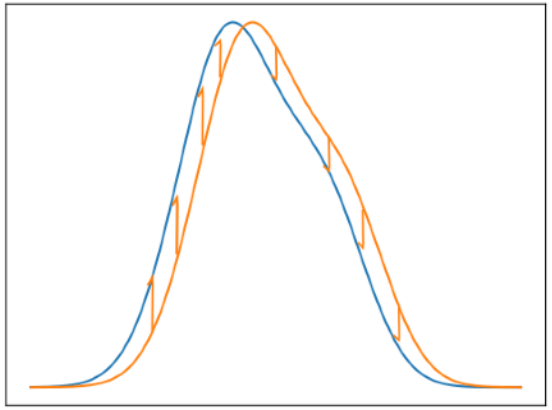
Unbinned multi-dimensional unfolding and uncertainty estimation



← A typical binary classifier to distinguish two sets

What it actually did: learn the differences in the distributions → (likelihood ratio)





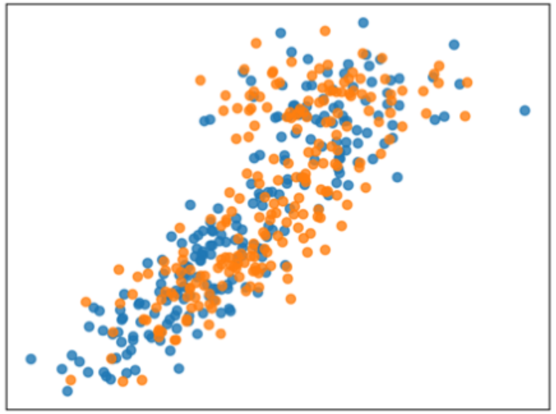
← We can use the classification scores to weight **MC** to **data**, and **nominal sample** to **systematic variations**

Event-wise unfolding → the result independent of binning

The actual unfolding in iterations:

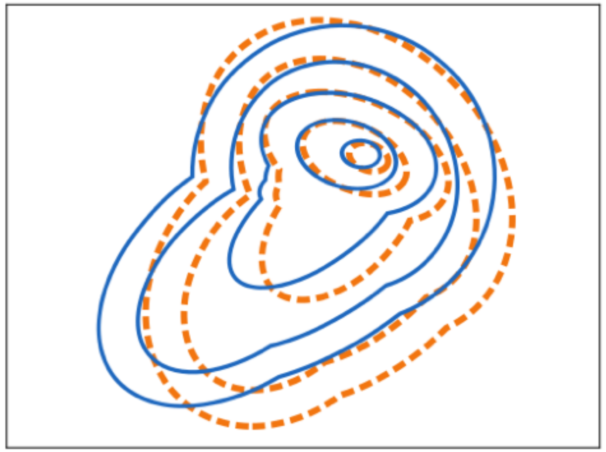
- Step 1: weight **MC** to **data**, at detector level
- Step 2: pull back the weights to particle(truth) level
- Extra 2 steps added to deal with the selection efficiency and signal acceptance

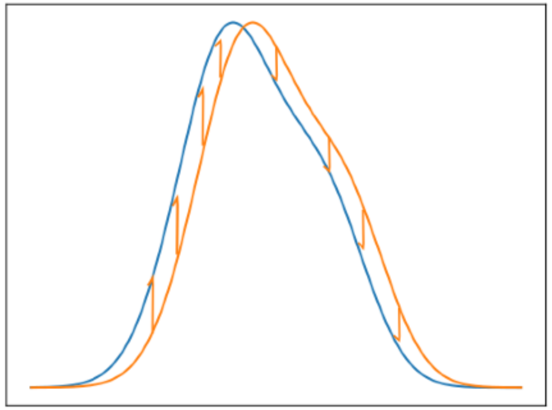
Unbinned multi-dimensional unfolding and uncertainty estimation



← A typical binary classifier to distinguish two sets

What it actually did: learn the differences in the distributions → (likelihood ratio)





← We can use the classification scores to weight **MC** to **data**, and **nominal sample** to **systematic variations**

Event-wise unfolding → the result independent of binning

The actual unfolding in iterations:

- Step 1: weight **MC** to **data**, at detector level
- Step 2: pull back the weights to particle(truth) level
- Extra 2 steps added to deal with the selection efficiency and signal acceptance

Event-wise uncertainty template → unbinned unfolding uncertainty & covariance

Uncertainty template construction with ML-based weighting

Example: **Mismodelling of observables used directly in unfolding** (bias from unfolding prior)
→ Uncertainty estimated from modelling differences between **nominal** and **alternative MC**

Derive the templates by **weighting nominal MC** to **alternative MC** at the **particle-level**
→ output: **weighted nominal MC events**
- same **particle-level distribution** as **alternative MC**
- keeps the **gen. → reco. migration** of the **nominal MC**

Uncertainty template construction with ML-based weighting

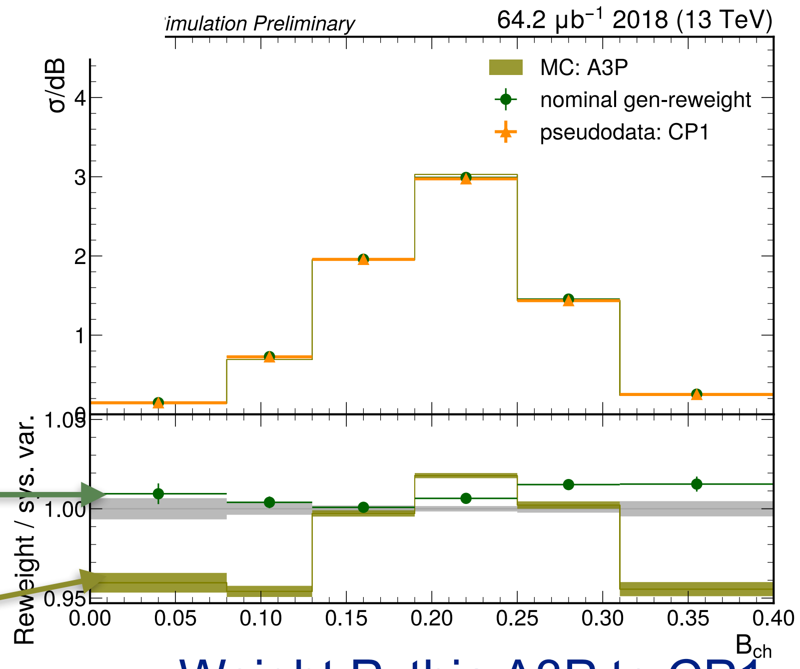
Example: **Mismodelling of observables used directly in unfolding** (bias from unfolding prior)
→ Uncertainty estimated from modelling differences between **nominal** and **alternative MC**

Derive the templates by **weighting nominal MC** to **alternative MC** at the **particle-level**
→ output: **weighted nominal MC events**
- same **particle-level distribution** as **alternative MC**
- keeps the **gen. → reco. migration** of the **nominal MC**

Example:
particle-level broadening
before & **after** weighting

After reweighting at the gen-level

Nominal MC



Weight Pythia A3P to CP1

Uncertainty template construction with ML-based weighting

Example: **Track reconstruction efficiency** uncertainty

- Step1: Randomly drop 2.1%(1%) tracks with pT<20 GeV (>20 GeV) in nominal MC*
- Step2: **weight** the nominal MC to Step1 output at **particle-** and **detector-level**

* The uncertainty of track reco. eff. is given by D* analysis: https://cds.cern.ch/record/2810814/

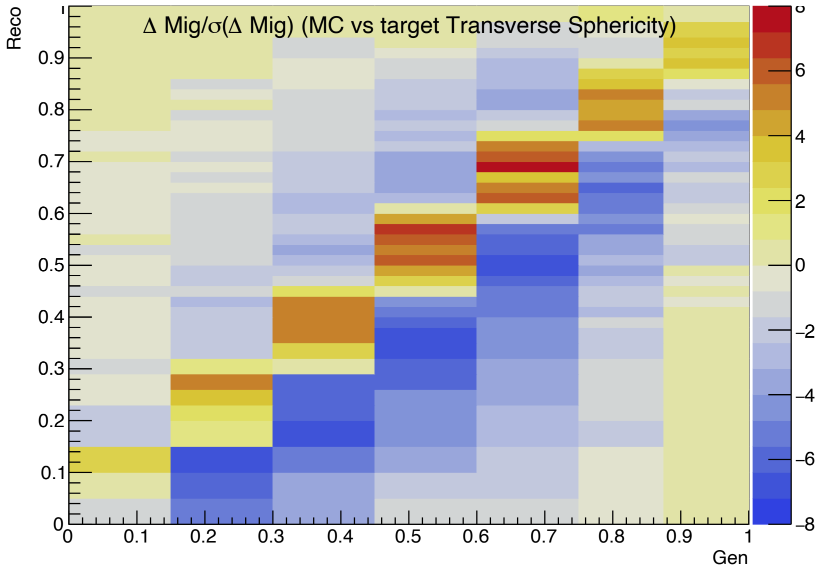
Uncertainty template construction with ML-based weighting

Example: **Track reconstruction efficiency** uncertainty

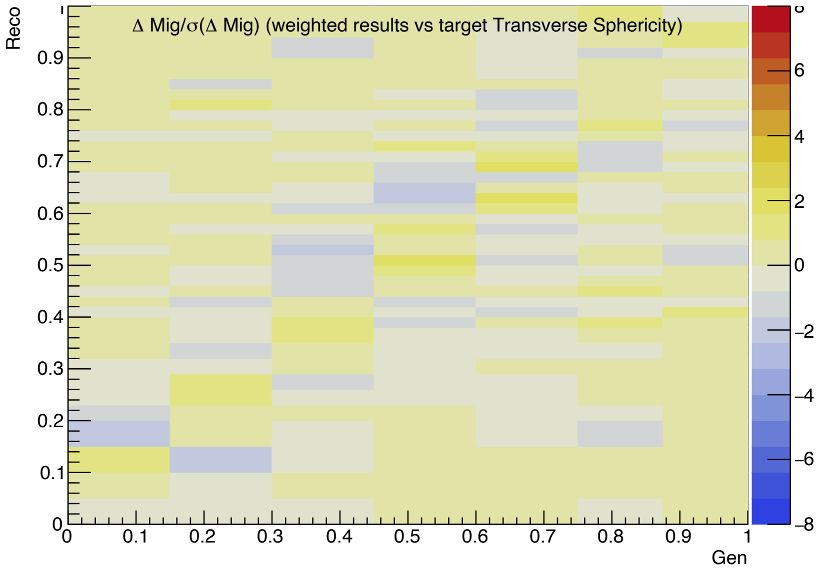
- Step1: Randomly drop 2.1%(1%) tracks with pT<20 GeV (>20 GeV) in nominal MC*
- Step2: **weight** the nominal MC to Step1 output at **particle-** and **detector-level**

Example:
Gen → reco migration
of transverse sphericity

Difference between nominal **MC** and **target**
before weighting



After weighting

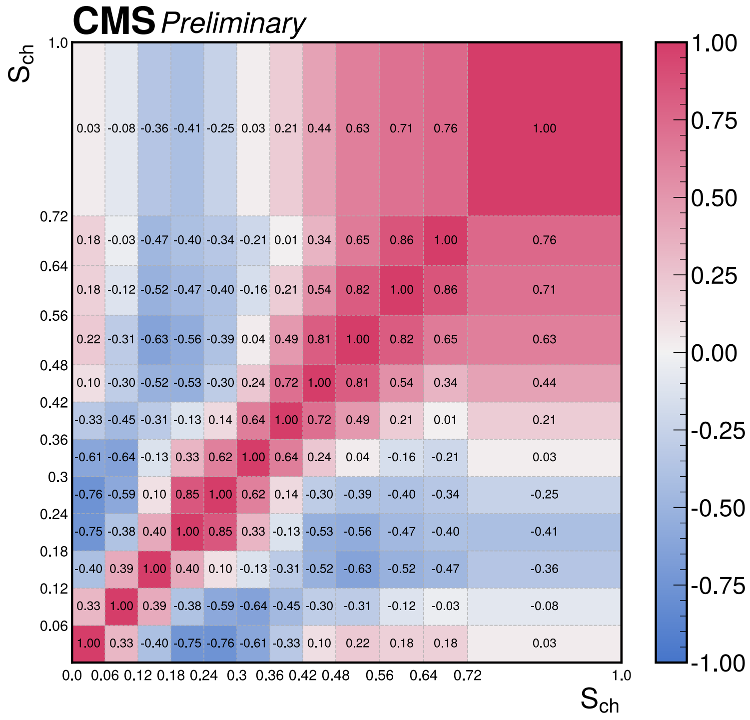


* The uncertainty of track reco. eff. is given by D* analysis: https://cds.cern.ch/record/2810814/

Unbinned uncertainty estimation

ML-based reweighting → **Uncertainty templates** as sets of **weights on nominal MC**
→ **Continuous nuisance** parameters can be assigned to the **event-weights**
→ Uncertainty **covariance** can be estimated from **toy experiments**

- Unfold with **“bootstraps” of MC** with **variations of nuisance parameters → Syst. Unc + Covariance**
- Unfold with “**bootstraps” of resampled data → Stat. Unc. + Covariance**



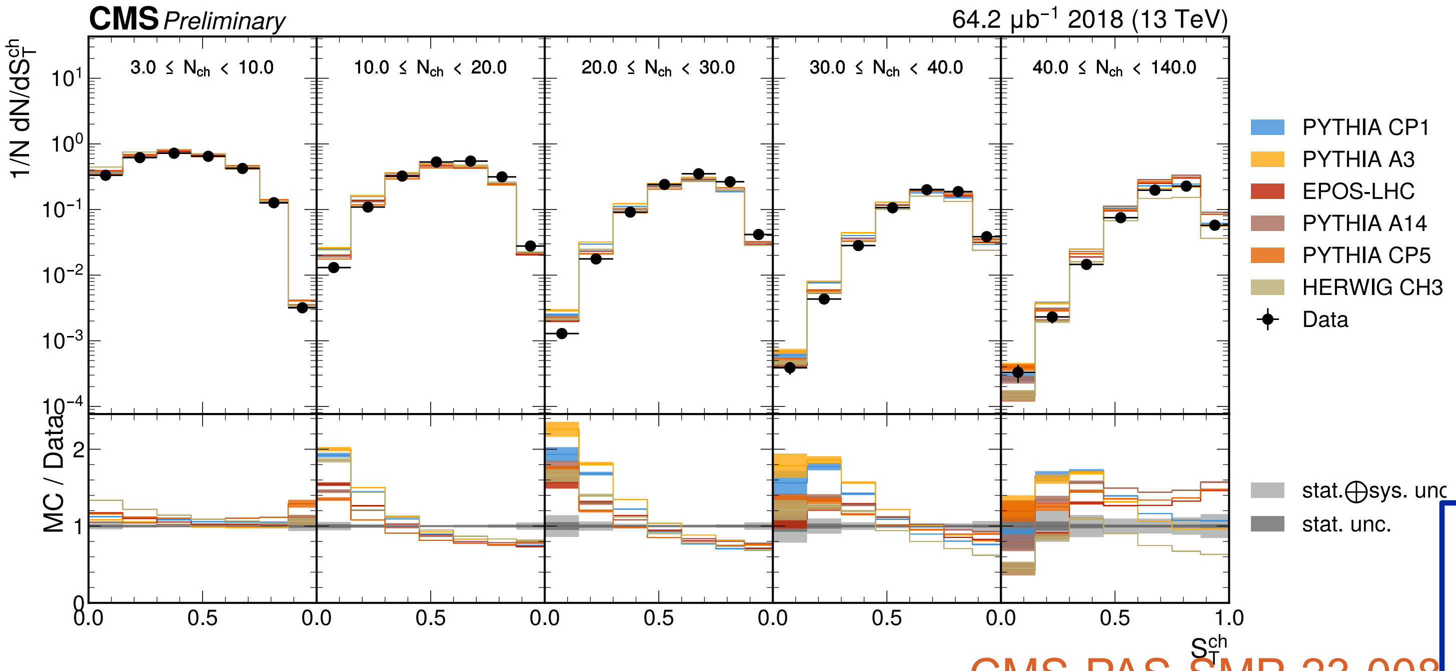
Uncertainties+Covariance on the event-wise unfolded data

CMS-PAS-SMP-23-008

Example: correlation of the syst. unc. of sphericity

Unfolding results

Unfolding results as **weighted MC events**



CMS-PAS-SMP-23-008

Example: 2D visualisation of transverse sphericity in charged particle multiplicity slices

Customise binning and **variable choices** are supported with the **event-wise unfolded data**

Simultaneously unfold all the variables for ML-based weighting

Add a variable to the unfolding:

Methods based on **binned** histograms:

Add **another dimension** in binning

→ require **higher statistics**

→ more **computation** in simulation and unfolding

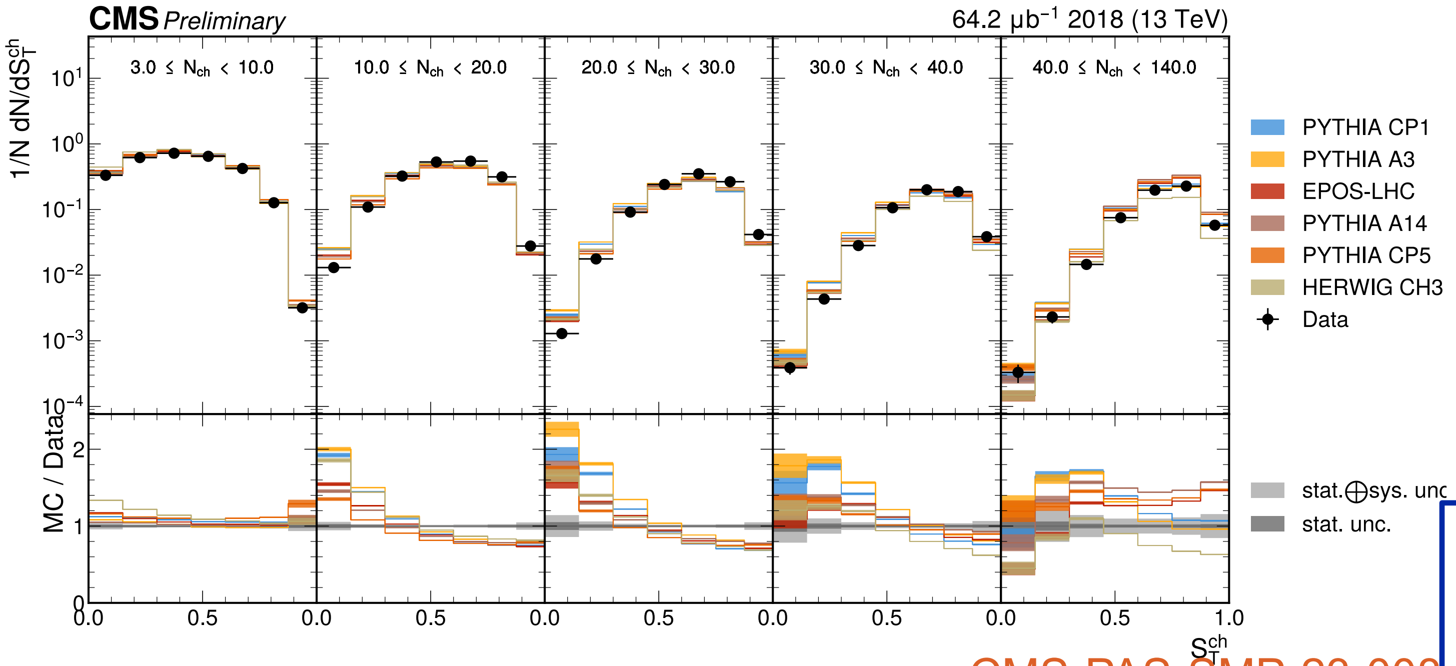
This method:

Add **a feature** in the ML training and evaluation

→ much easier to scale up the dimensions

Unfolding results

Unfolding results as **weighted MC events**



CMS-PAS-SMP-23-008

Example: 2D visualisation of transverse sphericity in charged particle multiplicity slices

More isotropic data than MC:
→ We provide the unfolded results for theoretical interpretation

Customise binning and **variable choices** are supported with the **event-wise unfolded data**

Simultaneously unfold all the variables for ML-based weighting

Add a variable to the unfolding:

Methods based on **binned** histograms:
Add **another dimension** in binning
→ require **higher statistics**
→ more **computation** in simulation and unfolding

This method:
Add **a feature** in the ML training and evaluation
→ much easier to scale up the dimensions

Summary

We have applied a novel **ML-based unfolding algorithm** (Multifold) to CMS analysis

- **Simultaneous** unfolding of **8 event shape observables**
- Unfold at **per-event** level

Customise binning and **variable choices** are supported with the **event-wise unfolded data**

Summary

We have applied a novel **ML-based unfolding algorithm** (Multifold) to CMS analysis

- **Simultaneous** unfolding of **8 event shape observables**
- Unfold at **per-event** level

Customise binning and **variable choices** are supported with the **event-wise unfolded data**

+

ML-based uncertainty template construction and **uncertainty estimation** at **per-event** level

Uncertainties+Covariance on the results

Summary

We have applied a novel **ML-based unfolding algorithm** (Multifold) to CMS analysis

- **Simultaneous** unfolding of **8 event shape observables**
- Unfold at **per-event** level

Customise binning and **variable choices** are supported with the **event-wise unfolded data**

+

ML-based uncertainty template construction and **uncertainty estimation** at **per-event** level

Uncertainties+Covariance on the results

The way to improve the usability of **unfolded results**

- Publish the **unbinned** results on **event-level**
- Publish the **weight sets** from **toy experiments**
  → **Unc. + Covariance**

Unbinned fit for **theoretical interpretation**
Unbinned generator tuning
(Or **any binning** chosen by the user)

Backup

Estimation of statistical uncertainty

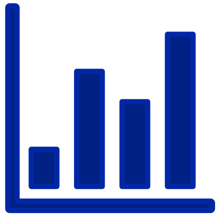
Statistical uncertainty from data

Estimation method: **pseudo-experiments**

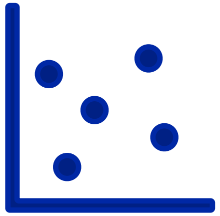
Unfold resampled data to estimate the effects on unfolding

- Assign the data events with **weights ~ Poisson(1)**
  → Alternative data samples (bootstraps) with statistical fluctuations
- **Unfold** these “**bootstrap**” data samples
  → a set of MC weights for each bootstrap
- **Standard deviation**s of these **unfolding results**
  → **unfolding stat. unc.**
- **Covariance** of the histograms of these results
  → covariance of the unfolding stat. unc.

Estimation of systematic uncertainty



Binned unfolding:
- **Systematic templates** as alternative MC **histograms**
- **Nuisance parameters** quantify the deviation from nominal MC to systematic template histograms



Extrapolate to unbinned unfolding:
- **Systematic templates** as alternative **weights on nominal MC** events (nominal: weight=1)
- **Nuisance parameters** quantify the deviation from the nominal weight 1 to the alternative weights

Aim in systematic uncertainty estimation:
- Construct the **templates as weights on nominal MC events**
- **Continuous nuisance parameters** applied on the weights
  → continuous deviation from nominal MC to systematic templates
  → enables uncertainty estimation with pseudo-experiments (unfolding with “bootstrap” MC)

Systematic uncertainty estimation based on unbinned reweighting

4. **Mismodelling of other observables which may change detector response**
   Derive the templates with two-step weighting

Systematic uncertainty estimation based on unbinned reweighting

4. **Mismodelling of other observables which may change detector response**
   Derive the templates with two-step weighting

- **Step 1**: weight the **alternative MC** to **nominal MC** at the **particle-level**
  → output: **weighted alternative MC**
    - with migration function of alternative MC
    - particle-level distributions of nominal MC

Systematic uncertainty estimation based on unbinned reweighting

4. **Mismodelling of other observables which may change detector response**
Derive the templates with two-step weighting

- **Step 1**: weight the **alternative MC** to **nominal MC** at the **particle-level**
  → output: **weighted alternative MC**
    - with migration function of alternative MC
    - particle-level distributions of nominal MC
- **Step 2**: weight the **nominal MC** to the **Step 1 output** at **particle- and detector-level**
  → output: **weighted nominal MC**
    - with migration function of alternative MC
    - particle-level distributions of nominal MC

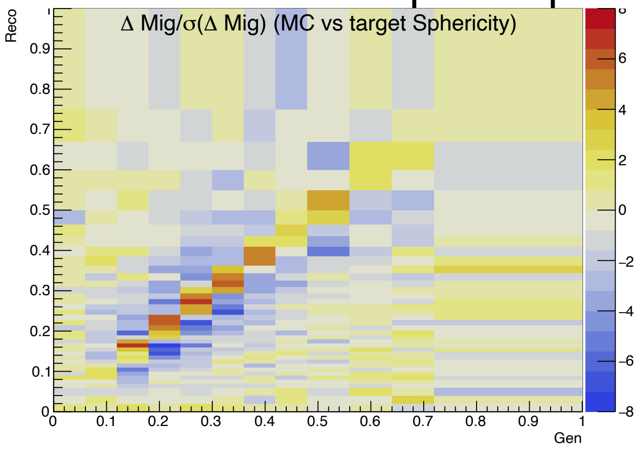
Systematic uncertainty estimation based on unbinned reweighting

4. **Mismodelling of other observables which may change detector response**
Derive the templates with two-step weighting

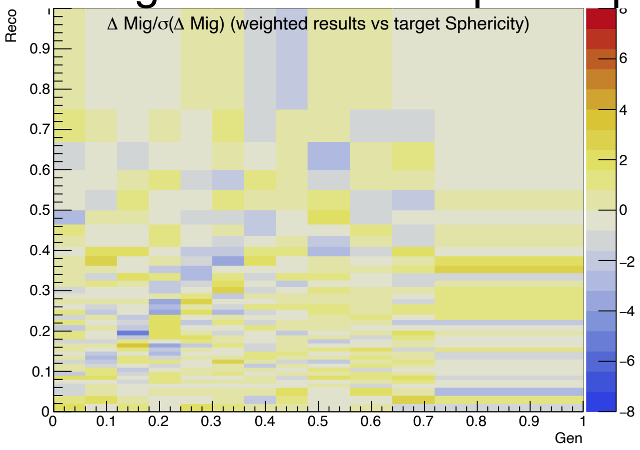
- **Step 1**: weight the **alternative MC** to **nominal MC** at the **particle-level**
  → output: **weighted alternative MC**
  - with migration function of alternative MC
  - particle-level distributions of nominal MC
- **Step 2**: weight the **nominal MC** to the **Step 1 output** at **particle- and detector-level**
  → output: **weighted nominal MC**
  - with migration function of alternative MC
  - particle-level distributions of nominal MC

Example:
Gen → reco migration
of spherocity

Before step 2 weighting:
nominal MC - step 1 output



After step 2 weighting:
weighted result - step 1 output

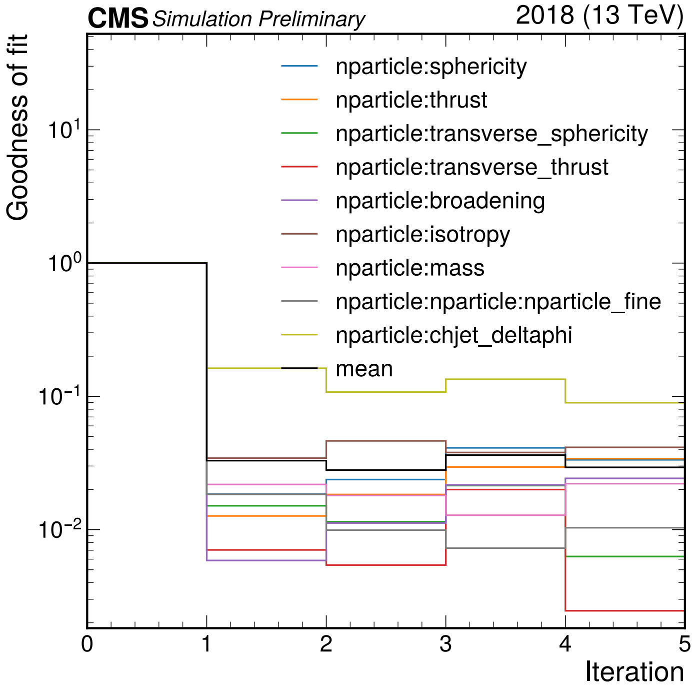


Validation of the unfolding

Validation: unfold the pseudo-data from Pythia A14 tune

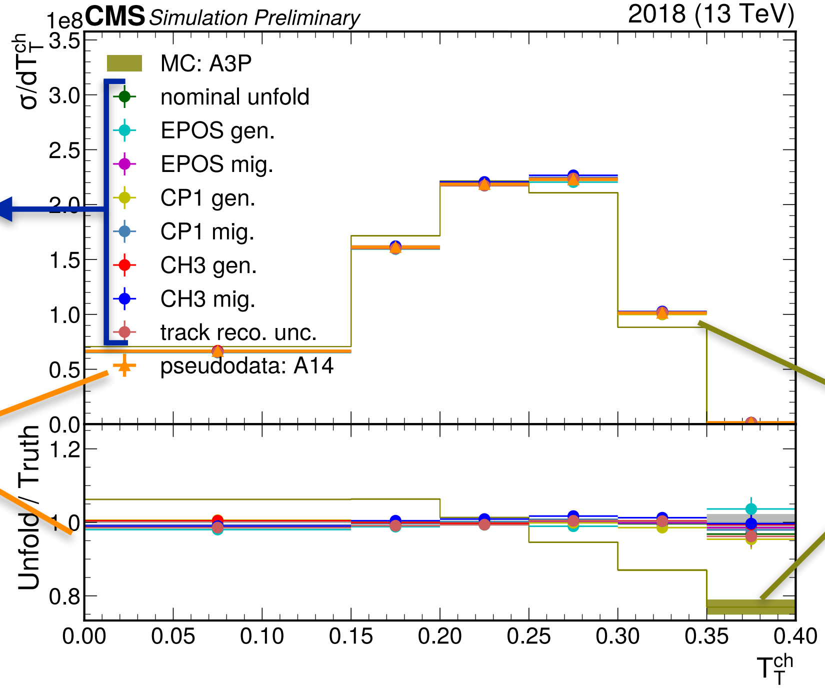
Alternative MC from Pythia A14, CP5 and CUETP8M1 tunes → **pseudo-data**

Unfold the **pseudo-data** with **nominal MC** and the **systematic templates** → Test the closure



χ^2/χ^2(0th iteration) between the unfolded histograms & pseudo-data truth

Unfolding with nominal MC and its systematic variations



Pseudodata truth

Nominal MC

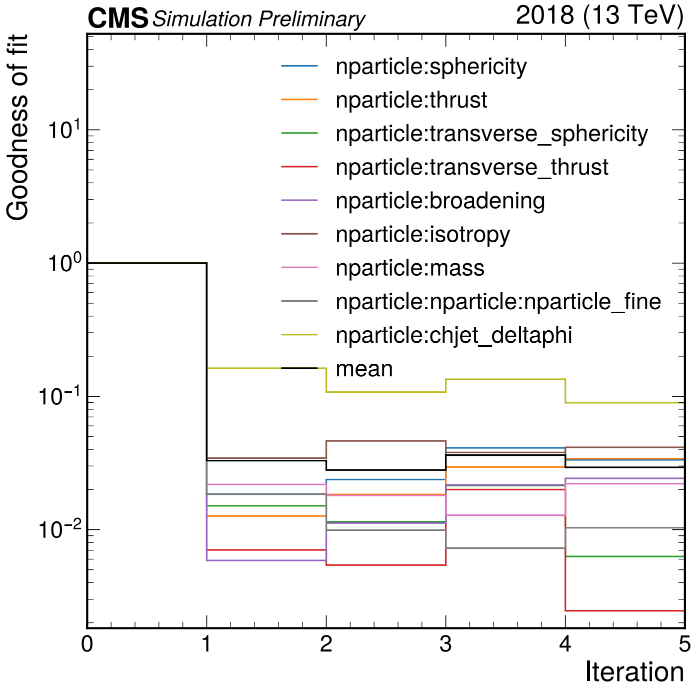
Particle-level
MC, **unfold**, and **pseudo-data truth**
transverse thrust at iteration 2

Validation: unfold the pseudo-data from Pythia A14 tune

Alternative MC from Pythia A14, CP5 and CUETP8M1 tunes → **pseudo-data**

Unfold the **pseudo-data** with **nominal MC** and the **systematic templates** → Test the closure

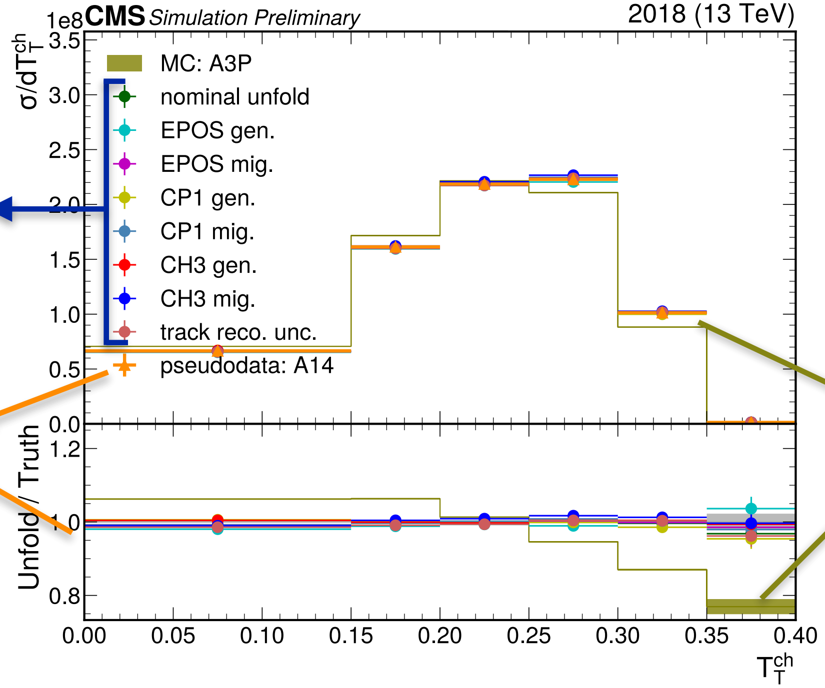
Example: Unfold the Pythia **A14** sample
(plots of other observables, reco-level plots, efficiency and acceptance in backup)



χ^2/χ^2(0th iteration) between the unfolded histograms & pseudo-data truth

Unfolding with nominal MC and its systematic variations

Pseudodata truth



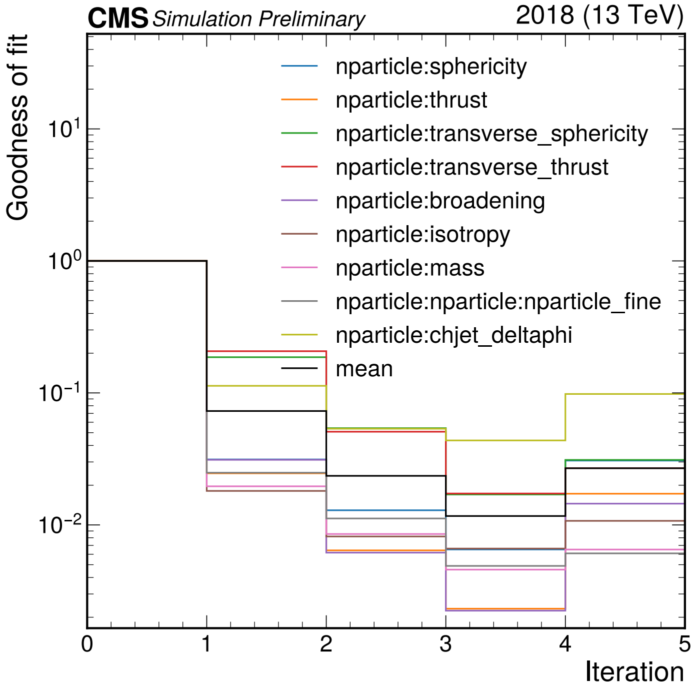
Nominal MC

Particle-level
MC, **unfold**, and **pseudo-data truth**
transverse thrust at iteration 2

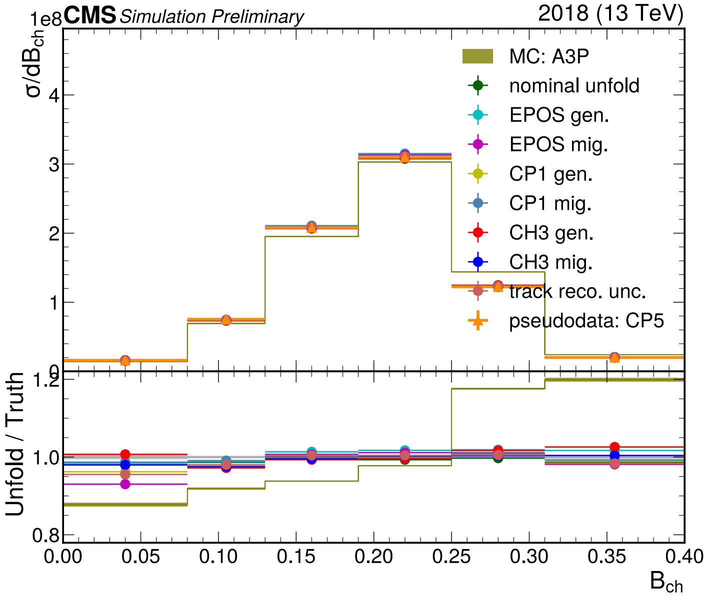
Validation: unfold the pseudo-data from Pythia CP5 tune

Alternative MC from Pythia A14, CP5 and CUETP8M1 tunes → **pseudo-data**

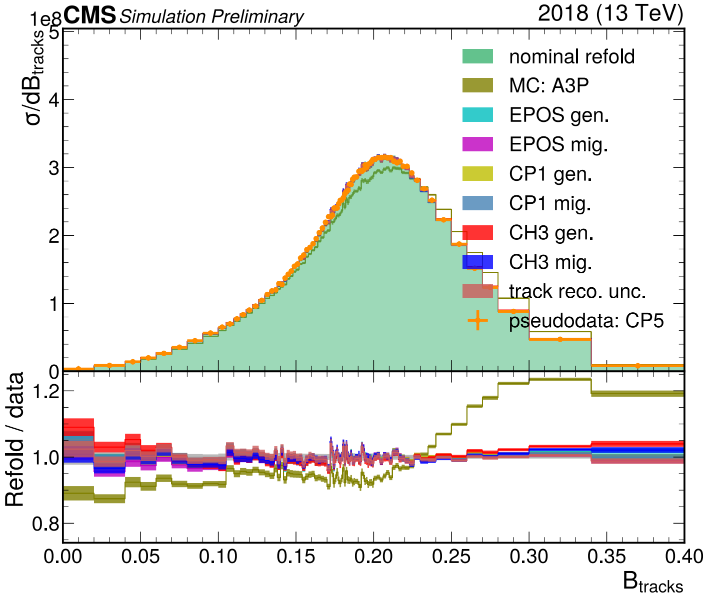
Unfold the **pseudo-data** with **nominal MC** and the **systematic templates** → Test the closure



χ^2/χ^2(0th iteration) between the unfolded histograms & pseudo-data truth



Particle-level
MC, **unfold**, and **pseudo-data truth** broadening at iteration 2



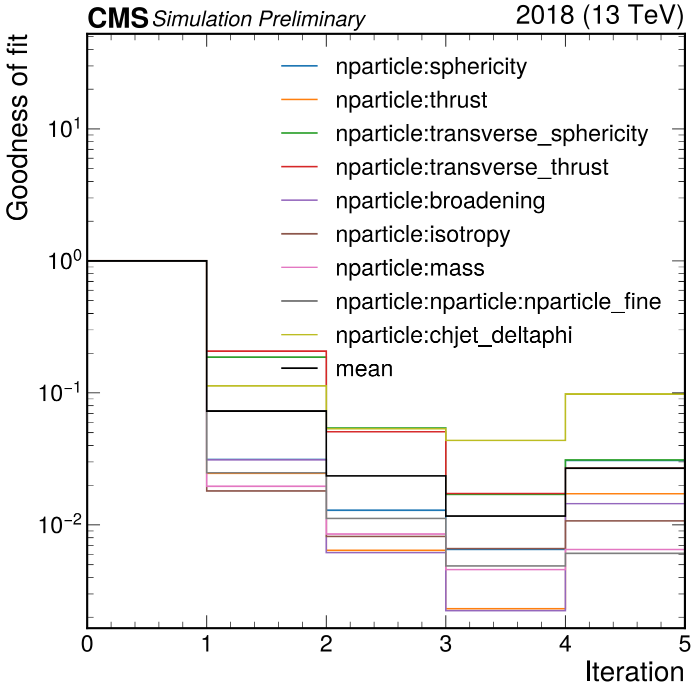
Detector-level
MC, **refold**, and **pseudo-data** broadening at iteration 2

Validation: unfold the pseudo-data from Pythia CP5 tune

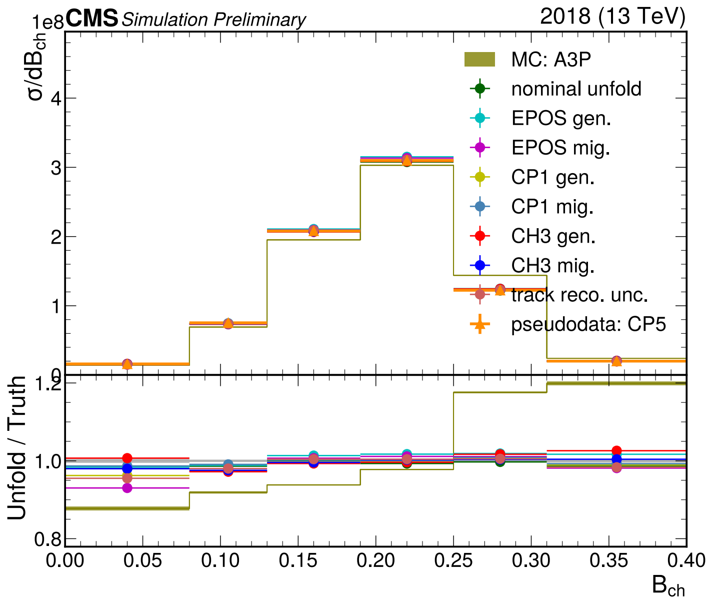
Alternative MC from Pythia A14, CP5 and CUETP8M1 tunes → **pseudo-data**

Unfold the **pseudo-data** with **nominal MC** and the **systematic templates** → Test the closure

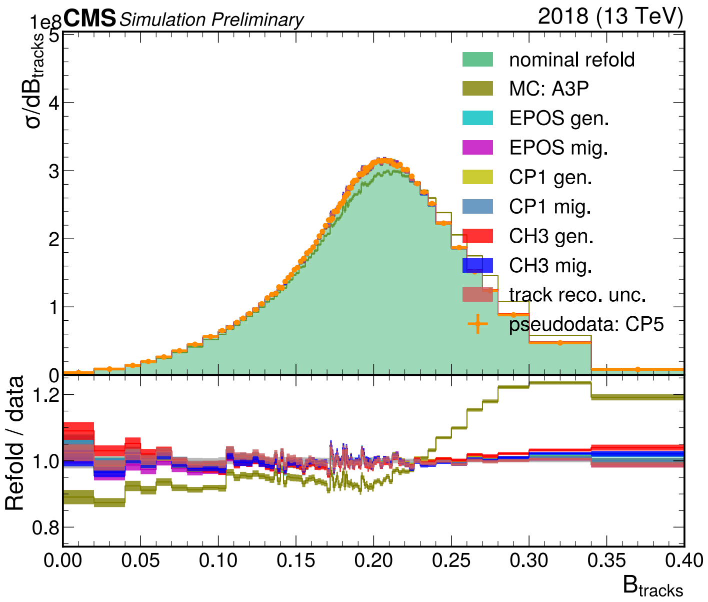
Example: Unfold the Pythia **CP5** sample



χ^2/χ^2(0th iteration) between the unfolded histograms & pseudo-data truth



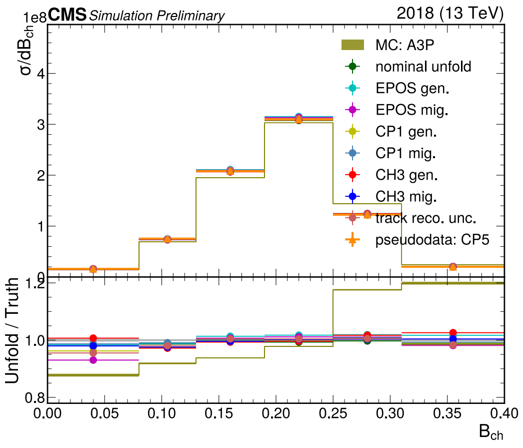
Particle-level
MC, **unfold**, and **pseudo-data truth** broadening at iteration 2

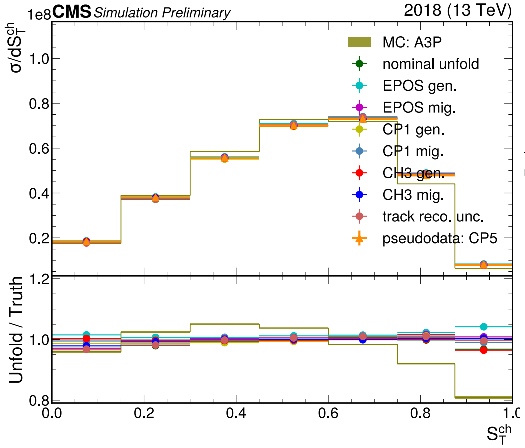


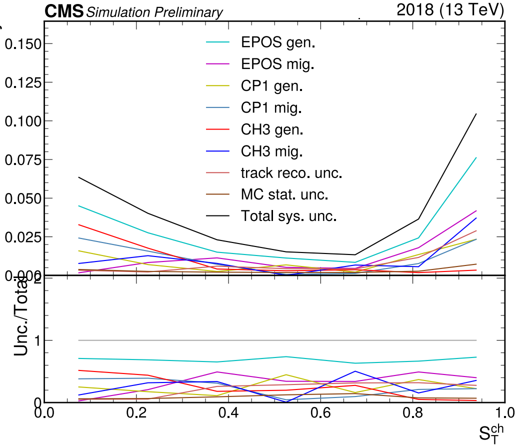
Detector-level
MC, **refold**, and **pseudo-data** broadening at iteration 2

Validation: unfold the pseudo-data with other systematic templates

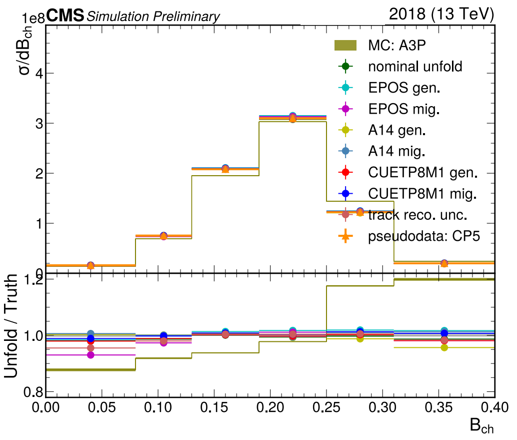
Particle-level
MC, **unfold**, and **pseudo-data truth**

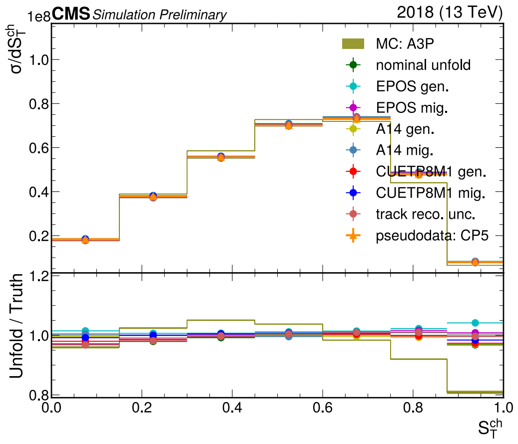


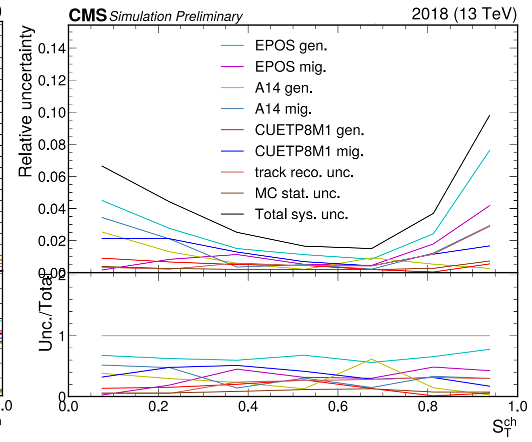




Systematic templates derived from EPOS, **Pythia CP1, Herwig CH3**







Systematic templates derived from EPOS, **Pythia A14, Pythia CUETP8M1**

Robustness test of MC **choices for systematic templates**
- The unfolding from alternative systematic templates also recovers the truth
- Uncertainties from gen-bias & migration functions are at a similar level

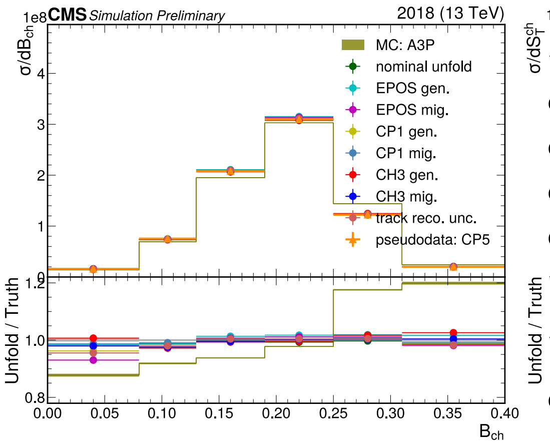
Broadening
unfold v.s. **truth**

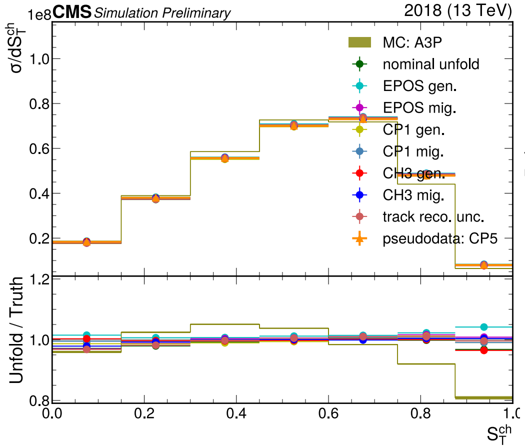
Transverse sphericity
unfold v.s. **truth**

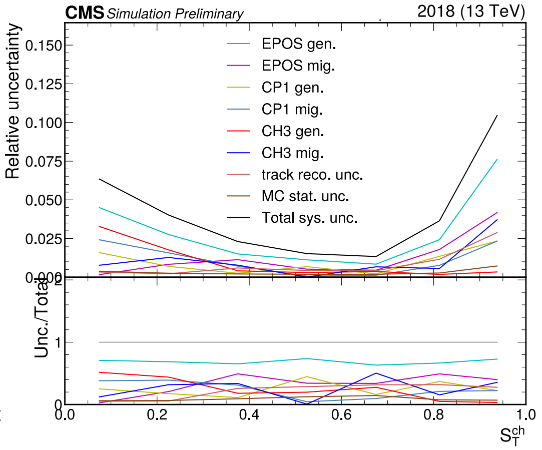
Transverse sphericity
uncertainty decomposition

Validation: unfold the pseudo-data with other systematic templates

Example: Unfold the Pythia **CP5** sample



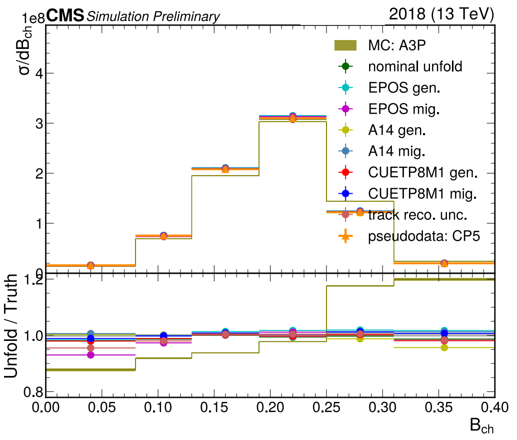


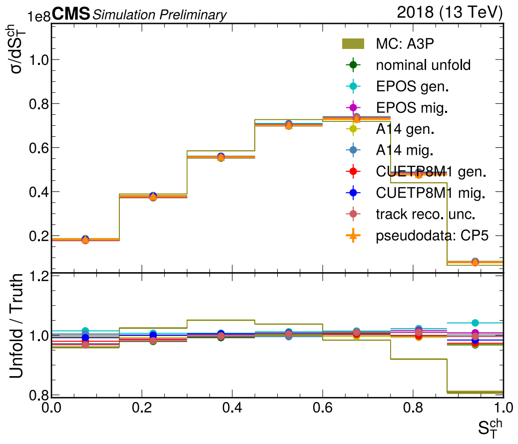


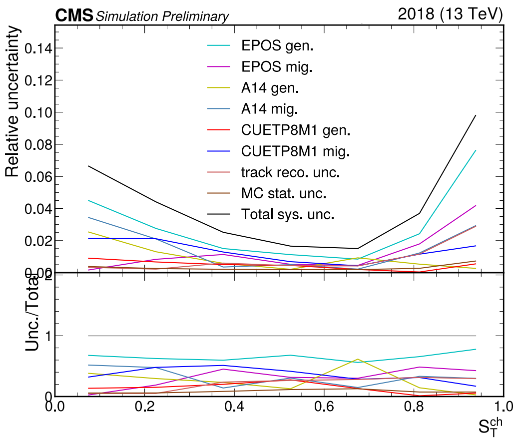
Particle-level

MC, **unfold**, and **pseudo-data truth**

Systematic templates derived from EPOS, **Pythia CP1, Herwig CH3**







Systematic templates derived from EPOS, **Pythia A14, Pythia CUETP8M1**

Robustness test of MC **choices for systematic templates**

- The unfolding from alternative systematic templates also recovers the truth
- Uncertainties from gen-bias & migration functions are at a similar level

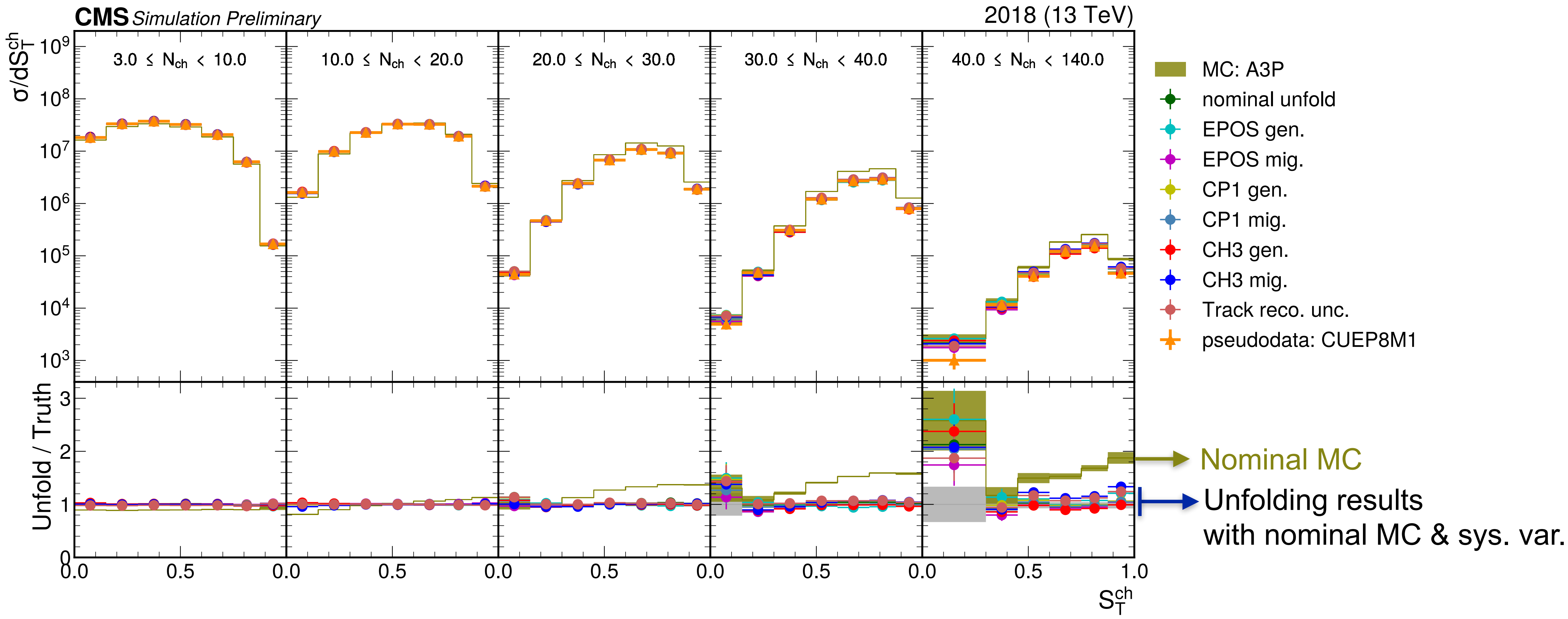
Broadening
unfold v.s. **truth**

Transverse sphericity
unfold v.s. **truth**

Transverse sphericity
uncertainty decomposition

Validation: unfold the pseudo-data

Test the unfolding on **2D distributions**

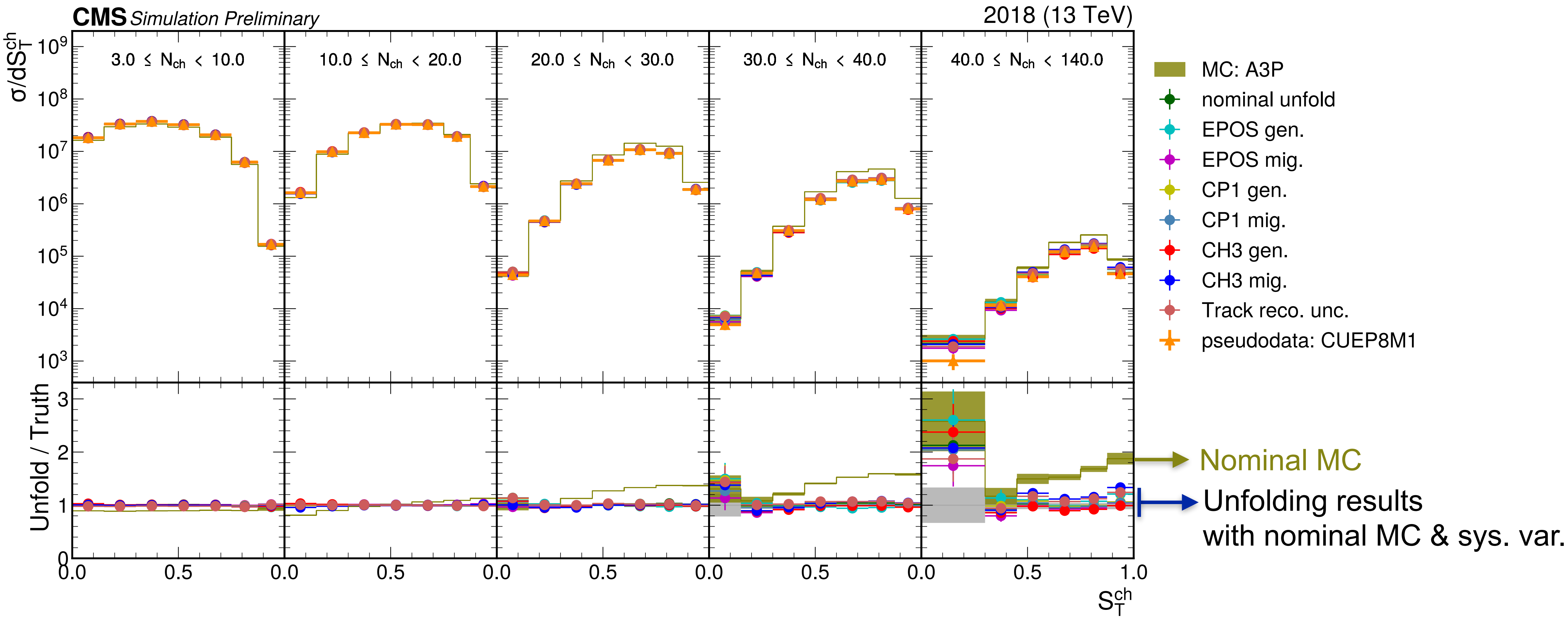


2D test also shows closure between unfolding results and the pseudo-data truth

Validation: unfold the pseudo-data

Test the unfolding on **2D distributions**

Example: Unfold the Pythia **CUETP8M1** sample, transverse sphericity in slices of N_{ch}



2D test also shows closure between unfolding results and the pseudo-data truth

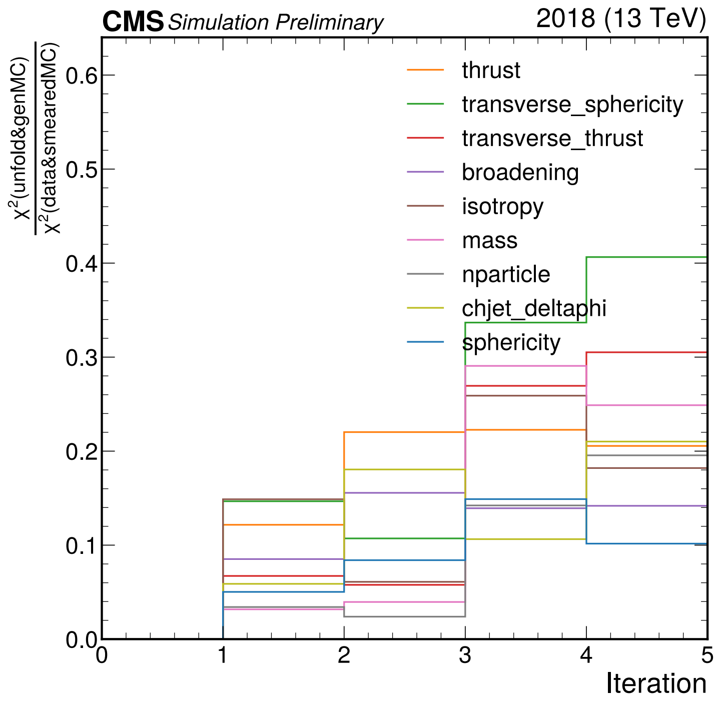
Validation: bottom-line test

Information loss during unfolding
→ the distinction between the **unfolded results** & **MC truth** < the distinction between **(pseudo-)data** & **smeared MC**
→ **bottom-line test**: the χ^2 between **unfolded results** (bias & MC stat. unc.) & **MC truth**
< the χ^2 between **(pseudo-)data** & **smeared MC**

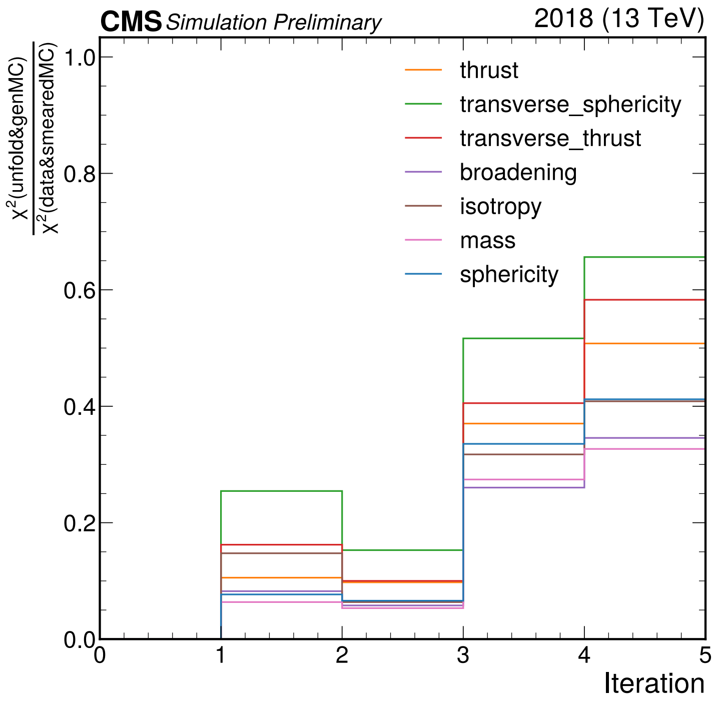
Example: χ^2 (unfold&gen-MC) / χ^2 (data&smeared MC) when unfolding CUEPT8M1 pseudo-data

Ideal case: χ^2 ratios ~ 1

information loss or
conservative unc. estimation
→ χ^2 ratios < 1



χ^2 ratios of 1D histograms



χ^2 ratios of 2D histograms
(event shape obs. in Nch slices)

Validation: bias and coverage test

Unfold pseudo-data with **toy experiments** of **uncertainty** variations

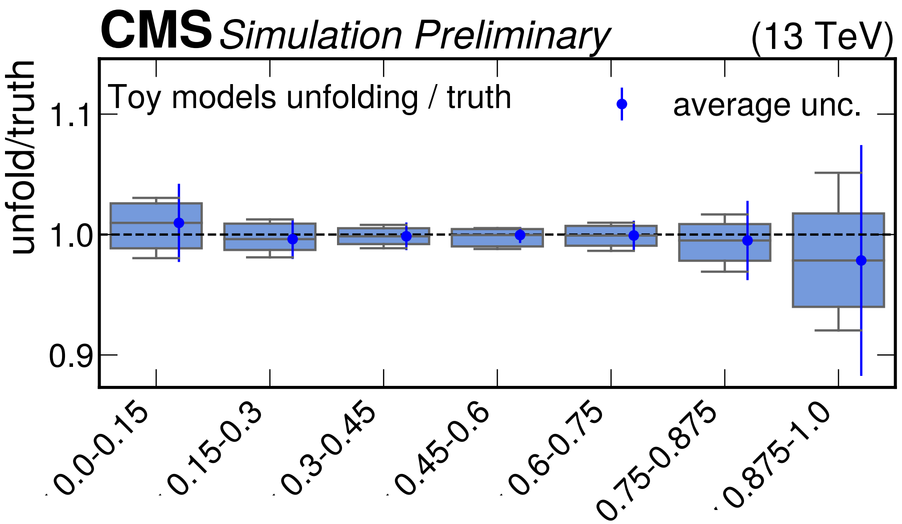
Bias test:

Bias of the unfolding results compared to pseudo-data truth

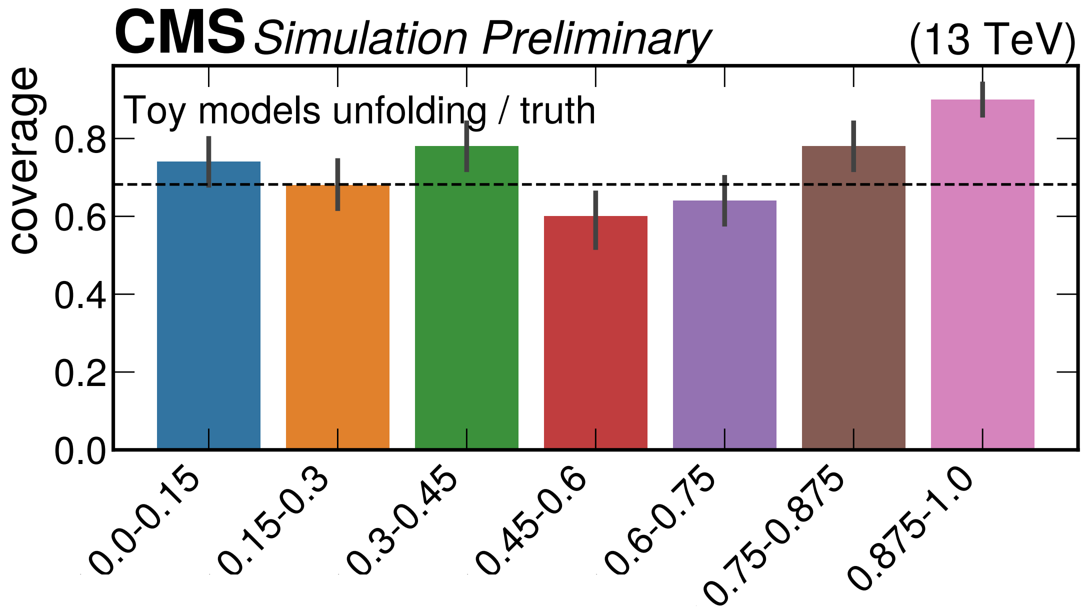
Coverage test:

How often does the unfold cover the pseudo-data?

Box-plot: 0.25, 0.5 and 0.75 quantile of 50 toys



Average coverage and its 68.2% confidence interval



Example: unfolding CUETP8M1 pseudo-data
Transverse sphericity distribution at iteration 2