University of Zurich[UZH]

Physics Institute

Event shape variables in pp collisions in CMS

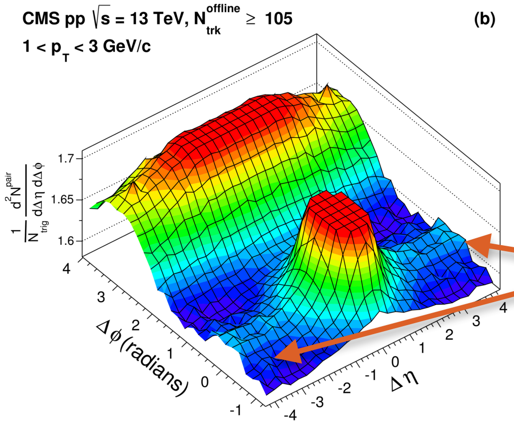
Weijie Jin

QCD@LHC 2024

Motivation

Motivation: previous event shape measurements

Existing observations of **unexpected effects in event shapes**

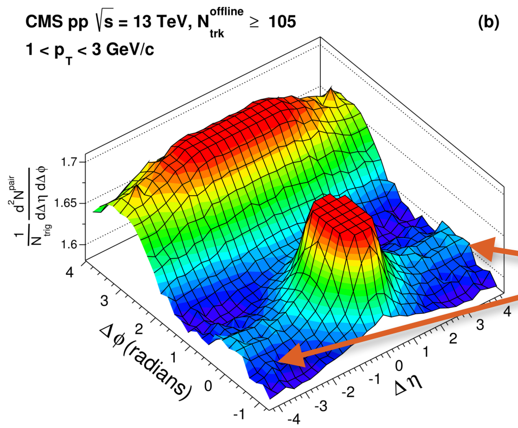


FSQ-15-002

Unexpected particle production across η, with ΔΦ ~ 0

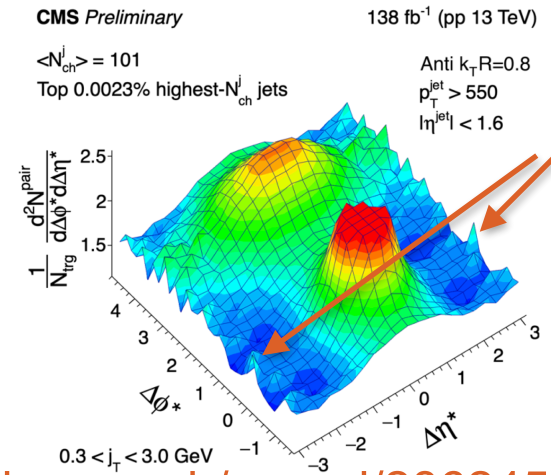
Motivation: previous event shape measurements

Existing observations of **unexpected effects in event shapes**



FSQ-15-002

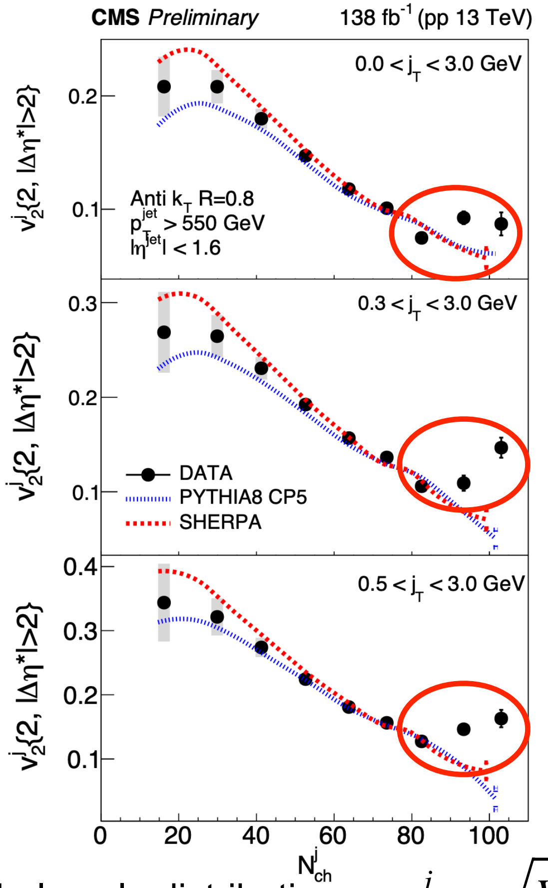
Unexpected particle production across η, with ΔΦ ~ 0



http://cds.cern.ch/record/2862457

Similar behavior observed in high multiplicity jets

The data-MC difference quantified by single-particle elliptic anisotropy*



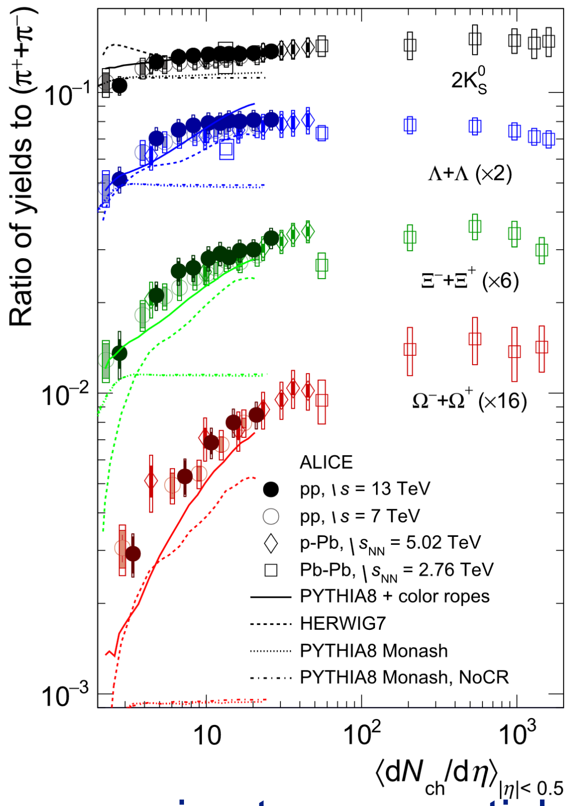
*related to the Fourier coefficient of two-particle azimuthal angle distribution as $v_2^j = \sqrt{V_{2\Delta}^j}$

Motivation: Strange hadron production in pp collisions

Mismodeling of **strangeness production in pp collisions**

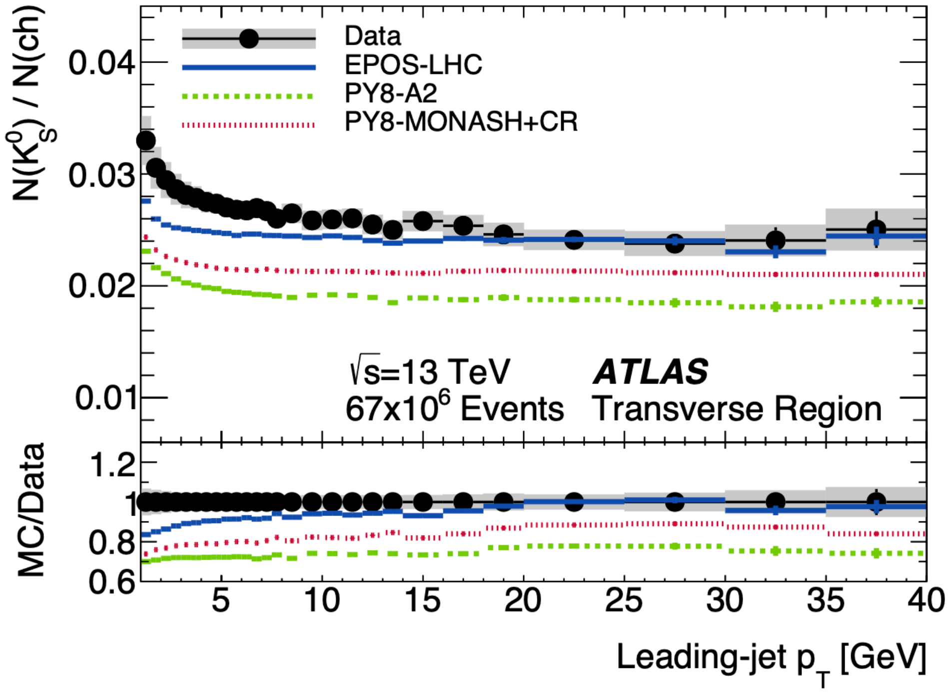
- Indicates the mismodeling in hadronization & potential quark-gluon plasma effects
- Affects the detector response and then the event shape measurement

Eur. Phys. J. C 80, 693 (2020)



Increase in strange particle as a function of particle multiplicity
→ no predicted by MC

https://arxiv.org/abs/2405.05048



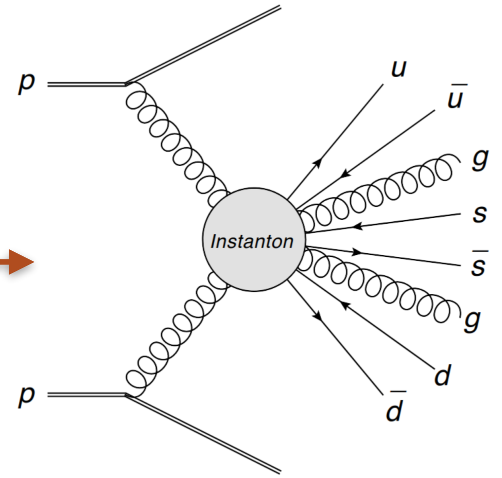
$N(K^0_S)/N_{ch}$ **fraction** v.s. **leading jet** p_T
in region transverse to the hard scattering (**underlying event** sensitive)
→ not predicted by MC

Motivation

QCD instantons

- Tunnelling process among discrete classical QCD vacuums which are topologically different
- A generic prediction of non-Abelian gauge theories

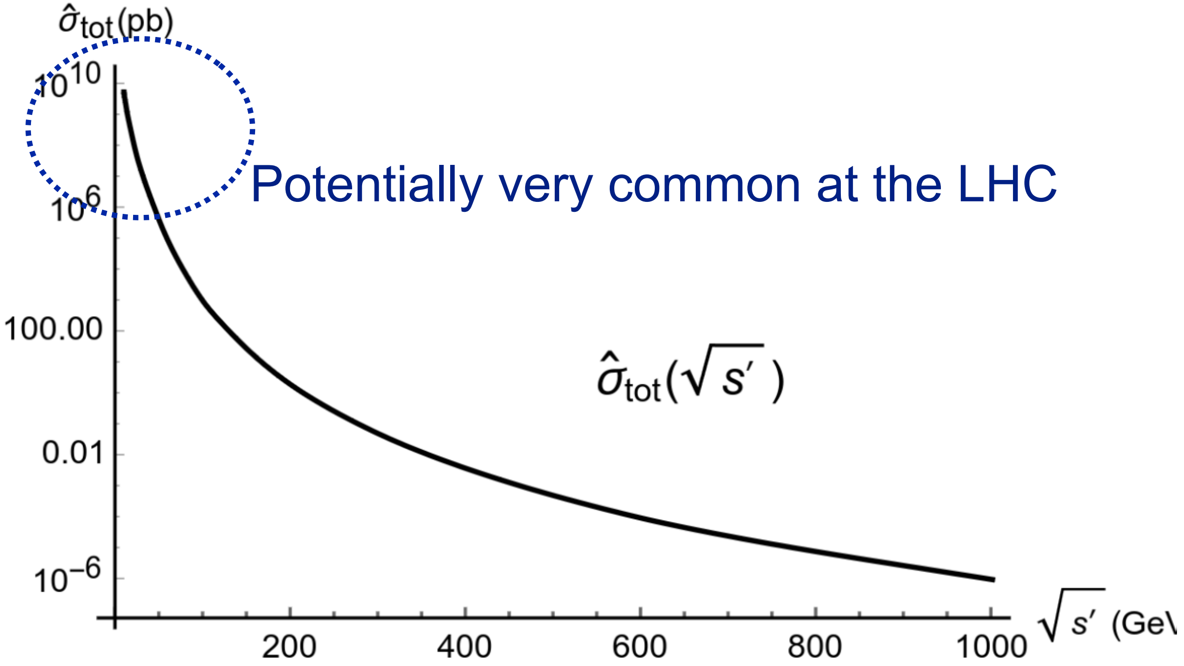
https://arxiv.org/abs/1911.09726

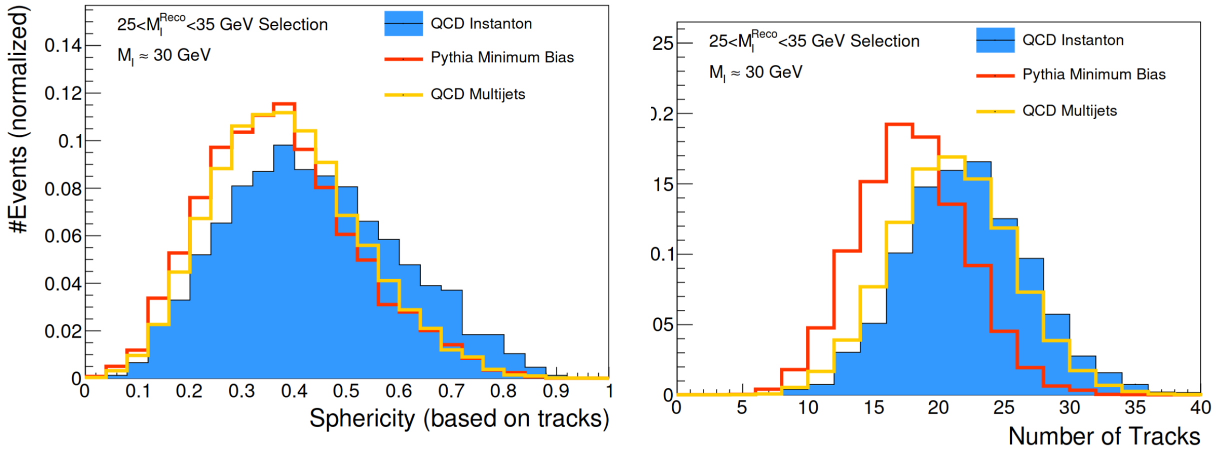


Final states from gluon fusion: $2N_f$ quarks + O(10) gluons

Signature: soft, isotropic events with high multiplicity

Potentially very common at the LHC





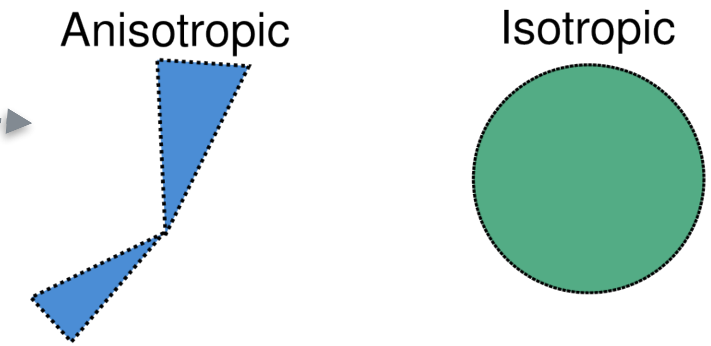
Event shape
as functions of charged particle momentum

Event shapes as functions of charged particle momentum

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**

Observables for measurement:

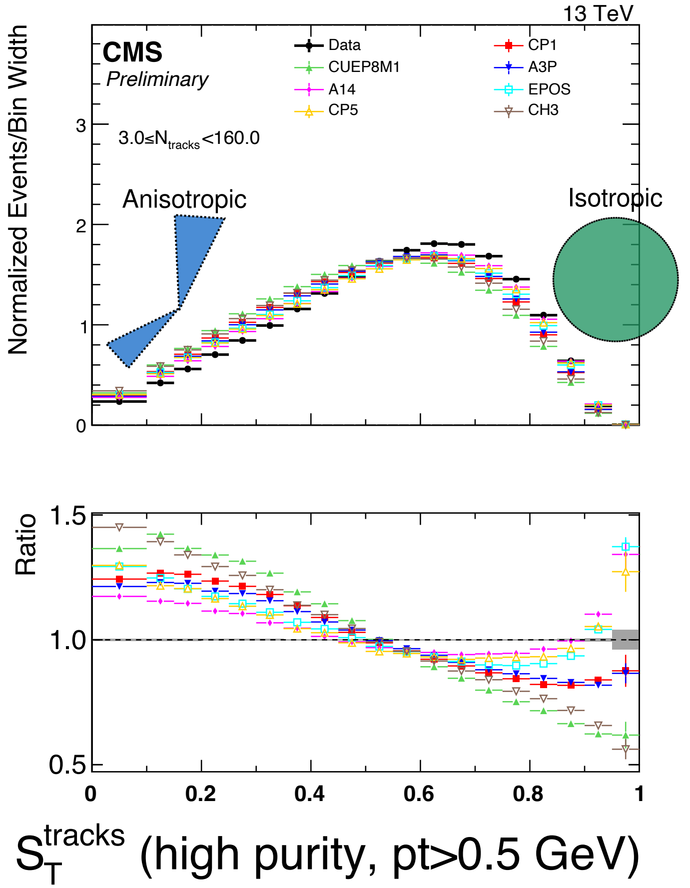
From charged particles p_T > 0.5 GeV, $|\eta|$< 2.4

- Charged particle multiplicity
- Invariant mass of charged particles

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

measures of momentum distributions

→ Example: transverse sphericity (detector level correspondence, to be discussed later)



Data & MC

Data: **Zerobias**, 2018 **low pileup** run, O(5M) events, ~64 μb^{-1}
MC: private minimum bias simulation **without pileup** (pileup effects given in backup)
→ minimal selections on primary vertices & tracks

Nominal samples and systematic variations

Pythia 8 CP1 (CMS), A3 (ATLAS)* ⟵ Different tunes, same MC model
EPOS-LHC ⟵ Regge-Gribov model, collective flow
Herwig 7 CH3 ⟵ Different shower& hadronization models

Validations and comparisons

Pythia CP5(CMS), CUETP8M1-NNPDF3.1(CMS), A14 (ATLAS) & its variations, CUETP8M1-NNPDF2.3(CMS), CUETP8M2T4, CUETP8M2T4-rope-hadronization&string-shoving, Pythia CP5 α_S(FSR) variations, Pythia CP5 color-reconnection tunes

*The ATLAS A3 tune was used as nominal MC for unfolding in the strategy development and validation. Later the nominal MC was changed to CMS CP1 tune for the data unfolding.

Uncertainty sources

Statistical uncertainty from data
$\longrightarrow$ fluctuations in the NN parameters
$\longrightarrow$ fluctuations of the unfolding output

Uncertainty sources

Statistical uncertainty from data
→ fluctuations in the NN parameters
→ fluctuations of the unfolding output

Systematic uncertainty from MC modelling

1. **MC statistics**
   → fluctuations in modelling

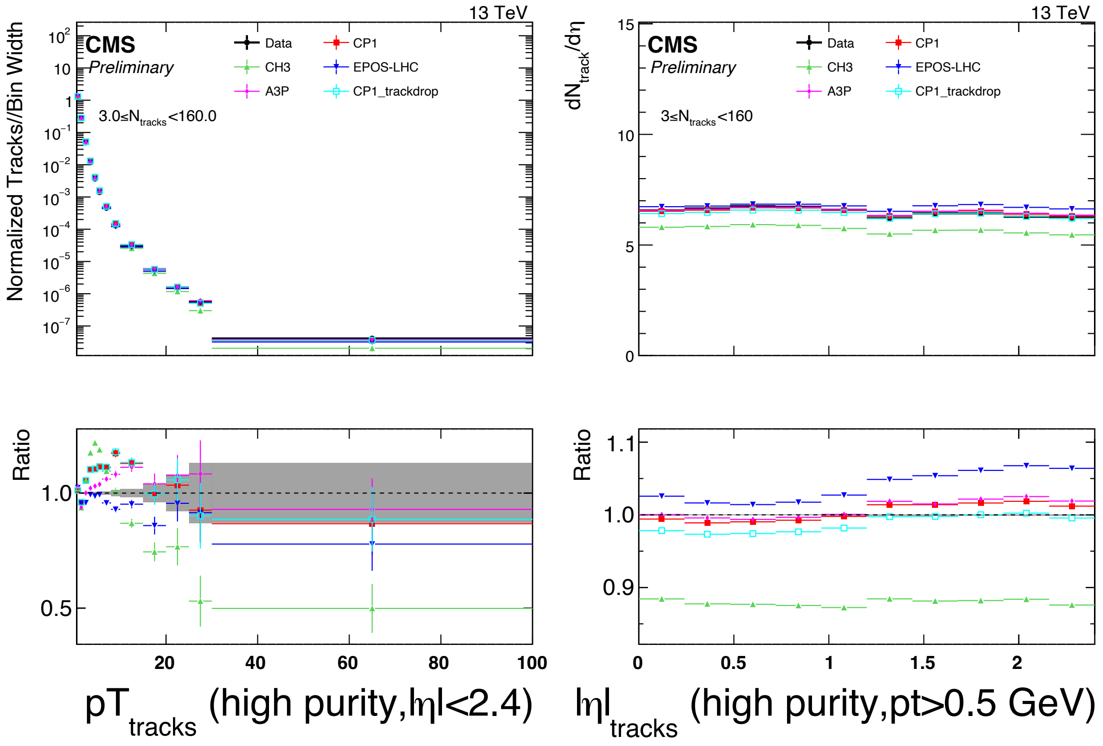
2. **Track reconstruction efficiency** uncertainty
   → differences between detector simulation and truth

3. **Mismodelling of observables used directly in unfolding**
   e.g. charged particle multiplicity, sphericity…
   → bias

4. **Mismodelling of other observables which may change detector response**
   e.g. track rapidity, particle composition, p_T
   → migration function uncertainty

Uncertainty sources

Statistical uncertainty from data
→ fluctuations in the NN parameters
→ fluctuations of the unfolding output



p_T and η of the particles
Not unfolded, but affect all the event shape obs.

Systematic uncertainty from MC modelling

1. **MC statistics**
   → fluctuations in modelling

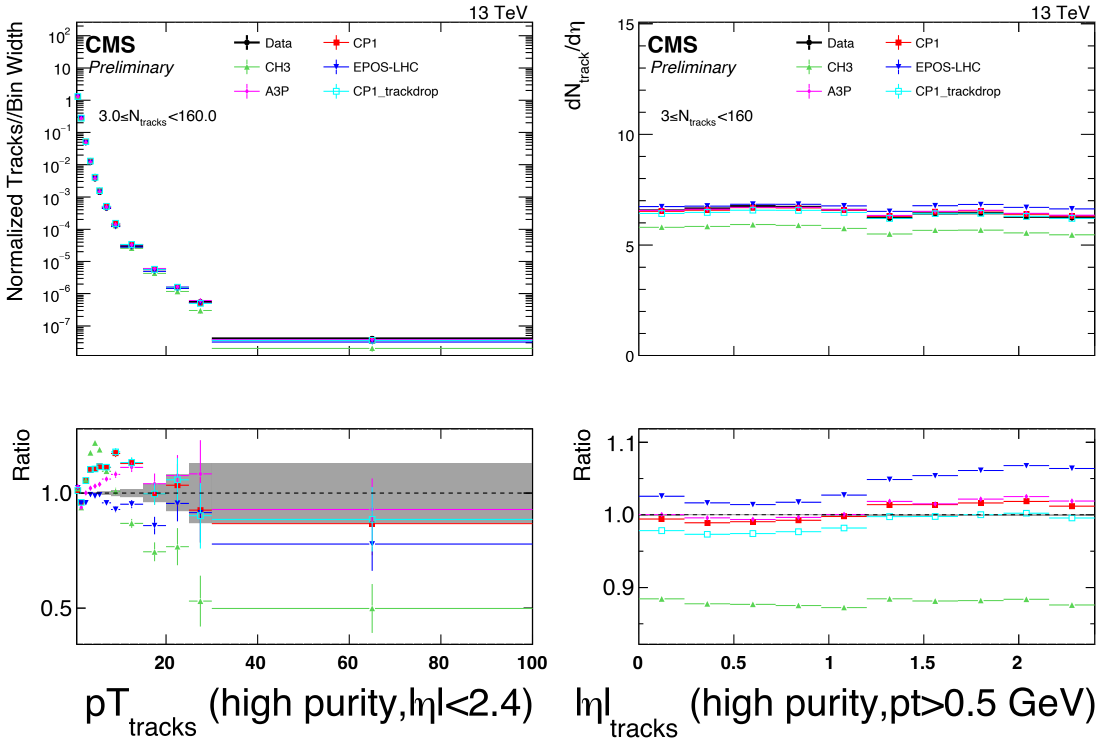
2. **Track reconstruction efficiency** uncertainty
   → differences between detector simulation and truth

3. **Mismodelling of observables used directly in unfolding**
   e.g. charged particle multiplicity, sphericity…
   → bias

4. **Mismodelling of other observables which may change detector response**
   e.g. track rapidity, particle composition, p_T
   → migration function uncertainty

Uncertainty sources

Statistical uncertainty from data
→ fluctuations in the NN parameters
→ fluctuations of the unfolding output



p_T and η of the particles
Not unfolded, but affect all the event shape obs.

Systematic uncertainty from MC modelling

1. **MC statistics**
   → fluctuations in modelling

2. **Track reconstruction efficiency** uncertainty
   → differences between detector simulation and truth

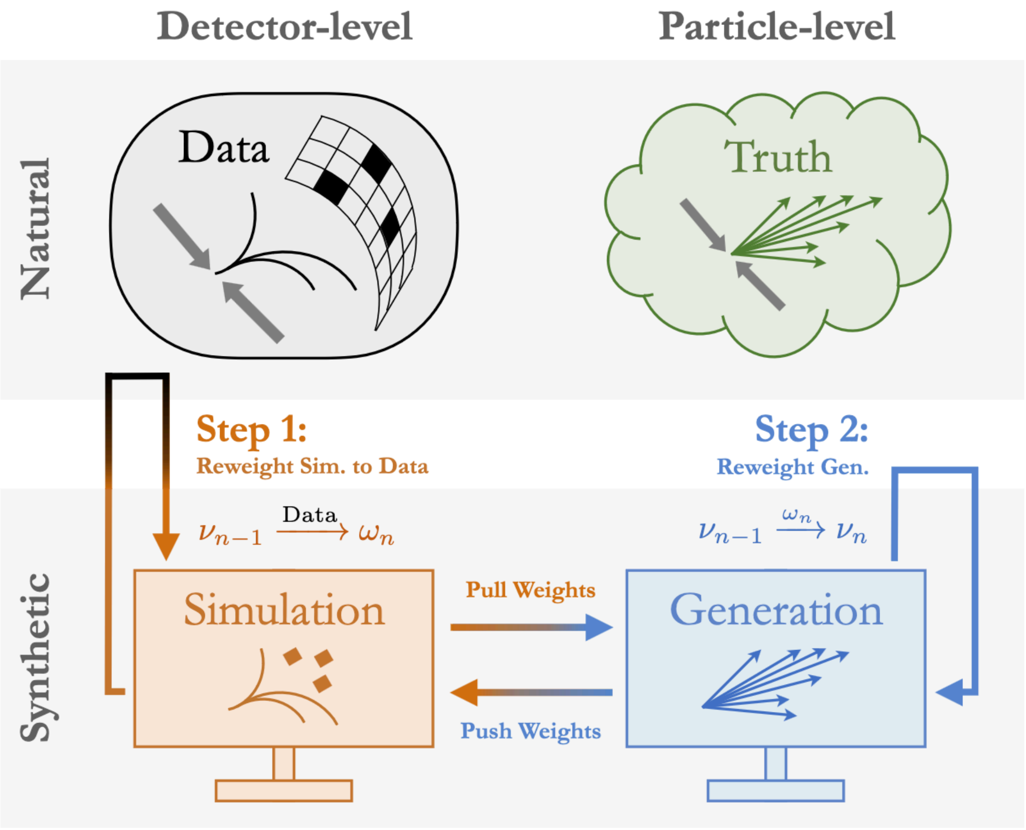
3. **Mismodelling of observables used directly in unfolding**
   e.g. charged particle multiplicity, sphericity…
   → bias

   → **We consider variations derived from 3 separate MC models**

4. **Mismodelling of other observables which may change detector response**
   e.g. track rapidity, particle composition, p_T
   → migration function uncertainty

Machine-learning-based unbinned unfolding & uncertainty estimation

Unfolding algorithm



Multifold *:

- **Input**: values of 8 observables for every event in simulation and data
- **Output**: reweighted simulated events approximating data

→ The result are **unbinned** weighted events, although we show binned histograms for visualisation

Two steps of unbinned reweighting:

1. Weight **MC** to **data** at **detector level**
2. Weight **original MC** to **reweighted MC** at **generator level**

   Extra 2 steps added to deal with the selection efficiency and signal acceptance

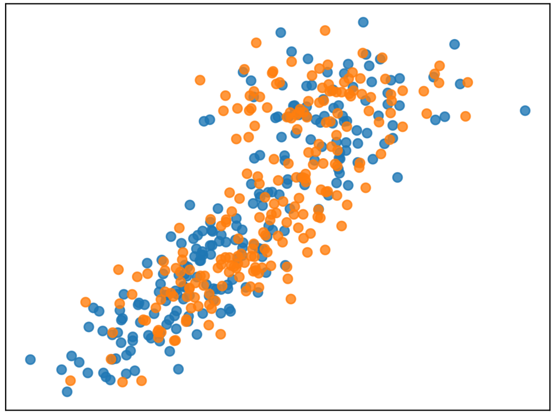
→ repeat in **iterations**

+

Unbinned weighting for uncertainty estimation

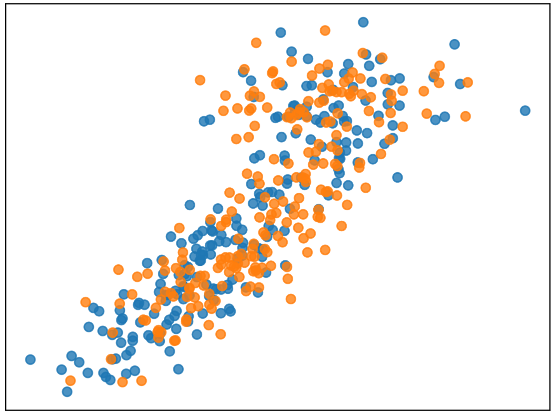
* https://arxiv.org/abs/1911.09107
https://arxiv.org/abs/2105.04448

Unbinned multi-dimensional unfolding and uncertainty estimation



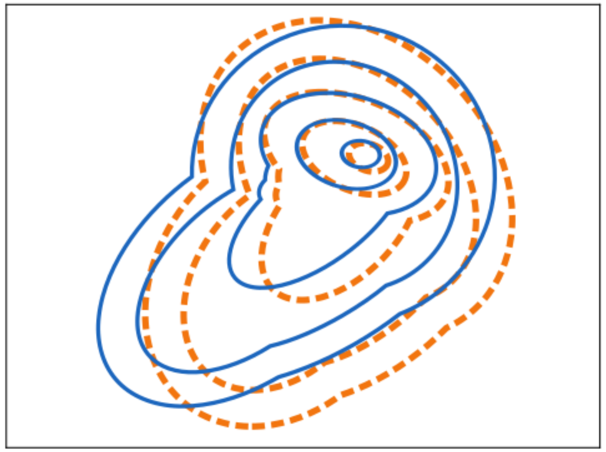
← A typical binary classifier to distinguish two sets

Unbinned multi-dimensional unfolding and uncertainty estimation

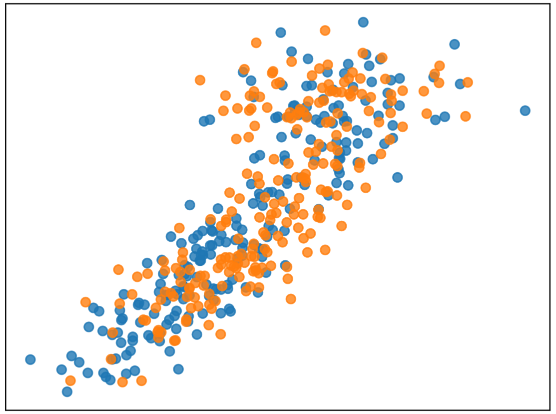


← A typical binary classifier to distinguish two sets

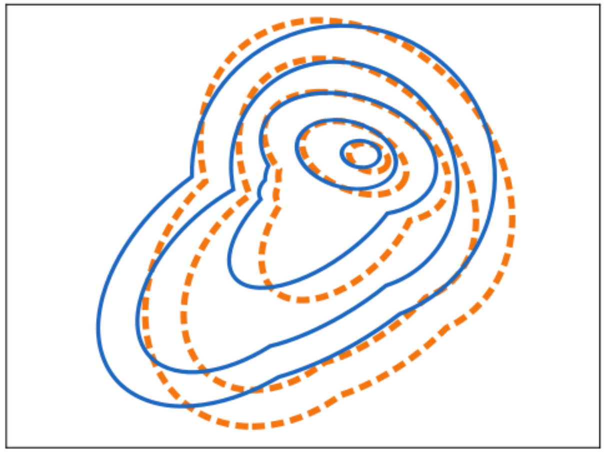
What it actually did: learn the differences in the distributions →



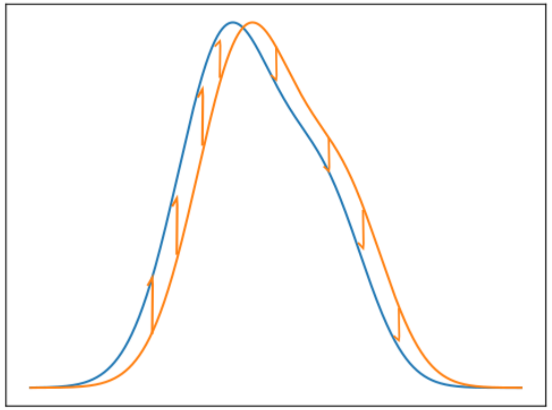
Unbinned multi-dimensional unfolding and uncertainty estimation



← A typical binary classifier to distinguish two sets

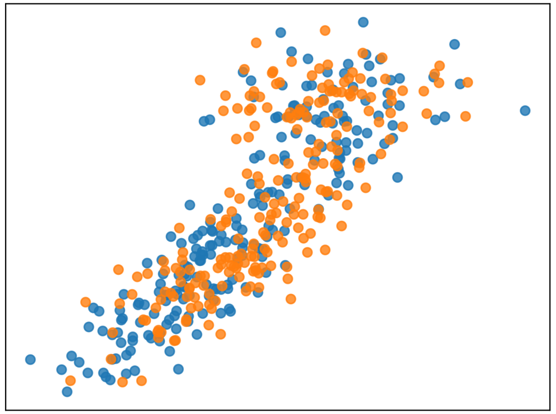


What it actually did: learn the differences in the distributions →



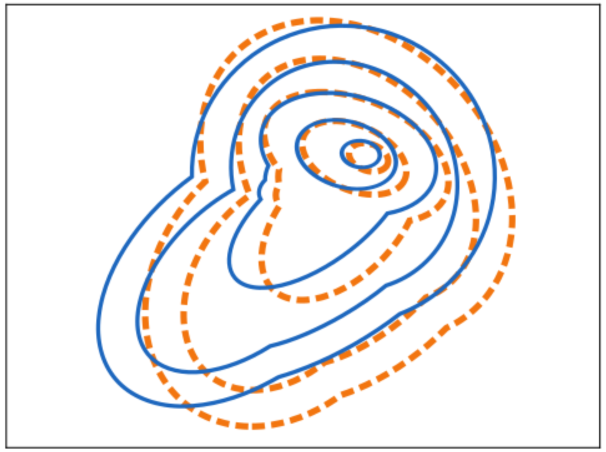
← We can use the classification scores to weight **MC** to **data**,

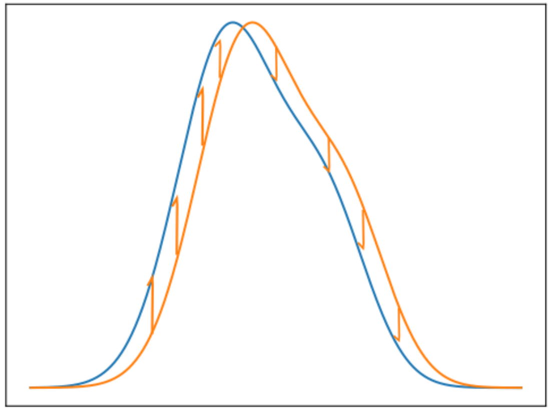
Unbinned multi-dimensional unfolding and uncertainty estimation



← A typical binary classifier to distinguish two sets

What it actually did: learn the differences in the distributions →





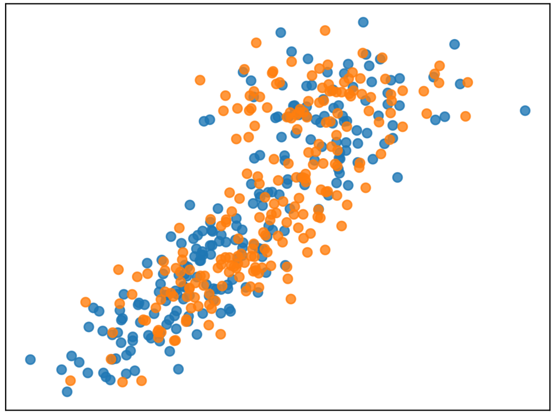
← We can use the classification scores to weight **MC** to **data**

Event-wise unfolding → the result independent of binning

The actual unfolding in iterations:

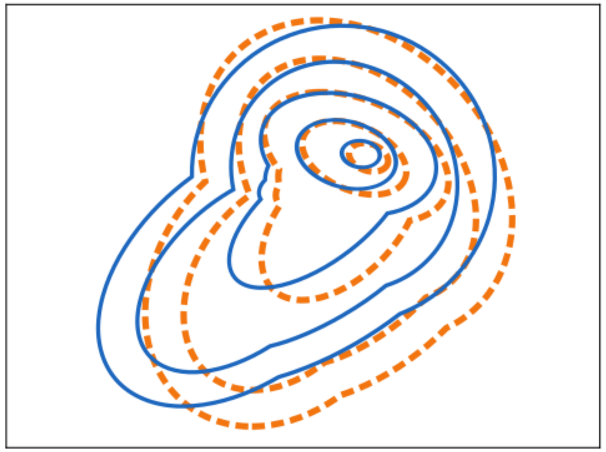
- Step 1: weight **MC** to **data**, at detector level
- Step 2: pull back the weights to particle(truth) level

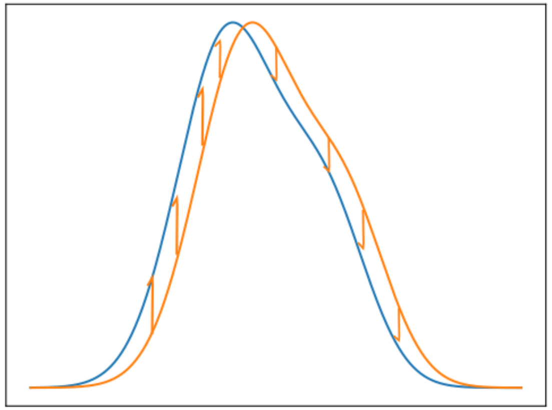
Unbinned multi-dimensional unfolding and uncertainty estimation



← A typical binary classifier to distinguish two sets

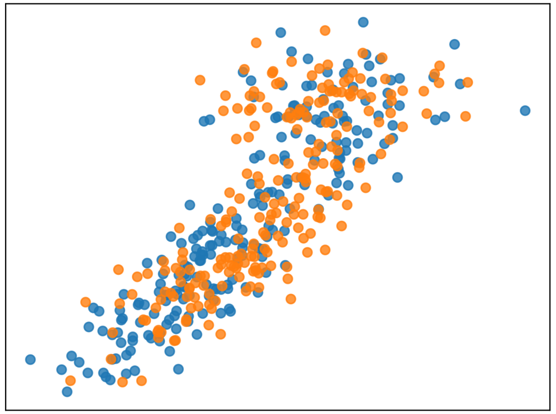
What it actually did: learn the differences in the distributions →





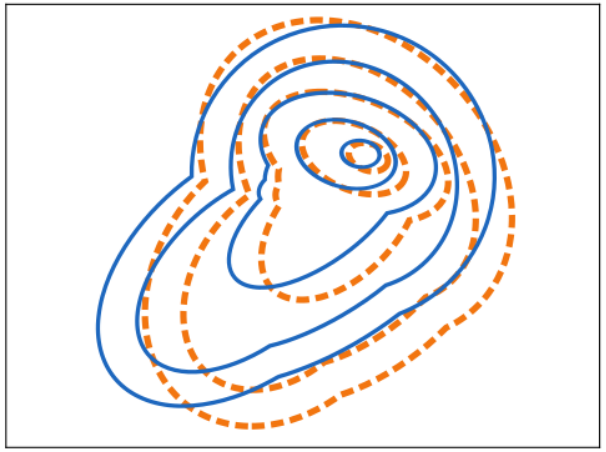
← We can also use the classification scores to weight **nominal MC sample** to **systematic variations**

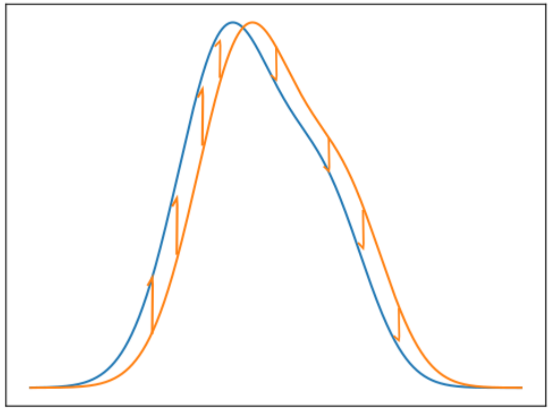
Unbinned multi-dimensional unfolding and uncertainty estimation



← A typical binary classifier to distinguish two sets

What it actually did: learn the differences in the distributions →





← We can also use the classification scores to weight **nominal MC sample** to **systematic variations**

Event-wise uncertainty template → unbinned unfolding uncertainty & covariance

Systematic uncertainty estimation based on unbinned reweighting

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/

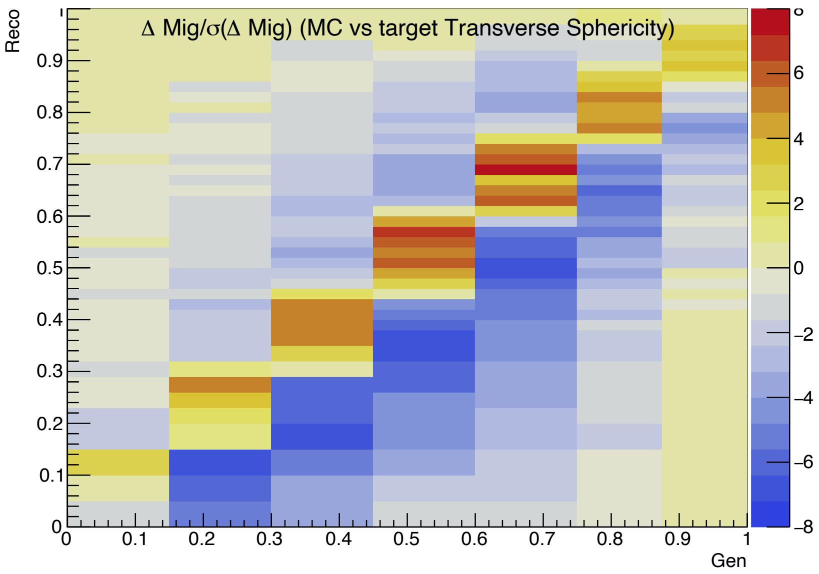
Systematic uncertainty estimation based on unbinned reweighting

Track reconstruction efficiency uncertainty

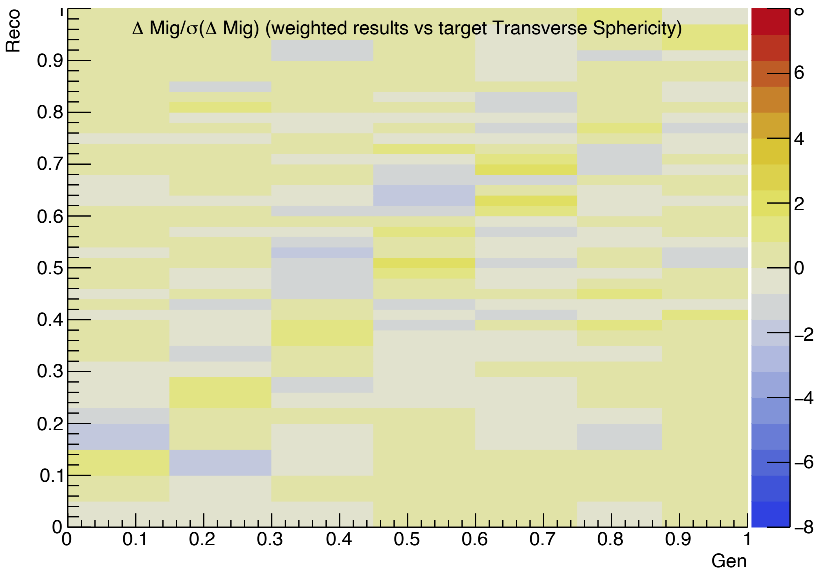
- Step1: Randomly drop 2.1%(1%) tracks with pT<20 GeV (>20 GeV) in nominal MC*
- Step2: weight the nominal MC to Step1 output at particle- and detector-level

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/

Systematic uncertainty estimation based on unbinned reweighting

Mismodelling of observables used directly in unfolding

Derive the templates by **weighting** **nominal MC** to **alternative MC** at the **particle-level**

→ **ML-based** unbinned weighting

→ output: **weighted nominal MC events**

- same **particle-level distribution** as **alternative MC**
- keeps the **gen. → reco. migration** of the **nominal MC**

Systematic uncertainty estimation based on unbinned reweighting

Mismodelling of observables used directly in unfolding

Derive the templates by **weighting** **nominal MC** to **alternative MC** at the **particle-level**

→ **ML-based** unbinned weighting

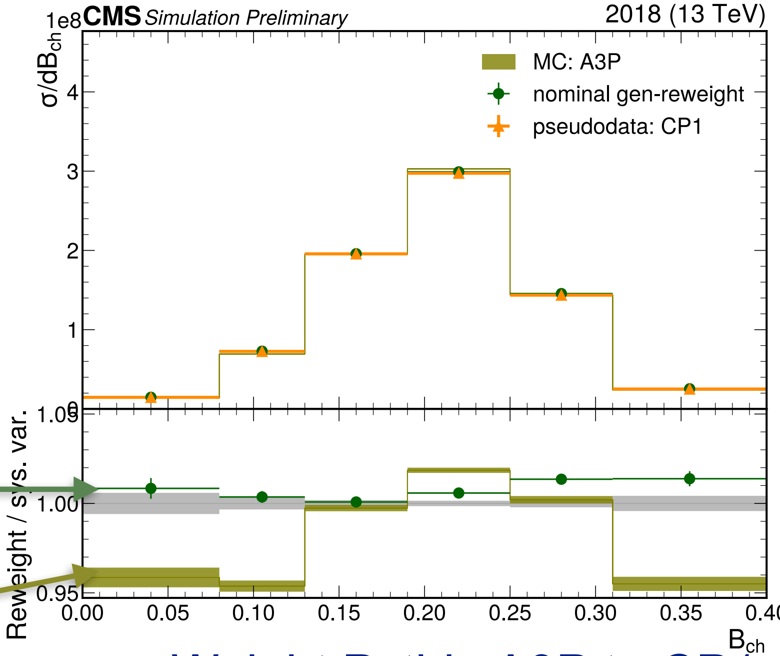
→ output: **weighted nominal MC events**

- same **particle-level distribution** as **alternative MC**
- keeps the **gen. → reco. migration** of the **nominal MC**

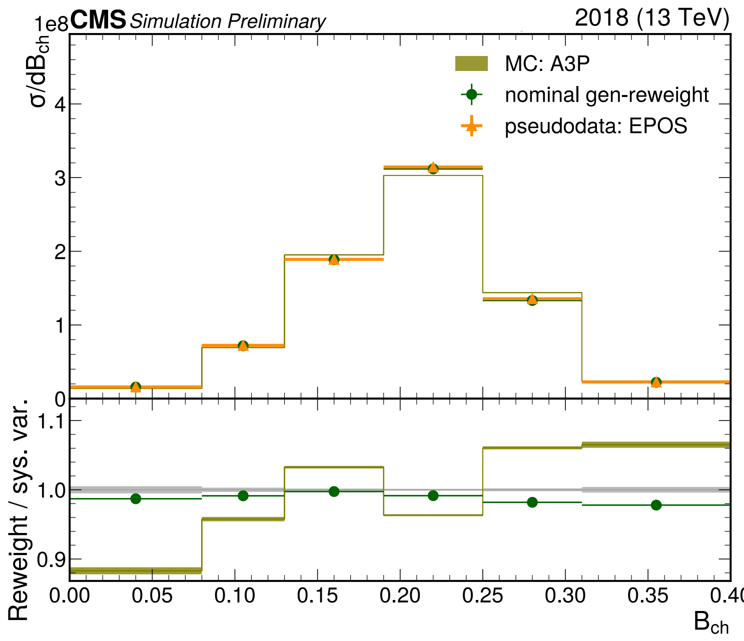
Example:
particle-level broadening
before & after weighting

After reweighting at the gen-level

Nominal MC



Weight Pythia A3P to CP1



Weight Pythia A3P to EPOS

Systematic uncertainty estimation based on unbinned reweighting

Mismodelling of other observables which may change detector response

Derive the templates with two-step weighting

Systematic uncertainty estimation based on unbinned reweighting

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

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

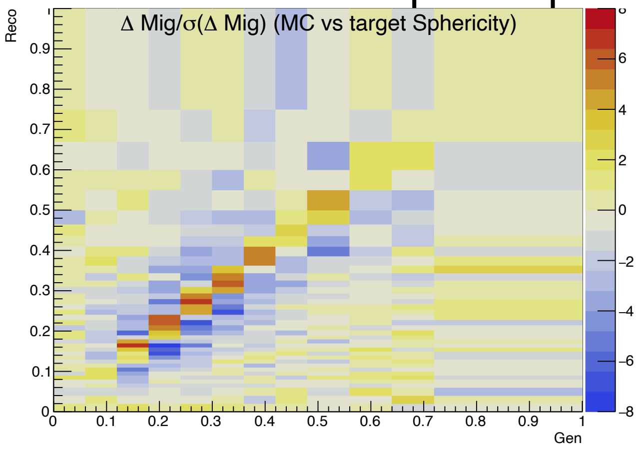
Mismodelling of other observables which may change detector response

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

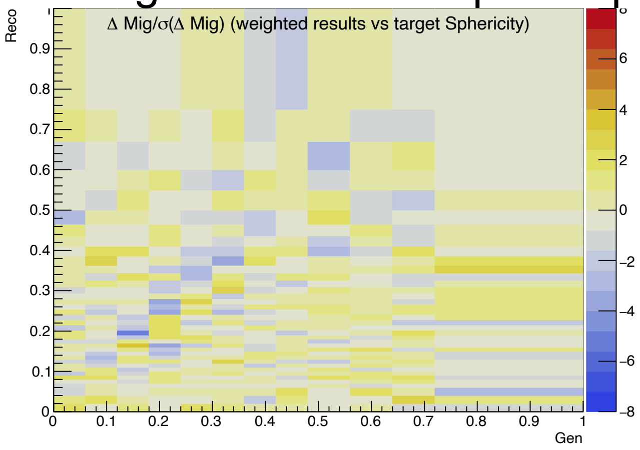
Example:
Gen → reco migration
of spherocity

Before step 2 weighting:
nominal MC - step 1 output



→

After step 2 weighting:
weighted result - step 1 output



Results

Unfolding results

Simultaneously unfold the 8 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

Simultaneously unfold the 8 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 as **weighted MC events**

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

Unfolding results

Simultaneously unfold the 8 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

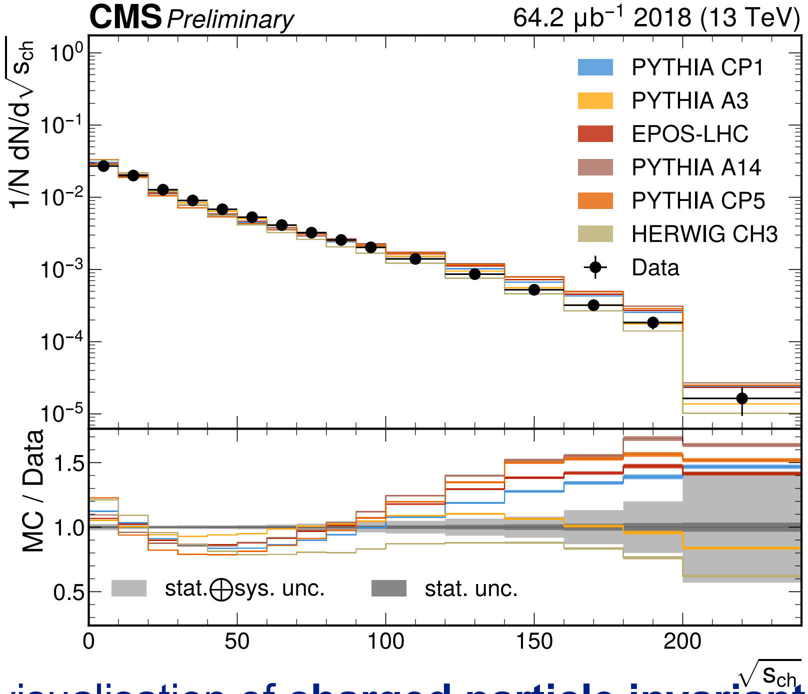
This method:

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

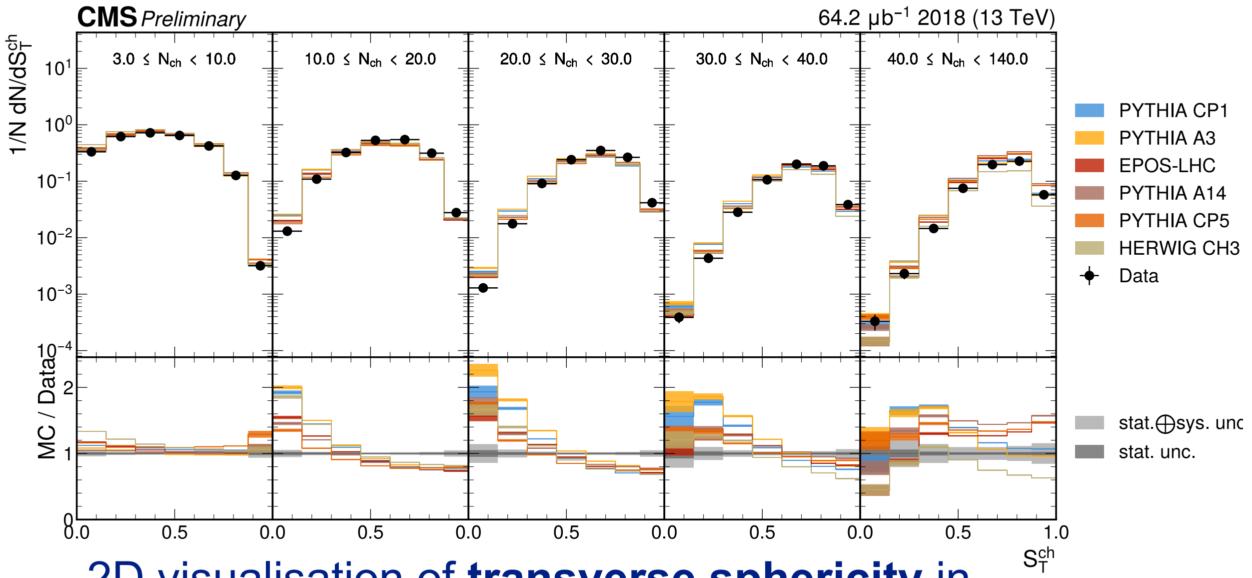
→ much easier to scale up the dimensions

Unfolding results as **weighted MC events**

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



1D visualisation of **charged particle invariant mass** distribution



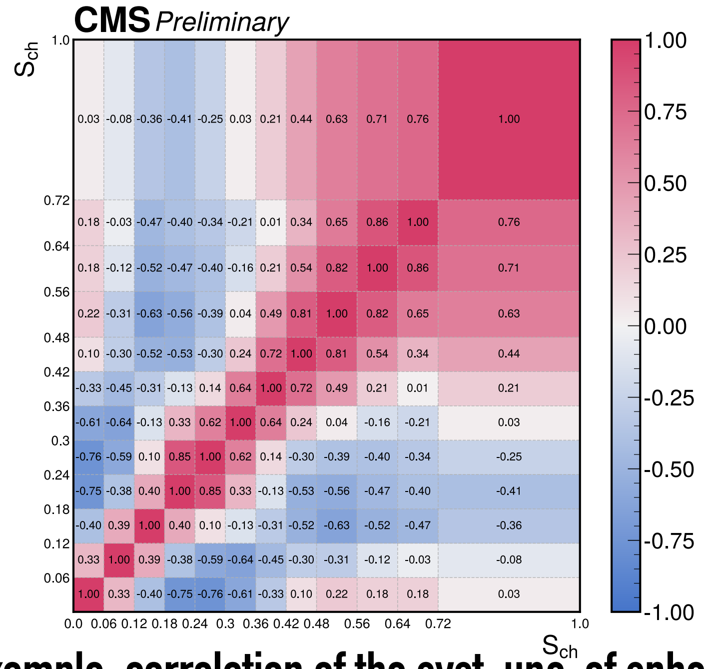
2D visualisation of **transverse sphericity** in **charged particle multiplicity** slices

CMS-PAS-SMP-23-008

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**



CMS-PAS-SMP-23-008

Example: correlation of the syst. unc. of sphericity

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

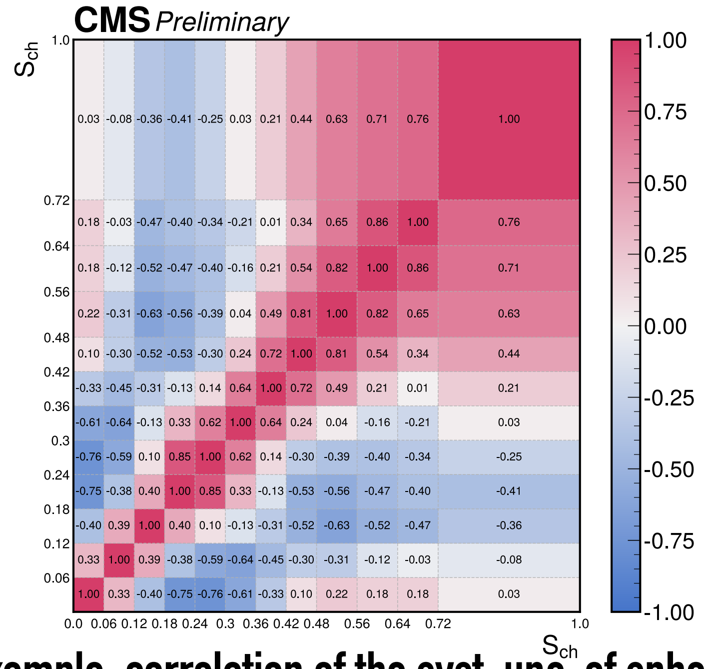
+

Uncertainties+Covariance on the results

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**



CMS-PAS-SMP-23-008

Example: correlation of the syst. unc. of sphericity

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

+

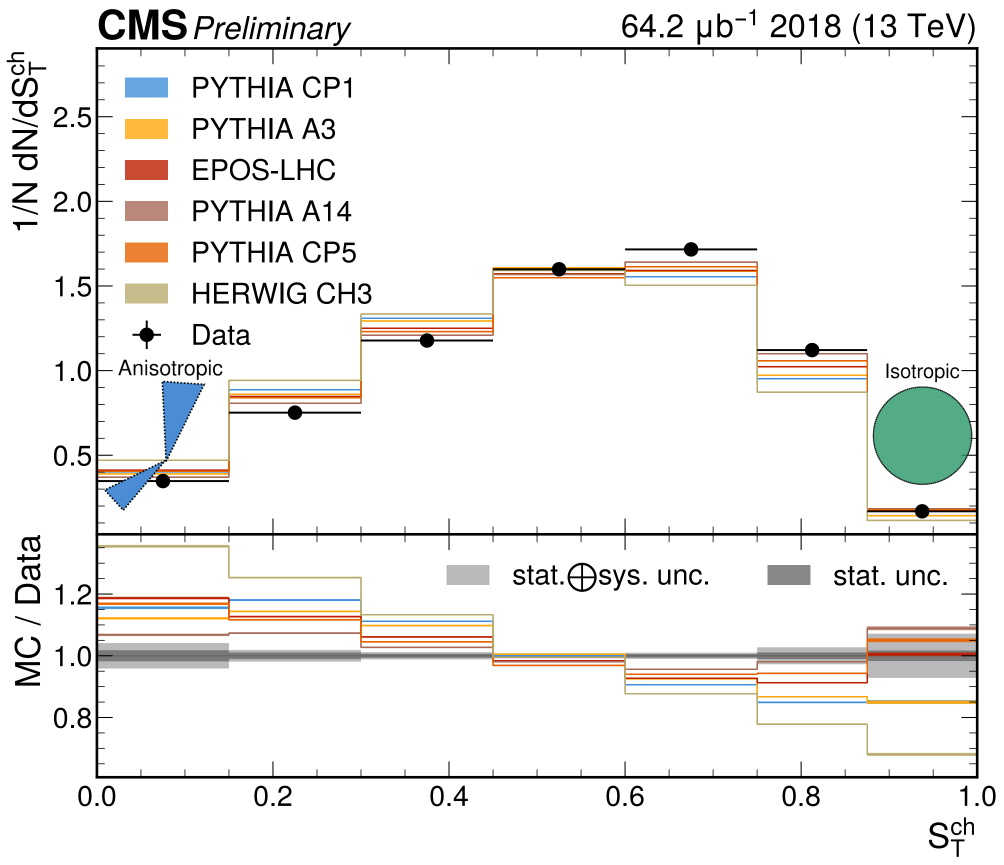
Uncertainties+Covariance on the results

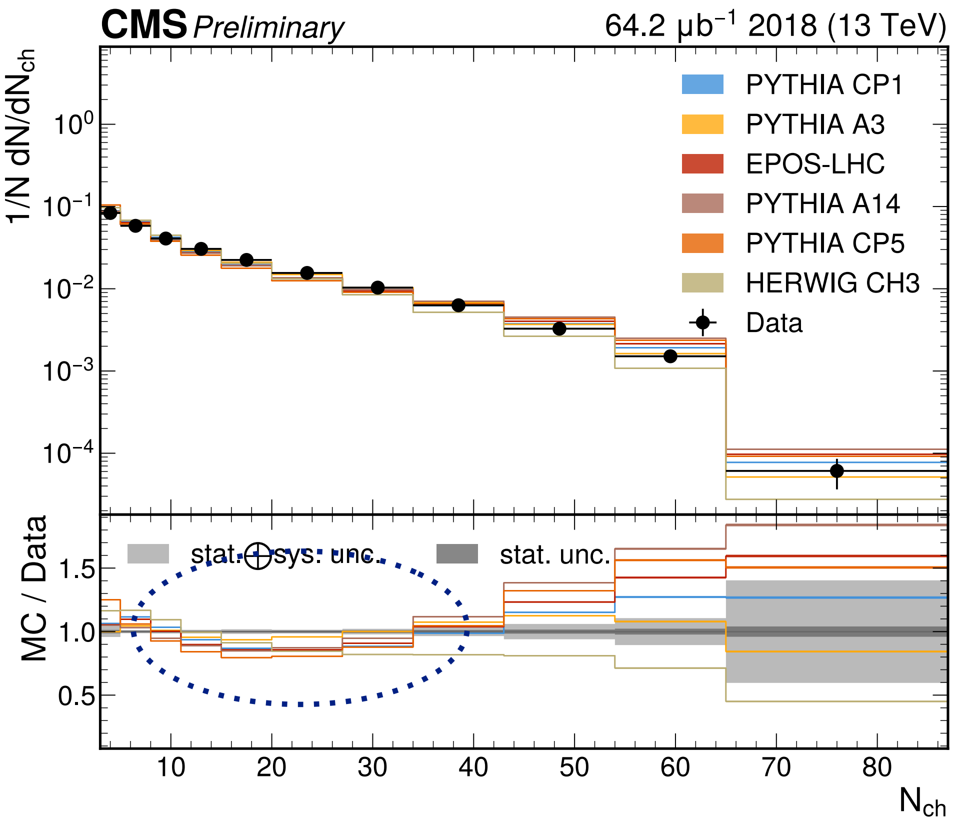
The way to improve the usability of **unfolded results**

- 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)

Interpretation of results

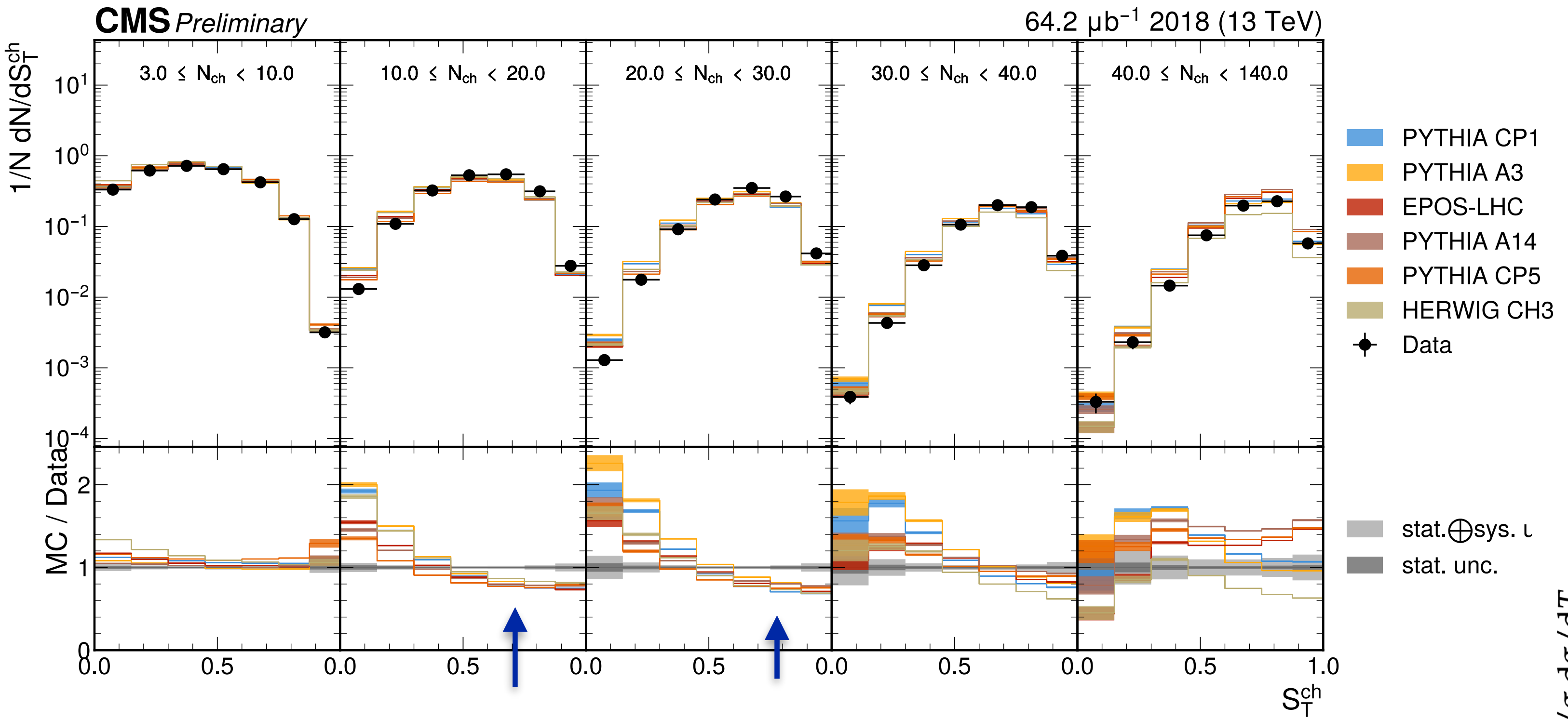


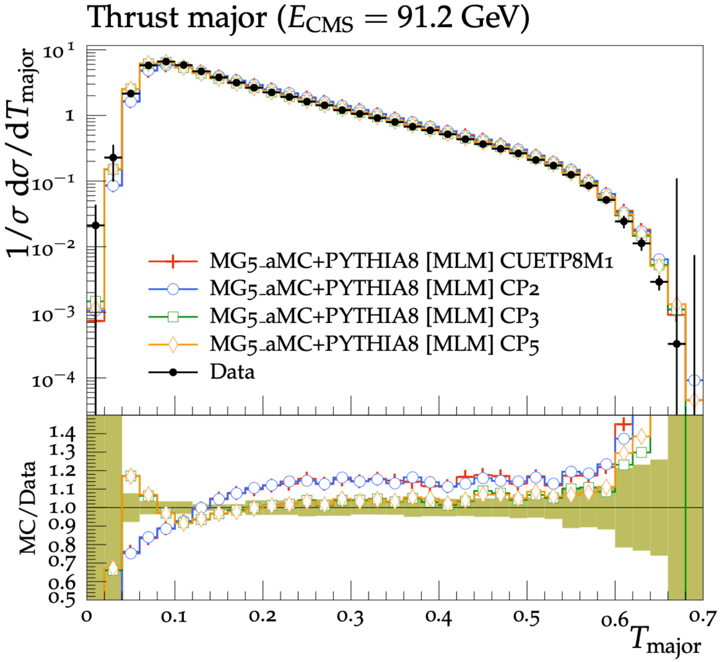


The **data** tends to be **more isotropic** than all the **MC** predictions

The **data** is **more accumulated** in the middle Nch region (~20)

Interpretation of results





The MC & data **discrepancy** in **event shape** is the **largest** in the **middle Nch** region

- The mismodeling is observed for all the event shape observables
- The mismodeling sustains under **variations of PDF, generator, UE tune, color-reconnection models, α_S(FSR)** (backup)
- **Opposite behaviours** of **pp** collision MC to the **e+e-** collision MC
- Missing QCD instanton effects or collective behaviour in this region?
- We provide the unfolded results for **further theoretical interpretation**

Summary

- **Minimum bias collisions** may be the home of **interesting physics** at the LHC
  - Observables that are not described by existing models → QGP or others?
  - Could be contributed by QCD instantons ← topological effects of non-abelian gauge fields
- **Event shapes** are important signatures of these physics effects

- We present a **measurement of these event shape observables** in CMS
  - **Unbinned high-dimensional unfolding** based on machine-learning models
  - **Unbinned uncertainty estimation** based on pseudo-experiments
  - Validations provided: pseudo-data unfolding, bias & coverage test, bottom-line test (backup)
  - The unfolding method is also used in an ATLAS measurement of Z+jets kinematics arxiv:2405.20041

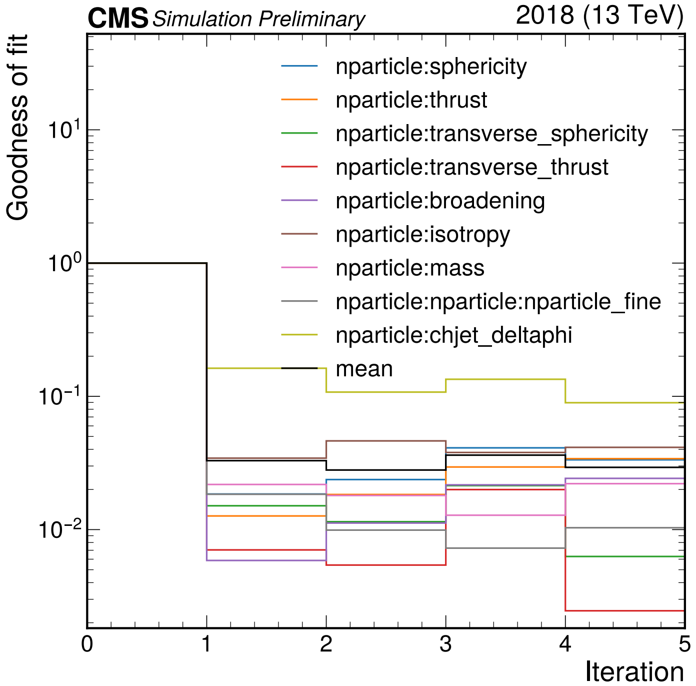
- Unfolding results as weighted MC events
  - Unbinned events → further usage of data does not depend on binning
  - Visualised by 1D or 2D histograms
  - Correlations of the uncertainties are provided (visualised by 2D histograms)
  - **Data is more isotropic than all the MC variations, especially in mid-Nch region**

Backup

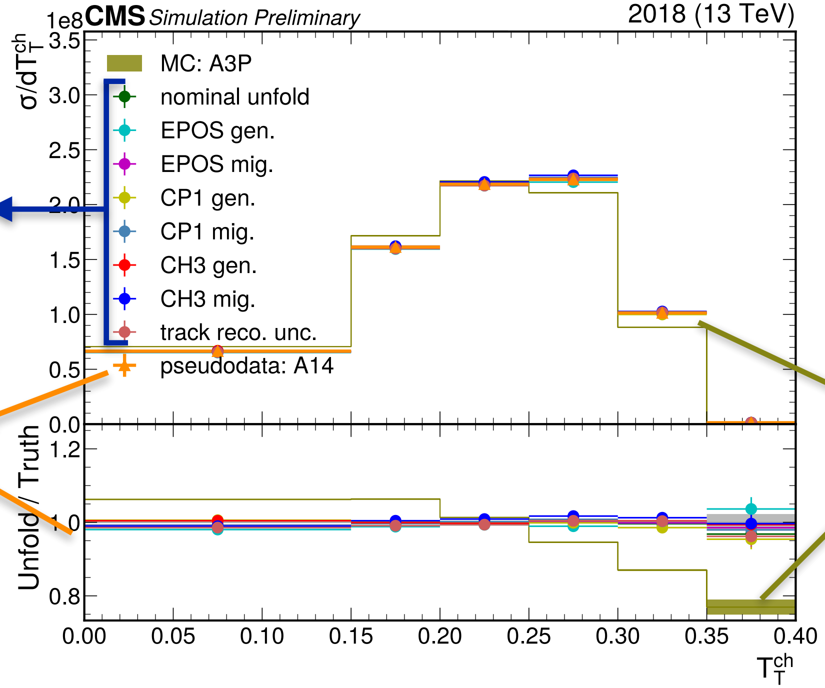
Validation: unfold the pseudo-data from Pythia A14 tune

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



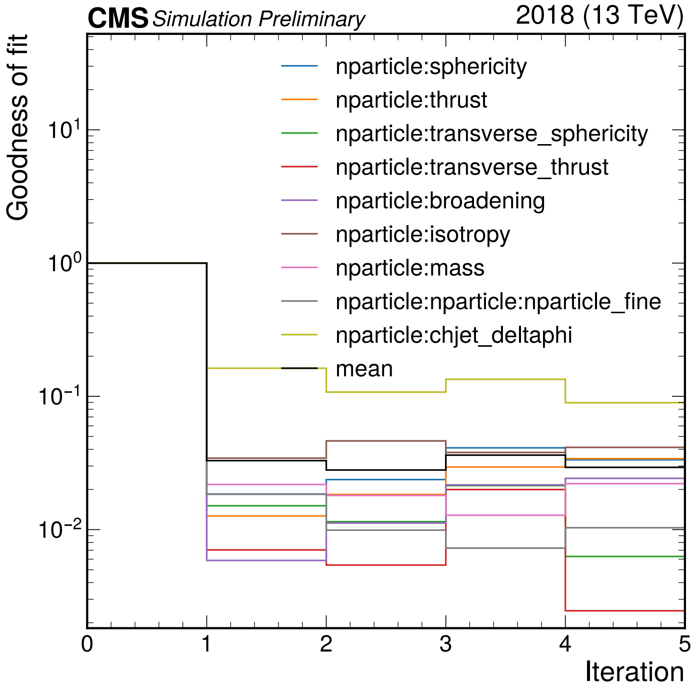
Particle-level
MC, **unfold**, and **pseudo-data truth**
transverse thrust at iteration 2

Validation: unfold the pseudo-data from Pythia A14 tune

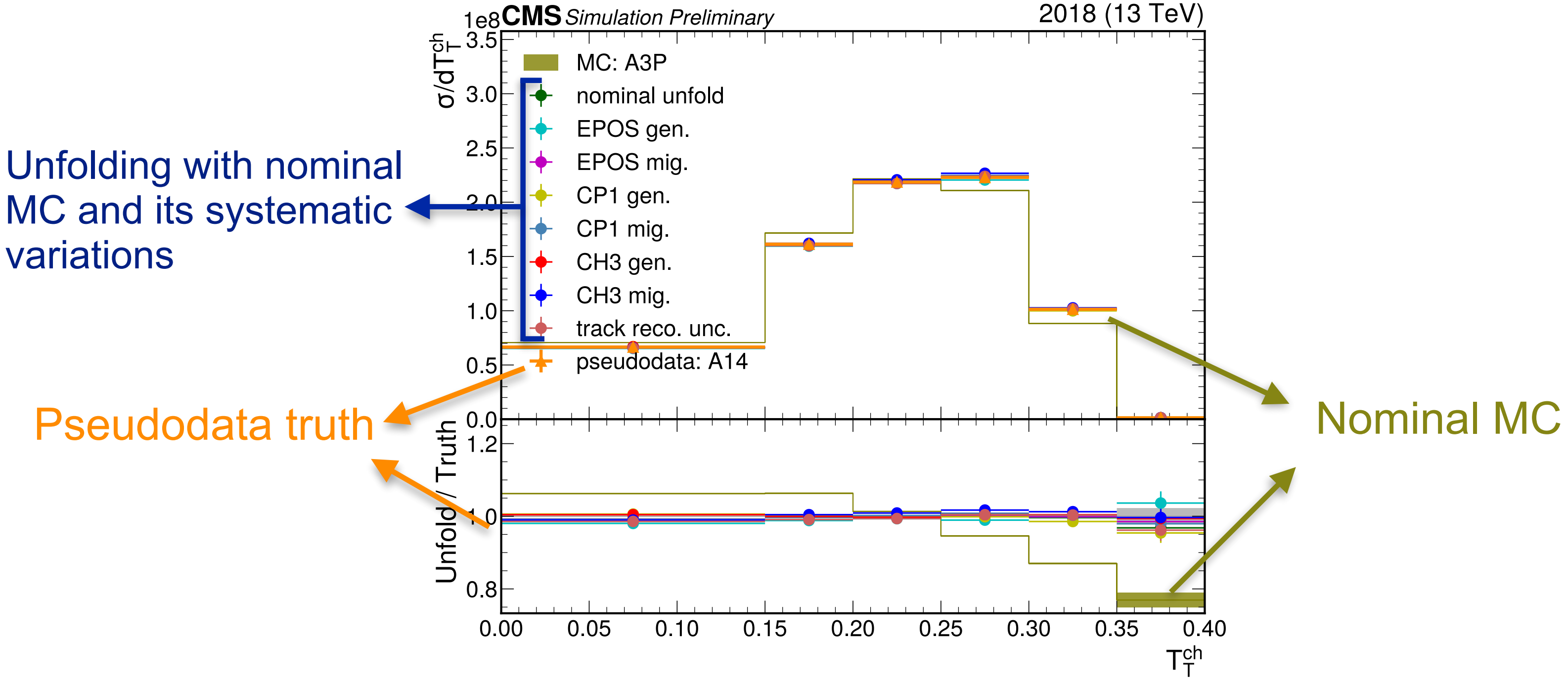
Alternative MC from Pythia A14, CP5 and CUETP8M1 tunes → **pseudo-data**

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

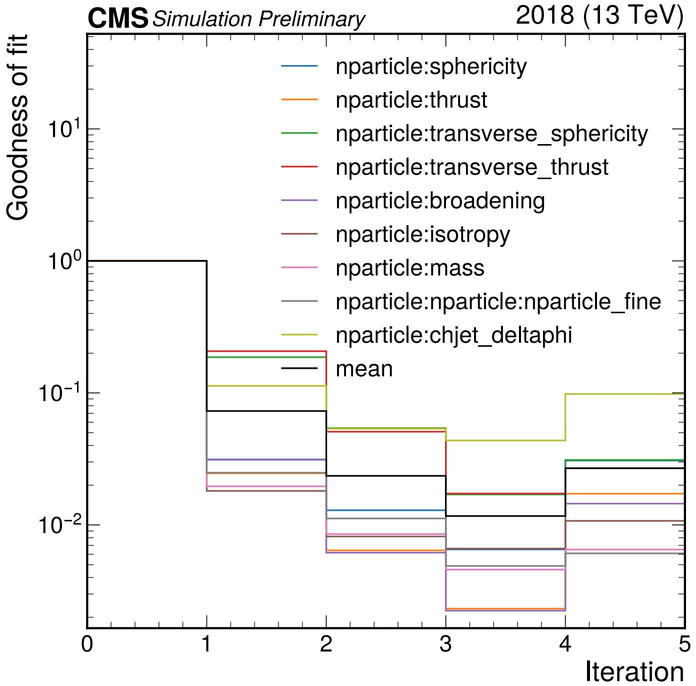


Particle-level
MC, **unfold**, and **pseudo-data truth**
transverse thrust at iteration 2

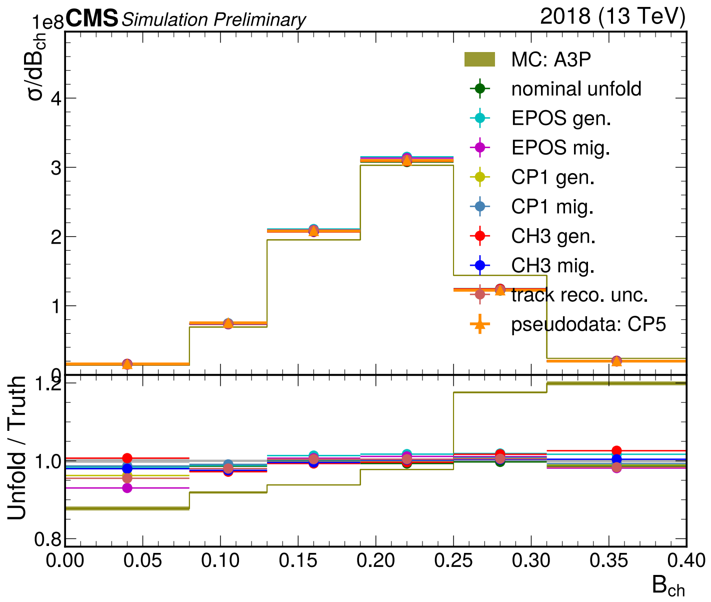
Validation: unfold the pseudo-data from Pythia CP5 tune

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

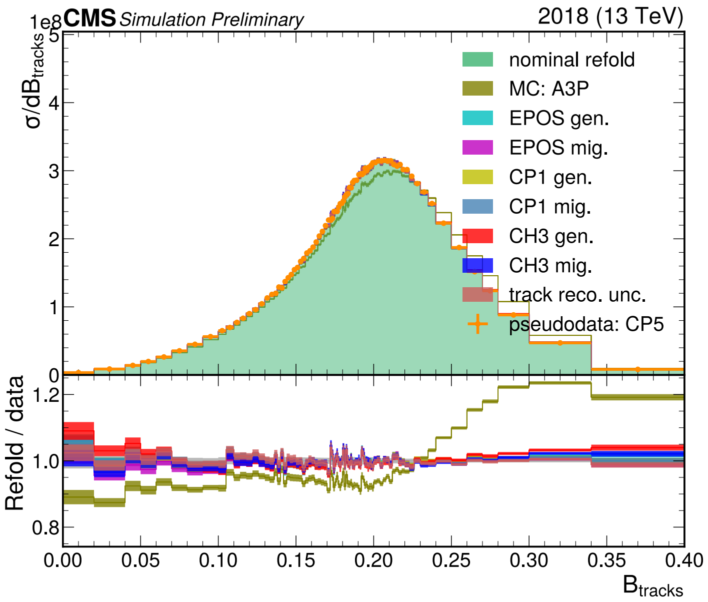


χ^2/χ^2(0th iteration) between the unfolded histograms & pseudo-data truth



Particle-level

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



Detector-level

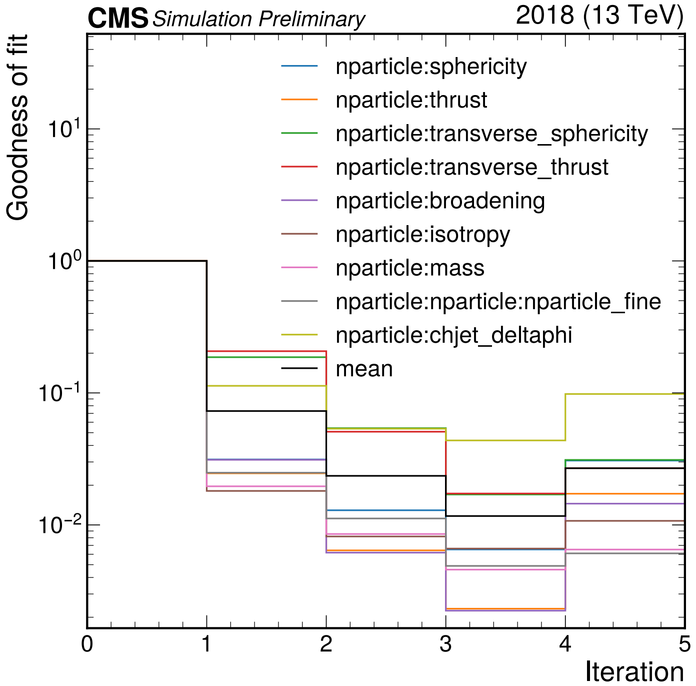
MC, **refold**, and **pseudo-data** broadening at iteration 2

Validation: unfold the pseudo-data from Pythia CP5 tune

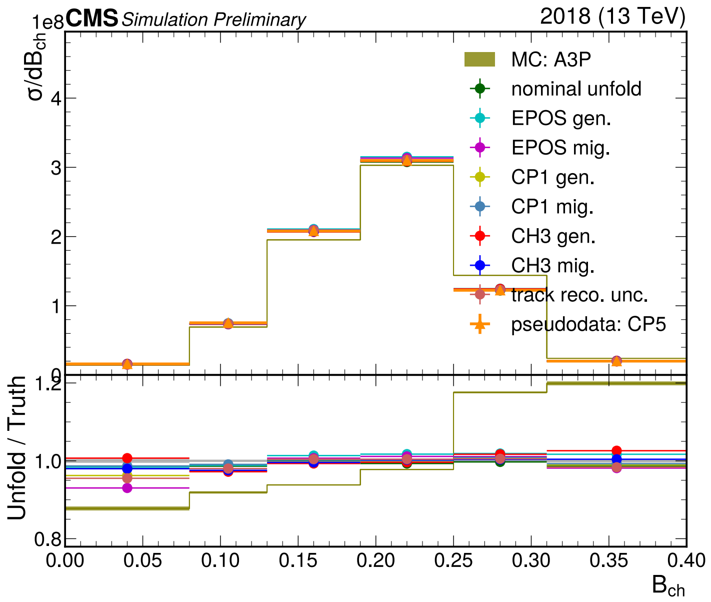
Alternative MC from Pythia A14, CP5 and CUETP8M1 tunes → **pseudo-data**

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

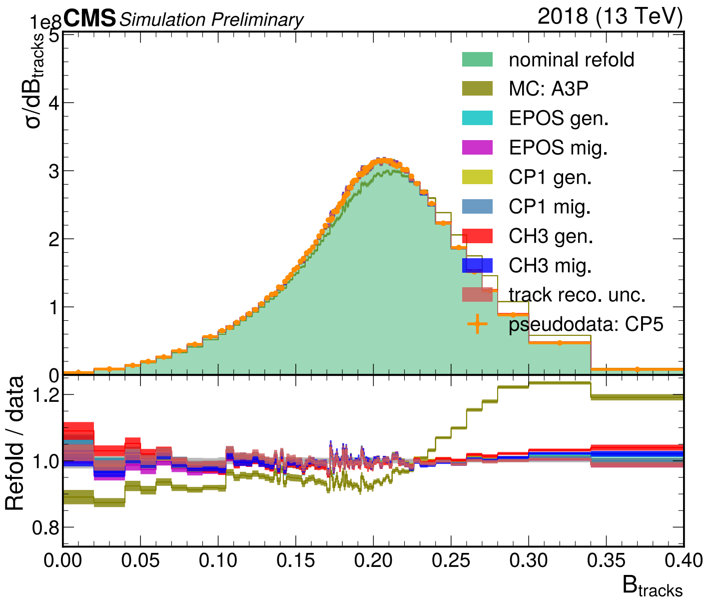
Example: Unfold the Pythia **CP5** sample



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



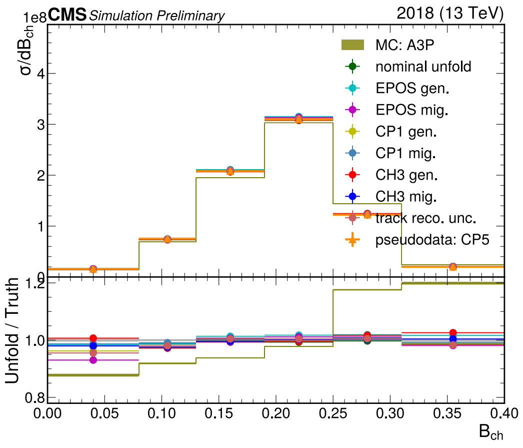
Particle-level
MC, **unfold**, and **pseudo-data truth** broadening at iteration 2

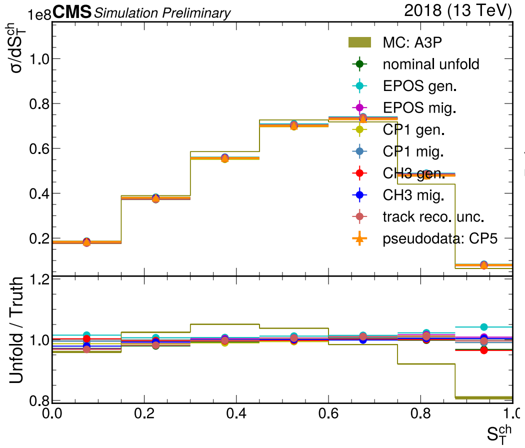


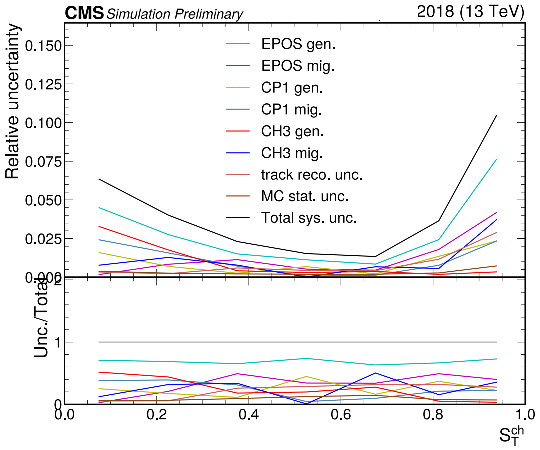
Detector-level
MC, **refold**, and **pseudo-data** broadening at iteration 2

Validation: unfold the pseudo-data with other systematic templates

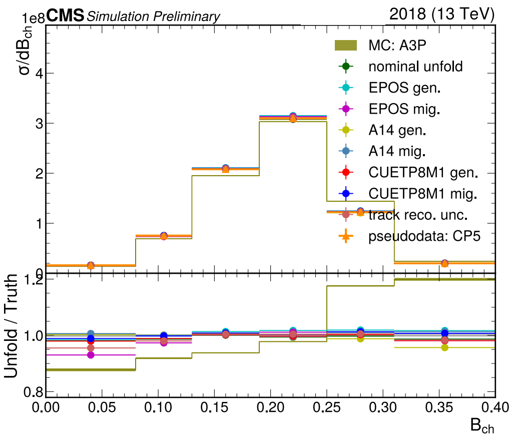
Particle-level
MC, **unfold**, and **pseudo-data truth**

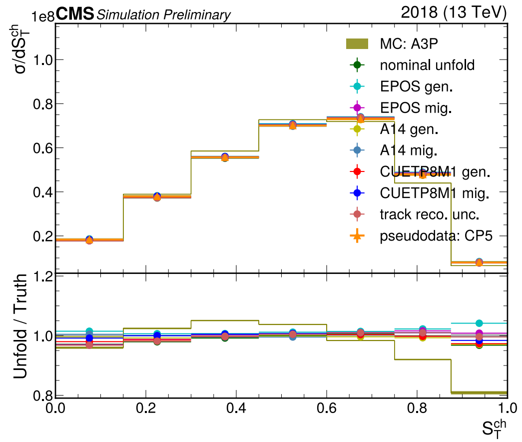


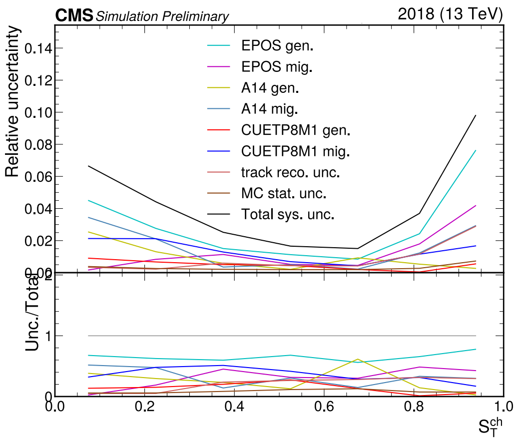




Systematic templates derived from EPOS, **Pythia CP1, Herwig CH3**







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**

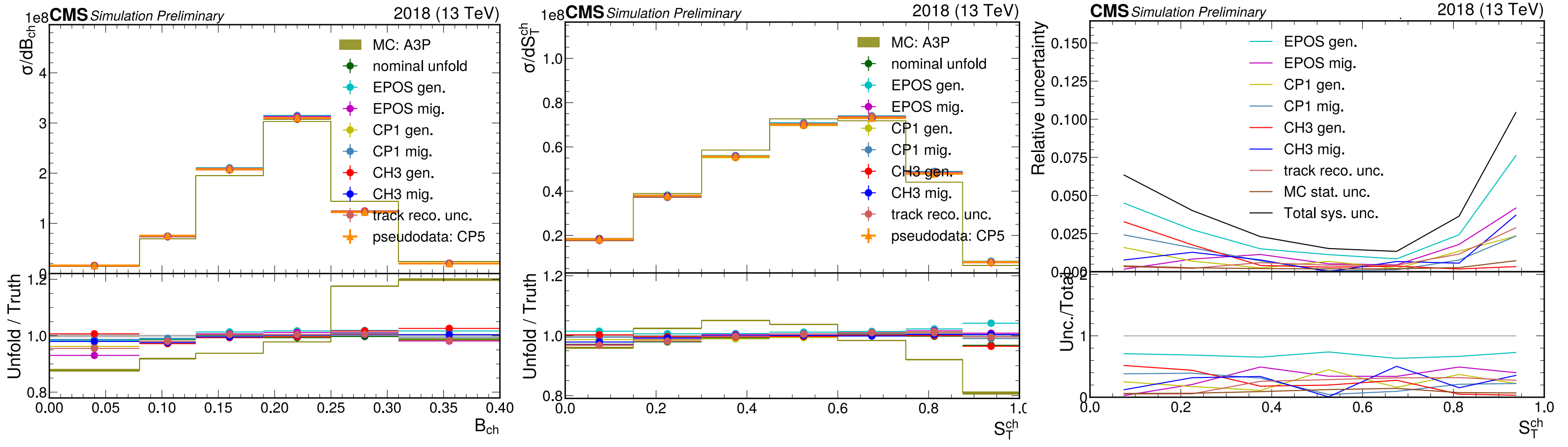
Transverse sphericity
unfold v.s. **truth**

Transverse sphericity
uncertainty decomposition

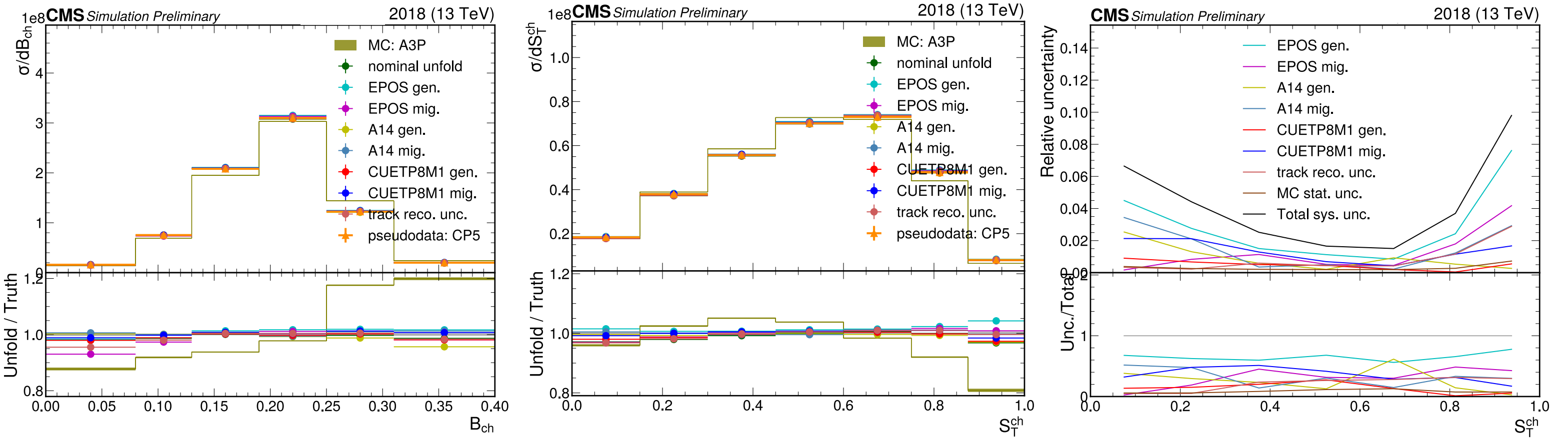
Validation: unfold the pseudo-data with other systematic templates

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

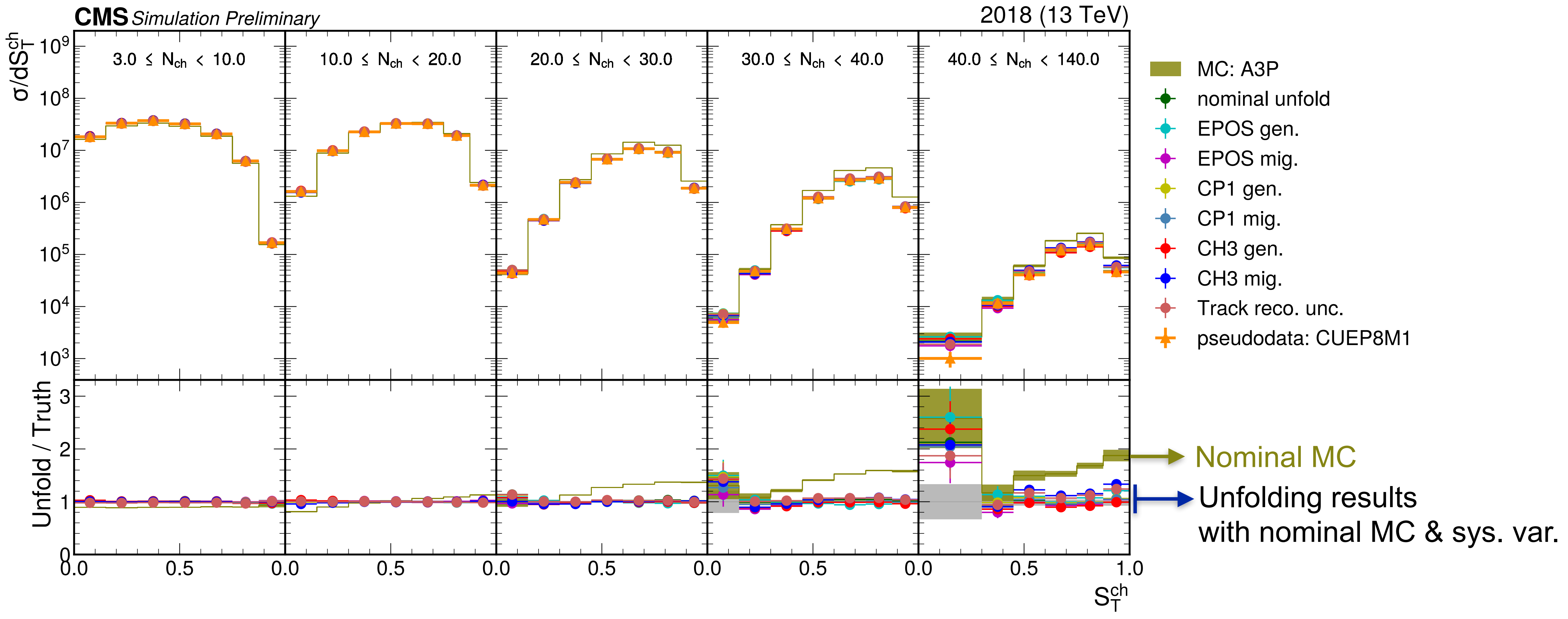
Broadening
unfold v.s. **truth**

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

Transverse sphericity
uncertainty decomposition

Validation: unfold the pseudo-data

Test the unfolding on **2D distributions**

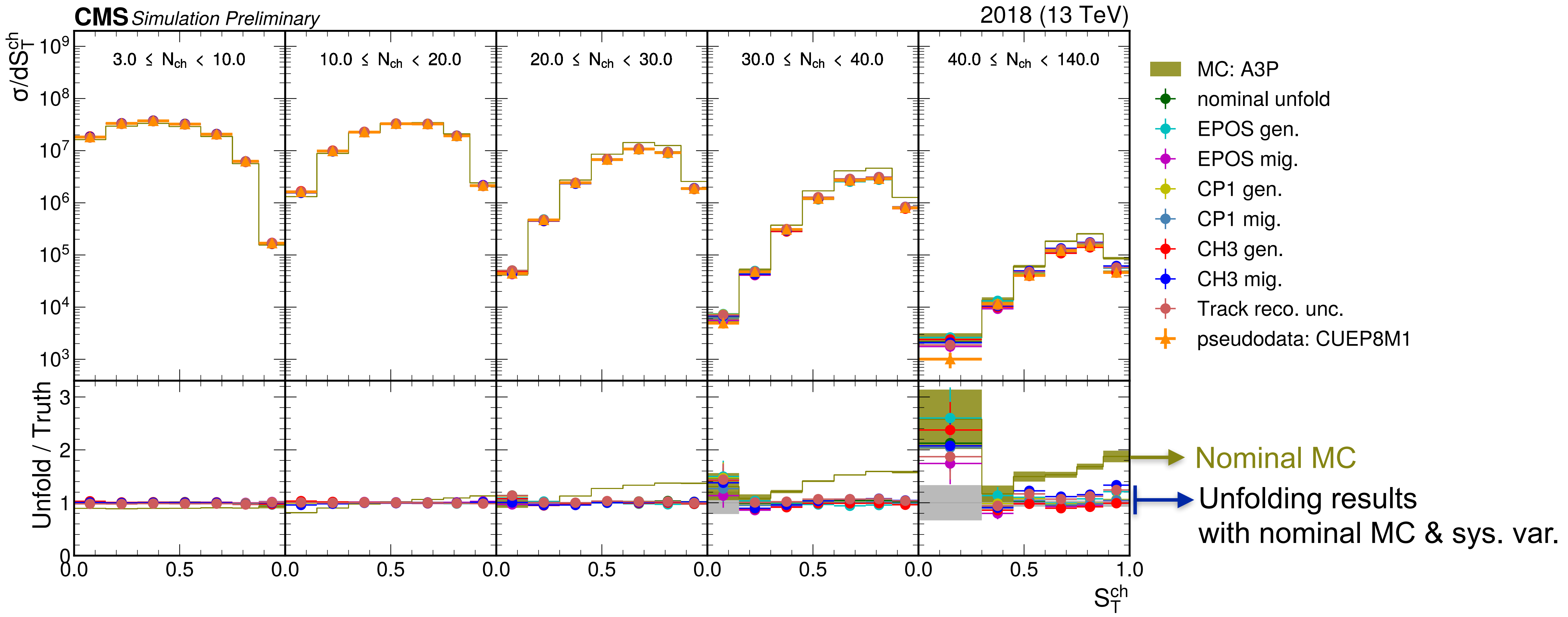


2D test also shows closure between unfolding results and the pseudo-data truth

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

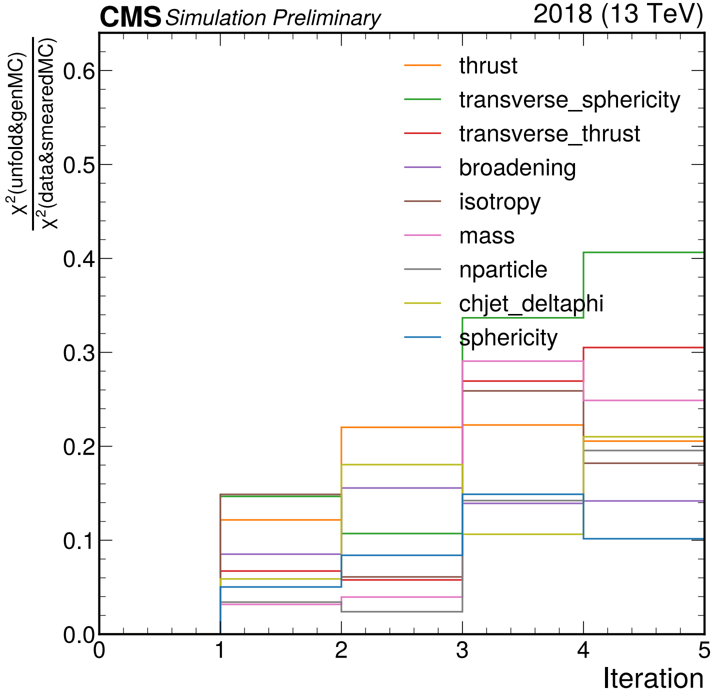
Validation: bottom-line test

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**

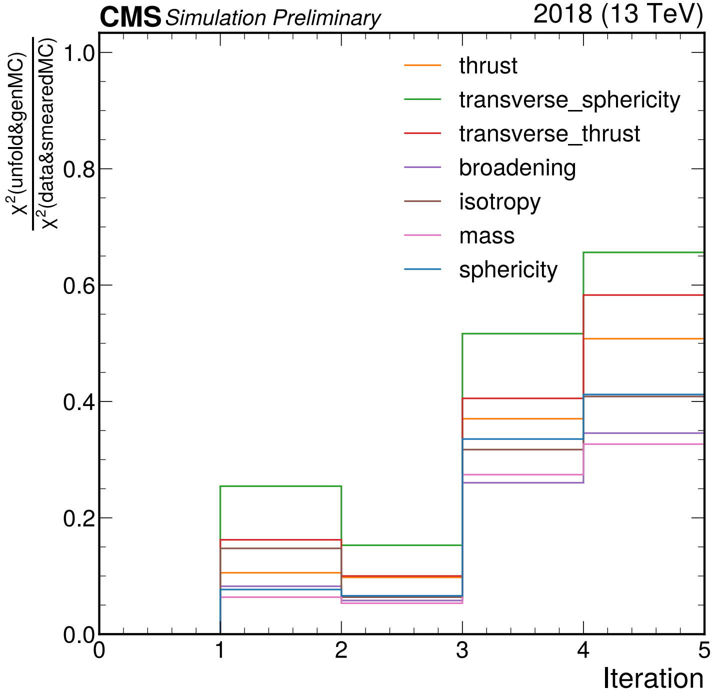
Example: χ^2 (unfold&gen-MC) / χ^2 (data&smeared MC) when unfolding CUEPT8M1 pseudo-data

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

Randomly **shift nominal MC** according to systematic variations
\+ **shift systematic templates** accordingly

$\xrightarrow{\text{unfold}}$

Resampled pseudo-data

Validation: bias and coverage test

Randomly **shift nominal MC** according to systematic variations
\+ **shift systematic templates** accordingly

unfold →

Resampled pseudo-data

Track reconstruction efficiency

Mismodeling of the observables to be unfolded

Mismodeling of other observables

Random samples of nuisance parameters for each sys. unc. θ_2 , θ_{3CP1} , θ_{3EPOS} , θ_{3CH3} , θ_{4CP1} , θ_{4EPOS} , θ_{4CH3} ~N(0,1)

Validation: bias and coverage test

Randomly **shift nominal MC** according to systematic variations
+ **shift systematic templates** accordingly

$\xrightarrow{\text{unfold}}$ **Resampled pseudo-data**

Track reconstruction efficiency

Mismodeling of the observables to be unfolded

Mismodeling of other observables

Random samples of nuisance parameters for each sys. unc. θ_2 , θ_{3CP1} , θ_{3EPOS} , θ_{3CH3} , θ_{4CP1} , θ_{4EPOS} , θ_{4CH3} ~N(0,1)

Deviation from nominal to sys. template: $\vec{w}_i^{\theta_i}$

Validation: bias and coverage test

Randomly **shift nominal MC** according to systematic variations $\xrightarrow{\text{unfold}}$ **Resampled pseudo-data**
+ **shift systematic templates** accordingly

Track reconstruction efficiency

Mismodeling of the observables to be unfolded

Mismodeling of other observables

Random samples of nuisance parameters for each sys. unc. θ_2 , θ_{3CP1} , θ_{3EPOS} , θ_{3CH3} , θ_{4CP1} , θ_{4EPOS} , θ_{4CH3} ~N(0,1)

Deviation from nominal to sys. template: $\overrightarrow{w}_i^{\theta_i}$

Poisson(1) weight on MC events (MC stat. unc.)

Validation: bias and coverage test

Randomly **shift nominal MC** according to systematic variations + **shift systematic templates** accordingly $\xrightarrow{\text{unfold}}$ **Resampled pseudo-data**

Track reconstruction efficiency

Mismodeling of the observables to be unfolded

Mismodeling of other observables

Random samples of nuisance parameters for each sys. unc. θ_2 , θ_{3CP1} , θ_{3EPOS} , θ_{3CH3} , θ_{4CP1} , θ_{4EPOS} , θ_{4CH3} ~N(0,1)

Deviation from nominal to sys. template: $\vec{w}_i^{\theta_i}$

Poisson(1) weight on MC events (MC stat. unc.)

Multiply the weights for all the sys. unc. sources

Validation: bias and coverage test

Randomly **shift nominal MC** according to systematic variations + **shift systematic templates** accordingly $\xrightarrow{\text{unfold}}$ **Resampled pseudo-data**

Track reconstruction efficiency

Mismodeling of the observables to be unfolded

Mismodeling of other observables

Random samples of nuisance parameters for each sys. unc. θ_2 , θ_{3CP1} , θ_{3EPOS} , θ_{3CH3} , θ_{4CP1} , θ_{4EPOS} , θ_{4CH3} ~N(0,1)

↓

Deviation from nominal to sys. template: $\vec{w}_i^{\theta_i}$

↓

Multiply the weights for all the sys. unc. sources

Poisson(1) weight on MC events (MC stat. unc.)

↓

MC bootstrap

Validation: bias and coverage test

Randomly **shift nominal MC** according to systematic variations + **shift systematic templates** accordingly $\xrightarrow{\text{unfold}}$ **Resampled pseudo-data**

Track reconstruction efficiency

Mismodeling of the observables to be unfolded

Mismodeling of other observables

Random samples of nuisance parameters for each sys. unc. θ_2 , θ_{3CP1} , θ_{3EPOS} , θ_{3CH3} , θ_{4CP1} , θ_{4EPOS} , θ_{4CH3} ~N(0,1)

↓

Deviation from nominal to sys. template: $\vec{w}_i^{\theta_i}$

↓

Multiply the weights for all the sys. unc. sources

Poisson(1) weight on MC events (MC stat. unc.)

↓

MC bootstrap

Poisson(1) weight on pseudo-data

↓

Pseudo-data bootstrap

Validation: bias and coverage test

Randomly **shift nominal MC** according to systematic variations
+ **shift systematic templates** accordingly

unfold

Resampled pseudo-data

Track reconstruction efficiency

Mismodeling of the observables to be unfolded

Mismodeling of other observables

Random samples of nuisance parameters for each sys. unc. θ_2 , θ_{3CP1} , θ_{3EPOS} , θ_{3CH3} , θ_{4CP1} , θ_{4EPOS} , θ_{4CH3} ~N(0,1)

Deviation from nominal to sys. template: $\vec{w}_i^{\theta_i}$

Poisson(1) weight on MC events (MC stat. unc.)

Multiply the weights for all the sys. unc. sources

MC bootstrap

Unfold

Poisson(1) weight on pseudo-data

Pseudo-data bootstrap

Validation: bias and coverage test

Randomly **shift nominal MC** according to systematic variations + **shift systematic templates** accordingly $\xrightarrow{\text{unfold}}$ **Resampled pseudo-data**

Track reconstruction efficiency

Mismodeling of the observables to be unfolded

Mismodeling of other observables

Random samples of nuisance parameters for each sys. unc. θ_2 , θ_{3CP1} , θ_{3EPOS} , θ_{3CH3} , θ_{4CP1} , θ_{4EPOS} , θ_{4CH3} ~N(0,1)

↓

Deviation from nominal to sys. template: $\vec{w}_i^{\theta_i}$

↓

Poisson(1) weight on MC events (MC stat. unc.)

Multiply the weights for all the sys. unc. sources

↓

MC bootstrap

↓

Further multiply the weights of sys. templates $\vec{w}_i$

↓

Systematic deviations on the MC bootstrap

Unfold

Poisson(1) weight on pseudo-data

↓

Pseudo-data bootstrap

Validation: bias and coverage test

Randomly **shift nominal MC** according to systematic variations
+ **shift systematic templates** accordingly $\xrightarrow{\text{unfold}}$ **Resampled pseudo-data**

Track reconstruction efficiency

Mismodeling of the observables to be unfolded

Mismodeling of other observables

Random samples of nuisance parameters for each sys. unc. θ_2 , θ_{3CP1} , θ_{3EPOS} , θ_{3CH3} , θ_{4CP1} , θ_{4EPOS} , θ_{4CH3} ~N(0,1)

Deviation from nominal to sys. template: $\vec{w}_i^{\theta_i}$

Poisson(1) weight on MC events (MC stat. unc.)

Multiply the weights for all the sys. unc. sources

MC bootstrap

Unfold

Poisson(1) weight on pseudo-data

Further multiply the weights of sys. templates $\vec{w}_i$

Pseudo-data bootstrap

Systematic deviations on the MC bootstrap

Unfold

Validation: bias and coverage test

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

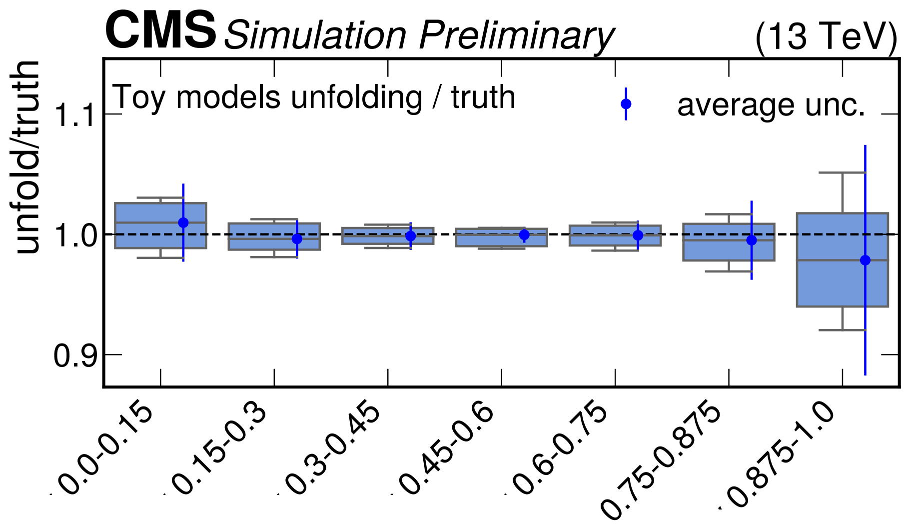
Bias test:

Bias of the unfolding results compared to pseudo-data truth

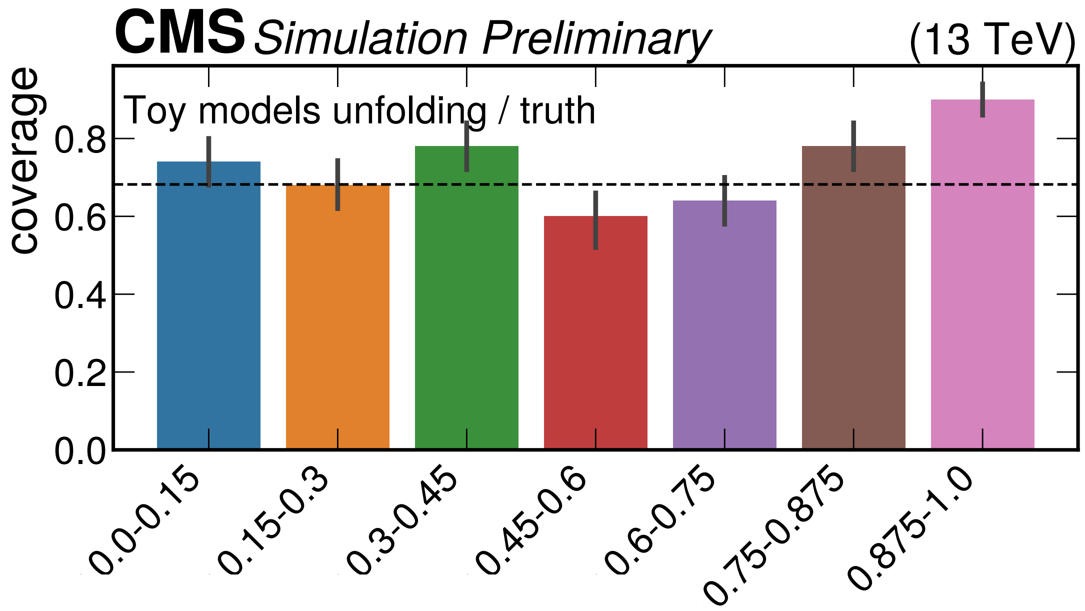
Coverage test:

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

Estimate systematic uncertainty with pseudo-experiments

1.MC statistics templates:
random **Poisson(1) weights** on MC

2. Track reco. eff. template:
weights $\overrightarrow{w_2}$ on nominal MC

3. Mismodel of obs. for unfolding template
weights $\overrightarrow{w}_{3A3}$, $\overrightarrow{w}_{3EPOS}$, $\overrightarrow{w}_{3CH3}$ on nominal MC

4. Mismodel of other obs. template:
weights $\overrightarrow{w}_{4A3}$, $\overrightarrow{w}_{4EPOS}$, $\overrightarrow{w}_{4CH3}$ on nominal MC

Estimate systematic uncertainty with pseudo-experiments

1.MC statistics templates:
random **Poisson(1) weights** on MC

2. Track reco. eff. template:
weights $\overrightarrow{w_2}$ on nominal MC

3. Mismodel of obs. for unfolding template
weights $\overrightarrow{w}_{3A3}$, $\overrightarrow{w}_{3EPOS}$, $\overrightarrow{w}_{3CH3}$ on nominal MC

4. Mismodel of other obs. template:
weights $\overrightarrow{w}_{4A3}$, $\overrightarrow{w}_{4EPOS}$, $\overrightarrow{w}_{4CH3}$ on nominal MC

Sample a nuisance parameters per sys. source θ2 , θ3CP1 , θ3EPOS , θ3CH3 , θ4A3 , θ4EPOS , θ4CH3 ~ N(0,1)
→ output: weights $\overrightarrow{w_i}^{\theta_i}$ mimicing the distribution of the systematic deviations,
i = 2, 3A3, 3EPOS, 3CH3, 4A3, 4EPOS, 4CH3

Estimate systematic uncertainty with pseudo-experiments

1.MC statistics templates:
random **Poisson(1) weights** on MC

2. Track reco. eff. template:
weights $\overrightarrow{w_2}$ on nominal MC

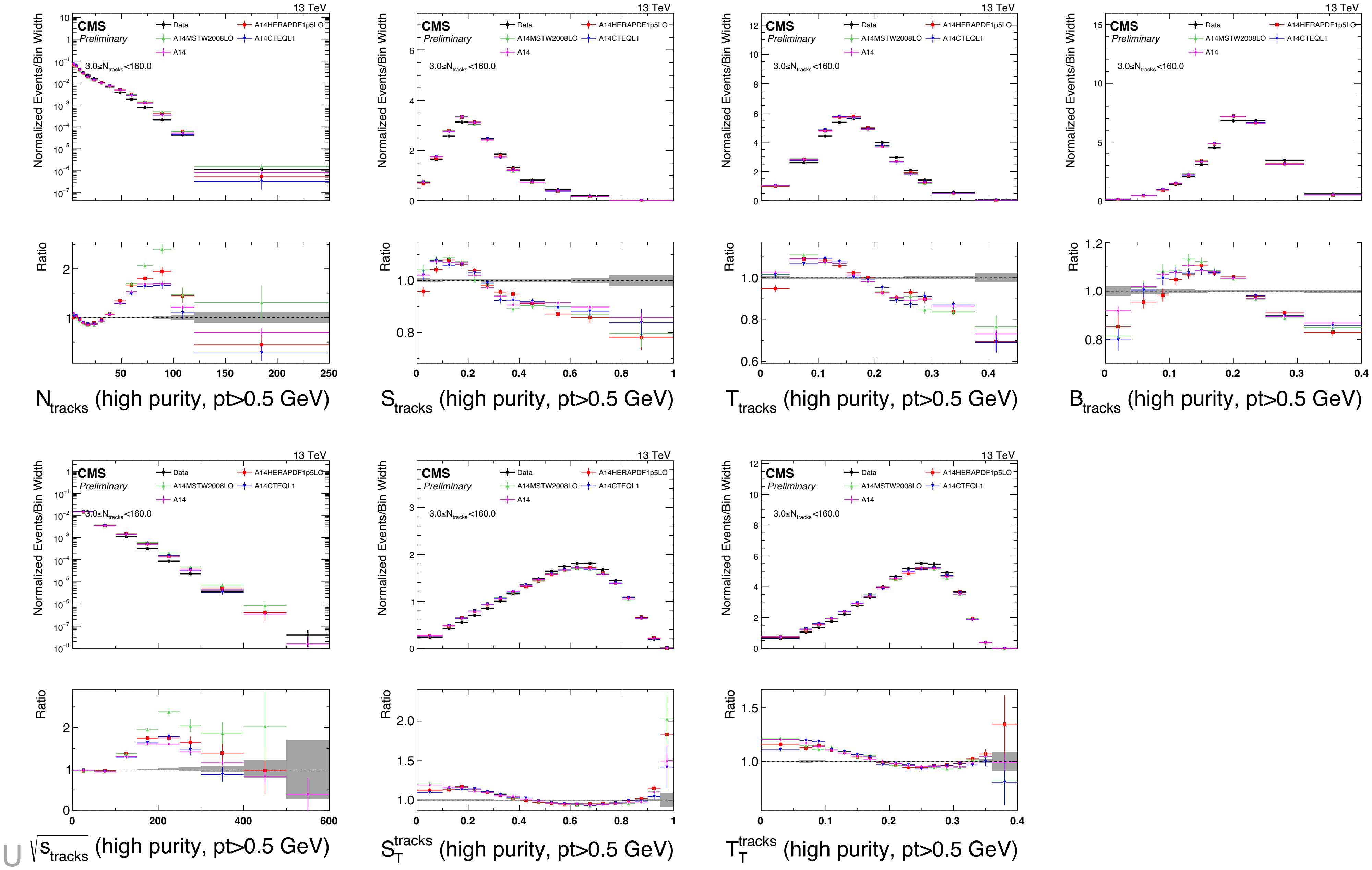
3. Mismodel of obs. for unfolding template
weights $\overrightarrow{w}_{3A3}$, $\overrightarrow{w}_{3EPOS}$, $\overrightarrow{w}_{3CH3}$ on nominal MC

4. Mismodel of other obs. template:
weights $\overrightarrow{w}_{4A3}$, $\overrightarrow{w}_{4EPOS}$, $\overrightarrow{w}_{4CH3}$ on nominal MC

Sample a nuisance parameters per sys. source θ_2 , θ_{3CP1} , θ_{3EPOS} , θ_{3CH3} , θ_{4A3} , θ_{4EPOS} , θ_{4CH3} ~ N(0,1)
→ output: weights $\overrightarrow{w_i}^{\theta_i}$ mimicing the distribution of the systematic deviations,
i = 2, 3A3, 3EPOS, 3CH3, 4A3, 4EPOS, 4CH3

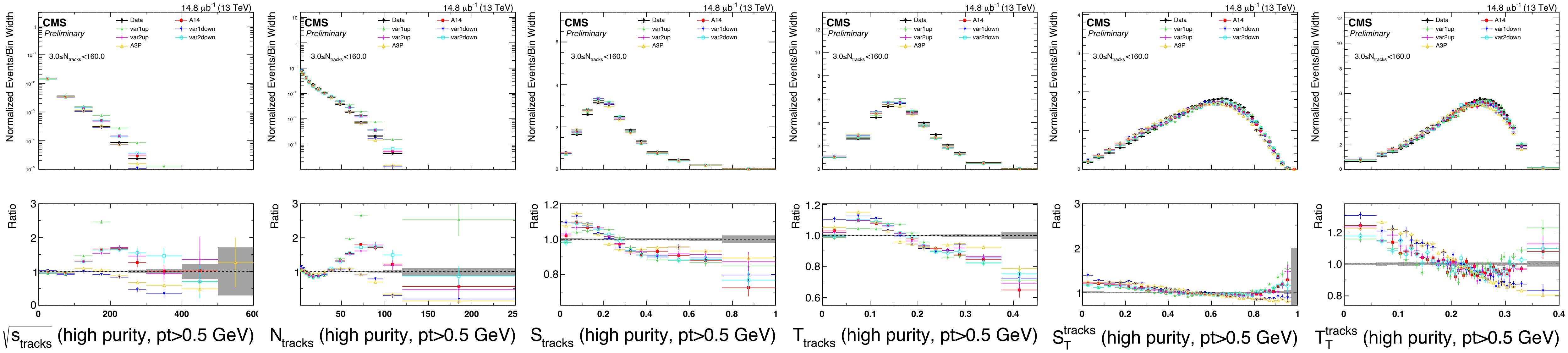
Multiply all the weights → a "**bootstrap**" MC set
→ **unfold** with various bootstrap MC sets
→ derive **systematic uncertainty** and **covariance** from the unfolding results

PDF variations of MC

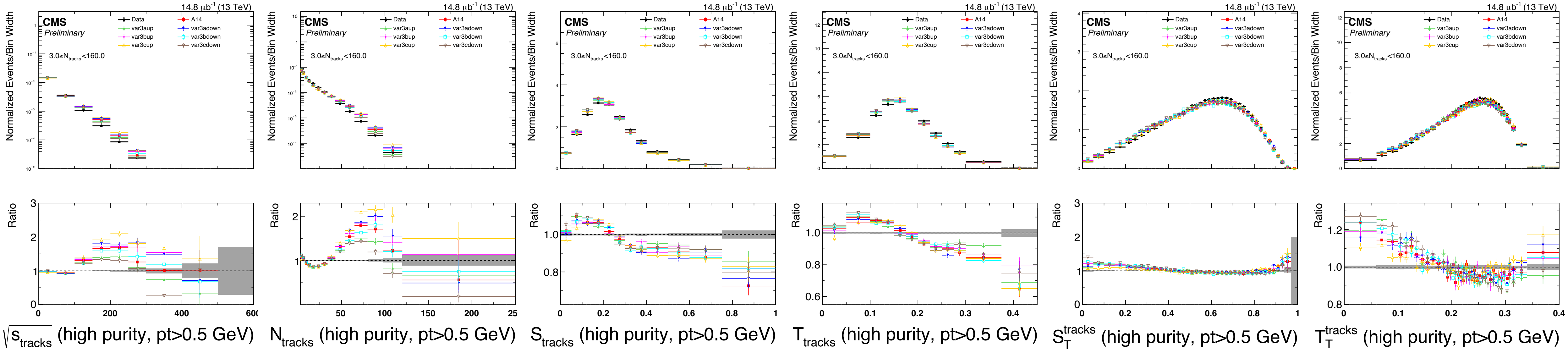


Tune variations of MC

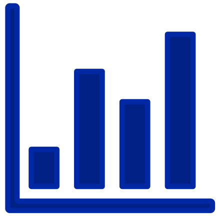
A14 tune eigen-variation 1 and 2



A14 tune eigen-variation 3a, 3b, 3c



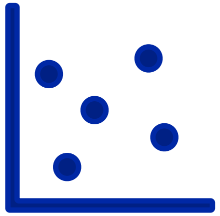
Estimation of systematic uncertainty



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)