PROCEEDINGS OF SCIENCE

Event shape variables in pp collisions in CMS

Soumyadip Barman[a,*] **On behalf of the CMS Collaboration**

[a] *Visva Bharati University, Santiniketan, India*

E-mail: soumyadip.barman@cern.ch

Event shape variables, constructed from the four-momenta of the final-state objects in an event, are sensitive to the models of Quantum Chromodynamics in hadronic final states. These are sensitive to the perturbative as well as non-perturbative aspects of Quantum Chromodynamics. The study of these variables can help us in understanding the geometrical structure of the hadronic final states. In this note, we present two such recent measurements by the CMS Collaboration and comparison with the predictions of QCD inspired theoretical models.

The European Physical Society Conference on High Energy Physics (EPS-HEP2025)
7-11 July 2025
Marseille, France

*Speaker

PoS(EPS-HEP2025)257

1. Introduction

Event shape variables (ESV), constructed from the four-momenta of the final-state objects in an event, are infrared and collinear safe and provide information of energy flow in the hadronic final states. They are sensitive to both perturbative and non-perturbative aspects of Quantum Chromodynamics (QCD), and can be used to test various aspects of QCD. They are robust against various sources of systematic uncertainties. In this note, we present two recent measurements of several ESVs by the CMS Collaboration using data of proton-proton collisions at $\sqrt{s}$ = 13 TeV collected by the CMS detector [1, 2], during the Run 2 of the Large Hadron Collider. CMS is a general-purpose detector with four sub-detectors, i.e., a tracker made of silicon pixels and strips, an electromagnetic calorimeter made of $PbWO_4$ crystals, a brass and plastic scintillator hadron calorimeter and a muon system. CMS uses particle flow approach to combine information from all the sub-detectors to reconstruct each single particle. Detector level distributions of the ESVs are corrected for the detector effects to obtain the true distributions. This makes comparison with the theoretical predictions and results from other experiments easier.

2. Measurement of event shape variables using charged particles inside jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV

2.1 Introduction

In this analysis [3], the ESVs are constructed using the four-momenta of the charged particles inside jets. Using particles instead of jets reduce theoretical issues, e.g. collinear and soft emissions of partons and the dependence of the jet clustering algorithm. However, events in Run 2 of LHC have large pile-up (PU) contamination and its effect is mitigated at the jet level but not at the particle level. On the other hand, charged particles are much less affected by PU, and hence, have been used for evaluating the ESVs.

The ESVs are defined such a way that they have a minimum value of zero for events with two back-to-back jets, and a higher value for spherical, multijet events where the jets are distributed isotropically. The five ESVs are defined below.

- **Complement of transverse thrust :** defined as $\tau_\perp \equiv 1 - T_\perp$ where, $T_\perp = \max_{\hat{n}_T} \left(\frac{\sum_i |\vec{p}_{T,i} \cdot \hat{n}_T|}{\sum_i |\vec{p}_{T,i}|} \right)$, $0 \geq \tau_\perp \geq 1 - 2/\pi$
- **Third-jet resolution parameter :** $Y_{2,3} = \frac{\max\left(p_{T,3}^2, \left\{\min(p_{T,i}, p_{T,j})^2 \cdot \frac{(\Delta R_{ij})^2}{R^2}\right\}\right)}{P_{12}^2}$
- **Jet broadening :** $B_T \equiv B_U + B_L$, where, $B_X = \frac{1}{2P_T} \sum_{i \in C_X} p_{T,i} \sqrt{(\eta_i - \eta_X)^2 + (\phi_i - \phi_X)^2}$, $\eta_X = \frac{\sum_{i \in C_X} p_{T,i}\,\eta_i}{\sum_{i \in C_X} p_{T,i}}$, , $\phi_X = \frac{\sum_{i \in C_X} p_{T,i}\,\phi_i}{\sum_{i \in C_X} p_{T,i}}$
- **Total jet mass :** $\rho_{Tot} \equiv \rho_U + \rho_L$ where, $\rho_x \equiv \frac{M_x^2}{P^2}$
- **Total transverse jet mass :** corresponding quantity of total jet mass in the transverse plane (ρ_{Tot}^T)

According to the definitions $\tau_\perp$ is sensitive to the initial perturbative part (hard scattering, hard gluon emission, etc.) whereas the jet masses and jet broadening depend more on the subsequent non-perturbative part (fragmentation, hadronization, etc.).

2.2 Event Selection and Samples

Five ESVs are studied using data corresponding to an integrated luminosity of $\mathcal{L}$ = 138 fb^{-1} collected during 2016-2018. Particle flow objects are combined to form jets using the anti-k_T algorithm [4] with R=0.4. Events are selected by requiring at least two jets with $p_T >$ 30 GeV and $|\eta| < 2.1$. Charged particles inside jets with $|\eta| < 2.5$ are used to evaluate the ESVs.

The data sample is divided into eight $H_{T,2}$ phase-space regions, depending on the trigger efficiency curves, where $H_{T,2}$ is defined as the average p_T of the jets having the highest (p_{T,j_1}) and the second highest (p_{T,j_2}) in an event. The eight $H_{T,2}$ ranges which are used in this analysis are (in GeV) – 83–109, 109–176, 176–247, 247–318, 318–387, 387–477, 477–573 and ≥ 573.

Three Monte Carlo generators Pythia8 v8.240 [5], MadGraph5_aMC@NLO v2.6.5 [6] and Herwig7 v7.2.2 [7] are used to simulate pp collision at $\sqrt{s}$ = 13 TeV. Simulated events are passed through Geant4 [8] package for considering CMS detector response. Simulated events are selected using the same criteria as collision events.

2.3 Unfolding & Uncertainties

In this analysis, two-dimensional (2D) unfolding is considered for each ESV involving the variable itself and $H_{T,2}$. The unfolding is performed using the TUnfold [9] framework to obtain the particle-level information. A number of closure tests have been done to ensure that the unfolding is done correctly.

For this analysis, possible sources of experimental systematic uncertainty are considered – jet energy scale (JES), jet energy resolution (JER), pileup, track reconstruction efficiency, track p_T resolution, etc. Monte Carlo model dependence of unfolding and parton density function (PDF) have been considered for theoretical uncertainties. The major contributions to the uncertainty are from JES (1–3%) and Monte Carlo model dependence (1–5%). The contribution from statistical uncertainty is small, < 1%, for all the five variables. The total relative uncertainty is 1-6% for all of them.

2.4 Results

Normalized unfolded distributions of the five ESVs are compared with the predictions of Pythia8, MadGraph5_aMC@NLO and Herwig7, see Figure 1. Pythia8 shows overall good agreement with data for $\tau_\perp$ and ρ^T_{Tot}, and overestimates the spherical, multijet events for ρ_{Tot}, Y_{23} and B_T. The agreement improves with increasing $H_{T,2}$ for all the ESVs. Herwig7 shows similar trends like Pythia8. The agreement between data and MadGraph5_aMC@NLO + Pythia8 is good in lower $H_{T,2}$ values, whereas it overestimates the share of 2-jet events and correspondingly underestimates the fraction of spherical, multijet events for all the ESVs except Y_{23} for higher $H_{T,2}$ values. Overall, the results indicate that our understanding of the energy flow in an event at different energy scales needs improvement.

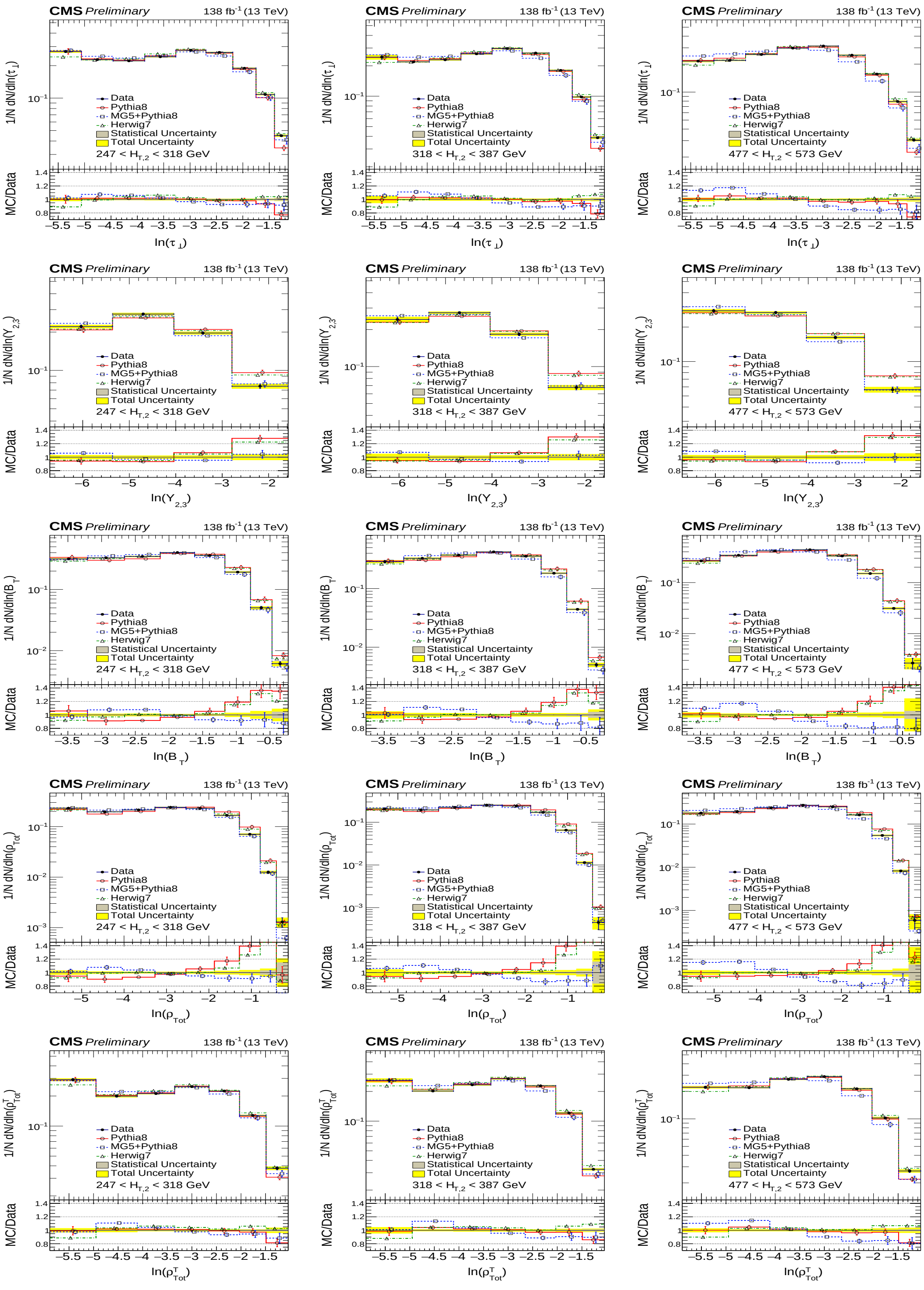

Figure 1: Normalized unfolded distributions of the $\tau_\perp$ (row 1), $Y_{2,3}$ (row 2), B_T (row 3), ρ_{Tot} (row 4), ρ^T_{Tot} (row 5) compared with the predictions of Pythia8 (CP5 tune) (red, solid), Herwig7 (CH3 tune) (green, dash-dotted) and MadGraph5_aMC@NLO + Pythia8 (blue, dotted) for three $H_{T,2}$ ranges (in GeV), i.e. 247-318, 318-387, 477-573. For each plot the inner grey band and the yellow band represent the statistical and total uncertainties and the lower panel shows the ratio of Monte Carlo to data. The error bars corresponding to Pythia8 (in red) and MadGraph5_aMC@NLO + Pythia8 (in blue) represent their respective variation due to the renormalization and factorization scales.

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

3.1 Introduction

The study of global event observable measurements in proton-proton collisions, e.g. correlations of outgoing particles [10] and multiplicity of strange hadrons [11] suggests difficulties in modelling. The models that attempt to describe these effects were developed for heavy ion collisions and include effects such as collective flow or rope hadronization. Detecting QCD instantons, which produce relatively isotropic distribution of particles is also of interest. These issues motivate the measurement of observables sensitive to nonpertubative effects in minimum-bias pp collisions. Moreover, dominant effect of soft QCD uncertainties suggests the need for experimental input to non-perturbative QCD. In this paper [12], eight event shape observables are studied using minimum-bias pp collisions.

The following ESVs are studied in this paper.

- **Particle multiplicity :** N
- **Total invariant mass :** $\sqrt{s}$
- **Sphericity :** measure of how isotropically the momenta are distributed, $S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2}$
- **Thrust :** measure of how highly collimated the momenta in an event along one particular axis, $T = 1 - \max_{\hat{n}} \left(\frac{\sum_i |\vec{p}_i \cdot \hat{n}|}{\sum_i |\vec{p}_i|} \right)$
- **Broadening :** measure of the fraction of energy perpendicular to the thrust axis, $B \equiv B_L + B_R$ where, left and right broadening is expressed as, $B_L = \sum_{i\in L} \frac{|\vec{p}_i \times \vec{n}|}{\sum_i |\vec{p}_i|}$, $B_R = \sum_{i\in R} \frac{|\vec{p}_i \times \vec{n}|}{\sum_i |\vec{p}_i|}$
- **Transverse sphericity :** sphericity in the transverse plane, calculated by setting p_i^z to 0 in the sphericity definition
- **Transverse thrust :** the thrust in the transverse plane, calculated by setting p_i^z to 0 in the thrust definition
- **Isotropy :** estimated using the distance measure $d_{ij} = 2(1 - \cos\theta_{ij})$, where θ_{ij} is the angle between two particles.

3.2 Event Selection and Samples

Eight ESVs are studied using low PU pp collisions recorded in 2018, with an average of about one inelastic pp collision per bunch crossing, corresponding to an integrated luminosity of $\mathcal{L} = 64$ μb^{-1}.

ESVs are calculated with tracks without clustering into jets. Events must have one primary vertex with ± 2 (± 24) cm in the transverse (longitudinal) direction with at least three tracks associated with it. All tracks must have $p_T > 0.5$ GeV and $|\eta| < 2.4$. The track impact parameter in the transverse (longitudinal) direction must be at most 8 (13) times the vertex resolution in the transverse (longitudinal) direction.

Simulated events from Monte Carlo event generators Pythia8 (CP1), Pythia8 (A3), Pythia8 (A14), Pythia8 (CP5), EPOS LHC [13] and Herwig7 (CH3) are used in this study. A minimally biased set of inelastic collisions is selected by requiring the MC events to have ≥ 2 charged particles with $p_T > 0.5$ GeV and $|\eta| < 2.4$ at the particle level.

3.3 Unfolding & Uncertainties

Multi-fold machine learning-based unfolding [14] is used in this study. This uses multiple reweighting steps on each iteration, first correcting the simulation towards the data at detector level, then correcting the original simulation towards this corrected simulation at the particle level. Additional sub-steps are also used to account for acceptance and efficiency effects. The unfolding procedure is iterated multiple times, with iteration serving as a form of regularization. Fewer iterations provide stronger regularization and bias towards the input simulation. The number of iterations was determined using goodness-of-fit tests on pseudodata, ensuring no significant improvement with additional iterations. All the observables are unfolded simultaneously using neural network, considering the correlation of the observables. Several tests are also conducted for validating the unfolding process. The final result of the unbinned unfolding is a set of reweighted MC events, estimating the probability density of real data in phase space. The results are binned after the unfolding to produce histograms of the differential cross-section as a function of the observable under study.

In this analysis, systematic uncertainties are larger than the statistical uncertainties. The uncertainty from tracking efficiency is the largest, ∼1.7%. The uncertainty from the migration functions for different physics models combined across all models is ∼1.6%. Whereas, the uncertainty from the regularization bias between the nominal model and other models is typically $O(1\%)$, and the combined uncertainty from regularization bias from all models is 1.2%. The statistical uncertainties are less than 1%. Total relative uncertainty for one-dimensional distribution is $O(5\%)$.

3.4 Results

Normalized unfolded distributions of eight ESVs are compared with the predictions of Pythia8 (CP1), Pythia8 (A3), Pythia8 (A14), Pythia8 (CP5), EPOS LHC and Herwig7 (CH3), see Figure 2. All generators overestimate the charged particle multiplicity and invariant mass in the low multiplicity region and underestimate at intermediate values. Moreover large difference is observed with data at the high multiplicity region. Other ESVs show a consistent trend where the unfolded data are more isotropic than the simulation. The discrepancy between data & MC is largest in the middle of the charged-particle multiplicity for the spherical distributions. Other ESVs also show similar trends. Variations of PDF, UE tune, colour-reconnection models, and α_s (FSR), etc., do not change the qualitative nature of the MC predictions vis-a-vis data. These investigations show that the data continue to prefer more isotropic distributions even within slices of charged-particle multiplicity and when considering only the plane transverse to the beam direction. The unfolded distributions from this measurement can be used to improve and develop existing models of pp collisions. These are also critical for understanding quark-gluon plasma and topological effects in non-abelian gauge theories.

PoS(EPS-HEP2025)257

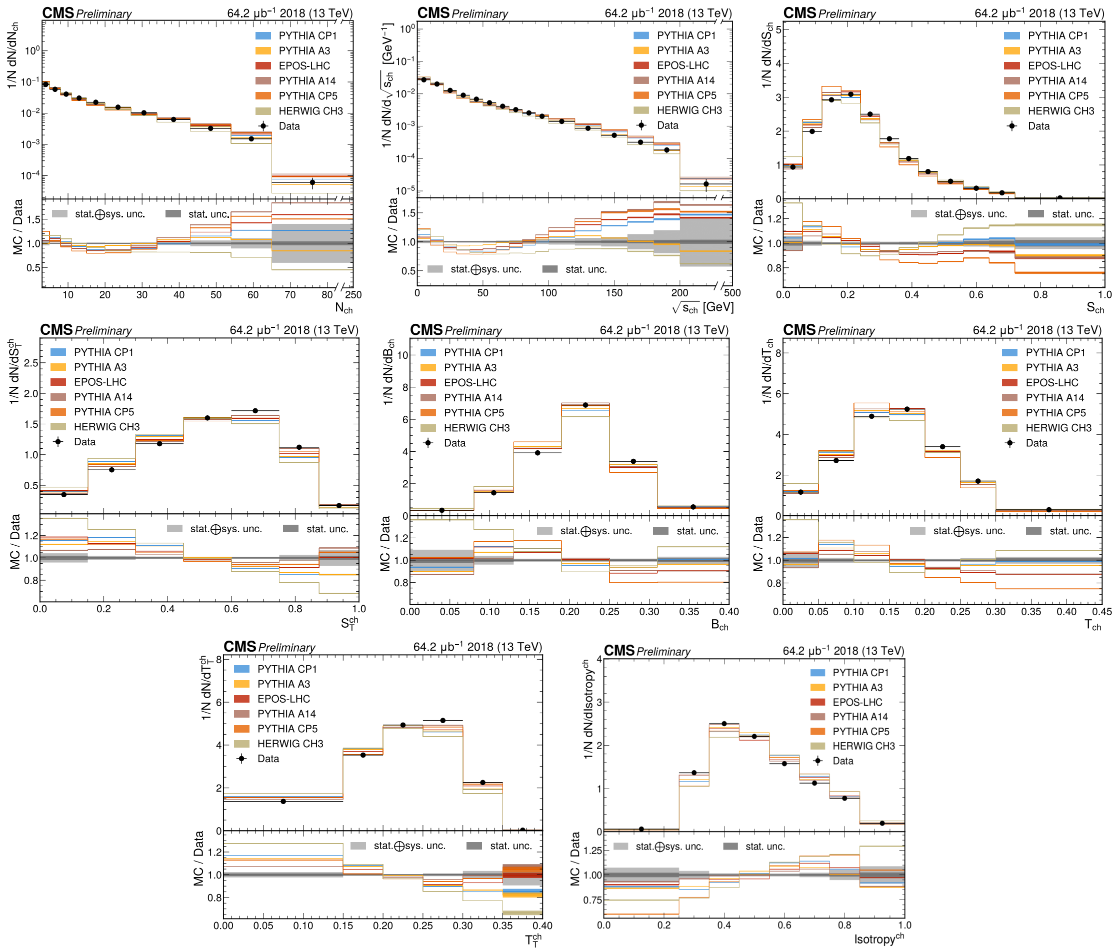

Figure 2: Normalized unfolded distributions of the N_{ch} (row 1, left), $\sqrt{s_{ch}}$ (row 1, middle), S_{ch} (row 1, right), S_T^{ch} (row 2, left), B_{ch} (row 2, middle) T_{ch} (row 2, right), T_T^{ch} (row 3, left) and $Isotropy^{ch}$ (row 3, right) are compared with the predictions PYTHIA8 (CP1 tune), PYTHIA8 (A3 tune), EPOS-LHC, PYTHIA8 (A14 tune), PYTHIA8 (CP5 tune) and HERWIG7 (CH3 tune).

4. Summary

The event shape variables are important in understanding both perturbative and non-perturbative effects in Quantum Chromodynamics at the Large Hadron Collider. In this note, we have presented two recent measurements of event shape variables using proton-proton collision data collected by the CMS detector. Detector-level distributions of several event shape observables are unfolded for correcting the detector effects using standard unfolding technique as well as machine learning based techniques. Both measurements show that the Monte Carlo models cannot represent the data completely. These measurements can be used to improve the theoretical predictions.

References

[1] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.

[2] CMS Collaboration. Development of the cms detector for the cern lhc run 3. *Journal of Instrumentation*, 19(05):P05064, may 2024. doi: 10.1088/1748-0221/19/05/P05064.

[3] CMS Collaboration. Measurement of event shape variables using charged particles inside jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV. CMS-PAS-SMP-22-004, 2025. URL https://cds.cern.ch/record/2929177.

[4] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.

[5] Torbjörn Sjöstrand et al. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015. doi: 10.1016/j.cpc.2015.01.024.

[6] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.

[7] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76:196, 2016. doi: 10.1140/epjc/s10052-016-4018-8.

[8] S. Agostinelli et al. GEANT4: A simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003. doi: 10.1016/S0168-9002(03)01368-8.

[9] Stefan Schmitt. TUnfold: an algorithm for correcting migration effects in high energy physics. *JINST*, 7:T10003, 2012. doi: 10.1088/1748-0221/7/10/T10003.

[10] Long- and short-range correlations and their event-scale dependence in high-multiplicity pp collisions at $\sqrt{s}$ = 13 tev. *Journal of High Energy Physics*, 2021(5), may 2021. ISSN 1029-8479. doi: 10.1007/jhep05(2021)290.

[11] Enhanced production of multi-strange hadrons in high-multiplicity proton–proton collisions. *Nature Physics*, 13(6):535–539, April 2017. ISSN 1745-2481. doi: 10.1038/nphys4111.

[12] CMS Collaboration. Measurement of event shapes in minimum bias events from pp collisions at 13 TeV. CMS-PAS-SMP-23-008, 2024. URL https://cds.cern.ch/record/2899591.

[13] T. Pierog, Iu. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner. Epos lhc: Test of collective hadronization with data measured at the cern large hadron collider. *Phys. Rev. C*, 92:034906, Sep 2015. doi: 10.1103/PhysRevC.92.034906.

[14] Anders Andreassen, Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler. Omnifold: A method to simultaneously unfold all observables. *Phys. Rev. Lett.*, 124: 182001, May 2020. doi: 10.1103/PhysRevLett.124.182001.