Suplemento de la Revista Mexicana de Física **4** 021116 (2023) 1–4

Extraction of multiparton interactions from ALICE pp collisions data using machine learning

E. Zepeda and A. Ortiz

Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Apartado Postal 70-543, Ciudad de México, 04510, México.

Received 19 March 2023; accepted 20 July 2023

Over the last years, Machine Learning methods have been successfully applied to a wealth of problems in high-energy physics. In this work, we discuss the extraction of the average number of Multiparton Interactions ($\langle N_{\rm mpi}\rangle$) from minimum-bias pp data at LHC energies using a regression based on Boosted Decision Trees. Using the available ALICE data on transverse momentum spectra as a function of multiplicity, we report that for minimum-bias pp collisions at $\sqrt{s}=7$ TeV the average $N_{\rm mpi}$ is 3.98 ± 1.01, which complements our previous results for pp collisions at $\sqrt{s}=5.02$ and 13 TeV. The comparisons indicated a modest center-of-mass energy dependence of $\langle N_{\rm mpi}\rangle$. The study is further extended extracting the multiplicity dependence of $N_{\rm mpi}$ for the three center-of-mass energies. These results are qualitatively consistent with the existing ALICE measurements sensitives to Multiparton Interactions. Through the regression applied to pp collisions at $\sqrt{s}=13$ TeV, we also show that computing the multiplicity in the forward region the extraction of $N_{\rm mpi}$ is improved. This result opens the possibility to extract the number of Multiparton Interactions event-by-event, and in this way study the particle production as a function of that quantity. Our results provide additional evidence of the presence of Multiparton Interactions in hadronic interactions and can help to the understanding of the heavy-ion-like behaviour observed in pp collisions data.

Keywords: LHC; Multiparton Interactions; pp collisions; heavy-ion collisions; machine learning.

DOI: https://doi.org/10.31349/SuplRevMexFis.4.021116

1. Introduction

In hadronic interactions it is possible to have Multiparton Interactions (MPI), *i.e.* several parton-parton interactions within the same hadron-hadron collision, given the composite nature of hadrons. The presence of MPI in pp collisions is supported by data [1-4], the study of its effects in pp collisions has recently attracted the attention of the heavy-ion community, because surprisingly, the high-multiplicity pp data unveiled heavy-ion-like features, *i.e.* azimuthal anisotropies [5], the enhancement of (multi-)strange hadrons [6], as well as radial flow patterns in the transverse momentum ($p_{\rm T}$) spectra of identified hadrons [7]. Besides the hydrodynamical approach [8, 9], MPI offers an alternative to explain the observed phenomena. For instance, colour reconnection and MPI can mimic radial flow patterns in pp collisions [10]. In this direction, we have proposed the extraction of the MPI activity from minimum-bias pp data using ML techniques [4, 11]. In this contribution, we summarize the main results including the multiplicity dependence of the average number of MPI extracted from the available ALICE pp data at the LHC [7, 12].

2. Analysis

Our approach relies on a Machine Learning (ML) regression technique based on Boosted Decision Trees (BDT). A regression tree is a binary tree structured regressor in which repeated yes or not decisions are taken on one single variable. In this way, the phase space is split into many regions where each output node represents a specific value of the target variable. For regression tasks the boosting algorithm used is the gradient boost, this procedure tries to minimize the loss-function which describes how the model is predictive with respect to the training data. The study is conducted using the Toolkit for Multivariate Analysis (TMVA) framework which provides a ROOT-integrated ML environment for the processing and parallel evaluation of multivariate classification and regression techniques [13].

The training is performed using pp collisions at $\sqrt{s}=13$ TeV simulated with PYTHIA 8.244 [14] tune 4C [15]. The choice of the input variables is based on their correlation with $N_{\rm mpi}$ [16]. We consider the event-by-event average transverse momentum and the mid-pseudorapidity charged particle multiplicity ($N_{\rm ch}$). Based on the kinematic restrictions from the ALICE data, these quantities are calculated for primary charged particles within $|\eta|<0.8$, and the average $p_{\rm T}$ considers tracks with transverse momentum above 0.15 GeV/c.

The systematic uncertainties take into account variations of the PYTHIA 8 model, as well as the MPI and hadronization models. To this end, tunes: 2C, 4C and Monash 2013 were used for training, and the effects of MPI and hadronization were investigated using the Monte Carlo (MC) generator HERWIG 7.2 [17] for training. Figure 1 shows the correlation between the self normalized number of MPI ($N_{\rm mpi}/\langle N_{\rm mpi}\rangle$) and the self normalized mid-pseudorapidity charged particle multiplicity ($N/\langle N_{\rm ch}\rangle$) in pp collisions at $\sqrt{s}=5.02$, 7 and 13 TeV. For $N_{\rm ch}/\langle N_{\rm ch}\rangle<3$, the self normalized $N_{\rm mpi}$ increases linearly with the event multiplicity.

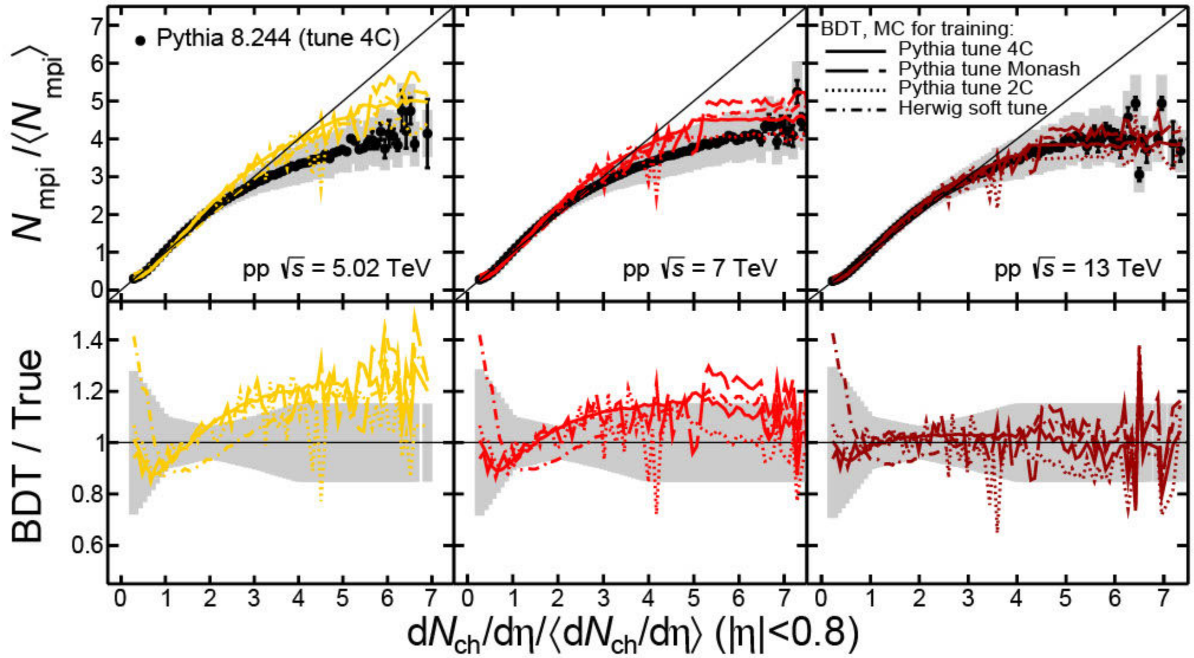

FIGURE 1. Monte Carlo closure test using pp collisions at $\sqrt{s} = 5.02$ (left), 7 (middle) and 13 TeV (right) simulated with PYTHIA 8 tune 4C. The top panels display the self normalized average number of MPI as a function of the self-normalized mid-pseudorapidity charged particle multiplicity. Ratios between ML results and the true values provided by PYTHIA are shown in the bottom panels.

On the other hand, for higher multiplicities, we observe a deviation of the self normalized $N_{\rm mpi}$ with respect to the linear trend. Figure 1 also displays the results obtained from regression (lines), and shows that one can recover the energy and multiplicity dependence using the ML-based regression.

Regarding the analysis using the available data, we built a toy MC in order to get the correlation between the event-by-event $\langle p_{\rm T}\rangle$ and $N_{\rm ch}$. The toy MC was constructed inside ROOT, and work as follows: for simplicity, each event class was simulated assuming its multiplicity spectrum as a Poisson distribution [18]. Their corresponding average multiplicity values as well as their contribution to the inelastic cross section were taken from [7, 12]. With this information, $N_{\rm ch}$ pseudo-particles were generated in each event, where each psuedo-particle had a transverse momentum which obeyed the $p_{\rm T}$ spectra reported by ALICE [7,12]. The information of all events generated with the toy MC was stored as a columnar dataset (TTree) using ROOT [19].

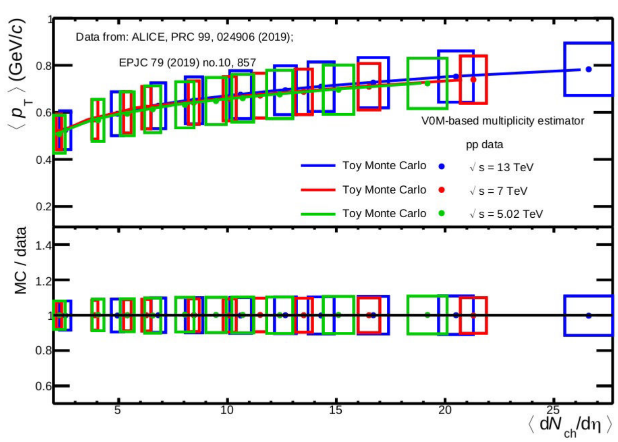

FIGURE 2. Mean transverse momentum as a function of the average charged-particle multiplicity density in pp collisions at $\sqrt{s} = 5.02$, 7 and 13 TeV. In the top panel, the ALICE data [7, 12] (solid markers) are compared with results from the toy MC (solid lines). The bottom panel displays the ratios between the toy MC and the data.

Figure 2 displays the mean transverse momentum as a function of the average charged-particle multiplicity density in pp collisions at $\sqrt{s} = 5.02$, 7 and 13 TeV. Within uncertainties, the toy MC reproduces the correlation between the $\langle p_{\rm T}\rangle$ and $\langle dN_{\rm ch}/d\eta\rangle$. In our approach, the event-by-event information produced by the toy MC was processed with the trained BDT in order to extract the MPI activity associated with the data.

3. Results

Using the ALICE data from pp collisions at $\sqrt{s} = 7$ TeV [7], we extract the average number of MPI, which is found to be $\langle N_{\rm mpi}\rangle = 3.98 \pm 1.01$. Figure 3 displays the average number of MPI as a function of the center-of-mass energy, for pp collision at $\sqrt{s} = 5.02$, 7 and 13 TeV. We obtain a regression value which is above unity, therefore, our results support the presence of MPI in pp collisions. We also observe a modest energy dependence, which is similar to that predicted by PYTHIA 8 [4].

Figure 4 displays the self normalized number of MPI ($N_{\rm mpi}/\langle N_{\rm mpi}\rangle$) as a function of the self-normalized mid-

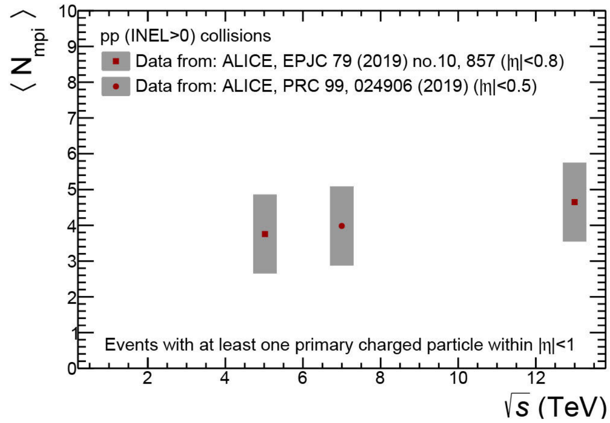

FIGURE 3. Average number of MPI as a function of the center-of-mass energy. Results for pp collisions at $\sqrt{s} = 7$ TeV, are compared to those for pp collisions at $\sqrt{s} = 5.02$ and 13 TeV reported in Ref. [4].

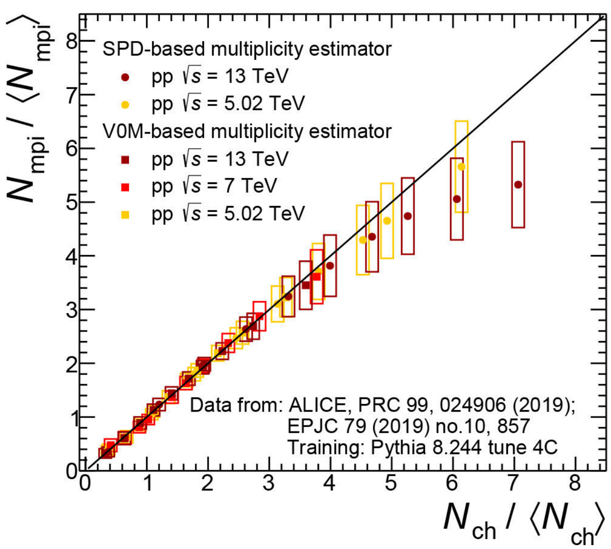

FIGURE 4. The self normalized average number of Multiparton Interactions as a function of the self normalized mid-pseudorapidity charged particle multiplicity is shown for pp collisions at $\sqrt{s} =$ 5.02, 7 and 13 TeV.

pseudorapidity charged-particle multiplicity ($N_{\rm ch}/\langle N_{\rm ch}\rangle$) in pp collisions at $\sqrt{s} = 5.02$, 7 and 13 TeV from ALICE data. We observe that for $N_{\rm ch} < 3\langle N_{\rm ch}\rangle$ the self normalized $N_{\rm mpi}$ increases linearly with the event multiplicity. Regarding higher multiplicities, we observe a deviation of the self normalized $N_{\rm mpi}$ with respect to the linear trend. This result qualitatively agrees with PYTHIA 8 (see Fig. 1).

Last but not least, we proposed to include more information in the BDT training. Using pp collisions data at $\sqrt{s} =$ 13 TeV generated with PYTHIA 8 Tune 4C we computed the event multiplicity in the forward-pseudorapidity regions, in order to determine if the event-by-event extraction of the $N_{\rm mpi}$ is improved. We trained three BDT sets which are described below:

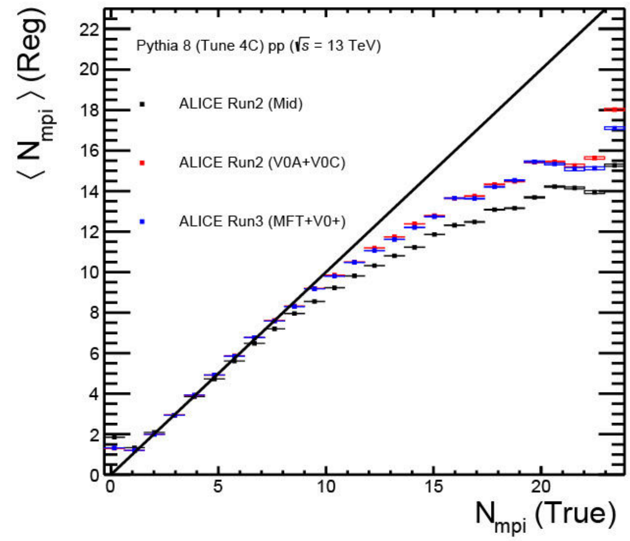

FIGURE 5. Average number of Multiparton Interactions ($\langle N_{\rm mpi}\rangle$ (Reg)) determined with the BDT as functions of the true value of Multiparton Interactions ($N_{\rm mpi}$ (True)) provided by PYTHIA simulations. Correlation displays when the multiplicity is computed in the mid-pseudorapidity range (black markers), as well as when is computed in the forward region defined by the V0A+V0C (red markers) and MFT+V0+ (blue markers) detectors.

- For the first set the multiplicity was computed in the $|\eta| < 0.8$ pseudorapidity range.
- For the second set the multiplicity was computed in the forward-pseudorapidity regions $2.8 < \eta < 2.5$ and $-3.6 < \eta < -1.7$, covered by the V0A and V0C arrays, respectively.
- Finally, for the third set the multiplicity was computed in the forward-pseudorapidity regions $-3.6 < \eta < -2.45$ and $2.2 < \eta < 5.1$, covered by the Muon Forward Tracker (MFT) and VZERO+ detectors respectively which are part of ALICE's Run3 upgrades planned from 2021 to 2023 [20].

Each BDT set was applied to pp collisions data at $\sqrt{s} =$ 13 TeV generated with PYTHIA 8 Tune 4C. Figure 5 displays the average number of Multiparton Interactions ($\langle N_{\rm mpi}\rangle$ (Reg)) determined with ML as a function of the number of Multiparton Interactions ($N_{\rm mpi}$ (True)) from pp collisions at $\sqrt{s} = 13$ TeV simulated with PYTHIA.

We observe that for $N_{\rm mpi}$ (True) ≈ 16, when the multiplicity is computed in the forward pseudorapidity range (ALICE Run2 V0A+V0C and ALICE Run3 MFT+V0+), the deviation of the $\langle N_{\rm mpi}\rangle$ (Reg) $/N_{\rm mpi}$ (True) ratio with respect to the case when the multiplicity is computed in the mid-pseudorapidity range (ALICE Run2 Mid), is around 13%. Therefore, including more information in the BDT training the extraction of the $N_{\rm mpi}$ is improved.

4. Conclusions

We report the extraction of the average number of MPI from pp data at the LHC energies using ML methods. We have found $\langle N_{\rm mpi}\rangle = 3.98 \pm 1.01$ for pp collisions at $\sqrt{s} = 7$ TeV. The comparisons with our previous results for pp collisions at $\sqrt{s} = 5.02$ and 13 TeV indicate a modest energy dependence of $N_{\rm mpi}$. This result provide experimental evidence of the presence of MPI in hadronic interactions. In addition, we also report the multiplicity dependence of $N_{\rm mpi}$ for the three center-of-mass energies. Finally, we have found that computing the event multiplicity in the forward-pseudorapidity regions the $N_{\rm mpi}$ extraction is improved, this result opens the possibility to extract the number of MPI event-by-event and in this way study the particle production as a function of MPI, this idea was explored on reference [4]. Our results are fully consistent with the so-called "mini-jet analysis" of ALICE [2], and suggest that high multiplicities (at mid-pseudorapidity) can only be reached by selecting events with many high-multiplicity jets.

Acknowledgments

Authors acknowledge Antonio Paz for providing the simulations with HERWIG 7.2. E. Z. acknowledges the fellowship of CONACyT.

1. B. Abelev *et al*., Transverse sphericity of primary charged particles in minimum bias proton-proton collisions at $\sqrt{s} = 0.9$, 2.76 and 7 TeV. *Eur. Phys. J. C*, **72** (2012) 2124. `https://doi.org/10.1140/epjc/s10052-012-2124-9`.
2. B. Abelev *et al*. Multiplicity dependence of two-particle azimuthal correlations in pp collisions at the LHC. *JHEP*, **09** (2013) 049, `https://doi.org/10.1007/JHEP09(2013)049`.
3. A. Ortiz, Experimental results on event shapes at hadron colliders. Adv. Ser. Direct. *High Energy Phys*., **29** (2018) 343, `https://www.worldscientific.com/doi/abs/10.1142/97898132277670016`.
4. A. Ortiz, A. Paz, J. D. Romo, S. Tripathy, E. A. Zepeda, and I. Bautista, Multiparton interactions in pp collisions from machine learning-based regression. *Phys. Rev. D*, **102** (2020) 076014, `https://link.aps.org/doi/10.1103/PhysRevD.102.076014`.
5. V. Khachatryan *et al.* Observation of Long-Range Near- Side Angular Correlations in Proton-Proton Collisions at the LHC. *JHEP*, **09** (2010) 091, `https://doi.org/10.1007/JHEP09(2010)091`.
6. J. Adam *et al*. Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions. *Nature Phys.*, **13** (2017) 535, `https://doi.org/10.1038/nphys4111`.
7. S. Acharya *et al*. Multiplicity dependence of light-flavor hadron production in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. C*, **99** (2019) 024906, `https://link.aps.org/doi/10.1103/PhysRevC.99.024906`.
8. P. Bożek, Collective flow in p-Pb and d-Pd collisions at TeV energies. *Phys. Rev. C* **85** (2012) 014911, `https://link.aps.org/doi/10.1103/PhysRevC.85.014911`.
9. J. L. Nagle and W. A. Zajc, Small System Collectivity in Relativistic Hadronic and Nuclear Collisions, *Ann. Rev. Nucl. Part. Sci.*, **68** (2018) 211, `https://doi.org/10.1146/annurev-nucl-101916-123209`.
10. A. Ortiz, P. Christiansen, E. Cuautle Flores, I. Maldonado Cervantes, and G. Paić, Color Reconnection and Flow-like Patterns in pp Collisions. *Phys. Rev. Lett.*, **111** (2013) 042001, `https://link.aps.org/doi/10.1103/PhysRevLett.111.042001`.
11. A. Ortiz and E.A. Zepeda, Extraction of the multiplicity dependence of multiparton interactions from LHC pp data using machine learning techniques. *J. Phys. G*, **48** (2021) 085014, `https://dx.doi.org/10.1088/1361-6471/abef1e`.
12. S. Acharya *et al*., Charged-particle production as a function of multiplicity and transverse spherocity in pp collisions at $\sqrt{s} = 5.02$ and 13 TeV. *Eur. Phys. J. C*, **79** (2019) 857, `https://dx.doi.org/10.1140/epjc/s10052-019-7350-y`.
13. H. Voss, Andreas Hocker, J. Stelzer, and F. Tegenfeldt, TMVA, the Toolkit for Multivariate Data Analysis with ROOT. *PoS*, **ACAT** (2007) 040, `https://doi.org/10.22323/1.050.0040`.
14. T. Sjöstrand *et al*., An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, **191** (2015) 159, `https://doi.org/10.1016/j.cpc.2015.01.024`.
15. R. Corke and T. Sjostrand, Interleaved Parton Showers and Tuning Prospects. *JHEP*, **03** (2011) 032, `https://doi.org/10.1007/JHEP03(2011)032`.
16. E. Cuautle, A. Ortiz, and G. Paic, Effects produced by multi-parton interactions and color reconnection in small systems. *Nucl. Phys. A*, **956** (2016) 749, `https://doi.org/10.1016/j.nuclphysa.2016.02.031`.
17. J. Bellm *et al*. Herwig 7.2 release note. *Eur. Phys. J. C*, **80** (2020) 452, `https://doi.org/10.1140/epjc/s10052-020-8011-x`.
18. A.I. Golokhvastov. Independent production and Poisson distribution. *Phys. Atom. Nucl.* **58** (1995) 1998.
19. I. Antcheva *et al*., ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization. *Comput. Phys. Commun.*, **180** (2009) 2499, `https://doi.org/10.1016/j.cpc.2009.08.005`.
20. W. H. Trzaska, New ALICE detectors for Run 3 and 4 at the CERN LHC, *Nucl. Instrum. Methods Phys. Res. A* **958** (2020) 162116, `https://doi.org/10.1016/j.nima.2019.04.070`.