# Quantum Chromodynamics Monte Carlo Tuning Studies in CMS

**Efe Yazgan**[*][†]

*Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, China*
*E-mail:* efe.yazgan@cern.ch

Recent QCD Monte Carlo tuning studies done in the CMS Collaboration are presented. Jet kinematics, jet substructure, and underlying event measurements in top quark pair events are discussed. New CMS PYTHIA 8 event tunes are presented, exploiting Monte Carlo configurations with consistent parton distribution functions and strong coupling parameter values in the matrix element and the parton shower, at leading order (LO), next-to-leading order (NLO) and next-to-next-to-leading order (NNLO). Predictions from PYTHIA 8 obtained with tunes based on NLO or NNLO PDFs are shown to reliably describe minimum-bias and underlying-event data with a similar level of agreement to predictions from tunes using LO PDF sets. The tunes are validated with a range of different measurements and matrix element-parton shower merged configurations.

---

[*]Speaker.

[†]On behalf of the CMS Collaboration

## 1. Introduction

Hadron collisions at particle colliders are modelled using quantum chromodynamics (QCD) Monte Carlo (MC) simulation codes. These codes consist of several parts. The hard scattering is modelled by simulating particles from the hadronization of partons kinematics of which are calculated using perturbation theory. Partons from initial-state radiation (ISR) and final-state radiation (FSR) are modelled by a parton shower (PS) algorithm. The underlying event (UE), a non-perturbative component, refers to the activity resulting from additional interactions of partons in the hadron collision. It consists of beam-beam remnants (BBR), and the particles from the additional relatively softer parton-parton scatterings referred to as the multiple-parton interactions (MPI). To obtain a reliable description of the UE, adjustable parameters in standard MC event generators, such as PYTHIA 8 [1], HERWIG [2], and SHERPA [3], need to be tuned to the data. Another non-perturbative ingredient is the parton distribution function (PDF) used both in the hard partonic matrix element (ME) calculation, the PS and the MPI model. $\alpha_S(M_Z)$ is a free parameter and its value used in simulations at LO or NLO are typically different. Different strategies are adopted in different collaborations; CMS [4] and ATLAS [5] tunes are traditionally based on LO PDFs, PYTHIA 8 [6] tunes are mostly based on LO PDFs, new tunes are based on NNLO PDFs, and HERWIG 7 [7] on NLO PDFs. Merging schemes, such as the *MLM* [8], allow the combination of predictions of jet production using ME calculations with those from PS emissions for soft and collinear parton radiation at leading-log (LL) accuracy without double counting or dead regions. Merged calculations capture some higher-order corrections with respect to the formal order of the ME calculation. Using the same PDF set and $\alpha_S$ value in the PS-ME merged calculations in the simulation of the various components is advocated in Ref. [9], and by the HERWIG 7 and SHERPA Collaborations. In the PYTHIA 8 tunes produced prior to Ref. [10], $\alpha_S(M_Z)$ values in general are not selected to be the same as those used in the PDFs. For example, in the Monash tune, the $\alpha_S^{FSR}(M_Z)$, set to 0.1365, is obtained by fitting standalone-PYTHIA 8 predictions to LEP event-shape measurements, $\alpha_S^{ISR}(M_Z)$ is assumed to be equal to $\alpha_S^{FSR}(M_Z)$, and the $\alpha_S(M_Z)$ values in hard scattering and MPI are set to 0.130 [6]. In this note, we discuss the preferred values of $\alpha_S^{ISR}(M_Z)$ and $\alpha_S^{FSR}(M_Z)$ obtained from jet multiplicity, substructure and UE event measurements in $t\bar{t}$ events and discuss new CMS [11] tunes based on consistent PDF and $\alpha_S(M_Z)$ values.

## 2. Revisiting Parton Shower Parameters and Tunes

The previous default CMS PYTHIA 8 tune, CUETP8M1 [4], does not describe the central values of $\sqrt{s} = 13$ TeV UE data very well [10] or the UE evolution from 1.96 to 7 TeV in UE in Z+jets [12]. It is also observed that the predictions of the POWHEG generator [13] interfaced with the PYTHIA 8 code with the CUETP8M1 tune overshoot the data for large multiplicities where jets are produced by the PS [14] for both $\sqrt{s} = 8$ and 13 TeV data.

### 2.1 Jet Multiplicity in $t\bar{t}$ events and $\alpha_S^{ISR}$

Tuning $\alpha_S^{ISR}$ using 8 TeV jet multiplicity in $t\bar{t}$ events using the PS dominated region yields $\alpha_S^{ISR} = 0.1108^{+0.0145}_{-0.0142}$ [14]. The tuned $\alpha_S^{ISR}(M_Z)$ value agrees with the PDG value of $\alpha_S^{ISR}(M_Z) =$

0.1181±0.0011 [15] well within uncertainties. It is found that to describe the number of jet distribution in t$\bar{\text{t}}$ events, the PYTHIA 8 ISR parameter *Rapidity Ordering* needs to be used [10]. Rapidity ordering imposes an additional constraint on the $p_{\text{T}}$ -ordered emissions reducing the parton emission phase space. The parton emission probability is mainly constrained by the jet activity and the interplay between the hard and soft parts of the parton emissions. However, it does not strongly constrain the UE. Therefore, $\alpha_S^{ISR}(M_Z)$ constrained by t$\bar{\text{t}}$ jet kinematics is used as a fixed input parameter in deriving a new UE tune, CUETP8M2T4 [14]. This new tune, with a lower $\alpha_S^{ISR}(M_Z)$, with respect to the CUETP8M1 tunes is found to improve the description of t$\bar{\text{t}}$ kinematics, as as the overall description of observables at $\sqrt{s} = 8$ and 13 TeV.

## 2.2 Jet substructure in t$\bar{\text{t}}$ events and $\alpha_S^{FSR}(M_Z)$

Jet substructure observables are measured using t$\bar{\text{t}}$ events in the lepton+jets channel at $\sqrt{s} = 13$ TeV [16] using the CUETP8M2T4 tune for the t$\bar{\text{t}}$ signal and CUETP8M1tune for other simulated samples. Jets are studied with different substructure observables using charged and neutral particles and charged-particle-only jet constituents. Inclusive jet samples as well as samples enriched in bottom, light-quark, or gluon jets are selected. Substructure observables at particle level are compared to NLO predictions from POWHEG +PYTHIA 8, HERWIG 7, SHERPA and DIRE 2 [17]. Predictions with the default tunes do not yield a good overall description of the data, in particular, for particle multiplicity-related observables. These measurements help tune models. As a first step in this direction, the angle between the groomed subjets is used to extract $\alpha_S^{FSR} = 0.115^{+0.015}_{-0.013}$ at LO+LL accuracy and using CMW rescaling [18]. This value agrees with the PDG value of $\alpha_S^{ISR}(M_Z)$ well within uncertainties. A more precise $\alpha_S^{FSR}(M_Z)$ could be obtained once calculations of top-quark decays with multiple emissions at higher order with PS at approximate NLL become available. These measurements also provide the data to improve analytical QCD calculations with higher-order corrections, for infrared-and/or collinear-safe observables.

## 2.3 Underlying event in t$\bar{\text{t}}$ events and $\alpha_S^{FSR}(M_Z)$

UE activity in t$\bar{\text{t}}$ dilepton events are measured by CMS at $\sqrt{s} = 13$ TeV. The measurements are based in particle-flow reconstruction. The UE contribution is isolated by removing charged particles associated with the decay products of the t$\bar{\text{t}}$ event candidates as well as with pileup interactions from the set of reconstructed charged particles for each event. The observables and categories chosen for the measurements enhance the sensitivity to t$\bar{\text{t}}$ modeling, MPI, color reconnection (CR) and the choice of $\alpha_S(M_Z)$ in the PYTHIA 8.

Most of the comparisons indicate a fair agreement between the data and the POWHEG + PYTHIA 8 setup with the CUETP8M2T4tune [14], but disfavor the setups in which MPI and CR are switched off. The data also disfavor the default configurations in POWHEG +HERWIG ++, POWHEG +HERWIG 7, and SHERPA . It has been furthermore verified that, as expected, the choice of the next-to-leading-order matrix-element generator does not impact significantly the expected characteristics of the UE by comparing predictions from POWHEG and MG5_aMC [19], both interfaced with PYTHIA 8.

The UE measurements in t$\bar{\text{t}}$ events test the hypothesis of universality of UE at an energy scale two times the top quark mass which is considerably higher than the ones at which UE models have

been studied in detail. The results also show that a value of $\alpha_S^{FSR}(M_Z) = 0.120 \pm 0.006$ is consistent with the data and the corresponding uncertainties translate to a variation of the renormalization scale by a factor of $\sqrt{2}$ that has already been adopted in CMS $t\bar{t}$ measurements.

## 2.4 New CMS UE tunes

A set of new 13 TeV PYTHIA 8.226 UE tunes is obtained with different choices of values of $\alpha_S(M_Z)$ used in the modeling of the ISR, FSR, hard scattering, and MPI, as well as the order of its evolution as a function of the four-momentum squared $Q^2$. We distinguish the new tunes according to the order of the NNPDF3.1 PDF set [20] used: LO-PDF, NLO-PDF, or NNLO-PDF. The tunes are labeled as CPi, where i=1,2,...,5. CP1 and CP2 are based on LO, CP3 on NLO, and CP4 and CP5 on NNLO NNPDF3.1 set. We fit charged-particle and $p_T^{\ sum}$ densities, measured in transMIN and transMAX regions as a function of $p_T^{\ max}$, as well as the charged-particle multiplicity as a function of pseudorapidity $\eta$, measured by CMS at $\sqrt{s} = 13$ TeV [21, 22]. In addition, we also use the charged-particle and $p_T^{\ sum}$ densities as a function of the leading charged-particle $p_T$, measured in transMIN and transMAX by CMS at $\sqrt{s} = 7$ TeV [23] and by CDF at $\sqrt{s} = 1.96$ TeV [24].

The parameters related to the simulation of the hadronization and beam remnants are not varied in the fits and are kept fixed to the values of the Monash tune. Only five parameters related to the simulation of MPI, to the overlap matter distribution function [25], and to the amount of CR are constrained for the new CMS tunes. In all tunes, we use the MPI-based CR model [26]. The overlap distribution between the two colliding protons is modeled according to a double-Gaussian functional form with the parameters `coreRadius` and `coreFraction`.

In the LO NNPDF3.1 set, $\alpha_S(M_Z) = 0.130$, whereas for the NLO and NNLO NNPDF3.1 sets, $\alpha_S(M_Z) = 0.118$. Irrespective of the specific PDF used, predictions from the new tunes reproduce well the UE measurements at center-of-mass energies $\sqrt{s} = 1.96$ and 7 TeV. A significant improvement in the description of UE measurements at 13 TeV is observed with respect to predictions from the previous tunes that were extracted using data at lower collision energies. For the first time, predictions based on higher-order PDF sets are shown to give a reliable description of minimum-bias (MB) and UE measurements, with a similar level of agreement as predictions from tunes using LO PDF sets. Predictions of the new tunes agree well with the data for MB observables measured in the central ($|\eta| < 2.4$) and forward ($3.2 < |\eta| < 4.7$) regions. The CP tunes simultaneously describe the number of charged particles produced in diffractive processes and MB collisions. Neither the CP tunes nor the CUETP8M1 tune describe the very forward region ($-6.6 < \eta < -5.2$) well. Measurements sensitive to double-parton scattering contributions are reproduced better by predictions using the LO PDF set in the UE simulation, without ISR rapidity ordering.

The UE simulation provided by the new tunes can be interfaced to higher-order and multileg matrix element generators, such as POWHEG and MG5_aMC, without degrading the good description of UE observables. Such predictions also reproduce well observables measured in multijet final states, Drell–Yan, and top quark production processes. The central values of the normalized $t\bar{t}$ cross section in bins of the number of additional jets predicted by POWHEG +PYTHIA 8 overestimate the data when a high value of $\alpha_S^{ISR}(M_Z) \sim 0.13$ is used (CMS PYTHIA 8 CP1 and CP2 tunes). Even when $\alpha_S^{ISR}(M_Z) = 0.118$ is used, the CP4 tune overestimates the data at high jet multiplicities. This is cured by the ISR rapidity ordering in the CP5 tune (as in the CUETP8M2T4 tune). Measurements of azimuthal dijet correlations are also better described when a value of $\alpha_S^{ISR}(M_Z) = 0.118$ is used in

predictions obtained with POWHEG merged with PYTHIA 8. All details about the CP tunes can be found in [10].

## 3. Conclusions

Monte Carlo event modeling is studied using jet multiplicity, jet substructure, and underlying event measurements in top quark pair events. These studies showed that the data is consistent with a lower value of strong coupling parameter for both initial and final state radiation in the parton shower when merged configurations are used. New CMS PYTHIA 8 multipurpose tunes, aiming for a consistent description of UE and MB observables at several collision energies and a reliable prediction of the UE simulation in various processes when merged with higher-order ME calculations are presented. These new tunes exploit Monte Carlo configurations with consistent parton distribution functions and strong coupling parameter values in the matrix element and the parton shower, at leading order (LO), next-to-leading order (NLO) and next-to-next-to-leading order (NNLO). Predictions from PYTHIA 8 obtained with tunes based on NLO or NNLO PDFs are shown to reliably describe minimum-bias and underlying-event data with a similar level of agreement to predictions from tunes using LO PDF sets along with a wide range of different measurements.

## References

[1] T. Sjöstrand, S. Ask, J. R., Christiansen, et al. , *An introduction to Pythia 8.2*, Comput. Phys. Commun. **191** (2015) 159 [arXiv:1410.3012]

[2] J. Bellm, et al., *Herwig 7.0/Herwig++ 3.0 release note* Eur. Phys. J. C **76** (2016) 196 [arXiv:1512.01178]

[3] T. Gleisberg, S. Hoeche, F. Krauss, et. al., *Event generation with SHERPA 1.1* J. High Ene. Phys. **02** (2009) 7 [arXiv:0811.4622]

[4] V. Khachatryan, et al. [CMS Collaboration], *Event generator tunes obtained from underlying event and multiparton scattering measurements* Eur. Phys. J. C **76** (2016) 155 [arXiv:1512.00815]

[5] ATLAS Collaboration, *ATLAS PYTHIA 8 tunes to 7 TeV data* ATL-PHYS-PUB-2014-021 (2015)

[6] P. Skands, S. Carrazza, and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 tune* Eur. Phys. J. C **74** (2014) 3024 [arXiv:1404.5630]

[7] S. Gieseke, F. Loshaj, and P. Kirchgaeber, *Soft and diffractive scattering with the cluster model in Herwig* Eur. Phys. J. C **77** (2017) 156 [arXiv:1612.04701]

[8] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions* **53** (2008) 473 [arXiv:0706.2569]

[9] B. Cooper, et al. *Importance of a consistent choice of $\alpha_S$ in the matching of AlpGen and PYTHIA* Eur. Phys. J. C **72** (2012) 2078 [arXiv:1109.5295]

[10] A. M. Sirunyan, et al. [CMS Collaboration], *Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements* CMS-GEN-17-001, CERN-EP-2019-007 (2019) [arXiv:1903.12179]

[11] CMS Collaboration, Jour. of Inst. **3** (2008)S08004

[12] A. M. Sirunyan, at al. [CMS Collaboration], *Measurement of the underlying event activity in inclusive Z boson production in proton-proton collisions at $\sqrt{s} =$ 13 TeV* J. High Ene. Phys. **07** (2018) 032 [arXiv:1711.04299]

[13] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX* J. High Ene. Phys. **06** (2010) 43 [arXiv:1002.2581]

[14] CMS Collaboration, *Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of t$\bar{t}$ at $\sqrt{s} =$ 8 and 13 TeV* CMS-PAS-TOP-16-021 (2016)

[15] M. Tanabashi, et al. [Particle Data Group], *Review of particle physics* Phys. Rev. D **98** (2018) 30001

[16] A. M. Sirunyan, at al. [CMS Collaboration], *Measurement of jet substructure observables in t$\bar{t}$ events from proton-proton collisions at $\sqrt{s} =$ 13 TeV* Phys. Rev. D **98** (2018) 092014 [arXiv:1808.07340]

[17] S. Hoche, and S. Prestel, *The midpoint between dipole and parton showers* Eur. Phys. J. C **75** (2015) 461 [arXiv:1506.05057]

[18] S. Catani, B. Webber, and G. Marchesini, *QCD coherent branching and semiinclusive processes at large x* Nucl. Phys. B **349** 635 (1991)

[19] 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* J. High Ene. Phys. **07** (2014) 79 [arXiv:1405.0301]

[20] R. D. Ball, et al., NNPDF Collaboration *Parton distributions from high-precision collider data* Eur. Phys. J. C **77** (2017) 603 [arXiv:1706.00428]

[21] CMS Collaboration, *Underlying event measurements with leading particles and jets in pp collisions at $\sqrt{s} =$ 13 TeV*, CMS Physics Analysis Summary (2015) CMS-PAS-FSQ-15-007

[22] CMS Collaboration, *Pseudorapidity distribution of charged hadrons in proton-proton collisions at $\sqrt{s}$ = 13 TeV* Phys. Lett. B **751** 143 [arXiv:1507.05915]

[23] CMS Collaboration, *Measurement of the underlying event activity in pp collisions at the LHC at 7 TeV and comparison with 0.9 TeV*, CMS Physics Analysis Summary (2012) CMS-PAS-FSQ-12-020

[24] CDF Collaboration, *Study of the energy dependence of the underlying event in proton-antiproton collisions* **92** (2015) 092009 [arXiv:1508.05340]

[25] T. Sjöstrand, and M. van Zijl, *A multiple interaction model for the event structure in hadron collisions* Phys. Rev. D **36** (1987) 2019

[26] T. Sjöstrand, and P. Z. Skands, *Multiple interactions and the structure of beam remnants* J. High Ene. Phys. **03** (2004) arXiv:0402078

PoS(DIS2019)080