Monte Carlo modeling and tuning in CMS

Gerrit Van Onsem on behalf of the CMS Collaboration[a,*]

[a]*CERN,*
Esplanade des Particules 1, Meyrin, Switzerland

E-mail: Gerrit.Van.Onsem@cern.ch

The underlying event in a proton-proton collision consists of multiple-parton interactions and beam remnants, which typically involve low energy scales. The modeling of these aspects of a collision is not always calculable in perturbative quantum chromodynamics, and is governed by phenomenological parameters that need to be "tuned" to observed data. The results of such an underlying-event tuning performed by the CMS experiment for the Pythia 8 and Herwig 7 generators are presented. Various assumptions in the proton parton distribution functions and parton shower development are tested. The derived tunes are validated using data of a wide range of physics processes, and provide a good description of CMS data at a center-of-mass energy $\sqrt{s} = 13$ TeV.

PoS(ICHEP2020)515

40th International Conference on High Energy physics - ICHEP2020
July 28 - August 6, 2020
Prague, Czech Republic (virtual meeting)

---

[*]Speaker

1. Introduction

Monte Carlo (MC) generators are used to simulate a multitude of different aspects of a proton-proton collision. These include the hard-scattering matrix element (ME), the parton showers and radiation, the hadronization of the partons, and the subsequent hadron decays. Additional interactions may take place in a collision, referred to as multiple-parton interactions (MPI). Furthermore, the remnants of the protons also still undergo interactions and hadronization. The combination of MPI and beam remnants is typically referred to as the underlying event (UE). In the simulation of the color exchange between partons in the event, the color fields may be rearranged – a mechanism known as color reconnection. In the CMS experiment [1] at the CERN LHC, commonly used MC generator programs include POWHEG [2] and MADGRAPH5_aMC@NLO [3] for ME calculations, and the PYTHIA 8 [4] and HERWIG 7 [5] generators to interface parton shower and UE simulations to the matrix elements. The UE, the parton shower development, and the hadronization typically involve lower energy scales, and their modeling is governed by phenomenological parameters that may be tuned to observed data. The CMS experiment performed such a tuning for the PYTHIA 8 generator (v. 8.226) [6] and the HERWIG 7 generator (v. 7.14) [7][1], for a set of MPI and color reconnection parameters.

2. Tuning procedure

In the MPI modeling, we consider a transverse momentum (p_T) threshold parameter $p_{T,0}$ that governs the transition between soft and hard interactions. The lower this threshold, the more MPI are generated, which results in more UE activity. This parameter has an exponential dependence on the center-of-mass energy $\sqrt{s}$, with the exponent a tunable parameter. Parameters related to the overlap distribution between the two colliding protons are considered as well, where a larger or denser overlap corresponds to more MPI. Finally, we consider a parameter related to the color reconnection probability. A large value of this parameter increases the amount of color reconnection and tends to reduce final particle multiplicities. The assumed NNPDF3.1 parton distribution functions (PDFs) and the strong coupling α_S also influence these parts of the simulation, and therefore affect the predictions. Technically the tuning is performed by fitting binned predictions of observables to data using the RIVET [9] and PROFESSOR [10] frameworks. More specifically, a χ^2 function is minimized with respect to the set of tuning parameters, using the measured and observed bin contents.

The tunes are extracted from relevant sensitive observables measured in minimum-bias (MB) data. The observables include the charged-particle multiplicity N_{ch} and scalar p_T sum densities as a function of the p_T of the leading track or charged-particle jet. These are measured in transverse regions. For instance, a transMax (transMin) region is defined according to where the most (least) charged particles or highest (lowest) scalar p_T sum is located. Another observable used in the tuning is N_{ch} as a function of the pseudorapidity η. These data are recorded at various $\sqrt{s}$ by the CMS and CDF detectors. Components of the hard scattering such as hadronization and initial- and final-state radiation (ISR and FSR) may affect these UE observables. For instance, the transMax region is more sensitive to ISR and FSR, and the transMin region more to MPI and beam remnants.

---

[1]The corresponding paper preprint [8] of the HERWIG 7 tuning performed by the CMS experiment has been submitted to the journal after the ICHEP conference.

Predictions for the distributions of the observables are obtained by simulating non-diffractive and diffractive inelastic events for each choice of tune parameters in a multidimensional scan.

3. Results of the Pythia 8 tuning

Results are derived for five CMS Pythia 8 tunes, labeled as CP1 to CP5. In these tunes, different assumptions are made for the order of the PDF set, and the $\alpha_S(m_Z)$ value and running. As we typically use higher-order PDFs in the next-to-leading-order (NLO) matrix elements, the usage of higher-order PDFs in the parton shower and UE may be motivated. The different assumptions in the various CP tunes allow us to check the consistency of matching the order of PDFs in the ME and the tunes. The CP1 and CP2 tunes assume a leading-order (LO) PDF; CP3 an NLO PDF; and CP4 and CP5 a next-to-next-to-leading-order (NNLO) PDF, where CP5 differs from CP4 by enabling a rapidity ordering of ISR. To be consistent with the NLO calculation of α_S, the $\alpha_S(m_Z)$ value in these higher-order PDF based tunes (CP3 to CP5) is chosen lower ($\alpha_S(m_Z) = 0.118$ across all parts of the simulation) than the leading-order tunes (CP1 assumes $\alpha_S(m_Z) = \alpha_S^{\mathrm{MPI}}(m_Z) = 0.130$ and $\alpha_S^{\mathrm{ISR/FSR}}(m_Z) = 0.1365$, and CP2 assumes $\alpha_S(m_Z) = \alpha_S^{\mathrm{ISR/FSR}}(m_Z) = \alpha_S^{\mathrm{MPI}}(m_Z) = 0.130$). The extracted values of the tuned parameters differ most significantly between LO and higher-order PDF based tunes, in the values of $p_{\mathrm{T},0}$ and its energy dependence, as well as the amount of color reconnection. These differences may be explained by the different shapes of gluon densitites at small x in the assumed PDFs, and by the different assumed α_S running.

Figure 1 shows example distributions of the N_{ch} density observable in the transMin region used in the Pythia 8 tuning, comparing the CMS data to the predictions from LO PDF based tunes (left) and from higher-order PDF based tunes (right). The rising part of the spectrum, where diffraction processes become more important, is not always well described, but this low-energy region below 3 GeV was not included in the tuning. All CP tunes describe the MB and UE observables that we studied similarly well.
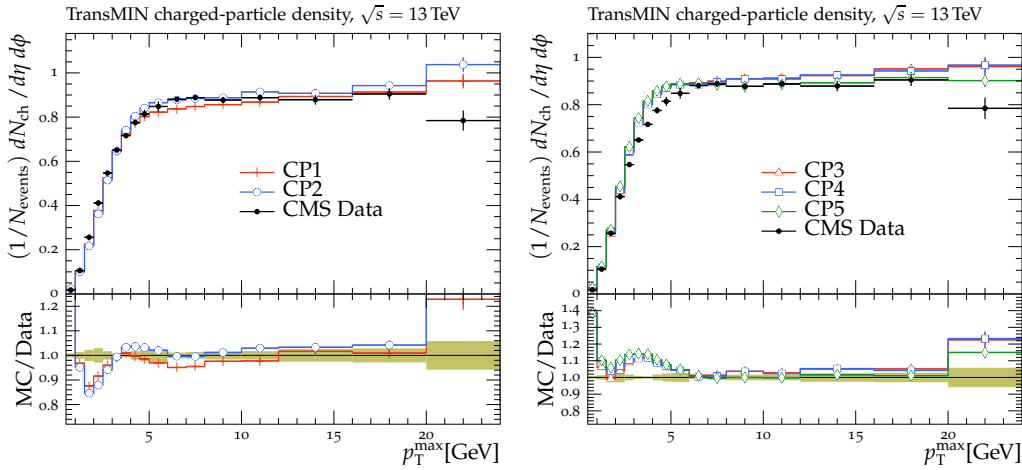

Figure 1: The CMS data at $\sqrt{s} = 13$ TeV on the normalized N_{ch} density distributions in the transMin region, as a function of the p_T of the leading charged particle. The data are compared to the predictions from the CMS Pythia 8 tunes CP1 and CP2 (left) and CP3, CP4, and CP5 (right) [6].

3

Many validations of the PYTHIA 8 tunes are performed, as ideally, an UE tune is universal and describes a wide range of processes and observables. Event shape observables measured at the LEP collider are of interest since, due the leptonic initial state, these observables are particularly sensitive to the α_S value assumed in FSR. The thrust observable T is shown in Fig. 2 (left), where higher values of T indicate the event is more dijet-like, and lower values more isotropic. We observe that the tunes assuming a lower value of α_S, resulting in less isotropic events, describe the data better. The angular separation between the two leading jets in multijet events recorded by the CMS experiment is checked as well, where the predictions are obtained from NLO dijet MEs from POWHEG merged with PYTHIA 8. We observe that tunes based on NLO α_S running are better than tunes with LO running, since the former result in a lower amount of FSR and thus less jet decorrelation. The tune performance is also checked using $t\bar{t}$ events, as shown in Fig. 2 (right) for the number of additional jets in the event, where the predictions are obtained from interfacing NLO MEs from powheg with PYTHIA 8. We find that an NLO α_S value and ISR rapidity ordering switched on, as assumed in the CP5 tune, is favored in these calculations.
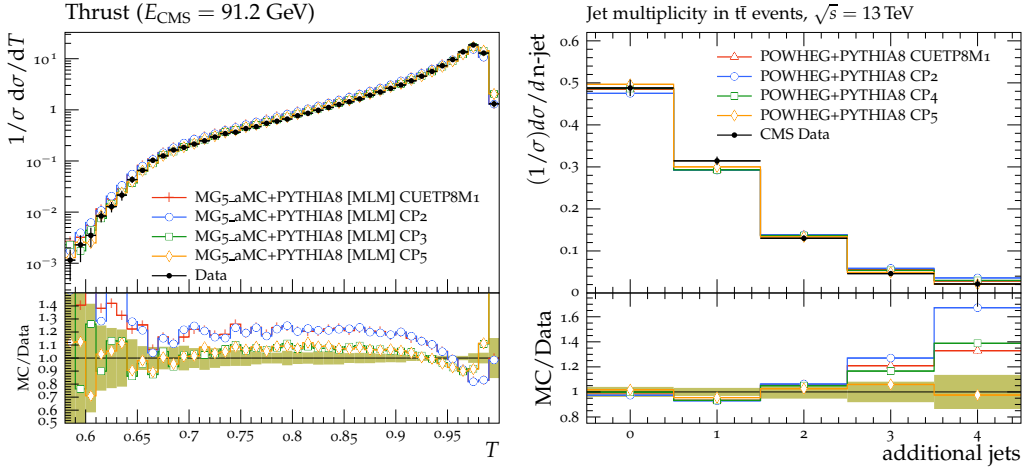

Figure 2: Validations of the CMS PYTHIA 8 tunes for the thrust observable T in $e^+e^- \rightarrow q\bar{q}$ events at the LEP collider (left), and for the additional jet multiplicity in $t\bar{t}$ events recorded by the CMS detector [6].

4. Results of the Herwig 7 tuning

In the CMS HERWIG 7 tuning, different assumptions are made in the UE only, for the order of the PDF set and the $\alpha_S(m_Z)$ value and running. The CH1 tune assumes $\alpha_S(m_Z) = 0.118$ and an NNLO PDF in the UE, which is changed to an LO PDF in the CH2 tune. The CH3 tune has the same assumptions as CH2 except for a higher value of $\alpha_S(m_Z) = 0.130$ in the UE, more compatible with the assumed LO PDF. In all of the CH tunes, an NNLO PDF in the parton shower is assumed, consistent with the PDF we typically use in ME calculations. The CH tunes are compared to a default tune, SoftTune, extracted by the HERWIG 7 authors using ATLAS data, in which an LO PDF in both the UE and parton shower is assumed. We observe that the CH tunes have a lower $p_{T,0}$ and energy exponent compared to SoftTune, which generally leads to an increased amount of MPI. From the overall χ^2 values corresponding to the tunes, we deduce that an LO PDF for the UE, as assumed in CH2 and CH3, is preferred, while the choice of α_S in the PDF seems less important.

Figure 3 shows example observables used in the HERWIG 7 tuning: the charged-hadron multiplicity as a function of η (left), where an increased amount of MPI in the CH tunes is apparent compared to the SoftTune, and the charged-particle p_T sum density in the transMin region (right). The CH tunes show typically good agreement in the plateau of these p_T sum density spectra, but discrepancies appear in the rising part of the spectrum, which was not included in the tuning.
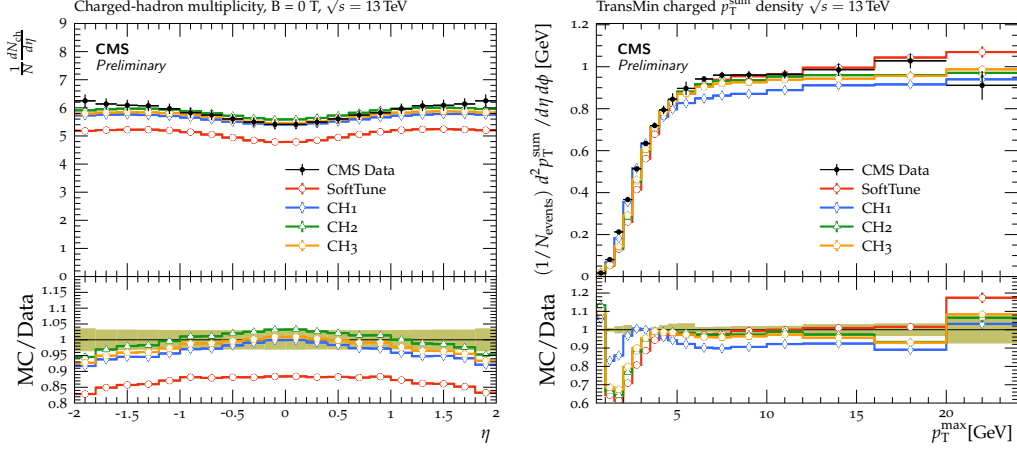

Figure 3: The CMS data at $\sqrt{s} = 13$ TeV on the normalized η distribution of charged hadrons (left), and on the normalized charged-particle p_T sum density distribution in the transMin region, as a function of the p_T of the leading track (right). The data are compared to predictions from HERWIG 7 with the SoftTune and CH tunes [7].

Various validations in different processes and observables are performed. For instance, kinematic properties in $t\bar{t}$ events are generally well described by the NLO ME calculation from POWHEG interfaced with HERWIG 7. The CH tunes do tend to underestimate the high jet multiplicity, as shown in Fig. 4 (left), but shower scale uncertainties are expected to be higher in this region. In vector boson production, where predictions are obtained from NLO ME calculations using MAD-GRAPH5_aMC@NLO interfaced with HERWIG 7, we observe a similarly good description by all CH tunes. Figure 4 (right) shows the charged-particle p_T sum density as a function of the p_T of the dimuon system in Z+jets production. This observable is sensitive to the UE description at higher momentum scales than MB data, and we observe that the CH tunes describe the data well.

5. Summary

New sets of tunes for the underlying-event simulation in proton-proton collisions at $\sqrt{s} = 13$ TeV are extracted and validated by the CMS experiment. The new CMS PYTHIA 8 tunes improve on older tunes extracted at lower $\sqrt{s}$. For the first time, predictions based on higher-order PDF sets are shown to give a reliable description. The first HERWIG 7 tunes with CMS data have also been derived, and these tunes are shown to describe the data generally well. Overall the underlying-event simulation from the new tunes interfaced with higher-order and multileg matrix elements provide a good description, as demonstrated using a wide range of processes. The extracted tunes, in particular the CP5 tune (PYTHIA 8) and the CH3 tune (HERWIG 7), are now widely used in the simulations in the $\sqrt{s} = 13$ TeV measurements and searches in the CMS experiment.
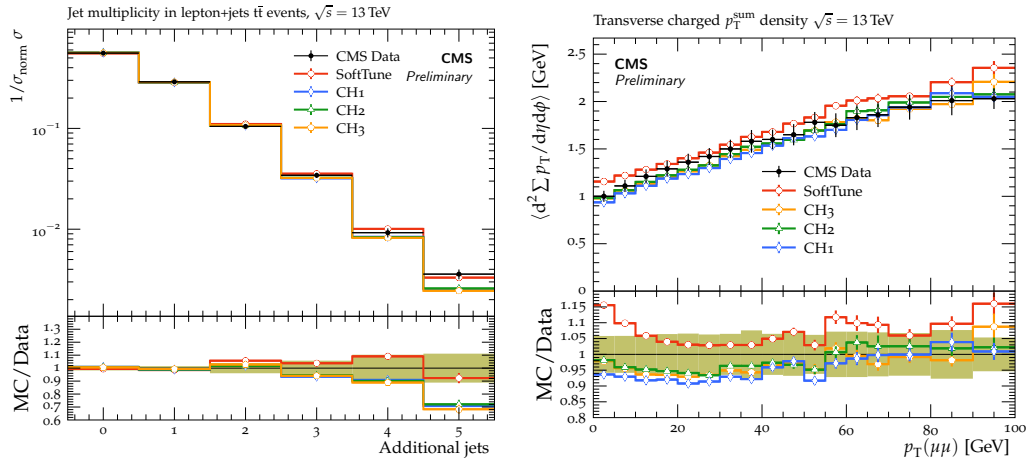
Figure 4: Validations of the CMS Herwig 7 tunes for the additional jet multiplicity in $t\bar{t}$ events (left), and for the charged-particle p_T sum density as a function of the p_T of the dimuon system in Z+jets events (right) [7].

References

[1] CMS collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[2] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043 [1002.2581].

[3] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer 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** (2014) 079 [1405.0301].

[4] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [1410.3012].

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

[6] CMS collaboration, *Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements*, *Eur. Phys. J. C* **80** (2020) 4 [1903.12179].

[7] CMS collaboration, *Extraction and validation of a set of HERWIG 7 tunes from CMS underlying-event measurements*, Tech. Rep. CMS-PAS-GEN-19-001, CERN, Geneva (2020).

[8] CMS collaboration, *Development and validation of HERWIG 7 tunes from CMS underlying-event measurements*, *Submitted to Eur. Phys. J. C* (2020) [2011.03422].

[9] A. Buckley, J. Butterworth, L. Lonnblad, D. Grellscheid, H. Hoeth, J. Monk et al., *Rivet user manual*, *Comput. Phys. Commun.* **184** (2013) 2803 [1003.0694].

[10] A. Buckley, H. Hoeth, H. Lacker, H. Schulz and J.E. von Seggern, *Systematic event generator tuning for the LHC*, *Eur. Phys. J. C* **65** (2010) 331 [0907.2973].