IL NUOVO CIMENTO **44 C** (2021) 79
DOI 10.1393/ncc/i2021-21079-4

Communications: SIF Congress 2020

Monte Carlo tuning in CMS

G. Sorrentino
Università di Trieste and INFN - Trieste, Italy

received 14 January 2021

Summary. — In this work, new sets of CMS underlying-event parameters for the PYTHIA8 event generator ("tunes") will be presented. The tunes extraction techniques will be described, together with the key observables and the relative measurements that have been performed to obtain the different parameter sets. The tunes validation will be also shown, using different physics scenarios, comparing the obtained predictions with the CMS experimental results.

1. – Introduction

Monte Carlo (MC) simulation is used to describe hadron-hadron collisions, according to models that are based on two main components: the hard scattering and the underlying event (UE). The first component consists of particles whose kinematics is predicted using perturbative matrix elements (MEs), together with partons from initial-state radiation (ISR) and final-state radiation (FSR), simulated with dedicated showering algorithms. The second component consists of beam-beam remnants (BBR) and particles from soft multiple-parton interactions (MPI). A tune defines a set of adjustable parameters that controls the behavior of the event modeling in standard MC event generators.

2. – Extraction of the new CMS PYTHIA8 tunes

2˙1. *Observables sensitive to underlying event.* – A new set of tunes has been extracted only for the UE simulation of PYTHIA8 [1] by fitting the charged-particle multiplicity and the charged-particle scalar-p_T sum densities. Two geometrical UE-sensitive regions are defined in the η-ϕ space, according to the azimuthal separation between the charged particles and the leading object, $\Delta\phi = \phi - \phi_{max}$. The regions are labelled as "toward" ($|\Delta\phi| \leq 60°$), "away" ($|\Delta\phi| > 120°$), and "transverse" ($60° < |\Delta\phi| \leq 120°$). The leading object is defined as the charged particle with the largest p_T, ϕ_{max} is the azimuth of the leading object and ϕ is the azimuth of an outgoing charged particle. The tunes have been extracted in transMAX and transMIN regions, the two transverse regions having

Table I. – *Parameters in the PYTHIA8 MC event generator together with the parameter ranges used for the fits* [2].

Parameter description	Name in PYTHIA8	Range
MPI threshold [GeV]	`MultipartonInteractions:pT0Ref`	1.0–3.0
Exponent of $\sqrt{s}$ dependance, ϵ	`MultipartonInteractions:ecmPow`	0.0–0.3
Matter fraction contained in the core	`MultipartonInteractions:coreFraction`	0.1–0.95
Radius of the core	`MultipartonInteractions:coreRadius`	0.1–0.8
Range of color reconnection probability	`ColorReconnection:range`	1.0–9.0

the maximum and minimum of either the number of charged particles, or charged-particle scalar-p_T sum densities (p_T^{sum}).

2˙2. *New CMS PYTHIA8 (CP) tunes at 13 TeV*. – The new tunes are distinguished according to the order of the parton distribution functions (PDF) set used. The CP1 tune uses the NNPDF3.1 PDF [3] set at leading order (LO). The strong coupling constant (α_S) values used for simulation of MPI, hard scattering, FSR, and ISR are, respectively, 0.13, 0.13, 0.1365, and 0.1365. The CP2 tune is the same as CP1 but with α_S values all equal to 0.13. The CP3 tune uses the NNPDF3.1 PDF set at next-to-leading order (NLO), with α_S all equal to 0.118, while the CP4 has the same values of α_S but NNPDF3.1 PDF set at next-to-next-to-leading order. The last tune, CP5, differs from CP4 only for the ISR ordering. For the extraction of the new tunes, only five parameters are constrained (table I), while those related to the hadronization and BBR are kept fixed to the values of the Monash 2013 tune [4]. The first threshold parameter is needed for the regularization of the divergence of the cross section at low p_T. The ϵ parameter is the exponent of the power law function that parameterizes the MPI energy dependence. The third and fourth parameters model the overlap distribution between the two colliding protons, described with a double-Gaussian functional form. This modeling allows to identify an inner, denser part, called core, and an outer less dense part. The `coreRadius` and `coreFraction` parameters represent, respectively, the width of the core and the fraction of quark and gluon content enclosed in it. The last parameter determines the amount of simulated color reconnection (CR): small (large) values tend to increase (reduce) the final particle multiplicities. The tunes are extracted by varying the above parameters and generating different sets of predictions, using the RIVET [5] and the PROFESSOR [6] frameworks. The obtained predictions are then fitted, minimizing the following χ^2 function:

$$\chi^2 = \sum_{O_j} \sum_i \frac{(f_{i,O_j}(p) - R_{i,O_j})^2}{\Delta^2_{i,O_j}}. \tag{1}$$

The sum is performed over all the i bins for each observable O_j. The $f_{i,O_j}(p)$ functions parameterize the dependence of the predictions in bin i on the tuning parameters, R_{i,O_j} is the value of the measured observable in such bin, and Δ^2_{i,O_j} is the total experimental uncertainty on R_{i,O_j}. In fig. 1 the TransMAX charged p_T^{sum} densities are shown, as a function of the transverse momentum of the leading charged particle (p_T^{max}). Predictions from LO tunes are slightly better than higher-order tunes in describing the energy de-

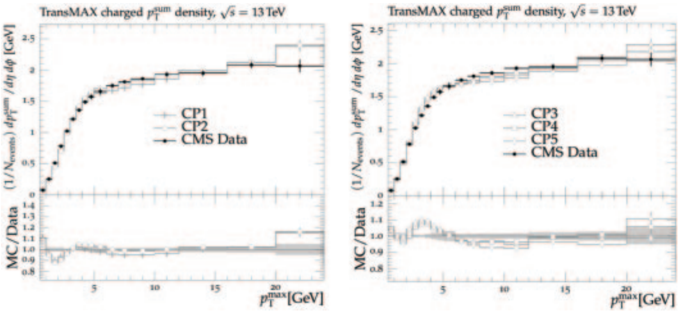

Fig. 1. – TransMAX p_T^{sum} densities, from the $\sqrt{s} = 13$ TeV analysis [7]. The data are compared with the CP1 and CP2 (left) and with CP3, CP4, CP5 (right) tunes. The MC/Data ratios are also shown, where the shaded bands represent the total experimental uncertainty in the data. Vertical lines on the data points refer to the total uncertainty in the data. Vertical lines drawn on the MC points refer to the statistical uncertainty in the predictions.

pendence of the considered UE measurements. For p_T^{max} values smaller than 3 GeV the predictions do not always reproduce the measurements, and exhibit discrepancies up to 20%, due to contributions from diffractive processes.

2˙3. *Tunes validation.* – Validation is performed comparing the predictions obtained with the new tunes with various experimental measurements performed by CMS.

- Top quark production [7]. In fig. 2 the $\mathrm{t\bar{t}}$ invariant mass (left) and the jet multiplicity (middle) distributions are shown. In the left plot, tune predictions under 1200 GeV are quite similar, while above 1200 GeV the CP4 tune provides the best agreement with the data. Looking at the jet multiplicity, a global overestimation of the data is observed increasing the number of jets, for all the tunes except CP5.
- Double parton scattering [8]. In fig. 2 (right) the $\Delta S = \arccos\left(\frac{\vec{p}_{T,1}\cdot\vec{p}_{T,2}}{|\vec{p}_{T,1}||\vec{p}_{T,1}|}\right)$ distribution is shown for $\mathrm{pp} \to 2\mathrm{b} + 2\mathrm{j} + \chi$ events: $\vec{p}_{T,1}$ refers to the transverse momentum of the hard-jet or bottom jet pair system, while $\vec{p}_{T,2}$ to that of the soft-jet or light-flavor jet pair system. Best predictions are provided by CP2 tune.

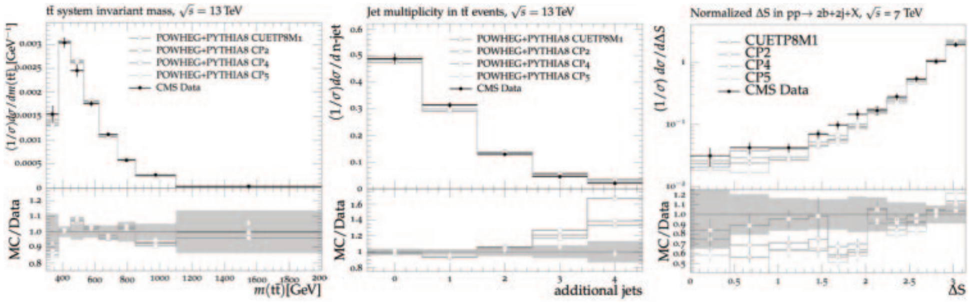

Fig. 2. – The invariant mass of the $\mathrm{t\bar{t}}$ system (left) and the number of additional jets (middle) from CMS $\sqrt{s} = 13$ TeV analysis [7]. The data are compared with POWHEG predictions, while the PS simulation is done with the PYTHIA8 tunes CUETP8M1, CP2, CP4, or CP5. The correlation observable ΔS measured in 2b+2j production (right) is compared to predictions of PYTHIA8 tunes CUETP8M1, CP2, CP4, and CP5, from the CMS $\sqrt{s} = 7$ TeV analysis [8].

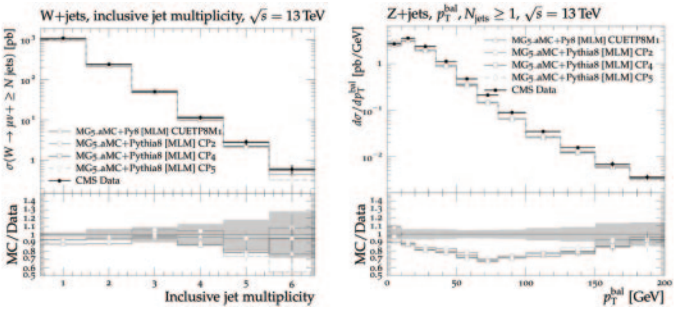

Fig. 3. – Comparison with the measurement of the inclusive jet multiplicity in W+jets (left) events and with the measurement of the p_T balance in Z+jets (right) predicted by MG5_aMC+PYTHIA8 with k_T-MLM merging [9, 10], for the different tunes. Tunes CP1 and CP3 are not shown but present a similar behavior as, respectively, tunes CP2 and CP4.

- W+jets and Z+jets production [9, 10]. In fig. 3 (left) the jet multiplicity in W+jets is shown. This distribution has little sensitivity to UE tunes, since only high jet-multiplicity is generated by PS, and all the different tune sets provide a good description of this observable, with a slightly better agreement for the CP2. In fig. 3 (right) the p_T balance (p_T^{bal}) distribution is shown, for Z+jets events with $N_{jets} \geq 1$, where $p_T^{bal} = |\vec{p}_T(Z) + \sum_{jets} \vec{p}_T(j_i)|$. Differences between the tunes are significant below ≈20 GeV, with CP2 tune providing better description of data.

3. – Summary and conclusions

The extraction of new tunes for the UE simulation of PYTHIA8 generator has been presented, and a significant improvement in the description of UE measurements at 13 TeV has been observed with respect to old Monash-based tune CUETP8M1, that showed a ≈10% disagreement in the 13 TeV transMIN region with $p_T^{max} > 5$ GeV, and did not provide a good fit to the jet multiplicity in $\mathrm{t\bar{t}}$ production. In addition, for the first time, tunes based on higher-order PDF sets have given a reliable description of UE measurements, with similar level of agreement to predictions from LO tunes. CMS has chosen the CP5 tune for the official MC Run2 production.

REFERENCES

[1] Sjöstrand T. *et al.*, *Comput. Phys. Commun.*, **191** (2015) 159.
[2] CMS Collaboration, *Eur. Phys. J. C*, **80** (2020) 4.
[3] NNPDF Collaboration, *Eur. Phys. J. C*, **77** (2017) 663.
[4] Skands P., Carrazza S. and Rojo J., *Eur. Phys. J. C*, **74** (2014) 3024.
[5] Andy Buckley D. G. and Jonathan Butterworth, *Comput. Phys. Commun.*, **184** (2013) 2803.
[6] Buckley A. *et al.*, *Eur. Phys. J. C*, **65** (2010) 331.
[7] CMS Collaboration, *Phys. Rev. D*, **95** (2017) 092001.
[8] CMS Collaboration, *Phys. Rev. D*, **94** (2016) 112005.
[9] CMS Collaboration, *Phys. Rev. D*, **96** (2017) 72005.
[10] CMS Collaboration, *Eur. Phys. J. C*, **78** (2018) 965.