

Top Quark Modelling and Tuning at CMS

EMYR CLEMENT
ON BEHALF OF THE CMS COLLABORATION*University of Bristol*

Recent measurements dedicated to improving the understanding of modelling top quark pair ($t\bar{t}$) production at the LHC are summarised. These measurements, performed with proton-proton collision data collected by the CMS detector at $\sqrt{s} = 13$ TeV, probe the underlying event in $t\bar{t}$ events, and use the abundance of jets in $t\bar{t}$ events to study the sub-structure of jets. A new set of tunes for PYTHIA 8, and their performance with $t\bar{t}$ data, are also discussed.

PRESENTED AT

11th International Workshop on Top Quark Physics
Bad Neuenahr, Germany, September 16–21, 2018

1 Introduction

Accurate modelling of top quark production at the LHC is a crucial aspect of top quark measurements. This is particularly true as modelling uncertainties often dominate over statistical and experimental sources of uncertainty in our measurements, and become a limiting factor in the precision of the measurements. The abundance of data collected by the CMS detector [1] allows extensive studies of top quark production to be performed with the aim of probing the accuracy of various aspects of Monte Carlo (MC) generators.

In these proceedings, two recent measurements, dedicated to improving the understanding of the modelling of top quark pair ($t\bar{t}$) production, are presented. The predictions of a new set of tunes for the PYTHIA 8 parton shower (PS) generator [2] are also presented, along with their performance when compared to top quark data.

2 Simulation of $t\bar{t}$ events

In the two measurements presented here, the nominal predictions for $t\bar{t}$ production are obtained using the POWHEG (v2) matrix element (ME) generator [3, 4, 5, 6], interfaced with PYTHIA 8 for the simulation of the PS. The ME calculations are at next-to-leading-order (NLO) accuracy, and PYTHIA 8 employs the CUETP8M2T4 tune [7]. This tune was derived to address a mismodelling of the jet multiplicity observed with the previous tune, CUETP8M1 [8]. The data are also compared with predictions obtained with other state-of-the-art ME and PS generators, such as MADGRAPH_MC@NLO [9], SHERPA [10], and HERWIG 7 [11, 12].

3 The Underlying Event in $t\bar{t}$ events

Measurements of the underlying event (UE) in the LHC environment are typically performed using minimum bias (MB) or DY events. Using $t\bar{t}$ events to study the UE probes the universality of the UE modelling at higher scales, of around twice the top mass (m_t). The properties of the UE, namely the multiplicity, momentum flux, and UE shape, have been measured by CMS with 35.9 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13 \text{ TeV}$ [13]. The measurement uses $t\bar{t}$ events with two leptons (e^\pm or μ^\pm) and two b -jets in the final state, which allows these products from the hard scatter, and those associated with pileup interactions, to be separated from the charged particles associated with the UE. The measured properties of the UE are unfolded to particle level, to a phase space that mimics that used to select the events.

Figure 1 shows the charged particle multiplicity of the UE, along with comparisons to different MC generators. The predictions of the POWHEG+PYTHIA 8 generator

configuration are also shown after varying several parameters sensitive to the modelling of $t\bar{t}$ production, such as the ISR and FSR scales, and the model used to simulate colour reconnection (CR). The simulation of multiparton interactions (MPI) and CR are found to be important in achieving an accurate description of the measured distributions, demonstrated by the effect of turning off the simulation of these processes. The data also favours a lower value of the FSR scale, $\alpha_S^{\text{FSR}}(M_Z)$, used in PYTHIA 8.

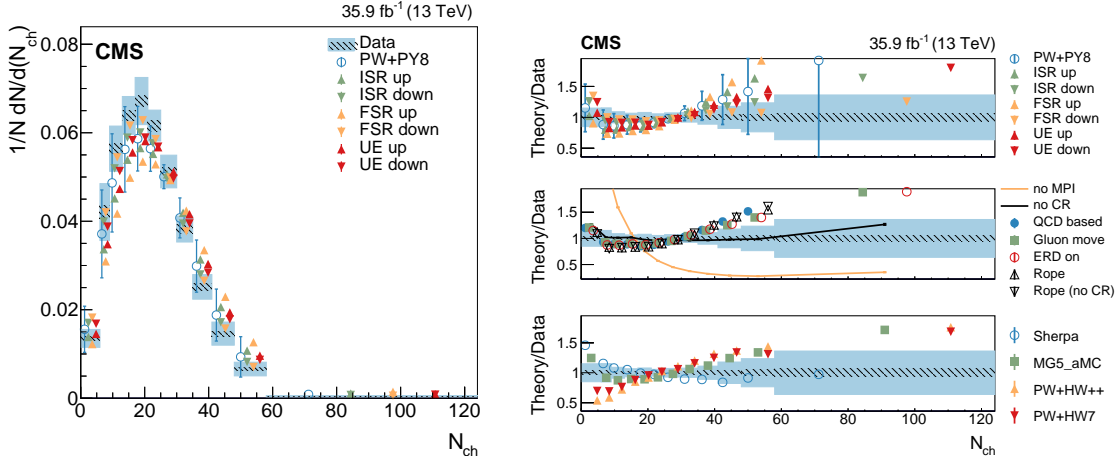


Figure 1: The charged particle multiplicity of the UE in $t\bar{t}$ events [13]. The data are compared to several different MC generators.

In order to enhance the sensitivity of the measured distributions to different aspects of the simulation, the data are also categorised into regions of different jet multiplicity, dilepton transverse momentum (p_T), and dilepton invariant mass (m_{ll}). The difference of the azimuthal angle, $\Delta\Phi$, between a charged particle in the UE and the dilepton momentum vector, \vec{p}_{ll} , is also used to define regions in the transverse plane. These are the "towards" ($\Delta\Phi < 60^\circ$), "transverse" ($120^\circ < \Delta\Phi < 60^\circ$), and "away" regions ($\Delta\Phi > 120^\circ$). The mean p_T of charged particles, $|\bar{p}_T|$, in the UE, as a function of the jet multiplicity in these different regions is shown in Figure 2. In each bin of jet multiplicity, the mean of the $|\bar{p}_T|$ distribution in that bin is shown.

4 Jet Substructure using $t\bar{t}$ events

Top quark pair events with one lepton in the final state provide an abundance of jets. The substructure of these jets has been studied by CMS using 35.9 fb^{-1} of data at $\sqrt{s} = 13 \text{ TeV}$ [14], testing the applicability of the PS models in the LHC environment. Several families of observables are measured, such as generalised angularities, energy correlation functions, and observables related to the soft drop algorithm. The charged

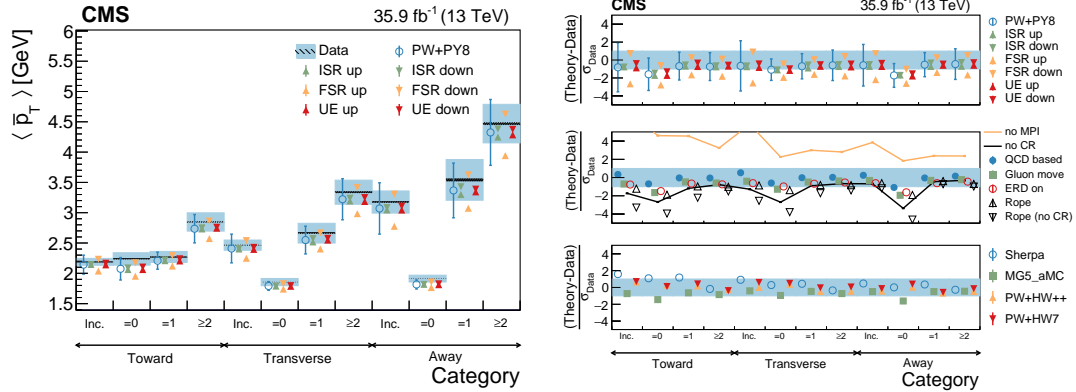


Figure 2: The mean $|\bar{p}_T|$ of the UE in $t\bar{t}$ events with different jet multiplicity, in the "towards", "transverse", and "away" regions [13]. The data are compared to several different MC generators.

particle multiplicity inside the jets, unfolded to particle level in a phase space similar to that used at reconstruction level, is shown in Figure 3. Samples of jets enriched in bottom quark, light quark, and gluon jets are also obtained, and the charged particle multiplicity for these different samples of jets is also shown in Figure 3. The distributions measured in the analysis typically show that the data again favours a lower value of $\alpha_S^{\text{FSR}}(M_Z)$ in PYTHIA 8. In addition to this, the agreement of the different predictions with the data is better for the light quark and gluon jet enriched samples, and worse for the b -jet enriched sample.

5 Extraction of $\alpha_S^{\text{FSR}}(M_Z)$

Both measurements presented in Sections 3 and 4 observe that a lower value of $\alpha_S^{\text{FSR}}(M_Z)$ is favoured by the data, compared to that used in CUETP8M2T4 PYTHIA 8 tune. As a consequence of this, both measurements extract a value for $\alpha_S^{\text{FSR}}(M_Z)$ by performing a scan of the χ^2 between the data and predictions from POWHEG+PYTHIA 8 as a function of $\alpha_S^{\text{FSR}}(M_Z)$. The most sensitive variable from the former analysis was found to be $|\bar{p}_T|$, and the extracted value was $\alpha_S^{\text{FSR}}(M_Z) = 0.120^{+0.006}_{-0.006}$. In the latter analysis, the angle between groomed subjets, ΔR_g , which is related to the jet width, was found to be the most sensitive variable, and the value extracted was $\alpha_S^{\text{FSR}}(M_Z) = 0.130^{+0.016}_{-0.020}$. Both extracted values are somewhat lower than that in the CUETP8M2T4 tune ($\alpha_S^{\text{FSR}}(M_Z) = 0.1365$). In addition to this, the uncertainties in the extracted value from the $|\bar{p}_T|$ distribution are approximately equivalent to a $\sqrt{2}$ variation in the renormalization scale. This is smaller than the factor 2 variation typically considered by top quark measurements performed by CMS when estimating an uncertainty in this source.

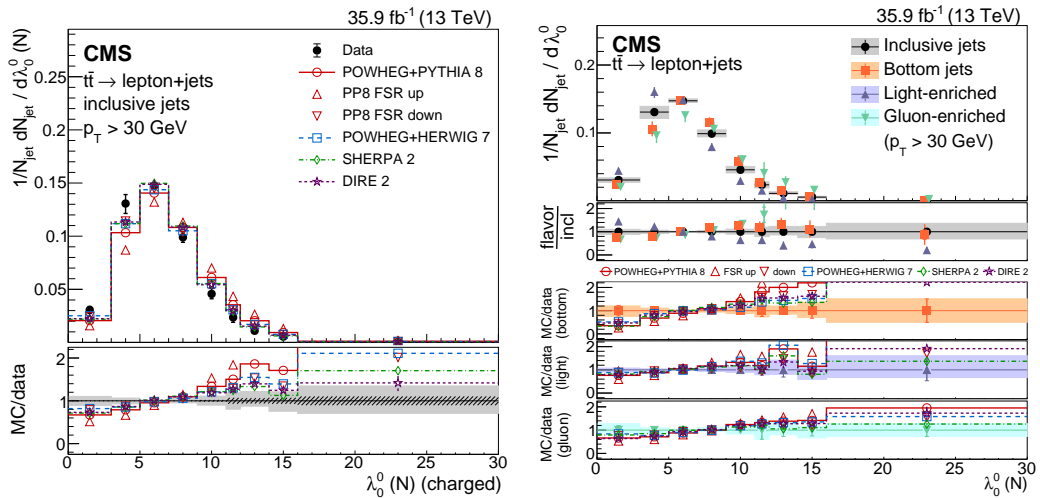


Figure 3: The charged particle multiplicity in jets [14]. On the left, the data are compared to different MC generators. On the right, the distributions are shown for data samples that are enriched in b -jets, light jets, and gluon jets.

6 New PYTHIA 8 tunes

A set of PYTHIA 8 tunes were derived by CMS where the same PDF set and α_S value are used in all aspects of the simulation, such as the hard process, ISR, FSR, and the PS [15]. The tunes were derived using UE and MB data, and the performance was tested with top data, along with UE, MB, and DY data.

Figure 4 compares the predictions for the leptonically decaying top quark p_T and jet multiplicity from POWHEG+PYTHIA 8 using the CUETP8M1, CP2, CP4, and CP5 tunes, with data. CP2 uses the NNPDF3.1 LO PDF set [16], and $\alpha_S = 0.130$. Both CP4 and CP5 uses NNPDF3.1 NNLO PDF set with $\alpha_S = 0.118$, however CP5 also enables rapidity ordering for ISR. The predictions from each setup are found to be able to describe the top quark kinematics, however only CP5 provides an accurate description for the whole range of additional jet multiplicity.

References

- [1] CMS Collaboration, JINST **3**, S08004 (2008).
- [2] T. Sjostrand, S. Mrenna and P. Z. Skands, Comput. Phys. Commun. **178**, 852 (2008).
- [3] P. Nason, JHEP **0411**, 040 (2004).
- [4] S. Frixione, P. Nason and C. Oleari, JHEP **0711**, 070 (2007).

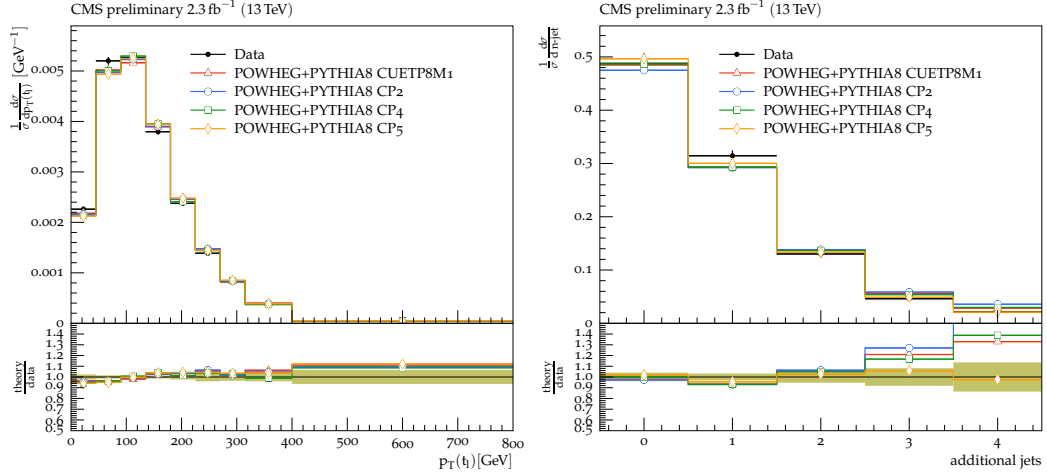


Figure 4: The p_T of the leptonically decaying top quark (left) and jet multiplicity (right) in $t\bar{t}$ events [15]. The data are compared to predictions from POWHEG+PYTHIA 8, with various PYTHIA 8 tunes.

- [5] S. Alioli, P. Nason, C. Oleari and E. Re, JHEP **1006**, 043 (2010).
- [6] S. Frixione, M. L. Mangano, P. Nason and G. Ridolfi, Adv. Ser. Direct. High Energy Phys. **15**, 609 (1998).
- [7] CMS Collaboration, CMS-PAS-TOP-16-021 (2016), <http://cds.cern.ch/record/2235192>.
- [8] CMS Collaboration, Eur. Phys. J. C **76**, no. 3, 155 (2016).
- [9] J. Alwall et al., JHEP **1407**, 079 (2014).
- [10] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert and J. Winter, JHEP **0902**, 007 (2009).
- [11] M. Bahr et al., Eur. Phys. J. C **58**, 639 (2008).
- [12] J. Bellm et al., Eur. Phys. J. C **76**, no. 4, 196 (2016).
- [13] CMS Collaboration, arXiv:1807.02810 (Submitted to EPJC).
- [14] CMS Collaboration, Phys. Rev. D **98**, no. 9, 092014 (2018)
- [15] CMS Collaboration, CMS-PAS-GEN-17-001 (2018), <http://cds.cern.ch/record/2636284>.
- [16] R. D. Ball et al. (NNPDF Collaboration), Eur. Phys. J. C **77**, no. 10, 663 (2017).