Distributions of topological observables in inclusive three- and four-jet events in pp collisions at $\sqrt{s} = 7$ TeV

CMS Collaboration

CERN, 1211 Geneva 23, Switzerland

Abstract This paper presents distributions of topological observables in inclusive three- and four-jet events produced in pp collisions at a centre-of-mass energy of 7 TeV with a data sample collected by the CMS experiment corresponding to a luminosity of 5.1 fb$^{-1}$. The distributions are corrected for detector effects, and compared with several event generators based on two- and multi-parton matrix elements at leading order. Among the considered calculations, MadGraph interfaced with PYTHIA6 displays the overall best agreement with data.

1 Introduction

In proton-proton collisions at the LHC, interactions take place between the partons of the colliding protons. The scattered partons from hard collisions fragment and hadronize into collimated groups of particles called jets. The study of jets with high transverse momentum ($p_T$) provides a test for physics beyond the standard model. While parton scattering is an elementary QCD process that can be calculated from first principles, predictions of jet distributions require an accurate hadronization model. In this paper, several hadronization models are examined.

High-$p_T$ parton production is described by perturbative QCD (pQCD) in terms of the scattering cross section convolved with a parton distribution function (PDF) for each parton that parametrizes the momentum distribution of partons within the proton. The hard-scattering cross section itself can be written as an expansion in the strong coupling constant $\alpha_s$. The leading term in this expansion corresponds to the emission of two partons. The next term includes diagrams where an additional parton is present in the final state as a result of hard-gluon radiation (e.g. $gg \rightarrow gg$). Cross sections for such processes diverge when any of the three partons becomes soft or when two of the partons become collinear. Finally, pQCD predicts three classes of four-jet events that correspond to the processes $qq/\rightarrow qqqg$, $qq/\rightarrow qgq$, and $qg/\rightarrow qggg$, where $q$ stands for both quarks and anti-quarks. Processes with two or more gluons in the final state receive a contribution from the triple-gluon vertex, a consequence of the non-Abelian structure of QCD.

We are studying distributions of topological variables, which are sensitive to QCD color factors, the spin structure of gluons, and hadronization models. These topological variables were studied widely in the earlier LEP [1,2] and the Tevatron [3,4] experiments and help to validate theoretical models implemented in various Monte Carlo (MC) event generators.

The distributions of multijet variables are sensitive to the treatment of the higher-order processes and approximations involved. Many MC event generators make use of leading order (LO) matrix elements (ME) in the primary $2 \rightarrow 2$ process. A good agreement between the measurements and MC predictions can establish the validity of the treatment of higher-order effects, and any large deviation may lead to large systematic uncertainties in searches for new physics.

The multijet observables presented here are based on hadronic events from 7 TeV pp collision data recorded with the CMS detector corresponding to an integrated luminosity of 5.1 fb$^{-1}$. The kinematic and angular properties of these events are computed from the jet momentum four-vectors. Unfolding techniques are used to correct for the effects of the detector resolution and efficiency. Systematic uncertainties resulting from the limited knowledge of the jet energy scale (JES), jet energy and angular resolution (JER), unfolding, and event selection are estimated, and the unfolded distributions are compared with predictions of several QCD-based MC models.

In this paper, the CMS detector is briefly described in Sect. 2. Sections 3 and 4 summarize the MC models used and the variables studied in this paper. Event selection and measurements are described in Sects. 5 and 6, respectively. The correction of the distributions due to detector effects is

* e-mail: cms-publication-committee-chair@cern.ch
discussed in Sect. 7. Sections 8 and 9 describe the estimation of systematic uncertainties and the final results. The overall summary is given in Sect. 10.

2 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the field volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. Extensive forward calorimetry complements the coverage provided by the barrel and endcap detectors. The barrel and endcap calorimeters cover a pseudorapidity region $\eta < 3.0$. Pseudorapidity is defined as $\eta = -\ln \tan(\theta/2)$, where $\theta$ is the polar angle. The transition between barrel and endcaps happens at $|\eta| = 1.475$ for the ECAL and $|\eta| = 1.15$ for the HCAL. The first level (L1) of the CMS trigger system, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select the most interesting events in a fixed time interval of less than 4 $\mu$s. The high-level trigger (HLT) processor farm further decreases the event rate from around 100 kHz to around 400 Hz before data storage. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [5].

3 Monte Carlo models

The MC event generators rely on models using modified LO QCD calculations. The elementary hard process between the partons is computed at LO. The parton shower (PS), used to simulate higher-order processes, follows an ordering principle motivated by QCD. Nevertheless, the parton shower models can differ in the ordering of emissions and the event generators can also have different treatments of beam remnants and multiple interactions.

The PYTHIA 6.4.26 [6] event generator uses a PS model to simulate higher-order processes [7–9] after the LO ME from pQCD calculations. The PS model, ordered by the $p_T$ of the emissions, provides a good description of event shapes when the emitted partons are close in phase space. Events are generated with the Z2 tune [10] for the underlying event. This tune is identical to the Z1 tune [11], except that it uses CTEQ6L1 [12] PDFs. The partons are hadronized (process of converting the partons into measured particles) using the Lund string model [13,14].

The PYTHIA 8.153 [15] event generator also uses a PS model with the successive emissions of partons ordered in $p_T$ and the Lund string model for hadronization. The main difference between the two PYTHIA versions is the description of multiparton interactions (MPI). In PYTHIA8, initial state radiation (ISR), final state radiation (FSR), and MPI are interleaved in the $p_T$ ordering, while in PYTHIA6, only ISR and FSR are interleaved. The tune 4C [16] is used with this generator. This tune uses CTEQ6L1 PDFs with parameters using CDF as well as early LHC measurements.

The HERWIG++ 2.4.2 [17] tune 23 [18] program takes the LO ME and simulates a PS using the coherent branching algorithm with angular ordering [19] of the showers. The partons are hadronized in this model using a cluster model [20] and the underlying event is simulated using the eikonal multiple partonic scattering model.

In the case of MADGRAPH 5.1.5.7 [21], multiparton final states are also computed at tree level. The parton shower and nonperturbative parts for MADGRAPH 5.1.5.7 simulation sample is handled by PYTHIA 6.4.26 with Z2 tune. The MLM matching procedure [22] is used to avoid double counting between the ME and the PS. The MadGraph samples are created in four bins of the variable $H_T$, the scalar sum of the parton $p_T$. The matching between ME and PS has been studied in detail and has been validated using inclusive jet $p_T$ distributions. Several samples are generated using different matching parameters and are used in estimating systematic uncertainty in the theoretical prediction.

These MC programs are the most commonly used models to describe multi-partonic final states and are normally used to describe QCD background in searches within CMS. The events produced from these models are simulated using a CMS detector simulation program based on GEANT4 [23] and reconstructed with the same program used for the data. These MC events are used for the comparison with the measurements as well as to correct the distributions for detector effects.

4 Definition of variables

4.1 Three-jet variables

The topological variables used in this study are defined in the parton or jet centre-of-mass (CM) system. The topological properties of the three-parton final state in the CM system can be described in terms of five variables [3]. Three of the variables reflect partition of the CM energy among the three final-state partons. There are three angles, which define the spatial orientation in the plane containing the three partons, but only two are independent.

It is convenient to introduce the notation $1 + 2 \rightarrow 3 + 4 + 5$ for the three-parton process. Here, numbers 1 and 2 refer to
incoming partons while the numbers 3, 4, and 5 label the outgoing partons in a descending order in energies in the parton CM frame, i.e. $E_3 > E_4 > E_5$ (Fig. 1). The final-state parton energy is an obvious choice for the topological variable for the three-parton final state. For simplicity, $E_i$ ($i = 3, 4, 5$) is often replaced by the scaled variable $x_i$ ($i = 3, 4, 5$), which is defined by $x_i = 2E_i/\sqrt{s_{345}}$, where $\sqrt{s_{345}}$ is the CM energy of the hard-scattering process. It is also referred to as the mass of the three-parton system, and by definition, $x_3 + x_4 + x_5 = 2$.

The internal structure of the three-parton final state is determined by any two scaled parton energies. The third one is calculated using Eq. 1. It needs two angular variables which fix the event orientation. In total, five independent kinematic variables are needed to describe the topological properties of the three-parton final state. In this analysis, however, the study is restricted to three variables: $\sqrt{s_{345}}$, $x_3$, and $x_4$, while the angular variables are not included.

### 4.2 Four-jet variables

To define a four-parton final state in its CM frame, eight independent parameters are needed. Two of these define the overall event orientation, while the other six fix the internal structure of the four-parton system. In contrast to the three-parton final state, there is no simple relationship between the scaled parton energies and the opening angles between partons. Consequently, the choice of topological variables is less obvious in this case. Variables are defined here in a way similar to those investigated for the three-parton final state. The four partons are ordered in descending energy in the parton CM frame and labeled from 3 to 6. The variables include the scaled energies and the polar angles of the four partons with respect to the beams.

In addition to the four-parton CM energy or the mass of the four-parton system ($\sqrt{s_{3456}}$), two angular distributions characterizing the orientation of event planes are investigated. One of these is the Bengtsson–Zerwas angle ($\chi_{BZ}$) [24] defined as the angle between the plane containing the two leading jets and the plane containing the two nonleading jets:

$$\cos \chi_{BZ} = \frac{(p_3 \times p_4) \cdot (p_5 \times p_6)}{|p_3 \times p_4| |p_5 \times p_6|}. \tag{2}$$

The second variable is the cosine of the Nachtmann–Reiter angle ($\cos \theta_{NR}$) [25] defined as the angle between the momenta of the two leading jets and the two nonleading jets:

$$\cos \theta_{NR} = \frac{(p_3 - p_4) \cdot (p_5 - p_6)}{|p_3 - p_4| |p_5 - p_6|}. \tag{3}$$

Figure 2 illustrates the definitions of $\chi_{BZ}$ and $\theta_{NR}$ variables. Historically, $\chi_{BZ}$ and $\theta_{NR}$ were proposed for $e^+e^-$ collisions to study gluon self-coupling. Their interpretation in pp collisions is more complicated, but the variables can be used as a tool for studying the internal structure of the four-jet events.

### 5 Data samples and event selection

Jets are reconstructed from particle-flow (PF) objects [26, 27] using the anti-$k_T$ clustering algorithm [28] with the distance parameter $R = 0.5$, as calculated with FASTJET 2.0 [29]. The PF algorithm utilizes the best energy measurements of each particle candidate from the most suitable combination of the detector components. A cluster is formed from all the particle-flow candidates that satisfy the chosen distance parameter. The four-momentum of the jet is then defined as the sum of four-momenta of the corresponding particle-flow candidates, which results in jets with nonzero mass.
Table 1  Prescales, integrated luminosity and offline $p_T$ threshold of the leading jet for different trigger paths. The terminology for Level 1 (L1) triggers as well as HLT includes the jet $p_T$ threshold (in GeV) applicable to the trigger

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2011A</td>
<td>L1 prescale</td>
<td>1–300</td>
<td>1–10</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>HLT prescale</td>
<td>15–180</td>
<td>1–5000</td>
<td>1–60</td>
<td>1–24</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$\int\mathcal{L}$ (pb$^{-1}$)</td>
<td>0.29</td>
<td>6.16</td>
<td>114.7</td>
<td>392.2</td>
<td>2328</td>
</tr>
<tr>
<td>2011B</td>
<td>L1 prescale</td>
<td>50–400</td>
<td>1–20</td>
<td>1–10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>HLT prescale</td>
<td>80–84</td>
<td>80–1000</td>
<td>10–100</td>
<td>4–30</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$\int\mathcal{L}$ (pb$^{-1}$)</td>
<td>0.12</td>
<td>1.12</td>
<td>40.2</td>
<td>136.0</td>
<td>2767</td>
</tr>
<tr>
<td>Overall</td>
<td>$\int\mathcal{L}$ (pb$^{-1}$)</td>
<td>0.41</td>
<td>7.29</td>
<td>154.8</td>
<td>528.2</td>
<td>5096</td>
</tr>
<tr>
<td>$p_T$ threshold</td>
<td>110 GeV</td>
<td>190 GeV</td>
<td>300 GeV</td>
<td>360 GeV</td>
<td>500 GeV</td>
<td></td>
</tr>
</tbody>
</table>

Table 2  Threshold of the leading jet $p_T$ for different HLT paths. This paper shows results from two representative trigger paths HLT110 and HLT370

<table>
<thead>
<tr>
<th>HLT</th>
<th>HLT60</th>
<th>HLT110</th>
<th>HLT190</th>
<th>HLT240</th>
<th>HLT370</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leading jet $p_T$ (GeV)</td>
<td>110–190</td>
<td>190–300</td>
<td>300–360</td>
<td>360–500</td>
<td>&gt;500</td>
</tr>
</tbody>
</table>

The JES correction applied to jets used in this analysis is based on high-$p_T$ jet events generated by PYTHIA6 and then simulated using GEANT4, and in situ measurements with dijet and photon + jet events [30]. An average of ten minimum bias interactions occur in each pp bunch crossing (pileup), and this requires an additional correction to remove the extra energy deposited by these pileup events. The size of the correction depends on the $p_T$ and $\eta$ of the jet. The correction appears as a multiplicative factor to the jet energy, and is typically less than 1.2 and approximately uniform in $\eta$.

Events passing single-jet HLT requirements are used in this analysis. These triggers require jets reconstructed from calorimetric information with the anti-$k_T$ clustering algorithm and with energy corrections applied. Jets are ordered in decreasing jet $p_T$, and the leading jet $p_T$ is required to be above a certain threshold. As offline jets are reconstructed with the PF algorithm, this may result in a trigger not being fully efficient near the threshold. Trigger efficiencies are studied as a function of the leading jet $p_T$ for all trigger thresholds. Values of the leading jet $p_T$, where the trigger efficiency is determined to be larger than 99 %, are listed in Table 1. It also summarizes the prescale factors and the effective integrated luminosities collected using the different HLT thresholds.

Jets are selected with restrictive criteria on the neutral energy fractions (both electromagnetic and hadronic components), and all the jets are required to have $p_T > 50$ GeV and absolute rapidity, $y = (1/2) \ln((E + p_z)/(E - p_z))$, $|\eta| \leq 2.5$. The jet with the highest $p_T$ is required to be above a threshold as given by the requirement from the trigger turn-on curve. To avoid overlap of events from two different HLT paths, the $p_T$ of the leading jet is also required to be less than an upper value. The overall criteria are summarized in Table 2. Though data from all the five trigger paths are studied, figures from two representative trigger paths (the highest $p_T$ threshold and a lower one with good statistical accuracy) are presented in this paper.

Events are selected with at least three jets passing the selection criteria as stated above. Additional selection requirements are also applied to reduce backgrounds due to beam halo, cosmic rays and detector noise. The event must have at least one good reconstructed vertex [31]. Missing transverse energy, $E_T^{miss}$, is required to be less than 0.3 $\sum E_T$, where the summation is over all PF jets. The quantities $E_T^{miss}$ and $\sum E_T$ are obtained from negative vector sum and scalar sum of the transverse momenta of the jets, respectively. A number of event filters [32] accept only those events that have negligible noise in the detector. The jets are ordered in decreasing $p_T$, and an event with at least three (four) jets satisfying the jet selection criteria is classified as a three-jet (four-jet) event.

6 Measurements

The 4-momenta of all the jets in the three- or four-jet event category are transformed into the CM frame of the three- or four-jet system. The jets are then ordered in decreasing energy. The three- and four-jet variables as described in Sect. 4 are then calculated from the kinematic and angular information of the jets. Since detector resolution varies over the potential kinematic ranges, variable bin widths are adopted for the jet masses and the scaled jet energies, while for angular variables constant bin widths are used.
6.1 Detector-level distributions

The measured distributions of the three- and four-jet variables are compared with predictions from two MC generators (PYTHIA6 and MADGRAPH + PYTHIA6), simulated using the identical detector condition as that in the data. The identical pileup condition is obtained by reweighting the MC simulation to match the spectrum of pileup interactions observed in the data. The size of the reweighting correction is typically less than 1 %. The agreement between the data and the MC predictions is reasonable, so these MC generators are used to correct the measured distributions.

Figures 3 and 4 show the normalized three- and four-jet mass distributions. The data are compared with two different MC programs: PYTHIA6 and MADGRAPH + PYTHIA6, each with two different HLTs with \( p_T \) thresholds above 110 and 370 GeV. As can be seen from the figures, there is agreement within a few percent between the data and the predictions of these two simulations. The difference between the predictions and the data varies typically from 4 to 10 %. However, there is a systematic deviation observed at high masses where the simulations are higher than the data.

7 Corrections for detector effects

Multijet variables obtained from MC samples may differ from data because of the detector resolution and acceptance. Before comparisons with other experiments or theoretical predictions can be made, detector effects are unfolded into distributions at the final-state particle level. The basic component of the unfolding is the response function, where experimental observables are expressed as a function of theoretical observables. For simplicity, observables are taken in discrete sets, and the response function is replaced by a response matrix. The observed distribution is then unfolded with the inverse of response matrix to obtain a distribution corrected for detector effects. Matrix inversion has potential complications, because it cannot handle large statistical fluctuations and the matrix itself could be singular. Instead, we use the RooUnfold package [33] with the D’Agostini iterative method [34] as the default algorithm and the singular value decomposition method [35] for cross-checks.

The default response matrix is obtained using the PYTHIA6 event generator. Statistical uncertainties are estimated from the square root of the covariance matrix obtained from a variation of the results generated by simulated experiments.

The corrections for detector resolution and acceptance change the shape of the three-jet mass distributions by approximately 10 %, less than 5 % for the scaled energy of nonleading jets, and up to 20 % for the scaled energy of the leading jet. For four-jet variables, corrections applied are of the order of 20 % for the four-jet mass, 10 % for \( x_{\text{BZ}} \), and less than 5 % for cos \( \theta_{\text{NR}} \).

Fig. 3 The upper panels display the normalized distributions of the reconstructed three-jet mass for events where the most forward jet has \(|y| < 2.5\). Figures differ by \( p_T \) ranges of the leading jet: 190–300 GeV (a), and above 500 GeV (b) for data (before correction due to detector effects) and predictions from MC generators. The bottom panel of each plot shows the ratio of MC predictions to the data. The data are shown with only statistical uncertainty.

8 Systematic uncertainties

The leading sources of systematic uncertainty are due to the JES, the JER, and the model dependence of corrections to
the data. The distributions are presented in this analysis as normalized distributions, thus the absolute scale uncertainty of energy measurement does not play a significant role. There are insignificant contributions due to resolution of $\gamma$. The main contribution of JES or JER to the uncertainty in the measurements is due to the migration of events from one category of jet multiplicity to the other.

The effect of pileup in the measured distributions has been studied as a function of the number of reconstructed vertices in the event. None of the variables show any significant dependence on the pileup condition, so systematic uncertainty due to pileup can be neglected.

8.1 Jet energy scale

One of the dominant sources of systematic uncertainty is due to the jet energy scale corrections. The JES uncertainty has been estimated to be $2–2.5\%$ for PF jets [30], depending on the jet $p_T$ and $\eta$. In order to map this uncertainty to the multijet variables, all jets in the selected events are systematically shifted by the respective uncertainties, and a new set of values for the multijet variables is calculated. This causes a migration of events from an event category of a given jet multiplicity to a different jet multiplicity. The migration could be as high as $20\%$ for some of the event categories. The corresponding distributions are then unfolded using the standard procedure as described in Sect. 7. The difference of these values from the central unfolded results is a measure of the uncertainty owing to the JES.

Uncertainties owing to the JES are found to be between $0.2–5.5\%$ in the three-jet mass, and $0.3–10\%$ in the four-jet mass. The systematic uncertainties are the largest at both ends of the mass spectra. The systematic uncertainties in scaled energy are between $0.1$ and $2.0\%$, and those in angular variables are in the range $0.1–3.0\%$. There is a small increase in the uncertainty for distributions where there is at least one jet in the endcap region of the detector.

8.2 Jet energy resolution

The JER is measured in data using the $p_T$ balance in dijet events [36]. Based on these measurements, the resolution effects are corrected using simulated events. To study the effect of the difference between the simulated and the measured resolution, several sets of unfolded distributions are obtained using response matrices from the default resolution matrix and changing the jet resolution within its estimated uncertainty. Alternatively, the response matrix is constructed by convolving the generator level distribution with the measured resolution. The measured distribution is unfolded by this response matrix vis-a-vis the response matrix determined using fully simulated sample of PYTHIA6 events. These two estimates provide independent descriptions of the detector modeling and the difference is used as a measure of the systematic uncertainty due to detector performance. Position resolution affects the measurement of the jet direction, and it is estimated using simulated multijet events and validated with data.

Uncertainties owing to the JER are found to be between $0.1–10\%$ in the three-jet mass, $0.3–15\%$ in the four-jet mass,
0.1–10 % in the scaled jet energies and 0.2–8.2 % in the angular variables.

8.3 Model dependence in unfolding

Unfolded distributions are obtained using two different response matrices derived from PYTHIA6 and MADGRAPH+PYTHIA6 simulations. The difference in the unfolded values, due to the choice of response functions, gives a measure of the systematic uncertainty. The uncertainties are at the level of 0.1–6.0 % in the three-jet mass, 0.1–3.0 % in the scaled energy distributions of the three-jet variables, 0.1–8.0 % in the four-jet mass and 0.1–6.2 % in the angular variables in the four-jet samples. The uncertainties in the scaled jet energy increases by a few percent for the samples with lower values of leading jet $p_T$.

Unfolding has been carried out using PYTHIA6 and MADGRAPH + PYTHIA6 samples, which has the same hadronization model. To test the effects of different hadronization models, MC samples from HERWIG++, which provides a different PS and hadronization approach, are used. However, the simulated event sample generated using HERWIG++ is statistically inadequate to be used in a complete unfolding procedure. The difference between bin-by-bin correction factors obtained with PYTHIA6 and HERWIG++ is found to be somewhat larger than the uncertainty due to the difference in the unfolding matrices: 0.1–12 % in the scaled energy distributions of three-jet variables, 0.1–7.7 % in the angular variables in the four-jet samples and 0.1–11.6 % in the jet masses. The larger values from the two estimates are chosen as the systematic uncertainties due to unfolding.

8.4 Event selection

Jet candidates are required to pass certain criteria [37] designed to reduce unwanted detector effects. This analysis uses jets identified with very restrictive criteria on the ratio of the energy carried by neutral to that carried by charged particles. The effect of using these criteria is tested by reevaluating the same distributions with jets selected after relaxing the selection on the fractions of the energy carried by the neutral and the charged particles. Also, the selection on $E_T^{\text{miss}}$ is changed, and the effect of this is estimated from the difference in the observed distributions. The uncertainty due to the event selection is found to be below 0.2 %.

8.5 Overall uncertainty

The first three sources mentioned above are the dominant sources of systematic uncertainty. The contributions to the uncertainty from the selection requirements and pileup effects are found to be negligible. The uncertainties are calculated for each bin of the measured distributions and are added in quadrature. The overall systematic uncertainty is found to be smaller than the statistical uncertainty for most of the bins. Typical uncertainties for the six variables studied in this analysis are summarized in Table 3.

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>Uncertainty (%) for leading jet $p_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>190–300 GeV</td>
</tr>
<tr>
<td>Three-jet mass</td>
<td></td>
</tr>
<tr>
<td>Jet and event selection</td>
<td>0.1</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>0.3–5.0</td>
</tr>
<tr>
<td>Jet resolution</td>
<td>0.1–10.0</td>
</tr>
<tr>
<td>Model dependence in unfolding</td>
<td>0.2–11.0</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>0.3–12.7</td>
</tr>
<tr>
<td>Statistical uncertainty</td>
<td>1.4–14.5</td>
</tr>
<tr>
<td>Scaled energy of the leading jet</td>
<td></td>
</tr>
<tr>
<td>Jet and event selection</td>
<td>0.1</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>0.1–1.9</td>
</tr>
<tr>
<td>Jet resolution</td>
<td>0.2–6.2</td>
</tr>
<tr>
<td>Model dependence in unfolding</td>
<td>0.1–6.0</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>0.8–7.2</td>
</tr>
<tr>
<td>Statistical uncertainty</td>
<td>1.6–17.2</td>
</tr>
<tr>
<td>Scaled energy of the second-leading jet</td>
<td></td>
</tr>
<tr>
<td>Jet and event selection</td>
<td>0.1</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>0.1–2.0</td>
</tr>
<tr>
<td>Jet resolution</td>
<td>0.1–5.0</td>
</tr>
<tr>
<td>Model dependence in unfolding</td>
<td>0.4–9.0</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>1.0–8.3</td>
</tr>
<tr>
<td>Statistical uncertainty</td>
<td>1.3–16.4</td>
</tr>
<tr>
<td>Four-jet mass</td>
<td></td>
</tr>
<tr>
<td>Jet and event selection</td>
<td>0.1</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>0.4–6.9</td>
</tr>
<tr>
<td>Jet resolution</td>
<td>0.4–11.7</td>
</tr>
<tr>
<td>Model dependence in unfolding</td>
<td>0.3–7.0</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>0.4–13.7</td>
</tr>
<tr>
<td>Statistical uncertainty</td>
<td>3.1–30.9</td>
</tr>
<tr>
<td>Bengtsson–Zerwas angle</td>
<td></td>
</tr>
<tr>
<td>Jet and event selection</td>
<td>0.1</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>0.1–3.0</td>
</tr>
<tr>
<td>Jet resolution</td>
<td>0.4–5.4</td>
</tr>
<tr>
<td>Model dependence in unfolding</td>
<td>0.3–3.5</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>1.4–5.9</td>
</tr>
<tr>
<td>Statistical uncertainty</td>
<td>5.1–8.4</td>
</tr>
<tr>
<td>Nachtmann–Reiter angle</td>
<td></td>
</tr>
<tr>
<td>Jet and event selection</td>
<td>0.1</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>0.1–1.0</td>
</tr>
<tr>
<td>Jet resolution</td>
<td>0.1–4.6</td>
</tr>
<tr>
<td>Model dependence in unfolding</td>
<td>0.2–2.1</td>
</tr>
</tbody>
</table>
### Table 3 continued

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>Uncertainty (%) for leading jet $p_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>190–300 GeV</td>
</tr>
<tr>
<td>Total systematic</td>
<td>0.9–5.0</td>
</tr>
<tr>
<td>Statistical uncertainty</td>
<td>3.4–4.2</td>
</tr>
</tbody>
</table>

### Results

#### 9.1 Comparison with models

The normalized differential distributions, corrected for detector effects, are plotted as a function of the three- and four-jet inclusive variables and compared with predictions from the four MC models: 

- **pythia** 6,
- **pythia** 8,
- **MadGraph** + **pythia** 6,
- **herwig**++.

The variables considered for these comparisons are three-jet mass, scaled energies of the leading and next-to-leading jet in the three-jet sample in the three-jet CM frame, four-jet mass, and the two angles $\chi_{\text{BZ}}$ and $\theta_{\text{NR}}$.

For the comparison plots (Figs. 5, 6, 7, 8, 9), the upper panel shows the data and the model predictions with the corresponding statistical uncertainty. The lower panels in each plot show the ratio of MC prediction to the data for each model. Comparisons are made for two different ranges of the leading jet $p_T$: $190 < p_T < 300$ GeV and $p_T > 500$ GeV.

Figure 5 shows the normalized corrected differential distribution as a function of the three-jet mass for two ranges of the leading-jet $p_T$. The three-jet mass distribution broadens for larger $p_T$ thresholds. The models show varying degrees of success for the different ranges of leading-jet $p_T$. Most models differ from the data in the low-mass spectrum. The **pythia** 6 simulation provides a good description of the data in the lower $p_T$ bin, while it has a larger deviation in the higher $p_T$ bin. The mean difference is at the level of 1.8–4.0 %.

**herwig**++ provides the worst agreement among the four models – the mean difference is at the level of 4.0–15 %.

Figure 6 shows the corrected normalized differential distribution as a function of the scaled leading-jet energy in the inclusive three-jet sample. The distributions peak close to 1 and the peaks get sharper for higher leading-jet $p_T$ range. The scaled leading-jet energy $x_3$ is expected to follow a linear rise from $\frac{1}{3}$ to 1 for a phase space model, which has only energy-momentum conservation, while QCD predicts indicate the statistical and systematic uncertainties combined in quadrature. The lower panels of each plot show the ratios of MC predictions to the data. The ratios are shown with statistical uncertainty in the data as well as in the MC, while the band shows combined statistical and systematic uncertainties.
Fig. 6 Corrected normalized distribution of scaled energy of the leading-jet in the inclusive three-jet sample. The other explanations are the same as Fig. 5

a deviation from linearity at higher values of \( x_3 \). This feature is observed in the data, particularly for higher \( p_T \) bins. Only MADGRAPH + PYTHIA6 provides a consistent description of the data. The agreement improves for the sample with leading-jet \( p_T \) above 500 GeV. The difference between the predictions from MADGRAPH + PYTHIA6 and the data are at the level of 3.5–6.1%.

Figure 7 shows the corrected normalized differential distribution as a function of the scaled energy of the second-leading jet, \( x_4 \), in the inclusive three-jet sample. For kinematic reasons, \( x_4 \) is expected to lie between 1/2 and 1. The distribution peaks around 0.65 for the low \( p_T \) threshold sample. The peak shifts to higher values of \( x_4 \) and the distribution becomes broader for the larger \( p_T \) threshold sample. Predic-
Corrected normalized distribution of four-jet mass. The other explanations are the same as Fig. 5

Figure 8 shows comparisons of the corrected normalized differential distribution as a function of the four-jet mass for the four MC models. The distribution broadens at higher minimum $p_T$ value. As can be seen from the figure, HERWIG++ provides the worst comparison. The average deviations are at the level of 15% for many of the distributions, particularly for the sample with leading-jet $p_T$ between 190 and 300 GeV. The level of agreement for the other three MC models is better than 10% over the entire $p_T$ region.

The sub-leading jets in the four-jet event category are predominantly due to the secondary splitting of partons. In case of gluon splitting, they can be due to a $q\bar{q}$ pair or gluons. Both the angular distributions, $\theta_{NR}$ and $\chi_{BZ}$, are different for these two scenarios and are representative of the colour factors for these couplings.

Figure 9 shows similar comparisons for the Bengtsson–Zerwas angle. Because the azimuthal angle is not defined for the back-to-back jets, the opening angle between the two leading and two nonleading jets is required to be less than $160^\circ$. As can be seen from the average deviation of the ratios from unity, predictions from MADGRAPH + PYTHIA6 and HERWIG++ represent the data well, while those from PYTHIA6 do poorly.

Figure 10 shows the corrected normalized differential distribution as a function of the cosine of the Nachtmann–Reiter angle in the inclusive four-jet sample. Most of the models follow the broad features of the data. However, the degree of agreement with data is different among models. MADGRAPH + PYTHIA6 provides the best description of the data; HERWIG++ with angular ordering in the parton shower is close to the data (the agreement is better than 5%), while PYTHIA6 has the largest deviation (the agreement is typically between 10–12%).

9.2 Effect of hadronization, underlying event, and PDFs

The disagreement between data and the MC models may arise from the implementation of nonperturbative components in the simulation due to the fragmentation model or the choice of PDF set. These effects have been investigated by studying the uncertainties due to hadronization model and PDF parametrization.

The MC models have different ways of modeling the underlying events and hadronization of the partons into hadrons. This may result in different predictions of the distributions of multijet variables depending on whether they are computed at the hadron or at the parton level. This effect has been investigated by studying two different MC models: PYTHIA6 and HERWIG++. This is done by evaluating the distributions at the parton and hadron level. PYTHIA6 uses the LUND string model, while HERWIG++ uses the cluster model. Also, colour reconnections are done differently in the two models. A generator-level study is carried out for both these models, where the effect of hadronization is studied using distributions from jets at parton- and hadron-level. The ratio of the parton- to the hadron-level distribution is then com-
pared. The mean difference between the two hadronization models is typically less than 5 %.

Comparisons are also made to different tunes of the underlying event models within PYTHIA6. The tunes (D6T, DW, P0, Z1, Z2, Z2*) [10,11,38–40] differ in the cutoff used to regularize the $1/p_T^4$ divergence for final-state partons, the ordering of the showers (virtuality ordering vs. $p_T$ ordering), multiparton interaction model, PDFs, and data sets used in the

tune. The resulting distributions agree typically within 5 %, so the disagreements with the data cannot be fully explained by this effect.

The MC models use CTEQ6 as the default PDF parametrization. There are many different PDF sets, which are based on different input data, assumptions, and parametrizations. Thus any calculation of a cross section or distributions in the simulation depends on the choice of PDF set.
Also, each PDF set has its own errors from its parametric assumptions and data input to fitting. The effect of the PDF set choice on the multijet variables is calculated according to the recommendation of PDF4LHC group [41,42]. Since comparisons are made only with leading order Monte Carlo models in this paper, only two leading order PDF sets are used in this comparison: CTEQ6l and MSTW2008lo68cl [43]. The uncertainties are found to be typically at the level of 1.0–2.0 % depending on the variable type and $p_T$ range considered.

10 Summary

Distributions of topological variables for inclusive three- and four-jet events in pp collisions measured with the CMS detector at a centre-of-mass energy of 7 TeV were presented using a data sample corresponding to an integrated luminosity of 5.1 fb$^{-1}$. The distributions were corrected for detector effects, and systematic uncertainties were estimated. These corrected distributions were compared with the predictions from four LO MC models: PYTHIA6, PYTHIA8, HERWIG++, and MADGRAPH + PYTHIA6.

Distributions of three- and four-jet invariant mass from all models show significant deviation from the data at high mass. The fact that all models have a common PDF suggests that the PDF errors at high mass are underestimated. The PDFs at high invariant mass have recently been constrained by CMS using dijet $p_T$ distributions[44].

The MADGRAPH simulations are based on tree-level calculations for two-, three-, and four-parton final states, while PYTHIA and HERWIG++ can have only two partons in the final state before showering. Not surprisingly, the three-jet predictions of MADGRAPH + PYTHIA6 give a more consistent description of the distributions studied in this analysis. The notable exception is at high $x_T$ (the next-to-leading jet), where two jets carry most of the CM energy. The difference is probably due to a double counting of three-parton with two-parton (with a parton from showering) final states.

The PYTHIA and HERWIG++ models give poor descriptions of the energy fractions in the three-jet final state. In particular, the distributions of $x_3$ (the leading jet) show large shape differences between data and theory that are inconsistent with PDFs or hadronization model uncertainties. Since the distributions from MADGRAPH + PYTHIA6 agree with those from the data, the discrepancies with PYTHIA and HERWIG++ are likely due to missing higher multiplicity ME, which are present in MADGRAPH.

All the models compared in this study do remarkably well describing the four-jet Bengtsson–Zerwas angle. The PYTHIA models have some systematic deviation from the data in describing the Nachtmann–Reiter angle. Parton showers with angular ordering, as implemented in HERWIG++, yield a better agreement with the measured data for these angular variables.

Acknowledgments We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centres and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: the Austrian Federal Ministry of Science, Research and Economy and the Austrian Science Fund; the Belgian Fonds de la Recherche Scientifique, and Fonds voor Wetenschappelijk Onderzoek; the Brazilian Funding Agencies (CNPq, CAPES, FAPERJ, and FAPESP); the Bulgarian Ministry of Education and Science; CERN; the Academy of Sciences of Armenia, Ministry of Science and Technology, and National Natural Science Foundation of China; the Colombian Funding Agency (COLCIENCIAS); the Croatian Ministry of Science, Education and Sport, and the Croatian Science Foundation; the Research Promotion Foundation, Cyprus; the Ministry of Education and Research, Estonian Research Council via IUT23-4 and IUT23-6 and European Regional Development Fund, Estonia; the Academy of Finland, Finnish Ministry of Education and Culture, and Helsinki Institute of Physics; the Institut National de Physique Nucléaire et de Physique des Particules/CNRS, and Commissariat à l’Énergie Atomique et aux Énergies Alternatives/CEA, France; the Bundesministerium für Bildung und Forschung, Deutsche Forschungsgemeinschaft, and Helmholtz-Gemeinschaft Deutscher Forschungszentren, Germany; the General Secretariat for Research and Technology, Greece; the National Scientific Research Foundation, and National Innovation Office, Hungary; the Department of Atomic Energy and the Department of Science and Technology, India; the Institute for Studies in Theoretical Physics and Mathematics, Iran; the Science Foundation, Ireland; the Istituto Nazionale di Fisica Nucleare, Italy; the Ministry of Science, ICT and Future Planning, and National Research Foundation (NRF), Republic of Korea; the Lithuanian Academy of Sciences; the Ministry of Education, and University of Malaya (Malaysia); the Mexican Funding Agencies (CINVESTAV, CONACYT, SEP, and UASLP-FAI); the Ministry of Business, Innovation and Employment, New Zealand; the Pakistan Atomic Energy Commission; the Ministry of Science and Higher Education and the National Science Centre, Poland; the Fundação para a Ciência e a Tecnologia, Portugal; JINR, Dubna; the Ministry of Education and Science of the Russian Federation, the Federal Agency of Atomic Energy of the Russian Federation, Russian Academy of Sciences, and the Russian Foundation for Basic Research; the Ministry of Education, Science and Technological Development of Serbia; the Secretaría de Estado de Investigación, Desarrollo e Innovación y Programa Consolider-Ingenio 2010, Spain; the Swiss Funding Agencies (ETH Board, ETH Zurich, PSI, SNF, UnizH, Canton Zurich, and SER); the Ministry of Science and Technology, Taipei; the Thailand Center of Excellence in Physics, the Institute for the Promotion of Teaching Science and Technology of Thailand, Special Task Force for Activating Research and the National Science and Technology Development Agency of Thailand; the Scientific and Technical Research Council of Turkey, and Turkish Atomic Energy Authority; the National Academy of Sciences of Ukraine, and State Fund for Fundamental Researches, Ukraine; the Science and Technology Facilities Council, UK; the US Department of Energy, and the US National Science Foundation. Individuals have received support from the Marie-Curie programme and the European Research Council and EPLANET (European Union); the Leventis Foundation; the A. P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l’Industrie et dans l’Agriculture.
(FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Council of Science and Industrial Research, India; the HOMING PLUS programme of Foundation for Polish Science, cofinanced from European Union, Regional Development Fund; the Compagnia di San Paolo (Torino); the Consorzio per la Fisica (Trieste); MIUR Project 2010RTX4TM (Italy); the Thalys and Aristeia programmes cofinanced by EU-ESF and the Greek NSRF; and the National Priorities Research Program by Qatar National Research Fund.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. Funded by SCOAP³.

References

CMS Collaboration

Yerevan Physics Institute, Yerevan, Armenia
V. Khachatryan, A. M. Sirunyan, A. Tumasyan

Institut für Hochenergiephysik der OeAW, Vienna, Austria

National Centre for Particle and High Energy Physics, Minsk, Belarus
V. Mossolov, N. Shumeiko, J. Suarez Gonzalez

Universiteit Antwerpen, Antwerp, Belgium

Vrije Universiteit Brussel, Brussels, Belgium

Université Libre de Bruxelles, Brussels, Belgium

Ghent University, Ghent, Belgium

Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Université de Mons, Mons, Belgium
N. Belyi, T. Caebergs, E. Daubie, G. H. Hammad

Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil

Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil

Universidade Estadual Paulista, Universidade Federal do ABC, São Paulo, Brazil

Institute for Nuclear Research and Nuclear Energy, Sofia, Bulgaria
A. Aleksandrov, V. Genchev, R. Hadijiiska, P. Iaydjiev, A. Marinov, S. Piperov, M. Rodozov, G. Sultanov, M. Vutova

University of Sofia, Sofia, Bulgaria
A. Dimitrov, I. Glushkov, L. Litov, B. Pavlov, P. Petkov

Institute of High Energy Physics, Beijing, China
National Taiwan University (NTU), Taipei, Taiwan

Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok, Thailand
B. Asavapibhop, G. Singh, N. Srimanobhas, N. Suwonjandee

Cukurova University, Adana, Turkey

Physics Department, Middle East Technical University, Ankara, Turkey

Bogazici University, Istanbul, Turkey
A. Albayrak, E. Gülmese, M. Kaya, O. Kaya, T. Yetkin

Istanbul Technical University, Istanbul, Turkey
K. Cankocak, F. I. Vardarlı

National Scientific Center, Kharkov Institute of Physics and Technology, Kharkov, Ukraine
L. Levchuk, P. Sorokin

University of Bristol, Bristol, UK

Rutherford Appleton Laboratory, Didcot, UK

Imperial College, London, UK

Brunel University, Uxbridge, UK
J. E. Cole, P. R. Hobson, A. Khan, P. Kyberd, D. Leggat, D. Leslie, I. D. Reid, P. Symonds, L. Teodorescu, M. Turner

Baylor University, Waco, USA
J. Dittmann, K. Hatakeyama, A. Kasmi, H. Liu, T. Scarborough

The University of Alabama, Tuscaloosa, USA

University of California, Davis, USA
The University of Iowa, Iowa City, USA

Johns Hopkins University, Baltimore, USA

The University of Kansas, Lawrence, USA

Kansas State University, Manhattan, USA
I. Chakaberia, A. Ivanov, S. Khalil, M. Makouski, Y. Maravin, L. K. Saini, N. Skhirtladze, I. Svintradze

Lawrence Livermore National Laboratory, Livermore, USA
J. Gronberg, D. Lange, F. Rebassoo, D. Wright

University of Maryland, College Park, USA

Massachusetts Institute of Technology, Cambridge, USA

University of Minnesota, Minneapolis, USA

University of Mississippi, Oxford, USA
J. G. Acosta, S. Oliveros

University of Nebraska-Lincoln, Lincoln, USA

State University of New York at Buffalo, Buffalo, USA
J. Dolen, A. Godshalk, I. Iashvili, A. Kharchilava, A. Kumar, S. Rappoccio

Northeastern University, Boston, USA

Northwestern University, Evanston, USA

University of Notre Dame, Notre Dame, USA

The Ohio State University, Columbus, USA

Princeton University, Princeton, USA
4: Also at National Institute of Chemical Physics and Biophysics, Tallinn, Estonia
5: Also at Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
6: Also at Universidade Estadual de Campinas, Campinas, Brazil
7: Also at Laboratoire Leprince-Ringuet, Ecole Polytechnique, IN2P3-CNRS, Palaiseau, France
8: Also at Joint Institute for Nuclear Research, Dubna, Russia
9: Also at Suez University, Suez, Egypt
10: Also at Cairo University, Cairo, Egypt
11: Also at Fayoum University, El-Fayoum, Egypt
12: Also at British University in Egypt, Cairo, Egypt
13: Now at Ain Shams University, Cairo, Egypt
14: Also at Université de Haute Alsace, Mulhouse, France
15: Also at Brandenburg University of Technology, Cottbus, Germany
16: Also at Institute of Nuclear Research ATOMKI, Debrecen, Hungary
17: Also at Eötvös Loránd University, Budapest, Hungary
18: Also at University of Debrecen, Debrecen, Hungary
19: Also at University of Visva-Bharati, Santiniketan, India
20: Now at King Abdulaziz University, Jeddah, Saudi Arabia
21: Also at University of Ruhuna, Matara, Sri Lanka
22: Also at Isfahan University of Technology, Isfahan, Iran
23: Also at University of Tehran, Department of Engineering Science, Tehran, Iran
24: Also at Plasma Physics Research Center, Science and Research Branch, Islamic Azad University, Tehran, Iran
25: Also at Laboratori Nazionali di Legnaro dell’INFN, Legnaro, Italy
26: Also at Università degli Studi di Siena, Siena, Italy
27: Also at Centre National de la Recherche Scientifique (CNRS)-IN2P3, Paris, France
28: Also at Purdue University, West Lafayette, USA
29: Also at Universidad Michoacana de San Nicolas de Hidalgo, Morelia, Mexico
30: Also at Institute for Nuclear Research, Moscow, Russia
31: Also at St. Petersburg State Polytechnical University, St. Petersburg, Russia
32: Also at National Research Nuclear University ‘Moscow Engineering Physics Institute’ (MEPhI), Moscow, Russia
33: Also at California Institute of Technology, Pasadena, USA
34: Also at Faculty of Physics, University of Belgrade, Belgrade, Serbia
35: Also at Facoltà Ingegneria, Università di Roma, Rome, Italy
36: Also at Scuola Normale e Sezione dell’INFN, Pisa, Italy
37: Also at University of Athens, Athens, Greece
38: Also at Paul Scherrer Institut, Villigen, Switzerland
39: Also at Institute for Theoretical and Experimental Physics, Moscow, Russia
40: Also at Albert Einstein Center for Fundamental Physics, Bern, Switzerland
41: Also at Gaziosmanpasa University, Tokat, Turkey
42: Also at Adiyaman University, Adiyaman, Turkey
43: Also at Cag University, Mersin, Turkey
44: Also at Anadolu University, Eskisehir, Turkey
45: Also at Ozyegin University, Istanbul, Turkey
46: Also at Izmir Institute of Technology, Izmir, Turkey
47: Also at Necmettin Erbakan University, Konya, Turkey
48: Also at Mimar Sinan University, Istanbul, Istanbul, Turkey
49: Also at Marmara University, Istanbul, Turkey
50: Also at Kafkas University, Kars, Turkey
51: Also at Yildiz Technical University, Istanbul, Turkey
52: Also at Rutherford Appleton Laboratory, Didcot, UK
53: Also at School of Physics and Astronomy, University of Southampton, Southampton, UK
54: Also at Faculty of Physics, Vinca Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia
55: Also at Argonne National Laboratory, Argonne, USA
56: Also at Erzincan University, Erzincan, Turkey
57: Also at Texas A&M University at Qatar, Doha, Qatar
58: Also at Kyungpook National University, Daegu, Korea