

**MEASUREMENT OF MULTIJET CROSS-SECTION RATIOS IN  
PROTON-PROTON COLLISIONS WITH THE CMS DETECTOR AT  
THE LHC**

A THESIS

Submitted to the  
FACULTY OF SCIENCE  
PANJAB UNIVERSITY, CHANDIGARH  
for the degree of

**DOCTOR OF PHILOSOPHY**

**2017**

**Anterpreet Kaur**

DEPARTMENT OF PHYSICS  
CENTRE OF ADVANCED STUDY IN PHYSICS  
PANJAB UNIVERSITY, CHANDIGARH  
INDIA



*Dedicated to  
my Grand-Parents*

*&*

*Parents*









# Contents

List of Figures	x
-----------------	---

List of Tables	xiii
----------------	------

---

1 Measurement of the Differential Inclusive Multijet Cross Sections and their Ratio	1
1.1 Data Samples . . . . .	2
1.1.1 Monte Carlo samples . . . . .	3
1.2 Event Selection . . . . .	4
1.2.1 Trigger Selection . . . . .	4
1.2.1.1 Primary vertex selection . . . . .	6
1.2.1.2 Jet Identification . . . . .	8
1.2.2 Pile-up reweighting . . . . .	9

List of publications	13
----------------------	----

Reprints	15
----------	----







# List of Figures

1.1	Trigger efficiencies turn-on curves for the single jet trigger paths used in the analysis. To determine the 99% efficiency threshold, the trigger turn-on curves are fitted using a sigmoid function taking into account the uncertainties using Clopper-Pearson confidence intervals. . . . .	7
1.2	Missing transverse energy fraction of the total transverse energy per event in data and simulated events. To remove background and noise, events with a fraction exceeding a certain threshold, here indicated with the dashed line, are rejected. . . . .	9
1.3	The fractions of jet constituents as observed in data and simulated events for different types of PF candidates for inclusive 2-jet events. Data and simulation are normalized to the same number of events. The distributions are shown after the application of the jet ID. . . . .	10
1.4	The fractions of jet constituents as observed in data and simulated events for different types of PF candidates for inclusive 3-jet events. Data and simulation are normalized to the same number of events. The distributions are shown after the application of the jet ID. . . . .	11
1.5	Number of reconstructed vertices in data and simulated events before (left) and after (right) the pile-up reweighting. . . . .	12



# List of Tables

1.1	Four data sets collected in run periods A, B,C and D during 2012, along with the corresponding run numbers and luminosity. . . . .	3
1.2	The official MC production samples generated in phase space slices in $H_T$ with the generator MADGRAPH5 and interfaced to PYTHIA6 for the parton shower and hadronization of the events. The cross section and number of events generated are mentioned for each sample. . . .	4
1.3	List of all single jet trigger paths used in the analysis. The column $H_{T,2}/2$ , 99% indicates the value of $H_{T,2}/2$ at which each trigger exhibits an efficiency larger than 99%. The last column reports the effective luminosity seen by each trigger. This number, divided by the total integrated luminosity of $19.71 \text{ fb}^{-1}$ , gives the effective prescale applied on a trigger over the whole run period. . . . .	5

# Chapter 1

## Measurement of the Differential Inclusive Multijet Cross Sections and their Ratio

The inclusive  $n$ -jet event samples include the events with number of jets  $\geq n$ , where  $n = 2$  and 3 in the current study. The inclusive multijet event yields are transformed into a differential cross section which is defined as :

$$\frac{d\sigma}{d(H_{T,2}/2)} = \frac{1}{\epsilon \mathcal{L}_{\text{int,eff}}} \frac{N_{\text{event}}}{\Delta(H_{T,2}/2)} \quad (1.1)$$

where  $N_{\text{event}}$  is the number of inclusive 2- or 3-jet events counted in an  $H_{T,2}/2$  bin,  $\epsilon$  is the product of the trigger and jet selection efficiencies, which are greater than 99%,  $\mathcal{L}_{\text{int,eff}}$  is the effective integrated luminosity, and  $\Delta(H_{T,2}/2)$  are the bin widths. The measurements are reported in units of (pb/GeV).

The differential inclusive multijet cross sections are measured as a function of the average transverse momentum,  $H_{T,2}/2 = \frac{1}{2}(p_{T,1} + p_{T,2})$ , where  $p_{T,1}$  and  $p_{T,2}$  denote the transverse momenta of the two leading jets. The cross section ratio  $R_{32}$ , defined in Eq. 1.2 is obtained by dividing the differential cross sections of inclusive

3-jet events to that of inclusive 2-jet one, for each bin in  $H_{T,2}/2$ .

$$R_{32} = \frac{\frac{d\sigma_{3-jet}}{d(H_{T,2}/2)}}{\frac{d\sigma_{2-jet}}{d(H_{T,2}/2)}} \quad (1.2)$$

For inclusive 2-jet events ( $n_j \geq 2$ ) sufficient data are available up to  $H_{T,2}/2 = 2 \text{ TeV}$ , while for inclusive 3-jet events ( $n_j \geq 3$ ) and the ratio  $R_{32}$ , the accessible range in  $H_{T,2}/2$  is limited to  $H_{T,2}/2 < 1.68 \text{ TeV}$ .

## 1.1 Data Samples

This measurement uses the data collected at the center of mass energy of 8 TeV by CMS experiment in the 2012 run period of the LHC. The 2012 data is taken in four periods A, B, C, D and the data sets are divided into samples according to the run period. Further each sample is grouped into subsets based on the trigger decision. For run B-D, the **JetMon** stream datasets contain prescaled low trigger threshold paths (HLTPFJet40, 80, 140, 200 and 260) while the **JetHT** stream datasets contain unprescaled high threshold trigger paths (HLT PFJet320 and 400). For run A, the **Jet** stream contains all the above mentioned trigger paths. The data to be used in physics analysis must satisfy a certain criteria which include proper performance of all detector subsystems as well as the passing of data quality monitoring (DQM) steps during the validation process. CMS uses JSON (Java Script Object Notation) format files to store the range of good lumi sections within a run. In the current analysis, the applied certification file<sup>1</sup> is based on the final event reconstruction of the 2012 CMS data sets. The datasets used in the current study are mentioned in the Table 1.1 along with the luminosity of each dataset.

---

<sup>1</sup>Cert\_190456-208686\_8TeV\_22Jan2013ReReco\_Collisions12\_JSON

Table 1.1: Four data sets collected in run periods A, B,C and D during 2012, along with the corresponding run numbers and luminosity.

Run	Run range	Data set	Luminosity $\text{fb}^{-1}$
A	190456-193621	/Jet/Run2012A-22Jan2013-v1/AOD	0.88
B	193834-196531	/Jet[Mon,HT]/Run2012B-22Jan2013-v1/AOD	4.41
C	198022-203742	/Jet[Mon,HT]/Run2012C-22Jan2013-v1/AOD	7.06
D	203777-208686	/Jet[Mon,HT]/Run2012D-22Jan2013-v1/AOD	7.37

The data sets have the LHC luminosity increasing with period, full data sample of 2012 corresponds to an integrated luminosity of  $19.71 \text{ fb}^{-1}$ .

### 1.1.1 Monte Carlo samples

To have a comparison of data results with the simulated events, the MADGRAPH5 [1] Monte-Carlo event generator has been used. The MADGRAPH5 generates matrix elements for High Energy Physics processes, such as decays and  $2 \rightarrow n$  scatterings. The underlying event is modeled using the tune Z2\*. It has been interfaced to PYTHIA6 [2] by the LHE event record [3], which generates the rest of the higher-order effects using the Parton Showering (PS) model. Matching algorithms ensure that no double-counting occurs between the tree-level and the PS-model-generated partons. The MC samples are processed through the complete CMS detector simulation to allow studies of the detector response and compare to measured data on detector level.

The cross section measured as a function of the transverse momentum  $p_T$  or the scalar sum of the transverse momentum of all jets  $H_T$  falls steeply with the increasing  $p_T$ . So in the reasonable time, it is not possible to generate a large number of high  $p_T$  events. Hence, the events are generated in the different phase-space region binned in  $H_T$  or the leading jet  $p_T$ . Later on, the different phase-space regions are added together in the data analyses by taking into account the cross section of the different phase-space regions. The official CMS MADGRAPH5 + PYTHIA6 MC samples used in



this analysis are generated as slices in the  $H_T$  phase-space are tabulated in Table 1.2 along with their cross sections and number of events generated.

Table 1.2: The official MC production samples generated in phase space slices in  $H_T$  with the generator MADGRAPH5 and interfaced to PYTHIA6 for the parton shower and hadronization of the events. The cross section and number of events generated are mentioned for each sample.

Generator	Sample	Events	Cross Section pb
MADGRAPH5 + PYTHIA 6	/QCD_HT-100To250_TuneZ2star_8TeV-madgraph-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	50129518	$1.036 \times 10^7$
	/QCD_HT-250To500_TuneZ2star_8TeV-madgraph-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	27062078	$2.760 \times 10^5$
	/QCD_HT-500To1000_TuneZ2star_8TeV-madgraph-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	30599292	$8.426 \times 10^3$
	/QCD_HT-1000ToInf_TuneZ2star_8TeV-madgraph-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	13843863	$2.040 \times 10^2$

## 1.2 Event Selection

To yield a multijet sample with high purity and high selection efficiency, the events are selected according to several quality criteria. This event selection also reduces beam induced background, detector-level noise and jets arising from fake calorimeter energy deposits.

### 1.2.1 Trigger Selection

In this analysis single jet trigger paths are used which consists of one L1 trigger seed and multiple HLT filters and are tabulated in Table 1.3. HLT\_PFJetX implies that there is at-least one jet in the event, whose  $p_T > X$  (GeV). The L1 trigger has a lower threshold to ensure full efficiency versus  $p_T$  of the HLT trigger. The  $p_T$  spectrum is steeply falling and hence the rates for low- $p_T$  jets are very high. So it is not feasible to use a single unscaled trigger for the selection of all required events. To collect sufficient data in the lower part of the  $p_T$  spectrum, different five prescaled low- $p_T$  trigger paths, each with different prescale value, are used. Also,

one unprescaled trigger i.e. HLT\_Jet320 is used in the high  $p_T$  region, in which the rate is sufficiently small to collect and store all events.

Table 1.3: List of all single jet trigger paths used in the analysis. The column  $H_{T,2}/2$ , 99% indicates the value of  $H_{T,2}/2$  at which each trigger exhibits an efficiency larger than 99%. The last column reports the effective luminosity seen by each trigger. This number, divided by the total integrated luminosity of  $19.71 \text{ fb}^{-1}$ , gives the effective prescale applied on a trigger over the whole run period.

Trigger Path	L1 threshold GeV	HLT threshold GeV	$H_{T,2}/2$ , 99% GeV	Eff. Lumi $\text{fb}^{-1}$
HLT_PFJet80	36	80	120.0	$0.21 \times 10^{-2}$
HLT_PFJet140	68	140	187.5	$0.56 \times 10^{-1}$
HLT_PFJet200	92	200	262.5	0.26
HLT_PFJet260	128	260	345.0	1.06
HLT_PFJet320	128	320	405.0	19.71

The efficiency of each trigger, as a function of the measured observable, is described by the turn-on curves with a rising part, where the trigger is partly inefficient, until a plateau region, corresponding to the region of full efficiency of the trigger. Hence it is necessary to determine the threshold above which a trigger becomes fully efficient. It is defined as the value at which the efficiency exceeds 99%. In the assumption that the reference trigger HLT\_PFJetX is fully efficient in the considered region of the phase space, the trigger efficiency for HLT\_PFJetY is defined as Eq. 1.3. The value of X is chosen previous to that of Y in  $p_T$  ordering from the trigger list so that the higher trigger condition can be emulated from the lower trigger path.

$$\epsilon_{\text{HLT\_PFJetY}} = \frac{H_{T,2}/2 \left( \text{HLT\_PFJetX} + (\text{L1Object\_p}_T > Z) + (\text{HLTObject\_p}_T > Y) \right)}{H_{T,2}/2(\text{HLT\_PFJetX})} \quad (1.3)$$

where Y indicates the  $p_T$  threshold of HLT\_PFJetY and Z is the L1 seed value corresponding to the trigger path HLT\_PFJetY. The denominator represents the number of events for which the reference trigger path HLT\_PFJetX has been fired.

The numerator is the number of events for which HLT\_PFJetX has been fired along the  $p_T$  of L1Object  $\geq Z$  and the  $p_T$  of HLTOBJECT  $\geq Y$ . For example, in order to obtain turn-on curve for HLT\_PFJet260, the reference HLT path HLT\_PFJet200 is chosen, the  $p_T$  cut on L1Object is 128 GeV and  $p_T$  cut on HLTOBJECT is 260 GeV. The uncertainty on the efficiency is indicated by error bars which represent Clopper-Pearson confidence intervals [4]. To determine the point, at which the trigger efficiency is larger than 99%, the turn-on distribution is fitted using a sigmoid function described in Eq. 1.4. The trigger turn-on curves as a function of  $H_{T,2}/2$  can be seen in Figure 1.1.

$$f_{fit}(x) = \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{x - \mu}{\sqrt{2}\sigma} \right) \right) \quad (1.4)$$

### 1.2.1.1 Primary vertex selection

A primary vertex (PV) is identified by a collection of tracks, measured in the tracker with a good fit quality between the hits and compatible with the beam line. The tracks are clustered according to the z-coordinate of their point of closest approach to the beam axis. Each event is required to have at least one good primary vertex which is well reconstructed within a distance of  $|z(PV)| < 24$  cm to the nominal interaction point of the detector.

- Require the radius in x-y plane of PV to be  $\rho < 2$  cm.
- Require the vertex fit no. of degrees of freedom (ndf)  $> 4$ . Thus, at least four tracks must be present in order to perform a valid vertex fit.

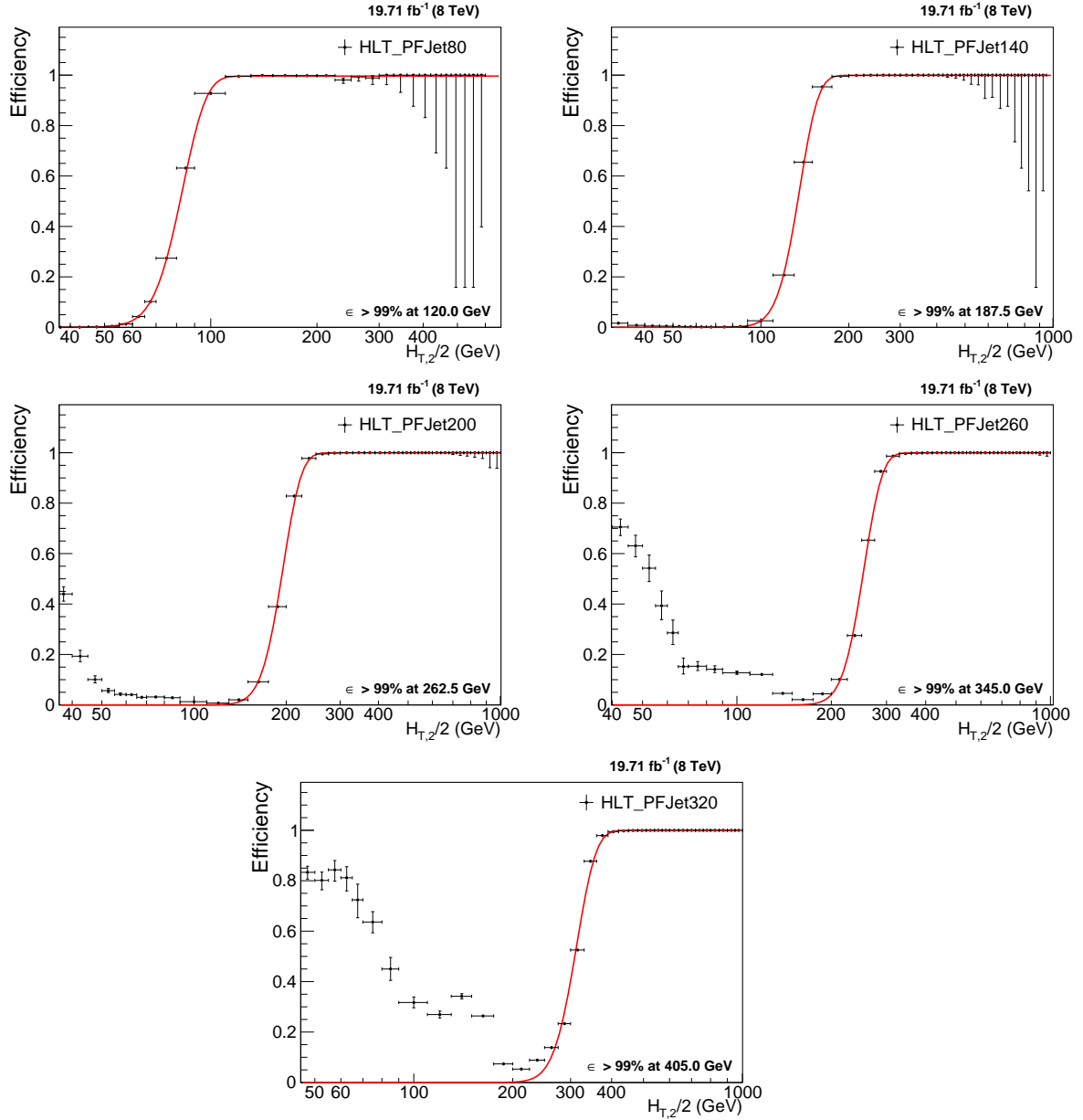


Figure 1.1: Trigger efficiencies turn-on curves for the single jet trigger paths used in the analysis. To determine the 99% efficiency threshold, the trigger turn-on curves are fitted using a sigmoid function taking into account the uncertainties using Clopper-Pearson confidence intervals.

### 1.2.1.2 Jet Identification

In order to suppress noise and non-physical jets, jet identification criteria (ID) has been applied. The official tight jet ID has been used which are based on the number of constituents inside the reconstructed jets. Instead of applying it event-wise, it is applied it on each jet and the following cuts need to be fulfilled by the jets :

- Each jet should contain at least two particles, one of which should be a charged hadron.
- The fraction of neutral hadrons and photons should be  $< 0.90$  to remove HCAL noise and ECAL noise, respectively.
- Muons that are falsely identified and clustered as jets are removed by requiring the muon fraction  $< 0.80$
- Based on information of the tracker, additional selection cuts are enforced in the region  $\eta < 2.4$ . Jets clustered from misidentified electrons are removed with the condition on charged electromagnetic fraction to be  $< 0.90$ . Also, the fraction of charged hadrons in the jet must be larger than zero.

Finally the event selection is as follows :

- All jets having  $p_T > 150$  GeV and  $|y| < 5$  are selected.
- Events with at least two jets are selected.
- The two leading jets should have  $|y| < 2.5$  and further jets are counted only, if they lie within the same central rapidity range of  $|y| < 2.5$ .
- In QCD, pure jet events are balanced in  $p_T$  and thus exhibit a low level of missing transverse energy, which predominantly is caused by jet calibration and resolution effects of the detector. Therefore, the ratio of missing transverse energy to the total transverse energy  $\frac{E_T^{\text{miss}}}{\sum E_T}$ , both derived from the whole event

information, is required to be less than 0.3 to select well measured jet events, as shown in Figure 1.2.

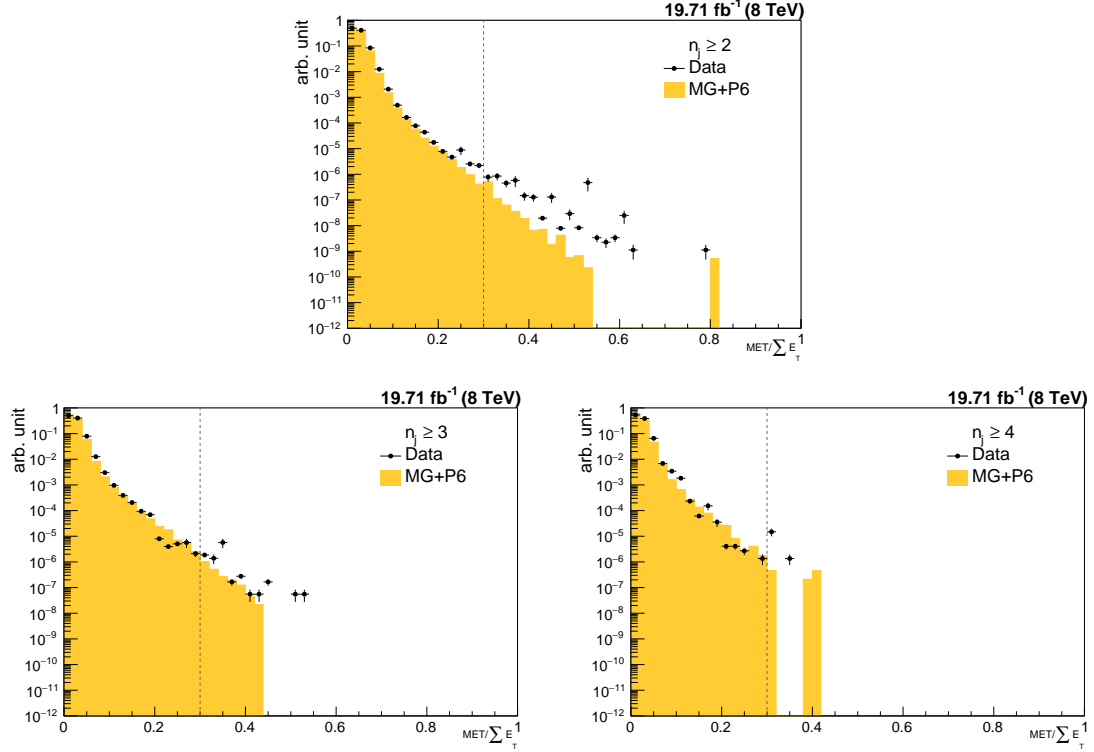


Figure 1.2: Missing transverse energy fraction of the total transverse energy per event in data and simulated events. To remove background and noise, events with a fraction exceeding a certain threshold, here indicated with the dashed line, are rejected.

The Figures 1.3- ?? show the distributions of the jet constituents observed in data and simulated events for inclusive 2-jet, 3-jet and 4-jet events respectively.

### 1.2.2 Pile-up reweighting

The official Monte-Carlo samples are generated with distributions for the number of pileup interactions which is meant to roughly cover, though not exactly match the conditions expected for each data-taking period. To still get comparable pile-up distributions in data and simulated events, the simulated events are reweighted with

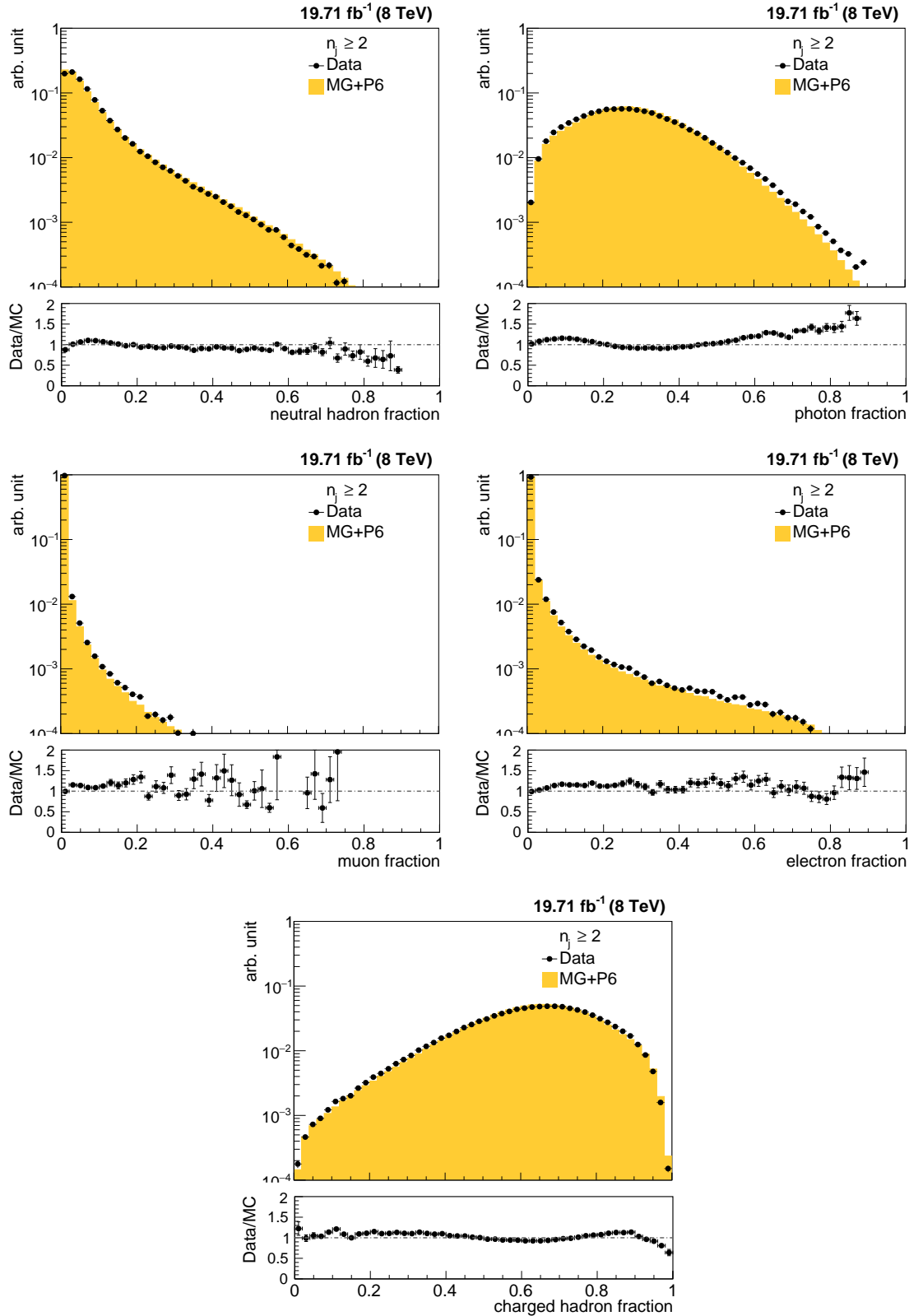


Figure 1.3: The fractions of jet constituents as observed in data and simulated events for different types of PF candidates for inclusive 2-jet events. Data and simulation are normalized to the same number of events. The distributions are shown after the application of the jet ID.

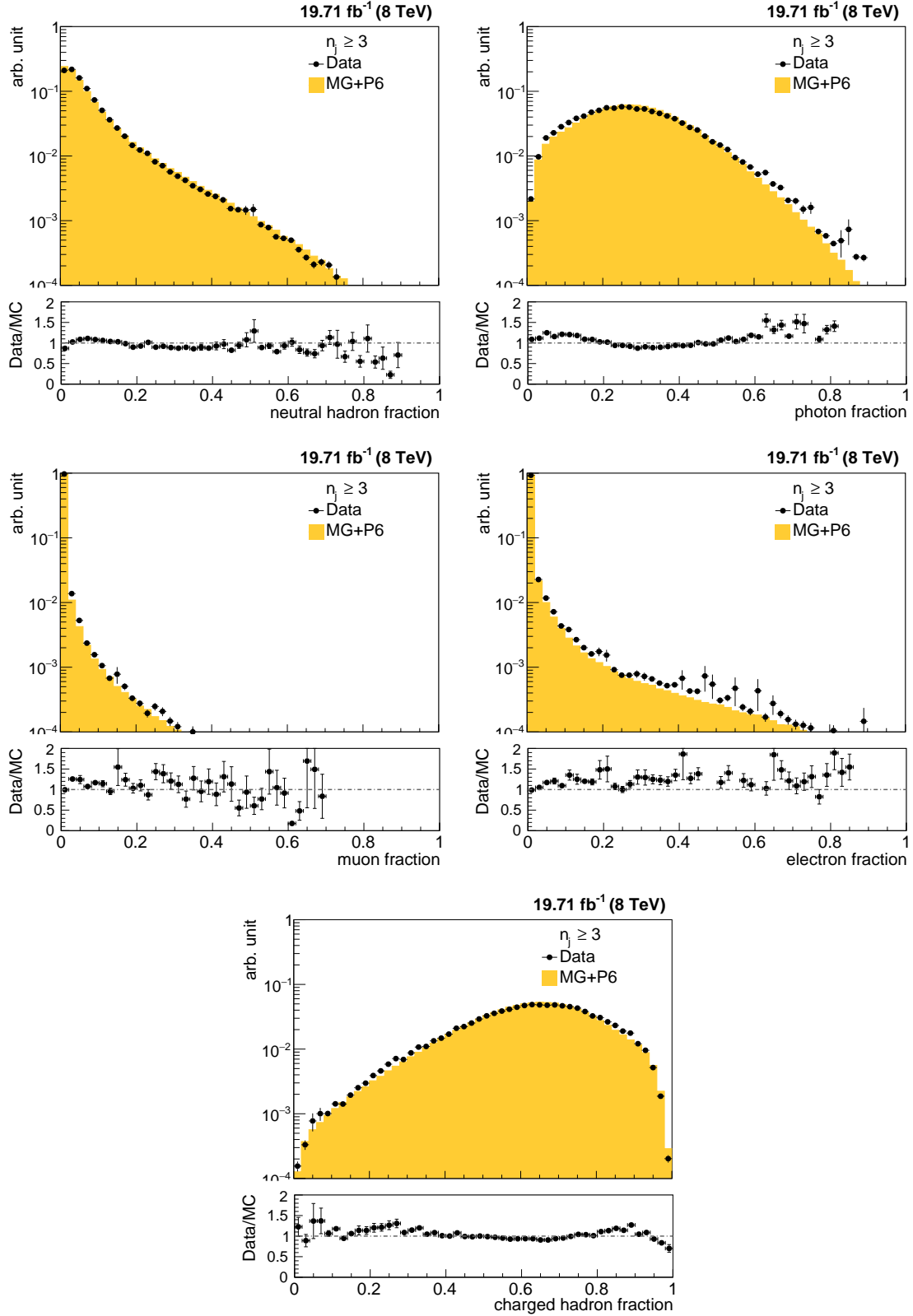


Figure 1.4: The fractions of jet constituents as observed in data and simulated events for different types of PF candidates for inclusive 3-jet events. Data and simulation are normalized to the same number of events. The distributions are shown after the application of the jet ID.



a weight  $w_{\text{PU}}$  to match the distribution in data :

$$w_{\text{PU}} = \frac{N_{\text{data}}(N_{\text{PU, est.}}) / \sum N_{\text{data}}}{N_{\text{MC}}(N_{\text{PU, truth}}) / \sum N_{\text{MC}}} \quad (1.5)$$

Figure 1.5 shows the number of reconstructed vertices before and after reweighting. The significant mismatch of the pile-up distributions in data and simulated events, which is present before the reweighting, completely vanishes.

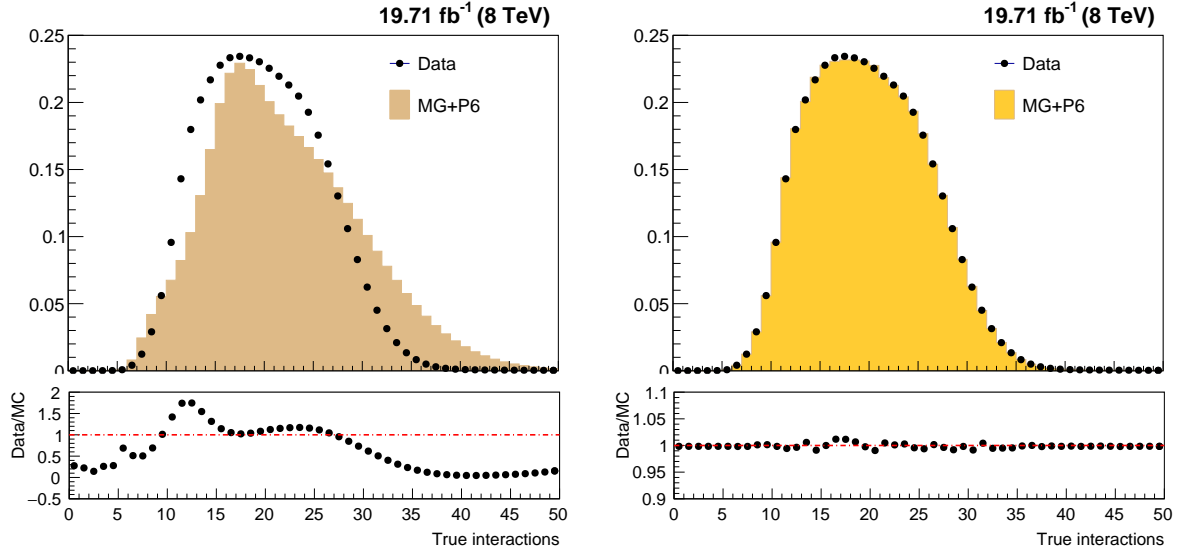


Figure 1.5: Number of reconstructed vertices in data and simulated events before (left) and after (right) the pile-up reweighting.

# Bibliography

- [1] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, “MadGraph 5 : Going Beyond,” *JHEP*, vol. 06, p. 128, 2011.
- [2] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual,” *JHEP*, vol. 05, p. 026, 2006.
- [3] J. Alwall *et al.*, “A Standard format for Les Houches event files,” *Comput. Phys. Commun.*, vol. 176, pp. 300–304, 2007.
- [4] C. J. Clopper and E. S. Pearson, “The use of confidence or fiducial limits illustrated in the case of the binomial,” *Biometrika*, vol. 26, no. 4, pp. 404–413, 1934.



*Selected*  
*Reprints*

