STUDY OF HADRONIC JET STRUCTURE AND SUBJET MULTIPLICITY IN p-p COLLISIONS AT LHC

A THESIS

Submitted to the

FACULTY OF SCIENCE

PANJAB UNIVERSITY, CHANDIGARH

for the degree of

DOCTOR OF PHILOSOPHY

2011

Manuk Zubin Mehta

DEPARTMENT OF PHYSICS

CENTRE OF ADVANCED STUDY IN PHYSICS

PANJAB UNIVERSITY, CHANDIGARH

INDIA

Dedicated to

my Grand-Parents

&

Parents

Contents

List of Figures

List of Tables

Chapter 1

Theory predictions

Predictions at NLO accuracy in pQCD are computed with the NLOJET++ program version 4.1.3 [?,?]. The results are provided within the framework of FASTNLO version 2.3 [?] for use within fits. The renormalization and factorization scales μ_r and μ_f are chosen equal to $H_{\mathrm{T},2}/2$. PDF sets at NLO available for a series of different assumptions on $\alpha_s(M_Z)$ via the LHAPDF6 package [?] are listed in Table 2.7. All sets employ a variable-flavour number scheme with at most five or six flavours apart from the ABM11 PDFs, which use a fixed-flavour number scheme with $N_F = 5$.

Out of these eight PDF sets the following three will not be considered further:

- At NLO, predictions based on ABM11 do not describe LHC jet data at small jet rapidity, cf. Refs. [?, ?, ?, ?].

- The HERAPDF2.0 set exclusively fits HERA DIS data with only weak constraints on the gluon PDF.

- The range in values available for $\alpha_s(M_Z)$ is too limited for the NNPDF3.0 set.

PDF uncertainties are evaluated according to the prescriptions given for each PDF set. Uncertainties on $\alpha_s(M_Z)$ are not considered, since this value is later on determined from a fit to the data. The PDF uncertainty as derived with the CT10

PDF set ranges from 2% to 30% for inclusive 2- and 3-jet cross sections and from 2% to 7% for R_{32}.

Table 1.1: NLO PDF sets available via LHAPDF6 for comparisons to data with various assumptions on the value of $\alpha_s(M_Z)$. Sets existing already in LHC Run 1 (upper rows) and newer sets for Run 2 (lower rows) are listed together with the corresponding number of flavours N_F, the assumed masses M_t and M_Z of the top quark and the Z boson, respectively, the default values of $\alpha_s(M_Z)$, and the range in $\alpha_s(M_Z)$ variation available for fits. A [*] behind the $\alpha_s(M_Z)$ values signifies that the parameter was fixed, not fitted.

Base set	Refs.	N_F	M_t ()	M_Z ()	$\alpha_s(M_Z)$	$\alpha_s(M_Z)$ range
ABM11	[?]	5	180	91.174	0.1180	0.110–0.130
CT10	[?]	≤5	172	91.188	0.1180[*]	0.112–0.127
MSTW2008	[?, ?]	≤5	10^{10}	91.1876	0.1202	0.110–0.130
NNPDF2.3	[?]	≤6	175	91.1876	0.1180[*]	0.114–0.124
CT14	[?]	≤5	172	91.1876	0.1180[*]	0.113–0.123
HERAPDF2.0	[?]	≤5	173	91.1876	0.1180[*]	0.110–0.130
MMHT2014	[?]	≤5	10^{10}	91.1876	0.1180[*]	0.108–0.128
NNPDF3.0	[?]	≤5	173	91.2	0.1180[*]	0.115–0.121

The uncertainty related to unknown higher orders of the perturbative series is evaluated with the conventional recipe of varying the default scale $H_{\mathrm{T},2}/2$ chosen for μ_r and μ_f independently in the following six combinations: $(\mu_r/H_{\mathrm{T},2}/2,\ \mu_f/H_{\mathrm{T},2}/2)$ = (1/2,1/2), (1/2,1), (1,1/2), (1,2), (2,1) and (2,2). The maximal upwards and downwards deviations in cross section from the central prediction are taken as scale uncertainty. This uncertainty ranges for inclusive 2-jet events from 5% to 13%, for inclusive 3-jet events from 11% to 17% and for their ratio R_{32} from 6% to 8%.

The computation of the NLO predictions with NLOJET++ is also subject to statistical fluctuations from the numerical integrations. For the inclusive 2-jet event cross sections this uncertainty is smaller than about one per mille, while for the inclusive 3-jet event cross section it amounts to 1–9 per mille.

Higher order effects of electroweak origin affect jet cross sections at large jet p_{T}. These electroweak (EWK) corrections have been calculated for the inclusive 1-jet and 2-jet case, cf. Ref. [?], but are not yet known for 3-jet production. Therefore,

they are considered for the 2-jet events, while for the 3-jet event cross section and for the ratio they have been neglected.

The impact of nonperturbative (NP) effects, i.e. from multiple-parton interactions (MPI) and hadronization, are evaluated by using samples obtained from different MC event generators with a simulation of parton-shower and underlying-event (UE) contributions. The leading order (LO) MC event generators HERWIG++ [?] with the default tune of version 2.3 and PYTHIA6 [?] with tune Z2* are considered, and the dijet NLO prediction from POWHEG [?,?,?] interfaced to PYTHIA8 with tune CUETS1 [?] for full event generation. The cross section ratios between a nominal event generation and a sample without hadronization and MPI effects are taken as correction separately for inclusive 2-, and 3-jet events, and as their ratio for R_{32}. This ratio is fitted by a power-law function. The differences in the correction factors obtained from the various MC event generators are assigned as an uncertainty. The central correction factors C^{NP} are determined by the centre of the envelope which covers all predictions and half of the spread is taken as the uncertainty.

The NP corrections are shown in Fig. 2.29 for the inclusive 2-jet (top) and 3-jet event cross sections (top) as well for R_{32} (bottom). They amount to $\approx$ 4–5% for inclusive 2-jet and 3-jet events and $\approx$ 1% for R_{32} at $H_{\mathrm{T},2}/2 \approx 300\,\mathrm{GeV}$ and decrease for increasing $H_{\mathrm{T},2}/2$. The uncertainty assigned to the NP corrections is of the order of 1–2%. The non-perturbative effects are reduced in the cross section ratio.

The total theoretical uncertainties are evaluated as the quadratic sum of the scale, PDF, NP, and statistical uncertainties. Figure 2.31 presents an overview of the theoretical uncertainties affecting the cross section prediction for inclusive 2-jet (top) and 3-jet events (top) and their ratio R_{32} (bottom), using the CT10 PDF set.
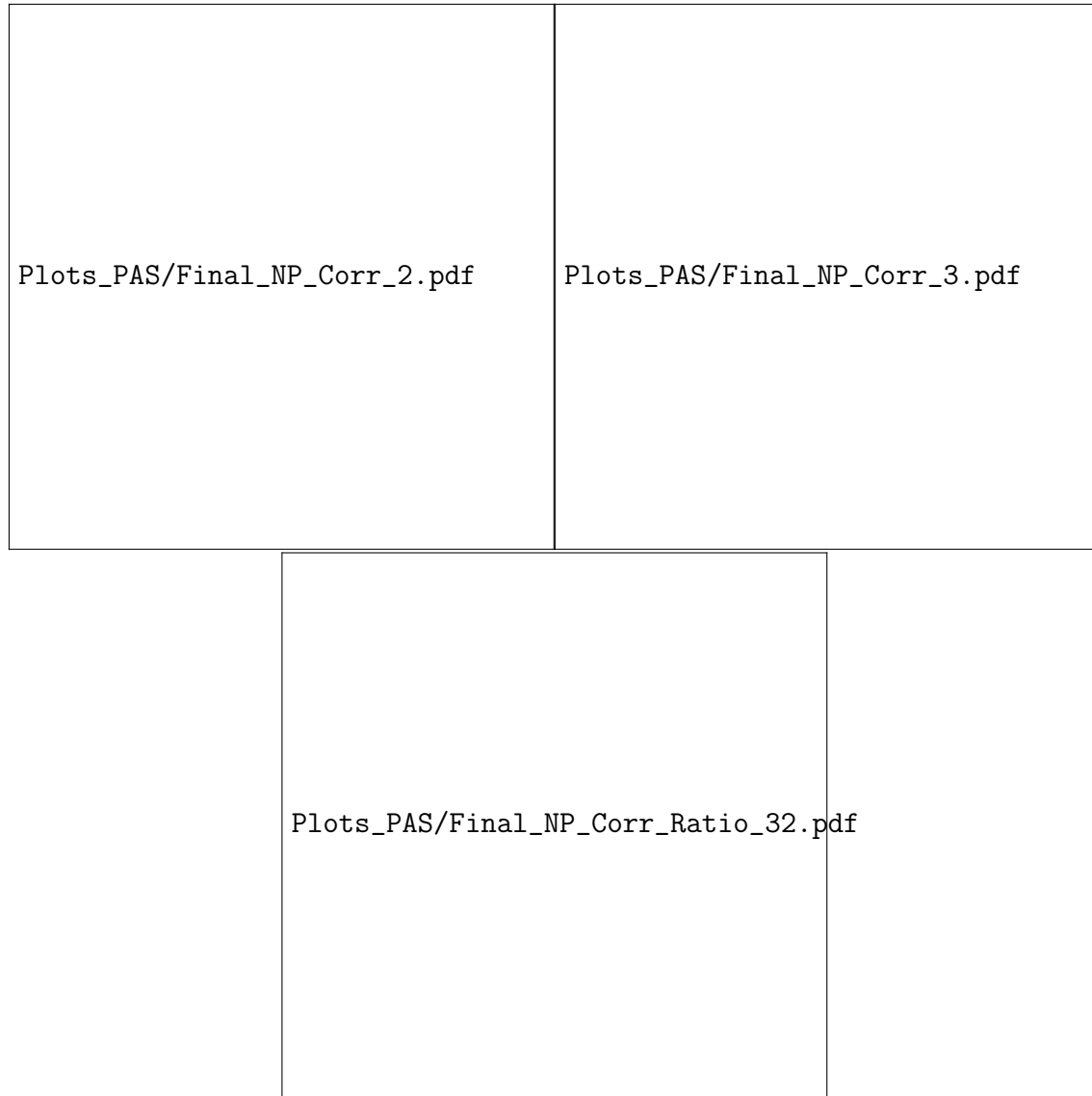
Figure 1.1: Fits to the nonperturbative corrections obtained for inclusive 2-jet (top) and 3-jet (top) event cross sections and their ratio R_{32} (bottom) as a function of $H_{\mathrm{T},2}/2$ within $|y| < 2.5$ for the three investigated MC event generators.

Figure 1.2: Overview of theoretical uncertainties affecting the cross section prediction for inclusive 2-jet (top) and 3-jet events (top) and their ratio R_{32} (bottom), using the CT10 PDF set. The total uncertainty is calculated by adding in quadrature the individual sources of uncertainty. The statistical uncertainties of the NLO computations are too small to be visible and are not shown.

Chapter 2

Measurement of the Inclusive Differential Multijet Cross Section

The inclusive differential multijet cross sections are measured as a function of the average transverse momentum, $H_{\mathrm{T},2}/2 = \frac{1}{2}(p_{\mathrm{T},1} + p_{\mathrm{T},2})$, where $p_{\mathrm{T},1}$ and $p_{\mathrm{T},2}$ denote the transverse momenta of the two leading jets.

2.1 Cross Section Definition

The inclusive mutijet event yields are transformed into a differential cross section which is defined as :

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\big(H_{\mathrm{T},2}/2\big)} = \frac{1}{\epsilon \, \mathcal{L}_{\mathrm{int,eff}}} \frac{N_{\mathrm{event}}}{\Delta\big(H_{\mathrm{T},2}/2\big)} \tag{2.1}$$

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

For inclusive 2-jet events sufficient data are available up to $H_{\mathrm{T},2}/2 = 2\,\mathrm{TeV}$, while for inclusive 3-jet events (and the ratio R_{32}) the accessible range in $H_{\mathrm{T},2}/2$ is limited to $H_{\mathrm{T},2}/2 < 1.68\,\mathrm{TeV}$. In the following, results for the inclusive 2-jet and 3-jet event selections will be labelled as $n_j \geq 2$ and $n_j \geq 3$, respectively.

2.2 Data Samples

During 2012, CMS collected data at the center of mass energy $\sqrt{s} = 8$ TeV in four periods A, B, C, D. The datasets are divided into samples according to the run period. For run B-D, the `JetMon` stream datasets contain prescaled low trigger threshold paths (HLT PFJet40, 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 datasets used in the current study are mentioned in the Table 2.1 along with the luminosity of each dataset :

Table 2.1: Datasets used along with the corresponding run numbers and luminosity.

Run	Run Range	Dataset	Luminosity
A	190456-193621	/Jet/Run2012A-22Jan2013-v1/AOD	0.88 fb^{-1}
B	193834-196531	/Jet[Mon,HT]/Run2012B-22Jan2013-v1/AOD	4.49 fb^{-1}
C	198022-203742	/Jet[Mon,HT]/Run2012C-22Jan2013-v1/AOD	7.06 fb^{-1}
D	203777-208686	/Jet[Mon,HT]/Run2012D-22Jan2013-v1/AOD	7.37 fb^{-1}

The data sets have the LHC luminosity increasing with period, full data sample of 2012 corresponds to an integrated luminosity of 19.7 fb^{-1}.

2.2.1 Monte Carlo samples

To have a comparison of data results with the simulated events, the MADGRAPH5 [?] 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 [**?**] by the LHE event record, 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 were generated as slices in the H_T phase-space :

- The MADGRAPH5 + PYTHIA6 MC : /QCD_HT-xxxtoxxxx_TuneZ2star_8TeV-madgraph-pythia6/
  Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM

2.3   Event Selection

To select good events, certain event selection criteria is applied to yield a multijet event sample with high selection efficiency. It also reduces beam induced background, detector-level noise and jets arising from fake calorimeter energy deposits. The event selection cuts for each event are following -

2.3.1   Trigger Selection

The CMS Trigger and Data Acquisition System (TriDAS) is designed to inspect the detector information at the full crossing frequency and to select events at a maximum rate of $O(10^2)$ Hz for archiving and later offline analysis. The required rejection rate of $O(10^5)$ Hz is too large to be achieved in a single processing step, if a high efficiency is to be maintained for the physics phenomena, CMS plans to study. For this reason, the full selection task is split into two steps : Level-1 Trigger, (L1 - hardware trigger) and Higher Level Trigger, (HLT - software trigger).

In this analysis the trigger paths which are used are single jet triggers which consists of one L1 trigger seed and multiple HLT filters. The current 2012 data samples contain several jet triggers at the level of HLT which are presented in Table 2.2 mentioning the trigger thresholds, along with the corresponding L1 trigger seeds. Also the trigger thresholds vs p_{T} [?] are shown. As expected, $H_{\mathrm{T},2}/2$ thresholds ($|y| < 2.5$) are comparable but lower than p_{T} thresholds ($|y| < 3.0$). The prescaled low p_{T} triggers having different prescales are used to have sufficient data in the lower part of the p_{T} spectrum. In addition unprescaled trigger HLT PFJet320 is also used in the region where the jet rate is sufficiently small to collect all the events. During the reconstruction of the spectrum, the prescales have been taken into the account.

Table 2.2: HLT trigger thresholds and corresponding L1SingleJet trigger.

HLT Path	PFJet80	PFJet140	PFJet200	PFJet260	PFJ
L1 seed	L1SingleJet36	L1SingleJet68	L1SingleJet92	L1SingleJet128	L1Singl
$H_{\mathrm{T},2}/2$, 99% (GeV)	120.0	187.5	262.5	345.0	40
$H_{\mathrm{T},2}/2$ range (GeV)	$120 - 188$	$188 - 263$	$263 - 345$	$345 - 406$	$405 -$
p_{T} range (GeV)	$133 - 220$	$220 - 300$	$300 - 395$	$395 - 507$	$507 -$
Integrated lumi (pb^{-1})	2.12	5.57 x 10	2.61 x 10^2	1.06 x 10^3	1.97

In order to obtain the lower threshold with maximum efficiency of the jet triggers we produce trigger turn on curves for each HLT trigger paths. The efficiency of each trigger is not calculated as a function of jet p_{T}, which was used in the trigger

decision, but versus $H_{\mathrm{T},2}/2$. The triggers show a turn-on behaviour which as can been seen in Figure 2.1.
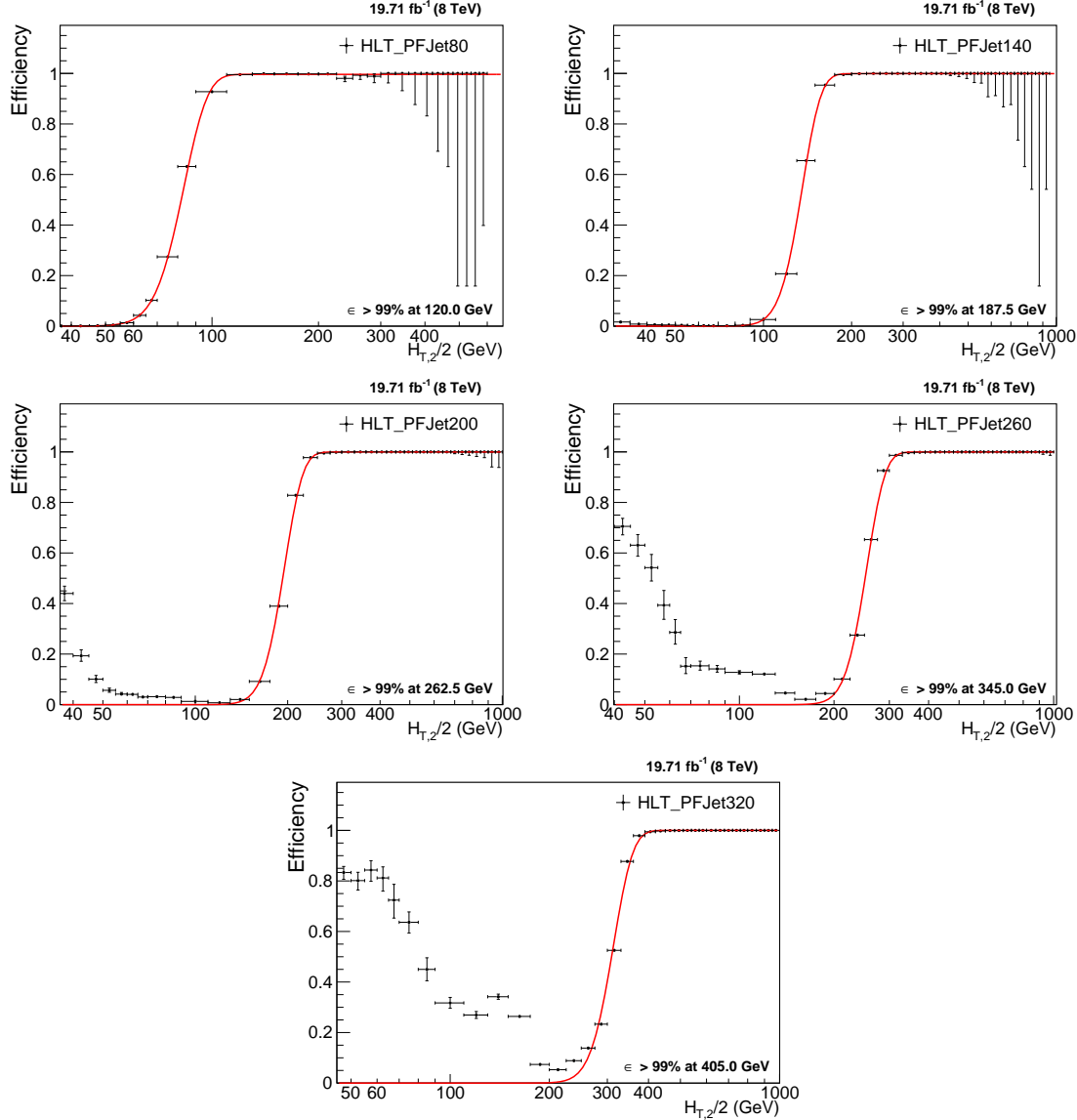
Figure 2.1: Trigger efficiencies turn-on curves for the single jet trigger paths used in the analysis.

The trigger efficiency for HLT_PFJetY is defined as :

$$\mathrm{HLT\_PFJet}_{\mathrm{eff}}\mathrm{Y} = \frac{\mathrm{H}_{\mathrm{T},2}/2(\mathrm{HLT\_PFJetX} + \mathrm{L1Object\_p_T} > \mathrm{Z} + \mathrm{HLTObject\_p_T} > \mathrm{Y})}{\mathrm{H}_{\mathrm{T},2}/2(\mathrm{HLT\_PFJetX})}$$

$$(2.2)$$

Here the denominator represents the number of events for which the trigger path HLT_PFJetX has been fired. Here 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. The numerator is the number of events for which HLT_PFJetX has been fired and the p_T of L1Object corresponding to the trigger path HLT_JetX is $\geq$ Z (where Z is the L1 seed value corresponding to the trigger path HLT_PFJetY) and the p_T of HLTObject corresponding to the trigger path HLT JetX is $\geq$ Y. For example, in order to obtain turn on curve for HLT_PFJet260, the immediate HLT path of lower threshold HLT_PFJet200 is chosen, the p_T cut on L1Object corresponding to the trigger path HLT PFJet260 is 128 GeV and p_T cut on HLTObject corresponding to the trigger path HLT Jet260 is 260 GeV. The uncertainty on the efficiency is indicated by error bars which represent Clopper-Pearson confidence intervals. To determine the point, at which the trigger efficiency is larger than 99%, the turn-on distribution is fitted using a sigmoid function, that describes the turn-on behaviour of the trigger paths.

$$f_{fit}(x) = \frac{1}{2}\left(1 + erf\left(\frac{x - \mu}{\sqrt{2}\sigma}\right)\right) \tag{2.3}$$

2.3.1.1   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 $\left|z(PV)\right|$ < 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.

2.3.1.2 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 $p_{\mathrm{T}} > 150$ 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 $\frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sum E_{\mathrm{T}}}$, 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 2.2.
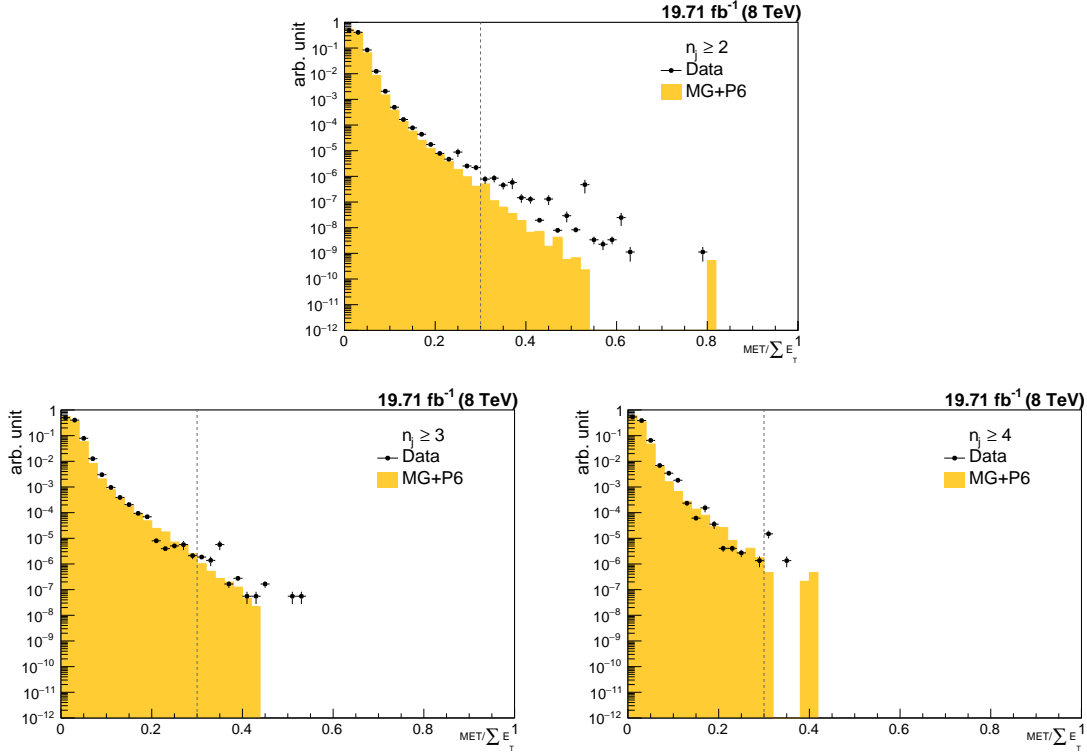
Figure 2.2: Missing transverse energy fraction of the total transverse energy per event in data and simulated events. To remove background and noise, events with a fraction exceeding a certain threshold, here indicated with the dashed line, are rejected.

The Figures 2.3- **??** show the distributions of the jet constituents observed in data and simulated events for inclusive 2-jet, 3-jet and 4-jet events respectively.

2.3.2 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
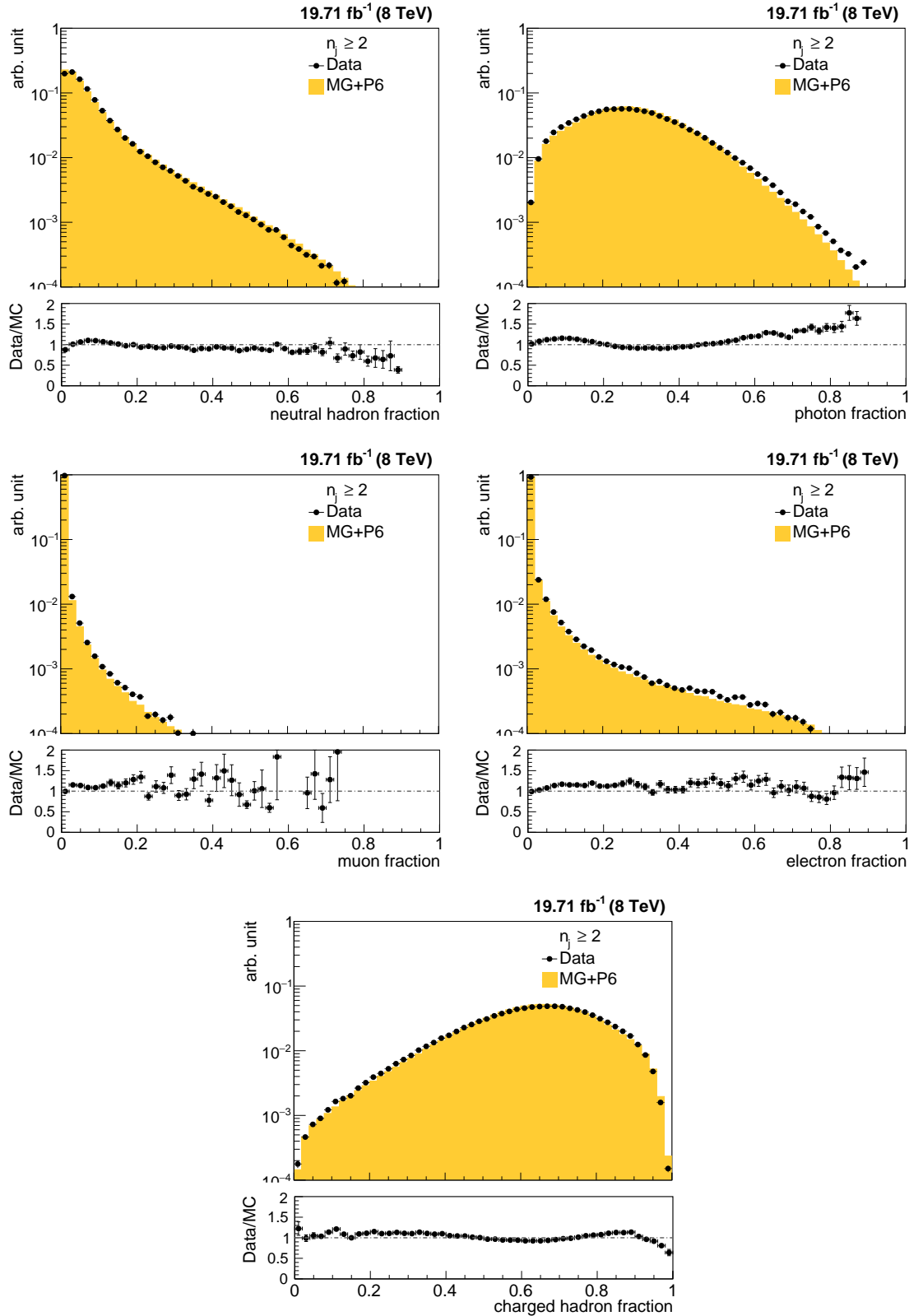
Figure 2.3: The fractions of jet constituents as observed in data and simulated events for different types of PF candidates for inclusive 2-jet events. Data and simulation are normalized to the same number of events. The distributions are shown after the application of the jet ID.
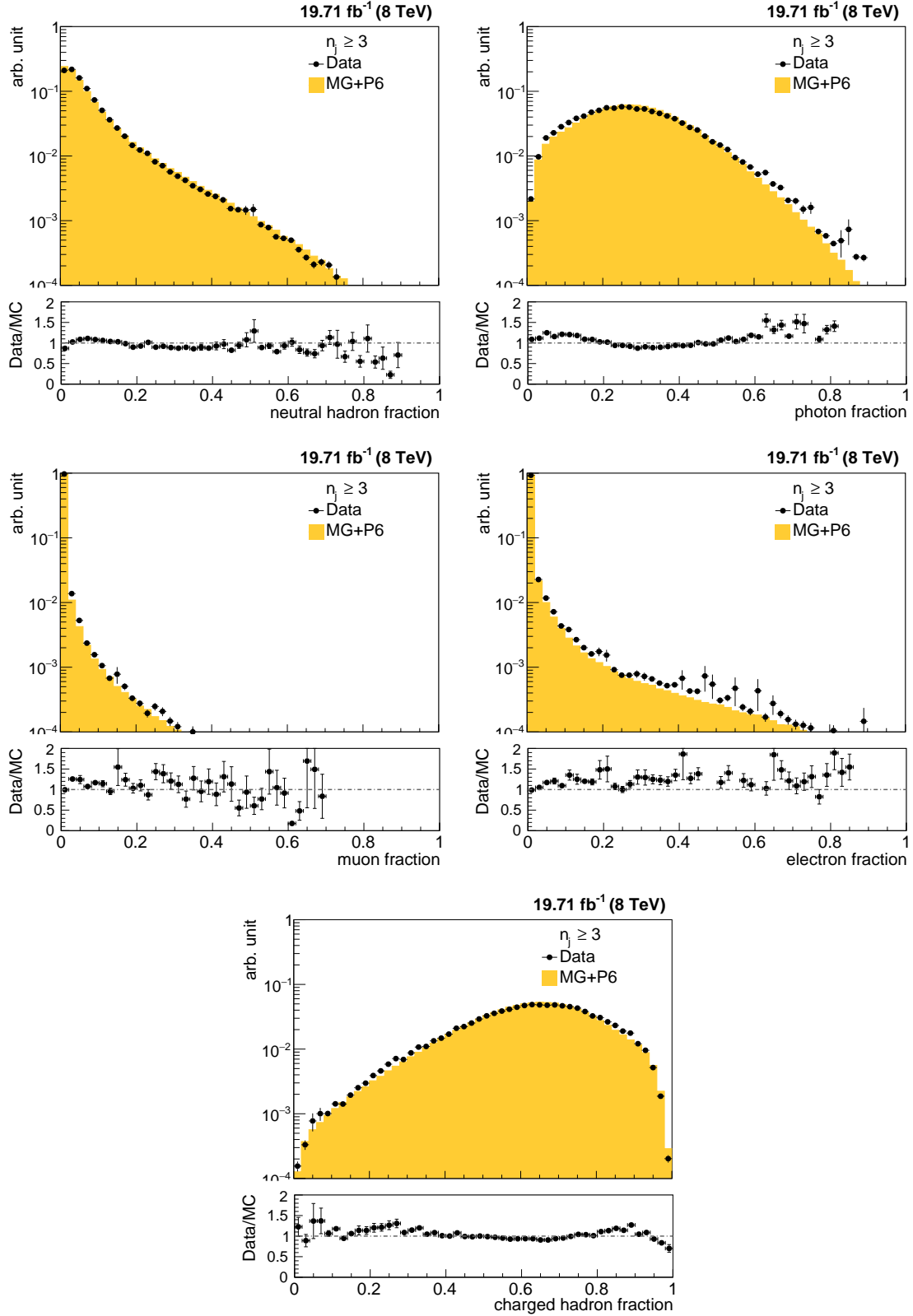
Figure 2.4: The fractions of jet constituents as observed in data and simulated events for different types of PF candidates for inclusive 3-jet events. Data and simulation are normalized to the same number of events. The distributions are shown after the application of the jet ID.

distributions in data and simulated events, the simulated events are reweighted with a weight w_{PU} to match the distribution in data :

$$w_{PU} = \frac{N_{data}(N_{PU}, est.)/\sum N_{data}}{N_{MC}(N_{PU}, truth)/\sum N_{MC}} \tag{2.4}$$

Figure 2.5 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.
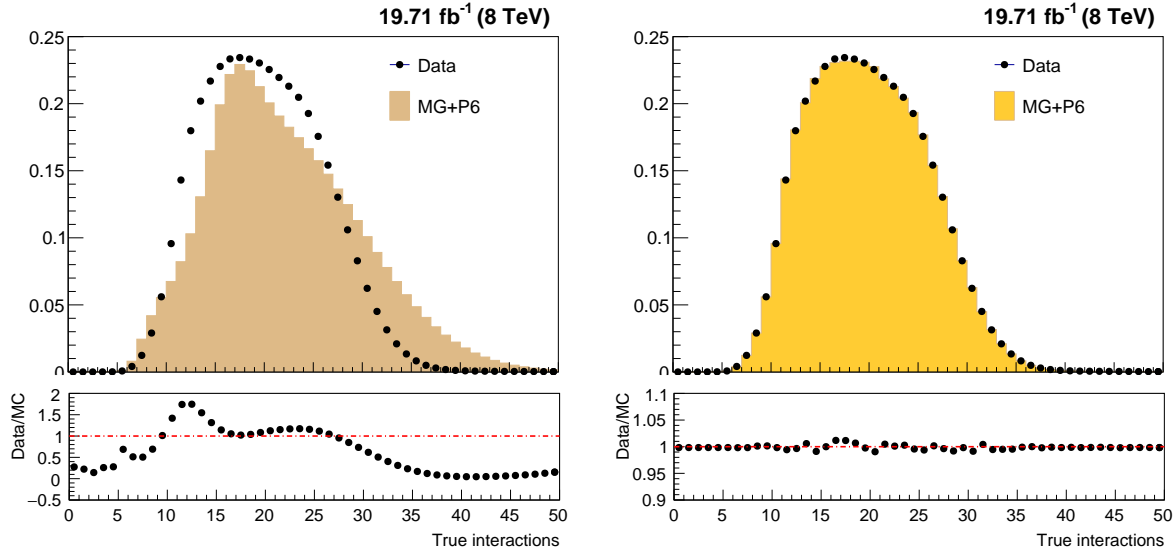
Figure 2.5: Number of reconstructed vertices in data and simulated events before (left) and after (right) the pile-up reweighting.

2.4 Kinematic Distributions

2.4.1 Differential cross section and ratios

The inclusive differential 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, and is defined by the basic formula

:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\left(H_{\mathrm{T},2}/2\right)} = \frac{1}{\epsilon \mathcal{L}_{\mathrm{int,eff}}} \frac{N}{\Delta\left(H_{\mathrm{T},2}/2\right)} \tag{2.5}$$

where

- N is the number of jets counted in a bin,

- $\mathcal{L}_{\mathrm{int,eff}}$ is the effective integrated luminosity,

- ϵ is the product of the trigger and jet selection efficiencies (greater than 99%),

- $\Delta\left(H_{\mathrm{T},2}/2\right)$ are the bin widths in $H_{\mathrm{T},2}/2$.

The measurements are reported in units of (pb/). The inclusive 2-jet events results are represented by n$_j$ $\geq$2, inclusive 3-jet events results by n$_j$ $\geq$3 and inclusive 4-jet events results by n$_j$ $\geq$4 in the figures. The differential cross section is studied for inclusive 2-jet, inclusive 3-jet and inclusive 4-jet events. Then ratio R_{mn} as a function of $H_{\mathrm{T},2}/2$ is studied which is the ratio of the differential cross section of selected inclusive m-jet events to that of inclusive n-jet events in each $H_{\mathrm{T},2}/2$ bin (m > n and m $\neq$ n),

$$R_{mn} = \frac{\left[\frac{\mathrm{d}\sigma}{\mathrm{d}\left(H_{\mathrm{T},2}/2\right)}\right]_{(m\text{-}jet)}}{\left[\frac{\mathrm{d}\sigma}{\mathrm{d}\left(H_{\mathrm{T},2}/2\right)}\right]_{(n\text{-}jet)}} \tag{2.6}$$

In this section we have detector level comparison of differential cross section of 2012 full data with NLO predictions as well as MADGRAPH5 + PYTHIA6 MC simulations. Figure 2.6 shows the comparison of differential cross section as a function of $H_{\mathrm{T},2}/2$ for inclusive 2-jet events (top left) and inclusive 3-jet events (top right), for data (black solid circles), MADGRAPH5 + PYTHIA6 MC (red solid circles) and NLO

(histogram). At present, for inclusive 4-jet events (bottom), the results from data and MC are available. NLO predictions are yet to be done for this case. Each plot also shows the ratio between the MC predictions and the data as well as between the NLO predictions and the data. Di-jet production is known to suffer from large corrections from soft gluon radiation which requires re-summation beyond fixed order perturbation theory. Theoretical predictions at NLO including the parton shower (NLO+PS) allow to account for these effects and obtain a better description of the available data. 150-200 bins are not included to avoid the infrared sensitivity for the bins next to min. p_{T} cut in NLO calculations for inclusive 2-jet events.

2.4.2   Jet Energy Resolution (JER)

To evaluate jet energy resolution (JER), CMS detector simulation based on MAD-GRAPH5 + PYTHIA6 MC event generators, is used. The jets clustered from stable generator particles (Gen jets) as well as from particle flow candidates reconstructed from the simulated detector output (Reco jets) are used. The studies of the JERC working group at CMS has shown that the jet energy resolution in data is actually worse than in simulation. So the reconstructed jet transverse momentum needs to be smeared additionally to match the resolution in data [?]. Table 2.3 shows the scaling factors (c) which need to be applied on the transverse momentum of simulated reconstructed jets. The scaling factors depend on the absolute η of the jet. The uncertainty on these measured scaling factors needs to be taken into account in a physics analysis. This is done by smearing the reconstructed jets with two additional sets of scaling factors, c_{up} and c_{down}, that correspond to varying the factors up and down respectively, by one sigma and evaluating the impact of these new sets.

The reconstructed jet p_{T} () is smeared randomly using a gaussian width widened by the scaling factor (c)

$$p_{\mathrm{T}} \rightarrow Gauss(\mu = p_{\mathrm{T}}, \sigma = \sqrt{c^2 - 1} \cdot JER^{MC_{p_{\mathrm{T}}}}) \tag{2.7}$$
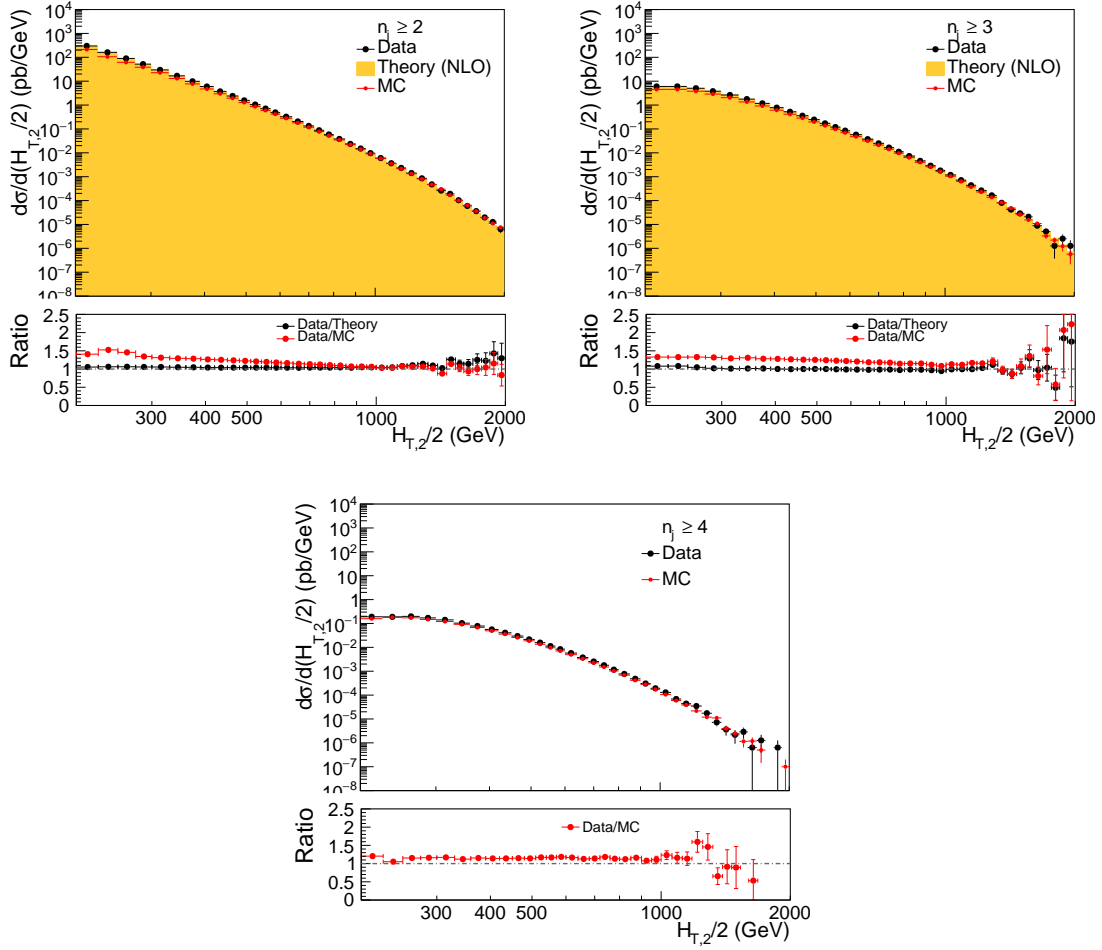
Figure 2.6: The differential cross section as a function $H_{T,2}/2$ for inclusive 2-jet events (top left), for inclusive 3-jet events (top right) and for inclusive 4-jet events (bottom), for data (black solid circles), MADGRAPH5 + PYTHIA6 MC (red solid circles) and NLO (histogram). Ratios between the MC predictions and the data as well as between the NLO predictions and the data are shown in bottom panel of each plot.

Table 2.3: The scaling factors to be applied to the reconstructed jet transverse momentum in simulated events to match the resolution in data.

η	$0.0 - 0.5$	$0.5 - 1.1$	$1.1 - 1.7$	$1.7 - 2.3$	$2.3 - 2.8$
$c_{central}$	1.079	1.099	1.121	1.208	1.254
c_{down}	1.053	1.071	1.092	1.162	1.192
c_{up}	1.105	1.127	1.150	1.254	1.316

where $JER^{MC_{p_{\mathrm{T}}}}$ is jet p_{T} absolute resolution

After smearing transverse momentum of each reco jet, $H_{\mathrm{T},2}/2$ is calculated from both generator particle jets and the particle flow jets. Then the resolution is calculated as :

$$R = \frac{Gen\ H_{\mathrm{T},2}/2}{Reco\ H_{\mathrm{T},2}/2} \tag{2.8}$$
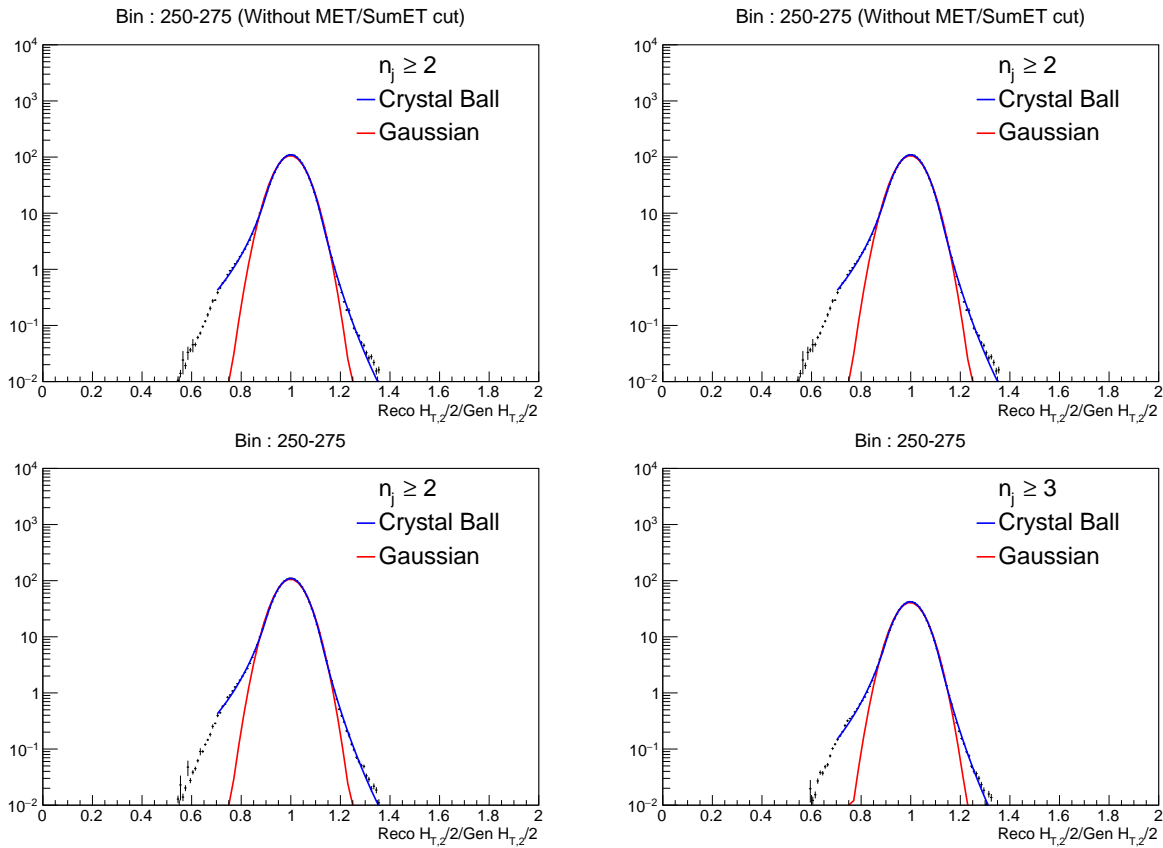
Figure 2.7: Fitting of the resolution distribution as a function of $H_{\mathrm{T},2}/2$ for inclusive 2-jet (left) and for inclusive 3-jet events (right). The blue solid line shows the Crystal Ball fit of Reco $H_{\mathrm{T},2}/2$/Gen $H_{\mathrm{T},2}/2$ in each Gen $H_{\mathrm{T},2}/2$ bin overlayed by Gaussian fitting the core of the resolution (red solid line). Top row shows the results without the $\frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sum E_{\mathrm{T}}}$ cut whereas the bottom results are after applying the $\frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sum E_{\mathrm{T}}}$ cut.

The width of the resolution distribution in a given Gen $H_{\mathrm{T},2}/2$ bin is interpreted as the resolution. The jet resolution can in good approximation be described by the σ of a Gaussian fit to the core of the jet resolution distribution. To describe

the tails of the jet response distribution, a double-sided Crystal-Ball function is used.

A fit example for one Gen $H_{\mathrm{T},2}/2$ bin is shown in Figure 2.7 for inclusive 2-jet (left) and inclusive 3-jet events(right). Here the black dots represent the jet resolution distribution, the double-sided Crystal-Ball fit (blue solid line) is overlayed by the Gaussian fit (red solid line). Top row shows the results without the $\frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sum E_{\mathrm{T}}}$ cut whereas the bottom row results are after applying the $\frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sum E_{\mathrm{T}}}$ cut. It is observed that this cut is not affecting the Reco/Gen tails.

The width of the distributions in each Gen $H_{\mathrm{T},2}/2$ bin is then plotted as a function of Gen $H_{\mathrm{T},2}/2$. From Figure 2.8, it has been observed that the Crystal Ball function better describes the measured distributions, especially in the low-$H_{\mathrm{T},2}/2$ region where the non- Gaussian tails are more pronounced. So we have switched to the Crystal Ball function to determine the resolution.

Figure 2.9 shows the final relative resolution which is described by a modified version of the NSC formula as mentioned in Equation 2.9. Also, the extrapolation of the fit function upto 2 TeV is shown. The formula is based on the usual NSC formula which describes the resolution in terms of noise N, a stochastic component S and a constant term C. Especially in the low $H_{\mathrm{T},2}/2$ region in which the tracking has a non-negligible influence on the resolution due to the particle flow algorithm, a slightly better fit is obtained by using the modified resolution formula. Table 2.4 gives the parameters of the fit for the inclusive 2-jet and inclusive 3-jet events.

$$\frac{\sigma(x)}{x} = \sqrt{sgn(N) \cdot \frac{N^2}{x^2} + S^2 \cdot x^{s-1} + C^2} \qquad (2.9)$$

While calculating JER where we are using one large rapidity bin, it has been observed that from Figure 2.10, when JER extracted from simulated MGP6
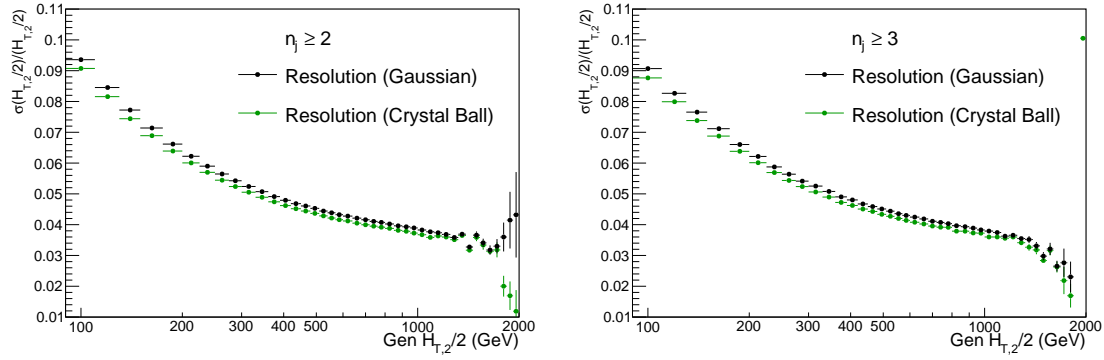
Figure 2.8: Comparison of resolution calculated using Gaussian and Crystal-Ball fit functions for inclusive 2-jet events (left) and for inclusive 3-jet events (right).
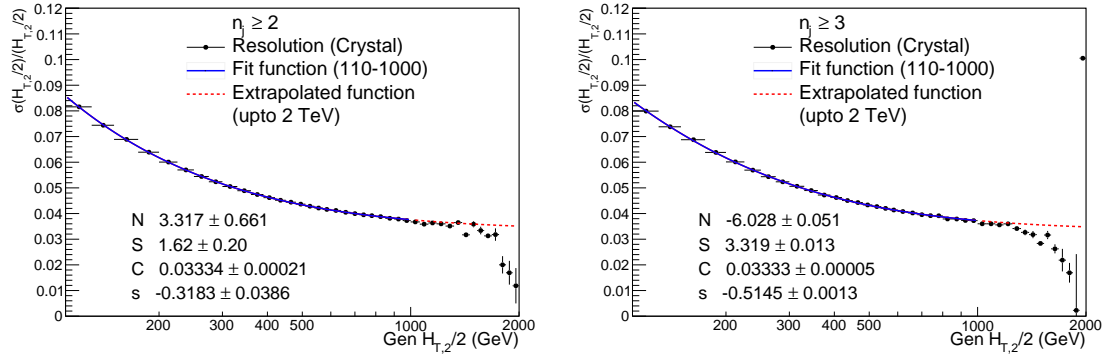
Figure 2.9: Resolution as a function of Gen $H_{\mathrm{T},2}/2$ for inclusive 2-jet events (left) and for inclusive 3-jet events (right).

Reco/MGP6 Gen is used in toyMC for smearing, it smears the fastNLO too much
(red curve). The extracted JER also smears MC Gen more as seen from Smeared
MGP6 Gen/MGP6 Gen (black dashed curve). When the 30% reduced JER is used to
smear MGGen, the Smeared MGGen/MGP6 Gen, matches with simulated MGP6
Reco/MGP6 Gen. Also, toyMC Gen smeared with 30% reduced JER, Smeared
FastNLO/Gen FastNLO. So an additional unfolding uncertainty is attributed by
comparison to 30% reduced JER.
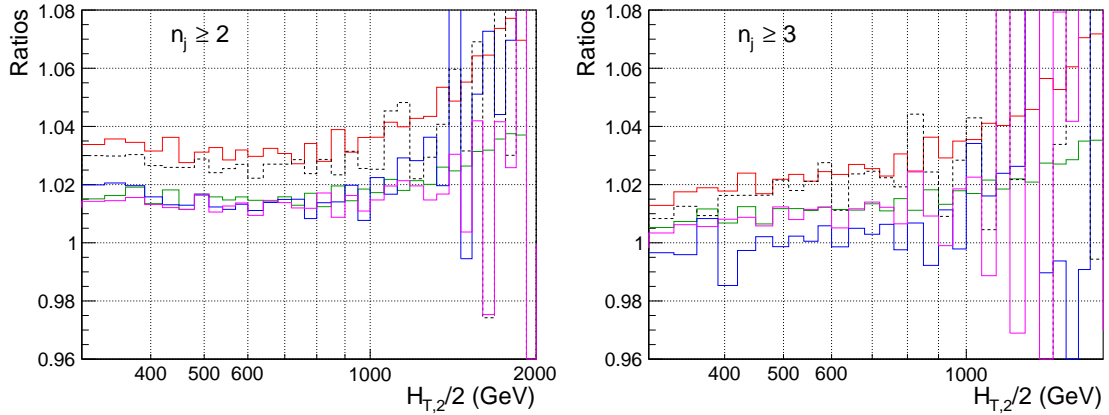
Figure 2.10: Simulated MGP6 Reco/MGP6 Gen to extract JER, Smeared
FastNLO/Gen FastNLO using extracted JER, Smeared MGP6 Gen/MGP6 Gen us-
ing extracted JER, smeared MGP6 Gen/MGP6 Gen using 30% reduced extracted
JER, Smeared FastNLO/Gen FastNLO using 30% reduced extracted JER; for in-
clusive 2-jet (left) and inclusive 3-jet events (right).

Table 2.4: The fit parameters of the resolution for inclusive 2-jet and inclusive 3-jet
events.

	N	S	C	s
Inclusive 2-jet	3.32	1.62	0.0333	-0.318
Inclusive 3-jet	-6.03	3.32	0.0333	-0.515

2.4.3 Unfolding

One of the main goal in the experimental measurement is to do the comparison
between data and theory predictions or with the results from other experiments.
But the finite detector resolution and the steeply falling jet p_T spectrum distorts

the physical quantities. As a result, the measured observables are different from the corresponding true values. Each p_T bin content contains the migrated events from neighbouring bins along with the original events. So an unfolding process of the data for detector effects should be followed. Unfolding uses a response matrix that maps the true distribution onto the measured one. The response matrix is usually derived from simulated Monte Carlo (MC) samples (eg. PYTHIA6 tune Z2 MC), which uses as input the true distribution from MC and introduces the smearing effects by taking into account the detector resolution. Then this response matrix is used to unfold the measured data spectrum. There are several drawbacks of constructing response matrix from the MC as MC is not perfectly calibrated. Also dealing with the systematic uncertainties is practically impossible with this technique.

However, there is an indirect way of constructing the response matrix which uses a custom Toy Monte Carlo by utilizing NLO prediction whose details are given in section 2.4.3.1. The particle level $H_{\mathrm{T},2}/2$ spectrum is obtained from the falling fitted function used to fit the theoretically predicted $H_{\mathrm{T},2}/2$ spectrum. The reconstructed level $H_{\mathrm{T},2}/2$ spectrum is obtained by smearing the particle level $H_{\mathrm{T},2}/2$ spectrum with the jet energy resolution (JER) parameters.

In this analysis, the measured cross sections are corrected for detector smearing effects and unfolded to stable particle level and then the ratio R_{32} is constructed. We have used the iterative D'Agostini Bayesian method with four iterations [?], implemented in the RooUnfold software package [?].

2.4.3.1   Response matrices

The NLO spectrum of differential cross section is fitted with the following two different functions :

- **Function I :**

$$f(p_T) = N[x_T]^{-a}[1 - x_T]^b \times exp[-c/x_T] \qquad (2.10)$$

where N is normalization factor and a, b, c are fit parameters.

This function is derived from the below function from **Phys.Rev.Lett. 107, 132001 (2011)**

$f(p_T; \alpha, \beta, \gamma) = N_0[p_T]^{-\alpha}[1 - \frac{1}{\sqrt{s}} 2p_T \cosh(y_{\min})]^{\beta} \times \exp[-\gamma/p_T]$, where N_0 is a normalization factor, α, β, γ are fit parameters, and $y_{\min}$ is the low-edge of the rapidity bin y under consideration.

using

$$\alpha = a, \quad \beta = b, \quad \gamma = c * \sqrt{s}/2, x_T = \frac{2 * p_T * cosh(y_{min})}{\sqrt{s}} = \frac{2 * p_T}{\sqrt{s}} \quad (2.11)$$

where transverse scaling variable x_T corresponds to the proton fractional momentum x for dijets with rapidity y=0, $\sqrt{s} = 8000$ GeV and y_{min} is low-edge of the rapidity bin y under consideration (Here y_{min} is taken equal to 0)

- **Function II :**

$$f(H_{T,2}/2) = A_0\Big(1 - \frac{H_{T,2}/2}{A_6}\Big)^{A_7} \times 10^{F(x)}, \text{where} \quad F(x) = \sum_{i=1}^{5} A_i\Big(\log\big(\frac{x}{A_6}\big)\Big)^i$$

$$(2.12)$$

where the parameter A_6 is fixed to $\frac{\sqrt{s}}{2cosh(y_{min})}$, where $\sqrt{s} = 8000$ GeV and y_{min} is the minimum rapidity. The other parameters are derived from the fitting.

Figure 2.11 shows the fitted NLO spectrum of differential cross section as a function of $H_{\text{T},2}/2$ using Function I (top) and using Function II (bottom) : for inclusive 2-jet events (left) and for inclusive 3-jet events (right). Function I is used primarily to generate response matrices and perform the closure tests and Function
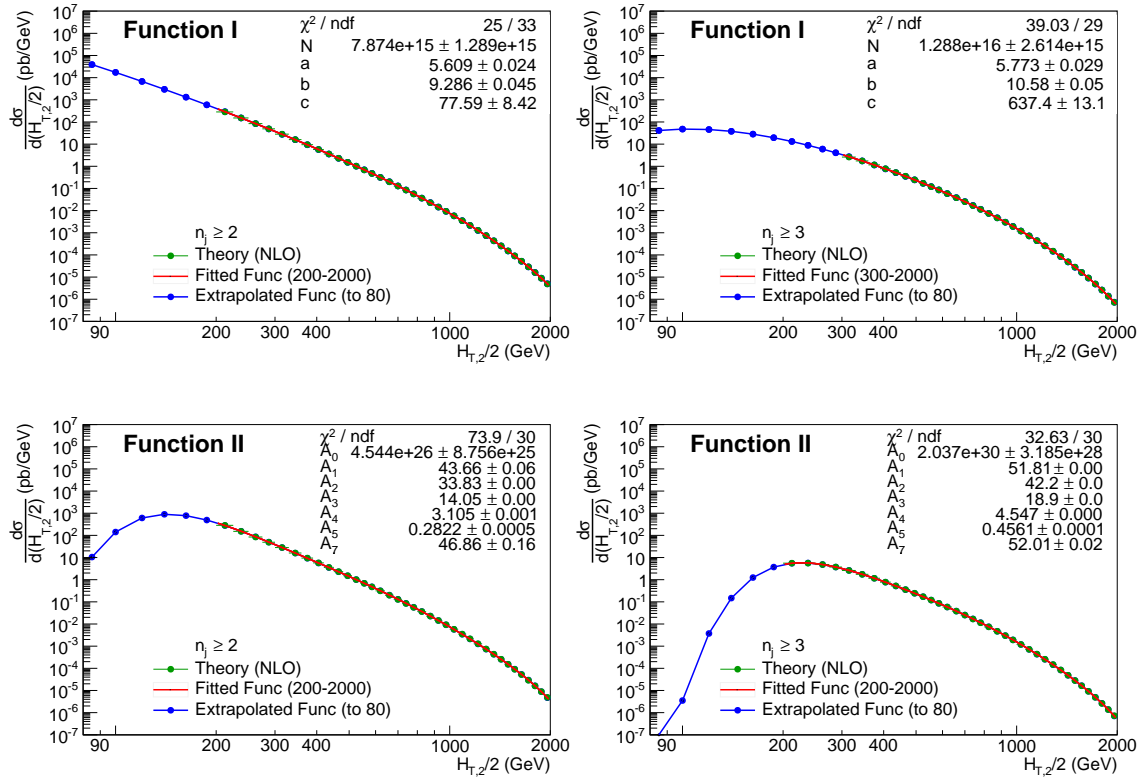
Figure 2.11: Fitted NLO spectrum of differential cross section as a function of $H_{T,2}/2$ using Function I (top) and using Function II (bottom) : for inclusive 2-jet events (left) and for inclusive 3-jet events (right).

II is used as an alternative function to calculate unfolding uncertainty, described in section 2.4.4.4. To include the migration to lower bins, the fitted functions are extrapolated to the lower $H_{T,2}/2$ values.
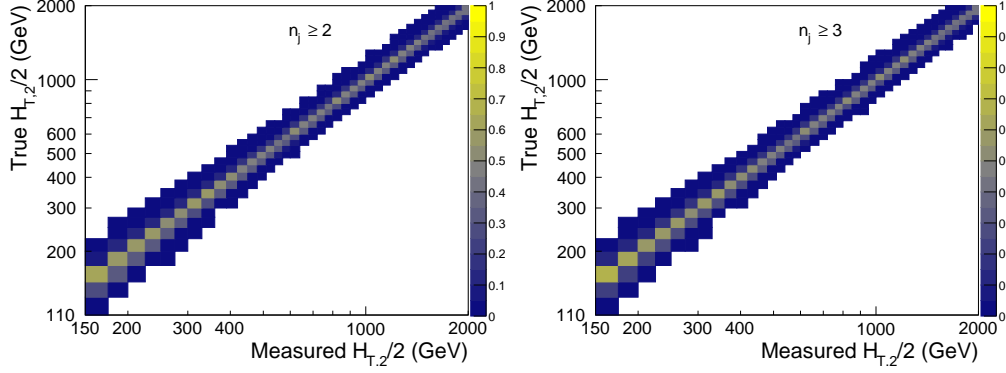
Figure 2.12: The response matrices derived using the Toy MC for inclusive 2-jet (left) and inclusive 3-jet events (right).
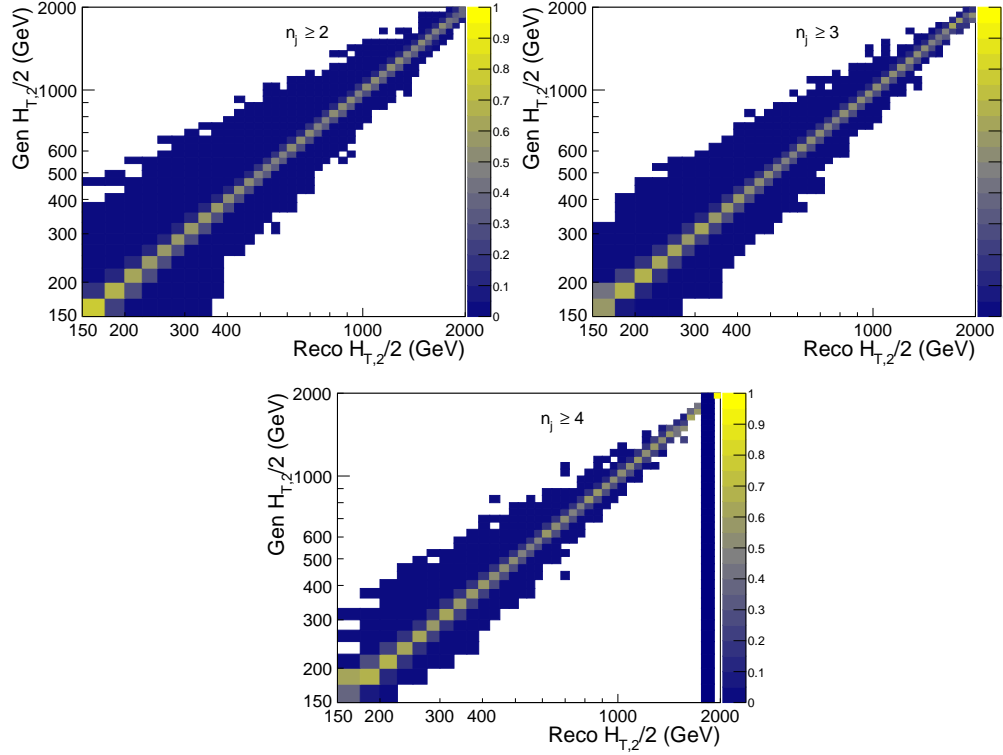
Figure 2.13: The response matrices constructed from MADGRAPH5 + PYTHIA6 MC inclusive 2-jet (top left), inclusive 3-jet events (top right) and inclusive 4-jet events (bottom).

A flat $H_{T,2}/2$ spectrum is generated and the fit parameters obtained from the NLO spectrum provides weights to the flat spectrum. A total of ten million events

are generated (in $H_{\mathrm{T},2}/2$ range 80-2000). These generated values are smeared with a Gaussian function, where σ of the Gaussian is determined from the relative resolution parametrization as a function of $H_{\mathrm{T},2}/2$ calculated from NSC formula mentioned in equation 2.9. The parameters N, S, C used for smearing are taken from Table 2.4.

These generated and smeared values are used to fill the response matrices by using the RooUnfold package. Figure 2.12 shows the response matrices derived using the Toy MC for inclusive 2-jet (left) and inclusive 3-jet events (right). The matrices are normalized to the number of events in each column. The response matrices are diagonal with small off-diagonal migrations between close- by $H_{\mathrm{T},2}/2$ bins.

Also, the response matrices are constructed from MADGRAPH5 + PYTHIA6 MC using the distributions from PFJets (Jets based on individual particles reconstructed by the Particle Flow Algorithm) and GenJets (Jets found by applying a jet algorithm on MC stable particles which are produced at the end of the hadronization stage). Figure 2.13 shows the response matrices constructed from MADGRAPH5 + PYTHIA6 MC for inclusive 2-jet (top left), inclusive 3-jet events (top right) and inclusive 4-jet events (bottom).

2.4.3.2 Closure test

First we unfold the smeared spectrum obtained from Toy MC, using the constructed response matrices shown in Figure 2.12 to check how good the unfolding procedure is working. Figure 2.14 shows that after unfolding, the smeared spectrum matches exactly with the randomly generated spectrum as the ratio of these distributions is perfectly flat at one for both inclusive 2-jet (left) as well as inclusive 3-jet events (right).

Also we unfold the distribution from PFJets obtained from MADGRAPH5 + PYTHIA6 MC using the constructed response matrices shown in Figure 2.13. Figure 2.15 shows that after unfolding, the spectrum from PFJets matches exactly with

that from GenJets as expected.

Also we unfold the distribution from PFJets obtained from MADGRAPH5 + PYTHIA6 MC with the toyMC response matrices shown in Figure 2.12. While performing this closure test, it has been observed that when 30% reduced JER is used to unfold MADGRAPH5 + PYTHIA6 Reco MC, a good closure is obtained as seen in Figure 2.16.

2.4.3.3 Unfolding data

After the validity of the unfolding method, the data is unfolded using the above reconstructed response matrices : both from NLO as well as MC. The unfolded data is compared to that of measured. Figure 2.17 shows the ratio of data unfolded to the measured data using response matrices from NLO (black solid circles), from NLO but 30% reduced JER (green solid circles) and from MC (red open circles); for inclusive 2-jet (left) and inclusive 3-jet events (right). The unfolding using MC and reduced JER NLO response matrices, give similar results within statistical uncertainties. The unfolding does not working well for bins near to minimum p_T cut for inclusive 2-jet events because two-jet rate is very sensitive to soft gluon emission while the higher jet multiplicities are less affected. The event yield for each bin in $H_{\mathrm{T},2}/2$ is tabulated in Appendix in Table **??**.

2.4.4 Experimental uncertainties

The experimental sources of uncertainty which affect the cross section measurement includes -

- the statistical uncertainty : propagated through the unfolding,

- the jet energy corrections uncertainty (JEC) : uncertainty due to the calibration of the jet energy,
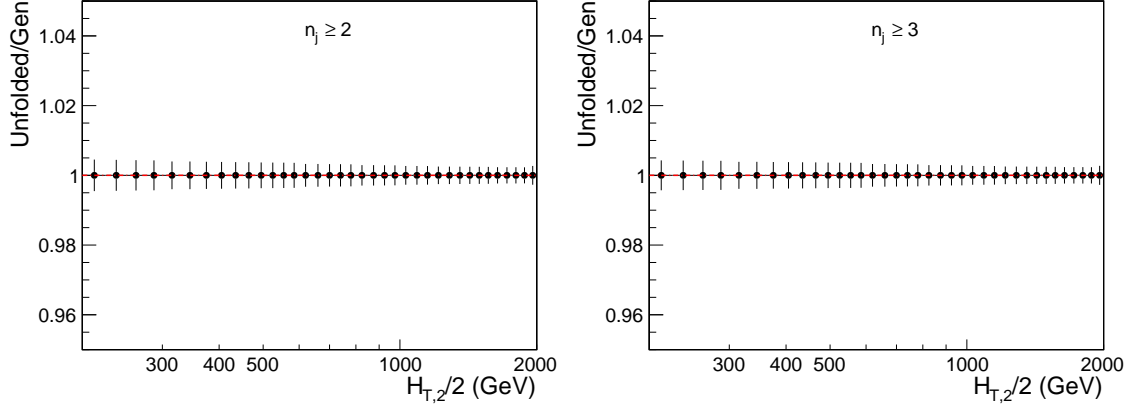
Figure 2.14: Closure test with response matrices from NLO for inclusive 2-jet (left) and inclusive 3-jet events (right).
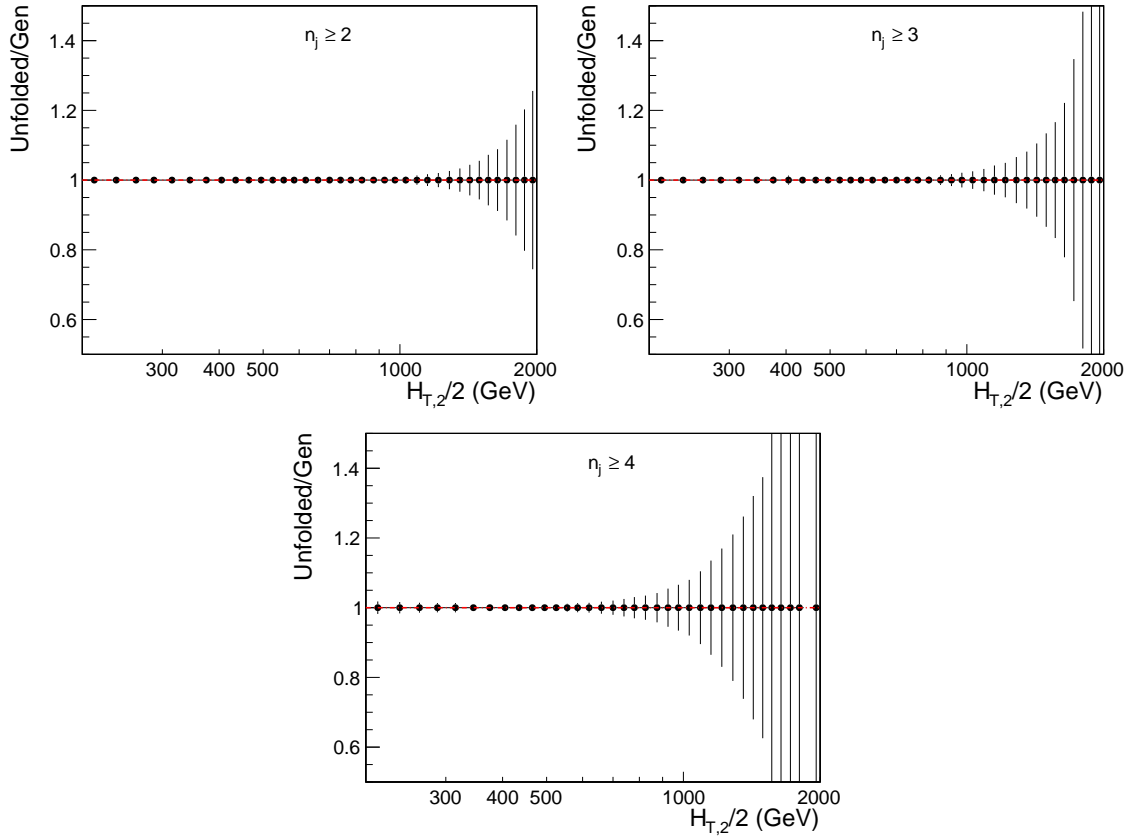
Figure 2.15: Closure test with response matrices from MADGRAPH5 + PYTHIA6 MC for inclusive 2-jet (top left), inclusive 3-jet events (top right) and inclusive 4-jet events (bottom).
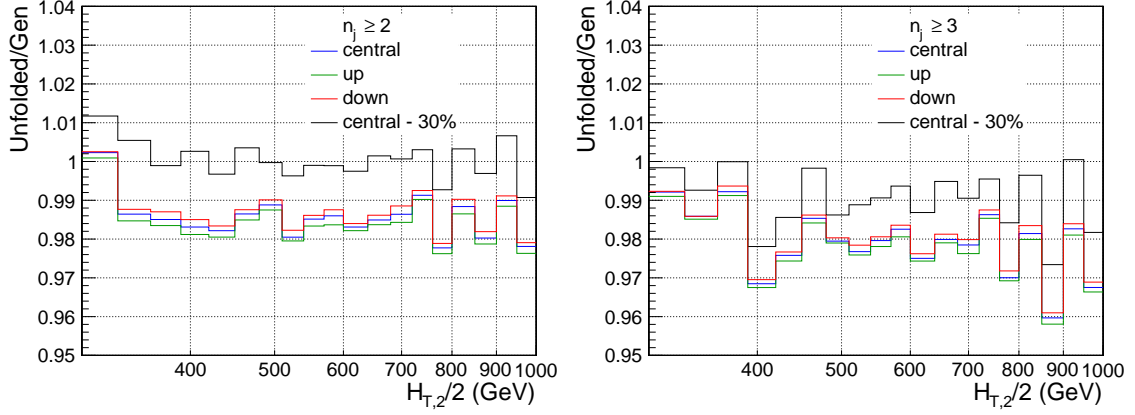
Figure 2.16: Closure test with unfolding MadGraph5 + pythia6 Reco MC with response matrices from NLO for inclusive 2-jet (left) and inclusive 3-jet events (right).
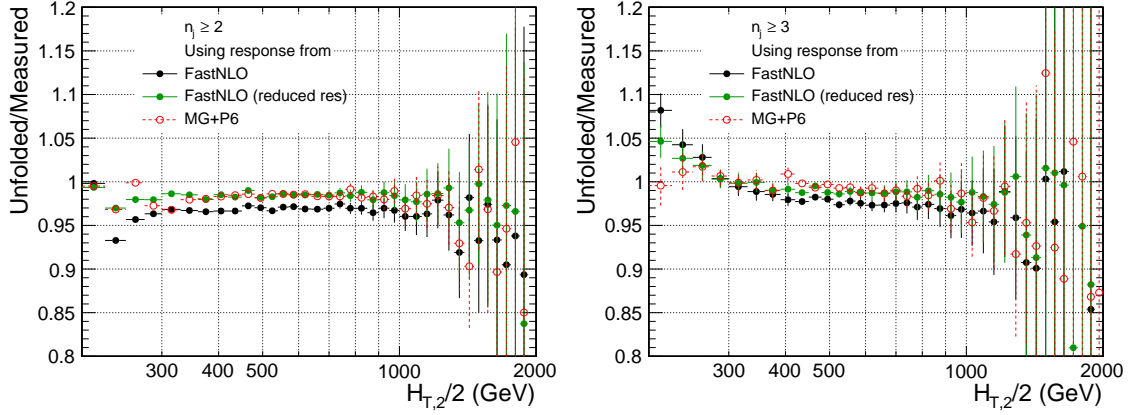
Figure 2.17: The ratio of data unfolded with that of measured using response matrices from NLO (black solid circles), from NLO but 30% reduced JER (green solid circles) and from MC (red open circles); for inclusive 2-jet (left) and inclusive 3-jet events (right).

- the unfolding uncertainty : attributed by unfolding procedure,

- the luminosity uncertainty : arising from the imperfection of the CMS luminosity,

- an uncorrelated residual uncertainty of 1% : which comprises the trigger efficiency uncertainty, the jet id efficiency and further residual effects.

2.4.4.1   Uncertainty on luminosity measurement

The uncertainty on the luminosity measurement for the 2012 LHC run is estimated to be 2.5% (syst.) and 0.5% (stat.) [?]. As the luminosity uncertainty translates into a normalization uncertainty on any absolute cross section measurement, a combined systematic uncertainty of 2.6% is assigned which is fully correlated across all the $H_{\mathrm{T},2}/2$ bins.

2.4.4.2   Statistical uncertainty

Statistical uncertainties of data points are propagated through the unfolding using a toy MC technique in which the measured data points are smeared within their statistical uncertainties and the unfolding procedure is repeated multiple times for each smeared spectra. One million of such toy spectra are used to propagate the statistical uncertainty. Figure 2.18 shows the relative statistical uncertainty before and after the unfolding procedure.

The uncertainty slightly increases during the unfolding process. Furthermore, the unfolding introduces a correlation between bins due to event migrations. These correlations are significant for neighbouring bins in $H_{\mathrm{T},2}/2$ and negligible between bins far off in the phase space. Figure 2.19 shows the correlations of the statistical uncertainty after the unfolding. We studied the correlations by performing unfolding with taking 4, 5 and 10 iterations and choosing the unfolding iteration parameter as
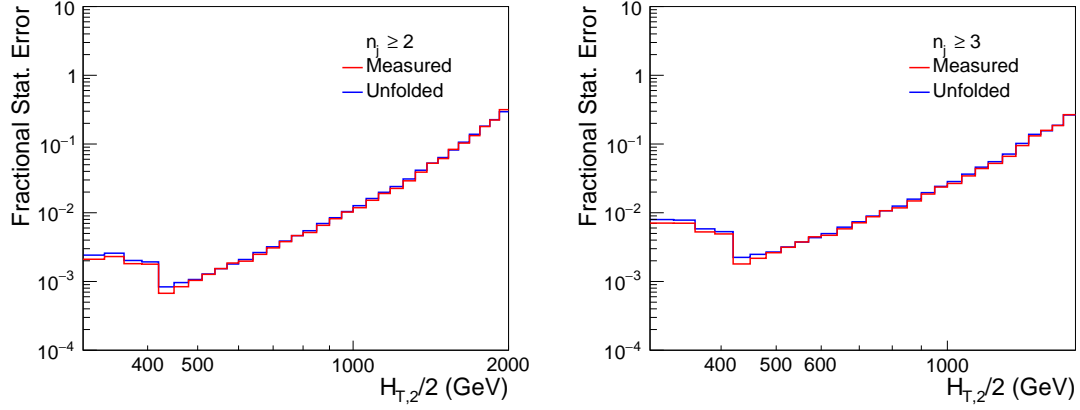
Figure 2.18: The fractional statistical uncertainties of the measured and the unfolded data for inclusive 2-jet events (left) and for inclusive 3-jet events (right). Depending on the unfolding procedure, the uncertainties can slightly increase, which is observed.

"4" ensures that the statistical errors in the unfolded distributions are greater than that of the measured distributions.
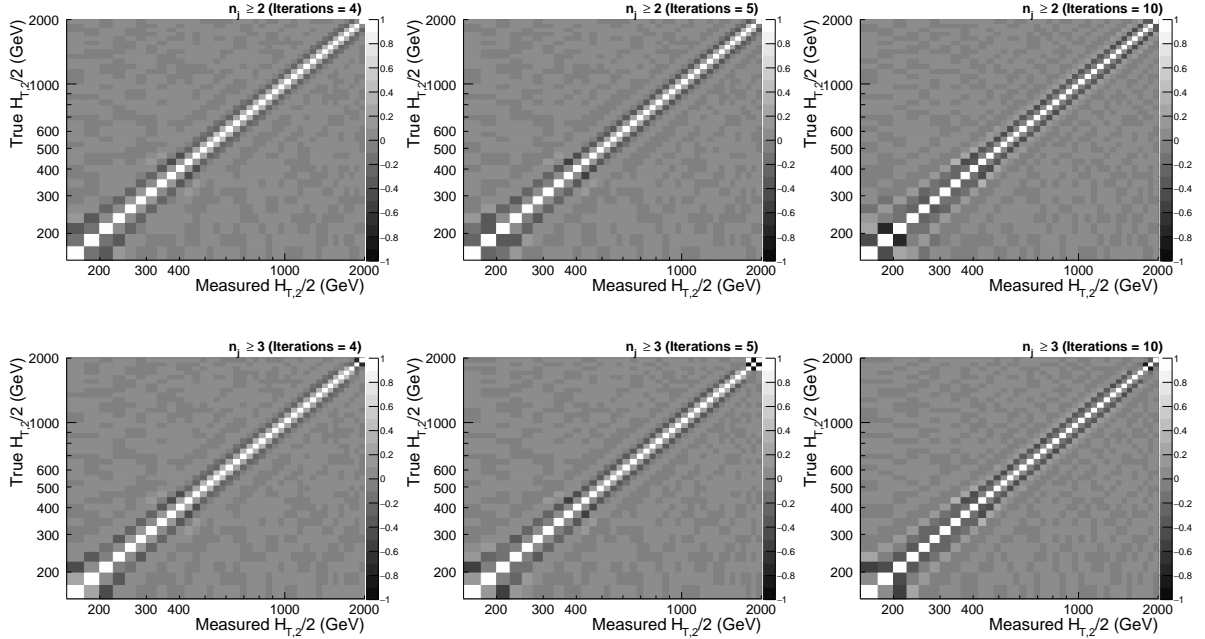
Figure 2.19: Correlations of the statistical uncertainty introduced by the unfolding procedure for inclusive 2-jet events (Top) and for inclusive 3-jet events (Bottom) with 4 iterations (left), 5 iterations (middle) and 10 iterations (right). Neighbouring bins have a significant correlation or anti-correlation through bin migrations.

2.4.4.3 Jet Energy Scale uncertainty

The systematic uncertainty on the measured cross sections is asymmetric and dominated by the uncertainty on the jet-energy corrections (JEC). This is estimated by shifting the jet p_T according to the JEC uncertainty. The JEC uncertainties are split into 25 mutually independent sources of uncertainty, in which each source is fully correlated in p_T and η but uncorrelated to all other sources and presents a 1σ shift. $H_\mathrm{T,2}/2$ is calculated by varying p_T and the difference from central $H_\mathrm{T,2}/2$ gives JEC uncertainty as a function of $H_\mathrm{T,2}/2$. As these uncertainties can be asymmetric, the upwards and downwards variation of each source are treated separately. The sum in quadrature of all sources yields the total JEC uncertainty.

The sources of uncertainty are grouped in four categories according to their origin and details about the jet energy corrections and uncertainties are given in [**?**] and in :

https://twiki.cern.ch/twiki/bin/viewauth/CMS/

JECUncertaintySources#Winter14_uncertainties.

The 25 individual uncertainty sources are the following : AbsoluteStat, AbsoluteScale, AbsoluteFlavMap, AbsoluteMPFBias, Fragmentation, SinglePionECAL, SinglePionHCAL, FlavorQCD, RelativeJEREC1, RelativeJEREC2, RelativeJERHF, RelativePtBB, RelativePtEC1, RelativePtEC2, RelativePtHF, RelativeFSR, RelativeStatFSR, RelativeStatEC2, RelativeStatHF, PileUpDataMC, PileUpPtRef, PileUpPtBB, PileUpPtEC1, PileUpPtEC2 and PileUpPtHF.

The AbsoluteFlavMap uncertainty is exactly zero for the 8 TeV and can be ignored. In this way practically 24 uncertainties are considered to calculate JEC uncertainty. For the four sources : RelativeJERHF, RelativePtHF, RelativeStatHF, PileUpPtHF, the JEC uncertainty is exactly zero because of $|y| < 2.5$ cut used in the analysis. So only 20 sources contribute to the total JEC uncertainty. The Figures **??**-**??** show the JEC uncertainty from each source separately for inclusive

2-jet (left) and 3-jet events (middle), respectively. The bin-wise values (in %) are tabulated in the Tables **??**, **??** for inclusive 2-jet and **??**, **??** for inclusive 3-jet events.

2.4.4.4   Unfolding uncertainty

The unfolding uncertainty is comprised of three uncertainties :

1. **JER uncertainty :** The jet energy resolution, which was derived in Section 2.4.2, is used to produce the response matrix using a forward smearing technique in the unfolding procedure. Therefore, a dependence on the jet energy resolution on the unfolded cross section is introduced and a further uncertainty source due to the uncertainty on the jet energy resolution is introduced. Table 2.3 shows the scaling factors, which were applied on reconstructed simulated events to obtain the actual resolution in data. The official recommendations include offset variations of these scaling factors to estimate the uncertainty on the resolution and are also given in Table 2.3. The determination of the resolution is repeated with the upwards and downwards variation of the resolution scaling factor applied. The unfolding procedure is also reiterated using the variations of the resolution and the differences of the obtained cross section to the nominal cross section are accounted for as a systematic uncertainty.

2. **Model dependence :** As explained earlier that to construct the response matrix by Toy MC method, the fitting of the theoretical predictions is performed by using function given in Equation 2.10. To use alternative function for this fitting gives the model dependence of the theoretical $H_{\mathrm{T},2}/2$ spectrum which affects the response matrix and thus the unfolding. The NLO $H_{\mathrm{T},2}/2$ spectrum is fitted by the another function described by Equation 2.12. The procedure mentioned in Section 2.4.3 is repeated to get the new unfolded cross sections. The differences in unfolded cross sections using the functions given by equations 2.10 and 2.12 gives the unfolding uncertainty.

3. **Additional uncertainty :** As explained in Section 2.4.2, an additional uncertainty is added from the difference in unfolding on comparing with reduced JER.

All the three uncertainties are added in quadrature to account the unfolding uncertainty which is 1-2% on cross-sections.

2.4.4.5   Total experimental uncertainty

After calculating the uncertainties from all different sources, total experimental uncertainty is obtained by adding in quadrature the uncertainties from individual sources. The bin-wise values (in %) of uncertainties from each source as well as total uncertainty are tabulated in Table **??** and Table **??** for inclusive 2-jet and inclusive 3-jet events, respectively. Figure 2.20 show the uncertainties from all sources of experimental uncertainty as well as the total uncertainty for inclusive 2-jet (left) and inclusive 3-jet (right) events. The systematic uncertainty on the measured cross sections is asymmetric and dominated by the uncertainty on the jet-energy scale at lower $H_{\mathrm{T},2}/2$ values and by statistical uncertainty at higher $H_{\mathrm{T},2}/2$ values. The experimental uncertainties from each source as well as total uncertainty are quoted in Table 2.5.

Table 2.5: Overview of all experimental uncertainties affecting the cross section measurement.

Uncertainty Source	Inclusive 2-jet	Inclusive 3-jet
Statistical	< 1 to 30%	< 1 to 40%
JEC	3 to 10%	4 to 12%
Unfolding	1-2%	1-2%
Luminosity	2.6%	2.6%
Residual uncorrelated	1%	1%
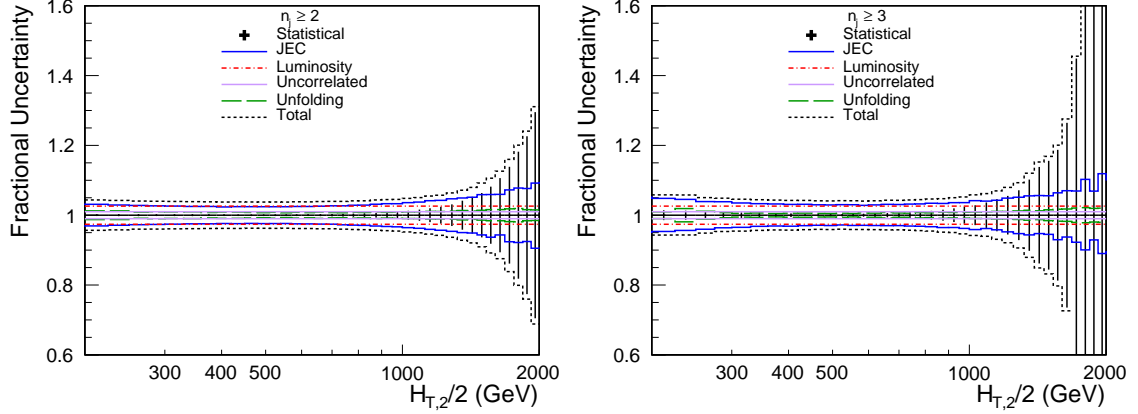
Figure 2.20: Overview of all experimental uncertainties affecting the cross section measurement for inclusive 2-jet (left) and inclusive 3- jet (right). The error bars indicate the statistical uncertainty after unfolding. The colored lines give the uncertainties resulting of jet energy scale, luminosity, unfolding and residual effects. The total uncertainty is calculated by adding in quadrature the individual sources of uncertainty.

2.5 Measurement of cross section ratio, unfolding and experimental uncertainties

2.5.1 Measurement of cross section ratio

The ratio of cross sections, R_{mn} as a function of $H_{T,2}/2$, is extracted by dividing the cross section of selected inclusive m-jet events to that of inclusive n-jet events at any given bin size of $H_{T,2}/2$. The ratios R_{32}, R_{43} and R_{42} are calculated. In cross section ratios, the numerator and denominator are not independent samples. So to calculate the statistical uncertainty for the cross section ratios before unfolding, the Wilson score interval is used which takes into account the correlation between the numerator and the denominator and give asymmetric errors. For example, the bin-wise inclusive 2-jet and 3-jet differential cross sections as well as cross section ratio R_{32}, calculated at detector level, along with statistical uncertainty (in %) are tabulated in Table **??**.

Figure 2.21 shows the ratio R_{mn} as a function of $H_{T,2}/2$: R_{32} data (black solid circles), R_{32} MADGRAPH5 + PYTHIA6 MC (red solid circles) and R_{32} NLO (green

solid line), R_{43} data (black solid triangles (up)), R_{43} MADGRAPH5 + PYTHIA6 MC (red solid triangles (up)) and R_{42} data (black solid triangles (down)), R_{42} MAD-GRAPH5 + PYTHIA6 MC (red solid triangles (down)). The ratios R_{mn} from data are in well agreement with that from MC as well as NLO predictions.
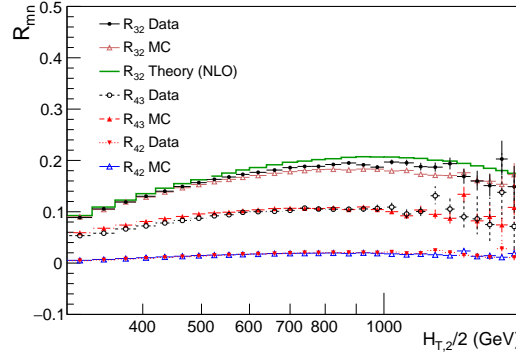
Figure 2.21: Cross section ratios R_{mn} as a function of $H_{T,2}/2$. The error bars give the statistical uncertainty, calculated by the Wilson score interval which takes into the account the correlation between the numerator and the denominator.

2.5.2   Unfolding

The measured ratio R_{32} as a function of $H_{T,2}/2$, is then corrected for detector smearing effects and unfolded to particle level. There can be two ways to obtain unfolded cross section ratio :

1. To unfold separately the inclusive 2-jet and 3-jet measured cross sections and then construct the ratio R_{32}

2. To unfold directly the ratio R_{32}

In the first method, we have ratios of unfolded cross sections by using response matrices constructed using Toy MC method i.e. smeared NLO matrices (Figure 2.12) as well as by using response matrices from MADGRAPH5 + PYTHIA6 MC (Figure 2.13). Figure 2.22 left shows the comparison of R_{32}, obtained from data unfolded using smeared NLO matrix (black solid circles) and the one unfolded

using MC matrix (red solid circles), with R_{32} from NLO prediction (red line). As the matrices from MC are also available for inclusive 4-jet events, so these can used to unfold the ratios R_{43} and R_{42} as shown in Figure 2.22 with black solid triangles (up) and black solid triangles (down), respectively.

In further analysis, cross section ratio obtained from data, unfolded using smeared NLO matrix is considered. Since the theory NLO predictions are available for inclusive 2-jet and 3-jet events and not for inclusive 4-jet events yet, so only ratio R_{32} is used in further analysis. Figure 2.22 right gives the ratio of unfolded R_{32} calculated from ratio of unfolded cross sections to that of measured one using central JER (black solid circles) as well as reduced JER (green solid circles). We have used first method to calculate the systematic uncertainties.
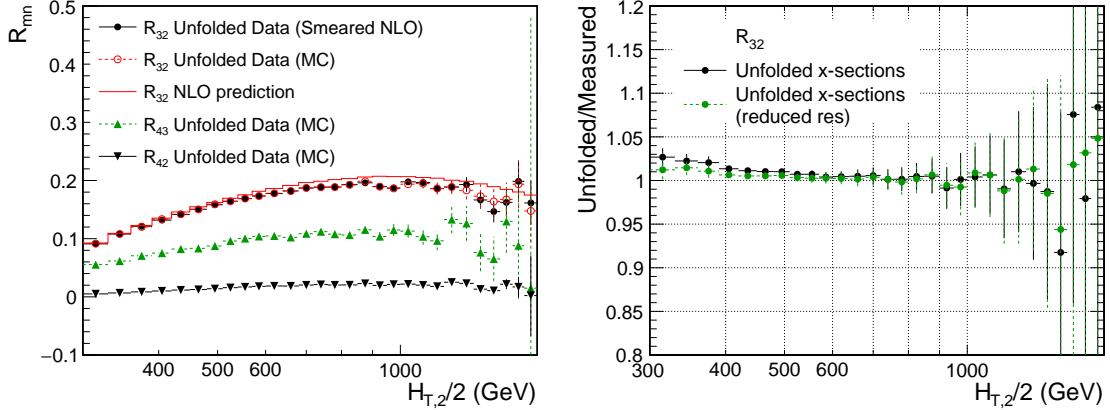
Figure 2.22: Left figure shows the unfolded R_{32}, obtained from data unfolded using smeared NLO matrix (black solid circles), from data unfolded using MC matrix (red solid circles) and from NLO prediction (red line); R_{43} from the data unfolded using MC matrix (black solid triangles (up)) and R_{42} from data unfolded using MC matrix (black solid triangles (down)). Right gives the ratio of unfolded R_{32} calculated from ratio of unfolded cross sections to that of measured one using central JER (black solid circles) as well as reduced JER (green solid circles).

2.5.2.1   Response matrix

To unfold directly the ratio R_{32}, we need to construct the response matrix using Toy MC method as explained in section 2.4.3. To obtain the true spectrum for R_{32}, the ratio of theory predictions, extrapolated with function I (equation 2.10, for inclusive

3-jet to that of 2-jet is taken, as shown in Figure 2.11 (Top). Then this ratio is fitted

using the polynomial function of degree 8 as shown in Figure 2.23.
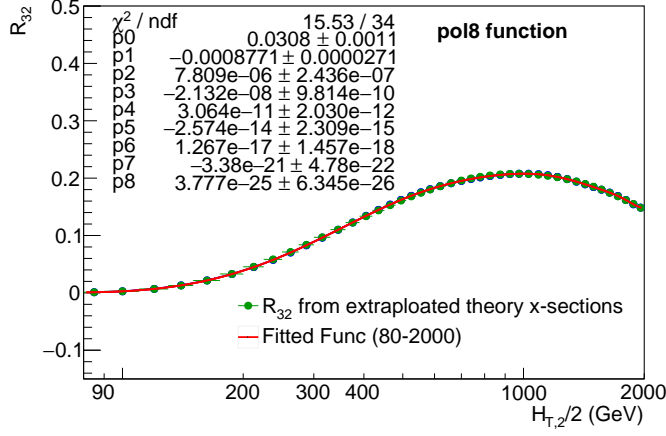
Figure 2.23: Fitted NLO spectrum of cross section ratio R_{32} as a function of $H_{\mathrm{T},2}/2$ using polynomial function of degree 8.

A flat $H_{\mathrm{T},2}/2$ spectrum is generated and the fit parameters obtained from

polynomial fit provides weights to the flat spectrum. A total of ten million events

are generated (in $H_{\mathrm{T},2}/2$ range 80-2000). These generated values are smeared with a

Gaussian function, where σ of the Gaussian is determined from the relative resolution

parametrization as a function of $H_{\mathrm{T},2}/2$ calculated from NSC formula mentioned in

equation 2.9. The parameters N, S, C used for smearing are taken to be same for

inclusive 3-jet events, as mentioned in Table 2.4.

These generated and smeared values are used to fill the response matrices by

using the RooUnfold package. Figure 2.24 (left) shows the response matrix derived

using the Toy MC for ratio R_{32}. The matrix is normalized to the number of events

in each column. The response matrix is diagonal with small off-diagonal migrations

between close-by $H_{\mathrm{T},2}/2$ bins.

First we unfold the smeared spectrum obtained from Toy MC to perform the

closure test. Figure 2.24 (right) shows that after unfolding, the smeared spectrum

matches exactly with the randomly generated spectrum as expected. The bottom

plot gives the ratio of unfolded
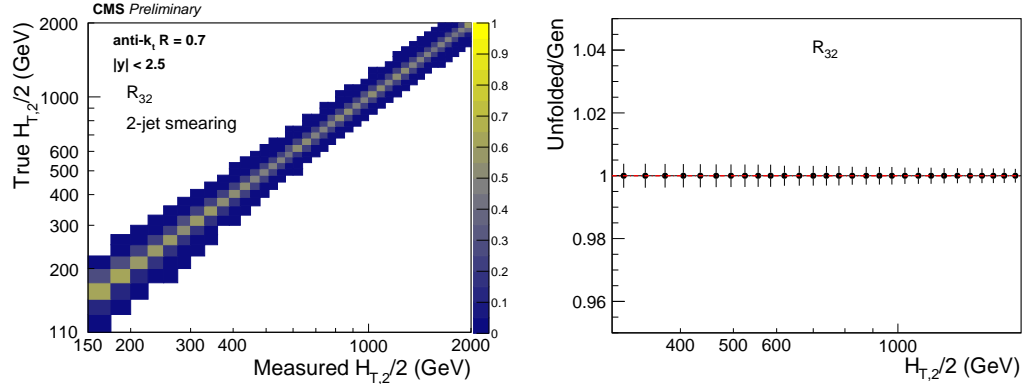
Figure 2.24: The response matrix derived using the Toy MC for ratio R_{32} (left) and Closure test (right).

2.5.3 Experimental uncertainties

In cross section ratio, the uncorrelated and luminosity uncertainties got cancel. The experimental sources of uncertainty which affect the cross section ratio measurement includes -

- the statistical uncertainty : propagated through the unfolding,

- the jet energy correction uncertainty (JEC) : uncertainty due to the calibration of the jet energy,

- the unfolding uncertainty : unfolding procedure.

2.5.3.1 Statistical uncertainty

The statistical uncertainties for ratio R_{32} are calculated in the same way as it was done for inclusive 2-jet and 3-jet cross sections, explained in section 2.4.4.2. Figure 2.25 shows the relative statistical uncertainty before and after the unfolding procedure.

The uncertainty slightly increases during the unfolding process. Also, the unfolding introduces a correlation between bins due to event migrations. To find the correlations of the statistical uncertainty introduced by the unfolding procedure
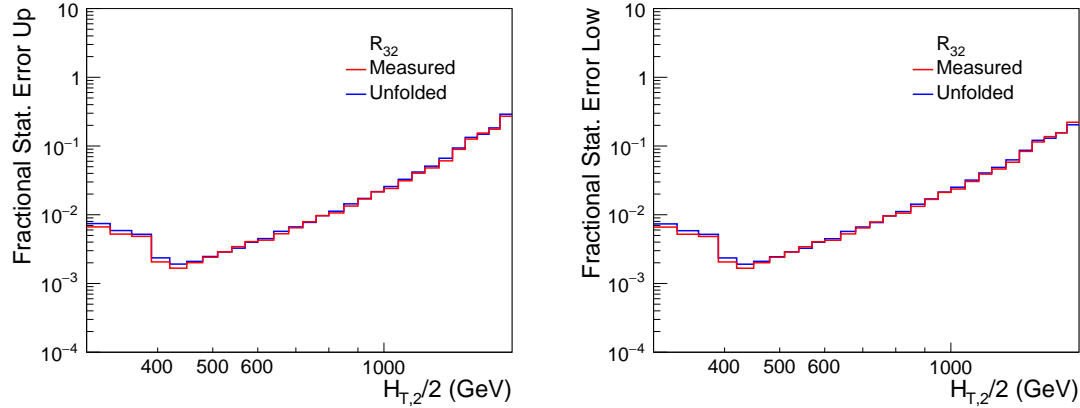
Figure 2.25: The fractional statistical uncertainties Up (left) and Low (right), of the unfolded as well as measured cross section ratio R_{32}. Depending on the unfolding procedure, the uncertainties can slightly increase, which is observed.

for ratio, R_{32}, firstly ratio is calculated and then unfolded this ratio to particle level. Figure 2.26 shows the correlations of the statistical uncertainty after the unfolding. We studied the correlations by performing unfolding with taking 4, 5 and 10 iterations and choosing the unfolding iteration parameter as "4" ensures that the statistical errors in the unfolded distributions are greater than that of the measured distributions.
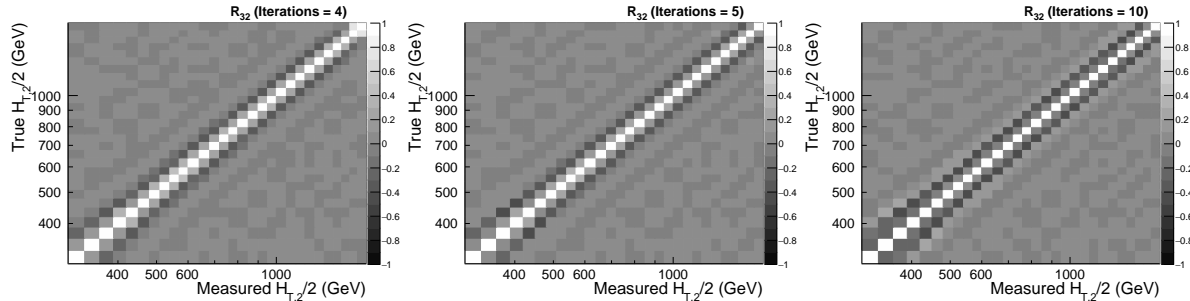
Figure 2.26: Correlations of the statistical uncertainty introduced by the unfolding procedure for ratio R_{32} with 4 iterations (left), 5 iterations (middle) and 10 iterations (right). Neighbouring bins have a significant correlation or anti-correlation through bin migrations.

2.5.3.2   Jet Energy Scale uncertainty

The systematic uncertainty on the measured cross sections is asymmetric and dominated by the uncertainty on the jet-energy correction (JEC) at low $H_{T,2}/2$. To

calculate JEC uncertainty for ratio R_{32}, the inclusive 2-jet and 3-jet events cross sections are measured as a function of $H_{T,2}/2$ by shifting the jet p_T according to the JEC uncertainty for each source separately. Then the ratio of these cross sections, R_{32} is taken and the difference from the central ratio R_{32}, gives the JEC uncertainty for each source. As these uncertainties can be asymmetric, the upwards and downwards variation of each source are treated separately. The sum in quadrature of all sources yields the total JEC uncertainty.

The Figures **??-??** show the JEC uncertainty from each source separately for ratio R_{32} (right). As expected, the JEC uncertainty for ratio is less than that for inclusive cross sections. The bin-wise values (in %) are tabulated in the Tables **??** and **??**.

2.5.3.3   Unfolding uncertainty

As explained earlier in section 2.4.4.4, the three uncertainties are considered for R_{32}:

1. **JER uncertainty :** To determine JER uncertainty, ratio R_{32} is calculated from the cross sections as a function of $H_{T,2}/2$ measured with the upwards and downwards variation of the resolution scaling factor as described in section 2.4.4.4. The differences of the obtained cross section ratio to the nominal cross section ratio are accounted as a systematic uncertainty.

2. **Model dependence :**   As explained earlier in section 2.4.4.4, the differences in cross section ratios obtained from unfolded cross sections using the functions given by equations 2.10 and 2.12 gives the unfolding uncertainty.

3. **Additional uncertainty :**   The additional uncertainty resulting from differences in unfolded cross-sections with central JER and 30% reduced JER is added.

All the three uncertainties are added to the unfolding uncertainty which is 1% for R_{32}.

2.5.3.4   Total experimental uncertainty

After calculating the uncertainties from all different sources, total experimental uncertainty is obtained by adding in quadrature the uncertainties from individual sources. The bin-wise values (in %) of uncertainties from each source as well as total uncertainty are tabulated in Table **??**. Figure 2.27 show the uncertainties from all sources of experimental uncertainty as well as the total uncertainty. The systematic uncertainty on the measured cross sections is asymmetric and dominated by the uncertainty on the jet-energy scale at lower $H_{\mathrm{T},2}/2$ values and by statistical uncertainty at higher $H_{\mathrm{T},2}/2$ values. The experimental uncertainties from each source as well as total uncertainty are quoted in Table 2.5.
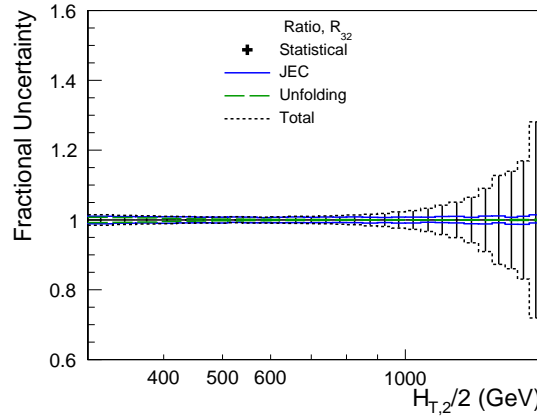
Figure 2.27: Overview of all experimental uncertainties affecting the cross section R_{32} as a function of $H_{\mathrm{T},2}/2$. The error bars indicate the statistical uncertainty. The colored lines give the uncertainties resulting of jet energy scale and unfolding. The total uncertainty is calculated by adding in quadrature the individual sources of uncertainty.

Table 2.6: Overview of all experimental uncertainties affecting the cross section ratio, R_{32}.

Uncertainty Source	Ratio R_{32}
Statistical	< 1 to $> 50\%$
JEC	1 to 2%
Unfolding	1%
Luminosity	cancels
Residual uncorrelated	cancels

2.6 Theory predictions

2.6.1 Fixed order NLO calculations

Predictions at NLO accuracy in pQCD are computed with the NLOJET++ program [?, ?]. The results are provided within the framework of FASTNLO [?] for use within fits. The renormalization and factorization scales μ_r and μ_f are chosen equal to $H_{\mathrm{T},2}/2$,

$$\mu_r = \mu_f = H_{\mathrm{T},2}/2 \tag{2.13}$$

PDF sets at NLO available for a series of different assumptions on $\alpha_s(M_Z)$ via the LHAPDF6 package [?] are listed in Table 2.7. All sets employ a variable-flavour number scheme with at most five or six flavours apart from the ABM11 PDFs, which use a fixed-flavour number scheme with $N_F = 5$.

Out of these eight PDF sets the following three will not be considered further:

- At NLO, predictions based on ABM11 do not describe LHC jet data at small jet rapidity, cf. Refs. [?, ?].

- The HERAPDF2.0 set exclusively fits HERA DIS data with only weak constraints on the gluon PDF.

- The range in values available for $\alpha_s(M_Z)$ is too limited for the NNPDF3.0 set.

PDF uncertainties are evaluated according to the prescriptions given for each PDF set. Uncertainties on $\alpha_s(M_Z)$ are not considered, since this value is later on determined from a fit to the data. The PDF uncertainty as derived with the CT10 PDF set ranges from 2% to 30% for inclusive 2- and 3-jet events.

To check the impact of higher-order contributions to the perturbative QCD prediction, the differences between LO prediction and NLO prediction are studied,

Table 2.7: NLO PDF sets available via LHAPDF6 for comparisons to data with various assumptions on the value of $\alpha_s(M_Z)$. Sets existing already in LHC Run 1 (upper rows) and newer sets for Run 2 (lower rows) are listed together with the corresponding number of flavours N_f, the assumed masses M_t and M_Z of the top quark and the Z boson, respectively, the default values of $\alpha_s(M_Z)$, and the range in $\alpha_s(M_Z)$ variation available for fits. A * behind the $\alpha_s(M_Z)$ values signifies that the parameter was fixed, not fitted.

Base set	Refs.	N_F	M_t ()	M_Z ()	$\alpha_s(M_Z)$	$\alpha_s(M_Z)$ range
ABM11	[?]	5	180	91.174	0.1180	0.110–0.130
CT10	[?]	≤ 5	172	91.188	0.1180^*	0.112–0.127
MSTW2008	[?, ?]	≤ 5	10^{10}	91.1876	0.1202	0.110–0.130
NNPDF2.3	[?]	≤ 6	175	91.1876	0.1180^*	0.114–0.124
CT14	[?]	≤ 5	172	91.1876	0.1180^*	0.113–0.123
HERAPDF2.0	[?]	≤ 5	173	91.1876	0.1180^*	0.110–0.130
MMHT2014	[?]	≤ 5	10^{10}	91.1876	0.1180^*	0.108–0.128
NNPDF3.0	[?]	≤ 5	173	91.2	0.1180^*	0.115–0.121

given by the k-factors, k_{NLO} defined as :

$$k_{NLO} = \frac{\sigma_{NLO}}{\sigma_{LO}}, \ k_{NLO}^{R_{32}} = \frac{k_{NLO}^{3-jet}}{k_{NLO}^{2-jet}} \tag{2.14}$$
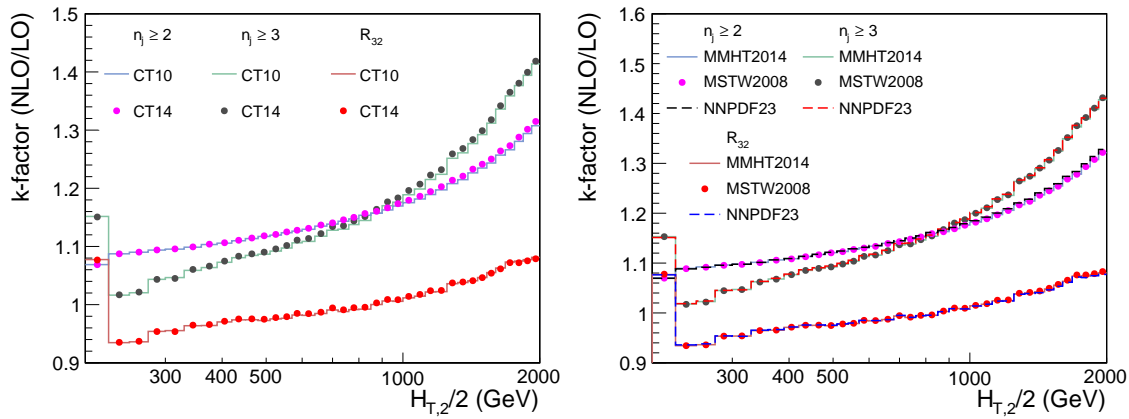
Figure 2.28: k-factors of the NLOJET++ cross section calculations for inclusive 2-jet and inclusive 3-jet cross sections and cross section ratio R_{32}, using the above mentioned scale choice and different PDF sets.

The size of NLO corrections gives an estimate about the effect of these higher-order corrections. If they are small, the LO result already describes the observable

cross sub- section precisely. Figure 2.28 shows the k-factors of the NLOJET++ cross section calculations for inclusive 2-jet and inclusive 3-jet cross sections and cross section ratio R_{32}, using the above mentioned scale choice and different PDF sets. k-factor for R_{32} is the ratio of k-factors for inclusive 3-jet cross sections to that of inclusive 2-jet as shown in Equation 2.14. The k-factors are similar for all the PDF sets in the lower region, but the differences increase in regions with larger $H_{T,2}/2$.

2.6.2   Non-Perturbative corrections

The fixed-order pQCD calculations predict the parton-level cross section and do not include additional soft QCD effects and hence cannot be directly compared to unfolded data. These calculations should be corrected for non-perturbative effects (NP) before comparison with the measurement at particle level. The impact of NP effects, i.e. from multiple-parton interactions (MPI) and hadronization, are evaluated by using samples obtained from different MC event generators with a simulation of parton-shower and underlying-event (UE) contributions. The leading order (LO), HERWIG++ [?] with the default tune of version 2.3 and PYTHIA6 [?] with tune Z2*, and the NLO, POWHEG [?,?,?], MC event generators are considered. The matrix-element calculation performed with POWHEG is interfaced to PYTHIA8 with tune CUETM1 [?] for the UE simulation. The cross section ratios between a nominal event generation interfaced to the simulation of UE contributions and a sample without hadronization and MPI effects are taken as correction separately for inclusive 2-jet, 3-jet events and ratio R_{32}, defined as in Equation 2.15. Equation 2.16 is used to calculate the NP correction factor for R_{32}. The correction is then applied as a bin-by-bin correction factor to the parton-level NLO cross section.

$$C^{NP} = \frac{\sigma^{PS+HAD+MPI}}{\sigma^{PS}} \qquad (2.15)$$

$$C_{R_{32}}^{NP} = \frac{\left(\frac{\sigma_{3\text{-}jet}}{\sigma_{2\text{-}jet}}\right)^{PS+HAD+MPI}}{\left(\frac{\sigma_{3\text{-}jet}}{\sigma_{2\text{-}jet}}\right)^{PS}} \tag{2.16}$$

$$f(H_{\text{T},2}/2) = a \cdot \left(H_{\text{T},2}/2\right)^{b} + c \tag{2.17}$$

This ratio is fitted by a power-law function defined in Equation 2.17. Since the correction factors obtained from different MC generators have large differences, an uncertainty is assigned to the correction factor. The correction factors C^{NP} are determined by the average of the envelope which covers all the differences and half of it is taken as the uncertainty. The NP corrections are shown in Figure 2.29 for the inclusive 2-jet (top left) and 3-jet event cross sections (top right), as well as for ratio R_{32} (bottom). They amount to $\sim 5\%$ for inclusive 2-jet, $\sim 7\text{--}8\%$ for inclusive 3-jet events and $\sim 4\%$ for ratio R_{32}, for $H_{\text{T},2}/2 \sim 200\,\text{GeV}$ and decrease rapidly for increasing $H_{\text{T},2}/2$. The uncertainty assigned to the NP corrections is of the order of 1–2%. The non-perturbative effects are reduced in the cross section ratio.

2.6.3   Electroweak corrections (EWK)

The fixed-order QCD calculations do not include contributions arising from virtual exchanges of massive W and Z bosons. The electroweak corrections (EWK) are calculated by [?] and are applied as a bin-by-bin correction factor to the fixed-order calculation of NLOJET++ as well as the MC predictions of MADGRAPH5 PYTHIA6. The correction factors in the phase space of the measurement are given in Figure 2.30 for inclusive 2-jet events. For inclusive 3-jet events are not calculated yet. The EWK increases up to 5% for $H_{\text{T},2}/2 > 1$ TeV and significantly improves the agreement between data and prediction. Since the guess from theory side is that EWK for inclusive 2-jet and 3-jet will be similar, so for R_{32}, it is assumed to be equal to the factor of 1.
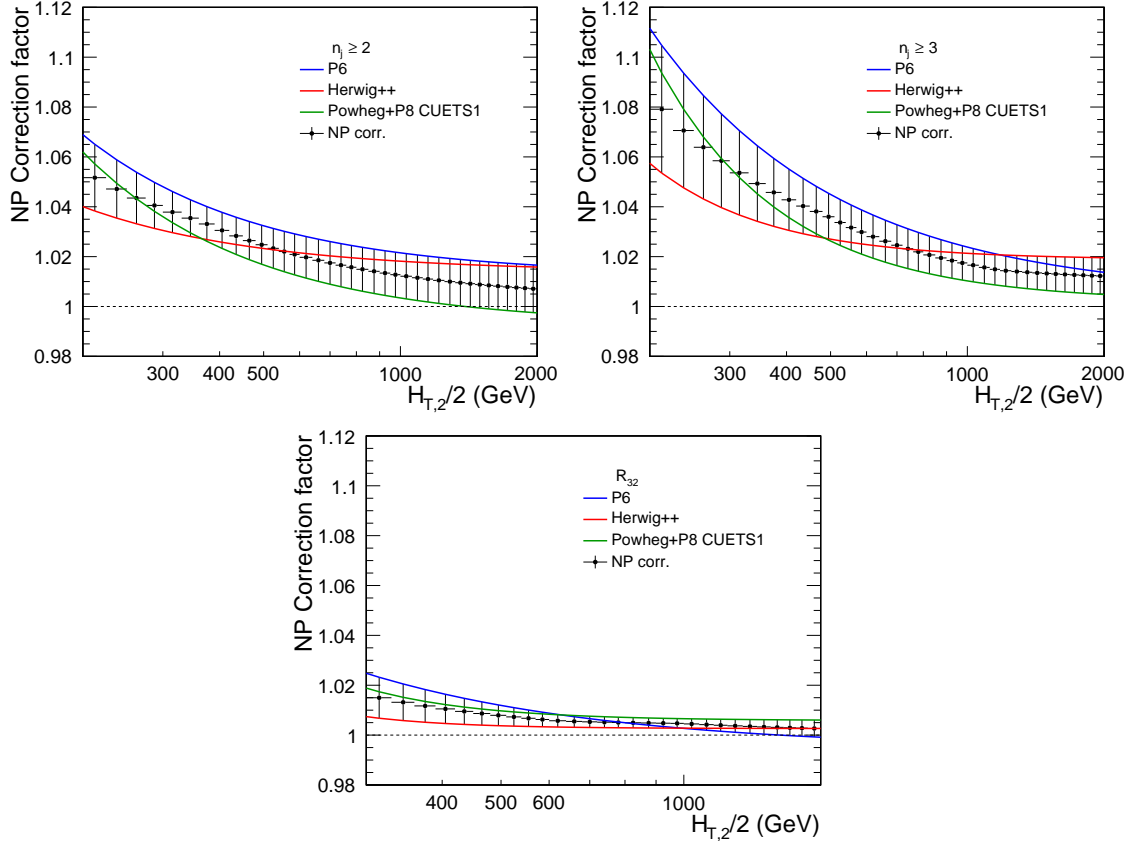
Figure 2.29: Fits to the nonperturbative corrections obtained for inclusive 2-jet (top left) and 3-jet (top right) event cross sections, as well as ratio R_{32}, as a function of $H_{T,2}/2$ for $|y| < 2.5$.
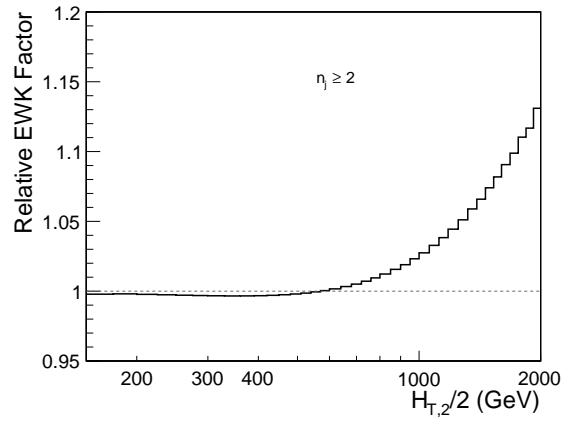
Figure 2.30: The electroweak corrections for inclusive 2-jet as a function of $H_{T,2}/2$.

2.6.4   Theory uncertainties

In this subsection, the derivation of theoretical uncertainties from various sources have been described. The total systematic theoretical uncertainties are evaluated as the quadratic sum of the scale, PDF, and NP uncertainties, and are tabulated in the table 2.8. Figure 2.31 gives the overview of systematic theoretical uncertainties affecting the cross section measurement for inclusive 2-jet (top left) and 3-jet events (top right) and their ratio R_{32} (bottom), using CT10 PDF set. The uncertainties on non-perturbative corrections have been already presented together with the obtained correction factors in Section 2.6.2.

The computation of the NLO predictions with NLOJET++ is also subject to statistical fluctuations from the complex numerical integrations. For the inclusive 2-jet event cross sections this uncertainty is smaller than about a per mille, while for the inclusive 3-jet event cross section it amounts to 1–9 per mille.

Higher orders of electroweak origin affect jet cross sections at large jet p_{T}. These electroweak corrections (EWK) have been calculated for the inclusive 1-jet and 2-jet case, cf. Ref. [**?**], but are not yet known for 3-jet production. Therefore, they are considered for the 2-jet events, while for the 3-jet event cross section and for the ratio they have been neglected.

Table 2.8: Overview of systematic theoretical uncertainties affecting the cross section measurement for inclusive 2-jet and 3-jet events using CT10 PDF set.

Uncertainty Source	Inclusive 2-jet	Inclusive 3-jet	R_{32}
Scale	5 to 13%	11 to 17%	6 to 8%
PDF	2 to 30%	5 to 50%	2 to 7%
NP	1%	1 to 2%	< 1%

2.6.4.1   Scale uncertainties

The uncertainty related to unknown higher orders of the perturbative series is evaluated with the conventional recipe of varying the default scale $H_{\mathrm{T},2}/2$ chosen for
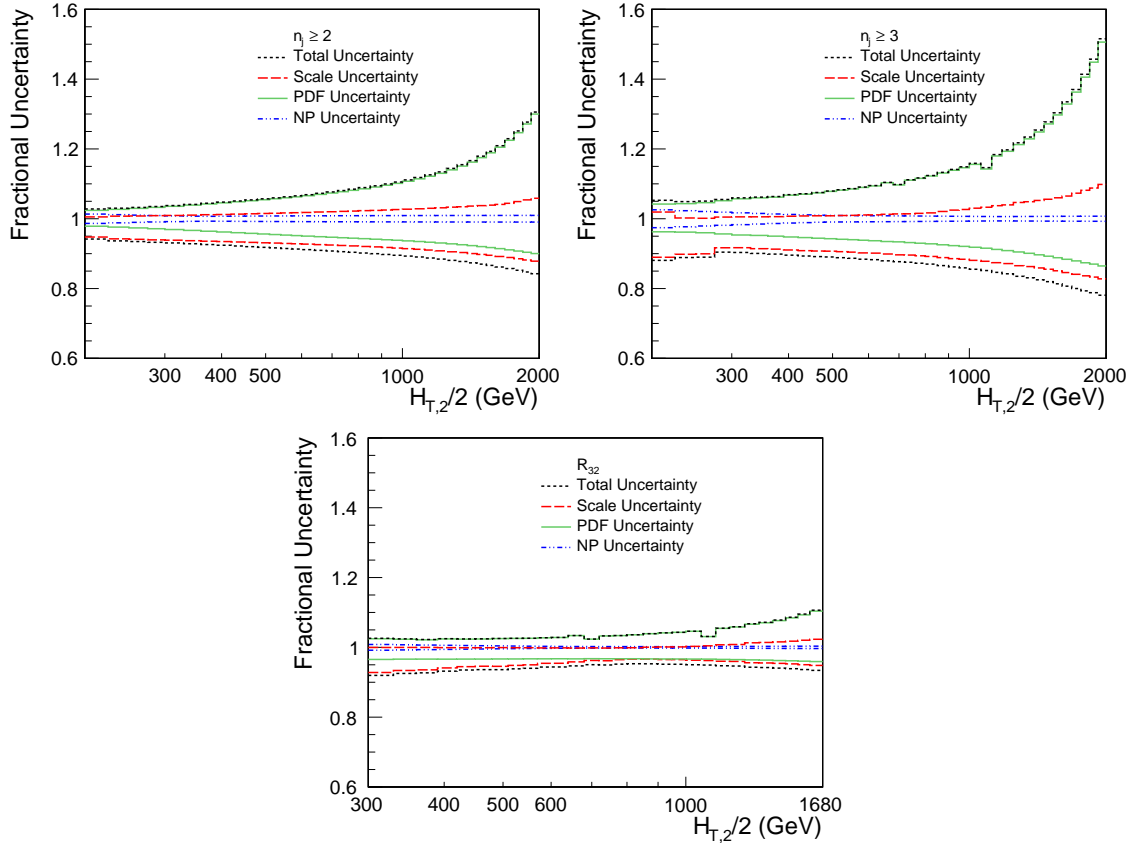
Figure 2.31: Overview of systematic theoretical uncertainties affecting the cross section measurement for inclusive 2-jet (top left) and 3-jet events (top right) and their ratio R_{32} (bottom) using CT10 PDF set. The total uncertainty is calculated by adding in quadrature the individual sources of uncertainty.

μ_r and μ_f independently in the following six combinations: $(\mu_r/H_{\mathrm{T},2}/2, \mu_f/H_{\mathrm{T},2}/2)$ = (1/2,1/2), (1/2,1), (1,1/2), (1,2), (2,1) and (2,2). The maximal upwards and downwards deviations in cross section from the central prediction are taken as scale uncertainty. This uncertainty ranges for inclusive 2-jet events from 5% to 13%, for inclusive 3-jet events from 11% to 17% and R_{32} from 6% to 8%.

2.6.4.2   PDF uncertainties

PDF uncertainties are evaluated according to the prescriptions given for each PDF set. Uncertainties on $\alpha_s(M_Z)$ are not considered, since this value is later on determined from a fit to the data. The PDF uncertainty as derived with the CT10 PDF set ranges from 2% to 30% for inclusive 2-jet events and from 5% to 50% for inclusive 3-jet events and from 2% to 7% for R_{32}.

2.7   Comparison of theory to data

Figure 2.32 shows the measured inclusive 2-jet and 3-jet event cross sections as a function of $H_{\mathrm{T},2}/2$ after unfolding for detector effects. On the left, the measurements are compared to the NLOJET++ predictions using the CT10 PDF set, corrected for NP effects and in addition for EWK effects in the 2-jet case. On the right, the comparison is made to the predictions from MADGRAPH5 + PYTHIA6 with tune Z2*, corrected for EWK effects in the 2-jet case. On a logarithmic scale, the data are in agreement with the NLO predictions over the whole range of $H_{\mathrm{T},2}/2$ from $300\,\mathrm{GeV}$ up to 2.0 (2-jet) and $1.68\,\mathrm{TeV}$ (3-jet) respectively.

For better visibility the ratios of data over the NLOJET++ predictions using the CT10 PDF set are shown in Figure 2.33. The data are well described by the predictions within their uncertainty, which is dominated at large $H_{\mathrm{T},2}/2$ by PDF effects in the upwards and by scale variations in the downwards direction. A trend towards an increasing systematic excess of the 2-jet data with respect to theory,
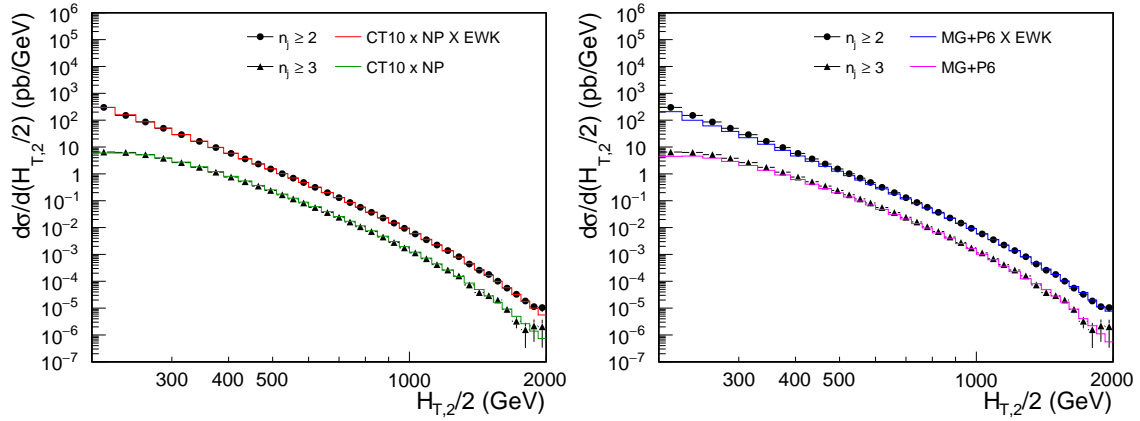
Figure 2.32: Comparison of the inclusive 2-jet and 3-jet event cross sections as a function of $H_{T,2}/2$ to theoretical predictions. On the (left), the data (points) are shown together with NLOJET++ predictions (line) using the CT10 PDF set, corrected for NP and EWK (2-jet) or only NP effects (3-jet). On the (right), the data (points) are compared to predictions from MADGRAPH5 + PYTHIA6 with tune Z2* (line), corrected for EWK effects in the 2-jet case. The error bars correspond to the total uncertainty, for which the statistical and systematic uncertainties are added in quadrature.

starting at about 1 TeV in $H_{T,2}/2$, is remedied by the inclusion of EWK corrections. In the 3-jet case the statistical precision of the data and the reach in $H_{T,2}/2$ is insufficient to observe any effect. The alternative PDF sets MSTW2008 and NNPDF2.3 exhibit a small underestimation of the cross sections at high $H_{T,2}/2$.

As for the NP corrections, the POWHEG framework providing a NLO dijet calculation matched to the parton showers of PYTHIA8 is used for a comparison. Here, POWHEG + PYTHIA8 are employed with the CUETS1 and CUETM1 tunes. The ratios of data over theory from POWHEG + PYTHIA8 with tune CUETS1 are shown in Figure 2.34. For comparison, the LO prediction from PYTHIA6, the tree-level multi-leg improved prediction by MADGRAPH5 + PYTHIA6, and the matched NLO prediction from POWHEG + PYTHIA8 with tune CUETM1 are shown as well. Significant discrepancies, which are cancelled to a large extent in the ratio R_{32}, are visible in the comparison with the LO prediction from MADGRAPH5 + PYTHIA6, in particular for small $H_{T,2}/2$. In contrast, the employed dijet MC PYTHIA8 and POWHEG + PYTHIA8 better describe the 2-jet event cross section, but fail for the
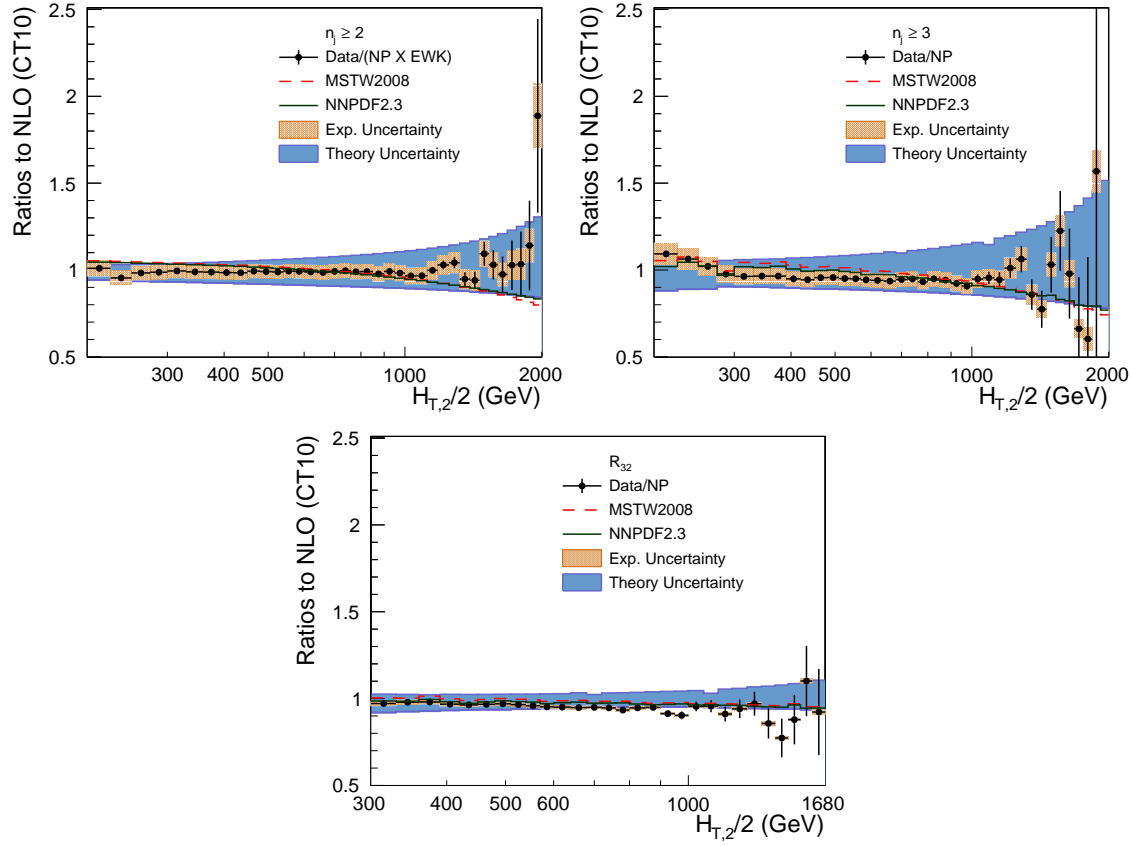
Figure 2.33: Ratio of data over theory using the CT10 PDF set for inclusive 2-jet (top left) and inclusive 3-jet event cross sections (top right) and their ratio R_{32} (bottom). For comparison predictions employing two other PDF sets are also shown. The error bars correspond to the statistical uncertainty of the data and the shaded rectangles to the total experimental systematic uncertainty. The shaded band around unity represents the total uncertainty of the theory.

3-jet ones. A more satisfactory description can be expected from 3-jet POWHEG predictions matched to PYTHIA8 for showering and hadronization.
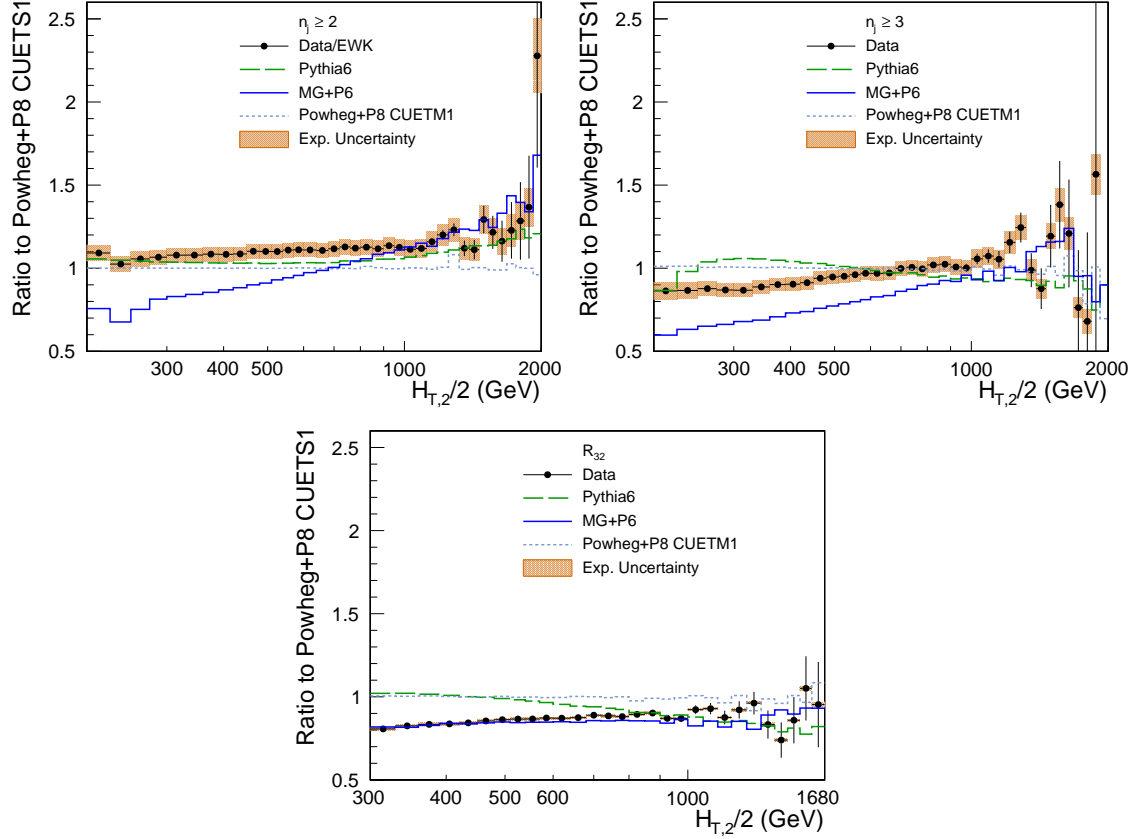
Figure 2.34: Ratio of data over the prediction from POWHEG + PYTHIA8 with tune CUETS1. For comparison the alternative tune CUETM1 of POWHEG + PYTHIA8, the tree-level multi-leg improved prediction by MADGRAPH5 + PYTHIA6 with tune Z2*, and the the LO MC predictions from PYTHIA6 tune Z2* are shown as well. The error bars correspond to the statistical uncertainty of the data and the shaded rectangles to the total experimental systematic uncertainty. EWK corrections have been accounted for in this comparison in the 2-jet case.

Figure 2.35 presents the cross section ratio R_{32}, as obtained from unfolded cross sections (blue soled circles), in comparison to that from NLO pQCD (CT10 PDF) (red dashed line) and to the ratio previously measured at 7 TeV [?].
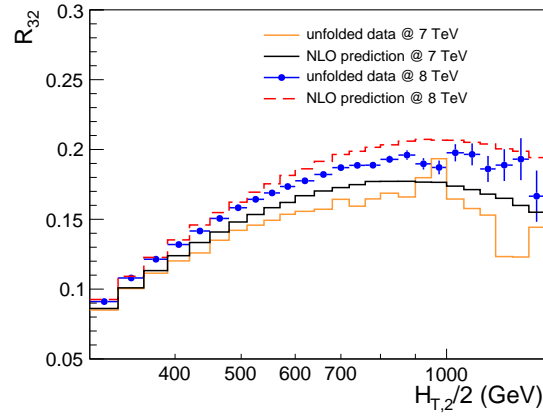
Figure 2.35: Cross section ratios R_{32} obtained from unfolded cross sections (blue solid circles), from NLO pQCD (CT10 PDF) (red dashed line), as a function of $H_{\mathrm{T},2}/2$ in comparison with the previously measured at 7.

Chapter 3

Extraction of the strong coupling constant

The inclusive jet production measurements at hadron colliders provides a direct probe to measure the strong coupling constant α_S. A previous measurement of α_S has been carried out from inclusive jet measurement at $\sqrt{s} = 7\text{TeV}$ [?] and also from Tevatron experiments [?,?]. The current analysis also broadly follows the same procedure for extraction of α_S.

The extraction of α_S is performed by fitting chi-square between measured $H_{\text{T},2}/2$ and theory prediction.

The differential inclusive jet production cross section upto at NLO is given by [?]:

$$\frac{d\sigma}{d(H_{\text{T},2}/2)} = \alpha_S^2(\mu_r)\hat{X}^{(0)}(\mu_f,(H_{\text{T},2}/2))[1 + \alpha_S(\mu_r)K1(\mu_r,\mu_f,(H_{\text{T},2}/2))]\,, \quad (3.1)$$

where $\frac{d\sigma}{d(H_{\text{T},2}/2)}$ is the differential inclusive-jet production cross section as a function of jet $H_{\text{T},2}/2$, μ_r and μ_f are the renormalization and factorization scales set equal to jet $H_{\text{T},2}/2$, $\alpha_S^2(\mu_r)\hat{X}^{(0)}(\mu_f,(H_{\text{T},2}/2))$ is the leading or-

der contribution to the differential inclusive-jet production cross section and $\alpha_S^3(\mu_r)\hat{X}^{(0)}(\mu_f, (H_{\mathrm{T},2}/2))K1(\mu_r, \mu_f, (H_{\mathrm{T},2}/2))$ is the NLO contribution. Equation 3.1 shows how the inclusive jet production cross section varies with $\alpha_S(\mu_r)$. This NLO cross section is corrected with the electroweak correction factors and NP correction factors.

For a fixed choice of μ_r, μ_f, different input values of α_S will lead to different theory predictions of the differential cross section distribution, giving an estimate of the sensitivity of the theory predictions with varying input value of α_S. A comparison with the measured spectrum gives an estimate of the input value of α_S for which the theory distribution has the closest matching with data.

The measured inclusive 2-jet and 3-jet event cross sections can be used for a determination of the strong coupling constant $\alpha_s(M_Z)$. In the following section, the sensitivity of the inclusive differential jet cross section and ratio, R_{32} to $\alpha_s(M_Z)$ is demonstrated and the fitting procedure is explained in details before presenting the outcome of the diverse fits of $\alpha_s(M_Z)$.

3.0.1 Sensitivity of the inclusive differential jet cross sections and their ratio, R_{32} to $\alpha_s(M_Z)$

First we have studied the sensitivity of the theory spectrum with varying α_S inputs for each PDF set. Figures 3.1, 3.2 and 3.4 present the ratio of data to the theory predictions for all variations in $\alpha_s(M_Z)$ available for the PDF sets CT10, CT14, MSTW2008, MMHT2014 and NNPDF2.3 at NLO evolution order as specified in Table 2.7, for inclusive 2-jet, inclusive 3-jet and ratio R_{32} respectively. The $\alpha_s(M_Z)$ value is varied in the range 0.112-0.127, 0.111-123, 0.110-0.130, 0.108-0.128 and 0.114-0.124 in steps of 0.001 for the CT10, CT14, MSTW2008, MMHT2014 and NNPDF2.3PDF sets, respectively. The error bars correspond to the total experimental uncertainty derived as quadratic sum from all uncertainty sources. For

brevity, the relative factor of NP between data and theory has been indicated as "Data/NP" in the legend. In Figure 3.1, the theory predictions are also corrected for EWK.

A small slope increasing with $H_{\mathrm{T},2}/2$ is visible for most PDFs in both cross sections. This effect is largely cancelled in the cross section ratio. R_{32} exhibits a flat behaviour with respect to the predictions for all five PDF sets in the whole range of $H_{\mathrm{T},2}/2$ up to 1.68 TeV. Therefore, these data can be used to determine the strong coupling constant, although only up to 1 TeV for the cross sections as long as electroweak corrections are not taken into account.

Moreover, the different sensitivity to $\alpha_s(M_Z)$ caused by the leading power in α_S in the expansion of the 2-jet inclusive ($\propto \alpha_S^2$) and the 3-jet inclusive cross section ($\propto \alpha_S^3$), and their ratio ($\propto \alpha_S^1$) is clearly visible from the spread between the calculations for the smallest and largest value of $\alpha_s(M_Z)$ within the same PDF set when passing through Figures 3.1–3.4. This also demonstrates the potential of ratios R_{mn} with $m - n > 1$.

3.0.2 The fitting procedure and fit results

As discussed in the previous section, the measured inclusive 2-jet and 3-jet event cross sections and their ratio R_{32} can be used for a determination of the strong coupling constant $\alpha_s(M_Z)$. The value of $\alpha_s(M_Z)$ is determined by minimizing the χ^2 between the experimental measurement and the theoretical predictions. The fit procedure here follows closely the one previously used in Refs. [?] and [?]. The χ^2 is defined as:

$$\chi^2 = M^T C^{-1} M \,, \tag{3.2}$$

where M is the vector of the differences between the data (D^i) and the theo-
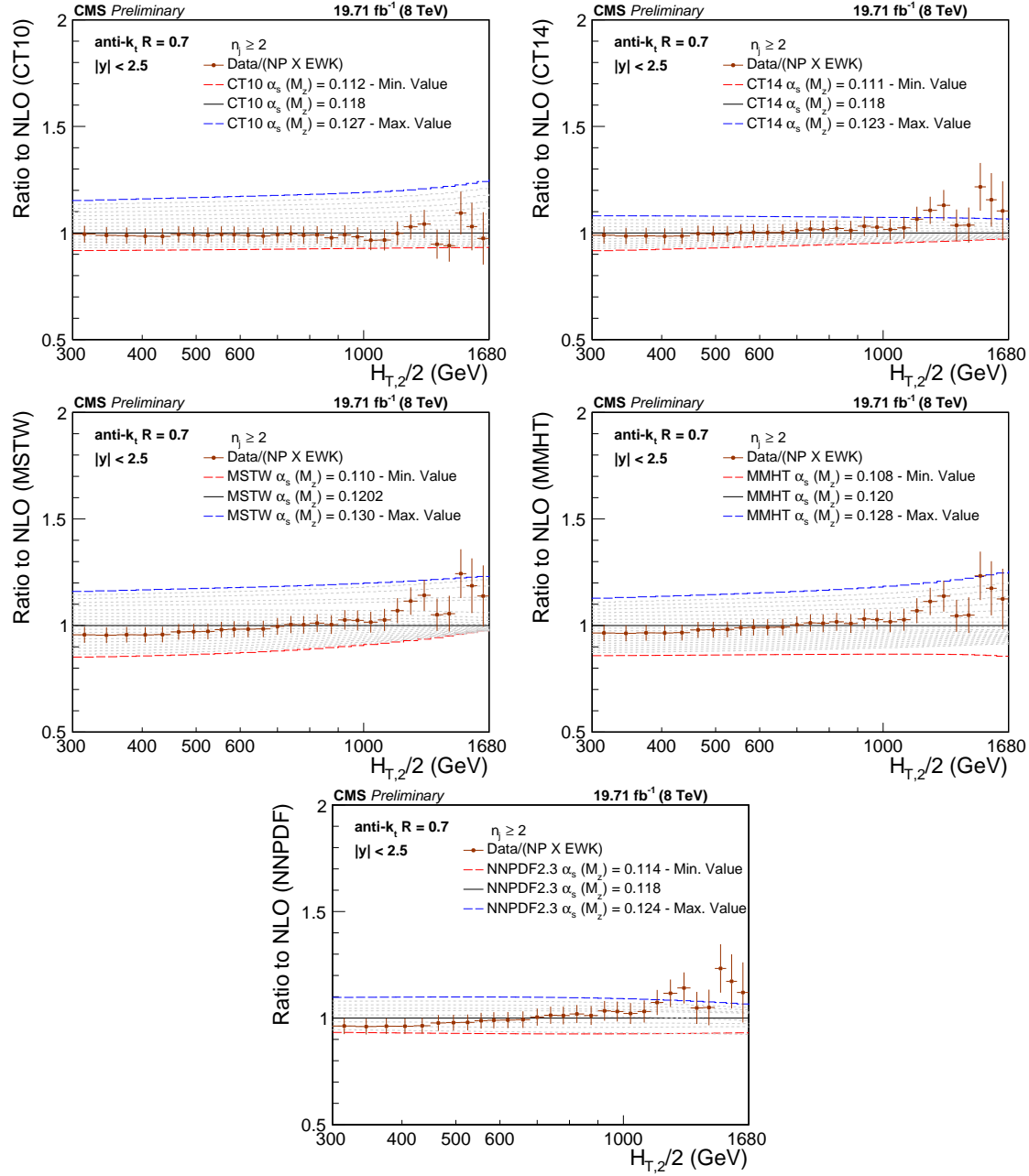
Figure 3.1: Ratio of the inclusive 2-jet differential cross section to theory predictions using the CT10 (top left), the CT14 (top right), the MSTW2008 (middle left), the MMHT2014 (middle right) and NNPDF2.3 (bottom) NLO PDF sets for a series of values of $\alpha_s(M_Z)$. The $\alpha_s(M_Z)$ value is varied in the range 0.112-0.127, 0.111-123, 0.110-0.130, 0.108-0.128 and 0.114-0.124 in steps of 0.001 for the CT10, CT14, MSTW2008, MMHT2014 and NNPDF2.3 NLO PDF sets, respectively. The error bars correspond to the total experimental uncertainty. For brevity, the relative factor of NP and EWK between data and theory has been indicated as "Data/(NP x EWK)" in the legend.
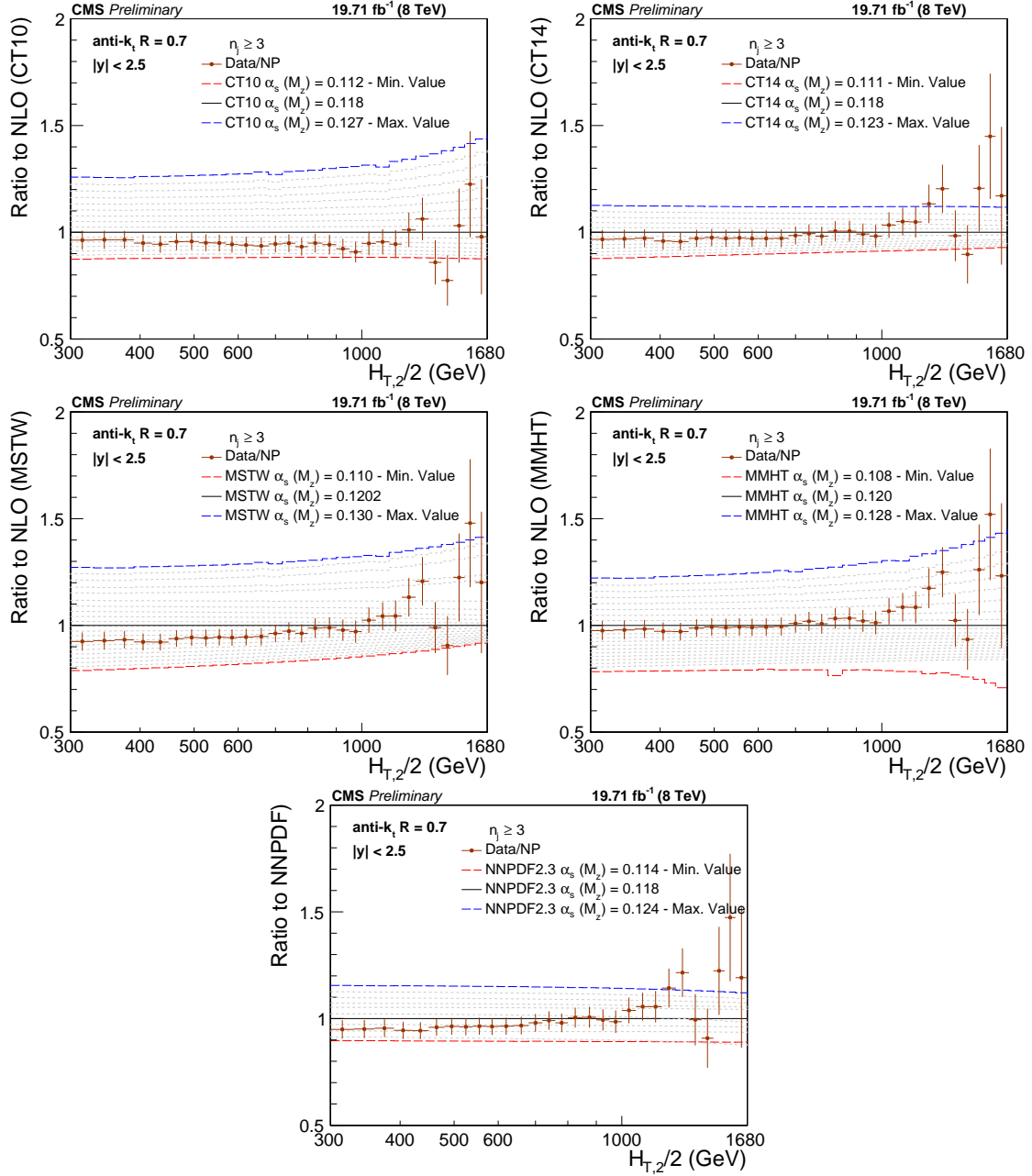
Figure 3.2: Ratio of the inclusive 3-jet differential cross section to theory predictions using the CT10 (top left), the CT14 (top right), the MSTW2008 (middle left), the MMHT2014 (middle right) and NNPDF2.3 (bottom) NLO PDF sets for a series of values of $\alpha_s(M_Z)$. The $\alpha_s(M_Z)$ value is varied in the range 0.112-0.127, 0.111-123, 0.110-0.130, 0.108-0.128 and 0.114-0.124 in steps of 0.001 for the CT10, CT14, MSTW2008, MMHT2014 and NNPDF2.3 NLO PDF sets, respectively. The error bars correspond to the total experimental uncertainty. For brevity, the relative factor of NP between data and theory has been indicated as "Data/NP" in the legend.
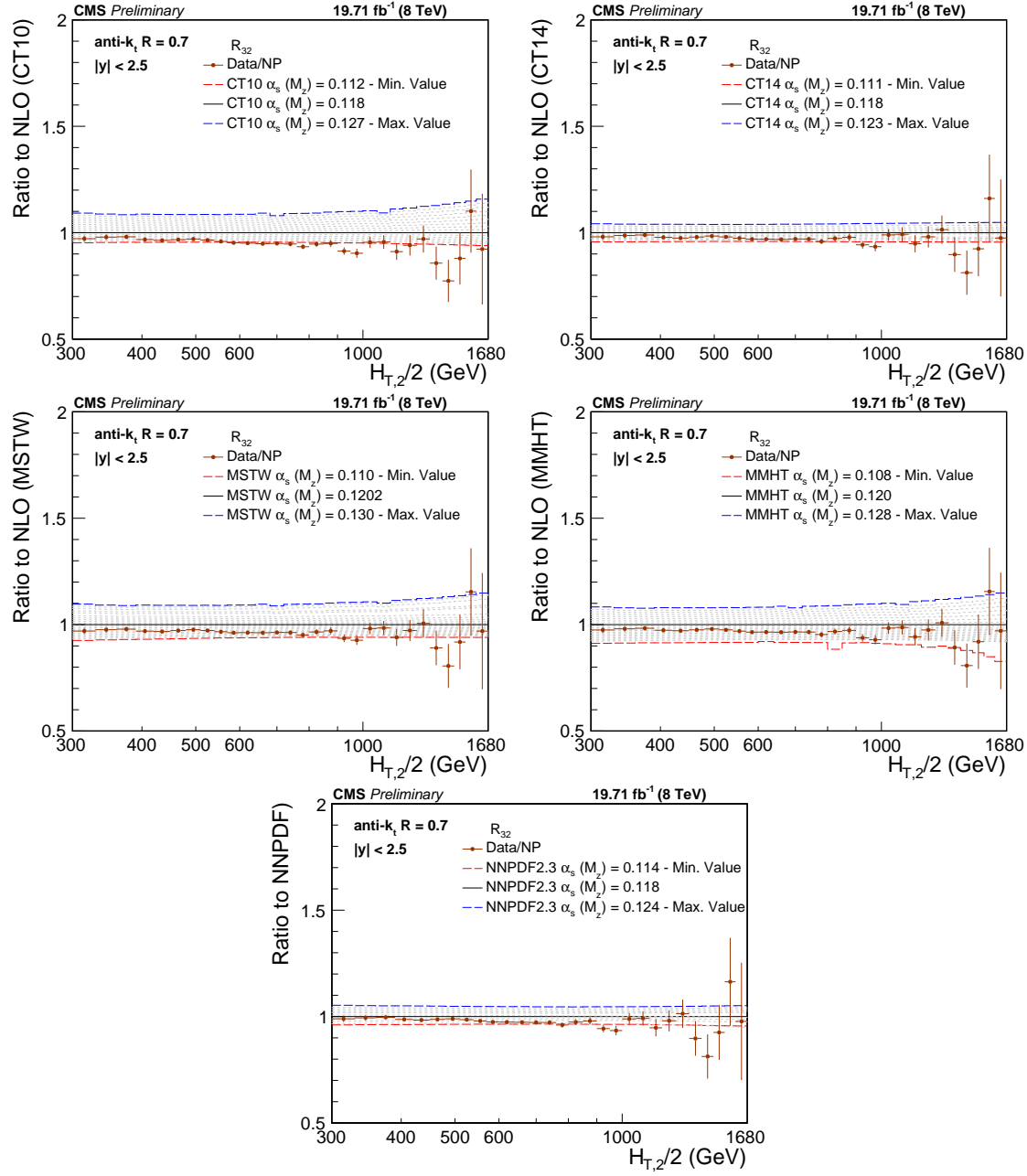
Figure 3.3: Ratio of the ratio, R_{32} to theory predictions using the CT10 (top left), the CT14 (top right), the MSTW2008 (middle left), the MMHT2014 (middle right) and NNPDF2.3 (bottom) NLO PDF sets for a series of values of $\alpha_s(M_Z)$. The $\alpha_s(M_Z)$ value is varied in the range 0.112-0.127, 0.111-123, 0.110-0.130, 0.108-0.128 and 0.114-0.124 in steps of 0.001 for the CT10, CT14, MSTW2008, MMHT2014 and NNPDF2.3 NLO PDF sets, respectively. The error bars correspond to the total experimental uncertainty. For brevity, the relative factor of NP between data and theory has been indicated as "Data/NP" in the legend.
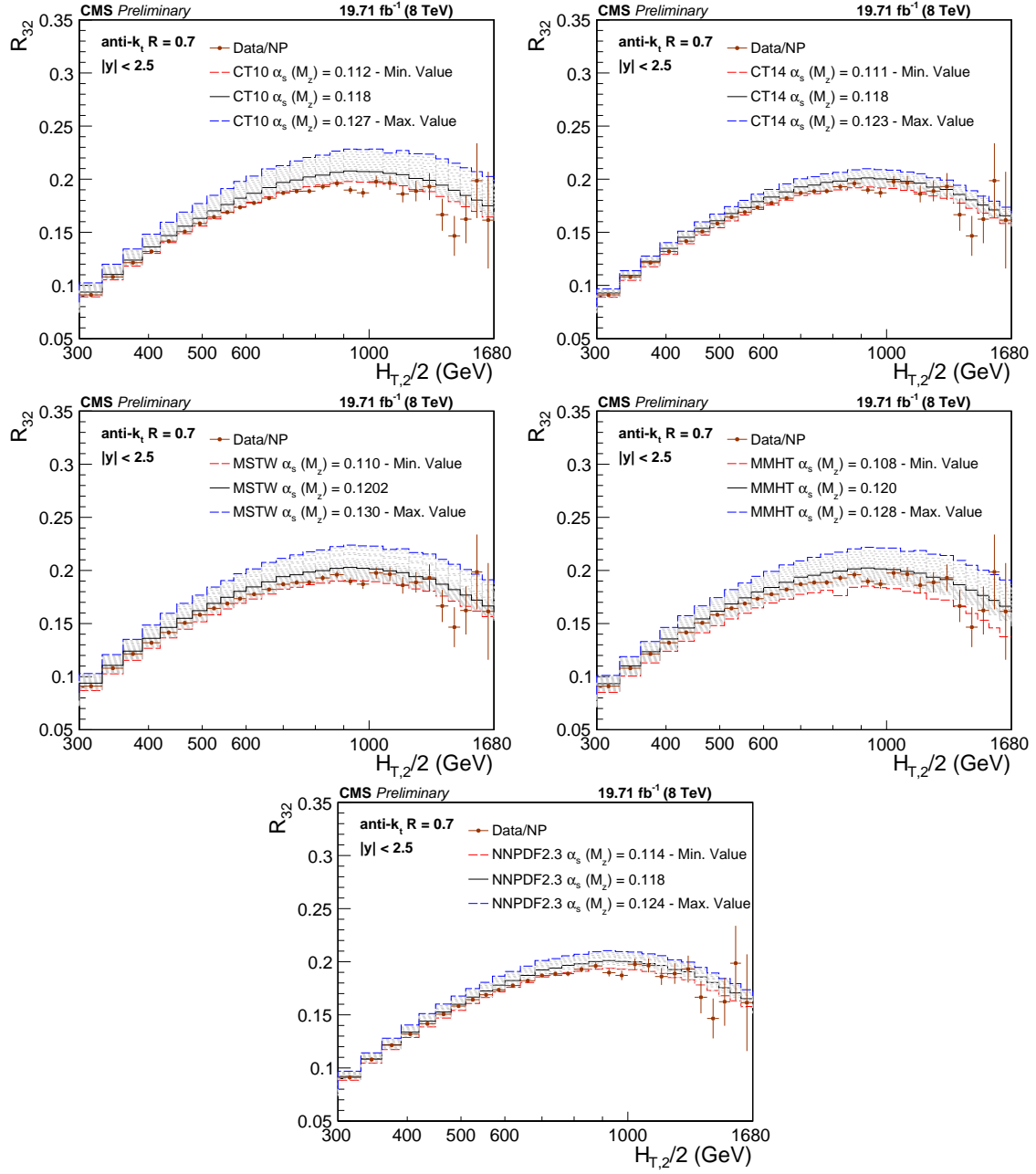
Figure 3.4: The NLO predictions using the CT10 (top left), the CT14 (top right), the MSTW2008 (middle left), the MMHT2014 (middle right) and NNPDF2.3 (bottom) NLO PDF sets for a series of values of $\alpha_s(M_Z)$, together with the measured R_{32}. The $\alpha_s(M_Z)$ value is varied in the range 0.112-0.127, 0.111-123, 0.110-0.130, 0.108-0.128 and 0.114-0.124 in steps of 0.001 for the CT10, CT14, MSTW2008, MMHT2014 and NNPDF2.3 NLO PDF sets, respectively. The error bars correspond to the total experimental uncertainty. For brevity, the relative factor of NP between data and theory has been indicated as "Data/NP" in the legend.

retical values (T^i) in each bin i,

$$M^i = D^i - T^i \tag{3.3}$$

and C is the covariance matrix including all experimental uncertainties as described in Section **??** and some theoretical uncertainties. More precisely, $C = C_{\mathrm{exp}} + C_{\mathrm{theo}}$ is defined as the sum of covariances of experimental and theoretical sources of uncertainty as follows

$$C_{\mathrm{exp}} = \mathrm{Cov}^{\mathrm{ExpStat}} + \sum \mathrm{Cov}^{\mathrm{JEC}} + \mathrm{Cov}^{\mathrm{Unfolding}} + \mathrm{Cov}^{\mathrm{Lumi}} + \mathrm{Cov}^{\mathrm{Uncor}} , \tag{3.4}$$

$$C_{\mathrm{theo}} = \mathrm{Cov}^{\mathrm{TheoStat}} + \mathrm{Cov}^{\mathrm{NP}} + \mathrm{Cov}^{\mathrm{PDF}} , \tag{3.5}$$

where the labelled covariance matrices account for the following effects:

- $\mathrm{Cov}^{\mathrm{ExpStat}}$: statistical uncertainty of the data including correlations introduced by the unfolding,

- $\mathrm{Cov}^{\mathrm{JEC}}$: the JEC systematic uncertainty,

- $\mathrm{Cov}^{\mathrm{Unfolding}}$: the unfolding systematic uncertainty including the JER,

- $\mathrm{Cov}^{\mathrm{Lumi}}$: the luminosity uncertainty,

- $\mathrm{Cov}^{\mathrm{Uncor}}$: a residual uncorrelated systematic uncertainty summarizing individual causes such as small trigger and identification inefficiencies, time dependence of the jet p_{T} resolution, and uncertainty on the trigger prescale factors,

- $\mathrm{Cov}^{\mathrm{TheoStat}}$: statistical uncertainty caused by numerical integrations in the cross section computations,

- $\mathrm{Cov}^{\mathrm{NP}}$: the systematic uncertainty of the NP corrections, and

- $\mathrm{Cov}^{\mathrm{PDF}}$: the PDF uncertainties.

The Unfolding, JEC, Lumi and PDF and NP systematics are treated as 100% correlated among $H_{T,2}/2$ bins. If δ_i is the total uncertainty on the double differential cross section, for the i-th $H_{T,2}/2$ bin, for any of these fully correlated sources, then the i, j-th element of the corresponding covariance matrix is given by

$$COV_{ij} = \delta_i \times \delta_j$$

In fits of the ratio R_{32}, the luminosity and residual uncorrelated uncertainties cancel completely. Partial cancellations between the other sources of uncertainty are taken into account in the fit.

The scale uncertainty is obtained by varying the renormalization and factorization scale by the six combinations of scale factors : $(\mu_F/\mu, \mu_R/\mu) = (0.5, 0.5)$, $(2, 2)$, $(1, 0.5)$, $(1, 2)$, $(0.5, 1)$, $(2, 1)$, where $\mu = H_{T,2}/2$. The largest upwards and downwards deviations from the default factors are defined as the scale uncertainty.

The treatment of PDF uncertainty depends on the individual PDF set. The PDF covariance matrix construction varies among different PDF sets. CT10 NLO PDF set has N_{ev} =26 eigen vectors with two PDF members viz. positive('+') and negative('-'). The PDF sets are denoted as $S_k^{\pm}$ for a given eigen vector k. Symmetric uncertainties due to PDF are then calculated using the formula [?]:

$$(\Delta X)^2 = \frac{1}{4} \sum_{k=1}^{N_{ev}} [X(S_k^+) - X(S_k^-)]^2 , \tag{3.6}$$

where $X(S_k^{\pm})$ are the double differential inclusive jet cross section for the $S_k^{\pm}$ PDF eigenvectors respectively. The CT10 uncertainties are downscaled by a factor of 1.64 in order to have the uncertainties at the 68.3% confidence level CL(1σ) instead of 90% CL(2σ).

The NNPDF2.3 PDF set comes with hundred different replicas (N_{rep}) instead

of different eigenvectors, as for the previously discussed PDF sets. The mean uncertainty is calculated as average uncertainty from 100 different replicas. Following the prescription given in [**?**], the PDF uncertainty is calculated as :

$$(\Delta X)^2 = \frac{1}{N_{rep} - 1} \sum_{k=1}^{N_{rep}} [X(S_k) - \langle X(S_k) \rangle]^2 \,, \tag{3.7}$$

where ΔX is the uncertainty on predicted double differential cross section, $X(S_k)$ is the double differential cross section for k-th replica and $\langle X \rangle$ is the average double differential cross section from all the replicas.

The MSTW-2008 NLO PDF sets are also used to cross-check the $\alpha_s(M_Z)$ extraction. The PDF uncertainties are derived in the same way as CT10, using 20 '+' and '-' eigenvectors.

First, fits to the cross sections are performed, where the range in $H_{\mathrm{T},2}/2$ is restricted to be between $300\,\mathrm{GeV}$ and $1\,\mathrm{TeV}$ to avoid the region close to the minimal p_T threshold of $150\,\mathrm{GeV}$ for each jet at low p_T and the onset of electroweak effects at high p_T, which are available for the dijet case only. The results are reported in Table 3.1 for the 2-jet and 3-jet event cross sections and in Table 3.2 for a simultaneous fit of both cross sections and for their ratio. In the ratio, EWK effects are assumed to cancel as do the luminosity and the uncorrelated uncertainty.

All cross section fits give compatible values for $\alpha_s(M_Z)$ in the range of 0.115–0.118; for the ratio R_{32} somewhat smaller values are obtained. A common issue, except for the ratio fits, is the rather small χ^2/n_{dof}. A possible explanation is an overestimation of the residual uncorrelated uncertainty of 1% that is cancelled for R_{32}. If the fits are repeated with an assumed uncertainty of 0.25% instead, the χ^2/n_{dof} values lie around unity while the $\alpha_s(M_Z)$ values are still compatible with the previous results but with slightly reduced uncertainties.

To investigate how the EWK corrections affect the fit results for $\alpha_s(M_Z)$, the

Table 3.1: Determination of $\alpha_s(M_Z)$ from the inclusive 2-jet and 3-jet event cross sections using five PDF sets at NLO. Only total uncertainties without scale variations are quoted. The results are obtained from a simultaneous fit to all 19 $H_{\mathrm{T},2}/2$ bins in the restricted range of $0.3 < H_{\mathrm{T},2}/2 < 1.0\,\mathrm{TeV}$.

PDF set	2-jets			3-jets		
	$\alpha_s(M_Z)$	$\pm\Delta\alpha_s(M_Z)$	χ^2/n_{dof}	$\alpha_s(M_Z)$	$\pm\Delta\alpha_s(M_Z)$	χ^2/n_{dof}
CT10	0.1174	0.0032	3.0/18	0.1169	0.0027	5.4/18
CT14	0.1160	0.0035	3.5/18	0.1159	0.0031	6.1/18
MSTW2008	0.1159	0.0025	5.3/18	0.1161	0.0021	6.7/18
MMHT2014	0.1165	0.0034	5.9/18	0.1166	0.0025	7.1/18
NNPDF2.3	0.1183	0.0025	9.7/18	0.1179	0.0021	9.1/18

Table 3.2: Determination of $\alpha_s(M_Z)$ from the inclusive 2-jet and 3-jet event cross sections simultaneously and from their ratio R_{32} using five PDF sets at NLO. Only total uncertainties without scale variations are quoted. The results are obtained from a simultaneous fit to all 38 (19) $H_{\mathrm{T},2}/2$ bins in the restricted range of $0.3 < H_{\mathrm{T},2}/2 < 1.0\,\mathrm{TeV}$.

PDF set	2- & 3-jets			R_{32}		
	$\alpha_s(M_Z)$	$\pm\Delta\alpha_s(M_Z)$	χ^2/n_{dof}	$\alpha_s(M_Z)$	$\pm\Delta\alpha_s(M_Z)$	χ^2/n_{dof}
CT10	0.1170	0.0026	8.2/37	0.1141	0.0028	19./18
CT14	0.1161	0.0029	9.1/37	0.1139	0.0032	15./18
MSTW2008	0.1161	0.0021	11./37	0.1150	0.0023	21./18
MMHT2014	0.1168	0.0025	11./37	0.1142	0.0022	19./18
NNPDF2.3	0.1188	0.0019	15./37	0.1184	0.0021	12./18

range in $H_{\mathrm{T},2}/2$ is extended to $0.3 < H_{\mathrm{T},2}/2 < 1.68\,\mathrm{TeV}$. Table 3.3 reports the values

obtained for $\alpha_s(M_Z)$ from fits to the 2-jet event cross section in this range with or

without EWK factors. The largest impact is a reduction in χ^2/n_{dof}, which indicates

a better agreement when EWK effects are included. In addition, a tendency to

slightly smaller $\alpha_s(M_Z)$ values is observed without the EWK corrections. For the

ratio R_{32} it is expected that these effects are much reduced.

Table 3.3: Determination of $\alpha_s(M_Z)$ from the inclusive 2-jet event cross section using five PDF sets at NLO with (right) and without (left) EWK corrections. Only total uncertainties without scale variations are quoted. The results are obtained from a simultaneous fit to all 29 $H_{\mathrm{T},2}/2$ bins in the range of $0.3 < H_{\mathrm{T},2}/2 < 1.68\,\mathrm{TeV}$.

PDF set	2-jets, no EWK			2-jets, with EWK		
	$\alpha_s(M_Z)$	$\pm\Delta\alpha_s(M_Z)$	χ^2/n_{dof}	$\alpha_s(M_Z)$	$\pm\Delta\alpha_s(M_Z)$	χ^2/n_{dof}
CT10	0.1163	0.0034	15./28	0.1165	0.0032	14./28
CT14	0.1137	0.0033	24./28	0.1144	0.0033	17./28
MSTW2008	0.1093	0.0028	27./28	0.1133	0.0023	19./28
MMHT2014	0.1127	0.0032	32./28	0.1141	0.0032	21./28
NNPDF2.3	0.1162	0.0024	31./28	0.1168	0.0024	23./28

From Figure 3.4 follows that only the PDF sets MSTW2008 and MMHT2014

provide a large enough range in $\alpha_s(M_Z)$ values to ensure fits without extrapolation.

The other three PDF sets are at the limit such that reliable fits cannot be performed

for all scale settings and/or bins in scale $Q = H_{\mathrm{T},2}/2$. Tables 3.4–3.6 give the

complete results for MSTW2008 and MMHT2014 for the full range in $H_{\mathrm{T},2}/2$ of

$0.3\,\mathrm{TeV}$ up to $1.68\,\mathrm{TeV}$, for scale variations in this range, and for subranges in $H_{\mathrm{T},2}/2$.

Using the MSTW PDF set, which dates from before the LHC start, the strong

coupling constant finally is determined to

$$\alpha_s(M_Z) = 0.1150 \pm 0.0010\,(\mathrm{exp})\ \pm 0.0013\,(\mathrm{PDF})\ \pm 0.0015\,(\mathrm{NP})\,^{+0.0050}_{-0.0000}\,(\mathrm{scale}) \tag{3.8}$$

$$= 0.1150 \pm 0.0023\,(\text{all except scale})\,^{+0.0050}_{-0.0000}\,(\mathrm{scale})\,.$$

The MMHT PDF set, although using LHC jet data to determine the PDF parame-

ters, leads to a very similar result of

$$\alpha_s(M_Z) = 0.1142 \pm 0.0010 \,(\text{exp}) \,\pm\, 0.0013 \,(\text{PDF}) \,\pm\, 0.0014 \,(\text{NP}) \,^{+0.0049}_{-0.0006} \,(\text{scale})\tag{3.9}$$

$$= 0.1142 \pm 0.0022 \,(\text{all except scale}) \,^{+0.0049}_{-0.0006} \,(\text{scale}) \,.$$

Table 3.4: Determination of $\alpha_s(M_Z)$ from the ratio R_{32} using the two most compatible PDF sets MSTW2008 and MMHT2014 at NLO. The results are obtained from a simultaneous fit to all 29 $H_{\mathrm{T},2}/2$ bins in the full range of $0.3 < H_{\mathrm{T},2}/2 < 1.68\,\mathrm{TeV}$.

PDF set	R_{32}: $\Delta\alpha_s(M_Z)\cdot 1000$						
	$\alpha_s(M_Z)$	exp.	PDF	NP	all exc. scale	scale	χ^2/n_{dof}
MSTW2008	0.1150	± 10	± 13	± 15	± 23	$^{+50}_{-0}$	26./28
MMHT2014	0.1142	± 10	± 13	± 14	± 22	$^{+49}_{-6}$	24./28

Table 3.5: Fitted values of $\alpha_s(M_Z)$ using R_{32} in the $H_{\mathrm{T},2}/2$ range from 0.3 up to 1.68 TeV at the central scale and for the six scale factor combinations for the two PDF sets MSTW2008 and MMHT2014.

$\mu_r/H_{\mathrm{T},2}/2$	$\mu_f/H_{\mathrm{T},2}/2$	MSTW2008		MMHT2014	
		$\alpha_s(M_Z)$	χ^2/n_{dof}	$\alpha_s(M_Z)$	χ^2/n_{dof}
1	1	0.1150	26./28	0.1142	24./28
1/2	1/2	0.1165	77./28	0.1160	73./28
2	2	0.1120	18./28	0.1191	18./28
1/2	1	0.1150	53./28	0.1136	48./28
1	1/2	0.1150	30./28	0.1142	28./28
1	2	0.1155	23./28	0.1147	22./28
2	1	0.1180	19./28	0.1175	19./28

Table 3.6: Uncertainty composition for $\alpha_s(M_Z)$ from the determination of α_S from the jet event rate R_{32} in bins of $H_{\mathrm{T},2}/2$. The statistical uncertainty of the NLO computation is negligible in comparison to any of the other sources of uncertainty. Electroweak corrections, significant only at high $H_{\mathrm{T},2}/2$, are assumed to cancel between the numerator and denominator.

$H_{\mathrm{T},2}/2$	MSTW2008: $\Delta\alpha_s(M_Z)\cdot 1000$					MMHT2014: $\Delta\alpha_s(M_Z)\cdot 1000$				
(GeV)	$\alpha_s(M_Z)$	exp.	PDF	NP	scale	$\alpha_s(M_Z)$	exp.	PDF	NP	scale
300–420	0.1157	± 15	± 14	± 19	$^{+53}_{-0}$	0.1158	± 14	± 10	± 19	$^{+52}_{-0}$
420–600	0.1153	± 11	± 14	± 18	$^{+57}_{-0}$	0.1154	± 11	± 12	± 17	$^{+56}_{-0}$
600–1000	0.1134	± 13	± 16	± 19	$^{+52}_{-0}$	0.1140	± 12	± 12	± 18	$^{+45}_{-0}$
1000–1680	0.1147	± 29	± 17	± 18	$^{+63}_{-11}$	0.1154	± 25	± 14	± 15	$^{+56}_{-11}$
300–1680	0.1150	± 10	± 13	± 15	$^{+50}_{-0}$	0.1142	± 10	± 13	± 14	$^{+49}_{-6}$

Selected

Reprints