Forward–Central Jet Correlations in *pp* Collisions at CMS

Dissertation

zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von

Pedro Miguel Ribeiro Cipriano

aus Aveiro

Hamburg
2014

Gutachter/in der Dissertation: Dr. Hannes Jung
Dr. Thomas Schöener–Sadenius

Gutachter/in der Disputation: Dr. Thomas Schöener–Sadenius
Prof. Dr. Peter Schleper

Datum der Disputation: 28. Januar 2014

Vorsitzender des Prüfungsausschusses: Prof. Dr. Georg Steinbrück

Vorsitzender des Promotionsausschusses: Prof. Dr. Peter H. Hauschildt

Dekan des Fachbereichs Physik: Prof. Dr. Daniela Pfannkuche

Dekan der MIN-Fakultät: Prof. Dr. Heinrich Graener

Abstract

The azimuthal correlation between central and forward jets has been measured in proton–proton collisions at the LHC recorded with the CMS detector at a centre-of-mass energy of 7 TeV. Both jets were required to have a minimum transverse momentum of 35 GeV. The forward jet was reconstructed in the hadronic forward calorimeter within the pseudo–rapidity range $3.2 < |\eta| < 4.7$, while the central jet was limited to $|\eta| < 2.8$. The measurement is performed inclusively and also differentially for different separations in pseudo–rapidity, the largest separation allowed by the CMS detector being $\Delta\eta = 7.5$ units. The analysis is carried out for inclusive dijet events which is then divided into two sub-samples. In one of these sub-samples, an additional jet with $p_T > 20$ GeV in the pseudo–rapidity interval between the forward and the central jet is required, while in the second sub-sample events containing such jets are explicitly vetoed. Furthermore, observables describing additional jets outside of the pseudo-rapidity interval spanned by the dijet system are measured. Most of the results are found to be well described by DGLAP–based Monte Carlo event generators.

Zusammenfassung

Die azimutalen Korrelationen zwischen Jets im zentralen und Vorwärtsbereich wurden in Proton–Proton Kollisionen am LHC mit dem CMS Detektor und einer Schwerpunktsenergie von 7 TeV gemessen. Von beiden Jets wurde ein minimaler Transversalimpuls von 35 GeV verlangt. Der vorwärts produzierte Jet wurde im hadronischen Vorwärtskalorimeter in einem Pseudorapiditätsbereich von $3.2 < |\eta| < 4.7$ rekonstruiert, während der zentrale Jet im Bereich von $|\eta| < 2.8$ rekonstruiert wurde. Die Messung wurde inklusiv und differentiell durchgeführt, für verschiedene Bereiche in Jets Pseudorapidität. Der grösst mögliche Unterschied in Pseudorapidität erlaubt, durch die Geometrie des CMS Detektors, ist $\Delta\eta = 7.5$. Die Analyse wurde für inklusive Zwei–Jet–Ereignisse und für zwei Unterkategorien durchgeführt. In der ersten Unterkategorie wird ein zusätzlicher Jet mit $p_T > 20$ GeV im Pseudorapiditätsbereich zwischen dem zentralen und dem Jet im Vorwärtsbereich, verlangt. In der zweiten Unterkategorie werden die Ereignisse mit solchen zusätzlichen Jets explizit nicht ausgewählt. Weiterhin werden zusätzliche Grössen gemessen, die Jets ausserhalb des Pseudorapiditätsbereichs, definiert durch die zwei Jets, beschreiben. Die DGLAP–basierten Monte Carlo Ereignisgeneratoren liefern eine gute Beschreibung der Messungen.

Contents

Introduction

The human beings, or rather the presumptuous *homo sapiens sapiens* [1] classification they give to themselves, are mainly conditioned by one single factor. It may seem very hard to believe that this species, which has mastered many harsh environments with clever tools and solutions, could be so strongly influenced by this factor. The factor has a deep impact on all dimensions of their lives, from the most simple movement to the higher levels of abstraction. This factor is present from the youngest offspring to the most ancient specimen. It is always there!

This factor is fear.

Some humans fear death, some fear the future, some fear strangers, some fear darkness. Some spend a lifetime trying to overcome and conquer their fears. Humans fear many things, which they, in the end, bring together in one entity. All human beings, with different degrees of intensity, fear the unknown [2].

This fear is unsettling and disrupting, influencing even the simplest human task. Since the dawn of civilization, religions and other spiritual movements have created answers to this fear. For centuries, this highly polarized and comforting approach has been enough; however, when the renaissance movements changed the old ways, the power of religion decreased and with it also the human degree of belief in these answers. This fear cannot be cured. The only treatment for this fear of the unknown is knowledge.

When it comes to get reliable knowledge, the scientific method has been one of the major achievements of the human species. As the famous aphorism says: knowledge is power. Nowadays, science has many branches, which are impossible to master in one lifetime. Nevertheless, the most important questions are still unanswered. One of the biggest questions, almost as big as the meaning of life, is: what is the universe? This question sounds naive, simple and free from philosophical considerations, but the answer is not straight forward. Against this question, the scientist themselves, contend almost in vain! Hopefully, they do not contend in vain literally, as the knowledge is built on top of previous experiences, which then result in other experiences.

In attempt to partially answer this overwhelming question, physicists have developed the Standard Model of particle physics [3]. This theory is compatible with the results of many experiments, which attempt to explain most of the phenomena known to humans at MeV scale and smaller. And to better scrutinize this Standard Model they have build many experiments, in which billions of Euro have been spent and thousands of scientists have worked for decades. Despite the huge effort, it was not enough, so new machines had to be built and new ideas had to be put into practice. The Large Hadron Collider is one of these machines and it was built to improve the knowledge of the Standard Model while probing the TeV scale, by colliding protons at the energy and intensity frontier. Such collisions are recorded by state–of–the–art detectors and analysed extensively.

The proton, often represented with three quarks and three connecting gluons, may look very simple. However, it only looks simple in a broad perspective, because a closer and careful investigation of the proton will show a complex structure. When colliding protons, or rather their fundamental constituents, there is a certain probability to produce new particles if enough energy is supplied. In this respect, the Large Hadron Collider can create the largest energy density of the solar system, thus enabling the study of phenomena previously inaccessible.

The discovery of these new phenomena requires a good understanding of the strong force, which drives most of the phenomena present at the Large Hadron Collider. These new and unobserved effects are, sometimes, hidden among billions of other effects. Therefore, a good knowledge of these common effects is vital to find new physics. In addition, measuring these processes improves the knowledge of the Standard Model parameters. Among many effects, one of not–yet–solved riddles is the parton evolution dynamics. For given phase–space region, there are a set of equations to describe the parton evolution. However, only one of these set has been observed so far and it is expected that the Large Hadron Collider could reach new phase–spaces, so different evolution schemes can be tested.

One way to study these parton evolution dynamics is to measure the correlations between bunches of highly collimated particles, called jets, produced by the collisions. These correlations can be translated into a difference in the azimuthal angle between two jets, which is an observable quantity. These evolution equations yield different predictions for these correlations.

In this thesis, correlations between the two leading jets are investigated. One of the leading jets is required to be in the central region of the detector, another is required to be in the forward region. The study is further extended to three topologies: the inclusive, with no further requirements, the presence and the veto of an extra jet between these two leading jets. The phase–space is divided according to the separation in direction of these two leading jets and the measurement is carried as a function of the separation and the topologies. Furthermore, the transverse momenta spectra and the direction are measured for the jet between the two leading jets and the jet outside of the two leading jets.

A theoretical overview of the Standard Model and Quantum Chromodynamics, as well as the motivations for these measurements are presented in chapter 1. The experimental setup, comprising the Large Hadron Collider and its experiments, with a special focus on the Compact Muon Solenoid detector, are introduced in chapter 2. The chapter 3 gives an overview on the event generation and reconstruction techniques. The event selection and control distributions are presented in chapter 4. The methods applied to correct for detector effects and their results are shown in chapter 5. The unfolding techniques and outcome are introduced in chapter 6. In chapter 7, an extensive study of the systematic uncertainties is given. The results with comparisons to several Monte Carlo predictions are presented in chapter 8. In the last pages, a conclusion and outlook are given.

Chapter 1

The Standard Model and QCD

In this chapter, a brief presentation of the Standard Model (see section 1.1), which includes Quantumchromodynamics, the theory of strong interactions (see section 1.2) is given. More detailed information about renormalization (see section 1.3), factorization (see section 1.4), evolution equations (see section 1.5), hadronization (see section 1.6) and the underlying event (see section 1.7) is included. Finally, an overview on jets is presented in section 1.8 and motivation for azimuthal correlations and a summary of earlier measurements is shown in section 1.9.

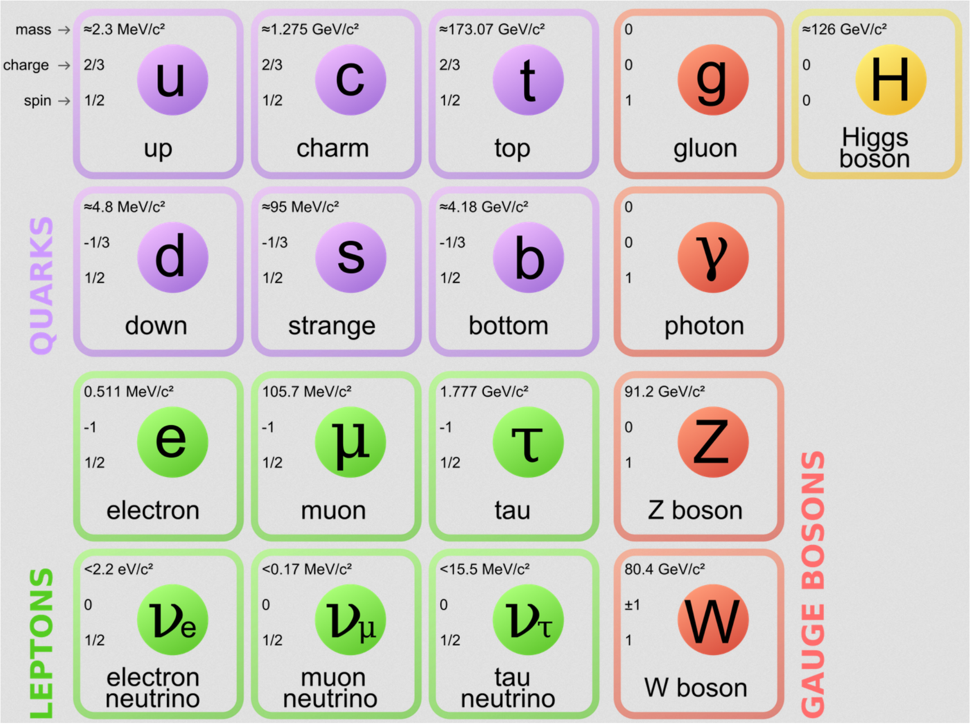
1.1 The Standard Model

The SM (**S**tandard **M**odel) [4–6] of particle physics is the most successful theory model ever made. The SM is based on the mathematical framework of Quantum Field Theory [7,8] and it was developed between the 50's and the 70's by many scientists. While the theoretists developed models, the experimentalists provided data to falsify or collect evidence of those models. The SM aims to be a complete theory, at the MeV scale and of higher energies, describing the electromagnetic, weak and strong interactions. It predicted the existence of W and Z bosons and the third generation of quarks and neutrinos many years before they were observed and so far it is consistent with all the experimental measurements. The Higgs boson [9–12] was the last particle predicted by the SM to be found [13,14] in 2012, thus completing the model. Despite its success, there are still many phenomena which it is not able to describe such as gravity, matter and anti–matter asymmetry, dark matter and dark energy.

The SM postulates the existence of the 17 elementary particles presented in figure 1.1. Twelve of them are fermions, particles with spin 1/2 which can combine to form the ordinary matter like stars and planets, represented in purple and green, and five are bosons, integer spin particles responsible for mediating the interactions between particles, represented in yellow and red.

The fermions are divided into groups according to how they interact: quarks and leptons. Represented in purple in figure 1.1, the quarks carry electrical and color charge, which allows them to interact via the strong force. A more detailed description is found in section 1.2. There are six quark flavours: up, down, charm, strange, top and bottom. The up, the charm and the top have electrical charge 2/3, while the down, the strange and the bottom have electrical charge -1/3. There are also six leptons: electron, electron neutrino, muon, muon neutrino, tau and tau neutrino, which do not carry color charge and are represented in green in figure 1.1. The neutrinos do not decay or carry electrical charge and only interact via weak force and consequently are hard to observe. Electrons, muons and taus have electrical charge and interact mainly through the electromagnetic force.

In figure 1.1, each one of the three fermion generations are represented by a column. The first generation is composed by up and down quarks, electron and electron neutrino. Except for neutrinos, each generation has similar properties, for example mass and decaying properties.



mass → charge → spin →	≈2.3 MeV/c² 2/3 1/2 u up	≈1.275 GeV/c² 2/3 1/2 c charm	≈173.07 GeV/c² 2/3 1/2 t top	0 0 1 g gluon	≈126 GeV/c² 0 0 H Higgs boson
QUARKS	≈4.8 MeV/c² -1/3 1/2 d down	≈95 MeV/c² -1/3 1/2 s strange	≈4.18 GeV/c² -1/3 1/2 b bottom	0 0 1 γ photon	
	0.511 MeV/c² -1 1/2 e electron	105.7 MeV/c² -1 1/2 μ muon	1.777 GeV/c² -1 1/2 τ tau	91.2 GeV/c² 0 1 Z Z boson	GAUGE BOSONS
LEPTONS	<2.2 eV/c² 0 1/2 ν_e electron neutrino	<0.17 MeV/c² 0 1/2 ν_μ muon neutrino	<15.5 MeV/c² 0 1/2 ν_τ tau neutrino	80.4 GeV/c² ±1 1 W W boson	

Figure 1.1: Particles of the Standard Model of particle physics [15].

The mass increases and the half–life time decreases with increasing generation. Each fermion row is grouped by electrical charge.

Bosons are the force carriers which mediate the particle interactions represented in figure 1.2 as lines connecting the particles. Photons enable electromagnetic interactions while gluons are responsible for the strong force, both are massless, have spin 1 and no electrical charge. The W and Z are spin 1 massive bosons which carry the weak force; the Z has no electrical charge, while the W has ±1 charge. The Higgs boson is massive, has no spin nor electrical charge and is responsable for the Higgs mechanism [9–12], which generates the particles' mass. A summary of the charateristics of the Standard Model bosons is presented in figure 1.1.

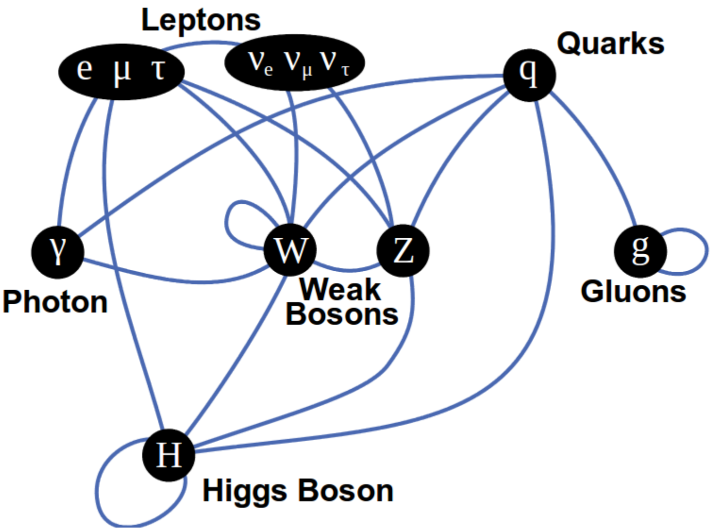


Figure 1.2: Interactions between the particles of the Standard Model [16].

In the SM, the interactions are described by local symmetry transformations U of the unitary group $SU(n)$. The unitary group is the group of $n \times n$ unitary matrixes, meaning that $UU^T = 1$ and $det(U) = 1$. Thus, an operator can be defined:

$$U(x) = e^{\left(i \sum_{a=1}^{n^2-1} \theta^a(x) T_a\right)}, \tag{1.1}$$

where θ^a are $n^2 - 1$ free parameters and T_a are generators. The generators are represented as traceless hermitian matrices, meaning that $tr(T_a) = 0$ and $T_a = T_a^+$, and they obey the commutation relation:

$$[T_a, T_b] = i f^c_{ab} T_c. \tag{1.2}$$

If the structure constant f^c_{ab} is zero, the generators comute, the corresponding bosons are not charged and the algebra is called *abelian*. The Standard Model obeys an $SU_C(3) \times SU_L(2) \times U_Y(1)$ symmetry, in which the $SU_L(2) \times U_Y(1)$ part corresponds to the electroweak sector.

1.2 QCD

The measurements carried out in this thesis belong to the category of QCD (**Q**uantum**C**hromo**D**ynamics) studies.

The symmetry group describing QCD is the non–abelian SU(3) group [17,18] responsible for the strong force. The generators do not commute and the mediators of the strong force carry colour charge [19]. The QCD SU(3) group has 8 free parameters, leading to eight gluons. Each gluon carries both a colour and an anti–colour charge. The nineth combination of one of the three colours and its anti–colour would lead to a colourless gluon.

The QCD Lagrangian [20] is given by:

$$\mathcal{L} = \sum_q \overline{\psi}_{q,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} \mathcal{A}^C_\mu - m_q \delta_{aq})\psi_{a,b} - \frac{1}{4} F^A_{\mu\nu} F^{A\,\mu\nu}, \tag{1.3}$$

where repeated indices are summed over. The γ^μ are the Dirac γ–matrices. The $\psi_{q,a}$ are quark–field spinors for a quark with flavour q, mass m_q and a colour index a, which runs from $a = 1$ to $N_C = 3$, corresponding to the three colours. The $\mathcal{A}^C_\mu$ correspond to the gluon fields, with C running from 1 to $N_C^2 - 1 = 8$, corresponding to the eight possible combinations of colour and anti-colour of the gluons. The t^C_{ab} are the eight 3×3 Gell–Mann matrices and are the $SU(3)$ generators. The field tensor $F^A_{\mu\nu}$ is given by:

$$F^A_{\mu\nu} = \delta_m u \mathcal{A}^A_{nu} - \delta_m u \mathcal{A}^A_{mu} - g_s f_{ABC} \mathcal{A}^B_\mu \mathcal{A}^C_\nu, \tag{1.4}$$

where the f_{ABC} are the structure constants of the $SU(3)$ group:

$$[t^A, t^B] = i f_{ABC} t^C. \tag{1.5}$$

Both quarks and gluons have never been observed as free particles. Due to the QCD property of confinement, they form hadrons, which are colour neutral composite particles. Hadrons made of two quarks are called mesons and hadrons with three quarks are baryons. However, at small distances, the quarks can be treated as free, which is a property called asymptotic freedom. If enough energy is supplied to a pair of particles, as in a hard collision, rather than breaking free, the pair creates and binds to the new particles, forming new pairs.

A more detailed description of QCD theory can be found in references [21,22].

1.3 Renormalization

Perturbative QCD describes the interaction of fundamental quarks using Feynman rules. When taking hadrons into account, non–pertubative QCD is necessary to describe them, because at the low–scale of the hadronization, it is not possible to distinguish between the different partons in a hadron.

In the perturbative regime, a partonic level cross–section of a specific process is given by:

$$\sigma = \int \frac{1}{f} \times |\mathcal{M}|^2 \times d\phi \tag{1.6}$$

where σ is the partonic cross–section, f is the incoming particle flux, $\mathcal{M}$ is the matrix element computed by Feynman rules and $d\phi$ is the phase–space volume integration variable.

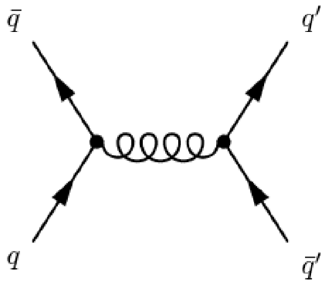


Figure 1.3: Leading–order Feynman Diagram for the $q\bar{q} \rightarrow q\bar{q}$ process.

The matrix element $\mathcal{M}$ describes the physics involved which can be visualised with Feynman diagrams. In figure 1.3 the leading order $q\bar{q} \rightarrow q\bar{q}$ process is presented. However, higher order corrections such as loops, and real and virtual corrections have to be taken in account. These additional contributions cause ultraviolet and infrared divergences because of the small–angle and soft radiation.

There are several regularization methods to cure the ultraviolet divergences that come from virtual and loop diagrams. One of the common approaches is the dimensional regularization. Each loop or virtual correction contributes with a term proportional to the strong coupling $\alpha_s = g_s/4\pi$. As each propagator of a loop or correction can contain loops or additional virtual corrections inside, diagrams in all orders of α_s contribute. The resulting expressions have a finite and a divergent part. In each order of α_s a counter term needs to be added to the Lagrangian to cancel the divergent parts, in a process called renormalization. The different renormalization methods differ in the choices of the arbitrary counter terms, which are specific for each application.

The result of a calculation using regularization depends on an arbitrary scale parameter μ, the renormalization scale. As a consequence, the strong coupling and quark masses become dependent on μ, but also any observable computed in all orders of α_s is independent on μ. Thus, renormalization accounts for effects at very small distances and by adjusting for example $\alpha_s(\mu)$ it is possible to remove physics happening at time scales $t \ll 1/\mu$ from the calculation. Therefore, the physical effects at scales smaller than $1/\mu$ are resummed.

1.3.1 Running Coupling

The renormalization group equation describes the dependence of the parameters influenced by the renormalization on μ and permits to compute the evolution of any parameter to a particular order. The dependence of α_s can be expressed as:

$$Q^2 \frac{d\alpha_s(Q^2)}{dQ^2} = \beta(\alpha_s(Q^2)) = -\alpha_s^2(\beta_0 + \beta_1\alpha_s + \beta_2\alpha_s^2 + ...), \tag{1.7}$$

where for μ the virtuality $Q^2 = k_T^2/(1-x)$ is used. The Taylor expansion coefficients β_i are used to derive $\alpha_s(Q^2)$ in pertubation theory up to a specified order. The value of α_s and its corrections in a given order are obtained by solving the renormalization group equation.

Using Λ_{QCD} as the reference scale, which is the scale where pertubative QCD breaks down, by setting the bonduary condition $\alpha_s(\Lambda_{QCD}) = \infty$, the NLO (**N**ext–to–**L**eading **O**rder) expression for α_s is:

$$\alpha_s(Q^2) = \frac{1}{1+\beta_0\omega(Q^2)}\left(1 - \frac{\beta_1 \log(\omega(Q^2))}{\beta_0^2\omega(Q^2))} + \frac{\beta_1^2}{\beta_0^4\omega(Q^2)}\left[\left(\log(\omega(Q^2)) - \frac{1}{2}\right)^2 - \frac{5}{4}\right]\right), \tag{1.8}$$

where $\omega(Q^2) = \log(Q^2/\Lambda_{QCD}^2)$.

The values of α_s can only be obtained from data, while the evolution can be computed analytically. At LO (**L**eading **O**rder), the running coupling $\alpha_s(Q^2)$ can be defined as:

$$\alpha_s(Q^2) = \frac{\alpha_s(Q_0^2)}{1+\beta_0\alpha_s(Q_0^2)\log(Q^2/Q_0^2)}, \tag{1.9}$$

where Q_0^2 is the reference scale to derive the values at other scales. Usually, M_Z^2 is used as reference value.

Often, the scale in α_s is chosen to be the virtuality Q^2 of the process of interest. If $\beta_0 > 0$, as in the case of six flavours, $\alpha_s(Q^2)$ increases as Q^2 decreases. As a smaller Q^2 corresponds to a larger distance, α_s increases with the distance of two coloured particles. This leads to the above defined confinement, in which free colour charges cannot be observed with a distance greater than $1/\Lambda_{QCD}$ due to the increasing force between them.

1.3.2 Higher–Order Corrections

In dimensional regularization, as the gluons are massless, in the case of virtual and real corrections, there is the possibility of soft emissions, in which the gluon may have a vanishing energy $E_g \to 0$. In addition, for real emissions the possibility of collinear contributions exists, in which the angle between the emitter and the gluon is very close to 0. Computing the diagrams for a certain order reveal infrared divergencies both on real and virtual parts. The divergencies cancel if both the real and the virtual contributions are taken in account. The order of the calculation of an observable is given by the inclusion of all diagrams with real and virtual corrections up to that order in α_s.

1.4 Factorization

The factorization theorem [23] states that the cross–section for a particular process can be reformulated as a convolution of the partonic cross–section with the parton density function. For a generic process, this theorem can be translated into:

$$\frac{\sigma(s,\tau)}{\sigma_0(s)} = f_a \otimes \hat{\sigma}^{ab} \otimes f_b = \int\int dx_1 dx_2 \sum_{a,b} f_a(x_1,Q)\,\hat{\sigma}^{ab}(x_1x_2\tau,\alpha_s(Q))\,f_b(x_2,Q) \tag{1.10}$$

where s is center–of–mass energy squared, while x_1 and x_2 are the momentum fraction of the incoming partons. The scaling variable for the physical process is $\tau = Q^2/s$, $f_{a,b}(x,Q)$ are parton distributions and $\hat{\sigma}^{ab}(\tau, \alpha_s(Q))$ is the corresponding cross–section for scattering of the two partons (a, b) at partonic center–of–mass energy squared equal to $s = x_1 x_2 s$. In this formula, all the factorization scales (μ) are equal to the physical large scale Q.

While investigating higher–order corrections in hadron–hadron collisions, one finds that infrared divergencies are not cancelled by the virtual corrections. The collinear gluon emissions, discussed in the previous section, come from a quark which is almost on its mass shell. These low–virtuality quarks are able to travel a long distance without any emission. Such emissions are not considered to be part of the hard interaction, but rather part of the incoming or outgoing radiation and thus treated separately.

In a complete α_s calculation, including all diagrams, up to a certain order, only the collinear emissions are not cancelled by the counter terms. The idea of factorization is to distinguish between the long distance physics, coming from almost real particles, and short distance, such as hard processes.

1.4.1 PDFs

One needs to define PDFs (**P**arton **D**ensity **F**unctions) in order to include the long distance effects, and thus, absorb the divergent part of the collinear emissions. As there is some arbitariness in the choice of the constant terms of the PDF, consequence of the different regularization schemes, one needs to introduce μ_F, the factorization scale, which separates between small and long distance physics. Very often, the regularization and the factorization scale are chosen to be the same. The PDF depends on regularization scheme and on the factorization scale. The parton splittings are dependent on the fraction z of the total momentum of the hadron.

One can look at the emission factor in the cross-section for two partons coming from a collinear splitting as a function of the relative transverse momentum, p_T (see definition section 2.3.1):

$$\frac{\alpha_s(p_\mathrm{T}^2)}{2\pi} \int\limits_{\mu_F^2}^{s} \frac{dp_\mathrm{T}^2}{p_\mathrm{T}^2} = \frac{\alpha_s(p_\mathrm{T}^2)}{2\pi} \log\left(\frac{s}{\mu_F^2}\right), \tag{1.11}$$

where $\sqrt{s}$ is the center–of-mass energy. Using this form, the the low–p_T contribution is factorized into the PDF.

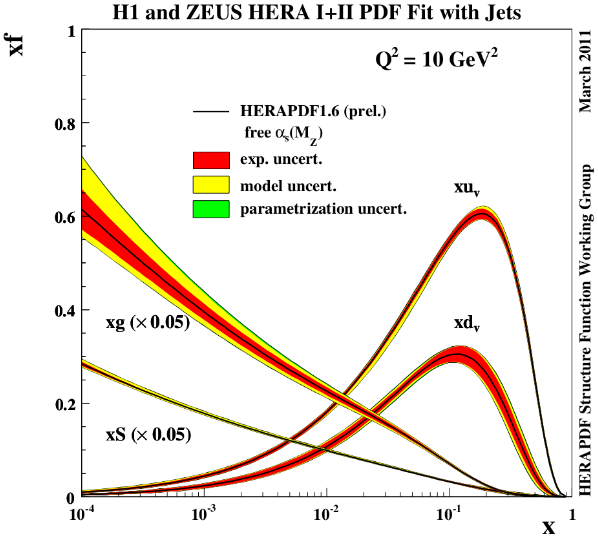


Figure 1.4: HERAPDF 1.6 PDFs obtained from the analysis of the ZEUS and H1 combined data at $Q^2 = 10$ GeV and $\alpha_s(M_Z)$ [24].

In practice, the PDFs are determined from data and used in the calculations as a set of values, dependent on x, Q^2 and parton flavour. There are different sets of PDF values available, obtained with distinct approximations and data samples, such as HERAPDF [25], CTEQ [26] and MSTW [27]. One example of such sets is presented in figure 1.4, which shows the parton density for different types of partons as function of x.

1.5 Evolution Equations

The parton evolution equations give the dominant contributions for the parton radiation. This is later modeled by the parton showers. There are several approaches, depending on the assumptions and on the evolution variable: DGLAP [29–33], BFKL [34–36] and CCFM [37–40]. The first two, which are schematized in figure 1.5, are discussed in this section.

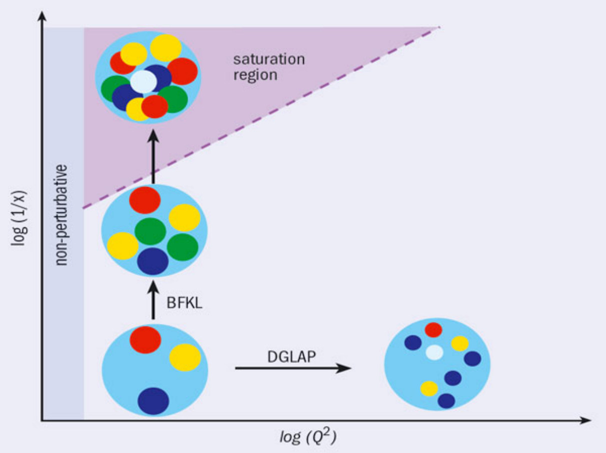


Figure 1.5: Diagram of the parton evolution equations [28].

1.5.1 DGLAP

The DGLAP evolution scheme is named after its authors (**D**okshitzer, **G**ibov, **L**ipatov, **A**ltarelli and **P**arisi) [29–33]. In the DGLAP scheme, the parton densities are evolved in the variable $Q^2 = k_T^2/(1-x)$, and only Feynman diagrams that give leading contributions in Q^2 or $\log(Q^2)$ are taken in to account. The DGLAP approximations is valid when $\log(Q^2) \gg \log(1/x)$.

As seen in figure 1.5, the parton density functions evolve in Q^2:

$$\frac{df_i(x,t)}{d\log(t)} = \frac{\alpha_s(t)}{2\pi} \sum_j \int_x^1 \frac{dx'}{x'} f_i(x',t) P_{i/j}(\frac{x}{x'}) \tag{1.12}$$

where the sum is over all parton types j that the parton i can branch to. $f_i(x', t)$ is the PDF of the flavour i carring a fraction x' of the original momentum at a given scale t. Every parton j has a probability $P_{i/j}(\frac{x}{x'})$ of originating from a parton of type i, where x is the fraction of the proton momentum carried by the parton i.

The plus distribution, denoted with the index + is defined so that its integral with any sufficient smooth distribution is:

$$\int_0^1 \frac{f(x)}{(1-x)_+} dx = \int_0^1 \frac{f(x) - f(1)}{1-x} dx \tag{1.13}$$

and

$$\frac{1}{(1-x)_+} = \frac{1}{1-x} \quad for \quad 0 \le x \le 1. \tag{1.14}$$

In the DGLAP scheme, the splitting functions $P_{i/j}(\frac{x}{x'})$ are:

$$P_{q/q}(x) = \frac{4}{3}\left[\frac{1+x^2}{(1-x)_+} + \frac{3}{2}\delta(1-x)\right], \tag{1.15a}$$

$$P_{q/g}(x) = \frac{1}{2}[x^2 + (1-x)^2], \tag{1.15b}$$

$$P_{g/q}(x) = \frac{4}{3}\left[\frac{1+(1-x)^2}{x}\right], \tag{1.15c}$$

$$P_{g/g}(x) = 6\left[\frac{x}{(1-x)_+} + \frac{1-x}{x} + x(1-x)\right] + \delta(1-x)\left[\frac{33-n_f}{6},\right] \tag{1.15d}$$

where $n_f = 6$ is the number of flavours and the $\delta(1-x)$ term arises from the virtual corrections. By imposing that the sum rules are not dependent on Q^2, the $\delta(1-x)$ term can be fixed with:

$$\int_0^1 P_{q/q}(x)dx = 0. \tag{1.16}$$

In the LLA (**L**eading **L**ogarithmic **A**pproximation), terms of the form $\alpha_s \log(Q^2)$ are typical results of the DGLAP evolution integrations. A DGLAP evolution to order n in α_s may be considered a summation over a ladder of n consecutive parton branchings, in which the propagating partons are strongly ordered in virtuality. This virtuality can be expressed as:

$$Q^2 = \frac{k_T^2}{1-x}. \tag{1.17}$$

For small values of x, a strong ordering in virtuality is also a strong ordering in transverse momentum squared:

$$k_{T_0}^2 \ll k_{T_1}^2 \ll ... \ll k_{T_{n-1}}^2 \ll k_{T_n}^2, \tag{1.18}$$

in which, the parent parton n carries a much larger transverse momentum than the originated parton $n-1$, having all parton transverse momenta virtualities limited below a fixed value Q^2. The $\alpha_s \log(Q^2)$ terms represent the contribution from the parton splitting diagrams in which the partons are strongly ordered in transverse momentum squared, k_T^2, with respect to the polar angle to the scattering.

1.5.2 BFKL

The BFKL evolution equations are named after their authors (**B**alitsky, **F**adin, **K**uraev and **L**ipatov) [34–36]. Unlike the DGLAP approach, no requirement on strong ordering in virtuality is made within the BFKL scheme. The evolution is made in x or $\log(\frac{1}{x})$, as can be seen in figure 1.5, for a ladder of partons strongly ordered in fractional momentum x_i:

$$x_0 \gg x_1 \gg ... \gg x_{n-1} \gg x_n. \tag{1.19}$$

At LLA, the BFKL equations are expected to contribute mainly in evolution at low x and are valid for the region $\log(1/x) \gg \log(Q^2)$.

The BFKL evolution equation has the form:

$$\frac{df(x,k_T)}{d\log(\frac{1}{x})} = \int dk_T'^2 K(k_T^2, k_T'^2) f(x, k_T'^2), \tag{1.20}$$

where k_T is the transverse momentum of the parent parton in a given splitting, while k_T' is the transverse momentum of its originated parton. $K(k_T^2, k_T'^2)$ is the splitting operator or kernel. The function $f(x, k_T^2)$ is the unintegrated gluon density:

$$xg(x,Q^2) = \int_0^{Q^2} \frac{dk'^2}{k'^2} f(x, k'^2) \tag{1.21}$$

in which, $g(x, Q^2)$ is the gluon density of the proton.

A random walk (see subsection 3.1.2) may lead to a diffusion of the ladder into the infrared region. As no limits are set on the possible energy of the emissions in the LLA evolution, energy conservation could be violated.

1.6 Hadronization

Hadronization is a process which transforms the coloured partons into colour–singlet hadrons. The transition happens at the hadronization scale, μ_h, which has a construction similar to the infrared cut of the parton showers. The two hadronization models more widely used are the String Model and the Cluster Model, which are briefly discussed in this section.

1.6.1 The String Model

At distances greater than one femtometer, the potential of the colour–dipole field between a charge and an anti–charge grows linearly with the separation of the charges. Such a property is called linear confinement and is the inspiration for the string model of hadronization [41, 42]. From that family of hadronization models, the most widely used is the Lund Model [43].

Consider a colour–connected quark and anti–quark pair, coming from the parton shower, with no intermediary gluons. As the partons move apart, linear confinement implies that a potential $V(r) = \kappa r$ is reached for large distances r. As the string connecting the parton stretches, the potential energy between the pair increases. The non–perturbative creation of an additional quark anti–quark pair can break the string, thus converting this potential energy into particles. This is illustrated in figure 1.6.

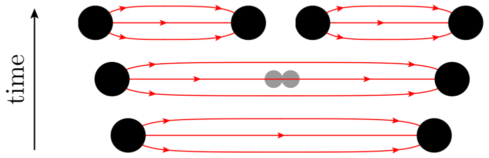


Figure 1.6: Illustration of string breaking by pair production in the string model.

In this model, the gluons are represented as kinks and build a transverse structure. The soft gluons are smoothly absorbed into the string. Gluons have one string on each side, increasing their relative energy loss by a factor of two. In the Lund Model, the string breaks are generated, starting with the leading outermost hadrons, containing the endpoint quarks. The iteration is done inwards, towards the center of the string, alternating randomly between the left and right side. In each step, a single on–shell hadron is produced, which then can decay.

1.6.2 The Cluster Model

The Cluster Model is based on the preconfinement effect [44, 45]. It was observed that the colour structure of a perturbative QCD shower at any scale μ_0 is such that colour–singlet subsystems of partons occur with a universal invariant mass distribution which only depends on μ_0 and Λ_{QCD} and not on the starting scale μ, when $\mu \gg \mu_0 \gg \Lambda_{QCD}$. This mass distribution is power–supressed at large masses. The HERWIG and HERWIG ++ generators use a cluster model for hadronization [46] based on this effect [47,48].

In the model, in addition to preconfinement, all gluons are forced to split into quark anti–quark pairs at the end of the parton shower. Comparing with the string model, this is equivalent to treating the gluons as seeds for string breaks rather than kinks. After the splitting, a new set of low mass colour–singlet quark plus anti–quark clusters is obtained.

If these clusters have an energy above the cutoff, they are forced to sequential cluster breakups along an axis defined by the constituent partons of the original cluster, until all clusters fall below the cutoff value. Then, all clusters are allowed to decay, starting with the lower masses.

1.7 The Underlying Event

One important characteristic of the pp collisions is the UE (**U**nderlying **E**vent). The UE is everything present in the event except the particles stemming from the hardest interaction. At hadron colliders, the main contributions are: pile–up, multiparton interactions, beam-beam remnants and initial- and final-state radiation. A schematic view is presented in figure 1.7.

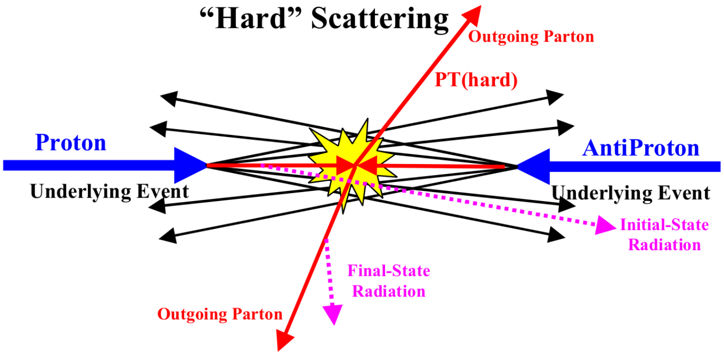


Figure 1.7: Underlying event in a hard scattering. Beam remnants are represented in black. The initial– and final–state radiation is represented in purple [59].

In figure 1.8, a schematic view of the MPI effect (**M**ulti**P**arton **I**nteraction) [60] is shown, which happens when there is more than one partonic interaction in an event. Usually, the extra interaction is softer, characterized by low p_T values, but it mixes with the final products of the hard scattering. MPI is invoked to explain the cross–sections of multi-jet production or the survival probability of large rapidity gaps in hard diffraction and event generators require it to sucessfully describe the UE at low p_T.

Pile–up takes place when more than a pair of protons collide. The number of pile–up collisions in an event is the number of proton pairs which have collided minus one, assuming that at least one pair collided. These extra collisions may yield also hard scattering products. A good tracker is required to separate the various collisions, by measuring the position of each of them with precision. The Pile–up is fundamentally different from the MPI, as the MPI is due to multiple interactions between the partons of the same proton and pile–up collisions are extra events happening in the same bunch crossing.

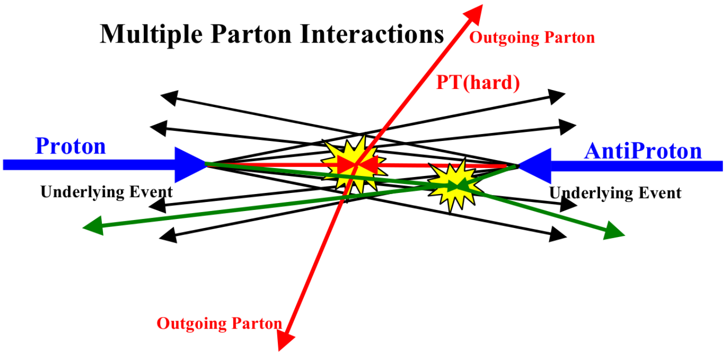


Figure 1.8: Multiparton interaction. The hard scattering is represented in red and the MPI in green [59].

After a scattering, the remaining quarks and gluons, coming from the colliding hadrons, will fly mostly towards the forward direction, the original beam direction, creating beam remnants, which carry the remaining energy of the colliding particles.

The initial– and final–state radiation will interfere with the hard scattering products by adding softer particles to the event.

Very often, the UE in hadron collisions is just a background and a nuissance, but it is still necessary to understand it properly to be able to measure the processes under study with a good level of precision.

1.8 Jets

Jet algorithms are tools to combine the large number of particles coming from hadronic final states into a smaller number of jets, making the analysis of the data easier. A jet is a highly collimated bunch of particles, which in many cases can even be identified in a visual inspection of the event. To make a comparison between experiments and also with theoretical models easier, algorithms were developed to make sure that equivalent quantities were compared. Thus, the jet algorithm aims to define quantities which have a relation with the parton level, by doing the inverse process from the parton shower. More detailed reviews on *jetography* in pp collisions can be found in [49, 50].

1.8.1 Algorithm Requirements

A jet algorithm should satisfy the following requirements [51]:

- Infrared and collinear safety: the result should not change if the original particle radiates a soft particle or if it splits into two collinear particles.
- Correspondence between the direction and energy of the original parton and the final jet.
- Small hadronization corrections and experimental uncertainties.
- Suppression of beam remnants contribution.
- Small experimental uncertainties.
- Simple to use both in experimental and theoretical analyses without any further modifications.

There are two main classes of jet algorithms: cone algorithms and sequential recombination algorithms.

1.8.2 Cone Algorithms

Cone algorithms were used for a long time at hadron colliders. Most of the iteractive cone algorithms start with some seed particle i, sum the momenta of all particles j within a cone of opening angle R, typically defined in terms of rapidity and azimuthal angle. The vectorial sum of all the j components is used as new seed and the process is repeated until the cone is stable. Then, if the resulting p_{T} is above a certain threshold, the sum of the elements j are considered a jet.

The variants of the cone algorithm differ in the set of seeds used and in the manner in which they ensure a one–to–one correspondence between particles to jets, as two cones may overlap. The use of seed particles is not compatible with infrared and collinear safety, causing problems with higher–order, and sometimes even leading–order, QCD calculations, in particular in multi–jet events. The cone algorithm jets are also prone to large non–pertubative corrections and instabilities. Thus, they are no longer recommendend, except for the seedless alternative: SISCone [52].

1.8.3 Sequential Recombination Algorithms

Sequential recombination algorithms were used in e^+e^- and ep experiments and are used at the LHC. In this type of algorithms one needs to define a distance between a pair of particles i, j:

$$d_{ij} = \frac{min(k_{t,i}^{2p}, k_{t,j}^{2p})\Delta_{ij}^2}{R^2}, \tag{1.22}$$

where Δ_{ij}^2 is the distance in the rapidity–azimuthal plane, k_t is the transverse momentum and R a free parameter. A "beam" distance is also defined as:

$$d_{iB} = k_{t,i}^{2p}. \tag{1.23}$$

The smallest d_{ij} and d_{iB} are identified. If it is a d_{ij}, the particles i and j are merged into a new pseudo–particle. If the smallest distance is d_{iB}, then i is considered a jet and removed from the list of particles. This process is repeated until all particles are part of a jet. As in the cone algorithms, only jets above a certain threshold are considered. The parameter p determines the type of algorithm:

- The inclusive k_t algorithm uses $p = 1$ [53–56];
- The Cambridge–Aachen algorithm sets $p = 0$ [57, 58];
- The anti–k_t algorithm is defined by $p = -1$ [65].

All these variants are infrared and collinear safe. While the two first ones lead to irregular jet shapes, the third gives cone–like boundaries. Thus, the anti–k_t algorithm became the standard choice in the LHC experiments.

1.9 Azimuthal Correlations

The azimuthal dijet correlations are measured in this thesis. These correlations can be observed as the azimuthal angle difference $\Delta\phi = |\phi_{jet1} - \phi_{jet2}|$. One jet is required to be in the central region of the detector and one in the forward region. The measurement of forward jets provides an important testing ground for QCD predictions of the Standard Model in the low–x region. The LHC can reach Q^2 and x values previously inaccessible to HERA [79], as

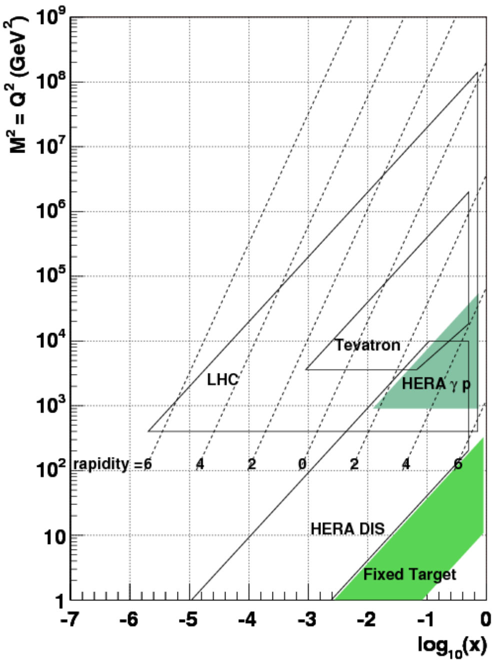


Figure 1.9: Kinematic phase–space accessible to fixed target, HERA, TEVATRON, LHC and fixed target experiments [80].

displayed in figure 1.9. To access the low–x region one must look at high pseudo–rapidity, which is equivalent to the forward region of the detector.

The central–forward dijets are an excellent probe for the parton ladder [61–64]. The multi–parton mechanism shifts a significant amount of gluon emisions to larger values of x, thus it can be distinguighed from a ladder with many emissions, which can display low–x effects. The azimuthal correlation, $\Delta\phi$ is sensitive to the k_T of the interacting parton. k_T comes from the parton evolution and can be different depending on k_T ordered or unordered cascade, and especially at small $\Delta\phi$, one expects to see effects of large k_T.

Azimuthal correlations in the central region of the detector are originated by high–x gluon processes, which are not sensitive to BFKL or CCFM effects. Forward jets are mostly originated from quark ladders and central jet from gluon ladders. Mueller–Navelet jets are a pair of jets, over a certain threshold, with largest rapidity separation. These Mueller–Navelet pairs can either originate from gluon, quark or mixed ladders, while the central jets come mostly from gluons. The effects from the parton cascade might be more clearly observed in the central and forward case, because there is a jet from a quark ladder and another form a gluon ladder, thus, both low and high–x scales can be probed simultaneously.

Earlier measurements of the azimuthal correlations in the central region of the detector have been preformed at the Tevatron [66] and by both the ATLAS [67] and the CMS [68] collaborations at the LHC.

At CMS, dijet azimuthal correlations over large dijet rapidity separations and for different p_{T} regions have been measured [68]. In general, the measured distributions are found to be well described by commonly-used DGLAP-based event generators and also by next-to-leading order QCD calculations, with the exception of the region of small values of dijet azimuthal difference $\Delta\phi_{dijet}$, which is dominated by events with many jets in the final state. Mueller–Navelet jets have been measured in several regions of the pseudo–rapidity [69], in which the DGLAP–based generators performed much better than the CCFM– or BFKL–based ones. The cross–section dijet ratio at large pseudo–rapidity separation $\Delta\eta_{dijet} = |\eta_{jet1} - \eta_{jet2}|$ has also been measured [70]. It was also found that the DGLAP–based event generators were able to describe the data. A review these of measurements can be found at [71].

At the TeV scale, jet rapidity and jet transverse momenta are well described by NLO calculations in perturbative QCD using the DGLAP approach and collinear factorization. When the collision energy $\sqrt{s}$ is considerably larger than the hard scattering scale given by the jet transverse momentum, p_T, calculations in perturbative QCD require a resummation of large $\log(1/x)$ terms. This leads to the prediction of new dynamic effects, expected to be described by Balitsky-Fadin-Kuraev-Lipatov (BFKL) evolution and k_T factorization [72–74]. An effective theory has been developed which describes strong interactions in this kinematic domain [75]. This description is particularly useful in events with several jets with large rapidity separation, which are not well described by DGLAP predictions.

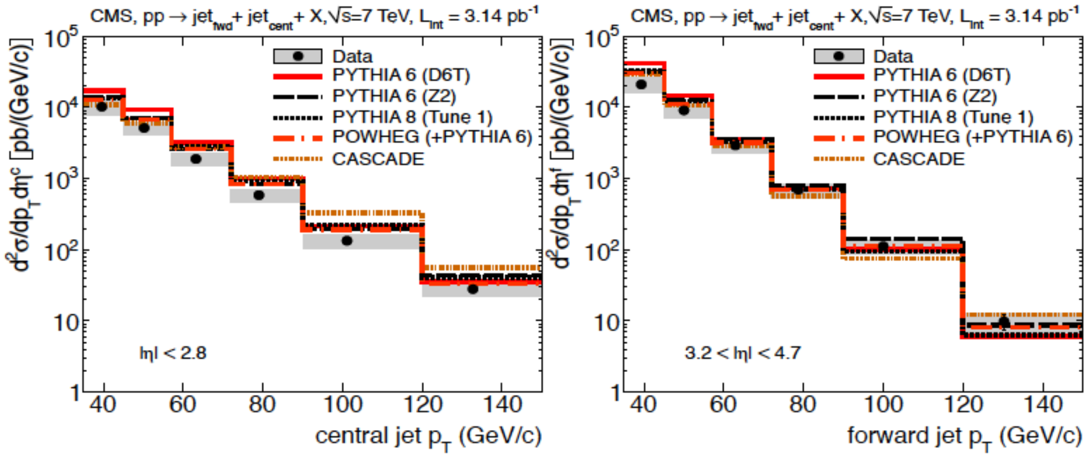


Figure 1.10: Central and forward jet p_{T} spectra at $\sqrt{s} = 7$ TeV measured with CMS [76].

This thesis focuses on a study of dijets which can have a large separation in pseudo–rapidity. In figure 1.10, an earlier CMS measurement of jet p_{T} spectra of such events is presented [76]. The advantage of this experimental setup is that the sensitivity to the details of the parton radiation along the parton ladder is maximised and that, consequently, different parton evolution schemes, such as BFKL and CCFM, might be experimentally confirmed or rejected. This is especially true for studies of the dijet decorrelation differentially in the pseudorapidity separation of the jets. The azimuthal correlation is expected to vary with the different pseudo–rapidity separations, showing an increased decorrelation as the central and forward jet are more apart and allow more phase–space for extra radiation. These decorrelations allow to test the general properties of models in a more or less well understood phase–space (small dijet separation) and then to look for effects of increasingly long parton ladders and many more possible emissions, which could reveal discrepancies between the different parton evolution approaches. Therefore, the measurement of the azimuthal correlation is done inclusively and also differentially in different pseudo–rapidity separations of the two jets constituting the dijet system.

More detailed studies of the hadronic final state, other than just dijet azimuthal correlations, might shed additional light on the emission dynamics along the parton ladder. One example are studies of the additional jet activity in the events. Measuring a hard emission between the two leading jets or outside them would allow the influence of such radiation on the dijet correlation to be investigated. The p_{T} and the pseudo–rapidity of these jets can also be measured. These observables are expected to exhibit additional sensitivity to different parton shower algorithms and to other QCD features like colour coherence effects or non-global observables [77, 78].

Chapter 2

The LHC and the CMS Detector

The LHC (**L**arge **H**adron **C**ollider) [81–84] is the only particle accelerator in operation which is able to explore the physics at the TeV scale. It is the machine that was used to produce the data for this high energy physics measurement.

In this chapter, the LHC machine and the LHC experiments are described, with special emphasis on the CMS (**C**ompact **M**uon **S**olenoid) [85–87] detector. In section 2.1, the technical details of the LHC are described and a summary of its operational history is given. A brief introduction on the LHC experiments is given in section 2.2. A detailed overview of the CMS detector is presented in section 2.3, where the components relevant for this analysis are described.

At the LHC, the particles do not travel in a continuous beam, but rather form groups, called bunches, which collide with a frequency f. When discussing collider physics, one must first define the instantaneous luminosity $\mathcal{L}$:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}, \tag{2.1}$$

where n_1 and n_2 are the number of particles in the two colliding bunches, respectively. The root–mean–square of the transverse beam sizes in the horizontal and vertical directions is characterized by σ_x and σ_y.

The number of delivered events, N, is the product of the cross-section σ and the integral in time over the instantaneous luminosity $\mathcal{L}$, called integrated luminosity:

$$N = \sigma \times \int \mathcal{L}\, \mathrm{d}t \tag{2.2}$$

Therefore, the number of collected events depends on the physical process, machine parameters and time. The luminosity is commonly measured in barns: 1 barn = 10^{-28} m^{-2}.

2.1 The LHC machine

CERN (European Organization for Nuclear Research, also known as **C**onseil **E**uropéen pour la **R**echerche **N**ucléaire), hosts the LHC. The LHC is used to accelerate and collide both protons and heavy ions at energy and luminosity frontiers. It has been assembled between 2000 and 2008 in the same tunnel as the former LEP (**L**arge **E**lectron **P**ositron) collider [88], built between 1984 and 1989. LEP has been in operation between 1989 and 2000 and the tunnel is located near to the Swiss–French border. The tunnel has a lenght of 26,659 km, a inside diameter of 3.8 m and is located between 45 and 170 m below the surface.

The protons for the collisions are supplied from electron–stripped hydrogen atoms and are accelerated by a chain of accelerators before they are injected into the LHC (see figure 2.1):

- in the LINAC 2 (**Lin**ear **Ac**celerator), the protons are accelerated to 50 MeV;

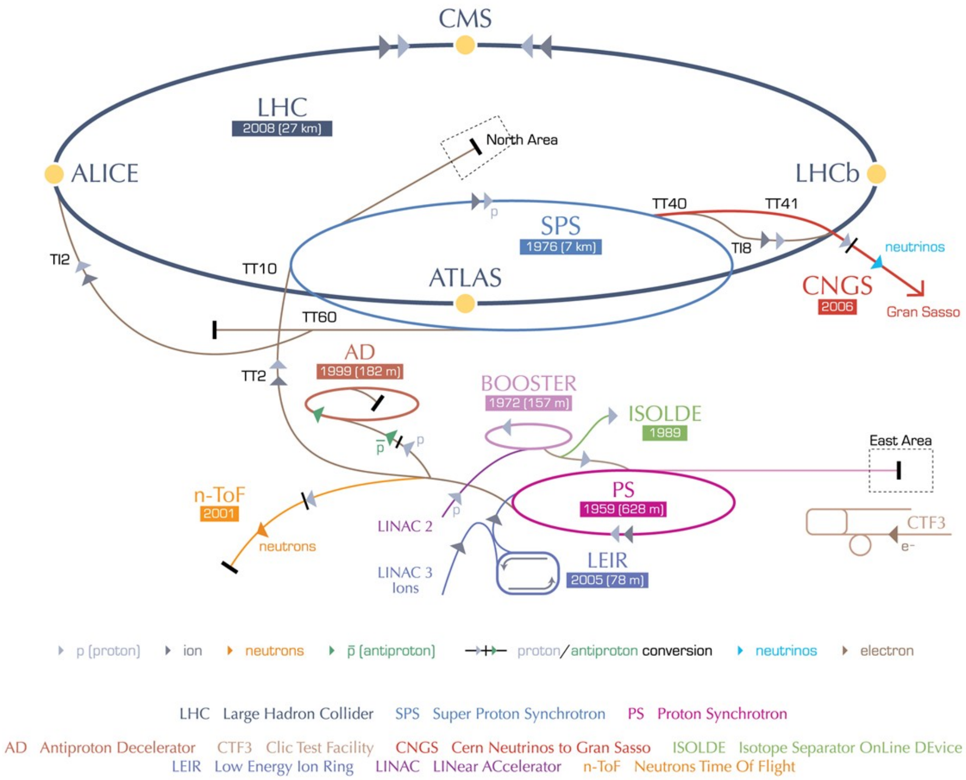


Figure 2.1: A schematic view of the CERN accelerator complex in 2008. Below each machine the name and year of completion is written. In case of circular accelerators the diameter is given [89].

- in the PSB (**P**roton **S**ynchrotron **B**ooster), the protons are accelerated from 50 MeV to 1.4 GeV;
- in the PS (**P**roton **S**ynchrotron), the protons are accelerated from 1.4 to 26 GeV;
- in the SPS (**S**uper **P**roton **S**ynchrotron), the protons are accelerated from 26 to 450 GeV.

When LHC collides ions, they are supplied by the Linac 3, accelerated in the LEIR (**L**ow **E**nergy **I**on **R**ing) and passed directly to PS.

The particles in both beams travel inside separate pipes. Nevertheless, both beam pipes are embedded in the same magnets, a decision that was made to save both space and construction costs. The beams cross at four interaction points, IP 1,2,5 and 8, where the large experiments are located: ATLAS, ALICE, CMS and LHCb. To prevent unwanted scattering, the beam pipes are kept at a vacuum of about 10^{-10} mbar. The energy of the protons is increased by 485 keV per turn by the RF (**R**adio **F**requency) cavities. Given that the protons travel almost at the speed of light, the frequency of revolution is about 11 kHz. In addition to the RF cavities, 1232 NbTi (Niobium-titanium) superconducting dipole magnets, cooled by super–fluid Helium II at 1.9 K, keep the protons in the correct trajectory. The task of keeping the beam focused is performed by 392 quadrupole magnets. Finally, at the crossing points, there are specific sextupoles and octopoles which steer and decrease the beam size in order to increase the probability of interaction. A summary of the LHC design parameters is presented in table 2.1.

The LHC can also collide lead ions up to a center of mass energy of 5.5 TeV or protons with heavy ions at $\sqrt{s} = 5$ TeV. The original schedule predicts that the LHC should run for about 10 years. Plans to upgrade both the luminosity up to 10^{35} cm^{-2} s^{-1} (HL-LHC - **H**igh **L**uminosity LHC) [92, 93] and then later the energy up to 33 TeV (HE-LHC - **H**igh **E**nergy LHC) [94] are under study. These upgrades are planned to take place in 2022, when the second LHC phase is over. The schematic view of the plan can be seen in figure 2.2.

Characteristic	Design	Achieved in 2010 [91]
Center–of–mass energy for *pp* collisions	14 TeV	7 TeV
Nominal magnetic field	8.33 T	≈ 4.55 T
Instantaneous luminosity	10^{34} cm^{-2} s^{-1}	10^{32} cm^{-2} s^{-1}
Number of bunches in the ring	2808	1326
Bunch crossing rate	40 MHz	20 MHz
Protons per bunch	$1.15\ .\ 10^{11}$	$7\ .\ 10^{10}$

Table 2.1: LHC design parameters

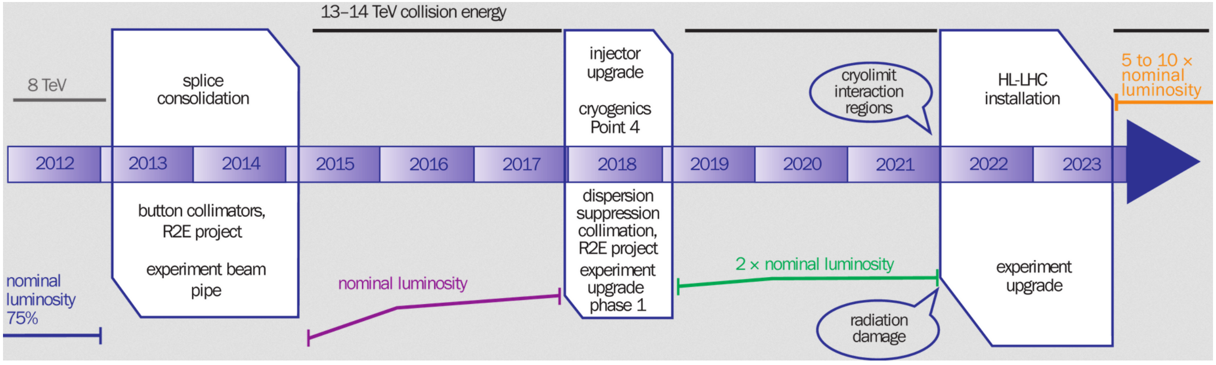


Figure 2.2: Schematic view of the HL-LHC long shutdown plan [95].

Every beam pair injected to the LHC, is called a *fill*. If beams are stable and the detector functional, data are taken. A period of data collection with similar detector conditions and without interruption is called *run*. The runs are divided in *lumi sections*, which typically last a few seconds.

2.1.1 Operational history

During the beam commissioning, on September 10th 2008, a defect in the superconducting connection between magnets led to a break of the liquid helium containment [90]. The incident caused a sudden expansion of a large volume of helium producing extensive damage to the machine. The repairing and prevention against further accidents took around a year to be completed, consequently delaying the physics program for several months.

The first pp collisions after the accident happened in November 2009 at a center–of–mass energy of $\sqrt{s} = 900$ GeV. Later in that year, the center–of–mass energy was raised to 2.36 TeV, breaking the world record on the center–of–mass energy for hadron colliders, previously held by the Tevatron.

In March 2010, the LHC set again a world record by reaching a center–of–mass energy of $\sqrt{s} = 7$ TeV. The design center–of–mass energy was not achieved because there was not enough confidence on the quenching protection system designed, which perhaps could not be able to prevent damage in the machine if the magnets would loose their superconducting properties. During that year the LHC delivered around 0.045 fb^{-1}, as presented in figure 2.3–left, of *pp* collisions at a bunch crossing rate of 50 ns. Also, it delivered a total integrated luminosity of about 9 μb^{-1} in *PbPb* at $\sqrt{s} = 2.76$ TeV per nucleon, as shown in figure 2.3–right.

The instantaneous luminosity was further increased in 2011, achieving another world record on April 21st 2011 with a value of 4.67 $\times 10^{32}$ cm^{-2} s^{-1}. The total luminosity delivered in that year was around 6 fb^{-1}, as presented in figure 2.4–left, in *pp* collisions and near to 160 μb^{-1} in *PbPb*, as shown in figure 2.4–right.

The center–of–mass energy was further increased in March 2012 to $\sqrt{s} = 8$ TeV. The risk for the machine was carefully evaluated before and weighted against the gain in luminosity and

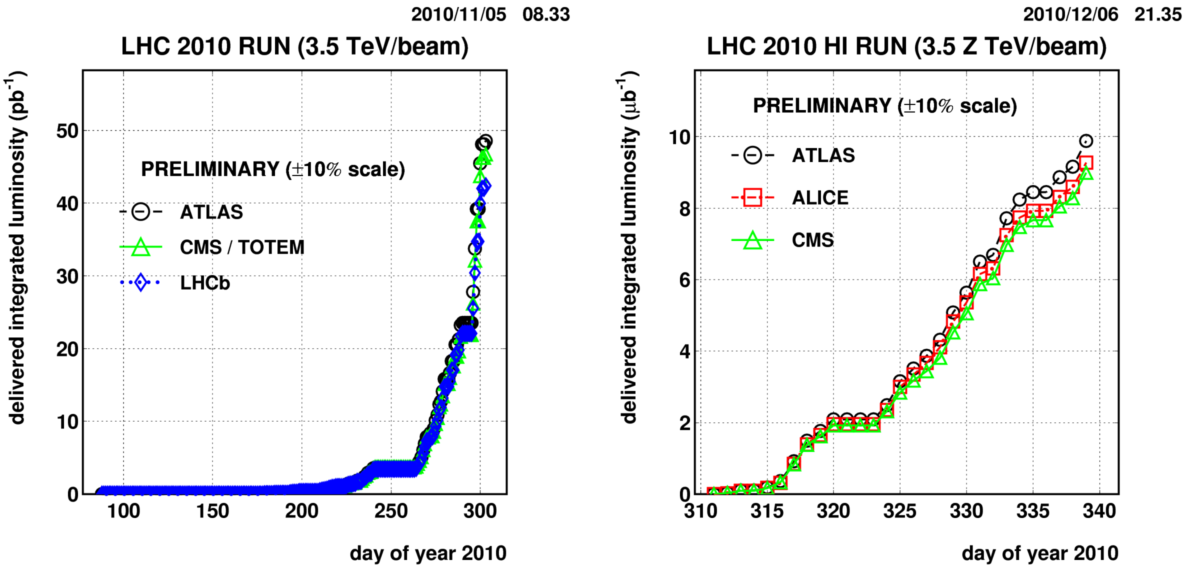


Figure 2.3: Luminosity delivered by LHC to its experiments in 2010. Delivered integrated luminosity as function of time for proton–proton colisions at $\sqrt{s} = 7$ TeV (left) [96] and Pb–Pb collisions at $\sqrt{s} = 2.76$ TeV (right) [97].

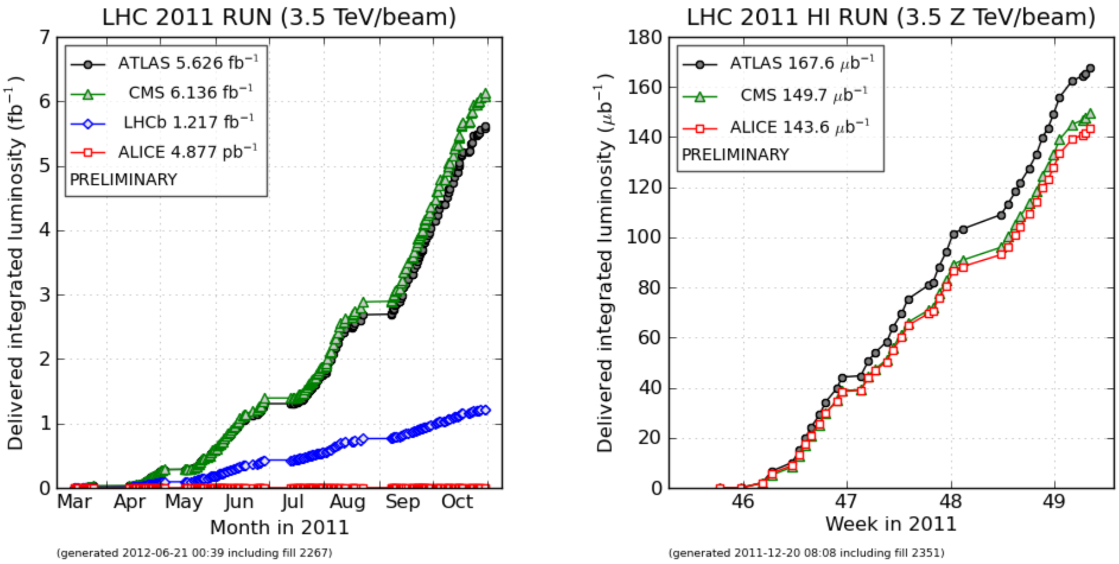


Figure 2.4: Luminosity delivered by LHC to its experiments in 2011. Delivered integrated luminosity as function of time for proton–proton colisions at $\sqrt{s} = 7$ TeV (left) [98] and Pb–Pb collisions at $\sqrt{s} = 2.76$ TeV (right) [99].

physics reach potential. The machine has operated at that energy since then, steadily increasing the delivered luminosity. In 2012, the LHC was able to provide around 23.3 fb^{-1}, as presented in figure 2.5–left, to both ATLAS and CMS (see section 2.2).

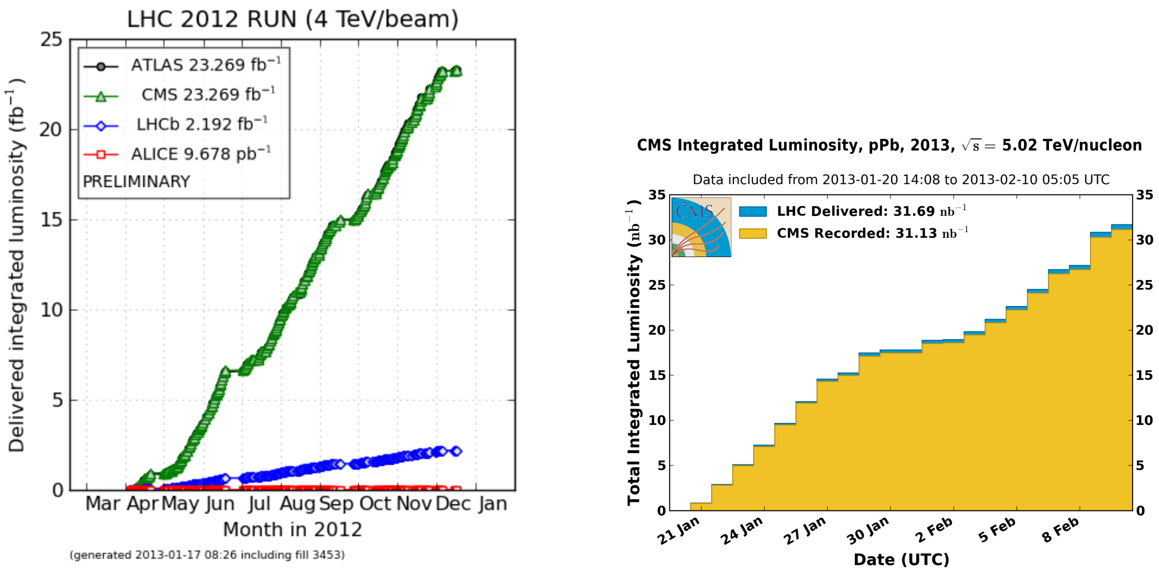


Figure 2.5: Luminosity delivered by LHC to its experiments in 2012 and 2013. Delivered integrated luminosity as function of time for proton–proton colisions at $\sqrt{s} = 8$ TeV in 2012 (left) [100] and p–Pb collisions at $\sqrt{s} = 5.02$ TeV per nucleon in 2013 (right) [101].

In January 2013, there was a pPb run at an energy of 5.02 TeV per nucleon. Around 31.7 nb^{-1} of data were delivered and CMS was able to collect 31.1 nb $^{-1}$, as shown in figure 2.5–right. There was also a *pp* run at $\sqrt{s} = 2.76$ TeV in which 5.4 pb^{-1} were collected.

After the end of Run 1, the data collection period from 2009 to 2013, the machine is going through a shut down period which started at March 2013 up to April of 2015. Among other improvements, the quench protection system will be upgraded, allowing to run at the design center–of–mass energy. The experiments will use this technical stop for maintenance and upgrade work.

2.2 LHC experiments

To investigate the physics at the TeV scale, several LHC experiments were designed to accomplish distinct tasks in different fields of particle physics. The main goal of the LHC experiments are the search for the Higgs boson and for new physics beyond the Standard Model, while improving the measurements of Standard Model processes.

As can be seen in figure 2.6, the LHC experiments are located in four of the eight straight sections of the tunnel, which have a length of aproximately 530m. The remaining sections are used by the LHC team for beam–maintenance related tasks such as beam acceleration, cleaning, injection and dumping [102].

For the two larger experiments, new caverns had to be excavated. The two larger experiments are 4π hermetic multi-purpose detectors: CMS, installed in Point 5, and Atlas (**A** **T**oroidal **L**HC **A**pparatu**S**) [103–105], located at Point 1. Both of them are based on the same general "onion–like" concept of consecutive layers of subdetector systems: the tracker, the calorimeters and the muon detectors. They are multi–purpose detectors, specialized for the search of the Higgs boson and any new physics, such as SUSY (**Su**per **Sy**mmetry), and other exotic processes.

The CMS detector will be discussed in detail in section 2.3.

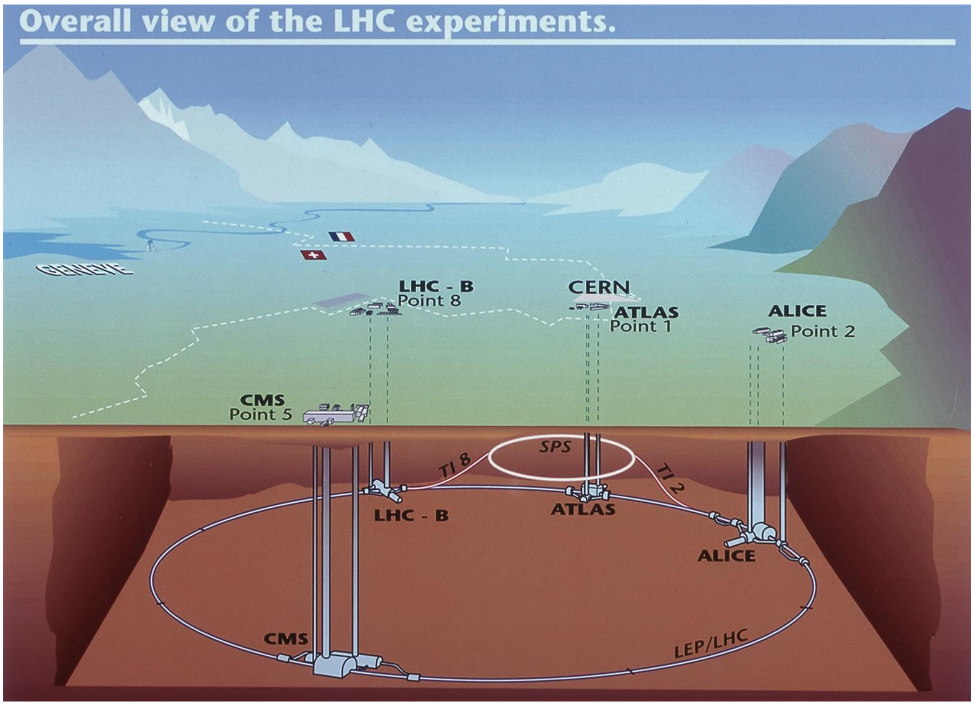


Figure 2.6: Schematic view of the LHC machine and its main experiments. ATLAS is located at Point 1, ALICE at Point 2, CMS at Point 5 and LHCb in Point 8 [102].

ATLAS is the largest multi–purpose particle detector of the world, having a length of 44 m and a diameter of 25 m. The magnetic field is created by toroid magnets (0.5 to 1 T) for the muon spectrometer and a small solenoid (2T) for the tracker. ATLAS has a high–granularity liquid Argon sampling electromagnetic calorimeter and a tile hadronic calorimeter. ATLAS has a better energy resolution in hadronic calorimenter and a worse momentum resulution of the charged tracks compared to CMS. However, the final performance for most of the physics analysis related objects such as muons, photons and jets is very similar, because the information of different subdetectors is combined in the reconstruction (see section 3.4).

ALICE (**A** **L**arge **I**on **C**ollider **E**xperiment) [106,107] is located at Point 2, where previously the LEP's L3 [108] experiment was installed. It is an asymmetric detector focussed on heavy–ion physics and it was built to investigate the quark gluon plasma, an exotic phase of matter with very high energy density. In addition, ALICE is looking for the origin of confinement and the mass–generation mechanism. To cope with the large particle multiplicities occurring in $PbPb$ collisions, ALICE has a silicon vertex detector, a TPC (**T**ime **P**rojection **C**hamber) and a transition radiation detector. However, this design made the detector response very slow and thus the TPC is not able to cope with the high event rates of the design luminosity for pp runs. Thus, LHC delivers reduced luminosity at Point 2.

LHC-b (**L**arge **H**adron **C**ollider **b**eauty) [109, 110] is an asymmetric detector assembled at Point 8, the place once occupied by DELPHI [111]. Its main physics goal is to investigate CP violation and rare decays of B and D mesons, and thus trying to explain the large matter-antimatter asymmetry in the universe. At the LHC center–of–mass energy, such mesons are predominatly produced in the forward and backward directions. Given that, and due to economical constraints, the spectrometer is a single arm, covering an angle between 10 and 250 mrad along the beam line. Since the cross-section for production of $b\bar{b}$ pairs is at the order of μb, the experiment can afford to run on lower luminosity to avoid pile-up, in a situation similar to ALICE.

The LHCf (**L**arge **H**adron **C**ollider **f**orward) [112,113] experiment consists of two very forward imaging calorimeters made of tungsten plates, plastic scintillators and position–sensitive sensors at 140 m from the interaction Point 1. The LHCf experiment aims to measure in the pseudorapidity range of $\eta > 8.4$. It is used to measure the luminosity provided by the LHC and to measure forward particles in order to simulate cosmic rays in laboratory conditions.

The TOTEM (**Tot**al **E**lastic and diffractive cross-section **M**easurement) detector [114, 115] consists of silicon–sensor several modules up to 220 m away from CMS. The pseudo-rapity range is $3.1 \leq \eta \leq 6.5$. TOTEM aims at measuring the total cross-section of elastic scattering and diffractive processes.

2.3 The CMS detector

The Compact Muon Solenoid is located about 100 m underground. Built in a cylindrical shape around the beampipe, which runs through the center of the detector, it is symmetric both radially around the beam pipe and in the z–direction along the beampipe. CMS is 21.6 m long and has a diameter of 14.6 m, weighting 12.5 ktons. CMS has been designed with special emphasis on an excellent muon reconstruction, which are a clean probe on new physics, as mentioned in the experiment's name. The different CMS components are arranged in a layer–like structure.

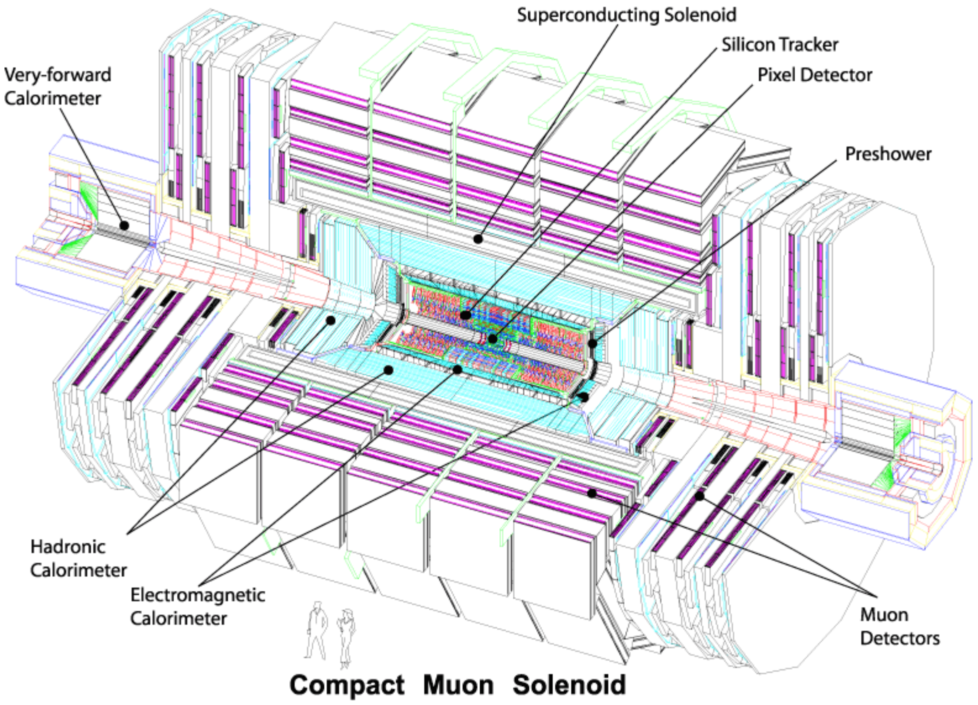


Figure 2.7: Schematic view of the CMS detector with two human figures as scale. All major components are labeled. The ZDC and CASTOR sub–detectors are not displayed in the figure [116].

A schematic view of the detector and its components is presented in figure 2.7. The Inner Tracking System (see subsection 2.3.3) surrounds the beampipe and includes the Vertex Detector and the Silicon Strip Detector. The next two layers are the Electromagnetic and the Hadronic Calorimeters (see subsection 2.3.4). All of these layers are installed within the superconducting Solenoid (see subsection 2.3.2). Hadronic Outer Calorimeter and the Muon Tracking System are located outside of the Solenoid. In addition, there are two hadronic endcaps calorimeters outside the iron structure. The CMS detector can be separated into five disk–shaped segments, allowing the inner parts of the detector to be accessed.

2.3.1 Coordinate system

In this section, the coordinate system of CMS which is used through out this thesis is described. The origin is at the interaction point at the center of the detector. The x–axis points horizontally inwards to the center of the LHC ring, while the y–axis points upwards. The z–axis points in counter–clockwise direction of the LHC. Cylindrical coordinates are commonly used. The coordinate ϕ is defined as the angle in the x–y azimuthal plane, starting from the x–axis. The coordinate θ is defined as the angle to the z–axis. Finally, r is the transverse distance from the z–axis.

In pp collisions, the initial momentum of the colliding partons is unknown but one can measure that the incoming protons have very small transverse momentum when compared to its total momentum. Therefore, variables which are Lorentz invariant under boosts along the beam axis are used. The rapidity y is defined as:

$$y = \frac{1}{2}\log\left(\frac{E+p_z}{E-p_z}\right), \tag{2.3}$$

where E is the particle's energy and p_z the momentum component in the z–direction. The pseudo–rapidity η is defined as

$$\eta = -\log\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{2.4}$$

In the relativistic limit, $m = 0$, one have $\eta = y$. The transverse projections are used to quantify the momentum and energy. The transverse momentum p_T is

$$p_T = p\sin(\theta), \tag{2.5}$$

where p is the momentum of the particle. The transverse energy E_T is

$$E_T = E\sin(\theta) = \frac{E}{|\cosh(\eta)|}. \tag{2.6}$$

2.3.2 Solenoid

A strong magnetic field, which is needed to enable high–precision measurements of the momenta of high–energy particles, was a driving force behind the detector design [117], its importance even contributing to the name of the experiment (Compact Muon **Solenoid**).

Most of the calorimenters and all tracking detectors are located inside the superconducting solenoid. The solenoid creates a homogeneous magnetic field of up to 3.8 T. The magnet is made of 4 layers of stabilized and re–enforced NbTi conductor, having a cold mass of 220 tons. It operates at 4.6 K and has a lenght of 12.5 m, a radius of 3 m and a thickness of 3.9 radiation lengths, which makes it the largest magnetic coil ever built. At the full current of 19.14 kA, it stores an energy of 2.6 GJ or 11.6 KJ/kg, being both the highest energy stored and the highest ratio of energy per cold mass when compared with other detector magnets. The field causes mechanical deformations in the detector of about 0.15%.

Outside of the solenoid, the muon chambers and outer hadronic calorimeter see a 2 T field of reversed polarity. This field is created by a 10 kt iron return yoke, which represents 80% of the CMS mass. It is composed by 5 wheels (6000 t) and 2 endcaps (2000 t each) with 3 disks each, having a diameter of 14 m and a length of 13 m.

2.3.3 Inner Tracking System

A schematic overview of the Inner Tracker is presented in figure 2.8. It has a length of 5.8 m and a diameter of 2.5 m, surrounding the interaction point. It is designed to provide a precise and efficient measurement of the trajectories of charged particles. It is composed of a pixel detector located between 4.4 cm and 10.2 cm in the radial direction from the beam axis and a silicon strip detector located from 20 cm to 1.16 m. Both barrel systems are complemented by endcaps, extending the acceptance up to $|\eta| < 2.5$.

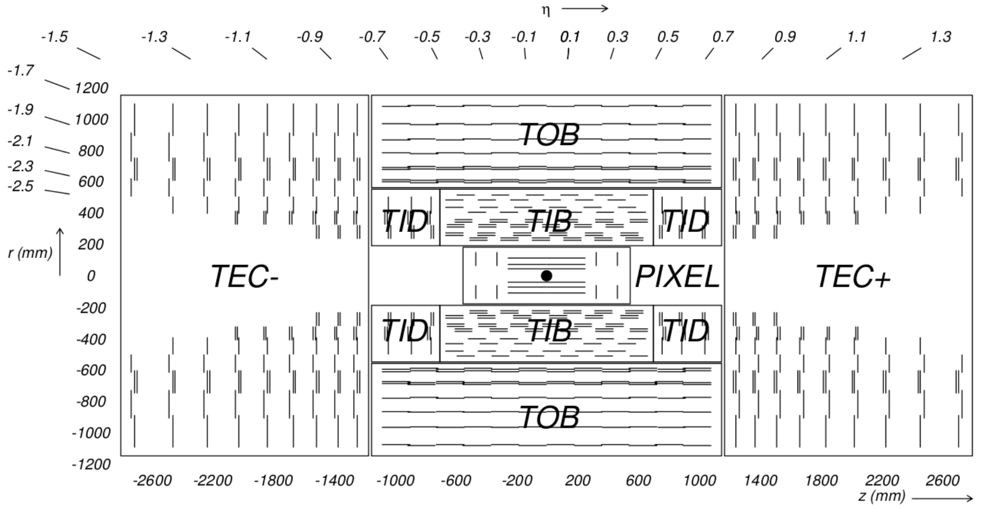


Figure 2.8: Schematic view of the Inner Tracking system of the CMS detector [87] with all of its subsystems labeled. Each line represents a silicon or pixel or strip detector module. Double lines indicate back-to-back modules which deliver stereo hits.

Charged–particle trajectories are bent due to the magnetic field. Their transverse momenta, p_{T}, in case they are travelling with the speed of light, c, are given by:

$$p_T = 0.3B\rho, \tag{2.7}$$

where p_{T} is the transverse momentum in GeV, B is the magnetic field in Tesla and ρ is the radius curvature in meters. The Inner Tracking System can measure the momentum of particles with p_{T} = 100 MeV [119] and achieves a relative p_{T} resolution of up to 1.5% for tracks with momenta of 100 GeV in the barrel area, decreasing with increasing η, towards the forward region. This subdetector also provides input to the high–level trigger algorithms.

Pixel Detector

The main goal of the Pixel Detector is to provide exact measurements of the starting point of any charged track, which is required for an exact reconstruction of vertices and tracks. It separates primary and secondary vertices, coming from short–lived particles, and due to this function, it is sometimes called Vertex Detector. It can identify heavy–quark decays and is useful for the suppression of pile–up events.

The Pixel Detector consists of the BPIX (**B**arrel **PIX**el) and FPIX (**F**orward **PIX**el) components. The barrel has three layers of modules, called ladders, and the endcaps have two layers, which are located at $z \simeq 35.5$ cm and $z \simeq 46.5$ cm form the interaction point. The detector blades are rotated by $20^{\underline{\mathrm{o}}}$ to compensate for particle deflection. The blades cover the pseudo–rapidity range up to $|\eta| \leq 2.4$. Some tracks can not be measured for $2.4 \leq |\eta| \leq 2.5$, as they do not hit all the pixel layers.

The silicon sensors are n–doped wafers on which p–doped pixels are made with the size of 100 μm × 150 μm and a thickness of 150 μm. The hit rate density is 1 MHz/mm^2 at 4 cm from the beam, requiring 66 million pixels to achieve a reasonable occupancy. The Pixel Detector is divided into 1440 modules and the total area corresponds to 1 m^2. The hit–position precision is 10 μm in $r\phi$ and 15 μm in z for trajectories with at least 2 hits in different layers [118].

Silicon Strip Tracker

With 15148 silicon micro–chip modules, comprising a total of 9.3 million strips, the Silicon Strip Detector aims to provide an additional momentum measurement of the produced particles detected in the Pixel Detector. In contrast to pixels, which are point–like detectors, the strips are line–like. The tracker endcaps extend its coverage up to $|\eta| \leq 2.4$ with a design overall occupancy between 1 and 3%.

The tracker in the barrel area is composed of the TIB (**T**racker **I**nner **B**arrel) and the TOB (**T**racker **O**ther **B**arrel). The TIB consists of four layers of strip detectors, each strip with the size of 10 cm × 80 μm and a thickness of 320 μm. The TOB consists of 6 layers, where the strips have a size of 25 cm × 180 μm and thickness of 500 μm. The TIB has a hit–position resolution of 230 μm, while TOB has a resolution of 530 μm. The first 2 layers in both tracker barrels are mounted as stereo, meaning they are double–sided with a relative angle of 100 mrad (5.3 degrees), allowing to measure also the z coordinate.

The tracker endcap is composed of the TID (**T**racker **I**nner **D**isk) and the TEC (**T**racker **E**nd**C**ap). The TID has 3 layers on each side of the interaction point and a thickness of 320 μm, while the TEC has 9 discs on each side of the interaction point, with a distance between 124 and 282 cm from the interaction point and a thickness of 500 μm. The hit–position resolution depends on $|\eta|$, being in the same order of magnitude as for the pixel detector for $|\eta| \leq 1.6$ and degrading as $|\eta|$ increases.

2.3.4 Calorimeters

The purpose of a calorimeter is to measure particle energies by full absorption. In order to perform a precision measurement, it is vital to absorb neutral particles, which could not be detected in the tracker. In addition, the calorimeters allow carrying out a second and independent measurement of charged particle energy.

When discussing calorimeters, one must first define the radiation length X_0, which is the average distance an electron needs, in order to have radiated enough by bremsstrahlung, to reduce its energy by a factor of $1/e$. Also one needs to define the interaction length, λ_1, in which the energy of a relativistic particle, due to nuclear interactions, is reduced by a factor of $1/e$.

The CMS calorimetry main components are the ECAL (**E**lectromagnetic **CAL**orimeter) and the HCAL (**H**adronic **CAL**orimeter). In the forward region CASTOR and ZDC (**Z**ero **D**egree **C**alorimeter) are located. The CMS calorimeters have a granular design to allow a position measurement and a matching with the particles detected in the Tracking System.

Electromagnetic Calorimeter

A schematic overview of the Electromagnetic Calorimenter [120] is shown in figure 2.9. It is composed of EB (**E**CAL **B**arrel) and EE (**E**CAL **E**ndcap). The Electromagnetic Calorimeter was built for precision energy measurements of photons and electrons, which create electromagnetic showers inside the high–density material via emissions of Bremsstrahlung–photons and conversion into e^+e^-–pairs. It uses lead–tungstate crystals ($PbWO_4$), summing up to 61200 units in EB and 7324 in EE. The $PbWO_4$ high–density and radiation–resistent crystal is also a fast calorimeter, emitting 80% of its light in the range 420–430 nm within 25 ns.

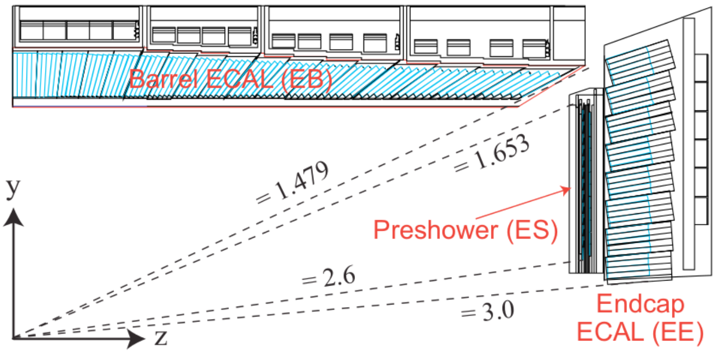


Figure 2.9: Longitudinal view of the positive x–y quarter of the CMS Electromagnetic Calorimeter [85].

The EB crystals have a radiation depth of 25.8 X_0, corresponding to a length of 23 cm. The crystals are 2.2 cm × 2.2 cm, which translates into $\Delta\phi \times \Delta\eta = 0.0175 \times 0.0175$. They are read out by photomultipliers and cover a pseudo–rapidity up to $|\eta| \leq 1.479$.

The crystals in the EE have a length of 22 cm, which corresponds to 24.7 X_0. Each crystal has a size of 2.86 cm × 2.86 cm, corresponding to $\Delta\phi \times \Delta\eta = 0.05 \times 0.05$ and is read out by phototriodes. EE is located at $z = \pm 3.154$ m and covers the absolute pseudo–rapidity between 1.453 and 3.0.

The Preshower Detector is located between the TEC and the EE, covering a pseudo–rapidity range $1.653 \leq |\eta| \leq 2.6$, with a total radiation depth of about 5 X_0. It consists of silicon strip detectors behind disks of lead. It helps distinguish photons and electrons from minimal ionizing particles, which have similar signatures, by starting and detecting the showering process of the former.

The energy resolution of the ECAL can be parametrized as

$$\frac{\sigma(E)}{E} = \frac{(2.8 \pm 0.3)}{\sqrt{E}} \oplus \frac{0.124}{E} \oplus 0.26 \tag{2.8}$$

where the values are derived from test beam results and simulation [86]. The first term represents the stochastic term, the second describes the electronic noise, digitalization and pile–up, while the third refers to the non–uniformity, energy leakage and errors in intercalibration. Hence, ECAL measures unconverted photons with $E_T = 100$ GeV with a relative energy resolution of $\simeq$0.5%.

Hadronic Calorimeter

The HCAL [121] aims to identify and measure the energy of hadronic particles and missing transverse energy up to $|\eta| = 5.2$. A schematic view is presented in figure 2.10. Is is composed of the HB (**H**CAL Barrel) and the HE (**H**CAL Endcap) inside the limited space between the solenoid, and the HO (**H**CAL **O**ther) and the HF (**H**CAL Forward) outside of the magnet.

Both HB and HE are composed of alternating layers of brass, an absorber material which is non–magnetic and has a short hadronic radiation length, and plastic scintillators. Adding the 1.1 λ_1 radiation lengths of the ECAL, the HB has a depth between 5.82 and 10 λ_1, while the HE has 10 λ_1. Each scintillator plate has a thickness of 3.7 mm, except for the first layer where the thickness is 9 mm, and is connected by wavelength–shifting fibres to multi–channel hybrid photodiodes.

The HB covers $|\eta| < 1.4$ and each crystal tower has a segmentation of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$, adding up to a total of 2304 towers. It has 14 brass layers between 50.4 and 56.6 mm thickness. Due to structural reasons, there are additional innermost and outermost layers made

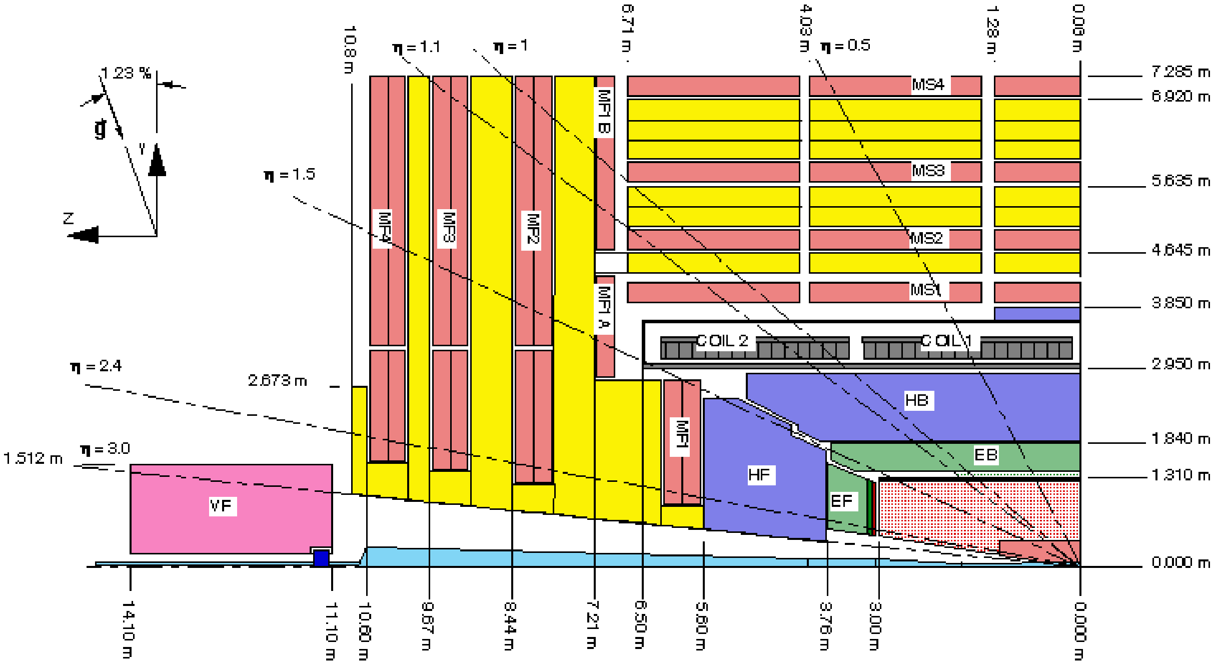


Figure 2.10: A longitudinal view of a quadrant of CMS. The HCAL is represented in purple and pink and labeled as HB, HF and VF. HB is the Hadronic Barrel and HF represents the Hadronic Endcap, while VF represents the Hadronic Forward Calorimeter. The purple section on top of coil 1 is Hadronic Outer Calorimeter.

of stainless steel with thicknesses of 4.0 and 7.5 cm. The HE covers $1.3 < |\eta| < 3.2$, each tower covers $\Delta\phi \times \Delta\eta = 0.17 \times 0.17$. There are 18 layers, made from 7.9 cm of brass, followed by 0.9 cm thick scintillators. They add up to the same total of 2304 towers.

The HO is located between the solenoid and the muon system. It compensates for the lack of radiation lengths in the central region, providing further containment of hadronic showers. Along the z–direction, the HO is divided into eight central rings with two layers, which increases the depth of hadronic section to 10 λ_1, adding to a total of 11.8 λ_1. The other rings, twelve on each side, have just one layer.

The HF covers the pseudo–rapidity range $2.9 < |\eta| < 5.2$ and it is located at $z = 11.2$ m, with r between 12.5 and 130 cm. It is designed to withstand the large particle flux from the forward direction, and therefore the design focussed on radiation hardness. The absorbing material is steel, while the fibers are made from of quartz. The Cherenkov light emitted is detected by photomultipliers. Half of the fibers go through the whole 165 cm depth of the detector, while the other half starts only 22 cm away from the surface. This difference can be used to distinguish between electromagnetic and hadronic showers, because the former would have been mostly absorbed in the first part of the detector, while the latter are more likely to develop into the second part.

The energy resolution for jets reconstructed with combined ECAL and HCAL information in the central region of the detector can be parametrized as [86]:

$$\frac{\sigma(E)}{E} = \frac{100}{\sqrt{E}} \oplus 4.5. \tag{2.9}$$

Thus, the energy resolution for jets with $E_T = 50$ GeV is less than 3% for $|\eta| < 3.0$. It increases to around 5% for $3.0 < |\eta| < 5.2$ [122].

CASTOR

CASTOR (**C**entauro **A**nd **S**trange **O**bject **R**esearch) is a quartz–tungsten sampling calorimeter. As can be seen from figure 2.11, it is located in the very forward region, installed at 14.38 m from the interaction point, covering the pseudo–rapidity range -5.2 > η > -6.6.

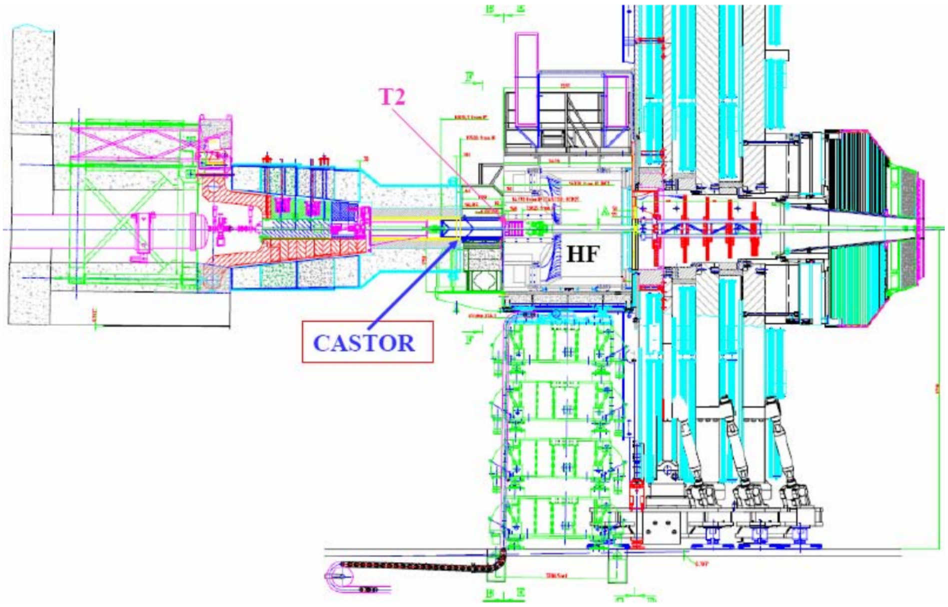


Figure 2.11: CASTOR position relative to CMS [87].

CASTOR is divided into two halves, surrounding the beam pipe when closed. The detection method and radiation–hardness design is very similar to the HF. However, in this case, tungsten is the absorbing material. The electromagnetic (EM) section has two modules while the hadronic (HAD) section has twelve modules, as shown in figure 2.12. Each module has five quartz plates and five tungsten plates, and in the EM section the tungsten plates have a thicknesses of 5.0 mm and the quartz plates of 2.0 mm. In the HAD section the thickness are 10.0 mm and 4.0 mm, respectively. To maximize the Cherenkov light output, all plates are inclined by 45 degrees with respect to the beam axis. The EM section is 20.1 X_0 deep, while the HAD section has a depth of 9.24 λ_1, adding up to a total of 10 λ_1. The detector has 16 segmentations in ϕ, making a total of 224 channels. The light produced in the quartz fibers is collected in light–guides and measured by photomultipliers.

The parametrization CASTOR energy resolution can be obtained by replacing the constant term by 1% and the stochastic term by $\simeq$50% in equation 2.9. The resolution of both the EM and HAD sections were estimated using electrons, pions and muons in the GeV energy range. A technical part of the thesis was carried out with CASTOR, details are given in appendix A.

2.3.5 Other detector components

The Muon System [123] is embedded into the return yoke of the solenoid and aims to deliver a high–precision momentum measurement of minimal ionizing particles, such as muons. It is composed of three different sub-detectors: DT (**D**rift **T**ube Chambers), CSC (**C**athode **S**trip **C**hambers) and RPC (**R**esistive **P**late **C**hambers). The DT covers $|\eta| < 1.2$ and is filled with a mixture of 85% CO_2 and 15% Argon. The CSC is made of multiwire proportional chambers covering a range of $0.9 < |\eta| < 2.4$. The RPCs are made of parallel–plate detectors filled with a gas mixture of 96% $C_2H_2F_4$, 3.5% C_4H_{10} and 0.3% SF_6. The RPC is able to detect up to $|\eta| = 1.6$.

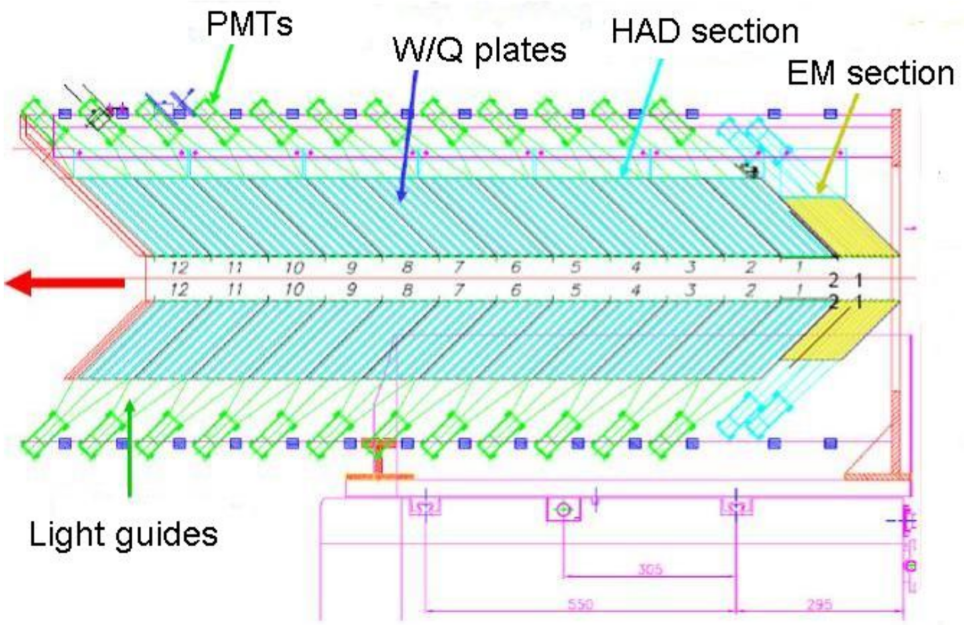


Figure 2.12: A longitudinal view of CASTOR with its elements labeled [87].

There are two ZDC, one on each side of CMS, located at 140 m from the interaction point. They cover the pseudo–rapidity $|\eta| \geq 8.3$ for neutral particles. They have two independent sections: an electromagnetic and a hadronic. The design and materials are very similar to the ones used by HF and CASTOR, with special emphasis on radiation hardness.

2.3.6 Luminosity measurement

For a precision measurement of the cross–section, the luminosity collected by CMS needs to be very well known, as pointed out in equation 2.2. At CMS, the well–known rate of inelastic proton–proton scattering is used to determine the luminosity [124], which is monitored using three different methods: online, offline and Van der Meer scans.

The online method uses the HF in two modes: by estimating the rate from the average of empty towers or by counting the average deposited transverse energy per tower, which has a linear relation with the luminosity.

The offline method uses both HF and Tracking. The HF method counts events with energy depositions which sum up to $\sum E_T \geq 1$ GeV coinciding within 8 ns in both backward and forward directions. The tracking method estimates the luminosity by counting the number of events with at least 2 tracks coming from same vertex and close to the interaction point. Other methods include TOTEM or the production rate of W and Z bosons.

Van der Meer [125] scans are performed to measure the beam profile, and thus, to estimate the luminosity using equation 2.1.

2.3.7 Trigger

The event rate at the LHC can be up to 40 MHz and there is a large number of read–out channels which can yield a total event size of about 1 MegaByte, making it impossible to store all events. Therefore, only the most interesting events can be recorded. With a crossing rate of about 50 ns, the trigger has to decide quickly whether an event is worth to be saved. At CMS the trigger is part of the DAQ (**D**ata **AQ**uisition) system [127]. With the trigger the event rate can be reduced by six orders of magnitude. The DAQ consists of the detector electronics and readout network.

The trigger is divided in two steps: L1 (**L**evel–**1** Trigger) and HLT (**H**igh **L**evel **T**rigger).

L1 Trigger

The L1 trigger [126] architecture is presented in figure 2.13. The L1 trigger is hardware–based, being made up of FPGAs (**F**ield **P**rogrammable **G**ate **A**rray), ASICs (**A**pplication **S**pecific **I**ntegrated **C**ircuit) and programable lookup tables. The main feature is speed: the L1 trigger is located near the detector and is able to make a decision in up to 3.2 μs, out of which 1 μs is due to computation and 2.2 μs are due to cable latency.

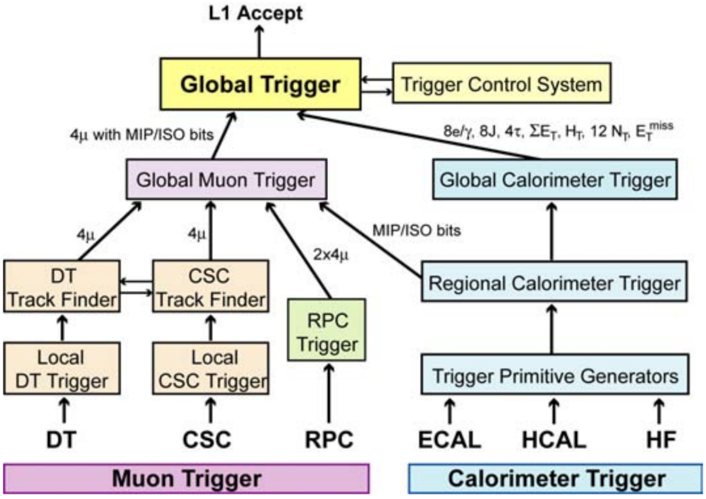


Figure 2.13: Architecture of Level–1 Trigger [87].

The granular readout information of the calorimeters and muon chambers are used to reduce the event rate to 100 kHz, which corresponds to 100 GB/s. Using calorimeter signals, a fast algorithm is used to reconstruct electrons, photons, jets and missing E_T, called trigger primitives. Information from the Muon System and minimum ionizing signatures in the calorimeters is used for muon reconstruction. The outputs from the global calorimeter trigger and the global muon trigger are collected in the the global trigger, which makes the final decision. While the decision is made, the information is pipelined, ready to be read and forwarded to HLT when the event is accepted.

HLT

The HLT [127, 128] is software–based and has the task to reduce the event rate from 10 kHz to 100 Hz. It reads the whole detector information from front–end or interfacing electronics to a computer farm, using zero supression, the skipping of null signals, and data compression techniques to reduce the event size. The readout network collects informations of the subdetectors in the Event Manager and the Event Builder merges them. The DAQ architecture is illustrated in figure 2.14.

The Event Builder runs the HLT algorithms in the computer farm, comprising about 2500 CPUs, reconstructing and filtering the events using a specialized software. Each CPU in the farm is an independent node and tries to reject the events as soon as possible, only including additional information of the tracker and the full detector information when required. As the HLT is purely software–implemented, it allows for easy changes and develops with the growing knowledge and understanding of the detector. Each physics group has its own set of trigger paths, targeting the specific objects and thresholds needed for their analyses. All accepted events, and a small fraction of rejected ones, are forwarded to data quality monitoring and storage. The information

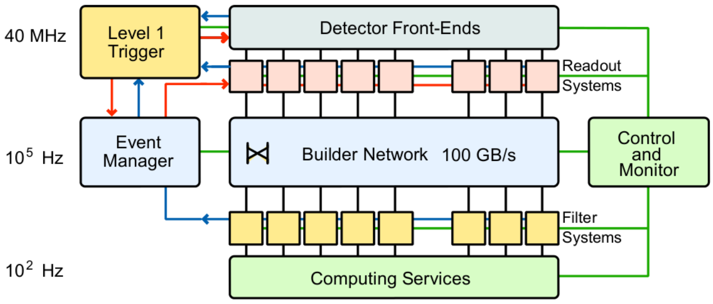


Figure 2.14: Architecture of CMS DAQ [87].

on the trigger decision is also stored and it is used to separate the events by categories based on the physics goals.

Trigger Prescales

Because of the limitation on the rate at which events can be recorded, not all triggered events can be saved. Thus, a prescaling factor has been introduced for those trigger paths with high rates. For example, with a prescale value of 5, only every 5th triggered event is recorded. Some trigger paths require more than one object, being able to sustain a lower prescale than from any of the individual objects without consuming much of the event rate bugdet. These prescales usually depend on the transverse momentum of the object and instantaneous luminosity, growing as the momentum threshold decreases or luminosity increases.

Taking prescales into account, the measured cross–section, σ_{meas}, can be writen as

$$\sigma_{meas} = \frac{N_{rec} \cdot d}{\mathcal{L}_{rec}} \tag{2.10}$$

where d is the prescale value, $\mathcal{L}_{rec}$ is the recorded luminosity and N_{rec} the number of recorded events with that trigger.

Computing System

Thousands of scientists arround the world expect to analyze about 15 Petabyte of data from the LHC per year. The processing power needed for storing and analyzing this data cannot be provided by a centralized computing system [129]. CMS divides their tasks through a global grid called WLCG (**W**orldwide **LHC** **C**omputing **G**rid) [130,131], which was developed together with the other LHC experiments and consists of 140 sites in 35 countries.

A tiered struture benefits from existing resources and expertise, providing redundancy and robustness. The tasks are distributed between the three different layers, which are Tier 0, Tier 1 and Tier 2. The dataflow in CMS is represented in figure 2.15.

Tier 0 is located at CERN and collects the raw data directly from the DAQ. It is connected with the CAF (**C**ERN **A**nalysis **F**acility), providing offline DQM (**D**ata **Q**uality **M**onitoring) as well as alignment and calibration constants. Tier 0 is also responsible for prompt data reconstruction.

The raw and reconstructed data are copied to at least one of the seven Tier 1 centers, which are large off–site computer farms connected with fast links to Tier 0. They are also used to re-reconstruct the data and to create AOD datasets (**A**nalysis **O**bject **D**ata), which feature only physics objects, thus having a reduced size.

About 50 Tier 2 centers are responsible for the user–related tasks such as: Monte Carlo production, offline calibration and data analysis. They also store data samples for direct access.

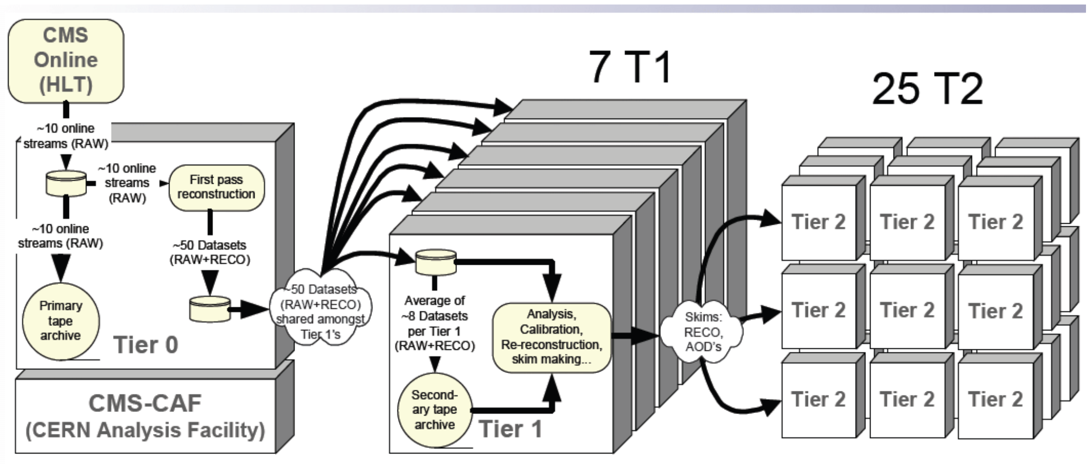


Figure 2.15: Schematic view of the dataflow in CMS [132].

A small computer farm or laptop is sometimes referred to as Tier 3. Although they can access the data, they have no responsibilities concerning processing or storing.

2.3.8 Data Quality Monitoring

Before the data can be used in physics analyses, the DQM [133] is applied to check that all detector components were working properly and the reconstruction software has no obvious problems. The DQM workflow is organized in two steps: online and offline.

In the online step, a few key histograms are filled from the data taken directly from the HLT filter at a rate of 100 MHz. The rest of the histograms are filled from a dedicated DQM stream which works with a 10–15 Hz rate. Different monitoring algorithms are then run and the results are used to determine the quality of the data acquired in real time. For example, DQM helps to detect noisy and bad channels during data taking.

The offline DQM monitors the fully reconstructed data, and sometimes the re-reconstructed data, already with high–level physics objects such as jets or electrons. Usually, this step is done some days after the data taking as the prompt reconstruction takes up to two days and re-reconstruction can take some weeks. The output of the offline DQM defines which runs and lumi–sections can be used for physics analysis.

The histograms can be visualized in a dedicated web service [134] accessible through a regular web browser. This service is able to handle up to 50000 histograms per run in a reliable way. All online shifts are done at CERN while the offline shifts are distributed between FNAL, DESY and CERN.

Chapter 3

Event Generation and Reconstruction

In this chapter, the Monte Carlo techniques (see section 3.1), the structure of a typical high energy physics event generator (see section 3.3), an overview of the event generators used in this analysis (see section 3.3) and a brief reference on simulation and reconstruction (see section 3.4) are presented.

3.1 Monte Carlo Techniques

In many cases in particle physics, like the evaluation of integrals or sampling of random variables governed by complicated probability density functions can only be done using Monte Carlo techniques.

A random number generator is necessary for the Monte Carlo techniques. Such an algorithm is expected to generate statistically independent uniformly distributed values in the interval [0,1[[135]. These random number generator implementations are available in several software libraries such as CERNLIB [136] or ROOT [137]. There are several algorithm available: RANLUX based on Lüscher's method [138], the maximally equidistributed combined Tausworthe [139] and the Mersenne twister algorithm of Matsumoto and Nishimura [140] and many others.

All of these methods and implementations are not true random number generators, as true random number can only be achieved through a physical process such as radioactive decay or atmospheric noise. However, the performance of these numerical implementations of pseudo–random number generators can be estimated using several statistical tests.

3.1.1 Variable Sampling and Integral Computation

There are many methods to sample variables according to a given distribution. The simplest way is the Von Neumann acceptance and rejection method. It consists in defining an area which encloses or encapsulates completely the function to be integrated. This area A can be defined as a rectangle or another geometric shape, a combination of those forms, one for each interval, or even a function whose integral is known, $A = \int g(x)dx$. An example of the later approach is sketched in figure 3.1.

The defined area is filled with points whose coordinates are defined by the pseudo–random number generator. If a given point is under the curve, it is accepted. The integral is then:

$$\int f(x)dx = \epsilon A, \tag{3.1}$$

where ϵ is the fraction of accepted points and $f(x)$ the function to be integrated.

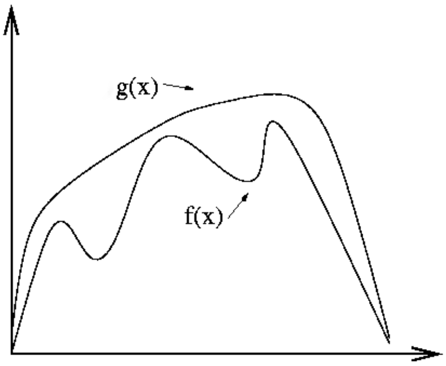


Figure 3.1: Example of the application of the Von Neumann acceptance and rejection method, $g(x)$ is the encapsulating function and $f(x)$ is the function to be integrated [141].

3.1.2 Markov Chains

In some cases, for example the generation of random numbers following a multivariate distribution with a high number of dimensions, the acceptance–rejection method may yield a too low efficiency to be practical. If independent random values are not required, MCMC (**M**arkov **C**hain **M**onte **C**arlo) methods can be used [142].

The Markov Chain method generates random points around the equilibrium distribution of the integrand function using the previous point as reference. In this way and given a sufficiently high number of random walk steps, the whole phase space can be explored. The only limitation of this method is that, for some unfortunate constructions of the chain, the number of steps can be impraticable. A simple and widely used implementation of MCMC is the Metropolis–Hastings algorithm [143].

The Markov Chains are specially useful for the parton shower, as the evolution of the BFKL parton ladder can be simulated using a random walk process.

3.2 Event Simulation

The event simulation in high energy physics using MC methods is factorized into several steps. The cross–section of the hard scattering process is calculated on a matrix–element level. High–order QCD effects coming from initial– and final–state radiation are taken into account by the parton showers. After this step, one has a parton–level prediction. The hadronization merges the partons into colourless hadrons, creating the so called hadron level. An additional step to add the underlying event may be required.

The generated events may be fed through a detector simulation. In this step, a full simulation of the event as seen by the detector is done using GEANT4 [144], which simulates the passage of particles through matter. After this step, which can include the trigger simulation as well, the prediction is called *digitalization level*. The simulated analog signals are collected and converted to digital signals. The data are then ready to be reconstructed, after this step it is called *detector level*.

3.3 Event Generators

The GPMC (**G**eneral–**P**urpose **M**onte **C**arlo) generators provide a fully exclusive modelling of high–energy collisions. Some of the most widely known and used GPMCs are HERWIG 6.5 [145], HERWIG++ [146], PYTHIA 6 [147] and PYTHIA 8 [148, 149]. These GPMC are used together with detector simulations to provide a realistic estimation of the detector response to high–energy processes. They are essential to go beyond fixed–order perturbative QCD, providing estimates for signals and backgrounds of the process under study.

BFKL– (HEJ [150–152]), CCFM–based (CASCADE [153]) and dijet NLO (POWHEG [154]) MC predictions, were not ready to be included in this thesis.

In this section, a brief summary of these differerent MC generators, their models, assumptions and applications is given.

3.3.1 HERWIG 6.5

HERWIG 6.5 is specialized for a detailed simulation of QCD parton showers. HERWIG is written in Fortran and uses DGLAP dynamics and is built with angular ordering. HERWIG takes into account the initial– and final–state QCD parton evolution with gluon interference. Colour coherence of the initial and final partons is implemented in all hard processes. Azimuthal correlations within and between jets are taken into account with gluon interference and polarization. HERWIG includes a cluster model for jet hadronization based on non-pertubative gluon splitting [44, 45]. An optional colour rearrangement model based on the space–time structure is provided [145]. HERWIG has no MPI model by its own, but uses the package JIMMY [155] to simulate the MPI. In this analysis, HERWIG 6.5 is run together with the CTEQ6L PDF [156].

3.3.2 HERWIG++

HERWIG++ 2.5 is built on the platform of ThePEG [157], using the C++ language, and has special emphasis on the accurate simulation of QCD radiation with DGLAP dynamics. The initial– and final–state QCD jet evolution is taken into account via angular ordering and soft gluon interference. HERWIG++ provides a detailed treatment of the suppression of QCD radiation from massive particles. HERWIG++ includes an eikonal model [146] for multiple partonic scattering to describe the underlying event. HERWIG++ features a cluster model of the hadronization of jets based on non–perturbative gluon splitting. A sophisticated model of hadron and tau decays using matrix elements to compute the momenta of the decay products for many modes is included, which deals with off–shell effects and spin correlations. In this work, HERWIG++ is run together with MRST2001 PDF [158] and the EE-3 tune [159] is used.

3.3.3 PYTHIA 6

PYTHIA 6 is based on the DGLAP evolution equation for initial– and final–state parton–showers, which are transverse–momentum–ordered. Written in Fortran, PYTHIA 6 uses the Lund string model [43] to describe the hadronization process. Final–state parton showers are matched to a number of different first–order matrix elements for the gluon emission. Colour reconnection [147] is included.

The UE has been extensively tuned and several tunes exist. These tunes differ in the choice of flavour, fragmentation and UE parameters. The latter set of parameters includes parameters for the parton showers, cut-off parameters for MPI, parameters determining the matter overlap between the incoming protons, and probabilities for colour reconnection. Some of the tunes have common flavour and fragmentation parameters, which have been determined using data from LEP. This analysis uses the following tunes:

- The Z2* tune [160] is a retune of the minimum transverse momenta of the MPI in the original Z2 [161] tune, using an automated tuning approach, CTEQ6L PDF and underlying event data at different center–of–mass energies. There are two predictions for this tune, one with MPI with fully detector simulation and reconstruction, which was used to estimate corrections and systematic uncertainties, and one without MPI on the hadron level.

- The P11 tune [162] is run with the CTEQ6L PDF.
- The AMBT1 tune [163] is run together with the MRST2007lomod PDF [164]. The AMBT1 tune has already been used by ATLAS in dijet correlations measurements [67].

3.3.4 PYTHIA 8

PYTHIA 8 is also based on DGLAP and uses the Lund string model for hadronization. It uses an improved MPI model which is interleaved with parton showering. In this analysis, PYTHIA 8 is run with the CTEQ6L PDF and tune 4C [165], which is the CMS recommended tune for the LHC. There was an emphasis on providing a good description of initial– and final–state parton showers and MPI. The parton showers are p_T–ordered and colour–reconnection is present. PYTHIA 8 has also been used to estimate corrections and systematic uncertainties.

3.4 Event Reconstruction

In this section, the event reconstruction is described, in which special emphasis is given to jet reconstruction and correction.

Events coming from trigger, described in subsection 2.3.7, need to be reconstructed. The reconstruction is performed by the application of a choir of algorithms: tracking algorithms turn the different hits recorded in the tracker system into trajectories of particle candidates, the calorimenter information of the single cells is combined into energy deposits to discriminate between different particles using shower shapes, and the muon system provides the muon candidates.

The PF (**P**article **F**low) idea was introduced at ALEPH [166] and is used as the standard reconstruction method in CMS [167]. Although the concept is simple and it was used in *ep* and e^+e^- colliders, its application is not trivial at hadron collider experiments because of the large particle multiplicity, preventing a generalized use. The information from all of the sub–detectors is combined to create coherent PF–particle candidates. The tracker resolution scales with p_T while the calorimeters scales with $1/\sqrt{E}$. This combination joins the best of both and helps to reduce the uncertainties because a given quantity is measured at least twice with different methods, tracker and calorimetry, and the final uncertainty is smaller than any individual uncertainty.

The same procedures are applied on both raw data and simulation–level MC events.

3.4.1 Jets Reconstruction and Correction

In this analysis, the jets are reconstructed using the anti-k_T jet clustering algorithm, with a distance parameter R of 0.5. Combined information from the calorimeters and trackers are used as input for the CMS particle–flow jet reconstruction algorithm. The reconstructed jets required only small additional energy corrections derived from the PYTHIA 6 MC generator and in-situ measurements with photon+jet and dijet events [168]. The corrections applied both on data and MC account for pile–up and jet response as function of η and p_T. Data required a further residual correction for η and p_T response. The contribution of miss-reconstructed jets is kept at a negligible level by means of the procedures discussed in [169, 170]. Additionally, the standard tight JetID criteria are required:

- neutral hadron fraction < 0.90;
- neutral EM fraction < 0.90;
- number of constituents > 1.

For $|\eta| < 2.4$, the additional criteria for JetID were:

- charged hadron fraction > 0;
- charged multiplicity > 0;
- charged EM fraction < 0.99.

These requirements ensure a low rate for other physics objects, for example photons, to be selected as a jet.

Chapter 4

Event Selection and Control Distributions

In this chapter the topologies under study (see section 4.1), the data samples used in the analysis (see section 4.2), the pile–up reweighting procedure (section 4.3), the event selection criteria (section 4.4), and some relevant control distributions (section 4.5) are presented.

4.1 Topologies Under Study

In this analysis, events with at least one forward and one central jet with $p_T > 35$ GeV are selected. This selection, defined as *inclusive selection*, is illustrated by the Feynman diagram in figure 4.1–left. The pseudo–rapidities of the central and the forward jet axes are required to be $|\eta| < 2.8$ and $3.2 < |\eta| < 4.7$, respectively.

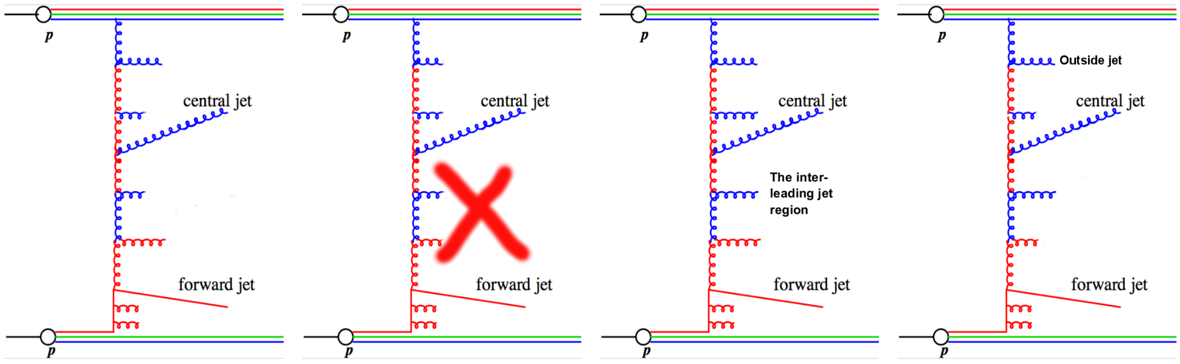


Figure 4.1: Feynman diagrams illustrating the different topologies studied. From left to right: inclusive scenario, inside–jet veto scenario, inside–jet tag scenario and outside–jet tag scenario.

The *inside region*, labeled in the Feynman diagrams of figure 4.1 as *inter–leading jet region* is defined as the pseudo–rapidity region between the leading forward and the leading central jets. The *outside region* is defined as the pseudo–rapidity region outside the dijet system. The $\Delta\phi$ measurement is repeated for two subsamples: events with at least one additional jet with $p_T > 20$ GeV in the inside region and events without any jet with $p_T > 20$ GeV in the region. The two subsamples are referred to as inside–jet tag and inside–jet veto. The jets in the inside region are called inside–jets and the jets in the outside region are called outside–jets. A fourth topology with at least one outside jet with $p_T > 20$ GeV is also studied. The motivation for these configurations is given in section 1.9.

4.1.1 Observables

The main observable of this analysis is the absolute value of the azimuthal difference between the leading central and leading forward jet:

$$\Delta\phi = |\phi_{central} - \phi_{forward}|, \tag{4.1}$$

where $\phi_{central}$ and $\phi_{forward}$ are the azimuthal coordinates of the leading jets. $\Delta\phi$ is further measured for the different scenarios and in slices of the pseudo–rapidity difference, $\Delta\eta$:

$$\Delta\eta = |\eta_{central} - \eta_{forward}|. \tag{4.2}$$

The leading jet p_T in the inside region, which is used to define the scenarios derived from the main selection, is measured. The deviation in pseudo–rapidity of the inside–jet from the mid-point between the leading central and forwards jets is defined as η^*:

$$\eta^* = \eta_{inside} - (\eta_{central} + \eta_{forward})/2, \tag{4.3}$$

where η_{inside} is the rapidity of the leading inside–jet. The leading jet p_{T} outside the dijet system is measured, as well as $\Delta\eta^{out}$ between the leading outside jet and outhermost jet in the dijet system:

$$\Delta\eta^{out} = \min(|\eta_{outside-jet} - \eta_{central-jet}|, |\eta_{outside-jet} - \eta_{forward-jet}|). \tag{4.4}$$

4.2 Data Samples

In this analysis data recorded with the CMS detector in 2010 are used. An overview of the datasets, run intervals and statistics is presented in table 4.1.

Table 4.1: Data samples used in the analysis

Sample	Runs	Statistics
/JetMETTau/Run2010A-Apr21ReReco-v1/AOD Short name: JetMETTau_2010A	135821 - 141887	284.2 nb^{-1}
/JetMET/Run2010A-Apr21ReRecov1/AOD Short name: JetMET_2010A	141950 - 144114	2.896 pb^{-1}
Total	-	3.180 pb^{-1}

The events stem from pp collisions at $\sqrt{s}$=7 TeV with low pile–up conditions. The period under analysis has several millions of events recorded due to low trigger prescales for low p_T jets. Late 2010 and the whole 2011 period were not used because of the higher pile-up, which would complicate the task of isolating a single hard interaction, and the high trigger prescales would prevent an increase in effective statistics.

The data have been reprocessed in April 2012 with CMSSW [86] version 4_2_1_patch2, which used the most up-to-date detector calibration. It corresponds to a total integrated luminosity of 3.18 pb^{-1}.

4.2.1 MC Samples

Both official CMS and privately produced MC samples have been used in this analysis. Official MC samples with large statistics are produced by the collaboration and can be used by many analyses. A private production is done by the analysers to get extra predictions which don't need full detector simulation and are too specific to be used by other analyses.

Official Production with Hadron and Detector Level

Monte Carlo simulation samples from the Summer 2012 official production have been used, which are generated and simulated with CMSSW 4_2_4_patch2, including the low pile–up conditions from 2010 data. Two different generators have been used: PYHTIA 8 with Tune 4C and PYTHIA 6 with Tune Z2*, both of them with pile–up simulated to match the 2010 conditions. In addition to hadron information, the detector level prediction is present, which has full detector simulation performed with GEANT 4 and reconstruction. An overview over the datasets used is presented in Table 4.2.

Table 4.2: Monte Carlo samples used in this analysis

Sample	Statistics
QCD_Pt_XtoY_fwdJet_TuneZ2star_HFshowerLibrary_7TeV_pythia6/ Summer12-LowPU2010_DR42-PU_S0_START42_V17B-v1/AODSIM	2×10^7 Events
QCD_Pt_XtoY_fwdJet_Tune4C_HFshowerLibrary_7TeV_pythia8/ Summer12-LowPU2010_DR42-PU_S0_START42_V17B-v1/AODSIM	4×10^7 Events

The p_{T} slices, denoted above as XtoY, avaliable for both MC samples are 10to25, 25to40, 40to80, 80to150. For PYTHIA 6, there was additionally 150toInf. To improve access, these samples have been turned into NTuples using the same code used by many Forward and Small-x QCD Physics group analyses [171]. NTuples are skimmed datasets, which discard all information and events not relevant for the analysis, thus greatly reducing the disk space used. The Ntuple is stored at EOS, a disk pool project hosted by CERN, where all group members have direct access. The data samples include a generator–level pre–selection requiring at least one forward jet. The NTuple preparation includes a selection step requiring at least one jet with $p_T > 15$ GeV and $|\eta| < 5$.

Hadron–level MC Prediction

The following hadron–level predictions compared to the fully corrected data at hadron–level:

- PYTHIA 6 Tunes Z2* (with and without MPI), P11 and AMBT1;
- PYTHIA 8 Tune 4C;
- HERWIG++ 2.5 Tune UE3;
- HERWIG 6.5.

Further details on these generators and tunes are given in section 3.3.

4.3 Pile–up Reweighting

The distribution of the number of reconstructed vertices of the generated MC samples in general does not describe any of the data periods. PYTHIA 6 and PYTHIA 8 predictions agree with each other, both displaying an excess of pile-up. This excess was added to allow a reweigthing to match different pile–up conditions. This difference between the pile–up conditions of 2010 data and the MC as function of number of reconstructed vertices, after the inclusive selection (see section 4.4), is shown in figures 4.2 and 4.3.

The vertex distribution varies between the datasets, but even the sample with the highest pile-up contribution is not described by the avaliable MC samples. The discrepancy increases with increasing vertex multiplicity and can reach several orders of magnitude.

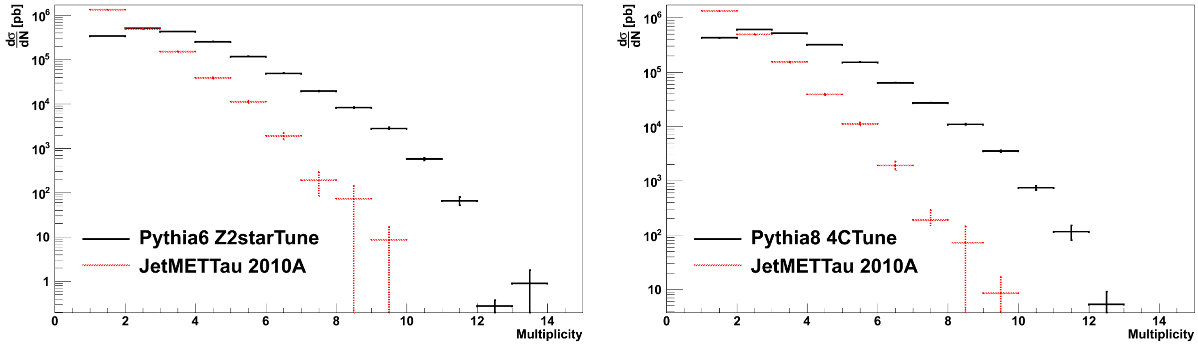


Figure 4.2: Number of vertices for the JetMETTau_2010A dataset compared with (left) PYTHIA 6 Tune Z2* and (right) PYTHIA 8 Tune 4C after the inclusive selection.

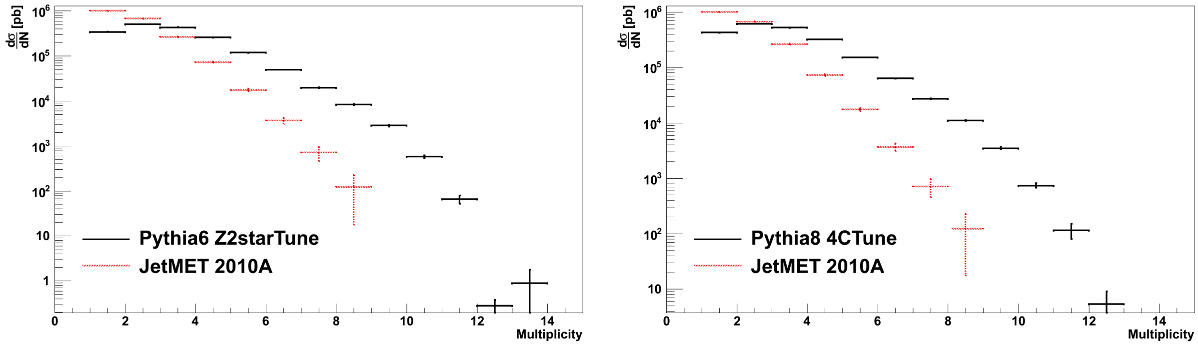


Figure 4.3: Number of vertices for the JetMET_2010A dataset compared with (left) PYTHIA 6 and (right) PYTHIA 8 after the inclusive selection.

The offical recipe demands the use of the official script to estimate the average pile-up. However, this recommendation is not extended to 2010, because the necessary pile–up information is missing for the 2010 period. As a further argument for using a simpler method, the average pile–up in 2010 data is around 1.2, a small effect compared to > 10 seen in 2011/12.

To deal with the problem, a vertex reweighting procedure was applied. A factor was computed:

$$Ratio_{bin\ i} = \frac{\sigma^{Data}_{bin\ i}}{\sigma^{MC}_{bin\ i}} \tag{4.5}$$

where $Ratio_{bin\ i}$ is the ratio as function of number of vertices, $\sigma^{Data}_{bin\ i}$ and $\sigma^{MC}_{bin\ i}$ are the cross-sections for a given number of vertices for data and Monte Carlo, respectively. That factor is then multiplied by the generated pile–up in the MC as an event weight, in both official MC samples and for each data-taking period, using the following relation:

$$Ratio_{bin\ i} \rightarrow g_{pile-up}(bin\ i-1) \tag{4.6}$$

where $g_{pile-up}(bin\ i-1)$ is the factor applied on simulated pile–up.

Due to detector inefficiences and limited resolution, vertex multiplicity and simulated pile–up are not 100% correlated, therefore an iterative approach has been used. The described process is thus repeated 5 times, using the factor obtained in the last iteration to be the starting point for the next one, as an event weight, to allow the method to converge. The evolution of the difference between iterations can be found in the figures 4.4 and 4.5 for different pile–up periods.

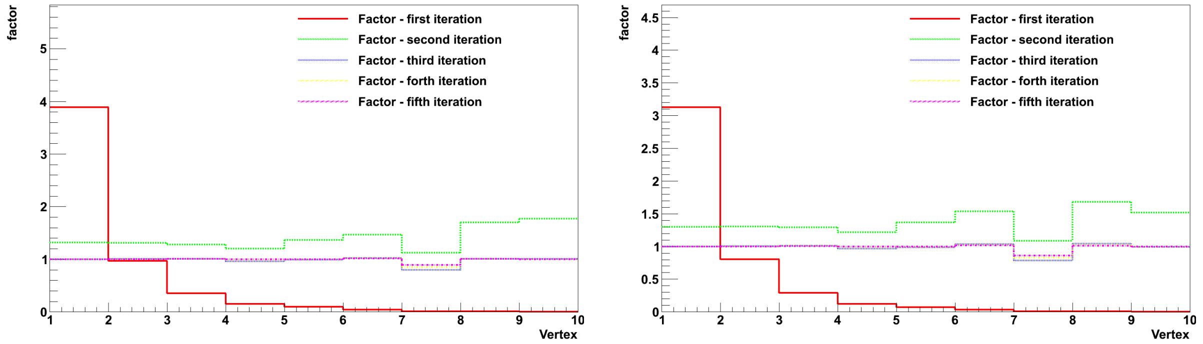


Figure 4.4: Pile–up reweight factor variation between iterations for (left) Pythia 6 Tune Z2* and (right) Pythia 8 Tune 4C using the JetMETTau_2010A dataset.

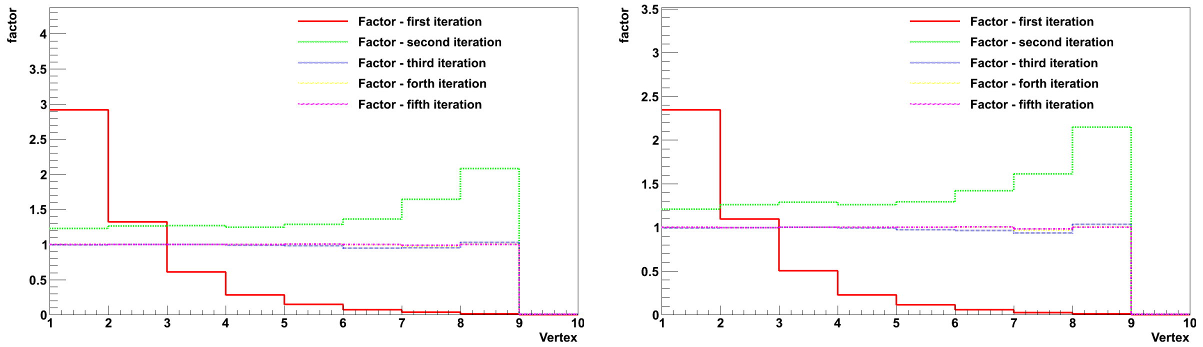


Figure 4.5: Pile–up reweight factor variation between iterations for Pythia 6 Tune Z2* (left) and Pythia 8 Tune 4C (right) using the JetMET_2010A dataset.

The evolution is similar for both MC samples reweighted with both data samples, supporting the convergence of the method, as the factors between iterations tend to 1. The event weight factors obtained after the fifth iteration are presented in the figure 4.6.

The reweighting factors are consistent for the two data samples. They illustrate the slightly different pile-up distributions obtained for both data taking periods. The difference between the two last iterations can be assigned as uncertainty for this method, which is in order of 0.1%. Therefore, it is considered to be negligible and it is not propagated to the final uncertainty.

4.4 Event Selection

4.4.1 Event Selection on the Ntuple

This analysis has been performed by using the QCD Ntuple made by the FSQ (**F**orward and **S**mall–x **QCD**) Group using Kousouris's code [171] and the CMS software version 4_2_4_patch2. This code transforms the reconstructed data samples into simple root trees. It contains all the information required to do a jet analysis with the most up-to-date detector reconstruction and subsequent corrections (see chapter 5). The common event selection of the NTuple is the following:

1. An official JSON (**J**ava**S**cript **O**bject **N**otation) file, containing the list of runs in which the whole the detector was working properly, as certified by the experts of each sub–system, from the Apr21ReReco reconstruction campaign is used.

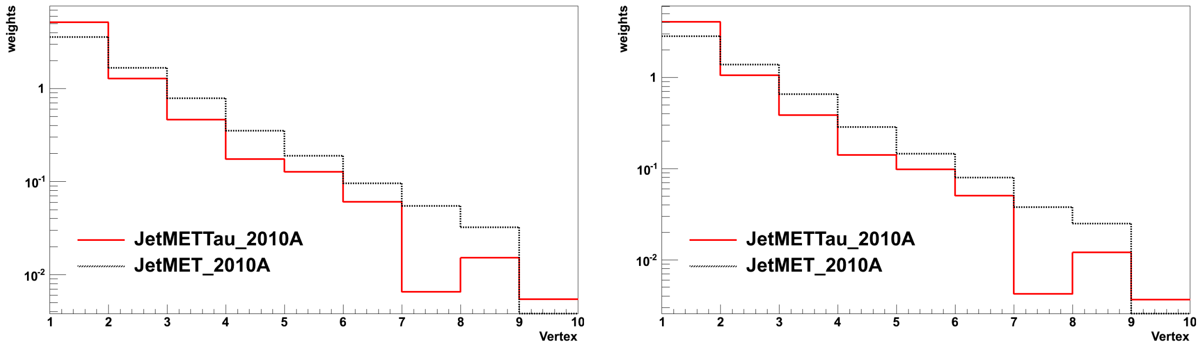


Figure 4.6: Reweight factor for the analysed data samples for Pythia 6 Tune Z2* (left) and Pythia 8 Tune 4C (right).

2. At least one jet with $p_{\mathrm{T}} > 15$ GeV within the rapidity range $0 < |\eta| < 5$ was selected. All jets with uncorrected transverse momentum $p_{\mathrm{T}} < 5$ GeVare omitted in the Ntuple.

4.4.2 Trigger Selection

For this analysis a signal from at least one of the following prescaled single jet triggers was required: HLT_Jet15U, HLT_30U or HLT_50U. The number after *JET* is the p_{T} threshold in GeV and the following *U* denotes that it is an uncorrected quantity. A detailed explanation of the trigger efficiency studies and combination method are presented in section 5.1.

4.4.3 Primary Vertex

Only events with at least one primary, non–fake vertex with $|z| < 24$ cm are accepted. the primary vertex was reconstructed with deterministic annealing [172] using at least 4 degrees of freedom. Fake vertices are generated from hadronic decays. In this analysis only non–fake vertices, which contain tracks coming directly from the hard scattering, are used.

4.4.4 Physics Selection

The four different topologies, discussed in section 4.1, required different physics selection. There selection is applied on top of the cuts referred above.

1. **Inclusive Scenario**: at least one central jet, $|\eta| < 2.8$, with $p_{\mathrm{T}} > 35$ GeV and at least one forward jet, $3.2 < |\eta| < 4.7$, with $p_{\mathrm{T}} > 35$ GeV.

2. **Inside–Jet Veto Scenario**: Inclusive Scenario requirements and no inside–jets with $p_{\mathrm{T}} > 20$ GeV.

3. **Inside–Jet Tag Scenario**: Inclusive Scenario requirements and at least one inside–jet with $p_{\mathrm{T}} > 20$ GeV.

4. **Outside–Jet Scenario**: Inclusive Scenario requirements and at least one outside–jet with $p_{\mathrm{T}} > 20$ GeV.

4.4.5 Results of the Selection

The results of the selection procedure described above are presented in table 4.3.

From the total 28485152 events in the NTuple, 8134200 from the JetMETTau_2010A and 20350952 from the JetMET_2010A samples, 27834769 passed the Primary Vertex cut and, from

Table 4.3: Selection results

Sample	JetMETTau_2010A	JetMET_2010A	Total
Events read	8134200	20350952	28485152
Events after the PV cut	7951080	19883689	27834769
Events after the trigger cut	4179779	10683202	14862981
Events after inclusive scenario	61286	136803	198089
Events in the inside–jet tag scenario	17700	56880	74580
Events in the inside–jet veto scenario	43586	79923	123509
Events in the outside–jet tag scenario	17872	54980	72852

those, 14862981 were triggered. The inclusive scenario selection was fulfilled for 198089 events, in which 74580 belong to the inside–jet tag topology while 123509 belong to the inside–jet veto scenario. The outside–jet tag scenario have 72852 events.

4.5 Control Distributions

In this section, the various distributions that demostrate how well the MC simulation describes the data are presented. These distributions are obtained after the main selection and vertex reweighting. They show the number of events divided by bin width and luminosity and they are not corrected for detector effects (detector level).

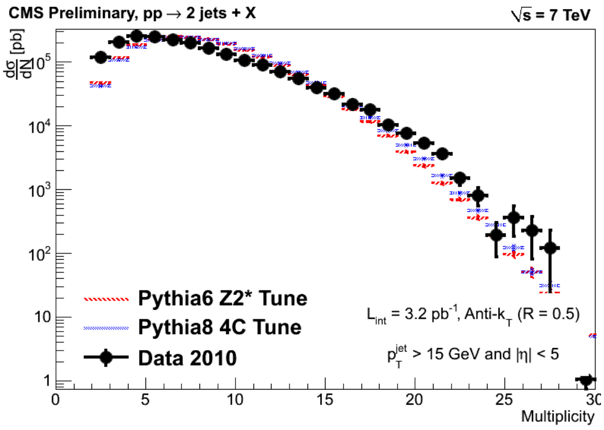


Figure 4.7: Detector level MC and uncorrected data comparison for the jet multiplicity after the inclusive selection for jets with $p_\mathrm{T} > 20$ GeV.

A comparison between data and Monte Carlo simulation on detector level for the jet multiplicity is presented in figure 4.7, jets with $p_T > 20$ GeV are required both on the data and MC. The spectra start at 2 and rise up to around 5 and then fall exponentially. The predictions deviate for low (2-5) and sligthly for high (> 18) jet multiplicity. Both Pythia 6 and Pythia 8 agree in the predicted multiplicity.

A comparison of the vertex multiplicity in data and MC is presented on figure 4.8. As expected by construction, after reweighting the steeply falling vertex distribution is fully described by MC. The z–coordinate of the primary vertex, which has a gaussian behaviour centered at zero, is also well described, being a valuable cross–check for the simulation of the vertex detection.

The p_T of the leading central jet is presented in figure 4.9. It displays a power–like decreasing function with increasing p_T. Except for the middle p_T region, from 57 to 150 GeV, where the

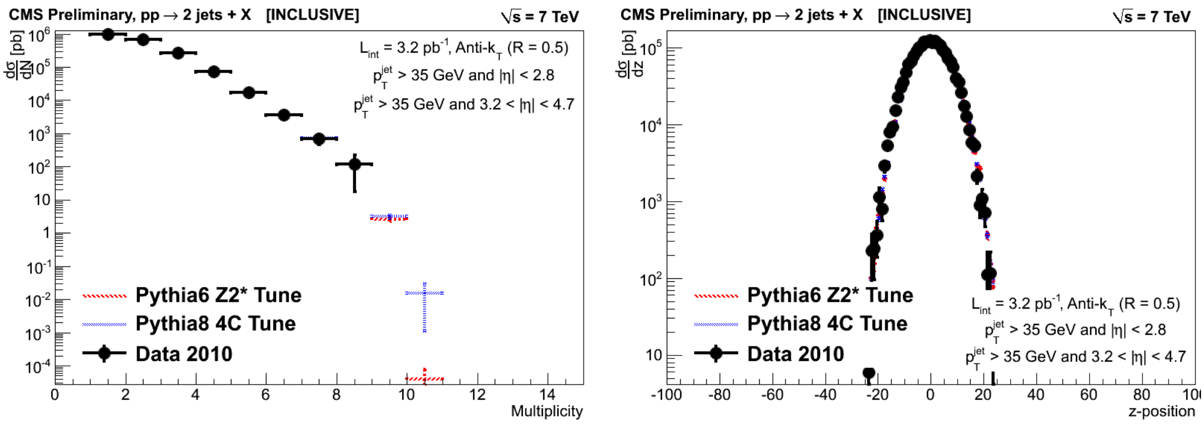


Figure 4.8: Detector level MC and uncorrected data comparison for (left) vertex multiplicity and (right) z–position after the inclusive selection and using the vertex reweighting.

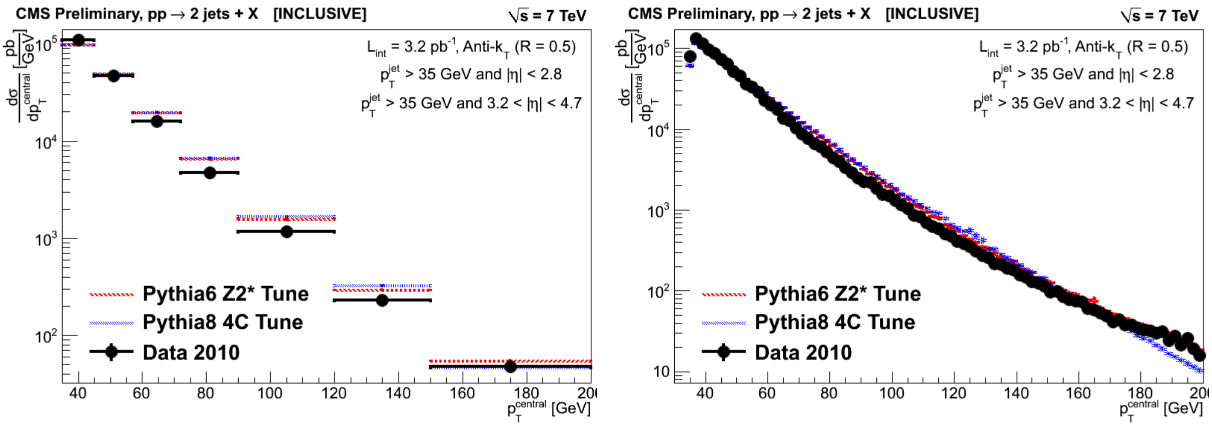


Figure 4.9: Detector level MC and uncorrected data comparison for leading central jet p_T after inclusive selection. Broad binning (left) and fine binning (right).

MC has a small excess when compared to data, both the broader and fine binning show a good agreement.

In figure 4.10 the p_{T} of the leading forward jet is presented. As expected, the forward jet spectrum falls more steeply than of the central jet. The shape does not agree with MC, showing a large deviation for higher p_{T}. Such behaviour has already been observed [76]. Both generators yield very similar predictions in the lower and medium p_{T} regions, and the difference in the high p_{T} region is due to the lack of events generated by PYTHIA 8 in that phase-space.

The pseudo–rapidity of the leading forward and central jets is shown in figure 4.11. The leading central jet η spectrum decreases with increasing pseudo-rapidity and is described by both MCs. The leading forward jet η, which is a steeply falling distribution, shows larger discrepancies with increasing η where MC understimates the data.

The azimuthal angle of the leading forward and central jets is shown in figure 4.12. As expected the distributions are flat within uncertainties and described by the MC predictions.

The pseudo–rapidity separation between the two leading jets, $\Delta\eta$, is presented in the figure 4.13. The distribution has a maximum for $\eta \simeq 3$, due to the selected phase–space. The data is very well described by MC, except for the last bin, where a small difference in the order of 10% is found.

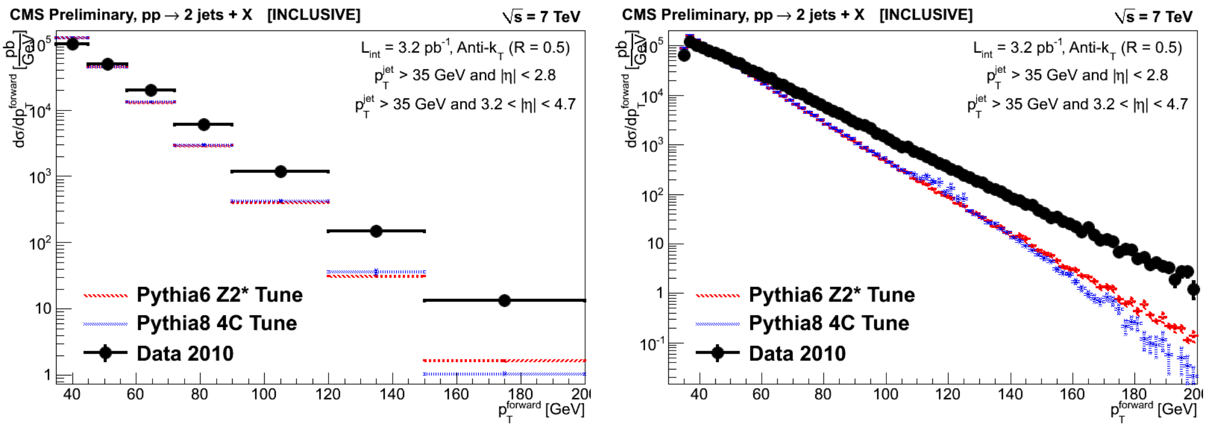


Figure 4.10: Detector level MC and uncorrected data comparison for leading forward jet p_T after inclusive selection. Broad binning (left) and fine binning (right).

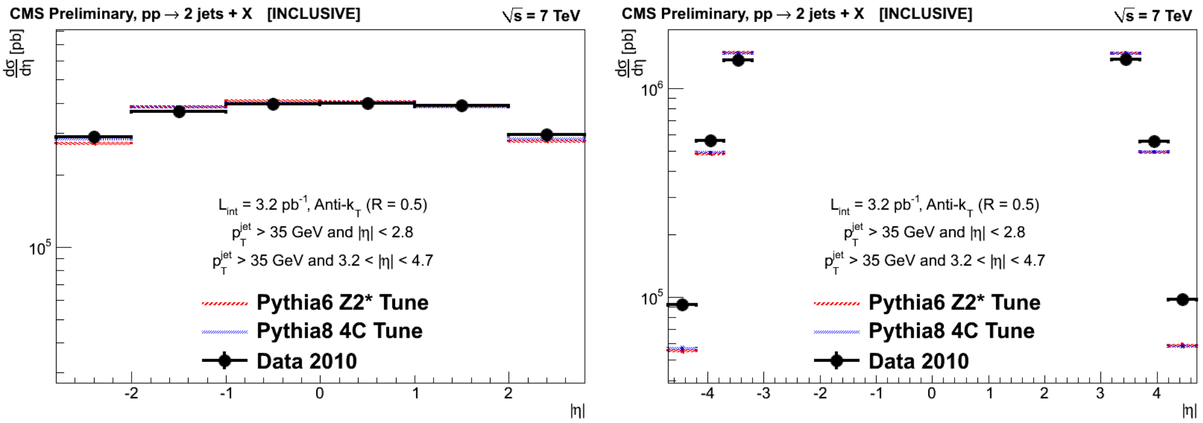


Figure 4.11: Detector level MC and uncorrected data comparison for leading jets η after main selection. Central jets (left) and forward jets (right).

The pseudo-rapidity separation between the two leading jets $\Delta\eta$ in the inside–jet veto and inside–jet tag scenarios is presented in figure 4.14. In the absence of an inside–jet, the two first bins have almost the same value and the distribution decreases for $\eta > 3.5$. The agreement between the MC and data is very good. For the inside–jet tag scenario the distribution peaks at $\eta > 3.5$. Therefore, the discrepancy present in the last $\Delta\eta$ bin in the figure 4.13 is caused by the presence of the inside-jet, as can be observed in the last bin on right side of figure 4.14.

Figure 4.15 shows the leading η of the inside–jet on the left panel and outside–jet on the right panel. In the case of the inside–jet, the distribution is almost flat for the central region $|\eta| < 3$, then falls exponentially, except for a sligthly dip at the center. The agreement is almost perfect, except for the highest η region, where the MC underestimates the data, as seen for the forward jet. For the outside–jet, a three-peak structure is found, the bigger peak being in the center and the other two located at $\eta \simeq 3.5$. The MC overestimates the data for the central region and then underestimates for the forward region.

Figure 4.16 shows the ϕ distribution of the leading inside–jet (left) and leading outside–jet (right). As expected from ideal uniform distribution, both distributions are nearly flat. For the inside–jet the MC tends to slightly underestimate the cross-section while for the outside–jet it overestimates the data.

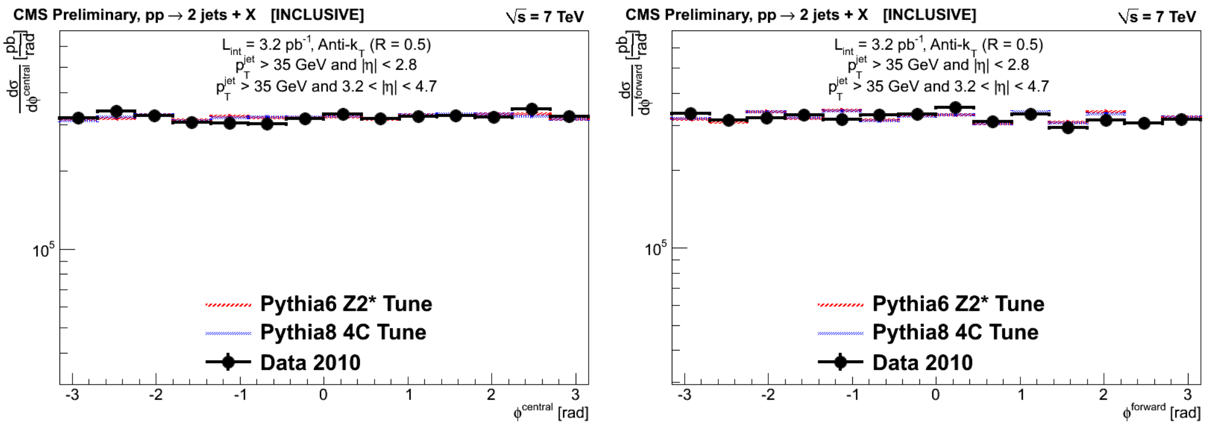


Figure 4.12: Detector level MC and uncorrected data comparison for leading jets ϕ after the inclusive selection. Central jets (left) and forward jets (right).

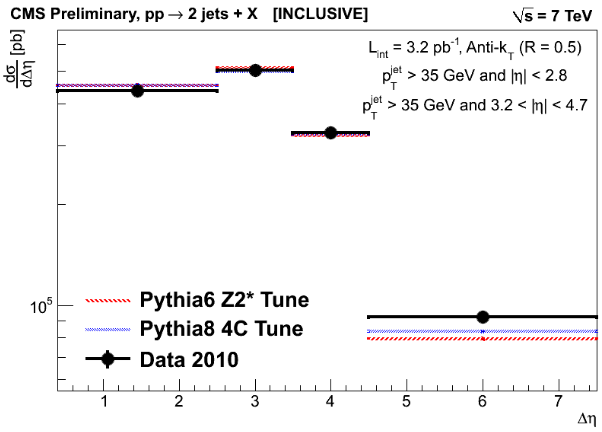


Figure 4.13: Detector level MC and uncorrected data comparison for $\Delta\eta$ after the inclusive selection.

The jet multiplicity is presented in figure 4.17 for the inside-jets (left) and outside–jets (right). The multiplicity for the inside–jets agrees reasonably with data and a small discrepancy in the first two bins and in the last bin is observed. The predicted shape for the outside–jets multiplicity is not described by MC but, at least, the two generators predictions agree within statistical uncertainties, represented as error bars.

The total jet p_T for the inside–jets (left) and outside–jets (right) is shown in figure 4.18. The total p_{T} is the sum of the p_{T} modules of all jets in a given region. The p_{T} distribution of the inside–jet has a maximum at $p_T \simeq 45$ GeV. The distribution is well described by the MC except for $p_T > 90$ GeV, where the MC underestimates the data. The outside–jet p_{T} data distribution has a maximum at $p_{\mathrm{T}} \simeq 55$ GeV while the MC predicts the maximum for $p_{\mathrm{T}} \simeq 45$ GeV. Thus, the cross-section is described except for the central region between 35 and 90 GeV where the MC overestimates the data.

The $\Delta\phi$ distribution for the inclusive selection is presented in figure 4.19. The distribution is steeply rising towards larger values of $\Delta\phi$. Both MC predictions agree with each other in their predictions but yield a too strong correlation. The MC underestimates the data for the lower $\Delta\phi$ region, $\Delta\phi < 2$.

The $\Delta\phi$ distribution is presented in figure 4.20 for the inside–jet veto (left) and inside–jet tag (right) scenarios. Both distributions are steeply rising towards larger values of $\Delta\phi$. In the

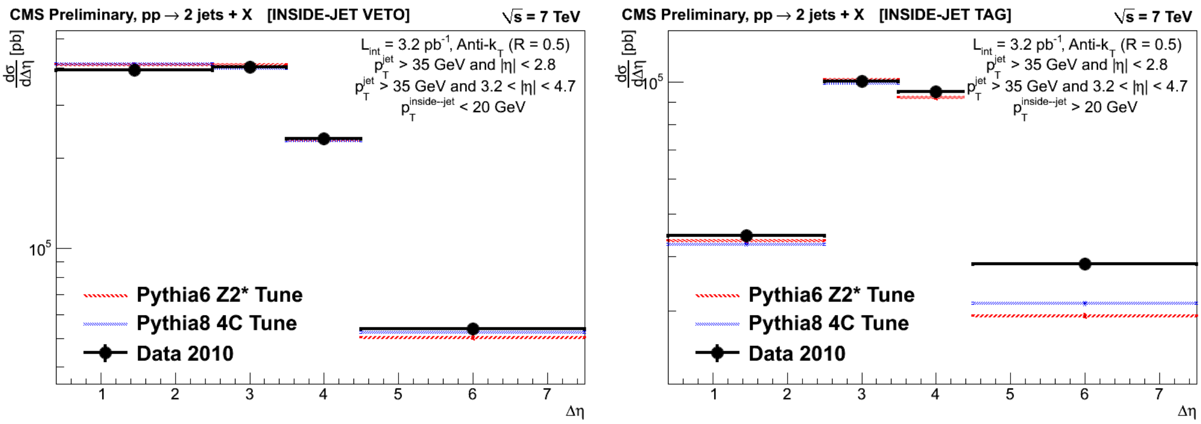


Figure 4.14: Detector level MC and uncorrected data comparison for $\Delta\eta$ for different scenarios: inside–jet veto (left) and inside–jet tag (right).

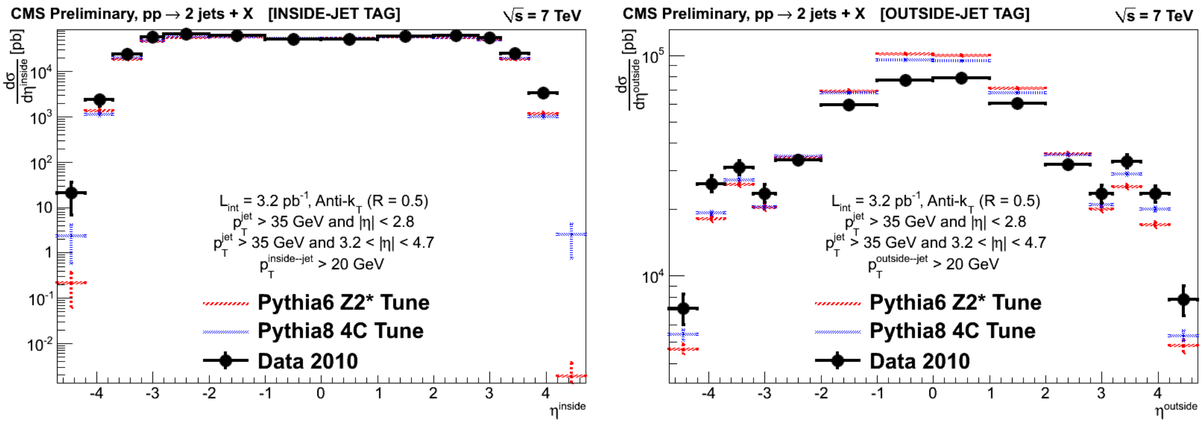


Figure 4.15: Detector level MC and uncorrected data comparison for leading inside–jet (left) and leading outside–jet η (right).

presence of the inside–jet the distribution flattens out and starts at higher value for the first bin. The MC in the inside–jet veto scenario predicts a too strong correlation while in the inside–jet tag scenario the MC is close to the data.

The $\Delta\phi$ distribution for different regions of pseudo-rapidity separation in the inclusive scenario are presented in figure B.1 in appendix B. In general, across all the $\Delta\eta$ slices, the same trends as for the inclusive measurement are found. The slope is very similar and the cross-section decreases with increasing $\Delta\phi$. The MC fails to describe the low $\Delta\phi$ region, underestimating the data.

The $\Delta\phi$ distribution for different regions of pseudo-rapidity separation in the inside–jet veto scenario is shown in figure B.2 in appendix B. MC does not describe data as the $\Delta\eta$ separation increases because it predics an increasing slope when compared to data. That discrepancy mainly present in the low $\Delta\phi$ region.

The $\Delta\phi$ distribution for different regions of pseudo-rapidity separation in the inside–jet tag scenario is presented in figure B.3 in appendix B. Overall, the slope is much flatter than in the main scenario and is almost constant as function of $\Delta\eta$. In general, MC describes data, except for the larger rapidity separation region, where the simulation gives too low values.

The leading inside–jet p_{T} (left) and $\eta*$ (right) are presented in figure 4.21. The p_{T} spectrum is a steeply falling distribution. PYTHIA 8 fails to describe the data for higher p_{T}, while PYTHIA

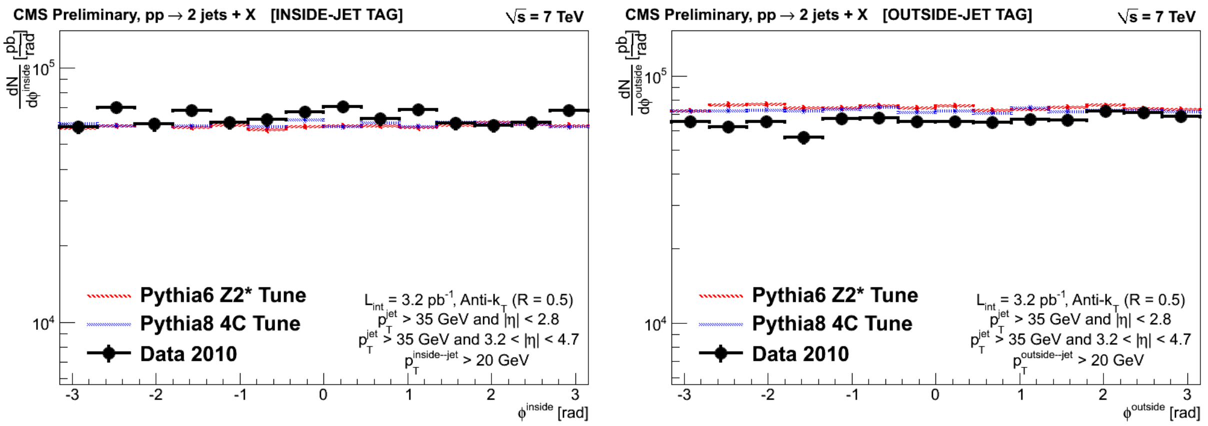


Figure 4.16: Detector level MC and uncorrected data comparison for the ϕ distribution in the leading inside–jet (left) and leading outside–jet (right).

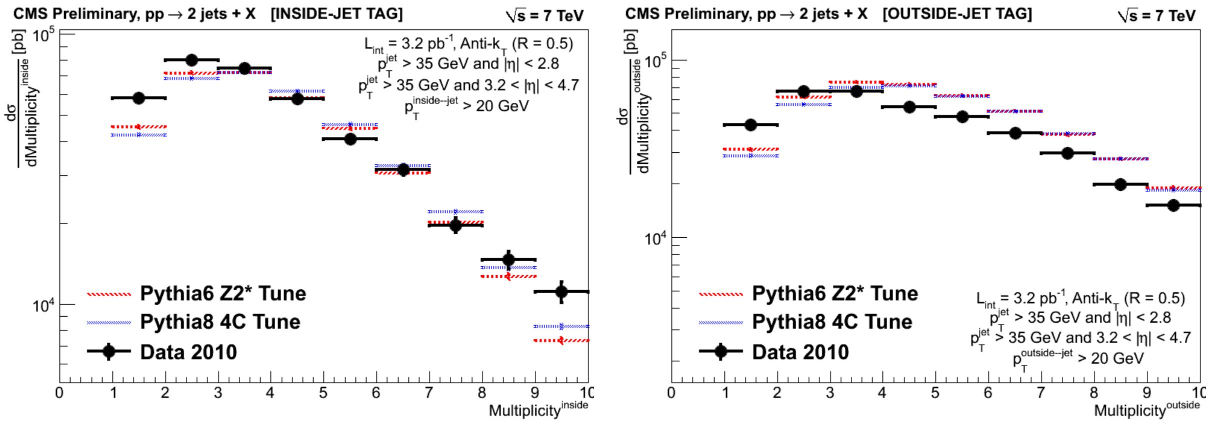


Figure 4.17: Detector level MC and uncorrected data comparison for jet multiplicity for the inside (left) and outside (right) regions.

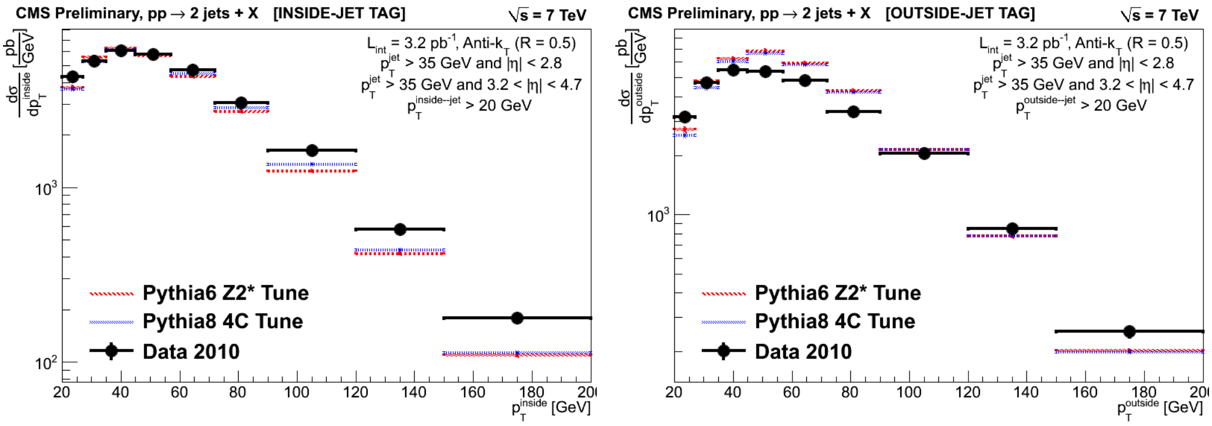


Figure 4.18: Detector level MC and uncorrected data comparison of total jet p_T for inside (left) and outside (right) regions.

6 provides a very good prediction. The η^* distribution has a gaussian–like structure centered at 0 and both MC samples slightly underestimate the η^* in the central region.

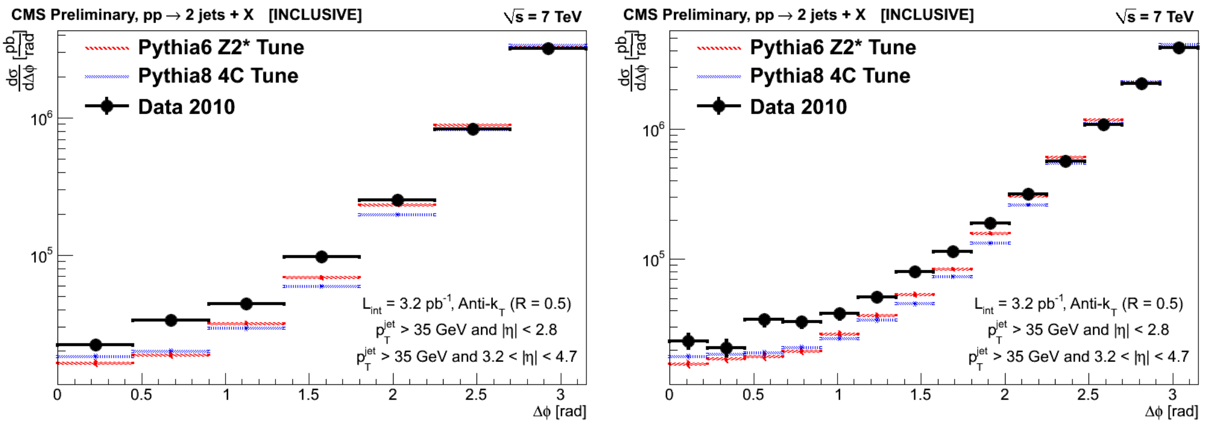


Figure 4.19: Detector level MC and uncorrected data comparison for $\Delta\phi$ in the inclusive selection: broad (left) and fine binning (right).

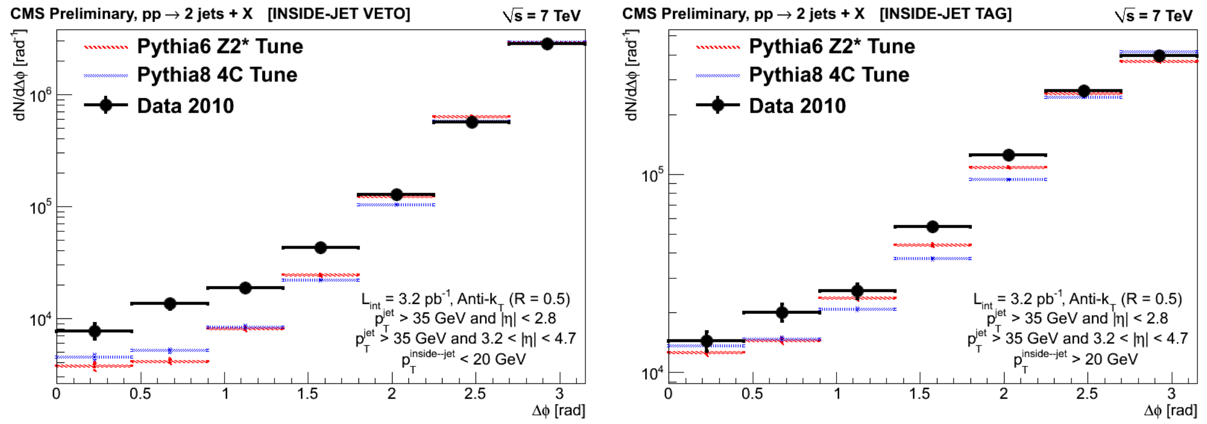


Figure 4.20: Detector level MC and uncorrected data comparison for $\Delta\phi$ in the inside–jet veto (left) and inside–jet tag (right) scenarios.

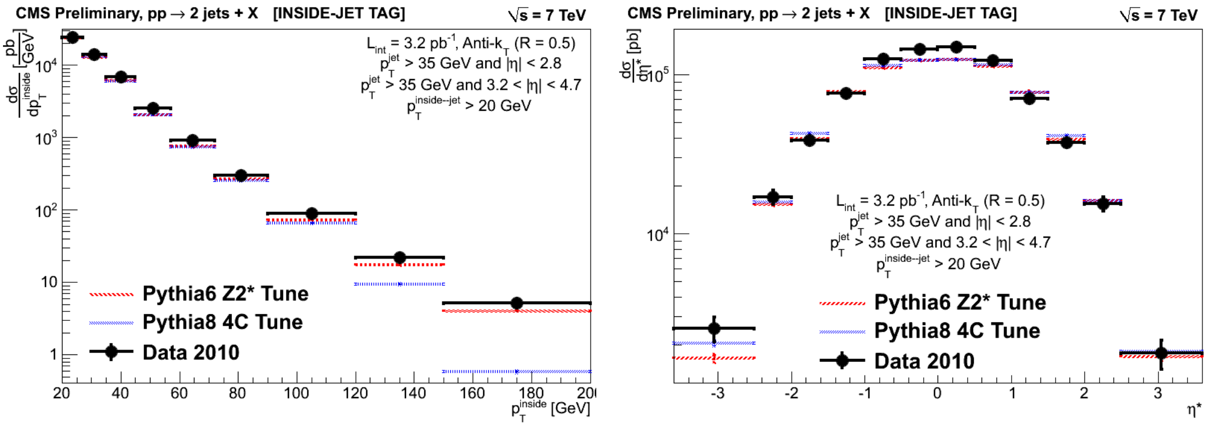


Figure 4.21: Detector level MC and uncorrected data comparison for leading jet p_{T} (left) and $\eta*$ (right) for the inside jet region.

The outside–jet leading p_{T} (left) and $\Delta\eta^{out}$ (right) are shown in figure 4.22. The slope of the distribution is steeply falling and the high–p_T cross-section is underestimated by PYTHIA 8

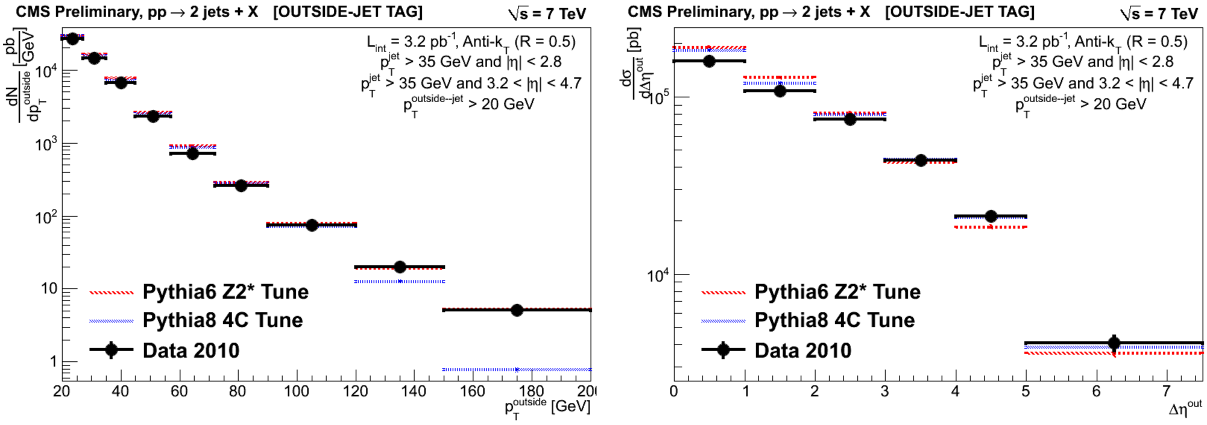


Figure 4.22: Detector level MC and uncorrected data comparison for leading jet p_T (left) and $\Delta\eta^{out}$ (right) for the outside region.

and well described by Pythia 6. The $\Delta\eta^{out}$ spectrum decreases exponentially and both MCs predict a slightly different shape for $\Delta\eta^{out}$, especially in the first bin.

Overall, MC predictions describe relevant observables in a reasonable way, while other quantities, like multiplicity, still pose a serious challenge for Pythia generators.

Chapter 5

Correction of Detector Effects

In this chapter the estimation and correction of various detector effects are discussed. The trigger efficiency (see section 5.1), the resolution of the observables (see section 5.2) and the migration matrices (see section 5.3) are determined. Purity, stability, acceptance, background (see section 5.4) and correction factors (see section 5.5) are computed.

5.1 Trigger

This analysis deals with highly prescaled triggers which are not fully efficient or do not yield significant statistics in the whole phase space. A detailed explanation of the trigger system of the CMS detector has been already presented in the subsection 2.3.7.

5.1.1 Individual Trigger Efficiency

The simplest approach to determine the trigger efficiency of a given trigger is to use an orthogonal trigger. An trigger orthogonal to other trigger, is a trigger which is independent of the other trigger because it is based on a different physics object, for example, an electron trigger can be used as orthogonal to jet triggers. The easiest way is to evaluate the ratio of number of events which fired both the orthogonal and the under study trigger over the number of events which fired the under study trigger:

$$\varepsilon_{trigger} = \frac{N_{events}[fired(trigger) \wedge fired(orthogonal)]}{N_{events}[fired(orthogonal)]}, \tag{5.1}$$

where ε is the efficiency, *trigger* is the under study trigger and *ortogonal* is an orthogonal trigger to the under study trigger.

In this analysis, because of the high values of the prescale, this approach could not be used, since the orthogonal trigger is either prescaled or does not yield enough statistics sop that a turn–on curve cannot be obtained. A turn–on curve shows the transition between the region where the trigger does not fire, $\varepsilon = 0$, to the region where it is fully efficient, $\varepsilon = 1$. Usually it is plotted as a function of p_{T} or energy.

The emulation or reconstruction of the trigger decision using the trigger elements, which describes how the triggers perceive a given physics object, is an alternative method to evaluate the efficiency which can be used under these conditions. The information required to use this method is contained in the NTuple, being the HLT and L1 objects, which are used to make the trigger decision. The efficiency for the emulated trigger is defined as:

$$\varepsilon_{HLT\_JetY} = \frac{N_{emulated}}{N_{all}} \tag{5.2a}$$

$$= \frac{InclusiveRecoJet_{p_T}(HLT\_JetX \wedge L1 - Object_{P_T} > Z \wedge HLT - Object_{P_T} > Y)}{InclusiveRecoJet_{p_T}(HLT\_JetX)}, \tag{5.2b}$$

where the denominator is the number of events which fired the emulator trigger, defined as N_{all}. The value, in p_{T}, of X is chosen lower to that of Y from the trigger list so a trigger with higher threshold is emulated from a lower one. For example, HLT_Jet30U should be emulated with HLT_Jet15U as emulator. The value of Z is the L1 the threshold corresponding to the trigger path of HLT_Jet$Y$. The numerator is the number of events which fired the HLT_Jet$X$ and the p_{T} of the L1 object corresponding to the trigger path is HLT_Jet$X \geq Z$ and the p_{T} of the HLT object corresponding to the trigger path is HLT_Jet$X \geq Y$, defined as $N_{emulated}$. The values of X, Y and Z for each trigger used in this analysis are presented in table 5.1.

Table 5.1: Triggers used in the analysis and respective emulation thresholds.

Trigger	Monitor (X)	L1 threshold (Z)	HLT threshold (Y)
HLT_Jet15U	HLT_L1Jet6U	6 GeV	15 GeV
HLT_Jet30U	HLT_Jet15U	20 GeV	30 GeV
HLT_Jet50U	HLT_Jet30U	30 GeV	50 GeV

The results for the trigger efficiency as a function of the leading–jet p_{T} are presented in figures 5.1 and 5.2. They are obtained using the formula 5.2 with binomial errors.

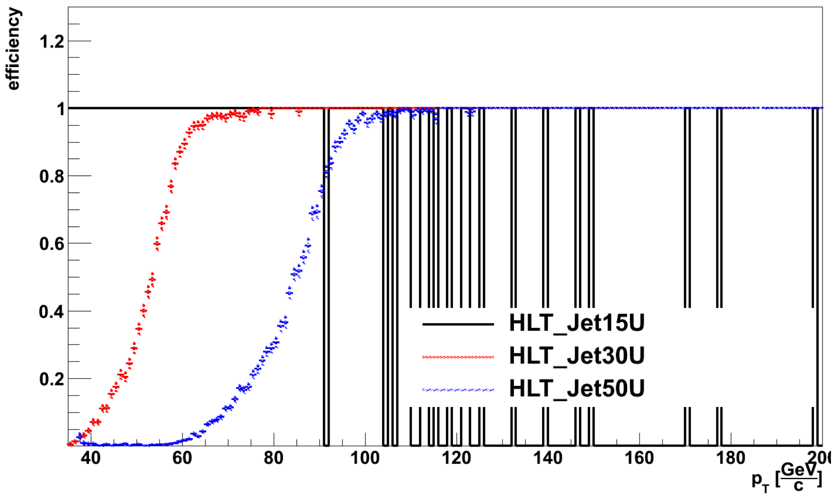


Figure 5.1: Trigger efficiency of HLT_Jet15U (solid black line), HLT_Jet30U (dotted red line) and HLT_Jet50U (dashed blue line) as a function of leading–jet p_{T} for the JetMETTau_2010A dataset.

The curves are very similar for the two datasets. HLT_Jet15U is efficient in the whole phase-space, having its turn–on point for $p_{\mathrm{T}} \leq 35$ GeV. HLT_Jet30U is efficient (more than 95% of events are selected after the emulation) for $p_{\mathrm{T}} > 70$ GeV and HLT_Jet50U is efficient ($> 95\%$) for $p_{\mathrm{T}} > 110$ GeV.

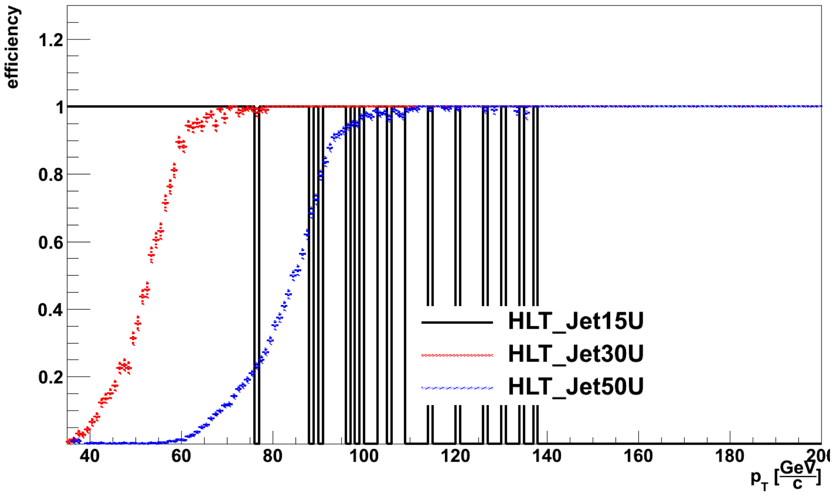


Figure 5.2: Trigger efficiency of HLT_Jet15U (solid black line), HLT_Jet30U (dotted red line) and HLT_Jet50U (dashed blue line) as a function of leading–jet p_{T} for the JetMET_2010A dataset.

5.1.2 Combination Method

As each of the trigger paths is seeded from an unique L1 trigger which is not fully efficient, one can combine the triggers using the division method for the trigger items in one–level systems described in [173]. This method allows to combine several trigger streams by dividing the phase space in regions. In each of these distinct regions, defined in terms of the appropriate variables, only one trigger is used and all other events not triggered by this particular trigger are discarded. To obtain the smallest statistical uncertainty in a given region, the trigger which is fully efficient and provides the larger number of events must be used in that region. This trigger is usually the one with the lower prescale. With this approach the combination is simplified as the analysis can be performed as if several independent subsamples, one for each trigger, are used.

In this analysis, the phase space was divided in three regions according to the leading–jet p_{T}. In each of these regions a fully efficient jet trigger providing the largest possible statistics is used. These subsamples are combined according to the recorded luminosity. The triggers used, leading–jet p_{T} regions and recorded luminosity are presented in table 5.2.

Table 5.2: Trigger Combination

Trigger	Leading–jet p_{T} region (GeV)	Recorded Luminosity (nb^{-1})	
		JetMETTau_2010A	JetMET_2010A
HLT_Jet15U	$35 \leq p_{\mathrm{T}} < 70$	13.799	9.645
HLT_Jet30U	$70 \leq p_{\mathrm{T}} < 110$	117.223	192.895
HLT_Jet50U	$p_{\mathrm{T}} \geq 110$	278.789	2869

5.1.3 Combined Trigger Efficiency

The trigger efficiency for the combination of triggers is evaluated with the following formula:

$$\varepsilon_{combination} = \frac{\sum\limits_{triggers} N^{trigger}_{emulated}}{\sum\limits_{triggers} N^{trigger}_{all}} \tag{5.3}$$

where $N^{trigger}_{emulated}$ and $N^{trigger}_{all}$ are $N_{emulated}$ and N_{all} computed from formula 5.2 for a given trigger.

In figure 5.3 the efficiency of the trigger combination is presented. The left side shows the efficiency as a function of $\Delta\phi$ and the right side as a function of $\Delta\eta$. The overall efficiency is larger than 99%. In the $\Delta\phi$ distribution, the three first bins lie on the 100% mark while the other four are between 0.99 and 1.00. For the last bin in that distribution, the two datasets have a difference bigger than the error, while the other bins agree within the statistical uncertainty. In the $\Delta\eta$ distribution all bins are between 0.99 and 1.00. They efficiency in both datasets agree; only in the second bin the difference is bigger than the statistical error. The efficiencies for the $\Delta\phi$ distribution in slices of $\Delta\eta$ for the main scenario is presented in figure C.1 in the appendix C.

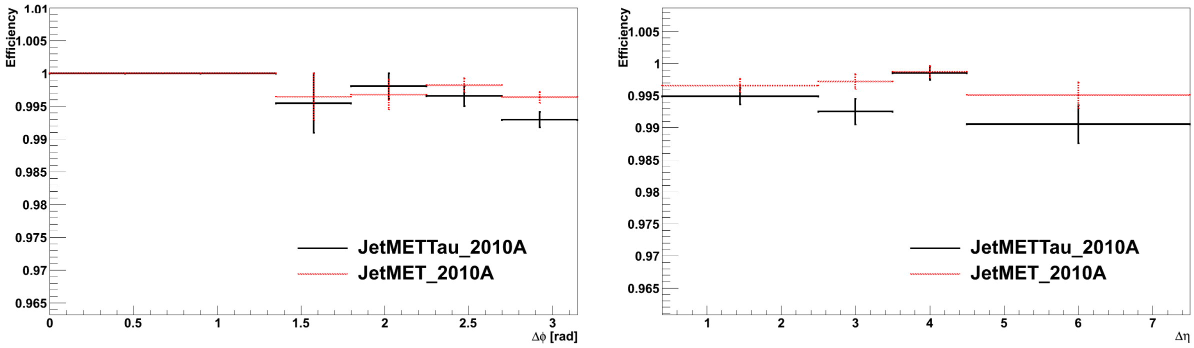


Figure 5.3: Efficiency of the trigger combination for the inclusive scenario in the JetMETTau_2010A (solid black line) and JetMET_2010A datasets (dashed red line). The trigger efficiency as a function of $\Delta\phi$ is presented on the left side and as function of $\Delta\eta$ on the right side.

The trigger efficiency of the trigger combination for the inside–jet veto scenario is presented in figure 5.4 as a function of $\Delta\phi$ and $\Delta\eta$. Both datasets agree within statistical uncertainties, except for the 5th and 7th bins of the left–side distribution. The first four bins in $\Delta\phi$ distribution are fully efficient, while the last three are between 0.99 and 1.00. The efficiency for the $\Delta\eta$ distribution is $> 99\%$ for all bins. The efficiency for the $\Delta\phi$ distributions as a function of $\Delta\eta$ for the inside–jet veto scenario is shown in figure C.2 in the appendix C.

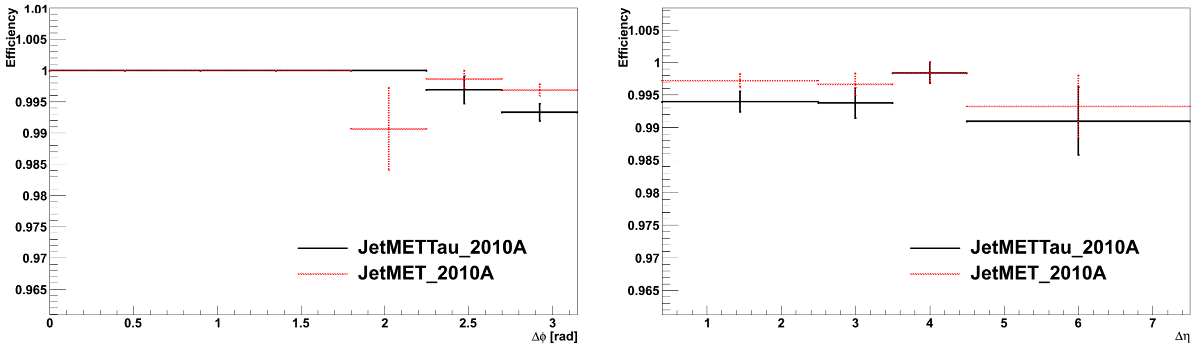


Figure 5.4: Efficiency of the trigger combination for the inside–jet veto scenario in the JetMETTau_2010A (solid black line) and JetMET_2010A datasets (dashed red line). The trigger efficiency as a function of $\Delta\phi$ is presented on the left side and as function of $\Delta\eta$ on the right side.

The combination of the trigger efficiency for the inside–jet tag scenario is shown in figure 5.5, in the left for the $\Delta\phi$ distribution and on the right for the $\Delta\eta$ distribution. Both datasets show an efficiency $> 99\%$ in both distributions. The efficiency for the $\Delta\phi$ distribution as a function of $\Delta\eta$ for the inside–jet tag scenario is presented in figure C.3 in the appendix C.

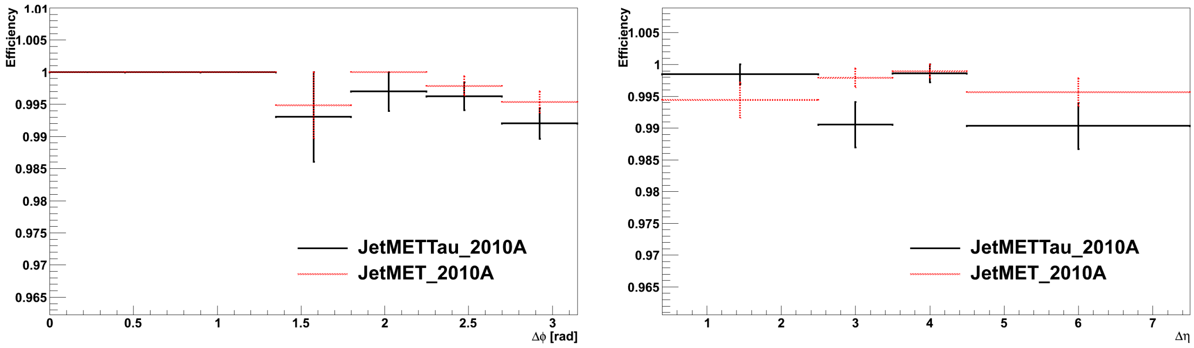


Figure 5.5: Efficiency of the trigger combination for the inside–jet tag scenario in the JetMETTau_2010A (solid black line) and JetMET_2010A datasets (dashed red line). The trigger efficiency as a function of $\Delta\phi$ is presented on the left side and as function of $\Delta\eta$ on the right side.

The efficiency for the trigger combination for the inside–jet distributions is presented in figure 5.6. The efficiency as a function of leading $p_{\mathrm{T}}{}^{inside}$ is shown in the left side, the value is $> 98\%$ for p_{T} between 20 and 27 GeV, $> 99\%$ for p_{T} between 27 and 90 and 100% for $p_{\mathrm{T}} > 100$ GeV. Except for the first bin, the efficiency is similar for both datasets. The efficiency as a function of η^* is presented on the right side of figure 5.6. Except for the last two bins, all the efficiencies are equal or above 99%. Except for the 6th and last bin, the efficiency in both datasets is similar.

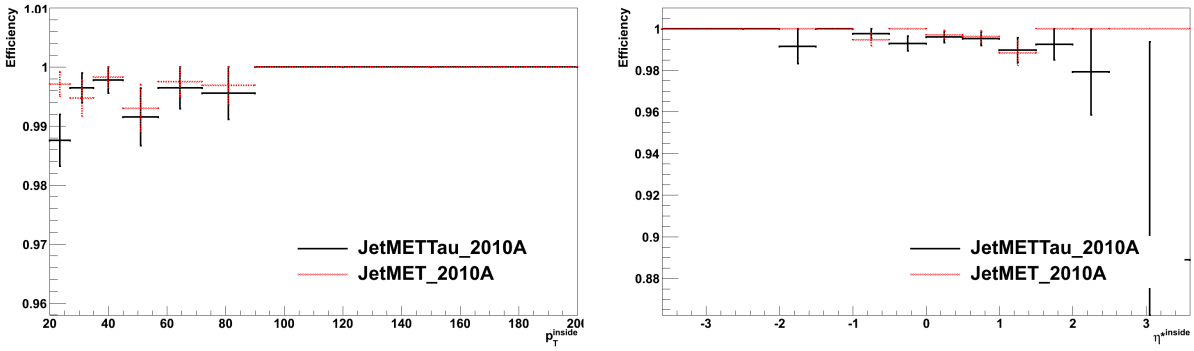


Figure 5.6: Efficiency of the trigger combination for the inside–jet tag scenario in the JetMETTau_2010A (solid black line) and JetMET_2010A datasets (dashed red line). The trigger efficiency as a function of leading p_{T} of the inside jet is presented on the left side and as function of $\eta*$ on the right side.

In figure 5.7 the efficiency for the outside–jet distributions is presented. On the left, the $p_{\mathrm{T}}{}^{outside}$ efficiency is $> 98\%$ for p_{T} between 20 and 90 GeV and fully efficient for p_{T} larger than 90 GeV. The datasets agree for most of the bins. In the right side the $\Delta\eta^{out}$ is displayed. The efficiency is estimated to be larger than 99% for all the bins and in the last bin the value is 100%. The estimates for the different datasets agree for most of the bins.

As the efficiencies are very close to 100%, there is no need for further corrections when using this combination method. The uncertainty of the method is as large as the difference to the full efficiency. This combination scheme will be used in this analysis.

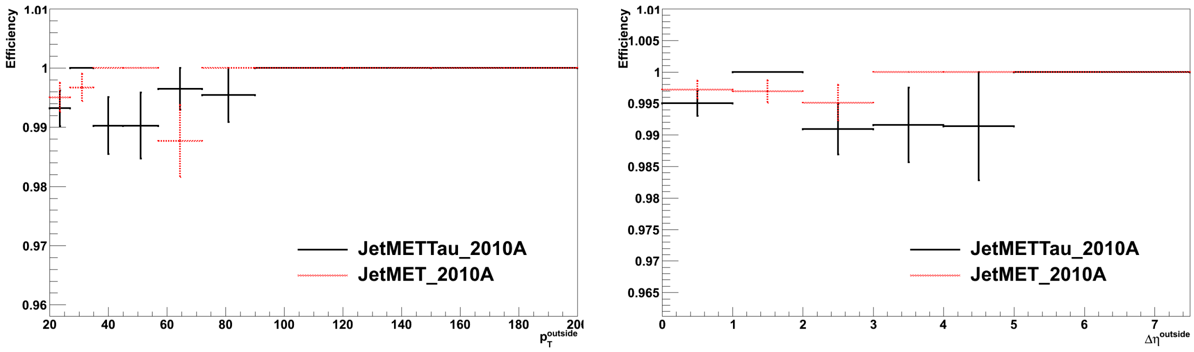


Figure 5.7: Efficiency of the trigger combination for the outside–jet tag scenario in the JetMETTau_2010A (solid black line) and JetMET_2010A datasets (dashed red line). The trigger efficiency as a function of leading p_{T} of the outside jet is presented on the left side and as function of $\Delta\eta^{out}$ on the right side.

5.1.4 Towards the Efficiency Correction

In an attempt to increase the effective statistics, an alternative combination method has been tested. Even though it was not applied to the data, the study developed is presented in this section. It assumes that all triggers are fully efficient for the whole phase space. The probability of an event j to fire a trigger item i is:

$$P_{ij} = \frac{r_{ij}}{d_i} \tag{5.4}$$

where r_{ij} is the trigger bit, with value which is either 0 or 1, and the d_i is the prescale factor. If one assumes that all trigger decisions are independent, the probability P_j that at least one of the events is accepted is:

$$P_j = 1 - \prod_{i=1}^{N_{items}} \left(1 - \frac{r_{ij}}{d_i}\right) \tag{5.5}$$

where N_{items} is the list of triggers used. The effective weight w_j for an event j is then:

$$w_j = \frac{1}{P_j}. \tag{5.6}$$

The weight factor is then computed and applied independently to each event. As seen in figures 5.1 and 5.2, only the HLT_Jet15U trigger path is fully efficient in the whole phase–space. Thus, the efficiency functions for the leading jet for the other two triggers are fitted to derive a correction factor, so this combination methd can be used.

The result from fitting a eight–degree polynomial is shown in figures 5.8 and 5.9. The fit intervals are adjusted to optimize the χ^2. In all cases the fit was successful in describing the turn–on curve and the features on the high–p_{T} edge were removed.

Using the parameters from the fit, a correction factor as a function of the leading jet p_{T} was constructed from the inverse of the polynomial function. The factor was applied between 35 and 80 GeV for the HLT_Jet30U path and between 35 and 116 GeV for the HLT_Jet50U trigger path. To minimize the poor description of the function after the turn-on point, where the deviation of the fitted function from the data was larger than the inefficiency, no correction was used. For the low–p_{T} region, before the turn–on point, where the value of the function was too small (< 0.02) or negative, it was replaced by 0.02 to avoid numerical problems. A cutoff was also applied when the function value was 1.01 or when it started to decrease.

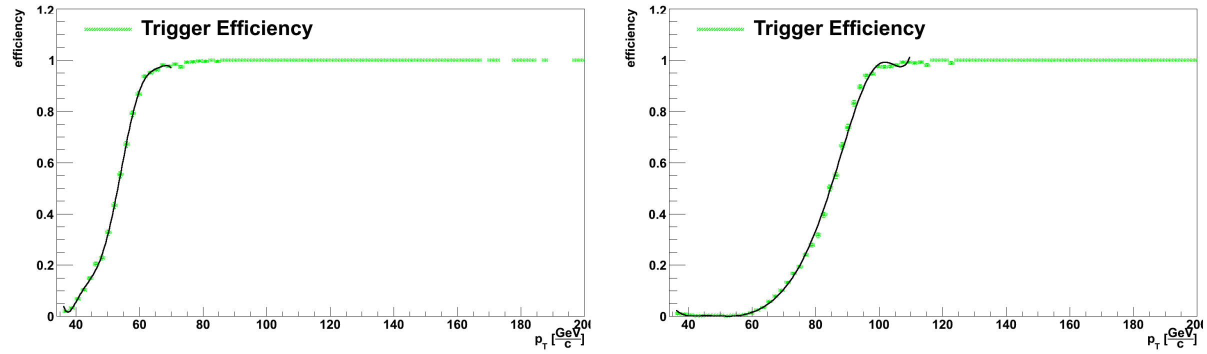


Figure 5.8: Leading Jet $p_{\rm T}$ efficiency of the JetMETTau_2010A dataset for the triggers HLT_Jet30U (left) and HLT_Jet50U (right). The black curve is a fitted 8–degree polynomial.

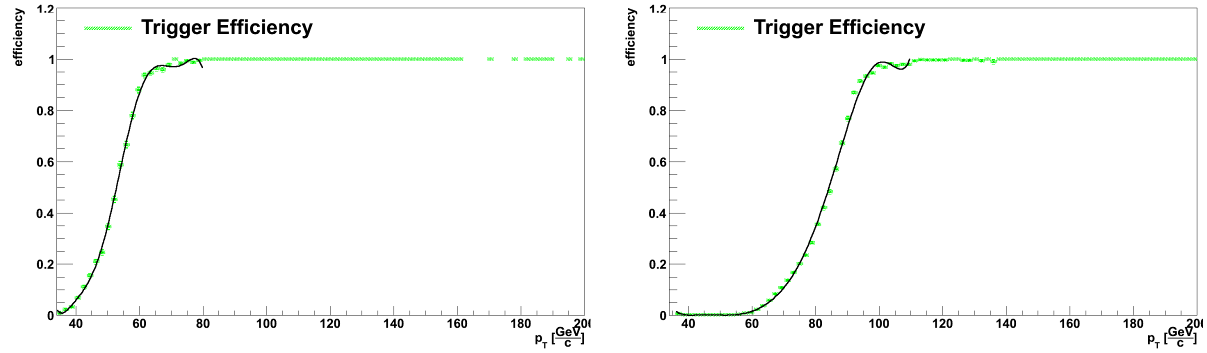


Figure 5.9: Leading Jet $p_{\rm T}$ efficiency of the JetMET_2010A dataset for the triggers HLT_Jet30U (left) and HLT_Jet50U (right). The black curve is a fitted 8–degree polynomial.

In figure 5.10 the efficiency as a function of leading jet $p_{\rm T}$ before and after the trigger correction using the HLT_Jet30U trigger for the JetMETTau_2010A dataset is presented. In the left side, a typical turn–on efficiency curve is seen. On the right side the efficiency after the correction is presented. For $p_{\rm T} < 60$ GeV the correction was too small, leading to an efficiency smaller than 100%, while for $p_{\rm T}$ between 60 and $\simeq 110$ GeV the correction was too large, causing the efficiency to be larger than 100%. The deficit can be attributed to the difficulty to describe the $p_{\rm T}$ curve for low values, while for larger values the presence of the HLT_Jet50U path interfers with the correction, causing double counting. Thus, this correction can not be used.

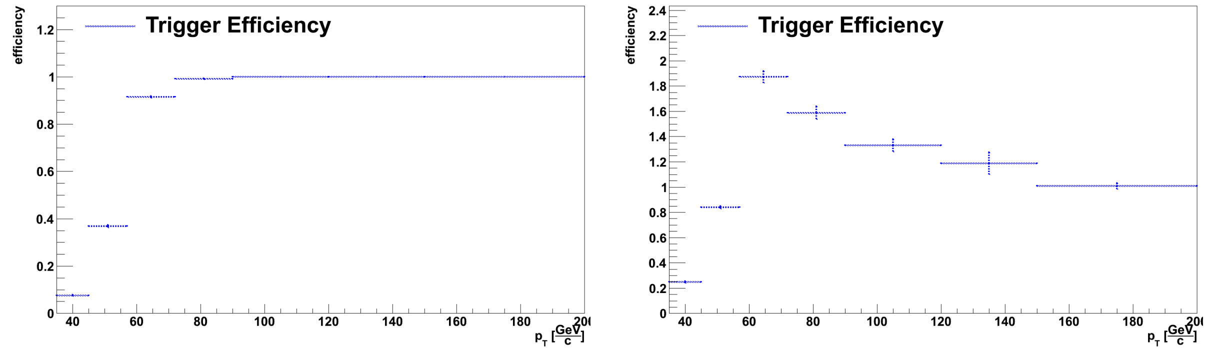


Figure 5.10: Leading Jet $p_{\rm T}$ efficiency before (left) and after (right) trigger correction for the JetMETTau_2010A dataset using HLT_Jet30U trigger.

The efficiency as a function of ϕ of the leading jet, before and after trigger correction, using the HLT_Jet30U trigger path for JetMETTau_2010A dataset is presented in the figure 5.11. In the left side, the curve before the correction is shown, which is flat with an average efficiency close to 50%. After the correction, on the right side, the curve is still almost flat with an estimated efficiency around 100%. No bias from the correction is seen in this variable.

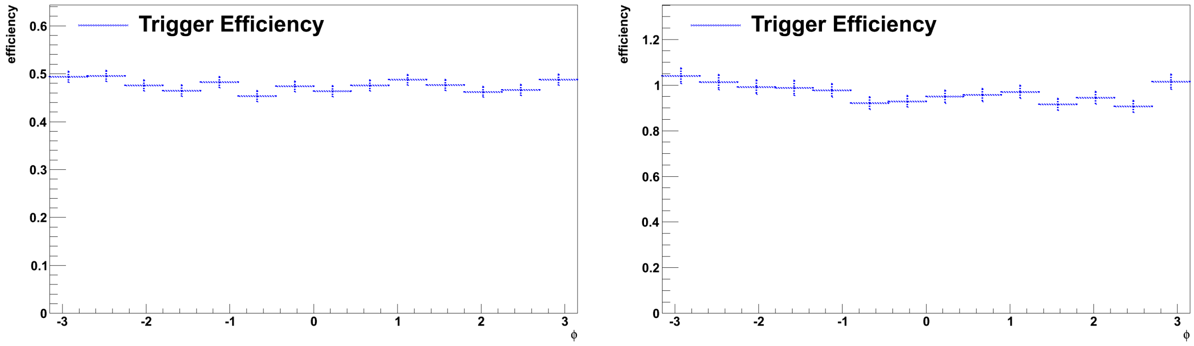


Figure 5.11: Efficiency as a function of leading Jet ϕ before (left) and after (right) trigger correction for the JetMETTau_2010A dataset using HLT_Jet30U trigger.

The efficiency as a function of $\Delta\eta$ before and after trigger correction using HLT_Jet30U trigger path for JetMETTau_2010A dataset is shown in the figure 5.12. As expected and similarly to the previous distribution, the efficiency is almost flat for both plots. It goes from ≃50%, before, to a bit under 100% after the correction. This variable displays no bias from the correction.

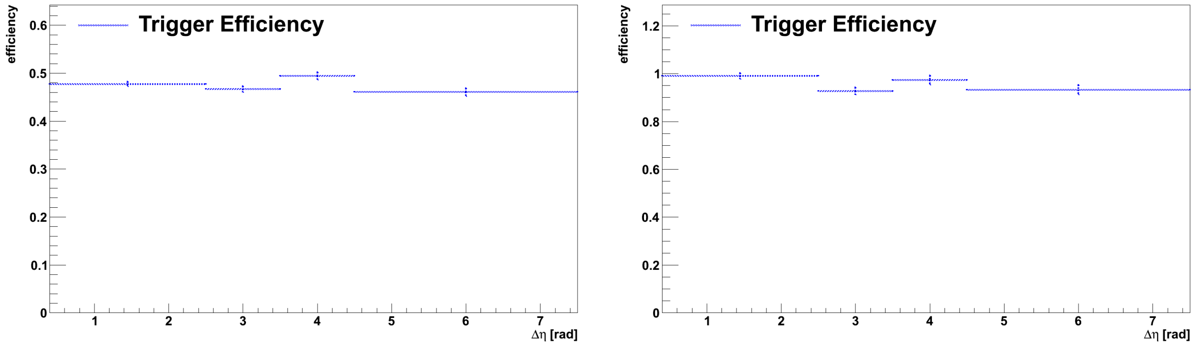


Figure 5.12: Efficiency as a function of $\Delta\eta$ before (left) and after (right) the trigger correction for the JetMETTau_2010A dataset using the HLT_Jet30U trigger.

The efficiency as a function of $\Delta\phi$ before and after the trigger correction using the HLT_Jet30U trigger path for the JetMETTau_2010A dataset is presented in the figure 5.13. Before the correction the distribution does not seem entirely flat, having a slightly better efficiency for the lower $\Delta\phi$ region, with an overall average close to 50%. After the correction the overall efficiency is close to 100%, but the tendency toward the low $\Delta\phi$ region is inverted, the value for the first bin is 80% and a fit of a polynomial of the first order up to the last bin at $\simeq$ 100% is possible. Therefore one can conclude that a small bias from the correction is present in this variable and thus, this correction is likely to bias the final result.

The efficiency as a function of the leading jet p_{T} before and after the trigger correction using the HLT_Jet50U trigger for the JetMETTau_2010A dataset is presented in the figure 5.14. In the left side the expected turn–on curve is displayed, which rises from almost 0 to 1, having the turn–on point $\simeq$ 120 GeV. On the right side, the distribution after the correction grows from 0 at 35 GeV to 7 at ≃110 GeV and then decreases to around 1.5 in the last bin. This behaviour

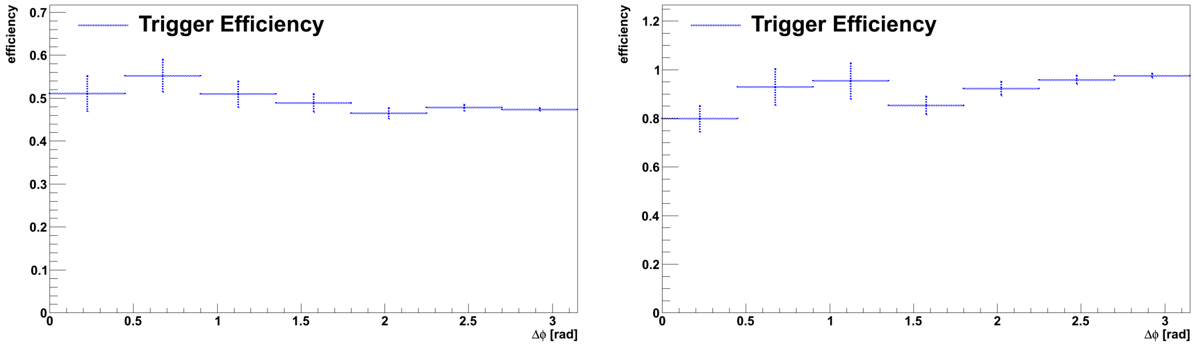


Figure 5.13: The efficiency as a function of $\Delta\phi$ before (left) and after (right) the trigger correction for the JetMETTau_2010A dataset using the HLT_Jet50U trigger.

can only be explained by the bad quality of the fit and double counting. Thus, a huge bias is seen in this distribution.

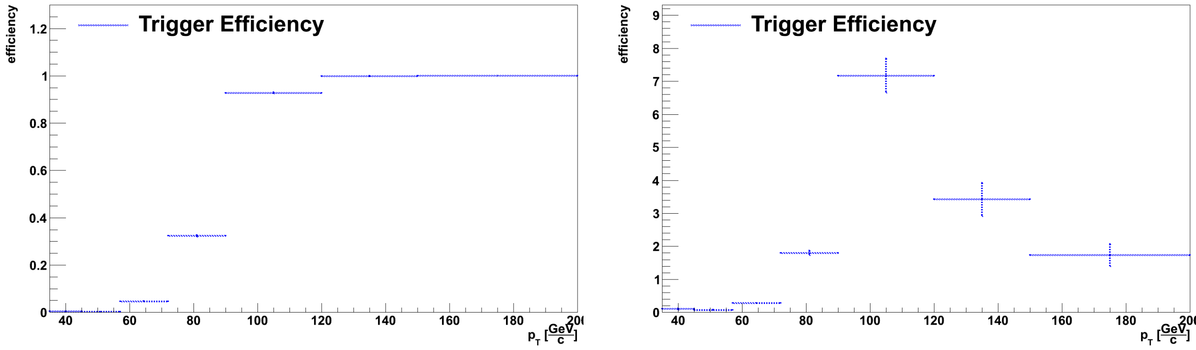


Figure 5.14: The efficiency as a function of leading jet p_{T} before (left) and after (right) the trigger correction for the JetMETTau_2010A dataset using the HLT_Jet50U trigger.

The efficiency as a function of the leading jet ϕ before and after the trigger correction using the HLT_Jet50U trigger for the JetMETTau_2010A dataset is shown in the figure 5.15. The distribution is almost flat before, with an average of 16%, and after the correction, with an average of 100%. Therefore, this variable displays no visible bias from the correction.

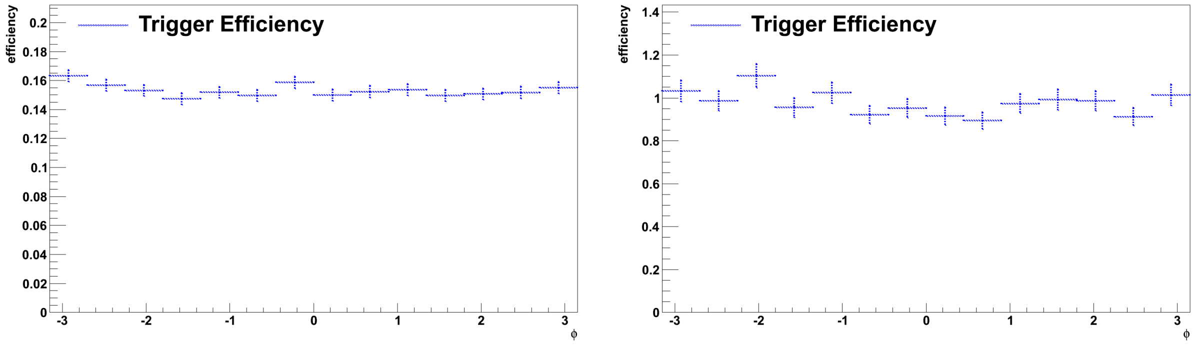


Figure 5.15: The efficiency as a function of leading jet ϕ before (left) and after (right) the trigger correction for the JetMETTau_2010A dataset using the HLT_Jet50U trigger.

The efficiency as a function of $\Delta\eta$ before and after the trigger correction using the HLT_Jet50U trigger for the JetMETTau_2010A dataset is presented in the figure 5.16. Before the correction the function shows a difference between low and high $\Delta\eta$ regions. The average is around 16%. After the correction the distribution is flatter and the difference is on the order of the statistical uncertainty with an average efficiency of $\simeq$100%. No bias seems to be introduced by the correction.

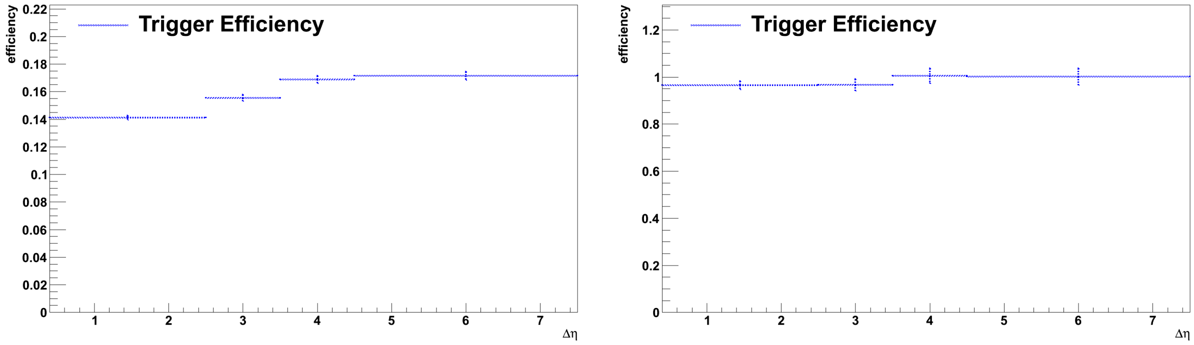


Figure 5.16: The efficiency as a function of $\Delta\eta$ before (left) and after (right) the trigger correction for the JetMETTau_2010A dataset using the HLT_Jet50U trigger.

The efficiency as a function of $\Delta\phi$ before and after the trigger correction using the HLT_Jet50U trigger path for JetMETTau_2010A dataset is presented in the figure 5.17. Before the correction the efficiency looks like a 1st degree polynomial decreasing from 0.32 for the first bin to 0.15 in the last. The average is around 0.16. After the correction, despite the fact that the average efficiency is 1 and it is 1 for the last 4 bins, it varied between 1.5 in the 3rd bin to 2.4 in the 1st. Thus, this correction introdused a huge bias in our observable.

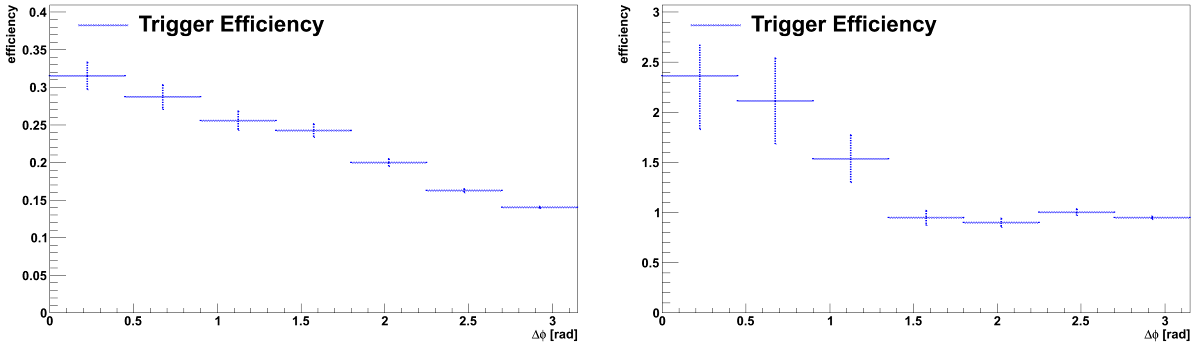


Figure 5.17: $\Delta\phi$ efficiency before (left) and after (right) the trigger correction for the JetMETTau_2010A dataset using HLT_Jet50U trigger.

The same trends are found for the JetMET_2010A dataset. It was found that the efficiency is η–dependent and has a not understood structure making it impossible to achieve better corrections based on the pseudo–rapidity. There was not enough statistics to do a 2D–correction on p_{T} and η. Attempts to do such correction in two steps, first p_{T} and then η, failed because of the bias of the first variable causes on the second. Further attemps to use the leading central or leading forward jet to correct for the efficiency yield similar results. Correcting the observables directly would introduce an uncontrollable bias and thus this option was not tested.

Therefore, because of all these unphysical efficiencies, the combination using only the efficient regions described in section 5.1.2 was adopted instead of the one described in this section.

5.2 Resolution

To choose the bin width of the distributions to be measured, the resolution must be studied. Unless one had access to a test beam, this kind of test is usally carried on MC samples where both generator and detector information is present. Thus, the resolution can be computed from MC by calculating the difference between the true and the reconstructed value of a given quantity. The most common quantity is the absolute resolution:

$$R_{absolute} = Observable\ _{true} - Observable\ _{measured}, \tag{5.7}$$

where the $Observable_{true}$ comes from generator information while $Observable_{measured}$ corresponds to detector level. When the observable spans several orders of magnitude, the absolute value of the resolution is not a good reference as the difference is likely to scale with the value. In that case, and if the value of the observable never approaches 0, the best option is to use the relative resolution:

$$R_{relative} = \frac{Observable\ _{true} - Observable\ _{measured}}{Observable\ _{true}}. \tag{5.8}$$

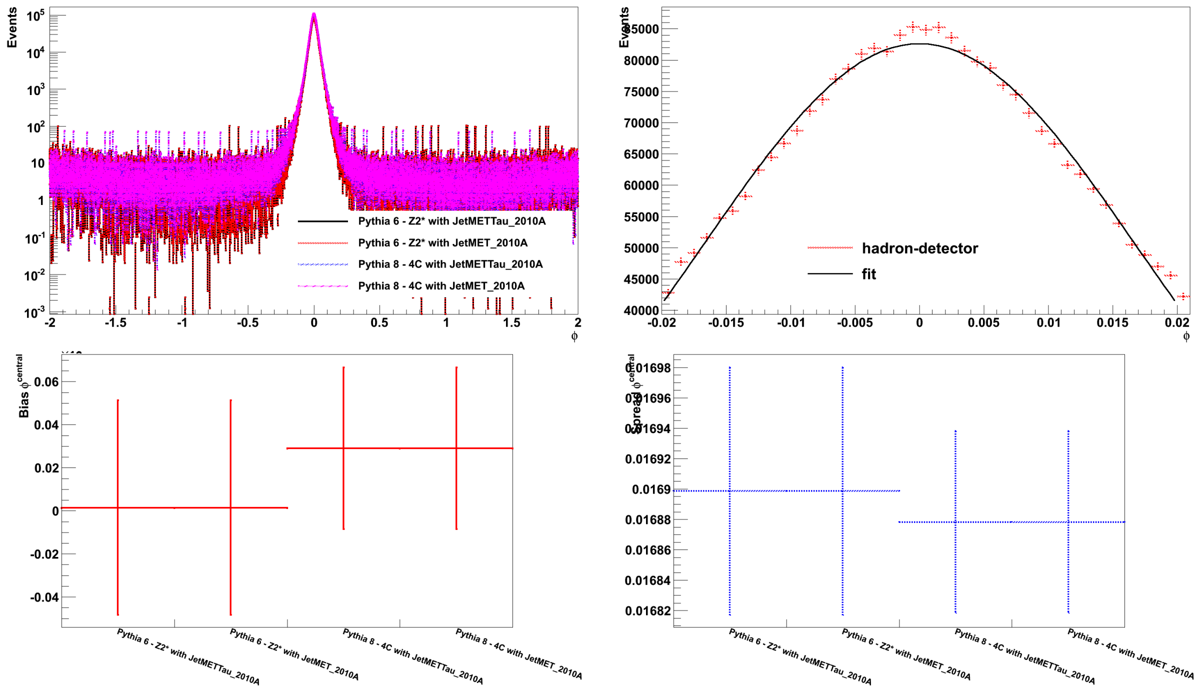


Figure 5.18: Example resolution study plots for the leading central jet ϕ. Overall view (top-left), peak fitted with a Gaussian function (top-right), fitted peak position (please note the y–scale should be multiplied 10^{-3}) (bottom-left) and resolution, which is the fitted peak width (bottom-right).

In this thesis, the following method is used to estimate the resolution:

1. For each variable, the log–scale plot of the resolution, obtained from figures 5.7 or 5.8, is checked for bumps and strange features in the tails of the distribution, to make sure there is only a single peak and that the resolution was a Gaussian–shape (see figure 5.18 top-left).

2. A Gaussian function is fitted to the peak, using points up to two Gaussian–widths away from the peak (see figure 5.18 top-right).

3. The errors and convergence of the fit are checked.
4. The bias of the observable is obtained from the fitted peak position (see figure 5.18 bottom-left).
5. The resolution is obtained from the fitted peak width (see figure 5.18 bottom-right).
6. The information is condensed in one plot (see figure 5.19-left), where the positions at the y-axis are the shift and the errors are the resolution.

Absolute Resolution

In this sub–section, the absolute resolution of que quantities used in this thesis is presented.

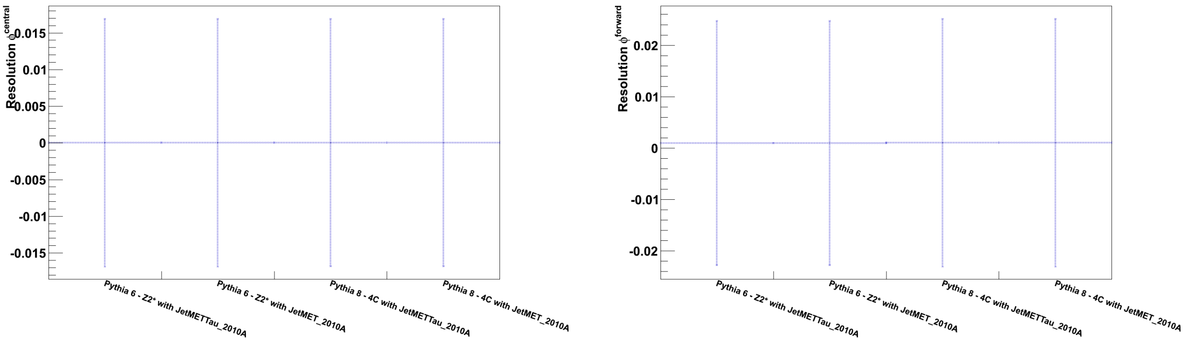


Figure 5.19: Leading jet ϕ absolute resolution and bias. Leading central jets (left) and leading forward jets (right).

In figure 5.19 the resolution as a function of the azimuthal angle, ϕ, for the leading central and leading forward jets are presented. In both cases, the bias is negligible (less than 10^{-3}). The width is 1.7×10^{-2} for central jets and 2.4×10^{-2} for forward jets. The results from both MCs simulations and from different pile-up scenarios are similar and consistent.

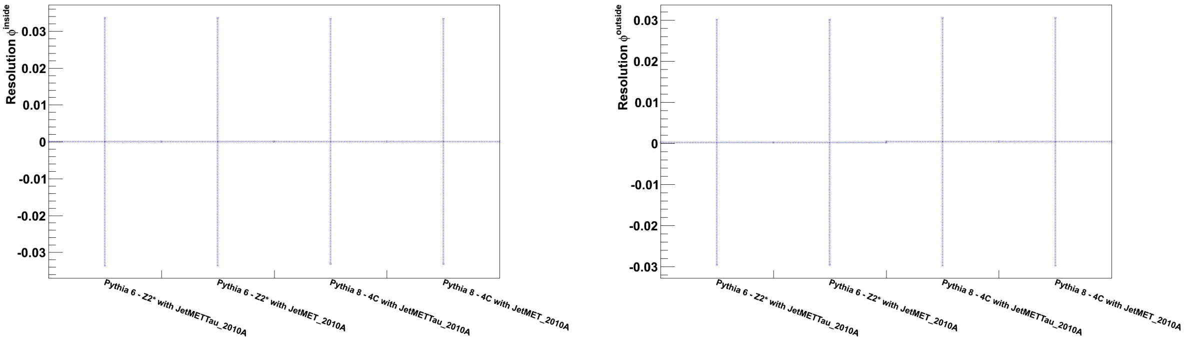


Figure 5.20: Leading jet ϕ absolute resolution and bias. Leading inside jets (left) and leading outside jets (right).

The resolution as a function of azimuthal angle, ϕ, for the leading inside and leading outside jets is presented in the figure 5.20. The shift is smaller than 10^{-3}. The width is 3.3×10^{-2} for inside jets and 3.0×10^{-2} for outside jets. There is an increase in the width when compared with the leading central and leading forward jets, which could be explained by the discontinuity between the Hadronic Barrel and Hadronic Forward calorimeters, present in the pseudo–rapidities from 3.3 to 3.7. Such discontinuity is expected to affect an inside jet more than an outside jet because it is more likely for an inside jet to be in this gap. The expected difference between the inside and outside jets is observed, thus supporting this assumption.

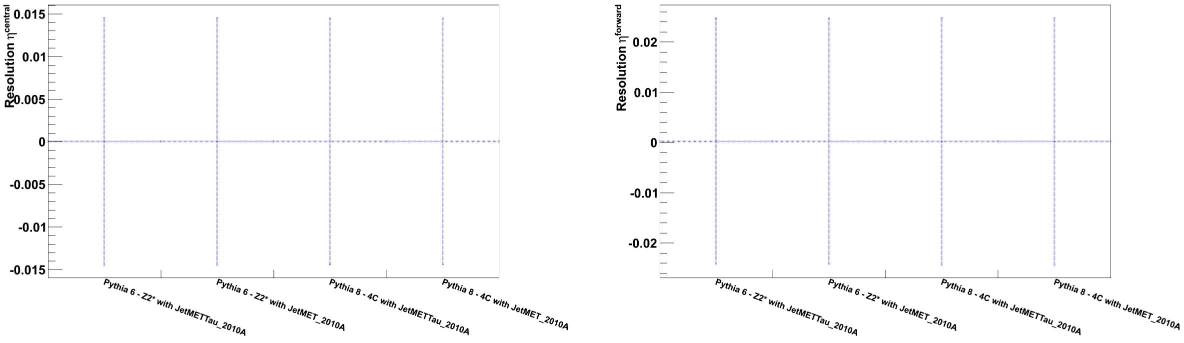


Figure 5.21: Leading jet η absolute resolution and bias. Leading central jets (left) and leading forward jets (right).

In the figure 5.21 the resolution as a function of the pseudo–rapidity, η, for the leading central and leading forward jets is shown. In both cases, the shift is negligible ($< 10^{-3}$). The width is 1.4×10^{-2} for central jets and 2.5×10^{-2} for forward jets. The conclusions are similar as for the ϕ distributions.

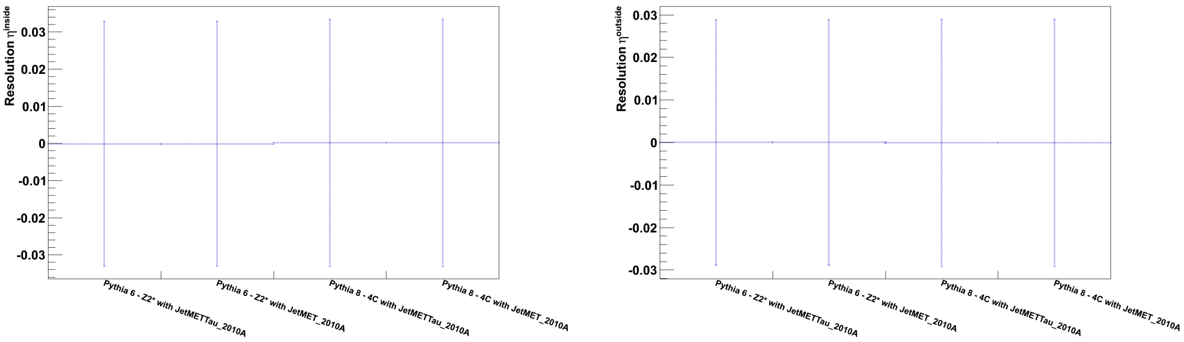


Figure 5.22: Leading jet η absolute resolution and bias. Leading inside jets (left) and leading outside jets (right).

The resolution as a function of the pseudo–rapidity, η, for the leading central and leading forward jets is presented in figure 5.22. The shift is smaller than 10^{-3}, the width is 3.3×10^{-2} for inside jets and 2.9×10^{-2} for outside jets. The same conclusions as for ϕ can be drawn.

Table 5.3: Absolute Resolution Summary

Observable	Shift	Resolution
Leading Central Jet ϕ	$< 10^{-3}$	0.017
Leading Forward Jet ϕ	$< 10^{-3}$	0.024
Leading Inside Jet ϕ	$< 10^{-3}$	0.033
Leading Outside Jet ϕ	$< 10^{-3}$	0.030
Leading Central Jet η	$< 10^{-3}$	0.014
Leading Forward Jet η	$< 10^{-3}$	0.025
Leading Inside Jet η	$< 10^{-3}$	0.033
Leading Outside Jet η	$< 10^{-3}$	0.029

The results for the relative resolution are summarized in table 5.3. All shifts are negligible and all widths are smaller than 0.04.

Relative Resolution

In this sub–section, the relative resolution of que quantities used in this thesis is presented.

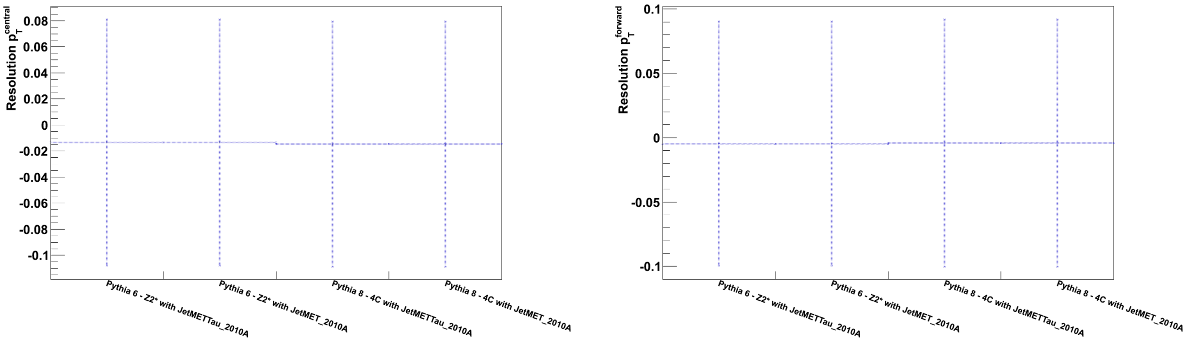


Figure 5.23: Leading jet p_T relative resolution and bias. Leading central jets (left) and leading forward jets (right).

In figure 5.23 the resolution as a function of the transverse momentum. p_T, for the leading central and the leading forward jet is presented. The shift is -1.4% for the leading central and -0.5% for the leading forward jet. The width is 9.4% for the leading central and 9.6% for the leading forward jet.

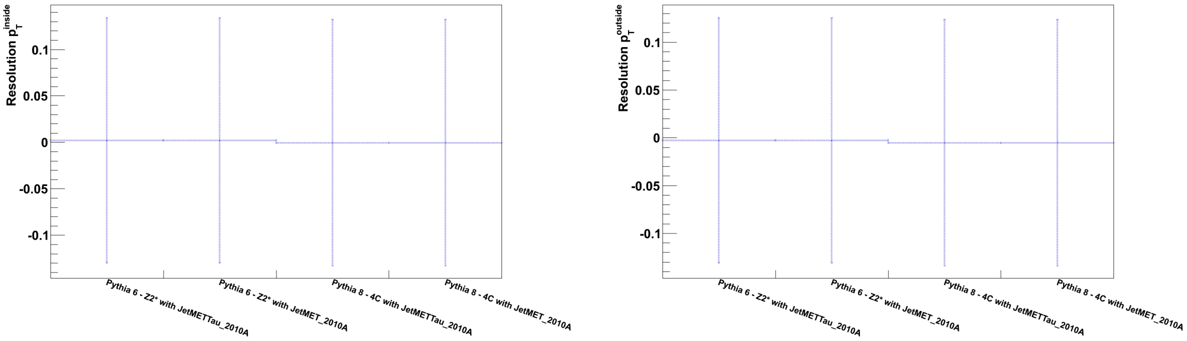


Figure 5.24: Leading jet p_T relative resolution and bias. Leading inside jets (left) and leading outside jets (right).

In figure 5.24 the resolution as a function of the transverse momentum, p_T, for the leading inside and leading outside the gap jet is shown. The same effect seen in the η and ϕ width is present here. The shift for the inside jet is negligible and for the outside jet is less than 0.5%. The width is around 13% and is consistent with the predictions of different MCs.

In table 5.4 the relative resolution summary is presented. Except for the leading central jet, all shifts are negligible. The resolution width is 9% for the central and forward jets, and 13% for the inside and outside jets.

Observables Resolution

In this sub–section, the resolution for the observables measured in this thesis is presented.

Table 5.4: Relative Resolution Summary

Observable	Shift	Resolution
Leading Central Jet p_{T}	-1.42%	9.43%
Leading Forward Jet p_{T}	-0.45%	9.56%
Leading Inside–Jet p_{T}	0.08%	13.23%
Leading Outside–Jet p_{T}	-0.41%	12.85%

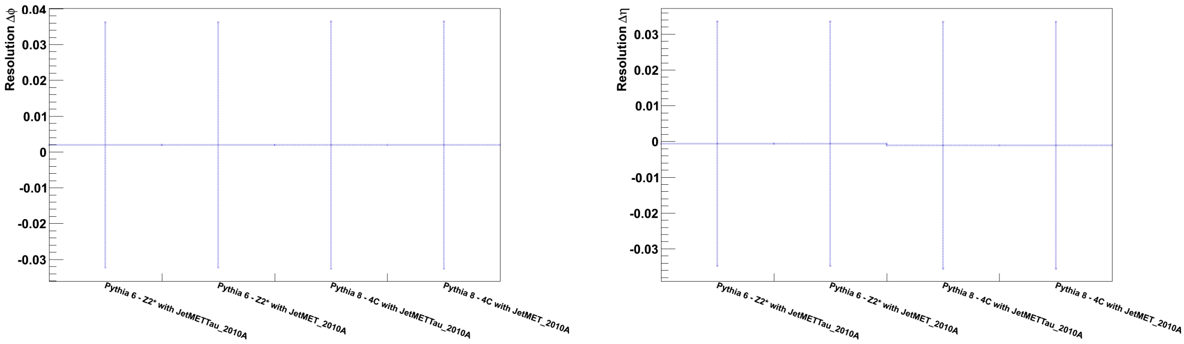


Figure 5.25: Resolution as function of $\Delta\phi$ (left) and $\Delta\eta$ (right) between the leading central and the forward jets in the inclusive scenario.

The resolution as a function of $\Delta\phi$ and $\Delta\eta$ between the leading central and the forward jets are presented in figure 5.25. The shift for $\Delta\phi$ is 1.9×10^{-3} and the width is 3.4×10^{-2}. The shift for $\Delta\eta$ is 0.9×10^{-3} and the width is 3.4×10^{-2}. As expected, the shift and the resolution are larger than for the central or forward jets alone, and consistent between different MC predictions. The resolution for $\Delta\phi$ in the inclusive scenario in different $\Delta\eta$ slices is presented in the figure D.1 in the appendix D. The bias for all slices is around 2×10^{-3}. The resolution width lies between 3.3 and 3.6 $\times 10^{-2}$. No fundamental differences are observed between the slices with respect to the main distribution.

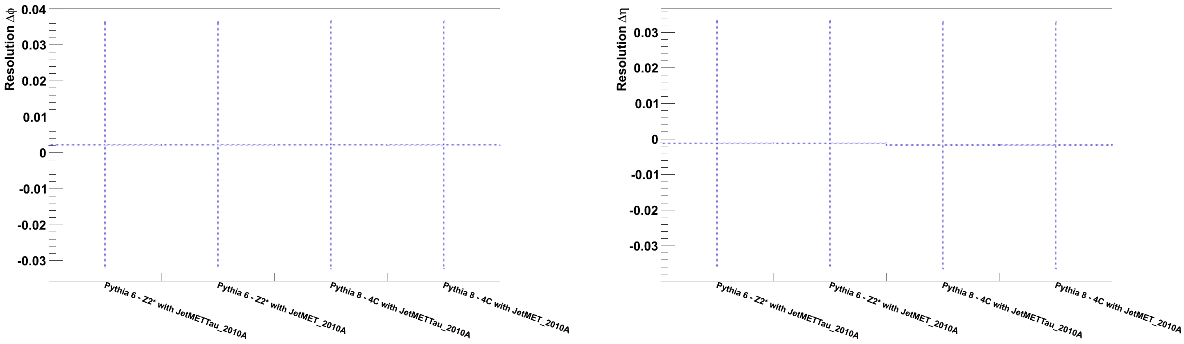


Figure 5.26: Resolution as function of $\Delta\phi$ (left) and $\Delta\eta$ (right) between the leading central and the forward jets in the inside–jet veto scenario.

The resolution as a function of $\Delta\phi$ and $\Delta\eta$ between the leading central and the forward jets for the inside–jet veto scenario are shown in figure 5.26. The shift is a bit more pronounced than in the inclusive scenario: 2.2×10^{-3} for $\Delta\phi$ and -1.5×10^{-3} for $\Delta\eta$. The values for the width are nearly the same as for the main scenario. The resolution for $\Delta\phi$ in the inside–jet veto scenario in different $\Delta\eta$ slices is shown in figure D.2 in appendix D. The shift values are a bit higher, but still around 2×10^{-3} and the resolution width is between 3.2 and 3.5 $\times 10^{-2}$. No significant differences from the main scenario were found.

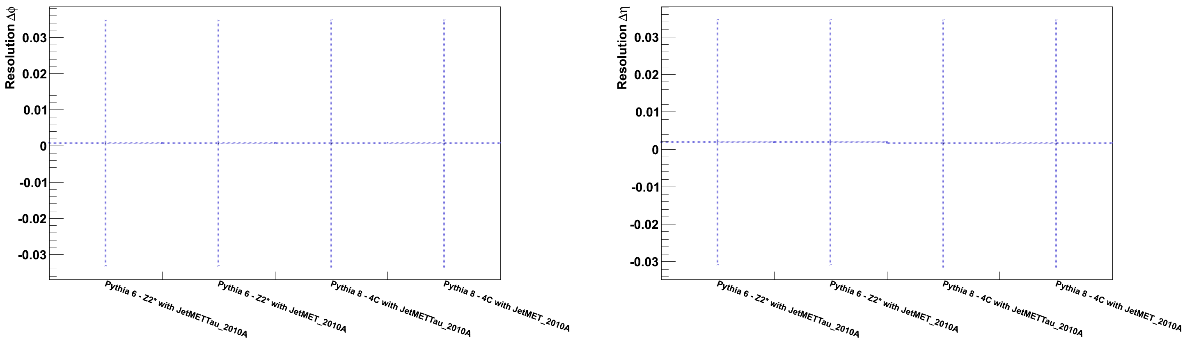


Figure 5.27: Resolution as function of $\Delta\phi$ (left) and $\Delta\eta$ (right) between the leading central and the forward jets in the inside–jet tag scenario.

The resolution as a function of $\Delta\phi$ and $\Delta\eta$ between the leading central and the forward jets for the inside–jet veto scenario are presented in figure 5.27. The shift for $\Delta\phi$ is smaller than 10^{-3} and 1.7×10^{-3} for $\Delta\eta$, which is the opposite sign when compared to the inclusive scenario. The values of the width are in the same range as for the other scenarios. The figure D.3 in the appendix D presents the resolutions for $\Delta\phi$ in $\Delta\eta$ slices for the inside–jet tag scenario. The shift is smaller than in the main scenario, around 1×10^{-3}, and the width has similar values as for the other scenarios.

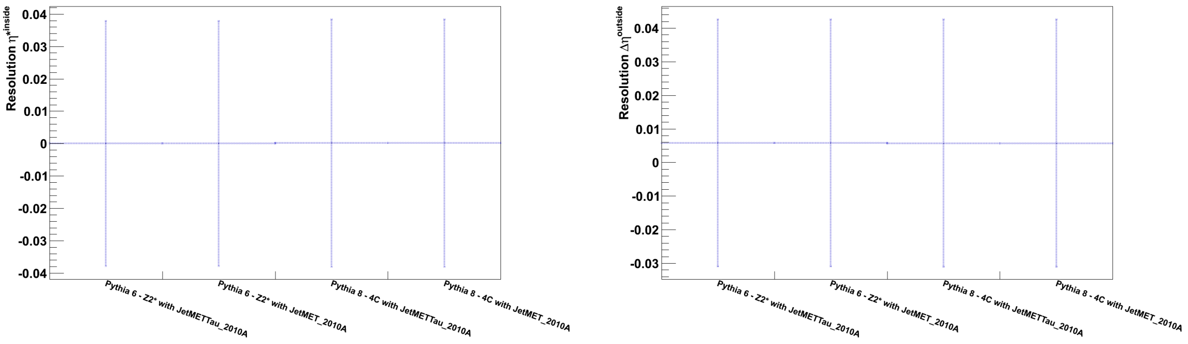


Figure 5.28: Resolution as function of $\eta*$ for the leading inside–jet (left) and $\Delta\eta^{out}$ for the leading outside–jet (left).

The resolution as a function of η^* inside the inter–leading jet region and $\Delta\eta^{out}$figure is shown in figure 5.28. The shift is much smaller than 10^{-3} and the resolution is 3.8×10^{-2} for η^*. $\Delta\eta^{out}$ has a shift of 5.8×10^{-3} and a width of 3.7×10^{-2}.

A resolution summary of the measured observables is presented in table 5.5. All widths are around 3×10^{-2}, while the shifts are no bigger than 6×10^{-3}. The p_{T} observables are discussed in the previous section. This was expected, as the observables are based in η and ϕ, whose resolutions equally excellent due to the finely grained detectors and good reconstruction.

Taking into account resolution results and effective statistics, the bins used in this analysis are defined as the following:

- Dijet $\Delta\phi$ bins: $\{0.00, 0.45, 0.90, 1.35, 1.80, 2.25, 2.70, 3.15\}$ (inclusive, inside–jet tag and inside–jet veto scenarios)
- Dijet $\Delta\eta$ bins: $\{0.4, 2.5, 3.5, 4.5, 7.5\}$ (inclusive, inside–jet tag and inside–jet veto scenarios)

Table 5.5: Observables Resolution Summary

Observable	Shift	Resolution
$\Delta\phi$ (main scenario)	1.91×10^{-3}	3.44×10^{-2}
$\Delta\eta$ (main scenario)	-0.87×10^{-3}	3.44×10^{-2}
$\Delta\phi$ (main scenario) - $0.4 > \Delta\eta >= 2.5$	2.30×10^{-3}	3.55×10^{-2}
$\Delta\phi$ (main scenario) - $2.5 > \Delta\eta >= 3.5$	1.63×10^{-3}	3.31×10^{-2}
$\Delta\phi$ (main scenario) - $3.5 > \Delta\eta >= 4.5$	1.56×10^{-3}	3.27×10^{-2}
$\Delta\phi$ (main scenario) - $4.5 > \Delta\eta >= 7.5$	1.46×10^{-3}	3.45×10^{-2}
$\Delta\phi$ (inside–jet veto)	2.20×10^{-3}	3.43×10^{-2}
$\Delta\eta$ (inside–jet veto)	-1.53×10^{-3}	3.45×10^{-2}
$\Delta\phi$ (inside–jet veto) - $0.4 > \Delta\eta >= 2.5$	2.45×10^{-3}	3.53×10^{-2}
$\Delta\phi$ (inside–jet veto) - $2.5 > \Delta\eta >= 3.5$	1.86×10^{-3}	3.29×10^{-2}
$\Delta\phi$ (inside–jet veto) - $3.5 > \Delta\eta >= 4.5$	1.99×10^{-3}	3.28×10^{-2}
$\Delta\phi$ (inside–jet veto) - $4.5 > \Delta\eta >= 7.5$	2.06×10^{-3}	3.40×10^{-2}
$\Delta\phi$ (inside–jet tag)	0.70×10^{-3}	3.41×10^{-2}
$\Delta\eta$ (inside–jet tag)	1.74×10^{-3}	3.29×10^{-2}
$\Delta\phi$ (inside–jet tag) - $0.4 > \Delta\eta >= 2.5$	1.00×10^{-3}	3.59×10^{-2}
$\Delta\phi$ (inside–jet tag) - $2.5 > \Delta\eta >= 3.5$	0.49×10^{-3}	3.31×10^{-2}
$\Delta\phi$ (inside–jet tag) - $3.5 > \Delta\eta >= 4.5$	0.74×10^{-3}	3.29×10^{-2}
$\Delta\phi$ (inside–jet tag) - $4.5 > \Delta\eta >= 7.5$	0.61×10^{-3}	3.40×10^{-2}
Leading inside–jet p_{T}	0.08%	13.23%
Leading inside–jet $\eta*$	0.08×10^{-3}	3.80×10^{-2}
Leading outside–jet p_{T}	-0.41%	12.85%
$\Delta\eta^{out}$	5.77×10^{-3}	3.81×10^{-2}

- Leading p_{T} inside and outside bins: $\{20, 27, 35, 45, 57, 72, 90, 120, 150, 200\}$ (inside–jet tag and outside–jet tag scenarios)
- Leading η^* inside bins: $\{-3.6, -2.5, -2.0, -1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.6\}$ (inside–jet tag scenario)
- Leading $\Delta\eta^{out}$ jet bins: $\{0.0, 1.0, 2.0, 3.0, 4.0, 5.0, 7.5\}$ (outside–jet tag scenario)

The bins are usually at least ten times larger than the resolution to prevent any sizable bin migration effects.

5.3 Migration Matrix

Although unfoling techniques are used in this thesis, it is necessary to compute the bin–by–bin corrections, because they can be used to compare with the unfolding result. The bin–by–bin correction factors are computed from MC detector level to MC stable particle level:

$$C_i = \frac{\sigma_{i\ stable\ particle\ level}}{\sigma_{i\ detector\ level}} \tag{5.9}$$

where C_i is the correction factor for bin i, $\sigma_{i\ stable\ particle\ level}$ is the MC hadron level cross–section measured in bin i and $\sigma_{i\ detector\ level}$ is the MC detector level cross–section measured in bin i.

To define the quality of the correction for the detector effects, matching the jets between hadron and detector level on MC is necessary to estimate the migrations as well the contributions from background and lack of efficiency. In this analysis, the rapidity difference between the hadron and detector level, $d\eta$, needs to be computed:

$$d\eta = (\eta_{detector\ level} - \eta_{hadron\ level}) \tag{5.10}$$

The matching cannot be done using the observable, thus ϕ was not included. The $d\eta$ difference should be less than a certain constant, R, which defines how close the two jets need to be in the θ–plane to be matched:

$$d\eta \leq R \tag{5.11}$$

In a successful matching, this condition must be fulfilled for both leading central and leading forward jets.

Two small cross-checks have been carried out in smaller samples in order to know the influence of the matching radius, R, presented in table 5.6, and of the MC samples, shown in table 5.7. The results for the matching radius show that the matching is approximately independent of the chosen radius and of the MC generator, since all results are very close and the difference in the number of events with matched jets is less than 1%. The matching efficiency using different samples lies between 80 and 82%, suggesting that the generator has no sizable influence.

Table 5.6: Matched events as function of the matching radius using the PYTHIA 8 -Tune 1 sample.

Radius	Events with matched central and forward jets
0.5	167994
1.0	168617
1.5	169046
2.0	169494

Table 5.7: Matching efficiency for different MC samples.

Sample	Total on detector level	Total matched	Efficiency
PYTHIA 8 - Tune 1	206322	168617	82%
PYTHIA 6 - Tune Z2	604957	497516	82%
HERWIG 6	323910	259647	80%

In figure 5.29 a typical $\Delta\phi$ migration matrix of the matched events is presented. Most of the elements are in the matrix's diagonal, which mean that the migrations are small. This result is observed also for the other generators and datasets. The results for the other scenarios and $\Delta\eta$ slices are very similar.

The migration matrix after matching as a function of leading inside–jet p_{T} is presented in figure 5.30. Most of the events are in the matrix's diagonal. A similar result was obtained for the leading p_{T} of the outside jet, as well across different MCs and different datasets. The fact that most of the off–diagonal bins have little or no entries shows that very rarely a high p_{T} jet is measured as a low p_{T} and vice–versa. Therefore, the p_{T} migrations are small.

The migration matrix after matching as a function of leading inside–jet η^* is shown in figure 5.31. As expected, due to the fact that the matching is carried using η, only the bins next to the matrix's diagonal are filled as this forces the migrations to be small. The result is not much different without matching.

The migration matrix after matching as a function of $\Delta\eta^{out}$ is shown in figure 5.32. The result and conclusions are similar to η^*.

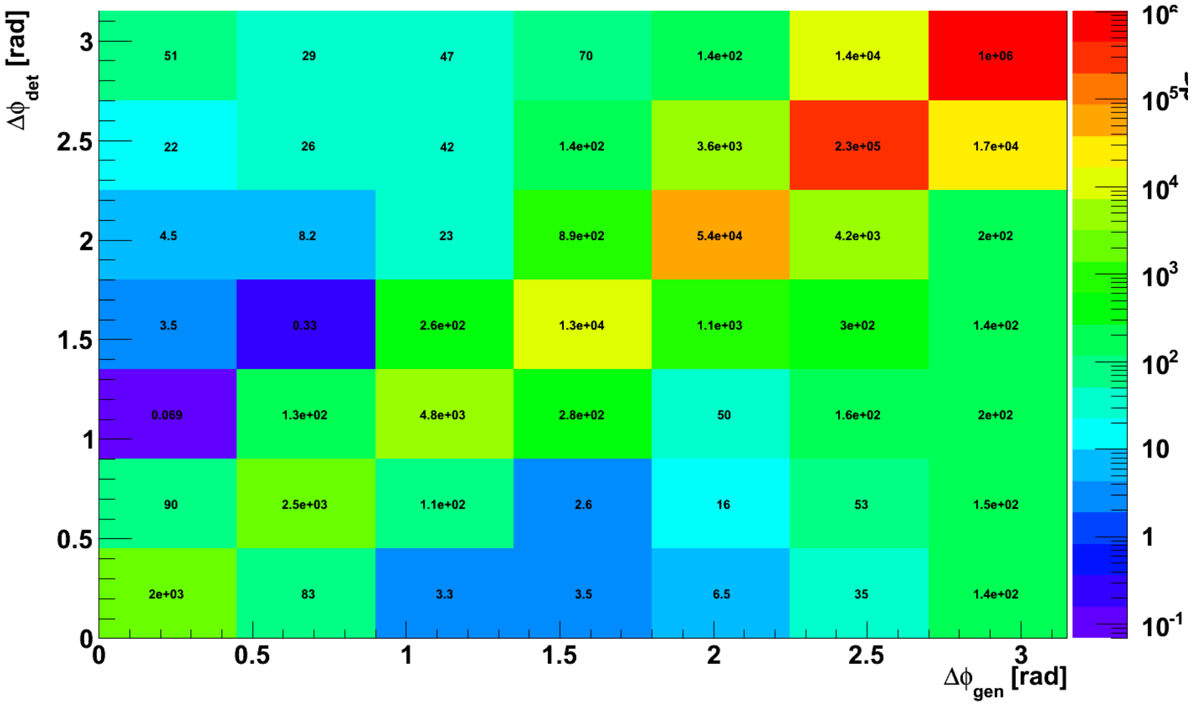


Figure 5.29: Migration matrix for matched jets as a function of $\Delta\phi$ using Pythia 6 - Z2* Tune with pile–up re–weight factors from JetMET_2010A dataset for the inclusive scenario.

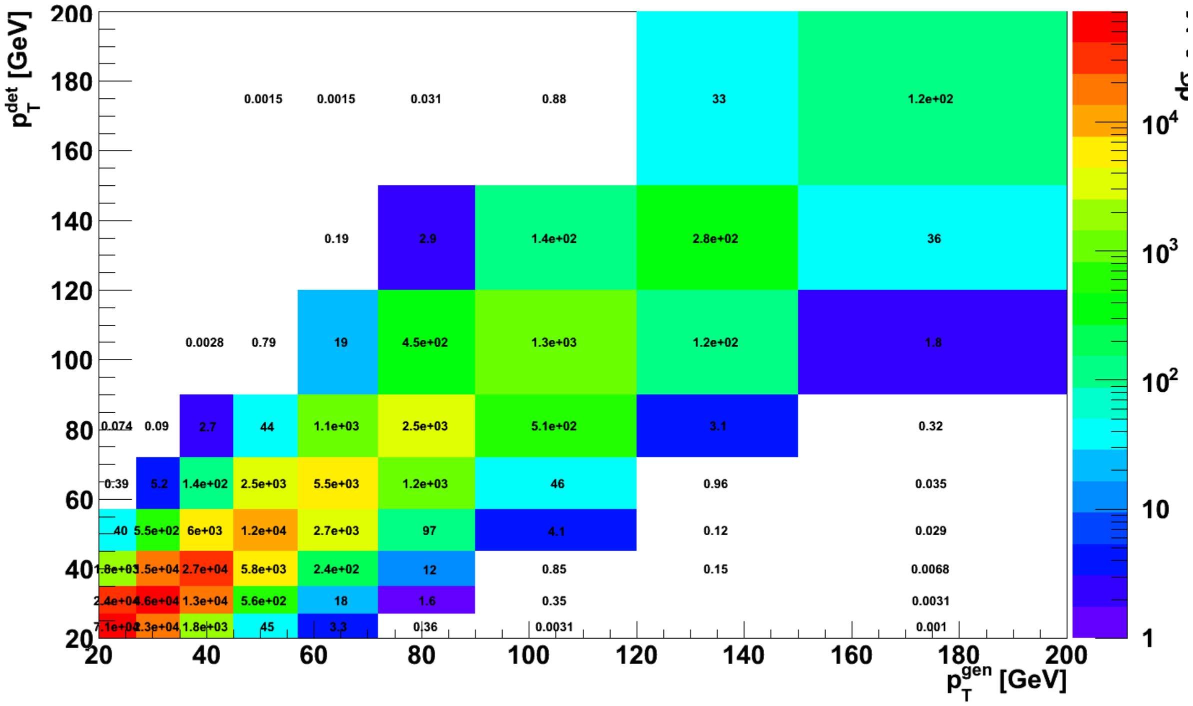


Figure 5.30: Migration matrix for matched jets as a function of leading inside–jet p_T using Pythia 6 - Z2* Tune with pile–up re–weight factors from JetMET_2010A dateset.

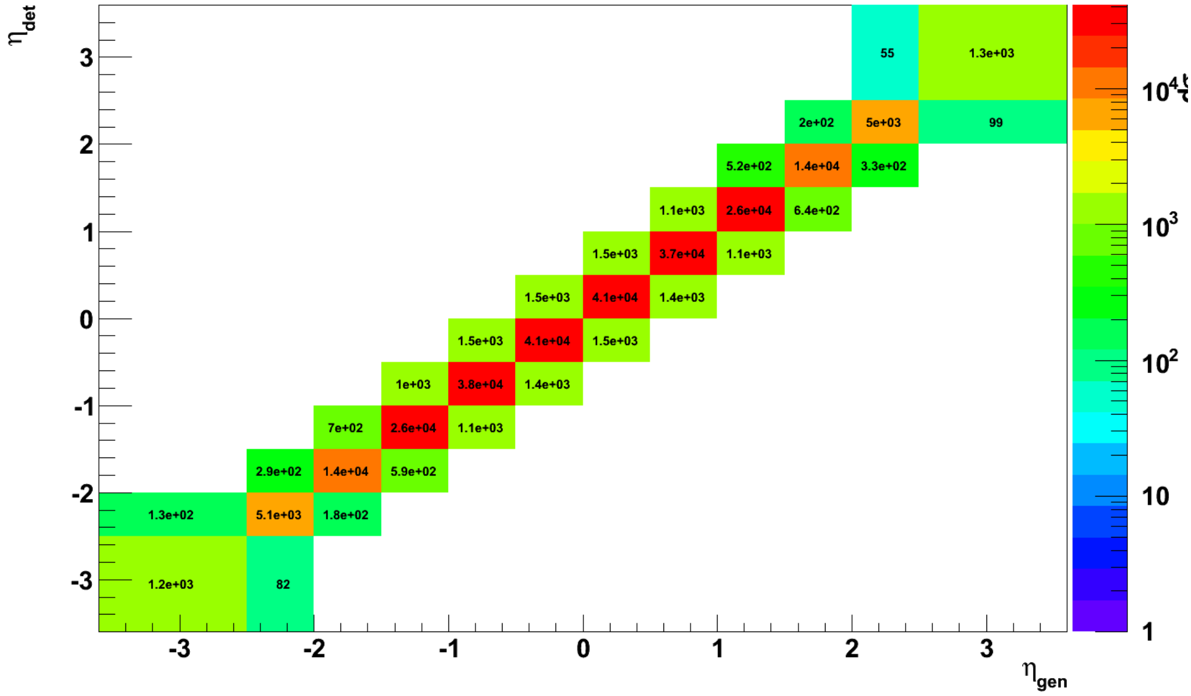


Figure 5.31: Migration matrix for matched jets as a function of leading inside–jet η^* using Pythia 6 - Z2* Tune with pile–up re–weight factors from JetMET_2010A dateset.

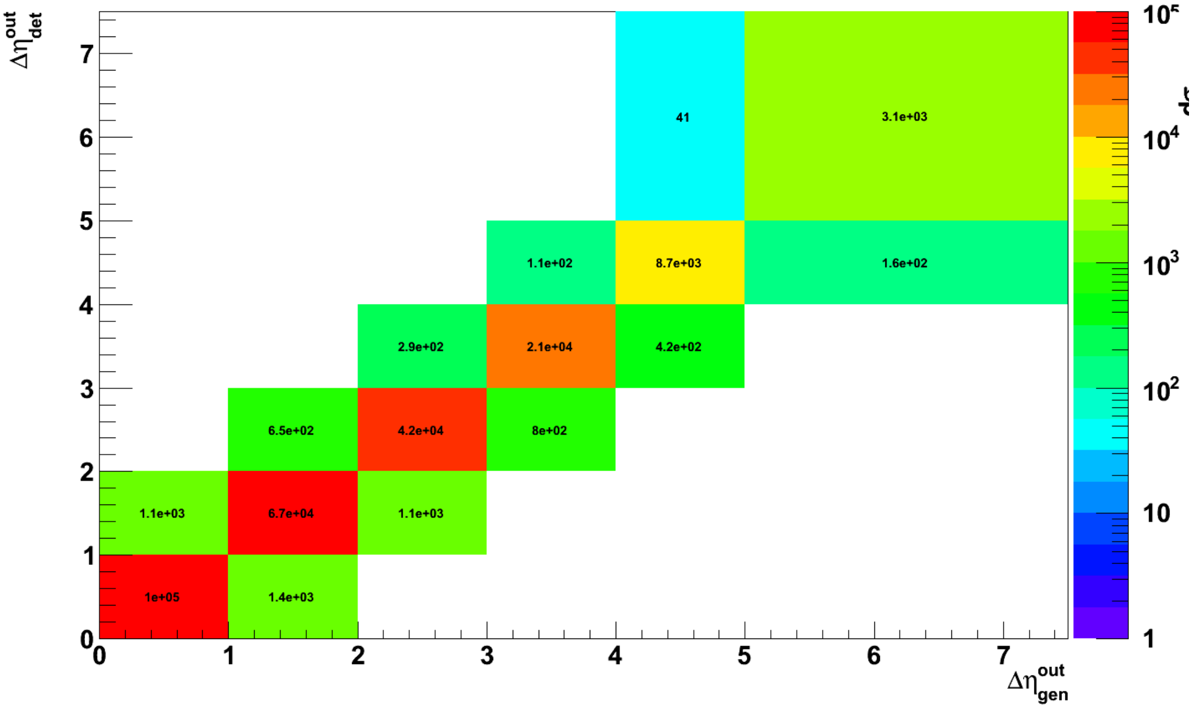


Figure 5.32: Migration matrix for matched jets as a function of $\Delta\eta$ using Pythia 6 - Z2* Tune with pile–up re–weight factors from JetMET_2010A dateset.

The results obtained for unmatched migration matrices are similar, except in the η^* and $\Delta\eta^{out}$ cases, where there is also a small amount of off–diagonal events. This means that the reconstruction is good and the migrations are small, even without the matching requirement.

5.4 Acceptance and Background, Purity, Stability

The quality of the correction factors for the detector effects can be determined by the quality of the matching and the amount of migrations. The quality of the matching procedure can be assessed by computing the acceptance (equation 5.12) and background (equation 5.13):

$$A_i^{MC} = \frac{N^{MC}_{matched\ jets}(E^{MC}_{had} \in bin\ i)}{N^{MC}_{all\ jets}(E^{MC}_{had} \in bin\ i)} \tag{5.12}$$

$$B_i^{MC} = 1 - \frac{N^{MC}_{matched\ jets}(E^{MC}_{det} \in bin\ i)}{N^{MC}_{all\ jets}(E^{MC}_{det} \in bin\ i)} \tag{5.13}$$

The migrations between bins can be quantified in terms of purity (equation 5.14) and stability (equation 5.15)

$$P_i^{MC} = \frac{N^{MC}_{matched\ jets}(E^{MC}_{det} \in bin\ i) \wedge (E^{MC}_{had} \in bin\ i)}{N^{MC}_{matched\ jets}(E^{MC}_{det} \in bin\ i)} \tag{5.14}$$

$$S_i^{MC} = \frac{N^{MC}_{matched\ jets}(E^{MC}_{had} \in bin\ i) \wedge (E^{MC}_{det} \in bin\ i)}{N^{MC}_{matched\ jets}(E^{MC}_{had} \in bin\ i)} \tag{5.15}$$

The bin–by–bin correction factor can be derived from these PSBA variables (**P**urity, **S**tability, **B**ackground and **A**cceptance) by:

$$C_i^{MC} = \frac{N^{MC}_{all\ jets}(E^{MC}_{had} \in bin\ i)}{N^{MC}_{all\ jets}(E^{MC}_{det} \in bin\ i)} = \frac{(1 - B_i^{MC})}{A_i^{MC}} \cdot \frac{P_i^{MC}}{S_i^{MC}} \tag{5.16}$$

The PSBA variables are a valuable tool to determine the quality of the correction factors. A consistency test for these factors can be obtained when comparing this correction factor with the one from the direct division of hadron level by detector level, which is represented in the middle member of equation 5.16.

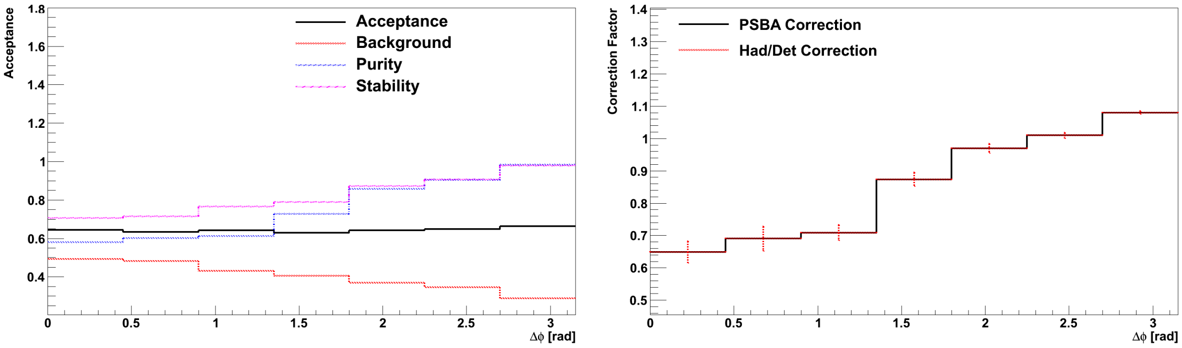


Figure 5.33: Acceptance, background, purity and stability (left) and closure test (right) for the bin–by–bin correction performed on the JetMETTau_2010A datasample corrected with PYTHIA 6 - Z2* Tune in the inclusive scenario.

In figure 5.33 the PSBA variables for $\Delta\phi$ in the main scenario are shown on the left side, while the closure test is presented on the right side. The acceptance is around 0.65 and constant over the whole range. The background is 0.5 for the lower $\Delta\phi$ region and decreases to 0.3 in the highest $\Delta\phi$ bin. The purity and stability start at 0.6 and grow to almost 1.0. There is no difference between the correction factors derived from direct division and PSBA variables. The variation of the correction factor can be explained by the fraction of the fakes present in the bin, which is greater in the lower $\Delta\phi$ region. The correction factor grows with $\Delta\phi$, being almost 1 in the higher $\Delta\phi$ region. In figures E.1 and E.2 in appendix E the PSBA variables for the $\Delta\eta$ slices and respective closure test are shown. For these distribution, it was found that the acceptance is a bit lower, while the background is bigger. Both purity and stability show the same structure and are above 0.8 for almost all bins.

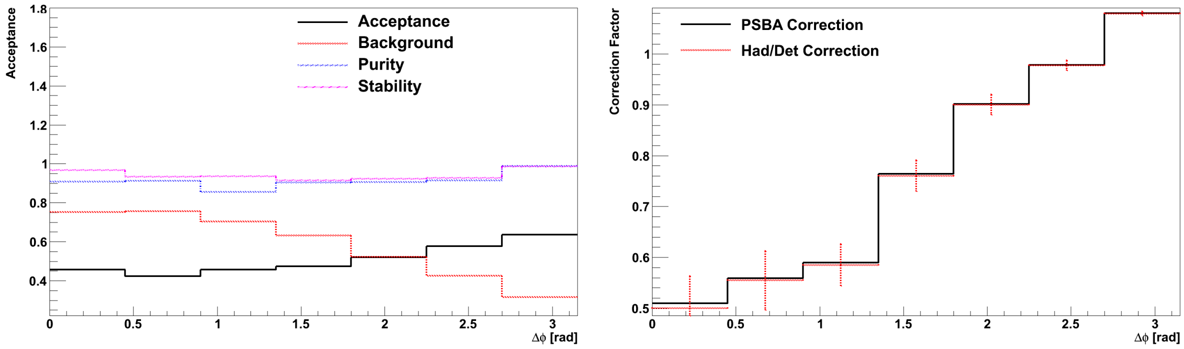


Figure 5.34: Acceptance, background, purity and stability (left) and closure test (right) for the bin–by–bin correction performed on the JetMETTau_2010A datasample corrected with Pythia 6 - Z2* Tune in the inside–jet veto scenario.

In figure 5.34 the PSBA variables for $\Delta\phi$ in the inside–jet veto scenario are presented on the left side, while the closure test is shown on the right side. The purity and stability are over 0.9 for most of the bins and don't exhibit any particular structure. The background starts at 0.75 and decreases to 0.3, while the acceptance grows from 0.45 to 0.65. The correction factors are consistent except for a little difference in the first bin. In figures E.3 and E.4 in appendix E the PSBA variables for the $\Delta\eta$ slices and respective closure test are presented. Except for the lower values of acceptance and for the higher values of background, all the PSBA variables have similar behaviour as for the integrated distribution.

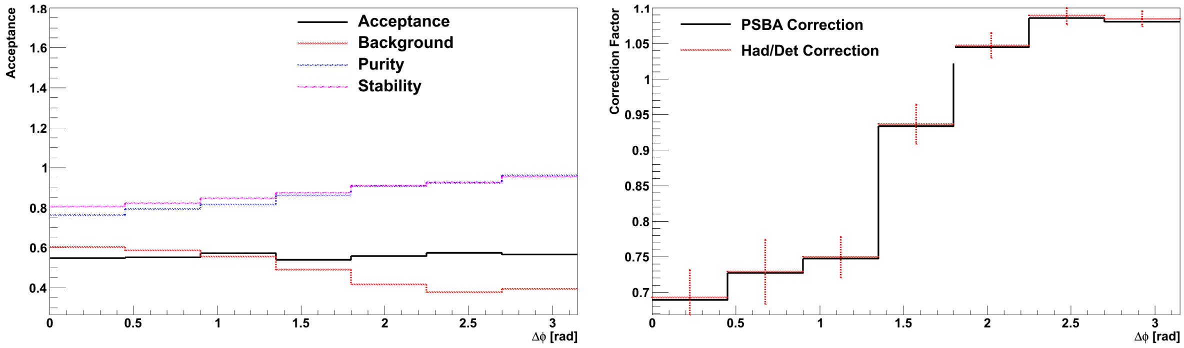


Figure 5.35: Acceptance, background, purity and stability (left) and closure test (right) for the bin–by–bin correction performed on the JetMETTau_2010A datasample corrected with Pythia 6 - Z2* Tune in the inside–jet tag scenario.

The PSBA variables for $\Delta\phi$ in the inside–jet tag scenario are presented on the left side of figure 5.35 and the closure test is shown on the right side. The acceptance is roughly constant over the whole range, with a value around 0.55. The background starts at 0.6 for low $\Delta\phi$ and decreases to 0.4 in the higher region. The purity and stability grow from 0.75 and 0.8 respectively to close to 1.0. There is a good agreement between both correction factors. The PSBA variables for $\Delta\phi$ in the inside–jet tag scenarion in slices of $\Delta\eta$ and closure test are presented in figures E.5 and E.6 in apendix E. In general, the same trends are found for the $\Delta\eta$ slices. The acceptance is slightly lower and background is bigger for some bins. Purity and stability start at lower $\Delta\phi$ region with high values and grow to near 1.0. The closure test shows a good agreement.

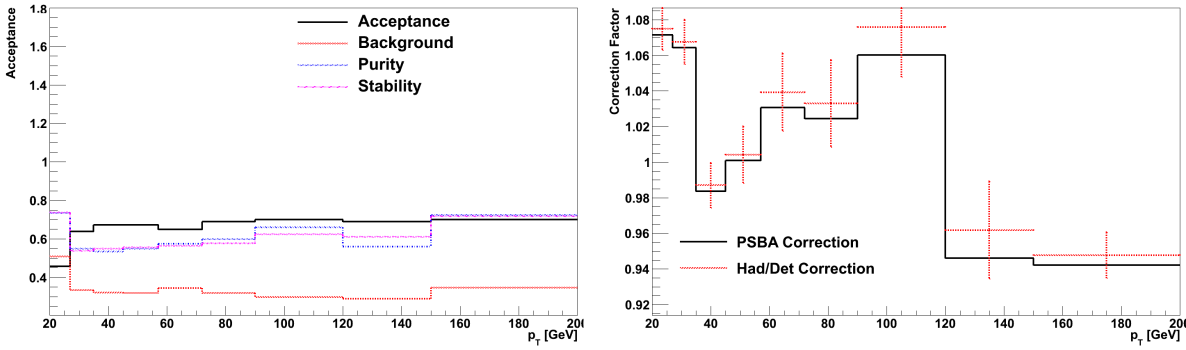


Figure 5.36: Acceptance, background, purity and stability (left) and closure test (right) for the bin–by–bin correction performed on the JetMETTau_2010A datasample corrected with PYTHIA 6 - Z2* Tune in the inside–jet tag scenario.

The PSBA variables for the leading inside–jet p_{T} are presented on the left side of figure 5.36 and the closure test is shown on the right side. The background starts at 0.5 for the first bin and then decreases to about 0.3 for the other bins. The acceptance starts at 0.45 in the first bin and jumps to 0.65 in the second, growing steedly to 0.75 toward the higher p_{T} region. Purity and stability start at 0.75, then decrease to 0.55 and grow to 0.75 towards the last bin. Both correction factor from PSBA and bin–by–bin are consistent.

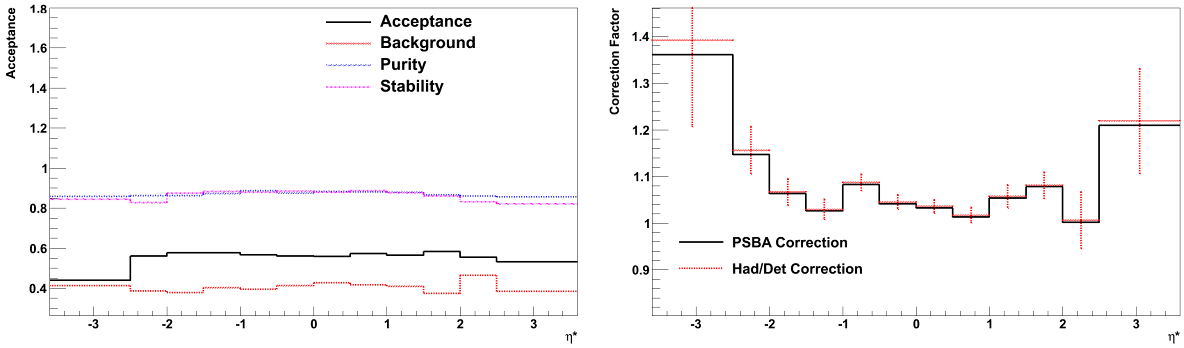


Figure 5.37: Acceptance, background, purity and stability (left) and closure test (right) for the bin–by–bin correction performed on the JetMETTau_2010A datasample corrected with Pythia 6 - Z2* Tune in the inside–jet tag scenario.

The PSBA variables for the leading inside–jet $\eta*$ are shown on the left side of figure 5.37 and the closure test is presented on the right side. Purity and stability are close to 0.9 in the center and decrease to 0.85 in the edges. The acceptance is 0.6 in the central region and decreases to between 0.05 and 0.15 in the edges of the distribution. The background is almost constant, with a value around 0.4. The correction factors agree.

In figure 5.38 the PSBA variables for leading outside–jet p_{T}, on the left, and the closure test, on the right, are presented. The background starts at 0.55 in the first bin and decreases to about 0.3 for the rest of the distribution. The acceptance is 0.75 for the first bin, decreasing to 0.7 towards the third bin, being 0.65 and 0.45 in the second and first bins respectively. Purity and stability have values of about 0.75 in the first bin, decrease to 0.55 in the second and then increase steadly up to 0.75 in the last bin. There is a good agreement between the correction factors obtained with PSBA variables and generator and detectors levels ratio.

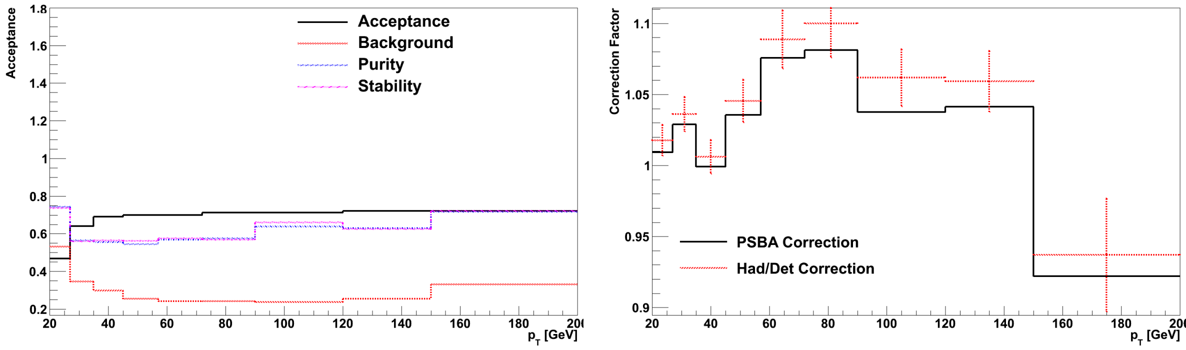


Figure 5.38: Acceptance, background, purity and stability (left) and closure test (right) for the bin–by–bin correction performed on the JetMETTau_2010A datasample corrected with Pythia 6 - Z2* Tune in the outside–jet tag scenario.

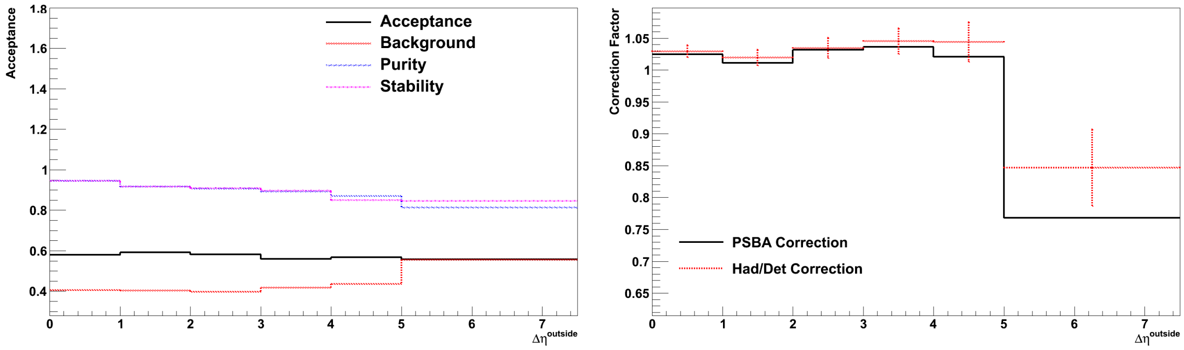


Figure 5.39: Acceptance, background, purity and stability (left) and closure test (right) for the bin–by–bin correction performed on the JetMETTau_2010A datasample corrected with Pythia 6 - Z2* Tune in the outside–jet tag scenario.

In figure 5.39 the PSBA variables as a function of $\Delta\eta^{out}$, on the left, and the closure test, on the right, are presented. The acceptance is between 0.55 and 0.6 for the whole range. The background grows from 0.4 to 0.45, jumping to 0.55 in the last bin. Both purity and stability are 0.85 for the first bin, decreasing to 0.8 in the last. Except for the last bin, the correction factors agree.

In general, the values of PSBA variables a within the acceptable range. Purity and stability are over 0.6 for most of the bins. The background rarely exceeds 0.5 and the acceptance very seldom is lower than 0.5. Finding many values outside these ranges would mean that the bin–to–bin or the inside–outside migrations are too big to use the bin–by–bin method. The closure tests were successful, allowing the conclusion that bin–by–bin correction can also be used in this analysis without any major problems.

5.5 Bin–by–Bin Correction

In this section the bin–by–bin correction factors for all measured observables are shown. The bin–by–bin correction factor is the average between PYTHIA 6 - Z2* Tune and PYTHIA 8 - 4C Tune. There is a set of factors for each pile–up condition. These factors will be, later on, compared with unfolding results.

The figure 5.40 show the correction factors for the $\Delta\phi$ distribution in the inclusive scenario. The plots display a dependence on $\Delta\phi$, larger correction factors are found in the lower $\Delta\phi$ region. This $\Delta\phi$ comes from higher low–p_T jet multiplicity present in the lower $\Delta\phi$ region, which have higher background and lower purity and stability. The factors varied between 0.6 and 1.1. In

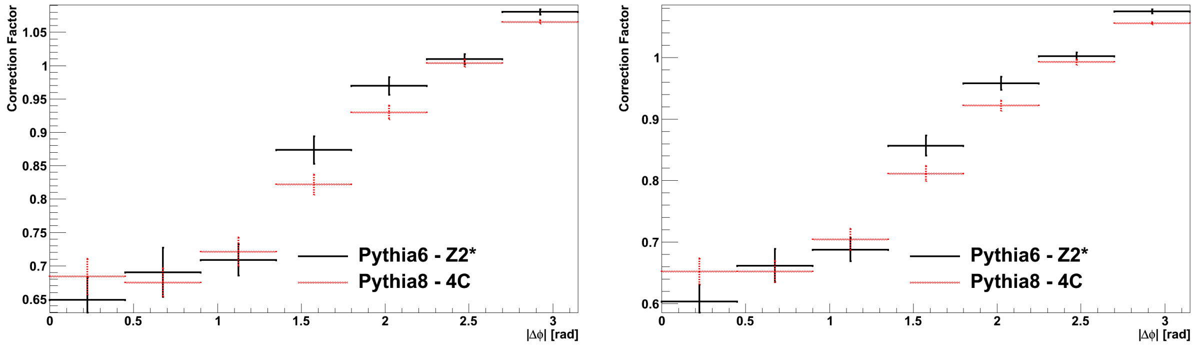


Figure 5.40: Correction factors to stable particle level for $\Delta\phi$ in the main selection for the JetMETTau_2010A (left) and JetMET_2010A (right) datasets.

most of the bins both MC agree within statistical uncertainty. The results are consistent between the different datasets, showing that pile–up does not play a significant role.

The $\Delta\phi$ correction factors for for different regions of $\Delta\eta$ to hadron level for the inclusive scenario are presented in figures F.1 and F.2 in appendix F. In overall, the correction factors have a larger difference between the different MCs than in the inclusive case and range from 0.60 to 1.15. The lower $\Delta\phi$ bins have the biggest statistical errors and also the larger difference between the MC samples, however they still agree within the statistical uncertainty. The trend is similar for both datasets.

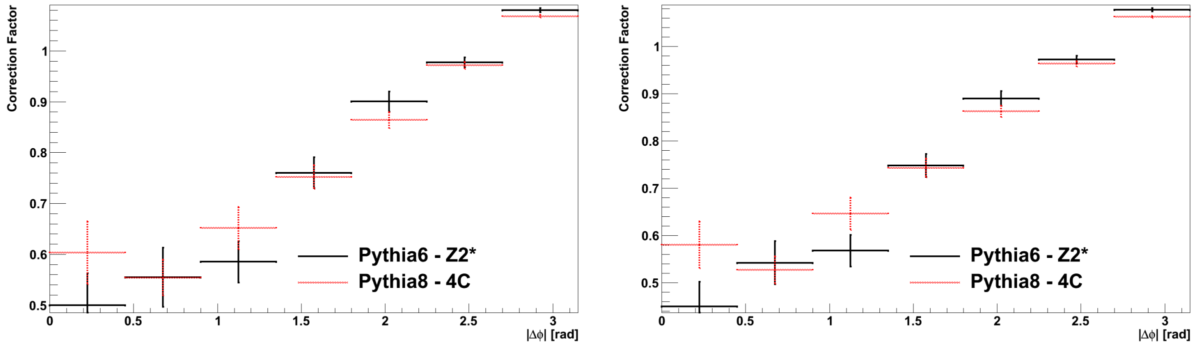


Figure 5.41: Correction factors to stable particle level for $\Delta\phi$ in the inside–jet veto scenario for the JetMETTau_2010A (left) and JetMET_2010A (right) datasets.

The correction factors for the $\Delta\phi$ distribution in the inside–jet veto scenario are shown in figure 5.41. A stronger dependence of the correction factors on $\Delta\phi$ is found, the values grow with $\Delta\phi$ from 0.4 to 1.1. The differences between the MC predictions are, in most bins, covered by the statistical uncertainties and are consistent between datasets.

The correction factors for the $\Delta\phi$ distribution in the inside–jet veto scenario in different regions of $\Delta\eta$ are shown in figures F.4 and F.4 in the appendix F. The correction factors varied from 0.4 to 1.1 and have the largest deviation from 1 for the lower $\Delta\phi$ region. Except for the lower $\Delta\phi$ region in the slice $4.5 > \Delta\eta \geq 7.5$, the discrepancies between the MCs are covered by statistics. The results for both datasets are consistent.

The correction factors for the $\Delta\phi$ distribution in the inside–jet tag scenario are presented in figure 5.42. The dependence with $\Delta\phi$ is stronger than in the main selection, showing two distinct regimes for low and high $\Delta\phi$ regions, creating two plateaus which include all bins except for the middle one. The correction factors varied between 0.65 and 1.10. Within statististics, both MC samples agree for most of the bins in both datasets.

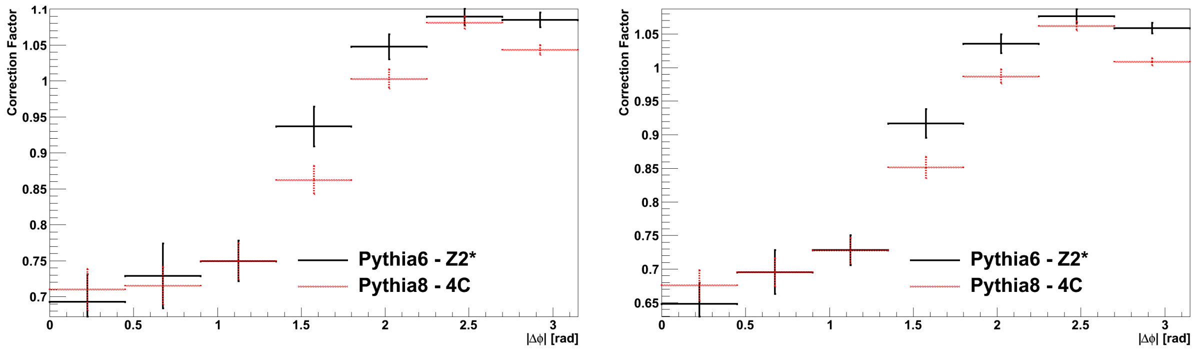


Figure 5.42: Correction factors to stable particle level for $\Delta\phi$ in the inside–jet tag scenario for the JetMETTau_2010A (left) and JetMET_2010A (right) datasets.

The correction factors for the $\Delta\phi$ distribution in the inside–jet tag scenario in different regions of $\Delta\eta$ are presented in figures F.5 and F.6 in appendix F. The trends are very similar as for the inclusive case. No significant discrepancies between different MCs samples and datasets are found. The factors range from 0.6 to 1.15.

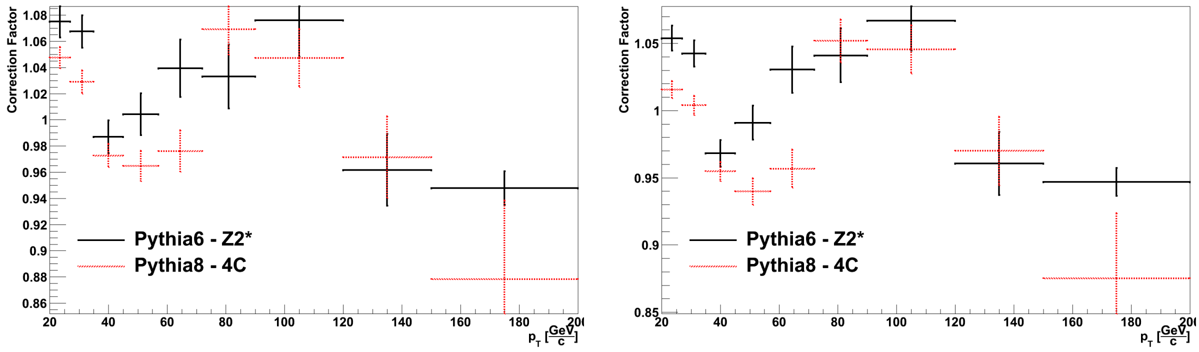


Figure 5.43: Correction factors to stable particle level for leading inside–jet p_{T} for the JetMETTau_2010A (left) and JetMET_2010A (right) datasets.

In figure 5.43 the correction factors as a function of the leading inside–jet p_{T} are presented. The corrections are very small and the values range only from 0.85 in higher p_{T} region to 1.10. A wave–like structure, dependent of p_{T}, is visible in all datasets. A small difference between the two MC samples is found in the lower p_{T} region, where PYTHIA 8 - 4C Tune tends to give factors closer to 1.0.

In figure 5.44 the correction factors as a function of the leading inside–jet η^* is presented. The factors run from 0.95 to 1.40, being smaller in the central region and growing towards the edges of the plot. A sligtly asymmetry is observed, which is too small to have any impact on the final result. The behaviour is similar across datasets and MC samples, except for the first bin, where the two MC are different in both datasets due to limited statistics.

In figure 5.45 the correction factors as a function of leading outside–jet p_{T} are shown. The values vary from 0.8 and 1.1, being bigger in the centre, $60 \geq p_{\mathrm{T}} \geq 100$ GeV. Except for the last bin, the correction factors are close to 1 and no significant structure is found. The two MC samples show a small difference in the central p_{T} region, where PYTHIA 8 - 4C Tune predict correction factors closer to 1.0.

In figure 5.46 the correction factors for the $\Delta\eta^{out}$ distribution is presented. The correction factors lie between 1.00 and 1.05 for all bins, except for the last where they fall to around 0.8. They agree between MC samples and datasets.

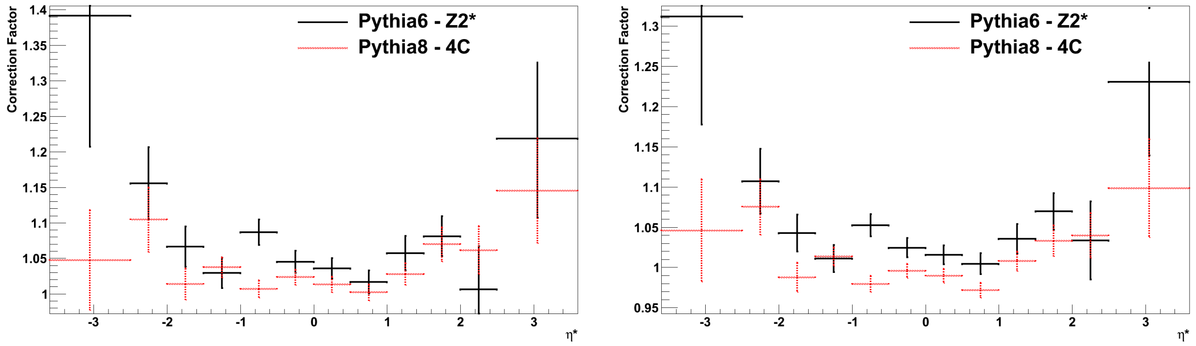


Figure 5.44: Correction factors to stable particle level for leading inside–jet η^* for the JetMETTau_2010A (left) and JetMET_2010A (right) datasets.

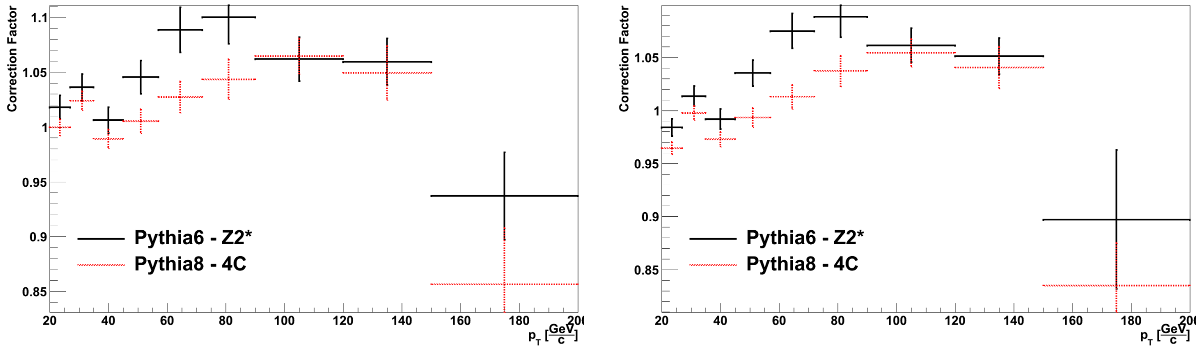


Figure 5.45: Correction factors to stable particle level for leading outside–jet p_{T} for the JetMETTau_2010A (left) and JetMET_2010A (right) datasets.

All correction factors have values around 1 and, as expected, did not result in big corrections. There is a clear dependence of the correction factors on $\Delta\phi$, which is explained by the larger fraction of low–p_{T} jets in the lower $\Delta\phi$ region. Also, this larger correction at lower $\Delta\phi$ is the pile–up contribution.

Although it is not necessary in this analysis, unfolding is used to produce the final corrected cross–section. A detailed exposition is given in the next chapter.

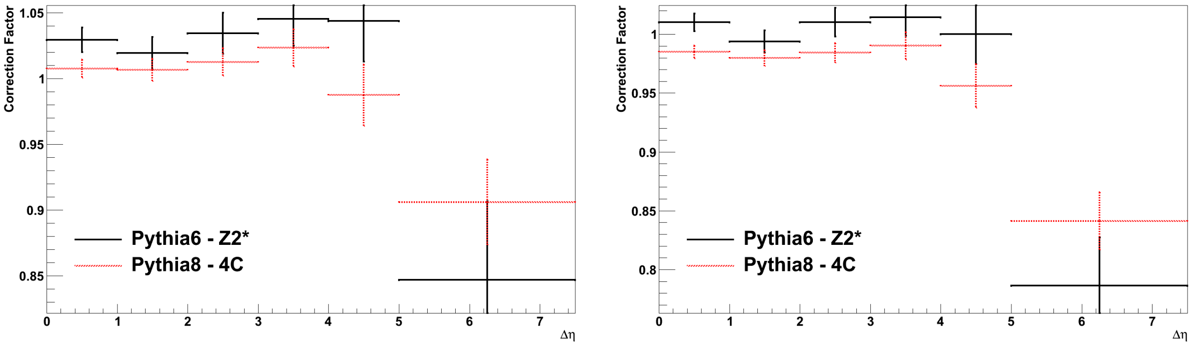


Figure 5.46: Correction factors to stable particle level for leading outside–jet $\Delta\eta^{out}$ for JetMETTau_2010A (left) and JetMET_2010A (right) datasets.

Chapter 6

Unfolding

In this chapter the inherent theory of unfolding (see section 6.1), the methods (see section 6.2), general benchmarking tests (see section 6.3) and results from the unfolding applied to the analysis of this thesis (see section 6.4) are presented.

6.1 Theory

Due to limited acceptance, finite resolution and migration effects, the measurement of a distribution does not yield the true result. A transformation of the measured distribution is necessary to obtain a result with a direct physical meaning, without the specific effects from the used detector. The result should allow a direct comparison with other experiments and theoretical distributions.

6.1.1 The direct and the inverse process

The transition from the true distribution $f(x)$ to the measured distribution $g(y)$ is described by the Fredholm integral [175]:

$$\int_{\Omega} A(y,x)f(x)dx + b(y) = g(y), \tag{6.1}$$

where $b(y)$ represents the backgroung contribution. The function $A(y,x)$ gives the response of the measurement device to the distribution $f(x)$. Two types of process are based on the equation:

- **direct process:** true distribution $f(x) \rightarrow g(y)$ measured distribution;
- **inverse process:** measured distribution $g(y) \rightarrow f(x)$ true distribution.

In the direct process the function $f(x)$ is converted into the detector signal $g(y)$. The direct processes can be simulated with Monte Carlo. Due to migration effects in the measurement, the measured distribution is usually smoother than the original distribution.

The process of the reconstruction of the true distribution $f(x)$ from the measured $g(y)$ is called the inverse process and belongs to the class of ill-posed problems. Due to the smoothing properties of the response function $A(y,x)$, the inverse process is usually unstable, which means that a small change in $g(y)$ may lead to large changes in the obtained $f(x)$.

The differences between true and measured distributions are ilustrated in figure 6.1. This differences come typically from the particle–physics detector response and are due to several effects:

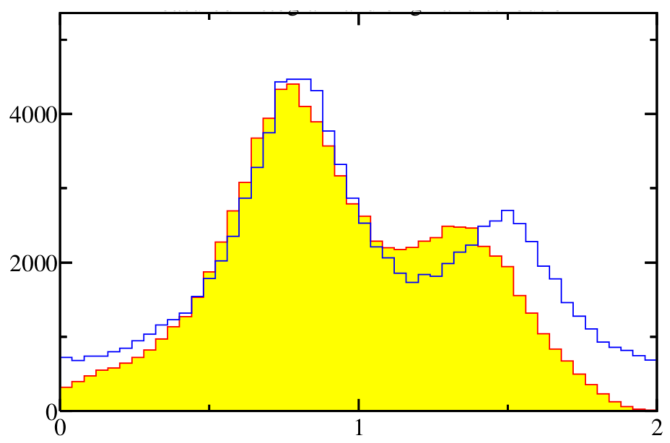


Figure 6.1: Example of true (blue) and measured distributions (red and yellow filled area) [175].

- In each bin there are statistical fluctuations from Poisson statistics, which can increase or decrease the number of counts [175].
- Due to limited resolution, migration between bins will happen, causing a given count to be recorded in the wrong bin.
- The limited acceptance and reduced efficiency can decrease the number of entries in a given bin
- Due to non-linear detector response, there can be an average shift in a certain direction.

Fakes are events outside of the phase–space which contaminate the measured distribution, while *missed* events are those that should be in the measured distribution and are not. For example, in a selection which has an energy threshold, events with E below the cut which are nevertheless reconstructed with E greater than the cut are called *fakes*. Events with E over the threshold which are measured with E smaller than the threshold or not measured at all, are *missed* events.

6.1.2 Reimann-Lebesgue lemma

As dicussed earlier, the integration of the response function $A(y,x)$ inside the Fredholm integral equation of first kind has a smoothing effect on the function $f(x)$. The reduction of the intensity of the high–frequency component is formulated in the Riemann–Lebesgue theorem. Using $f_p(x) = \sin(2\pi px)$ with $p = 1, 2, \ldots$ the result of the folding or the direct process is given by [175]:

$$\int_\Omega A(y,x)\sin(2\pi px)dx = g(y) \qquad p = 1, 2, \ldots \tag{6.2}$$

The Riemann–Lebesgue theorem states:

$$g(y) \to 0 \qquad when \qquad p \to \infty \tag{6.3}$$

As the high frequency contributions are supressed in the direct process, they are amplified in the inverse process, allowing for a small random pertubation of $g(y)$ to create a large effect in the unfolding result $f(x)$, which limits the capacity to reconstruct narrow structures, for example a mass peak.

6.1.3 Discretization of linear inverse problems

The equation 6.1 can be solved using numerical methods, by replacing the functions by sets of discrete data. Usually, in high energy physics, the measured distributions are given in the form of histograms with statistical errors. One can define the number of measured bins as m and the number of true bins as n. The measured values are transformed into a m–size vector **y** and the errors replaced by a m–size vector ϵ, while the response function $A(y, x)$ is translated into a m–by–n matrix **A** *response matrix*. Thus, the equation 6.1 becomes linear:

$$Ax + \epsilon = y, \tag{6.4}$$

assuming a case without background, where **x** is the n–size vector containing the unfolded result.

6.1.4 Solution by naive inversion

When the matrix A is quadratic, $m = n$, and if it has an inverse, the unfolded result x can be written as:

$$x = A^{-1}y \tag{6.5}$$

Using the standard law of uncertainties with the covariance matrix V_y taken from data, one could compute the covariance matrix of the estimate V_x according to:

$$V_x = A^{-1}V_y(A^{-1})^T \tag{6.6}$$

As a system of linear equation has a unique solution for $m = n$ equations and n unknowns, the covariance matrix V_y has no influence on the result **x**. Although the naive inversion is simple to perform, its results are usually not accurate. In addition, the case $m = n$ is not recommended, because of instabilities of the inverse processes. Therefore, more complex methods need to be applied.

6.2 Unfolding Methods

In the package RooUnfold 1.1.1 [174], used in this thesis, several methods are avalaible. In this section an overview of the bin–by–bin (see subsection 6.2.1), Bayesian (see subsection 6.2.2), SVD (see subsection 6.2.3) and TUnfold methods (see subsection 6.2.4) are presented.

6.2.1 Bin–by–bin Method

The bin–by–bin method is a simple case of unfolding which takes into account only the diagonal elements of the response matrix. This was applied in section 5.5. For this method, the response matrix **A** is always quadratic, $m = n$, and if one assumes that there is no background, the unfolding problem is reduced to:

$$x_i = \frac{1}{A_{i,i}} y_i. \tag{6.7}$$

The obvious shortcoming of this method is the fact that it does not take into account the correlations between bins, assuming an over–simplified picture of the measurement. As discussed in section 5.4, the confidence of this method is determined by the acceptance, background, purity and stability values. To be able to use the bin–by–bin method, one should ensure that the acceptance is very high and the background is very low, otherwise one must correct for it or the method will likely introduce a bias in the result. The migrations can not be corrected directly, thus in the case of very low purity or stability, another unfolding method is required. As an additional requirement to use the bin–by–bin method, the magnitude of the off–diagonal elements when compared with the diagonal has to be small.

6.2.2 Bayesian Method

The Bayesian unfolding method was proposed by d'Agostini [176]. One can formulate Bayes' theorem in terms of several independent causes $C_i, i = 1, 2, ..., n_c$ which can produce the effect E. If one assumes the initial probability of the causes, $P(C_i)$, to be known as well as the conditional probability of the ith cause to produce the effect $P(E|C_i)$, Bayes formula can be written as:

$$P(C_i|E) = \frac{P(E|C_i).P(C_i)}{\sum_{l=1}^{n_C} P(E|C_l).P(C_l)} \tag{6.8}$$

When one event is observed, the probability that it has been due to the ith cause is proportional to the probability of the cause multiplied by the probability of the cause to produce the effect. Bayes' formula increases the knowledge of $P(C_i)$ as the number of the observations increases. However, $P(C_i|E)$ depends on the initial probability of the causes, which has to be known a priori. The final distribution depends on $P(E|C_i)$, which must be estimated with Monte Carlo methods.

When observing $n(E)$ events with effect E, the expected number of events from each cause is

$$n(C_i) = n(E).P(C_i|E). \tag{6.9}$$

The outcome of a measurement has several possible effects $E_j, j = 1, 2, ..., n_E$ for a given cause C_i. For each effect, the $P(C_i|E_j)$ can be computed. These conditional probabilities $P(C_i|E_j)$ come from the response matrix A. Thus, one can write the equation 6.8 as:

$$P(C_i|E_j) = \frac{P(E_j|C_i).P_0(C_i)}{\sum_{l=1}^{n_C} P(E_j|C_l).P_0(C_l)}, \tag{6.10}$$

being P_0 the initial probability. The equation 6.10 should satisty:

- $\sum_{i=1}^{n_C} P_0(C_i) = 1.$
- If the probability of a cause is initially set to zero, it can never change.
- The normalization condition is $\sum_{i=1}^{n_C} P(C_i|E_j) = 1.$
- $0 \leq \epsilon_i \leq 1$ for $\epsilon_i = \sum_{j=1}^{n_E} P(E_j|C_i)$: there is no need for each cause to produce at least one of the effects taken into consideration, with ϵ_i being the efficiency of detection of cause C_i in any of the possible effects.

After N_{obs} experimental observations the expected number of events assigned to each of the causes, due to observed events, can be computed to each effect from equation 6.9:

$$n(C_i)_{obs} = \sum_{j=1}^{n_e} n(E_j)P(C_i|E_j). \tag{6.11}$$

If one takes the measurement efficiency into account, the equation 6.11 can be written as:

$$n(C_i) = \frac{1}{\epsilon_i} \sum_{j=1}^{n_e} n(E_j)P(C_i|E_j) \quad \text{with} \quad \epsilon_i \neq 0 \tag{6.12}$$

From these unfolded events, the true total number of events (see equation 6.13), the final probabilities of the causes (see equation 6.14) and the overall efficiency (see equation 6.15) can be estimated:

$$N_{true} = \sum_{i=1}^{n_C} n(C_i), \tag{6.13}$$

$$P(C_i) \equiv P(C_i|n(E)) = \frac{n(C_i)}{N_{true}}, \tag{6.14}$$

$$\epsilon = \frac{N_{obs}}{N_{true}}. \tag{6.15}$$

The obtained ϵ may be different from the initial efficiency ϵ_0 computed from MC simulations:

$$\epsilon_0 = \frac{N_{meas}}{N_{true}} = \frac{\sum_{i=1}^{n_C} \epsilon_i P_0(C_i)}{\sum_{i=1}^{n_C} P_0(C_i)} \tag{6.16}$$

The closer the initial distribution $P_0(C)$ is to the data, the better the agreement with the final distribution will be. It can be demostrated with a toy MC that $P(C)$ lies between $P_0(C)$ and the true one, suggesting to proceed iteratively. Thus, the Bayesian unfolding can be performed with the following steps:

1. choose the initial distribution $P_0(C)$ and the expected number of events $n_0(C_i) = P_0(C_i).N_{obs}$ from the best knowledge of the process under study;
2. compute $n(C)$ and $P(C)$;
3. calculate the χ^2 between $n(C)$ and $n_0(C)$;
4. replace $P_0(C)$ by $P(C)$ and $n_0(C)$ by $n(C)$ and go to step 2. If, after the second iteration, the value of χ^2 is arbitrary small enough, stop iteration.

6.2.3 SVD Method

The SVD (**S**ingular **V**alue **D**ecomposition) method proposes an alternative method to solve equation 6.4 [177]. An SVD of a real $m \times n$ response matrix **A** can be numerically factorized in the form [178, 179]:

$$A = USV^T \tag{6.17}$$

where **U** is an $m \times m$ orthogonal matrix, **V** is an $n \times n$ orthogonal matrix and **S** is an $m \times n$ diagonal matrix with non–negative diagonal elements. These matrices have to obey to the following conditions:

- $UU^T = U^TU = I$;
- $VV^T = V^TV = I$;
- $S_{ij} = 0$ for $i \neq j$;
- $S_{ii} \equiv s_i \geq 0$.

The quantities s_i are called singular values of the matrix A, while columns of **U** and **V** are called the left and right singular vectors. If **A** is itself orthogonal, all its singular values are equal to 1, while a degenerate matrix will have at least one 0 among the singular values. If some singular values are much smaller than other singular values or equal to 0, it could be difficult to invert the matrix. Thus, SVD is a possible method to treat such problem and the solution is given by:

$$x = VS^{-1}U^Ty. \tag{6.18}$$

One of the avaliable implementations is TSVD [180].

6.2.4 TUnfold Method

The TUnfold method [182] solves the inverse problem defined in equantion 6.4 by minimizing the χ^2 using the following equation:

$$\chi^2 = (y - Ax)^T V_{yy}^{-1}(y - Ax) + \tau^2 (L(x - x0))^T L(x - x0) + \lambda \sum_i (y_i - (Ax)_i) \tag{6.19}$$

where the first term is the MC input, the second is the regularization and the third is responsable for the preservation of the distribution area. τ is the parameter defining the regularisation strength, L is an $n \times l$ matrix with the regularisation conditions, x_0 is the bias distribution and λ is a Langragian multiplier.

The χ^2 can be minimized as a function of x for fixed $\lambda = 0$ or as a function of x and λ.

The regularization is necessary because the simple least squares solution is dominated by contributions from data and rounding errors. The choice of the regularization strength τ is done using the L–curve, which takes the name from its shape and is defined from λ and τ dependent functions [181]. An example of an L–curve is found in figure 6.2. The best L–curve value of the regularization is found at the kink of the L–shaped plot, avoiding a too weak or too strong regularization.

The interface to the TUnfold method [182] implementation in RooUnfold has 4 different regularization methods:

- **None**: No regularization is applied;
- **Size**: The size of $(x - x_0)$ is minimized;
- **Derivative**: The first derivative of $(x - x_0)$ is minimized;
- **Curvature**: The second derivative of $(x - x_0)$ is minimized.

6.3 Benchmarking tests

In this thesis, benchmarking tests were performed to ensure a good implementation and understanding of a given unfolding method and its implementation. The tests comprised the influence of the binning, influence of the training function and correlations.

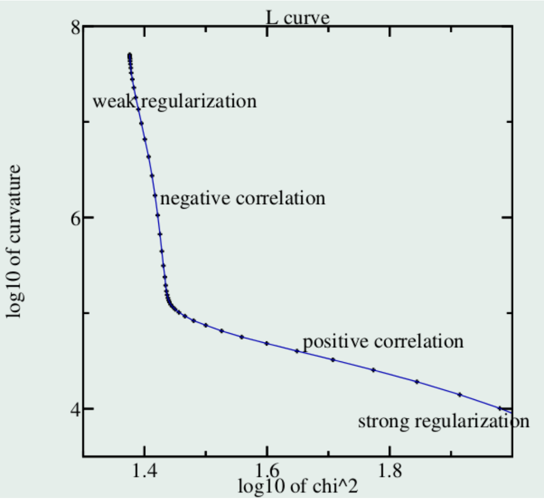


Figure 6.2: Example of an L–Curve [181].

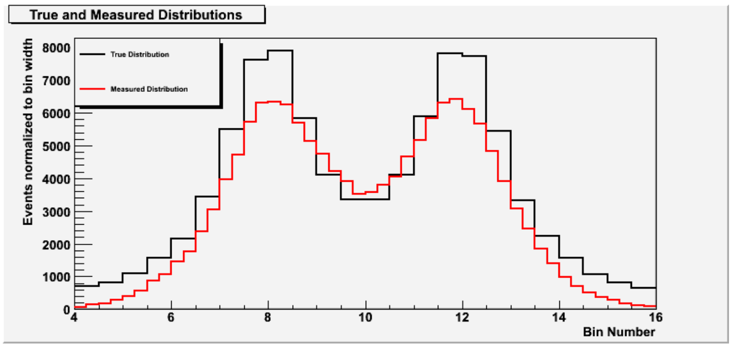


Figure 6.3: True (back) and measured (red) distributions used in the benchmarking test.

In this tests, true and measured distributions have been generated and are presented in figure 6.3. The true data were constructed from 10^5 events distributed according to a double Breit–Wigner function. The measured distribution was obtained by smearing each event with a Gaussian function with no bias and a width of 0.5. There was an additional acceptance function applied described by

$$\epsilon = 1 - \frac{2(x-10)}{36}. \tag{6.20}$$

The true distribution has 24 equidistant bins, with a bin width of 0.25, while the measured distribution has 48 bins with a bin width of 0.25.

An unbiased response matrix was generated from 10^6 uniformly distributed events. This unbiased matrix is necessary to remove any possible effect of the distribution on the final result. The training functions of the response matrix are presented in figure 6.4. In this case, as the training true distribution is uniform, the training measured distribution has the same shape as the acceptance.

The measured distribution was unfolded with TUnfold algorithm using the size regularization and the result is presented in figure 6.5. The unfolded result does not agree with the true distribution and sometimes is more away than the measured distribution. Thus, as expected, a blind unfolding cannot be performed, as the inverse problems are highly instable, so one must control how the method behaves.

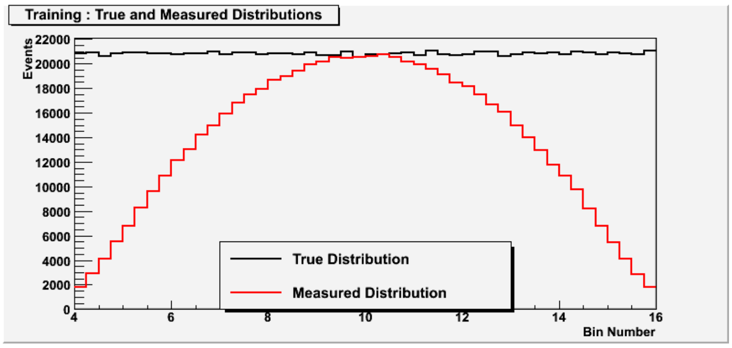


Figure 6.4: Training distributions used to create the unbiased response matrix in the benchmarking test: true (black) and measured (red).

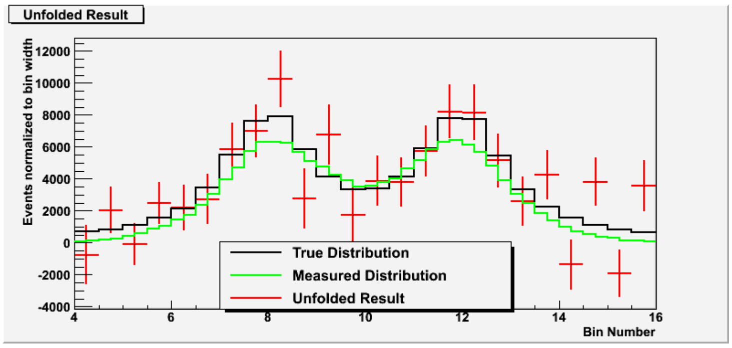


Figure 6.5: Unfolded result for the benchmarking test: true (black), measured (green) and unfolded (red) distributions.

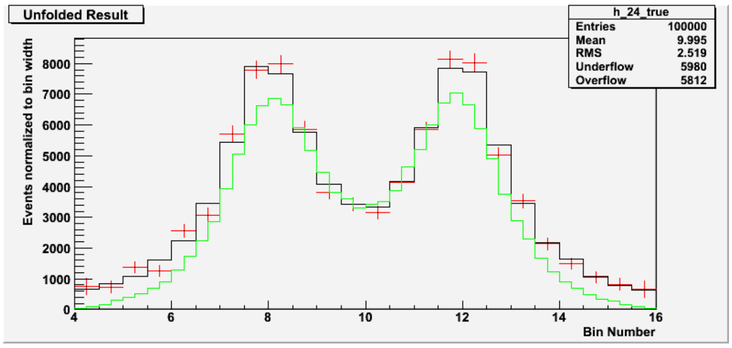


Figure 6.6: Unfolded result for the benchmarking test with smearing smaller than bin width: true (black), measured (green) and unfolded (red) distributions.

The study was repeated using a smearing width equal to 0.3, which is smaller than the bin width. The result is presented in figure 6.6. This test proved that, when the smearing is smaller than the bin width, the unfold result is compatible with the true distribution.

The bin correlations are presented in figure 6.7 with smearing equal to 0.5 on the left side and equal to 0.3 on the right side. It can be concluded that the correlations smear out as the smearing increases. The neighbouring bins are strongly correlated. In both cases, the correlation, ρ, sign alternates and the correlation decreases between i–next neighbour to the $i+1$–next neighbour.

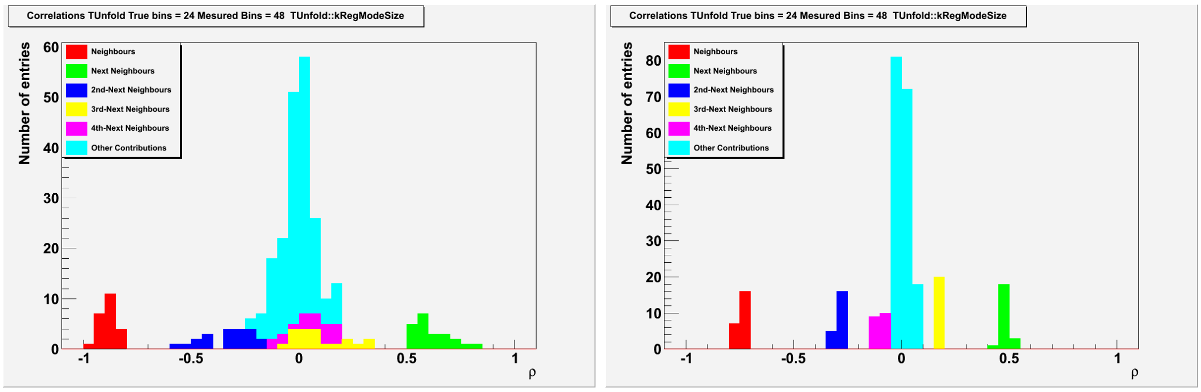


Figure 6.7: Correlations between bins for the benchmarking test: smearing equal to bin width (left) and smaller than the bin width (right).

6.4 Unfolding results

In this section, the response matrices (see subsection 6.4.1), the closure test (see subsection 6.4.2), the cross–closure test (see section 6.4.3) and unfolded distributions (see subsection 6.4.4) of this analysis are presented.

6.4.1 Response matrices

To construct the response matrices, the number of true bins was always twice of the measured ones. This is necessary to a better modeling of the bin migrations.

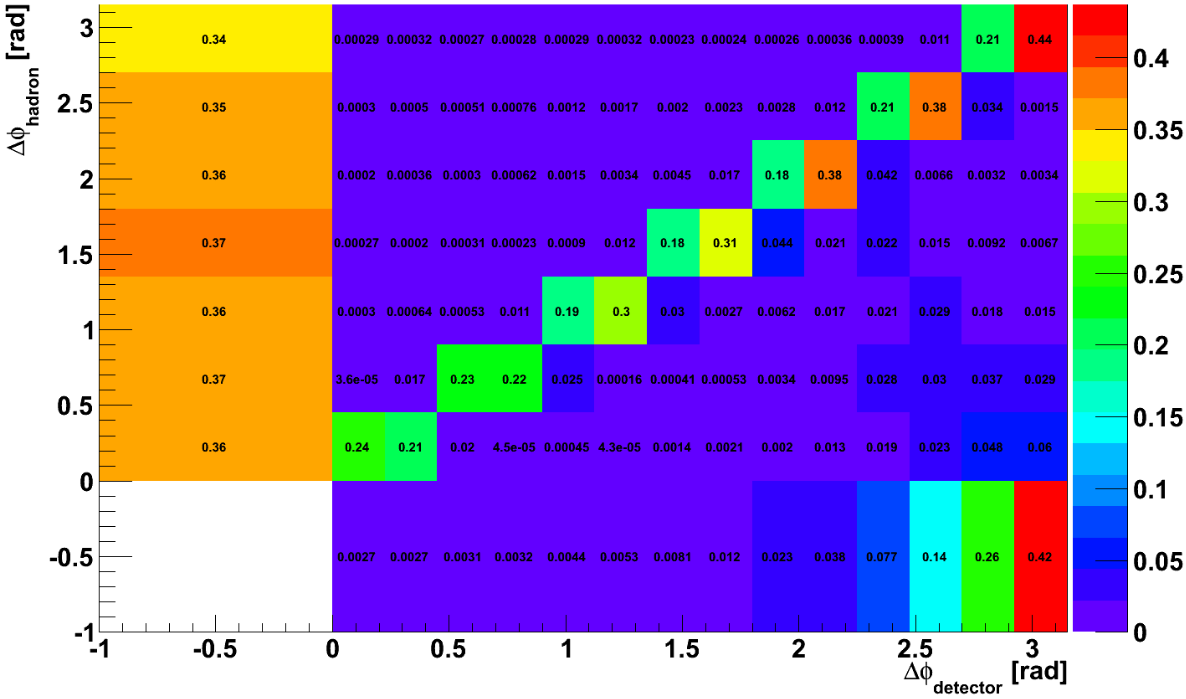


Figure 6.8: Response Matrix for $\Delta\phi$ in inclusive scenario for the JetMETTau_2010A dataset obtained with Pythia 6 - Tune Z2*. Fakes are represented in the negative bin of $\Delta\phi_{hadron}$. Missed eventsare represented in the negative bin of $\Delta\phi_{detector}$.

The response matrix for $\Delta\phi$ in the inclusive scenario, which includes fake and missed events, can be found in figure 6.8. Most of the events sit on the diagonal, except for a significant amount of fake and missed events, which are represented in the underflow bins of the hadron and detector axis respectively. A similar behavior is found for the response matrices obtained with Pythia 8, for the other datasets and $\Delta\phi$ distributions.

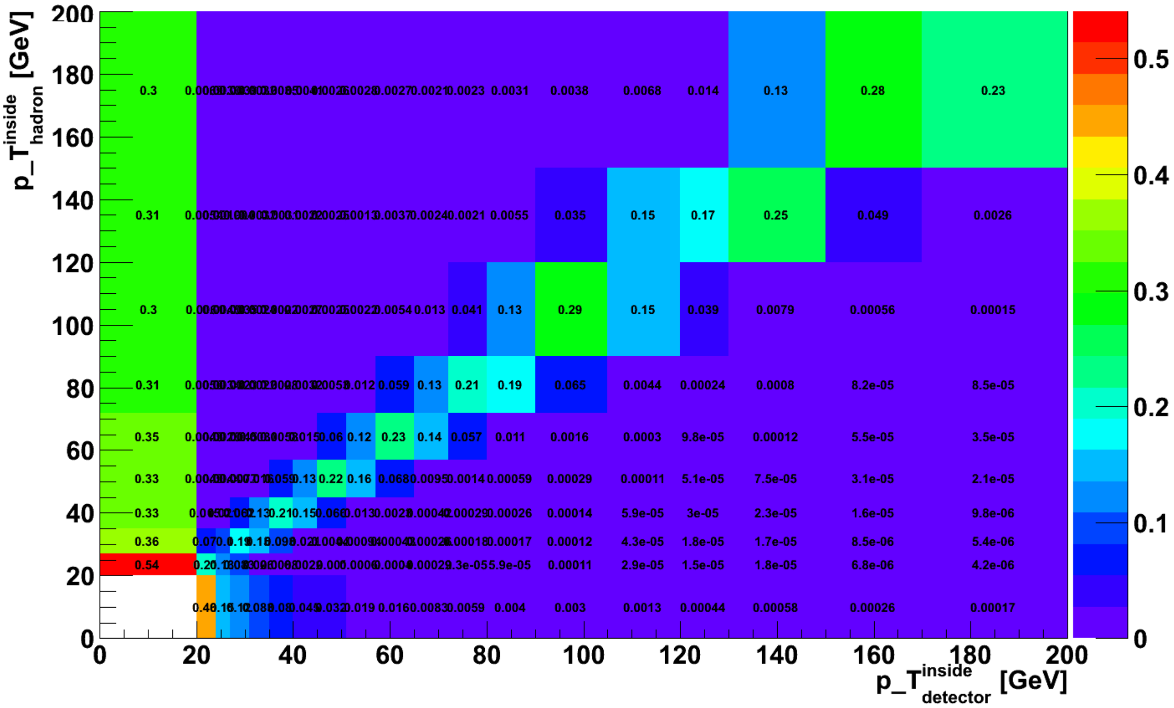


Figure 6.9: Response Matrix for the leading inside–jet p_T for the JetMETTau_2010A dataset obtained with Pythia 6 - Tune Z2*. Fake events are represented in the negative bin of $p_T{}^{inside}_{hadron}$. Missed events are represented in the negative bin of $p_T{}^{inside}_{detector}$.

In figure 6.9 the response matrix for the leading inside–jet p_T is presented. Except for a significant fraction of fake and missed events, most events are on the diagonal or in next–to–diagonal bins. This portion of the fake and missed events is due to the p_T cut on the leading inside–jet and one can conclude that the bin migrations in p_T are small when compared to the migrations into and out of the sample. Similar results were obtained for the leading outside–jet p_T distribution and also across different MC predictions and datasets.

The response matrix for the leading inside–jet η^* is shown in figure 6.10. Although most of the events are on the diagonal, the fraction of fake and missed events is considerable. This fraction is, once more, due to the p_T cut. The same conclusions can be drawn for both MCs and datasets.

The response matrix for the leading $\Delta\eta^{outside}$ is presented in figure 6.11. Once again, most of the elements are on the diagonal, except for a portion of fake and missed events due to the p_T cut. Similar effects were found for the other datasets and MC predictions.

By analysing the response matrix, one concludes that the migrations in and out of the sample are the major problem, rather than bin migrations. This was already indicated by the large values of purity and stability and low values of acceptance and high values of background. Thus, unfolding is expected to be able to deal with this problem in a slightly more efficient way than the simple bin–by–bin approach.

6.4.2 Consistency test

To make sure that the unfolding methods perform as expected, a consistency test has been carried out for each distribution. It consists of unfolding the MC prediction on detector level with the response matrix obtained with the same prediction. The result should be close to the

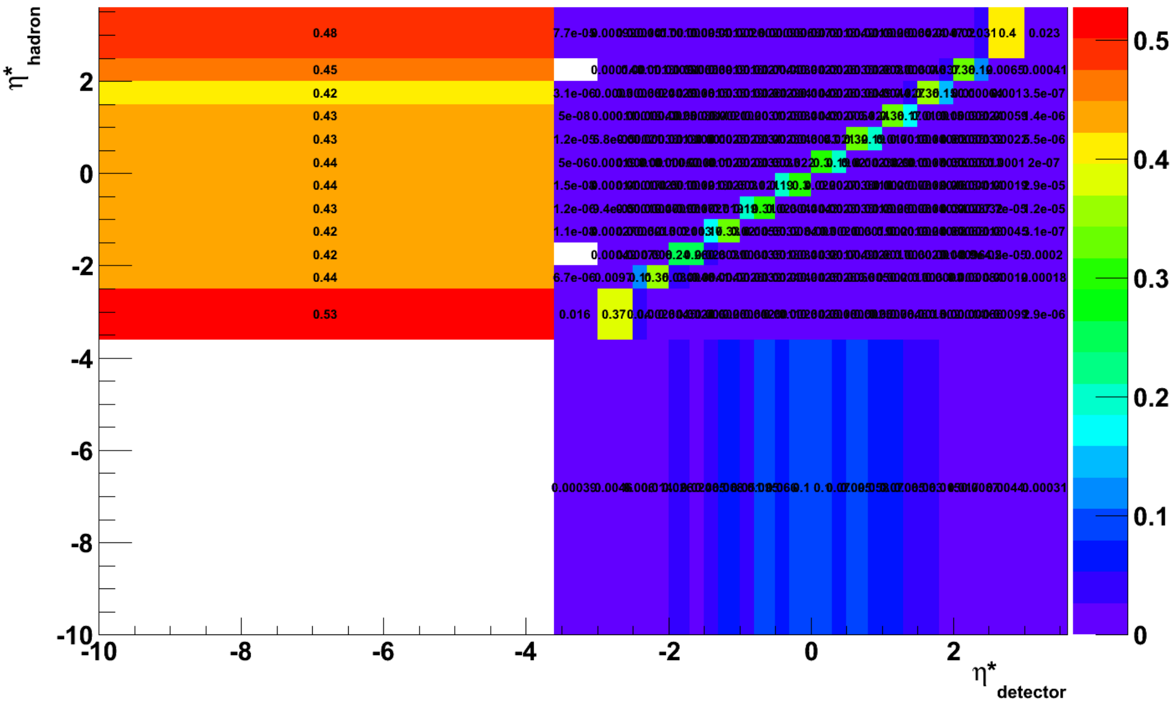


Figure 6.10: Response Matrix for the leading inside–jet $\eta*$ for the JetMETTau_2010A dataset obtained with PYTHIA 6 - TUNE Z2*. Fakes are represented in the negative bin of $\eta*_{hadron}$. Missed events are represented in the negative bin of $\eta*_{detector}$.

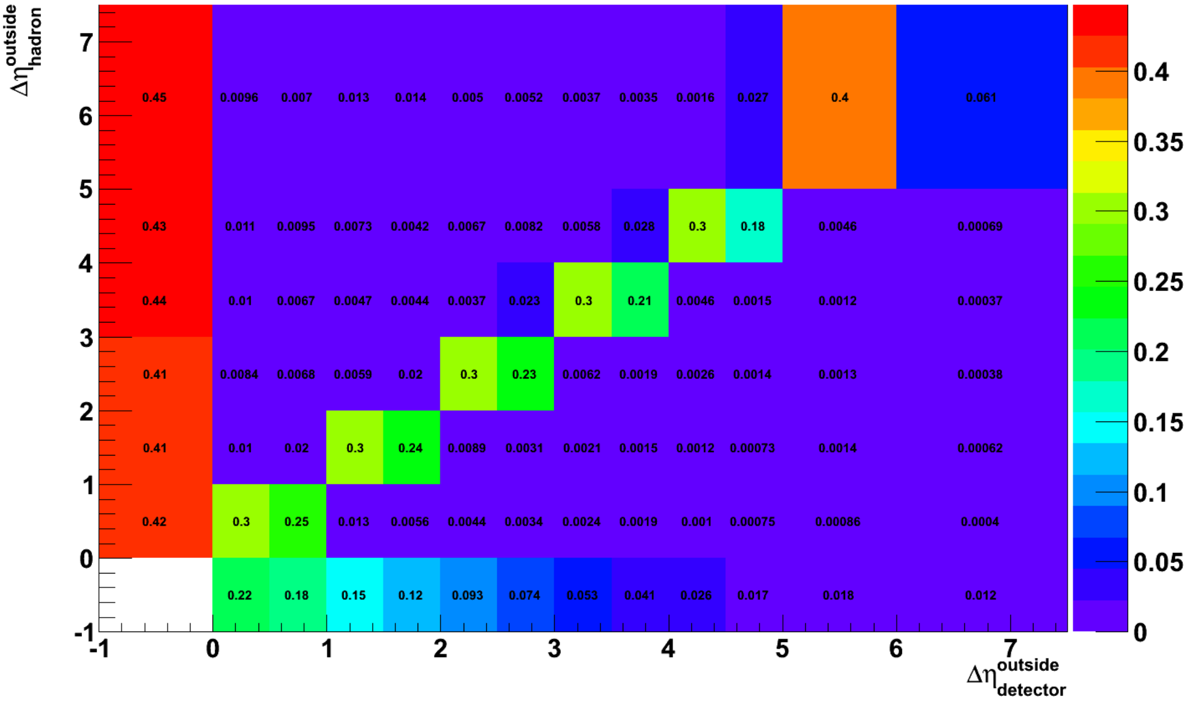


Figure 6.11: Response matrix for $\Delta\eta^{out}$ in inclusive scenario for the JetMETTau_2010A dataset obtained with PYTHIA 6 Z2*tune. Fakes are represented in the negative bin of $\Delta\eta^{out}_{hadron}$. Missed events are represented in the negative bin of $\Delta\eta^{out}_{detector}$.

hadron–level predition, otherwise there is some problem with the method, implementation or input.

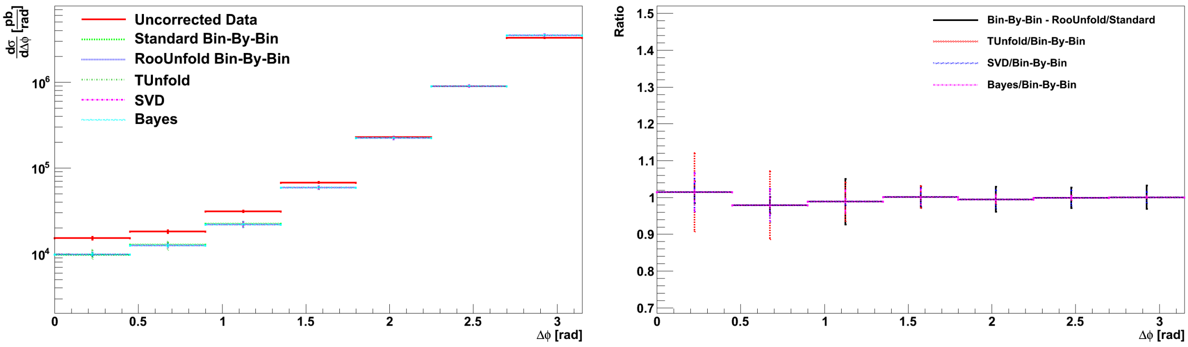


Figure 6.12: $\Delta\phi$ in the inclusive scenario for the JetMETTau_2010A daset. The Pythia 6 - Tune Z2* on detector level unfolded with Pythia 6 - Tune Z2* response matrix. Cross-section (left) and ratio (right). Pythia 6 - Tune Z2* on detector level is represented with a red line, denoted as uncorrected data. The MC truth is labeled as standard bin-by-bin and is represented with a green line. The other curves represent the unfolding results.

The consistency test for $\Delta\phi$ in the inclusive scenario is presented in figure 6.12. One can see that when the Pythia 6 response matrix is used to unfold Pythia 6 on detector level with Bayes method, one gets the hadron–level cross-section back. The small differences, in the order of 2%, were expected because of the diference in the background subtraction method and negligible since as they are much smaller than the uncertainty of the measurement. Similar results were obtained for the other $\Delta\phi$ distributions and also when using Pythia 8 to unfold itself.

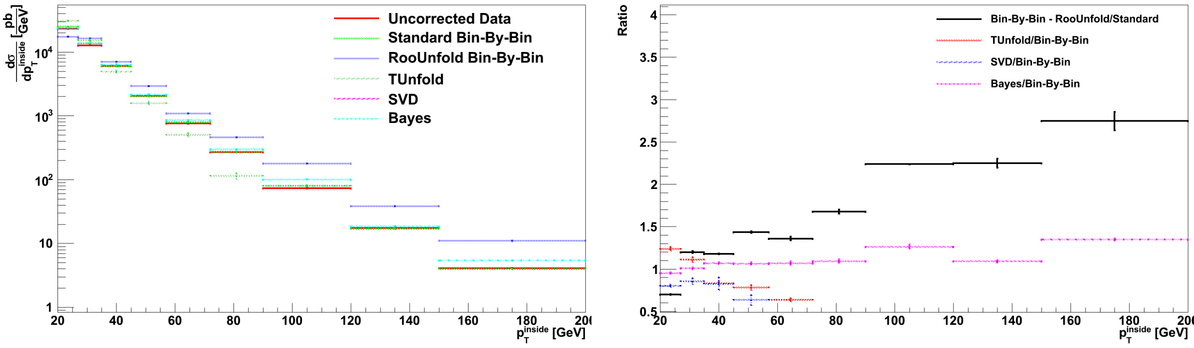


Figure 6.13: Pythia 6 - Tune Z2* on detector level unfolded with Pythia 6 - Tune Z2* response matrix for the leading inside–jet p_T for the JetMETTau_2010A dataset. Cross-section (left) and ratio (right). The Pythia 6 - Tune Z2* on detector level is represented with a red line, denoted as uncorrected data. The hMC truth is labeled as standard bin-by-bin and is represented with a green line. The other curves represent the unfolding results.

The consistency test for the leading inside–jet p_T is shown in figure 6.13. One can see that all methods, except for Bayes, have trouble to reconstruct the hadron–level prediction in the medium– and high– p_T region. On the other hand, the Bayesian method deviates just a few % to the bin–by–bin result. Similar results were obtained for the other datasets and also with the Pythia 8 prediction unfolded with itself. Similar trends were found for the leading outside–jet p_T distribution.

In figure 6.14 the unfolding consistency test for the leading inside–jet $\eta*$ distribution is presented. All unfolding methods perform in an acceptable way in the central region, but deviate in the edges. The Bayesian method yields less difference from the hadron level prediction than any other unfolding method. Similar conclusions can be taken from the Pythia 8 distribution unfolded with Pythia 8 response matrix as well with the weights of the different datasets.

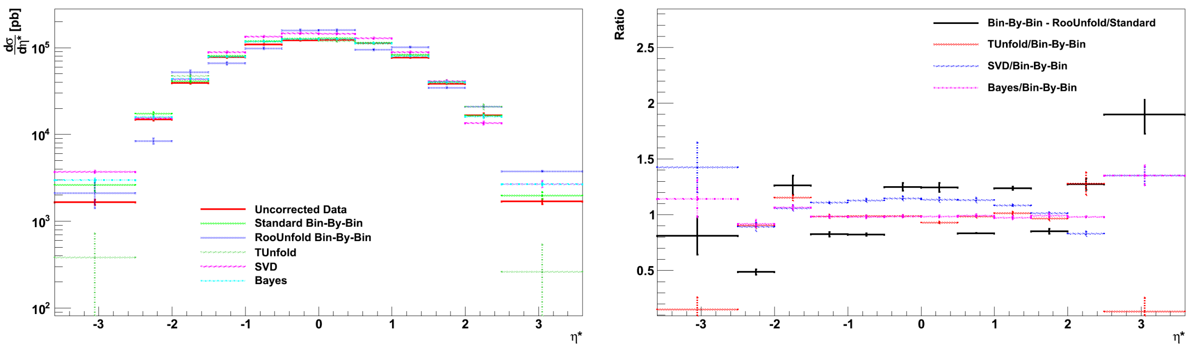


Figure 6.14: Pythia 6 - Tune Z2* on detector level unfolded with Pythia 6 - Tune Z2* response matrix for the leading inside–jet η^* for the JetMETTau_2010A dataset. Cross-section (left) and ratio (right). The Pythia 6 - Tune Z2* on detector level is represented with a red line, denoted as uncorrected data. The MC truth is labeled as standard bin-by-bin and is represented with a green line. The other curves represent the unfolding results.

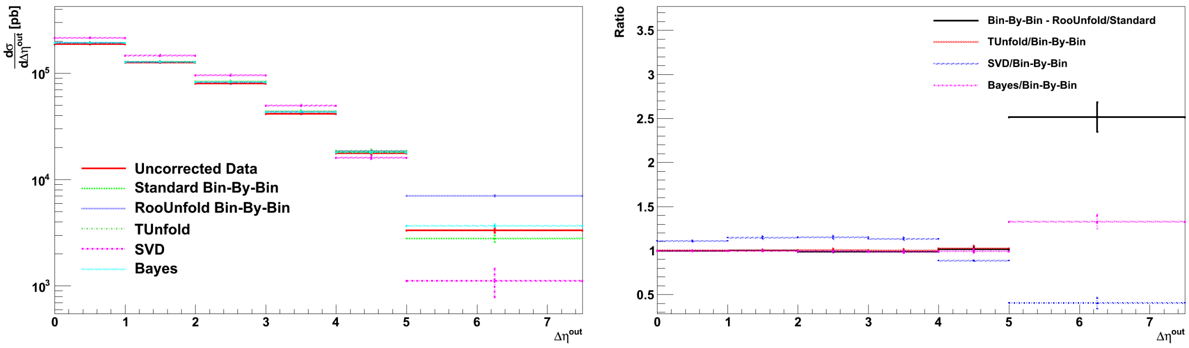


Figure 6.15: Pythia 6 - Tune Z2* on detector level unfolded with Pythia 6 - Tune Z2* response matrix for the leading $\Delta\eta^{out}$ for the JetMETTau_2010A dataset. Cross-section (left) and ratio (right). The Pythia 6 - Tune Z2* on detector level is represented with a red line, denoted as uncorrected data. The MC truth is labeled as standard bin-by-bin and is represented with a green line. The other curves represent the unfolding results.

In figure 6.15 the consistency test for $\Delta\eta^{out}$ distribution is shown. Except for the last bin, all methods could reconstruct the distribution on hadron level, but only Bayes' method provides a consistent result for all the bins. This outcome was also found for Pythia 8 and in other datasets.

The consistency test showed that the Bayesian method deliver the best results.

6.4.3 Closure test

A closure test is an additional cross–check to the unfolding process, because it provides a statistical independent test. It can be performed by dividing the MC events in two sub–samples, using one half to obtain the response matrix and the other half to obtain the measured distribution. Statistical independent samples can also be obtained by using a MC generator to compute the response matrix and another to provide the measured distribution. This analysis used the latter closure test.

The unfolding closure test for $\Delta\phi$ in the inclusive scenario is presented in figure 6.16. When using the Pythia 8 response matrix to unfold the Pythia 6 result, one gets a result close to the original values. The discrepancy, which is around 5% for most of the bins, is in line with what one could expect because the response matrix and the measured distribution are generated by different MC. Similar results were obtained when the opposite was done, using the PythiaA 6

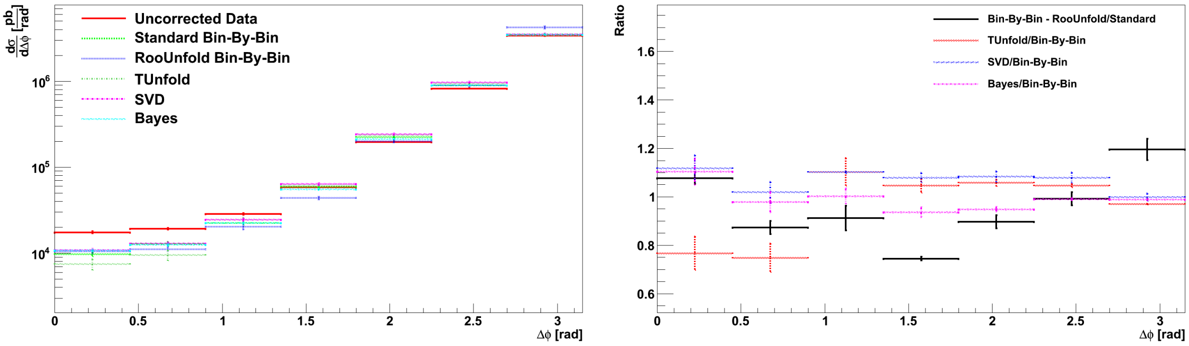


Figure 6.16: Pythia 6 Z2* tune on detector level unfolded with Pythia 8 4C tune response matrix for $\Delta\phi$ in the inclusive scenario for the JetMETTau_2010A dataset. Cross-section (left) and ratio (right). The Pythia 6 Z2* tune on detector level is represented with a red line, denoted as uncorrected data. The MC truth is labeled as standard bin-by-bin and is represented with a green line. The other curves represent the unfolding results.

response matrix to unfold the Pythia 8 distribution. The outcome obtained for other datasets and other $\Delta\phi$ distributions was similar.

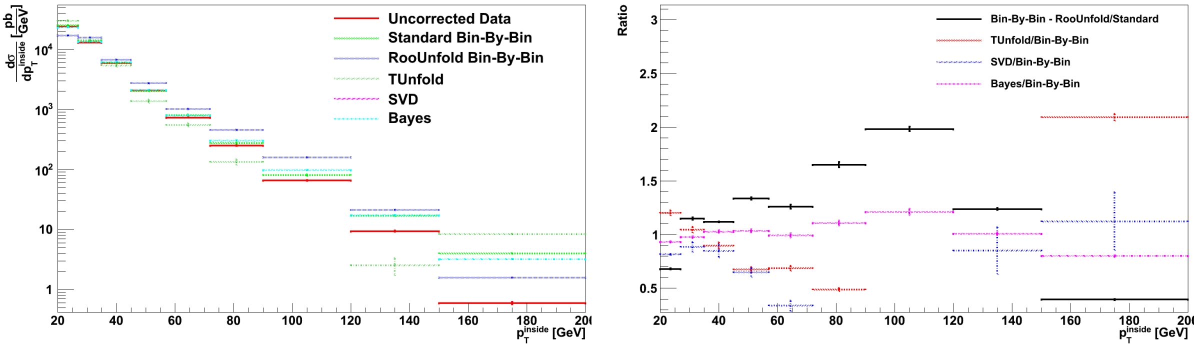


Figure 6.17: Pythia 6 Z2* tune on detector level unfolded with Pythia 8 4C tune response matrix for the leading inside–jet p_T for the JetMETTau_2010A dataset. Cross-section (left) and ratio (right). The Pythia 6 Z2* tune on detector level is represented with a red line, denoted as uncorrected data. The MC truth is labeled as standard bin-by-bin and is represented with a green line. The other curves represent the unfolding results.

The unfolding closure test for leading inside–jet p_T is shown in figure 6.17. The Bayesian unfolding seems to be the only method which performs in an acceptable way, yielding results compatible with the bin–by–bin method on the level of few %. The other methods fail in the middle– and high–p_T regions, as they did for the consistency test. The same conclusions can be drawn from the leading outside–jet p_T distribution. Similar results were obtained for different MC predictions and datasets.

In figure 6.18, the unfolding closure test for leading inside–jet η^* is shown. In the central region, the TUnfold and Bayes methods yield good results, being only a few % away from the true result. All methods have troubles in the edges, but Bayes' method gives the most reasonable results. Similar outcomes were found for the other MC predictions and datasets.

In figure 6.19, the unfolding closure test for $\Delta\eta^{out}$ is presented. Although all methods describe reasonably the low $\Delta\eta^{out}$ region, the best result is provided by Bayes method. TUnfold and SVD are away in the last two bins. Similar results have been obtained for other MC predictions and datasets.

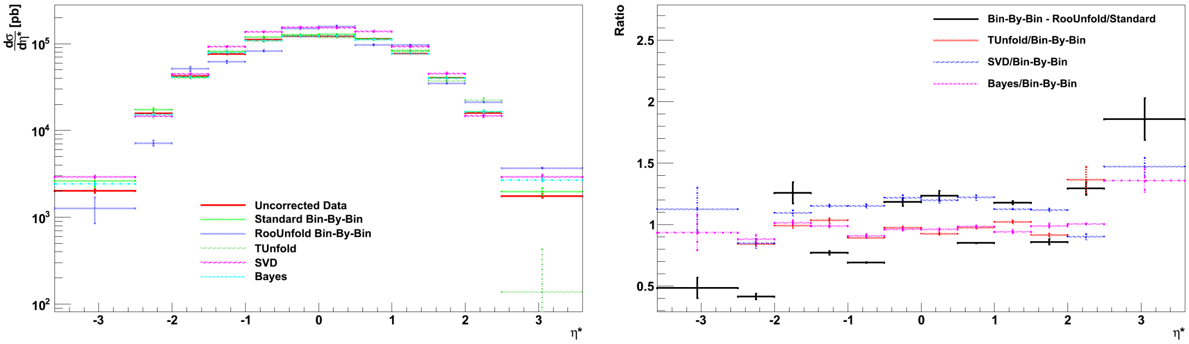


Figure 6.18: PYTHIA 6 Z2* tune on detector level unfolded with PYTHIA 8 4C tune response matrix for the leading inside–jet η^* for the JetMETTau_2010A dataset. Cross-section (left) and ratio (right). The PYTHIA 6 Z2* tune on detector level is represented with a red line, denoted as uncorrected data. The MC truth is labeled as standard bin-by-bin and is represented with a green line. The other curves represent the unfolding results.

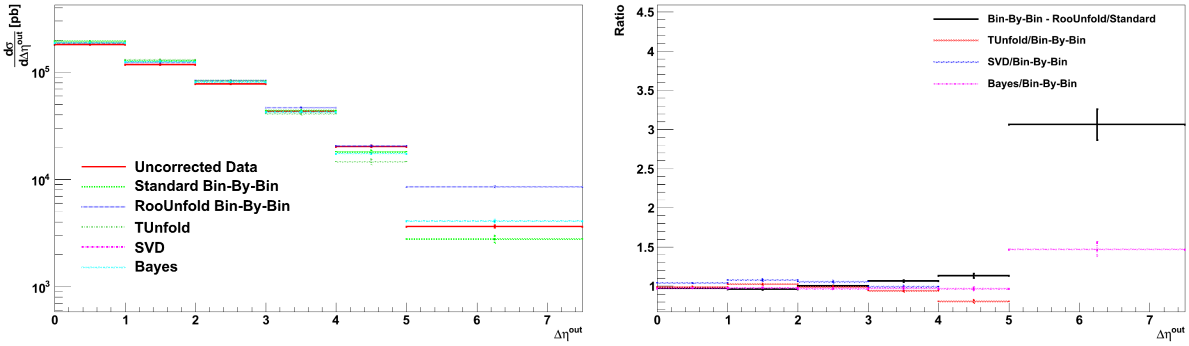


Figure 6.19: PYTHIA 6 Z2* tune on detector level unfolded with PYTHIA 8 4C tune response matrix for the $\Delta\eta^{out}$ for the JetMETTau_2010A dataset. Cross-section (left) and ratio (right). PYTHIA 6 - Z2* tune on detector level is represented with a red line, denoted as uncorrected data. The MC truth is labeled as standard bin-by-bin and is represented with a green line. The other curves represent the unfolding results.

Bayes method proved to give the best results without too much tuning of the method's parameters. It should have been possible to get good results with other methods too, but as the unfolding is used just as a cross–check it was not considered necessary.

6.4.4 Unfolded distributions

Taking into account the consistency and the closure test results, the measured data have been unfolded only with Bayes' method.

The unfolding result for $\Delta\phi$ in the inclusive scenario is presented in figure 6.20. The obtained cross-section is very similar to the bin–by–bin result for the higher $\Delta\phi$ region and deviates towards lower values, up to 0.9, for the lower $\Delta\phi$ region. Similar results were obtained for other datasets and MC generators.

The results of the $\Delta\phi$ unfolding in the inclusive scenario for different $\Delta\eta$ slices are presented in figures G.1 and G.2 in the appendix G. The results are very similar to the inclusive scenario, as well across datasets and MC predictions.

Figure 6.21 shows the unfolding result for $\Delta\phi$ in the inside–jet veto scenario. The difference between bin–by–bin and Bayes is smaller than in the inclusive scenario, around 2%, and is

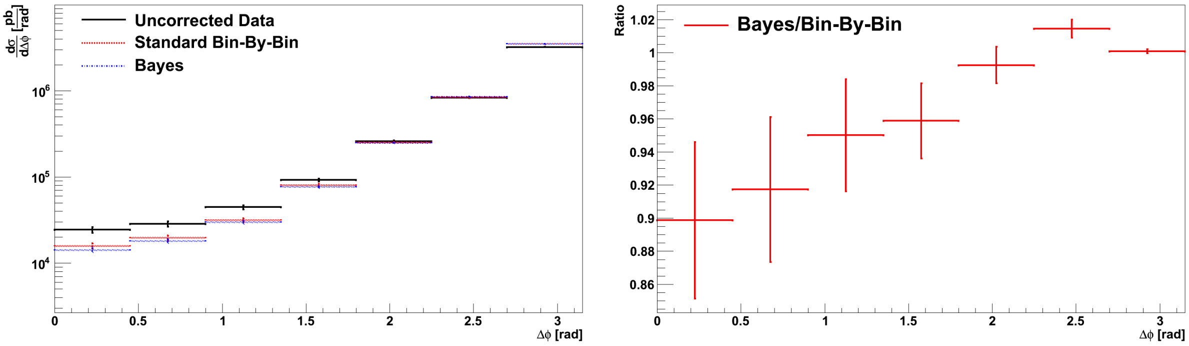


Figure 6.20: Data unfolded with Pythia 6 - Tune Z2*. Cross-section (left) and ratio (right) for $\Delta\phi$ in the inclusive scenario for the JetMETTau_2010A dataset.

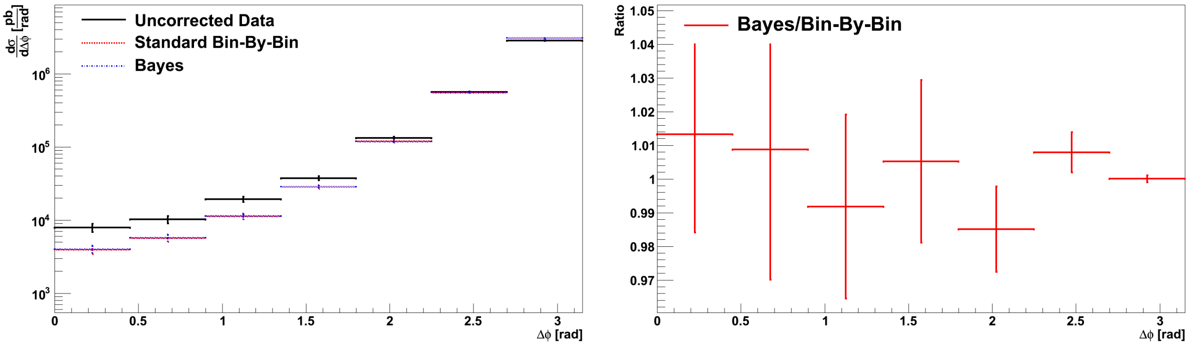


Figure 6.21: Data unfolded with Pythia 6 - Tune Z2* for $\Delta\phi$ in the inside–jet veto scenario. Cross-section (left) and ratio (right) for the JetMETTau_2010A dataset.

constant over the whole distribution. Similar results were found for the other MC predictions and datasets.

The results for $\Delta\phi$ in the inside–jet veto scenario in slices of $\Delta\eta$ are shown in figures G.3 and G.4 in the appendix G. No significant difference between bin–by–bin and Bayes' method is observed except for the very low $\Delta\phi$ region in the wider rapidity separation. The results are consistent for the other datasets and MC predictions.

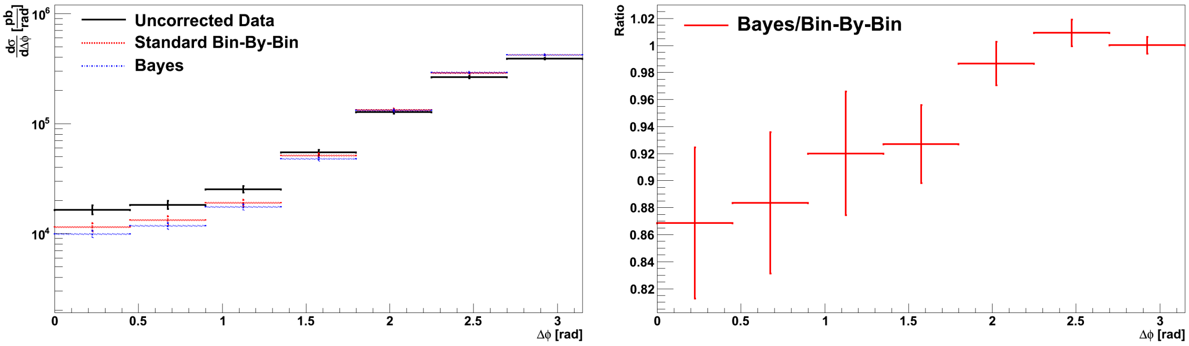


Figure 6.22: Data unfolded with Pythia 6 - Tune Z2* for $\Delta\phi$ in the inside–jet tag scenario for the JetMETTau_2010A dataset. Cross-section (left) and ratio (right).

Figure 6.22 presents the unfolding result for the inside–jet tag scenario. As expected, the ratio is very small in the higher $\Delta\phi$ region and goes down gradually as far as 0.87 in the lower $\Delta\phi$ region. Very similar results were obtained for other MC samples and datasets.

The results for $\Delta\phi$ in the inside–jet tag scenario in slices of $\Delta\eta$ are shown in figures G.5 and G.6 in the appendix G.Small discrepancies between Bayes' method and bin–by–bin, of the order of 10%, are found, especially in the lower $\Delta\phi$ region. The conclusions are the same for the other datasets and MC predictions.

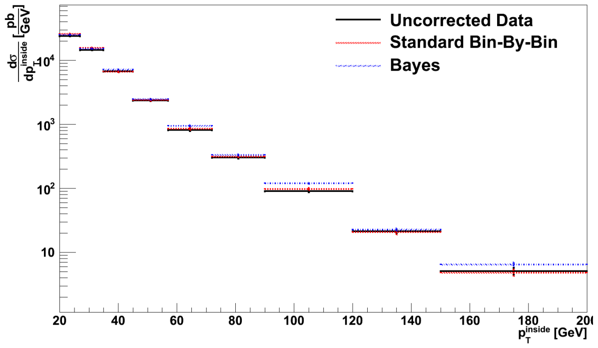


Figure 6.23: Data unfolded with PYTHIA 6 - TUNE Z2* for leading inside–jet p_T for the JetMETTau_2010A dataset. Cross-section (left) and ratio (right).

In figure 6.23 the unfolding results for the leading inside–jet p_T are presented. In the lower p_T region Bayes and bin–by–bin agree. For the higher p_T region the Bayes method yields a growing cross–section ratio with p_T, up to 40%. When using PYTHIA 8 the effect is the opposite, so the average still agrees with bin–by–bin. Similar results were found for both datasets.

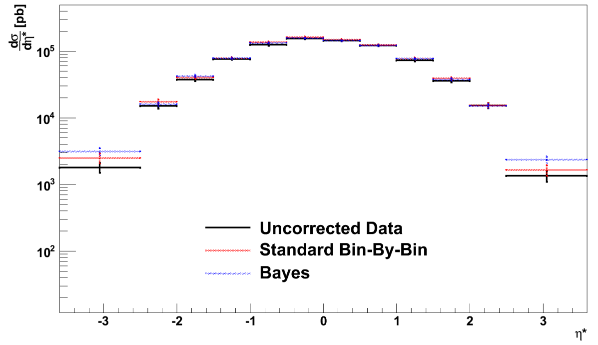


Figure 6.24: Data unfolded with PYTHIA 6 - TUNE Z2* for leading inside–jet η^* in the JetMETTau_2010A dataset. Cross-section (left) and ratio (right).

In figure 6.24 the unfolding results for the leading inside–jet η^* are shown. The results are compatible with bin-by-bin in the central region. At the edges, where the statistics is smaller, the results deviate up to 40%. The opposite trend had been found when using PYTHIA 8. The agreement is consistent between MC predictions and datasets.

In figure 6.25 the unfolding result for the leading outside–jet p_T is presented. The conclusions across the different datasets and MC predictions are the same as for the leading inside–jet.

In figure 6.26 the unfolding result for the leading $\Delta\eta^{out}$ is shown. Except for the last bin, the results agree between bin–by-.bin and Bayes method, up to few %. The opposite trend in the last bin is found for PYTHIA 8. The result is similar for both datasets.

In general the Bayes' method results agree very well with bin–by–bin. The unfolded data are averaged according to effective statistics, accepting equal contributions from both MC, and used as hadron–level corrected data for the final results.

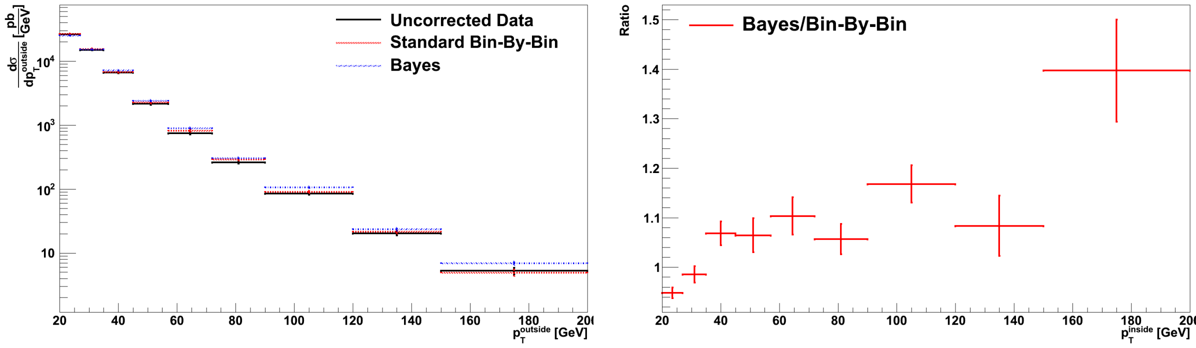


Figure 6.25: Data unfolded with Pythia 6 - Tune Z2* for leading outside–jet p_T for the JetMETTau_2010A dataset. Cross-section (left) and ratio (right).

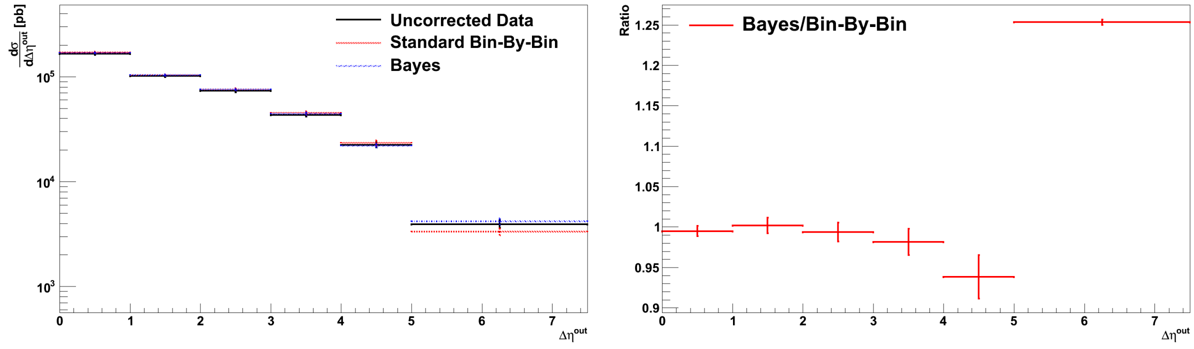


Figure 6.26: Data unfolded with Pythia 6 - Tune Z2* for leading $\Delta\eta^{out}$ in the JetMETTau_2010A dataset. Cross-section (left) and ratio (right).

Chapter 7

Uncertainties

In this chapter, the determination of the measurement uncertainties is discussed. In section 7.1, the statistical uncertainty due to the limited number of events and the statistical uncertainty coming from the unfolding procedure are compared. The correlated uncertanties, which include the uncertainties due to the jet energy scale, the luminosity and the trigger inefficiency, are presented in section 7.2. The uncorrelated uncertainties are described in section 7.3 and include statistical uncertainties and the uncertainties due to the model dependence and pile–up. The total experimental uncertainty, which combines all correlated and uncorrelated uncertainties, is presented in section 7.4. The y–range of all plots was choosen to be constant to provide a better comparison across distributions.

7.1 Statistical and Unfolding Uncertainties

When using the Bayes unfolding method, the statistical uncertainty from the propagated statistical uncertainties decreases with the correction procedure. This is easily understood because with each unfold iteration of the Bayes method more information is added to the distribution. However, the effect is regarded as controversial by many statisticians and heavily discussed without consensus [183] and, thus, it is important to estimate the its size. The unfolding uncertainty is used as statistical uncertainty.

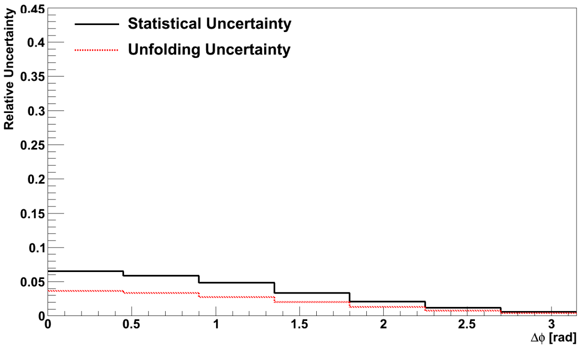


Figure 7.1: Relative statistical (black–solid) and unfolding (red–dashed) uncertainties of the $\Delta\phi$ distribution in the inclusive scenario.

In figure 7.1 the statistical and unfolding uncertainties of $\Delta\phi$ in the inclusive scenario are presented. In the following, the statistical uncertainty due to the limited number of events is called statistical uncertainty, while the statistical uncertainy coming from the unfolding is refered as unfolding uncertainty. The reason why the statistical uncertainty is reduced almost to half in some bins is due, as explained earlier, to the information added by the Bayes method iterations.

This reduction happens strongly in the lower $\Delta\phi$ region. A similar behaviour is observed for the $\Delta\eta$ slices, as shown in figure H.1 in the appendix H, with reductions between half and one third in the first bin.

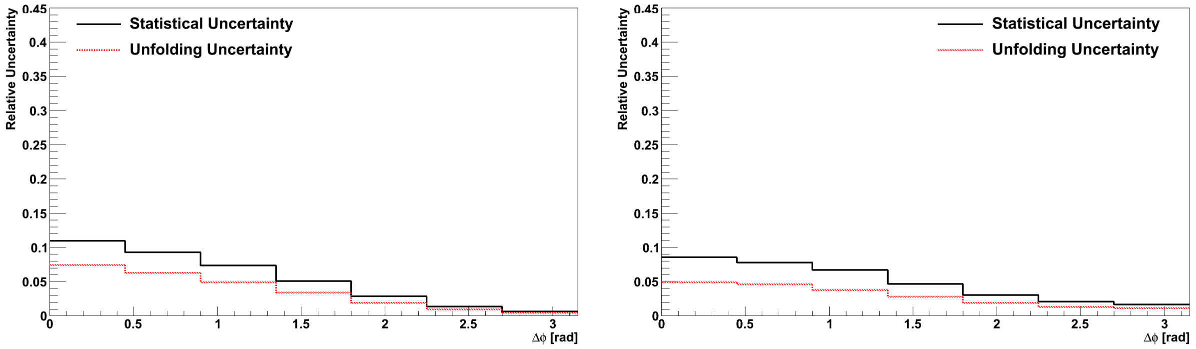


Figure 7.2: Relative statistical (black–solid) and unfolding (red–dashed) uncertainties of the $\Delta\phi$ distribution in the inside–jet veto (left) and inside–jet tag (right) scenarios.

In figure 7.2, the statistical before unfolding and unfolding uncertainties for $\Delta\phi$ in the inside–jet veto and inside–jet tag scenarios are shown. The inside–jet veto scenario shows a reduction of one third for the lower $\Delta\phi$ region and almost no reduction for the higher $\Delta\phi$ region. The inside–jet tag scenario shows a decrease of 50% between the statistical and unfolding errors for the lower $\Delta\phi$ region and a tiny decrease for the last bins. The same conclusions can be drawn for the $\Delta\eta$ slices, as presented in figures H.2 and H.3 in the appendix H, were the reductions are proportionally smaller for the larger pseudo–rapidity separations.

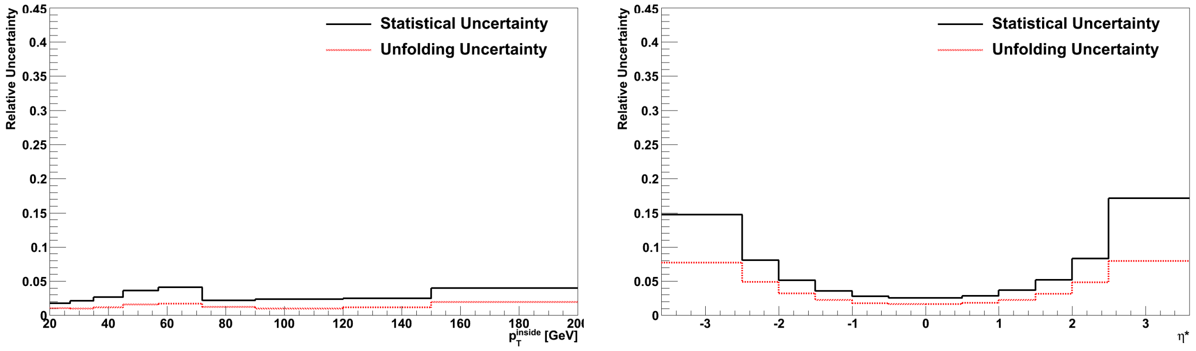


Figure 7.3: Relative statistical (black–solid) and unfolding (red–dashed) uncertainties of the leading inside–jet p_{T} (left) and η^* (right) distributions.

The statistical and unfolding uncertainties of the leading inside–jet p_{T} and η^* distributions are presented in figure 7.3. In the leading inside–jet p_{T} distribution, the unfolding smoothed out the structures at $p_{\mathrm{T}} = 80$ GeV, which were due to the transition between two different trigger streams (see subsection 5.1.2). In the leading inside–jet η^*, the Bayes method reduces the uncertainty in the central region by a small amount and in the edges up to 50%. Thus, the bins which suffer more from small statistics are the ones that experience a larger reduction.

In figure 7.4, the statistical and unfolding uncertainties for the leading outside–jet p_{T} and $\Delta\eta^{out}$ distributions are presented. The outside–jet p_{T} structures are smoothed out, creating a more uniform distribution and reducing the uncertainty to 50%. The statistical uncertainty grows with $\Delta\eta^{out}$ as does the reduction, which reduces as much as 50% of the initial uncertainty.

A summary of the statistical uncertainties is presented in table 7.1 and a summary of the unfolding uncertainties is shown in table 7.2. The average uncertainty is an unweighted uncer-

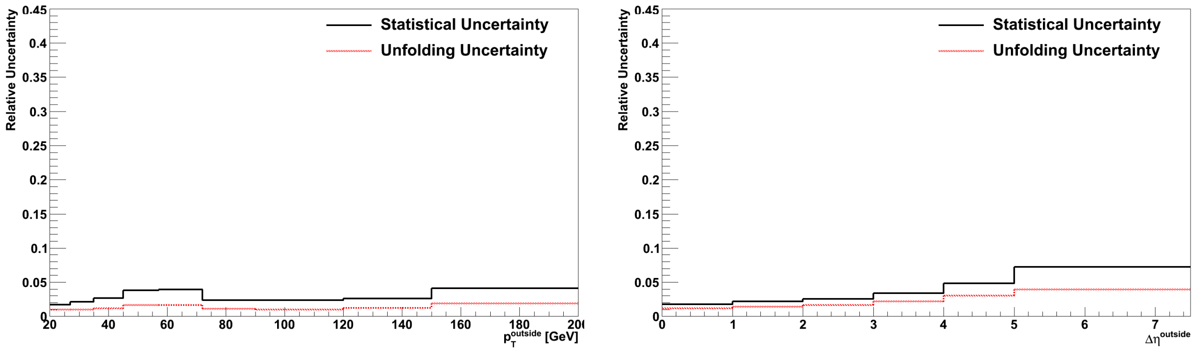


Figure 7.4: Relative statistical (black–solid) and unfolding (red–dashed) uncertainties of the leading outside–jet p_{T} (left) and $\Delta\eta^{out}$ (right) distributions.

tainty average between bins. The p_{T} distributions have the lowest uncertainty in both cases. As expected, the double differentially distributions, $\Delta\phi$ in slices of $\Delta\eta$, yield the largest uncertainties because of the small number of events. In some cases, the information added by the Bayes' unfolding method can reduce the statistical uncertainty up to 50%, specially where the statistical uncertainties are bigger.

7.2 Correlated Uncertainties

A correlated uncertainty is such that it affects all the data points to the same direction. The correlated uncertainties of this analysis are the jet energy scale uncertainty, luminosity uncertainty and trigger inefficiency uncertainty. To estimate the total correlated uncertainty, these uncertainties are added in quadrature.

In CMS, the luminosity is measured with an accuracy of 4% [184]. This has a direct impact on the normalisation of the measured cross-section and, therefore, it is added to the total uncertainty as a constant.

As discussed in section 5.1, the trigger combination is overall more than 99% efficient, thus a 1% in the upward direction is added to cover any potential inefficiencies. This addition is not an error estimation of the trigger uncertainty, but rather a conservative estimation of possible inefficiencies.

The dominant correlated uncertainty comes from the jet energy scale estimation. Since the p_{T} spectra are steeply falling distribution functions, most of the events are close to the p_{T}–threshold, which would turn a small calibration uncertainty into a major migration effect. The uncertainty of the jet energy scale calibration varies over the pseudorapidity in the detector: typical values for central jets are 2%, increasing to 6% for the most forward region. The effect on the measurement is estimated by varying the energy of all detector–level jets in the event by the corresponding uncertainty, both upwards and downwards, before applying the jet p_{T} selection.

The correlated uncertainty of the $\Delta\phi$ distribution in the inclusive scenario is presented in figure 7.5. A clear asymmetry between the downwards and upwards variation is visible, on average is -38% and + 48%, being larger in the upward direction. Trigger inefficiency and luminosity do not yield a significant contribution to the total correlated uncertainty. The total correlated uncertainty decreases with increasing $\Delta\phi$, because of the contributions on the lower p_{T} jet in the lower $\Delta\phi$ region. The higher jet energy scale uncertainty values in the lower $\Delta\phi$ region are due to the greater fraction of jets near the p_{T} threshold in these bins, which thus increase their migration probability.

In figures I.1 and I.2 in the appendix I, the correlated uncertainty of the $\Delta\phi$ distributionfor the inclusive scenario in slices of $\Delta\eta$ is presented. The values of the jet energy scale are larger than

Table 7.1: Relative statistical uncertainty

Distribution	Average	Minimum	Maximum
$\Delta\phi$ inclusive scenario	3.47%	0.58%	6.52%
$0.4 < \Delta\eta < 2.5$	5.91%	0.82%	11.01%
$2.5 < \Delta\eta < 3.5$	7.29%	1.18%	13.74%
$3.5 < \Delta\eta < 4.5$	8.06%	1.52%	15.56%
$4.5 < \Delta\eta < 7.5$	7.19%	1.76%	13.16%
$\Delta\phi$ inside–jet veto scenario	5.36%	0.61%	10.96%
$0.4 < \Delta\eta < 2.5$	7.23%	0.84%	14.27%
$2.5 < \Delta\eta < 3.5$	10.98 %	1.23%	21.50%
$3.5 < \Delta\eta < 4.5$	16.95%	1.62%	44.02%
$4.5 < \Delta\eta < 7.5$	16.32%	1.93%	34.50%
$\Delta\phi$ inside–jet tag scenario	4.90%	1.63%	8.54%
$0.4 < \Delta\eta < 2.5$	10.94%	3.09%	18.35%
$2.5 < \Delta\eta < 3.5$	10.39%	3.21%	18.41%
$3.5 < \Delta\eta < 4.5$	9.98%	3.40%	17.65%
$4.5 < \Delta\eta < 7.5$	8.50%	3.36%	14.86%
Leading inside–jet p_T	2.81%	1.79%	4.10%
Leading inside–jet η^*	6.38%	2.55%	17.14%
Leading outside–jet $\Delta\eta^{out}$	3.65%	1.74%	7.22%
Leading outside–jet p_T	2.85%	1.69%	4.08%

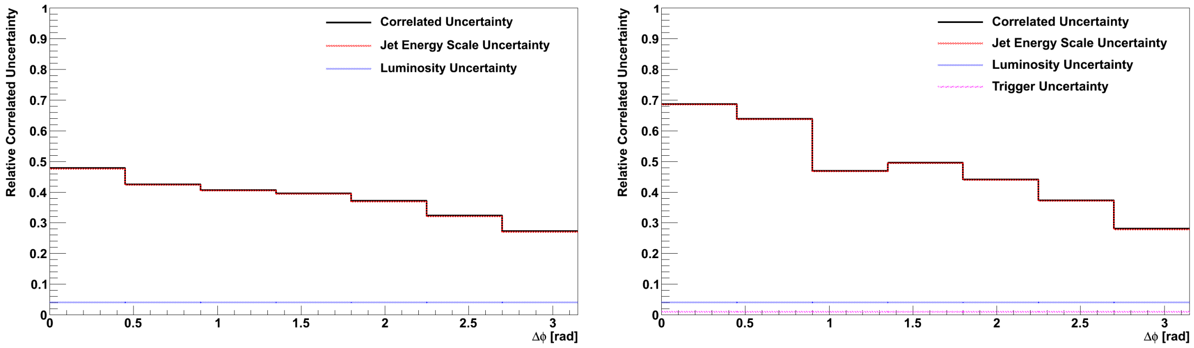


Figure 7.5: Relative correlated uncertainties of $\Delta\phi$ in the inclusive scenario in case of downward (left) and upward (right) variation.

in the single differential distribution, because the method is also sensitive to the low statistics, causing a larger total correlated uncertainty. The other features of the distributions are similar to the inclusive case.

The $\Delta\phi$ distribution of the correlated uncertainty for the inside–jet veto scenario is shown in figure 7.6. In the case of upward variation, there is a significant difference between the first two bins and the rest, the former have an uncertainty of about 80% and the latter about 40%. By comparison, the downward estimation is flatter and the uncertainty decreases with increasing $\Delta\phi$. The upwards uncertainty is larger than downwards, which was expected, as the migrations inside are bigger than the migration outside.

In figures I.3 and I.4 in appendix I, the $\Delta\phi$ correlated uncertainty for the inside–jet veto scenario in slices of $\Delta\eta$ is shown. There is a huge variation in the lower $\Delta\phi$ region for the upwards uncertainty which was caused by a lack of statistics of the estimation in this phase–space. The

Table 7.2: Relative unfolding uncertainty

Distribution	Average	Minimum	Maximum
$\Delta\phi$ inclusive scenario	2.01%	0.40%	3.64%
$0.4 < \Delta\eta < 2.5$	3.67%	0.57%	6.76%
$2.5 < \Delta\eta < 3.5$	4.73%	0.82%	8.98%
$3.5 < \Delta\eta < 4.5$	5.19%	1.06%	9.93%
$4.5 < \Delta\eta < 7.5$	4.63%	1.22%	8.41%
$\Delta\phi$ inside–jet veto scenario	3.58%	0.42%	7.42%
$0.4 < \Delta\eta < 2.5$	4.81%	0.59%	9.48%
$2.5 < \Delta\eta < 3.5$	7.54%	0.86%	15.15%
$3.5 < \Delta\eta < 4.5$	11.16%	1.13%	27.86%
$4.5 < \Delta\eta < 7.5$	12.05%	1.34%	27.36%
$\Delta\phi$ inside–jet tag scenario	2.89%	1.08%	4.90%
$0.4 < \Delta\eta < 2.5$	6.62%	2.08%	11.13%
$2.5 < \Delta\eta < 3.5$	6.69%	2.19%	12.03%
$3.5 < \Delta\eta < 4.5$	6.53%	2.32%	11.55%
$4.5 < \Delta\eta < 7.5$	5.50%	2.27%	9.53%
Leading inside–jet p_T	1.30%	0.98%	1.91%
Leading inside–jet η^*	3.59%	1.65%	7.96%
Leading outside–jet $\Delta\eta^{out}$	2.22%	1.16%	3.94%
Leading outside–jet p_T	1.29%	0.97%	1.89%

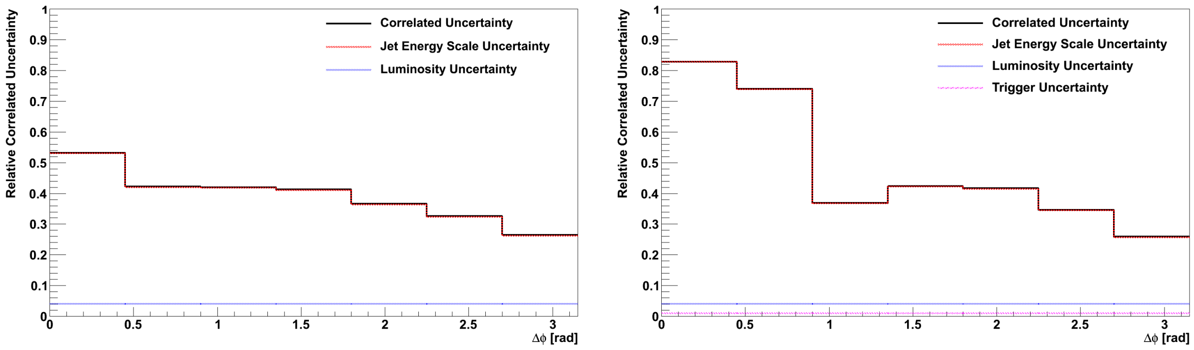


Figure 7.6: Relative correlated uncertainties of $\Delta\phi$ in the inside–jet veto scenarioin case of downward (left) and upward (right) variation.

other features observed in the distributions have a similar behaviour as the corresponding cases in the inclusive scenario.

The $\Delta\phi$ distribution of the correlated uncertainty for the inside–jet tag scenario is presented in figure 7.7. As in the previous scenarios, the upwards variation is larger than the downwards and structures due to statistical fluctuations appear. The difference between the bins is smaller for both upward and downward directions, being smaller than in the inside–jet veto scenario for the lower $\Delta\phi$ region and larger for the higher $\Delta\phi$ region. Even for the lower values of the jet energy scale, in the higher $\Delta\phi$ region, the other uncertainties do not contribute in a significant way to the total correlated uncertainty.

The $\Delta\phi$ correlated uncertainties for the inside–jet tag scenario in slices of $\Delta\eta$ are presented in figures I.5 and I.6 in the appendix I. The correlated uncertainty is much flatter than in any of the other scenarios. Otherwise, similar trends as in the corresponding distributions of the other scenarios are found.

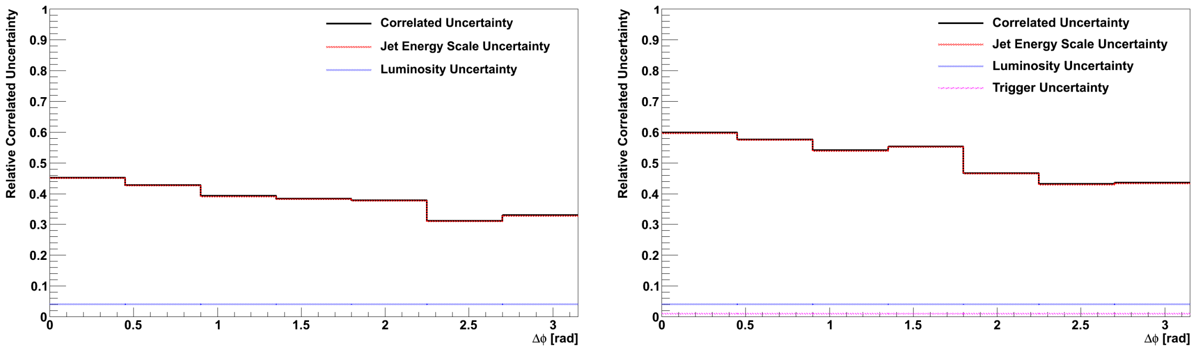


Figure 7.7: Relative correlated uncertainties of $\Delta\phi$ in the inside–jet tag scenario in case of downward (left) and upward (right) variation.

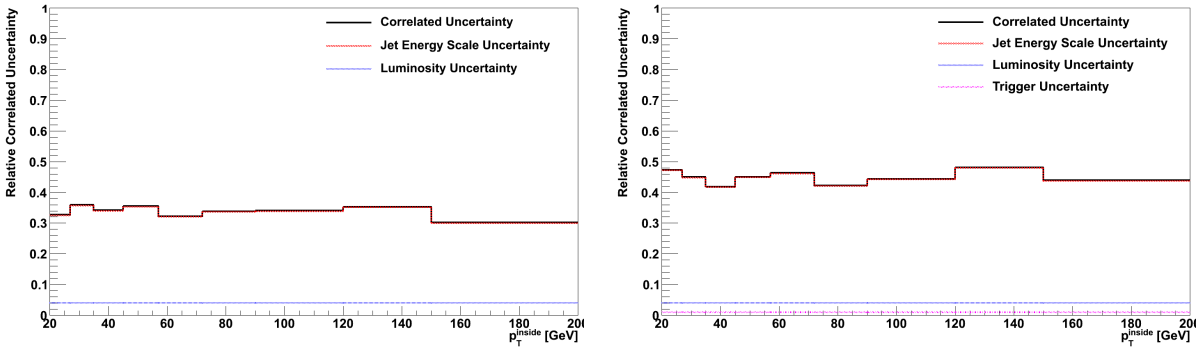


Figure 7.8: Relative correlated uncertainties for the leading inside–jet p_{T}. Downwards (left) and upwards (right) variation.

In figure 7.8, the correlated uncertainties for the leading inside–jet p_{T} are presented. The jet energy scale uncertainty is still the dominating uncertainty, displaying a variation of -30% +40%, and the contributions of the other uncertainties are negligible. The total correlated uncertainty is almost flat for both directions, without any evident structures, being slightly larger for the upward direction. The absolute values are, in unweighted average, smaller than for the $\Delta\phi$ distributions.

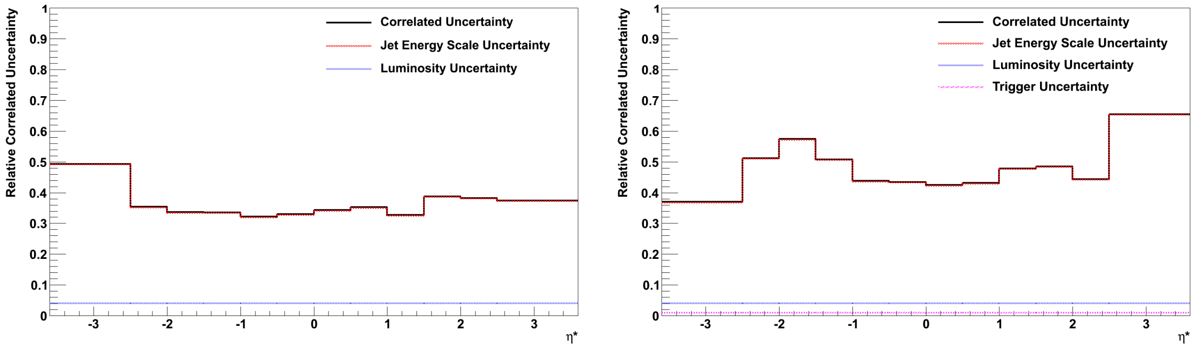


Figure 7.9: Relative correlated uncertainties for the leading inside–jet η^*. Downwards (left) and upwards (right) variation.

In figure 7.9, the correlated uncertainties for the leading inside–jet η^* are shown. Except for the increase in the first bin, the correlated uncertainty is almost flat in the downward direction.

One small peaky structure appear in the upwards variation, causing an increase up to 20% in the uncertainty of these bins. The fluctuations near the edges can be explained by the lack of statistics of the estimation, which are of the same order as statistical uncertainties (see figure 7.3–right).

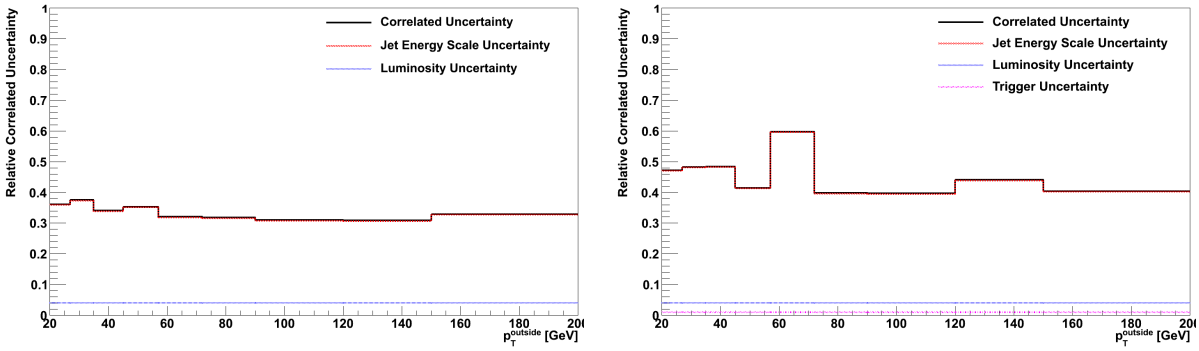


Figure 7.10: Relative correlated uncertainties for the leading outside–jet $p_{\rm T}$. Downwards (left) and upwards (right) variation.

The correlated uncertainties for the leading outside–jet $p_{\rm T}$ are presented in figure 7.10. Except for the difference in the fifth bin of the upwards distribution, which is due to a statistical fluctuation, the uncertainties are flat in both directions. The upwards direction is 10% higher than the downwards one.

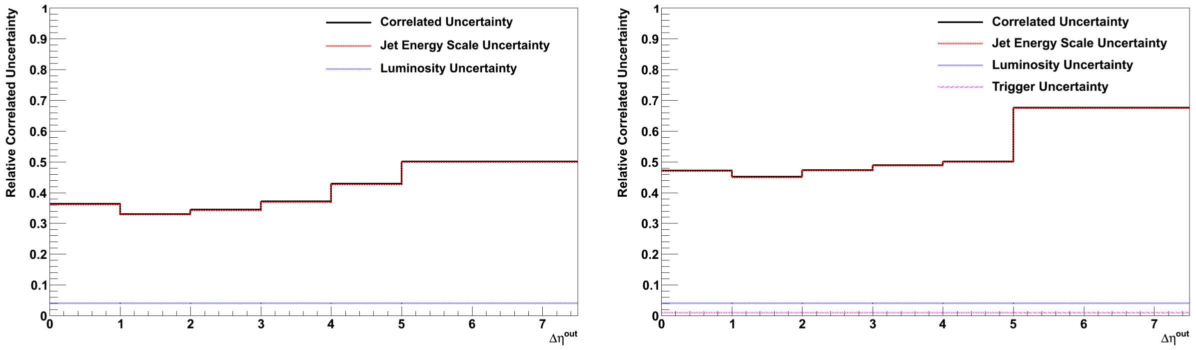


Figure 7.11: Relative correlated uncertainties for $\Delta\eta^{out}$. Downwards (left) and upwards (right) variation.

The correlated uncertainties for $\Delta\eta^{out}$ are shown in figure 7.11. The uncertainty grows steadily, around 15% in total, from the lower to the higher $\Delta\eta^{out}$ region. The last bin of the upwards variation has a value 10% higher than expected by the assumption of steady growth, which is just a statistical fluctutation of the method.

A summary of the jet energy scale uncertainty is presented in table 7.3 and a summary of the total correlated uncertainties is shown in table 7.4. The average is unweighted between bins. The jet energy scale uncertainty drives the total value of the correlated uncertainty in all distributions and all bins. The contributions of trigger inefficiency and luminosity are less than 0.5% of the total correlated uncertainty.

7.3 Uncorrelated Uncertainties

The uncorrelated uncertainties do not change the data points in a given direction but are independent for each data point. The uncorrelated uncertainties of this analysis are due to the

Table 7.3: Relative jet energy scale uncertainty

Distribution	Average	Minimum	Maximum
$\Delta\phi$ inclusive scenario	-37.94% +48.16%	-26.99% +27.80%	-47.60% +68.55%
$0.4 < \Delta\eta < 2.5$	-38.24% +51.28%	-26.06% +26.75%	-45.91% +83.59%
$2.5 < \Delta\eta < 3.5$	-35.59% +46.48%	-27.53% +27.78%	-44.81% +63.38%
$3.5 < \Delta\eta < 4.5$	-36.57% +48.12%	-26.99% +29.27%	-49.60% +70.74%
$4.5 < \Delta\eta < 7.5$	-41.41% +47.89%	-29.88% +30.48%	-52.68% +62.21%
$\Delta\phi$ inside–jet veto scenario	-38.95% +48.13%	-26.17% +25.66%	-53.00% +82.73%
$0.4 < \Delta\eta < 2.5$	-39.80% +53.43%	-25.39% +25.29%	-53.32% +101.85%
$2.5 < \Delta\eta < 3.5$	-35.12% +48.22%	-26.89% +25.55%	-41.93% +75.24%
$3.5 < \Delta\eta < 4.5$	-41.27% +63.14%	-19.16% +26.45%	-72.87% +165.46%
$4.5 < \Delta\eta < 7.5$	-43.72% +36.69%	-28.39% +22.78%	-62.01% +57.63%
$\Delta\phi$ inside–jet tag scenario	-38.03% +51.24%	-30.94% +42.93%	-45.00% +59.62%
$0.4 < \Delta\eta < 2.5$	-37.14% +52.33%	-28.49% +37.55%	-46.32% +66.80%
$2.5 < \Delta\eta < 3.5$	-36.05% +48.55%	-29.93% +40.11%	-47.51% +66.98%
$3.5 < \Delta\eta < 4.5$	-35.64% +50.23%	-26.56% +35.77%	-46.27% +61.02%
$4.5 < \Delta\eta < 7.5$	-41.15% +53.86%	-32.68% +40.41%	-49.72% +72.40%
Leading inside–jet p_T	-33.54% +44.71%	-30.00% +41.63%	-35.71% +47.93%
Leading inside–jet η^*	-35.94% +47.75%	-31.93% +36.71%	-49.16% +65.34%
Leading outside–jet $\Delta\eta^{out}$	-38.78% +50.84%	-32.82% +44.94%	-59.93% +67.37%
Leading outside–jet p_T	-33.28% +45.25%	-30.61% +39.42%	-37.29% +59.59%

pile–up modeling, model dependence and the statistical uncertainties due to the limited number of events. The total uncorrelated uncertainty is obtained by adding in quadrature all uncorrelated uncertainties. All the plots of this section have the same y–axis scale to make the comparisons across the distributions easier.

The estimation of the pile–up in data has been accomplished with an uncertainty much smaller than 1%, as seen from the values obtained with different MC (see section 4.3). As this uncertainty may have an effect on the unfolding, 1% was attributed to this effect.

The statistical and unfolding uncertainties have been discussed in section 7.1. For the total uncorrelated uncertainty, the unfolding uncertainty has been used as the statistical contribution.

The model uncertainty is estimated by comparing the correction given the detector simulation of two distint hadron–level predictions using a different physics model which, in pratice, is usually performed by using two different MC generators. In this analysis, the difference of the unfolding correction yielded by PYTHIA 6 and PYTHIA 8 were used as model dependence.

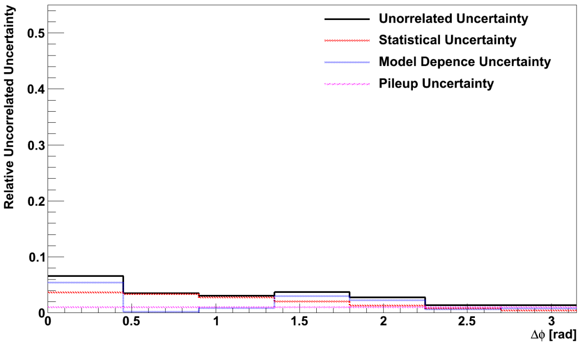


Figure 7.12: Relative uncorrelated uncertainties for $\Delta\phi$ distribution in the inclusive scenario.

Table 7.4: Relative total correlated uncertainty

Distribution	Average	Minimum	Maximum
$\Delta\phi$ inclusive scenario	-38.16% +48.35%	-27.29% +28.10%	-47.76% +68.67%
$0.4 < \Delta\eta < 2.5$	-38.45% +51.47%	-26.36% +27.06%	-46.09% +83.69%
$2.5 < \Delta\eta < 3.5$	-35.81% +46.67%	-27.81% +28.08%	-44.99% +63.52%
$3.5 < \Delta\eta < 4.5$	-36.79% +48.30%	-27.29% +29.56%	-49.76% +70.86%
$4.5 < \Delta\eta < 7.5$	-41.61% +48.08%	-30.14% +30.76%	-52.83% +62.34%
$\Delta\phi$ inside–jet veto scenario	-39.17% +48.34%	-26.47% +25.98%	-53.15% +82.84%
$0.4 < \Delta\eta < 2.5$	-40.02% +53.63%	-25.71% +25.63%	-53.47% +101.93%
$2.5 < \Delta\eta < 3.5$	-35.36% +48.42%	-27.19% +25.88%	-42.12% +75.36%
$3.5 < \Delta\eta < 4.5$	-41.50% +63.32%	-19.57% +26.77%	-72.98% +165.51%
$4.5 < \Delta\eta < 7.5$	-43.91% +36.94%	-28.67% +23.15%	-62.14% +57.78%
$\Delta\phi$ inside–jet tag scenario	-38.24% +51.41%	-31.20% +43.13%	-45.18% +59.76%
$0.4 < \Delta\eta < 2.5$	-37.36% +52.50%	-28.77% +37.77%	-46.49% +66.93%
$2.5 < \Delta\eta < 3.5$	-36.28% +48.73%	-30.20% +40.32%	-47.67% +67.10%
$3.5 < \Delta\eta < 4.5$	-35.87% +50.41%	-26.86% +36.01%	-46.45% +61.16%
$4.5 < \Delta\eta < 7.5$	-41.35% +54.02%	-33.22% +40.62%	-49.88% +72.52%
Leading inside–jet p_T	-33.78% +44.89%	-30.26% +41.83%	-35.93% +48.11%
Leading inside–jet η^*	-36.16% +47.93%	-32.18% +36.94%	-49.32% +65.47%
Leading outside–jet $\Delta\eta^{out}$	-38.99% +51.01%	-33.07% +45.13%	-50.09% +67.50%
Leading outside–jet p_T	-33.52% +45.44%	-30.88% +39.63%	-37.50% +59.73%

In figure 7.12 the uncorrelated uncertainties for the $\Delta\phi$ distribution in the inclusive scenario are presented. The total uncorrelated uncertainty grows towards the lower $\Delta\phi$ region, mainly driven by the model dependence contribution, except in the second and third bin where the statistical uncertainty dominates over the other uncertainties, because the model dependence is much smaller in those bins. The total uncorrelated uncertainty only exceeds 5% in the first bin.

In figure J.1 in the appendix J, the uncorrelated uncertainties of the $\Delta\phi$ distribution in the inclusive scenario as a function of $\Delta\eta$ are shown. The total uncorrelated uncertainty grows with increasing $\Delta\eta$ separation. Except for the wider pseudo–rapidity separation, where both statistical and model dependence have a similar magnitude, the uncertainty is dominated by the statistical uncertainty.

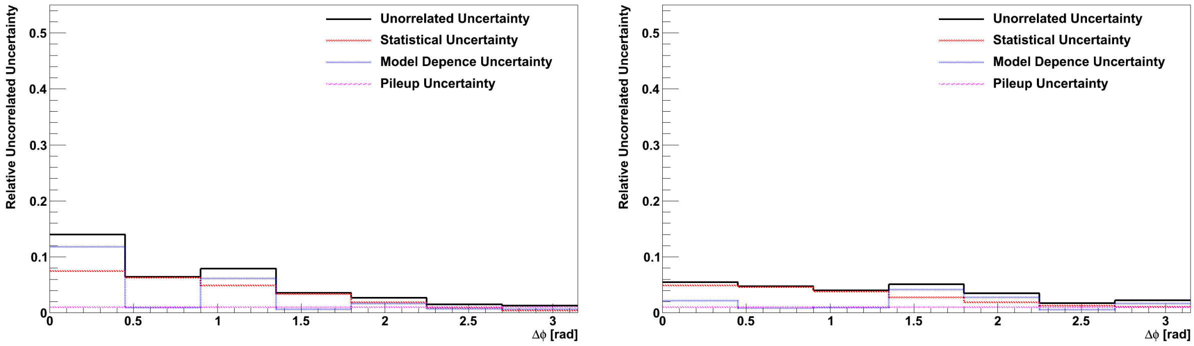


Figure 7.13: Relative uncorrelated uncertainties for $\Delta\phi$ distribution in the inside–jet veto (left) and inside–jet tag (right) scenarios.

The uncorrelated uncertainties for the $\Delta\phi$ distribution in the inside–jet veto and inside–jet tag scenarios are presented in figure 7.13. In the inside–jet veto scenario, the model dependence is the dominant uncertainty in the first and third bin, while statistical uncertainties dominates the second and forth; the remaining bins have similar values for all the contributions. For the inside–jet tag scenario, the statistical uncertainty dominates in the lower $\Delta\phi$ region and the uncertainties yield similar values for the higher $\Delta\phi$ region.

The uncorrelated uncertainties for the $\Delta\phi$ distribution in the inside–jet veto scenario in slices of $\Delta\eta$ are shown in figure J.2 in the appendix J. The uncorrelated uncertainties grow with increasing pseudo–rapidity separation. While the statistical uncertainty grows slowly, the model dependence grows quite fast yielding a large contribution in the lower $\Delta\phi$ region in the distribution with wider $\Delta\eta$ separation.

In figure J.3 in the appendix J, the uncorrelated uncertainties for $\Delta\phi$ in the inside–jet tag scenario in slices of $\Delta\eta$ are shown. The total uncorrelated uncertainty is almost constant in the $\Delta\eta$ slices. The statistical uncertainty dominates all the distributions in the lower $\Delta\phi$ region.

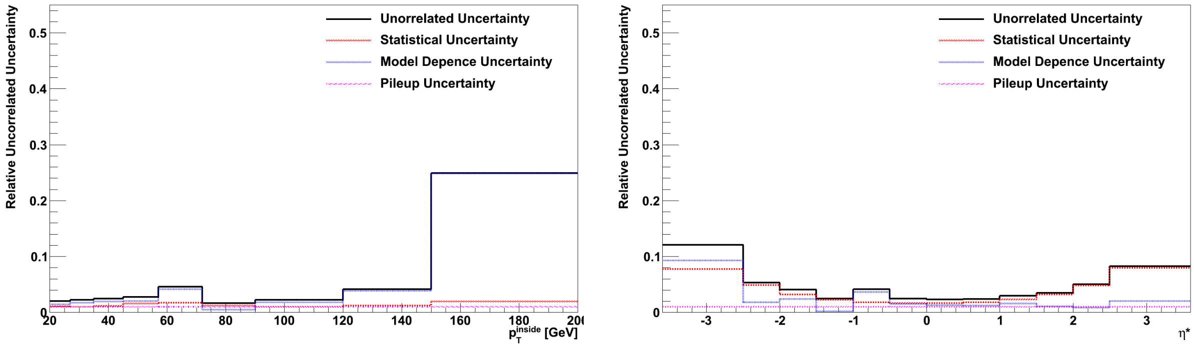


Figure 7.14: Relative uncorrelated uncertainties for the leading inside–jet p_T (left) and η^* (right) distributions.

In figure 7.14, the uncorrelated uncertainties for the leading inside–jet p_T and η^* are presented. The leading inside–jet p_T distribution is dominated by the model dependence uncertainty. With exception of the last bin, which is due to a missing p_T region in PYTHIA 8, the values of the total uncertainty are smaller than 5%. As expected, the statistical uncertainty for the leading inside–jet η^* distribution is symmetric, growing in both upwards and downwars directions with increasing $|\eta^*|$. The model dependence seems to be larger in the left side than in the right side, but that does not have a significant impact in the total uncorrelated uncertainty.

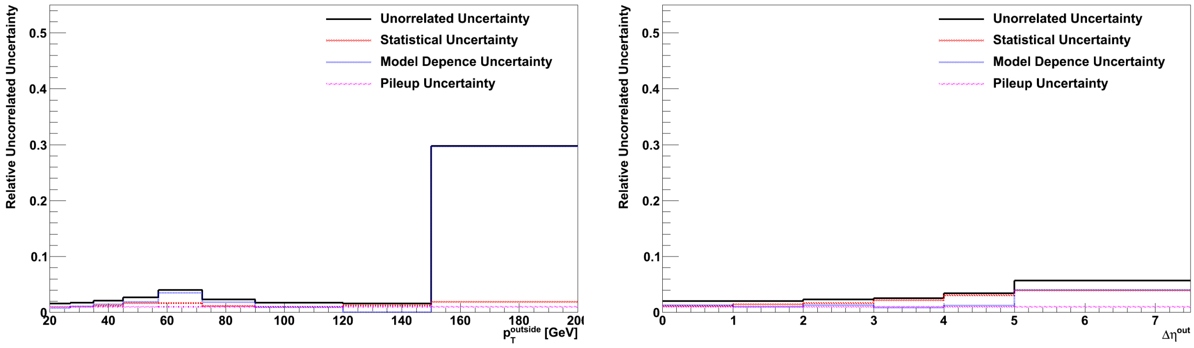


Figure 7.15: Relative uncorrelated uncertainties for the leading outside–jet p_T (left) and $\Delta\eta^{out}$ (right) distributions.

The uncorrelated uncertainties for leading outside–jet p_T and $\Delta\eta^{out}$ are presented in figure 7.15. The total uncorrelated uncertainty in the p_T distribution is smaller than 2% for most of the bins, having similar contributions from the three uncertainties, except for the last bin where the model dependence blows up the total uncertainty value, because it has a contribution of

30%. In the $\Delta\eta^{out}$ distribution, the total uncorrelated uncertainty grows with $\Delta\eta^{out}$ up to 6%. Both statistical and model dependence grow with $\Delta\eta^{out}$, having similar contributions to the total uncorrelated uncertainty.

In tables 7.5 and 7.6, respectively, the model dependence and total uncorrelated uncertainties are presented. The average is unweighted between bins. In some specific phase–spaces, the model dependence uncertainty dominates over the statistical. However, the statistical uncertainty drives the total, showing that has no significant dependence on the MC simulation used for the unfolding.

Table 7.5: Relative model dependence uncertainty

Distribution	Average	Minimum	Maximum
$\Delta\phi$ inclusive scenario	±1.85%	±0.17%	±5.38%
$0.4 < \Delta\eta < 2.5$	±2.81%	±0.65%	±4.88%
$2.5 < \Delta\eta < 3.5$	±1.86%	±0.60%	±4.74%
$3.5 < \Delta\eta < 4.5$	±1.81%	±0.58%	±3.93%
$4.5 < \Delta\eta < 7.5$	±3.66%	±0.28%	±9.14%
$\Delta\phi$ inside–jet veto scenario	±3.18%	±0.59%	±11.75%
$0.4 < \Delta\eta < 2.5$	±3.05%	±0.56%	±7.83%
$2.5 < \Delta\eta < 3.5$	±6.58%	±0.45%	±19.01%
$3.5 < \Delta\eta < 4.5$	±3.78%	±0.20%	±10.39%
$4.5 < \Delta\eta < 7.5$	±14.68%	±0.36%	±48.95%
$\Delta\phi$ inside–jet tag scenario	±1.85%	±0.56%	±4.17%
$0.4 < \Delta\eta < 2.5$	±3.20%	±0.95%	±6.53%
$2.5 < \Delta\eta < 3.5$	±1.28%	±0.41%	±3.43%
$3.5 < \Delta\eta < 4.5$	±3.37%	±0.65%	±6.43%
$4.5 < \Delta\eta < 7.5$	±3.08%	±0.28%	±7.20%
Leading inside–jet p_T	±4.68%	±0.48%	±24.82%
Leading inside–jet η^*	±2.21%	±0.16%	±9.24%
Leading outside–jet $\Delta\eta^{out}$	±1.59%	±0.83%	±3.97%
Leading outside–jet p_T	±4.55%	±0.05%	±29.69%

7.4 Total Uncertainty

The total uncertainty is obtained by adding all the correlated and uncorrelated uncertainties in quadrature. The correlated uncertainty is represented by a yellow band while the uncorrelated is displayed as error bars.

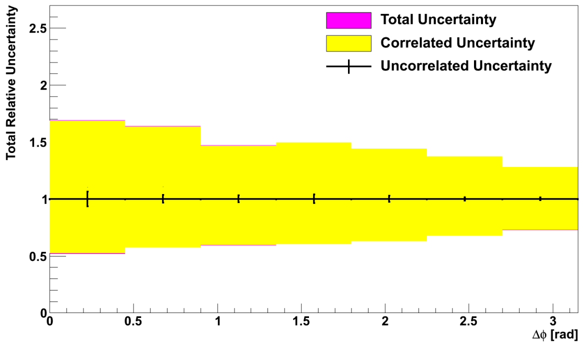


Figure 7.16: Relative total uncertainty of the $\Delta\phi$ distribution in the inclusive scenario.

Table 7.6: Relative total uncorrelated uncertainty

Distribution	Average	Minimum	Maximum
$\Delta\phi$ inclusive scenario	±3.18%	±1.34%	±6.57%
$0.4 < \Delta\eta < 2.5$	±4.87%	±1.32%	±8.40%
$2.5 < \Delta\eta < 3.5$	±5.33%	±1.59%	±9.63%
$3.5 < \Delta\eta < 4.5$	±5.76%	±1.80%	±10.73%
$4.5 < \Delta\eta < 7.5$	±6.16%	±1.86%	±12.46%
$\Delta\phi$ inside–jet veto scenario	±5.31%	±1.26%	±13.93%
$0.4 < \Delta\eta < 2.5$	±5.90%	±1.29%	±12.33%
$2.5 < \Delta\eta < 3.5$	±10.67%	±1.60%	±22.23%
$3.5 < \Delta\eta < 4.5$	±11.96%	±1.54%	±29.75%
$4.5 < \Delta\eta < 7.5$	±20.17%	±1.83%	±53.37%
$\Delta\phi$ inside–jet tag scenario	±3.81%	±1.72%	±5.45%
$0.4 < \Delta\eta < 2.5$	±7.71%	±2.77%	±11.22%
$2.5 < \Delta\eta < 3.5$	±6.96%	±2.56%	±12.10%
$3.5 < \Delta\eta < 4.5$	±7.73%	±3.06%	±13.25%
$4.5 < \Delta\eta < 7.5$	±6.51%	±2.80%	±11.98%
Leading inside–jet p_T	±5.22%	±1.65%	±24.91%
Leading inside–jet η^*	±4.57%	±2.33%	±12.07%
Leading outside–jet $\Delta\eta^{out}$	±2.98%	±1.98%	±5.68%
Leading outside–jet p_T	±5.26%	±1.56%	±29.76%

The total uncertainty for the $\Delta\phi$ distribution in the inclusive scenario is presented in figure 7.16. The main contribution comes from the correlated uncertainties, while the contribution of the uncorrelated uncertainties are almost negligible. The average values are 38% for the downward and 48% for the upward variation.

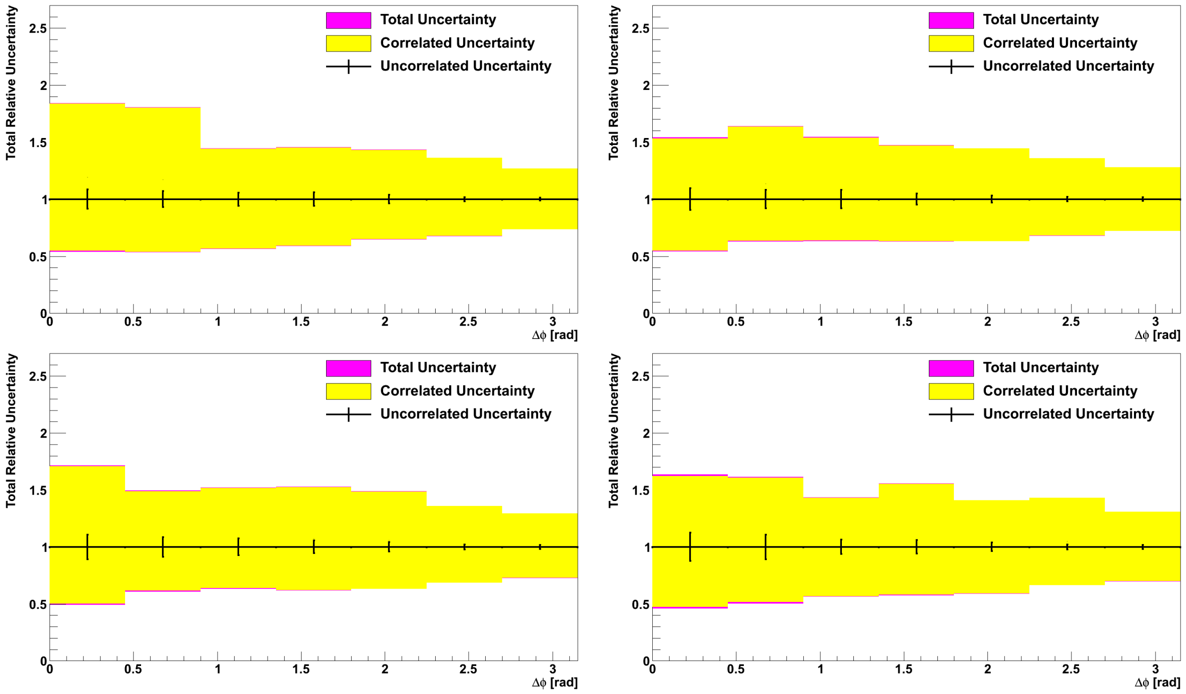


Figure 7.17: Relative total uncertainty of the $\Delta\phi$ in the inclusive scenario in slices of $\Delta\eta$: $0.4 > \Delta\eta \geq 2.5$ (top-left), $2.5 > \Delta\eta \geq 3.5$ (top-right), $3.5 > \Delta\eta \geq 4.5$ (bottom-left) and $4.5 > \Delta\eta \geq 7.5$ (bottom-right).

In figure 7.17, the total uncertainty of the $\Delta\phi$ distribution in the inclusive scenario in slices of $\Delta\eta$ is presented. The total uncertainty is dominated by the correlated uncertainties and the small contribution from the uncorrelated uncertainties is almost only visible in the lower $\Delta\phi$ region.

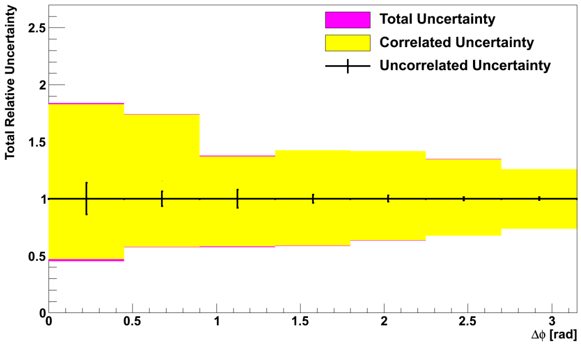


Figure 7.18: Relative total uncertainty of the $\Delta\phi$ distribution in the inside–jet veto scenario.

The total uncertainty of the $\Delta\phi$ distribution in the inside–jet veto scenario is shown in figure 7.18. Appart from the increased asymmetry between the downwards and upwards directions, the conclusions are analogue as for the inclusive scenario.

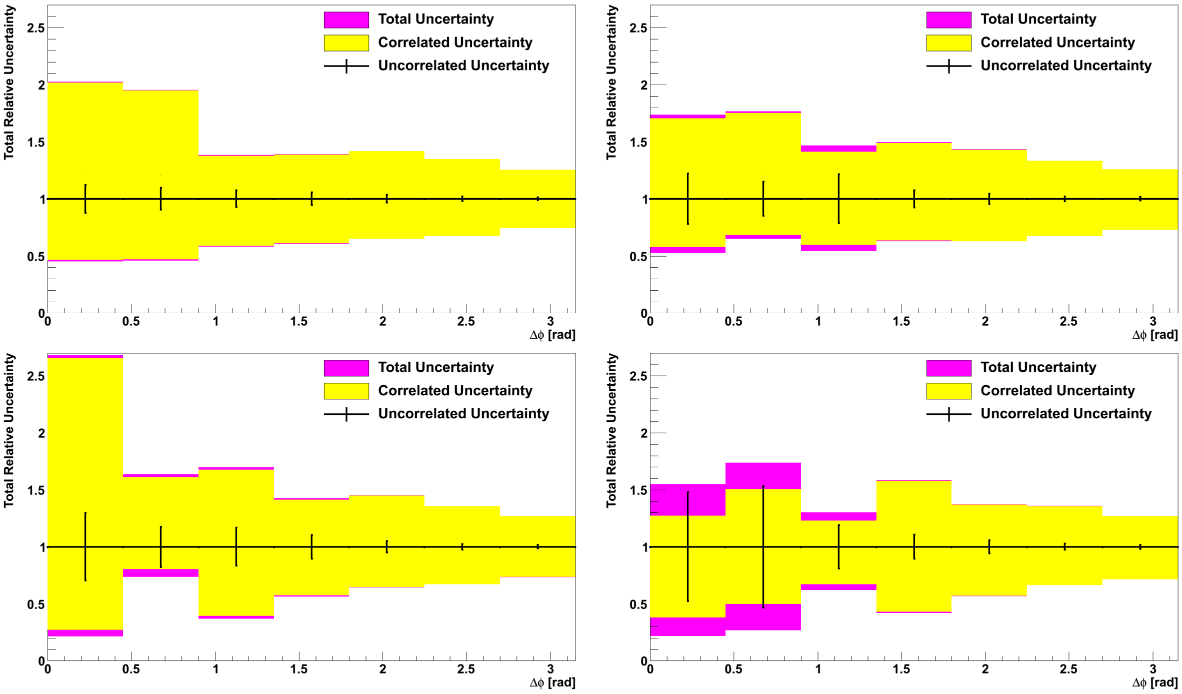


Figure 7.19: Relative total uncertainty of the $\Delta\phi$ distributions in the inside–jet veto scenario as function of $\Delta\eta$: $0.4 > \Delta\eta \geq 2.5$ (top-left), $2.5 > \Delta\eta \geq 3.5$ (top-right), $3.5 > \Delta\eta \geq 4.5$ (bottom-left) and $4.5 > \Delta\eta \geq 7.5$ (bottom-right).

In figure 7.19, the total uncertainty of the $\Delta\phi$ distribution in the inside–jet veto scenario is shown in slices of $\Delta\eta$. The total correlated uncertainty in the upwards direction is bigger in the first bin when compared with the second of the $2.5 > \Delta\eta \geq 3.5$ distribution, this is due to the larghe jet energy scale uncertainty in this bin, which is caused by low statistics used in the estimation. The total uncorrelated uncertainty is quite large for the first two bins of the $4.5 > \Delta\eta \geq 7.5$ distribution, mainly because of the model dependence uncertainty.

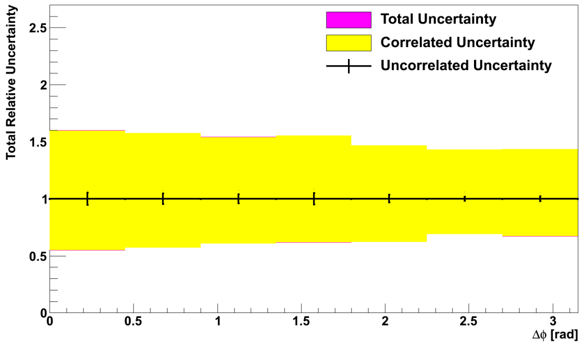


Figure 7.20: Relative total uncertainty of the $\Delta\phi$ distribution in the inside–jet tag scenario.

In figure 7.20, the total uncertainty of the $\Delta\phi$ distribution in the inside–jet tag scenario is shown. The distribution is flatter and more symmetric than in the previous scenarios, having almost no contribution from the uncorrelated uncertainties. This is due to a more uniform distribution of the statistics between the bins and better agreement of the different models used for the unfolding.

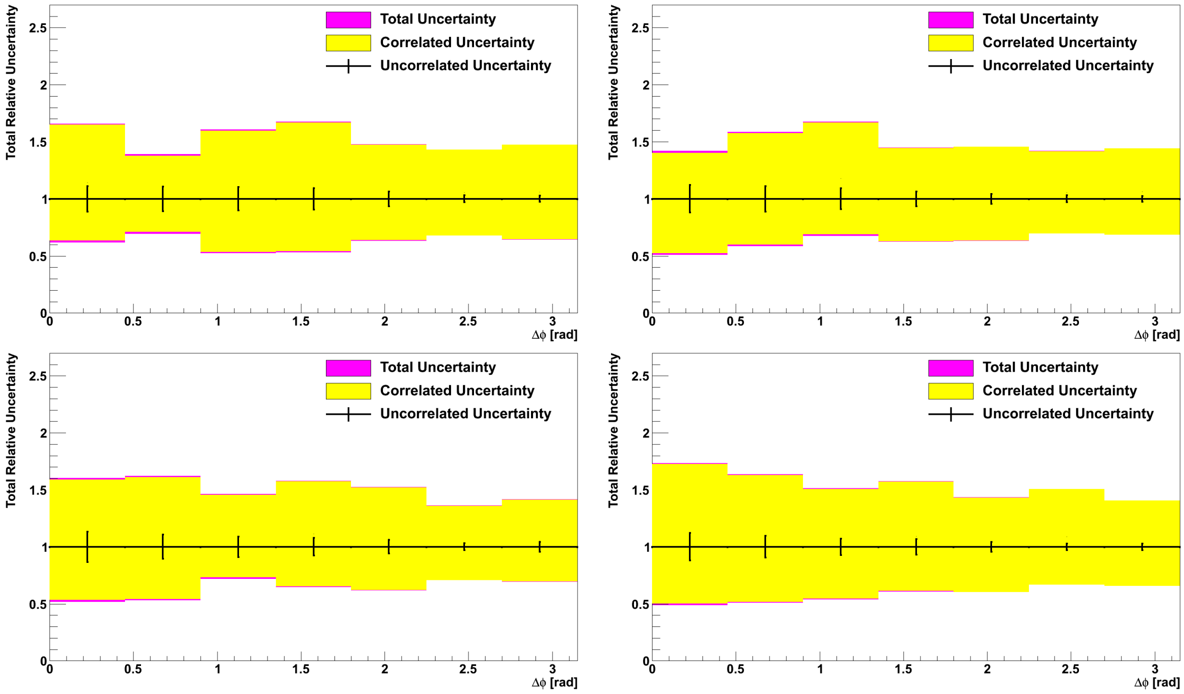


Figure 7.21: Relative total uncertainty of the $\Delta\phi$ distributions in the inside–jet tag scenario as function of $\Delta\eta$: $0.4 > \Delta\eta \geq 2.5$ (top-left), $2.5 > \Delta\eta \geq 3.5$ (top-right), $3.5 > \Delta\eta \geq 4.5$ (bottom-left) and $4.5 > \Delta\eta \geq 7.5$ (bottom-right).

In figure 7.21, the total uncertainty of the $\Delta\phi$ distribution in the inside–jet veto scenario in slices of $\Delta\eta$ is presented. The contributions from the uncorrelated uncertainties are almost negligible. There is no big difference between the different $\Delta\eta$ slices. The asymmetry between the lower and higher $\Delta\phi$ region is less pronounced. The effects of the upward and downward variations are less asymmetric than in the other cases.

In figure 7.22, the total uncertainty for inside–jet p_{T} and η^{*} is presented. The total uncertainty for the inside–jet p_{T} is flat with p_{T} with correlated uncertainties dominating the overall uncertainty. The last bin of p_{T} is an exception, having a larger uncorrelated uncertainty value

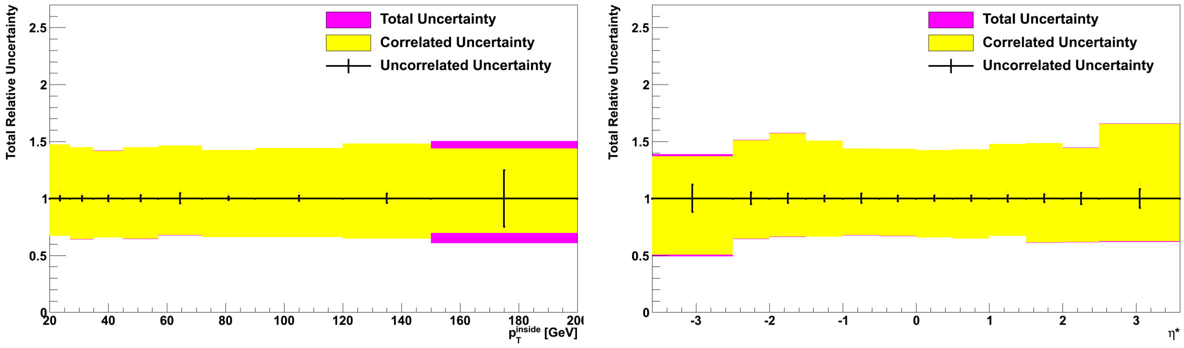


Figure 7.22: Relative total uncertainty of the inside–jet p_{T} (left) and η^* (right) distributions.

due to the model dependence, which yields a significant contribution to the total uncertainty. The inside–jet η^* total uncertainty is dominated by the correlated uncertainties.

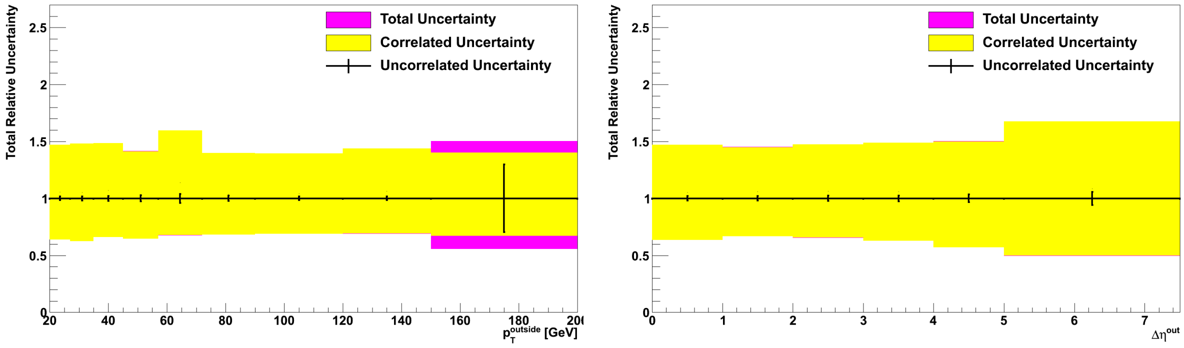


Figure 7.23: Relative total uncertainty of the outside–jet p_{T} (left) and $\Delta\eta^{out}$ (right) distributions.

In figure 7.23, the total uncertainty for the outside–jet p_{T} and $\Delta\eta^{out}$ is displayed. Except for the presence of an excess caused by statistical fluctuations in the upward direction of the fifth bin, $57 < p_{\mathrm{T}} < 72$ GeV, for the outside–jet p_{T} distribution, the conclusions are the same as for the inside–jet p_{T} distribution. The $\Delta\eta^{out}$ total uncertainty is dominated by the correlated uncertainties.

The summary of the total uncertainties is presented in table 7.7. Except for the few cases discussed earlier, the jet energy scale uncertainty dominate the total uncertainty of the analysis. The typical values are -40% in the downwards and +50% in the upwards variation.

Table 7.7: Relative total uncertainty

Distribution	Average	Minimum	Maximum
$\Delta\phi$ inclusive scenario	-38.31% +48.47%	-27.32% +28.13%	-48-21% +68.98%
$0.4 < \Delta\eta < 2.5$	-38.80% +51.72%	-26.40% +27.10%	-46.63% +84.11%
$2.5 < \Delta\eta < 3.5$	-36.30% +47.02%	-27.86% +28.13%	-46.01% +64.04%
$3.5 < \Delta\eta < 4.5$	-37.31% +48.69%	-27.35% +29.61%	-50.91% +71.66%
$4.5 < \Delta\eta < 7.5$	-42.15% +48.54%	-30.20% +30.81%	-54.28% +63.58%
$\Delta\phi$ inside–jet veto scenario	-39.64% +48.70%	-26.50% +26.01%	-54.95% +84.00%
$0.4 < \Delta\eta < 2.5$	-40.52% +53.99%	-25.73% +25.66%	-54.87% +102.68%
$2.5 < \Delta\eta < 3.5$	-37.55% +50.00%	-27.24% +25.93%	-47.62% +76.83%
$3.5 < \Delta\eta < 4.5$	-43.88% +64.59%	-26.37% +26.80%	-78.81% +168.17%
$4.5 < \Delta\eta < 7.5$	-50.48% +45.38%	-28.72% +27.01%	-78.36% +73.72%
$\Delta\phi$ inside–jet tag scenario	-38.44% +51.56%	-31.24% +43.17%	-45.50% +60.01%
$0.4 < \Delta\eta < 2.5$	-38.30% +53.16%	-30.74% +39-29%	-47.63% +67.58%
$2.5 < \Delta\eta < 3.5$	-37.05% +49.34%	-30.35% +41-91%	-49.18% +57.72%
$3.5 < \Delta\eta < 4.5$	-36.78% +51.05%	-28.35% +36.14%	-48.30% +62.08%
$4.5 < \Delta\eta < 7.5$	-41.92% +54.46%	-33.34% +40.72%	-51.30% +73.51%
Leading inside–jet p_T	-34.88% +45.71%	-32.59% +41.90%	-39.20% +50.48%
Leading inside–jet η^*	-36.51% +48.25%	-32.45% +38.87%	-50.78% +65-99%
Leading outside–jet $\Delta\eta^{out}$	-39.11% +51.10%	-33.12% +45.17%	-50.41% +67.74%
Leading outside–jet p_T	-34.87% +46.57%	-30.91% +39.66%	-44.39% +59.86%

Chapter 8

Results

In this chapter, the final results of the analysis are presented in the form of fully unfolded cross–sections compared to stable particle–level MC predictions. The $\Delta\phi$ distributions for the inclusive scenario are shown in section 8.1. The distributions for the inside–jet veto scenario are presented in section 8.2 and the results for the inside–jet tag scenario are displayed in section 8.3. The cross–sections for the outside–jet tag scenario are shown in section 8.4. A summary of the scenarios is presented in section 8.5.

In all plots of this chapter, the data are presented as dots with the error bars indicating the uncorrelated uncertainties and the yellow band the correlated uncertainties added in quadrature. The statistical uncertainty on the MC is represented by the error bars connected to the lines. All cross–sections are normalized to bin width.

8.1 Inclusive scenario

The results of the inclusive measurement are shown in figures 8.1, 8.2, 8.3, as well as in K.1 and K.2 in the appendix K.

In figure 8.1 the cross–section is presented as a function of the azimuthal difference between the forward and the central jet, $\Delta\phi$. The curve is a steeply falling distribution, spanning over 2.5 orders of magnitude. In the top–left and bottom–left plots, the data are compared to PYTHIA 6 predictions with different tunes, as well as to a PYTHIA 6 prediction performed without MPI. PYTHIA 6 predicts events with jets which are somewhat more correlated than in the data, and no large difference between the various PYTHIA 6 tunes is observed. However, overall the data are described by PYTHIA 6 when considering the fairly large systematic uncertainties. The exclusion of the contribution from MPI results in a larger correlation between the forward and central jets, since the radiation from MPI mixed with the hard scattering, its absense results in a worse agreement. In the right–top and right–bottom plots in figure 8.1, the data are compared to different MC generators. HERWIG++ describes the shape and normalisation very well. HERWIG 6.5 predicts the correct slope and slightly underestimates data, while PYTHIA 8 predicts a slightly steeper $\Delta\phi$ distribution. The best description is provided by HERWIG++.

When presenting the $\Delta\phi$ cross–sections differentially in the pseudorapidity separation between the forward and central jets, $\Delta\eta$, the $\Delta\phi$ distributions of the jets tend to flatten out with increasing $\Delta\eta$. This is shown in figures 8.2, 8.3, K.1 and K.2. Such behavior is expected from kinematics, because a large separation between the jets allows more QCD radiation between them, leading to a larger decorrelation between the forward and central jet. In figure 8.2 the data are compared to PYTHIA 6 predictions with different tunes and without MPI. As for the $\Delta\eta$-integrated measurement, there is no striking difference between the tunes. The contribution from MPI is independent of $\Delta\eta$. The deviation between data and MC is of about the same order as the inclusive measurement and, when considering the correlated uncertainties, the data are

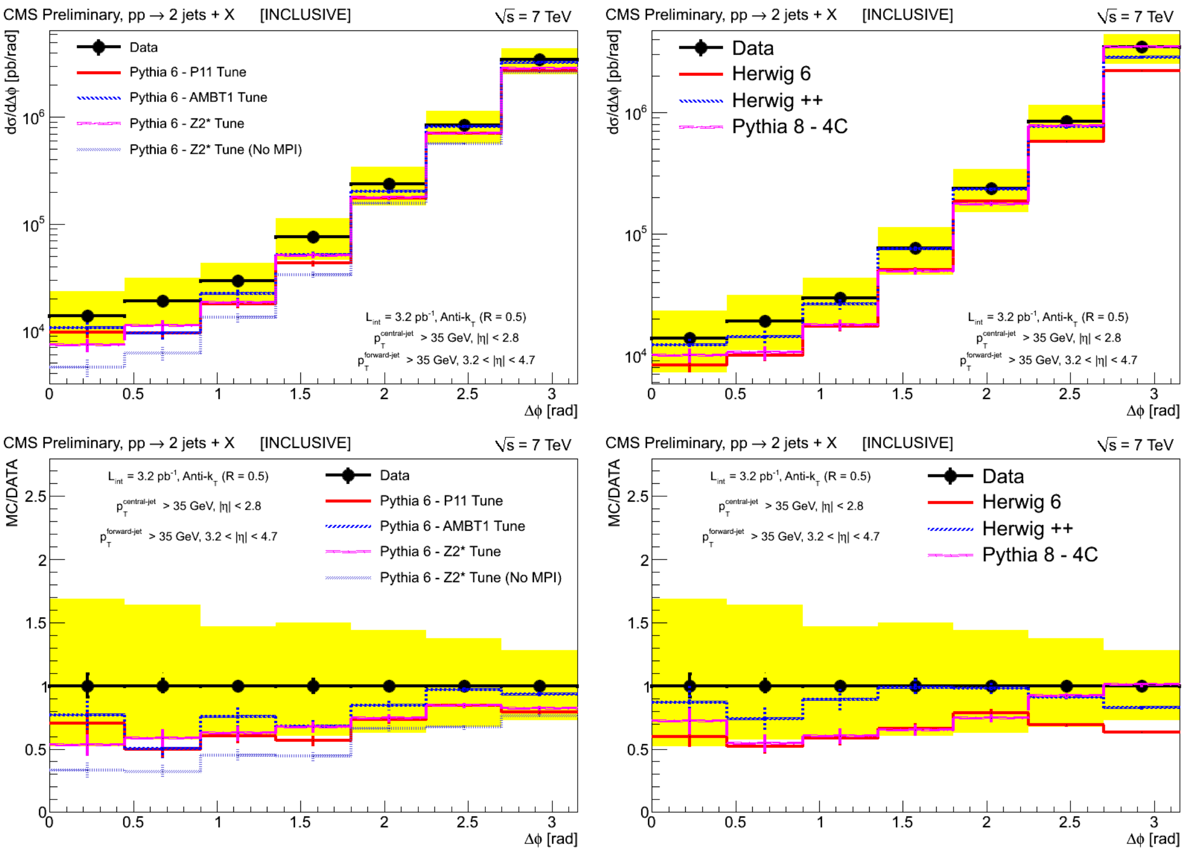


Figure 8.1: Cross–section in bins of $\Delta\phi$ (top) and ratios between MC and data (bottom) in the inclusive scenario. Data are compared to Pythia 6 predictions with different tunes and without MPI (left), and to various MC generators (right).

described by Pythia 6. When comparing the inclusive double-differential measurement to the different MC generators, in figure 8.3, one again sees that Herwig++ describes the data very well. With Herwig++, the shape is slightly better reconstructed than the other generators at low pseudo–rapidity separation, $\Delta\eta$, as can be seen in figure 8.3–bottom. In general, there are no large differences between the inclusive and the double-differential measurement in terms of MC–data comparison.

8.2 Inside–jet veto scenario

In figures 8.4, 8.5, 8.6, and in K.3 and K.4 in the appendix K, the azimuthal jet correlations for the inside–jet veto scenario are shown, in which events with inside–jets above $p_T = 20$ GeV between the forward and the central jets have been rejected.

The azimuthal correlation in the inside–jet veto scenario is presented in figure 8.4. The curve is steeper than in the inclusive scenario, spannning three orders of magnitude. The PYTHIA 6 predictions are further away from the data, predicting a too strong correlation, and the prediction without MPI is even further away, having a difference of up to one order of magnitude for the lower $\Delta\phi$ region. On the other hand, the PYTHIA 6 tunes agree with each other within the statistical uncertainties. Both HERWIG 6 and PYTHIA 8 yield a better description of the data than PYTHIA 6, but HERWIG++ provides the best overall description. HERWIG 6 underestimates the total cross-section and PYTHIA 8 predicts a too steep slope.

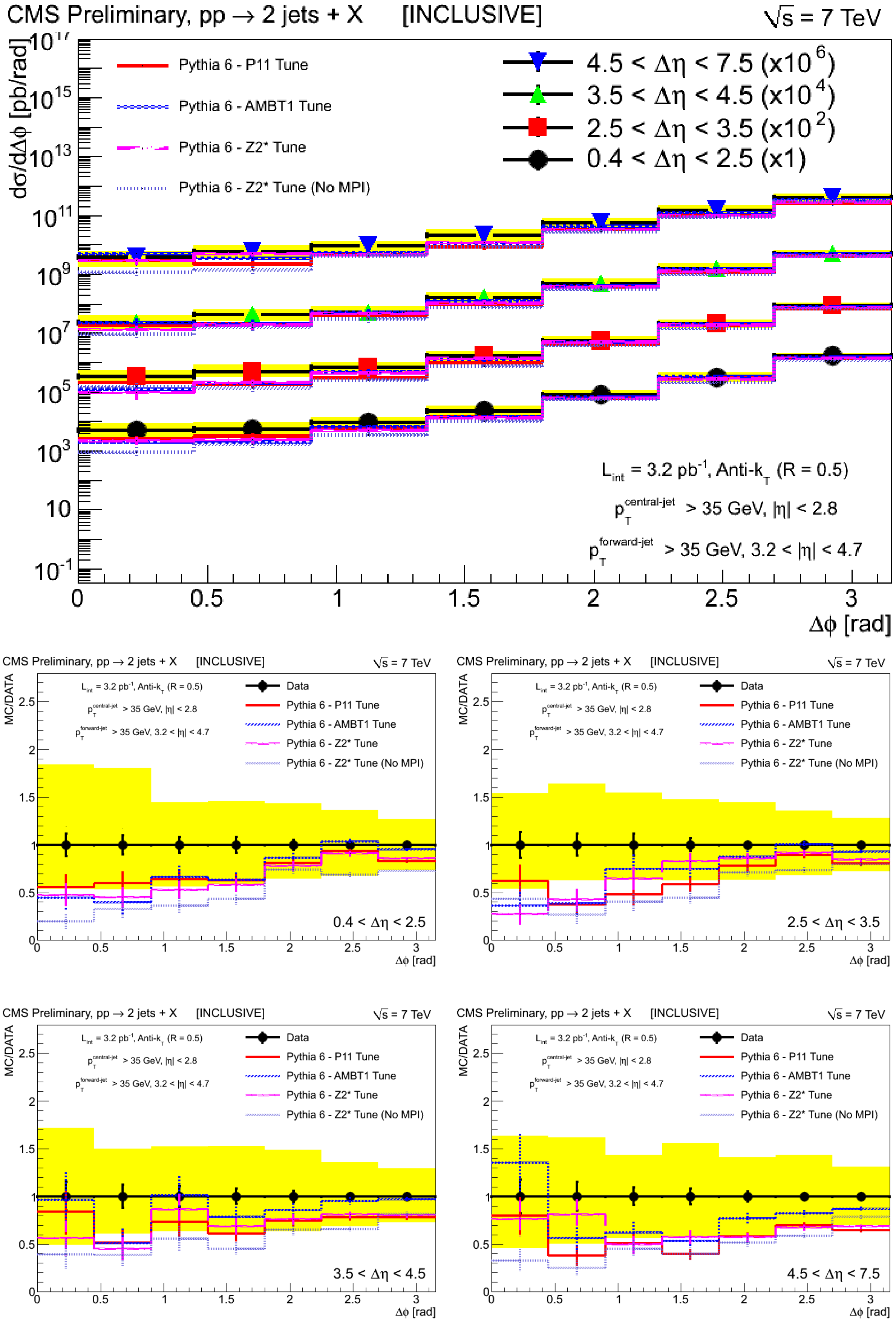


Figure 8.2: Double–differential cross–section for the inclusive scenario in bins of the difference in azimuthal angle between the forward and the central jet, $\Delta\phi$, in bins of the pseudorapidity separation between the two jets (top), $\Delta\eta$, and the ratio between MC and data (bottom). Data are compared to Pythia6.

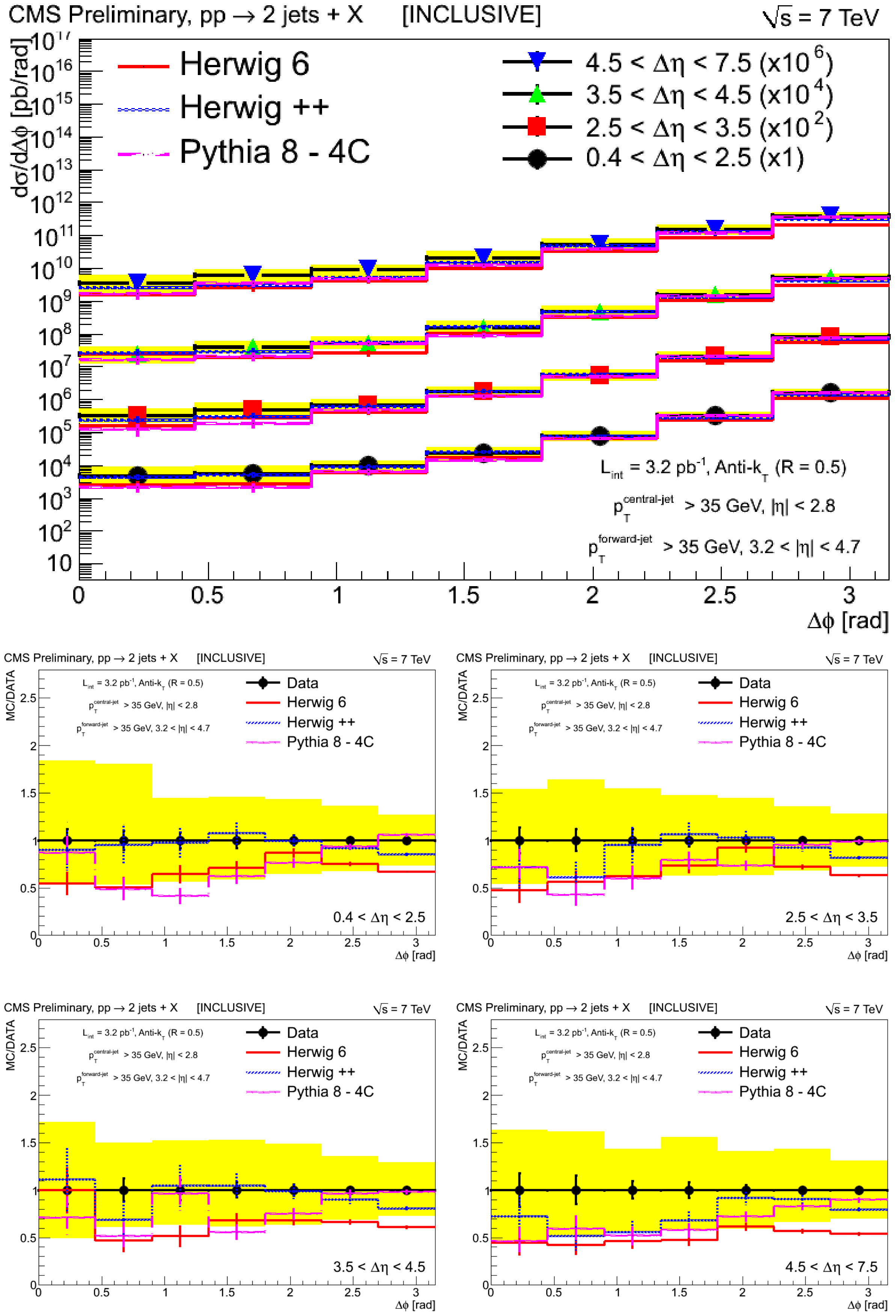


Figure 8.3: Double–differential cross–section for the inclusive scenario in bins of the difference in azimuthal angle between the forward and the central jet, $\Delta\phi$, as a function of the pseudorapidity separation between the two jets (top), $\Delta\eta$, and the ratio between MC and data (bottom). Data are compared to various MC generators.

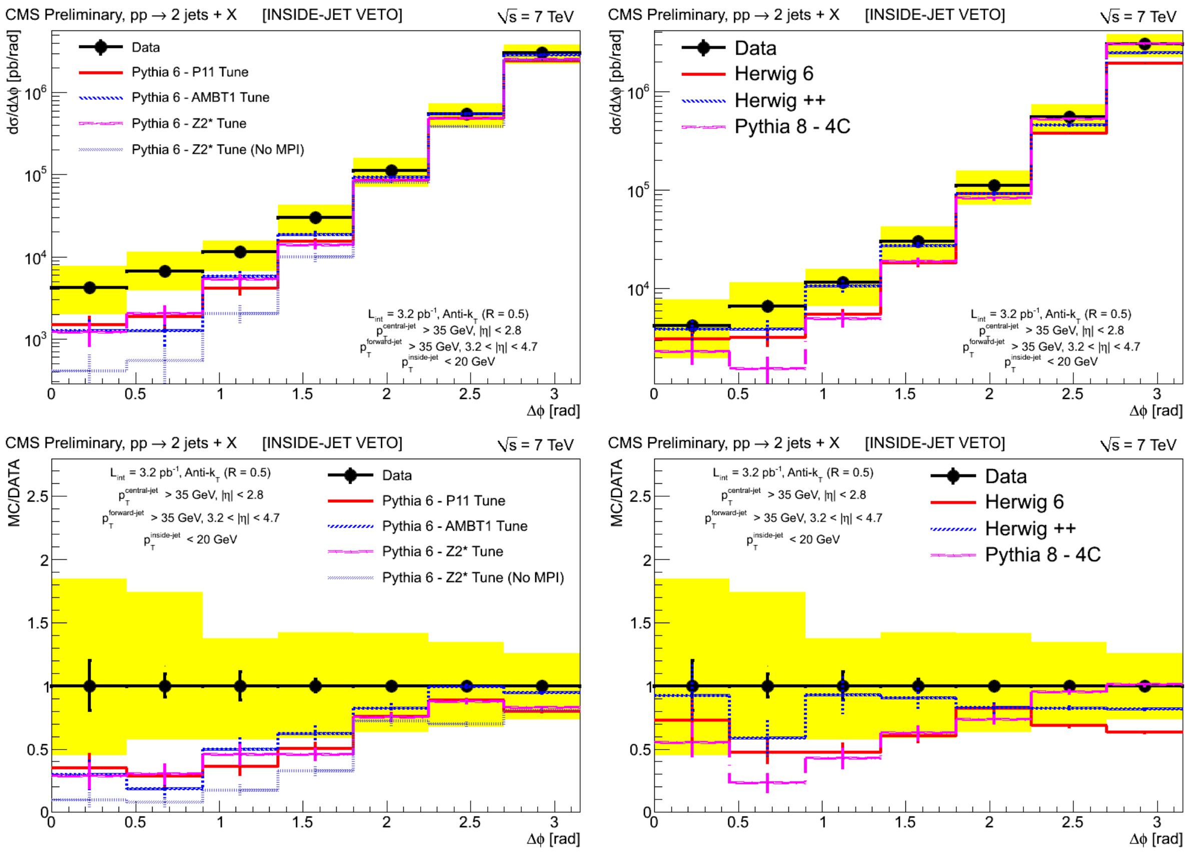


Figure 8.4: Cross–section for the inside–jet veto scenario in bins of the difference in azimuthal angle between the forward and the central jet, $\Delta\phi$, (top) and ratios between MC and data (bottom).

In figures 8.5 and 8.6, and in K.3 and K.4 in the appendix K, the $\Delta\phi$ distribution in the inside–jet veto scenario in different regions of of $\Delta\eta$ are presented. There is no significant variation of the slope with $\Delta\eta$, all distributions display the same three orders difference between the first and last bin. This outcome was expected as the absence of an inside–jet would not destroy the inherent correlations between the leading jets. Except for the prediction without MPI, the PYTHIA 6 tunes agree with each other. The PYTHIA 6 tune without MPI is up to one order of magnitude lower than data in the lower $\Delta\phi$ region. This can be understood if the lower p_{T} jets are taken into account, which do not interfers with the scenario requirement but still flatten the distribution. The predictions for PYTHIA 6 are inside the error band for the higher $\Delta\phi$ region and outside for the lower $\Delta\phi$ region. Except for the statistical fluctuation in the wider $\Delta\eta$ case in the lower $\Delta\phi$ region, the behaviour is similar for all PYTHIA 6 tunes in all $\Delta\eta$ slices. With the exception of a few statistical fluctuations in the lower $\Delta\phi$ region, the conclusions regarding PYTHIA 8 and HERWIG 6 are the same as for the integrated inside–jet veto scenario. The best prediction is still provided by HERWIG++, showing only small differences in the lower $\Delta\phi$ region.

8.3 Inside–jet tag scenario

The leading inside–jet p_{T} is presented in figure 8.7 and the leading inside–jet η^* is shown in figure 8.8. In figures 8.9, 8.10 and 8.11, also in K.5 and K.6 in the appendix K, the azimuthal jet

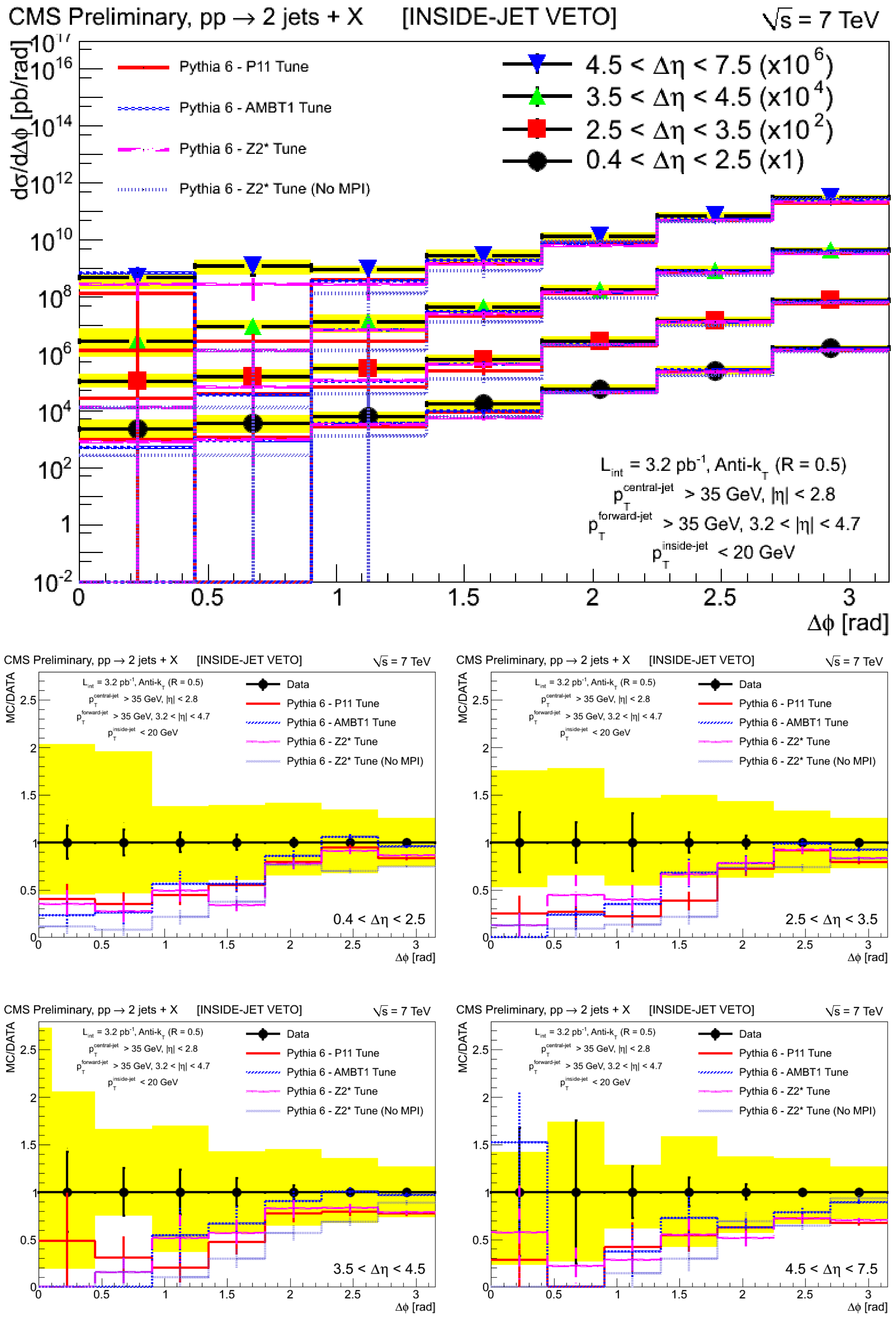


Figure 8.5: Double–differential cross–section for the inside-jet veto scenario in bins of the difference in azimuthal angle between the forward and the central jet, $\Delta\phi$, (top) as function of the pseudorapidity separation between the two jets, $\Delta\eta$, and ratios between MC and data (bottom).

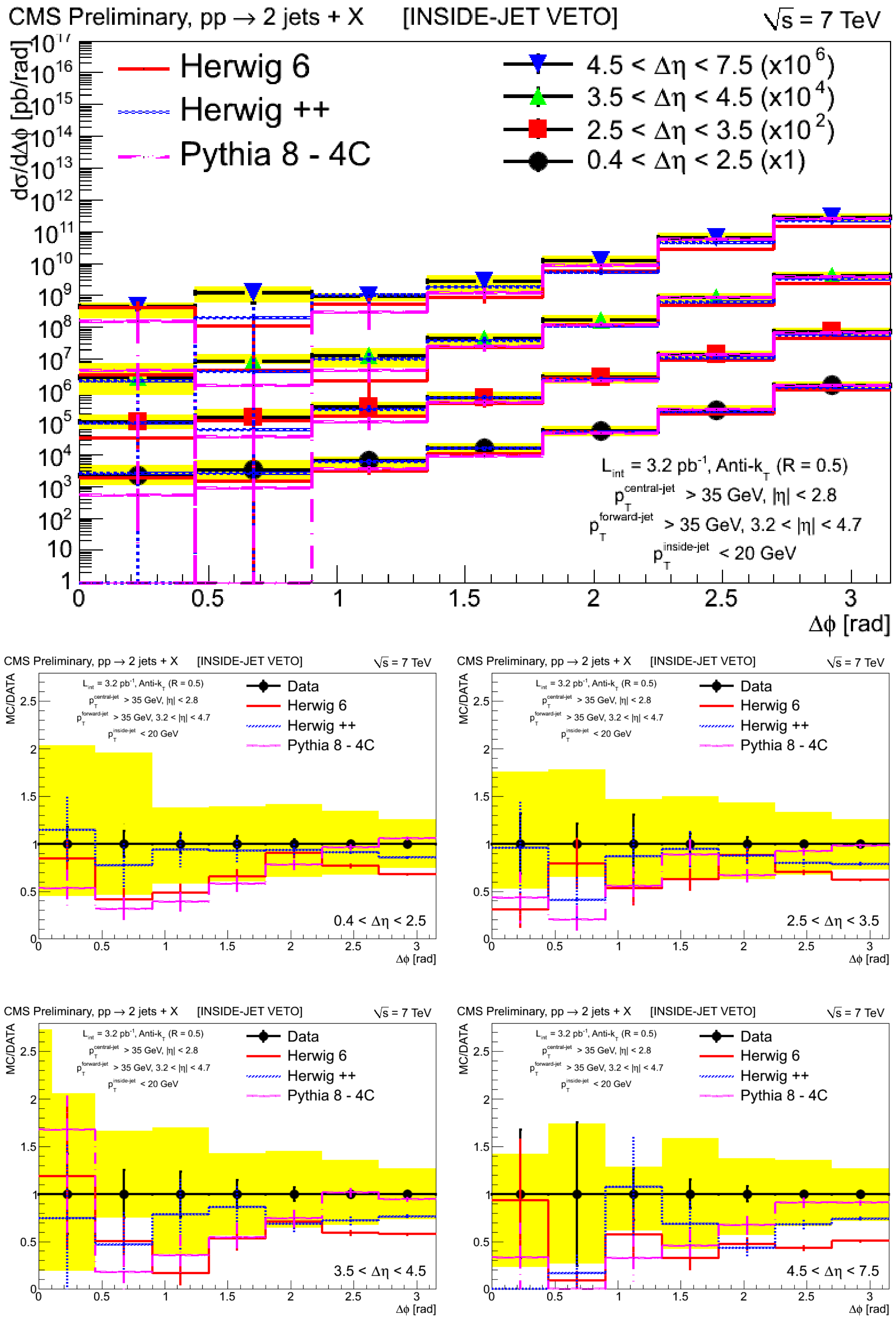


Figure 8.6: Double–differential cross–section for the inside-jet veto scenario in bins of the difference in azimuthal angle between the forward and the central jet, $\Delta\phi$, (top) as a function of the pseudorapidity separation between the two jets, $\Delta\eta$, and ratios between MC and data (bottom).

correlations for the inside–jet tag scenario is shown, in which only events with inside–jets above $p_{\mathrm{T}} = 20$ GeV between the forward and the central jets have been accepted.

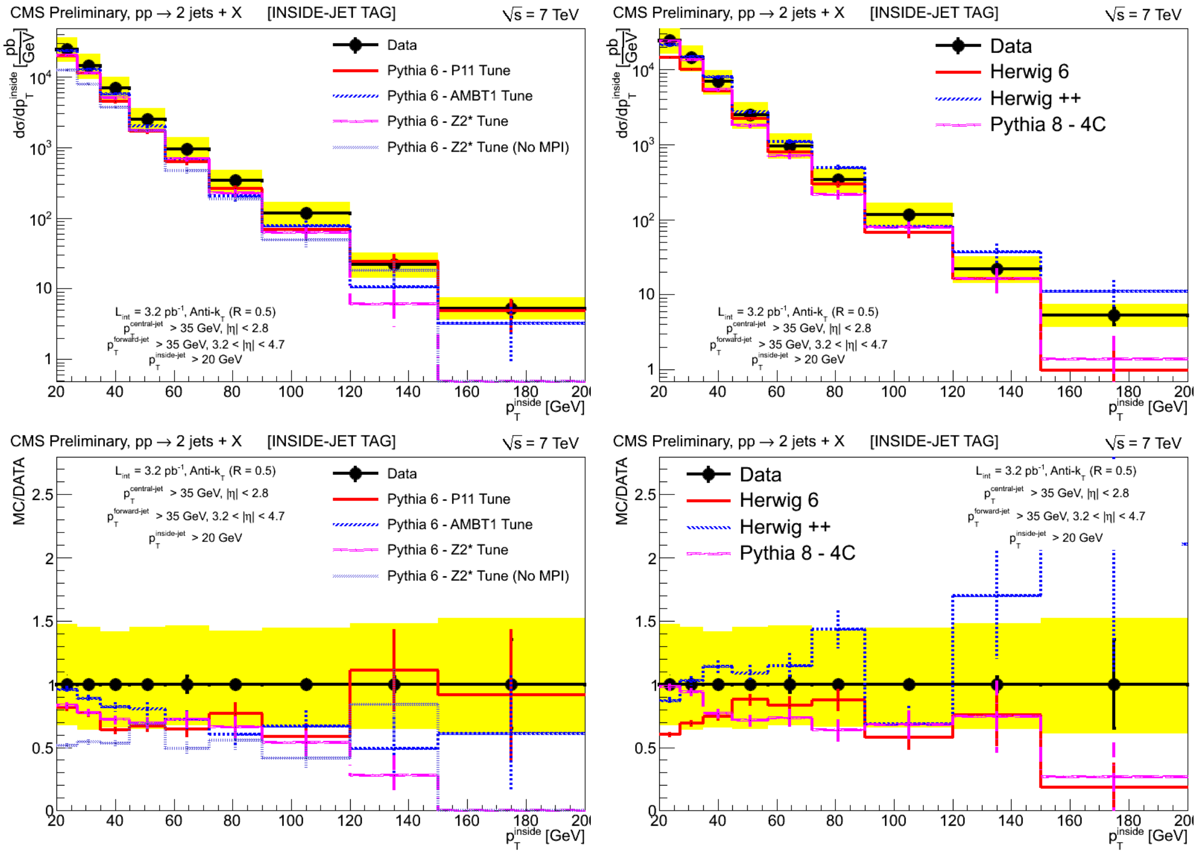


Figure 8.7: Cross–section as a function of p_{T} of the inside-jet. Data are compared to PYTHIA6 (left) and various MC generators (right).

The leading inside–jet p_{T} is a falling power–law distribution towards the higher p_{T} region and spans almost over four orders of magnitude. The various PYTHIA 6 tunes describe the distribution reasonably well for the lower p_{T} region, agreeing with each other, except for the tune without MPI. The PYTHIA 6 without MPI predicts a lower cross–section in the lower p_{T} region. The Z2* tune yields a very low cross–section for higher p_{T} region. PYTHIA 8, HERWIG 6 and HERWIG ++ display an acceptable prediction for the lower p_{T} region. HERWIG ++ deviates for the higher–p_{T} region, showing different shape than any other generator, having a flatter slope. The best description is given by PYTHIA 6 with P11 tune, which has similar results to the AMBT1 tune.

The leading inside–jet η^*cross–section is shown in figure 8.8. Most of the events are concentrated in the central region of the distribution, meaning that most of the leading inside–jets are near to the mid–point of both leading jets. The PYTHIA 6 prediction without MPI shows a significant deficit in the central region. The other MC predictions are within the uncertainties for most of the bins, not being able to fully describe the edges and giving an overall slighty lower cross–section. The best prediction is given by HERWIG++, describing the data within uncertainties, closely followed by PYTHIA 8.

The $\Delta\phi$ cross–section is presented in figure 8.9. As expected from kinematics, the events with an inside–jet give a flatter distribution, spanning only one and half orders of magnitude, because the jets are less correlated due to the fact that the additional inside–jet removes momentum from the forward–central jet system. When comparing the data to PYTHIA6, it is seen that the description of the shape by the MC is better than in the inclusive scenario. All tunes agree with

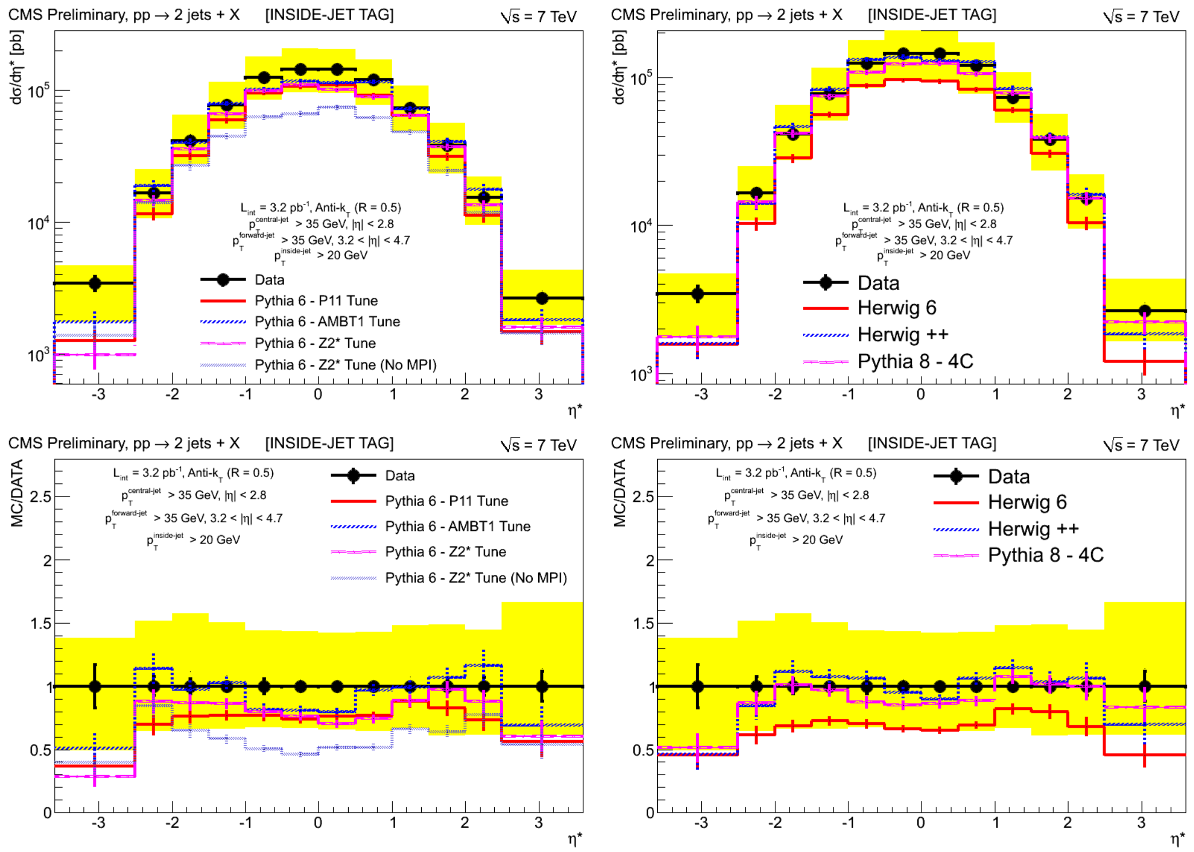


Figure 8.8: Cross–sections as a function of η^* of the inside-jet. Data are compared to PYTHIA6 (left) and various MC generators (right)

each other, except for the prediction without MPI, which is half an order of magnitude lower. HERWIG 6 yields a slightly lower cross–section. The best description is given by HERWIG++, followed closely by the PYTHIA 8, PYTHIA 6 AMBT1 and Z2* tunes.

In figures 8.5, 8.6, also in K.3 and K.4 in the appendix K, the $\Delta\phi$ distribution in the inside–jet veto scenario as a function of $\Delta\eta$ are shown. The slope is smaller than in the other scenarios and as $\Delta\eta$ increases, decreases from two to one and half orders of magnitude. This is expected, as the presence of extra radiation washes out the correlation as the pseudo–rapidity diference, $\Delta\eta$, between the leading jets increases and the phase–space for such radiation increases. The prediction without MPI yields consistently a lower cross–section in all slices. The other PYTHIA 6 Tunes agree with each other and they describe the data within uncertainties, yielding a similar performance. HERWIG 6 tends to underestimate the total cross–section. PYTHIA 8 yields a reasonable description, but HERWIG ++ gives the best description.

8.4 Outside–jet tag scenario

In figures 8.12 and 8.13, the measurements of the leading outside–jet p_T and $\Delta\eta^{out}$ cross–sections are shown.

The leading outside–jet p_T is presented in figure 8.12. The distribution is steeply falling spanning three orders of magnitude. Although the prediction without MPI is on the edge for the first bin, which reflects that low p_T jets contribute to some extent in this region, the leading inside–jet p_T is well described by all predictions in the lower $\Delta\phi$ region. In the higher $\Delta\phi$ region,

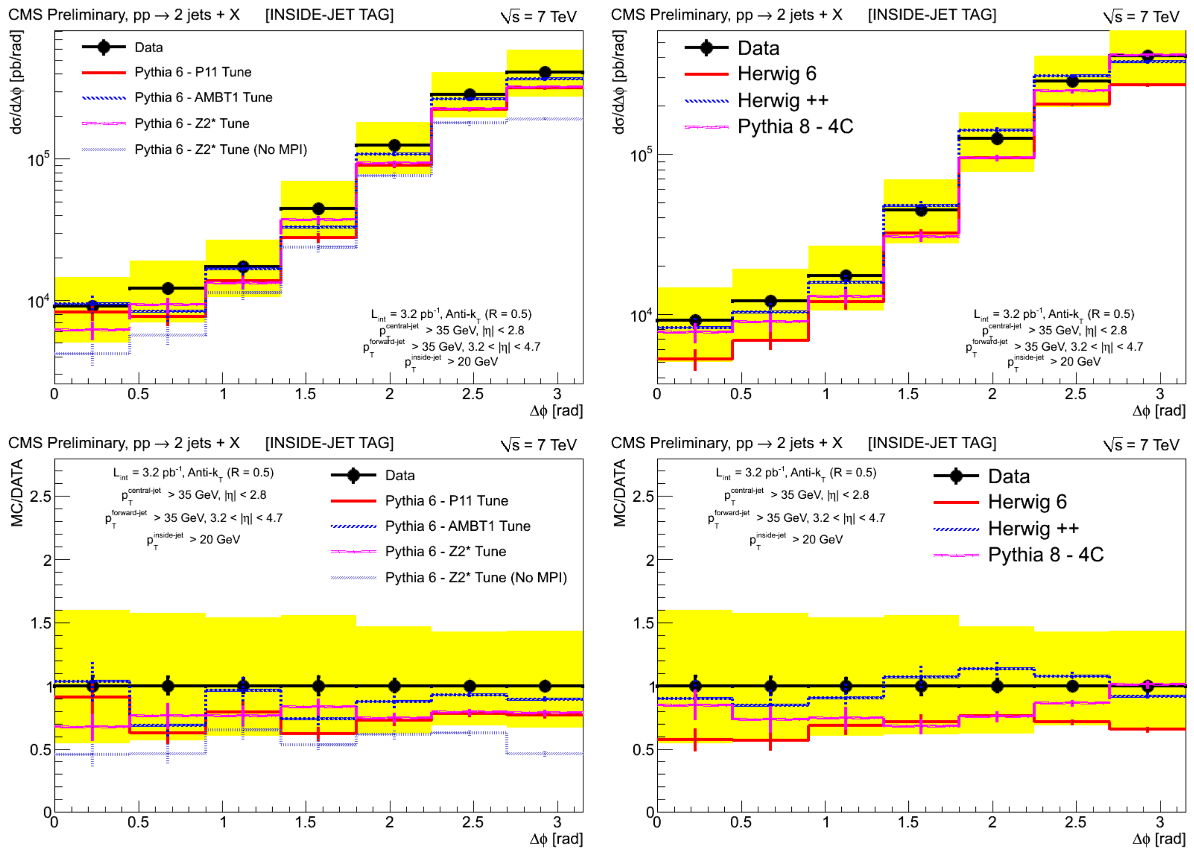


Figure 8.9: Cross–section for the inside–jet tag scenario as a function of the difference in azimuthal angle between the forward and the central jet, $\Delta\phi$, (top) and the ratio between MC and data (bottom).

the prediction given by the PYTHIA 6 AMBT1 tune is too low. The best description is given by the PYTHIA 6 Z2* tune.

The $\Delta\eta^{out}$ distribution is presented in figure 8.13. The distribution decreases towards the higher $\Delta\eta^{out}$ region, spanning around one and half orders of magnitude. The PYTHIA 6 prediction without MPI yields a smaller deficit for the lower $\Delta\eta^{out}$ region when comparing to other PYTHIA 6 Ttunes, which agree with each other. Except for the last bin, where the predictions are located near the lower edge of the correlated uncertainty band, the PYTHIA 6 tunes give a fair description of the data. The shape prediction by HERWIG 6 is compatible with the data and although the cross–section is slightly lower, but still inside the uncertainty band. The best description is given by the PYTHIA 8 and HERWIG++ generators.

8.5 Comparison of scenarios

The azimuthal correlations were measured in different scenarios and for different rapidity separations and compared with different Monte Carlo predictions. Overall Herwig 6.5 and Herwig++ perform better than Pythia 6 and Pythia 8, giving the best shape description. The best overall description is provided by Herwig++. The presence of MPI in Pythia 6 Z2* tune enhaces the prediction, which otherwise was almost one order of mangitude away from the data in the low–x phase–space. The different Pythia 6 Tunes yield very similar results, showing that appart from MPI, the modelation of remaining UE effects do not contribute in

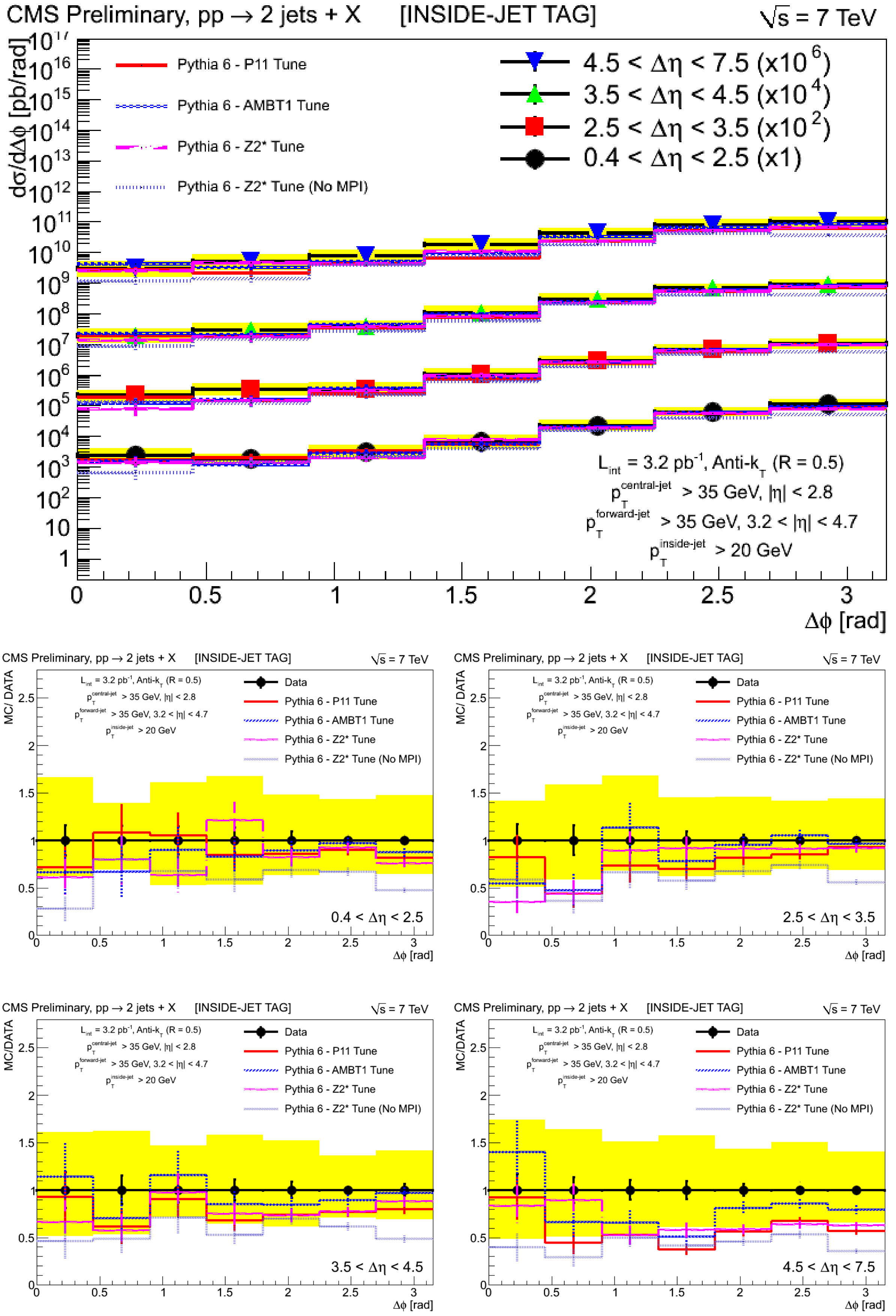


Figure 8.10: Double–differential cross–section for the inside–jet veto scenario as a function of the difference in azimuthal angle between the forward and the central jet, $\Delta\phi$, (top) as a function of the pseudorapidity separation between the two jets, $\Delta\eta$ and ratios between MC and data (bottom).

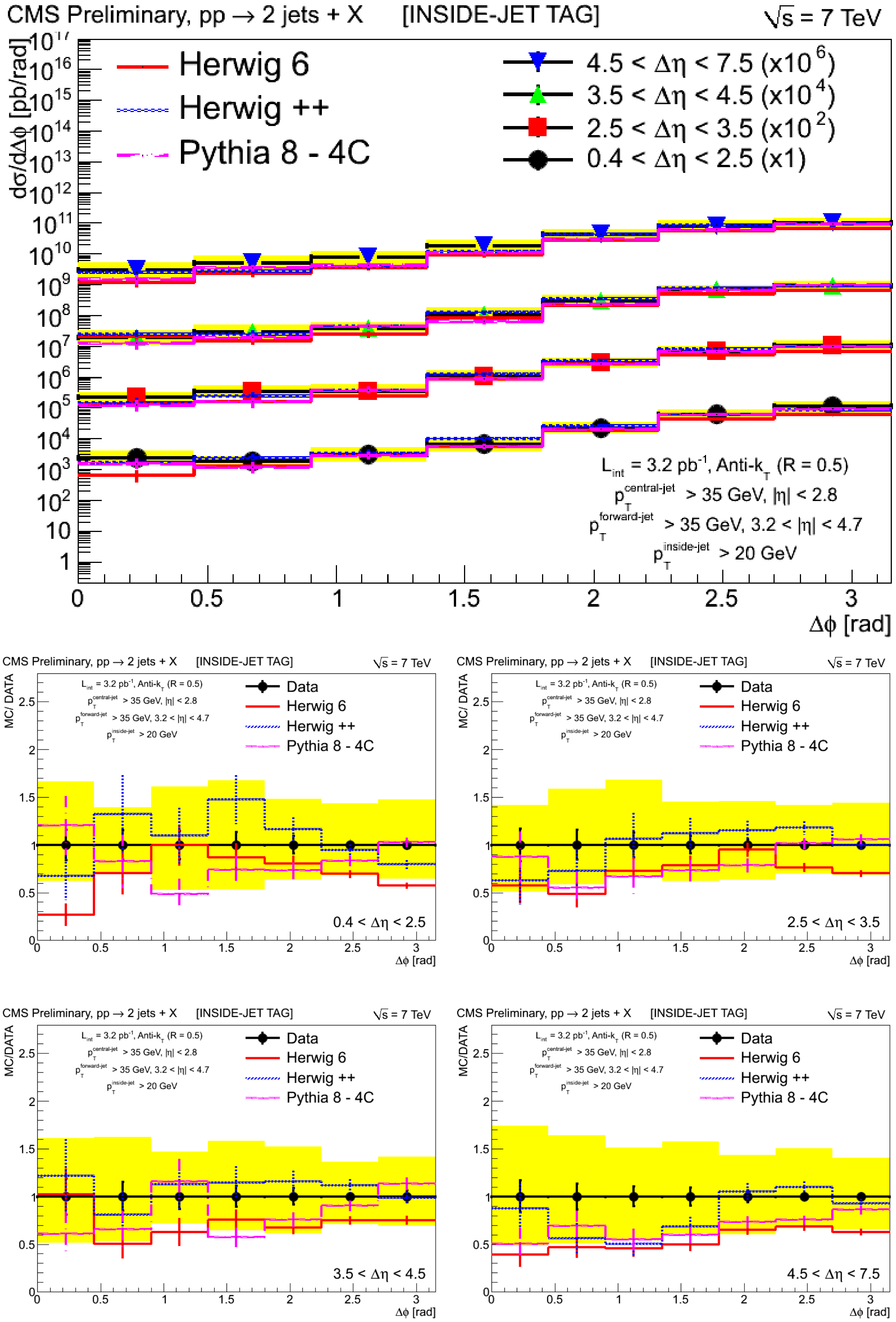


Figure 8.11: Double–differential cross–section for the inside–jet veto scenario in bins of the difference in azimuthal angle between the forward and the central jet, $\Delta\phi$, (top) as a function of the pseudorapidity separation between the two jets, $\Delta\eta$ and ratios between MC and data (bottom).

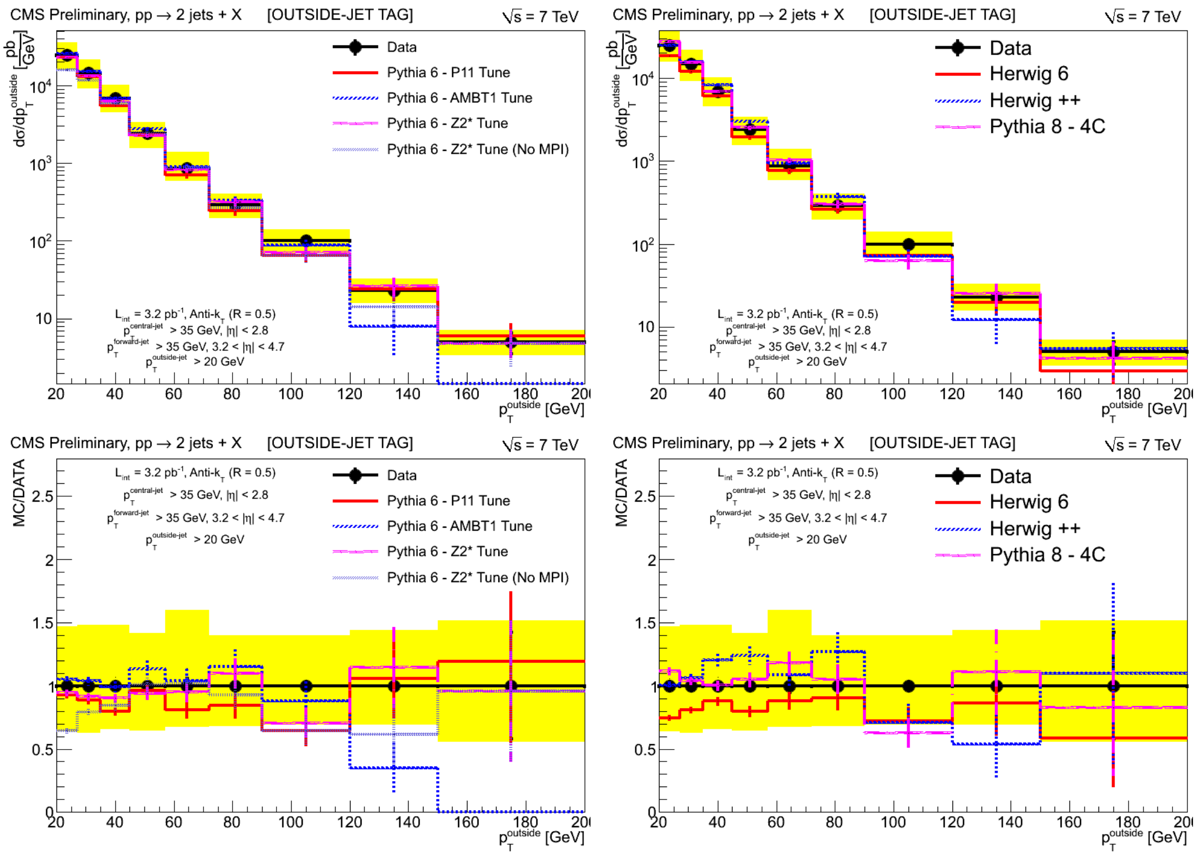


Figure 8.12: Cross-sections as a function of p_{T} of the outside–jet. Data are compared to PYTHIA6 (left) and various MC generators (right).

a significant way. It was found that DGLAP MCs describe the observables very well and no significant deviation from the DGLAP–regime was found.

The slope variation between the scenarios is further illustrated in figure 8.14 where the integrated measurements are compared in one plot. The plots highlight the differences in the slopes, thus emphasising the correlation dependence on the scenarios, making them into a perfect summary of the analysis.

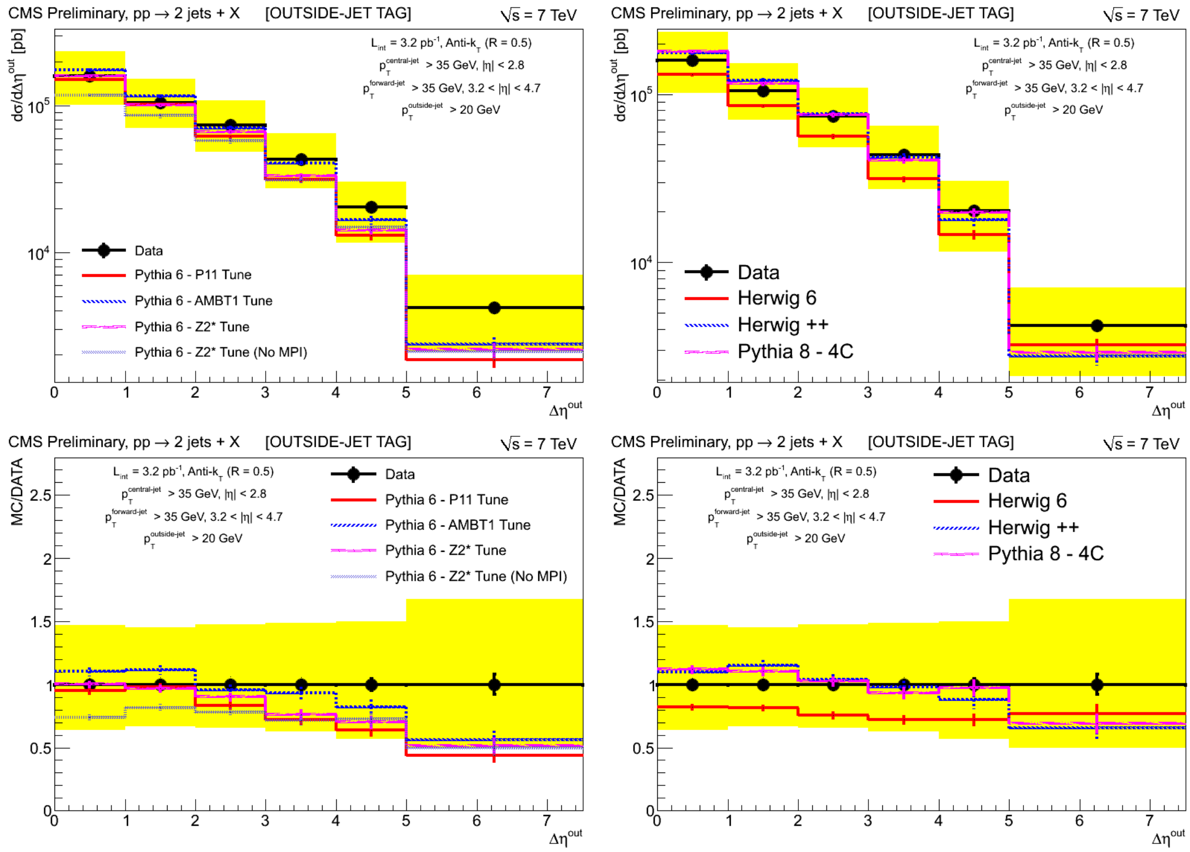


Figure 8.13: Cross-sections asa function of $\Delta\eta^{out}$ of the outside–jet. Data are compared to PYTHIA6 (left) and various MC generators (right).

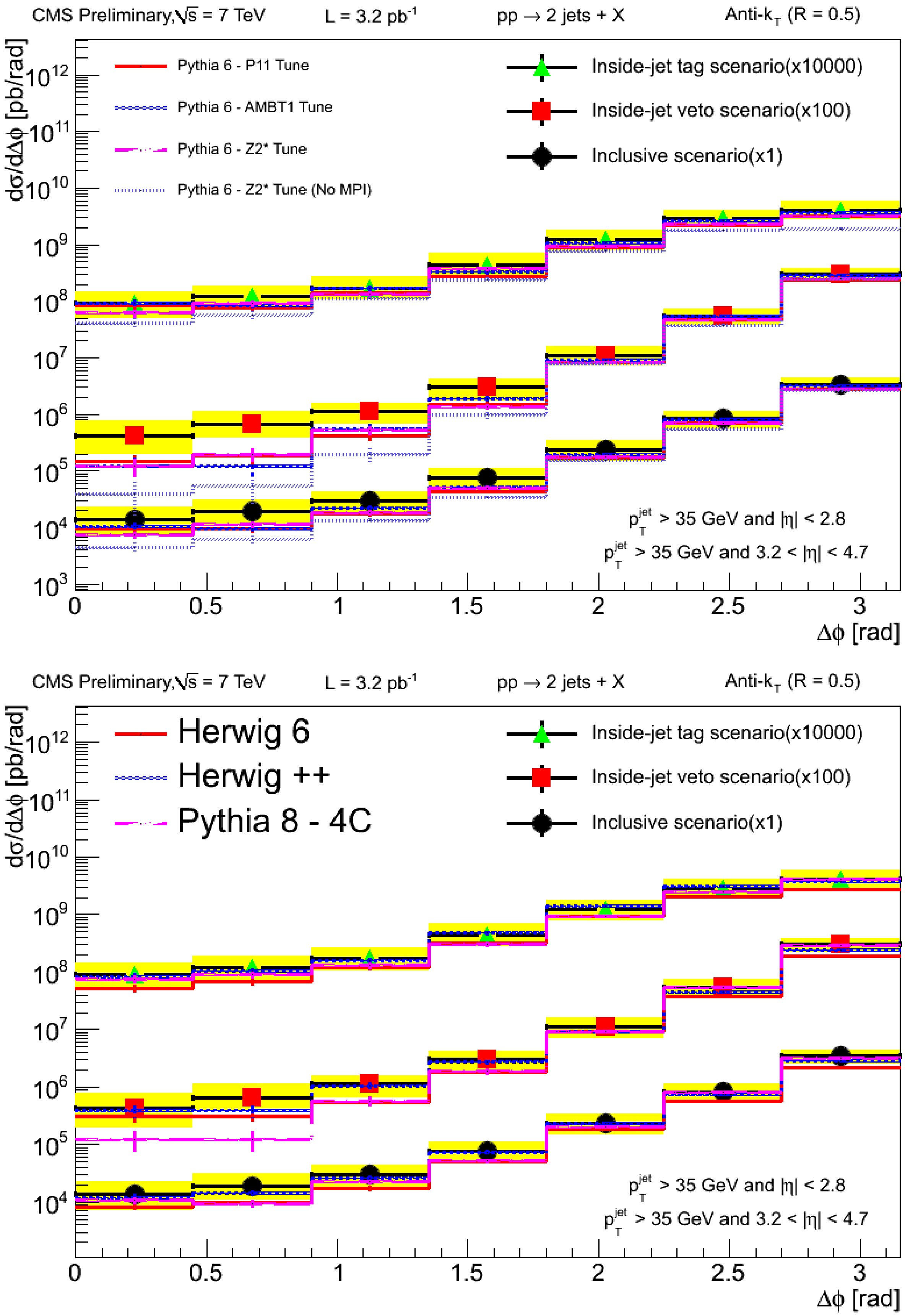


Figure 8.14: Comparison between the cross–sections for the three different scenarios investigated in this analysis.

Conclusions and Outlook

The azimuthal difference $\Delta\phi$ between forward ($3.2 < |\eta| < 4.7$) and central jets ($|\eta| < 2.8$) has been measured at a centre-of-mass energy of $\sqrt{s} = 7$ TeV. The measurement is done inclusively and differentially in the pseudo–rapidity separation, $\Delta\eta$, between the jets. In addition, the measurement is repeated for events with a third jet, with $p_{\mathrm{T}} > 20$ GeV, between the forward and the central jet in pseudo–rapidity, and for events with this inter-leading jet activity vetoed. The tendency is that the correlation between the jets decreases with increasing $\Delta\eta$ and when a third jet is required between the forward and the central jet. This effect is expected since the additional radiation in the event will naturally destroy the simple back-to-back topology of pure dijet events. Consequently, the correlation is strongest when the third-jet activity is vetoed.

The tested Monte Carlo generators and tunes in general describe the measurement within the systematic uncertainties. HERWIG++ provides by far the best predictions describing both the shape and normalization of the data.

Similar conclusions apply for the case of the measurement of the p_{T} and η^* of the leading jet inside the forward-central dijet system. Again, HERWIG++ describes best the overall shape of the spectrum. The p_{T} of the leading jet outside of the leading forward-central dijet system is also in shape best described by HERWIG++. The pseudo–rapidity separation between the outside-jet and the forward-central dijet system is remarkably well described by both PYTHIA 8 and HERWIG++, while PYTHIA 6 fails to describe both shape and normalization.

The different PYTHIA 6 Tunes do not differ much in their predictions. When the contribution from MPI is turned off, the description of the dijet correlation is worse, because the increased activity in the event which interfers with the leading pair is not present.

In summary, the DGLAP–based generators describe well the complex observables of this analysis. No conclusion about evidence for BFKL–like parton dynamics, or any other evolution scheme beyond DGLAP, can be drawn from the presented measurement.

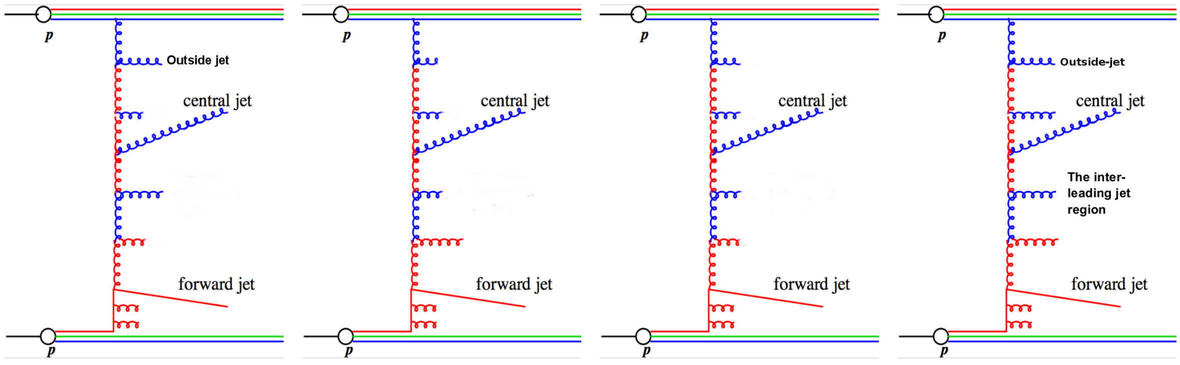


Figure 8.15: Feynman diagrams illustrating the different topologies. From left to right: outside–jet tag scenario, outside–jet veto scenario, exclusive scenario and inside–jet tag and outside–jet tag scenario.

Comparison with MC using different evolution equations are being prepared, thus the BFKL and CCFM [37–40] prescriptions will be tested. The Sherpa [185, 186], Cascade [153] and Pohweg [154] predictions will be compared to data.

The next logical step, in which interesting results can be obtained, is to measure the correlations in the presence and absence of an outside–jet. This outside–jet could then define two scenarios: outside–jet veto and outside–jet tag, see figure 8.15, in which $\Delta\phi$ could be measured both as single differential and as a function of $\Delta\eta$. Unfortunately, because of the lack of statistics for the low pile–up data, it is only possible to measure the exclusive dijet case, presented in figure 8.15, as a function of $\Delta\eta$ but it is not possible to measure it for the scenario with both an inside and an outside–jet, shown in figure 8.15. A summary of the statistics avaliable for these scenarios is shown in table 8.1. With all these scenarios together, it would be possible to draw a more accurate picture on the influence of extra radiation in the dijet correlations and thus, better understand the subtle effects of the parton evolution.

Table 8.1: Number of events in each scenario to studied

Sample	JetMETTau_2010A	JetMET_2010A	Total
Outside–jet tag scenario	17872	54980	72852
Outside–jet veto scenario	43414	81823	125237
Exclusive scenario	31054	46051	77105
Inside+outside–jet tag scenario	5340	21108	26448

Because of the higher pile–up conditions foreseen in the next LHC phase, some of these measurements will become extremely difficult. Thus, inclusive cross–sections may not be the easiest way to study the QCD phenomena, but rather a difference of cross-sections [187].

I must not fear
Fear is the mind-killer
Fear is the little death that brings total obliteration
I will face my fear
I will permit it to pass over me and through me
And when it has gone past I will turn the inner eye to see its path
Where the fear has gone there will be nothing
Only I will remain

- Bene Gesserit Litany Against Fear, Dune, Frank Herbert

Acknowledgements

First of all, I would like to thank my office mates at DESY: Mira, Pavel, Zlatka, Nil, an elderly emeritus professor which I don't remember the name, Paolo and Samantha. I would also like to thank my office mates at CERN: Ivan, Ekatherina, David, Walter and Panos. They have shared (many of them not a voluntary basis) most of my happy and sad moments with outstanding patience, answering questions and helping out in the most difficult moments. Most of you know that it is not easy to write a thesis (if not, you would soon enough!). This happy end would not have been possible without you.

I would like to thank my supervisors and mentors: Albert Knutsson, Hannes Jung and Thomas Schörner–Sadenius. Albert taught me how to perform an analysis, while Hannes make sure i did not stray away and Thomas tried to keep everything together. They made me fell welcome and at easy in the CMS community. Their contributions and helpful comments made me progress, not only in the analysis but also in the scientific world. Both Hannes and Thomas supported me (although DESY actually paid for) when I was willing to participate in a school or conference, which helped to expand my knowledge in particle physics.

It was a pleasure to work in the QCD group at DESY: Hannes, Thomas, Albert, Martin, Nastja, Mira, Zlatka, Igor, Panos, Alex, Nil, Paolo, Samantha and Hans. It was very nice to have the memorable Physics and Cookies every week (even when there was no cookies or tea). I will always remember those nice dinners at Blau Blumen.

In the CMS group at DESY, I would like to thank Agni, Tayler, Carmen, Jasone, Caroline, Francesco, Ivan, Wolf, Luigi, Gregor, David, Armin, Maria, Artur, Nazar, Mathias, Eleni, Altan, Ozgür, Dirk, Roberval, Johannes and Jan (I hope I didn't forgot anyone!). At some point or other you have contributed to my thesis or my general happiness (in most cases for both even without knowing or willing).

I can not forget the people (other than my supervisors) that gave me feedback while writing/correcting the thesis: Paolo, Samantha, Jasone, Francesco, Mathias, Agni, Carmen, Krzysztof and Eleni. Your comments and suggestions greatly improved the overall quality of the thesis and because of that I owe you all very much. I would like to thank the chair of the disputation Prof. Dr. Georg Steinbrueck for the nice introduction of me. I want to thank my supervisors (Dr. Hannes Jung and Dr. Thomas Schörner–Sadenius) for being the referees for the written part and Dr. Thomas Schörner–Sadenius and Prof. Dr. Peter Schleper for being the referees for the disputation. I would like to thank all those who (in more or less voluntary basis) were willing to ask question during the disputation: Dra. Carmen Diez Pardos, Dra. Eleni Ntomari, Dr. Altan Cakir, Dr. Francisco Pedro, Dr. Simon Plätzer, Dr. Johannes Hauk, Dr. Roberval Walsh and Dr. Panos Katsas. I would also like to thank all those who attended the disputation, in special the one who prepared the nice hat I got: Ivan, Carmen, Samantha, Paolo, Jenia, Ganna, Jasone, Maikel, Marco, Stefano and Ozgür (I again hope that I didn't forgot anyone).

Apart for the DESY–CMS group, there are some people to which I'm very much in debt: Gero, Stefano, Francisco, Marco, Maikel, Halil, Henning and many others. There are also many nice persons I meet at CERN and in the 69th Scotish Summer School of Particle Physics: Resmus, Colin, Xenophon, Wim, Varun, Flavia, Roxana, Zhen, Suyog, Donnchadha, Marzieh, Chris, Jonathan, Shreyashi, Laurent, Ipsita, Robert, Esben, Maik, Adrian and Daniel.

Outside DESY, I would like to thank my life partner: Ana Filipa Guerra Silva Gomes da Piedade. She is like a pillar in my life with enduring patience and resourcefulness. Filipa help me through the whole PhD by being always present. She made sure that I would not go nuts during the most stressful phases. There is no words to express my gratitude and love for her. Mi amas vin!

I would like to thank my family to support my decision of pursuing my academic career so far away from home.

Finally, I would like to thank some very nice people which I meet in/lived with Hamburg during the last three years and 6 months: Dennis, Tom, Daniel, Jan, Felipe, Pri, Yingying, Manvidas, Alvin, Annika, Bihn, Bugra, Clemens, Dominik, Gamze, Hannah, James, Jennifer, Jens, Jessica, Jiyoung, João, Kim, Klaus, Konrad, Svetlana, Allan, Martina, Nasia, Natasa, Niklas, Nina, Rasmus, Sara, Sasha, Sebastian, Tigran and Valentina.

And, once again, thank you very much!

Appendix A

Service Work - Castor

Every scientist joining the CMS Collaboration is expected to perform the necessary service to the experiment. The technical part of the thesis focused on the CASTOR sub–detector, which is described in detail in subsection 2.3.4. The calibration (see section A.1), upgrade (see section A.2) and DQM (see section A.3) activities, in which I took part, are summarized in this chapter.

A.1 Castor Calibration

The absolute CASTOR jet energy calibration was performed, during 2011, using the dijet p_{T}–balacing technique. Events with at least one central jet, $|\eta| < 2.5$, and at least one jet in CASTOR, $-5.2 < \eta < -6.6$, were selected. In a balanced dijet event in a perfect calibrated detector, the transverse momenta of the two jets are in average equal. The non–uniform jet energy response as function of pseudo–rapidity spoils the balance. Thus, the scale of a dijet event is defined by the average transverse momentum of the two jets:

$$p_{\mathrm{T}}^{dijet} = \frac{p_{\mathrm{T}}^{central} + p_{\mathrm{T}}^{castor}}{2} \tag{A.1}$$

The balance, B, is the difference between the p_{T} of the two jets:

$$B^{dijet} = \frac{p_{\mathrm{T}}^{central} - p_{\mathrm{T}}^{castor}}{2} \tag{A.2}$$

The balance can be used to estimate the relative response of the jet in CASTOR with respect to the central jet:

$$r = \frac{2 + \langle B \rangle}{2 - \langle B \rangle} \tag{A.3}$$

For exactly one event, the estimator r is just $p_{\mathrm{T}}^{castor}/p_{\mathrm{T}}^{central}$, while for many events it is the least biased estimator of the relative response.

The dijet p_{T}–balance for jets in the back–to–back configuration in the azimuthal plane, $\Delta\phi > 2.7$ rad, is presented in figure A.1. There is a good agreement between the Pythia 6, Pythia 8 , Herwig++ predictions and the data.

A.2 Castor Upgrade

During 2012, CASTOR went through an upgrade phase, which was intended to prepare the detector for the pPb run and for the low pile–up pp run in the beginning of 2013, before the first long shut–down. The Photomultiplier Tubes characterization, discussed in section A.2.1, and replacement was one of the main tasks of this upgrade.

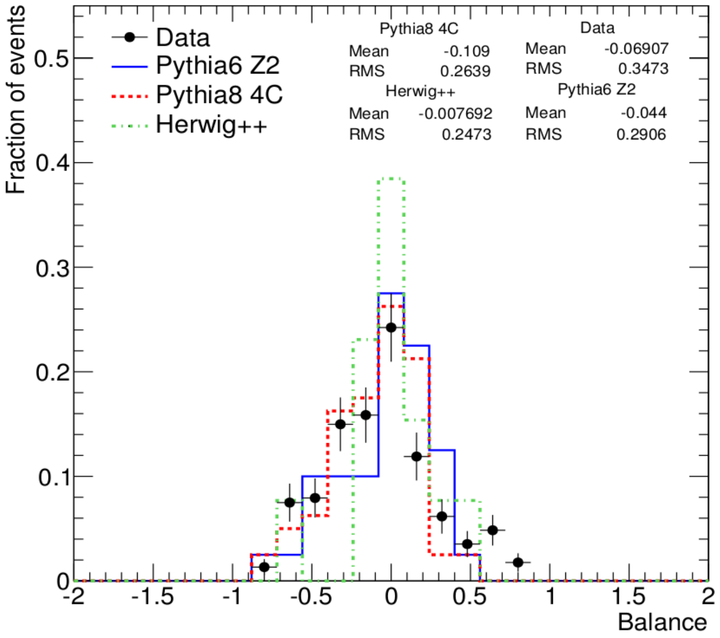


Figure A.1: Dijet p_{T}–balance between a central jet and a CASTOR jet compared with different MC predictions.

A.2.1 PMT Characterization

The PMT (**P**hoto**M**ultiplier **T**ube) characterization was carried out during the summer of 2012. All PMTs in sectors 1 to 10, and 4 in the sector 11 were removed and tested. This was necessary to estimate the radiation damage on the PMTs and to decide the ones in need of replacement. The test consisted in illuminating the PMT with light from various wavelengths and measure its cathode and anode currents. This was done in a dark box and the LEDs had the following wavelengths: 400, 470, 500 and 517 nm. The measurement has been carried for the working voltages of 800, 900, 1000, 1200, 1400, 1600 and 1800 V.

A typical measurement would span around three hours, and an example can be seen in figure A.2. The procedure was the following:

1. It started with a rest phase of about two minutes, in which no voltage or light were fed into the PMT.
2. A working current was appplied to the PMT and the system was allowed to rest for around ten minutes. This is important to stabilize the PMT and reduce the dark current, which is the cathode current due the applied voltage.
3. An LED light was switched on for about ten minutes and then switched off.
4. The box was left without any LED light for around five minutes.
5. If no more LED have to be tested, then the working current is changed. Go to step 2. If another LED have to be tested, then go to step 3.
6. If all LED and the working currents have been tested, then all LEDs and the working current are switched off.
7. After waiting two minutes, the measurement is finished.

The cathode, working voltage and LED switched on, as well the anode currents for the PMT under test and the reference PMT, which was one PMT not yet damaged by the radiation, were

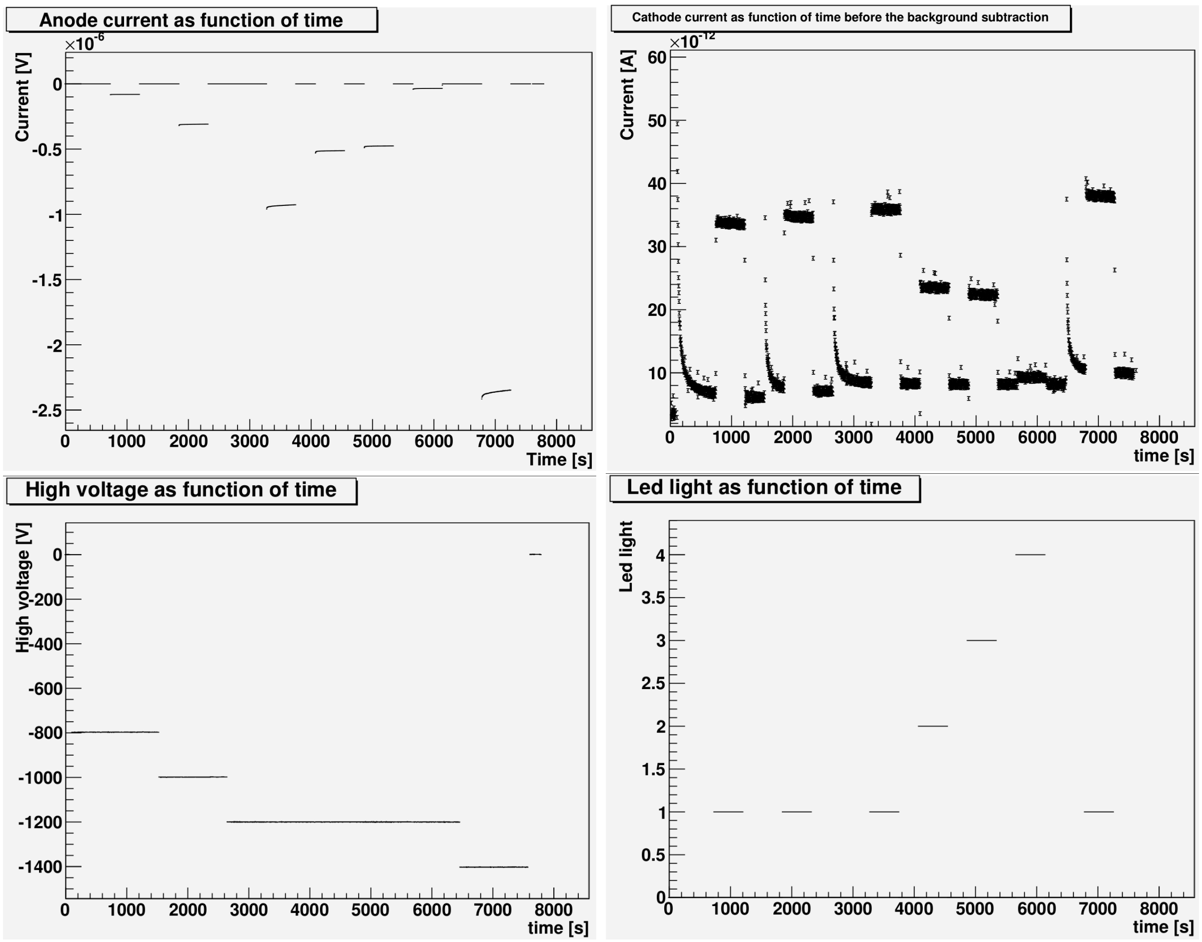


Figure A.2: A typical PMT measurement: anode current (top–left), high voltage feed to the PMT (bottom–left), cathode current (top–right) and LED light (bottom–right) as function of time.

recorded. Points in the cathode current with a sudden and unexplained variation, denonimated spikes, were removed from the subsequent treatment of the data. To ensure the robustness of the method, several PMT were measured more than once. In general, the repetitions yield similar results. A function was fitted to the cathode current without any LED light turned on, the dark current:

$$y = p_0 + p_1^{-\frac{x}{p_2}}, \tag{A.4}$$

where y is the current in Ampere, x is the time in seconds and p_0, p_1 and p_2 are fitting parameters. An example of such a fit is presented in figure A.3. The interval without points in the plot corresponds to the removed points due to the LED light.

If the fit was sucessful, the value of the resulting function was subtracted from all the data points corresponding to a single working voltage, as can be seen in figure A.4. In the case of a good fit, the base current is around 0. If the fit was not sucessful, the points around where the LED was switched on and off were sampled and averaged to estimate the jump. This value or the absolute value of the cathode current with the LED light, when the subtraction was done, have been used as cathode current.

Thus, the gain is defined as:

$$gain = \frac{anode\ current}{cathode\ current}. \tag{A.5}$$

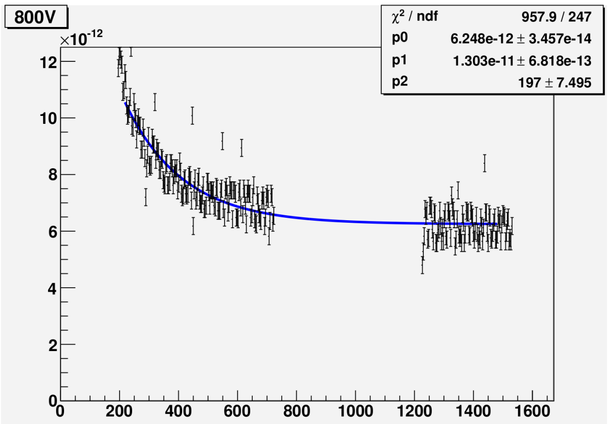


Figure A.3: An example of the fit of leakage current in the cathode.

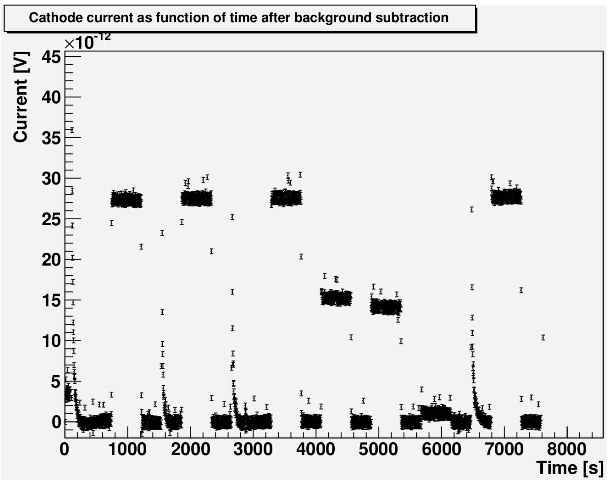


Figure A.4: Cathode current after leakage current subtraction.

The quantum efficiency is defined as:

$$QE = \frac{cathode\ current}{anode_{reference}\ current}. \tag{A.6}$$

The values of the gain, quantum efficiency and ratio of the gain of the first and third led, all at 1200 V, as well as the number of measurements and number of bad fits ($\chi^2/NDF > 100$) and were mapped for both old and new PMTs.

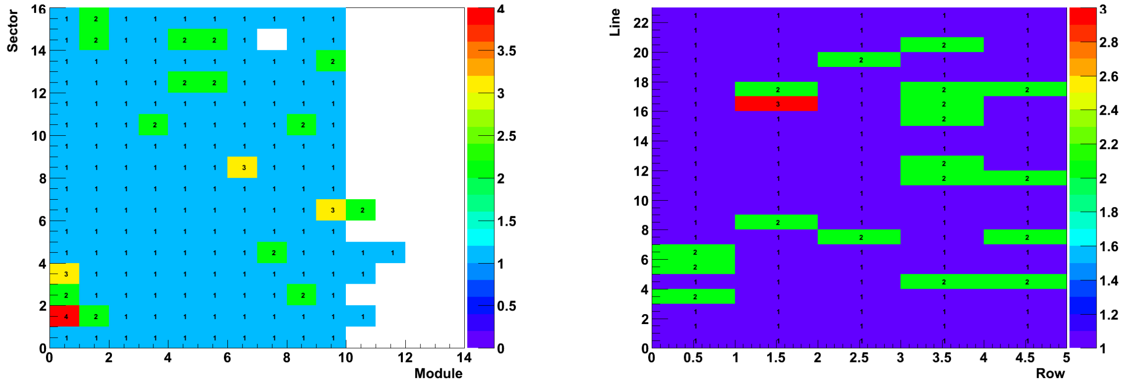


Figure A.5: Number of measurements for old (left) and new (right) PMTs.

The number of measurements for each PMT is shown in figure A.5. Some of the multiple measurements were either incomplete or made to test the consistence of the method, or to have an extra input when the characterization of the PMT was dubious.

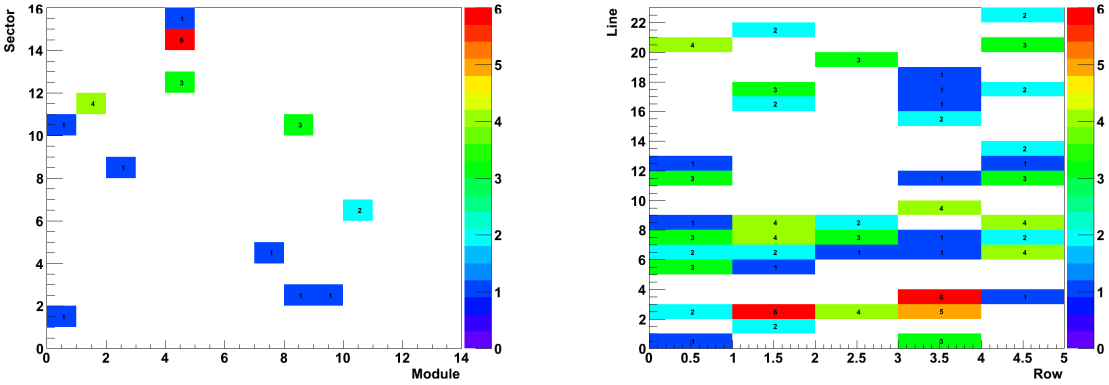


Figure A.6: Number of bad fits for old (left) and new (right) PMTs.

The number of bad fits for each PMT is presented in figure A.6. There were more bad fits in the new PMTs than in the old ones because the dark current was much higher in the new ones. Even in the measurement with bad fits it was possible to extract the cathode current for the background.

The gain is shown in figure A.7. One can see that the radiation has not affected all the PMTs in the same way. One could expected more damage in the first modules, but the gains do not yield this kind of structure.

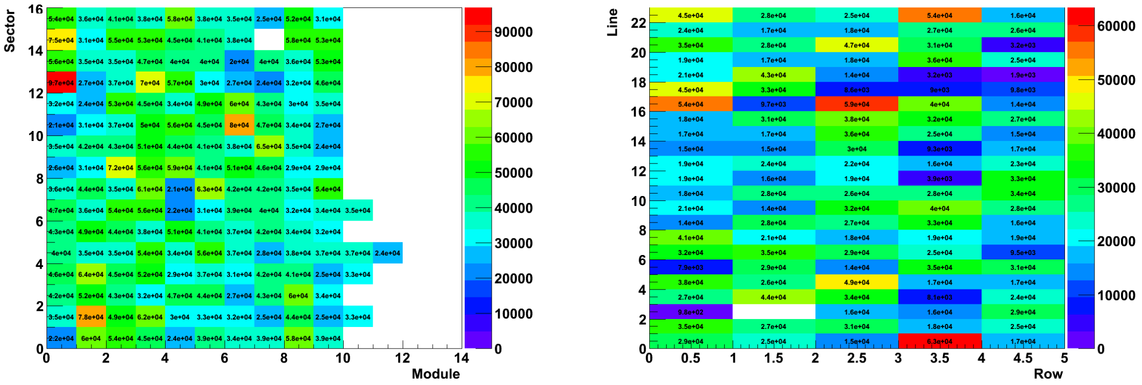


Figure A.7: Gain at 1200V for old (left) and new (right) PMTs.

In figure A.8, the quantum efficiency at 1200V is presented. The values for the new PMTs were very similar to each other and relatively high, except for one defective PMT, located in third line of the second row. As expected, the values of the quantum efficiency for the old PMTs were lower but they did not follow any particular pattern.

In figure A.9, the ratio of the quantum efficiency at 1200V between the first and the third LED lights is shown. This ratio reflects the darkening of the cathode glass as function of the wavelenght due to radiation damage. Again, no particular structures are present for the old PMTs. The variation among the new PMTs was smaller than in the old ones.

Taking in account the quantum efficiencies and gains at 1200V, a proposal for the replacement and rearrangement of the PMTs was done. The gains and the quantum efficiencies of the installed PMTs are presented in figure A.11. The PMTs with higher gain were installed in sectors 7 and 8, to compensate for the maximum of the magnetic field present in that region. The highest quantum efficiencies were selected to be installed in the first modules, where the radiation damage

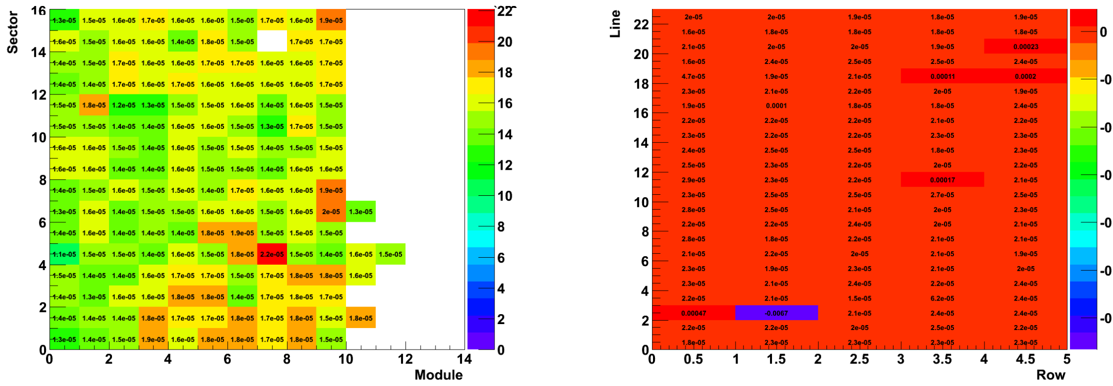


Figure A.8: Quantum efficiency at 1200V for old (left) and new (right) PMTs.

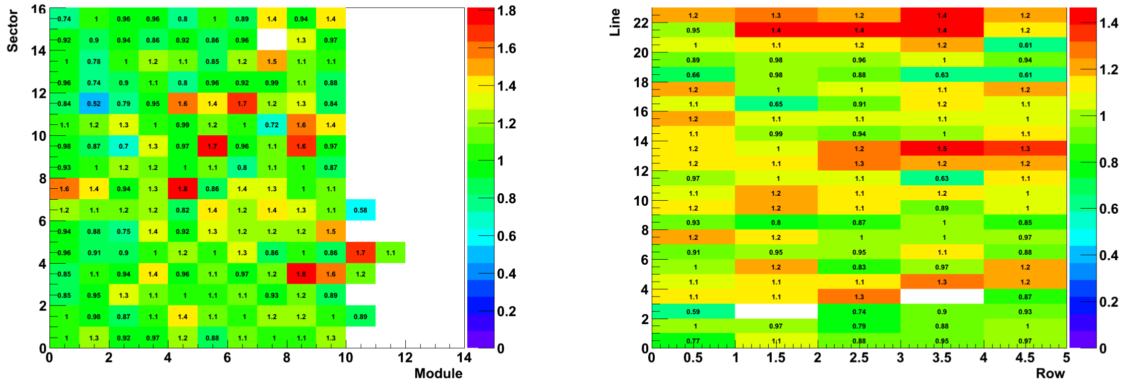


Figure A.9: Ratio of quantum efficiency at 1200 V between the first and the third LEDs for old (left) and new (right) PMTs.

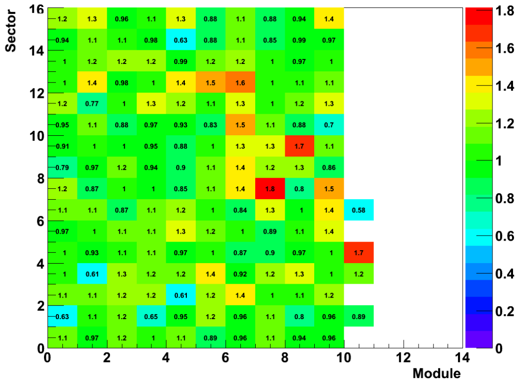


Figure A.10: Ratio of quantum efficiency at 1200 V between the first and the third LEDs for installed PMTs.

is expected to be more severe. The ratio of the quantum efficiency at 1200V between the first and the third LED is shown in figure A.10. One can see that the distribution is smoother and more homogeneneous than in the old PMTs, meaning that the values between neighbours are more similar.

This rearrengement was then applied in CASTOR for the data taking in the beginning of 2013.

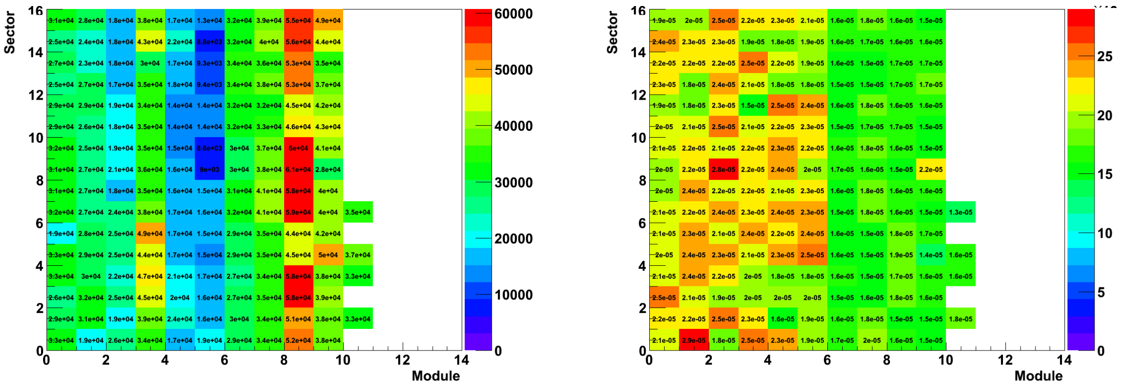


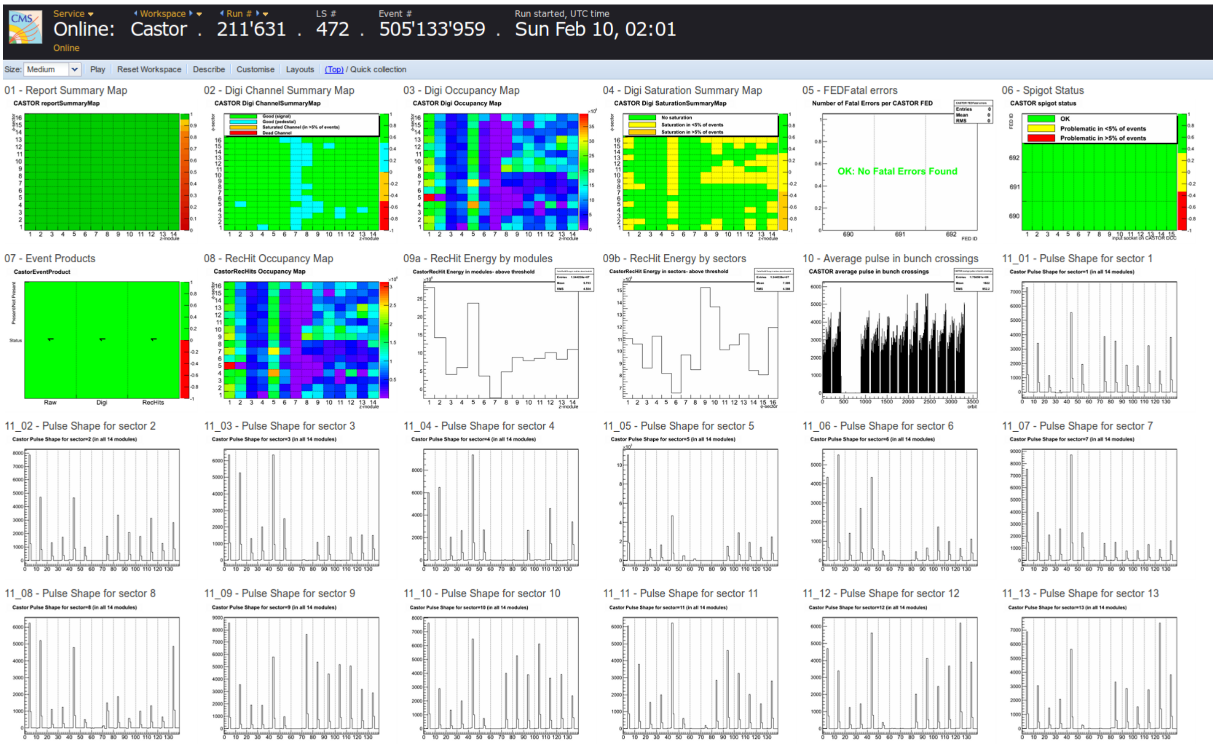
Figure A.11: Gain (left) and quantum efficiency (right) at 1200V for installed PMTs.

A.3 Castor DQM

From the autumn 2012, I was in charge of the CASTOR DQM development and data certification.

A.3.1 DQM development

Because of its upgrade, CASTOR was not part of the 2012 data taking. For that reason, part of the code had to be rewritten to match the new CMSSW requirements as the framework is always evolving. This oportunity was used to update the modules and the GUI, specially to improve the organization of the output.


Service: Online: Workspace: Castor . Run #: 211'631 . LS #: 472 . Event #: 505'133'959 . Run started, UTC time: Sun Feb 10, 02:01

Size: Medium

OK: No Fatal Errors Found

Figure A.12: CASTOR new GUI

The new GUI layout is shown in figure A.12. It was designed to include all the relevant information required for the data certification for both shifters and experts in one single page. This, together with an updated Twiki [188], lead to the inclusion of CASTOR into the official certification chain for the first time. Also, new consistency and sanity checks were implemented to make sure that the performance of the detector and the data collected is well understood.

A.3.2 Data Certification in 2013

In early 2013, CASTOR took part in the pPb and low pile–up pp runs. In the pPb runs, 31.13 nb^{-1} were collected with >99.8% efficiency: only few seconds of data taking could not be recovered because trigger and HCAL were not working. In the pp reference runs at 2.76 TeV, 5.41 pb^{-1} were collected. Although a hot channel, S4M11, was found, 100% of the data can be used. A typical report summary for the pp reference run can be found in figure A.13, in which the hot channel is displayed in orange.

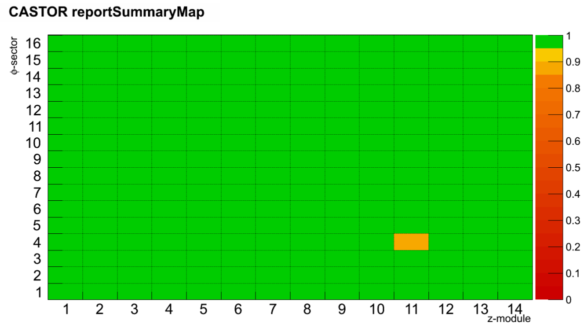


Figure A.13: A typical CASTOR report summary in a pp reference run in 2013.

A.3.3 LS1 software upgrade

During the LS1 (**L**ong **S**hutdown **1**), the DQM software is going through a major upgrade. The single–thread concept is no longer viable for the processing power needed in the next LHC phase, thus the multi–thread processing will be used. This require a complete rewrite of the code, which is on progress, to face the demands of the new framework requirements. In addition, several old and useless artifacts were removed, to decrease the memory usage and execussion time. This task is expected to be finished in mid 2014.

Appendix B

Control plots for $\Delta\phi$ in slices of $\Delta\eta$

In this appendix, the control plots for the double differencial distributions in the different scenarios are presented.

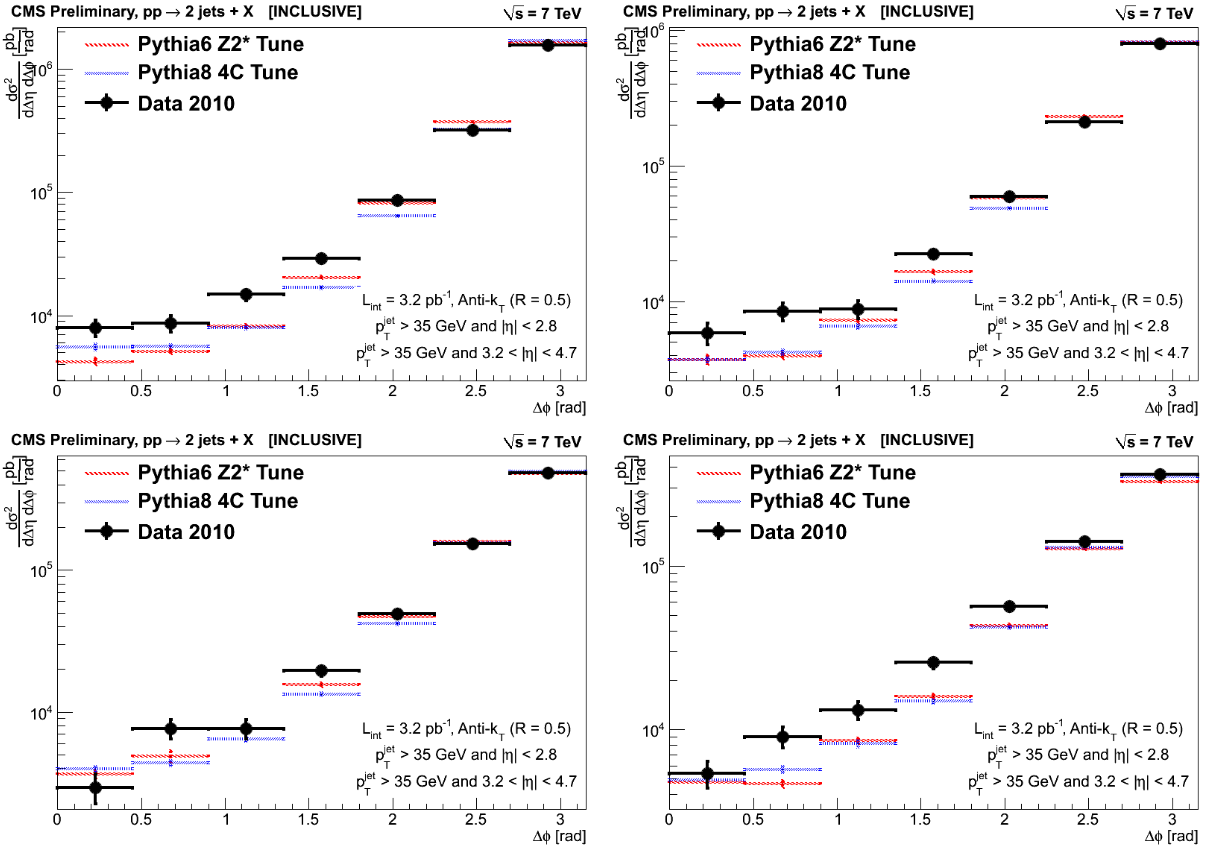


Figure B.1: Control plots for the $\Delta\phi$ distribution in different slices of $\Delta\eta$ in the inclusive scenario: $0.4 < \Delta\eta \leq 2.5$ (top–left), $2.5 < \Delta\eta \leq 3.5$ (top–right), $3.5 < \Delta\eta \leq 4.5$ (bottom–left) and $4.5 < \Delta\eta \leq 7.5$ (bottom–right).

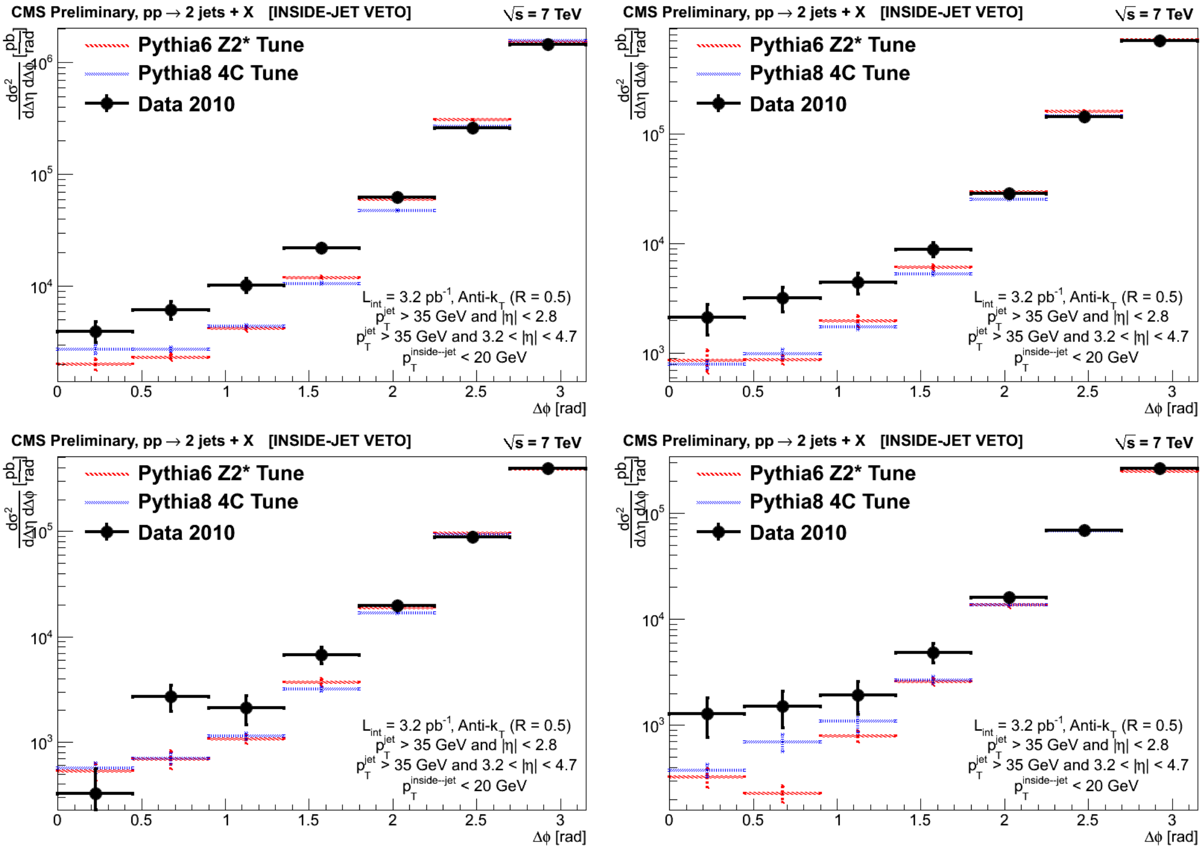


Figure B.2: Control plots for the $\Delta\phi$ distribution in different slices of $\Delta\eta$ in the inside–jet veto scenario: $0.4 < \Delta\eta \leq 2.5$ (top–left), $2.5 < \Delta\eta \leq 3.5$ (top–right), $3.5 < \Delta\eta \leq 4.5$ (bottom–left) and $4.5 < \Delta\eta \leq 7.5$ (bottom–right).

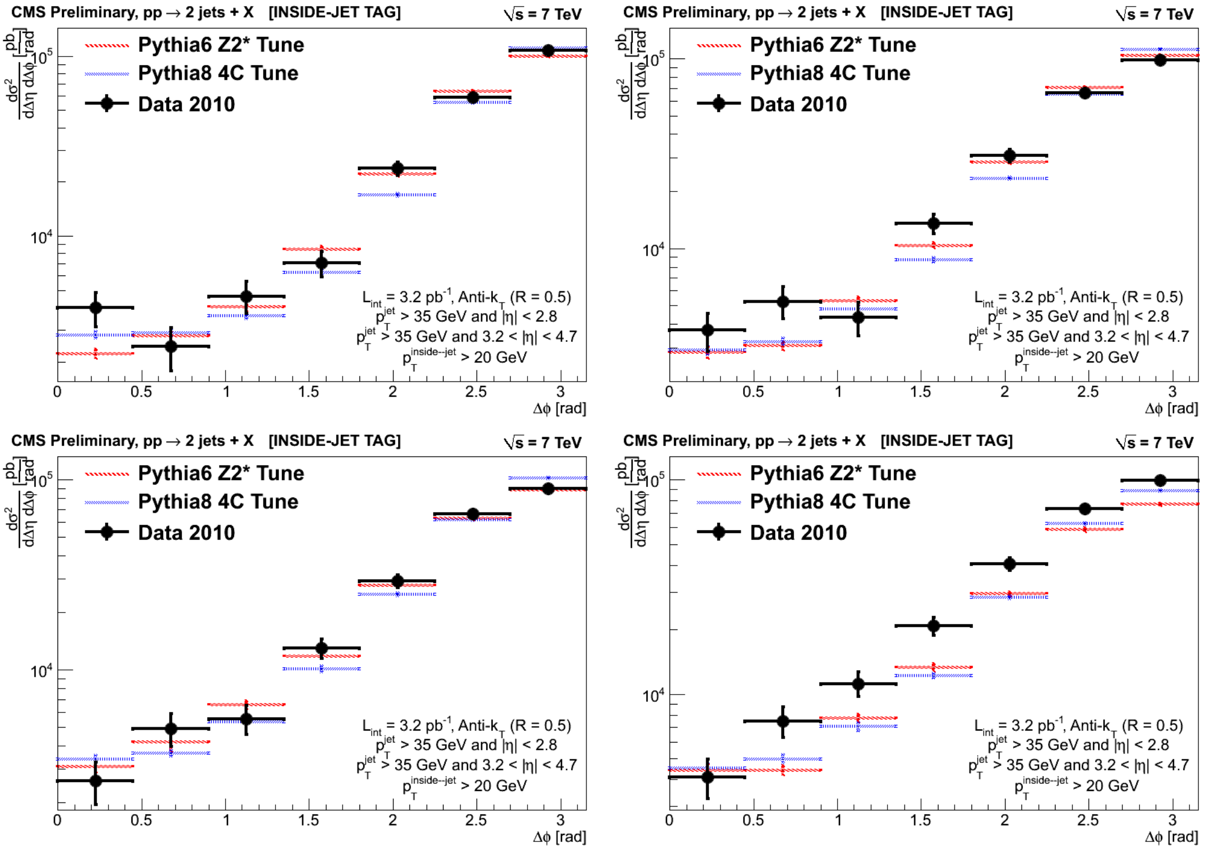


Figure B.3: Control plots for the $\Delta\phi$ distribution in different slices of $\Delta\eta$ in the inside–jet tag scenario: $0.4 < \Delta\eta \leq 2.5$ (top–left), $2.5 < \Delta\eta \leq 3.5$ (top–right), $3.5 < \Delta\eta \leq 4.5$ (bottom–left) and $4.5 < \Delta\eta \leq 7.5$ (bottom–right).

Appendix C

Trigger Efficiency for the Trigger Combination Using the Division Method

In this appendix, the efficiency for the trigger combination is presented for the different scenarions in $\Delta\eta$ slices.

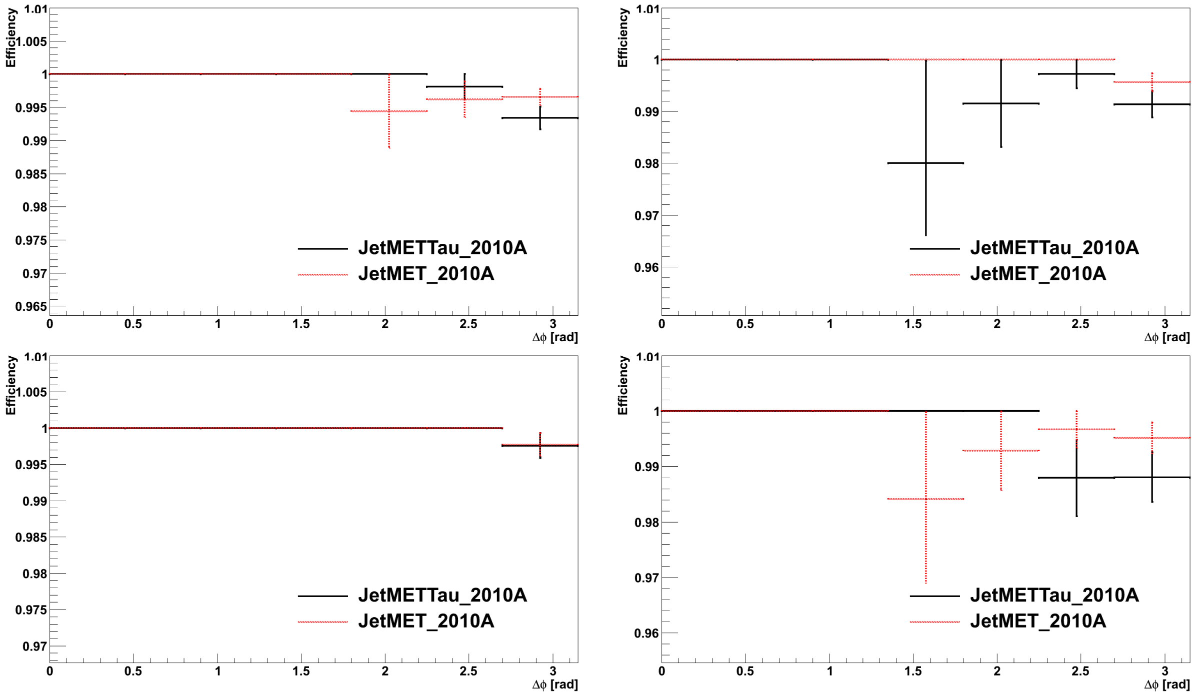


Figure C.1: Efficiency of the trigger combination for the main scenario as function of $\Delta\eta$ for the JetMETTau_2010A and JetMET_2010A datasets.

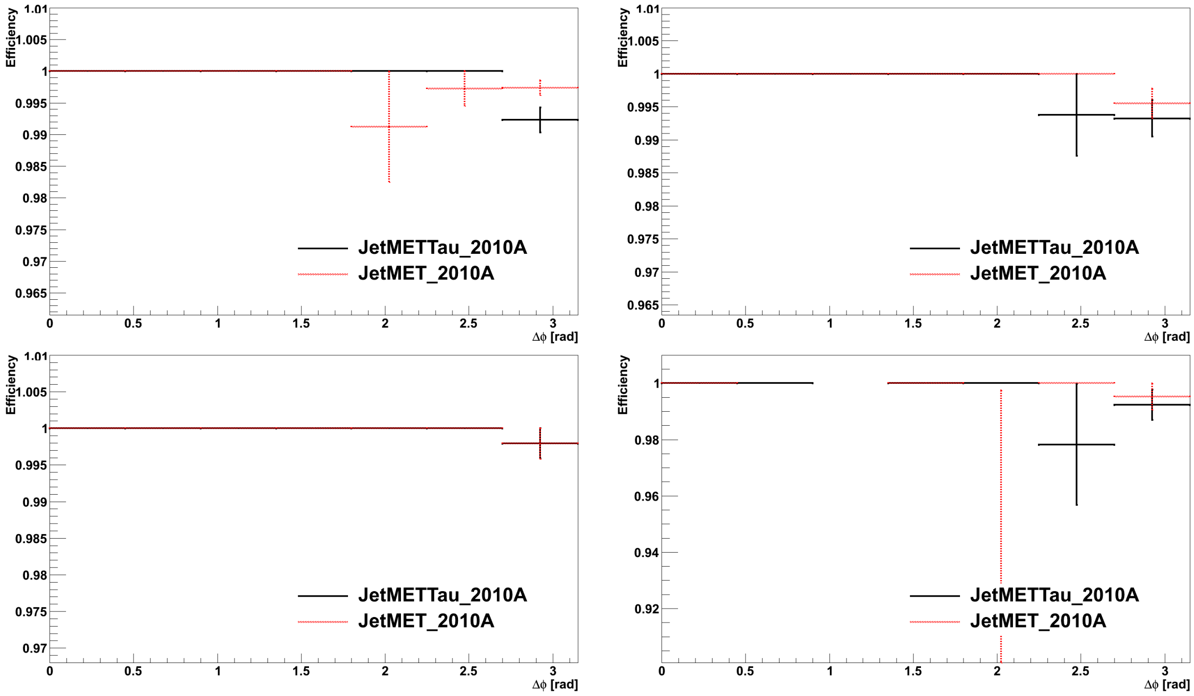


Figure C.2: Efficiency of the trigger combination for the inside–jet veto scenario as function of $\Delta\eta$ for the JetMETTau_2010A and JetMET_2010A datasets.

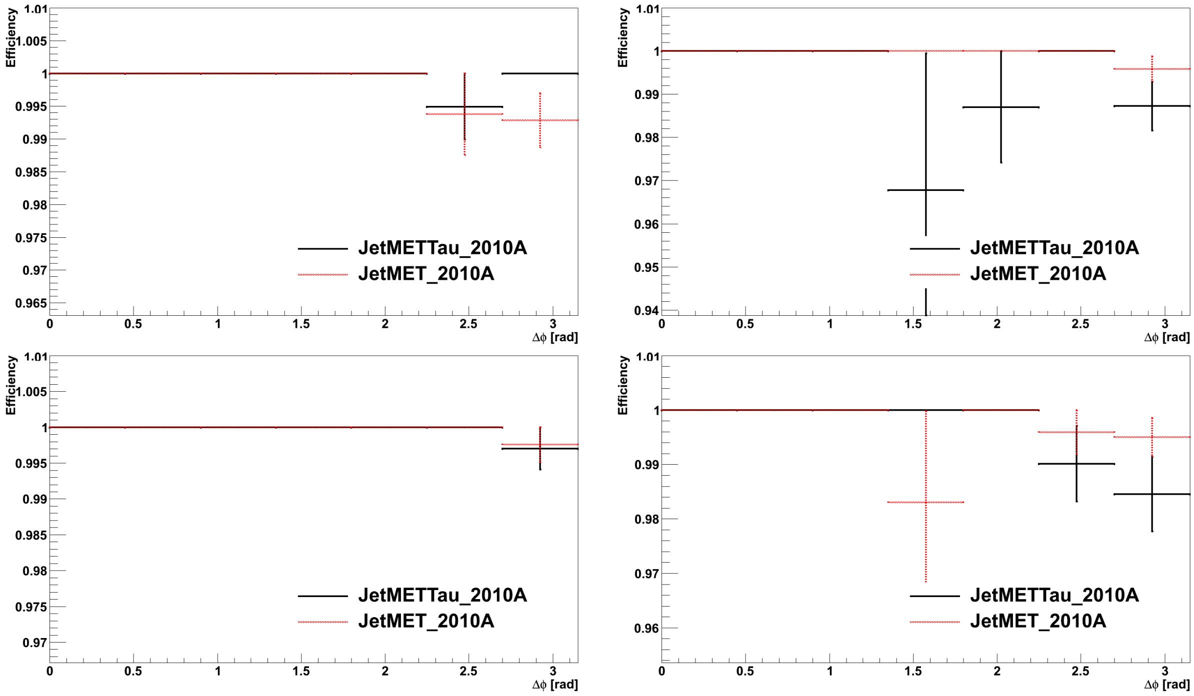


Figure C.3: Efficiency of the trigger combination for the inside–jet tag scenario as function $\Delta\eta$ for the JetMETTau_2010A and JetMET_2010A datasets.

Appendix D

Resolution for the $\Delta\eta$ Slices

In this appendix, the $\Delta\phi$ resolution in $\Delta\eta$ slices is presented for the different scenarios.

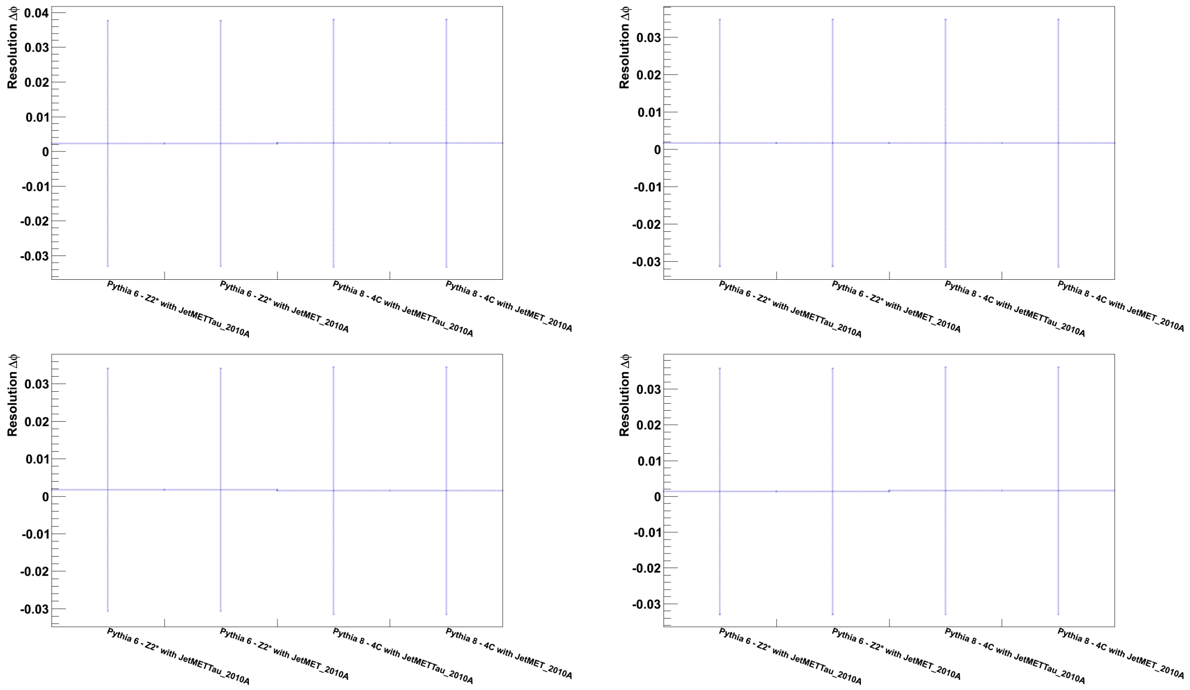


Figure D.1: $\Delta\phi$ resolution between the leading central and the leading forward jets in the inclusive scenario for different $\Delta\eta$ slices: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right)

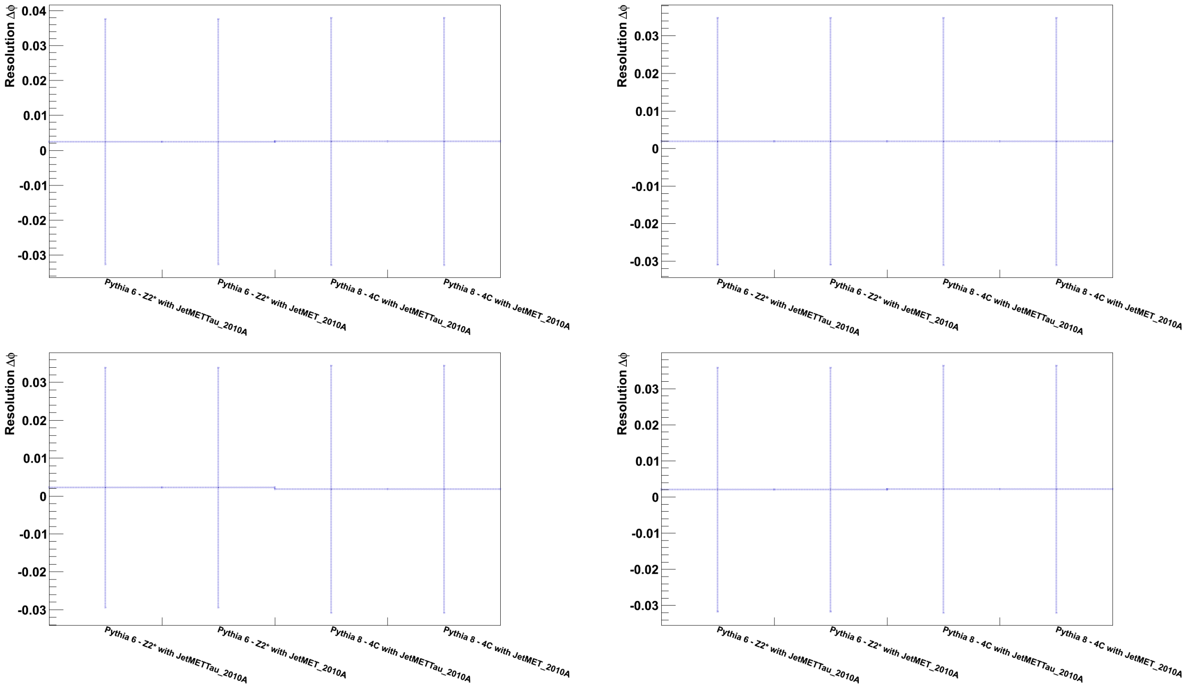


Figure D.2: $\Delta\phi$ resolution between the leading central and the leading forward jets for the inside–jet veto scenario in different $\Delta\eta$ slices: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right)

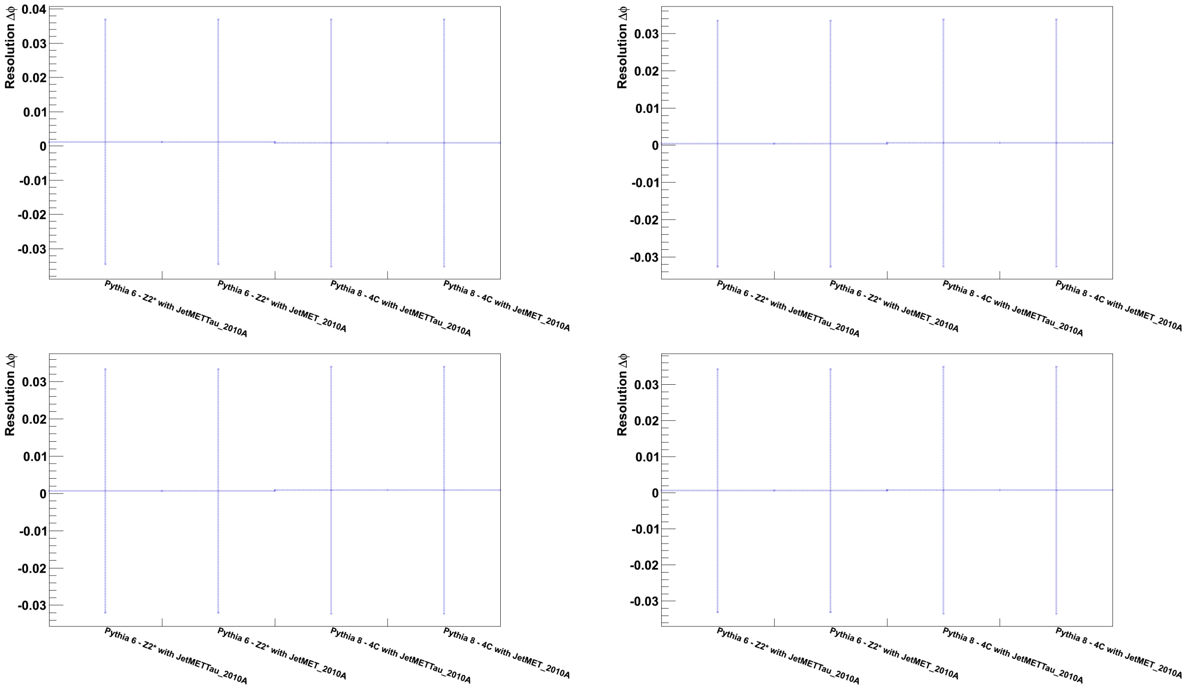


Figure D.3: $\Delta\phi$ resolution between the leading central and the leading forward jets for the inside–jet tag scenario in different $\Delta\eta$ slices: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right)

Appendix E

Acceptance, Background, Purity and Stability in Slices of $\Delta\eta$

In this appendix, the acceptance, background, purity and stability as function of $\Delta\phi$ in $\Delta\eta$ slices is presented for the different scenarios performed on the JetMETTau_2010A datasample.

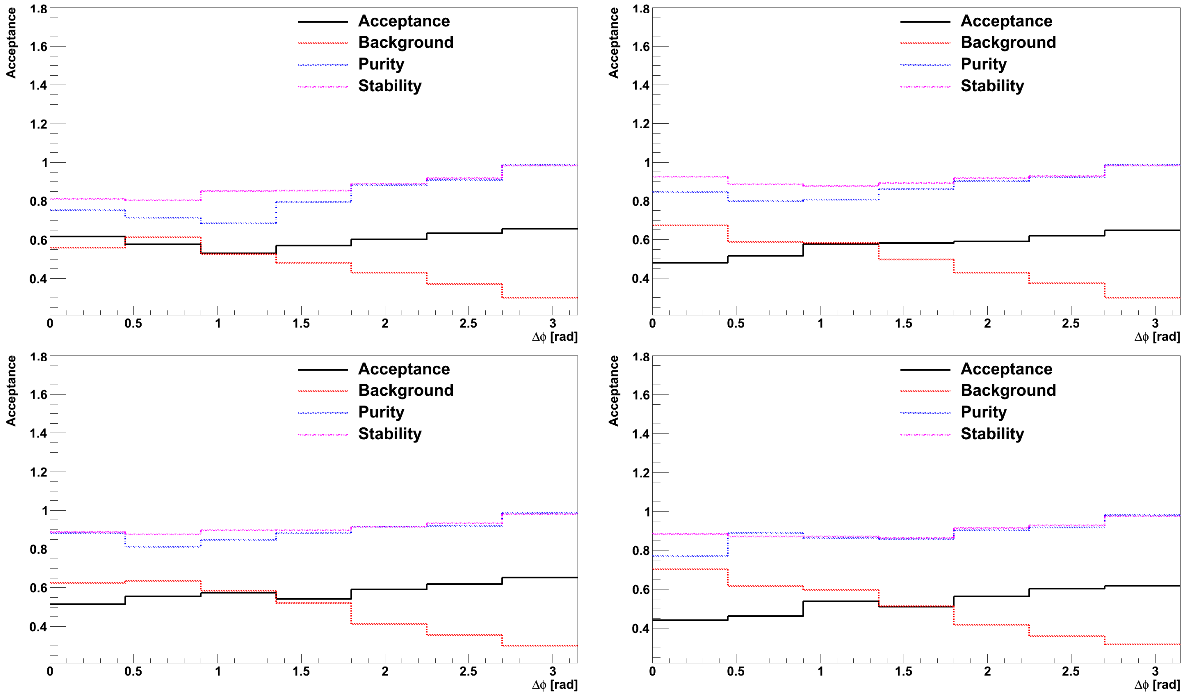


Figure E.1: Acceptance, background, purity and stability for the bin–by–bin correction with Pythia 6 - tune Z2* in the main scenario for different $\Delta\eta$ slices.

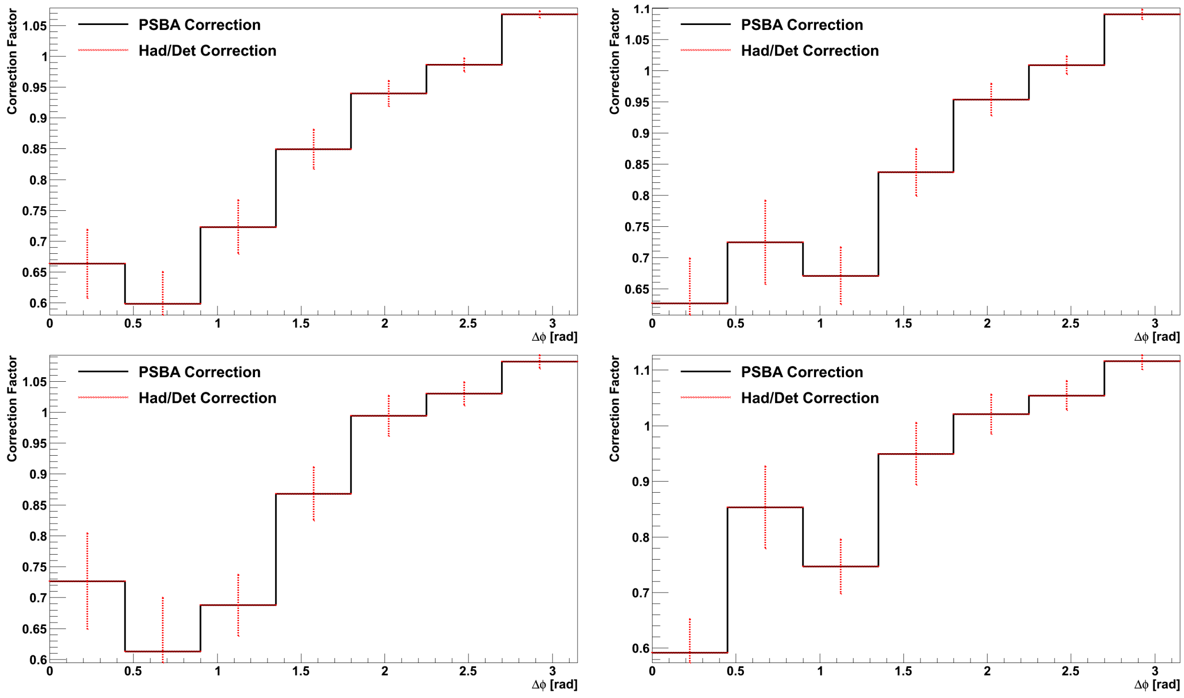


Figure E.2: Closure test for the bin–by–bin correction performed on the JetMETTau_2010A datasample corrected with Pythia 6 - tune Z2* in the main scenario for different $\Delta\eta$ slices.

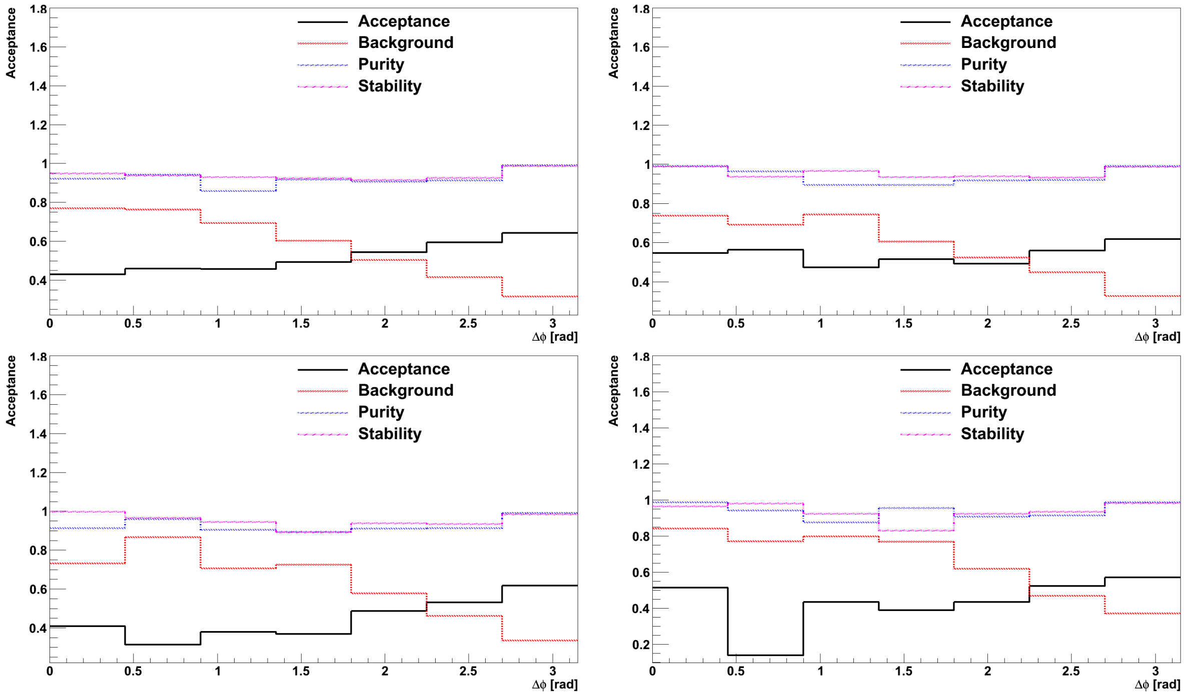


Figure E.3: Acceptance, background, purity and stability for the bin–by–bin correction with Pythia 6 - tune Z2* in the inside–jet veto scenario for different $\Delta\eta$ slices.

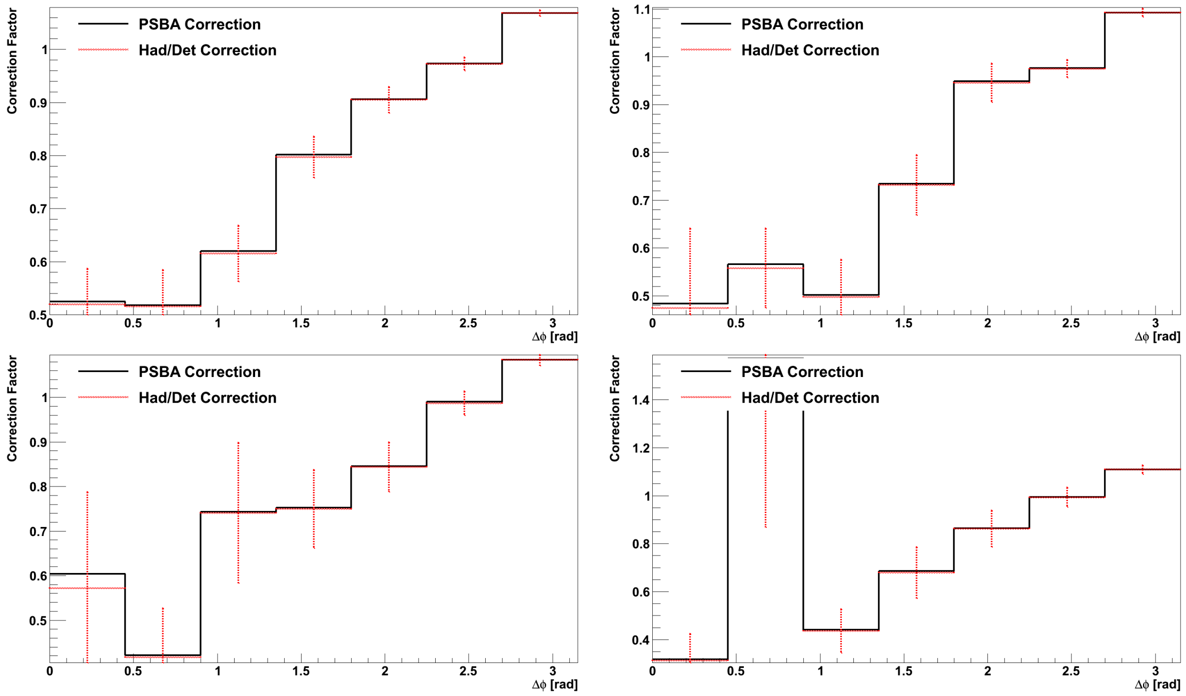


Figure E.4: Closure test for the bin–by–bin correction with Pythia 6 - tune Z2* in the inside–jet veto scenario for different $\Delta\eta$ slices.

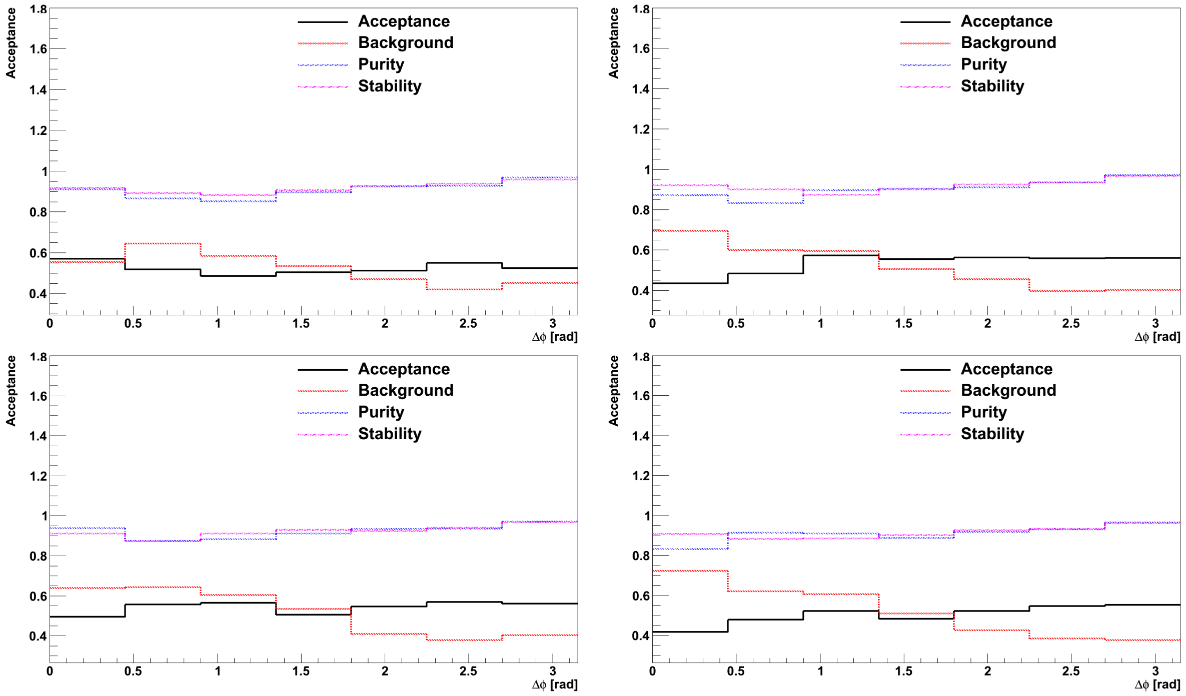


Figure E.5: Acceptance, background, purity and stability for the bin–by–bin correction with Pythia 6 - tune Z2* in the inside–jet tag scenario for different $\Delta\eta$ slices.

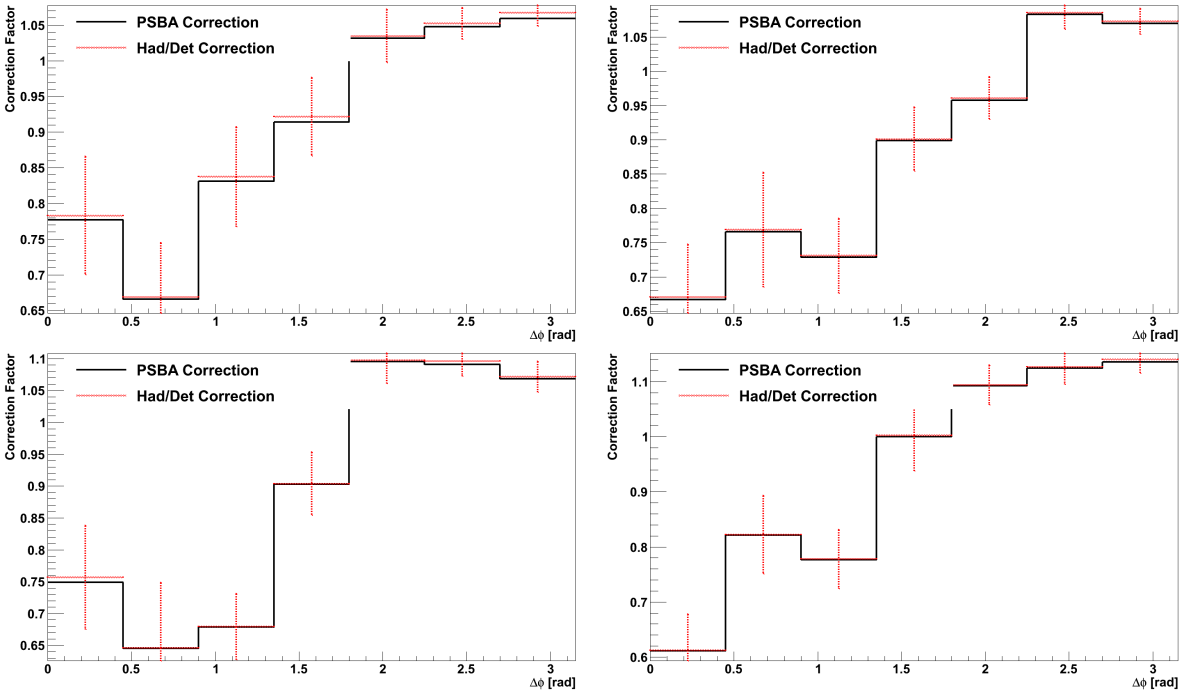


Figure E.6: Closure test for the bin–by–bin correction with Pythia 6 - tune Z2* in the inside–jet tag scenario for different $\Delta\eta$ slices.

Appendix F

Correction Factors

In this appendix, the correction factors for $\Delta\phi$ as function of $\Delta\eta$ for the different scenarios for the JetMETTau_2010A dataset are presented.

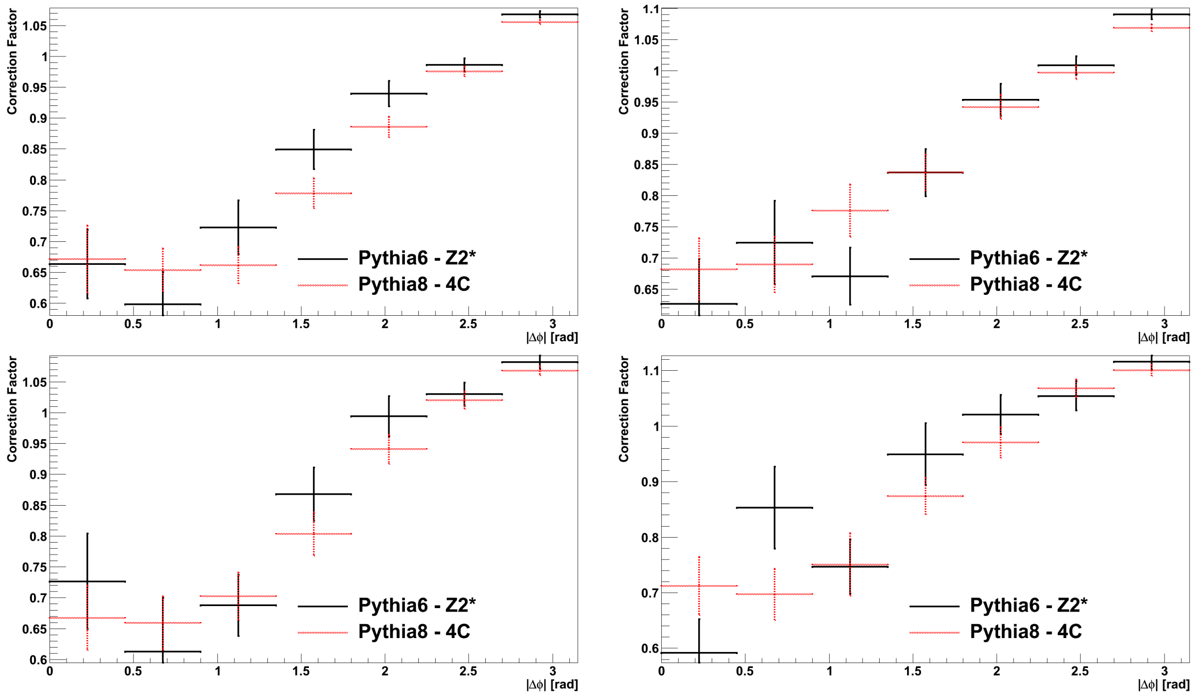


Figure F.1: Stable particle level correction factors for $\Delta\phi$ in the main selection in different $\Delta\eta$ slices: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right)

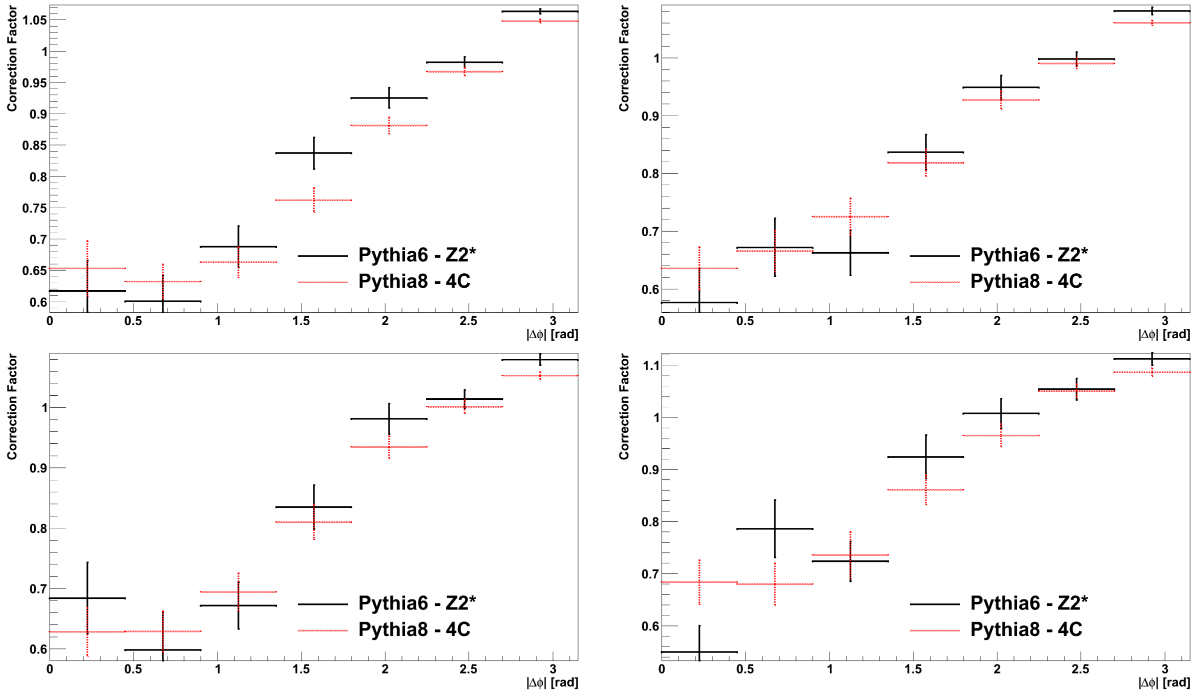


Figure F.2: Stable particle level correction factors for $\Delta\phi$ in the main selection in different $\Delta\eta$ slices: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right

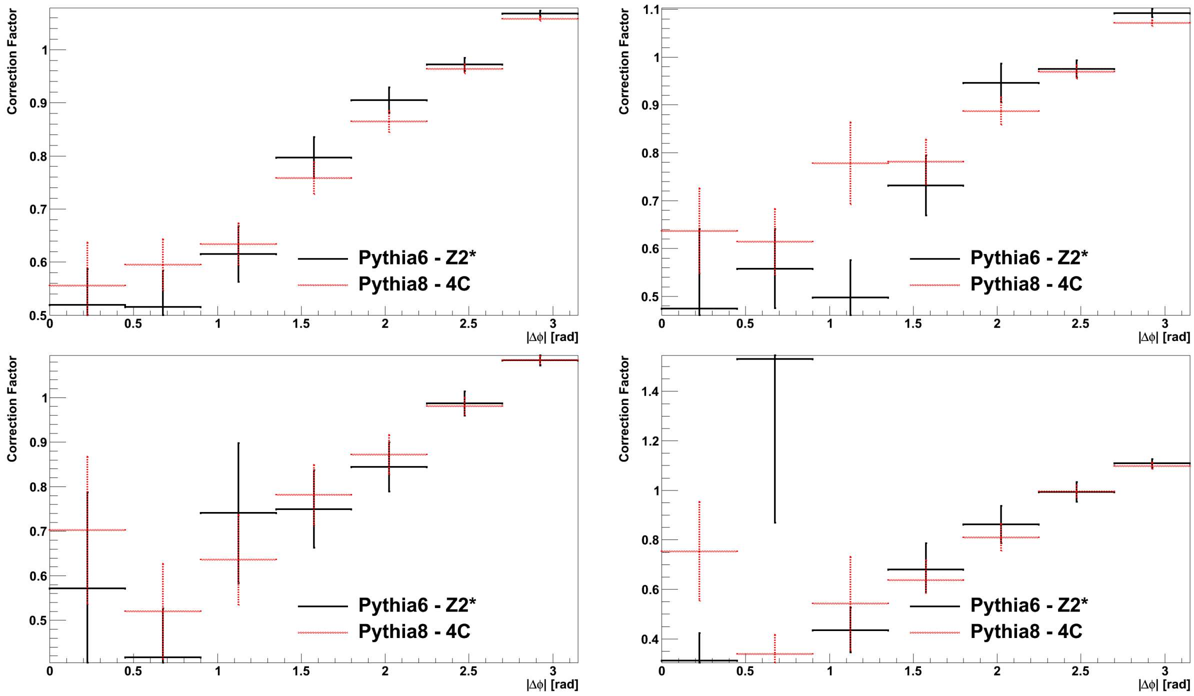


Figure F.3: Stable particle level correction factors for $\Delta\phi$ in the inside–jet veto scenario in different $\Delta\eta$ slices: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right)

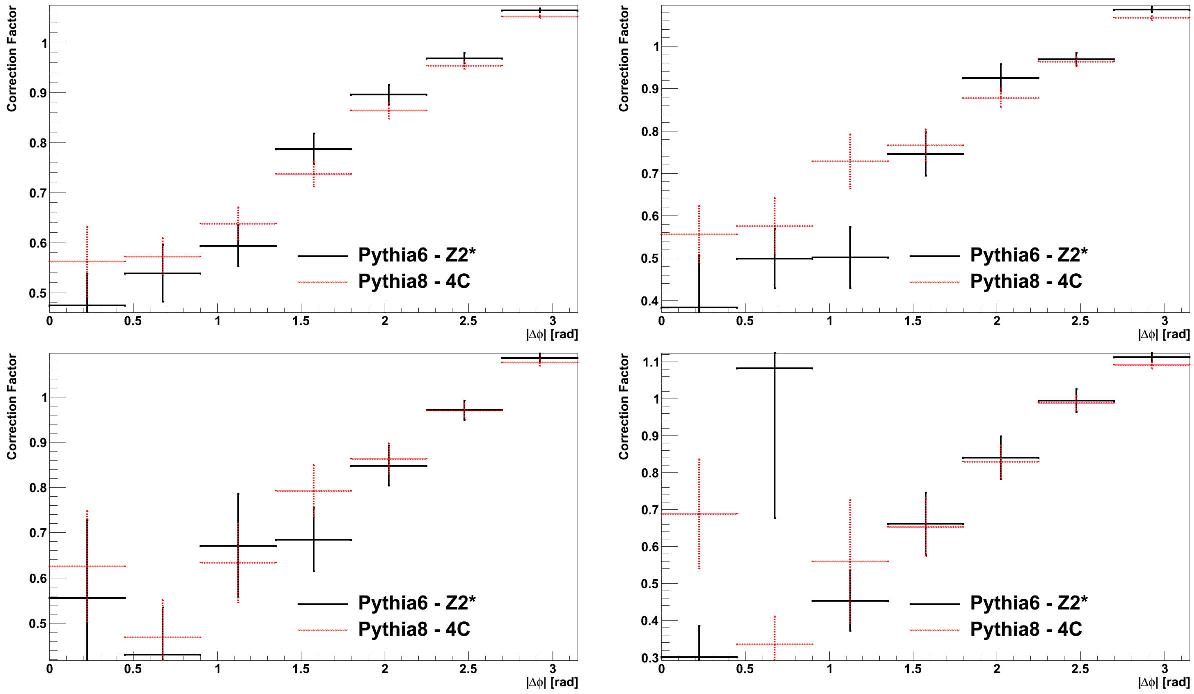


Figure F.4: Stable particle level correction factors for $\Delta\phi$ in the inside–jet veto scenario in different $\Delta\eta$ slices: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right)

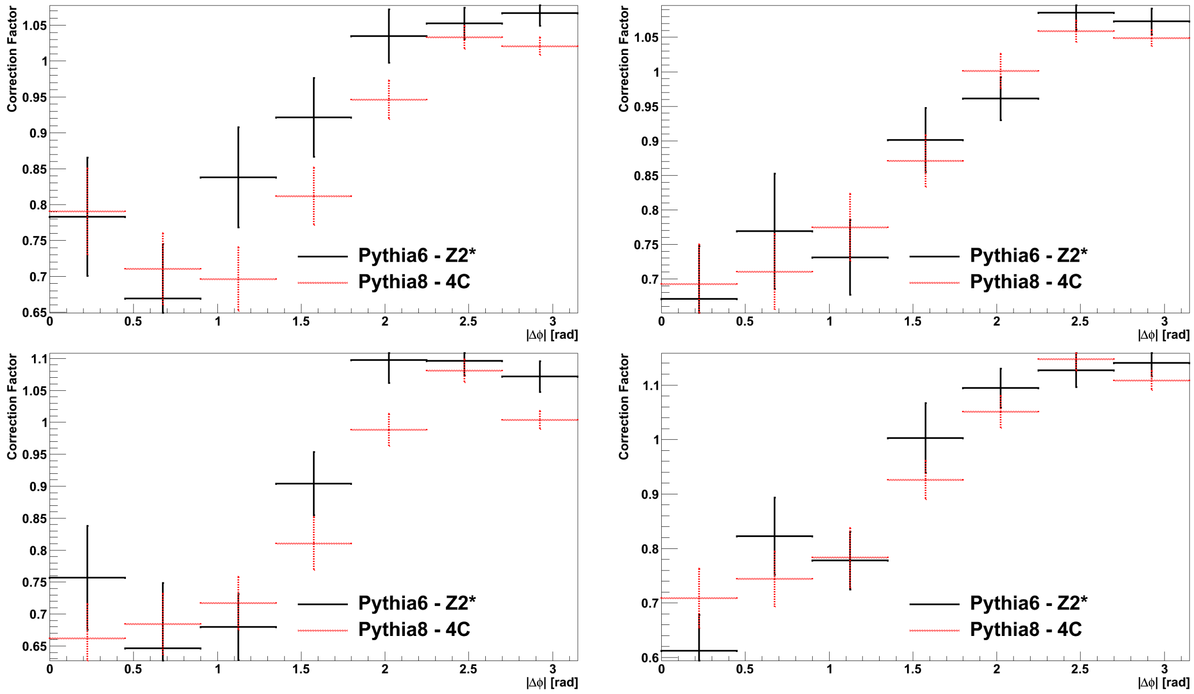


Figure F.5: Stable particle level correction factors for $\Delta\phi$ in the inside–jet tag scenario in different $\Delta\eta$ slices: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right)

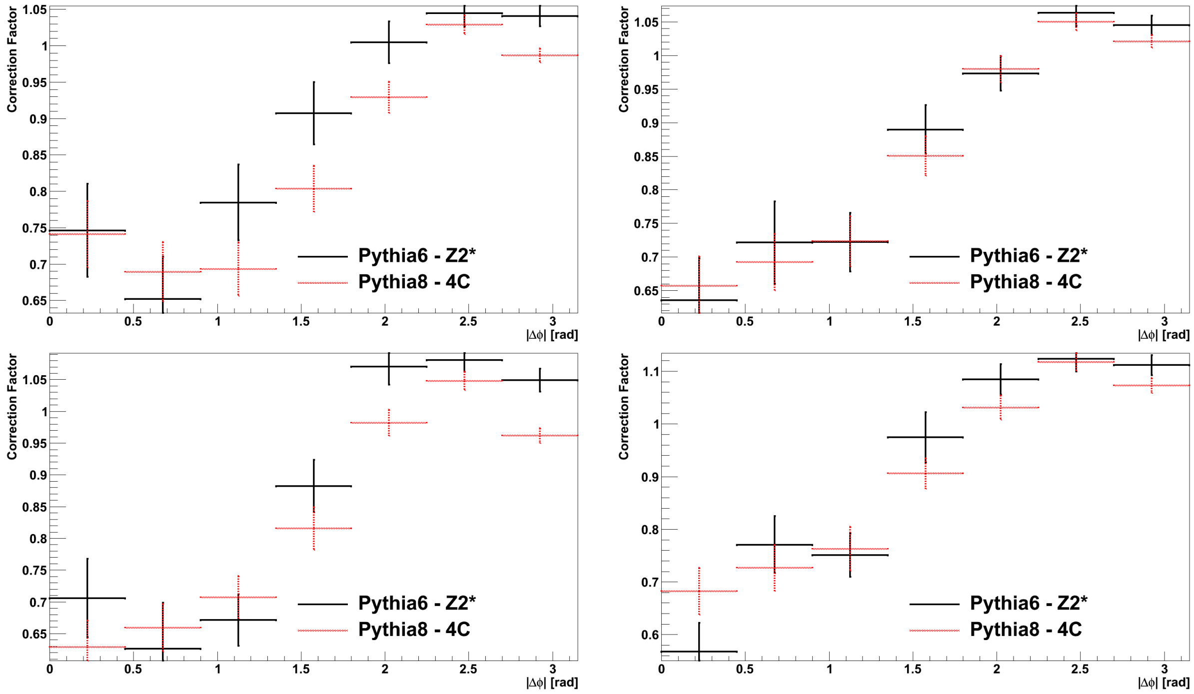


Figure F.6: Stable particle level correction factors for $\Delta\phi$ in the inside–jet tag scenario in different $\Delta\eta$ slices: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right

Appendix G

Unfolding Results

In this appendix the unfolding results for $\Delta\phi$ in $\Delta\eta$ slices for the different scenarios are presented.

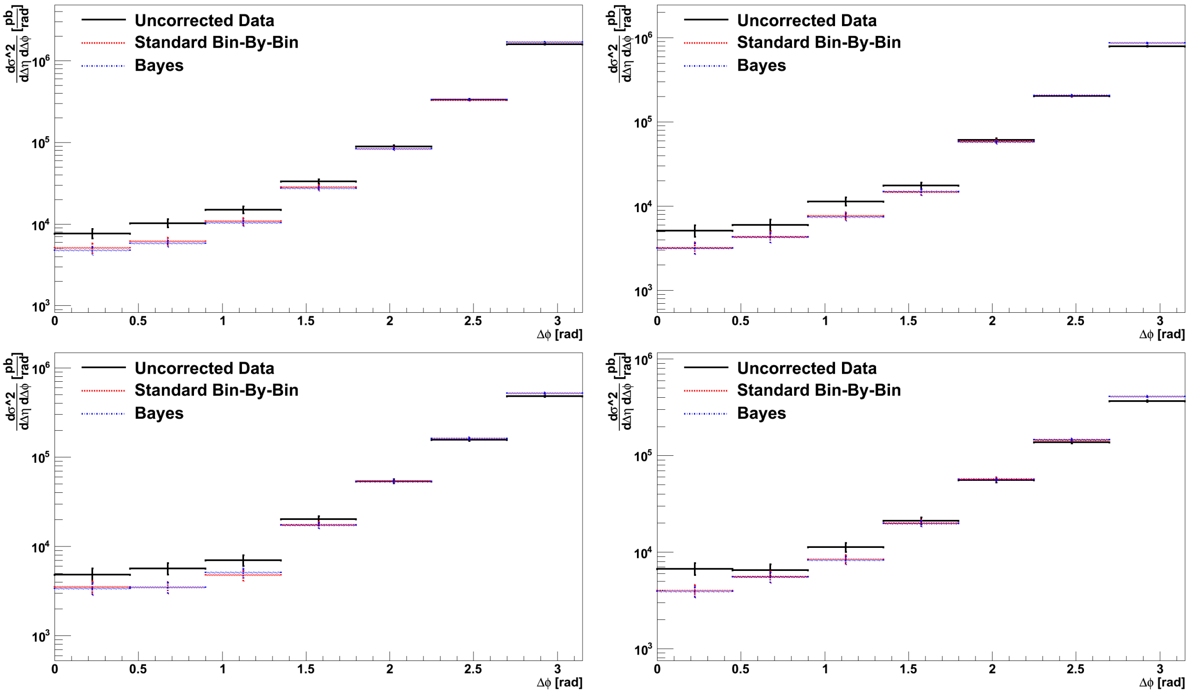


Figure G.1: Data unfolded with Pythia 6 – tune Z2* for $\Delta\phi$ in the inclusive scenario for the JetMETTau_2010A dataset in slices of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right

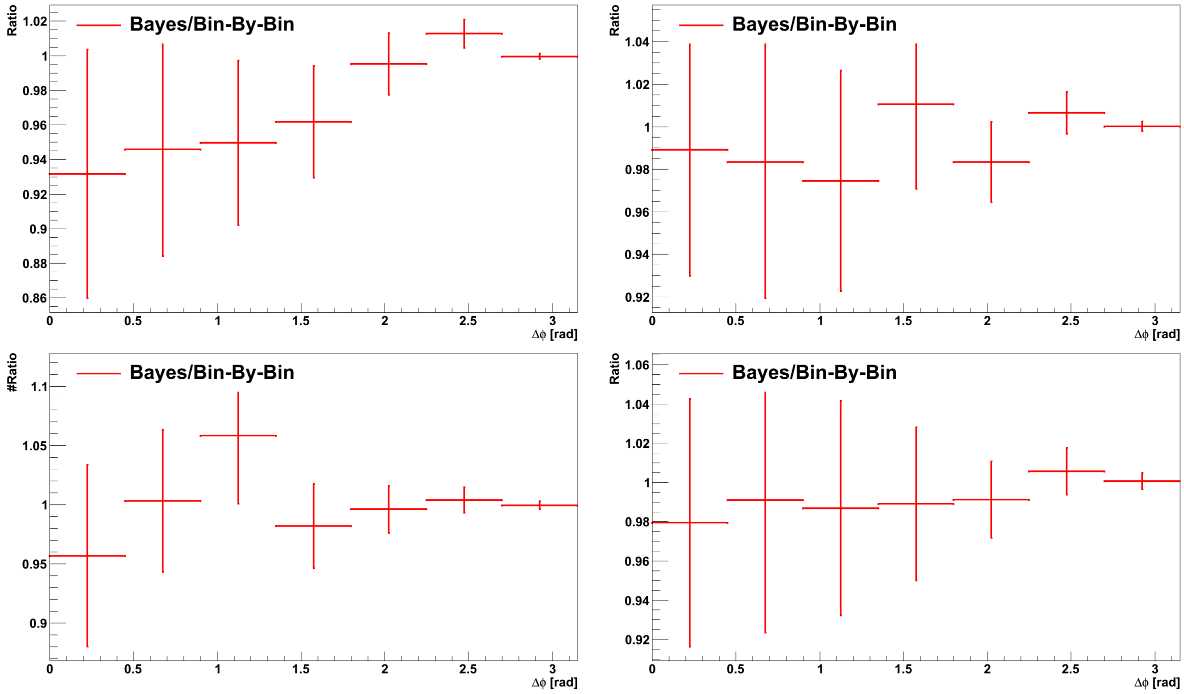


Figure G.2: Ratio of the unfolded data with Pythia 6 – tune Z2* for $\Delta\phi$ in the inclusive scenario for the JetMETTau_2010A dataset in slices of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right

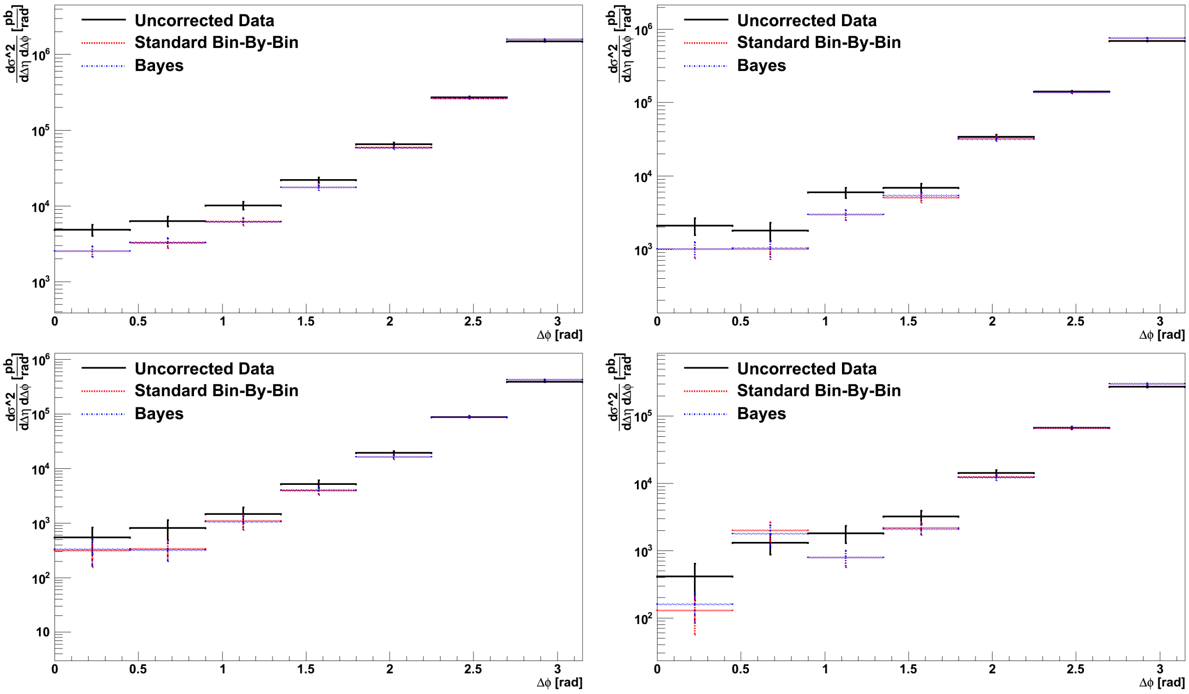


Figure G.3: Unfolded data with Pythia 6 – tune Z2* for $\Delta\phi$ in the inside–jet veto scenario for the JetMETTau_2010A dataset in slices of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right

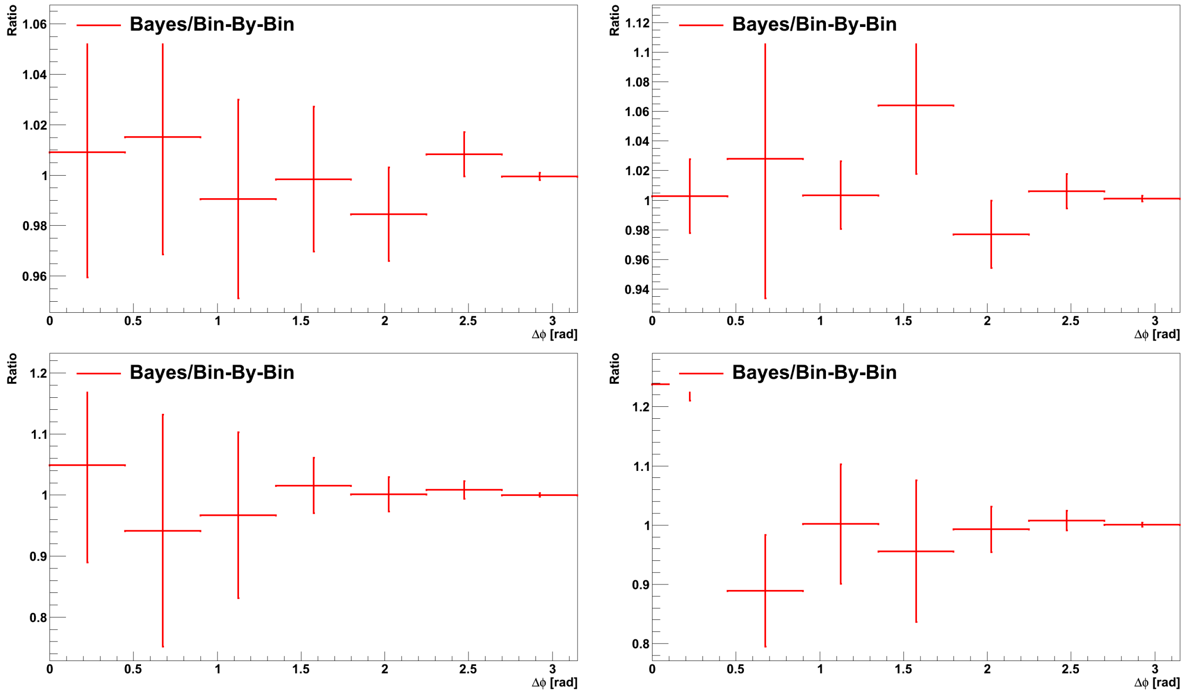


Figure G.4: Ratio of the unfolded data with Pythia 6 – tune Z2* for $\Delta\phi$ in the inside–jet veto scenario for the JetMETTau_2010A dataset in slices of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right

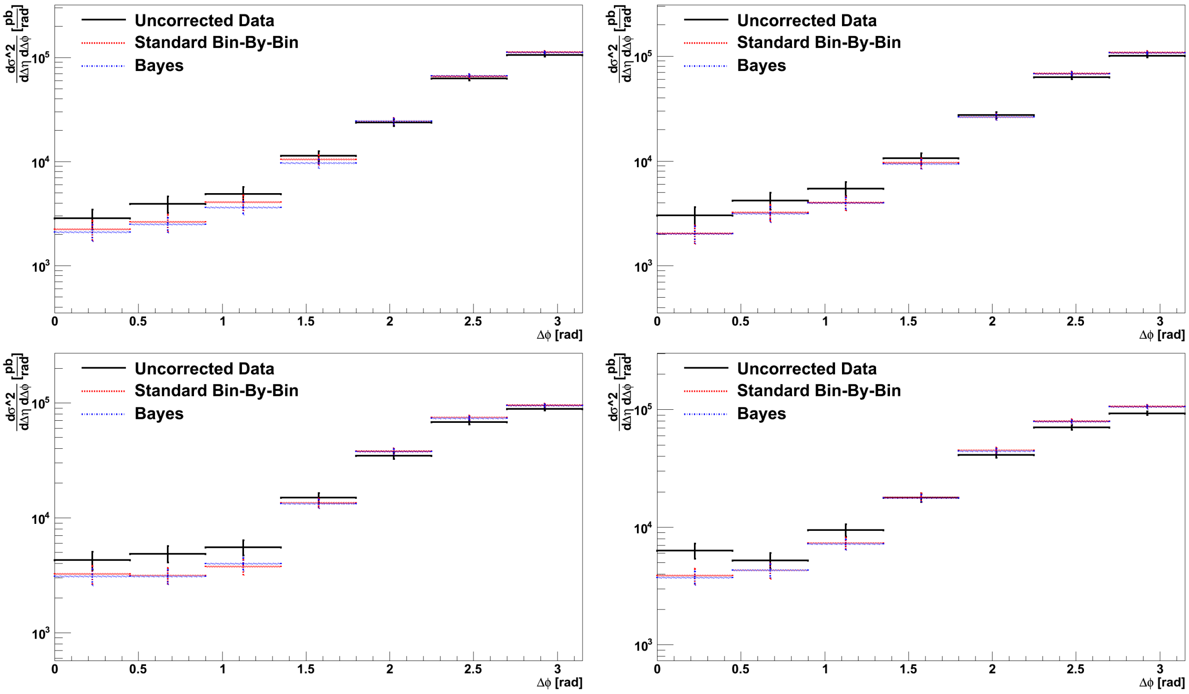


Figure G.5: Unfolded data with Pythia 6 – tune Z2* for $\Delta\phi$ in the inside–jet tag scenario for the JetMETTau_2010A dataset in slices of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right

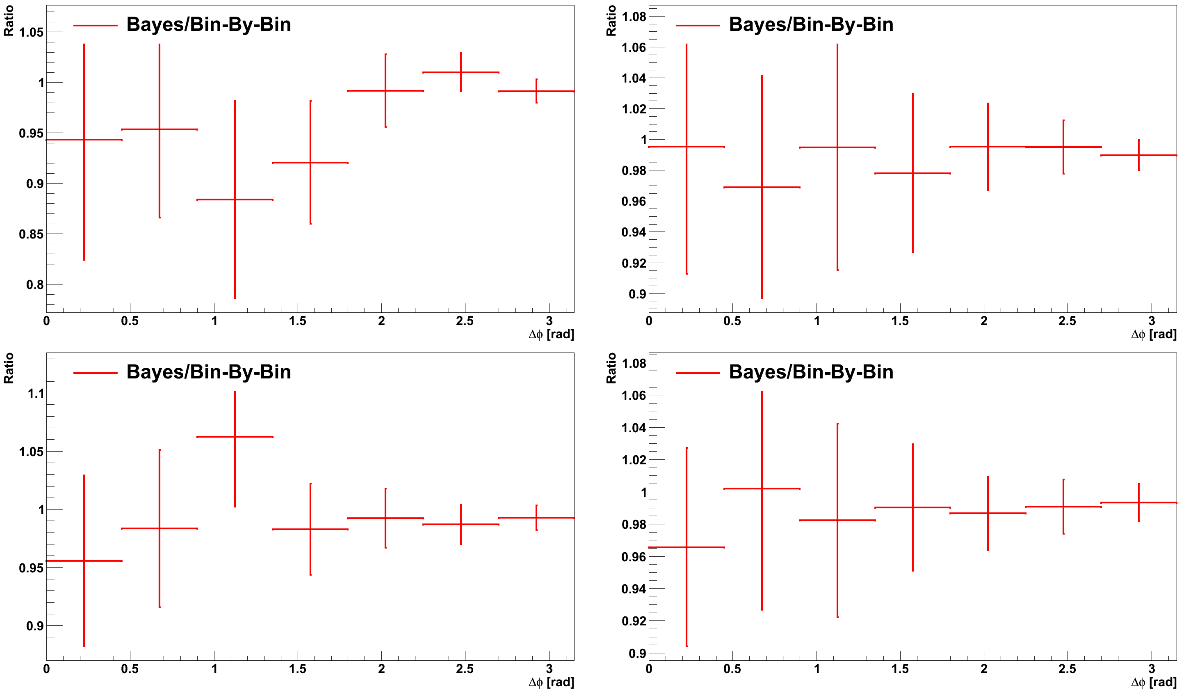


Figure G.6: Ratio of the unfolded data with Pythia 6 – tune Z2* for $\Delta\phi$ in the inside–jet tag scenario for the JetMETTau_2010A dataset in slices of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right

Appendix H

Statistical and Unfolding Uncertainties

In this appendix the statistical and the unfolding uncertainties for the $\Delta\phi$ distributions, as function of $\Delta\eta$, for the different scenarios are shown.

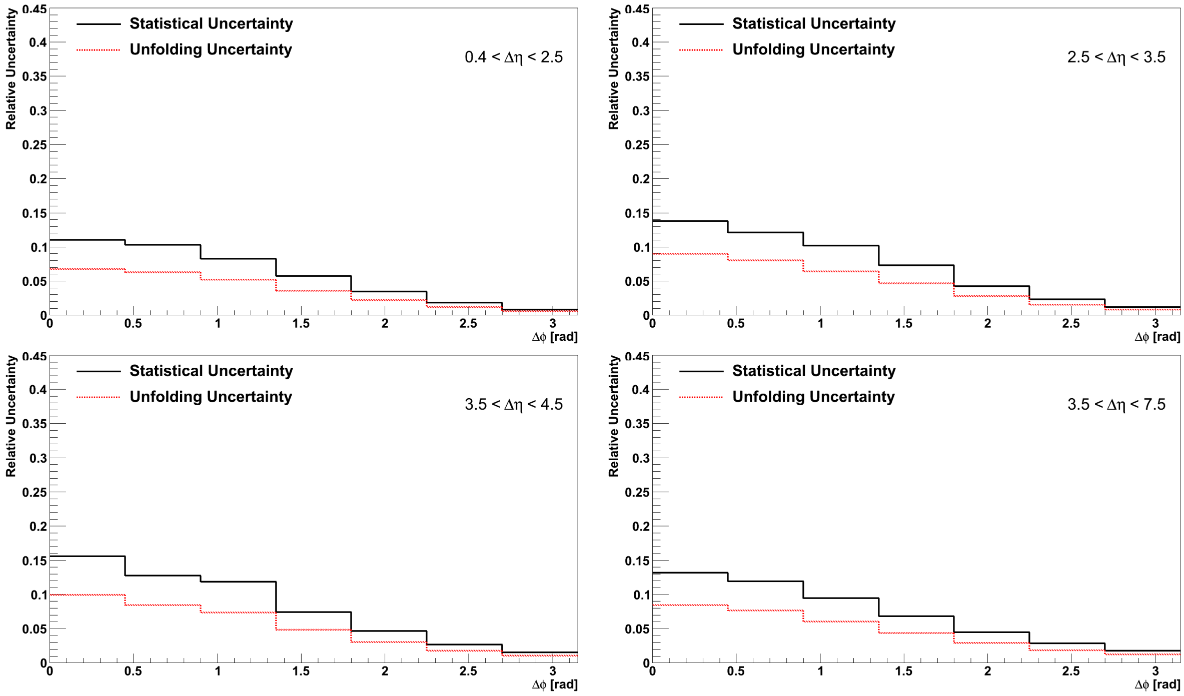


Figure H.1: Relative statistical (black) and unfolding (red) uncertainties of the $\Delta\phi$ distributions in the inclusive scenario as function of $\Delta\eta$.

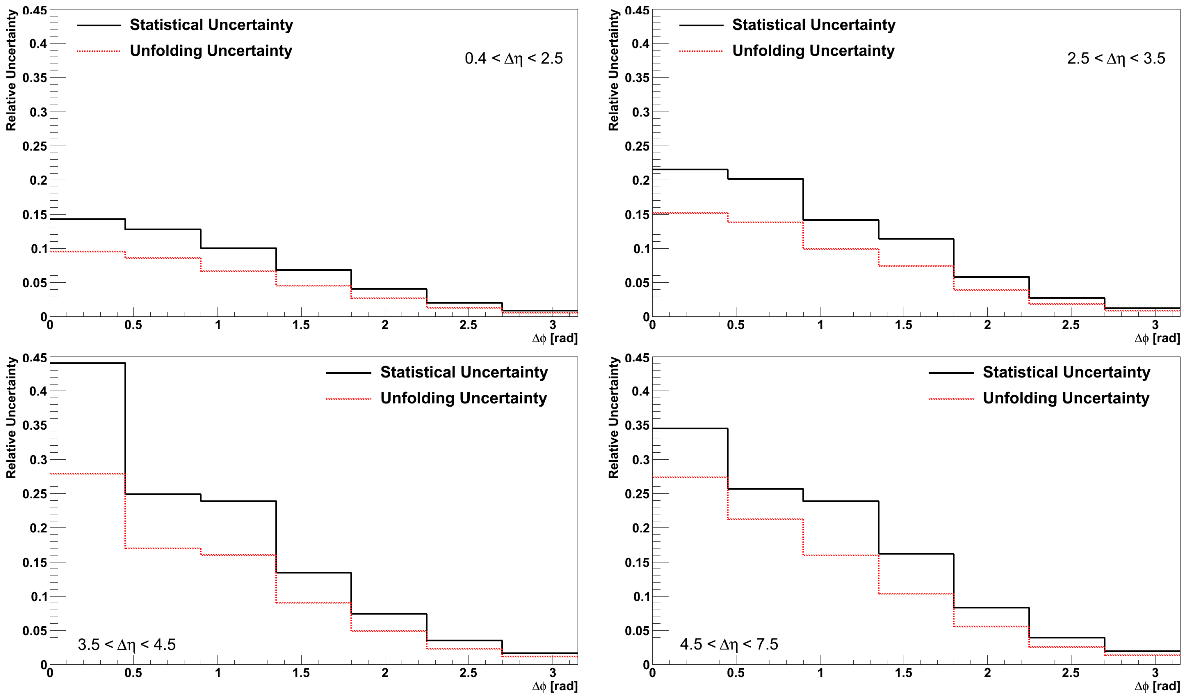


Figure H.2: Relative statistical (black) and unfolding (red) uncertainties of the $\Delta\phi$ distributions in the inside–jet veto scenario as function of $\Delta\eta$.

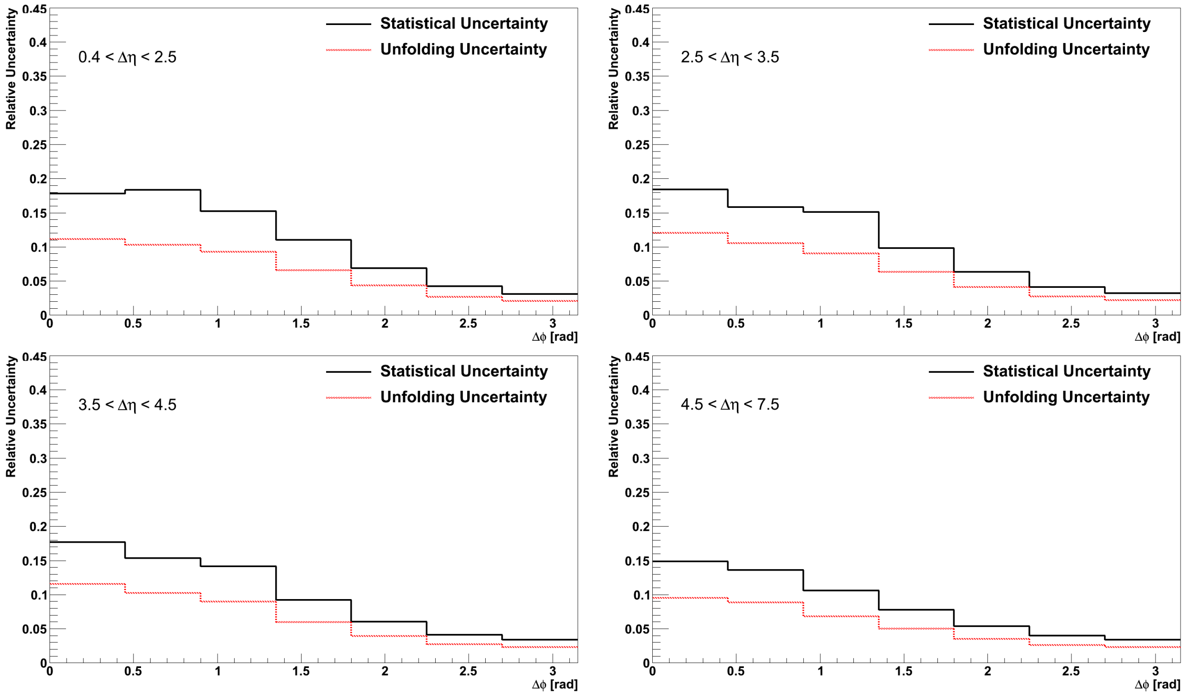


Figure H.3: Relative statistical (black) and unfolding (red) uncertainties of the $\Delta\phi$ distributions in the inside–jet tag scenario as function of $\Delta\eta$.

Appendix I

Correlated Uncertainties

In this appendix the correlated uncertainties for the $\Delta\phi$ distributions, as function of $\Delta\eta$, for the different scenarios are presented.

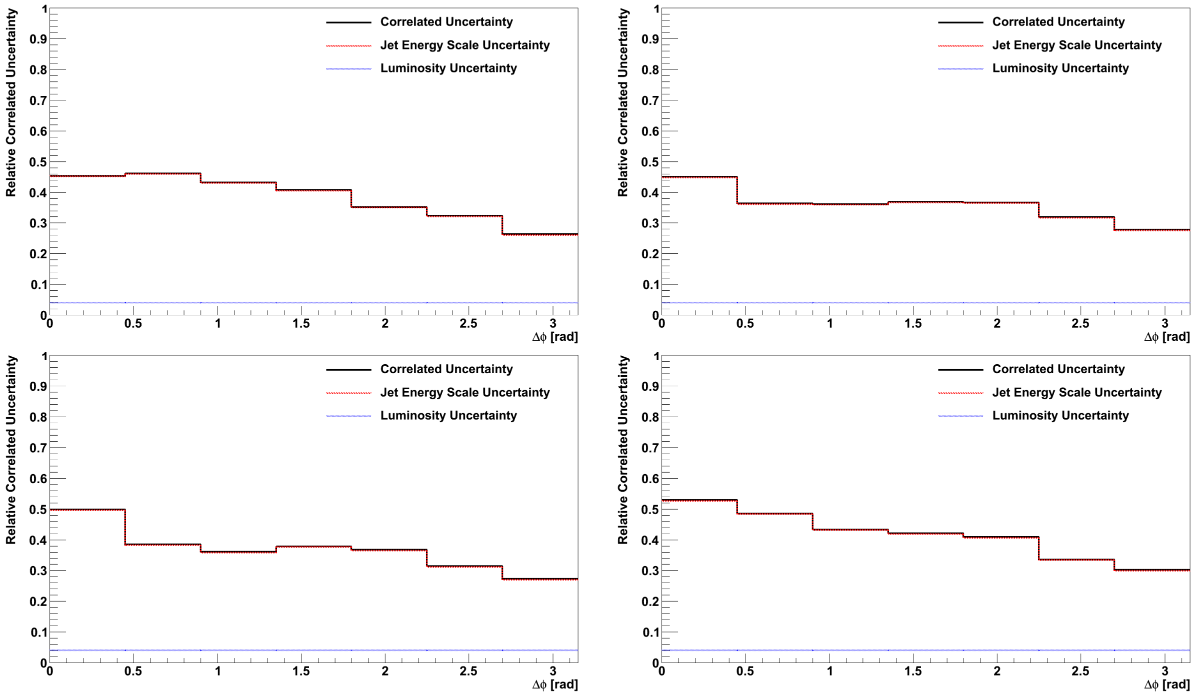


Figure I.1: Relative downwards correlated uncertainties of the $\Delta\phi$ distribution in the inclusive scenario as function of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right).

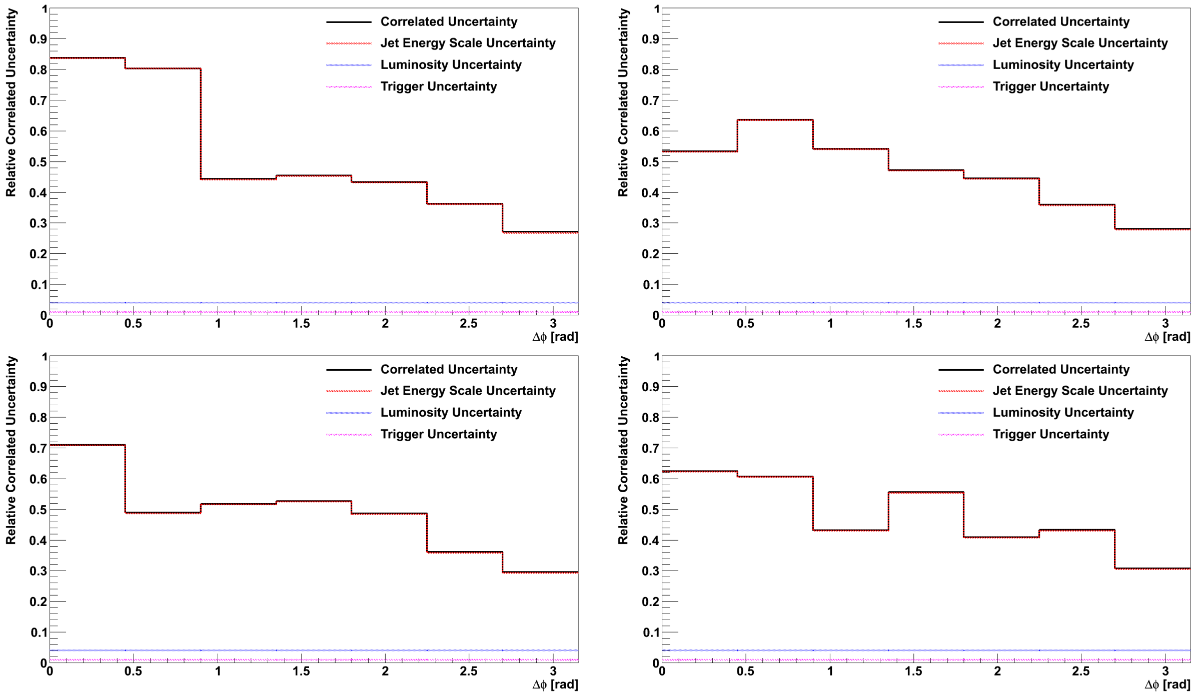


Figure I.2: Relative upwards correlated uncertainties of the $\Delta\phi$ distribution in the inclusive scenario as function of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right).

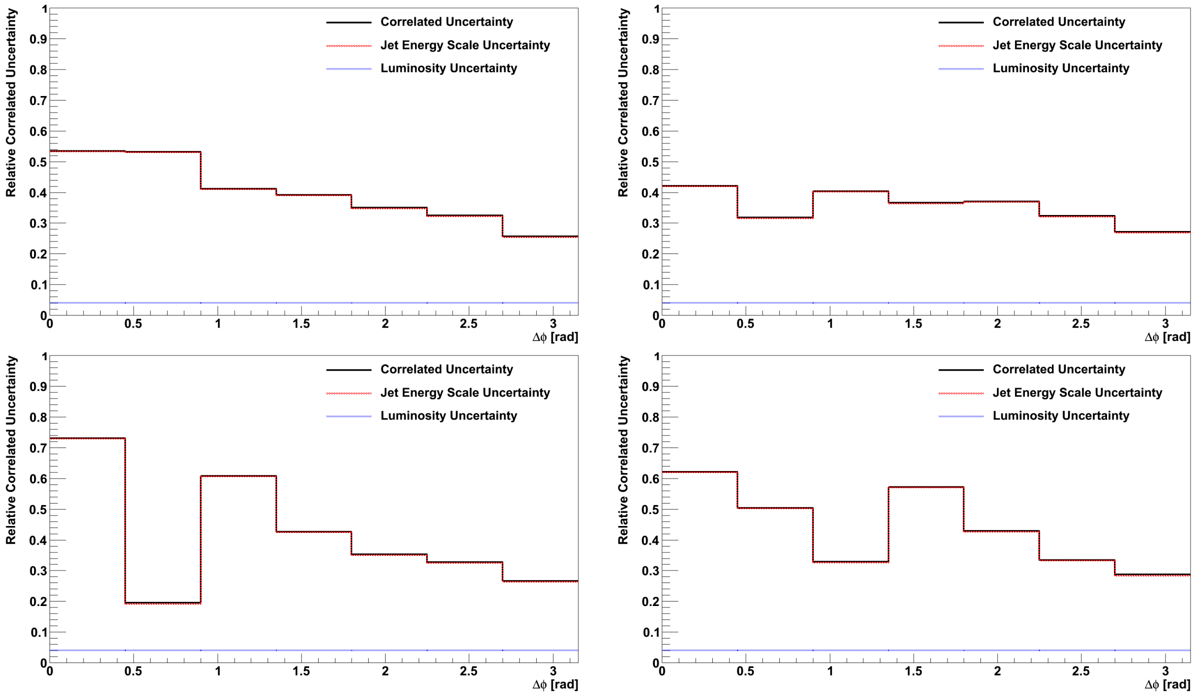


Figure I.3: Relative downwards correlated uncertainties of the $\Delta\phi$ distributions in the inside–jet veto scenario as function of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right).

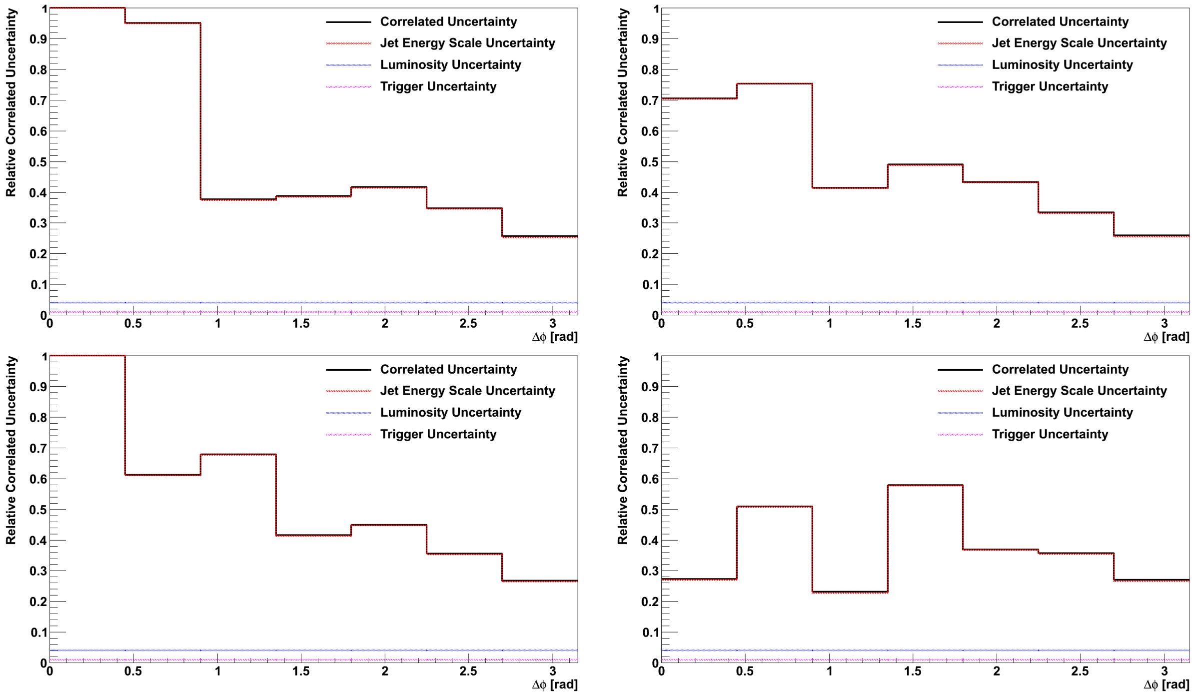


Figure I.4: Relative upwards correlated uncertainties of the $\Delta\phi$ distributions in the inside–jet veto scenario as function of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right).

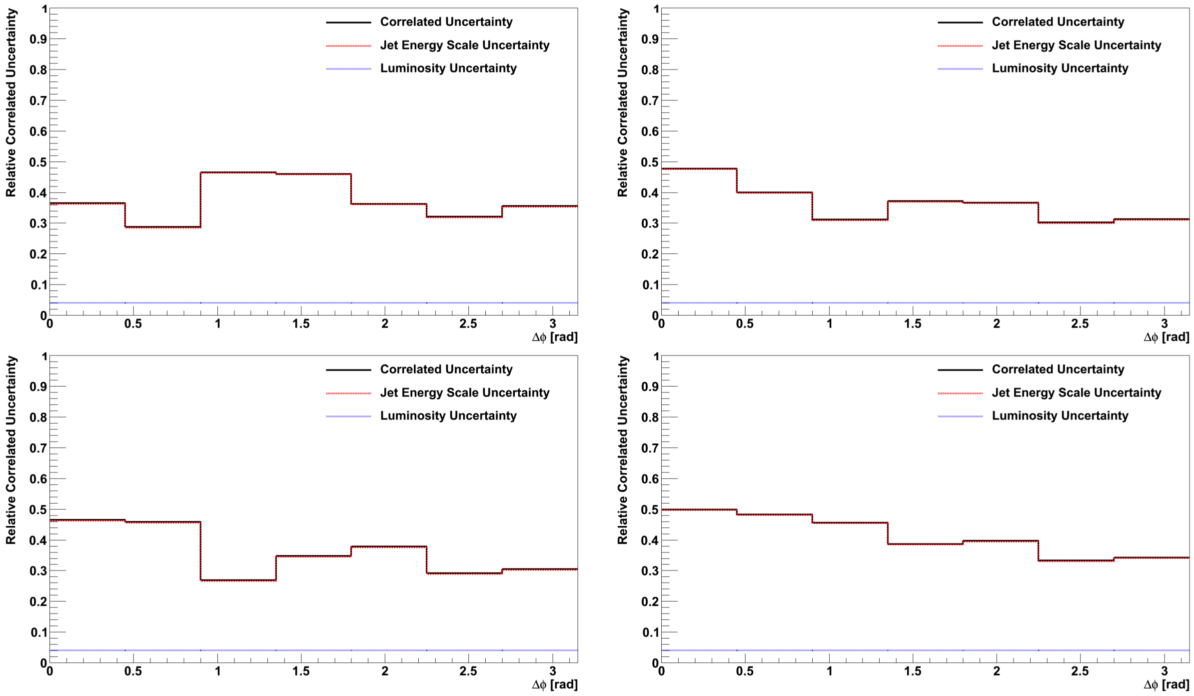


Figure I.5: Relative downwards correlated uncertainties of the $\Delta\phi$ distributions in the inside–jet tag scenario as function of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right).

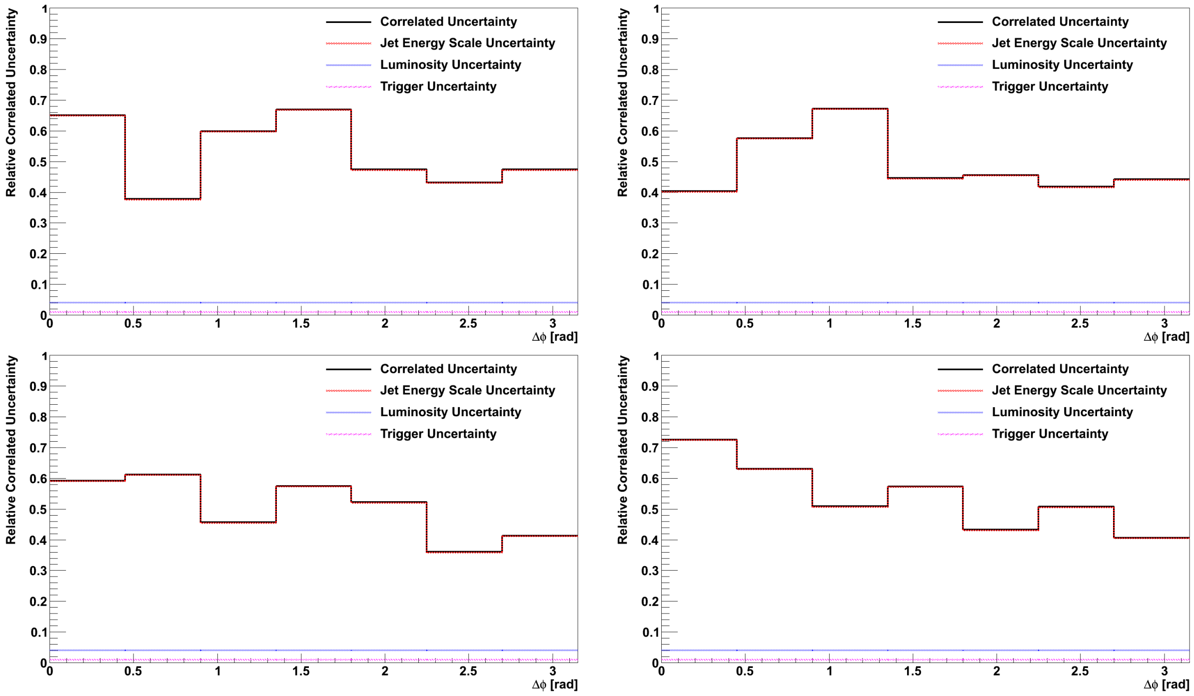


Figure I.6: Relative upwards correlated uncertainties of the $\Delta\phi$ distributions in the inside–jet tag scenario as function of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right).

Appendix J

Uncorrelated Uncertainties

In this appendix the uncorrelated uncertainties for the $\Delta\phi$ distributions, as function of $\Delta\eta$, for the different scenarios are presented.

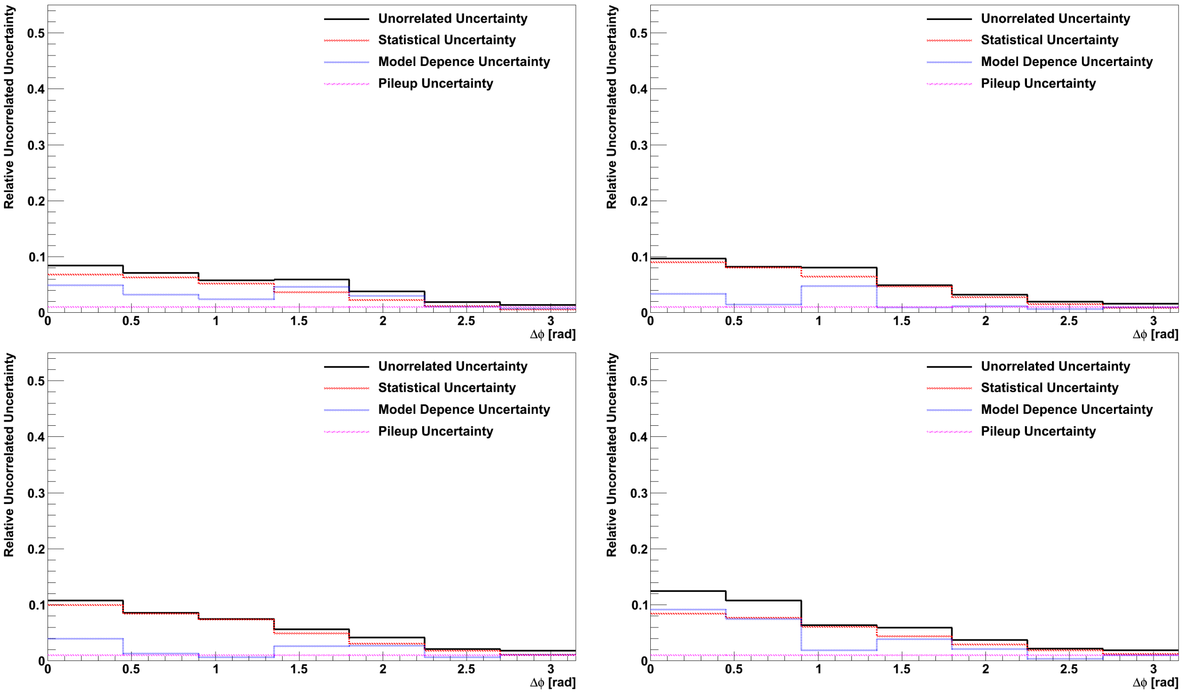


Figure J.1: Relative uncorrelated uncertainties of the $\Delta\phi$ distributions in the inclusive scenario as function of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right).

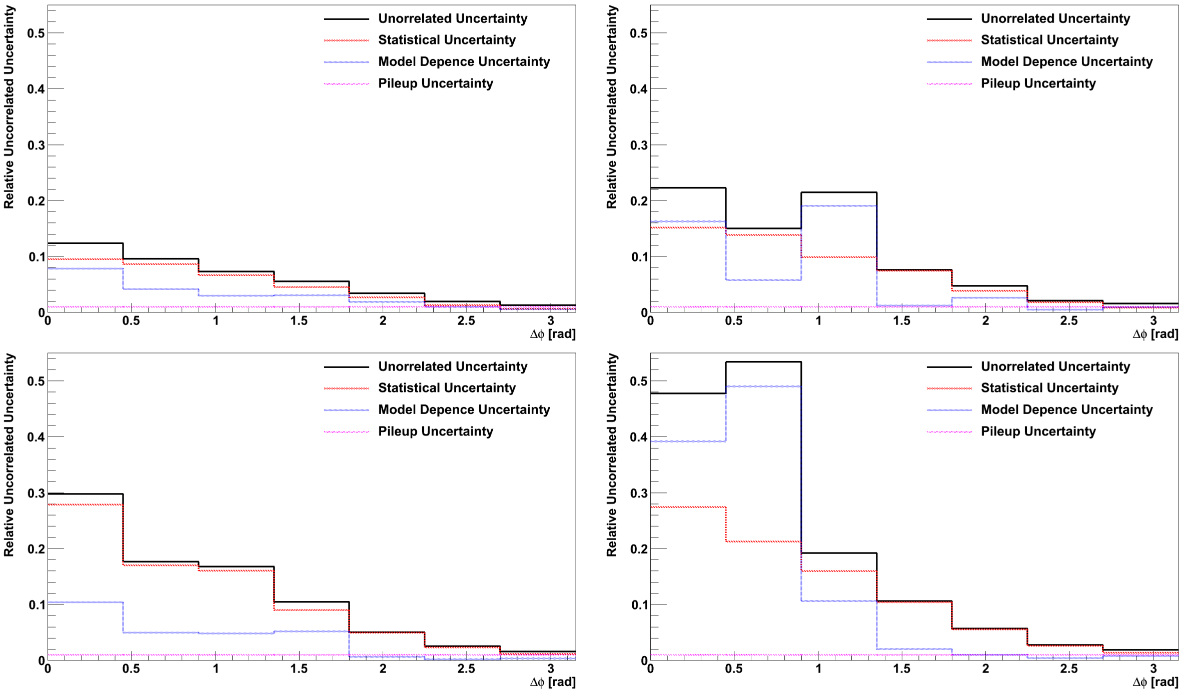


Figure J.2: Relative uncorrelated uncertainties of the $\Delta\phi$ distributions in the inside–jet veto scenario as function of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right).

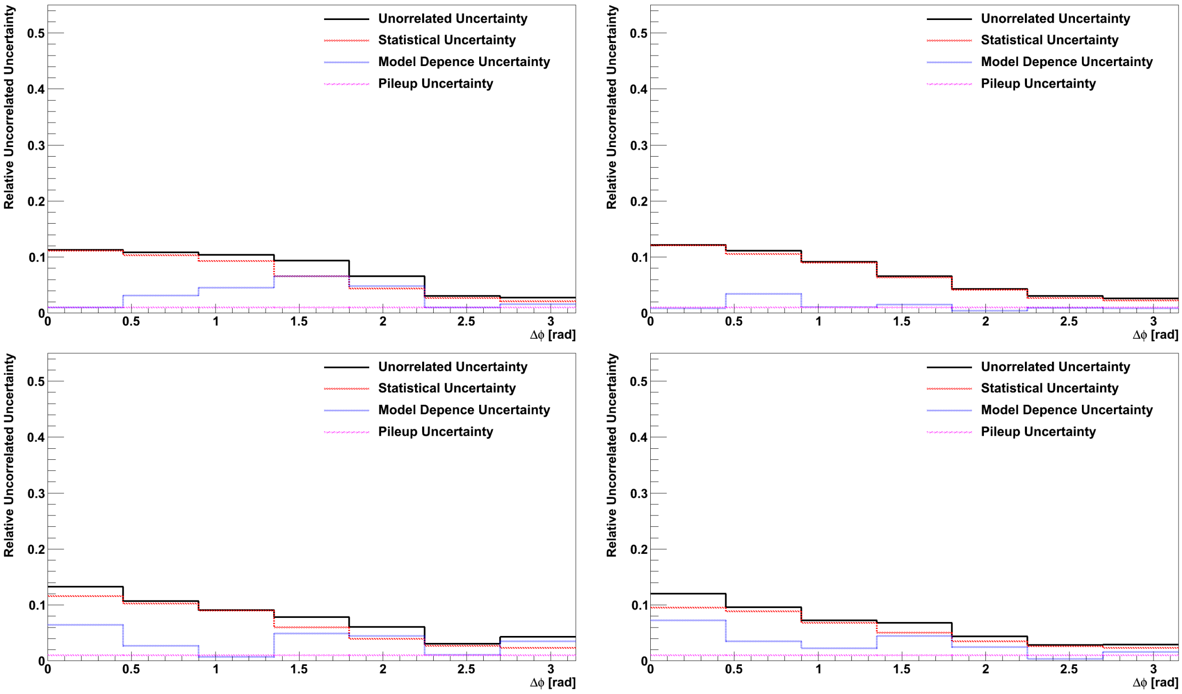


Figure J.3: Relative uncorrelated uncertainties of the $\Delta\phi$ distributions in the inside–jet tag scenario as function of $\Delta\eta$: $0.4 < \Delta\eta \leq 2.5$ (top-left), $2.5 < \Delta\eta \leq 3.5$ (top-right), $3.5 < \Delta\eta \leq 4.5$ (bottom-left) and $4.5 < \Delta\eta \leq 7.5$ (bottom-right).

Appendix K

Double differential results

In this appendix, the $\Delta\phi$ unfolded cross–section as a function of $\Delta\eta$ for the different scenarios is compared with several MC predictions. The data are shown as dots with the error bars indicating the uncorrelated uncertainties and the yellow band showing the correlarated uncertainties added in quadrature. The statistical uncertainty on the MC is represented by the error bars connected to the lines.

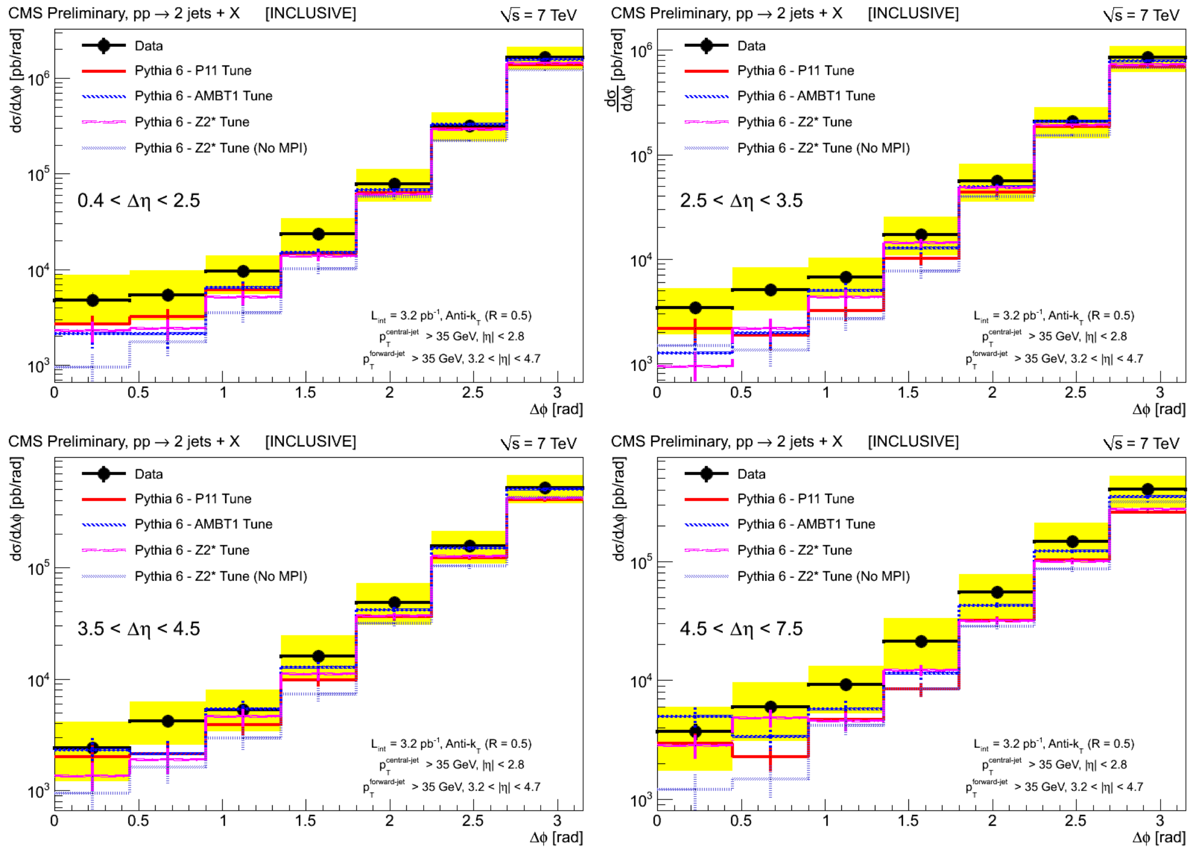


Figure K.1: Double–differential cross–section for the inclusive scenario in bins of the difference in the azimuthal angle between the forward and the central jet, $\Delta\phi$, as a function of the pseudorapidity separation between the two jets $\Delta\eta$. Data are compared to PYTHIA6.

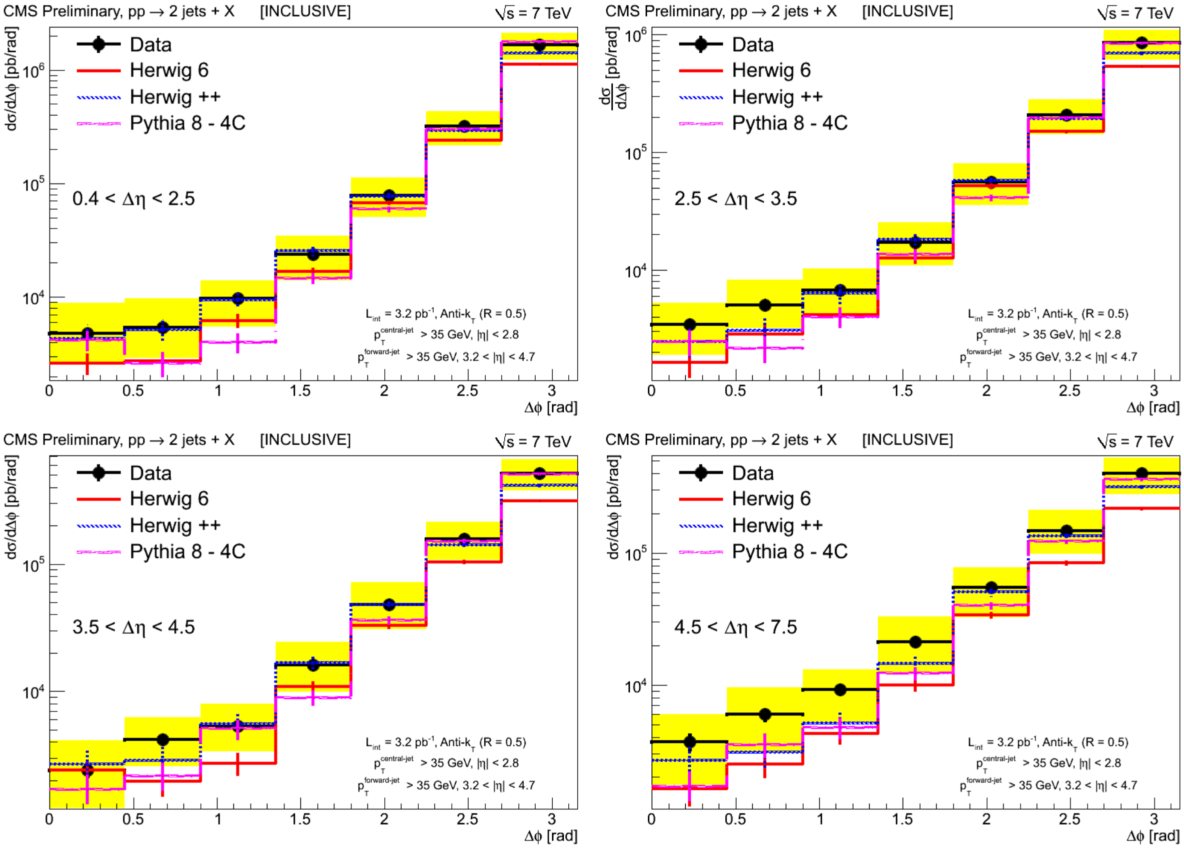


Figure K.2: Double–differential cross–section for the inclusive scenario in bins of the difference in the azimuthal angle between the forward and the central jet, $\Delta\phi$, as a function of the pseudorapidity separation between the two jets $\Delta\eta$. Data are compared to several MC predictions.

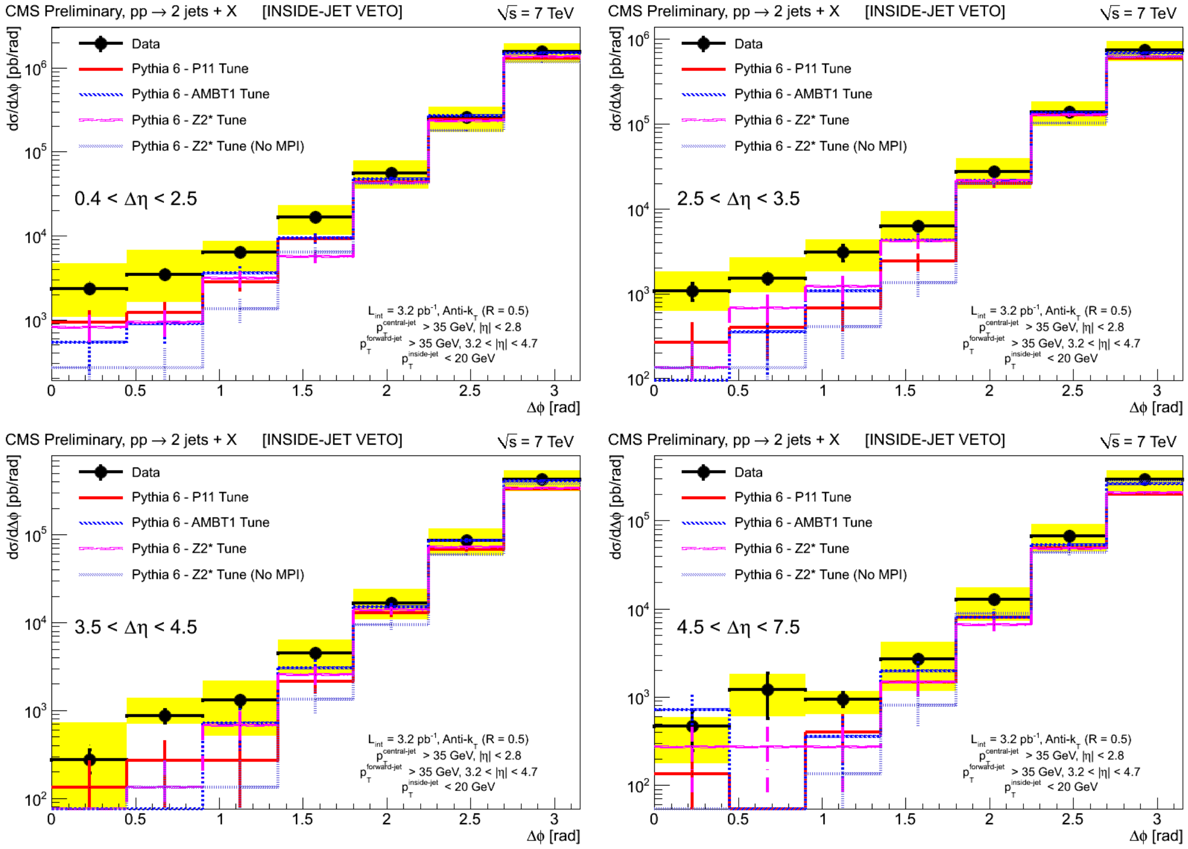


Figure K.3: Double differential cross section for the inside–jet veto scenario in bins of the difference in the azimuthal angle between the forward and the central jet, $\Delta\phi$, as a function of the pseudorapidity separation between the two jets $\Delta\eta$. Data are compared to Pythia 6.

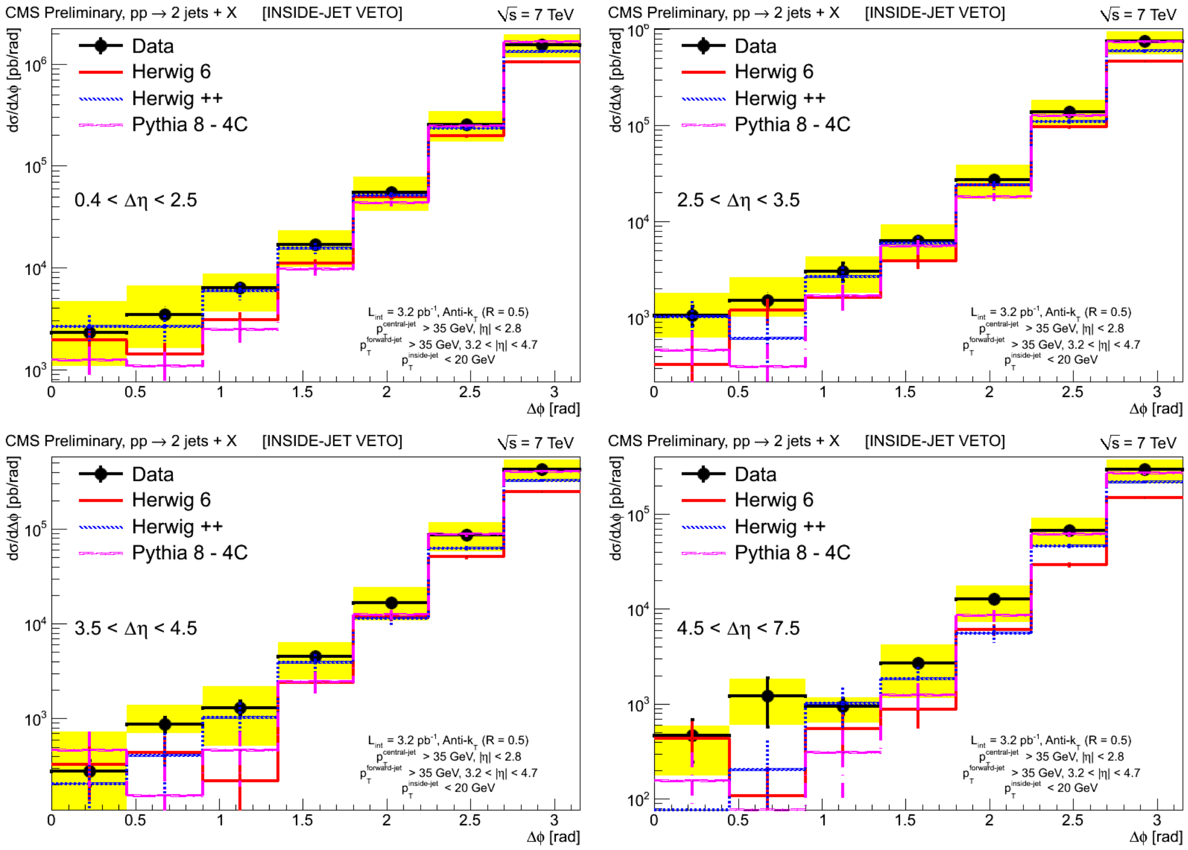


Figure K.4: Double–differential cross–section for the inside–jet veto scenario in bins of the difference in the azimuthal angle between the forward and the central jet, $\Delta\phi$, as a function of the pseudorapidity separation between the two jets $\Delta\eta$. Data are compared to several MC predictions.

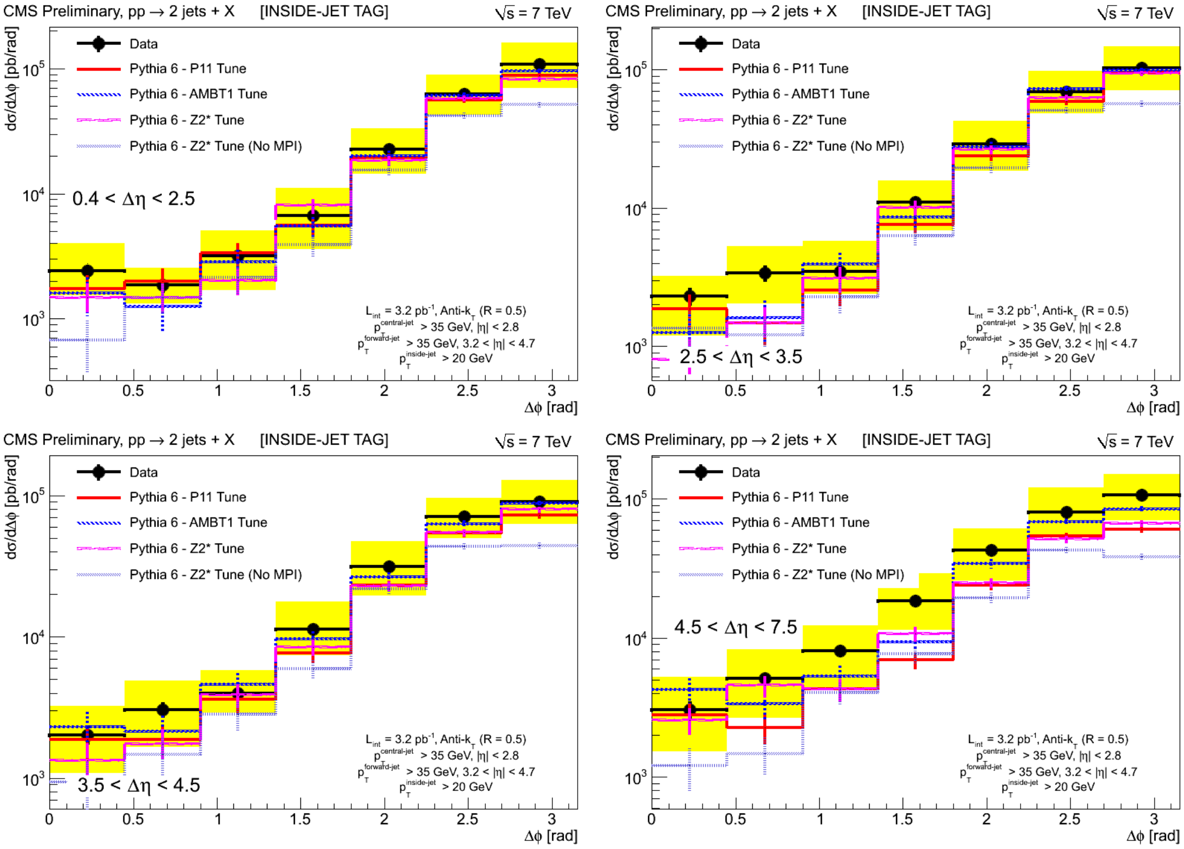


Figure K.5: Double–differential cross–section for the inside–jet tag scenario in bins of the difference in the azimuthal angle between the forward and the central jet, $\Delta\phi$, as a function of the pseudorapidity separation between the two jets $\Delta\eta$. Data are compared to Pythia 6.

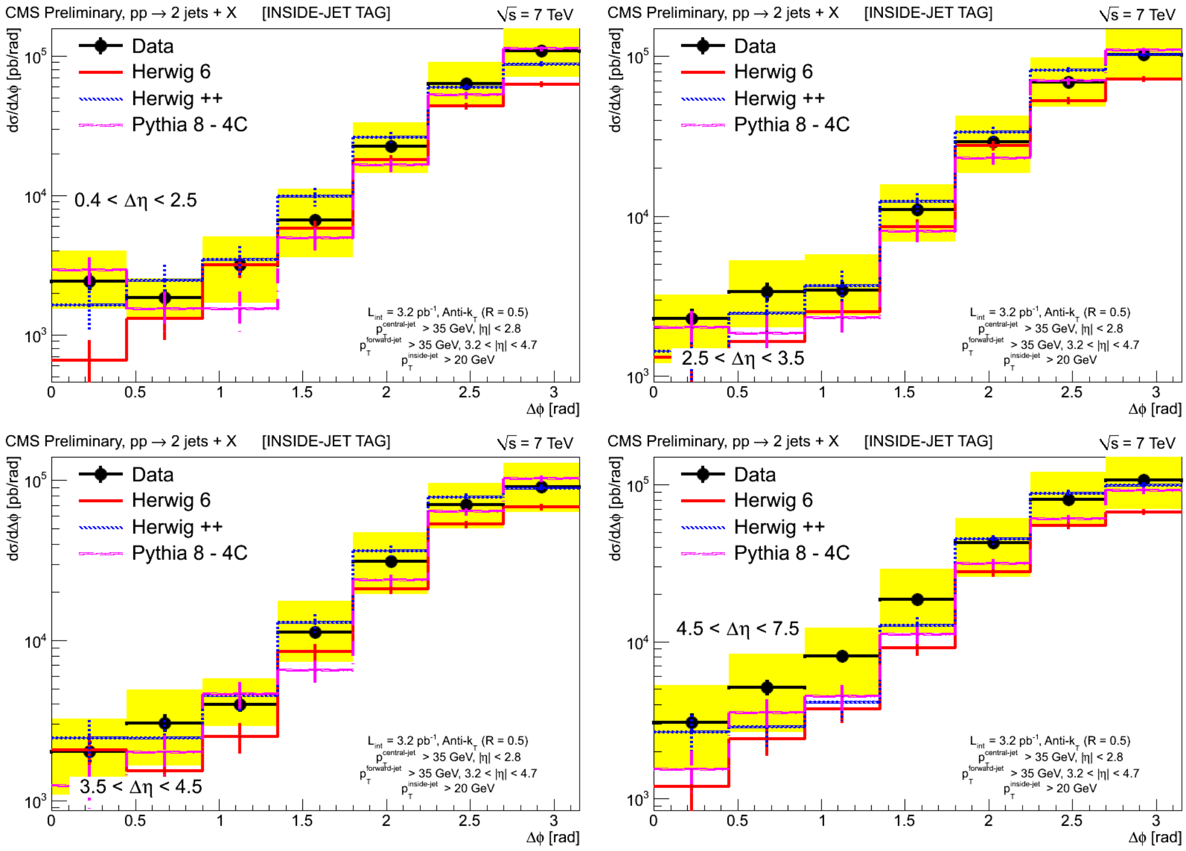


Figure K.6: Double–differential cross–section for the inside–jet tag scenario in bins of the difference in the azimuthal angle between the forward and the central jet, $\Delta\phi$, as a function of the pseudorapidity separation between the two jets $\Delta\eta$. Data are compared to several MC predictions.

Bibliography

Introduction

[1] M. Goodman, D. Tagle, D. Fitch, et al., “Primate evolution at the DNA level and a classification of hominoids“, J Mol Evol 30, 260—266, doi:10.1007/BF02099995 (1990)

[2] B. Gallagher, “Fear of the unknown“, Safer Communities 7 ,3, 22–24. doi:10.1108/17578043200800026 (2008)

[3] R. Oerter, “The Theory of Almost Everything: The Standard Model, the Unsung Triumph of Modern Physics“, Penguin Group (2006)

Standard Model and QCD

[4] S. L. Glashow, “Partial–symmetries of weak interactions“, *Nuclear Physics* 22 (4), 579 – 588 (1961).

[5] S. Weinberg, “Elementary particle theory of composite particles“, *Phys. Rev.* 130, 776 – 783, April 1963.

[6] S. Weinberg, “A model of leptons“, *Phys. Rev. Lett.* 130, 1264 – 1266, November 1967.

[7] C. Itzykson and J. B. Zuber, “Quantum Field Theory“, McGraw-Hill, ISBN 0-07-032071-3 (1980)

[8] D. McMahon, “Quantum Field Theory“, McGraw-Hill, ISBN 978-0-07-154382-8 (2008)

[9] P. W. Higgs, “Broken symmetries, massless particles and gauge fields“, *Phys. Lett.* 12, 132 – 133, 1964.

[10] P. W. Higgs, “Broken symmetries and the masses opf gauge bosons“, *Phys. Rev. Lett.* 13, 508 – 509, October 1964.

[11] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons“, *Phys. Rev. Lett.* 13, 321 – 323, August 1964.

[12] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, “Global conservation laws and massless particles“, *Phys. Rev. Lett.* 13, 585 – 587, November 1963.

[13] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC“, *Phys.Lett.* B716, 30 – 61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235, Jul 2012

[14] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC“, *Phys.Lett.* B716, 1 – 29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214, Jul 2012

[15] `http://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`

[16] `http://upload.wikimedia.org/wikipedia/commons/4/4c/Elementary_particle_interactions.svg`

[17] D. Gross and F. Wilczek, “Ultravioler Behaviour of Nonabelian Gauge Theories“, Phys. Rev. Lett. 30, 1343–1346, doi:10.1103/PhysRevLett.30.1343

[18] H. Politzer, “Reliable Pertubative Results for String Interactions¿‘, Phys. Rev. Lett. 30, 1346–1349, doi:10.1103/PhysRevLett.30.1346

[19] G. Altarelli, “A QCD Primer“, AIO Conf. Proc 631, 70–111 (2002)

[20] Michael E. Peskin and Daniel V. Schroeder, “An Introduction to Quantum Field Theory“, Westview Press (1995)

[21] R. K. Ellis, W. J. Stirling and B. R. Webber, “QCD and Collider Physics“, Cambridge University Press (1996)

[22] Günther Dissertori, Ian Knowles and Michael Shmelling “Quantum Chromodynamics, High Energy Experiments and Theory“, Oxford University Press (2003)

[23] John C. Collins, Davison E. Soper and George Sterman, “Factorization of Hard Processes in QCD“, Adv. Ser. Direct. High Energy Phys. 5:1-91, arXiv:hep-ph/0409313 (1988)

[24] Amanda Cooper–Sarkar, on behalf of H1 and Zeus Collaborations “PDF Fits at HERA“, Proceedings of the 2011 Europhysics Conference on High Energy Physics, HEP 2011, arXiv:1112.2107 (2011)

[25] H1 and Zeus Collaborations “Combined Measurement and QCD Analysis of the Inclusive ep Scattering Cross Sections at HERA“, JHEP 1001, 109, doi:10.1007/JHEP01(2010)109, arXiv:0911.0884 (2010)

[26] Daniel Stump, Joey Huston, Jon Pumplin, et al., “Inclusive Jet Production, Parton Distributions, and the Search for New Physics“, JHEP 0310, 046, doi:10.1088/1126-6708/2003/10/046, arXiv:hep-ph/0303013 (2003)

[27] A.D. Martin, W.J. Stirling, R.S. Thorne and G. Watt, “Parton distributions for the LHC“, Eur.Phys.J. C63, 189–285, doi:10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002 (2009)

[28] `http://images.iop.org/objects/ccr/cern/50/6/17/CCqcd3_06_10.jpg`

[29] V.N. Gribov and L.N. Lipatov, “Deep inelastic *ep* scattering in pertubarion theory“, Sov. J Nucl. Phys. 15, 438 (1972)

[30] V.N. Gribov and L.N. Lipatov, “e+ e- pair annihilation and deep inelastic ep scattering in perturbation theory”, Sov. J. Nucl. Phys. 15, 675 (1972).

[31] L.N. Lipatov, “The parton model and perturbation theory“, Sov. J Nucl. Phys. 20, 94 (1975)

[32] Yuri L. Dokshitzer, “Calculation of structure functions of deep-inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics“, Sov. Phys. JETP 46, 641 (1977)

[33] G. Altarelli and G.Parisi, “Asymptotic Freedom in Parton Language“, Nucl. Phys. B126, 294 (1977)

[34] E.A. Kuraev, L.N. Lipatov and V.S. Fadin, “Multi - Reggeon Processes in the Yang-Mills Theory“, Sov. Phys. JETP 44, 443 (1976)

[35] E.A. Kuraev, L.N. Lipatov and V.S. Fadin, “The Pomeranchuk Singularity in Nonabelian Gauge Theories“, Sov. Phys. JETP 45, 199 (1977)

[36] Ya. Ya. Balitski and L.N. Lipatov, “The Pomeranchuk Singularity in Quantum Chromodynamics“, Sov. J. Nucl. Phys. 28, 822 (1978)

[37] Marcello Ciafaloni, “Coherence Effects in Initial Jets at Small q^2/s“, Nucl. Phys. B296, 49 (1988)

[38] S. Catani, F. Fiorani and G. Marchesini, “QCD Coherence in Initial State Radiation“, Phys. Lett. B234, 339 (1990)

[39] S. Catani, F. Fiorani and G. Marchesini, “Small x Behavior of Initial State Radiation in Perturbative QCD“, Nucl. Phys. B336, 18 (1990)

[40] Giuseppe Marchesini, “QCD coherence in the structure function and associated distributions at small x“, Nucl. Phys. B445, 49–80, arXiv:9412327 (1995)

[41] X. Artru and G. Mennessier “String model and multiproduction“, Nucl. Phys. B70, 93 (1974)

[42] Andy Buckley, Jonathan Butterworth, et al., “General-purpose event generators for LHC physics“, Phys. Reports 504, 145, arXiv:1101.2599 (2011)

[43] B. Andersson et al., “Parton fragmentation and string dynamics“, Physics Reports 97, 2-3, 31, doi:10.1016/0370-1573(83)90080-7 (1983)

[44] D. Amati and G. Veneziano “Preconfinement as a property of perturbative QCD“, Phys. Lett. B83, 87 (1979)

[45] A. Bassetto, M. Ciafaloni and G. Marchesini, “Color singlet distributions and mass damping in perturbative QCD“, Phys. Lett. B83, 207 (1979)

[46] B.R. Webber “A QCD Model for Jet Fragmentation Including Soft Gluon Interference“, Nucl. Phys. B238, 492, doi:10.1016/0550-3213(84)90333-X (1984)

[47] G. C. Fox and S. Wolfram “A Model for Parton Showers in QCD“, Nucl. Phys. B168, 285–295 (1980)

[48] R. D. Field and S. Wolfram “A QCD Model for e+e- Annihilation“, Nucl. Phys. B213, 65–84 (1983)

[49] Gavin P. Salam, “Towards Jetography“, Eur. Phys. J. C67, 637, arXiv:0906:1833 (2010)

[50] S.D. Ellis, J. Huston, K. Hatakeyama, P. Loch and M. Toennesmann “Jets in Hadron-Hadron Collisions‘, Prog. Part. Nucl. Phys. 60, 484–551, doi:10.1016/j.ppnp.2007.12.002, arXiv:0712.2447 (2008)

[51] S. V. Chekanov “Jet algorithms: a minireview“, 14th Topical conference in Hadron Collider Physics (2002)

[52] Gavin P. Salam and Gregory Soyez “A practical Seedless Infrared-Safe Cone jet algorithm“, JHEP 9808, 001, arXiv:0704.0292m doi:10.1088/1126-6708/2007/05/086 (2007)

[53] S. Catani, Yuri L. Dokshitzer, M. Olsson et al., “New clustering algorithm for multi–jet cross-sections in $e+e-$ annihilation“, Phys. Lett. B269, 432–438, doi:10.1016/0370-2693(91)90196-W (1991)

[54] S. Catani, Yuri L. Dokshitzer and B. R. Webber, “The K-perpendicular clustering algorithm for jets in deep inelastic scattering and hadron collisions“, Phys. Lett. B285, 291–299, doi:10.1016/0370-2693(92)91467-N (1992)

[55] S. Catani, Yuri L. Dokshitzer, M. H. Seymour and B. R. Webber, “Longitudinally invariant k_t clustering algorithms for hadron hadron collisions“, Nucl. Phys. B406, 187–224, doi:10.1016/0550-3213(93)90166-M (1993)

[56] Stephen D. Ellis and Davision E. Soper “Successive Combination Jet Algorithm For Hadron Collisions“, Phys. Rev. D48, 3160–3166, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266 (1993)

[57] Yuri L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber “Better Jet Clustering Algorithms“, JHEP 9708, 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323 (1997)

[58] M. Wobisch and T. Wengler, “Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering“, arXiv:hep-ph/9907280 (1999)

[59] R. D. Field, “The Underlying Event in Hard Scattering Processes“, eConfC010630:P501, UFIFT-HEP-01-19, arXiv:hep-ph/0201192 (2001)

[60] P. Bartalini, E. L. Berger, B. Blok, et al., “Multi-Parton Interactions at the LHC“, ANL-HEP-PR-11-65, CMS-CR-2011-048, DESY 11-185, LTH929, KA-TP-32-2011, TTK-11-52, arXiv:1111.0469 (2011)

[61] M. Deak, F. Hautmann, H. Jung and K. Kutak “Forward Jet Production at the Large Hadron Collider“, JHEP 0909:121, doi:10.1088/1126-6708/2009/09/121, arXiv:0908.0538 (2009)

[62] M. Deak, F. Hautmann, H. Jung and K. Kutak, “Forward-Central Jet Correlations at the Large Hadron Collider‘, DESY-10-179, OUTP-10-07-P, arXiv:1012.6037 (2010)

[63] M. Deak, F. Hautmann, H. Jung and K. Kutak, “Forward Jets and Energy Flow in Hadronic Collisions“, arXiv:1112.6354 (2011)

[64] K. Kutak and S. Sapeta, “Gluon saturation in dijet production in p-Pb collisions at Large Hadron Collider“, Phys. Rev. D **86**, 094043, arXiv:1205.5035 (2012)

[65] M. Cacciari, G. P. Salam and G. Soyez, “The anti–k_T jet clustering algorithm”, JHEP **0804** 063, doi:10.1088/1126-6708/2008/04/063 (2008)

[66] D0 Collaboration, “Measurement of dijet azimuthal decorrelations at central rapidities in p anti-p collisions at $\sqrt{s} = 1.96$ TeV”, Phys. Rev. Lett. 94, 221801, arXiv:hep-ex/0409040, doi:10.1103/PhysRevLett.94.221801 (2005)

[67] ATLAS Collaboration, “Measurement of Dijet Azimuthal Decorrelations in pp Collisions at $\sqrt(s)$=7 TeV”, Phys. Rev. Lett. 106, 172002, doi:10.1103/PhysRevLett.106.172002, arXiv:1102.2696 (2011)

[68] CMS Collaboration, “Dijet Azimuthal Decorrelations in pp Collisions at $\sqrt{s} = 7$ TeV”, Phys. Rev. Lett. 106, 122003, arXiv:1101.5029, doi:10.1103/PhysRevLett.106.122003 (2011)

[69] CMS Collaboration, “Mueller-Navelet dijet azimuthal decorrelations in p-p at 7 TeV”, CMS–PAS–FSQ–12–002 (2012)

[70] CMS Collaboration, “Ratios of dijet production cross sections as a function of the absolute difference in rapidity between jets in proton-proton collisions at $\sqrt{s}$=7 TeV”, Eur. Phys. J. C 72, 2216, arXiv:1204.0696, doi:10.1140/epjc/s10052-012-2216-6 (2012)

[71] Pedro Cipriano, “Small-x QCD physics probed with jets in CMS”, Presented at the Low x workshop, May 30 - June 4, Rehovot and Eilat, Israel (2013)

[72] E.M. Levin et al., “Heavy quark production in semihard nucleon interactions”, Sov. J. Nucl. Phys. 53 657 (1991).

[73] S. Catani, M. Ciafaloni and F. Hautmann, “High-energy factorization and small-x heavy flavour production”, Nucl. Phys. B 366 135, doi:doi:10.1016/0550-3213(91)90055-3 (1991).

[74] J.C. Collins and R.K. Ellis, “Heavy quark production in very high-energy hadron collisions”, Nucl. Phys. B 360 3, doi:10.1016/0550-3213(91)90288-9 (1991)

[75] L.N. Lipatov, “Small x physics in perturbative QCD”, Phys. Rept. C 286 131. doi:10.1016/S0370-1573(96)00045-2 (1997)

[76] CMS Collaboration, “Measurement of the inclusive production cross sections for forward jets and for dijet events with one forward and one central jet in pp collisions at $\sqrt{s}$=7 TeV”, JHEP **1206** 36, doi:10.1007/JHEP06(2012)036, arXiv:1202.0704 (2012)

[77] Jeffrey R. Forshaw, A. Kyrieleis and M. H. Seymour, “Super-leading logarithms in non-global observables in QCD”, JHEP 0608, 059 doi:10.1088/1126-6708/2006/08/059, arXiv:hep-ph/0604094 (2006)

[78] M. Dasgupta and G.P. Salam, “Resummation of nonglobal QCD observables”, Phys. Lett. B, 512, 323-330, doi:10.1016/S0370-2693(01)00725-0, arXiv:hep-ph/0104277 (2001)

[79] “HERA - A Proposal for a Large Electron Proton Colliding Beam Facility at DESY”, DESY-HERA-81-10, (1981)

[80] J.M. Butterworth and T. Carli, “QCD uncertainties at the LHC and the implications of HERA”, C04-04-14, 22–37, arXiv:hep-ph/0408061 (2004)

LHC and CMS detector

[81] O.S. Brunning (Ed.) et al, “LHC Design Report. Vol I: The LHC Main Ring“, CERN-2004-003-V-1

[82] O.S. Brunning (Ed.) et al, “LHC Design Report. Vol II: The LHC Infrastructure and General Services”, CERN-2004-003-V-2

[83] O.S. Brunning (Ed.) et al, “LHC Design Report. Vol III: The LHC Injector Chain”, CERN-2004-003-V-3

[84] Lyndon Evans and Philip Bryant (Ed.), “LHC Machine”, JINST, **3**, August 2008, doi:1748-0221/3/08/S08001

[85] CMS Collaboration, “CMS Physics Technical Design Report Volume I : Detector Performance and Software”, CMS-TDR-008-1, 2006, CERN-LHCC-2006-001

[86] CMS Collaboration, “CMS physics Technical Design Report, Volume II : Physics Performance”, CMS-TDR-008-2, 2006, CERN-LHCC-2006-021

[87] CMS Collaboration, “The CMS experiment at the CERN LHC”, JINST, **3**, August 2008, doi:1748-0221/3/08/S08004

[88] “LEP design report, Volume 2: The LEP main ring”, CERN-LEP-84-01, `http://cds.cern.ch/record/102083?ln=pt` (1984)

[89] Christiane Lefèvre, "The CERN accelerator complex", `http://cdsweb.cern.ch/record/1260465`

[90] M. Lamont, "LHC: status and commissioning plans", Moriond 2009 QCD session, arXiv:0906.0347

[91] R. Assmann, M. Ferro-Luzzi, M. Giovannozzi, et al., "LHC beam parameters for the physics run at 3.5 TeV", `https://edms.cern.ch/document/1059898/`, 2009

[92] Q. Xu, T. Nakamoto, T. Ogitsu, et al., "Conceptual Design of a Large-aperture Dipole Magnet for HL-LHC Upgrade", CERN-ATS-2012-051 (2012)

[93] G. Dissertori, D. Luckey, F. Nessi-Tedaldi, et al., "A visualization of the damage in Lead Tungstate calorimeter crystals after exposure to high-energy hadrons", ETHZ-IPP-2012-01, arXiv:1202.0403 (2012)

[94] Ralph Assmann, Roger Bailey, Oliver Brüning, et al., "First Thoughts on a Higher–Energy LHC", CERN–ATS–2010–177 (2010)

[95] High Luminosity LHC Colaboration urlhttp://hilumilhc.web.cern.ch/HiLumiLHC/about/, accessed in December 2013

[96] LHC Collaboration "Public luminosity results for pp collisions in 2010", November 2010, `http://lpc-afs.web.cern.ch/lpc-afs/LHC/2010/lui_days_liny.png`

[97] LHC Collaboration "Public luminosity results for Pb-Pb collisions in 2010", December 2010, `http://lpc-afs.web.cern.ch/lpc-afs/LHC/2010/lui_days_liny_ion.png`

[98] LHC Collaboration "Public luminosity results for pp collisions in 2011", June 2012, `http://lpc-afs.web.cern.ch/lpc-afs/LHC/lui_days_liny.png`

[99] LHC Collaboration "Public luminosity results for Pb-Pb collisions in 2011", December 2011, `http://lpc-afs.web.cern.ch/lpc-afs/LHC/lui_days_liny_ion.png`

[100] LHC Collaboration "Public luminosity results for pp collisions in 2012", January 2013, `http://lpc-afs.web.cern.ch/lpc-afs/LHC/2012/luminosity_integrated_date.png`

[101] LHC Collaboration "Public luminosity results for p-Pb collisions in 2013", February 2013, `http://cms-service-lumi.web.cern.ch/cms-service-lumi/publicplots/int_lumi_per_day_cumulative_ppb_2013.png`

[102] ATLAS Collaboration, "Overall view of the LHC experiments", June 1999, `http://www.atlas.ch/photos/atlas_photos/selected-photos/lhc/9906026_01_layout_sch.jpg`

[103] ATLAS Collaboration, "ATLAS Detector and Physics Performance, Technical Design Report, Volume I", ATLAS-TDR-14, May 1999, CERN-LHCC-99-014

[104] ATLAS Collaboration, "ATLAS Detector and Physics Performance, Technical Design Report, Volume II", ATLAS-TDR-15, May 1999, CERN-LHCC-99-015

[105] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", JINST, **3**, August 2008, doi:1748-0221/3/08/S08003

[106] ALICE Collaboration, "Alice: Technical Proposal for A Large Ion Collider Experiment at the CERN LHC", LHCC-P-3, December 1995, CERN-LHCC-95-071

[107] ALICE Collaboration, "The ALICE experiment at the CERN LHC", JINST, **3**, August 2008, doi:1748-0221/3/08/S08002

[108] L3 Collaboration, "Technical proposal : L3", LEPC-P-4, May 1983, CERN-LEPC-83-5

[109] LHCb Collaboration, "LHCb - Technical Proposal, A Large Hadron Collider Beauty Experiment for Precision Measurements of CP Violation and Rare Decays", LHCC-P-4, 1998, CERN-LHCC-98-004

[110] LHCb Collaboration, "The LHCb Detector at the LHC", JINST, **3**, August 2008, doi:1748-0221/3/08/S08005

[111] DELPHI Collaboration, "DELPHI : technical proposal", DELPHI-83-66, May 1983, CERN-LEPC-83-3

[112] LHCf Collaboration, "LHCf experiment : Technical Design Report", LHCF-TDR-001, February 2006, CERN-LHCC-2006-004

[113] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", JINST, **3**, August 2008, doi:1748-0221/3/08/S08006

[114] TOTEM Collaboration, "TOTEM, Total Cross Section, Elastic Scattering and Diffraction Dissociation at the LHC : Technical Proposal", LHCC-P-5, March 1999, CERN-LHCC-99-007

[115] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider", JINST, **3**, August 2008, doi:1748-0221/3/08/S08007

[116] CMS Collaboration, "CMS TDR Volume 1 figure gallery", `http://cmsdoc.cern.ch/cms/cpt/tdr/ptdr1-figs/Figures/Figure_001-002.pdf`

[117] CMS Collaboration, "The Magnet Project: Technical Design Report", CERN-LHCC-97-10

[118] CMS Collaboration "Description and Performance of the CMS Track and Primary Vertex Reconstruction", CMS Physics Analysis Summary CMS-TRK-11-001 (2012)

[119] CMS Collaboration "Pseudorapidity and leading transverse momentum distributions of charged particles in pp collisions at $\sqrt{s} = 8$ TeV", PAS-FSQ-12-026, `http://cms-physics.web.cern.ch/cms-physics/public/FSQ-12-026-pas.pdf`

[120] CMS Collaboration, "The Electromagnetic Calorimenter: Technical Design Report", LHCC–97–33

[121] CMS Collaboration, "The Hadron Calorimenter Project: Technical Design Report", CERN–LHCC–97–31 (1997)

[122] CMS Collaboration "Determination of jet energy calibration and transverse momentum resolution in CMS", JINST 6, Vol 11, P11002 (2011)

[123] CMS Collaboration, "The Muon Project: Technical Design Report", CERN–LHCC–97–32 (1997)

[124] CMS Collaboration "Measurement of CMS Luminosity", CMS Physics Analysis Summary, CMS–PAS–EWK–10–004 (2010)

[125] S. van der Meer "Calibration of the effective Beam Height in the ISR", CERN–ISR–PO, 68, 31, `http://cdsweb.cern.ch/record/296752`

[126] CMS Collaboration, "The TriDAS Project, Technical Design Report, Volume 1: The Trigger Systems", CERN–LHCC–2000–038, CMS–TDR–006, `http://cdsweb.cern.ch/record/706847`

[127] CMS Collaboration "The TriDAS Project, Technical Design Report, Volume 2: Data Acquisition and High–Level Trigger", CERN–LHCC–2002–026, CMS–TDR–006–add–2, `http://cdsweb.cern.ch/record/578006`

[128] CMS Collaboration "The CMS high level trigger", EPJC **46** 3, 605–667 `doi:10.1140/epjc/s2006-02495-8`.

[129] CMS Collaboration "The CMS Computing Project – Technical Design Report", CERN–LHCC–2005–23 (2005)

[130] CMS Collaboration "CMS Computing Model", CERN–LHCC–2004–035 (2004)

[131] `http://lcg.web.cern.ch/lcg`

[132] `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookComputingModel`

[133] L. Tuura, A. B. Mayer, I. Segoni et al., "CMS data quality monitoring: systems and experiences", JPCS 219 072020 (2010), doi:10.1088/1742-6596/219/7/072020

[134] L. Tuura, G. Eulisse and A. B. Mayer, "CMS data quality monitoring web service", JPCS 219 072055 (2010), doi:10.1088/1742-6596/219/7/072055

Generators

[135] F. James, "A review of pseudorandom number generators", Comp. Phys. Comm. 60, 329 (1990)

[136] `cernlib.web.cern.ch/cernlib`

[137] `root.cern.ch`

[138] F. James, "RANLUX: A Fortran implementation of the high-quality pseudo-random number generator of Luscher", Comp. Phys. Comm. 79, 111–114 (1994)

[139] P. L'Ecuyer, "Maximally Equidistributed Combined Tausworthe Generators", Mathematics of Computation 65, 213 (1996)

[140] M. Matsumoto and T. Nishimura, "Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator", ACM Transactions on Modeling and Computer Simulation Vol.8, No. 1, 3–30 (1998)

[141] `http://omega.albany.edu:8008/cdocs/summer99/lecture4/env90.gif`

[142] John Kemeny and J. Laurie Snell, "Finite Markov Chains", Springer–Verlag (1976)

[143] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications", Biometrika 51, 1 97–109 (1970)

[144] S. Agostinelli, J. Allison, K. Amako et al., "Geant4 Collaboration, Geant4—a simulation toolkit", Nuc. Inst. Met. Phys. A, 506, 3, 250–303, doi:10.1016/S0168-9002(03)01368-8 (2003)

[145] G. Corcella, I.G. Knowles, G. Marchesini et al., "HERWIG 6.5", JHEP 0101 (2001), hep-ph/0210213

[146] M. Bahr, S. Gieseke, M. A. Gigg et al., "Herwig++ Physics and Manual", Eur. Phys. J. C58, 639–707 (2008)

[147] T. Sjostrand, Stephen Mrenna and Peter Skands, "PYTHIA 6.4 Physics and Manual", JHEP 05, 026, arXiv:0603175, doi: 10.1088/1126-6708/2006/05/026 (2006)

[148] T. Sjostrand, Stephen Mrenna and Peter Skands, "A Brief Introduction to PYTHIA 8.1", Comput. Phys. Commun. 178, 852–867, arXiv:0710.3820, doi:10.1016/j.cpc.2008.01.036 (2008)

[149] Torbjörn Sjöstrand, et al, "PYTHIA 8 Physics and Manual", `http://home.thep.lu.se/~torbjorn/pythia81html/Welcome.html` (2013)

[150] Jeppe R. Andersen and Jennifer M. Smillie, "The Factorisation of the t-channel Pole in Quark-Gluon Scattering Phys.Rev.D81:114021, arXiv:0910.5113 (2010)

[151] Jeppe R. Andersen and Jennifer M. Smillie "Constructing All-Order Corrections to Multi-Jet Rates", JHEP 1001:039, arXiv:0908.2786 (2010)

[152] Jeppe R. Andersen and Jennifer M. Smillie, "Multiple Jets at the LHC with High Energy Jets", JHEP06 010, arXiv:1101.5394 (2011)

[153] Hannes Jung and G.P. Salam, "Hadronic final state predictions from CCFM: The Hadron level Monte Carlo generator CASCADE", Eur. Phys. J. C19, 351-360 (2001)

[154] Simone Alioli, Paolo Nason, Carlo Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", JHEP 06, 043 (2010)

[155] J. M. Butterworth and M. H. Seymour, "JIMMY4: Multiparton Interactions in Herwig for the LHC", October 2004

[156] Pumplin, J. and others, "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis", J. High Energy Phys. 07, 012, doi:10.1088/1126-6708/2002/07/012 (2002)

[157] L. Lönnblad, "ThePEG, PYTHIA7, Herwig++ and ARIADNE', Nucl. Instrum. Meth. A559, 246–248 (2006)

[158] Alan D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, "MRST2001: Partons and α_s from precise deep inelastic scattering and Tevatron jet data", Eur. Phys. J. C 23, 73, doi:10.1007/s100520100842 (2002)

[159] `https://herwig.hepforge.org/trac/wiki/MB_UE_tunes`

[160] `http://indico.cern.ch/getFile.py/access?contribId=13&sessionId=2&resId=0&materialId=slides&confId=140054`

[161] Field, Rick, "Early LHC Underlying Event Data - Findings and Surprises", arXiv:1010.3558 (2010)

[162] Peter Zeiler Skands, "Tuning Monte Carlo Generators: The Perugia Tunes", Phys. Rev. D 82, 074018, arXiv:1005.3457, doi:10.1103/PhysRevD.82.074018 (2010)

[163] ATLAS Collaboration, "Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC", New J.Phys. 13, 053033, doi:10.1088/1367-2630/13/5/053033, arXiv:1012.5104 (2011)

[164] A. Sherstnev and R. S. Thorne, "Parton Distributions for LO Generators", Eur. Phys. J. C 55, 553, doi:10.1140/epjc/s10052-008-0610-x, arXiv:0711.2473 (2008)

[165] Richard Corke and Torbjörn Sjöstrand, “Interleaved Parton Showers and Tuning Prospects”, JHEP 1103, 032, doi:10.1007/JHEP03(2011)032, arXiv:1011.1759 (2011)

[166] ALEPH Collaboration, “Performance of the ALEPH detector at LEP”, Nuc. Inst. Met. A360, 481 (1995)

[167] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}”, CMS Physics Analysis Summary PFT-09-001 (2009)

[168] CMS Collaboration, “Determination of the jet energy scale in CMS with pp collisions at $\sqrt{s} = 7$ TeV”, CMS Physics Analysis Summary JME-10-010 (2010)

[169] CMS Collaboration, “Jets in 0.9 and 2.36 TeV p–p Collisions”, CMS Physics Analysis Summary JME-10-001 (2010)

[170] CMS Collaboration, “Jets Performance in 7 TeV p–p Collisions”, CMS Physics Analysis Summary JME-10-003 (2010)

Selection

[171] K. Kousouris, `http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/UserCode/KKousour/`

[172] Rose K., “Deterministic Annealing for Clustering, Compression, Classification Regretion and related Optimization Problems”, Proceedings of the IEEE **Vol. 86, Issue 11**, (1998)

Corrections

[173] Lendermann, V., Haller, J. et al., “Combining Triggers in HEP Data Analysis”, Nucl. Instrum. Meth. **A604** 707, arXiv:0901.4118, doi:10.1016/j.nima.2009.03.173 (2009),

Unfolding

[174] Tim Adye, `http://hepunx.rl.ac.uk/~adye/software/unfold/RooUnfold.html`

[175] Volker Blobel, “Unfolding – Linear Inverse Problems”, Notes for the Terrascale workshop at DESY, `https://www.wiki.terascale.de/images/a/a8/Blobel_Unfold_Note_20100524.pdf`, May 2010

[176] G. D’Agostini, “A multidimensional unfolding method based on Bayes’ theorem”, Nucl. Instrum. Meth. A362, 487–498 (1995)

[177] Andreas Hoecker and Vakhtang Kartvelishvili “SVD Approach to Data Unfolding”, Nucl. Instrum. Meth A372, 469–481 (1996), `doi:10.1016/0168-9002(95)01478-0`, `arXiv:hep-ph/9509307`

[178] G. E. Forsythe, M. A. Malcon and C. B. Moler “Computer Methods for Mathematical Computations”, Prentice–Hall Inc., Englewood Cliffs, (1977)

[179] C. E. Lawson and R.J. Hanson, “Solving Least Squares Problems”, Prentice–Hall Inc., Englewood Cliffs, (1974)

[180] `http://root.cern.ch/root/html/TSVDUnfold.html`

[181] P. C. Hansen, “The L-Curve and its Use in the Numerical Treatment of Inverse Problems”, in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering, 119–142 (1974)

[182] `http://root.cern.ch/root/html/TUnfold.html`

Uncertainties

[183] Panel discussion on unfolding, Terascale Statistics School 2010, Göttingen, `https://indico.desy.de/conferenceDisplay.py?confId=3086`

[184] CMS Collaboration, “Luminosity Estimation for 2010 Data - Update”, CMS Physics Analysis Summary, EWK-11-001 (2011)

Conclusion

[185] T. Gleisberg, S. Hoeche, F. Krauss et al., “SHERPA 1.alpha, a proof-of-concept version”, JHEP 0402, 056 (2004)

[186] Tanju Gleisberg, Stefan Höche, Frank Krauss, et al., `http://ppewww.physics.gla.ac.uk/~ahgemmell/Sherpa/113Manual.pdf`

[187] P. Cipriano, S. Dooling, A. Grebenyuk, P. Gunnellini, F. Hautmann, H. Jung and P. Katsas, “Higgs as a gluon trigger”, Phys. Rev. D, vol. 88, Issue 9, id. 097501, doi:10.1103/PhysRevD.88.097501, DESY-13-139, arXiv:1308.1655 (2013)

Castor

[188] `https://twiki.cern.ch/twiki/bin/viewauth/CMS/DQMShiftCastor`

List of Figures

List of Tables