DESY Summer Student Program 2010

Transverse momentum of inclusive forward Jets in CMS with $\sqrt{s} = 7\ TeV$

Pablo Bueno Gómez[1]

Universidad de Oviedo

supervised by **Albert Knutsson**

DESY CMS Group

9th September 2010

Abstract

In this work we describe a comparison between the corrected experimental data corresponding to the transverse momentum of the forward jets measured in the CMS experiment at LHC with $\sqrt{s} = 7\ TeV$), and the predictions from various Monte Carlo simulations. After describing the analysis routine, which was coded as a part of the summer school project, we study the effects of the modification of several Pythia parameters and tunes. Also predictions from the Herwig MC generator are compared to the data.

[1]UO189399@uniovi.es

Contents

Introduction

The main part of our report consists of the analysis routine that has been developed, and the discussion of the results, trying to determine the accuracy of the simulators to reproduce the experimental data.

After a brief introduction to the CMS experiment, the first part of this report is devoted to the explanation of some fundamental physical concepts. The second section includes some comments about Pythia and Herwig, as well as about the software used to run them (Agile), and to analyze (Rivet) their simulated data as well as an explanation of the analysis routine written. In the next section, we show and discuss the results obtained. We compare the experimental data with the MC predictions from several Pythia and Herwig configurations and tunes. And finally, the last chapter includes the conclusions.

The CMS Experiment [1]

The CMS (Compact Muon Solenoid) is one of the four detectors built at LHC. It is, together with ATLAS, one of the two main experiments at the proton-proton collider.

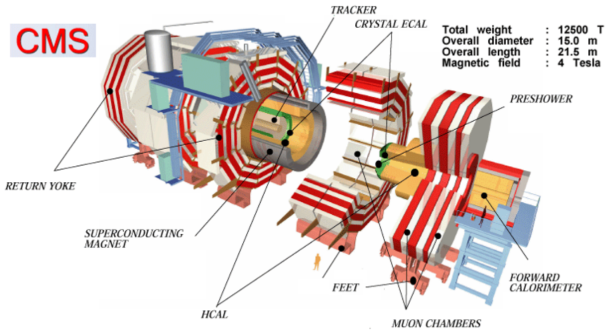

Figure 1: *Compact Muon Solenoid*

The main objective of the CMS detector is to make research in the experimental frontier of High Energy Physics. This includes, for instance, Higgs boson search, SUSY, extra dimensions research, but also more fundamental research such as QCD, which is dealt with in this report.

In order to fulfill all this physical aims, CMS is a multi-purpose detector that has been designed in fifteen separate sections. Each one of the layers is designed to stop, track or measure a different type of particle emerging from proton-proton and heavy ion collisions.

The main components of the machine are (listed from the inside to the outside) a full silicon tracker, a scintillating electromagnetic calorimeter, a hadron calorimeter, a superconducting solenoid magnet, and a muon detection system.

More than 2000 scientist are working or collaborating in CMS, coming from 155 institutes in 37 countries.

1 Physics

1.1 Detector Coordinates. Transverse Momentum. Forward Region.

There are several physical quantities that are of importance in the analysis of high energy events in a detector such as the pseudo rapidity, the transverse momentum, or the scattering angle ϕ. The coordinates system that we use to characterize these observables is the one showed in Figure 2:

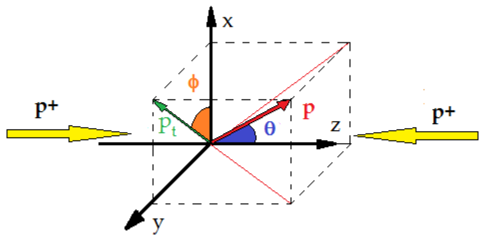

Figure 2: *Detector Coordinates and p_t, ϕ and θ definitions*

Transverse momentum is the component of the momentum of a particle in the transverse plane (the one perpendicular to the beam axis). Pseudo rapidity is a spacial coordinate describing the angle of a particle relative to the beam axis defined as:

$\eta = -\ln(\tan\frac{\theta}{2})$

Where θ is the polar angle of the direction of the particle with respect to the anticlockwise beam direction.

In our work here, we mainly concentrated our efforts in the study of the transverse momentum for the forward jets produced at LHC after the hadronization of partons.

The forward region of a detector is that corresponding to the region placed close (low angle) to the line in which the beams collide. Thus, there are two forward regions, one on each side of the collision place. They usually are referred as forward and backward regions. A whole area of Physics (known as Forward Physics) studies the characteristics of all the phenomena occurring in this region of a HEP collision detector (not necessarily in the CMS).

1.2 Jets.

A jet is a collimated flow of particles that form around a high-energy quark or gluon. Because of QCD confinement, particles carrying a color charge, such as quarks, cannot exist in free form. Therefore they fragment into hadrons before they can be directly detected, becoming jets. Thus, jets are produced by the hadronization of partons in QCD hard scattering processes. These hadronization processes are usually explained using phenomenological models like the Lund string fragmentation model. The interest in jets comes from the fact that they can be observed in particle detectors, rather than quarks.

The basic process for the creation of a jet in hadron-hadron collisions is described schematically in Figure 3.

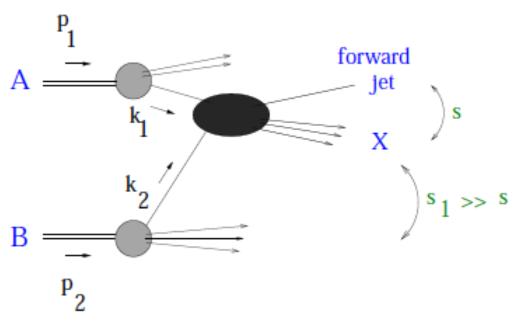

Figure 3: *Jet production in the forward rapidity region in hadron-hadron collisions*

Working at polar angles that are small but sufficiently far from the beam axis not to be affected by beam remnants, gives a high sensitivity to non-standard QCD effects. For example, the measurement of azimuthal plane correlations between hight-p_t events widely separated in rapidity is possible. [2]

2 Analysis routine.

2.1 Monte Carlo collision generators.

Rivet is a C++ class library, which provides the infrastructure and calculational tools for simulation-level analyses for high energy collider experiments, enabling physicists to validate event generator models and tunings with minimal effort and maximum portability[3]. AGILe is a Generator Interface Library (and executable), i.e. a uniform object oriented C++ interface for a variety of Fortran-based Monte Carlo event generators. The role of AGILe is to provide a standard steering interface for Fortran generator codes which usually do not come with an executable, usually need to be recompiled to change parameter settings, and cannot write output into the C++ HepMC event record: AGILe remedies all of these defects. The agile-runmc executable provides a very powerful yet simple command-line interface for steering a variety of generators[4]. In our work we have used two of them:

- PYTHIA6:422
- HERWIG:6.510

Pythia and Herwig are two Monte Carlo packages for high energy particle collisions. They are able to generate high-energy-physics events, i.e. sets of outgoing particles produced in the interactions between two incoming particles.[5][6]

In order to systematically work with the output from these programs, an analysis routine is necessary. Rivet library includes several Standard analyses that can be used to study events from a variety of generators and tunes. However, our work required the development of a new one.

2.2 Analysis routine.

As we have said, in order to compare the experimental data with the simulated one, we designed an analysis routine to be used by Agile with Pythia and Herwig generators.

Our aim is to create histograms of the transverse momentum of all the jets in the forward region ($3.2 < |\eta| < 4.7$) and determine the accuracy of the Monte Carlo simulators trying to reproduce the results.

The experimental data, taken at CMS with a center of mass energy of 7 TeV have been fully corrected i.e., the detector effects have been removed (bin-by-bin correction has been applied). As we have already mentioned, the η cut, that defines the forward region, is, in our case: $3.2 < |\eta| < 4.7$.

Moreover, the centre of binning is: $(40, 51, 64.5, 81, 105, 135)(GeV)$. Jets have been reconstructed with the Anti-k_t algorithm, with $R = 0.5$ (where $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is the radius of the jet cone in the $\eta - \phi$ space).

Our analysis routine consists of three functions.

The first one (init) is in turn divided in three. First, we pre-select the range of particles that we want in the final state. We choose those with pseudorapidity between -25 and 25 as this is a large enough range to include all the particles that could be relevant. Moreover, we establish that there is no limit of minimum energy for the particles. That is to say, we take into account all the particles with energy larger than 0 GeV.

Afterwards, we add the projections that we are going to use in our routine: fs (final state), and FastJets, for which we select the algorithm Anti-k_t with $R = 0.5$.

In the third and last part of init, we create the histogram for the transverse momentum of the Forward Jets. The binning of these histograms are choosen to be the same as for the measured data.

The second and main part of our routine is called once for each generated event. Here, we select jets whith transverse momentum larger than 20 GeV (The first bin centre is 40 GeV).

The analysis process is very simple: The histogram is filled for each jet generated with tranverse momentum larger than 20 GeV, and pseudorapidity η in the forward region ($3.2 < |\eta| < 4.7$).

The last part of the routine has the main aim of normalizing our histogram. We must take into account that we want to normalize to both luminosity and to the binwidth of η and p_t. Thus, on the Y axis of our histogram the variable that will be represented will be the second derivative of the cross section with respect to η and the transverse momentum p_t. For the luminosity normalization, we calculate the inverse luminosity, defined as:

$\mathcal{L}^{-1} = \frac{1}{\int Ldt} = \frac{\sigma}{N}$

Where σ is the total cross section for the MC run, and N the total number of events generated. Moreover, we must normalize it to the range in η. In order to do this, we divide all the values of the histogram by 3. That value is the total range in pseudorapidity that is included in the two forward regions ($(4.7 - 3.2)2 = 3$).

The binwidth normalization for p_t is done automatically in Rivet when using the so called AIDA format.

3 Results.

3.1 Settings.

In order to determine the accuracy of Pythia reproducing the experimental data, we will modify several parameters of the simulator, verifying the agreement of the plots with the experience in each case. These parameters are:

- The existence of multiple interactions ($MSTP(81) = 1$ for multiple interactions ON or $MSTP(81) = 0$ for multiple interactions OFF). It is expected that the number of jets with a certain value of p_t is larger in every region of the detector (not only in the forward one) when $MSTP(81) = 1$ than in the case of $MSTP(81) = 0$, because a larger number of particles can be produced when we have several partonic interaction in each pp collision. Thus, we must expect that Pythia reproduce experimental data with quite better accuracy when this parameter is activated.

- We can also modify the minimum value of the p_t to be generated by Pythia in the hard scattering region. The parameter that allow us to do this is $CKIN$, and its units are GeV. Thus, in the case that $CKIN$ is low, we will be sure about that we are including all the jets in our histogram, but it will take much more time to get good statistics, because a lot of events will be useless for us. On the other hand, if we choose, for instance, $CKIN = 40$, our event generation efficiency will be higher but it may happen that we exclude jets with p_t lower than 40 GeV.

- Moreover, we will try to compare four Phythia Tunes with the experimental data, to know which one is the best describing the forward jets production at LHC for $\sqrt{s} = 7\ TeV$.

3.2 Results.

The main objective of our work, as we have already said, is to establish the level of accuracy of the MC simulators, trying to reproduce the experimental data measured at LHC. We show our best result first. This one has been provided by Herwig after 7 million events, and more than 30 hours of simulation:

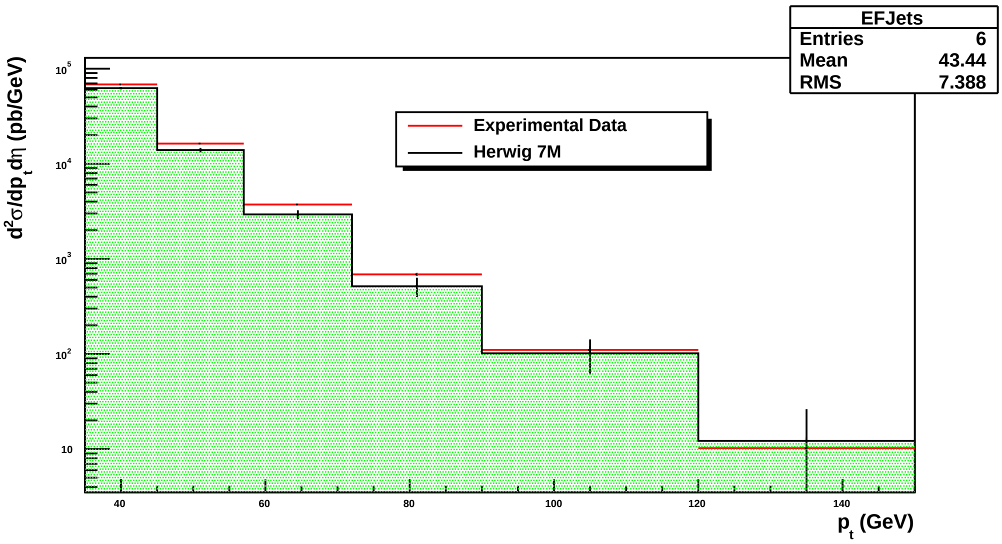

Figure 4: *Herwig MC simulation for the transverse momentum of forward jets at LHC with* $\sqrt{s} = 7\ TeV$ *compared to the CMS measurement.*

As we can see, Herwig simulation describes acceptably well the measured data although the bar errors do not actually include them for some of the bins.

Our second result concerns the importance of the multiple partonic interactions parameter in Pythia simulator in order to fit the measured data.

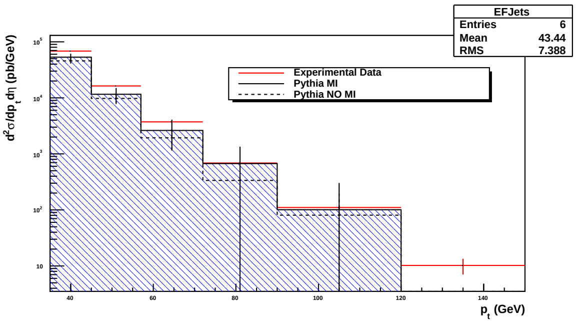

Figure 5: *Simulated Pythia results for the transverse momentum of forward jets at LHC with* $\sqrt{s} = 7\ TeV$ *and with the MI turned ON and OFF.*

Although our statistics are not very good in this histogram, we can see a significative difference between the results obtained with $MSTP(81) = 1$ and the ones with $MSTP(81) = 0$. For all the bins, the exclusion of the multiple interactions parameter produces a diminution in the cross section, causing a worse description of the experimental data. As it was expected, the multiple interactions parameter must be switched on in order to describe the data.

The next plot tries to determine the importance of the modification of the minimum value of the p_t to be generated by Pythia in the hard scattering region:

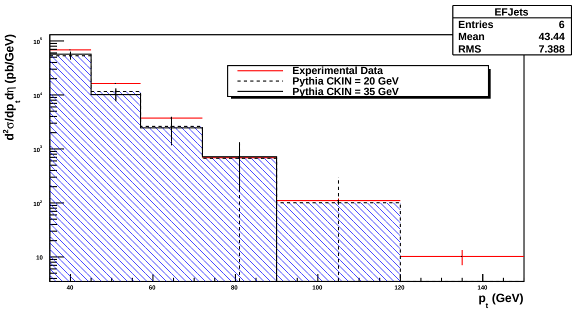

Figure 6: *Pythia results changing the CKIN parameter*

The results confirm that our event generation efficiency is higher when $CKIN$ has larger values, and that there is no difference in the accuracy of the fitting.

It is interesting to establish a comparison between the results provided by our two Monte Carlo simulators. First of all, we must say that Pythia is 5 times faster than Herwig in the event generation. However, the statistics provided by Herwig are much better than the Pythia ones, even with the fifth part of events.

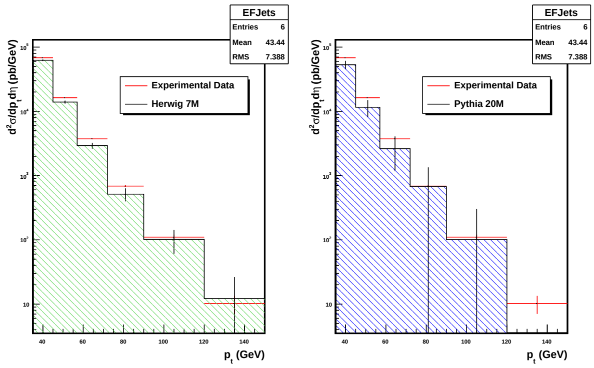

Figure 7: Comparison between Herwig (7 M events) and Pythia (20 M events)

We can observe that Herwig seems to describe better the data than Pythia in the low p_t region, although we can not be totally sure about that because of the poor Phythia statistics (and its big error bars). We have the same problem (or even more significative) with the high p_t region, where the Phythia error bars are so big that we can not say almost anything about its precision fitting the measured data.

The next and last result concerns four Pythia Tunes. Plotting them in the same histogram, we try to determine which one is the most precise, and also to establish a general range of uncertainty for the Pythia simulator in this experience:

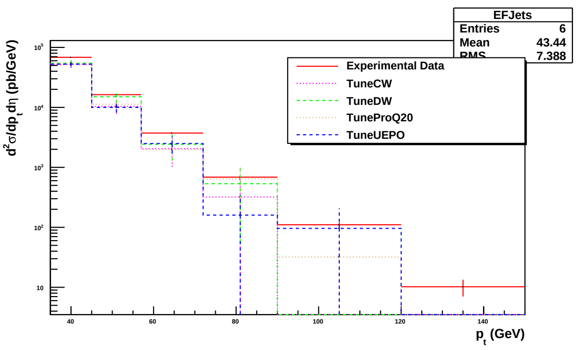

Figure 8: Results for 4 Pythia tunes

No one of the Pythia Tunes describes very correctly the measured data.

However, we can see that the total range covered by their error bars includes the experimental data almost in all cases.

4 Conclusions.

After the analysis of our results, we can extract the following conclusions:

- Both Monte Carlo generators, Pythia and Herwig, give an acceptable description of the data measured at LHC.
- Multiple interactions parameter must be activated in Pythia to obtain a more faithful fit.
- Herwig provides better statistics than Pythia, with less events, in the same running time. The accuracy of the description of both simulators seems to be similar. Only Herwig seems to be a little bit better in the low p_t region.

Aknowledgements.

I want to express my acknowledgement and appreciation to my supervisor Dr Albert Knutsson for his valuable guidance and patience, and also for all the things that I have learnt from him during the Programme. I am also thankful to Dr Mira Kraemer for her supervision and help during the first weeks.

I also express my recognition to Dr Joachim Meyer, Mrs Andrea Schrader and all the people who are behind the Summer Student Programme for their work making possible this great experience.

Special thanks go to my officemates Samantha Dooling, Irina Cioara and Paolo Gunnellini for their help, advise, and mutual encouragement.

References

[1] http://public.web.cern.ch/public/en/lhc/CMS-en.html.

[2] arXiv:0910.3400v2 [hep-ph]

[3] http://projects.hepforge.org/rivet/rivet-manual.pdf

[4] http://projects.hepforge.org/agile/

[5] http://hepwww.rl.ac.uk/theory/seymour/herwig/

[6] http://projects.hepforge.org/pythia6/