Tuning of the Herwig++ MC generator using CMS and ATLAS data

Krzysztof Bożek

DESY summer student 2011

Supervisors:

Albert Knutsson, DESY CMS group,
Simon Plätzer, DESY Theory group.

Abstract

Tuning of Herwig++ 2.5.1 to MPI parameter was done to ATLAS and CMS underlying event data, CMS charged hadron spectrum data, as well as to ATLAS jet shapes and CMS event shapes data. It was done using the Professor toolkit. Five Herwig++ parameters have been considered: $p_{\perp,min}$, colour reconnection probability (P_{reco}), inverse hadron radius squared (μ), probability of colour disruption ($\mathrm{P_{disrupt}}$), and Intrinsic $p_\perp$ Gaussian (IntPtGauss). The tuning was performed using default Herwig++'s PDFs set (MSTW LO**) and CTEQ61L. No sensitivity was seen for probability of colour disruption, and Intrinsic $p_\perp$ Gaussian. Jet shapes and event shapes data gave no constraint to investigated parameters. Tension between charged hadron spectrum and ATLAS UE data was discovered. Significant improvement in describing data was achieved using CTEQ61L PDFs set.

DESY, Hamburg 2011

Contents

1 Introduction

Natural consequence of physics approximation and phenomenological models in Monte Carlo (MC) generators are several free parameters in the model which cannot be derived from basic principles and must be determined from experimental data. Such parameters appear in most of generators parts. MC generators must be tuned to measurements. Fortunately most of parameters, like hadronization parameters, could be determined once and should not depend on energy scale. These parameters could be obtained from very precise LEP or BELLE colliders data. However there are also few parameters which should be obtained for each colliding energy, mainly multi parton interaction (MPI) parameters in hadron–hadron collisions.

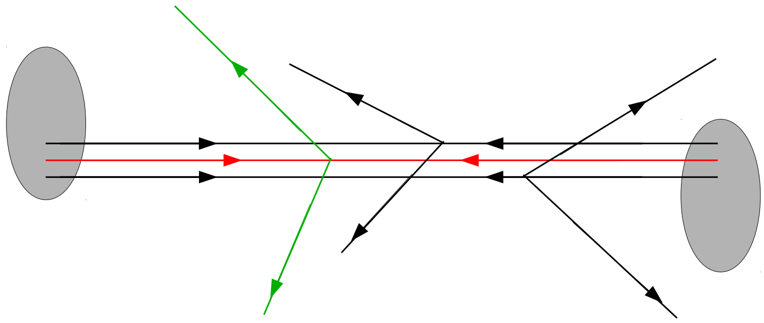


Figure 1: Proton–proton collision in context of MPI only (without parton shower). The coloured lines present hard collision, and the black ones multiple interactions (which could be both hard and soft).

Because of that the proton is a composite object, during collision of two protons more than one pair of partons can interact. This is so called multiple partonic interaction. It can be seen on figure 1. In p–p collision hard collision with high transverse momentum transfer can take place. This is which are mostly interested in, because in such process new particles could be produced. Accompanying activity is the so called underlying event (UE). Underlying event can be hard or soft, and to perform precise measurements of Standard Model (SM) and search for physics beyond SM in LHC we need to have a good understanding of what happens between the interacting protons. Underlying event data are very sensitive to MPI parameters, and can could be used to constrain the parameters.

When measuring UE properties we would like to avoid influence of hard primary scatters. To deal with this the detector volume is usually divided into three regions: toward, away and transverse with respect to leading (hard) track. Generally in toward region we expect activity dominated by particles from hard primary scattering. On the other hand, from momentum conservation, we expect some hard activity, which balance the leading track. Only in the transverse region there should be more or less only the underlying event. So considering this region gives us the best underlying event data.

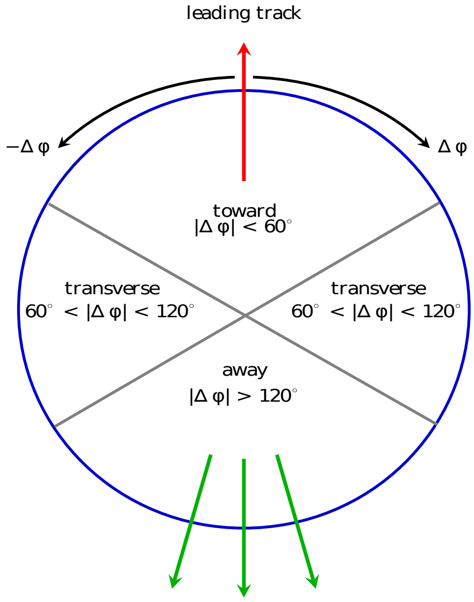


Figure 2: Transverse plane with respect to the beam line. Detector is divided according to the leading track in event into toward, away and transverse region [1].

1.1 Herwig++

The MC generator which I tuned is Herwig++ [3]. It is general purpose MC generator for the simulation for hard lepton–lepton, lepton–hadron, and hadron–hadron collisions. The parton–shower approach is used to simulate initial– and final–state QCD radiation, including colour coherence effects, with special emphasis on the correct description of radiation from heavy particles. The underlying event is simulated using an eikonal multiple parton-parton scattering model. The formation of hadrons from the quarks and gluons produced in the parton shower is described using the cluster hadronization model. In the tuning I have considered five parameters: three from the underlying event model – $\mathrm{p}_{\perp,\mathrm{min}}$, inverse hadron radius squared (μ), and probability of colour disruption ($\mathrm{P}_{\mathrm{disrupt}}$), one from shower evolution, one from shower and hadronization models, respectively – Intrinsic $p_\perp$ Gaussian, and colour reconnection probability [4].

1.2 Professor

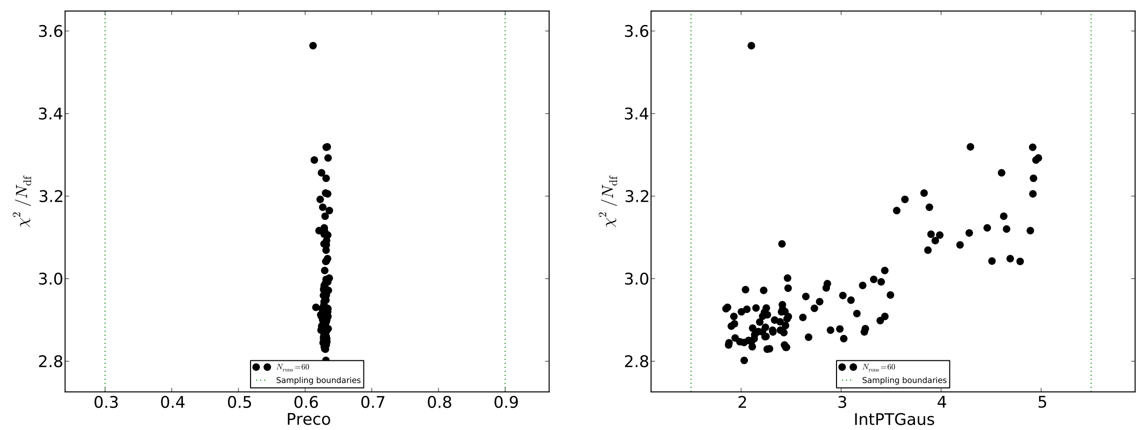
Professor is a tuning tool for Monte Carlo generators to experimental data. The main idea is to parameterise the MC generator response to parameter variations, then numerically finding parameters set, which corresponds to the best fit of the parameterised function to data. This approach needs relatively small number of MC runs (compared to other tuning methods like brute-force scanning parameters space), which demand a lot of CPU power. Furthermore results are quite reliable, and could be tested by Professor-integrated tools. Below I will discuss how I obtained and tested results using Professor tools. [5, 7].

First step to get a tune is to generate MC runs which will be used as anchor points to get the parametrisation. The minimal number of anchor points N_{min}, which we need, depends on the number of parameters which we want to tune, and the type of parameterising function.

Professor provides polynomial of 2nd and 3rd order to parametrise the parameter space. In all my tunes I used a cubic interpolation. Together with fact that I tuned 3,4 and 5 parameters I needed at least 20, 35, 56 anchor points [7]. These numbers are minimal required number of points to obtain interpolation functions parameters. This is described in detail in [5], because of the complicated true MC response it is recommended to give more freedom to interpolation function and use more anchor points. In my tunes I used at least $4 \cdot N_{\min}/3$ points to interpolate response functions.

Furthermore it is important to check if the interpolating function doesn't strongly depend on the choice of anchor points. To verify it different subsets of anchor points should be used for interpolation and tuning. Then results should be compared. I usually used 100 different subsets. Ideally they should be independent, but because of CPU limitation I usually generate only 3-4 times more points than number of points used to interpolate MC response.

Concluding I generated ≈ 300 MC response points distributed randomly in parameter space. Then from these points I chose randomly 100 subsets, each with ≈ 80 points. To each subset I calculated parametrisation and tuning. Finally, using Professor tool I made scatter figure which present how results from different subsets are scattered. Examples of that are presented in figure 3. If everything is correct and parameter is sensitive (choice of anchor points, parametrisation function is sufficient) the results should lie on a vertical straight line, like in figure 4 a). In case that observables are not sensitive to parameter, results are scattered, like in the figure 3 b).



(a) Ideal behaviour of results. Parameter is well constrained by tune.

(b) This figure shows the insensitivity to IntPTGaus parameter

Figure 3: Example of scatter plots obtained from tuning. Observable on left figure is sensitive to tuned data, but observables on right figure not.

2 Tuning

I have done several tunes to underlying event, jet and event shapes data. Each time I used different set of observables to which I tuned, or different anchor points set (especially with different parameters range). In next subsection I will discuss the data analyses that were used, then I will shortly describe all performed tunes. At the end I discuss results and summarise.

2.1 Description of observables.

Because I wanted to tune multiple interaction parameters of Herwig++ the obvious choice for observables was underlying event data from ATLAS and CMS collaborations:

- Track-based underlying event at 900 GeV and 7 TeV as measured by ATLAS (ATLAS_2010_S8894728). This analysis contains underlying event observables like distribution of number of charged particles (N_{chg}), charged particles $p_\perp$ sum ($\Sigma p_\perp$) and their averages with respect to the transverse momentum of the leading track ($p_\perp$(leading)). There are separate distributions for toward, transverse and away region defined by the leading track. Although the analysis contains data for $\sqrt{\text{s}} = 900$ GeV and 7 TeV I used only that with higher energy.

- Charged particles transverse momentum and pseudo-rapidity spectra from proton-proton collisions at 7000 GeV (CMS_2010_S8656010). This analysis contains distributions of number of charged hadron N_{chg} with respect to their transverse momentum $p_\perp$ in different pseudo-rapidity regions $|\eta|$. Data contain only non–single–diffractive events.

- Track-based underlying event at $\sqrt{\text{s}}$=0.9 and 7 TeV from CMS: charged particles analysis transverse to the leading jets (CMS-QCD-10-010). It contains charged particles distributions (N_{ch}, $p_{\perp,ch}$, $\sum p_{\perp,ch}$) with respect to transverse momentum of the leading track. This analysis requires non single diffractive events.

- Measurement of the non–single–diffractive (NSD) charged particles multiplicity at $\sqrt{s} =$ 0.9, 2.36, and 7 TeV with the CMS detector. (CMS_2011_S8884919). Analysis contains primary charged hadron multiplicity distributions for non-single-diffractive events in different pseudo-rapidity ranges. There are data from $\sqrt{s} = 900\,\mathrm{GeV},\ 2.36\,\mathrm{GeV}, \mathrm{and},\ 7\mathrm{TeV}$, but I used only that for the highest energy.

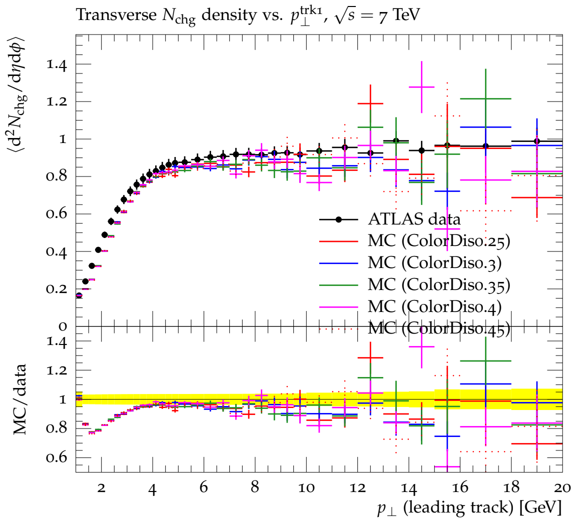
Jet shapes and event shapes can in principle be influenced by the underlying event, so I included in tuning also these measurements. I considered two sets of data:

- Event shapes (CMS_2011_S8957746). This analysis contains distributions of two event-shapes variables – the central transverse thrust τ_C and the central thrust minor $\tau_{m,C}$. The data are divided into three bins of the leading jet transverse momentum $p_\perp^{jet1}$.

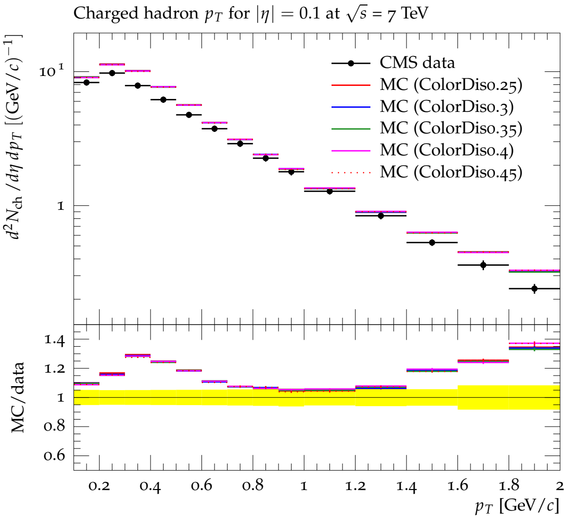
- Jet shapes at 7 TeV as measured by ATLAS (ATLAS_2011_S8924791). It contains distributions of differential and integrated jet shapes $\rho(r)$ and $\Psi(r)$ respectively, in different jet $p_\perp$ and pseudo-rapidity value. $\rho(r)$ is defined as a average fraction of the jet transverse momentum that lies inside an annulus of inner radius $r - \Delta r/2$ and outer radius $r + \Delta r/2$ around the jet axis. $\Psi(r)$ is a average fraction of the jet transverse momentum that lies inside a cone of radius r.

2.2 Small dependence on probability for colour disruption

I scanned the parameter space to investigate which parameters have an influence on considered observables. I varied one parameter at a time and looked at the spread of the predictions. From these study I realized that changes in probability for colour disruption in range 0.2–0.4 have no significant influence on the underlying event, jet and event shapes observables. In figure 4 examples of these variations are shown. Shown are MC runs with different Colour Disruption



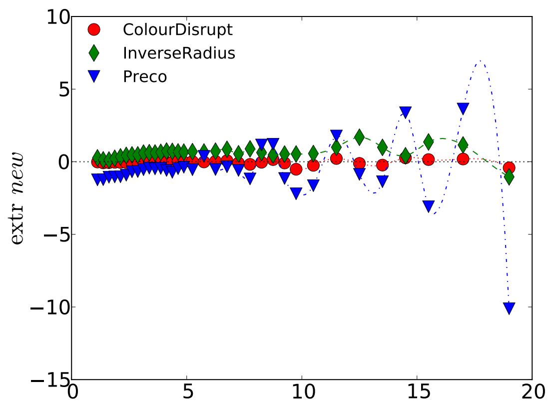
(a) Number of charged particles (N_{ch}) density with respect to leading track transverse momentum $p_{\perp}$(leading) in transverse region of detector(see figure 2) from ATLAS experiment.



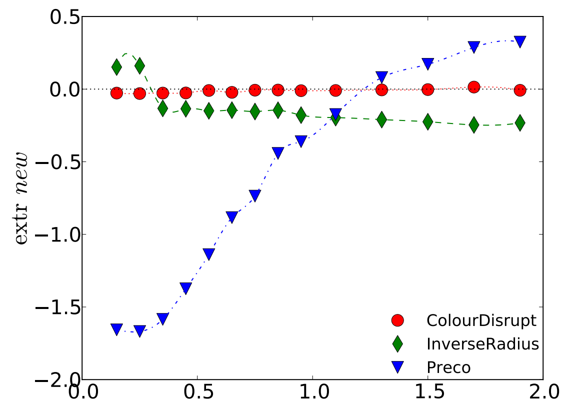
(b) Transverse momentum p_T distribution of charged hadron in $|\eta| < 0.1$ region from CMS experiment.

Figure 4: MC predictions with different value of colour disruption probability. Difference between the choices are compatible with statistic fluctuations.

probability and other parameters kept at default value. Besides this I made sensitivity plots for underlying event observables. Sensitivity plots are based on MC runs with value of $P_{disrupt}$ selected randomly from 0 to 1. In figure 5 can be seen that there is no significant influence of $P_{disrupt}$ on observable. Because of that I decided to not include $P_{disrupt}$ in further tunes and set it to the default Herwig++ value of $P_{disrupt} = 0.3493643$.



(a) Sensitivity of three parameters to density of N_{ch} with respect to $p_\perp$ (The same observables like in figure 4 (a). High deviation of P_{reco} sensitivity curve is caused by statistical fluctuations.)



(b) Sensitivity of three parameters to distribution of p_T for charged hadron in $|\eta|$ <0.1 region (see figure 4 (b))

Figure 5: Examples of sensitivity observable on parameters values. The deviation from zero is a measure of sensitivity, with zero indicating no sensitivity.

2.3 Different tuning

In this report I present results of the following tunes.

- "5th tune" family – based on anchor points set, in which I varied three MPI parameters: inverse hadron radius squared (μ), $p_{\perp,\mathrm{min}}$, and colour reconnection probability (P_{reco}) in ranges shown in table 1. Only the underlying event observables have been considered. In total I generated 292 MC points(each contains 500k events), from which I chose 100 subsets each with 61 points for interpolation. I performed three tunes with these sets. First, including only ATLAS underlying events data (analysis ATLAS_2010_S8894728 [1]; Used observables are listed in appendix A.2). I will call that tune A_UE tune. Second, using CMS charged hadron spectra (CHS) data (CMS_2010_S8656010 [8]; observables are in appendix A.3). Note that Herwig++ has so far only been tuned to the ATLAS underlying event data. I will call that tune CMS_CHS tune. At the end I did combined tuning to ATLAS and CMS CHS data, using the same observables, with equal weights, like in previous described tunings. I named it A_UE_CMS_CHS tune.

- the "5thCTEQ6L" family – same as in the 5th tune's family except that I used CTEQ6L PDF set instead of default Herwig++ one, MSTW LO**.

- "8th tune" family. In this case I varied four parameters – μ, $p_{\perp,min}$, and P_{reco} and in addition the width of the Gaussian of the intrinsic $p_\perp$ (IntPtGauss). Production ranges are shown in table 1. Moreover I used underlying event, and also jet shapes(ATLAS_2011_S8924791 [2]) and event shapes analyses (CMS_2011_S8957746 [9]). In case of jet and event shapes analyses I used QCD $2 \to 2$ matrix elements instead of the minimum bias matrix element recommended for underlying event simulation. Furthermore I applied generator cut on the jet $p_\perp$ in order to get sufficient number of events in $p_\perp$ range studied in these analyses. In appendix A.1 is describe exactly what cuts I have been using for which observables. In total I generated 197 MC points(each contains 400k events), from which I chose 100 subsets each with 60 points for the interpolation. I performed six tunes with these sets. The first three were tunes to underlying event and charged hadron spectra (CHS) like in 5th tune's family ("ATLAS UE" , "CMS CHS (2010)", and combined "ATLAS, CMS CHS" tune). Additionally I performed tunes to CMS event shapes data (I will call it "CMS EShapes"), combined CMS event shapes and ATLAS Jet shapes data (called JetEventS), combined ATLAS UE, Jet and Event shapes data (called ATLAS UE & JetEventS), and combined ATLAS, CMS 2010 UE, Jet and Event shapes data (called ATLAS CMS UE & JetEventSS). Observables used for the tuning are listed in tables 4, 5, 6, and 7 the appendix.

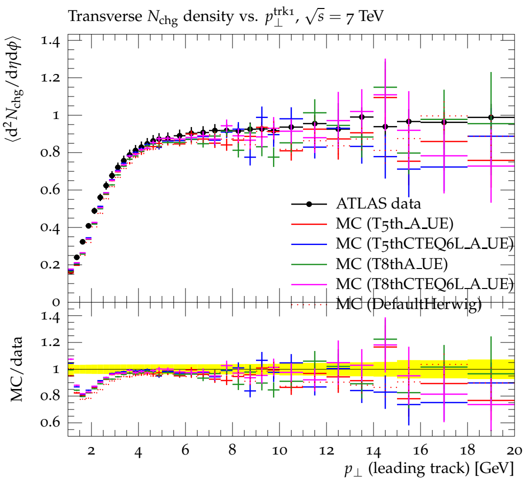
- "8thCTEQ6L1" family. The only difference between these and "8th tune" family are using CTQ6L PDF set instead of default one. Also chose different anchor points, and unfortunately because of model constraints on the parameters, I got only 57 anchor points at the end for that family. Therefore results for this family may need more statistics. However the same tunings as in "8th tune" family were done.

- "10th tune" family. These tunes (together with 10thCTEQ6L1) are last tunes. Parameters ranges were chosen according to previous experience. Therefore I varied three parameters – μ, ptMin, and P_{reco}, in ranges shown in table 1. I performed only underlying event analyses. In total I generated 149 MC points, from which I chose 100 subsets each with 60 points for interpolation. For each parameters set I computed 1.5 millions events I performed six tunes with these sets. I used more events in each points than in previous tunes to have enough statistic for CMS UE (CMS_QCD_10_010 [6]) data. Therefore I was able to perform tunes to CMS UE data. In general in "10th tune" family I performed six different tunes: to ATLAS underlying event data (A_UE); CMS underlying event data (CMS_UE_all); selected CMS underlying event data (CMS_UEsel); CMS charged hadron spectra data (CMS_CHS); combined CMS and ATLAS underlying event data (A_UE_CMS_UEsel); combined CMS underlying event and charged hadron spectrum data (CMS_UEsel_CHS)

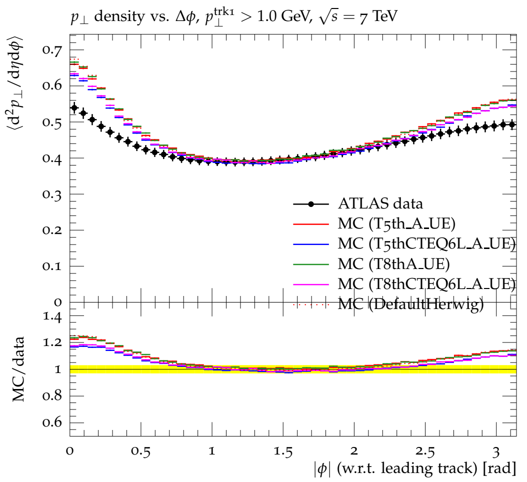
- "10thCTEQ6L1" family. The only difference between these and "10th tune" family are using CTQ6L1 PDF set instead of default one. In fact I also chose different anchor points. I got 171 anchor points at the end for that family. Each point has 1.5 millions events, and the same tunings as in "10th tune" family were done.

	Parameters range					
	$p_{\perp}$	μ^2	P_{reco}	IntPtGauss	PDF set	MC points
5th	[3.0 − 5.0]	[0.6 − 2.0]	[0.5 − 0.8]	fixed 2.2	default	292
5thCTEQ6L1	[3.0 − 5.0]	[0.6 − 2.0]	[0.5 − 0.8]	fixed 2.2	CTEQ6L	292
8th	[3.0 − 5.0]	[0.6 − 2.0]	[0.3 − 0.9]	[1.5 − 5.5]	default	197
8thCTEQ6L1	[3.0 − 5.0]	[0.6 − 2.0]	[0.3 − 0.9]	[1.5 − 5.5]	CTEQ6L	57
10th	[2.8 − 4.2]	[0.5 − 1.5]	[0.45 − 0.75]	fixed 2.2	default	149
10thCTEQ6L1	[2.8 − 4.2]	[0.5 − 1.5]	[0.45 − 0.75]	fixed 2.2	CTEQ6L	171

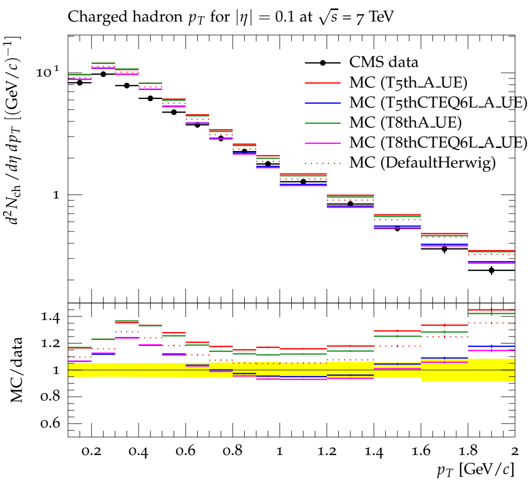
Table 1: Settings used in different tunes.



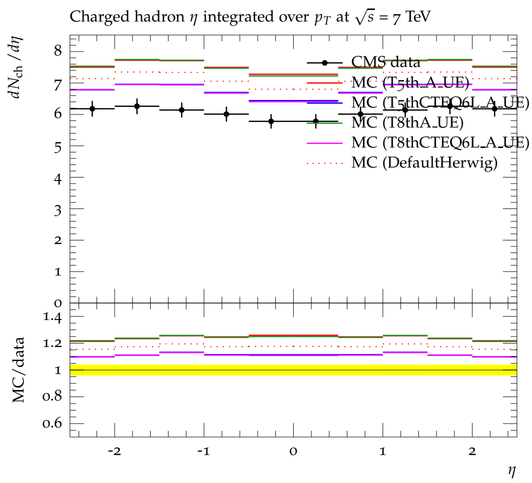
(a) N_{ch} density to leading track $p_\perp$.



(b) Transverse momentum $p_\perp$ angular density from Atlas detector.

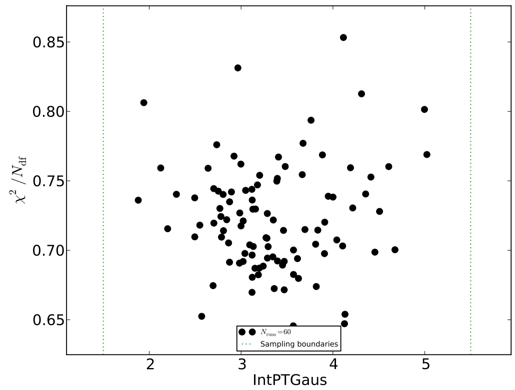


(c) Distribution of charged hadron transverse momentum for $|\eta| < 0.1$ from CMS CHS analysis.

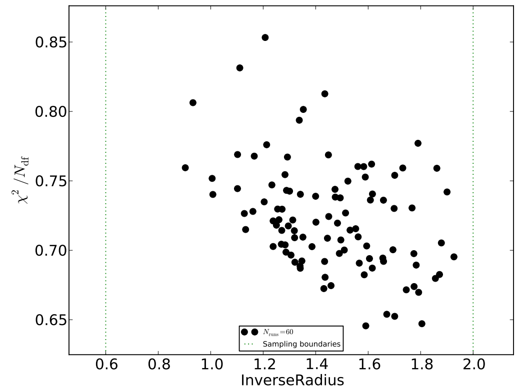


(d) Number of charged hadrons in different η region.

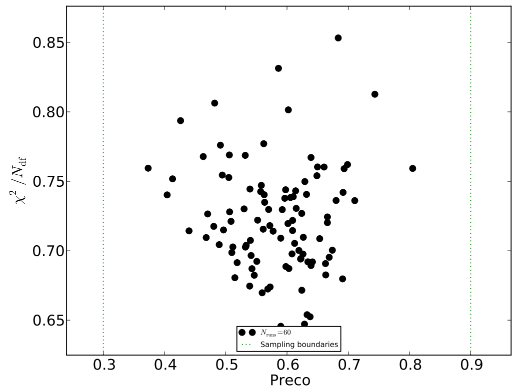
Figure 6: Predictions from tunes compared to ATLAS UE and CMS CHS data. Generally all MC describe ATLAS UE data very well, but that with CTEQ61L PDFs set describe CMS CHS data much better than other. MC prediction based on MSTW LO** PDFs set (default PDFs set) are worst than default Herwig++ parameters set. It means that there is a tension between ATLAS UE and CMS CHS data (ATLAS and CMS data prefer different parameter values). This tension is smaller using CTEQ6L1 PDFs set, and predictions are better than default Herwig++.



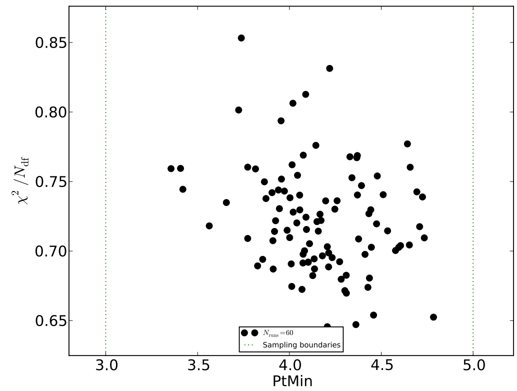
(a) Scatter plot for intrinsic $p_\perp$ Gaussian.



(b) Scatter plot for inverse hadron radius squared μ.



(c) Scatter plot for reconnection probability.



(d) Scatter plot for $p_{\perp,min}$.

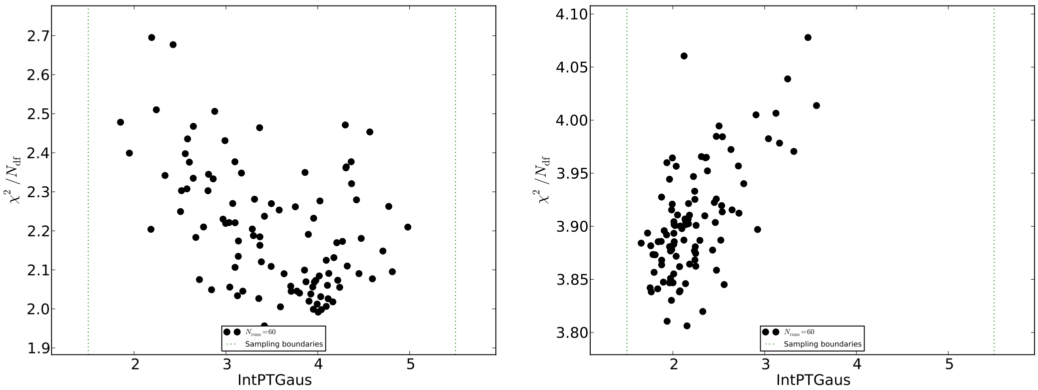
Figure 7: Scatter plots made from tune from jet shapes and event shapes data. All plots show results which are very scattered. It means that none of parameters are constrained by that data.

2.4 Discussion of results

Table 2 shows results of different tunes. The following conclusions can be drawn:

- Using the CTEQ6L PDF set instead of default gives generally lower χ^2 value. The better agreement with data can be seen in figure 6. Some examples of underlying events data from ATLAS and CMS are shown. All tunes shown in figure 6 were done only to ATLAS UE, and they describe slightly better ATLAS UE data than default Herwig++ tune. Significant improvement has been observed in describing CMS UE data by using the CTEQ6L PDF set. Figures 6 c) and d) show examples of this. One can see that tension [1] between ATLAS_2010_S8894728 and CMS_2010_S8656010 data are smaller .

[1] ATLAS and CMS data prefer different parameters values.



(a) Scatter plot from ATLAS UE tune from "8th family"

(b) Scatter plot from combined ATLAS UE and CMS CHS tune from "8th family"

Figure 8: IntPtGauss scatter plots obtained from ATLAS UE tune, and combined ATLAS UE and CMS CHS tune. Results are very scattered, what means that intrinsic $p_\perp$ Gaussian can not be obtained from considered data.

- Jet and event shape data have small impact on tuning results. This is shown in figure 7. It shows scatter plots obtained from tune from jet shapes and event shapes data (to "8th family").

- According to table 2, tunes performed with the CTEQ6L PDF set prefer smaller value of $p_\perp$. It means that for CTEQ6L perturbative QCD is valid to smaller energy scale. It is also visible that generally results of $p_\perp$ after tuning are in boundary of chosen range ($[3 \rightarrow 5]$). In future tunes range of $p_\perp$ should be smaller (without jet and event shapes data it should be something like $[2.7 - 4.0]$).

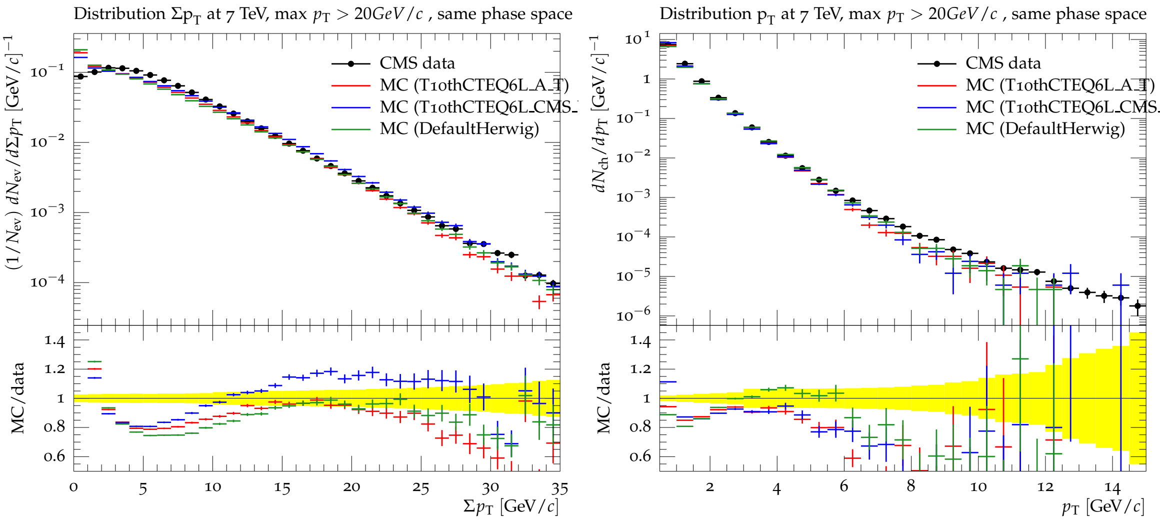
- Tuning only to CMS UE data (CMS_2010_S8656010) gave unphysical results. It is probably caused by the method of reducing diffractive collisions in CMS data: single-diffractive dissociation events were distinguished from non-single-diffractive events based on simulation in PYTHIA MC generator.

- It looks like intrinsic $p_\perp$ Gaussian has very small (if any) influence on underlying event data. There is almost no difference in χ^2 between tunings with fixed (5thTune and 5thCTEQ6LTune) and varied IntPtGauss (8thTune and CTEQ6LBoth). At the same time, errors on IntPtGauss are very large. The insensitivity can also be seen from the scatter plots in figure 8. In future tuning IntPtGauss could be omitted from the tuning, unless a more constraining observable like the Drell-Yan $p_\perp$ spectrum is included as well.

- From table 2 it is seen that there is almost no tension between ATLAS and CMS (selected[2]) underlying event data (taking into account errors of predicted parameter). Even errors on obtained parameters are on the same level for both cases. However is worth to mention that tune from CMS selected UE data describe better CMS CHS data, ATLAS UE data worse, than default, or my tune from ATLAS UE data. Figure 9 shows this behaviour. It presents comparison of ATLAS UE, CMS UE and CMS CHS

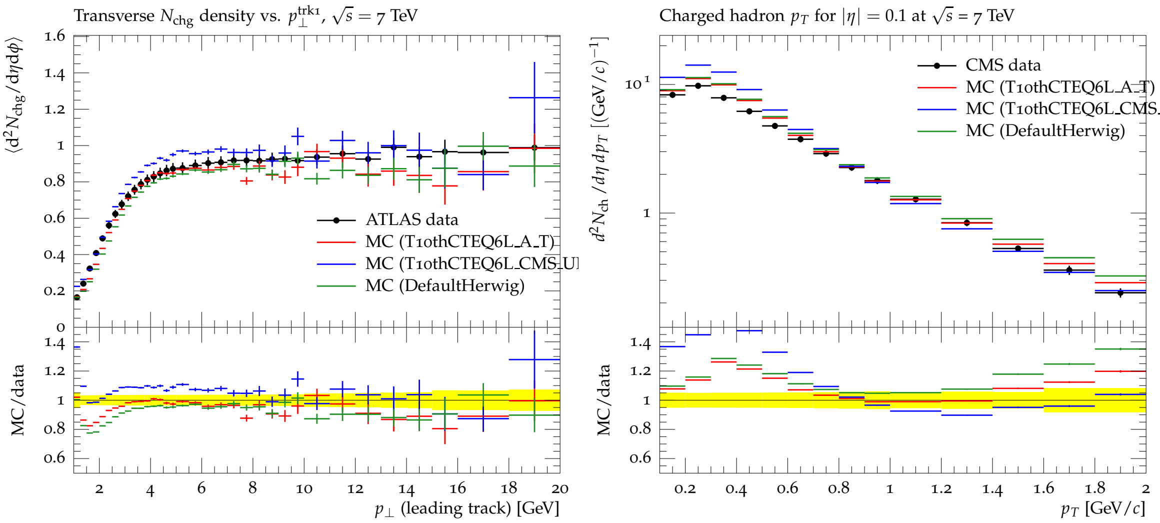
[2]Only observable listed in table 9 were used.

data to MC tunes. Generally CMS UE data are not very well predicted by MCs. It can be seen from high χ^2 value (table 2), as well as from comparison to data in figure 9. Problem is probably the same as in tune from CMS CHS data, the method of reducing diffractive collisions in CMS data.



(a) Distribution of total transverse momentum $\Sigma p_\perp$. All tunes have problem with describing the data.

(b) [Distribution of transverse momentum $\Sigma p_\perp$. MC in this case have low statistic in tail of distribution.



(c) Number of charged particles (N_{ch}) density with respect to leading track transverse momentum $p_\perp$(leading) in transverse region of detector from ATLAS.

(d) Number of charged hadrons in different η region. Tune from CMS selected UE data describes data the best.

Figure 9: Some example of CMS UE, ATLAS UE and CMS CHS data with MC prediction (different tunes – tune from ATLAS UE data (using CTEQ6L1 PDFs set – T10thCTEQ_A_T on legends), CMS selected UE data (using CTEQ6L1 PDFs set – T10thCTEQ_CMS_UEsel), or default one – DefaultHerwig). Generally tune from CMS UE data predict only CMS CHS data better than other tunes.

		Parameters				
		$p_\perp$	μ	P_{reco}	IntPtGauss	χ^2/Ndf
ATLAS UE						
	5th	3.152815 ± 0.5396797	0.7428927 ± 0.1604285	0.5846110 ± 0.09905312	2.2	≈ 2.0
	5thCTEQ6L1	3.030393 ± 0.4988861	0.9737598 ± 0.2407679	0.6336742 ± 0.09257186	2.2	≈ 1.7
	8th	3.207499 ± 0.5028249	0.7593066 ± 0.1768789	0.5809417 ± 0.1035801	3.602890 ± 3.892025	≈ 2.2
	8thCTEQ6L1	3.108931 ± 0.4681086	1.030902 ± 0.2478281	0.6313351 ± 0.09679149	4.176520 ± 2.256789	≈ 1.7
	10th	3.129892 ± 0.4254336	0.7375031 ± 0.1448087	0.5789712 ± 0.1037525	2.2	≈ 2.1
	10thCTQ6L1	2.923713 ± 0.4164518	0.9251335 ± 0.2272152	0.6264389 ± 0.1018729	2.2	≈ 1.5
CMS UE all data						
	10th	3.470722 ± 0.2587991	0.9273738 ± 0.1129382	0.5054889 ± 0.03505769	2.2	≈ 35
	10thCTQ6L1	3.192352 ± 0.1819538	1.134238 ± 0.08307842	0.5362138 ± 0.03607070	2.2	≈ 30
CMS UEsel						
	10th	3.007353 ± 0.4550958	0.8609347 ± 0.1424959	0.4677092 ± 0.1111187	2.2	≈ 30
	10thCTQ6L1	2.867026 ± 0.2831136	1.009589 ± 0.1547239	0.6027205 ± 0.08834702	2.2	≈ 30
ATLAS UE, CMS CHS						
	5th	3.739264 ± 0.7611599	0.9469332 ± 0.3065160	0.6313965 ± 0.07341302	2.2	≈ 3.9
	5thCTEQ6L1	3.073268 ± 0.4580006	1.017645 ± 0.2307892	0.6604675 ± 0.07710487	2.2	≈ 2.0
	8th	3.733241 ± 0.7584721	0.9487589 ± 0.3108400	0.6312224 ± 0.08129311	2.111966 ± 7.093361	≈ 3.9
	8thCTEQ6L1	3.129202 ± 0.4330447	1.069229 ± 0.2225444	0.6583470 ± 0.07929201	4.072058 ± 2.058187	≈ 2.1
CMS CHS						
	5th	3.818088 ± 0.3245146	$-0.07415699 \pm 0.2691693$	0.5494709 ± 0.07857292	2.2	≈ 2.5
	5thCTEQ6L1	3.059173 ± 0.2732315	0.1340038 ± 0.4367015	0.6105759 ± 0.08409025	2.2	≈ 1.4
	8th	3.898064 ± 0.4717944	0.05413113 ± 0.3190096	0.5801110 ± 0.1102095	2.094103 ± 5.271108	≈ 3.0
	8thCTEQ6L1	2.995983 ± 0.3426049	0.5765091 ± 0.3387419	0.6224081 ± 0.07966010	4.192534 ± 4.206630	≈ 1.4
	10th	4.131145 ± 0.8265671	0.1373492 ± 0.1974078	0.6353554 ± 0.1059367	2.2	≈ 3.5
	10thCTEQ6L1	3.219499 ± 0.4203372	0.3847307 ± 0.4556989	0.6483182 ± 0.05962080	2.2	≈ 2.8
CMS UEsel_CHS						
	10th	3.805796 ± 0.5237726	1.150616 ± 0.1995292	0.6380649 ± 0.05554572	2.2	≈ 21
	10thCTEQ6L1	2.966628 ± 0.2608620	$1.182656 \pm 0.1.628795$	0.6843106 ± 0.05086262	2.2	≈ 12
ATLAS_UE_Jet_CMS_EventS						
	8th	3.392380 ± 1.073847	0.8165268 ± 0.4245646	0.5913418 ± 0.1301148	3.909508 ± 4.994173	≈ 1.9
	8thCTEQ6L1	3.176116 ± 0.5772612	1.050827 ± 0.3490147	0.6300968 ± 0.1195389	3.982970 ± 1.107424	≈ 1.5
CMS EventS						
	8th	4.012561 ± 1.464581	1.433025 ± 0.8428927	0.6895451 ± 0.5690749	3.647995 ± 4.099829	≈ 0.9
	8thCTEQ6L1	4.122858 ± 1.022083	1.164676 ± 0.6624232	0.6726206 ± 0.2134214	2.092204 ± 1.496788	≈ 0.8
JetEventS						
	8th	4.229340 ± 3.593653	1.493113 ± 1.380206	0.5886013 ± 0.5949717	3.230931 ± 5.314123	≈ 0.7
	8thCTEQ6L1	4.160737 ± 1.540725	1.387515 ± 0.6708315	0.5435066 ± 0.4816282	2.711399 ± 5.234753	≈ 0.9
ATLAS, CMS CHS						
& JetEventS	8th	3.780009 ± 1.485674	0.9575673 ± 0.5603852	0.6291066 ± 0.09246440	2.885267 ± 4.631621	≈ 3.0
	8thCTEQ6L1	3.248976 ± 0.4523929	1.100353 ± 0.2814493	0.6546816 ± 0.09386104	3.751962 ± 1.141261	≈ 1.8

Table 2: Results of performed tunes. They are divided with respect of observables used in tunes. Presented parameters are: $p_{\perp,min}$ – PtMin, inverse hadron radius – μ, colour reconnection probability – P_{reco}, and Intrinsic $p_\perp$ Gaussian – IntPtGauss. Observables sets used in tunes: ATLAS underlying event data (ATLAS UE); CMS underlying event data (CMS_UE_all); selected CMS underlying event data (CMS_UEsel); ATLAS underlying event and CMS charged hadron spectrum data (ATLAS_UE,CMS_CHS);CMS charged hadron spectrum data (CMS_CHS); combined CMS charged hadron spectrum and underlying event data (CMS UE_CHS); combined ATLAS underlying event, jet shapes, and CMS event shapes data (ATLAS_UE_Jet_CMS_EventS); CMS event shapes (CMS_EventS); combined ATLAS jet shapes and CMS event shapes data (JetEventS); combined ATLAS_UE, CMS_CHS, and JetEventS data (ATLAS_UE_CMS_CHS_JetEventS)

2.5 Conclusions and Outlook

Main outcome of my work is that the CTEQ6L1 PDFs set gives better results than default MSTW LO**. The similar observation was report also in case of PYTHIA MC generator. Maybe the CTEQ6L1 PDFs set should be default PDFs set.

During my work general problem with describing CMS data was discovered (without problem with ATLAS data). It is probably caused by method of reducing diffractive collisions in CMS data. It appeared that data with operational definition of diffraction reduction, like $N_{ch} \geq 6$, $p_{\perp}(\text{leadingtrack}) > 500\text{MeV}$ are preferred.

Moreover it turns out that jet and event shapes data are not sensitive to $p_{\perp}$, μ, and P_{reco}, intrinsic $p_{\perp}$ Gaussian parameters. These parameters are weakly constrained by jet and event shapes data.

Furthermore parameter intrinsic $p_{\perp}$ Gaussian has no influence on underlying event data, as well as on charged hadron spectrum data.

General conclusion is that more tunes, especially with CMS UE data should be performed. Number of events in this case must be higher, or different methods must be applied to minimise statistic fluctuation (like in figure 9 (b)).

Acknowledgements

I would like to thank Albert Knutsson and Simon Plätzer, who supervised me during DESY Summer Students Program 2011. I would not have gotten results without their help and support. I would like to thank especially for their patience and forbearance. Moreover I am grateful for hours they spent discussing with me, solving my problems. Last but not least I would like to thank for all coffee, which we had together. It was a pleasure for me to work with them.

On the other hand I would like to thank the organisers of DESY Summer Students Program 2011 for inviting me and doing everything that I would enjoy time in DESY. Special thanks to Olaf Behnke for his idea what we (summer students) can do in Hamburg and around.

Finally I would like to thank all of summer students for time which we spend together.

A Observables used in tunings

A.1 Used jet $p_\perp$ cuts

In case of jet shapes and event shapes data I used several jet transverse momentum $p_{\perp,min}$ cuts. In table 3 cuts and corresponding observables are listed for ATLAS jet event shapes (ATLAS_2011_S8924791) and CMS event shapes data (CMS_2011_S8957746).

cut	observable from ATLAS jet shapes data	observable from CMS event shapes data
20	d01-x0*-y01 d01-x0*-y02 d02-x0*-y01 d02-x0*-y02 d03-x0*-y01 d03-x0*-y02 d04-x0*-y01 d04-x0*-y02	—
80	d05-x0*-y01 d05-x0*-y02 d06-x0*-y01 d06-x0*-y02	d01-x01-y01 d01-x01-y02 d01-x02-y01 d01-x02-y02
190	d07-x0*-y01 d07-x0*-y02 d08-x0*-y01 d08-x0*-y02 d09-x0*-y01 d09-x0*-y02	d01-x03-y01 d02-x03-y01
380	d10-x0*-y01 d10-x0*-y02 d11-x06-y01 d11-x06-y02	—

Table 3: Cuts and corresponding observables from ATLAS jet shapes and CMS event shapes data.

A.2 ATLAS UE Tunes

Tunes based on "Track-based underlying event at 900 GeV and 7 TeV in ATLAS (ATLAS_2010_S8894728)" analysis with $\sqrt{s} = 7$ TeV. Used observables are listed on table 4

Observable	Weight
Transverse N_{chg} density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Toward N_{chg} density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Away N_{chg} density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Transverse $\sum p_\perp$ density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Toward $\sum p_\perp$ density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Away $\sum p_\perp$ density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Std. dev. Transverse N_{chg} density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Std. dev. Transverse $\sum p_\perp$ density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Transverse $\langle p_\perp \rangle$ vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Toward $\langle p_\perp \rangle$ vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Away $\langle p_\perp \rangle$ vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV	1.0
Transverse $\langle p_\perp \rangle$ vs. N_{chg}, $\sqrt{s} = 7$ TeV	1.0
Toward $\langle p_\perp \rangle$ vs. N_{chg}, $\sqrt{s} = 7$ TeV	1.0
Away $\langle p_\perp \rangle$ vs. N_{chg}, $\sqrt{s} = 7$ TeV	1.0
N_{chg} density vs. $\Delta\phi$, $p_\perp^{\text{trk1}} > 1.0$ GeV, $\sqrt{s} = 7$ TeV	1.0
N_{chg} density vs. $\Delta\phi$, $p_\perp^{\text{trk1}} > 2.0$ GeV, $\sqrt{s} = 7$ TeV	1.0
N_{chg} density vs. $\Delta\phi$, $p_\perp^{\text{trk1}} > 3.0$ GeV, $\sqrt{s} = 7$ TeV	1.0
N_{chg} density vs. $\Delta\phi$, $p_\perp^{\text{trk1}} > 5.0$ GeV, $\sqrt{s} = 7$ TeV	1.0
$p_\perp$ density vs. $\Delta\phi$, $p_\perp^{\text{trk1}} > 1.0$ GeV, $\sqrt{s} = 7$ TeV	1.0
$p_\perp$ density vs. $\Delta\phi$, $p_\perp^{\text{trk1}} > 2.0$ GeV, $\sqrt{s} = 7$ TeV	1.0
$p_\perp$ density vs. $\Delta\phi$, $p_\perp^{\text{trk1}} > 3.0$ GeV, $\sqrt{s} = 7$ TeV	1.0
$p_\perp$ density vs. $\Delta\phi$, $p_\perp^{\text{trk1}} > 5.0$ GeV, $\sqrt{s} = 7$ TeV	1.0
Transverse N_{chg} density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV, $p_\perp > 100$ MeV	1.0
Toward N_{chg} density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV, $p_\perp > 100$ MeV	1.0
Away N_{chg} density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV, $p_\perp > 100$ MeV	1.0
Transverse $\sum p_\perp$ density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV, $p_\perp > 100$ MeV	1.0
Toward $\sum p_\perp$ density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV, $p_\perp > 100$ MeV	1.0
Away $\sum p_\perp$ density vs. $p_\perp^{\text{trk1}}$, $\sqrt{s} = 7$ TeV, $p_\perp > 100$ MeV	1.0
Transverse N_{chg} density vs. $\lvert\eta^{\text{trk1}}\rvert$, $\sqrt{s} = 7$ TeV, $p_\perp > 100$ MeV	1.0
Transverse $\sum p_\perp$ density vs. $\lvert\eta^{\text{trk1}}\rvert$, $\sqrt{s} = 7$ TeV, $p_\perp > 100$ MeV	1.0

Table 4: Observables used in ATLAS tunes.

A.3 CMS CHS tunes

Tunes based on "Charged particles transverse momentum and pseudo-rapidity spectra from proton-proton collisions at 7000 GeV (CMS_2010_S8656010)" analysis. Used observables are listed on table 5

Observable	Weight
Charged hadron p_T for $\lvert\eta\rvert = 0.1$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 0.3$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 0.5$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 0.7$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 0.9$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 1.1$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 1.3$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 1.5$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 1.7$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 1.9$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 2.1$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert = 2.3$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron p_T for $\lvert\eta\rvert < 2.4$ at $\sqrt{s} = 7$ TeV	1.0
Charged hadron η integrated over p_T at $\sqrt{s} = 7$ TeV	1.0

Table 5: Observables used in CMS_CHS tunes.

A.4 Jet shapes observables

Observables from "Jet shapes at 7 TeV in ATLAS (ATLAS_2011_S8924791)" analysis used in "JetEventS", "ATLAS UE & JetEventS", and "ATLAS, CMS UE & JetEventS" tunes (table 6).

Observable	Weight
Jet shape Ψ for $p_\perp \in$ 30–40 GeV, $y \in$ 0.0–0.3	1.0
Jet shape ρ for $p_\perp \in$ 30–40 GeV, 40–60 GeV, 60–80 GeV, 80–110 GeV, 110–160 GeV, 160–210 GeV, 210–260 GeV, 260–310 GeV, 310–400 GeV, 400–500 GeV, 500–600 GeV, $y \in$ 0.3–0.8	each bin 1.0
Jet shape Ψ for $p_\perp \in$ 30–40 GeV, 40–60 GeV, 60–80 GeV, 80–110 GeV, 110–160 GeV, 160–210 GeV, 210–260 GeV, 260–310 GeV, 310–400 GeV, 400–500 GeV, 500–600 GeV, $y \in$ 0.3–0.8	each bin 1.0
Jet shape ρ for $p_\perp \in$ 30–40 GeV, 40–60 GeV, 60–80 GeV, 80–110 GeV, 110–160 GeV, 160–210 GeV, 210–260 GeV, 260–310 GeV, 310–400 GeV, 400–500 GeV, 500–600 GeV, $y \in$ 0.8–1.2	each bin 1.0
Jet shape Ψ for $p_\perp \in$ 30–40 GeV, 40–60 GeV, 60–80 GeV, 80–110 GeV, 110–160 GeV, 160–210 GeV, 210–260 GeV, 260–310 GeV, 310–400 GeV, 400–500 GeV, 500–600 GeV, $y \in$ 0.8–1.2	each bin 1.0
Jet shape ρ for $p_\perp \in$ 30–40 GeV, 40–60 GeV, 60–80 GeV, 80–110 GeV, 110–160 GeV, 160–210 GeV, 210–260 GeV, 260–310 GeV, 310–400 GeV, 400–500 GeV, 500–600 GeV, $y \in$ 1.2–2.1	each bin 1.0
Jet shape Ψ for $p_\perp \in$ 30–40 GeV, 40–60 GeV, 60–80 GeV, 80–110 GeV, 110–160 GeV, 160–210 GeV, 210–260 GeV, 260–310 GeV, 310–400 GeV, 400–500 GeV, 500–600 GeV, $y \in$ 1.2–2.1	each bin 1.0
Jet shape ρ for $p_\perp \in$ 30–40 GeV, 40–60 GeV, 60–80 GeV, 80–110 GeV, 110–160 GeV, 160–210 GeV, 210–260 GeV, 260–310 GeV, 310–400 GeV, 400–500 GeV, 500–600 GeV, $y \in$ 2.1–2.8	each bin 1.0
Jet shape Ψ for $p_\perp \in$ 30–40 GeV, 40–60 GeV, 60–80 GeV, 80–110 GeV, 110–160 GeV, 160–210 GeV, 210–260 GeV, 260–310 GeV, 310–400 GeV, 400–500 GeV, 500–600 GeV, $y \in$ 2.1–2.8	each bin 1.0
Jet shape ρ for $p_\perp \in$ 30–40 GeV, 40–60 GeV, 60–80 GeV, 80–110 GeV, 110–160 GeV, 160–210 GeV, 210–260 GeV, 260–310 GeV, 310–400 GeV, 400–500 GeV, 500–600 GeV, $y \in$ 0.0–2.8	each bin 1.0
Jet shape ρ for $p_\perp \in$ 30–600 GeV, $y \in$ 0.0–0.3	1.0

Table 6: Used observables from ATLAS Jet shapes analysis.

A.5 Event shapes observables

Tunes based on "Event shapes (CMS_2011_S8957746)" analysis. Used observables are listed on table 7

Observable	Weight
Central Transv. Thrust, $90\,\text{GeV} < p_\perp^{\text{jet }1} < 125\,\text{GeV}$, $\sqrt{s} = 7\,\text{TeV}$	1.0
Central Transv. Minor, $125\,\text{GeV} < p_\perp^{\text{jet }1} < 200\,\text{GeV}$, $\sqrt{s} = 7\,\text{TeV}$	1.0
Central Transv. Minor, $p_\perp^{\text{jet }1} > 200\,\text{GeV}$, $\sqrt{s} = 7\,\text{TeV}$	1.0
Jet shape ρ for $p_\perp \in$ 30–40 GeV, $y \in$ 0.0–0.3	1.0
Jet shape Ψ for $p_\perp \in$ 30–40 GeV, $y \in$ 0.0–0.3	1.0
Jet shape ρ for $p_\perp \in$ 30–40 GeV, $y \in$ 0.3–0.8	1.0
Jet shape Ψ for $p_\perp \in$ 30–40 GeV, $y \in$ 0.3–0.8	1.0

Table 7: Used observables from ATLAS Jet shapes analysis.

A.6 CMS underlying event observables

Tunes based on "Underlying event data (CMS_QCD_10_010)" analysis. Used observables are listed on table 8

Observable	Weight
ProfileN$_{\text{ch}}$at$\sqrt{s}$=7TeV,60 $< \lvert\Delta\phi\rvert <$ 120,$\lvert\eta\rvert < 2$	1.0
ProfileΣp$_{\text{T}}$at$\sqrt{s}$=7 TeV,60 $< \lvert\Delta\phi\rvert <$ 120,$\lvert\eta\rvert < 2$	1.0
DistributionN$_{\text{ch}}$at $\sqrt{s}$=7 TeV,maxp_{T} >20 GeV/c, same phase space	1.0
DistributionΣp$_{\text{T}}$at $\sqrt{s}$=7 TeV,maxp_{T} >20 GeV/c, same phase space	1.0
Distributionp$_{\text{T}}$at $\sqrt{s}$=7 TeV,maxp_{T} >20 GeV/c, same phase space	1.0
DistributionN$_{\text{ch}}$at $\sqrt{s}$=7 TeV,maxp_{T} >3 GeV/c, same phase space	1.0
DistributionΣp$_{\text{T}}$at $\sqrt{s}$=7 TeV,maxp_{T} >3 GeV/c, same phase space	1.0
Distributionp$_{\text{T}}$at $\sqrt{s}$=7 TeV,maxp_{T} >3 GeV/c, same phase space	1.0

Table 8: Used observables from CMS underlying event data in "CMS underlying event" tunes.

A.7 CMS selected underlying event observables

Tunes based on "Underlying event data (CMS_QCD_10_010)" analysis, but observables were chosen according to these with relatively small statistic errors. Used observables are listed on table 9. These observables were used in "CMS_UEsel", "CMS_UEsel_CHS", and seven times bigger in "A_UE_CMS_UEsel" tunes.

Observable	Weight
DistributionN$_{\text{ch}}$at $\sqrt{s}$=7 TeV,maxp_{T} >20 GeV/c, same phase space	7.0
DistributionΣp$_{\text{T}}$at $\sqrt{s}$=7 TeV,maxp_{T} >20 GeV/c, same phase space	7.0
Distributionp$_{\text{T}}$at $\sqrt{s}$=7 TeV,maxp_{T} >20 GeV/c, same phase space	7.0
DistributionN$_{\text{ch}}$at $\sqrt{s}$=7 TeV,maxp_{T} >3 GeV/c, same phase space	7.0

Table 9: Used observables from CMS underlying event data in "CMS selected underlying event " and combined "CMS selected UE and charged hadron spectrum" tunes.

References

[1] Georges Aad et al. Measurement of underlying event characteristics using charged particles in pp collisions at $\sqrt{s} = 900 GeV$ and 7 TeV with the ATLAS detector. *Phys. Rev.*, D83:112001, 2011.

[2] Georges Aad et al. Study of Jet Shapes in Inclusive Jet Production in pp Collisions at sqrt(s) = 7 TeV using the ATLAS Detector. *Phys. Rev.*, D83:052003, 2011.

[3] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J.*, C58:639–707, 2008.

[4] Manuel Bahr, Stefan Gieseke, and Michael H. Seymour. Simulation of multiple partonic interactions in Herwig++. *JHEP*, 07:076, 2008.

[5] Andy Buckley, Hendrik Hoeth, Heiko Lacker, Holger Schulz, and Jan Eike von Seggern. Systematic event generator tuning for the LHC. *Eur. Phys. J.*, C65:331–357, 2010.

[6] Serguei Chatrchyan et al. Measurement of the Underlying Event Activity at the LHC with sqrt(s)= 7 TeV and Comparison with sqrt(s) = 0.9 TeV. *Accepted by J. High Energy Phys.*, 2011.

[7] Professor Collaboration. *Professor toolkit documentation.* Web page: http://projects.hepforge.org/professor/docs.sphinx/index.html.

[8] Vardan Khachatryan et al. Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at sqrt(s) = 7 TeV. *Phys. Rev. Lett.*, 105:022002, 2010.

[9] Vardan Khachatryan et al. First Measurement of Hadronic Event Shapes in pp Collisions at sqrt(s)=7 TeV. *Phys. Lett.*, B699:48–67, 2011.