

The Compact Muon Solenoid Experiment **CMS Note** Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland



02 Feb 2006

# Sensitivity of the Muon Isolation Cut Efficiency to the Underlying Event Uncertainties

S. Abdullin<sup>1</sup>), D. Acosta<sup>2</sup>), P. Bartalini<sup>2</sup>), R. Cavanaugh<sup>2</sup>), A. Drozdetskiy<sup>2</sup>), A. Korytov<sup>2</sup>), G. Mitselmakher<sup>2</sup>), Yu. Pakhotin<sup>2</sup>), B. Scurlock<sup>2</sup>), A. Sherstnev<sup>3</sup>)

CMS collaboration

#### Abstract

In the context of the H  $\rightarrow$  ZZ  $\rightarrow 4\mu$  analysis, uncertainties in predicting the muon isolation cut efficiency are studied by varying the PYTHIA parameters responsible for simulation of the underlying event. The processes considered are H  $\rightarrow$  ZZ  $\rightarrow 4\mu$ , ZZ  $\rightarrow 4\mu$ , and  $t\bar{t} \rightarrow 4\mu + X$ . We also show that an inclusive Z data sample will allow for direct experimental measurement of the 4-muon isolation cut efficiencies associated with H  $\rightarrow$  ZZ  $\rightarrow 4\mu$  and ZZ  $\rightarrow 4\mu$  events with a systematic uncertainty of less than 2%.

<sup>&</sup>lt;sup>1)</sup> Fermi National Laboratory, Chicago, IL, USA

<sup>&</sup>lt;sup>2)</sup> University of Florida, Gainesville, FL, USA

<sup>&</sup>lt;sup>3)</sup> Moscow State University, Moscow, Russia

# **1** Introduction

In future searches for the Higgs boson at the LHC via its 4-muon decay channel,  $H \rightarrow ZZ \rightarrow 4\mu$ , the muon isolation cut, i.e., the requirement of a low energy flow in the muon vicinity, plays a key role in suppressing many otherwise dominating backgrounds where all or some muons originate from hadronic decays (tt and Zbb are the most important backgrounds in this category). Assuming that these reducible backgrounds indeed can be brought well below the level of the irreducible ZZ-background, the question of evaluating the level of ZZ background becomes of paramount importance. Having reduced the tt and Zbb backgrounds to a negligible level, we also have suppressed the ZZ background and signal. Therefore, one must worry about the efficiency of the muon isolation cut with respect to the ZZ background and Higgs boson signal and, even more, about the sensitivity of this efficiency to large theoretical uncertainties associated with a poor understanding of the underlying event (UE) physics. The UE is defined as [1] all the remnant activity from the same proton-proton interaction. The goal of the studies presented in this note was not to optimize the muon isolation cut in order to maximize the signal-over-background significance, but rather to answer the following two questions:

- How well can we predict the isolation cut efficiency using the current Monte Carlo generators?
- Can we measure the isolation cut efficiency using the experimental data themselves and, if yes, would the associated experimental systematic uncertainties be smaller than the Monte-Carlo-based theoretical uncertainties?

CMS uses both a tracker-based isolation cut and a calorimeter-based isolation cut. In these generator-level studies, we looked only at the tracker-based isolation cut; we believe the relative sensitivity of the calorimeter-based muon isolation cut to the UE uncertainties must be correlated with that for the tracker-based muon isolation. Calorimeter-based muon isolation by itself depends more than the tracker-based one on an accurate detector realization and that was the reason we avoided its implementation for this generator-level analysis.

Although this study is done as a part of the  $H \rightarrow ZZ \rightarrow 4\mu$  analysis, the results and discussion presented here are of general interest to all analyses making use of lepton isolation cuts.

The analysis presented in this note is done in accordance with the official CMS guidelines described in [1] concerning UE for a particular Monte Carlo generator with a particular set of model parameters. Only first-order effects influencing the UE in this model are considered.

# 2 Event Generation Parameters for PYTHIA

Higgs boson,  $t\bar{t}$  and Z-inclusive data samples were generated with PYTHIA 6.223 [2]. The ZZ data sample was generated at the matrix-element level with CompHEP [3], and then PYTHIA was used to complete the event simulation (parton shower development, UE, hadronization, and particle decays). The PYTHIA parameters that drive the UE simulation were consistently chosen to match those selected for the Data Challenge 2005 (DC05) CMS official production (see Table 1). Detailed discussion of the associated phenomenology and the corresponding references can be found elsewhere [1].

parameter	CDF	ATLAS	CMS (DC04)	CMS (DC05)	comment	
PARP(82)	2	1.8	1.9	2.9	regularization scale of PT spectrum for MI	
PARP(84)	0.4	0.5	0.4	0.4	parameter of matter distribution inside hadrons	
PARP(85)	0.9	0.33	0.33	0.33	probability in MI for two gluons with color connections	
PARP(86)	0.95	0.66	0.66	0.66	probability in MI for two gluons (as a closed loop)	
PARP(89)	1800	1000	1000	14000	reference energy scale	
PARP(90)	0.25	0.16	0.16	0.16	power of the energy-rescaling term	
pt <sub>cut-off</sub>	3.34	2.75	2.90	2.90	fi nal pt <sub>cut-off</sub>	

Table 1: Parameters in PYTHIA for multi-parton interactions (MI) and UE for CDF, ATLAS and CMS.

The most critical parameter affecting the UE activity is  $pt_{cut-off}$ , the lowest transverse momentum (PT) allowed for multi-parton interactions. The smaller the value of  $pt_{cut-off}$  the larger the number of tracks associated with the underlying event. The  $pt_{cut-off}$  value and its evolution with the center of mass energy of proton-proton collisions are defined via the following formula:

 $pt_{cut-off} = PARP(82) * (14000 \text{GeV}/PARP(89))^{PARP(90)}.$ 

The three parameters, PARP(82,89,90), have meaning only in this combination. The parameters PARP(89) and PARP(90) are fixed at 14,000 and 0.16, correspondingly. We decided to vary  $pt_{cut-off}$  by  $\pm 3\sigma$ , or  $\pm 0.5$  GeV, which seems to be a sensible estimation of the theoretical uncertainties arising from UE modeling [4]. Note that  $pt_{cut-off} = 3.34$  GeV, as extracted from CDF's Tune A of PYTHIA MI parameters, differs from the default values used by ATLAS (2.75 GeV) and CMS (2.9 GeV) by ~ 0.5 GeV because it was done using a different PYTHIA parameter tuning model. It is listed in Table 1 for completeness only.

## **3** UE Analysis and Results

#### 3.1 Monte Carlo sample production

Processes used in these studies were:  $t\bar{t}$  (PYTHIA parameter MSEL = 6); Higgs boson signal (m<sub>H</sub> = 150 GeV, PYTHIA parameters MSEL = 0, MSUB(102,123,124) = 1 with H allowed to decay to  $Z/\gamma *$  only,  $Z/\gamma *$  allowed to decay to  $e/\mu/\tau$  pair only, and  $\tau$  allowed to decay to  $e/\mu$  only); ZZ (PYTHIA parameters MSEL = 0, MSUB(1) = 22 with  $Z/\gamma *$  allowed to decay to  $e/\mu/\tau$  pair only, and  $\tau$  allowed to decay to  $e/\mu$  only); Z-inclusive (PYTHIA parameters MSEL = 0, MSUB(1) = 1 with Z allowed to decay to muon pair only). For Higgs boson signal, we used PHOTOS as a generator of bremsstrahlung photons.

Generator-level cuts:

- $t\bar{t}$ : at least four muons with PT > 7 GeV and  $|\eta| < 2.4$ ;
- Higgs boson signal: at least four muons with PT > 7 GeV and  $|\eta| < 2.4$ ;  $5 < M_{inv}(\mu^+\mu^-) < 150 \text{ GeV}$  for 2 intermediate resonances  $(Z/\gamma *)$ ;
- ZZ-sample: same as for signal;
- Z-inclusive: no user defined cuts.

#### 3.2 Events selection

Events-selection cuts were further imposed on the produced Monte Carlo samples. These cuts were chosen to mimic those optimized for the future data analysis. There are two distinct sets of such cuts.

First, only "good muons" were selected. A muon was considered to be "good" if it had PT > 7 GeV in the barrel region ( $|\eta| < 1.1$ ) or P > 9 GeV in the endcaps ( $1.1 < |\eta| < 2.4$ ). This ensures that the muon reconstruction efficiencies are at their plateau, which helps minimize systematic uncertainties on the muon reconstruction efficiency.

Then, event-selection cuts similar to the full analysis cuts were applied. They are:

- At least 2 pairs of opposite sign muons with the invariant masses of all  $\mu^+\mu^-$  pair permutations being greater than 12 GeV (this cut suppresses heavy-quark resonances).
- The PT of all four selected muons must be greater than 10 GeV (signal-over-background optimization).
- Invariant mass of the four muons must be greater than 110 GeV and less than 700 GeV (a Higgs boson with M < 114.4 GeV is excluded at LEP, a Higgs boson with mass over 700 GeV is strongly disfavored by theory and, also, would have too low a production cross section).
- ISOL =  $\sum PT_i$  (PT with respect to the beam direction) should be less or equal to 0, 0, 1, 2 GeV for the four muons when the muons are sorted by the ISOL parameter. The sum runs over only charged particle tracks with PT greater then 0.8 GeV and inside a cone of radius  $R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2} = 0.3$  in the azimuth-pseudorapidity space. A PT threshold of 0.8 GeV roughly corresponds to the PT for which tracks start looping inside the CMS Tracker. Muon tracks are not included in the calculation of the ISOL parameter.

#### 3.3 Track multiplicity

Variations of the pt<sub>cut-off</sub> parameter mainly affect the UE track multiplicity (Fig. 1) and leave the transverse momenta of these tracks basically unchanged (Fig. 2).





Figure 1: Inclusive  $dN/d|\eta|$ -distribution for charged particle tracks (other than muons) for the Higgs boson Monte Carlo sample. Normalization is to the average number of muon tracks per event. Clearly seen is the variation in multiplicity of tracks for three different cases: the black upper dashed line (higher multiplicity) is for the downward  $-3\sigma$  variation on the pt<sub>cut-off</sub> value, the blue middle line is for the default case, and the red lower dashed line is for the upward  $+3\sigma$  variation.

Figure 2: Inclusive  $dN/dP_T$ -distribution for charged particle tracks (other than muons) for the Higgs boson Monte Carlo sample. Normalization is to the average number of muon tracks per event. Three different cases shown: the black upper dashed line (higher multiplicity) is for the downward  $-3\sigma$  variation on the  $pt_{cut-off}$  value, the blue middle line is for the default case, and the red lower dashed line is for the upward  $+3\sigma$  variation. The shape of the distributions remains unchanged.

Figures 3, 5 and 7 show how the density of tracks around the muons changes with the MI  $pt_{cut-off}$  value for the three different sources of muons. Enlarged versions of the same three plots are given in Figs. 4, 6 and 8. One can see from the  $t\bar{t}$  background plots that the default cone size R = 0.3 is quite reasonable.

The distributions of the ISOL parameter values (and their variation with MI  $pt_{cut-off}$ ) are shown in Figs. 9, 10, 11 and 12.

#### **3.4** Comparison of the generator-level and the full simulation results

To see whether the conclusions derived from these generator-level studies can be applied to the full simulation analysis, the following cross-checks were performed.

Figure 13 shows the distribution of the tracker-based muon isolation cut parameter for Higgs boson sample muons from the full simulation analysis and from these generator-level studies. One can see that the distributions are reasonably close to each other, despite the inevitable differences in the two analyses (e.g., inefficiency of track reconstruction in the full simulation, pile-up events, etc.)

One can also compare the final event selection efficiencies, for which we measure  $(23 \pm 2)\%$  in the full simulation analysis and  $(21 \pm 2)\%$  in these generator-level studies. (These efficiencies include all cuts described in subsection 3.2.)

These comparisons show that the tracker-based isolation studies carried out at the generator level are similar to the full-detector simulation results (this comparison is meant to show that the generator level ISOL parameters and corresponding efficiencies are consistent with the full simulation values). Therefore, all relative effects such as differences in the efficiencies (different UE event models, ZZ vs Z samples) as obtained in our generator-level studies can be applied to the full-detector simulation analyses.



Figure 3: dN/dR distribution for UE tracks around "good" muons (Higgs boson events). The distribution is normalized per "good" muon. The blue (middle) solid line is for the default MI  $pt_{cut-off}$ , the black (upper) dashed line is for downward  $-3\sigma$  variation of  $pt_{cut-off}$  value, the red (lower) dashed line is for upward  $+3\sigma$  variation.



Figure 5: Similar to Figure 3 for muons from  $t\bar{t}$ , hadronic decay.



Figure 7: Similar to Figure 3 for muons from  $t\bar{t}$ , W- and  $W - \tau$ -decay.



Figure 4: Enlarged version of Fig. 3. The cone size R = 0.3 used for calculations of the ISOL parameter (see text) is shown by the arrow.



Figure 6: Enlarged version of Figure 5.



Figure 8: Enlarged version of Figure 7.



Figure 9: Distribution of the tracker isolation parameter ISOL (Higgs boson events). Normalization is per "good" muon. The blue (middle) solid line is for the default MI  $pt_{cut-off}$ , the black dashed line is for downward  $-3\sigma$  variation of  $pt_{cut-off}$  value, the red dashed line is for upward  $+3\sigma$  variation.



Figure 11: Similar to Figure 9 for muons from  $t \overline{t}$  and hadronic decay.



Figure 10: Similar to Figure 9 for muons from  $t\bar{t}$ , W- and  $W - \tau$ -decay.



Figure 12: Enlarged version of Fig. 11.



Figure 13: Comparison of the tracker isolation parameter distributions for the full-detector simulation (blue dashed line) and for this generator-level study (red solid line). Default UE simulation settings were used.

#### 3.5 Tracker-based muon isolation cut efficiency

#### 3.5.1 Prediction uncertainties

Figures 14, 15 and 16 show the muon isolation cut efficiency averaged over all "good" muons (see section 3.2) for the tt sample and the Higgs boson. For the tt background, we show two plots: one for muons originating from  $W \rightarrow \mu\nu$  and  $W \rightarrow \tau\nu \rightarrow \mu\nu\nu\nu$  decays and the other for muons originating from hadronic decays (typically, the former would tend to be isolated and the latter non-isolated). The average isolation efficiency per "good" muon is calculated as the ratio of the number of "good" muons with the isolation parameter ISOL below a particular threshold to the total number of "good" muons. Figure 17 shows the isolation cut efficiency for the least isolated muon out of four (Higgs boson sample). We use a cut at ISOL=2 GeV for such muons. One can see that this cut alone will have ~ 80% efficiency, with  $\pm 5\%$  uncertainty due to the UE variations.





Figure 14: Muon isolation cut efficiency averaged over selected muons whose parents are W bosons ( $t\bar{t}$  events). The blue middle line is for the default MI pt<sub>cut-off</sub>, the black upper line is for downward  $-3\sigma$  variation of pt<sub>cut-off</sub> value, the red lower line is for upward  $+3\sigma$  variation.





Figure 16: Similar to Fig. 14 for Higgs boson events.



Figure 17: Muon isolation cut efficiency for the least isolated muon from 4 selected ones in Higgs boson events.

Figure 18 compares the muon isolation cut efficiency curves for the main irreducible ZZ background and for the Higgs boson events. Clearly, these efficiencies are very similar.

#### 3.5.2 Sensitivity to kinematical cuts

Figure 19 demonstrates another very important feature of the tracker-based muon isolation cut: its efficiency is not very sensitive to the kinematical analysis cuts. The figure has two sets of efficiency curves: one is obtained for "good" muons and another for "good" muons passing further event selection cuts as described in section 3.2. One can hardly see any difference. Therefore, the conclusions of this analysis will not depend on the choice of the final event selection cuts.





Figure 18: Muon isolation cut efficiency averaged over 4 selected muons for Higgs boson events (solid lines, Fig. 16) and ZZ background (dashed lines). The blue middle line is for the default MI  $pt_{cut-off}$ , the black upper line is for downward  $-3\sigma$  variation of  $pt_{cut-off}$  value, the red lower line is for upward  $+3\sigma$  variation.

Figure 19: Muon isolation cut efficiency averaged over 4 selected muons for Higgs boson events. Solid lines are for good muons from events after analysis cuts (same as Fig. 16) and dashed lines are for good muons from events before analysis cuts. There is no difference between the two sets of points at the level of the statistical precision. The blue middle line is for the default MI pt<sub>cut-off</sub>, the black upper line is for downward  $-3\sigma$  variation of pt<sub>cut-off</sub> value, the red lower line is for upward  $+3\sigma$  variation.

#### 3.5.3 Evaluation of the muon isolation cut efficiency from data using random-cone directions

Figures 20 and 21 show the isolation cut efficiency and the isolation cut parameter distribution as calculated for random directions uniformly distributed in  $\eta - \phi$  space ( $|\eta| < 2.4$ ). The algorithm of the ISOL parameter calculation is the same as for "real" MC muons, except that now the ISOL parameter takes into account the sum of PT for tracks around random directions in the acceptance region. The Higgs boson Monte Carlo sample was used to make these plots. We see that the graphs obtained for the random cone (solid lines) and for "real" muons (dashed line; identical to Figures 16 and 19) look very similar. In fact, they agree within statistical uncertainties.

This observation motivated us to investigate whether we can measure the isolation cut efficiency by using some distinct reference data sample and applying the random-cone technique. The reference data sample must have a large cross section (to provide good statistics), be relatively clean from backgrounds, and have a similar underlying structure to ZZ events. Inclusive  $Z \rightarrow \mu\mu$  seems to be just what we need. The cross section according to PYTHIA output is ~ 1.6 nb, and  $Z \rightarrow \mu\mu$  has a very clean signature.

Figure 22 shows the isolation cut efficiencies computed for random-cone directions from a Z-inclusive Monte Carlo generation sample. One can see that the isolation cut efficiencies for muons in the ZZ sample are very well mimicked by the efficiencies calculated for random cones in the Z-inclusive sample. The variations in the UE  $pt_{cut-off}$  have nearly identical effects on both data samples.

For comparison, we also show the random-cone-based efficiencies using the  $t\bar{t}$  sample (Fig. 23). Differences of up to 5% with the ZZ efficiencies are seen. Clearly, the  $t\bar{t}$ -sample is not good at mimicking the UE of the ZZ sample; calibration with the  $t\bar{t}$ -sample would give as large an error as the UE theoretical uncertainty.





Figure 20: Muon isolation cut efficiency for randomcone directions (solid lines) and for muons (dashed lines) for signal events. The blue middle lines are for the default MI pt<sub>cut-off</sub>, the black upper lines are for downward  $-3\sigma$  variation of pt<sub>cut-off</sub> value, the red lower lines are for upward  $+3\sigma$  variation.

Figure 21: Distribution of the tracker-based muon isolation parameter ISOL for random-cone directions (blue solid line) and for muons (red dashed line). Higgs boson events with the default UE simulation parameters are used.



Figure 22: Muon isolation cut efficiency for randomcone directions for Z-inclusive (dashed lines) and for ZZ (solid lines) events. The blue middle lines are for the default MI  $pt_{cut-off}$ , the black upper lines are for downward  $-3\sigma$  variation of  $pt_{cut-off}$  value, the red lower lines are for upward  $+3\sigma$  variation.

Figure 23: Similar to Fig. 22 but using random-cone directions in  $t\bar{t}$  background events.

#### **3.5.4** $4\mu$ isolation cut efficiency per event

Efficiencies per event for the different samples with the three different  $pt_{cut-off}$  values are listed in Table 2. The values for the Signal, ZZ-background, and Z-inclusive samples using the random-cone technique are in agreement with each other for all three UE scenarios. The range of efficiencies for the ZZ-background spans from ~ 0.72 to ~ 0.84. This range of  $\pm 6\%$  absolute of the central value can be associated with the variations in the 4-muon isolation cut efficiency arising from theoretical uncertainties on the UE parameters in PYTHIA.

Table 2: Efficiency per event using different events samples: Higgs boson signal with  $m_H = 150$  GeV, ZZ background, Z-inclusive (4 RND muons),  $t\bar{t}$  background. "4 RND muons" means that for a particular process in each event 4 random cone directions were used to calculate the ISOL parameter and the corresponding values were treated as ones for "real" muons.

process/case	effi ciency (default)	efficiency $(-3\sigma)$	effi ciency $(+3\sigma)$
signal, $m_H = 150 \text{ GeV}$	$0.775 \pm 0.004$	$0.707\pm0.005$	$0.812 \pm 0.004$
ZZ background	$0.780 \pm 0.004$	$0.721 \pm 0.005$	$0.838 \pm 0.004$
4 RND muons, Z-inclusive events	$0.762\pm0.007$	$0.706\pm0.007$	$0.821 \pm 0.006$
$t\overline{t}$ background	$0.016\pm 0.001$	$0.013 \pm 0.001$	$0.015\pm0.001$

On the other hand, it appears possible to use the Z-inclusive sample to gauge the UE activity and evaluate the 4-muon isolation cut efficiency experimentally. There might be a small systematic shift of the order of ~ 2% in efficiencies between the ZZ and Z-inclusive samples. This shift for the introduced random-cone calibration from data technique, which makes the result to a large degree independent of a particular UE scenario. For the three different UE simulations of used in these studies, we obtain the following offsets:  $0.018 \pm 0.008$ ,  $0.015 \pm 0.009$ ,  $0.017 \pm 0.007$ . Much larger Monte Carlo samples would be needed to measure the shift more accurately. Meanwhile, conservatively, one may just ignore this correction and assign a 2% systematic uncertainty on the Z-sample-based estimate of the 4-muon isolation cut efficiency for ZZ-background and Higgs boson signal events. This uncertainty is already much smaller in comparison to the other systematics such as the experimental uncertainties on the muon reconstruction efficiency, theoretical uncertainties associated with the choice of PDF's and QCD scale, etc.

The efficiency for accepting  $t\bar{t}$ -events is of the order of  $0.015 \pm 0.001$ . Its sensitivity to the UE could not be studied due to lack of statistics, but it is not expected to be too large as it is dominated by the jet activity. If the reducible  $t\bar{t}$ - and Zbb-backgrounds could not be suppressed well below the ZZ-background, one would need to study their sensitivity to the UE physics, as well as to the jet fragmentation modeling.

### 4 Summary

The isolation cut efficiency per muon due to uncertainties in the considered UE models vary as much as  $\pm 5\%$  (the efficiency itself and its uncertainty strongly depend on how tight the ISOL cut is). The 4-muon isolation cut efficiency per event for ZZ  $\rightarrow 4\mu$  background is measured to be  $(78 \pm 6)\%$ .

To decrease these large uncertainties to a negligible level with respect to other systematic uncertainties, one can calibrate the isolation cut efficiency from data using Z-inclusive events  $(Z \rightarrow 2\mu)$  and the random-cone technique. We show that this indeed significantly decreases the uncertainties associated with a poor understanding of the UE physics. There might be ~ 2% systematic shift in the 4-muon isolation cut efficiencies obtained this way. In principle, one could correct for this shift, but it does not appear to be necessary as this uncertainty is already smaller than other systematic and statistical errors.

The results and techniques described in this note may be of interest to other analyses relying on lepton isolation cuts.

## **5** Acknowledgments

We would like to thank M. Aldaya, P. Arce, J. Caballero, B. Cruz, G. Dissertori, T. Ferguson, U. Gasparini, P. Garcia, J. Hernandez, I. Josa, M. Konecki, P. Moisenz, E.R. Morales, N. Neumeister, A. Nikitenko, F. Palla and I. Vorobiev for their active participation in the analysis discussions and comments on the text of this note.

## References

- [1] P. Bartalini et al., "Guidelines for the evaluation of the theory error systematics at the LHC", CMS Note 2005/013
- [2] T. Sjostrand *et al.*: "The Complete PYTHIA 6.206 Manual", hep-ph/0108264, http://www.thep.lu.se/ torb-jorn/Pythia.html
- [3] CompHEP collaboration: "CompHEP a package for evaluation of Feynman diagrams and integration over multi-particle phase space. User's manual for version 33", hep-ph/9908288, http://theory.sinp.msu.ru/parser/parser.php?file=/phpcms/comphep
- [4] P. Nason *et al.*, Proceedings of the Workshop on Standard Model Physics (and more) at the LHC, hepph/0003142. CERN 2000/004. p 293.