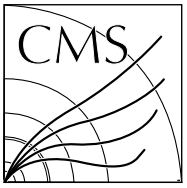
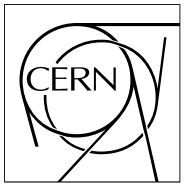


The Compact Muon Solenoid Experiment

CMS Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland



10 May 2006

Measuring Muon Reconstruction Efficiency from Data

D. Acosta[2)], P. Bartalini[2)], A. Drozdetskiy[1)2)], A. Korytov[2)], G. Mitselmakher[2)]

CMS collaboration

Abstract

We suggest a method of measuring the global muon reconstruction efficiency ϵ directly from data, which largely alleviates uncertainties associated with our ability to monitor and reproduce in Monte Carlo simulation all details of the underlying detector performance. With the data corresponding to an integrated luminosity $L = 10$ fb^{-1}, and using this method, the precision of measuring ϵ for muons in the P_T range of 10-100 GeV is expected to be better than 1%.

version: 10 May 2006

[1)] Corresponding author, e-mail: Alexey.Drozdetskiy@cern.ch

[2)] University of Florida, Gainesville, FL, USA

1 Introduction

Understanding the muon reconstruction efficiency (MRE) with a sub-percent precision at the earlier stages of the CMS operation may be a challenging, if not formidable, task if one attempts to evaluate it by building into the detector simulation all relevant details of detector performance and associated uncertainties. Such simulation would have to include realistic modeling of: geometry, detector edge effects, dead or noisy channels/boards, corrupted data, detectors with turned off or reduced high voltage, luminosity and beam halo, trigger tables. One would need to monitor and incorporate in simulation all, often intermittent, changes in the detector performance.

On the other hand, we can devise a strategy to measure MRE from data in such a way that it would take into account the real detector performance automatically. By choosing a reference process with a large production cross-section (e.g., inclusive Z production), we would be able to reliably measure MRE at the early stage of operation. In this note, we propose such a method and explore its potential in the context of $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 4\mu$ analysis [3], where, for obvious reasons, the question of muon efficiency is of very high importance. Reference integrated luminosity used in the note corresponds to the amount of data with which a discovery of the Standard Model Higgs boson may become possible.

Shown below is a feasibility study, which defines a general strategy. The results to be obtained by using this methodology will be applicable to all analyses involving muons of moderate transverse momenta P_T in the range of 10-100 GeV.

2 Strategy for Muon Efficiency Reconstruction from Data

The global muon reconstruction, GMR (the standard CMS algorithm for combining information from the Tracker and Muon sub-systems), is based on *independent* reconstruction of a muon in a standalone Muon system and matching it with a track of similar kinematic parameters in the Tracker. Details on all reconstruction algorithms used in the analysis including GMR can be found in the CMS Physics Technical Design Report, volume 1 [1].

To measure MRE we use a sample of events with at least one muon of $P_T > 20$ GeV. This value is higher than the High Level Trigger threshold for single muons of 19 GeV [2]. Throughout the paper we call these high P_T muons "HLT muons". This large data sample consists mostly of $W \rightarrow \mu\nu$ decays with about 10% of $Z \rightarrow \mu\mu$ [2]. The production rate is about 10 Hz at $L = 2 \times 10^{33}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

Therefore, for the purposes of these studies, we used the inclusive W and Z Monte Carlo (MC) samples from the official CMS production. The inclusive W and Z samples were produced in several $\hat{P_T}$ bins. For this feasibility study, we selected a sample generated at $\hat{P_T}$ interval $50 - 85$ GeV. Details of the corresponding cross-sections (W/Z were forced to leptonic decay modes) and numbers of expected and simulated events are given in Table 1. Including all other $\hat{P_T}$ bins would further increase the number of events: full cross section for W/Z-inclusive production is about $1.5 \cdot 10^5/3.9 \cdot 10^4$ pb, which means we should expect of the order of 10^4 times larger statistics than number of generated events. Throughout the note we refer to $10\ \mathrm{fb}^{-1}$ of integrated luminosity for the expected number of events, although the statistical errors are based on the number of simulated events we used shown in Table 1.

Table 1: The W/Z samples used in the study ($\hat{P_T}$ interval $50 - 85$ GeV). Full inclusive cross sections are about $1.5 \cdot 10^5/3.9 \cdot 10^4$ pb.

	inclusive W	inclusive Z
Cross-section $pp \rightarrow W/Z + X$	$1.014 \cdot 10^4$ pb	$1.456 \cdot 10^3$ pb
Number of events for $L = 10\ \mathrm{fb}^{-1}$	$1.014 \cdot 10^8$	$1.456 \cdot 10^7$
Events used in these studies	115,995	93,996
Events in these studies with a muon $P_T > 20$ GeV	18,141	20,247

Starting from this data sample with at least one high-P_T muon (HLT muon), we can reconstruct and histogram invariant masses of the clean trigger muon paired one by one with all tracks found in the Tracker (Fig. 1), all Standalone muons (Fig. 2), and Globally Reconstructed Muons (Fig. 3). The distributions are expected to and do show a distinct peak at $M_{inv} = M_Z$; note the log scale in the Figures. By extrapolating the spectrum from sidebands, we can evaluate the number of $Z \rightarrow \mu\mu$ events: $N_{Z(TRK)}$, $N_{Z(SAM)}$, $N_{Z(GMR)}$ in our data sample.

Since the GMR relies on a match between the Stand Alone Muon system and the Tracker, the efficiencies for the three categories are related:

$$\epsilon_{GMR} = \epsilon_{TRK} \cdot \epsilon_{SAM}, \quad (1)$$

where ϵ_{GMR} is a Global Muon Reconstruction efficiency, ϵ_{TRK} is a track finding efficiency in the Tracker, and ϵ_{SAM} is the efficiency of finding a muon in the Stand Alone Muon system. No correlations between Tracker and Muon systems are included as they are two independent systems.

Then, we can write:

$$N_{Z(TRK)} = \epsilon_{HLT} \cdot \epsilon_{TRK} \cdot N_Z, \quad (2)$$
$$N_{Z(SAM)} = \epsilon_{HLT} \cdot \epsilon_{SAM} \cdot N_Z, \quad (3)$$
$$N_{Z(GMR)} = \epsilon_{HLT} \cdot \epsilon_{GMR} \cdot N_Z \quad (4)$$

where ϵ_{HLT} is the common, *unknown* efficiency for detecting high P_T muons and N_Z is the total *unknown* number of $Z \to \mu\mu$ events in our sample. These three equations in combination with Eq.1 can be easily solved for the muon reconstruction efficiency:

$$\epsilon_{GMR} = (N_{Z(GMR)})^2/(N_{Z(TRK)}N_{Z(SAM)}), \quad (5)$$
$$\epsilon_{TRK} = N_{Z(GMR)}/N_{Z(SAM)}, \quad (6)$$
$$\epsilon_{SAM} = N_{Z(GMR)}/N_{Z(TRK)}. \quad (7)$$

Note that there are two types of efficiencies which enter Eq's 2-4. The first efficiency ϵ_{HLT} is for the high-P_T muon preselected by the trigger and our analysis cuts:

- $P_T > 20$ GeV, $|\eta| < 2.4$
- Muon isolation: $ISOL_{TRK} < 0.5$ GeV; $ISOL_{CAL} < 2$ GeV.

Here Tracker-based isolation is defined as $\mathrm{ISOL_{TRK}} = \sum \mathrm{P_T^i}$, where the sum runs over charged particle tracks inside a cone of radius $\mathrm{R} = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.3$ in the azimuth-pseudorapidity space around the muon axis (P_T of tracks is measured with respect to the beam direction). Calorimeter-based isolation is defined via summing over calorimeter tower E_T's in the same cone.

The other efficiencies refer to the offline muon reconstruction. It is those efficiencies that we attempt to measure. We impose the following cuts on the probe muons/tracks (those muons/tracks for which we are calculating efficiencies): $\mathrm{P_T} > 7$ GeV in the barrel region ($|\eta| < 1.1$) or $\mathrm{P} > 13$ GeV in the endcaps ($1.1 < |\eta| < 2.4$). Only such muons are used in the $\mathrm{H} \to \mathrm{ZZ}^{(*)} \to 4\mu$ analysis to ensure that the muon reconstruction efficiencies are close to their plateau efficiency level (see Figs. 4 and 5), which helps minimize systematic uncertainties.

Also, P_T and P ranges of muons in the central and forward directions in the inclusive Z sample (probe muons only) are very similar to those in the $ZZ \to 4\mu$ process, the dominant background in the 4-muon Higgs boson decay search (see Figs. 6 and 7), which makes the efficiencies reconstructed in Z-inclusive data samples directly applicable to the ZZ background. Indeed, even average efficiencies for all muons are nearly identical: 0.972 ± 0.001 for Z-inclusive muons and 0.978 ± 0.001 for the ZZ sample (Figs. 6 and 7). Here additional restrictions on two muon invariant mass applied to have similar $\mathrm{P_T}$-range as in our $\mathrm{H} \to \mathrm{ZZ}^{(*)} \to 4\mu$ analysis, see Ref. [3].

3 Results

Distributions of invariant masses $M_{inv}(\mu_{HLT} + track)$, $M_{inv}(\mu_{HLT} + \mu_{standalone})$, $M_{inv}(\mu_{HLT} + \mu_{GMR})$ are shown in Figs. 1, 2 and 3. To calculate $N_{Z(TRK)}$, $N_{Z(SAM)}$, $N_{Z(GMR)}$, the distributions are fit with an exponential background and a Gaussian signal distribution. In order to reduce the dependence on the shape of the signal N_Z was estimated from the difference of N_{TOT}, the total number of entries in the signal region, and N_B as derived from the background part of the fit. The signal region was extending to $\pm 6\sigma$ GeV around the fitted mean value of the Gaussian distribution. Expectations for N_{TOT}, N_B and $N_Z = N_{TOT} - N_B$ are summarized in

Table 2: Number of expected events at $L = 10$ fb^{-1} in the in the M_Z peak signal region, N_{TOT}; background (accidental) pairings, N_B; and reconstructed $Z \to 2\mu$ events, N_Z. The error on N_B is for the actual statistics used in these studies ($L \sim 0.01$ fb^{-1}). See text for details.

	HLT+GMR	HLT+SAM	HLT+TRK
N_{TOT}	2,070,000	2,270,000	4,360,000
N_B	150,000	340,000	2,410,000
$N_Z = N_{TOT} - N_B$	1,920,000	1,930,000	1,950,000
δN_B	$\sim 30,000$	$\sim 30,000$	$\sim 100,000$

Table 2. In future studies systematic uncertainties in the determination of the background should be determined, either by varying the functions used in the fit or by comparing N_B with the number obtained from simulation.

The errors on N_B, δN_B, are defined by the actual statistics used in these studies, approximately corresponding to $L = 0.01$ fb^{-1}. It is these errors that drive the total statistical uncertainty on the measured efficiencies. We calculate the statistical uncertainty on the efficiencies assuming independent measurements of the event counts:

$$\frac{\delta\epsilon_{GMR}}{\epsilon_{GMR}} = 2\frac{\delta N_{B(GMR)}}{N_{Z(GMR)}} \oplus \frac{\delta N_{B(TRK)}}{N_{Z(TRK)}} \oplus \frac{\delta N_{B(SAM)}}{N_{Z(SAM)}}, \tag{8}$$

$$\frac{\delta\epsilon_{TRK}}{\epsilon_{TRK}} = \frac{\delta N_{B(GMR)}}{N_{Z(GMR)}} \oplus \frac{\delta N_{B(SAM)}}{N_{Z(SAM)}}, \tag{9}$$

$$\frac{\delta\epsilon_{SAM}}{\epsilon_{SAM}} = \frac{\delta N_{B(GMR)}}{N_{Z(GMR)}} \oplus \frac{\delta N_{B(TRK)}}{N_{Z(TRK)}} \tag{10}$$

Given the sample we use in these studies, the measured Global Muon Reconstruction efficiency is $\epsilon_{GMR} = 0.983 \pm 0.062$. Table 3 shows that the measured efficiency agrees very well with the efficiency that we can reconstruct from comparing the reconstructed muons with generator MC truth muons. This is the main result of these studies. With more than $10^3 - 10^4$ times larger statistics (increase in statistics expected for $L = 10$ fb^{-1} with respect to the number of generated events we use in this study), one can sub-divide the full statistics in a grid of η, ϕ and P_T regions and still be able to measure efficiency in each of them with a sub-percent precision. Such sub-division will also automatically show all "cracks" in $\eta - \phi - P_T$ space in the detector sensitivity, where the detector sensitive parts in $\eta - \phi - P_T$ space can be defined as sensitive areas for GMR HLT isolated (high quality) muons. Note that the P_T/P and ϕ dependencies of muon efficiency are almost flat and would not need fine binning. Note also that the method automatically includes possible sources of systematic uncertainties like fakes, misidentified muons and all other sorts of "background contamination" as those tracks do not have a distinct feature of peaking around Z^0 invariant mass and hence will be subtracted by the procedure as we perform calibration to side bands.

Table 3: Measured muon efficiencies ($L = 0.01$ fb^{-1}) and those obtained from comparing reconstructed muons with MC truth muons.

	GMR	SAM	TRK
ϵ from Z inclusive data	0.983 ± 0.062	0.986 ± 0.054	0.997 ± 0.022
ϵ in Z inclusive MC sample	0.972 ± 0.001	N/A	N/A
ϵ in ZZ MC sample	0.978 ± 0.001	N/A	N/A

4 Summary

A method of measuring the global muon reconstruction efficiency ϵ directly from data was studied. With the data corresponding to an integrated luminosity $L = 10$ fb^{-1}, the precision of measuring ϵ for muons in the P_T range of $10 - 100$ GeV is expected to be better than 1%, potentially much better. The method largely alleviates uncertainties associated with our ability to monitor and reproduce in Monte Carlo simulation all details of the underlying detector performance.

The CMS simulation/reconstruction software is undergoing a major changeover in preparations for data taking. As the new Monte Carlo samples produced in the new software framework become available, the studies will be

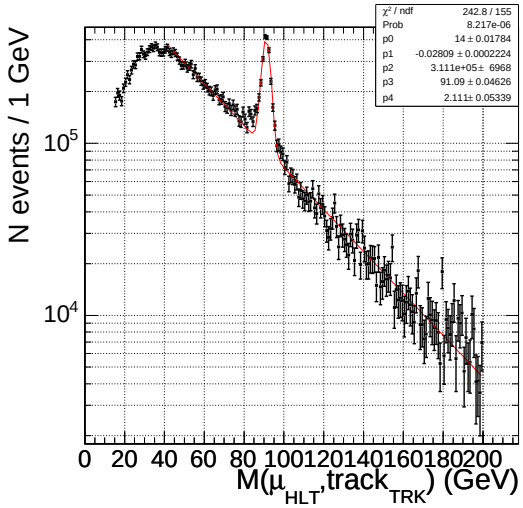


Figure 1: Invariant mass distribution: $M(\mu_{HLT}, track_{TRK})$ for HLT muon and Tracker track.

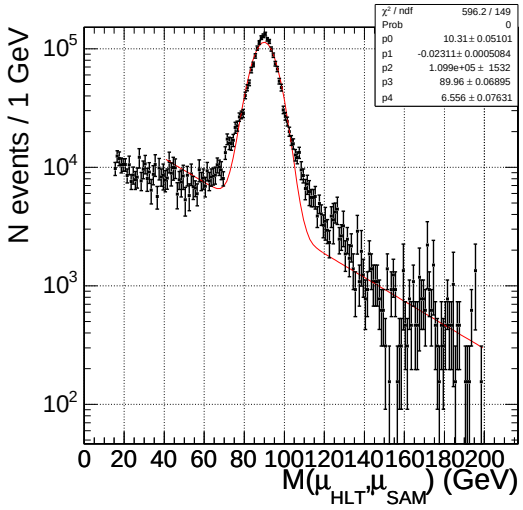


Figure 2: Invariant mass distribution: $M(\mu_{HLT}, \mu_{SAM})$ for HLT muon and Standalone Muon Reconstructor muon.

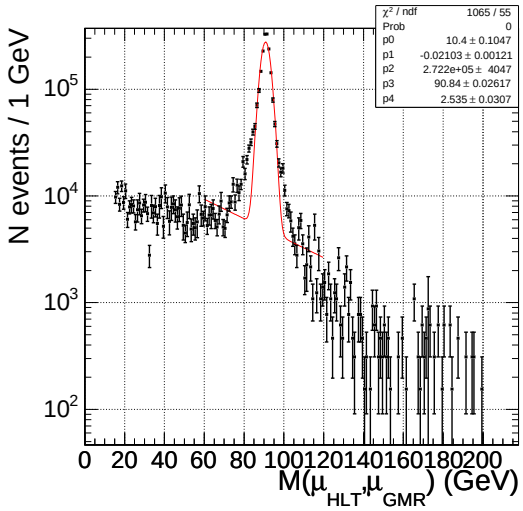


Figure 3: Invariant mass distribution: $M(\mu_{HLT}, \mu_{GMR})$ for HLT muon and another Global Muon Reconstructor muon.

updated with much larger statistics that will also allow us to evaluate possible systematic errors at sub-percent level.

5 Acknowledgments

We would like to thank W. Adam, T. Cox, J. Mumford, N. Neumeister and T. Todorov for useful discussions.

References

[1] CMS Collaboration, *"The CMS Physics Technical Design Report, Volume 1"*, CERN/LHCC, 2006.

[2] CMS Collaboration, *"The Trigger and Data Acquisition project"*, v.2, p.308, 2002 (see Fig. 15-21)

[3] S. Abdullin et al., *"Search for* $H \to ZZ^{(*)} \to 4\mu$ *Using* $M(4\mu)$*-Dependent Cuts"*, CMS Note in preparation, 2006.

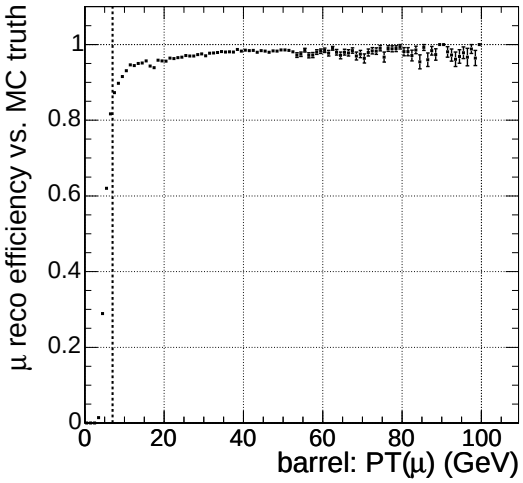


Figure 4: Global Muon Reconstruction efficiency calculated from matching reconstructed and Monte Carlo truth muons in the barrel region for ZZ events.

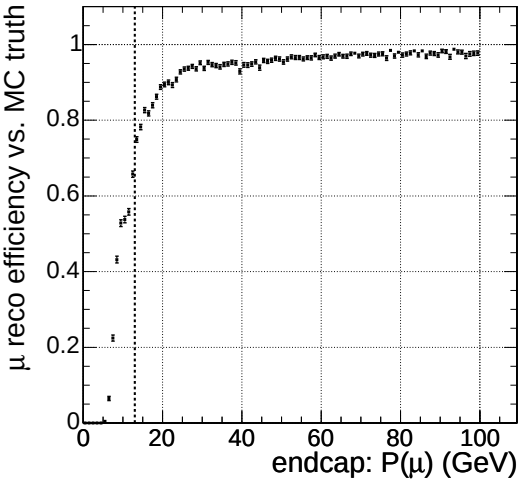


Figure 5: Global Muon Reconstruction efficiency calculated from matching reconstructed and Monte Carlo truth muons in the endcap region for ZZ events.

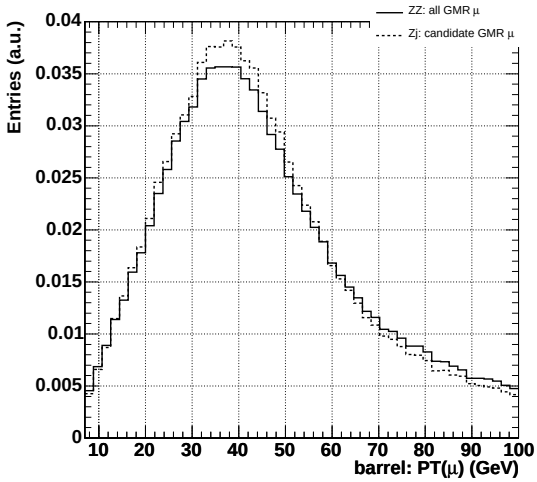


Figure 6: Muon P_T spectrum in the barrel region for ZZ (solid line) and Z (dashed lines) events.

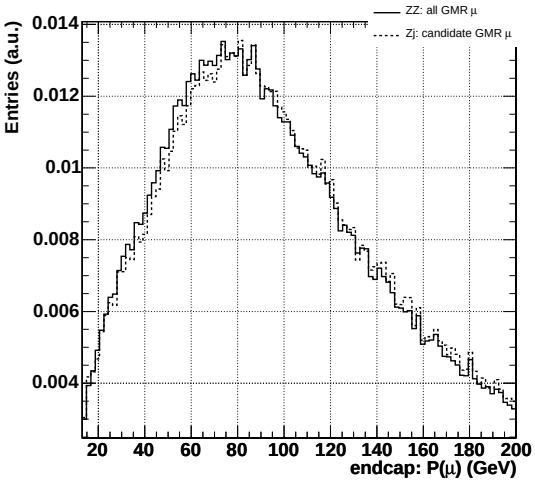


Figure 7: Muon P spectrum in the endcap region for ZZ (solid line) and Z (dashed lines) events.