Available on CMS information server

CMS NOTE 2010/version1

The Compact Muon Solenoid Experiment

CMS Note

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

February 28, 2010

Primary Vertex Resolution Measured with Data Driven Two-Vertex Method at $\sqrt{s} = 0.9/2.36$ TeV

J. Bernardini, F. Fiori, F. Palla

INFN Sezione di Pisa, Pisa, Italy

K. Burkett, Y. Gao

Fermi National Accelerator Laboratory, Batavia, USA

G. Hanson, G.Y. Jeng

University of California, Riverside, Riverside, USA

Abstract

This note introduces a data driven algorithm, referred as two-vertex method, to measure the primary vertex spatial resolution. The resolution is measured as a function of the number of tracks used in fitting the vertex and the average $P_T$of those tracks. The method is validated on Monte Carlo simulated data and applied to the first LHC collision data collected at $\sqrt{s}$ = 0.9 and 2.36 TeV in CMS. The primary vertex resolution in mininum bias data is measured to be close to 50 μm in x and y, and 70 μm in z.

Contents

1 Introduction

The precise estimate of the primary vertex poistion on the event basis is crucial for many physics measurements (lifetime, b-tagging etc) in p-p collision. It is especially important in the early data taking at low energy $\sqrt{s} = 900$ GeV, when both beam transverse width and longitudinal length are large. It also helps in separating the interesting parton level *hard interactions* from the huge background due to long distance diffractive interact ions of protons.

Several vertex fitting algorithms [1] are implemented and studied in CMS reconstruction framework. Among them, the *Adaptive Vertex Filter* algorithm is the most robust one and hence being used in the default offline reconstruction. It starts from the *generalTrack* track collection. The prompt tracks are selected with cuts on the track transverse impact parameter, number of hits, and the χ^2 per degree of freedom. These selected tracks are then clustered in z. Clusters are split when there is a gap over a thresold. The tracks in each cluster are then fit with an Adaptive Vertex Fit, which is based on an iterative re-weighted least-squares fit based on Kalman filter. Each track in the vertex is assigned a weight between 0 and 1 based on their compatibility with the common vertex. The outlier tracks with larger distance to the vertex postion are down weighted significantly, which makes the algorithm robust against the outliers.

The primary vertex resolution is given by the uncertainty reported by the primary vertex algorithm. It is tightly coupled to the impact parameter resolution of the input tracks. It also strongly depends on the number of tracks used in fitting the vertex and the P_T of those tracks. Since each vertex is independent and has its own uncertainty, it is not mathematically possible to define a primary vertex resolution for a specific sample. Therefore, the best we can do is to try to measure an average resolution of an event ensemble.

In this note, we introduce a data-driven method, referred as two-vertex method, to measure the vertex resolution dependence on the number of tracks and averge P_T of those tracks used in the primary vertex. In this method, we measure the primary vertex position of each event with two independent sets of tracks. The vertex resolution is obtained by comparing the relative position of the two vertices in x, y and z. The distribution of the difference in the fitted vertex positions can then be used to extract the resolution and pulls.

2 Data Samples and Event Selections

This analysis is performed with the data collected at the center-of-mass energy of 900 GeV and 2360 GeV.

2.1 Data Samples

We use *December* 19^{th} reprocessing of the collected data and MC datasets:

- /MinimumBias/BeamCommissioning09-BSCNOBEAMHALO-Dec19thSkim_336p3_v1/
- /MinBias/Summer09-STARTUP3X_V8K_900GeV-v1/
- /MinBias/Summer09-STARTUP3X_V8L_2360GeV-v1/

For the collision dataset, the alignment parameters of the silicon tracker were computed with about two million of cosmic ray tracks collected in November 2009 and the nominal values of the alignment parameter errors (APE) have been used in the reconstruction. The simulated events used in this note are minimum-bias events produced with PYTHIA 6.4 [2] event generator at center-of-mass energies of 900 GeV and 2360 GeV and processed with a simulation of the CMS detector response based on GEANT 4 [3]. The applied misalignment, miscalibration and dead channel map correspond to the expected start-up conditions. The longitudinal distribution of the primary collision vertices has been tuned to match the real data. The signal in the silicon strip tracker was simulated in *peak* mode in agreement with the mode used in the readout chips, during the data taking [**?**].

To select the minimum bias collsion events, the skimmed collision dataset has the following technical trigger bits requirements:

- BSC trigger: technical trigger bit 40 or 41
- Veto BeamHalo: Triggers: 36, 37, 38, 39

In addition, we apply technical trigger bit 0 to pick up the correct bunch crossing for data. Note that this is not simulated in MC.

2.2 Event Selection

To reduce further the background from non-collision events and to select useful events for tracking studies, we have the following event selections:

- At least one real primary vertex reconstructed
- fraction of *highPurity* tracks larger than 20% if the number of reconstructed tracks is larger than 10.

Table 1 shows the selection effcidency and final number of events analysis in each datasets.

Table 1: List of datasets and selection efficiency. The selection efficiency on n-th colomn is obtained in addition to the cuts in the preceding n-1 columns.

Dataset name	Trigger	primary vertex	track high purity	Events left
900 GeV Dec19thSkim	100%	94%	100%	260,000
900 GeV STARTUP3X_V8K	67%	93%	100%	256,000
2360 GeV Dec19thSkim	100 %	95%	100%	12,959
2360 GeV STARTUP3X_V8L	68%	94%	100%	12,491

3 Primary Vertex Resolution Measured with Two-Vertex Method

3.1 Two-vertex Method Description

To measure the primary vertex resolution, we measure the primary vertex position of each event with two independent sets of tracks. We obtain the algorithm resolution by comparing the relative position of the two vertices in x, y and z. The distribution of the difference in the fitted vertex positions can then be used to extract the resolution and pulls. Figure 1 illustrates the basic idea of this method.

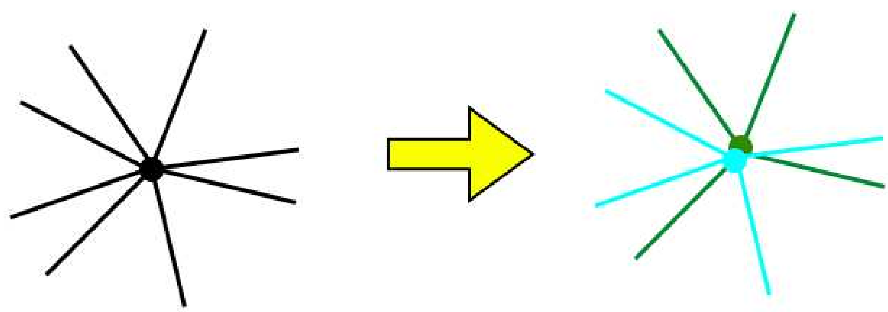

Figure 1: *Illustration of the two-vertex method. Please see text for more details.*

The method proceeds as follows.

1. Divide tracks used in the primary vertex in two independent sets (A and B)
   - We take tracks used the "offlinePrimaryVertices" and sort them in descending order of P_T.
   - Starting from highest-$P_T$track, we take pairs of tracks and randomly choose which track goes into which set (A or B). This ensures that there is no asymmetry of the track P_T spectrum in the two vertices. If the number of the tracks are odd, we drop the lowest-$P_T$track.
2. Vertex fit each trackset (A and B) independently using "offlinePrimaryVertices" algorithm.

   We monitor the basic track distributions (P_T, η, ϕ, dxy and dz) of vertices reconstructed with split tracksets. Figure 2 show the comparison of these variables between the tracks from original vertex and two split vertices, in the mininum bias collision data taken at $\sqrt{s}$ =900GeV. We see that the tracks used in all three cases have consistent kinematics.

3. Extract resolution and pull from the difference in two fitted vertices

   When comparing the two fitted vertices, it is important that the number of tracks used in the final fitted vertices and the ΣP_T and z of the two vertices do not differ by too much. As we sort the tracks according the P_T prior to splitting, the difference of the ΣP_T between the two vertices are negligible.

   We apply additional selections on the following two variables to ensure the two vertices we compare bear similar kinmematics.

   - Relative difference of the number of tracks used in the two vertices, $|\frac{\mathtt{nTrk(vtx1)-nTrk(vtx2)}}{\mathtt{nTrk(vtx1)+nTrk(vtx2)}}| < 0.1$
   - The separation signficance in z, defined as $\frac{|z(vtx1)-z(vtx2)|}{\mathtt{max}(\sigma\mathtt{z(vtx1)},\sigma\mathtt{z(vtx2)})} > 5$

   Figure 3 shows the distribution of these two variables before applying the cuts. The selection efficiency of the two cuts is found to be 74% in both data and MC.

   We analyze the difference between the two selected matching vertices and and extract the resolution and pull dependence of the number of tracks used in the primary vertex as follows.

   - The difference between the coordinates of the two vertices in x, y and z is histogrammed for each number of tracks. We fit a single gaussian to CORE of the distribution in the range of $\pm 2\times$RMS. The resolution is defined as the gaussian width σ divided by $\sqrt{2}$.
   - The pull distribution at each number of track bin is filled with the quantity $\frac{x_1-x_2}{\sqrt{\sigma x_1^2+\sigma x_2^2}}$, where x_i and σx_i are the vertex position and errors of the two vertices. We fit a single gaussian to each distribution in the range of $\pm 2\times$RMS.

Before moving forward, it is important to point out that this method provides an estimate of the resolution and pulls comparing vertices formed with average number of tracks that are about half of the number of tracks that are used in the actual primary vertex fitter. As the intrinsic vertex resolution is well known to decrease with the number tracks used in the primary vertex fitter. Therefore, the average resolutin obtained from this method will always be larger than the actual resolution in a given sample. More details on the systematic uncertainties are described in section 6.

3.2 Resolution and Pulls in Different P_T Ranges

The primary vertex resolution depends strongly on the $P_T$of the tracks used in the fit or better on the average $P_T$of the track set. To study this dependence in collision data we use a variant of the method described above, since the poor statistics available does not permit to sample different $P_T$ranges. To override this problem, for each event we perform a *multiple* split of the selected tracks in pairs of sets with the same number, requiring that the difference in the average $P_T$for each pair of sets differe less than 10%. In other words from the original collection of N tracks we form randomly (N/2-1) pairs of sets each containing a number of tracks ranging from 2 to N/2. In this way it is possible to improve the statistics in the low number of tracks bins using events with higher number of tracks.

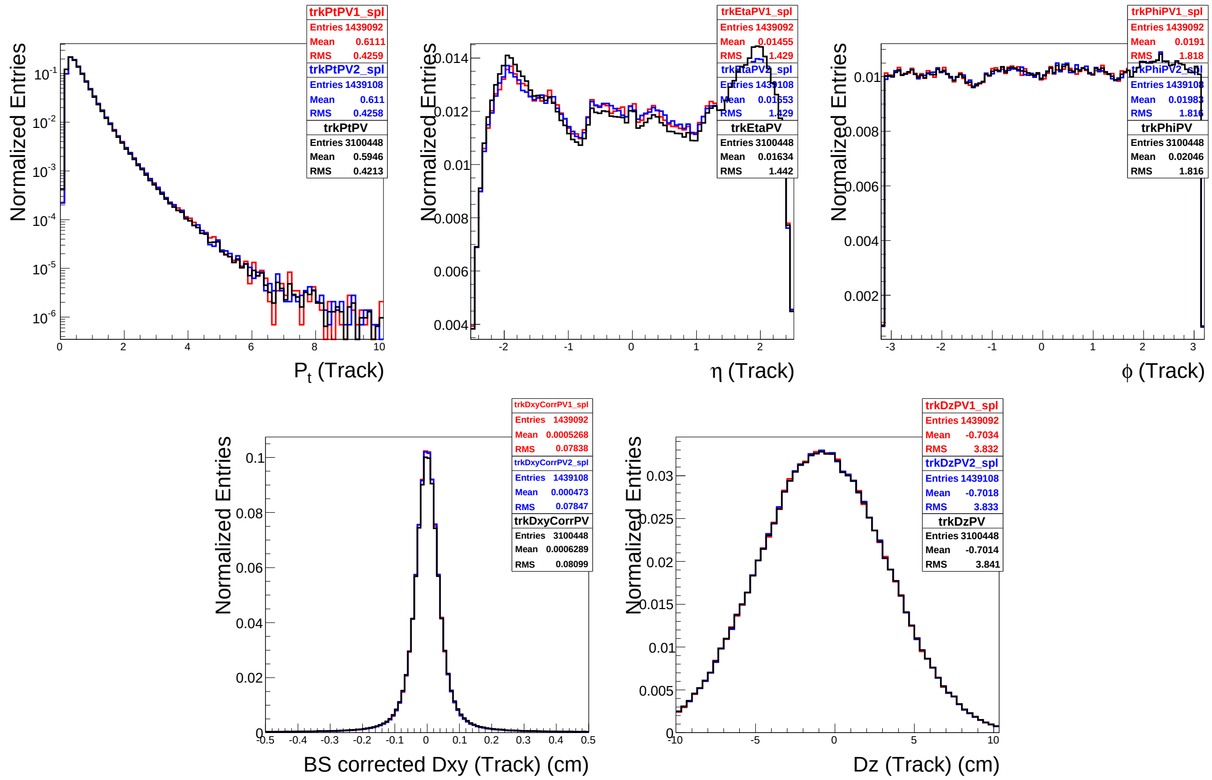

Figure 2: *Track P_T, η, Φ, dxy and dz distributions of the tracks used in the orginal primary vertex (black), split vertex A (red) and split vertex B (blue).*

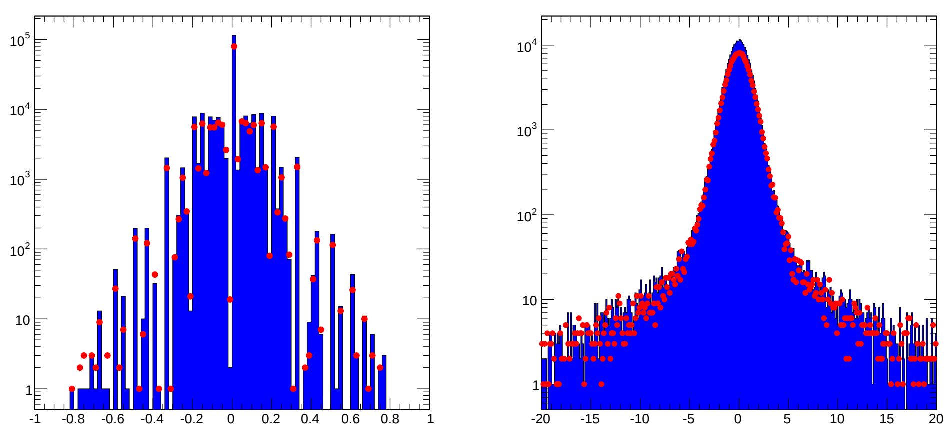

Figure 3: *Left: The relative difference of number of tracks used in the two vertex fit; Right: The z-signficance of the two split vertices fit. Please see text for the definition. Data is shown with red dots while MC is shown as histogram filled in blue.*

4 Two-vertex Method Validation on MC Samples

We validate the two-vertex method on MC data, by comparing the resolution and pulls obtained between two-vertex method and MC method. In MC method, the primary vertex resolution is evaluated by comparing the reconstructed and simulated positions $x_{\mathtt{rec}} - x_{\mathtt{sim}}$.

4.1 Resolution and Pulls vs Number of Tracks

Figure 4- 5 show the resolution and pull as a function of the nubmer of tracks used in the primary vertex for four different methods. In each plot, the results of the following four methods are overlaid:

- Black: MC method applied on the original unsplit vertex collection
- Red: MC method applied on the split vertex collection A
- Green: MC method applied on the split vertex collection B
- Blue: Two-vertex method

We see good agreement between the results obtained with MC and two-vertex methods. There is slight difference at large number of track bins (>10) between the resolutions obtained by applying MC method to the unsplit trackset and the rest of the methods. It is because of the average P_T of the split track sets are in general 100 MeV larger than the original *generalTracks*. At low energy at $\sqrt{s} = 900$ GeV, the average track P_T is close to 0.5 GeV, this small difference in P_T introduces difference in the track impact parameter resolution, hence translated into the difference in vertex resolution. This effect is expected to become much smaller on high energy collisions.

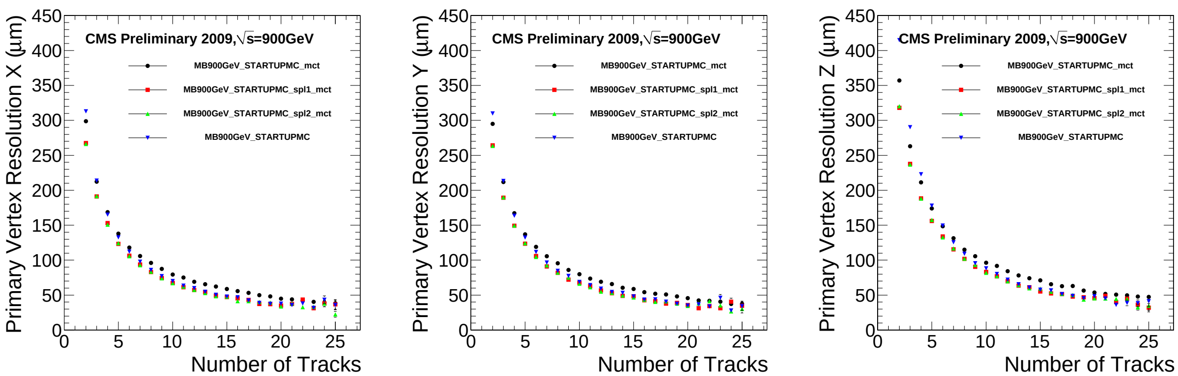

Figure 4: *Primary vertex resolution as a function of the number of tracks used in the fitted vertex.*

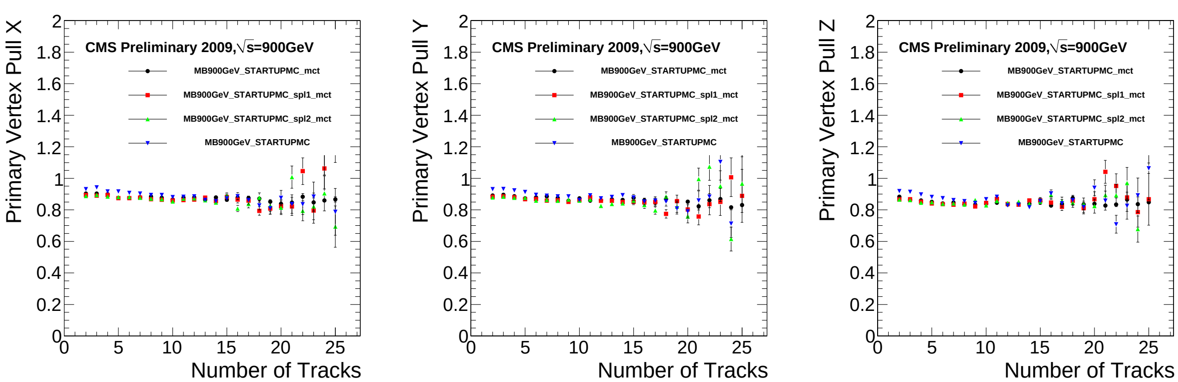

Figure 5: *Fitted pulls from primary vertex distribution as a function of the number of tracks used in the fitted vertex.*

4.2 Resolution and Pulls vs Number of Tracks at Different P_T Ranges

Using the alternative split algorithm described in section 3.2 we can compare the two-vertex metod outcomes with the MC truth in different $\bar{P}_T$bins. Here we consider three o.4 GeV/c $\bar{P}_T$bins, from 0 to 1.2 GeV/c.

Figure 6 shows the comparison of the resolutions obtained with the split sets and the MC true vertex. The difference in the resolution at high number of tracks discussed in 4.1 disappear comparing the MC truth and the data driven metho at the same $\bar{P}_T$. We have to notice there there is still a relevant difference between the MC truth and the data driven method for small numbers of tracks (< 4). For the lowest $\bar{P}_T$bin the difference in resolution is about 100 μ m in X,Y and 150 μ m in Z in the case that two tracks are used in the primary vertex fit. This is to be ascribed to the different *physical* content of the events considered in the two cases: for the MC the primary vertex is originally reconstructed with only two tracks, hence with the same weight in the fit [1]. On the other side, for the two vertex method, the original vertex in the event has been reconstructed using four tracks at least, that in principle can have very different weights in the fit. A further discussion about the effect of the weights in the resolution is presented in section 6.2. However once the number of tracks increase this effect is ruled out.

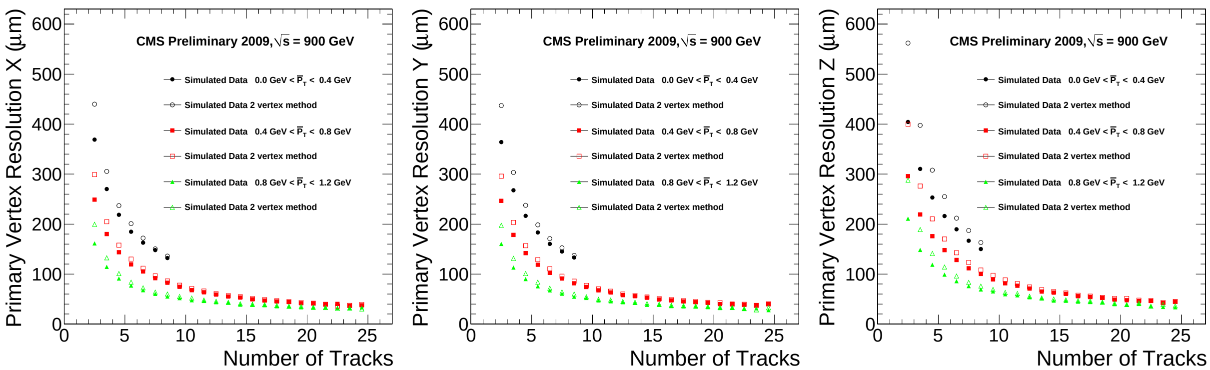

Figure 6: *Primary vertex resolution as a function of the number of tracks used in the fit in three diffetent $\bar{P}_T$bins.*

[1] The Adaptive Vertex Filter in the case of only two tracks reduces to the Kalman Vertex Filter.

5 Resolution and Pull on Collision Data

5.1 Resolution and Pull vs Number of Tracks

Figure 7 shows the measured primary vertex resolution as a function of the number of tracks in x (left), y (middle), and z (right). Results are shown for both the December data and the MC and a good agreement in the curves is seen. As shown in Figure 8, we compare the two-vertex resolution with the default vertex error of"offlinePrimaryVertices" using all tracks on the collision data. The two results match very well in the plateu region, with average difference less than 5 μm. Figure 9 shows the measured pulls on the primary vertex, using the difference in the measured position and uncertainties reported by the fit. The pulls are roughly flat and close to unity.

The primary vertex resolution measured are much smaller than the beam spot resontruction. As shown in Ref. [4], the transverse beam width is measured to be $\sigma_x \sim 200\mu m$ and $\sigma_y \sim 120\mu m$ in the 900GeV data and $\sigma_{x,y} \sim 120\mu m$ in the 2.36 TeV data. The length in z is found to be $\sigma_z \sim 4$ cm in the 900GeV data and $\sigma_z \sim 2.8$ cm in the 2.36TeV data.

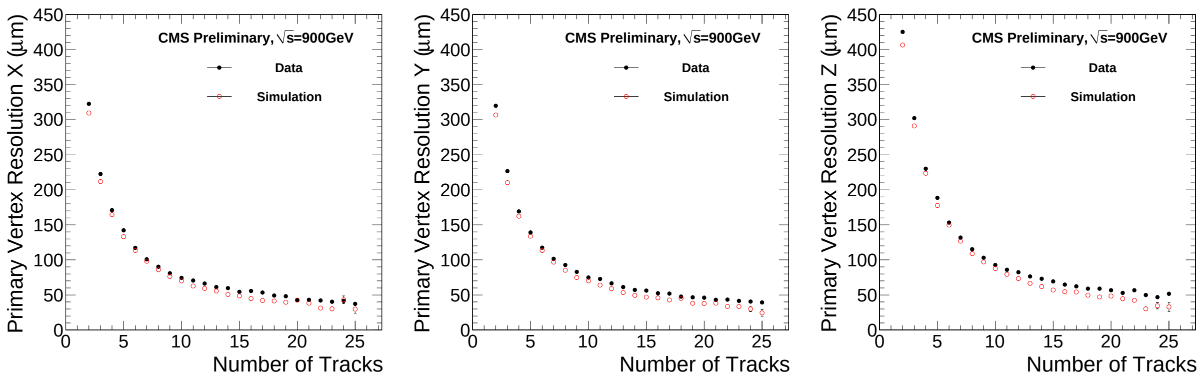

Figure 7: *Primary vertex resolution as a function of the number of tracks used in the fitted vertex.*

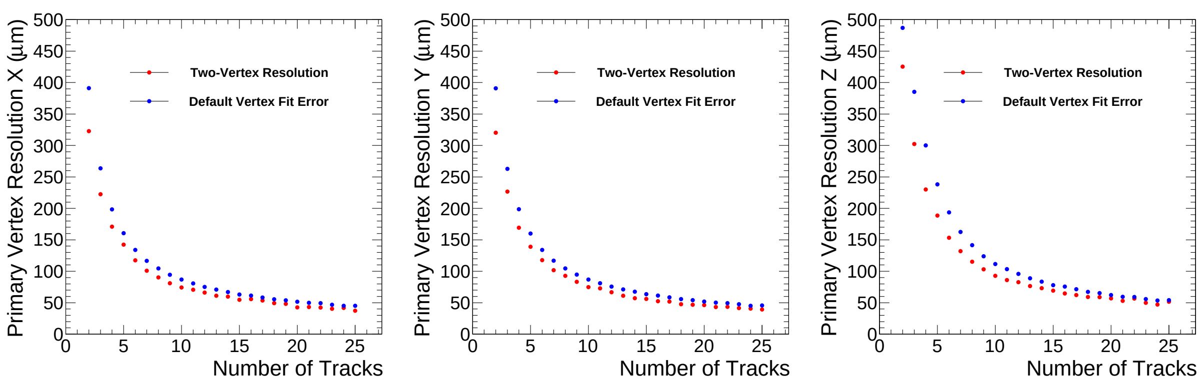

Figure 8: *Comparison of the two-vertex resolution with the default vertex error of"offlinePrimaryVertices" using all tracks on the collision data with* $\sqrt{s} = 900 GeV$.

5.2 Resolution and Pull vs Number of Tracks at Different P_T Ranges

Figures 10 and 11 show the resolution and pulls in three different $\bar{P}_T$ranges for MC and 900 GeV collisons data. The agreement with MC is very good for all the different $\bar{P}_T$bins. The pulls are falt and close to unity in the whole range.

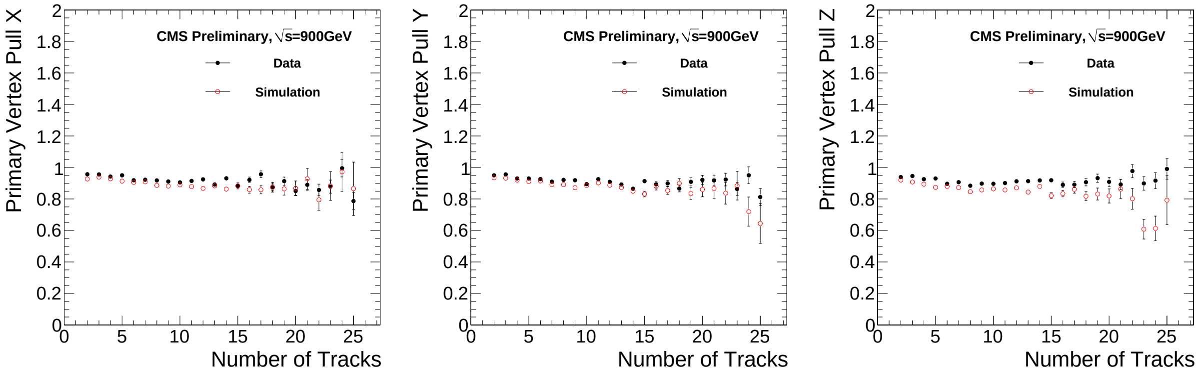

Figure 9: *Fitted pulls from primary vertex distribution as a function of the number of tracks used in the fitted vertex in 900GeV data.*

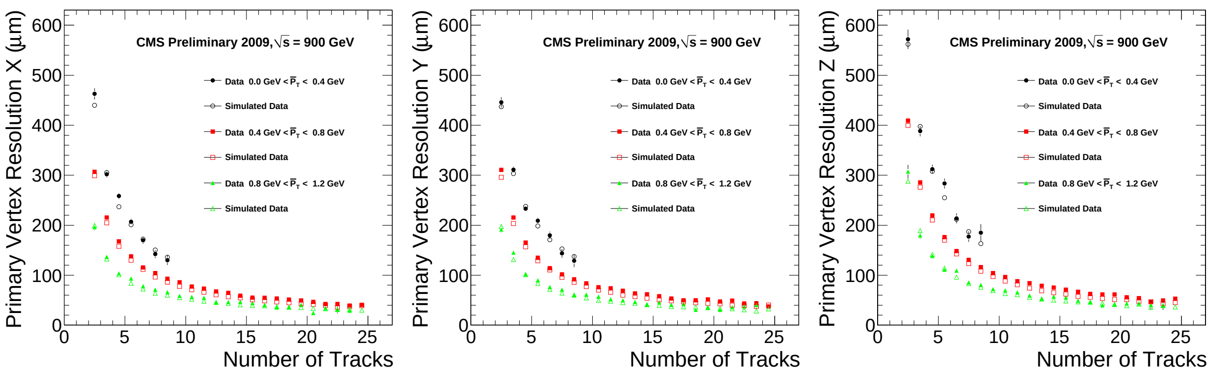

Figure 10: *Primary vertex resolution as a function of the number of tracks used in the fitted vertex in three different* $\bar{P}_T$ *bins.*

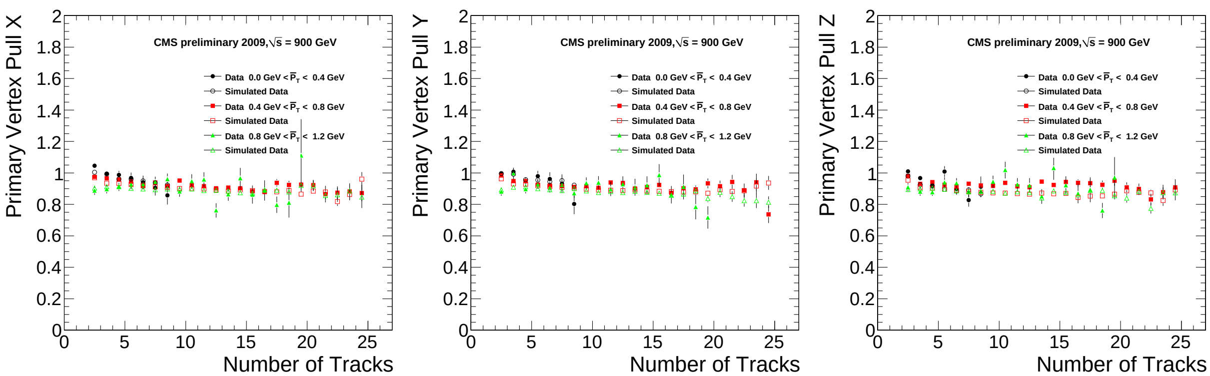

Figure 11: *Fitted pulls from primary vertex distribution as a function of the number of tracks used in the fitted vertex for three different* $\bar{P}_T$ *bins.*

5.3 Compare Resolution and Pulls of 900 and 2360 GeV Data

We apply the two-vertex method on the 2.36 TeV data. Figure 12-13 show the resolution and pulls results comparing $\sqrt{s} = 900$GeV and 2360GeV data. We observe that the resolutions taken at these two energies are consistent.

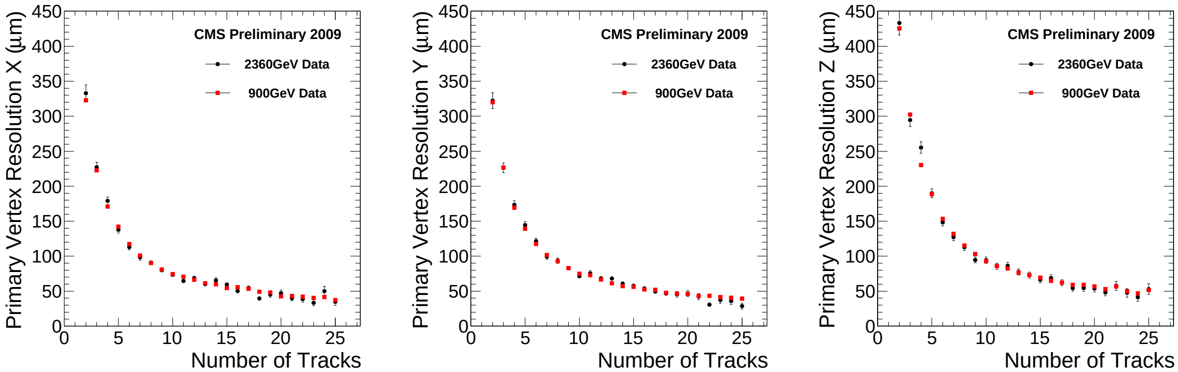

Figure 12: *Primary vertex resolution as a function of the number of tracks used in the fitted vertex.*

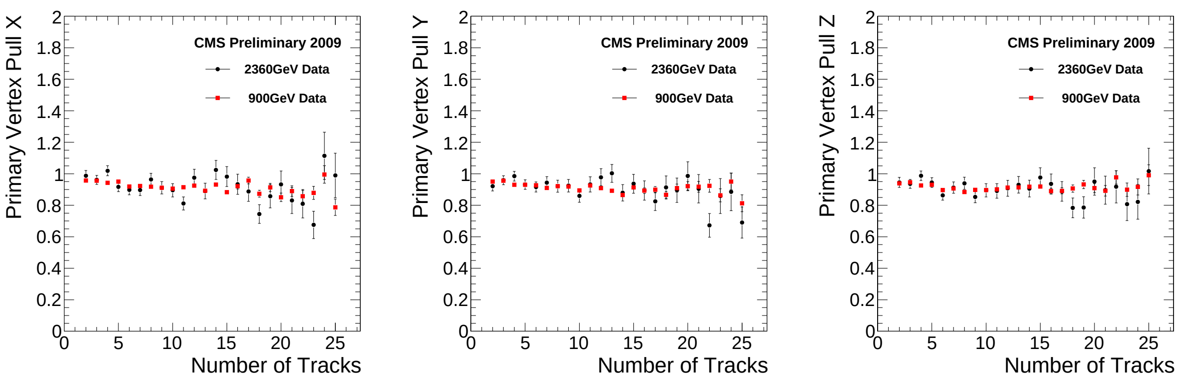

Figure 13: *Fitted pulls from primary vertex distribution as a function of the number of tracks used in the fitted vertex.*

6 Systematic Effects and Cross Checks

In this section, we discuss briefly the systematic uncertainties in the primary vertex resolution measurement with two-vertex method.

6.1 Effect of the Alignment Scenario

The primary vertex resolution depends on the track impact parameter resolution, which is affected by the tracker alignment. At the early stage of data taking, the alignment senario may not describe the actual value. The systematic uncertainties on the track impact parameters induced by the misalignment will be translated to the systematic uncertainties of primary vertex resolution.

To estimate the effect of misalignment on the primary vertex resolution, we repeat the study on the MC data simulated with perfectly aligned tracker and with the alignment position error (APE) set to 0. Figure 14- 15 show the resolution and pulls versus number of tracks results comparing 900 GeV collision data, MC with misalignment scenario matching the data taking and the MC with ideal geometry and APE.

From this study, we estimate systematic uncertainty due to the mis-alignment is within 5 μm when the number of tracks exceed 5. The difference in resolution can be as large as around 20-30 μm when number of tracks are less than 5. This difference also shows up in the corresponding bins of pull distributions. Besides the expected larger uncertainties on these low number of tracks vertices, we also find that the single gaussian fit does not describe the residual well. It is sensitive to the fit range. Nevertheless, in all number of track bins, the mis-aligned systematic uncertainties are within 10% of the resolution itself.

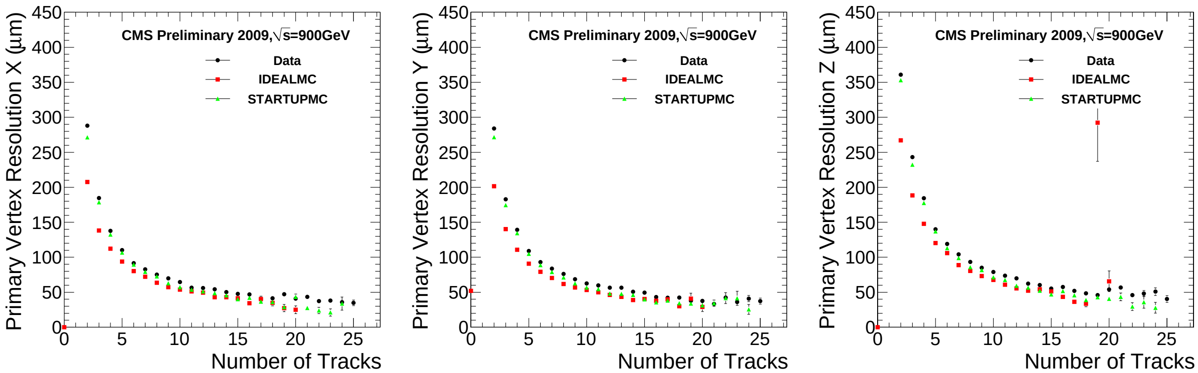

Figure 14: *Primary vertex resolution as a function of the number of tracks used in the fitted vertex, comparing the collison data taken at* $\sqrt{s} = 900 GeV$ *(Black), mininum bias MC with startup (green) and ideal (red) alignment.*

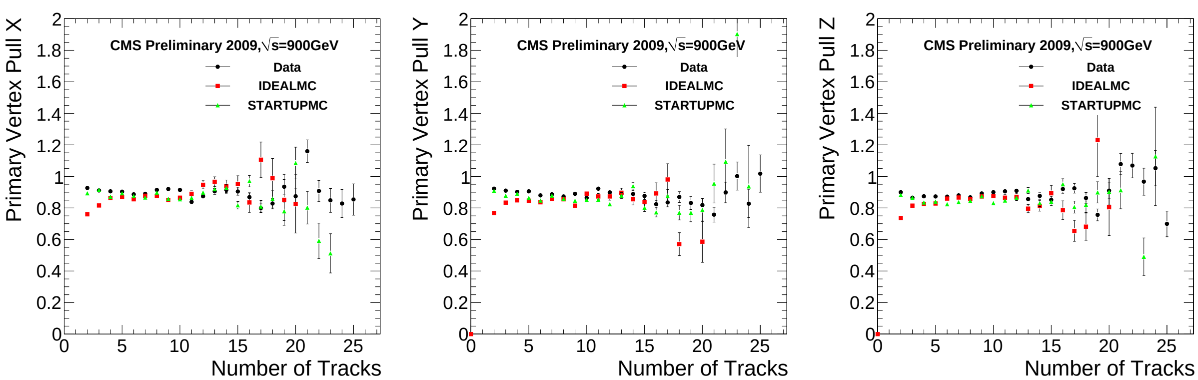

Figure 15: *Fitted pulls from primary vertex distribution as a function of the number of tracks used in the fitted vertex, comparing the collison data taken at* $\sqrt{s} = 900 GeV$ *(Black), mininum bias MC with startup (green) and ideal (red) alignment.*

6.2 Effect of the split vertex weight difference

In the Adaptiver filter, all tracks in a given cluster are used in the vertex fit. The outlier tracks with larger distance to the vertex postion are down weighted significantly thus contribute little to the fitted vertex positions or errors. Thus as we compare the two split vertices, though the number of tracks do not differ, the number of tracks with high weights (> 0.5) (or the numbers of degree of freedom) may differ. In that case, the difference between the two vertex positions does not represent the vertex error in either vertex.

To cross check this effect, we repeat the study with resolution and pull versus the number of high weighted tracks instead of the total number of tracks. Figure 16- 17 show the comparison between the two approaches. The difference between the two methods are expected as the number of high weighted tracks is always smaller than the number of total tracks in the vertex. On the other hand, the resolution at the stable (or tail) region of number of tracks are consistent with each. This indicates that the low weight tracks do not affect the measurement. Besides, the pull distributions are consistent in all number of tracks bins.

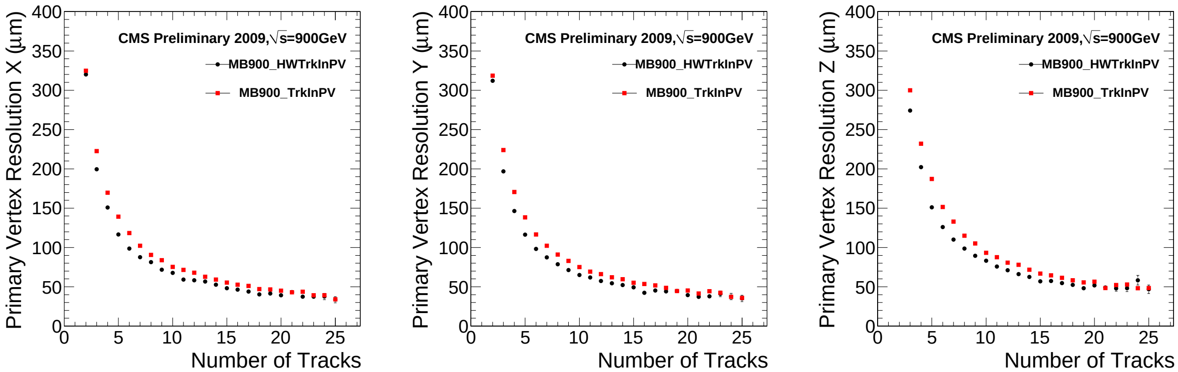

Figure 16: *Primary vertex resolution as a function of the number of tracks used in the fitted vertex.*

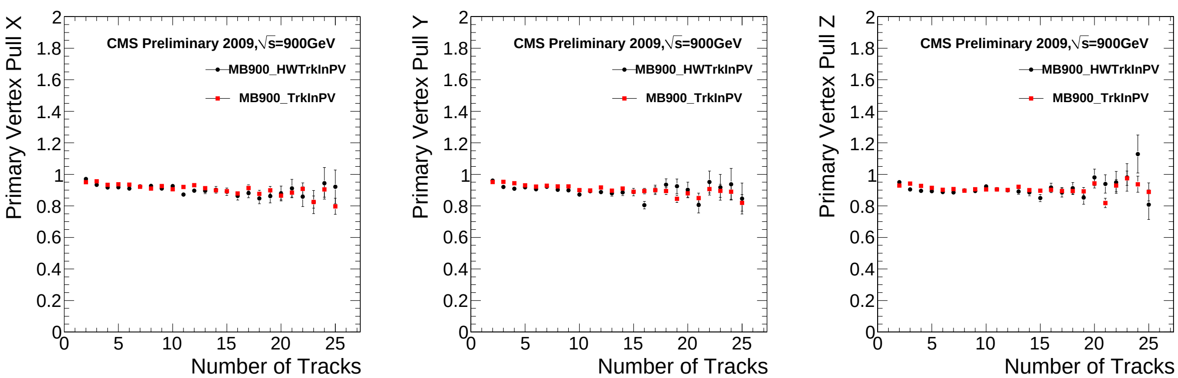

Figure 17: *Fitted pulls from primary vertex distribution as a function of the number of tracks used in the fitted vertex.*

7 Conclusion

In conclusion, we have developped a data-driven algorithm, referred as two-vertex method, to measure the primary vertex resolution. We have validated the method on the MC simulated data. The results obtained using the two-vertex method are consistent with the results obtained by comparing the reconstructed and simulated vertex positions.

We apply the two-vertex method to the mininum bias data taken at $\sqrt{s}$ =900 GeV and 2360 GeV, reprocessed in the Dec19th cycle. The resolution and pull distributions versus the number of tracks used in the primary vertex and the average P_T of those tracks are consistent with the results obtained by applying the two-vertex method on the MC data.

The primary vertex resolutions are measured to be much smaller than the beam width in both 900 and 2360GeV Data. More specifically, the primary vertex resolution in x and y are measured to be close to 50 μm, while the transverse beam width is measured to be $\sigma_x \sim 200\mu m$ and $\sigma_y \sim 120\mu m$ in the 900GeV data and $\sigma_{x,y} \sim 120\mu m$ in the 2.36 TeV data. The primary vertex resolution in z is measured as close to 70 μm with more than 10 tracks, while the beam length in z is found to be $\sigma_z \sim 4$ cm in the 900GeV data and $\sigma_z \sim 2.8$ cm in the 2.36TeV data.

References

[1] CMS NOTE-2006/032, “Vertex Fitting in the CMS Tracker” CMS NOTE-2004/002, “Sensitivity of Robust Vertex Fitting Algorithms”

[2] Pythia generator package, http://home.thep.lu.se/ torbjorn/Pythia.html

[3] Allison, J. et, al, *IEEE Transaction on Nuclear Science*, **53 No.1 (2006) 270.**

[4] CMS PAS TRK-10-001, “Tracking and Vertexing Results from First Collisions”