LHCb-INT-20XX-YYY
April 15, 2016

Primary Vertex reconstruction for Run II of data taking

Agnieszka Dziurda[1], Vladimir Vava Gligorov[2], Mariusz Witek[1]

[1] *Institute of Nuclear Physics PAN (IFJ), Krakow,*
[2] *European Organization for Nuclear Research (CERN), Geneva*

Abstract

We report...

1 Introduction

The LHCb detector reconstruction and real-time data analysis ("trigger") are sensitive to the LHC running conditions, in particular to the multiplicity of the particles produced in the *pp* collisions and their momenta. As the LHC increased its energy from the $\sqrt{s} = 7(8)$ TeV achieved in 2011 (2012) to $\sqrt{s} = 13$ TeV for Run II, (2015-2018), the LHCb reconstruction had to be reoptimized. Additionally, the power of the Event Filter Farm roughly doubled between 2012 and 2015, which significantly increased the time budget for the reconstruction and allowed a more offline-like reconstruction inside the trigger. This documents reports the reoptimization of the Primary Vertex (PV) reconstruction. This new optimization enables an offline-quality reconstruction to be used in the online system, results in a better PV finding efficiency and lower fake rate. Natural units are used throughout.

2 Principles of PV reconstruction

The objective of the primary vertex (PV) reconstruction is to determine the position and composition of the proton-proton interactions which occur within LHCb's vertex detector. It is composed of two steps : seeding and fitting.

The position of the *pp* collision is determined using a set of input tracks. A search is performed in a loop over tracks and for each of them a number of *close* tracks is determined. A track is defined as *close* when the distance of the closest approach to the reference track is less then 1 mm. A selected group of tracks is called a *seed* and it creates a candidate for a PV. For each pair of tracks in the *seed*, the distance of the closest approach is calculated and their weighted average determines the final position of the PV candidate.

In the second step, the position of the PV is found by minimising the the χ^2 defined as:

$$\text{Vertex}\chi^2 = \sum_{i=1}^{\text{nTracks}} \text{IP}\chi_i^2 \cdot w_i, \tag{1}$$

where nTracks denotes the number of tracks associated to the vertex, $\text{IP}\chi_i^2$ is a single track impact parameter χ^2 (defined as the change in the vertex χ^2 when said track is included in the vertex, relative to when it is not included) and w_i is its weight, which parametrizes the significance of the given track in the vertex fit. The weights w_i are obtained for each track i based on the adaptive weighted least square method with Tukey biweights [1] and are given by:

$$w = \begin{cases} (1 - \chi_{IP}^2/C_T^2)^2, & \text{if } \chi_{IP} < C_T, \\ 0, & \text{if } \chi_{IP} \geq C_T, \end{cases} \tag{2}$$

where C_T is Tukey's constant. The χ^2 defined in Eq. 1 has several advantages over a simple least-squares optimization. Firstly, it minimizes the impact of secondary particles which do not originate directly from the PV, but are likely assigned to the PV. This could lead to a systematic shift towards the secondary vertex from which these displaced particles

originate. Secondly, the fraction of badly reconstructed tracks (ghosts) is minimized, which are dangerous because they have an underestimated error and could therefore bias the PV position.

A more detailed description of the algorithm procedure can be found in [2–4].

3 Inputs to the algorithm

The PV finding algorithm begins with a set of input tracks. In 2010-2012 data taking two algorithms were in use: online and offline. Since the time budget in the first trigger stage (HLT1) was limited, the online reconstruction was based on unfitted VELO tracks, which didn't have an accurate covariance matrix. More precise results were obtained by the second, offline, reconstruction based on Kalman fitted tracks [5]. The online (offline) algorithm was executed in Moore (Brunel).

Thanks to developments in the trigger, the time budget is increased and it is possible to perform a Kalman fit in HLT1. Therefore the aim of these developments is to demonstrate that a single algorithm can meet the online timing constraints without sacrificing any offline performance. For this purpose the new 2015 online algorithm uses Kalman fitted VELO tracks.

The agreement between the algorithms can be quantified by the decorrelation variable given by:

$$\frac{A_{OFF} - A_{ON}}{\sigma_{OFF}}, \tag{3}$$

where $A = \{x, y, z\}$ denotes coordinates, $A_{OFF}(A_{ON})$ is position of PVs found by the offline (online) algorithm and σ_{OFF} is the position uncertainty for the offline approach. The obtained distribution is parametrized by a triple Gaussian:

$$PDF_{tG}(\mu, \sigma_1, \sigma_2, \sigma_3, f_1, f_2; X) = f_1 G(\mu, \sigma_1; X) + f_2 G(\mu, \sigma_2; X) + (1 - f_1 - f_2) G(\mu, \sigma_3; X), \tag{4}$$

where $G(\mu, \sigma_i; X)$ denotes a single Gaussian distribution with common mean μ and width $\sigma = \sigma_i$ and X is an observable. The average width $\bar{\sigma}$ is defined as:

$$\overline{\sigma}^{tG} = \sqrt{f_1 \sigma_1^2 + (1 - f_1) f_2 \sigma_2^2 + (1 - f_1)(1 - f_2) \sigma_3^2}. \tag{5}$$

Finally, the difference d between approaches is described by:

$$d = \sqrt{\overline{\sigma}^{tG} + \sigma_N} - 1, \tag{6}$$

where σ_N denotes the nominal value equals to 1. The initial difference between the online algorithm with unfitted VELO tracks and the offline algorithm is 64% (58%) for the $\mathbf{x}(\mathbf{z})$ coordinate [1]. The agreement is improved by using Kalman fitted VELO tracks for the online algorithm, where the results are 21% and 16% for the $\mathbf{x}$ and $\mathbf{z}$ coordinates,

[1] Unless explicitly stated otherwise, the performance for the $\mathbf{y}$ coordinate is similar to $\mathbf{x}$ throughout this document, and is not quoted.

respectively. This difference can be reduced even more by including Long tracks with a transverse momentum (p_{T}) above 500 MeV/c. However, as we shall see later, a further reduction in this decorrelation does not in fact improve the actual resolution of the PV algorithm, but simply reduces the stochastic differences between the algorithms. Therefore, the PV algorithm which uses Kalman fitted VELO tracks is considered in further studies as the nominal candidate for Run II datataking. This will be referred to as the "online 2015" algorithm in what follows. The summary of the decorrelation results is collected in Tab. 1.

Table 1: The differences between the offline and online primary vertex algorithms.

Parameter	online 2010-2012 unfitted Velo tracks	online 2015 Kalman fitted Velo tracks	online 2015 Kalman fitted Velo + Long $p_T > 500$ MeV/c
$\overline{\sigma}_x^{tG}$	1.294	0.679	0.414
μ_x	0.007	-0.002	-0.001
d [%]	64	21	8.4
$\overline{\sigma}_z^{tG}$	1.123	0.599	0.478
μ_z	-0.002	0.004	0.0003
d [%]	58	16	10

3.1 Performance

The quality of the PV algorithm can be studied in terms of its resolution and pseudo pulls. All studies are based on a $b\bar{b}$-inclusive simulation sample in nominal 2015 conditions; a data-driven measurement of the resolution is presented later in Sec. 5.

3.1.1 Global resolution

The global resolution is determined from the distribution of the difference between the fitted and generated position of the PV, ΔA, where A denotes a given coordinate. The resolution is obtained by fitting a triple Gaussian given by Eq. 4 and taking its averaged width as the parameter in question. The results are collected in Tab. 2 and an example fit for the online 2015 algorithm is presented in Fig. 1. The resolution for the **z** coordinate is compatible within uncertainty for all considered approaches, while the resolution for the **x** coordinate is worse by (2-3)% with respect to offline algorithm. However, since the **z** resolution completely dominates, this is not a critical loss in performance.

The Kalman fit uses the p_{T} of the track when computing the multiple scattering correction. The p_{T} of the VELO tracks used in the online 2015 algorithm is unknown, therefore to enable the fit, it is set to an average value of 400 MeV. The influence of the global resolution on this average value is studied. Fig. 2 shows the global resolution as a function of average p_T and the minimum is found to be at 300 MeV.

Table 2: The global resolution for the online 2015 and offline primary vertex reconstructions.

Parameter [mm]	online 2015	offline
Δx	0.0205 ± 0.0001	0.0198 ± 0.0001
Δz	0.1432 ± 0.0006	0.1433 ± 0.0006

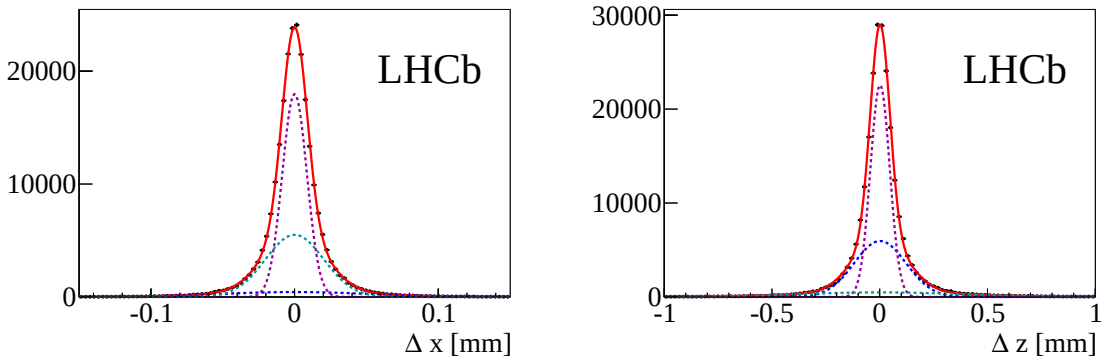

Figure 1: The resolution obtained in the $b\bar{b}$-inclusive sample for the online 2015 algorithm, for the (left) **x** coordinate and (right) **z** coordinate. The solid line corresponds to a triple Gaussian fit, while the dashed lines denote its respective components.

3.1.2 Pseudo pulls

The pseudo pulls for the primary vertex are given by the global resolution divided by reconstructed position uncertainty. This distribution is parametrized by a double Gaussian:

$$PDF_{dG}(\mu, \sigma_1, \sigma_2, f_1; P) = f_1 G(\mu, \sigma_1; P) + (f_1 - 1) G(\mu, \sigma_2; P), \tag{7}$$

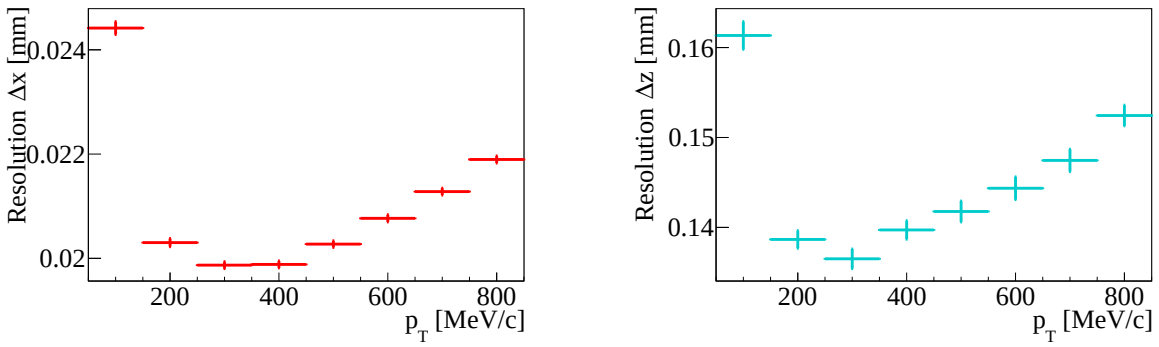

Figure 2: Global resolution as a function of the average VELO track p_{T}, for the (left) **x** coordinate and (right) **z** coordinate.

where $G(\mu, \sigma_i; t)$ denotes a single Gaussian distribution with common mean μ and width $\sigma = \sigma_i$. The P denotes a pseudo pull observable. The average width $\overline{\sigma}^{dG}$ is defined as:

$$\overline{\sigma}^{dG} = \sqrt{f_1\sigma_1^2 + (1 - f_1)\sigma_2^2}. \tag{8}$$

An average width $\overline{\sigma}^{dG} = 1$ and $\mu = 0$ are signatures of unbiased results. The obtained results are collected in Tab. 3, while the fits are shown for the online 2015 algorithm in Fig. 3. The mean μ_x for both online 2015 and offline approaches is consistent with zero within uncertainty. A small bias is observed for the **z** coordinate, where the results for the online 2015 algorithm are less shifted. In addition, the average widths for the online 2015 algorithm describe the uncertainties accurately, while the offline algorithm underestimates the uncertainties.

Table 3: The pseudo pulls for the online 2015 and offline primary vertex algorithms.

Parameter	online 2015	offline
μ_x	0.001 ± 0.002	-0.001 ± 0.002
$\overline{\sigma}_x^{dG}$	1.04± 0.02	1.17 ± 0.04
μ_z	0.030± 0.002	0.038 ± 0.002
$\overline{\sigma}_z^{dG}$	1.06± 0.02	1.13 ± 0.02

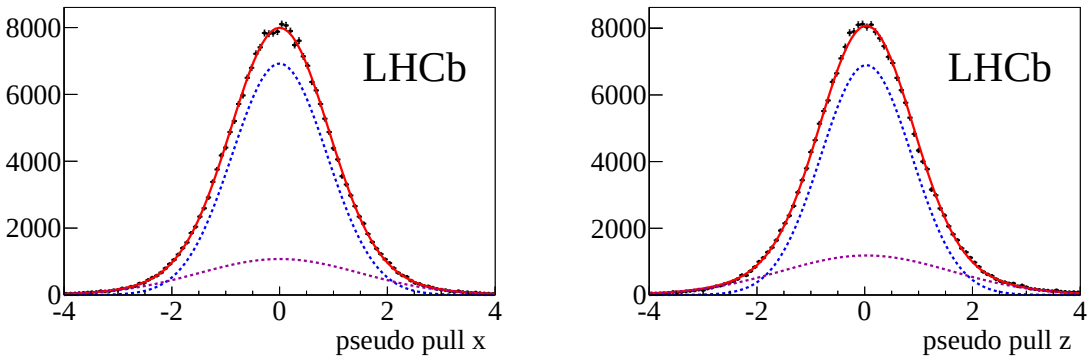

Figure 3: The pseudo pull obtained in the $b\bar{b}$-inclusive sample for the online 2015 algorithm, for the (left) **x** coordinate and (right) **z** coordinate. The solid line corresponds to a double Gaussian fit, while the dashed lines denote its respective components.

3.1.3 Impact parameter distributions

Another quantity which has to be considered is the track's impact parameter, which is defined as the minimum perpendicular distance between the reconstructed track and the primary vertex. The magnitude of the impact parameter is calculated for all tracks from

the best container. Fig. 4 shows the obtained results which are split according to the p_{T} of the particle associated to the track: below and above 1 GeV. This represents 85% and 15% of tracks in the container, respectively. In the first case, below 1 GeV, the distributions are in good agreement for various algorithms. A discrepancy is seen for tracks above 1 GeV, where the impact parameter calculated with respect to the online 2015 primary vertices is shifted upwards. Most likely, it can be explained by the influence of assigning an average $p_T = 400$ MeV to all VELO tracks as described above. Since this value is much less than 1 GeV, these tracks have a lower weight in the online 2015 algorithm compared to the offline algorithm, and therefore the position of the reconstructed primary vertex is slightly shifted away from them. To study this effect the Landau distribution is fitted with mean μ_L and width σ_L, and the results are listed in Tab. 4. As can be seen the Landau widths are in good agreement, whereas the mean is shifted by 9% upwards. Since the mean can be interpreted as the expected resolution, this implies 9% worse impact parameter resolution for 15% of the tracks. The previous behaviour could be recovered by adding Long tracks with $p_T > 500$ MeV, as also shown in Fig. 4.

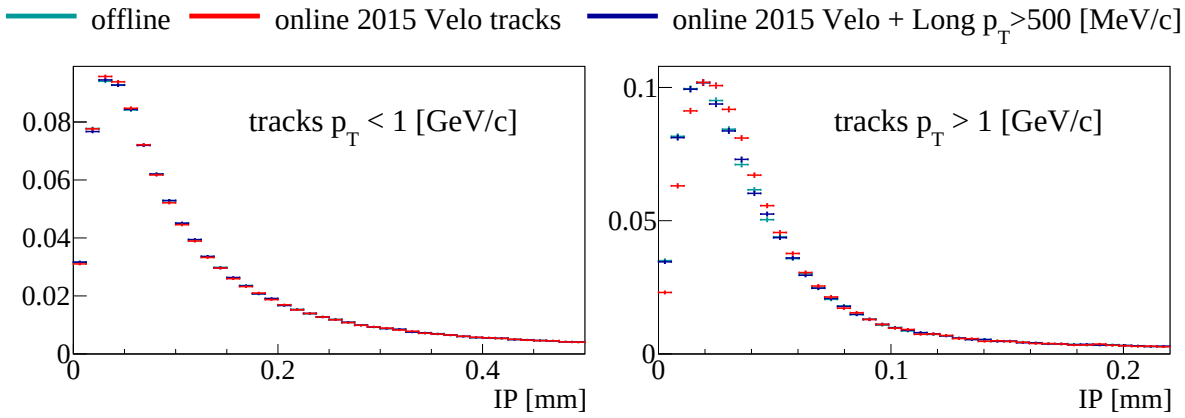

Figure 4: The impact parameter calculated with respect to the primary vertex for tracks with (left) $p_T < 1$ GeV, and (right) $p_T > 1$ GeV.

Table 4: Result of fitting a Landau distribution to the impact parameter calculated with respect to the online 2015 and offline primary vertices for tracks with $p_T > 1$ GeV.

Parameter	online 2015	offline
σ_L	0.0097 ± 0.0004	0.0096 ± 0.0004
μ_L	0.0228 ± 0.0006	0.0208 ± 0.0006

3.1.4 Influence on the decay-time resolution

A possible influence on the decay-time resolution is studied based on the signal $B_s^0 \to \phi^0\phi^0$ simulation sample, generated with the nominal 2015 conditions. The decay-time resolution

is defined by:

$$\Delta t = t_{rec} - t_{true}, \tag{9}$$

where t_{rec} is the reconstructed decay-time and t_{true} corresponds to the true decay-time generated in the simulation sample. The resulting distribution is parametrized by a triple Gaussian given by Eq. 4. The results are collected in Tab. 5 and presented in Fig. 5. The average of the decay-time resolution is in good agreement with zero, and no influence of changing the primary vertex reconstruction is found.

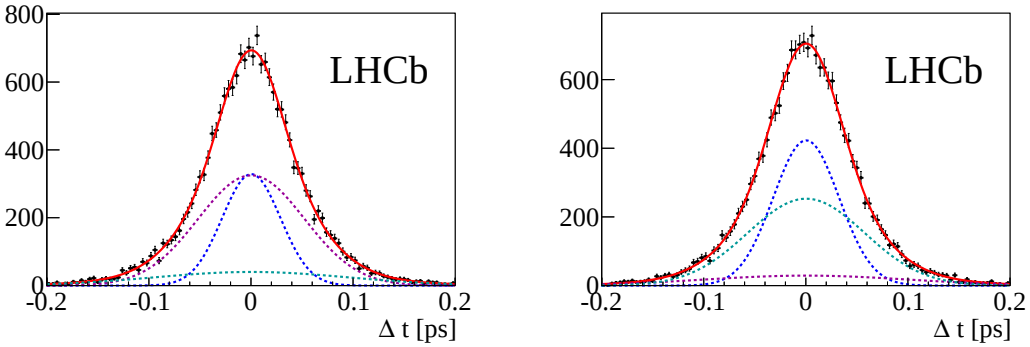

Figure 5: Triple Gaussian time resolution fitted to the $B_s^0 \to \phi^0\phi^0$ simulation sample. The solid curve corresponds to the full fit, while the dashed curves are components. Left: online 2015 reconstruction, Right: offline reconstruction.

Table 5: The fit parameters for the decay-time resolution.

Parameters	online 2015	offline
μ [ps]	0.00031 ± 0.00035	0.00033 ± 0.00035
f_1	0.23 ± 0.18	0.31 ± 0.10
f_2	0.64 ± 0.09	0.82 ± 0.13
σ_1 [ps]	0.028 ± 0.005	0.028 ± 0.003
σ_2 [ps]	0.044 ± 0.008	0.052 ± 0.006
σ_3 [ps]	0.077 ± 0.007	0.095 ± 0.025
$\bar{\sigma}^{tG}$ [ps]	0.0528 ± 0.007	0.0536 ± 0.009

3.1.5 Conclusion on inputs to the PV algorithm

Based on presented studies it has been decided that the new, nominal PV algorithm for Run II will use only VELO Kalman fitted tracks. The same algorithm is used online and offline, which results in offline quality PV finding in real-time in the trigger.

4 Optimization

With the inputs determined, the reconstruction algorithm is optimized in order to maximize its efficiency and minimise the fake rate. This optimization is performed using the same $b\bar{b}$-inclusive sample as earlier [2]. The PV algorithm used in these studies corresponds to the online 2015 approach.

The reconstruction efficiency is defined as the ratio of the number of reconstructed PVs to the number of reconstructible generated PVs. The latter is a pp interaction vertex which contains at least a minimal number of reconstructed VELO tracks. In the presented studies the minimal number of reconstructed tracks ($\min_{Tracks}$) varies between 3-5 and is used in the optimization procedure.

Reconstructed PVs can be split into two categories: vertices which are matched by any generated PV (so-called true PV) and the remaining, not matched vertices (so-called false PV). The PV is matched when the distance to any reconstructible PV is $|\Delta z| < 5\sigma_z$, where σ_z denotes the estimated position uncertainty along z axis. The fake rate is defined as the ratio of the number of false PVs and all reconstructible generated PVs.

The optimization is considered in a two dimensional space:

- minimum number of tracks in vertex: $\min_{Tracks} \in [3, 4, 5]$,
- maximum χ^2 to accept track: $\max_{\chi^2} \in [0.1, 40.0]$.

In the past (2010-2012 data taking) the nominal setting were: $\min_{Tracks} = 5$ and $\max_{\chi^2} = 9.0$. The efficiency and fake rate as a function of the above variables are shown in Fig. 6. For $\min_{Tracks} = 4$ the efficiency increases by about 3% with respect to the nominal conditions, which can be considered as the new optimal value. At the same time, the fake rate rises twice. To prevent the growing fake rate, a radial distance cut is used, as described later. An additional 1.5% in efficiency can be gained for $\min_{Tracks} = 3$, however in this condition the fake rate increases three times with respect to the nominal conditions, thus $\min_{Tracks} = 3$ is excluded from further studies. A more detailed optimisation is then considered in the narrower $\max_{\chi^2}$ range $[6.0, 12.0]$ and for $\min_{Tracks} \in [4, 5]$.

4.1 Radial distance optimization

In order to reduce the fake rate several requirements were considered:

- radial distance of PV: $r = \sqrt{x^2 + y^2}$, where (x,y) is PV's position,
- minimal radial distance of track: $\min RD_{Track} = \min(\sqrt{x^2 + y^2})$ where (x,y) is the track position closest to the beamline for tracks coming from the PV,
- maximal radial distance of track: $\max RD_{Track} = \max(\sqrt{x^2 + y^2})$ where (x,y) is the track position closest to the beamline for tracks coming from the PV.

[2] As a cross-check studies have been performed for the $c\bar{c}$-inclusive simulation sample, and the obtained results are in good agreement

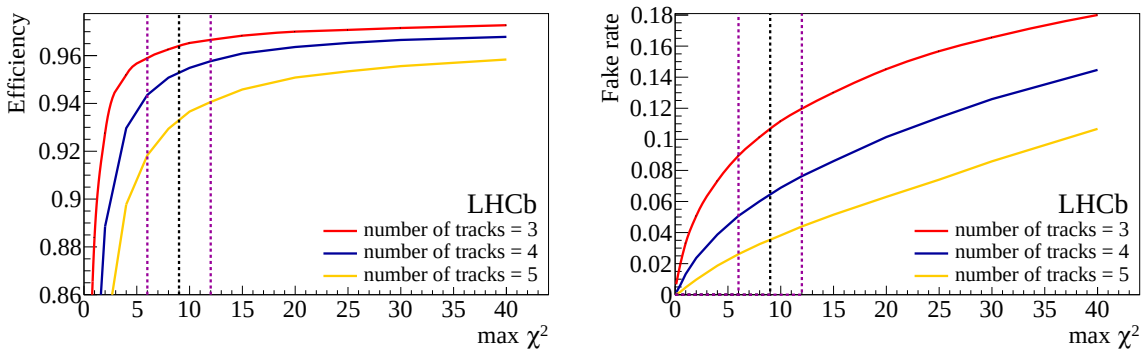

Figure 6: Efficiency (left) and fake rate (right) as a function of the minimal track multiplicity and maximal χ^2 to accept the track. The default value: $\max_{\chi^2} = 9.0$ is denoted as a black dashed line. The red, blue and orange curves correspond to $\min_{Tracks} \in [3, 4, 5]$, respectively. The violet vertical lines correspond to the most interesting range which is considered in the further studies.

Fig. 7 shows the efficiency versus fake rate rejection, so-called ROC curve, for radial variables. As can be seen, the radial distance of the PV is the most powerful variable to reduce the fake rate, and it is therefore optimized further.

The radial distance distribution for false and true PVs is shown in Fig. 8. The distributions for PVs reconstructed using $\min_{Tracks} = 4$ and $\max_{\chi^2} \in [6, 9, 12]$ are in good agreement. In addition, the radial distance for most true PVs is below 0.2 mm, while for false PVs it extends until 8 mm.

The optimization of the radial distance maximizes the purity multiplied by the true PV efficiency, where the purity is defined as a ratio of the number of true PVs to all reconstructed PVs found in the event. The figure of merit is maximized for $\min_{Tracks} = 4$ and $\max_{\chi^2} \in [6, 9, 12]$. The optimal working point for the radial distance criterion is found to be $r < 0.2$ mm. independently of the maximal χ^2 to accept the track. Figure 9 shows the optimization function with all components, as an example, for $\min_{Tracks} = 4$ and $\max_{\chi^2} = 9$.

4.2 Final optimization

Figure 10 shows the influence of applying the radial distance cut $r < 0.2$ to the reconstruction efficiency and fake rate for the following algorithm settings: $\min_{Tracks} = [4, 5]$ and $\max_{\chi^2} \in [6, 12]$. The radial cut is 99.5% efficient, while the fake rate is reduced by 85%.

In addition, the reconstruction using $\max_{\chi^2} = 12$ allows more tracks to form a PV and thus provides better resolution for PVs with low track multiplicity. Table 6 presents the resolution and for two reconstructions: $\max_{\chi^2} = 9$ and $\max_{\chi^2} = 12$. The additional tracks don't bias the vertex position. Therefore the optimal working point is set to be $\max_{\chi^2} = 12$.

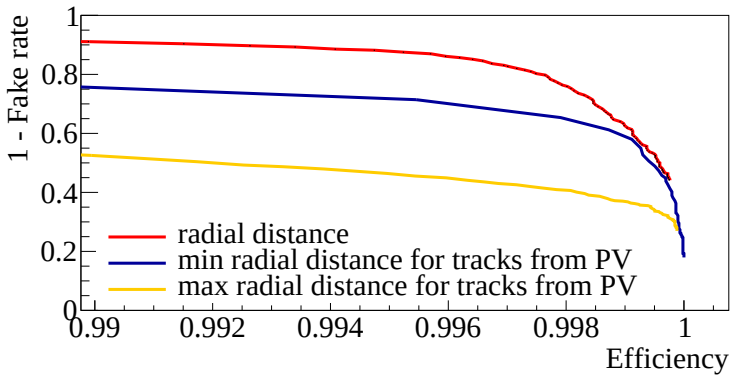

Figure 7: Efficiency versus fake rate rejection (ROC curve) for radial variables.

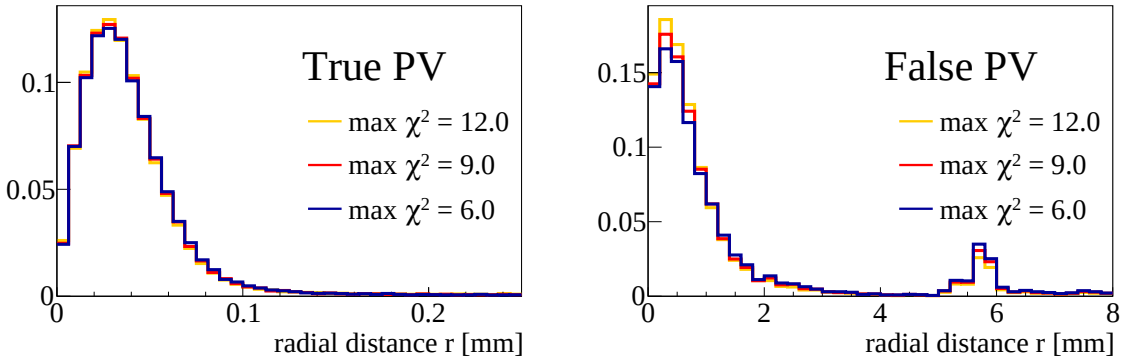

Figure 8: Radial distance distributions for true PVs (left) and fake PVs (right) for the reconstruction with $\min_{Tracks} = 4$.

Table 6: Global resolution for PVs reconstructed with $\min_{Tracks} = 4$ and $\max_{\chi^2} = [9, 12]$.

Parameter [mm]	$max_{\chi^2} = 9$	$max_{\chi^2} = 12$
μ_x	2.88×10^{-5}	0.05×10^{-5}
Δx	0.0199	0.0195
μ_z	0.0016	0.0018
Δz	0.0140	0.0136

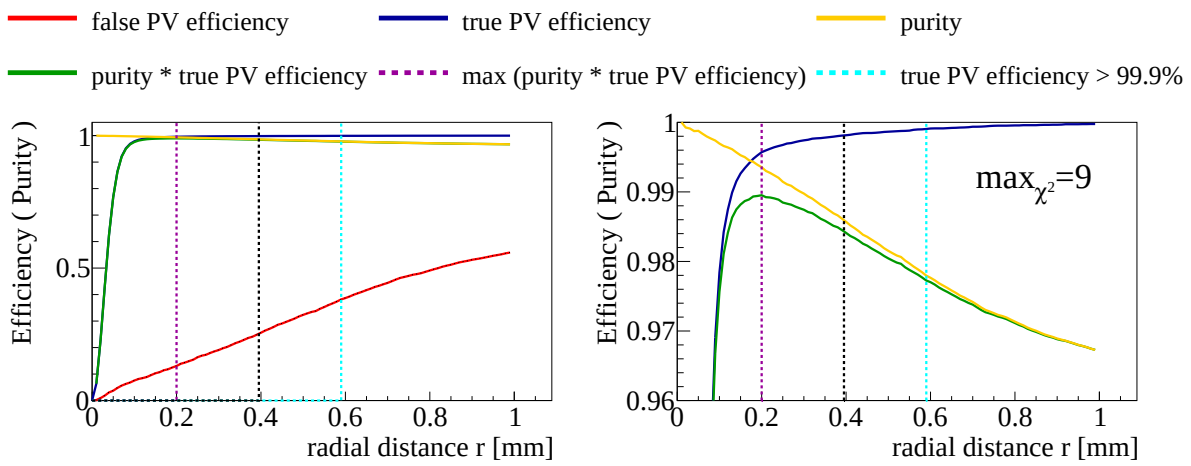

Figure 9: Optimization of the radial distance cut for $\min_{Tracks} = 4$ and $\max_{\chi^2} = 9$ in the (left) full and (right) zoomed range. Different contributions are shown as coloured lines as described in the legend placed in the top.

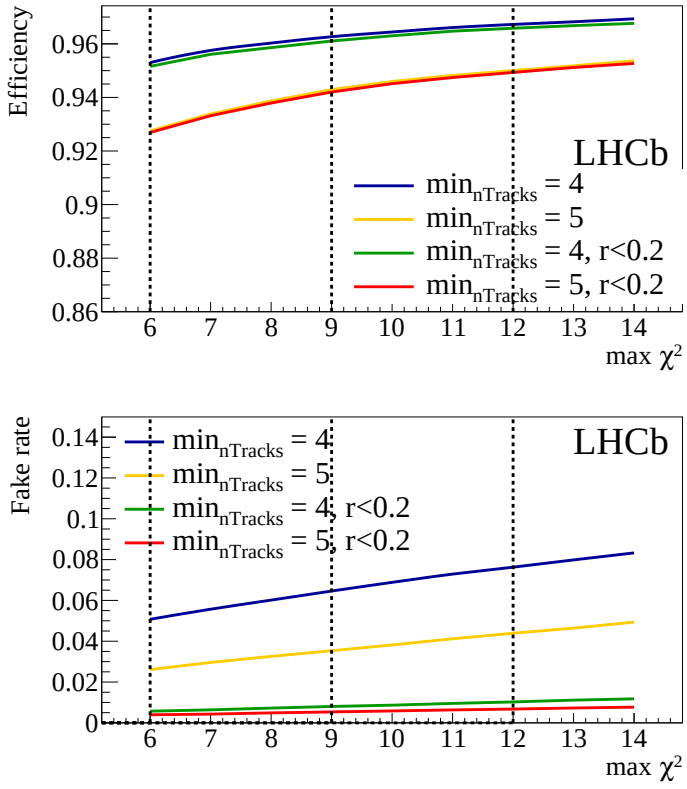

Figure 10: Efficiency and fake rate as a function of minimal track multiplicty and maximal χ^2 to accept the track before and after applying the radial distance cut $r < 0.2$.

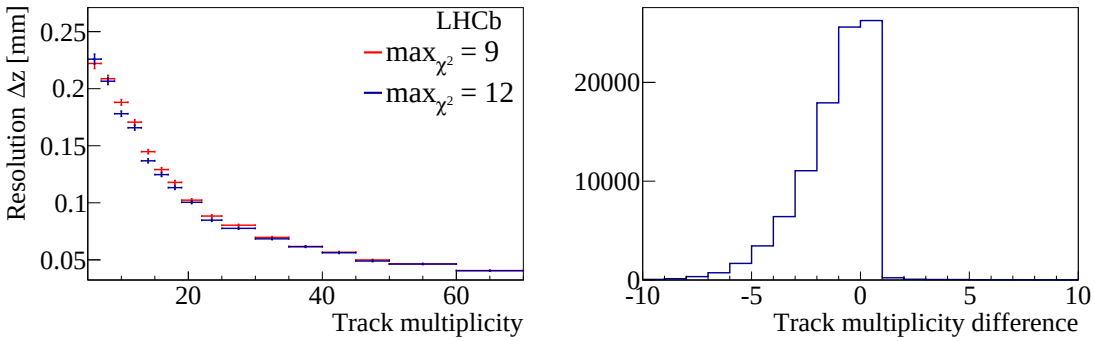

Figure 11: Left: Resolution as a function of track multiplicity for reconstruction with $\min_{Tracks} = 4$ and $\max_{\chi^2} = [9, 12]$. Right: Difference between track multiplicity in the PV for the reconstruction using $\max_{\chi^2} = 9$ and $\max_{\chi^2} = 12$.

4.3 Comparison

The true PVs can be split into several subgroups depending on their properties such as:

- distance to the closest other reconstructible PVs:
  - Isolated - with distance to closest reconstructible PV $|\Delta z| > 10$ mm,
  - Non-Isolated - with distance to closest reconstructible PV $|\Delta z| < 10$ mm,
- multiplicity of tracks:
  - 1st - with the greatest track multiplicity,
  - 2nd - with 2nd greatest track multiplicity,
  - 3rd - with 3rd greatest track multiplicity,
- depending on daughter particles:
  - beauty - with a beauty hadron coming from it,
  - charm - with a charm hadron coming from it,
  - other.

In the same way, false PVs can be categorized depending on the misidentification into:

- charm - when the reconstructed PV is a charm decay vertex,
- beauty - when the reconstructed PV is a beauty decay vertex,
- other.

Table 7 presents the efficiency comparison for all PV categories between the nominal ($\min_{Tracks} = 5$, $\max_{\chi^2} = 9.0$) and optimal ($\min_{Tracks} = 4$, $\max_{\chi^2} = 12.0$,$r < 0.2$) conditions for the online 2015 algorithm. Similar results for the fake rate are collected in Tab. 8. A substantial gain is visible for each category of true PVs, in particular for Non-Isolated (3%), 3rd (3%) and other (4%). The global efficiency is increased by 2.7%. At the same time the fake rate reduces from 3.8% to 1.0%, with the most substantial reduction for beauty PV misidentification, from 2.7% to 0.9%.

The reconstruction efficiency depends on the track multiplicity associated to the vertex, which varies with the data samples used. Moreover, a non-negligible effect occurs due to different numbers of *pp* interactions. In the case of multiple *pp* interactions the efficiency of close vertices is degraded; the lower multiplicity PV are affected due to track migration from the low multiplicity PVs to high multiplicity one. The efficiency of all true PVs as a function of track multiplicity is shown in Fig. 12 for both the optimal and nominal conditions of the online 2015 algorithm. A substantial gain with respect to nominal conditions is observed for low track multiplicity.

Table 7: Comparison of the reconstruction efficiency for nominal ($\min_{Tracks} = 5$, $\max_{\chi^2} = 9.0$) and optimal ($\min_{Tracks} = 4$, $\max_{\chi^2} = 12.0$, $r < 0.2$ mm) conditions for the online 2015 algorithm split by true PVs categories.

	Nominal	Optimal
true PV category	Efficiency [%]	
All	94.1	96.8
Isolated	95.8	97.7
Non-Isolated	86.0	89.0
1st	95.9	97.5
2nd	93.2	95.7
3rd	92.1	95.1
charm	98.3	98.9
beauty	98.8	99.2
other	88.3	92.8

Table 8: Comparison of the fake rate for nominal ($\min_{Tracks} = 5$, $\max_{\chi^2} = 9.0$) and optimal ($\min_{Tracks} = 4$, $\max_{\chi^2} = 12.0$,$r < 0.2$) conditions for the online 2015 algorithm split by false PV categories.

	Nominal	Optimal
false PV category	Fake rate [%]	
All	3.8	1.1
charm	0.2	0.1
beauty	2.7	0.9
other	0.9	0.1

5 PV resolution on data

The PV resolution can be computed in a data-driven manner by comparing two independent positions of the PV in the same event. The VELO tracks found in the event are split randomly into two samples, and the PV reconstruction is performed in both subsamples. The resolution is defined as the difference between the two sets of reconstructed PVs. The width of this difference is given by the RMS of the distribution and corrected by a factor of $1/\sqrt{2}$, which corresponds to the PV resolution for the full set of tracks. The quantity is strongly correlated with the number of tracks associated to the vertex ($nTracks$) and can be described using the function:

$$\sigma(nTracks) = \frac{A}{(nTracks)^B} + C, \tag{10}$$

where A, B, C denote free parameters. In principle, the resolution for the infinite track multiplicity is given by parameter C, while parameter A corresponds to the resolution for low track multiplicity.

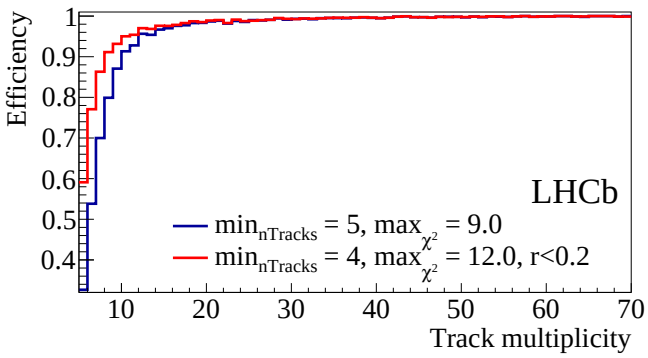

Figure 12: Reconstruction efficiency as a function of the track multiplicity for nominal ($\min_{Tracks} = 5$, $max_{\chi^2} = 9.0$) and optimal ($\min_{Tracks} = 4$, $\max_{\chi^2} = 12.0$, $r < 0.2$ conditions and all reconstructed PVs.

The PV resolution obtained on 2015 data is shown in Fig. 13 together with the fit result. The dashed orange and red curve indicate the PV resolution found for 2011 and 2012 data, respectively. We found a similar dependence of the resolution on the track's multiplicity in 2011, 2012, and 2015 for the **x** coordinate. For the **z** coordinate the 2015, 2012, and 2011 resolutions are the same for very low and very high track multiplicities, while for middle multiplicities the 2015 reconstruction performs better. To illustrate this, the comparison of the resolution for PVs in the bin corresponding to 25 tracks in collected is Tab. 9. The 2015 PV resolution is better by about 7% for the **x** coordinate and 20% for the **z** coordinate.

In addition, we checked the agreement of PV resolution between data and MC, which is shown in Fig. 14, where the MC resolution is measured using truth information. There is a very good agreement between the data and MC resolutions.

Table 9: The PV resolution for 25 tracks.

	$\sigma_x(25)$	$\sigma_z(25)$
data 2011 ($\sqrt{s} = 7$ TeV)	0.0139 ± 0.002	0.089 ± 0.009
data 2012 ($\sqrt{s} = 8$ TeV)	0.0135 ± 0.001	0.093 ± 0.006
data 2015 ($\sqrt{s} = 13$ TeV)	0.0129 ± 0.001	0.073 ± 0.001
simulation 2015 ($\sqrt{s} = 13$ TeV)	0.0127 ± 0.001	0.073 ± 0.001

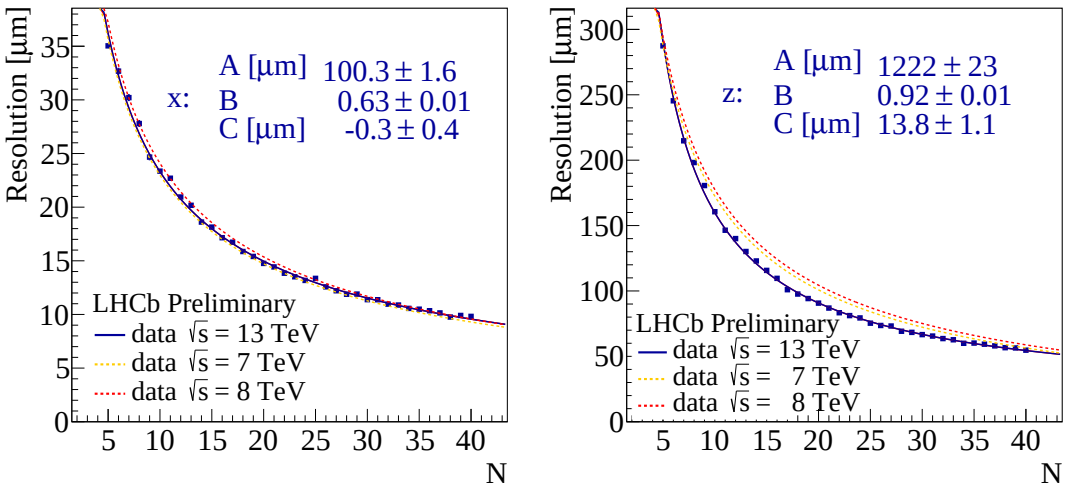

Figure 13: The PV resolution measured on data. The solid blue line corresponds to 2015 data taking, while dashed orange and red correspond to 2011 and 2012, respectively. The resolution is parameterized by Eq. 10, and the result of the fit is indicated in the plot.

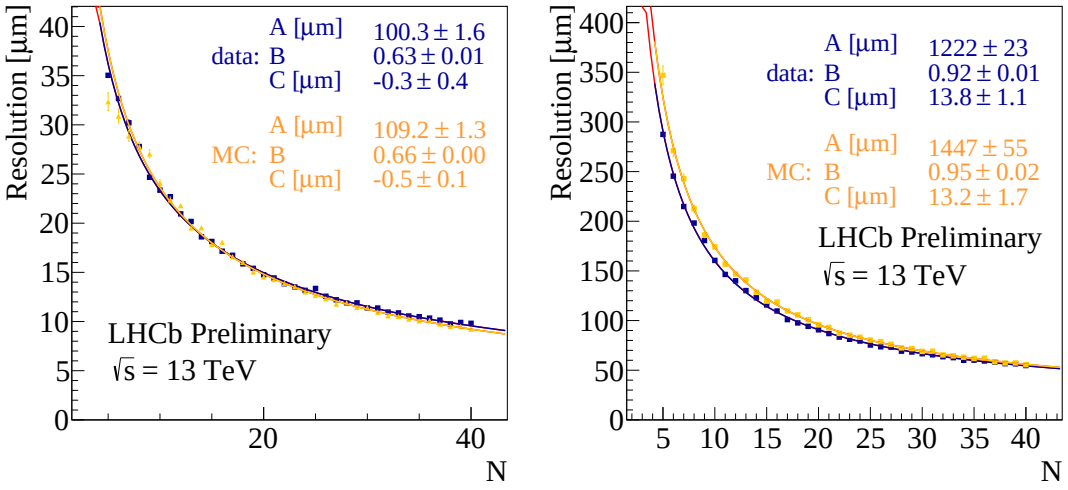

Figure 14: Comparison between the PV resolution measured on data and MC. The solid blue line corresponds to 2015 data taking, while orange corresponds to the MC. The resolution is parameterized by Eq. 10, and the result of the fit is indicated in the plot.

6 Double threshold for 2016 data onwards

When stable beams are declared the VELO is closed and centered around the collision region. This region is determined by reconstructing the PV position using only tracks in one or the other half of the VELO and comparing the results. After the closing procedure, the beam position should agree with the resolvers at the level of <50 μm and it should be centered around the beam within 200 μm. The radial distance described in Sec. 4.1 is calculated with respect to the resolver position of the VELO, and thus a small deviation with respect to the collision region can be introduced.

Figure 15 shows the track multiplicity as a function of radial distance in case of a perfect centering of the VELO (left) and in the case where some deviations are introduced (right). Since the nominal cut value for the radial distance is $r < 0.2$ mm, this kind of deviation can affect the efficiency of the PV reconstruction.

While a problem in the closing and centering procedure is very rare, it did happen during 2015 datataking. Therefore, to make the PV reconstruction insensitive to it, for 2016 data taking the radial distance has been changed to $r < 0.2$ mm for PVs with fewer then 10 tracks, and $r < 0.4$ mm for other PVs. The relative efficiency of the PV algorithm as function of the shift in both the **x** and **y** coordinates ($x_{shift} = [-0.25, 0.25]$ mm, $y_{shift} = [-0.25, 0.25]$ mm) is shown in Fig. 16. The bottom plots indicate the regions with the relative efficiency greater than 99.9% (red), 99%(orange), 98%(green), 97%(cyan). We see that the PV finding with a double radial threshold is less sensitive to problems with the VELO centering, with a minimal efficiency of 80%. The maximal radial shift which gives an efficiency greater than 99.9% is found to be $r_{shift} < 0.07$ mm for the PV reconstruction with a double radial threshold. This maximal radial shift is used in order to define an alarm threshold in the online PV monitoring.

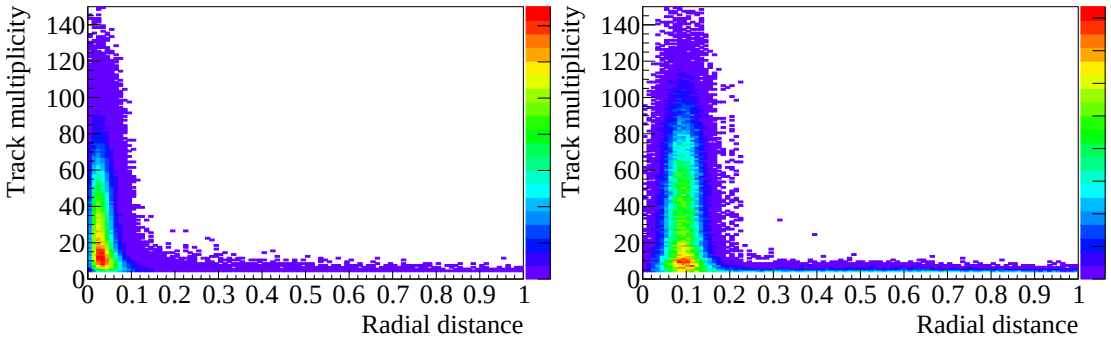

Figure 15: Track multiplicity as function of radial distance for (left) perfect centering of the VELO and (right) imperefect centering of the VELO.

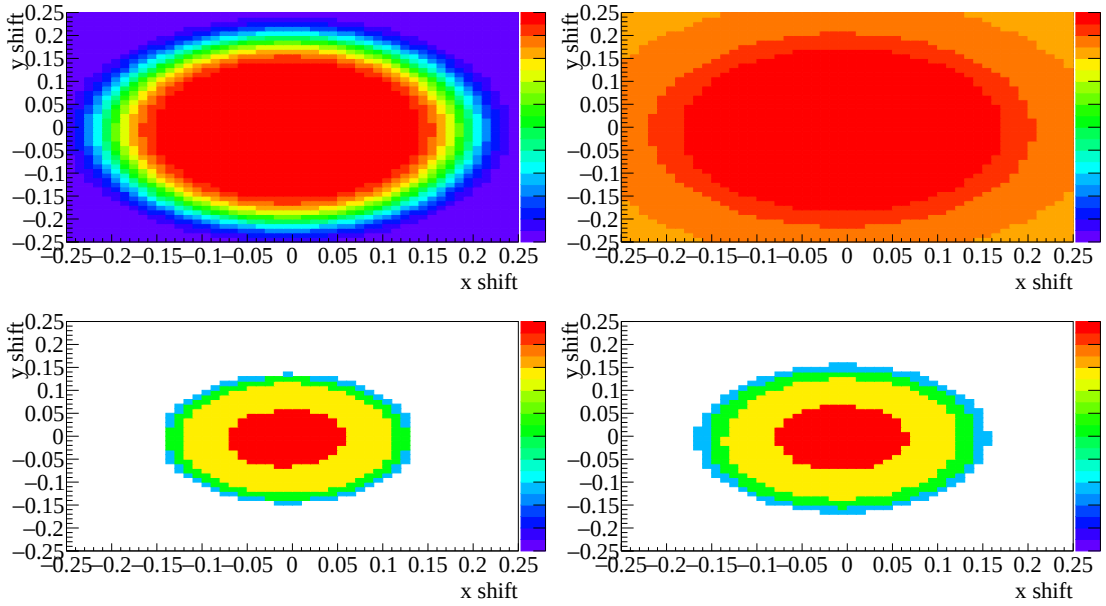

Figure 16: Relative efficiency as function of the shift in both the **x** and **y** coordinates. The top plots show the entire range of the relative efficiency from 0% (blue) to 100% (red). The bottom plots highlight the regions with the relative efficiency grather than 99.9% (red), 99%(orange), 98%(green), 97%(cyan).

7 Conclusion

Detailed studies of the primary vertex reconstruction have been performed. It is shown that using Kalman fitted VELO tracks allows the same algorithm to be used online and offline with no loss in performance. In addition, the algorithm has been reoptimized in order to maximize the reconstruction efficiency and minimise the fake rate. The new working point is found to be $\min_{Tracks} = 4$, $\max_{\chi^2} = 12.0$, $r < 0.2$ mm with a 3% gain for efficiency and ∼70% relative fake rate reduction. For 2016 data taking the radial distance cut has been relaxed to be less $r < 0.2$ mm for PVs with less then 10 tracks associated to it, and $r < 0.4$ for other PVs. This new approach is less senstive to any deviations which occur due to imperfect VELO centering.

References

[1] J. W. Tukey, *Approximate weights*, Ann. Math. Statist. **19** (1948) 91.

[2] M. Krasowski *et al.*, *Primary vertex reconstruction*, Tech. Rep. LHCb-2007-011. CERN-LHCb-2007-011, CERN, Geneva, Sep, 2007.

[3] M. Kucharczyk, P. Morawski, and M. Witek, *Updated Primary Vertex Reconstruction*, Tech. Rep. LHCb-INT-2012-006. CERN-LHCb-INT-2012-006, CERN, Geneva, Feb, 2012.

[4] M. Kucharczyk, P. Morawski, and M. Witek, *Primary Vertex Reconstruction at LHCb*, Tech. Rep. LHCb-PUB-2014-044. CERN-LHCb-PUB-2014-044, CERN, Geneva, Sep, 2014.

[5] R. Kalman, *A new approach to linear filtering and prediction problems*, Journal of Basic Engineering **82** (1960) 35.