UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea in Fisica

Tesi di Laurea

Reconstruction of $D_s^{\pm}$ Mesons with the CMS Experiment

Relatore

Prof. Paolo Ronchese

Laureando

Enrico Lupi

Anno Accademico 2021/2022

Table of Contents

Introduction

The CMS experiment at the LHC [7] observes products of proton-proton collisions at an energy in the center of mass system $\sqrt{s} = 13\,TeV$. In these collisions a large number of particles are produced: among them, B mesons, composed of a b (beauty) quark and a light quark, and particularly B_s^0, whose second quark is a s (strange) quark, are especially interesting.

For neutral B mesons, B_d^0 and B_s^0, mass eigenstates are mixtures of flavour eigenstates, leading to the phenomenon of particle-antiparticle oscillations, so that a meson created as B_s^0 can decay as a $\bar{B}_s^0$. When mass eigenstates have equal contents of the flavour states CP symmetry is conserved in the mixing, otherwise a violation is observed; a violation could also appear in the decay. When studying CP violation in B_s^0 decay, the capability to determine the flavor (B = ± 1) of the meson at its production and/or decay time is thus essential.

In the decay of B_s^0 and $\bar{B}_s^0$ mesons, $D_s^{\pm}$ mesons are often produced and their charge allows to infer the flavor (B = ± 1) of the decaying B_s^0. $D_s^{\pm}$ mesons are composed of a quark c (charm) and a quark s and decay in turn: among all the possibilities, the one with two opposite charged kaons ($K^+ K^-$) and one charged pion ($\pi^{\pm}$) is one of the easiest to reconstruct exclusively.

In this work a preliminary study is done, regarding the reconstruction of $D_s^{\pm}$ mesons decay in the $\phi\,\pi^{\pm}$ channel, followed by the cascade decay $\phi \rightarrow K^+ K^-$.

Due to LHC's high luminosity very large combinatorial background noise is expected: this would not only be an unacceptable nuisance when doing physics studies using that decay, but also an enormous waste of computational power, as the estimation of the decay point from the reconstructed particle trajectories is a CPU intensive operation, and its application to all $K^+ K^- \pi^{\pm}$ combinations would lead to a process time explosion. The aim of this work is therefore to find useful criteria to implement a preselection in order to identify candidates to use for a full reconstruction.

1. The CMS Experiment

The Compact Muon Solenoid (CMS) experiment [7] is a general purpose particle detector built on the Large Hadron Collider (LHC) at CERN; it studies proton-proton collisions that, during Run 2 (2016-2018), reached a center-of-mass energy of 13 TeV.

It has a cylindrical symmetry around the direction of the proton beam. In cartesian coordinates, using the nominal collision point as the origin, this axis is usually identified as the z-axis, while the y-axis is defined as the vertical direction pointing upwards and the x-axis as the direction pointing towards the center of the accelerator ring. In order to describe particle momenta, though, it is usually preferred to use a polar coordinate system with the (p_T, ϕ, η) tern; p_T is the absolute value of the projection of the total momentum on the xy-plane, ϕ is the azimuthal angle, or the angle between the momentum and the x-axis in the xy-plane, and η is the pseudorapidity, defined as a function of the polar angle θ relative to the z-axis: $\eta = -\log(\tan(\theta/2))$.

The CMS detector is comprised of different layers, as shown in fig. 1.1; going from the innermost to the outermost part we find:

- the tracker

- the calorimeters

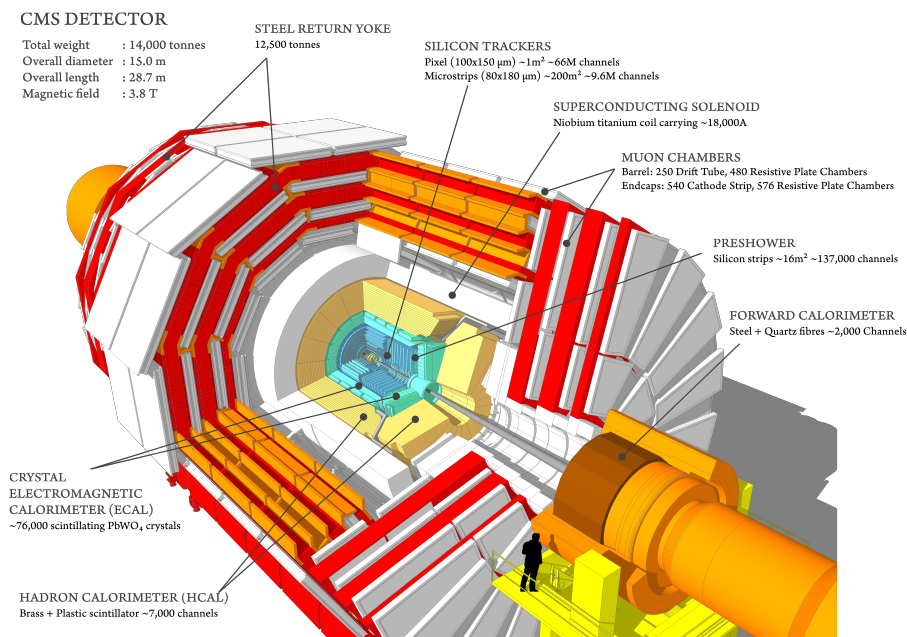- the magnet coil

- the muon detectors

Figure 1.1: Exploded view of the CMS.

1.1. The Tracker

The tracker layer [4] is tasked with accurately measuring the position of particles that traverse it in order to reconstruct their momenta and the position of primary and secondary vertexes.

The CMS tracker is made of two parts. At the core of the detector ($r < 20\,cm$), where the particle flux is at its highest, silicon pixel detectors are placed in three barrel layers with two endcap disks on each side, as shown in fig 1.2; these pixels reach a spatial resolution of about $10\,\mu m$ for the r-ϕ measurement and about $20\,\mu m$ for the z measurement. On the outside ($r > 20\,cm$), the low flux makes it possible to use silicon microstrips instead.

The tracker has been designed in order to guarantee a high resolution for particles with pseudorapidity $|\eta| < 2.4$.
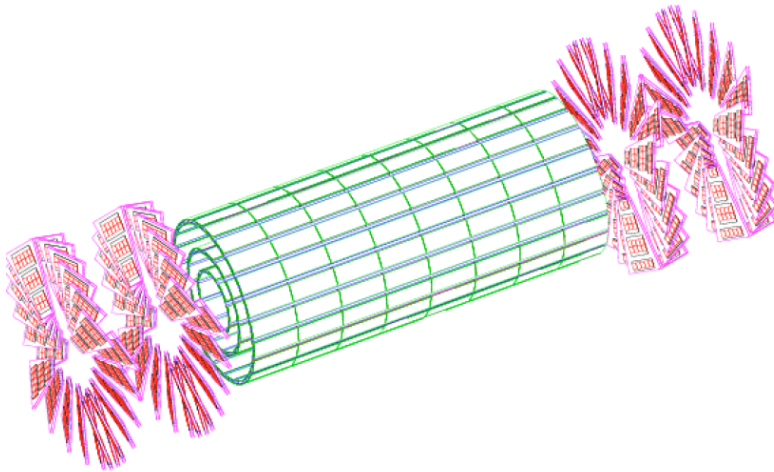

Figure 1.2: Pixel tracker layout.

1.2. The Calorimeters

The CMS detector has two different concentric calorimeters: the inner one is electromagnetic (ECAL) while the outer is hadronic (HCAL).

The ECAL [5], shown in fig. 1.3, is a homogeneous calorimeter whose main goal is to absorb electrons and photons in order to measure their energy. It is composed of lead tungstate $PbWO_4$ crystals mounted in a central barrel part with two endcaps; these crystals have short radiation length, emit light fast and are radiation hard. As photodetectors, it uses silicon avalanche photodiodes in the barrel and vacuum phototriodes in the endcaps.

The HCAL [6] is divided in four parts. The hadron barrel (HB) and hadron endcap (HE), covering the pseudorapidity regions $|\eta| < 1.4$ and $1.4 < |\eta| < 3.0$ respectively, are located between the ECAL and the magnet coil and use plastic scintillators as the active medium and brass as the absorber, due to the short interaction length and non-magnetic properties; the hadron forward (HF), covering the region $3.0 < |\eta| < 5.0$, uses quartz fibers to collect Cherenkov light; lastly, the hadron outer (HO), for $|\eta| < 1.26$, lies outside the coil and samples the energy from penetrating hadron showers leaking through the rear of the calorimeters.
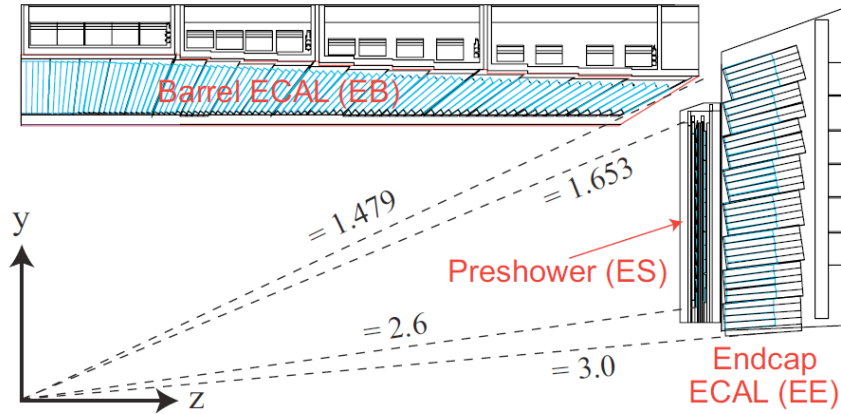
Figure 1.3: Transverse section through the ECAL.

1.3. The Magnet Coil

CMS uses a 13-m-long, $5.9\,m$ inner diameter, superconducting niobium-titanium solenoid, capable of creating a $4\,T$ magnetic field: such a strong magnetic field is necessary in order to achieve good momentum resolution within a compact spectrometer without making stringent demands on the muon-chamber.

1.4. The Muon Detectors

Depending on the region and the radiation environment, three types of gaseous detectors are used to identify and measure muons [9], as shown in fig. 1.4.

In the Barrel Detector, that consists of four concentric stations inside the magnet return yoke of CMS ($|\eta| < 1.2$), drift tube (DT) chambers are used as the neutron induced background is small, the muon rate is low and the residual magnetic field in the chambers is low. In the 2 endcaps, where the muon rate, the neutron induced background and the magnetic field get higher, it is preferred to deploy cathode strip chambers (CSC); these cover the region up to $|\eta| < 2.4$. Finally, in both regions resistive plate chambers (RPC) are present: they cover the region $|\eta| < 1.6$ and provide a fast response with good time resolution (albeit with a coarser position resolution than DTs or CSCs).
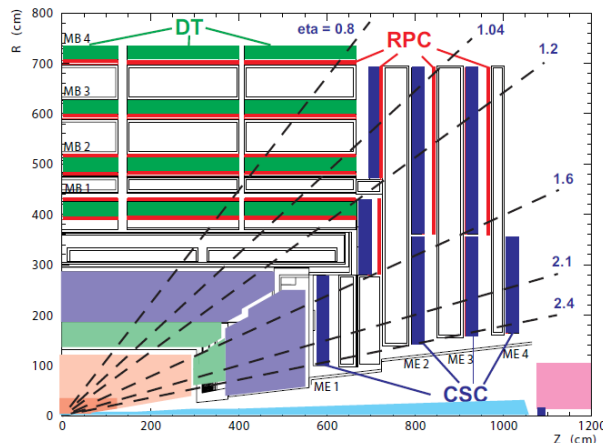

Figure 1.4: Layout of one quarter of the CMS muon system.

4

1.5. Trigger System

In the LHC, at the achieved luminosity of $L \gtrsim 10^{34} \, cm^{-2} s^{-1}$, proton bunches cross every $25 \, ns$ and dozens of interactions occur at each bunch crossing: the amount of data produced is beyond the capacity of any archival system so a system to select only interesting events and reject the large background must be implemented. This task is achieved by the trigger system [8], divided in two steps.

First, we have the L1 trigger, a hardware system with fixed latency; it receives trigger primitives from the calorimeters and muon chambers and evaluates this information in a global trigger to decide whether to accept the event or not. This trigger lowers the event rate to $100 \, kHz$.

Secondly, there is the High Level Trigger (HLT), a software system that for each event reconstructs objects such as electrons, muons, and jets and then applies identification criteria in order to select only the most interesting ones for data analysis. It lowers the rate of stored events by three orders of magnitude, resulting in a frequency of $1 \, kHz$.

2. Data Samples

The goal of this section is to find suitable $D_s^\pm$ candidates for signal and background events. The preliminary work on the preselection of interesting events has been done on data containing B mesons decays, created from a Monte Carlo simulation reproducing the 2018 LHC run conditions. Events have been simulated using PYTHIA [12], B decays with EVTGEN [11] and the detector response with GEANT [1] .

The data are organised in ROOT Ntuples [3] containing all the parameters of interest: the generation point of simulated particles, the 3-momenta $(p_x,\ p_y,\ p_z)$ of simulated particles and reconstructed tracks, impact parameters in the transverse xy-plane and along the z-axis, the position of reconstructed secondary vertices for several decays, including $\phi \to K^+K^-$, and lastly the invariant mass of particles associated to secondary vertices.

Secondary vertices of $\phi \to K^+K^-$ candidates have been reconstructed considering all track pairs of particles having an opposite charge, having an invariant mass M_{inv}^ϕ inside a $\pm 150\,MeV/c^2$ range around the ϕ mass $M^\phi = 1020\,MeV/c^2$ [13] (in the hypothesis that they are kaons) and being compatible with coming from a common vertex, which was found applying a Kalman Filter algorithm [10].

2.1. Signal

First, a sample of simulated $D_s^\pm$ decaying in the $\phi\,\pi^\pm$ channel where the ϕ decays in turn as $\phi \to K^+K^-$ has been studied, and the simulated kaons and pions have been associated to their corresponding reconstructed tracks. For each simulated particle, tracks with the same charge of the particle in question and satisfying the condition $|\Delta z| < 1.5\,cm$ were considered, with Δz being the difference between the track impact parameter and the z-coordinate of the creation point of the simulated particle. Among all track candidates, the one that minimizes the squared difference of track and simulated particle 3-momenta

$$\Delta p^2 = \left(p_x^{trk} - p_x^{sim}\right)^2 + \left(p_y^{trk} - p_y^{sim}\right)^2 + \left(p_z^{trk} - p_z^{sim}\right)^2$$

was chosen. The distributions of Δp^2 of the chosen particle-track associations for kaon and pion candidates is shown in fig. 2.1; each association was accepted when $\Delta p^2 < 2 \times 10^{-3}\ (GeV/c)^2$, otherwise it was rejected.

The position of the $D_s^\pm$ decay vertex was then estimated taking the midpoint of the segment joining the pion track and the ϕ decay vertex; the $D_s^\pm$ 3-momentum was calculated as the sum of the three daughter particles' reconstructed momenta and the invariant mass $M_{inv}^{D_s}$ was computed from the same reconstructed momenta assuming the decay products are two kaons and a pion.

A total of 50,745 events satisfied all aforementioned criteria and were thus used in the following analysis: henceforth, they will be referred to as "simulated signal".

Afterwards, the matching between simulated $\phi \to K^+K^-$ and the reconstructed secondary vertices was checked. When the two daughter particles of the simulated $\phi \to K^+K^-$ decay were associated to the two tracks used to build a vertex the event was taken as matching; 139 events that satisfy this additional requirement have been selected. In the following, we will refer to the sample of matched events as "reconstructed signal".
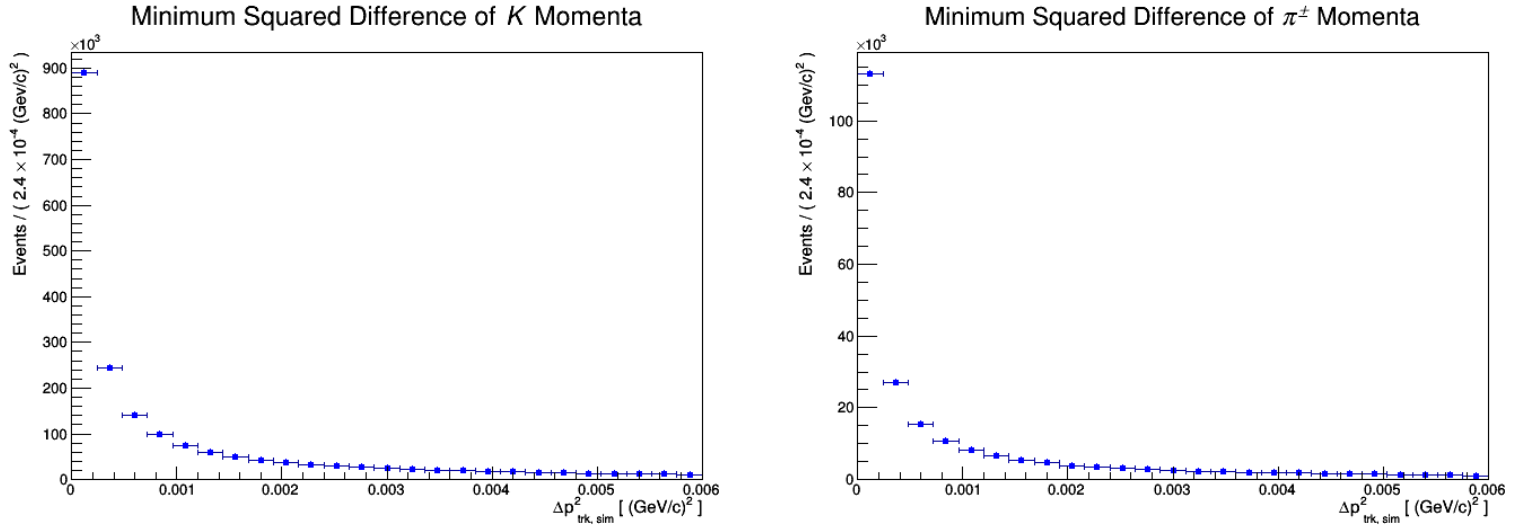
Figure 2.1: *Squared difference of track and simulated particle 3-momenta for kaons (left) and pions (right).*

Due to the very small number of reconstructed signal events, the "full" simulated sample was used in the following analysis. For the events in the simulated sample the $\phi \to K^+ K^-$ secondary vertex was of course not available, so the ϕ decay point was estimated by taking the midpoint of the segment joining the closest approach points in the two tracks associated to the kaons. The distance between the closest approach midpoint and the reconstructed vertex for the events in the reconstructed signal sample is shown in fig. 2.2.
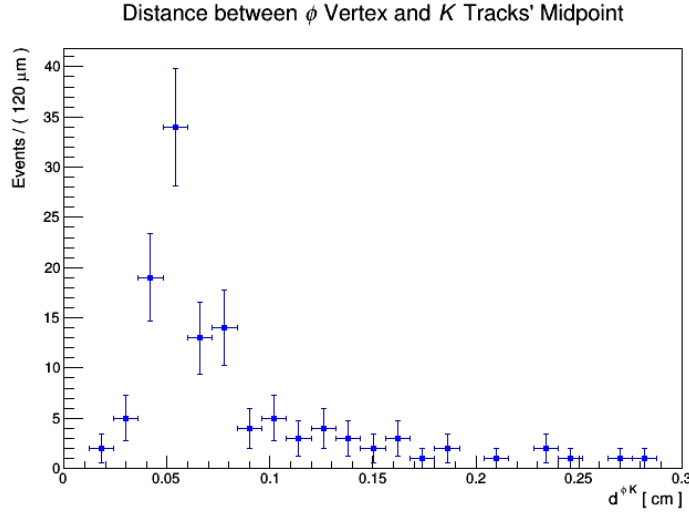

Figure 2.2: *Distance between reconstructed ϕ vertex and K tracks' closest approach midpoint.*

2.1.1 Simulated and Reconstructed Signal Events Compatibility

To check the compatibility between "simulated signal" and "reconstructed signal" events, the distributions of a set of quantities have been compared for the two samples. The following quantities have been considered:

- $d^{K^+ K^-}$, the distance between K tracks

7

- $d^{K\pi}$, the distance between the K tracks' midpoint and the pion track

- $d_{xy}^{K^+}$, the distance between the K^+ track and the z-axis

- $d_{xy}^{K^-}$, the distance between the K^- track and the z-axis

- p_{max}^{K}, the modulus of the momentum of the fastest K track

- p_{min}^{K}, the modulus of the momentum of the slowest K track

- p^{π}, the modulus of the momentum of the π track

- $\theta^{K\pi}$, the angle between the pion direction and the direction orthogonal to the plane containing the two kaon directions

$$\theta^{K\pi} = \frac{(\vec{p}_{K^+} \times \vec{p}_{K^-}) \cdot \vec{p}_\pi}{|\vec{p}_{K^+} \times \vec{p}_{K^-}||\vec{p}_\pi|}$$

- M_{inv}^{ϕ}, the invariant mass of the ϕ

- θ_{hel}^{π}, the pion helicity angle, i.e. the angle between $\vec{p}_\pi$ in the rest frame of $D_s^{\pm}$ and $\vec{p}_{D_s}$

- $d_{xy}^{D_s}$, the distance between the line identified by the position of the $D_s^{\pm}$ decay vertex and its momentum, and the z-axis

- $d_{fly}^{D_s}$, the transverse flight distance of $D_s^{\pm}$, calculated as the distance between the position of the $D_s^{\pm}$ decay vertex and the z-axis

- $M_{inv}^{D_s}$, the invariant mass of the $D_s^{\pm}$

For each variable, the distributions obtained from the two samples along with their ratio are shown in fig. 2.3, 2.4, 2.5.
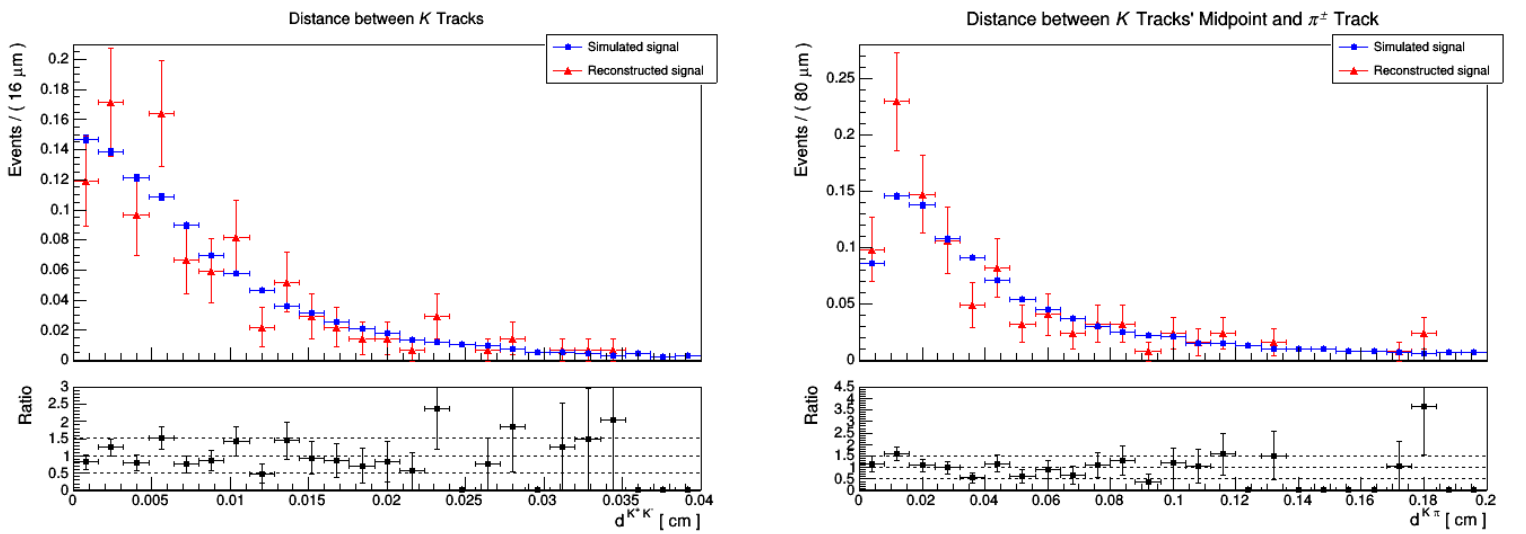

Figure 2.3: Comparison between simulated and reconstructed signal events for $d^{K^+K^-}$ and $d^{K\pi}$.
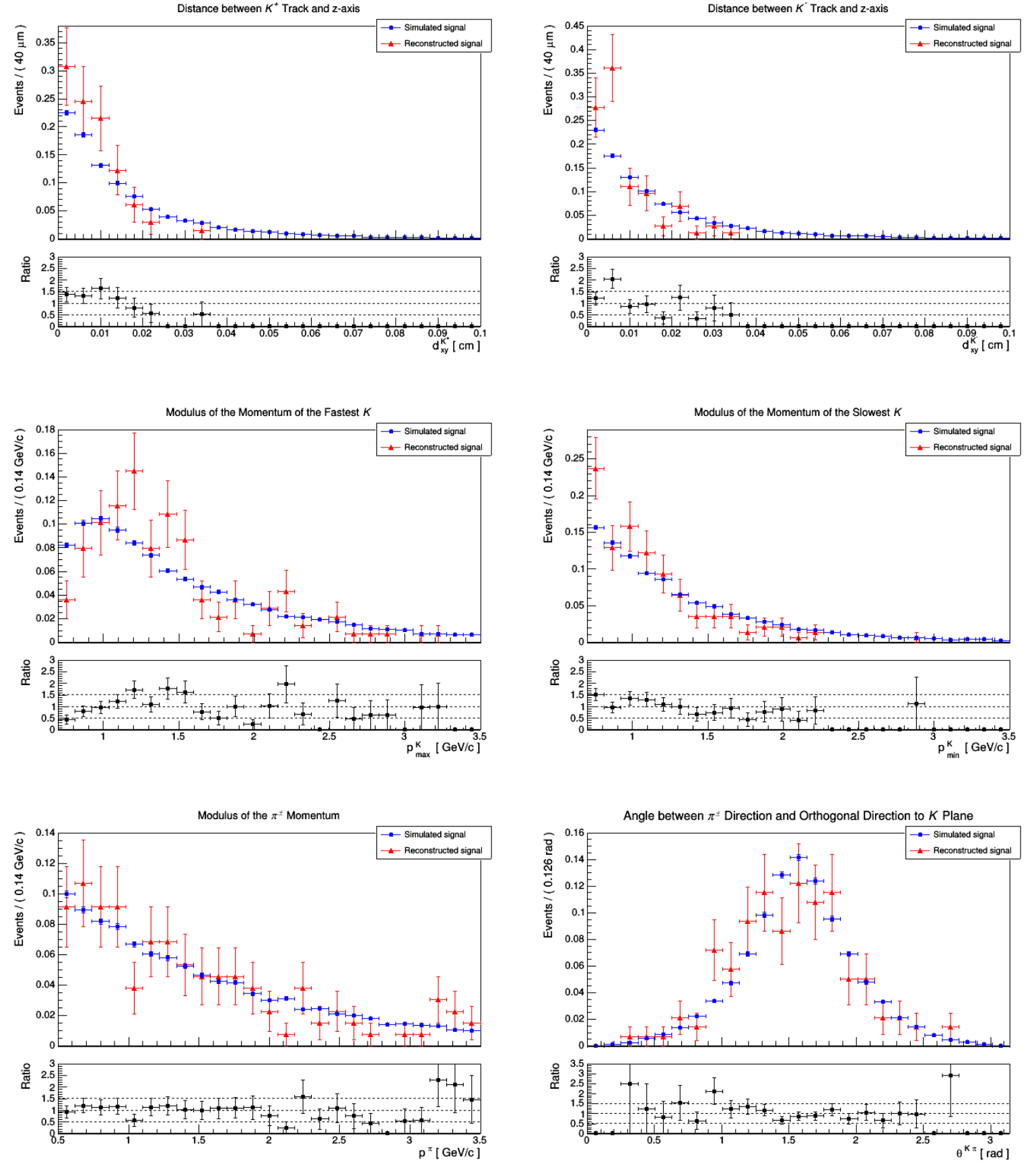
Figure 2.4: Comparison between simulated and reconstructed signal events for $d_{xy}^{K^+}$, $d_{xy}^{K^-}$, p_{max}^K, p_{min}^K, p^π and $\theta^{K\pi}$.
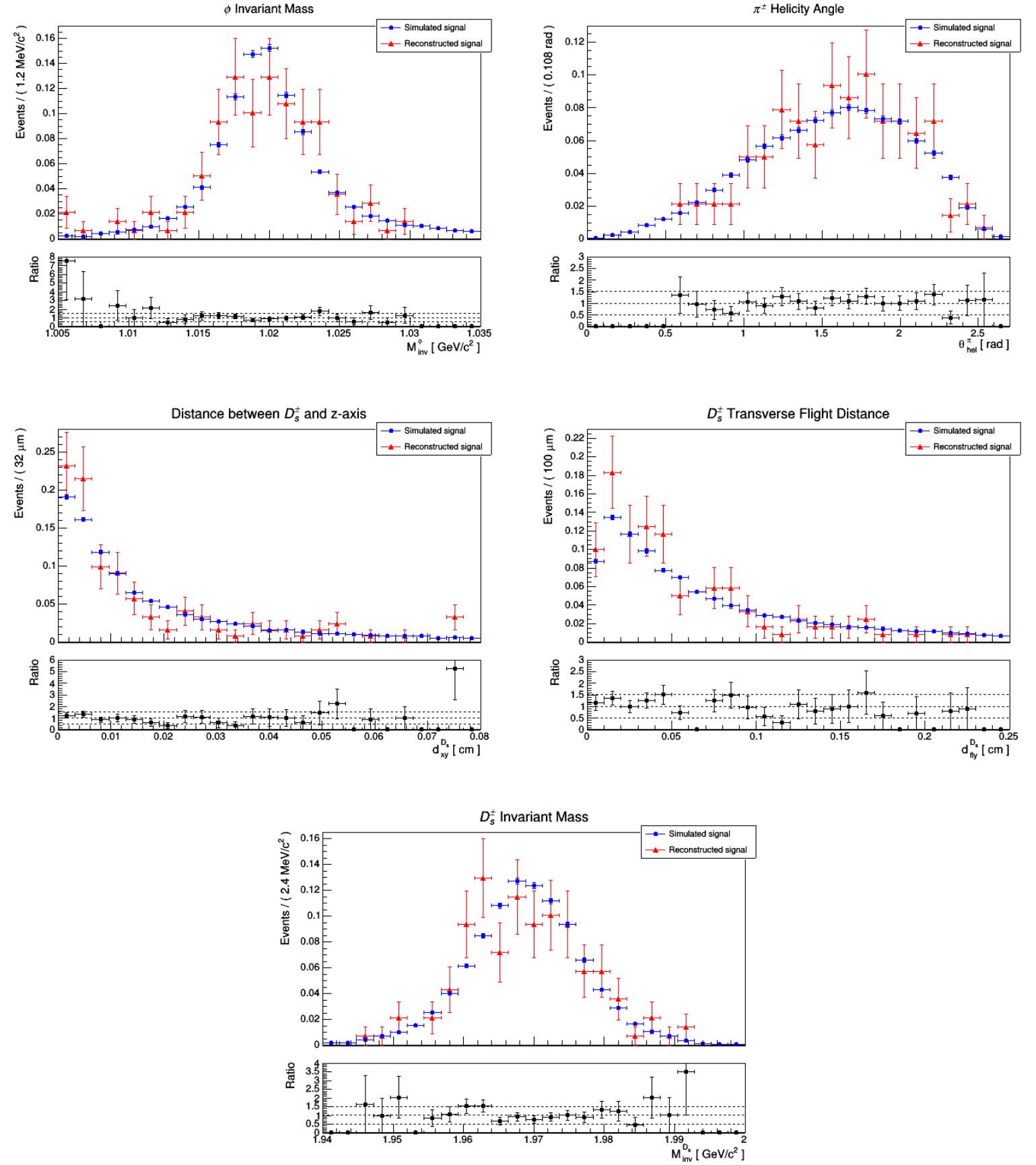
Figure 2.5: Comparison between simulated and reconstructed signal events for M_{inv}^{ϕ}, θ_{hel}^{π}, $d_{xy}^{D_s}$, $d_{fly}^{D_s}$ and $M_{inv}^{D_s}$.

The distributions from the larger "simulated signal" set appear compatible with the ones from the smaller "reconstructed signal" set. The ratio of the pairs of histograms hovers near 1 for most of the bins and oscillates only due to the statistical fluctuations in the second dataset.

2.2. Background

In order to study background events, we used a sample of events without any simulated ϕ where, however, the combinatorial background was reconstructed as a $\phi \to K^+ K^-$ vertex decay, and tracks from pions that, together with a ϕ, could be thought to derive from a $D_s^\pm$.

In background events it was of course not possible to define a set of reconstructed particles matching the simulated particles in a $D_s^\pm \to \phi \pi^\pm$ decay, because its presence was excluded when defining the event sample. Candidate decays of $D_s^\pm$ mesons in the $\phi \pi^\pm$ channel in background sample have been thus reconstructed taking, for each event, the reconstructed $\phi \to K^+ K^-$ vertices and choosing for each one an additional track, assumed to correspond to a pion, such that the total invariant mass of the reconstructed $D_s^\pm$ lies in the [1.8; 2.1] GeV/c^2 range. The distribution of $|\Delta z|$, with Δz now defined as the distance between the pion track impact parameter and the z-coordinate of the K closest approach midpoint, is shown in fig. 2.6; each pion candidate was taken if $|\Delta z| < 0.1\,cm$, otherwise it was rejected. Among all possible tracks fulfilling these requirements, the one with the minimum total distance from the K closest approach midpoint was taken.

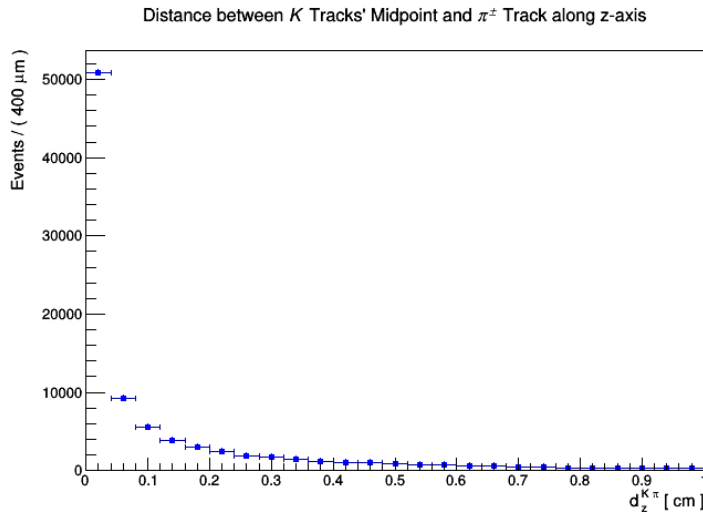The position of the $D_s^\pm$ decay vertex and its 3-momentum were then estimated as in sec. 2.1.

Figure 2.6: Distance between K tracks' closest approach midpoint and $\pi^\pm$ track impact parameter along the z-axis.

A total of 63,169 $D_s^\pm \to \phi \pi^\pm$ candidates were reconstructed; from now on, these events will be referred to simply as "background".

3. Likelihood Discriminator

The signal and background events samples described above have been used to implement a likelihood discriminator to classify new events.

A likelihood discriminator is a simple type of multivariate analysis. It works by first defining a set of N discriminating variables x_j, with $1 \leq j \leq N$, each one distributed with probability density functions $P_j^{sig}(x_j)$ and $P_j^{bkg}(x_j)$ for signal and background events. For each event the discriminating variable is defined as

$$D = \prod_{j=1}^{N} P_j(x_j) = \frac{\prod_j P_j^{sig}(x_j)}{\prod_j P_j^{sig}(x_j) + \prod_j P_j^{bkg}(x_j)}$$

The discriminating variable is contained in the [0; 1] range: if it is close to 1, the event is probably signal, while on the contrary, if it is close to 0, it is probably background.

The P_j functions have been estimated from histograms of the x_j variables, normalized at 1, in two samples of signal and background events. The D variable is then computed using the histogram contents for the bins corresponding to the x_j quantities in that event.

The chosen discriminating variables are the ones introduced in subsec. 2.1.1.

3.1. Training

The likelihood discriminator has been trained on half of the datasets discussed before, for a total of 25,332 events for signal and 31,599 events for background. The resulting normalised histograms are shown in fig. 3.1, 3.2, 3.3:
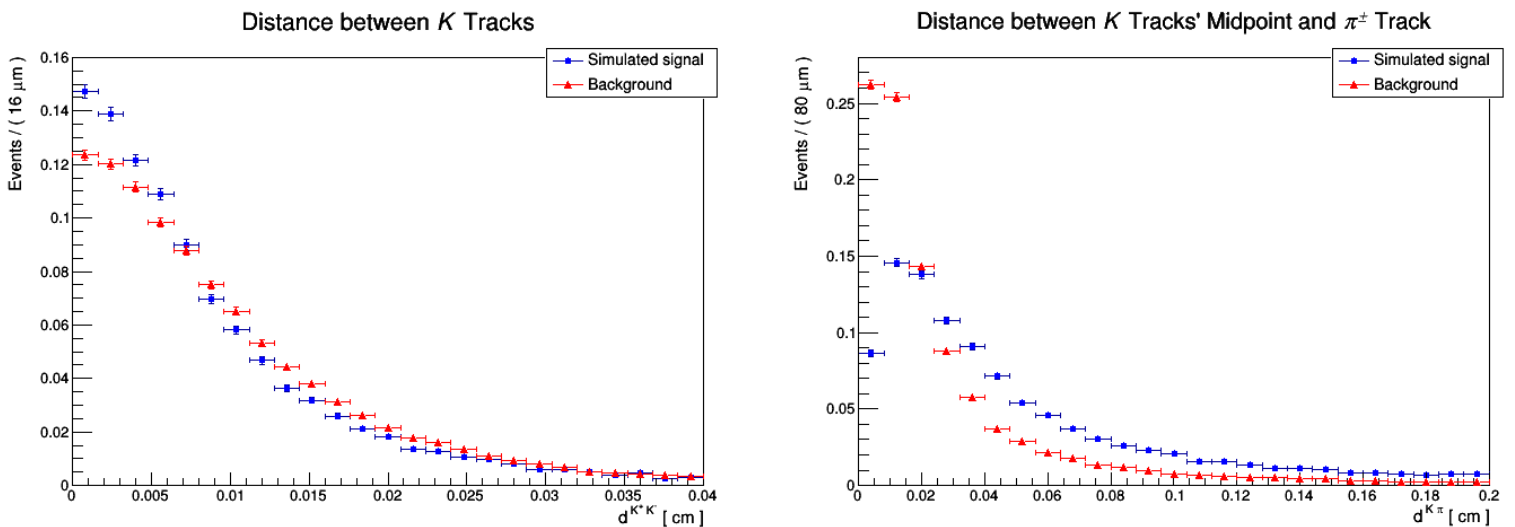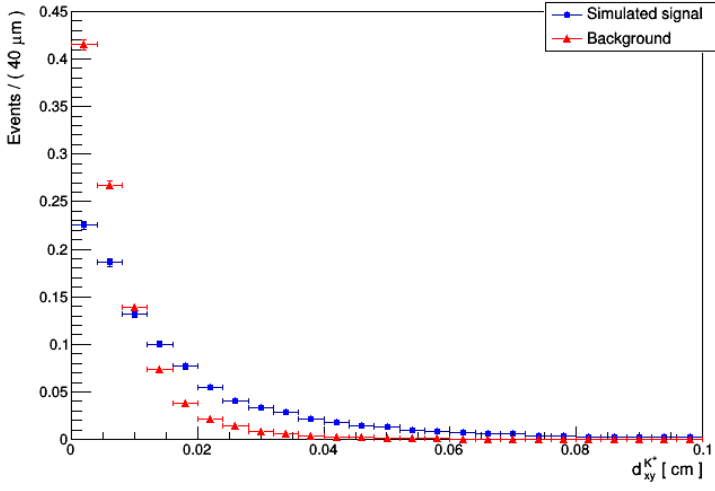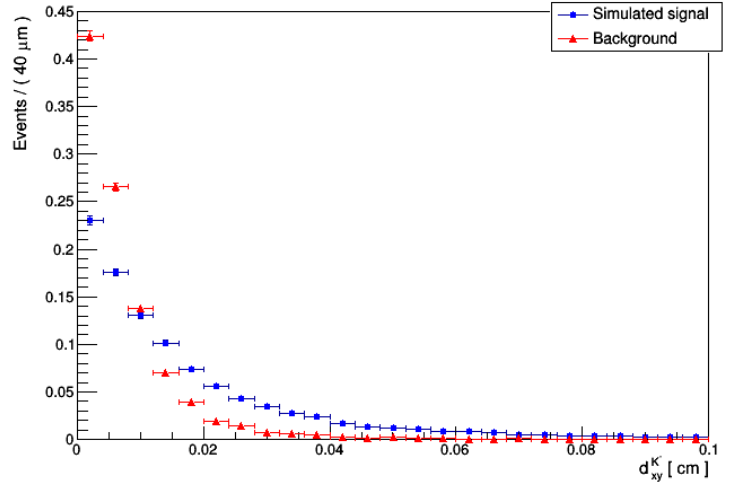

Figure 3.1: Probability density functions of $d^{K^+K^-}$ and $d^{K\pi}$ for signal and background events.
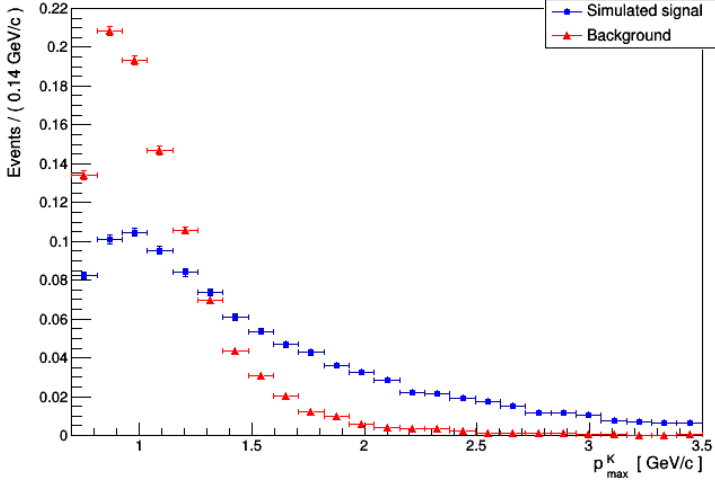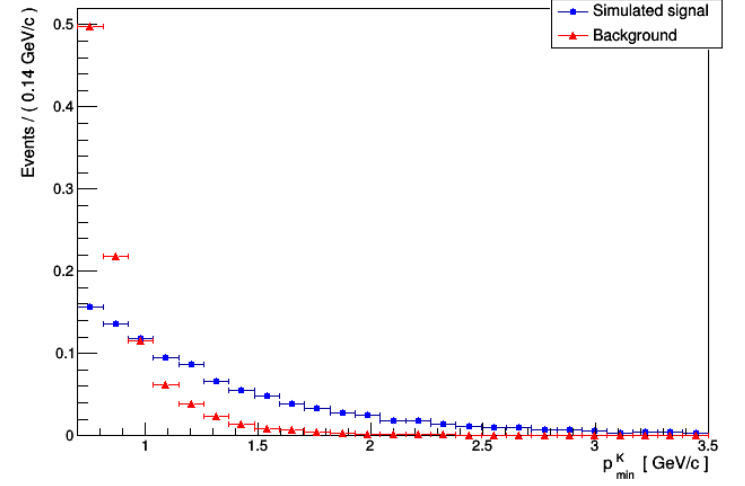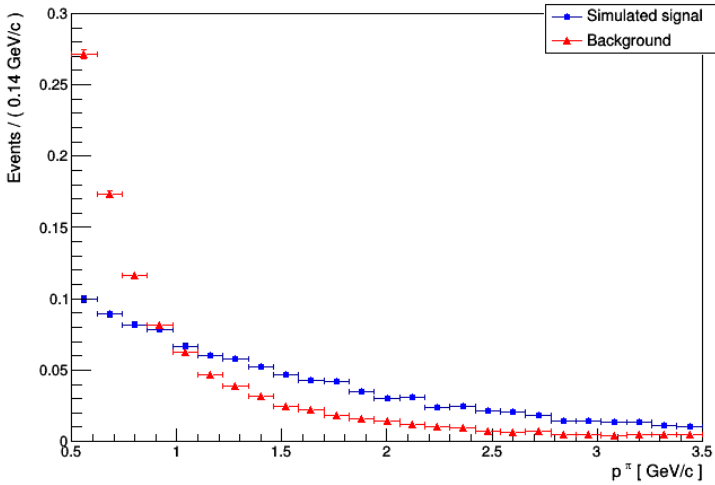
Figure 3.2: *Probability density functions of $d_{xy}^{K^+}$, $d_{xy}^{K^-}$, p_{max}^K, p_{min}^K, p^π and $\theta^{K\pi}$ for signal and background events.*
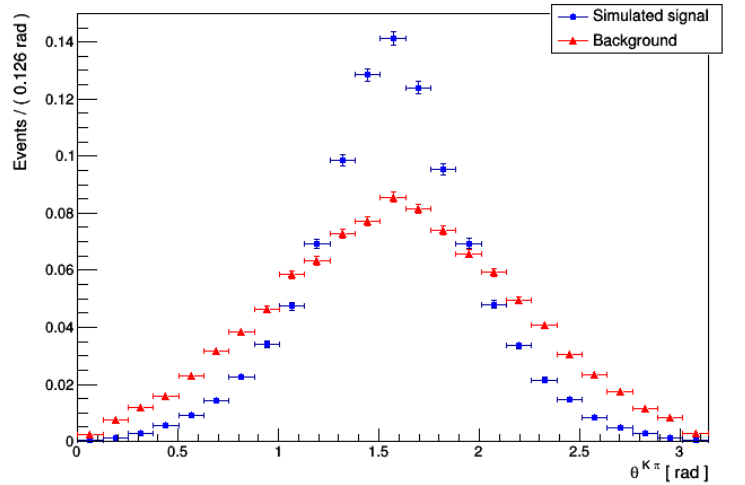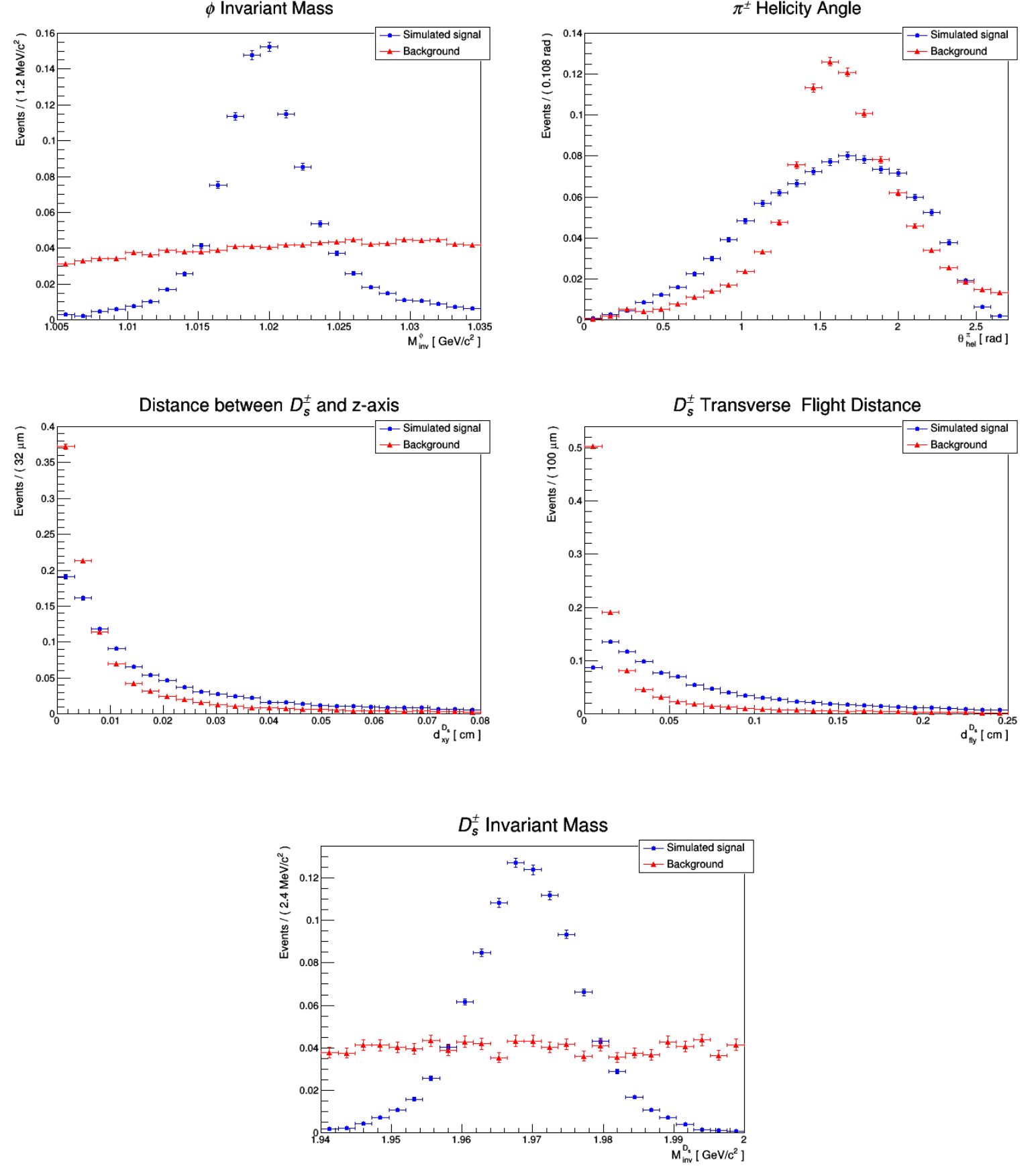
Figure 3.3: Probability density functions of M_{inv}^{ϕ}, θ_{hel}^{π}, $d_{xy}^{D_s}$, $d_{fly}^{D_s}$ and $M_{inv}^{D_s}$ for signal and background events.

The comparison between signal and background events shows that the first set usually presents larger tails, with the only exceptions being $\theta^{K\pi}$, M_{inv}^{ϕ} and $M_{inv}^{D_s}$. In the $\theta^{K\pi}$ histogram, signal events tend to gather to the central value of $\frac{\pi}{2}$ whereas background events have a wider distribution; in the two mass histograms, signal events correctly show a resonance while the background is mostly uniformly distributed.

3.2. Test

Three different discriminators have been created, each one using a different set of discriminating variables: the first one uses all of them, the second one does not use $M_{inv}^{D_s}$ and the third one utilizes neither $M_{inv}^{D_s}$ nor M_{inv}^{ϕ}. The likelihood discriminators have been tested on the second halves of the two datasets, obtaining fig. 3.4.
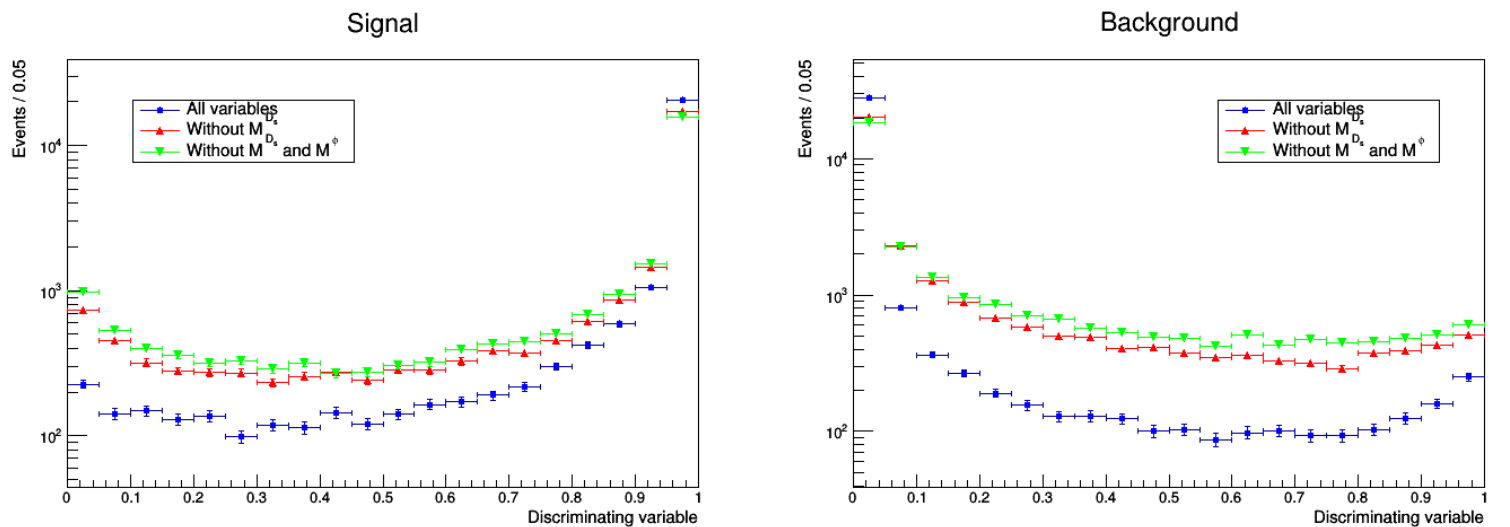

Figure 3.4: Discriminating variables for signal (left) and background (right) events.

An estimate of the discriminators' performance can be obtained by studying the Receiver Operating Characteristic (ROC) curve [2].
The ROC curve is created by varying the value of the discriminating variable at which the event is considered signal or background and then plotting the true positive rate, also known as sensitivity, on the y-axis and the false positive rate, obtained as $1-$ specificity, on the x-axis for each threshold setting. A quantitative estimation is then given by the area under the curve (AUC): the closer the area is to 1, the better the discriminator's performance. The ROC of a random classifier is a diagonal straight line and the corresponding AUC is 0.5.

The ROC for the three discriminators described above is shown in fig. 3.5: the curves have been obtained by connecting 21 points corresponding to D values ranging from 0 to 1 with increments of 0.05.
The area under the curves has been computed, obtaining 0.987 for the first discriminator, 0.943 for the second one and 0.920 for the last one. As expected, the first discriminator is the most powerful, as the two invariant masses are arguably the variables where signal and background events differ the most and therefore their inclusion greatly enhances the discrimination capability.
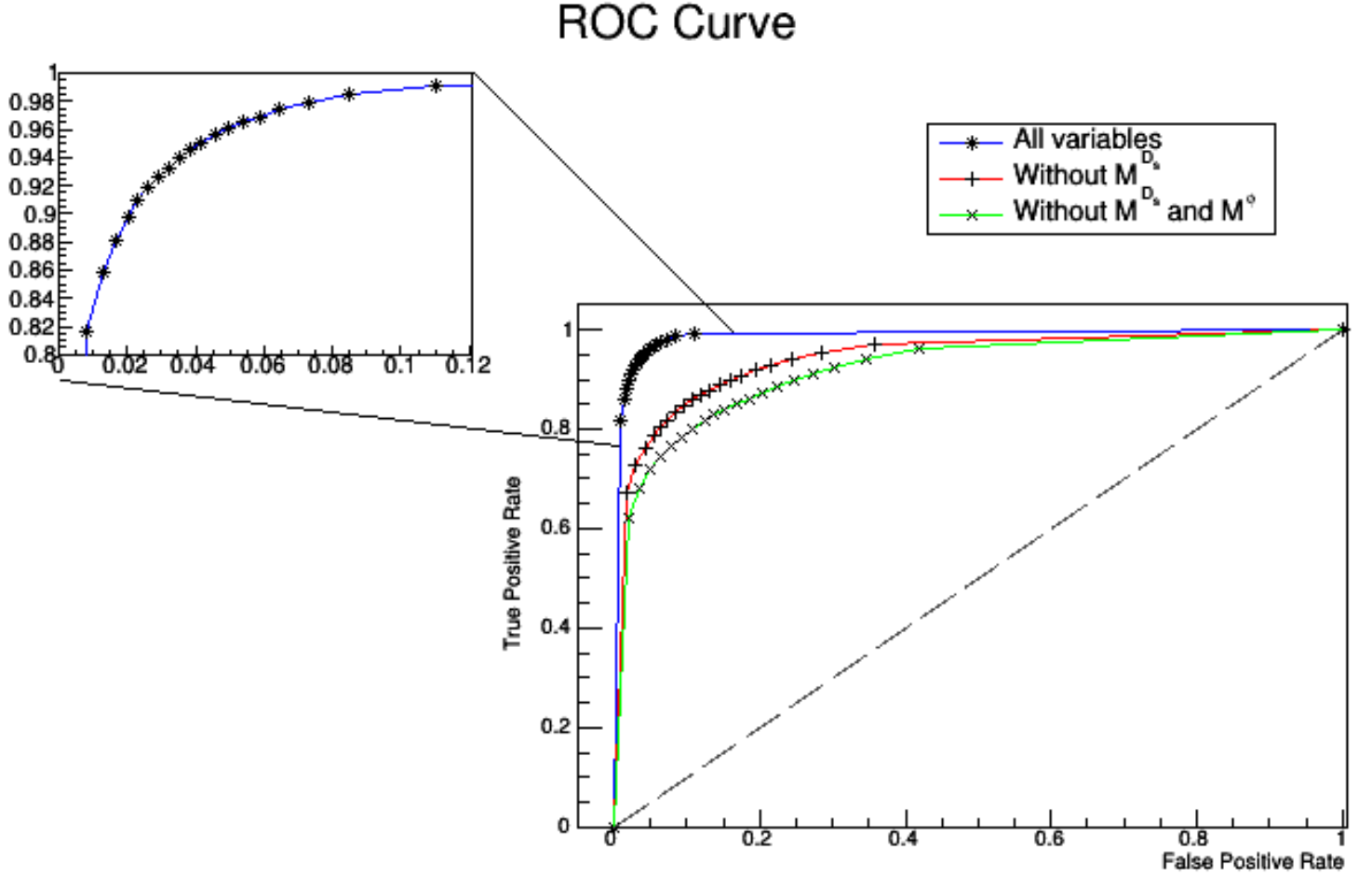
Figure 3.5: ROC curve for the three different discriminators and zoom for the first one.

Choosing a threshold setting of 0.9 for D to discriminate between signal and background, the efficiency of the discriminators results in 85.8% for the one that uses all variables, 72.8% for the one that excludes $M_{inv}^{D_s}$ and and 68.2% for the one that uses neither $M_{inv}^{D_s}$ nor M_{inv}^{ϕ}.
On the other hand, the percentages of background events that pass as signal are 1.3%, 3.0% and 3.5% respectively, thus reducing the size of the background of more than one order of magnitude.

In a following analysis using this discriminator, when producing invariant mass histograms to highlight the presence of a signal, $M_{inv}^{D_s}$ should be excluded all the same in order to avoid an obvious bias in the obtained distribution. This is also true, albeit to a smaller extent, for the discriminator that utilizes M_{inv}^{ϕ}, as of course the ϕ mass is strongly correlated to the $D_s^{\pm}$ mass.

Conclusion

$D_s^\pm$ mesons are useful when studying flavor physics as their charge allows to infer the flavor of a decaying B_s^0. The aim of this work was thus to implement a preselection so as to identify $D_s^\pm$ decay candidates in the $\phi\,\pi^\pm$ channel, followed by the cascade decay $\phi \to K^+ K^-$, to use for a full reconstruction.

A set of 13 variables to be used in the preselection was chosen, and their distributions were studied in two simulated data samples containing signal and background events respectively. In the signal sample each event contained a $D_s^\pm \to \phi\,\pi^\pm$ decay and the variables were computed using the reconstructed tracks matching the simulated particles coming from that decay. In the background sample no ϕ mesons were present among the simulated samples and fake candidates were reconstructed choosing opposite charged track pairs, compatible with coming from a common vertex and having an invariant mass near the ϕ mass, plus an additional charged track. Using the distribution of those 13 variables a "likelihood discriminator" was built; its ROC curve was then plotted and the area under the curve (AUC) was computed.
The obtained AUC was 0.987; when removing $M_{inv}^{D_s}$ from the set of variables the AUC was reduced to 0.943, removing also M_{inv}^{ϕ} the AUC decreased to 0.920.

References

[1] J. Allison et al. "Recent developments in Geant4". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 835 (2016) 186-225. `doi:10.1016/j.nima.2016.06.125`.

[2] A.P. Bradley. "The use of the area under the ROC curve in the evaluation of machine learning algorithms". *Pattern Recognition*, 30 (1997) 1145-1159. `doi:10.1016/S0031-3203(96)00142-2`.

[3] R. Brun and F. Rademakers. "ROOT — An object oriented data analysis framework". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389 (1997) 81-86. `doi:10.1016/S0168-9002(97)00048-X`.

[4] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". *Journal of Instrumentation*, 9 (2014) P10009. `doi:10.1088/1748-0221/9/10/p10009`.

[5] The CMS Collaboration. "Performance and operation of the CMS electromagnetic calorimeter". *Journal of Instrumentation*, 5 (2010) T03010. `doi:10.1088/1748-0221/5/03/t03010`.

[6] The CMS Collaboration. "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data". *Journal of Instrumentation*, 5 (2010) T03012. `doi:10.1088/1748-0221/5/03/t03012`.

[7] The CMS Collaboration. "The CMS experiment at the CERN LHC". *Journal of Instrumentation*, 3 (2008) S08004. `doi:10.1088/1748-0221/3/08/s08004`.

[8] The CMS Collaboration. "The CMS trigger system". *Journal of Instrumentation*, 12 (2017) P01020. `doi:10.1088/1748-0221/12/01/p01020`.

[9] The CMS Collaboration. "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7\,TeV$ at the LHC". *Journal of Instrumentation*, 8 (2013) P11002. `doi:10.1088/1748-0221/8/11/p11002`.

[10] R. Frühwirth. "Application of Kalman filtering to track and vertex fitting". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262 (1987) 444-450. `doi:10.1016/0168-9002(87)90887-4`.

[11] D.J. Lange. "The EvtGen particle decay simulation package". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462 (2001) 152-155. `doi:10.1016/S0168-9002(01)00089-4`.

[12] T. Sjöstrand, S. Mrenna and P. Skands. "A brief introduction to PYTHIA 8.1". *Computer Physics Communications*, 178 (2008) 852-867. `doi:10.1016/j.cpc.2008.01.036`.

[13] P.A. Zyla et al. "Review of Particle Physics". *PTEP*, 2020 (2020) 083C01. `doi:10.1093/ptep/ptaa104`.