Characterising Particle ID Performance for the LHCb Calibration Sample $K_S^0 \to \pi^+\pi^-$ in Run 3 Data

Candidate number: 1061541 **Supervisor:** Prof. Sneha Malde **Project:** PP04

Abstract

The LHCb data analysis strategy relies on calibration samples for PID calibration. The use of the $K_S^0 \to \pi^+\pi^-$ calibration sample over the existing $D^{*+} \to D^0\pi^+$ sample in the momentum range $P \in [1.5, 30]$ GeV/c was investigated. PID efficiencies of the two samples were produced, and compared as a cross-check. Finding likely agreement, signal yields between the two samples were compared, showing that the $K_S^0 \to \pi^+\pi^-$ sample has the necessary statistics to be used. We conclude that the $K_S^0 \to \pi^+\pi^-$ sample could be made available as a more reliable sample for PID in this kinematic range, where the $D^{*+} \to D^0\pi^+$ sample suffers from bad purity, and has a background that is more difficult to model. Furthermore, the performances of the PID discrimination variables Delta-Log-Likelihood and ProbNN were compared. It was found that for $P \in [15, 30]$ GeV/c, ProbNN showed higher misidentification rates of π as μ and K, indicating that the Delta-Log-Likelihood should be used in this momentum range to reduce uncertainties in future analyses.

1 Introduction & background

The LHCb experiment is comprised of a series of detectors at the LHC, focusing on precision measurements of heavy flavour physics, including measurements of CP violation in heavy flavour hadrons [1]. The ability to correctly identify particles in the detector is essential for both the trigger (the system that discards uninteresting or background events) and for analysing the data collected by the experiment [2]. The Ring Imaging Cherenkov (RICH) system provides the particle ID (PID) functionality for identifying charged particles in LHCb [3].

The RICH system is comprised of Cherenkov radiators, mirrors, and photon detectors. In the radiators, fast moving particles radiate photons in a Cherenkov cone of opening angle $\theta_C = (n\beta)^{-1}$, with β the particle velocity in units of c, and n the refractive index [4]. The mirrors remove the Cherenkov photons from the main detector acceptance, and focus them onto rings incident on the photon detectors. The momentum of the particle is fixed from the bending radius of the track in the presence of a magnetic field. Therefore, different mass hypotheses for the same track will result in different values of the calculated velocity and of the corresponding θ_C. Rings corresponding to different mass hypotheses are fitted to the hits registered by the photon detectors, with the fit quality giving the likelihood $\mathcal{L}(X)$ of the track being a particular species of particle X. These likelihoods are combined to produce the variable Delta-Log-Likelihood (DLL), used to compare whether a certain PID hypothesis should be chosen over another (a *PID discrimination variable*). It is given by $DLL(X-Y) = \log\mathcal{L}(X) - \log\mathcal{L}(Y)$, i.e. positive values indicate that the hypothesis X is more favourable.

Alongside the DLL, a second PID discrimination variable, ProbNN(X), can be obtained from the LHCb data. This is a neural network-informed probability of the hypothesis X [5]. ProbNN uses correlations between the different detector systems in the experiment to provide an (in principle) improvement over the DLL [6]. In practice, PID discrimination variables are used to make PID cuts. These are selections of events ('cuts') based on the level of confidence in the identification of each event with a particular species of particle. For example, a PID cut might be the selection of events that satisfy $DLL(K-\pi) < 0$, i.e. those events that are more likely to be pions than kaons. PID cuts are used in heavy-flavour data analysis to reduce background to the level required by the analysis at hand: the PID discrimination variables for each event inform how likely it is to be background, so a PID cut based on these variables rejects background. Any PID cut will involve some amount of true signal being rejected, and some amount of misidentification of background as signal (which would be allowed through the cut). The rates at which these processes happen must be understood in order to interpret the results of analyses. The proportion of true signal events passing a PID is known as the PID efficiency, while the rate of background events incorrectly allowed through as signal is known as the misidentification rate. These rates depend on the cut chosen, on momentum and on the angle of the track with respect to the beam direction (amongst other variables), and they are obtained through PID calibration.

LHCb is reliant on data-driven calibration: while simulations could in theory be used to do PID calibration, they are not currently of satisfactory quality. Currently they generate incorrect multiplicities (how 'busy' the detector is when each event is recorded), and time-saving shortcuts have been made in the simulation of photons in the RICH. Therefore, PID calibration is done with calibration samples, which consist of events where the particle species can be deduced unambiguously without the use of the experiment's PID systems, e.g. by using the kinematic structure of the event [2]. This gives a sample of events with known particle content (up to a small amount of background), which can then be used to assess the response of the experiment's PID systems. In other words, as calibration samples are comprised almost entirely of signal events, they can be used to obtain PID efficiencies and misidentification rates, e.g. by performing a PID cut and counting the fraction of sample events that pass the cut. In this way a measure of the PID performance of LHCb can be obtained.

This investigation focuses on assessing the pion PID performance of LHCb using the $K_S^0 \to \pi^+\pi^-$ calibration sample, with the Run 3 data. This data was collected by the experiment in 2024 and had not yet been analysed. In particular, the K_S^0 sample had not yet been investigated as a candidate calibration sample prior to this work, so its purity (level of contamination by background) was unknown. There was also no knowledge of any potential challenges in modelling the signal and background of the sample. We show that the sample is high purity, and the modelling of signal and background was straightforward, with scope to be improved with a simulation sample. Unfortunately, it was not possible to produce a sufficiently large simulation sample, generated in realistic conditions, on the timescale of this project.

Like the K_S^0 sample, the $D^{*+} \to D^0\pi^+$ calibration sample produces a set of events known to correspond to pions, and is the sample currently in use for pion PID. The kinematic ranges of the D^{*+} decay products overlaps with that of the K_S^0, but also extends to higher momenta. It is of interest to compare which calibration sample is more reliable for pion PID in the momentum range of the K_S^0 sample, namely 1.5 to 30 GeV/c. This is the main focus of this investigation.

Doing analysis using a PID discrimination variable that has a lower misidentification rate will reduce backgrounds. Therefore, it is of importance to compare the PID performance of the DLL and ProbNN variables. In particular, we investigate misidentification rates of pions as muons or kaons. The decay $B^0 \to \mu^+\mu^-$ is highly suppressed in the standard model, but may be enhanced in beyond the standard model theories [7]. Background processes include $B^0 \to \pi^+\pi^-$ and $B_s^0 \to \pi^+\pi^-$: lowering the misidentification rate of pions as muons would allow for better rejection of these backgrounds. Lowering the misidentification rate of pions as kaons reduces uncertainties in studies of many hadronic processes.

Section 2 investigates the purity of the K_S^0 sample and its kinematic range. We also describe how PID performance measures can be obtained from calibration sample data. The resulting efficiencies for the K_S^0 sample are discussed in Sec. 3. Section 4 compares pion PID efficiencies from the K_S^0 sample to those of the D^{*+} sample, as a cross-check to ensure there is agreement between them. We compare the signal yields of the two samples, in order to evaluate which sample should be the dominant one in this kinematic range, or otherwise if the K_S^0 sample can be used to augment the calibration provided by the D^{*+} sample in this range. Section 5 shows the results of a comparison study between the performance of the DLL and ProbNN. Finally, concluding thoughts are given in Sec. 6.

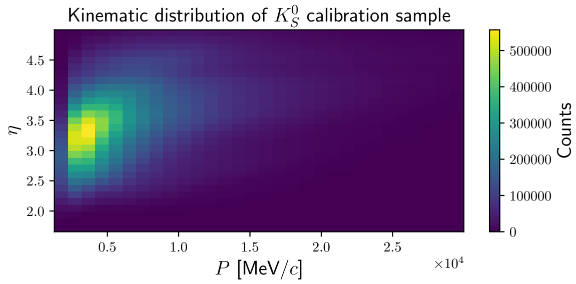
2 Methodology

2.1 Kinematic distribution of the K_S^0 and $D^{*+} \to D^0\pi^+$ calibration samples

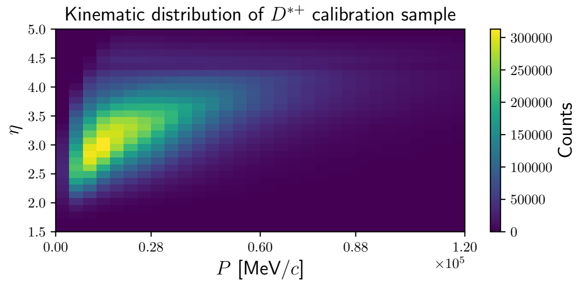
The kinematic distributions (Fig. 1) were parametrised by the momentum P and the pseudorapidity η, which is related to the angle of the track with respect to the beam direction (θ) by $\eta = -\ln\tan(\theta/2)$. Cuts were applied to the K_S^0 sample dataset such that $P \in [0.15, 3] \times 10^4$ MeV/c and $\eta \in [1.5, 5]$. These cuts were used in all analysis presented in this work.

2.2 Contributions from background processes

The presence in the data of events due to the background processes $\Lambda^0 \to p\pi$, $K^* \to K\pi$, $D^0 \to K\pi$, and $\phi \to KK$ was investigated. This is effectively an exhaustive list of possible contributions that are two body decays, with branching fractions large enough to be significant, which could feasibly result in a reconstructed mass close enough to the K_S^0 mass to contribute to the background. With the original hypothesis being $K_S^0 \to \pi\pi$, alternative mass hypotheses were tested by changing the timelike component of the 4-momentum $\sqrt{m^2 + p^2}$ of one or both of the daughter particles. The invariant mass



(a) K_S^0 calibration sample



(b) D^{*+} calibration sample

Figure 1: Comparison of pion kinematic distributions. The D^{*+} sample extends to higher momenta. The sPlot method (see Sec. 2.3) was used to reject background events when producing these plots.

was then reconstructed from the sum of the daughter 4-momenta. Of the four dominant background processes investigated, a peak was only observed for $\Lambda^0 \to p\pi$, located at the mass of the Λ^0 (Fig. 2). The structure visible was a small peak sitting in a valley, and was determined to be due to an incorrectly functioning veto on the Λ^0 mass in the trigger of LHCb. Here the algorithm to reject $\Lambda^0 \to p\pi$ events failed at a low rate, meaning some, but not all such events were rejected. Cuts were placed either side of the Λ^0 peak to remove the contribution from this process. This was done by computing the invariant mass for the $p\pi$ hypothesis and removing any events with invariant mass falling within the range $[1104, 1128]$ MeV/c^2.

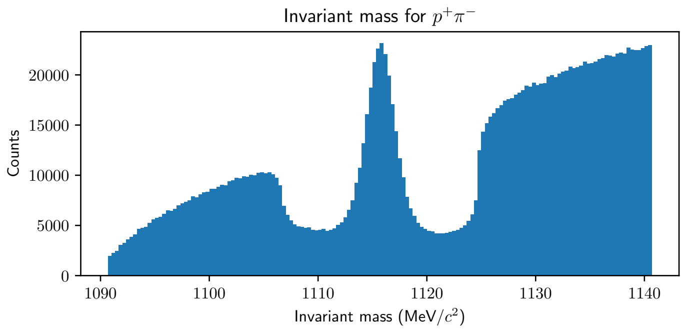


Figure 2: Invariant mass distribution of the data under the $p\pi$ hypothesis. The peak visible at the Λ^0 mass (1116 MeV/c^2) was due to an incorrectly functioning veto in the LHCb trigger, meaning some Λ^0 background was allowed into the K_S^0 calibration sample.

2.3 Obtaining PID performance measures from calibration sample data

As we have a calibration sample of pions, the pion PID efficiencies are the fraction of signal events in the sample that pass a given PID cut. For example, an efficiency could be obtained by counting the number k of signal events passing the cut $DLL(K - \pi) < 0$ (i.e. the pion hypothesis is favoured) and taking the ratio k/N with N the total number of signal events before the cut. Note that a discrimination of signal events from background events is necessary. The sPlot technique [8] allows us to discriminate signal events from background events in variables where we have no knowledge of the signal and background PDFs (so-called *control variables*), which include the DLL, ProbNN, P, and η.

This technique requires a variable for which we do have knowledge of the signal and background PDFs (a *discriminating variable*), and can model these with fits. Here we use the invariant mass of the $\pi\pi$ system as this variable. These signal and background PDFs are then used to compute a weight, referred to as an sWeight, which is assigned to each event. When events are binned, the contributions from events with negative sWeights partially cancel with those from events with positive sWeights, in such a way that leaves only contributions from signal events in each bin. Hence, by binning sWeighted data and summing sWeights in each bin, the per-bin signal yield N can be obtained. The number k of events passing a PID cut can be obtained by summing the sWeights in each bin after the cut; the efficiency in each bin is given by $\epsilon = k/N$.

It should be noted that the sPlot technique can be used only under the assumption that the joint PDF describing all variables in the dataset, f_{tot}, factorises into functions of the discriminating/control variables only. This is to say, for a single discriminating variable m and control variable t, that $f_{\text{tot}}(m, t) = g(m)h(t)$. Applied to this study, this means that the mass distribution should not vary in different regions of the P, η parameter space.

The fit-and-count (FAC) technique is an alternative way of computing efficiencies, without using sWeights. In this procedure, the data is binned in P and η, and a fit to the $\pi\pi$ invariant mass is performed on the data in each bin. The quantity N for each bin is taken from the signal yield returned by this fit. Then, a PID cut is made, and a second mass fit performed on the data passing the cut. The quantity k is found from the signal yield of this post-cut fit, with the efficiency again given by $\epsilon = k/N$.

The sPlot method offers better scalability compared to the FAC method [2]. It also allows the analyst to avoid having to do fits to determine PID efficiencies and misidentification rates: these can be labour intensive to carry out, especially for complex fits, e.g. the 2D simultaneous fits used in the $D^{*+} \to D^0\pi^+$ calibration sample. The FAC method can struggle to model how the contamination rates of different sources of background change before and after a PID cut. Nevertheless, as it is independent of the sWeights, it serves as a useful cross-check for the validity of the sPlot method, in light of the assumption for its validity set out above. Here, the usage of the FAC method is valid as the K^0_S sample is high purity and does not require a complex fit to model the signal and background. In this study, both methods were used. All results here are based on the sPlot method, with the FAC method used to cross-check the results and provide a measure of systematic bias (see Sec. 2.3.2).

2.3.1 Fits and computation of sWeights

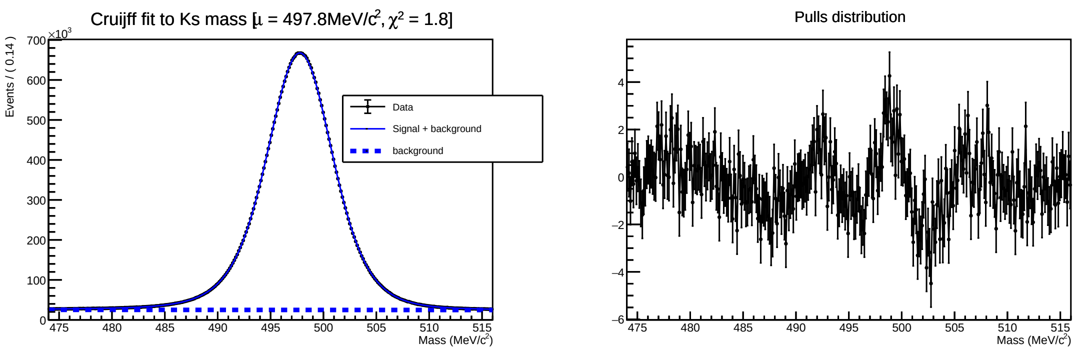


Figure 3: The mass distribution of the K^0_S, with signal and background fits (left). Note the high purity of the sample. Also shown is the pulls distribution (right): note the presence of an oscillatory component, indicating imperfect modelling of detector resolution.

Fits to the invariant mass of the $\pi\pi$ system were made in order to determine signal and background PDFs, which in turn were used to compute sWeights. The signal model used was a Gaussian PDF $f_G(m; \mu, \sigma)$ plus an asymmetrical Gaussian-like PDF $f_C(m; \mu, \sigma, \alpha_L, \alpha_R, \beta)$, known as the Cruijff distribution, which is given (up to a normalisation factor) by

$$f_C(m; \mu, \sigma, \alpha_L, \alpha_R, \beta) = \begin{cases} \exp\left(-\dfrac{(m-\mu)^2(1+\beta(m-\mu)^2)}{2\sigma^2 + \alpha_L(m-\mu)^2}\right) & m < \mu \\ \exp\left(-\dfrac{(m-\mu)^2(1+\beta(m-\mu)^2)}{2\sigma^2 + \alpha_R(m-\mu)^2}\right) & m > \mu. \end{cases} \tag{1}$$

The parameters α_L and α_R of f_C account for asymmetry due to e.g. radiative effects, while also modelling the effect of the experimental resolution of the LHCb [9]. The parameter β of f_C ensures a Gaussian tail when $\alpha_{L,R}(m-\mu)^2 \gg 2\sigma^2$. A linear combinatorial background f_B was included such that the fit PDF was given by $f_G + f_C + f_B$. It was found that the presence of f_G was not necessary to achieve a reasonable fit, so it was removed. All parameters (including signal and background yields) were allowed to vary, resulting in a fit with $\chi^2_\nu = 1.8$. While this χ^2_ν is reasonably high, without access to simulations which could be used to model the signal and background more accurately (e.g. constrain the tails of the mass distribution), it was not possible to lower the χ^2_ν further without overfitting. The fit and distribution of pulls (i.e. error-normalised residuals) is shown in Fig. 3.

Models other than the Cruijff distribution were investigated, namely a double-Gaussian PDF and an asymmetric Crystal Ball PDF [10]. Due to the high statistics of the sample and the complex detector resolution, it proved challenging to obtain a satisfactory fit from these models. For this reason, the investigation proceeded with the Cruijff distribution, despite the oscillatory component visible in the pulls (see the right of Fig. 3). This meant that confidence in the quality of the residuals was sacrificed in favour of a better modelling (and separation) of signal and background events by the fit with the Cruijff distribution.

The signal and background PDFs from the aforementioned fits were used to compute sWeights using the `RooStats::sPlot` class of the software `ROOT` [11]. These were then used to compute efficiencies using the procedure outlined in Sec. 2.3.

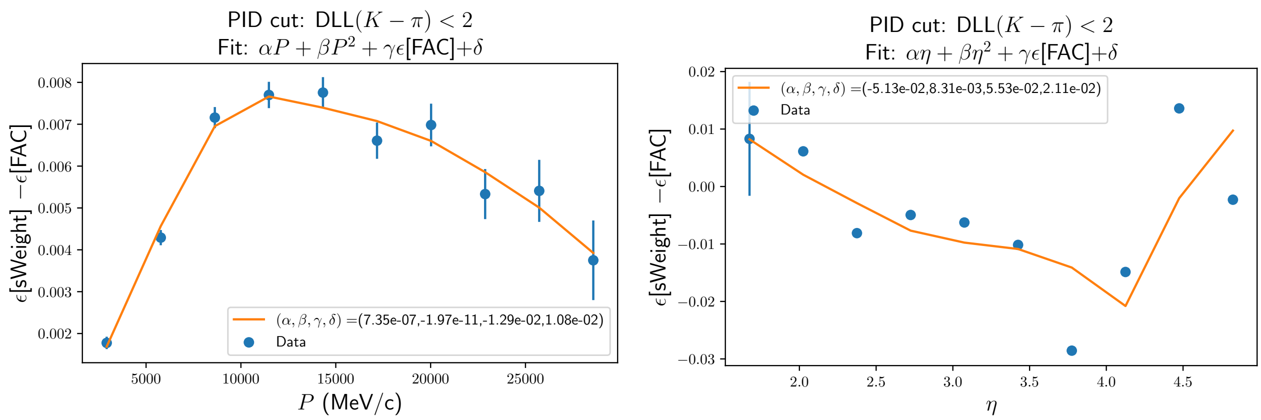
2.3.2 Consideration of uncertainties

Statistical errors for the efficiencies were calculated by assuming that the PID cuts performed on the data could be modelled by a binomial process with true efficiency (i.e. probability of success) ϵ_{true}. To a good approximation, we may replace ϵ_{true} with our efficiency $k/N \equiv \epsilon$. Thus the statistical error in the efficiency is given [12] by $\sigma_{\text{stat}} = \sqrt{k(1-k/N)}/N$.

A measure of systematic bias was calculated from the difference between the PID efficiencies, as computed from the sPlot procedure ($\epsilon_{\text{sWeight}}$) and from the fit-and-count (ϵ_{FAC}) procedure. As the FAC procedure does not require the assumption that the PDF factorises into discriminating and control variables, it can be used to show the presence of any systematic errors in the sWeights. We define the differences $r_i = \epsilon_{\text{sWeight},i} - \epsilon_{\text{FAC},i}$, with i indexing each bin in P, η. A model of systematic bias $b_i = \alpha x + \beta x^2 + \gamma \epsilon_{\text{FAC},i} + \delta$ was fitted (via a weighted least-squares method) to these differences for each variable $x \in \{P, \eta\}$. This fit was used to quantify additional scatter σ_{add} not explained by statistical scatter σ_{stat}, which was estimated with the formula

$$\sum_{\text{bins } i} \frac{(r_i - b_i)^2}{\sigma^2_{\text{stat}} + \sigma^2_{\text{add}}} = \text{degrees of freedom} = \text{number of bins}. \tag{2}$$

This additional scatter was added in quadrature with the statistical scatter to produce a measure of overall scatter. An example of the modelling of bias with this procedure is shown in Fig. 4.



(a) Differences as a function of P (b) Differences as a function of η

Figure 4: The systematic bias present in the results of Fig. 5, shown in the differences between PID efficiencies produced by the sPlot and FAC methods. The error bars show the statistical scatter σ_{stat}; these results were used to obtain the additional scatter σ_{add} not explained by σ_{stat}, more visible in Fig. 4b. This additional scatter was added in quadrature with σ_{stat} to give the error bars in Fig. 5.

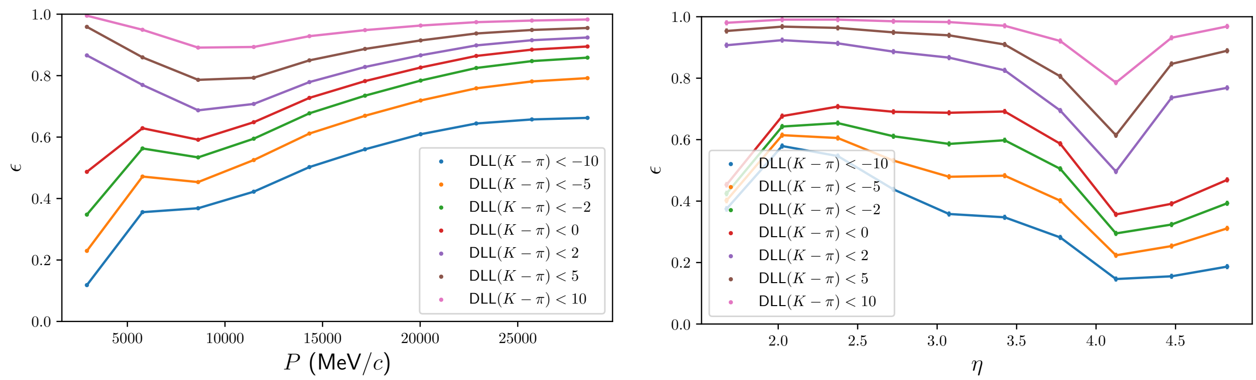
Note that all systematic biases were left uncorrected, but are quoted should corrections be required in future work.

3 Pion PID efficiency results

3.1 Asymmetry of $\pi^{\pm}$ PID efficiencies

The PID efficiency $\epsilon[\pi^{\pm}]$ was compared for the two pion charges, to identify any potential charge asymmetry. The PID efficiencies for both π^+ and π^- were calculated as functions of P, η for PID cuts of $DLL(K-\pi) < x$ such that $x \in [-10, -5, -2, 0, 2, 5, 10]$ (these efficiencies are presented in Sec. 3.2). Then, the asymmetry $A_i = |\epsilon[\pi^+]_i - \epsilon[\pi^-]_i|$ was averaged via an inverse-variance weighted mean over each bin i, giving an average $\bar{A}$ for each cut. Across all cuts, the maximum observed value of $\bar{A}$ was 0.25(9)%, i.e. a very small asymmetry. In the following work we restrict our investigation to π^+ only.

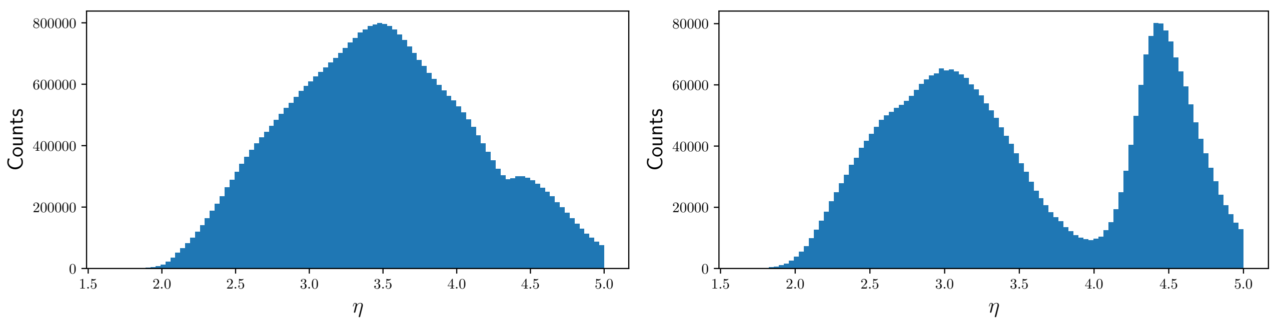
3.2 PID efficiency of π^+ using a DLL cut.



(a) π^+ PID efficiency in P (b) π^+ PID efficiency in η

Figure 5: Plots showing the π^+ PID efficiency (ϵ) for a range of DLL cuts. Error bars show $\sigma_{\text{stat}} \oplus \sigma_{\text{add}}$, and are too small to see on Fig. 5a. The additional scatter σ_{add} was calculated for a single cut, namely $DLL(K-\pi) < 2$, but is assumed to be a reasonable representation of the additional scatter present for all cuts shown here. Figure 4 gives more details on the calculation of systematic bias and σ_{add} for these results.

Figure 5 shows plots of the PID efficiency for the π^+ in P, η, for a range of cuts on the DLL. We identify two features in these plots. Firstly, on both plots in Fig. 5, we note the splitting of the curves into two families around $DLL(K-\pi) = 0$, in the low P and low η regions. At low P, there are fewer Cherenkov photons produced, leading to less PID information being available from the RICH system. At low η, i.e. high track angle, there are reductions in the optical acceptance for the RICH. The lack of PID information leads to less certainty in the particle hypothesis, and the DLL collapses towards zero in both cases. A large peak at $DLL(K-\pi) = 0$ means that a cut on one side of the peak will include a lot more data than a cut on the other, separating the curves into two families. Secondly, in Fig. 5b, we note another split into two families of curves at $\eta \approx 4$. The data featured a bump in the plot of its η distribution, beginning at this value of η (see Fig. 6). This bump consisted solely of



(a) No cut to DLL (b) Cut: $0.01 > DLL(K-\pi) > -0.01$

Figure 6: Distribution of η for π^+ (Fig. 6a) with a bump visible at $\eta \approx 4.5$, which was enhanced by restricting the DLL to just around zero (Fig. 6b). Hence, events in the bump have $DLL(K-\pi) \approx 0$.

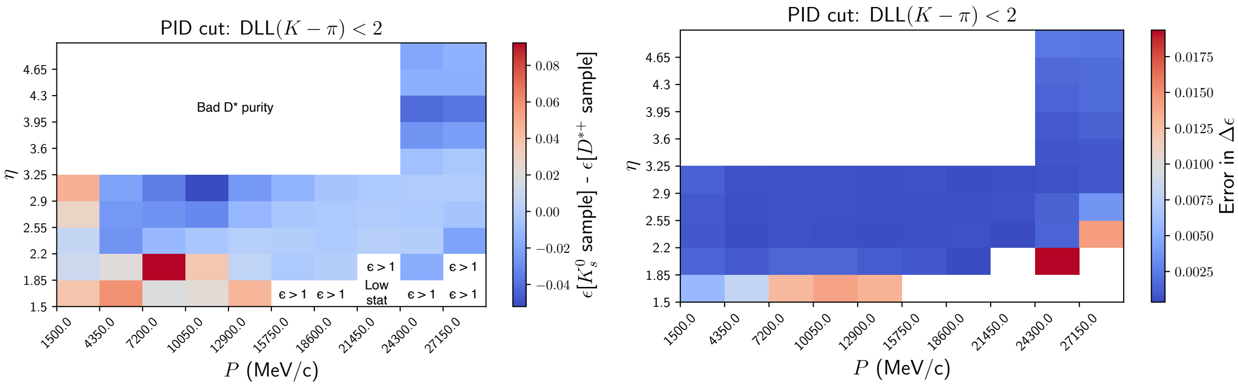
events with $DLL(K-\pi)=0$, meaning that at this η there was a significant contribution from events with no meaningful PID information. Again, a large spike at $DLL(K-\pi)=0$ caused a separation into two families of curves.

The modelling of systematic bias using the FAC cross check is shown in Fig. 4: we observed a maximum disagreement between the two methods of 0.8% for the data in P and 3% for the data in η. These were the highest levels of disagreement between the FAC and sPlot methods observed over the whole investigation. Future work could investigate whether the assumption required for the validity of using the sPlot method is violated by a small amount here.

4 Comparison to the $D^{*+} \to D^0\pi^+$ calibration sample

4.1 Comparison of PID efficiencies

We expect agreement between the PID efficiencies produced with the K_S^0 sample and the D^{*+} sample, as the PID performance should depend only on the detector response to pions, not on the decay channel that produced them. Figure 7 shows a 2D plot of the difference in PID efficiencies produced from the $K_S^0 \to \pi^+\pi^-$ and the $D^{*+} \to D^0\pi^+$ calibration sample. The difference was relatively large in some bins, reaching around 8%.



(a) Plot of $\Delta\epsilon \equiv \epsilon[K_S^0$ sample$]\,-\epsilon[D^{*+}$ sample]

(b) Statistical errors for the plot in Fig. 7a

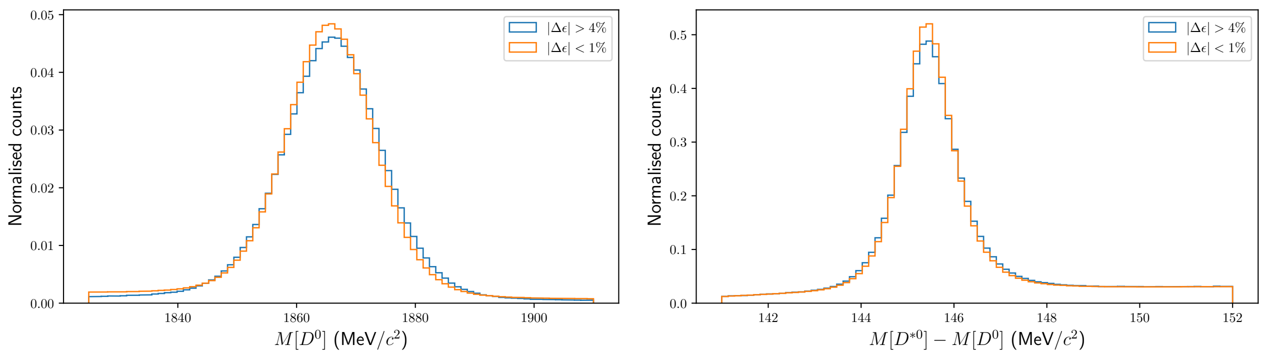
Figure 7: Plots showing the difference in PID efficiency between the $K_S^0 \to \pi^+\pi^-$ and the $D^{*+} \to D^0\pi^+$ calibration samples, as well as associated errors. Some bins had relatively high differences ($\sim 8\%$). Note the bins removed for having bad purity for the D^{*+} calibration sample, and the bins removed for having an efficiency larger than 1 for the K_s^0 sample due to imperfect cancellation of sWeights. A single bin for the K_S^0 sample with $N=1.55, k=1.14$ and $\sigma_{\text{stat}}=0.35$ was also removed, and is marked as being low-statistics.

In order to investigate the origin of this discrepancy, mass distributions for the D^{*+} sample were plotted, for bin regions with a significant disagreement in PID efficiency ($|\Delta\epsilon|>4\%$). These distributions were compared to those in regions with good agreement ($|\Delta\epsilon|<1\%$). The results are shown in Fig. 8. The difference in these mass distribution between the regions with good vs. bad agreement indicates that the validity of the sPlot method is called into question for this calibration sample (see Sec. 2.3 for an outiline of the necessary assumptions when using the sPlot technique). Therefore, this provides a plausible explanation for the larger than expected difference in efficiencies between the two calibration samples, namely systematic errors in the sWeights. It is likely that once these are taken into account, there will be agreement between the two sets of PID efficiencies, but further investigation is required.

Note that a FAC procedure to cross-check the $D^{*+} \to D^0\pi^+$ efficiencies was not possible. This was due to difficulties in accounting for the contamination rates of different D meson decays changing before and after the PID cut. Further, a 2D FAC would be extremely time consuming to execute, requiring 106 fits to be done (2 per bin).

4.2 Comparison of yields

Having carried out the cross-check by comparing the efficiencies, a comparison of signal yields was performed, to determine if the K_S^0 sample has the necessary statistics to augment the D^{*+} sample



(a) D^0 mass distribution

(b) Distribution of the D^{*+}, D^0 mass difference.

Figure 8: Plots showing comparisons between mass distributions for bin regions with good and bad agreement in PID efficiency between the two calibration samples. Note the differences, showing that the total PDF does not cleanly factorise into functions of the discriminating and control variables. This calls into question the validity of using the sPlot technique for the D^{*+} calibration sample.

for pion PID calibration in the Run 3 data. For the $K^0_S \to \pi^+\pi^-$ sample, the signal yield was taken from the fit performed in Sec. 2.3.1. For the $D^{*+} \to D^0\pi^+$ sample, cuts were made around the peak of the D^{*+} mass distribution and that of $M[D^{*+}] - M[D^0]$ (M denotes the mass distribution), both over the kinematic range of the $K^0_S \to \pi^+\pi^-$ sample. This simple procedure was used instead of a fit-based procedure as there is high purity around the peaks in the D^{*+} sample, and the fits for this sample are complex. Hence, this rough estimate is suitable for our purposes, and avoids having to do time-consuming fits.

The LHCb trigger uses a prescaling factor f_i so that it randomly selects a fraction f_i of events of a particular type i, and rejects the rest. This avoids unnecessarily large volumes of uninteresting data being collected. It was found that, with the existing prescale factors applied, the yields were comparable ($\sim 5 \times 10^7$), indicating that even with significant prescaling the K^0_S sample is a viable addition to the LHCb calibration samples. With the prescale factors removed the $K^0_S \to \pi^+\pi^-$ yield would be almost 3 orders of magnitude larger (yield$_{K^0_S} \sim 8 \times 10^{10}$, yield$_{D^{*+}} \sim 10^8$). This means that the pion yields in Run 4 could be increased, or the pion momentum range could be enhanced, through the use of graduated prescales at different momentum thresholds.

5 Comparison of the PID discrimination variables DLL and ProbNN

5.1 Calculation of misidentification rates

Table 1: The global cuts used to give a roughly 5% misidentification rate for π as K and μ.

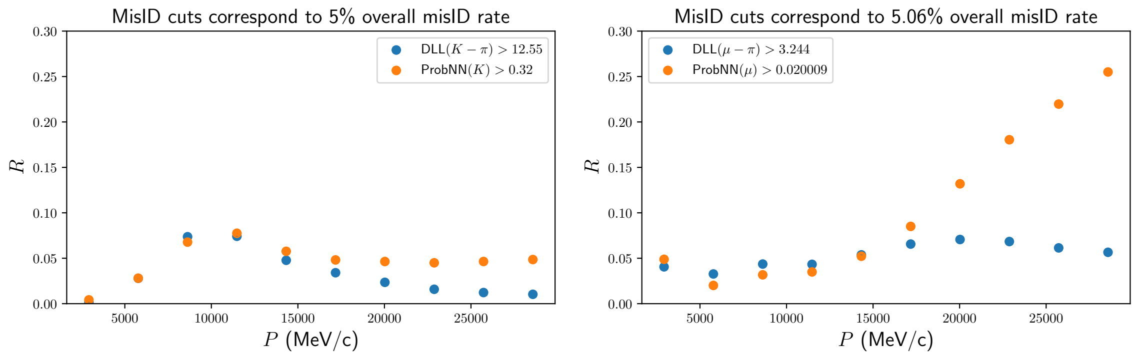
π^+ misidentified as	DLL cut	ProbNN cut	Global misidentification rate
K^+	$DLL(K-\pi) > 12.55$	ProbNN$(K) > 0.32$	5.0%
μ^+	$DLL(\mu-\pi) > 3.244$	ProbNN$(\mu) > 0.020009$	5.06%

Cuts on the DLL and ProbNN were found that gave a roughly 5% misidentification rate over the whole sample. These cuts are given in Table 1. The method for calculating misidentification rates using the sPlot technique is analogous to that outlined in Sec. 2.3. The data was first binned in P and η, and sWeights summed in each bin as before, giving the quantity N. The cuts listed in Table 1 were then applied to each bin, and the sWeights in each bin re-counted after these cuts to give the quantity k. The misidentification rate in each bin was then given by the ratio k/N. As before, the misidentification rates were cross-checked by an analogous fit-and-count procedure, which was also used to estimate systematic bias.

5.2 Misidentification rates

Figure 9 shows the misidentification rate of π^+ as K^+ and μ^+ as a function of P. There was no significant observed difference in the misidentification rate as a function of η between using a cut to

DLL or ProbNN, so we do not discuss it here. We observe that in the momentum range 15000-30000 MeV$/c$, the PID performance of ProbNN is worse than that of DLL, with approximate consistency below this range. Details of the systematic biases present in these results are given in Appendix A.



(a) Misidentification rate of π^+ as K^+

(b) Misidentification rate of π^+ as μ^+

Figure 9: Misidentification rates (R) of π^+ as μ^+ and K^+. ProbNN showed worse PID performance compared to the DLL at higher momenta. Error bars show $\sigma_{\text{stat}} \oplus \sigma_{\text{add}}$, and are too small to see. See Appendix A for more details on the calculation of systematic bias and σ_{add} for these results.

The cause of the worse performance of ProbNN relative to the DLL is unclear. It is possible that ProbNN performed better in LHCb Run 2, but the simulation samples for training the neural network in Run 3 require adjustment to provide better performance.

6 Conclusion

There are two notable limitations of this work. Firstly, measures of PID performance were only considered as functions of P and η. We would expect PID performance measures to also depend on the multiplicity, as a busier detector will have more photons in the RICH system. This is because the presence of more hits on the photon detectors makes it harder to definitively say which hits belong to which ring, leading to reduced confidence in the PID hypothesis. An analysis of PID performance with photon detector occupancy was beyond the scope of this investigation, and analyses of this sort with other calibration samples reveal less dependence and fewer interesting features compared to analyses that focus on P and η. As a result, additional dependencies were not considered here.

Secondly, it is possible that the errors and systematic biases were underestimated for the efficiency results shown in Sec. 3.2, as only a single PID cut was used to compute the errors and systematic biases in those results. In the interests of time it was assumed that these errors and biases were representative of those in all cuts considered: future work could perform a more in-depth calculation. It would also be worth investigating whether the assumption required for the usage of the sPlot method is violated at a low level here, given the slightly higher disagreement between results from the sPlot and FAC methods, compared to other results in this study.

To conclude, this work has been the first look into the PID performance of the $K^0_S \to \pi^+\pi^-$ calibration sample. Alongside giving insights into the purity and kinematic distribution of the sample, we have shown that the $K^0_S \to \pi^+\pi^-$ calibration sample is suitable for use in place of (or to augment) the $D^{*+} \to D^0\pi^+$ sample. The latter suffers from bad purity in the low-momentum region, and has a background that is difficult to model, requiring a complex 2D fit. The $K^0_S \to \pi^+\pi^-$ sample does not have these issues, and therefore may be used as a more reliable source of pion PID calibration in this momentum range. This can be implemented by tuning the existing prescale factors such that $K^0_S \to \pi^+\pi^-$ becomes the dominant sample in the momentum range considered here, although it is too late to do so ahead of the 2026 run of the experiment. Nevertheless, following simulation studies to improve the signal modelling and to cross-check the analysis, the $K^0_S \to \pi^+\pi^-$ calibration sample could be released for use in analyses of data collected from 2024-2026. In future work, it would be useful to investigate the disagreement between the PID efficiencies of the two samples, perhaps by running a detailed simulation. While it is likely they agree once systematic errors in the sWeights are accounted for, a simulation study could confirm this.

The comparison study between the DLL and ProbNN presented in this work will enable a more informed choice of PID discriminating variable to be made during analysis, reducing systematic errors. While we showed the DLL outperforms ProbNN for $P \in [15, 30]$ GeV/c, it is possible that ProbNN performs better at higher momentum. This should be verified, and if it is found to be the case, analyses should proceed by using the DLL to perform PID cuts in the momentum region investigated here, and ProbNN in the regions where it performs better.

References

[1] The LHCb Collaboration et al. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005. URL: https://doi.org/10.1088/1748-0221/3/08/S08005.

[2] Roel Aaij et al. "Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2". In: *EPJ Tech. Instrum.* 6.1 (2019), p. 1. DOI: 10.1140/epjti/s40485-019-0050-z. arXiv: 1803.00824. URL: https://cds.cern.ch/record/2308409.

[3] M. Adinolfi et al. "Performance of the LHCb RICH detector at the LHC". In: *The European Physical Journal C* 73.5 (May 2013). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-013-2431-9. URL: http://dx.doi.org/10.1140/epjc/s10052-013-2431-9.

[4] G Barr et al. "Particle Physics in the LHC Era". In: Oxford University Press, 2016. Chap. 4.3.4.

[5] Denis Derkach et al. "Machine-Learning-based global particle-identification algorithms at the LHCb experiment". In: *Journal of Physics: Conference Series* 1085.4 (Sept. 2018), p. 042038. DOI: 10.1088/1742-6596/1085/4/042038. URL: https://doi.org/10.1088/1742-6596/1085/4/042038.

[6] LHCb Collaboration. "LHCb detector performance". In: *International Journal of Modern Physics A* 30.07 (Mar. 2015), p. 1530022. ISSN: 1793-656X. DOI: 10.1142/s0217751x15300227. URL: http://dx.doi.org/10.1142/S0217751X15300227.

[7] Mateusz Czaja and Mikolaj Misiak. *Current Status of the Standard Model Prediction for the $B_s \to \mu^+\mu^-$ Branching Ratio.* 2024. arXiv: 2407.03810 [hep-ph]. URL: https://arxiv.org/abs/2407.03810.

[8] M. Pivk and F.R. Le Diberder. "sPlot: A statistical tool to unfold data distributions". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1–2 (Dec. 2005), pp. 356–369. ISSN: 0168-9002. DOI: 10.1016/j.nima.2005.08.106. URL: http://dx.doi.org/10.1016/j.nima.2005.08.106.

[9] R. Aaij et al. "Measurement of the CKM angle γ in $B^\pm \to DK^\pm$ and $B^\pm \to D\pi^\pm$ decays with $D \to K^0_S h^+h^-$". In: *Journal of High Energy Physics* 2021.2 (Feb. 2021). ISSN: 1029-8479. DOI: 10.1007/jhep02(2021)169. URL: http://dx.doi.org/10.1007/JHEP02(2021)169.

[10] Tomasz Skwarnicki. "A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances". PhD thesis. Cracow, INP, 1986.

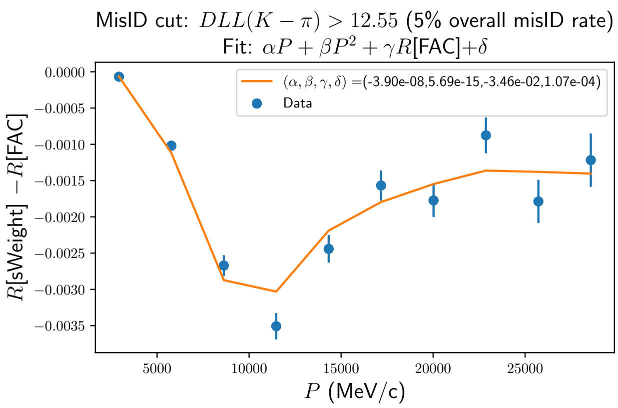
[11] URL: https://root.cern.ch/doc/master/classRooStats_1_1SPlot.html.

[12] Marc Paterno. "Calculating efficiencies and their uncertainties". In: (Dec. 2004). DOI: 10.2172/15017262.

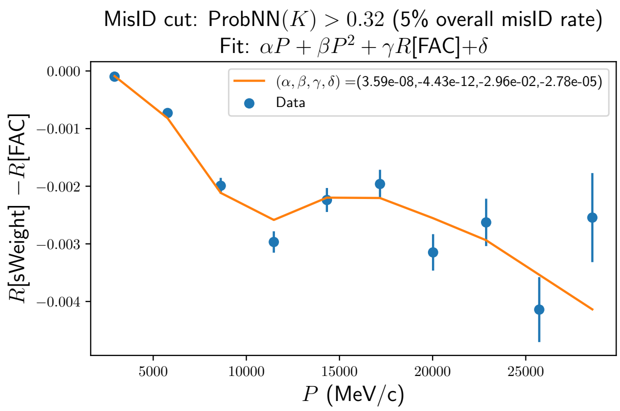
Appendix

A Details of systematic biases in the misidentification rates

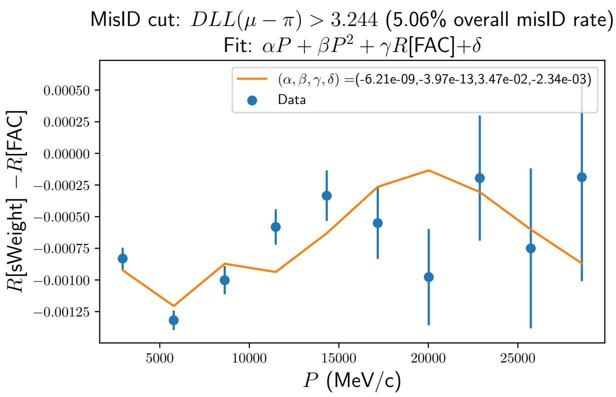
We provide details of the systematic biases present in the results of Sec. 5.2. The procedure used to calculate these biases is outlined in Sec. 2.3.2. Figure 10 shows the weighted least-squares fits performed on the differences $r_i = R_{\text{sWeight},i} - R_{\text{FAC},i}$ between the misidentification rates R as computed by the sPlot method and those from the FAC method. Here i indexes each momentum bin. The fit parameters $\alpha, \beta, \gamma, \delta$ are given in Fig. 10.



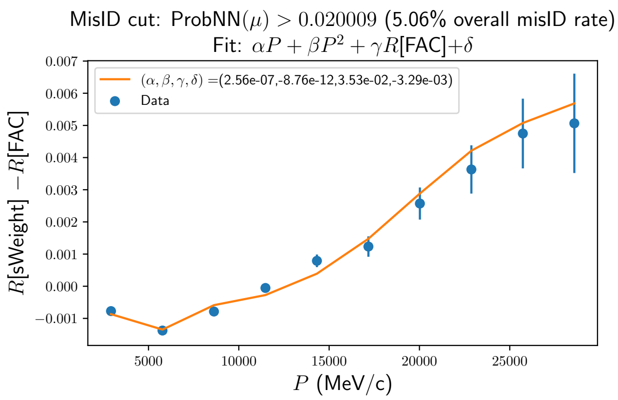
(a) Misidentification as K, DLL cut



(b) Misidentification as K, ProbNN cut



(c) Misidentification as μ, DLL cut



(d) Misidentification as μ, ProbNN cut

Figure 10: The systematic bias present in the results of Fig. 9, shown in the differences between the misidentification rates produced by the sPlot and FAC methods. This bias was modelled following the procedure set out in Sec. 2.3.2. The error bars show the statistical scatter σ_{stat}; these results were used to obtain a measure of additional scatter σ_{add} not explained by σ_{stat}, which was added in quadrature to the statistical errors in Fig. 9.