

1

BELLE NOTE 1486

2

Measurement of the decay

3

$B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$ with B2BII

4

Matic Lubej

5

Ljubljana, 2018

Changelog

- 4. May: First submission of the note.
- 16. May: Added ROE validation section.
- 7. July: update of whole BNOTE, B2BII chapter, finishing control fits.
- 19. August:
 - More details on the fit process and background splitting.
 - Adding sources of systematics
 - Ready for box opening
- 5. October:
 - Added fits to data
 - Added all sources of systematics
 - New model for uncertainty study
 - q^2 calculation description
 - efficiency as a function of m_{KK} and q^2
 - Fit yields in bins of m_{KK} and q^2

Contents

	Page
1. Introduction	1
2. Data and Monte-Carlo Samples	6
2.1. Signal MC Production	6
2.2. Control Decay	8
3. Belle to Belle II Format Conversion	9
3.1. Conversion Procedure	9
3.2. Validation	10
4. Event Reconstruction	14
4.1. Final State Particles Selection	14
4.2. Pre-selection of First B Meson Candidates	20
4.3. Loose Neutrino Reconstruction	23
4.3.1. q^2 calculation	26
4.4. Final Stage Optimization	27
4.5. Charge product categorization	30
4.6. Selection Summary	31
5. Rest of Event Clean-up	33
5.1. Clusters Clean-up	33
5.1.1. π^0 MVA Training	33
5.1.2. γ MVA Training	35
5.2. Tracks Clean-up	37
5.2.1. Tracks from Long-lived Particles	38
5.2.2. Duplicate Tracks	39
5.3. Belle Clean-up	46
5.4. Clean-up Results	46
5.5. ROE Clean-up Validation	50
6. Background Suppression	54
6.1. Resonant Background	54
6.2. Continuum Suppression	55
6.2.1. Characteristic Variables	56
6.2.2. MVA Training	58
6.3. $B\bar{B}$ Suppression	59
6.3.1. Boosting to Uniformity	61
6.4. Selection Optimization	62
6.4.1. $B\bar{B}$ Background Composition and Lepton Veto	65

56	6.5. Data and MC Agreement	69
57	6.5.1. Off-resonance Data	70
58	6.5.2. On-resonance Data	71
59	7. Extraction of Physical Parameters	73
60	7.1. Fit Setup	74
61	7.1.1. Control Fit	74
62	7.1.2. Signal Fit	77
63	7.2. Adaptive Binning Algorithm	80
64	7.3. Toy MC Experiments	81
65	7.3.1. Pseudo-Experiment: Expected Signal Yield	81
66	7.3.2. Pseudo-Experiment: Linearity Test	82
67	8. Results	84
68	8.1. Signal MC Fit Results	84
69	8.2. Control Fit Result	87
70	8.2.1. Branching Ratio Measurement for Control Decay	90
71	8.3. Signal Fit Results to Data	92
72	8.3.1. Signal Yield	92
73	8.3.2. Branching Ratio	95
74	8.3.3. Signal Distribution in bins of m_{KK}	95
75	8.3.4. Signal Distribution in bins of q^2	96
76	9. Systematic Uncertainty	98
77	9.1. Contributions to Systematics	98
78	9.1.1. PID Efficiency Correction	98
79	9.1.2. Fit Bias and Binning Effects	99
80	9.1.3. Gaussian Constraints	99
81	9.1.4. Fit Template Smearing and Offset	100
82	9.1.5. Effects of a Finite MC sample	100
83	9.1.6. MVA Selection Efficiencies	101
84	9.1.7. Model Uncertainty Effects	102
85	9.2. Summary of Systematics	105
86	A. ROE MVA Control Plots	109
87	A.1. ROE Clean-up π^0 Training	109
88	A.2. ROE Clean-up γ Training	113
89	A.3. ROE Clean-up Duplicate Pair Training	116
90	A.4. ROE Clean-up Duplicate Track Training	121
91	B. MVA Control Plots	125
92	B.1. $q\bar{q}$ Suppression Training	125
93	B.2. Standard $B\bar{B}$ Suppression Training	130
94	B.3. Uniformity Boosted $B\bar{B}$ Suppression Training	133
95	C. Other Plots	135
96	C.1. Signal Fits in m_{KK}	135
97	C.2. Signal Fits in q^2	138

Chapter 1.

Introduction

Particle physics is an established branch of physics with a rich history in theory and experiments ever since the beginning of the 20th century. So far the experimental and theoretical research have shown us hand in hand that the universe consists of particles and force carriers. Particles of matter, or elementary particles, are divided into two groups, quarks, and leptons. The quarks that we know today are called u (up), d (down), s (strange), c (charm), b (bottom) and t (top). Leptons are further split into two groups; charged leptons e (electron), μ (muon), τ (tau lepton) and their corresponding neutrinos ν_e (electron neutrino), ν_μ (muon neutrino), ν_τ (tau neutrino). Particles of force are known as gauge bosons and they are γ (photon), g (gluon), W^\pm (charged weak bosons) and Z^0 (neutral weak boson). Theoretical calculations also predicted the recently discovered Higgs boson (H), which is responsible for the mass of all particles. Some of the particles above also have a mirrored version of themselves, called antiparticles, which exhibit somewhat different properties as their un-mirrored versions.

Combinations of quarks such as $q_1 q_2 q_3$ (hadrons) or $q_1 \bar{q}_2$ (mesons) can make up heavier particles that we see today. Such particles are protons and neutrons, but also heavier particles which can be produced in processes involving very high energies. Such heavy particles are unstable and decay into lighter ones via forces of nature. Together with the elementary particles and force carriers, three out of four of these forces are joined in a theoretical model called the Standard Model (SM) [1, 2, 3, 4], which is shown in Figure 1.1. They are the electromagnetic, weak nuclear and strong nuclear force. Gravity is not included in the current version of the Standard Model due to mathematical singularities associated with the point-like nature of elementary particles. We can afford to ignore it due to its weakly interacting nature. Researching such processes in large experiments enables us to study the mechanism of how elementary particles interact. By doing so we are able to learn the secrets of the universe and how it all began.

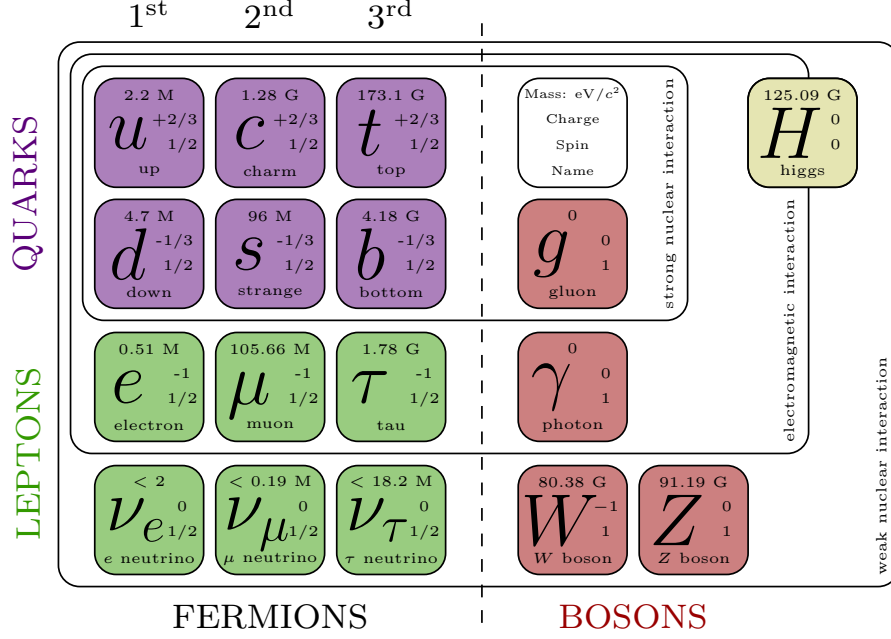


Figure 1.1.: A schematic of the Standard Model.

This analysis revolves around decays of the so-called B mesons, which are particles that consist of a b quark and a light \bar{u} or \bar{d} quark (or vice-versa). One of the most surprising features of the universe that can be studied with decays of B mesons is the CP symmetry violation (\mathcal{CP}). CP symmetry is a combination of the C symmetry (charge conjugation) and the P symmetry (spatial inversion). It states that there is no reason why processes of particles and mirrored processes of antiparticles would be different. Today we know that this does not hold for all cases and we, in fact, find processes which violate this postulate. We also know that \mathcal{CP} is very closely related to the weak nuclear force. Here lies our motivation for studying decays of B mesons, since they exhibit a rich spectrum of decays which proceed via the weak nuclear force.

One of the most important properties of the weak nuclear force is that it can change the flavor of particles. Flavor is a quantum number which is conserved for each type of quark, so changing a flavor of a quark means changing the quark itself. Such processes are forbidden for the electromagnetic and the strong nuclear force, but not for the weak one. All the information about quark transitions and transition probabilities can be merged into a form of a complex matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [5, 6]

$$V_{CKM} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix}. \quad (1.1)$$

The CKM matrix is a unitary matrix and has only four free parameters which are not described by theory. Its unitarity provides us with several mathematical identities, out of which the most famous one is

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \quad (1.2)$$

It can be represented by a triangle in the complex plane, called the unitarity triangle, shown in Figure 1.2. The sides and the angles of the unitarity triangle are

150 closely connected to the free parameters of the CKM matrix. It is important to
 151 mention that all experimental measurements depend only on these four parameters,
 152 so it is possible to determine them by measuring the angles and sides of the unitarity
 153 triangle. This way the unitarity triangle offers us a unique way to test the consistency
 154 of the SM. The ultimate goal is to then join all such measurements and overconstrain
 155 the unitarity triangle to check if all the sides meet. By improving such measurements
 156 one can check whether the SM is consistent, or if there are some contributing physics
 157 processes that we do not yet understand. Such processes are commonly referred to
 158 as "new physics" (NP). The measurements of the sides and angles of the triangle
 159 are done by using different decays out of which a large portion are B meson decays.
 160 Here lies another motivation for using B mesons in the analysis.

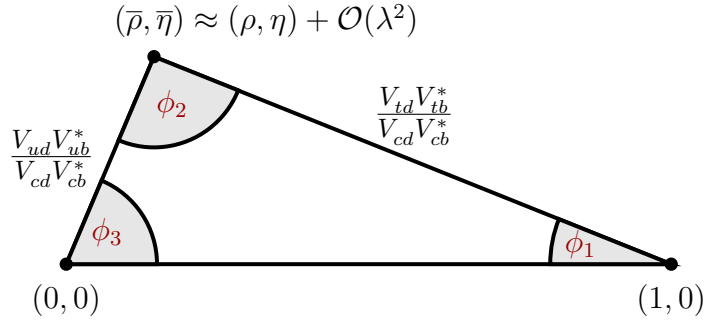


Figure 1.2.: The unitarity triangle with λ , η , ρ and A (not shown) as free parameters of the CKM matrix, shown in the Wolfenstein parametrization [7].

161 In this analysis, we focus on the V_{ub} CKM matrix element, which corresponds to
 162 $b \rightarrow u$ quark transitions. It has the smallest absolute value of all the CKM matrix
 163 elements and the largest error, so it offers the most room for improvement. Such
 164 quark transitions are present in charmless semileptonic B meson decays of the form

$$B^+ \rightarrow X_u^0 \ell^+ \nu_\ell, \quad (1.3)$$

165 where X_u^0 represents a charmless hadron with a u quark and ℓ is one of the charged
 166 leptons e , μ or τ . Measuring the decay rate of the B meson in such decays paves the
 167 way for the CKM matrix element determination. Decay rates are directly connected
 168 to the V_{ub} element as

$$d\Gamma \propto G_F^2 |V_{ub}|^2 |L^\mu \langle X_u | \bar{u} \gamma_\mu \frac{1}{2} (1 - \gamma_5) b | B \rangle|^2, \quad (1.4)$$

169 where Γ is the decay width, G_F is the Fermi coupling constant, L^μ is the leptonic
 170 current and the expression in the Dirac brackets is the hadronic current. The factor
 171 $|V_{ub}|^2$ represents the probability for the $b \rightarrow u$ quark transition. Measurement
 172 of the V_{ub} CKM matrix element can be performed in two possible ways, with the
 173 exclusive or the inclusive method, both of which are described below. Both methods
 174 require different experimental and theoretical techniques, so they provide largely
 175 independent determinations of $|V_{ub}|$. Currently, both methods also have comparable
 176 accuracies.

177 In the exclusive method, one studies the decays of B mesons to a specific charmless
 178 hadronic final state, such as $B \rightarrow \pi \ell \nu$. Clean determination of the V_{ub} is possible due

to precise experimental measurements along with reliable theoretical calculations. However, theoretical calculations are more challenging for decays to a specific final state, since hadronization of quarks has to be taken into account. There are also two main experimental challenges in this method. One has to reduce the abundant background from $B \rightarrow X_c \ell \nu$ processes since the $b \rightarrow c$ quark transition is much more common. The second experimental challenge is to separate the B meson decay with the specific charmless hadronic final state from other $B \rightarrow X_u \ell \nu$ decays since it roughly populates the same regions of the phase-space as the signal decay.

In the inclusive method, one studies the decays of B mesons to any charmless hadronic final state $B \rightarrow X_u \ell \nu$. In this case, the total decay rate for $b \rightarrow u \ell \nu$ can be calculated accurately since hadronization does not have to be taken into account. The greater challenge with this method is again the experimental measurement of the total decay rate due to the $B \rightarrow X_c \ell \nu$ background. Experimental sensitivity to V_{ub} is highest where $B \rightarrow X_c \ell \nu$ decays are less dominant. Theory and experiment have to compromise and limit the V_{ub} determination to a region where the signal-to-background ratio is good. Theoretical calculations take this into account by reliably calculating the partial decay rate $\Delta\Gamma$, which is more challenging than the total decay rate. One possible and often used approach to reduce $b \rightarrow c$ background is to reject all events with K particles, or kaons, present in the final particle selection. The procedure is called a K -veto. Kaons consist of an s quark, which is mainly produced in $c \rightarrow s$ transitions. This means that if a kaon is found in the event, it is very likely that it originates from a particle with a c quark, indicating the $b \rightarrow c$ process.

If V_{ub} is determined with both these methods, the values can be compared. It turns out that the consistency between these two results is only marginal, where the difference is at a level of 3σ . The current world averages [8] of the exclusive (from $B^0 \rightarrow \pi^- \ell^+ \nu$) and inclusive (GGOU collab. [9]) are

$$|V_{ub}|_{\text{excl.}} = (3.65 \pm 0.09 \pm 0.11) \times 10^{-3}, \quad (1.5)$$

$$|V_{ub}|_{\text{incl.}}^{\text{GGOU}} = (4.52 \pm 0.15 \pm_{-0.14}^{+0.11}) \times 10^{-3}, \quad (1.6)$$

where the first and the second errors are the experimental and the theoretical error, respectively. We see that inclusive measurements prefer higher values than exclusive ones. This is known as the V_{ub} puzzle. It is necessary to make further research as to why this difference occurs. The reason could be an unknown experimental or theoretical error, or it is even possible that some NP contributions occur. This analysis will focus on a possible reason that could be hidden in the selection mentioned before. By performing a K -veto, one discards all events with kaons in the final state in order to suppress $b \rightarrow c$ contributions. In this analysis, we focus on the charged $B \rightarrow K K \ell \nu$ decay, which is very similar to the $B \rightarrow \pi \ell \nu$, except for a production of an $s\bar{s}$ quark pair, which then combines with final state quarks to form kaons, as shown in Figure 1.3. In this case, we have kaons in the final state where the B meson decayed via a $b \rightarrow u$ process. Such decays were discarded in previous V_{ub} determinations with the inclusive method, but in principle, they contribute to the result and should be taken into account. The results of this analysis should help us make a step closer to solving the V_{ub} puzzle.

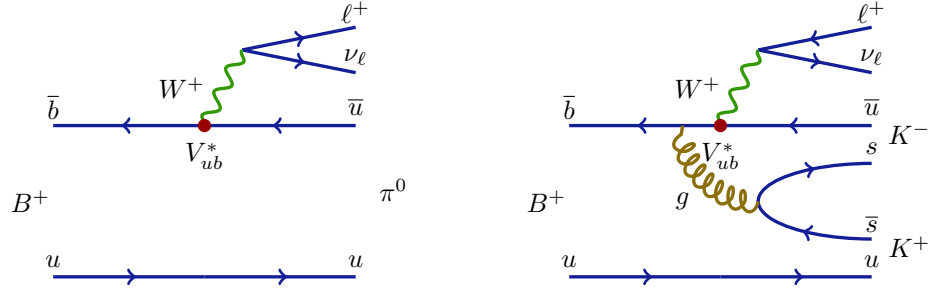


Figure 1.3.: Feynman diagrams for the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decay (left) and the $B^+ \rightarrow K^- K^+ \ell^+ \nu_\ell$ decay (right).

217 Specifically, we will be focusing on decays of the charged B mesons of the form
218 $B^+ \rightarrow K^+ K^- \ell^+ \nu$, since it includes two charged kaons, as opposed to the case of the
219 neutral B meson decay. The reason for this is a simpler decay chain and a higher
220 reconstruction efficiency. All further occurrences of $B \rightarrow KK\ell\nu$ automatically imply
221 decays of the form $B^+ \rightarrow K^+ K^- \ell^+ \nu$ and its charge conjugated counterpart.

Chapter 2.

Data and Monte-Carlo Samples

The Belle detector acquired a dataset of about $L_0 \approx 710 \text{ fb}^{-1}$ of integrated luminosity in its lifetime at the $\Upsilon(4S)$ energy of 10.58 GeV, which corresponds to about $771 \times 10^6 B\bar{B}$ meson pairs. Additionally, several streams of Monte-Carlo (MC) samples were produced, where each stream of MC corresponds to the same amount of data that was taken with the detector. The main focus of this and other similar analyses is to study a rare signal decay, which means that the amount of such decays in the existing MC is not abundant enough. In such cases, it is a common practice to produce specific samples of signal MC, where the abundance of signal decays is much larger, enabling us to study its properties in greater detail.

The following samples were used in this analysis

- data

- Belle on-resonance dataset of about L_0 integrated luminosity, measured at $\Upsilon(4S)$ resonance energy,
- Belle off-resonance dataset of about $1/10 \times L_0$ integrated luminosity, measured at 60 MeV below $\Upsilon(4S)$ resonance energy,

- signal MC, corresponding to about $400 \times L_0$,

- other MC

- generic on-resonance, 10 streams of B^+B^- and $B^0\bar{B}^0$ (denoted as **charged** and **mixed**) and 6 streams of $q\bar{q}$ produced at $\Upsilon(4S)$ resonance energy, where each stream corresponds to L_0 ,
- generic off-resonance, 6 streams of $q\bar{q}$ produced at 60 MeV below $\Upsilon(4S)$ resonance energy, where each stream corresponds to $1/10 \times L_0$,
- $B \rightarrow X_u \ell \nu$ (denoted as **ulnu**), not included in previous MC samples, equal to an amount of $20 \times L_0$,
- other rare B meson decays (denoted as **rare**), not included in previous MC samples, equal to an amount of $50 \times L_0$.

2.1. Signal MC Production

The signal MC sample of $B^+ \rightarrow K^+ K^- \ell \nu_\ell$ and the charge conjugated B^- decays was produced using the `mcproduzh` [10, 11] package for producing Belle MC. The package accepts a decay file, which describes the decays to be generated. The decay file used for signal MC generation was the same as for the **ulnu** sample since it includes the decays of interest. An additional skim was applied in order to select

only events of interest with at least 2 kaons and a light lepton, all coming from the same particle. This decreases the CPU consumption during the detector simulation and reconstruction.

The relevant processes which contribute to our signal decay are

- $B^+ \rightarrow a_{00}\ell^+\nu_\ell$,
- $B^+ \rightarrow a_{20}\ell^+\nu_\ell$,
- $B^+ \rightarrow f_2\ell^+\nu_\ell$,
- $B^+ \rightarrow f_0\ell^+\nu_\ell$,
- $B^+ \rightarrow X_u^0\ell^+\nu_\ell$,

where a_{00} , a_{20} , f_2 and f_0 are light unflavored states which include further decays into a K^+K^- pair, and X_u^0 represents a generic $u\bar{u}$ quark pair, which further hadronizes based on the PYTHIA quark hadronization model [12]. Figure 2.1 shows the invariant mass of the KK pair from various contributions of the MC generator. The light unflavored states have small contributions with resonant structures, while KK pairs from the X_u^0 state are more frequent and follow a wider and smoother distribution.

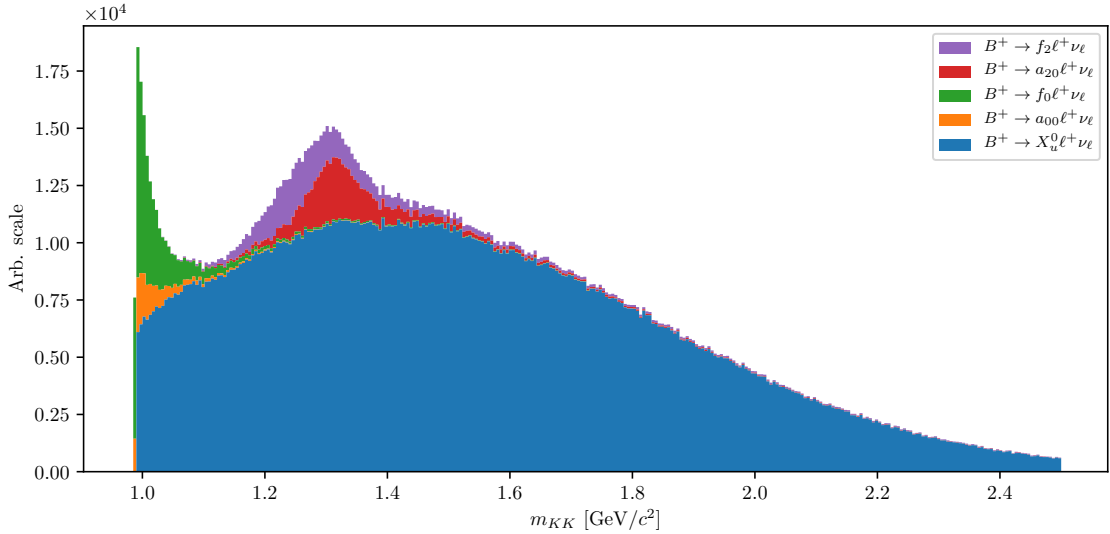


Figure 2.1.: Invariant mass of the KK pair from various contributions of the MC generator. The light unflavored states have small contributions with resonant structures, while KK pairs from the X_u^0 state are more frequent and follow a wider and smoother distribution.

The produced signal MC sample contains decays of the form $B \rightarrow KK\ell\nu$ as well as $B \rightarrow KKX\ell\nu$, where X can be any hadron as long as it satisfies all the selection rules of the decay. It is possible to calculate the MC branching ratios for each channel by making combinations of the particles directly from the generator. Table 2.1 shows some of the most prominent channels, which are similar to our signal decay, as well as their relative fraction. It is clear that our signal decay is the most abundant one, with a relative contribution of about 28 %, while other

channels contribute only up to about 8 % or less. Additionally, our signal decay is the cleanest, while other decays include neutral particles like π^0 , which are harder to reconstruct and suffer from a decrease in efficiency due to reconstruction effects.

Channel	Ratio [%]	Channel	Ratio [%]
K^+K^-	28.14	$K^+K^-\rho^0$	1.93
$K^+K^-\pi^0$	8.94	$K^+\bar{K}^0\rho^-$	1.84
$K^+\bar{K}^0\pi^-$	8.71	$K^0K^-\rho^+$	1.83
$K^0K^-\pi^+$	8.70	$K^0\bar{K}^0\rho^0$	0.00
$K^+K^-\pi^+\pi^-$	4.15	$K^+K^-\pi^0\pi^0$	0.86
$K^0\bar{K}^0$	3.32	$K^+K^-\pi^+\rho^-$	0.69
$K^0\bar{K}^0\pi^0$	3.26	$K^+K^-\rho^+\pi^-$	0.68
$K\bar{K}$ pair with η	7.08		
$K\bar{K}$ pair with ω	5.33		
Other	14.53		

Table 2.1.: Relative branching ratios of $B \rightarrow KKX\ell\nu$ decays by channel.

We generate about 1.3×10^9 events of the form $B \rightarrow X_u\ell\nu$, which corresponds to an integrated luminosity of about $L = 400 \times L_0$, where this value was obtained by normalizing the signal MC to the amount of signal in the $B \rightarrow X_u\ell\nu$ MC sample. This amounts to a total of about 9.37×10^6 generated signal events, and to a branching ratio

$$\mathcal{B}(B^+ \rightarrow K^+K^-\ell^+\nu_\ell)_{MC} = 1.53 \times 10^{-5}, \quad (2.1)$$

where ℓ is e or μ . During analysis, the abundant signal MC sample is scaled down to correspond to the amount of data taken with the Belle detector.

2.2. Control Decay

In this analysis, we are also able to define another B meson decay which occupies almost the same phase space as our signal decay. This process can be used for the monitoring of our analysis steps, which are applied to both measured and simulated data. Any kind of difference between the two might indicate our procedure to be fine-tuned to simulated data, or some other similar problem.

We define a control decay of the form

$$B^+ \rightarrow \bar{D}^0\ell^+\nu, \quad D^0 \rightarrow K^+K^-,$$

which is much more abundant and, most importantly, easy to suppress since it only populates a very narrow region in the kaon invariant mass spectrum. Due to no extra particles in the D^0 decay, the kaon invariant mass is equal to $m_{KK} \approx m_{D^0}$ up to very good precision. By excluding this narrow region we discard the majority of the control candidates while discarding only a small amount of the signal candidates. A more quantitative description of suppressing control and other background candidates is written in chapter 6.

Chapter 3.

Belle to Belle II Format Conversion

3.1. Conversion Procedure

The Belle experiment finished its data-taking run of 10 years at the end of 2010 after collecting a dataset of about 1 ab^{-1} . That year the Belle detector was shut down and the Belle II experiment has started in its place. While the focus moved to the construction of the Belle II detector and the development of the Belle II Analysis Framework (BASF2) [13], Belle analyses are still on-going and Belle data is still being used today. BASF2 software with its modular structure has a more intuitive approach to performing analyses, however, since it was rewritten completely from scratch, it was designed for the incoming Belle II data and therefore out-of-the-box usage of Belle data is outside of its scope.

In the Belle II Collaboration, a task force was created in order to convert Belle data into Belle II format (B2BII) [14]. The B2BII package was developed as a part of BASF2 in order to convert data and MC of the Belle experiment and make it available within BASF2. In addition to the convenience of Belle data being processed in the more intuitive and advanced BASF2 framework, B2BII allows for estimation and validation of performances of various advanced algorithms being developed for Belle II. The conversion itself, however, is considered non-trivial. Although the conversion of the raw detector data would be possible, the reconstruction algorithms of BASF2 are optimized for Belle II and cannot be effectively applied to Belle data. To bypass this problem, reconstructed objects from PANTHER tables, a custom solution of the Belle collaboration based on C/C++ and Fortran, are mapped to their corresponding representations in BASF2. In this analysis, we use the developed converter package in order to analyze Belle data with the Belle II software.

The conversion in the B2BII package is divided into three BASF2 modules. The first module opens the Belle input files and reads the events into memory in the form of PANTHER tables. This module consists predominantly of reused BASF code. The second module applies various calibration factors, such as experiment- and run-dependent factors, to the beam energy, particle identification information, error matrices of the fitted tracks, etc. The module also applies some low-level cuts to reproduce removing background events as done within BASF. The actual conversion and the mapping of reconstructed objects are done in the last module. For more information see [15].

3.2. Validation

In order to make sure the conversion was successful and without errors, a thorough validation should be performed. This is done by comparing histograms of all physical quantities of the reconstructed objects on simulated and recorded events, processed with BASF and BASF2.

Our signal decay mode consists of three charged tracks, track conversion should perform flawlessly. Additionally, energy measurement is also very important in our untagged analysis in order to successfully determine the missing 4-momentum in the event, which is why we also need a correct conversion of the ECL clusters for photons and π^0 particles. Figures 3.1 to 3.3 show the basic physical properties of converted tracks, photons and π^0 particles, obtained with BASF and BASF2, and their difference, which is (up to numerical precision) equal to 0. The plots indicate that the conversion is successful in all aspects and we can proceed with the analysis in the framework of BASF2.

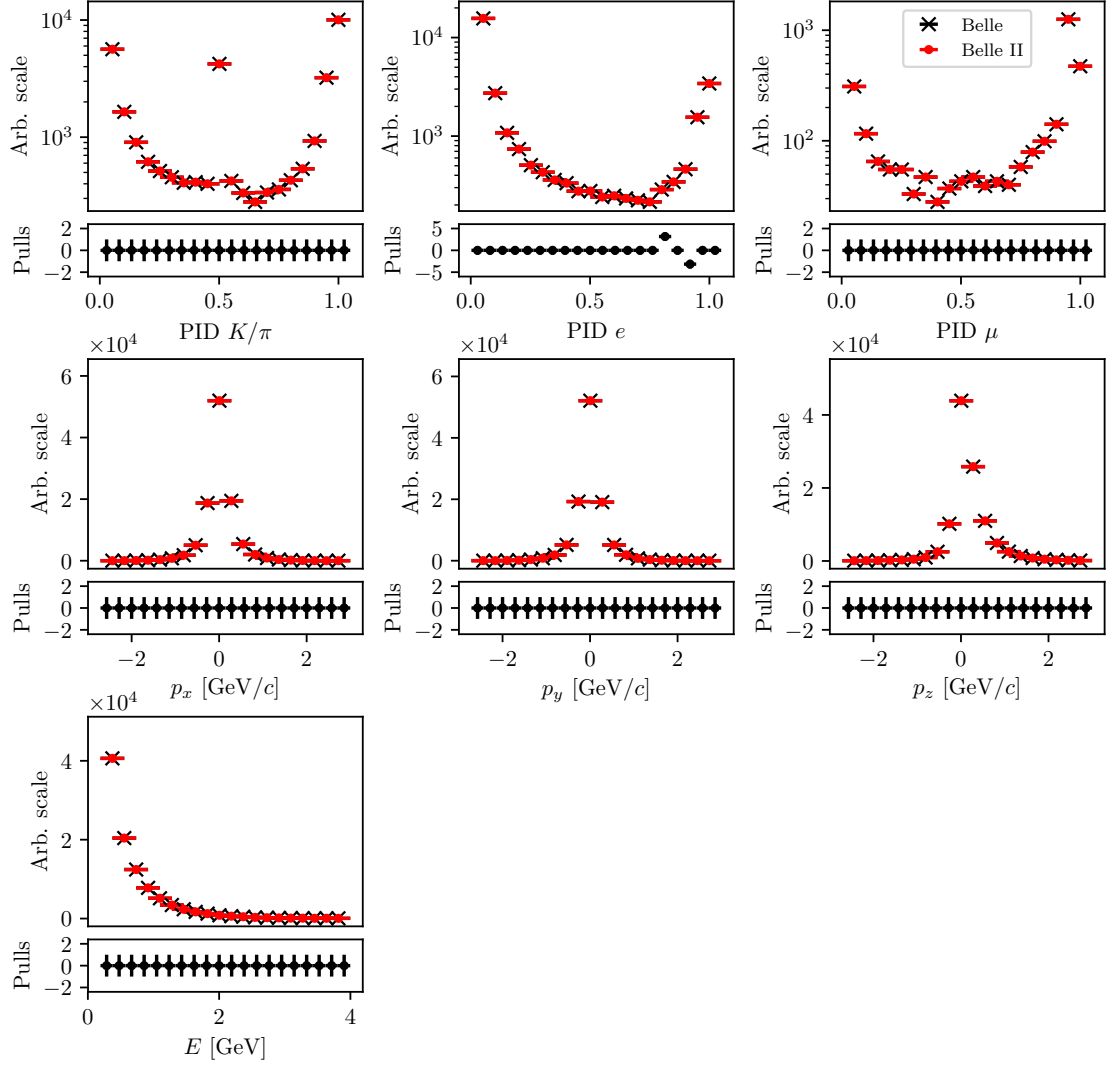


Figure 3.1.: Some of the more important physical properties of tracks for Belle and Belle II in the conversion process. The histograms seem to overlap and the conversion is assumed to be successful.

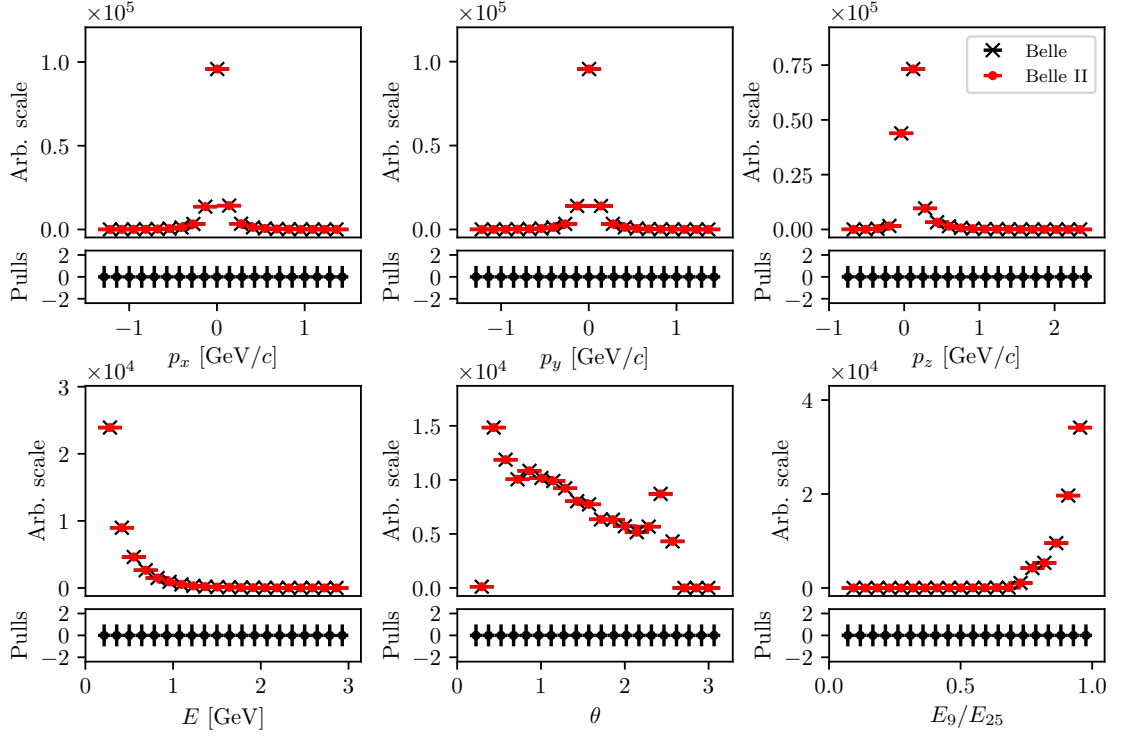


Figure 3.2.: Some of the more important physical properties of photons for Belle and Belle II in the conversion process. The histograms seem to overlap and the conversion is assumed to be successful.

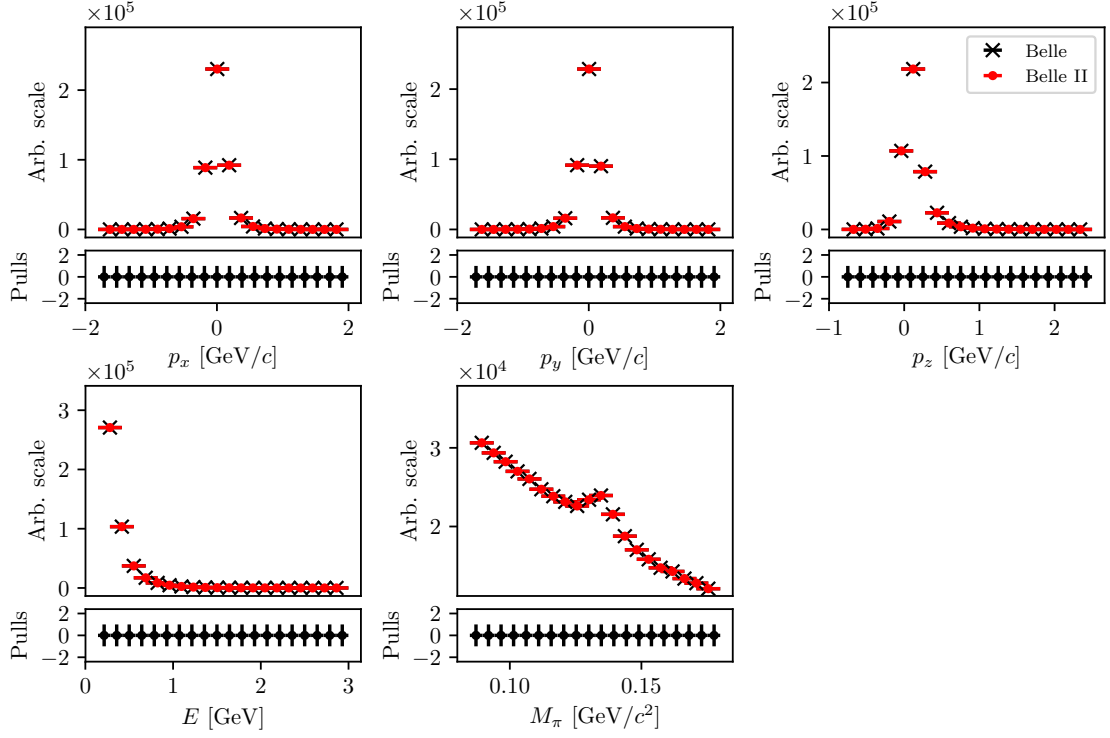


Figure 3.3.: Some of the more important physical properties of π^0 particles for Belle and Belle II in the conversion process. The histograms seem to overlap and the conversion is assumed to be successful.

Chapter 4.

Event Reconstruction

In this chapter the procedure for event reconstruction of the B meson decay $B \rightarrow KK\ell\nu$ is shown, starting with final state particle selection and then combining them to obtain B meson candidates.

4.1. Final State Particles Selection

Since the neutrino escapes detection, we can only reconstruct the charged tracks in the decay, which are the two charged kaons (K) and the light lepton, which is the electron (e) or muon (μ). These are some of the particles which are commonly referred to as final state particles (FSP). Final state particles have a long lifetime and are usually the particles that we detect when they interact with the material in the detector.

It is important to limit our selection of FSP particles in order to cut down the number of particle combinations, and consequentially computation time and file sizes.

At this point in the analysis, we do not apply any intelligent cuts yet, which results in a large number of available particles and their combinations. In order to minimize this effect, we perform this part of the study on a smaller subset of the available generic MC, experiment no. 23 and 31, which correspond to an integrated luminosity of 6.273 fb^{-1} and 17.725 fb^{-1} , respectively. We chose these two experiments to get closer to the appropriate ratio of SVD1 and SVD2 data in the full Belle MC.

Leptons

Figures 4.1 and 4.2 show the impact parameters d_0 and z_0 , the momentum in $\Upsilon(4S)$ center-of-mass system (CMS), and the PID information for true and fake electrons and muons from any source, where an extra category for true electrons/muons from the signal B meson decay is shown. The difference between the true leptons from any source and those from signal decays is due to the distinct kinematics of the parent's decay to the electron.

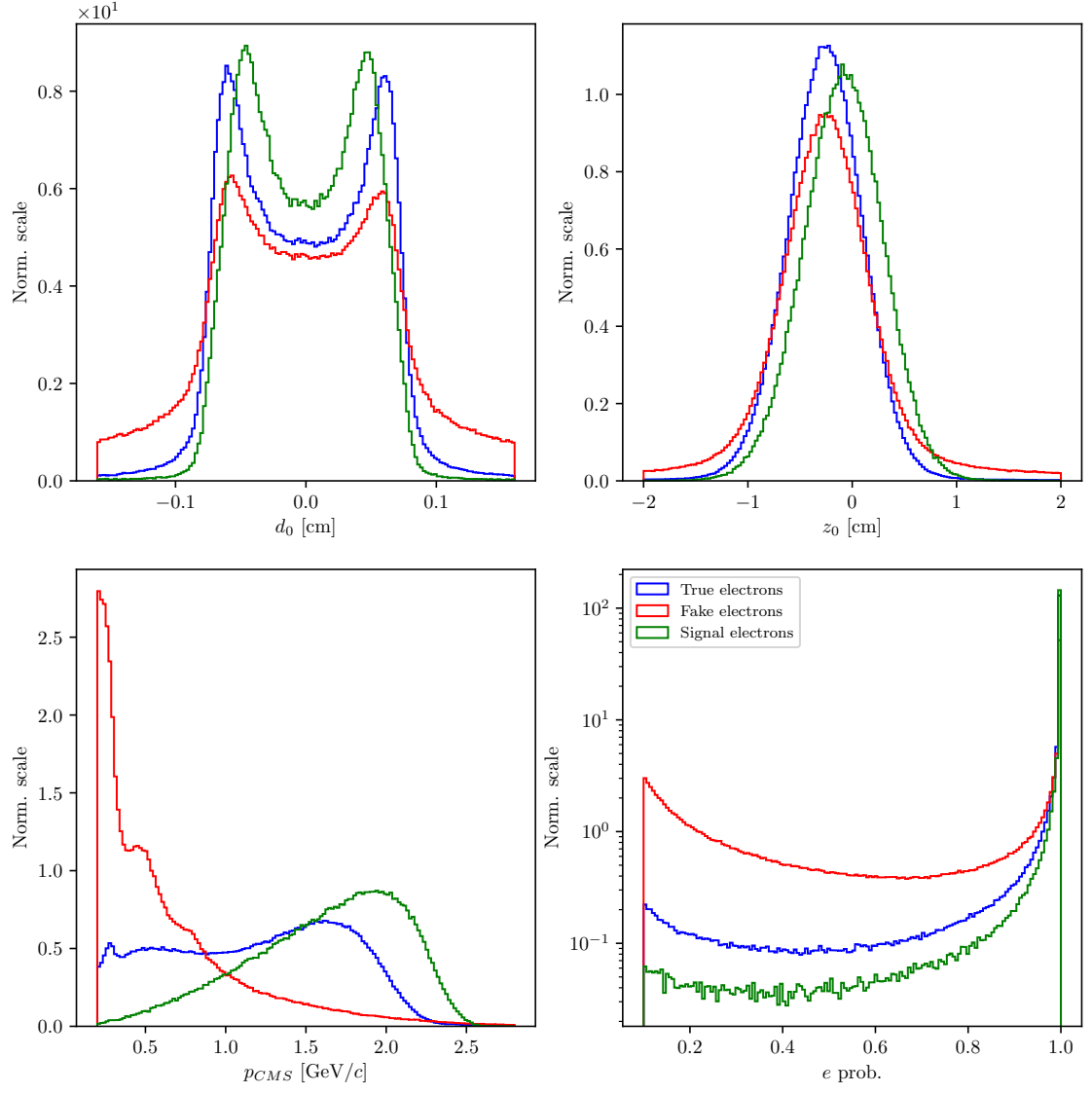


Figure 4.1.: Normalized properties of true (blue), fake (red) electrons from any source, and true electrons from signal B candidates (green).

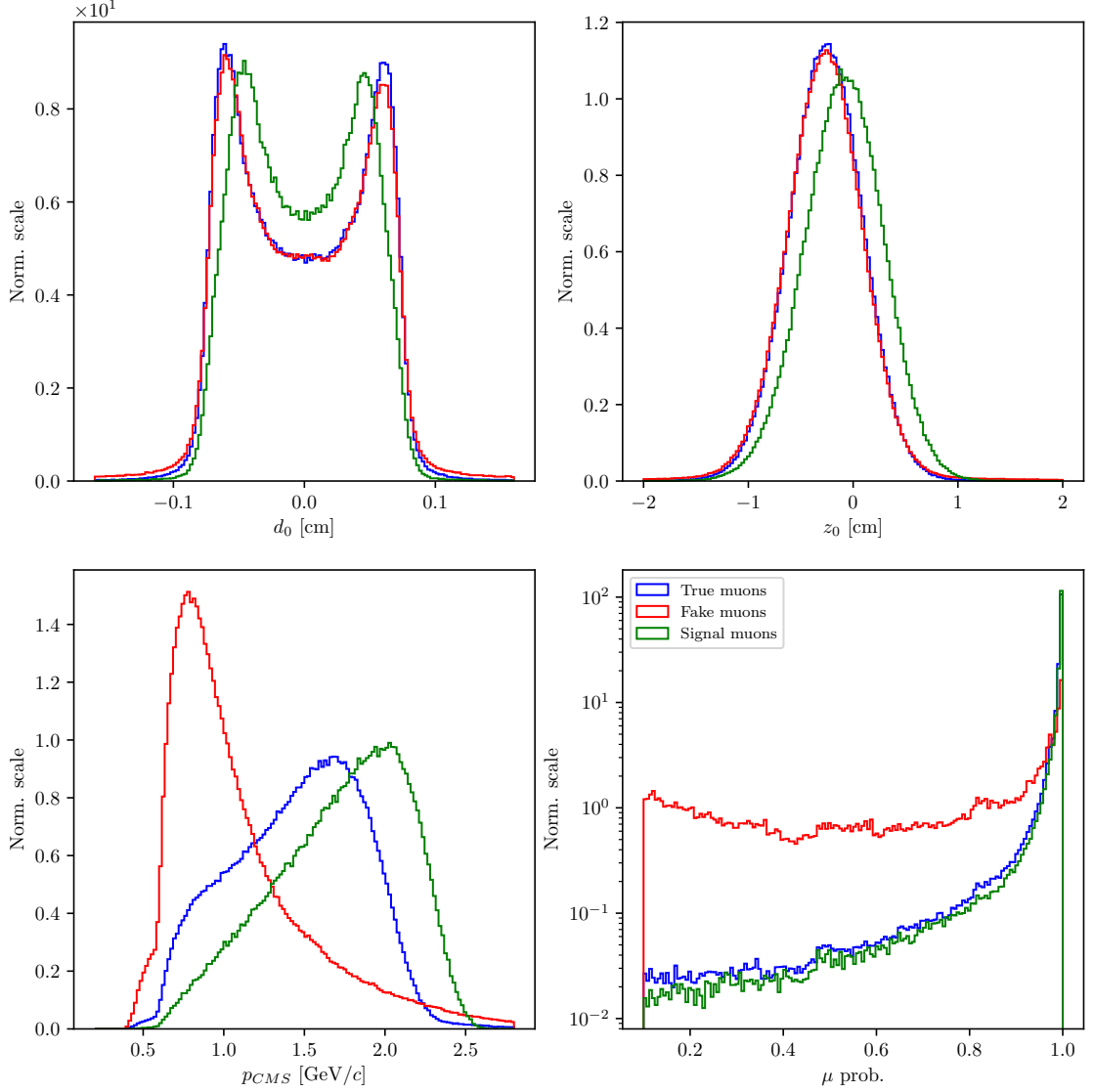


Figure 4.2.: Normalized properties of true (blue), fake (red) muons from any source, and true muons from signal B candidates (green).

Based on these distributions, we can define a set of cuts

- $|d_0| < 0.1$ cm,
- $|z_0| < 1.5$ cm,
- $p_{CMS} \in [0.4, 2.6]$ GeV/ c for electrons,
- $p_{CMS} \in [0.6, 2.6]$ GeV/ c for muons.

After this selection we can determine the optimal PID cuts for electrons and muons, where we optimize the selection by maximizing the standard definition of *figure of merit* (FOM), defined in Eq. (4.1)

$$\text{FOM} = \sqrt{\mathcal{E}\mathcal{P}} \propto \frac{S}{\sqrt{S+B}}, \quad (4.1)$$

where the argument in the square root is the product of the efficiency (\mathcal{E}) and the purity (\mathcal{P}) function. The definitions of signal (S) and background (B) are somewhat fluid throughout the analysis and need to be defined for each FOM separately. In this section we define two representations of S and B . In FOM_1 the signal S represents correctly reconstructed final state particles, while in FOM_2 the signal S represents correctly reconstructed final state particles which also come from a correctly reconstructed B meson candidate. In both cases B represents all other particle candidates which do not satisfy the conditions of S .

The FOM plots are shown in Figures 4.3 and 4.4. The cut values are based on PID cuts used for PID efficiency calibration. The optimal value for the PID cuts is equal to the largest available value, regardless of the leptons coming from signal decays or not. The optimized PID cuts for leptons are

- e prob. > 0.9 for electrons,
- μ prob. > 0.97 for muons.

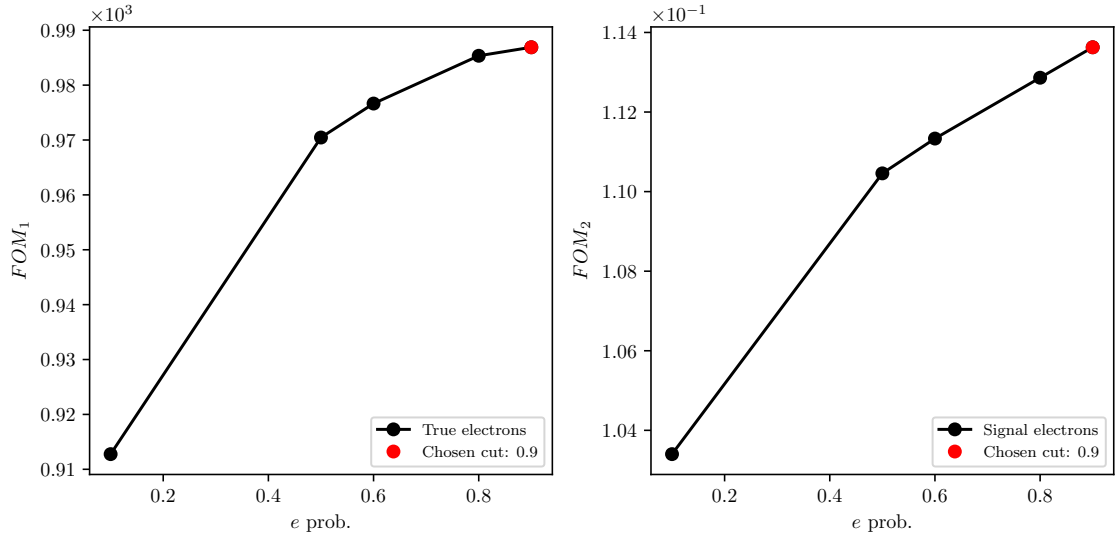


Figure 4.3.: FOM optimizations of the PID probability cuts for true electrons (left) and true electrons from signal B candidates (right).

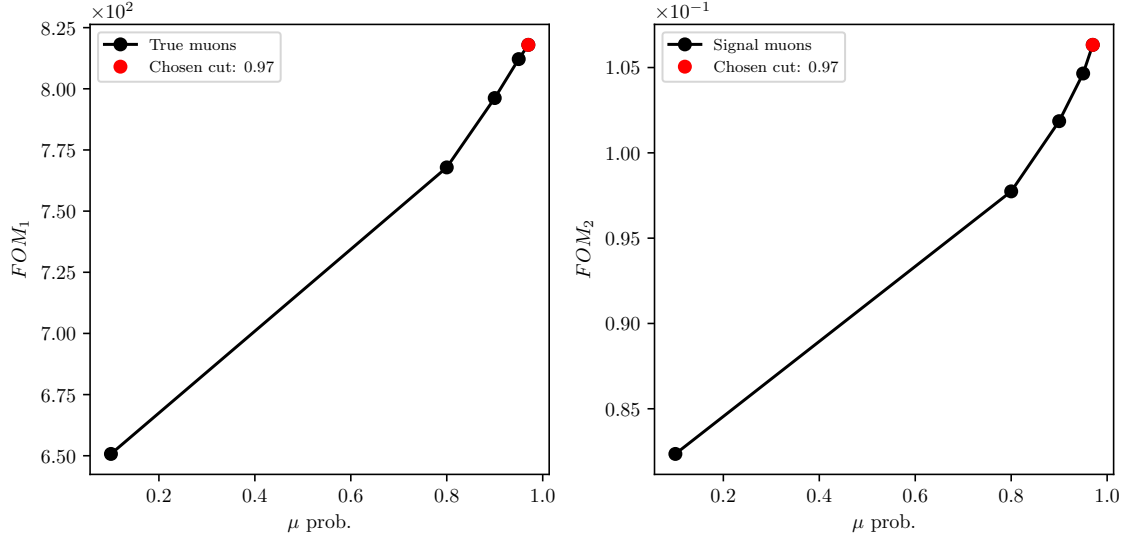


Figure 4.4.: FOM optimizations of the PID probability cuts for true muons (left) and true muons from signal B candidates (right).

400 **Kaons**

401 We repeat the procedure for both kaons. Figure 4.5 shows the impact parameters
 402 d_0 and z_0 , the momentum in $\Upsilon(4S)$ center-of-mass system (CMS), and the PID
 403 information for true and fake kaons, where an extra category for true kaons from
 404 the signal decay is shown.

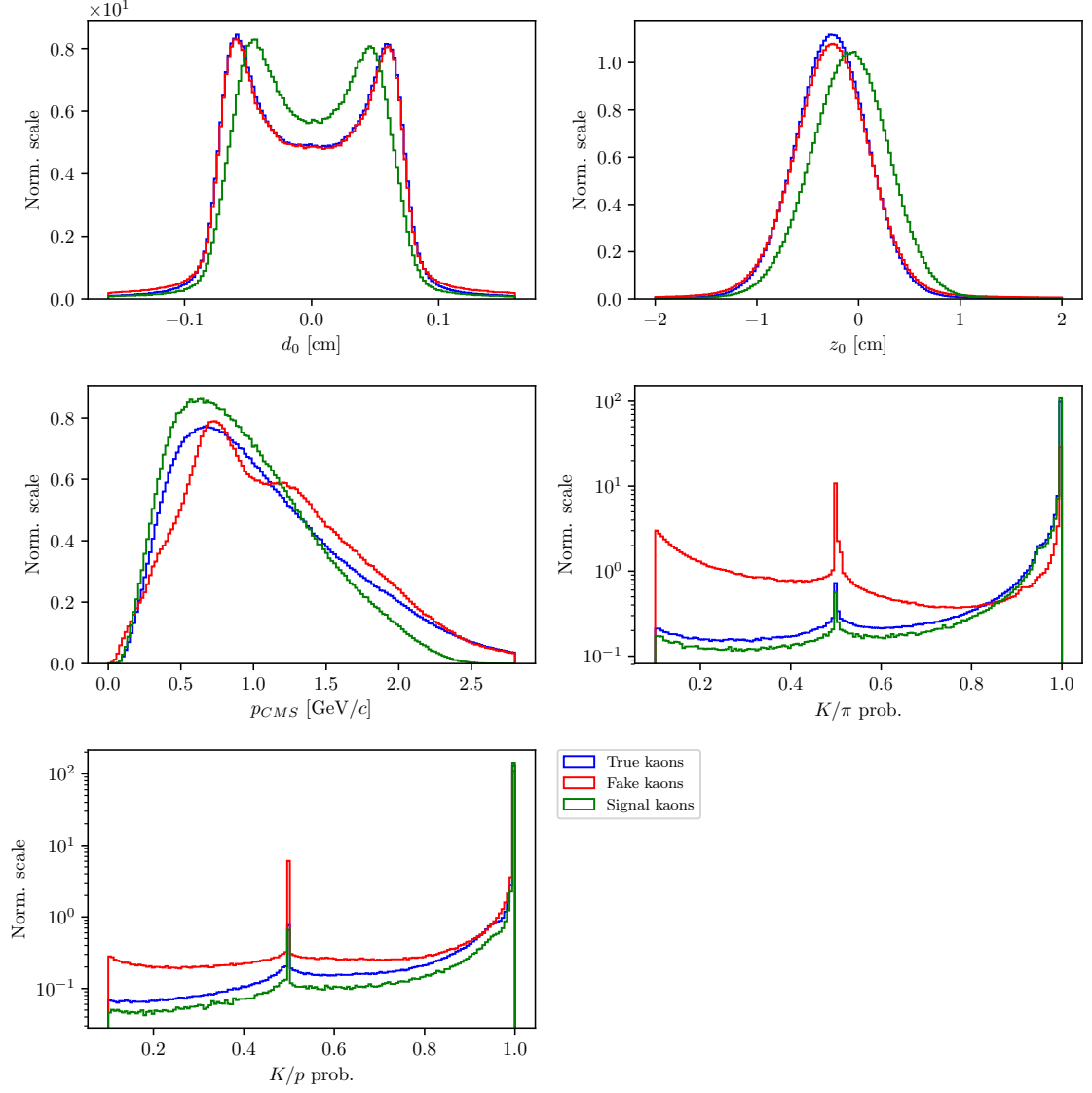


Figure 4.5.: Normalized properties of true (blue), fake (red) and true kaons (green) from signal B candidates.

We define the kaon cuts in the same manner as in the case for leptons

- $|d_0| < 0.15$ cm,
- $|z_0| < 1.5$ cm,
- $p_{CMS} \in [0, 2.5]$ GeV/ c .

The PID optimization, in this case, is taken in two steps. First, we optimize the cut on K/π , and after that the K/p separation probability. Figure 4.6 shows the optimization procedure for PID cuts on kaon candidates.

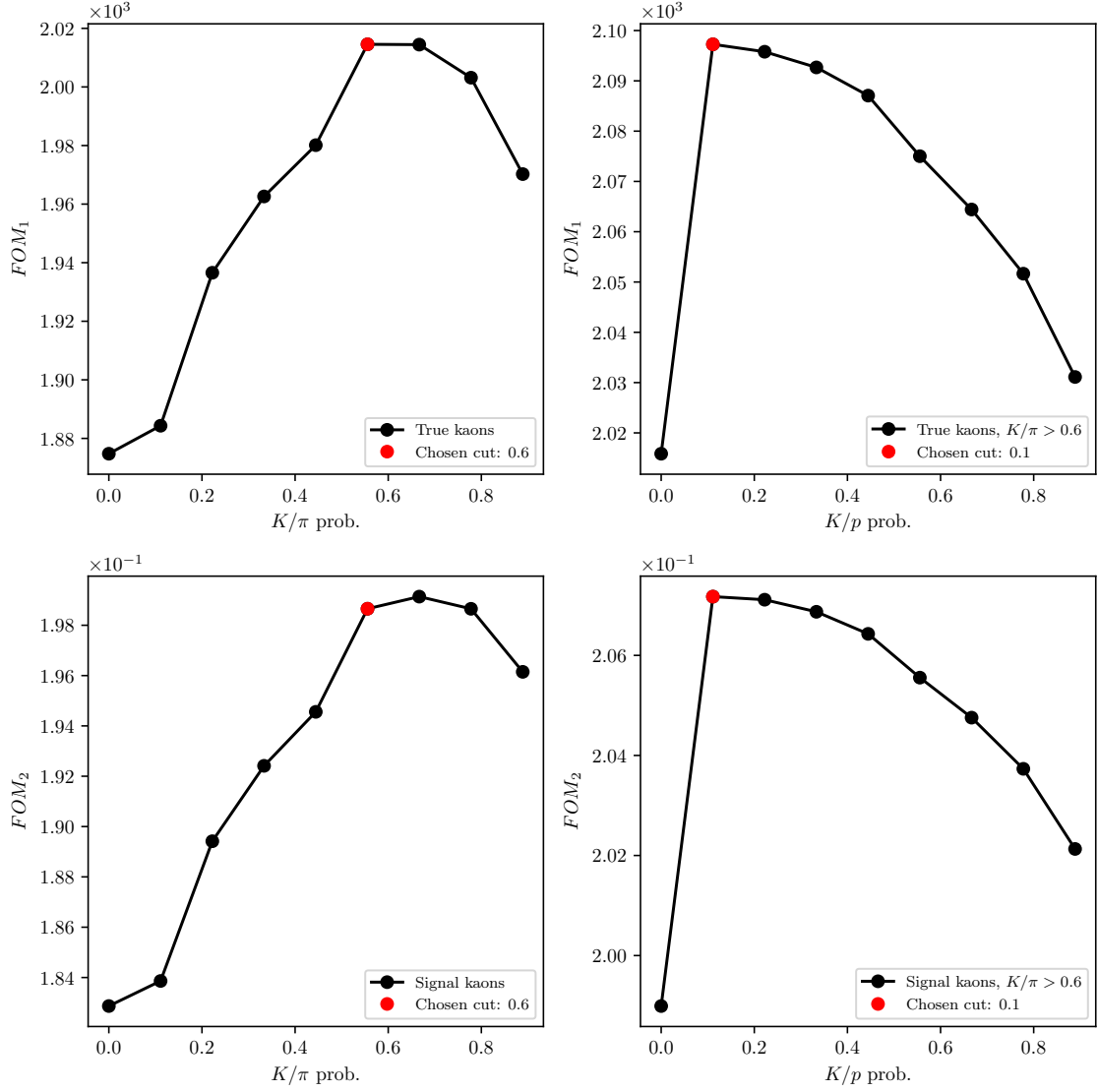


Figure 4.6.: FOM optimizations of the PID probability cuts for true kaons (top) and true kaons from signal B candidates (bottom). The plots on the left show the optimization of the first step for the K/π probability cut, and the plot on the right the K/p probability cut.

The optimized PID cuts for kaons are

- $K/\pi > 0.6$,
- $K/p > 0.1$.

4.2. Pre-selection of First B Meson Candidates

In this section, we use the charged particle candidates from the previous section to make particle combinations, which correspond to B meson candidates. When a B meson candidate is selected, additional features can be calculated and used for background rejection. Since we are still performing this part of the study on

420 a smaller subset of the full available MC sample, we will perform under-optimized
 421 cuts based on the FOM optimization in order to optimize them later on the full
 422 Belle MC sample.

Since the missing neutrino escapes detection, we reconstruct the B mesons in the following two channels

$$\begin{aligned} B^+ &\rightarrow K^+ K^- e^+, \\ B^+ &\rightarrow K^+ K^- \mu^+, \end{aligned}$$

423 and similarly for B^- . When an arbitrary combination is obtained, we perform a
 424 vertex fit of the three tracks in order to discard combinations with a low probability
 425 of tracks coming from the same point. B mesons have a relatively long lifetime and
 426 decay along the z axis of the detector in the direction of the boost, so the vertex fit
 427 is enforced with an IPTUBE constraint, which constrains the vertex to an elongated
 428 ellipsoid along beam direction. We demand that the fit converged and apply a cut on
 429 the minimal fit probability. The fit probability for signal and background B meson
 430 candidates is shown in Figure 4.7 (left). We perform a FOM cut optimization of
 431 this variable, which is shown in Figure 4.7 (right) for the subset of the Belle MC
 432 sample. In this and in the following cases, the definition of S from Eq. (4.1) are
 433 correctly reconstructed B meson candidates with a missing neutrino which are not
 434 coming from a $b \rightarrow c$ transition.

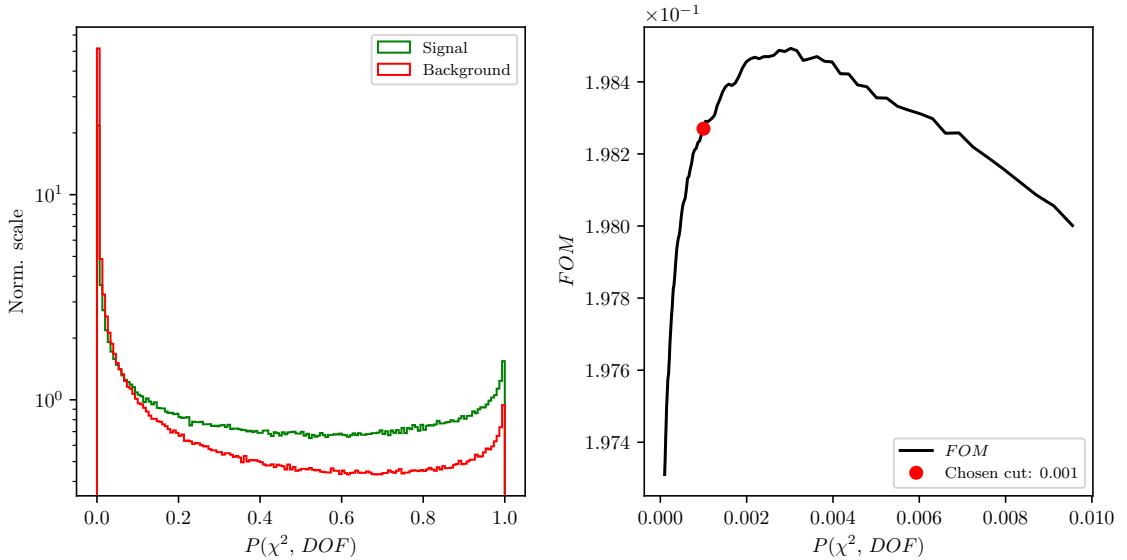


Figure 4.7.: Normalized vertex fit probability distribution for signal and background B meson candidates in logarithmic scale(left) and FOM optimization of the vertex fit probability (right) for the subset of the full Belle MC sample.

435 The chosen pre-cut on the fit probability is

- 436 • $P(\chi^2, NDF) > 1.0 \times 10^{-3}$.

437 With the neutrino being the only missing particle on the reconstructed side, it
 438 is possible to determine the angle between the direction of the reconstructed B
 439 (denoted as $Y \rightarrow KK\ell$) and the nominal B , as

$$p_\nu = p_B - p_Y, \quad (4.2)$$

$$p_\nu^2 = m_\nu^2 = m_B^2 + m_Y^2 - 2E_BE_Y + 2\vec{p}_B \cdot \vec{p}_Y \approx 0, \quad (4.3)$$

$$\cos(\theta_{BY}) = \frac{2E_BE_Y - m_B^2 - m_Y^2}{2|\vec{p}_B||\vec{p}_Y|}, \quad (4.4)$$

where p denotes a scalar, \vec{p} a vector, and p a 4-vector. All the energy and momenta above are calculated in the CMS frame. The mass of the neutrino is equal to 0 to a very good precision, so we use it in Eq. (4.3). In addition, we substitute the unknown energy and momentum magnitude, E_B and $|\vec{p}_B|$, of the B meson in Eq. (4.4), with quantities from the well known initial conditions

$$E_B = E_{CMS}/2, \quad (4.5)$$

$$|\vec{p}_B| = p_B = \sqrt{E_B^2 - m_B^2}, \quad (4.6)$$

440 where E_{CMS} is the total energy of the e^+e^- collision in the CMS frame and m_B is
 441 the nominal mass of the B meson. This improves the resolution of $\cos(\theta_{BY})$, which
 442 leads to better signal-to-background discrimination.

443 For the correctly reconstructed candidates, this variable lies in the $[-1,1]$ region,
 444 though only to a certain precision, due to the finite detector resolution. Background
 445 candidates, however, populate also the non-physical regions, as shown in Figure 4.8
 446 (left).

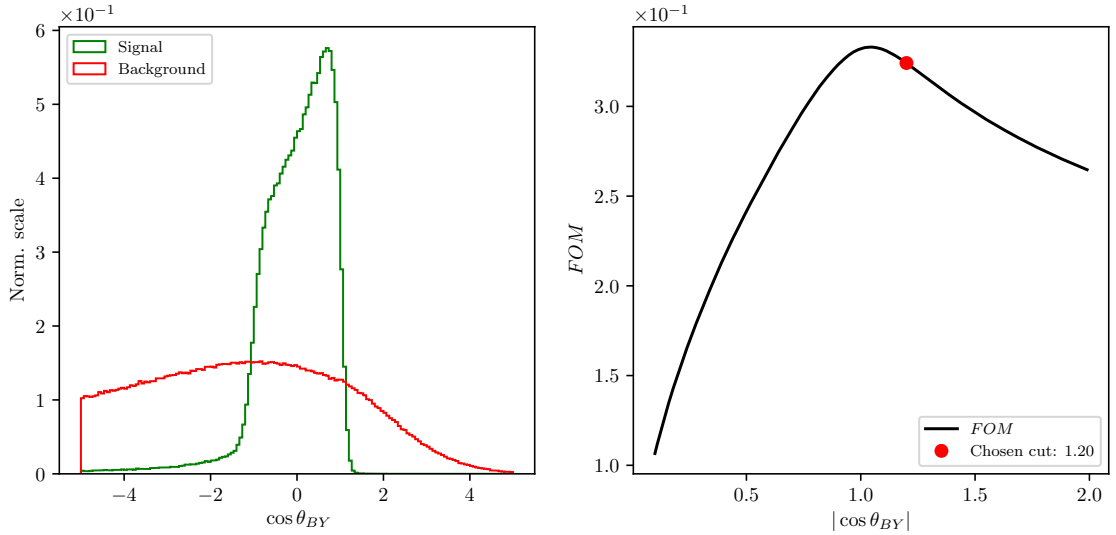


Figure 4.8.: Normalized $\cos\theta_{BY}$ distribution for signal and back-
 ground B meson candidates (left) and FOM optimiza-
 tion of the $\cos\theta_{BY}$ variable (right) for the subset of the
 full Belle MC.

447 We again impose an under-optimized cut on this variable from Figure 4.8 (right)
 448 to discard a large amount of background on this subset of the full Belle MC

449 • $|\cos(\theta_{BY})| < 1.20$.

4.3. Loose Neutrino Reconstruction

The signal-side neutrino escapes detection, so we cannot directly determine its four-momentum. However, due to the detectors geometry, which almost completely covers the full solid angle, and due to well known initial conditions of the $\Upsilon(4S)$ meson, it is possible to determine the kinematics of the missing neutrino via indirectly reconstructing the companion B meson by summing up the four-momenta of all the FSP particles in the event which were not used in the reconstruction of the signal side B meson. This is known as the *untagged* method since we are not using any kind of tagging method to reconstruct the companion B meson. The particles used in the indirect companion B meson reconstruction are also said to belong to the *rest of the event* (ROE).

Due to the beam background in the detector, material interactions, or other processes, random tracks and clusters enter our event and get reconstructed as part of the physics process we want to study. These tracks and clusters are not interesting and further spoil the data we measure. In order to remedy this, we perform an extensive clean-up of the tracks and clusters in the ROE side before calculating the four-momentum of the missing part of the event. Here we see the motivation for the ROE clean-up since our signal candidate reconstruction depends on tracks and clusters in the ROE side. The clean-up procedure is performed separately on tracks and clusters and uses multiple steps with multivariate analysis (MVA) algorithms in order to separate good tracks and clusters from the bad ones, which populate the ROE. Then, for each ROE object, a ROE mask is created for tracks and clusters, which narrates the use of this object in the final calculations of the ROE four-momentum. From this point on we assume the ROE to be efficiently cleansed of extra tracks and clusters. A more detailed description of the ROE clean-up can be found in Chapter 5.

The total missing four-momentum in the event can be determined as

$$p_{miss} = p_{\Upsilon(4S)} - \sum_i^{\text{Event}} (E_i, \vec{p}_i), \quad (4.7)$$

$$p_{miss} = p_{\Upsilon(4S)} - \left(p_Y - \sum_i^{\text{Rest of event}} (E_i, \vec{p}_i) \right), \quad (4.8)$$

where the summation runs over all charged and neutral particles in the defined set with

$$p_i^{\text{neutral}} = (p_i, \vec{p}_i) \quad \text{and} \quad p_i^{\text{charged}} = \left(\sqrt{m_i^2 + p_i^2}, \vec{p}_i \right), \quad (4.9)$$

where we assumed all neutral particles to be massless photons. For charged tracks in the ROE a mass hypothesis needs to be defined in order to determine the track's energy. After the ROE clean-up we make the following procedure of choosing the mass hypothesis

1. e , if e prob. $> \mu$ prob. and e prob. > 0.9 ,
2. otherwise μ , if μ prob. $> e$ prob. and μ prob. > 0.97 ,
3. otherwise K , if K/π prob. > 0.6 ,

485 4. otherwise π .

We define the square of the missing mass, m_{miss}^2 , which is consistent with zero, if signal-side neutrino is the only missing particle in the event, as shown in Eq. (4.11).

$$p_\nu = p_{miss} = (E_{miss}, \vec{p}_{miss}), \quad (4.10)$$

$$m_{miss}^2 = p_{miss}^2 = p_\nu^2 = m_\nu^2 \approx 0. \quad (4.11)$$

486 Since the detector is not perfect, the distribution of the m_{miss}^2 variable has a non-
 487 zero width. Additionally, a tail is introduced due to missing particles like neutrinos,
 488 other neutral undetected particles such as K_L^0 , or simply missing tracks due to
 489 detection failure. Figure 4.9 shows the distribution of m_{miss}^2 as defined with the
 490 missing four-momentum in Eq. (4.10). Correctly reconstructed candidates, which
 491 come from events where the other B meson decayed via a hadronic decay mode, peak
 492 at zero. If this is not the case, candidates are shifted to larger values of this variable,
 493 even if the event in question is a signal event. For this purpose, we define a subset of
 494 all signal candidates, which come from events where the companion B meson decayed
 495 hadronically and all of its particles were taken into account correctly. We only allow
 496 for missing photons, since they are frequently radiated due to bremsstrahlung effects
 497 from final-state electrons and they do not have such a big impact on the 4-momentum
 498 of the final candidate. We denote this subset as *perfect* signal. This subset is used
 499 to correctly define the clean-up parameters and is not used in any reconstruction
 500 steps since we cannot know in data which neutral particles are actually missing.

501 Due to this fact, we impose a cut on the m_{miss}^2 variable in order to partially discard
 502 candidates with spoiled properties, even if it was in principle a correct combination
 503 of FSP particles on the signal side. The cut was chosen based on the optimization
 504 of FOM, where in this case the definition of S were perfectly reconstructed signal
 505 candidates. The chosen cut value is

506 • $|m_{miss}^2| < 3.0 \text{ GeV}/c^2$,

507 which is also under-optimized at this point due to the same reasons as in the cases
 508 above.

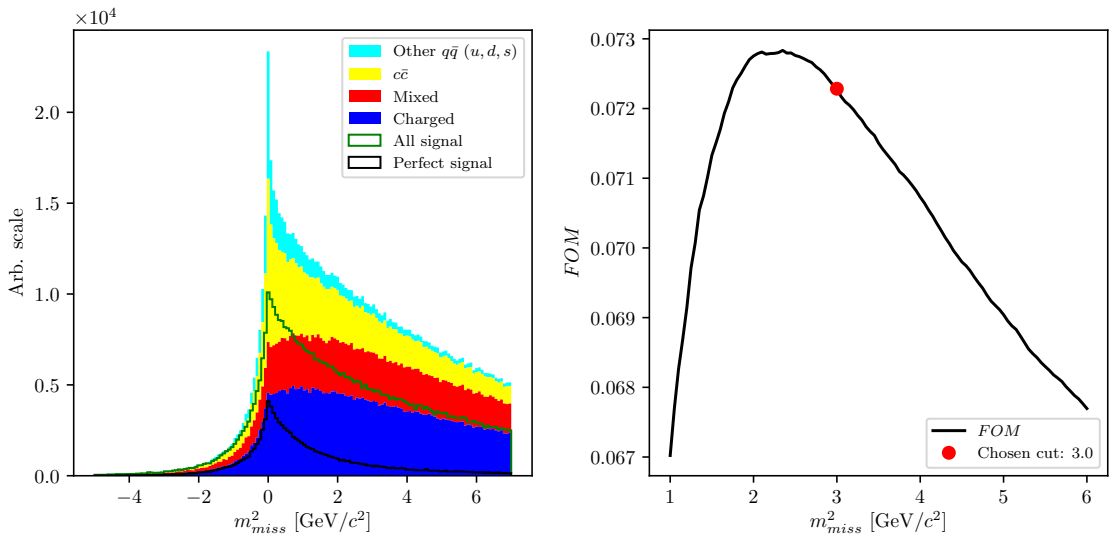


Figure 4.9.: Squared missing mass distribution for signal and various types of background. All signal (green) and perfect signal (black) are scaled up equally.

509 The main uncertainty of the neutrino four-momentum, defined in Eq. (4.10),
 510 comes from energy uncertainty. It is a common practice to substitute the miss-
 511 ing energy with the magnitude of the missing momentum, since the momentum
 512 resolution from the measurement is much better, thus redefining the neutrino four-
 513 momentum to

$$p_\nu = (|\vec{p}_{miss}|, \vec{p}_{miss}), \quad (4.12)$$

514 which fixes the neutrino mass to 0 GeV/c².

The newly defined neutrino four-momentum can be added to the four-momentum of the $Y(KK\ell)$ candidate to obtain the full B meson four-momentum and calculate the traditional M_{BC} and ΔE variables

$$\Delta E = E_B - E_{CMS}/2, \quad (4.13)$$

$$M_{BC} = \sqrt{(E_{CMS}/2)^2 - |\vec{p}_B|^2}. \quad (4.14)$$

515 Since the final fit will be performed over ΔE and M_{BC} , we define the fit region

- 516 • $M_{BC} \in [5.1, 5.295]$ GeV/c²,
- 517 • $\Delta E \in [-1.0, 1.3]$ GeV.

518 Figure 4.10 shows the distributions of ΔE (left) and M_{BC} (right) for signal and
 519 major types of background after the pre-cuts. Both signal components are scaled up
 520 with respect to the background components but are in proper scale one to another.
 521 The effects of missing particles are clearly seen based on the shape difference between
 522 all and perfect signal.

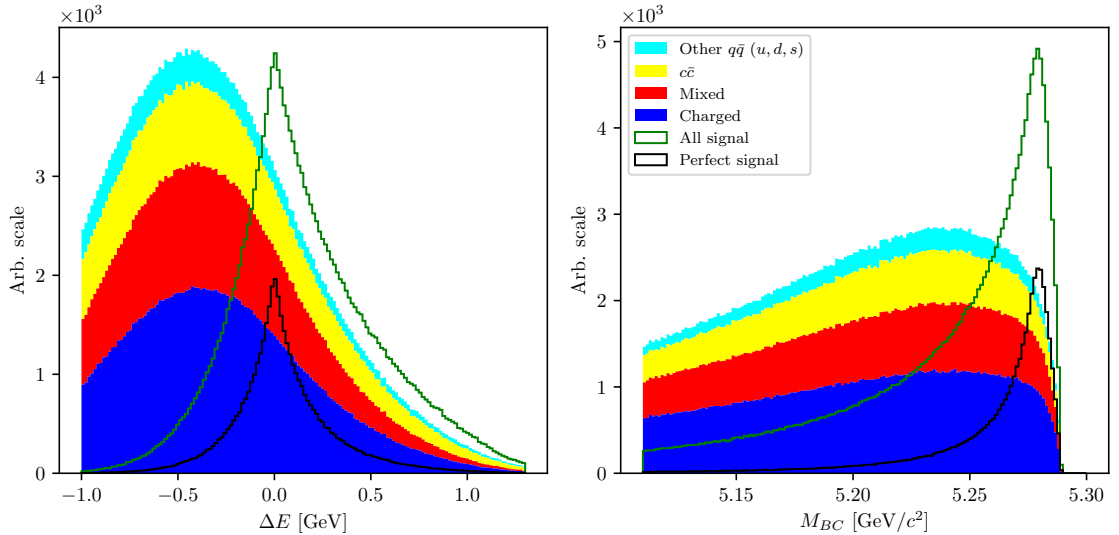


Figure 4.10.: Distributions of ΔE (left) and M_{BC} (right) for signal and major types of background after the precuts. Both signal components are scaled up with respect to the background components, but are in proper scale one to another. The perfect signal has a much better resolution in both distributions, since the event is perfectly reconstructed.

4.3.1. q^2 calculation

q^2 is the squared Lorentz invariant of the four-momentum which is transferred from the B meson to the W boson, also known as "momentum transfer squared". There are several possible calculations of this variable which offer a different resolution. In this analysis, we follow the calculation procedure of q^2 from [16] which yields the best resolution.

For correctly reconstructed events Eq. (4.13) satisfies the condition $\Delta E \approx 0$ within precision. It is possible to rescale the neutrino energy in such way that we fix ΔE to zero, meaning

$$\Delta E' = (E_Y + \alpha E_\nu) - E_{CMS}/2 = 0. \quad (4.15)$$

and calculate and adapted version of M_{BC}

$$M'_{BC} = \sqrt{(E_{CMS}/2)^2 - |\vec{p}_Y + \alpha \vec{p}_\nu|^2}. \quad (4.16)$$

The neutrino momentum resolution dominates the ΔE uncertainty [], so the correction factor α reduces this source of uncertainty.

A second correction can be applied by rotating the direction of the neutrino momentum by a small angle with respect to the reconstructed one. Such an angle is chosen in order fix the value of M'_{BC} to the nominal mass of the B meson, m_B .

The corrected neutrino momentum is then fixed to expected values of ΔE and M_{BC} and solely used for the q^2 calculation. With p_ℓ as the reconstructed lepton four-momentum, we define the q^2 as

$$q^2 = q^2 = (p_\ell + p_\nu)^2. \quad (4.17)$$

The q^2 distributions and the corresponding resolutions are shown in Figure 4.11, along with additional versions of q^2 .

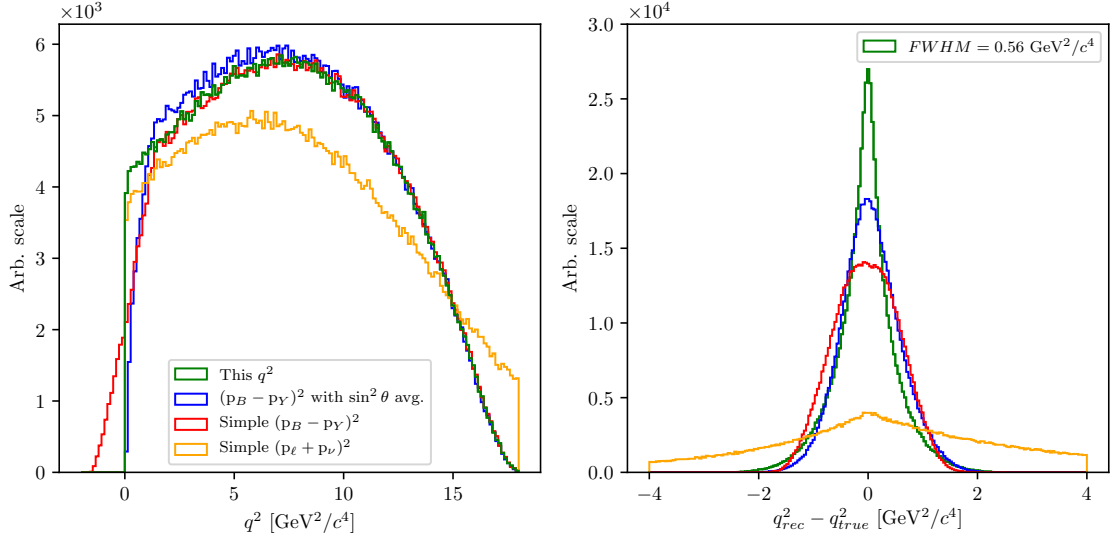


Figure 4.11.: Distributions of q^2 (left) and q^2 resolution (right) for various methods of q^2 calculation. The blue distribution follows the procedure in [16], whereas the red and the orange ones are straight-forward calculations with available information in the reconstruction. The q^2 calculation in red assumes a resting B meson in the CMS frame, and the calculation in orange uses the neutrino four-momentum from Eq. (4.10).

One must bear in mind that even though this calculation yields a precise result, this says nothing about the correctness of the q^2 model which was used (ISGW2 [17]). Since this decay has not been observed yet, we do not have a good description of the decay model, which is also a source of systematics in this analysis.

4.4. Final Stage Optimization

With the charge particle selection and a rough selection of the B meson particles in place, we can now afford to run the reconstruction procedure over all of the 10 streams of the full available Belle generic MC. After obtaining the full reconstructed sample, the first task is to optimize the under-optimized cuts from the pre-selection stage. Repeating the procedure on the full sample results in the FOM shapes shown in Figure 4.12, while the optimal cut values are

- $P(\chi^2, NDF) > 6.0 \times 10^{-3}$,
- $|\cos(\theta_{BY})| < 1.05$.

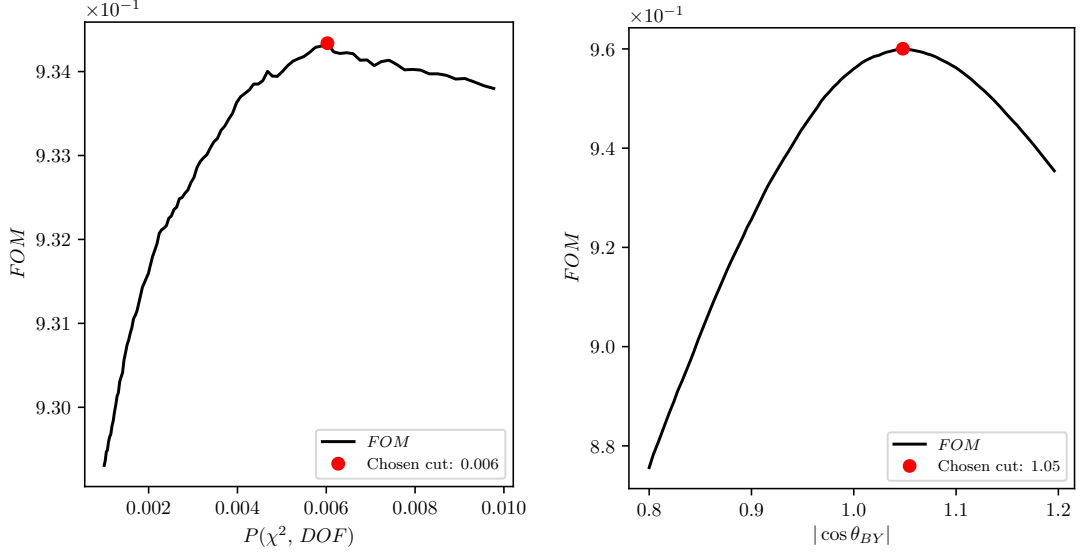


Figure 4.12.: FOM optimization of the vertex fit probability (left) and the $\cos \theta_{BY}$ variable (right) for the full Belle MC sample.

With further optimizations, we will be fine-tuning the signal to background ratio. The most prominent and distinguishable part of our signal is the perfectly reconstructed signal. For this purpose, we define a signal region in ΔE and M_{BC} , where most of our perfectly reconstructed candidates lie. We use this region for all of the following optimization steps in this chapter and also in the background suppression chapter 6 since this allows us to better improve the signal to background ratio where it counts most. The 2D FOM optimization of the optimal ΔE and M_{BC} signal region is shown in Figure 4.13.

The signal region is defined as

- $M_{BC} > 5.271 \text{ GeV}/c^2$,
- $|\Delta E| < 0.126 \text{ GeV}$.

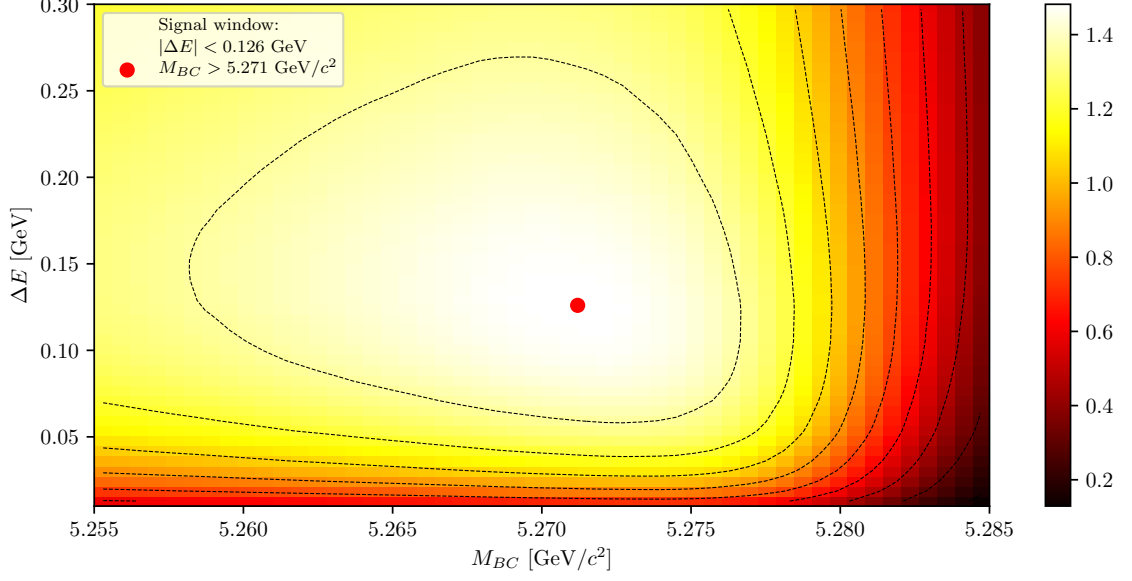


Figure 4.13.: 2D FOM optimization of the signal region definition, where the signal in the optimization was represented by perfectly reconstructed candidates.

567 With the signal window defined, we can tighten the cut on m_{miss}^2 , which we
 568 intentionally left loose before the signal categorization. With the FOM optimization
 569 of perfectly reconstructed candidates inside the signal region, shown in Figure 4.14,
 570 the optimal cut on m_{miss}^2 is

571 • $|m_{miss}^2| < 0.975 \text{ GeV}/c^2$.

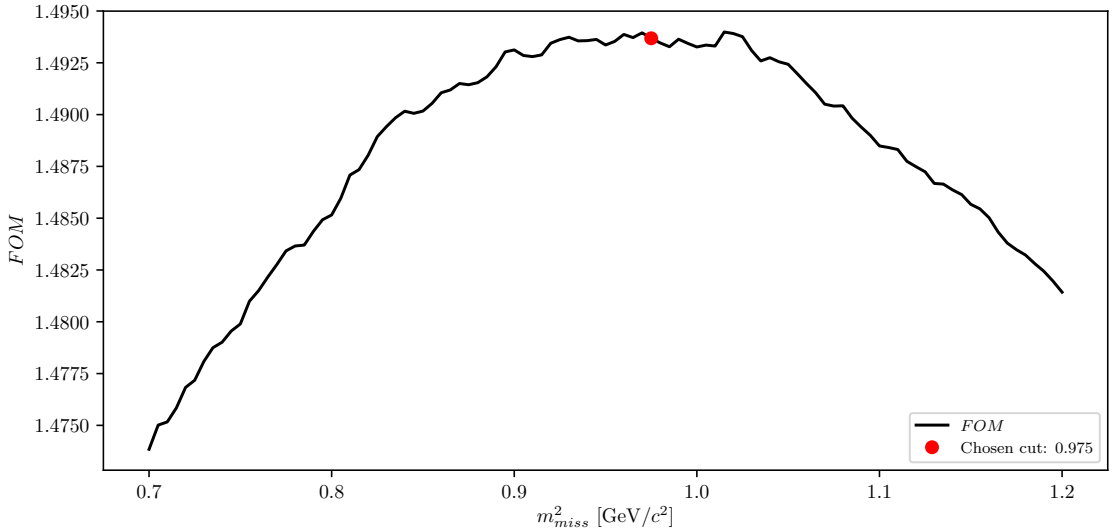


Figure 4.14.: FOM optimization of the optimal m_{miss}^2 cut in the signal region.

4.5. Charge product categorization

The missing information due to an escaping neutrino in our reconstructed channel is replaced by information from the companion B meson. Since this is an untagged reconstruction, the quality of the companion B meson affects the properties of the signal candidate. Perfect reconstruction of a hadronically decayed companion B meson results in pronounced peaks at $\Delta E \approx 0$, $m_{miss}^2 \approx 0$ and $M_{BC} \approx m_B$, while imperfect reconstruction due to any kind of missing particles produces tails, shift or simply a worse resolution of the mentioned distributions. These effects are undesired since they make it harder to separate signal from background.

To remedy this, we look at the charge product of the reconstructed B meson and the ROE object. For correctly reconstructed events, this should have a value of

$$q_{B^\pm} q_{B^\mp} = -1, \quad (4.18)$$

however, this value is distributed due to missing charged particles in the reconstruction. Figure 4.15 shows various signal distributions of ΔE and M_{BC} in arbitrary (left) and normalized (right) scales, with the relative ratios of 67.86 % and 32.14 % for correct and wrong values of the charge product, respectively. We see that correctly reconstructed events represent the majority of signal and also have the best resolution in ΔE and M_{BC} , so we proceed with the analysis by imposing the cut in Eq. 4.18.

While this cut introduces a drop in the signal efficiency of about 32.14 %, it improves the resolution of our signal ΔE and M_{BC} distributions and also the signal to background ratio, where the latter changes from 0.95×10^{-3} to 1.09×10^{-3} .

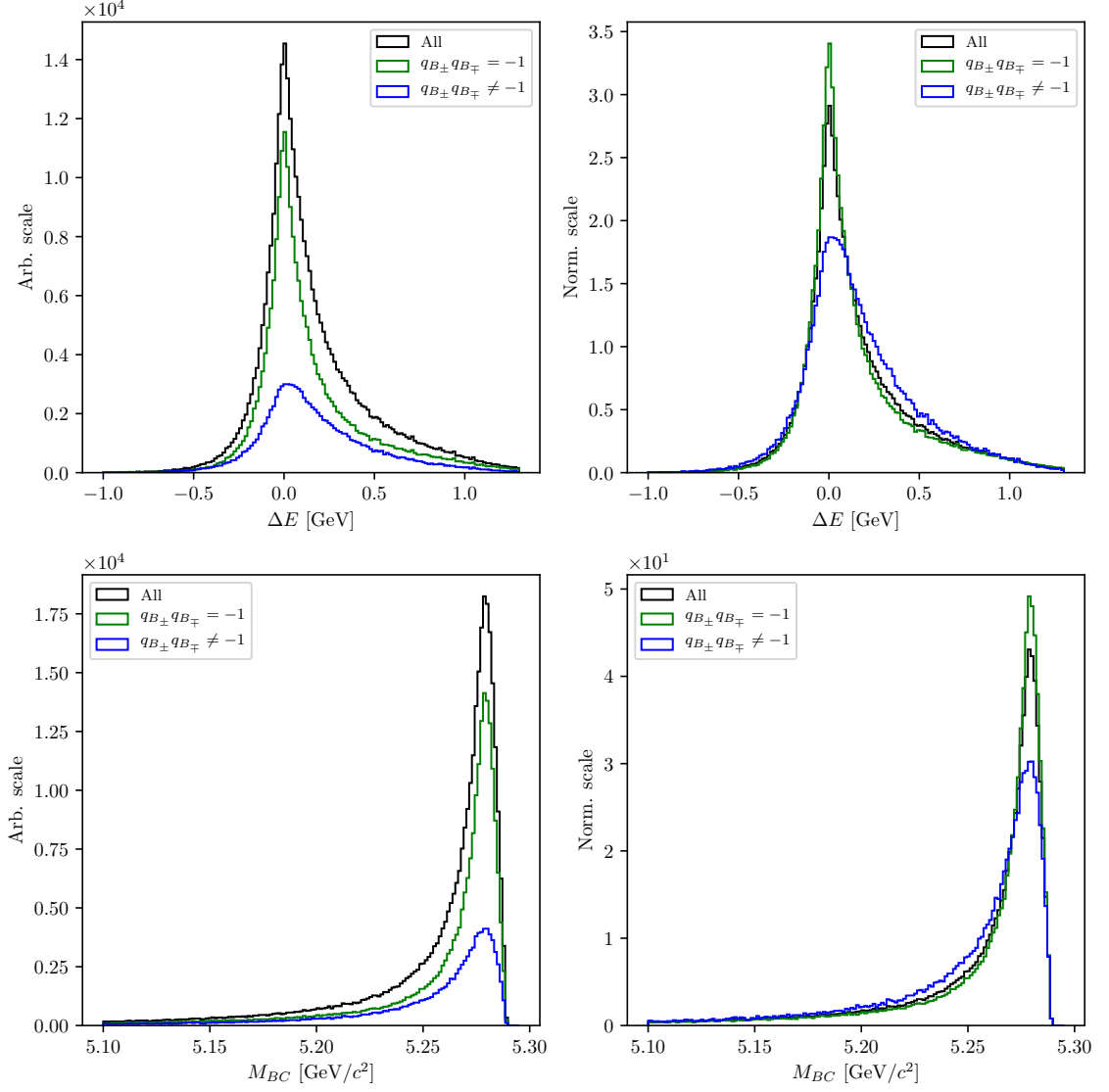


Figure 4.15.: Signal distributions of ΔE and M_{BC} based on the charge product of both B mesons in the event. The plot on the left shows the distributions in an arbitrary scales, while the plot on the right shows the normalized distributions.

4.6. Selection Summary

In this section, one can find the full summary of all final selection cuts in the event reconstruction, from FSP particles up to the B meson.

- FSP particles:

- electrons: $|d_0| < 0.1$ cm, $|z_0| < 1.5$ cm, $p > 0.6$ GeV/ c ,
 $p_{CMS} \in [0.4, 2.6]$ GeV/ c , $eID > 0.9$,

- muons: $|d_0| < 0.1$ cm, $|z_0| < 1.5$ cm, $p_{CMS} \in [0.6, 2.6]$ GeV/ c ,
 $\mu ID > 0.97$,

- 601 – kaons: $|d_0| < 0.15 \text{ cm}, |z_0| < 1.5 \text{ cm}, p_{CMS} < 2.5 \text{ GeV}/c,$
- 602 $K/\pi \text{ ID} > 0.6, K/p \text{ ID} > 0.1,$
- 603 • B meson candidates:
- 604 – standard cuts: $P(\chi^2, DOF) > 6 \times 10^{-3}, |\cos \theta_{BY}| < 1.05, |m_{miss}^2| <$
- 605 $0.975 \text{ GeV}/c^2,$
- 606 – fit region cuts: $\Delta E \in [-1.0, 1.3] \text{ GeV}, M_{BC} \in [5.1, 5.295] \text{ GeV}/c^2,$
- 607 – signal region cuts: $|\Delta E| < 0.126 \text{ GeV}, M_{BC} > 5.271 \text{ GeV}/c^2,$
- 608 – charge categorization: $q_{B^\pm} q_{B^\mp} = -1.$

Chapter 5.

Rest of Event Clean-up

Continuing from section 4.3, the description of the ROE clean-up process is described here.

Training the MVA classifiers follows the same recipe for all the steps in this chapter. For each step, we run B meson reconstruction on Signal MC with a generic companion B meson. This way the produced weight files are less likely to be signal-side dependent and can be used also for untagged analyses of other decays. For every correctly reconstructed signal B meson we save the necessary information for each MVA step (i.e. properties of ROE clusters). Only correctly reconstructed B candidates are chosen here, to prevent leaks of information from the signal side to the ROE side.

More information about the MVA training, hyper-parameter optimization and feature importance for each MVA step in this chapter can be found in the Appendix A.

5.1. Clusters Clean-up

Photons originate from the IP region, travel to the ECL part of the detector in a straight line and produce a cluster. The direction of the photon is determined via the location of the cluster hit in the ECL and the energy of the photon is directly measured via the deposited energy. This way the four-momentum of photons is determined and used in Eq. (4.8).

Most of the photons in events with B mesons come from $\pi^0 \rightarrow \gamma\gamma$ decays. However, a lot of hits in the ECL are also created by photons coming from the beam-induced background or secondary interactions with the detector material. Photons of the first kind should be taken into account when calculating the missing 4-momentum, while the latter kind adds extra energy and momentum which spoils our measured quantities. In the first step of the clusters clean-up, we train an MVA which recognizes good π^0 candidates and apply this information to the daughter photons. This represents a sort of a π^0 origin probability, which peaks at or is equal to 0 for photons not coming from π^0 particles, and peaks at 1 otherwise. This information is used as an additional classifier variable in the next step of the clean-up, where we train to recognize good photons in an event.

5.1.1. π^0 MVA Training

The training dataset of π^0 candidates contains

- 183255 target candidates,

644

- 200000 background candidates,

645

646

647

648

649

650

where the definition of a target is that both photon daughters that were used in the reconstruction of the π^0 are actual photons and real daughters of the π^0 particle. We use π^0 candidates from the converted Belle particle list and select those with invariant mass in the range of $M \in [0.10, 0.16]$ GeV. After that we perform a mass-constrained fit on all candidates, keeping only the ones for which the fit converged.

The input variables used in this MVA are

651

652

653

654

655

656

657

658

659

660

661

662

- p and p_{CMS} of π^0 and γ daughters,
- fit prob. of the mass-constrained fit, invariant mass and significance of mass before and after the fit,
- angle between the photon daughters in the CMS frame,
- cluster quantities for each daughter photon
 - E_9/E_{25} ,
 - theta angle,
 - number of hit cells in the ECL,
 - highest energy in cell,
 - energy error,
 - distance to closest track at ECL radius.

The classifier output variable is shown in Figure 5.1.

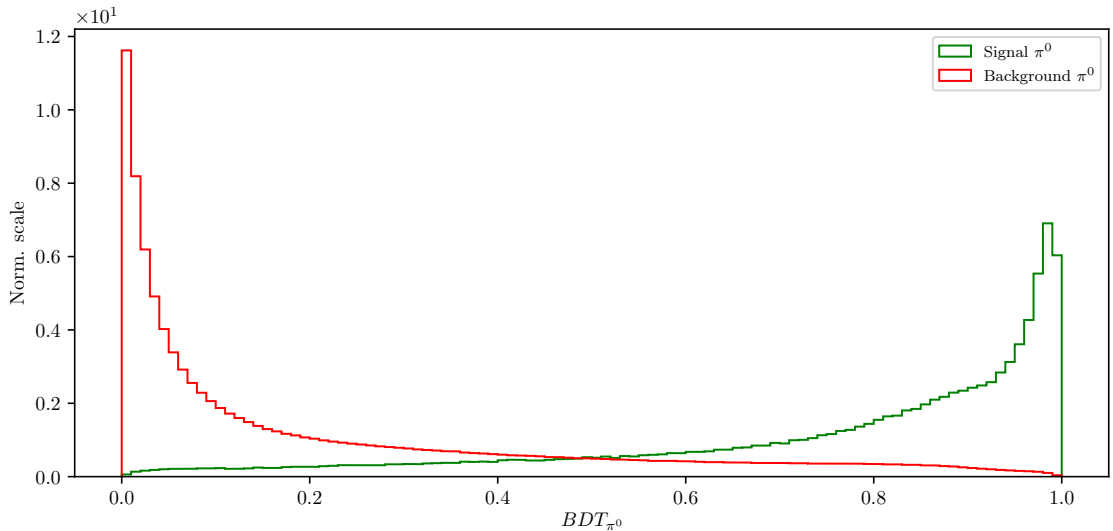


Figure 5.1.: Classifier output of the π^0 training for signal and background π^0 candidates.

663

664

The distributions for all input variables and their correlations for signal and background candidates can be found in Appendix A for all steps of the ROE clean-up.

5.1.2. γ MVA Training

In this MVA training, we take the π^0 classifier output of the previous training as an input in order to train a classifier to distinguish between good and bad photons. The π^0 probability information from the previous step is applied to all photon pairs which pass the same π^0 cuts as defined in the previous step. Since it is possible to have overlapping pairs of photons, the π^0 probability is overwritten in the case of a larger value, since this points to a greater probability of a correct photon combination. On the other hand, some photon candidates fail to pass the π^0 selection, these candidates have a fixed value of π^0 probability equal to zero.

The training dataset of γ candidates contains

- 171699 target candidates,
- 177773 background candidates,

where the definition of a target is that the photon is an actual photon which is related to a primary MC particle. This tags all photon particles from secondary interactions as background photons. We use the converted γ candidates from the existing Belle particle list.

The input variables used in this MVA are

- p and p_{CMS} of γ candidates,
- π^0 probability,
- cluster quantities
 - E_9/E_{25} ,
 - theta angle,
 - number of hit cells in the ECL,
 - highest energy in cell,
 - energy error,
 - distance to closest track at ECL radius.

The classifier output variable is shown in Figure 5.2.

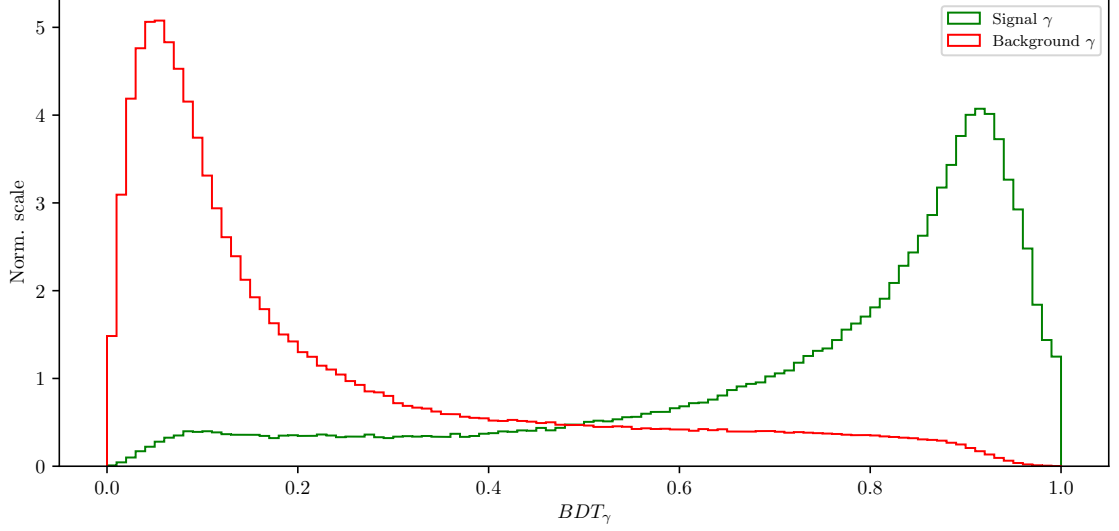


Figure 5.2.: Classifier output of the γ training for signal and background γ candidates.

692 With the final weights for photon classification in hand, we apply them to the
 693 photon particle list. The cut optimization is shown in Figure 5.3 (left), with the
 694 optimal cut on the γ classifier output at

- 695 • $BDT_\gamma > 0.5045$.

696 Figure 5.3 (right) shows the LAB frame momentum of the photons before and
 697 after the cut in logarithmic scale. The signal efficiency and background rejection at
 698 this clean-up cut are

- 699 • Signal efficiency: $\epsilon_{SIG} = 83.2 \%$,
- 700 • Background rejection: $1 - \epsilon_{BKG} = 81.2 \%$.

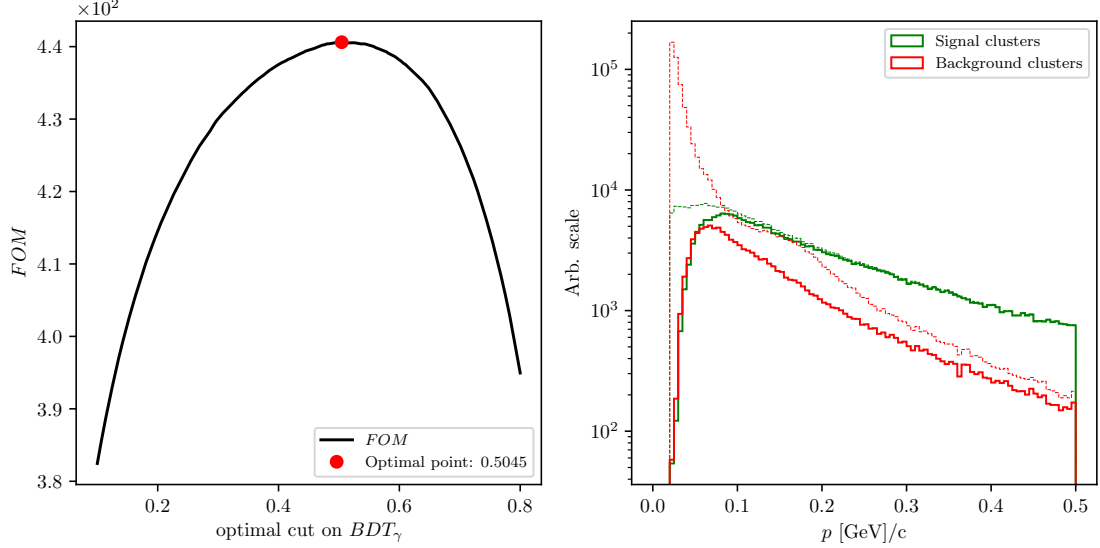


Figure 5.3.: The FOM of the classifier output optimization (left) and momentum magnitude in the LAB frame of signal and background photon candidates before and after the optimal cut (right).

701 The event is now considered to be clean of extra clusters.

702 5.2. Tracks Clean-up

703 Charged particles leave hits in the detector, which are then grouped into tracks
 704 by advanced tracking algorithms. The track is fitted and the track momentum is
 705 determined. With the help of particle identification information (PID), we are able
 706 to make an intelligent decision about the mass hypothesis of the particle and thus
 707 reconstruct the charged particle's four-momentum, which is then added in the loop
 708 in Eq. (4.8).

709 Most of the quality (good) tracks, which come from physics event of interest,
 710 come from the IP region, where the collisions occur. Cleaning up the tracks is a
 711 more complex procedure than cleaning up the clusters. The following facts need to
 712 be taken into account

- 713 (a) good tracks can also originate away from the IP region, due to decays of long-
 714 lived particles, such as $K_S^0 \rightarrow \pi^+ \pi^-$,
- 715 (b) charged particles from background sources produce extra tracks or duplicates,
- 716 (c) low momentum charged particles can curl in the magnetic field and produce
 717 multiple tracks,
- 718 (d) secondary interactions with detector material or decays of particles in flight can
 719 produce "kinks" in the flight directory, resulting in multiple track fit results per
 720 track.

721 Schematics of all the cases mentioned above are shown in Figure 5.4.

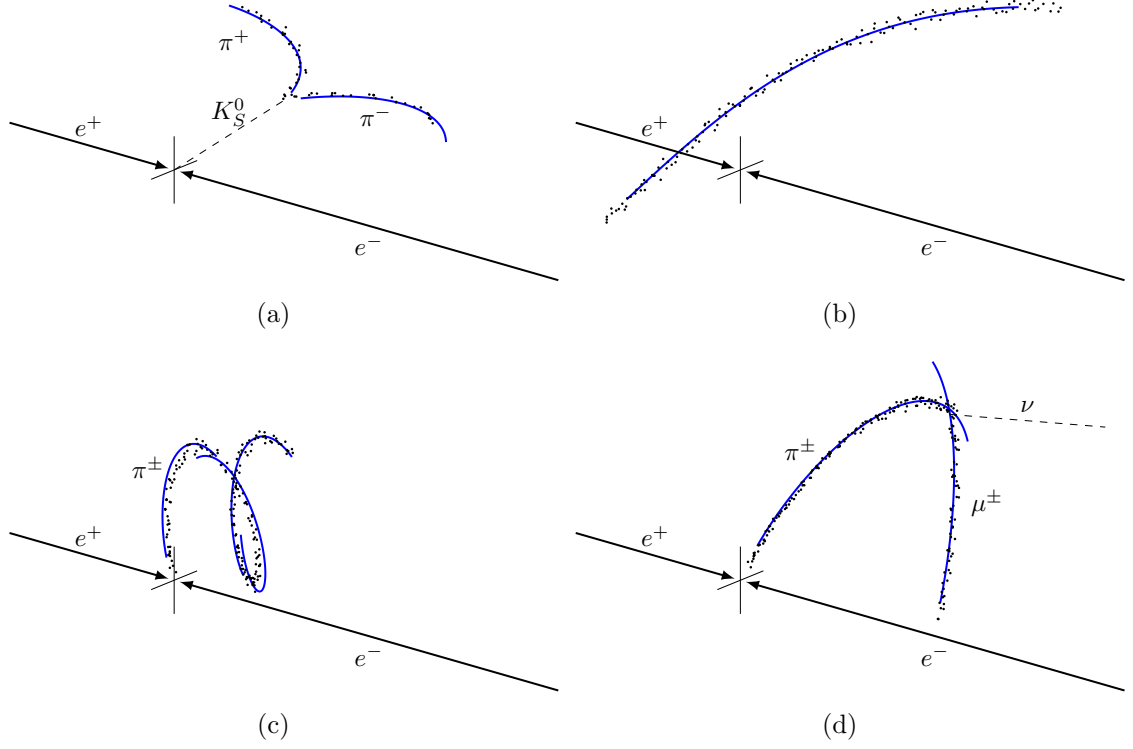


Figure 5.4.: (a) Tracks from long-lived neutral particles, which decay away from the IP region, (b) Random tracks from background which are reconstructed, (c) Low-momentum particles which curl in the magnetic field, (d) in-flight decays of particles, which produce a kink in the trajectory.

It is obvious that tracks from the same momentum source should only be taken into account once, or, in the case of background tracks, not at all. Such tracks will from this point on be denoted as *extra* tracks, because they add extra four-momentum to our final calculations in Eq. (4.8). At the same time, we have to take care that we do not identify *good* tracks as *extra* tracks. Both of these cases have negative impacts on the final resolution of all variables which depend on information from ROE.

5.2.1. Tracks from Long-lived Particles

The first step in tracks clean-up is taking care of tracks from long-lived particles, such as $K_S^0 \rightarrow \pi^+\pi^-$, $\gamma \rightarrow e^+e^-$ and decays of Λ baryons. Here we only focus on K_S^0 , since they are the most abundant. This step is necessary because of the π^\pm particles, coming from the K_S^0 decays, have large impact parameters, which is usually a trait of background particles. In order to minimize confusion from the MVA point-of-view, these tracks are taken into account separately.

We use the converted K_S^0 candidates from the existing Belle particle list and use a pre-trained Neural Network classifier in order to select only the good K_S^0 candidates. Figure 5.5 shows the distribution of the K_S^0 invariant mass for signal and background candidates, before and after the classifier cut. The momentum of

740 selected K_S^0 candidates is added to the ROE, while the daughter tracks are discarded
 741 from our set.

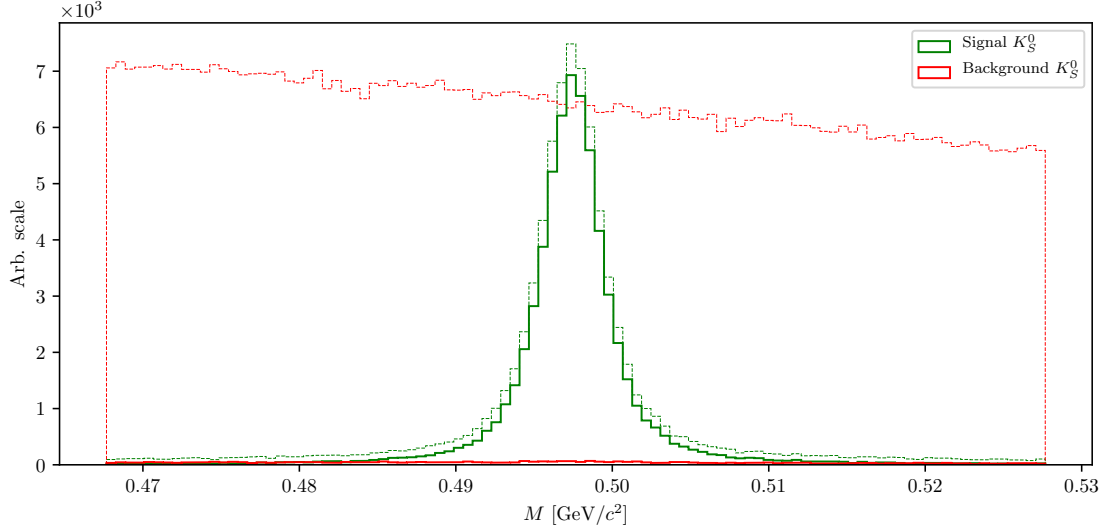


Figure 5.5.: Invariant mass of the K_S^0 candidates before (dashed lines) and after (solid lines) the cut on the Neural Network classifier for signal (green) and background candidates (red). Signal peaks at nominal K_S^0 mass, while background covers a wider region.

742 The signal efficiency and background rejection for K_S^0 candidates after this cut
 743 and on the full range are

- 744 • Signal efficiency: $\epsilon_{SIG} = 80.7 \%$,
- 745 • Background rejection: $1 - \epsilon_{BKG} = 99.4 \%$.

746 5.2.2. Duplicate Tracks

747 All good tracks at this point should be coming from the IP region, since we took
 748 care of all the good tracks from long-lived particle decays, therefore we apply a cut
 749 on impact parameters for all the remaining tracks

- 750 • $|d_0| < 10 \text{ cm}$ and $|z_0| < 20 \text{ cm}$

751 and proceed with the clean-up of track duplicates.

752 Defining a duplicate track pair

753 In this step, we wish to find a handle on secondary tracks from low momentum
 754 curlers and decays in flight. The main property for these cases is that the 3D
 755 opening angle between such two tracks is very close to 0° or 180° , since tracks
 756 deviate only slightly from the initial direction, but can also be reconstructed in the
 757 opposite way. Figure 5.6 shows the distribution of the angle between two tracks in
 758 a single pair for random track pairs and duplicate track pairs, where the latter were
 759 reconstructed as two same-sign or opposite-sign tracks.

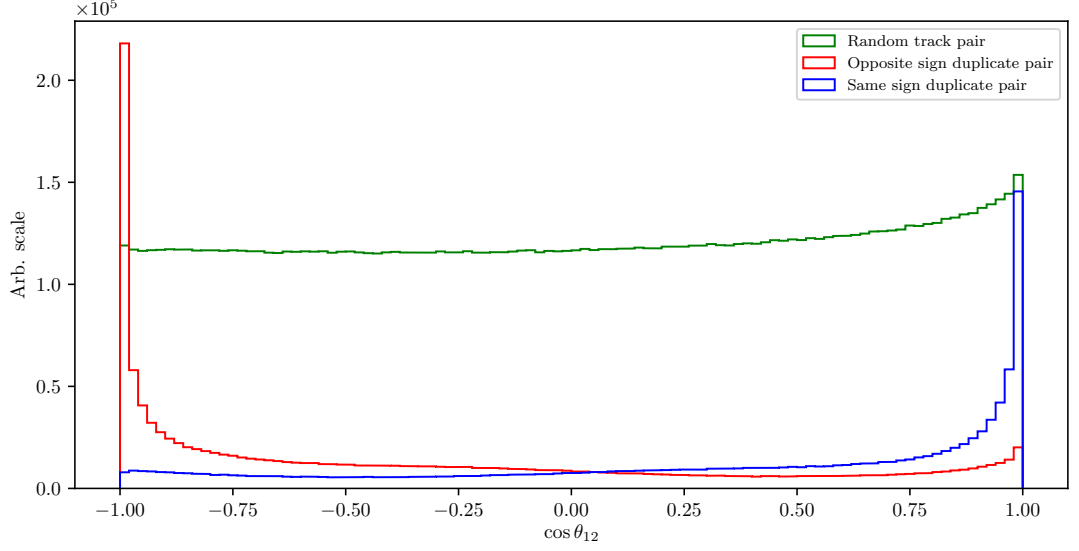


Figure 5.6.: Distribution of the angle between two tracks in a single pair for random track pairs (green) and duplicate track pairs, where the latter were reconstructed as two same-sign (blue) or opposite-sign tracks (red).

760 If the particle decayed mid-flight or produced multiple tracks due to being a low-
761 momentum curler, then, as the name suggests, these particles most likely had low
762 momentum in the transverse direction, p_T . Since both tracks originate from the
763 same initial particle, the momentum difference should also peak at small values.
764 Figure 5.7 shows the momentum and momentum difference of tracks which belong
765 to a random or a duplicate track pair.

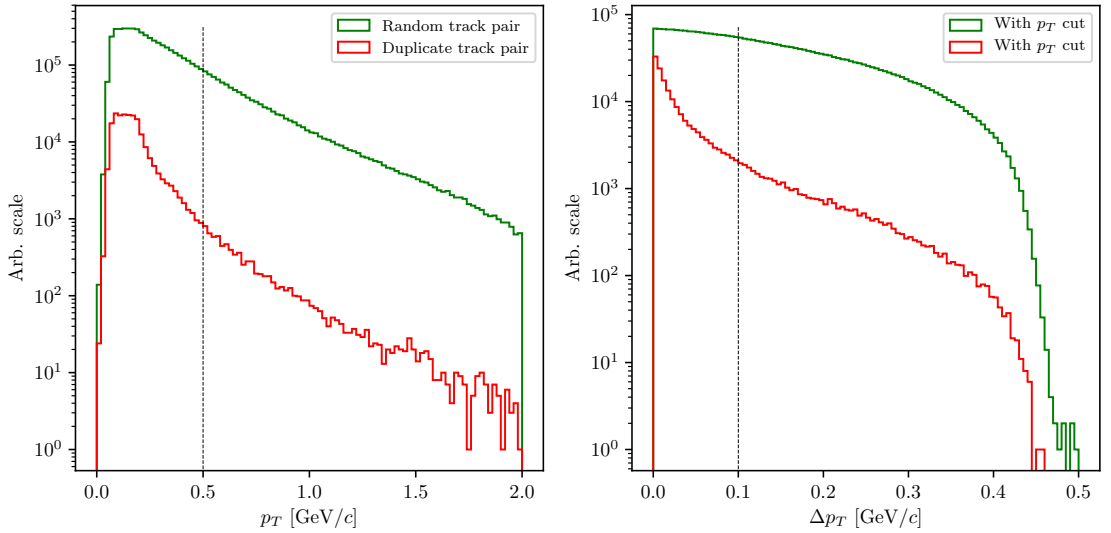


Figure 5.7.: Distribution of transverse momentum p_T (left) and transverse momentum difference Δp_T (right) for all tracks coming from random (green) or duplicate track pairs (red). The plot on the right already includes the cut on p_T from the plot on the left.

766 We impose a cut of

- 767 • $p_T < 0.5 \text{ GeV}/c$,
- 768 • $|\Delta p_T| < 0.1 \text{ GeV}/c$,

769 in order to cut down the number of random track pairs, while retaining a high
 770 percentage of duplicate track pairs. After all the cuts defined in this chapter, the
 771 final distribution of the angle between two tracks is shown in Figure 5.8.

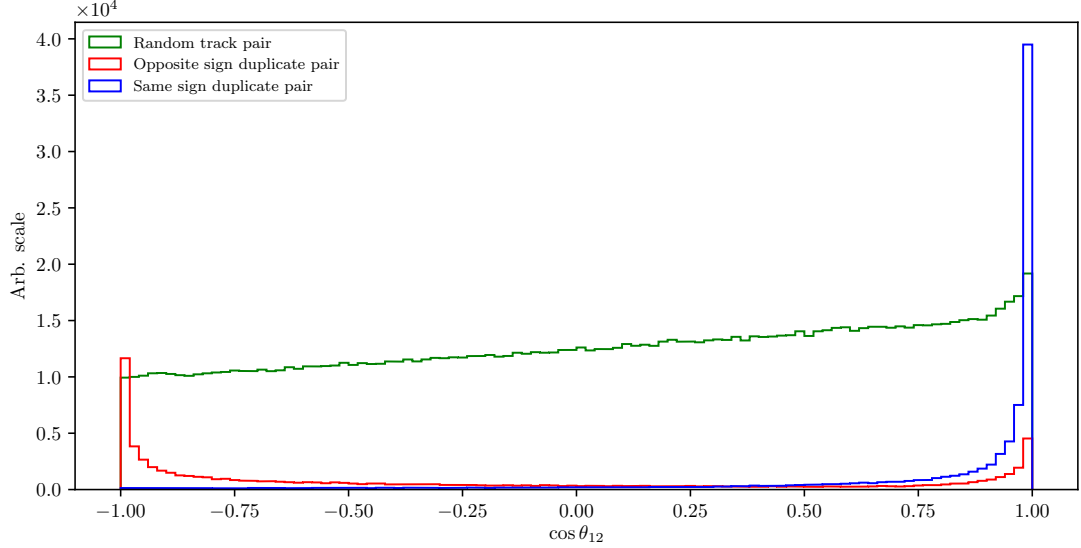


Figure 5.8.: Distribution of the angle between two tracks in a single pair after applying the selection cuts defined in this subsection. The distributions are shown for random track pairs (green) and duplicate track pairs, where the latter were reconstructed as two same-sign (blue) or opposite-sign tracks (red).

772 Training the duplicate track pair MVA

773 This final sample of track pairs is now fed into an MVA, which is trained to recognize
 774 duplicate track pairs over random ones. The training dataset contains

- 775 • 113707 target candidates,
- 776 • 190314 background candidates,

777 where the definition of a target is that the track pair is a duplicate track pair.

778 The input variables used in this MVA are

- 779 • angle between tracks,
- 780 • track quantities
 - 781 – impact parameters d_0 and z_0 ,
 - 782 – transverse momentum p_T ,
 - 783 – helix parameters and helix parameter errors of the track,

- 784 – track fit p -value,
- 785 – number of hits in the SVD and CDC detectors

786 The classifier is able to distinguish between random and duplicate track pairs in
 787 a very efficient manner, as shown in Figure 5.9.

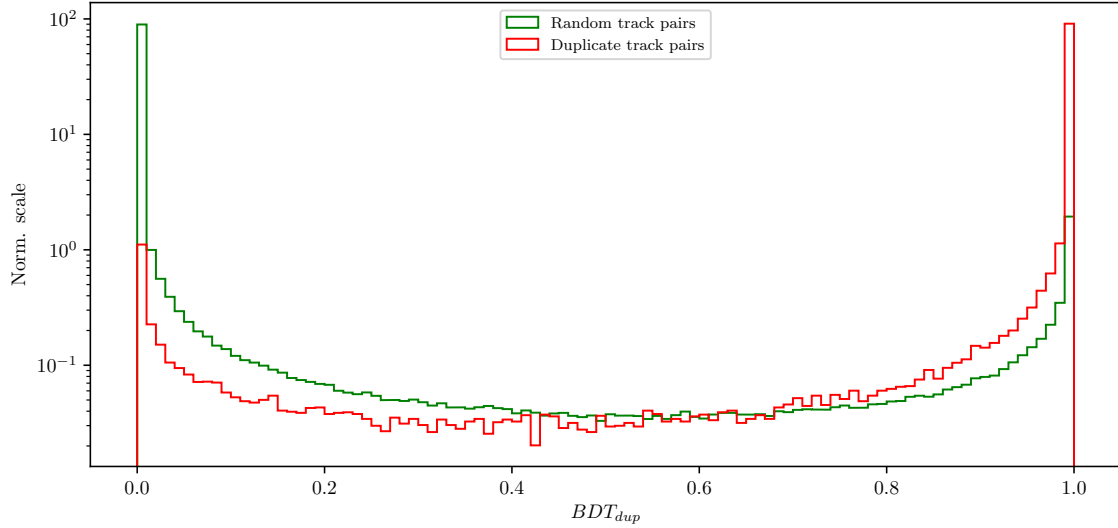


Figure 5.9.: Classifier output of the track pair training for random track pairs and duplicate track pairs.

788 The FOM function for optimal cut selection is shown in Figure 5.10 (left), along
 789 with the angle between the two tracks before and after the optimal cut (right). The
 790 optimal cut for duplicate track selection is

- 791 • $BDT_{duplicate} > 0.9985$.

792 The signal efficiency and background rejection for duplicate pair candidates after
 793 this cut is

- 794 • Signal efficiency: $\epsilon_{SIG} = 87.2 \%$,
- 795 • Background rejection: $1 - \epsilon_{BKG} = 98.8 \%$,

796 where signal and background represent duplicate and random track pairs, respec-
 797 tively.

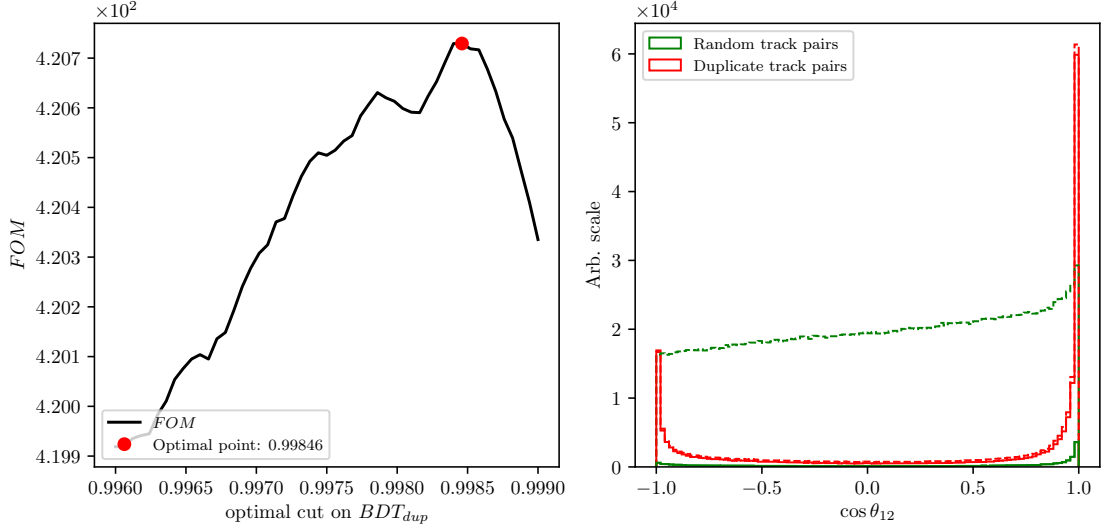


Figure 5.10.: The optimization of the FOM function for the cut on classifier output (left) and distribution of the angle between two tracks in a single pair before (dashed) and after (solid) applying the optimal cut on the output classifier for random and duplicate track pairs (right).

Defining duplicate tracks

What remains now is to decide which track from the duplicate track pair to keep and which to discard. For this purpose, we apply duplicate pair-level information to each track in the pair in the form of

$$\Delta f = f_{this} - f_{other}, \quad (5.1)$$

where f is an arbitrary variable from the list of track quantities in section 5.2.2. From the point-of-view of *this* track, a track is more *duplicate*-like if the following is true

- $\Delta d_0, \Delta z_0 > 0$ (*this* track further away from the IP region),
- $\Delta p_T, \Delta p_Z < 0$ (*this* track has lower momentum),
- $\Delta N_{SVD}, \Delta N_{CDC} < 0$ (*this* track has less hits in the SVD and CDC),

Additionally we define an MC truth variable

$$\Delta \chi^2 = \chi_{this}^2 - \chi_{other}^2, \quad \chi^2 = \sum_{i=x,y,z} \frac{(p_i - p_i^{MC})^2}{\sigma(p_i)^2}, \quad (5.2)$$

where we compare all components of track momentum to the true values. If $\Delta \chi^2 > 0$, then *this* track has a higher probability of being a duplicate track and should be discarded.

However, it turns out that solving this problem is not as simple as discarding one track and keeping the other one. An additional complication here is that we can have more than one extra track from the same initial particle, which leads to track

pairs where both tracks are track duplicates. For example, if we have the following case

$$\begin{aligned}
t_1 &: \text{good track}, \\
t_2 &: \text{extra track}, \\
t_3 &: \text{extra track}, \\
\text{pair}_1 &: (t_1, t_2), \\
\text{pair}_2 &: (t_1, t_3), \\
\text{pair}_3 &: (t_2, t_3),
\end{aligned}$$

where t_1 is the original track and t_2 and t_3 are extra tracks, with t_3 being even more duplicate-like with respect to t_2 . Here tracks t_2 and t_3 should be discarded while t_1 should be kept. We can achieve this if we overwrite existing pair-level information in the tracks for cases where the variable difference Δf is more duplicate-like. If we follow the same example, we could fill information about the property f in six different orders.

$$\begin{aligned}
1. (t_1, t_2^*) &\rightarrow (t_1, t_3^*) \rightarrow (t_2^*, t_3^*), \\
2. (t_1, t_2^*) &\rightarrow (t_2^*, t_3^*) \rightarrow (t_1, t_3^*), \\
3. (t_1, t_3^*) &\rightarrow (t_2, t_3^*) \rightarrow (t_1, t_2^*), \\
4. (t_1, t_3^*) &\rightarrow (t_1, t_2^*) \rightarrow (t_2^*, t_3^*), \\
5. (t_2, t_3^*) &\rightarrow (t_1, t_3^*) \rightarrow (t_1, t_2^*), \\
6. (t_2, t_3^*) &\rightarrow (t_1, t_2^*) \rightarrow (t_1, t_3^*),
\end{aligned}$$

812 where the "*" symbol denotes when a track is recognized as a duplicate track
813 with respect to the other track. We see that no matter the order, both t_2 and t_3 get
814 recognized as duplicate tracks correctly.

815 Training the duplicate track MVA

816 The training procedure is similar as before. The sample of tracks from duplicate
817 track pairs is now fed into an MVA, which is trained to distinguish duplicate tracks
818 from good tracks. The training dataset contains

- 819 • 84339 target candidates,
- 820 • 68280 background candidates,

821 where the definition of a target is that the track is a duplicate track.

822 The input variables used in this MVA are

- 823 • theta angle of the track momentum,
- 824 • track quantities
 - 825 – impact parameters d_0 and z_0 and their errors,
 - 826 – CMS frame momentum p_{CMS} and momentum components p_T and p_z
 - 827 – number of hits in the SVD and CDC detectors
 - 828 – track fit p -value,

- 829 • pair-level information
- 830 – $\Delta d_0, \Delta z_0, \Delta N_{CDC}, \Delta N_{SVD}, \Delta p_T, \Delta p_z, \Delta p\text{-value}.$

831 The weights from this training are applied to the tracks, where now each track
 832 has a certain probability of being a duplicate track. We then compare these values
 833 between both tracks in each track pair as

$$\Delta BDT_{final} = BDT_{final}^{this} - BDT_{final}^{other}, \quad (5.3)$$

834 which is again applied to all track pairs and overwritten for tracks which are more
 835 duplicate-like. The classifier output and the classifier output difference for each
 836 track are shown in Figure 5.11.

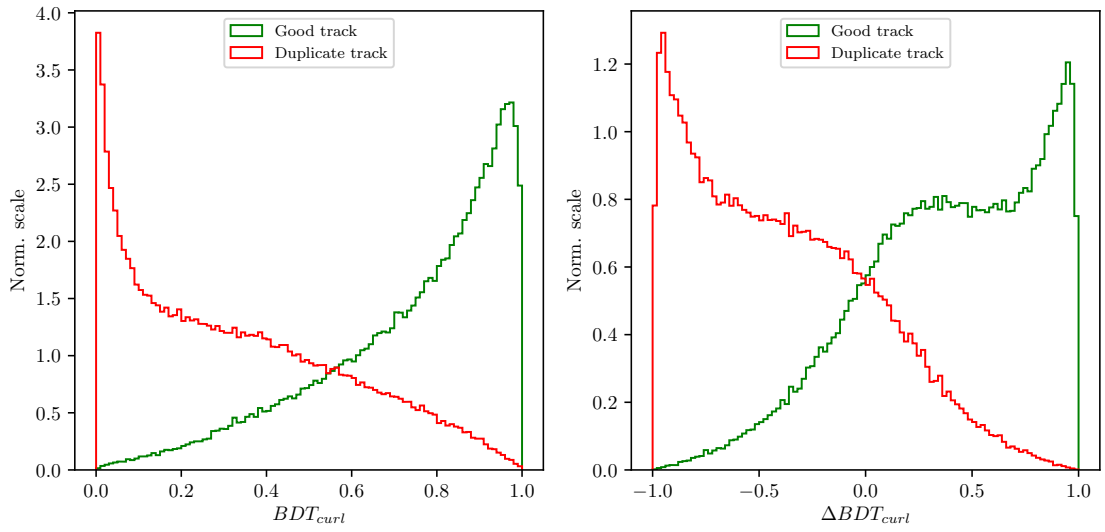


Figure 5.11.: Classifier output of the MVA training for curling track recognition (left) and difference of the classifier output, calculated for each track in a track pair (right).

837 Finally, we select all duplicate tracks which survive the cut

$$\Delta BDT_{final} > 0 \quad (5.4)$$

838 and discard them from our ROE. We can check the performance of our duplicate
 839 track classifier by applying the procedure to a validation sample of duplicate track
 840 pairs and compare the predicted result with the truth, based on Eq. (5.2). Table 5.1
 841 shows the performance of the duplicate track recognition in the form of percentages
 842 of correctly and incorrectly identified duplicate and original tracks. The model seems
 843 to perform well and the event is now considered to be clean of duplicate tracks.

	Predicted duplicate track	Predicted good track
Duplicate track	83.07 %	22.62 %
Good track	16.93 %	77.38%

Table 5.1.: Ratios of correctly classified and misclassified tracks.

5.3. Belle Clean-up

For comparison, we define the Belle clean-up, used standardly at Belle, which is much simpler and relies only on a set of rectangular cuts for neutral particles as well as charged ones. This clean-up procedure is not applied in addition to our ROE clean-up, but separately, only for comparison.

In the case of photons, a single cut on photon energy is applied, depending on the region where the photon hit the relevant part of the detector. The photon cuts are summarized in Table 5.2.

	$17^\circ < \theta < 32^\circ$	$32^\circ < \theta < 130^\circ$	$130^\circ < \theta < 150^\circ$
E_γ	$> 100 \text{ MeV}$	$> 50 \text{ MeV}$	$> 150 \text{ MeV}$

Table 5.2.: Photon selection for the Belle clean-up procedure. Different cuts are applied on photons in different parts of the detector

In case of tracks, pairs are selected which satisfy the following criteria:

- $p_T < 275 \text{ MeV}/c$,
- $\Delta p = |\vec{p}_1 - \vec{p}_2| < 100 \text{ MeV}/c$,
- $\cos \theta(\vec{p}_1, \vec{p}_2) < 15^\circ$ for same sign,
- $\cos \theta(\vec{p}_1, \vec{p}_2) > 165^\circ$ for opposite sign.

Of the two tracks, the one with a larger value of formula in Eq. 5.5 is discarded. The remaining tracks in the event then need to satisfy the conditions described in Table 5.3.

$$(\gamma|d_0|)^2 + |z_0|^2, \quad \gamma = 5. \quad (5.5)$$

	$p_T < 250 \text{ MeV}/c$	$250 \text{ MeV}/c < p_T < 500 \text{ MeV}/c$	$p_T > 500 \text{ MeV}/c$
$ d_0 $	$< 20 \text{ cm}$	$< 15 \text{ cm}$	$< 10 \text{ cm}$
$ z_0 $	$< 100 \text{ cm}$	$< 50 \text{ cm}$	$< 20 \text{ cm}$

Table 5.3.: Photon selection for the Belle clean-up procedure. Different cuts are applied on photons in different parts of the detector

5.4. Clean-up Results

In this section, the results of the ROE clean-up are shown. It is obvious that cleaning up the event affects the shape of various distributions, especially ΔE and M_{BC} , which we are most interested in. Since the reconstruction procedure includes cuts on the cleaned-up variables, the clean-up also affects the efficiency of the reconstructed sample, not only the resolution.

We compare the clean-up setup, defined in this analysis, to the standard clean-up used by Belle, and to a default case, where no clean-up was applied at all. We apply

868 the clean-up procedure to our signal MC sample with all the applied cuts, defined in
 869 section 4.6, except for the signal categorization cuts. Figure 5.12 (left) shows signal
 870 candidate distributions of ΔE and M_{BC} for various clean-up setups. Focusing on
 871 the ROE clean-up, we see an improvement in resolution in both observed variables
 872 and an overall decrease in efficiency. The efficiency decrease is expected since the
 873 cleaned-up variables are able to better isolate the perfectly reconstructed candidates
 874 and discard the non-perfect candidates. In fact, the efficiency of the perfectly re-
 875 constructed candidates increases after the ROE clean-up, as shown in Figure 5.12
 876 (right). The signal MC sample after in case of the Belle clean-up also shows a slight
 877 improvement in the resolution, but looking at the perfectly reconstructed candi-
 878 dates we see that this clean-up procedure is not optimal. Table 5.4 shows ratios
 879 of efficiencies and $FWHM$'s of the clean-up procedures for the perfect signal with
 880 respect to the default case, based on the ΔE distribution. While both the Belle and
 881 ROE clean-up improve the resolution, ROE clean-up performs significantly better
 882 and also increases the amount of the perfectly reconstructed candidates in the final
 883 sample.

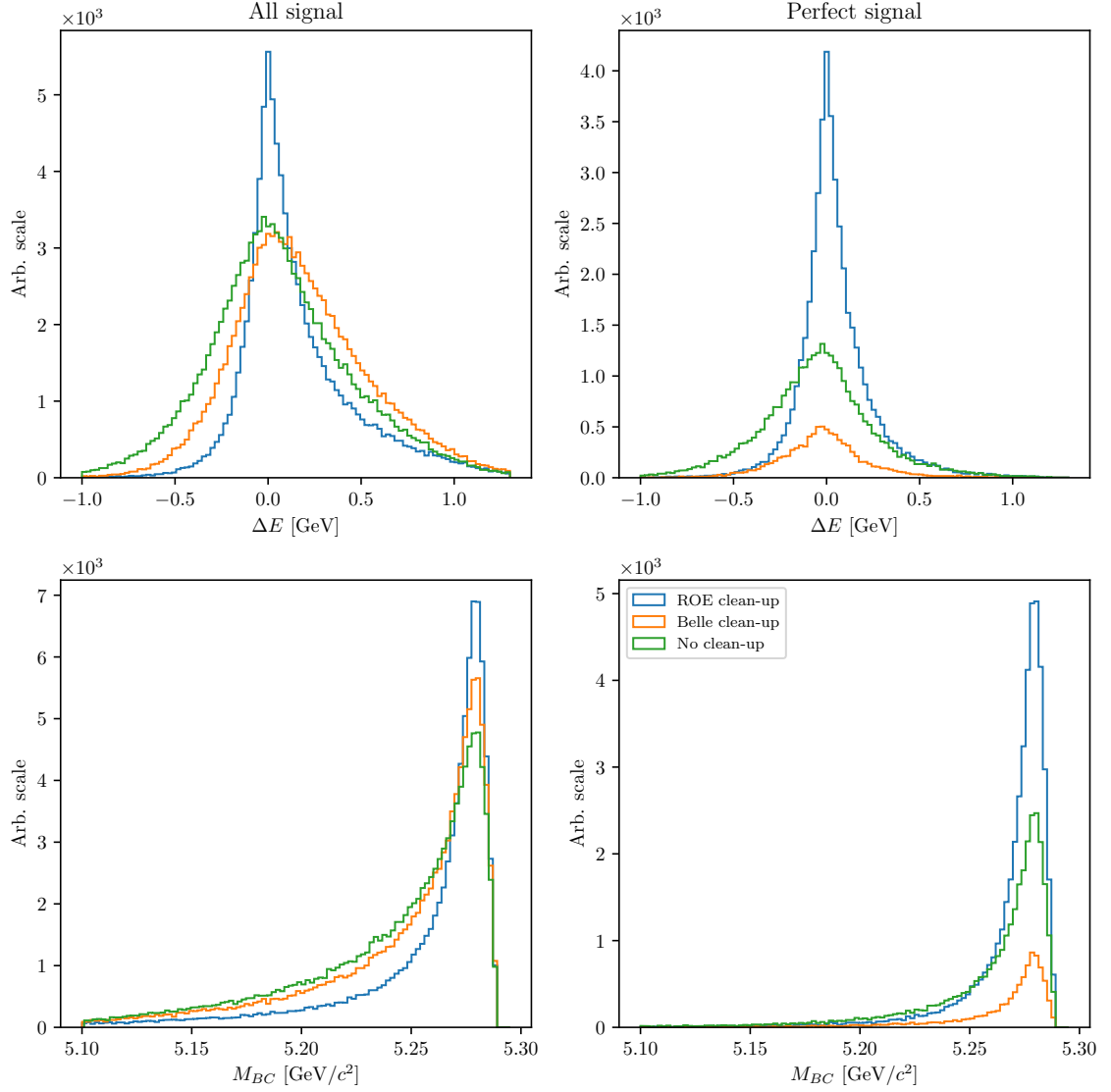


Figure 5.12.: ΔE and M_{BC} distributions for various types of clean-up procedures. The figures on the left are shown for the full signal sample after the stated cuts, while the figures on the right are shown for the perfectly reconstructed signal candidates. For ROE clean-up, the procedure seems to improve resolution as well as increase the amount of perfectly reconstructed candidates, relative to the default case.

	Efficiency ratio	FWHM ratio
Belle clean-up	28.5 %	75.0 %
ROE clean-up	140.1 %	35.0 %

Table 5.4.: Comparison of efficiencies and $FWHM$'s of ROE and Belle clean-up setups with respect to the default case (no clean-up).

885 the signal and companion B meson candidate, already defined in Eq. (4.18), shown
 886 in Figure 5.13 for various clean-up procedures. The figure shows an improved reso-
 887 lution of the distribution, which means that candidates migrate to the correct value
 888 of the charge product after the clean-up. Looking at the perfectly reconstructed
 889 candidates we again see the increase in the bin corresponding to the correct charge
 890 product. As a cross-check, we can also look at ΔE and M_{BC} variables for each
 891 value of the charge product. These plots are shown for the full signal MC sample in
 892 Figure 5.14 and they show a clear resolution improvement for the correct value of
 893 the charge product in the case of the ROE clean-up. For other values of the charge
 894 product, there also seems to be a small improvement for both cases of clean-up, but
 895 it is negligible compared to the plots in the second column. This supports our choice
 896 of signal categorization, defined in section 4.5, where we select only candidates with
 897 the correct value of the charge product.

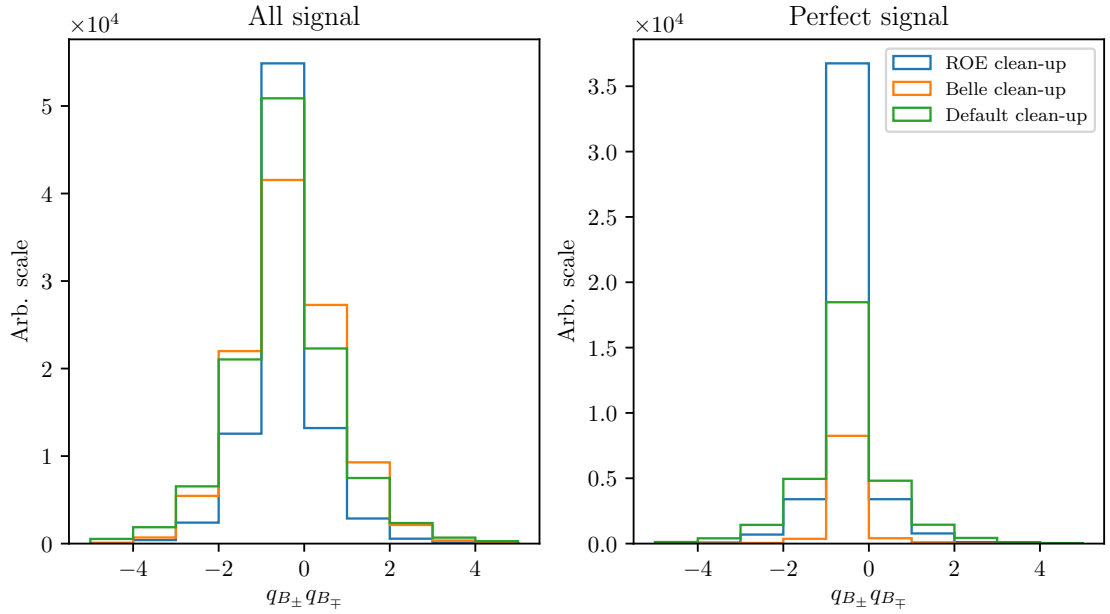


Figure 5.13.: Distribution of the charge product of both B mesons for various types of clean-up procedures, shown on the full signal MC (left) and for the perfectly reconstructed signal candidates (right). For ROE clean-up, the procedure seems to increase the number of perfectly reconstructed candidates.

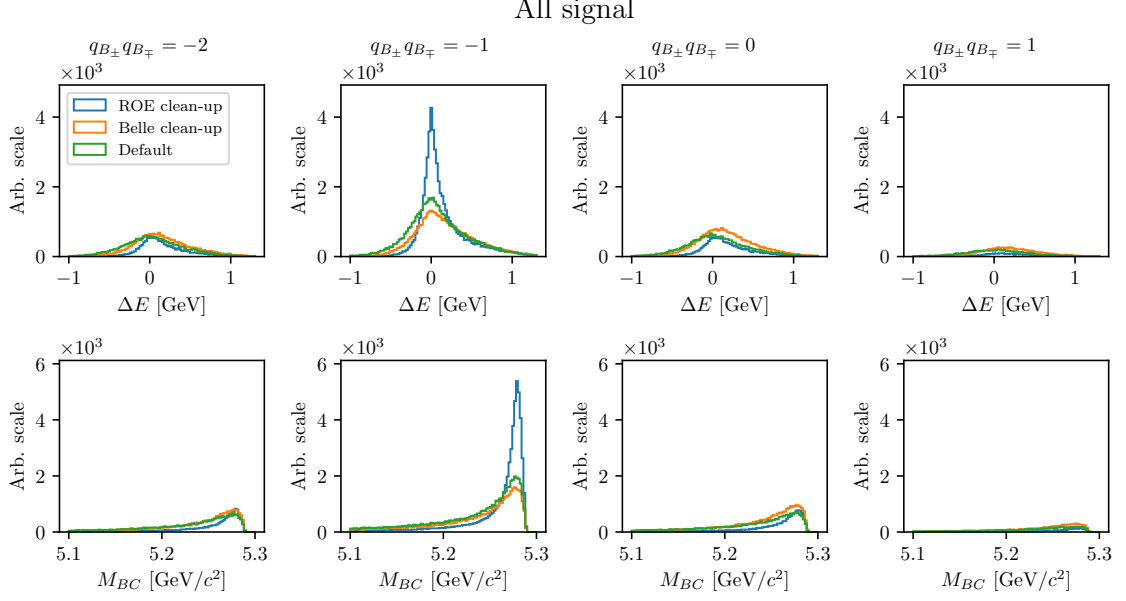


Figure 5.14.: Distributions of ΔE (top) and M_{BC} (bottom) for various types of clean-up procedures, split by specific values of the charge product, shown for the full signal MC. There is a significant improvement in resolution after ROE-clean-up for the case of the correct value of the charge product.

5.5. ROE Clean-up Validation

The ROE clean-up seems to perform well on signal MC based on the results in the previous section. However, it is necessary to make sure that this procedure performs as well on other simulated and measured data, which is done in this section. The clean-up procedure is validated on the control sample,

$$B^+ \rightarrow \bar{D}^0 \ell^+ \nu, \quad D^0 \rightarrow K^+ K^-,$$

which was already defined in section 2.2. The control candidates are reconstructed in the same manner as the signal candidates. In addition to the same cuts applied as in the previous section, we also apply a selection to make the control sample more significant. We discard all candidates which fail to pass the cut on invariant mass of the two kaons

$$1.849 \text{ GeV}/c^2 < m_{KK} < 1.879 \text{ GeV}/c^2 \quad (5.6)$$

as shown in Figure 5.15.

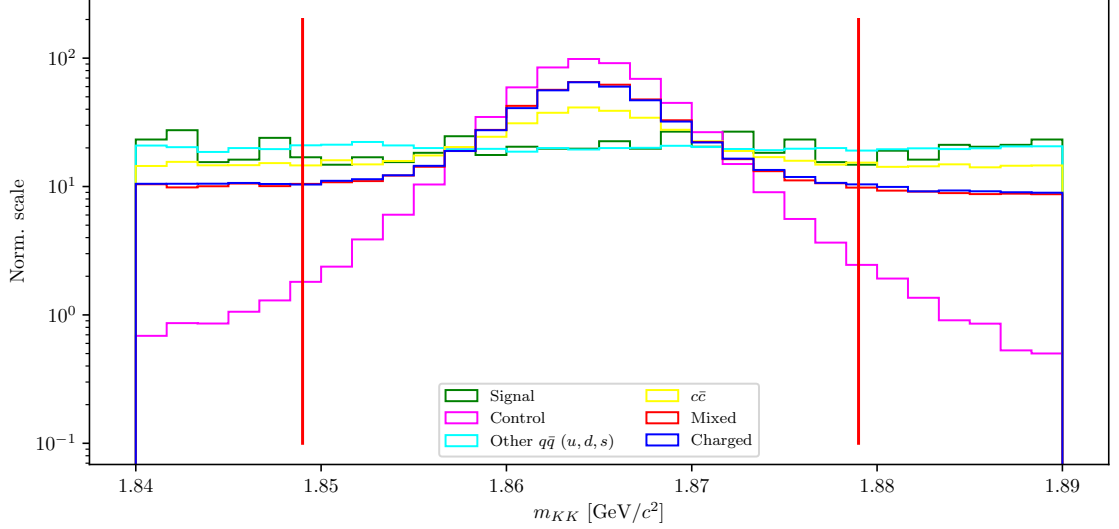


Figure 5.15.: Normalized distributions of m_{KK} for the full MC dataset. The red lines represent the cut on region in the m_{KK} distribution where the control sample is enhanced. The $m_K K$ distribution drops quickly for the case of the control decay, while staying uniform for other contributions.

With the control sample selection determined, we now run the reconstruction with and without the ROE clean-up procedure on MC and data. For the purpose of this validation, we only run the reconstruction over 1 stream of the full available generated MC.

The effects of the ROE clean-up are shown in Figure 5.16, where we overlay the data points to a stacked histogram of MC contributions for ΔE and M_{BC} . We see that data and MC agree well. A slight systematic trend can be seen in the ΔE variable, which is addressed in Section 7.1.1. The control sample resolution seems very poor in the case without the clean-up, but it improves significantly if the clean-up procedure is applied, as expected. The simulated background also seems to gain an improvement in the resolution, but this is likely due to the background consisting of similar candidates as the control sample. This means that the clean-up performs as expected due to the nature of the decays and does not arbitrarily shape the background to be more signal like. Additionally, it should be pointed out that, after the clean-up, the simulated background resolution is worse compared to the control decay resolution, while this is not the case if the clean-up procedure is not performed.

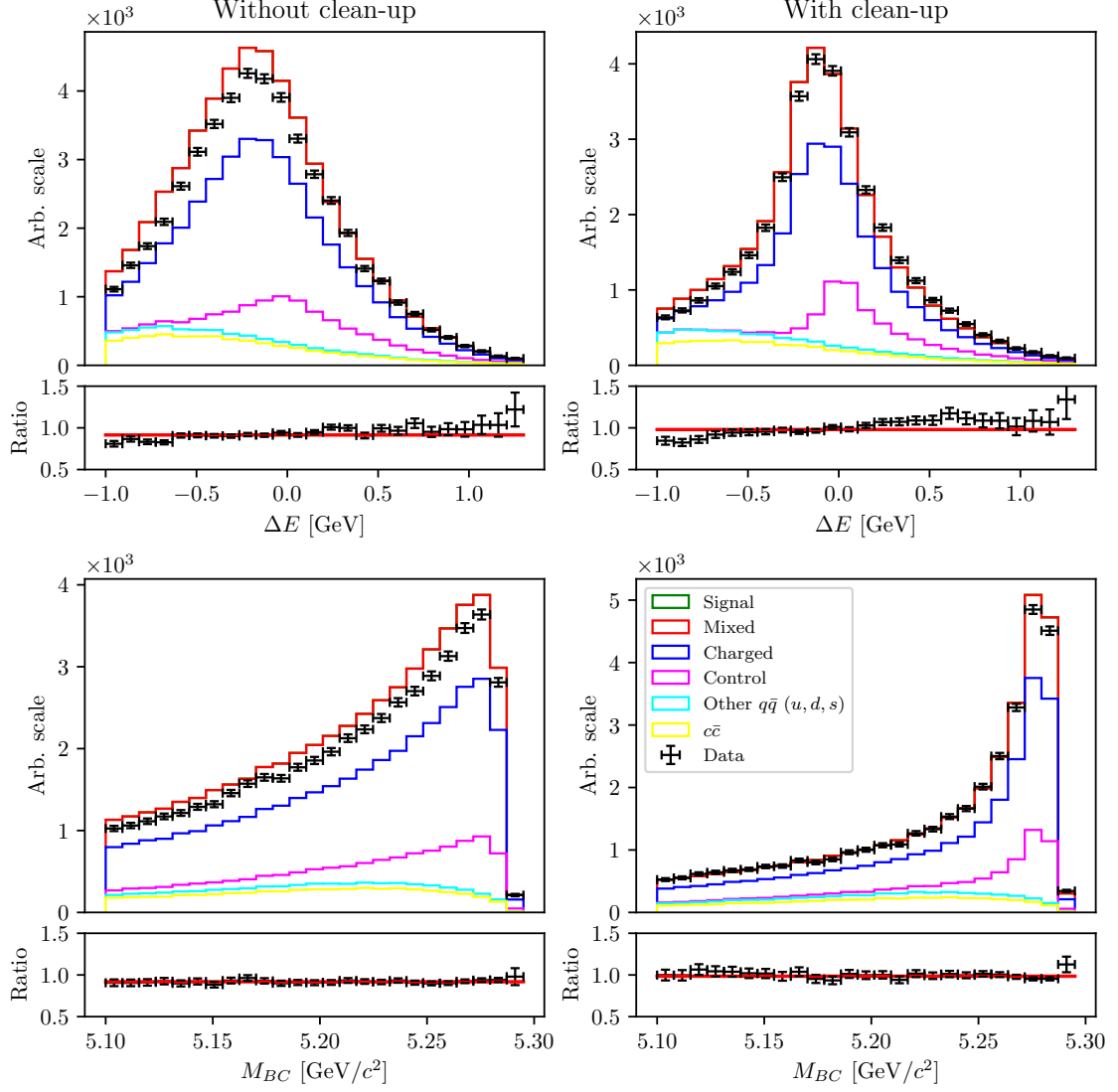


Figure 5.16.: Distributions of ΔE (top) and M_{BC} bottom for the case without (left) and with ROE clean-up (right). The resolution of the control sample is improved and the MC and data agree well in all aspects. While the simulated background resolution is also improved, it is worse compared to the resolution of the control sample.

922 To perform the clean-up validation in greater detail we also compare the data and
 923 MC agreement in bins of the charge product of the two B mesons. Figure 5.17 shows
 924 the cleaned-up versions of ΔE and M_{BC} for each charge product bin in the same
 925 manner as shown in the previous section. We see that the MC and data agreement
 926 persists in all cases.

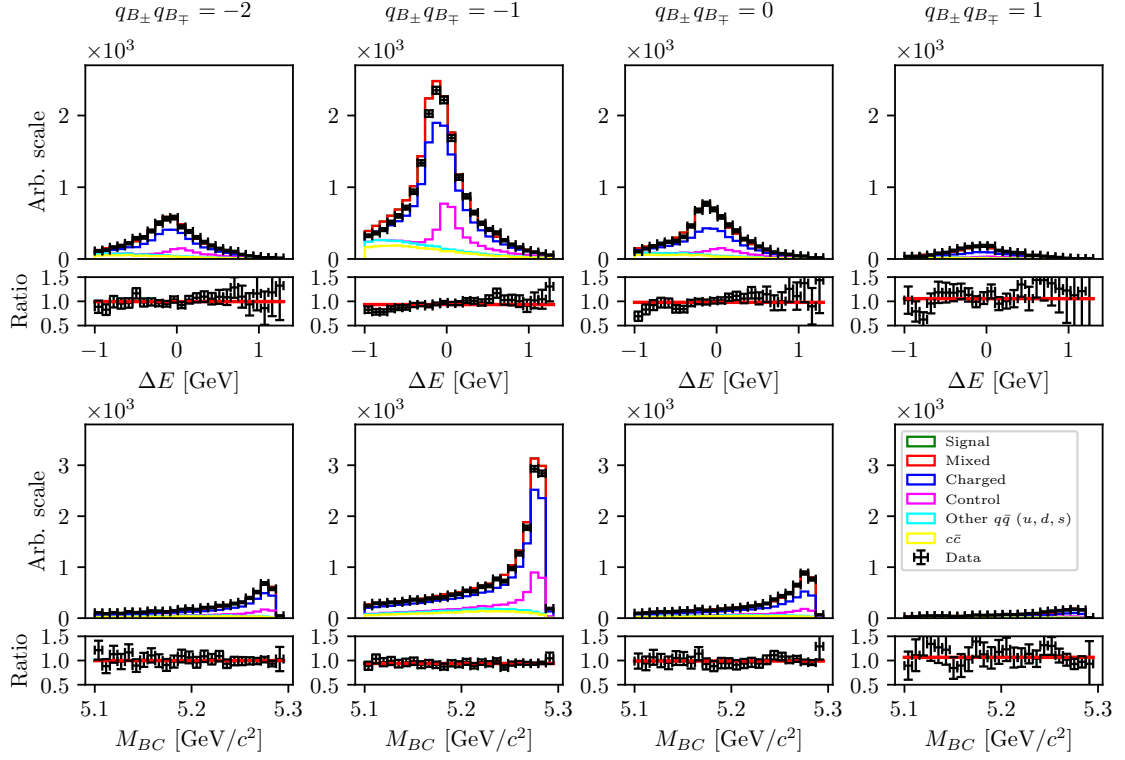


Figure 5.17.: Distributions of ΔE (top) and M_{BC} (bottom) split in bins of the charge product of the two B mesons.

927 The ROE clean-up procedure seems to perform well. It significantly improves the
 928 resolution of the signal/control candidates and increases the amount of perfectly
 929 clean events. The clean-up procedure was also applied to data and no disagreement
 930 with respect to the simulated MC samples was found. This means that the procedure
 931 does not differ between MC and data and does not affect them differently. The
 932 procedure was therefore validated in great detail and is suitable to be used in this
 933 analysis.

Chapter 6.

Background Suppression

This chapter shows the procedure in suppressing various kinds of backgrounds by applying cuts to MVA classifier outputs. More information about the MVA training, feature importance and hyper-parameter optimization for each MVA step in this chapter can be found in the Appendix B.

6.1. Resonant Background

In this analysis we study decays with kaons in the final state. This means that standard procedures in $b \rightarrow u$ analyses in order to suppress $b \rightarrow c$ backgrounds, such as K -veto, are not possible. As a consequence, our final sample consists of combinations of K pairs coming also from $b \rightarrow c$ sources, such as $D^0 \rightarrow K^+ K^-$. Such candidates usually have resonance-like properties in the two-kaon invariant mass spectrum. Figure 6.1 shows this invariant mass spectrum of two kaons, m_{KK} , where obvious resonant structures are present from sources like

- $\phi \rightarrow K^+ K^-$ (sharp resonance at $\sim 1.019 \text{ GeV}/c^2$),
- $D^0 \rightarrow K^+ K^-$ (sharp peak at $\sim 1.864 \text{ GeV}/c^2$),
- $D^0 \rightarrow K^+ \pi^-$ (wide, shifted peak, due to kaon miss-identification).

In order to suppress these resonant backgrounds, while studying signal or control decay, we impose a set of the following cuts

- Signal cut: $|m_{KK} - m_\phi| > \Delta_\phi$, $|m_{KK} - m_{D^0}| > \Delta_{D^0}$, $|m_{K\pi} - m_{D^0}| > \Delta_{D^0}$,
- Control cut: $|m_{KK} - m_{D^0}| \leq \Delta_{D^0}$, $|m_{K\pi} - m_{D^0}| > \Delta_{D^0}$,

where m_{KK} is the KK invariant mass and $m_{K\pi}$ is the invariant mass of KK , where the kaon's mass, which has the same charge as the B meson, was given the mass of the π^0 particle, and where $m_\phi \approx 1.019 \text{ GeV}/c^2$ and $m_{D^0} \approx 1.864 \text{ GeV}/c^2$ are nominal masses of the ϕ and D^0 mesons, and $\Delta_\phi \approx 8 \times 10^{-3} \text{ GeV}/c^2$ and $\Delta_{D^0} \approx 1.5 \times 10^{-2} \text{ GeV}/c^2$ are symmetric cut widths around the nominal mass values for the ϕ and D^0 mesons, respectively. By imposing the signal or control cut on our data, we are able to efficiently isolate the desired subset, which is very useful for further studies of the control decay. Table 6.1 shows the subsample efficiency after applying either of the cuts.

	$\epsilon(\text{Signal cand.})$	$\epsilon(\text{Control cand.})$	$\epsilon(\phi \text{ resonance cand.})$
Signal cut	95.4%	4.0%	13.6%
Control cut	1.9%	96.0%	0.0%

Table 6.1.: Various subset efficiencies after imposing the signal or control cut on the KK invariant mass.

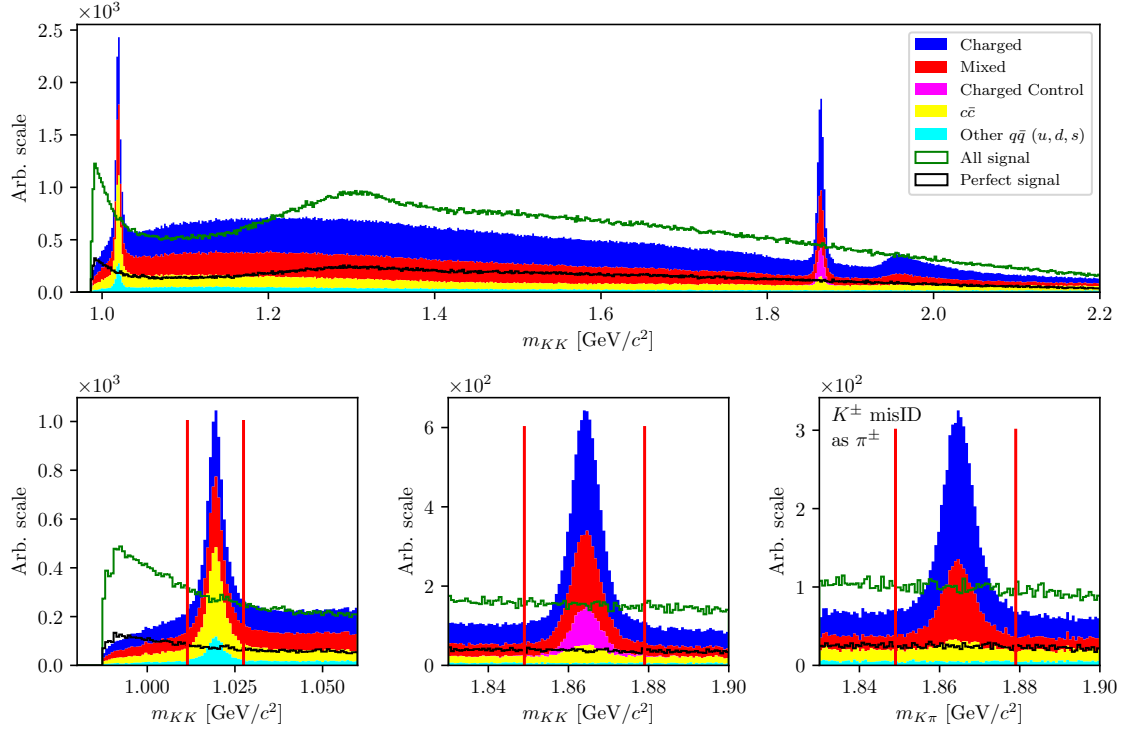


Figure 6.1.: Invariant mass of two correctly reconstructed kaons (left) and invariant mass of two kaons, where one was miss-identified as a pion (right). Signal (green) and perfect signal (black) are equally scaled up.

6.2. Continuum Suppression

Continuum background are physics processes where continuum states are produced in electron and positron collisions

$$e^+e^- \rightarrow q\bar{q},$$

where $q = u, d, s$ or c , and are a sizable contribution to $B\bar{B}$ events. Additionally to kinematic constraints to separate $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ decays from $e^+e^- \rightarrow q\bar{q}$, properties of the "event shape" are also often used, because phase-space distributions of decayed particles differ for these two processes. Continuum background events are generated in a back-to-back way in the CMS frame, so hadrons produced in the quark fragmentation possess only a small transverse momentum compared to the initial momentum magnitude. This leads to a spatially confined, jet-like structure. On the other hand, B mesons from $B\bar{B}$ events are produced almost at rest in the

973 CMS frame. Their decay products form an isotropic distribution in the detector,
 974 which yields a spherical event shape.

975 6.2.1. Characteristic Variables

976 Information on the phase-space distribution of decay particles is obtained in a num-
 977 ber of different ways. In this subsection different characteristic variables are pre-
 978 sented which are used in the MVA training. They all focus on kinematic and shape
 979 differences between the two processes, which we want to discriminate.

980 Thrust and Related Variables

981 It is possible to define a thrust axis \vec{T} for a collection of N momenta p_i as a unit
 982 vector along which their total projection is maximal. Thrust axis \vec{T} can be obtained
 983 by maximizing the expression

$$\vec{T} = \frac{\sum_i |\vec{T} \cdot \vec{p}_i|}{\sum_i |\vec{p}_i|}. \quad (6.1)$$

984 In this case, a related variable is $|\cos \theta_T|$, where θ_T is the angle between the thrust
 985 axis of the momenta from B meson decay particles and the thrust axis of all particles
 986 in the ROE. Since both B mesons in $B\bar{B}$ events are produced at rest, their decay
 987 particles, and consequentially their thrust axes, are uniformly distributed in the
 988 range $[0, 1]$. On the other hand, decay particles from continuum events follow the
 989 direction of the jets in the event. As a consequence, the thrusts of both the B meson
 990 and the ROE are strongly directional and collimated, which results in a large peak
 991 at $|\cos \theta_T| \approx 1$. Additionally, one can also use the variable $|\cos \theta_{TB}|$, which is the
 992 thrust axis between the B candidate and the beam axis. For B candidates from
 993 $B\bar{B}$ events, this distributions is again uniformly distributed, while for candidates
 994 from continuum events this distribution follows the distribution of the jets with the
 995 function $1 + \cos^2 \theta_{T,B}$. Practically, such a distribution exhibits a drop at $|\cos \theta_{TB}| \approx$
 996 1 , due to the acceptance loss of the detector in the direction of the beam pipes.
 997 Figure 6.2 shows the distributions of $|\cos \theta_T|$ (left) and $|\cos \theta_{T,B}|$ (right) for different
 998 B meson candidates.

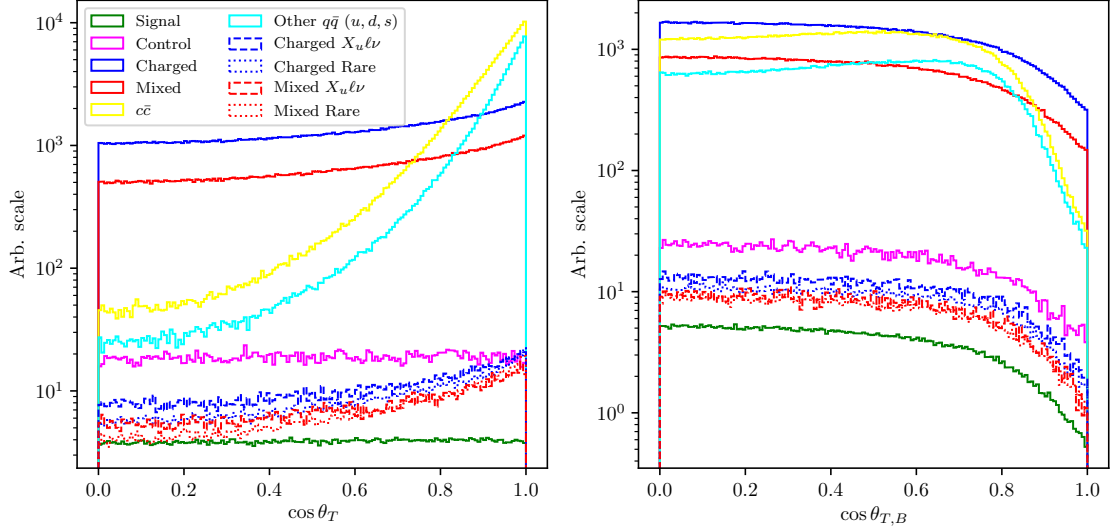


Figure 6.2.: Distributions of $|\cos \theta_T|$ (left) and $|\cos \theta_{T,B}|$ (right) for different B meson candidates.

999 CLEO Cones

1000 CLEO cones have been introduced by the CLEO collaboration [18] and are an ad-
 1001 dditional specific tool to provide optimal background discrimination. They are nine
 1002 variables corresponding to the momentum flow around the thrust axis of the B me-
 1003 son candidate, binned in nine cones of 10° around the thrust axis, as illustrated in
 1004 Figure 6.3.

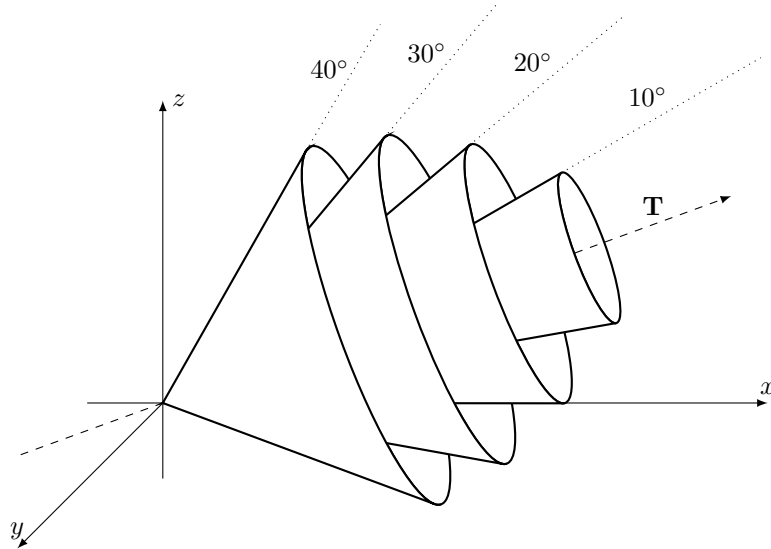


Figure 6.3.: Concept of CLEO cones. \vec{T} denotes the thrust axis of the B meson candidates in the event. Each variable corresponds to a momentum flow around the thrust axis in steps of 10° .

1005 KSFW Moments

1006 Fox-Wolfram moments are another useful parametrization of phase-space distribu-
 1007 tion of energy flow and momentum in an event. For a collection of N momenta p_i ,
 1008 the k -th order normalized Fox-Wolfram moment R_k is defined as

$$R_k = \frac{H_k}{H_0} = \frac{1}{H_0} \sum_{i,j} |p_i||p_j| P_k(\cos \theta_{ij}), \quad (6.2)$$

1009 where θ_{ij} is the angle between p_i and p_j , and P_k is the k -th order Legendre poly-
 1010 nomial. For events with two strongly collimated jets, R_k takes values close to 0 (1)
 1011 for odd (even) values of k , so these moments provide a convenient discrimination
 1012 between $B\bar{B}$ and continuum events.

1013 Belle developed a refined generation of Fox-Wolfram moments, called Kakuno-
 1014 Super-Fox-Wolfram (KSFW) moments to further suppress the continuum back-
 1015 ground. There are 17 different KSFW moments which are grouped into R_k^{so} , R_k^{oo}
 1016 and R_k^{ss} [19]. The latter ones are excluded due to correlations with B meson specific
 1017 variables.

1018 6.2.2. MVA Training

1019 Most of the characteristic variables, described in section 6.2.1, were taken together in
 1020 order to train a single MVA classifier for continuum suppression. All characteristic
 1021 variables were checked for possible q^2 , M_{BC} or ΔE correlation. Variables with sig-
 1022 nificant correlation or complex shapes in the correlation distribution were discarded
 1023 from the training set, since they would have introduced unwanted dependence on
 1024 the unreliable model, ISGW2, used for signal MC generation. Additionally, all of
 1025 the characteristic variables in our set do not depend on the signal mode, they only
 1026 differ in the kinematic and topological aspects of $B\bar{B}$ and continuum background
 1027 events.

1028 The training dataset consisted of 2×10^5 candidates, where 50 % of the candidates
 1029 are correctly reconstructed signal events, 25 % are $u\bar{u}$, $d\bar{d}$ and $s\bar{s}$ background with
 1030 expected proportions, and 25 % is $c\bar{c}$ background. Since the full Belle dataset is
 1031 experiment dependent, we construct the training dataset by proportionally sampling
 1032 each MC dataset, corresponding to the appropriate experiment number.

1033 The training variable set consisted of

- 1034 • B meson direction and thrust related variables
 - 1035 – magnitude of thrust axes of B and ROE ,
 - 1036 – cosine of the angle between the thrust axis of B and thrust axis of ROE ,
 - 1037 – cosine of the angle between the thrust axis of B and beam direction,
 - 1038 – reduced Fox-Wolfram moment R_2 ,
- 1039 • all 9 CLEO Cones
- 1040 • KSFW Moments
 - 1041 – R_{01}^{so} , R_{02}^{so} , R_{03}^{so} , R_{04}^{so} ,
 - 1042 – R_{10}^{so} , R_{12}^{so} , R_{14}^{so} ,

- 1043 $- R_{20}^{so}, R_{22}^{so}, R_{24}^{so},$
- 1044 $- R_0^{oo}, R_1^{oo}, R_2^{oo}, R_3^{oo}, R_4^{oo},$
- 1045 • FlavorTagging variables
- 1046 $- qp$ of $e, \mu, \ell,$
- 1047 $- qp$ of intermediate $e, \mu, \ell,$
- 1048 $- qp$ of $K, K/\pi,$ slow pion, fast hadron,
- 1049 $- qp$ of maximum $P^*, \Lambda,$ fast-slow-correlated (FSC),
- 1050 • Other
- 1051 $- \Delta z, \Delta t.$

1052 Figure 6.4 shows the classifier output for various types of background, all in ex-
 1053 pected MC proportions. B meson candidates from continuum background are dom-
 1054 inant at lower values, while candidates from $B\bar{B}$ events populate the region with
 1055 higher values.

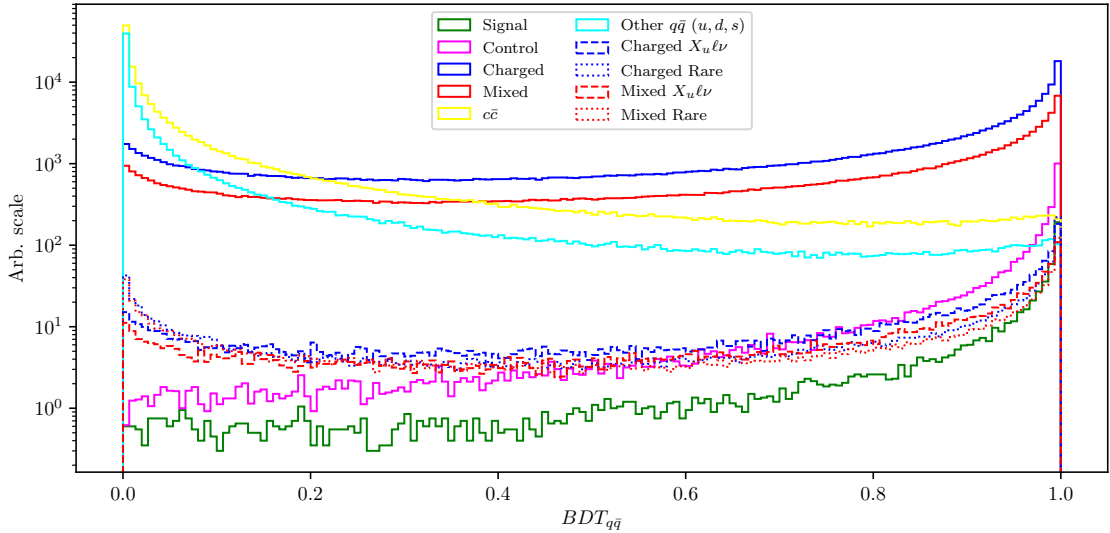


Figure 6.4.: Continuum suppression classifier output for signal and various types of background. B candidates from continuum events dominate the lower region, while candidates from $B\bar{B}$ dominate in the upper region of the classifier output.

1056 6.3. $B\bar{B}$ Suppression

1057 After separating continuum background from $B\bar{B}$ events, the next step is to train
 1058 an MVA classifier to recognize our signal candidates among the candidates of other
 1059 $B\bar{B}$ background. $B\bar{B}$ background consists of

- 1060 • $b \rightarrow c\ell\nu$ background,
- 1061 • $b \rightarrow u\ell\nu$ background,

- Other rare decays (radiative, penguin, rare 2- and 3-body decays, ...).

Similarly, the training dataset for this classifier consisted of 2×10^5 candidates, where 50 % of the candidates are correctly reconstructed signal events. The remaining part of the training dataset consists of all background, not including the control sample, because we are not interested in suppressing it directly. The background part of the dataset consists of 75 % charged and neutral $B\bar{B}$ events in equal proportions, whereas the remaining 25 % is equally populated with charged and neutral $B\bar{B}$ events from $b \rightarrow u\ell\nu$ and other rare decays. The training dataset was proportionally sampled in the same manner as described in section 6.2.2.

In order to separate this kind of background, we must be careful not to introduce correlations with the fit variables (ΔE , M_{BC}) or any kind of model dependence (correlation with q^2). This means that we can not use any information of the decay particles or the candidate, which is of kinematics nature, such as decay particles momenta, decay angles or other variables with such behavior.

The training variable set consisted of

- fit probability of $P(\chi^2, DOF)$ of the signal candidate and the ROE side, separately,
- $\cos \theta_{BY}$ from Eq. (4.4),
- cos of the angle between momentum and vertex of X , where $X \in [KK, KK\ell, KK\ell\nu]$,
- FlavorTagger variables for the two signal-side kaons,
- number of kaons, tracks and distant tracks in ROE,
- θ angle of the ROE momentum in CMS frame,
- ξ_Z from [20]
- Δz ,
- m_{miss}^2 from Eq. (4.11),
- $B \rightarrow D^*\ell\nu$ veto variables,

where distant tracks are all tracks in ROE which satisfy the condition of $|d_0| > 10.0$ cm or $|z_0| > 20.0$ cm. The last entry, $B \rightarrow D^*\ell\nu$ veto variables, are a set of variables where we partially reconstruct the D^* candidate 4-momentum via a linear combination of the π_s^\pm 4-momentum in the $D^* \rightarrow D\pi_s^\pm$ decay. It helps discard the most dominant $B \rightarrow D^*\ell\nu$ background. It is most efficient in the $B^0 \rightarrow D^{*-}\ell^+\nu$ decay, where D^{*-} further decays via $D^{*-} \rightarrow \bar{D}^0\pi_s^-$. Other decays do not contain a charged π_s particle and are harder to reconstruct with good precision. This results in larger suppression of the neutral $B\bar{B}$ background only. Figures 6.5 shows the veto variable with a partial reconstruction of a charged π_s^\pm .

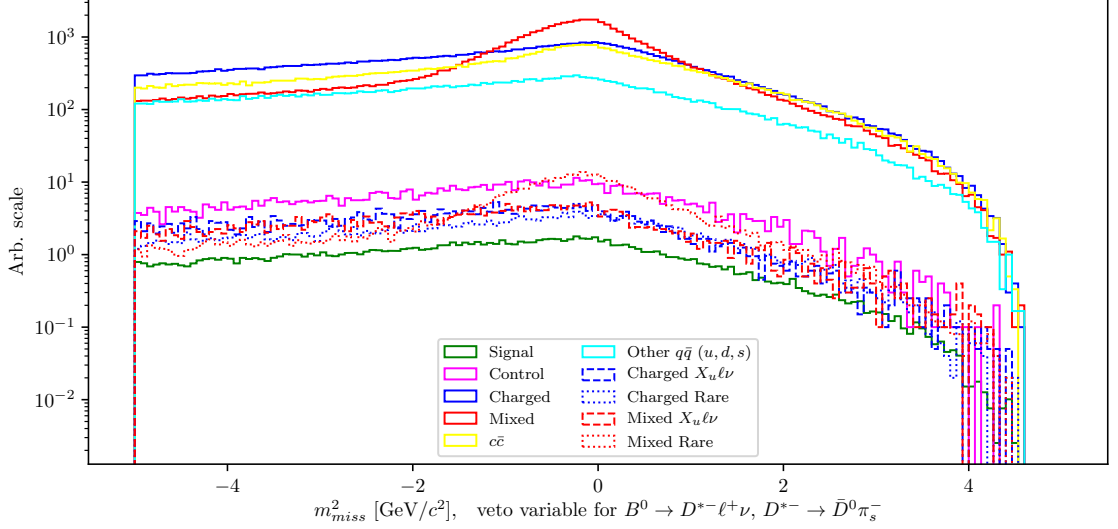


Figure 6.5.: Distribution of m^2_{miss} for partially reconstructed $B^0 \rightarrow D^{*-} \ell^+ \nu$ decays, which serves as a veto.

When the training is finished and the hyper-parameters of the classifier are optimized, the classifier output, as shown in Figure 6.6 (left), can be used for background suppression. B meson candidates from $B\bar{B}$ background are dominant at lower values, while candidates from $B\bar{B}$ events populate the region with higher values. Since the differences between signal and background $B\bar{B}$ events are smaller than $B\bar{B}$ and $q\bar{q}$ events, the resulting classifier has a smaller separation power than in the previous section.

6.3.1. Boosting to Uniformity

The selection approach with standard classifiers is optimal for counting experiments, as it by construction produces the optimal selection for observing an excess of signal over background events. Today's BDT algorithms, which work in this way, produce non-uniform selection efficiencies and may, as a consequence, shape background distributions to look like signal. In order to minimize such behavior, it is possible to discard variables, which are correlated with the variable of interest (in our case ΔE and M_{BC}), from the training set. This, however, decreases the classifiers discriminating power. Another approach is to use a novel boosting method, uBoost, which is trained to optimize an integrated FOM under the constraint that the BDT selection efficiency for the desired class must be uniform. The uBoost algorithm balances the biases to produce the optimal uniform selection [21].

The training set used in this training is the same as described at the beginning of this chapter, along with the same set of training variables. It will be seen later that the standard BDT classifier shapes the background to look like signal mostly in the M_{BC} picture, therefore we train the uBDT classifier with a uniformity constraint in the M_{BC} variable for background candidates with the uBoost algorithm. The resulting classifier output is shown in Figure 6.6 (right). For this classifier, the separation power between signal and background seems worse, however, the shapes of backgrounds differ significantly, which greatly affects the performance of signal extraction.

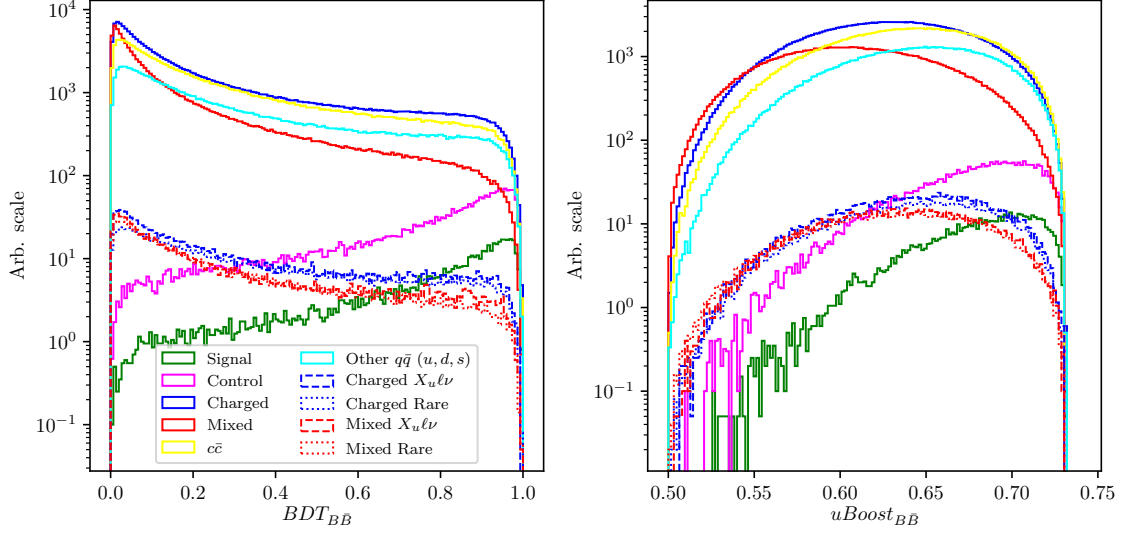


Figure 6.6.: $B\bar{B}$ suppression classifier output for signal and various types of background for the standard BDT classifier (left) and the uBDT classifier (right). B candidates from $B\bar{B}$ background events dominate the lower region, while signal and control candidates dominate in the upper region of the classifier output.

6.4. Selection Optimization

Instead of two separate $q\bar{q}$ and $B\bar{B}$ FOM optimizations, it is more efficient to do a simultaneous 2D FOM optimization, since the two classifiers are not completely uncorrelated. In the same manner, as before, FOM is optimized for perfectly reconstructed signal candidates in the signal window, after all the pre-cuts, signal categorization, and after cutting out the background resonances and the control decay. The FOM plot with the optimal point for both $B\bar{B}$ MVA classifiers is shown in Figure 6.7.

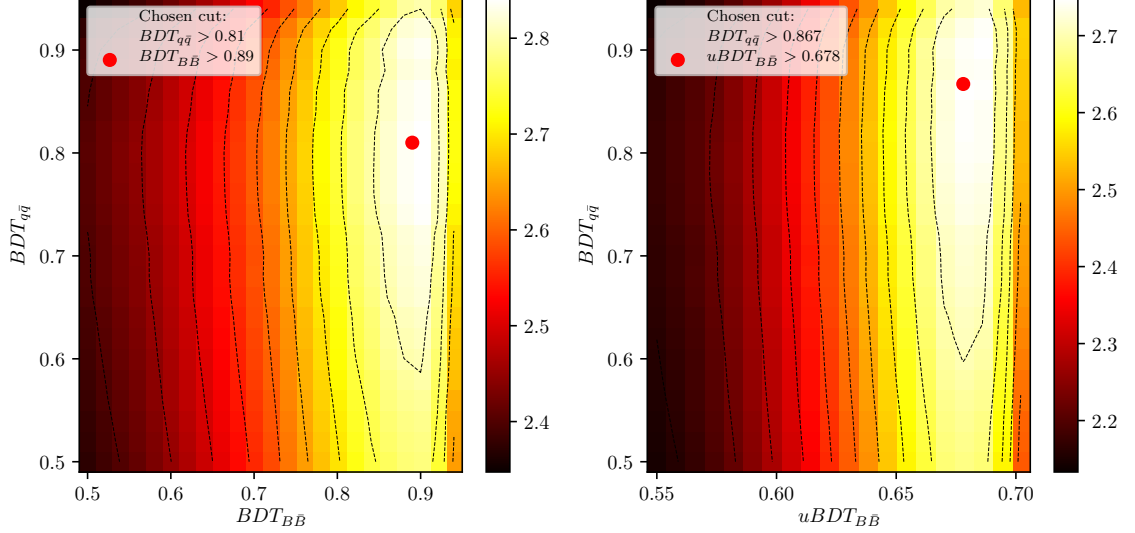


Figure 6.7.: 2D FOM optimization of continuum suppression classifier and the standard BDT (left) and uBDT (right) $B\bar{B}$ suppression classifier.

We can compare signal and major background distributions of ΔE and M_{BC} after the 2D FOM optimization for both classifiers. Figure 6.8 shows the arbitrary (left) and normalized scale (right) for ΔE (top) and M_{BC} (bottom) for the final sample optimized with the standard BDT classifier, while Figure 6.9 shows similarly for the final sample optimized with uBDT classifier. We can see that there is considerably more background in the latter case, however, also shapes of background and signal distributions differ greatly, meaning there is less room for correlation. The biggest change seems to be in the shape of the M_{BC} distribution, where the background component is much more signal like in the final sample optimized with the standard BDT classifier than in the other case. Additionally, the shapes are more easily constrained in the latter case, since they are present in regions where no signal is expected. The total numbers of expected signal candidates and the signal-to-noise ratios for both classifiers are:

- Standard BDT: $N_{sig} = 176$, $N_{sig}/N_{bkg} = 4.83 \%$,
- uBDT: $N_{sig} = 264$, $N_{sig}/N_{bkg} = 1.33 \%$.

Due to the large difference in ΔE and M_{BC} shape, we will continue the analysis with the uBDT classifier, although the comparison between both methods will be shown for the final fit result in the next chapter.

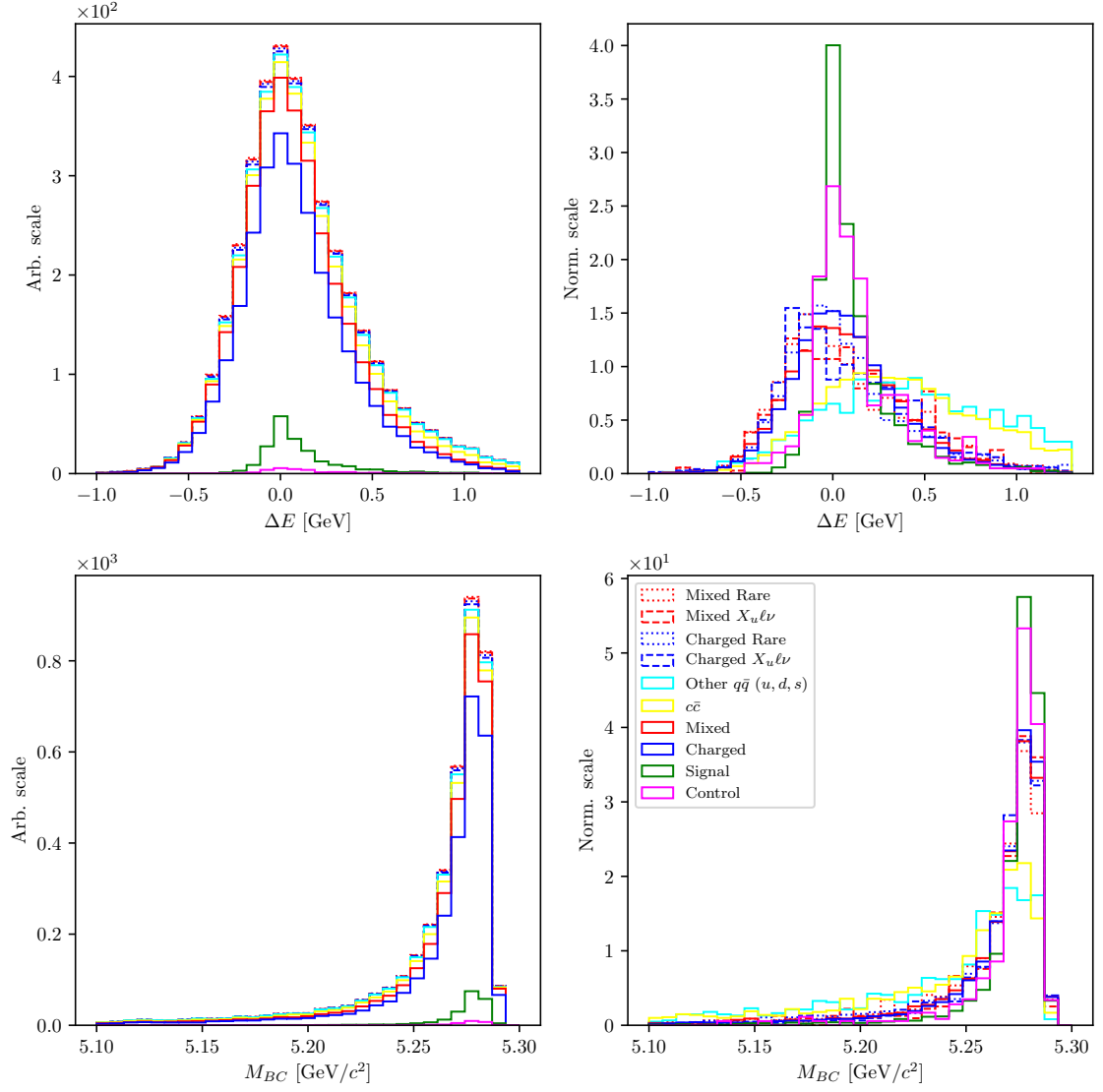


Figure 6.8.: Arbitrary (left) and normalized scale (right) for ΔE (top) and M_{BC} (bottom) for the final sample optimized with the standard BDT classifier.

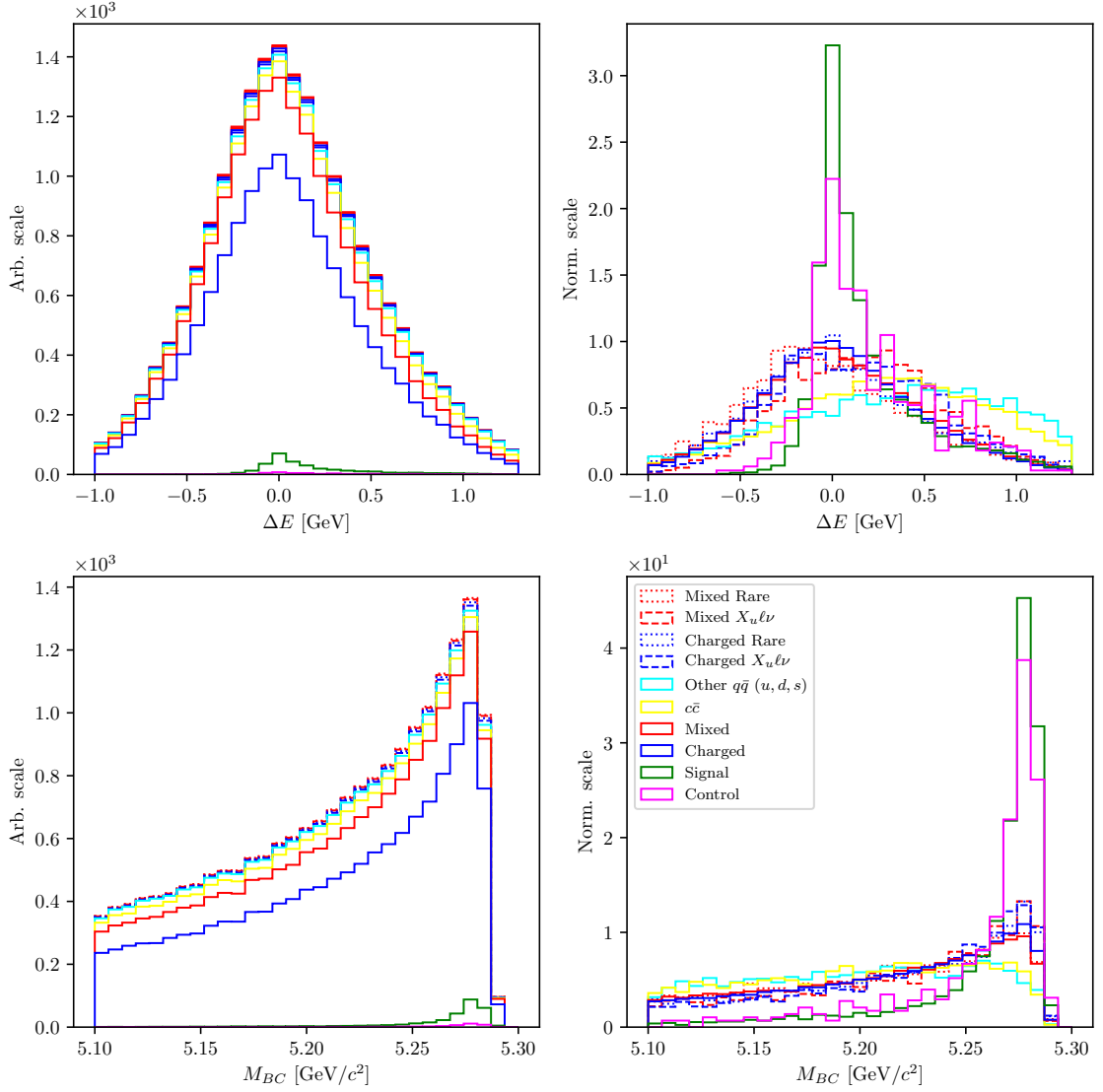


Figure 6.9.: Arbitrary (left) and normalized scale (right) for ΔE (top) and M_{BC} (bottom) for the final sample optimized with the uBDT classifier for $B\bar{B}$ suppression.

6.4.1. $B\bar{B}$ Background Composition and Lepton Veto

The majority of background candidates after the final selection is represented by candidates from $B\bar{B}$ events. In order to suppress this background even further, we need to take a look at its structure and recognize various contributions to this part of the background. Figure 6.10 shows ΔE and M_{BC} for the most significant contributions, along with m_{KK} . While most of the candidates come from events where all reconstructed charged particles in the signal decay do not come from one B meson, but both of them, these candidates are not so problematic, because their distribution is rather smooth and frequent in regions where we expect no signal. On the other hand, there are also contributions from B meson decays which produce more signal-like distributions. We will denote the first kind of background as $\Upsilon(4S)$ -matched and the second kind as the B -matched $B\bar{B}$ background. Fortunately, these decays are well known and well measured, so their yields can be constrained. Especially

1165 problematic is the double semileptonic decay $B \rightarrow \bar{D}^{(*)}\ell^+\nu$, $D^+ \rightarrow \bar{K}^-\ell^+\nu$, where
 1166 the secondary lepton is misidentified as a kaon. Even though the decay has two neu-
 1167 trinos, these events survive the m_{miss}^2 selection cut and produce peaks at the same
 1168 positions as signal distributions, while exhibiting only a slightly worse resolution.

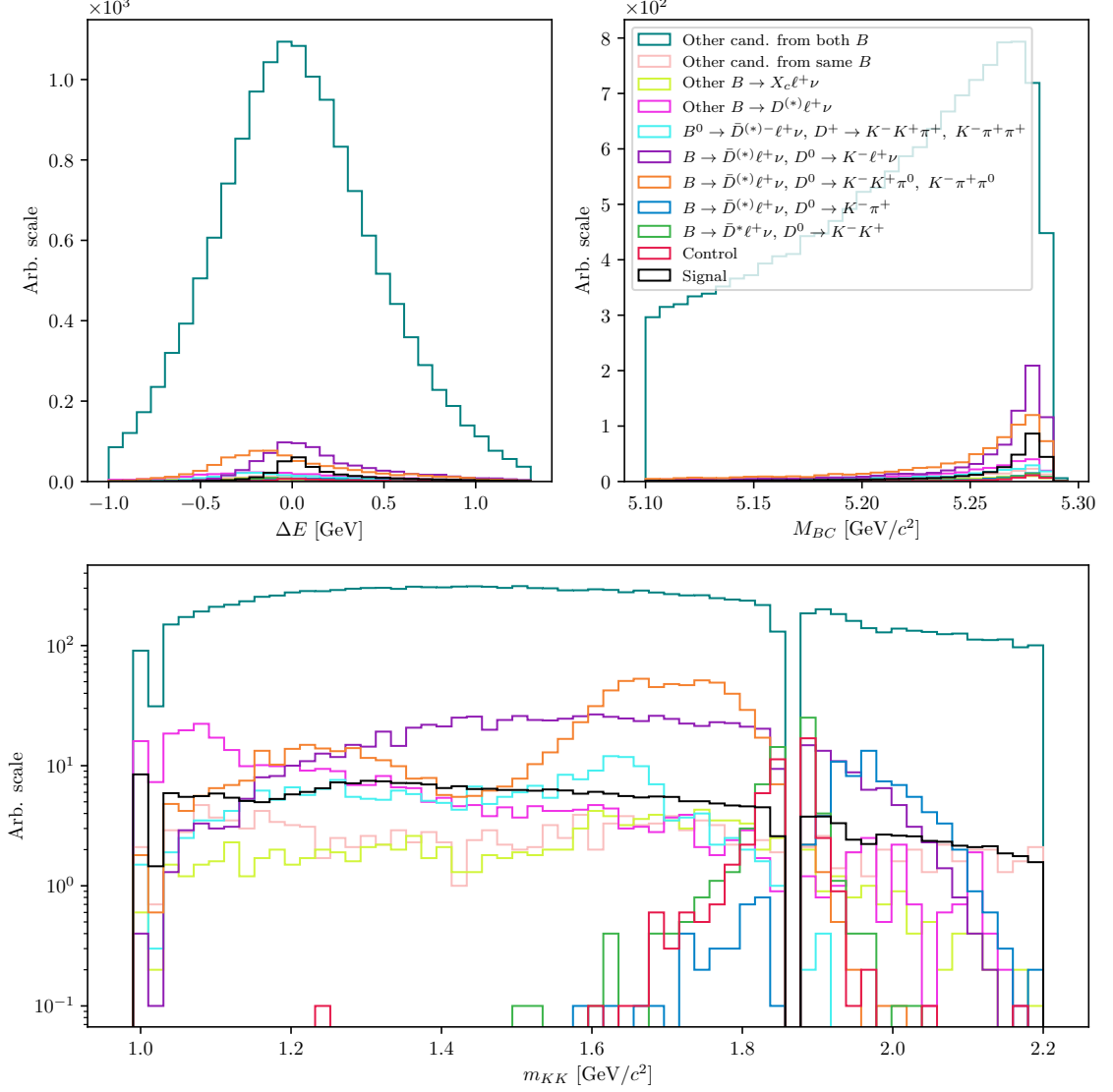


Figure 6.10.: ΔE (left), M_{BC} (right) and m_{KK} (bottom) for major contributions to the $B\bar{B}$ background in the signal cut region. $\Upsilon(4S)$ -matched backgrounds represents the majority, but has a smooth and wide background, which has a distinguishable shape from signal. B -matched contributions show a peak in M_{BC} and sometimes in ΔE , but can be constrained using existing measurements.

1169 In order to suppress these candidates, a lepton veto can be applied to both kaons,
 1170 stating that neither of the kaons should exhibit lepton-like properties. On the can-
 1171 didates passing the final selection, we optimize the eID and μID PID cuts, where
 1172 S and B in Eq. 4.1 are represented by perfect signal candidates and by background

1173 candidates, respectively, whereas in the latter case a lepton has been misidentified
 1174 as a kaon. 2D FOM plots for both kaons are shown in Figure 6.11, where it can be
 1175 seen that in the majority of the cases, an electron is misidentified as the kaon with
 1176 the opposite charge to the B meson. With the optimal cuts of

- 1177 • K_0 : $eID < 0.8$,
- 1178 • K_1 : $eID < 0.1$, $\mu ID < 0.8$,

1179 we reject 77.5% of candidates from the double semileptonic decays, while efficiency
 1180 loss of signal and other types of $B\bar{B}$ background is about 5 – 6%. The $B\bar{B}$ back-
 1181 ground after the lepton veto cuts is shown in Figure 6.12 for the signal cut region,
 1182 and in Figure 6.13 for the control cut region.

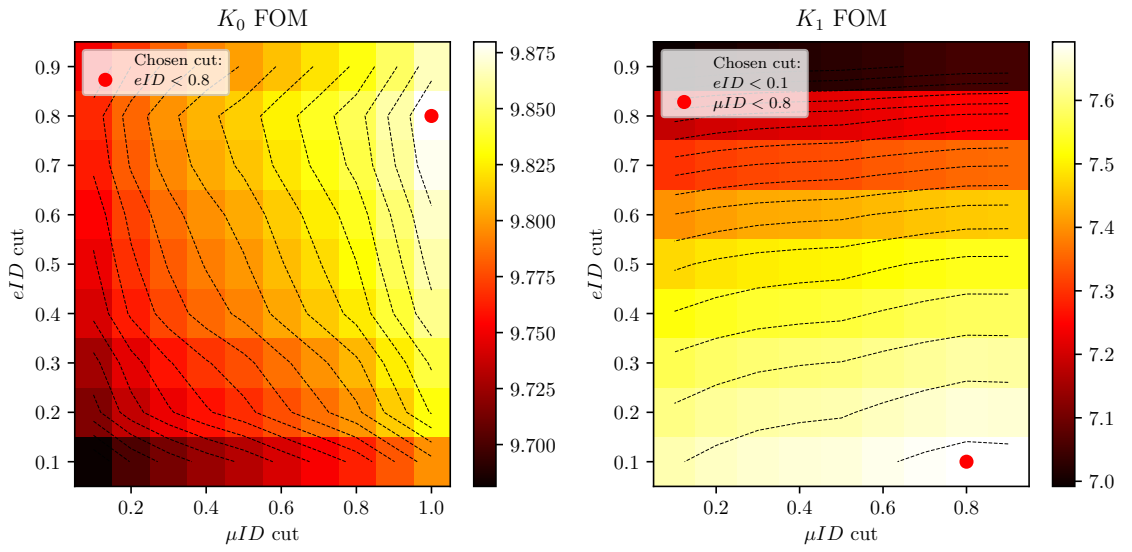


Figure 6.11.: 2D FOM for optimal eID and μID cuts on same-sign (left) and opposite-sign (right) kaons with respect to the B meson charge. For double semileptonic background component, in most cases an electron is misidentified as the opposite-sign kaon in the reconstruction chain.

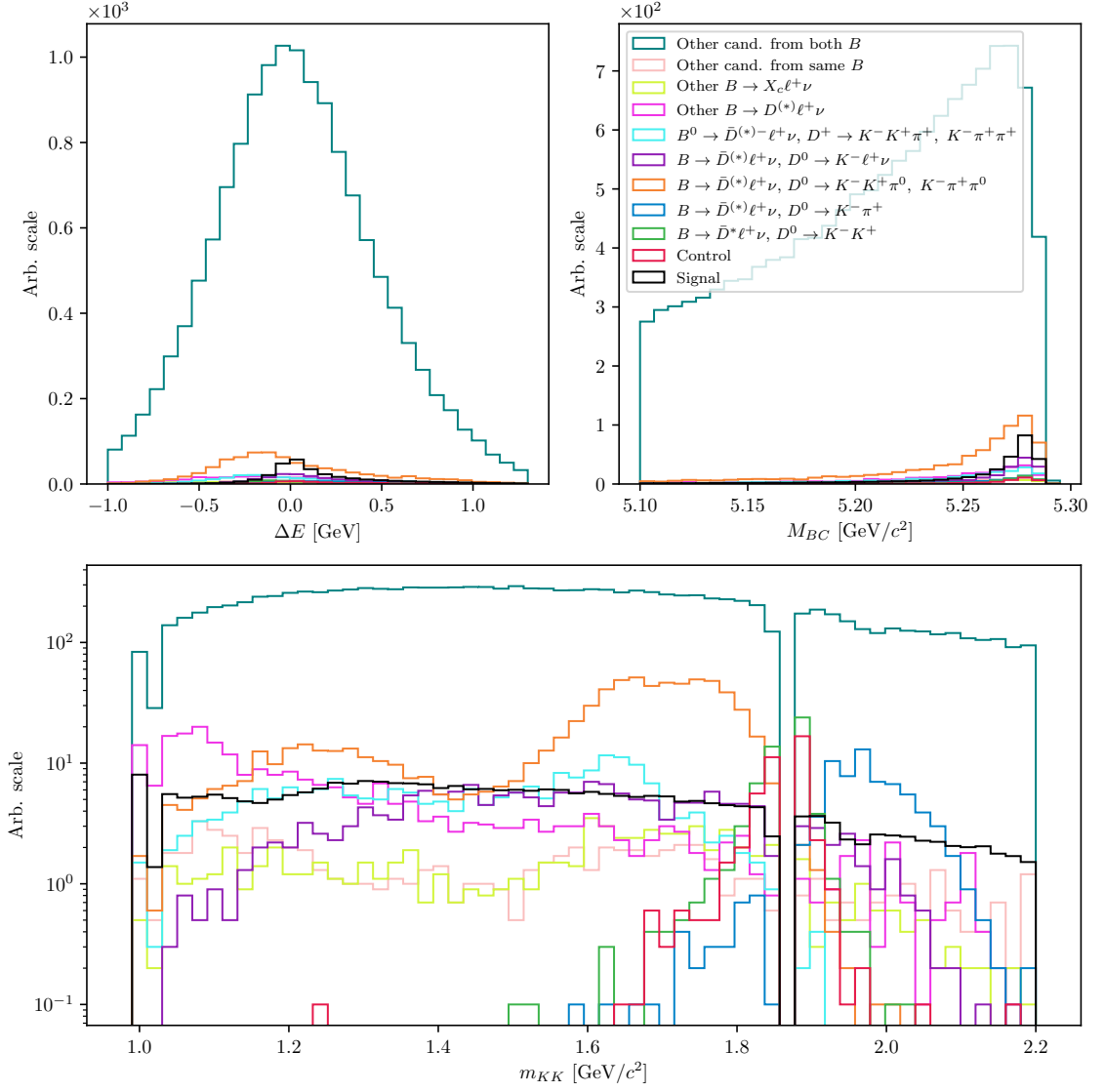


Figure 6.12.: ΔE (left), M_{BC} (right) and m_{KK} (bottom) for major contributions to the $B\bar{B}$ background in the signal cut region after the lepton veto. The double semileptonic background component is suppressed by a factor of 4 – 5.

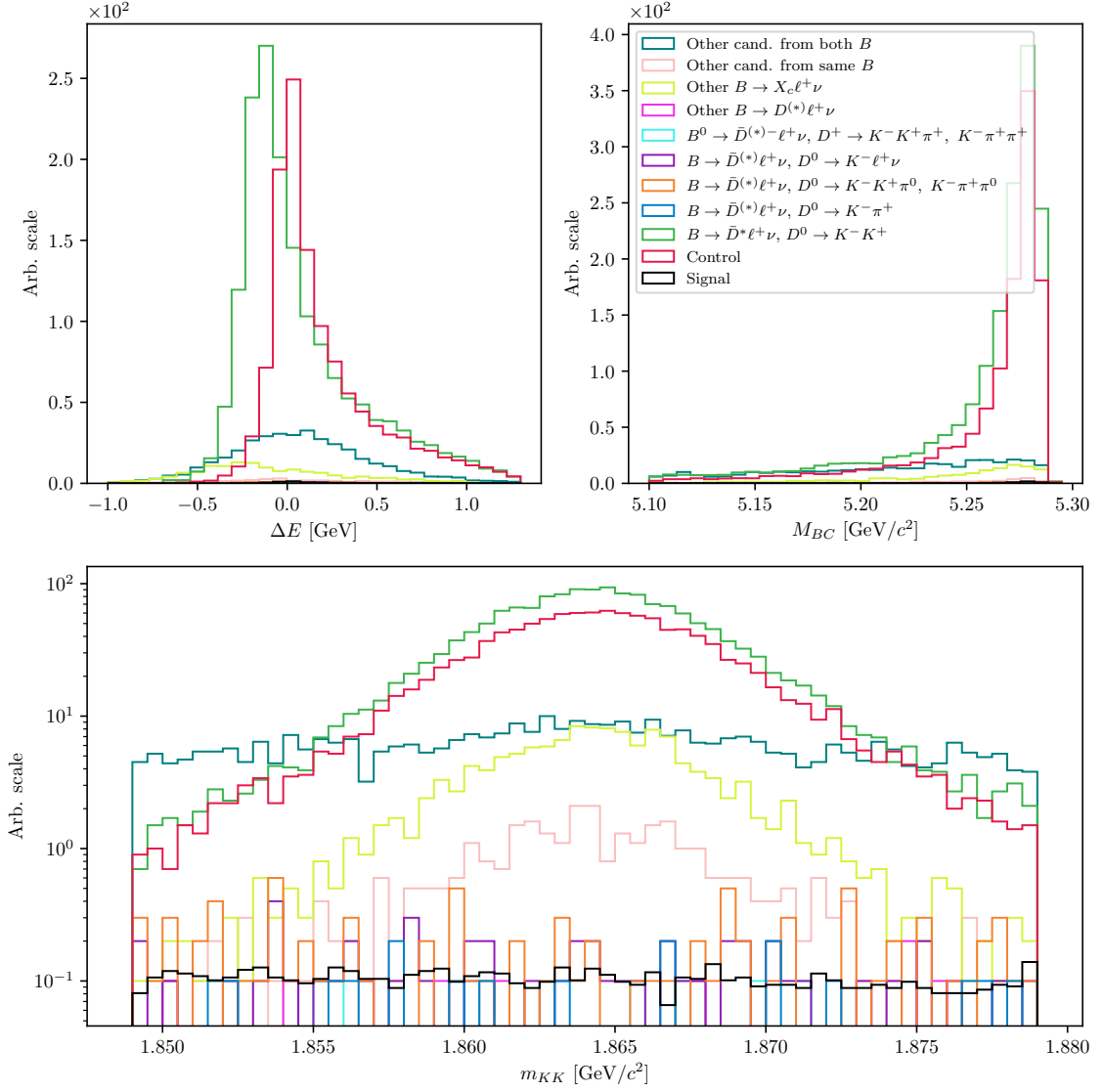


Figure 6.13.: ΔE (left), M_{BC} (right) and m_{KK} (bottom) for major contributions to the $B\bar{B}$ background in the control cut region after the lepton veto. The major component in this case are other $B \rightarrow D^* \ell + \nu, D \rightarrow K^+ K^-$ decays, besides the control decay.

1183 6.5. Data and MC Agreement

1184 With the final selection in place, we can check the data and MC agreement by
 1185 checking the control decay region in on- and off- resonance data. Off-resonance
 1186 samples provide the ability to check the agreement of the $q\bar{q}$ background component,
 1187 while on-resonance samples can be used to check the validity of the control MC
 1188 sample and, consequentially, the signal MC sample.

1189 6.5.1. Off-resonance Data

1190 The off-resonance data were collected at 60 MeV below the $\Upsilon(4S)$ resonance peak
1191 energy in order to determine the non- $B\bar{B}$ background. It, therefore, offers a direct
1192 view of the $q\bar{q}$ background data sample, which we can compare to the off-resonance
1193 MC sample. Figure 6.14 shows ΔE , M_{BC} and the $q\bar{q}$ classifier output, $BDT_{q\bar{q}}$, for
1194 off-resonance data and stacked MC in the control region, before any MVA cuts,
1195 where the MC sample was scaled down by a factor of 6, due to 6 streams of MC.
1196 These figures do not show a fit to data, but merely an overlay of data and stacked MC
1197 distributions, and show good data and MC agreement of the off-resonance sample
1198 already before the fit. More importantly than the normalization, the shape of data
1199 and MC also seems to be the same, so further corrections of ΔE and M_{BC} on MC
1200 are not necessary. This is also consistent by the flatness of the ratio function for ΔE
1201 and M_{BC} , shown in the same Figure. There seems to be a difference in the classifier
1202 performance for the continuum background suppression on data and MC. This leads
1203 to further differences between data and MC in the $q\bar{q}$ sample after the classifier cut,
1204 but we estimate that these differences are negligible, since a relatively small amount
1205 of continuum background passes the selection, compared to other background types.

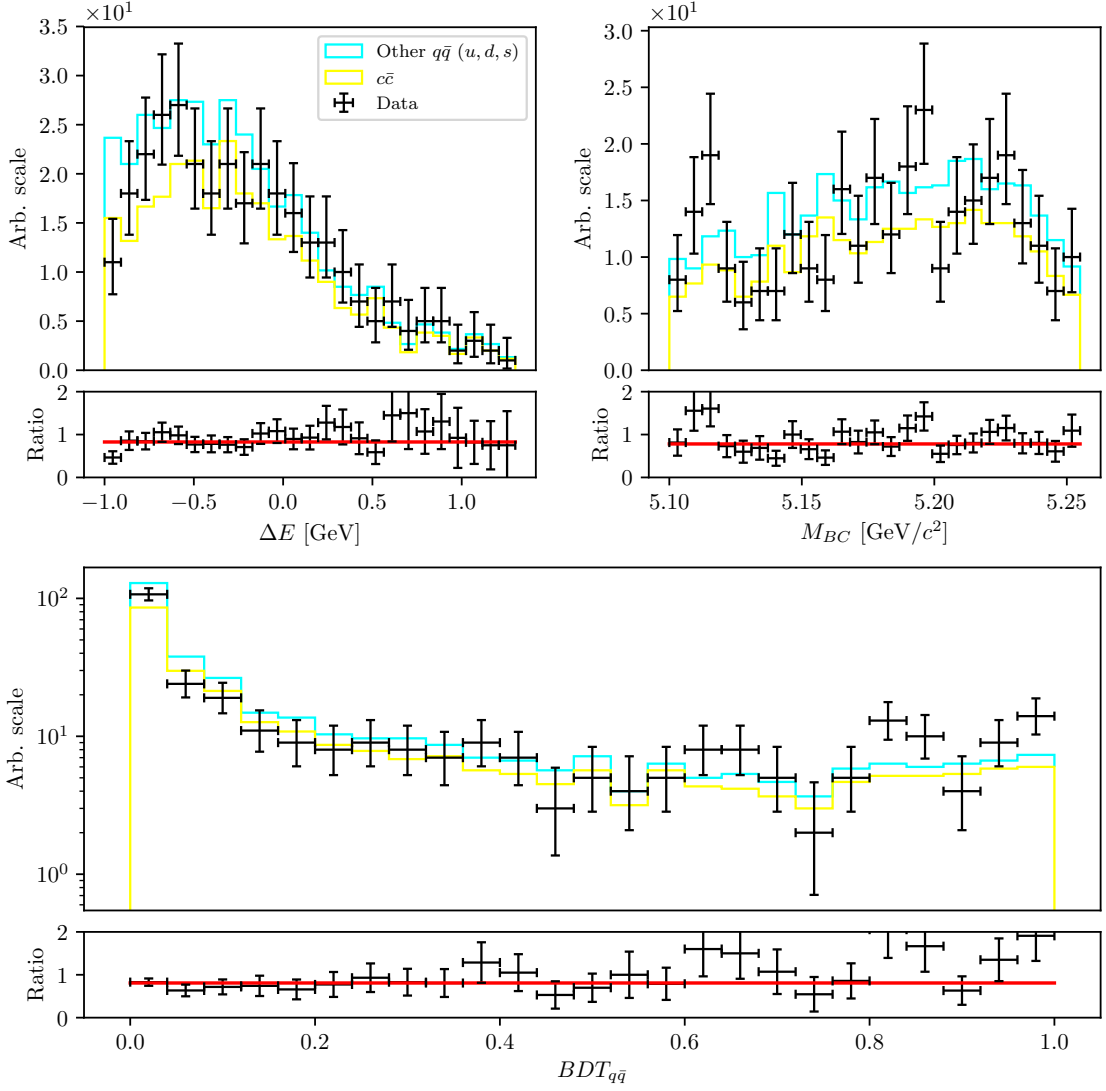


Figure 6.14.: ΔE (left), M_{BC} (right) and the $q\bar{q}$ classifier output (bottom), for off-resonance data and MC in the control region before any MVA cuts.

6.5.2. On-resonance Data

We can repeat the check on on-resonance data. Figure 6.15 shows ΔE , M_{BC} and $BDT_{q\bar{q}}$, where one can see inconsistencies between data and stacked MC on the lower spectrum, where continuum background is dominant. These figures again do not show the fit, but merely an overlay of data to the stacked MC distributions, where we see that the MC is over-estimated in the lower region of the continuum suppression MVA output, most likely due to additional disagreements from other sources. On the other hand, data and MC seem to agree well in the upper part of the spectrum, where $B\bar{B}$ events are dominant. Overall, data and MC seem to agree well already off-the-shelf after all the pre-cuts and without any corrections. This means that the modeling of this MC sample is very precise in this particular region of data and that there are no significant differences between data and MC for the control sample and the signal sample.

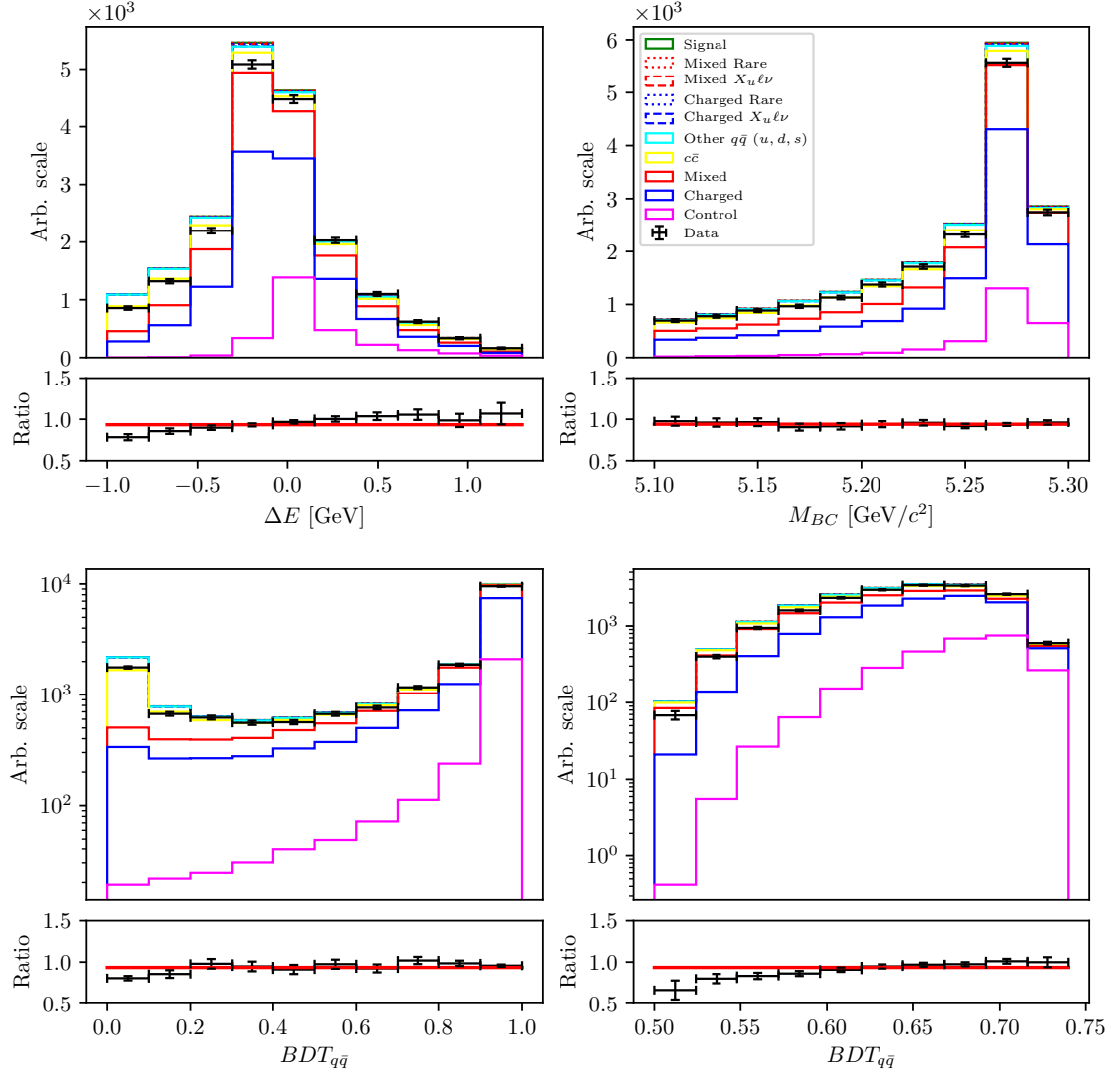


Figure 6.15.: ΔE (left), M_{BC} (right) and the $q\bar{q}$ classifier output (bottom), for on-resonance data and MC in the control region before any MVA cuts.

Chapter 7.

Extraction of Physical Parameters

In this chapter, the procedure for signal yield extraction is presented. We use the framework of `RooFit` [22] where we define 2D histogram templates in ΔE and M_{BC} , based on MC, for signal and several types of background. Using these templates, the independent full sample is fitted with the binned extended maximum likelihood (ML) fit, so that the individual template ratios and their sum describe the fitted sample as best as possible. In particle physics we are often dealing with low numbers of events and need to account for the Poisson nature of the data, therefore we use the likelihood fit, since it takes the Poisson errors into account, unlike the χ^2 fit, where the errors are assumed to be Gaussian. In this procedure, we attempt to find the parameter values that maximize the likelihood function, given the observations.

If $P(n|\vec{\alpha})$ is the probability of measuring n candidates, where $\vec{\alpha}$ is a set of parameters on which P depends, we can define the likelihood function L for a series of such measurements (i.e., bins in histogram) n_i based on Poisson statistics as

$$L(\vec{\alpha}) = \prod_{i=1} P(n_i|\vec{\alpha}) = \prod_{i=1} \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}, \quad (7.1)$$

where μ_i is the expected value for each measurement. It is also common to search for the minimum of the negative value of $\ln L$, or negative log-likelihood (NLL), as

$$\mathcal{L}(\vec{\alpha}) = -\ln L(\vec{\alpha}) = -\sum_i \ln \left(\frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!} \right) = \sum_i \ln(n_i!) + \mu_i - n_i \ln(\mu_i). \quad (7.2)$$

Maximizing L or minimizing \mathcal{L} gives us a maximum likelihood estimate of the set of parameters $\vec{\alpha}_{ML}$ which best describe the observed data.

The ML method provides a method to estimate the fit uncertainty. This is especially useful if the log-likelihood has a non-parabolic shape, which leads to asymmetric errors. We calculate the errors using the `MINOS` algorithm from the `MINUIT` package [23], which is implemented in `RooFit`. The algorithm follows the log-likelihood function out of the minimum to find the appropriate intervals of confidence for each parameter, taking the parameter correlations into account.

To estimate the goodness of the likelihood fit, one option is to generate toy MC experiments and obtain the expected log-likelihood distribution. Likelihood fits, however, also offer another way to test the goodness of fit via the likelihood ratio (LR), where we compare the likelihood obtained under the ML parameters $\vec{\alpha}_{ML}$, to the likelihood obtained under the null hypothesis parameters $\vec{\alpha}_{H_0}$. This determines how likely the data is under one model than the other. We define the LR test as

$$\lambda = -2 \ln \left(\frac{L(\vec{\alpha}_{ML})}{L(\vec{\alpha}_{H_0})} \right) = -2 [\ln L(\vec{\alpha}_{ML}) - L(\vec{\alpha}_{H_0})] \sim \chi_q^2, \quad (7.3)$$

1250 which asymptotically behaves as the χ_q^2 distribution with $q = m - n$ degrees of free-
1251 dom, where m and n are degrees of freedom of $L(\vec{\alpha}_{ML})$ and $L(\vec{\alpha}_{H_0})$, respectively. In
1252 particle physics we usually study a specific decay and try to perform measurements
1253 of the signal yield, so the null hypothesis in this case is that we expect to observe
1254 no signal. This means that for the null hypothesis we fix the expected signal yield
1255 parameter to zero, while leaving the other parameters of $\vec{\alpha}_{H_0}$ the same as in $\vec{\alpha}_{ML}$,
1256 which results in $n = m - 1$ degrees of freedom and in their difference $q = m - n = 1$.
1257 For such a simple LR test of a single parameter the LR test then follows the χ^2
1258 distribution with 1 degree of freedom. In this case we can define the fit significance
1259 from the χ^2 value in units of σ as

$$\text{Significance} = \sqrt{\lambda} = \sqrt{\chi^2}. \quad (7.4)$$

1260 7.1. Fit Setup

1261 We perform 10 fits to each stream of MC, where 9 streams were used for the creation
1262 of the templates and the remaining stream was used as fitted data. When fitting
1263 real data, all available MC was used for creating the templates. The full signal MC
1264 sample was used for the signal template definition in case of MC as well as data fits.
1265 The signal part of the `ulnu` sample was not used in template construction, it was
1266 only used as a part of the fitted sample.

1267 The same MC samples are used for template construction as described in chapter
1268 2,

- 1269 • signal MC,
- 1270 • 10 streams of `charged` and `mixed` $B\bar{B}$ background,
- 1271 • 6 streams of $c\bar{c}$ (`charm`) and other $q\bar{q}$ (`uds`) background,
- 1272 • `ulnu` sample, corresponding to $20\times$ integrated luminosity of the full Belle
1273 dataset,
- 1274 • `rare` sample, corresponding to $50\times$ integrated luminosity of the full Belle
1275 dataset.

1276 7.1.1. Control Fit

1277 $B\bar{B}$ background composition in control region is shown in Figure 6.13. Due to the
1278 strict cut of the m_{KK} around the D^0 mass window, most of the decays with a D^0
1279 proceed via $D^0 \rightarrow K^+K^-$ decay. In this case, the following fit templates are chosen

- 1280 • signal template,
- 1281 • $q\bar{q}$ template,
- 1282 • C_0 : $B^+ \rightarrow \bar{D}^0 \ell^+ \nu$, $D^0 \rightarrow K^- K^+$ (control decay),
- 1283 • C_1 : $B \rightarrow \bar{D}^* \ell^+ \nu$, $D^0 \rightarrow K^- K^+$,
- 1284 • other $B\bar{B}$ BKG template.

1285 In control fits, all template shapes are fixed and the yields of all templates are floated,
 1286 except for the $q\bar{q}$ template in cases after the $BDT_{q\bar{q}}$ cuts and signal template in all
 1287 cases, where both are expected to be very close to zero and are instead fixed to the
 1288 expected MC yield. Additionally, since the C_0 and C_1 decays are well known and
 1289 measured, we make use of this fact in the form of a ratio N_{C_1}/N_{C_0} , which is fixed to
 1290 the MC value in case of the MC fit and constrained to the measured value in case
 1291 of fits to real data. The ratio is implemented based on the decay channels shown in
 1292 Table 7.1 and is defined as

$$r_1 = \frac{\left(\sum_j N_{1j} \times \rho_{1j}\right)}{N_{00} \times \rho_{00}}, \quad (7.5)$$

1293 where j runs over all channels in the category C_1 and where ρ_{ij} is the branching
 1294 ratio correction factor for the specific channel N_{ij} , which incorporates information
 1295 from world measurements. It is defined as

$$\rho = \frac{\mathcal{B}^{PDG}}{\mathcal{B}^{GEN}}, \quad (7.6)$$

1296 where \mathcal{B}^{PDG} is the measured branching ratio and \mathcal{B}^{GEN} is the branching ratio value
 1297 used in MC generation. The branching ratio correction factor has been implemented
 1298 due to differences between measured and MC branching ratio values. Each branch-
 1299 ing ratio measurement serves as a constraint used in the fit. All branching ratio
 1300 constraints in the control fit are shown in Table 7.2. The measured values are
 1301 cited only for the B^0 decay mode, where isospin symmetry has been assumed. The
 1302 corresponding B^+ branching ratios can be calculated as

$$\mathcal{B}(B^+) = \mathcal{B}(B^0) \times \tau_{B^+/B^0}, \quad (7.7)$$

1303 where τ_{B^+/B^0} is the ratio of B -meson decay times, which is measured to be [8]

$$\tau_{B^+/B^0} = 1.076 \pm 0.004. \quad (7.8)$$

Category	Channel	B Decay Mode	D Decay Mode	N_{MC}
C_0	N_{00}	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	1184 ± 34
C_1	N_{10}	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	1458 ± 38
	N_{11}	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	186 ± 16

Table 7.1.: Well defined decay channels used for constraining the control fits.

1304 In case of MC fits, the fitted sample is also generated with MC, so $\mathcal{B}_i^{PDG} = \mathcal{B}_i^{GEN}$
 1305 and Eq. (7.5) simplifies to a simple MC yield ratio. On fits to real data, expected MC
 1306 yields and branching ratio measurements are implemented as independent Gaussian
 1307 constraints in order to properly account for correlations in Eq. (7.5).

ID	Decay	\mathcal{B}_{GEN}	\mathcal{B}_{PDG}	ρ	Ref.
0	$B^0 \rightarrow D^- \ell^+ \nu$	2.13×10^{-2}	$(2.13 \pm 0.09) \times 10^{-2}$	1.00 ± 0.04	[8]
1	$B^0 \rightarrow D^{*-} \ell^+ \nu$	5.33×10^{-2}	$(4.88 \pm 0.11) \times 10^{-2}$	0.92 ± 0.02	
2	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	2.31×10^{-2}	$(2.29 \pm 0.10) \times 10^{-2}$	0.99 ± 0.04	[calc.]
3	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	5.79×10^{-2}	$(5.25 \pm 0.12) \times 10^{-2}$	0.91 ± 0.02	
4	$D^0 \rightarrow K^- K^+$	3.90×10^{-3}	$(3.97 \pm 0.07) \times 10^{-3}$	1.02 ± 0.02	[24]

Table 7.2.: MC and measured values of branching ratios along with the calculated correction factors used for constraining the control fit.

Smearing and Offset Parameters

With simulated data, we are able to perform detailed studies prior to looking at the measured data. However, simulated data often does not describe real data perfectly. Out of variables ΔE and M_{BC} , ΔE is especially prone to a lack of precision in energy measurements. This can introduce either overestimation of resolution on MC as well as a possible shift in the measured energy in either direction. Due to this fact, we introduce a modification of the ΔE variable by applying a smearing factor and an offset. These are applied by simple transformations of

$$f_{\text{offset}} : x \mapsto x + a, \quad (7.9)$$

$$f_{\text{smearing}} : x \mapsto \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (7.10)$$

where x goes over all entries in the ΔE distribution, a is the energy offset and f_{smearing} corresponds to the normal distribution with mean μ and standard deviation σ . In the case of the M_{BC} variable the mentioned effects are not as prominent, so the smearing and offset for the latter variable are omitted.

The following parameter phase space is scanned in order to determine the best parameter values

- smearing factor in range $[0.0, 0.08]$ GeV in steps of 8×10^{-3} ,
- offset in range $[0.0, 0.003]$ GeV in steps of 1.5×10^{-4} ,

where for each parameter pair the likelihood ratio test is performed to estimate the goodness of the fit. Figure 7.1 shows the contour plot of the likelihood ratio λ , as defined in Eq. (7.3), for 2 d.o.f., for MC (left) and data (right). The scan over MC serves the purpose of a consistency check, where we expect the best fit to occur in the phase space where neither smearing nor offset are applied. In the case of data, we see that we obtain a better fit by introducing some level of smearing and offset. In both cases, the two parameters have shown no sign of significant correlation, so we treat them independently. The likelihood ratio test allows us to estimate the parameter values in the 1σ confidence interval, where we obtain the optimal parameter set

- Smearing: 40_{-17}^{+15} MeV,
- Offset: $6_{-6}^{+4.6}$ MeV.

1329 We apply this transformation to our MC samples in all cases when fitting real data.

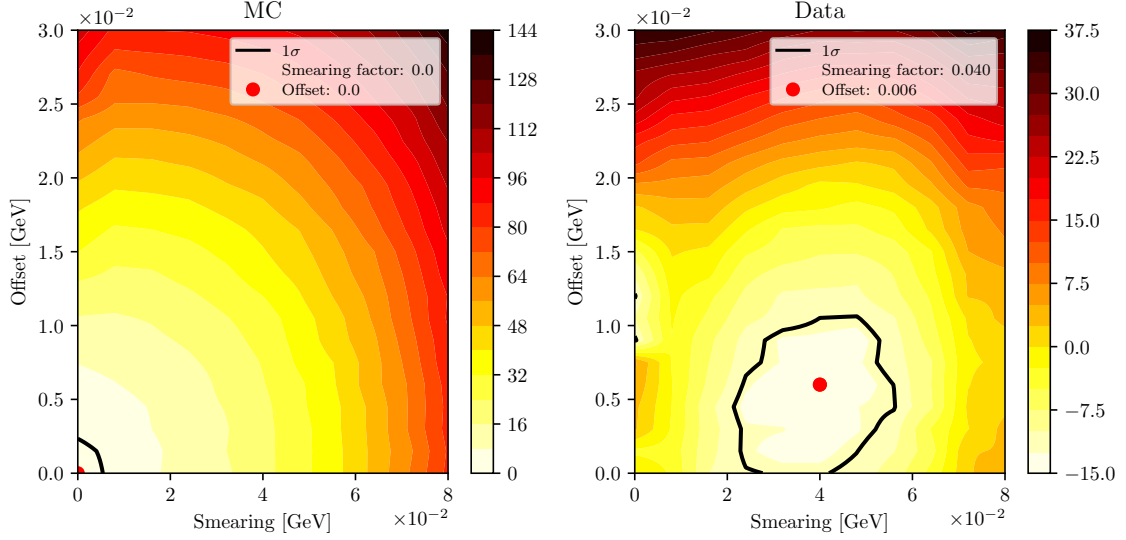


Figure 7.1.: Likelihood ratio test of an additional smearing and offset parameter to MC (left) and data (right).

1330 7.1.2. Signal Fit

1331 The motivation for the choice of signal fit templates comes from Figure 6.12. The
1332 following histogram templates were defined

- 1333 • signal template,
- 1334 • $q\bar{q}$ template,
- 1335 • a series of well defined templates from $B\bar{B}$ background:
 - 1336 – C_0 : $B^+ \rightarrow \bar{D}^0 \ell^+ \nu$, $D^0 \rightarrow K^- K^+$ (control decay),
 - 1337 – C_1 : $B \rightarrow \bar{D}^* \ell^+ \nu$, $D^0 \rightarrow K^- K^+$,
 - 1338 – C_2 : $B \rightarrow \bar{D}^{(*)} \ell^+ \nu$, $D^0 \rightarrow K^- \pi^+$,
 - 1339 – C_3 : $B \rightarrow \bar{D}^{(*)} \ell^+ \nu$, $D^0 \rightarrow K^- K^+ \pi^0$, $K^- \pi^+ \pi^0$,
 - 1340 – C_4 : $B \rightarrow \bar{D}^{(*)} \ell^+ \nu$, $D^0 \rightarrow K^- \ell^+ \nu$,
 - 1341 – C_5 : $B^0 \rightarrow D^{(*)-} \ell^+ \nu$, $D^+ \rightarrow K^- K^+ \pi^+$, $K^- \pi^+ \pi^+$,
 - 1342 – C_6 : other $B \rightarrow \bar{D}^{(*)} \ell^+ \nu$ decays,
- 1343 • remaining $B\bar{B}$ background template.

1344 As mentioned in chapter 6, the majority of the background comes from $B\bar{B}$ events.
1345 Various processes (C_0 to C_6) contribute to this background which are well known and
1346 measured, so we make use of these measurements by fixing their yields in MC fits and
1347 appropriately constraining them in real data fits. The remaining $B\bar{B}$ background is
1348 merged into a single template. In this case, the shape of all templates is fixed as well,
1349 while the yields are floated for all templates except for the constrained background

1350 templates. The yield constraints are based on the channels shown in Table 7.3 and
 1351 defined for each template category as

$$Y_i = \eta_{\text{norm.}} \times \frac{\left(\sum_j N_{ij} \times \rho_{ij}\right)}{\rho_{00}}, \quad (7.11)$$

1352 where j runs over all channels in the category C_i and where ρ_{ij} is defined in Eq.
 1353 (7.6). The first factor, $\eta_{\text{norm.}}$, serves as a normalization factor in order to scale the
 1354 number of generated $B\bar{B}$ events to the number of $B\bar{B}$ events in measured data. We
 1355 define it as

$$\eta_{\text{norm.}} = \frac{N_{\text{control}}^D}{N_{\text{control}}^{MC}}, \quad (7.12)$$

1356 where N_{control}^D and N_{control}^{MC} are control yields in the control fit for data and MC,
 1357 respectively.

1358 In addition to branching ratio constraints in Table 7.2, further constraints are
 1359 defined in Table 7.4 due to more decay channels. In case of the category C_6 , we
 1360 have no firm handle on the D meson decay, therefore no correction for this branching
 1361 ratio can be introduced, so we set a correction factor of 1 with a 100% error for the D
 1362 meson decay branching ratio. As most of the correction factors used for constraints
 1363 have deviations (including the errors) from nominal values well below 100% this
 1364 value is very conservative.

Category	Channel	B Decay Mode	D Decay Mode	Expected MC Yield
C_0	N_{00}	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	44 ± 7
C_1	N_{10}	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	53 ± 7
	N_{11}	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	6 ± 2
C_2	N_{20}	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+$	23 ± 5
	N_{21}	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+$	41 ± 6
	N_{22}	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+$	6 ± 2
C_3	N_{30}	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- K^+ \pi^0$	102 ± 10
	N_{31}	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+ \pi^0$	210 ± 15
	N_{32}	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- K^+ \pi^0$	135 ± 12
	N_{33}	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+ \pi^0$	266 ± 16
	N_{34}	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- K^+ \pi^0$	19 ± 4
	N_{35}	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+ \pi^0$	35 ± 6
C_4	N_{40}	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- e^+ \nu$	47 ± 7
	N_{41}	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- \mu^+ \nu$	7 ± 3
	N_{42}	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- e^+ \nu$	98 ± 10
	N_{43}	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- \mu^+ \nu$	10 ± 3
	N_{44}	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- e^+ \nu$	14 ± 4
	N_{45}	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- \mu^+ \nu$	3 ± 2
C_5	N_{50}	$B^0 \rightarrow D^- \ell^+ \nu$	$D^+ \rightarrow K^- K^+ \pi^+$	102 ± 10
	N_{51}	$B^0 \rightarrow D^- \ell^+ \nu$	$D^+ \rightarrow K^- \pi^+ \pi^+$	63 ± 8
	N_{52}	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^+ \rightarrow K^- K^+ \pi^+$	31 ± 6
	N_{53}	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^+ \rightarrow K^- \pi^+ \pi^+$	21 ± 5
C_6	N_{60}	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	Other D^0 and D^+ decays	69 ± 8
	N_{61}	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$		94 ± 10
	N_{62}	$B^0 \rightarrow D^- \ell^+ \nu$		63 ± 8
	N_{63}	$B^0 \rightarrow D^{*-} \ell^+ \nu$		35 ± 6

Table 7.3.: Well defined decay channels used for constraining the signal fits.

ID	Decay	\mathcal{B}_{GEN}	\mathcal{B}_{PDG}	ρ	Ref.
5	$D^0 \rightarrow K^- \pi^+$	3.82×10^{-2}	$(3.89 \pm 0.04) \times 10^{-2}$	1.02 ± 0.01	[24]
6	$D^0 \rightarrow K^- K^+ \pi^0$	2.36×10^{-3}	$(3.37 \pm 0.15) \times 10^{-3}$	1.43 ± 0.06	
7	$D^0 \rightarrow K^- \pi^+ \pi^0$	13.08×10^{-2}	$(14.2 \pm 0.5) \times 10^{-2}$	1.09 ± 0.04	
8	$D^0 \rightarrow K^- e^+ \nu$	3.41×10^{-2}	$(3.53 \pm 0.028) \times 10^{-2}$	1.04 ± 0.01	
9	$D^0 \rightarrow K^- \mu^+ \nu$	3.41×10^{-2}	$(3.31 \pm 0.13) \times 10^{-2}$	0.97 ± 0.04	
10	$D^+ \rightarrow K^- K^+ \pi^+$	9.06×10^{-3}	$(9.51 \pm 0.34) \times 10^{-3}$	1.05 ± 0.04	
11	$D^+ \rightarrow K^- \pi^+ \pi^+$	9.51×10^{-2}	$(8.98 \pm 0.28) \times 10^{-2}$	0.94 ± 0.03	

Table 7.4.: Additional MC and measured values of D meson branching ratios along with the calculated correction factors used for constraining the signal fit.

7.2. Adaptive Binning Algorithm

The fit templates contain areas of low statistics, which are populated with bins with zero content. This is a direct consequence of having a finite MC sample and represent a liability in ML fits. Due to low statistics in the edge regions, the locations of these empty bins can vary for the templates and the fitted sample. A problem occurs if all templates have an empty bin where the fitted sample does not. In the scope of ML fits, this effectively means that there are entries in bins where the probability of having them is 0. We will call such bins *problematic* because in these cases the fit does not converge.

The ideal solution for this problem would be to increase the MC statistics. Since this is not an option, we pursue other solutions, such as decreasing the number of bins. While this solves the problem, the drawback of it is a decrease in the template resolution in densely populated regions, where good resolution is most needed. The optimal solution here seems to be a choice of variable bins, with fine binning in the densely populated regions and larger bins in the regions with low statistics.

We have devised an algorithm, which compares the templates and the fitted sample, and defines a variable binning so that there are no more problematic bins in the end. Figure 7.2 shows an example of how the procedure works. The algorithm takes an argument for the initial number of uniform bins in each dimension and does the following

1. define uniform binning in both dimensions with the provided argument,
2. create a 2D histogram from MC templates with expected yields,
3. define an *optimal* region, where most of the 2D integral is contained and where all bins have a non-zero content (this region does not change throughout the process),
4. compare the histograms for the expected and the fitted sample, find the problematic bins,
5. loop until all problematic bins disappear
 - a) find the problematic bin, which is nearest to the maximum bin,
 - b) change the binning from N to $N - 1$ from that bin and in the direction away from the maximum bin.

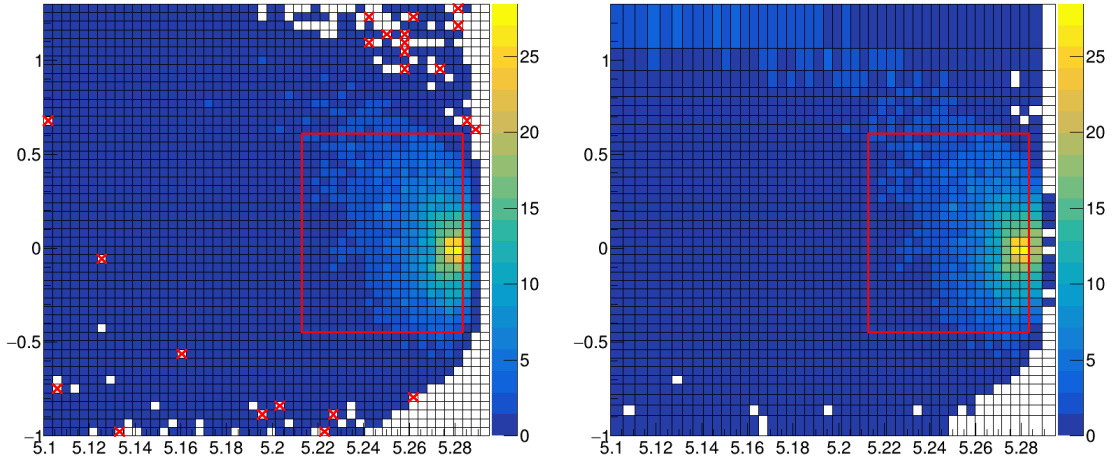


Figure 7.2.: Steps taken in the adaptive binning algorithm. Left image shows the initial 2D histogram with the defined optimal region and the problematic bins, the right image shows the final binning with the unchanged optimal region, while the problematic bins are gone due to the new binning choice.

1396 An additional problem occurs in the plotting of the fitted templates and the sample
 1397 with variable binning. It would seem that RooFit does not take the bin widths into
 1398 account when plotting, while everything works as expected for the fit itself. This
 1399 was bypassed by extracting the fitted yields and applying them to templates and
 1400 samples with uniform binning, which were then used for drawing.

1401 7.3. Toy MC Experiments

1402 For the chosen final selection and fit procedure, toy MC pseudo-experiments were
 1403 performed in order to confirm the behavior of the fit setup. The fit behavior is also
 1404 checked as a function of the signal yield in the form of a linearity test. A detailed
 1405 description of toy MC experiments is written in this section.

1406 With toy MC experiments in this section, we study the yields, errors and the pulls
 1407 of the signal fit by generating our own pseudo-datasets, according to the MC. This
 1408 significantly reduces the time we would need to produce the datasets in the standard
 1409 way, while still reliably describing the underlying physics behind the pseudo-dataset.
 1410 All available MC was used for pseudo-dataset generation as well as creating tem-
 1411 plates.

1412 7.3.1. Pseudo-Experiment: Expected Signal Yield

1413 We constructed 3×10^3 pseudo-datasets, where each dataset was generated with the
 1414 expected amount of each template category, distributed according to the Poisson
 1415 distribution. All fits were performed with the optimal initial uniform binning of
 1416 19×19 bins in ΔE and M_{BC} .

Figure 7.3 shows distributions of the fit yields, errors and the pull distribution of all pseudo-fits. The fits seem to be under control, although there is a slight bias present in the negative direction, which can also be seen in the pull distribution plot. The latter follows a normal distribution with a mean of (-0.11 ± 0.02) and a standard deviation of (1.01 ± 0.01) . The mean (\bar{X}) and the standard deviation (S) were calculated in the usual way, while their errors $\sigma_{\bar{X}}$ and σ_S were calculated as [25]

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{N}}, \quad \sigma_S = \frac{S}{\sqrt{2(N-1)}}, \quad (7.13)$$

where N is the number of performed measurements. It seems that there is a small bias present in the pseudo-fits. A decision has been made to put this difference as one of the contributions to the overall systematic uncertainty.

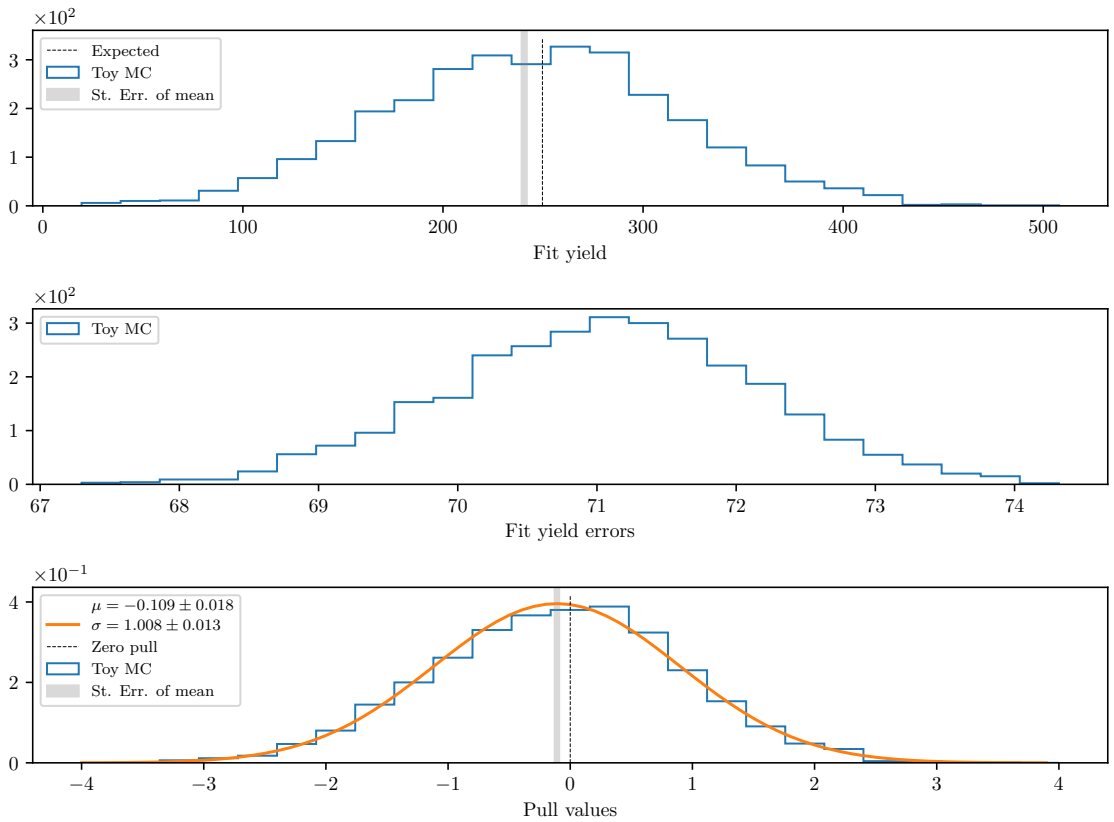


Figure 7.3.: Toy MC fits of pseudo-data showing the fit yield (top), fit errors (center) and the pull distribution of the fits (bottom).

7.3.2. Pseudo-Experiment: Linearity Test

Linearity test is used for determining the sensitivity of the fit to the amount of signal in the fitted sample. Since this is the first measurement of this decay channel, MC modeling is not reliable and could be very different from reality, so we need to perform this test in order to determine our sensitivity to smaller, as well as larger amounts of expected signal.

The pseudo-datasets are generated in the same way as in the previous subsection, with the exception of signal, which is generated in various amounts. 50 steps from

1434 [0.1, 10] in the logarithmic scale are taken for fractions of signal amount and for
 1435 each fraction we generate 500 pseudo-datasets according to Poisson statistics.

1436 Figure 7.4 shows the difference between mean fit yield and the expected yield,
 1437 mean pull and the mean significance at each signal fraction value. The expected
 1438 MC result lies at the fraction value $10^0 = 1$. The plots show a small bias with
 1439 respect to the expected signal yield, while the pulls seem to be described by the
 1440 normal distributions throughout the fraction range with a slight offset, which will
 1441 be incorporated as a contribution to the overall systematic uncertainty. At expected
 1442 value, we are at about 3.6σ significance.

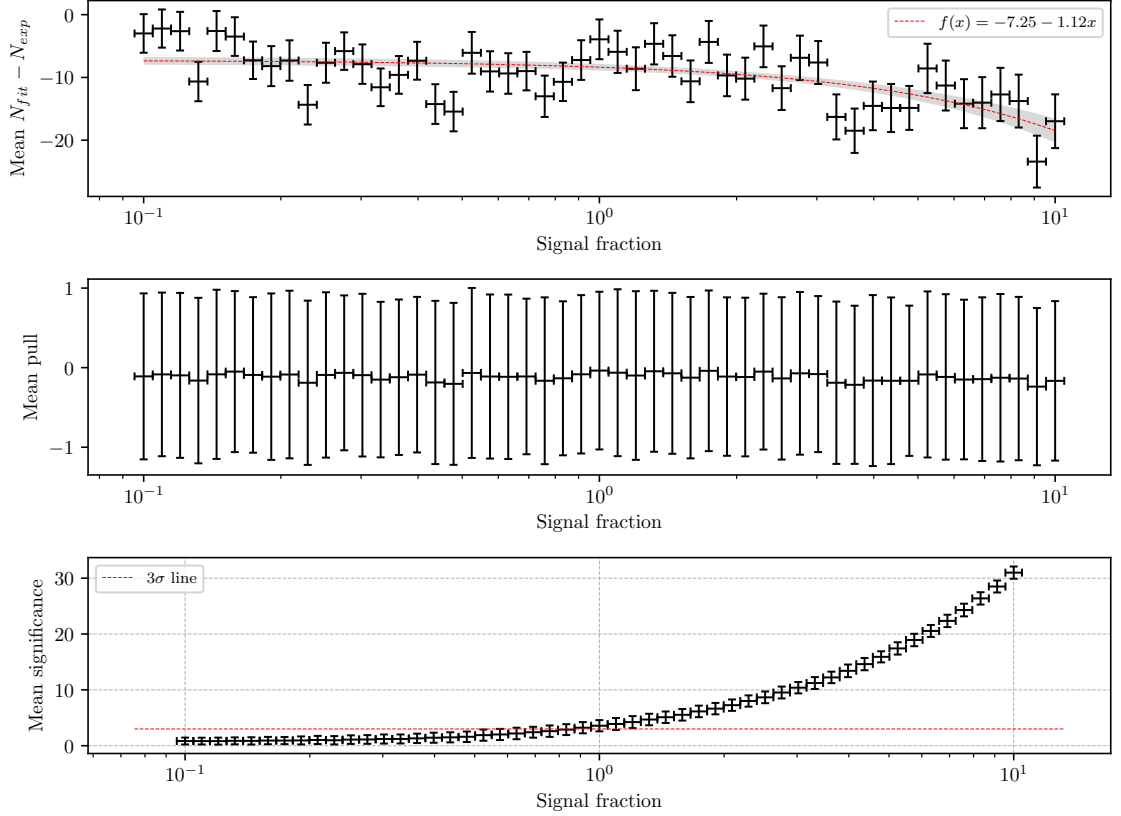


Figure 7.4.: Mean fit yield and expected yield difference (top), mean pull (center) and mean significance (bottom) as a function of signal fraction.

Chapter 8.

Results

In this section, we present the first results of signal and control fits on MC as well as data, along with the control decay branching ratio measurement. We also show results of the signal fit on MC and data.

8.1. Signal MC Fit Results

With the signal fit setup described in section 7.1.2, we proceed to fit the 10 streams of MC. To compare both methods of $B\bar{B}$ bar suppression, two different samples were prepared and used in the fit. Since the choice of initial uniform binning is not obvious, we perform fits to all streams of MC for each initial binning choice in the range $N \times N$, $N \in [4,30]$. Figure 8.1 shows the expected yield differences, pulls and fit significances for both final samples for each binning case. The difference between fitted and expected signal yield should be equal to 0 to ensure no bias is present in the fit, while the average pull distribution for each bin case should have a central value at 0, with a width of 1. It can be seen from the plots that this is not exactly the case, as a slight offset is present in the other direction as in the case of the pseudo-experiments. This difference is also accounted for as a contribution to the overall systematic uncertainty. The pull distribution seems to be closer to the normal distribution for the case of the uBDT classifier, as well as the fit significance appears to be larger by approximately 1σ , so we conclude that the uBDT classifier fit setup outperforms the standard BDT fit setup, which determines our choice of the final selection. The binning in ΔE and M_{BC} is chosen at the plateau of the significance, where no significant bias is present and is somewhere in the region of 20 bins in each dimension.

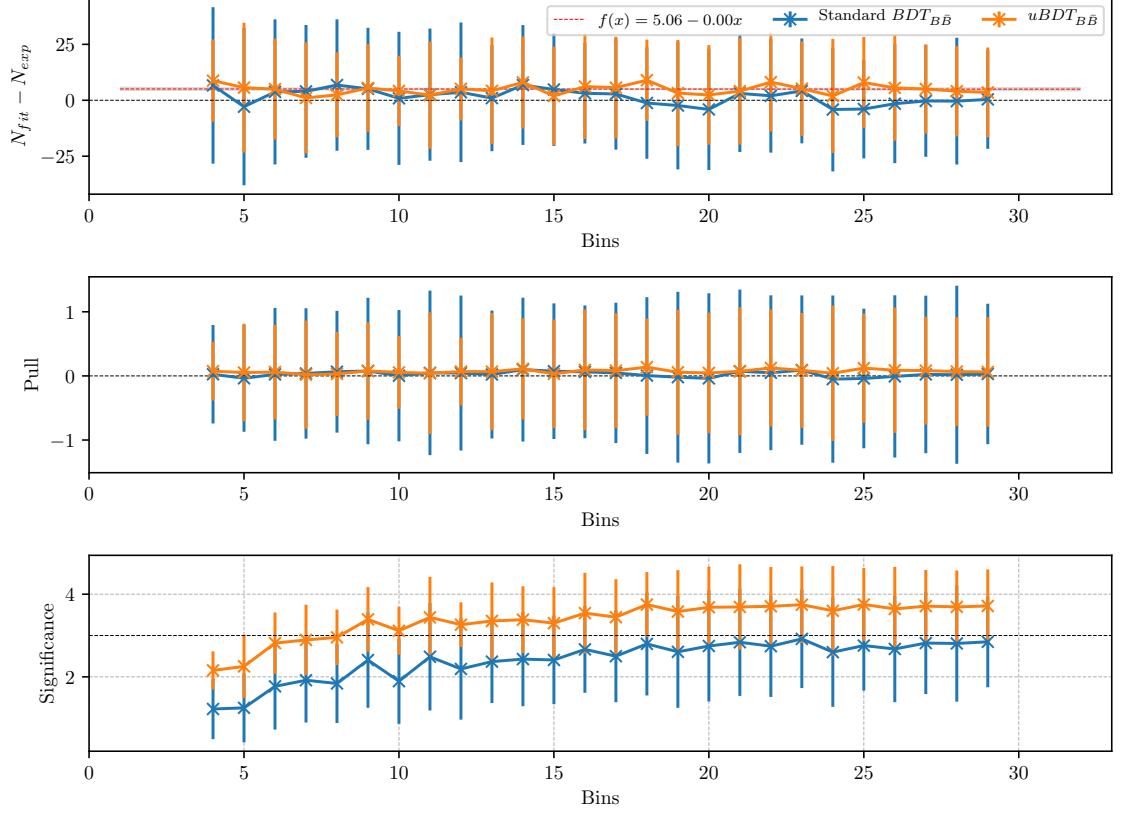


Figure 8.1.: Fitted yield and expected yield difference (top), pulls (center) and fit significance (bottom) as a function of binning in ΔE and M_{BC} for the final sample, optimized with the standard BDT classifier (blue) and the uBDT classifier (orange).

For the chosen binning of 19×19 in ΔE and M_{BC} we perform the 10 stream MC fits, where an average stream fit is shown in Figure 8.2, while all fit results are shown in Figure 8.3. All stream fit results were fitted with a 0th-degree polynomial. The global result seems to describe the expected value in a precise manner, with the bias much smaller than the average statistical error. The normalized χ^2 value with $10 - 1 = 9$ degrees of freedom of the global fit is $\chi^2_9 = 1.23$, while the average significance of the fits is around 3.5σ .

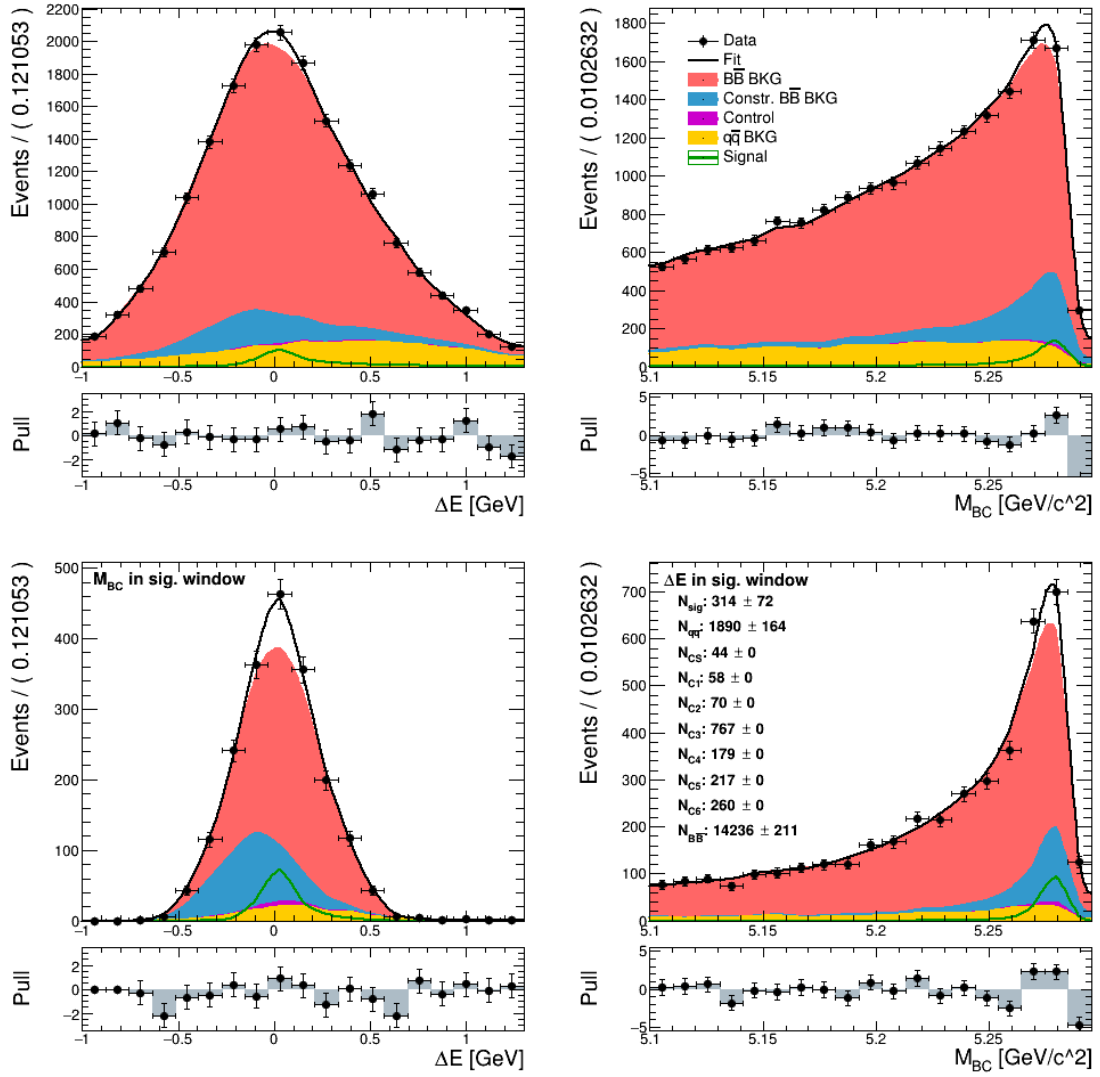


Figure 8.2.: An example fit to one stream of MC. Left column shows the M_{BC} and the right column shows the ΔE distribution of the full fitted sample in the full fit region (top) and the signal region (bottom).

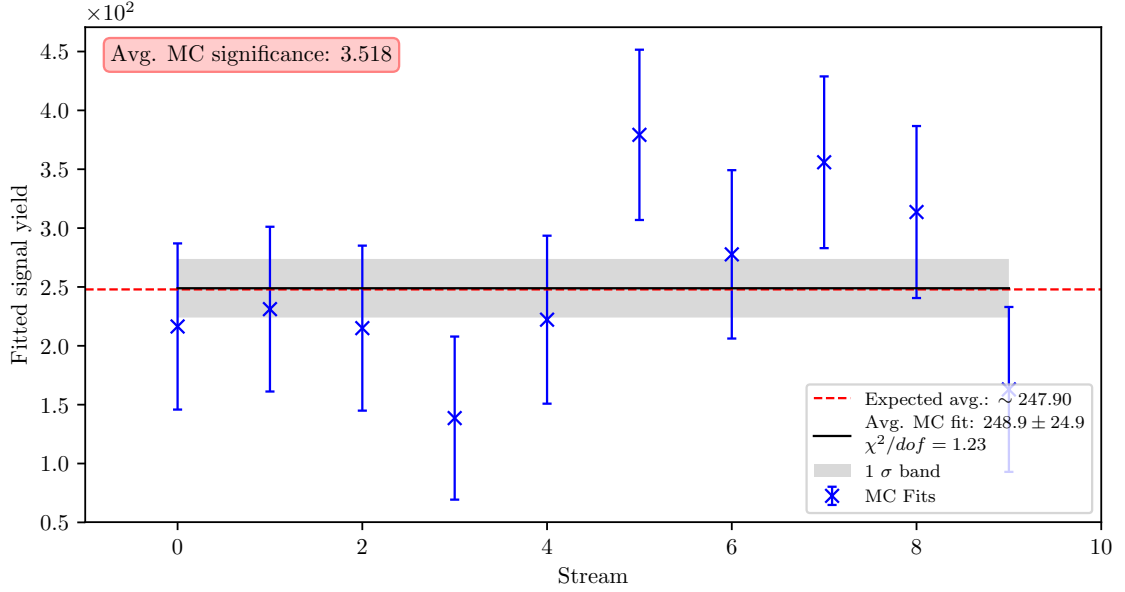


Figure 8.3.: Fits to all 10 streams of MC and the global fit with a zero-th degree polynomial. The red line shows the mean value of the global fit and the gray band shows the 1σ confidence interval.

8.2. Control Fit Result

With the control fit setup described in section 7.1.1, we proceed to fit the control sample after the final selection for 10 streams of MC and 1 stream of data. An average MC stream fit is shown in Figure 8.4 for MC and Figure 8.5 for data, while all fit results are shown in Figure 8.6, where all streams of MC are fitted with a 0th degree polynomial. The control fit results for split and joined lepton modes are shown in Table 8.1.

	N^{MC}	N^{data}
$\ell = e \text{ or } \mu$	1182 ± 11	1187 ± 44
$\ell = e$	591 ± 8	583 ± 28
$\ell = \mu$	592 ± 7	613 ± 30

Table 8.1.: Control sample fit results for MC and data for various lepton final state modes.

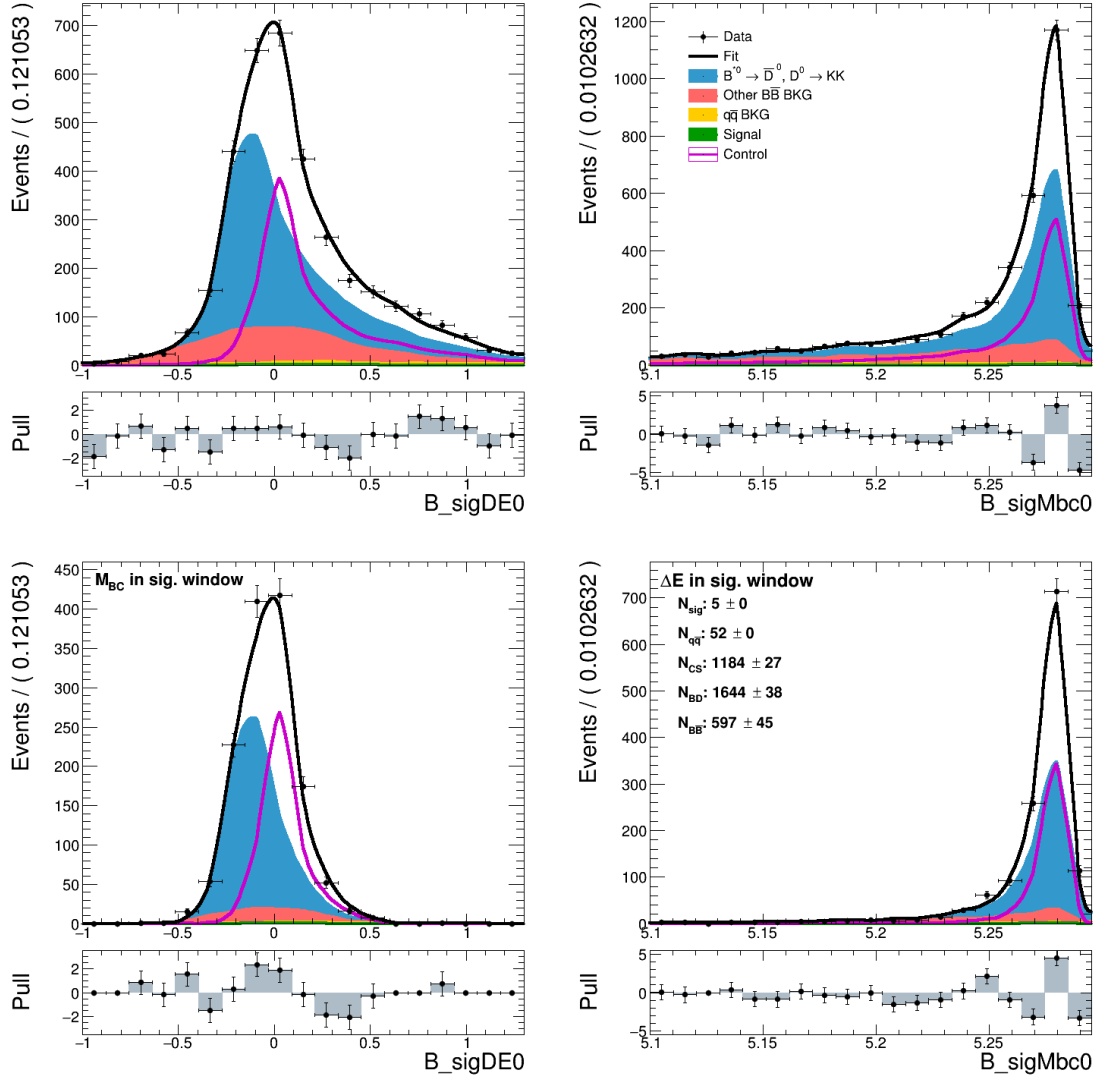


Figure 8.4.: Control fit result on one stream of MC. Left column shows the M_{BC} and the right column shows the ΔE distribution in the full fit window (top) and the signal window (bottom).

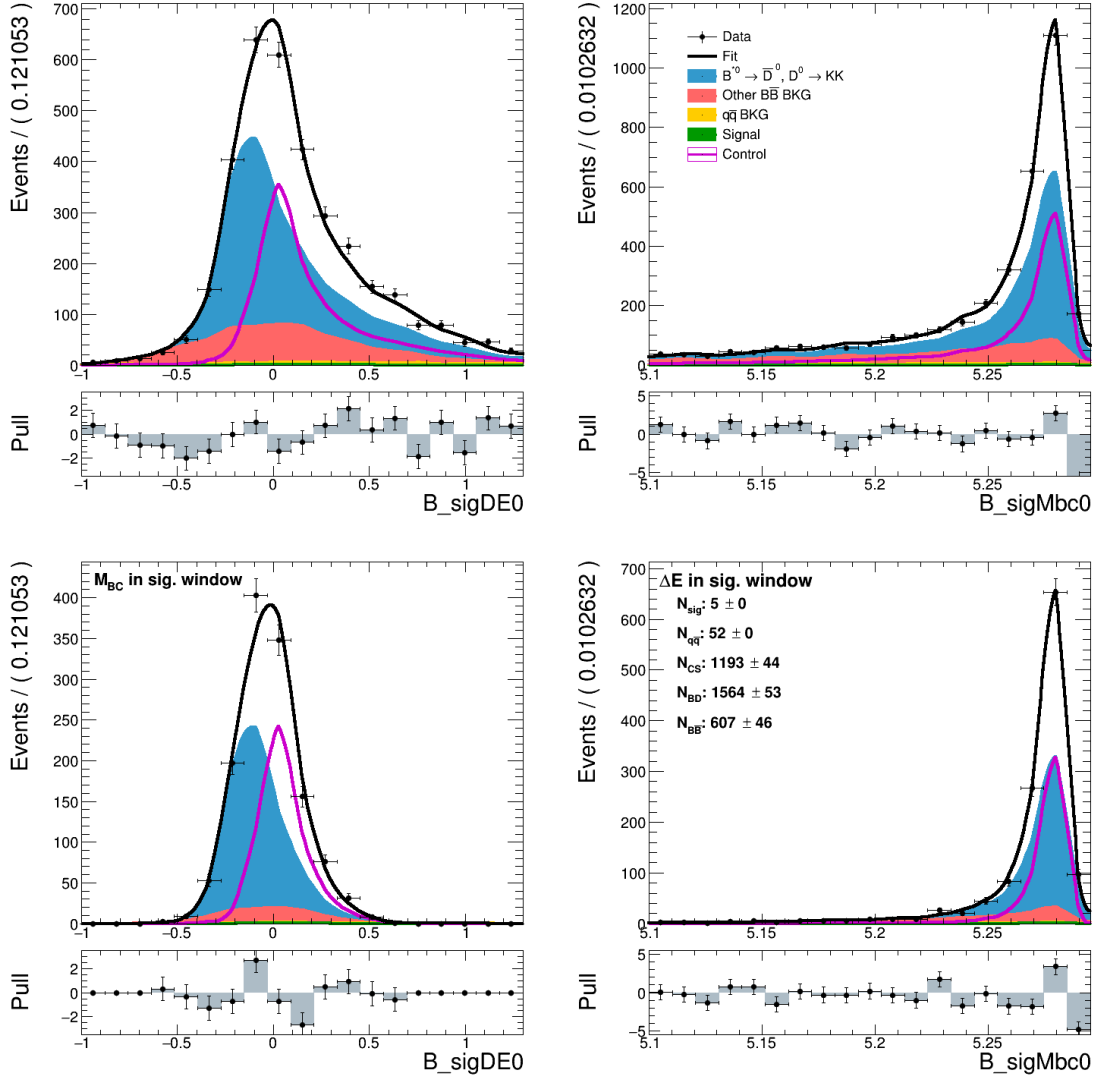


Figure 8.5.: Control fit result on real data. Left column shows the M_{BC} and the right column shows the ΔE distribution in the full fit window (top) and the signal window (bottom).

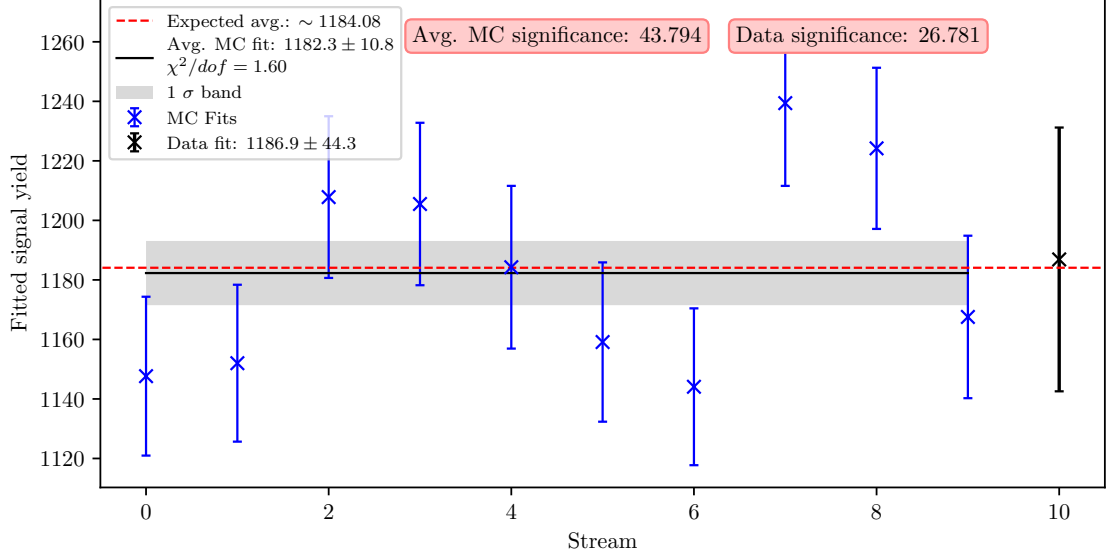


Figure 8.6.: Control fit to data and to all 10 streams of MC. The red line shows the mean value of the global MC fit with a 0th degree polynomial. The gray band shows the 1σ confidence interval around the global MC fit.

1481 8.2.1. Branching Ratio Measurement for Control Decay

1482 After acquiring the fit results on MC and data, we are able to determine the branch-
 1483 ing ratio of the control decay, which is defined as

$$\mathcal{B}_{\text{control}}^{\text{MC}} = \frac{N_{\text{control}}^{\text{MC}} \times \epsilon_{\text{MC}}}{2N_{B\bar{B}}^{\text{MC}}}, \quad (8.1)$$

$$\mathcal{B}_{\text{control}} = \frac{N_{\text{control}} \times \epsilon_{\text{MC}} \times \rho_{\text{PID}}}{2N_{B\bar{B}}}, \quad (8.2)$$

1484 where $N_{\text{control}}^{\text{MC}}$ and N_{control} are yields of the control fit on MC and data, ϵ_{MC} is the
 1485 MC efficiency of the control sample, ρ_{PID} the PID correction factor, and $N_{B\bar{B}}^{\text{MC}}$ and
 1486 $N_{B\bar{B}}$ are the numbers of $B\bar{B}$ meson pairs on MC and data, respectively. The factor
 1487 of 2 in the denominator comes from the fact that there are 2 B mesons in each $B\bar{B}$
 1488 meson pair ($\times 1/2$), where only about 50% of the B meson pairs are charged B^+B^-
 1489 meson pairs ($\times 2$), and from the fact that we are interested in the branching ratio
 1490 to the lepton final state of either e or μ , and not their sum ($\times 1/2$).

The control sample efficiency was determined on a separate signal MC sample of the control decay, where we generated 5×10^6 $B^+\bar{B}^-$ pairs, with one B always decaying via $B^+ \rightarrow \bar{D}^0 \ell^+ \nu$, $D^0 \rightarrow K^+ K^-$. After applying the final selection, the full and split efficiencies with regard to the lepton final state were determined to be

$$\begin{aligned} \epsilon_{\text{MC}} &= (8.61 \pm 0.04) \times 10^{-3}, \\ \epsilon_{\text{MC}}^e &= (4.25 \pm 0.03) \times 10^{-3}, \\ \epsilon_{\text{MC}}^\mu &= (4.35 \pm 0.03) \times 10^{-3}, \end{aligned}$$

where the efficiency error was estimated based on a formula from [26].

$$\epsilon_{MC} = \frac{1}{N} \sqrt{n(1 - \frac{n}{N})},$$

1491 where n is a subset of the full set N .

The PID correction factor is obtained by taking into account the known PID efficiency differences between data and MC. It is described in detail in section 9.1.1 and is determined to be

$$\rho_{PID} = 0.99 \pm 0.02$$

1492 for the e and μ mode, as well as both of them together.

The number of $B\bar{B}$ meson pairs can be counted on MC and has been measured for data by the collaboration. The values are

$$\begin{aligned} N_{B\bar{B}}^{MC} &= 765.98 \times 10^6, \\ N_{B\bar{B}} &= (771.58 \pm 10.56) \times 10^6. \end{aligned}$$

1493 Finally, we can determine the branching ratios based on the calculations in Eq.
1494 (8.2). The obtained values are shown in Table 8.2 and graphically shown in Figure
1495 8.7, along with the MC generated value and the current PDG world average. Both
1496 MC and measured determinations of the control decay branching ratio are in agree-
1497 ment with the expected and the world average values. One should note that the
1498 black error bars correspond to statistical uncertainty. Of all the systematic uncer-
1499 tainties, only the PID systematics are included in this results. Other contributions of
1500 systematics are not included since this measurement is not the goal of our analysis.

	$\mathcal{B}_{GEN} [\times 10^{-5}]$	$\mathcal{B}_{PDG} [\times 10^{-5}]$	$\mathcal{B}^{MC} [\times 10^{-5}]$	$\mathcal{B}^{\text{data}} [\times 10^{-5}]$
$\ell = e \text{ or } \mu$	9.01	9.10 ± 0.42	8.97 ± 0.09	9.03 ± 0.40
$\ell = e$			9.07 ± 0.14	9.02 ± 0.49
$\ell = \mu$			8.88 ± 0.12	9.18 ± 0.50

Table 8.2.: Control sample fit results for MC and data for various lepton final state modes.

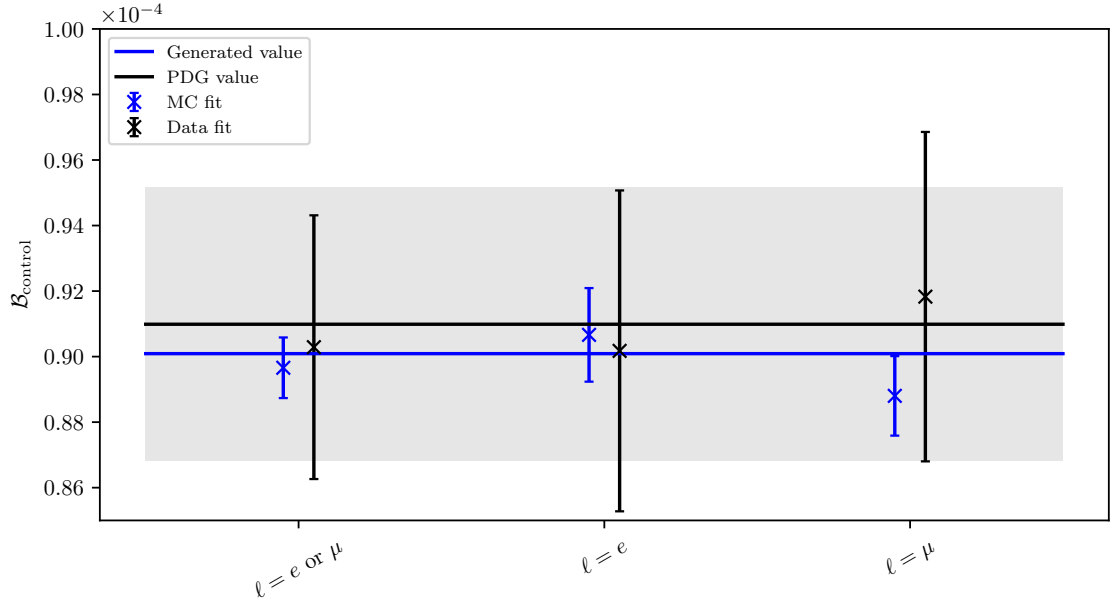


Figure 8.7.: Various branching ratio determinations for the control decay of our analysis.

8.3. Signal Fit Results to Data

8.3.1. Signal Yield

Finally, after making sure that the analysis procedure makes sense on the control MC, control data and signal MC sample, we can continue to perform the signal extraction process on the full Belle $\Upsilon(4S)$ data sample. Figure 8.8 shows the fit result in projections of ΔE and M_{BC} for both fit regions. The extracted signal yield on data as well as the yields of the remaining contributions are shown in Table 8.3. Values of all the constraints are shown in Table 8.4.

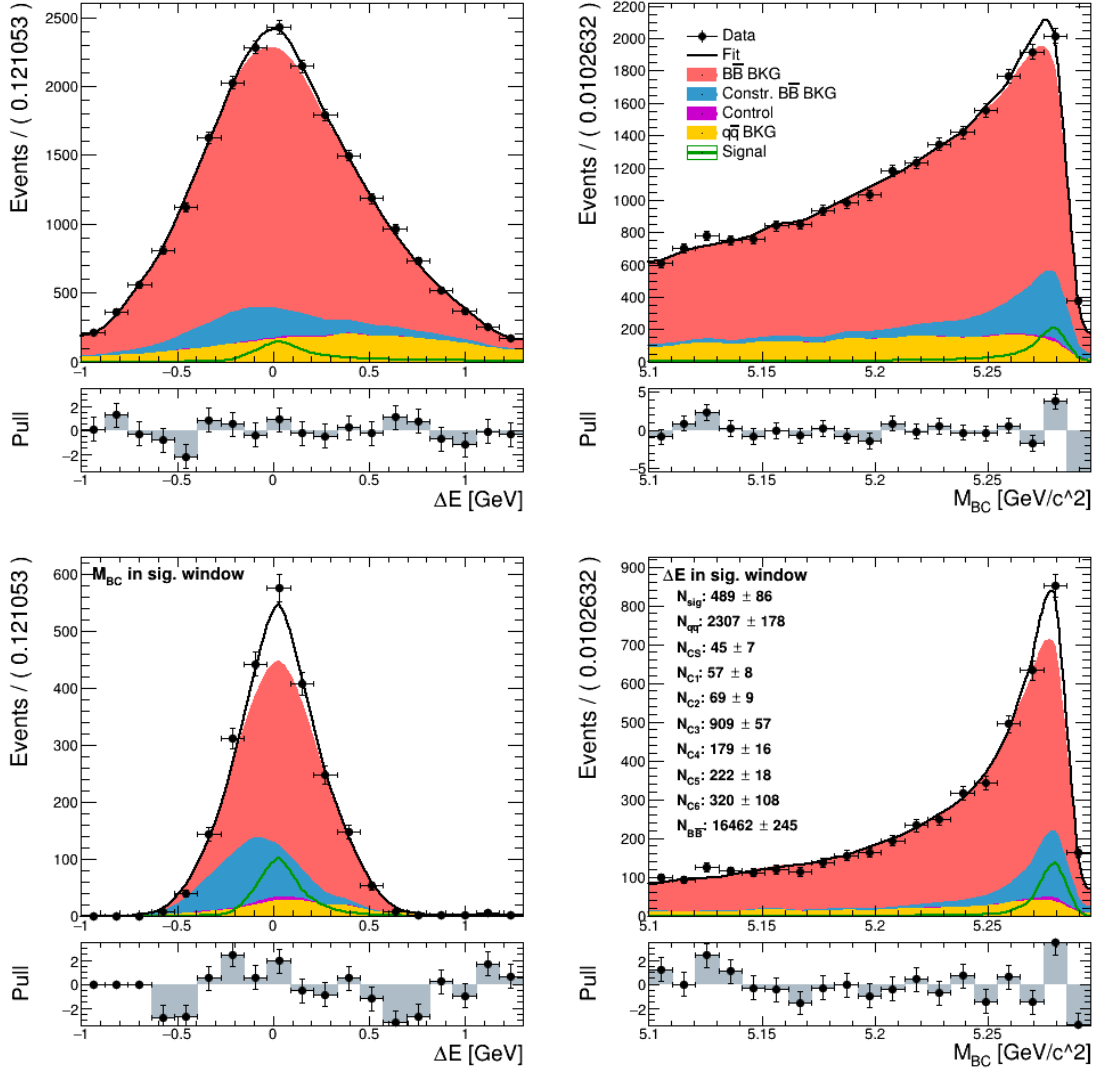


Figure 8.8.: Signal fit result on real data. Left column shows the M_{BC} and the right column shows the ΔE distribution in the full fit window (top) and the signal window (bottom).

Category	Fit Yield
Signal	489 ± 86
$q\bar{q}$ background	2307 ± 178
C_0	45 ± 7
C_1	57 ± 8
C_2	69 ± 9
C_3	909 ± 57
C_4	179 ± 16
C_5	222 ± 18
C_6	320 ± 108
Other $B\bar{B}$ background	16462 ± 245

Table 8.3.: Yields of all signal fit contributions on data.

Constraint	Value	Constraint	Value
\mathcal{B}_0	0.049 ± 0.001	N_{30}	103 ± 10
\mathcal{B}_1	1.076 ± 0.004	N_{31}	212 ± 14
\mathcal{B}_2	0.021 ± 0.001	N_{32}	136 ± 12
\mathcal{B}_3	1.014 ± 0.018	N_{33}	268 ± 16
\mathcal{B}_4	1.018 ± 0.010	N_{34}	20 ± 4
\mathcal{B}_5	1.439 ± 0.063	N_{35}	35 ± 6
\mathcal{B}_6	1.092 ± 0.038	N_{40}	48 ± 7
\mathcal{B}_7	1.035 ± 0.008	N_{41}	7 ± 3
\mathcal{B}_8	0.971 ± 0.038	N_{42}	99 ± 10
\mathcal{B}_9	1.052 ± 0.038	N_{43}	10 ± 3
\mathcal{B}_{10}	0.945 ± 0.029	N_{44}	14 ± 4
\mathcal{B}_{11}	1.253 ± 0.443	N_{45}	3 ± 2
N_{control}^{MC}	1181 ± 11	N_{50}	103 ± 10
$N_{\text{control}}^{Data}$	1214 ± 43	N_{51}	64 ± 8
N_{00}	44 ± 7	N_{52}	31 ± 6
N_{10}	54 ± 7	N_{53}	22 ± 5
N_{11}	6 ± 2	N_{60}	69 ± 8
N_{20}	23 ± 5	N_{61}	94 ± 10
N_{21}	41 ± 6	N_{62}	63 ± 8
N_{22}	6 ± 2	N_{63}	35 ± 6

Table 8.4.: Mean values and standard deviations of constraints after the fit.

In this fit setup, we perform a random smearing of the ΔE variable according to the Poisson distribution. The result is similar to a convolution with a Gaussian function, but the application is different in the sense that it returns a slightly different result after each fit. We perform 500 data fits in order to obtain the mean values of the signal yield and fit error. The averaged result is

$$\bar{N}_{sig} = 489 \pm 85, \quad \text{fit significance} = 5.9\sigma,$$

1509 where the error represents the joint value of the statistical error and the systematic
1510 error of the Gaussian constraints used in the fit. This means that our initial fit is a
1511 good representation of the average fit result.

1512 It is possible to estimate the contribution of the systematic uncertainty to the fit
1513 error above, which is found to be

$$\sigma_{\text{stat}} = 81, \quad \sigma_{\text{sys}}^{GC} = 26, \quad \delta_{\text{sys}}^{GC} = 5.3\%, \quad (8.3)$$

1514 where GC stands for Gaussian constraint. Further details about systematic uncer-
1515 tainties can be found in chapter 9.

1516 8.3.2. Branching Ratio

Similarly as for the control decay, we are able to calculate the branching ratio of the signal decay via the formulas

$$\mathcal{B}_{\text{sig}}^{MC} = \frac{N_{\text{sig}}^{\text{MC}} \times \epsilon_{MC}}{2N_{B\bar{B}}^{MC}}, \quad (8.4)$$

$$\mathcal{B}_{\text{sig}} = \frac{N_{\text{sig}} \times \epsilon_{MC} \times \rho_{PID}}{2N_{B\bar{B}}}, \quad (8.5)$$

1517 where $N_{\text{sig}}^{\text{MC}}$ and N_{sig} are yields of the signal fit on MC and data, ϵ_{MC} is the MC
1518 efficiency of the signal sample, ρ_{PID} the PID correction factor, and $N_{B\bar{B}}^{MC}$ and $N_{B\bar{B}}$
1519 are the numbers of $B\bar{B}$ meson pairs on MC and data, respectively.

The signal efficiency was determined on the same signal MC sample as was used throughout the analysis. The full signal efficiency is determined to be

$$\epsilon_{MC} = (1.052 \pm 0.003) \times 10^{-2},$$

1520 where the efficiency error was calculated in the same manner as in section 8.2.1.
1521 The PID correction factors for signal and the numbers of $B\bar{B}$ meson pairs on MC
1522 and data are the same as in the case of control decay branching ration measurement

1523 Finally, we can determine the branching ratios based on the calculations in Eq.
1524 (8.5). The obtained values are shown in Table 8.5. We see that the measured value
1525 is twice as large as the MC value. The error in case of the MC result is statistical
1526 only, while the first and the second error in the data result are the statistical and
1527 the systematic errors, respectively.

1528 With the statistical significance of about 5.9σ this is the first observation of the
1529 charmless semileptonic decay $B^+ \rightarrow K^+ K^- \ell^+ \nu$. It indicates that the MC contribu-
1530 tion to our simulated samples is underestimated. This analysis has shown that the
1531 branching fraction of the decay seems considerable enough in order to affect results
1532 in precision physics in cases where they are ignored.

	$\mathcal{B}_{GEN} [\times 10^{-5}]$	$\mathcal{B}^{MC} [\times 10^{-5}]$	$\mathcal{B}^{\text{data}} [\times 10^{-5}]$
$\ell = e \text{ or } \mu$	1.57	1.55 ± 0.15	$3.04 \pm 0.54_{-0.71}^{+0.74}$

Table 8.5.: Signal decay branching ratio results for MC and data.

1533 8.3.3. Signal Distribution in bins of m_{KK}

1534 It is possible to take a deeper look in the signal distribution over the m_{KK} variable
1535 by performing the signal fit in bins of the m_{KK} distribution instead of a single fit
1536 over the whole region. This offers deeper insight into the real m_{KK} distribution and
1537 provides more detail about the reliability of our MC samples. Results can be used to
1538 check theoretical predictions or to update existing MC generators. Table 8.6 shows
1539 the selected regions in m_{KK} , along with the corresponding signal yields on MC and
1540 data. The results are graphically shown in Figure 8.9. Figures for each m_{KK} window
1541 fit can be found in Appendix C.1. In the given results, the m_{KK} regions around the
1542 ϕ and the D^0 resonances has been excluded (see section 6.1).

Region	Expected	MC fit	Data fit
$0.980 < m_{KK} < 1.357$	80	77 ± 12	220 ± 40
$1.357 < m_{KK} < 1.735$	96	94 ± 19	234 ± 55
$1.735 < m_{KK} < 2.112$	50	49 ± 8	81 ± 51
$2.112 < m_{KK} < 2.490$	18	18 ± 6	1 ± 20
$2.490 < m_{KK} < 2.867$	5	12 ± 4	-11 ± 8
$2.867 < m_{KK} < 3.245$	1	-3 ± 2	-7 ± 6
$3.245 < m_{KK} < 3.623$	1	-2 ± 1	-7 ± 5
$3.623 < m_{KK} < 4.000$	0	-1 ± 1	-4 ± 4

Table 8.6.: Various signal fit yields for for each of the defined m_{KK} windows.

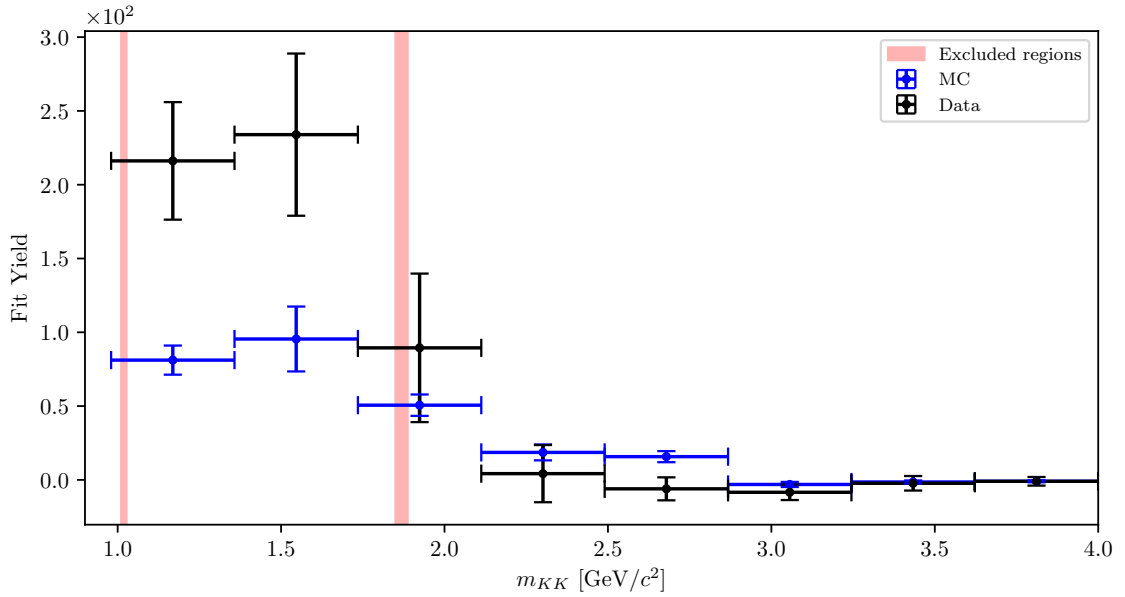


Figure 8.9.: Signal yield distribution as a function of bins in m_{KK} .

1543 8.3.4. Signal Distribution in bins of q^2

1544 Similarly, we look at the signal distribution over the square momentum transfer to
1545 the lepton pair, q^2 , by performing the signal fit in bins of the q^2 distribution instead
1546 of a single fit over the whole region. In addition to the results offering insight for
1547 updating the MC generators, we are able to extract the V_{ub} parameter of the CKM
1548 matrix, provided that the appropriate form factors of the decay are available. This
1549 was not performed in this analysis. Table 8.7 shows the selected regions in q^2 , along
1550 with the corresponding signal yields on MC and data. The results are graphically
1551 shown in Figure 8.10. Figures for each q^2 window fit can be found in Appendix C.2.

Region	Expected	MC fit	Data fit
$0.000 < q^2 < 2.250$	41	37 ± 14	-9 ± 39
$2.250 < q^2 < 4.500$	49	44 ± 6	92 ± 38
$4.500 < q^2 < 6.750$	46	44 ± 7	192 ± 38
$6.750 < q^2 < 9.000$	40	40 ± 8	103 ± 36
$9.000 < q^2 < 11.250$	35	35 ± 7	59 ± 33
$11.250 < q^2 < 13.500$	25	25 ± 6	44 ± 25
$13.500 < q^2 < 15.750$	13	11 ± 3	18 ± 16
$15.750 < q^2 < 18.000$	2	-0 ± 2	10 ± 22

Table 8.7.: Various signal fit yields for for each of the defined q^2 windows.

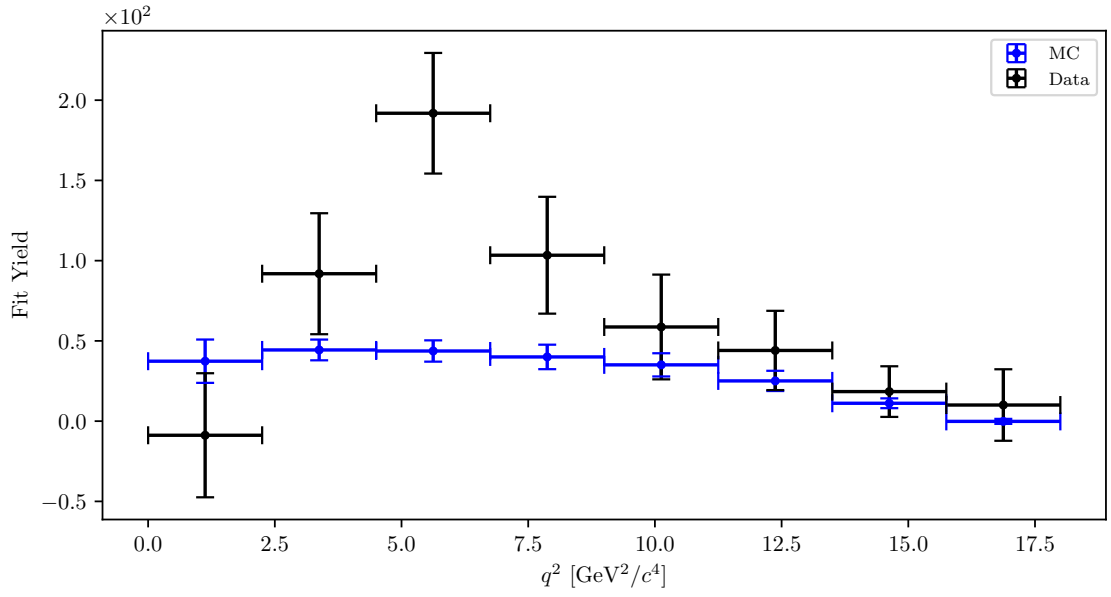


Figure 8.10.: Signal yield distribution as a function of bins in q^2 .

Chapter 9.

Systematic Uncertainty

In this chapter, the systematic errors of the analysis are discussed. These uncertainties arise due to various reasons, some of them being the difference between real and simulated data, or due to the nature of the approaches taken in a specific analysis. Depending on their type, some uncertainties are generic and prepared beforehand in order to be used in all analyses, while others are analysis specific and possible sources need to be thought through thoroughly. All systematic uncertainties contributions which are cited in the final result in section 8.3.2 are summarized at the end of this chapter.

9.1. Contributions to Systematics

9.1.1. PID Efficiency Correction

The PID selection efficiency for the three charged particles in our signal decay needs to be corrected on MC due to various differences when comparing to data. The Belle PID group has prepared correction factors and systematics tables for PID efficiencies for all charged particles. In case of kaon ID and lepton ID, the tables are binned in experiment numbers, particle momentum and in $\cos\theta$ of the particle direction, where, for each bin, a ratio of efficiencies between MC and data is provided, as well as the systematic errors. Each particle's correction factor and error is shown in Table 9.1, as well as the corresponding entry for all 3 particles. The entries are shown for both signal and control region, where the differences are related to the kinematic variations between the signal and control modes.

The central values were obtained with a weighted average over all experiments, where 100% correlation for error calculation was assumed. A full correlation was also assumed when calculating the KK correction, as both K use the same PID information.

The final PID efficiency systematic error on the full MC sample is determined to be

$$\sigma_{\text{sys}}^{\text{PID}} = 10, \quad \delta_{\text{sys}}^{\text{PID}} = 2.0\%, \quad (9.1)$$

for the signal as well as the control decay.

PID correction and systematics	Control decay	Signal decay
Same sign K (w.r.t the B meson)	1.005 ± 0.009	1.007 ± 0.010
Opposite sign K (w.r.t the B meson)	1.004 ± 0.009	1.006 ± 0.009
e	0.977 ± 0.011	0.976 ± 0.011
μ	0.985 ± 0.009	0.980 ± 0.009
ℓ	0.981 ± 0.007	0.980 ± 0.007
KKe	0.986 ± 0.021	0.988 ± 0.022
$KK\mu$	0.994 ± 0.020	0.993 ± 0.021
$KK\ell$	0.991 ± 0.019	0.990 ± 0.020

Table 9.1.: PID correction factors and systematic error for various charged particles and their combinations.

1581 9.1.2. Fit Bias and Binning Effects

Signal and background templates in our analysis are not perfectly distinct from one another and may potentially cause some over- or underestimation of the signal fit yield. In order to study this problem, we estimate the bias from the binning study performed in section 8.1 as well as the linearity test toy MC study in section 7.3.2. The two bias functions describe a bias in each direction and are approximated as

$$f_{\min}(x) = -7.25 - 1.12x - \sigma_{f_{\min}}(x), \quad (9.2)$$

$$\sigma_{f_{\min}}(x) = \sqrt{0.050x^2 - 0.175x + 0.410}, \quad (9.3)$$

$$f_{\max}(N_b) = 5.06 + \sigma_{f_{\max}}(N_b), \quad (9.4)$$

$$\sigma_{f_{\max}}(N_b) = \sqrt{0.004N_b^2 - 0.113N_b + 1.112} \quad (9.5)$$

1582 where x represents the signal yield fraction of the data fit and N_b represents the
1583 binning choice of the fit. Values of 1σ intervals have been added to the bias functions
1584 in order to be more conservative. The extracted signal yield on data with the fit
1585 setup of $N_b = 19$ bins was determined to be $N_{sig} = 487$, which leads to $x =$
1586 $N_{sig}/N_{sig}^{MC} = 487/250 \approx 2$. The bias interval is therefore

$$\sigma_{\text{sys}}^{\text{bias}} = {}^{+6}_{-10}, \quad \delta_{\text{sys}}^{\text{bias}} = {}^{+1.2\%}_{-2.0\%}. \quad (9.6)$$

1587 9.1.3. Gaussian Constraints

As mentioned in section 8.3, it is possible to estimate the size of the systematic error of the Gaussian constraints. By fixing the constraints to the nominal values, determined by the data fit, we obtain the pure statistical error, which can then be subtracted from the average fit error in order to determine the systematic uncertainty contribution due to using Gaussian constraints. For the fixed and non-fixed case we perform 500 fits and calculate with the mean values. The split errors are then

$$\bar{N}_{sig} = 489 \pm 85, \quad (9.7)$$

$$\bar{N}_{sig}^{fix} = 489 \pm 81, \quad (9.8)$$

$$\sigma_{\text{stat}} = 81, \quad \sigma_{\text{sys}}^{GC} = 26, \quad \delta_{\text{sys}}^{GC} = 5.0\%, \quad (9.9)$$

1588 where GC stands for Gaussian constraint. We see that the constrained channels are
 1589 very well defined and introduce little to none additional uncertainty.

1590 9.1.4. Fit Template Smearing and Offset

1591 The smearing and offset of the ΔE variable was discussed in section 7.1.1, where we
 1592 have estimated the central value of the parameters as well as their range in the 1σ
 1593 confidence level. We have to perform a study of the effects of different smearing and
 1594 offset parameter values on the final value of the signal yield. From section 7.1.1, the
 1595 parameter values are

- 1596 • Smearing: 40^{+15}_{-17} MeV,
- 1597 • Offset: $6^{+4.6}_{-6}$ MeV.

1598 Since the two parameters are largely uncorrelated, we are able to perform the study
 1599 in the form of signal fits with four different combinations of parameters in the given
 1600 range. For each parameter setting 500 fits are performed and the following results
 1601 are obtained

- 1602 • Set: [smearing, offset]: [23 MeV, 6 MeV], Result: $\bar{N}_{sig} = 448$,
- 1603 • Set: [smearing, offset]: [55 MeV, 6 MeV], Result: $\bar{N}_{sig} = 540$,
- 1604 • Set: [smearing, offset]: [40 MeV, 0 MeV], Result: $\bar{N}_{sig} = 550$,
- 1605 • Set: [smearing, offset]: [40 MeV, 12.6 MeV], Result: $\bar{N}_{sig} = 439$,

which results in the following estimate of systematic uncertainties for smearing and
 offset parameters

$$\sigma_{\text{sys}}^{\text{sm.}} = \begin{matrix} +51 \\ -41 \end{matrix}, \quad \delta_{\text{sys}}^{\text{sm.}} = \begin{matrix} +10.4\% \\ -8.3\% \end{matrix}, \quad (9.10)$$

$$\sigma_{\text{sys}}^{\text{off.}} = \begin{matrix} +61 \\ -48 \end{matrix}, \quad \delta_{\text{sys}}^{\text{off.}} = \begin{matrix} +12.4\% \\ -9.9\% \end{matrix}. \quad (9.11)$$

1606 9.1.5. Effects of a Finite MC sample

1607 The shape of signal and backgrounds templates in our analysis is fixed and only
 1608 their normalization is considered as a floating parameter in the fit. Due to the
 1609 finite size of the MC sample, the template shape introduces an additional source of
 1610 uncertainty, as it may differ if produced in a separate, equal-sized MC sample. Since
 1611 the bins in these 2D histogram templates are statistically independent, we take the
 1612 content of each bin and vary the value according to the Poisson distribution. This
 1613 procedure is repeated 500 times and the width of the fit yield distribution is taken
 1614 as the uncertainty estimate. Since the systematic effects of the Gaussian constraints
 1615 are included in the obtained distribution, we appropriately discard them to evaluate
 1616 this systematic contribution. The resulting finite MC sample contribution of the
 1617 systematic uncertainty is

$$\sigma_{\text{sys}}^{MC} = 11, \quad \delta_{\text{sys}}^{MC} = 2.4\%. \quad (9.12)$$

9.1.6. MVA Selection Efficiencies

Control sample fits allow us to check the behavior of optimized MVA cuts on MC as well as data and see if any of the MVA steps introduce a possible disagreement between the two. We compare control yields, their ratios, and ratios of cut efficiencies (double ratios). The following cut scenarios are studied

- (a) final selection before any MVA step,
- (b) (a) + $BDT_{q\bar{q}}$ cut,
- (c) (a) + $uBDT_{B\bar{B}}$ cut,
- (d) (a) + $BDT_{q\bar{q}}$ + $uBDT_{B\bar{B}}$ cut (final selection).

The results for control fit yields, their ratios, and double ratios are shown in Figure 9.1. The plot shows that the yield ratios and cut efficiency ratios are consistent with 1. This means that data and MC are in agreement before as well as after applying the final selection cuts. This is an important check since the behavior of our analysis on the control sample suggests that the final selection is not over-optimized to signal MC.

We estimate the systematic error due to the MVA selection steps as the standard deviation of double ratio entries around the nominal value for each step in the MVA selection, except for the final two values for e and μ modes, since we are performing the inclusive lepton measurement. The systematic error estimate for this contribution is

$$\sigma_{\text{sys.}}^{\text{MVA}} = 5, \quad \sigma_{\text{sys.}}^{\text{MVA}} = 1.0\%. \quad (9.13)$$

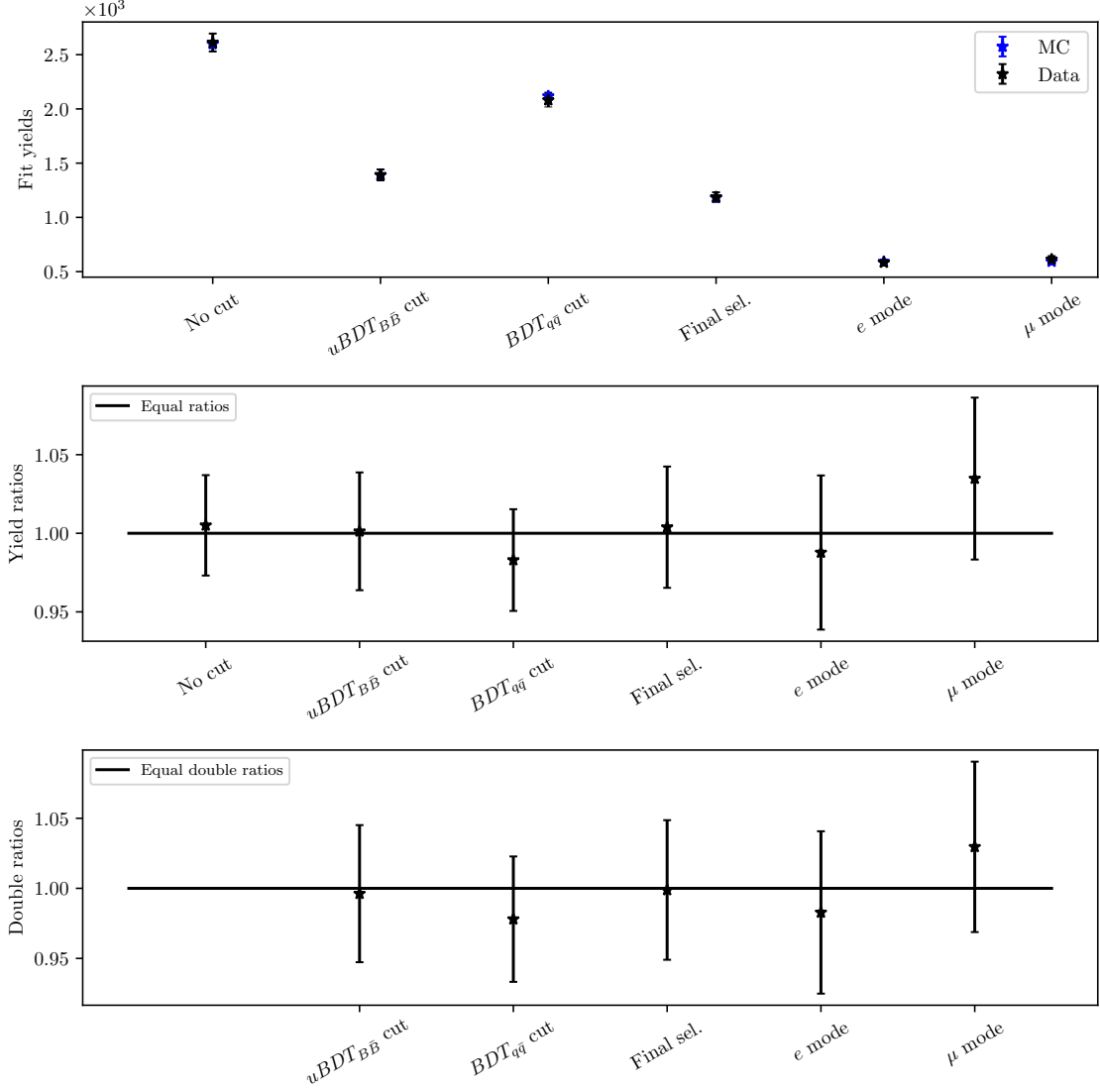


Figure 9.1.: Fit yields, their ratios and ratios of cut efficiencies (double ratios) for the control sample fits to data and MC.

9.1.7. Model Uncertainty Effects

The signal decay model used in the generation step is **ISGW2** [17], which is known to result in unrealistic predictions and poor agreement with data, so it is not the most precise model for our signal MC sample. Due to this model unreliability, our analysis has been set up as model-independent as possible via means of not using variables, which exhibit model dependence. Such variables are squared momentum transfer to the lepton pair (q^2), the invariant mass of the two kaon daughters (m_{KK}) or decay angle between any two charged particles in the final state. In order to test the effects of model dependency on our final result, we prepare three additional signal MC samples, produced with three extreme scenarios of decay model choice. In the first scenario we generate the signal MC sample with a generic phase-space decay mode **PHSP** [10], which results in continuum-like distributions of q^2 and m_{KK} . In the other scenario, only resonant-like contributions of m_{KK} are used. The third

1651 scenario is obtained by sampling the first model in a non-uniform way so that the
 1652 final distribution is more dominant in the higher region of the q^2 distribution. These
 1653 scenarios act as extreme cases of decay model choice and present a reasonable, if
 1654 not a conservative measure of the model uncertainty.

1655 Shape Effects of the Model

1656 The shape effect is evaluated by using signal templates, constructed from different
 1657 signal MC samples. We perform 500 fits for each case. The differences between
 1658 mean values of fit yields serve as an estimate of the model uncertainty. Figure 9.2
 1659 shows the generated m_{KK} and q^2 distributions of the three mentioned decay models
 1660 as well as distributions of ΔE and M_{BC} after the final selection. Figures show a
 1661 good agreement of ΔE and M_{BC} for different cases of the generator model.

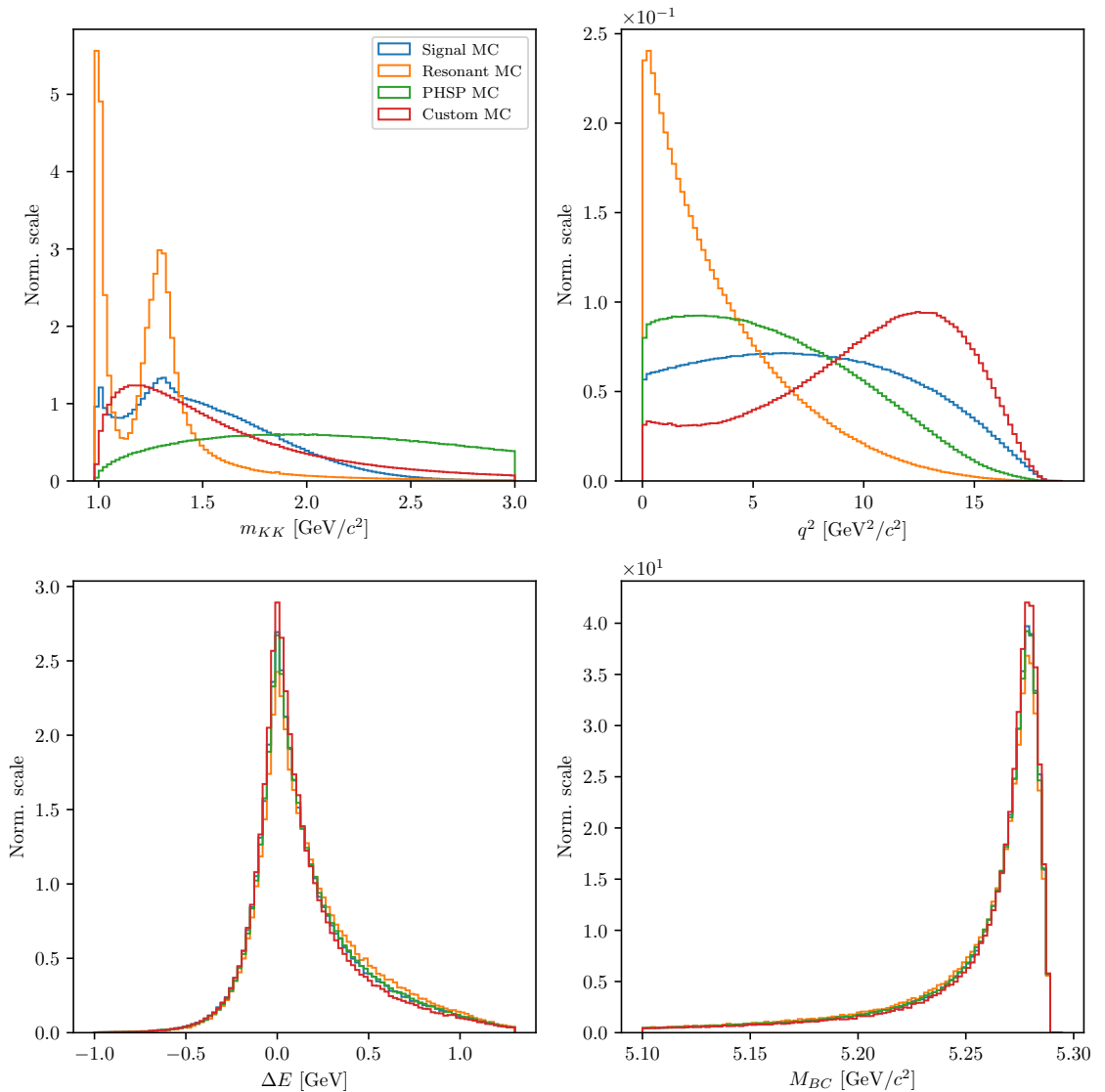


Figure 9.2.: m_{KK} (top left), q^2 (top right), ΔE (bottom left) and M_{BC} (bottom right) for the main and three extreme cases of signal MC generator choice.

The resulting average signal yields for the three model choices are

$$\bar{N}_{sig} = 487, \quad (9.14)$$

$$\bar{N}_{sig}^{\text{PHSP}} = 532, \quad (9.15)$$

$$\bar{N}_{sig}^{\text{res.}} = 490. \quad (9.16)$$

1662 We see that in general the model with a worse resolution in ΔE and M_{BC} will likely
 1663 result in a larger yield, and vice versa for a model with a better resolution. Overall
 1664 these models give a conservative estimation of the systematic uncertainty due to the
 1665 shape effects with a value of

$$\sigma_{\text{sys}}^{\text{mod.}} = {}^{+45}_{-41}, \quad \delta_{\text{sys}}^{\text{mod.}} = {}^{+9.1\%}_{-8.5\%}. \quad (9.17)$$

1666 Model Efficiency

1667 The second way that the choice of a model can affect our analysis is the model
 1668 efficiency. Due to the different properties of the model, we can expect different
 1669 efficiencies in model dependent variables like m_{KK} and q^2 . The efficiency as a
 1670 function of these variables is shown in Figure 9.3 for all discussed generator models.
 1671 In a perfect scenario, a model-independent analysis should produce a flat efficiency
 1672 as a function of the model dependent variables. While this is rarely achieved in
 1673 practice, the overall discrepancy of the efficiency functions from very different models
 1674 is relatively small.

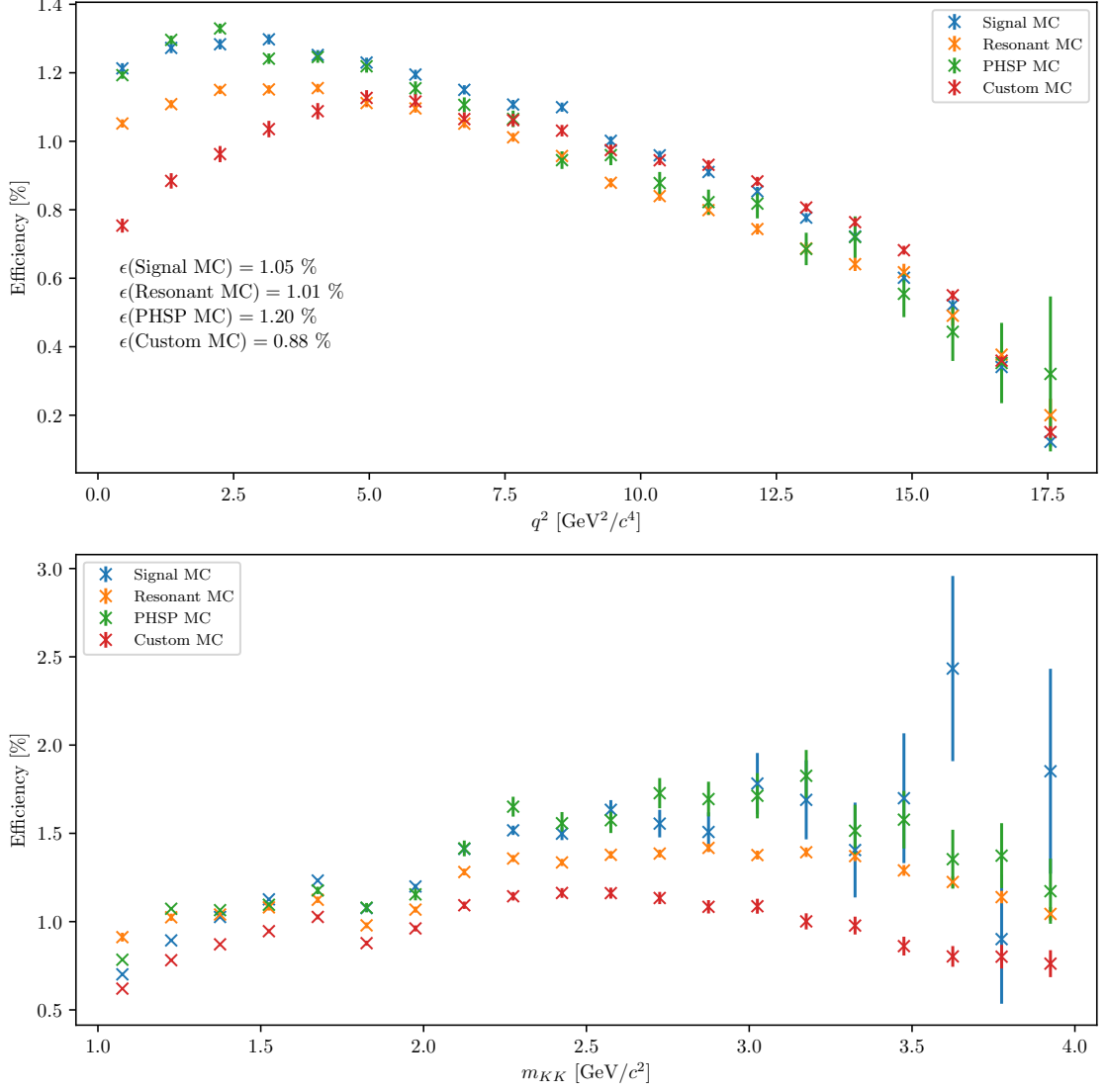


Figure 9.3.: Efficiency as a function of q^2 (top) and m_{KK} (bottom) for the main and three extreme cases of signal MC generator choice.

1675 To estimate the systematic uncertainty due to model efficiency effects, we compute
 1676 the overall efficiencies and take the extreme cases of relative differences with regard
 1677 to the main signal MC model. This results in

$$\sigma_{\text{sys}}^{\text{mod.}} = {}^{+70}_{-79}, \quad \delta_{\text{sys}}^{\text{mod.}} = {}^{+14.3\%}_{-16.2\%}. \quad (9.18)$$

1678 9.2. Summary of Systematics

1679 The summary of all systematic uncertainties is shown in Table 9.2. The full estimate
 1680 of the systematic uncertainty is summed up in quadrature and applied to the result
 1681 in section 8.3.2.

Source	σ	δ [%]
PID	10	2.0
Fit Bias	+6 -10	+1.2 -2.0
Gaussian Constraints	26	5.0
Template Smearing	+51 -41	+10.4 -8.3
Template Offset	+61 -48	+12.4 -9.9
Finite MC Effects	11	2.4
MVA Selection	5	1.0
Model Shape	+45 -41	+9.1 -8.5
Model Efficiency	+70 -79	+14.3 -16.2
Total	+119 -114	+24.3 -23.3

Table 9.2.: Summary of systematic uncertainties in this analysis.

Bibliography

- 1683 [1] Sheldon L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear*
1684 *Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: [https://doi.org/](https://doi.org/10.1016/0029-5582(61)90469-2)
1685 [10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2). URL: [http://www.sciencedirect.com/](http://www.sciencedirect.com/science/article/pii/0029558261904692)
1686 [science/article/pii/0029558261904692](http://www.sciencedirect.com/science/article/pii/0029558261904692).
- 1687 [2] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 1967),
1688 pp. 1264–1266. DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264). URL: [https://link.](https://link.aps.org/doi/10.1103/PhysRevLett.19.1264)
1689 [aps.org/doi/10.1103/PhysRevLett.19.1264](https://link.aps.org/doi/10.1103/PhysRevLett.19.1264).
- 1690 [3] Abdus Salam. “Weak and electromagnetic interactions”. In: *Selected Papers*
1691 *Of Abdus Salam: (With Commentary)*. World Scientific, 1994, pp. 244–254.
- 1692 [4] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-
1693 Hadron Symmetry”. In: *Phys. Rev. D* 2 (7 1970), pp. 1285–1292. DOI: [10.](https://doi.org/10.1103/PhysRevD.2.1285)
1694 [1103/PhysRevD.2.1285](https://doi.org/10.1103/PhysRevD.2.1285). URL: [https://link.aps.org/doi/10.1103/](https://link.aps.org/doi/10.1103/PhysRevD.2.1285)
1695 [PhysRevD.2.1285](https://link.aps.org/doi/10.1103/PhysRevD.2.1285).
- 1696 [5] Nicola Cabibbo. “Unitary symmetry and leptonic decays”. In: *Physical Review*
1697 *L* 10.12 (1963), p. 531.
- 1698 [6] Makoto Kobayashi and Toshihide Maskawa. “CP-violation in the renormal-
1699 izable theory of weak interaction”. In: *Progress of Theoretical Physics* 49.2
1700 (1973), pp. 652–657.
- 1701 [7] Lincoln Wolfenstein. “Parametrization of the Kobayashi-Maskawa Matrix”. In:
1702 *Phys. Rev. Lett.* 51 (21 Nov. 1983), pp. 1945–1947. DOI: [10.1103/PhysRevLett.](https://doi.org/10.1103/PhysRevLett.51.1945)
1703 [51.1945](https://doi.org/10.1103/PhysRevLett.51.1945). URL: [https://link.aps.org/doi/10.1103/PhysRevLett.51.](https://link.aps.org/doi/10.1103/PhysRevLett.51.1945)
1704 [1945](https://link.aps.org/doi/10.1103/PhysRevLett.51.1945).
- 1705 [8] Y. Amhis et al. “Averages of b -hadron, c -hadron, and τ -lepton properties as of
1706 summer 2016”. In: *Eur. Phys. J. C* 77 (2017). updated results and plots avail-
1707 able at <https://hflav.web.cern.ch>, p. 895. DOI: [10.1140/epjc/s10052-](https://doi.org/10.1140/epjc/s10052-017-5058-4)
1708 [017-5058-4](https://doi.org/10.1140/epjc/s10052-017-5058-4). arXiv: [1612.07233](https://arxiv.org/abs/1612.07233) [hep-ex].
- 1709 [9] P. Gambino, P. Giordano, et al. “Inclusive semileptonic B decays and the
1710 determination of $|V_{ub}|$ ”. In: *JHEP* 10 (2007), p. 058. DOI: [10.1088/1126-](https://doi.org/10.1088/1126-6708/2007/10/058)
1711 [6708/2007/10/058](https://doi.org/10.1088/1126-6708/2007/10/058). arXiv: [0707.2493](https://arxiv.org/abs/0707.2493) [hep-ph].
- 1712 [10] David J Lange. “The EvtGen particle decay simulation package”. In: *Nuclear*
1713 *Instruments and Methods in Physics Research Section A: Accelerators, Spec-*
1714 *trometers, Detectors and Associated Equipment* 462.1-2 (2001), pp. 152–155.
- 1715 [11] Sea Agostinelli, John Allison, et al. “GEANT4 a simulation toolkit”. In:
1716 *Nuclear instruments and methods in physics research section A: Accelerators,*
1717 *Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–
1718 303.
- 1719 [12] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. “PYTHIA 6.4 physics
1720 and manual”. In: *Journal of High Energy Physics* 2006.05 (2006), p. 026.

- [13] T. Kuhr, C. Pulvermacher, et al. “The Belle II Core Software”. In: (2018). arXiv: [1809.04299](https://arxiv.org/abs/1809.04299) [physics.comp-ph].
- [14] Keck. T et al. “B2BII Data conversion from Belle to Belle II”. In: (). To appear.
- [15] Thomas Keck. “Machine learning algorithms for the Belle II experiment and their validation on Belle data”. Karlsruher Institut fr Technologie, Diss., 2017. Dr. Karlsruher Institut fr Technologie, 2017, 240 pages. URL: <https://ekp-invenio.physik.uni-karlsruhe.de/record/48940>.
- [16] D. M. Asner, K. W. Edwards, and others. “Study of exclusive charmless semileptonic B decays and extraction of $|V_{ub}|$ at CLEO”. In: *Phys. Rev. D* 76 (1 2007), p. 012007. DOI: [10.1103/PhysRevD.76.012007](https://doi.org/10.1103/PhysRevD.76.012007). URL: <https://link.aps.org/doi/10.1103/PhysRevD.76.012007>.
- [17] Daryl Scora and Nathan Isgur. “Semileptonic meson decays in the quark model: An update”. In: *Phys. Rev. D* 52 (1995), pp. 2783–2812. DOI: [10.1103/PhysRevD.52.2783](https://doi.org/10.1103/PhysRevD.52.2783). arXiv: [hep-ph/9503486](https://arxiv.org/abs/hep-ph/9503486) [hep-ph].
- [18] D. M. Asner, M. Athanas, et al. “Search for exclusive charmless hadronic B decays”. In: *Physical Review D* 53.3 (1996), p. 1039.
- [19] A. J. Bevan, B. Golob, et al. “The physics of the B factories”. In: *The European Physical Journal C* 74.11 (2014), p. 3026.
- [20] “Study of $B \rightarrow \pi l \nu$ and $B \rightarrow \rho l \nu$ decays and determination of $|V_{ub}|$, author = del Amo Sanchez, P. and Lees, J. P. and Poireau, V. and Prencipe, E. and others, collaboration = BABAR Collaboration, journal = Phys. Rev. D, volume = 83, issue = 3, pages = 032007, numpages = 45, year = 2011, month = 2, publisher = American Physical Society, doi = 10.1103/PhysRevD.83.032007, url = <https://link.aps.org/doi/10.1103/PhysRevD.83.032007>”. In: ().
- [21] Justin Stevens and Mike Williams. “uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers”. In: *Journal of Instrumentation* 8.12 (2013), P12013.
- [22] Wouter Verkerke and David Kirkby. “The RooFit toolkit for data modeling”. In: *Statistical Problems in Particle Physics, Astrophysics and Cosmology*. World Scientific, 2006, pp. 186–189.
- [23] F. James. “MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1”. In: (1994).
- [24] M. Tanabashi. “Review of particle physics”. In: *Phys. Rev. D* 98 (2018), p. 030001.
- [25] Sangtae Ahn and Jeffrey A Fessler. “Standard errors of mean, variance, and standard deviation estimators”. In: *EECS Department, The University of Michigan* (2003), pp. 1–2.
- [26] Marc Paterno. *Calculating efficiencies and their uncertainties*. Tech. rep. Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), 2004.

Appendix A.

ROE MVA Control Plots

A.1. ROE Clean-up π^0 Training

A.1.1. Variable Importance

	Name	Alias	Importance
0	chiProb	v_0	0.280
1	useCMSFrame(daughterAngleInBetween(0,1))	v_1	0.203
2	daughter(0,useCMSFrame(p))	v_2	0.073
3	InvM	v_3	0.072
4	daughter(1,clusterHighestE)	v_4	0.061
5	daughter(1,clusterTheta)	v_5	0.049
6	daughter(1,p)	v_6	0.047
7	daughter(0,clusterHighestE)	v_7	0.029
8	daughter(0,clusterTheta)	v_8	0.024
9	daughter(0,clusterE9E25)	v_9	0.018
10	daughter(0,minC2HDist)	v_{10}	0.018
11	daughter(1,minC2HDist)	v_{11}	0.017
12	daughter(1,clusterE9E25)	v_{12}	0.016
13	useRestFrame(daughterAngleInBetween(0,1))	v_{13}	0.014
14	daughter(1,clusterNHits)	v_{14}	0.013
15	daughter(0,clusterNHits)	v_{15}	0.011
16	daughter(0,clusterErrorE)	v_{16}	0.009
17	daughter(1,clusterErrorE)	v_{17}	0.009
18	SigMBF	v_{18}	0.007
19	useCMSFrame(p)	v_{19}	0.006
20	daughter(0,p)	v_{20}	0.005
21	SigM	v_{21}	0.005
22	daughter(1,useCMSFrame(p))	v_{22}	0.005
23	useLabFrame(daughterAngleInBetween(0,1))	v_{23}	0.005
24	p	v_{24}	0.003

Table A.1.: Variable names, aliases and importance in the scope of π^0 MVA training for ROE clean-up.

A.1.2. Variable Distributions

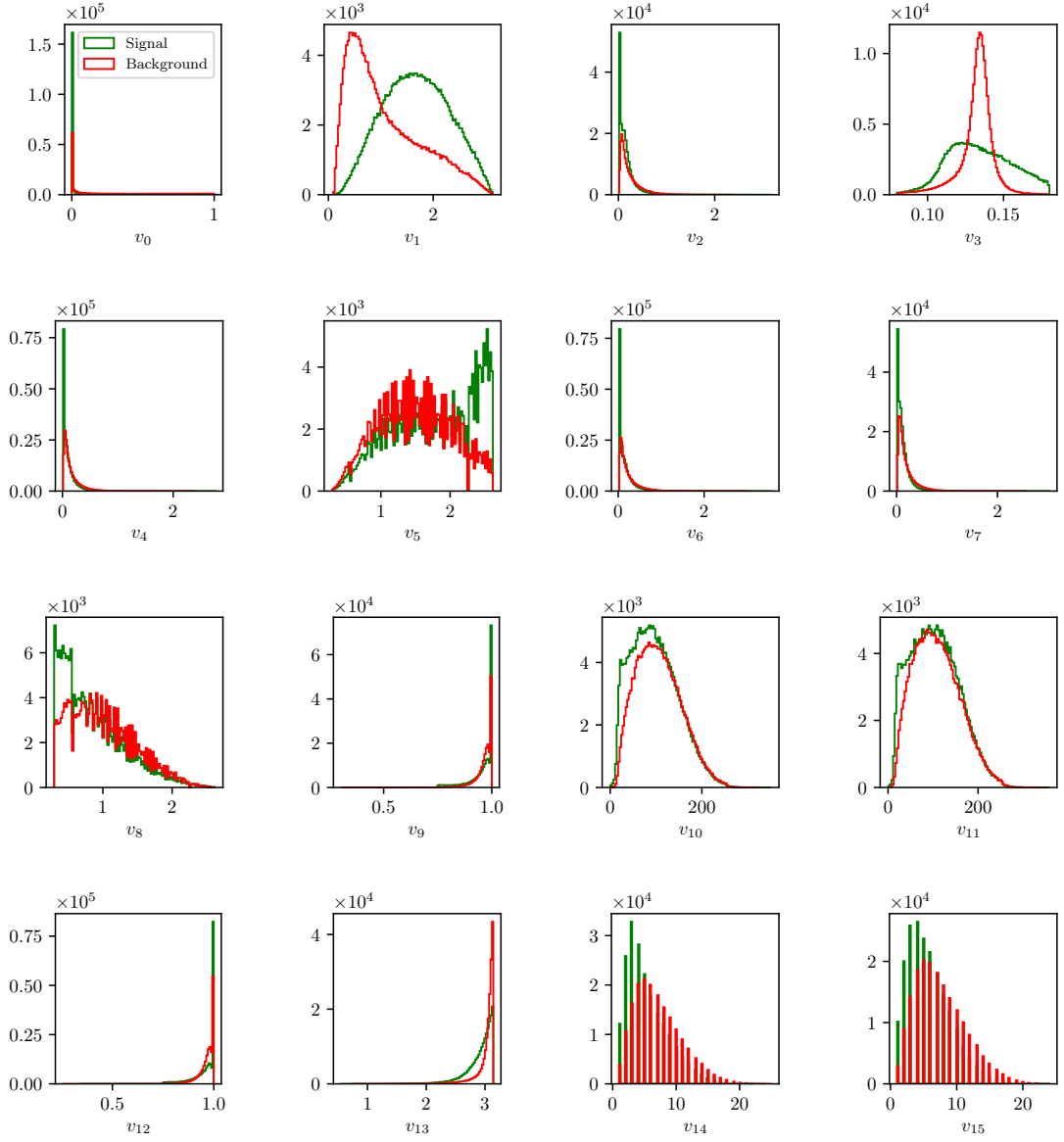


Figure A.1.: Feature distributions for MVA training of π^0 candidates in the scope of ROE clean-up.

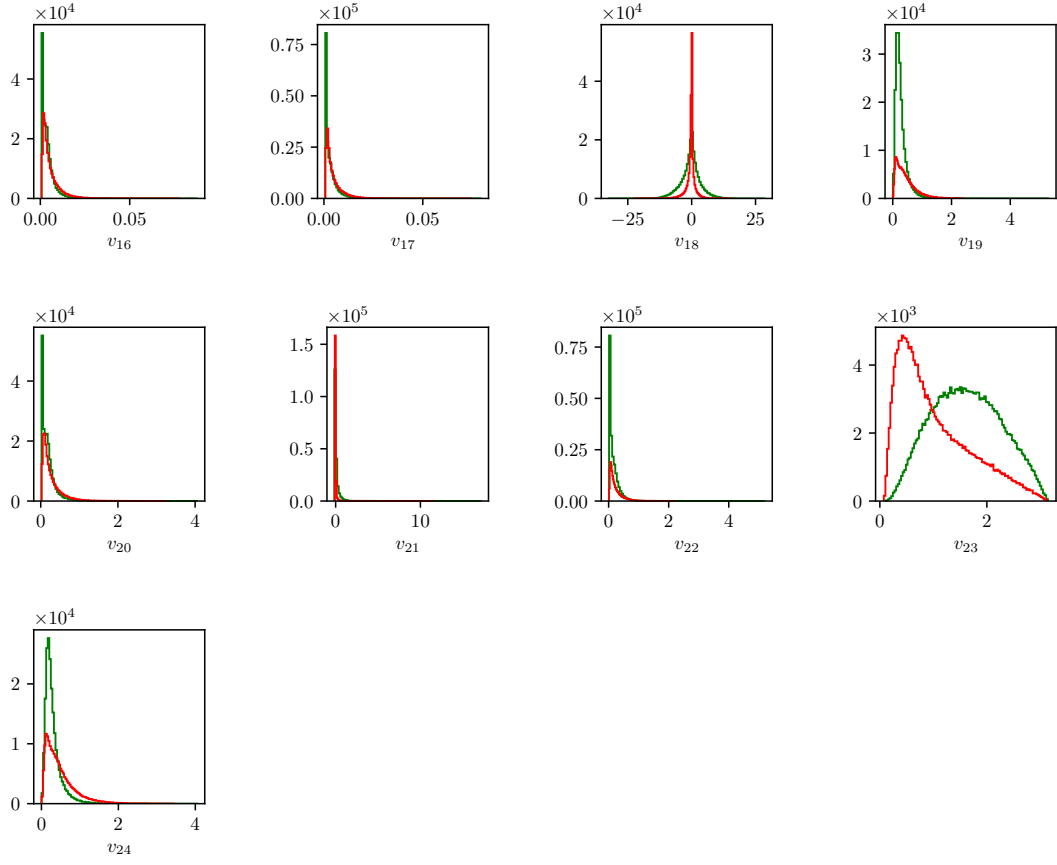


Figure A.1.: Feature distributions for MVA training of π^0 candidates in the scope of ROE clean-up.

A.1.3. Hyper-parameter Optimization

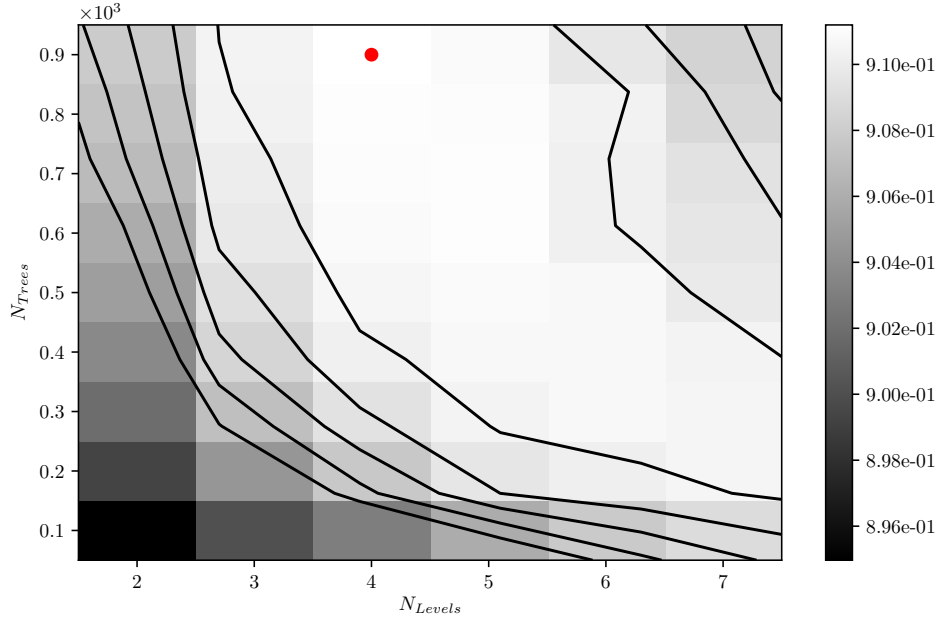


Figure A.2.: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of π^0 candidates in the scope of the ROE clean-up.

A.1.4. Results

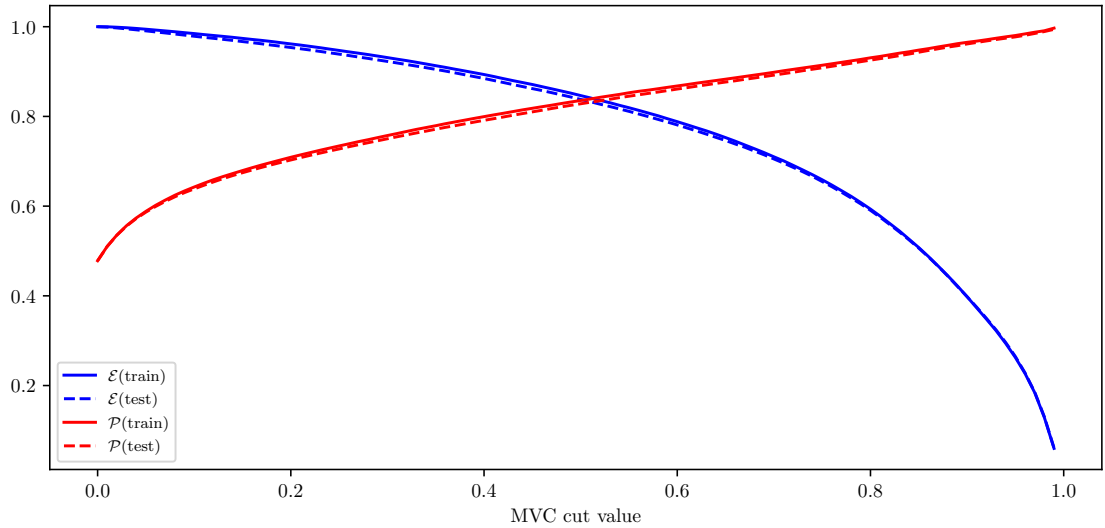


Figure A.3.: Efficiency (\mathcal{E}) and purity (\mathcal{P}) of the MVA classifier output for π^0 candidates training on the train (solid) and test (dashed) samples.

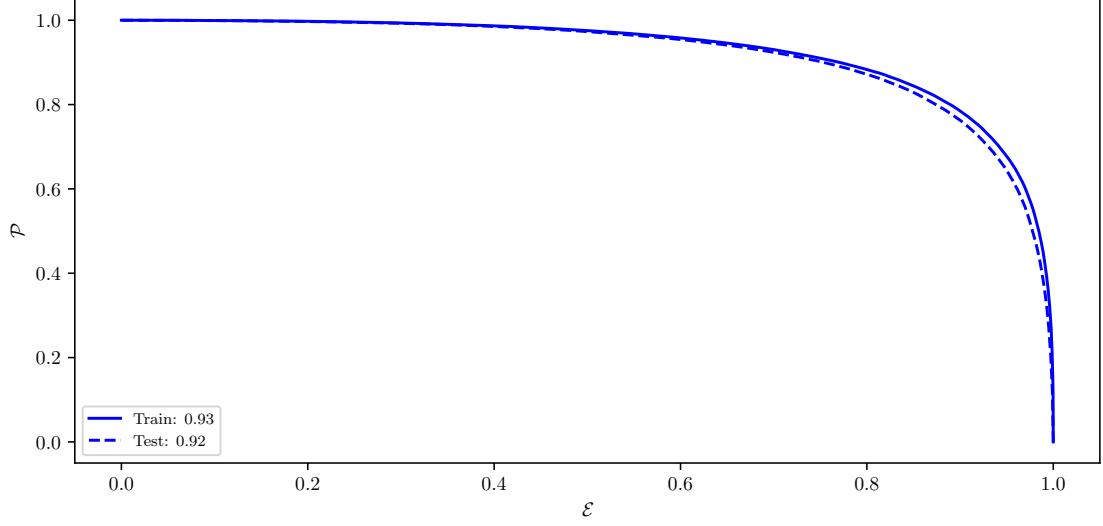


Figure A.4.: ROC curves of the MVA classifier output for π^0 candidates training on the train (solid) and test (dashed) samples.

1768 A.2. ROE Clean-up γ Training

1769 A.2.1. Variable Importance

	Name	Alias	Importance
0	p	v_0	0.327
1	pi0p	v_1	0.243
2	clusterHighestE	v_2	0.226
3	minC2HDist	v_3	0.052
4	cosTheta	v_4	0.036
5	clusterE9E25	v_5	0.031
6	clusterNHits	v_6	0.025
7	clusterUncorrE	v_7	0.022
8	clusterR	v_8	0.015
9	useCMSFrame(p)	v_9	0.013
10	clusterErrorE	v_{10}	0.010
11	clusterReg	v_{11}	0.000

Table A.2.: Variable names, aliases and importance in the scope of γ MVA training for ROE clean-up.

A.2.2. Variable Distributions

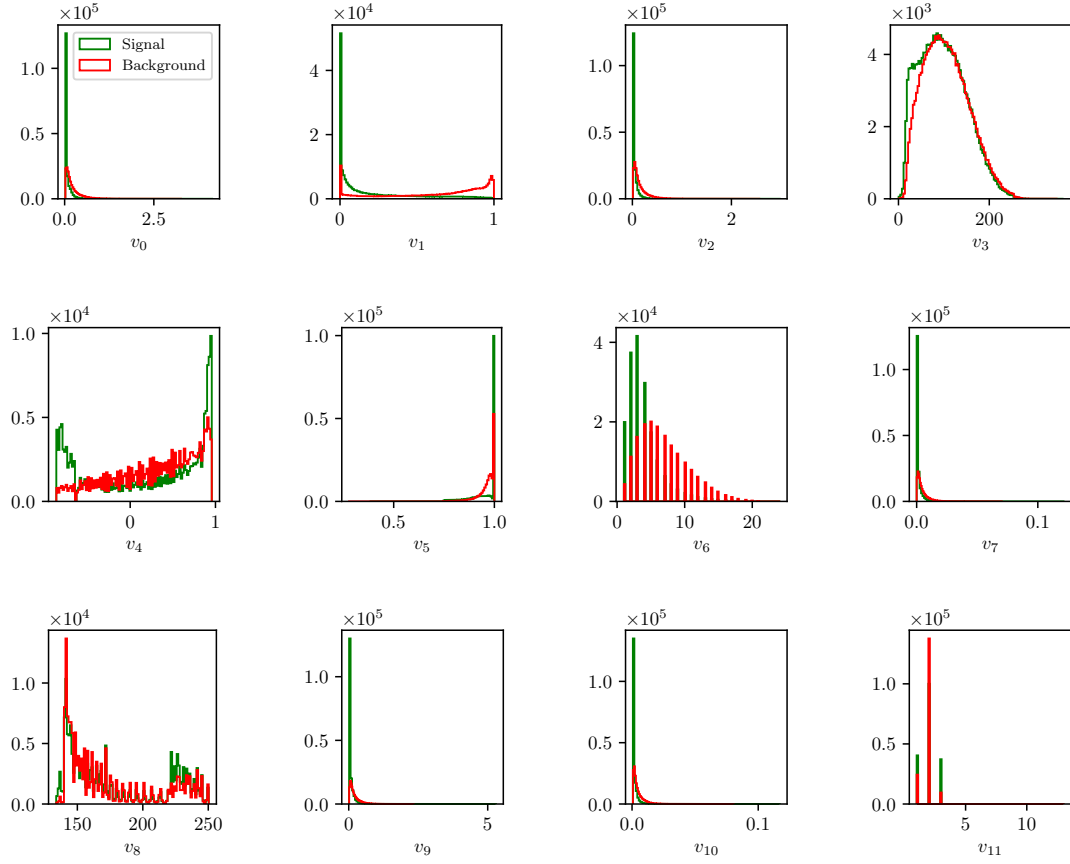


Figure A.5.: Feature distributions for MVA training of γ candidates in the scope of ROE clean-up.

1771 A.2.3. Hyper-parameter Optimization

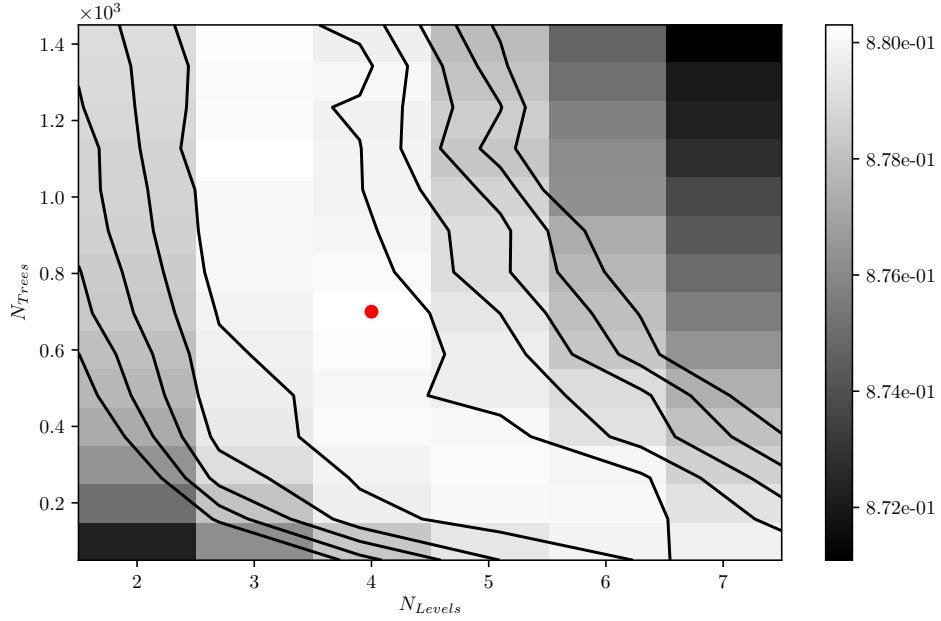


Figure A.6.: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of γ candidates in the scope of the ROE clean-up.

1772 A.2.4. Results

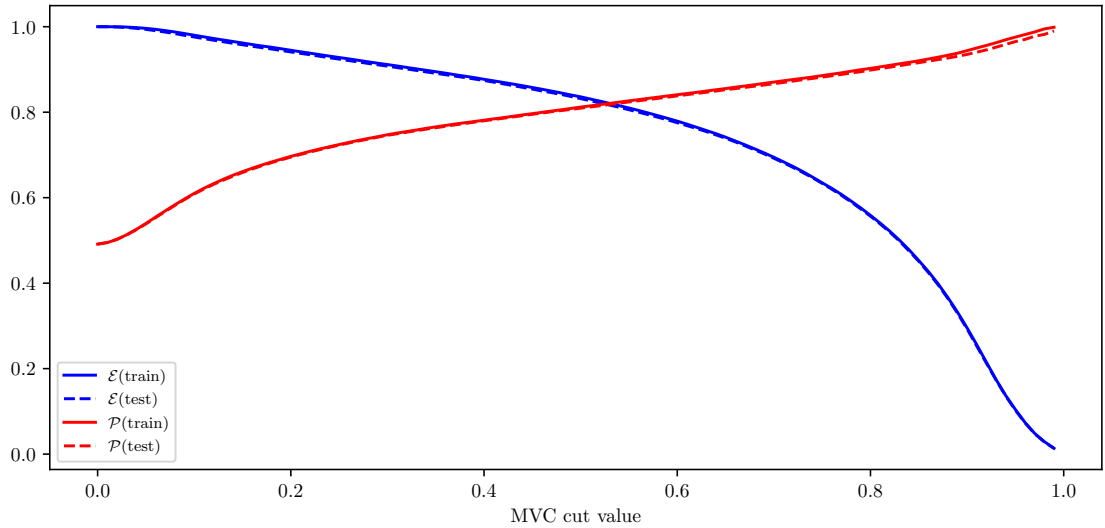


Figure A.7.: Efficiency (\mathcal{E}) and purity (\mathcal{P}) of the MVA classifier output for γ candidates training on the train (solid) and test (dashed) samples.

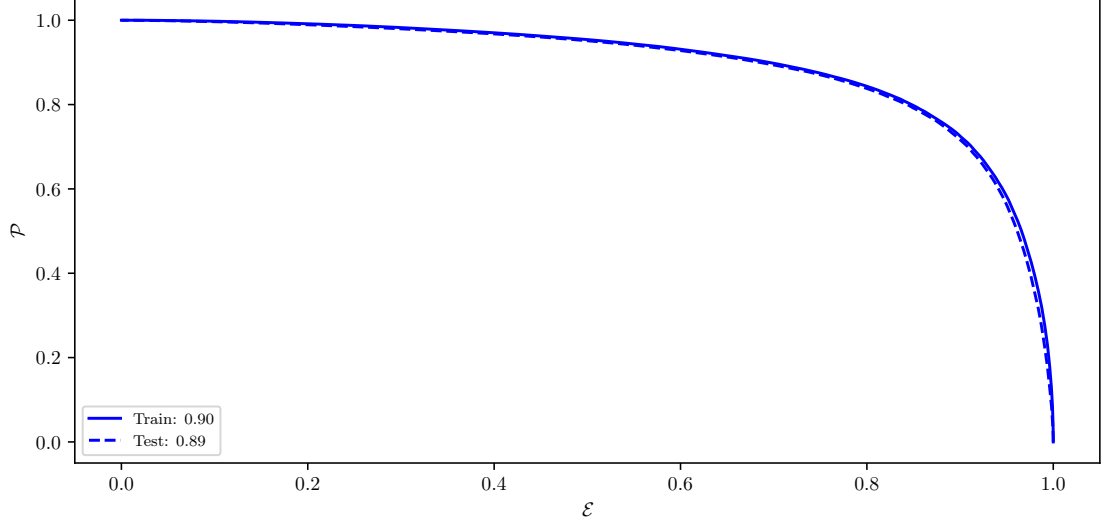


Figure A.8.: ROC curves of the MVA classifier output for γ candidates training on the train (solid) and test (dashed) samples.

1773 A.3. ROE Clean-up Duplicate Pair Training

1774 A.3.1. Variable Importance

	Name	Alias	Importance
0	useCMSFrame(daughterAngleInBetween(0,1))	v_0	0.132
1	daughter(0,phi0Err)	v_1	0.082
2	useLabFrame(daughterAngleInBetween(0,1))	v_2	0.055
3	daughter(1,d0)	v_3	0.051
4	daughter(1,phi0Err)	v_4	0.051
5	daughter(0,d0)	v_5	0.050
6	daughter(1,nCDCHits)	v_6	0.040
7	daughter(1,d0Err)	v_7	0.037
8	daughter(0,nCDCHits)	v_8	0.034
9	daughter(1,z0)	v_9	0.032
10	daughter(0,z0)	v_{10}	0.030
11	daughter(0,d0Err)	v_{11}	0.028
12	daughter(0,nSVDHits)	v_{12}	0.028
13	daughter(1,pz)	v_{13}	0.027
14	daughter(1,useCMSFrame(p))	v_{14}	0.024
15	extraInfo(decayModeID)	v_{15}	0.023
16	daughter(0,pz)	v_{16}	0.020
17	daughter(1,nSVDHits)	v_{17}	0.020
18	daughter(0,pValue)	v_{18}	0.020
19	daughter(1,tanlambda)	v_{19}	0.018
20	daughter(1,pValue)	v_{20}	0.018
21	daughter(0,tanlambda)	v_{21}	0.017
22	daughter(0,phi0)	v_{22}	0.016

23	daughter(1,phi0)	v_{23}	0.016
24	daughter(0,useCMSFrame(p))	v_{24}	0.015
25	daughter(0,z0Err)	v_{25}	0.014
26	daughter(1,omega)	v_{26}	0.013
27	daughter(0,omega)	v_{27}	0.013
28	daughter(1,z0Err)	v_{28}	0.012
29	daughter(0,pt)	v_{29}	0.011
30	daughter(0,omegaErr)	v_{30}	0.011
31	daughter(1,omegaErr)	v_{31}	0.010
32	daughter(1,pt)	v_{32}	0.009
33	daughter(0,tanlambdaErr)	v_{33}	0.009
34	daughter(1,tanlambdaErr)	v_{34}	0.009
35	useRestFrame(daughterAngleInBetween(0,1))	v_{35}	0.003
36	daughter(1,charge)	v_{36}	0.000
37	daughter(0,charge)	v_{37}	0.000

Table A.3.: Variable names, aliases and importance in the scope of duplicate track pair MVA training for ROE clean-up.

A.3.2. Variable Distributions

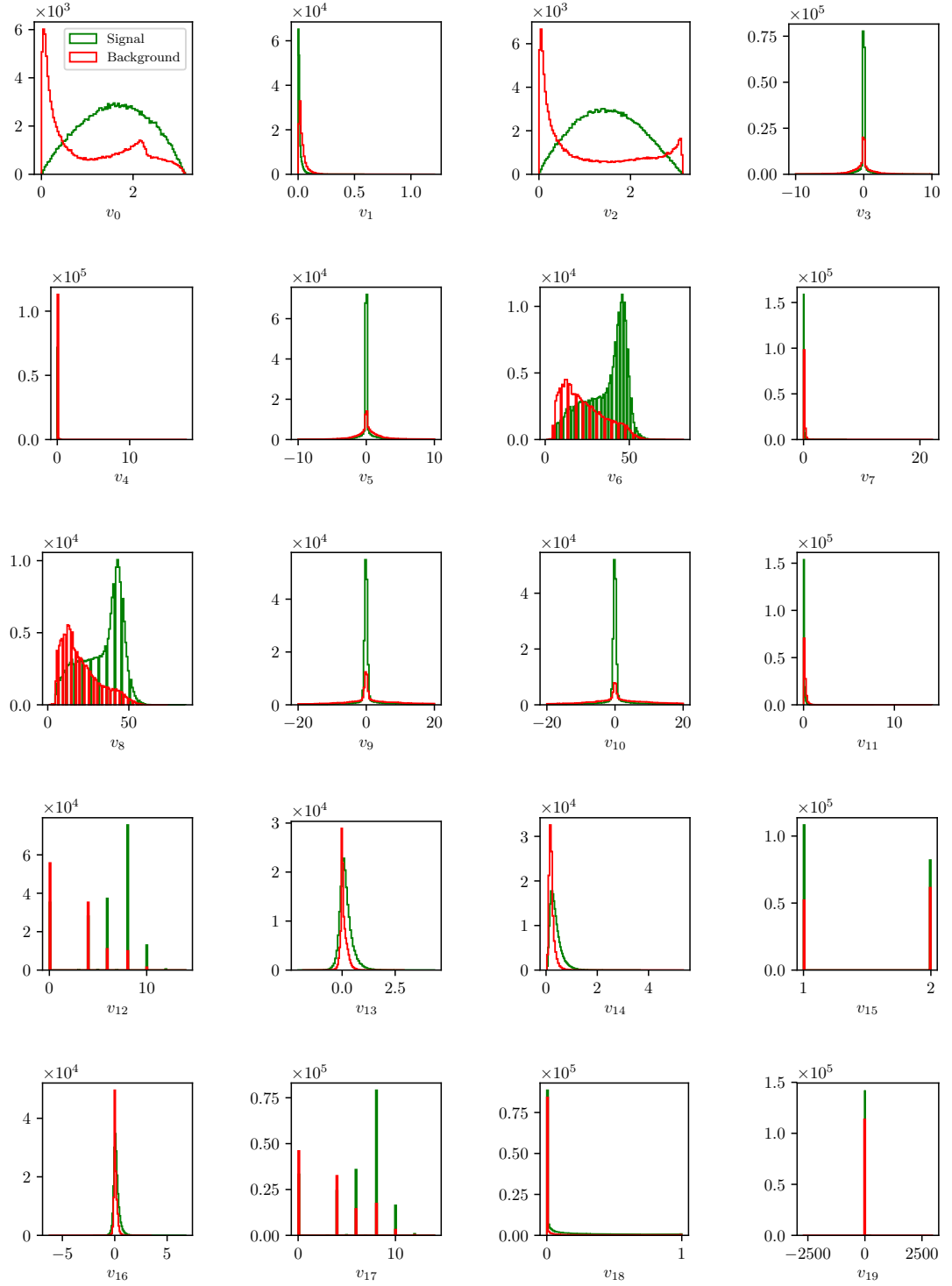


Figure A.9.: Feature distributions for MVA training of duplicate track pair candidates in the scope of ROE clean-up.

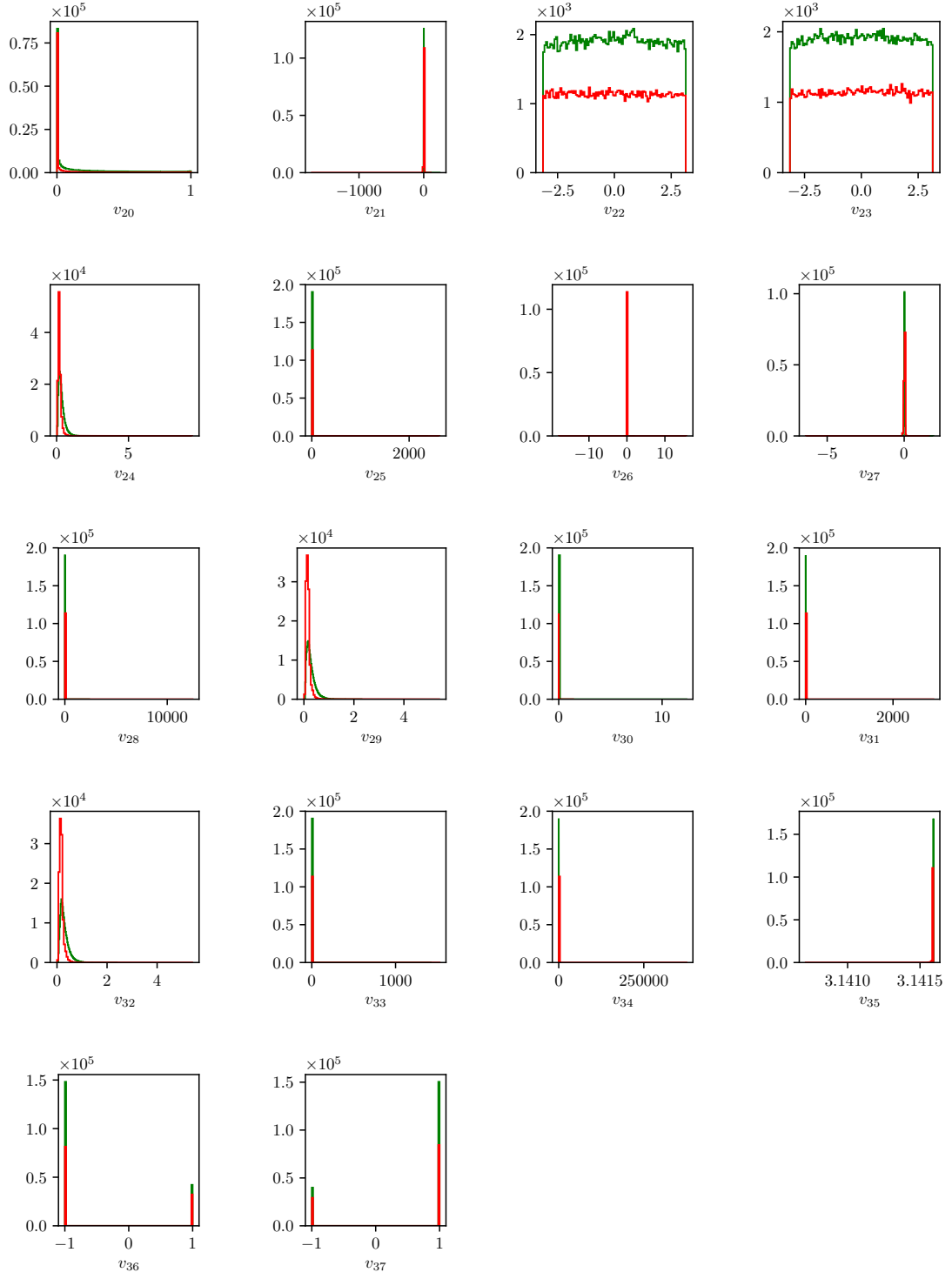


Figure A.9.: Feature distributions for MVA training of duplicate track pair candidates in the scope of ROE clean-up.

A.3.3. Hyper-parameter Optimization

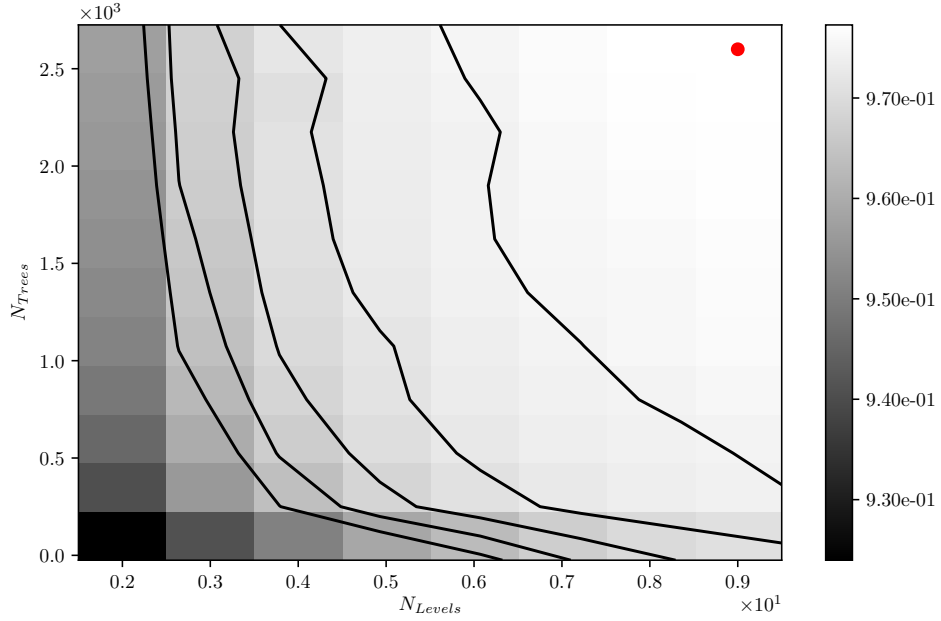


Figure A.10.: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of duplicate track pair candidates in the scope of the ROE clean-up.

A.3.4. Results

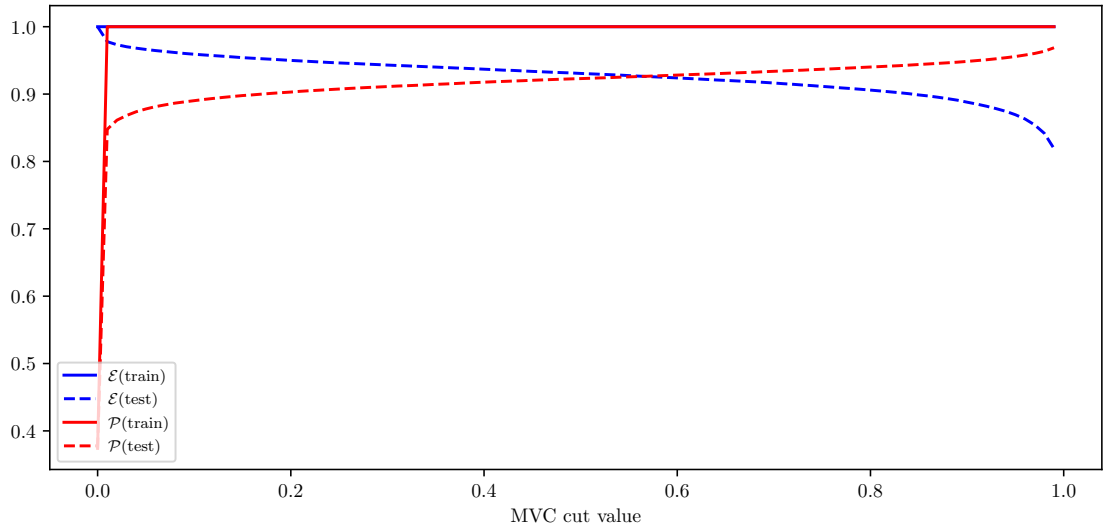


Figure A.11.: Efficiency (\mathcal{E}) and purity (\mathcal{P}) of the MVA classifier output for duplicate track pair candidates training on the train (solid) and test (dashed) samples.

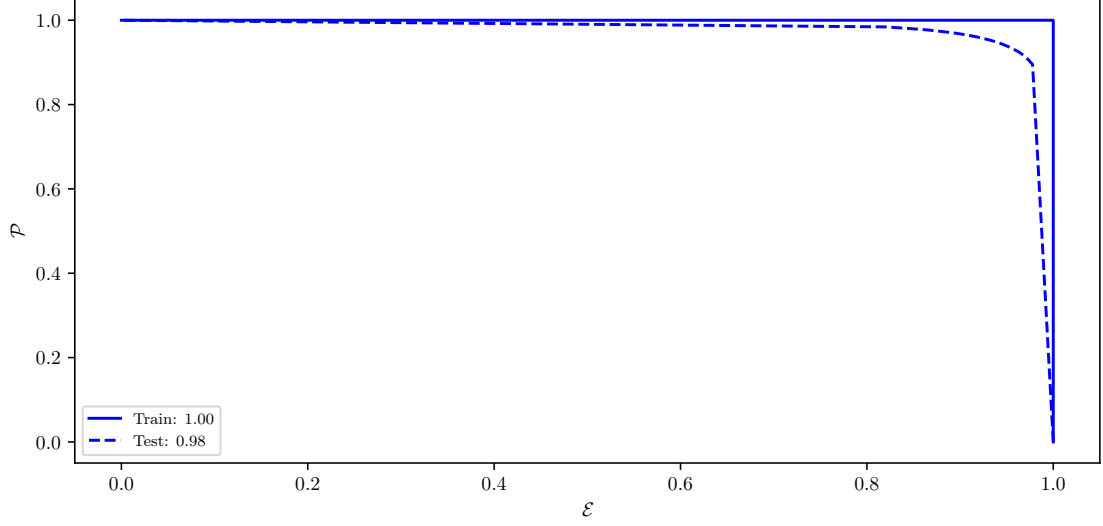


Figure A.12.: ROC curves of the MVA classifier output for duplicate track pair candidates training on the train (solid) and test (dashed) samples.

1778 A.4. ROE Clean-up Duplicate Track Training

1779 A.4.1. Variable Importance

	Name	Alias	Importance
0	extraInfo(d0Diff)	v_0	0.214
1	extraInfo(z0Diff)	v_1	0.087
2	d0	v_2	0.069
3	extraInfo(pValueDiff)	v_3	0.060
4	z0	v_4	0.058
5	phi0Err	v_5	0.056
6	extraInfo(pzDiff)	v_6	0.055
7	extraInfo(ptDiff)	v_7	0.045
8	z0Err	v_8	0.043
9	extraInfo(nCDCHitsDiff)	v_9	0.037
10	extraInfo(nSVDHitsDiff)	v_{10}	0.034
11	pt	v_{11}	0.032
12	d0Err	v_{12}	0.030
13	pValue	v_{13}	0.029
14	nCDCHits	v_{14}	0.028
15	nSVDHits	v_{15}	0.028
16	pz	v_{16}	0.025
17	cosTheta	v_{17}	0.024
18	phi0	v_{18}	0.023
19	useCMSFrame(p)	v_{19}	0.021

Table A.4.: Variable names, aliases and importance in the scope of duplicate track MVA training for ROE clean-up.

A.4.2. Variable Distributions

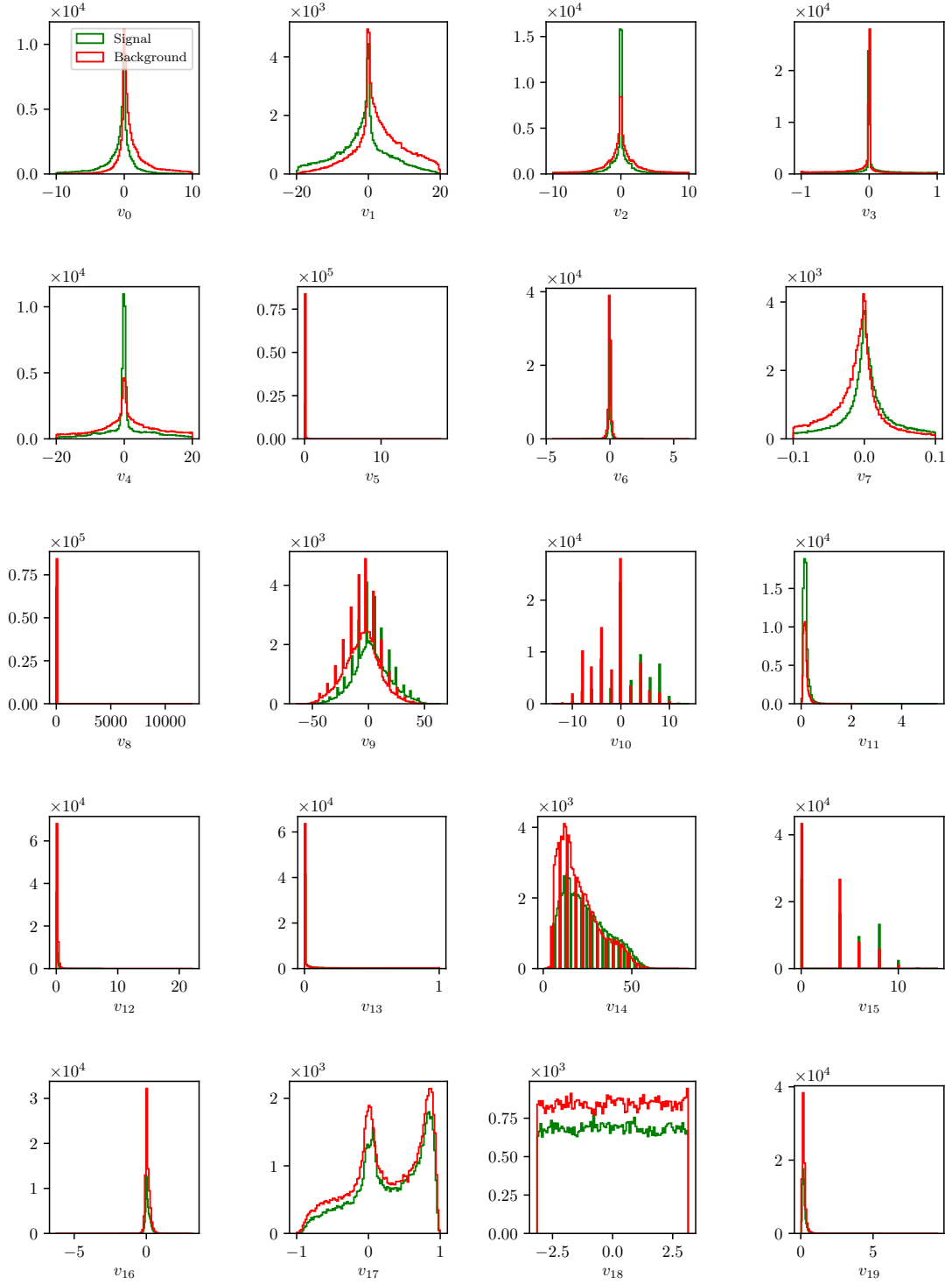


Figure A.13.: Feature distributions for MVA training of duplicate track candidates in the scope of ROE clean-up.

1781 A.4.3. Hyper-parameter Optimization

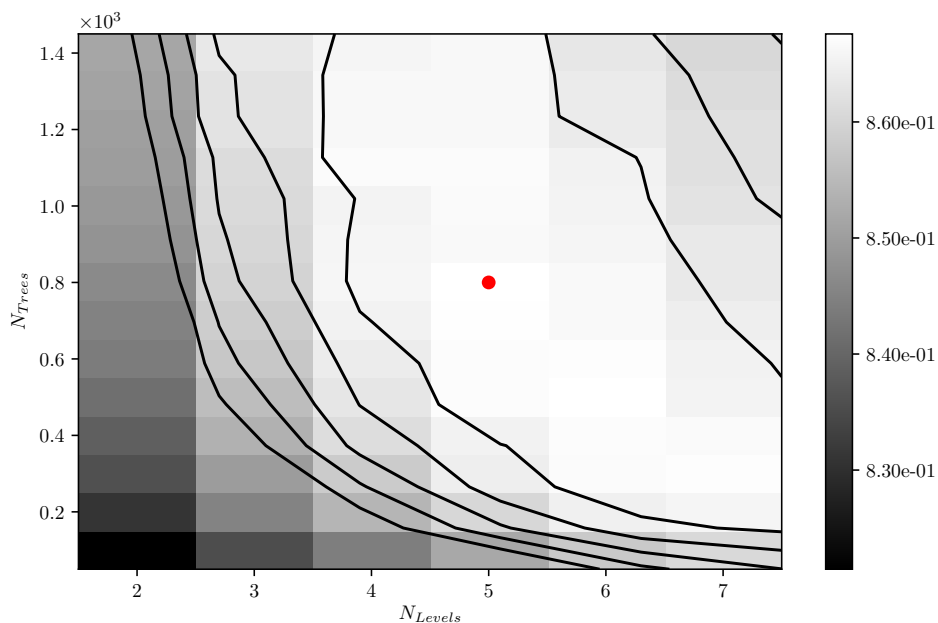


Figure A.14.: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of duplicate track candidates in the scope of the ROE clean-up.

1782 A.4.4. Results

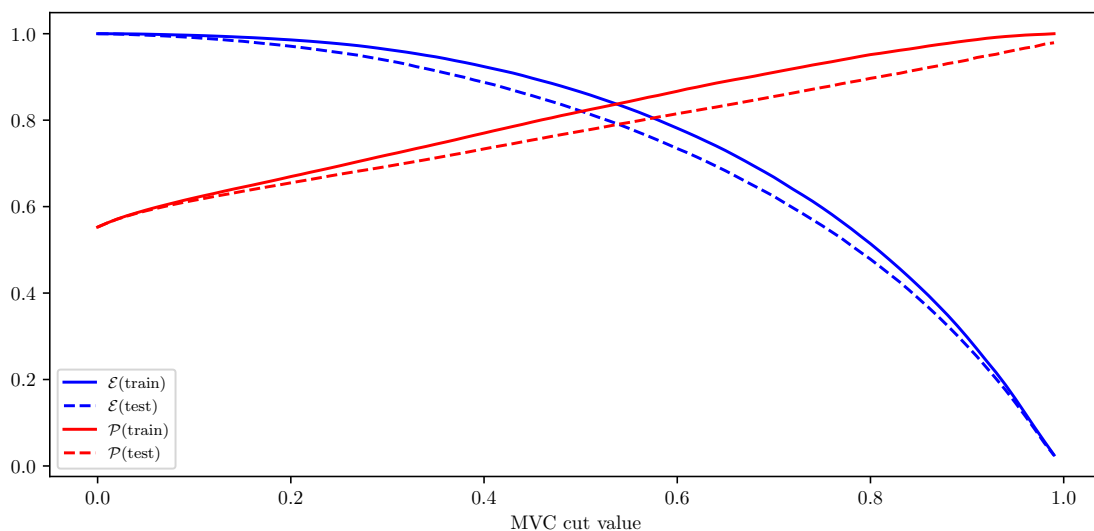


Figure A.15.: Efficiency (\mathcal{E}) and purity (\mathcal{P}) of the MVA classifier output for duplicate track candidates training on the train (solid) and test (dashed) samples.

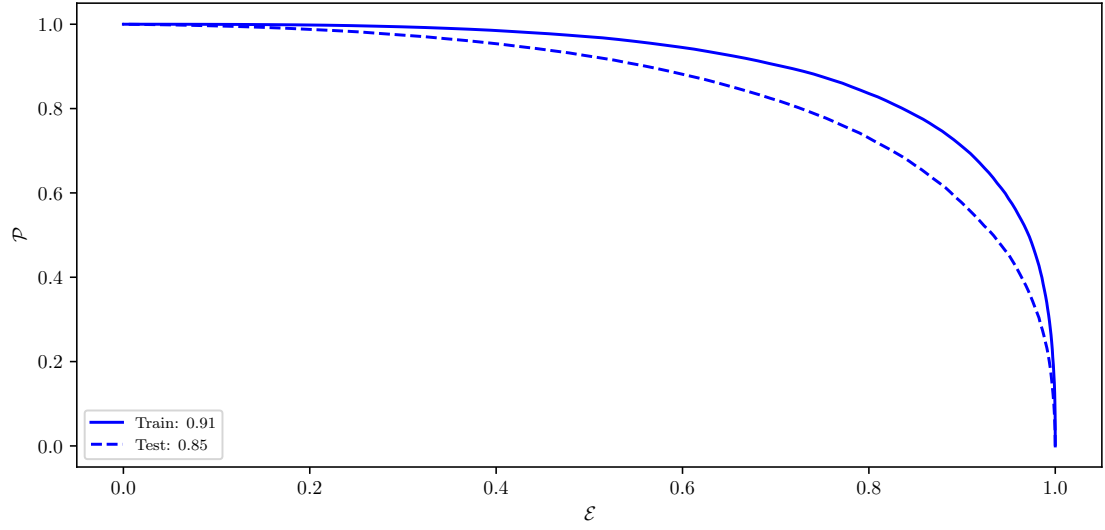


Figure A.16.: ROC curves of the MVA classifier output for duplicate track candidates training on the train (solid) and test (dashed) samples.

Appendix B.

MVA Control Plots

B.1. $q\bar{q}$ Suppression Training

B.1.1. Variable Importance

	Name	Alias	Importance
0	B_CosTBT0	v_0	0.291
1	B_hso02	v_1	0.096
2	B_ThrustB	v_2	0.096
3	B_roeFit_dz	v_3	0.075
4	B_R2	v_4	0.054
5	B_hso12	v_5	0.044
6	B_hoo2	v_6	0.032
7	B_Thrust0	v_7	0.027
8	B_qpKaon	v_8	0.024
9	B_cc2_CcROE	v_9	0.023
10	B_hoo0	v_{10}	0.019
11	B_cc3_CcROE	v_{11}	0.019
12	B_cc4_CcROE	v_{12}	0.016
13	B_CosTBz	v_{13}	0.015
14	B_hso01	v_{14}	0.015
15	B_cc1_CcROE	v_{15}	0.015
16	B_cc5_CcROE	v_{16}	0.013
17	B_cc6_CcROE	v_{17}	0.012
18	B_qpFastHadron	v_{18}	0.012
19	B_cc7_CcROE	v_{19}	0.010
20	B_cc9_CcROE	v_{20}	0.010
21	B_cc8_CcROE	v_{21}	0.010
22	B_qpMaximumPstar	v_{22}	0.008
23	B_hso10	v_{23}	0.008
24	B_hso04	v_{24}	0.007
25	B_qpLambda	v_{25}	0.006
26	B_hoo1	v_{26}	0.006
27	B_qpKaonPion	v_{27}	0.006
28	B_hoo4	v_{28}	0.006
29	B_qpSlowPion	v_{29}	0.006
30	B_hso03	v_{30}	0.005
31	B_hso14	v_{31}	0.004

32	B_qpFSC	v_{32}	0.004
33	B_hoo3	v_{33}	0.004

Table B.1.: Variable names, aliases and importance in the scope of $q\bar{q}$ suppression MVA training.

B.1.2. Variable Distributions

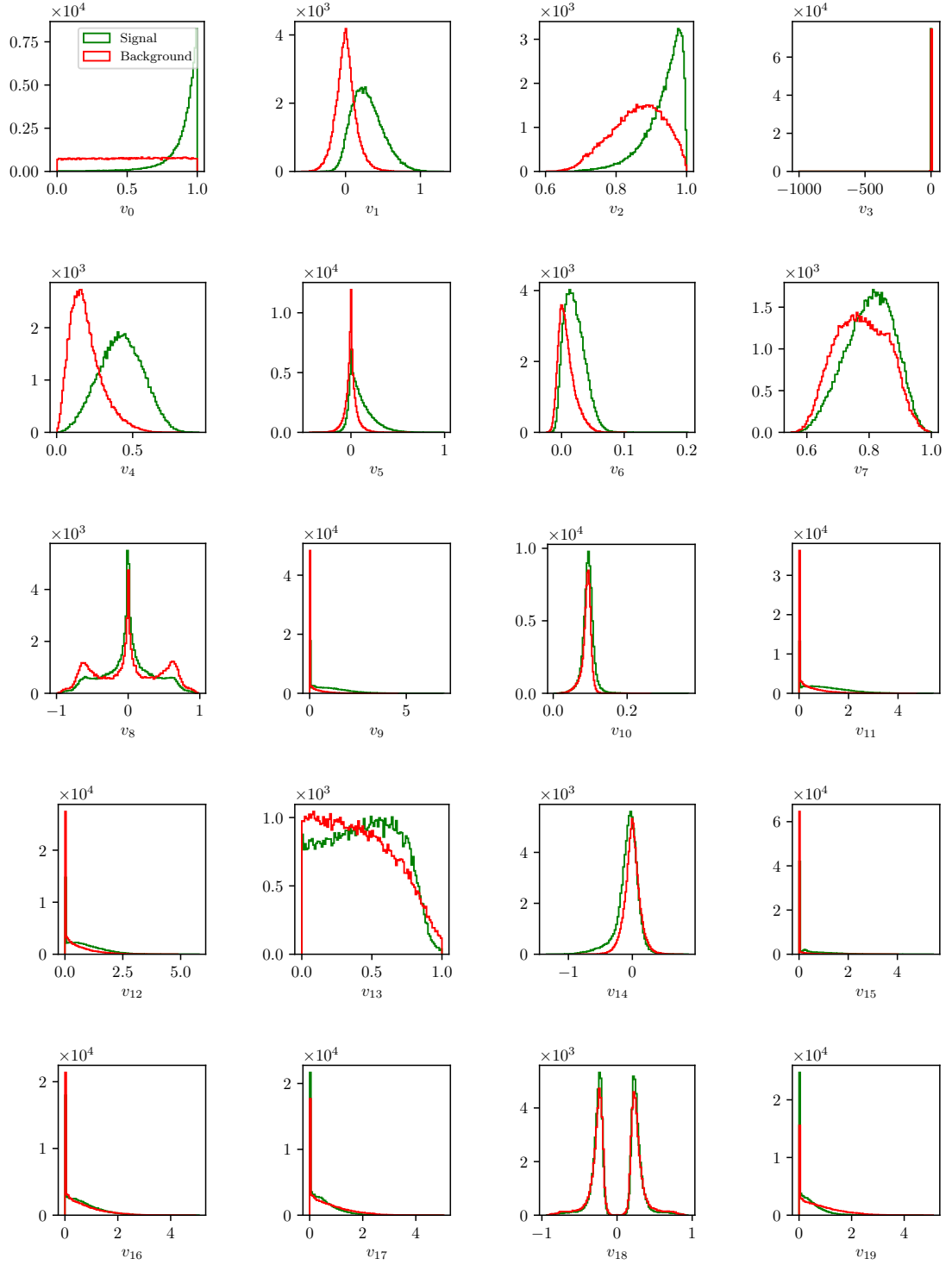


Figure B.1.: Feature distributions for MVA training of $q\bar{q}$ background suppression.

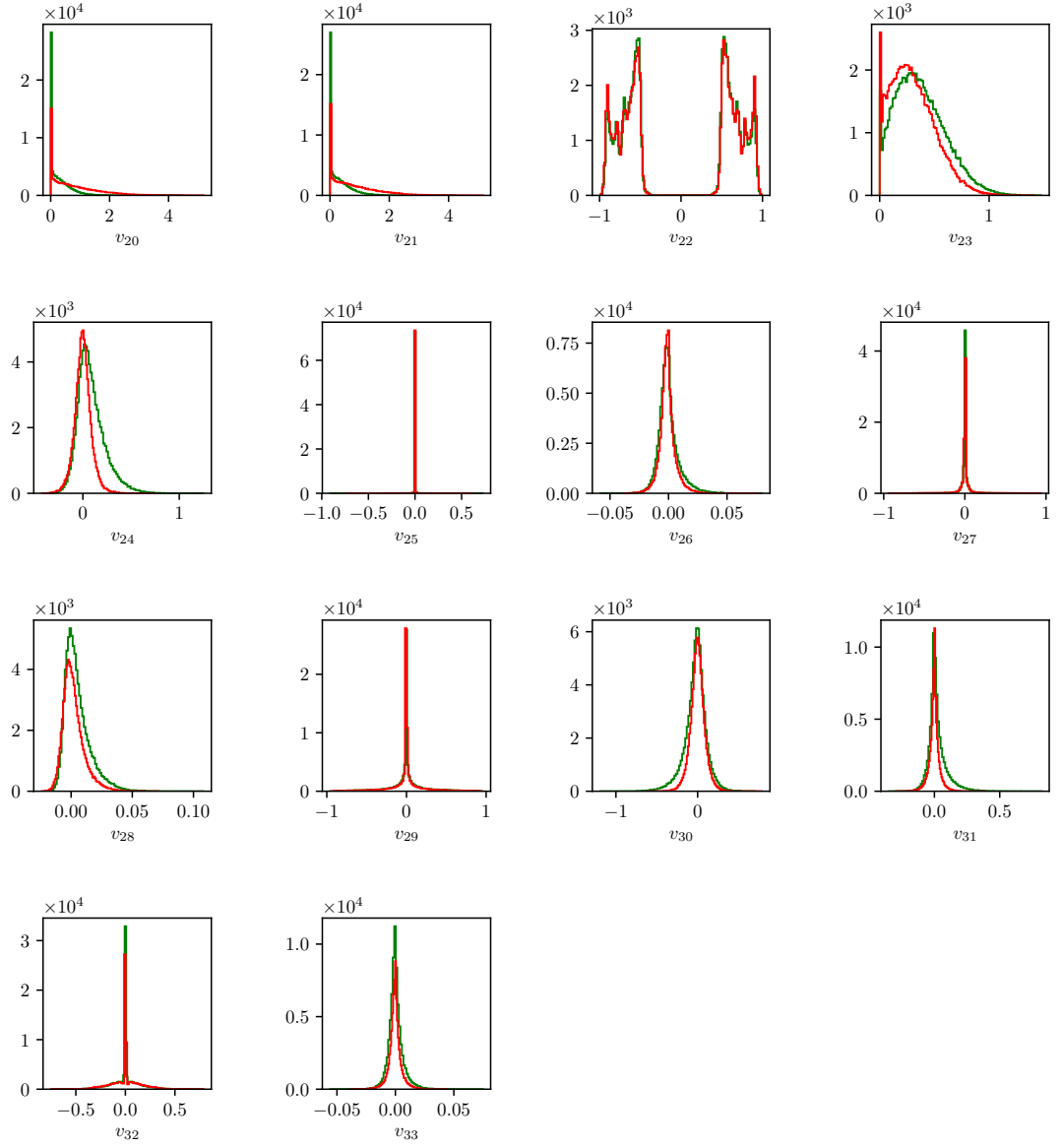


Figure B.1.: Feature distributions for MVA training of $q\bar{q}$ background suppression.

B.1.3. Hyper-parameter Optimization

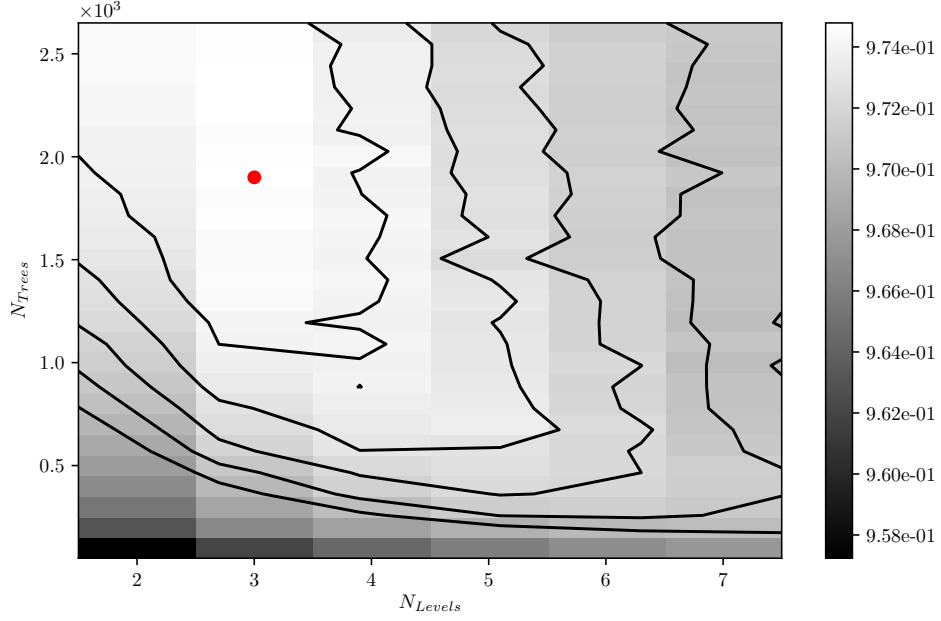


Figure B.2.: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of $q\bar{q}$ background suppression.

B.1.4. Results

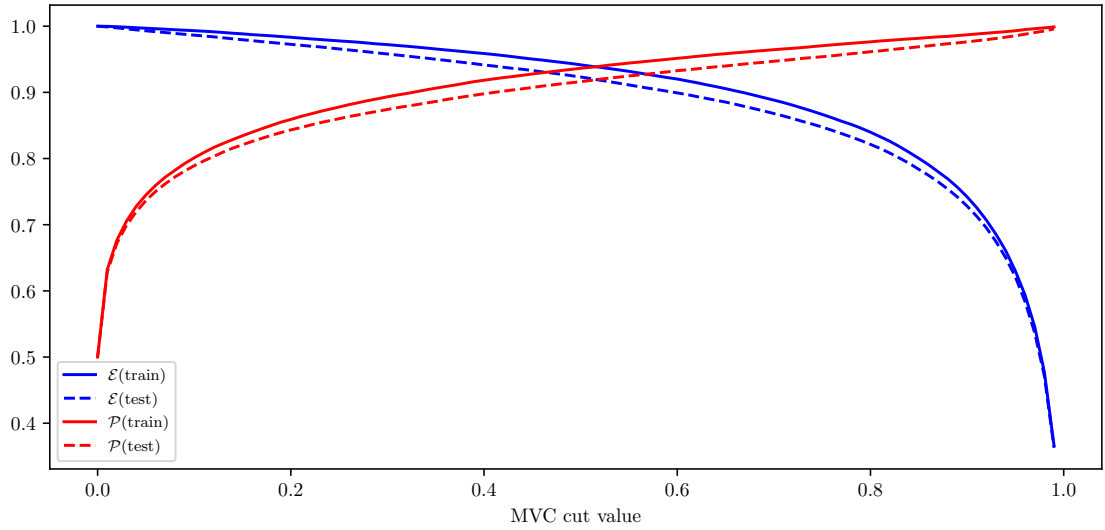


Figure B.3.: Efficiency (\mathcal{E}) and purity (\mathcal{P}) of the MVA classifier output for $q\bar{q}$ background suppression training on the train (solid) and test (dashed) samples.

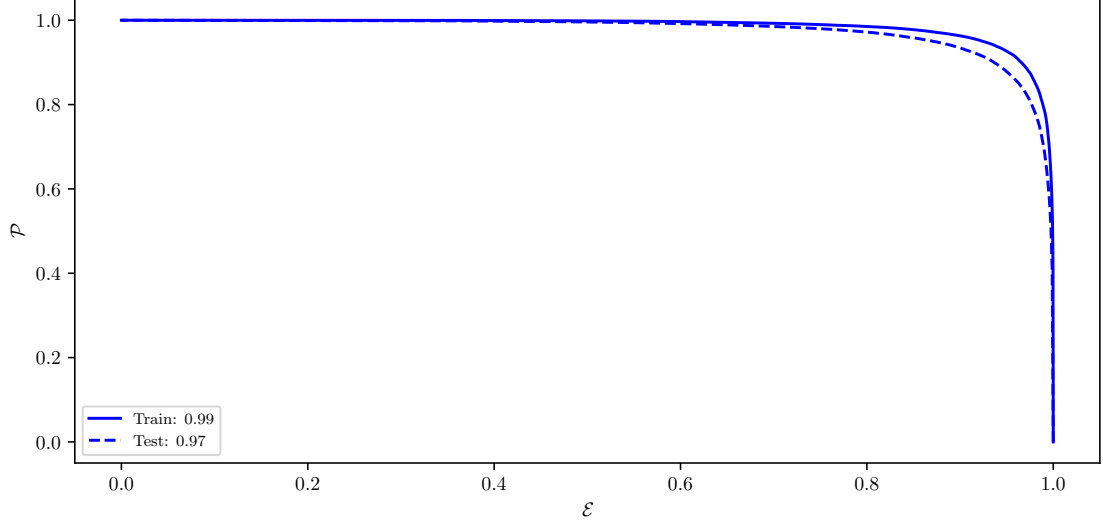


Figure B.4.: ROC curves of the MVA classifier output for $q\bar{q}$ background suppression training on the train (solid) and test (dashed) samples.

1790 B.2. Standard $B\bar{B}$ Suppression Training

1791 B.2.1. Variable Importance

	Name	Alias	Importance
0	B_cosMomVtxKKlnu	v_0	0.372
1	B_ROE_PThetacms0	v_1	0.096
2	B_nROETrk0	v_2	0.079
3	B_K1FT	v_3	0.063
4	B_cosBY	v_4	0.051
5	B_roeFit_dz	v_5	0.047
6	B_xiZ0	v_6	0.043
7	B_cosMomVtx	v_7	0.038
8	B_chiProb	v_8	0.031
9	B_nKaonInROE	v_9	0.028
10	B_missM2Veto1	v_{10}	0.026
11	B_missM2Veto2	v_{11}	0.021
12	B_nROEDistTrk	v_{12}	0.018
13	B_cosMomVtxKK	v_{13}	0.018
14	B_K0FT	v_{14}	0.017
15	B_QVeto1	v_{15}	0.016
16	B_missM20	v_{16}	0.015
17	B_TagVPvalue	v_{17}	0.012
18	B_QVeto2	v_{18}	0.010

Table B.2.: Variable names, aliases and importance in the scope of $B\bar{B}$ background suppression.

B.2.2. Variable Distributions

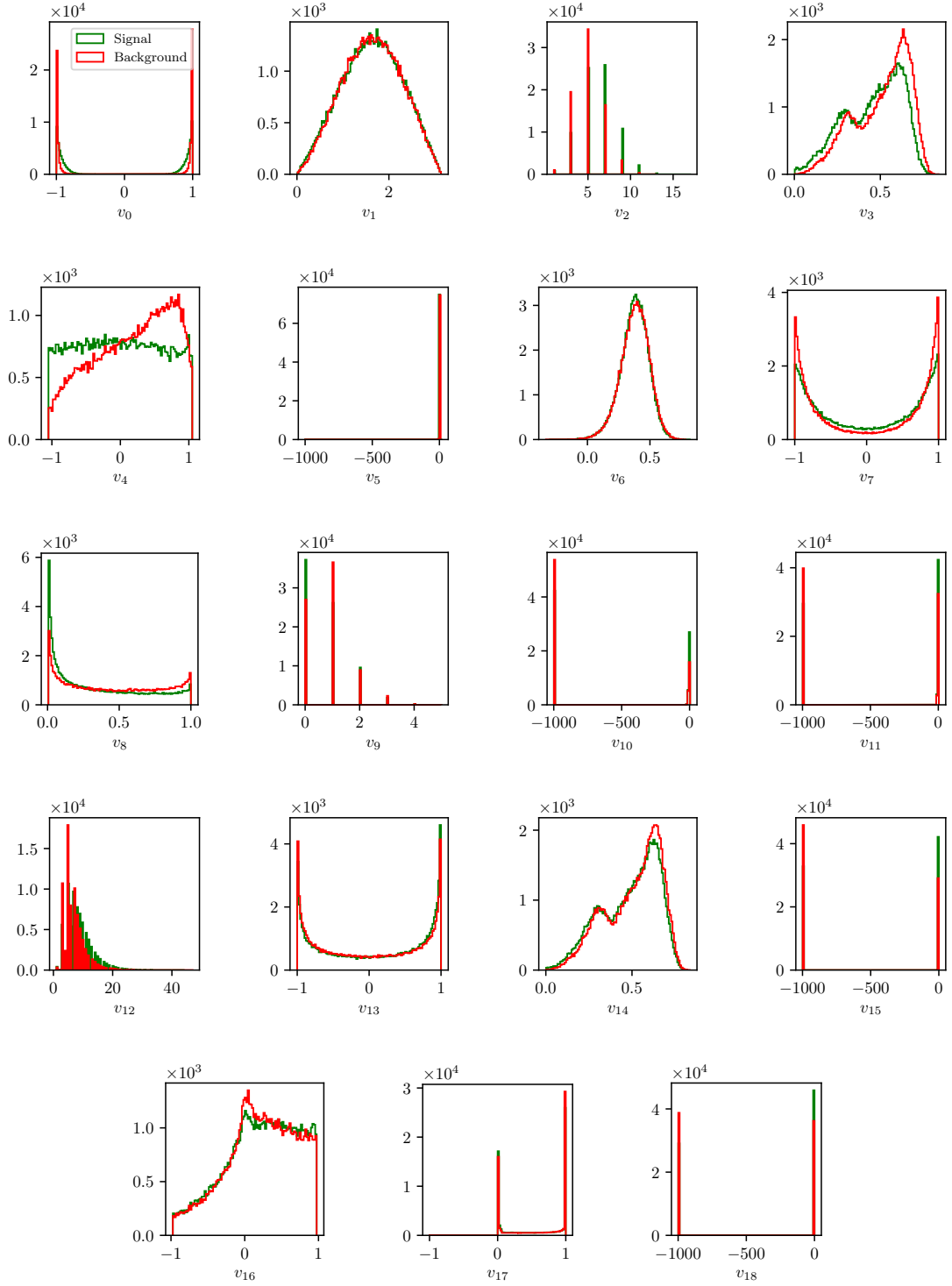


Figure B.5.: Feature distributions for MVA training of $B\bar{B}$ background suppression.

B.2.3. Hyper-parameter Optimization

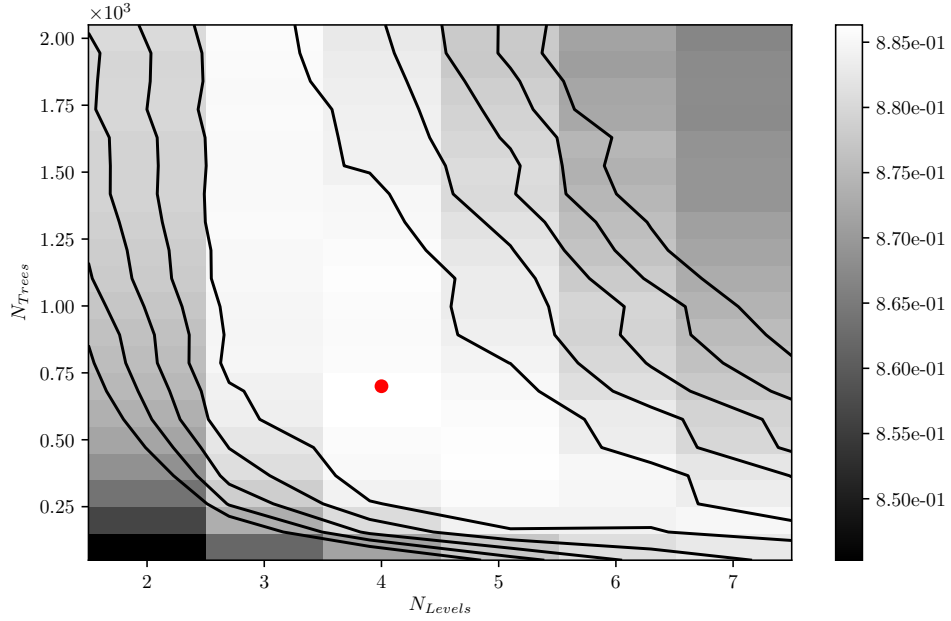


Figure B.6.: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of $B\bar{B}$ background suppression.

B.2.4. Results

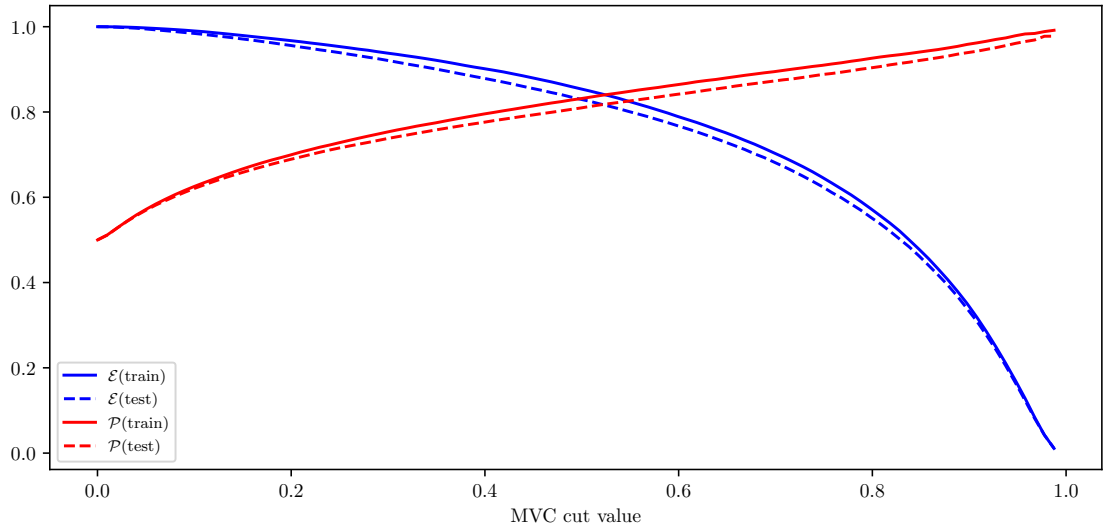


Figure B.7.: Efficiency (\mathcal{E}) and purity (\mathcal{P}) of the MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.

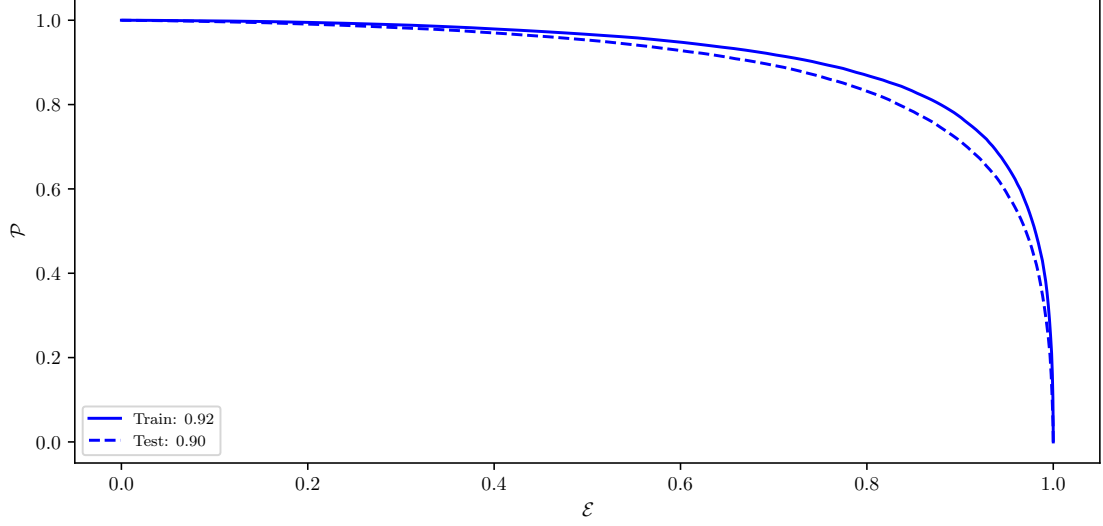


Figure B.8.: ROC curves of the MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.

1795 **B.3. Uniformity Boosted $B\bar{B}$ Suppression** 1796 **Training**

1797 **B.3.1. Hyper-parameter Optimization**

1798 Hyper-parameters were not optimized due to the large CPU time consumption of
1799 the algorithm. The following set up of the hyper-parameters was chosen

- 1800 • nTrees: 300
- 1801 • nLevels: 4

B.3.2. Results

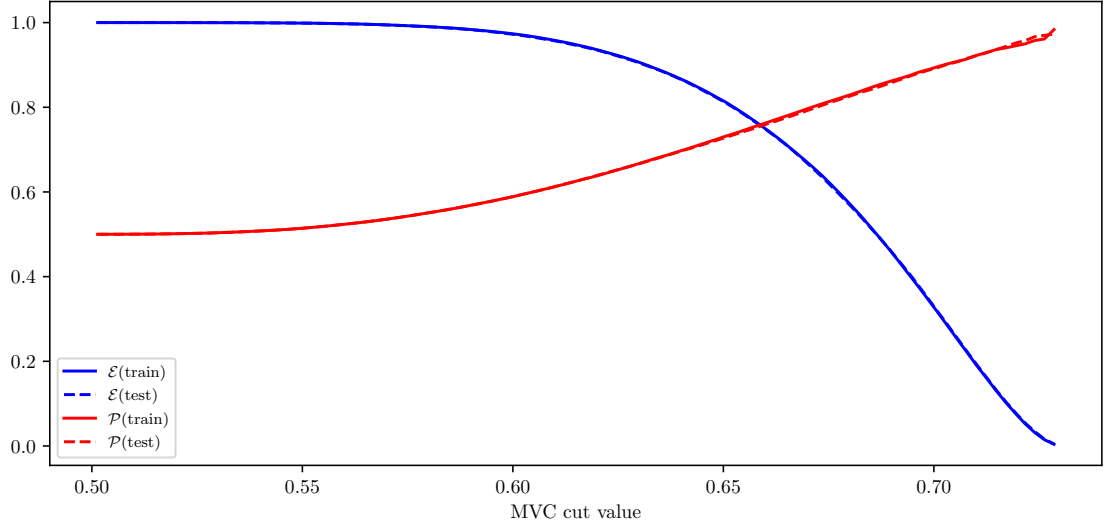


Figure B.9.: Efficiency (\mathcal{E}) and purity (\mathcal{P}) of the uniformity boosted MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.

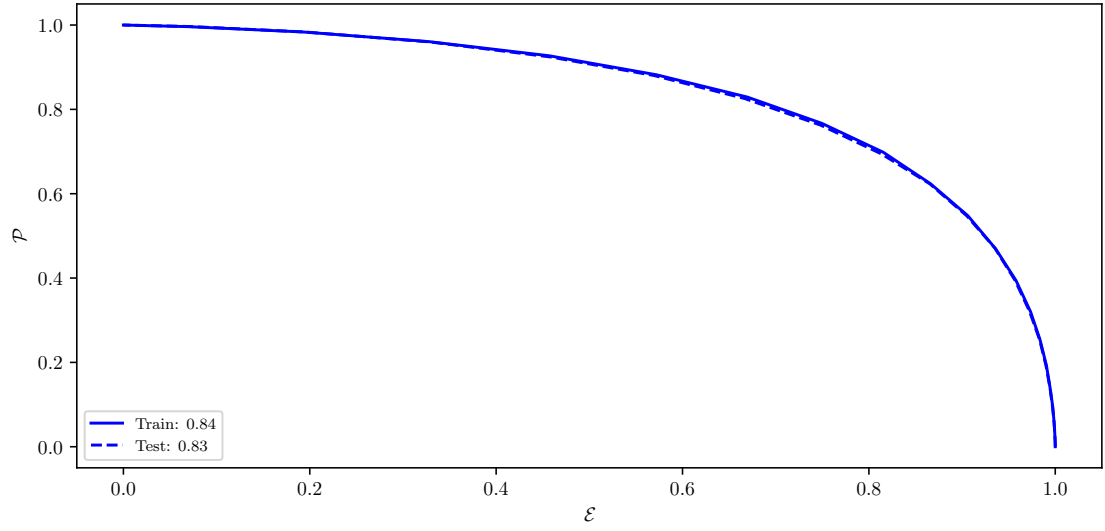


Figure B.10.: ROC curves of the uniformity boosted MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.

Appendix C.

Other Plots

C.1. Signal Fits in m_{KK}

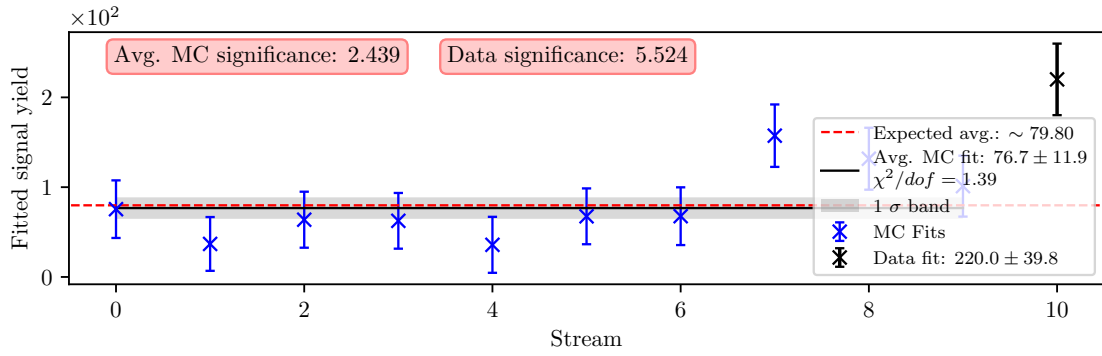


Figure C.1.: Signal fit result for the 1st m_{KK} window for MC and data in the range $0.980 < m_{KK} < 1.011$.

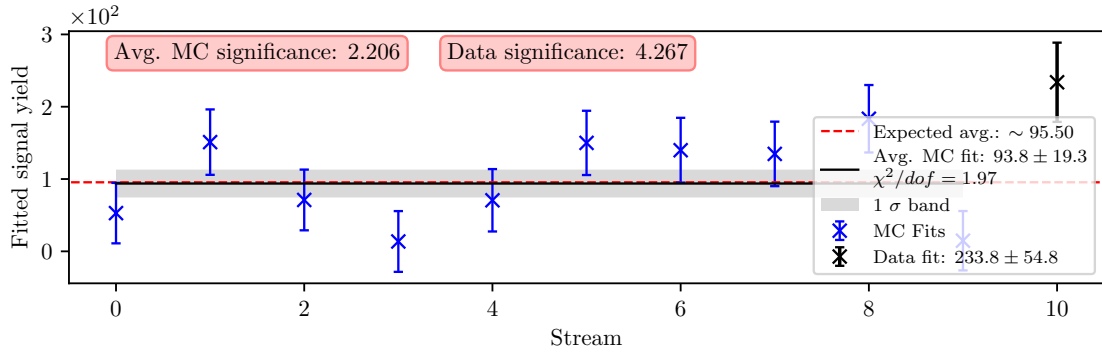


Figure C.2.: Signal fit result for the 2nd m_{KK} window for MC and data in the range $1.011 < m_{KK} < 1.027$.

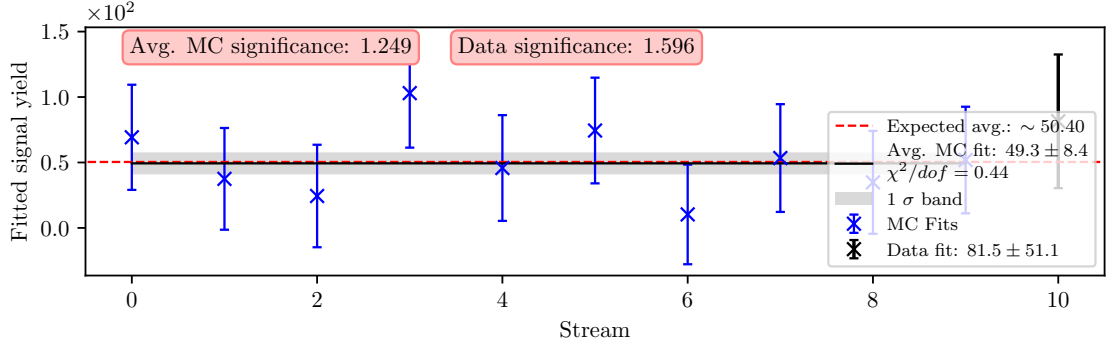


Figure C.3.: Signal fit result for the 3rd m_{KK} window for MC and data in the range $1.027 < m_{KK} < 1.187$.

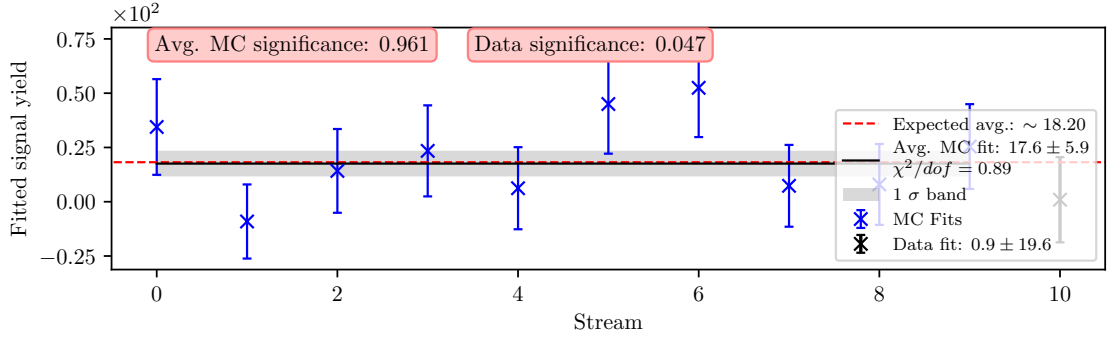


Figure C.4.: Signal fit result for the 4th m_{KK} window for MC and data in the range $1.187 < m_{KK} < 1.342$.

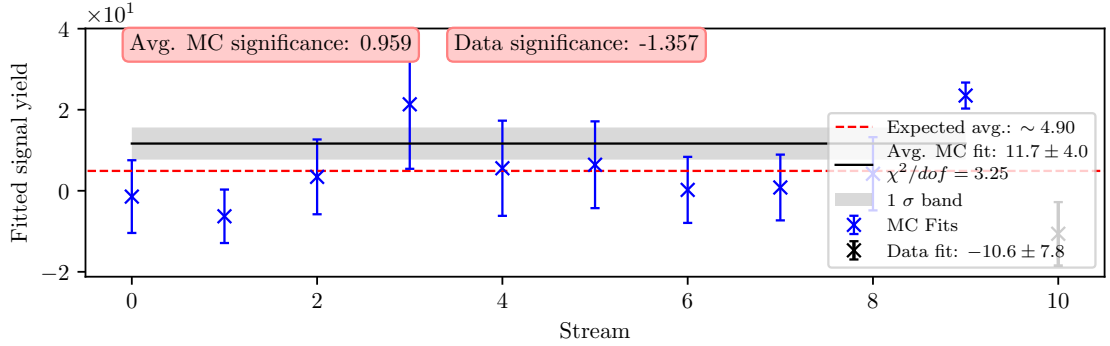


Figure C.5.: Signal fit result for the 5th m_{KK} window for MC and data in the range $1.342 < m_{KK} < 1.497$.

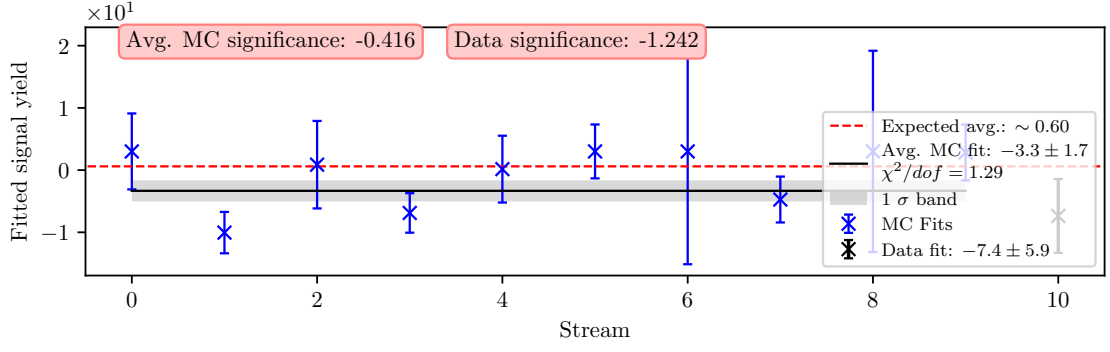


Figure C.6.: Signal fit result for the 6th m_{KK} window for MC and data in the range $1.497 < m_{KK} < 1.647$.

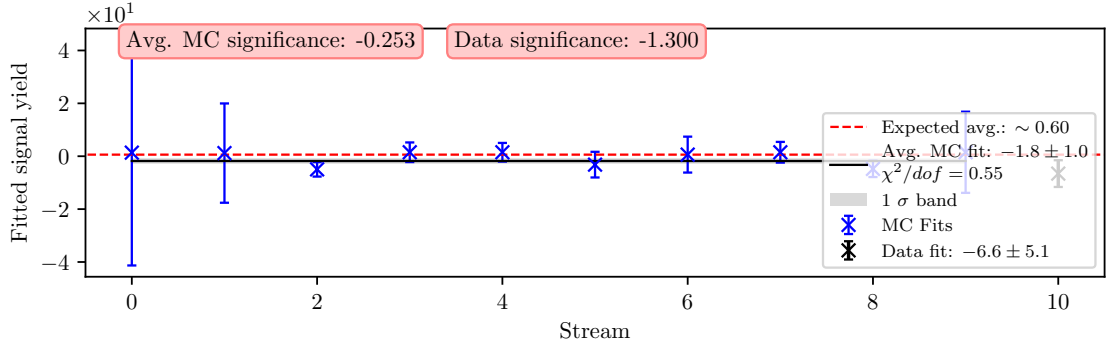


Figure C.7.: Signal fit result for the 7th m_{KK} window for MC and data in the range $1.647 < m_{KK} < 1.850$.

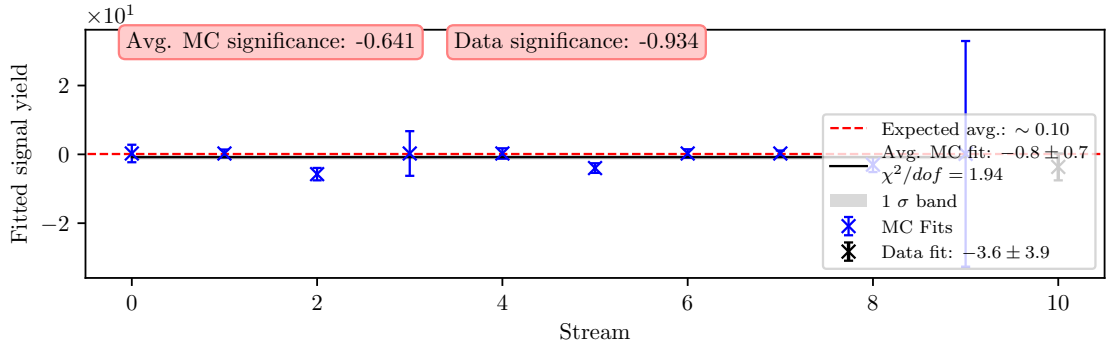


Figure C.8.: Signal fit result for the 8th m_{KK} window for MC and data in the range $1.850 < m_{KK} < 1.880$.

C.2. Signal Fits in q^2

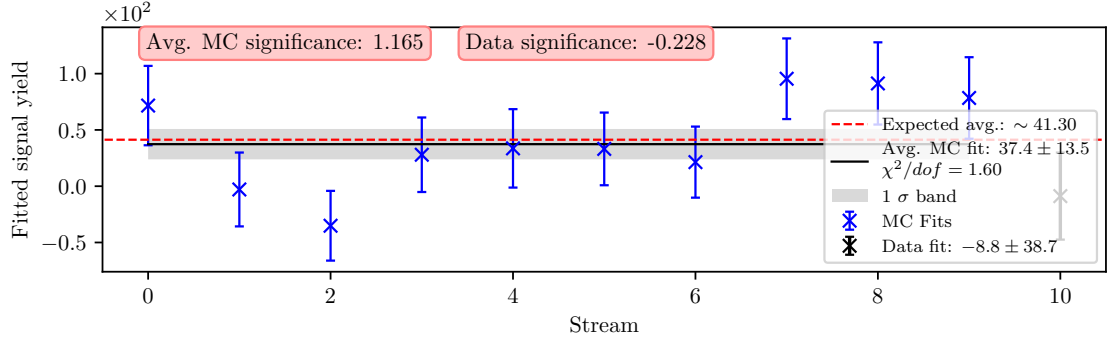


Figure C.9.: Signal fit result for the 1st q^2 window for MC and data in the range $0.000 < q^2 < 2.250$.

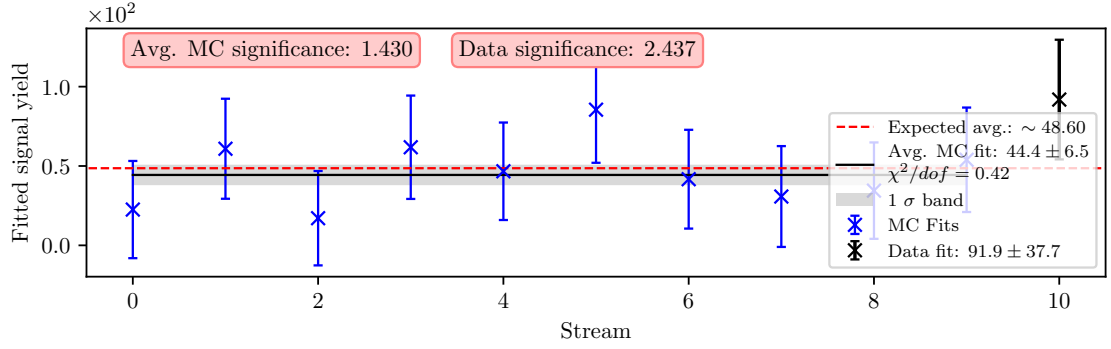


Figure C.10.: Signal fit result for the 2nd q^2 window for MC and data in the range $2.250 < q^2 < 4.500$.

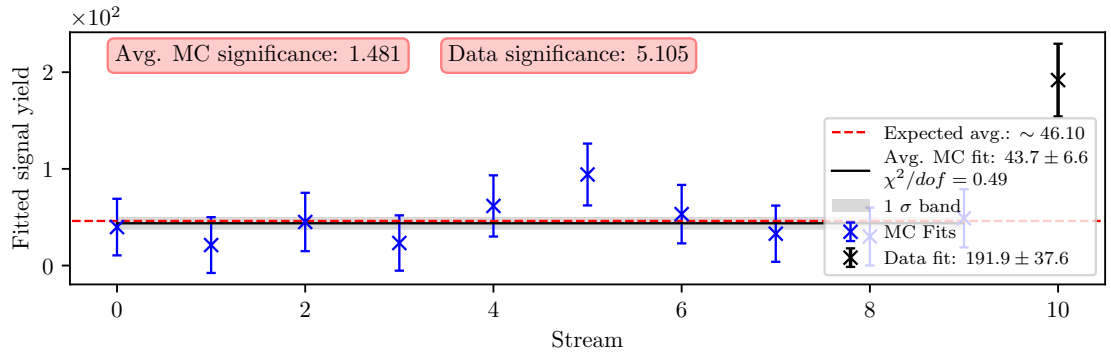


Figure C.11.: Signal fit result for the 3rd q^2 window for MC and data in the range $4.500 < q^2 < 6.750$.

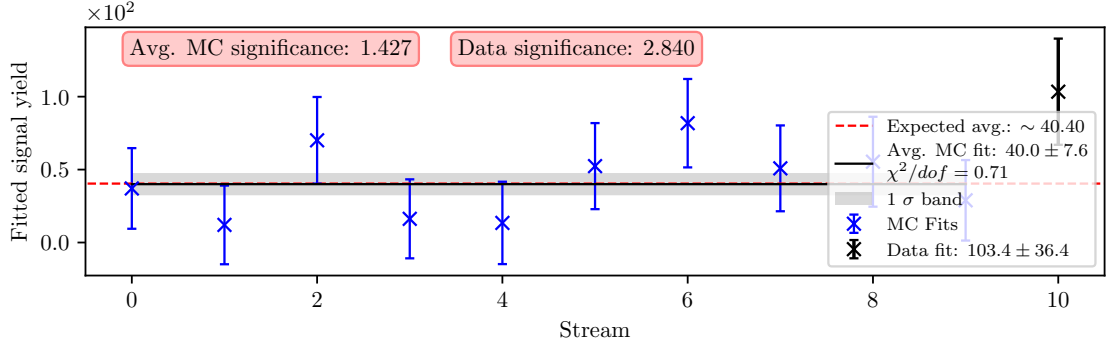


Figure C.12.: Signal fit result for the 4th q^2 window for MC and data in the range $6.750 < q^2 < 9.000$.

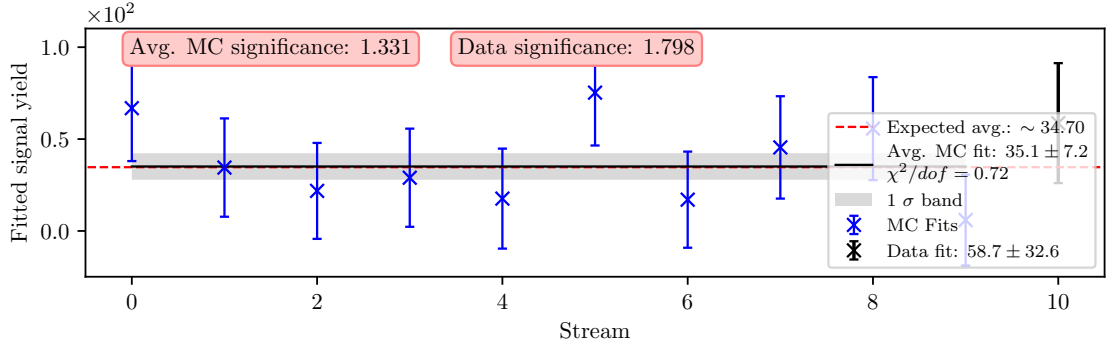


Figure C.13.: Signal fit result for the 5th q^2 window for MC and data in the range $9.000 < q^2 < 11.250$.

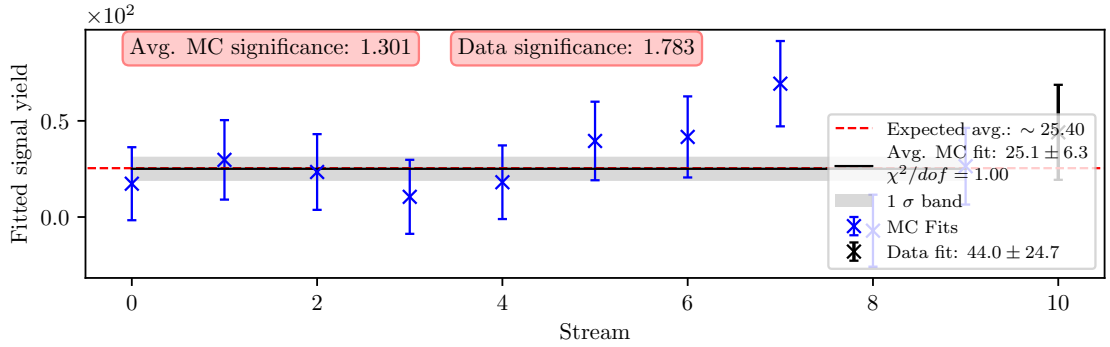


Figure C.14.: Signal fit result for the 6th q^2 window for MC and data in the range $11.250 < q^2 < 13.500$.

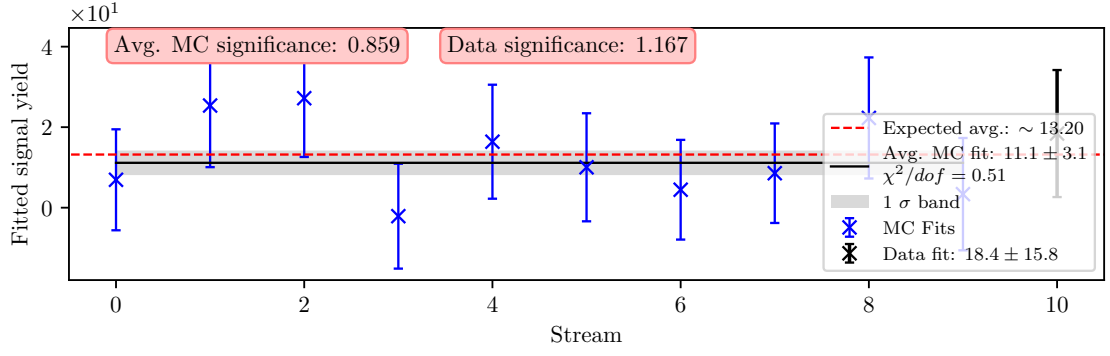


Figure C.15.: Signal fit result for the 7th q^2 window for MC and data in the range $13.500 < q^2 < 15.750$.

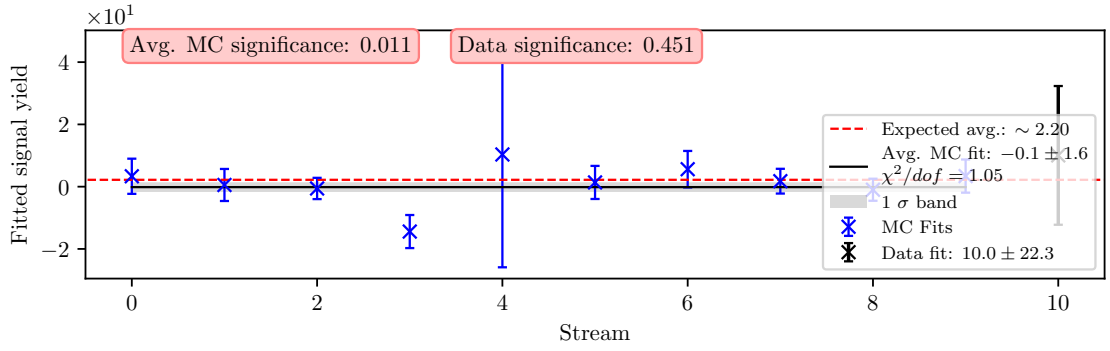


Figure C.16.: Signal fit result for the 8th q^2 window for MC and data in the range $15.750 < q^2 < 18.000$.