BELLE NOTE 1486

Measurement of the decay $B^+ \to K^+ K^- \ell^+ \nu_\ell$ with B2BII

Matic Lubej

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.

Contents

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 (gluons), $W^\pm$ (charged weak bosons) and Z^0 (neutral weak boson). The theory 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), 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 its complex and 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 secrets of the universe and how it all began.

1

QUARKS

2.2 M	1.28 G	173.1 G
u +2/3 1/2	c +2/3 1/2	t +2/3 1/2
up	charm	top

4.7 M	96 M	4.18 G
d -1/3 1/2	s -1/3 1/2	b -1/3 1/2
down	strange	bottom

LEPTONS

0.51 M	105.66 M	1.78 G
e -1 1/2	μ -1 1/2	τ -1 1/2
electron	muon	tau

< 2	< 0.19 M	< 18.2 M
ν_e 0 1/2	ν_μ 0 1/2	ν_τ 0 1/2
e neutrino	μ neutrino	τ neutrino

Mass: eV/c^2
Charge
Spin
Name

strong nuclear interaction

| 0 |
| g 0 1 |
| gluon |

| 0 |
| γ 0 1 |
| photon |

80.38 G	91.19 G
W -1 1	Z 0 1
W boson	Z boson

electromagnetic interaction

| 125.09 G |
| H 0 0 |
| higgs |

weak nuclear interaction

FERMIONS   BOSONS

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 ($C\!\!\!/P\!\!\!/$). 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 $C\!\!\!/P\!\!\!/$ is very closely related to the the weak nuclear force. Here lies our motivation for studying decays of B mesons, since they exhibit a rich spectrum of decays, many of which underway 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 of the information regarding quark transitions and transition probabilities can be merged into a form of a complex matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [1, 2]

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

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

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

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 closely connected to the free parameters of the CKM matrix. It is important to mention that all experimental measurements depend only on these four parameters, so it is possible to determine them by measuring the angles and sides of the unitarity triangle. This way the unitarity triangle offers us a unique way to test the consistency of the SM. The ultimate goal is to then join all such measurements and overconstrain the unitarity triangle to check if all the sides meet. By improving such measurements one can check whether the SM is consistent, or if there are some contributing physics processes that we do not yet understand. Such processes are commonly referred to as "new physics" (NP). The measurements of the sides and angles of the triangle are done by using different decays of which a large portion are B meson decays. Here lies another motivation for using B mesons in the analysis.
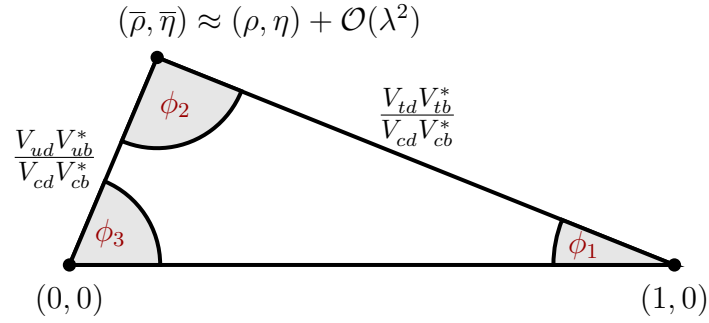
Figure 1.2: The unitarity triangle with λ, η, ρ and A (not shown) as free parameters of the CKM matrix.

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

$$B^+ \to X_u^0 \ell^+ \nu_\ell, \tag{1.3}$$

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

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

3

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

In the exclusive method one studies the decays of B mesons to a specific charmless hadronic final state, such as $B \to \pi\ell\nu$. Clean determination of the V_{ub} is possible due to precise experimental measurement 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 $B \to X_c\ell\nu$ processes, since the $b \to c$ 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 $B \to X_u\ell\nu$ 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 $B \to X_u\ell\nu$. In this case, the total decay rate for $b \to u\ell\nu$ 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 $B \to X_c\ell\nu$ background. Experimental sensitivity to V_{ub} is highest where $B \to X_c\ell\nu$ 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. Theory takes 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 $b \to c$ 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 $c \to s$ 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 $b \to c$ process.

If V_{ub} is determined with both these methods, the values can be compared. It turns out that consistency between these two results is only marginal, where the difference is at a level of 3σ. The current world averages [3] of the exclusive (from $B^0 \to \pi^-\ell^+\nu$) and inclusive (GGOU collab. [4]) are

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

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

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 $b \to c$ contributions. In this analysis we focus on the charged $B \to KK\ell\nu$ decay, which is very similar to the $B \to \pi\ell\nu$, 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 $b \to u$ 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.
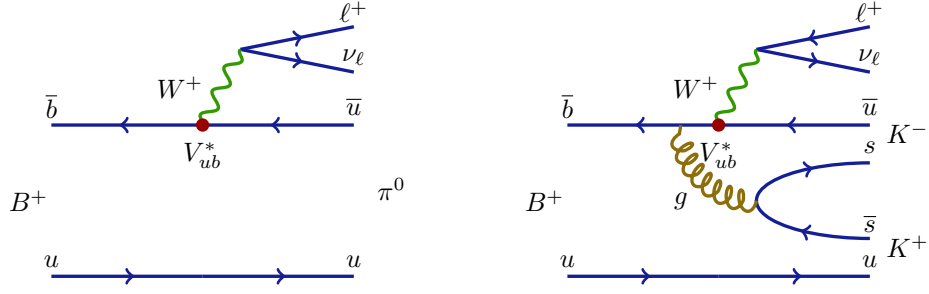
Figure 1.3: Feynman diagrams for the $B^+ \to \pi^0 \ell^+ \nu_\ell$ decay (left) and the $B^+ \to K^- K^+ \ell^+ \nu_\ell$ decay (right).

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

Chapter 2

Data and Monte-Carlo samples

The Belle detector acquired a dataset of about $L_0 \approx 710$ fb^{-1} of integrated luminosity in it's lifetime, which corresponds to about 771×10^6 $B\bar{B}$ 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 it's 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 $B\bar{B}$ (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 $1/10 \times L_0$,

  - $B \to X_u \ell \nu$ (denoted as `ulnu`), not included in previous MC samples, equal to an amount of $20 \times L_0$,

209   – other rare B meson decays (denoted as `rare`), not included in previous MC
210   samples, equal to an amount of $50 \times L_0$.

2.1 Signal MC production

The signal MC sample of $B^+ \to K^+ K^- \ell \nu_\ell$ and the charge conjugated B^- decays was produced using the `mcproduzh` 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

- $B^+ \to a_{00} \ell^+ \nu_\ell$,

- $B^+ \to a_{20} \ell^+ \nu_\ell$,

- $B^+ \to f_2 \ell^+ \nu_\ell$,

- $B^+ \to f_0 \ell^+ \nu_\ell$,

- $B^+ \to X_u^0 \ell^+ \nu_\ell$,

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 [5]. 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 structurer, 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 $B \to KK\ell\nu$ as well as $B \to KKX\ell\nu$, 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 28.14 %, 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 decrease in efficiency due to reconstruction effects.

Channel	Ratio [%]
K^+K^-	28.14
$K^+K^-\pi^0$	8.94
$K^+\bar{K}^0\pi^-$	8.71
$K^0K^-\pi^+$	8.70
$K^+K^-\pi^+\pi^-$	4.15
$K^0\bar{K}^0$	3.32
$K^0\bar{K}^0\pi^0$	3.26
$K^+K^-\rho^0$	1.93
$K^+\bar{K}^0\rho^-$	1.84
$K^0K^-\rho^+$	1.83

Channel	Ratio
$K^+K^-\pi^0\pi^0$	0.86
$K^+K^-\pi^+\rho^-$	0.69
$K^+K^-\rho^+\pi^-$	0.68
$K^0\bar{K}^0\rho^0$	0.00
$K\bar{K}$ pair with η	7.08
$K\bar{K}$ pair with ω	5.33
Other	14.53

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

We generate about 1.3×10^9 events of the form $B \to X_u\ell\nu$, 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 $B \to X_u\ell\nu$ MC sample. This amounts to a total of about 9.37×10^6 generated signal events, and to a branching ratio

$$\mathcal{B}\left(B^+ \to K^+K^-\ell^+\nu_\ell\right)_{MC} = 1.53 \times 10^{-5}, \tag{2.1}$$

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

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

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

B2BII conversion

The predecessor of the Belle II experiment was the Belle experiment, which finished its data taking run of 10 years at 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 was born from the ashes, where even some of the old detector components were reused. This moved focus from Belle analyses and Belle Analysis Framework (BASF) to the construction of the Belle II detector and the development of Belle II Analysis Framework (BASF2), which was written completely from scratch, making the BASF2 software incompatible with Belle data. This resulted in gradual loss of knowledge on the maintenance and operation of the BASF software. The construction of the Belle II detector today is still an ongoing process, although first collisions were already recorded in April 2018. By the year 2025 it is foreseen that Belle II will have recorded about 50 ab^{-1} of data, which is about 50 times more than in case of Belle.

However, this is still in the distant future and in principle we need to wait for data in order to start doing analyses. On the other hand, even though the Belle experiment finished collecting data, the data itself is still relevant and has the potential for interesting physics analyses today. In the Belle II Collaboration, a task force was created in order to convert Belle data into Belle II format (B2BII). 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

11

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 is done in the last module. For more information see [6].

3.1 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. 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.
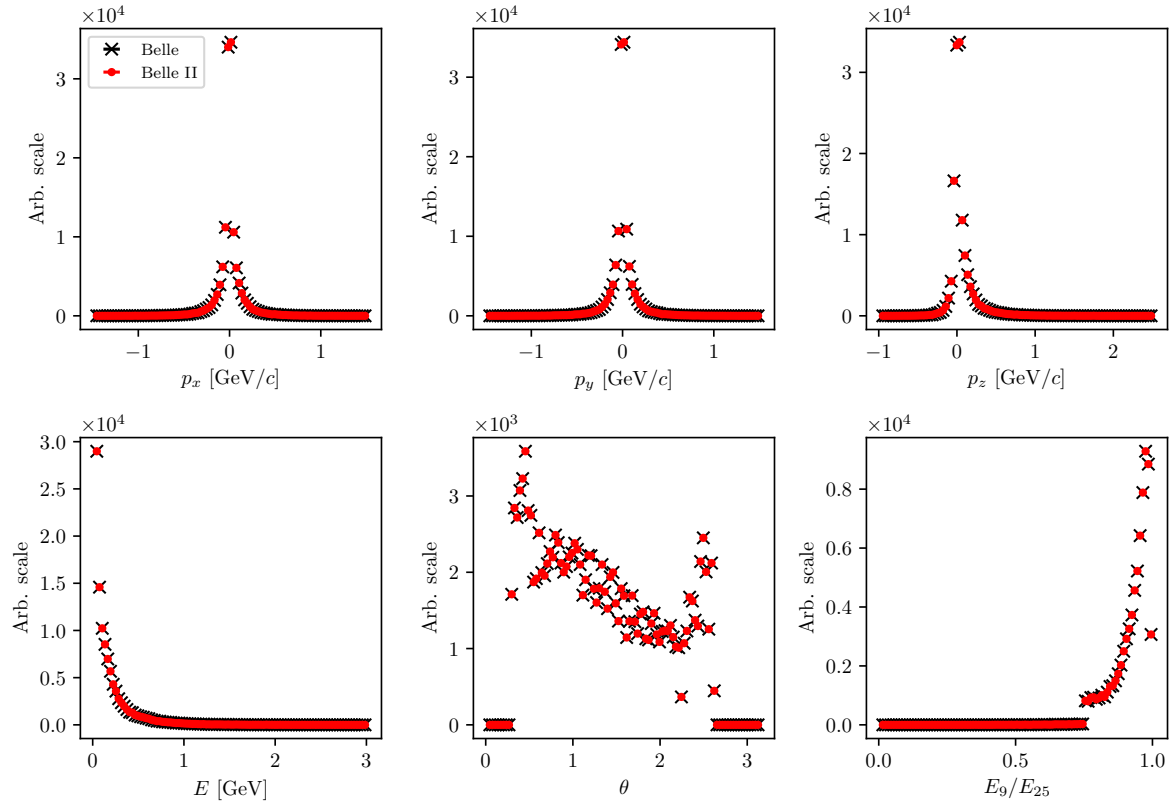
Figure 3.2: Some of the more important physical properties of photons for Belle and Belle II in the conversion process. The histograms seem to overlap and the conversion is assumed to be successful.

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 $B \to KK\ell\nu$ 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 of 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 te 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, where an extra category for true electrons/muons from the signal decay is shown.

16

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

Figure 4.2: Normalized properties of true (blue), fake (red) 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)

$$\text{FOM} = \sqrt{\mathcal{EP}} \propto \frac{S}{\sqrt{S+B}}, \tag{4.1}$$

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 candidatess (right).
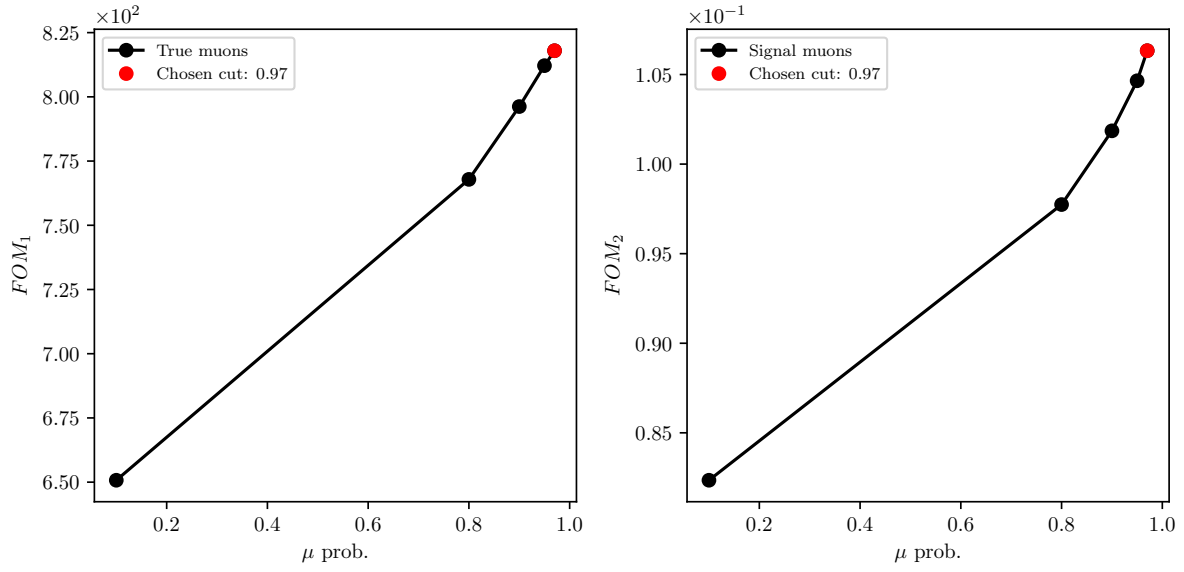
19

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

Kaons

We repeat the procedure for both kaons. Figure 4.5 shows 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 kaons, where an extra category for true kaons from the signal decay is shown.

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

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

362   • $|d_0| < 0.15$ cm,

363   • $|z_0| < 1.5$ cm,

364   • $p_{CMS} \in [0, 2.5]$ GeV/c.

365   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
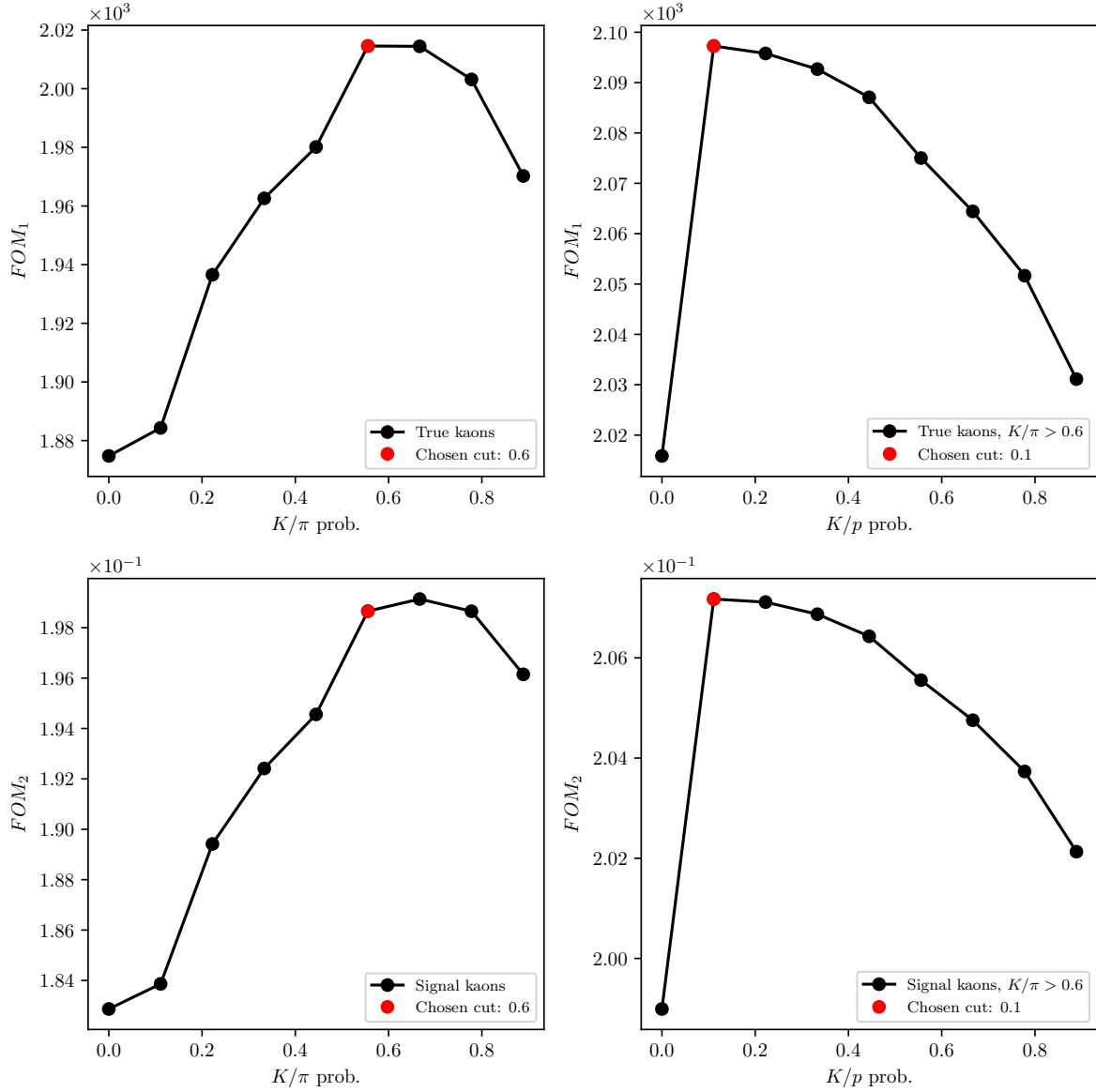procedure for PID cuts on kaon candidates.

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.

368 The optimized PID cuts for kaons are

369 • $K/\pi > 0.6$,

370 • $K/p > 0.1$.

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

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

$$B^+ \to K^+ K^- e^+,$$
$$B^+ \to K^+ K^- \mu^+,$$

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

The chosen pre-cut on the fit probability is

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

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

$$\mathrm{p}_\nu = \mathrm{p}_B - \mathrm{p}_Y, \tag{4.2}$$

$$\mathrm{p}_\nu^2 = m_\nu^2 = m_B^2 + m_Y^2 - 2E_B E_Y + 2\vec{p}_B \cdot \vec{p}_Y \approx 0, \tag{4.3}$$

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

where 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 can 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

$$E_B = E_{CMS}/2, \tag{4.5}$$

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

where E_{CMS} is the total energy of the e^+e^- collision in the CMS frame and m_B is the nominal mass of the B meson.

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

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

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

- $|\cos(\theta_{BY})| < 1.20$.

4.3 Loose neutrino reconstruction

The signal-side neutrino escapes detection, so we cannot directly determine it's 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

$$\mathrm{p}_{miss} = \mathrm{p}_{\Upsilon(4S)} - \sum_i^{\text{Event}} \left(E_i, \, \vec{p}_i \right), \tag{4.7}$$

$$\mathrm{p}_{miss} = \mathrm{p}_{\Upsilon(4S)} - \left(\mathrm{p}_Y - \sum_i^{\text{Rest of event}} \left(E_i, \, \vec{p}_i \right) \right), \tag{4.8}$$

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

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

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,

4. otherwise π.

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

$$p_\nu = p_{miss} = (E_{miss}, \vec{p}_{miss}),$$ (4.10)
$$m^2_{miss} = p^2_{miss} = p^2_\nu = m^2_\nu \approx 0.$$ (4.11)

Since the detector is not perfect, the distribution of the m^2_{miss} variable has a non-zero width. Additionally, a tail is introduced due to missing particles like neutrinos, other neutral undetected particles such as K^0_L, or simply missing tracks due to detection failure. Figure 4.9 shows the distribution of m^2_{miss} as defined with the missing four-momentum in Eq. (4.10). Correctly reconstructed candidates, which come from events where the other B meson decayed via a hadronic decay mode, peak at zero. If this is not the case, candidates are shifted to larger values of this variable, even if the event in question is a signal event. For this purpose we define a subset of all signal candidates, which come from events where the companion B meson decayed hadronically and all of it's particles were taken into account correctly. We only allow for missing photons, since they are frequently irradiated due to brehmsstrahlung effects and they do not have such a big impact on the 4-momentum of the final candidate. We denote this subset as *perfect* signal.

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

- $|m^2_{miss}| < 3.0 \text{ GeV}/c^2$,

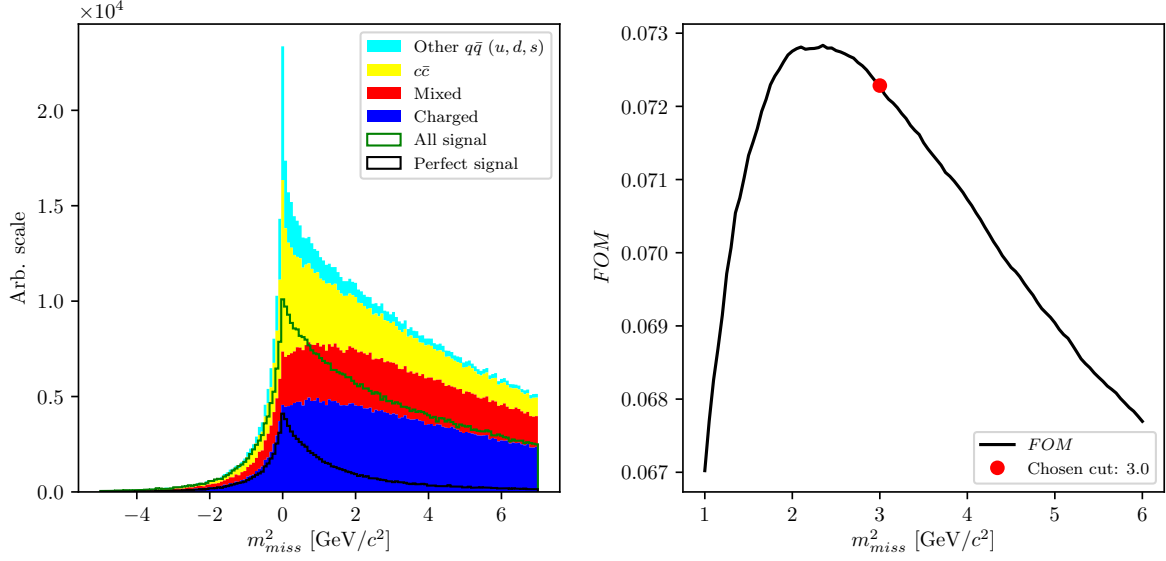which is also under-optimized at this point due to the same reasons as in the cases above.

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.

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

$$\mathrm{p}_\nu = \left(|\vec{p}_{miss}|, \, \vec{p}_{miss}\right), \tag{4.12}$$

which fixes the neutrino mass to 0 GeV/c^2.

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

$$\Delta E = E_B - E_{CMS}/2, \tag{4.13}$$

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

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

- $M_{BC} \in [5.1, \, 5.295]$ GeV/c^2,

- $\Delta E \in [-1.0, \, 1.3]$ GeV.

Figure 4.10 shows the distributions of ΔE (left) and M_{BC} (right) for signal and major
types of background. Both signal components are scaled up with

respect to the background components, but are in proper scale one to another. The effects of missing particles are clearly seen based on the shape difference between all and perfect signal.
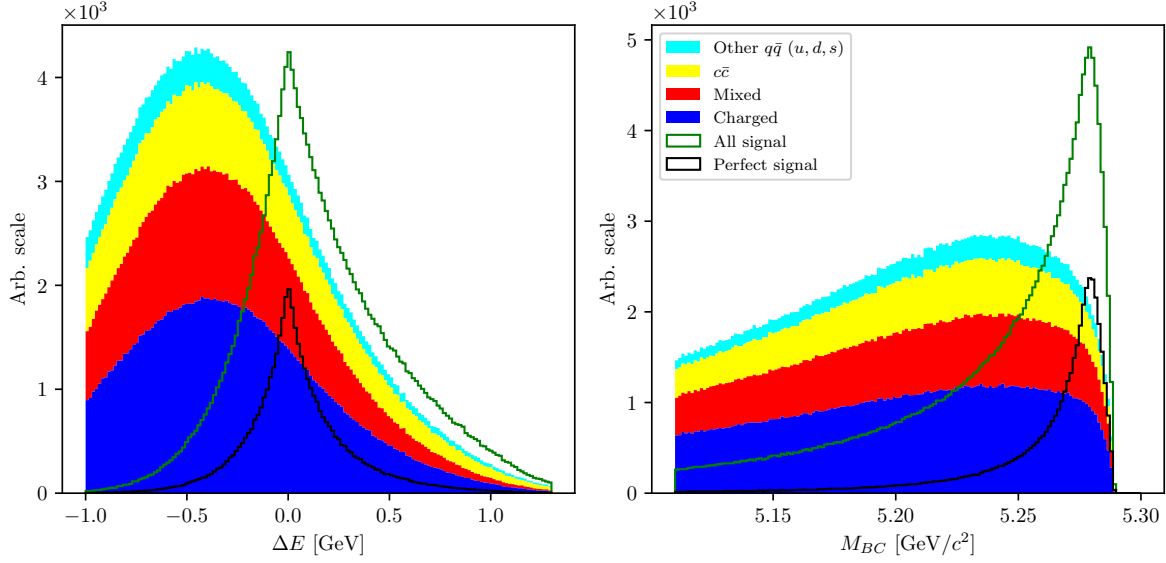

Figure 4.10: Distributions of ΔE (left) and M_{BC} (right) for signal and major types of background after the precuts. Both signal components are scaled up with respect to the background components, but are in proper scale one to another. The perfect signal has a much better resolution in both distributions,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

4.4 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 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 4.11, while the optimal cut values are

- $P(\chi^2, NDF) > 6.0 \times 10^{-3}$,

- $|\cos(\theta_{BY})| < 1.05$.

29

Figure 4.11: 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 4.12.

The signal region is defined as

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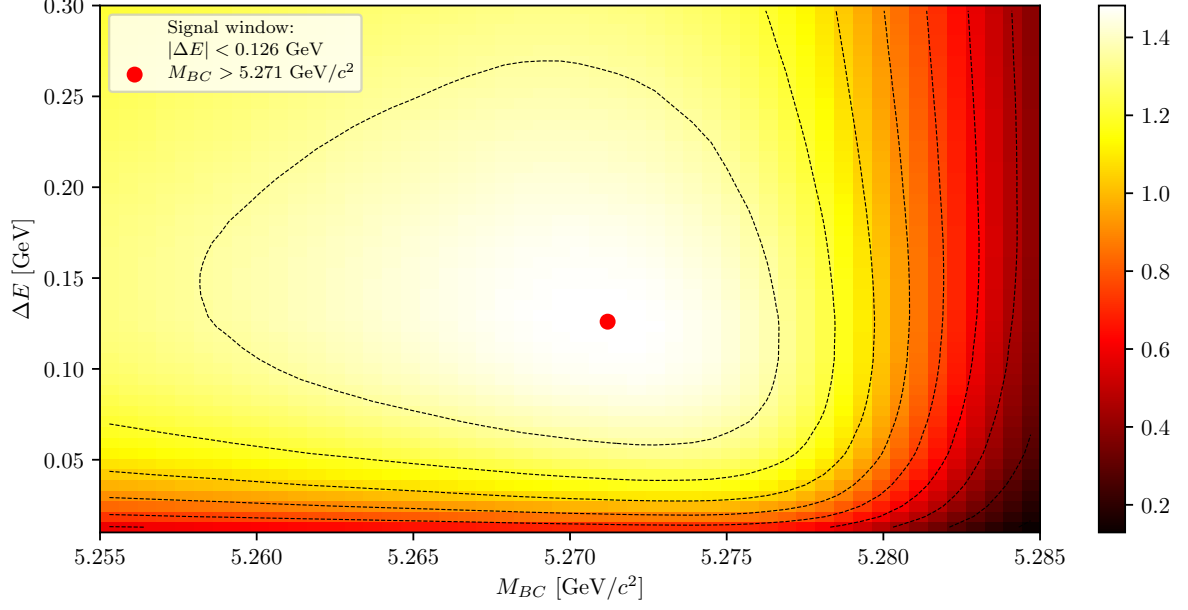
- $|\Delta E| < 0.126$ Gev.

Figure 4.12: 2D FOM optimization of the signal region definition, where the signal in the optimization was represented by perfectly reconstructed candidates.

With the signal window defined, we can tighten the cut on m_{miss}^2, which we intentionally left loose before the signal categorization. With the FOM optimization of perfectly reconstructed candidates inside the signal region, shown in Figure 4.13, the optimal cut on m_{miss}^2 is

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

Figure 4.13: 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 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

$$q_{B^{\pm}}q_{B^{\mp}} = -1, \tag{4.15}$$

however, this value is distributed due to missing charged particles in the reconstruction. Figure 4.14 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. 4.15.

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}.
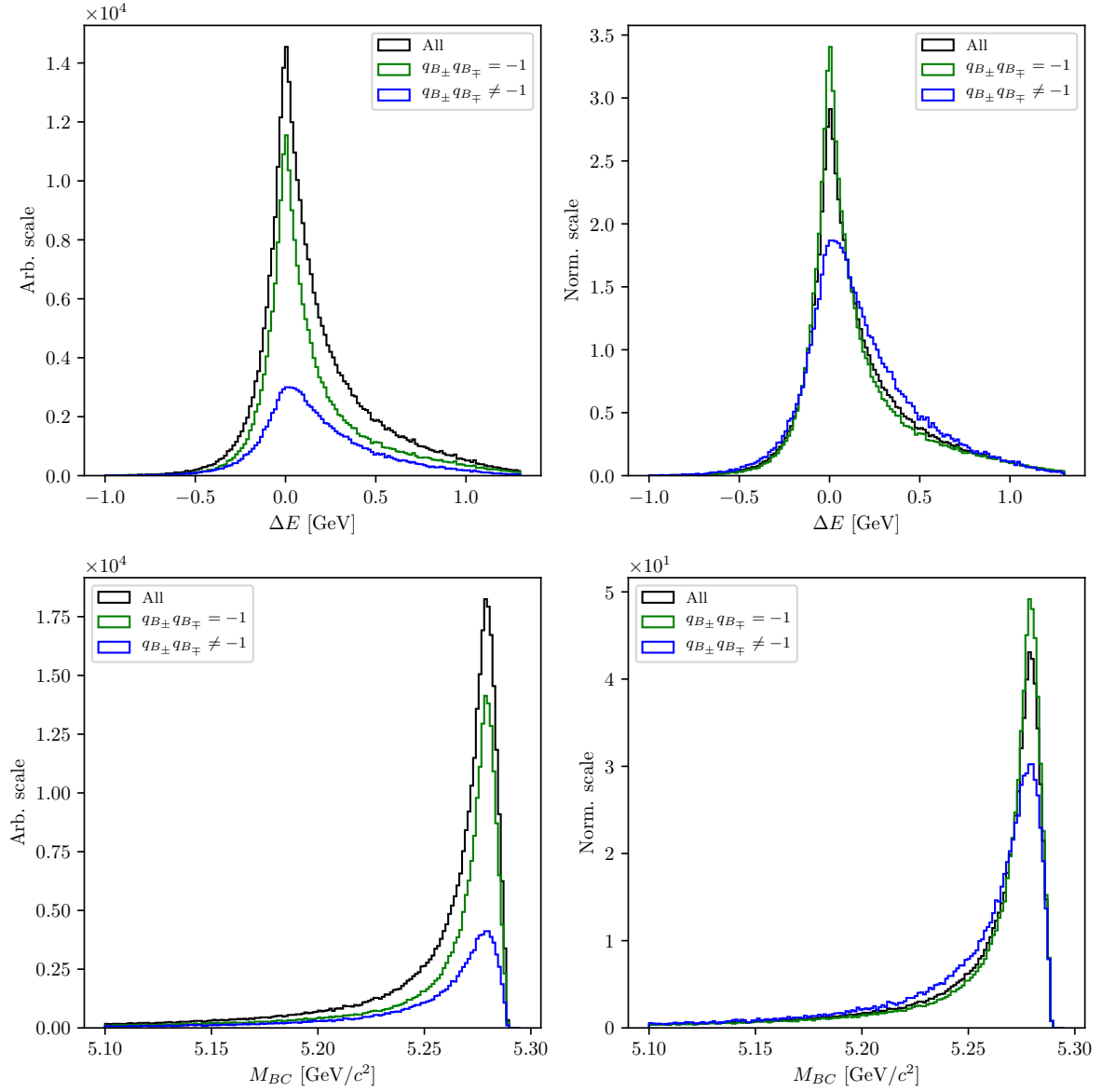

Figure 4.14: Signal distributions of ΔE and M_{BC} based on the charge product of both B mesons in the event. The plot on the left shows the distributions in an arbitrary scales, while the plot on the right shows the normalized distributions.

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, $\mu ID > 0.97$,

  - kaons: $|d_0| < 0.15$ cm, $|z_0| < 1, 5$ cm, $p_{CMS} < 2.5$ GeV/c, $K/\pi\ ID > 0.6$, $K/p\ ID > 0.1$,

- B meson candidates:

  - standard cuts: $P(\chi^2, DOF) > 6{\times}10^{-3}$, $|\cos\theta_{BY}| < 1.05$, $|m^2_{miss}| < 0.975$ GeV/c^2,

  - fit region cuts: $\Delta E \in [-1.0, 1.3]$ GeV, $M_{BC} \in [5.1, 5.295]$ GeV/c^2,

  - signal region cuts: $|\Delta E| < 0.126$ GeV, $M_{BC} > 5.271$ GeV/c^2,

  - 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.

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 $\pi^0 \to \gamma\gamma$ decays. However, a lot of hits in the ECL are also created by photons coming from 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.

35

5.1.1 π^0 MVA training

The training dataset of π^0 candidates contains

- 183255 target candidates,

- 200000 background candidates,

where the definition of 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

- 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.
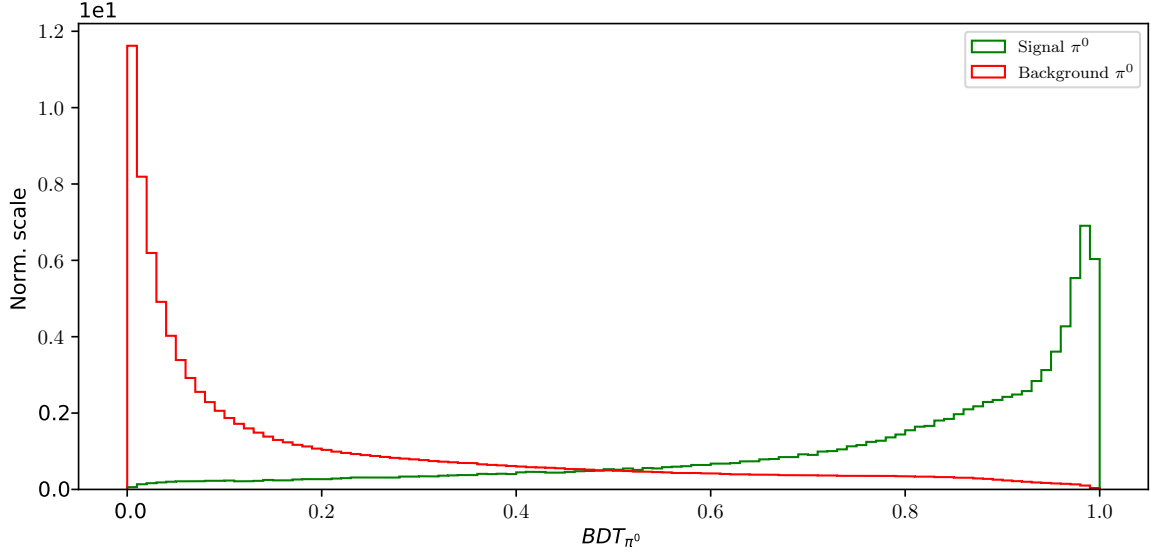
The classifier output variable is shown in Figure 5.1.

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

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's 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 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

594 list.

595 The input variables used in this MVA are

596   • p and p_{CMS} of γ candidates,

597   • π^0 probability,

598   • cluster quantities

599   – $E_9/E25$,

600   – theta angle,

601   – number of hit cells in the ECL,

602   – highest energy in cell,

603   – energy error,

604   – distance to closest track at ECL radius.

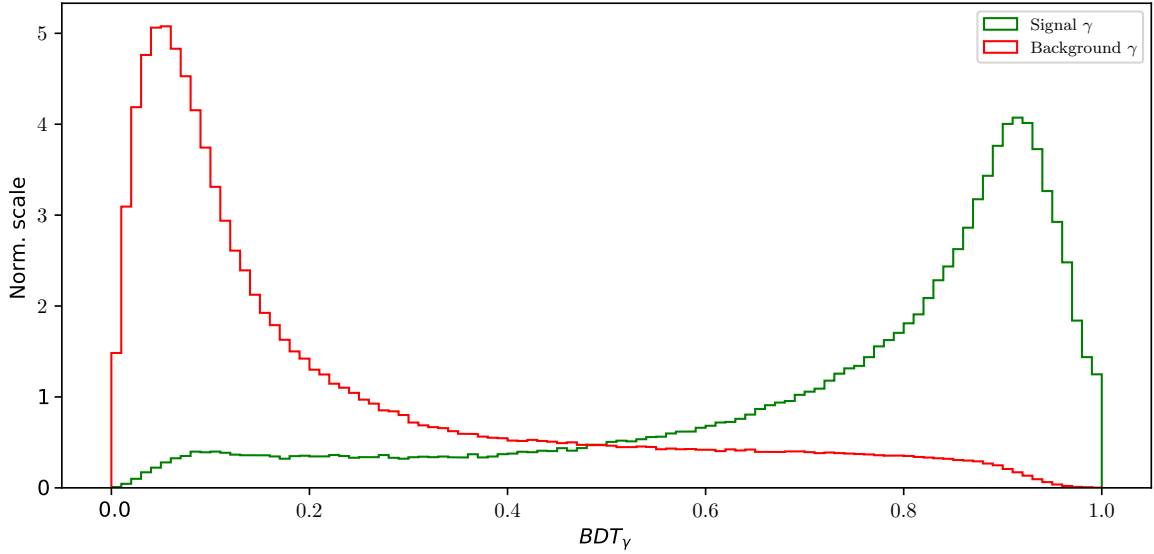605 The classifier output variable is shown in Figure 5.2.

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

606 With the final weights for photon classification in hand, we apply them to the photon
607 particle list. The cut optimization is shown in Figure 5.3 (left), with the optimal cut on

38

the γ classifier output at

- $BDT_\gamma > 0.5045$.

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

- Signal efficiency: $\epsilon_{SIG} = 83.2$ %,

- 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).

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

5.2 Tracks clean-up

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

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

(a) good tracks can also originate away from the IP region, due to decays of long-lived particles, such as $K_S^0 \to \pi^+\pi^-$,

(b) charged particles from background sources produce extra tracks, or duplicates,

(c) low momentum charged particles can curl in the magnetic field and produce multiple tracks,

(d) secondary interactions with detector material or decays of particles in flight can produce "kinks" in the flight directory, resulting in multiple track fit results per track.

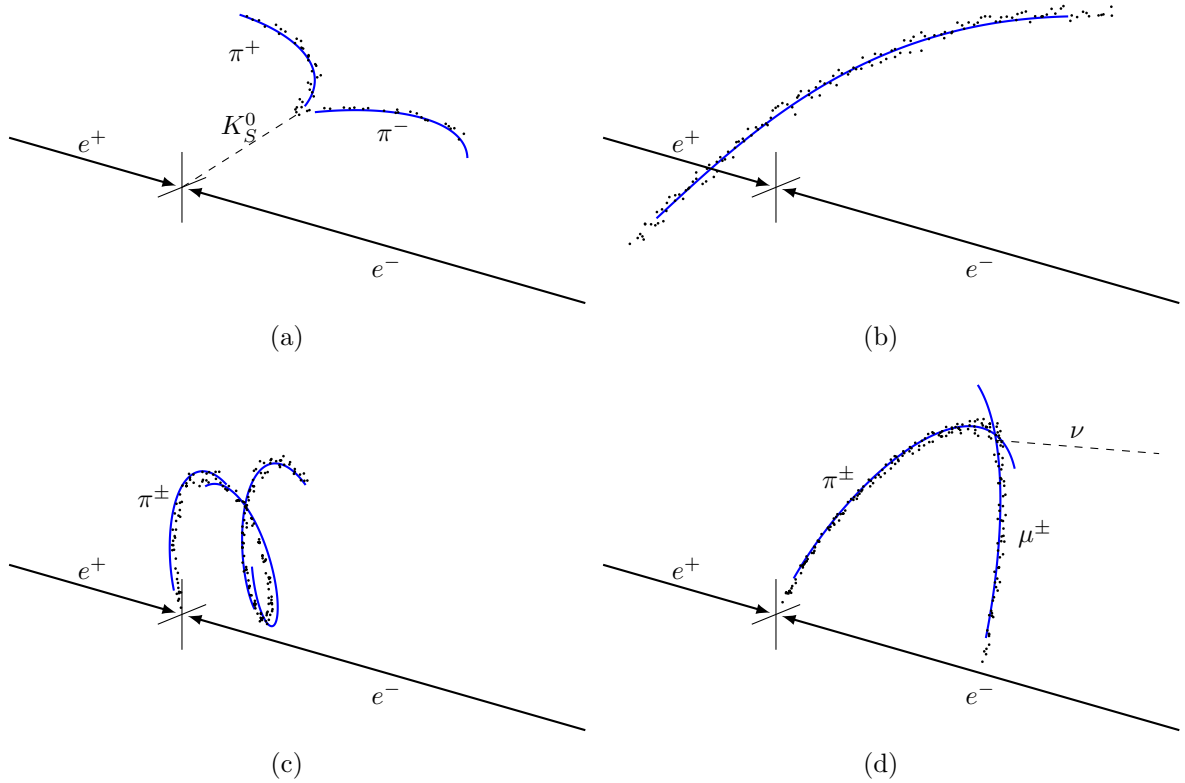Schematics of all the cases mentioned above are shown in Figure 5.4.

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 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 don't 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. Here we only focus on K_S^0, since they are the most abundant. This step is necessary because the $\pi^\pm$ 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 result 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 selected K_S^0 candidates is added to the ROE, while the daughter tracks are discarded from our set.
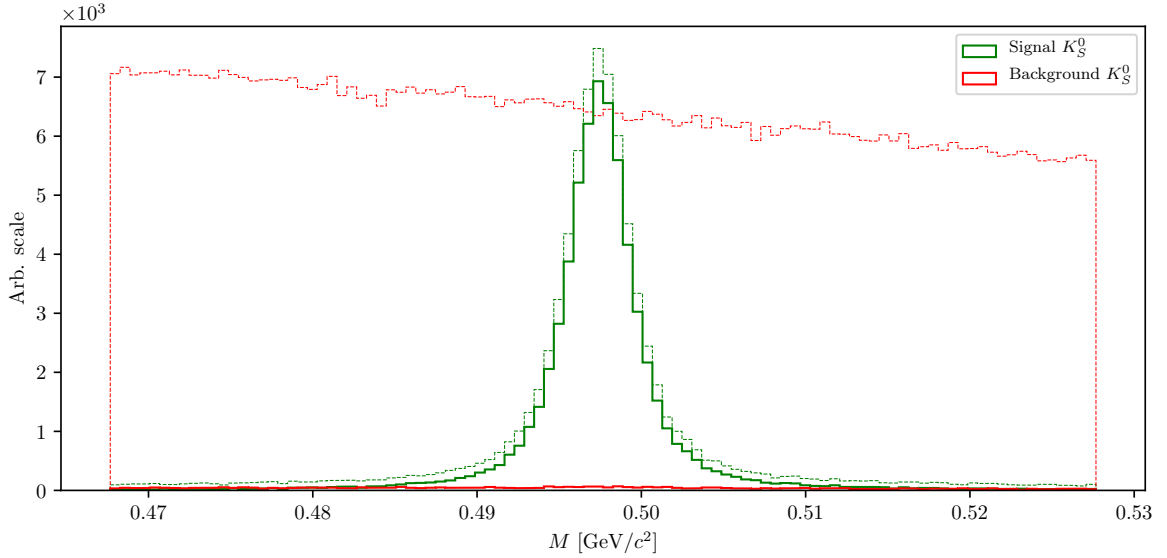

Figure 5.5: Invariant mass of the K_S^0 candidates before (dashed lines) and after (solid lines) the cut on the Neural Network classifier for signal (green) and background candidates (red). Signal peaks at nominal K_S^0 mass, while background covers a wider region.

41

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

- Signal efficiency: $\epsilon_{SIG} = 80.7$ %,

- Background rejection: $1 - \epsilon_{BKG} = 99.4$ %.

5.2.2 Duplicate tracks

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

- $|d_0| < 10$ cm and $|z_0| < 20$ cm

and proceed with the clean-up of track duplicates.

Defining a duplicate track pair

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

42

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).

If the particle decayed mid-flight or produced multiple tracks due to being a low-momentum curler, then, as the name suggests, these particles most likely had low momentum in the transverse direction, p_T. Since both tracks originate from the same initial particle, the momentum difference should also peak at small values. Figure 5.7 shows the momentum and momentum difference of tracks which belong to a random or a duplicate track pair.
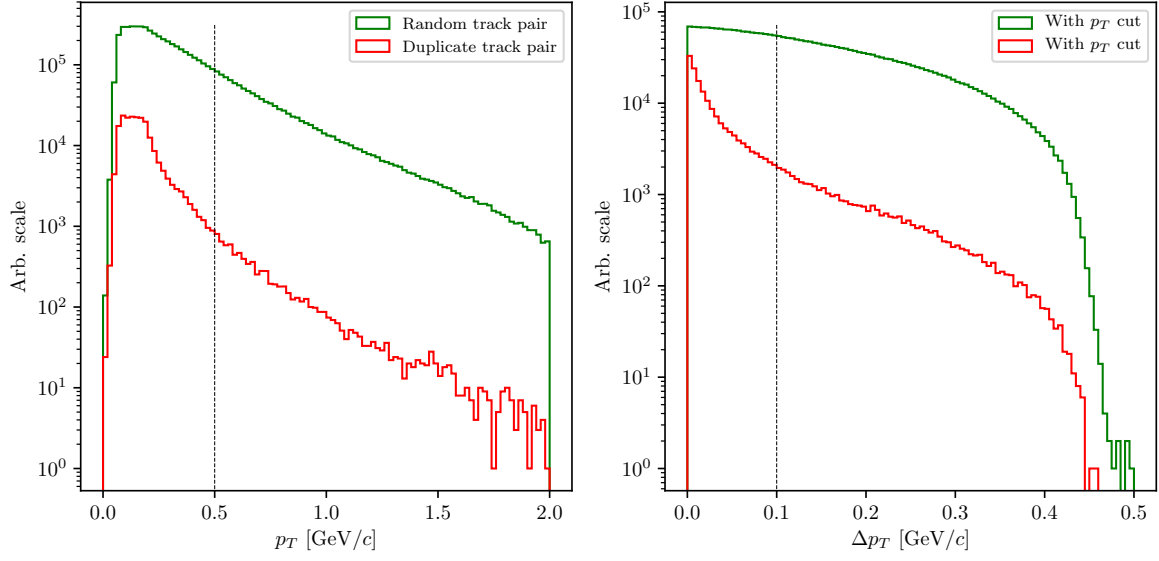
Figure 5.7: Distribution of transverse momentum p_T (left) and transverse momentum difference Δp_T (right) for all tracks coming from random (green) or duplicate track pairs (red). The plot on the right already includes the cut on p_T from the plot on the left.

We impose a cut of

- $p_T < 0.5$ GeV/c,

- $|\Delta p_T| < 0.1$ GeV/c,

in order to cut down the number of random track pairs, while retaining a high percentage of duplicate track pairs. After all the cuts defined in this chapter, the 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).

Training the duplicate track pair MVA

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

- 113707 target candidates,

- 190314 background candidates,

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

The input variables used in this MVA are

- angle between tracks,

- track quantities

    - impact parameters d_0 and z_0,

    - transverse momentum p_T,

45

– helix parameters and helix parameter errors of the track,

– track fit p-value,

– number of hits in the SVD and CDC detectors

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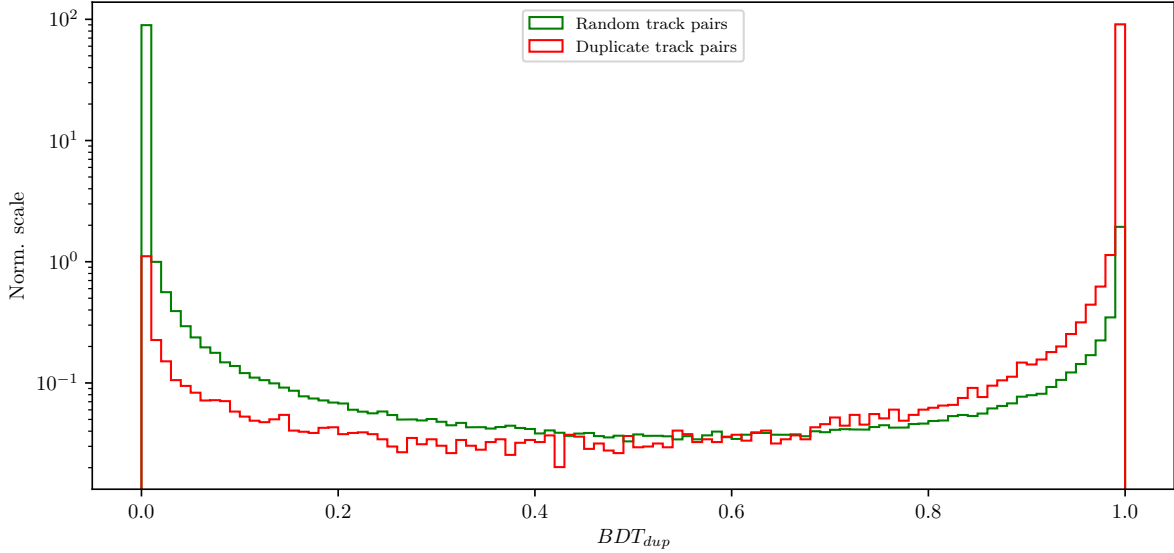efficient manner, as shown in Figure 5.9.

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

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

• $BDT_{duplicate} > 0.9985$.

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

• Signal efficiency: $\epsilon_{SIG} = 87.2$ %,

• Background rejection: $1 - \epsilon_{BKG} = 98.8$ %,

where signal and background represent duplicate and random track pairs, respectively.

Figure 5.10: The optimization of the FOM function for the cut on clas-
sifier 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 ran-
dom 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

$$\Delta f = f_{this} - f_{other}, \tag{5.1}$$

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

- Δd_0, $\Delta z_0 > 0$ (*this* track further away from the IP region),

- Δp_T, $\Delta p_Z < 0$ (*this* track has lower momentum),

- ΔN_{SVD}, $\Delta N_{CDC} < 0$ (*this* track has less hits in the SVD and CDC),

Additionally we define an MC truth variable

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

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

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

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.

$$1. \ (t_1, t_2*) \quad \rightarrow \quad (t_1, t_3*) \quad \rightarrow \quad (t_2*, t_3*),$$
$$2. \ (t_1, t_2*) \quad \rightarrow \quad (t_2*, t_3*) \quad \rightarrow \quad (t_1, t_3*),$$
$$3. \ (t_1, t_3*) \quad \rightarrow \quad (t_2, t_3*) \quad \rightarrow \quad (t_1, t_2*),$$
$$4. \ (t_1, t_3*) \quad \rightarrow \quad (t_1, t_2*) \quad \rightarrow \quad (t_2*, t_3*),$$
$$5. \ (t_2, t_3*) \quad \rightarrow \quad (t_1, t_3*) \quad \rightarrow \quad (t_1, t_2*),$$
$$6. \ (t_2, t_3*) \quad \rightarrow \quad (t_1, t_2*) \quad \rightarrow \quad (t_1, t_3*),$$

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

Training the duplicate track MVA

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

- 84339 target candidates,

- 68280 background candidates,

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

The input variables used in this MVA are

- theta angle of the track momentum,

- track quantities

  - impact parameters d_0 and z_0 and their errors,

  - CMS frame momentum p_{CMS} and momentum components p_T and p_z

  - number of hits in the SVD and CDC detectors

  - track fit p-value,

- pair-level information

  - Δd_0, Δz_0, ΔN_{CDC}, ΔN_{SVD}, Δp_T, Δp_z, Δp-value.

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

$$\Delta BDT_{final} = BDT_{final}^{this} - BDT_{final}^{other}, \tag{5.3}$$

which is again applied to all track pairs and overwritten for tracks which are more duplicate-like. The classifier output and the classifier output difference for each track is 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).

Finally, we select all duplicate tracks which survive the cut

$$\Delta BDT_{final} > 0 \tag{5.4}$$

and discard them from our ROE. We can check the performance of our duplicate track classifier by applying the procedure on a validation sample of duplicate track pairs and compare the predicted result with the truth, based on Eq. (5.2). Table 5.1 shows the performance of the duplicate track recognition in the form of percentages of correctly and incorrectly identified duplicate and original tracks. The model seems 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

The clean-up, used standardly at Belle, is much simpler and relies only on a set of rectangular cuts for neutral particles as well as charged ones. In 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.

	$17° < \theta < 32°$	$32° < \theta < 130°$	$130° < \theta < 150°$
E_γ	> 100 MeV	> 50 MeV	> 150 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:

- $p_T < 275$ MeV$/c$,

- $\Delta p = |\mathbf{p}_1 - \mathbf{p}_2| < 100$ MeV$/c$,

- $\cos\theta(\mathbf{p}_1, \mathbf{p}_2) < 15°$ for same sign,

- $\cos\theta(\mathbf{p}_1, \mathbf{p}_2) > 165°$ 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.

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

	$p_T < 250$ MeV$/c$	250 MeV$/c < p_T < 500$ MeV$/c$	$p_T > 500$ MeV$/c$
$|d_0|$	< 20 cm	< 15 cm	< 10 cm
$|z_0|$	< 100 cm	< 50 cm	< 20 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 distribution, 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 the clean-up procedure to our signal MC sample with all the applied cuts, defined in 4.6, except

₇₇₄ for the signal categorization cuts. Figure 5.12 (left) shows signal candidate distributions
₇₇₅ of ΔE and M_{BC} for various clean-up setups. Focusing on the ROE clean-up, we see an
₇₇₆ improvement in resolution in both observed variables and an overall decrease in efficiency.
₇₇₇ The efficiency decrease is expected, since the cleaned-up variables are able to better
₇₇₈ isolate the perfectly reconstructed candidates and discard the non-perfect candidates.
₇₇₉ In fact, the efficiency of the perfectly reconstructed candidates increases after the ROE
₇₈₀ clean-up, as shown in Figure 5.12 (right). The signal MC sample after in case of the Belle
₇₈₁ clean-up also shows a slight improvement in the resolution, but looking at the perfectly
₇₈₂ reconstructed candidates we see that this clean-up procedure is not optimal. Table 5.4
₇₈₃ shows ratios of efficiencies and $FWHM$'s of the clean-up procedures for the perfect
₇₈₄ signal with respect to the default case, based on the ΔE distribution. While both the
₇₈₅ Belle and ROE clean-up improve the resolution, ROE clean-up performs significantly
₇₈₆ better and also increases the amount of the perfectly reconstructed candidates in the
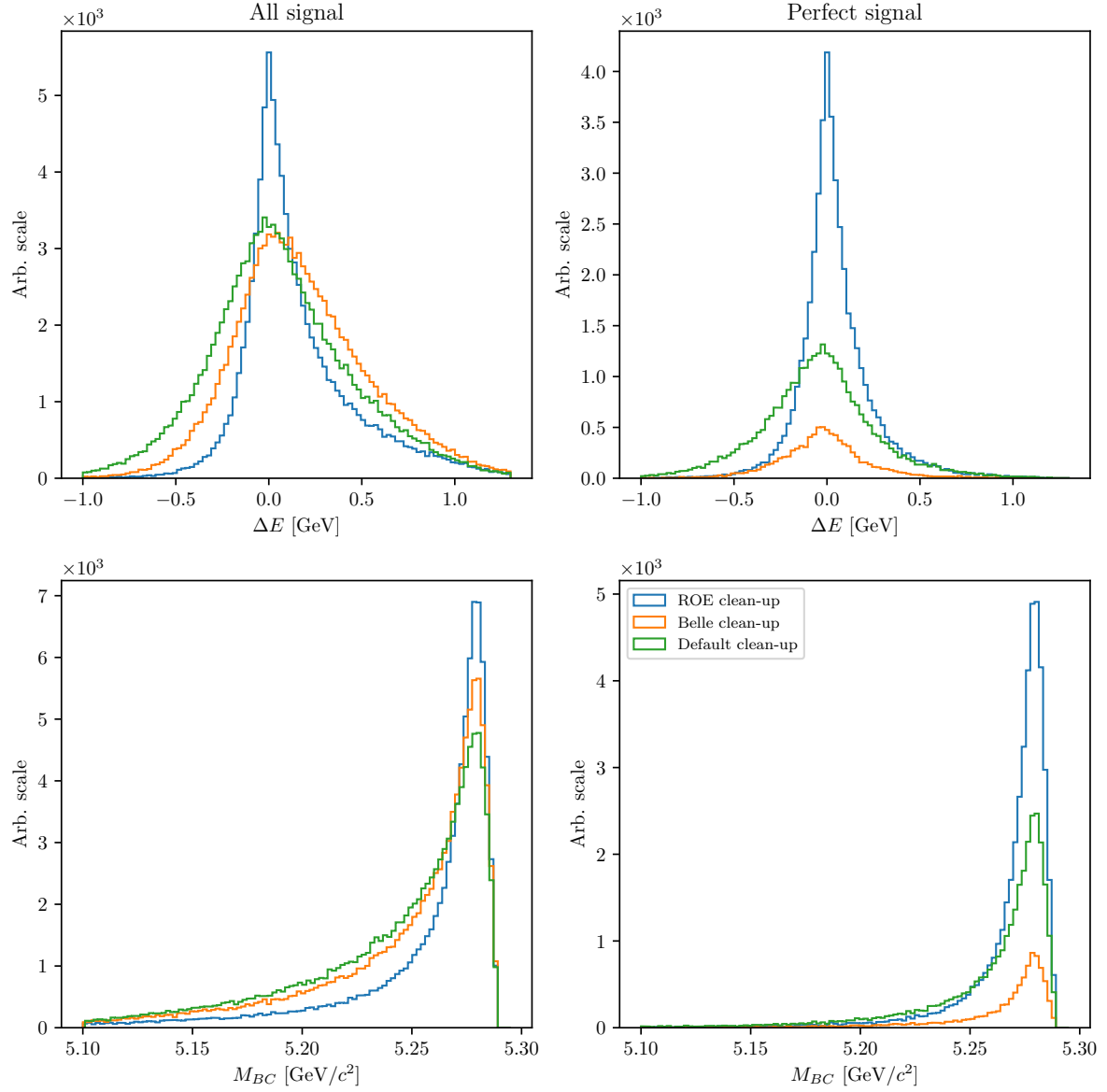₇₈₇ final sample.

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).

Another variable which heavily depends on the clean-up is the charge product of the signal and companion B meson candidate, already defined in Eq. (4.15), shown in Figure 5.13 for various clean-up procedures. The figure shows an improved resolution of the distribution, which means that candidates migrate to the correct value of the charge product after the clean-up. Looking at the perfectly reconstructed candidates we again see the increase in the bin corresponding to the correct charge product. As a cross-check, we can also look at ΔE and M_{BC} variables for each value of the charge product. These plots are shown for the full signal MC sample in Figure 5.14 and they show a clear resolution improvement for the correct value of the charge product in the case of the ROE clean-up. For other values of the charge product there also seems to be a small improvement for both cases of clean-up, but it is negligible compared to the plots in the second column. This supports our choice of signal categorization, defined in Section 4.5, where we select only candidates with 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.

All signal

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-cleanup 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,

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

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

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

as shown in Figure 5.15.

Figure 5.15: Distributions of m_{KK} for the full MC dataset. The black lines represent the cut on region in the m_{KK} distribution where the control sample is enhanced.

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 stram of the full available generated MC.

The affects of the ROE clean-up are shown in Figure 5.16. We see that data and MC agree well, and evidently even better after the clean-up procedure. 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.

57
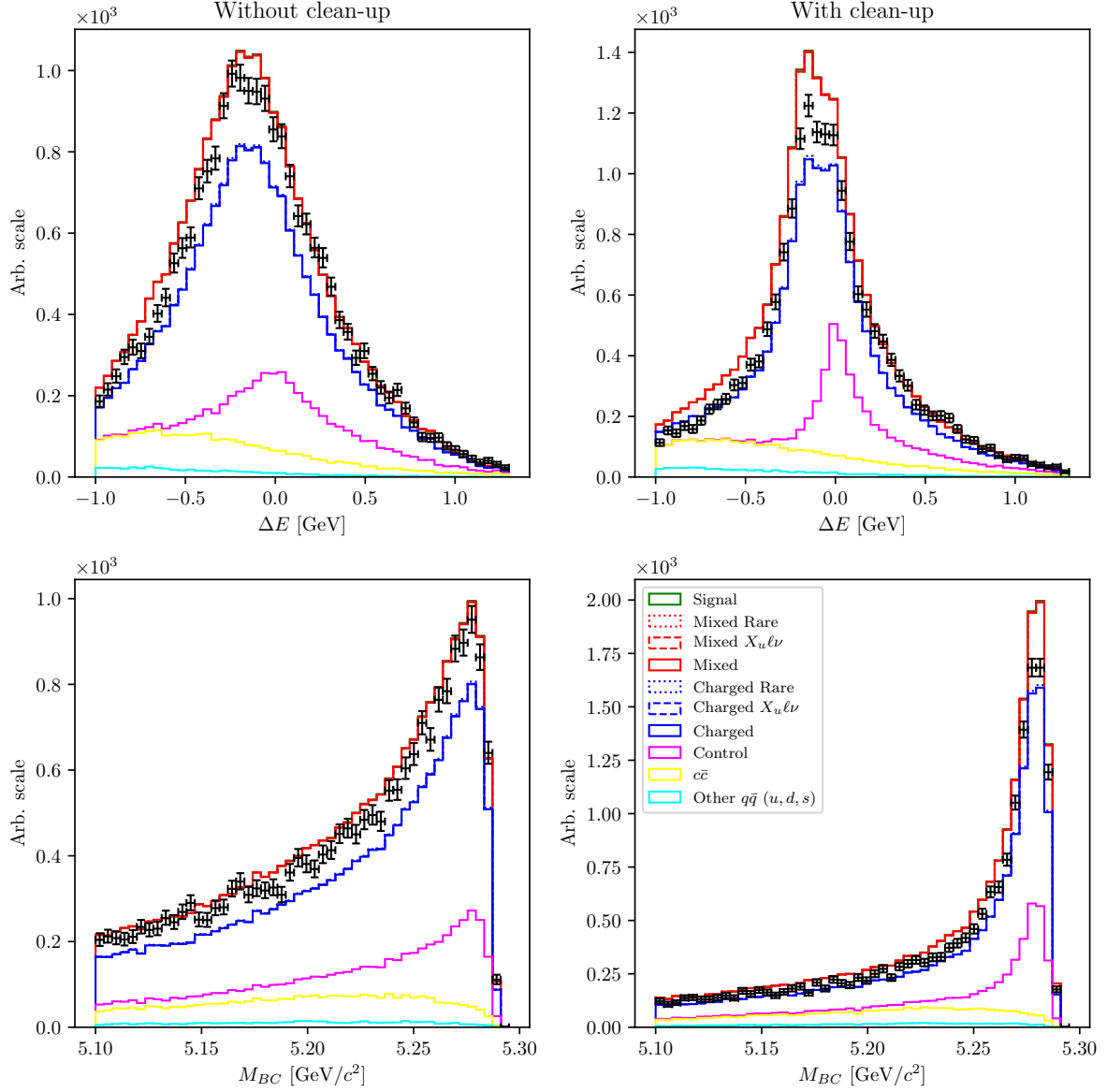
Figure 5.16: Distributions of ΔE (top) and M_{BC} bottom for the case without (left) and with ROE clean-up (right). The resolution of the control sample is improved and the MC and data agree well in all aspects. While the simulated background resolution is also improved, it is worse compared to the resolution of the control sample.

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

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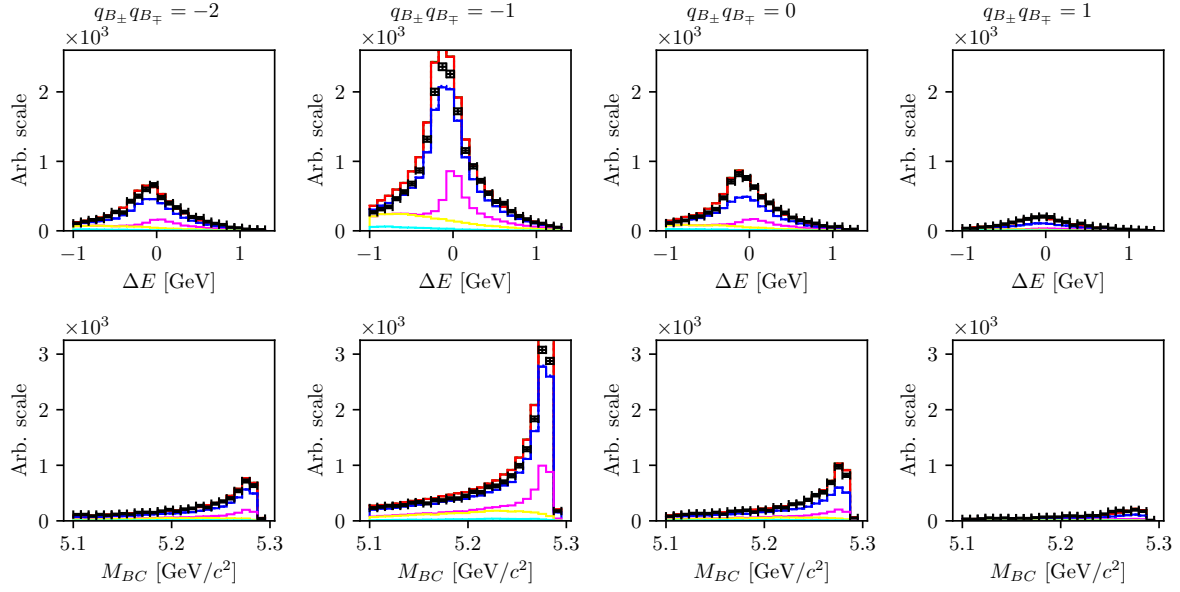
Figure 5.17: Distributions of ΔE (top) and M_{BC} (bottom) split in bins of the charge product of the two B mesons.

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

Chapter 6

Background suppression

This chapter shows the procedure in suppressing various kinds of background by applying cuts on MVA classifier outputs.

6.1 Resonant background

In this analysis we study decays with kaons in the final state. This means that standard procedures in $b \to u$ analyses in order to suppress $b \to c$ backgrounds, such as K-veto, are not possible. As a consequence, our final sample consists of combinations of K pairs coming also from $b \to c$ sources, such as $D^0 \to K^+ K^-$. 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

- $\phi \to K^+ K^-$ (sharp resonance at ~ 1.019 GeV/c^2),

- $D^0 \to K^+ K^-$ (sharp peak at ~ 1.864 GeV/c^2),

- $D^0 \to K^+ \pi^-$ (wide, shifted peak, due to kaon miss-identification).

In order to suppress these resonant backgrounds, while studying the 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 $m_\phi \approx 1.019$ GeV/c^2 and $m_{D^0} \approx 1.864$ GeV/c^2 are nominal masses of the ϕ and D^0 mesons, and $\Delta_\phi \approx 8 \times 10^{-3}$ GeV/c^2 and $\Delta_{D^0} \approx 1.5 \times 10^{-2}$ 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.

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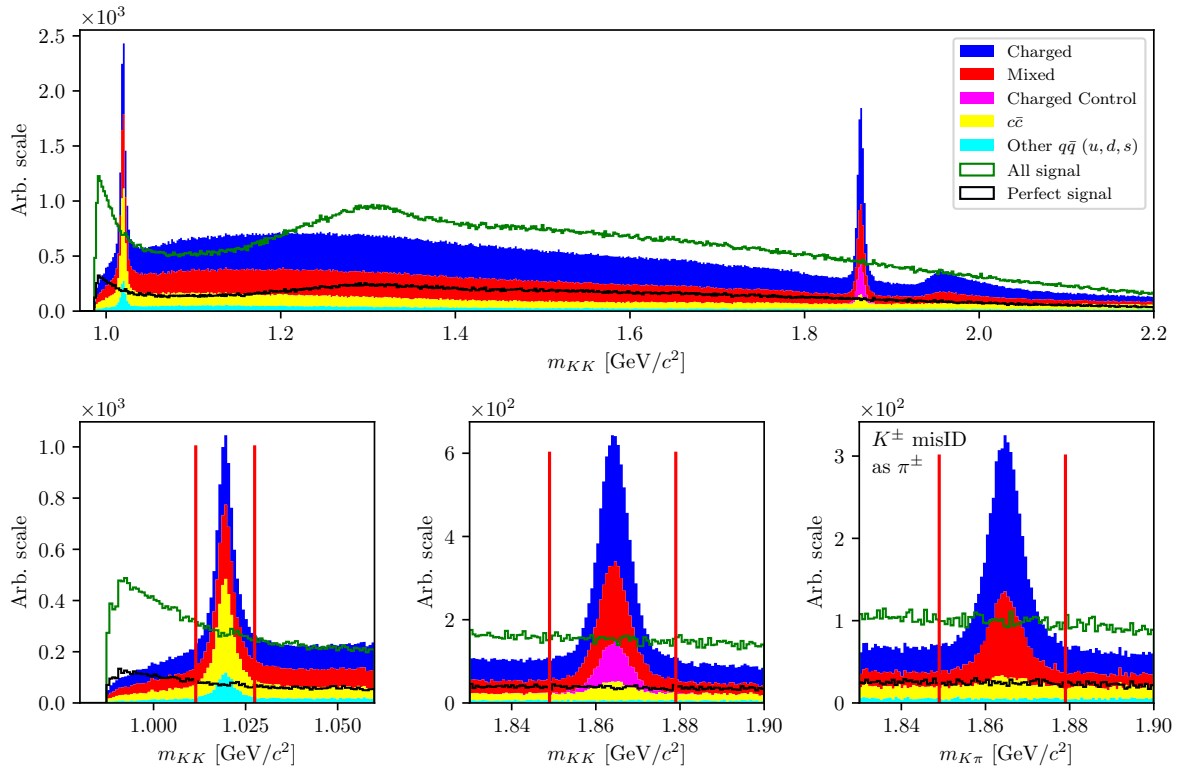
Figure 6.1: Invariant mass of two correctly reconstructed kaons (left) and invariant mass of two kaons, where one was miss-identified as a pion (right). Signal (green) and perfect signal (black) are equally scaled up.

6.2 Continuum suppression

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

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

where $q = u$, d, s or c, and are a sizable contribution to $B\bar{B}$ events. Additionally to kinematic constraints to separate $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ decays from $e^+e^- \to q\bar{q}$, 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 CMS frame. Their decay products from an isotropic distribution in the detector, which yields a spherical event shape.

6.2.1 Characteristic variables

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

Thrust and related variables

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

$$\mathbf{T} = \frac{\sum_i |\mathbf{T} \cdot \mathbf{p}_i|}{\sum_i |\mathbf{p}_i|}. \tag{6.1}$$

In this case, a related variable is $|\cos\theta_T|$, where θ_T is the angle between the thrust axis of the momenta from B meson decay particles, and the thrust axis of all particles in the ROE. Since both B mesons in $B\bar{B}$ events are produced at rest, their decay particles, and consequentially their thrust axes, are uniformly distributed in the range $[0,\ 1]$. On the other hand, decay particles from continuum events follow the direction of the jets in the event. As a consequence, the thrusts of both the B meson and the ROE are strongly directional and collimated, which results in a large peak at $|\cos\theta_T| \approx 1$. Additionally, one can also use the variable $|\cos\theta_{TB}|$, which is the thrust axis between the B candidate and the beam axis. For B candidates from $B\bar{B}$, this distributions is

again uniformly distributed, while for candidates from continuum events this distribution follows the distribution of the jets with the function $1 + \cos^2 \theta_{T,B}$. Figure 6.2 shows the distributions of $|\cos \theta_T|$ (left) and $|\cos \theta_{T,B}|$ (right) for different B meson candidates.
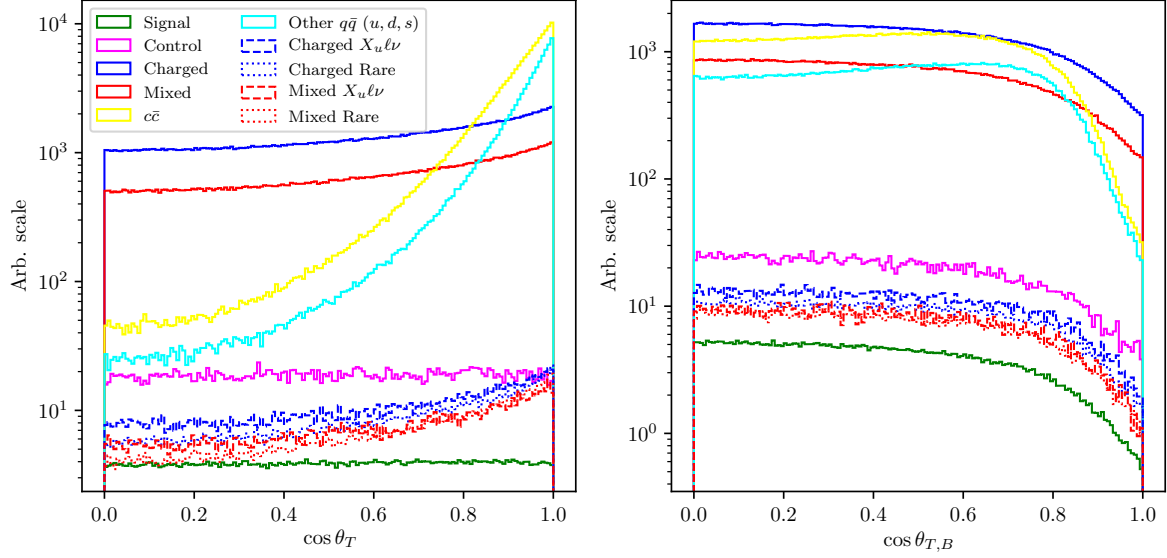

Figure 6.2: Distributions of $|\cos \theta_T|$ (left) and $|\cos \theta_{T,B}|$ (right) for different B meson candidates.

CLEO cones

CLEO cones have been introduced by the CLEO collaboration [7] and are an additional specific tool to provide optimal background discrimination. They are nine variables corresponding to the momentum flow around the thrust axis of the B meson candidate, binned in nine cones of 10° around the thrust axis, as illustrated in Figure 6.3.

63

Figure 6.3: Concept of CLEO cones. **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 $10°$.

KSFW moments

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

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

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

Belle developed a refined generation of Fox-Wolfram moments, called Kakuno-Super-Fox-Wolfram (KSFW) moments to further suppress the continuum background. There are 17 different KSFW moments which are grouped into R_k^{so}, R_k^{oo} and R_k^{ss} [8]. The latter ones are excluded due to correlations with B meson specific variables.

6.2.2 MVA training

Most of the characteristic variables, described in subsection 6.2.1, were taken together in order to train a single MVA classifier for continuum suppression. All characteristic

64

variables were checked for possible q^2 correlation. Variables with significant correlation or complex shapes in the 2D plot were discarded from the training set, since they would have introduced unwanted dependence on the unreliable model, ISGW2, used for signal MC generation. Additionally, all of the characteristic variables in our set do not depend on the signal mode, they only differ in the kinematic and topological aspects of $B\bar{B}$ and continuum background events.

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

The training variable set consisted of

- B meson direction and thrust related variables

    - magnitude of thrust axes of B and ROE,

    - cosine of angle between thrust axis of B and thrust axis of ROE,

    - cosine of angle between thrust axis of B and beam direction,

    - reduced Fox-Wolfram moment R_2 ,

- all 9 CLEO Cones

- KSFW Moments

    - R_{01}^{so}, R_{02}^{so}, R_{03}^{so}, R_{04}^{so},

    - R_{10}^{so}, R_{12}^{so}, R_{14}^{so},

    - R_{20}^{so}, R_{22}^{so}, R_{24}^{so},

    - R_0^{oo}, R_1^{oo}, R_2^{oo}, R_3^{oo}, R_4^{oo},

- FlavorTagging variables

    - qp of e, μ, ℓ,

    - qp of intermediate e, μ, ℓ,

    - qp of K, K/π, slow pion, fast hadron,

    - qp of maximum P^*, Λ, fast-slow-correlated (FSC),

• Other

– Δz, Δt.

Figure 6.4 shows the classifier output for various types of background, all in expected
MC proportions. B meson candidates from continuum background are dominant at
lower values, while candidates from $B\bar{B}$ events populate the region with 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.

6.3 $B\bar{B}$ suppression

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

• $b \to c\ell\nu$ background,

• $b \to u\ell\nu$ background,

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

Similarly as before, 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 $b \to u\ell\nu$ and other rare decays. The training dataset was proportionally sampled in the same manner as described in subsection 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 [9]

- Δz,

- m_{miss}^2 from Eq. (4.11),

- $B \to D^*\ell\nu$ 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, $B \to D^*\ell\nu$ veto variables, are a set of variables where we partially reconstruct the D^* candidate 4-momentum via a linear combination of the $\pi_s^{\pm}$ 4-momentum in the $D^* \to D\pi_s^{\pm}$ decay. It helps discard the most dominant $B \to D^*\ell\nu$ background. It is most efficient in the $B^0 \to D^{*-}\ell^+\nu$ decay, where D^{*-} further decays via $D^{*-} \to \bar{D}^0\pi_s^-$. 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 $\pi_s^{\pm}$.
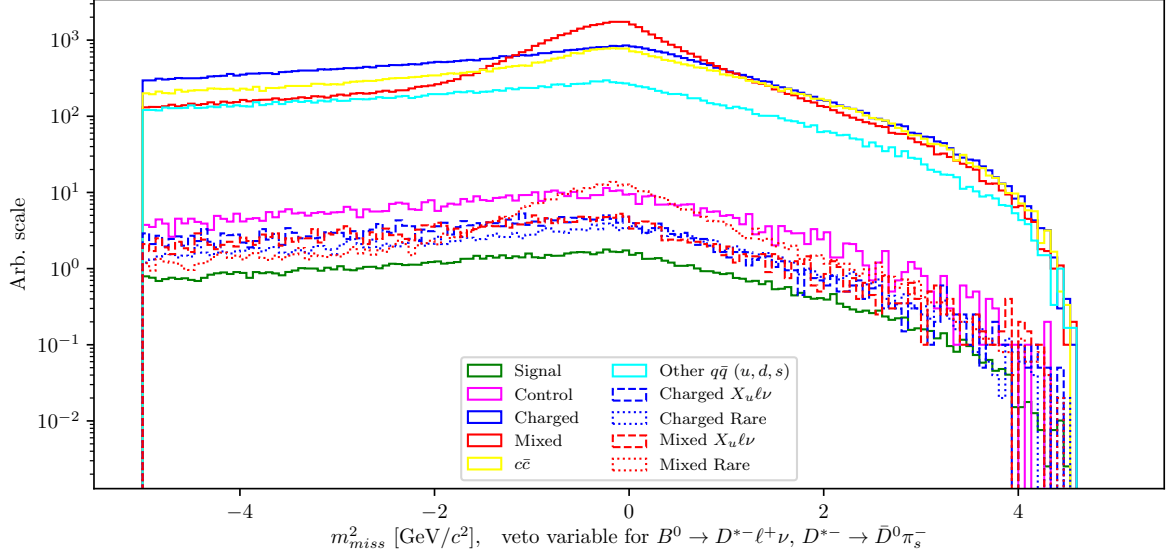
Figure 6.5: Distribution of m_{miss}^2 for partially reconstructed $B^0 \to D^{*-}\ell^+\nu$ decays, which serves as a veto.

When the training is finished and hyper-parameters of the classifier are optimized, the classifier output, 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 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 [10].

The training set used in this training is the same as described in the beginning of this chapter, along with the same set of training variables. It will be seen later that

68

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.
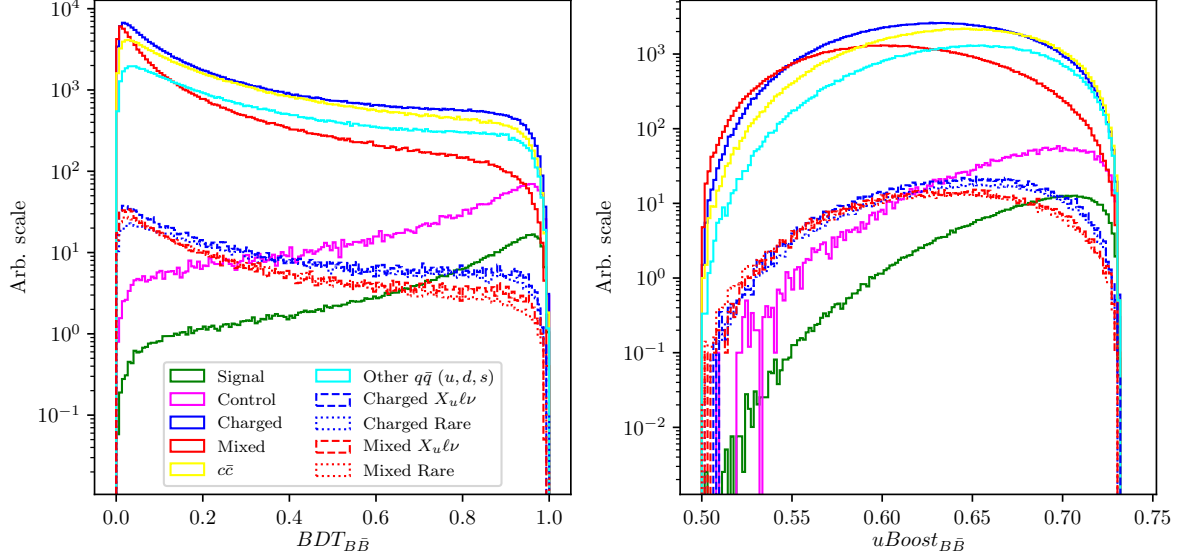

Figure 6.6: $B\bar{B}$ suppression classifier output for signal and various types of background for the standard BDT classifier (left) and the uBDT classifier (right). B candidates from $B\bar{B}$ background events dominate the lower region, while signal and control candidates dominate in the upper region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

69
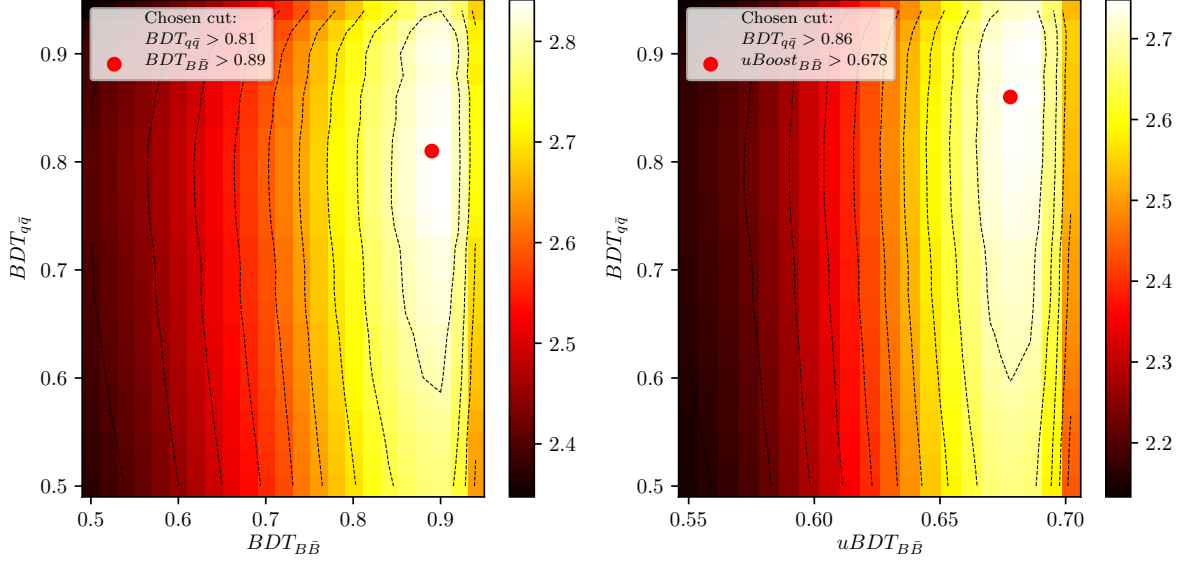
Figure 6.7: 2D FOM optimization of continuum suppression classifier and the standard BDT (left) and uBDT (right) $B\bar{B}$ suppression classifier.

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: $N_{sig} = 176$, $\quad N_{sig}/N_{bkg} = 4.83$ %,

- uBDT: $N_{sig} = 264$, $\quad N_{sig}/N_{bkg} = 1.33$ %.

Due to the large difference in ΔE and M_{BC} shape, we will continue the analysis with the uBDT classifier, although comparison between both methods will be shown for final fit result in the next chapter.
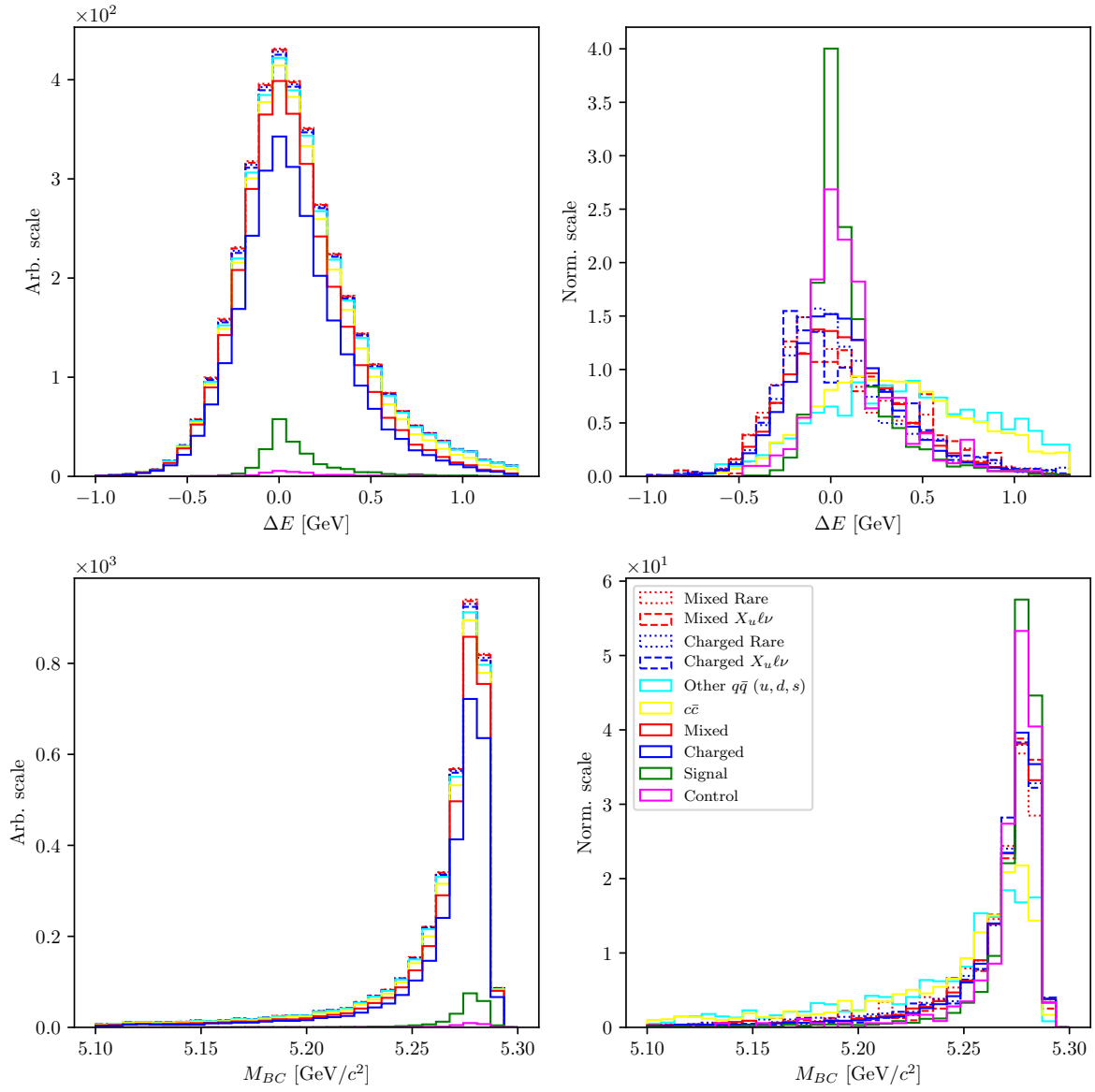
Figure 6.8: Arbitrary (left) and normalized scale (right) for ΔE (top) and M_{BC} (bottom) for the final sample optimized with the standard BDT classifier.
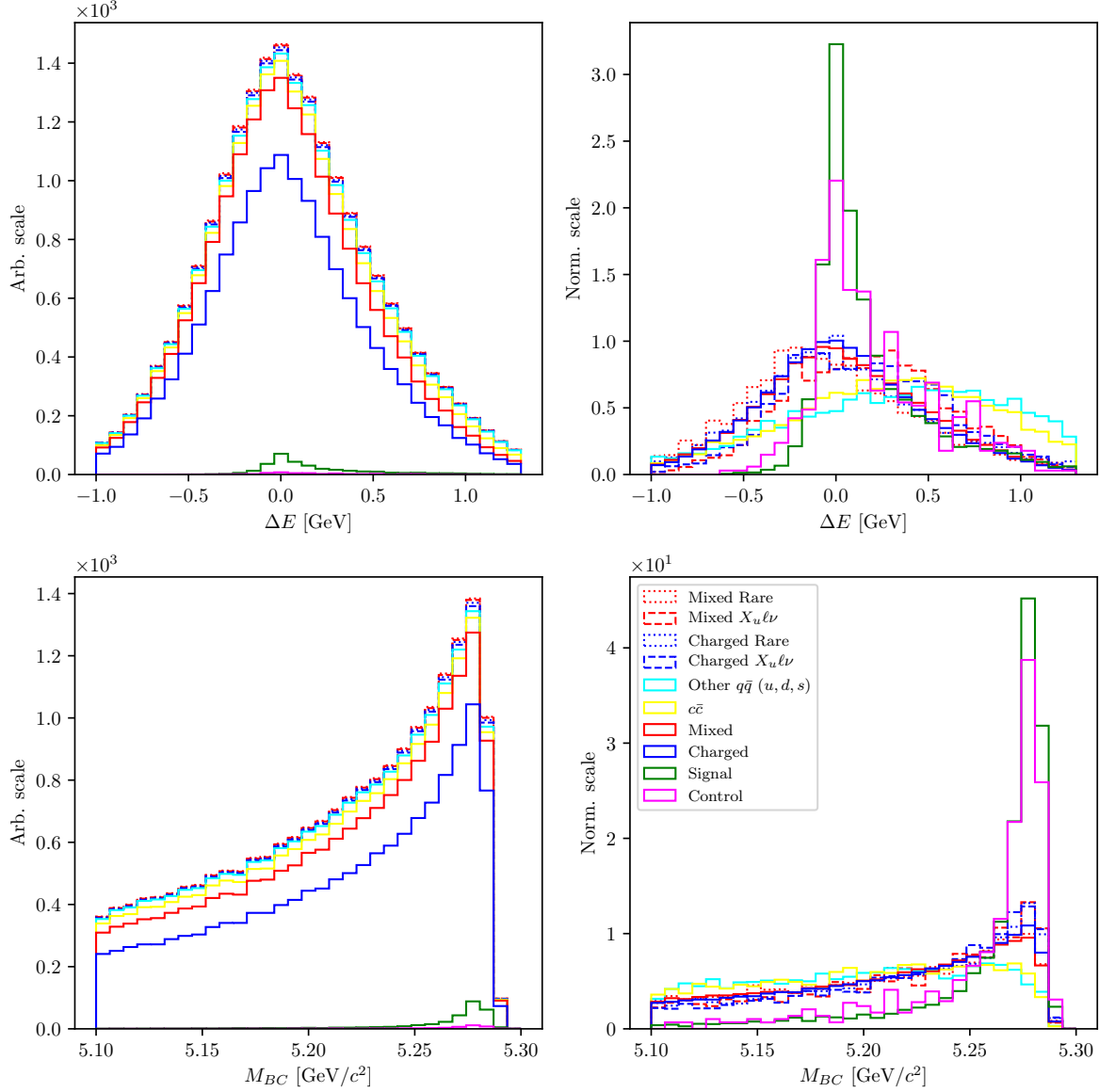
71

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 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 collaborative 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 problematic is the double semileptonic decay $B \to \bar{D}^{(*)}\ell^+\nu$, $D^+ \to \bar{K}^-\ell^+\nu$, where the secondary lepton is misidentified as a kaon. Even though the decay has two neutrinos, these events survive the m^2_{miss} selection cut and produce peaks at the same positions as signal distributions, while exhibiting only a slightly worse resolution.
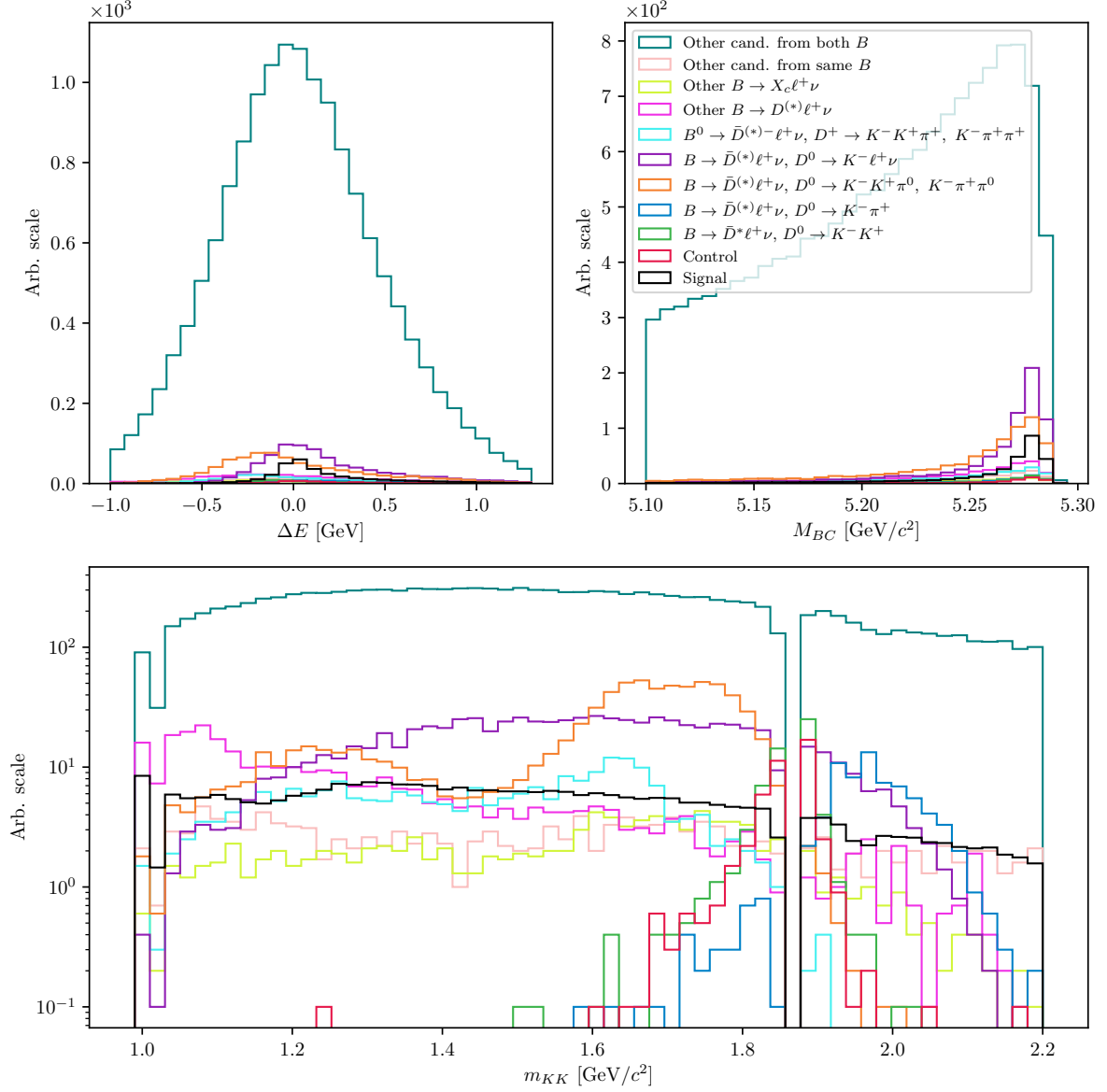
Figure 6.10: ΔE (left), M_{BC} (right) and m_{KK} (bottom) for major contributions to the $B\bar{B}$ background in the signal cut region. $\Upsilon(4S)$-matched backgrounds represents the majority, but has a smooth and wide background, which has a distinguishable shape from signal. B-matched contributions show a peak in M_{BC} and sometimes in ΔE, but can be constrained using existing measurements.

In order to suppress these candidates, a lepton veto can be applied on both kaons, stating that neither of the kaons should exhibit lepton-like properties. On the candidates passing the final selection, we optimize the eID and μID PID cuts, where S and B in Eq. 4.1

are represented by perfect signal candidates and by background candidates, respectively, where in the latter case a lepton has been missidentified as a kaon. 2D FOM plots for both kaons are shown in Figure 6.11, where it can be seen that in majority of the cases, an electron is missidentified as the kaon with the opposite charge to the B meson. With the optimal cuts of

- $K_0 :\ eID < 0.7,$

- $K_1 :\ eID < 0.1,\ \mu ID < 0.8,$

we reject 77.5% of candidates from the double semileptonic decays, while efficiency loss of signal and other types of $B\bar{B}$ background is about $5-6\%$. The $B\bar{B}$ background after the lepton veto cuts is shown in Figure 6.12 for the signal cut region, and in Figure 6.13 for the control cut region.
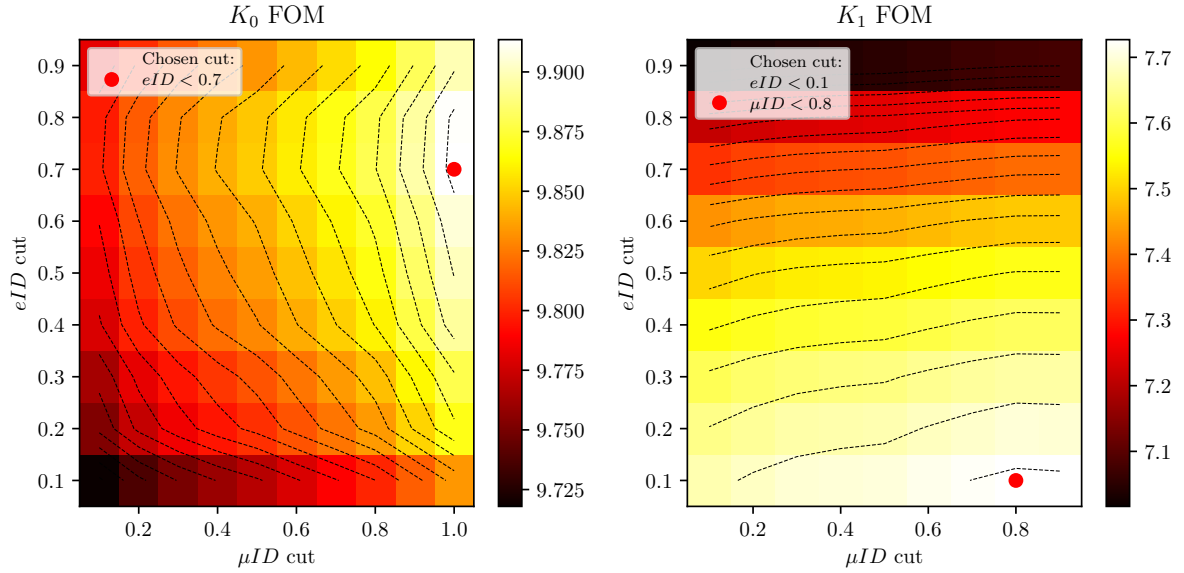
Figure 6.11: 2D FOM for optimal eID and μID cuts on same-sign (left) and opposite-sign (right) kaons with respect to the B meson charge. For double semileptonic background component, in most cases an electron is missidentified as the opposite-sign kaon in the reconstruction chain.
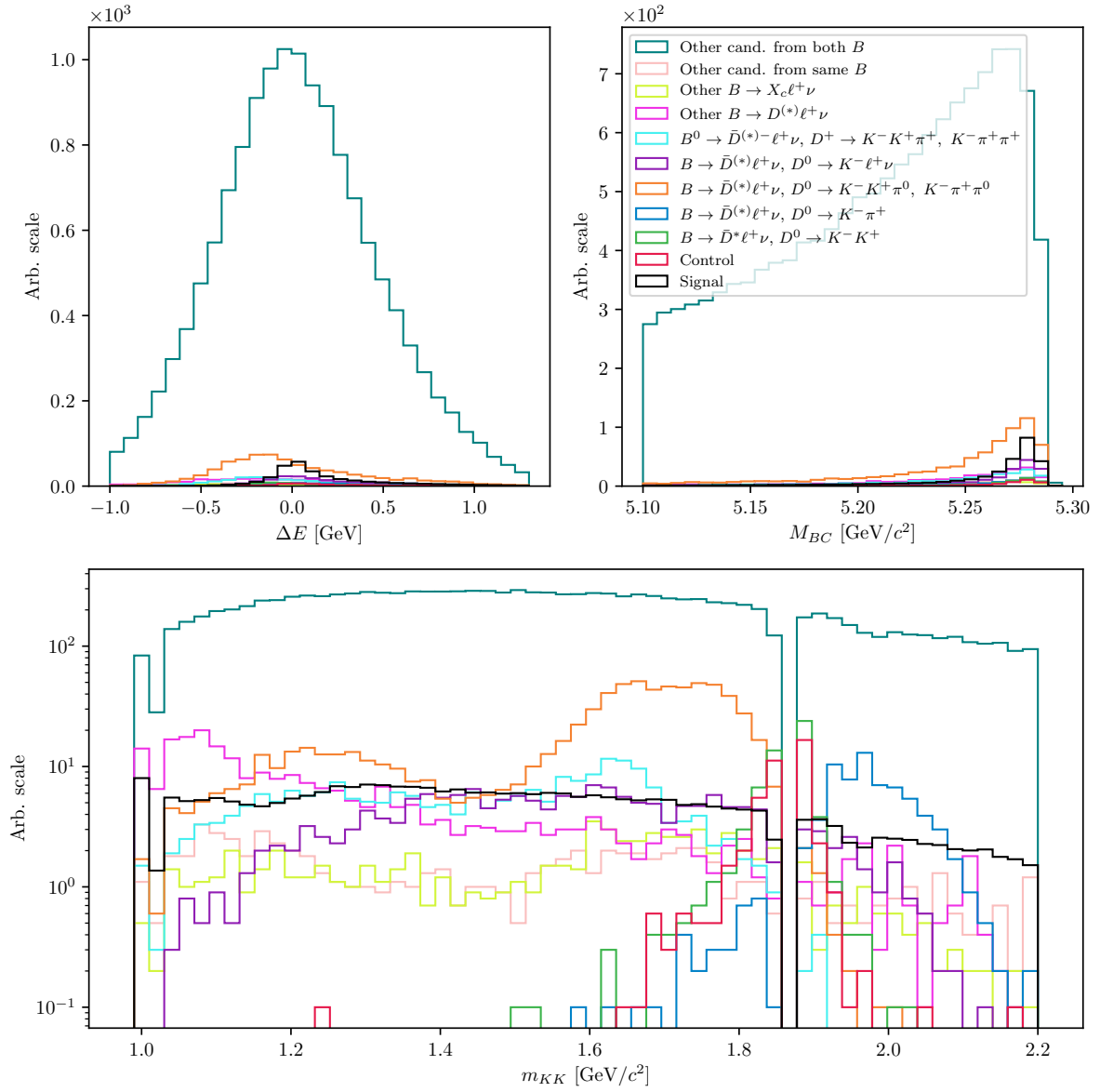
Figure 6.12: ΔE (left), M_{BC} (right) and m_{KK} (bottom) for major contributions to the $B\bar{B}$ background in the signal cut region after the lepton veto. The double semileptonic background component is suppressed by a factor of $4 - 5$.
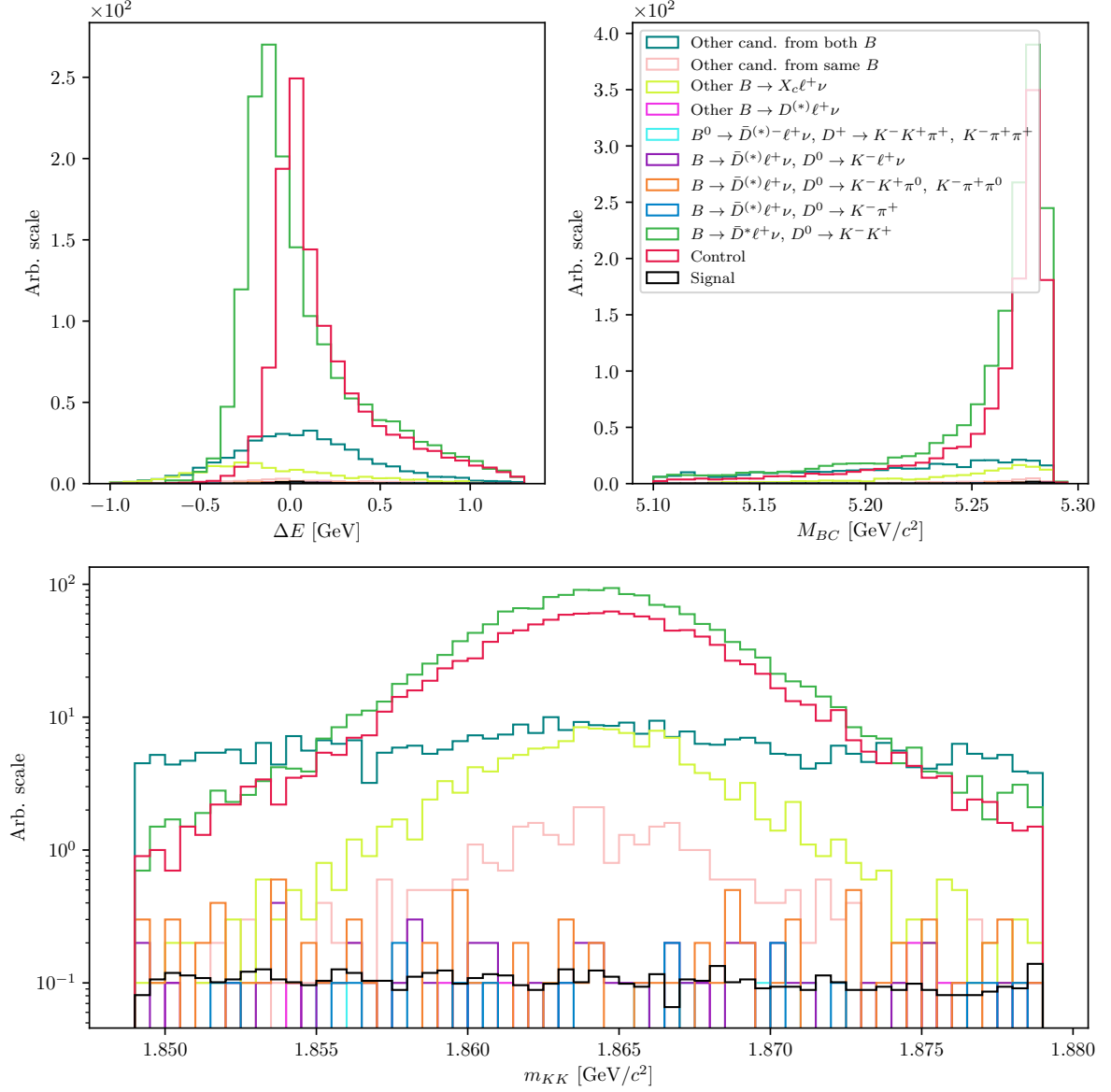
Figure 6.13: ΔE (left), M_{BC} (right) and m_{KK} (bottom) for major contributions to the $B\bar{B}$ background in the control cut region after the lepton veto. The major component in this case are other $B \to D^* \ell + \nu$, $D \to K^+ K^-$ decays, besides the control decay.

6.5 Data and MC agreement

With the final selection in place, we can check the data and MC agreement by checking the control decay region in on- and off- resonance data. Off-resonance samples provide

the ability to check the agreement of the $q\bar{q}$ background component, while on-resonance samples can be used to check the validity of the control MC sample and, consequentially, the signal MC sample.

6.5.1 Off-resonance data

The off-resonance data were collected at 60 MeV below the $\Upsilon(4S)$ resonance peak energy in order to determine the non-$B\bar{B}$ background. It therefore offers a direct view of the $q\bar{q}$ background data sample, which we can compare to the off-resonance MC sample. Figure 6.14 shows ΔE, M_{BC} and the $q\bar{q}$ classifier output, $BDT_{q\bar{q}}$, for off-resonance data and MC in the control region, before any MVA cuts, where the MC sample was scaled down by a factor of 6, due to 6 streams of MC. Figures show good data and MC agreement of the off-resonance sample. More importantly than the normalization, the shape of data and MC also seems to be the same, so further corrections of ΔE and M_{BC} on MC are not necessary. This is also argumented by the flatness of the ratio function for ΔE and M_{BC}, shown in the same Figure. There seems to be a difference in the classifier performance for the continuum background suppression on data and MC. This leads to further differences between data and MC in the $q\bar{q}$ sample after the classifier cut, but we estimate that these differences are negligible, since a relatively small amount of continuum background passes the selection, compared to other background types.
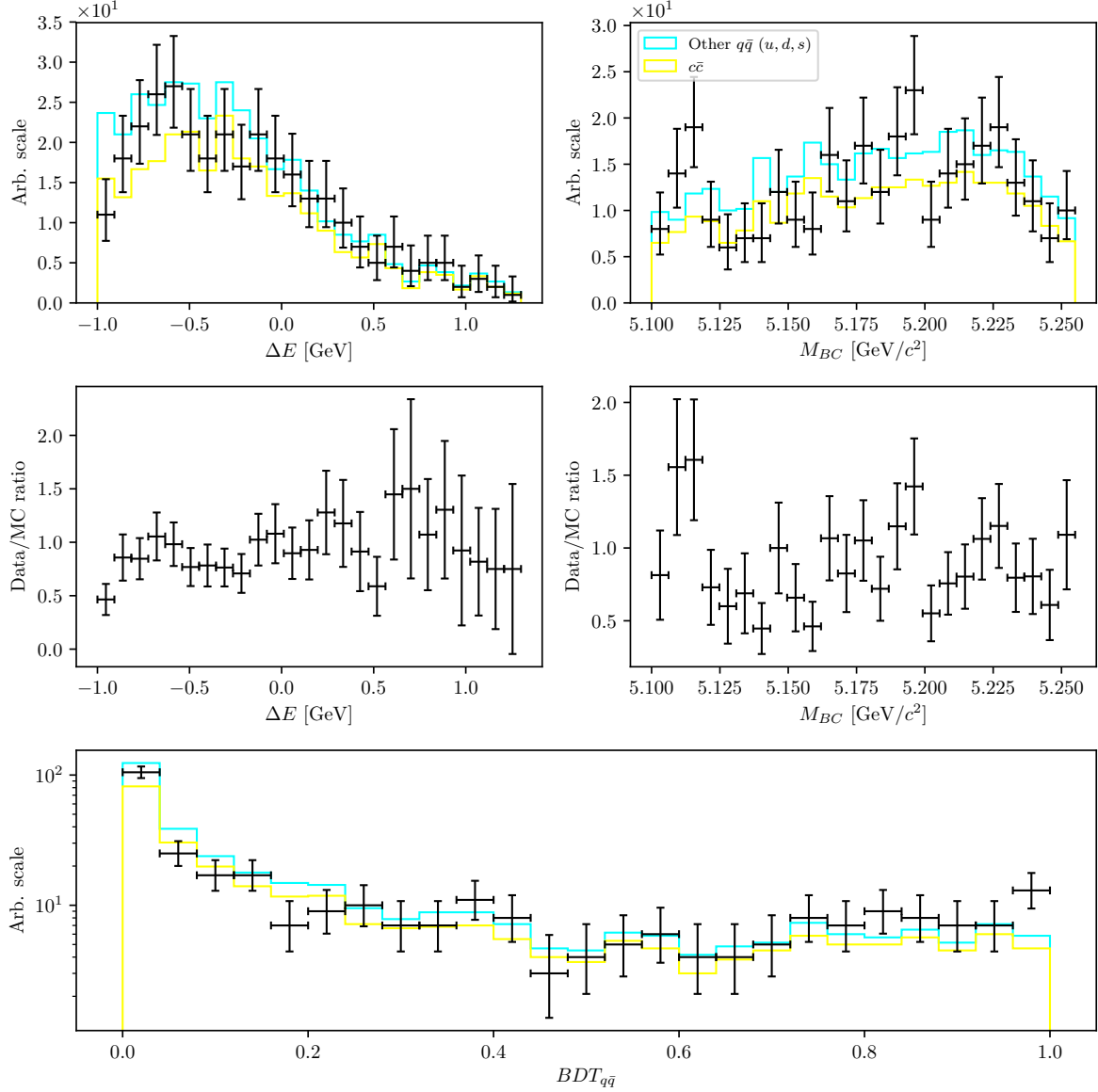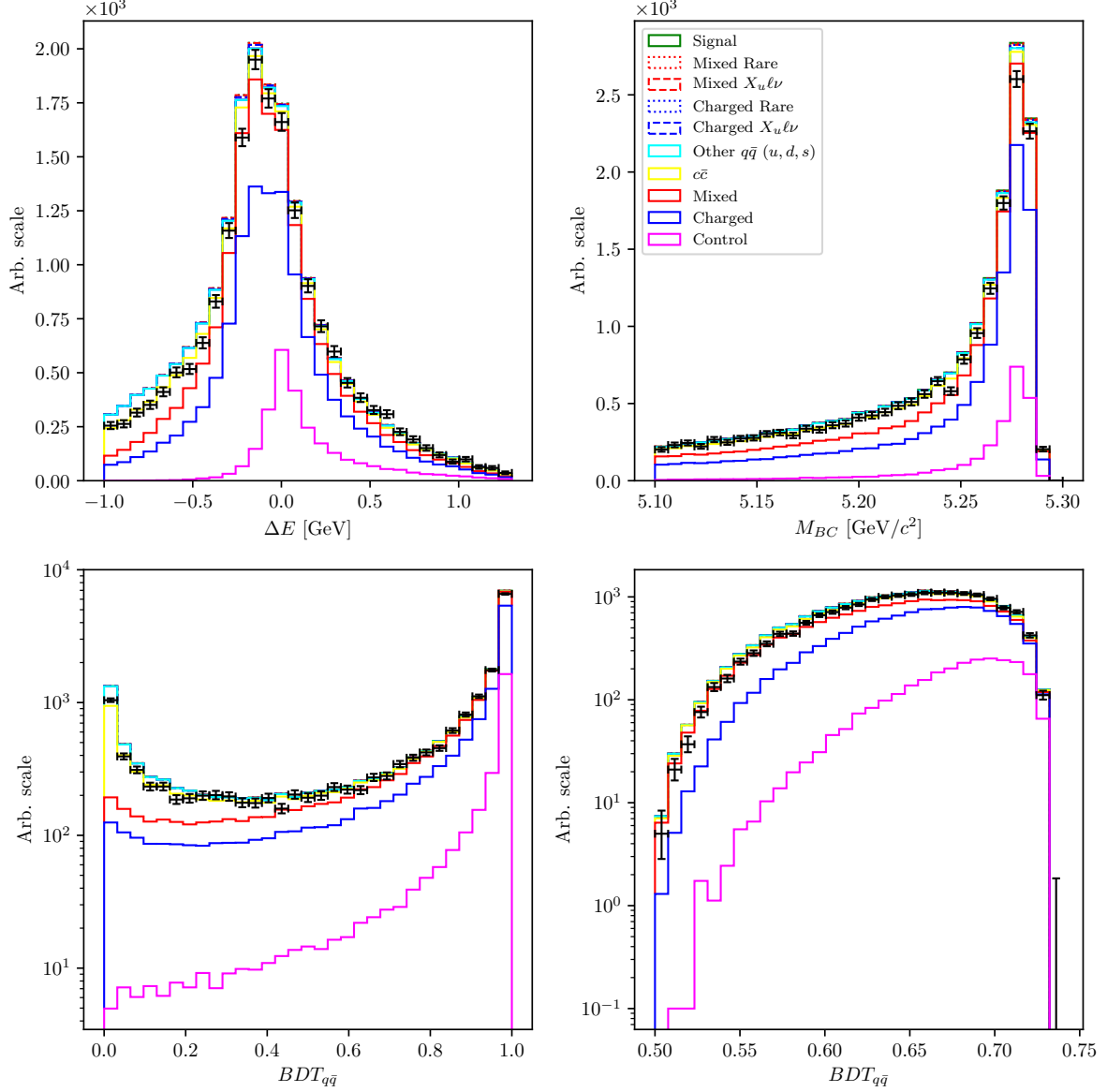
Figure 6.14: ΔE (left), M_{BC} (right) and the $q\bar{q}$ classifier output (bottom), for off-resonance data and MC in the control region before any MVA cuts.

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 MC on the lower spectrum, where continuum background is dominant. We see that MC is over-estimated in this region, 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 an 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

Signal extraction

In this chapter the procedure for signal yield extraction is presented. We use the framework of `RooFit` [11] 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 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

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

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

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

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 [12], 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

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

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

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

7.1 Fit Setup

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

The same MC samples are used for template construction as described in Chapter 2,

- signal MC,

- 10 streams of `charged` and `mixed` $B\bar{B}$ background,

- 6 streams of $c\bar{c}$ (`charm`) and other $q\bar{q}$ (`uds`) background,

- `ulnu` sample, corresponding to $20\times$ integrated luminosity of the full Belle dataset,

- `rare` sample, corresponding to $50\times$ integrated luminosity of the full Belle dataset.

7.1.1 Control Fit

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

- signal template,

- $q\bar{q}$ template,

- C_0 : $B^+ \to \bar{D}^0 \ell^+ \nu$, $D^0 \to K^-K^+$ (control decay),

- C_1 : $B \to \bar{D}^* \ell^+ \nu$, $D^0 \to K^-K^+$,

- other $B\bar{B}$ BKG template.

In control fits, all template shapes are fixed and the yields of all templates are floated, except for the signal template, which is expected to be very close to zero and is fixed to the expected MC yield. Additionally, since the C_0 and C_1 decays are well known and measured, we make use of this fact in the form of a ratio N_{C_1}/N_{C_0}, which is fixed to the MC value in case of the MC fit and constrained to the measured value in case of fits to real data. The ratio is implemented based on the decay channels shown in Table 7.1 and is defined as

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

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

$$\rho = \frac{\mathcal{B}^{PDG}}{\mathcal{B}^{GEN}}. \tag{7.6}$$

The branching ratio correction factor has been implemented due to differences between measured and MC branching ratio values. Each branching ratio measurement serves as a constraint used in the fit. All branching ratio constraints in the control fit are shown in Table 7.2. The measured values are cited only for the B^0 decay mode, where isospin symmetry has been assumed. The corresponding B^+ branching ratios can be calculated as

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

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

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

In case of MC fits, the fitted sample is also generated with MC, so $\mathcal{B}_i^{PDG} = \mathcal{B}_i^{MC}$ and Eq. (7.5) simplifies to a simple MC yield ratio. On fits to real data, expected MC yields and

Category	Channel	B Decay Mode	D Decay Mode	N_{MC}
C_0	N_{00}	$B^+ \to \bar{D}^0 \ell^+ \nu$	$D^0 \to K^- K^+$	1182 ± 34
C_1	N_{10}	$B^+ \to \bar{D}^{*0} \ell^+ \nu$	$D^0 \to K^- K^+$	1455 ± 38
	N_{11}	$B^0 \to D^{*-} \ell^+ \nu$	$D^0 \to K^- K^+$	186 ± 16

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

branching ratio measurements are implemented as independent Gaussian constraints in order to properly account for correlations in Eq. (7.5).

Decay	$\mathcal{B}_{GEN}$	$\mathcal{B}_{PDG}$	Reference
$B^0 \to D^- \ell^+ \nu$	2.13×10^{-2}	$(2.13 \pm 0.09) \times 10^{-2}$	[X]
$B^0 \to D^{*-} \ell^+ \nu$	5.33×10^{-2}	$(4.88 \pm 0.11) \times 10^{-2}$	[X]
$B^+ \to \bar{D}^0 \ell^+ \nu$	2.31×10^{-2}	$(2.29 \pm 0.10) \times 10^{-2}$	[X]
$B^+ \to \bar{D}^{*0} \ell^+ \nu$	5.79×10^{-2}	$(5.25 \pm 0.12) \times 10^{-2}$	[X]
$D^0 \to K^- K^+$	3.9×10^{-3}	$(3.97 \pm 0.07) \times 10^{-3}$	[X]

Table 7.2: MC and measured values of branching ratios used for constraining the control fit.

Smearing and Offset Parameters

With simulated data we are able to perform detailed studies prior to looking at 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 perform a scan over two additional parameters of offset and smearing, applied on the ΔE variable. In 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

- 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 case of data, we see that we obtain a better fit by introducing some level of smearing and offset. In neither of the cases the two parameters show any sign of significant correlation, so we treat them as independent parameters. 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^{+15}_{-17} MeV,

- Offset: $6^{+4.6}_{-6}$ MeV.

We apply this transformation to our MC samples in all cases when fitting real data.
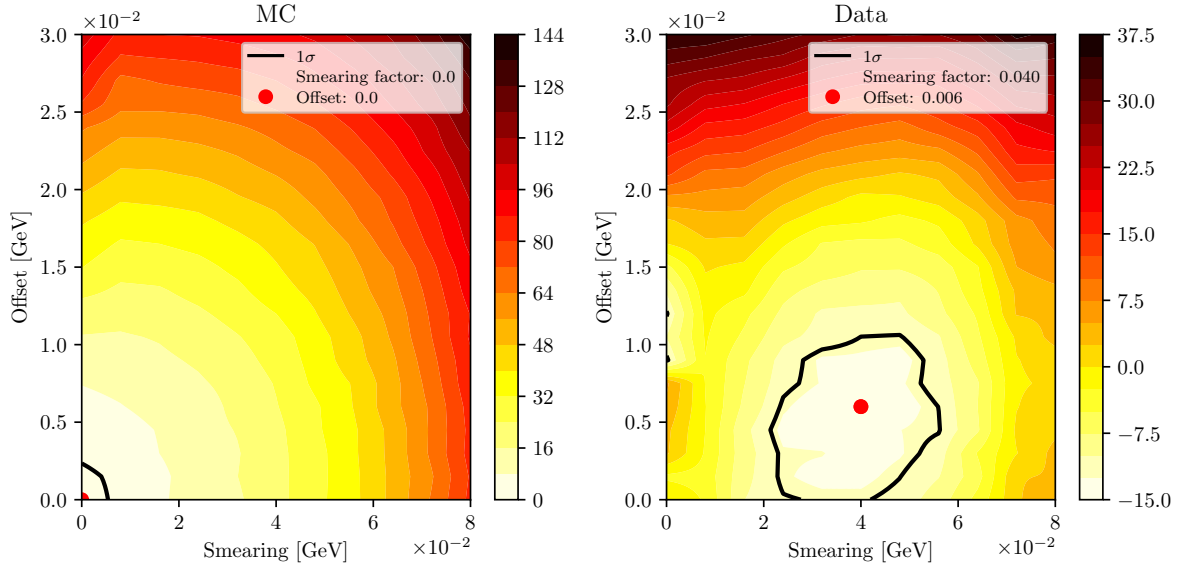
Figure 7.1: Likelihood ratio test of an additional smearing and offset parameter to MC (left) and data (right).

7.1.2 Signal Fit

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

- signal template,

- $q\bar{q}$ template,

- a series of well defined templates from $B\bar{B}$ background:

    - $C_0:\ B^+ \to \bar{D}^0\ell^+\nu,\ D^0 \to K^-K^+$ (control decay),

1209 $-\ C_1:\ B \to \bar{D}^* \ell^+ \nu,\ D^0 \to K^- K^+,$

1210 $-\ C_2:\ B \to \bar{D}^{(*)} \ell^+ \nu,\ D^0 \to K^- \pi^+,$

1211 $-\ C_3:\ B \to \bar{D}^{(*)} \ell^+ \nu,\ D^0 \to K^- K^+ \pi^0,\ K^- \pi^+ \pi^0,$

1212 $-\ C_4:\ B \to \bar{D}(*) \ell^+ \nu,\ D^0 \to K^- \ell^+ \nu,$

1213 $-\ C_5:\ B^0 \to D^{(*)-} \ell^+ \nu,\ D^+ \to K^- K^+ \pi^+,\ K^- \pi^+ \pi^+,$

1214 $-\ C_6:\ $ other $B \to \bar{D}^{(*)} \ell^+ \nu$ decays,

1215   • remaining $B\bar{B}$ background template.

1216 As mentioned in Chapter 6, the majority of the background comes from $B\bar{B}$ events.
1217 Various processes (C_0 to C_6) contribute to this background which are well known and
1218 measured, so we make use of these measurements by fixing their yields in MC fits and
1219 appropriately constraining them in real data fits. The remaining $B\bar{B}$ background is
1220 merged into a single template. In this case, the shape of all templates is fixed as well,
1221 while the yields are floated for all templates except for the constrained background
1222 templates. The yield constraints are based on the channels shown in Table 7.3 and
1223 defined for each template category as

$$Y_i = \eta_{\text{norm.}} \times \frac{\left(\sum_j N_{ij} \times \rho_{ij} \right)}{\rho_{00}}, \tag{7.9}$$

1224 where j runs over all channels in the category C_i and where ρ_{ij} is defined in Eq. (7.6).
1225 The first factor, $\eta_{\text{norm.}}$, serves as a normalization factor in order to scale the number of
1226 generated $B\bar{B}$ events to the number of $B\bar{B}$ events in measured data. We define it as

$$\eta_{\text{norm.}} = \frac{N_{\text{control}}^D}{N_{\text{control}}^{MC}}, \tag{7.10}$$

1227 where N_{control}^D and N_{control}^{MC} are control yields in the control fit for data and MC, respec-
1228 tively.

1229 In addition to branching ratio constraints in Table 7.2, further constraints are defined
1230 in Table 7.4 due to more decay channels. In case of the category C_6, we have no firm
1231 handle on the D meson decay, therefore no correction for this branching ratio can be
1232 introduced, so we set a correction factor of 1 with a 100% error for the D meson decay
1233 branching ratio.

Category	Channel	B Decay Mode	D Decay Mode	Expected MC Yield
C_0	N_{00}	$B^+ \to \bar{D}^0 \ell^+ \nu$	$D^0 \to K^- K^+$	44 ± 7
C_1	N_{10}	$B^+ \to \bar{D}^{*0} \ell^+ \nu$	$D^0 \to K^- K^+$	53 ± 7
	N_{11}	$B^0 \to D^{*-} \ell^+ \nu$	$D^0 \to K^- K^+$	6 ± 2
C_2	N_{20}	$B^+ \to \bar{D}^0 \ell^+ \nu$	$D^0 \to K^- \pi^+$	23 ± 5
	N_{21}	$B^+ \to \bar{D}^{*0} \ell^+ \nu$	$D^0 \to K^- \pi^+$	41 ± 6
	N_{22}	$B^0 \to D^{*-} \ell^+ \nu$	$D^0 \to K^- \pi^+$	6 ± 2
C_3	N_{30}	$B^+ \to \bar{D}^0 \ell^+ \nu$	$D^0 \to K^- K^+ \pi^0$	103 ± 10
	N_{31}	$B^+ \to \bar{D}^0 \ell^+ \nu$	$D^0 \to K^- \pi^+ \pi^0$	211 ± 15
	N_{32}	$B^+ \to \bar{D}^{*0} \ell^+ \nu$	$D^0 \to K^- K^+ \pi^0$	135 ± 12
	N_{33}	$B^+ \to \bar{D}^{*0} \ell^+ \nu$	$D^0 \to K^- \pi^+ \pi^0$	267 ± 16
	N_{34}	$B^0 \to D^{*-} \ell^+ \nu$	$D^0 \to K^- K^+ \pi^0$	19 ± 4
	N_{35}	$B^0 \to D^{*-} \ell^+ \nu$	$D^0 \to K^- \pi^+ \pi^0$	36 ± 6
C_4	N_{40}	$B^+ \to \bar{D}^0 \ell^+ \nu$	$D^0 \to K^- e^+ \nu$	48 ± 7
	N_{41}	$B^+ \to \bar{D}^0 \ell^+ \nu$	$D^0 \to K^- \mu^+ \nu$	7 ± 3
	N_{42}	$B^+ \to \bar{D}^{*0} \ell^+ \nu$	$D^0 \to K^- e^+ \nu$	98 ± 10
	N_{43}	$B^+ \to \bar{D}^{*0} \ell^+ \nu$	$D^0 \to K^- \mu^+ \nu$	10 ± 3
	N_{44}	$B^0 \to D^{*-} \ell^+ \nu$	$D^0 \to K^- e^+ \nu$	14 ± 4
	N_{45}	$B^0 \to D^{*-} \ell^+ \nu$	$D^0 \to K^- \mu^+ \nu$	3 ± 2
C_5	N_{50}	$B^0 \to D^- \ell^+ \nu$	$D^+ \to K^- K^+ \pi^+$	103 ± 10
	N_{51}	$B^0 \to D^- \ell^+ \nu$	$D^+ \to K^- \pi^+ \pi^+$	63 ± 8
	N_{52}	$B^0 \to D^{*-} \ell^+ \nu$	$D^+ \to K^- K^+ \pi^+$	31 ± 6
	N_{53}	$B^0 \to D^{*-} \ell^+ \nu$	$D^+ \to K^- \pi^+ \pi^+$	21 ± 5
C_6	N_{60}	$B^+ \to \bar{D}^0 \ell^+ \nu$	Other D^0 and D^+ decays	69 ± 8
	N_{61}	$B^+ \to \bar{D}^{*0} \ell^+ \nu$		95 ± 10
	N_{62}	$B^0 \to D^- \ell^+ \nu$		63 ± 8
	N_{63}	$B^0 \to D^{*-} \ell^+ \nu$		36 ± 6

Table 7.3

Decay	$\mathcal{B}_{GEN}$	$\mathcal{B}_{PDG}$	Reference
$D^0 \to K^- K^+$	3.9×10^{-3}	$(3.97 \pm 0.07) \times 10^{-3}$	[X]
$D^0 \to K^- \pi^+$	3.82×10^{-2}	$(3.89 \pm 0.04) \times 10^{-2}$	[X]
$D^0 \to K^- K^+ \pi^0$	2.36×10^{-3}	$(3.37 \pm 0.15) \times 10^{-3}$	[X]
$D^0 \to K^- \pi^+ \pi^0$	13.08×10^{-2}	$(14.2 \pm 0.5) \times 10^{-2}$	[X]
$D^0 \to K^- e^+ \nu$	3.41×10^{-2}	$(3.53 \pm 0.028) \times 10^{-2}$	[X]
$D^0 \to K^- \mu^+ \nu$	3.41×10^{-2}	$(3.31 \pm 0.13) \times 10^{-2}$	[X]
$D^+ \to K^- K^+ \pi^+$	9.06×10^{-3}	$(9.51 \pm 0.34) \times 10^{-3}$	[X]
$D^+ \to K^- \pi^+ \pi^+$	9.51×10^{-2}	$(8.98 \pm 0.28) \times 10^{-2}$	[X]

Table 7.4: Additional MC and measured values of D meson branching ratios 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 statistic 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),

88

4. compare the histograms for the expected and the fitted sample, find the problematic bins,

5. loop until all problematic bins disappear

   a) find problematic bin, which is nearest to the maximum bin,

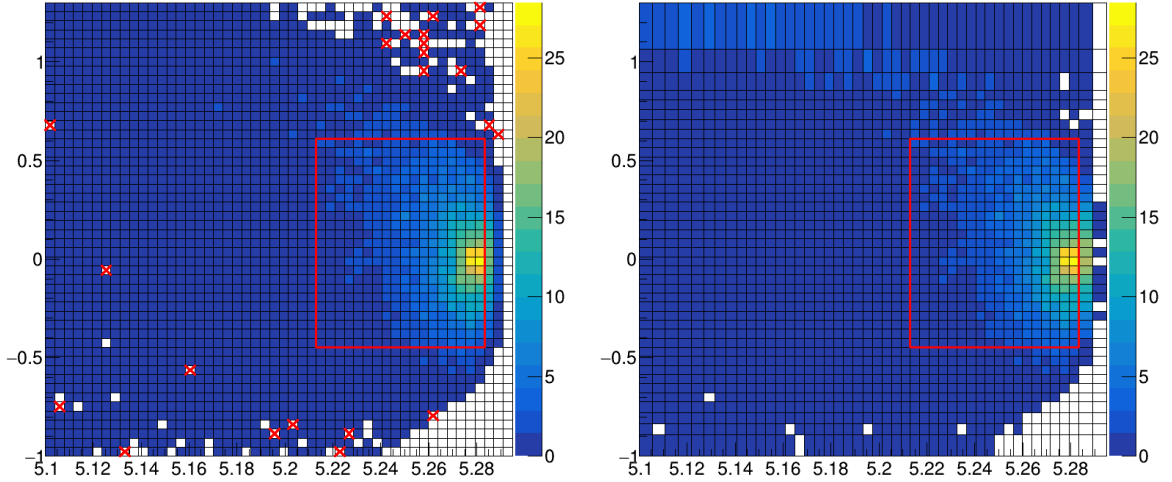   b) change the binning from N to $N - 1$ from that bin and in the direction away from the maximum bin.

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.

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

7.3 Fit Results

In this section we present the fir 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.

89

7.3.1 Signal MC fit results

With the signal fit setup described in 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 7.3 shows the expected yield differences, pulls and fit significances for both final samples for each binning case. The difference between fitted an 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. While both fit results seem to have no bias with respect to the binning choice, the pull distribution seems to be closer to the normal distribution and the significance of the fit better by 1σ in the case of the sample optimized with the uBDT classifier. This fixes our choice of the final selection. The binning in ΔE and M_{BC} is chosen at the plateau of the significance, where also no bias is present, and is somewhere in the region of 20 bins in each dimension.

For the chosen binning of 20×20 in ΔE and M_{BC} we perform the 10 stream MC fits, where an average stream fit is shown in Figure 7.4, while all fit results are shown in Figure 7.5. 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 $\chi^2_9 = 0.99$, while the average significance of the fits is around 3.57σ.

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

7.3.2 Control Fit Result

With the control fit setup described in 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 stream fit is shown in Figure 7.6 for MC and Figure 7.7 for data, while all fit results are shown in Figure 7.8, where all streams of MC were fitted with a 0th degree polynomial. The control fit results for split and joined lepton modes are shown in Table 7.5.
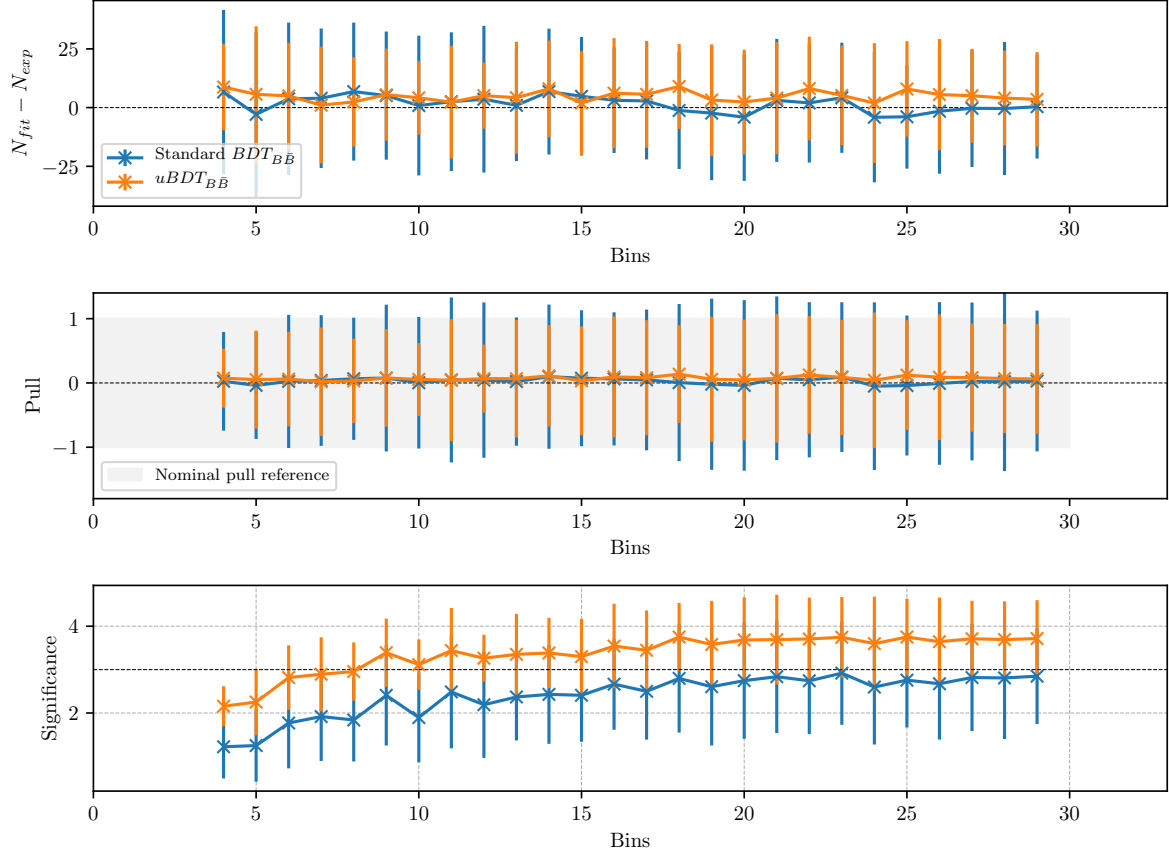
Figure 7.3: Fitted yield and expected yield difference (top), pulls (center) and fit significance (bottom) as a function of binning in ΔE and M_{BC} for the final sample, optimized with the standard BDT classifier (blue) and the uBDT classifier (orange).
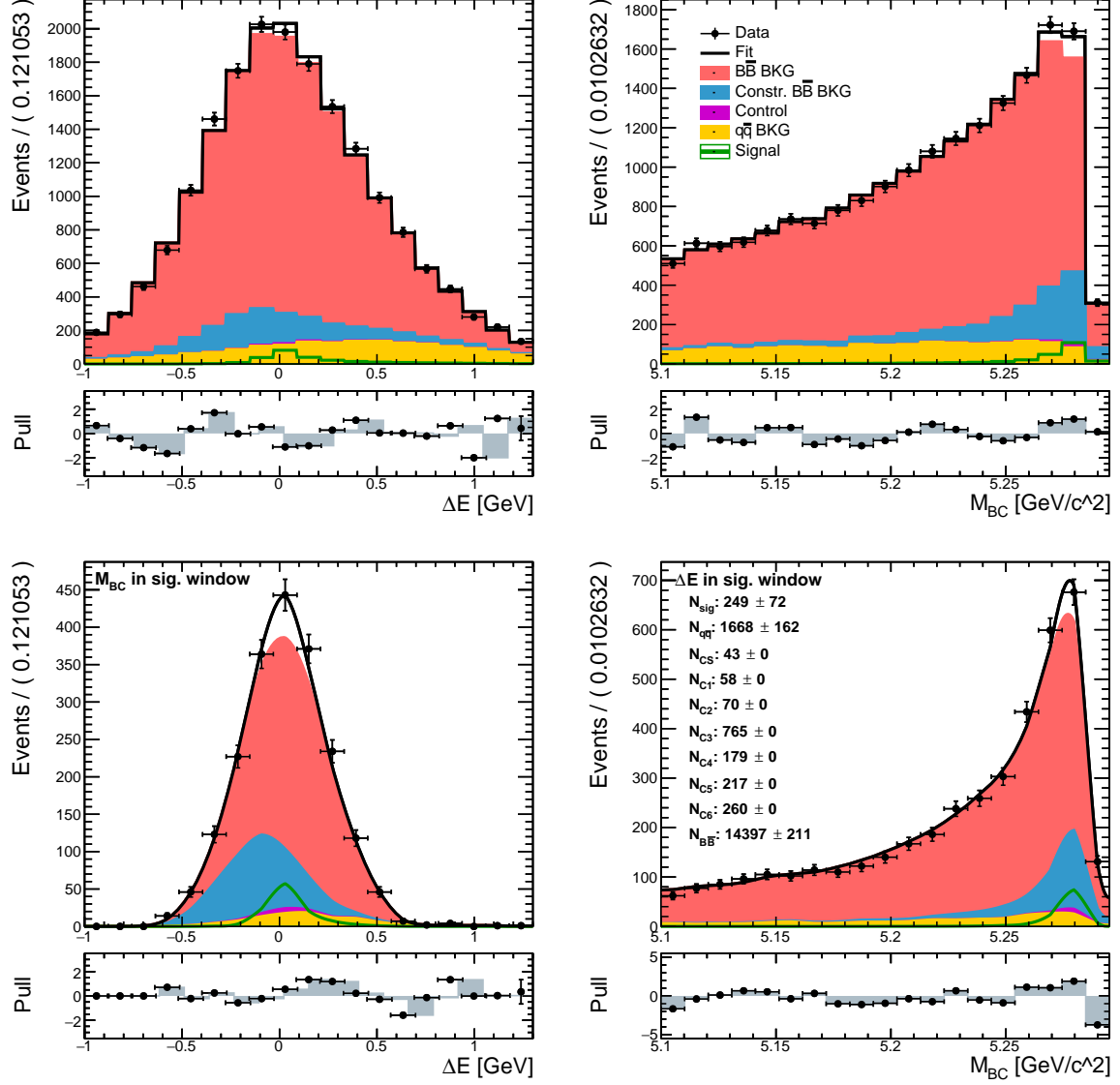
Figure 7.4: An example fit to one stream of MC. Left column shows the M_{BC} and the right column shows the ΔE distribution of the full fitted sample in the full fit region (top) and the signal region (bottom).
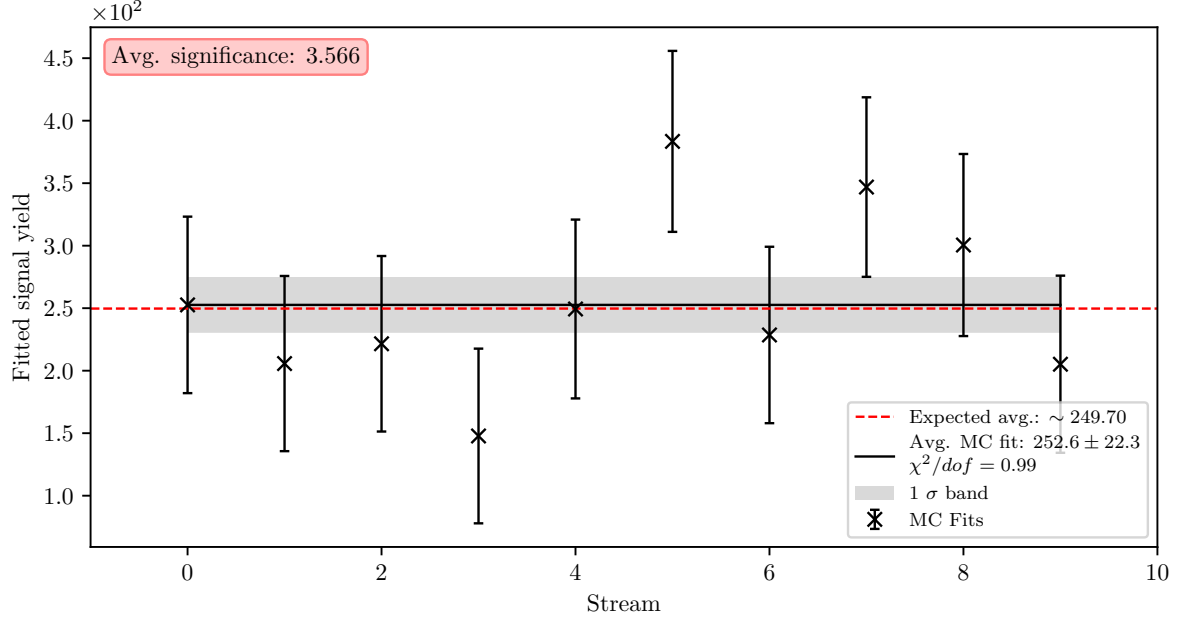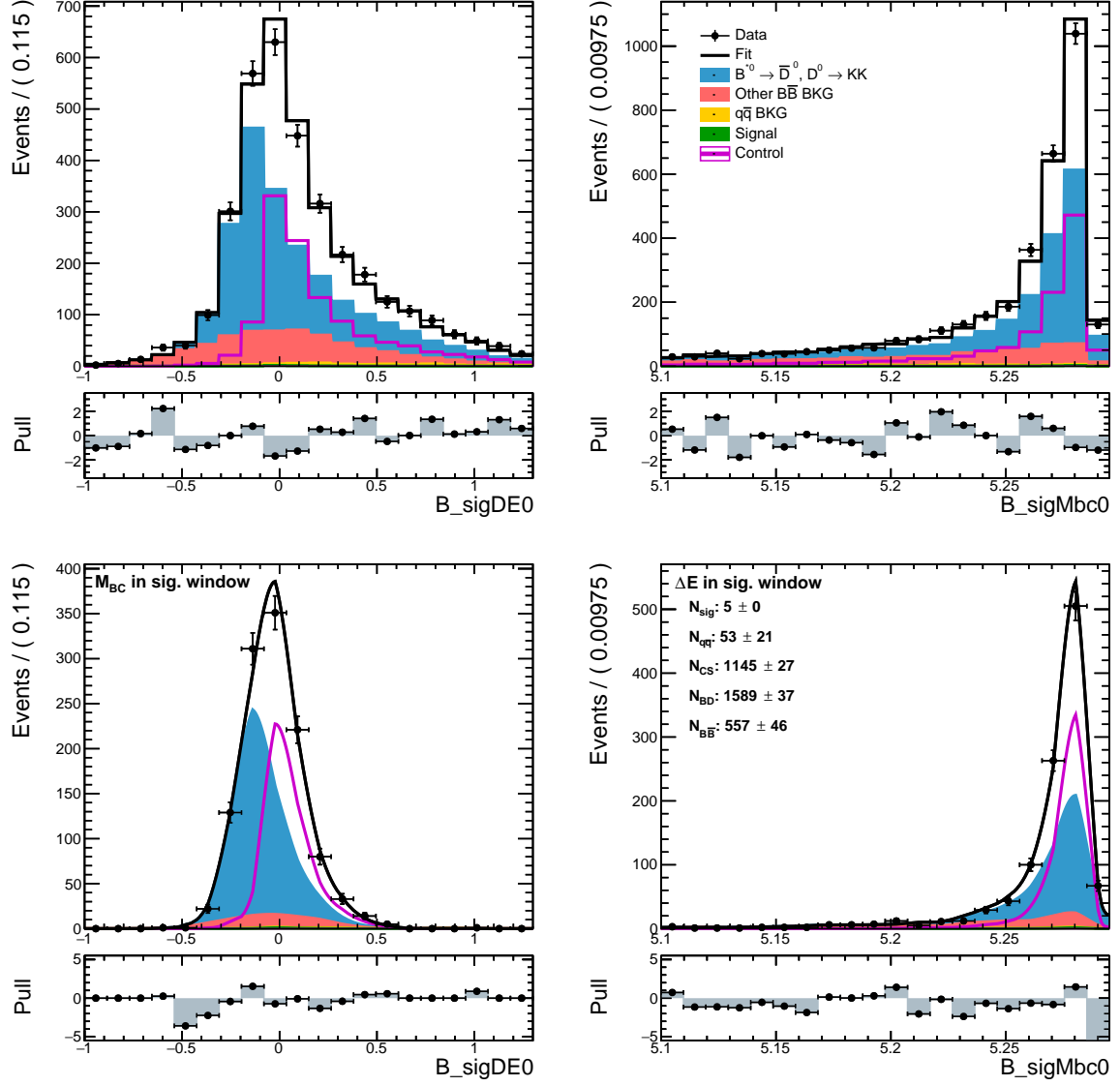
Figure 7.5: Fits to all 10 streams of MC and the global fit with a zero degree polynomial. The red line shows the mean value of the global fit and the gray band shows the 1σ confidence interval.

	N^{MC}	N^{data}
$\ell = e$ or μ	1180 ± 11	1196 ± 44
$\ell = e$	590 ± 8	581 ± 28
$\ell = \mu$	592 ± 7	618 ± 30

Table 7.5: Control sample fit results for MC and data for various lepton final state modes.

7.3.3 Branching Ratio Measurement for Control Decay

After acquiring the fit results on MC and data, we are able to determine the branching ratio of the control decay, which is defined as

$$\mathcal{B}_{\text{control}}^{MC} = \frac{N_{\text{control}}^{\text{MC}} \times \epsilon_{MC}}{2 N_{B\bar{B}}^{MC}}, \tag{7.11}$$

$$\mathcal{B}_{\text{control}} = \frac{N_{\text{control}} \times \epsilon_{MC} \times \rho_{PID}}{2 N_{B\bar{B}}}, \tag{7.12}$$

where $N_{\text{control}}^{\text{MC}}$ and N_{control} are yields of the control fit on MC and data, ϵ_{MC} is the MC efficiency of the control sample, ρ_{PID} the PID correction factor, and $N_{B\bar{B}}^{MC}$ and $N_{B\bar{B}}$ are

Figure 7.6: Control fit result on stream 9 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).
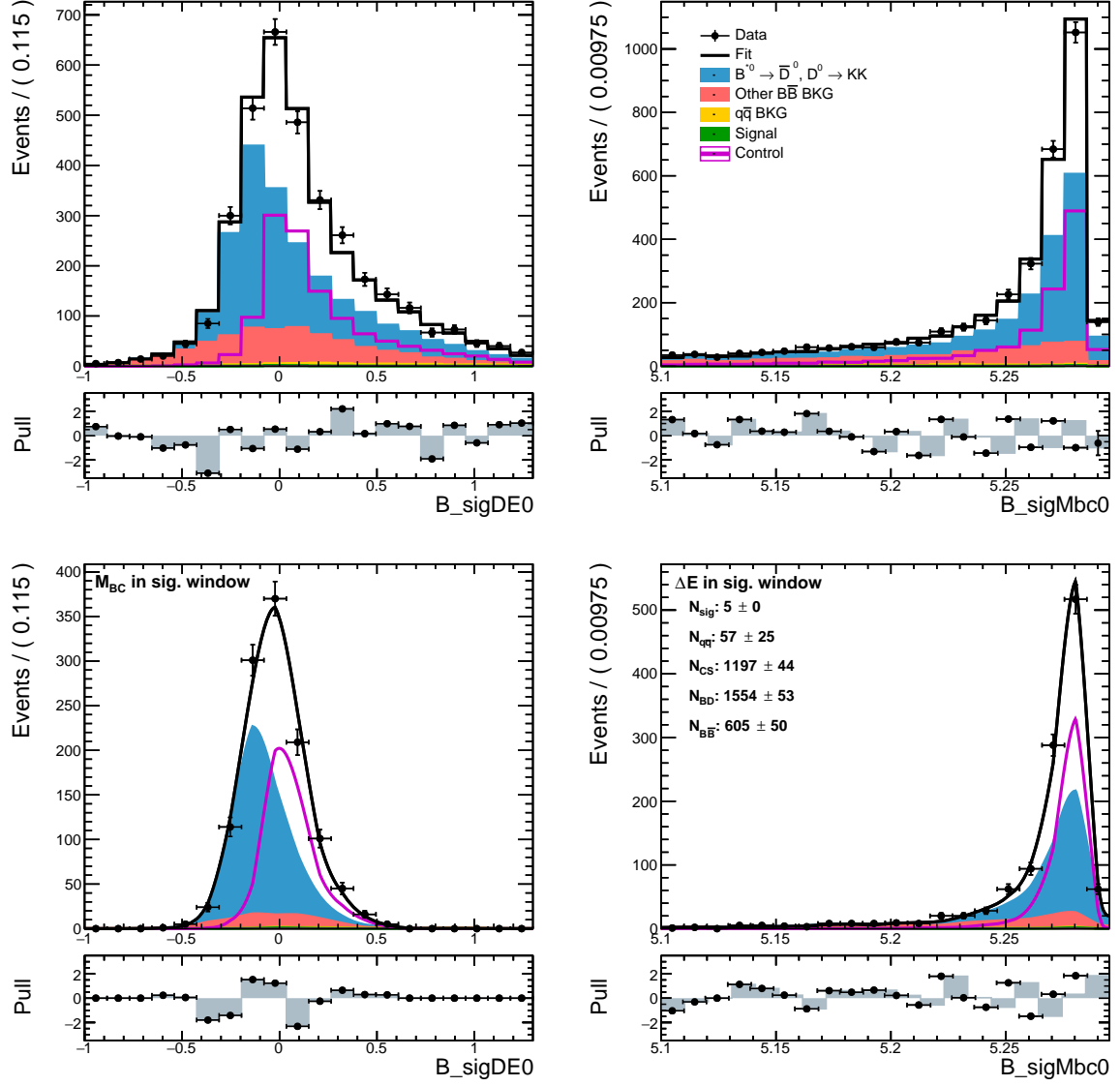
Figure 7.7: Control fit result on real data. Left column shows the M_{BC} and the right column shows the ΔE distribution in the full fit window (top) and the signal window (bottom).
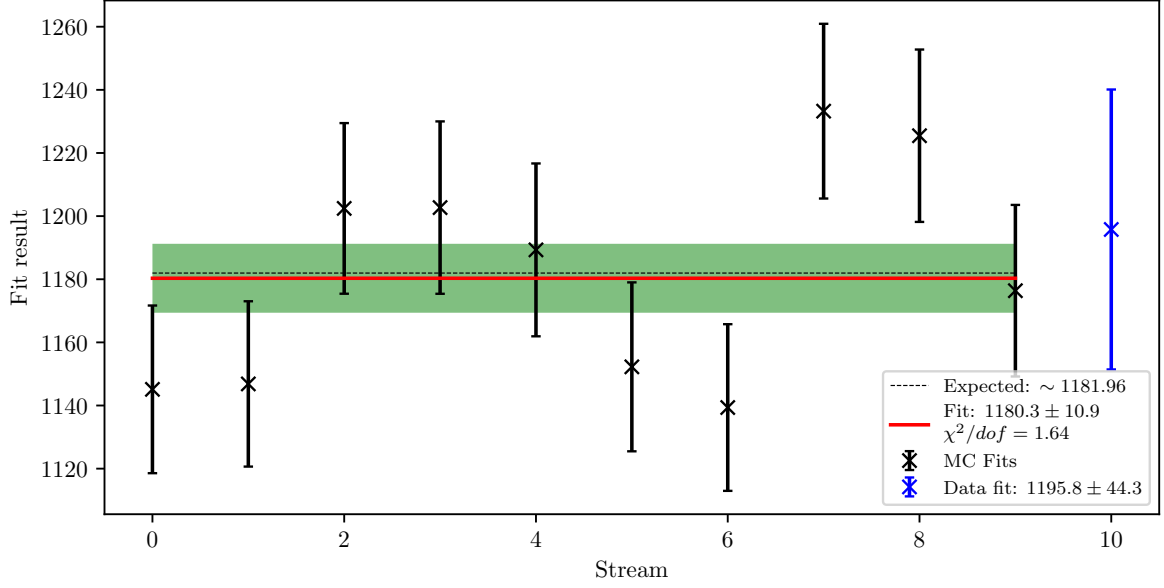
Figure 7.8: Control fit to data and to all 10 streams of MC. The red line shows the mean value of the global MC fit with a 0th degree polynomial. The gray band shows the 1σ confidence interval around the global MC fit.

the numbers of $B\bar{B}$ meson pairs on MC and in data, respectively. The factor of 2 in the denominator comes from the fact that there are 2 B mesons in each B meson pair ($\times 1/2$), where only about 50% of the B meson pairs are charged B^+B^- meson pairs ($\times 2$), and from the fact that we are interested in the branching ratio 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 $B^+ \to \bar{D}^0 \ell^+ \nu$, $D^0 \to K^+K^-$. After applying the final selection, the full and split efficiencies with regard to the lepton final state were determined to be

$$\epsilon_{MC} = (8.89 \pm 0.04) \times 10^{-3},$$
$$\epsilon_{MC}^e = (4.40 \pm 0.03) \times 10^{-3},$$
$$\epsilon_{MC}^\mu = (4.49 \pm 0.03) \times 10^{-3},$$

where the efficiency error was estimated based on a formula from [13].

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

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 X and is determined to be

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

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

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

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

Finally, we can determine the branching ratios based on the calculations in Eq. (7.12). The obtained values are shown in Table 7.6 and graphically shown in Figure 7.9, along with the MC generated value and the current PDG world average. Both MC and measured determination of the control decay branching ratio are in agreement with the expected and the world average values. One should note that the black error bars correspond to statistical uncertainty. Of all the systematic uncertainties, only the PID systematics are included in this results. Other contributions of systematics are not included, since this measurement is not the goal of our analysis.

$\mathcal{B}_{PDG}$	$(9.10 \pm 0.42) \times 10^{-5}$	
$\mathcal{B}_{GEN}$	9.01×10^{-5}	
	$\mathcal{B}^{MC}$	$\mathcal{B}^{\text{data}}$
$\ell = e$ or μ	$(8.94 \pm 0.09) \times 10^{-5}$	$(9.10 \pm 0.4) \times 10^{-5}$
$\ell = e$	$(9.04 \pm 0.14) \times 10^{-5}$	$(9.03 \pm 0.49) \times 10^{-5}$
$\ell = \mu$	$(8.85 \pm 0.12) \times 10^{-5}$	$(9.23 \pm 0.50) \times 10^{-5}$

Table 7.6: Control sample fit results for MC and data for various lepton final state modes.

7.3.4 Signal fit to data
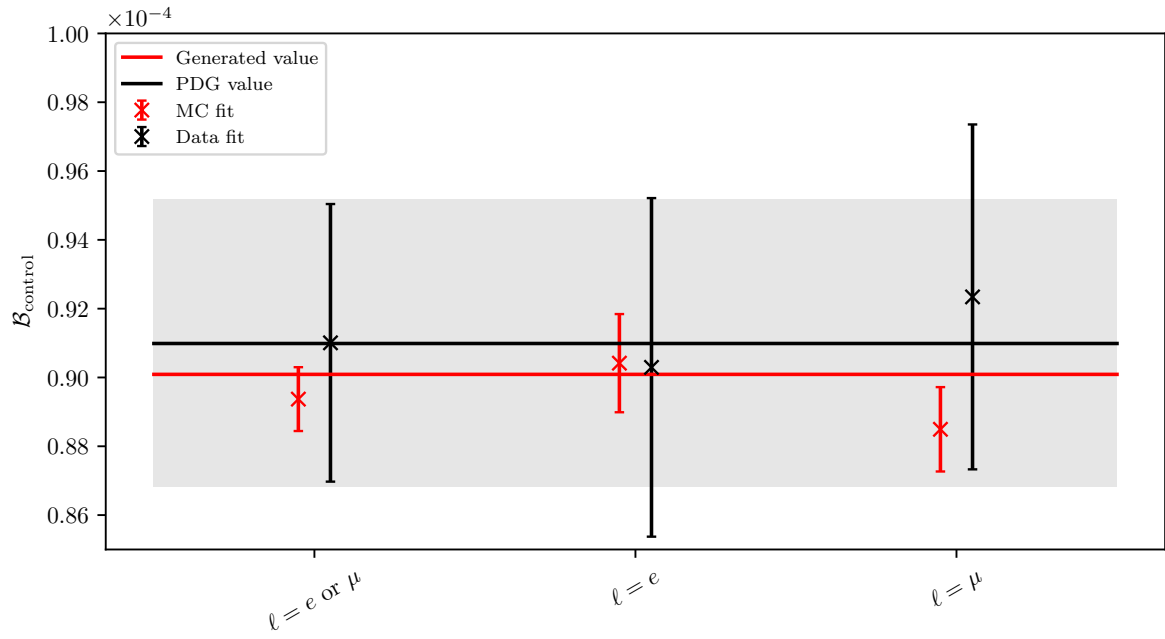
Figure 7.9: Various branching ratio determinations for the control decay of our analysis.

Chapter 8

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.

8.1 PID efficiency correction

The PID cut 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 factor and systematics tables for PID cut efficiencies of 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, 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 8.1, as well as the corresponding entry for all 3 particles. The entries are shown for both the signal and control region.

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

PID correction and systematics	Control region	Signal region
Same sign K (w.r.t the B)	1.005 ± 0.009	1.007 ± 0.010
Opposite sign K (w.r.t the B)	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
$KK\ell$	0.990 ± 0.019	0.990 ± 0.020

Table 8.1: PID correction factors and systematic error for various charged particles and their combinations.

8.2 Toy MC experiments

Toy MC experiments allow us to study the yields, errors and the pulls of the signal fit by generating our own pseudo datasets, according to the MC, and not producing it in the standard way, which consumes unaffordable amounts of CPU time. In order to test the fit behavior, a toy MC study was performed where all available MC was used for template creation.

8.2.1 Pseudo-experiments for expected signal yield

We constructed 3×10^3 pseudo datasets, where each dataset was generated with the expected amount of each template category, distributed according to the Poisson distribution. All fits were performed with the optimal initial uniform binning of 20×20 bins in ΔE and M_{BC}.

Figure 8.1 shows distributions of the fit yields, errors and the pull distribution of all pseudo fits. The fits seem to be under control since there is not any significant bias, and the pulls follow a normal distribution with a mean of $(1.1 \pm 1.8) \times 10^{-2}$ and standard deviation of (1.013 ± 0.013). The mean $(\bar{X})$ and the standard deviation (S) were calculated in the usual way, while the errors of these statistics were obtained by calculating the standard error of the mean $(\sigma_{\bar{X}})$ and standard deviation (σ_S), taken from [14], as

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{N}}, \quad \sigma_S = \frac{S}{\sqrt{2(N-1)}}, \tag{8.1}$$
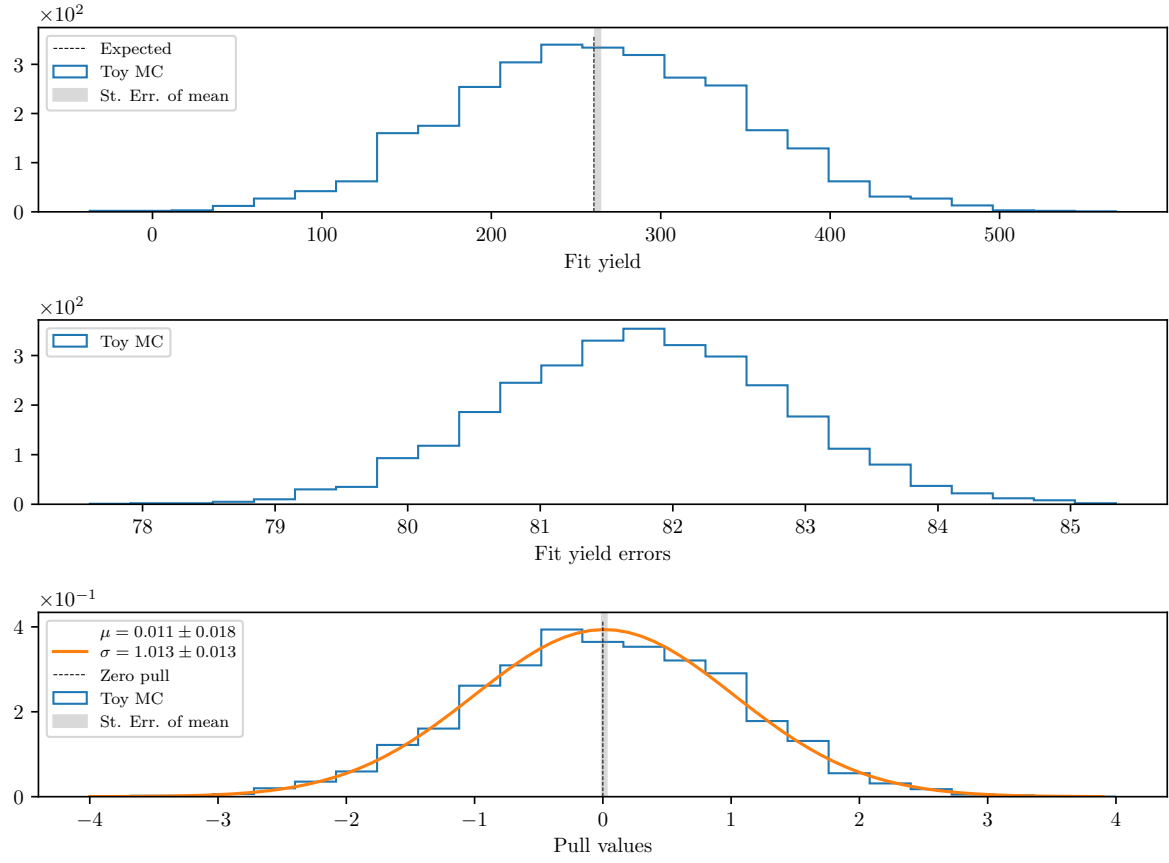
where N is the number of performed measurements.

Figure 8.1: Toy MC fits of pseudo data showing the fit yield (top), fit errors (center) and the pull distribution of the fits (bottom).

8.2.2 Toy MC linearity test

Linearity test is used for determining sensitivity of the fit to the amount of signal in the fitted sample. Since this is a 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 with various amounts. 50 steps from [0.1, 10] in the logarithmic scale are taken for fractions of signal amount and for each fraction we generate 500 pseudo datasets according to Poisson statistics.

Figure 8.2 shows the mean fit yield and expected yield difference, mean pull and the mean significance at each signal fraction value. The expected MC result is shown at fraction value 1. The plots show no significant bias with with respect to the signal fraction, while the pulls seems to be described by the normal distributions throughout the fraction range. At expected value we are at about 3.2σ significance, as already pointed out at the beginning of this section.

8.3 Fit Bias

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 estimated the fit bias as a function of the signal yield in the form of a linear approximation, as already shown in Figure X. The bias function is approximated as

$$bias function, \tag{8.2}$$

which leads to a systematic uncertainty of about X at the value of expected signal yield.

8.4 Fit Template Smearing and Offset

8.5 Effects of a Finite MC sample

The shape of signal and backgrounds templates in our analysis is fixed and only their normalization is considered in the fit. Due to the finite size of the MC sample, the template shape introduces an additional source of uncertainty, as it may differ if produced in a separate, equal-sized MC sample. Since the bins in these 2D histogram templates
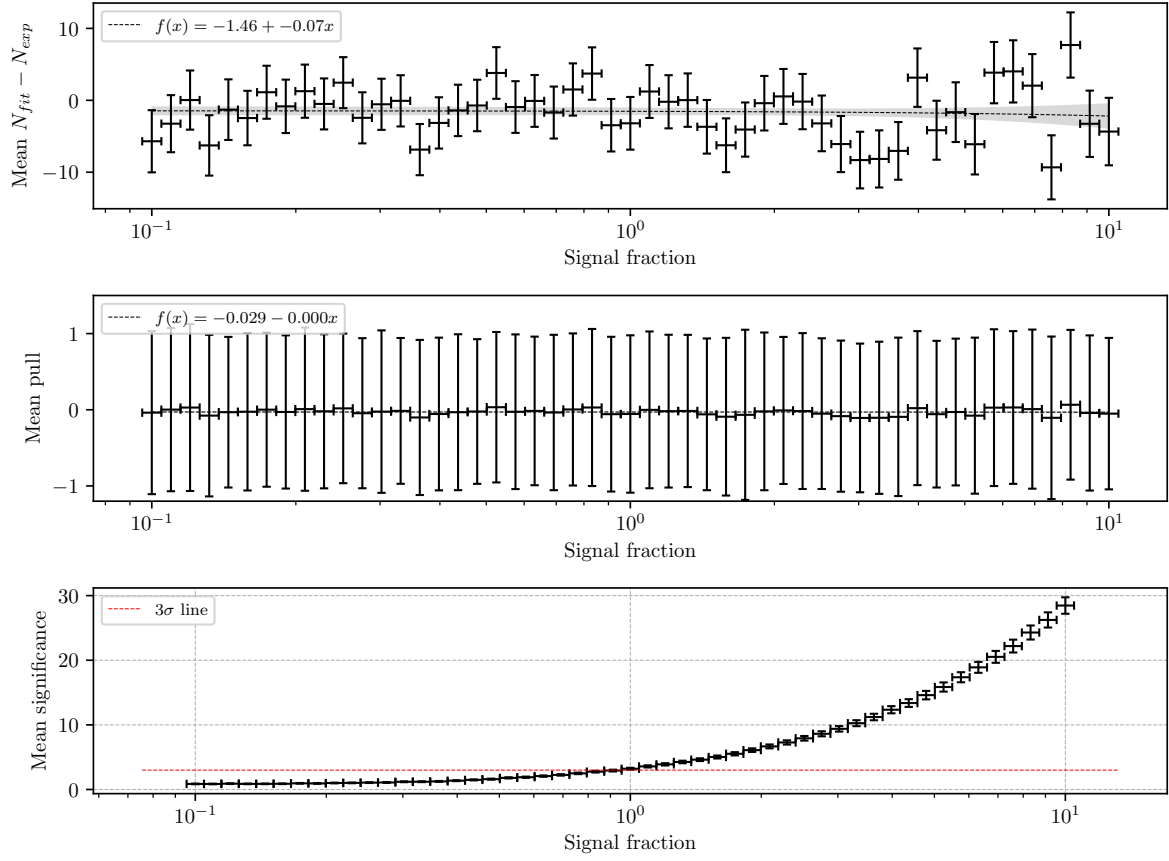
Figure 8.2: Mean fit yield and expected yield difference (top), mean pull (center) and mean significance (bottom) as a function of signal fraction.

are statistically independent, we can take each bin and vary the bin content according to the Poisson distribution. This procedure is repeated for X times and the width of the fit yield distribution, shown in Figure X, is taken as the uncertainty estimate.

8.6 MVA Selection Efficiencies

8.7 Model Uncertainty Effects

The used signal decay model in the generation step was ISGW2 [X], which is known not to be the most precise model for such decays (which ones X)? Due to this uncertainty, our analysis has been made as model independent as possible via means of not using variables, which might have a model dependent nature, in any kind of selection steps or MVA trainings. Such variables are i.e. squared momentum transfer to the lepton pair (q^2), 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 uncertainty on our final result, we prepare two additional signal MC samples, produced with an extreme scenario decay model. In the first scenario we generate the signal MC sample with a generic phase-space decay mode PHSP [X], which results in continuum-like distributions of q^2 and m_{KK}. In the other scenario only resonant-like contributions in m_{KK} are used. These two scenarios act as extreme cases of decay model choice and present a reasonable measure of the model uncertainty. Figure X shows the generated m_{KK} distribution of the three mentioned decay models and distributions of ΔE and M_{BC} after the final selection. The different signal templates are used and the range of the corresponding fit yields is taken as the uncertainty range.

PLOTS AND MORE TEXT

8.8 Summary of Systematics

Bibliography

[1] N. Cabibbo. "Unitary symmetry and leptonic decays". In: *Physical Review L* 10.12 (1963), p. 531.

[2] M. Kobayashi and T. Maskawa. "CP-violation in the renormalizable theory of weak interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657.

[3] Y. Amhis et al. "Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016". In: *Eur. Phys. J.* C77.12 (2017), p. 895. DOI: `10.1140/epjc/s10052-017-5058-4`. arXiv: `1612.07233 [hep-ex]`.

[4] P. Gambino et al. "Inclusive semileptonic B decays and the determination of —V(ub)—". In: *JHEP* 10 (2007), p. 058. DOI: `10.1088/1126-6708/2007/10/058`. arXiv: `0707.2493 [hep-ph]`.

[5] T. Sjöstrand, S. Mrenna, and P. Skands. "PYTHIA 6.4 physics and manual". In: *Journal of High Energy Physics* 2006.05 (2006), p. 026.

[6] T. 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`.

[7] D. Asner et al. "Search for exclusive charmless hadronic B decays". In: *Physical Review D* 53.3 (1996), p. 1039.

[8] A. Bevan et al. "The physics of the B factories". In: *The European Physical Journal C* 74.11 (2014), p. 3026.

[9] P. del Amo Sanchez et al. "Study of $B \to \pi l \nu$ and $B \to \rho l \nu$ decays and determination of $|V_{ub}|$". In: *Phys. Rev. D* 83 (3 2011), p. 032007. DOI: `10.1103/PhysRevD.83.032007`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.83.032007`.

[10] J. Stevens and M. Williams. "uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers". In: *Journal of Instrumentation* 8.12 (2013), P12013.

[11] W. Verkerke and D. Kirkby. "The RooFit toolkit for data modeling". In: *Statistical Problems in Particle Physics, Astrophysics and Cosmology*. World Scientific, 2006, pp. 186–189.

[12] F. James. "MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1". In: (1994).

[13]   M. Paterno. *Calculating efficiencies and their uncertainties.* Tech. rep. Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), 2004.

[14]   S. Ahn and J. A. Fessler. "Standard errors of mean, variance, and standard deviation estimators". In: *EECS Department, The University of Michigan* (2003), pp. 1–2.

A: MVA control plots

ROE clean-up π^0 training

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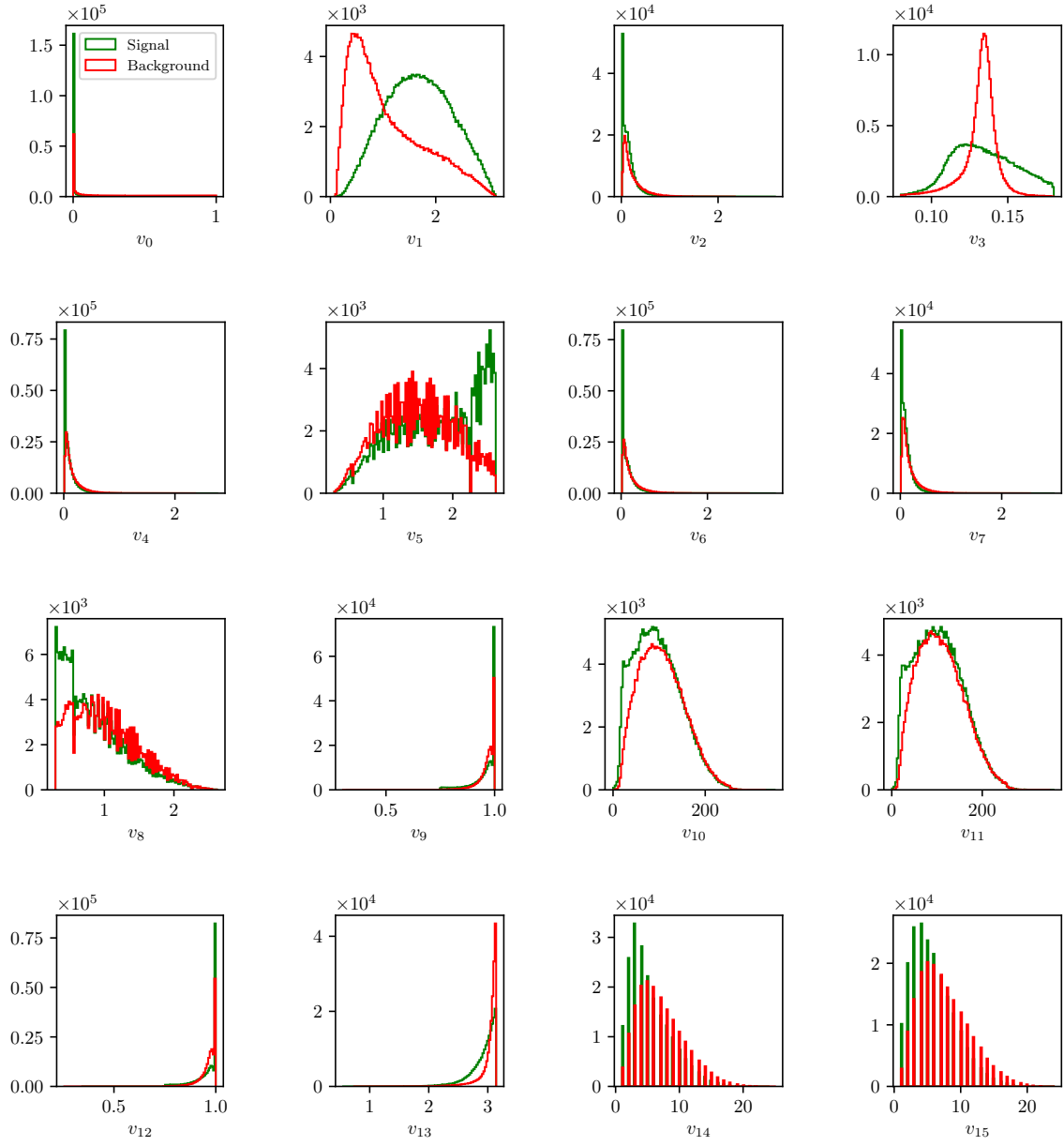
Variable distributions

Figure 8.3: Feature distributions for MVA training of π^0 candidates in the scope of ROE clean-up.

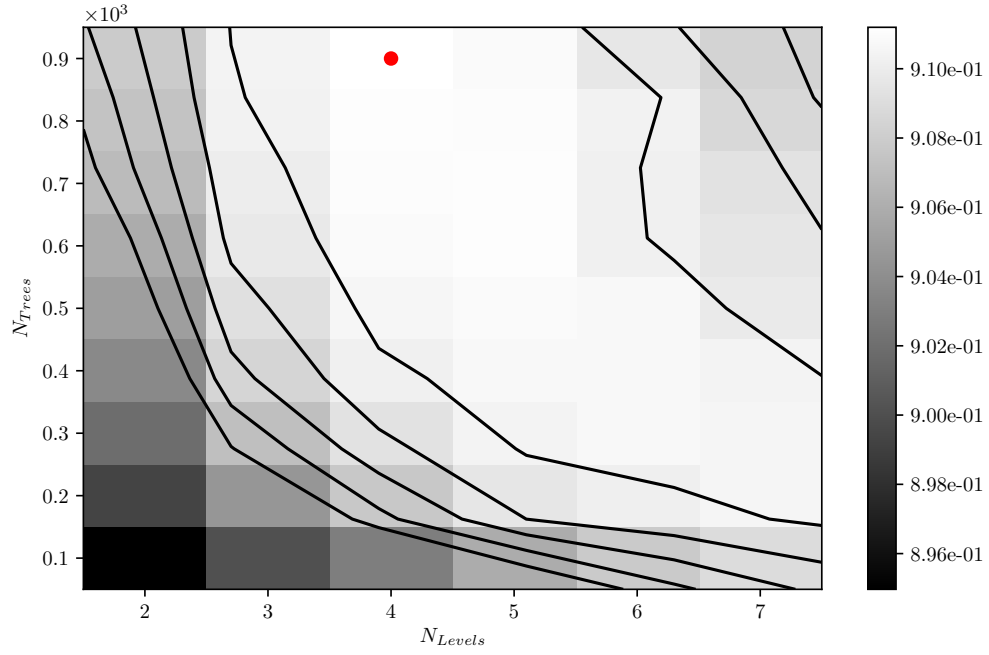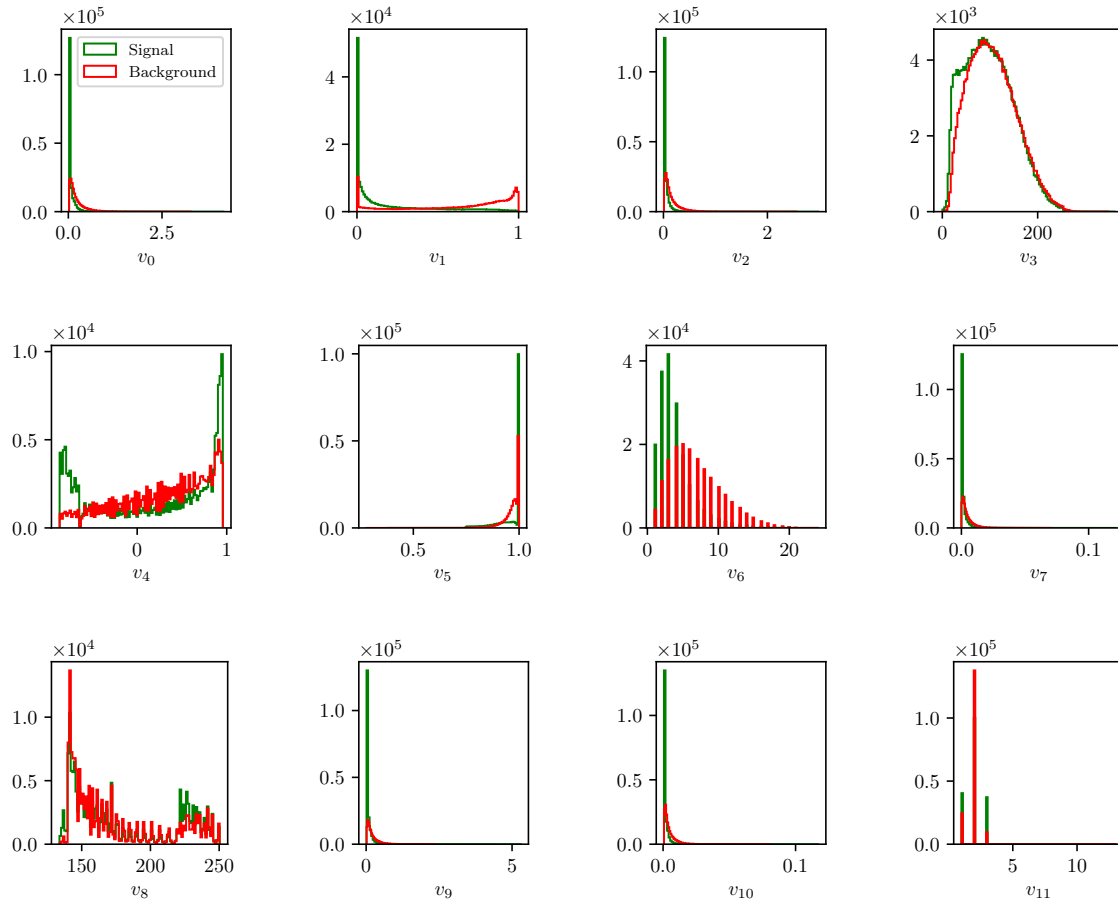Figure 8.3: Feature distributions for MVA training of π^0 candidates in the scope of ROE clean-up.
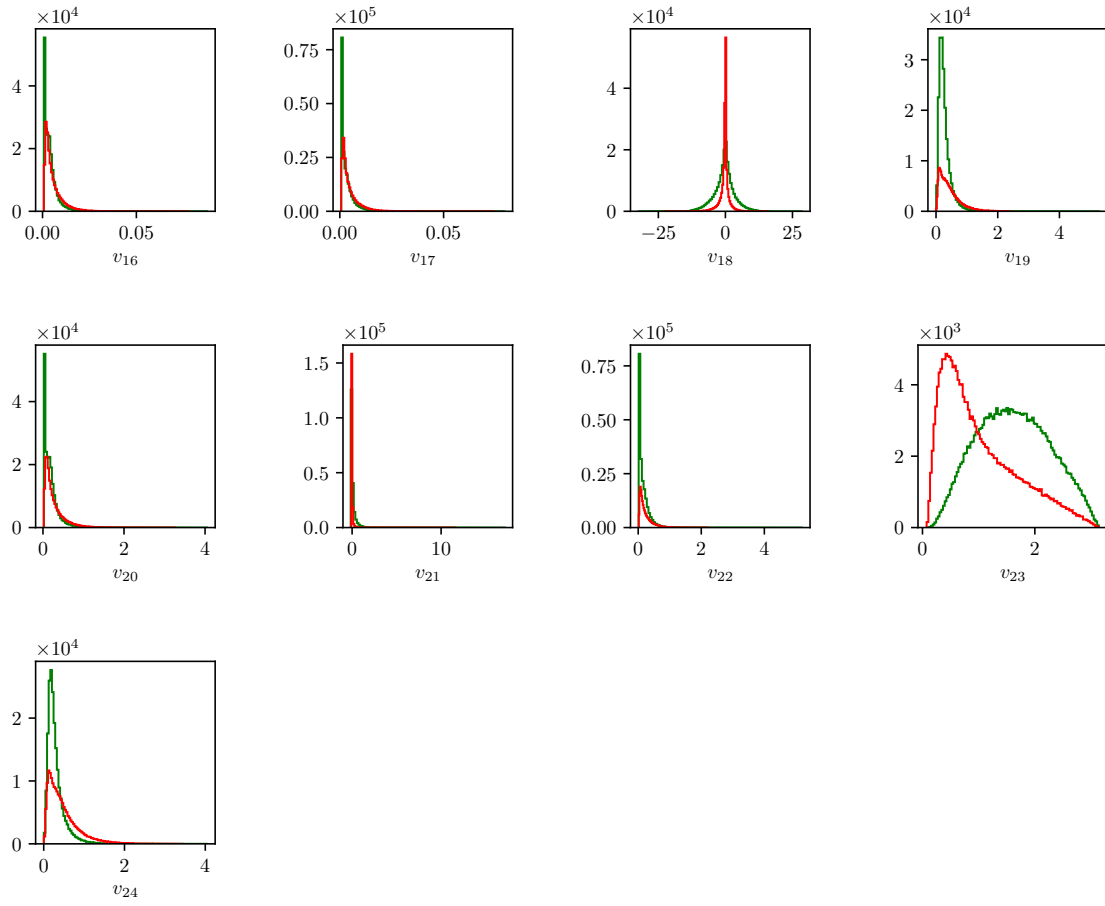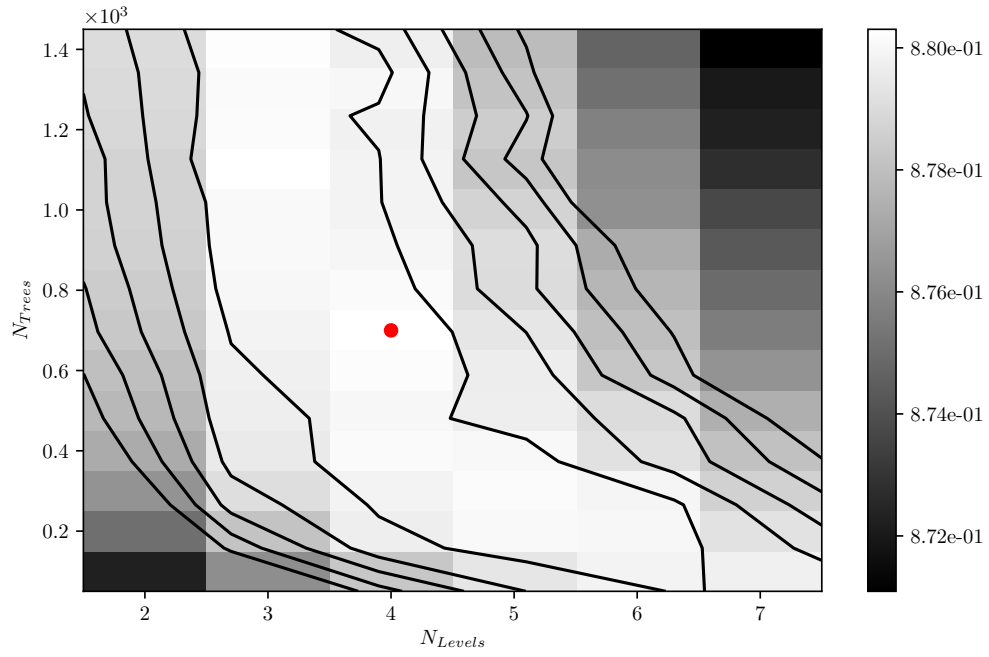
Hyper-parameter optimization

Figure 8.4: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of π^0 candidates in the scope of the ROE clean-up.
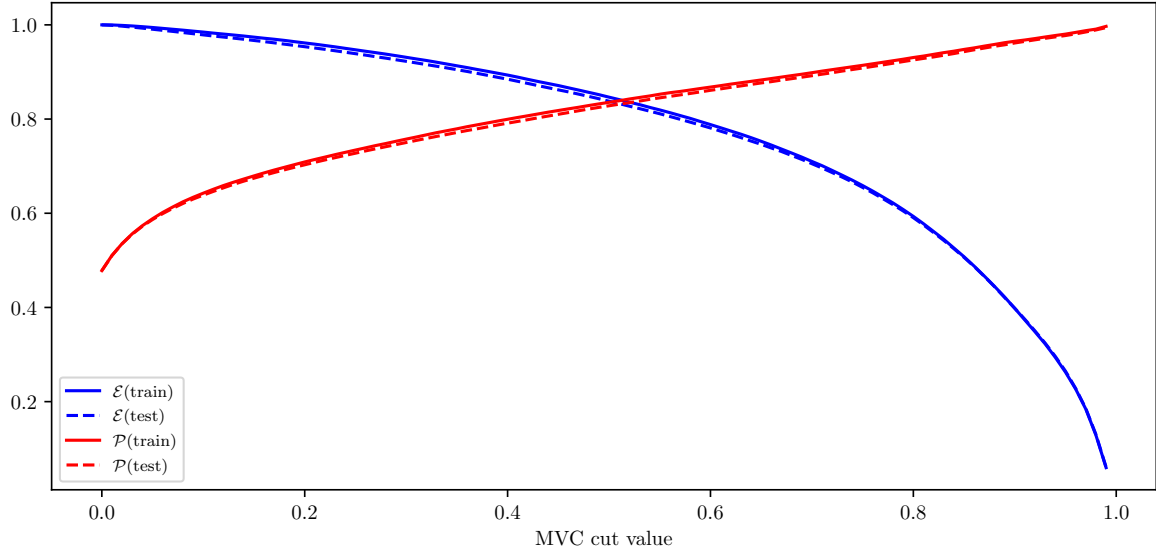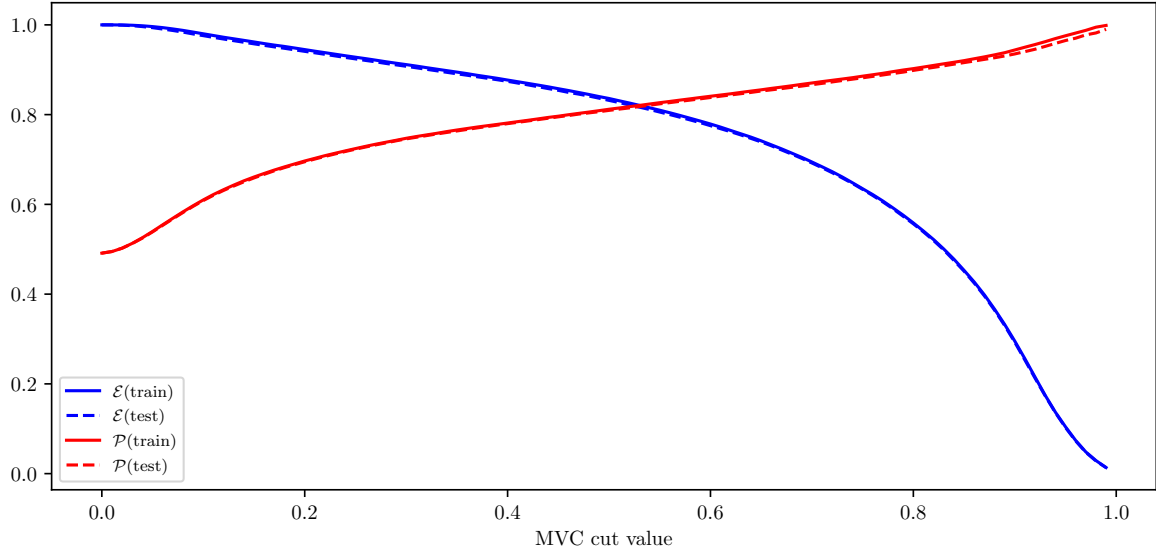
Figure 8.5: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the MVA classifier output for π^0 candidates training on the train (solid) and test (dashed) samples.
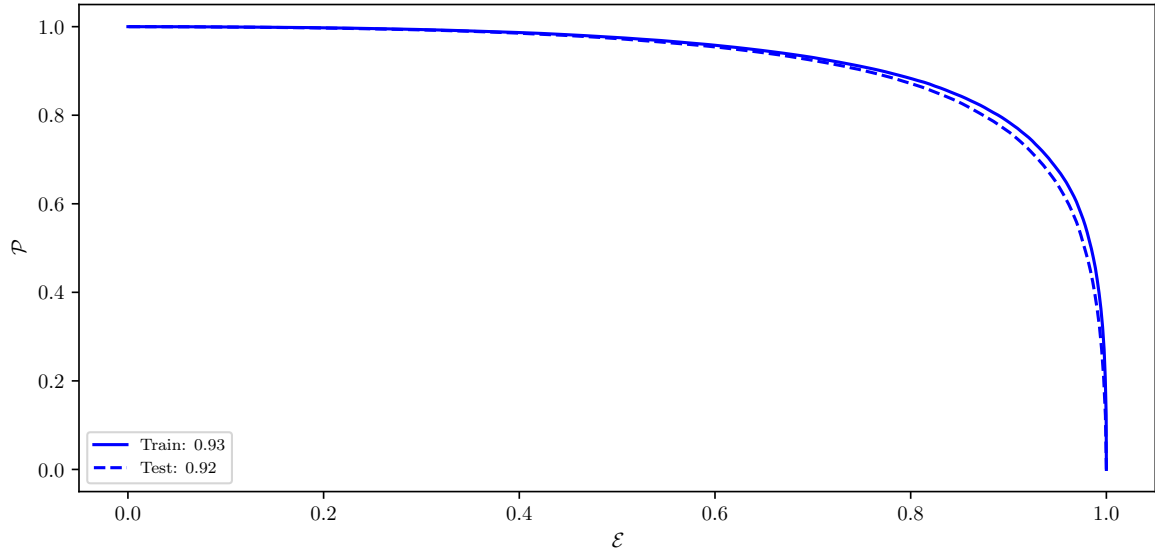
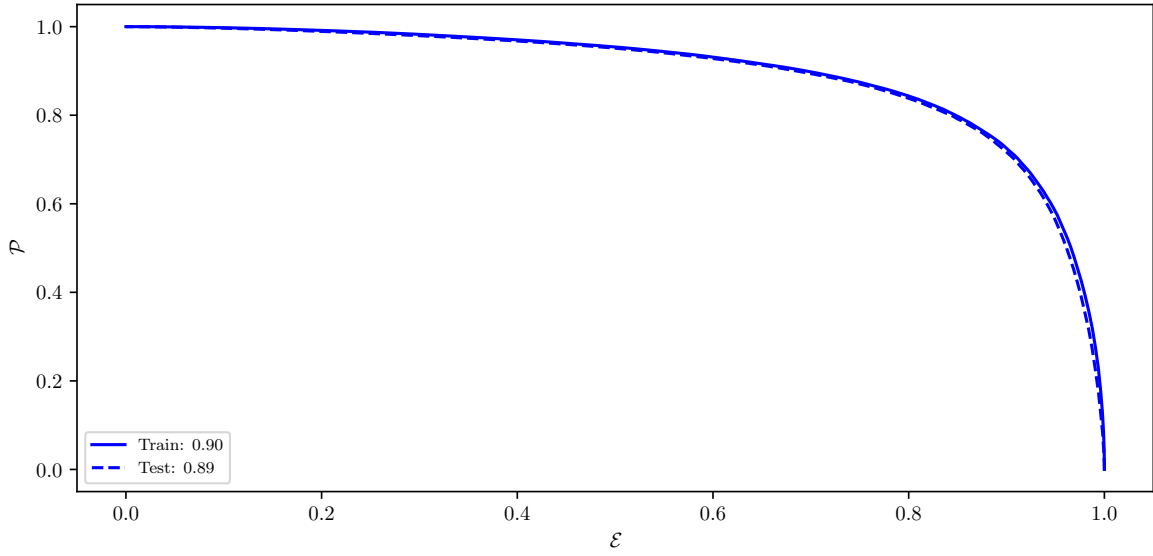Figure 8.6: ROC curves of the MVA classifier output for π^0 candidates training on the train (solid) and test (dashed) samples.

ROE clean-up γ training

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

Variable distributions

Figure 8.7: Feature distributions for MVA training of γ candidates in the scope of ROE clean-up.

Hyper-parameter optimization

Figure 8.8: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of γ candidates in the scope of the ROE clean-up.

Figure 8.9: 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 8.10: ROC curves of the MVA classifier output for γ candidates training on the train (solid) and test (dashed) samples.

ROE clean-up duplicate pair training

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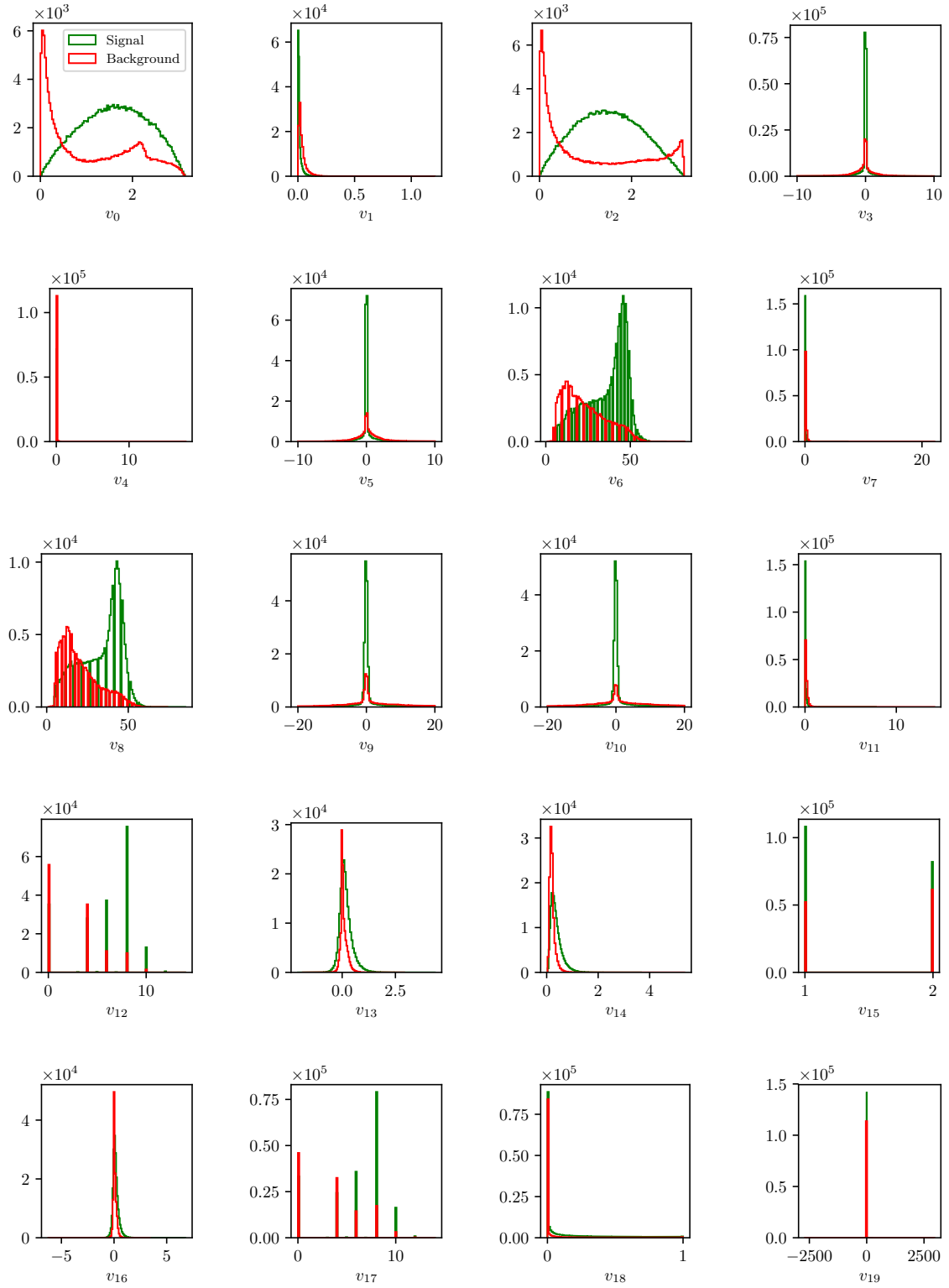
Variable distributions

Figure 8.11: Feature distributions for MVA training of duplicate track pair candidates in the scope of ROE clean-up.

119
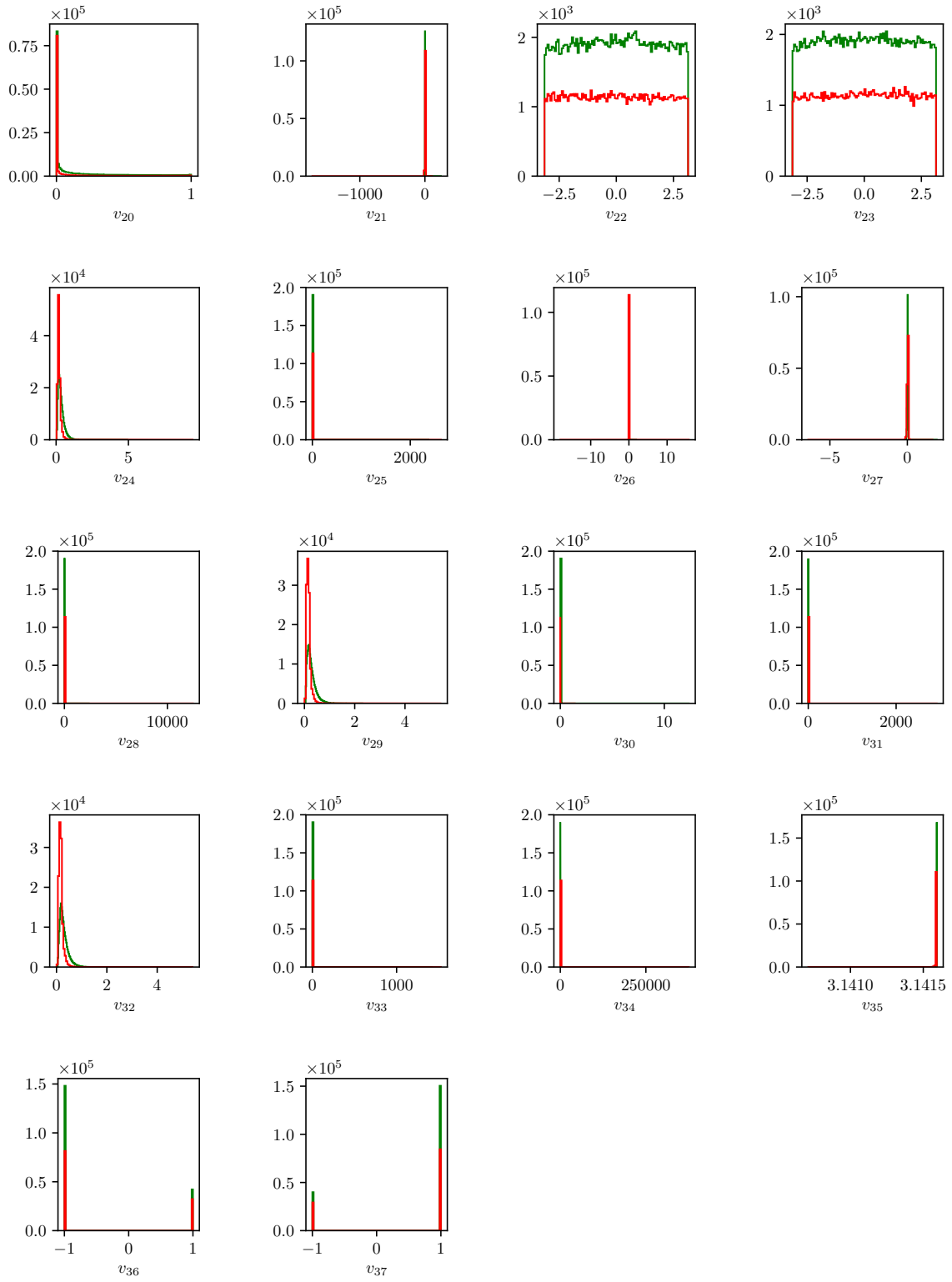
Figure 8.11: Feature distributions for MVA training of duplicate track pair candidates in the scope of ROE clean-up.
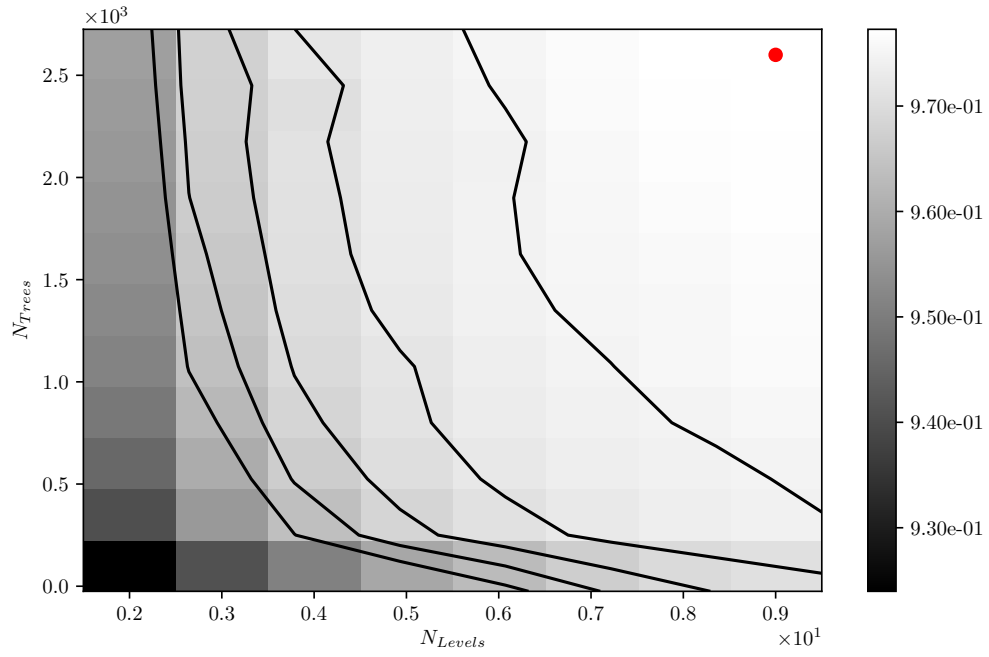
Figure 8.12: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of duplicate track pair candidates in the scope of the ROE clean-up.
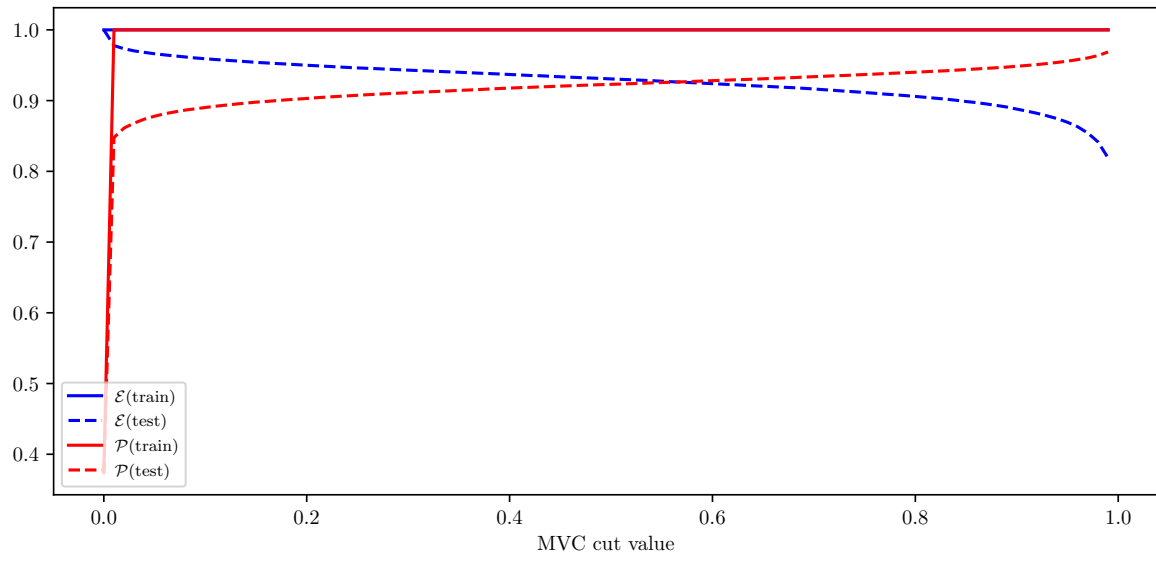
Figure 8.13: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the MVA classifier output for duplicate track pair candidates training on the train (solid) and test (dashed) samples.
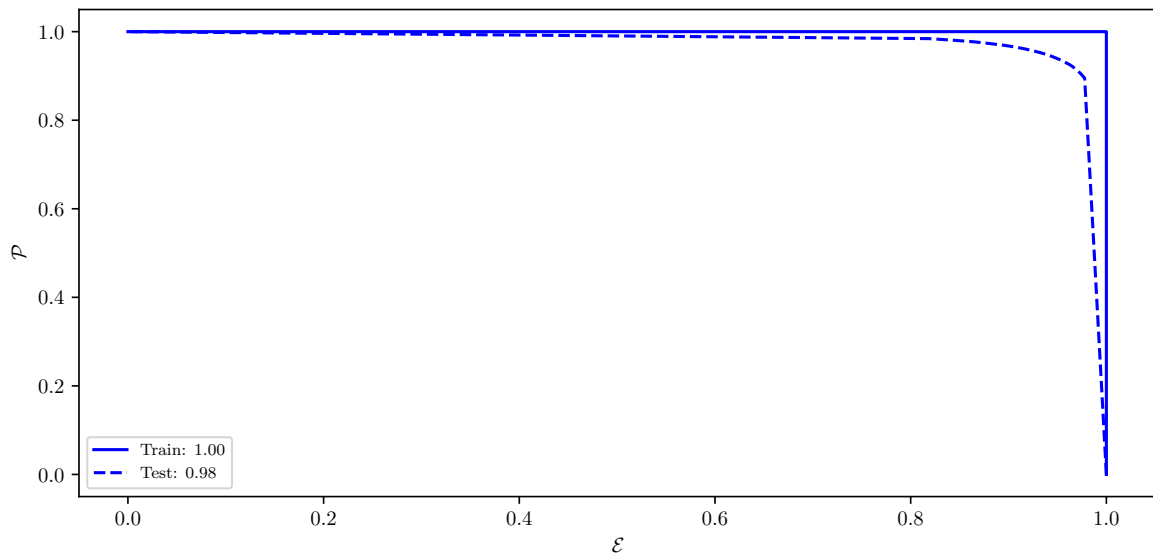
Figure 8.14: ROC curves of the MVA classifier output for duplicate track pair candidates training on the train (solid) and test (dashed) samples.

ROE clean-up duplicate track training

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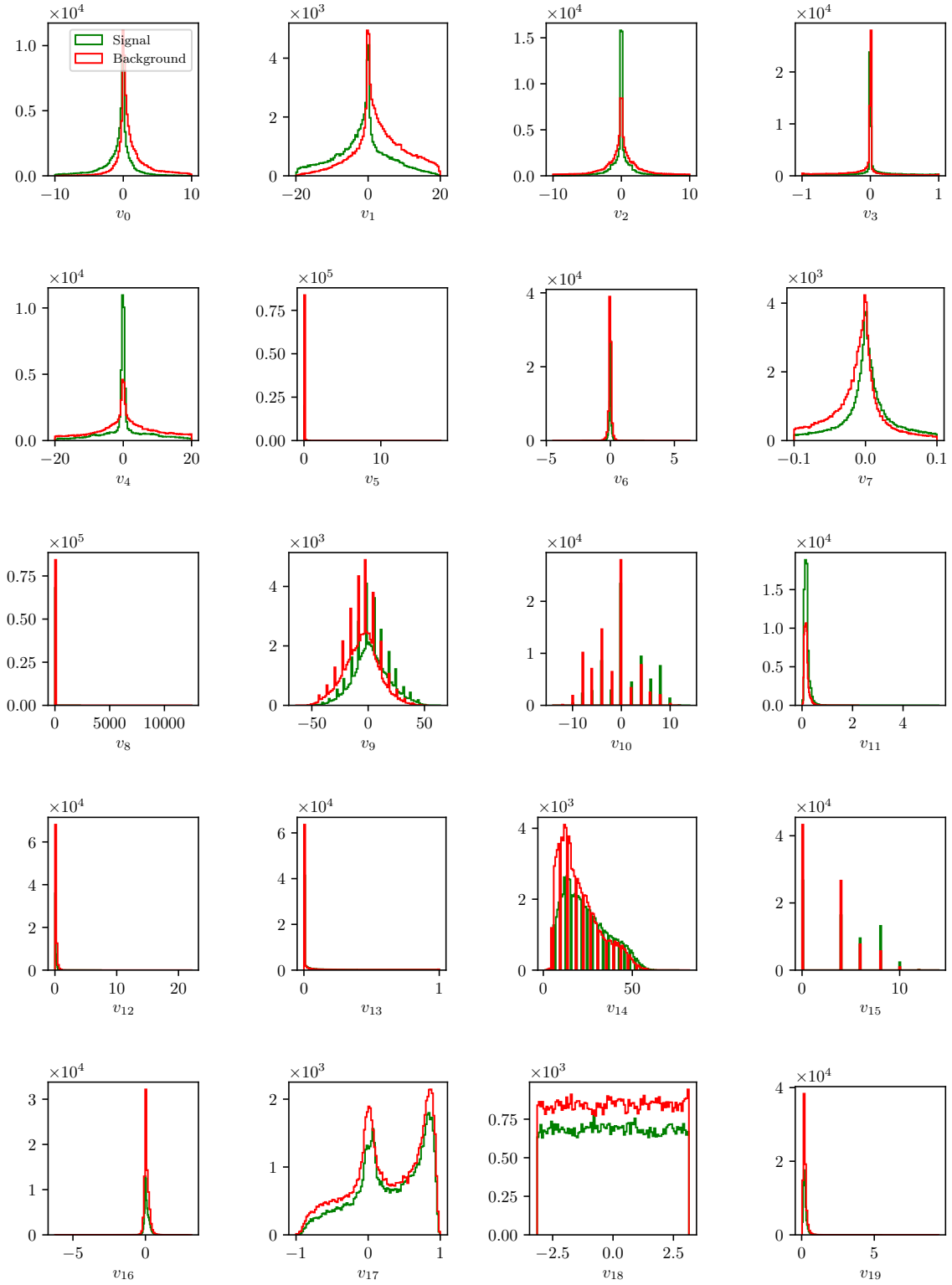
Variable distributions

Figure 8.15: Feature distributions for MVA training of duplicate track candidates in the scope of ROE clean-up.
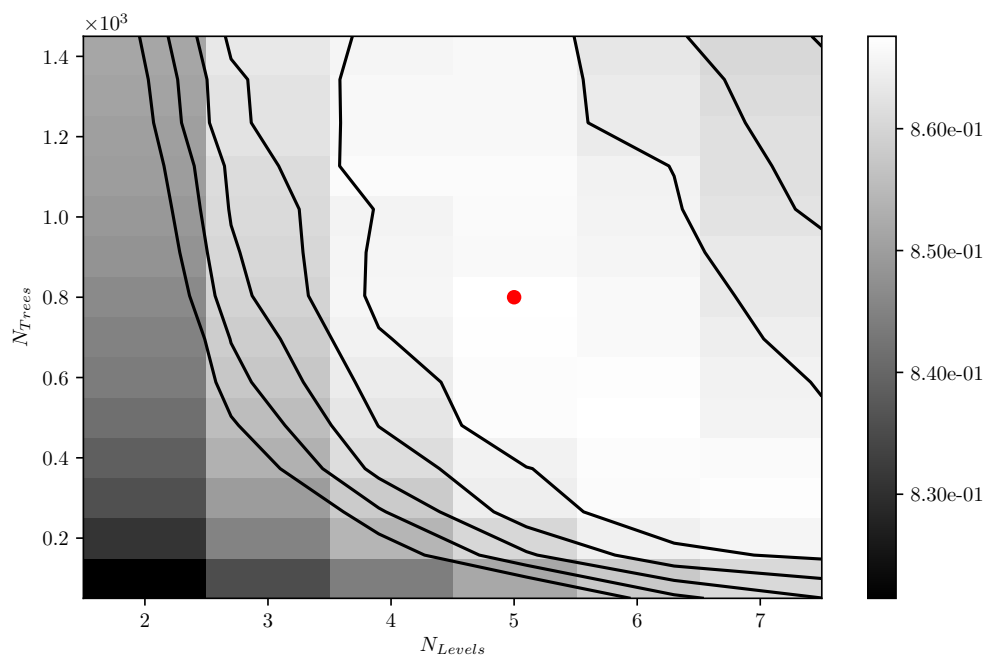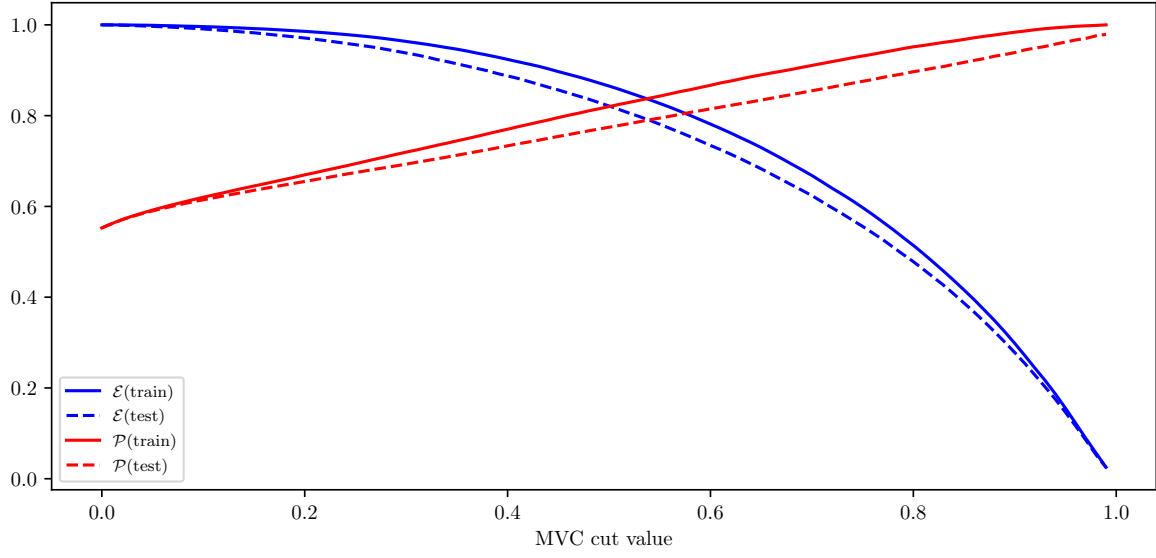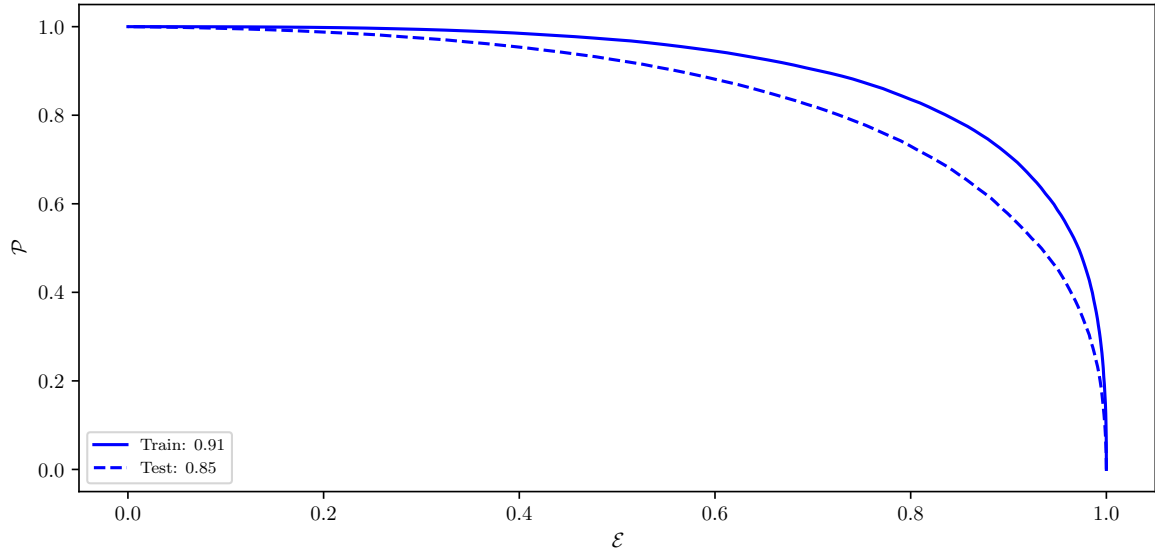
Hyper-parameter optimization

Figure 8.16: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of duplicate track candidates in the scope of the ROE clean-up.

Figure 8.17: 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 8.18: ROC curves of the MVA classifier output for duplicate track candidates training on the train (solid) and test (dashed) samples.

126

$q\bar{q}$ suppression training

Variable importance

	Name	Alias	Importance
0	B_CosTBTO	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 8.6: Variable names, aliases and importance in the scope of $q\bar{q}$ suppression MVA training.
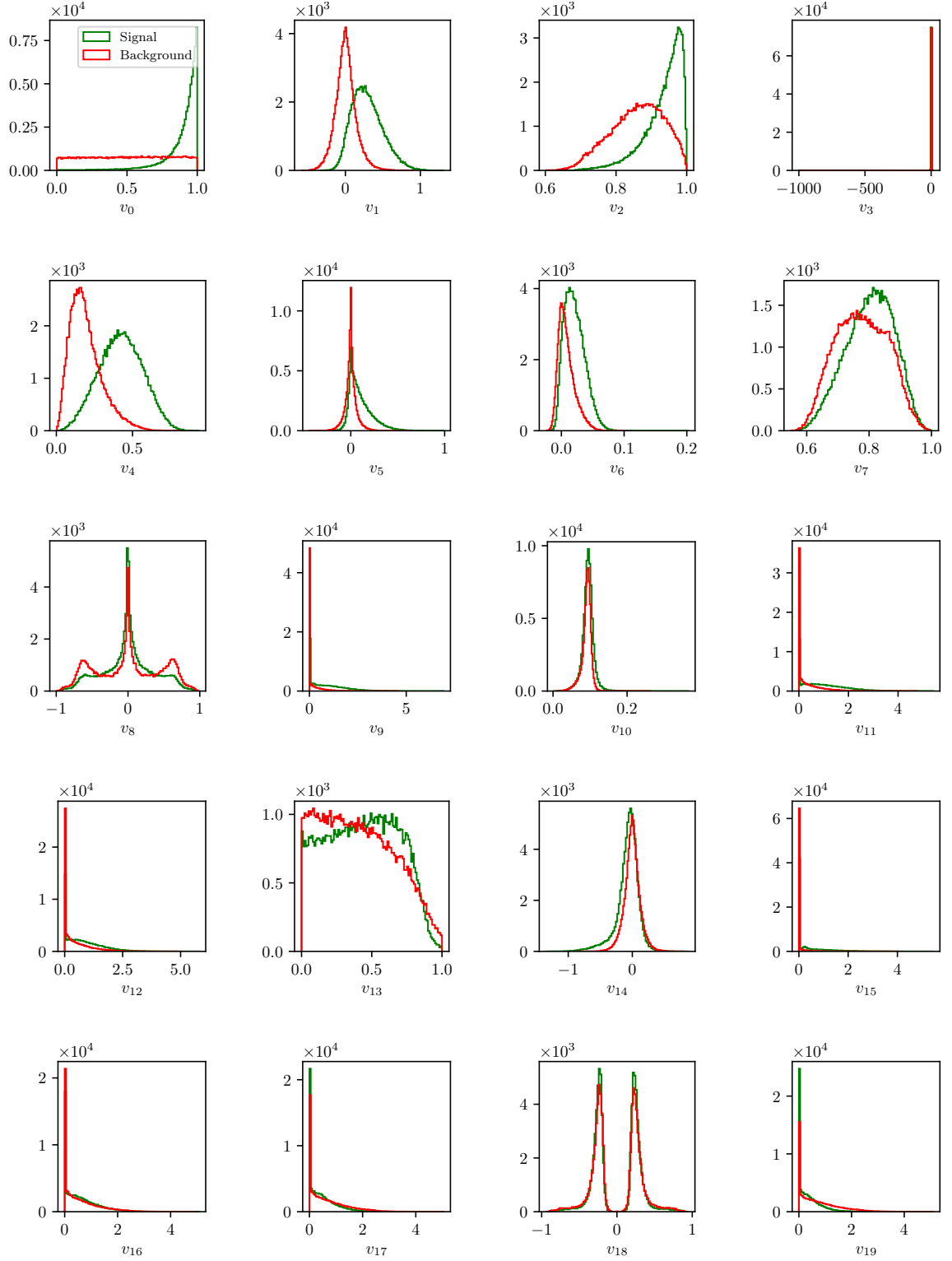
Variable distributions

Figure 8.19: Feature distributions for MVA training of $q\bar{q}$ background suppression.
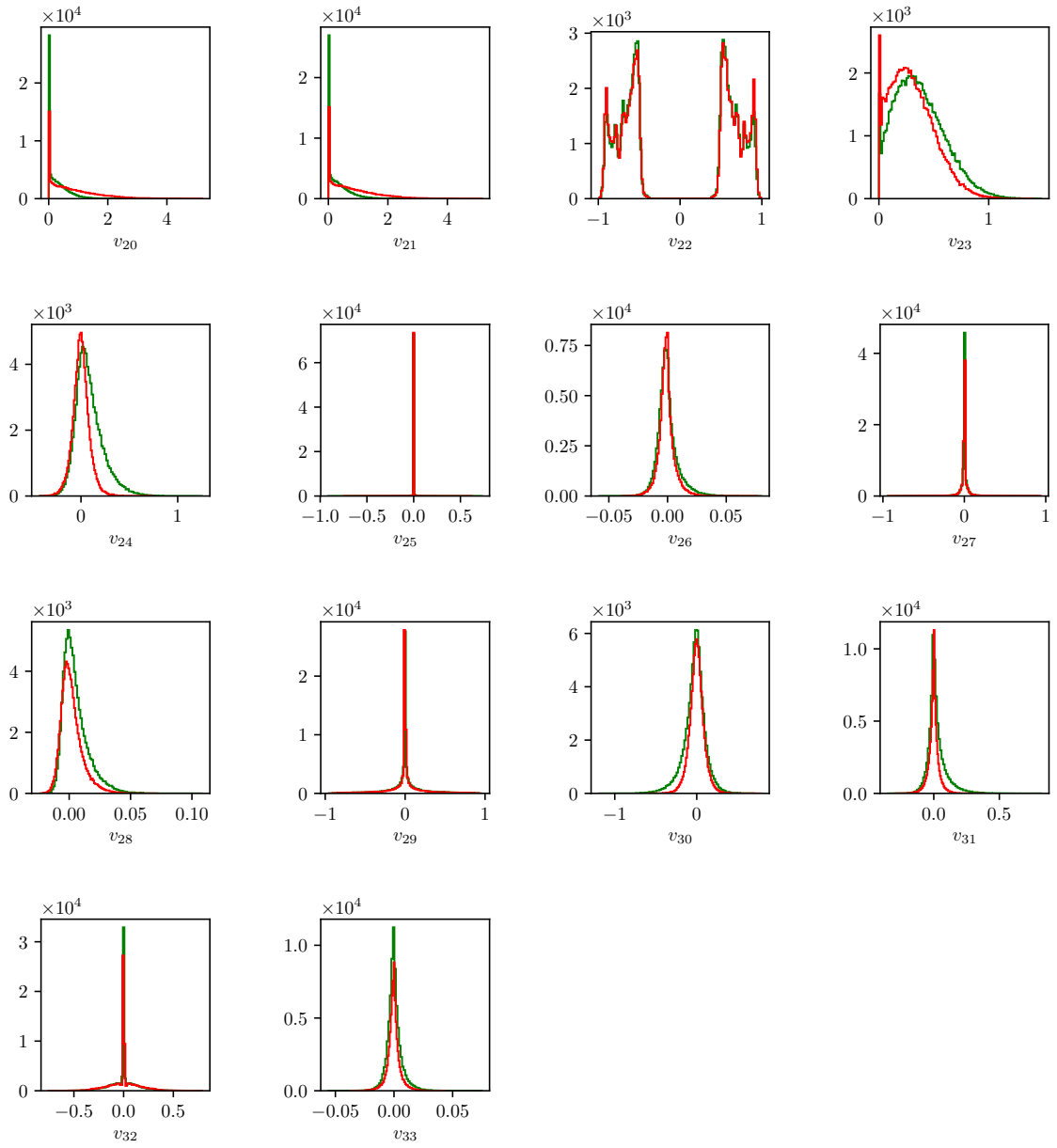
Figure 8.19: Feature distributions for MVA training of $q\bar{q}$ background suppression.
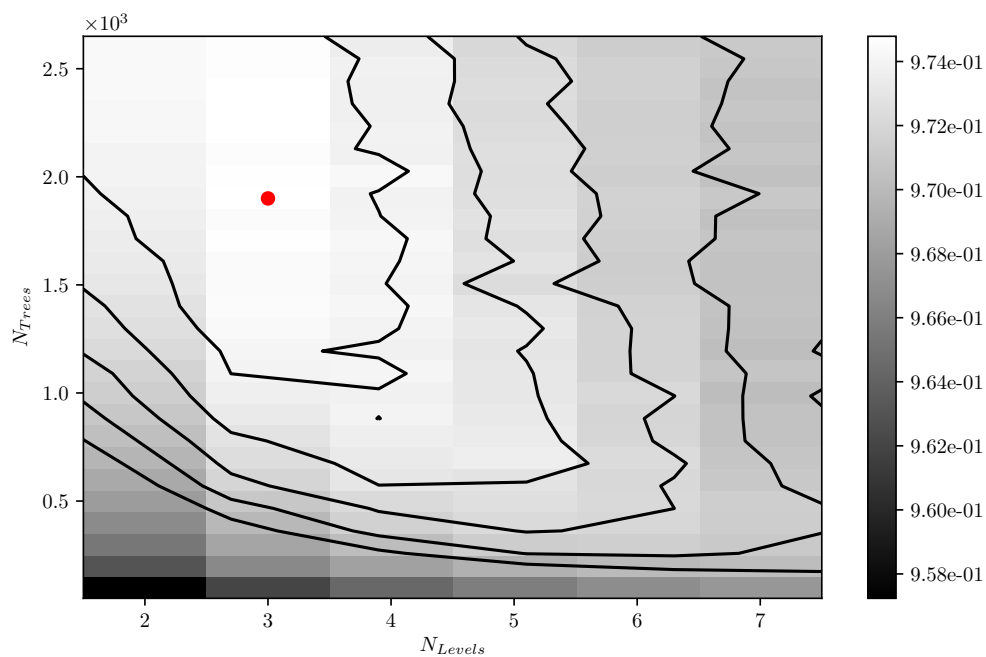
Hyper-parameter optimization

Figure 8.20: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of $q\bar{q}$ background suppression.
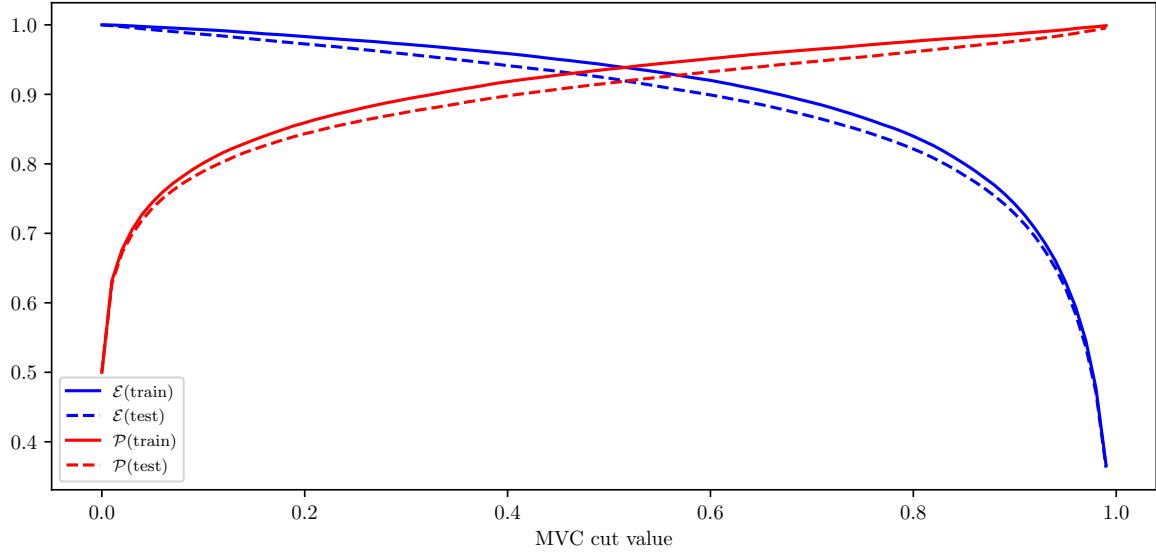
Figure 8.21: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the MVA classifier output for $q\bar{q}$ background suppression training on the train (solid) and test (dashed) samples.
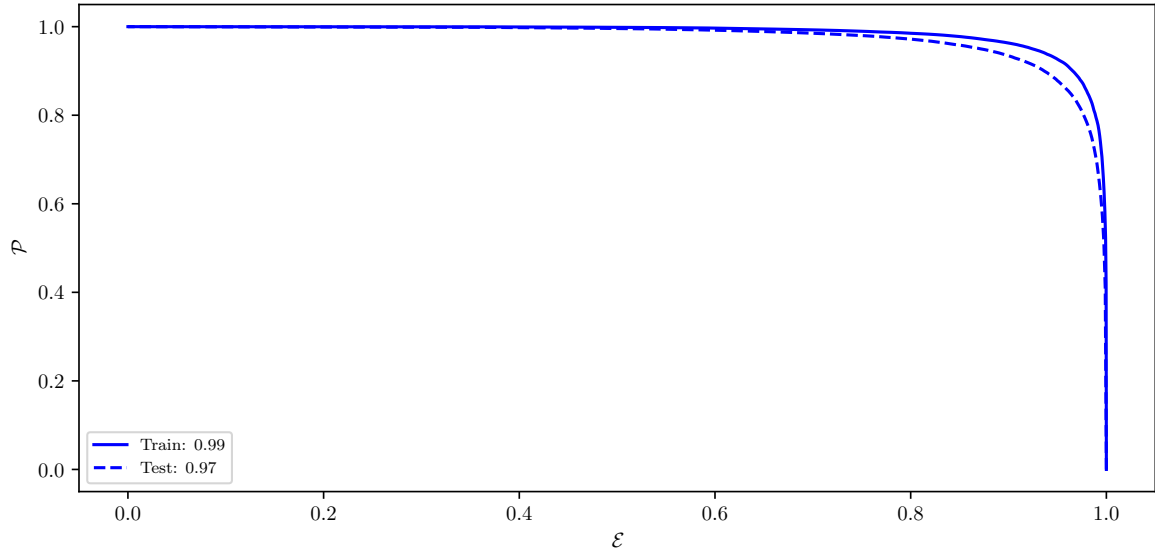

Figure 8.22: ROC curves of the MVA classifier output for $q\bar{q}$ background suppression training on the train (solid) and test (dashed) samples.

Standard $B\bar{B}$ suppression training

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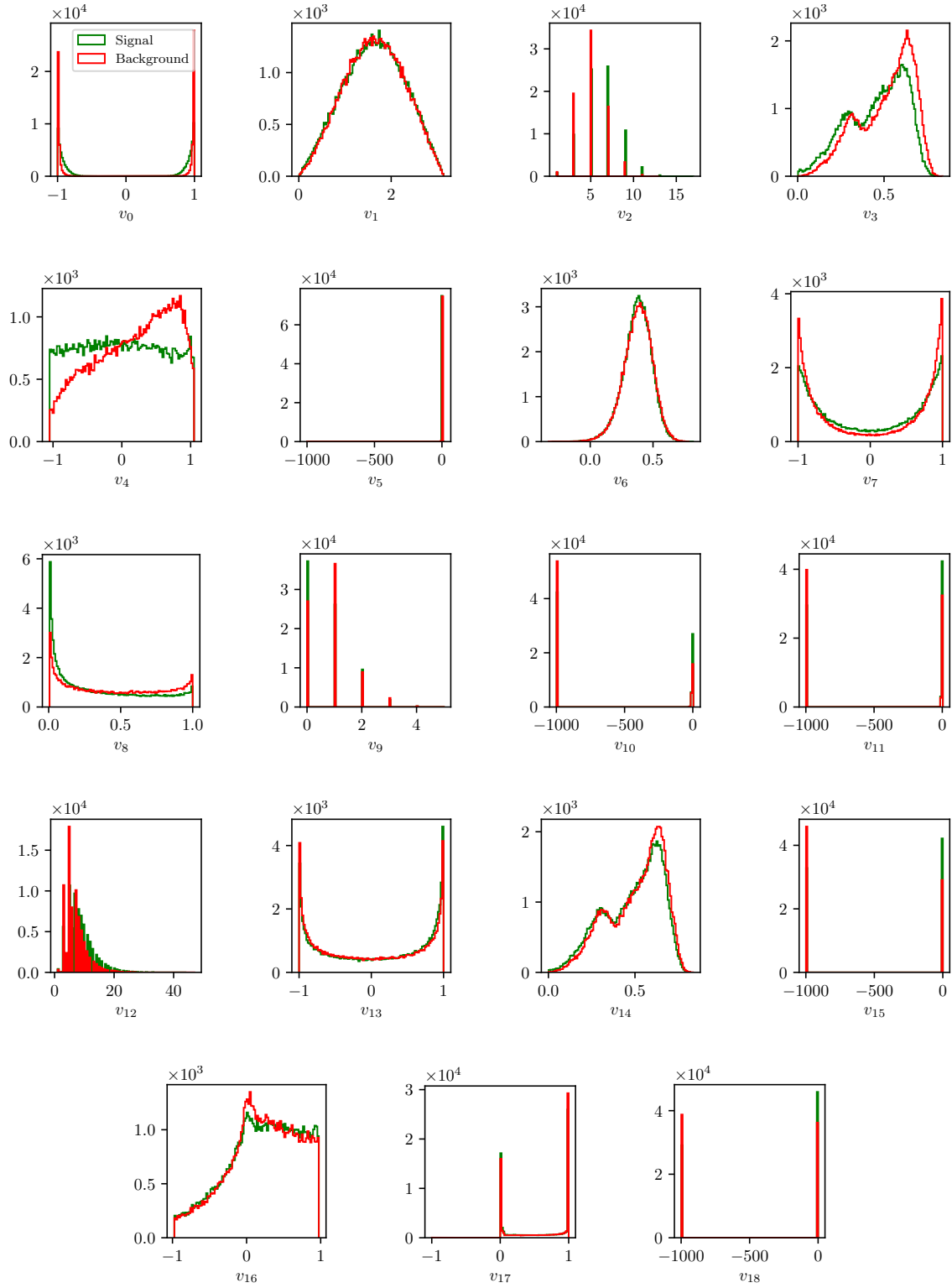
Variable distributions

Figure 8.23: Feature distributions for MVA training of $B\bar{B}$ background suppression.
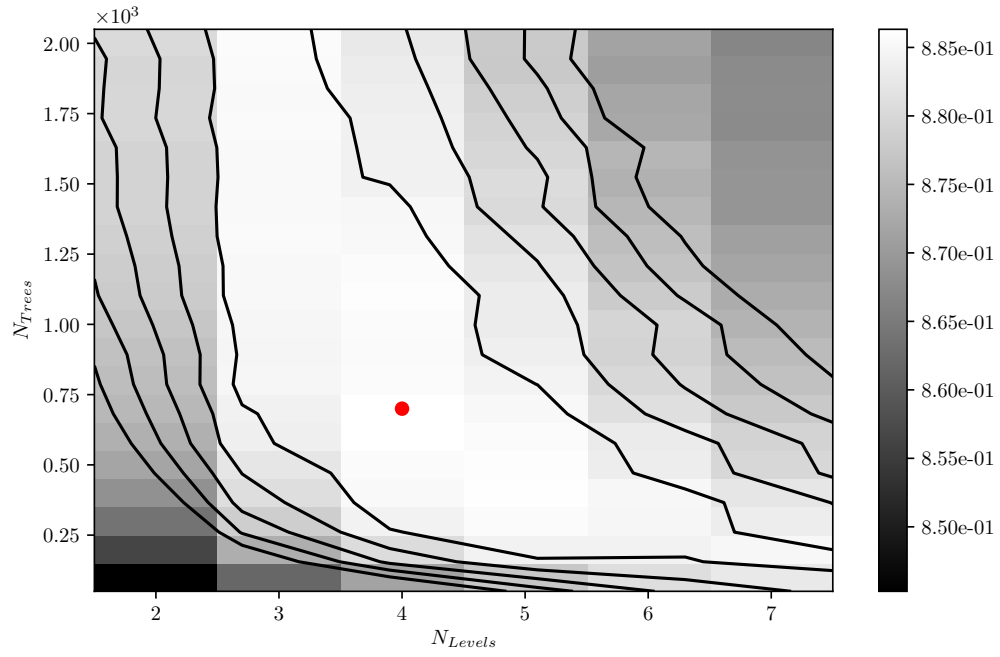
Hyper-parameter optimization

Figure 8.24: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of $B\bar{B}$ background suppression.
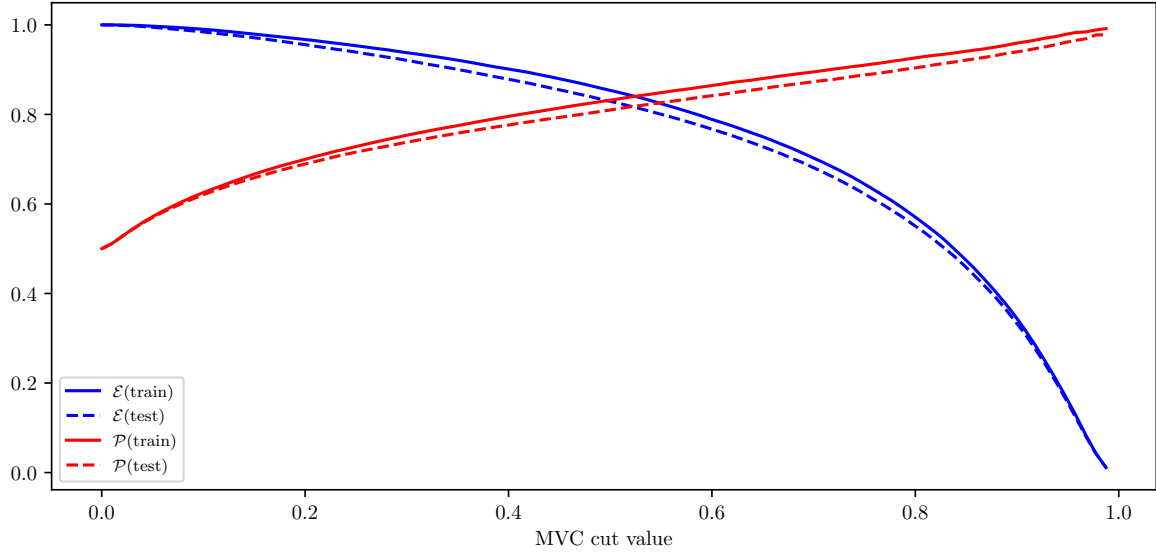
Figure 8.25: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.
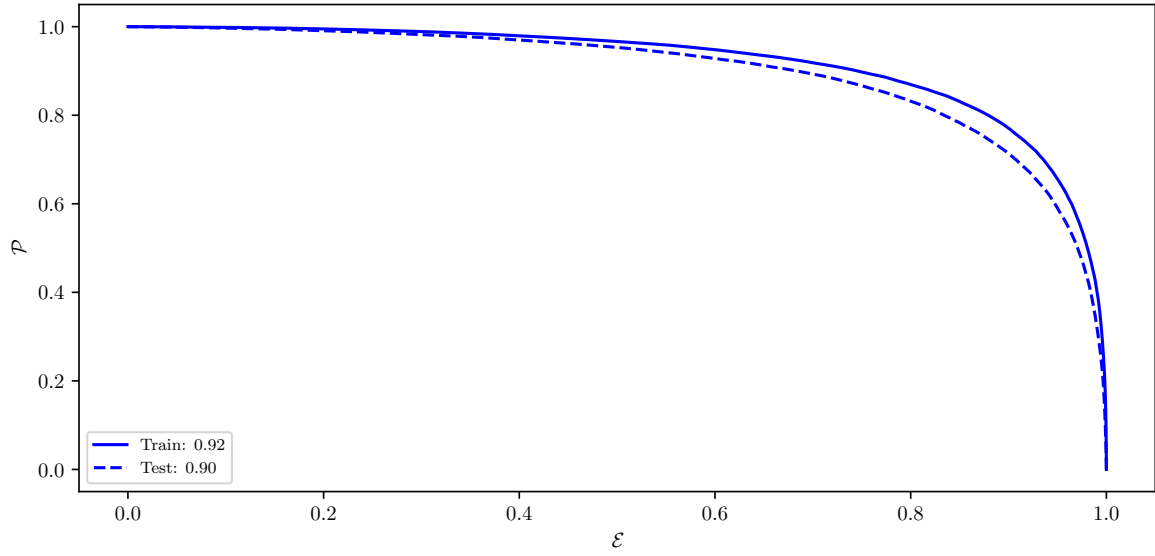

Figure 8.26: ROC curves of the MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.

136

Uniformity boosted $B\bar{B}$ suppression training

Hyper-parameter optimization

Hyper-parameters were not optimized due to the large CPU time consumption of the algorithm. The following set up of the hyper-parameters was chosen
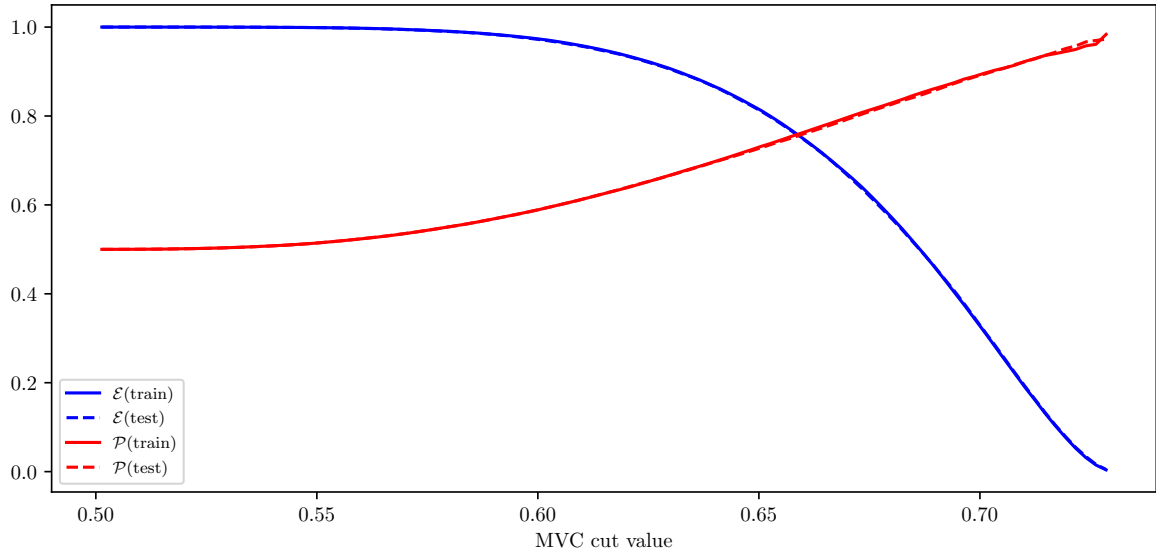
- `nTrees`: 300

- `nLevels`: 4

Results

Figure 8.27: 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.
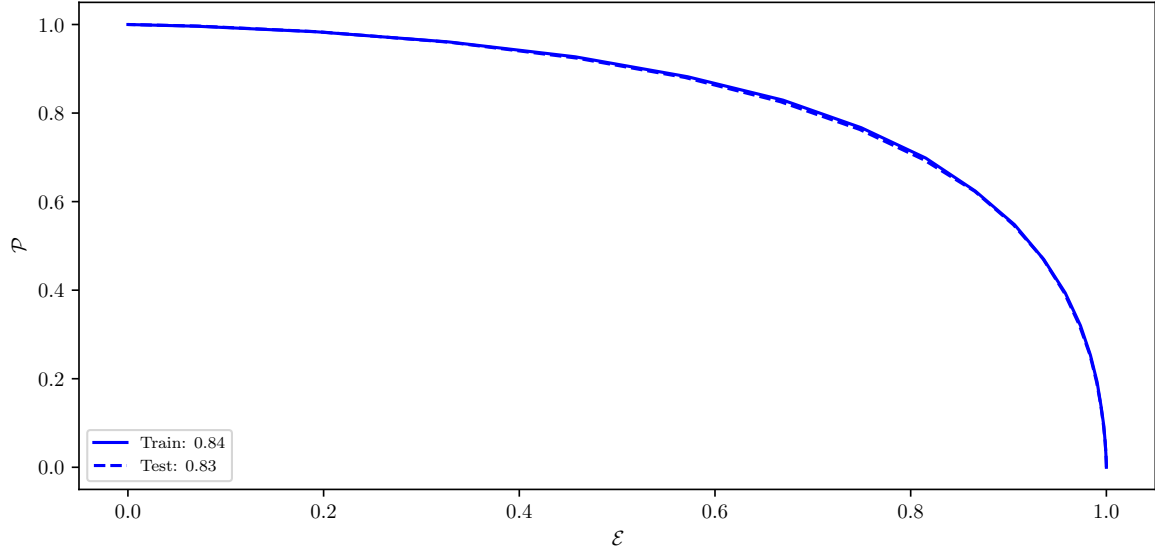
Figure 8.28: ROC curves of the uniformity boosted MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.