Exclusive untagged analysis of the semi-leptonic decay $B \rightarrow \rho\ \ell\ \nu$ in preparation for extracting $|V_{ub}|$ using early Belle II data

Lex Marinus Greeven

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

August 2019

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, ...
Date Signature

1. Gutachter: Prof. Dr. Jochen Dingfelder
2. Gutachter: Priv. Doz. Dr. Philip Bechtle

Acknowledgements

First of all, I want to thank Jochen Dingfelder for giving me the opportunity to write my master thesis within the Belle II analysis group. It has been a wonderful experience, made complete with the trip to Japan for presenting my results to the collaboration. Secondly, my thanks go out to Peter ‘vögeln’ Lewis for all his wise advice during the year and putting up with multiple drafts of my thesis, the KEK dorm party is something I won’t forget soon.

The Belle analysis office wouldn’t be the same without Stephan, thanks so much for being the babysitter for Svenja and myself during the last year. ‘What did you expect?’ has become part of our dictionary, and all the random discussions helped us through some painfully long grid downloading times. Arigato gozaimasu! That brings me to Svenja, my partner in crime. Often we were seen as one and the same person, but we prooved them wrong at B2GM. Taking over my analysis during your PhD, and switching from C++ to Python, what more can I ask for! I also want to thank the rest of the Belle analysis group: bread-baking Alexander, Christian the ice man, ‘just a bachelor student’ Jonathan, truth-matching Lu, and DB Jun. Work-related or not, there was always something to talk about during our Mensa lunches. You are all more than welcome in Amsterdam for a beer (or two), and let the LHCb vs Belle II battle begin. The same goes for my friends who I met along the way during my masters: Srijan, Georgios, Bence, and Guillermo. In the end, we did all pass AQT (sorta).

Outside the world of physics, I would like to say thank you to my parents and my brother. Your support over all the years has kept me going forward in my studies! Special thanks also goes to Matt and Amy, the two most wonderful housemates. Together we kept the kitchen clean! Finally, my thanks go out to Michelle. You put up with my complaining about changing Monte Carlo versions (twice!), long downloading times, the lack of data, and many other things. Without you, it wouldn’t have been the same.

Contents

1 Introduction **1**

2 Theory **3**

2.1 The Standard Model of particle physics . . . 3
2.2 Weak interaction . . . 4
2.3 $B\overline{B}$ production . . . 5
2.4 $B \to \rho\ell\nu$ . . . 5

3 SuperKEKB, Belle II, and Multivariate Analysis **9**

3.1 SuperKEKB accelerator . . . 9
3.2 Belle II detector . . . 10
3.2.1 Vertex detector . . . 10
3.2.2 Central Drift Chamber . . . 11
3.2.3 Particle Identification . . . 11
3.2.4 Electromagnetic Calorimeter . . . 11
3.2.5 K_L^0 and μ detector . . . 11
3.3 Multivariate analysis . . . 12

4 Particle reconstruction at Belle II **13**

4.1 Trigger system . . . 13
4.2 Track reconstruction . . . 13
4.3 ECL cluster reconstruction . . . 14

5 Signal selection **15**

5.1 Signal and background definitions . . . 15
5.2 Pre-selections and reconstruction improvements . . . 16
5.2.1 ROE selections . . . 17
5.2.2 $\theta_{\gamma\gamma}$ . . . 18
5.2.3 nCleanedTracks . . . 19
5.2.4 θ_{miss} . . . 20
5.2.5 ρ^0 vertex fit χ^2 . . . 21
5.2.6 Fit region . . . 22
5.3 Signal-specific selections . . . 24
5.3.1 $\cos\theta_{BY}$ . . . 24
5.3.2 Momentum selection . . . 26
5.3.3 m_ρ . . . 27

5.3.4 Efficiencies . . . 28
5.4 Best candidate selection . . . 28
5.5 Continuum and $B\overline{B}$ background suppression . . . 29
5.5.1 Continuum suppression . . . 29
5.5.2 $B\overline{B}$ suppression . . . 34
5.5.3 Figure of Merit optimization . . . 38

6 Signal extraction results and discussion 41
6.1 Used fitting method . . . 41
6.2 Fit results and discussion . . . 44
6.3 Possible systematic uncertainties . . . 46
6.4 Possible analysis improvements . . . 47

7 First look at early Belle II data 49
7.1 Pre-selections . . . 49
7.2 Pre-BDT selections . . . 53
7.3 BDT fit region . . . 55

8 Summary, Conclusion and Outlook 57
8.1 Summary . . . 57
8.2 Outlook . . . 58

Bibliography 59

A Appendix 63
A.1 $|\Delta t_{\gamma\gamma}|$ selection . . . 63
A.2 Fit region . . . 64
A.3 Momentum selection . . . 65
A.4 Efficiency distributions . . . 66
A.5 Efficiency tables . . . 67
A.6 D^* veto . . . 69
A.7 BDT . . . 73
A.7.1 BDT input variables . . . 73
A.7.2 Importance ranking . . . 83
A.7.3 Overtraining tests . . . 91
A.7.4 Correlation matrices . . . 92
A.7.5 FOM distributions . . . 108
A.8 Signal extraction fit . . . 112
A.8.1 Pull distributions . . . 112
A.8.2 Asimov fits . . . 116
A.8.3 Fit results . . . 120

List of Figures 125

List of Tables 129

CHAPTER 1

Introduction

Nature has always been a source of intrigue to humanity. Already in ancient times, the Greek philosophers and Arabic scientists wondered what the natural world consists of and which laws govern it. Unlike those philosophers and scientists, modern-day physicists have the experimental tools to test their theories. The curiosity to understand nature, however, has not changed. Fundamental questions like 'What is the universe made of?' and 'How did the universe start?' are still (partially) unanswered.

The field of particle physics tries to answer these questions by studying the fundamental building blocks of nature. Research in particle physics has come a long way since the early 1900's, when Thomson discovered the electron [1], Rutherford showed that atoms have a positively charged nucleus [2] and Bohr came up with his famous atomic model [3]. Further understanding of quantum mechanics and quantum field theory improved knowledge on the essence of particles; the concepts of spin, fermions and bosons were introduced [4]. During the the 20th century, experimental methods became more precise and higher energies could be reached, which ultimately lead to the Standard Model (SM) of particle physics during the 1970's [5]. It includes three generations of fundamental quarks and leptons (spin = $\frac{1}{2}$), four gauge bosons (spin = 1) covering the electromagnetic, weak and strong interaction, and a scalar Higgs boson. The SM has been extensively tested experimentally, with most recently the discovery of the Higgs boson by the ATLAS and CMS experiments at CERN [6].

Although the SM is very successful at describing particles and their interactions, there are physical observations that cannot be calculated and explained with it. Parameters like the masses of the particles, the number of particle generations, and the weak interaction mixing Cabbibo-Kobayashi-Maskawa (CKM) matrix are not given by the SM, but have to be determined experimentally. The existence of dark matter and dark energy, the baryogenesis problem, and gravity are all not included in the SM. To address these shortcomings, multiple 'Beyond the Standard Model' (BSM) physics models have been proposed, of which the Minimal Supersymmetric Standard Model is the most well-known [7]. Searching for BSM physics can be achieved by directly observing new BSM particle resonances, which is mostly done at the proton-proton collider experiments at CERN, or by performing high-precision measurements of SM parameters and comparing the experimental values to their SM counterpart. Any significant deviation between theory and experimental results then indicates the presence of BSM physics.

The Belle II experiment is an experiment dedicated to such high-precision SM measurements. It is focused on measurements on B mesons, using $\sqrt{s} = m_{\Upsilon(4S)}$, leading to a 96% pure production of B mesons [8]. One of the studied parameters in the experiment is the CKM matrix element V_{ub}. Since the theoretical SM value of V_{ub} is relatively small, BSM physics could have an observable effect on it. Furthermore, different experimental methods have yielded disagreeing results with a 3σ deviation [9]. These experimental methods differ by either integrating over all possible charmless final states or by selecting a specific one e.g. π or ρ, which are called inclusive and exclusive respectively.

This thesis focuses on the exclusive untagged analysis of the decay $B \to \rho\ell\nu$. Using Belle II Monte Carlo, we extract the branching fraction of the decay and set up the analysis for extracting V_{ub}. Firstly, we will shortly describe the involved theory in the analysis. Secondly, we will give a description of the Belle II detector and the SuperKEKB particle collider. Thirdly, we will provide a short introduction to particle reconstruction at Belle II. Fourthly, we will provide the performed selections in the analysis. We will show and discuss the obtained results and have a first look at early Belle II data. Finally, we present our conclusion and a further outlook.

CHAPTER 2

Theory

This chapter will provide an overview of the theory involved in this analysis. The first section will discuss the SM, the second the weak interaction, the third $B\overline{B}$ production, and the final section $B \to \rho\ell\nu$ decays. More detailed descriptions of the SM can be found in particle physics textbooks, e.g. *Modern Particle Physics* [10].

2.1 The Standard Model of particle physics

The Standard Model of particle physics is the quantum field theory describing the interactions of elementary particles. It describes three of four known fundamental forces of nature: the electromagnetic, weak, and strong interactions. A schematic overview of the SM is shown in figure 2.1. According to the SM, all matter is made out of three types of elementary particles: quarks, leptons and force mediators. The leptons and quarks have half-integer spin and are called fermions, while the force mediators have integer spin and are called bosons. An exception to this is the Higgs boson, which has spin zero.

Unlike quarks, the leptons do not interact strongly. Both the leptons and the quarks fall into three generations, which is often referred to as the mass hierarchy. Every generation has one quark with charge $+\frac{2}{3}e$, one quark with charge $-\frac{1}{3}e$, a lepton with charge $-e$ and a neutral, massless neutrino. Stable matter consists of only first generation particles. For every particle there is also a corresponding anti-particle with opposite (colour) charge.

There are six ‘flavours’ of quarks: u, d, s, c, b, and t. Each quarks comes in one of 3 colours: red, green or blue, representing the charge of the strong force. The strong force is described by Quantum Chromodynamics (QCD) and is an SU(3) symmetry group. The force carrier of the strong force is the massless, colour-charged gluon, of which there are 8 types. QCD leads to the concept of colour confinement, which implies that quarks cannot be observed free, but only in colour-neutral bound states. These states consist of either a quark and anti-quark, which are called mesons, or a combination of three (anti-)quarks, which are called baryons.

All charged particles interact with the electromagnetic interaction, described by Quantum Electrodynamics (QED). QED is a U(1) symmetry, which gives rise to the massless photon (γ) as its boson. The weak interaction is a SU(2) symmetry, with the massive $W^{\pm}$ and Z^0 bosons as force carriers. Neutrinos only interact weakly, since they do not have colour and charge.

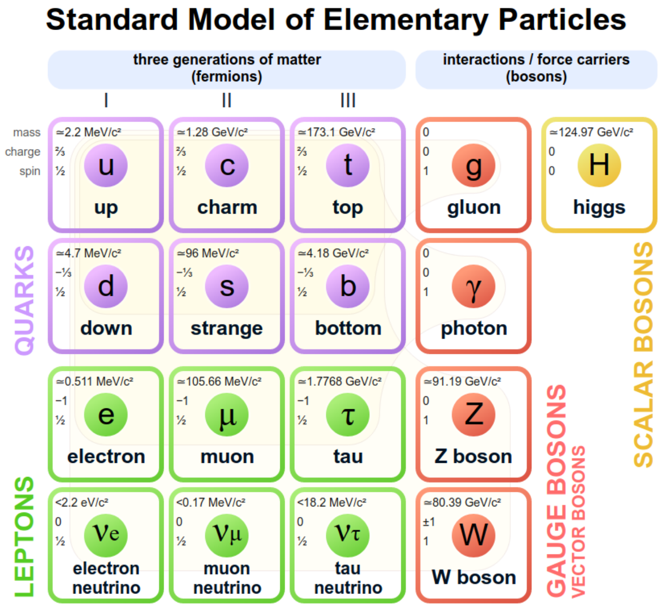


Figure 2.1: Schematic depiction of the Standard Model of particle physics [11].

2.2 Weak interaction

The weak interaction is different from the other forces in the SM in multiple ways. Firstly, it is the only force with massive gauge bosons, with $m_{W^{\pm}} \approx 80$ GeV and $m_{Z^0} \approx 90$ GeV. This is caused by the spontaneous symmetry breaking of the electroweak $\mathrm{U}(1) \times \mathrm{SU}(2)$ symmetry due to the Higgs mechanism [12]. Secondly, the weak interaction can violate parity and charge in decays due to its V-A nature [13].

Lastly, in the SM the weak interaction is the only force that can change quark and lepton flavour. This process is described by the Cabbibo-Kobayashi-Maskawa (CKM) matrix, which relates the mass eigenstates of the quarks to the weak interaction eigenstates [14]. This matrix is given in equation 2.1.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \text{ with } V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{2.1}$$

The values of the CKM matrix relate to the relative probability that a quark of a certain type decays into a different quark type. $|V_{ub}|^2$, for example, is the probability that a bottom quark decays into an up quark. The CKM matrix favours decays within the same quark generation (e.g. $u \longrightarrow d$), with the corresponding elements being of order unity. Decays between one generation (e.g. $c \longrightarrow d$) are suppressed, and decays between two generations (e.g. $b \longrightarrow u$) are doubly suppressed, with V_{ub} and V_{td} being $\mathcal{O}(10^{-3})$.

Decays involving the doubly suppressed CKM elements, V_{ub} and V_{td}, are intensively being studied in flavour physics. Since the SM predicts low branching fractions for these type of decays, it is a good probe for indirect BSM effects. However, these studies are experimentally difficult, since the decays are rare and therefore hard to measure.

2.3 B$\overline{\text{B}}$ production

At B-factories, $B\overline{B}$ pairs are produced by colliding a positron with an electron exactly at the mass of the Υ(4S) resonance, 10.579 GeV. The Υ(4S) consists of a $b\overline{b}$ pair, which then strongly decays into two B mesons. This process is shown in the Feynman diagram in figure 2.2.

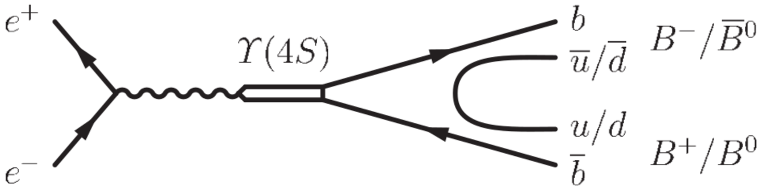
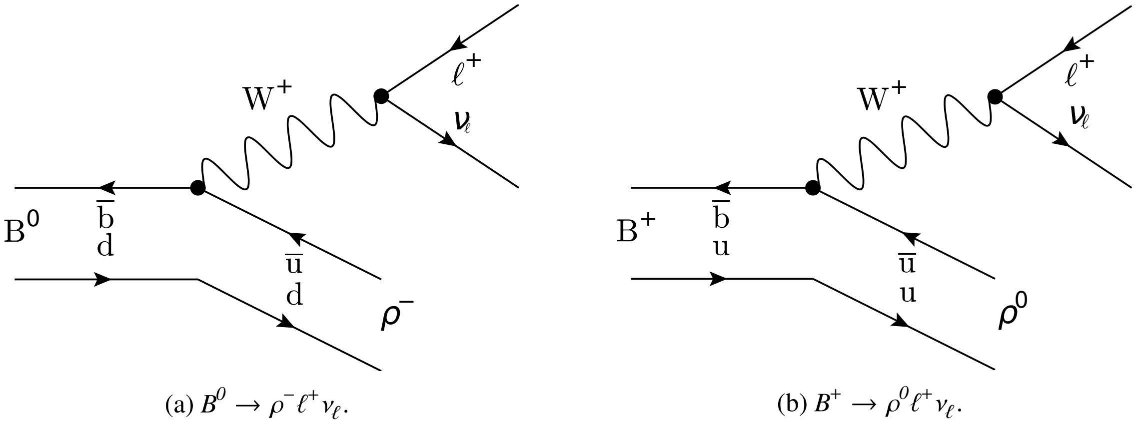


Figure 2.2: Feynman diagram for the process $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ [15].

The branching fraction of the decay $\Upsilon(4S) \to B^+B^-$ is $(51.4 \pm 0.6)\%$ and for $\Upsilon(4S) \to B^0\overline{B}^0$ it is $(48.6 \pm 0.6)\%$ [16]. Next to $e^+e^- \to \Upsilon(4S)$, the positron and electron can also annihilate into a $f\overline{f}$ pair. This is called continuum, and is a background to the $B\overline{B}$ decays.

2.4 B $\to \rho\ell\nu$

This thesis focuses on the decays $B^0 \to \rho^\pm(\to \pi^\pm\pi^0)\ell^\mp\nu_\ell$ and $B^\pm \to \rho^0(\to \pi^\pm\pi^\mp)\ell^\pm\nu_\ell$, where ℓ can be either an electron or a muon. The world-average branching fractions of these decays are $(1.58 \pm 0.11) \times 10^{-4}$ and $(2.94 \pm 0.21) \times 10^{-4}$ respectively [16]. The tree-level Feynman diagrams for both decays are shown in figure 2.3.



(a) $B^0 \to \rho^-\ell^+\nu_\ell$.

(b) $B^+ \to \rho^0\ell^+\nu_\ell$.

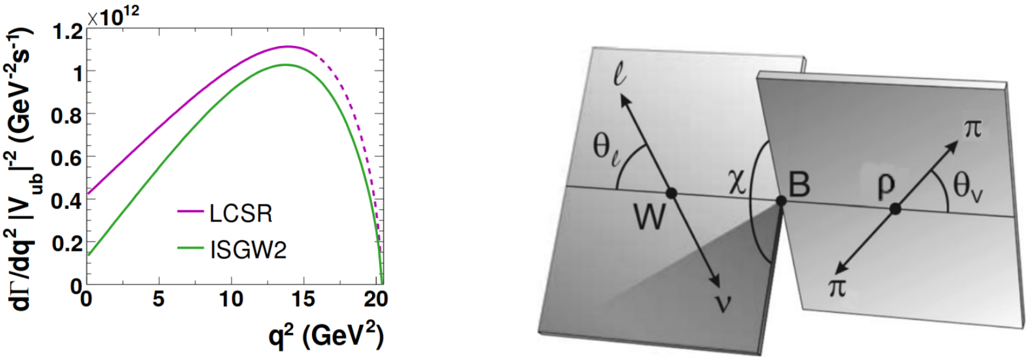
Figure 2.3: Tree-level Feynman diagrams for the studied decays.

The decays can be used to determine the magnitude of V_{ub}. In semileptonic B decays, the hadronic and leptonic components of the matrix element can be factorized [17]. The hadronic part can be described using so-called form factors, which provide an approximation of the $q - \overline{q}$ QCD interactions inside the B meson. Multiple methods exist to calculate the form factors, using quark-model calculations (ISGW2) [18], QCD light-cone sum rules (LCSR) [19], and, for scalar mesons, lattice QCD calculations (LQCD). The form factors are functions of the kinematic quantity q^2, also known as the momentum transfer, which is defined as

$$\begin{aligned} q^2 = m_W^2 &= (P_\ell + P_\nu)^2 \\ &= (P_B - P_\rho)^2 = M_B^2 + m_\rho^2 - 2M_B E_\rho \end{aligned} \tag{2.2}$$

where m_W^2 is the mass squared of the virtual W boson, P_ℓ and P_ν the four-momenta of the lepton and the neutrino, M_B the mass of the B meson and m_ρ, E_ρ, P_ρ are respectively the mass, energy and four-momentum of the ρ meson. In this analysis, the latter definition is used to calculate the reconstructed q^2. Figure 2.4(a) shows the predicted q^2 distributions for ISGW2 and LCSR calculations. In addition to q^2, the form factors depend on the three helicity angles: θ_l, θ_V and χ, shown in figure 2.4(b). They are defined as follows:

- θ_l is the angle between the W direction in B rest frame and the lepton direction in the W rest frame.
- θ_V is the angle between the ρ direction in the B rest frame and the π direction in the ρ rest frame.
- χ is the angle between the ρ and W decay planes.



(a) Predicted q^2 distributions based on QCD light-cone sum rules (purple) and quark-model calculations (green) [17]. The dashed line indicates extrapolation of the LCSR predictions into all q^2 regions.

(b) Schematic distribution of the helicity angles θ_l, θ_V and χ for the signal decay.

Figure 2.4: The form factor parameters.

Since the ρ is a vector meson, the polarization vector of the meson is important for the decay. For low mass leptons, the hadronic current can be simplified and only three form factors play a role: the axial form factors $A_1(q^2)$ and $A_2(q^2)$, and the vector form factor $V(q^2)$. It is more common to express the differential decay rate in terms of the helicity amplitudes, which correspond to the ρ meson's three helicity states:

$$\begin{aligned} H_{\pm}(q^2) &= (M_B + m_\rho)\left[A_1(q^2) \mp \frac{2M_B p_\rho}{(M_B + m_\rho)} V(q^2)\right], \\ H_0(q^2) &= \frac{M_B + m_\rho}{2m_\rho\sqrt{q^2}} \times \left[(M_B^2 - m_\rho^2 - q^2)A_1(q^2) - \frac{4M_B^2 p_\rho^2}{(M_B + m_\rho)^2} A_2(q^2)\right] \end{aligned} \tag{2.3}$$

where p_ρ is the momentum of the ρ meson. The differential decay rate can then be written as:

$$\frac{d\Gamma(B \rightarrow \rho \ell \nu)}{dq^2 \cos\theta_l} = |V_{ub}|^2 \frac{G_f^2 p_\rho q^2}{128\pi^3 M_B^2} \times \left[\sin\theta_l^2 |H_0|^2 + (1 - \cos\theta_l)^2 \frac{|H_+|^2}{2} + (1 + \cos\theta_l)^2 \frac{|H_-^2|}{2}\right] \tag{2.4}$$

where G_f is the Fermi coupling constant. The magnitude of V_{ub} can be determined by comparing the measured partial branching fraction to the theoretical decay rate prediction:

$$|V_{ub}| = \sqrt{\Delta B(q_{min}^2, q_{max}^2)/\tau_B \Delta\xi_{th}(q_{min}^2, q_{max}^2)} \tag{2.5}$$

with $\Delta B(q_{min}^2, q_{max}^2)$ the obtained partial branching fraction in the measured q^2 range, τ_B the mean lifetime of the B meson, and $\Delta\xi_{th}(q_{min}^2, q_{max}^2)$ the theoretical decay rate taken from ISGW2 or LCSR calculations.

CHAPTER 3

SuperKEKB, Belle II, and Multivariate Analysis

This chapter will provide a brief description of the SuperKEKB accelerator, the Belle II detector setup, and multivariate analysis. For a more detailed description of the experiment, we refer to the Belle II Technical Design Report [8] and the KEK Super *B* Factory Letter of Intent [20].

3.1 SuperKEKB accelerator

SuperKEKB is an electron-positron collider located at KEK in Tsukuba, Japan. It has been specifically developed for the Belle II experiment, with a center of mass (CMS) collision energy of 10.58 GeV, corresponding to the $\Upsilon(4S)$ resonance mass. The planned instantaneous luminosity is $\mathcal{L} = 8 \cdot 10^{35} \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, approximately 40 times higher than its predecessor KEKB. A schematic drawing of the accelerator is shown in figure 3.1.

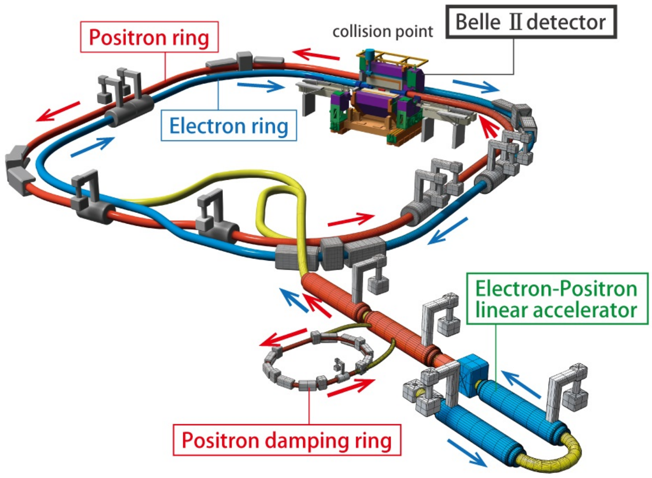


Figure 3.1: Schematic drawing of the SuperKEKB accelerator [21].

The beam energies of the collider are asymmetric, with $E_{e^-} \approx 7$ GeV and $E_{e^+} \approx 4$ GeV. This leads to the $\Upsilon(4S)$ CMS frame being Lorentz boosted with respect to the laboratory frame compared, which is needed to resolve the *B* decay vertices.

3.2 Belle II detector

The Belle II detector is a hermetic particle detector, consisting of different cylindrical and endcap detector component layers around the interaction point and a 1.5 T superconducting solenoid magnet which is required for particle tracking. A schematic overview of the detector is shown in figure 3.2.

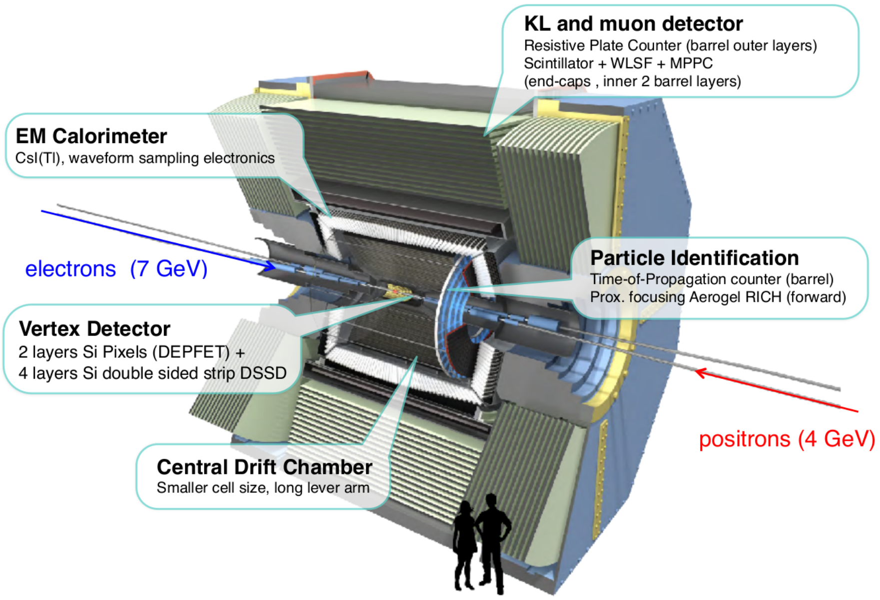


Figure 3.2: Schematic drawing of the Belle II detector, comparing to its predecessor Belle [8]. The most important upgrades compared to Belle are the improved PID system and the vertex detector.

The following sections will give a short description for each of the Belle II detector components.

3.2.1 Vertex detector

The heart of the detector contains two (during early data taking only one) layers of silicon Pixel Vertex Detectors (PXD), followed by four layers of silicon Double Sided Strip Detectors. These two systems combined are called the Vertex Detector (VXD). The PXD is based on the Depleted Field-Effect Transistor (DEPFET) technology, with a pixel size of 50×50 µm for the first layer and 50×70 µm for the second layer. The four DSSD layers consist of multiple ladders, which overlap $\sim 10\%$ warranting that a traversing particle always hits at least two ladder. The VXD has a hit resolution of approximately 10 µm.

3.2.2 Central Drift Chamber

Surrounding the VXD is the Central Drift Chamber (CDC), which, in combination with the VXD, is used for particle tracking. It is also used to measure the momentum of particles using the track curvature caused by the magnetic field from the superconducting solenoid. The CDC uses a gas mixture consisting of 50% He and 50% C_2H_6, and consists of 56 cylindrical layers of wires which are divided into 9 super-layers. The CDC radial resolution is $\sim$ 100 μm and the axial resolution is $\sim$ 2 mm.

3.2.3 Particle Identification

The CDC is followed by a Time-of-Propagation counter in the barrel region and an Aerogel Ring-Imaging Cherenkov detector (ARICH) at the forward region of the detector. Both detectors use Cherenkov radiation to determine the type of particle, since the angle at which the radiation is emitted depends on the particle's mass. The TOP consists of 16 quartz crystals, which reflect the Cherenkov light to photo-multiplier tubes that measure the position and time of the photons. The ARICH consists of an silica aerogel in which the Cherenkov radiation photons are created, an expansion volume, and an array of photo-detectors. It directly measures the circle of Cherenkov photons, with a smaller circle radius indicating a higher particle mass. The information from the TOP and ARICH, together with the CDC, is used to calculate a log-likelihood function for every particle type hypothesis: the particle PID. The functions return the relative probability that the particle hypothesis is correct.

3.2.4 Electromagnetic Calorimeter

Behind the PID system, the Electromagnetic Calorimeter (ECL) is located, which primarily provides energy measurements for photons. It consists of a barrel section and two end-caps at the front and back of the detector, and has an acceptance between 12° and 155°. In total, it is made up of 8736 thallium activated cesium iodide CsI(TI) crystals, in which particles deposit their energy due to electromagnetic showering. The resolution is empirically determined as

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{0.066\%}{E}\right)^2 + \left(\frac{0.81\%}{\sqrt[4]{E}}\right)^2 + (1.34\%)^2} \tag{3.1}$$

which corresponds to a resolution of $\sim$ 2% at $E = 100$ MeV.

3.2.5 $\mathbf{K^0_L}$ and μ detector

The final detector component is the K-Long and Muon detector (KLM). As the name suggests, it is used to measure the particles which are not stopped in the electromagnetic calorimeter: the K_L^0 and μ. It consists of plastic scintillators with silicon photo-multipliers, and Resistive Plate Chambers (RPC). The silicon photo-multipliers have a time resolution smaller than 1 ns. The RPCs contain a gas mixture composed of 62% HFC-134a, 30% argon, and 8% butane-silver.

3.3 Multivariate analysis

Multivariate analysis (MVA) tools are often used in particle physics. MVA methods map a multi-dimensional input vector to a single output classifier to separate signal from background. Compared to normal selections on variables, this has the advantage of the MVA method using correlations between the variables which might not be easily visible to the user. MVAs are trained on a statistically independent dataset, to obtain a non-biased classification results when applying to user data.

A commonly used MVA method is a Boosted Decision Tree (BDT). BDTs work using a series of binary selections, nodes, which form a tree structure. A schematic drawing of a BDT is shown in figure 3.3. These selections are performed consecutively on the input variables, dividing the the n-dimensional feature space into different regions (so-called leaves). The output classifier gives an indication of how 'signal-like' a candidate is. Gradient Boosting is used to improve the BDT performance by assigning weights to every event in the training [22]. Wrongly classified events will receive a higher weight, while correctly classified events get a lower weight. These steps are then repeated several hundred times. The final BDT classifier output is the weighted sum of all the trained decision trees outputs. This thesis uses FastBDT for the training of BDTs [23].

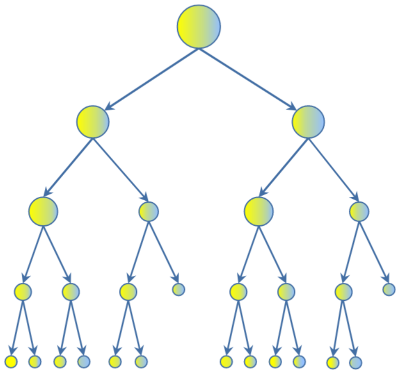


Figure 3.3: Schematic drawing of a Boosted Decision Tree [24]. In this example, the two colours indicate the background and signal categories. Each circle represents a leaf of the tree, while each arrow indicates a certain selection on one of the input variables. Each level in the diagram represents a decision layer in the tree.

CHAPTER 4

Particle reconstruction at Belle II

This chapter will provide a brief description of particle reconstruction at Belle II. Firstly, the Belle II trigger system will be described. Secondly, charged particle track reconstruction will be discussed. Finally, the reconstruction of neutral particles in the ECL will be explained. For a more detailed description, see the PhD thesis of Sara Pohl [25] and the Belle II Technical Design Report [8].

4.1 Trigger system

The trigger system tells the detector when events occur, if they should be processed and recorded, and when events can be ignored. This is necessary due to limitations in computing power and data storage, making the storing of every event impossible. The Belle II trigger system consists of two layers: the Level 1 trigger and the high level trigger. Physics events leave a signature in the detector, e.g. a number of ECL clusters, number of tracks in the CDC, and total energy deposit in the ECL. These parameters are then used in the trigger to decide if an event happened, and if it should be recorded or not. The triggers consist of multiple subtriggers: the track trigger, the ECL trigger, the TOP trigger and the KLM trigger. All these triggers are fed into a Global Decision Logic (GDL). By combining all the subtrigger information, the GDL then decides whether the data should be recorded or not. The GDL also ignores unwanted physics events, e.g. Bhabha scattering. If the event should be recorded, the GDL sends a trigger signal to the Data Acquisition system which stores the data. If the trigger conditions are not met, the event will not be stored and the detector components will be reset to their initial state.

4.2 Track reconstruction

Charged particles are reconstructed in the CDC using their charge deposit in the gas. Charge deposit occurs due to ionization, where the charged particles collide with the gas atoms and release an electron, which due to the strong electric field at the anode wires gets amplified via a charge avalanche, and is finally collected by the CDC wires. The CDC information, the amount of charge and location, is complemented using the VXD, where the charged particle creates electron-hole pairs, which are collected and read out. The VXD is used to find the production vertex of the particle, to e.g. distinguish between particles produced at the primary vertex or at a secondary decay. Instead of using individual

CDC hits (charge collected on one wire in the CDC), Belle II uses track segments. A track segment is defined as an arrangement of 5 CDC wire layers into a so-called ‘super layer’. An example of three different track segments is shown in figure 4.1. In the segment, one wire is the ‘first priority’ wire which is used to find the direction of the traversing particle by looking at the hits on the following wires. The advantage of using track segments instead of individual CDC wire hits is that less information has to be stored, while keeping information such as the left/right crossing of the track. It does, however, come with a reduced accuracy due to the low number of hits used in the segment.

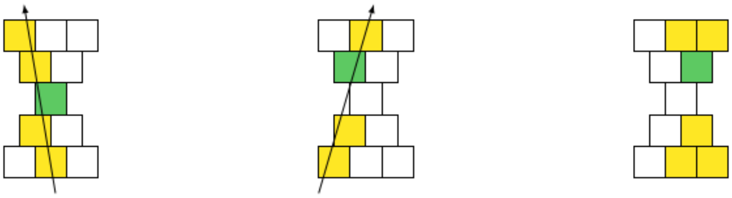


Figure 4.1: Three different track segment hits, with the priority wire coloured green [25]. Left: left passage. Middle: right passage. Right: undecided passage. The squares represent individual track segments, lined up in rows as a schematic view of the CDC. The black lines indicate the reconstructed track through the track segments.

From the track segments, the track is reconstructed using a Hough transformation to combine the individual segments into circles in the transverse plane [25]. From this, the track’s azimuth angle ϕ and the circle radius r are obtained. The radius of the track is related to the strength of the magnetic field, B, and the transverse momentum of the particle, p_T by:

$$r = \frac{p_T}{0.3B} \tag{4.1}$$

where r is given in meter, p_T in GeV and B in T. Two-dimensional least square fits are performed on the circles taking drift times into account, after which a three-dimensional fit is performed to obtain the polar angle θ and the z value. This process is performed for every track in the event, after which the (r,ϕ,θ,z) parameters are returned for every track in the event, together with the total number of tracks.

4.3 ECL cluster reconstruction

Photons are reconstructed using their energy deposit due to electromagnetic showering in the ECL. The ECL uses 4x4 crystal cells, called trigger cells, to match energy deposits to a particle. If a trigger cell satisfies the so-called Isolated Cluster Number (ICN), a 3x3 rectangle of trigger cells around the ICN is taken and the cell with the highest energy is selected. A 3x3 grid around this highest energy cell is selected, and the sum of the cells inside this grid equals the energy of the cluster. This is then returned as the energy of the particle. The position of the cluster is calculated by taking a weighted average over the cells in the grid, while the timing is taken from the highest energy cell. Next to reconstructing photons, tracks from charged particles, like e and π, can be matched with ECL clusters. This is done by extrapolating the track to the radius of the ECL and finding the cluster that is closest in the transverse direction. Some other neutral particles, like K_L^0, shower hadronically and can be detected by their energy deposit.

CHAPTER 5

Signal selection

This chapter will describe the signal selection performed for the $B \to \rho\ell\nu$ analysis. Since the branching fraction of the $B \to \rho\ell\nu$ decay is relatively small, $\mathcal{O}(10^{-4})$, not many events were expected. Background channels, like $B \to D^*\ell\nu$, were expected to dominate. It was therefore necessary to perform selections taking into account what distinguishes the signal decay from the background, e.g. the higher p^*_ℓ and the lower mass of the ρ compared to X_c hadrons. Especially early on in the Belle II experiment, the analysis will be limited by the statistics of the available data. Therefore it was important to maximize signal retention, while still rejecting enough background to obtain a visible signal, which lead to the use of MVA techniques instead of more ‘traditional’ selections.

The Monte Carlo (MC) simulation used in the analysis is the twelfth official MC campaign produced by the Belle II data production group [26], corresponding to 160 fb^{-1}. The MC consists of $B^0\overline{B}^0$ and B^+B^- MC and continuum MC consisting of $e^+e^- \to u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}, \tau^-\tau^+$. MC containing only signal decays was used for efficiency measurements and for MVA training. All selections are optimized for an expected early phase 3 data size of 10 fb^{-1} and applied in consecutive order, while graphs are shown for 3 fb^{-1}, corresponding to the available early phase 3 data size at the moment of writing this thesis [27].

5.1 Signal and background definitions

The signal and background categories are split based on the mother of the charged lepton. The used subdivision is explained below:

- Signal: two types where for both the lepton originates from a signal decay.
  - *True signal*: both pions originate from the signal decay.
  - *Combinatorial signal*: either one or both pions originate from the second B meson.
- Continuum: candidates originating from $e^+e^- \to q\bar{q}$ or $e^+e^- \to \ell^+\ell^-$ events.
- $B \to D^*\ell\nu$: the lepton originates from a $B \to D^*\ell\nu$ decay, pions can come from either B.
- $B \to \pi\ell\nu$: the lepton originates from a $B \to \pi\ell\nu$ decay, the pions can originate from either B meson.

- $B \to X_u \ell \nu$: the lepton candidate originates from a charmless semi-leptonic B decay that is not the signal decay mode or the π mode. The pions can originate from either B.
- Other $B\overline{B}$: remaining background sources from other B decays, categorized in order to prevent overlap.
  - *Fake lepton*: the lepton candidate is a misidentified charged hadron or wrong lepton type.
  - *Fake pion*: a pion candidate is a misidentified hadron or lepton.
  - *Secondary lepton*: the true lepton originates from a particle other than a B.
  - $B \to D\ell\nu$: the lepton originates from a $B \to D\ell\nu$ decay.
  - $B \to D^{**}\ell\nu$: the lepton originates from a $B \to D^{**}\ell\nu$ decay.
  - Other $B\overline{B}$ events, including D_s decays and baryonic decays.

5.2 Pre-selections and reconstruction improvements

The reconstruction of the decay is performed using certain pre-selections. The pre-selections for charged particles coming from pre-defined particle lists are shown in table 5.1. For the reconstruction of the neutral π^0, two photons were reconstructed with an invariant mass $0.124 < m_{\gamma\gamma} < 0.140$ GeV, leading to a π^0 reconstruction efficiency of 40%. A B vertex fit was performed using TreeFitter [28]. The vertex fit was required to converge, by requiring the χ^2 probability to exceed 0.

	$\pi^\pm$	$\ell^\pm$
PID	> 0.098	> 0.800
dr [cm]	< 0.5	< 0.5
dz [cm]	$-2.0 < dz < 2.0$	$-2.0 < dz < 2.0$
$\cos\theta$	In CDC acceptance	In CDC acceptance
Track χ^2 prob.	> 0.001	> 0.001
NCDC hits	> 20	> 20
p^* [GeV]	-	> 1.0

Table 5.1: Pre-selections for charged particles.

5.2.1 ROE selections

Unwanted background particles in the rest of event (ROE), all particles that do not belong to the reconstructed signal B candidate, can lead to wrong distributions in the ROE and continuum suppression variables. Ideally, the ROE would only consist of particles coming from the non-signal B in the event. However, the event also contains tracks and clusters coming from beam background, and including these in the ROE leads to wrongly calculated ROE variables. To remove these tracks and clusters, and to 'clean up' the ROE, selections were applied on the ROE tracks and clusters. The selections for the ROE clusters were:

- $p^* \geq 0.05$ GeV
- $p^* \leq 3.20$ GeV
- $E > 0.09$ GeV
- ECL cluster timing < 50 ns
- `trackMatchType == 0`: cluster has no matched track

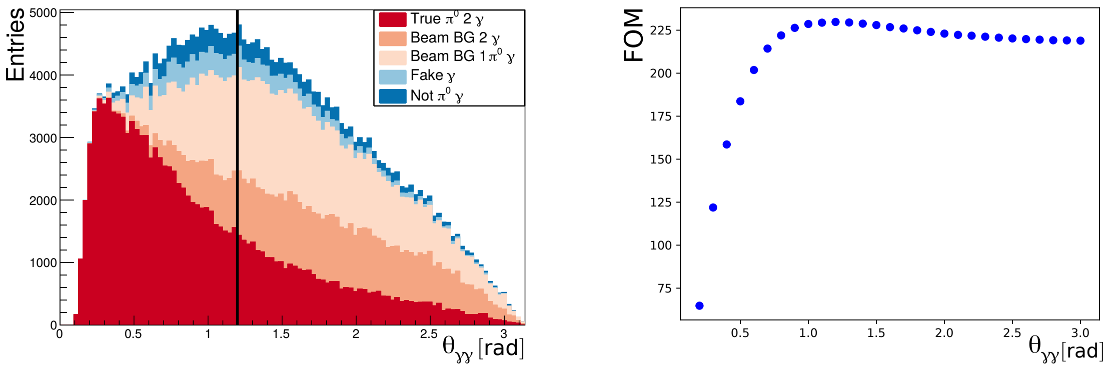
These selections removed low-energy and out-of-time clusters from beam background photons. Only clusters with no matched track were selected, since clusters coming from photons should not have an extrapolated track associated with their cluster. The selections for ROE tracks were:

- Number of CDC hits > 0
- $p^* \leq 3.2$ GeV
- $dr < 2.0$
- $|dz| < 4.0$

By selecting tracks only with a momentum smaller than 3.2 GeV, high-energy tracks coming from non-B decays were excluded. The dr and dz selection made sure only tracks consistent with coming from the interaction point were selected.

5.2.2 $\theta_{\gamma\gamma}$

After the above mentioned pre-selections and ROE selections, it was observed that the dominant background source was the reconstruction of fake π^0 candidates, where the reconstructed photons come from beam background or a combinatoric combination of physics and beam background photons. To improve the π^0 reconstruction, a study was performed on the angle between the two photons in the $\Upsilon(4S)$ center of mass (CMS) frame, $\theta_{\gamma\gamma}$. Since a π^0 from a real physics event will in general be boosted, a relatively small $\theta_{\gamma\gamma}$ is expected, compared to a more isotropic distribution for beam background photons and photons not from a common mother. The $\theta_{\gamma\gamma}$ distribution is shown in figure 5.1(a). To reduce fake π^0 candidates, a figure of merit (FOM) optimization was performed, with FOM= $N^{\text{true}}_{\pi^0}/\sqrt{N^{\text{true}}_{\pi^0} + N^{\text{misrecon}}_{\pi^0}}$. The FOM for different selections is shown in figure 5.1(b). Using the maximum FOM, $\theta_{\gamma\gamma}$ was required to be smaller than 1.20 rad. The selection has a true π^0 retention efficiency of 65% and a fake π^0 background rejection of 71%.



(a) $\theta_{\gamma\gamma}$ distribution and selection (indicated by the black line) for reconstructed π^0 candidates.

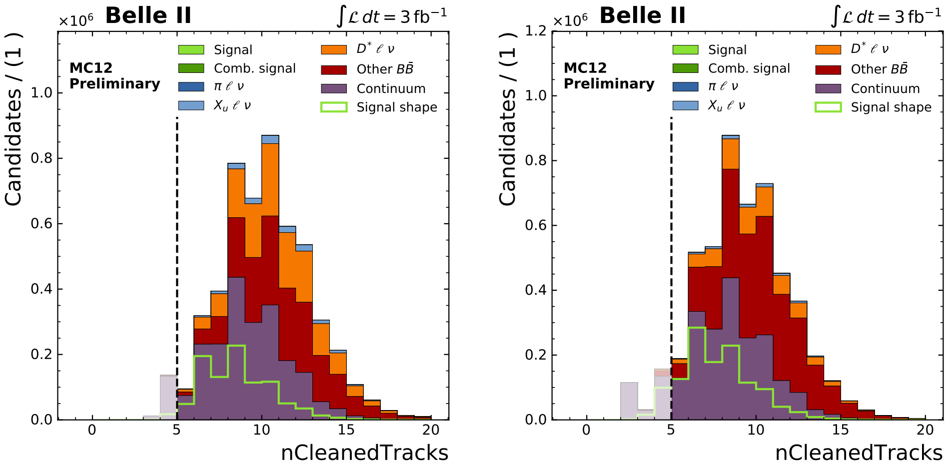
(b) FOM, calculated as $N^{\text{true}}_{\pi^0}/\sqrt{N^{\text{true}}_{\pi^0} + N^{\text{misrecon}}_{\pi^0}}$ for different $\theta_{\gamma\gamma}$ selections.

Figure 5.1: $\theta_{\gamma\gamma}$ selection for improved π^0 background rejection.

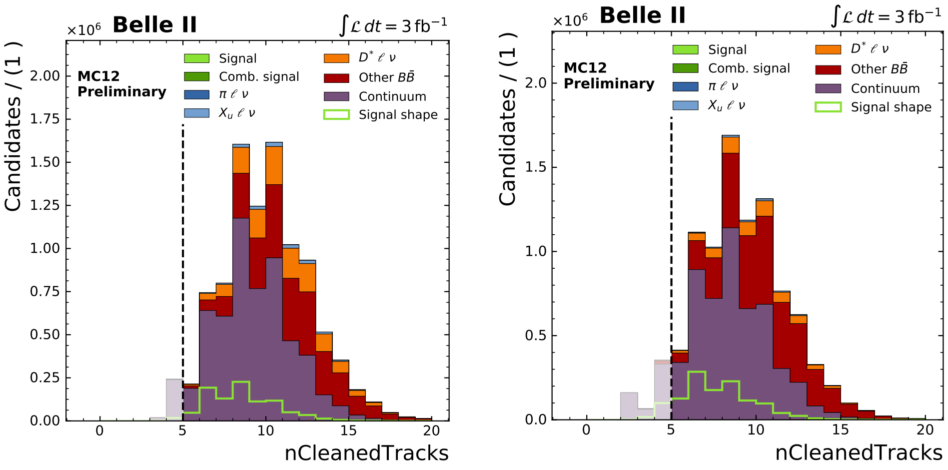
The difference in cluster timing of the π^0 candidate photons was studied as well, specifically targeting beam background photons. Due to the limited signal purity improvement gained with this selection and considering the potential mis-modelling of $|\Delta t_{\gamma\gamma}|$ in MC, no selection was performed. Distributions for $|\Delta t_{\gamma\gamma}|$ before and after the $\theta_{\gamma\gamma}$ selection are shown in appendix section A.1.

5.2.3 nCleanedTracks

To reduce background from $e^+e^- \longrightarrow \tau^+\tau^-$ continuum events and low-multiplicty events which are not included in the used MC simulation, a selection on the number of cleaned tracks was performed, where cleaned tracks are defined by the track $dr < 2.0$ cm and $|dz| < 4.0$ cm. τ particles often decay into e^- and μ^-, which from continuum events can fall in the same energy range as the leptons from semi-leptonic B decays. Almost all the background in the low nCleanedTracks region originates from these events. Since these continuum-type events have a low track multiplicity compared to the signal decay events, they can therefore be removed by requiring the number of cleaned tracks to exceed 4, as shown in figure 5.2 by the black line. The selection has a signal efficiency of 91%.



(a) $B^{\pm} \longrightarrow \rho^0 e^{\pm} \nu_e$ (b) $B^0 \longrightarrow \rho^{\pm} e^{\mp} \nu_e$

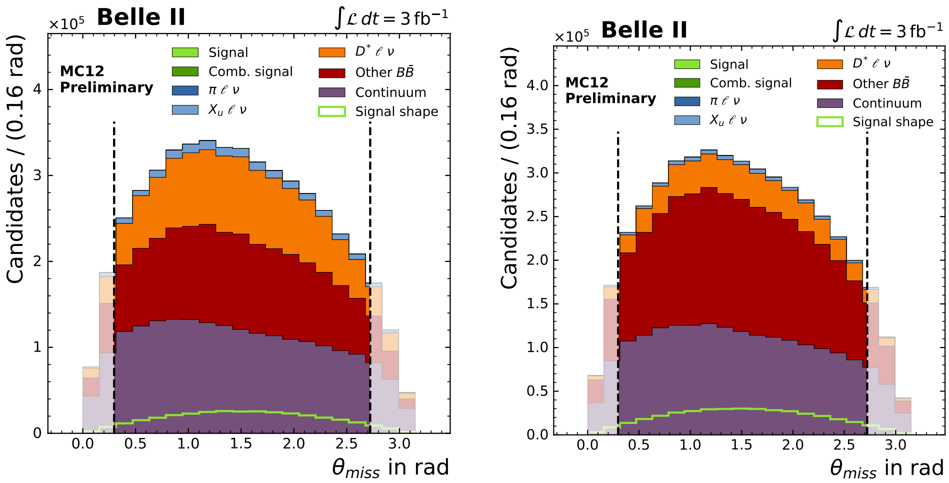


(c) $B^{\pm} \longrightarrow \rho^0 \mu^{\pm} \nu_{\mu}$ (d) $B^0 \longrightarrow \rho^{\pm} \mu^{\mp} \nu_{\mu}$

Figure 5.2: Number of cleaned tracks distributions. The black line indicates the selection of number of cleaned tracks bigger than 4.

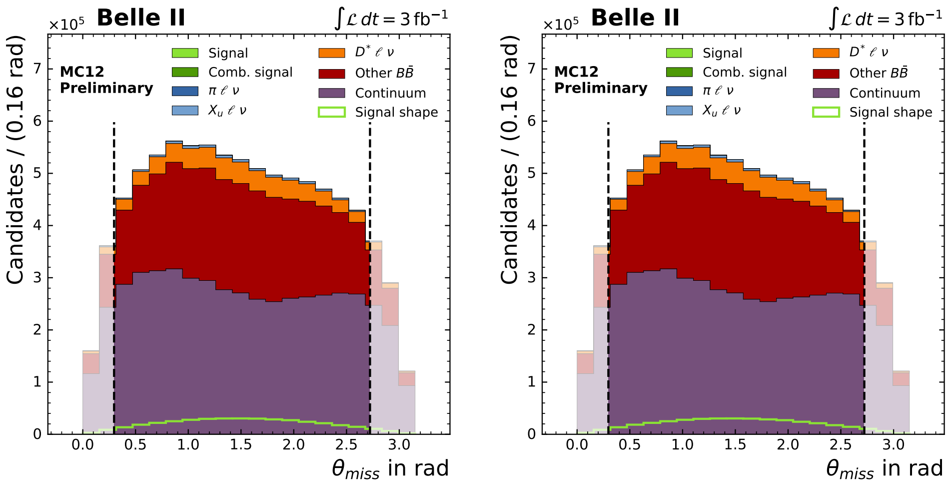
5.2.4 θ_{miss}

Events with a signal decay have at least one neutrino that is not picked up by the detector, and can only be indirectly measured by using the missing momentum of the event, described by $\vec{p}_{\text{miss}}$ and θ_{miss}. To reject candidates where the missing momentum does not come from a neutrino, but from a massive particle that was not picked up in the detector, the angle of the missing momentum vector θ_{miss} was required to be within the detector acceptance range: $17° < \theta_{\text{miss}} < 156°$. Distributions and the performed selections for θ_{miss} with a signal efficiency of 92% are shown in figure 5.3.



(a) $B^{\pm} \to \rho^{0} e^{\pm} \nu_{e}$

(b) $B^{0} \to \rho^{\pm} e^{\mp} \nu_{e}$



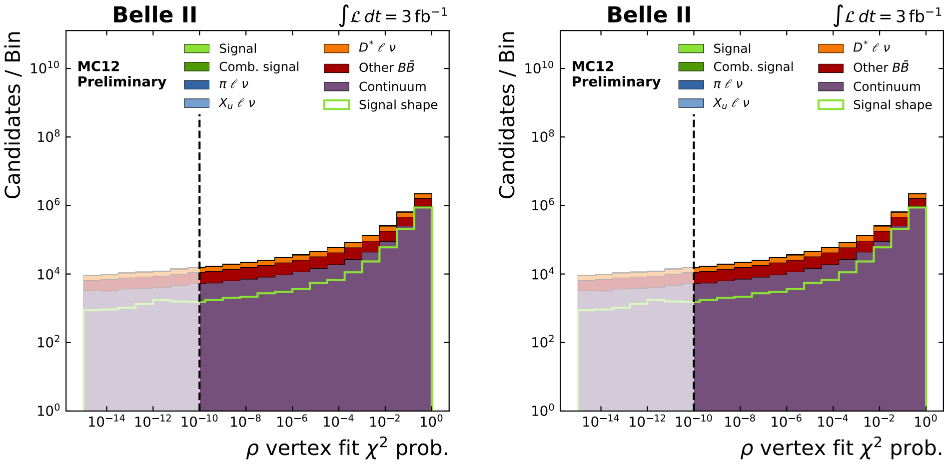
(c) $B^{\pm} \to \rho^{0} \mu^{\pm} \nu_{\mu}$

(d) $B^{0} \to \rho^{\pm} \mu^{\mp} \nu_{\mu}$

Figure 5.3: θ_{miss} distributions. The black lines indicates the selection on the detector acceptance region $17° < \theta_{\text{miss}} < 156°$.

5.2.5 ρ^0 vertex fit χ^2

ρ^0 candidates are reconstructed by matching two $\pi^\pm$ candidates together. For a true ρ^0, the two $\pi^\pm$ should come from the same decay vertex, and therefore the tracks should have a common origin. To improve the ρ^0 reconstruction and reduce the amount of combinatoric ρ^0 background, a decay vertex fit on the two reconstructed $\pi^\pm$ tracks was performed. The fit was done using RAVE fitter [29] integrated within the Belle II software framework. The χ^2 probability returned by the fitter was required to be bigger than 10^{-10}, leading to a signal efficiency of 97%, while rejecting 15% of the combinatoric ρ^0 background. The χ^2 probability distributions are shown in figure 5.4.



(a) $B^\pm \to \rho^0 e^\pm \nu_e$ (b) $B^\pm \to \rho^0 \mu^\pm \nu_\mu$

Figure 5.4: ρ^0 vertex fit χ^2 probability distributions. The black line indicates the selection at 10^{-10}.

5.2.6 Fit region

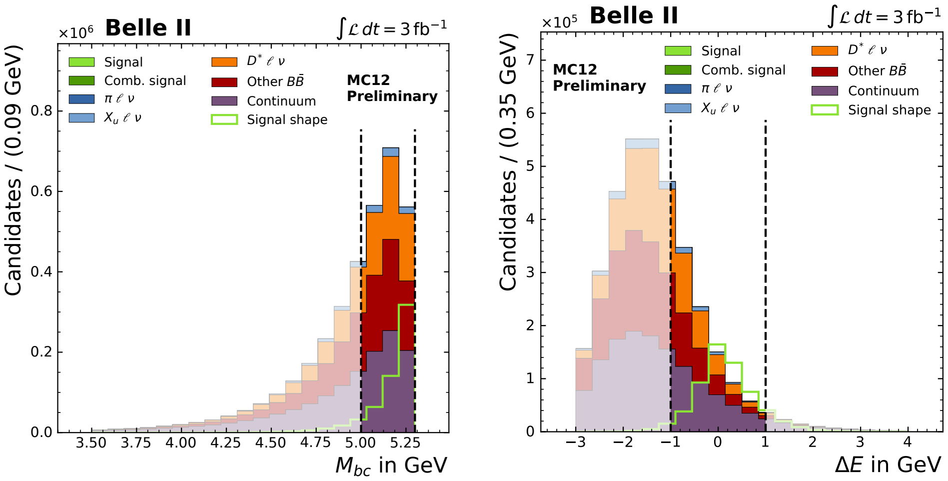
After these selections, the background was dominated by continuum and combinatorial $B\overline{B}$ background. Both these backgrounds can be suppressed by looking at how kinematically consistent the reconstructed B candidate is compared to a true B meson. The beam-constrained mass M_{bc} and the difference between the expected and reconstructed energy of the B candidate ΔE were used for this. In the laboratory frame, the variables are defined as

$$\Delta E = \frac{P_B \cdot P_{e^-e^+} - s/2}{\sqrt{s}} \tag{5.1}$$

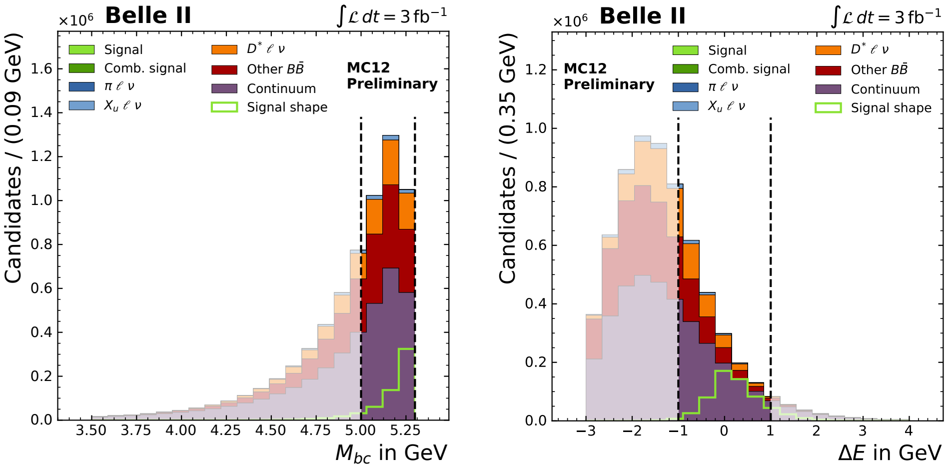
and

$$M_{bc} = \sqrt{\frac{(s/2 + \vec{p}_B \cdot \vec{p}_{e^-e^+})^2}{E^2_{e^-e^+}} - p_B^2} \tag{5.2}$$

where $p_B = \sqrt{E_B^2 - M_B^2}$ is the B momentum in the lab frame, $E_{e^-e^+}$ and $\vec{p}_{e^-e^+}$ are the total beam energy and momentum, and $\sqrt{s}$ is the CMS energy. Since the energy of the colliding e^+ and e^- are precisely known, the resolution of M_{bc} is better than the standard invariant mass of the B candidate [15]. A correctly reconstructed B candidate has $M_{bc} \approx m_B$ and ΔE approximately 0. Distributions for both these variables for $B^{\pm}$ are shown in figure 5.5, while the distributions for B^0 are shown in appendix A.2. Since the neutrino is only approximated by the missing momentum, leading to a lower resolution on $\vec{p}_B$ and therefore more spread out M_{bc} and ΔE distributions, the fit region is selected as $|\Delta E| < 1.0$ GeV and $5.0 < M_{bc} < 5.3$ GeV. The selection has a signal efficiency of 75% and removes 85% of all remaining background.



(a) M_{bc} for $B^{\pm} \longrightarrow \rho^0 e^{\pm} \nu_e$ (b) ΔE for $B^{\pm} \longrightarrow \rho^0 e^{\pm} \nu_e$



(c) M_{bc} for $B^{\pm} \longrightarrow \rho^0 \mu^{\pm} \nu_{\mu}$ (d) ΔE for $B^{\pm} \longrightarrow \rho^0 \mu^{\pm} \nu_{\mu}$

Figure 5.5: M_{bc} (left) and ΔE (right) distributions for $B^{\pm}$ signal decays with the defined fit region $|\Delta E| < 1.0$ GeV and $5.0 < M_{bc} < 5.3$ GeV.

5.3 Signal-specific selections

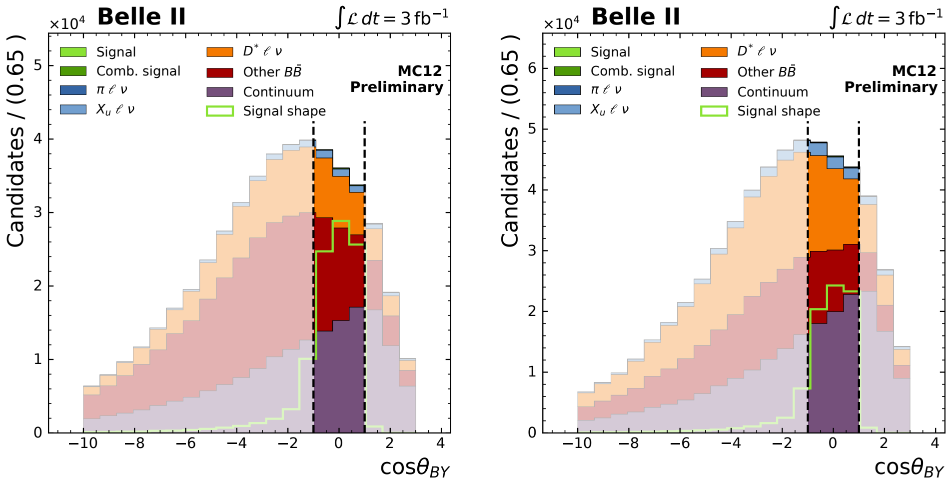
After all the above-mentioned pre-selections, the background was still dominated by continuum events and misreconstructed $B\overline{B}$ events. To further reduce these background, a set of sequentially applied selections was performed.

5.3.1 $\cos\theta_{BY}$

Another variable that was used to remove misreconstructed $B\overline{B}$ background was $\cos\theta_{BY}$ [15]. By combining the reconstructed lepton candidate with the ρ candidate, a so-called Y candidate can be formed. For a correctly reconstructed Y candidate coming from a semi-leptonic B decay, the angle between its momentum vector and the momentum vector of the B, $\cos\theta_{BY}$, can be calculated as:

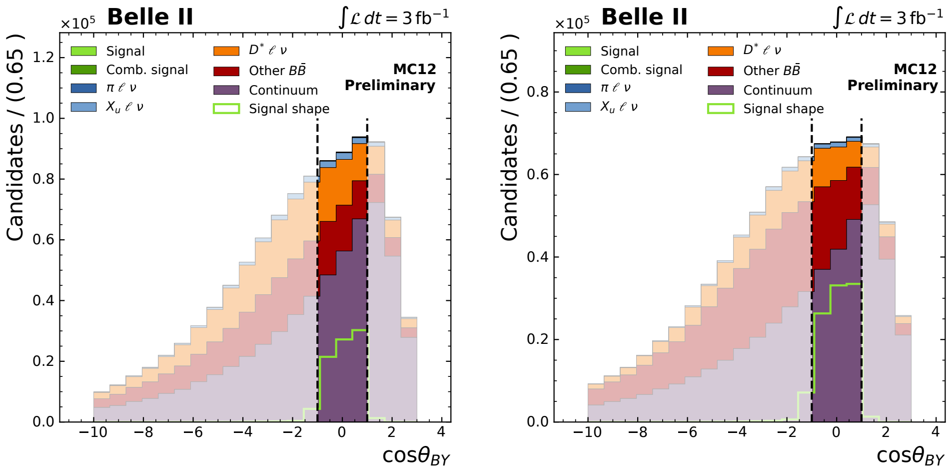
$$\cos\theta_{BY} = \frac{2E_B^* E_Y^* - M_B^2 - M_Y^2}{2p_B^* p_Y^*} \tag{5.3}$$

where $E_B^* = \sqrt{s}/2$ is the average CMS energy of the colliding beams. For signal events $|\cos\theta_{BY}| \leq 1.0$ is expected, but to allow for the detector resolution and photon radiation effects $|\cos\theta_{BY}| \leq 1.1$ was used as the selection. These radiative effects can be seen in the $e^\pm$ decay channel distributions, since the electron mass is lower than the muon mass, and therefore is more likely to undergo Brehmsstrahlung. For non-semi-leptonic B decays and misreconstructed signal decays, the angle θ_{BY} has no physical meaning and is therefore spread out much further than between -1 and 1, as can be seen in the distributions in figure 5.6. The selection has a signal efficiency of 90% and rejects 81% of the total background.



(a) $B^{\pm} \longrightarrow \rho^{0} e^{\pm} \nu_{e}$

(b) $B^{0} \longrightarrow \rho^{\pm} e^{\mp} \nu_{e}$



(c) $B^{\pm} \longrightarrow \rho^{0} \mu^{\pm} \nu_{\mu}$

(d) $B^{0} \longrightarrow \rho^{\pm} \mu^{\mp} \nu_{\mu}$

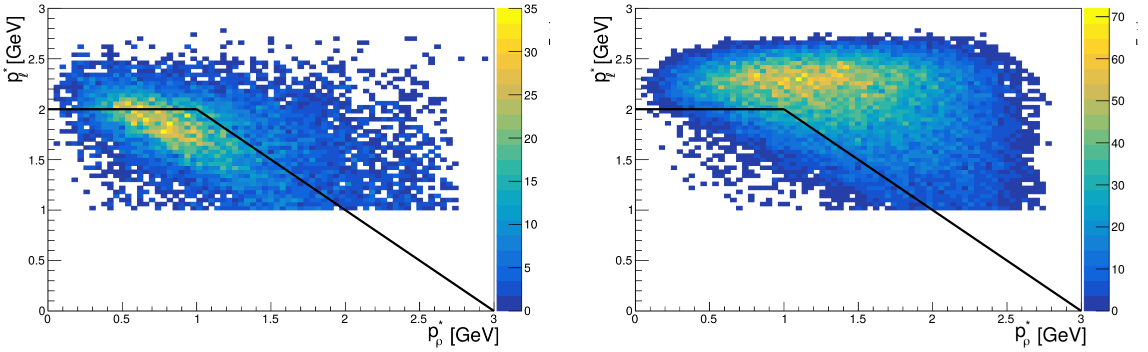
Figure 5.6: $\cos\theta_{BY}$ distributions and the selection between $-1.1 < \cos\theta_{BY} < 1.1$.

5.3.2 Momentum selection

After the $\cos\theta_{BY}$ selection, most background events consisted of continuum events and $B \to X_c \ell \nu$ combinatorial events, where either one or both π come from a different B than the ℓ. To reduce these backgrounds, a selection was performed on the momentum of the hadron (p^*_ρ) and lepton (p^*_ℓ) in the CMS frame. On average, the lepton from the $B \to \rho\ell\nu$ decay has a higher momentum than a lepton coming from a $B \to X_c \ell \nu$ decay due to the mass difference of the final state hadron. It was therefore necessary to select leptons with $p^*_\ell \geq 2.0$ GeV. To also maintain a constant q^2 efficiency, a two-point selection was introduced:

- $p^*_\ell \geq 2.0$ GeV *or*
- $p^*_\ell + p^*_\rho \geq 3.0$ GeV

The 2D distributions for p^*_ρ and p^*_ℓ in signal and background are shown in figure 5.7 for the $B^\pm \to \rho^0 e^\pm \nu_e$ channel, with the black line indicating the above-mentioned selection. The distributions for the other signal decay channels are shown in appendix section A.3. The signal efficiency is 90% while rejecting 60% of all remaining backgrounds.



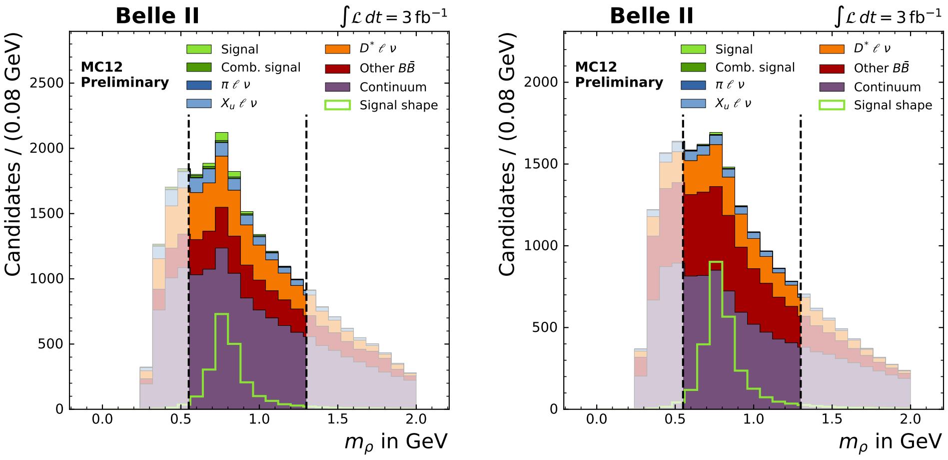
(a) Background

(b) Signal

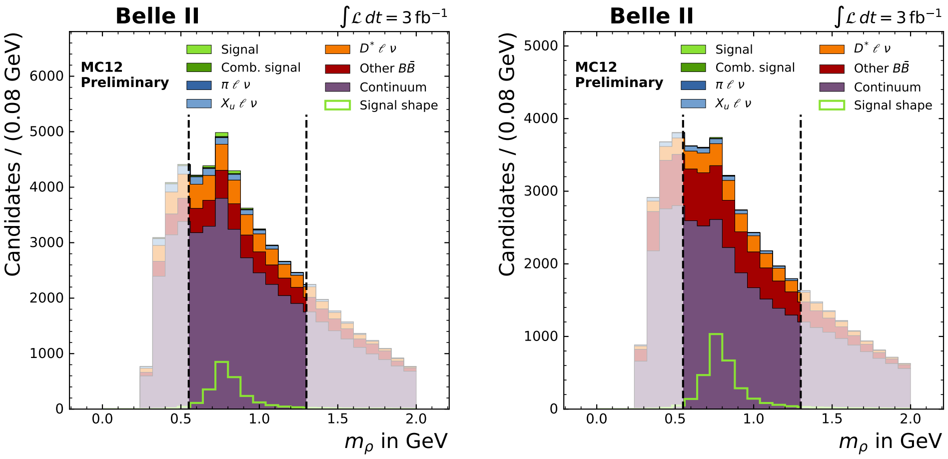
Figure 5.7: 2D distributions for p^*_ρ and p^*_ℓ for background (left) and signal (right) in $B^\pm \to \rho^0 e^\pm \nu_e$.

5.3.3 m_ρ

To further reduce the remaining background, a selection on the invariant mass of the reconstructed ρ was performed. The ρ is a relatively wide resonance, with $m_\rho = 775.26 \pm 0.25$ MeV and a decay width of $\Gamma = 149.1 \pm 0.8$MeV [16]. Consequentially, a tight selection on m_ρ is not possible due to a low signal efficiency. Taking this into consideration, the m_ρ was required to be between $0.55 \leq m_\rho \leq 1.30$ GeV, roughly 2.5× the natural decay width around the world-average mass, yielding a signal efficiency of 93% and background rejection of 44%. Distributions for m_ρ are shown in figure 5.8.



(a) $B^\pm \to \rho^0 e^\pm \nu_e$ (b) $B^0 \to \rho^\pm e^\mp \nu_e$

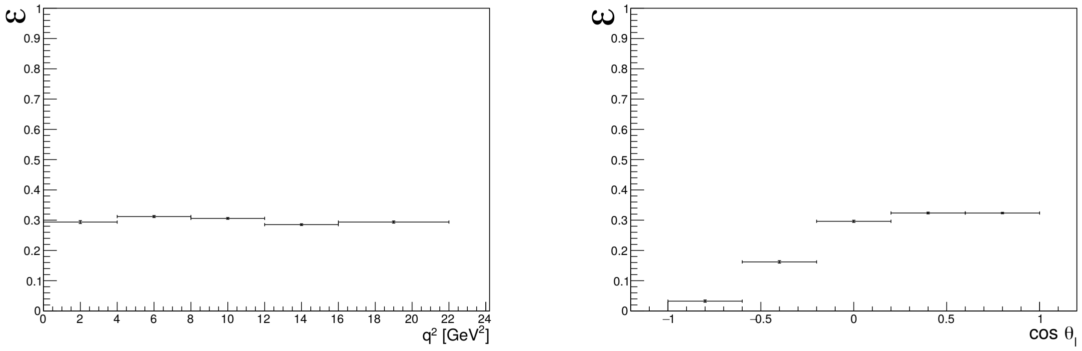


(c) $B^\pm \to \rho^0 \mu^\pm \nu_\mu$ (d) $B^0 \to \rho^\pm \mu^\mp \nu_\mu$

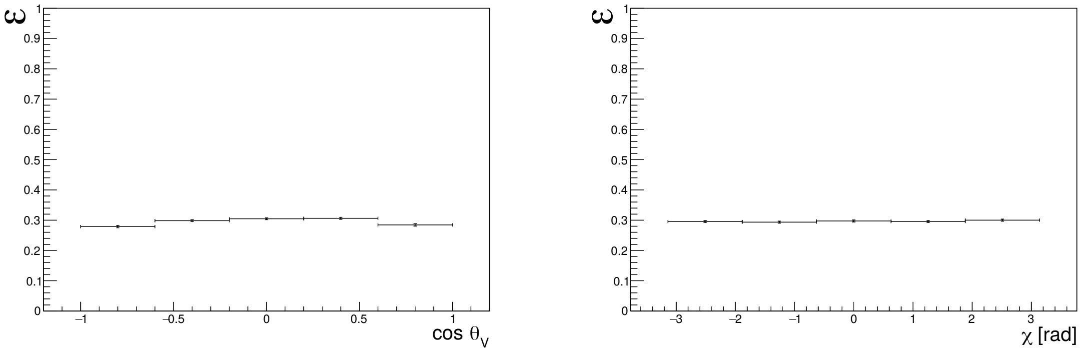
Figure 5.8: m_ρ distributions for the four signal decays and the selection $0.55 \leq m_\rho \leq 1.30$ GeV.

5.3.4 Efficiencies

After the pre-selections the signal selection efficiency for $B^{\pm}$ was 30% and for B^{0} 20%. The efficiency distributions for the four form factor parameters; q^2, $\cos\theta_l$, $\cos\theta_V$ and χ, are shown in figure 5.9. The same distributions for B^{0} can be found in appendix A.4. The four distributions are relatively constant, except for the first bin in $\cos\theta_l$ which has a lower efficiency. This is caused by the $p_l^* > 1.0$ GeV selection and the two-point momentum selection.



(a) q^2 efficiency. (b) $\cos\theta_l$ efficiency.



(c) $\cos\theta_V$ efficiency. (d) χ efficiency.

Figure 5.9: Signal efficiency distributions of q^2 and the three decay angles $\cos\theta_l$, $\cos\theta_V$ and χ, for $B^{\pm}$ after pre-selections.

5.4 Best candidate selection

After the application of all the above-mentioned selections, the average candidate multiplicity was 1.21. Out of these candidates, the best candidate was selected based on the sum of the PIDs of the reconstructed lepton and the pion(s). This had a signal efficiency of 99%.

5.5 Continuum and $\mathbf{B\overline{B}}$ background suppression

After the best candidate selection, the remaining background was dominated by continuum events in the low q^2 region, and by $B\overline{B}$ events in the high q^2 region, as can be seen in the ΔE and M_{bc} distributions shown in figure 5.10. To further reject this remaining background, two BDT types were trained. The first BDT type was trained to suppress background coming from continuum events, while the second type was trained to reject background events coming from $B\overline{B}$ events. To obtain the maximum efficiency in q^2, BDTs are trained in five q^2 bins with the ranges:

$q1$: $q^2 \leq 4.0\ \mathrm{GeV}^2$

$q2$: $4.0 < q^2 \leq 8.0\ \mathrm{GeV}^2$

$q3$: $8.0 < q^2 \leq 12.0\ \mathrm{GeV}^2$

$q4$: $12.0 < q^2 \leq 16.0\ \mathrm{GeV}^2$

$q5$: $q^2 > 16\ \mathrm{GeV}^2$

The BDTs were trained individually for the electron and the muon decay channels to maximize their performance. The BDTs were trained with a 50/50 ratio of truth-matched signal MC events and background events. Training was performed with the basf2 MVA package using the FastBDT library [23]. The BDTs consist of 150 trees, 3 layers, and a shrinkage of 0.2 was used during training.

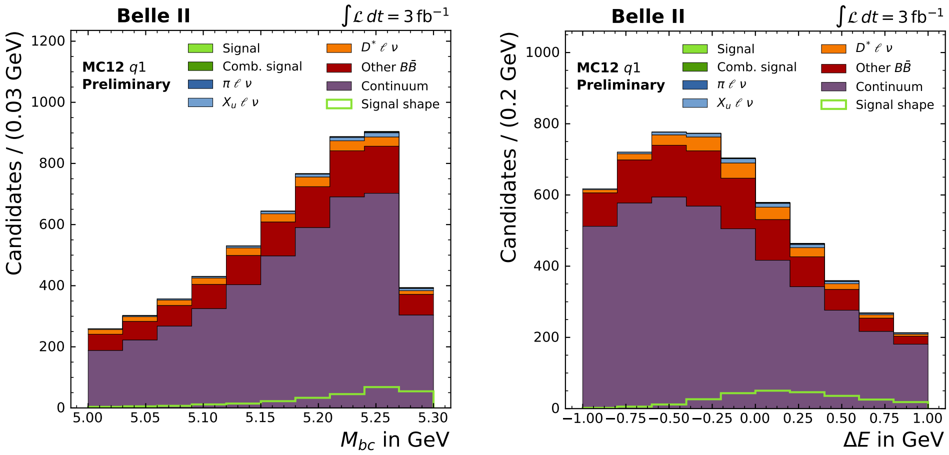
5.5.1 Continuum suppression

To reject background from continuum processes, BDTs were trained using Kakuno Super Fox-Wolfram (KFSW) moments. The KFSW moments describe the shape of an event, indicating how ‘jet-like’ or how spherical the event is, by parameterising the event shape in terms of spherical harmonics [30]. They are defined as

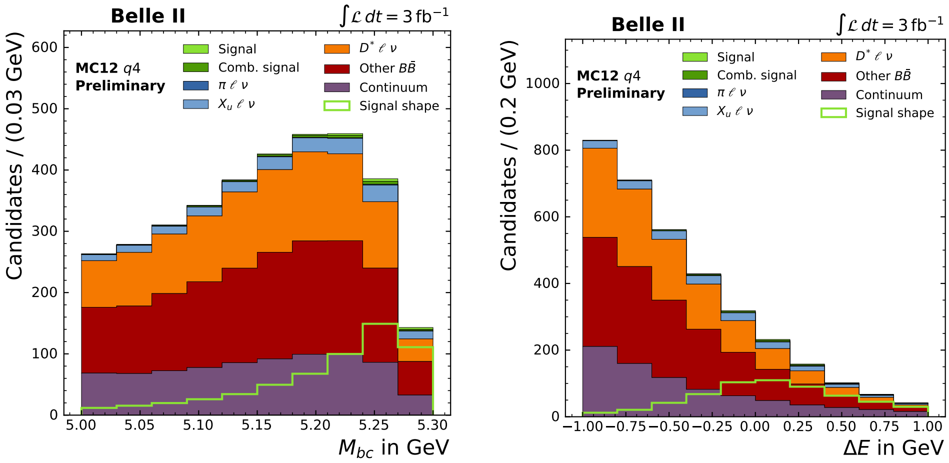
$$h_l^k = \frac{\sum_{m,n} |\vec{p}_m||\vec{p}_n| P_l(cos\theta_{mn})}{\sum_{m,n} |\vec{p}_m||\vec{p}_n|} \text{ and } R_l = \frac{h_l}{h_0}, \tag{5.4}$$

where $\vec{p}_m$ and $\vec{p}_n$ are the momenta of the particles m and n, $P_l(cos\theta_{mn})$ the l-th order Legendre polynomial of the cosine of the angle between the momenta of particles m and n. k is the category of the KSFW type: $k = so$ when m comes from the signal B and n from the ROE, and $k = oo$ when both m and n come from the ROE. For more detailed information about KSFW moments, see *The Physics of the B Factories* [31]. CLEO cones (named after the CLEO collaboration) were also used in the BDT [32]. CLEO cones are a set of cones (in steps of 10°) around the thrust axis of the reconstructed B candidate in which the scalar momentum flow is measured. Lastly, a collection of thrust axis variables from both the reconstructed B candidate as well as the rest of event (ROE) were used as input variables. $\cos\theta^{z}_{T_B}$ is the angle between the thrust axis of the B and the z-axis, $\cos\theta^{ROE}_{T_B}$ the angle between the thrust axis of the B and the ROE thrust axis, and $|\vec{T}_B|$ and $|\vec{T}_{ROE}|$ are the magnitude of the thrust vectors of the B and ROE respectively.

The exact variables used in the $B^{\pm} \to \rho^0 e^{\pm} \nu_e$ BDT in the $q1$ bin training, sorted by importance, are shown on the next page. For the variable ranking for all other BDTs, see appendix section A.7.2.



(a) M_{bc} in the $q1$ bin (b) ΔE in the $q1$ bin

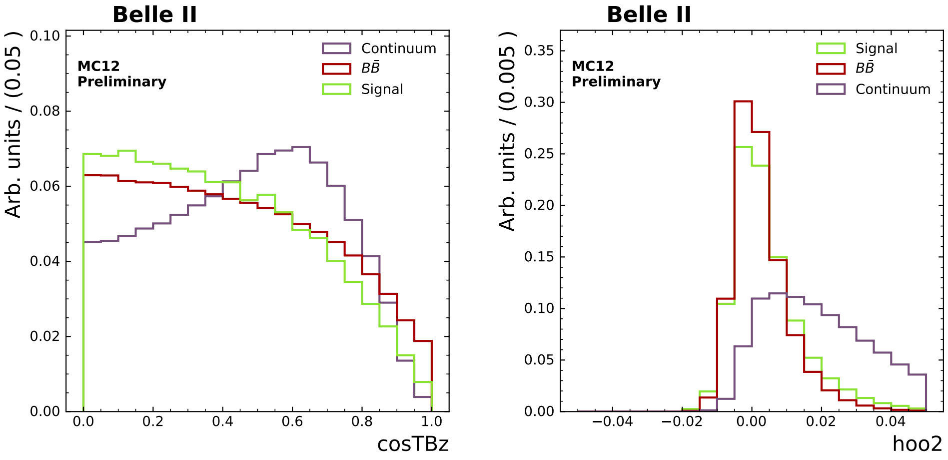


(c) M_{bc} in the $q4$ bin (d) ΔE in the $q4$ bin

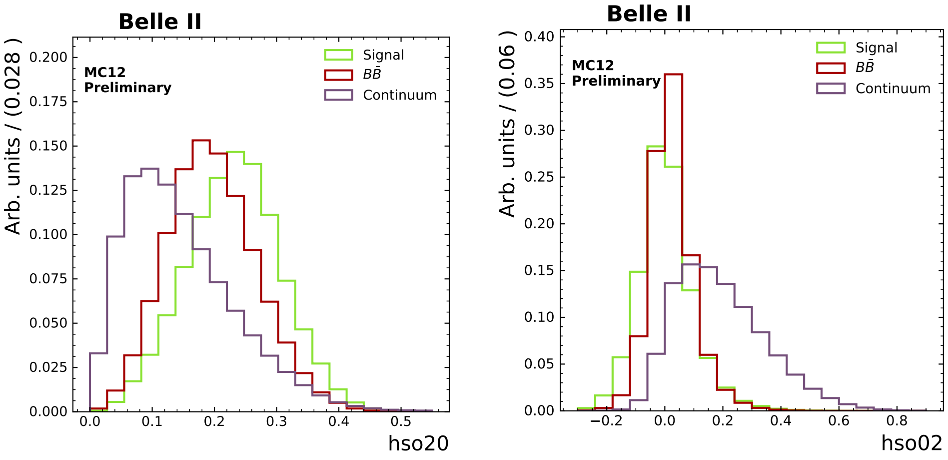
Figure 5.10: Distributions for M_{bc} and ΔE for the $B^0 \to \rho^\pm e^\mp \nu_e$ in the $q1$ bin (top), where the background is dominated by continuum events, and the $q4$ bin (bottom) where $B\bar{B}$ background dominates.

1. $\cos\theta^{z}_{T_B}$
2. h_2^{oo}
3. h_{20}^{so}
4. h_{02}^{so}
5. h_{22}^{so}
6. h_{12}^{so}
7. $cc1$
8. $|\vec{T}_B|$
9. $\cos\theta^{ROE}_{T_B}$
10. h_0^{oo}
11. $cc2$
12. $cc3$
13. R_2
14. $|\vec{T}_{ROE}|$

Normalized distributions for signal, $B\overline{B}$, and continuum events for the four highest ranking variables are shown in figure 5.11. The normalized distributions for the other variables in all decay channels are shown in appendix section A.7.1. A clear separation between the signal and continuum shapes is visible. The BDT also uses the correlations between the input variables, which are not easily visible by eye.



(a) $\cos\theta^{z}_{T_B}$ (b) hoo2



(c) hso20 (d) hso02

Figure 5.11: Normalized distributions for the four highest ranking variables in the $B^{\pm} \to \rho^{0} e^{\pm} \nu_e$ $q1$ bin continuum suppression BDT. From top left to bottom right: $\cos\theta^{z}_{T_B}$, h_2^{oo}, h_{20}^{so}, and h_{02}^{so}.

These correlations are shown in the correlation matrices for signal and background in figure 5.12. BDT performance on the training and testing data is shown in figure 5.13, together with the overtraining check. The p-test for the performance difference of the BDT between training and testing data shows there is no significance difference in performance, so overtraining was not present. The correlation matrices and overtraining checks for the other signal decay channels and q^2 bins are shown in appendix sections A.7.3 and A.7.4. Looking at the correlation matrix in figure 5.12, it can be seen that there are correlations between some higher-order KSFW variables, but non of the correlations are high enough to require removal from the BDT training.

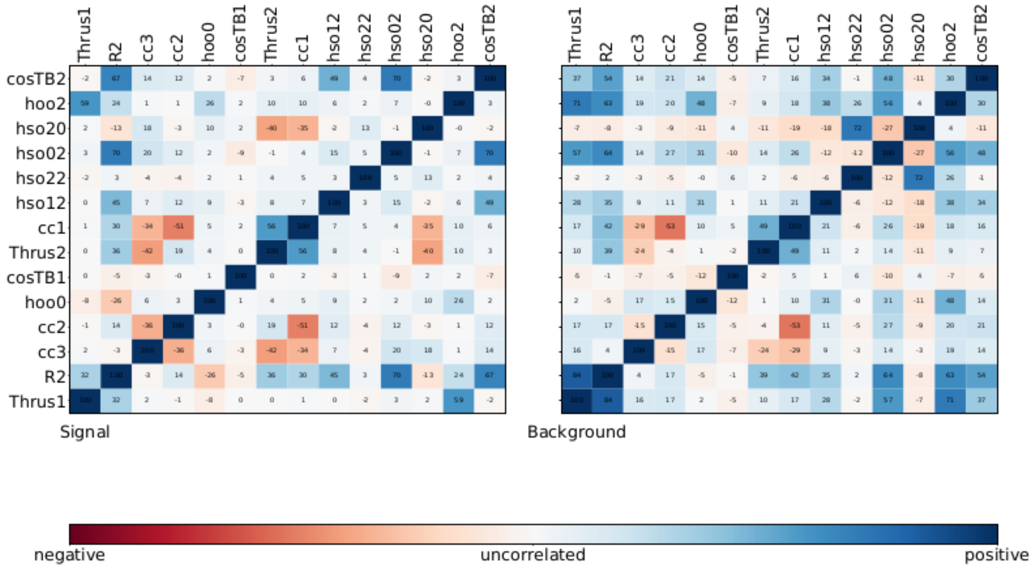


Figure 5.12: Correlation matrices for the $B^{\pm} \to \rho^0 e^{\pm} \nu_e$ continuum suppression BDT in the $q^2 \leq 4.0$ GeV^2 bin.

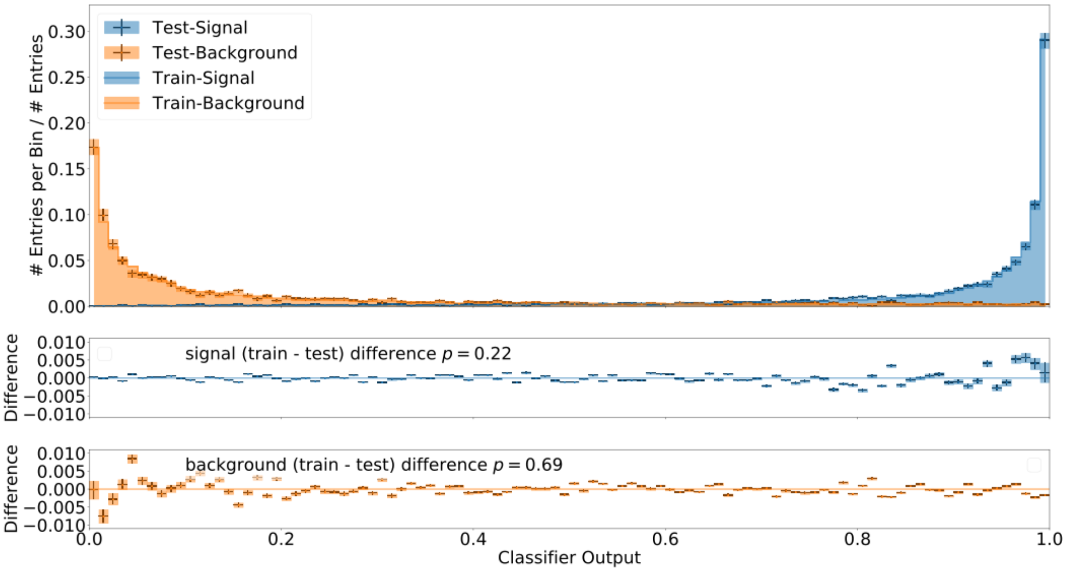
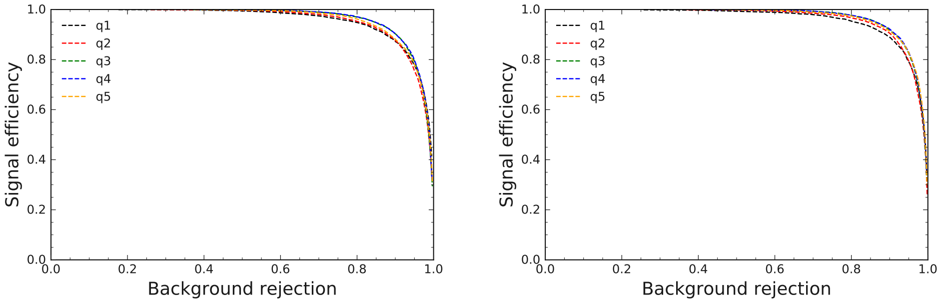


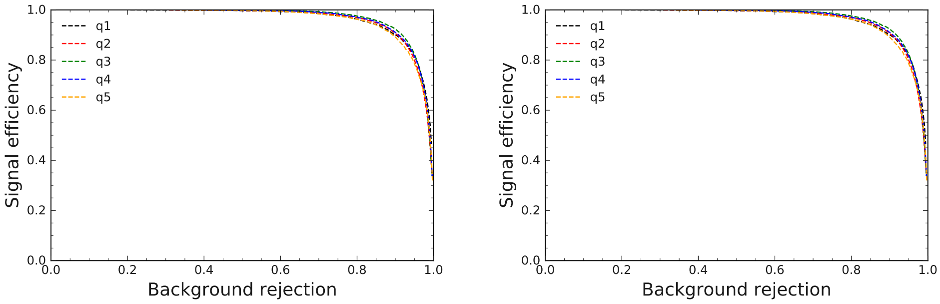
Figure 5.13: Overtraining check for the $B^{\pm} \to \rho^0 e^{\pm} \nu_e$ continuum suppression BDT in the $q^2 \leq 4.0$ GeV^2 bin. The p-value shows no significant difference between training and testing data, so overtraining is not present.

The receiver operating characteristic (ROC) curves for the trained BDTs are shown in figure 5.14. A ROC curve shown how well the BDT is performing in retaining signal and rejecting background events. Ideally, the BDT should reject all background events and only retain signal events, leading to a curve integral of 1. Practically this is not possible, as can be seen in the obtained ROC curves. All curves show a similar shape and have an integral of around 0.94, indicating a comparable performance in all q^2 bins and the four signal decay channels. The integral of the ROC curves is close to 1, indicating that the BDTs peform well in the separation of continuum and signal events.



(a) $B^{\pm} \rightarrow \rho^{0} e^{\pm} \nu_e$

(b) $B^{0} \rightarrow \rho^{\pm} e^{\mp} \nu_e$



(c) $B^{\pm} \rightarrow \rho^{0} \mu^{\pm} \nu_{\mu}$

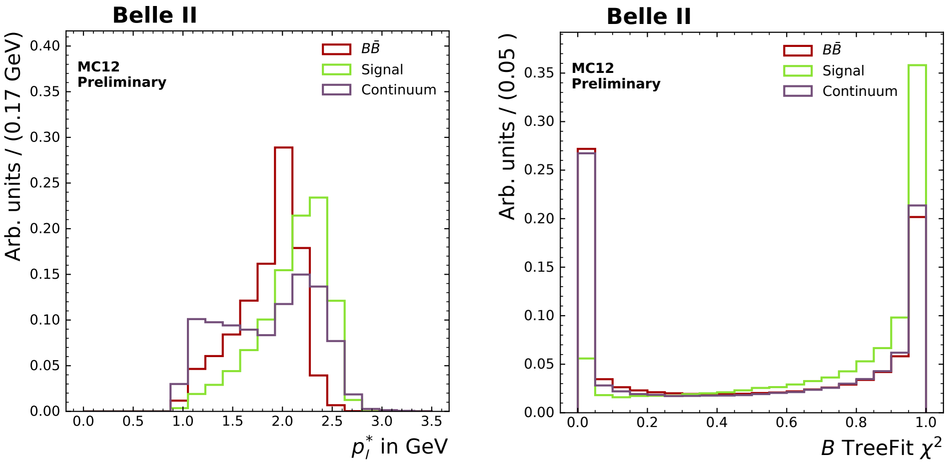
(d) $B^{0} \rightarrow \rho^{\pm} \mu^{\mp} \nu_{\mu}$

Figure 5.14: ROC curves for the trained BDTs in the five q^2 bins for all four decay channels. The curves all show a similar shape, indicating comparable BDT performance.

5.5.2 B$\overline{\text{B}}$ suppression

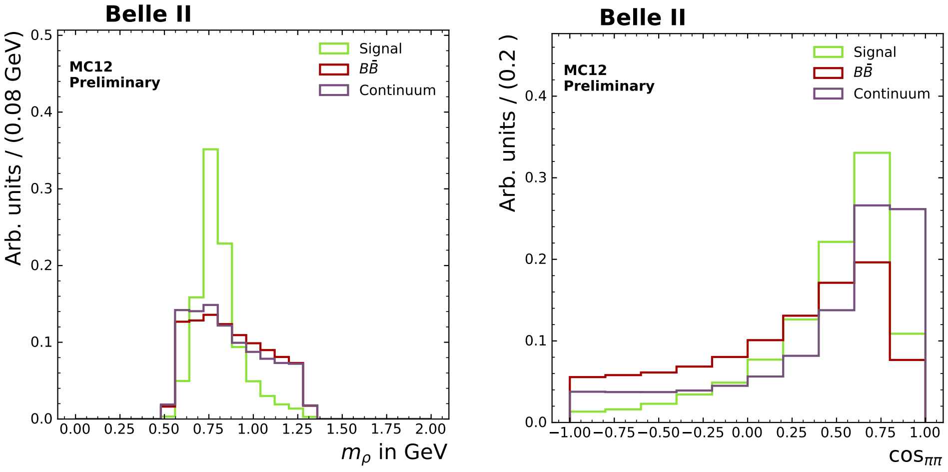
Now that continuum suppression is implemented, $B\overline{B}$ are the dominant background. To reject this background, $B\overline{B}$ suppression BDTs were trained using the same settings as the continuum suppression BDTs, but on a different set of parameters. The variables used in the BDTs include a range of kinematic variables, as well as the beam-constrained mass and ΔE for the ROE, m_{ρ}, the angle between the two reconstructed π in the ρ CMS frame $\theta_{\pi\pi}$, the angle between the B and the ℓ in the W CMS frame θ_l, the χ^2 probability of the B vertex fit, and a D^* veto. For more information on the D^* veto, see appendix section A.6. The variable ranking for $B^{\pm} \rightarrow \rho^{0} e^{\pm} \nu_e$ BDT in the $q1$ bin is shown below. Normalized distributions for the four highest ranking variables are shown in figure 5.15, all other distributions can be found in appendix section A.7.1. Once again, the distributions show that the signal shape is clearly separated from the targeted $B\overline{B}$ background shape, especially in the peaking B vertex fit and the m_{ρ} distributions. For the variable rankings of all the other trained BDTs, see appendix section A.7.2.

1. p_l^*
2. B vertex χ^2 prob.
3. m_ρ
4. M_{bc}^{ROE}
5. $\cos\theta_{\pi\pi}$
6. p_{miss}^{lab}
7. ΔE^{ROE}
8. $\cos\theta_{BY}$
9. p_ρ^*
10. θ_{miss}
11. $\cos\theta_l$
12. D^* veto



(a) p_l^*

(b) B vertex fit χ^2 prob.



(c) m_ρ

(d) $\cos\theta_{\pi\pi}$

Figure 5.15: Normalized distributions for the four highest ranking variables in the $B^\pm \to \rho^0 e^\pm \nu_e$ $q1$ bin $B\overline{B}$ suppression BDT. From top left to bottom right: p_l^*, B vertex fit χ^2 probability, m_ρ, and $\cos\theta_{\pi\pi}$.

The correlation matrices for signal and background are shown in figure 5.16. The performance on the training and testing data is shown in figure 5.17, together with the overtraining check. Similarly to the continuum suppression BDTs, no overtraining is present. The correlation matrices and overtraining checks for the other signal decay channels and q^2 bins are shown in appendix sections A.7.3 and A.7.4. Looking at the correlation matrix in figure 5.16, it can be seen that the input variables are relatively uncorrelated, with the exception of p_ℓ^* and $\cos\theta_\ell$, p_ℓ^* and p_{miss}^*, and m_ρ and $\cos\theta_{\pi\pi}$. This is not unexpected, since these are all kinematic quantities describing the decay, and are therefore correlated due to the momentum and energy conservation in the decay.

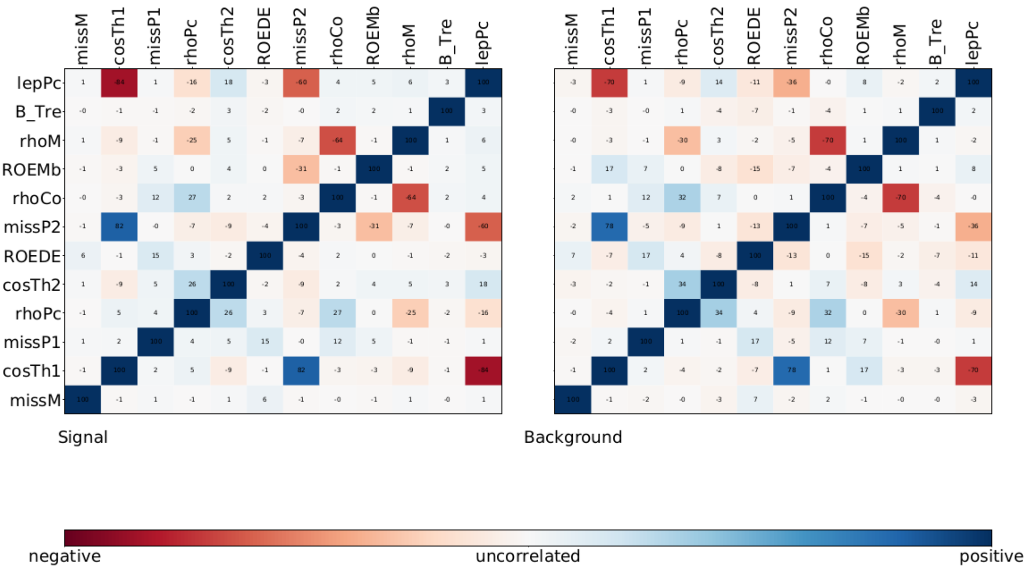


Figure 5.16: Correlation matrices for the $B^0 \to \rho^\pm e^\mp \nu_e$ $B\overline{B}$ suppression BDT in the $q^2 \leq 4.0\ \text{GeV}^2$ bin.

The ROC curves for the trained $B\overline{B}$ BDTs in all decay channels and q^2 bins are shown in figure 5.18. Compared to the continuum suppression ROC curves, the $B\overline{B}$ BDTs perform slightly worse. A possible explanation for this is the more diverse set of background events used in the training, compared to only one type of background in the continuum BDTs. Another potential explanation is that some background types (e.g. $B \to X_u \ell \nu$) have similar event kinematics, making them hard to distinguish from signal events. Overall, the BDTs have a worse performance in the low q^2 region, most likely due to the lower training statistics.

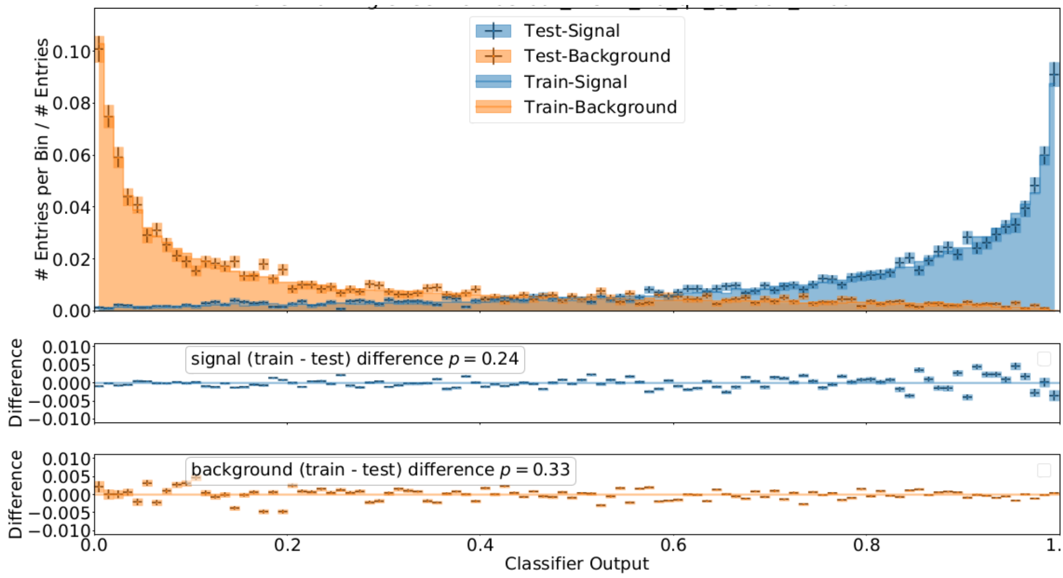
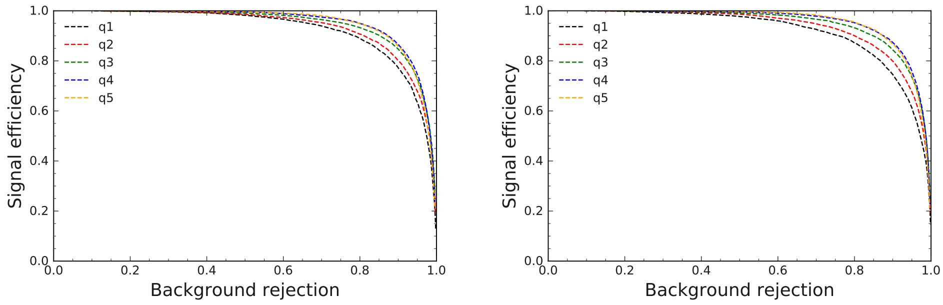
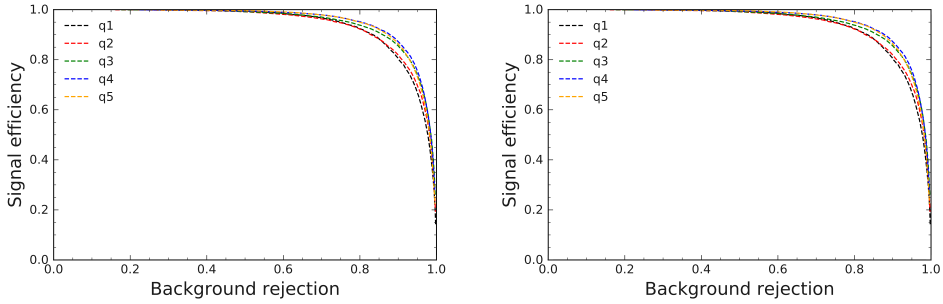


Figure 5.17: Overtraining check for the $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ $B\overline{B}$ suppression BDT in the $q^2 \leq 4.0\,\mathrm{GeV}^2$ bin. The p-test shows no significant difference between training and testing performance, so overtraining is not present.



(a) $B^{\pm} \to \rho^0 e^{\pm} \nu_e$

(b) $B^0 \to \rho^{\pm} e^{\mp} \nu_e$



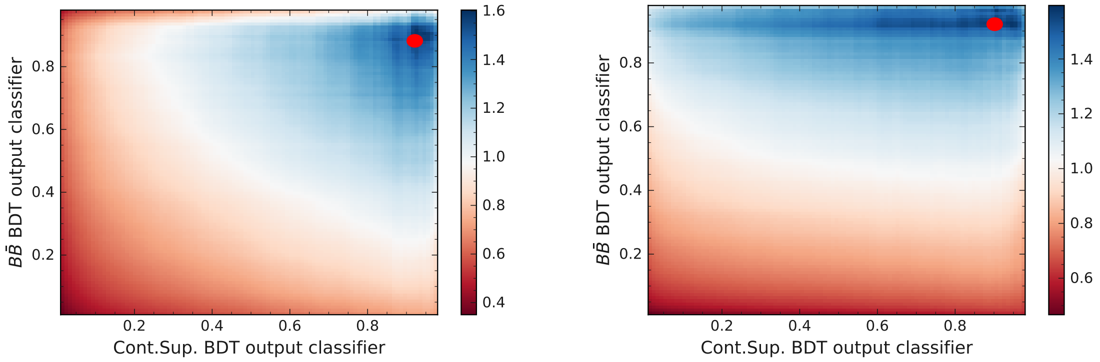
(c) $B^{\pm} \to \rho^0 \mu^{\pm} \nu_\mu$

(d) $B^0 \to \rho^{\pm} \mu^{\mp} \nu_\mu$

Figure 5.18: ROC curves for the trained $B\overline{B}$ BDTs in the five q^2 bins for all four decay channels.

5.5.3 Figure of Merit optimization

To obtain the best signal efficiency and background rejection, a 2D-FOM optimization was performed in every q^2 bin by simultaneously selecting on the continuum suppression BDT and the $B\overline{B}$ BDT. The FOM was calculated as $N_{\text{signal}}/\sqrt{N_{\text{signal}} + N_{\text{BG}}}$ in the region $5.21 \leq M_{bc} < 5.30$ GeV, where signal is expected to peak. Two example FOM distributions are shown in figure 5.19 for the $B^0 \to \rho^\pm e^\mp \nu_e$ $q1$ and $q4$ bins, with the red dot indicating the selections that maximize the FOM. All other FOM distributions are shown in appendix section A.7.5.



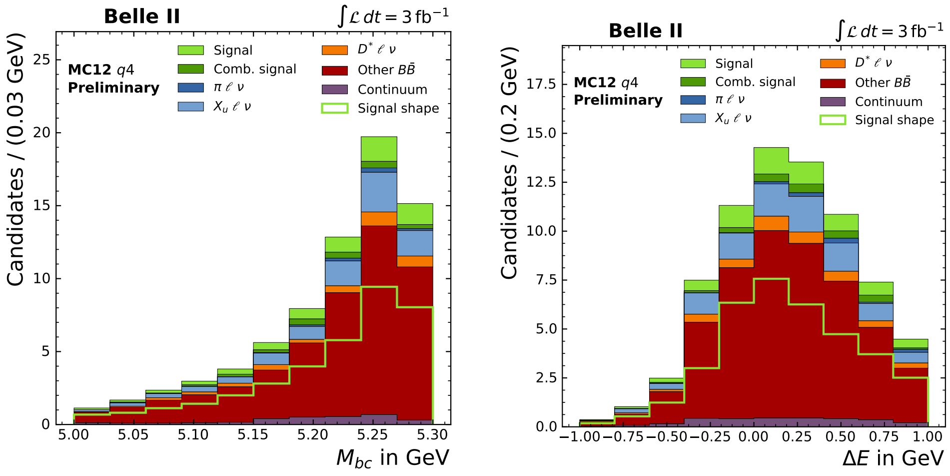
(a) $q^2 \leq 4.0$ GeV^2

(b) $12.0 < q^2 \leq 16.0$ GeV^2

Figure 5.19: 2D FOM distributions for the BDT selection optimization for the signal decay channel $B^0 \to \rho^\pm e^\mp \nu_e$.

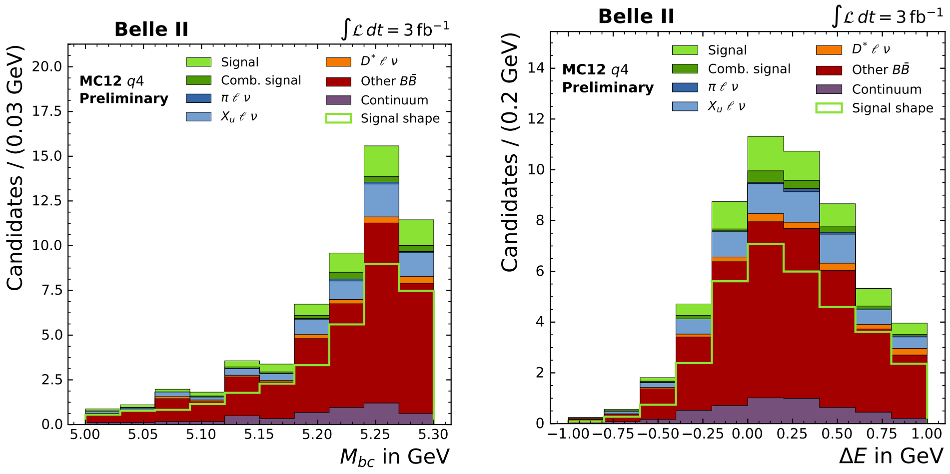
In all signal decay channels and q^2 bins, the FOM optimization leads to very tight selections on both BDT output classifiers, in the range from 0.88 to 0.98. This is caused by the fact that before applying any BDT selections, there is still a factor 300 times more background events than signal events. Because of this, the FOM is higher at very tight selections, since the removal of background events has a higher effect than the retention of signal events. This skews the FOM to selecting at high BDT output classifier values. Difference can be seen between the q^2 bins, in the low q^2 region tighter cuts on the continuum suppression BDT output classifier are preferred. This can be explained by most of the continuum events falling in the low q^2 region, so a tighter selection leads to a higher background rejection compared to the high q^2 region. The FOM BDT selection has an average efficiency of 30% with a background rejection of 99%.

Distributions of M_{bc} and ΔE for the B^0 $q4$ bin after applying the FOM BDT selections are shown in figure 5.20. Looking at the distributions, it can be seen that the shape of the background events is almost identical to the shape of the signal events. This is most likely caused by the tight selection on the BDT output classifiers, due to the large number of background events before applying any selection on the BDT output classifiers, leading to only ‘signal like’ events passing the selection and forcing the backgrounds to look like the signal shape. Another possible explanation is that not all the input variables in the BDTs are uncorrelated to M_{bc} and ΔE, which after a tight selection on the BDT output classifiers could make the distributions correlated as well. The BDT classifiers themselves all have a correlation of around 0.20-0.25 to M_{bc} and ΔE.



(a) $B^0 \to \rho^\pm e^\mp \nu_e$ M_{bc}

(b) $B^0 \to \rho^\pm e^\mp \nu_e$ ΔE



(c) $B^0 \to \rho^\pm \mu^\mp \nu_\mu$ M_{bc}

(d) $B^0 \to \rho^\pm \mu^\mp \nu_\mu$ ΔE

Figure 5.20: Distributions for M_{bc} and ΔE for the $B^0 \to \rho^\pm e^\mp \nu_e$ (top) and $B^0 \to \rho^\pm \mu^\mp \nu_\mu$ (bottom) in the bin $12.0 < q^2 \leq 16.0$ GeV^2. The shape of the signal events compared to the background events is almost identical.

Due to this, performing a simultaneous fit on M_{bc} and ΔE is not possible, since the signal and background yields would be almost maximally anti-correlated, and therefore a different approach was required. Instead of optimizing the selection on both BDT output classifiers, both BDTs were required to exceed 0.8, and a simultaneous fit on both BDT output classifiers was used. This will be more extensively discussed in chapter 6.

Table 5.2 shows the signal efficiencies for all the performed selections in the $q^2 < 4.0$ GeV^2 bin. The total signal efficiency for $B^\pm$ is around 15%, while for B^0 it is 8%. The efficiency tables for the other q^2 bins are shown in appendix section A.1. The difference between the $B^\pm$ and the B^0 efficiency stems mostly from the π^0 which is present in the B^0 decay. The π^0 reconstruction efficiency is worse than the reconstruction efficiency for the $\pi^\pm$, mostly due to the amount of beam background photons which lead to wrong π^0 reconstructions.

Selection	$\epsilon^{e}_{B^0}$	$\epsilon^{\mu}_{B^0}$	$\epsilon^{e}_{B^\pm}$	$\epsilon^{\mu}_{B^\pm}$
Acceptance and pre-selections	0.294 ± 0.004	0.271 ± 0.004	0.485 ± 0.002	0.478 ± 0.002
$\theta_{\gamma\gamma}$ < 1.20 rad	0.966 ± 0.019	0.969 ± 0.019	-	-
nCleanedTracks > 4	0.905 ± 0.019	0.907 ± 0.018	0.981 ± 0.022	0.982 ± 0.022
$\cos\theta_{miss}$ in acceptance	0.920 ± 0.020	0.923 ± 0.020	0.923 ± 0.021	0.908 ± 0.021
M_{bc} and ΔE selection	0.846 ± 0.020	0.841 ± 0.019	0.819 ± 0.020	0.795 ± 0.020
ρ vertex fit > 10^{-10}	-	-	0.973 ± 0.026	0.972 ± 0.026
$\lvert\cos\theta_{BY}\rvert < 1.1$	0.936 ± 0.023	0.970 ± 0.023	0.949 ± 0.025	0.982 ± 0.027
Combined p^*_l/p^*_ρ	0.994 ± 0.025	0.971 ± 0.023	0.994 ± 0.026	0.993 ± 0.027
$0.55 < m_\rho < 1.30$ GeV	0.865 ± 0.022	0.860 ± 0.022	0.874 ± 0.024	0.877 ± 0.025
Best candidate selection	0.998 ± 0.022	0.999 ± 0.022	0.999 ± 0.029	0.999 ± 0.029
BDT classifiers selection	0.477 ± 0.016	0.522 ± 0.017	0.507 ± 0.018	0.556 ± 0.019
Total	0.076 ± 0.002	0.087 ± 0.002	0.146 ± 0.002	0.159 ± 0.001

Table 5.2: Efficiencies per performed selection for all 4 decay channels in the $q1$ bin.

CHAPTER 6

Signal extraction results and discussion

This chapter will describe the signal extraction fit used to obtain the signal yield from 6.5 fb^{-1} MC, corresponding to the total collected early phase 3 Belle II data, and discuss the results. Since the data was not fully understood at the moment of writing this thesis, signal extraction on data was not possible. Firstly, the fit method and variables used for signal extraction will be introduced. Secondly, the obtained results will be shown and discussed. Thirdly, possible systematic uncertainty sources are mentioned. Finally, possible improvements to the performed analysis will be mentioned. The fits were only performed for the electron signal decay channels, since in early phase 3 data the μ PID was not working.

6.1 Used fitting method

The signal extraction fit was performed using the RooFit package [33]. A simultaneous maximum likelihood fit of the obtained post-selection MC templates of the continuum suppression and the $B\overline{B}$ BDT output classifiers was used on the MC sample corresponding to 6.5 fb^{-1}. The fit region was selected as continuum suppression BDT output classifier > 0.8 and the $B\overline{B}$ BDT output classifier > 0.8. The fit was performed separately in every q^2 bin, and the obtained signal yields were corrected using the measured signal efficiencies from MC. Two categories were fitted:

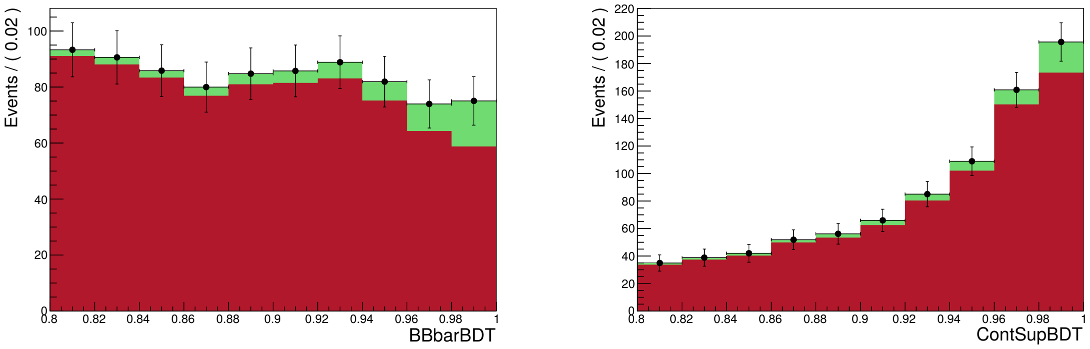
1. Signal and combinatorial signal
2. All other remaining background

Due to limited statistics and the similarity in shapes of the different background types, only two fit categories could be used. To obtain a maximized signal signal yield, signal and combinatorial signal were fitted together, since both categories scale according to the branching fraction of the signal decays, and therefore both contribute to the signal significance. Because of the limited statistics in the MC fit sample after applying all signal selections, the fit yields for both the signal and background categories were only allowed to be positive. From the obtained yields, the branching fraction was calculated using

$$\mathcal{B}(B \to \rho \ell \nu) = \frac{\sum_{i=1}^{5} \frac{N_{\text{sig}}^{qi}}{\epsilon_{\text{sig}}^{qi}}}{2N_{B\overline{B}}} \tag{6.1}$$

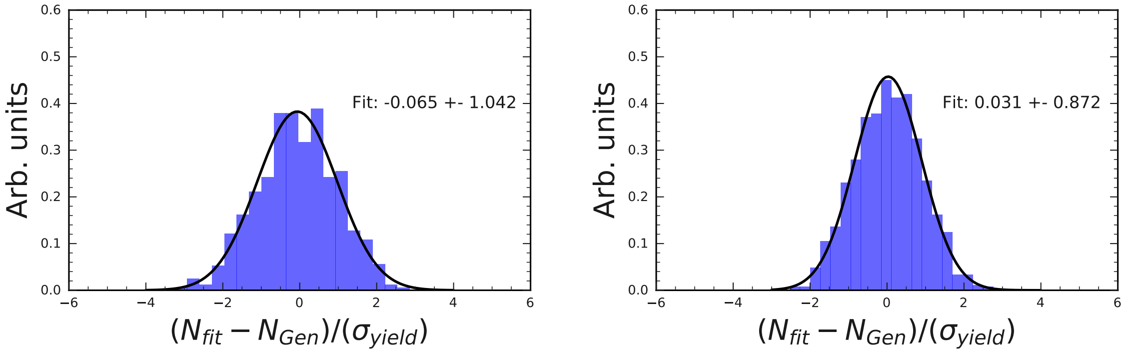
where N_{sig}^{qi} is the signal yield obtained from the fit in the q^2 bin qi, $\epsilon_{\mathrm{sig}}^{qi}$ the total signal selection efficiency in the i-th q^2 bin, and $N_{B\overline{B}}$ is the amount of produced $B\overline{B}$ pairs in the used MC sample. $N_{B\overline{B}}$ was taken from the MC generator information, and for 6.5 fb^{-1} corresponded to $N_{B^0\overline{B}^0} = 2,169,375$ and $N_{B^+B^-} = 2,299,375$ [26]. The factor 2 comes from the fact that both B mesons can decay into the signal mode. From this, $|V_{ub}|$ can be calculated using equation 2.5, this was however not done in this thesis due to the limited statistics in the individual q^2 bins.

The fit was verified by performing an Asimov fit, fitting the templates to themselves, and by generating 1000 toy MC sets from the templates and calculating the pull as $(N_{\mathrm{Fit}} - N_{\mathrm{Gen}})/\sigma_{\mathrm{Yield}}$. For a correctly working fit, the Asimov fit should yield the initial input values from the templates. The pull distribution should have a mean of 0.0 and a standard deviation of 1.0. An example Asimov fit and the obtained pull distributions for the signal and background yields in the $B^0 \longrightarrow \rho^{\pm} e^{\mp} \nu_e$ $q2$ bin are shown in figure 6.1. Both fitted BDT output classifier distributions exactly follow the data, indicating the fit works correctly. The signal pull has a mean of -0.065 and a standard deviation of 1.042, meaning that the fit gives a stable signal yield. The background yield also has a mean close to 0, but the standard deviation is 0.872, indicating that the uncertainty on the background yield is overestimated. This is potentially caused by the fact that all background components are fitted together, instead of fitting the individual background categories separately, due to possible (anti)-correlations between the different background types. All other Asimov fits and pull distributions are shown in appendix sections A.8.1 and A.8.2.



(a) Fitted $B\overline{B}$ BDT suppression output classifier.

(b) Fitted continuum suppression BDT output classifier.



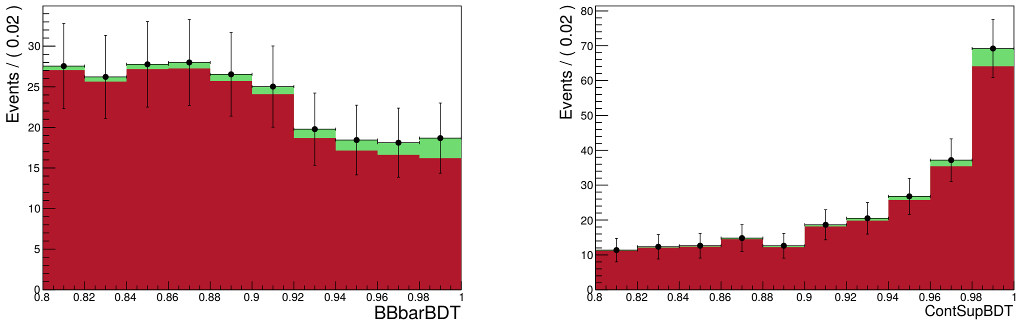
(c) Pull distribution for the signal yield (n = 1000).

(d) Pull distribution for the background yield (n = 1000).

Figure 6.1: Top: Asimov fit for the $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ $q2$ decay, scaled to 16 fb^{-1}. The green indicates the fitted signal, while the red is the fitted background shape. Bottom: Pull distributions and their fitted normal Gaussians for the $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ $q2$ fit.

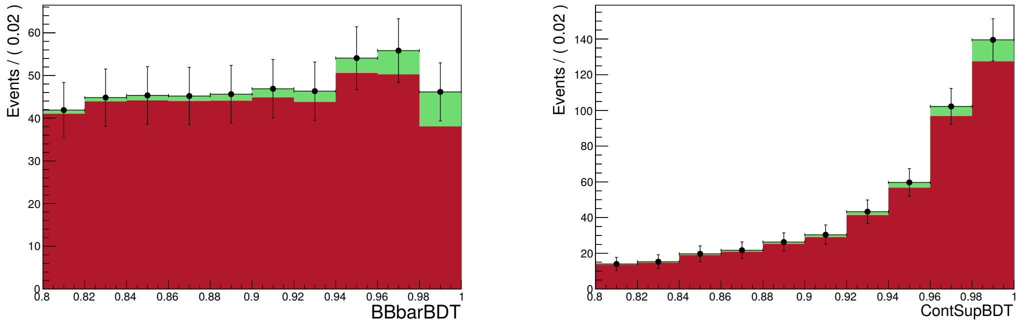
6.2 Fit results and discussion

The 6.5 fb^{-1} MC fitted signal and background yields are shown in table 6.1. The fitted distributions for $B^0 \longrightarrow \rho^{\pm} e^{\mp} \nu_e$ in the $q1$ and $q4$ bins are shown in figure 6.2, where green is the fitted signal and the red the fitted background shape. The other fitted distributions are shown in appendix section A.8.3.



(a) Fitted $B\overline{B}$ BDT suppression output classifier in the $q1$ bin.

(b) Fitted continuum suppression BDT output classifier in the $q1$ bin.



(c) Fitted $B\overline{B}$ BDT suppression output classifier in the $q4$ bin.

(d) Fitted continuum suppression BDT output classifier in the $q4$ bin.

Figure 6.2: Fit results for the $B^0 \longrightarrow \rho^{\pm} e^{\mp} \nu_e$ signal decay channel in the $q1$ (top) and $q4$ (bottom) bins.

Looking at the results in table 6.1, it can be seen that only some of the q^2 bins in the medium q^2 region have a significant signal yield. This matches with the predicted q^2 distributions in figure 2.4(a) from the LCSR and ISGW2 calculations, since most signal events are predicted to have a q^2 in this region. The signal selection efficiency in the higher q^2 region is also higher than in the low q^2 region, leading to more events passing the selections and a higher statistics in the fit. The calculated branching fractions from the signal yields in table 6.1 are shown in table 6.2.

Comparing the obtained branching fractions to the world-average PDG values mentioned in section 2.4, it can be seen that the values agree within 1σ. This is expected, since the fit was performed on MC that uses the PDG branching fractions in the event generation. The results are, however, also in agreement with 0, due to the large statistical uncertainty. This uncertainty can be explained by the low statistics of the MC distributions used in the fit, corresponding to the size of the collected Belle II early phase 3 data of 6.5 fb^{-1}. A bigger data sample is required to obtain a significant measurement of

$B^{\pm} \to \rho^0 e^{\pm} \nu_e$	$q1$	$q2$	$q3$	$q4$	$q5$
Signal	12 ± 23	26 ± 25	35 ± 28	33 ± 27	17 ± 25
Background	310 ± 26	430 ± 29	686 ± 33	556 ± 32	440 ± 29

$B^0 \to \rho^{\pm} e^{\mp} \nu_e$	$q1$	$q2$	$q3$	$q4$	$q5$
Signal	11 ± 17	23 ± 19	33 ± 23	27 ± 24	15 ± 22
Background	225 ± 20	318 ± 23	523 ± 28	445 ± 28	372 ± 25

Table 6.1: Obtained yields for the two fitted signal decay modes and q^2 bins, and their statistical uncertainties.

Decay channel	$\mathcal{B}$
$B^{\pm} \to \rho^0 e^{\pm} \nu_e$	$(2 \pm 1) \cdot 10^{-4}$
$B^0 \to \rho^{\pm} e^{\mp} \nu_e$	$(3 \pm 2) \cdot 10^{-4}$

Table 6.2: Obtained branching fractions for the e^- signal decay modes and their statistical uncertainty.

the branching fraction. To quantify the amount of data that would be required to obtain a significant result, MC toys were created from the obtained signal and background templates, and Asimov fits were done to obtain the signal yields with statistical uncertainties, shown in table 6.3. The fit distributions are shown in appendix section A.8.2. The calculated branching fractions from the Asimov fit yields are shown in table 6.4. The study showed that for 3σ statistical significance approximately 16 fb^{-1} is needed, assuming the MC correctly describes the data. This amount of data will most likely be reached, and hopefully exceeded, in the Belle II winter run period in 2019, making a statistically significant measurement possible.

$B^{\pm} \to \rho^0 e^{\pm} \nu_e$	$q1$	$q2$	$q3$	$q4$	$q5$
Signal	29 ± 36	63 ± 40	86 ± 43	80 ± 42	42 ± 38
Background	765 ± 41	1060 ± 45	1687 ± 53	1370 ± 43	1084 ± 45

$B^0 \to \rho^{\pm} e^{\mp} \nu_e$	$q1$	$q2$	$q3$	$q4$	$q5$
Signal	26 ± 26	57 ± 30	81 ± 36	67 ± 37	36 ± 33
Background	555 ± 31	1060 ± 45	1288 ± 44	1095 ± 33	915 ± 39

Table 6.3: Obtained Asimov yields for the e^- signal decay modes and q^2 bins using 16 fb^{-1} MC, and their statistical uncertainties.

Decay channel	$\mathcal{B}$
$B^{\pm} \to \rho^0 e^{\pm} \nu_e$	$(3 \pm 1) \cdot 10^{-4}$
$B^0 \to \rho^{\pm} e^{\mp} \nu_e$	$(1.6 \pm 0.5) \cdot 10^{-4}$

Table 6.4: Obtained branching fractions for the e^- signal decay modes and their statistical uncertainty using 16 fb^{-1} MC and the Asimov fit yields.

6.3 Possible systematic uncertainties

The measurements of the signal branching fractions not only have a statistical uncertainty, but also a systematic uncertainty. Due to time limitations, this could not be studied in the scope of this thesis, so only estimates based on comparable previous analyses are provided. One possible systematic uncertainty source is the type of form factors used in the MC. Since the form factor describes the signal decay kinematics, a different form factor calculation might yield e.g. a different q^2 spectrum. This would lead to different selection efficiencies and therefore a potentially different final result. Changing the form factors for the major background sources, e.g. $B \to D^* \ell \nu$, can lead to more or fewer background events passing the selections, and can therefore also influence the final result. To take this into account, form factor reweighting can be done. By assigning every event a new weight, such that the kinematic distributions match the new form factor, the effect can be quantified. In the $B \to \rho \ell \nu$ analysis by the BaBar collaboration, the systematic uncertainty introduced by the form factors was 9% at low q^2 and 6% at high q^2 [17]. Using newer, better understood form factor calculations could reduce the error, therefore the error is estimated to be of the order of 5%.

Another source of uncertainty is how well the used Monte Carlo simulation models the physics events and the detector response. A first look on Belle II data will be given in chapter 7, but a full study was not possible due to time limitations. Possible disagreements in detector response could come from mismodeling of the tracking, the performance of the PID, photon detection efficiency, and the B vertex fit. The Belle $B \to \pi \ell \nu$ quantified this effect as a 3.4% uncertainty [34]. The before-mentioned BaBar analysis found a disagreement between their continuum MC and data [17]. This was solved by assigning a weight to the continuum events, which introduced an extra uncertainty on the order of 6%. Similarly, the branching fractions for the $b \to u \ell \nu$ and $b \to c \ell \nu$ decays can be varied around their world averages by assigning a weight to these events and scaling them to the expected amount. The above-mentioned Belle analysis quantified this effect as 1.1% error, so a similar error is expected for this analysis [34].

Thirdly, a major expected systematic uncertainty is the modeling of the inclusive $B \to X_u \ell \nu$ MC. Since the higher X_u resonances are not well understood, these events are modeled with a so-called inclusive model, where the total decay rate is determined from Heavy Quark Expansion. The MC is characterized by a shape function and its parameters [35]. These parameters come with corresponding uncertainties, which can be used to vary the parameters and assessing the impact on the final analysis result. Past analyses show conflicting relative uncertainties, with almost an order of magnitude difference. The before-mentioned Belle analysis found an X_u modeling uncertainty of $\mathcal{O}(1\%)$ [34], while the BaBar $B \to \rho \ell \nu$ analysis had an uncertainty of $\mathcal{O}(10\%)$ [17]. For this analysis, a conservative estimate is made and the uncertainty is expected to be comparable to the BaBar analysis.

Finally, the use of BDTs adds an extra systematic uncertainty to the results. Since the BDTs are trained on MC, they could potentially perform worse on data due to the mismodeling in MC. Because of the limited available statistics, the performance of the trained BDTs on data could not be fully verified. A possible solution for this would be the training of similar BDTs for a decay with a higher branching fraction, e.g. $B \to D^* \ell \nu$. If these BDTs show a good data-MC agreement, it can be assumed that the BDTs for the signal decay channels also agree within an uncertainty. This was, however, outside the scope of this thesis, and therefore an estimate of the relative uncertainty cannot be given. Another possible solution for this is instead using MC for training the continuum suppression BDTs, off-resonance data could be used to train on. Off-resonance data is data taking with the Belle II experiment while the beam energy is lowered to just below the $\Upsilon(4S)$ threshold, such that only continuum events are present. This would remove the uncertainties coming from the mismodeling of the BDT input variables in the MC compared to data.

6.4 Possible analysis improvements

This analysis uses BDTs for the suppression of continuum and $B\overline{B}$ events, and for the signal extraction fit. BDTs are a relatively old machine learning technique, and outside the field of particle physics they have been replaced by neural networks and deep learning techniques, which often come with a better performance [36]. Within the Belle II experiment, the use of deep neural networks for continuum suppression has already been successfully studied [37]. The deep neural networks did not only lead to a comparable or better performance compared to BDTs, but also to a reduced training time. However, caution is advised when implementing these solutions, since with deep neural networks it is more difficult to understand what is happening within the MVA, which makes it complicated to convey the systematic uncertainties.

Since the signal extraction fit was performed in q^2 bins, another possible improvement could be optimizing more selections in q^2 bins. The two-point momentum and $\theta_{\gamma\gamma}$ selections could be improved in this way, leading to a potentially higher signal efficiency and background rejection. Furthermore, the ΔE and M_{bc} selections could be optimized further. Since the distributions were not used for the signal extraction fit, due to the high signal and background yield correlations in the fit, a tighter selection may have been beneficial. Both variables could also be included in the $B\overline{B}$ suppression BDT training, potentially increasing their performance.

The signal extraction fit currently only fits two categories, signal and all background, which was necessary because of the low statistics available. In possible future studies, the fit can be improved by floating the background categories separately and fixing well-understood background decay channels, e.g. $B \to D\ell\nu$ and $B \to D^*\ell\nu$ using their known branching fractions. This could potentially lead to a smaller uncertainty on the yield, and lower correlations between the signal and background yields. Ideally this analysis would be combined with the Belle II untagged $B \to \pi\ell\nu$ analysis, where a simultaneous fit for both decays mode could be used to constrain the cross-feed background events for both ρ and π modes.

Lastly, when more data becomes available, the selections performed in this thesis could be re-optimized. Since this analysis is statistically limited, increasing the data size should lead to improved results. FOM optimizations would potentially change, and selections would shift to tighter values due to the presence of more signal events in the data sample. Due to these tighter selections, more background would be removed, and the signal-to-background ratio would improve as well, leading to a better fit result.

CHAPTER 7

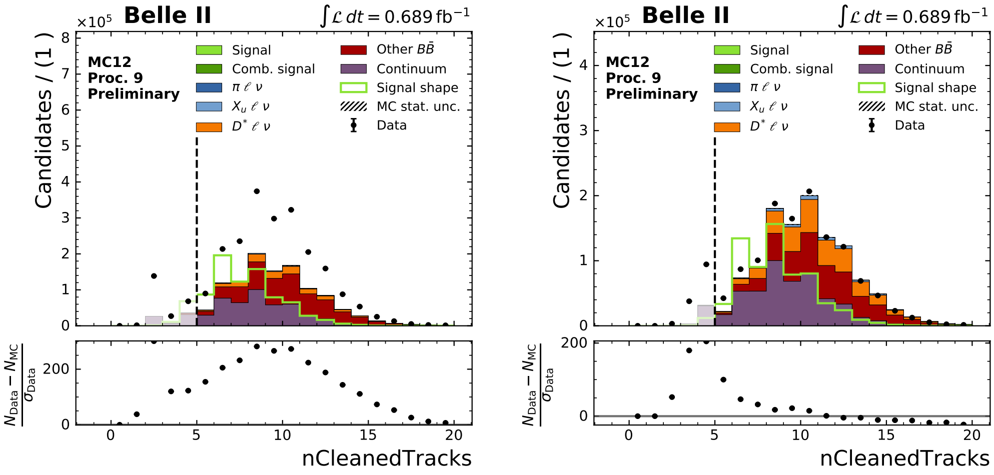
First look at early Belle II data

Due to time limitations, a full study on the data-MC agreement could not be done, however this chapter will describe a first look at the early phase 3 Belle II data. The Belle II experiment is in the early stages of its operation, and therefore the data-MC agreement is expected to be sub-optimal. To quantify the data-MC agreement, the pull was calculated as $(N_{\text{Data}} - N_{\text{MC}})/\sigma_{\text{Data}}$, indicating how significant the difference between data and MC is. Since the μ PID was not functional in early phase 3 data, the distributions are only shown for the electron signal decay channels. The data used is the proc9 data sample, consisting of experiment 3, 7, and 8, with a total integrated luminosity of 3.124 fb^{-1} [27]. Firstly, the data-MC agreement will be checked after only the pre-selections. Secondly, all the selections up until the BDT output classifiers were applied, and the agreement was checked again. Finally, the agreement in the BDT output classifier fit region was investigated.

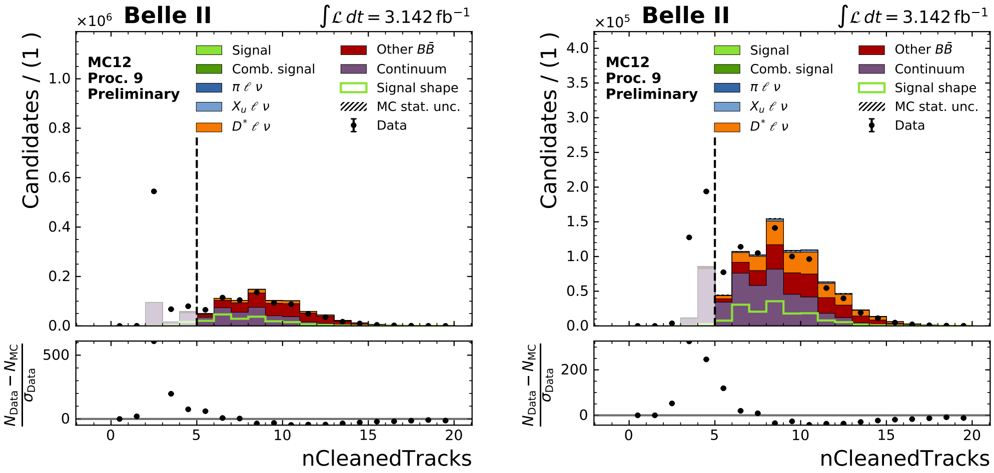
7.1 Pre-selections

Instead of using the whole proc9 data set, the data-MC distributions after the pre-selections are shown using the experiment 7 subset from the proc9 dataset, corresponding to 689 pb^{-1}. This was required due to the large candidate multiplicity, leading to very large filesizes which are difficult to process.

Figure 7.1 (top) shows the nCleanedTracks distributions for the experiment 7 data sample and the MC scaled to the expected luminosity. It can be seen that at nCleanedTracks equal to 2, the data and MC disagree significantly. This was caused by the fact that Bhabha scattering was not included in the used MC, however these events were removed by the nCleanedTrack selection described in section 5.2.3, and this disagreement should therefore not affect the analysis. Looking at the B^0 distribution, it can be seen that the data contains almost double the amount of events than the MC, compared to the $B^{\pm}$ distribution where there is a relatively good agreement between data and MC. A potential explanation for this is the presence of more beam background photons in the data compared to what was modeled in the MC, leading to more possible π^0 candidates and therefore a higher candidate multiplicity. This was checked by performing the best candidate selection described in section 5.4 after the pre-selections already, such that only 1 candidate per event was left. The nCleanedTracks distributions after this selection are shown in figure 7.1 (bottom). It can be seen that the agreement between data and MC improves significantly by this selection, confirming that the disagreement stems from combinatorial background events, which lead to a higher candidate multiplicity.



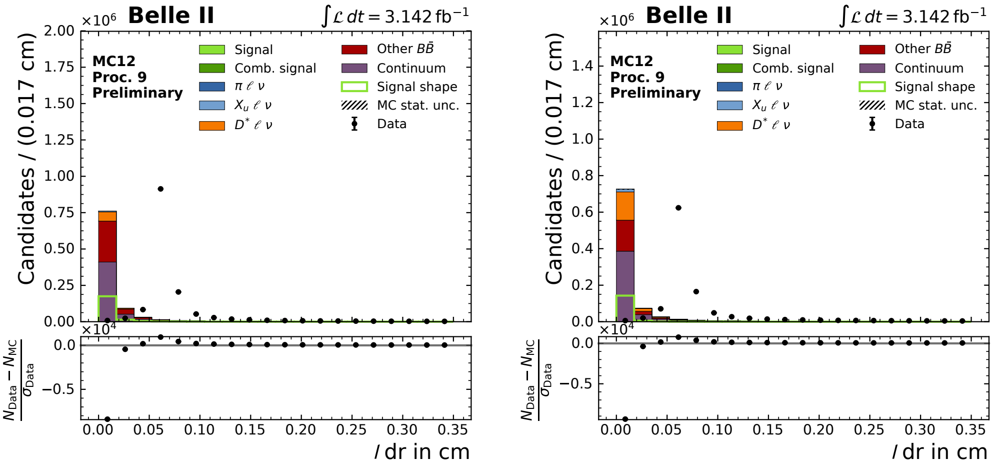
(a) $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ (b) $B^{\pm} \to \rho^0 e^{\pm} \nu_e$



(c) $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ (d) $B^{\pm} \to \rho^0 e^{\pm} \nu_e$

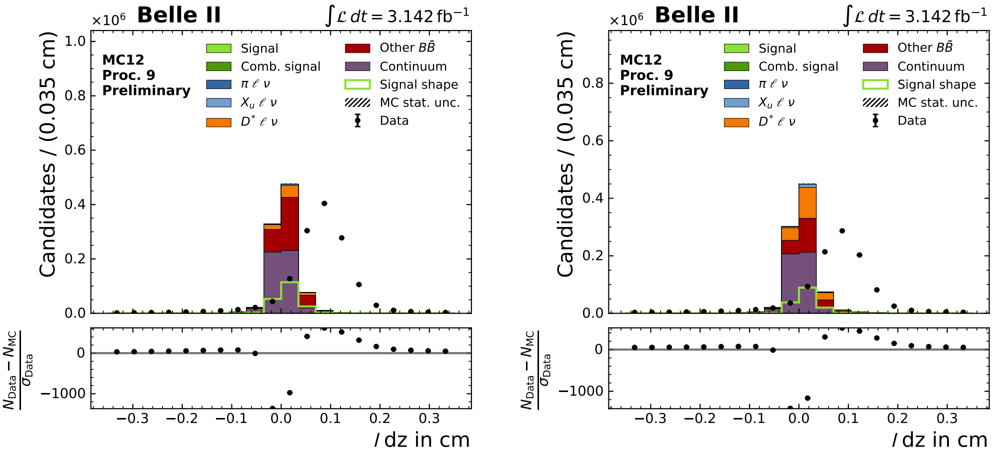
Figure 7.1: Distributions for nCleanedTracks using the 689 pb^{-1} experiment 7 data sample after pre-selections, before (top) and after (bottom) best candidate selection using the whole 3.124 fb^{-1} proc9 data sample. Large disagreements between data and MC are visible before the best candidate selection, mostly due to the presence of BhaBha scattering and the potentially higher of number beam background photons in data. The disargeement mostly disappears after the best candidate selection.

Distributions for dr and dz of the lepton candidate track after the pre-selections and best candidate selection are shown in figure 7.2. It can be seen that there is a shift in the position of the interaction point in data compared to MC, and that the shape of the MC template also does not match the data. This mismodeling could potentially lead to wrongly calculated kinematic variables, e.g. the momenta of particles and their decay angles. Another possible explanation for the difference is that some detector components are not performing as well as they are simulated, or are poorly calibrated.



(a) $B^0 \to \rho^\pm e^\mp \nu_e$ dr

(b) $B^\pm \to \rho^0 e^\pm \nu_e$ dr

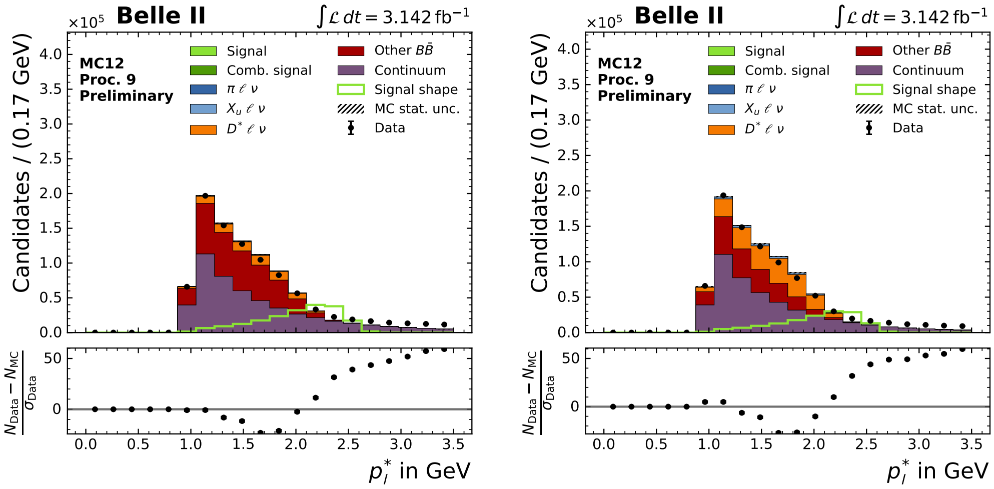


(c) $B^0 \to \rho^\pm e^\mp \nu_e$ dz

(d) $B^\pm \to \rho^0 e^\pm \nu_e$ dz

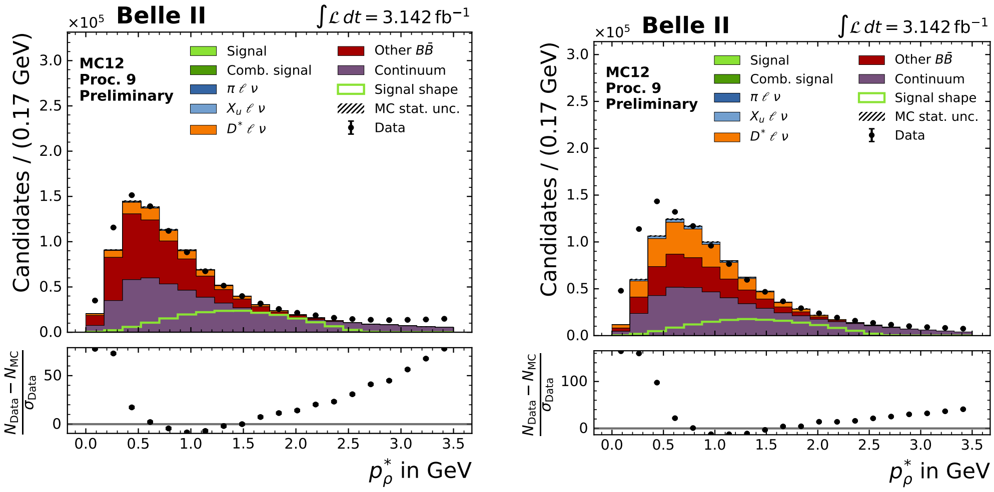
Figure 7.2: Lepton candidate dr and dz distributions for the proc9 data sample after pre-selections and best candidate selection. While the MC peaks at 0, the data shows a shifted interaction point in both radial and z-direction.

These effects can be seen in the p_ℓ^* and p_ρ^* distributions in figure 7.3, where both distributions show a deficit of data in the mid to high momentum region, where most $B\overline{B}$ events are expected. At the tail of the distributions, it can be seen that the continuum is not correctly modeled and that there seems to be more continuum in data than in the MC templates.



(a) $B^0 \rightarrow \rho^\pm e^\mp \nu_e\ p_\ell^*$

(b) $B^\pm \rightarrow \rho^0 e^\pm \nu_e\ p_\ell^*$



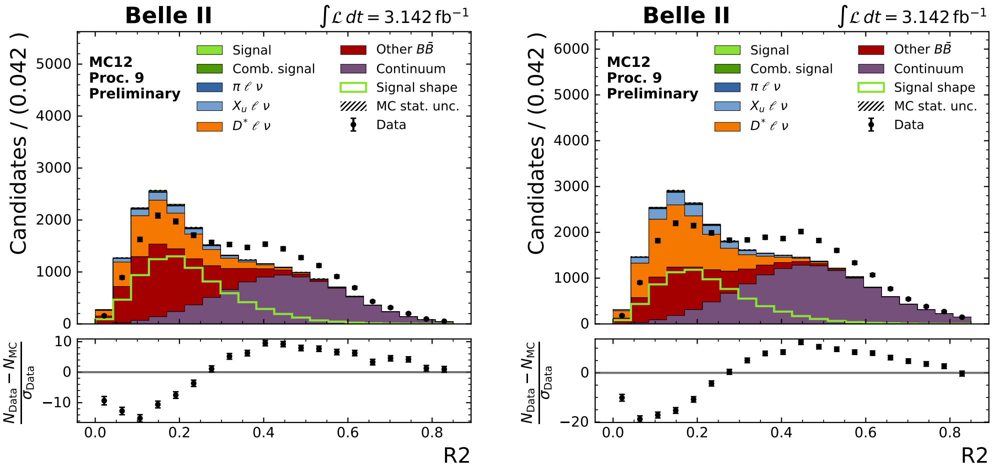
(c) $B^0 \rightarrow \rho^\pm e^\mp \nu_e\ p_\rho^*$

(d) $B^\pm \rightarrow \rho^0 e^\pm \nu_e\ p_\rho^*$

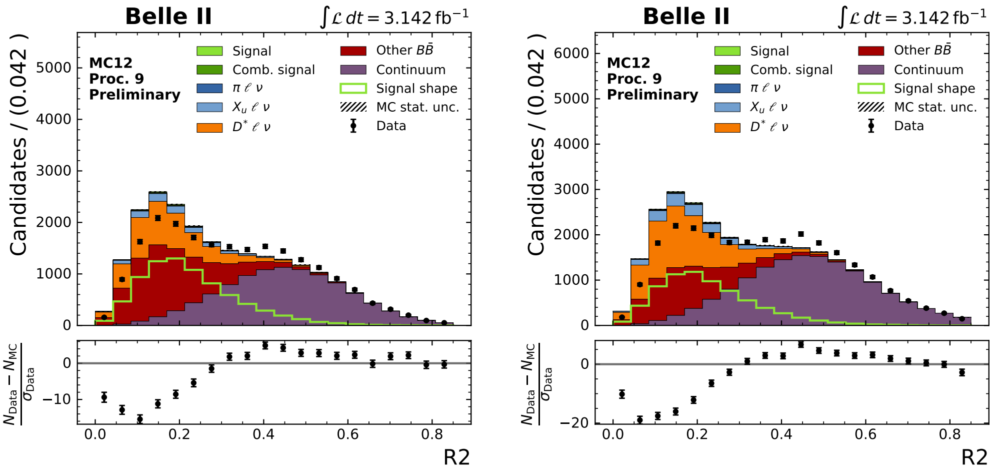
Figure 7.3: p_ℓ^* and p_ρ^* distributions. A deficit of data is visible in the region where MC predicts the most events.

7.2 Pre-BDT selections

After all selections up until the best candidate selection, the data-MC agreement was checked again. Distributions for R_2 are shown in figure 7.4 (top), and indicate a mismatch between the amount of continuum and $B\overline{B}$ events in the MC compared to the data. The data was lower than the MC in the low R_2 region, where more $B\overline{B}$ events are expected, compared to the high R_2 region, where the data exceeded the MC, indicating the presence of more continuum events. To verify this, the continuum was scaled up by 20% and the agreement was checked again, see figure 7.4 (bottom). This lead to an improved data-MC agreement, but for an optimal agreement a fit could be performed by floating the relative $B\overline{B}$ and continuum components to the data.



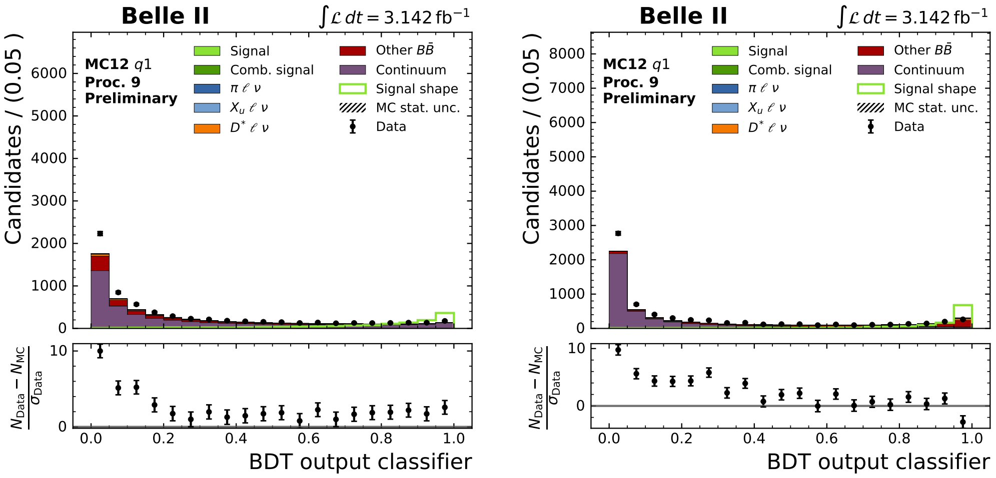
(a) $B^0 \to \rho^\pm e^\mp \nu_e$ before scaling (b) $B^\pm \to \rho^0 e^\pm \nu_e$ before scaling



(c) $B^0 \to \rho^\pm e^\mp \nu_e$ after scaling (d) $B^\pm \to \rho^0 e^\pm \nu_e$ after scaling

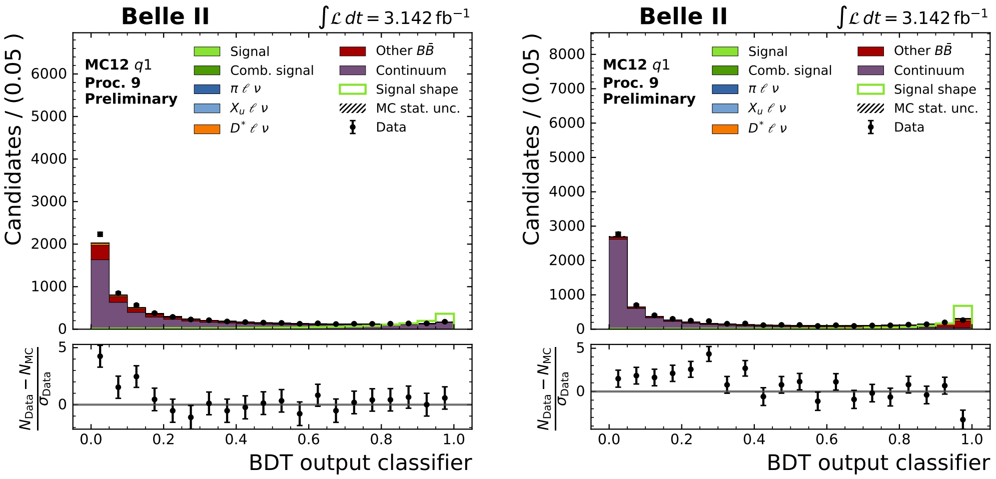
Figure 7.4: Distributions for R_2 using the whole proc9 data sample corresponding to 3.124 fb^{-1}. Top: unscaled continuum. Bottom: scaled-up continuum by 20%.

The same effect can be seen in the BDT output classifiers for the $B^0 \rightarrow \rho^{\pm} e^{\mp} \nu_e$ signal decay channel in the $q1$ bin, shown in figure 7.5 both with (bottom) and without (top) continuum scaling. The agreement significantly improves with the scaled continuum, yet in the $B\overline{B}$ suppression BDT output classifier there is a deficit of signal events in the high output classifier region.



(a) $B\overline{B}$ suppression BDT output classifier in the $q1$ bin before continuum scaling

(b) Continuum suppression BDT output classifier in the $q1$ bin before continuum scaling



(c) $B\overline{B}$ suppression BDT output classifier in the $q1$ bin after continuum scaling

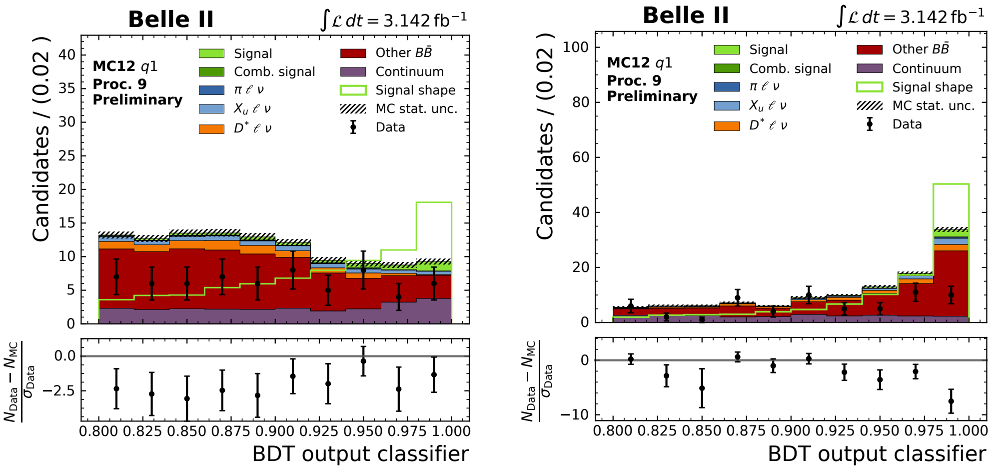
(d) Continuum suppression BDT output classifier in the $q1$ bin after continuum scaling

Figure 7.5: BDT output classifiers for the $B^0 \rightarrow \rho^{\pm} e^{\mp} \nu_e$ channel in the $q1$ bin, before (top) and after (bottom) scaling the continuum.

7.3 BDT fit region

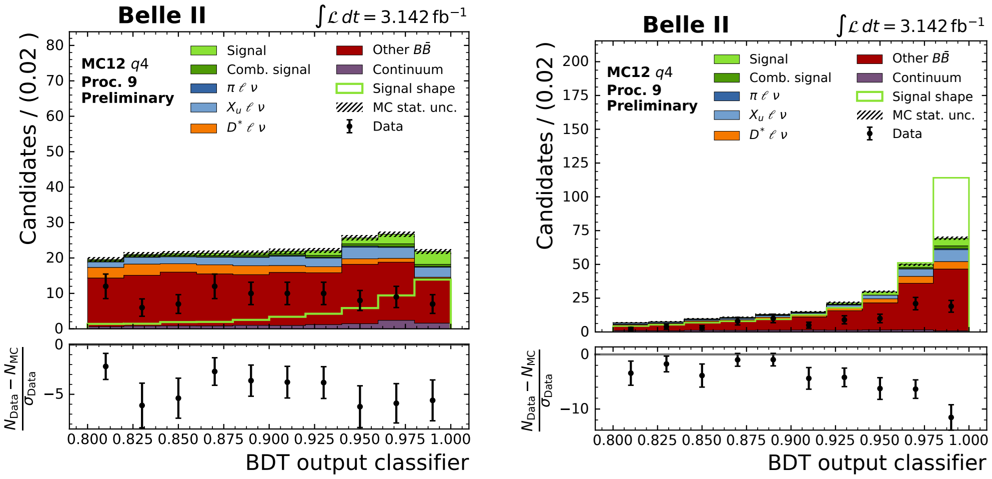
Finally, the data-MC agreement was checked in the BDT fit region, where both the continuum suppression and the $B\overline{B}$ suppression BDT output classifiers exceed 0.8. The BDT output classifiers for the $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ channel in the $q1$ and $q4$ bins are shown in figure 7.6. The data is in rough agreement, but overall lower than the MC in all the distributions. This is most likely the same deficit seen in the p^*_ℓ, p^*_ρ and R_2 distributions. Another factor adding to this could be the mismodeling of the BDT input variables in MC, like p^*_ℓ and p^*_ρ, leading to a worse performance on data, and therefore fewer events having a high output classifier. With the current data-MC agreement, a meaningful branching fraction extraction is not possible. Further study is required to fully understand the data, before a signal extraction fit can be performed.

It is important to note that this disagreement is most likely caused by low-level detector performance issues, which have to be addressed by the Belle II detector working groups. The disagreement is not caused by deficiencies in this analysis, and over time the agreement between data and MC should improve. Furthermore, new MC versions could potentially simulate the detector response better and therefore further improve the agreement.



(a) $B\overline{B}$ suppression BDT output classifier in the $q1$ bin

(b) Continuum suppression BDT output classifier in the $q1$ bin



(c) $B\overline{B}$ suppression BDT output classifier in the $q4$ bin

(d) Continuum suppression BDT output classifier in the $q4$ bin

Figure 7.6: BDT output classifiers for the $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ channel in the $q1$ and $q4$ bins.

CHAPTER 8

Summary, Conclusion and Outlook

In this thesis, the decay $B \longrightarrow \rho\ell\nu$ was studied using Belle II Monte Carlo simulation corresponding to the early Belle II data size of 6.5 fb^{-1}. Selections were performed to reduce background events, and BDTs were trained to further reject continuum and $B\overline{B}$ background events. It was found that performing a FOM selection on the BDT output classifiers lead to highly correlated signal and background shapes in ΔE and M_{bc}, so instead the BDT output classifiers were fitted. It was found that a significant branching fraction measurement was not possible with the current data set, and at least 16 fb^{-1} is required for the 'evidence' of $B \longrightarrow \rho\ell\nu$ at Belle II. Due to time limitations, further study on the systematic uncertainties of this analysis could not be performed, but the major sources are expected to be the MC modeling of the detector response,the MC modeling of continuum, and the modeling of inclusive X_u background. Early Belle II data showed disagreements compared to the MC that was used, indicating that further studies by the Belle II detector groups are required to fully understand the data before a meaningful signal extraction on data is possible performed.

8.1 Summary

The analysis was optimized for the expected early Belle II data, corresponding to 10 fb^{-1}. A set of reconstruction selections was used to reduce background from misreconstructed B candidates. Further selections on $\cos\theta_{BY}$, the momentum of the final state particles, and m_ρ were used to remove background from $B \longrightarrow X_c\ell\nu$ and combinatorial B events. Boosted Decision Trees were trained for each signal decay channel, split into five q^2 bins, for the suppression of continuum and $B\overline{B}$ events. A 2D figure of merit optimization of selections on the two BDT output classifiers in every q^2 bin was performed to maximize the signal yield. Results indicate that the signal and background yield in the fit were too correlated for a correctly working fit, so the BDT output classifiers were used for the signal extraction fit instead.

The signal extraction fit was a simultaneous maximum likelihood fit of the obtained post-selection MC templates of the continuum suppression and the $B\overline{B}$ BDT output classifiers using a MC sample corresponding to the actually obtained Belle II data size of 6.5 fb^{-1}. Two components were used in the fit, one containing signal and combinatorial signal events, and the other all other background categories combined. The fit was verified using Asimov fits and pull distributions obtained by throwing 1000 MC toys. Only the electron channels were fitted, since the μ PID was not working in the early Belle II data. The signal branching fractions with their statistical uncertainties were obtained as

$\mathcal{B}(B^{\pm} \to \rho^0 e^{\pm} \nu_e) = (2 \pm 1) \cdot 10^{-4}$ and $\mathcal{B}(B^0 \to \rho^{\pm} e^{\mp} \nu_e) = (3 \pm 2) \cdot 10^{-4}$, which agree with the world average PDG values of $(1.58 \pm 0.11) \cdot 10^{-4}$ for the B^0 channel and $(2.94 \pm 0.21) \cdot 10^{-4}$ for the $B^{\pm}$ channel respectively. These results are not significant, due to the statistically limited available data, indicating that a significant measurement of $\mathcal{B}(B \to \rho \ell \nu)$ is not possible with the current Belle II data set. A MC study was performed to find the amount of data needed to obtain a 3σ statistically significant result, which showed that at least 16 fb^{-1} is required. This amount of data will most likely be reached in the winter run of the Belle II experiment.

Due to time limitations, further study on the systematic uncertainties of this analysis could not be performed. Looking at similar previously performed analyses, the major sources of systematic uncertainty will come from the uncertainties on the form factors used in the Monte Carlo, the early Belle II detector performance, the uncertainties on the branching fractions of the signal and background channels, the modeling of the inclusive X_u MC, and the use of BDTs. Possible improvements to the analysis could be the use of (deep) neural networks instead of BDTs for the continuum and $B\overline{B}$ suppression, the optimization of all selections in q^2 bins, fixing certain background categories in the fit to their expected branching fractions, and combining this analysis with the Belle II untagged $B \to \pi \ell \nu$ analysis, allowing for a more sophisticated fitting method using crossfeed between the two decay modes.

Early Belle II data was inspected, and the data-MC agreement was checked at multiple selection stages. It was seen that the data had more combinatorial background events, most likely due to the presence of more beam background photons then what was modeled in the MC. The interaction point was shifted away from 0, which could potentially lead to wrong calculations of kinematic variables. The R_2 distributions showed that the continuum background was underestimated, which can possibly be solved by scaling the continuum events. With the current data-MC agreement, a meaningful signal extraction is not possible, due to the large disagreements between the data and the Monte Carlo templates.

8.2 Outlook

In the near future, this analysis will indeed be combined with the untagged $B \to \pi \ell \nu$ analysis, and with more data, the measurement of the branching fraction should be possible, if the data-MC agreement can be improved. Further study is required on the systematic uncertainties, since these are not understood at the moment. The analysis has been set up such that the efficiencies in all q^2 bins are roughly constant, however the $\cos\theta_l$ efficiency is not flat. This could be a problem for the extraction of $|V_{ub}|$, since not many events will be left in the first $\cos\theta_l$ bin. This could potentially be prevented by slightly changing the selections on the momentum of the ℓ and ρ. Additionally, with more available data some selections might change, and the FOM optimizations for the BDT output classifiers might shift to less tight selections, leading to the possibility of a combined ΔE and M_{bc} fit compared to a fit on the BDT output classifiers. Moreover, the use of fewer bins for the $\cos\theta_l$ distribution could also help, although this comes at a cost of reduced resolution.

In conclusion, ‘evidence’ of $B \to \rho \ell \nu$ at Belle II should be possible when more data becomes available in the near future, at roughly 16 fb^{-1}, however further study is still required to fully understand the data, the systematic uncertainties, and the use of a more sophisticated fitting method by possibly combining the extraction of $B \to \pi \ell \nu$ and $B \to \rho \ell \nu$ into one fit.

Bibliography

[1] J. J. Thomson, *XL. Cathode rays*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44** (1897) 293 (cit. on p. 1).

[2] H. Gegier and E. Marsden, *On a diffuse reflection of the alpha-particles*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **82** (1909) (cit. on p. 1).

[3] N. Bohr, *I. On the constitution of atoms and molecules*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **26** (1913) 1 (cit. on p. 1).

[4] P. A. M. Dirac, *The principles of quantum mechanics*, 27, Oxford university press, 1930 (cit. on p. 1).

[5] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579; A. Salam, “Weak and Electromagnetic Interactions”, *Elementary particle theory. Relativistic groups and analyticity. Proceedings of the Eighth Nobel Symposium*, ed. by N. Svartholm, Stockholm: Almquist & Wiksell, 1968 367; S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264, cit. on p. 1.

[6] G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, arXiv: `1207.7214 [hep-ex]`; S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, arXiv: `1207.7235 [hep-ex]`, cit. on p. 1.

[7] S. Dimopoulos and H. Georgi, *Softly broken supersymmetry and SU(5)*, Nuclear Physics B **193** (1981) 150, ISSN: 0550-3213, URL: `http://www.sciencedirect.com/science/article/pii/0550321381905228` (cit. on p. 1).

[8] T. Abe et al., *Belle II Technical Design Report*, arXiv e-prints, arXiv:1011.0352 (2010) arXiv:1011.0352, arXiv: `1011.0352 [physics.ins-det]` (cit. on pp. 1, 9, 10, 13).

[9] A. Petrella, *Inclusive and Exclusive* $|V_{ub}|$, arXiv e-prints, arXiv:0903.5180 (2009) arXiv:0903.5180, arXiv: `0903.5180 [hep-ex]` (cit. on p. 2).

[10] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 9781107034266, URL: `http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.2.T46::2013` (cit. on p. 3).

[11] Wikipedia, the free encyclopedia, *Standard Model of Elementary Particles*, [Online; accessed March 05, 2019], 2019, URL: `https://upload.wikimedia.org/wikipedia/commons/0/00/Standard_Model_of_Elementary_Particles.svg` (cit. on p. 4).

[12] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508, URL: `https://link.aps.org/doi/10.1103/PhysRevLett.13.508` (cit. on p. 4).

[13] C. S. Wu et al., *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (4 1957) 1413, URL: `https://link.aps.org/doi/10.1103/PhysRev.105.1413` (cit. on p. 4).

[14] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (1973) 652, ISSN: 0033-068X, eprint: `http://oup.prod.sis.lan/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf`, URL: `https://dx.doi.org/10.1143/PTP.49.652` (cit. on p. 4).

[15] J. Dingfelder and T. Mannel, *Leptonic and semileptonic decays of B mesons*, Rev. Mod. Phys. **88** (3 2016) 035008, URL: `https://link.aps.org/doi/10.1103/RevModPhys.88.035008` (cit. on pp. 5, 22, 24).

[16] M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** (3 2018) 030001, URL: `https://link.aps.org/doi/10.1103/PhysRevD.98.030001` (cit. on pp. 5, 27).

[17] P. Del Amo Sanchez et al., *Study of $B \to \pi l \nu$ and $B \to \rho l \nu$ decays and determination of $|V_{ub}|$*, Phys. Rev. D **83**, 032007 (2011) 032007, arXiv: `1005.3288 [hep-ex]` (cit. on pp. 5, 6, 46).

[18] N. Isgur et al., *Semileptonic B and D decays in the quark model*, Phys. Rev. D **39** (3 1989) 799, URL: `https://link.aps.org/doi/10.1103/PhysRevD.39.799` (cit. on p. 6).

[19] P. Ball and R. Zwicky, *$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form factors from light-cone sum rules reexamined*, Phys. Rev. D **71** (1 2005) 014029, URL: `https://link.aps.org/doi/10.1103/PhysRevD.71.014029` (cit. on p. 6).

[20] J. W. Flanagan and Y. Ohnishi, *Letter of Intent for KEK Super B Factory Part III: Accelerator Design*, 2004, URL: `http://www-superkekb.kek.jp/documents/LoI_accelerator.pdf` (cit. on p. 9).

[21] KEK, *SuperKEKB schematic drawing*, [Online; accessed March 08, 2019], 2016, URL: `https://phys.org/news/2016-04-particles-circulate-superkekb.html` (cit. on p. 9).

[22] J. H. Friedman, *Greedy function approximation: A gradient boosting machine.*, Ann. Statist. **29** (2001) 1189, URL: `https://doi.org/10.1214/aos/1013203451` (cit. on p. 12).

[23] T. Keck, *FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification*, arXiv e-prints, arXiv:1609.06119 (2016) arXiv:1609.06119, arXiv: `1609.06119 [cs.LG]` (cit. on pp. 12, 29).

[24] D. Ignatov and A. Ignatov, *Decision Stream: Cultivating Deep Decision Trees*, arXiv e-prints, arXiv:1704.07657 (2017) arXiv:1704.07657, arXiv: `1704.07657 [cs.LG]` (cit. on p. 12).

[25] S. Pohl and C. Kiesling, *Track Reconstruction at the First Level Trigger of the Belle II Experiment*, Presented on 11 04 2018, PhD thesis: Munich, Ludwig-Maximilians-Universität, 2018 (cit. on pp. 13, 14).

[26] Belle 2 data production group, *Data Production MC12*, [Online; accessed March 12, 2019], 2019, URL: `https://confluence.desy.de/pages/viewpage.action?spaceKey=BI&title=Data+Production+MC12` (cit. on pp. 15, 42).

[27] Belle 2 data production group, *Phase 3 data*, [Online; accessed July 8, 2019], 2019, URL: `https://confluence.desy.de/display/BI/Phase+3+data` (cit. on pp. 15, 49).

[28] F. Tenchini, "A Decay Tree Fitter for the Belle II Analysis Framework", 0206, 25th International Workshop on Vertex Detectors, La Biodola, Isola d'Elba, Italy (cit. on p. 16).

[29] W. Waltenberger, *RAVE—A Detector-Independent Toolkit to Reconstruct Vertices*, IEEE Transactions on Nuclear Science **58** (2011) 434 (cit. on p. 21).

[30] G. C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, Phys. Rev. Lett. **41** (23 1978) 1581, URL: `https://link.aps.org/doi/10.1103/PhysRevLett.41.1581` (cit. on p. 29).

[31] A. J. Bevan et al., *The Physics of the B Factories*, The European Physical Journal C **74** (2014) 3026, ISSN: 1434-6052, URL: `https://doi.org/10.1140/epjc/s10052-014-3026-9` (cit. on p. 29).

[32] D. M. Asner et al., *Search for exclusive charmless hadronic B decays*, Phys. Rev. D **53** (3 1996) 1039, URL: `https://link.aps.org/doi/10.1103/PhysRevD.53.1039` (cit. on p. 29).

[33] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, arXiv e-prints, physics/0306116 (2003) physics/0306116, arXiv: `physics/0306116 [physics.data-an]` (cit. on p. 41).

[34] H. Ha et al., *Measurement of the decay $B^0 \to \pi^-\ell^+\nu$ and determination of $|V_{ub}|$*, Phys. Rev. D **83** (7 2011) 071101, URL: `https://link.aps.org/doi/10.1103/PhysRevD.83.071101` (cit. on p. 46).

[35] F. D. Fazio and M. Neubert, *B→Xulbar nul decay distributions to order αs*, Journal of High Energy Physics **1999** (1999) 017, URL: `https://doi.org/10.1088%2F1126-6708%2F1999%2F06%2F017` (cit. on p. 46).

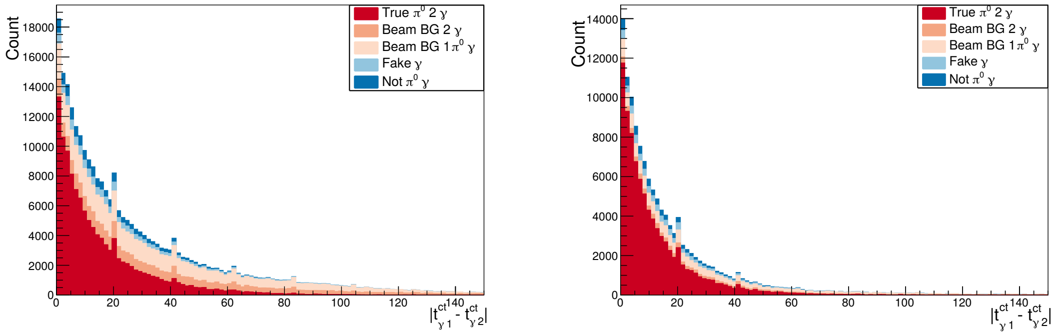
[36] I. Goodfellow, Y. Bengio and A. Courville, *Deep learning*, MIT press, 2016 (cit. on p. 47).

[37] D. Weyland,
Continuum Suppression with Deep Learning techniques for the Belle II Experiment,
Karlsruher Institut für Technologie (KIT), Masterarbeit, 2017,
MS: Karlsruher Institut für Technologie (KIT), 2017,
URL: `https://ekp-invenio.physik.uni-karlsruhe.de/record/48934` (cit. on p. 47).

APPENDIX A

Appendix

A.1 $|\Delta t_{\gamma\gamma}|$ selection

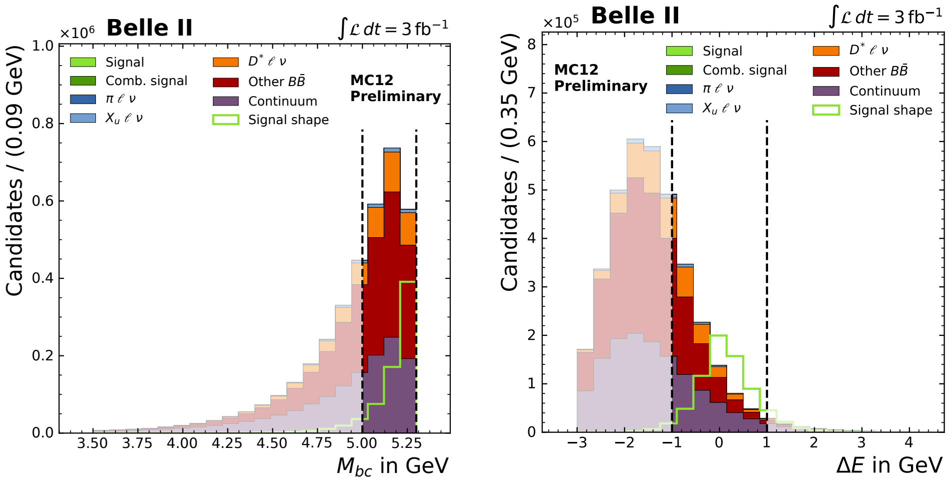


(a) $|\Delta t_{\gamma\gamma}|$ before $\theta_{\gamma\gamma}$ selection.

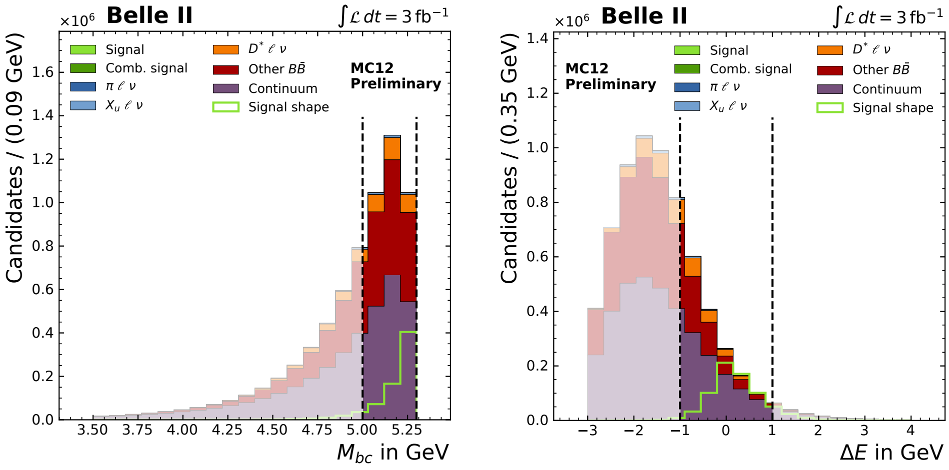
(b) $|\Delta t_{\gamma\gamma}|$ after $\theta_{\gamma\gamma}$ selection.

Figure A.1: $|\Delta t_{\gamma\gamma}|$ distributions for reconstructed π^0. The regular features are due to the timing resolution of the detector.

A.2 Fit region



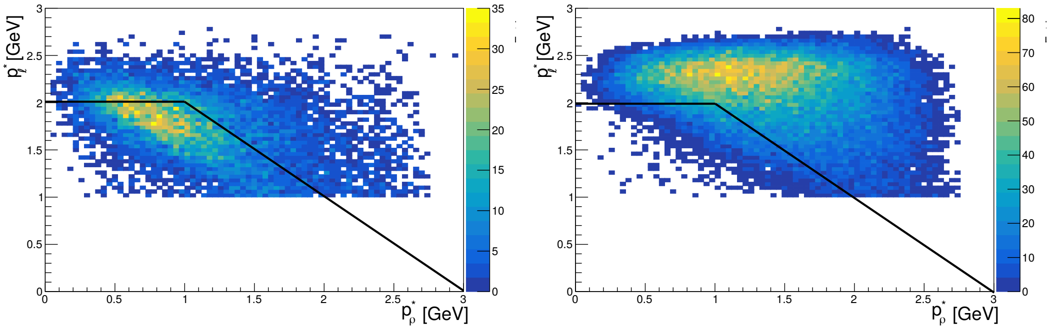
(a) M_{bc} for $B^0 \to \rho^\pm e^\mp \nu_e$ (b) ΔE for $B^0 \to \rho^\pm e^\mp \nu_e$



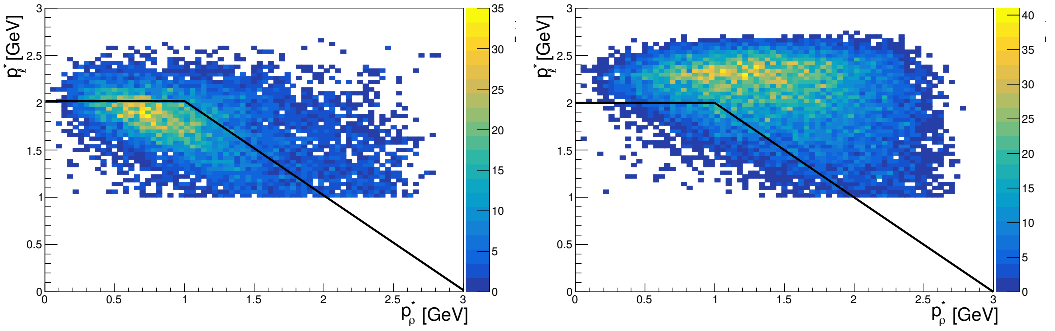
(c) M_{bc} for $B^0 \to \rho^\pm \mu^\mp \nu_\mu$ (d) ΔE for $B^0 \to \rho^\pm \mu^\mp \nu_\mu$

Figure A.2: M_{bc} (left) and ΔE (right) distributions for B^0 signal decays with the defined fit region $|\Delta E| < 1.0$ GeV and $5.0 < M_{bc} < 5.3$ GeV.

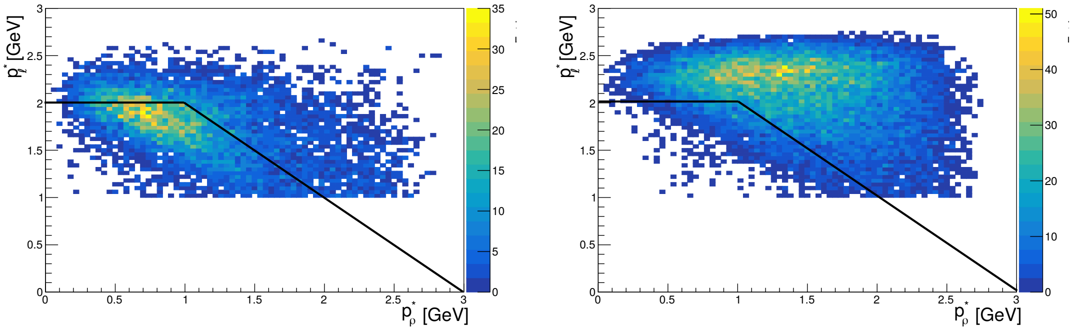
A.3 Momentum selection



(a) $B^{\pm} \to \rho^{0} \mu^{\pm} \nu_{\mu}$ Background (b) $B^{\pm} \to \rho^{0} \mu^{\pm} \nu_{\mu}$ Signal



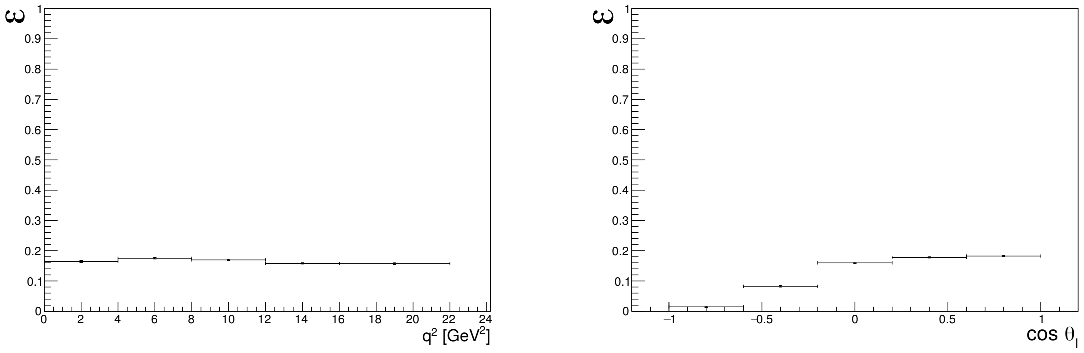
(c) $B^{0} \to \rho^{\pm} e^{\mp} \nu_{e}$ Background (d) $B^{0} \to \rho^{\pm} e^{\mp} \nu_{e}$ Signal



(e) $B^{0} \to \rho^{\pm} \mu^{\mp} \nu_{\mu}$ Background (f) $B^{0} \to \rho^{\pm} \mu^{\mp} \nu_{\mu}$ Signal

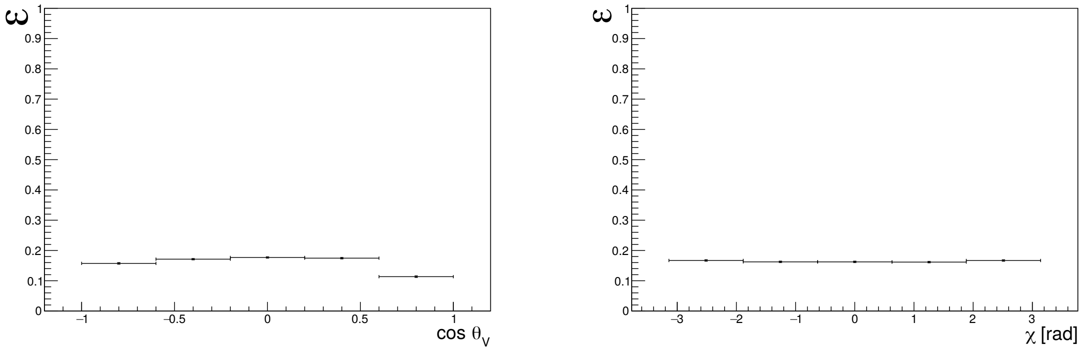
Figure A.3: 2D distributions for p^*_ρ and p^*_ℓ for background (left) and signal (right) in the remaining signal decay channels.

A.4 Efficiency distributions



(a) q^2 efficiency.

(b) $\cos\theta_l$ efficiency.



(c) $\cos\theta_V$ efficiency.

(d) χ efficiency.

Figure A.4: Signal efficiency distributions of q^2 and the three decay angles $\cos\theta_l$, $\cos\theta_V$ and χ, for B^0 after pre-selections.

A.5 Efficiency tables

Selection	$\epsilon^e_{B^0}$	$\epsilon^\mu_{B^0}$	$\epsilon^e_{B^\pm}$	$\epsilon^\mu_{B^\pm}$
Acceptance and pre-selections	0.294 ± 0.004	0.271 ± 0.004	0.485 ± 0.002	0.478 ± 0.002
$\theta_{\gamma\gamma} < 1.20$ rad	0.975 ± 0.012	0.973 ± 0.013	-	-
nCleanedTracks > 4	0.908 ± 0.012	0.906 ± 0.012	0.981 ± 0.014	0.979 ± 0.015
$\cos\theta_{miss}$ in acceptance	0.917 ± 0.012	0.917 ± 0.013	0.918 ± 0.014	0.916 ± 0.014
M_{bc} and ΔE selection	0.844 ± 0.013	0.851 ± 0.013	0.833 ± 0.014	0.840 ± 0.014
ρ vertex fit > 10^{-10}	-	-	0.976 ± 0.017	0.976 ± 0.017
$\lvert\cos\theta_{BY}\rvert < 1.1$	0.848 ± 0.014	0.947 ± 0.015	0.871 ± 0.016	0.973 ± 0.017
Combined p^*_l/p^*_ρ	0.983 ± 0.017	0.981 ± 0.016	0.982 ± 0.018	0.985 ± 0.016
$0.55 < m_\rho < 1.30$ GeV	0.888 ± 0.015	0.883 ± 0.014	0.887 ± 0.017	0.885 ± 0.016
Best candidate selection	0.999 ± 0.018	0.999 ± 0.017	0.998 ± 0.020	0.999 ± 0.019
BDT classifiers selection	0.522 ± 0.011	0.542 ± 0.011	0.557 ± 0.013	0.583 ± 0.013
Total	0.077 ± 0.002	0.089 ± 0.002	0.150 ± 0.002	0.176 ± 0.002

Table A.1: Efficiencies per performed selection for all 4 decay channels in the $q2$ bin.

Selection	$\epsilon^e_{B^0}$	$\epsilon^\mu_{B^0}$	$\epsilon^e_{B^\pm}$	$\epsilon^\mu_{B^\pm}$
Acceptance and pre-selections	0.294 ± 0.004	0.271 ± 0.004	0.485 ± 0.002	0.478 ± 0.002
$\theta_{\gamma\gamma} < 1.20$ rad	0.964 ± 0.011	0.965 ± 0.011	-	-
nCleanedTracks > 4	0.901 ± 0.010	0.905 ± 0.011	0.980 ± 0.013	0.981 ± 0.013
$\cos\theta_{miss}$ in acceptance	0.922 ± 0.010	0.926 ± 0.011	0.928 ± 0.012	0.928 ± 0.012
M_{bc} and ΔE selection	0.840 ± 0.011	0.858 ± 0.011	0.831 ± 0.012	0.841 ± 0.011
ρ vertex fit > 10^{-10}	-	-	0.973 ± 0.014	0.973 ± 0.014
$\lvert\cos\theta_{BY}\rvert < 1.1$	0.831 ± 0.011	0.955 ± 0.013	0.838 ± 0.013	0.967 ± 0.014
Combined p^*_l/p^*_ρ	0.983 ± 0.013	0.983 ± 0.013	0.985 ± 0.016	0.986 ± 0.015
$0.55 < m_\rho < 1.30$ GeV	0.932 ± 0.014	0.928 ± 0.014	0.930 ± 0.016	0.933 ± 0.014
Best candidate selection	0.999 ± 0.015	0.999 ± 0.014	0.999 ± 0.017	0.999 ± 0.014
BDT classifiers selection	0.582 ± 0.010	0.571 ± 0.009	0.581 ± 0.011	0.600 ± 0.011
Total	0.087 ± 0.004	0.010 ± 0.003	0.146 ± 0.002	0.192 ± 0.001

Table A.2: Efficiencies per performed selection for all 4 decay channels in the $q3$ bin.

Selection	$\epsilon^{e}_{B^0}$	$\epsilon^{\mu}_{B^0}$	$\epsilon^{e}_{B^\pm}$	$\epsilon^{\mu}_{B^\pm}$
Acceptance and pre-selections	0.294 ± 0.004	0.271 ± 0.004	0.485 ± 0.002	0.478 ± 0.002
$\theta_{\gamma\gamma} < 1.20$ rad	0.957 ± 0.011	0.954 ± 0.011	-	-
nCleanedTracks > 4	0.901 ± 0.010	0.900 ± 0.011	0.979 ± 0.013	0.981 ± 0.013
$\cos\theta_{miss}$ in acceptance	0.938 ± 0.011	0.941 ± 0.010	0.939 ± 0.012	0.941 ± 0.013
M_{bc} and ΔE selection	0.836 ± 0.011	0.855 ± 0.011	0.826 ± 0.012	0.843 ± 0.012
ρ vertex fit $> 10^{-10}$	-	-	0.973 ± 0.014	0.974 ± 0.014
$\lvert\cos\theta_{BY}\rvert < 1.1$	0.820 ± 0.012	0.957 ± 0.013	0.834 ± 0.013	0.969 ± 0.014
Combined p^*_l/p^*_ρ	0.964 ± 0.014	0.963 ± 0.013	0.957 ± 0.017	0.961 ± 0.014
$0.55 < m_\rho < 1.30$ GeV	0.968 ± 0.015	0.970 ± 0.014	0.971 ± 0.016	0.972 ± 0.015
Best candidate selection	0.999 ± 0.015	0.999 ± 0.014	0.998 ± 0.011	0.998 ± 0.015
BDT classifiers selection	0.583 ± 0.010	0.579 ± 0.010	0.584 ± 0.011	0.571 ± 0.010
Total	0.088 ± 0.003	0.010 ± 0.002	0.162 ± 0.002	0.190 ± 0.002

Table A.3: Efficiencies per performed selection for all 4 decay channels in the $q4$ bin.

Selection	$\epsilon^{e}_{B^0}$	$\epsilon^{\mu}_{B^0}$	$\epsilon^{e}_{B^\pm}$	$\epsilon^{\mu}_{B^\pm}$
Acceptance and pre-selections	0.294 ± 0.004	0.271 ± 0.004	0.485 ± 0.002	0.478 ± 0.002
$\theta_{\gamma\gamma} < 1.20$ rad	0.929 ± 0.015	0.927 ± 0.014	-	-
nCleanedTracks > 4	0.904 ± 0.014	0.900 ± 0.014	0.979 ± 0.016	0.981 ± 0.016
$\cos\theta_{miss}$ in acceptance	0.955 ± 0.014	0.953 ± 0.016	0.951 ± 0.016	0.956 ± 0.016
M_{bc} and ΔE selection	0.828 ± 0.014	0.843 ± 0.014	0.819 ± 0.016	0.831 ± 0.015
ρ vertex fit $> 10^{-10}$	-	-	0.970 ± 0.018	0.967 ± 0.018
$\lvert\cos\theta_{BY}\rvert < 1.1$	0.818 ± 0.015	0.953 ± 0.017	0.828 ± 0.016	0.957 ± 0.018
Combined p^*_l/p^*_ρ	0.931 ± 0.018	0.932 ± 0.017	0.930 ± 0.020	0.938 ± 0.018
$0.55 < m_\rho < 1.30$ GeV	0.975 ± 0.019	0.976 ± 0.019	0.969 ± 0.021	0.972 ± 0.019
Best candidate selection	0.999 ± 0.021	0.998 ± 0.019	0.998 ± 0.022	0.999 ± 0.020
BDT classifiers selection	0.599 ± 0.014	0.570 ± 0.013	0.569 ± 0.015	0.554 ± 0.013
Total	0.086 ± 0.002	0.097 ± 0.003	0.151 ± 0.002	0.177 ± 0.002

Table A.4: Efficiencies per performed selection for all 4 decay channels in the $q5$ bin.

A.6 D* veto

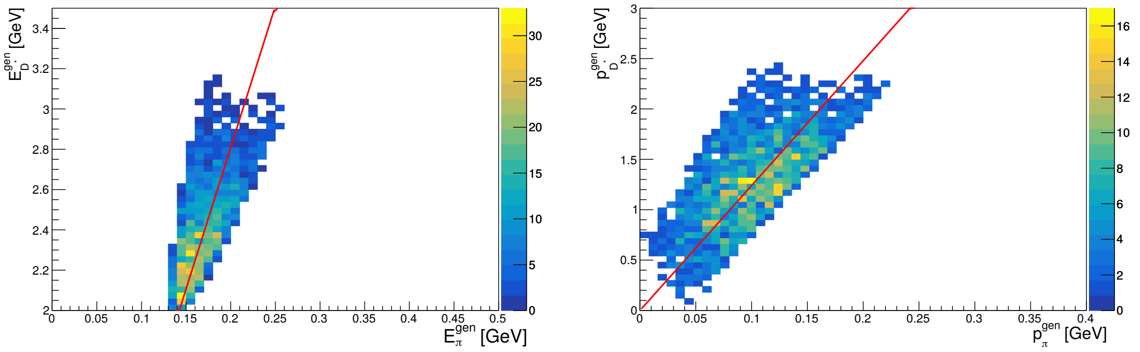
To suppress background coming from $B \longrightarrow D^*\ell\nu$ events, a D^* veto was created. The D^* decays via $D^* \longrightarrow D\pi_{\text{slow}}$, so a slow π is present in the event. With the performed selections in this analysis, slow π are rarely reconstructed on the signal side and will end up in the ROE. To veto events containing a D^*, a π_{slow} was reconstructed in the ROE using the following selections:

$\pi^{\pm}$: $p^*_\pi < 0.4$ GeV

γ: $E_\gamma > 0.05$ GeV

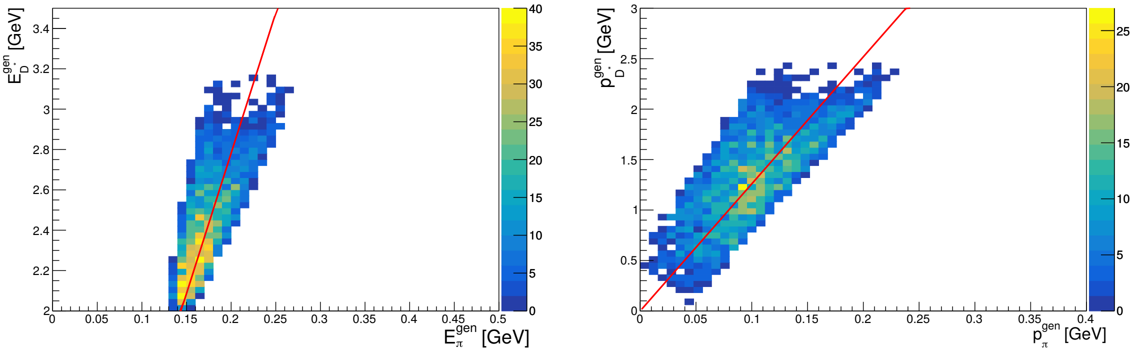
π^0: $0.08 < m_{\pi^0} < 0.20$ GeV and $p^*_\pi < 0.4$ GeV

The energy and momentum of the reconstructed π were used to calculate the energy and momentum of a D^* candidate using relations obtained from $B \longrightarrow D^*\ell\nu$ MC generator level information. Figure A.5 shows the 2D distributions and the $y = ax$ linear fits for $p^{\text{gen}}_{D^*}$ and p^{gen}_π for both $B^0 \longrightarrow D^*\ell\nu$ and $B^{\pm} \longrightarrow D^*\ell\nu$ decays.



(a) $B^{\pm} \longrightarrow D^*\ell\nu$ fit $a = 14.03 \pm 0.13$ ($\chi^2/ndf = 6.32$).

(b) $B^{\pm} \longrightarrow D^*\ell\nu$ fit $a = 12.40 \pm 0.17$ ($\chi^2/ndf = 1.44$).



(c) $B^0 \longrightarrow D^*\ell\nu$ fit $a = 13.92 \pm 0.12$ ($\chi^2/ndf = 6.09$).

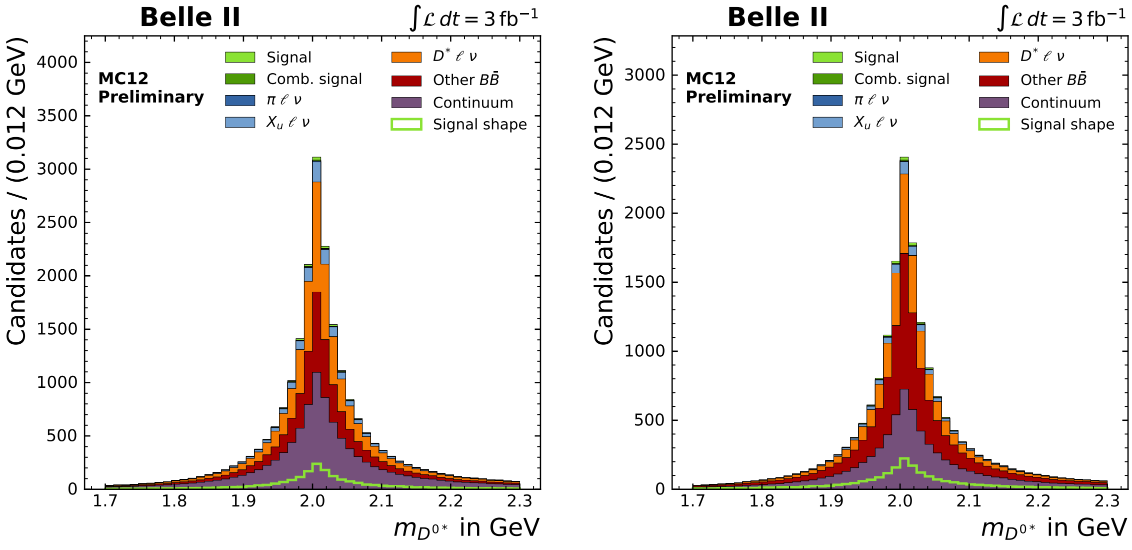
(d) $B^0 \longrightarrow D^*\ell\nu$ fit $a = 12.59 \pm 0.15$ ($\chi^2/ndf = 2.08$).

Figure A.5: Generator $E^*_{D^*}$ and E^*_π (left) and $p^*_{D^*}$ and p^*_π (right) distributions and fitted values.

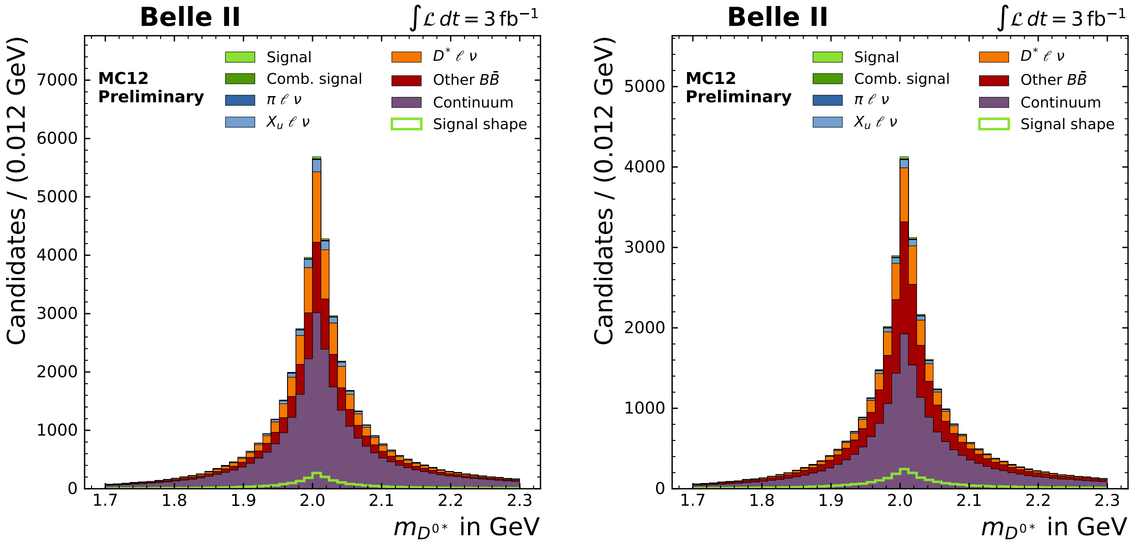
Using the measured $p^*_{\pi_{\text{slow}}}$, $E^*_{\pi_{\text{slow}}}$ and the obtained fit results, the invariant mass of the D^* candidate was calculated using

$$m_{D^*} = \sqrt{(a_E^{fit} E^*_{\pi_{\text{slow}}})^2 - (a_p^{fit} p^*_{\pi_{\text{slow}}})^2}, \tag{A.1}$$

where a_E^{fit} and a_p^{fit} are the obtained fit results from the MC generator information. Distributions for the reconstructed $m_{D^{*\pm}}$ before the BDT selections are shown in figure A.7 and for $m_{D^{*0}}$ in figure A.6. Looking at the distributions, no clear distinction between signal and background is visible, yet in the $B\overline{B}$ suppression BDT it has separation power.

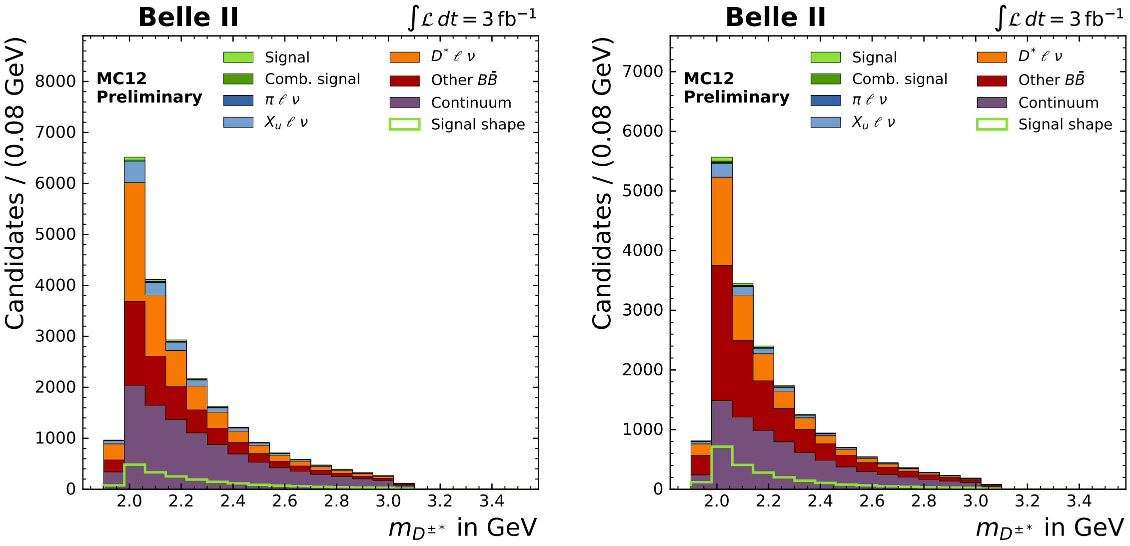


(a) $B^{\pm} \to \rho^{0} e^{\pm} \nu_e$ (b) $B^{0} \to \rho^{\pm} e^{\mp} \nu_e$

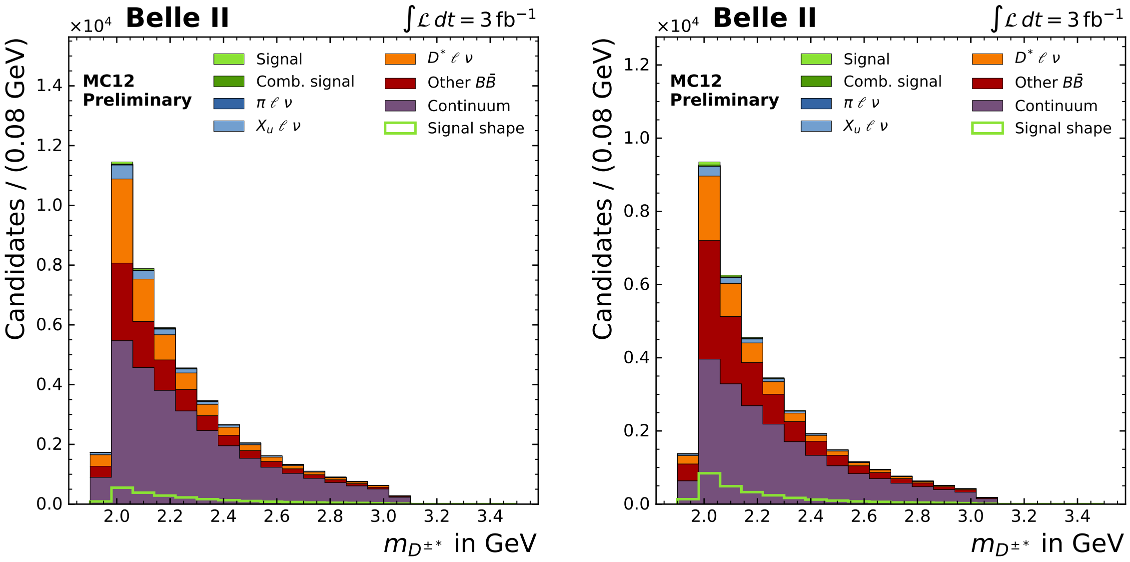


(c) $B^{\pm} \to \rho^{0} \mu^{\pm} \nu_{\mu}$ (d) $B^{0} \to \rho^{\pm} \mu^{\mp} \nu_{\mu}$

Figure A.6: Reconstructed $m_{D^{*0}}$ from the D^* veto.



(a) $B^{\pm} \longrightarrow \rho^0 e^{\pm} \nu_e$ (b) $B^0 \longrightarrow \rho^{\pm} e^{\mp} \nu_e$



(c) $B^{\pm} \longrightarrow \rho^0 \mu^{\pm} \nu_{\mu}$ (d) $B^0 \longrightarrow \rho^{\pm} \mu^{\mp} \nu_{\mu}$

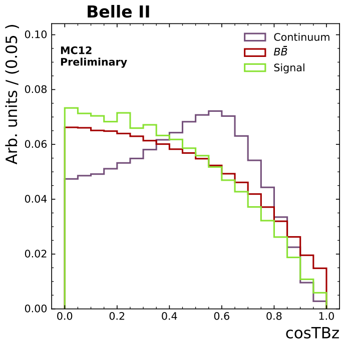
Figure A.7: Reconstructed $m_{D^{*\pm}}$ from the D^* veto.

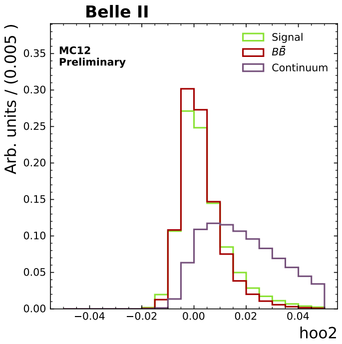
A.7 BDT

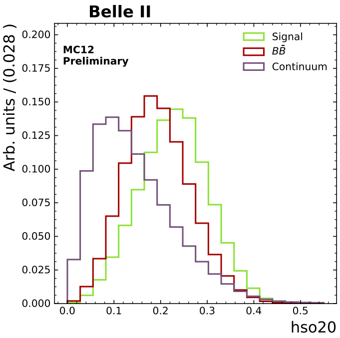
For the full BDT training reports, see
https://drive.google.com/open?id=1A3e_PsOCxXMiYX2yXFjF723Ff4WqiPeS.

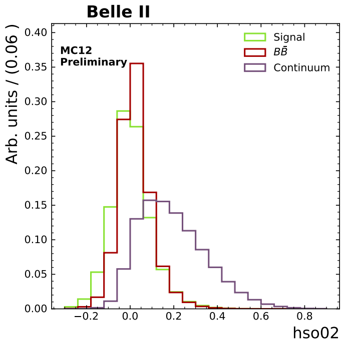
A.7.1 BDT input variables

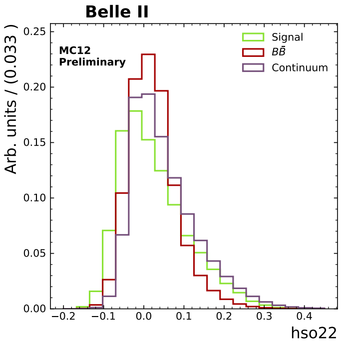
Continuum suppression

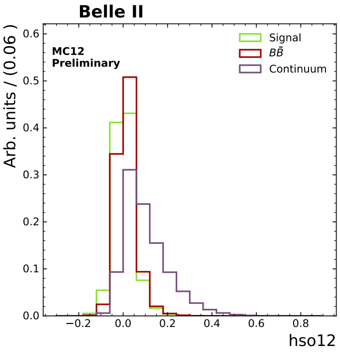


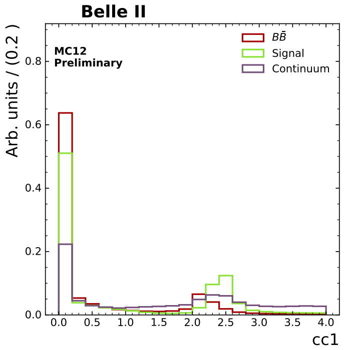


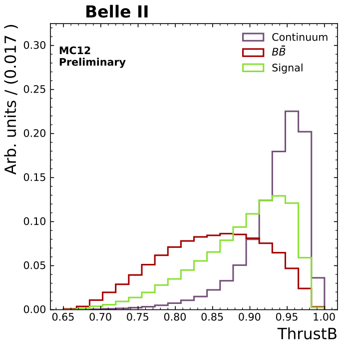


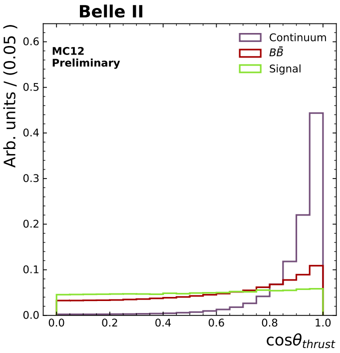












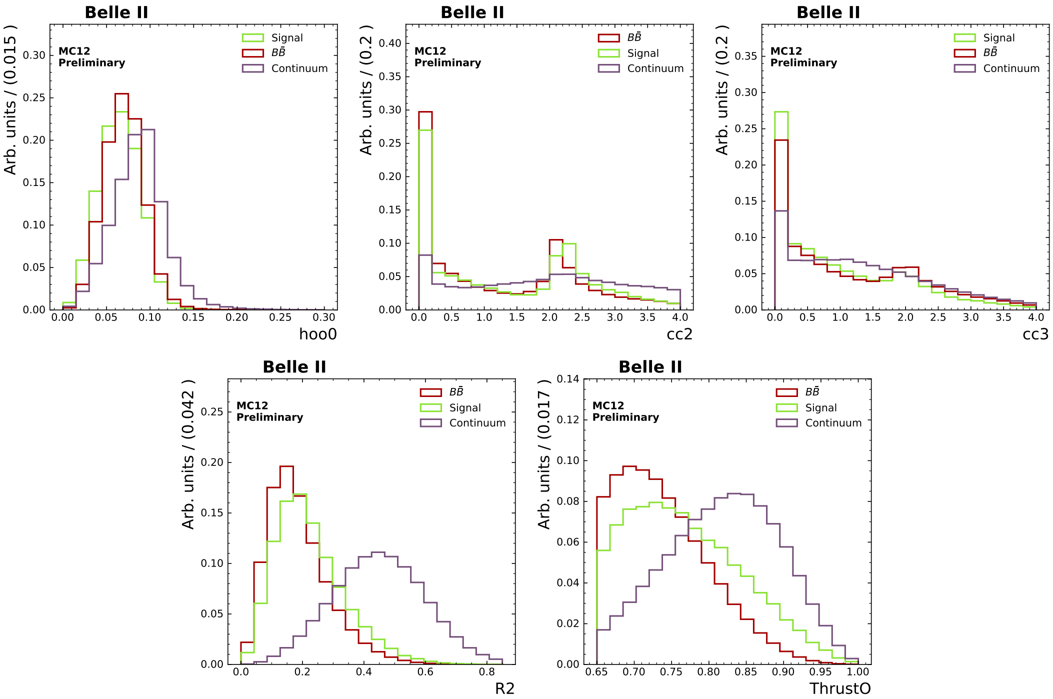
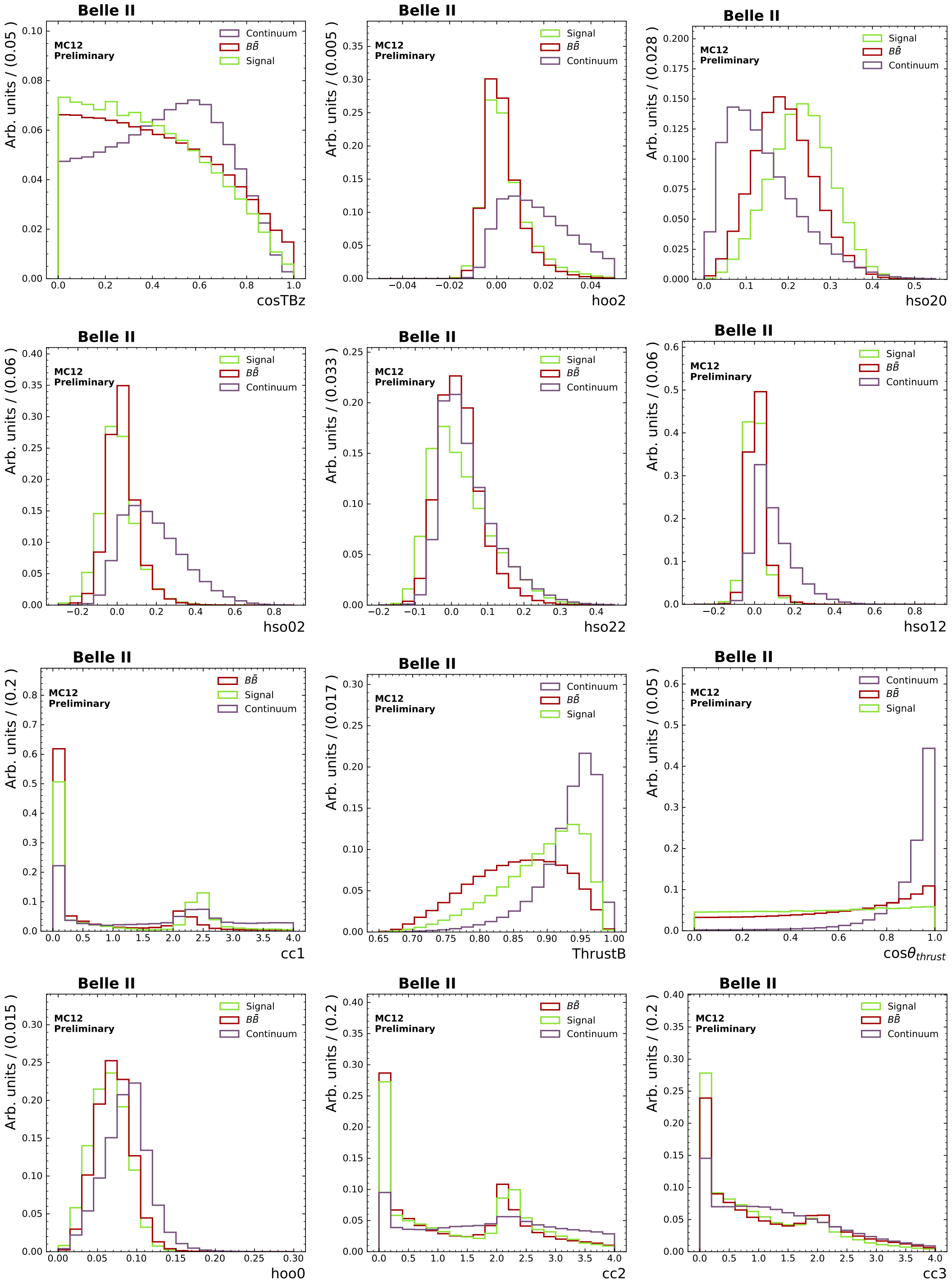


Figure A.8: Normalized shapes for the $B^0 \to \rho^\pm e^\mp \nu_e$ continuum suppression BDT input variables.


Belle II
MC12
Preliminary
Continuum
$B\bar{B}$
Signal
Arb. units / (0.05)
cosTBz
Arb. units / (0.005)
hoo2
Arb. units / (0.028)
hso20
Arb. units / (0.06)
hso02
Arb. units / (0.033)
hso22
Arb. units / (0.06)
hso12
Arb. units / (0.2)
cc1
Arb. units / (0.017)
ThrustB
Arb. units / (0.05)
$\cos\theta_{thrust}$
Arb. units / (0.015)
hoo0
Arb. units / (0.2)
cc2
Arb. units / (0.2)
cc3

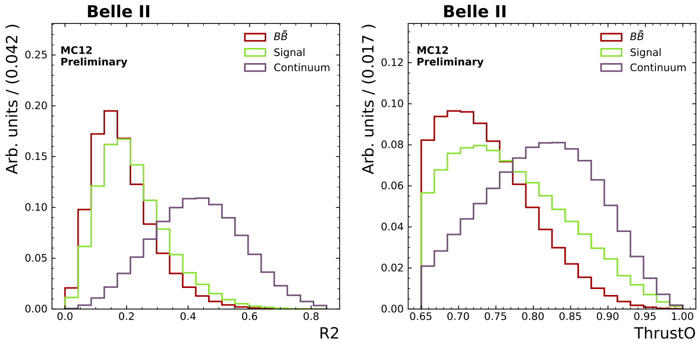
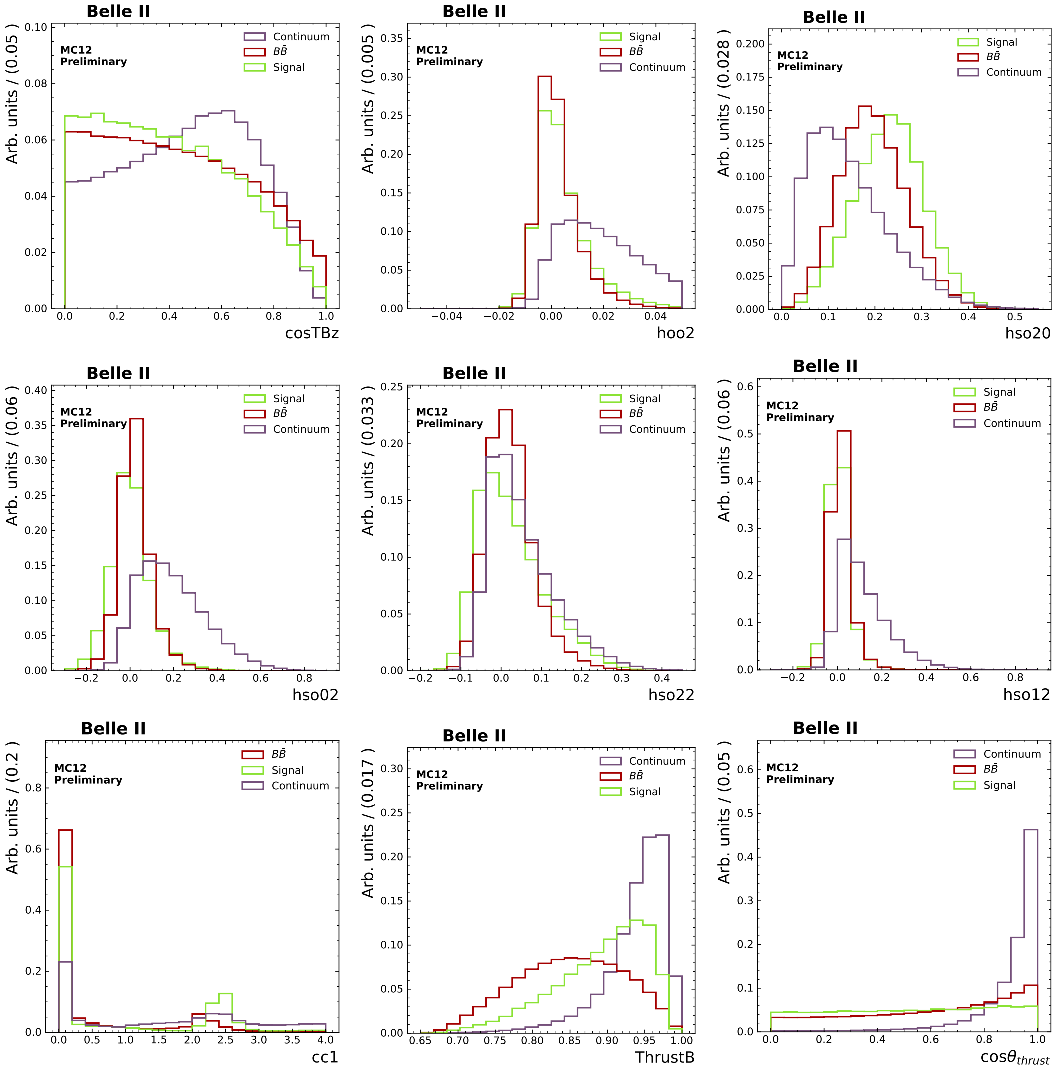


Figure A.9: Normalized shapes for the $B^0 \to \rho^\pm \mu^\mp \nu_\mu$ continuum suppression BDT input variables.



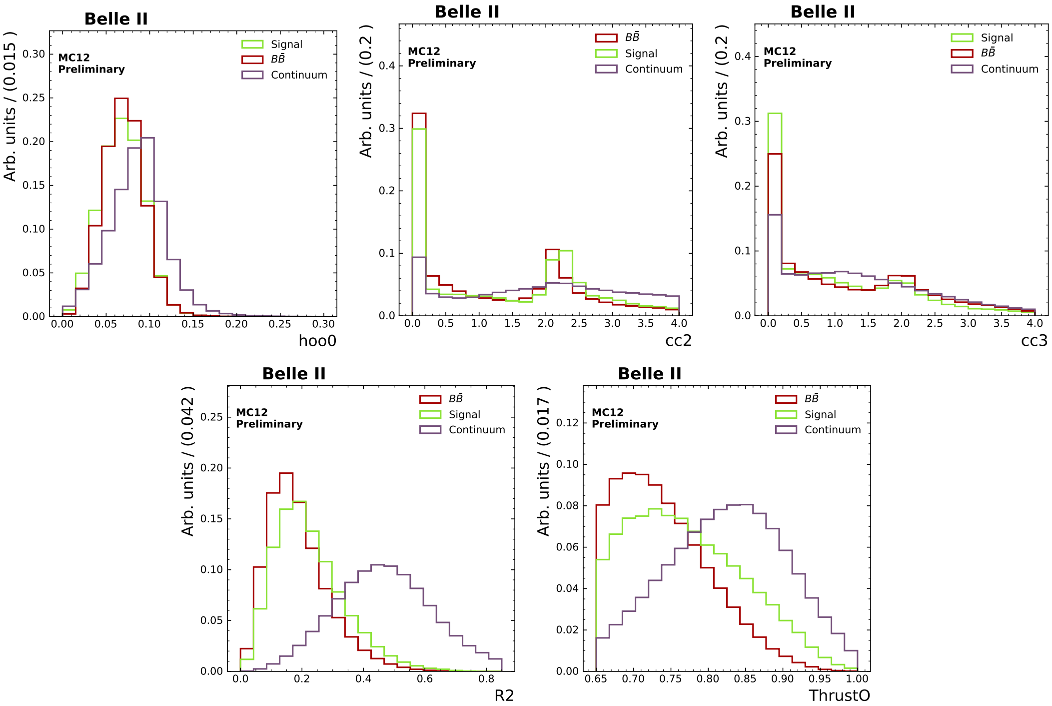
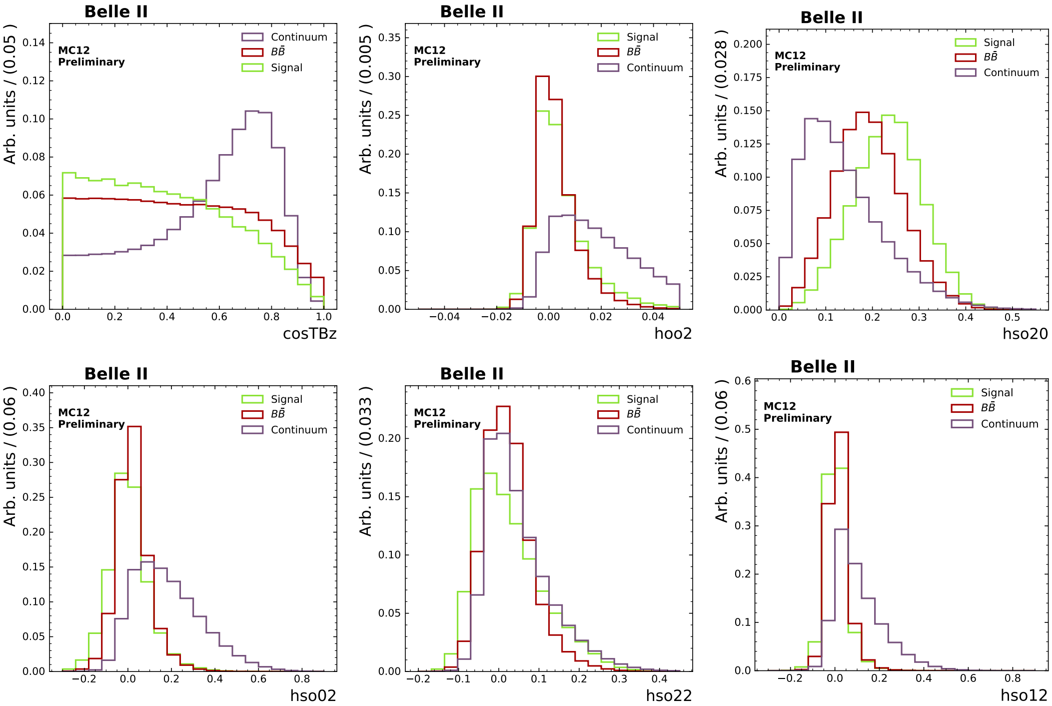


Figure A.10: Normalized shapes for the $B^{\pm} \to \rho^{0} e^{\pm} \nu_{e}$ continuum suppression BDT input variables.



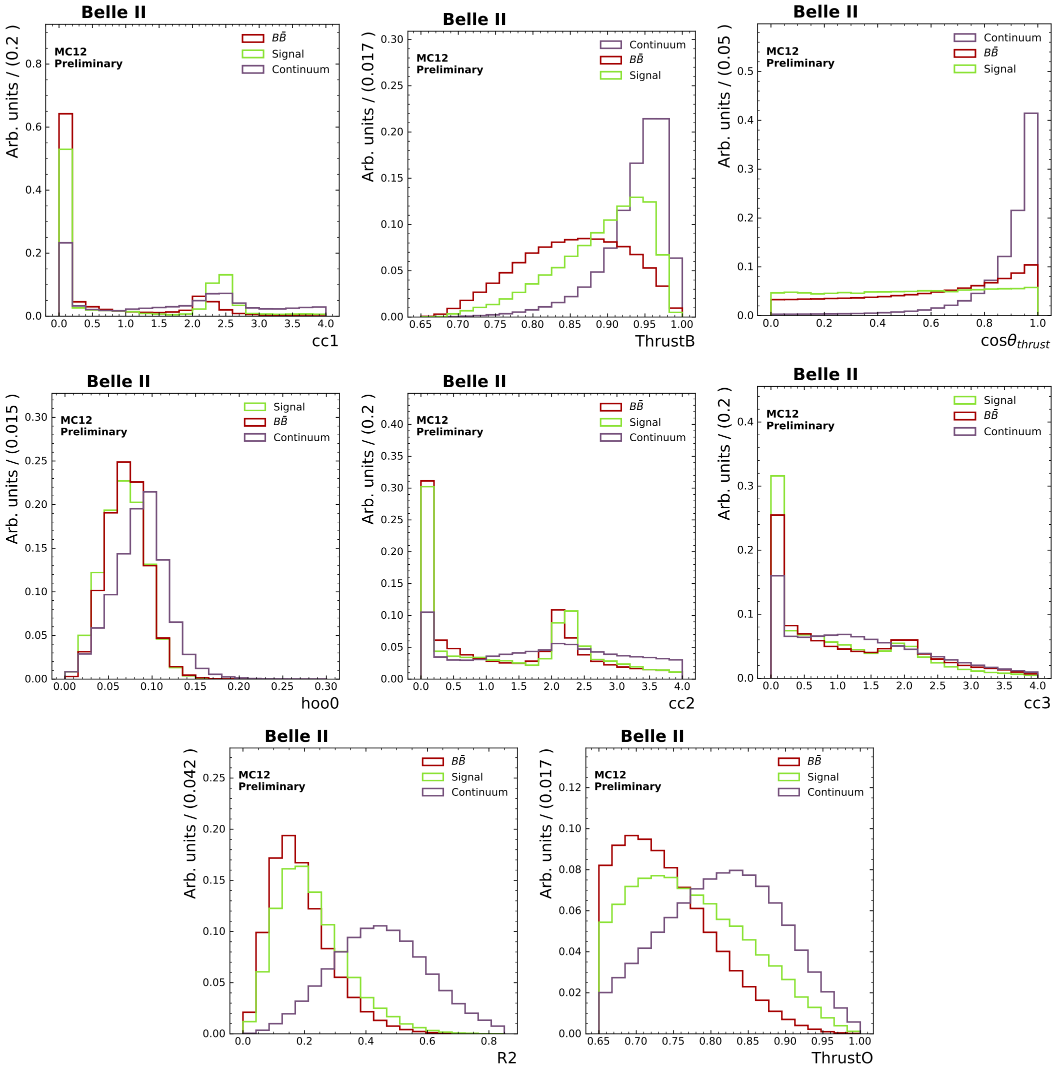


Figure A.11: Normalized shapes for the $B^{\pm} \rightarrow \rho^{0}\mu^{\pm}\nu_{\mu}$ continuum suppression BDT input variables.

$B\overline{B}$ suppression

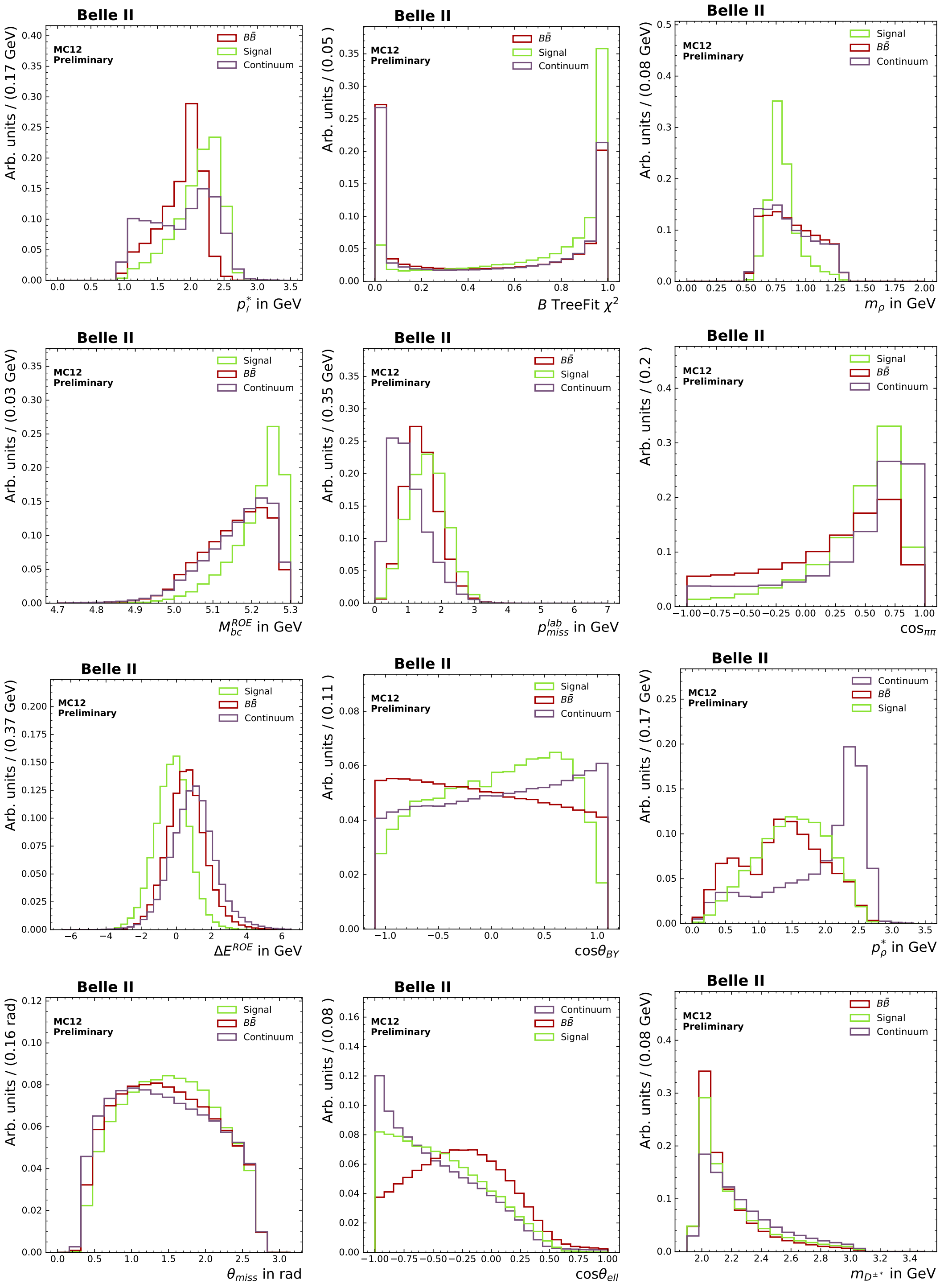


Figure A.12: Normalized shapes for the $B^0 \to \rho^\pm e^\mp \nu_e$ $B\overline{B}$ suppression BDT input variables.

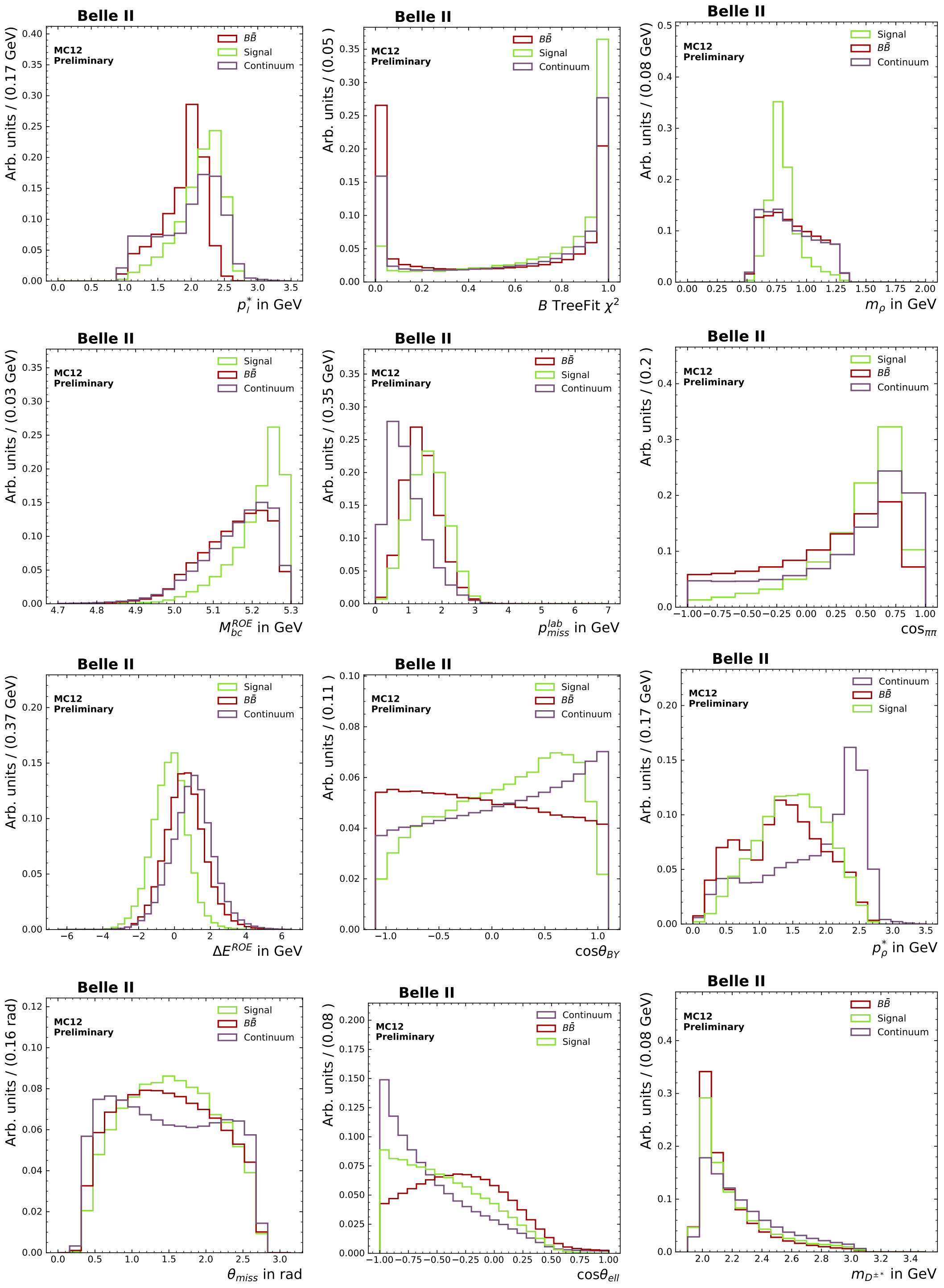


Figure A.13: Normalized shapes for the $B^0 \to \rho^\pm \mu^\mp \nu_\mu$ $B\overline{B}$ suppression BDT input variables.

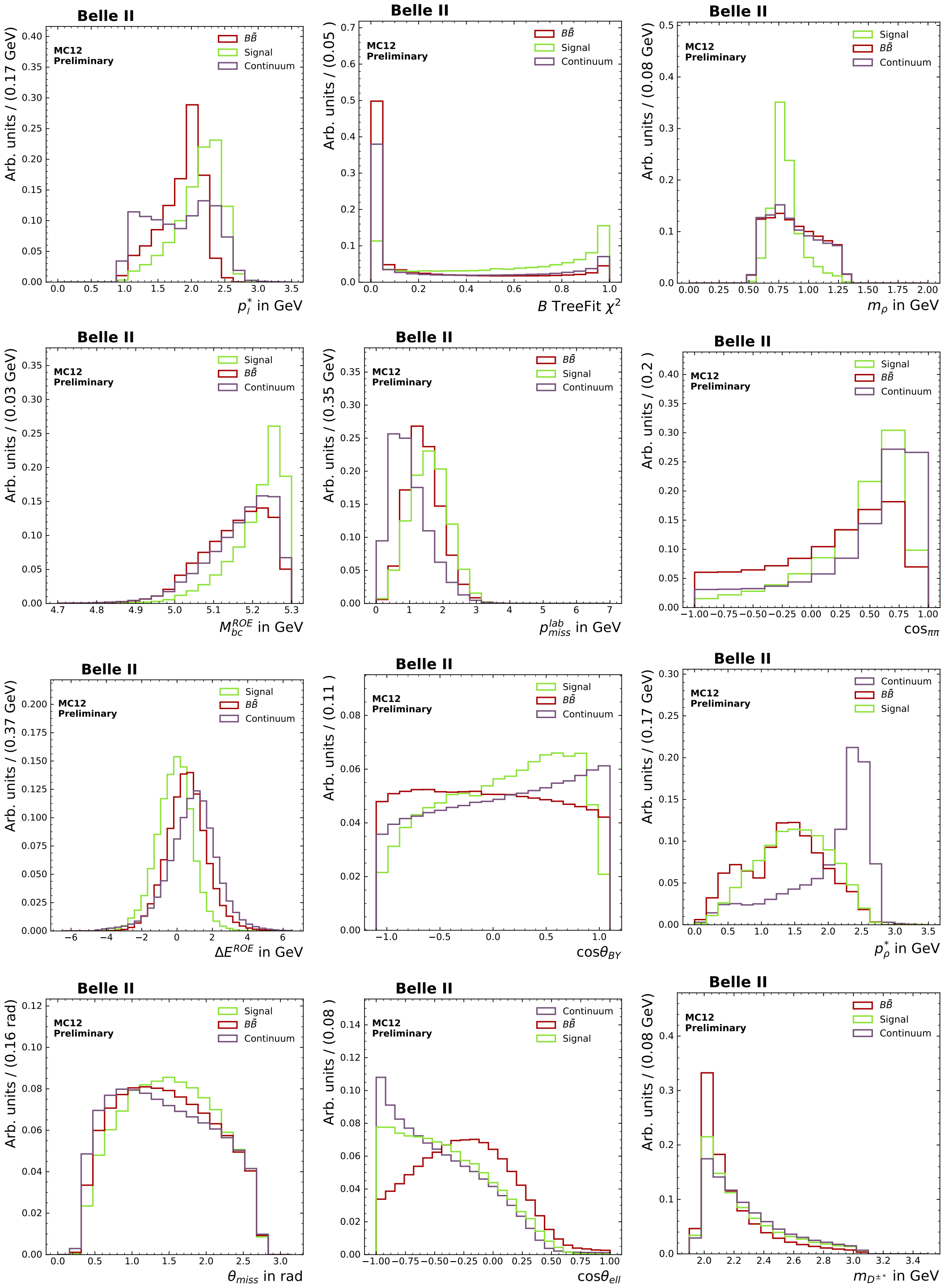


Figure A.14: Normalized shapes for the $B^\pm \to \rho^0 e^\pm \nu_e$ $B\overline{B}$ suppression BDT input variables.

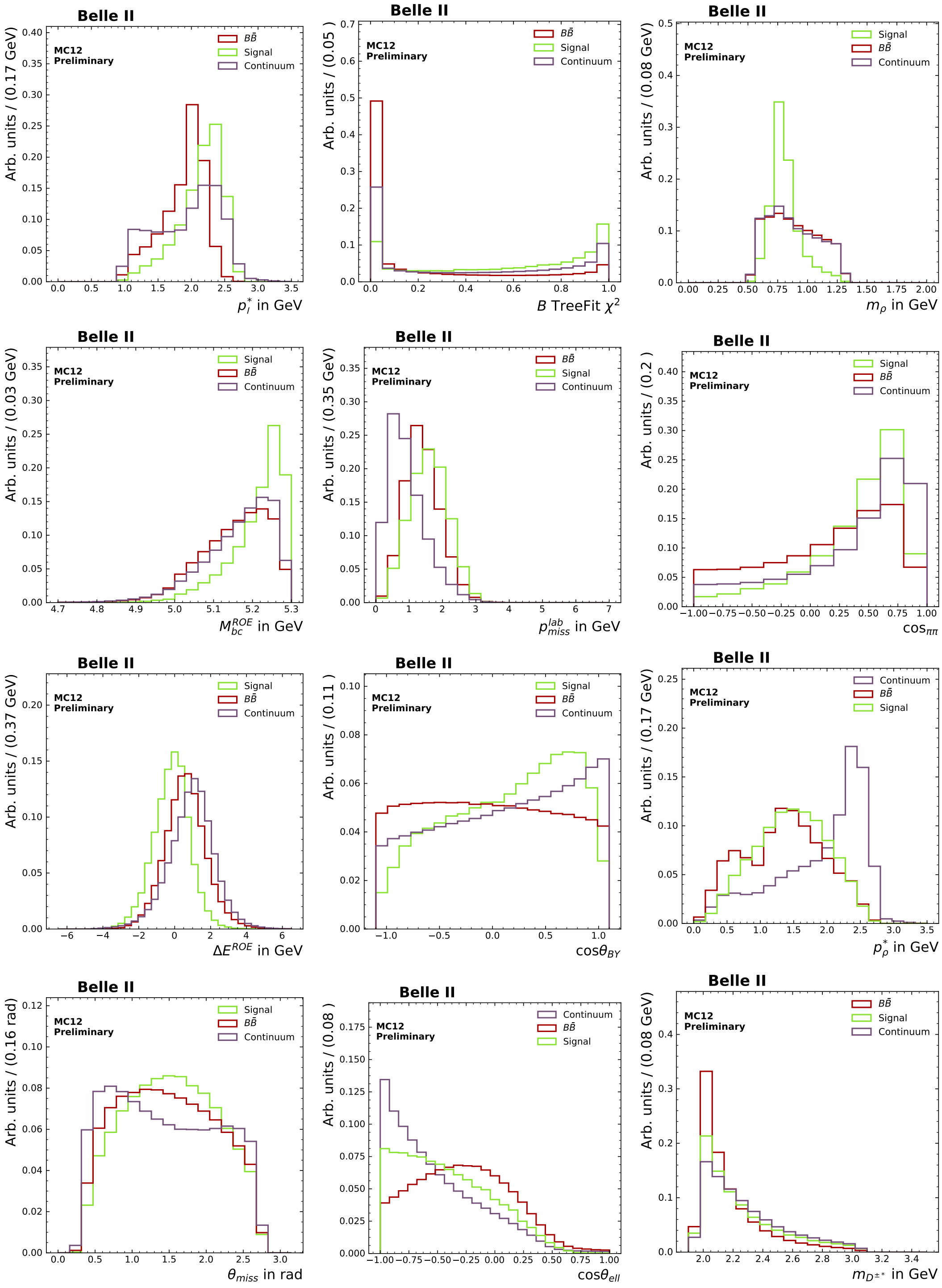


Figure A.15: Normalized shapes for the $B^{\pm} \rightarrow \rho^{0}\mu^{\pm}\nu_{\mu}$ $B\overline{B}$ suppression BDT input variables.

A.7.2 Importance ranking

This section lists all the important ranking for the trained BDTs. Firstly the continuum suppression rankings are shown, secondly the $B\overline{B}$ rankings are listed.

Continuum suppression

$B^0 \rightarrow \rho^{\pm} e^{\mp} \nu_e$: $q^2 < 4.0\,\text{GeV}^2$

1. $\cos\theta^{z}_{T_B}$
2. h^{oo}_{2}
3. h^{so}_{20}
4. h^{so}_{02}
5. h^{so}_{22}
6. h^{so}_{12}
7. $cc1$
8. $|\vec{T}_B|$
9. $\cos\theta^{ROE}_{T_B}$
10. h^{oo}_{0}
11. $cc2$
12. $cc3$
13. R_2
14. $|\vec{T}_{ROE}|$

$B^0 \rightarrow \rho^{\pm} e^{\mp} \nu_e$: $4.0 \leq q^2 < 8.0\,\text{GeV}^2$

1. $\cos\theta^{z}_{T_B}$
2. h^{oo}_{2}
3. h^{so}_{20}
4. h^{so}_{02}
5. h^{so}_{12}
6. h^{so}_{22}
7. h^{oo}_{0}
8. $|\vec{T}_B|$
9. $cc1$
10. $\cos\theta^{ROE}_{T_B}$
11. $cc2$
12. R_2
13. $|\vec{T}_{ROE}|$
14. $cc3$

$B^0 \rightarrow \rho^{\pm} e^{\mp} \nu_e$: $8.0 \leq q^2 < 12.0\,\text{GeV}^2$

1. h^{so}_{20}
2. $\cos\theta^{z}_{T_B}$
3. h^{oo}_{0}
4. h^{oo}_{2}
5. h^{so}_{12}
6. R_2
7. h^{so}_{02}
8. $\cos\theta^{ROE}_{T_B}$
9. $|\vec{T}_B|$
10. $cc1$
11. h^{so}_{22}
12. $cc2$
13. $|\vec{T}_{ROE}|$
14. $cc3$

$B^0 \rightarrow \rho^{\pm} e^{\mp} \nu_e$: $12.0 \leq q^2 < 16.0\,\text{GeV}^2$

1. h^{so}_{20}
2. $\cos\theta^{z}_{T_B}$
3. R_2
4. h^{oo}_{0}
5. h^{oo}_{2}
6. $|\vec{T}_B|$
7. $\cos\theta^{ROE}_{T_B}$
8. h^{so}_{12}
9. $cc1$
10. h^{so}_{02}
11. h^{so}_{22}
12. $cc2$
13. $|\vec{T}_{ROE}|$
14. $cc3$

$B^0 \to \rho^\pm e^\mp \nu_e$: $q^2 \geq 16.0\,\mathrm{GeV}^2$

1. $\cos\theta^{z}_{T_B}$
2. h^{so}_{20}
3. R_2
4. $|\vec{T}_B|$
5. h^{oo}_{0}
6. h^{oo}_{2}
7. $\cos\theta^{ROE}_{T_B}$
8. h^{so}_{12}
9. $cc1$
10. h^{so}_{22}
11. h^{so}_{02}
12. $|\vec{T}_{ROE}|$
13. $cc2$
14. $cc3$

$B^0 \to \rho^\pm \mu^\mp \nu_\mu$: $q^2 < 4.0\,\mathrm{GeV}^2$

1. $\cos\theta^{z}_{T_B}$
2. h^{oo}_{2}
3. $\cos\theta^{ROE}_{T_B}$
4. h^{so}_{20}
5. h^{so}_{12}
6. h^{so}_{02}
7. h^{oo}_{0}
8. h^{so}_{22}
9. $cc1$
10. $|\vec{T}_B|$
11. $cc3$
12. $|\vec{T}_{ROE}|$
13. R_2
14. $cc2$

$B^0 \to \rho^\pm \mu^\mp \nu_\mu$: $4.0 \leq q^2 < 8.0\,\mathrm{GeV}^2$

1. $\cos\theta^{z}_{T_B}$
2. h^{so}_{20}
3. h^{oo}_{2}
4. $\cos\theta^{ROE}_{T_B}$
5. h^{oo}_{0}
6. h^{so}_{22}
7. h^{so}_{12}
8. h^{so}_{02}
9. $|\vec{T}_B|$
10. $cc1$
11. $cc2$
12. $|\vec{T}_{ROE}|$
13. $cc3$
14. R_2

$B^0 \to \rho^\pm \mu^\mp \nu_\mu$: $8.0 \leq q^2 < 12.0\,\mathrm{GeV}^2$

1. h^{so}_{20}
2. $\cos\theta^{z}_{T_B}$
3. $\cos\theta^{ROE}_{T_B}$
4. h^{oo}_{0}
5. R_2
6. h^{so}_{12}
7. h^{oo}_{2}
8. h^{so}_{02}
9. h^{so}_{22}
10. $|\vec{T}_B|$
11. $cc1$
12. $cc2$
13. $|\vec{T}_{ROE}|$
14. $cc3$

$B^0 \to \rho^\pm \mu^\mp \nu_\mu$: $12.0 \leq q^2 < 16.0\,\mathrm{GeV}^2$

1. h^{so}_{20}
2. $\cos\theta^{z}_{T_B}$
3. $\cos\theta^{ROE}_{T_B}$
4. R_2
5. h^{oo}_{0}
6. h^{oo}_{2}
7. $|\vec{T}_B|$
8. h^{so}_{12}
9. $cc1$
10. h^{so}_{02}
11. h^{so}_{22}
12. $|\vec{T}_{ROE}|$
13. $cc2$
14. $cc3$

$B^0 \to \rho^\pm \mu^\mp \nu_\mu$: $q^2 \geq 16.0\ \mathrm{GeV}^2$

1. $\cos\theta^{z}_{T_B}$
2. h^{so}_{20}
3. $\cos\theta^{ROE}_{T_B}$
4. $|\vec{T}_B|$
5. R_2
6. h^{oo}_2
7. h^{oo}_0
8. h^{so}_{12}
9. $cc1$
10. h^{so}_{22}
11. h^{so}_{02}
12. $|\vec{T}_{ROE}|$
13. $cc2$
14. $cc3$

$B^\pm \to \rho^0 e^\pm \nu_e$: $q^2 < 4.0\ \mathrm{GeV}^2$

1. $\cos\theta^{z}_{T_B}$
2. h^{oo}_2
3. h^{so}_{20}
4. h^{so}_{02}
5. h^{so}_{22}
6. h^{so}_{12}
7. $cc1$
8. $|\vec{T}_B|$
9. $\cos\theta^{ROE}_{T_B}$
10. h^{oo}_0
11. $cc2$
12. $cc3$
13. R_2
14. $|\vec{T}_{ROE}|$

$B^\pm \to \rho^0 e^\pm \nu_e$: $4.0 \leq q^2 < 8.0\ \mathrm{GeV}^2$

1. $\cos\theta^{z}_{T_B}$
2. h^{so}_{20}
3. h^{oo}_2
4. h^{so}_{02}
5. h^{so}_{22}
6. h^{so}_{12}
7. $|\vec{T}_B|$
8. h^{oo}_0
9. $cc1$
10. $cc2$
11. $\cos\theta^{ROE}_{T_B}$
12. R_2
13. $|\vec{T}_{ROE}|$
14. $cc3$

$B^\pm \to \rho^0 e^\pm \nu_e$: $8.0 \leq q^2 < 12.0\ \mathrm{GeV}^2$

1. h^{so}_{20}
2. $\cos\theta^{z}_{T_B}$
3. h^{oo}_2
4. h^{oo}_0
5. R_2
6. h^{so}_{02}
7. h^{so}_{12}
8. $|\vec{T}_B|$
9. $cc1$
10. $\cos\theta^{ROE}_{T_B}$
11. h^{so}_{22}
12. $cc2$
13. $|\vec{T}_{ROE}|$
14. $cc3$

$B^\pm \to \rho^0 e^\pm \nu_e$: $12.0 \leq q^2 < 16.0\ \mathrm{GeV}^2$

1. h^{so}_{20}
2. $\cos\theta^{z}_{T_B}$
3. R_2
4. h^{oo}_0
5. h^{oo}_2
6. $|\vec{T}_B|$
7. $\cos\theta^{ROE}_{T_B}$
8. $cc1$
9. h^{so}_{12}
10. h^{so}_{02}
11. $cc2$
12. h^{so}_{22}
13. $|\vec{T}_{ROE}|$
14. $cc3$

$B^{\pm} \to \rho^{0} e^{\pm} \nu_{e}$: $q^{2} \geq 16.0\ \mathrm{GeV}^{2}$

1. $\cos\theta^{z}_{T_B}$
2. h^{so}_{20}
3. R_2
4. h^{oo}_{0}
5. $|\vec{T}_B|$
6. h^{oo}_{2}
7. $\cos\theta^{ROE}_{T_B}$
8. $cc1$
9. h^{so}_{02}
10. h^{so}_{22}
11. $cc2$
12. h^{so}_{12}
13. $|\vec{T}_{ROE}|$
14. $cc3$

$B^{\pm} \to \rho^{0} \mu^{\pm} \nu_{\mu}$: $q^{2} < 4.0\ \mathrm{GeV}^{2}$

1. $\cos\theta^{z}_{T_B}$
2. h^{so}_{20}
3. h^{oo}_{2}
4. $\cos\theta^{ROE}_{T_B}$
5. h^{so}_{02}
6. h^{so}_{22}
7. h^{so}_{12}
8. $cc1$
9. $|\vec{T}_B|$
10. h^{oo}_{0}
11. $cc3$
12. $cc2$
13. $|\vec{T}_{ROE}|$
14. R_2

$B^{\pm} \to \rho^{0} \mu^{\pm} \nu_{\mu}$: $4.0 < q^{2} < 8.0\ \mathrm{GeV}^{2}$

1. $\cos\theta^{z}_{T_B}$
2. h^{so}_{20}
3. h^{so}_{22}
4. h^{oo}_{2}
5. $\cos\theta^{ROE}_{T_B}$
6. h^{so}_{02}
7. h^{oo}_{0}
8. h^{so}_{12}
9. $|\vec{T}_B|$
10. $cc1$
11. $cc2$
12. $|\vec{T}_{ROE}|$
13. R_2
14. $cc3$

$B^{\pm} \to \rho^{0} \mu^{\pm} \nu_{\mu}$: $8.0 < q^{2} < 12.0\ \mathrm{GeV}^{2}$

1. h^{so}_{20}
2. $\cos\theta^{z}_{T_B}$
3. $\cos\theta^{ROE}_{T_B}$
4. h^{oo}_{0}
5. R_2
6. h^{so}_{02}
7. h^{oo}_{2}
8. h^{so}_{22}
9. h^{so}_{12}
10. $cc1$
11. $|\vec{T}_B|$
12. $cc2$
13. $|\vec{T}_{ROE}|$
14. $cc3$

$B^{\pm} \to \rho^{0} \mu^{\pm} \nu_{\mu}$: $12.0 < q^{2} < 16.0\ \mathrm{GeV}^{2}$

1. h^{so}_{20}
2. $\cos\theta^{z}_{T_B}$
3. $\cos\theta^{ROE}_{T_B}$
4. R_2
5. h^{oo}_{0}
6. h^{oo}_{2}
7. $|\vec{T}_B|$
8. $cc1$
9. h^{so}_{12}
10. h^{so}_{02}
11. h^{so}_{22}
12. $|\vec{T}_{ROE}|$
13. $cc2$
14. $cc3$

$B^{\pm} \rightarrow \rho^{0} \mu^{\pm} \nu_{\mu}$: $q^2 \geq 16.0\ \text{GeV}^2$

1. $\cos\theta^{z}_{T_B}$
2. h^{so}_{20}
3. $\cos\theta^{ROE}_{T_B}$
4. R_2
5. $|\vec{T}_B|$
6. h^{oo}_{2}
7. h^{oo}_{0}
8. $cc1$
9. $|\vec{T}_{ROE}|$
10. h^{so}_{12}
11. h^{so}_{02}
12. h^{so}_{22}
13. $cc2$
14. $cc3$

$B\overline{B}$ suppression

$B^{0} \rightarrow \rho^{\pm} e^{\mp} \nu_{e}$: $q^2 < 4.0\ \text{GeV}^2$

1. p^{*}_{l}
2. B vertex χ^2 prob.
3. m_{ρ}
4. M^{ROE}_{bc}
5. $\cos\theta_{\pi\pi}$
6. p^{lab}_{miss}
7. ΔE^{ROE}
8. $\cos\theta_{BY}$
9. p^{*}_{ρ}
10. θ_{miss}
11. $\cos\theta_{l}$
12. D^{*} veto

$B^{0} \rightarrow \rho^{\pm} e^{\mp} \nu_{e}$: $4.0 \leq q^2 < 8.0\ \text{GeV}^2$

1. p^{*}_{l}
2. B vertex χ^2 prob.
3. m_{ρ}
4. M^{ROE}_{bc}
5. p^{lab}_{miss}
6. ΔE^{ROE}
7. $\cos\theta_{\pi\pi}$
8. $\cos\theta_{BY}$
9. $\cos\theta_{l}$
10. D^{*} veto
11. p^{*}_{ρ}
12. θ_{miss}

$B^{0} \rightarrow \rho^{\pm} e^{\mp} \nu_{e}$: $8.0 \leq q^2 < 12.0\ \text{GeV}^2$

1. p^{*}_{l}
2. B vertex χ^2 prob.
3. p^{lab}_{miss}
4. ΔE^{ROE}
5. m_{ρ}
6. M^{ROE}_{bc}
7. $\cos\theta_{\pi\pi}$
8. $\cos\theta_{BY}$
9. D^{*} veto
10. p^{*}_{ρ}
11. $\cos\theta_{l}$
12. θ_{miss}

$B^{0} \rightarrow \rho^{\pm} e^{\mp} \nu_{e}$: $12.0 \leq q^2 < 16.0\ \text{GeV}^2$

1. p^{*}_{l}
2. p^{lab}_{miss}
3. B vertex χ^2 prob.
4. ΔE^{ROE}
5. m_{ρ}
6. M^{ROE}_{bc}
7. $\cos\theta_{\pi\pi}$
8. $\cos\theta_{BY}$
9. p^{*}_{ρ}
10. D^{*} veto
11. θ_{miss}
12. $\cos\theta_{l}$

$B^{0} \rightarrow \rho^{\pm} e^{\mp} \nu_{e}$: $q^2 \geq 16.0\ \text{GeV}^2$

1. p^{*}_{l}
2. ΔE^{ROE}
3. p^{lab}_{miss}
4. B vertex χ^2 prob.
5. m_{ρ}
6. M^{ROE}_{bc}
7. p^{*}_{ρ}
8. $\cos\theta_{BY}$
9. $\cos\theta_{l}$
10. D^{*} veto
11. θ_{miss}
12. $\cos\theta_{\pi\pi}$

$B^0 \to \rho^{\pm}\mu^{\mp}\nu_{\mu}$: $q^2 < 4.0\ \mathrm{GeV}^2$

1. p_l^*
2. B vertex χ^2 prob.
3. m_ρ
4. ΔE^{ROE}
5. $\cos\theta_{\pi\pi}$
6. M_{bc}^{ROE}
7. p_{miss}^{lab}
8. $\cos\theta_{BY}$
9. $\cos\theta_l$
10. p_ρ^*
11. D^* veto
12. θ_{miss}

$B^0 \to \rho^{\pm}\mu^{\mp}\nu_{\mu}$: $4.0 \leq q^2 < 8.0\ \mathrm{GeV}^2$

1. p_l^*
2. B vertex χ^2 prob.
3. m_ρ
4. M_{bc}^{ROE}
5. ΔE^{ROE}
6. p_{miss}^{lab}
7. $\cos\theta_{\pi\pi}$
8. $\cos\theta_{BY}$
9. $\cos\theta_l$
10. D^* veto
11. θ_{miss}
12. p_ρ^*

$B^0 \to \rho^{\pm}\mu^{\mp}\nu_{\mu}$: $8.0 \leq q^2 < 12.0\ \mathrm{GeV}^2$

1. p_l^*
2. B vertex χ^2 prob.
3. p_{miss}^{lab}
4. ΔE^{ROE}
5. M_{bc}^{ROE}
6. m_ρ
7. $\cos\theta_{\pi\pi}$
8. $\cos\theta_{BY}$
9. p_ρ^*
10. D^* veto
11. $\cos\theta_l$
12. θ_{miss}

$B^0 \to \rho^{\pm}\mu^{\mp}\nu_{\mu}$: $12.0 \leq q^2 < 16.0\ \mathrm{GeV}^2$

1. p_l^*
2. p_{miss}^{lab}
3. ΔE^{ROE}
4. B vertex χ^2 prob.
5. M_{bc}^{ROE}
6. m_ρ
7. $\cos\theta_{\pi\pi}$
8. $\cos\theta_{BY}$
9. p_ρ^*
10. D^* veto
11. θ_{miss}
12. $\cos\theta_l$

$B^0 \to \rho^{\pm}\mu^{\mp}\nu_{\mu}$: $q^2 \geq 16.0\ \mathrm{GeV}^2$

1. p_l^*
2. p_{miss}^{lab}
3. ΔE^{ROE}
4. B vertex χ^2 prob.
5. m_ρ
6. M_{bc}^{ROE}
7. p_ρ^*
8. $\cos\theta_{BY}$
9. $\cos\theta_l$
10. $\cos\theta_{\pi\pi}$
11. θ_{miss}
12. D^* veto

$B^{\pm} \rightarrow \rho^{0} e^{\pm} \nu_e$: $q^2 < 4.0\,\mathrm{GeV}^2$

1. p_l^*
2. m_ρ
3. M_{bc}^{ROE}
4. B vertex χ^2 prob.
5. p_{miss}^{lab}
6. ΔE^{ROE}
7. $\cos\theta_{\pi\pi}$
8. D^* veto
9. $\cos\theta_{BY}$
10. $\cos\theta_l$
11. p_ρ^*
12. θ_{miss}

$B^{\pm} \rightarrow \rho^{0} e^{\pm} \nu_e$: $4.0 \leq q^2 < 8.0\,\mathrm{GeV}^2$

1. p_l^*
2. m_ρ
3. M_{bc}^{ROE}
4. p_{miss}^{lab}
5. B vertex χ^2 prob.
6. ΔE^{ROE}
7. D^* veto
8. $\cos\theta_{\pi\pi}$
9. $\cos\theta_{BY}$
10. $\cos\theta_l$
11. p_ρ^*
12. θ_{miss}

$B^{\pm} \rightarrow \rho^{0} e^{\pm} \nu_e$: $8.0 \leq q^2 < 12.0\,\mathrm{GeV}^2$

1. p_l^*
2. p_{miss}^{lab}
3. m_ρ
4. M_{bc}^{ROE}
5. B vertex χ^2 prob.
6. ΔE^{ROE}
7. D^* veto
8. $\cos\theta_{\pi\pi}$
9. p_ρ^*
10. $\cos\theta_{BY}$
11. $\cos\theta_l$
12. θ_{miss}

$B^{\pm} \rightarrow \rho^{0} e^{\pm} \nu_e$: $12.0 \leq q^2 < 16.0\,\mathrm{GeV}^2$

1. p_l^*
2. p_{miss}^{lab}
3. m_ρ
4. ΔE^{ROE}
5. M_{bc}^{ROE}
6. B vertex χ^2 prob.
7. D^* veto
8. p_ρ^*
9. $\cos\theta_{\pi\pi}$
10. $\cos\theta_{BY}$
11. $\cos\theta_l$
12. θ_{miss}

$B^{\pm} \rightarrow \rho^{0} e^{\pm} \nu_e$: $q^2 \geq 16.0\,\mathrm{GeV}^2$

1. p_l^*
2. p_{miss}^{lab}
3. ΔE^{ROE}
4. m_ρ
5. M_{bc}^{ROE}
6. B vertex χ^2 prob.
7. p_ρ^*
8. D^* veto
9. $\cos\theta_{BY}$
10. $\cos\theta_l$
11. θ_{miss}
12. $\cos\theta_{\pi\pi}$

$B^\pm \to \rho^0 \mu^\pm \nu_\mu$: $q^2 < 4.0\ \mathrm{GeV}^2$

1. p_l^*
2. m_ρ
3. p_{miss}^{lab}
4. M_{bc}^{ROE}
5. B vertex χ^2 prob.
6. ΔE^{ROE}
7. $\cos\theta_{BY}$
8. $\cos\theta_{\pi\pi}$
9. D^* veto
10. $\cos\theta_l$
11. p_ρ^*
12. θ_{miss}

$B^\pm \to \rho^0 \mu^\pm \nu_\mu$: $4.0 \le q^2 < 8.0\ \mathrm{GeV}^2$

1. p_l^*
2. m_ρ
3. p_{miss}^{lab}
4. M_{bc}^{ROE}
5. B vertex χ^2 prob.
6. ΔE^{ROE}
7. $\cos\theta_{\pi\pi}$
8. D^* veto
9. $\cos\theta_{BY}$
10. $\cos\theta_l$
11. p_ρ^*
12. θ_{miss}

$B^\pm \to \rho^0 \mu^\pm \nu_\mu$: $8.0 \le q^2 < 12.0\ \mathrm{GeV}^2$

1. p_l^*
2. p_{miss}^{lab}
3. m_ρ
4. M_{bc}^{ROE}
5. ΔE^{ROE}
6. B vertex χ^2 prob.
7. D^* veto
8. $\cos\theta_{\pi\pi}$
9. $\cos\theta_{BY}$
10. p_ρ^*
11. $\cos\theta_l$
12. θ_{miss}

$B^\pm \to \rho^0 \mu^\pm \nu_\mu$: $12.0 \le q^2 < 16.0\ \mathrm{GeV}^2$

1. p_l^*
2. p_{miss}^{lab}
3. ΔE^{ROE}
4. m_ρ
5. M_{bc}^{ROE}
6. B vertex χ^2 prob.
7. D^* veto
8. $\cos\theta_{\pi\pi}$
9. p_ρ^*
10. $\cos\theta_{BY}$
11. $\cos\theta_l$
12. θ_{miss}

$B^\pm \to \rho^0 \mu^\pm \nu_\mu$: $q^2 \ge 16.0\ \mathrm{GeV}^2$

1. p_l^*
2. p_{miss}^{lab}
3. ΔE^{ROE}
4. m_ρ
5. M_{bc}^{ROE}
6. p_ρ^*
7. B vertex χ^2 prob.
8. D^* veto
9. θ_{miss}
10. $\cos\theta_{BY}$
11. $\cos\theta_l$
12. $\cos\theta_{\pi\pi}$

A.7.3 Overtraining tests

	$q1$	$q2$	$q3$	$q4$	$q5$
$e^{\pm}$ signal	0.52	0.62	0.41	0.09	0.20
$e^{\pm}$ background	0.26	0.63	0.93	0.66	0.04
$\mu^{\pm}$ signal	0.72	0.48	0.23	0.12	0.01
$\mu^{\pm}$ background	0.37	0.25	0.68	0.83	0.66

Table A.5: P-tests for difference between training and testing performance of the trained B^0 continuum suppression BDTs. No overtraining is present.

	$q1$	$q2$	$q3$	$q4$	$q5$
$e^{\pm}$ signal	0.22	0.85	0.28	0.28	0.06
$e^{\pm}$ background	0.69	0.53	0.41	0.32	0.08
$\mu^{\pm}$ signal	0.56	0.27	0.59	0.87	0.67
$\mu^{\pm}$ background	0.44	0.45	0.83	0.72	0.59

Table A.6: P-tests for difference between training and testing performance of the trained $B^{\pm}$ continuum suppression BDTs. No overtraining is present.

	$q1$	$q2$	$q3$	$q4$	$q5$
$e^{\pm}$ signal	0.24	0.04	0.31	0.53	0.01
$e^{\pm}$ background	0.33	0.39	0.94	0.19	0.40
$\mu^{\pm}$ signal	0.72	0.24	0.70	0.84	0.30
$\mu^{\pm}$ background	0.07	0.70	0.07	0.50	0.70

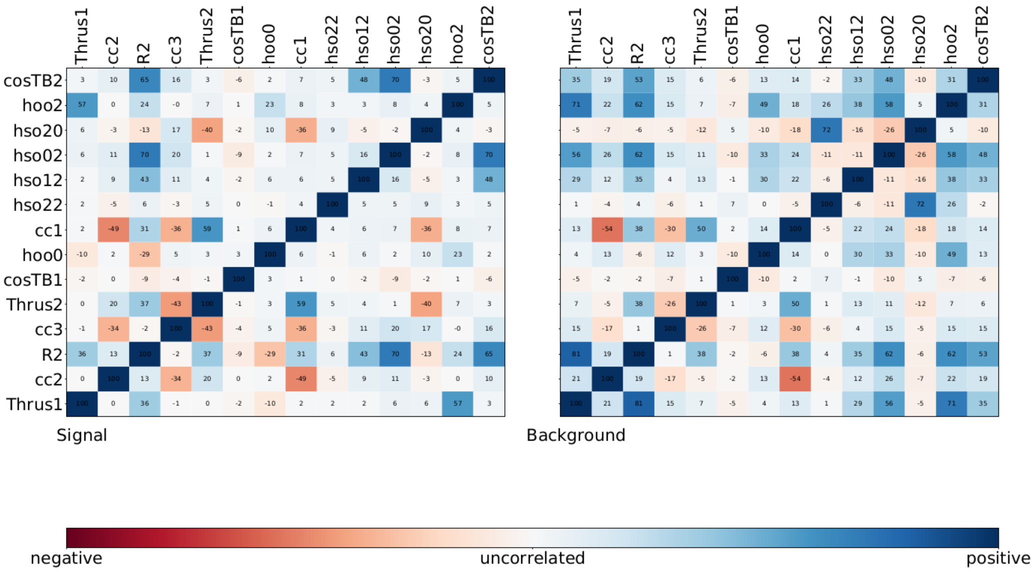
Table A.7: P-tests for difference between training and testing performance of the trained B^0 $B\overline{B}$ suppression BDTs. No overtraining is present.

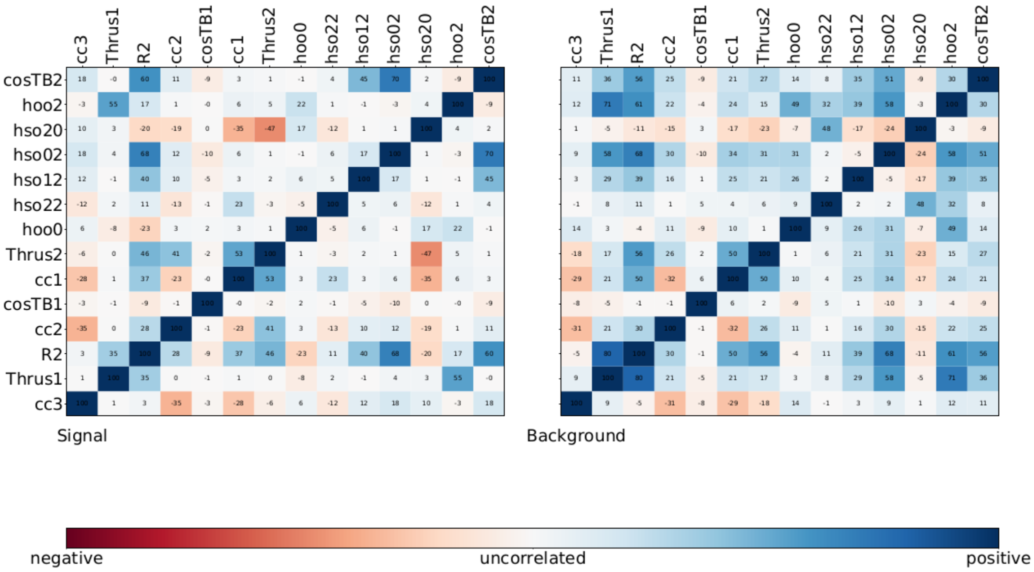
	$q1$	$q2$	$q3$	$q4$	$q5$
$e^{\pm}$ signal	0.01	0.47	0.74	0.03	0.29
$e^{\pm}$ background	0.67	0.96	0.72	0.11	0.62
$\mu^{\pm}$ signal	0.09	0.43	0.94	0.81	0.04
$\mu^{\pm}$ background	0.01	0.16	0.46	0.58	0.38

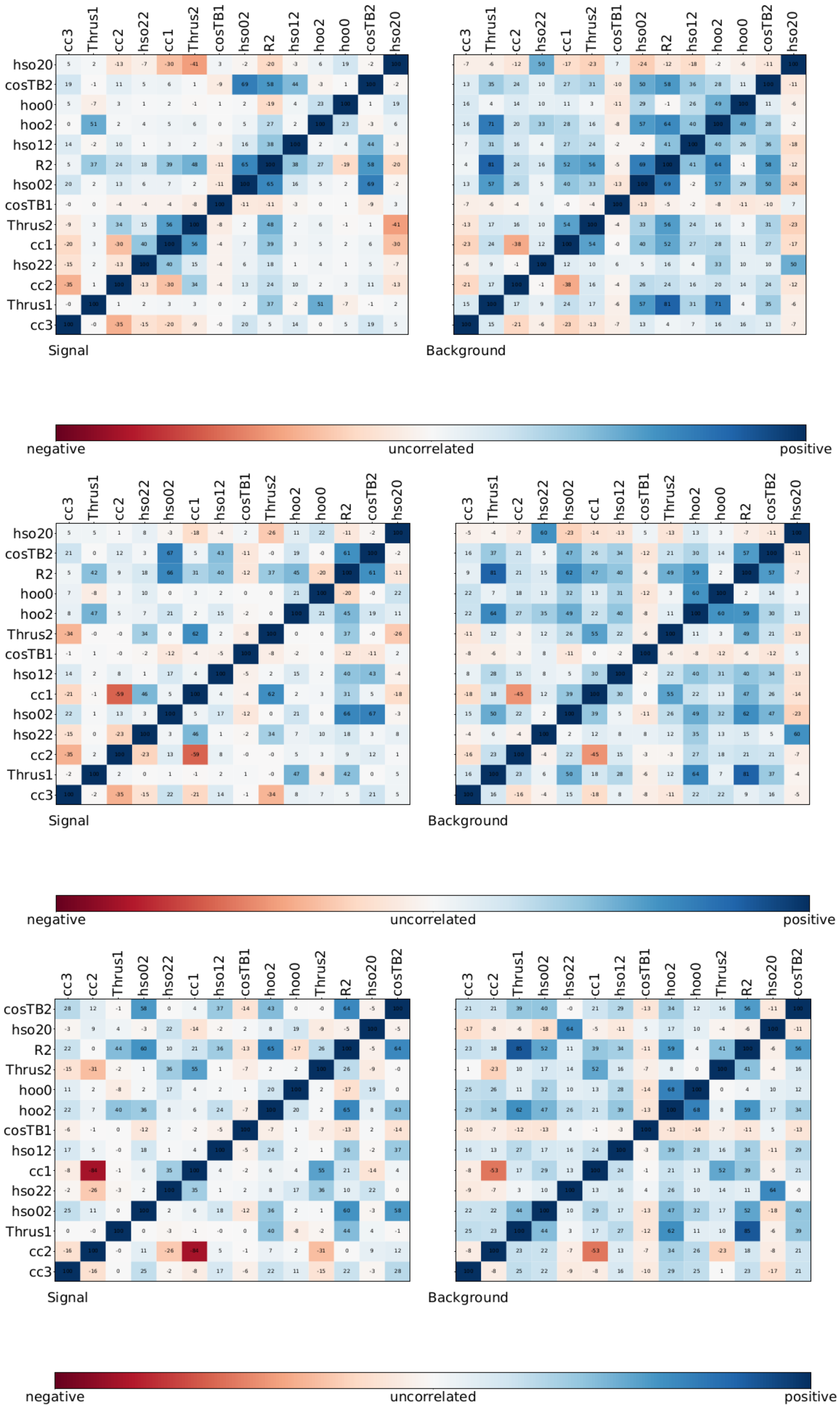
Table A.8: P-tests for difference between training and testing performance of the trained $B^{\pm}$ $B\overline{B}$ suppression BDTs. No overtraining is present.

A.7.4 Correlation matrices

The next few pages show the correlation matrices for all the trained continuum suppression and $B\overline{B}$ suppression BDTs. The correlations are relatively constant in the different q^2 bins per signal decay channel. First the continuum suppression matrices will be shown, afterwards the $B\overline{B}$ matrices.







Signal

	cc3	Thrus1	cc2	hso22	cc1	Thrus2	cosTB1	hso02	R2	hso12	hoo2	hoo0	cosTB2	hso20
hso20	5	2	-13	-7	-30	-41	3	-2	-20	-3	6	19	-2	100
cosTB2	19	-1	11	5	6	1	-9	69	58	44	-3	1	100	-2
hoo0	5	-7	3	1	2	-1	1	2	-19	4	23	100	1	19
hoo2	0	51	2	4	5	6	0	5	27	2	100	23	-3	6
hso12	14	-2	10	1	3	2	-3	16	38	100	2	4	44	-3
R2	5	37	24	18	39	48	-11	65	100	38	27	-19	58	-20
hso02	20	2	13	6	7	2	-11	100	65	16	5	2	69	-2
cosTB1	-0	0	-4	-4	-4	-8	100	-11	-11	-3	0	1	-9	3
Thrus2	-9	3	34	15	56	100	-8	2	48	2	6	-1	1	-41
cc1	-20	3	-30	40	100	56	-4	7	39	3	5	2	6	-30
hso22	-15	2	-13	100	40	15	-4	6	18	1	4	1	5	-7
cc2	-35	1	100	-13	-30	34	-4	13	24	10	2	3	11	-13
Thrus1	-0	100	1	2	3	3	0	2	37	-2	51	-7	-1	2
cc3	100	-0	-35	-15	-20	-9	-0	20	5	14	0	5	19	5

Background

cc3	Thrus1	cc2	hso22	cc1	Thrus2	cosTB1	hso02	R2	hso12	hoo2	hoo0	cosTB2	hso20
-7	-6	-12	50	-17	-23	7	-24	-12	-18	-2	-6	-11	100
13	35	24	10	27	31	-10	50	58	36	28	11	100	-11
16	4	14	10	11	3	-11	29	-1	26	49	100	11	-6
16	71	20	33	28	16	-8	57	64	40	100	49	28	-2
7	31	16	4	27	24	-2	-2	41	100	40	26	36	-18
4	81	24	16	52	56	-5	69	100	41	64	-1	58	-12
13	57	26	5	40	33	-13	100	69	-2	57	29	50	-24
-7	-6	-4	6	-0	-4	100	-13	-5	-2	-8	-11	-10	7
-13	17	16	10	54	100	-4	33	56	24	16	3	31	-23
-23	24	-38	12	100	54	-0	40	52	27	28	11	27	-17
-6	9	-1	100	12	10	6	5	16	4	33	10	10	50
-21	17	100	-1	-38	16	-4	26	24	16	20	14	24	-12
15	100	17	9	24	17	-6	57	81	31	71	4	35	-6
100	15	-21	-6	-23	-13	-7	13	4	7	16	16	13	-7

negative uncorrelated positive

Signal

	cc3	Thrus1	cc2	hso22	hso02	cc1	hso12	cosTB1	Thrus2	hoo2	hoo0	R2	cosTB2	hso20
hso20	5	5	1	8	-3	-18	-4	2	-26	11	22	-11	-2	100
cosTB2	21	0	12	3	67	5	43	-11	-0	19	-0	61	100	-2
R2	5	42	9	18	66	31	40	-12	37	45	-20	100	61	-11
hoo0	7	-8	3	10	0	3	2	0	0	21	100	-20	-0	22
hoo2	8	47	5	7	21	2	15	-2	0	100	21	45	19	11
Thrus2	-34	-0	-0	34	0	62	2	-8	100	0	0	37	-0	-26
cosTB1	-1	1	-0	-2	-12	-4	-5	100	-8	-2	0	-12	-11	2
hso12	14	2	8	1	17	4	100	-5	2	15	2	40	43	-4
cc1	-21	-1	-59	46	5	100	4	-4	62	2	3	31	5	-18
hso02	22	1	13	3	100	5	17	-12	0	21	0	66	67	-3
hso22	-15	0	-23	100	3	46	1	-2	34	7	10	18	3	8
cc2	-35	2	100	-23	13	-59	8	-0	-0	5	3	9	12	1
Thrus1	-2	100	2	0	1	-1	2	1	-0	47	-8	42	0	5
cc3	100	-2	-35	-15	22	-21	14	-1	-34	8	7	5	21	5

Background

cc3	Thrus1	cc2	hso22	hso02	cc1	hso12	cosTB1	Thrus2	hoo2	hoo0	R2	cosTB2	hso20
-5	-4	-7	60	-23	-14	-13	5	-13	13	3	-7	-11	100
16	37	21	5	47	26	34	-12	21	30	14	57	100	-11
9	81	21	15	62	47	40	-6	49	59	2	100	57	-7
22	7	18	13	32	13	31	-12	3	60	100	2	14	3
22	64	27	35	49	22	40	-8	11	100	60	59	30	13
-11	12	-3	12	26	55	22	-6	100	11	3	49	21	-13
-8	-6	-3	8	-11	0	-2	100	-6	-8	-12	-6	-12	5
8	28	15	8	5	30	100	-2	22	40	31	40	34	-13
-18	18	-45	12	39	100	30	0	55	22	13	47	26	-14
15	50	22	2	100	39	5	-11	26	49	32	62	47	-23
-4	6	-4	100	2	12	8	8	12	35	13	15	5	60
-16	23	100	-4	22	-45	15	-3	-3	27	18	21	21	-7
16	100	23	6	50	18	28	-6	12	64	7	81	37	-4
100	16	-16	-4	15	-18	8	-8	-11	22	22	9	16	-5

negative uncorrelated positive

Signal

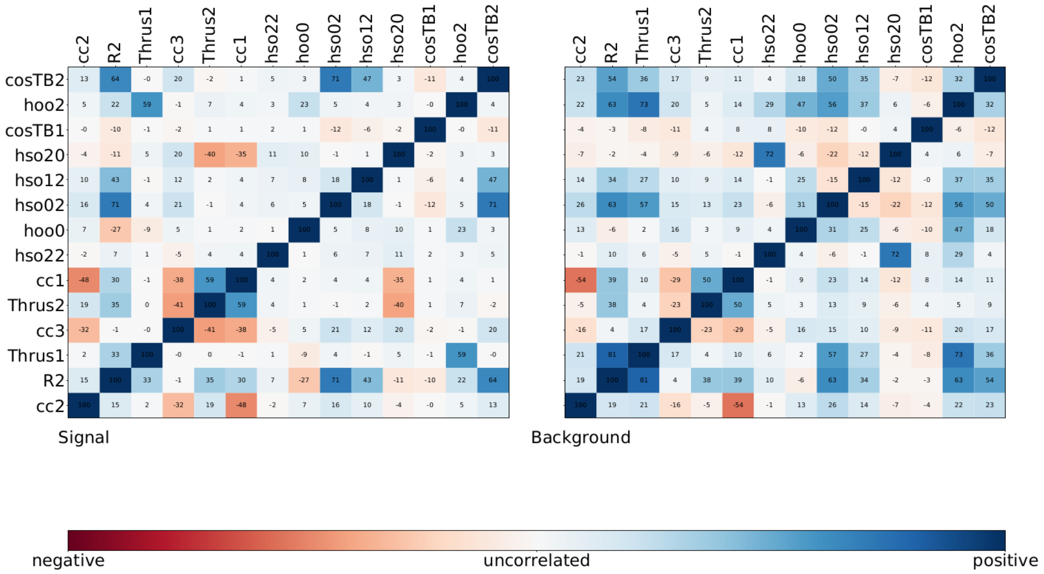
	cc3	cc2	Thrus1	hso02	hso22	cc1	hso12	cosTB1	hoo2	hoo0	Thrus2	R2	hso20	cosTB2
cosTB2	28	12	-1	58	0	4	37	-14	43	0	-0	64	-5	100
hso20	-3	9	4	-3	22	-14	-2	2	8	19	-9	-5	100	-5
R2	22	0	44	60	10	21	36	-13	65	-17	26	100	-5	64
Thrus2	-15	-31	-2	1	36	55	1	-7	2	2	100	26	-9	-0
hoo0	11	2	-8	2	17	4	2	1	20	100	2	-17	19	0
hoo2	22	7	40	36	8	6	24	-7	100	20	2	65	8	43
cosTB1	-6	-1	0	-12	2	-2	-5	100	-7	1	-7	-13	2	-14
hso12	17	5	-0	18	1	4	100	-5	24	2	1	36	-2	37
cc1	-8	-84	-1	6	35	100	4	-2	6	4	55	21	-14	4
hso22	-2	-26	-3	2	100	35	1	2	8	17	36	10	22	0
hso02	25	11	0	100	2	6	18	-12	36	2	1	60	-3	58
Thrus1	0	-0	100	0	-3	-1	-0	0	40	-8	-2	44	4	-1
cc2	-16	100	-0	11	-26	-84	5	-1	7	2	-31	0	9	12
cc3	100	-16	0	25	-2	-8	17	-6	22	11	-15	22	-3	28

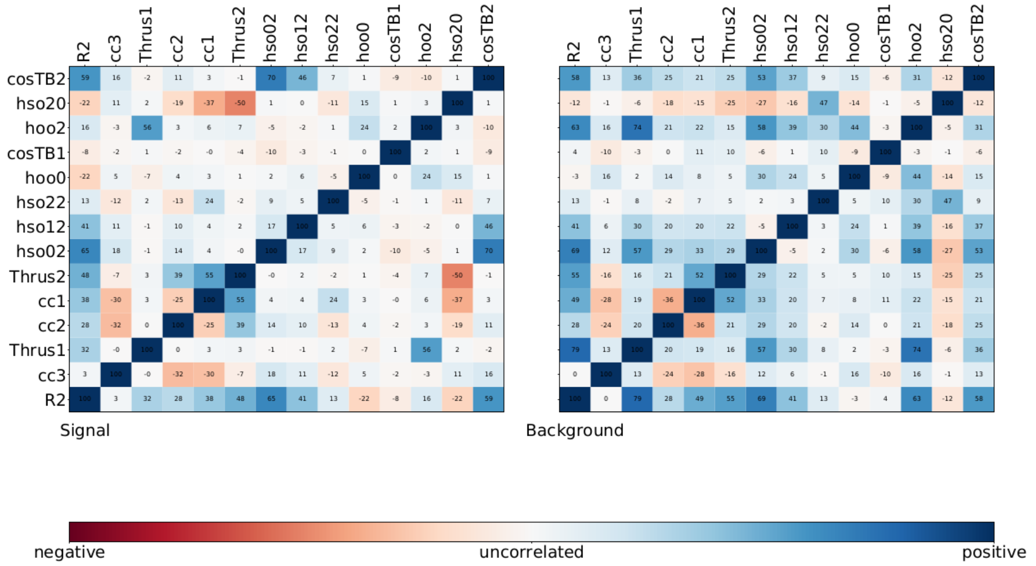
Background

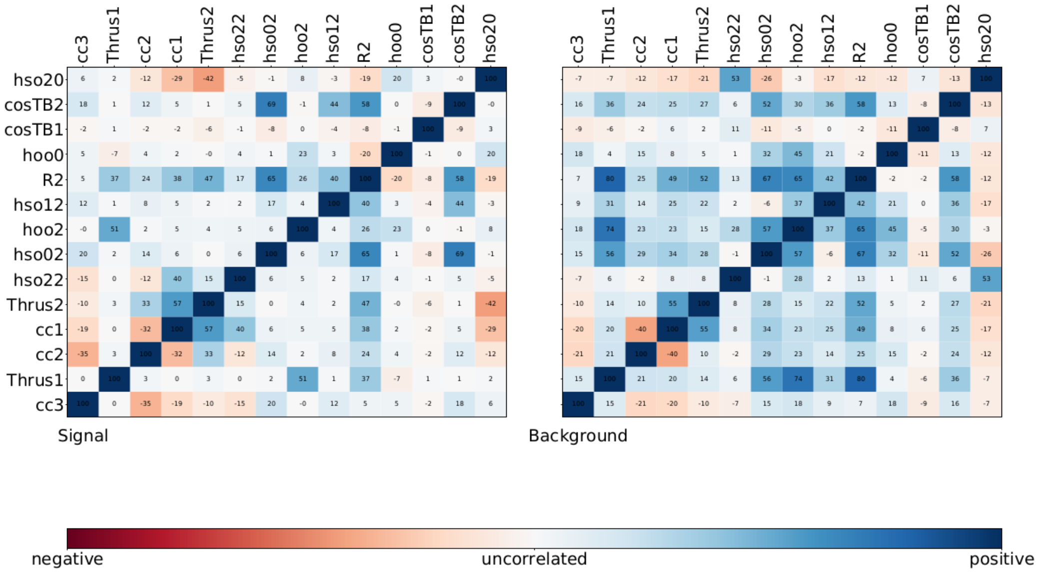
cc3	cc2	Thrus1	hso02	hso22	cc1	hso12	cosTB1	hoo2	hoo0	Thrus2	R2	hso20	cosTB2
21	21	39	40	-0	21	29	-13	34	12	16	56	-11	100
-17	-8	-6	-18	64	-5	-11	5	17	10	-4	-6	100	-11
23	18	85	52	11	39	34	-11	59	4	41	100	-6	56
1	-23	10	17	14	52	16	-7	8	0	100	41	-4	16
25	26	11	32	10	13	28	-14	68	100	0	4	10	12
29	34	62	47	26	21	39	-13	100	68	8	59	17	34
-10	-7	-12	-13	4	-1	-3	100	-13	-14	-7	-11	5	-13
16	13	27	17	16	24	100	-3	39	28	16	34	-11	29
-8	-53	17	29	13	100	24	-1	21	13	52	39	-5	21
-9	-7	3	10	100	13	16	4	26	10	14	11	64	-0
22	22	44	100	10	29	17	-13	47	32	17	52	-18	40
25	23	100	44	3	17	27	-12	62	11	10	85	-6	39
-8	100	23	22	-7	-53	13	-7	34	26	-23	18	-8	21
100	-8	25	22	-9	-8	16	-10	29	25	1	23	-17	21

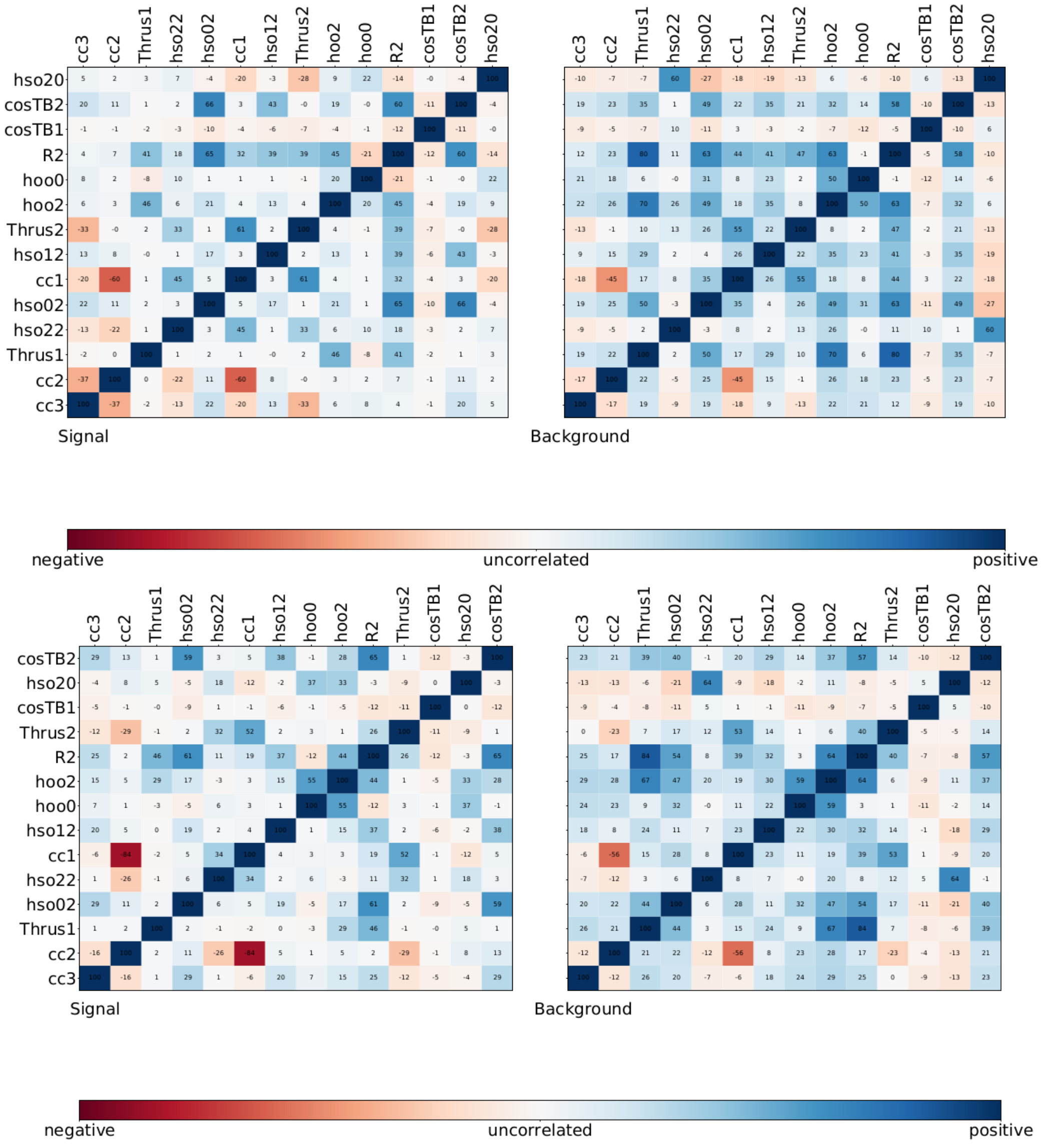
negative uncorrelated positive

Figure A.16: Correlation matrices for the $B^0 \rightarrow \rho^{\pm} e^{\mp} \nu_e$ continuum suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$.


cc2
R2
Thrus1
cc3
Thrus2
cc1
hso22
hoo0
hso02
hso12
hso20
cosTB1
hoo2
cosTB2
Signal
Background
negative
uncorrelated
positive


R2
cc3
Thrus1
cc2
cc1
Thrus2
hso02
hso12
hso22
hoo0
cosTB1
hoo2
hso20
cosTB2
Signal
Background
negative
uncorrelated
positive


cc3
Thrus1
cc2
cc1
Thrus2
hso22
hso02
hoo2
hso12
R2
hoo0
cosTB1
cosTB2
hso20
Signal
Background
negative
uncorrelated
positive



	cc3	cc2	Thrus1	hso22	hso02	cc1	hso12	Thrus2	hoo2	hoo0	R2	cosTB1	cosTB2	hso20
hso20	5	2	3	7	-4	-20	-3	-28	9	22	-14	-0	-4	100
cosTB2	20	11	1	2	66	3	43	-0	19	-0	60	-11	100	-4
cosTB1	-1	-1	-2	-3	-10	-4	-6	-7	-4	-1	-12	100	-11	-0
R2	4	7	41	18	65	32	39	39	45	-21	100	-12	60	-14
hoo0	8	2	-8	10	1	1	1	-1	20	100	-21	-1	-0	22
hoo2	6	3	46	6	21	4	13	4	100	20	45	-4	19	9
Thrus2	-33	-0	2	33	1	61	2	100	4	-1	39	-7	-0	-28
hso12	13	8	-0	1	17	3	100	2	13	1	39	-6	43	-3
cc1	-20	-60	1	45	5	100	3	61	4	1	32	-4	3	-20
hso02	22	11	2	3	100	5	17	1	21	1	65	-10	66	-4
hso22	-13	-22	1	100	3	45	1	33	6	10	18	-3	2	7
Thrus1	-2	0	100	1	2	1	-0	2	46	-8	41	-2	1	3
cc2	-37	100	0	-22	11	-60	8	-0	3	2	7	-1	11	2
cc3	100	-37	-2	-13	22	-20	13	-33	6	8	4	-1	20	5

Signal

	cc3	cc2	Thrus1	hso22	hso02	cc1	hso12	Thrus2	hoo2	hoo0	R2	cosTB1	cosTB2	hso20
	-10	-7	-7	60	-27	-18	-19	-13	6	-6	-10	6	-13	100
	19	23	35	1	49	22	35	21	32	14	58	-10	100	-13
	-9	-5	-7	10	-11	3	-3	-2	-7	-12	-5	100	-10	6
	12	23	80	11	63	44	41	47	63	-1	100	-5	58	-10
	21	18	6	-0	31	8	23	2	50	100	-1	-12	14	-6
	22	26	70	26	49	18	35	8	100	50	63	-7	32	6
	-13	-1	10	13	26	55	22	100	8	2	47	-2	21	-13
	9	15	29	2	4	26	100	22	35	23	41	-3	35	-19
	-18	-45	17	8	35	100	26	55	18	8	44	3	22	-18
	19	25	50	-3	100	35	4	26	49	31	63	-11	49	-27
	-9	-5	2	100	-3	8	2	13	26	-0	11	10	1	60
	19	22	100	2	50	17	29	10	70	6	80	-7	35	-7
	-17	100	22	-5	25	-45	15	-1	26	18	23	-5	23	-7
	100	-17	19	-9	19	-18	9	-13	22	21	12	-9	19	-10

Background

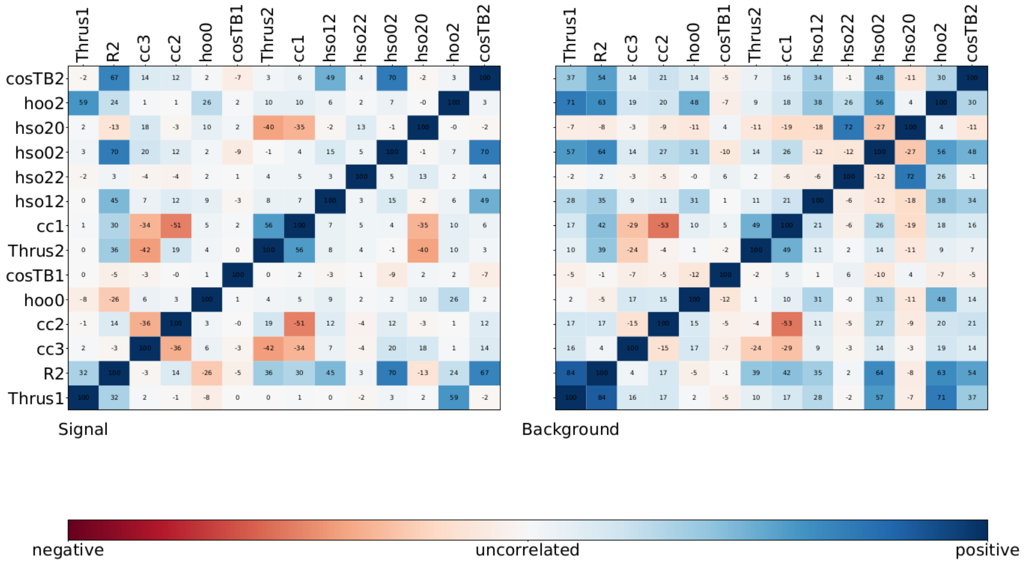
	cc3	cc2	Thrus1	hso02	hso22	cc1	hso12	hoo0	hoo2	R2	Thrus2	cosTB1	hso20	cosTB2
cosTB2	29	13	1	59	3	5	38	-1	28	65	1	-12	-3	100
hso20	-4	8	5	-5	18	-12	-2	37	33	-3	-9	0	100	-3
cosTB1	-5	-1	-0	-9	1	-1	-6	-1	-5	-12	-11	100	0	-12
Thrus2	-12	-29	-1	2	32	52	2	3	1	26	100	-11	-9	1
R2	25	2	46	61	11	19	37	-12	44	100	26	-12	-3	65
hoo2	15	5	29	17	-3	3	15	55	100	44	1	-5	33	28
hoo0	7	1	-3	-5	6	3	1	100	55	-12	3	-1	37	-1
hso12	20	5	0	19	2	4	100	1	15	37	2	-6	-2	38
cc1	-6	-84	-2	5	34	100	4	3	3	19	52	-1	-12	5
hso22	1	-26	-1	6	100	34	2	6	-3	11	32	1	18	3
hso02	29	11	2	100	6	5	19	-5	17	61	2	-9	-5	59
Thrus1	1	2	100	2	-1	-2	0	-3	29	46	-1	-0	5	1
cc2	-16	100	2	11	-26	-84	5	1	5	2	-29	-1	8	13
cc3	100	-16	1	29	1	-6	20	7	15	25	-12	-5	-4	29

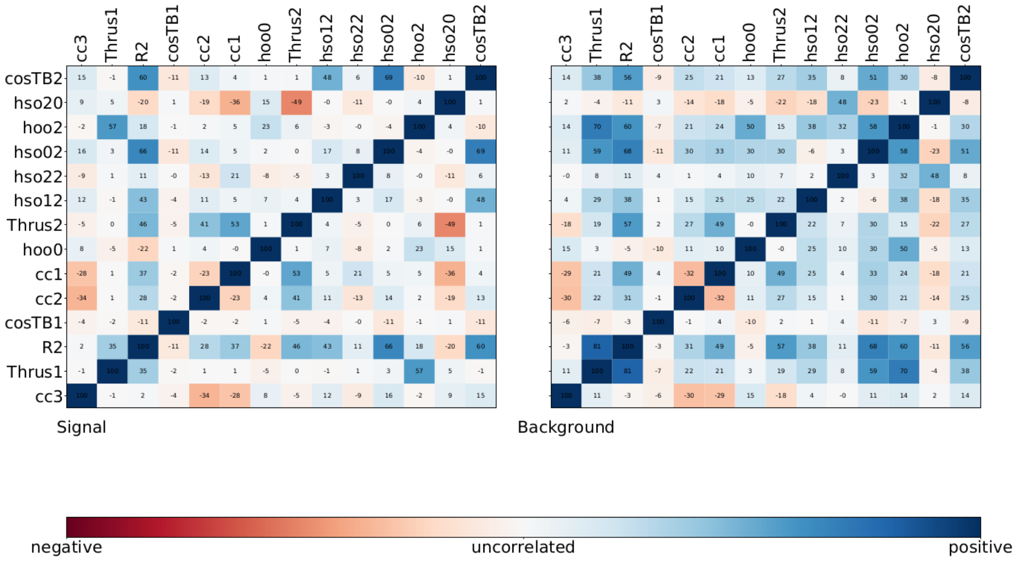
Signal

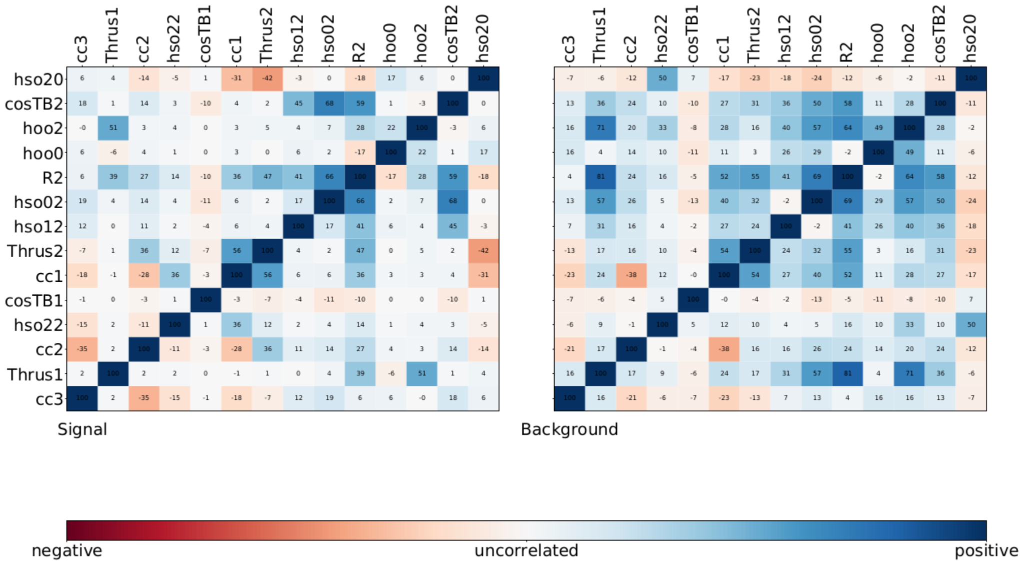
	cc3	cc2	Thrus1	hso02	hso22	cc1	hso12	hoo0	hoo2	R2	Thrus2	cosTB1	hso20	cosTB2
	23	21	39	40	-1	20	29	14	37	57	14	-10	-12	100
	-13	-13	-6	-21	64	-9	-18	-2	11	-8	-5	5	100	-12
	-9	-4	-8	-11	5	1	-1	-11	-9	-7	-5	100	5	-10
	0	-23	7	17	12	53	14	1	6	40	100	-5	-5	14
	25	17	84	54	8	39	32	3	64	100	40	-7	-8	57
	29	28	67	47	20	19	30	59	100	64	6	-9	11	37
	24	23	9	32	-0	11	22	100	59	3	1	-11	-2	14
	18	8	24	11	7	23	100	22	30	32	14	-1	-18	29
	-6	-56	15	28	8	100	23	11	19	39	53	1	-9	20
	-7	-12	3	6	100	8	7	-0	20	8	12	5	64	-1
	20	22	44	100	6	28	11	32	47	54	17	-11	-21	40
	26	21	100	44	3	15	24	9	67	84	7	-8	-6	39
	-12	100	21	22	-12	-56	8	23	28	17	-23	-4	-13	21
	100	-12	26	20	-7	-6	18	24	29	25	0	-9	-13	23

Background

Figure A.17: Correlation matrices for the $B^0 \to \rho^{\pm}\mu^{\mp}\nu_{\mu}$ continuum suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$.


Thrus1 R2 cc3 cc2 hoo0 cosTB1 Thrus2 cc1 hso12 hso22 hso02 hso20 hoo2 cosTB2
cosTB2 -2 67 14 12 2 -7 3 6 49 4 70 -2 3 100
hoo2 59 24 1 1 26 2 10 10 6 2 7 -0 100 3
hso20 2 -13 18 -3 10 2 -40 -35 -2 13 -1 100 -0 -2
hso02 3 70 20 12 2 -9 -1 4 15 5 100 -1 7 70
hso22 -2 3 -4 -4 2 1 4 5 3 100 5 13 2 4
hso12 0 45 7 12 9 -3 8 7 100 3 15 -2 6 49
cc1 1 30 -34 -51 5 2 56 100 7 5 4 -35 10 6
Thrus2 0 36 -42 19 4 0 100 56 8 4 -1 -40 10 3
cosTB1 0 -5 -3 -0 1 100 0 2 -3 1 -9 2 2 -7
hoo0 -8 -26 6 3 100 1 4 5 9 2 2 10 26 2
cc2 -1 14 -36 100 3 -0 19 -51 12 -4 12 -3 1 12
cc3 2 -3 100 -36 6 -3 -42 -34 7 -4 20 18 1 14
R2 32 100 -3 14 -26 -5 36 30 45 3 70 -13 24 67
Thrus1 100 32 2 -1 -8 0 0 1 0 -2 3 2 59 -2
Signal
Thrus1 R2 cc3 cc2 hoo0 cosTB1 Thrus2 cc1 hso12 hso22 hso02 hso20 hoo2 cosTB2
37 54 14 21 14 -5 7 16 34 -1 48 -11 30 100
71 63 19 20 48 -7 9 18 38 26 56 4 100 30
-7 -8 -3 -9 -11 4 -11 -19 -18 72 -27 100 4 -11
57 64 14 27 31 -10 14 26 -12 -12 100 -27 56 48
-2 2 -3 -5 -0 6 2 -6 -6 100 -12 72 26 -1
28 35 9 11 31 1 11 21 100 -6 -12 -18 38 34
17 42 -29 -53 10 5 49 100 21 -6 26 -19 18 16
10 39 -24 -4 1 -2 100 49 11 2 14 -11 9 7
-5 -1 -7 -5 -12 100 -2 5 1 6 -10 4 -7 -5
2 -5 17 15 100 -12 1 10 31 -0 31 -11 48 14
17 17 -15 100 15 -5 -4 -53 11 -5 27 -9 20 21
16 4 100 -15 17 -7 -24 -29 9 -3 14 -3 19 14
84 100 4 17 -5 -1 39 42 35 2 64 -8 63 54
100 84 16 17 2 -5 10 17 28 -2 57 -7 71 37
Background
negative
uncorrelated
positive


cc3 Thrus1 R2 cosTB1 cc2 cc1 hoo0 Thrus2 hso12 hso22 hso02 hoo2 hso20 cosTB2
cosTB2 15 -1 60 -11 13 4 1 1 48 6 69 -10 1 100
hso20 9 5 -20 1 -19 -36 15 -49 -0 -11 -0 4 100 1
hoo2 -2 57 18 -1 2 5 23 6 -3 -0 -4 100 4 -10
hso02 16 3 66 -11 14 5 2 0 17 8 100 -4 -0 69
hso22 -9 1 11 -0 -13 21 -8 -5 3 100 8 -0 -11 6
hso12 12 -1 43 -4 11 5 7 4 100 3 17 -3 -0 48
Thrus2 -5 0 46 -5 41 53 1 100 4 -5 0 6 -49 1
hoo0 8 -5 -22 1 4 -0 100 1 7 -8 2 23 15 1
cc1 -28 1 37 -2 -23 100 -0 53 5 21 5 5 -36 4
cc2 -34 1 28 -2 100 -23 4 41 11 -13 14 2 -19 13
cosTB1 -4 -2 -11 100 -2 -2 1 -5 -4 -0 -11 -1 1 -11
R2 2 35 100 -11 28 37 -22 46 43 11 66 18 -20 60
Thrus1 -1 100 35 -2 1 1 -5 0 -1 1 3 57 5 -1
cc3 100 -1 2 -4 -34 -28 8 -5 12 -9 16 -2 9 15
Signal
cc3 Thrus1 R2 cosTB1 cc2 cc1 hoo0 Thrus2 hso12 hso22 hso02 hoo2 hso20 cosTB2
14 38 56 -9 25 21 13 27 35 8 51 30 -8 100
2 -4 -11 3 -14 -18 -5 -22 -18 48 -23 -1 100 -8
14 70 60 -7 21 24 50 15 38 32 58 100 -1 30
11 59 68 -11 30 33 30 30 -6 3 100 58 -23 51
-0 8 11 4 1 4 10 7 2 100 3 32 48 8
4 29 38 1 15 25 25 22 100 2 -6 38 -18 35
-18 19 57 2 27 49 -0 100 22 7 30 15 -22 27
15 3 -5 -10 11 10 100 -0 25 10 30 50 -5 13
-29 21 49 4 -32 100 10 49 25 4 33 24 -18 21
-30 22 31 -1 100 -32 11 27 15 1 30 21 -14 25
-6 -7 -3 100 -1 4 -10 2 1 4 -11 -7 3 -9
-3 81 100 -3 31 49 -5 57 38 11 68 60 -11 56
11 100 81 -7 22 21 3 19 29 8 59 70 -4 38
100 11 -3 -6 -30 -29 15 -18 4 -0 11 14 2 14
Background
negative
uncorrelated
positive


cc3 Thrus1 cc2 hso22 cosTB1 cc1 Thrus2 hso12 hso02 R2 hoo0 hoo2 cosTB2 hso20
hso20 6 4 -14 -5 1 -31 -42 -3 0 -18 17 6 0 100
cosTB2 18 1 14 3 -10 4 2 45 68 59 1 -3 100 0
hoo2 -0 51 3 4 0 3 5 4 7 28 22 100 -3 6
hoo0 6 -6 4 1 0 3 0 6 2 -17 100 22 1 17
R2 6 39 27 14 -10 36 47 41 66 100 -17 28 59 -18
hso02 19 4 14 4 -11 6 2 17 100 66 2 7 68 0
hso12 12 0 11 2 -4 6 4 100 17 41 6 4 45 -3
Thrus2 -7 1 36 12 -7 56 100 4 2 47 0 5 2 -42
cc1 -18 -1 -28 36 -3 100 56 6 6 36 3 3 4 -31
cosTB1 -1 0 -3 1 100 -3 -7 -4 -11 -10 0 0 -10 1
hso22 -15 2 -11 100 1 36 12 2 4 14 1 4 3 -5
cc2 -35 2 100 -11 -3 -28 36 11 14 27 4 3 14 -14
Thrus1 2 100 2 2 0 -1 1 0 4 39 -6 51 1 4
cc3 100 2 -35 -15 -1 -18 -7 12 19 6 6 -0 18 6
Signal
cc3 Thrus1 cc2 hso22 cosTB1 cc1 Thrus2 hso12 hso02 R2 hoo0 hoo2 cosTB2 hso20
-7 -6 -12 50 7 -17 -23 -18 -24 -12 -6 -2 -11 100
13 36 24 10 -10 27 31 36 50 58 11 28 100 -11
16 71 20 33 -8 28 16 40 57 64 49 100 28 -2
16 4 14 10 -11 11 3 26 29 -2 100 49 11 -6
4 81 24 16 -5 52 55 41 69 100 -2 64 58 -12
13 57 26 5 -13 40 32 -2 100 69 29 57 50 -24
7 31 16 4 -2 27 24 100 -2 41 26 40 36 -18
-13 17 16 10 -4 54 100 24 32 55 3 16 31 -23
-23 24 -38 12 -0 100 54 27 40 52 11 28 27 -17
-7 -6 -4 5 100 -0 -4 -2 -13 -5 -11 -8 -10 7
-6 9 -1 100 5 12 10 4 5 16 10 33 10 50
-21 17 100 -1 -4 -38 16 16 26 24 14 20 24 -12
16 100 17 9 -6 24 17 31 57 81 4 71 36 -6
100 16 -21 -6 -7 -23 -13 7 13 4 16 16 13 -7
Background
negative
uncorrelated
positive

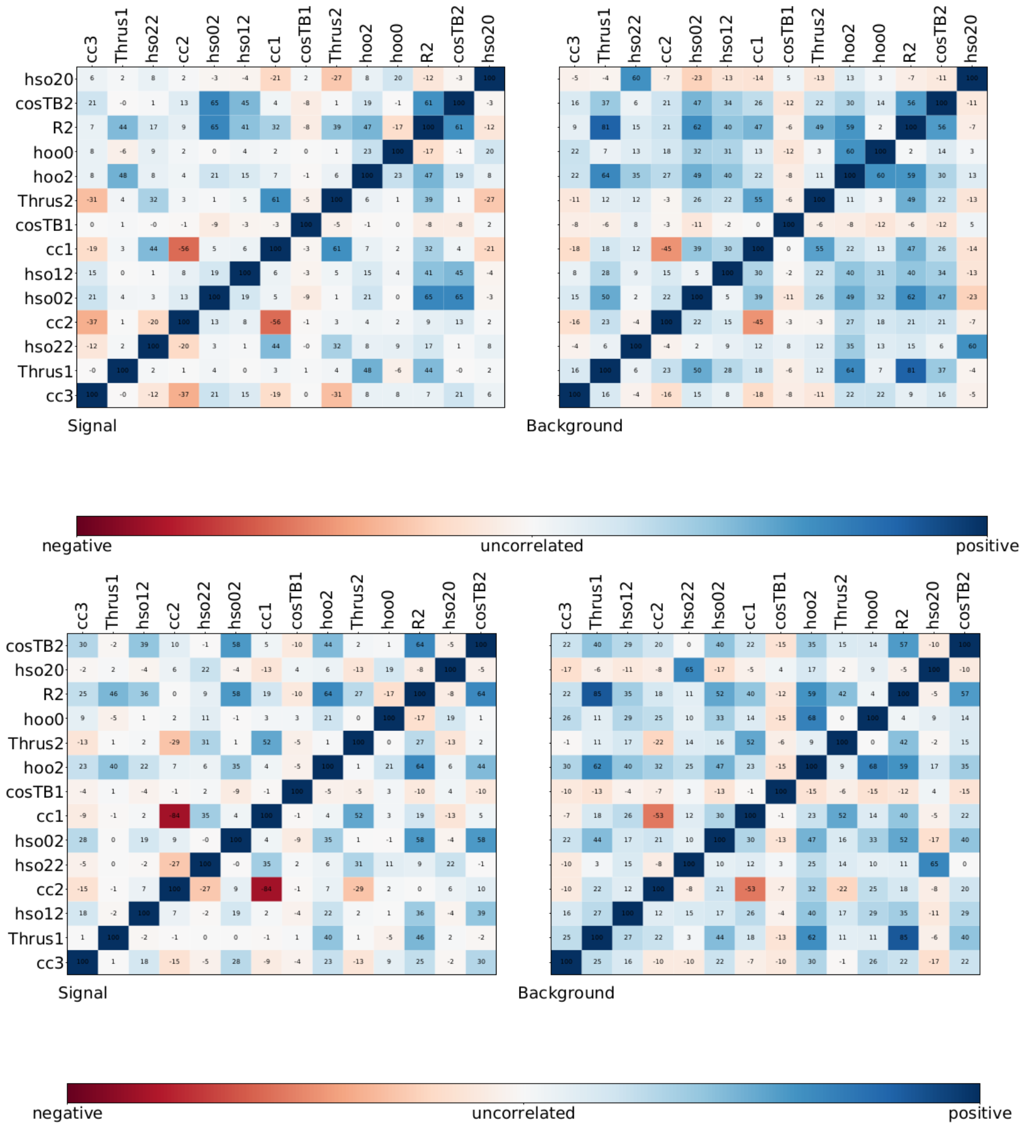
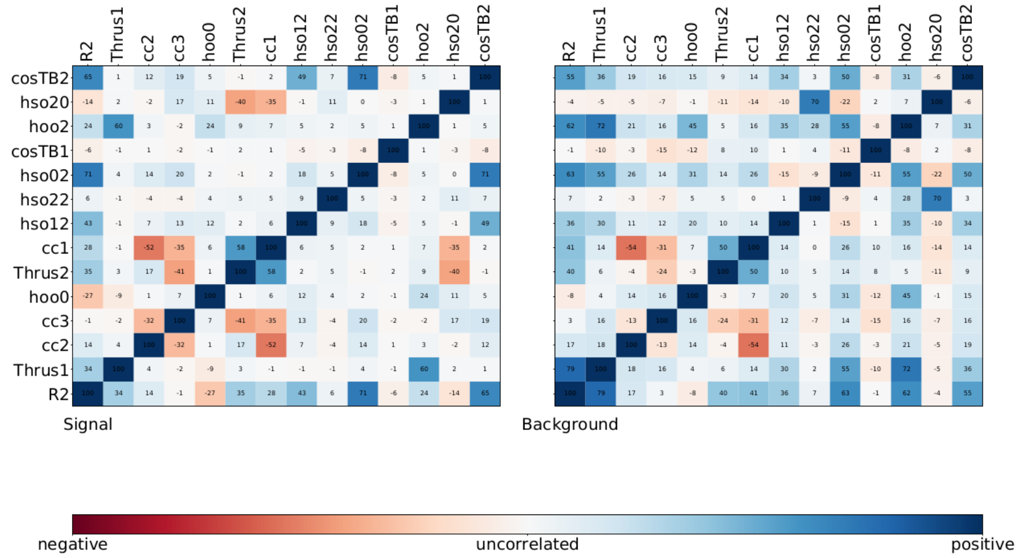
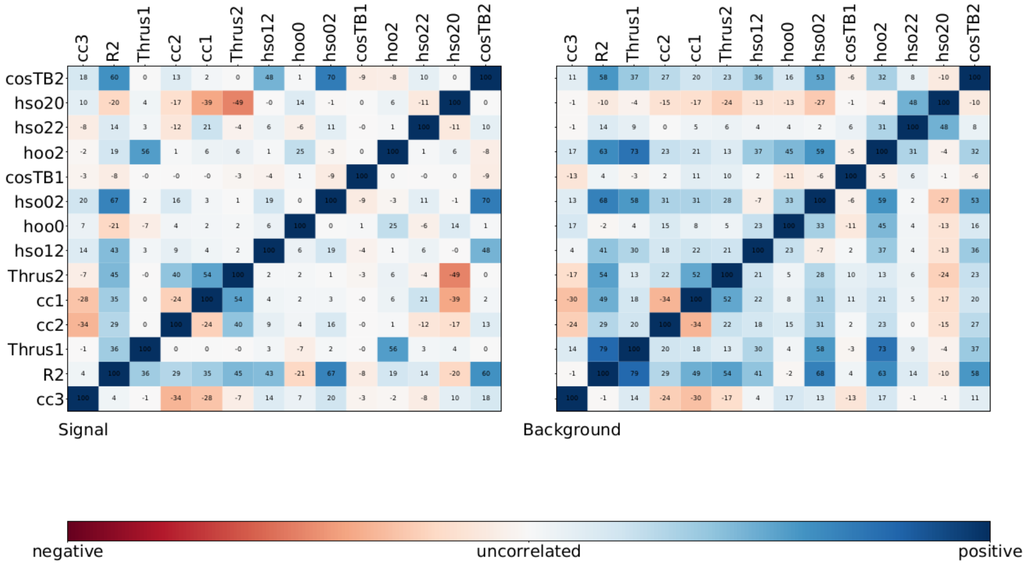
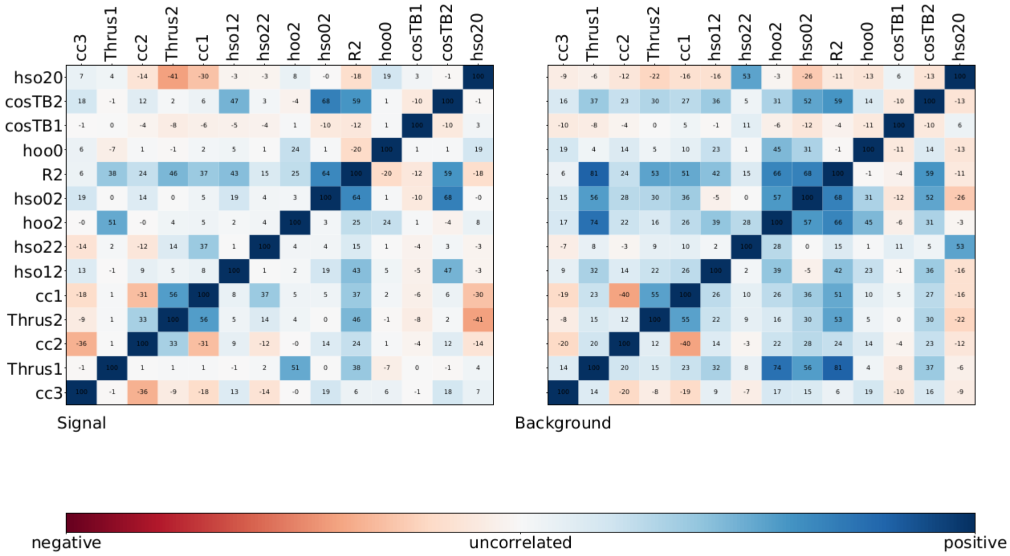
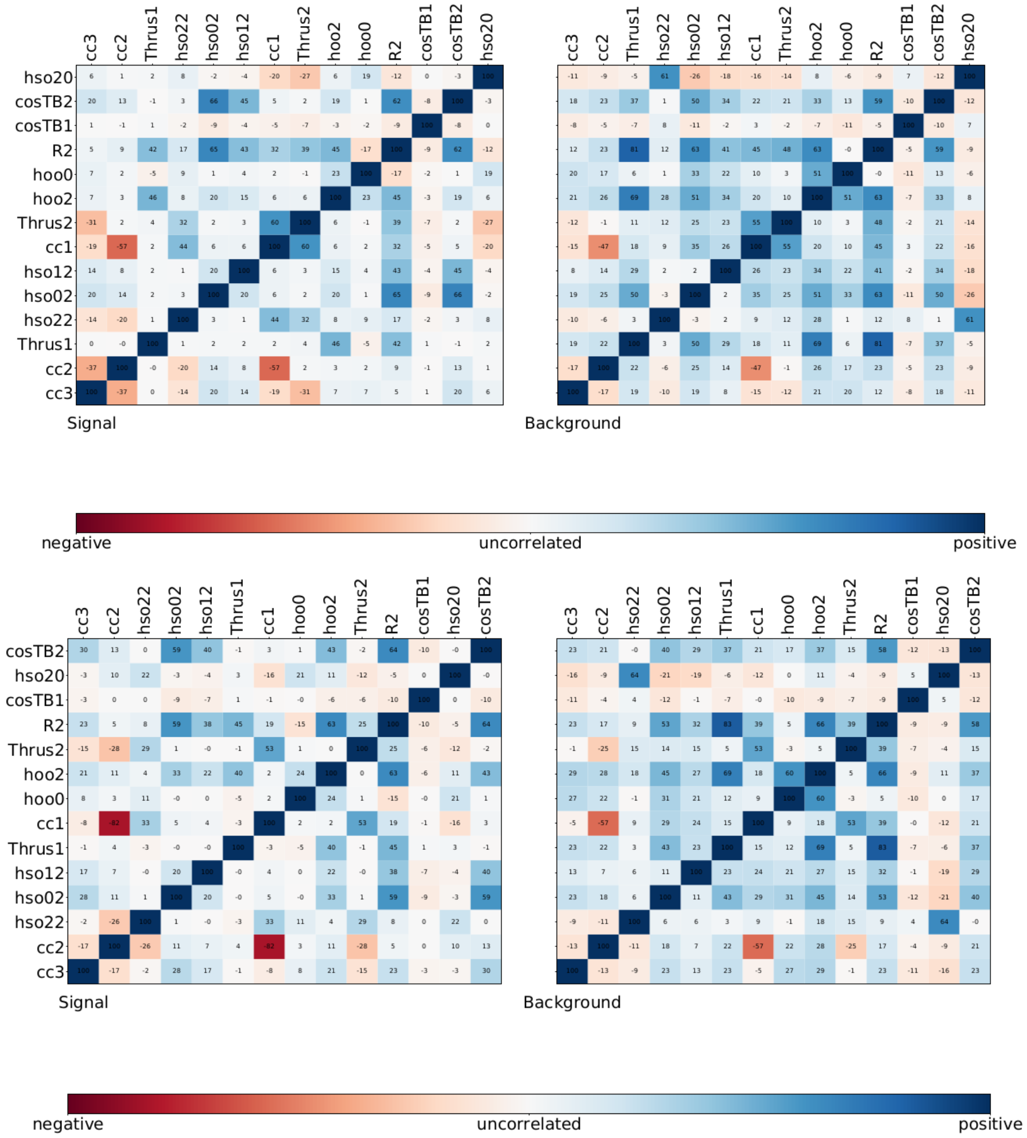


Figure A.18: Correlation matrices for the $B^{\pm} \rightarrow \rho^{0} e^{\pm} \nu_{e}$ continuum suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$.


R2
Thrus1
cc2
cc3
hoo0
Thrus2
cc1
hso12
hso22
hso02
cosTB1
hoo2
hso20
cosTB2
Signal
Background
negative
uncorrelated
positive


cc3
R2
Thrus1
cc2
cc1
Thrus2
hso12
hoo0
hso02
cosTB1
hoo2
hso22
hso20
cosTB2
Signal
Background
negative
uncorrelated
positive


cc3
Thrus1
cc2
Thrus2
cc1
hso12
hso22
hoo2
hso02
R2
hoo0
cosTB1
cosTB2
hso20
Signal
Background
negative
uncorrelated
positive



Signal

	cc3	cc2	Thrus1	hso22	hso02	hso12	cc1	Thrus2	hoo2	hoo0	R2	cosTB1	cosTB2	hso20
hso20	6	1	2	8	-2	-4	-20	-27	6	19	-12	0	-3	100
cosTB2	20	13	-1	3	66	45	5	2	19	1	62	-8	100	-3
cosTB1	1	-1	1	-2	-9	-4	-5	-7	-3	-2	-9	100	-8	0
R2	5	9	42	17	65	43	32	39	45	-17	100	-9	62	-12
hoo0	7	2	-5	9	1	4	2	-1	23	100	-17	-2	1	19
hoo2	7	3	46	8	20	15	6	6	100	23	45	-3	19	6
Thrus2	-31	2	4	32	2	3	60	100	6	-1	39	-7	2	-27
cc1	-19	-57	2	44	6	6	100	60	6	2	32	-5	5	-20
hso12	14	8	2	1	20	100	6	3	15	4	43	-4	45	-4
hso02	20	14	2	3	100	20	6	2	20	1	65	-9	66	-2
hso22	-14	-20	1	100	3	1	44	32	8	9	17	-2	3	8
Thrus1	0	-0	100	1	2	2	2	4	46	-5	42	1	-1	2
cc2	-37	100	-0	-20	14	8	-57	2	3	2	9	-1	13	1
cc3	100	-37	0	-14	20	14	-19	-31	7	7	5	1	20	6

Background

cc3	cc2	Thrus1	hso22	hso02	hso12	cc1	Thrus2	hoo2	hoo0	R2	cosTB1	cosTB2	hso20
-11	-9	-5	61	-26	-18	-16	-14	8	-6	-9	7	-12	100
18	23	37	1	50	34	22	21	33	13	59	-10	100	-12
-8	-5	-7	8	-11	-2	3	-2	-7	-11	-5	100	-10	7
12	23	81	12	63	41	45	48	63	-0	100	-5	59	-9
20	17	6	1	33	22	10	3	51	100	-0	-11	13	-6
21	26	69	28	51	34	20	10	100	51	63	-7	33	8
-12	-1	11	12	25	23	55	100	10	3	48	-2	21	-14
-15	-47	18	9	35	26	100	55	20	10	45	3	22	-16
8	14	29	2	2	100	26	23	34	22	41	-2	34	-18
19	25	50	-3	100	2	35	25	51	33	63	-11	50	-26
-10	-6	3	100	-3	2	9	12	28	1	12	8	1	61
19	22	100	3	50	29	18	11	69	6	81	-7	37	-5
-17	100	22	-6	25	14	-47	-1	26	17	23	-5	23	-9
100	-17	19	-10	19	8	-15	-12	21	20	12	-8	18	-11

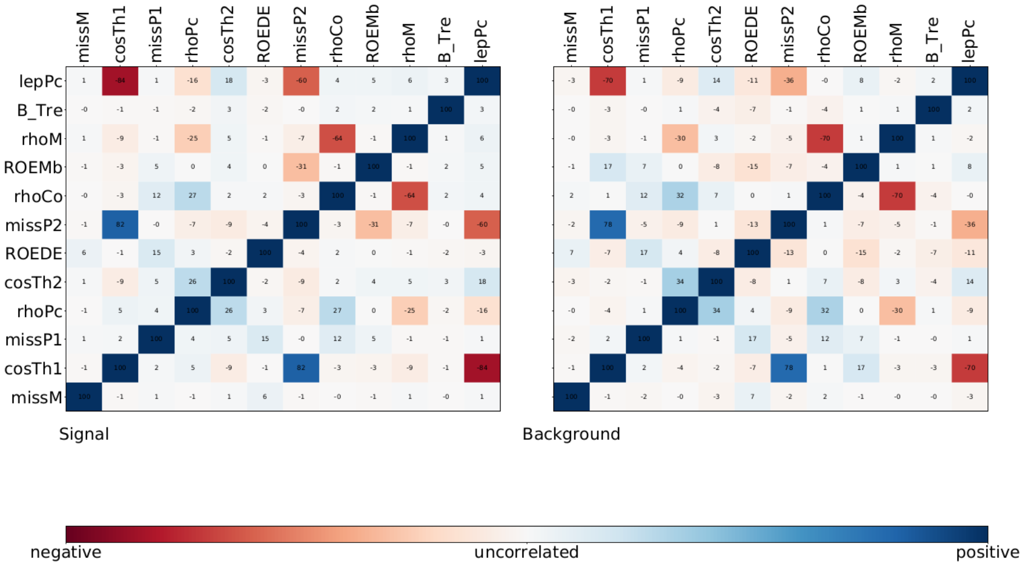
Signal

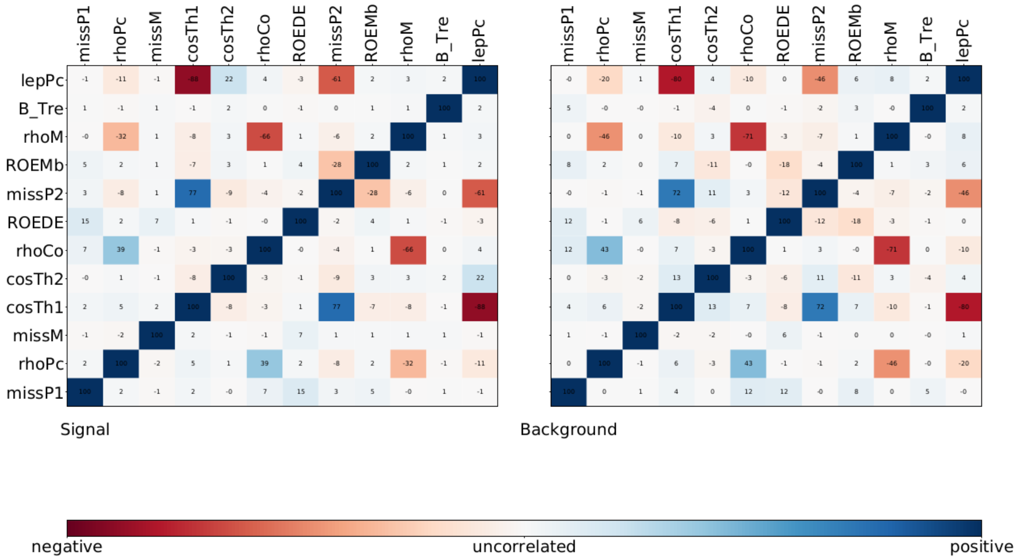
	cc3	cc2	hso22	hso02	hso12	Thrus1	cc1	hoo0	hoo2	Thrus2	R2	cosTB1	hso20	cosTB2
cosTB2	30	13	0	59	40	-1	3	1	43	-2	64	-10	-0	100
hso20	-3	10	22	-3	-4	3	-16	21	11	-12	-5	0	100	-0
cosTB1	-3	0	0	-9	-7	1	-1	-0	-6	-6	-10	100	0	-10
R2	23	5	8	59	38	45	19	-15	63	25	100	-10	-5	64
Thrus2	-15	-28	29	1	-0	-1	53	1	0	100	25	-6	-12	-2
hoo2	21	11	4	33	22	40	2	24	100	0	63	-6	11	43
hoo0	8	3	11	-0	0	-5	2	100	24	1	-15	-0	21	1
cc1	-8	-82	33	5	4	-3	100	2	2	53	19	-1	-16	3
Thrus1	-1	4	-3	-0	-0	100	-3	-5	40	-1	45	1	3	-1
hso12	17	7	-0	20	100	-0	4	0	22	-0	38	-7	-4	40
hso02	28	11	1	100	20	-0	5	-0	33	1	59	-9	-3	59
hso22	-2	-26	100	1	-0	-3	33	11	4	29	8	0	22	0
cc2	-17	100	-26	11	7	4	-82	3	11	-28	5	0	10	13
cc3	100	-17	-2	28	17	-1	-8	8	21	-15	23	-3	-3	30

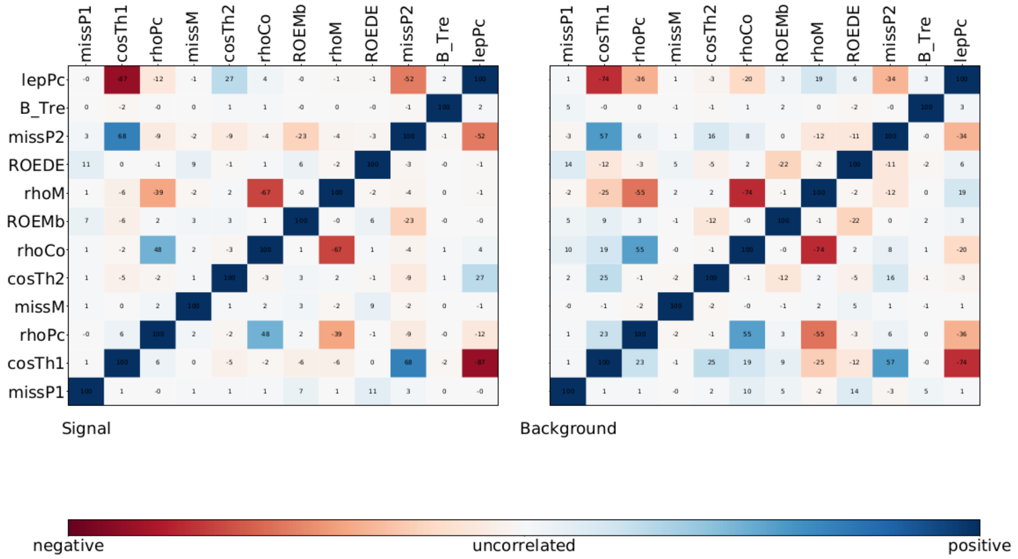
Background

cc3	cc2	hso22	hso02	hso12	Thrus1	cc1	hoo0	hoo2	Thrus2	R2	cosTB1	hso20	cosTB2
23	21	-0	40	29	37	21	17	37	15	58	-12	-13	100
-16	-9	64	-21	-19	-6	-12	0	11	-4	-9	5	100	-13
-11	-4	4	-12	-1	-7	-0	-10	-9	-7	-9	100	5	-12
23	17	9	53	32	83	39	5	66	39	100	-9	-9	58
-1	-25	15	14	15	5	53	-3	5	100	39	-7	-4	15
29	28	18	45	27	69	18	60	100	5	66	-9	11	37
27	22	-1	31	21	12	9	100	60	-3	5	-10	0	17
-5	-57	9	29	24	15	100	9	18	53	39	-0	-12	21
23	22	3	43	23	100	15	12	69	5	83	-7	-6	37
13	7	6	11	100	23	24	21	27	15	32	-1	-19	29
23	18	6	100	11	43	29	31	45	14	53	-12	-21	40
-9	-11	100	6	6	3	9	-1	18	15	9	4	64	-0
-13	100	-11	18	7	22	-57	22	28	-25	17	-4	-9	21
100	-13	-9	23	13	23	-5	27	29	-1	23	-11	-16	23

Figure A.19: Correlation matrices for the $B^{\pm} \rightarrow \rho^{0}\mu^{\pm}\nu_{\mu}$ continuum suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$.


missM
cosTh1
missP1
rhoPc
cosTh2
ROEDE
missP2
rhoCo
ROEMb
rhoM
B_Tre
lepPc
lepPc
B_Tre
rhoM
ROEMb
rhoCo
missP2
ROEDE
cosTh2
rhoPc
missP1
cosTh1
missM
Signal
Background
negative
uncorrelated
positive


missP1
rhoPc
missM
cosTh1
cosTh2
rhoCo
ROEDE
missP2
ROEMb
rhoM
B_Tre
lepPc
lepPc
B_Tre
rhoM
ROEMb
missP2
ROEDE
rhoCo
cosTh2
cosTh1
missM
rhoPc
missP1
Signal
Background
negative
uncorrelated
positive


missP1
cosTh1
rhoPc
missM
cosTh2
rhoCo
ROEMb
rhoM
ROEDE
missP2
B_Tre
lepPc
lepPc
B_Tre
missP2
ROEDE
rhoM
ROEMb
rhoCo
cosTh2
missM
rhoPc
cosTh1
missP1
Signal
Background
negative
uncorrelated
positive

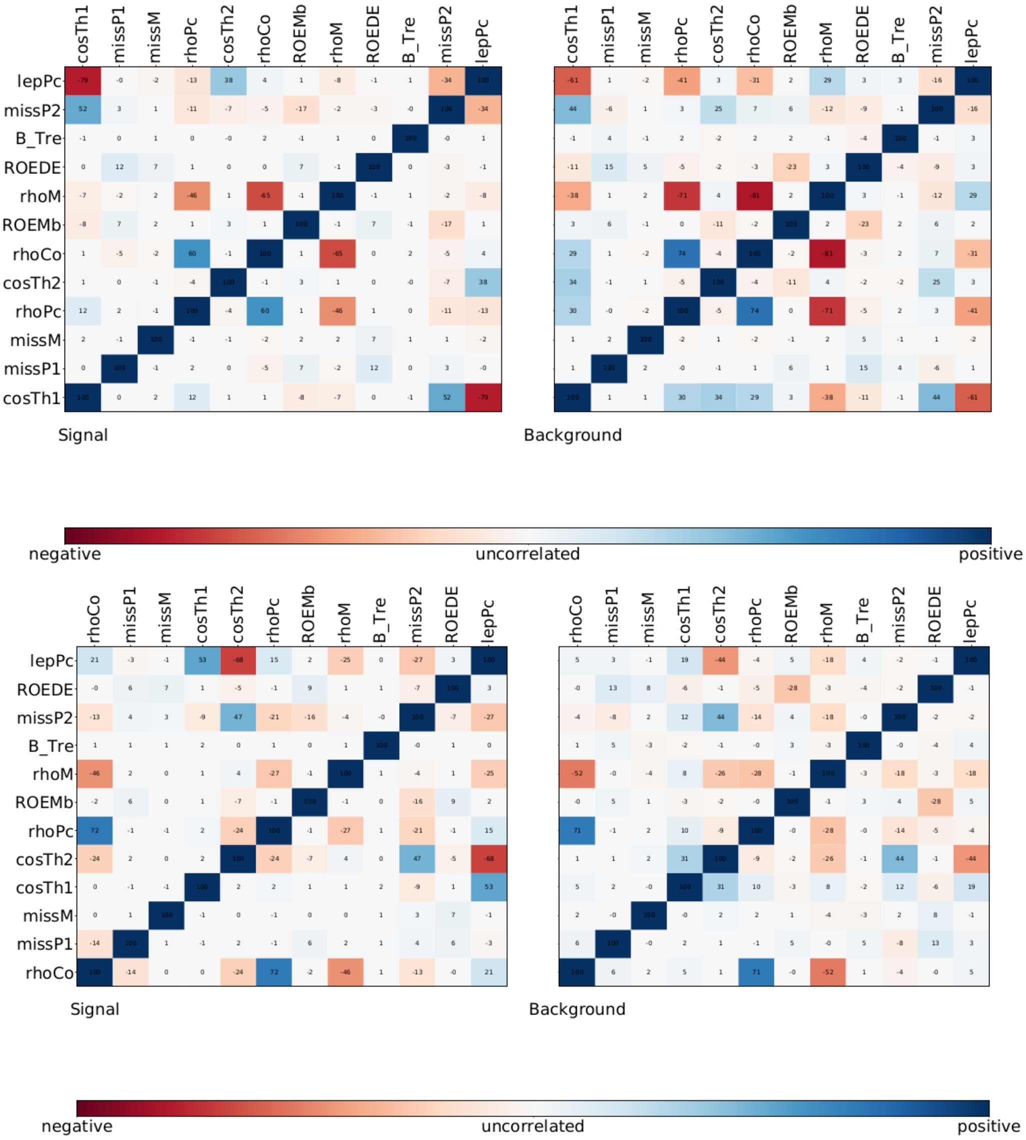
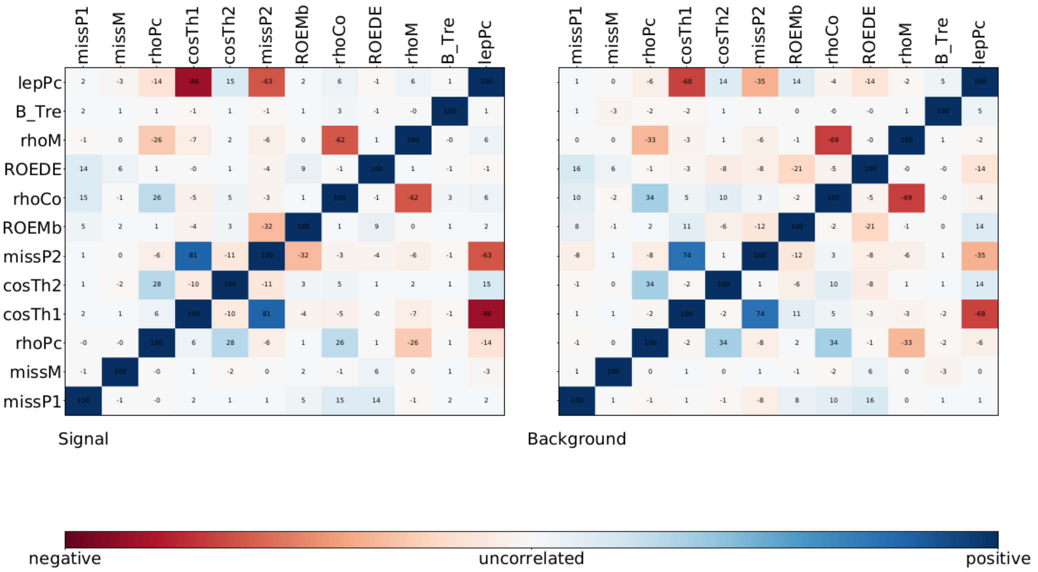
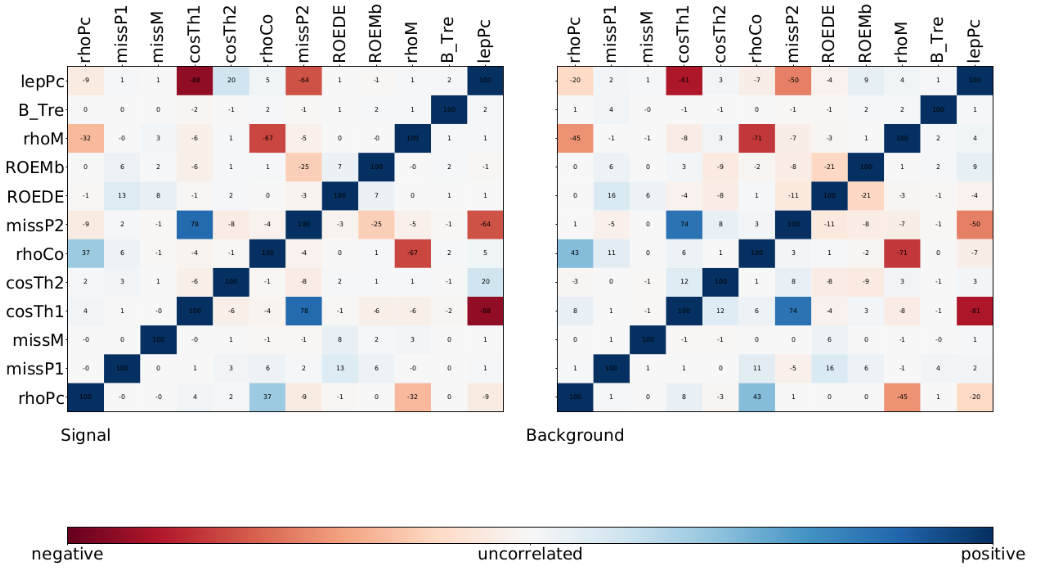
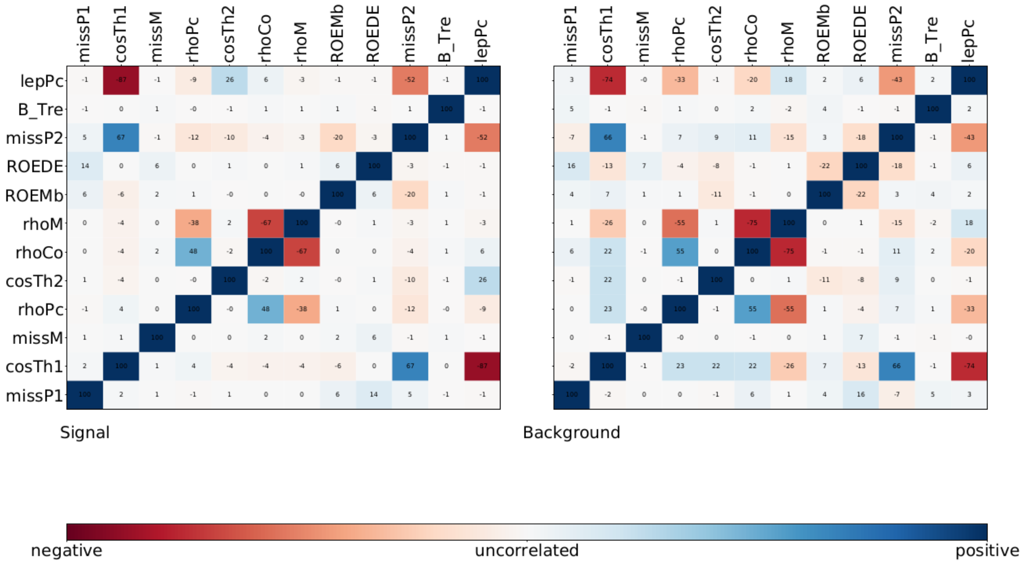


Figure A.20: Correlation matrices for the $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ $B\overline{B}$ suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$.


missP1
missM
rhoPc
cosTh1
cosTh2
missP2
ROEMb
rhoCo
ROEDE
rhoM
B_Tre
lepPc
lepPc
B_Tre
rhoM
ROEDE
rhoCo
ROEMb
missP2
cosTh2
cosTh1
rhoPc
missM
missP1
Signal
Background
negative
uncorrelated
positive


rhoPc
missP1
missM
cosTh1
cosTh2
rhoCo
missP2
ROEDE
ROEMb
rhoM
B_Tre
lepPc
lepPc
B_Tre
rhoM
ROEMb
ROEDE
missP2
rhoCo
cosTh2
cosTh1
missM
missP1
rhoPc
Signal
Background
negative
uncorrelated
positive


missP1
cosTh1
missM
rhoPc
cosTh2
rhoCo
rhoM
ROEMb
ROEDE
missP2
B_Tre
lepPc
lepPc
B_Tre
missP2
ROEDE
ROEMb
rhoM
rhoCo
cosTh2
rhoPc
missM
cosTh1
missP1
Signal
Background
negative
uncorrelated
positive

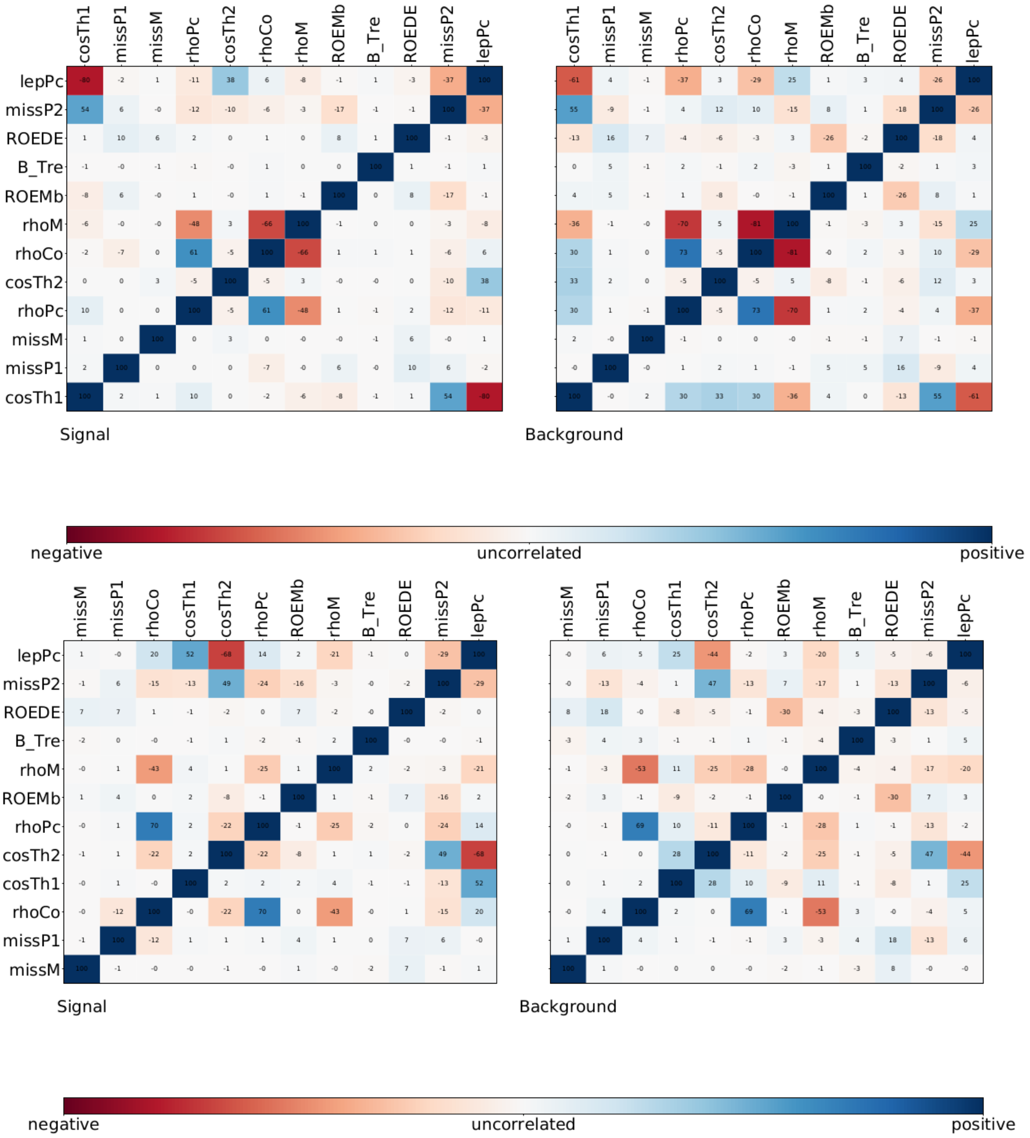
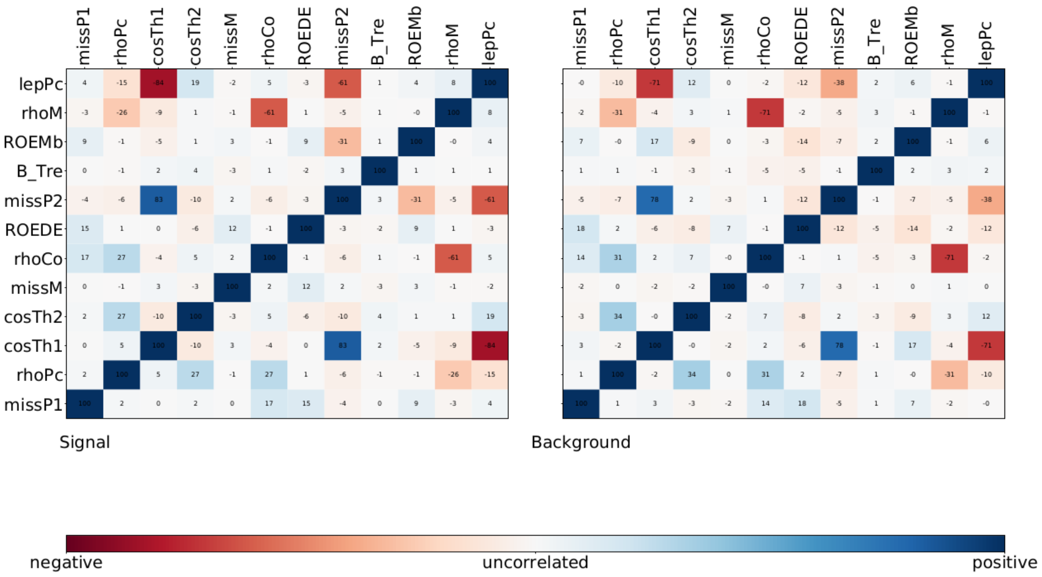


Figure A.21: Correlation matrices for the $B^0 \to \rho^{\pm}\mu^{\mp}\nu_\mu$ $B\overline{B}$ suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$.

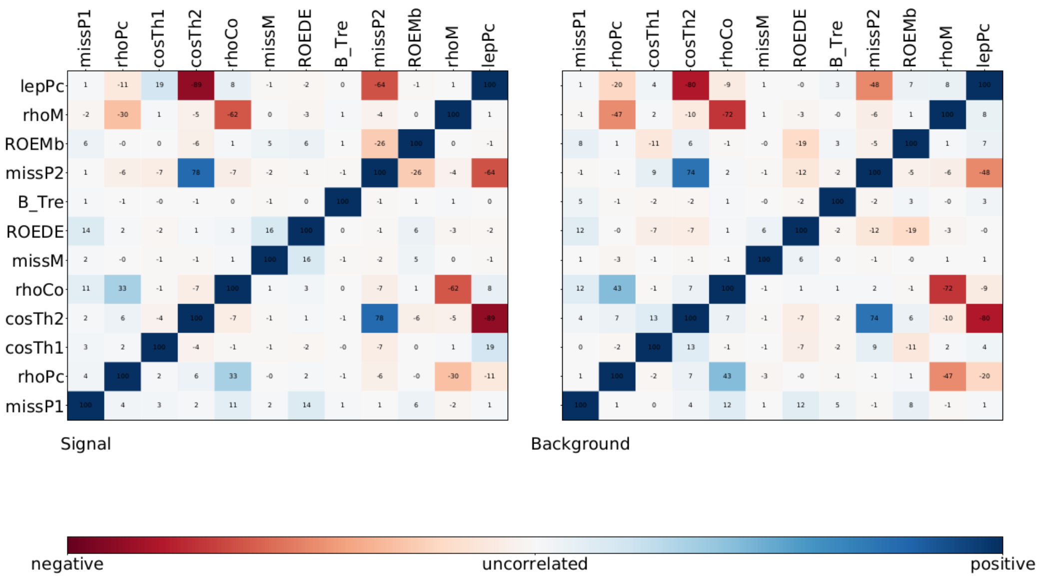


	missP1	rhoPc	cosTh1	cosTh2	missM	rhoCo	ROEDE	missP2	B_Tre	ROEMb	rhoM	lepPc
lepPc	4	-15	-84	19	-2	5	-3	-61	1	4	8	100
rhoM	-3	-26	-9	1	-1	-61	1	-5	1	-0	100	8
ROEMb	9	-1	-5	1	3	-1	9	-31	1	100	-0	4
B_Tre	0	-1	2	4	-3	1	-2	3	100	1	1	1
missP2	-4	-6	83	-10	2	-6	-3	100	3	-31	-5	-61
ROEDE	15	1	0	-6	12	-1	100	-3	-2	9	1	-3
rhoCo	17	27	-4	5	2	100	-1	-6	1	-1	-61	5
missM	0	-1	3	-3	100	2	12	2	-3	3	-1	-2
cosTh2	2	27	-10	100	-3	5	-6	-10	4	1	1	19
cosTh1	0	5	100	-10	3	-4	0	83	2	-5	-9	-84
rhoPc	2	100	5	27	-1	27	1	-6	-1	-1	-26	-15
missP1	100	2	0	2	0	17	15	-4	0	9	-3	4

Signal

	missP1	rhoPc	cosTh1	cosTh2	missM	rhoCo	ROEDE	missP2	B_Tre	ROEMb	rhoM	lepPc
lepPc	-0	-10	-71	12	0	-2	-12	-38	2	6	-1	100
rhoM	-2	-31	-4	3	1	-71	-2	-5	3	-1	100	-1
ROEMb	7	-0	17	-9	0	-3	-14	-7	2	100	-1	6
B_Tre	1	1	-1	-3	-1	-5	-5	-1	100	2	3	2
missP2	-5	-7	78	2	-3	1	-12	100	-1	-7	-5	-38
ROEDE	18	2	-6	-8	7	-1	100	-12	-5	-14	-2	-12
rhoCo	14	31	2	7	-0	100	-1	1	-5	-3	-71	-2
missM	-2	0	-2	-2	100	-0	7	-3	-1	0	1	0
cosTh2	-3	34	-0	100	-2	7	-8	2	-3	-9	3	12
cosTh1	3	-2	100	-0	-2	2	-6	78	-1	17	-4	-71
rhoPc	1	100	-2	34	0	31	2	-7	1	-0	-31	-10
missP1	100	1	3	-3	-2	14	18	-5	1	7	-2	-0

Background

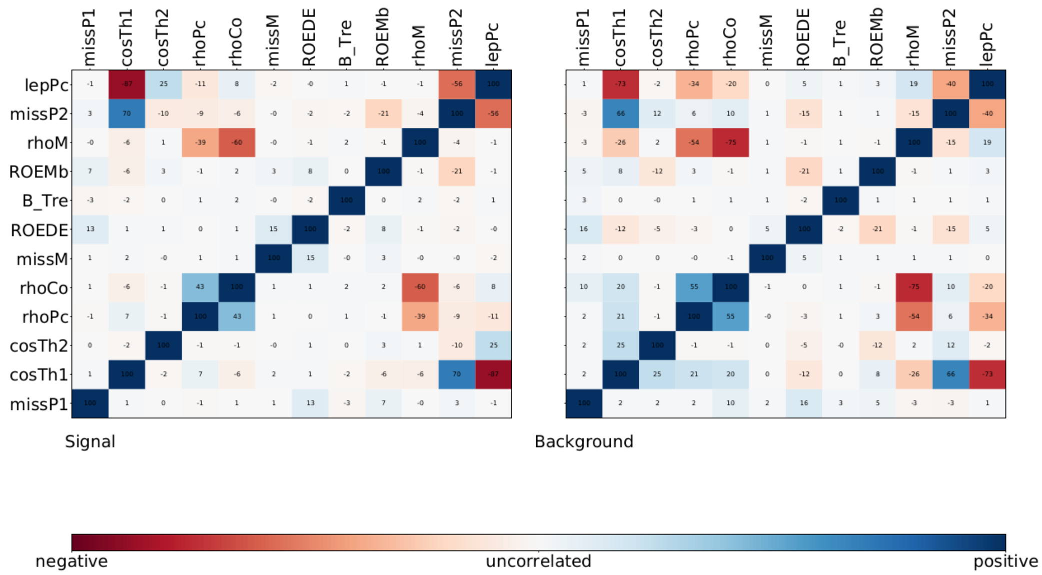


	missP1	rhoPc	cosTh1	cosTh2	rhoCo	missM	ROEDE	B_Tre	missP2	ROEMb	rhoM	lepPc
lepPc	1	-11	19	-89	8	-1	-2	0	-64	-1	1	100
rhoM	-2	-30	1	-5	-62	0	-3	1	-4	0	100	1
ROEMb	6	-0	0	-6	1	5	6	1	-26	100	0	-1
missP2	1	-6	-7	78	-7	-2	-1	-1	100	-26	-4	-64
B_Tre	1	-1	-0	-1	0	-1	0	100	-1	1	1	0
ROEDE	14	2	-2	1	3	16	100	0	-1	6	-3	-2
missM	2	-0	-1	-1	1	100	16	-1	-2	5	0	-1
rhoCo	11	33	-1	-7	100	1	3	0	-7	1	-62	8
cosTh2	2	6	-4	100	-7	-1	1	-1	78	-6	-5	-89
cosTh1	3	2	100	-4	-1	-1	-2	-0	-7	0	1	19
rhoPc	4	100	2	6	33	-0	2	-1	-6	-0	-30	-11
missP1	100	4	3	2	11	2	14	1	1	6	-2	1

Signal

	missP1	rhoPc	cosTh1	cosTh2	rhoCo	missM	ROEDE	B_Tre	missP2	ROEMb	rhoM	lepPc
lepPc	1	-20	4	-80	-9	1	-0	3	-48	7	8	100
rhoM	-1	-47	2	-10	-72	1	-3	-0	-6	1	100	8
ROEMb	8	1	-11	6	-1	-0	-19	3	-5	100	1	7
missP2	-1	-1	9	74	2	-1	-12	-2	100	-5	-6	-48
B_Tre	5	-1	-2	-2	1	-0	-2	100	-2	3	-0	3
ROEDE	12	-0	-7	-7	1	6	100	-2	-12	-19	-3	-0
missM	1	-3	-1	-1	-1	100	6	-0	-1	-0	1	1
rhoCo	12	43	-1	7	100	-1	1	1	2	-1	-72	-9
cosTh2	4	7	13	100	7	-1	-7	-2	74	6	-10	-80
cosTh1	0	-2	100	13	-1	-1	-7	-2	9	-11	2	4
rhoPc	1	100	-2	7	43	-3	-0	-1	-1	1	-47	-20
missP1	100	1	0	4	12	1	12	5	-1	8	-1	1

Background



	missP1	cosTh1	cosTh2	rhoPc	rhoCo	missM	ROEDE	B_Tre	ROEMb	rhoM	missP2	lepPc
lepPc	-1	-87	25	-11	8	-2	-0	1	-1	-1	-56	100
missP2	3	70	-10	-9	-6	-0	-2	-2	-21	-4	100	-56
rhoM	-0	-6	1	-39	-60	-0	-1	2	-1	100	-4	-1
ROEMb	7	-6	3	-1	2	3	8	0	100	-1	-21	-1
B_Tre	-3	-2	0	1	2	-0	-2	100	0	2	-2	1
ROEDE	13	1	1	0	1	15	100	-2	8	-1	-2	-0
missM	1	2	-0	1	1	100	15	-0	3	-0	-0	-2
rhoCo	1	-6	-1	43	100	1	1	2	2	-60	-6	8
rhoPc	-1	7	-1	100	43	1	0	1	-1	-39	-9	-11
cosTh2	0	-2	100	-1	-1	-0	1	0	3	1	-10	25
cosTh1	1	100	-2	7	-6	2	1	-2	-6	-6	70	-87
missP1	100	1	0	-1	1	1	13	-3	7	-0	3	-1

Signal

	missP1	cosTh1	cosTh2	rhoPc	rhoCo	missM	ROEDE	B_Tre	ROEMb	rhoM	missP2	lepPc
lepPc	1	-73	-2	-34	-20	0	5	1	3	19	-40	100
missP2	-3	66	12	6	10	1	-15	1	1	-15	100	-40
rhoM	-3	-26	2	-54	-75	1	-1	1	-1	100	-15	19
ROEMb	5	8	-12	3	-1	1	-21	1	100	-1	1	3
B_Tre	3	0	-0	1	1	1	-2	100	1	1	1	1
ROEDE	16	-12	-5	-3	0	5	100	-2	-21	-1	-15	5
missM	2	0	0	-0	-1	100	5	1	1	1	1	0
rhoCo	10	20	-1	55	100	-1	0	1	-1	-75	10	-20
rhoPc	2	21	-1	100	55	-0	-3	1	3	-54	6	-34
cosTh2	2	25	100	-1	-1	0	-5	-0	-12	2	12	-2
cosTh1	2	100	25	21	20	0	-12	0	8	-26	66	-73
missP1	100	2	2	2	10	2	16	3	5	-3	-3	1

Background

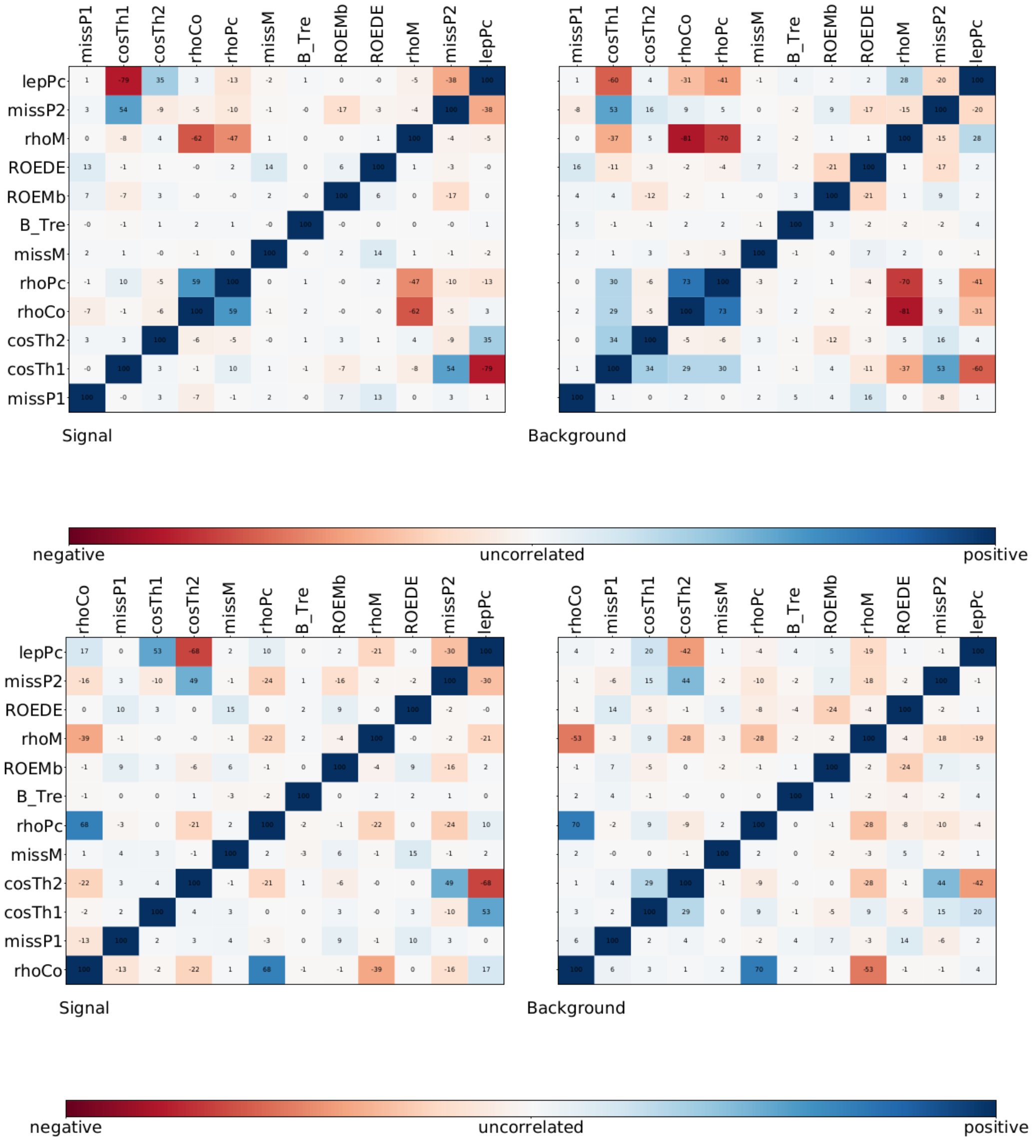
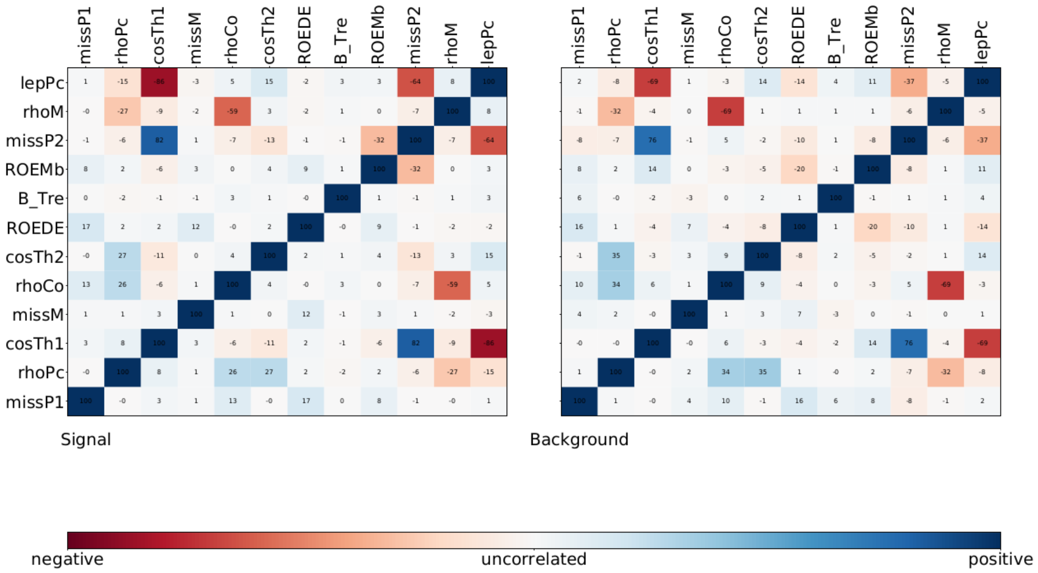
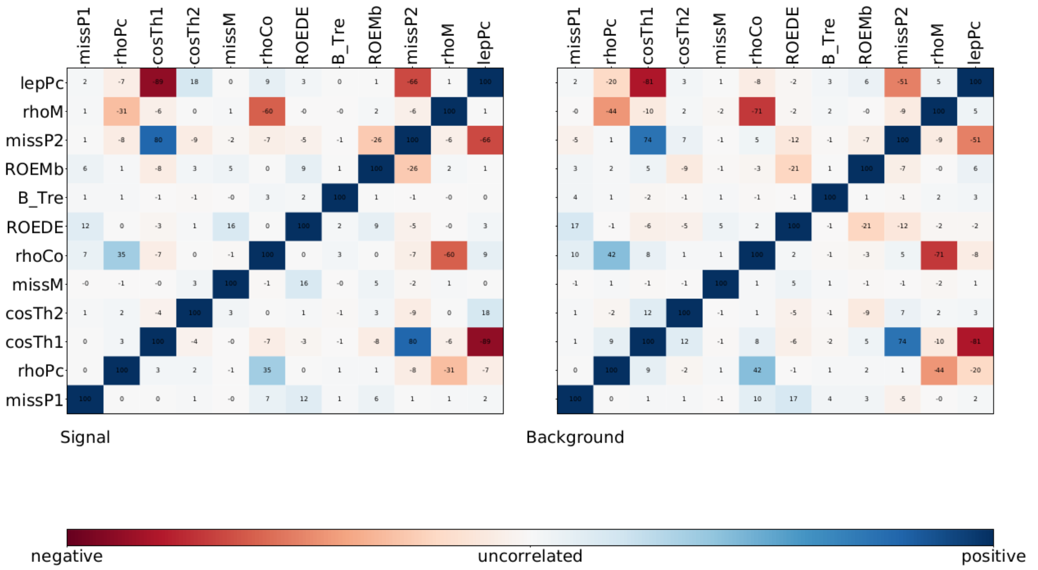
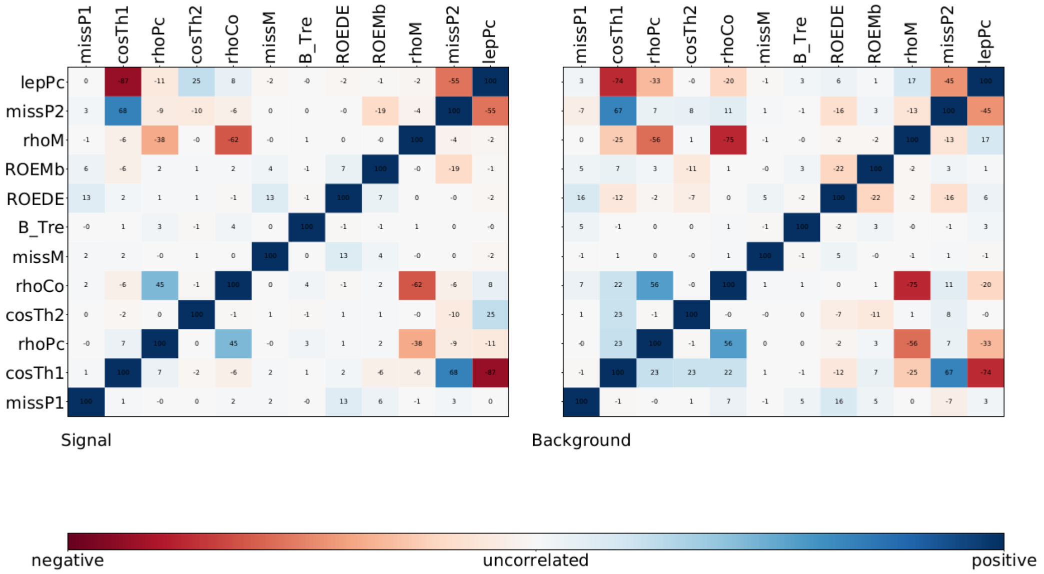
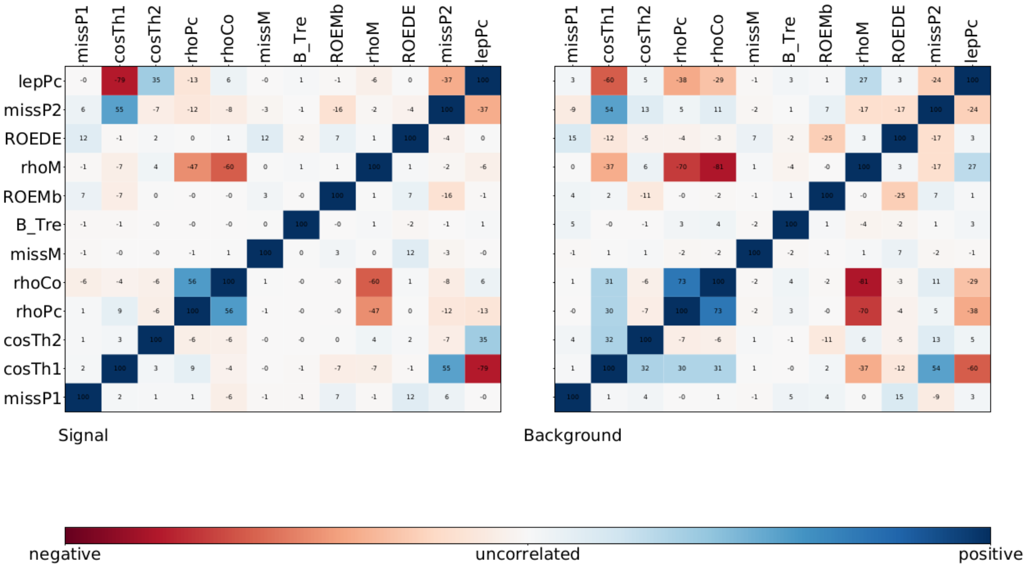


Figure A.22: Correlation matrices for the $B^{\pm} \to \rho^{0} e^{\pm} \nu_e$ $B\overline{B}$ suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$.


missP1
rhoPc
cosTh1
missM
rhoCo
cosTh2
ROEDE
B_Tre
ROEMb
missP2
rhoM
lepPc
Signal
Background
negative
uncorrelated
positive


missP1
rhoPc
cosTh1
cosTh2
missM
rhoCo
ROEDE
B_Tre
ROEMb
missP2
rhoM
lepPc
Signal
Background
negative
uncorrelated
positive


missP1
cosTh1
rhoPc
cosTh2
rhoCo
missM
B_Tre
ROEDE
ROEMb
rhoM
missP2
lepPc
Signal
Background
negative
uncorrelated
positive

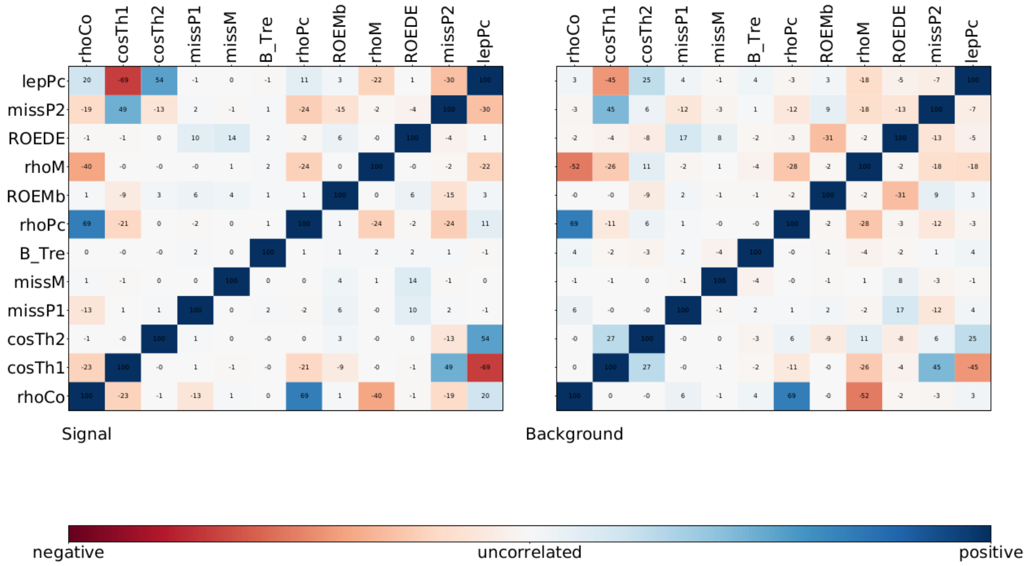


	missP1	cosTh1	cosTh2	rhoPc	rhoCo	missM	B_Tre	ROEMb	rhoM	ROEDE	missP2	lepPc
lepPc	-0	-79	35	-13	6	-0	1	-1	-6	0	-37	100
missP2	6	55	-7	-12	-8	-3	-1	-16	-2	-4	100	-37
ROEDE	12	-1	2	0	1	12	-2	7	1	100	-4	0
rhoM	-1	-7	4	-47	-60	0	1	1	100	1	-2	-6
ROEMb	7	-7	0	-0	-0	3	-0	100	1	7	-16	-1
B_Tre	-1	-1	-0	-0	-0	0	100	-0	1	-2	-1	1
missM	-1	-0	-0	-1	1	100	0	3	0	12	-3	-0
rhoCo	-6	-4	-6	56	100	1	-0	-0	-60	1	-8	6
rhoPc	1	9	-6	100	56	-1	-0	-0	-47	0	-12	-13
cosTh2	1	3	100	-6	-6	-0	-0	0	4	2	-7	35
cosTh1	2	100	3	9	-4	-0	-1	-7	-7	-1	55	-79
missP1	100	2	1	1	-6	-1	-1	7	-1	12	6	-0

Signal

	missP1	cosTh1	cosTh2	rhoPc	rhoCo	missM	B_Tre	ROEMb	rhoM	ROEDE	missP2	lepPc
lepPc	3	-60	5	-38	-29	-1	3	1	27	3	-24	100
missP2	-9	54	13	5	11	-2	1	7	-17	-17	100	-24
ROEDE	15	-12	-5	-4	-3	7	-2	-25	3	100	-17	3
rhoM	0	-37	6	-70	-81	1	-4	-0	100	3	-17	27
ROEMb	4	2	-11	-0	-2	-1	1	100	-0	-25	7	1
B_Tre	5	-0	-1	3	4	-2	100	1	-4	-2	1	3
missM	-1	1	1	-2	-2	100	-2	-1	1	7	-2	-1
rhoCo	1	31	-6	73	100	-2	4	-2	-81	-3	11	-29
rhoPc	-0	30	-7	100	73	-2	3	-0	-70	-4	5	-38
cosTh2	4	32	100	-7	-6	1	-1	-11	6	-5	13	5
cosTh1	1	100	32	30	31	1	-0	2	-37	-12	54	-60
missP1	100	1	4	-0	1	-1	5	4	0	15	-9	3

Background



	rhoCo	cosTh1	cosTh2	missP1	missM	B_Tre	rhoPc	ROEMb	rhoM	ROEDE	missP2	lepPc
lepPc	20	-69	54	-1	0	-1	11	3	-22	1	-30	100
missP2	-19	49	-13	2	-1	1	-24	-15	-2	-4	100	-30
ROEDE	-1	-1	0	10	14	2	-2	6	-0	100	-4	1
rhoM	-40	-0	-0	-0	1	2	-24	0	100	-0	-2	-22
ROEMb	1	-9	3	6	4	1	1	100	0	6	-15	3
rhoPc	69	-21	0	-2	0	1	100	1	-24	-2	-24	11
B_Tre	0	-0	-0	2	0	100	1	1	2	2	1	-1
missM	1	-1	0	0	100	0	0	4	1	14	-1	0
missP1	-13	1	1	100	0	2	-2	6	-0	10	2	-1
cosTh2	-1	-0	100	1	0	-0	0	3	-0	0	-13	54
cosTh1	-23	100	-0	1	-1	-0	-21	-9	-0	-1	49	-69
rhoCo	100	-23	-1	-13	1	0	69	1	-40	-1	-19	20

Signal

	rhoCo	cosTh1	cosTh2	missP1	missM	B_Tre	rhoPc	ROEMb	rhoM	ROEDE	missP2	lepPc
lepPc	3	-45	25	4	-1	4	-3	3	-18	-5	-7	100
missP2	-3	45	6	-12	-3	1	-12	9	-18	-13	100	-7
ROEDE	-2	-4	-8	17	8	-2	-3	-31	-2	100	-13	-5
rhoM	-52	-26	11	-2	1	-4	-28	-2	100	-2	-18	-18
ROEMb	-0	-0	-9	2	-1	-1	-2	100	-2	-31	9	3
rhoPc	69	-11	6	1	-0	-0	100	-2	-28	-3	-12	-3
B_Tre	4	-2	-3	2	-4	100	-0	-1	-4	-2	1	4
missM	-1	-1	0	-1	100	-4	-0	-1	1	8	-3	-1
missP1	6	-0	-0	100	-1	2	1	2	-2	17	-12	4
cosTh2	-0	27	100	-0	0	-3	6	-9	11	-8	6	25
cosTh1	0	100	27	-0	-1	-2	-11	-0	-26	-4	45	-45
rhoCo	100	0	-0	6	-1	4	69	-0	-52	-2	-3	3

Background

Figure A.23: Correlation matrices for the $B^{\pm} \to \rho^{0} \mu^{\pm} \nu_{\mu}$ $B\overline{B}$ suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$.

A.7.5 FOM distributions

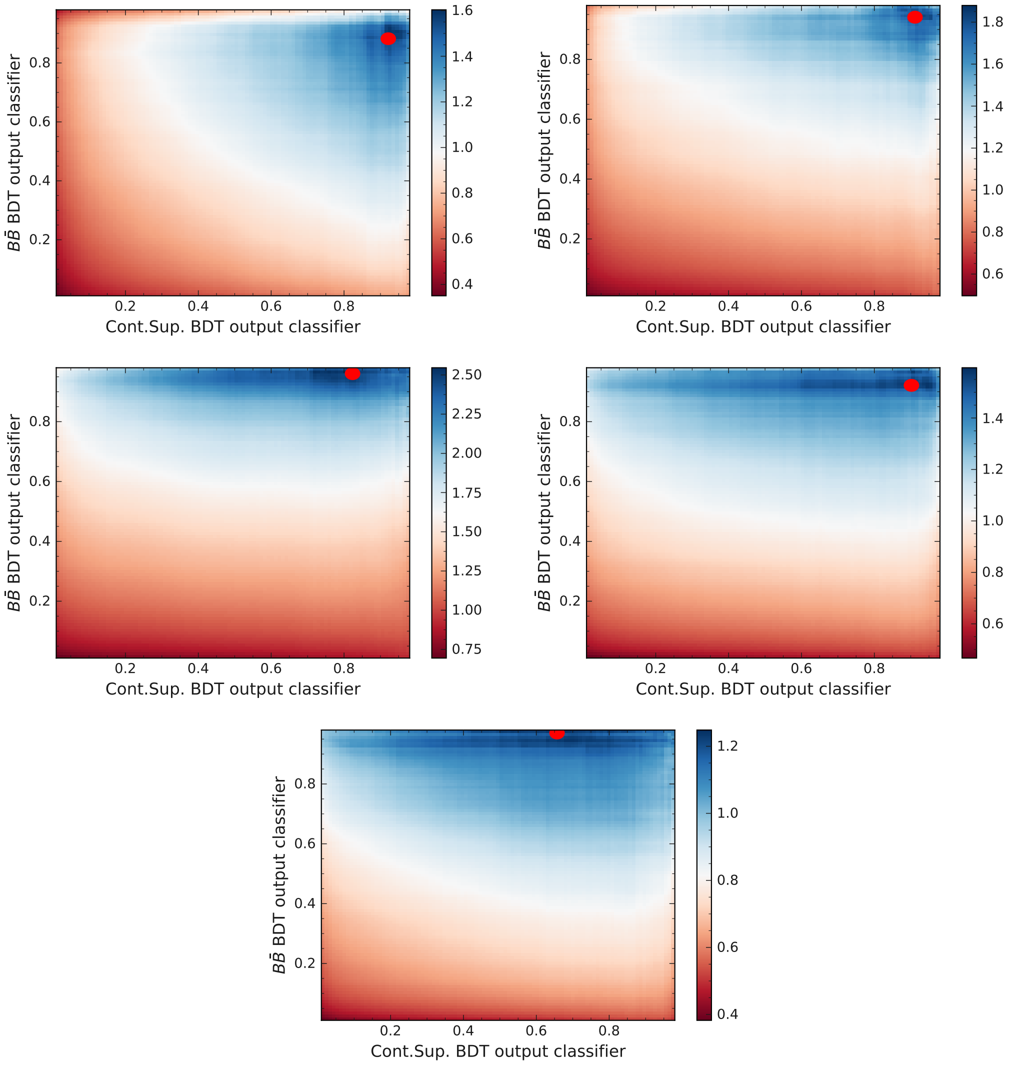


Figure A.24: BDT output classifier FOM distributions for the $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ signal decay channel. From left-to-right then top-to-bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. The red dot indicates the optimal selection.

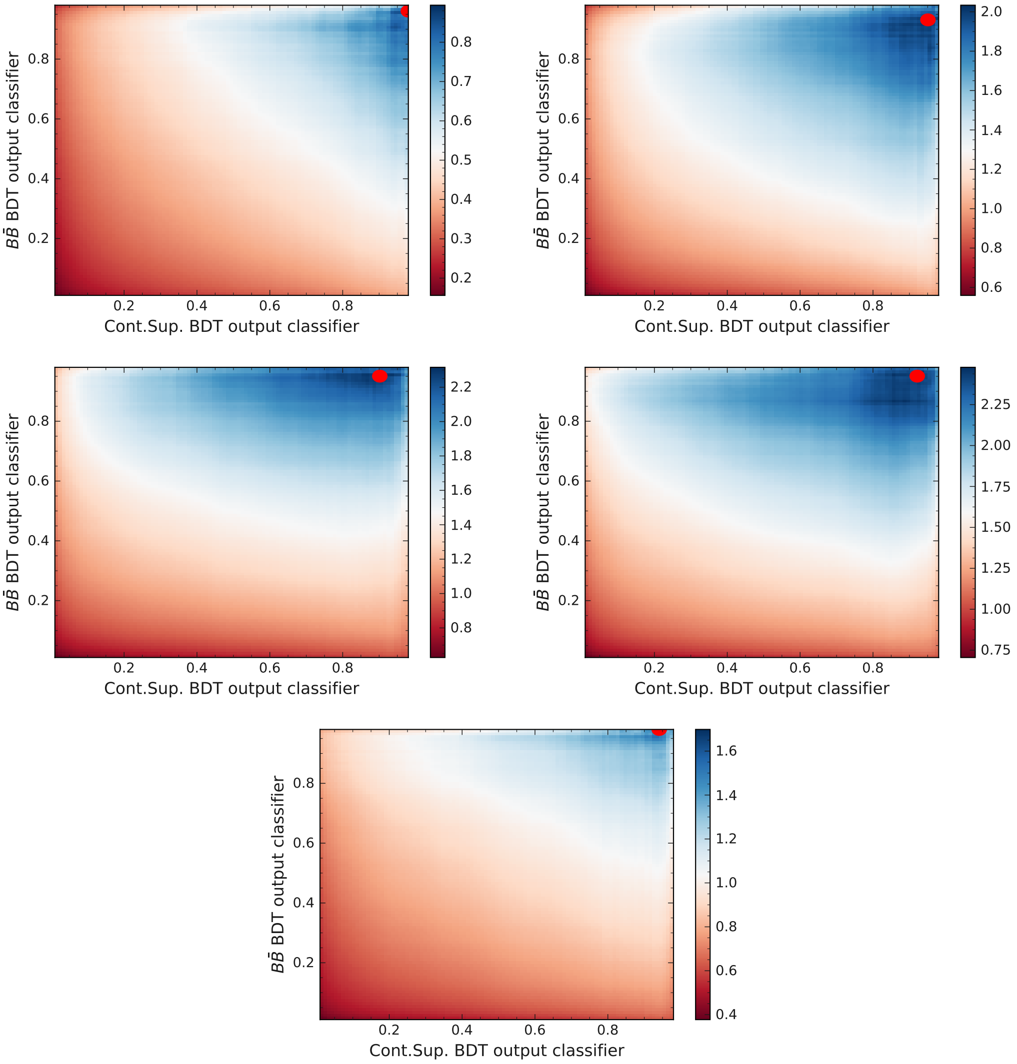


Figure A.25: BDT output classifier FOM distributions for the $B^0 \rightarrow \rho^\pm \mu^\mp \nu_\mu$ signal decay channel. From left-to-right then top-to-bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. The red dot indicates the optimal selection.

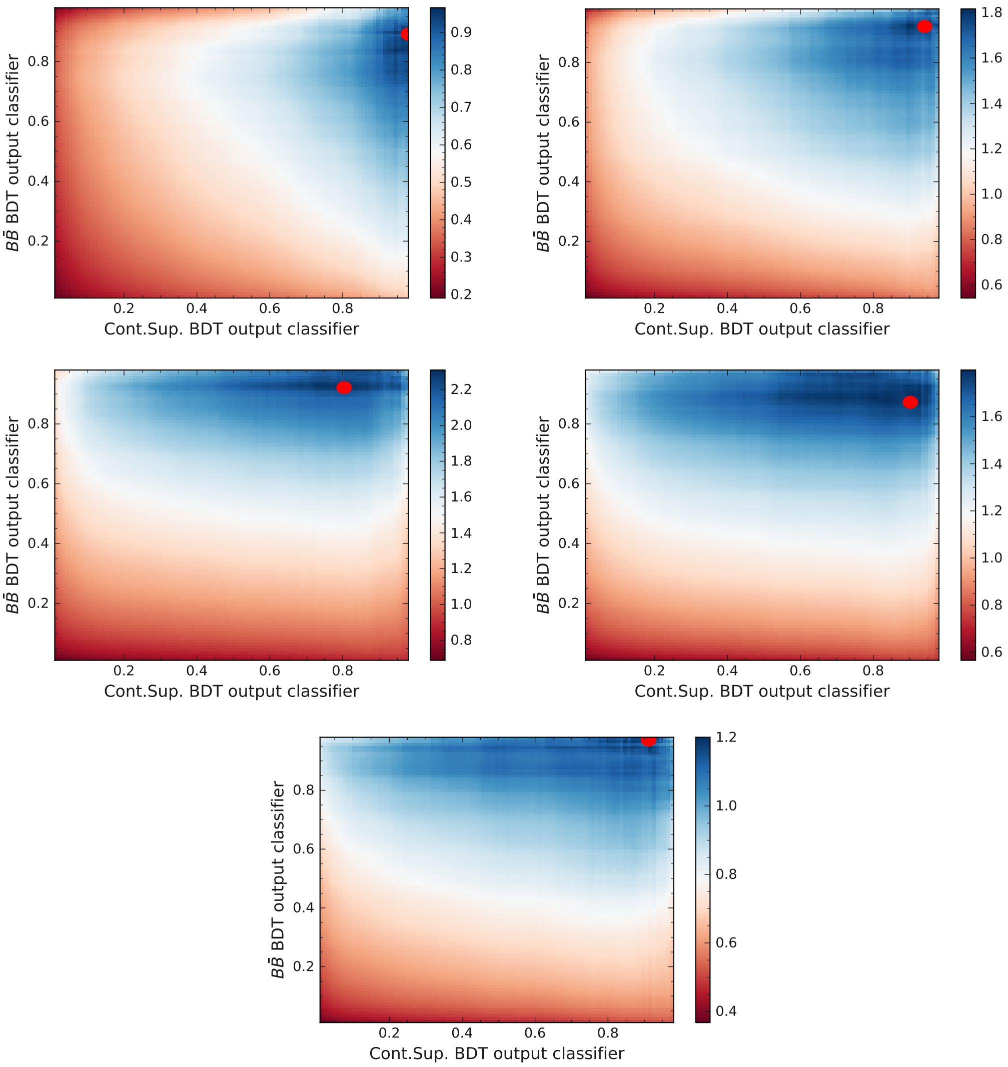


Figure A.26: BDT output classifier FOM distributions for the $B^{\pm} \to \rho^{0} e^{\pm} \nu_{e}$ signal decay channel. From left-to-right then top-to-bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. The red dot indicates the optimal selection.

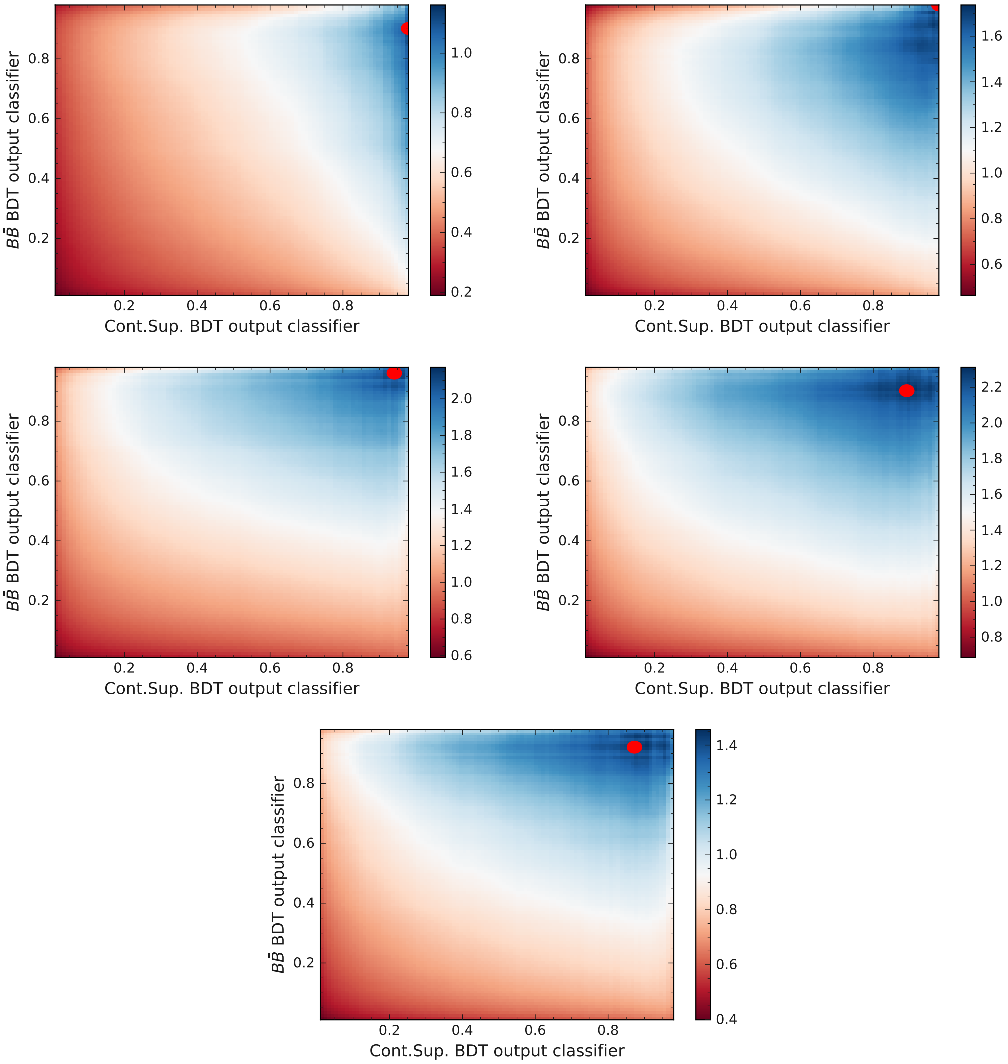
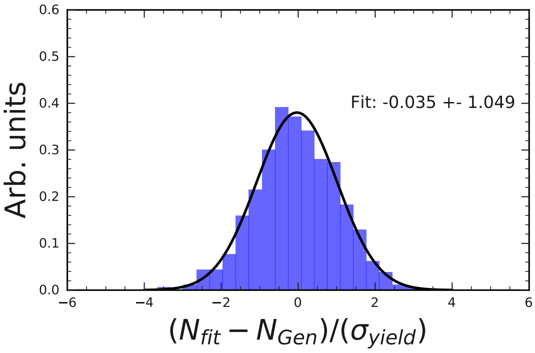
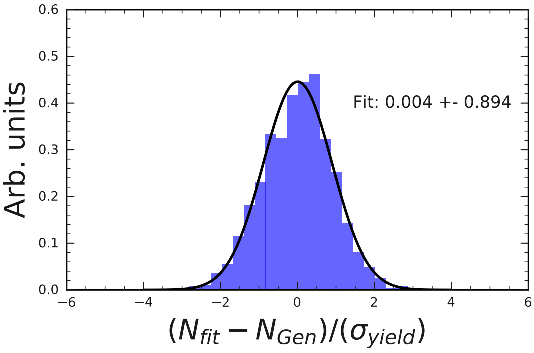


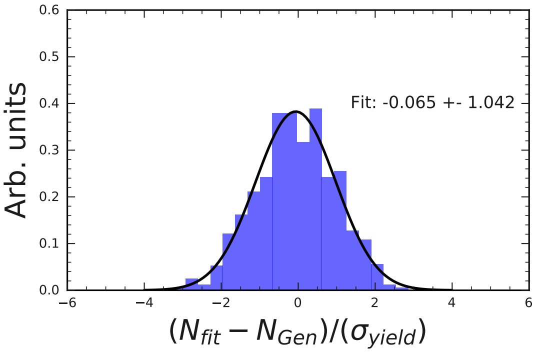
Figure A.27: BDT output classifier FOM distributions for the $B^{\pm} \to \rho^{0}\mu^{\pm}\nu_{\mu}$ signal decay channel. From left-to-right then top-to-bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. The red dot indicates the optimal selection.

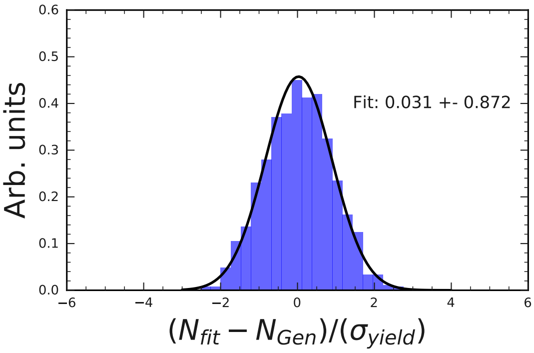
A.8 Signal extraction fit

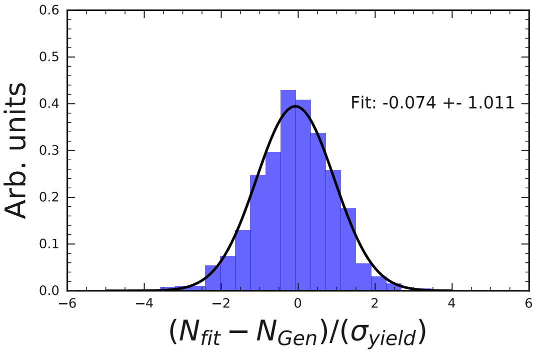
A.8.1 Pull distributions

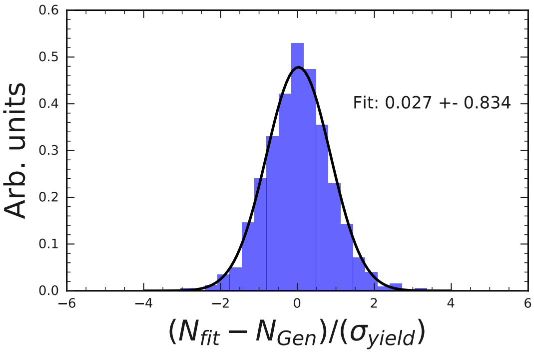












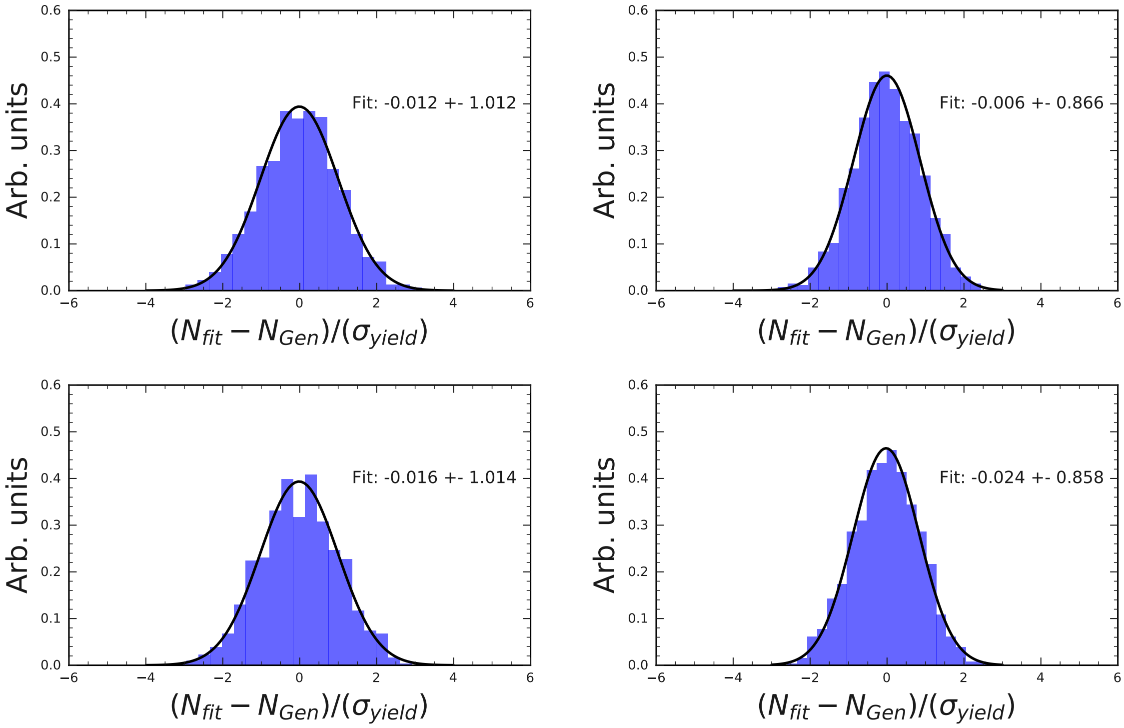
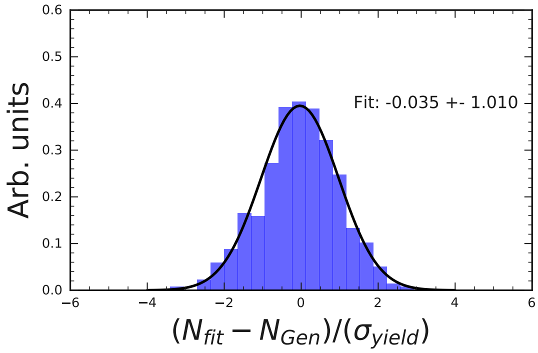
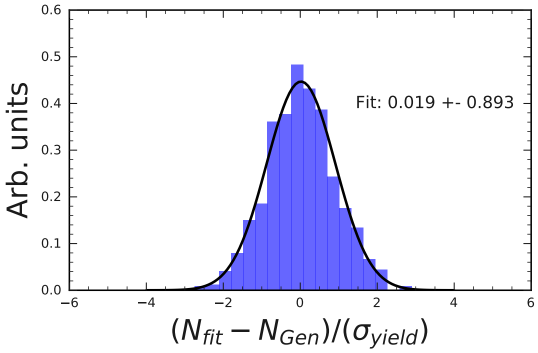
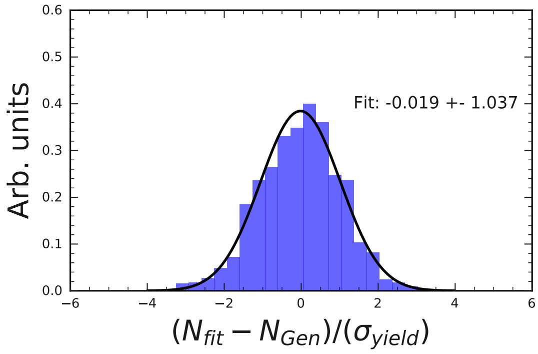
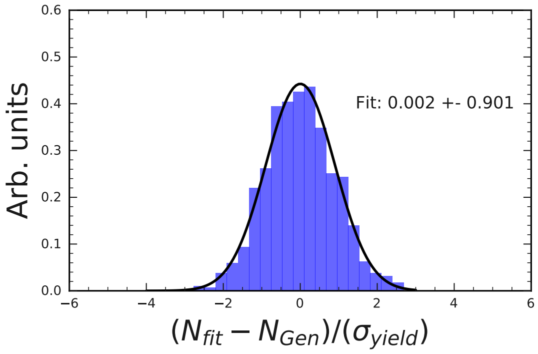


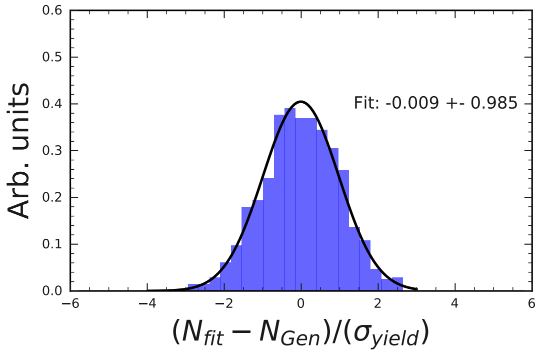
Figure A.28: Pull distributions and the fitted normal Gaussians for the $B^0 \to \rho^\pm e^\mp \nu_e$ fits, created using $n = 1000$ MC toys. Left: signal yield. Right: background yield. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$.

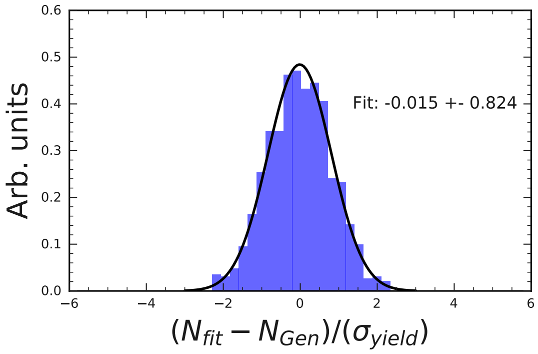

Fit: -0.035 +- 1.010
Arb. units
$(N_{fit} - N_{Gen})/(\sigma_{yield})$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−6
−4
−2
0
2
4
6

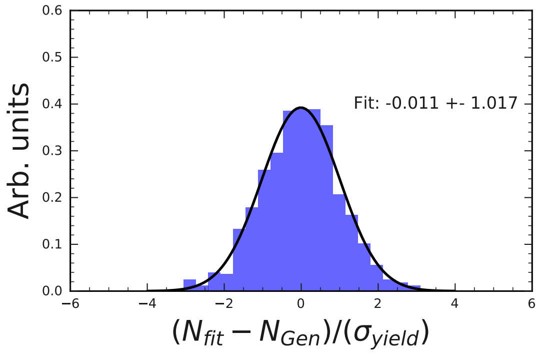

Fit: 0.019 +- 0.893
Arb. units
$(N_{fit} - N_{Gen})/(\sigma_{yield})$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−6
−4
−2
0
2
4
6

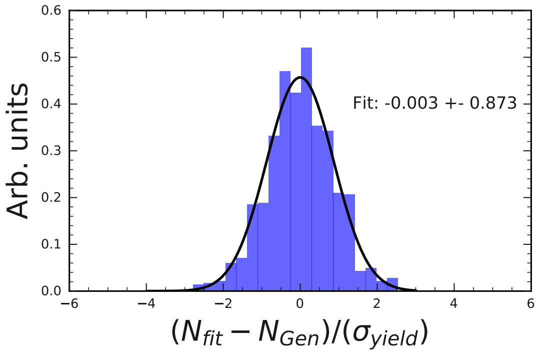

Fit: -0.019 +- 1.037
Arb. units
$(N_{fit} - N_{Gen})/(\sigma_{yield})$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−6
−4
−2
0
2
4
6


Fit: 0.002 +- 0.901
Arb. units
$(N_{fit} - N_{Gen})/(\sigma_{yield})$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−6
−4
−2
0
2
4
6


Fit: -0.009 +- 0.985
Arb. units
$(N_{fit} - N_{Gen})/(\sigma_{yield})$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−6
−4
−2
0
2
4
6


Fit: -0.015 +- 0.824
Arb. units
$(N_{fit} - N_{Gen})/(\sigma_{yield})$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−6
−4
−2
0
2
4
6


Fit: -0.011 +- 1.017
Arb. units
$(N_{fit} - N_{Gen})/(\sigma_{yield})$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−6
−4
−2
0
2
4
6


Fit: -0.003 +- 0.873
Arb. units
$(N_{fit} - N_{Gen})/(\sigma_{yield})$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−6
−4
−2
0
2
4
6

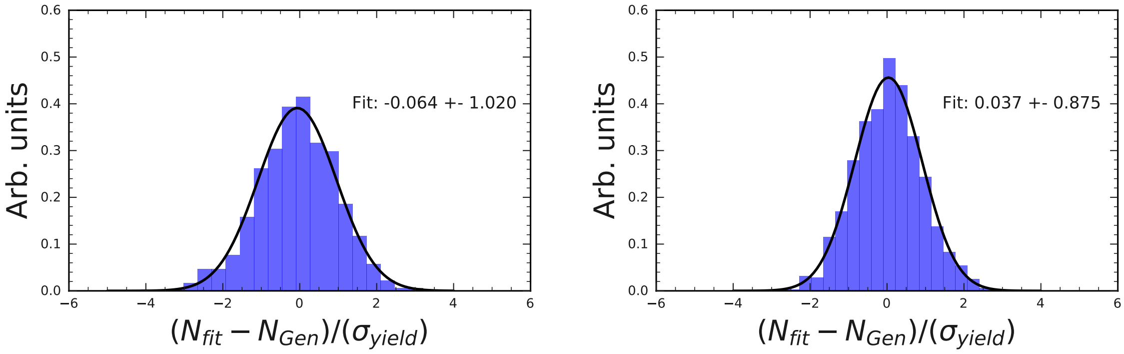
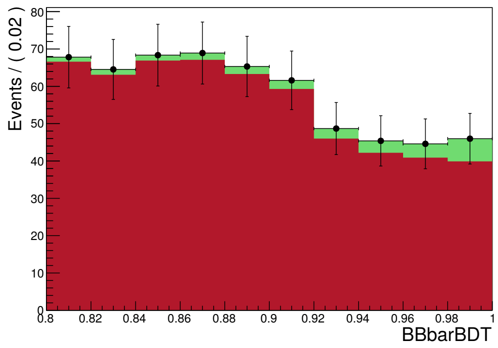
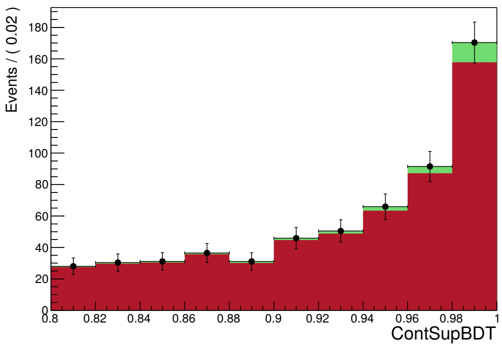
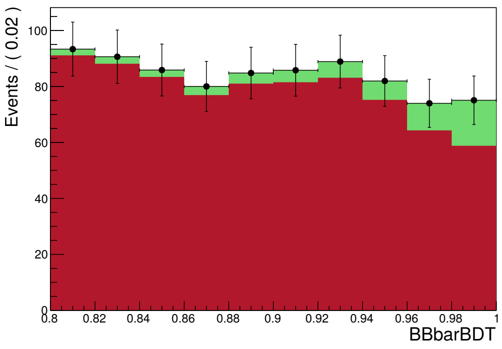


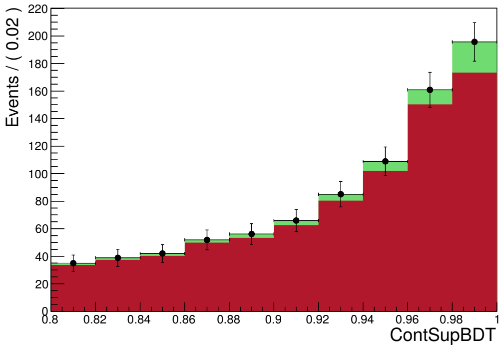
Figure A.29: Pull distributions and the fitted normal Gaussians for the $B^{\pm} \to \rho^{0} e^{\pm} \nu_e$ fits, created using $n = 1000$ MC toys. Left: signal yield. Right: background yield. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$.

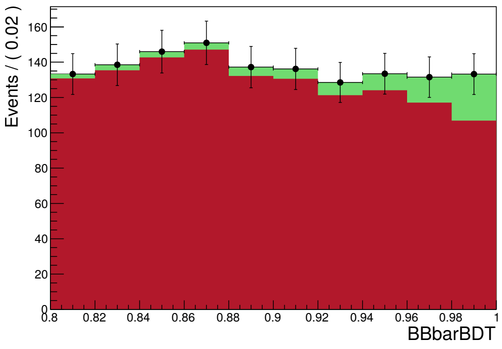
A.8.2 Asimov fits

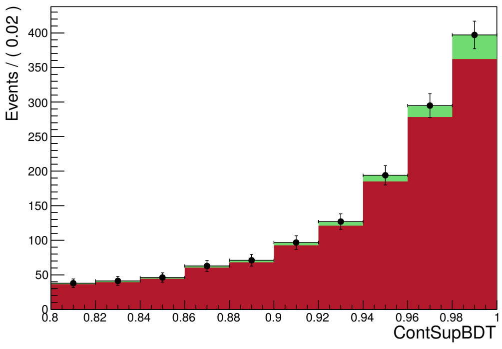












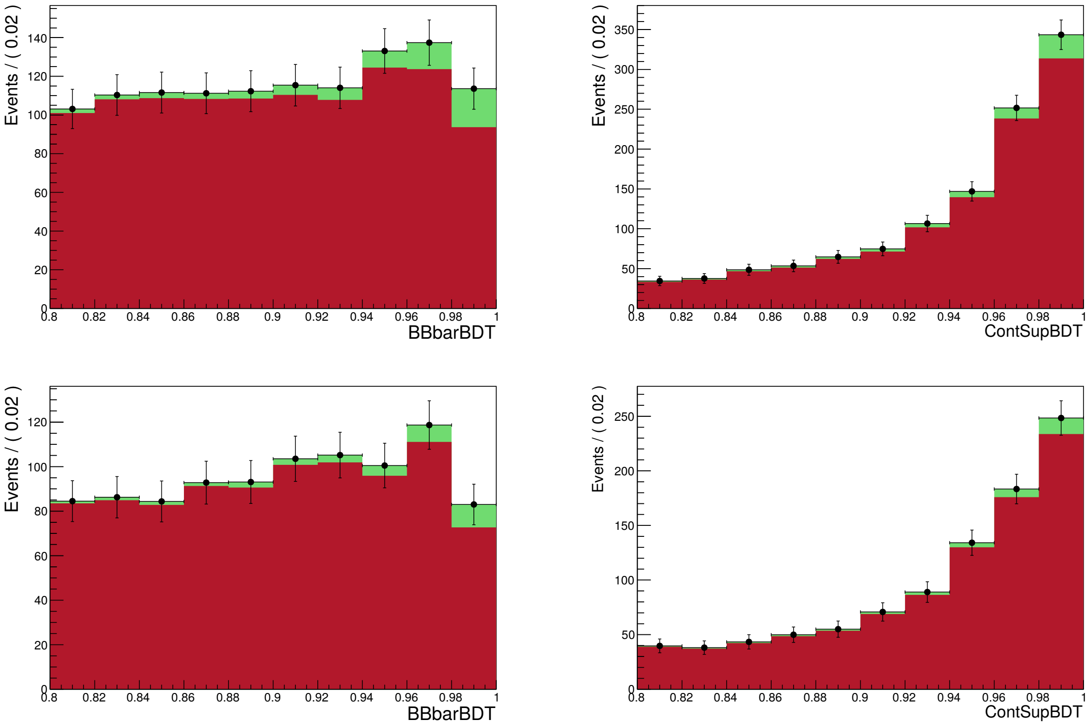
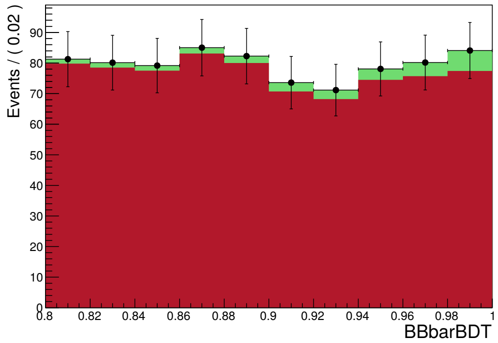
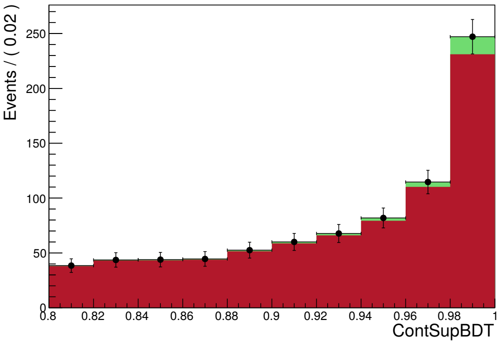
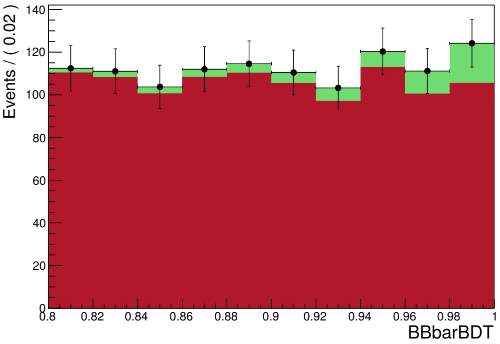
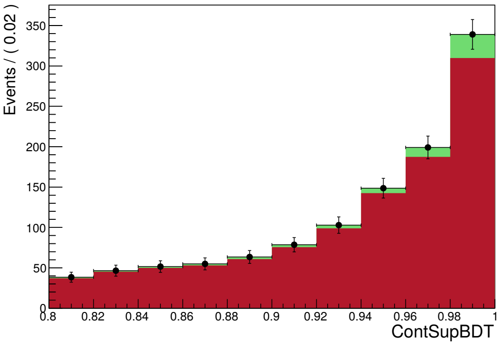


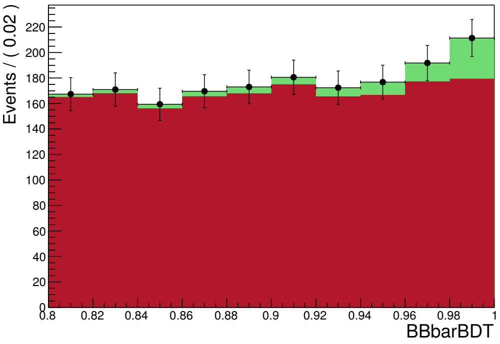
Figure A.30: Asimov fits for the $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ channel. Left: $B\overline{B}$ suppression BDT output classifier fit. Right: continuum suppression BDT output classifier fit. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$.

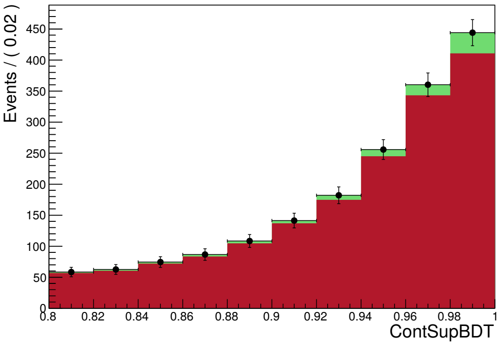

Events / (0.02)
0
10
20
30
40
50
60
70
80
90
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
BBbarBDT


Events / (0.02)
0
50
100
150
200
250
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
ContSupBDT


Events / (0.02)
0
20
40
60
80
100
120
140
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
BBbarBDT


Events / (0.02)
0
50
100
150
200
250
300
350
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
ContSupBDT


Events / (0.02)
0
20
40
60
80
100
120
140
160
180
200
220
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
BBbarBDT


Events / (0.02)
0
50
100
150
200
250
300
350
400
450
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
ContSupBDT

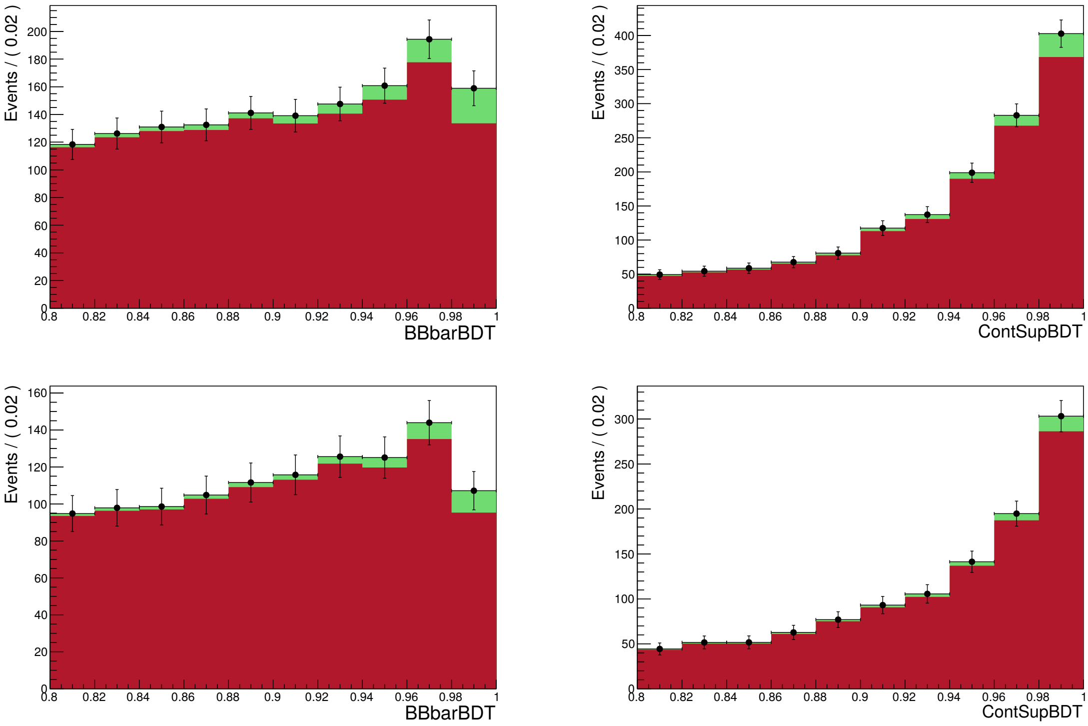
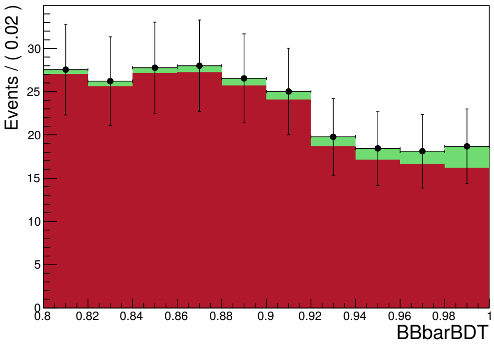
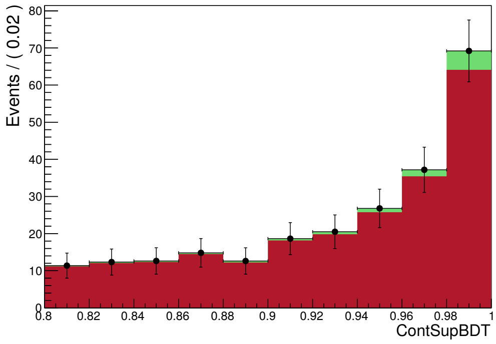
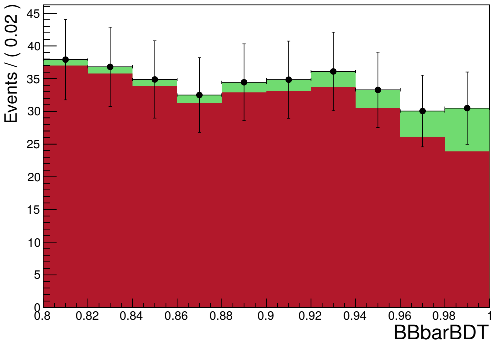


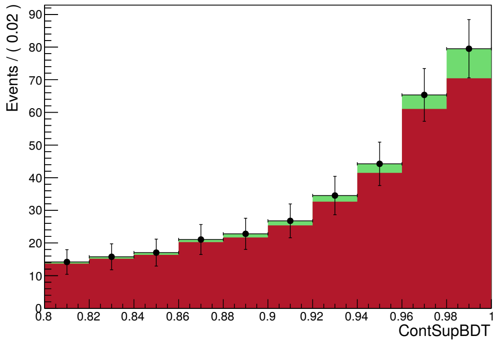
Figure A.31: Asimov fits for the $B^{\pm} \to \rho^{0} e^{\pm} \nu_e$ channel. Left: $B\overline{B}$ suppression BDT output classifier fit. Right: continuum suppression BDT output classifier fit. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$.

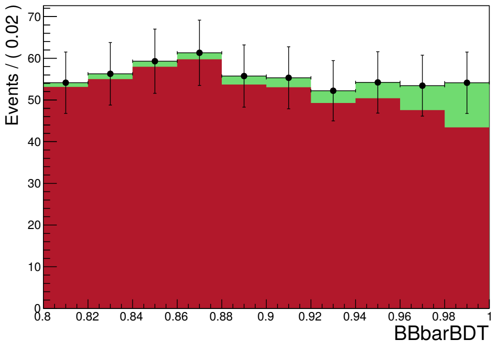
A.8.3 Fit results

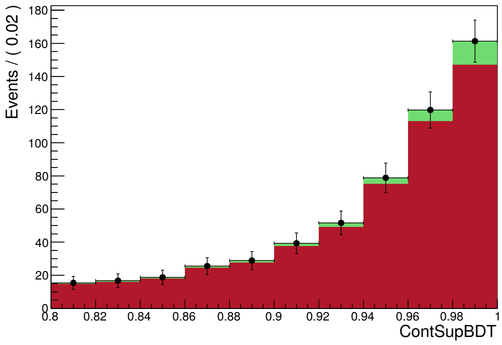












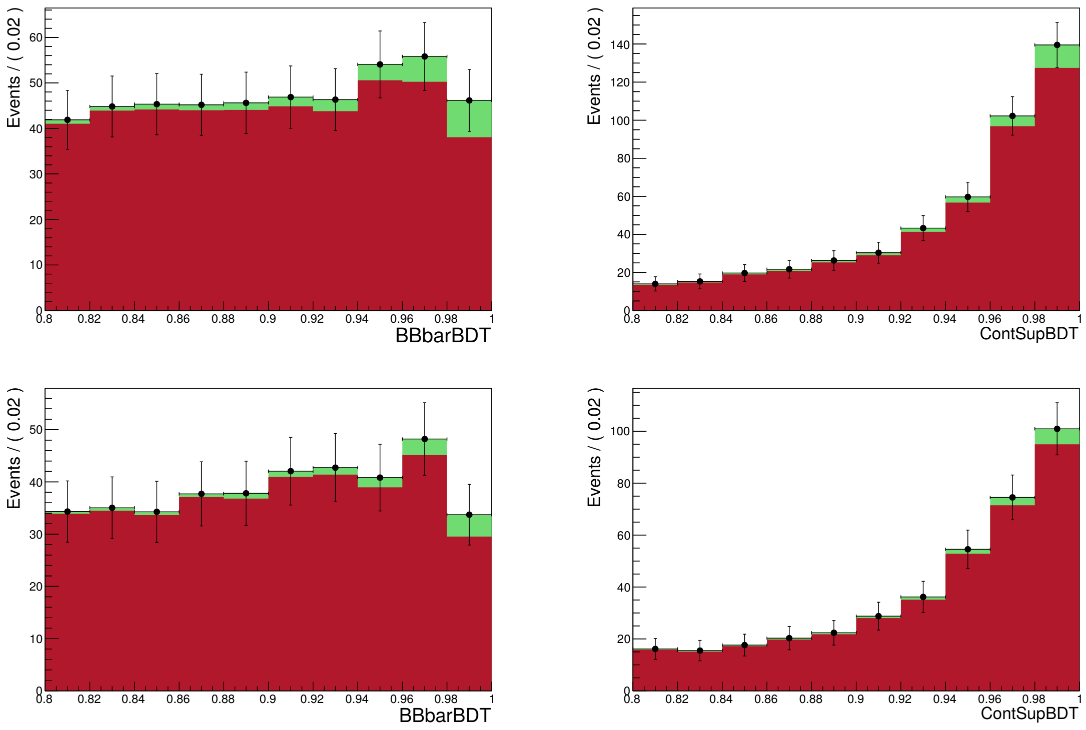
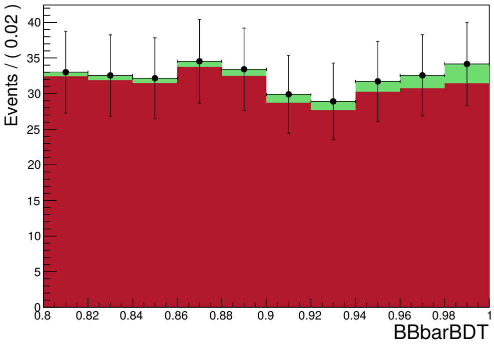
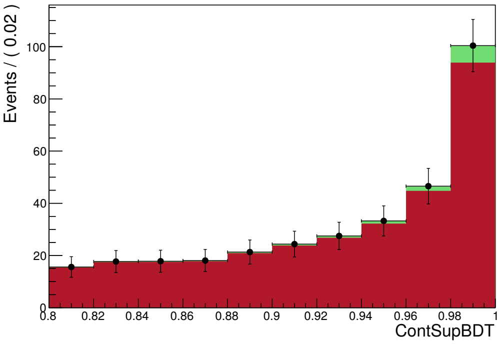
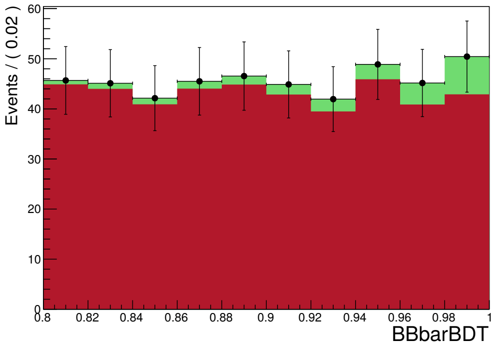
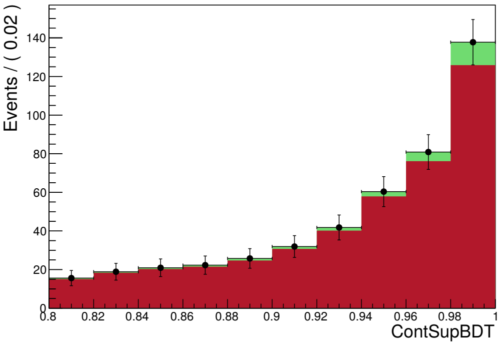


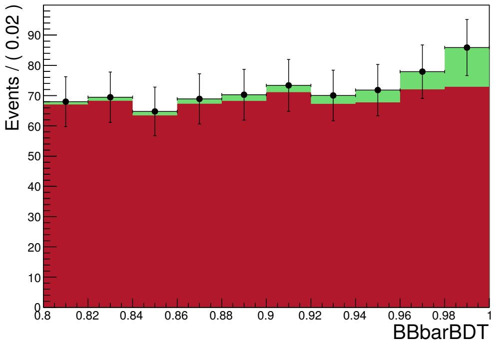
Figure A.32: Fit results for the $B^0 \to \rho^{\pm} e^{\mp} \nu_e$ channel. Left: $B\overline{B}$ suppression BDT output classifier fit. Right: continuum suppression BDT output classifier fit. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$.

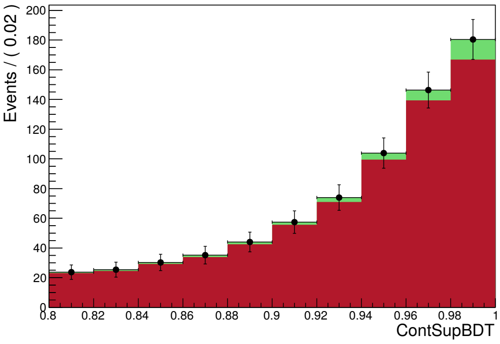

Events / (0.02)
0
5
10
15
20
25
30
35
40
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
BBbarBDT


Events / (0.02)
0
20
40
60
80
100
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
ContSupBDT


Events / (0.02)
0
10
20
30
40
50
60
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
BBbarBDT


Events / (0.02)
0
20
40
60
80
100
120
140
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
ContSupBDT


Events / (0.02)
0
10
20
30
40
50
60
70
80
90
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
BBbarBDT


Events / (0.02)
0
20
40
60
80
100
120
140
160
180
200
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1
ContSupBDT

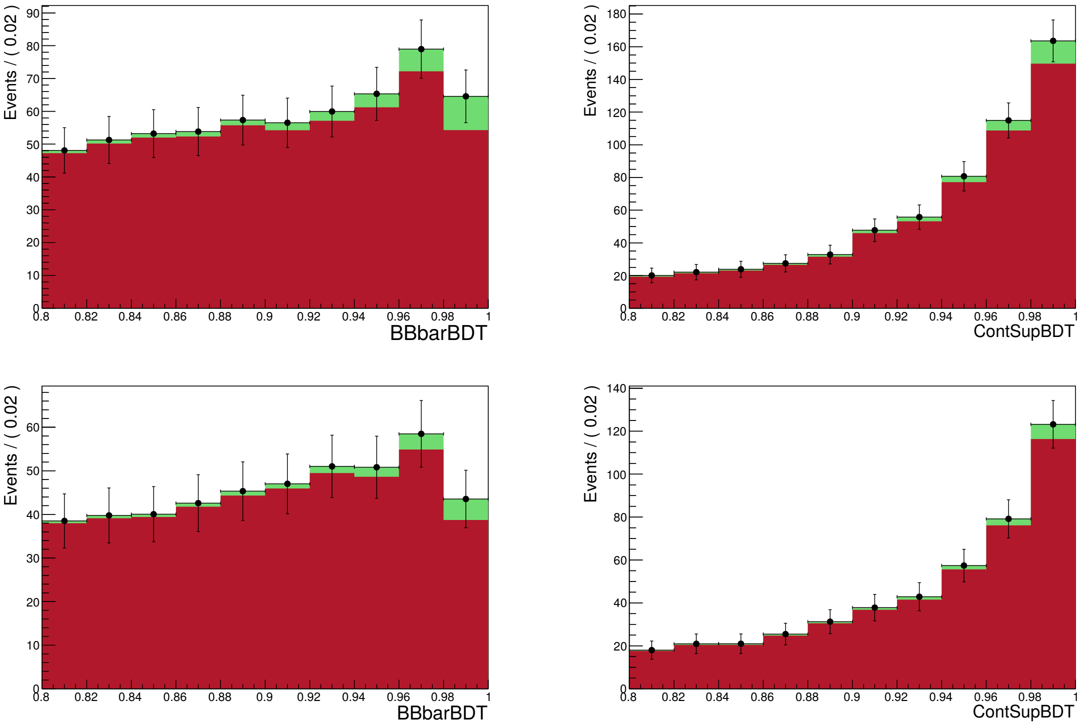


Figure A.33: Fitresults for the $B^{\pm} \rightarrow \rho^0 e^{\pm} \nu_e$ channel. Left: $B\overline{B}$ suppression BDT output classifier fit. Right: continuum suppression BDT output classifier fit. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$.

List of Figures

2.1 Schematic depiction of the Standard Model of particle physics [11]. 4
2.2 Feynman diagram for the process $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ [15]. 5
2.3 Tree-level Feynman diagrams for the studied decays. 5
2.4 The form factor parameters. 6

3.1 Schematic drawing of the SuperKEKB accelerator [21]. 9
3.2 Schematic drawing of the Belle II detector, comparing to its predecessor Belle [8]. The most important upgrades compared to Belle are the improved PID system and the vertex detector. 10
3.3 Schematic drawing of a Boosted Decision Tree [24]. In this example, the two colours indicate the background and signal categories. Each circle represents a leaf of the tree, while each arrow indicates a certain selection on one of the input variables. Each level in the diagram represents a decision layer in the tree. 12

4.1 Three different track segment hits, with the priority wire coloured green [25]. Left: left passage. Middle: right passage. Right: undecided passage. The squares represent individual track segments, lined up in rows as a schematic view of the CDC. The black lines indicate the reconstructed track through the track segments. 14

5.1 $\theta_{\gamma\gamma}$ selection for improved π^0 background rejection. 18
5.2 Number of cleaned tracks distributions. The black line indicates the selection of number of cleaned tracks bigger than 4. 19
5.3 θ_{miss} distributions. The black lines indicates the selection on the detector acceptance region $17° < \theta_{\text{miss}} < 156°$. 20
5.4 ρ^0 vertex fit χ^2 probability distributions. The black line indicates the selection at 10^{-10}. 21
5.5 M_{bc} (left) and ΔE (right) distributions for $B^\pm$ signal decays with the defined fit region $|\Delta E| < 1.0$ GeV and $5.0 < M_{bc} < 5.3$ GeV. 23
5.6 $\cos\theta_{BY}$ distributions and the selection between $-1.1 < \cos\theta_{BY} < 1.1$. 25
5.7 2D distributions for p^*_ρ and p^*_ℓ for background (left) and signal (right) in $B^\pm \to \rho^0 e^\pm \nu_e$. 26
5.8 m_ρ distributions for the four signal decays and the selection $0.55 \leq m_\rho \leq 1.30$ GeV. . 27
5.9 Signal efficiency distributions of q^2 and the three decay angles $\cos\theta_l$, $\cos\theta_V$ and χ, for $B^\pm$ after pre-selections. 28
5.10 Distributions for M_{bc} and ΔE for the $B^0 \to \rho^\pm e^\mp \nu_e$ in the $q1$ bin (top), where the background is dominated by continuum events, and the $q4$ bin (bottom) where $B\overline{B}$ background dominates. 30

5.11 Normalized distributions for the four highest ranking variables in the $B^{\pm} \to \rho^{0} e^{\pm} \nu_e$ $q1$ bin continuum suppression BDT. From top left to bottom right: $\cos\theta^{z}_{T_B}$, h_2^{oo}, h_{20}^{so}, and h_{02}^{so}. . . . 31
5.12 Correlation matrices for the $B^{\pm} \to \rho^{0} e^{\pm} \nu_e$ continuum suppression BDT in the $q^2 \leq 4.0\ \mathrm{GeV}^2$ bin. . . . 33
5.13 Overtraining check for the $B^{\pm} \to \rho^{0} e^{\pm} \nu_e$ continuum suppression BDT in the $q^2 \leq 4.0\ \mathrm{GeV}^2$ bin. The p-value shows no significant difference between training and testing data, so overtraining is not present. . . . 33
5.14 ROC curves for the trained BDTs in the five q^2 bins for all four decay channels. The curves all show a similar shape, indicating comparable BDT performance. . . . 34
5.15 Normalized distributions for the four highest ranking variables in the $B^{\pm} \to \rho^{0} e^{\pm} \nu_e$ $q1$ bin $B\overline{B}$ suppression BDT. From top left to bottom right: p_l^*, B vertex fit χ^2 probability, m_ρ, and $\cos\theta_{\pi\pi}$. . . . 35
5.16 Correlation matrices for the $B^{0} \to \rho^{\pm} e^{\mp} \nu_e$ $B\overline{B}$ suppression BDT in the $q^2 \leq 4.0\ \mathrm{GeV}^2$ bin. . . . 36
5.17 Overtraining check for the $B^{0} \to \rho^{\pm} e^{\mp} \nu_e$ $B\overline{B}$ suppression BDT in the $q^2 \leq 4.0\ \mathrm{GeV}^2$ bin. The p-test shows no significant difference between training and testing performance, so overtraining is not present. . . . 37
5.18 ROC curves for the trained $B\overline{B}$ BDTs in the five q^2 bins for all four decay channels. . . 37
5.19 2D FOM distributions for the BDT selection optimization for the signal decay channel $B^{0} \to \rho^{\pm} e^{\mp} \nu_e$. . . . 38
5.20 Distributions for M_{bc} and ΔE for the $B^{0} \to \rho^{\pm} e^{\mp} \nu_e$ (top) and $B^{0} \to \rho^{\pm} \mu^{\mp} \nu_\mu$ (bottom) in the bin $12.0 < q^2 \leq 16.0\ \mathrm{GeV}^2$. The shape of the signal events compared to the background events is almost identical. . . . 39

6.1 Top: Asimov fit for the $B^{0} \to \rho^{\pm} e^{\mp} \nu_e$ $q2$ decay, scaled to 16 fb^{-1}. The green indicates the fitted signal, while the red is the fitted background shape. Bottom: Pull distributions and their fitted normal Gaussians for the $B^{0} \to \rho^{\pm} e^{\mp} \nu_e$ $q2$ fit. . . . 43
6.2 Fit results for the $B^{0} \to \rho^{\pm} e^{\mp} \nu_e$ signal decay channel in the $q1$ (top) and $q4$ (bottom) bins. 44

7.1 Distributions for nCleanedTracks using the 689 pb^{-1} experiment 7 data sample after pre-selections, before (top) and after (bottom) best candidate selection using the whole 3.124 fb^{-1} proc9 data sample. Large disagreements between data and MC are visible before the best candidate selection, mostly due to the presence of BhaBha scattering and the potentially higher of number beam background photons in data. The disargeement mostly disappears after the best candidate selection. . . . 50
7.2 Lepton candidate dr and dz distributions for the proc9 data sample after pre-selections and best candidate selection. While the MC peaks at 0, the data shows a shifted interaction point in both radial and z-direction. . . . 51
7.3 p_ℓ^* and p_ρ^* distributions. A deficit of data is visible in the region where MC predicts the most events. . . . 52
7.4 Distributions for R_2 using the whole proc9 data sample corresponding to 3.124 fb^{-1}. Top: unscaled continuum. Bottom: scaled-up continuum by 20%. . . . 53

7.5 BDT output classifiers for the $B^0 \to \rho^\pm e^\mp \nu_e$ channel in the $q1$ bin, before (top) and after (bottom) scaling the continuum. 54
7.6 BDT output classifiers for the $B^0 \to \rho^\pm e^\mp \nu_e$ channel in the $q1$ and $q4$ bins. 56

A.1 $|\Delta t_{\gamma\gamma}|$ distributions for reconstructed π^0. The regular features are due to the timing resolution of the detector. 63
A.2 M_{bc} (left) and ΔE (right) distributions for B^0 signal decays with the defined fit region $|\Delta E| < 1.0$ GeV and $5.0 < M_{bc} < 5.3$ GeV. 64
A.3 2D distributions for p^*_ρ and p^*_ℓ for background (left) and signal (right) in the remaining signal decay channels. 65
A.4 Signal efficiency distributions of q^2 and the three decay angles $\cos\theta_l$, $\cos\theta_V$ and χ, for B^0 after pre-selections. 66
A.5 Generator $E^*_{D^*}$ and E^*_π (left) and $p^*_{D^*}$ and p^*_π (right) distributions and fitted values. . . 69
A.6 Reconstructed $m_{D^{*0}}$ from the D^* veto. 71
A.7 Reconstructed $m_{D^{*\pm}}$ from the D^* veto. 72
A.8 Normalized shapes for the $B^0 \to \rho^\pm e^\mp \nu_e$ continuum suppression BDT input variables. 74
A.9 Normalized shapes for the $B^0 \to \rho^\pm \mu^\mp \nu_\mu$ continuum suppression BDT input variables. 76
A.10 Normalized shapes for the $B^\pm \to \rho^0 e^\pm \nu_e$ continuum suppression BDT input variables. 77
A.11 Normalized shapes for the $B^\pm \to \rho^0 \mu^\pm \nu_\mu$ continuum suppression BDT input variables. 78
A.12 Normalized shapes for the $B^0 \to \rho^\pm e^\mp \nu_e$ $B\overline{B}$ suppression BDT input variables. 79
A.13 Normalized shapes for the $B^0 \to \rho^\pm \mu^\mp \nu_\mu$ $B\overline{B}$ suppression BDT input variables. 80
A.14 Normalized shapes for the $B^\pm \to \rho^0 e^\pm \nu_e$ $B\overline{B}$ suppression BDT input variables. 81
A.15 Normalized shapes for the $B^\pm \to \rho^0 \mu^\pm \nu_\mu$ $B\overline{B}$ suppression BDT input variables. 82
A.16 Correlation matrices for the $B^0 \to \rho^\pm e^\mp \nu_e$ continuum suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$. 93
A.17 Correlation matrices for the $B^0 \to \rho^\pm \mu^\mp \nu_\mu$ continuum suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$. 95
A.18 Correlation matrices for the $B^\pm \to \rho^0 e^\pm \nu_e$ continuum suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$. 97
A.19 Correlation matrices for the $B^\pm \to \rho^0 \mu^\pm \nu_\mu$ continuum suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$. 99
A.20 Correlation matrices for the $B^0 \to \rho^\pm e^\mp \nu_e$ $B\overline{B}$ suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$. 101
A.21 Correlation matrices for the $B^0 \to \rho^\pm \mu^\mp \nu_\mu$ $B\overline{B}$ suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$. 103
A.22 Correlation matrices for the $B^\pm \to \rho^0 e^\pm \nu_e$ $B\overline{B}$ suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$. 105
A.23 Correlation matrices for the $B^\pm \to \rho^0 \mu^\pm \nu_\mu$ $B\overline{B}$ suppression BDTs. From top to bottom: $q1$,$q2$,$q3$,$q4$ and $q5$. 107
A.24 BDT output classifier FOM distributions for the $B^0 \to \rho^\pm e^\mp \nu_e$ signal decay channel. From left-to-right then top-to-bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. The red dot indicates the optimal selection. 108

A.25 BDT output classifier FOM distributions for the $B^0 \longrightarrow \rho^{\pm}\mu^{\mp}\nu_{\mu}$ signal decay channel. From left-to-right then top-to-bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. The red dot indicates the optimal selection. 109
A.26 BDT output classifier FOM distributions for the $B^{\pm} \longrightarrow \rho^0 e^{\pm}\nu_e$ signal decay channel. From left-to-right then top-to-bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. The red dot indicates the optimal selection. 110
A.27 BDT output classifier FOM distributions for the $B^{\pm} \longrightarrow \rho^0 \mu^{\pm}\nu_{\mu}$ signal decay channel. From left-to-right then top-to-bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. The red dot indicates the optimal selection. 111
A.28 Pull distributions and the fitted normal Gaussians for the $B^0 \longrightarrow \rho^{\pm}e^{\mp}\nu_e$ fits, created using $n = 1000$ MC toys. Left: signal yield. Right: background yield. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. 113
A.29 Pull distributions and the fitted normal Gaussians for the $B^{\pm} \longrightarrow \rho^0 e^{\pm}\nu_e$ fits, created using $n = 1000$ MC toys. Left: signal yield. Right: background yield. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. 115
A.30 Asimov fits for the $B^0 \longrightarrow \rho^{\pm}e^{\mp}\nu_e$ channel. Left: $B\overline{B}$ suppression BDT output classifier fit. Right: continuum suppression BDT output classifier fit. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. 117
A.31 Asimov fits for the $B^{\pm} \longrightarrow \rho^0 e^{\pm}\nu_e$ channel. Left: $B\overline{B}$ suppression BDT output classifier fit. Right: continuum suppression BDT output classifier fit. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. 119
A.32 Fit results for the $B^0 \longrightarrow \rho^{\pm}e^{\mp}\nu_e$ channel. Left: $B\overline{B}$ suppression BDT output classifier fit. Right: continuum suppression BDT output classifier fit. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. 121
A.33 Fitresults for the $B^{\pm} \longrightarrow \rho^0 e^{\pm}\nu_e$ channel. Left: $B\overline{B}$ suppression BDT output classifier fit. Right: continuum suppression BDT output classifier fit. From top to bottom: $q1$, $q2$, $q3$, $q4$, and $q5$. 123

List of Tables

5.1 Pre-selections for charged particles. 16
5.2 Efficiencies per performed selection for all 4 decay channels in the $q1$ bin. 40

6.1 Obtained yields for the two fitted signal decay modes and q^2 bins, and their statistical uncertainties. 45
6.2 Obtained branching fractions for the e^- signal decay modes and their statistical uncertainty. 45
6.3 Obtained Asimov yields for the e^- signal decay modes and q^2 bins using 16 fb^{-1} MC, and their statistical uncertainties. 45
6.4 Obtained branching fractions for the e^- signal decay modes and their statistical uncertainty using 16 fb^{-1} MC and the Asimov fit yields. 45

A.1 Efficiencies per performed selection for all 4 decay channels in the $q2$ bin. 67
A.2 Efficiencies per performed selection for all 4 decay channels in the $q3$ bin. 67
A.3 Efficiencies per performed selection for all 4 decay channels in the $q4$ bin. 68
A.4 Efficiencies per performed selection for all 4 decay channels in the $q5$ bin. 68
A.5 P-tests for difference between training and testing performance of the trained B^0 continuum suppression BDTs. No overtraining is present. 91
A.6 P-tests for difference between training and testing performance of the trained $B^{\pm}$ continuum suppression BDTs. No overtraining is present. 91
A.7 P-tests for difference between training and testing performance of the trained B^0 $B\overline{B}$ suppression BDTs. No overtraining is present. 91
A.8 P-tests for difference between training and testing performance of the trained $B^{\pm}$ $B\overline{B}$ suppression BDTs. No overtraining is present. 91