Untagged Exclusive Analysis of the Semileptonic Decay $B \to \pi \ell \nu$ from Belle II Data in Preparation for $|V_{ub}|$ Extraction

Svenja Katharina Granderath

Masterarbeit in Physik
angefertigt im Physikalischen Institut

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

September 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

This thesis would not have been possible without the kind support of the people surrounding me. Here I would like to give a special mention to some of the people that have accompanied me during the last year and have made this such a special time.

First of all, I would like to thank my supervisor Jochen Dingfelder for introducing me to the world of B Physics. Little did I know that I was going to be doing my master thesis in flavour physics when I signed up for giving a talk on "Hot Topics in B Physics" under his supervision. Not to mention, that I was going to have the chance to present the status of my analysis at the collaboration meeting in Japan. I am very grateful to Jochen for the opportunity to write my master thesis in the Belle II group in Bonn.

I would also like to thank Peter Lewis, who joined the Belle II group shortly before me, but right from the beginning always had helpful advice on hand. No matter how many times a day I would go over to his office to ask a question or to tell him that I ran into a problem, he was always happy to explain, discuss and look for a solution. The other person that I could always ask for advice is Stephan Duell, and I would like to express my thanks to him. Especially, but not only, at the beginning of my analysis he had to endure countless questions, but never grew tired of answering them. But Stephan did not only take care of me – which brings me to Lex.

Lex and I bonded immediately over starting our analyses at the same time and being new to Belle II. I would like to thank him for all the discussions, laughs, solving problems together and, of course, our office-singing. I will not forget exploring Tokyo together. I would also like to thank the rest of the Belle II group for all the support and fun during the last year: Alexander, Christian, Mario, Lu, Jun and Jonathan. My thanks also go out to the friends that I met during my studies: Srijan, Beth, Asia, Georgios, Elle and Bence. Without you it would not have been the same.

Finally, I would like to say a huge thank you to my parents, my grandparents and my brother for all the support they have given me, not just only during the last year, but throughout all my studies. During the bad times they listen to my problems and cheer me up, and during the good times they celebrate with me. I'm very thankful that you play such a big role in my life.

Contents

CHAPTER 1

Introduction

With Thomson's discovery of the electron in 1897 a first big step in our understanding of the fundamental building blocks of our universe was taken [1]. By 1956, the proton, positron and neutrino had been discovered and in the 1970s a common theory called the Standard Model (SM) was formulated, containing the elementary particles discovered so far, predicting a three-quark model and describing the interactions between these particles. Within the same decade, the up, down, strange and charm quarks were discovered. The SM went on to successfully predict particles such as the bottom quark, which was first discovered through the $\Upsilon(1S)$ observation in 1977 [2]. The SM has withstood many stringent high-precision tests during the last 50 years. It also succesfully predicted the existence of its last missing puzzle piece: the Higgs boson, which was discovered in 2012 [3].

On the other hand, the SM lacks explanations for some pending questions in particle physics. For example, it is unable to incorporate the fourth fundamental force, gravity, and cannot explain the origins of dark energy and dark matter. The SM also fails to justify the existence of the matter-antimatter asymmetry in our universe. All of the above deficiencies may in the future be explained through extensions to the SM, but explanations could also lie in new physics beyond the Standard Model.

In order to constrain the SM, and therefore the possibility for the existence of new physics, precision measurements of the SM parameters are performed. This is the case, for example, for the elementary particle masses and the parameters of the Cabibbo-Kobayashi-Maskawa (CKM) matrix describing quark mixing with the weak interaction. The elements of the CKM matrix have been extensively studied by the B-factories BaBar and Belle, as well as by LHCb, and for some of the elements a tension between the exclusive and inclusive measurement results persists. These two determination methods rely on different model-dependent theoretical uncertainties. The inclusive method is not concerned with the specific identity of the final state, while the exclusive method reconstructs a specific final state. The CKM matrix elements $|V_{ub}|$ and $|V_{cb}|$ differ by $\sim 3.5\sigma$ and $\sim 3\sigma$, respectively, in the exclusive as compared to the inclusive measurement [4]. The $|V_{ub}|$ and $|V_{cb}|$ results from various exclusive modes, along with a two-dimensional fit to these results, are shown in Figure 1.1. An inclusive measurement is also displayed and the discrepancy between exclusive and inclusive measurements is discernible.

With the commissioning of the Belle II experiment at the SuperKEKB accelerator in March 2019, new inclusive and exclusive measurements of $|V_{ub}|$ and $|V_{cb}|$ become feasible in order to test whether the current tension prevails, and in that case to uncover possible origins. The tension could be caused by errors in the experimental analysis, or the theoretical predictions that go into the CKM matrix element measurement. However, it could also be a hint of new physics that is not currently described by

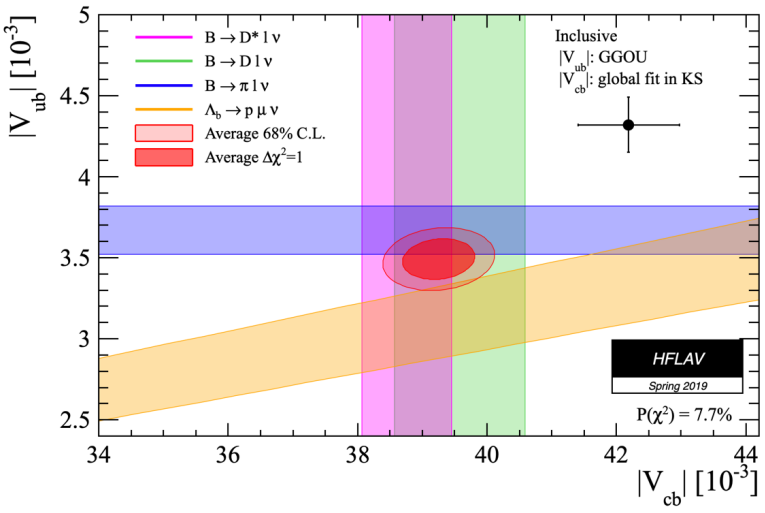


Figure 1.1: A two-dimensional plot of $|V_{ub}|$ and $|V_{cb}|$ measurements. The pink and green bands show exclusive $|V_{cb}|$ measurements from $B \to D^* \ell \nu$ and $B \to D \ell \nu$, respectively. The blue band gives an exclusive $|V_{ub}|$ measurement from $B \to \pi \ell \nu$ and the orange band represents an exclusive $|V_{ub}|/|V_{cb}|$ measurement from $\Lambda_b \to p \mu \nu$. A combined two-dimensional fit to these results has been performed and the 68% confidence level contour, as well as the $\Delta\chi^2 = 1$ contour, are shown in red. The χ^2 probability of the fit is 7.7%. An inclusive measurement of $|V_{ub}|$ performed with the GGOU shape function and an inclusive measurement of $|V_{cb}|$ from a global fit in the kinematic scheme (KS) are displayed. The tension between this point and the 2D fit result is visible [5].

the Standard Model. Especially interesting for the measurement of $|V_{ub}|$ is the charmless semileptonic decay $B \to \pi \ell \nu$. It is the most accessible decay, due to a clear experimental signature and relatively small theoretical uncertainties, and as shown in Figure 1.1, currently has the highest constraining power on $|V_{ub}|$.

The goal of this thesis is extracting the branching fraction of $B \to \pi \ell \nu$ (with $\ell = e, \mu$) using $5.6\,\text{fb}^{-1}$ of early Belle II data in order to prepare for $|V_{ub}|$ extraction. The discussed method is an untagged exclusive measurement. Since the theoretical models are parametrised by the momentum transfer of the decay, q^2, the branching fraction will be investigated in bins of q^2. It will be extracted from a simultaneous fit to the energy difference, ΔE, and the beam constrained mass, M_{bc}. These are variables that have high discriminating power between signal and background events due to the specific signal kinematics.

An introduction to relevant background information is given in Chapter 2 and the SuperKEKB accelerator and Belle II detector are discussed in Chapter 3. Chapter 4 describes the reconstruction of events. Signal selection and methods of background suppression are explained in Chapters 5 and 6, respectively. Next, the Data-Monte Carlo agreement is described in Chapter 7, before signal extraction and results are discussed in Chapters 8 and 9, respectively.

CHAPTER 2

Background Information

This chapter will give an introduction to theoretical models and experimental concepts that are relevant for this analysis. It will briefly explain the Standard Model (SM) of particle physics and then focus on weak decays as described by the SM. Special attention is given to semileptonic decays of B mesons. Lastly, the principles behind an experimental classifying method called multivariate analysis are outlined.

2.1 The Standard Model of Particle Physics

The Standard Model of particle physics is a quantum field theory that describes the elementary particles shown in Figure 2.1 and the fundamental forces between them, i.e. the strong, the electromagnetic and the weak force.

There are two main types of particles, called bosons and fermions, which are integer and half-integer spin particles, respectively, where spin is a particle's intrinsic angular momentum. The vector (spin-1) bosons are called gauge bosons and they are the carriers of the fundamental forces. The elementary fermions can be further categorised into leptons and quarks. Six flavours exist of both of these. There are three up-type quarks that have a charge of +2/3 and three down-type quarks of charge -1/3. Similarly, there exist three charged leptons of charge -1 and three chargeless neutrinos. The particles within these families have different masses but they share all other properties. The particles can therefore be categorised into three generations of increasing mass, where higher-generation particles decay into lower-generation particles. Each of the 12 fermions has an antiparticle with the same properties as the particle, except for having the inverse charge. Antiquarks also have the inverse colour charge, which is a particle property resulting from the strong interaction. There are six such colour charges (red, green, blue and their anticolours) and all particles that bear colour charge interact strongly.

The strong force is described by Quantum Chromodynamics (QCD) and the SU(3) symmetry. It is mediated by massless gluons that carry colour charge. Next to interacting with quarks, they also interact with each other (gluon self-interaction). At very high energy scales, when the quarks are bound inside a nucleon, the strength of the strong interaction decreases and the quarks possess asymptotic freedom [7]. On the other hand, when the quarks are separated, the strong interaction becomes strong enough to produce another quark pair, so that single quarks cannot be observed. This property is called colour confinement [8].

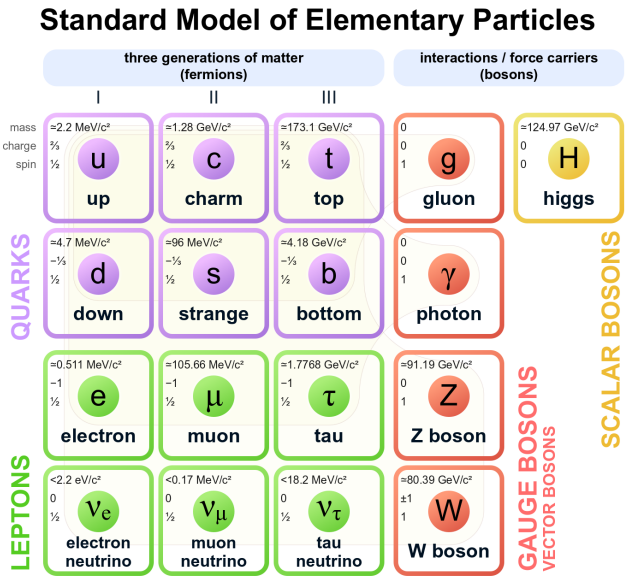


Figure 2.1: A graphical description of the Standard Model of particle physics. The quarks are coloured purple, the leptons green, the vector bosons red, and the scalar Higgs boson is coloured yellow. The particle masses, charges and spins are shown [6].

The electromagnetic force occurs between particles that carry electric charge and is described by Quantum Electrodynamics (QED) with a local U(1) symmetry. It is mediated by mass- and chargeless photons. Quarks and charged leptons interact electromagnetically. The weak force is weaker than the strong and the electromganetic forces. It is described by a local SU(2) symmetry and mediated by three massive bosons ($W^{\pm}$, Z^0). At high energies these two forces merge into the electroweak interaction described by the electroweak theory with an U(1) × SU(2) symmetry. The $W^{\pm}$, Z^0 and photon result from spontaneous electroweak symmetry breaking, involving the Higgs mechanism, giving the three weak gauge bosons their mass [9]. The scalar (spin-0) Higgs boson is also a product of the Higgs mechanism.

2.2 Weak Decays

The weak interaction is special in that is the only interaction that allows quarks to change between flavours. This is possible because the weak quark eigenstates are not the same as the mass quark eigenstates, but they are rotated with respect to the mass eigenstates. This results in mixing between the down-type eigenstates, as described by the 3 × 3 Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$\underbrace{\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix}}_{\text{Weak Eigenstates}} = \underbrace{\begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix}}_{\text{CKM Matrix}} \underbrace{\begin{bmatrix} d \\ s \\ b \end{bmatrix}}_{\text{Mass Eigenstates}}. \tag{2.1}$$

The CKM matrix can be parametrised by three angles and one complex phase. The probability of

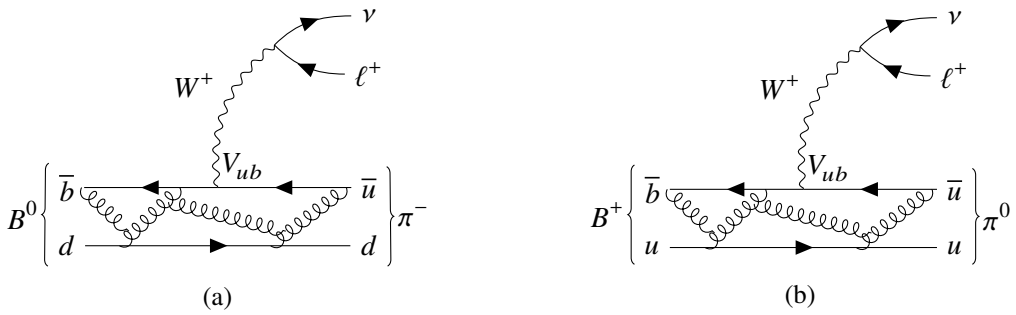


Figure 2.2: Feynman diagrams for the flavour-changing semileptonic weak decays (a) $B^0 \to \pi^- \ell \nu$ and (b) $B^+ \to \pi^0 \ell \nu$. The decay rate is proportional to $|V_{ub}|^2$. Examples of quark-gluon and gluon self-interactions are shown.

a quark of flavour i changing into a quark of flavour j is proportional to the CKM matrix element squared $|V_{ij}|^2$. The elements of the CKM matrix are hierarchical in magnitude. The diagonal elements are of order $\simeq 1$, $|V_{us}|$ and $|V_{cd}| \simeq 1/4$, $|V_{ts}|$ and $|V_{cb}| \simeq 1/20$, and $|V_{ub}|$ and $|V_{td}| \simeq 1/250$ [10]. Therefore, the decays described by the diagonal of the matrix occur most often. The CKM matrix is required to conserve probability upon changing flavours, and should only rotate the weak eigenstates, and due to this it is unitary.

2.3 Semileptonic B Decays

One process that contains a weak flavour-changing interaction is a semileptonic B decay, where a B meson decays into a hadron, lepton and a neutrino. One example is the decay $B \to \pi \ell \nu$. Feynman diagrams corresponding to B^0 and B^+ decays are shown in Figure 2.2. In detail, for the B^0 decay shown in (a), the $\overline{b}$ quark of the B meson changes into an $\overline{u}$ quark through emitting a W boson. The W boson decays into a lepton-neutrino pair and the spectator quark d combines with the $\overline{u}$ quark to form a pion. As explained in Section 2.2, the probability for this decay is proportional to $|V_{ub}|^2$. The decay kinematics can be described by the momentum transfer, q^2, which is given by

$$q^2 = (p_B - p_\pi)^2 = (p_\ell + p_\nu)^2, \tag{2.2}$$

where p_B, p_π, p_ℓ and p_ν are the four-momenta of the B meson, pion, lepton and neutrino, respectively.

Next to a weak component, the decay kinematics also contain a strong component due to quark-gluon and gluon self-interactions. Examples of these are shown in the Feynman diagrams in Figure 2.2. Essentially, each of the possible strong interactions results in an additional Feynman diagram. The QCD effects of the decay are to be calculated from these. However, this cannot be done analytically. The strong effects are therefore described by the form factor (FF), which is usually parametrised by q^2. The dependence of the differential decay rate on the weak and the strong components can be summarised by

$$\frac{d\Gamma(B \to \pi \ell \nu)}{dq^2} \propto |V_{ub}|^2 \times |\mathrm{FF}(q^2)|^2. \tag{2.3}$$

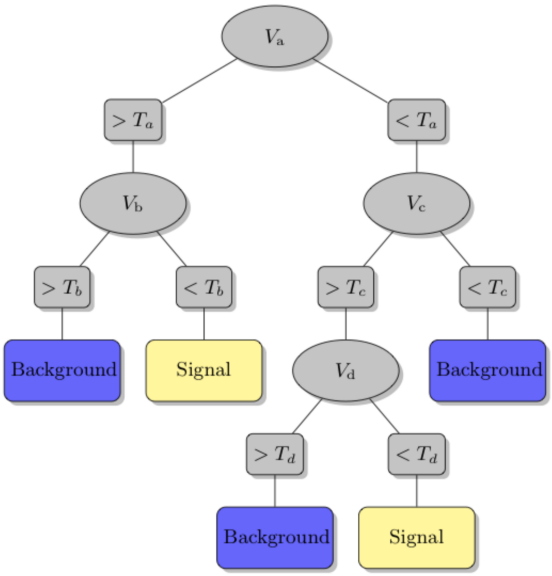


Figure 2.3: A single decision tree, starting with an unknown sample, V_a. At each node a cut, T_i, divides the sample into two subsamples, which at the last stage are classified as signal (yellow) or background (blue) [13].

The constant of proportionality is

$$k = \frac{G_F^2}{32} \int_{-1}^{1} p_\pi^3 \sin^2 \theta_{W\ell} \, d\cos\theta_{W\ell}, \tag{2.4}$$

where G_F is the Fermi constant, p_π is the pion momentum and $\theta_{W\ell}$ is the angle of the lepton with respect to the virtual W boson in the W boson rest frame [11]. If the theoretical predictions of the form factor are known and the differential decay rate of $B \to \pi\ell\nu$ is measured, $|V_{ub}|$ can be extracted.

2.4 Multivariate Analysis (MVA)

Multivariate analysis (MVA) is a set of techniques to investigate data dependent on multiple variables [12]. One application is distinguishing predefined classes of data, such as signal and background. In particle physics MVAs are often used in order to maximise the signal to background ratio. Different types of MVA algorithms exist to achieve this. While some are simply based on linear algebra, others employ machine learning (ML). The ML algorithms are able to automatically understand the underlying statistical distribution of a given dataset and exploit correlations between variables. Such ML-based MVAs can then be used to obtain a classifier between signal and background.

One type of ML-based MVA is a boosted decision tree (BDT). A decision tree is employed to classify an unknown sample, V_a, into signal and background by performing consecutive cuts, T_i, as shown in Figure 2.3. Each cut occurs at a node. The decision tree is trained on a known sample with background and signal labelled as such. The decision tree is able to pick up on the distinguishing features between these categories. The cut-flow with the highest signal probability (number of signal events in the bin/total number of events) is combined to form a one dimensional decision tree classifier.

This classifier is the BDT output and it embodies the discriminating power between signal and background. The BDT performance is evaluated on a testing sample.

If a single tree is trained on a complex model it is prone to incorporate specific statistical features of the training sample into the classifier. This is called overtraining. One method to check for overtraining relies on a hypothesis test, where the null hypothesis is that the testing sample is the same as the training sample within statistical variations. A p-value can then be constructed from the difference between the classifier distributions of the training and the testing sample. The p-value gives the probability that the testing sample is in agreement with the output classifier of the BDT. In physics it is conventional to reject null hypotheses due to overtraining, if they yield a p-value smaller than 0.05 [14]. On the other hand, if a very simple model is chosen, the output classifier will not encompass the features of the training sample. The testing data will then not be described by the null hypothesis. This is called undertraining and the p-value for such an undertrained BDT lies above the threshold of 0.95.

To reduce the probability of overtraining occuring, multiple trees can be trained using a boosting algorithm. This means that events are assigned weights depending on how well they were classified by a single tree. The weight is then taken into account as the next tree is trained. One boosting algorithm is a gradient boosted decision tree (GBDT) that builds the final classifier not on the signal probability, but on the combined weights. A stochastic gradient boosted decision tree (SGBDT), in addition, trains on subsets of the the initial sample, which reduces the probability of overtraining. For Belle II FastBDT is the standard boosting algorithm, which is a type of SGBDT [15].

A number of parameters (hyperparameters) can be set that affect the performance of the BDT and need to be optimised. One of these hyperparamters is the *number of trees*, which determines how many separate decision trees are combined into the BDT. This is inversely proportional to the rate of combining the trees, the hyperparameter *shrinkage*. The *depth of the trees* is another hyperparameter and determines the number of cut levels in each separate tree, while the maximum *number of cuts* in a tree is another hyperparameter. The size of the subset that each SGBDT is trained on is given by the hyperparameter called *sampling rate*.

The Receiver Operating Characteristic (ROC) curve is a standard method used for evaluating MVAs. It gives the ratio of background rejection to signal efficiency as a function of the BDT output classifier. The area under the ROC curve (AUC) is a measure of the BDT performance. In the ideal case, where the BDT can 100% separate signal from background, the AUC is 1. An example ROC curve for the case of an almost ideal BDT is shown in Figure 2.4. It also shows a ROC curve for the case of random guessing, where the AUC is 0.5, and a ROC curve for a well performing BDT.

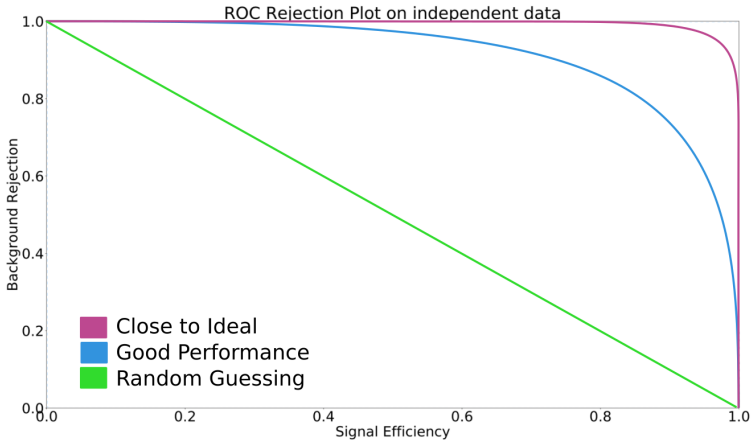


Figure 2.4: Example ROC curves giving the ratio between background rejection and signal efficiency for different classifier outputs for three different BDTs: Random guessing (green), good performance (blue) and almost ideal performance (pink).

CHAPTER 3

SuperKEKB and Belle II

This chapter will give an overview of the experimental environment that the analysis is based in. Firstly, the SuperKEKB accelerator and the resulting electron-positron collisions will be discussed. Next, the types of background occuring at SuperKEKB will be explained. Lastly, the Belle II detector, its subcomponents and the trigger will be described. Since this thesis discusses early Phase 3 data, a special focus will be laid on the proccesses dominant during this measurement period.

3.1 SuperKEKB

SuperKEKB is a two-ring asymmetric-energy electron-positron collider based at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. It is the successor of KEKB, which was in operation from 1999 to 2010. It was commisioned for Phase 1 in early 2016, during which single beams were studied. Phase 2 of operation started in early 2018, during which most of the detector had been assembled and the first electron-positron collisions occured. Phase 3 has the full detector installed and began in March 2019. The measurement period from that point until the summer shut-down in July 2019 is referred to as early Phase 3.

As its predecesssor, SuperKEKB runs at a centre of mass energy corresponding to the mass of the $\Upsilon(4S)$ resonance of $\sqrt{s}$ =10.58 GeV. Electrons are generated in bunches using an electron gun and a secondary tungsten target is used to generate positrons [16]. The electrons and positrons are first accelerated to 7.0 GeV and 4.0 GeV, respectively, using a linear accelerator. Positrons are guided through a damping ring to reduce their emittance (spatial and energy spread) and then both electrons and positrons are fed into the main accelerator ring. The positrons are stored in the low energy ring (LER) and the electrons in the high energy ring (HER). The Belle II detector is positioned at the collision point and will be explained further in Section 3.4. The accelerator structure at KEK is shown in Figure 3.1.

SuperKEKB aims to collect an integrated luminosity of 50 ab^{-1} by 2025 [16]. The instantaneous design luminosity is $\mathcal{L} = 8.0 \times 10^{35}\,cm^{-2}s^{-1}$, which is a factor of 40 greater than the peak luminosity of KEKB. This is planned to be achieved through (1) squeezing the beam, causing a factor 20 increase and (2) high beam currents, leading to a factor 2 increase in luminosity. (1) is achieved by strong focussing of the beams close to the collision point. For this, large beam crossing angles of 83 mrad need to be employed, leading to horizontal and vertical beam sizes of ~µm and ~nm, respectively.

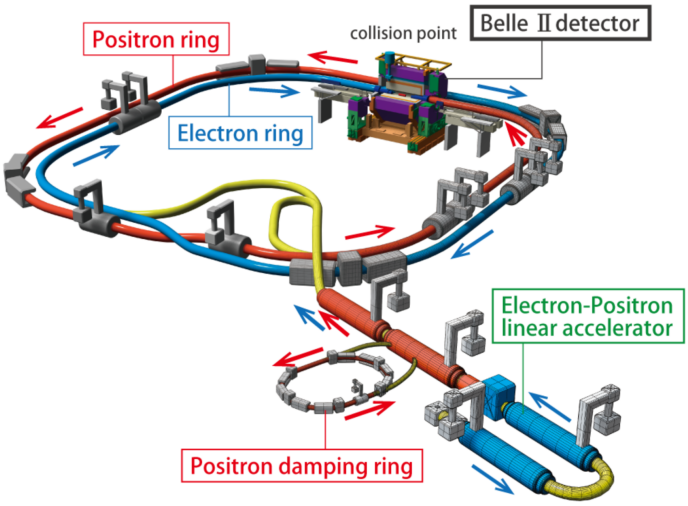


Figure 3.1: Layout of the SuperKEKB accelerator at KEK in Japan (not to scale). The linear accelerator, as well as the electron and positron rings, and the position of the Belle II detector at the collision point are shown. The positron damping ring reduces the emittance of the positrons [17].

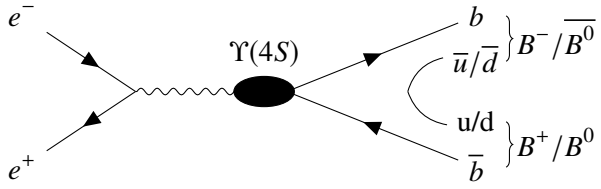


Figure 3.2: Diagram for resonant Υ(4S) production from an electron-positron intersction and subsequent decay to a $B\overline{B}$ pair.

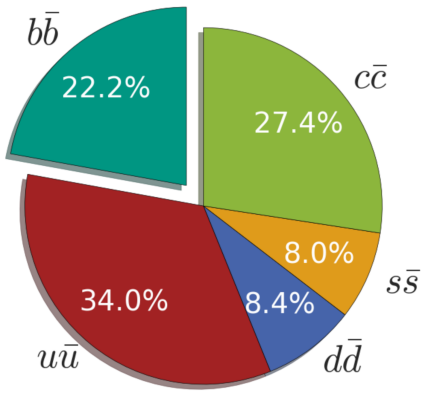
3.2 Particle Collisions

The centre of mass energy of 10.58 GeV, at which the electron and positron collide, corresponds to the mass of the Υ(4S) resonance. The cross section for Υ(4S) production at this energy is 1.11 nb [18]. Since the Υ(4S) energy lies just above the threshold for $B\overline{B}$ production ($E_{\Upsilon(4S)} > 2m_B$), the Υ(4S) decays to a $B\overline{B}$ pair in more than 96% of the cases. This production mechanism is shown in Figure 3.2.

The $B\overline{B}$ pairs are either $B^0\overline{B^0}$ or B^+B^- pairs, as not enough energy remains to strongly produce anything other than an $u\overline{u}$ or $d\overline{d}$ quark pair. The ratio of $B^0\overline{B^0}$ pairs to B^+B^- pairs is 48.6 : 51.4 [10]. The mass of the B mesons is 5.28 GeV and they have a mean lifetime of ~1.6 ps. The aymmetry in the beam energies of the HER and LER causes a boost of the B meson flight direction along the electron beam direction. The relatively long lifetime of the B meson then leads to displaced vertices of ~100 µm.

Process	Cross Section [nb]
$\Upsilon(4S)$	1.11
$u\bar{u}$	1.61
$d\bar{d}$	0.40
$s\bar{s}$	0.38
$c\bar{c}$	1.30
$e^+e^-(\gamma)$	300
$\mu^+\mu^-(\gamma)$	1.15
$\gamma\gamma(\gamma)$	4.99
$e^+e^-e^+e^-$	39.7
$e^+e^-\mu^+\mu^-$	18.9
$\tau^+\tau^-(\gamma)$	0.92

(a)



(b)

Figure 3.3: (a) Cross sections at the SuperKEKB beam energies as given by [18] and (b) the relative fractions of quark pair continuum production. $b\bar{b}$ production accounts for 22.2% [19].

3.3 Backgrounds at Belle II

Electron-positron collisions do not only produce $\Upsilon(4S)$. Other physics processes are considered backgrounds at Belle II. They are luminosity dependent and increase in importance at the increased luminosity of SuperKEKB. These backgrounds, along with beam backgrounds, will be discussed in the following sections. The cross sections of the discussed processes at the SuperKEKB beam energies are given in (a) of Figure 3.3.

3.3.1 Continuum Background

Instead of producing an $\Upsilon(4S)$, the electron-positron collision can produce non-resonant quark pairs. Here we will refer to these backgrounds as continuum. The cross sections are shown in (a) and the relative fractions are shown in (b) of Figure 3.3, respectively. They are called continuum background because quark pair production occurs at any energy, given that it is above the threshold of two times the respective quark mass. This also means that there is no resonance peak for these processes when scanning over the energy. At the Belle II centre of mass energy the possible quark pairs are $u\bar{u}$, $d\bar{d}$, $c\bar{c}$ and $s\bar{s}$, with $u\bar{u}$ and $c\bar{c}$ pair production the most dominant.

3.3.2 Low-Multiplicity Background

Low-multiplicity backgrounds are those that occur from QED electron-positron interactions and result in a low number of final-state particles (four or less). Feynman diagrams for four types of low-multiplicity processes are shown in Figure 3.4.

Electron (Bhabha) scattering is one type of low-multiplicity background. At the SuperKEKB centre of mass energy it has a cross section of 300 nb (~300 times greater than the $\Upsilon(4S)$ production cross section) and it is therefore a dominant process [18]. It is especially problematic when a photon is produced during the scattering process (radiative Bhabha scattering): $e^+e^- \to e^+e^-\gamma$, as shown in the Feynman diagram in (a) of Figure 3.4. The photon can interact with the detector material and

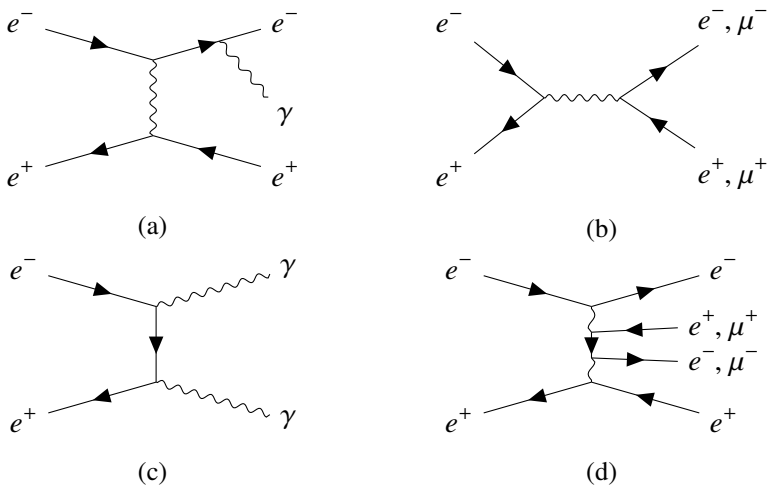


Figure 3.4: Feynman diagrams for low-multiplicity processes: (a) radiative Bhabha scattering, (b) electron/muon pair production, (c) annihilation into two photons and (d) two-photon process.

pair produce, resulting in false hits. The electron and positron can also pair produce two electrons ($e^+e^- \to e^+e^-$) or two muons ($e^+e^- \to \mu^+\mu^-$) or annihilate into two photons ($e^+e^- \to \gamma\gamma$). Feynman diagrams for these processes are given in (b) and (c) of Figure 3.4, respectively.

Another source of low-multiplicity background is the two-photon process:

$$\begin{aligned} e^+e^- \to e^+e^-\gamma^*\gamma^* &\to e^+e^-e^+e^- \\ &\to e^+e^-\mu^+\mu^-. \end{aligned} \tag{3.1}$$

A Feynman diagram for the dominant process is shown in (d) of Figure 3.4. The electron-positron collision produces two virtual photons which then pair produce low-momentum electrons, positrons and muons [20].

3.3.3 τ Pair Production

Tau pair production is also a continuum QED process that occurs at Belle II and its Feynman diagram is analogous to the one of electron/muon pair production shown in (b) of Figure 3.4. The τ leptons can decay in various ways which means they behave like low-multiplicity background in some cases and like continuum in others. Hadronic tau lepton decays often mimic quark pair continuum background. They can, however, also behave like low-multiplicity background, as is the case for leptonic tau decays.

3.3.4 Beam Background

Beam (or beam-induced) backgrounds do not originate from an electron-positron collision, but instead result, for example, from beam particles interacting with each other or with residual gas particles left in the beam pipe. They mimic the behaviour of signal hits and can therefore cause serious problems during event reconstruction. The expected dominant source of beam background in Phase 3 is Touschek scattering. This results from intra-bunch collisions between beam particles, causing particles to leave the bunch [21]. Since the rate of this process in inversely proportional to the beam

size, which is small for SuperKEKB, Touschek scattering is expected to increase in size compared to KEKB. However, it is also proportional to the beam energy, E^{-3}, and therefore the increase in the energy of the LER as compared to KEKB, reduces this effect.

The source of beam background that is dominant in early Phase 3 is beam-gas scattering. This results from beam particles interacting with residual gas particles, such as H_2 and CO, in the beam pipe. The interaction has components of both the Coulomb interaction and Bremsstrahlung. Similarly, these processes can result in a beam particle leaving the bunch and showering in the detector.

A third type of beam background comes from synchrotron radiation. This is radiation emitted radially outward as a particle is accelerated. This is especially problematic inside the detector where strong magnetic focusing enhances this process. The radiated low energy photons can interact with the detector and produce false hits. While beam background processes may occur anywhere along the beam pipe, they are most problematic when the beam loss occurs in the interaction region.

3.4 Belle II Detector

At the increased luminosity of SuperKEKB the number of both signal and background events increases and the detector needs to cope with these higher levels. For this reason the Belle II detector replaced the Belle detector and it was commisioned in March 2019. It is composed of a number of subdetectors that are placed around the collision point in the beam pipe, which has an inner radius of 10 mm [16]. The detector fully encloses the interaction region (IR). This is a region of approximately 4 m around the interaction point (IP).

The coordinate system of the detector is right-handed Cartesian with the origin at the IP by convention. The z-axis points along the beam pipe in the forward direction (HER), the y-axis points to the top of the detector and the x-axis points radially out of the detector ring. The detector can also be described in polar coordinates: r, θ and ϕ, where r is the radial distance from the IP, θ is the polar angle with respect to z and ϕ is the azimuthal angle in the x-y plane. The angular acceptance of the detector, i.e. the solid angle in which particles are detected, is $17° < \theta < 150°$ and $\phi \leq 2\pi$. The detector is divided into three regions of θ which are:

- forward region: $17° < \theta \leq 30°$
- barrel region: $30° < \theta \leq 125°$
- backward region: $125° < \theta \leq 150°$

The Belle II detector and the coordinate system are shown in Figure 3.5. The asymmetry in the beam energies is reflected in an asymmetry in the detector. The subdetectors have different purposes such as vertexing, tracking, particle identification, or charge, energy or momentum measurements, and will be explained in the following sections. A superconducting magnet producing a solenoid field of 1.5 T along the beam direction is installed outside all but the last detector component (KLM), allowing momentum and charge measurements to be performed [16].

3.4.1 Pixel Detector (PXD)

The two innermost detector components make up the Vertex Detector (VXD). The one closest to the beam pipe is the Pixel Detector (PXD). Its main goal is the vertexing of short lived particles that decay within the IR by allowing the precise extrapolation of their long lived decay particles' tracks into the

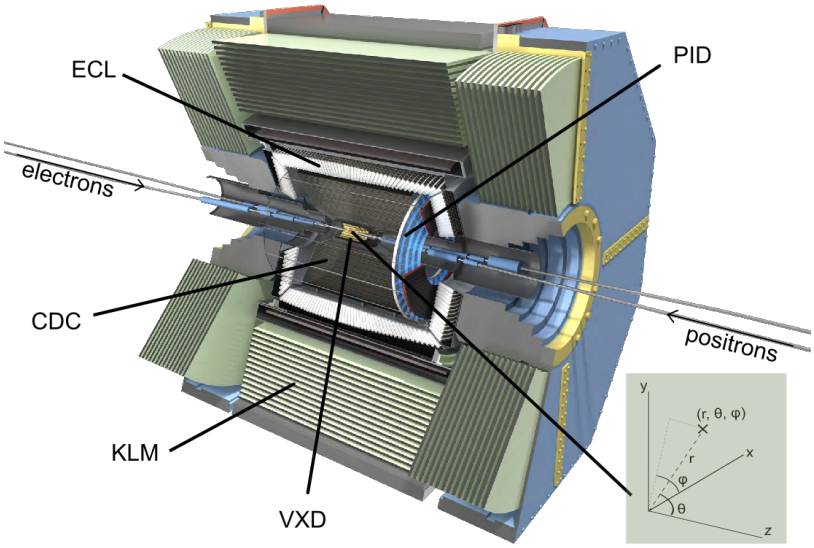


Figure 3.5: The Belle II detector and its subdetectors: the Vertex Detector (VXD), the Central Drift Chamber (CDC), the Electromagetic Calorimeter (ECL), the Particle Identification System (PID) and the K_L^0 and Muon Detector (KLM). The coordinate system at the interacton point is shown. Adapted from [22].

IR. In addition, the track momentum can be determined from the track curvature due to the magnet. The PXD consists of two layers depleted field effect transistors (DEPFET) silicon detectors placed at radii of 14 and 22 mm from the IP. There are 20 modules (16 in the first and 4 in the second layer), each composed of a matrix of 768 × 250 DEPFET sensors [16].

3.4.2 Silicon Vertex Detector (SVD)

The other VXD subdetector is the Silicon Vertex Detector (SVD). Similar to the PXD, the SVD allows vertexing of short lived particles and momentum measurements of their decay products. It is composed of four layers of double-sided silicon strip detectors positioned at radii between 39 and 135 mm from the IP [16]. These detectors are made of silicon bulk surrounded by layers of n-doped and p-doped material. Strips are chosen here instead of pixels since this reduces the cost and the number of readout channels.

3.4.3 Central Drift Chamber (CDC)

The central drift chamber (CDC) allows 3D tracking, momentum measurements from the track curvature in the magnetic field, and a determination of the energy loss. The CDC output is vital for the trigger system, which will be explained in Section 3.5, and energy loss measurements are important for particle identification.

A drift chamber is based on the principle that a charged particle traversing a gas mixture will ionise the gas. If an electric field is applied, the electrons will drift along the electric field. The strong electric field close to the wire causes secondary ionizations, which leads to avalanche multiplication, and a charge is induced in the drift chamber wire. If multiple such wires are placed parallel to the beam axis

(axial layers) the position of the particle in the x-y plane can be determined from the diffference in timing between the wires. By skewing some of the wires (stereo layers) a measurement of the position in the z plane becomes possible. The currents measured by the drift chamber are also proportional to the energy of the particle, and so the energy loss $\frac{dE}{dx}$ can be found.

The CDC has 14336 such sense wires made of tungsten [16]. Each of the sense wires is surrounded by 8 aluminium field wires. The gas mixture is made of 50% Helium and 50% Ethane, C_2H_6. The CDC extends from radii of 160 to 1 130 mm from the IP. The sense wires are aranged in alternating superlayers of axial and stereo layers. Except for the first superlayer, which is made of 8 layers, the superlayers are made of 6 layers. The sense wires in the stereo layers are skewed by −74 mrad to 70 mrad.

3.4.4 Time of Propagation Counter (TOP)

The next layer of subdetectors is called the Particle Identification System (PID). This is made of two subdetectors. The Time of Propagation Counter (TOP) is the part of the PID system that is installed around the barrel region of the IR and uses the Cherenkov principle to allow separation between $K^{\pm}$ and $\pi^{\pm}$, as well as protons and deuterons. Muons also leave a signal in the TOP. The Cherenkov principle states that particles moving through a material at a speed greater than the speed of light in the material will emit photons. The photons are emitted at the Cherenkov angle (θ_c) characteristic of the particle given by

$$\cos\theta_c = \frac{1}{\beta n(\lambda)}, \tag{3.2}$$

where β is the ratio of the velocity of the particle to the speed of light in vaccum. This is equivalent to the ratio of the momentum, p, to the energy, E, of the particle $\beta = \frac{p}{E}$. $n(\lambda)$ is the refractive index of the material at a given wavelength λ.

From measurements of the Cherenkov angle of a particle, the β of the particle can be found. The β for particles of different masses are different at the same momentum and therefore the TOP is able to resolve mass differences. The TOP is made of quartz crystals, that totally internally reflect Cherenkov photons upon their generation. They are guided until they reach photomultipliers at the end of the crystal [16]. $K^{\pm}$ are heavier than $\pi^{\pm}$, have a smaller β and from Equation 3.2 have a wider Cherenkov angle. Due to the wider angle, Cherenkov photons from $K^{\pm}$ are reflected fewer times inside the crystals than Cherenkov photons from $\pi^{\pm}$. They therefore have a smaller time of propagation between their creation and their arrival at the photomultiplier than $\pi^{\pm}$s.

3.4.5 Aerogel RICH Detector (ARICH)

The second component of the PID system is the Aerogel Ring-Imaging Cherenkov Detector (ARICH) and it is positioned at the forward endcap, with an acceptance of $14.78° < \theta < 33.71°$. In addition to being able to separate $K^{\pm}$ and $\pi^{\pm}$, it to some extent can also discriminate between muons, electrons and pions, if they have a momentum below 1 GeV [16]. In the ARICH the Cherenkov radiator is aerogel and the cone formed by the Cherenkov angle given in Equation 3.2 is projected onto a photon detector. The radius of the projected circle is related to the Cherenkov angle and so the β of the particle and subsequently the mass can be determined, if the momentum is known.

3.4.6 Electromagnetic Calorimeter (ECL)

The Electromagentic Calorimeter (ECL) gives measurements of the energy and angular coordinates of photons and allows the identification of electrons. It is especially important in the reconstruction of π^0s, which quickly decay to two photons. When a particle passes through the ECL, electromagnetic or hadronic showers, or ionisation and excitation processes, produce photons, which cause more showers. Almost the entire shower energy is recovered and from this, the energy of the particle can be inferred. This is not the case for mimimum ionizing particles, such as muons, that do not deposit their entire energy in the ECL.

The ECL is made of 8736 thalium doped scintillating Caesium Iodine CsI(Tl) crystals in the barrel and pure CsI in the endcaps, with a pair of photodiodes attached to the end of each crystal [16]. It covers a polar angle of $12.4° < \theta < 155.1°$.

3.4.7 $\mathbf{K_L^0}$ and Muon Detector (KLM)

The K_L and Muon Detector (KLM) is designed to identify K_L^0 and $\mu^\pm$ with a momentum greater than 0.6 GeV. It is placed just outside the superconducting magnet and covers a polar angle of $20° < \theta < 155°$. It consists of alternating layers of 4.6 cm-thick iron plates and active detector material. The iron also serves as a return yoke for the superconducting magnet. In the barrel region Resisitive Plate Chambers detect charged particles, while in the endcap regions scintillator strips are used [16]. Both muons and K_L leave hits in the active detector material of the KLM. They can be distinguished by matching the hits to tracks in the CDC for muons, and to clusters in the ECL for K_L.

3.5 Trigger

The bunches cross every 8 ns and so huge amounts of data are collected by the detector and would need to be recorded. However, not every bunch crossing produces interesting interactions and the trigger system is responsible for making the decision whether or not a certain collision is read out. The trigger system uses information from the subdetectors to perform track and cluster reconstruction. It aims to identify $\Upsilon(4S)$ events and to reject backgrounds, such as Bhabha scattering, which it is able to ignore mostly.

There are multiple hierarchical trigger levels and the first one is based on hardware. This is called the Level 1 (L1) trigger and it takes the information from a subset of the detector components (CDC and ECL) into account. If the L1 trigger decision is positive, the information from all subdetectors, except for the PXD, is read out and combined to build an event. During this stage the software High Level Trigger (HLT) decides whether to keep or discard the event [23]. The decision is based on the calculation of some event variables. If the decision is to keep the event, the PXD information is added. The event is then stored to disk, where it can be further processed using the Belle II Analysis software framework ($basf2$).

CHAPTER 4

Reconstruction

After an event has passed the trigger, offline event reconstruction is performed by combining subdetector information. Particles are reconstructed from tracks and clusters. The tracks are formed from hits in the CDC along with information from the VXD, while the clusters are combined energy deposits in the ECL or KLM. Similarly, Monte Carlo (MC) generated particles are reconstructed from simulated clusters and tracks. Here the MC simulates electron-positron collisions at the $\Upsilon(4S)$ energy.

In both cases, events are reconstructed by combining reconstructed particles. Selections can be performed on the particles, as well as on the event, during reconstruction . The loss of events due to these selections is quantified by the reconstruction efficiency. Since this analysis is performed in bins of reconstructed q^2, the reconstruction efficiency is found for each q^2 bin. This is the number of events reconstructed in a q^2 bin out of the number generated in that bin. However, since the form factor is parametrised by the true q^2, the ratio of the number of events in a certain reconstructed q^2 bin to the number of events in that true q^2 bin needs to be determined. This is called the bin migration factor and results from the difference between the reconstructed and true q^2, or the resolution in q^2.

In the following, first, the MC samples used in this analysis will be explained. After this, the event reconstruction, starting from the reconstruction of tracks and clusters and moving on to the reconstruction of the signal side and rest of event, will be discussed. Next, the selections applied during the reconstrution of events in this analysis will be described and finally, the reconstruction efficiency and bin migration factor and their origins will be explained.

4.1 Monte Carlo Simulation

Monte Carlo (MC) samples in particle physics are simulations of physics processes. For this analysis, MC samples containing the simulation of continuum, tau pair and $\Upsilon(4S) \rightarrow B\overline{B}$ events at the SuperKEKB beam energies are used. In the simulation of $B\overline{B}$ events, the B mesons can either decay generically or in a specific mode. Here we use 400 fb^{-1} of generic Monte Carlo, where both B mesons decay generically. This corresponds to $\sim$ 400 million $B\overline{B}$ events and $\sim$ 2000 million continuum events.

For the signal selection and the BDT training, described in Chapter 6, we use signal MC. In signal MC one of the B mesons decays in a specific decay mode, while the other B meson decays generically. In our signal MC one of the B mesons always decays into $X_u \ell \nu$. We used samples of 80 million $X_u \ell \nu$ events, for each B^0 and $B^{\pm}$, and extracted $B \rightarrow \pi \ell \nu$ events from these.

The $B\overline{B}$ events were generated using the event generator EvtGen [24]. The decays and their relative

frequencies are given to the program in the form of a decay file. The decay file also contains the theoretical predictions of the form factor (FF), given by the FF model. The FF model used in the generation of $B \to \pi \ell \nu$ events in the Belle II MC is the ISGW2 model. The FF model defines the shape of the q^2 spectrum in the MC and it is instrumental in the extraction of $|V_{ub}|$, as explained in Section 2.3.

The continuum and tau pair events were generated using PYTHIA and KKMC [25] [26]. The beam backgrounds are not generated, but are mixed into the Monte Carlo at a later stage, using background overlays extracted from measured data [18]. The propagation of the particles through the detector and the resulting interactions are simulated using Geant4 [27]. This step allows a comparison between MC generated variables and their reconstructed counterparts. Information taken from generator level will be referred to as "true" from this point forth.

4.2 Track Reconstruction

Track reconstruction can be divided into two steps, which are (1) building tracks from hits (track finding), and (2) extraction of track variables (track fitting). To perform track reconstruction, tracking algorithms are employed. These differ for the various tracking subdetectors due to different levels of background and spatial resolution. The VXD has good spatial resolution but suffers from high levels of beam background, while the CDC has worse spatial resolution but is affected less by beam backgrounds. The algorithms are based on fitting helix tracks and connecting neighbouring hits through pattern recognition.

During track fitting, the momentum, charge and other variables can be extracted from helix fits to the tracks. Two important variables are the track impact parameters, dr and dz. When the track is extrapolated towards the IP, dr and dz are the closest distances to the IP in the radial coordinate, r, and in the z coordinate, respectively.

Some problems encountered in tracking are caused by duplicate tracks, called clones. A clone could result from a curler, which is a very low momentum particle that is deflected strongly by the magnetic field. Parts belonging to the same spiral track might be reconstructed as two separate tracks. Ghosts are also problematic in early Phase 3. They are the results of combining hits into tracks, that have been assigned incorrect position information. Some tracks will also not be identified, even though they are in the acceptance region, due to faults in the algorithm or inefficiencies in the detector. This can be quantified by the tracking efficiency, which is the ratio of reconstructed tracks to the actual number of charged particles.

4.3 Cluster Reconstruction

Clusters are reconstructed from energy deposits in the ECL using a cluster algorithm. Crystals are picked as a seed for a connected region (CR) if they are local energy maxima with an energy greater than 10 MeV [18]. Neighbouring crystals are then added to the CR if they have an energy greater than 0.5 MeV. If two CRs overlap they are merged and, after weights are asigned to the overlapping crystals, divided into clusters.

4.4 Particle Identification

Next, track-cluster matching occurs and the results are combined with information from various subdetectors to test for consistency with different particle hypotheses. Generally, charged particles are identified by having tracks and corresponding clusters while neutral particles will not have a track, but a cluster.

Once a track has been matched to a cluster, a charged particle hypothesis is applied. This means the energy-momentum relationship of one of the charged long-lived particles ($e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, proton or deuteron) is assumed. For each particle hypothesis, α, and each PID-relevant detector, *det*, a particle likelihood is constructed, $\mathcal{L}_{\alpha,\mathrm{det}}$. The global PID likelihood for a particle type is then given by

$$\mathrm{PID}_{\alpha} = \frac{\prod_{\mathrm{det}} \mathcal{L}_{\alpha,\mathrm{det}}}{\sum_{\alpha}(\prod_{\mathrm{det}} \mathcal{L}_{\alpha,\mathrm{det}})}. \tag{4.1}$$

For the charged hadron likelihoods ($\pi^{\pm}$, $K^{\pm}$, p, d) the relevant PID information comes from the PID system and the CDC. The information from these detectors is also important for the electron likelihood, but the ECL information is the primary input. The muon likelihood is built from KLM information.

4.5 Event Reconstruction

From tracks that are matched to clusters, and the particle identification variables, charged particles can be reconstructed. Clusters that cannot be matched to tracks are checked for the kinematic consistency with photons or K^0_{L} using the cluster shape. While K^0_{L} will have clusters in the KLM and possibly in the ECL, photons will only have clusters in the ECL. Two clusters in the same event that are identified as coming from photons are combined to reconstruct a π^0.

The reconstructed particles can then further be combined through track vertexing and invariant mass reconstruction to reconstruct the decay chain. In a tagged measurement all the particles in an event are reconstructed (both B meson decay chains), while in an untagged measurement only the signal B meson side is reconstructed. In an untagged measurement all unused tracks and clusters in that event are then labelled as rest of event (ROE). The untagged reconstruction of the signal side and the ROE, along with imposed selections, will be explained in more detail below. The neutrino can be reconstructed indirectly from the ROE and this will also be discussed.

4.5.1 Signal-Side Reconstruction

The signal-side $B \to \pi \ell \nu$ events can be reconstructed from the reconstructed particles. Selections on the individual particles, as well as on the whole event, can be performed. While we place no requirements on the lepton nor on the whole event at this point, the selections on the pion candidates are explained below.

The Belle II Analysis software framework (*basf2*) provides standard particle lists of a given efficiency, which are the results of selections that retain that certain fraction of true particles. Standard lists are used to select the pion candidates. We select the 40% efficiency list for the π^0, which contains a requirement on the π^0 mass. For a correctly reconstructed π^0 the invariant mass should lie in the π^0 mass resonance around 135 MeV [10]. To allow for misreconstruction of the photon energies, a broad window of $0.124 < m_{\pi} < 0.140\,\mathrm{GeV}$ is selected.

For the charged pion, we choose the 85% efficiency standard list. The standard list selects on two CDC variables. It requires the number of hits in the CDC that are associated with a track to be greater than 20 and that the track lies within the CDC acceptance. There is also a requirement on the impact parameters ($dr < 0.5\,\text{cm}$ and $|dz| < 2\,\text{cm}$) to reject pions not coming from the IP. In addition, the standard list requires the pion ID to be greater than 0.098.

4.5.2 Rest of Event (ROE) Reconstruction

In an untagged analysis all tracks and clusters in an event that are not part of the signal side are part of the ROE. Assuming the signal side is correctly reconstructed, the ROE should only contain the decay products of the second B meson. However, often beam background or clones present in the event will also be part of the ROE. Therefore, selections on the ROE are especially, but not only, targeted at rejecting beam background. Selections on the ROE are track, cluster, as well as event-level requirements. In $basf2$ selections on tracks and clusters in the ROE are performed using an ROE mask. These selections remove those tracks and clusters from the ROE that do not pass the selections. This step is called ROE clean-up. The introduction of the ROE mask also leads to the application of particle mass hypotheses to the ROE objects in order to calculate event-level variables. The choice of mass hypothesis is given by the expected probability of a particle occuring in the ROE.

One of the selections in the ROE mask is based on the kinematics of our signal mode. If the signal side is correctly reconstructed, there should be no particle in the ROE with a momentum greater than 3.2 GeV. We, therefore, require the momentum for both tracks and clusters in the ROE to be $p \leq 3.2\,\text{GeV}$. For clusters we also choose a lower momentum ($p \geq 0.05\,\text{GeV}$) and a lower energy limit ($E > 0.09\,\text{GeV}$), below which the detector performance degrades. This additionally rejects beam backgrounds, since they often lie below these limits.

Additional requirements on tracks are placed on the impact parameters, $dr < 2\,\text{cm}$ and $|dz| < 4\,\text{cm}$. This rejects particles not coming from the IP, such as beam backgrounds and low-multiplicity backgrounds. It still allows for long lived particles, such as B and D mesons, but removes K_S^0 or K_L^0. Also, to identify a valid track, at least one hit in the CDC is needed. We want to reject clusters not corresponding to photons and so we require clusters to have no matching track. To reject beam background photons we require the time between the cluster and the collision to be less than 50 ns. To further exclude beam backgrounds, we reject clusters where the ratio between the energy in the inner 3×3 ECL crystals and the energy in the surrounding 5×5 crystals (exluding the corners) is less than 0.9.

4.5.3 Neutrino Reconstruction

While the neutrino cannot be detected and therefore not reconstructed, it can indirectly be reconstructed from the ROE through constructing a missing energy. This is based on the assumption that all final state particles from the $\mathrm{B\overline{B}}$ decays are reconstructed in either the signal side or the ROE. The missing energy is calculated from the $\Upsilon(4S)$ energy, $E_{\Upsilon(4S)}$, the total energy of the signal side, E_{signal}, and the total energy of the ROE, E_{ROE}, as

$$E_{\text{miss}} = E_{\Upsilon(4S)} - E_{\text{signal}} - E_{\text{ROE}}. \tag{4.2}$$

The missing energy includes the energy of all undetected particles, such as neutrinos, which cannot be detected by Belle II. This is especially interesting when a single neutrino is expected in the event,

which is true for the signal decay $B \to \pi \ell \nu$, assuming the other B meson decay did not involve a neutrino. The neutrino energy can then be reconstructed from $E_{\text{miss}} = E_\nu$, and since the neutrino is assumed massless, $E_\nu \approx p_\nu = p_{\text{miss}}$. When looking at the missing mass squared of such an event, i.e. $p_\nu^2 = m_{\text{miss}}^2$, it should be centred at 0 if there is only one neutrino.

4.6 Reconstruction Efficiency

4.6.1 Momentum Transfer (q^2) Binning

As was explained in Section 2.3, the analysis is performed in bins of reconstructed q^2 to allow for the extraction of $|V_{ub}|$. The reconstructed q^2 spectrum is divided into five bins with the following ranges:

- $q1$: $q^2 \leq 5\,\text{GeV}^2$
- $q2$: $5\,\text{GeV}^2 < q^2 \leq 10\,\text{GeV}^2$
- $q3$: $10\,\text{GeV}^2 < q^2 \leq 15\,\text{GeV}^2$
- $q4$: $15\,\text{GeV}^2 < q^2 \leq 20\,\text{GeV}^2$
- $q5$: $q^2 > 20\,\text{GeV}^2$

The reconstructed q^2 is calculated from the pion and B meson four-vectors as given in Equation 2.2. The B meson four-vector cannot be reconstructed, due to the neutrino, and we therefore assume that the B meson is at rest when it decays. This means its energy is equal to half the centre of mass energy and its momentum is zero.

4.6.2 Resolution

The true q^2, q^2_{true}, and reconstructed q^2, q^2_{rec}, distributions for correctly reconstructed signal events in the $B^\pm \to \pi^0 e\nu$ and $B^0 \to \pi^\pm e\nu$ modes are shown in (a) and (b) of Figure 4.1, respectively. For both modes the q^2_{rec} deviates from q^2_{true} at low q^2. This could be due to: (1) the incorrect assumption that the B meson is at rest when it decays or (2) pion reconstruction effects. To check for (1) we recalculate q^2, now without assuming that the B is at rest, but instead using the true B four-momentum known from generator information. This "corrected q^2" is also shown in Figure 4.1. For the B^0 modes this correction almost entirely removes the discrepancy at low q^2, while for the $B^\pm$ modes a difference at low q^2 persists.

Any leftover discrepancy between the true and reconstructed q^2 must then be due to (2): a mismeasurement of the pion momentum. Since the difference in the $B^\pm$ modes is a lot greater, the reconstruction of the neutral pions must be worse than that of the charged pions, which is expected due to the limited energy resolution of the photons.

The difference between reconstructed and true q^2 can be quantified by the resolution, $q^2_{\text{true}} - q^2_{\text{rec}}$. A 2D histogram of true versus reconstructed q^2 for correctly reconstructed signal events is shown in (c) of Figure 4.1. In the ideal case this would result in a line with a gradient of 1. The histogram, however, shows that the resolution in q^2 is not ideal. As an example, the q^2 resolution in the five q^2 bins of the $B^\pm \to \pi^0 e\nu$ mode is given in (d) of Figure 4.1. Due to reconstruction effects, the resolution in each bin is not Gaussian. The full width half maximum (FWHM) and peak position were determined for each q^2 bin in each mode and they are given in the appendix. For low q^2 bins there is a negative bias in the peak position, so that the reconstructed q^2 is on average greater than the true q^2. This bias

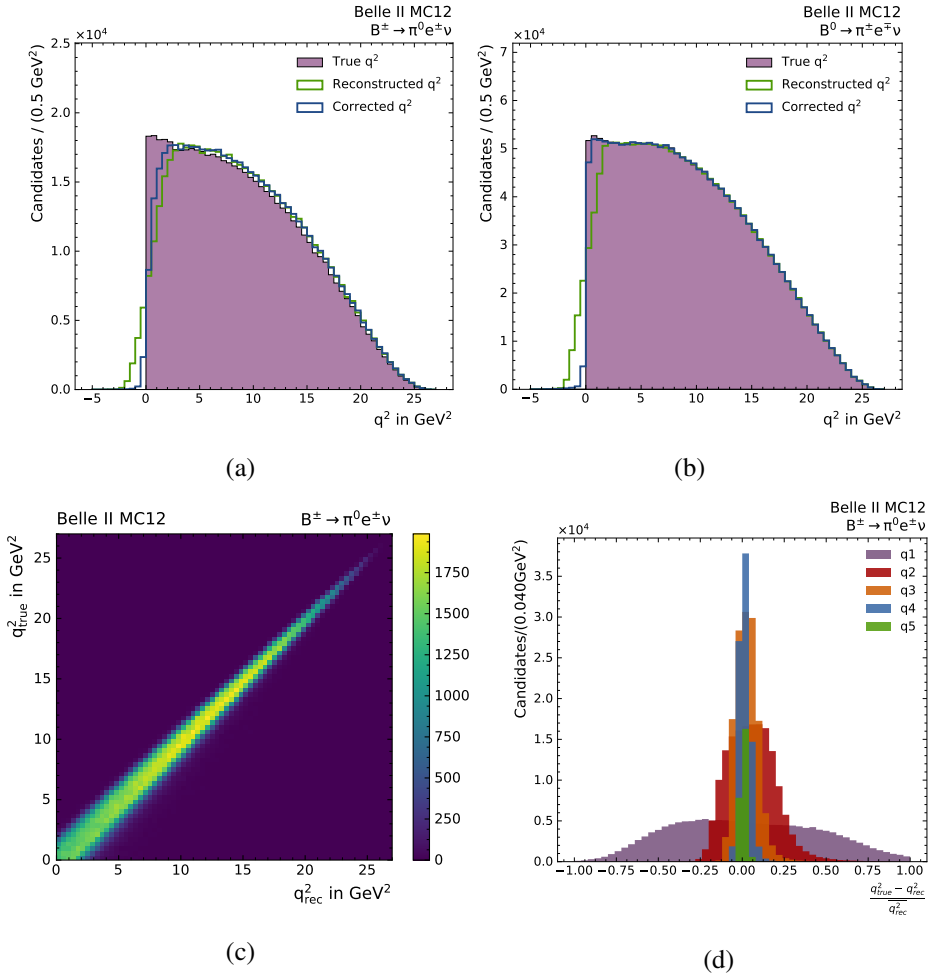


Figure 4.1: Plots for correctly reconstructed signal events from signal MC. True, reconstructed and corrected q^2 for (a) the $B^{\pm} \to \pi^0 e\nu$ mode and (b) the $B^0 \to \pi^{\pm} e\nu$ mode. (c) 2D histogram of true versus reconstructed q^2 in the $B^{\pm} \to \pi^0 e\nu$ mode. (d) The distributions of the normalised resolutions of q^2 for the five q^2 bins in the $B^{\pm} \to \pi^0 e\nu$ mode.

disappears for higher q^2 bins and the peak is centred closer to 0. The FWHM is greatest for the lowest q^2 bins and decreases for higher q^2 bins.

4.6.3 Reconstruction Efficiency

A reconstruction efficiency can be defined and found from MC to describe the loss of events due to the reconstruction described above. This is neccessary in order to determine the number of signal events in a certain q^2_{true} bin, j, from the number of correctly reconstructed events in a q^2_{rec} bin, i. The reconstruction efficiency is given by

$$\epsilon^i_{\text{rec}} = \frac{N^i_{\text{rec}}}{N^j_{\text{gen}}}. \tag{4.3}$$

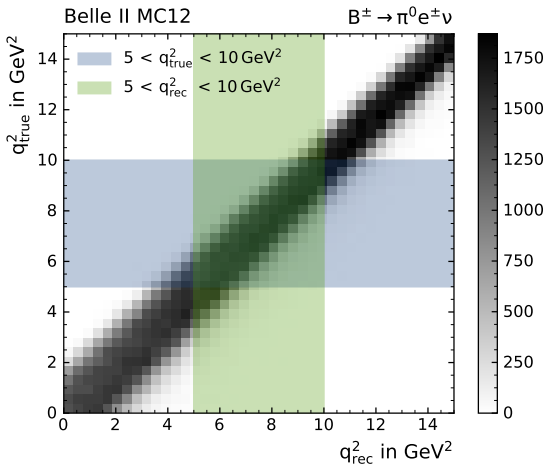


Figure 4.2: Illustration of the bin migration factor calculation from the 2D histogram of true versus reconstructed q^2 for correctly reconstructed signal events in the $B^{\pm} \to \pi^0 e\nu$ mode. The $q2$ bin for q^2_{rec} is coloured green and the same bin for q^2_{true} is highlighted in blue.

The results for the five q^2 bins are given in Table 4.1 and shown in (a) and (b) of Figure 4.3 for the $B^0 \to \pi^{\pm} e\nu$ and $B^{\pm} \to \pi^0 e\nu$ modes. The reconstruction efficiencies in the $B^{\pm}$ modes are ~30% lower than the ones in the B^0 modes. We expect that this is a result of the poor reconstruction of the π^0.

Mode	**Reconstruction Efficiency**				
	q1	***q2***	***q3***	***q4***	***q5***
$B^0 \to \pi^{\pm} e\nu$	0.738	0.786	0.815	0.807	0.798
$B^{\pm} \to \pi^0 e\nu$	0.455	0.529	0.551	0.548	0.490
$B^0 \to \pi^{\pm} \mu\nu$	0.738	0.792	0.824	0.819	0.804
$B^{\pm} \to \pi^0 \mu\nu$	0.454	0.532	0.559	0.553	0.496

Table 4.1: Reconstruction efficiencies for the five q^2_{rec} bins of the four modes.

4.6.4 Bin Migration Factor

In our case the reconstruction efficiency includes effects from reconstruction selections on the pion, the acceptance due to the detector geometry, the tracking efficiency and bin migration. Essentially, the effect of all these components can be investigated separately. We focus on the effects of bin migration on the reconstruction efficiency, since this varies if a different form factor model is used in the MC.

Bin migration occurs when an event lies in a certain q^2_{true} bin, j, but it lies in a q^2_{rec} bin, i, where $i \neq j$. This can be seen for the $q2$ bin in Figure 4.2. The events in the blue band are in the q^2_{true} $q2$ bin and the ones in the green band are in the q^2_{rec} $q2$ band. In the ideal case, where no bin migration occurs, all events would lie in the overlapping region, $N^i_{\text{rec}} = N^j_{\text{rec}} = N^{i+j}_{\text{rec}}$ and therefore $N^i_{\text{rec}}/N^j_{\text{rec}} = 1$. We call this fraction the bin migration factor, BMF.

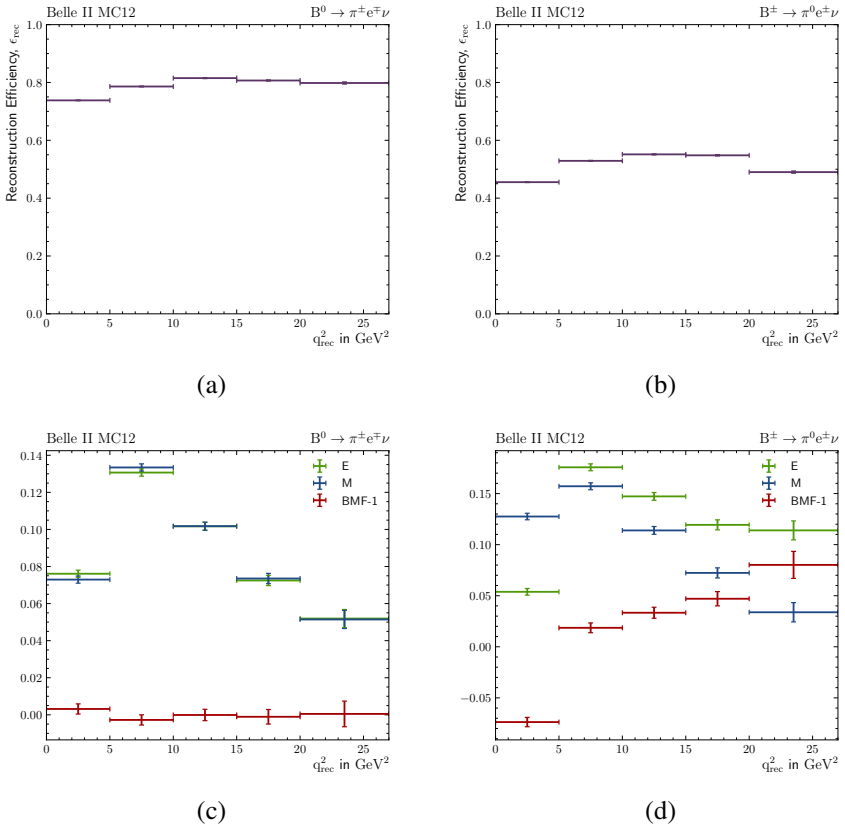


Figure 4.3: (a)-(b) The reconstruction efficiencies and (c) -(d) E, M and $BMF - 1$ across the five q^2_{rec} bins for the $B^0 \to \pi^\pm e\nu$ mode (left) and the $B^\pm \to \pi^0 e\nu$ mode (right) found from signal MC.

If the $BMF \neq 1$, bin migration occurs. If the $BMF > 1$, more events are in the q^2_{rec} bin than in the q^2_{true} bin, while if the $BMF < 1$, fewer events are in the q^2_{rec} bin than in the q^2_{true} bin. Going back to Figure 4.2, the number of events only lying in the green band are called N^i_{extra} and the number of events only lying in the blue band are called N^j_{missing}. The total number of events in the green and blue bands can then be written as $N^i_{\text{rec}} = N^{i+j}_{\text{rec}} + N^i_{\text{extra}}$ and $N^j_{\text{rec}} = N^{i+j}_{\text{rec}} + N^j_{\text{missing}}$, respectively. Substituting and dividing by N^j_{rec}, the BMF is given by

$$BMF = \frac{N^i_{\text{rec}}}{N^j_{\text{rec}}} = 1 + E - M, \tag{4.4}$$

where $M = \frac{N^j_{\text{missing}}}{N^j_{\text{rec}}}$ and $E = \frac{N^i_{\text{extra}}}{N^j_{\text{rec}}}$. E, M and $BMF - 1$ for the five q^2_{rec} bins are shown in (c) and (d) of Figure 4.3 for the $B^0 \to \pi^\pm e\nu$ and $B^\pm \to \pi^0 e\nu$ modes. The BMF is consistent with 1 throughout for the B^0 modes and E and M cancel approximately. This is not the case for the $B^\pm$ modes. Again, this must be due to the poor reconstruction of the π^0 momentum, which systematically lowers the q^2_{rec} due to lost energy in the ECL.

CHAPTER 5

Signal Selection

This chapter will discuss the steps involved in the selection of $B \to \pi\ell\nu$ signal events. The branching fraction of $B^{\pm} \to \pi^0\ell\nu$ is $(7.80 \pm 0.27) \times 10^{-5}$ and of $B^0 \to \pi^{\pm}\ell\nu$ is $(1.50 \pm 0.06) \times 10^{-4}$ [10]. Since this thesis looks at early Phase 3 data with an expected integrated luminosity of $\sim 6\,\text{fb}^{-1}$, corresponding to $\sim 10^6\ B\overline{B}$ events, the total expected number of signal events in each mode is ~1000 events. An untagged analysis is performed, due to this small number of expected signal events. While the signal purity (number of signal events/total number of events) achievable in an untagged analysis is lower than in a tagged one, untagged analyses retain a relatively high signal efficiency.

Characteristics of the signal mode will be discussed and selections on these along with other variables will be optimised for $6\,\text{fb}^{-1}$ on Monte Carlo. Before further background suppression with BDTs is discussed in the next chapter, a best candidate selection is performed. In the following, generally, if no significant difference between the modes is observed, an example plot for the $B^{\pm} \to \pi^0 e\nu$ mode will be shown. Where a selection is performed separately in the five q^2 bins, one example plot is shown in one of the q^2 bins of the $B^{\pm} \to \pi^0 e\nu$ mode. The plots for the other modes and q^2 bins is given in the appendix.

5.1 Signal Characteristics

Firstly, upon reconstructing the signal decay, two tracks are expected for the decay to a charged pion and one track is expected for the decay to a neutral pion. For $B^{\pm}$ decays the lepton needs to have the same charge as the B meson, while in B^0 decays, the pion and the lepton need to have opposite charges. There is also a correlation between the B^0-flavour and the lepton charge. The B^0 decays into a negatively-charged lepton and the $\overline{B^0}$ decays into a positively-charged lepton.

The kinematics of the signal decay can sufficiently be described by the momentum transfer and the lepton momentum. The q^2 of the signal decay $B \to \pi\ell\nu$ in terms of the lepton, pion, B meson and neutrino momentum can be written as

$$\begin{aligned} q^2 &= m_B^2 + m_\pi^2 - 2m_B E_\pi \\ &= m_\ell^2 - 2p_\ell p_\nu \end{aligned} \tag{5.1}$$

in the centre of mass frame of the B meson. The endpoint is $\sim 27.7\,\text{GeV}^2$, corresponding to the production of the pion at rest ($E_\pi = 0$), and is limited by the B meson and pion masses. The lower

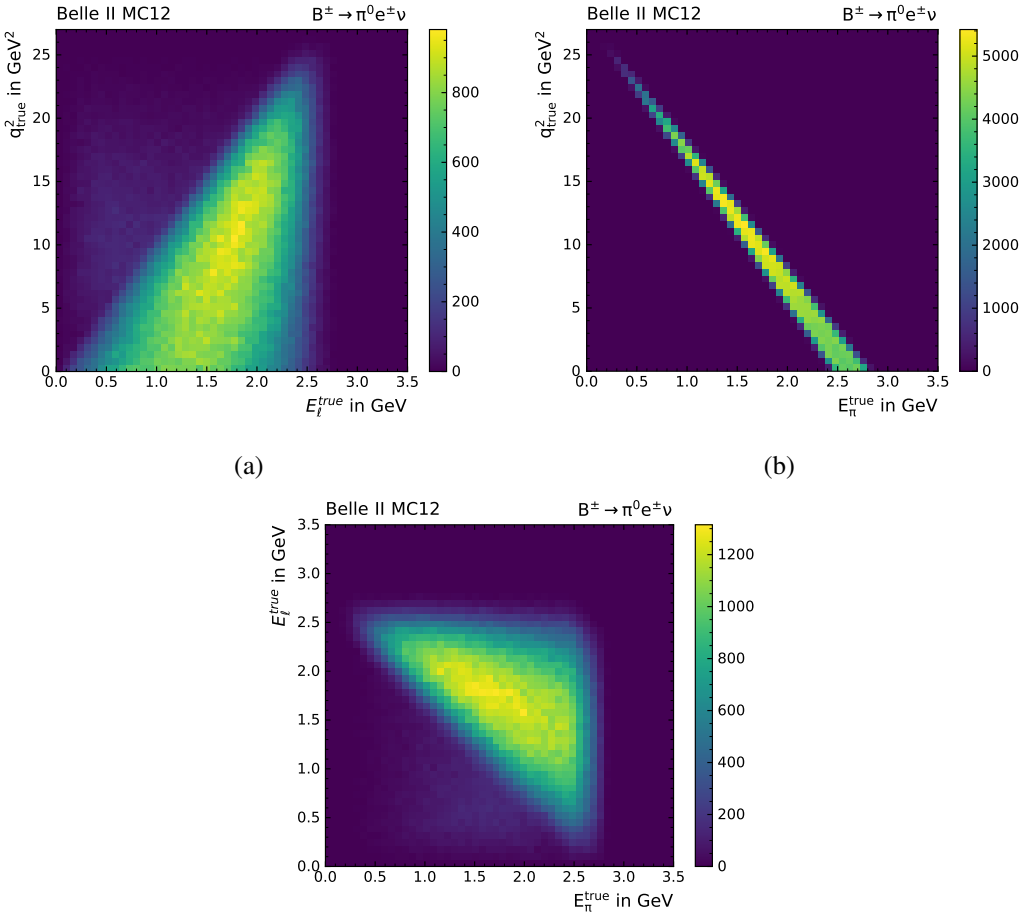


Figure 5.1: 2D histograms describing the phase space of the signal decay from signal MC. (a) and (b) are plots of the momentum transfer, q^2, versus the lepton, E_ℓ, and pion energy, E_π, respectively. (c) shows the 2D distribution of E_ℓ versus E_π.

bound occurs when there is just enough energy to produce a lepton and a neutrino at rest ($p_\ell = m_\ell$, $p_\nu = m_\nu = 0$). The lower bound is set by the lepton mass and it is ~ 0. The relationships between the true q^2, the true lepton energy, E_ℓ, and the true pion energy, E_π, are shown in Figure 5.1. As seen in (a) and (b), at high q^2 the lepton has a high energy and the pion little energy, while at low q^2 the lepton generally has less energy but the pion is very energetic. The phase space for signal in terms of E_ℓ and E_π is shown in (c). The sum of the energies is generally greater than 2.5 GeV.

Looking at $B^\pm \to \pi^0 \ell \nu$ events, where the π^0 further decays into two photons within ~ 100 as [10], additional kinematic information can be obtained from the π^0 decay. Since the π^0 from a signal B meson decay generally does not decay at rest, the two photons will be boosted and the angle between them should be close to 0.

The kinematic consistency of a reconstructed event with signal can be checked using the variables ΔE, M_{bc} and $\cos\theta_{BY}$. They are often chosen as signal extraction variables due to their high signal-

background separation power and will be explained next. Both ΔE and M_{bc} use the fact that the initial energy state of the $\Upsilon(4S)$ is known. The beam constrained mass, M_{bc}, is a measure of the invariant mass of the B candidate, calculated when the energy of the B meson is replaced with the beam energy, $E_{beam} = \frac{\sqrt{s}}{2}$, where $\sqrt{s}$ is the centre of mass energy. M_{bc} is given by

$$M_{bc} = \sqrt{E_{beam}^2 - p_B^2}, \tag{5.2}$$

and should be equal to the B meson mass for correctly reconstructed signal (5.28 GeV). Since this did not use the reconstructed B candidate energy, a second variable can be constructed, which is the difference between the reconstructed and the expected energy. This is called the energy difference, ΔE, and is given by

$$\Delta E = E_B - E_{beam}. \tag{5.3}$$

For correctly reconstructed B candidates ΔE should be consistent with 0.

In the decay $B \to \pi \ell \nu$ the lepton and pion can collectively be called a Y candidate. The recoil mass of a B meson that decays into a Y candidate is given by

$$\begin{aligned} m_r^2 &= (P_B - P_Y)^2 \\ &= m_B^2 + m_Y^2 - 2E_B E_Y - 2p_B p_Y \cos\theta_{BY}, \end{aligned} \tag{5.4}$$

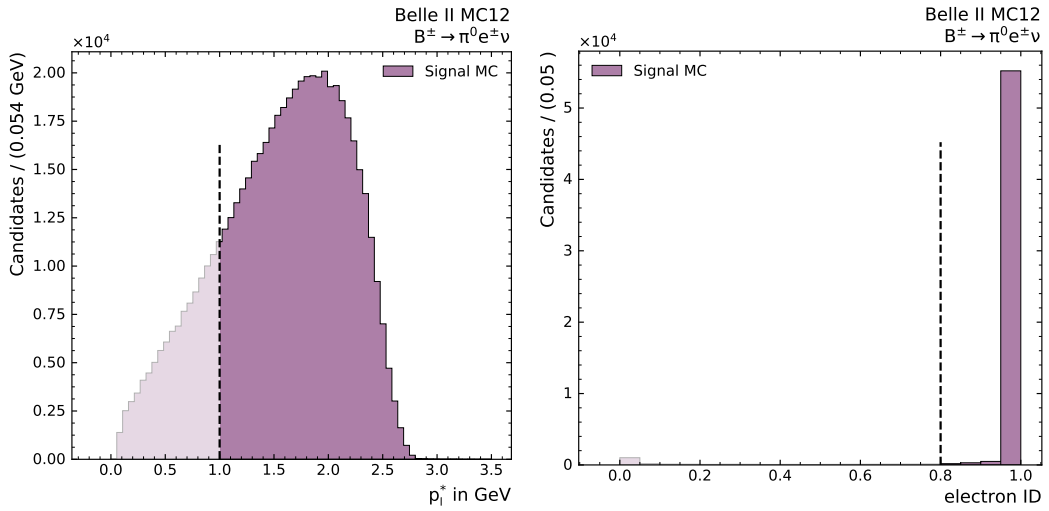
where P_B, p_B and m_B are the four-momentum, magnitude of the three-momentum and the mass of the B meson, respectively. P_Y and p_Y are the four-momentum and magnitude of the three-momentum of the Y candidate. m_Y is the invariant mass of the Y candidate. $\cos\theta_{BY}$ is the cosine of the angle between p_B and p_Y. The recoil mass should be that of the neutrino for correctly reconstructed signal and therefore be consistent with 0. This results in the following expression for $\cos\theta_{BY}$:

$$\cos\theta_{BY} = \frac{2E_B E_Y - m_B^2 - m_Y^2}{2p_B p_Y}. \tag{5.5}$$

If $\cos\theta_{BY}$ is calculated in this way, it should lie between -1 and 1 for signal, where only one neutrino is missing.

5.2 Lepton Selection

First we select electron and muon candidates. To validly use the global particle identification variables, the electron and muon ID, the centre of mass particle momentum needs to be greater than 1 GeV. We therefore require that the lepton centre of mass momentum, p_ℓ^*, is greater than 1 GeV, and that the electron ID is greater than 0.8 for electron candidates, and the muon ID is greater than 0.8 for muon candidates. This rejects low-momentum background and it also reduces the number of misidentified leptons. These are mainly pions, that have been reconstructed as leptons. The lepton momentum distribution and the 1 GeV requirement for the $B^\pm \to \pi^0 e\nu$ mode are shown in (a) of Figure 5.2. The electron ID distribution and the selection on the electron ID, after the lepton momentum selection has been performed, are shown in (b). Due to the strong correlation between lepton momentum and the momentum transfer, the selection on the lepton momentum mainly affects events in the lowest q^2 bin. The average efficiency of the lepton selections over all q^2 bins and all modes is $\approx 80\%$.



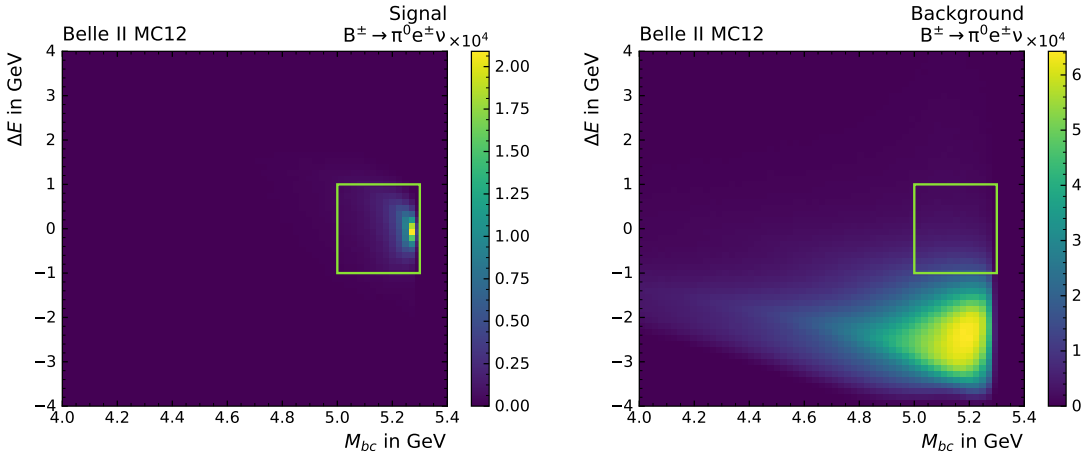
(a) (b)

Figure 5.2: The distributions in the $B^{\pm} \to \pi^0 e\nu$ mode from signal MC of (a) the reconstructed centre of mass lepton momentum and (b) the electron ID after the previous selection. The selections on the lepton momentum and the electron ID are shown.

5.3 Fit Region Selection

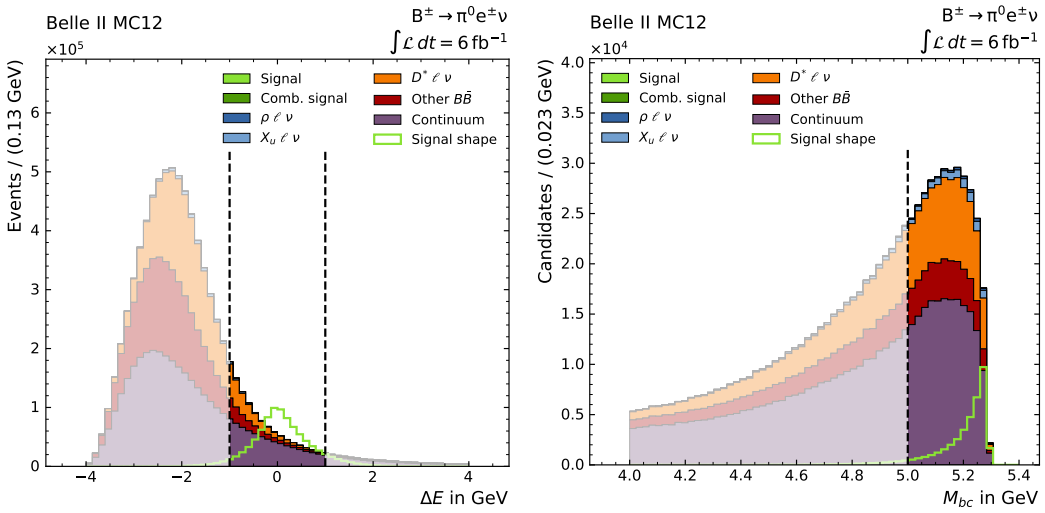
To extract signal we will perform a simultaneous fit of the energy difference, ΔE, and the beam constrained mass, M_{bc}, as defined in Section 5.1. At this point, we select a fit region in ΔE and M_{bc} that contains most of the signal events. For correctly reconstructed signal events ΔE should peak at 0 and M_{bc} at 5.28 GeV. Figure 5.3 shows 2D histograms of ΔE and M_{bc} for signal in (a) and background in (b) after the lepton selection. The selected fit region is given by $5.0 \leq M_{bc} \leq 5.3$ GeV and $|\Delta E| \leq 1.0$ GeV, and in (a) and (b) it is represented by the area inside the green box.

Before this selection, for the B^0 modes, the amount of background events is ~5000 times greater than the number of signal events, and for the $B^{\pm}$ modes, the number of background events is ~10000 times greater than the number of signal events. After the fit region selection the number of background events decreases to ~ 1000 times the number of signal events for all modes. At the same time, this selection retains approximately 80% of all signal events, while rejecting on average 93% of background. The distributions of and the selections on ΔE and M_{bc} are shown in (a) and (b) of Figure 5.4, where the M_{bc} distribution is the one after the ΔE selection has been performed. Both the reconstruction of ΔE and M_{bc} rely on the assumption that there is only one neutrino in the event, and the subsequent indirect reconstruction of the neutrino energy and momentum. Since the neutrino is reconstructed from the ROE, and this contains, for example, beam background particles, the distributions of ΔE and M_{bc} for signal are wider than those expected from an ideal reconstruction. At this point, the fraction of signal events is too small and signal cannot be seen. The background components will be explained in the next section.



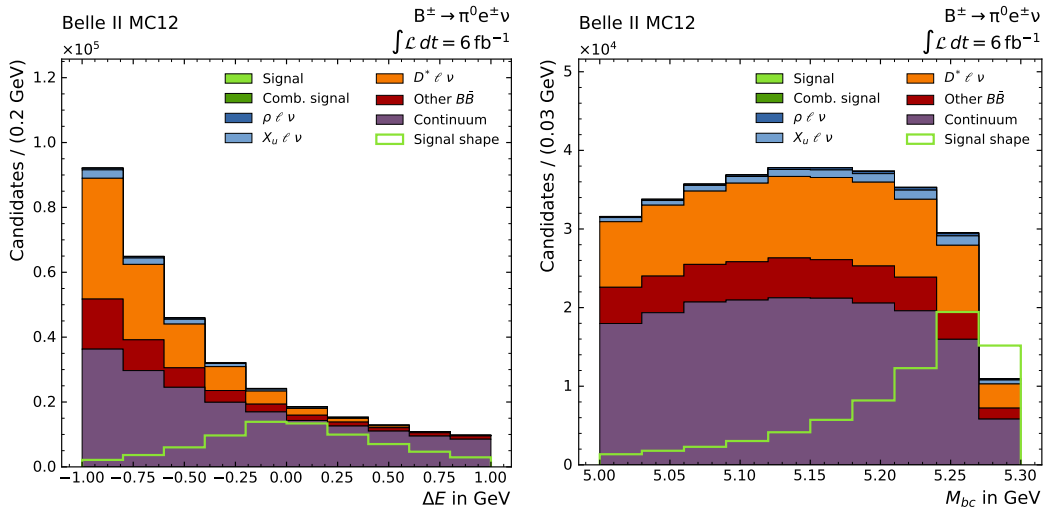
(a) (b)

Figure 5.3: 2D distributions of ΔE and M_{bc} for the $B^{\pm} \to \pi^0 e\nu$ mode in MC for (a) signal and (b) background. The fit region is given by the area inside the green box.



(a) (b)

Figure 5.4: (a) ΔE distribution for the $B^{\pm} \to \pi^0 e\nu$ mode with the selection on ΔE shown. (b) The M_{bc} distribution after the selection shown in (a). The selection on M_{bc} is displayed.



(a) (b)

Figure 5.5: The (a) ΔE and (b) M_{bc} distributions in MC for the $B^{\pm} \to \pi^0 e\nu$ mode after the lepton and fit region selections have been performed.

5.4 Background Classification

The ΔE and M_{bc} distributions for the $B^{\pm} \to \pi^0 e\nu$ mode after these selections have been performed are shown in Figure 5.5. The events can be split into three main categories: signal, continuum and $B\overline{B}$ background. These categories and their subcomponents are explained below. An arbitrarily scaled version of the signal shape is also shown and, as expected, signal peaks at 0 in ΔE and at 5.28 GeV in M_{bc}.

Signal

Signal is divided into correctly reconstructed $B \to \pi\ell\nu$ events, which is called signal, and into incorrectly reconstructed $B \to \pi\ell\nu$ events, which is called combinatorial signal. An incorrectly reconstructed $B \to \pi\ell\nu$ event, in this case, is an event in which the lepton has been correctly reconstructed and where it originated from a $B \to \pi\ell\nu$ decay, but the pion was misidentified. The number of these events scales with the signal branching fraction, and it is therefore called combinatorial signal.

Continuum

As explained in Section 3.3.1, instead of an $\Upsilon(4S)$, electron-positron collisions can produce quark pair continuum background. Since some τ pair production events behave similarly to continuum background, we include it in the continuum background category.

$\mathrm{B\overline{B}}$ Background

The $\mathrm{B\overline{B}}$ background is composed of $\Upsilon(4S) \to \mathrm{B\overline{B}}$ events that do not contain $B \to \pi\ell\nu$ decays. It includes events with fake or secondary leptons. A fake lepton is a track that has been reconstructed as a lepton, but that is not actually one. Secondary leptons, on the other hand, were correctly identified, but are not direct daughters of a B meson decay. These components are both part of the other $\mathrm{B\overline{B}}$ category.

All other events in the $\mathrm{B\overline{B}}$ background have a true lepton coming from a B meson decay, but the pion is misidentified, i.e. $B \to X\ell\nu$ events. We can split these into $B \to X_c\ell\nu$ and $B \to X_u\ell\nu$ events. X_c can be a D, D_s, D^* or higher-order resonances collectively called D^{**}. All but the $B \to D^*\ell\nu$ events, which have their own category, are also part of the other $\mathrm{B\overline{B}}$ histogram category. The $B \to X_u\ell\nu$ category is further split into a separate $B \to \rho\ell\nu$ and a $B \to X_u\ell\nu$ category, that contains all other possible semileptonic decays into X_u particles, such as ω, η and others. If any other process occurs it will be part of the $B \to X_u\ell\nu$ category.

Other Backgrounds

We do not include low-multiplicity backgrounds in our Monte Carlo samples, since we will employ a selection to remove the majority of it. The selection will be explained in the following section. On the other hand, beam background is included in the Monte Carlo, since it, as discussed in Section 4.1, is mixed into the MC samples. It does, however, not compose a separate background category as it occurs in all of the previously discussed categories.

5.5 Selections

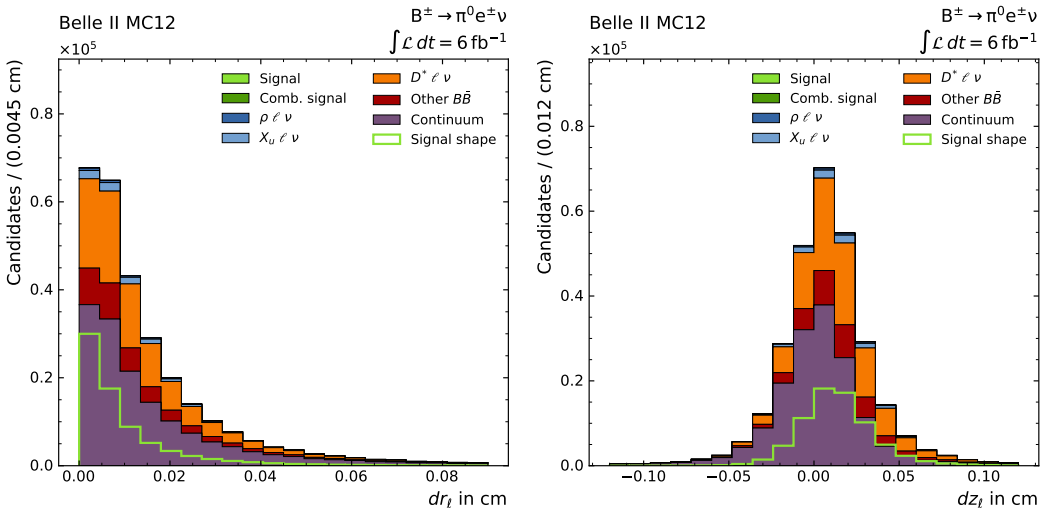
At this point, we perform some selections based on the properties of the detector and the physics of the signal decay. The selections aim at increasing the chosen figure of merit, FOM, given by

$$\mathrm{FOM} = \frac{N_{\mathrm{sig}}}{\sqrt{N_{\mathrm{sig}} + N_{\mathrm{bkg}}}}. \tag{5.6}$$

The FOM is a measure of the statistical significance of signal on top of background. N_{sig} and N_{bkg} are the numbers of signal and background events, respectively. Maximising the FOM for a certain selection maximises the statistical separation between signal and background. For the preselections we optimise the FOM for a MC sample of $6\,\mathrm{fb}^{-1}$ over all q^2 bins, while for any selections after that point we optimise the FOM in each q^2 bin separately.

In addition, we study the change in purity for each selection, defined as the number of signal events divided by the total number of events. At this point in the selection, the FOM ranges from approximately 0.5 in the $B^\pm$ modes to approximately 0.8 in the B^0 modes. The purity in the $B^\pm$ modes is ~0.07% and in the B^0 modes it is ~0.1%. The $B^0 \to \pi^\pm e\nu$ mode has the highest purity (0.14%) and the $B^\pm \to \pi^0\mu\nu$ mode has the lowest purity (0.06%).

We also monitor the signal efficiency and the background rejection of each of the selection steps. Special care is taken in observing the evolution of the signal efficiency in each of the q^2 bins. Ideally, this would be a flat distribution and the selections would affect all q^2 bins equally. This flat distribution is desired since we want to avoid sharp features in the q^2 distribution, as these would be problematic in the extraction of $|V_{ub}|$ using the form factor.



(a) (b)

Figure 5.6: Distributions of (a) dr and (b) dz of the electron in the $B^{\pm} \to \pi^0 e\nu$ mode in MC. The distributions for the pion impact parameters are similar.

5.5.1 Preselections

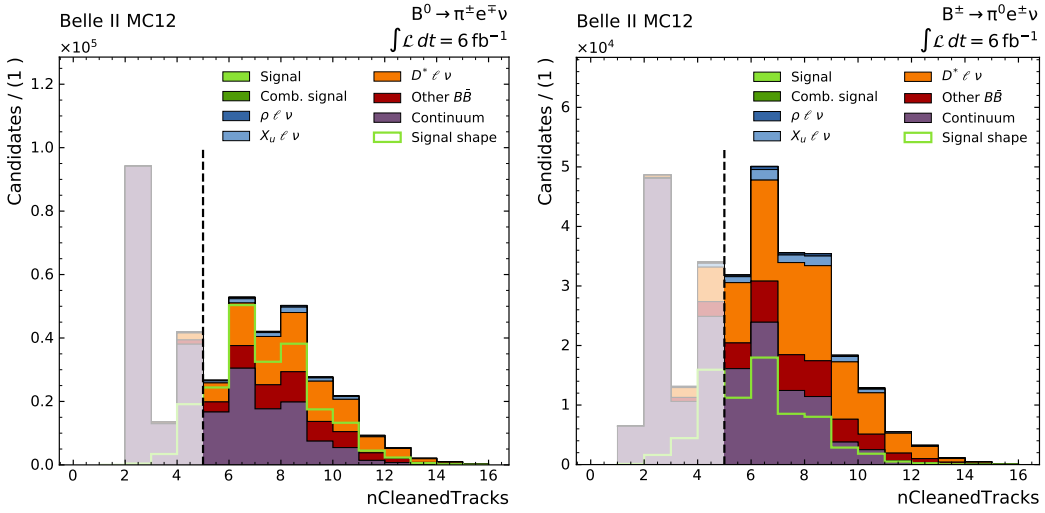
Impact Parameters

The impact parameters give a measure of whether a particle is coming from the IP. In addition to the standard list requirements on the $\pi^{\pm}$ impact parameters, explained in Section 4.5.1, we place a requirement of $dr < 0.5\,\mathrm{cm}$ and $|dz| < 2\,\mathrm{cm}$ on the lepton and π^0 impact parameters. This removes long lived particles and events with displaced vertices, but in terms of the collision point position it is still relatively loose. This can be seen in the plots of the lepton dr and dz in Figure 5.6, where the required values lie outside the axis range. We choose a loose cut, with signal efficiencies up to 99.99%, since in early Belle II the IP is moving and its position is not accurately known.

Number of Tracks

Looking at the number of tracks on the signal side, one track is expected for the $B^{\pm}$ modes and two tracks are expected for the B^0 modes. Except for some rare cases, where, for example, the second B meson also decays into two neutral particles, at least two tracks are expected from the other side. Therefore, in general, signal events should consist of at least four tracks. The average multiplicity of a B meson decay at the $\Upsilon(4S)$ energy is ~5.4, and so the number of tracks for signal events should peak around seven tracks [28].

Requiring at least four tracks per event removes most of the low-multiplicity background, such as Bhabha scattering or muon pair production, but it does not remove two-photon background events, which are expected to have four charged tracks. To also suppress this background a requirement of at least five tracks per event is imposed. This requirement is placed on the number of "cleaned" tracks (nCleanedTracks). nCleanedTracks is the number of tracks in an event that obey the same dr and dz requirements imposed on the signal tracks ($dr < 0.5\,\mathrm{cm}$ and $|dz| < 2\,\mathrm{cm}$). The distributions of



(a) (b)

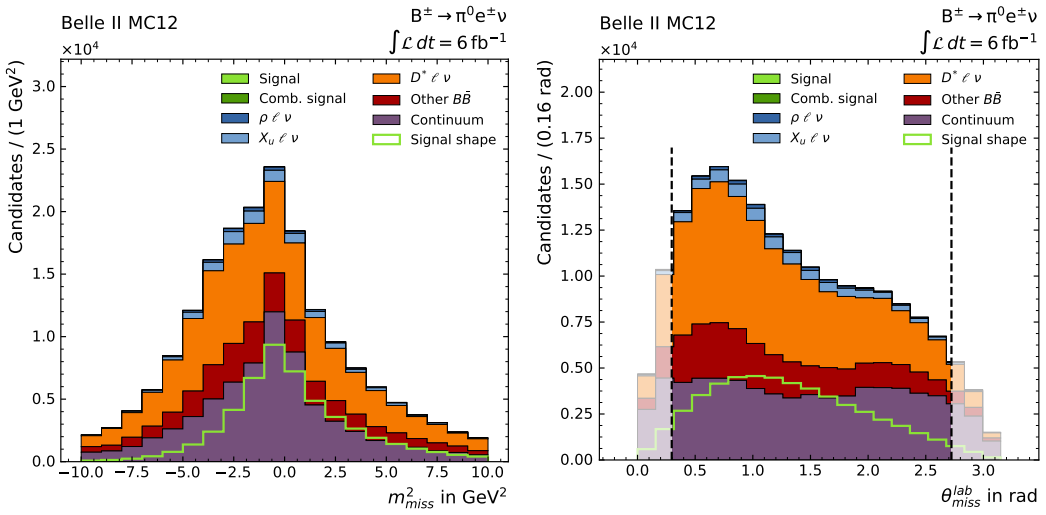
Figure 5.7: Distribution of the number of "cleaned" tracks, nCleanedTracks, in MC for the (a) $B^0 \to \pi^\pm e\nu$ and (b) $B^\pm \to \pi^0 e\nu$ modes with the requirement of nCleanedTracks > 4 shown. Low-multplicity background is not included in these distributions.

nCleanedTracks for the $B^0 \to \pi^\pm e\nu$ and $B^\pm \to \pi^0 e\nu$ modes with the imposed selections are shown in Figure 5.7.

For all modes the continuum MC strongly peaks at 2 in nCleanedTracks. This is mainly due to tau pair events. At the same time, the expected peak around 7 in nCleanedTracks is observed for the $B\overline{B}$ MC. The distributions for the $B^\pm$ modes are shifted to the left compared to those of the B^0 modes. This is expected, since the $B^\pm$ modes have one less track on the signal side due to the neutral pion. This is also mirrored in the signal efficiencies of the selection on nCleanedTracks. For the B^0 modes the signal efficiencies are ~89%, while for the $B^\pm$ modes they are ~70%. At the same time, the selection rejects 34% and 45% of background in the B^0 and $B^\pm$ modes, respectively. However, these numbers should be even higher if we include the rejection of low-multiplicity background, which is not shown.

Neutrino Reconstruction

The indirect reconstruction of neutrinos in semileptonic decays through missing momentum, p_{miss}, was described in Section 4.5.3. For signal the missing mass squared, m^2_{miss}, is expected to peak at 0. However, as seen in the plot of m^2_{miss} in the lab frame in (a) of Figure 5.8 the signal distribution is relatively broad. This means there are more undetected particles than expected, such as additional neutrinos from the second B meson decay, as well as extra tracks and clusters, that contribute to the missing momentum. Taking a look at the polar angle of the missing momentum, θ_{miss}, massive particles with a polar angle outside the detector acceptance will be undetected, counted towards the missing momentum, and thus bias the m^2_{miss} distribution. We therefore reject candidates with a missing momentum angle in the lab frame outside the detector acceptance and require $0.2967 \leq \theta_{\text{miss}} \leq 2.7227\,\text{rad}$. The distribution of θ_{miss} along with the selection are shown in (b) of Figure 5.8. The signal efficiciency across all modes is ~95%.



(a) (b)

Figure 5.8: Distribution of (a) the missing mass squared, m^2_{miss} and (b) the missing momentum angle in the lab frame, θ_{miss}. The selection is shown. Both distributions come from MC for the $B^\pm \to \pi^0 e\nu$ mode.

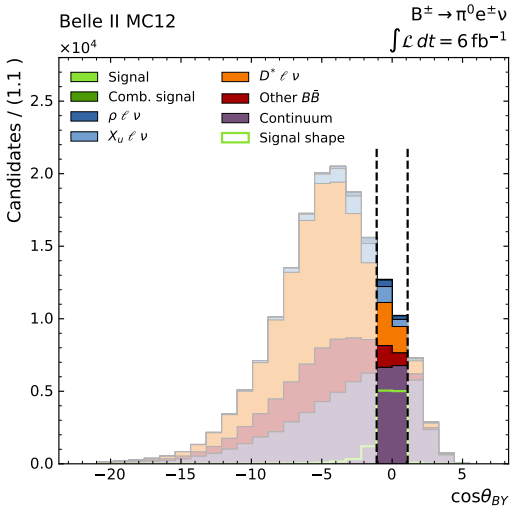


Figure 5.9: Distribution of $\cos\theta_{BY}$ for the $B^\pm \to \pi^0 e\nu$ mode in MC. The selection is shown.

Pion-Lepton Candidate

To reduce the amount of misreconstructed continuum and $B\overline{B}$ background events, the cosine of the angle between the B and the Y candidate, $\cos\theta_{BY}$, as discussed in Section 5.1, is considered. $\cos\theta_{BY}$ should lie between -1.0 and 1.0 for correctly reconstructed signal. To account for reconstruction effects, we require $|\cos\theta_{BY}| < 1.1$. The distribution of $\cos\theta_{BY}$ and the selection for the $B^\pm \to \pi^0 e\nu$ mode are shown in Figure 5.9. The selection rejects ~83% of background with a signal efficiency of ~88%. In addition, the purity in all modes increases by a factor 4.

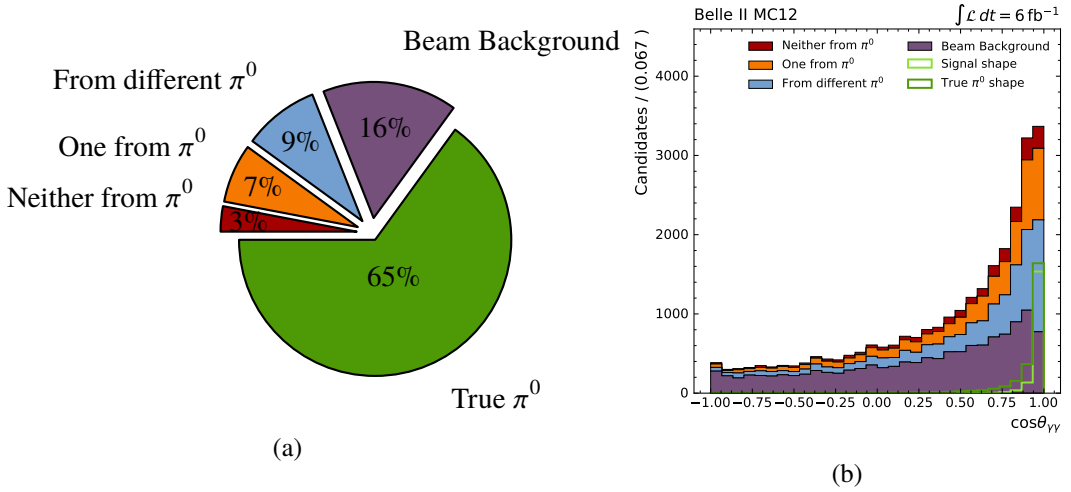


Figure 5.10: (a) Pie chart giving the relative fractions of true and misreconstructed π^0s obtained from MC. (b) The distribution of the cosine of the angle between the two photons that the π^0 is misreconstructed from in MC. The distribution is split into categories depending on the origin of the photons. The shape of the true pion and signal distributions is also overlayed.

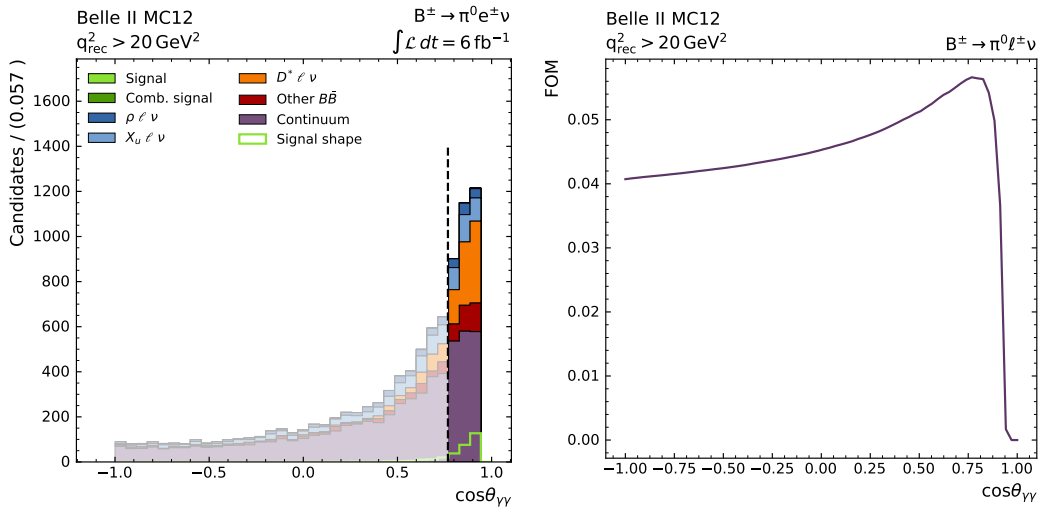
5.5.2 π^0 Reconstruction

At this point in the selection approximately one third of the reconstructed π^0s in the $B^\pm \to \pi^0 \ell \nu$ modes are misreconstructed pions, while only one tenth of the charged pions in the $B^0 \to \pi^\pm \ell \nu$ modes are misidentified. Since the π^0 is reconstructed from two photons and beam background contributes additional photons to events, the π^0 reconstruction suffers from combinatorics. Approximately 50% of the misreconstructed π^0s have been reconstructed from one or two beam background particles based on MC. The other half have been reconstructed from non-beam background photons. This category can be further split into π^0s reconstructed from two photons where both photons come from different π^0 decays, where one photon comes from a π^0 decay, and where neither comes from a π^0 decay. The relative fractions are shown in the pie chart shown in (a) of Figure 5.10.

As explained in Section 5.1, the photons coming from a signal $B^\pm \to \pi^0 \ell \nu$ event, with subsequent π^0 decay, should be boosted and therefore the cosine of the angle between them, $\cos\theta_{\gamma\gamma}$, is biased towards 1. The distribution of this variable for the categories described above is shown in (b) of Figure 5.10. While signal peaks very clearly at 1, the distribution is almost uniform for beam background. We optimise the selection on $\cos\theta_{\gamma\gamma}$ in each q^2 bin by maximising the FOM given in Equation 5.6. Here we optimise the FOM for a combined electron and muon mode: $B^\pm \to \pi^0 \ell \nu$.

The $\cos\theta_{\gamma\gamma}$ distribution is shown in (a) of Figure 5.11 for the highest q^2 bin, along with the selection given by the FOM optimisation. The FOM as a function of $\cos\theta_{\gamma\gamma}$ in this bin is shown in (b) of Figure 5.11. The value of $\cos\theta_{\gamma\gamma}$ corresponding to the maximum in the FOM is chosen as the selection limit. This results in the following selections for the five q^2 bins: 0.826 ($q1$), 0.855 ($q2$), 0.826 ($q3$), 0.739 ($q4$) and 0.768 ($q5$).

The reconstructed π^0s in the low q^2 bins are high momentum pions and the daughter photons are highly boosted, even in background. The separation power between signal and background is therefore limited in these bins. The $q5$ bin, on the other hand, contains low momentum π^0s, and as beam



(a) (b)

Figure 5.11: (a) Distribution of the cosine of the angle between the two photons that the π^0 is reconstructed from, $\cos\theta_{\gamma\gamma}$, for the $q5$ q^2 bin in MC for the $B^\pm \to \pi^0 e\nu$ mode. The selection marked in this plot is obtained from the FOM optimisation that is respresented by (b).

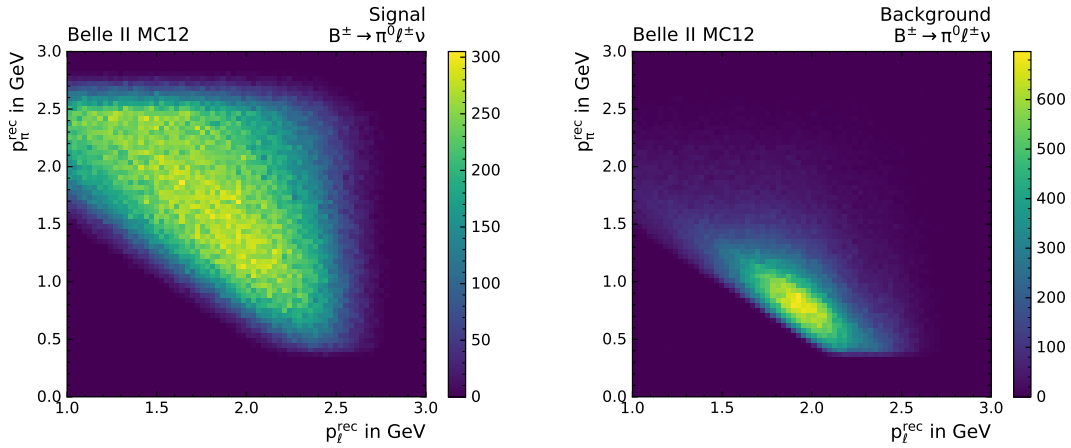
background usually consists of low momentum particles, most of the beam background contributes to the $q5$ bin. At the same time, the reconstructed π^0s in this bin, are barely boosted. Due to this, the selection is most effective in the $q5$ bin. The selection rejects on average 30% of the background, while retaining 96% of all signal events. The fraction of misreconstructed π^0s is reduced to one sixth of all reconstructed neutral pions.

5.5.3 Combined Momentum

After the selection on the angle between the photons, one third of the remaining background comes from $B\overline{B}$ events, of which approximately half are $B \to D^*\ell\nu$ events. As discussed in Section 5.1, a structure is present in the E_ℓ versus E_π distribution in signal. Looking at the 2D distributions of p_ℓ versus p_π for signal and $B\overline{B}$ background after all the above selections have been applied, in Figure 5.12, a diagonal structure can be observed in the $B\overline{B}$ background distributions.

We therefore apply a diagonal selection on $p_\ell + p_\pi$, optimising the selections again in each q^2 bin for combined electron and muon modes. The $p_\ell + p_\pi$ distribution for the $q4$ bin of the $B^\pm \to \pi^0 e\nu$ mode is shown in (a) of Figure 5.13. The selection in this bin is given by the maximum in the FOM as a function of $p_\ell + p_\pi$ displayed in (b). The selections on the 2D distributions of p_ℓ versus p_π in the $q4$ bin in signal and $B\overline{B}$ background are represented by the green line given in (c) and (d), respectively. Events to the right of the green line pass this selection.

The resulting selections for the q^2 bins given by the FOM optimisation are 0 ($q1$), 2.893 ($q2$), 2.893 ($q3$), 2.857 ($q4$) and 2.746 ($q5$) for the $B^0 \to \pi^\pm \ell\nu$ modes and 0 ($q1$), 2.893 ($q2$), 2.857 ($q3$), 2.857($q4$) and 2.709 ($q5$) for the $B^\pm \to \pi^0 \ell\nu$ modes. The selection is 0 in the first q^2 bin, as signal and background occupy the same phase space and there is no gain in the FOM for any selection. The selection on $p_\ell + p_\pi$, on average, rejects 50% of $B\overline{B}$ background and 25% of continuum background,



(a) (b)

Figure 5.12: 2D distributions of reconstructed p_ℓ versus p_π for (a) signal and (b) $B\overline{B}$ background from MC for the $B^{\pm} \to \pi^0 \ell \nu$ mode after all preceding selections.

leading to an overall background rejection of 33%. The signal efficiency is approximately 91%.

5.6 Best Candidate Selection

After the selection on $p_\ell + p_\pi$ the candidate multiplicity in the B^0 and $B^{\pm}$ modes is still 1.2 and 1.3, respectively. To ensure that there is only one candidate per event from this point on, we implement a best candidate selection based on particle ID variables. For the B^0 modes we choose the candidate with the highest sum of the lepton ID and pion ID, while for the $B^{\pm}$ modes we select the candidate with the highest lepton ID. The signal efficiencies of the best candidate selection are all on average greater than 93%.

5.7 Selection Efficiencies

The signal efficiencies for all the previously discussed selections averaged over the five q^2 bins for the $B^0 \to \pi^{\pm} e\nu$, $B^0 \to \pi^{\pm} \mu\nu$, $B^{\pm} \to \pi^0 e\nu$ and $B^{\pm} \to \pi^0 \mu\nu$ modes are given in Table 5.1. The average total signal efficiency due to these selections for each mode is also given. They lie between 29% and 45%. The FOM has increased by a factor of 3 for all modes. The highest purity mode is $B^0 \to \pi^{\pm} e\nu$ with a purity of 1.3%. The lowest purity is 0.4% in the $B^{\pm} \to \pi^0 \mu\nu$ mode. This is also reflected in the amount of background compared to signal present in these modes. While the number of background events is 75 times the amount of signal in the $B^0 \to \pi^{\pm} e\nu$ mode, in the $B^{\pm} \to \pi^0 \mu\nu$ mode this factor is 248, even though only 0.18% of the initial background in this mode is remaining.

The signal efficiency across the q^2 bins for the $B^{\pm} \to \pi^0 e\nu$ mode is shown for the lepton momentum selection and the selection on the photon angle in (a) and (b) of Figure 5.14, respectively. As explained in Section 5.2 and Section 5.5.2, the efficiency is low in the first q^2 bin for the lepton momentum selection and low in the highest q^2 bin for the selection on the angle between the photons. Overall, the

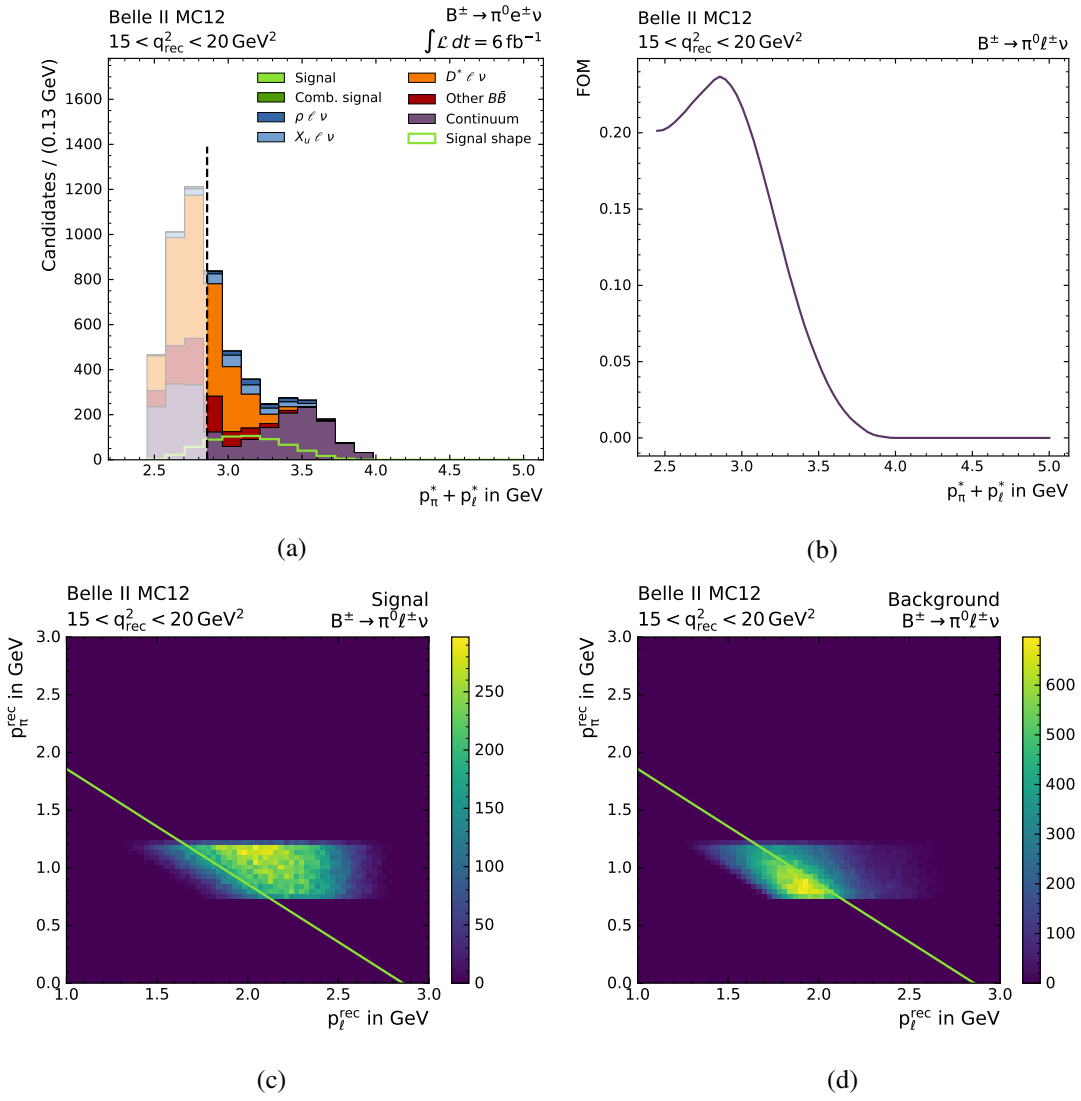
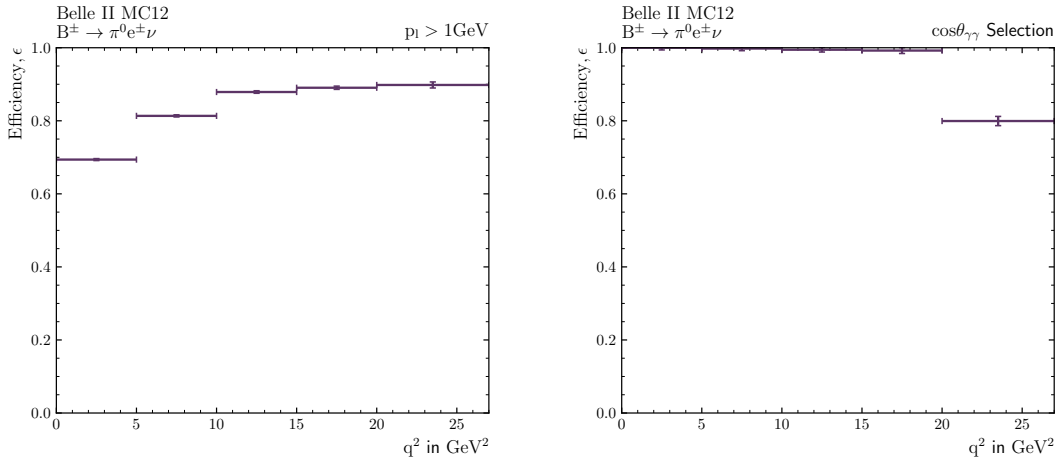


Figure 5.13: (a) Distribution of $p_\ell + p_\pi$ in MC in the $q4$ bin of the $B^\pm \to \pi^0 e\nu$ mode. The selection is found from the FOM optimisation for this bin represented by (b). This is the FOM as a function of $p_\ell + p_\pi$ in the $q4$ bin. The same selection for the $q4$ bin is represented by the green line in the 2D distributions of p_ℓ versus p_π for (c) signal and (d) $B\overline{B}$ background. Events to the right of the line pass the selection.

Selection	Mode			
	$B^0 \to \pi^\pm e\nu$	$B^0 \to \pi^\pm \mu\nu$	$B^\pm \to \pi^0 e\nu$	$B^\pm \to \pi^0 \mu\nu$
p^*_ℓ	0.830	0.881	0.835	0.887
Lepton ID	0.960	0.903	0.958	0.903
ΔE and M_{bc}	0.797	0.805	0.812	0.823
dr and dz	0.999	0.999	1.000	1.000
nCleanedTracks	0.889	0.889	0.698	0.698
θ_{miss}	0.944	0.945	0.945	0.947
$\cos\theta_{BY}$	0.849	0.943	0.828	0.924
$\cos\theta_{\gamma\gamma}$	1.000	1.000	0.957	0.957
$p_\ell + p_\pi$	0.907	0.914	0.927	0.931
Best Candidate	0.937	0.938	0.950	0.953
Total	0.384	0.435	0.299	0.342

Table 5.1: Efficiencies of the selections discussed in this chapter found from signal MC for all modes and averaged over the q^2 bins. The total efficiency is also given for each mode.



(a) (b)

Figure 5.14: Efficiencies across the five q^2 bins of the $B^\pm \to \pi^0 e\nu$ mode for (a) the lepton momentum selection and (b) the selection on the cosine of the angle between the two photons.

structure remains flat, since selections affecting different q^2 bins balance the effects. The plots for the other modes and selections are given in the appendix. Detailed signal efficiencies for all five q^2 bins and all four modes are also given in the appendix.

CHAPTER 6

Continuum and $B\overline{B}$ Background Suppression

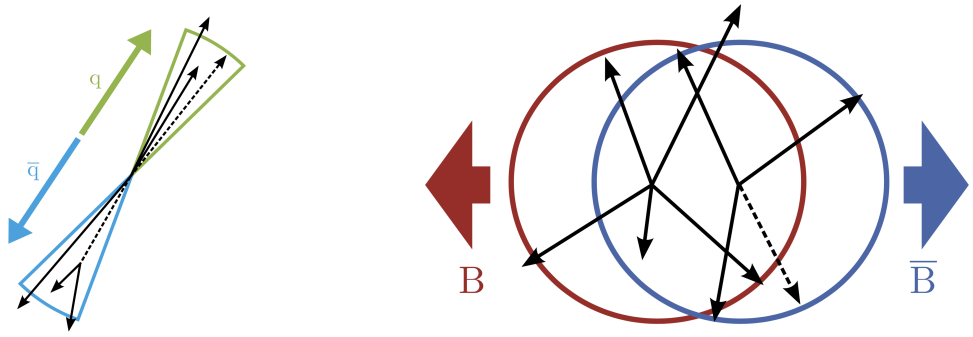
The ratio between $B\overline{B}$ and continuum background at this point in the selections is approximately 1:4. These two types of background have different characteristics, which, along with variables that exploit the characteristics to distinguish between signal and backgroud, will be explained below. We will separately train BDTs for distinguishing signal from these backgrounds for different q^2 bins.

In the following, the samples and settings used in the BDTs will be explained and the BDT performances will then be discussed separately for continuum and $B\overline{B}$ suppression. Finally, a selection on the output classifiers is performed. The selections are determined by optimising the figure of merit across the 2D distributions of the continuum suppression and $B\overline{B}$ suppression classifiers.

6.1 Background Characteristics

In the centre of mass frame of the colliding electrons and positrons, the quarks in a quark pair are produced back to back and they each carry half the centre of mass energy. Colour confinement causes the quarks to combine with other quarks and antiquarks to produce hadrons. In the lab frame these continuum events have a jet-like shape due to boosting, as depicted in (a) of Figure 6.1. On the other hand, in the case of $B\overline{B}$ production, the B mesons are at rest upon production in the centre of mass frame. This occurs since the $\Upsilon(4S)$ energy is just large enough to produce the two B mesons and results in isotropic angular distributions of the B meson decay products, as depicted in (b). To distinguish continuum from $B\overline{B}$ events, event shape variables are investigated and they are explained in the next section.

Distinguishing $B\overline{B}$ background events from signal events is a harder task, but there are some characteristics that allow separation. For one, the kinematics of the particles will be different depending on the kind of B meson decay they originated from. This means the daughter momenta and angular distributions of $B \to X\ell\nu$ decays, where $X \neq \pi$, will be different from those of signal decays. Next to that, in a $B\overline{B}$ background event at least one particle is reconstructed incorrectly. Thus variables that provide information about the reconstruction can be investigated, such as ROE variables. These will be explained in Section 6.3.



(a) (b)

Figure 6.1: Depiction of (a) the jet-like behaviour of continuum production and (b) the more spherical event shape of B$\overline{\text{B}}$ events. Adapted from [29].

6.2 Continuum Suppression Variables

The continuum suppression variables describe the event shape and will be described below.

Thrust and Thrust Angles

The thrust axis ($\vec{T}$) is defined as the axis for which the sum of the projected momenta of the tracks ($\vec{p}_i$) in the event is maximal:

$$\vec{T} = \max\left(\sum_i^N \vec{T} \cdot \vec{p}_i\right) \tag{6.1}$$

The cosine of the angle between the thrust axis and the axis corresponding to the sum of the momenta of all the particles in the event is called the cosine of the thrust angle, cosTBT, or also $\cos\theta_{\text{Thrust}}$. This is used to test the jet-likeness of an event. For continuum events the angle should be close to zero, since the two axes should coincide, and the distribution of the cosine of the angle should peak at 1. For B$\overline{\text{B}}$ events the distribution in the angle should be uniform, and therefore the distribution in the cosine of the angle should not have any peaking features. Another thrust angle of interest is the angle between the thrust axis and the beam axis, called cosTBz, or $\cos\theta_z$.

The thrust, T, is given by the sum of the projection of the momenta of all the tracks in the event onto the thrust axis, normalised by the sum of the momenta of the tracks [30]. It is therefore related to the thrust axis by

$$T = \frac{\sum_i^N |\vec{T} \cdot \vec{p}_i|}{\sum_i^N |\vec{p}_i|}. \tag{6.2}$$

For jet-like events the thrust should be close to 1, while for more spherical events the thrust takes values around 0.5, corresponding to an isotropic distribution.

Fox-Wolfram Moments

The Fox-Wolfram Moments (FWM) are event shape variables that describe the momentum and energy distributions of a system of N particles [31]. The ℓ-th order FWM is defined by

$$H_\ell = \sum_{i,j}^{N} |\vec{p}_i||\vec{p}_j| P_\ell(\cos\theta_{ij}), \tag{6.3}$$

where $|\vec{p_{i,j}}|$ is the magnitude of the momentum of the i-th or j-th particle, respectively. $P_\ell(\cos\theta_{ij})$ is the ℓ-th Legendre polynomial of the cosine of the angle between the i-th and j-th particle, $\cos\theta_{ij}$.

Different types of normalized Fox-Wolfram moments can be constructed, such as

$$R_\ell = H_\ell / H_0. \tag{6.4}$$

For jet-like events, such as continuum, R_ℓ is 0 when ℓ is odd and it is 1 when ℓ is even. Another type of normalized FWM is

$$h_\ell^k = \frac{\sum_{i,j}^{N} |\vec{p}_i||\vec{p}_j| P_\ell(\cos\theta_{ij})}{\sum_{i,j}^{N} |\vec{p}_i|.|\vec{p}_j|}, \tag{6.5}$$

where k defines whether the particles i and j are taken from the signal side or from the ROE [30]. If $k = oo$, both i and j are taken from the ROE. If $k = so$, particle i is taken from the signal side and particle j is taken from the ROE.

CLEO Cones

The CLEO cones were developed by the CLEO collaboration and they describe the momentum flow around the thrust axis. That is, how the momentum vectors of the particles in an event are distributed. 9 cones around the thrust axis are defined, whose opening angles increase in 10 degree steps. The magnitude of the momentum falling into each of these cones is measured [32]. Since the decay particle flight directions for $B\overline{B}$ events are not correlated, but for continuum they are, the momentum flow in these CLEO cones can distinguish continuum from $B\overline{B}$ events.

6.3 $B\overline{B}$ Suppression Variables

To separate $B\overline{B}$ background from signal, a number of kinematic variables can be investigated to check for misreconstruction. The variables can be split into signal-side and ROE variables.

Signal-Side Variables

As explained in Section 5.1, the phase space for the signal decay is very clearly defined by the identity of the decay products. Due to this, the distribution of the lepton and pion momentum will look different depending on whether or not they were coming from a signal decay. As before, the angle between the pion-lepton candidate and the B candidate, $\cos\theta_{BY}$, provides discriminating power.

The angle between the momentum vectors of the lepton and pion, $\cos\theta_{\pi\ell}$, is also distinctive for signal decays. For signal decays there is a correlation between the pion and lepton flight directions.

Especially at low q^2 the pion and lepton are often produced back to back in the centre of mass frame. This results in the distribution of $\cos\theta_{\pi\ell}$ peaking at -1 for signal. For misreconstructed B$\overline{\text{B}}$ background, on the other hand, the flight directions of the two reconstructed particles should be largely uncorrelated, thus resulting in a uniform distribution.

ROE Variables

For a correctly reconstructed event the ROE should only contain the decay products of the second B meson. This can be tested by constructing an energy difference and beam constrained mass for the ROE, ΔE^{ROE} and M_{bc}^{ROE}, similar to these variables on the signal side. If the second B meson is reconstructed correctly ΔE^{ROE} peaks at 0, and M_{bc}^{ROE} peaks at the B meson mass. The resolution in these variables depends on the application of the correct mass hypotheses in the ROE, and it is limited by the presence of neutrinos in the second B meson decay. Any neutrinos present in the decay of the second B meson are collectively assigned to the signal side and are included in ΔE and M_{bc} calculated on the signal side.

The neutrino reconstruction variables also allow to check for correctly reconstructed signal, such as the missing momentum, the missing momentum angle and the missing mass squared. A variable, $m^2_{\text{miss}}/2E_{\text{miss}}$, can be constructed from the missing mass squared and the missing energy. This removes the linear dependence of the missing mass squared resolution on the missing energy. Again this should be centred at 0 for signal decays.

6.4 BDT Samples and Settings

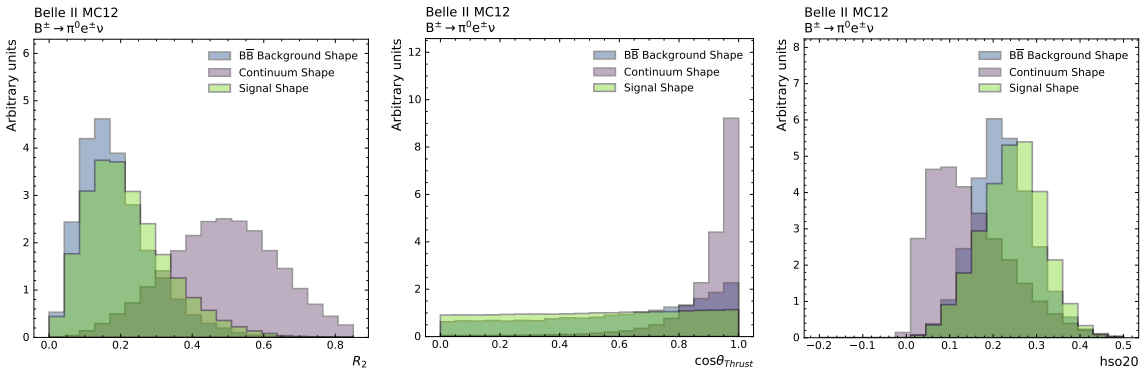
To suppress continuum background, samples of signal MC and continuum background MC are trained against each other. The B$\overline{\text{B}}$ background BDTs are trained on samples of signal MC and B$\overline{\text{B}}$ background MC. Due to limited statistics in some bins, especially in the low q^2 bins for B$\overline{\text{B}}$ background, and in the high q^2 bins for signal, we combine the two lowest q^2 bins into a $lowq$ bin and the two highest q^2 bins into a $highq$ bin. The intermediate bin is called $medq$. We therefore train 4 (modes) $\times$ 2 (background type) $\times$ 3 (q^2 bins) = 24 BDTs. The ratio between signal and background in each bin is set to 50:50 and a testing sample one ninth the size of the training sample is chosen for each bin. The type of BDT used is FastBDT. First, the BDTs are trained on a preliminary list of interesting variables using the default hyperparameter options, as explained in Section 2.4:

- Number of Trees: 150
- Depth of the Trees: 3
- Number of Cuts: 8
- Shrinkage: 0.2
- Sampling Rate: 0.5

The 10 highest ranking variables for continuum suppresion and the 8 highest ranking variables for B$\overline{\text{B}}$ suppression are selected in each q^2 bin. With these variables a grid search is performed to determine the optimal hyperparameters. The hyperparameters of the BDT with the highest area under the ROC curve (AUC) were chosen and they are given in the appendix for all 24 BDTs. Especially the number of trees affected the performance of the BDTs. For the continuum BDTs the best performance is achieved with 200 trees, while the B$\overline{\text{B}}$ BDTs perform best with 150 trees.

6.5 Continuum Suppression

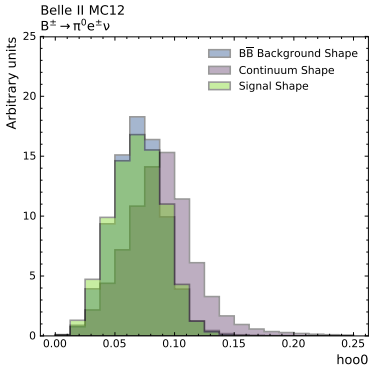
With the hyperparameter settings determined from the grid search the continuum suppression BDTs are trained again on the top 10 variables. The list of the top 10 variables varied between the q^2 bins, but they are all in the list: R_2, CLEO Cone 1, $\cos\theta_{\text{Thrust}}$, $\cos\theta_z$, Thrust, h_2^{so}, h_{12}^{so}, h_{20}^{so}, h_{22}^{so}, h_2^{oo}, h_0^{oo}. The exact list of the top 10 variables and the corresponding importance ranking for each BDT is given in the appendix.



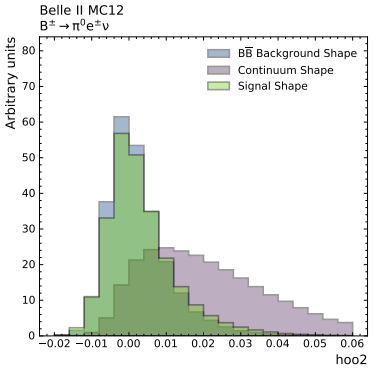
(a) R_2

(b) $\cos\theta_{\text{Thrust}}$

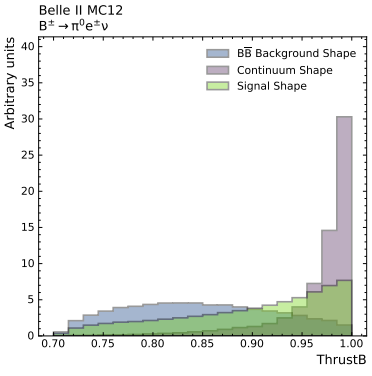
(c) h_{20}^{so}



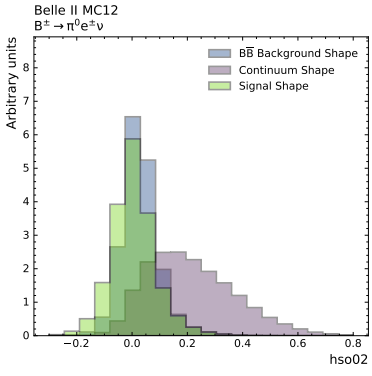
(d) h_0^{oo}



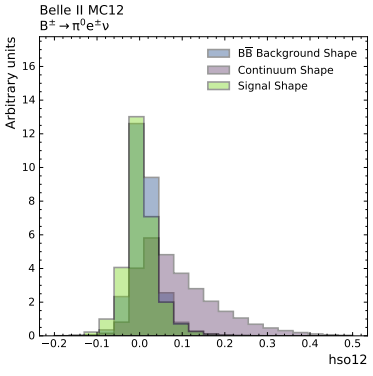
(e) h_2^{oo}



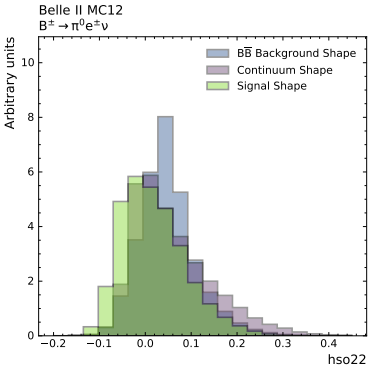
(f) Thrust



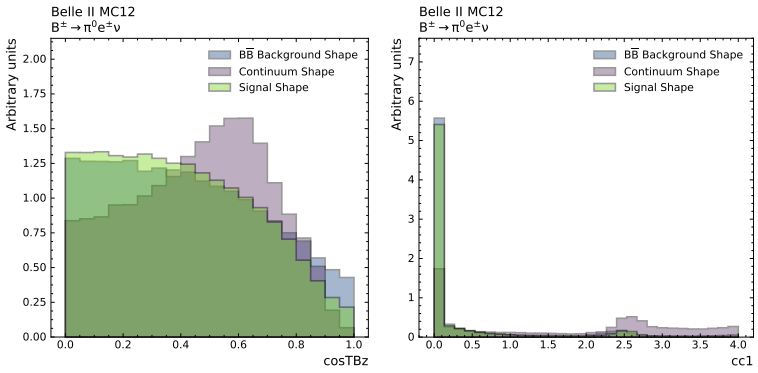
(g) h_2^{so}



(h) h_{12}^{so}



(i) h_{22}^{so}



(j) $\cos\theta_z$ (k) CLEO Cone 1

Figure 6.2: Normalised distributions of the continuum suppression BDT input variables for signal in green, continuum background in purple and B$\overline{\text{B}}$ background in blue for the $B^{\pm} \to \pi^0 e\nu$ mode.

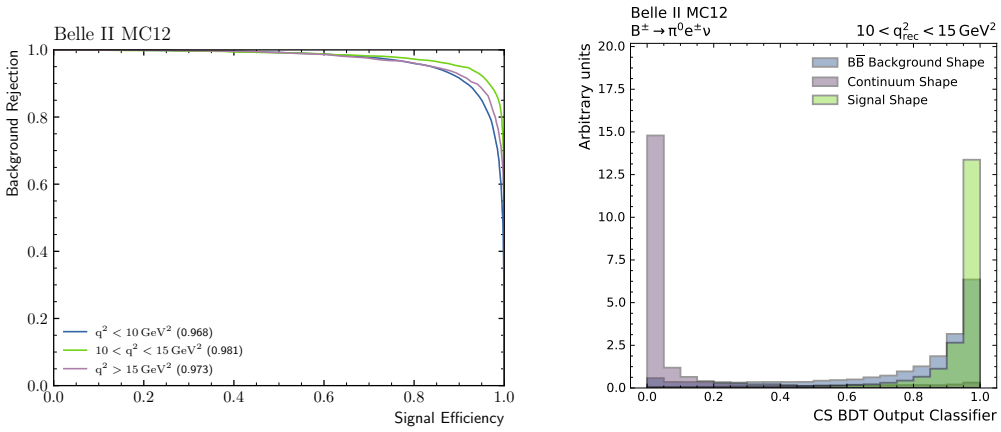
For most of the BDTs the normalized FWM R_2 is the most important variable, often followed by $\cos\theta_{\text{Thrust}}$ and h_{20}^{so}. The normalised distributions of all the variables for signal, B$\overline{\text{B}}$ and continuum background are shown in Figure 6.2. The correlation matrices for these variables show no exceptionally high correlations and they are given in the appendix.

The BDTs all have AUCs that are greater than 0.96, resulting in good separation power between continuum and signal. As an example, the ROC curves for the $lowq$, $medq$ and $highq$ q^2 bins in the $B^{\pm} \to \pi^0 e\nu$ mode are shown in (a) of Figure 6.3. The p-values for signal and background overtraining tests for all the BDTs are given in the appendix. There are some cases where the p-value for signal or background is greater than 0.9 or smaller than 0.1, but none lie outside the thresholds for overtraining or undertraining set in Section 2.4. The normalised output classifier for signal, continuum and B$\overline{\text{B}}$ background for the $q3$ q^2 is shown in (b) of Figure 6.3. Signal clearly peaks at 1, while continuum background peaks at 0.

6.6 B$\overline{\text{B}}$ Suppression

Similar to the continuum suppression BDTs in the previous section, the BDTs for B$\overline{\text{B}}$ background suppression are trained using the corresponding hyperparameters found in the grid search and the top 8 variables. As before, the top 8 variables are not the same in each q^2 bin, and they are given in the appendix, along with the importance ranking. They are in the list of the following variables: p_ℓ, p_π, $\cos\theta_{\pi\ell}$, $m^2_{\text{miss}}/2E_{\text{miss}}$, $\cos\theta_{BY}$, ΔE^{ROE}, M_{bc}^{ROE}, $p^{\text{lab}}_{\text{miss}}$ and $\theta^{\text{lab}}_{\text{miss}}$. The most important variables for most of the BDTs are ΔE^{ROE}, M_{bc}^{ROE}, $\cos\theta_{\pi\ell}$ and $p^{\text{lab}}_{\text{miss}}$. The normalised distributions of the BDT input variables for signal, continuum and B$\overline{\text{B}}$ background are shown in Figure 6.4.

The ROC curves for the $lowq$, $medq$ and $highq$ bins in the $B^{\pm} \to \pi^0 e\nu$ mode are shown in (a) of Figure 6.5. The AUCs for the B$\overline{\text{B}}$ suppression BDTs are all greater than 0.87, and they are on average lower than those of the continuum suppression BDTs. The BDTs therefore have lower separation power between background and signal, as compared to the continuum suppression BDTs. The p-values of the overtraining tests for signal and background are in the acceptable region, as are the correlations for the BDT input variables. The p-values and correlation matrices are given in the appendix.



(a) (b)

Figure 6.3: (a) ROC curves for the *lowq*, *medq* and *highq* bins in the $B^\pm \to \pi^0 e\nu$ mode. The AUC for each of the curves is given in the legend. (b) The normalised distribution of the continuum suppression BDT output classifier for signal is shown in green, continuum background in purple and $B\overline{B}$ background in blue for the $q3$ bin of the $B^\pm \to \pi^0 e\nu$ mode.

Mode	Selection on CS Classifier				
	$q1$	$q2$	$q3$	$q4$	$q5$
$B^0 \to \pi^\pm e\nu$	0.862	0.931	0.862	0.759	0.793
$B^\pm \to \pi^0 e\nu$	0.895	0.842	0.737	0.790	0.842
$B^0 \to \pi^\pm \mu\nu$	0.931	0.862	0.759	0.931	0.931
$B^\pm \to \pi^0 \mu\nu$	0.895	0.895	0.842	0.842	0.947

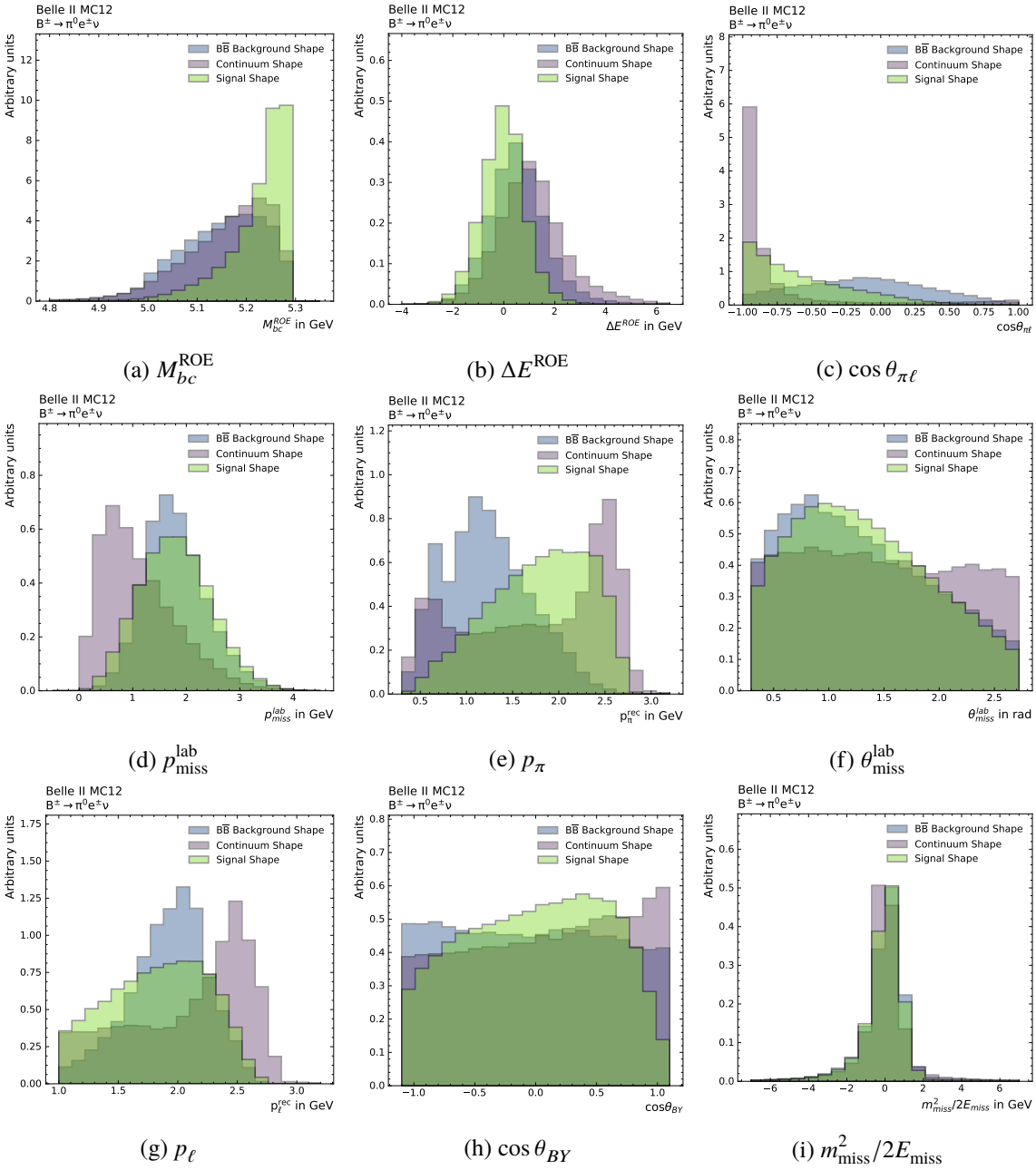
Table 6.1: Selections on the continuum suppression BDT output classifiers for the five q^2 bins in the four modes.

6.7 FOM Evaluation

As a last selection step we select on the BDT output classifiers to remove continuum and $B\overline{B}$ background. We perform a 2D optimisation of the FOM to determine the optimal selections. Plots of the FOM as a function of the continuum suppression and $B\overline{B}$ BDT outputs are shown for the $q1$ and $q3$ bins of the $B^0 \to \pi^\pm e\nu$ mode in Figure 6.6. The maximum points are marked and labelled. In the lowest q^2 bin continuum suppression has a higher weight, while in the higher q^2 bins $B\overline{B}$ background suppression is more important.

We realised, however, that selecting on the $B\overline{B}$ BDT output classifier resulted in the shapes of M_{bc} and ΔE in signal and background looking very similar. The selection introduced correlations of up to 99% into the M_{bc} and ΔE distributions. This makes a fit to these distributions impossible. We therefore decided to only perform a selection on the continuum suppression BDT output classifiers. The selections are given in Table 6.1 and the distributions of the output classifiers with the corresponding selections in the $q1$ and $q3$ bins are shown in Figure 6.7.

These selections have average signal efficiencies that range from 71% to 77%, with background

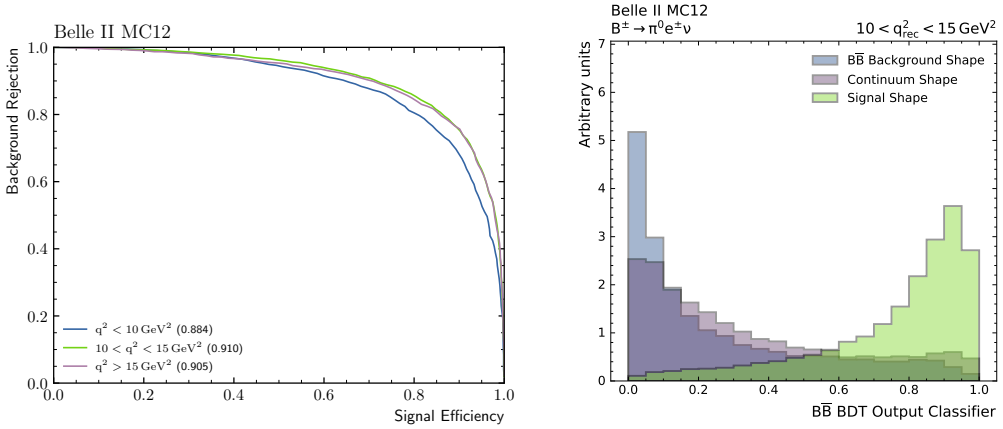


(a) M_{bc}^{ROE} (b) ΔE^{ROE} (c) $\cos\theta_{\pi\ell}$

(d) $p_{\text{miss}}^{\text{lab}}$ (e) p_π (f) $\theta_{\text{miss}}^{\text{lab}}$

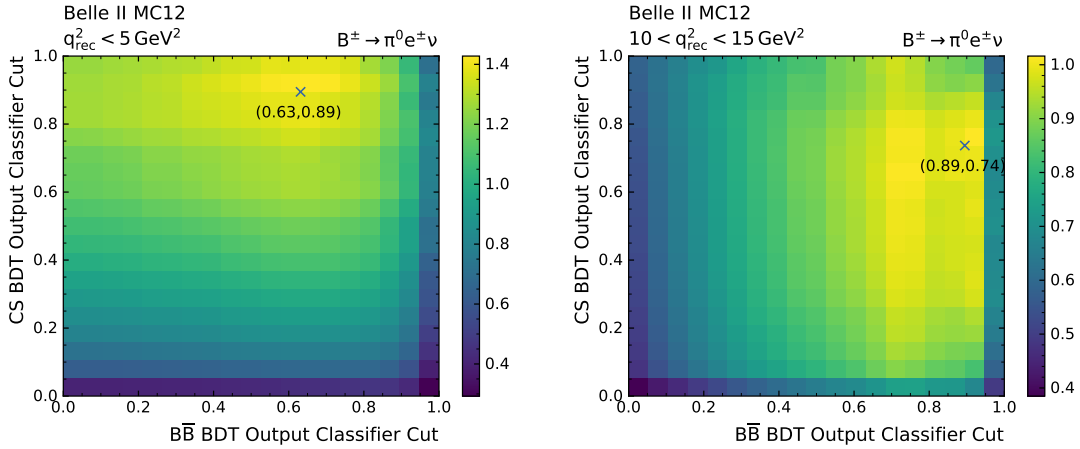
(g) p_ℓ (h) $\cos\theta_{BY}$ (i) $m_{\text{miss}}^2/2E_{\text{miss}}$

Figure 6.4: Normalised distributions of the B$\overline{\text{B}}$ BDT input variables for signal in green, continuum background in purple and B$\overline{\text{B}}$ background in blue for the $B^\pm \to \pi^0 e\nu$ mode.



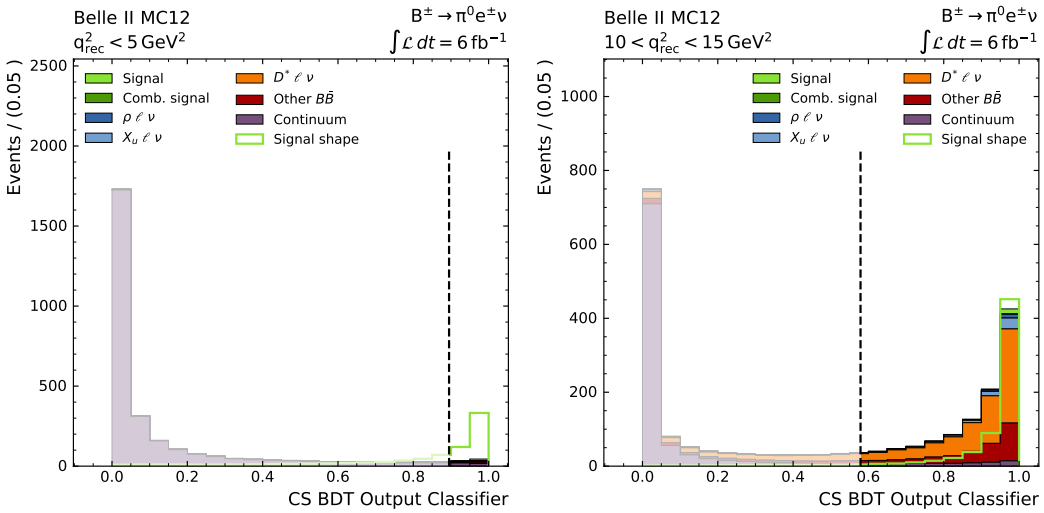
(a) (b)

Figure 6.5: (a) ROC curves for the *lowq*, *medq* and *highq* q^2 bins in the $B^\pm \to \pi^0 e\nu$ mode. The AUC for each of the curves is given in the legend. (b) The normalised distribution of the $B\overline{B}$ suppression BDT output classifier for signal is shown in green, continuum background in purple and $B\overline{B}$ background in blue for the $q3$ bin of the $B^\pm \to \pi^0 e\nu$ mode.



(a) (b)

Figure 6.6: 2D histograms of the FOM as a function of the continuum suppression and $B\overline{B}$ BDT output classifiers for the $B^\pm \to \pi^0 e\nu$ mode in (a) the $q1$ bin and (b) the $q3$ bin. The optimal selections are marked and labelled.



(a) (b)

Figure 6.7: Distribution of the continuum suppression output classifier in the (a) $q1$ and (b) $q3$ bins of the $B^\pm \to \pi^0 e\nu$ mode. The selections are shown.

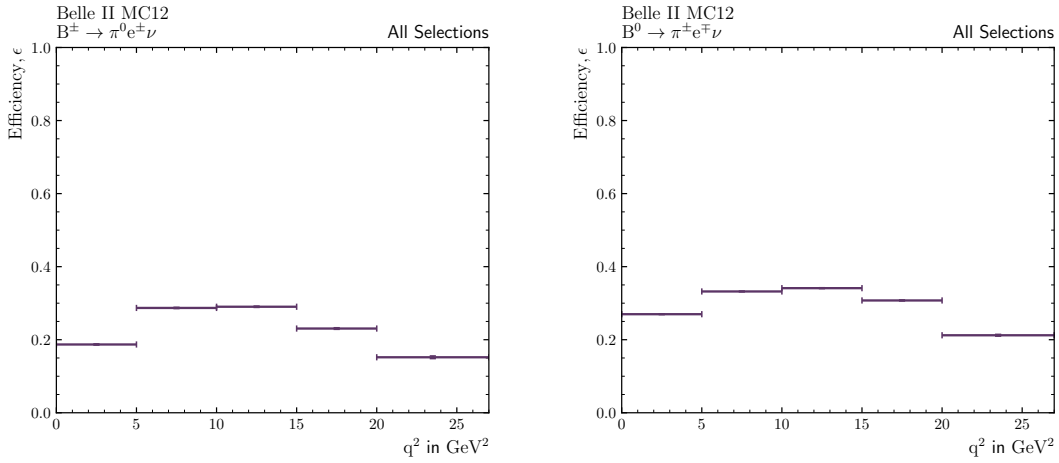
rejections of approximately 78% in the electron modes and approximately 90% in the muon modes. As expected, the selection on the continuum suppression BDT output classifier performs well at rejecting continuum background and 98% of the continuum background in each mode is rejected. The selection, in addition, removes approximately 50% of the B$\overline{\text{B}}$ background in each mode. Due to this, the purity increases by factors 4 and 7 in the electron and muon modes, respectively. Now, the dominant background events in all modes are the $B \to D^*\ell\nu$ events.

6.8 Total Selection Efficiencies

The total selection efficiencies for the five q^2 bins, after all selections discussed in this thesis have been performed, are given in Table 6.2. The total eficiencies are also shown in Figure 6.8 for the $B^0 \to \pi^\pm e\nu$ and $B^\pm \to \pi^0 e\nu$ modes. The efficiencies are lowest in the $q1$ and $q5$ bin, but across the q^2 bins they range from 20% to 40% and no sharp features are visible.

Mode	**Total Efficiency**				
	***q*1**	***q*2**	***q*3**	***q*4**	***q*5**
$B^0 \to \pi^\pm e\nu$	0.270	0.332	0.342	0.307	0.212
$B^\pm \to \pi^0 e\nu$	0.187	0.287	0.289	0.230	0.151
$B^0 \to \pi^\pm \mu\nu$	0.248	0.428	0.415	0.279	0.197
$B^\pm \to \pi^0 \mu\nu$	0.219	0.308	0.301	0.288	0.145

Table 6.2: Total selection effficiencies after all selections have been performed for the five q^2 bins in the four modes.



(a) (b)

Figure 6.8: Total efficiencies after all selections in this thesis have been performed for the five q^2 bins for the $B^{\pm} \to \pi^0 e\nu$ and $B^0 \to \pi^{\pm} e\nu$ modes.

The average FOM for each mode has improved by factors of 4 to 5, as compared to before the preselections. The FOM ranges from 1.4 to 3.3. The highest purity mode is still $B^0 \to \pi^{\pm} e\nu$ (4.5%), which is a 30-fold improvement relative to its value before the preselections. The $B^0 \to \pi^{\pm} \mu\nu$ mode remains the lowest purity mode with a purity of 2.5%. However, this is already a factor 45 improvement compared to what it was before the preselections. At this point, we are left with background in size about 20 to 40 times the size of signal in each mode.

CHAPTER 7

Data-MC Agreement

The available data sample consists of 5.594 fb^{-1} of early Phase 3 data, corresponding to Experiment 7 and Experiment 8. The following sections will discuss the agreement between data and MC for the kinematic variables, selection variables, the BDT input variables and output classifiers, and ΔE and M_{bc}. First, scaling factors will be introduced in order to consistently compare the shapes of these distributions.

The distributions will be shown sequentially, i.e. after the previously discussed selection has been performed. Since we do not include any low-multiplicity MC, however, all the selections up to and including the selection on the number of "cleaned" tracks have already been performed for all the distributions shown in this chapter.

7.1 Kinematic Variables and Scaling

As a first step the kinematic variables of the lepton and pion are investigated. The data-MC agreement for the momentum, the cosine of the θ angle, and the ϕ angle of the lepton and pion, and the respective pulls, in the $B^{\pm} \to \pi^0 e\nu$ mode are shown in Figure 7.1. Looking at the data-MC agreement of the lepton momentum distribution, the MC systematically underestimates data in the high momentum tail, which mainly contains continuum events. If a universal scaling factor is introduced, in order to compare the shapes between data and MC distributions from this point on, the $B\overline{B}$ MC will systematically overestimate data.

Instead, we determine separate scaling factors for continuum and $B\overline{B}$ MC. The variable R_2 has good separation power between these two components and the scaling factors are investigated from the data-MC agreement in that variable, given for the $B^{\pm} \to \pi^0 e\nu$ and $B^{\pm} \to \pi^0 \mu\nu$ modes in (a) and (b) of Figure 7.2. In the muon mode, the continuum MC largely overestimates the data, while in the electron mode it underestimates data. We expect that the noticeable difference between data and MC in the muon mode is due to problems with the muon ID.

An off-resonance data sample of 0.827 fb^{-1} is available. Off-resonance data is taken at a centre of mass energy that is just below the mass of the $\Upsilon(4S)$ resonance. Datsa taken at this energy contains continuum events, but no $B\overline{B}$ events. Thus, the size and shape of the off-resonance data can be compared to the continuum MC. The off-resonance data-MC agreement of R_2 is shown in Figure 7.3, and the same situation as described above is observed. We therefore define a continuum scaling factor, given by the ratio between the number of candidates in the continuum MC sample and the

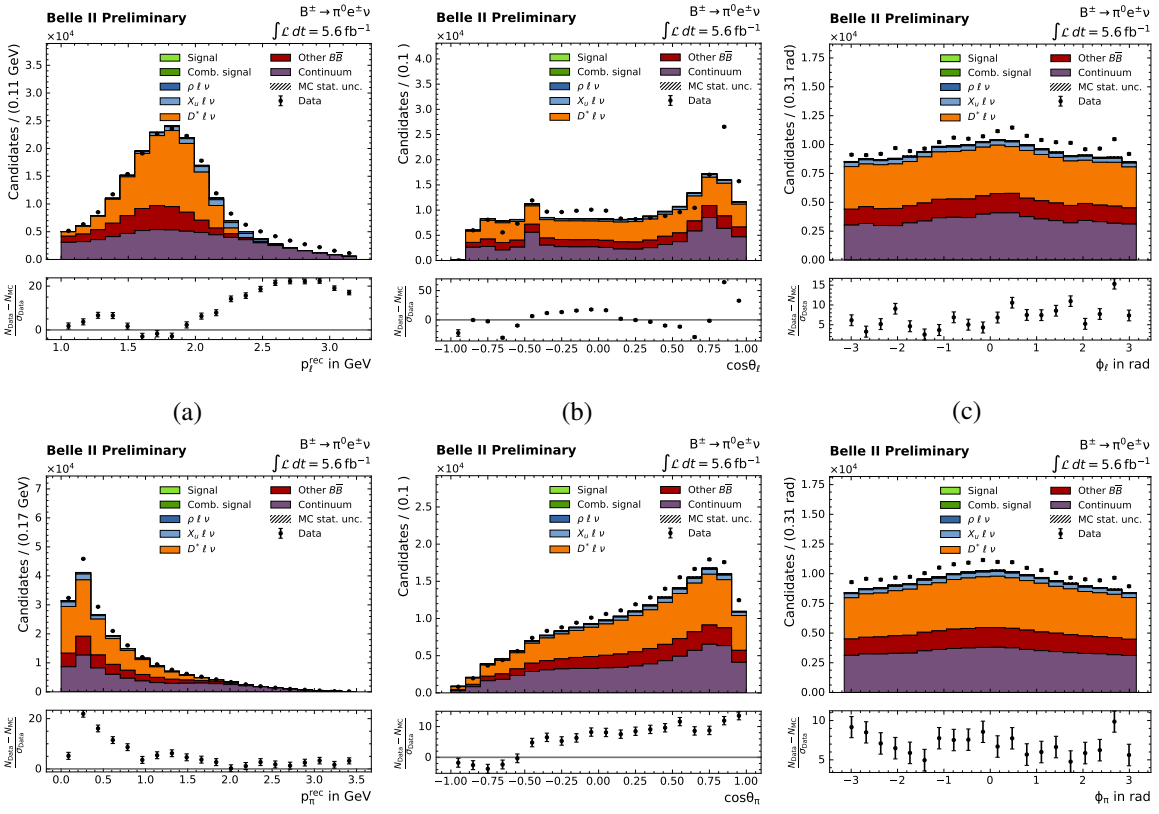
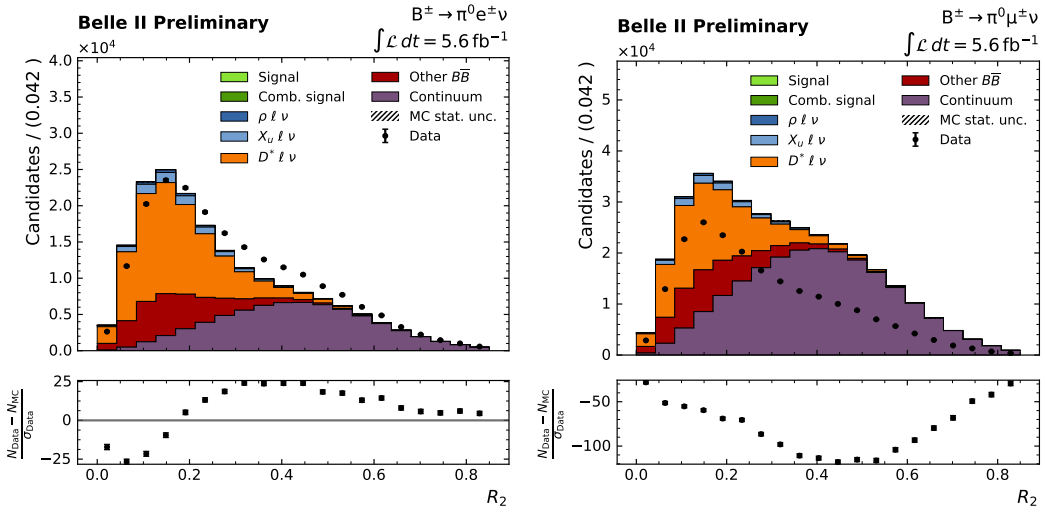


Figure 7.1: Data-MC agreement and pulls in the $B^{\pm} \to \pi^0 e\nu$ mode for the (a) lepton momentum, (b) cosine of the lepton θ angle, (c) lepton ϕ angle, (d) pion momentum, (e) cosine of the pion θ angle and (f) pion ϕ angle.

Scaling Factor	**Mode**			
	$B^0 \to \pi^{\pm} e\nu$	$B^0 \to \pi^{\pm} \mu\nu$	$B^{\pm} \to \pi^0 e\nu$	$B^{\pm} \to \pi^0 \mu\nu$
Continuum	1.412	0.385	1.440	0.371
BB	0.895	0.905	0.918	0.943

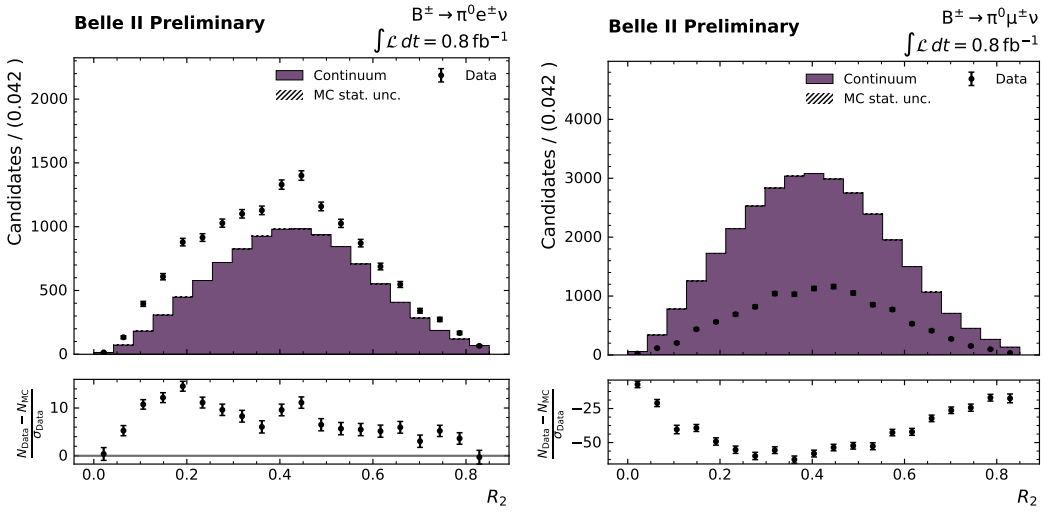
Table 7.1: Scaling factors for the continuum and $B\overline{B}$ MC as determined from the off-resonance data.

number of candidates in the off-resonance data sample. Next, we define a $B\overline{B}$ scaling factor as the the number of candidates in the $B\overline{B}$ MC divided by the difference in the number of candidates in the on-resonance and off-resonance data samples. The scaling factors for the four modes are given in Table 7.1. The distributions for R_2, now with these scaling factors applied, are shown in Figure 7.4. The pulls decrease by factors of approximately 2 and 4 in the electron and muon modes, respectively. We notice a rightward shift of data compared to MC, which we attribute to a rightward shift in data from $B\overline{B}$ events, possibly due to beam background. The scaling factors will be used for all of the plots in this chapter. Now shape differences that exist due to reasons beyond a mismatch in the ratio between $B\overline{B}$ and continuum MC are visible.



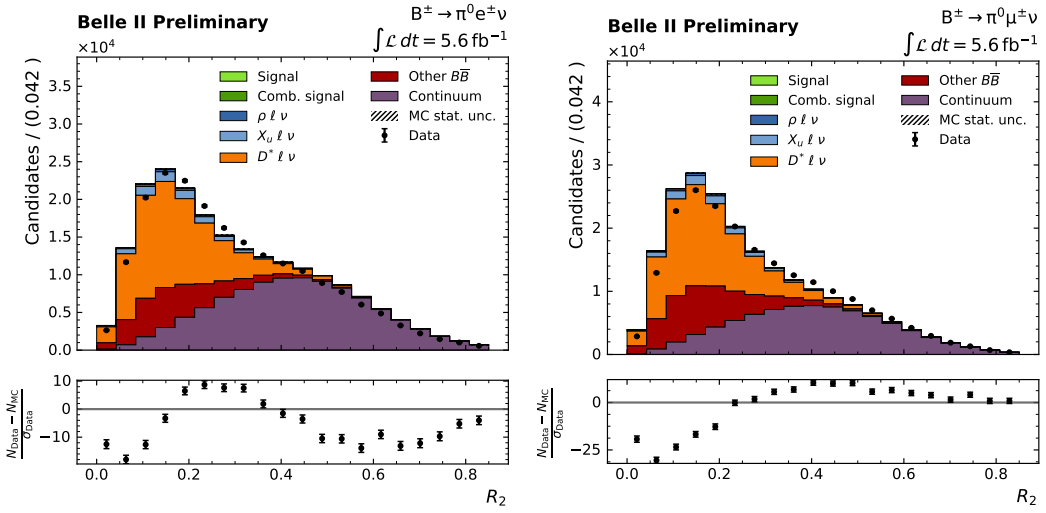
(a) (b)

Figure 7.2: Data-MC agreement and pulls for R_2 in the (a) $B^{\pm} \to \pi^0 e\nu$ and (b) $B^{\pm} \to \pi^0 \mu\nu$ without the application of a scaling factor.



(a) (b)

Figure 7.3: Agreement between off-resonance data and continuum MC in R_2 in the (a) $B^{\pm} \to \pi^0 e\nu$ and (b) $B^{\pm} \to \pi^0 \mu\nu$ modes.



(a) (b)

Figure 7.4: Data-MC agreement and pulls of R_2 in the (a) $B^{\pm} \to \pi^0 e\nu$ and (b) $B^{\pm} \to \pi^0 \mu\nu$ modes after the scaling factors have been applied to MC.

7.2 Selection Variables

Next, the data-MC agreement of the selection variables is investigated. The agreement for the lepton momentum, electron ID, muon ID, ΔE, M_{bc} and nCleanedTracks is shown in Figure 7.5. These are the distributions after all selections up to and including the selection on the number of "cleaned" tracks has been performed. The shape agreements seem reasonable.

The agreements for the lepton dr and dz are shown in (a) and (b) of Figure 7.6. The distributions for the pion dr and dz are similar. Data and MC do not agree in these variables and there is a clear shift in the peak, as well as a broadening. This must be due to a mismodelling of the interaction point position in the MC, as compared to where it really is in the detector. This will affect variables relying on tracking, such as momentum, but it is difficult to quantify the impact. A distribution where the effect can clearly be seen is shown in (c). This is the χ^2 distribution of a vertex fit of the pion and lepton to the interaction point performed with TreeFitter [33]. We initially required the fit to converge as a preselection step and, in addition, used this variable in the $B\overline{B}$ BDT training. Due to the disagreement between data and MC, this requirement was imposed no longer and the variable was excluded from the BDT training.

The data-MC agreement in the $B^{\pm} \to \pi^0 e\nu$ mode for θ_{miss} and $\cos\theta_{BY}$ is given in (a) and (b) of Figure 7.7. The data-MC agreement for one q^2 bin for $\cos\theta_{\gamma\gamma}$ and $p_\ell + p_\pi$ is shown in (c) and (d). Each of these distributions is shown after the selection on the previous variable has been performed. The general shapes agree for $\cos\theta_{BY}$, $\cos\theta_{\gamma\gamma}$ and $p_\ell + p_\pi$. We noticed that for the distribution of θ_{miss} the agreement is worse at higher angles.

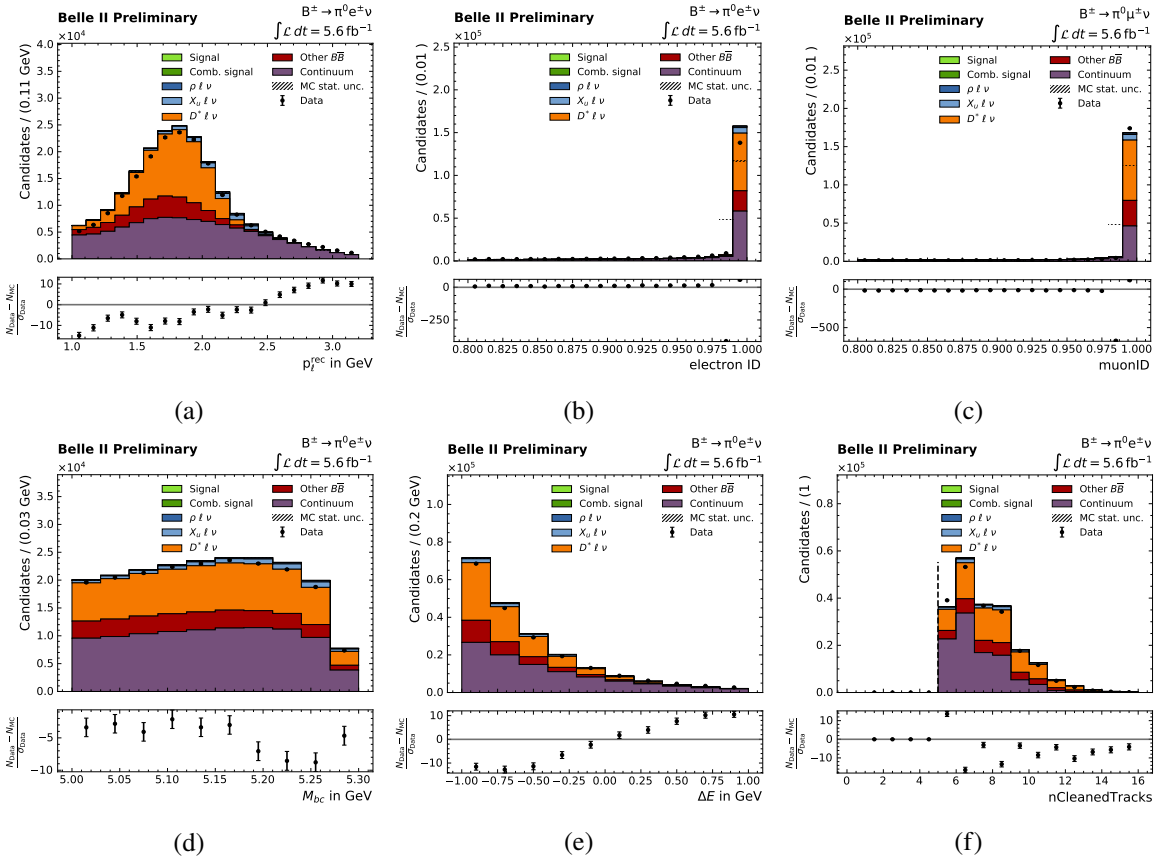


Figure 7.5: Data-MC agreement and pulls for (a) p_ℓ, (b) electron ID, (c) muon ID, (d) ΔE, (e) M_{bc} and (f) nCleanedTracks in the $B^\pm$ modes.

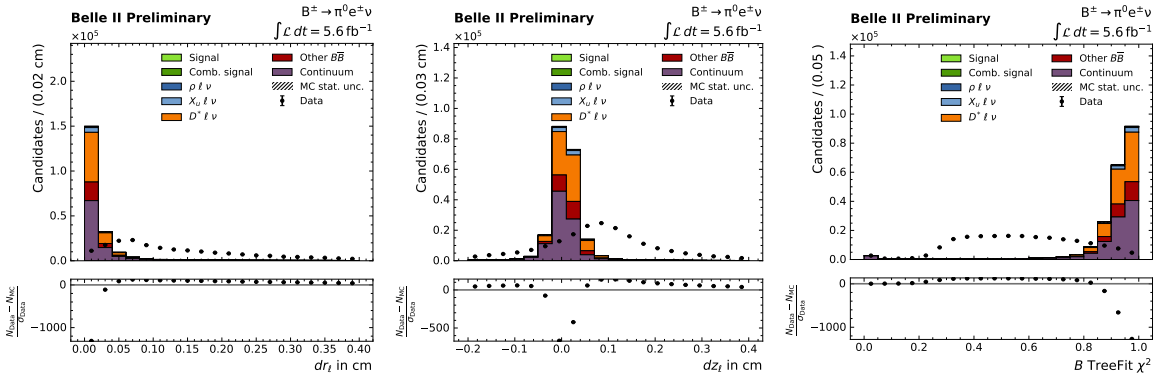


Figure 7.6: Data-MC agreement and pulls for the lepton (a) dr and (b) dz in the $B^\pm \to \pi^0 e\nu$ mode. (c) shows the data-MC agreement for the χ^2 probabilty of a vertex fit of the lepton and pion to the interaction point.

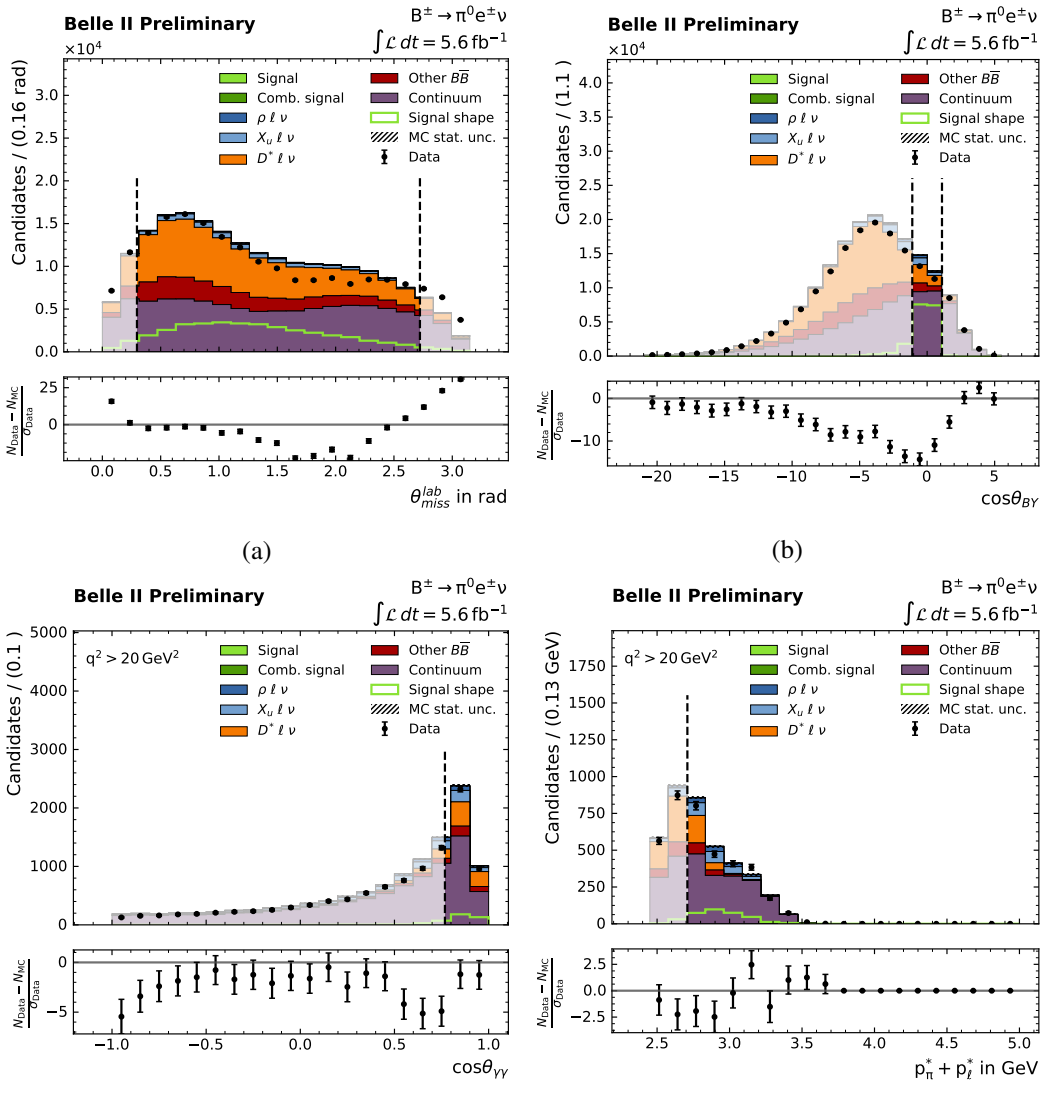
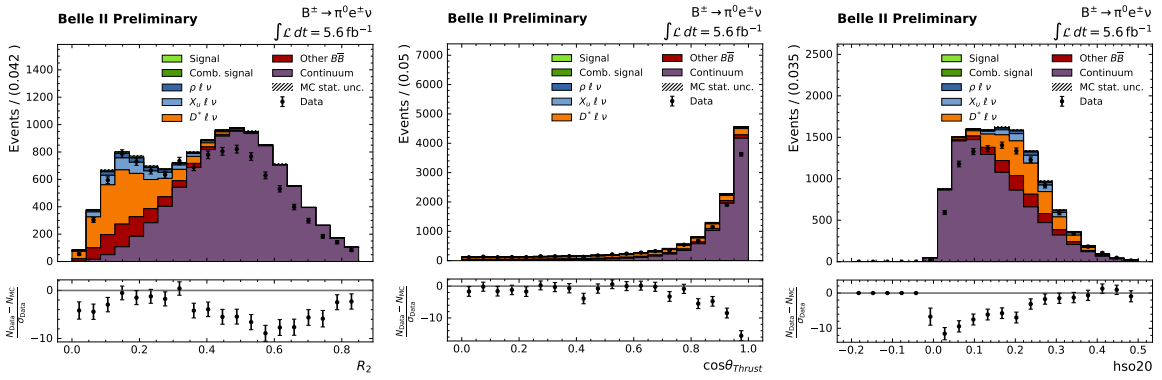


Figure 7.7: Data-MC agreement and pulls in the $B^{\pm} \to \pi^0 e\nu$ mode for (a) θ_{miss} and (b) $\cos\theta_{BY}$. The data-MC agreement in the $q5$ bin is shown for (c) $\cos\theta_{\gamma\gamma}$ and (d) $p_\ell + p_\pi$.

7.3 BDT Input Variables and Output Classifiers

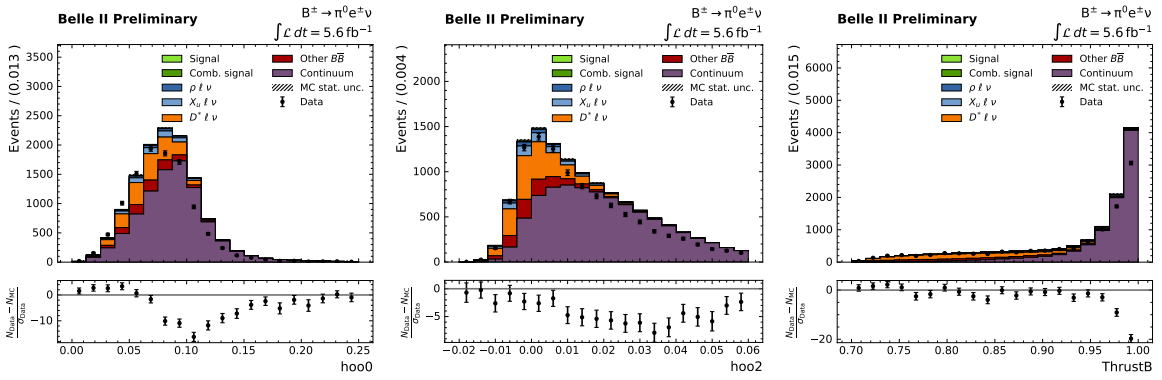
The data-MC agreement for the input variables of the continuum suppression BDTs before and after the selection on the output classifier in the $B^{\pm} \to \pi^0 e\nu$ mode are shown in Figure 7.8 and Figure 7.10, respectively. Before the BDT output classifier selection, continuum MC seems to overestimate data. However, the agreement between data and MC after the selection on the BDT output classifier, when there is little continuum background left, is good. The data-MC agreement of the BDT output classifier is shown for the $q1$ and $q3$ bins in the $B^{\pm} \to \pi^0 e\nu$ mode in Figure 7.9. Data and MC agree well at classifier values close to 1, where most of the signal lies, but the agreement is worse at values around 0.



(a) R_2

(b) $\cos\theta_{\mathrm{Thrust}}$

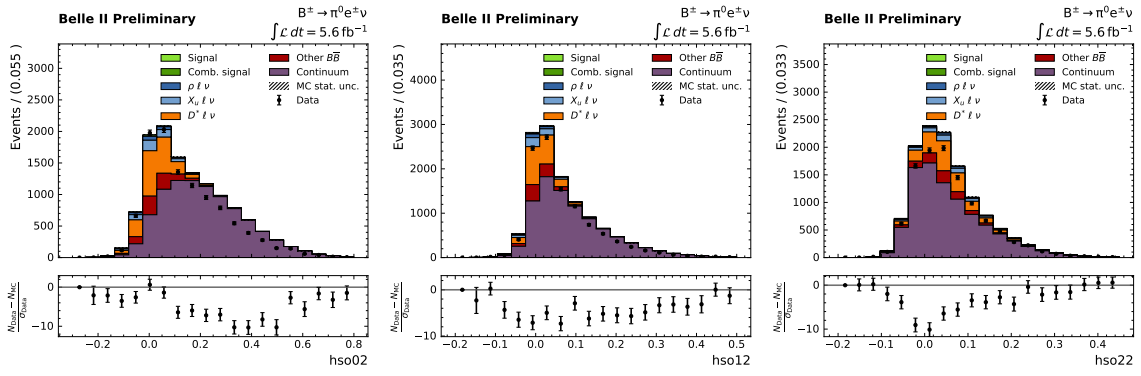
(c) h_{20}^{so}



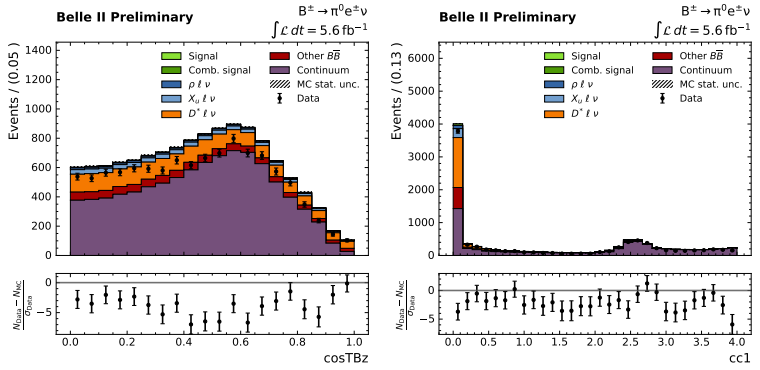
(d) h_0^{oo}

(e) h_2^{oo}

(f) Thrust

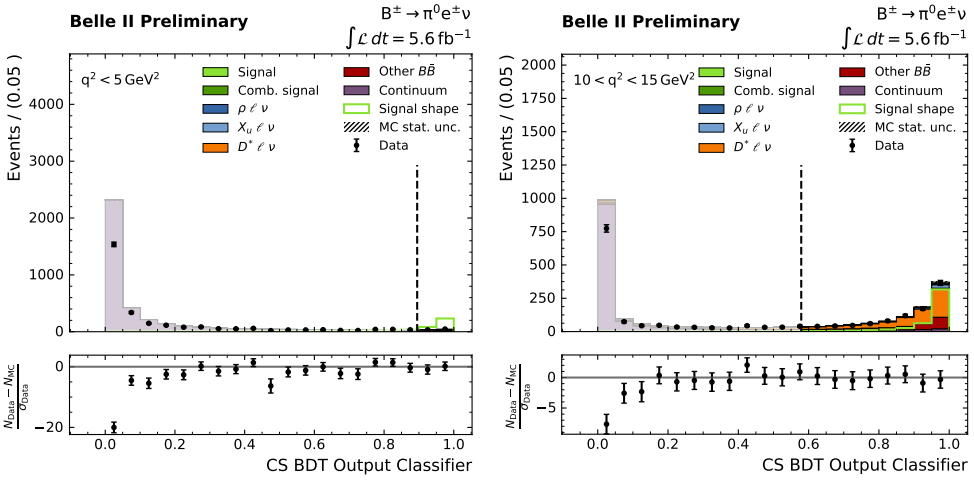


(g) h_2^{so} (h) h_{12}^{so} (i) h_{22}^{so}



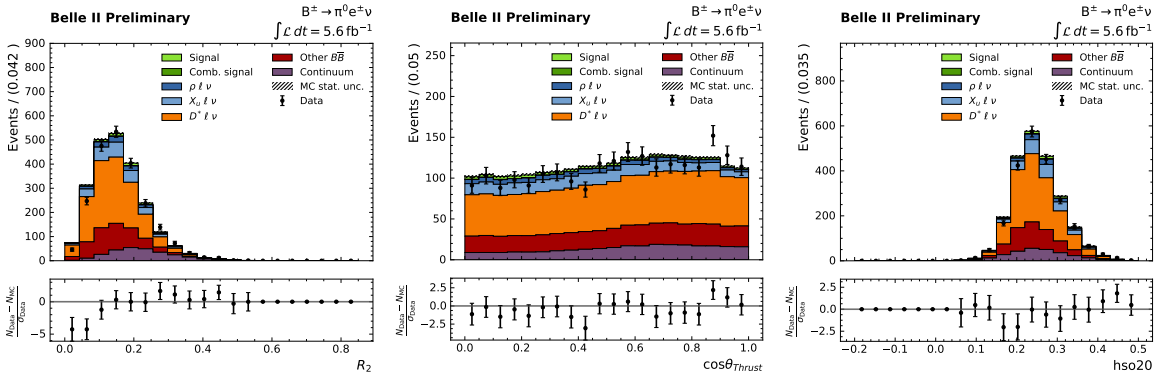
(j) $\cos\theta_z$ (k) CLEO Cone 1

Figure 7.8: Data-MC agreement and pulls for the continuum suppression BDT input variables integrated over all q^2 bins before a selection on the continuum suppression output classifier has been performed for the $B^{\pm} \to \pi^0 e\nu$ mode.



(a) (b)

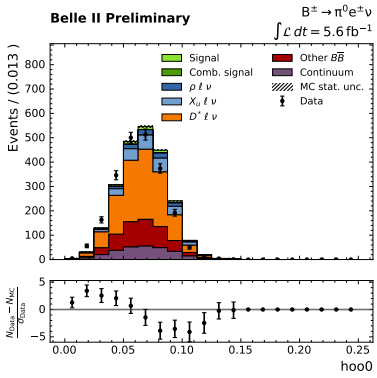
Figure 7.9: Data-MC agreement for the continuum suppression BDT output classifiers in the (a) $q1$ and (b) $q3$ bins of the $B^{\pm} \to \pi^0 e\nu$ mode.



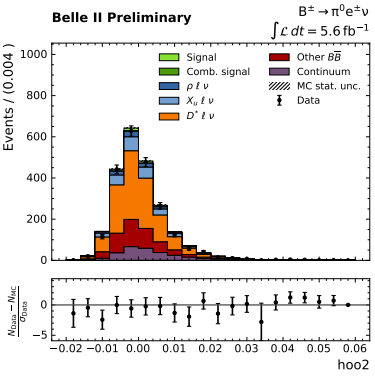
(a) R_2

(b) $\cos\theta_{\mathrm{Thrust}}$

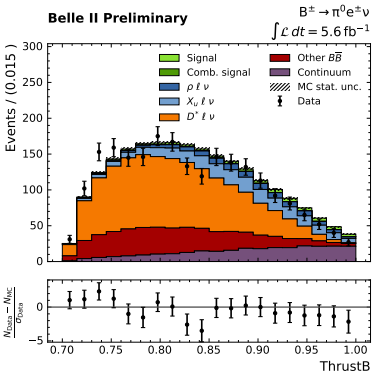
(c) h_{20}^{so}



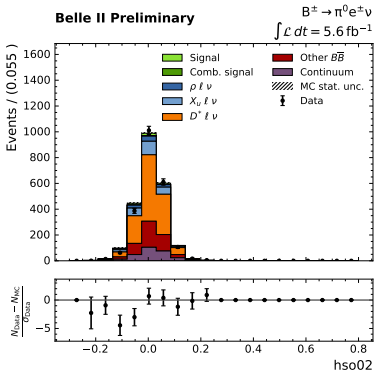
(d) h_0^{oo}



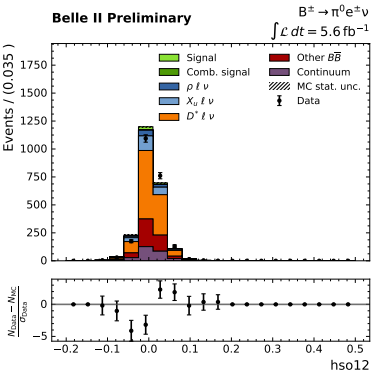
(e) h_2^{oo}



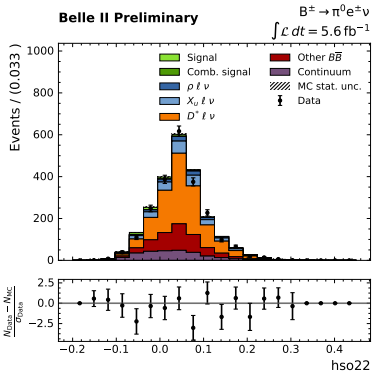
(f) Thrust



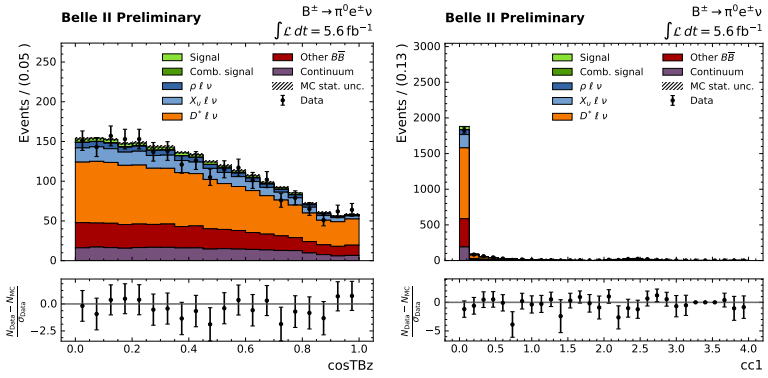
(g) h_2^{so}



(h) h_{12}^{so}

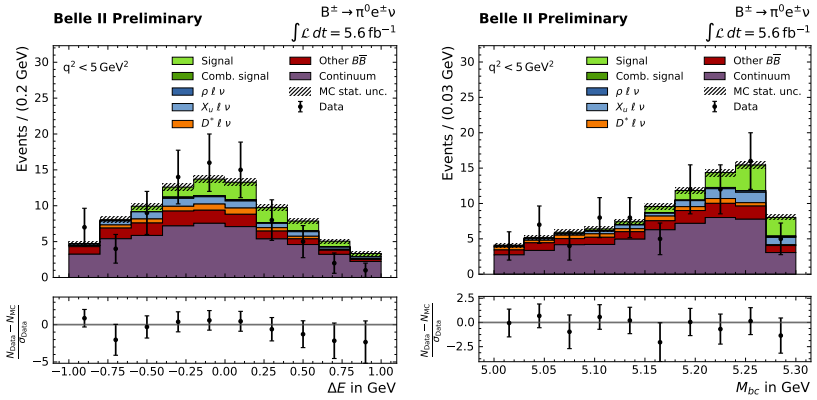


(i) h_{22}^{so}



(j) $\cos\theta_z$ (k) CLEO Cone 1

Figure 7.10: Data-MC agreement and pulls for the continuum suppression BDT input variables integrated over all q^2 bins after the selection on the continuum suppression output classifier has been performed.



(a) (b)

Figure 7.11: The data-MC agreement for the $q1$ bin of the $B^{\pm} \to \pi^0 e\nu$ mode for (a) ΔE and (b) M_{bc} after the selection on the BDT output classifier has been performed.

7.4 Pre-Fit ΔE and M_{bc}

The agreements between data and MC for ΔE and M_{bc} in the $q1$ bin of the $B^{\pm} \to \pi^0 e\nu$ mode are shown in (a) and (b) of Figure 7.11, respectively. At this point we are left with very low statistics. From MC we expect ~230 signal events per mode, split into 5 q^2 bins, in the B^0 modes. In the $B^{\pm}$ modes we only expect ~70 signal events. There are multiple reasons why there is a difference in these numbers between the charged and mixed B modes. One reason is that the branching fraction for $B^{\pm} \to \pi\ell\nu$ is smaller than $B^0 \to \pi\ell\nu$. In addition, the reconstruction and selection efficiencies are lower for the modes containing π^0s. Due to the low statistics, the statistical errors of the distributions are very high. Within the errrors, data and MC generally agree. In the muon modes we noticed some discrepancies in the lowest and highest q^2 bins. Continuum is the dominant background in the lowest q^2 bin, while in the highest q^2 bin $B \to X_u\ell\nu$ events are the biggest background. The intermediate q^2 bins are dominated by $B \to D^*\ell\nu$ background.

CHAPTER 8

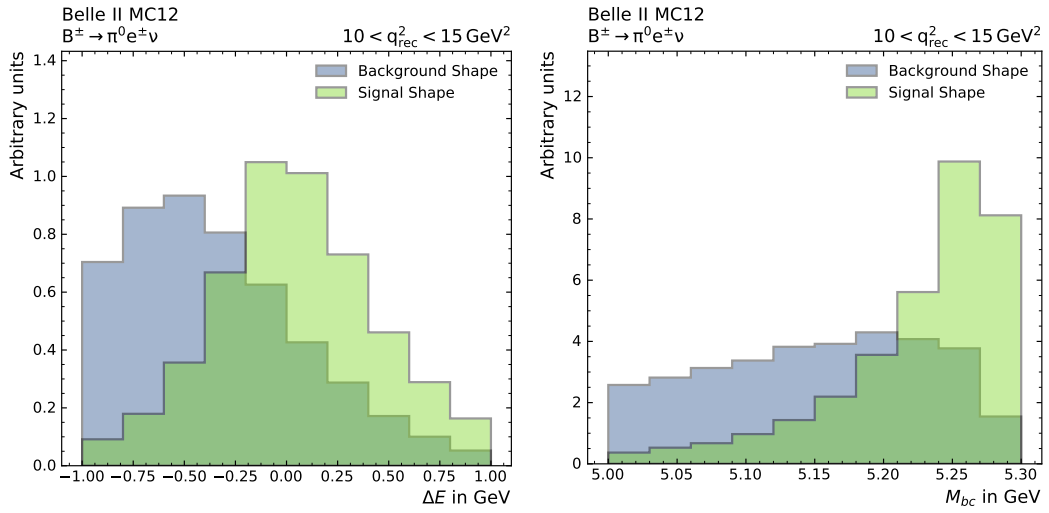
Fit

This chapter will discuss the extraction of signal from simultaneous extended binned maximum likelihood fits of ΔE and M_{bc} in q^2 bins. We will first validate the fit by looking at the pull distributions of toy studies. A toy is generated from a distribution by fluctuating the entries in each bin by the bin's statistical error. A fit to the toy can then be performed in the same way that a fit to data is performed. As will be seen, the available data sample is too small to perform a valid fit. Therefore, a toy study will be performed to estimate the integrated luminosity needed for a total branching fraction measurement to be significant.

8.1 Fit Setup

The goal is extracting the signal yields in each q^2 bin from simultaneous extended binned maximum likelihood fits of ΔE and M_{bc} to the data sample of $5.594\,\text{fb}^{-1}$. A binned maximum likelihood fit builds a likelihood function for the agreement between the to-be-fitted distribution and the distributions of the fit components. The distributions of the fit components are created from probability density functions (PDFs) of the fit component templates. Each template has an associated yield (number of entries). The fitter can either be told to keep the yield for a certain fit component fixed, or to let it float within a certain range. The fit result of a binned maximum likelihood fit is a yield for each fit component, corresponding to the maximum likelihood, while keeping the total yield fixed. An extended binned maximum likelihood fit, in addition, allows the total yield to vary within Poisson errors. We perform the fit using the ROOT fitting library RooFit [34] [35].

We have two fit components and let both their yields float. The fit components are signal and background. Since both yields scale with the $B \to \pi\ell\nu$ branching fraction, signal and combinatorial signal are combined into one component. We also combine all background categories into one fit component, due to the similarity in the shapes of the continuum and $\mathrm{B\overline{B}}$ backgrounds. The PDFs for the fit components are created from high statistics MC templates. The background templates contain all available generic MC, and the signal templates consist of all available signal MC. The sizes of these were described in Section 4.1. The normalised ΔE and M_{bc} distributions of the signal and background templates for the $q3$ bin of the $B^{\pm} \to \pi^0 e\nu$ mode are shown in Figure 8.1.



(a) (b)

Figure 8.1: Normalised distributions of the signal and background templates in (a) ΔE and (b) M_{bc} of the $q3$ bin in the $B^{\pm} \to \pi^0 e\nu$ mode.

8.2 Fit Validation

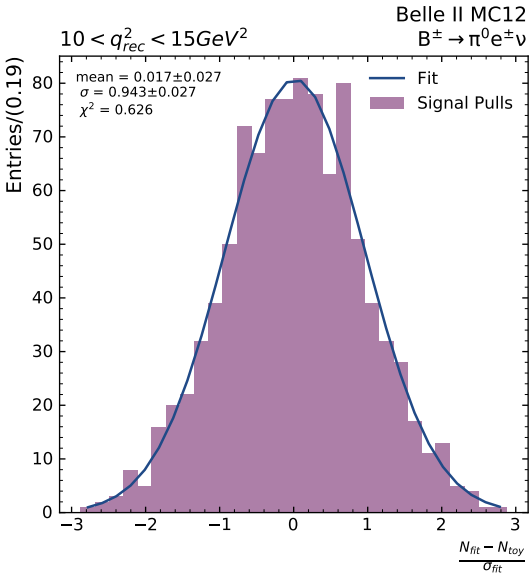
First we perform a fit validation by checking the fit stability. We also check that the fit does not introduce a bias into the fit result. In order to do this, we perform toy studies on toys of an integrated luminosity of $6\,\mathrm{fb}^{-1}$. Toy ΔE and M_{bc} distributions are generated by creating a histogram for each fit component, that is a result of letting the entries in each bin of the corresponding fit template fluctuate within the bin's Poissonian error, and combining the resulting histograms. The toys generated from the templates described above, are high statistics toys. Toys of $6\,\mathrm{fb}^{-1}$ are obtained by scaling of the high statistics toys. The original fit component distributions can then be fit to these toy distributions.

We generate 1000 such toys and look at the pull distributions of the fit results. The pull is the difference between the fit yield, N_{fit}, and the generated toy yield, N_{toy}, normalised by the fit error, σ_{fit}. It is given by

$$\mathrm{Pull} = \frac{N_{\mathrm{fit}} - N_{\mathrm{toy}}}{\sigma_{\mathrm{fit}}}. \tag{8.1}$$

The shape of the pull distributions should be Gaussian, with a mean of 0 and a standard deviation, σ, of 1. A deviation from a mean of 0 means the fit introduces a bias, while a devaition from a σ of 1 indicates that the error estimation is inaccurate. The pull distribution for signal found from the toy studies in the $q3$ bin in the $B^{\pm} \to \pi^0 e\nu$ mode is shown in (a) of Figure 8.2. A Gaussian fit is performed and the fit paramaters, as well as the reduced χ^2 value of the fit, are shown. The fit results of the means are generally biased in the negative direction by up to 9% . There is therefore an overestimation in the signal yield by the fit. The σ values also show a deviation from 1, up to 12% in some cases. The reduced χ^2 values of the fits lie between 0.8 and 1.8, which are acceptable values.

We similarly performed pull distribution studies for background and we encountered a deviation in the width of the pull distribution. The width is consistently smaller than 1 by $\sim$ 20%. The fit therefore



(a)

Figure 8.2: (a) Pull distribution of the toy study performed in the $q3$ bin of the $B^\pm \to \pi^0 e\nu$ mode. 1000 toys were generated.

systematically overestimates the error in the background yield. This has no direct effect on our signal yield measurement. One reason for these deviations could be that the error matrix of the fit to the toy data is not positive-definite and could not be calculated accurately. We noticed that this occured for some fits and the error matrix was consequently forced to be positive-definite, thus introducing an uncertainty into the fit results. Ideally, these cases would be excluded from the toy study.

While performing this toy study, we noticed that not all the fits to the toy data converged and that the correlations between signal and background were relatively high. Especially in the $q1$ and $q5$ bins correlations in the yields of up to 0.99 occured. This could be due to two reasons: (1) Introduction of correlations in ΔE and M_{bc} in signal and background due to selections or (2) statistical limits from the available data sample. As already noticed in selecting on the $B\overline{B}$ BDT output classifier, (1) occurs by performing cuts on variables that are correlated to ΔE and M_{bc}. Instead of relying solely on optimising the FOM, which focuses on the statistical significance, we might have to change our optimising figure so that it takes correlations to the signal extraction variables into account. On the other hand, we are also limited by (2). At this point, for 6 fb^{-1} toys, we end up with very low statistics toys. The histograms of the fit components resulting from a fit to such low statistics toys are likely to have bin-by-bin Poissonian errors that overlap with each other and so correlations will automatically be high for distributions this size. The reason for the failure of some fits is probabily a combination of (1) and (2). This is problematic when fitting to data, but in the above toy study only fits that did converge were included.

8.3 Fit to Data

Next, we attempted fitting to data in each q^2 bin. The correlations in the yields were very high, up to 0.99, and the fit failed for some q^2 bins. Even when the fit converged, often the error matrix was not determined accurately. All of the signal yields were consistent with 0; we are limited by statistics at this point.

8.4 Significance Prediction

Next, we performed a toy study to estimate the luminosity neded to reach significance in the total branching fraction for each mode. If the signal fit yield in one q^2 bin, i, is given by N^i_{fit}, the partial branching fraction in that bin is given by

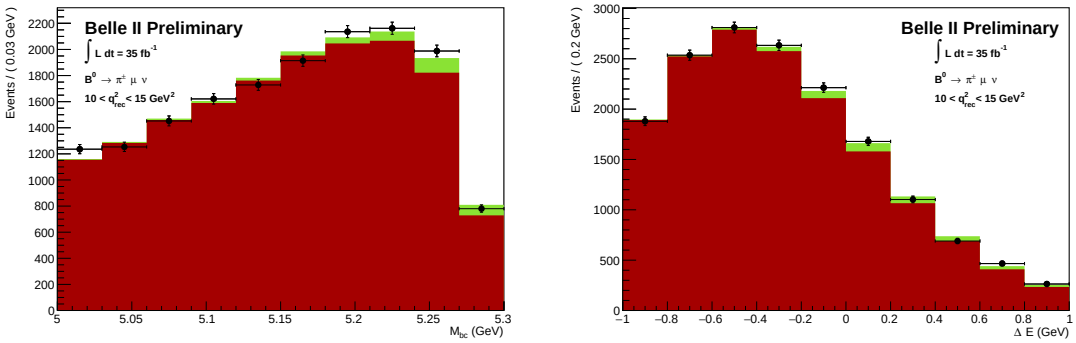
$$\text{BF}^i = \frac{N^i_{\text{fit}}}{2\epsilon^i_{\text{rec}}\epsilon^i_{\text{sel}} f_{\text{BB}} N_{\text{BB}}}, \tag{8.2}$$

where ϵ^i_{rec} and ϵ^i_{sel} are the reconstruction and selection efficiencies in bin i, respectively. f_{BB} gives the fraction of $\text{B}\overline{\text{B}}$ events that decay into $\text{B}^0\overline{\text{B}^0}$ or B^+B^-, with $f_{\text{BB}} = 0.514$ and $f_{\text{BB}} = 0.486$ for $B^\pm$ modes and B^0 modes, respectively. N_{BB} is the number of $\text{B}\overline{\text{B}}$ pairs produced at the specified luminosity. Once the partial branching fractions have been determined, the total branching fraction (BF) is given by the sum of the partial branching fractions:

$$\text{BF} = \sum_i \text{BF}^i. \tag{8.3}$$

In order to determine the integrated luminosity needed for significant results we generate toy distributions of different integrated luminosities, find the partial branching fractions in each q^2 bin, and from these the total BF with the corresponding significance, given by the BF divided by its error. We generate toys of 6, 15, 20, 25, 35, 50 and 75 fb^{-1}. An example fit to toy data of 35 fb^{-1} is given for the $q3$ bin of the $B^0 \to \pi^\pm \mu\nu$ mode in Figure 8.3. The correlations between the signal and background yields at integrated luminosities greater than 10 fb^{-1} for the B^0 mode, or 15 fb^{-1} for the $B^\pm$ mode, are lower than the ones observed at 6 fb^{-1}. They lie between 0 and 0.7. Once the total branching fractions at each integrated luminosity have been found, we plot the significance in the BF versus the integrated luminosity, as shown for all four modes in Figure 8.4.

We expect the significance to scale with the square root of the number of events, and therefore with the square root of the luminosity. However, since we are only looking at part of the spectrum, we assume the behaviour can be approximated by a linear function in this regime and we therefore fit a linear function to the data. From the fit we determine the luminosities needed to achieve 3σ and 5σ significance. These values are given in Table 8.1. We predict that a dataset approximately 3 and 4 times the size of the current dataset is needed for 3σ and 5σ significance in the B^0 modes. In the $B^\pm$ modes 6 and 10 times the dataset size is neccessary to achieve the same significance. We expect this is both due to the lower initial statistics in the $B^\pm$ modes, and due to misreconstruction of the π^0, resulting in higher levels of background. We need to note, however, that this is purely the statistical significance. No systematic factors have been accounted for in this estimation.



(a) (b)

Figure 8.3: Fit to toy distributions of 35 fb^{-1} for (a) M_{bc} and (b) ΔE in the $q3$ bin of the $B^0 \to \pi^{\pm}\mu\nu$ mode. Signal is coloured green and background is coloured red.

Mode	**Luminosity [fb^{-1}]**	
	3σ Significance	**5σ Significance**
$B^0 \to \pi^{\pm} e\nu$	13.6	19.8
$B^{\pm} \to \pi^0 e\nu$	35.1	54.3
$B^0 \to \pi^{\pm}\mu\nu$	14.8	21.9
$B^{\pm} \to \pi^0\mu\nu$	33.8	54.3

Table 8.1: Predicted integrated luminosity needed for 3σ and 5σ significance in the total branching fraction measurement for the four modes.

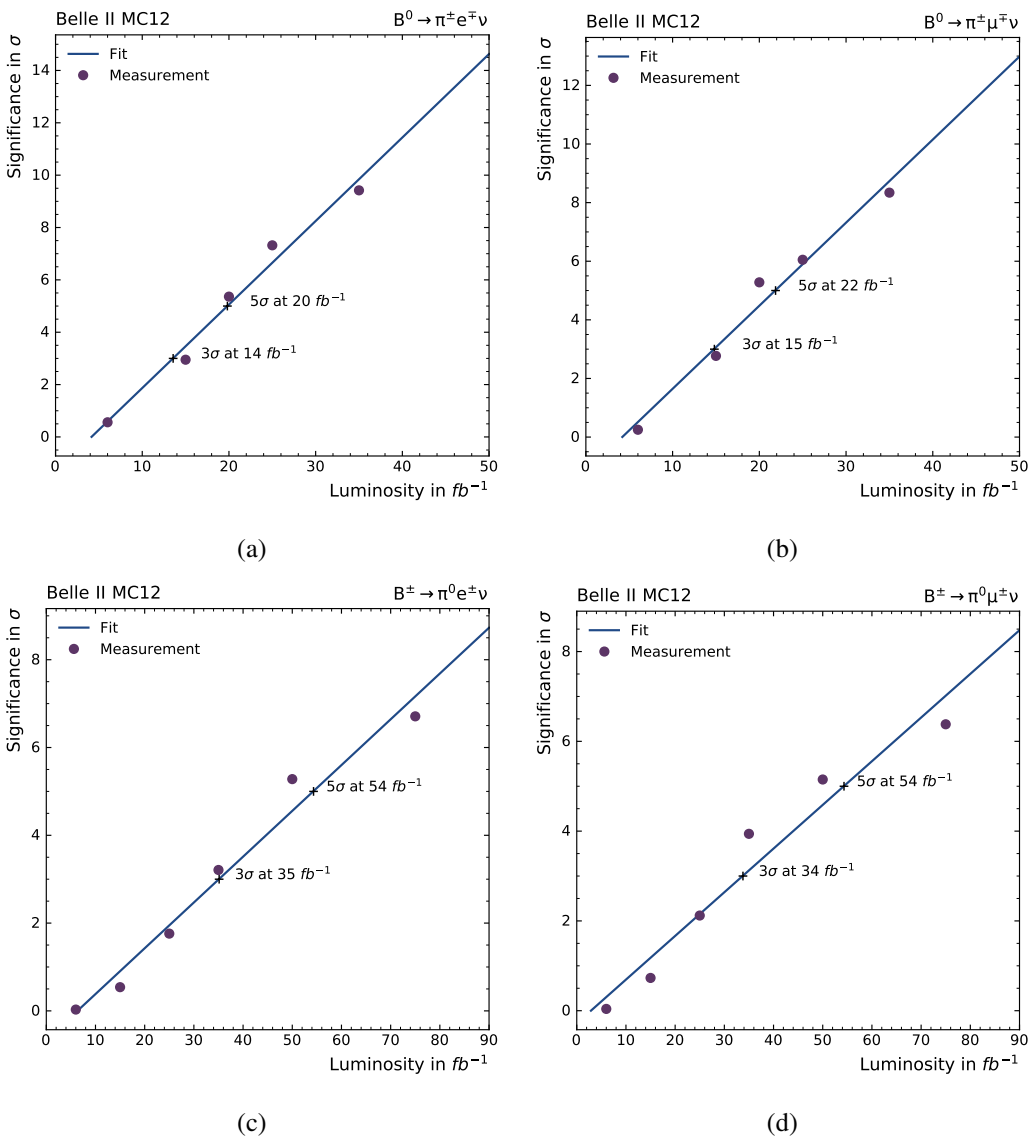


Figure 8.4: Plots of the significance in the total branching fraction versus the integrated luminosity as found from toy studies. The data was fit with a linear function to predict the integrated luminosities at which 3σ and 5σ significance is achieved. These points are marked with black crosses and labelled.

CHAPTER 9

Conclusion and Outlook

9.1 Conclusion

In this thesis an untagged exclusive analysis of the charmless semileptonic decay $B \to \pi \ell \nu$ using an integrated luminosity of $5.6\,\text{fb}^{-1}$ of early Belle II data was discussed. The analysis was performed in 5 bins of momentum transfer, which makes it sensitive to form factor parametrisations. This method was chosen to prepare for later $|V_{ub}|$ extraction using a larger dataset. The signal selections were separately optimised in the momentum transfer bins using a figure of merit (FOM) based on the statistical significance of signal on top of background. First, selections based on detector properties and the kinematics of the signal decay were performed. After that, boosted decision trees (BDTs) were employed to separate signal from continuum background, as well as from $\mathrm{B\overline{B}}$ background.

After selecting on the $\mathrm{B\overline{B}}$ BDT output classifier the shapes in signal and background of the signal extraction variables, M_{bc} and ΔE, were very similar and the correlations between the yields were high. Therefore, in the end, a selection was only placed on the continuum suppression BDT output classifier. Finally, we set up and validated a simultaneous extended binned maximum likelihood fit on the M_{bc} and ΔE distributions to extract signal in each momentum transfer bin.

Fitting to data was not possible, due to very low statistics and high correlations in the yileds of the templates. We performed toy studies to predict the luminosities necessary to achieve statistical significance of 3σ and 5σ in the total branching fraction for each of the modes. In the B^0 modes a luminosity of $\sim 15\,\text{fb}^{-1}$ and $\sim 20\,\text{fb}^{-1}$ is needed for 3σ and 5σ significance, respectively. For the $B^{\pm}$ modes higher luminosities are needed: $\sim 30\,\text{fb}^{-1}$ for 3σ significance and $\sim 50\,\text{fb}^{-1}$ for 5σ significance.

Just like the FOM, these predictions only take the statistical significance into account. In performing this analysis, we noticed that at this stage, with very low statistics, the chosen FOM is not valid. For example, it is not able to incorporate the effects due to shape similarities. We are also limited by other factors, such as the reconstruction efficiency or the shape disagreement between data and MC. The shape agreement is especially affected by mismodelling of both the size and the shape of beam background. In order to improve the analysis all the above need to be investigated further.

9.2 Outlook

This analysis will be continued and further optimised to run on a larger dataset once it becomes available. Systematic effects will be determined, such as the effects from beam background. Techniques

to reduce these will be investigated in the form of studies to clean the ROE, as well as a further analysis of ECL variables. Currently the perfomance of the timing between clusters is unreliable, but when its use becomes safe, we expect it to be a good discrimator. Another option is training a BDT to identify events containing beam background particles.

In addition to the systematic effects, the current analysis steps also need to be evaluated for their performance, and improved upon. As stated above, a new kind of FOM will need to be investigated, which in some way encompasses not just the statistical significance, but also the correlations. The fit method will also be optimised. Ideally, we want to split the background template into the separate background components, and separately look for ways to fix or constrain these fit components. One example could be using the branching fraction of $B \to D^* \ell \nu$ to constrain the corresponding fit component. Next to that, different fit methods, maybe less prone to suffer from high correlations in one-dimensional distributions, will be investigated. One possibility is a two dimensional fit of M_{bc} and ΔE. In addition, this analysis will be combined with the analysis of the semileptonic decay $B \to \rho \ell \nu$. This might offer oportunities to cross-feed results from one mode to the other and thus constrain fit components further.

Bibliography

[1] J. Thomson, *Cathode Rays*, Philosophical Magazine **40** (1897) (cit. on p. 1).

[2] S. W. Herb et al.,
Observation of a Dimuon Resonance at 9.5GeV in 400GeV Proton-Nucleus Collisions,
Phys. Rev. Lett. **39** (1977) (cit. on p. 1).

[3] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012),
arXiv: 1207.7214 (cit. on p. 1).

[4] C. Bouchard, L. Cao and P. Owen,
Summary of the 2018 CKM working group on semileptonic and leptonic b-hadron decay,
(2019), arXiv: 612.03765 (cit. on p. 1).

[5] Heavy Flavor Averaging Group, *Combined exclusive |Vub| and |Vcb|*,
[Online; accessed August 13, 2019], URL: https://hflav-eos.web.cern.ch/hflav-eos/semi/spring19/html/ExclusiveVub/exclVubVcb.html (cit. on p. 2).

[6] Wikipedia, the free encyclopedia, *Standard Model of Elementary Particles*,
[Online; accessed August 4, 2019], 2006,
URL: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (cit. on p. 4).

[7] D. H. Perkins, *Introduction to High Energy Physics*, 1982, ISBN: 9780521621960 (cit. on p. 3).

[8] F. Halzen and A. D. Martin,
Quarks and Leptons: An Introductory Course in Modern Particle Physics, Wiley, 1984,
ISBN: 9780471887416 (cit. on p. 3).

[9] A. Pich, *Electroweak Symmetry Breaking and the Higgs Boson*,
Acta Physica Polonica B **47** (2016), arXiv: 512.08749v1 (cit. on p. 4).

[10] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98** (3 2018)
(cit. on pp. 5, 10, 19, 25, 26).

[11] *Study of B to pi l nu and B to rho l nu decays and determination of |Vub|*,
Phys. Rev. D **83** (3 2011), arXiv: 1005.3288v2 (cit. on p. 6).

[12] T. Keck,
Machine learning algorithms for the Belle II experiment and their validation on Belle data,
ETP-KA/2017-31, PhD Thesis: Karlsruhe Institute of Technology, 2017 (cit. on p. 6).

[13] T. Adams et al., *Gravitational-Wave Detection using Multivariate Analysis*,
Phys. Rev. D **88** (2013), arXiv: 1305.5714 (cit. on p. 6).

[14] A. G. Patriota, *A classical measure of evidence for general null hypotheses*, Fuzzy Sets and Systems **233** (2013), arXiv: 1201.0400 (cit. on p. 7).

[15] T. Keck, *FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification*, CoRR **abs/1609.06119** (2016), arXiv: 1609.06119 (cit. on p. 7).

[16] Belle II Collaboration, *Belle II Technical Design Report*, 2010, arXiv: 1011.0352 (cit. on pp. 9, 13–16).

[17] K. Akai, K. Furukawa and H. Koiso, *SuperKEKB collider*, Nuclear Instruments and Method in Physics Research A **907** (2018), arXiv: 1809.01958 (cit. on p. 10).

[18] E. Kou et al., *The Belle II Physics Book*, 2018, arXiv: 1808.10567 (cit. on pp. 10, 11, 18).

[19] D. Weyland, *Continuum Suppression with Deep Learning techniques for the Belle II Experiment*, ETP-KA/2017-30, Master Thesis: Karlsruhe Institute of Technology, 2017 (cit. on p. 11).

[20] A. Moll, *Comprehensive study of the background for the Pixel Vertex Detector at Belle II*, PhD Thesis: Ludwig-Maximilians-Universitat Munchen, 2015 (cit. on p. 12).

[21] P. Lewis et al., *First Measurements of Beam Backgrounds at SuperKEKB*, Nuclear Instruments and Methods in Physics Research A **914** (2019), arXiv: 1802.01366 (cit. on p. 12).

[22] B. I. Collaboration, *Belle II Detector 3D model*, [Online; accessed August 6, 2019], URL: https://www.belle2.org/archives/ (cit. on p. 14).

[23] N. Braun, *Combinatorial Kalman Filter and High Level Trigger Reconstruction for the Belle II Experiment*, ETP-KA/2018-23, PhD Thesis: Karlsruhe Institute of Technology, 2018 (cit. on p. 16).

[24] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) (cit. on p. 17).

[25] T. Sjostrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual*, Journal of High Energy Physics (2006), arXiv: 0603175 (cit. on p. 18).

[26] S. Jadach, B. Ward and Z. Was, *The precision Monte Carlo event generator KK for two-fermion final states in e+e- collisions*, Computer Physics Communications **130** (2000), arXiv: 9912214 (cit. on p. 18).

[27] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) (cit. on p. 18).

[28] G. Brandenburg et al., *Charged track multiplicity in B meson decay*, Phys. Rev. D **61** (7 2000), arXiv: 9907057v1 (cit. on p. 32).

[29] M. Rohrken, *Time-Dependent CP Violation Measurements in Neutral B Meson to Double-Charm Decays at the Japanese Belle Experiment*, IEKP-KA-2012-13, PhD Thesis: Karlsruhe Institute of Technology, 2012 (cit. on p. 42).

[30] A. J. Bevan et al., *The Physics of the B Factories*, Eur. Phys. J. **C74** (2014), arXiv: 1406.6311 (cit. on pp. 42, 43).

[31] G. C. Fox and S. Wolfram,
Observables for the Analysis of Event Shapes in e+ e- Annihilation and Other Processes,
Phys. Rev. Lett. **41** (1978) (cit. on p. 43).

[32] D. M. Asner et al., *Search for exclusive charmless hadronic B decays*, Phys. Rev. **D53** (1996),
arXiv: hep-ex/9508004 (cit. on p. 43).

[33] F. Tenchini, *A Decay Tree Fitter for the Belle II Analysis Framework*,
25th International Workshop on Vertex Detectors (2016) (cit. on p. 56).

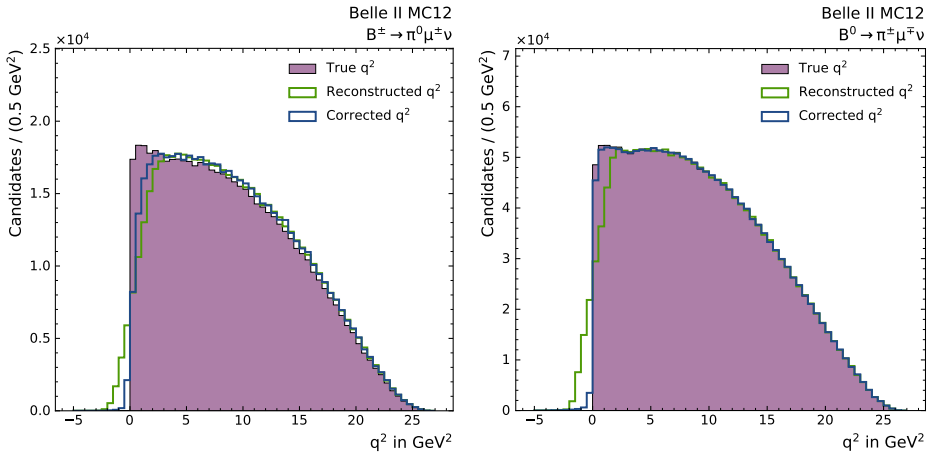
[34] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*,
Nucl. Instrum. Meth. **A389** (1997) (cit. on p. 63).

[35] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003),
arXiv: physics/0306116 (cit. on p. 63).

APPENDIX A

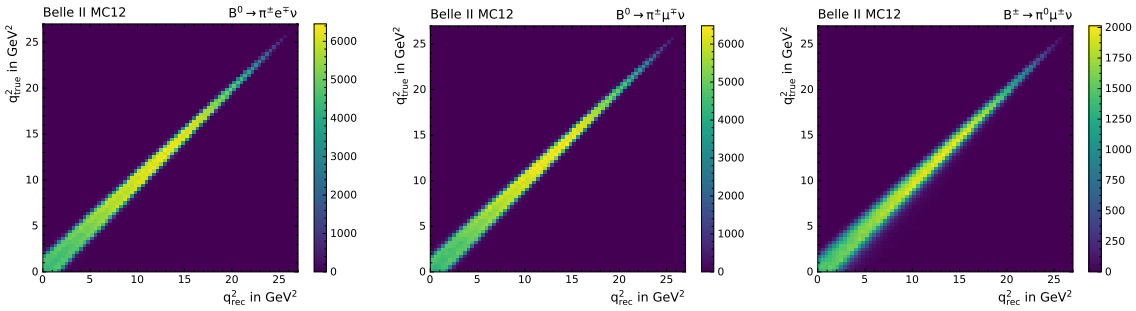
Reconstruction and Signal Characteristics

A.1 Reconstruction Effects



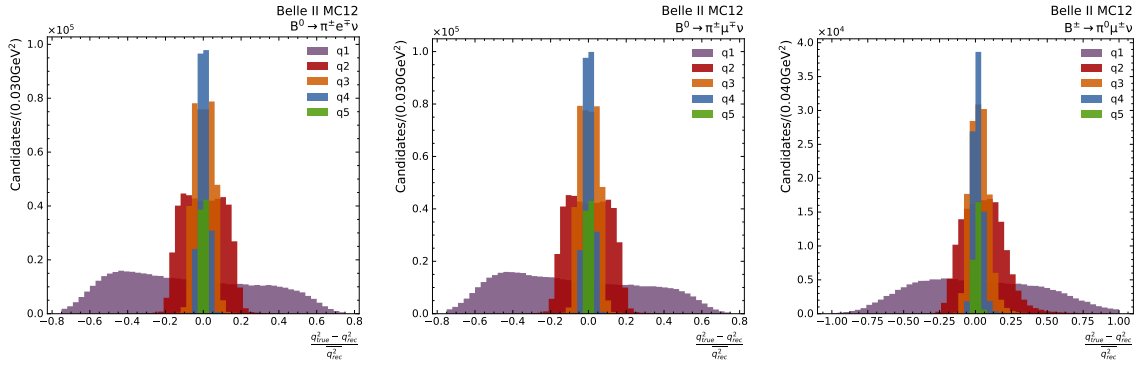
(a) (b)

Figure A.1: True, reconstructed and corrected q^2 for correctly reconstructed signal events from signal MC in the (a) $B^{\pm} \to \pi^0 \mu \nu$ and (b) $B^0 \to \pi^{\pm} \mu \nu$ modes.



(a) (b) (c)

Figure A.2: 2D histograms of true versus reconstructed q^2 for correctly reconstructed signal events found from signal MC in the (a) $B^0 \to \pi^\pm e\nu$ mode, (b) the $B^0 \to \pi^\pm \mu\nu$ mode and (c) the $B^\pm \to \pi^0 \mu\nu$ mode.

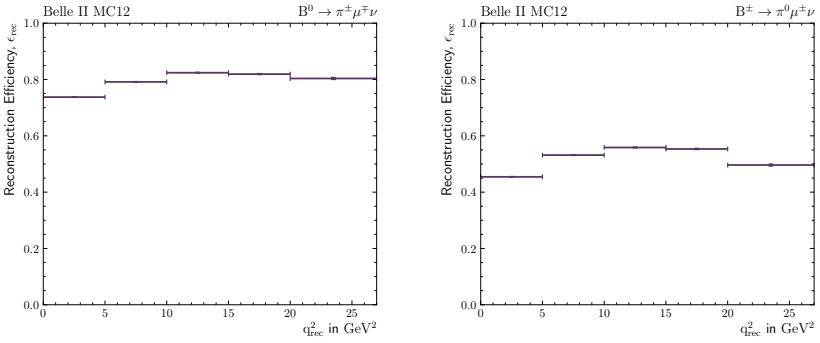


(a) (b) (c)

Figure A.3: The distribution of the normalised resolution of q^2 for the five q^2 bins from signal MC in the (a) $B^0 \to \pi^\pm e\nu$ mode, (b) the $B^0 \to \pi^\pm \mu\nu$ mode and (c) the $B^\pm \to \pi^0 \mu\nu$ mode.

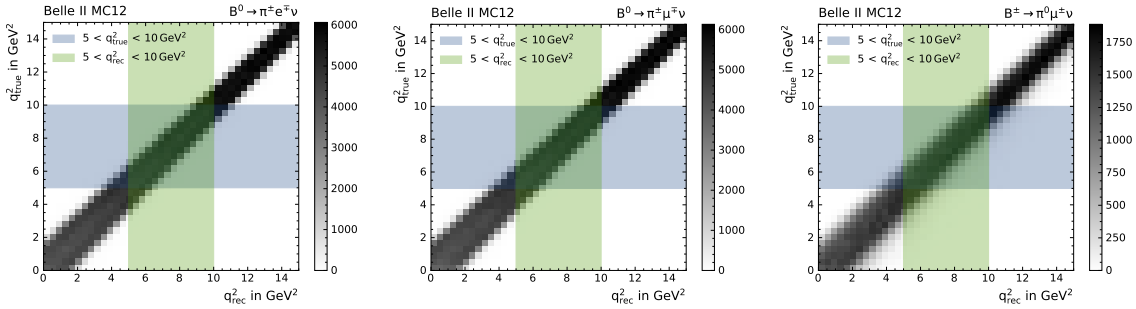
Mode	**Resolution**	**q^2 Bin**				
		q1	***q2***	***q3***	***q4***	***q5***
$B^0 \to \pi^\pm e\nu$	Centre	-0.369	-0.080	0.037	0.0180	0.006
	FWHM	1.050	0.324	0.148	0.072	0.032
$B^\pm \to \pi^0 e\nu$	Centre	-0.246	0.088	0.041	0.018	0.006
	FWHM	1.267	0.377	0.168	0.082	0.035
$B^0 \to \pi^\pm \mu\nu$	Centre	-0.365	-0.081	-0.035	0.001	0.006
	FWHM	1.043	0.324	0.148	0.065	0.032
$B^\pm \to \pi^0 \mu\nu$	Centre	-0.246	0.089	0.040	0.018	0.005
	FWHM	1.260	0.379	0.168	0.081	0.035

Table A.1: FWHM and peak positions of the distributions of the normalised q^2 resolution for the five q^2 bins in the four modes.



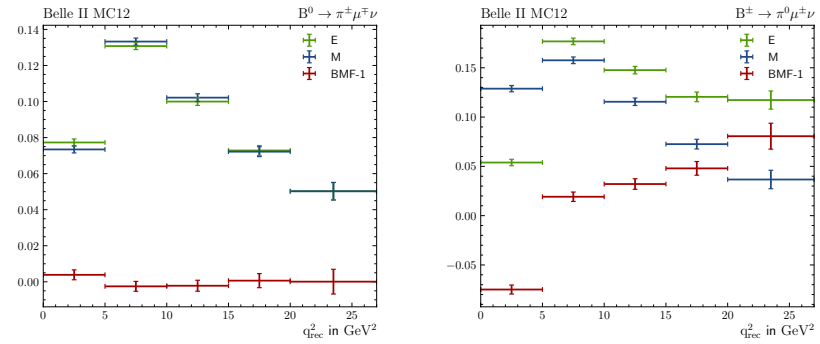
(a) (b)

Figure A.4: The reconstruction efficiencies across the five q^2 bins for the (a) $B^0 \to \pi^\pm \mu\nu$ and (b) $B^\pm \to \pi^0 \mu\nu$ mode found from signal MC



(a) (b) (c)

Figure A.5: Zoomed-in 2D histograms of true versus reconstructed q^2 for correctly reconstructed signal events found from signal MC in the (a) $B^0 \to \pi^\pm e\nu$ mode, (b) the $B^0 \to \pi^\pm \mu\nu$ mode and (c) the $B^\pm \to \pi^0 \mu\nu$ mode.



(a) (b)

Figure A.6: E, M and $BMF - 1$ for the five q^2 bins in the (a) $B^0 \to \pi^\pm \mu\nu$ and (b) $B^\pm \to \pi^0 \mu\nu$ modes.

A.2 Signal Characteristics

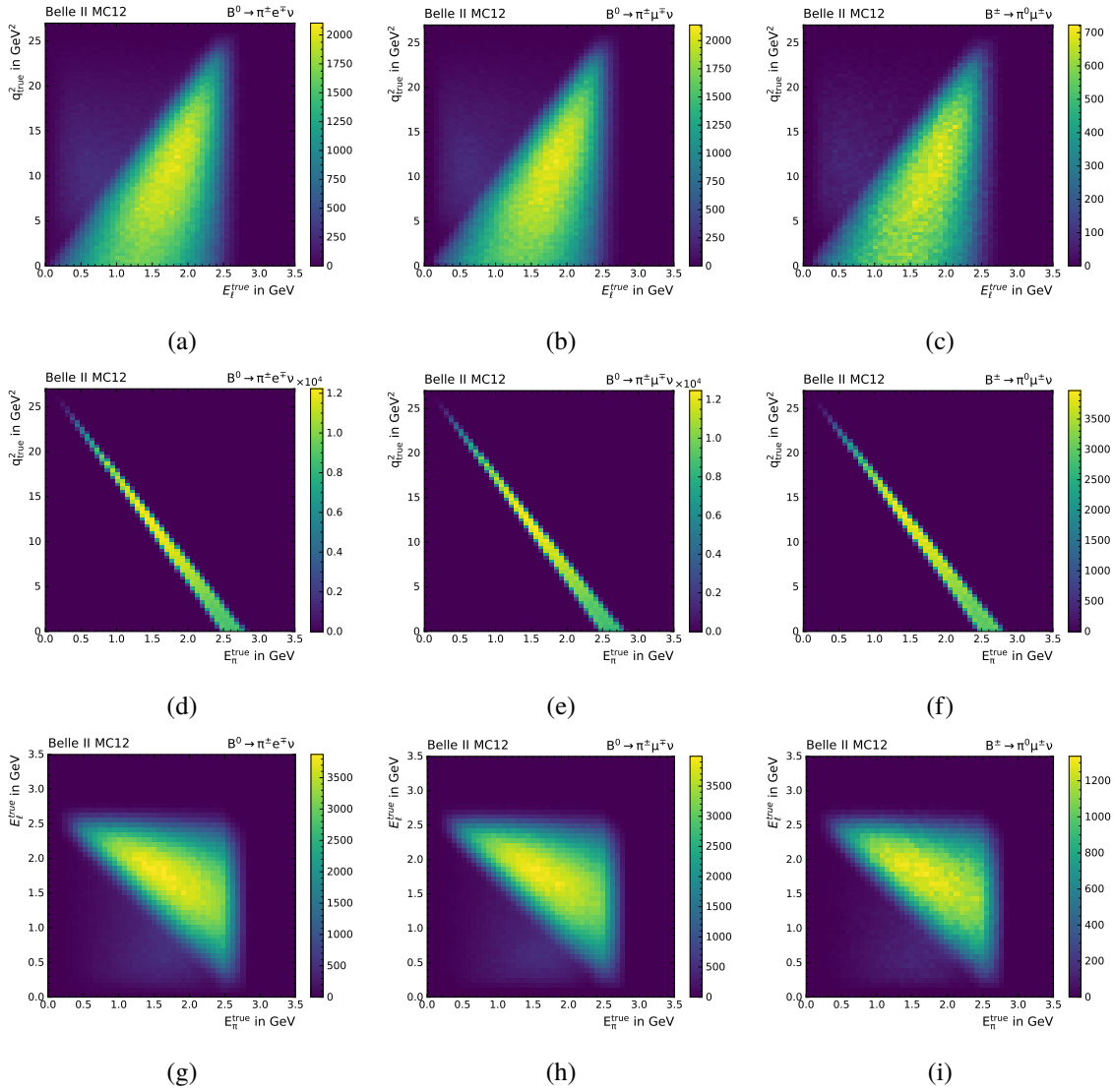


Figure A.7: 2D histograms for correctly reconstructed signal from signal MC. (a)-(c) The true momentum transfer, q^2, versus the true lepton energy, E_ℓ. (d)-(f) The true momentum transfer, q^2, versus the true pion energy, E_π. (e)-(f) The true lepton energy, E_ℓ, versus the true pion energy, E_π. The distributions are shown for the $B^0 \to \pi^\pm e\nu$ mode (left), the $B^0 \to \pi^\pm \mu\nu$ mode (middle) and the $B^\pm \to \pi^0 \mu\nu$ mode (right).

APPENDIX B

Selections and Data-MC Agreement

B.1 Lepton and Fit Region Selection

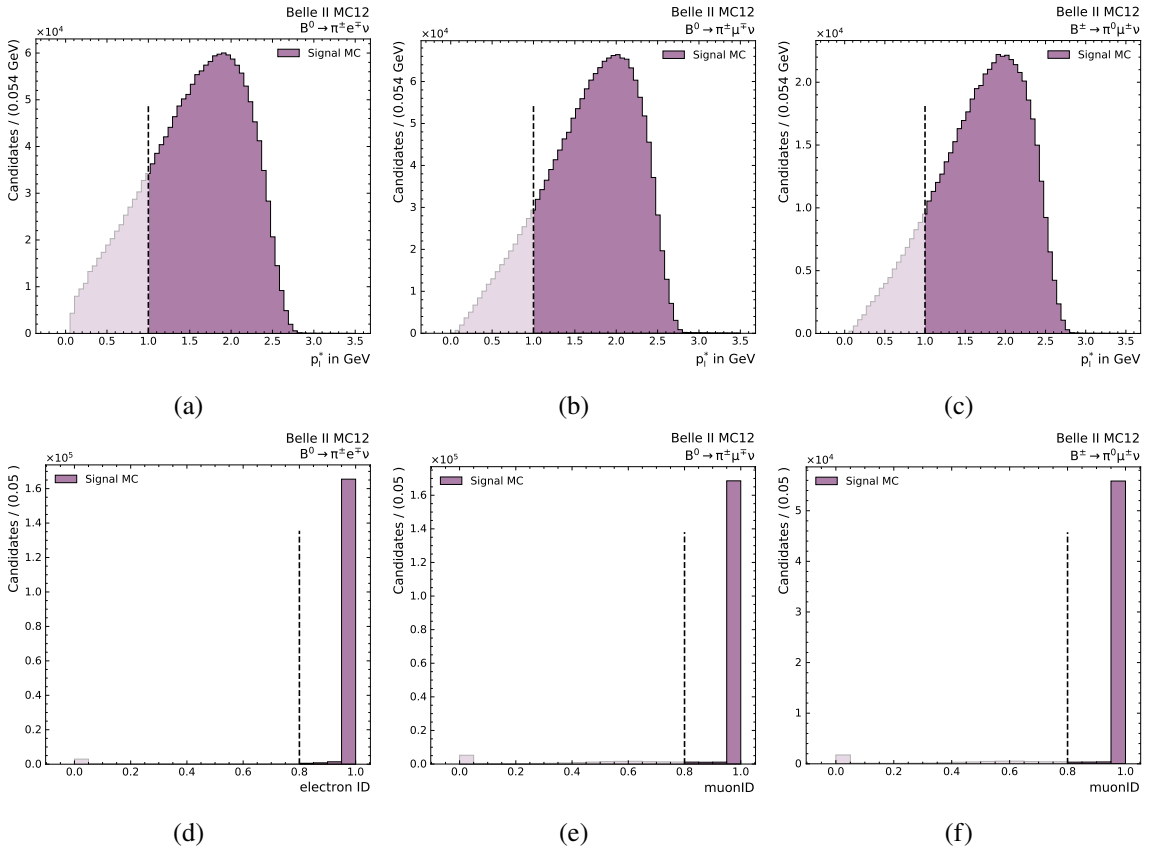


Figure B.1: The distributions from signal MC of the centre of mass lepton momentum (top) and of the electron/muon ID after the lepton momentum selection (bottom). The distributions and the corresponding selections on the lepton momentum and lepton ID are shown for the $B^0 \to \pi^\pm e\nu$ (left), $B^0 \to \pi^\pm \mu\nu$ (middle) and $B^\pm \to \pi^0 \mu\nu$ (right) modes.

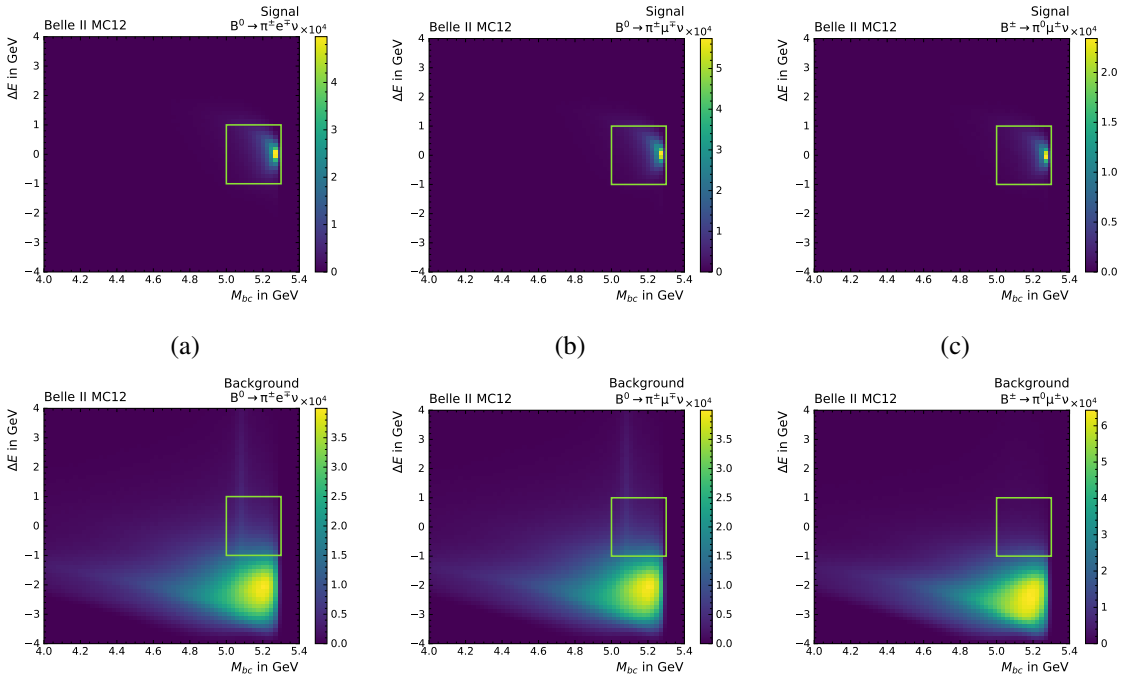


Figure B.2: 2D distributions of ΔE and M_{bc} in MC for signal (top) and background (bottom) in the $B^0 \to \pi^\pm e\nu$ (left), $B^0 \to \pi^\pm \mu\nu$ (middle) and $B^\pm \to \pi^0 \mu\nu$ (right) modes. The fit region is given by the green box.

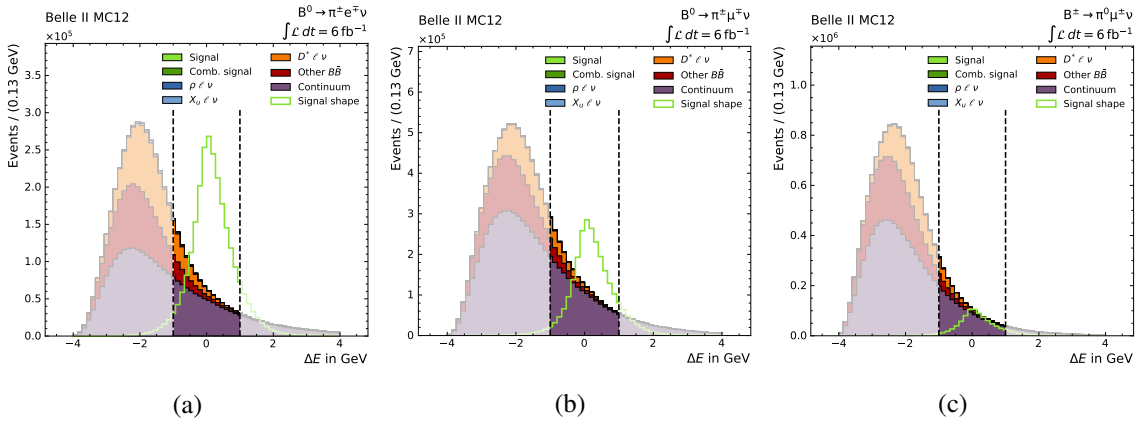
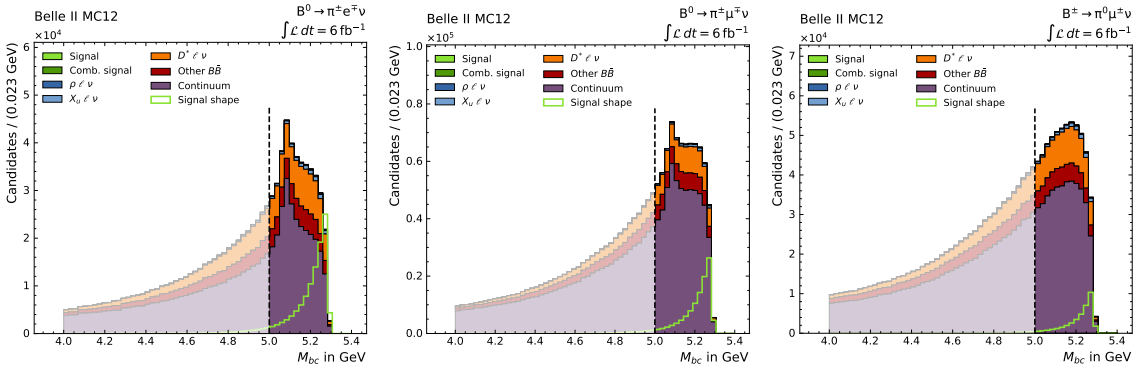
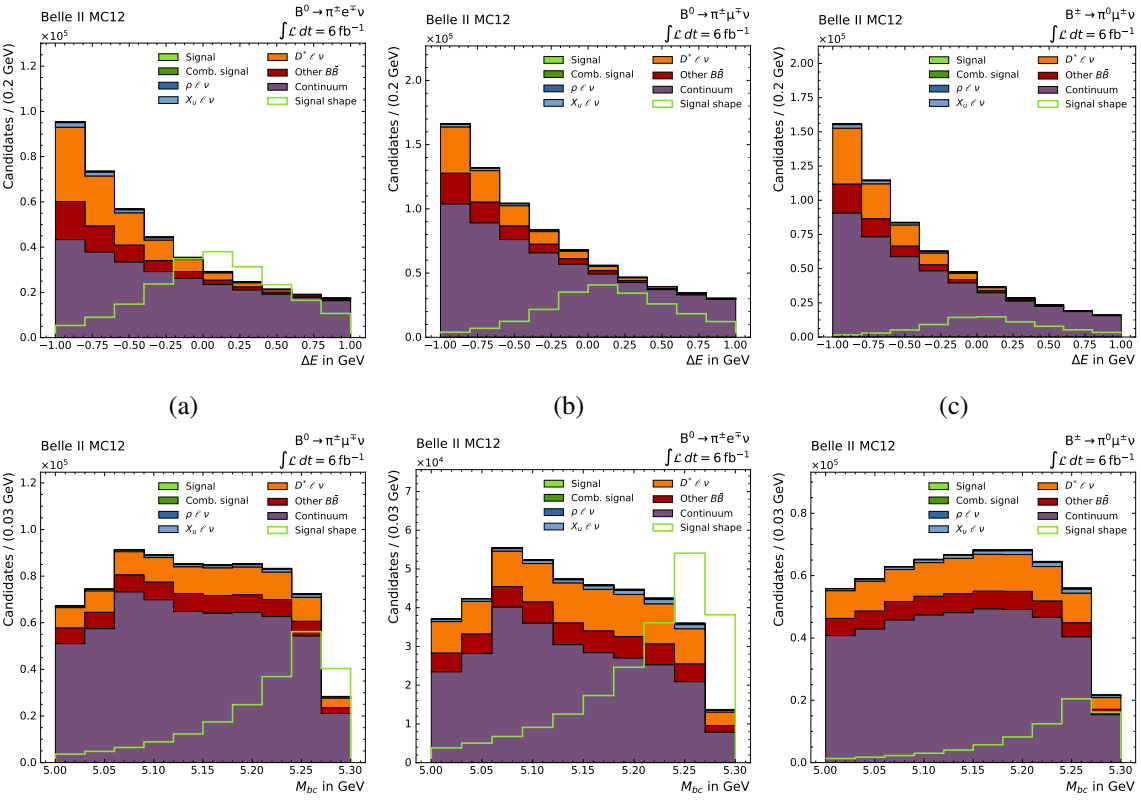


Figure B.3: The ΔE distributions for the (a) $B^0 \to \pi^\pm e\nu$, (b) the $B^0 \to \pi^\pm \mu\nu$ and (c) $B^\pm \to \pi^0 \mu\nu$ modes. The selections on ΔE are marked.



(a) (b) (c)

Figure B.4: The M_{bc} distributions after the ΔE selection has been performed. The distributions for the (a) $B^0 \to \pi^\pm e\nu$, (b) $B^0 \to \pi^\pm \mu\nu$ and (c) $B^\pm \to \pi^0 \mu\nu$ modes. The selections on M_{bc} are marked.



(a) (b) (c)

(d) (e) (f)

Figure B.5: The (a)-(c) ΔE and (d)-(f) M_{bc} distributions in MC for the $B^0 \to \pi^\pm e\nu$ (left), $B^0 \to \pi^\pm \mu\nu$ (middle) and $B^\pm \to \pi^0 \mu\nu$ modes (right) after the lepton and fit region selections have been performed.

B.2 Preselections

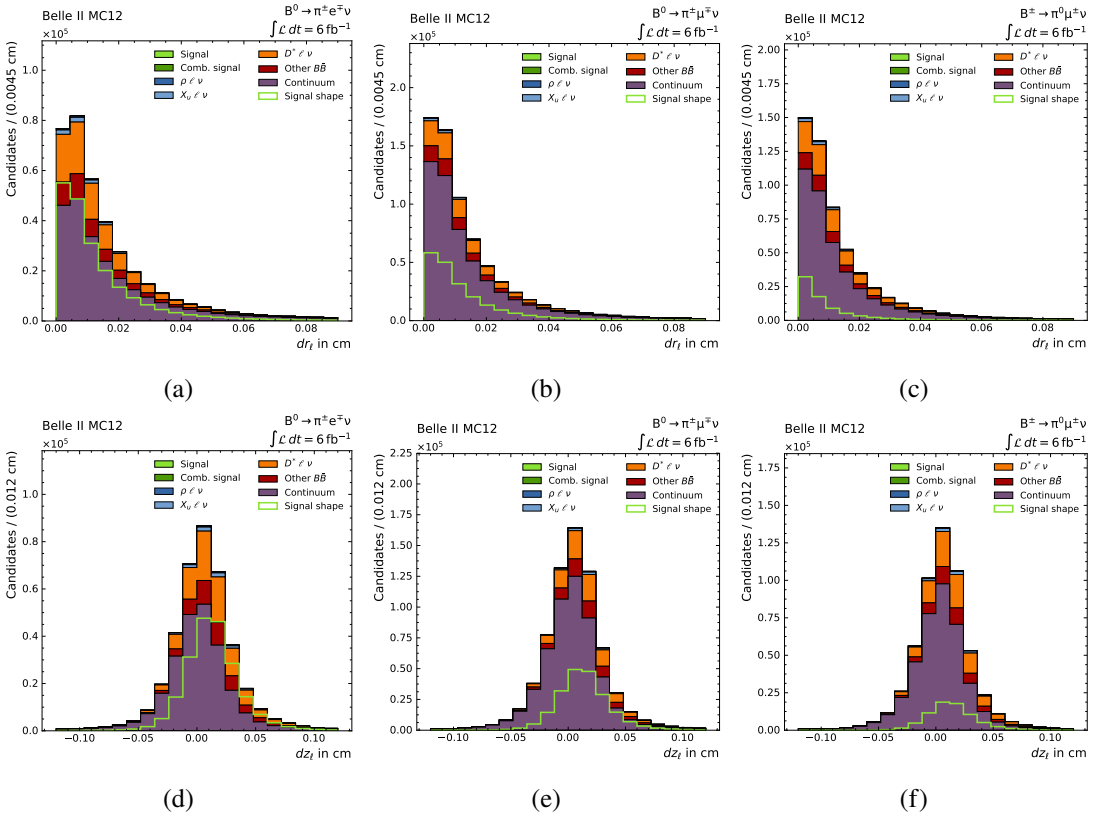


Figure B.6: Distributions of (a)-(c) dr and (d)-(f) dz of the lepton from MC in the $B^0 \to \pi^\pm e\nu$ (left), $B^0 \to \pi^\pm \mu\nu$ (middle) and $B^\pm \to \pi^0 \mu\nu$ mode (right).

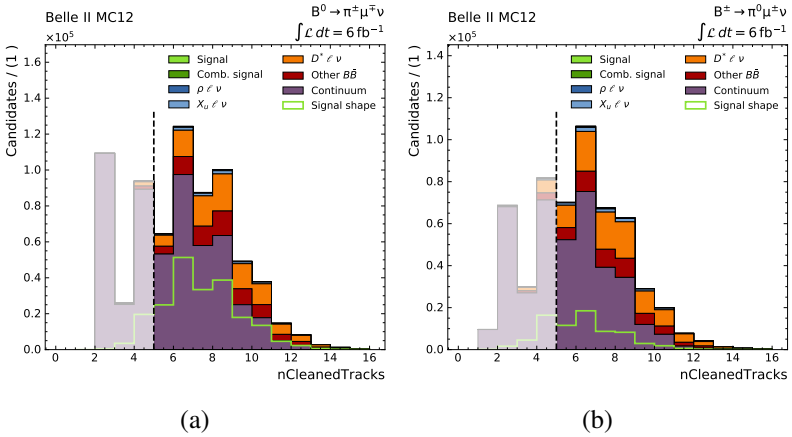
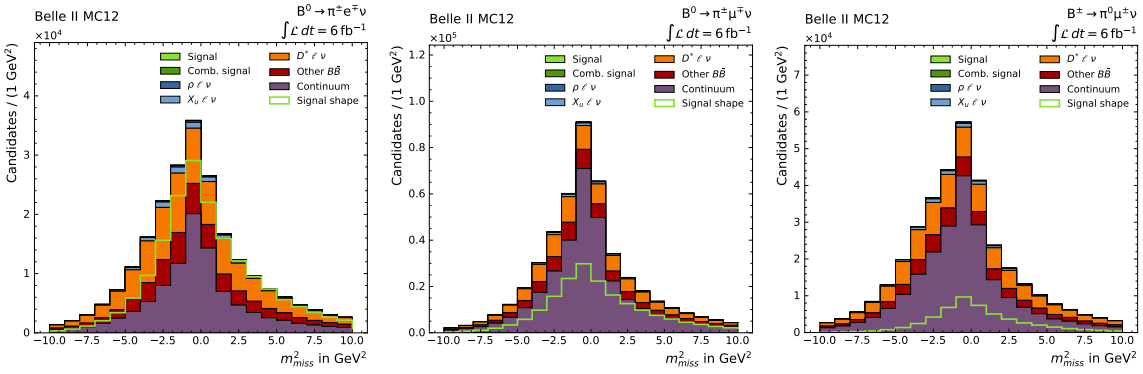
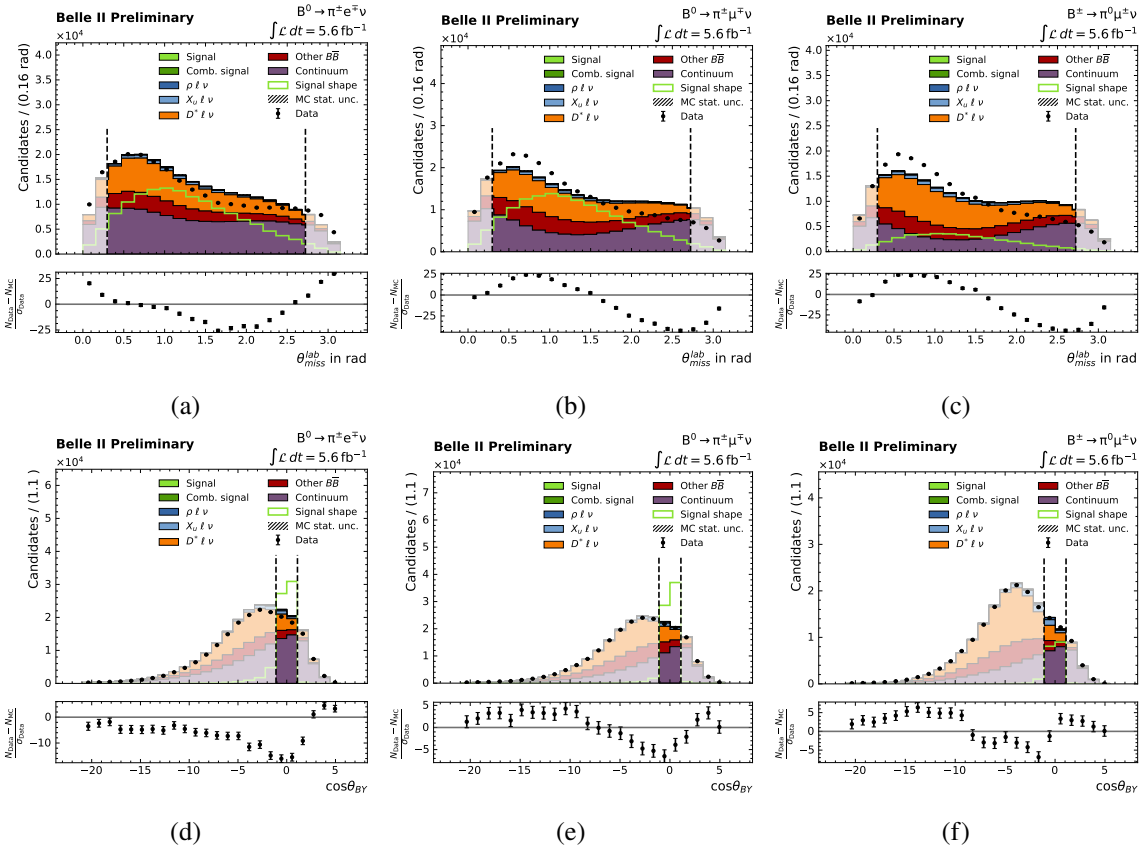


Figure B.7: Distribution of the number of cleaned tracks, nCleanedTracks, in MC for the (a) $B^0 \to \pi^\pm e^\mp \nu$ and (b) $B^\pm \to \pi^0 e^\pm \nu$ modes.



(a) (b) (c)

Figure B.8: Distribution of the missing mass squared, m^2_{miss}, from MC in the (a) $B^0 \to \pi^\pm e\nu$, (b) $B^0 \to \pi^\pm \mu\nu$ and (c) $B^\pm \to \pi^0 \mu\nu$ modes.



(a) (b) (c)

(d) (e) (f)

Figure B.9: Data-MC agreement and pulls for (a)-(c) θ_{miss} and (d)-(f) $\cos\theta_{BY}$ in the $B^0 \to \pi^\pm e\nu$ (left), $B^0 \to \pi\pm\mu\nu$ (middle) and $B^\pm \to \pi^0 \mu\nu$ (right) modes.

B.3 Photon Angle Selection

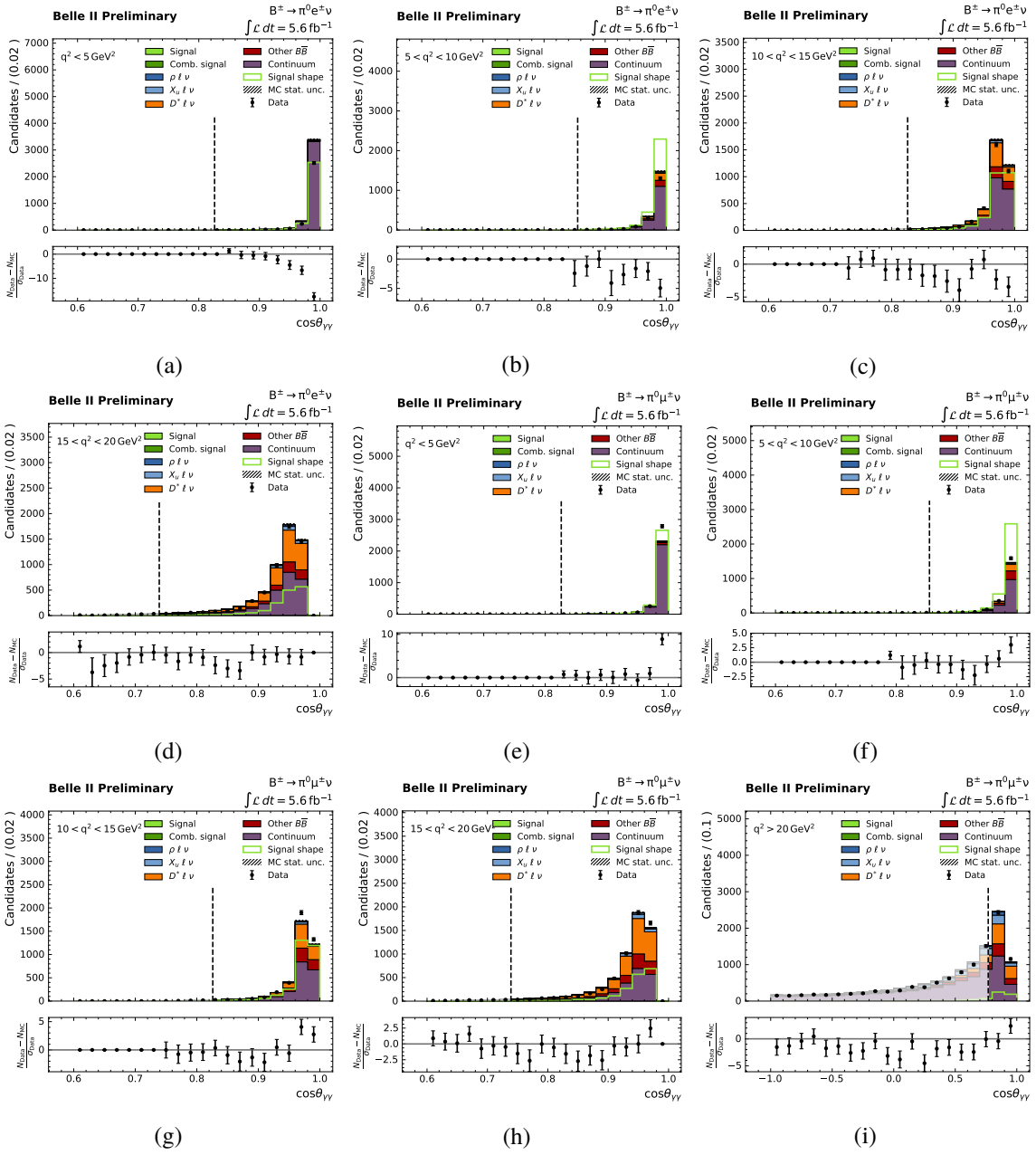
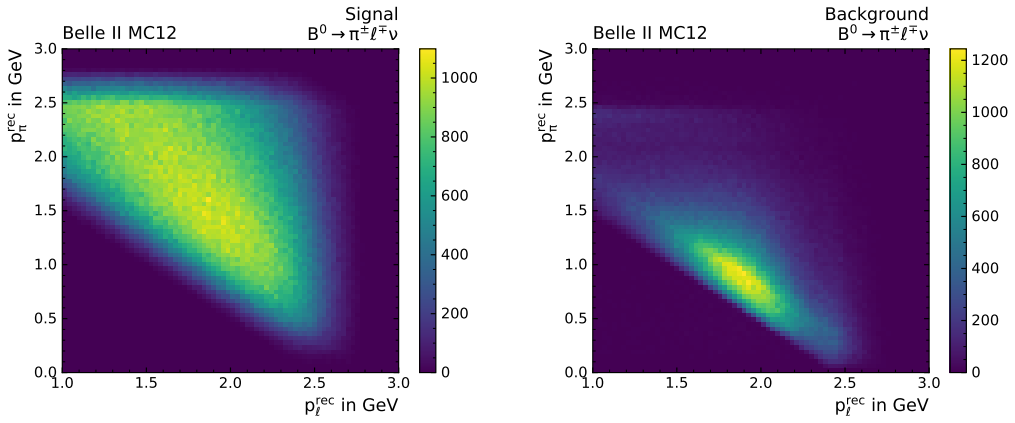


Figure B.10: Data-MC agreement for $\cos\theta_{\gamma\gamma}$ in the q^2 bins for the (a)-(d) $B^{\pm} \to \pi^0 e\nu$ and (e)-(i) $B^{\pm} \to \pi^0 \mu\nu$ modes.

B.4 Combined Momentum Selection



(a) (b)

Figure B.11: Distributions of p_ℓ versus p_π for (a) signal and (b) $B\overline{B}$ background from MC for the $B^0 \to \pi^\pm \ell \nu$ mode after all preceding cuts.

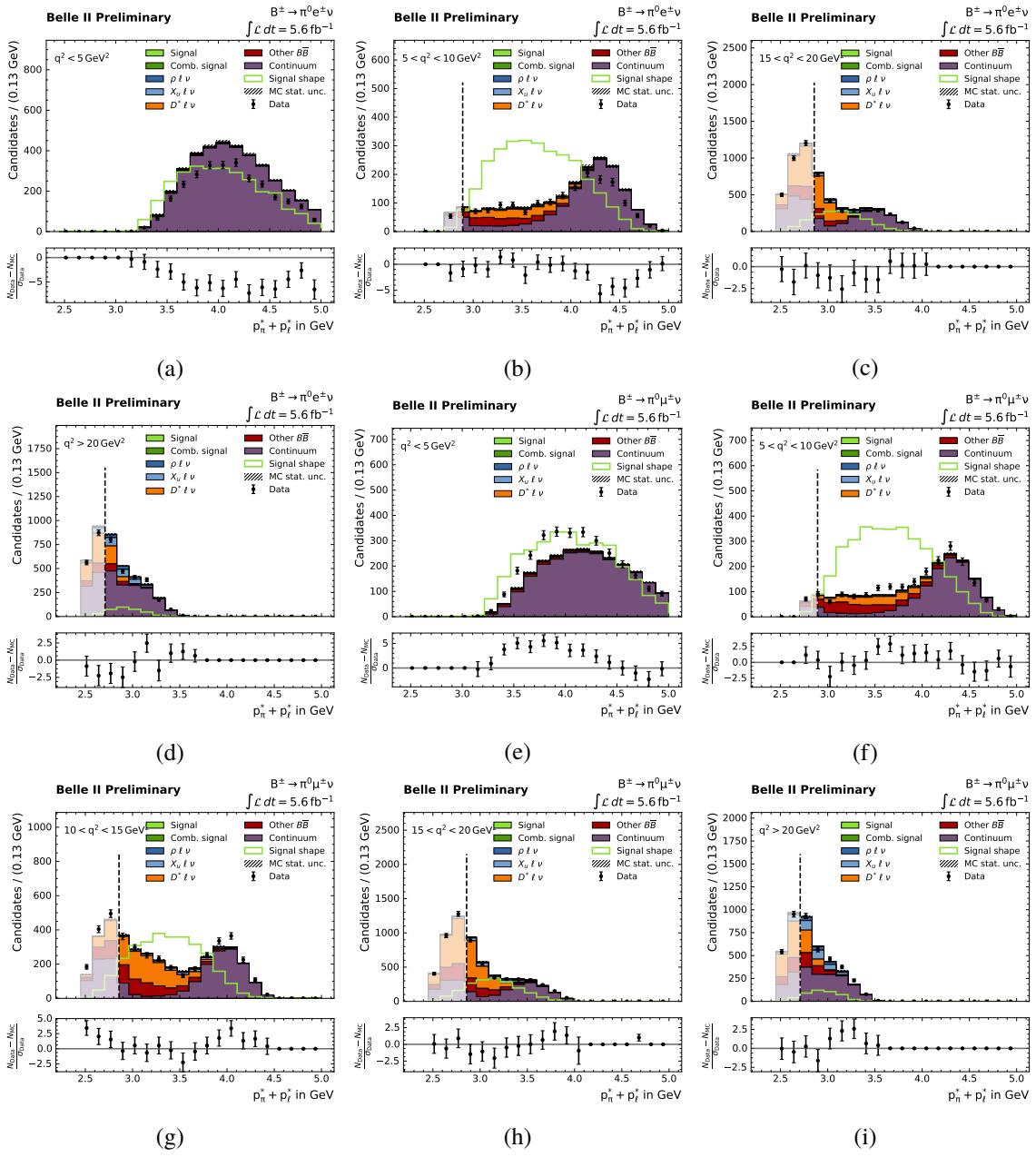


Figure B.12: Data-MC agreement for the sum of the lepton and pion momentum in the q^2 bins of the (a)-(d) $B^{\pm} \to \pi^0 e\nu$ and (e)-(i) $B^{\pm} \to \pi^0 \mu\nu$ modes.

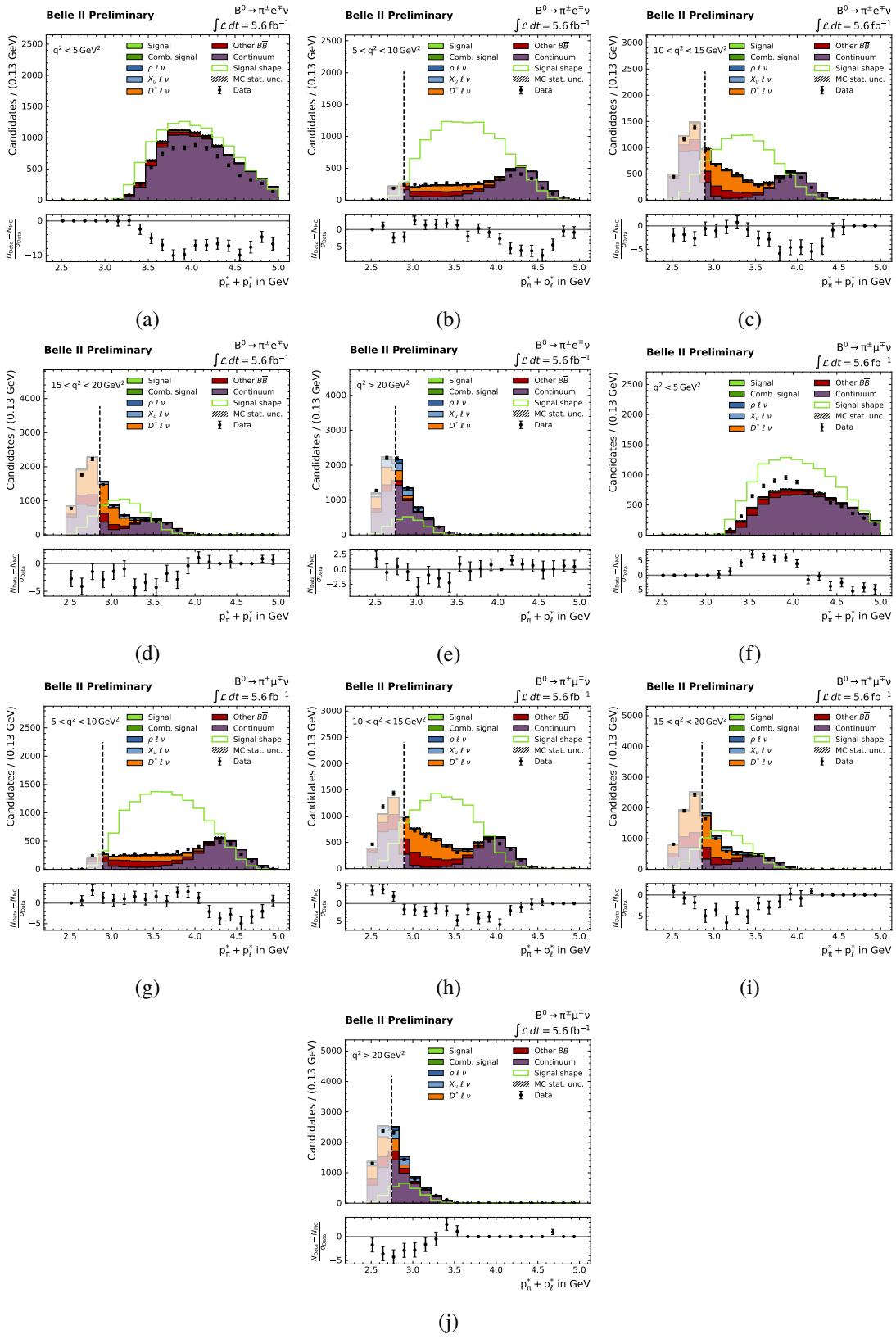


Figure B.13: Data-MC agreement for the sum of the lepton and pion momentum in the q^2 bins of the (a)-(e) $B^0 \to \pi^\pm e\nu$ and (f)-(j) $B^0 \to \pi^\pm \mu\nu$ modes.

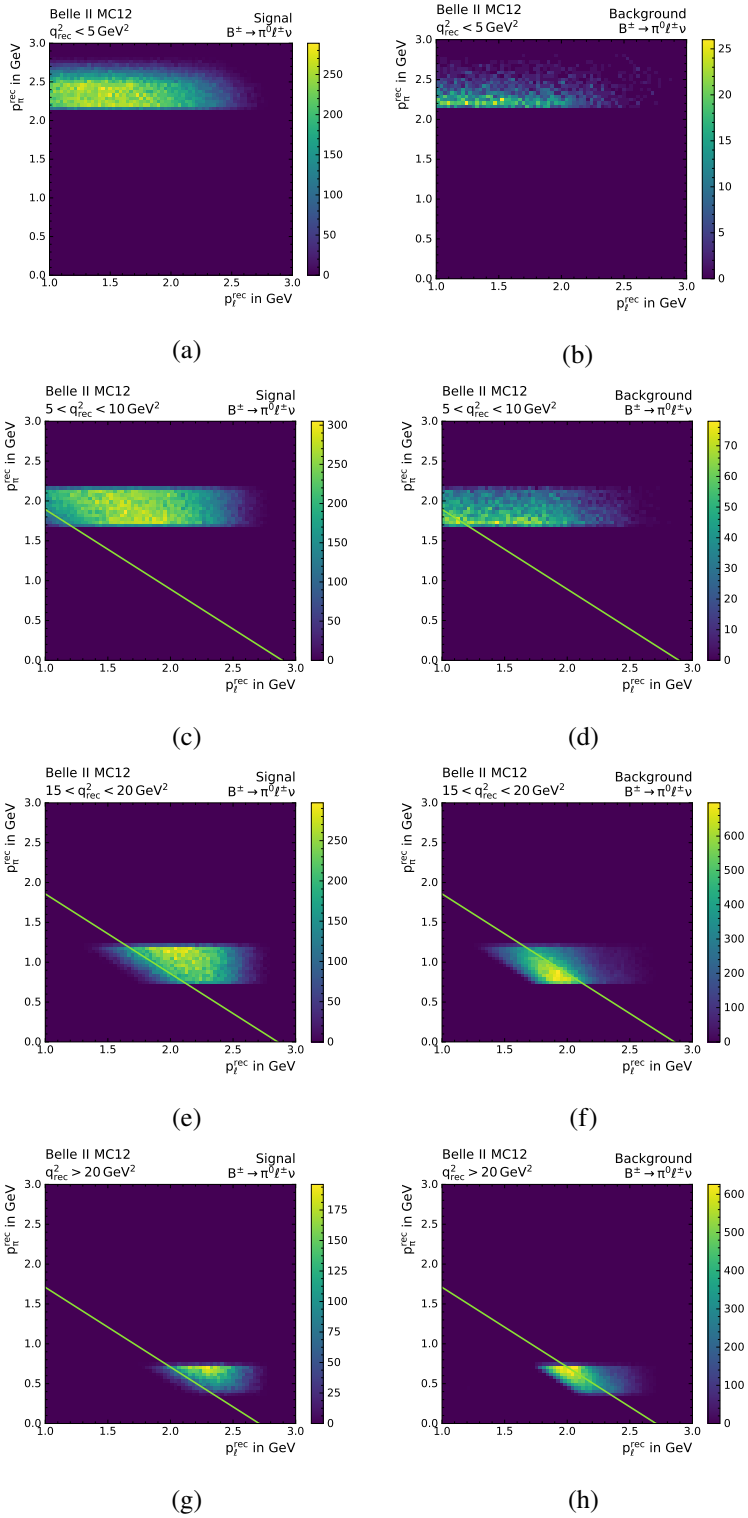


Figure B.14: 2D distributions of p_ℓ versus p_π for signal (left) and $B\overline{B}$ background (right) for the $B^\pm \to \pi^0 \ell \nu$ modes.

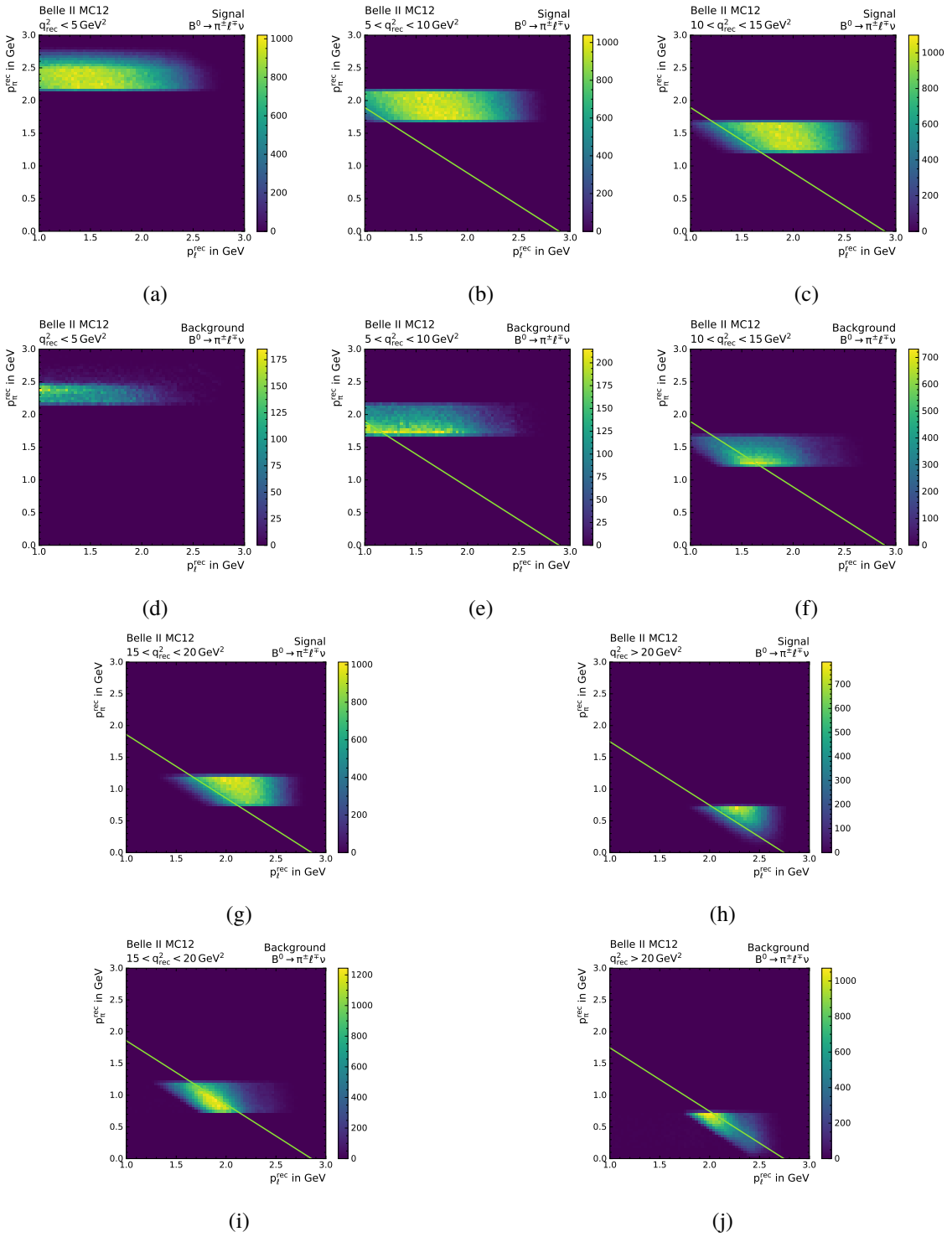


Figure B.15: 2D distributions of p_ℓ versus p_π for signal (top) and $B\overline{B}$ background (bottom) for the $B^0 \to \pi^\pm \ell \nu$ modes.

B.5 Data-MC Scaling

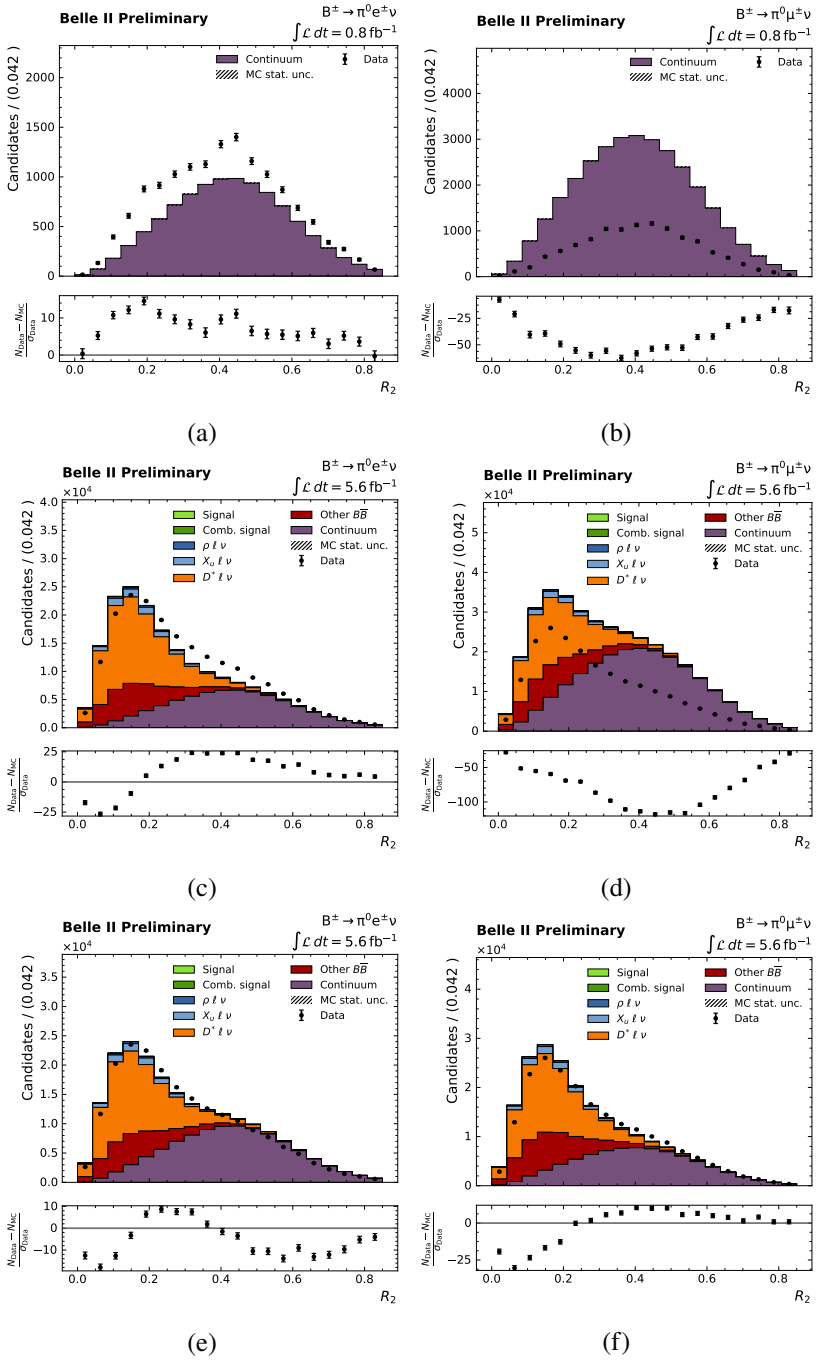
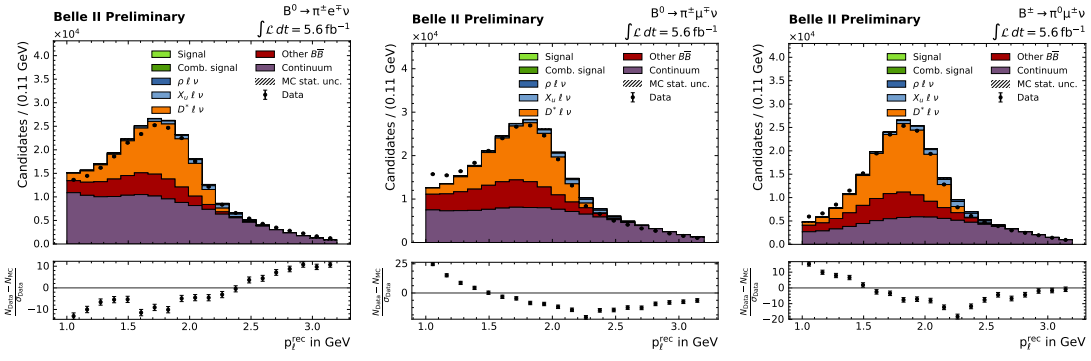


Figure B.16: Data-MC agreement in R_2 between off-resonance data and continuum MC (top) and data-MC agreement before (middle) and after(bottom) the scaling factors have been applied in the $B^{\pm} \to \pi^0 e\nu$ (left) and $B^{\pm} \to \pi^0 \mu\nu$ (right) modes.

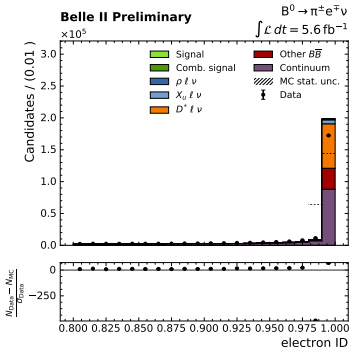
B.6 Selection Variables before nCleanedTracks Selection



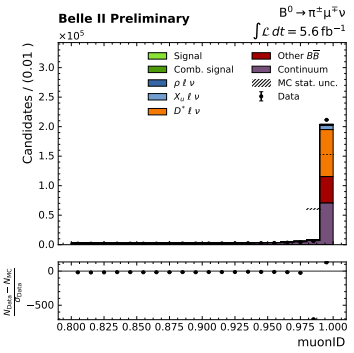
(a)

(b)

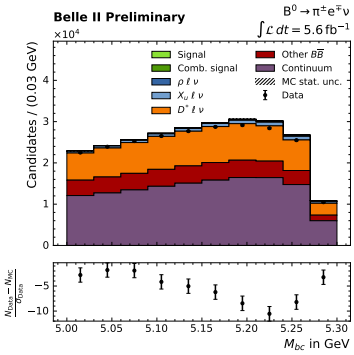
(c)



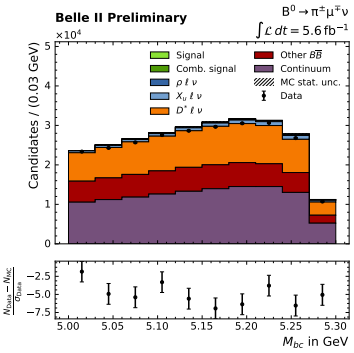
(d)



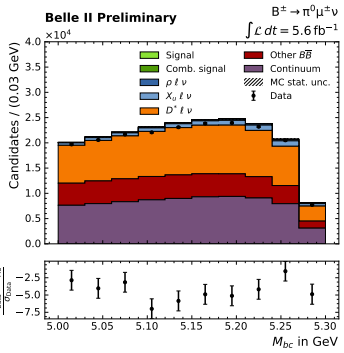
(e)



(f)



(g)



(h)

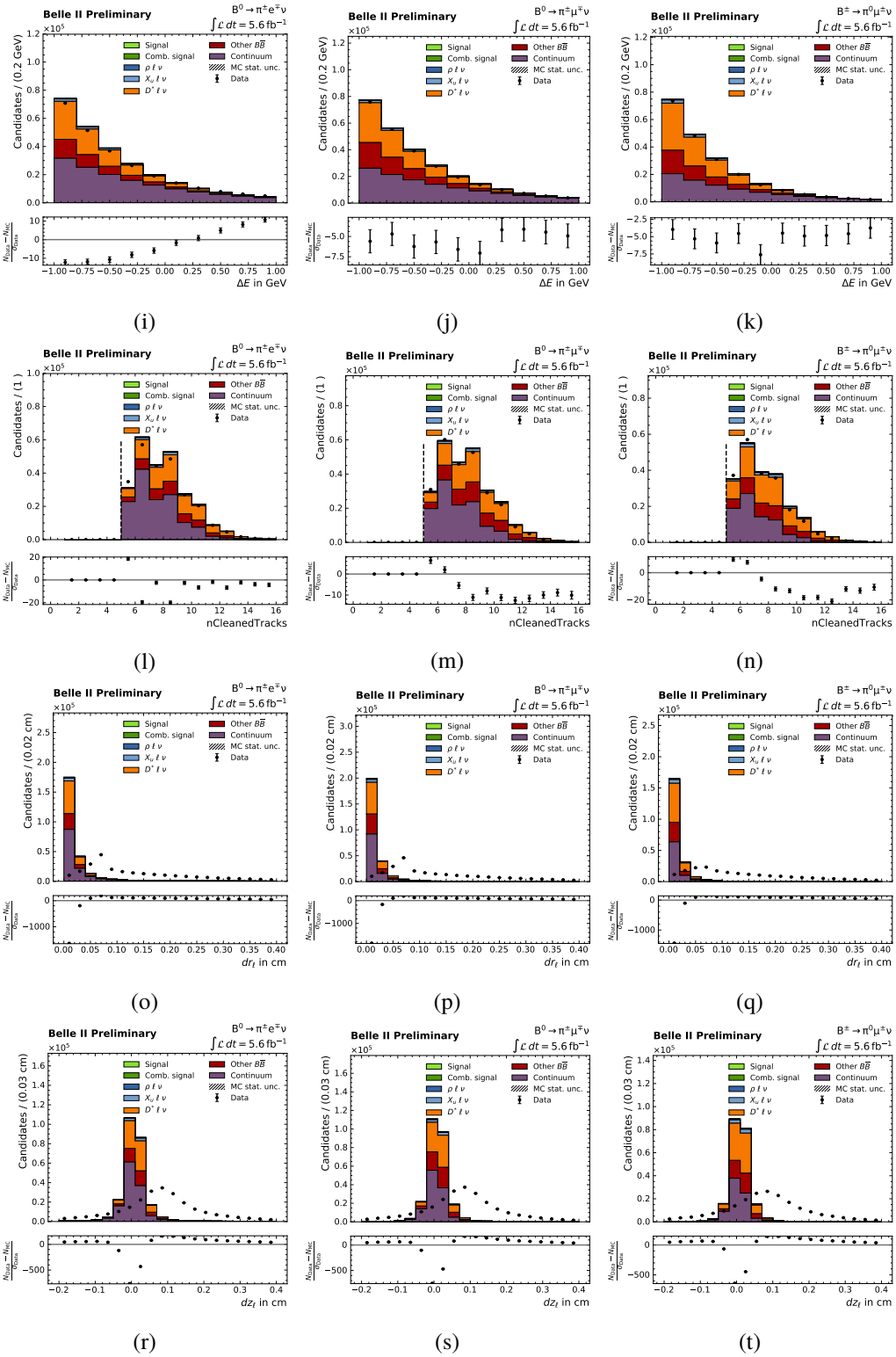
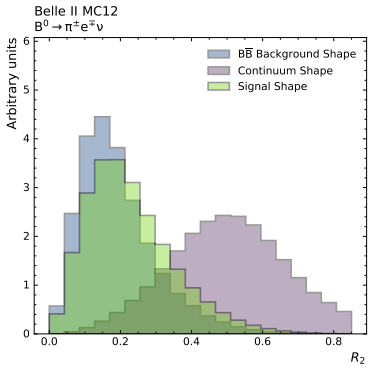


Figure B.17: Data MC agreement and pulls for (a)-(c) p_ℓ, (d) electron ID, (e) muon ID, (f)-(h) M_{bc}, (i)-(k) ΔE, (l)-(n) nCleanedTracks, (m)-(o) lepton dr and (p)-(r) lepton dz in the $B^0 \to \pi^\pm e\nu$ (left), $B^0 \to \pi\pm\mu\nu$ (middle) and $B^\pm \to \pi^0\mu\nu$ (right) modes.

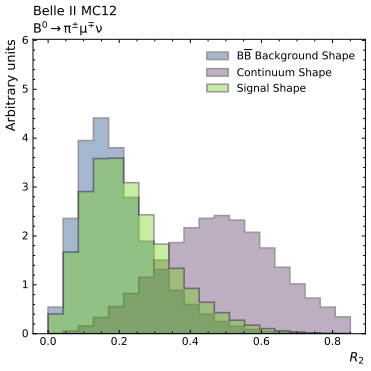
APPENDIX C

Boosted Decision Trees

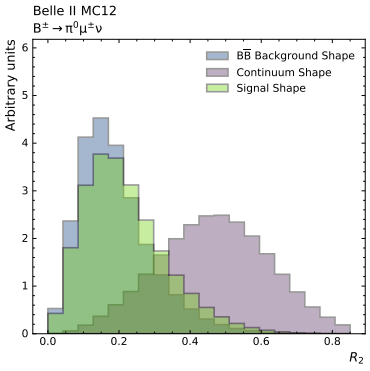
C.1 Continuum Suppression Input Variables



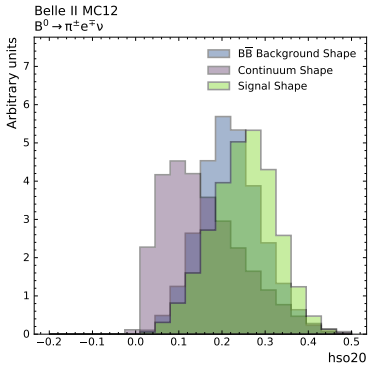
(a) R_2



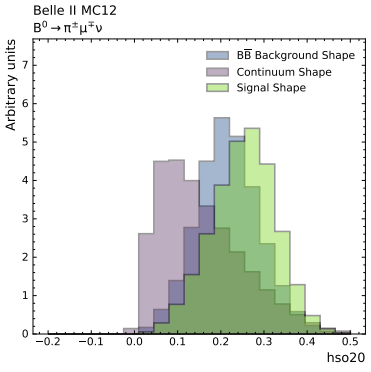
(b) R_2



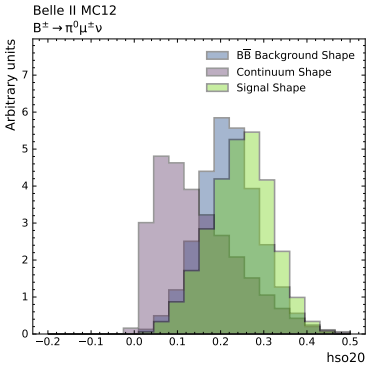
(c) R_2



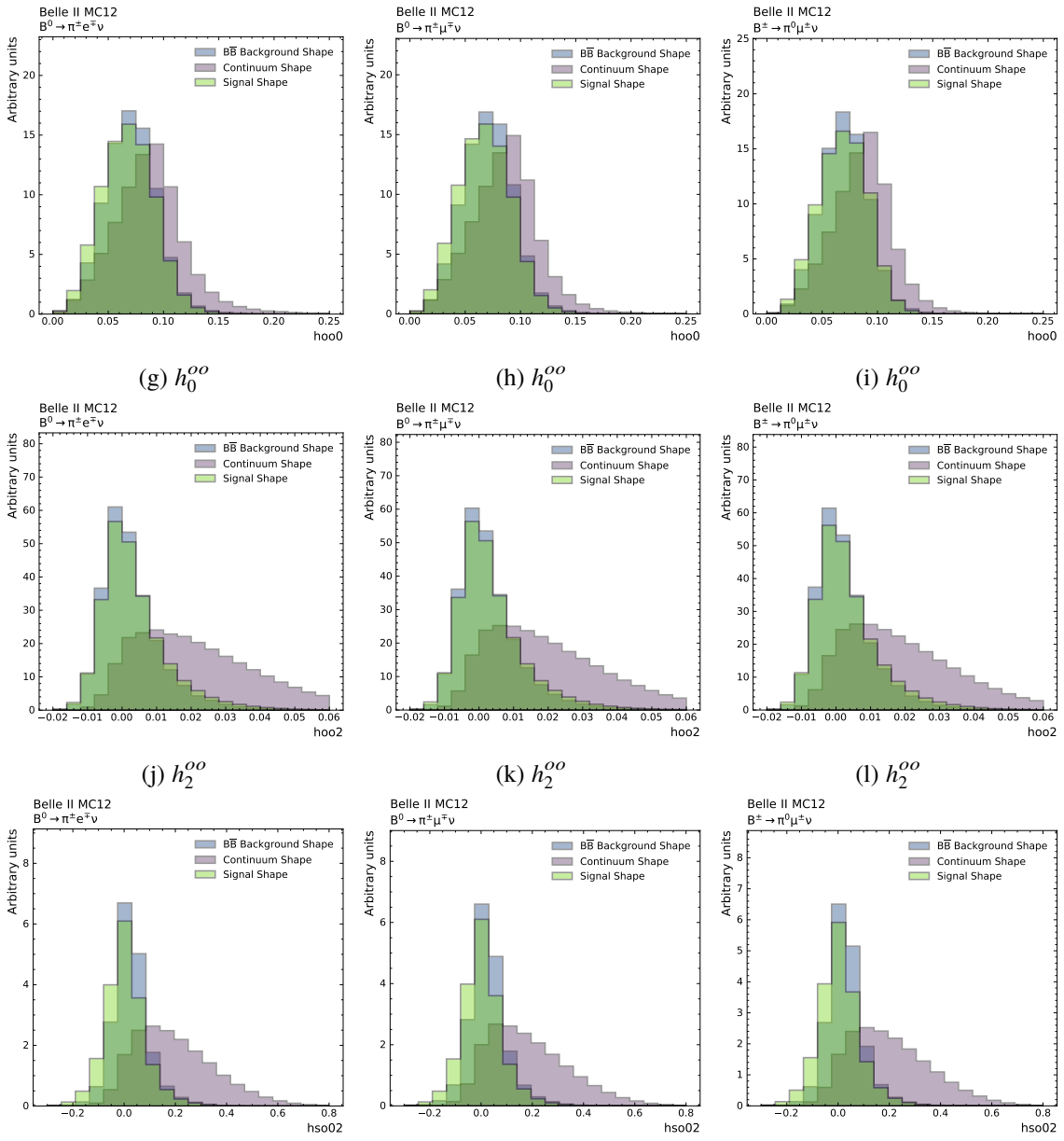
(d) h_{20}^{so}



(e) h_{20}^{so}



(f) h_{20}^{so}

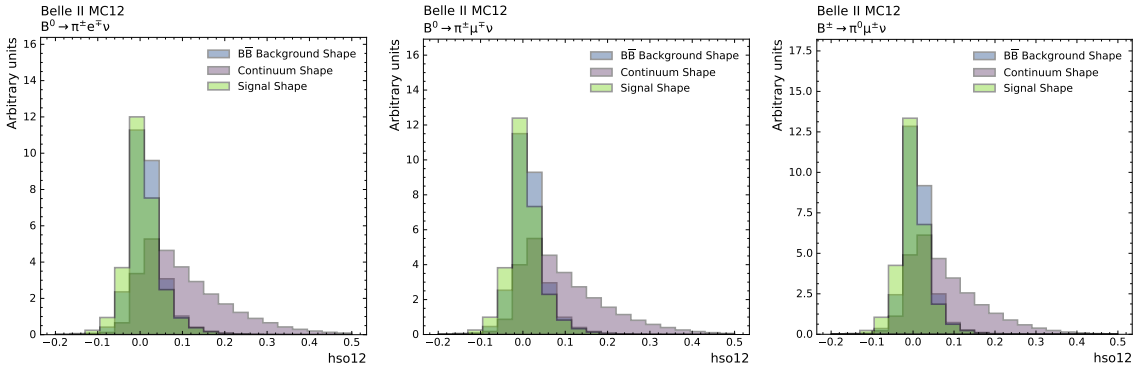


(g) h_0^{oo} (h) h_0^{oo} (i) h_0^{oo}

(j) h_2^{oo} (k) h_2^{oo} (l) h_2^{oo}

(m) h_2^{so} (n) h_2^{so} (o) h_2^{so}

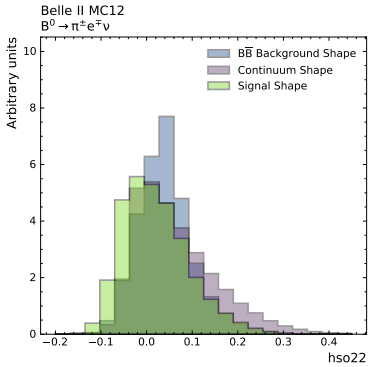
Figure C.1: Normalised distributions of the continuum suppression BDT input variables for signal in green, continuum background in purple and $B\overline{B}$ background in blue for the $B^0 \to \pi^{\pm} e\nu$ (left), $B^0 \to \pi\pm\mu\nu$ (middle) and $B^{\pm} \to \pi^0\mu\nu$ (right) modes.



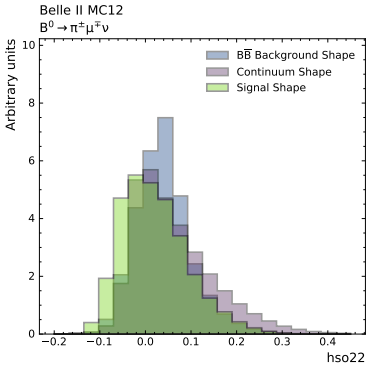
(a) h_{12}^{so}

(b) h_{12}^{so}

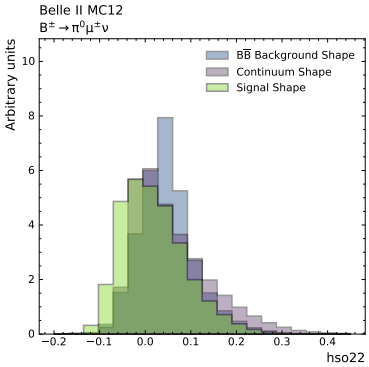
(c) h_{12}^{so}



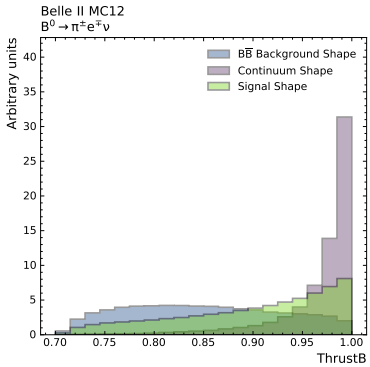
(d) h_{22}^{so}



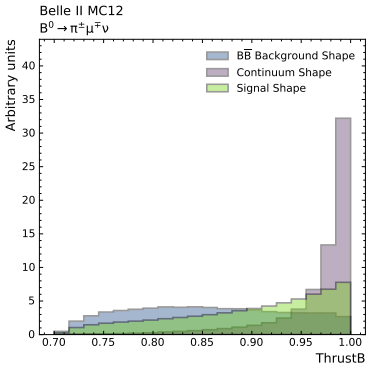
(e) h_{22}^{so}



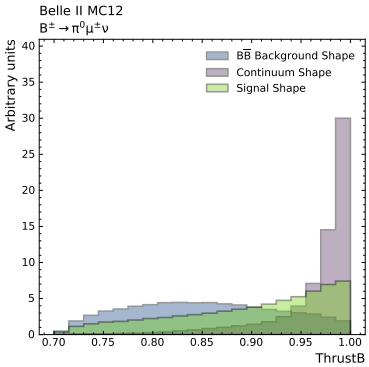
(f) h_{22}^{so}



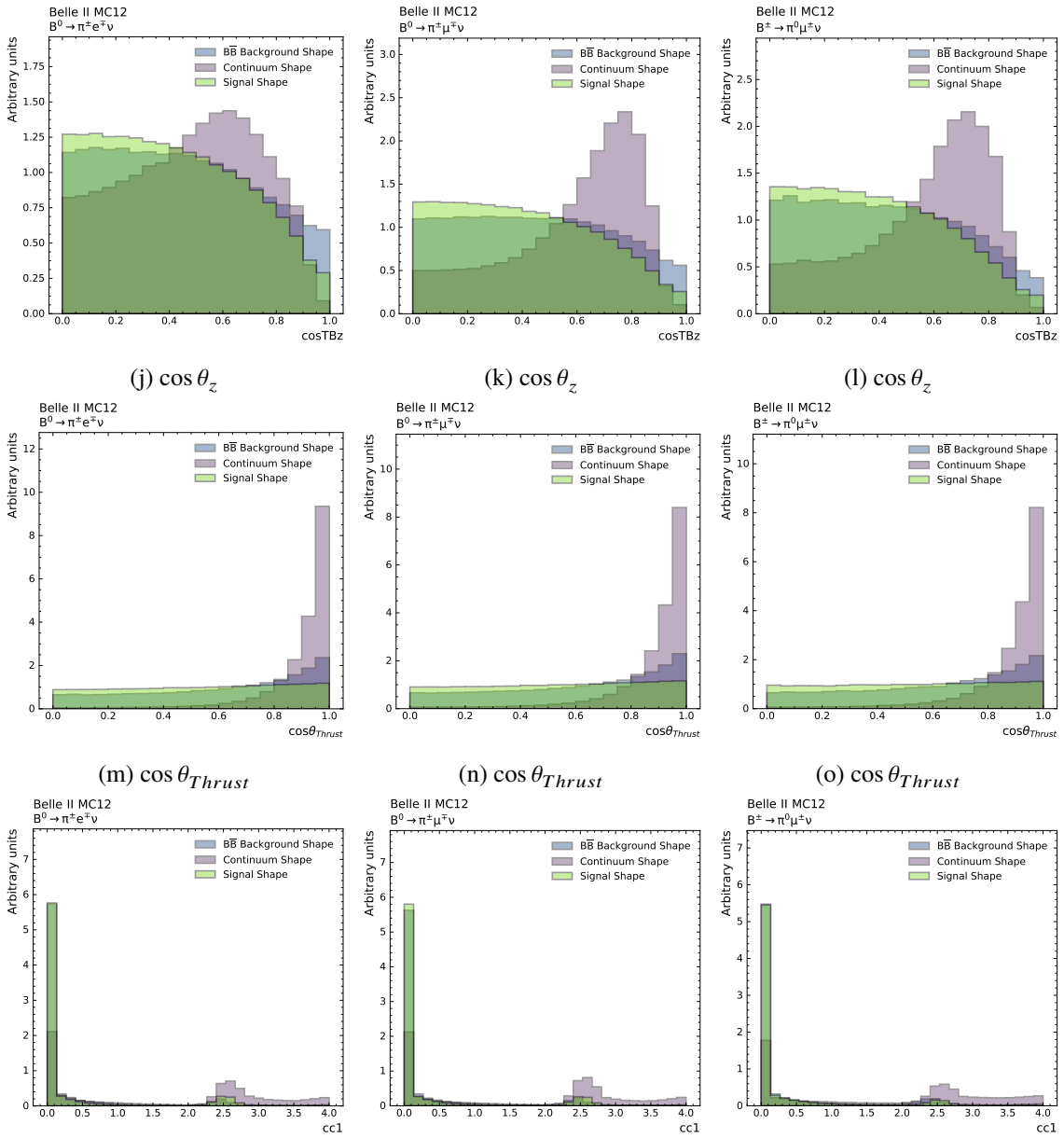
(g) Thrust



(h) Thrust



(i) Thrust

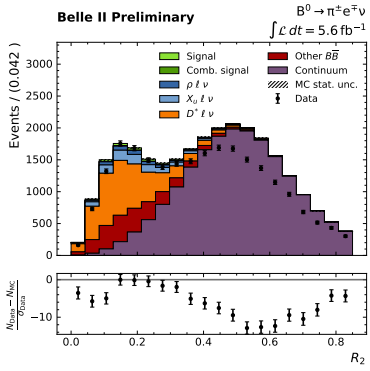


(j) $\cos\theta_z$ (k) $\cos\theta_z$ (l) $\cos\theta_z$

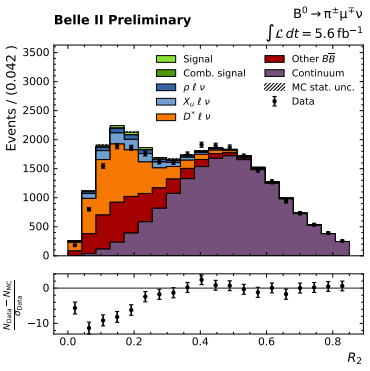
(m) $\cos\theta_{Thrust}$ (n) $\cos\theta_{Thrust}$ (o) $\cos\theta_{Thrust}$

(p) CLEO Cone 1 (q) CLEO Cone 1 (r) CLEO Cone 1

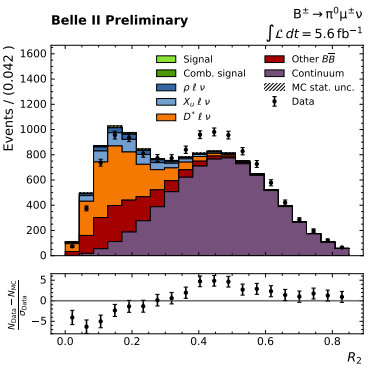
Figure C.2: Continued: Normalised distributions of the continuum suppression BDT input variables for signal in green, continuum background in purple and $B\overline{B}$ background in blue for the $B^0 \to \pi^{\pm} e\nu$ (left), $B^0 \to \pi\pm\mu\nu$ (middle) and $B^{\pm} \to \pi^0\mu\nu$ (right) modes.



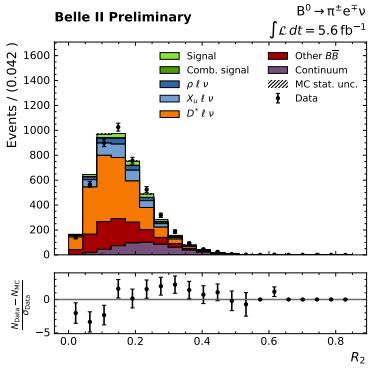
(a) R_2



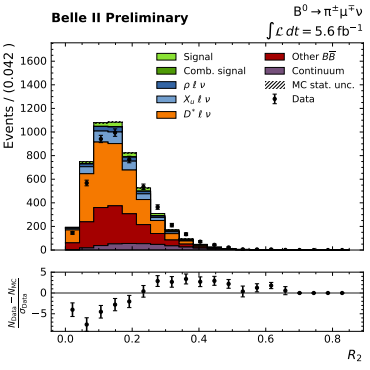
(b) R_2



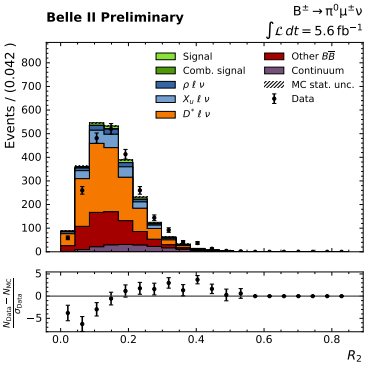
(c) R_2



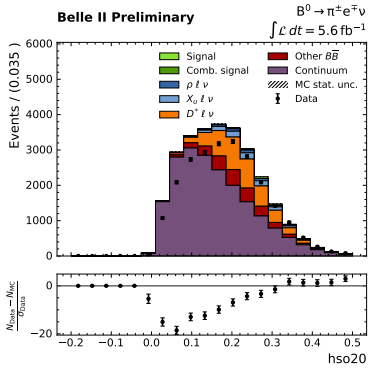
(d) R_2



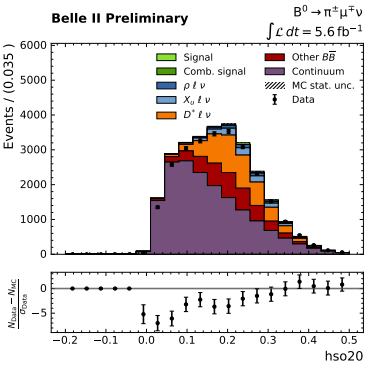
(e) R_2



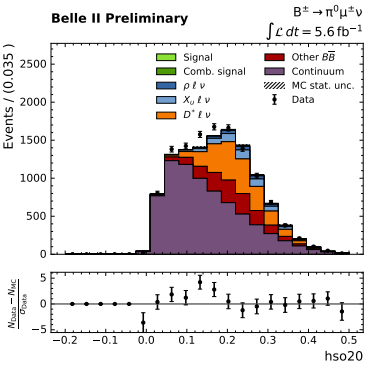
(f) R_2



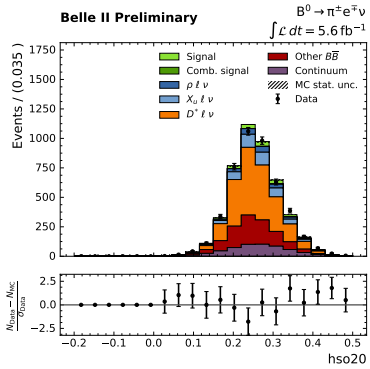
(g) h_{20}^{so}



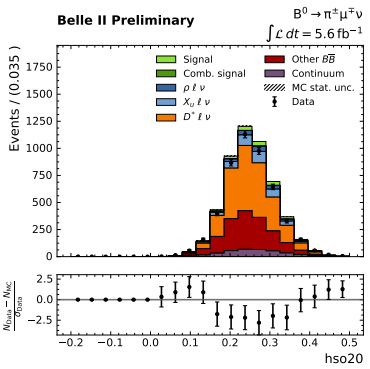
(h) h_{20}^{so}



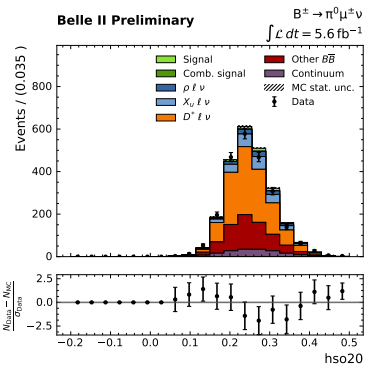
(i) h_{20}^{so}



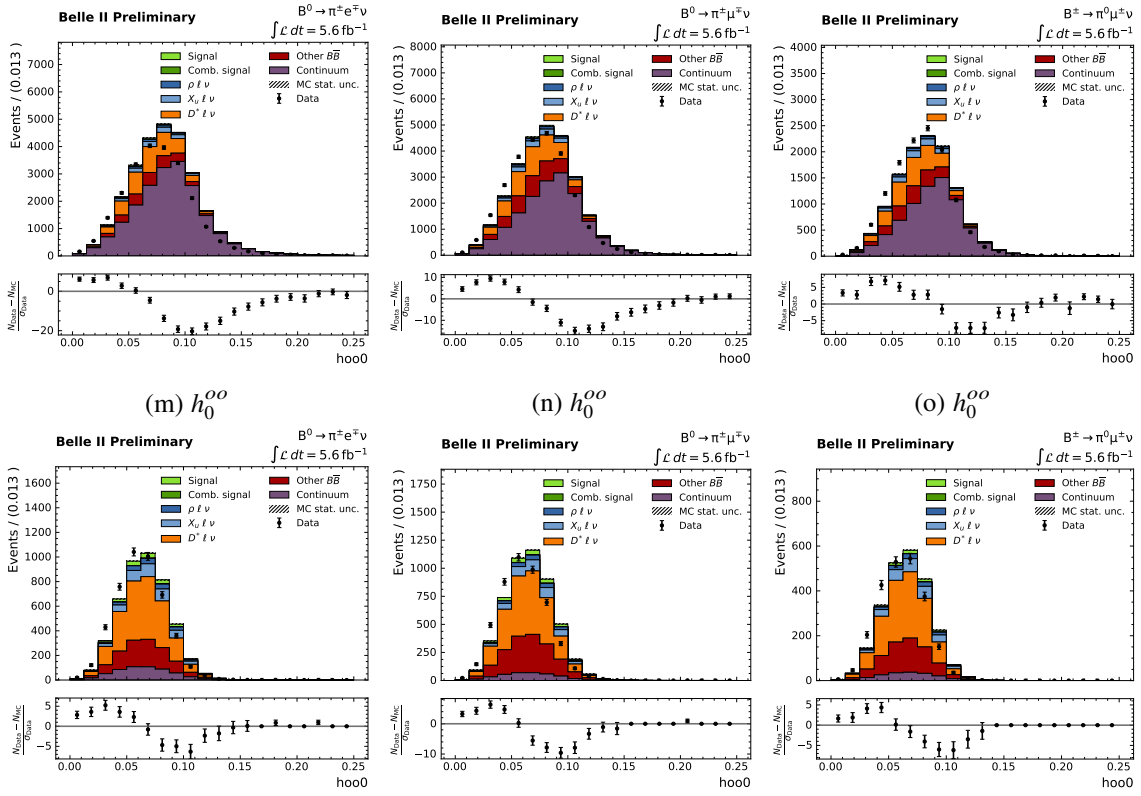
(j) h_{20}^{so}



(k) h_{20}^{so}



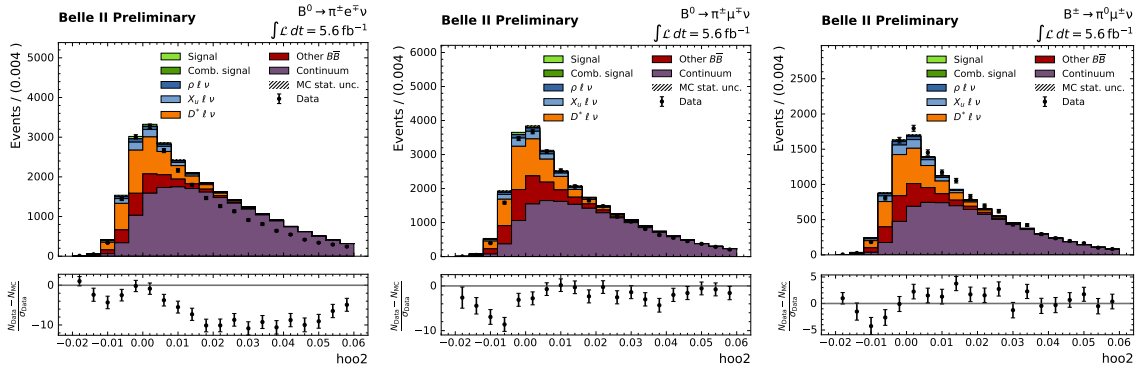
(l) h_{20}^{so}



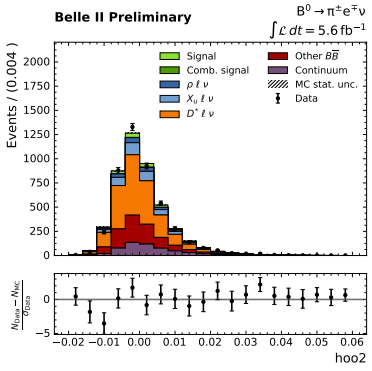
(m) h_0^{oo} (n) h_0^{oo} (o) h_0^{oo}

(p) h_0^{oo} (q) h_0^{oo} (r) h_0^{oo}

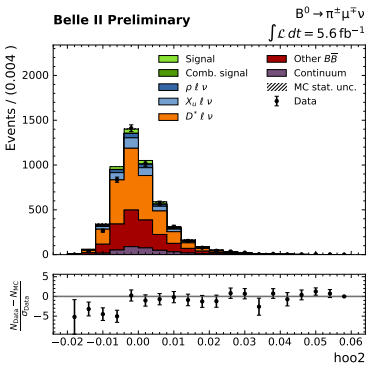
Figure C.3: Data-MC agreement for the continuum suppression BDT input variables integrated over all q^2 bins before (top) and after (below) a selection on the continuum suppression output classifier for the $B^0 \to \pi^\pm e\nu$ (left), $B^0 \to \pi\pm\mu\nu$ (middle) and $B^\pm \to \pi^0\mu\nu$ (right) mode has been placed.



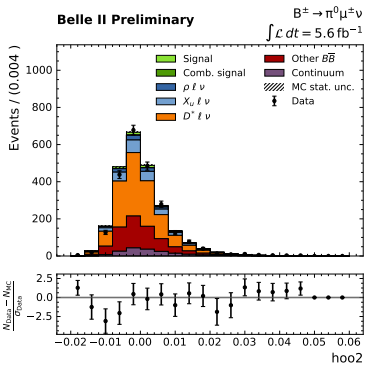
(a) h_2^{oo} (b) h_2^{oo} (c) h_2^{oo}



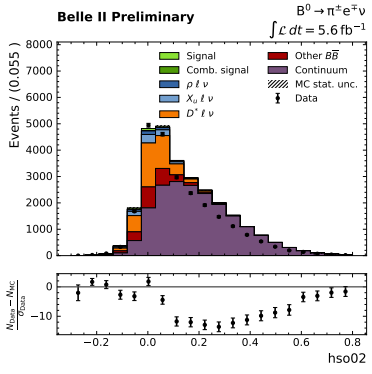
(d) h_2^{oo}



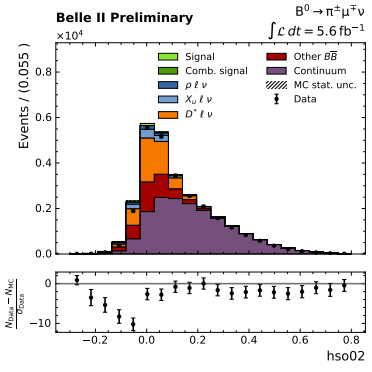
(e) h_2^{oo}



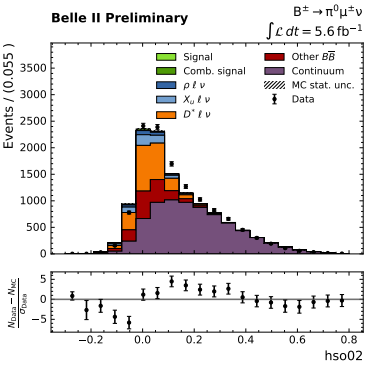
(f) h_2^{oo}



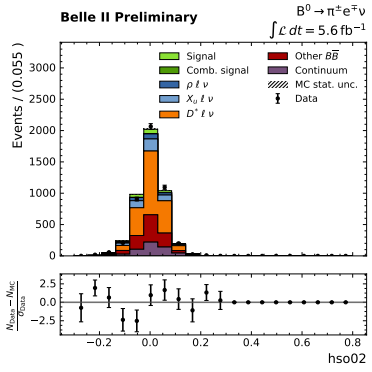
(g) h_2^{so}



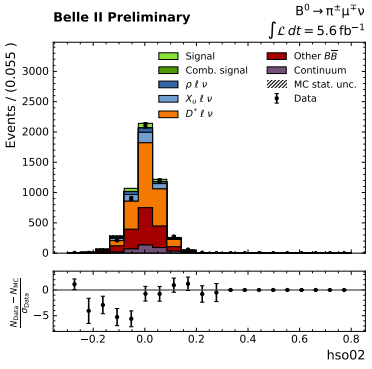
(h) h_2^{so}



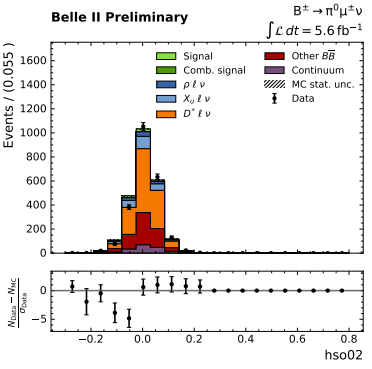
(i) h_2^{so}



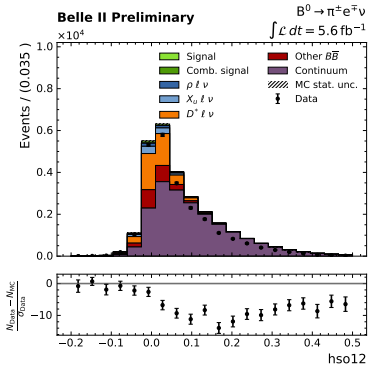
(j) h_2^{so}



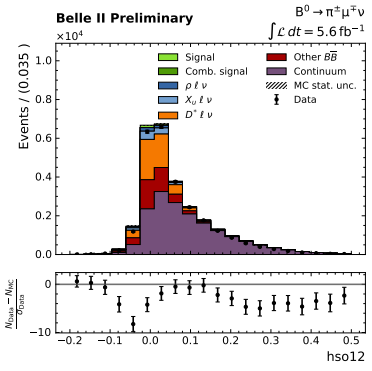
(k) h_2^{so}



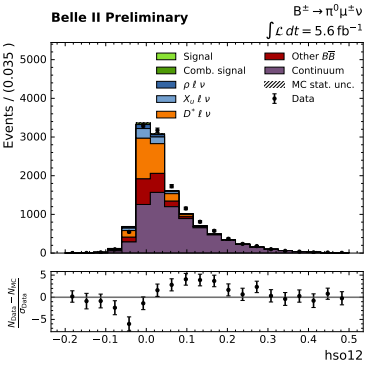
(l) h_2^{so}



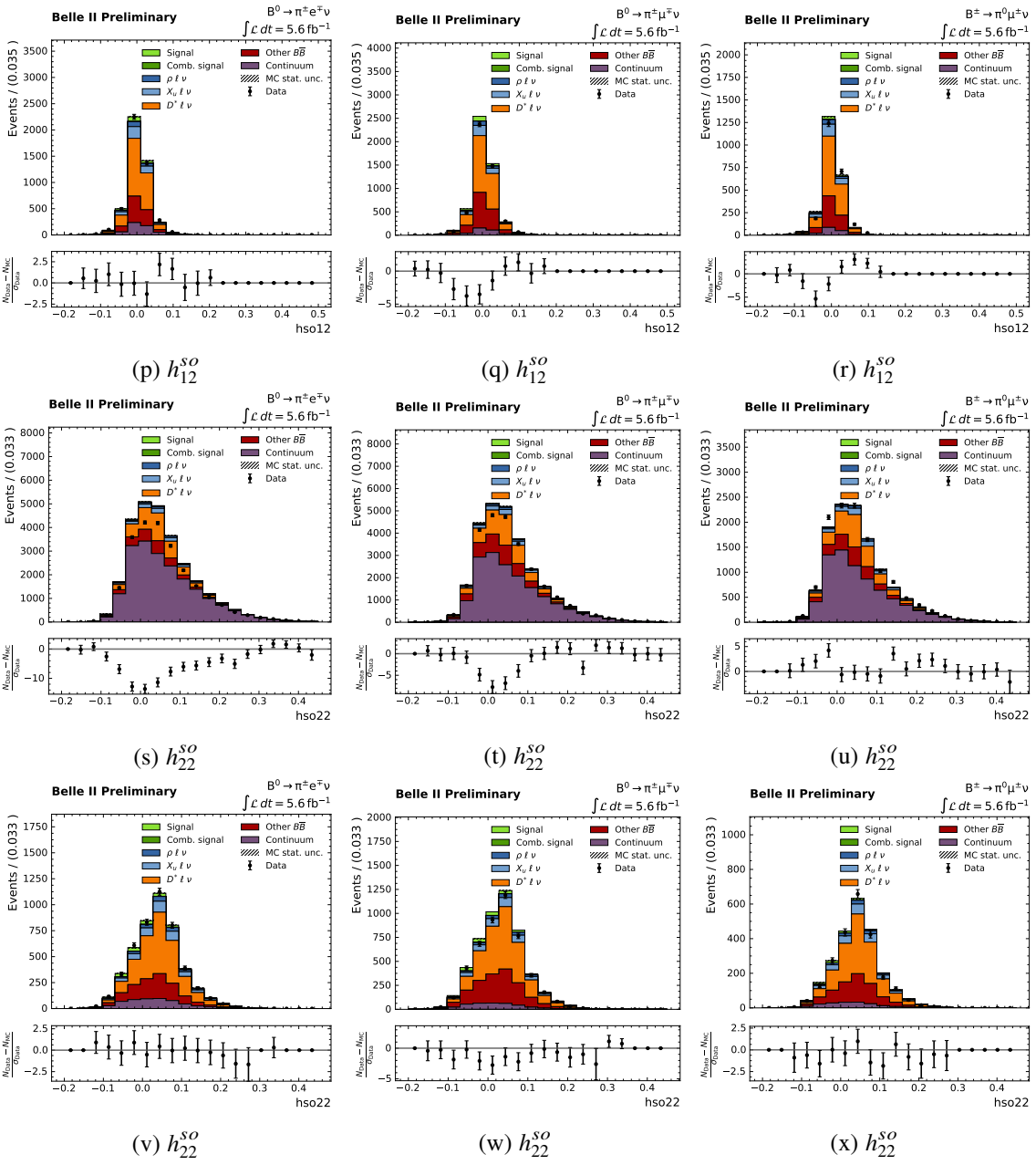
(m) h_{12}^{so}



(n) h_{12}^{so}



(o) h_{12}^{so}

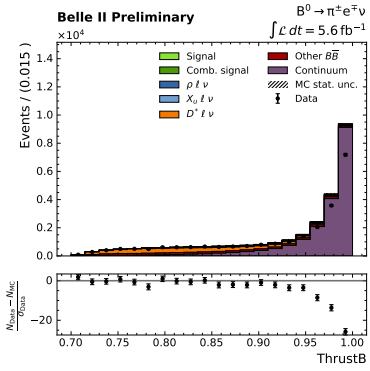


(p) h_{12}^{so} (q) h_{12}^{so} (r) h_{12}^{so}

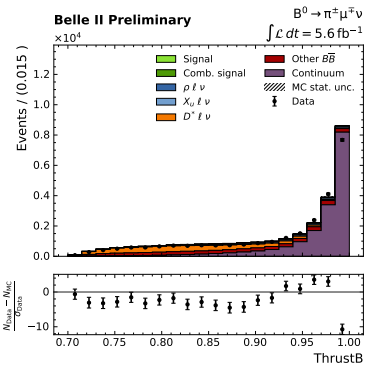
(s) h_{22}^{so} (t) h_{22}^{so} (u) h_{22}^{so}

(v) h_{22}^{so} (w) h_{22}^{so} (x) h_{22}^{so}

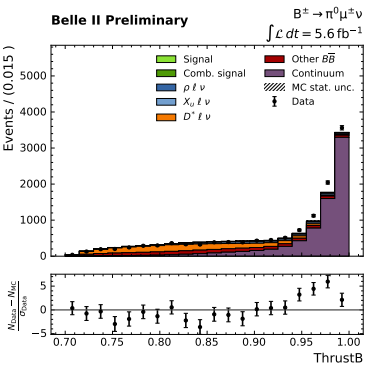
Figure C.4: Continued: Data-MC agreement for the continuum suppression BDT input variables integrated over all q^2 bins before (top) and after (below) a selection on the continuum suppression output classifier for the $B^0 \to \pi^\pm e\nu$ (left), $B^0 \to \pi\pm\mu\nu$ (middle) and $B^\pm \to \pi^0\mu\nu$ (right) mode has been placed.



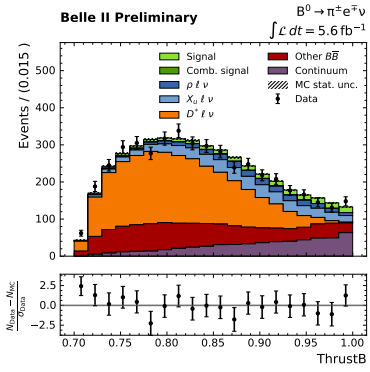
(a) Thrust



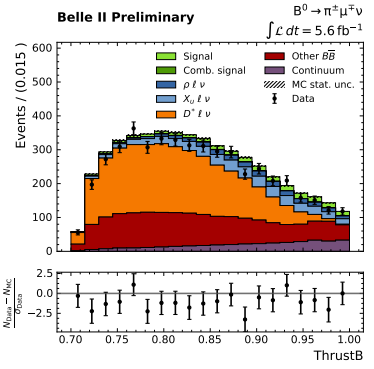
(b) Thrust



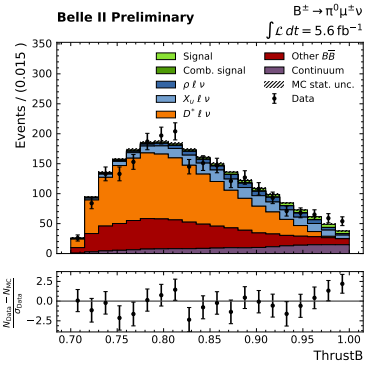
(c) Thrust



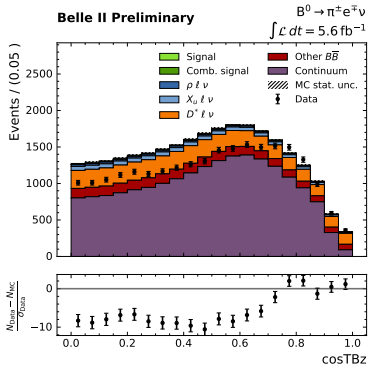
(d) Thrust



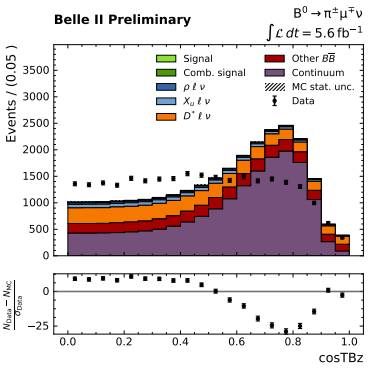
(e) Thrust



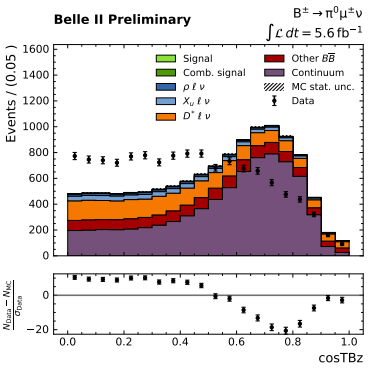
(f) Thrust



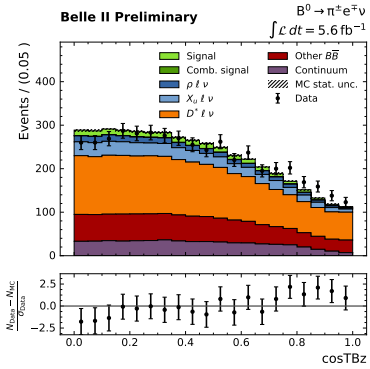
(g) $\cos\theta_z$



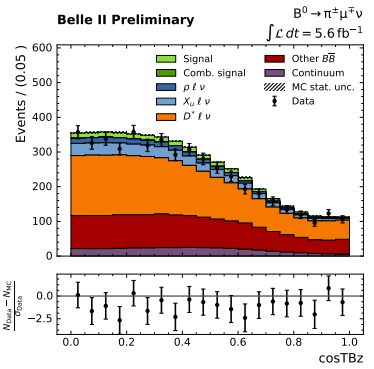
(h) $\cos\theta_z$



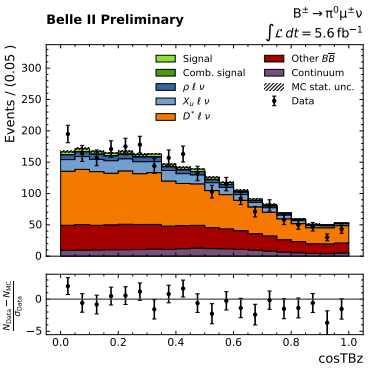
(i) $\cos\theta_z$



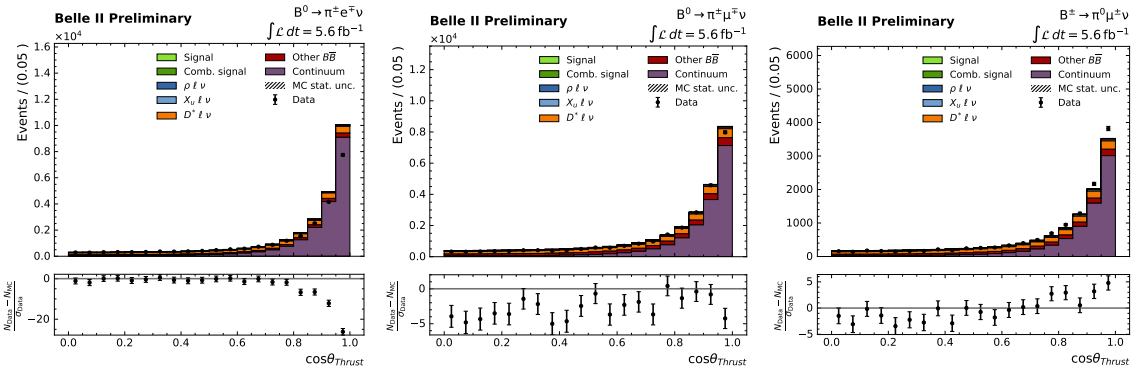
(j) $\cos\theta_z$



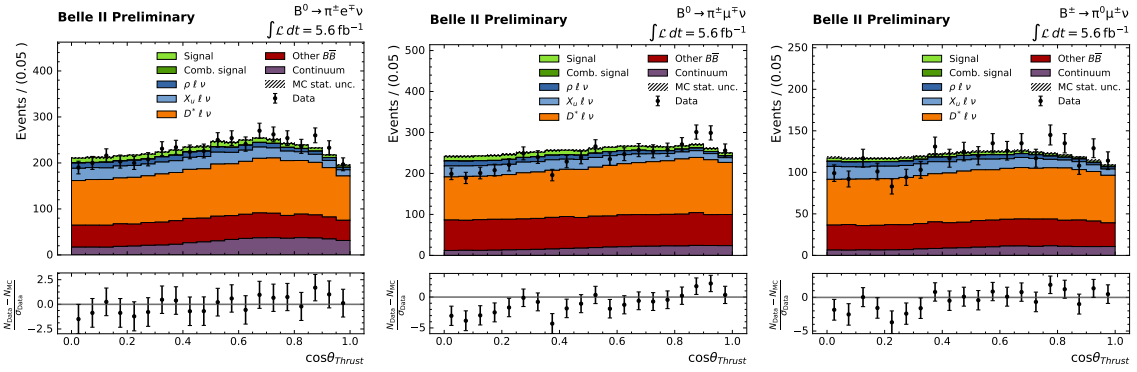
(k) $\cos\theta_z$



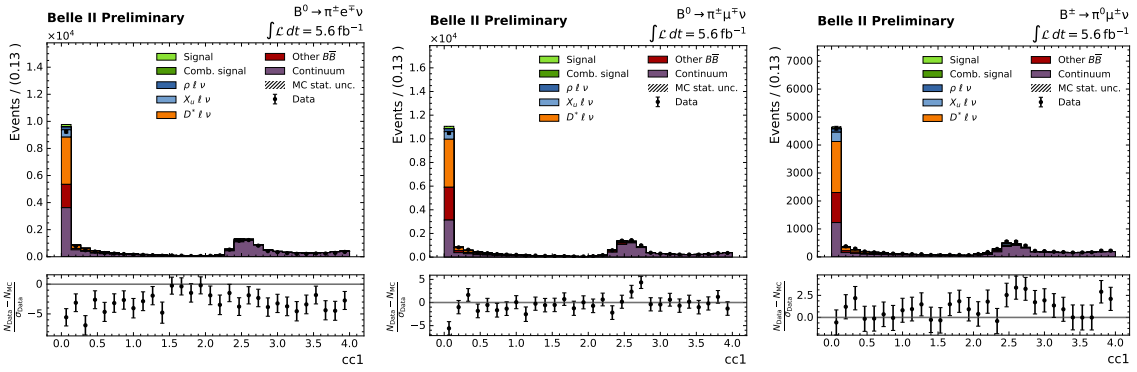
(l) $\cos\theta_z$



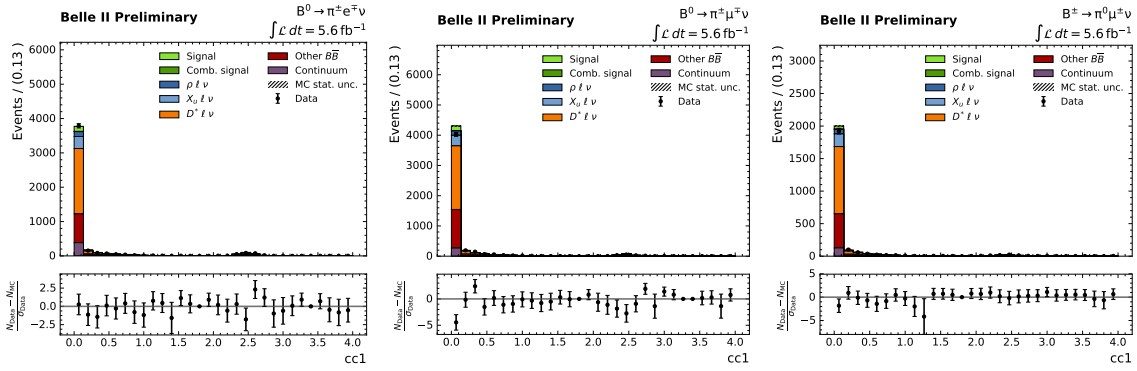
(m) $\cos\theta_{Thrust}$ (n) $\cos\theta_{Thrust}$ (o) $\cos\theta_{Thrust}$



(p) $\cos\theta_{Thrust}$ (q) $\cos\theta_{Thrust}$ (r) $\cos\theta_{Thrust}$



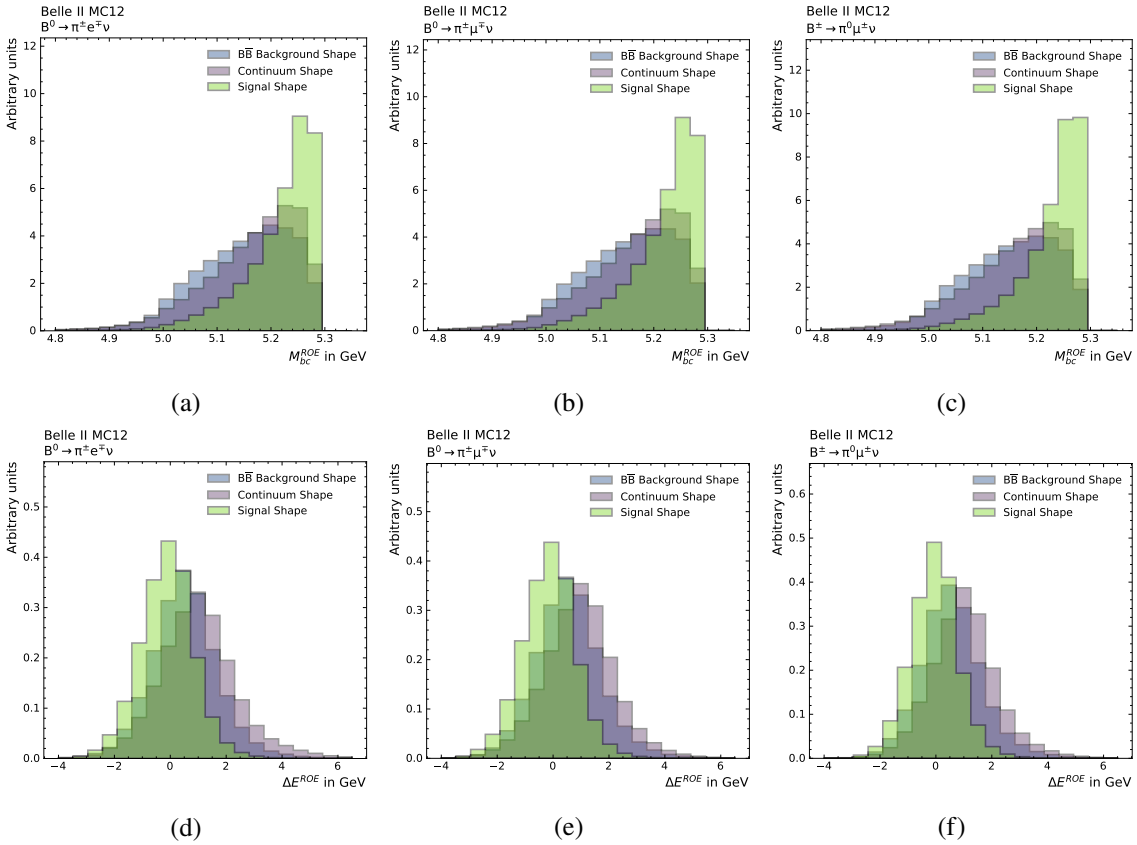
(s) CLEO Cone 1 (t) CLEO Cone 1 (u) CLEO Cone 1



(v) CLEO Cone 1 (w) CLEO Cone 1 (x) CLEO Cone 1

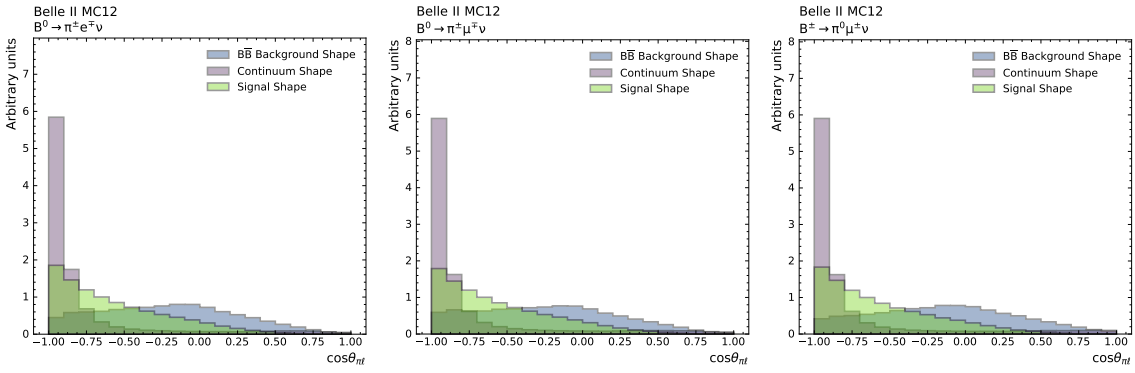
Figure C.5: Continued: Data-MC agreement for the continuum suppression BDT input variables integrated over all q^2 bins before (top) and after (below) a selection on the continuum suppression output classifier for the $B^0 \to \pi^\pm e\nu$ (left), $B^0 \to \pi\pm\mu\nu$ (middle) and $B^\pm \to \pi^0\mu\nu$ (right) mode has been placed.

C.2 BB̄ Suppression Input Variables

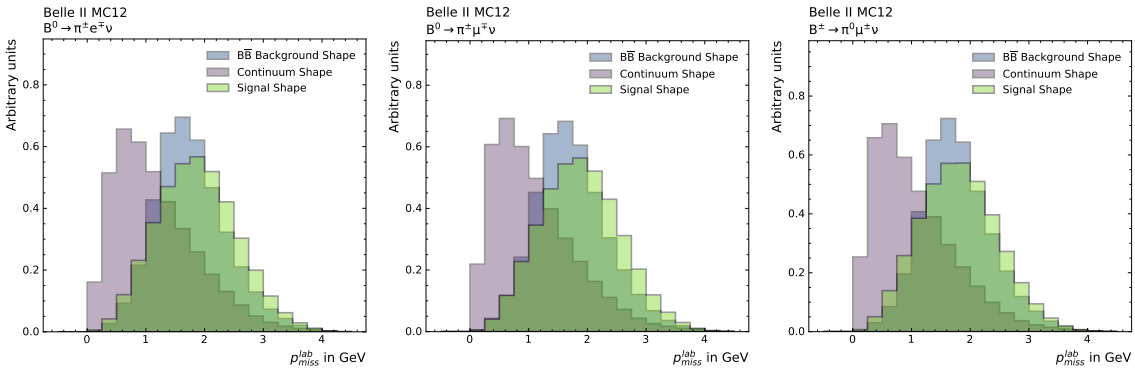


(a) (b) (c)

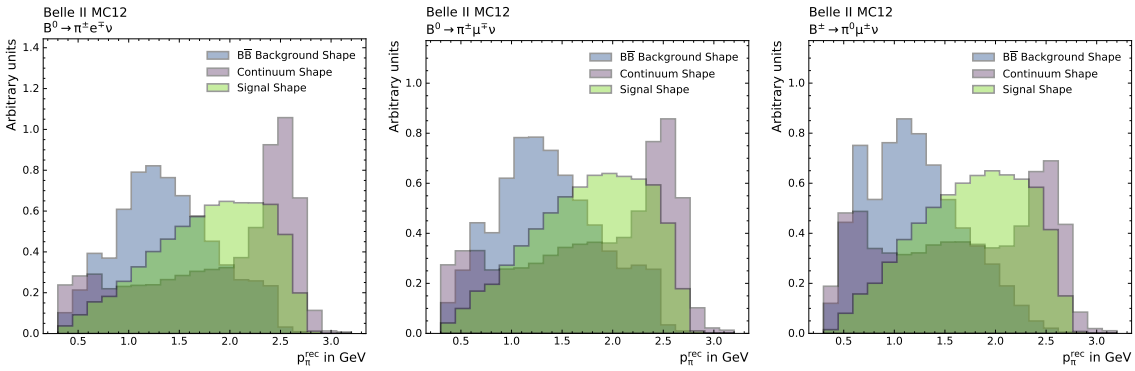
(d) (e) (f)



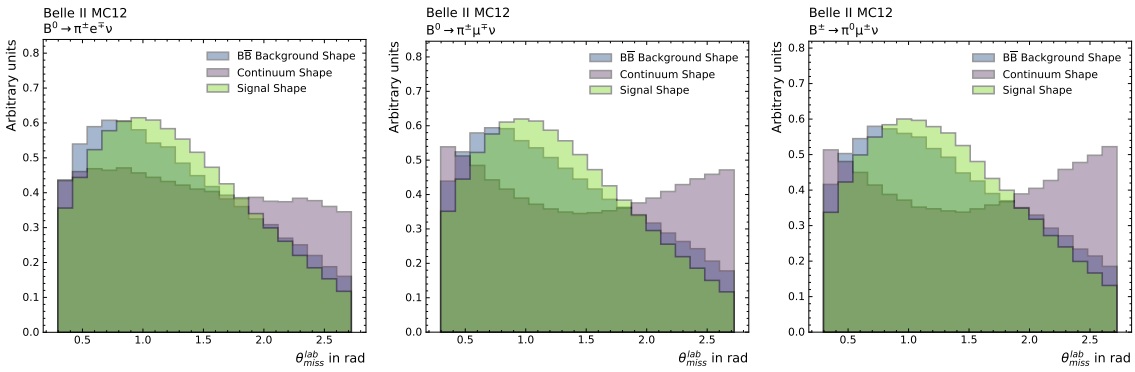
(g) h_0^{oo} (h) h_0^{oo} (i) h_0^{oo}



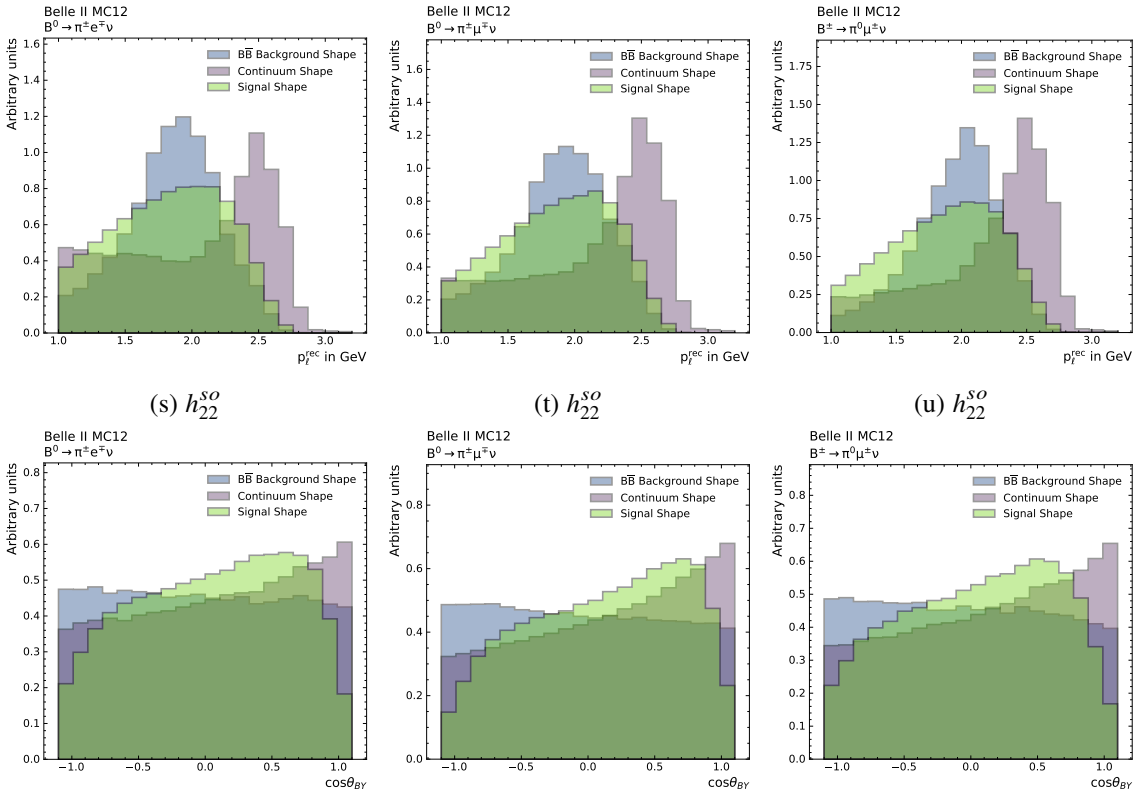
(j) h_2^{oo} (k) h_2^{oo} (l) h_2^{oo}



(m) h_2^{so} (n) h_2^{so} (o) h_2^{so}



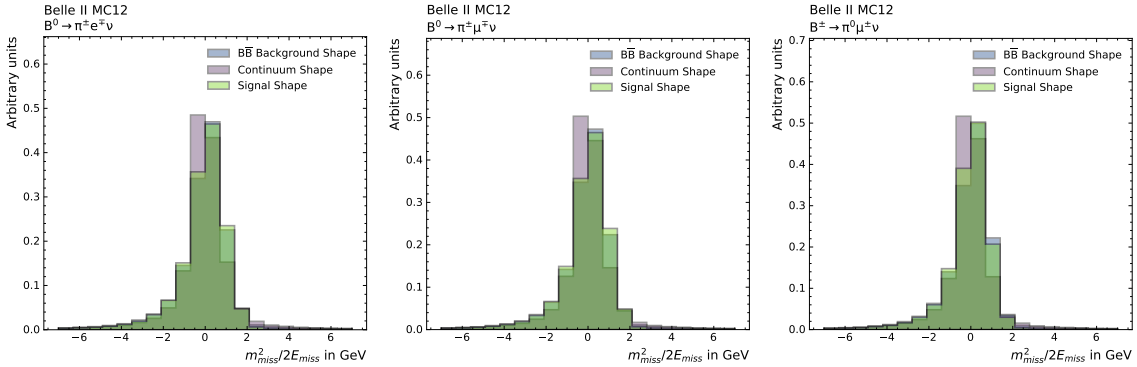
(p) h_{12}^{so} (q) h_{12}^{so} (r) h_{12}^{so}



(s) h_{22}^{so} (t) h_{22}^{so} (u) h_{22}^{so}

(v) Thrust (w) Thrust (x) Thrust

Figure C.6: Normalised distributions of the B$\overline{\text{B}}$ suppression BDT input variables for signal in green, continuum background in purple and B$\overline{\text{B}}$ background in blue for the $B^0 \to \pi^{\pm} e\nu$ (left), $B^0 \to \pi\pm\mu\nu$ (middle) and $B^{\pm} \to \pi^0 \mu\nu$ (right) modes.



(a) $\cos\theta_z$ (b) $\cos\theta_z$ (c) $\cos\theta_z$

Figure C.7: Continued: Normalised distributions of the B$\overline{\text{B}}$ suppression BDT input variables for signal in green, continuum background in purple and B$\overline{\text{B}}$ background in blue for the $B^0 \to \pi^{\pm} e\nu$ (left), $B^0 \to \pi\pm\mu\nu$ (middle) and $B^{\pm} \to \pi^0 \mu\nu$ (right) modes.

C.3 Grid Search

Mode	q^2 Bin	BDT Parameters				
		nTrees	nLevels	nCuts	Shrinkage	Ratio
$B^0 \to \pi^\pm e\nu$	$lowq$	200	5	6	0.15	0.6
	$medq$	200	4	6	0.15	0.9
	$highq$	200	4	7	0.25	0.8
$B^\pm \to \pi^0 e\nu$	$lowq$	250	5	8	0.1	0.8
	$medq$	200	5	7	0.1	0.6
	$highq$	200	5	10	0.05	0.6
$B^0 \to \pi^\pm \mu\nu$	$lowq$	200	5	6	0.1	0.6
	$medq$	200	5	9	0.15	0.8
	$highq$	200	4	10	0.15	0.7
$B^\pm \to \pi^0 \mu\nu$	$lowq$	200	4	9	0.15	0.8
	$medq$	200	5	8	0.1	0.5
	$highq$	200	3	9	0.2	0.8

Table C.1: Optimal hyperparameters as found from the grid search for the continuum suppression BDTs for the three q^2 bins of the four modes.

Mode	q^2 Bin	BDT Parameters				
		nTrees	nLevels	nCuts	Shrinkage	Ratio
$B^0 \to \pi^\pm e\nu$	$lowq$	150	2	8	0.1	0.8
	$medq$	150	2	6	0.1	0.6
	$highq$	150	2	6	0.05	0.2
$B^\pm \to \pi^0 e\nu$	$lowq$	150	2	6	0.1	0.1
	$medq$	150	2	7	0.25	0.6
	$highq$	150	2	6	0.25	0.2
$B^0 \to \pi^\pm \mu\nu$	$lowq$	150	2	7	0.2	0.4
	$medq$	150	2	6	0.05	0.3
	$highq$	150	2	6	0.05	0.5
$B^\pm \to \pi^0 \mu\nu$	$lowq$	150	2	6	0.15	0.2
	$medq$	150	2	6	0.25	0.8
	$highq$	150	2	6	0.25	0.9

Table C.2: Optimal hyperparameters as found from the grid search for the $B\overline{B}$ suppression BDTs for the three q^2 bin of the four modes.

C.4 Importance Rankings

Rank	$B^0 \to \pi^\pm e\nu$			$B^0 \to \pi^\pm \mu\nu$		
	lowq	***medq***	***highq***	***owq***	***medq***	***highq***
1	R_2	R_2	R_2	R_2	R_2	R_2
2	$\cos\theta_{Thrust}$	h_{20}^{so}	h_{20}^{so}	h_{20}^{so}	h_{20}^{so}	h_{20}^{so}
3	h_{20}^{so}	h_{22}^{so}	$\cos\theta_{Thrust}$	$\cos\theta_{Thrust}$	h_{22}^{so}	CLEO 1
4	h_2^{oo}	h_0^{oo}	CLEO 1	$\cos\theta_z$	CLEO 1	$\cos\theta_{Thrust}$
5	h_2^{so}	$\cos\theta_{Thrust}$	h_0^{oo}	h_2^{so}	$\cos\theta_{Thrust}$	$\cos\theta_z$
6	h_{12}^{so}	h_{12}^{so}	h_2^{oo}	h_{12}^{so}	h_0^{oo}	h_2^{oo}
7	h_{22}^{so}	h_2^{so}	$\cos\theta_z$	h_{22}^{so}	$\cos\theta_z$	h_0^{oo}
8	CLEO 1	CLEO 1	h_{12}^{so}	h_2^{oo}	h_{12}^{so}	h_{12}^{so}
9	h_0^{oo}	h_2^{oo}	h_{20}^{so}	h_0^{oo}	h_2^{so}	h_{22}^{so}
10	$\cos\theta_z$	$\cos\theta_z$	h_2^{so}	CLEO 1	h_2^{oo}	h_2^{so}

Table C.3: Ranking of the importance of the variables in the continuum suppression BDTs of the B^0 modes.

Rank	$B^\pm \to \pi^0 e\nu$			$B^\pm \to \pi^0 \mu\nu$		
	lowq	***medq***	***highq***	***lowq***	***medq***	***highq***
1	R_2	R_2	R_2	R_2	R_2	R_2
2	$\cos\theta_{Thrust}$	h_{20}^{so}	h_{20}^{so}	h_{20}^{so}	R_2	h_{20}^{so}
3	h_2^{so}	h_2^{so}	CLEO 1	h_2^{so}	h_{22}^{so}	$\cos\theta_{Thrust}$
4	h_{20}^{so}	h_{12}^{so}	$\cos\theta_{Thrust}$	$\cos\theta_{Thrust}$	h_0^{oo}	CLEO 1
5	h_{12}^{so}	h_0^{oo}	h_0^{oo}	h_{12}^{so}	CLEO 1	$\cos\theta_z$
6	h_2^{oo}	h_{22}^{so}	h_2^{oo}	$\cos\theta_z$	h_{12}^{so}	h_0^{oo}
7	h_{22}^{so}	$\cos\theta_{Thrust}$	h_{12}^{so}	h_{22}^{so}	$\cos\theta_{Thrust}$	h_2^{oo}
8	h_0^{oo}	Thrust	$\cos\theta_z$	h_2^{oo}	h_2^{so}	h_{22}^{so}
9	CLEO 1	CLEO 1	h_{22}^{so}	h_0^{oo}	Thrust	h_{12}^{so}
10	$\cos\theta_z$	h_2^{oo}	Thrust	CLEO 1	h_2^{oo}	Thrust

Table C.4: Ranking of the importance of the variables in the continuum suppression BDTs of the $B^\pm$ modes.

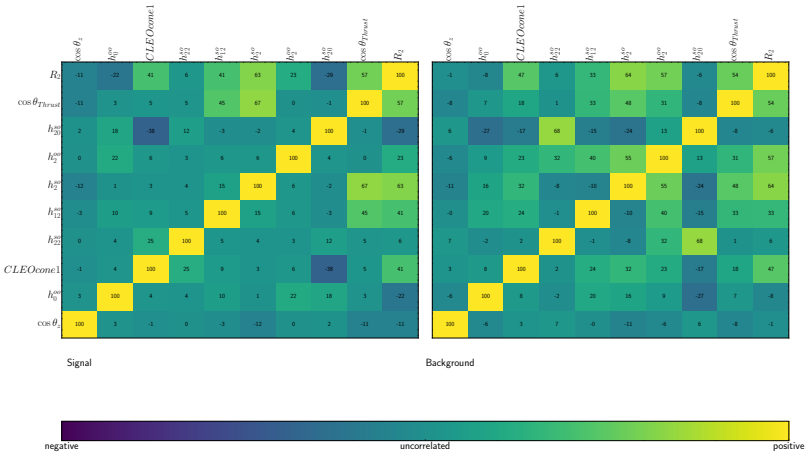
Rank	$B^0 \to \pi^\pm e\nu$			$B^0 \to \pi^\pm \mu\nu$		
	lowq	***medq***	***highq***	***lowq***	***medq***	***highq***
1	M_{bc}^{ROE}	M_{bc}^{ROE}	ΔE^{ROE}	M_{bc}^{ROE}	M_{bc}^{ROE}	M_{bc}^{ROE}
2	p_{miss}^{lab}	p_{miss}^{lab}	M_{bc}^{ROE}	ΔE^{ROE}	ΔE^{ROE}	ΔE^{ROE}
3	p_π	ΔE^{ROE}	p_{miss}^{lab}	p_{miss}^{lab}	p_{miss}^{lab}	p_{miss}^{lab}
4	ΔE^{ROE}	$\cos\theta_{\pi\ell}$	$\cos\theta_{\pi\ell}$	$\cos\theta_{BY}$	$\cos\theta_{\pi\ell}$	$\cos\theta_{\pi\ell}$
5	$\cos\theta_{\pi\ell}$	p_ℓ	p_ℓ	$\cos\theta_{\pi\ell}$	p_ℓ	$\cos\theta_{BY}$
6	$\cos\theta_{BY}$	$\cos\theta_{BY}$	$\cos\theta_{BY}$	p_π	$\cos\theta_{BY}$	p_ℓ
7	p_ℓ	$\frac{m_{miss}^2}{2E_{miss}}$	p_π	p_ℓ	$\frac{m_{miss}^2}{2E_{miss}}$	p_π
8	$\frac{m_{miss}^2}{2E_{miss}}$	p_π	$\frac{m_{miss}^2}{2E_{miss}}$	$\frac{m_{miss}^2}{2E_{miss}}$	p_π	$\frac{m_{miss}^2}{2E_{miss}}$,

Table C.5: Ranking of the importance of the variables in the $B\overline{B}$ suppression BDTs of the B^0 modes.

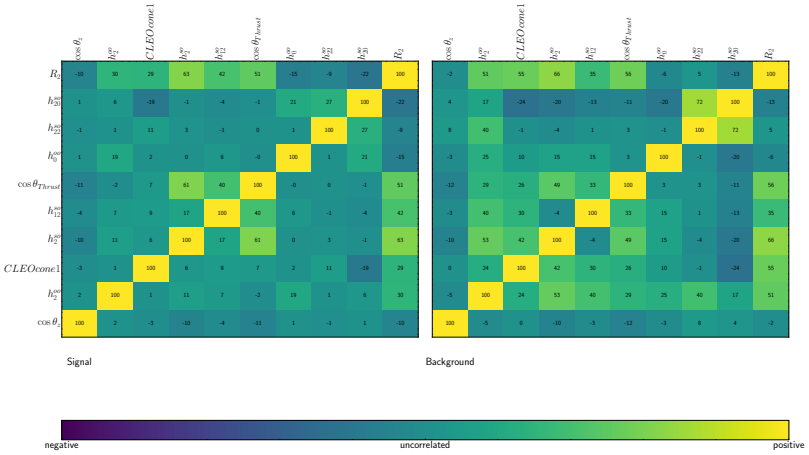
Rank	$B^\pm \to \pi^0 e\nu$			$B^\pm \to \pi^0 \mu\nu$		
	owq	***medq***	***highq***	***lowq***	***medq***	***highq***
1	M_{bc}^{ROE}	M_{bc}^{ROE}	M_{bc}^{ROE}	M_{bc}^{ROE}	M_{bc}^{ROE}	M_{bc}^{ROE}
2	p_π	p_{miss}^{lab}	p_{miss}^{lab}	p_π	p_{miss}^{lab}	p_{miss}^{lab}
3	$\cos\theta_{\pi\ell}$	$\cos\theta_{\pi\ell}$	$\cos\theta_{\pi\ell}$	p_{miss}^{lab}	$\cos\theta_{\pi\ell}$	$\cos\theta_{\pi\ell}$
4	p_{miss}^{lab}	ΔE^{ROE}	ΔE^{ROE}	$\cos\theta_{\pi\ell}$	ΔE^{ROE}	ΔE^{ROE}
5	ΔE^{ROE}	p_ℓ	$\cos\theta_{BY}$	ΔE^{ROE}	$\cos\theta_{BY}$	$\cos\theta_{BY}$
6	$\cos\theta_{BY}$	$\cos\theta_{BY}$	p_π	$\cos\theta_{BY}$	p_ℓ	p_π
7	$\frac{m_{miss}^2}{2E_{miss}}$	θ_{miss}^{lab}	θ_{miss}^{lab}	$\frac{m_{miss}^2}{2E_{miss}}$	θ_{miss}^{lab}	p_ℓ
8	p_ℓ	$\frac{m_{miss}^2}{2E_{miss}}$	p_ℓ	p_ℓ	$\frac{m_{miss}^2}{2E_{miss}}$	θ_{miss}^{lab}

Table C.6: Ranking of the importance of the variables in the $B\overline{B}$ suppression BDTs of the $B^\pm$ modes.

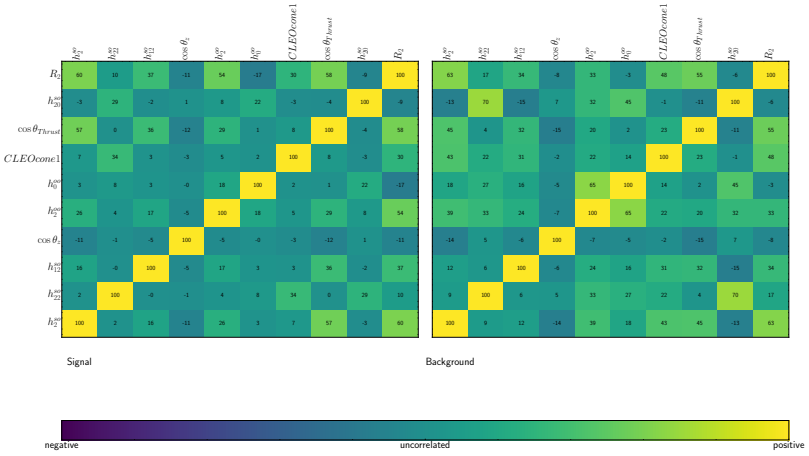
C.5 Correlation Matrices



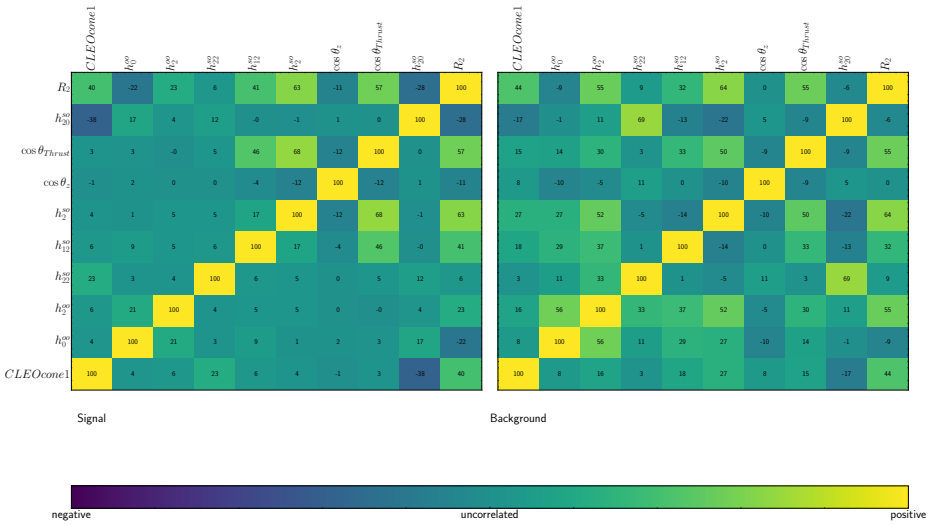
(a) $B^0 \to \pi^{\pm} e\nu : lowq$



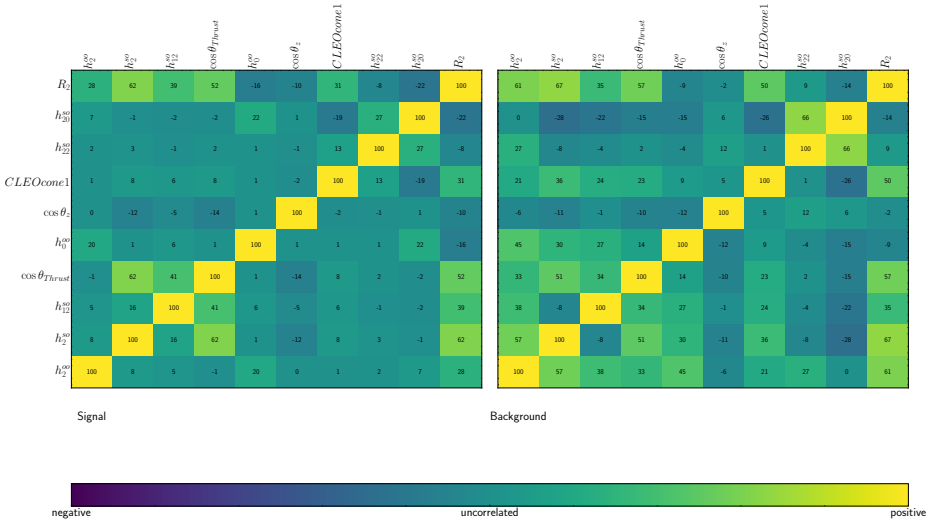
(b) $B^0 \to \pi^{\pm} e\nu : medq$



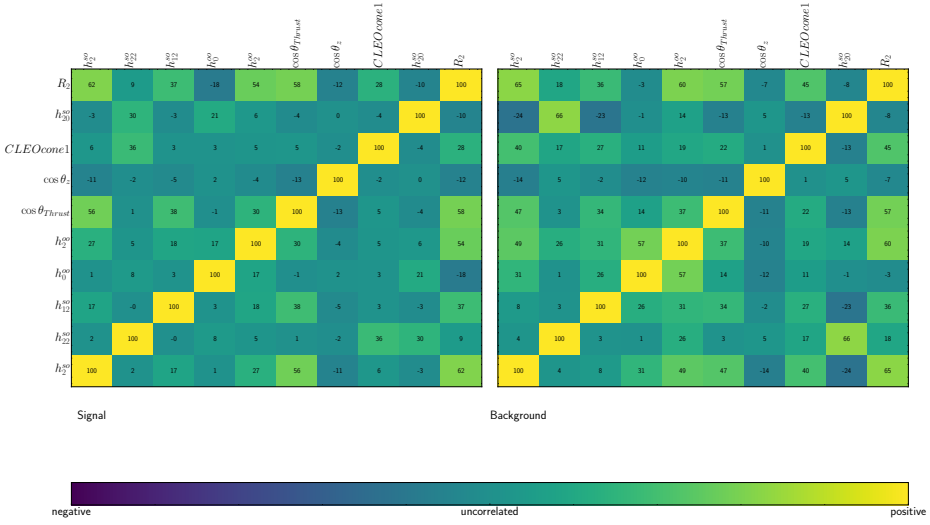
(c) $B^0 \to \pi^{\pm} e\nu : highq$



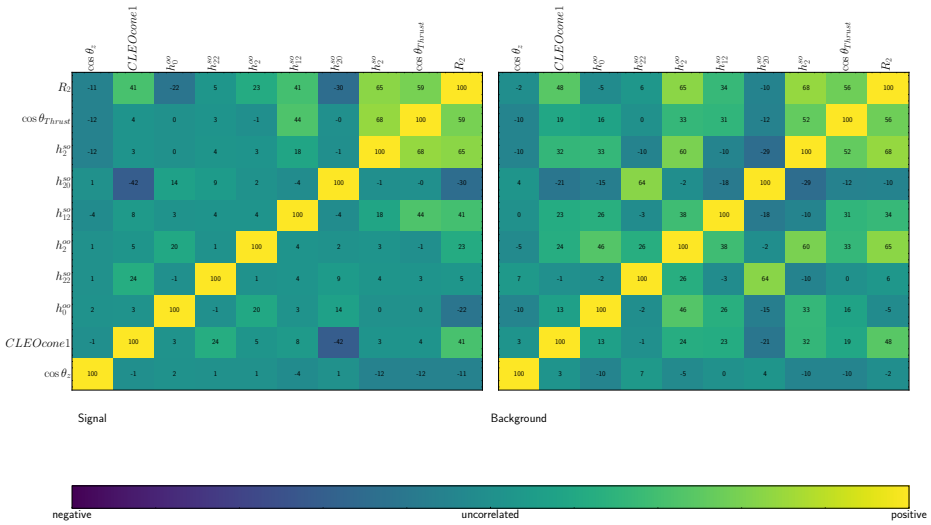
(d) $B^0 \to \pi^{\pm}\mu\nu$: *lowq*



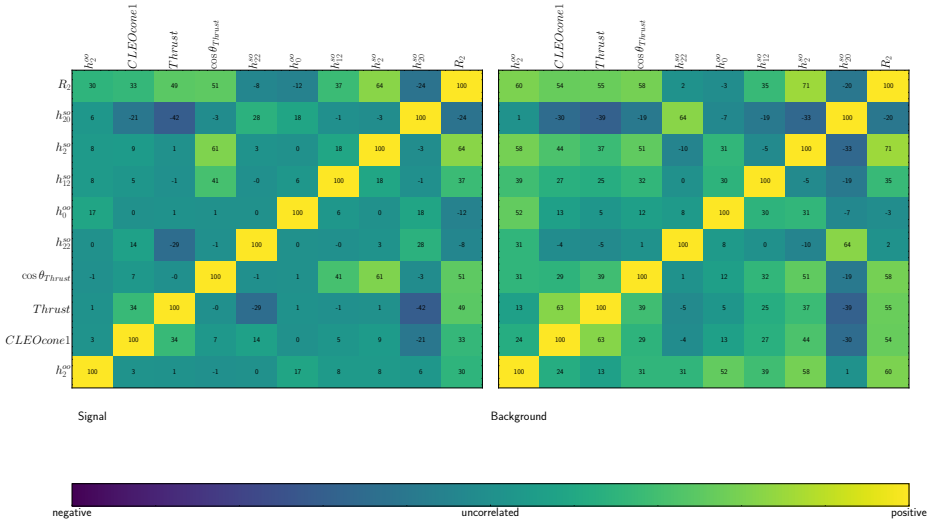
(e) $B^0 \to \pi^{\pm}\mu\nu$: *medq*



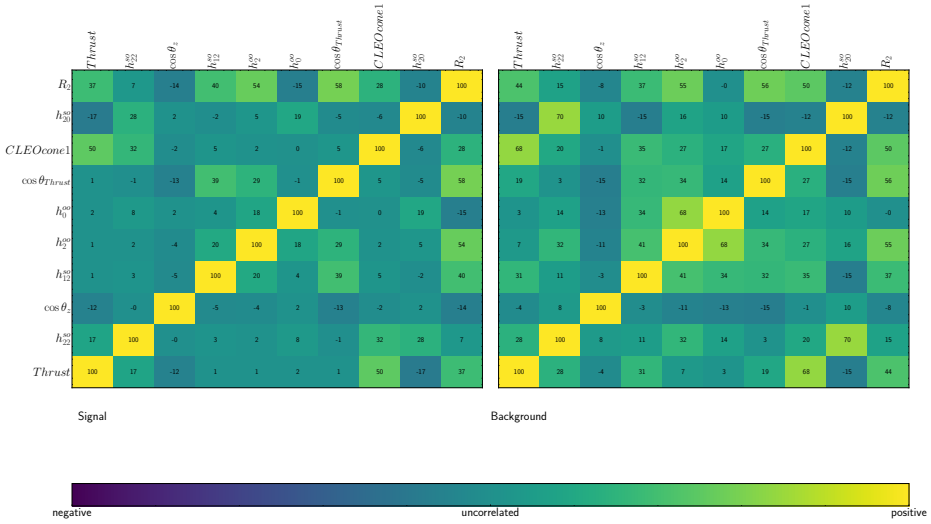
(f) $B^0 \to \pi^{\pm}\mu\nu$: *highq*



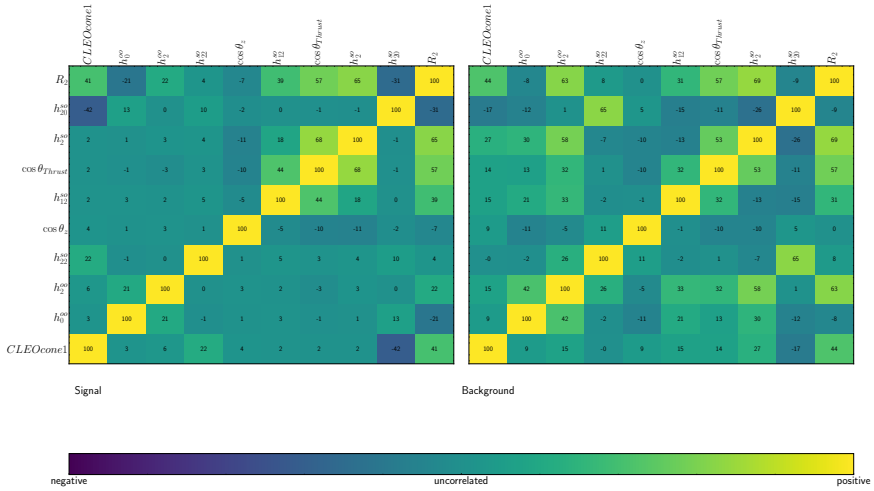
(g) $B^{\pm} \to \pi^0 ev : lowq$



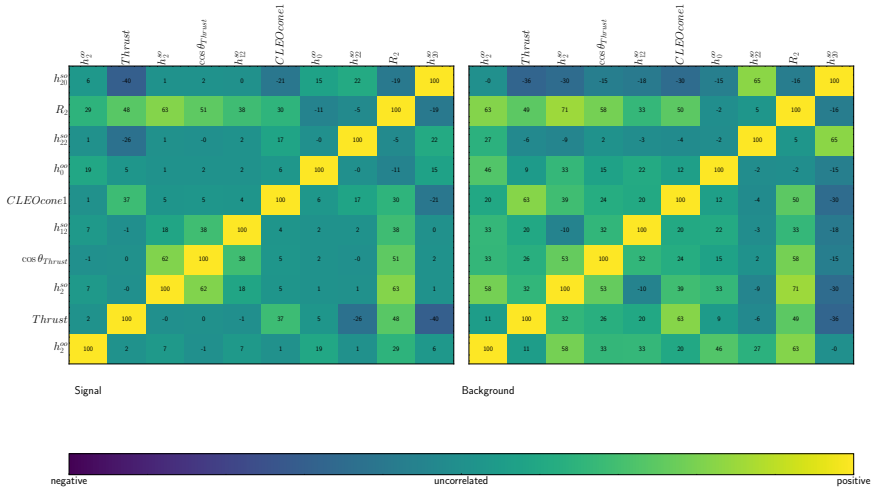
(h) $B^{\pm} \to \pi^0 ev : medq$



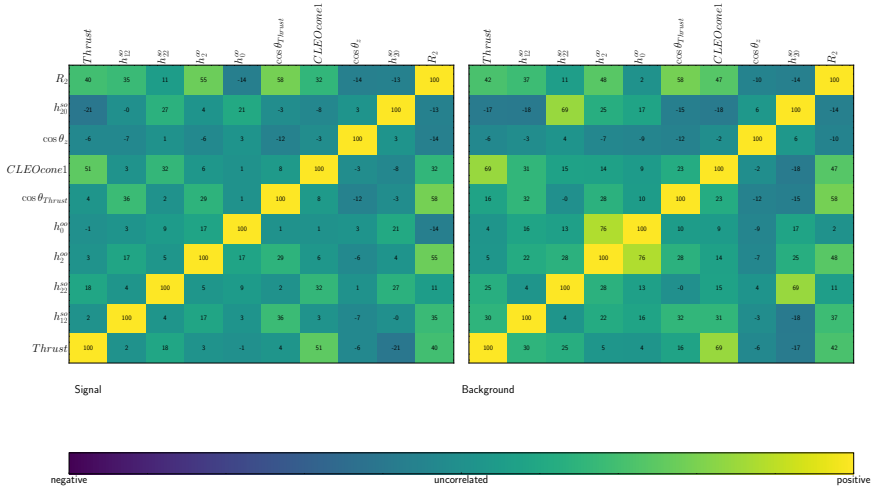
(i) $B^{\pm} \to \pi^0 ev : highq$



(j) $B^{\pm} \to \pi^0 \mu\nu : lowq$

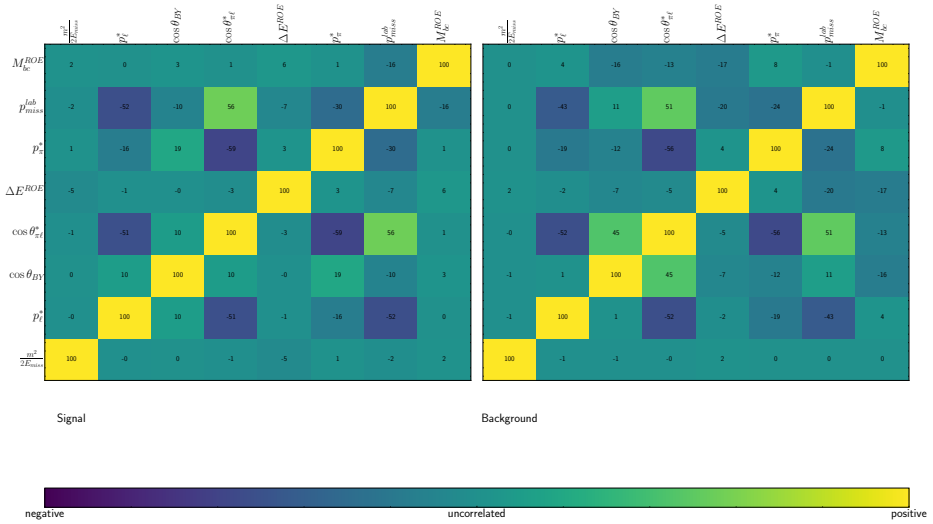


(k) $B^{\pm} \to \pi^0 \mu\nu : medq$

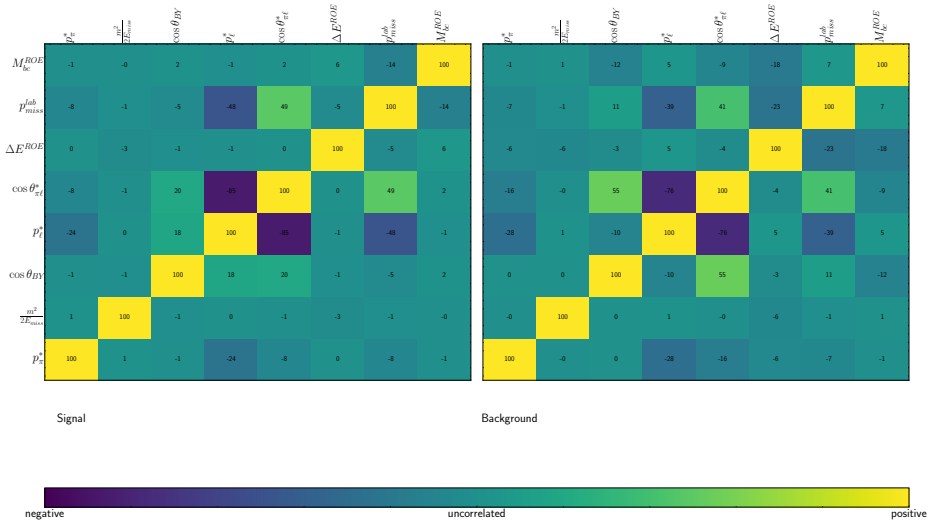


(l) $B^{\pm} \to \pi^0 \mu\nu : highq$

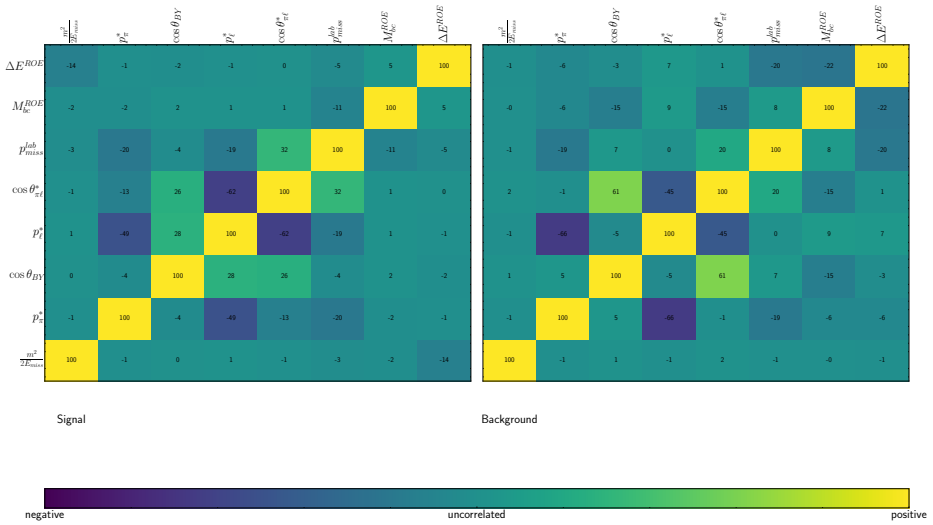
Figure C.8: Correlation matrices for the variables in the continuum suppression BDTs for each mode in the $lowq$, $medq$ and $highq$ q^2 bins.



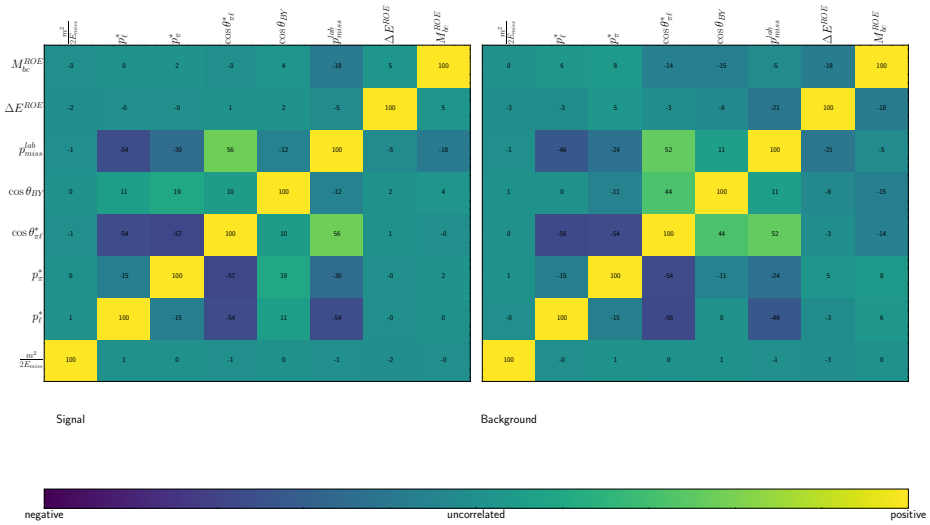
(a) $B^0 \to \pi^0 e\nu$: *lowq*



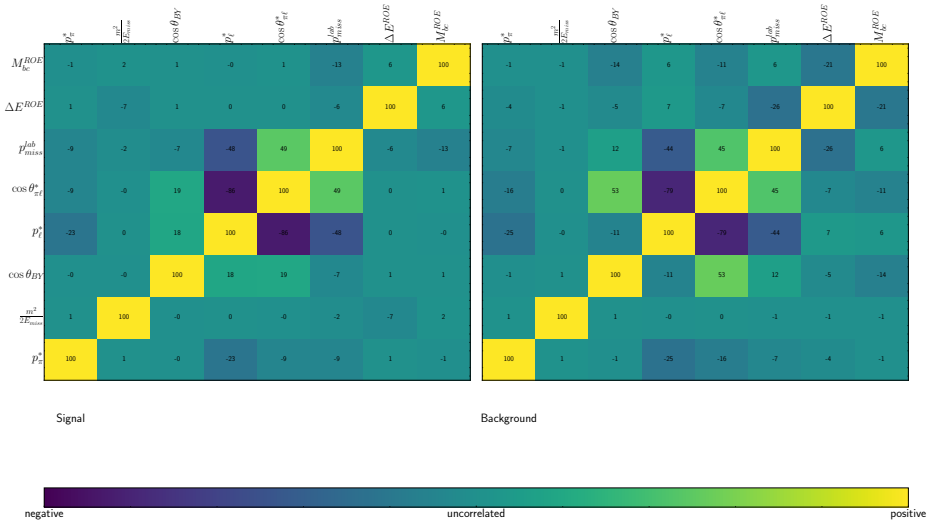
(b) $B^0 \to \pi^0 e\nu$: *medq*



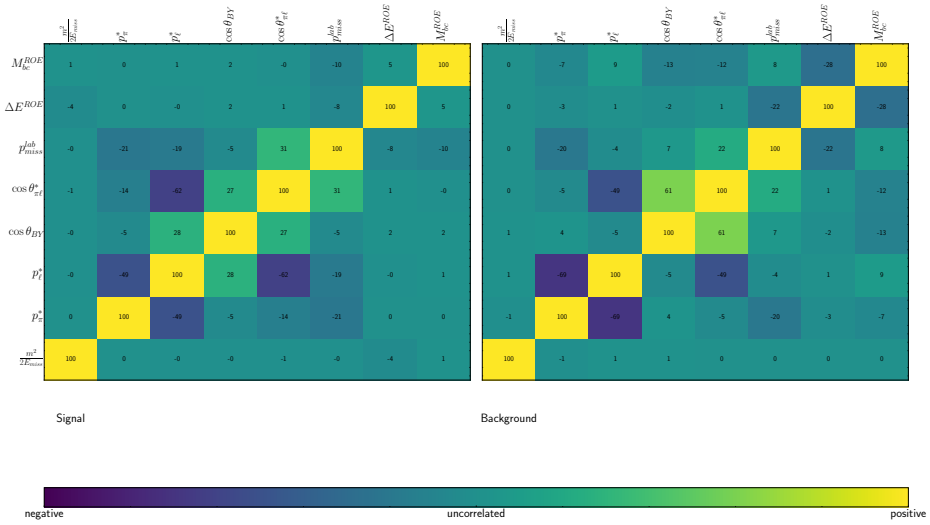
(c) $B^0 \to \pi^0 e\nu$: *highq*



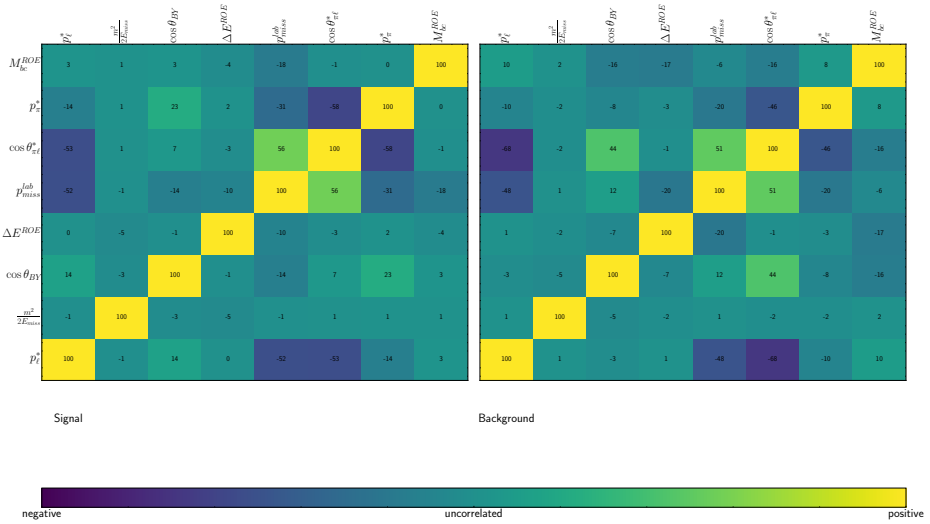
(d) $B^0 \to \pi^0 \mu \nu$: *lowq*



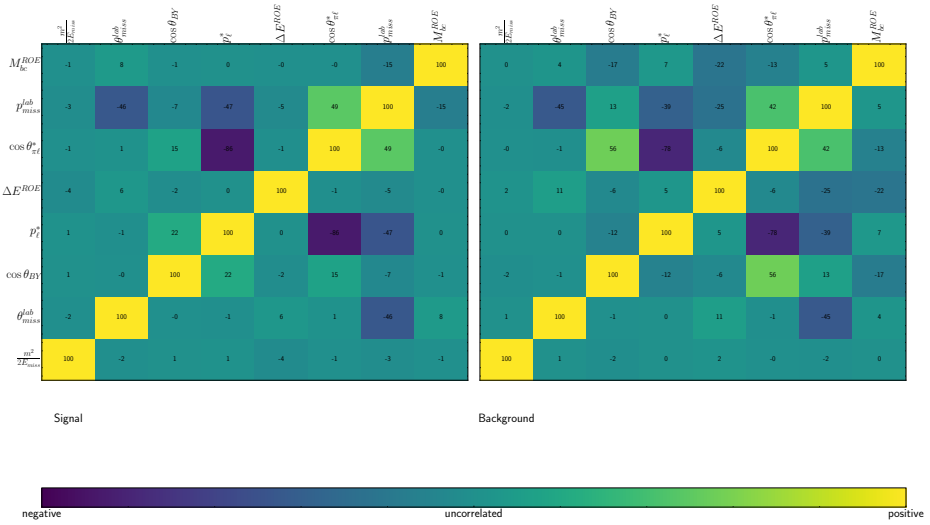
(e) $B^0 \to \pi^0 \mu \nu$: *medq*



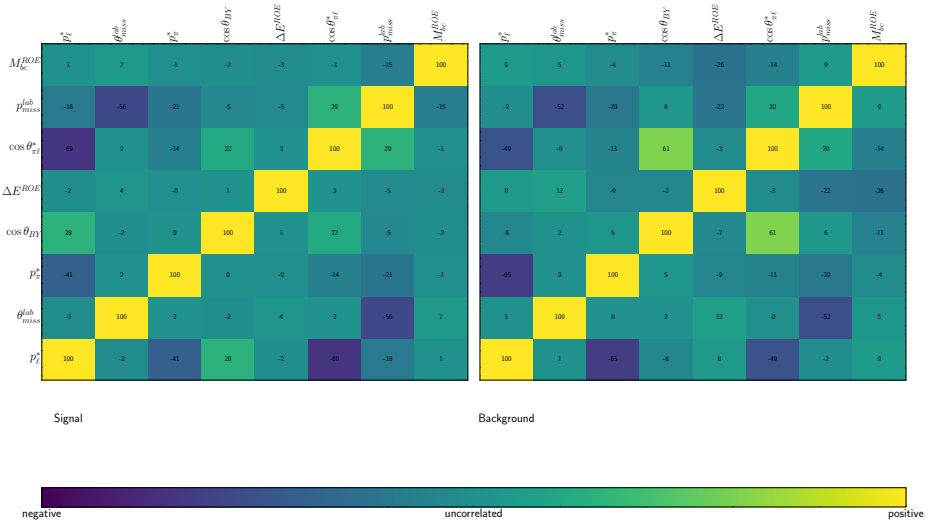
(f) $B^0 \to \pi^0 \mu \nu$: *highq*



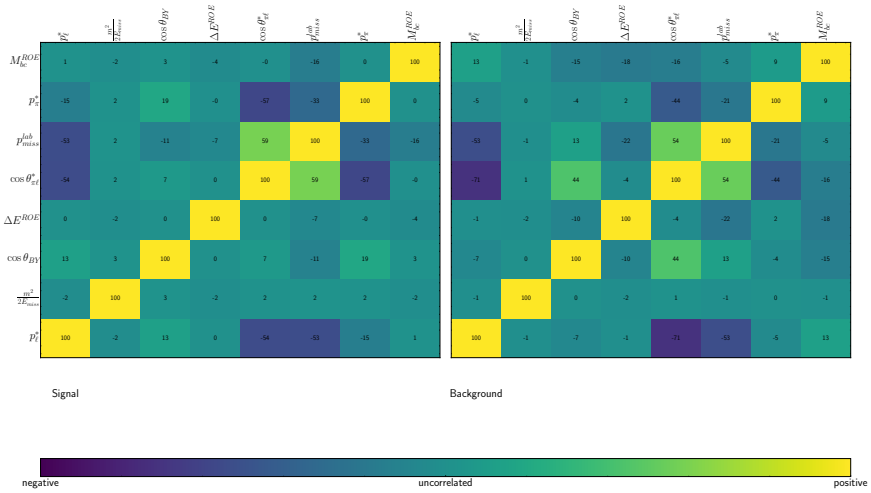
(g) $B^{\pm} \to \pi^0 ev : lowq$



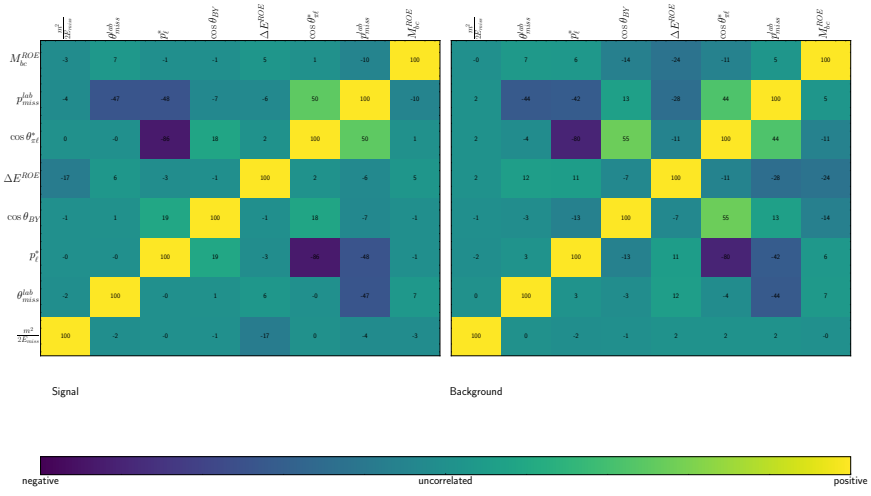
(h) $B^{\pm} \to \pi^0 ev : medq$



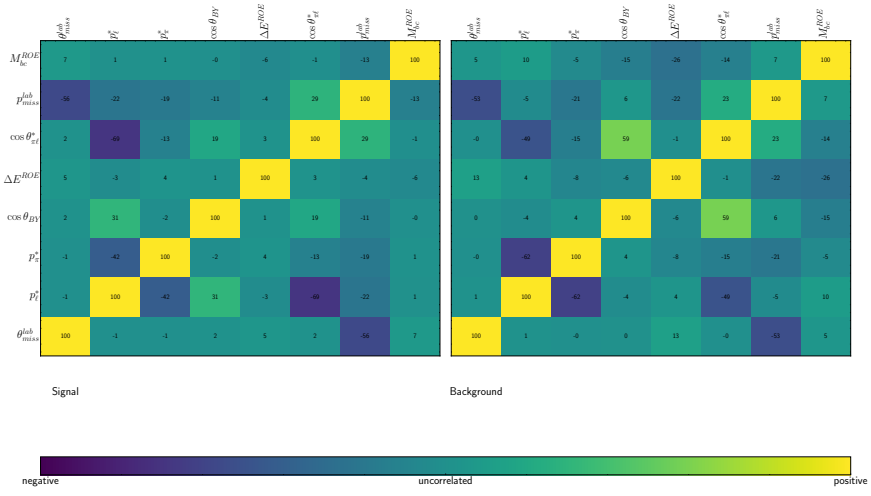
(i) $B^{\pm} \to \pi^0 ev : highq$



(j) $B^{\pm} \to \pi^0 \mu\nu$: *lowq*



(k) $B^{\pm} \to \pi^0 \mu\nu$: *medq*



(l) $B^{\pm} \to \pi^0 \mu\nu$: *highq*

Figure C.9: Correlation matrices for the variables in the $B\overline{B}$ suppression BDTs for each mode in the *lowq*, *medq* and *highq* q^2 bins.

C.6 ROC Curves

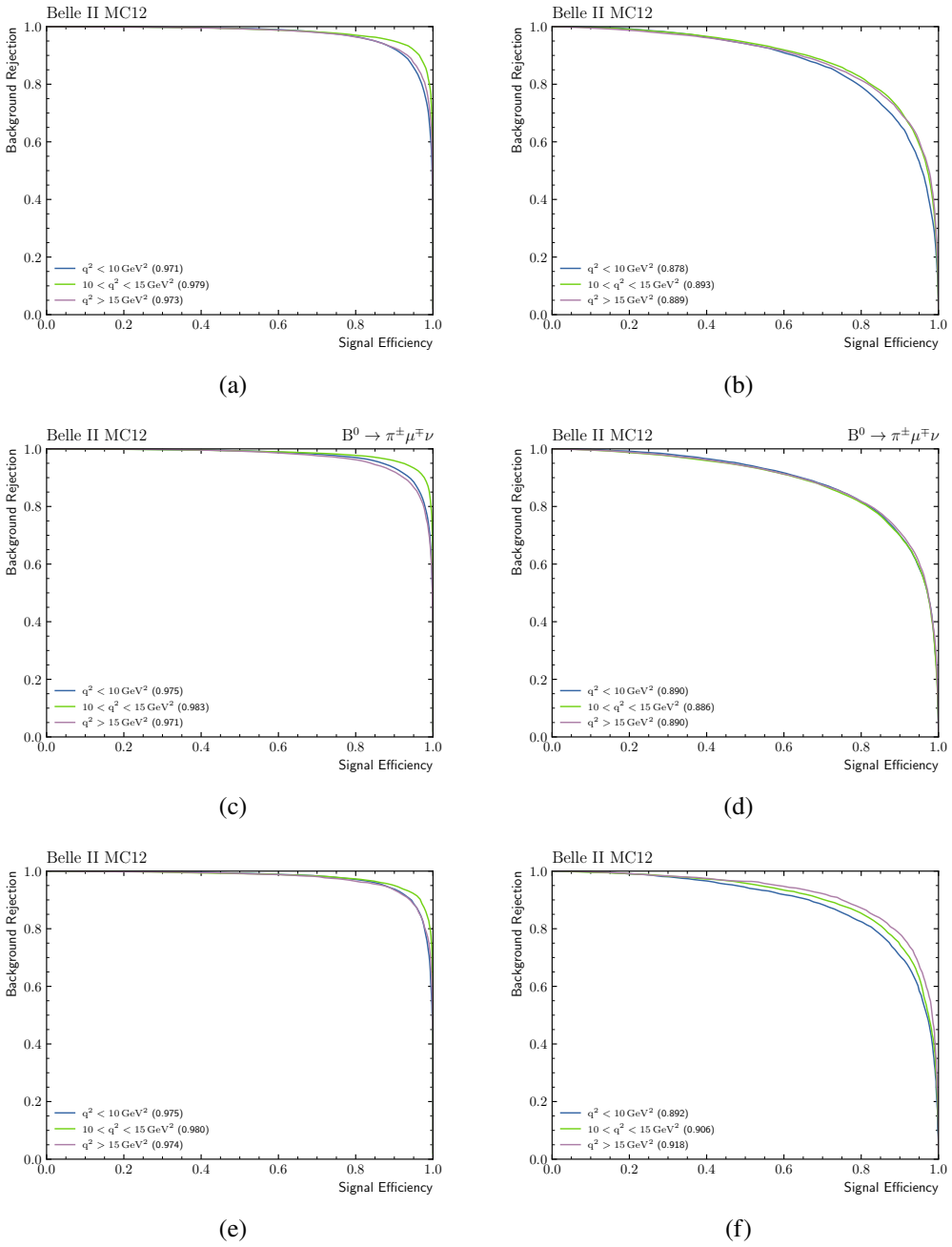


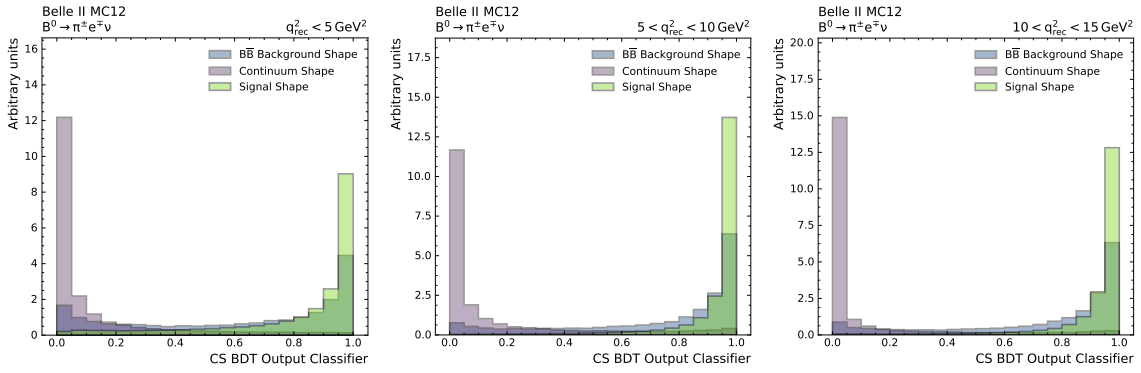
Figure C.10: ROC curves for the *lowq*, *medq* and *highq* q^2 bins for the continuum (left) and $B\overline{B}$ (right) suppression BDTs in the $B^0 \to \pi^\pm e\nu$ (top), $B^0 \to \pi^\pm \mu\nu$ (middle) and $B^\pm \to \pi^0 \mu\nu$ (bottom) modes. The AUC for each of the curves is given in the legend.

C.7 p-Values

Mode	q^2 Bin	CS		$B\overline{B}$	
		Signal	**Background**	**Signal**	**Background**
$B^0 \to \pi^\pm e\nu$	*lowq*	0.514	0.669	0.914	0.170
	medq	0.842	0.196	0.617	0.735
	highq	0.531	0.060	0.428	0.495
$B^\pm \to \pi^0 e\nu$	*lowq*	0.109	0.121	0.355	0.379
	medq	0.872	0.740	0.580	0.907
	highq	0.227	0.172	0.351	0.421
$B^0 \to \pi^\pm \mu\nu$	*lowq*	0.139	0.193	0.283	0.652
	medq	0.809	0.333	0.935	0.292
	highq	0.942	0.372	0.650	0.096
$B^\pm \to \pi^0 \mu\nu$	*lowq*	0.754	0.898	0.939	0.451
	medq	0.792	0.526	0.800	0.814
	highq	0.132	0.467	0.319	0.458

Table C.7: p-Values of signal and background of the continuum suppression and $B\overline{B}$ suppression BDTs for the *lowq*, *medq* and *highq* q^2 bins for the four modes.

C.8 BDT Classifiers



(a) (b) (c)

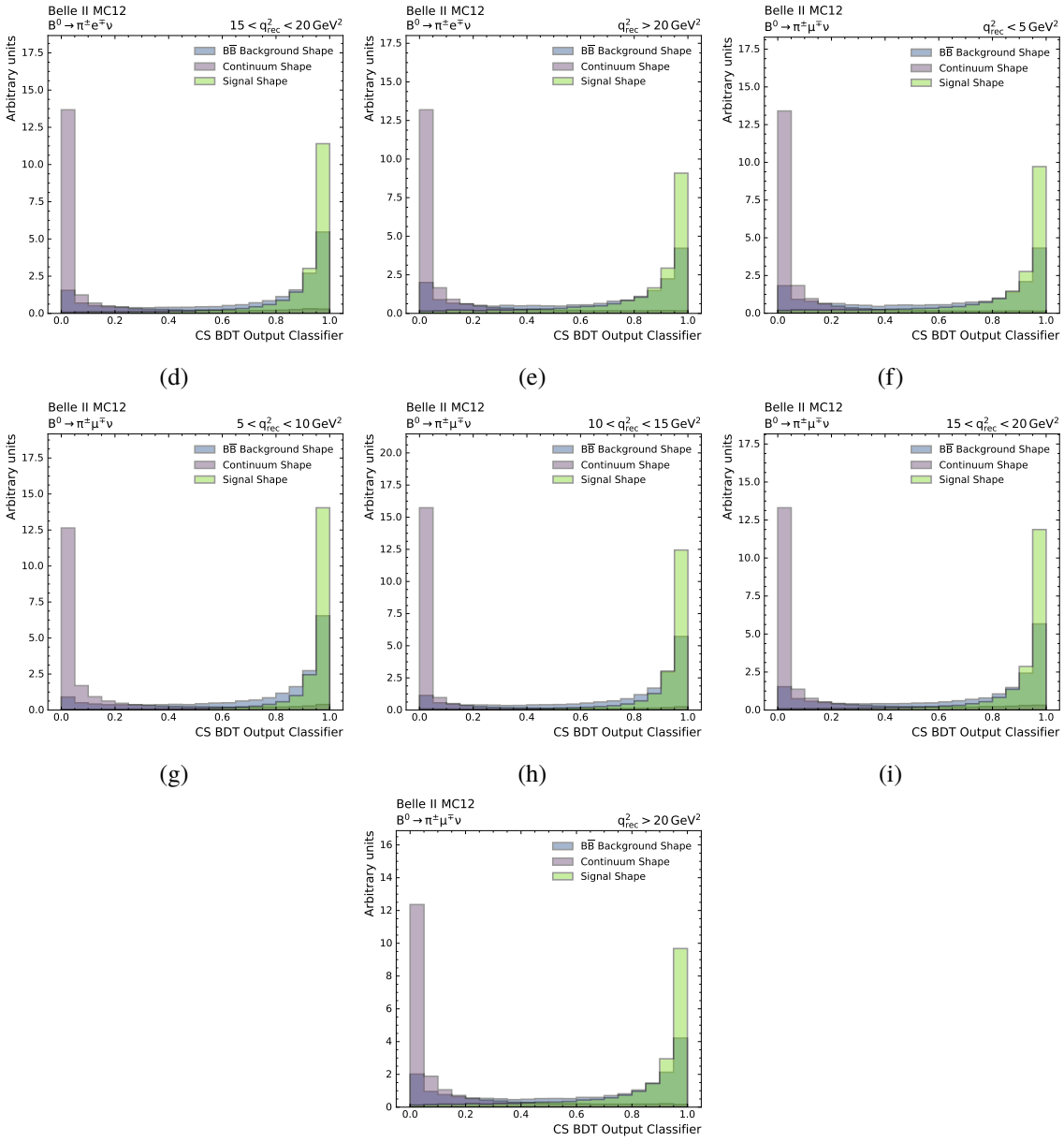


Figure C.11: The normalised distribution of the continuum suppression BDT output classifier for signal in green, continuum background in purple and $B\overline{B}$ background in blue for the $B^0 \to \pi^\pm \ell \nu$ mode.

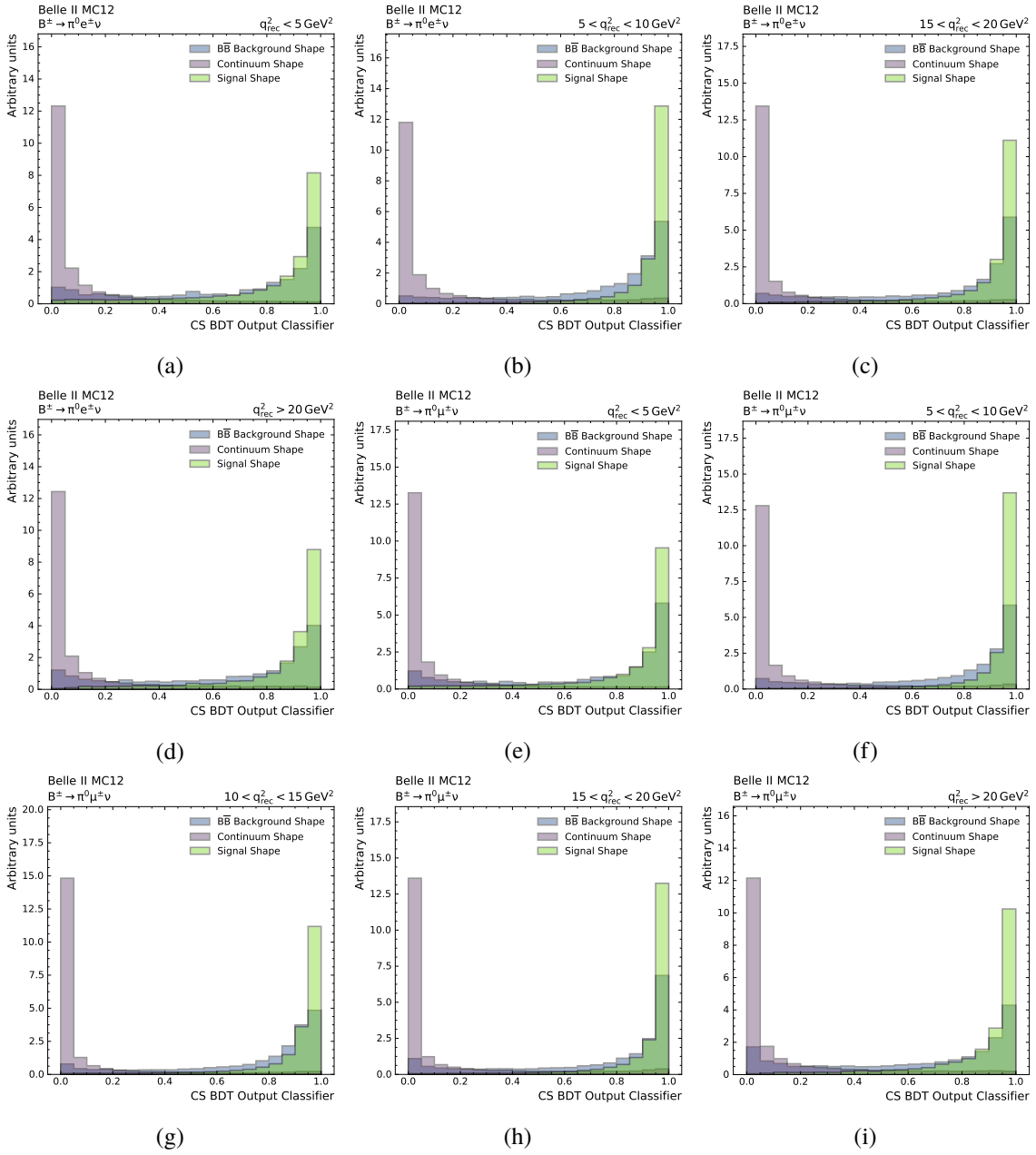


Figure C.12: The normalised distribution of the continuum suppression BDT output classifier for signal in green, continuum background in purple and $B\overline{B}$ background in blue for the $B^\pm \to \pi^0 \ell \nu$ modes.

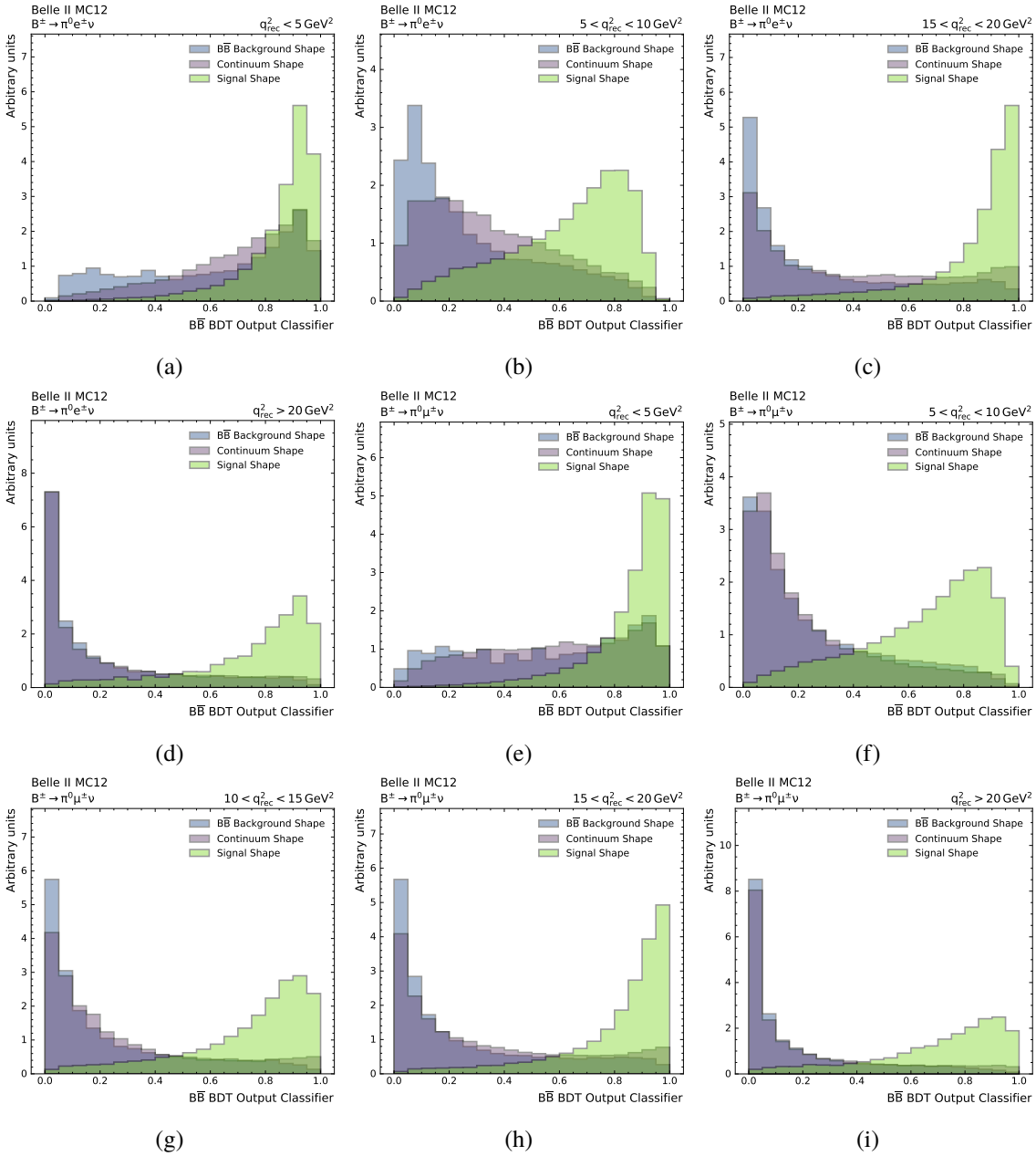
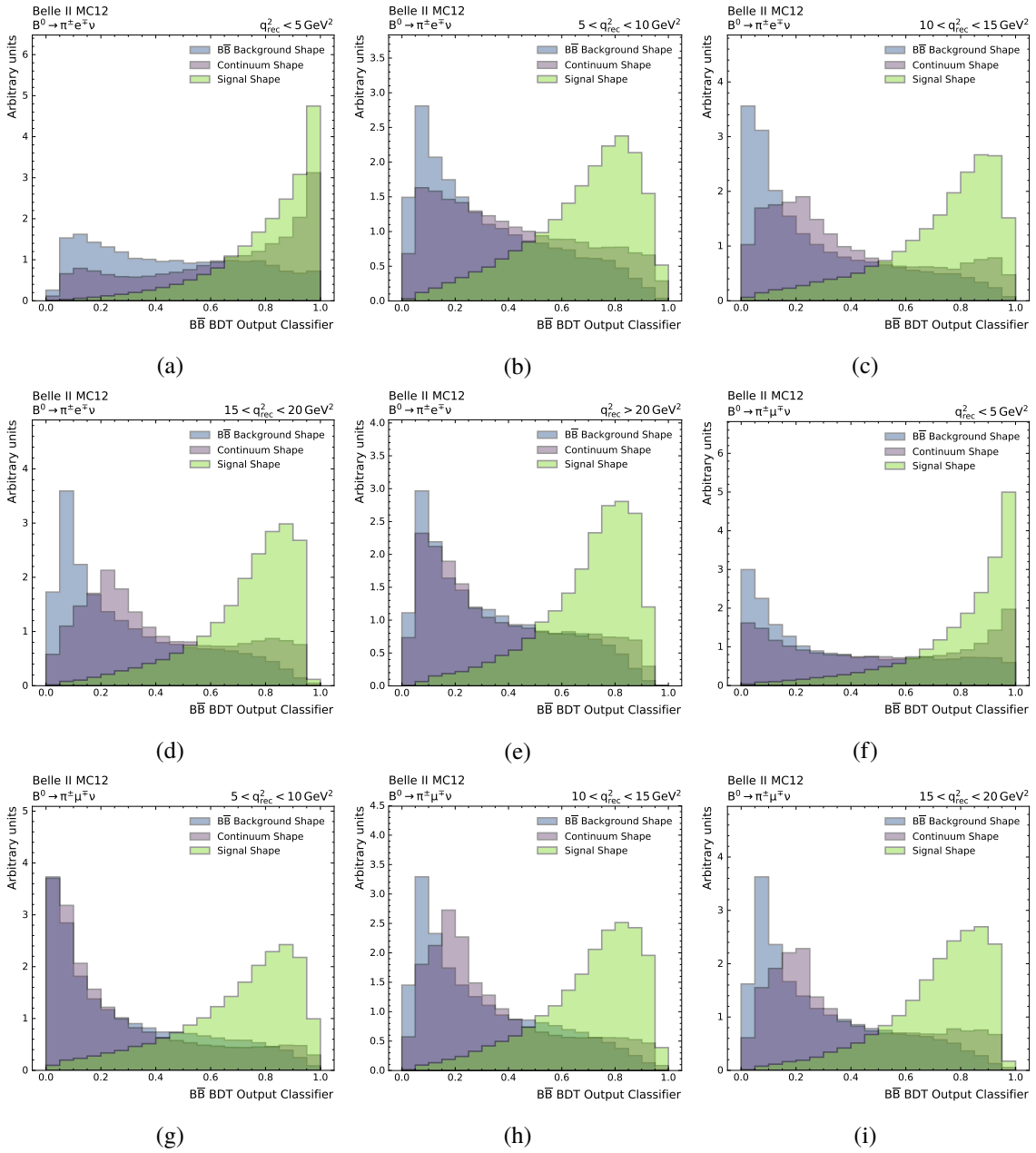
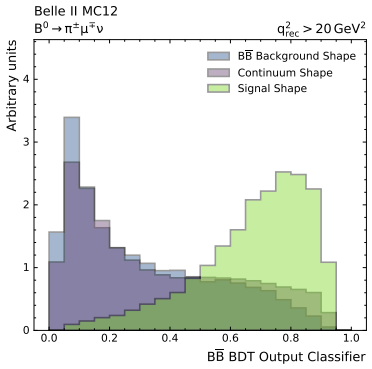


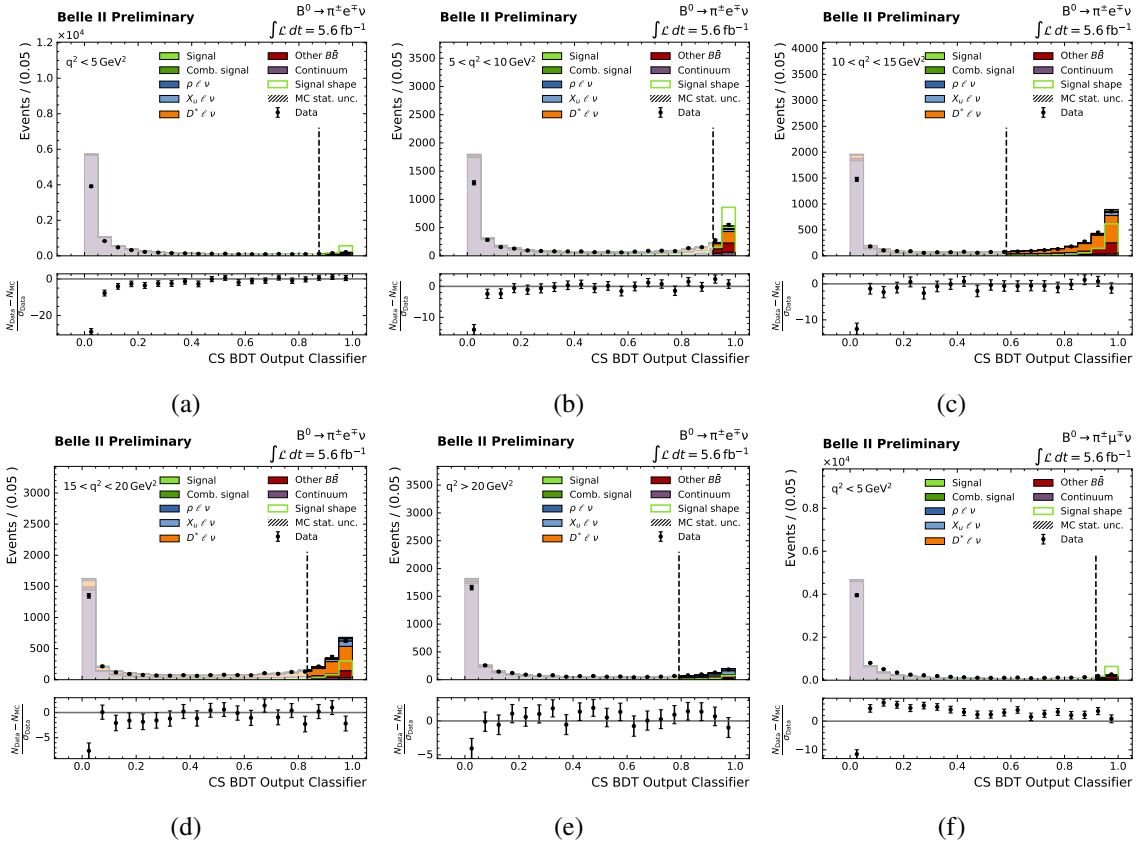
Figure C.13: The normalised distribution of the $B\overline{B}$ suppression BDT output classifier for signal in green, continuum background in purple and $B\overline{B}$ background in blue for the $B^\pm \to \pi^0 \ell \nu$ modes.


Belle II MC12
$B^0 \to \pi^\pm e^\mp \nu$
$q^2_{rec} < 5\,GeV^2$
$5 < q^2_{rec} < 10\,GeV^2$
$10 < q^2_{rec} < 15\,GeV^2$
$15 < q^2_{rec} < 20\,GeV^2$
$q^2_{rec} > 20\,GeV^2$
$B^0 \to \pi^\pm \mu^\mp \nu$
$B\bar{B}$ Background Shape
Continuum Shape
Signal Shape
Arbitrary units
$B\bar{B}$ BDT Output Classifier
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)



(j)

Figure C.14: The normalised distribution of the $B\overline{B}$ suppression BDT output classifier for signal in green, continuum background in purple and $B\overline{B}$ background in blue for the $B^0 \to \pi^\pm \ell \nu$ mode.



(a) (b) (c)

(d) (e) (f)

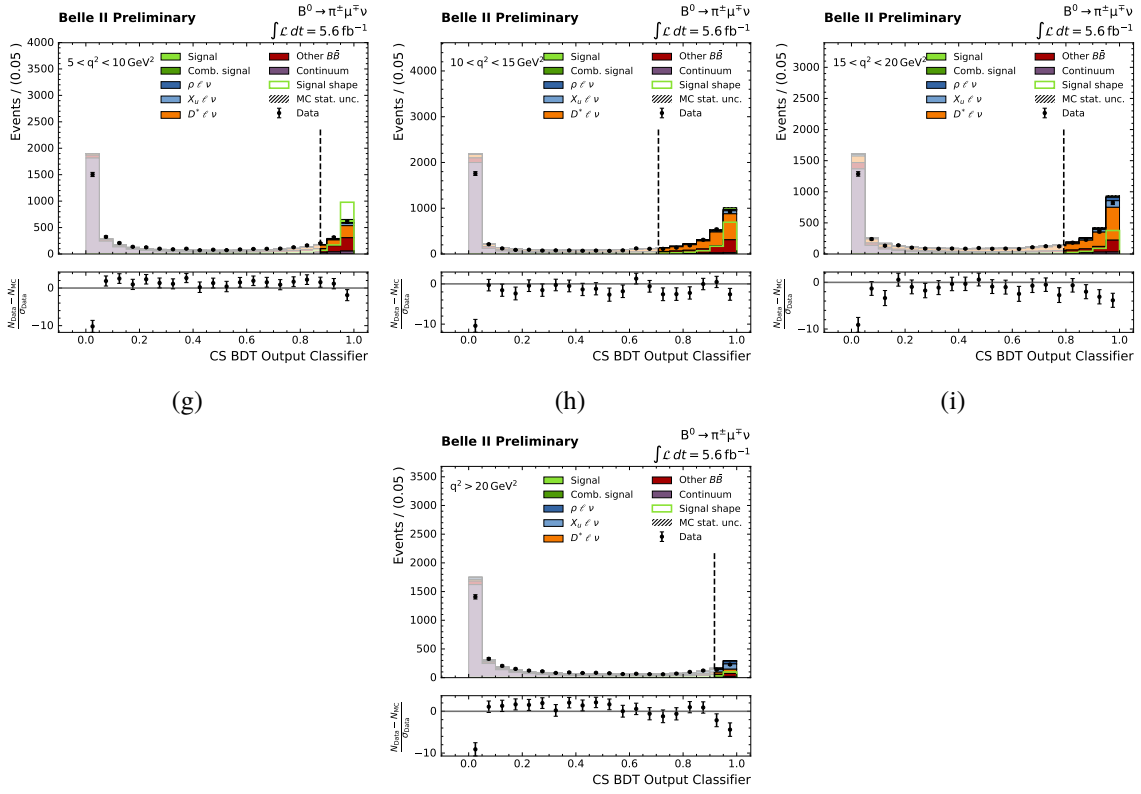
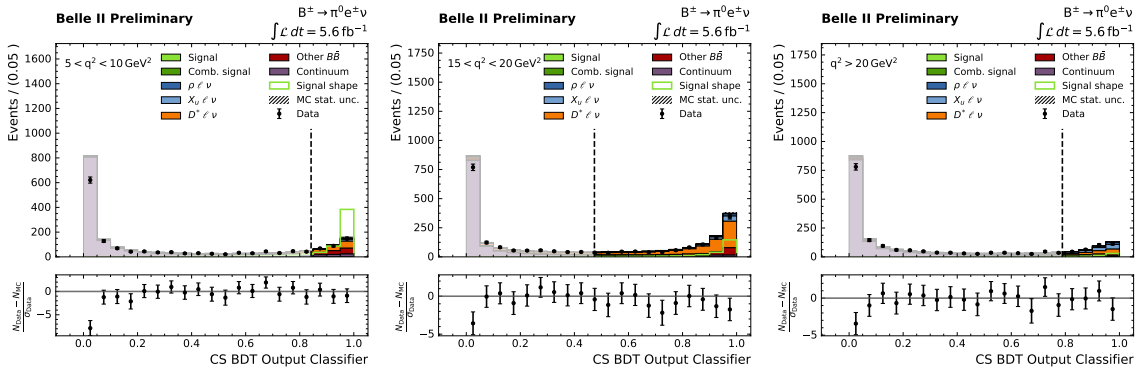


Figure C.15: Data MC agreement for the continuum suppression BDT output classifiers for the five q^2 bins of the B^0 modes.



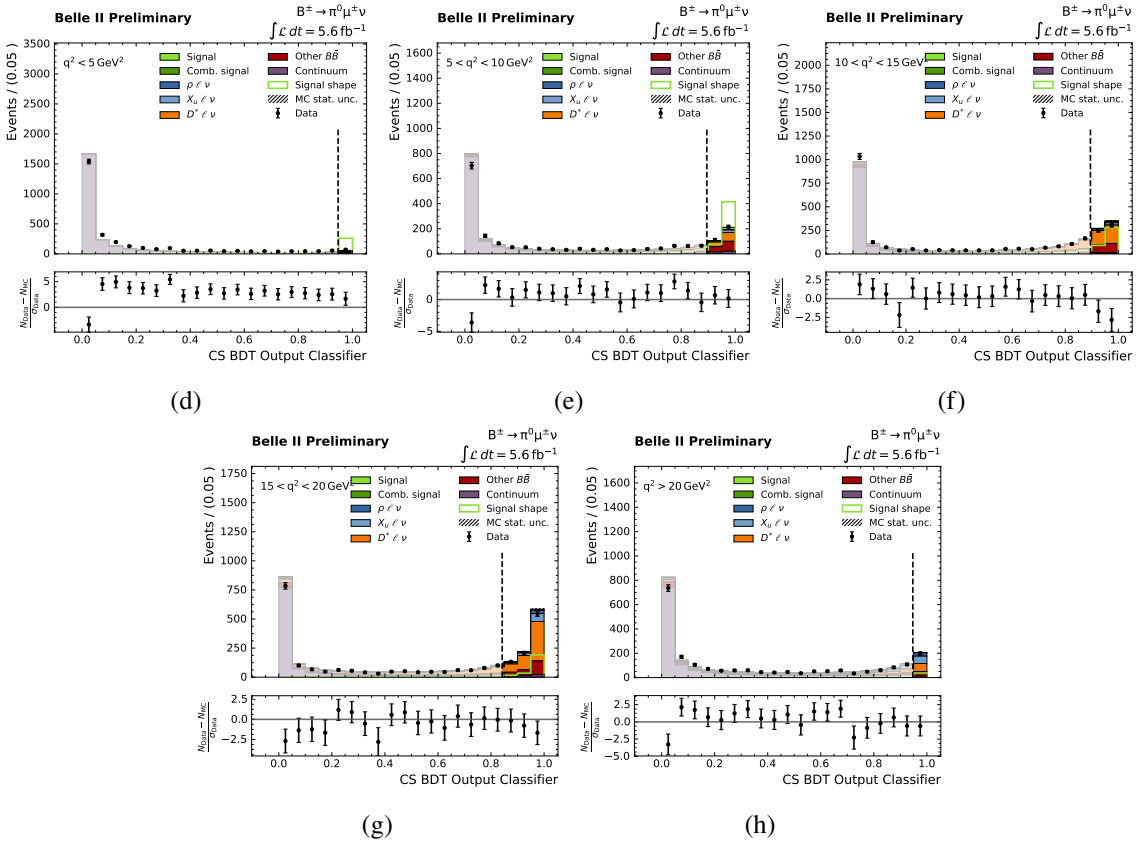
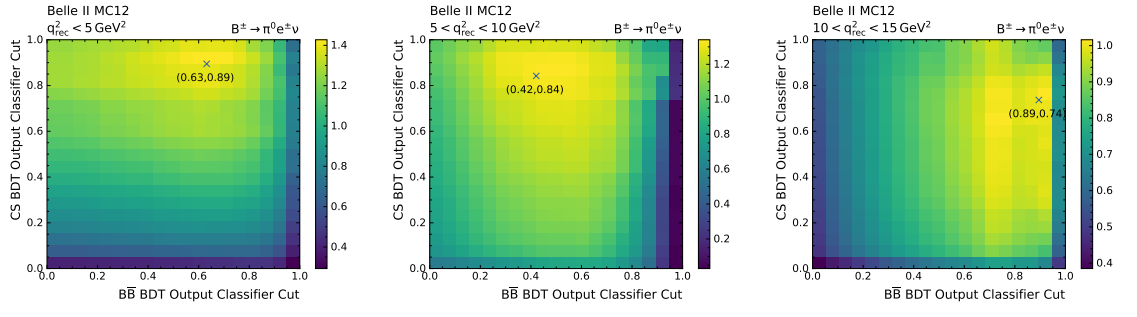


Figure C.16: Data MC agreement for the continuum suppression BDT output classifiers for the five q^2 bins of the $B^\pm$ modes.

C.9 FOM Optimization



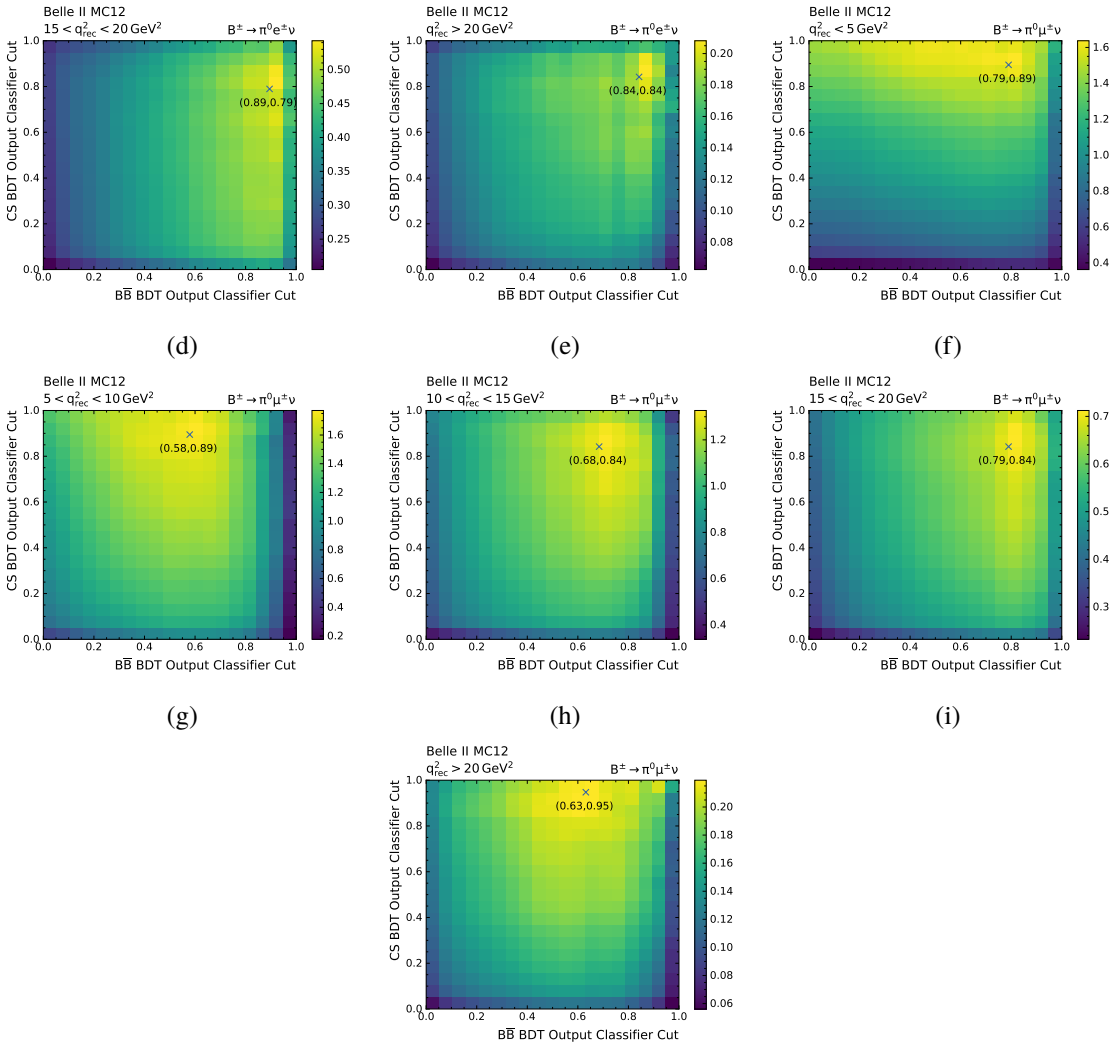


Figure C.17: 2D histograms of the FOM as a function of the continuum suppression and $B\overline{B}$ BDT output classifiers in the five q^2 bins for the $B^{\pm}$ modes. The optimal cuts are marked and labelled.

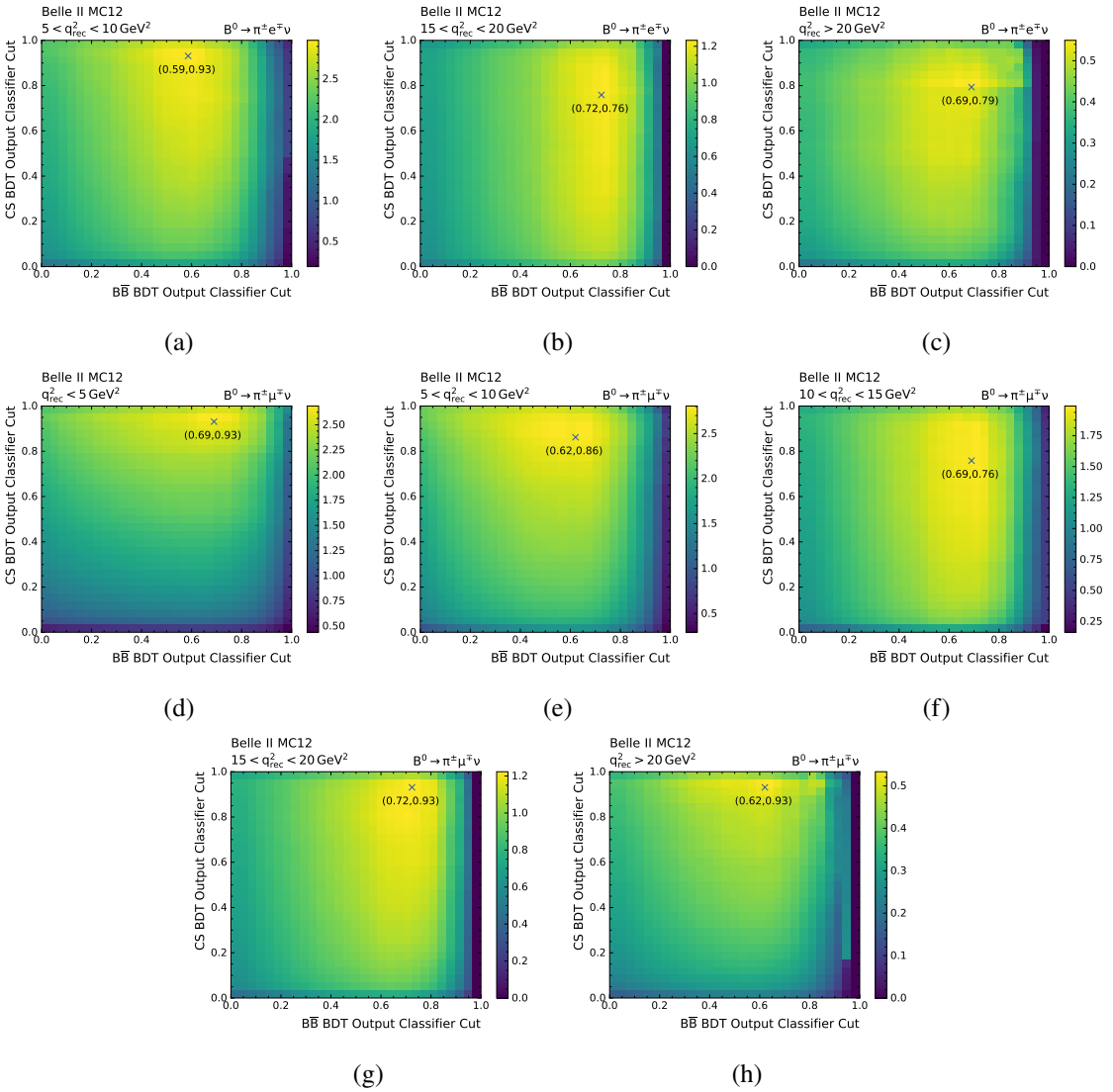
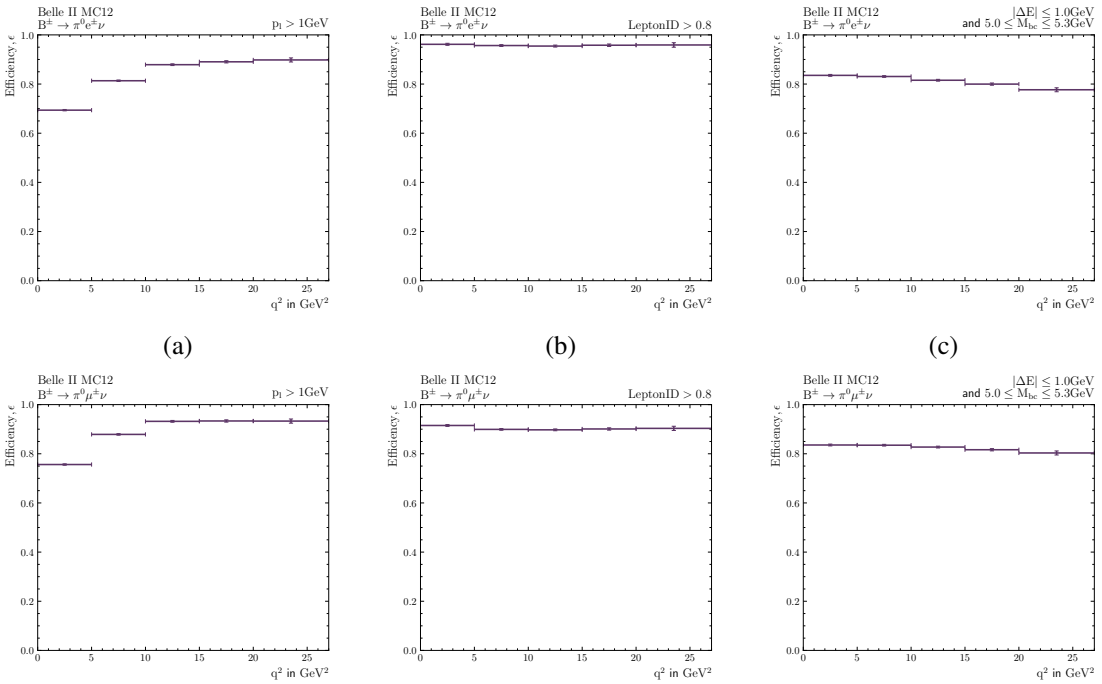


Figure C.18: 2D histograms of the FOM as a function of the continuum suppression and $B\overline{B}$ BDT output classifiers in the five q^2 bins for the B^0 modes. The optimal cuts are marked and labelled.

APPENDIX D

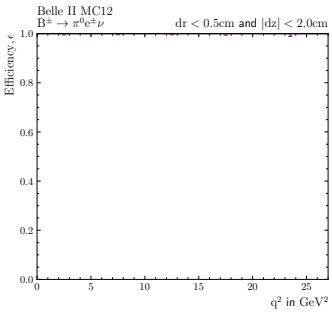
Efficiencies and Fit

D.1 Selection Efficiencies

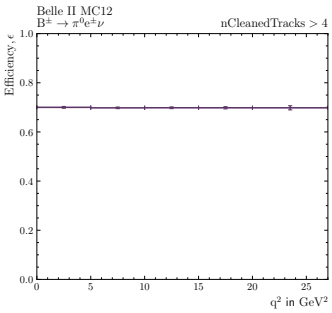


(a) (b) (c)

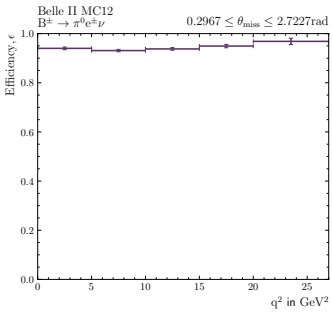
(d) (e) (f)



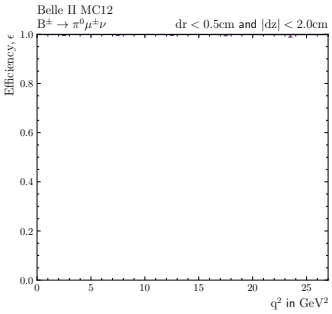
(g)



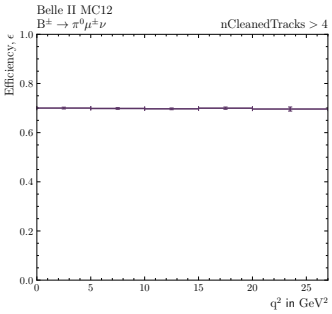
(h)



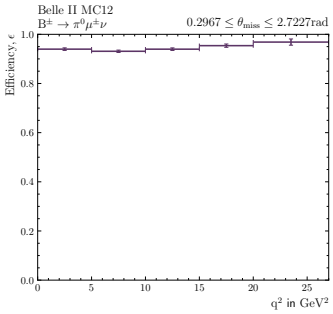
(i)



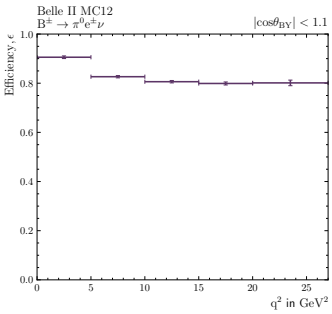
(j)



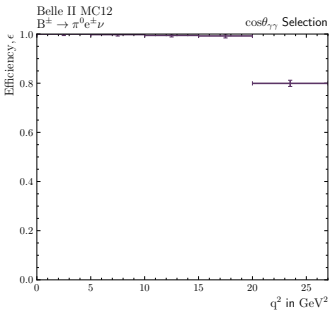
(k)



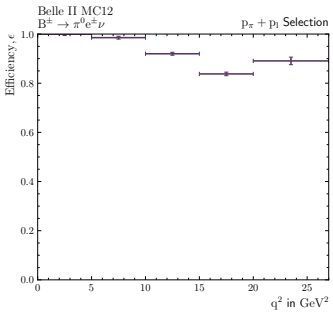
(l)



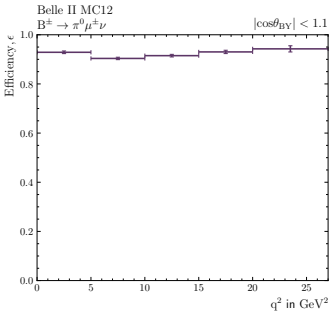
(m)



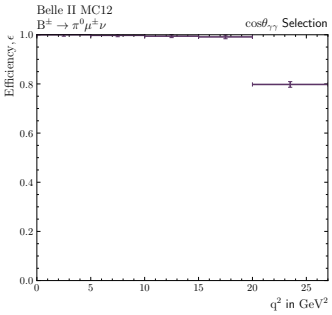
(n)



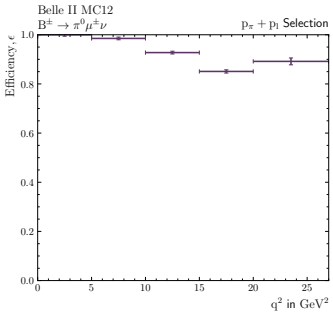
(o)



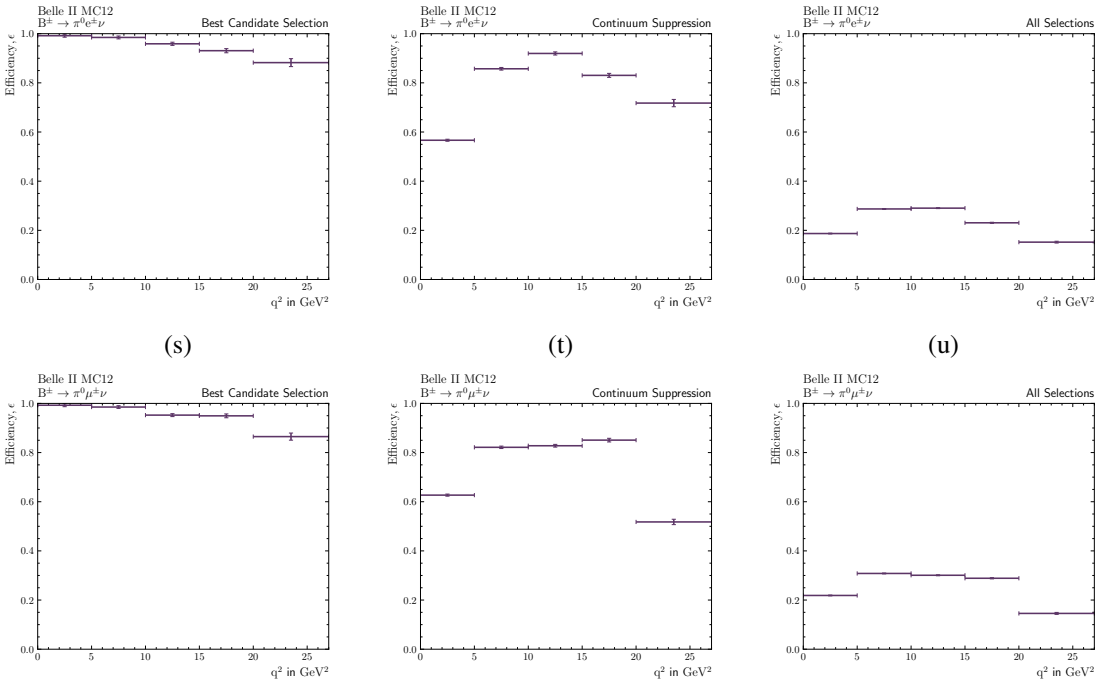
(p)



(q)



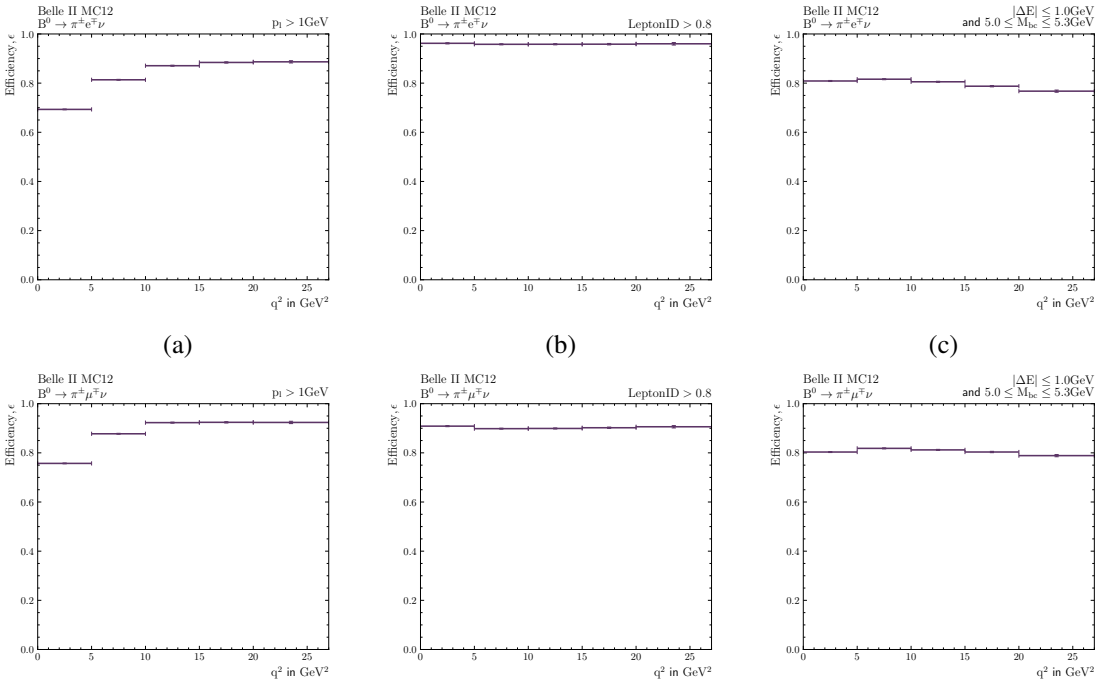
(r)



(s) (t) (u)

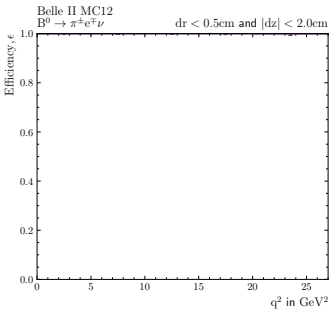
(v) (w) (x)

Figure D.1: Efficiencies across the five q^2 bins of the $B^{\pm}$ modes for all selections.

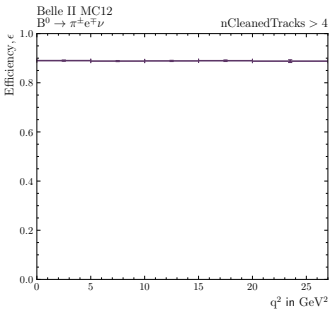


(a) (b) (c)

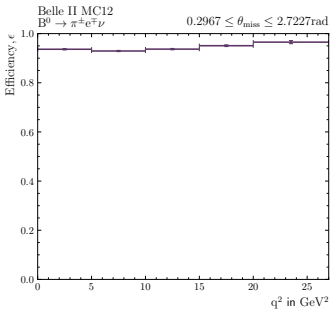
(d) (e) (f)



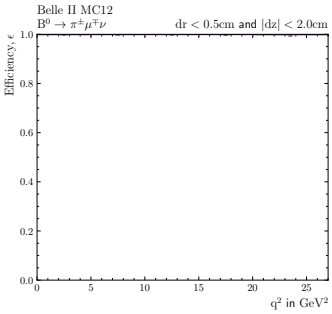
(g)



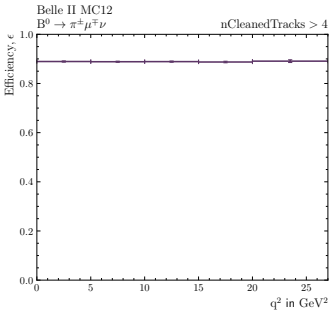
(h)



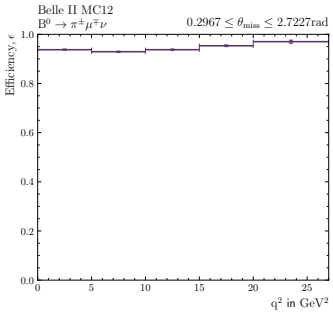
(i)



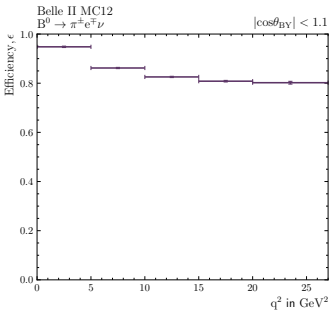
(j)



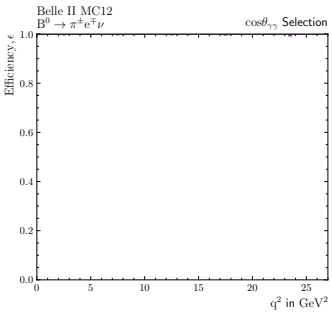
(k)



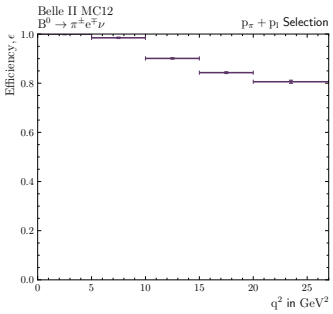
(l)



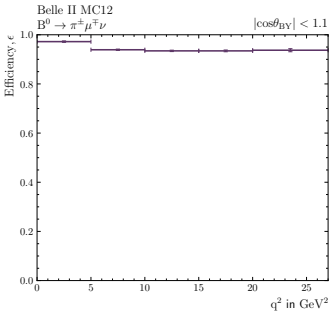
(m)



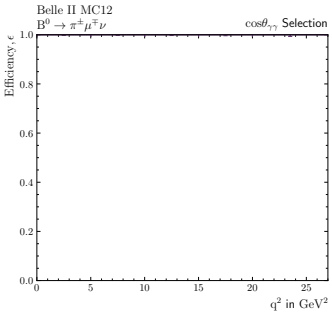
(n)



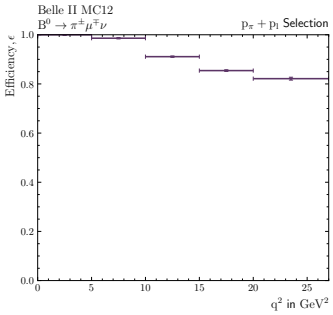
(o)



(p)



(q)



(r)

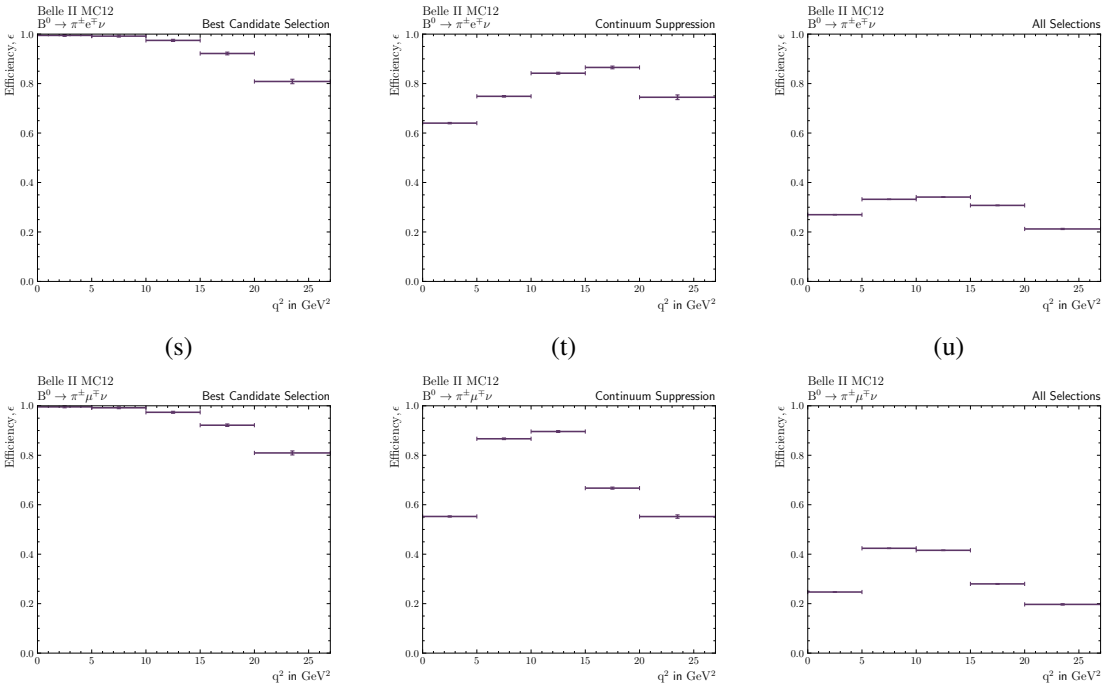


Figure D.2: Efficiencies across the five q^2 bins of the B^0 modes for all selections.

Selection	Efficiency				
	$q1$	$q2$	$q3$	$q4$	$q5$
p^*_ℓ	0.693	0.814	0.871	0.884	0.887
Lepton ID	0.962	0.958	0.958	0.958	0.960
ΔE and M_{bc}	0.809	0.816	0.805	0.787	0.767
dr and dz	0.999	0.999	1.000	0.999	0.999
nCleanedTracks	0.890	0.888	0.889	0.890	0.888
θ_{miss}	0.936	0.929	0.937	0.951	0.965
$\cos\theta_{BY}$	0.948	0.862	0.826	0.808	0.802
$p_\ell + p_\pi$	1.000	0.985	0.901	0.843	0.806
Best Candidate	0.994	0.993	0.975	0.925	0.800
Cont. Suppression	0.638	0.751	0.841	0.865	0.733

Table D.1: Efficiencies for the $B^0 \to \pi^\pm e\nu$ mode of the selections discussed and found from signal MC.

Selection	Efficiency				
	$q1$	$q2$	$q3$	$q4$	$q5$
p^*_ℓ	0.757	0.878	0.923	0.924	0.924
Lepton ID	0.909	0.898	0.899	0.902	0.906
ΔE and M_{bc}	0.803	0.818	0.812	0.803	0.789
dr and dz	0.999	0.999	1.000	1.000	0.999
nCleanedTracks	0.889	0.889	0.889	0.887	0.891
θ_{miss}	0.937	0.929	0.937	0.953	0.970
$\cos\theta_{BY}$	0.972	0.939	0.934	0.934	0.937
$p_\ell + p_\pi$	1.00	0.986	0.911	0.854	0.821
Best Candidate	0.995	0.992	0.973	0.922	0.809
Cont. Suppression	0.555	0.867	0.894	0.671	0.555

Table D.2: Efficiencies for the $B^0 \to \pi^\pm \mu\nu$ mode of the selections found from signal MC.

Selection	Efficiency				
	$q1$	$q2$	$q3$	$q4$	$q5$
p^*_ℓ	0.694	0.813	0.879	0.890	0.898
Lepton ID	0.962	0.957	0.954	0.958	0.959
ΔE and M_{bc}	0.835	0.831	0.815	0.800	0.777
dr and dz	1.000	1.000	1.000	1.000	1.000
nCleanedTracks	0.700	0.698	0.698	0.698	0.698
θ_{miss}	0.940	0.931	0.938	0.949	0.968
$\cos\theta_{BY}$	0.906	0.826	0.806	0.799	0.801
$\cos\theta_{\gamma\gamma}$	1.000	0.998	0.995	0.992	0.800
$p_\ell + p_\pi$	1.000	0.985	0.920	0.838	0.891
Best Candidate	0.993	0.983	0.957	0.937	0.881
Cont. Suppression	0.567	0.856	0.919	0.820	0.669

Table D.3: Efficiencies for the $B^\pm \to \pi^0 e\nu$ mode of the selections found from signal MC.

Selection	Efficiency				
	$q1$	$q2$	$q3$	$q4$	$q5$
p_ℓ^*	0.756	0.879	0.932	0.933	0.933
Lepton ID	0.915	0.899	0.897	0.901	0.903
ΔE and M_{bc}	0.836	0.835	0.827	0.816	0.803
dr and dz	1.000	1.000	1.000	1.000	0.999
nCleanedTracks	0.700	0.698	0.697	0.700	0.696
θ_{miss}	0.940	0.931	0.940	0.954	0.968
$\cos\theta_{BY}$	0.929	0.904	0.915	0.930	0.943
$\cos\theta_{\gamma\gamma}$	1.00	0.998	0.994	0.991	0.798
$p_\ell + p_\pi$	1.00	0.985	0.927	0.851	0.892
Best Candidate	0.993	0.985	0.951	0.947	0.890
Cont. Suppression	0.623	0.821	0.828	0.848	0.535

Table D.4: Efficiencies for the $B^\pm \to \pi^\pm \mu\nu$ mode of the selections found from signal MC.

D.2 Background Rejection

Selection	Mode			
	$B^0 \to \pi^\pm e\nu$	$B^0 \to \pi^\pm \mu\nu$	$B^\pm \to \pi^0 e\nu$	$B^\pm \to \pi^0 \mu\nu$
ΔE and M_{bc}	0.923	0.925	0.943	0.959
dr and dz	0.067	0.054	0.079	0.048
nCleanedTracks	0.385	0.318	0.450	0.439
θ_{miss}	0.125	0.201	0.117	0.197
$\cos\theta_{BY}$	0.820	0.778	0.867	0.819
$\cos\theta_{\gamma\gamma}$	0.000	0.000	0.279	0.310
$p_\ell + p_\pi$	0.344	0.301	0.326	0.262
Best Candidate	0.016	0.012	0.036	0.029
Cont. Suppression	0.794	0.894	0.767	0.888
Total	0.9991	0.9994	0.9997	0.9998

Table D.5: Background rejection of the selections in this thesis for all modes and averaged over the q^2 bins. The total background rejection is also given for each mode.

D.3 ΔE and M_{bc} Pre-Fit

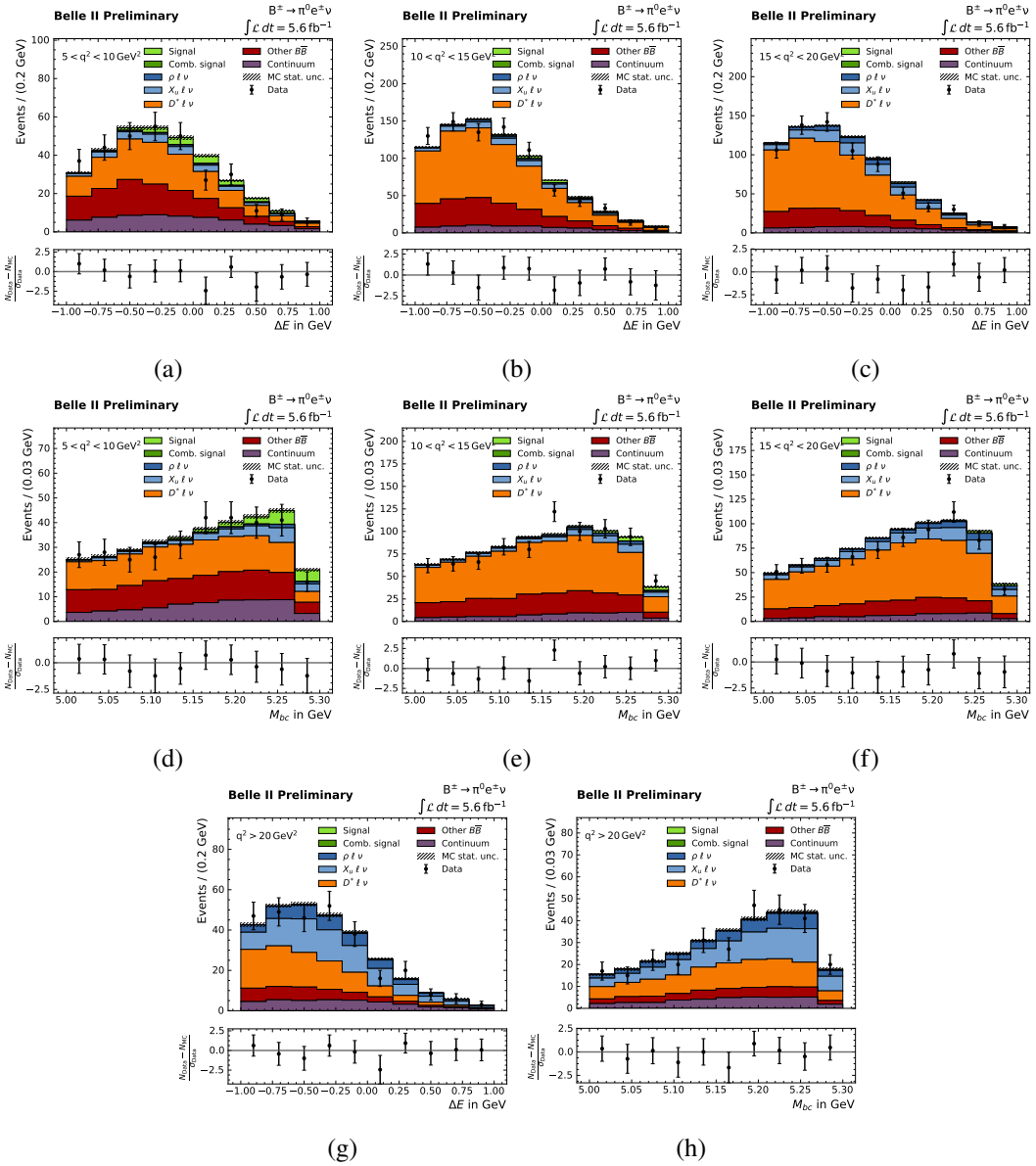


Figure D.3: Data-MC agreement for the five q^2 bins of the $B^{\pm} \to \pi^{\pm} e \nu$ mode for ΔE and M_{bc}.

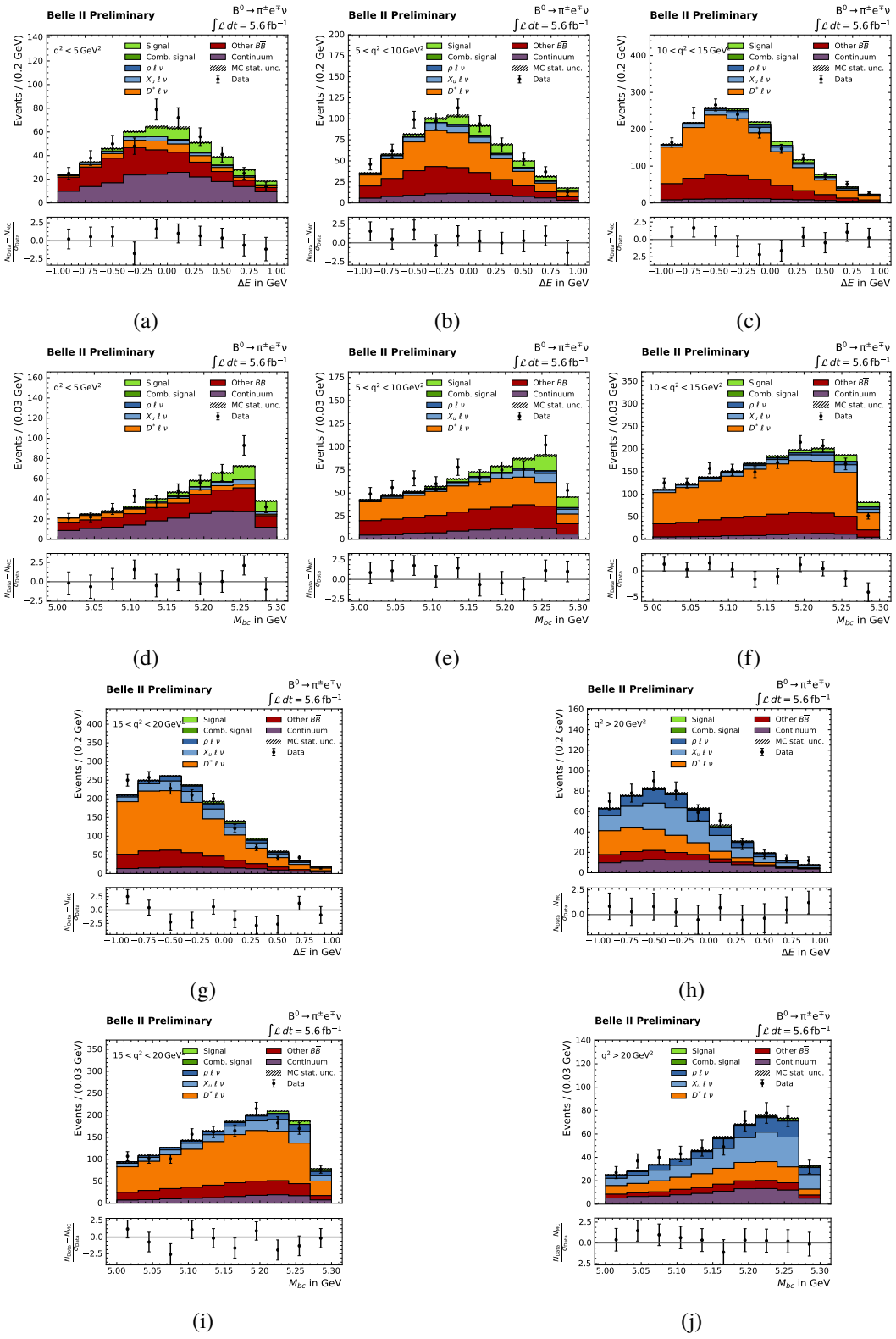


Figure D.4: Data-MC agreement for the five q^2 bins of the $B^0 \to \pi^\pm e\nu$ mode for ΔE and M_{bc}.

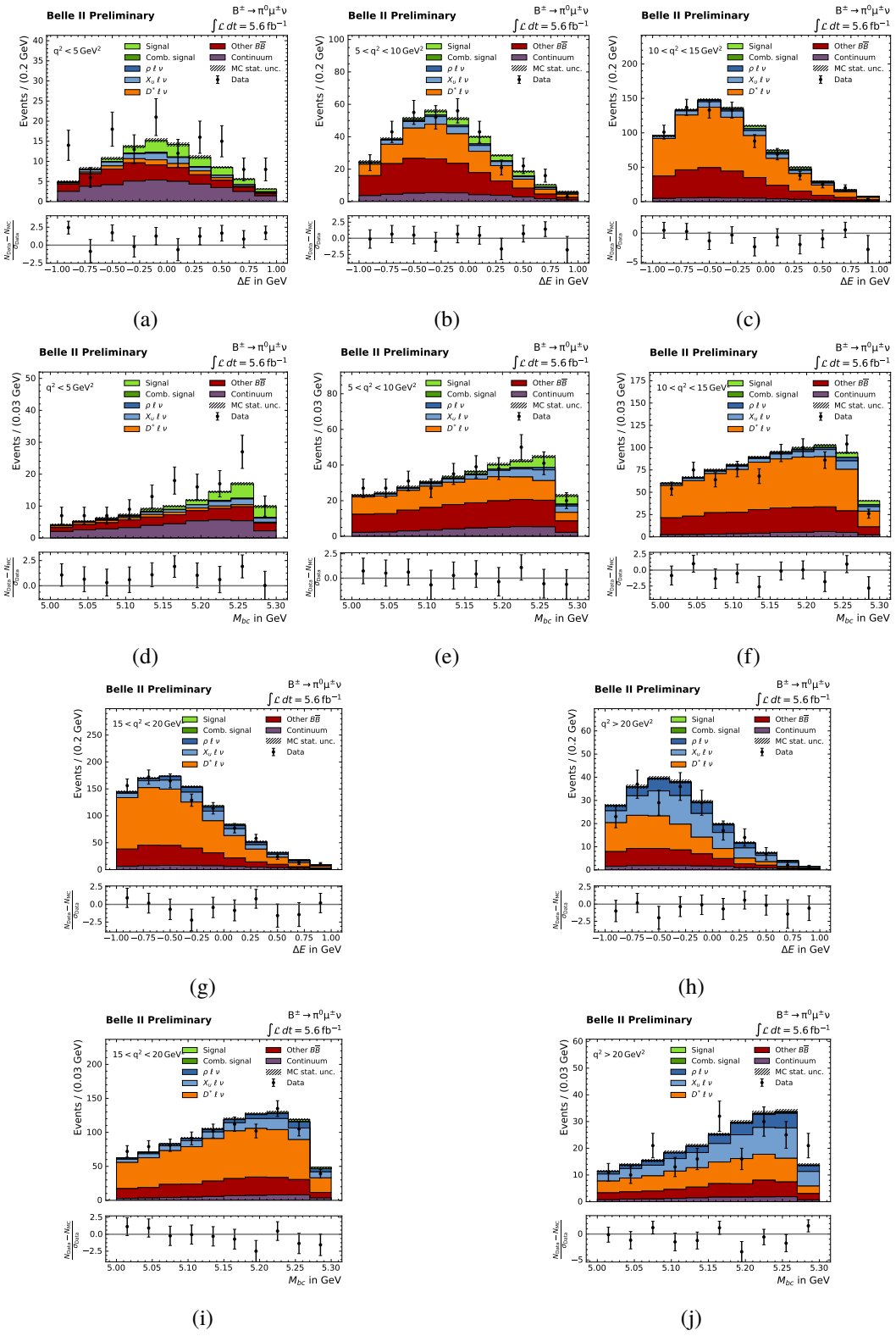


Figure D.5: Data-MC agreement for the five q^2 bins of the $B^{\pm} \to \pi^0 \mu \nu$ mode for ΔE and M_{bc}.

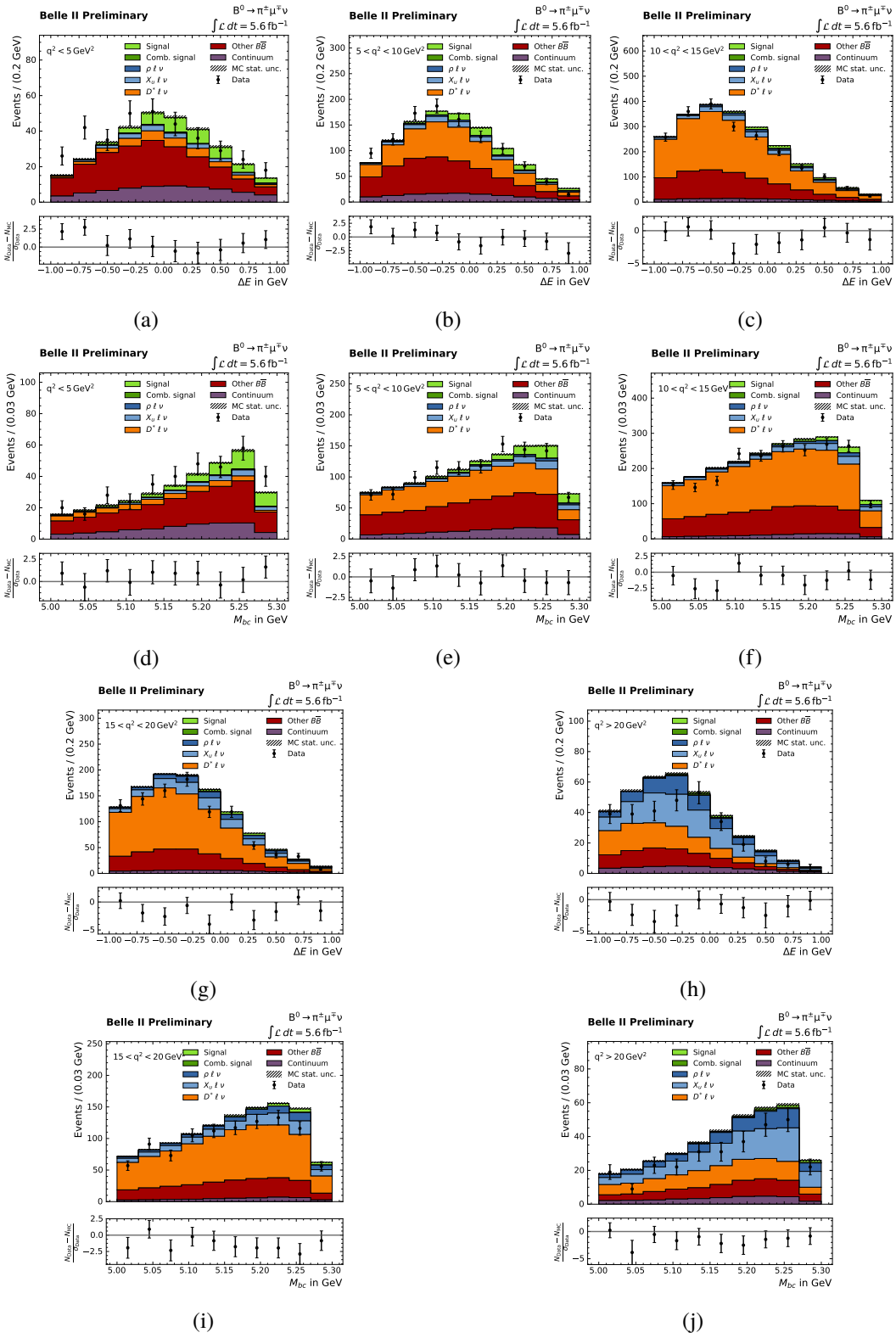


Figure D.6: The data MC agreement for the five q^2 bins of the $B^0 \to \pi^\pm \mu \nu$ mode for ΔE and M_{bc}.

D.4 Pull Distributions

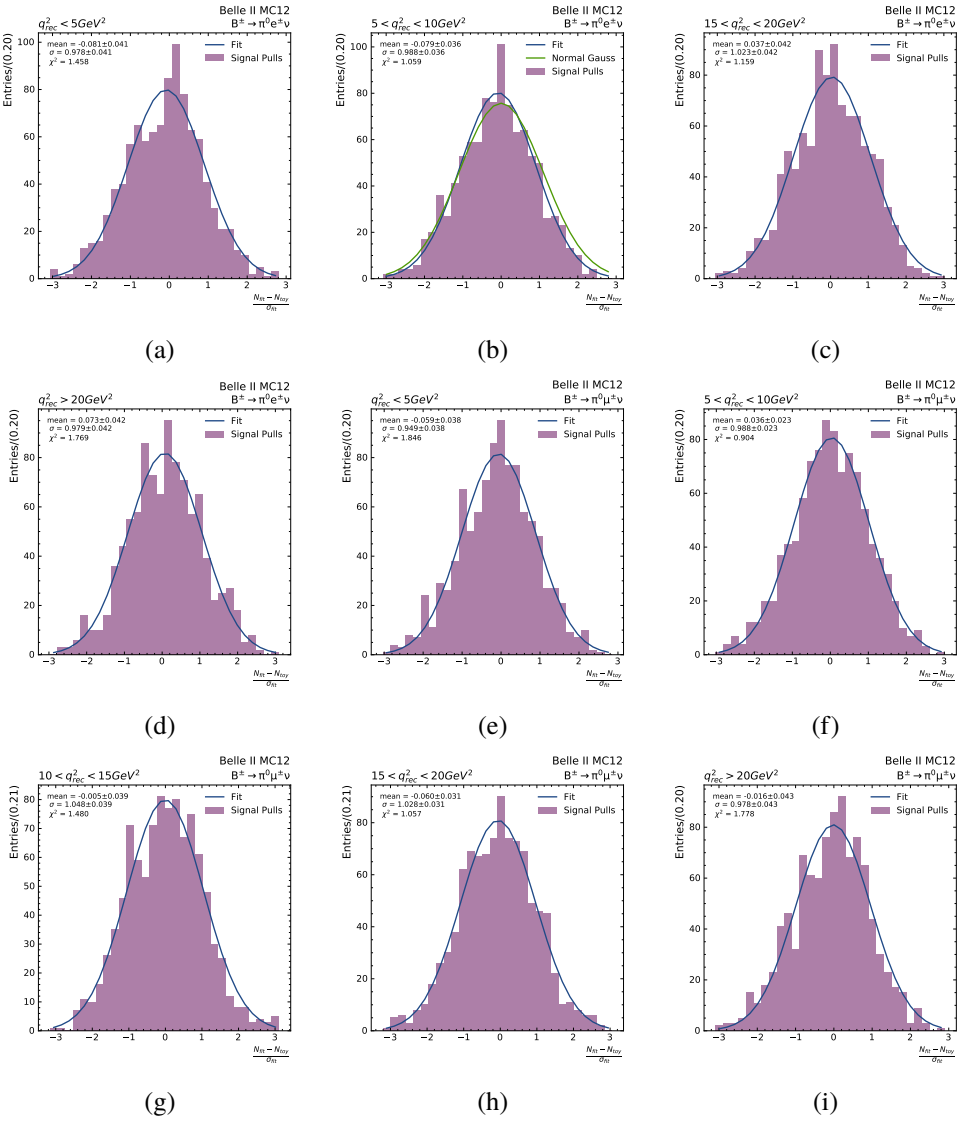


Figure D.7: Pull distributions of the five q^2 bins in the $B^\pm$ modes.

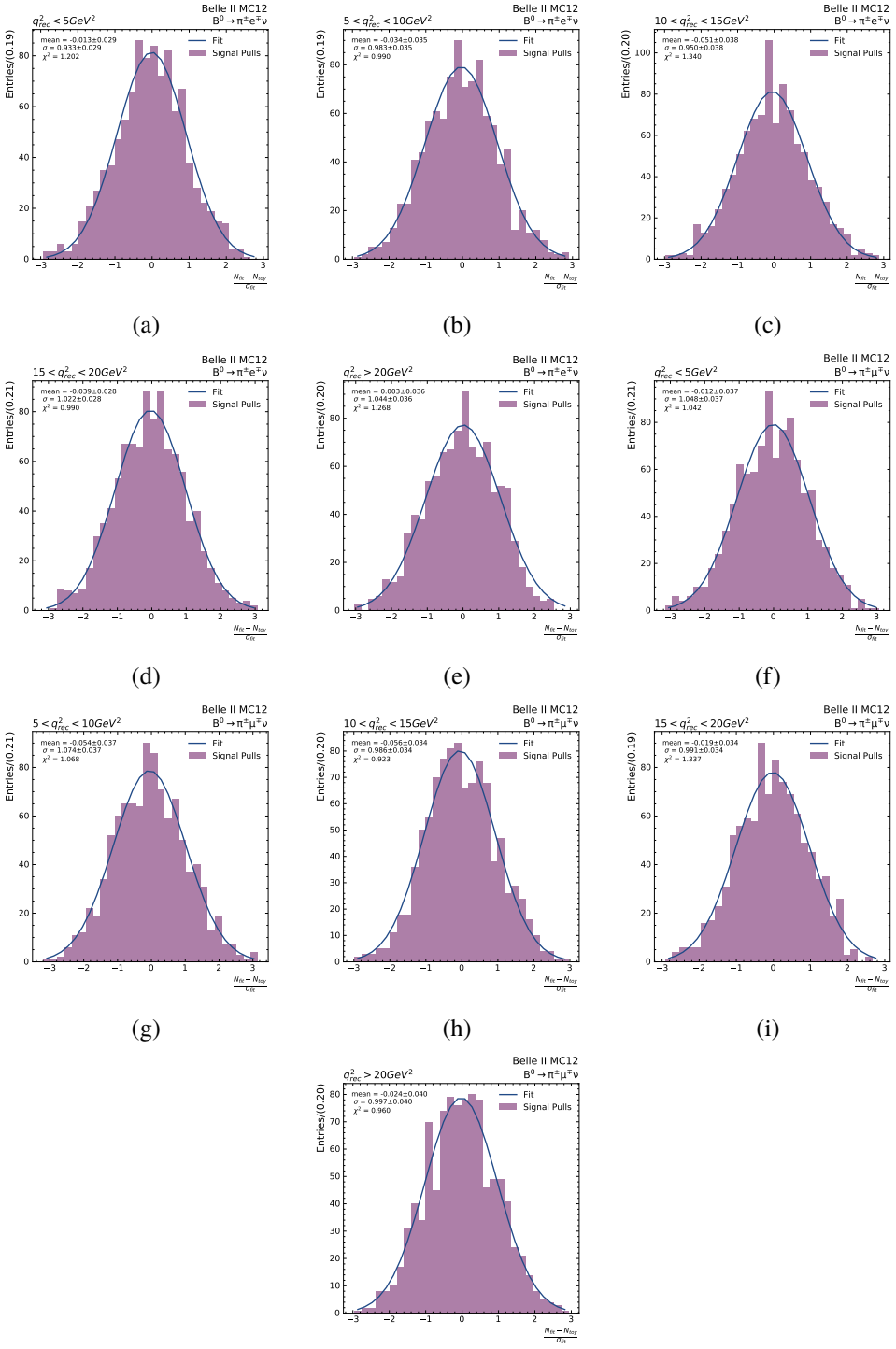


Figure D.8: Pull distributions of the five q^2 bins in the $B^\pm$ modes.

List of Figures

List of Tables