Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Sensitivity Study of $R(D^{(*)})$ with an Inclusive Tagging Approach

by
Tia Crane

Masters Thesis
at

Faculty of Mathematics, Informatics and Natural Sciences
Department of Physics
Universität Hamburg

March 16, 2026

Study Program:	M.Sc. Physics
Student ID:	7714053
First Supervisor:	Prof. Dr. Kerstin Tackmann
Second Supervisor:	Dr. Alexander Glazov

This version has been edited from the originally submitted thesis for clarity and completeness. No additional studies performed after the submission date are included.

Physics Department: Affidavit for final theses

The following formulation including signature must be included in your thesis.

Statutory Declaration

I declare that I have authored the Bachelor's/Master's thesis independently and have not used any aids other than those specified. I have clearly labelled all passages that are taken from other works in terms of wording or meaning as borrowed material, stating the exact source in each case. I confirm that I have not already submitted this thesis in the same or a similar version in another examination procedure.

If electronic aids based on generative artificial intelligence (gAI) were used in the course of the preparation of this Bachelor's/Master's thesis, I confirm that my own work was the main focus and that complete documentation of all aids used is available in accordance with good scientific practice. I am responsible for any incorrect or distorted content, incorrect references, violations of data protection and copyright law or plagiarism generated by the gAI.

I agree/disagree* that the Bachelor's/Master's thesis* may be published.

Hamburg, 14.12.2025

Place, date Signature

Delete or omit as appropriate.

Abstract

This thesis presents the first simultaneous analysis of $R(D)$ and $R(D^*)$ using inclusive tagging with Run I data at Belle II where a total of $275 fb^{-1}$ is used to obtain the reported sensitivity estimation. Current experimental measurements show a combined tension of 3.8σ with respect to the Standard Model, indicating an enhanced coupling of the charged weak $W^{\pm}$ current to the third lepton generation. Previous measurements relied on a fully reconstructed tagging method, in which both B mesons are reconstructed. This approach provides high resolution in kinematic observables but suffers from low signal efficiency, thus causing the reported values to be statistically limited.

In an attempt to increase statistical sensitivity, the inclusive tagging method is used. Inclusive tagging reconstructs only the signal B meson while allowing the companion B to decay freely. This increases signal efficiency at the cost of lower purity and smeared kinematic distributions. To address these challenges, two successive boosted decision trees (BDTs) are applied to reconstructed $B^{\pm} \to D^0 \mu \nu$ samples. The ratio $R(D^{(*)})$ is then extracted directly in bins of the second BDT and the momentum transfer q^2.

The blinded preliminary results presented here provide a conservative estimate of the potential statistical sensitivity, with 23% for $R(D)$ and 9% for $R(D^*)$, motivating the measurement of $R(D^{(*)})$ using inclusive tagging in combination with a more complete reconstruction of $B \to D\ell\nu$ decays.

Contents

List of Figures

List of Tables

1 Introduction

1.1 The Standard Model and Charged Current Flavour Physics

The Standard Model (SM) of particle physics provides a theoretical framework describing the fundamental particles and their interactions. The current formulation is based on the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge symmetry, which governs the strong, weak, and electromagnetic interactions. These fundamental interactions are mediated by the spin-1 gauge bosons: eight massless gluons, one massless photon, and three massive bosons, $W^\pm$ and Z^0. Each of these gauge bosons can be thought of as a degree of freedom introduced to preserve local gauge invariance through the covariant derivative, $D_\mu = \partial_\mu + igA_\mu$, where A_μ represents the respective gauge boson field. Properly accounting for this in the Lagrangian results in the addition of a new interaction term:

$$\mathcal{L} = \underbrace{\bar{\psi} i\gamma^\mu \partial_\mu \psi}_{\text{kinetic term}} - \underbrace{m\bar{\psi}\psi}_{\text{mass term}} - \underbrace{g\bar{\psi}\gamma^\mu \psi A_\mu}_{\text{interaction term}} , \tag{1}$$

where an arbitrary constant g is introduced. As g only appears in the interaction term, it is known as the coupling strength and designates how strongly the Dirac field ψ interacts with the gauge field A_μ. The covariant derivative appears in the Lagrangian for all fields charged under the gauge group, thus providing one *universal coupling strength*. [37] [34]

Of particular interest is the coupling strength between the charged weak current, ascribed by the exchange of the $W^\pm$ gauge bosons, and Dirac fermions ψ. Fermions appear in three generations, each of which follows the symmetry structure of local gauge invariance, exhibiting identical couplings with the $W^\pm$ bosons. Taking the tree-level $b \to c\ell\nu$ transition depicted in Figure 1.1, the transition probability, ascribed by the matrix element $\mathcal{M}$, may be written as:

$$\mathcal{M}(b \to c\ell\nu) \approx \frac{g}{\sqrt{2}} V_{cb}\, j^\mu \frac{-g_{\mu\nu} + q_\mu q_\nu / M_W^2}{q^2 - M_W^2} \frac{g}{\sqrt{2}} j^\nu , \tag{2}$$

where g represents the *universal coupling strength* of the $W^\pm$ bosons to a pair of fermions (quarks or leptons), $q_{\mu,\nu}$ is the associated four-momentum transfer, j^μ and j^ν are the hadronic quark ($b \to c$) and leptonic ($\bar{\nu}_\ell \to \ell^-$) currents, and V_{cb} is the associated Cabibbo–Kobayashi–Maskawa (CKM) matrix element. The CKM rotation matrix maps the quark mass-eigenstates onto the weak-eigenstates, providing proper adjustment to the transition probability between inter- or intra-generational quarks. [25] [37]

With the SM assumption of massless neutrinos, there is no such rotation matrix to modify the leptonic couplings. Therefore, the SM predicts unaltered universal couplings between all three lepton generations and the $W^\pm$ bosons.

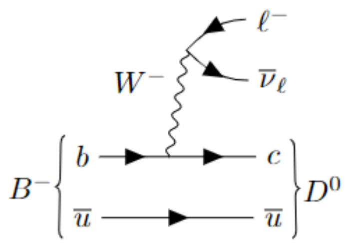


Figure 1.1: Semi-leptonic B decay.

Figure 1.1 not only displays the quark level $b \to c\ell\nu$ transition but the complete tree-level semi-leptonic decay of the charged B meson. Though the momentum transfer is relatively small ($q^2 << M_W^2$), the large $W^\pm$ mass (~80 GeV) induces a point-like interaction. Due to the short time and short range associated with the point-like interaction, the light $\bar{u}$ quark does not participate and can be treated as a passive spectator. This allows the decay to be factorized into hadronic and leptonic currents. The leptonic current is well-known, while the hadronic current may be parameterized with form factors, which encompass all the non-perturbative QCD effects in the interaction. [37]

These theoretically clean semi-leptonic B decays are also experimentally accessible (Section 2.1), making them a powerful probe of the leptonic couplings to the $W^\pm$ bosons. In particular, ratios of branching fractions:

$$R(D^{(*)}) = \frac{\mathcal{B}(B \to D^{(*)}\tau\nu)}{\mathcal{B}(B \to D^{(*)}\ell\nu)}\,, \quad \ell = e, \mu \tag{3}$$

provide a sharp observable of SM predictions as a large number of uncertainties associated with detector effects and hadronic currents cancel out.

1.2 Motivation and Context in Flavour Physics

Each of the three lepton generations carries a distinct flavour, which — owing to the absence of intra-generational couplings — is conserved in the weak interaction. Together with the universal coupling of the $W^\pm$ boson to all charged leptons, this leads to the principle of lepton-flavour universality (LFU): the assertion that charged-current weak interactions treat all lepton flavours identically except for kinematic effects from their masses. A particularly sensitive probe of LFU is provided by the two-dimensional observables $R(D)$ and $R(D^{(*)})$ (Equation 3).

Any deviation from the SM predictions for $R(D)$ and/or $R(D^*)$ would strongly indicate the need to modify the charged-current structure or introduce new interactions. Importantly, $R(D)$ and $R(D^*)$ serve as partially independent LFU observables, allowing different classes of new physics to be distinguished. An enhancement observed predominantly in $R(D)$ would suggest SM extensions such as those realized in two-Higgs-doublet models. Examples of which include the addition of a charged Higgs, which not only induces effects on $R(D^{(*)})$ but also the $\tau^{\pm}$ polarization [16]. Conversely, a simultaneous enhancement in both $R(D)$ and $R(D^{*})$ typically points to new physics models such as leptoquarks (LQs), which mediate interactions between leptons and quarks. Proposed models not only induce effects $R(D^{(*)})$ but predict sizable branching fractions for $b \to s\tau^{+}\tau^{-}$ [19].

1.3 Overview of Experimental Status

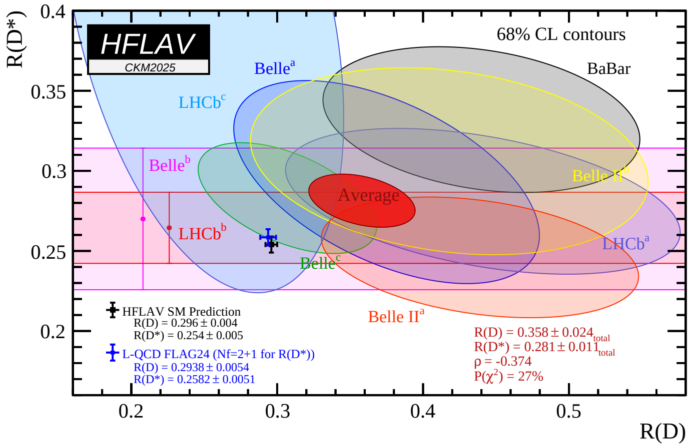


Figure 1.2: Current 68% confidence level (CL) contours for individual experimental results and HFLAV averages of $R(D^{(*)})$. [26]

Recent measurements display enhancement in the third-generation lepton coupling to the charged weak $W^{\pm}$ current. Averaging the independent experimental results yields:

$$R(D) = 0.281 \pm 0.011 \tag{4}$$
$$R(D^{*}) = 0.358 \pm 0.024\,, \tag{5}$$

which provide an excess of 2.5σ and 2.3σ from the (averaged) SM predictions:

$$R(D) = 0.296 \pm 0.004 \tag{6}$$
$$R(D^{*}) = 0.254 \pm 0.005\,. \tag{7}$$

Including the experimental average correlation between $R(D)$ and $R(D^{*})$ of $\rho = -0.374$, the resulting tension with the SM prediction is around 3.8σ. [26]

All reported measurements to date, require the explicit reconstruction of both sides of the $\Upsilon(4S) \to (B_{\rm sig} \to D^{(*)}\tau^{\pm}\nu)\,(B_{\rm tag} \to X)$ decay. Focusing on the recent Belle II results (Table 1.1), the yellow contour in Figure 1.2 reconstructs a semi-leptonic (SL)

B_{tag} decay and a leptonic $\tau^{\pm} \to \ell\nu\bar{\nu}$ decay [2]. The resulting measurement pushed the tension from 3.3σ to 3.8σ. The orange contour in Figure 1.2 results from reconstruction of a hadronic B_{tag} decay with a leptonic $\tau^{\pm}$ decay [39], which kept the current tension at 3.8σ. Additionally, on-going Belle II analyses include measurement of $R(D^*)$ and $\tau^{\pm}$ polarization with reconstruction of a hadronic B_{tag} paired with a one-prong hadronic $\tau^{\pm}$ decay. Discussions on tagged approaches are not pertinent to the remainder of this thesis and are thereby not discussed further. [26] [29]

$R(D)/R(D^*)$	Tension	B Tagging	τ Decays	References
$R(D) = 0.418 \pm 0.074 \pm 0.051$ $R(D^*) = 0.306 \pm 0.034 \pm 0.018$ $\rho = -0.22$	1.7σ	SL	$\tau \to \ell\nu\bar{\nu}$	[2]
$R(D) = 0.439 \pm 0.055 \pm 0.046$ $R(D^*) = 0.242 \pm 0.019 \pm 0.016$ $\rho = -0.20$	1.5σ	Hadronic	$\tau \to \ell\nu\bar{\nu}$	[39]

Table 1.1: Recent Belle II results for simultaneous $R(D^{(*)})$ analyses.

This thesis provides a sensitivity study of an inclusive B_{tag} where the available phase space is opened up by not requiring an explicit reconstruction of the B_{tag} decay. Inclusive tagging associates the sum of all remaining tracks and clusters, not associated with the reconstructed B_{sig}, with the companion "B_{tag}" meson, denoted as B_{comp} from this point forward. The resulting sensitivity of an inclusive-tag analysis is projected to exceed that of the combined sensitivity of hadronic and semi-leptonic tagged approaches. This is because the inclusive method not only accepts both semi-leptonic and hadronic decays, but is also not limited to the sum of the conventionally reconstructed B_{tag} decay.

2 Experimental and Statistical Foundations

2.1 B-Physics

In the early 1970s, it became increasingly clear that elementary fermions were organized into families of two quarks and two leptons. The discovery of the heavy third-generation lepton, the $\tau^{\pm}$, in the mid-1970s strongly suggested the existence of a corresponding third-generation quark doublet. Theoretical additions of a third quark generation were formalized shortly thereafter by Makoto Kobayashi and Toshihide Maskawa, who formulated a 3×3 quark-mixing matrix — now called the Cabibbo–Kobayashi–Maskawa

(CKM) matrix. Their framework showed that a complex phase, capable of generating a naturally arising CP violation mechanism, is only possible when three quark generations are present. [13]

The bottom quark, b, was experimentally discovered soon after in 1977 through the observation of the Υ resonances at Fermilab. These states were later confirmed to correspond to bound $b\bar{b}$ states, and in 1983 the CLEO experiment definitively identified B mesons containing a heavy b quark paired with a lighter quark. Unlike the extremely short-lived top quark, the b quark is sufficiently long-lived to form mesons whose weak decays provide a rich environment for flavor and CP-violation studies. [13]

Modern B-factories exploit this by operating at the $\Upsilon(4S)$ resonance, just above the threshold for producing $B\bar{B}$ pairs (Figure 2.1). The corresponding decay yields an exceptionally clean and well-understood sample of B mesons. Dominant backgrounds arise from continuum production of light-quark pairs ($q\bar{q} = u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$), which subsequently hadronize into collimated jet-like topologies. These jet-like topologies enable powerful separation from the nearly spherical $B\bar{B}$ topologies, thereby permitting strong continuum background suppression with event-shape observables. Any residual continuum background can be studied through the use of the off-resonance region (Figure 2.1).

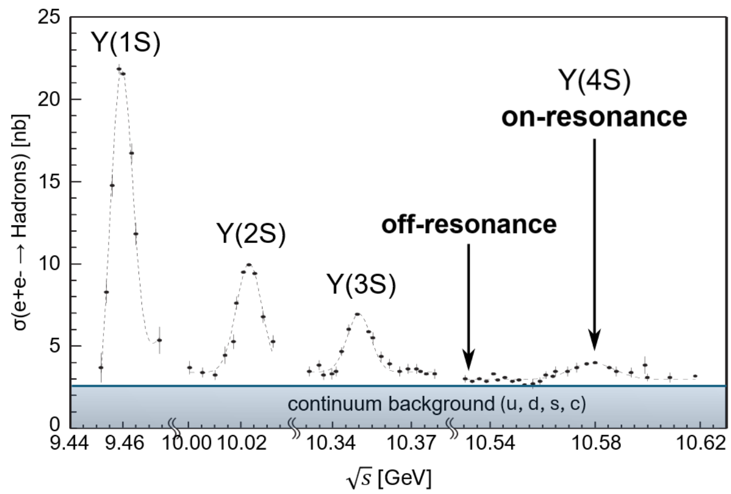


Figure 2.1: Cross section of Υ resonances. Modified from [6].

Provided the completion of proper continuum background removal and/or modeling studies, the clean $\Upsilon(4S)$ environment enables precision studies of rare or experimentally challenging B decays. Including those with multiple unobserved neutrinos, such as:

$$B \to D(\tau \to \ell\nu\bar{\nu})\nu \tag{8}$$

A defining feature of B-factories such as Belle II is the ability to fully reconstruct one of the two produced mesons (B_{tag}) while analyzing the decay of the other (B_{sig}). This strategy not only enabled observations of CP violation in the B-system but also opened a new era of high-precision flavor physics. In particular, decays such as Equation 8 — which have smeared kinematics due to the three undetected neutrinos — remain accessible with competitive precision.

2.2 The Belle II Experiment

2.2.1 SuperKEKB

SuperKEKB is the upgraded successor to the KEKB asymmetric-energy e^+e^- collider located in Tsukuba, Japan. Its design objective is to increase luminosity such that it maximizes the production rate of $B\bar{B}$ pairs at the $\Upsilon(4S)$ resonance. The instantaneous luminosity can be expressed schematically as:

$$L = \frac{N_b n_{e^-} n_{e^+} f}{A_{eff}} \tag{9}$$

where N_b is the number of bunches, $n_{e^\pm}$ the bunch populations, f the revolution frequency, and A_{eff} the effective transverse overlap area of the colliding beams. SuperKEKB achieves its target luminosity primarily through the use of large horizontal crossing angles at the interaction point (IP), a.k.a. "crab crossing", which provides an aggressive reduction of A_{eff}. [1] [4]

In addition to increasing the overall yield of $B\bar{B}$ events, the resulting small interaction region enhances the precision of vertex reconstruction. A well-constrained interaction point reduces uncertainties in observables that depend on missing energy or missing momentum, thereby improving the sensitivity of analyses involving final-state neutrinos. This capability is especially significant for semi-leptonic processes, such as $R(D^{(*)})$, in which multiple neutrinos render kinematic reconstruction intrinsically challenging.

2.2.2 Belle II Detector

Surrounding the IP is the Belle II detector, a multilayered apparatus made up of several specialized subdetectors (Figure 2.2). Belle II uses a right-handed Cartesian coordinate system where the z-axis is defined along the electron beam line. In Figure 2.2, the electron beam travels horizontally from left to right through the center of the detector, so the $+z$ direction corresponds to the right-hand side of the image. In this top-view configuration, the $+y$-axis points out of the page and the $+x$-axis points toward the top of the page.

The polar angle θ measures the angle of a particle's trajectory relative to the beam line. In this thesis, tracks are considered to have a detectable trajectory if they exhibit a polar angle within the range of $17° < \theta < 150°$, such that they traverse the primary tracking subdetector, the central drift chamber (CDC). The azimuthal angle ϕ describes the direction of the trajectory around the beam line, measured with respect to the x-axis. Finally, the transverse momentum p_T is the component of a particle's momentum projected onto the xy-plane.

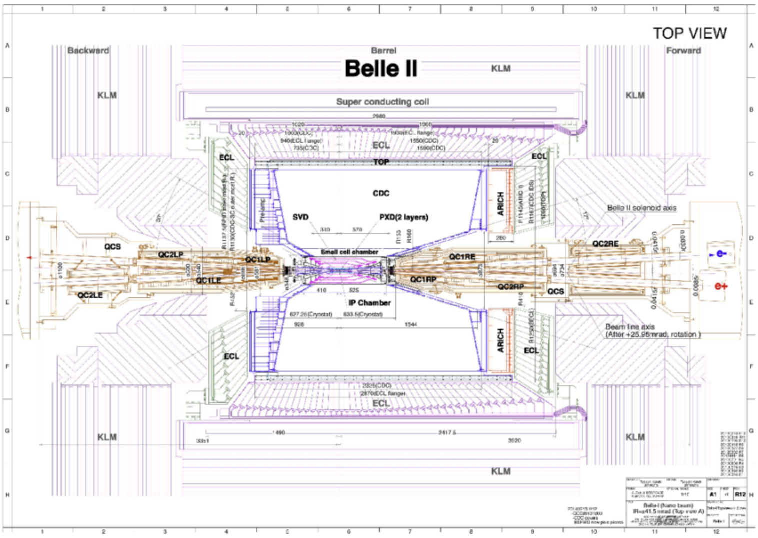


Figure 2.2: Schematic diagram of the Belle II detector. [18]

Tracking Detectors

Vertex Detectors (VXD): The vertex detector system, spanning radii from 14 mm to 140 mm, is crucial for reconstructing event kinematics by providing precise positions for vertex fits (Section 3.2.1). To achieve increased granularity near the IP and mitigate effects from overlapping particle hits (pile-up), the inner pixel detector (PXD) provides two-dimensional measurements with silicon pixels. At larger radii, the probability of having pile-up significantly decreases. Therefore, the outer vertex detector, the silicon strip detector (SVD), uses orthogonal silicon strips to provide one-dimensional position measurements. [1]

Central Drift Chamber (CDC): Following the vertex detectors is the main tracking detector, the CDC, which provides momentum and spatial information through a twisted wire configuration, suspended in $\mathrm{He} - \mathrm{C_2H_6}$ gas [1]. Charged particles traversing the gas ionize electrons along their path. The resulting free electrons further ionize electron-hole pairs as they drift toward the anode, causing an avalanche. The charge accumulation in the avalanche provides an amplified signal current, associated with a discrete "hit" in the detector. The hits are combined to extrapolate the track's trajectory and determine the incident particle's momentum $p = B[T]R[m]$, where R is the bending radius.

Low-Momentum Particle Identification

Central Drift Chamber (CDC): Beyond providing charged-particle tracking and momentum measurements, the CDC contributes to particle identification at low momentum through ionization energy loss, dE/dx (Figure 2.3). At momenta of order $\mathcal{O}(0.1)$ GeV, the Bethe–Bloch curves for different particle species are well separated, enabling effective discrimination. At higher momenta (>1 GeV), however, the ionization bands converge, and the dE/dx information provides minimal separation power.

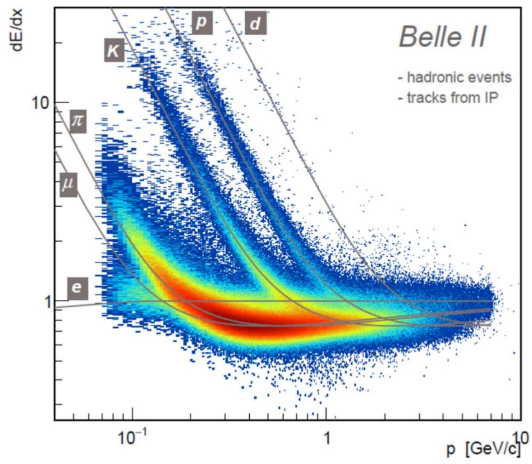


Figure 2.3: Ionization energy loss in CDC. [40]

High-Momentum Particle Identification

To combat the lack of discrimination power in high-momentum tracks, Cherenkov-based detectors are used in both the barrel and forward regions:

Time of Propagation Detector (TOP): The TOP detector achieves PID for high-momentum tracks in the barrel region through the use of a fused silica (quartz) radiator. [1] Upon interaction of an incident particle, Cherenkov photons are emitted with a Cherenkov angle θ_c dependent on the velocity:

$$\cos\,\theta_c = \frac{1}{n\beta}\ , \tag{10}$$

where n is the refractive index. The emitted photons are internally reflected within the quartz bar until reaching a readout photo-multiple tube (PMT) at its edge. Information on the required time for the Cherenkov photons to reach the PMT is combined with tracking information from previous subdetectors to determine θ_c, thereby providing PID information based on the resulting velocity measurement (Section 2.2.3).

Aerogel Ring-Imaging Cherenkov Detector (ARICH): The aerogel radiator, configured in the forward endcap, provides velocity-based PID information for incident particles. Upon interaction, the particles emit a ring of Cherenkov photons which expands before being detected by a sensor plate. To balance the need for further photon emission and single photon focusing, a non-homogeneous (dual-layer) radiator is used. The resulting emission of two overlapping rings reduces uncertainties related to the point of emission. [1] Upon detection, the radius of the Cherenkov rings r is measured to provide a velocity estimation, and thereby PID (Section 2.2.3), through a modified form of Equation 10:

$$\beta = \frac{1}{n\,\cos\theta_c} \approx \frac{1}{n\,\cos(2r/R)}\ , \tag{11}$$

where R is the distance between the aerogel radiator and the sensor plate.

Electromagnetic Calorimeter (ECL)

The electromagnetic calorimeter (ECL), constructed from homogeneous CsI(Tl) crystals, provides precise energy measurements for electrons and photons. These electromagnetic particles initiate well-contained showers with limited longitudinal diffusion, allowing the deposited energy to be reconstructed with high accuracy.

At SuperKEKB's luminosity, several backgrounds complicate the interpretation of ECL clusters. Beam-induced backgrounds generate numerous low-energy "fake" photons with broad and nearly flat timing distributions. Because genuine electromagnetic clusters are time-coincident with the bunch crossing and peak sharply at $t_0 \approx 0$, timing information provides a powerful handle to suppress such contamination. A second major source of background arises from Bhabha scattering ($e^+e^- \rightarrow e^+e^-(\gamma)$), which produces extremely energetic electrons that may partially leak into the outer KLM system. This leakage leads to degraded shower shapes and biased energy estimates that can mimic hadrons or introduce misidentified electron candidates. To reduce effects from poorly measured electrons, many analyses place requirements on the measured energy resolution. Finally, secondary photons produced in strong interactions between hadrons and the detector material ("hadronic split-off") can result in further "fake" photon clusters. These tend to exhibit smaller spacial separations between the cluster candidate and the closest reconstructed charged trajectory. In contrast, genuine signal photons are typically well isolated, enabling effective suppression of hadronic split-off. [6]

K_L and $\mu^\pm$ System (KLM)

The outermost subdector is the KLM, which not only serves as a return yolk for the magnetic field generated by the solenoid but also detects long-lived hadrons and muons. The KLM detector is composed of 14 (15) layers of highly dense iron absorbers which are sandwiched with glass-electrode resistive plate chambers (scintillators) that act as active detectors in the barrel (endcaps). Upon interaction of an incident particle, Upon interaction of long-lived incident particles, such as the K_L and $\mu^\pm$, hadronic showers are produced in the iron absorbers and detected in the active layers. These showers are reconstructed as KLM clusters, where clusters within 5° of one another are grouped into a single cluster. [1]

The KLM provides the primary means of $\mu^\pm$ and K_L^0 identification. $\mu^\pm$ identification proceeds by extrapolating a charged track through the PXD, SVD, and CDC with a pion mass hypothesis. If a KLM cluster is found near the track, the consistency between the range of the predicted and extrapolated trajectory determines if the track is identified as a muon. Although the KLM reconstruction algorithms are highly optimized, performance limitations remain due to the close similarity between pion and muon masses, energy losses, and ranges. Therefore, muon candidates tend to have relatively low purity and detection efficiency.

2.2.3 Particle Identification (PID)

Final state candidates ($e^{\pm}$, $\mu^{\pm}$, $K^{\pm}$, $\pi^{\pm}$, p, d) are distinguished by exploiting their characteristic interactions with the subdetectors described above. Each subdetector provides information that is encoded into a likelihood $\mathcal{L}_i^{\text{subdec}}$ for a given mass hypothesis i. A combined likelihood may then be formulated by taking the product over all contributing subdetectors. From this, a likelihood ratio (PID) is constructed to give a quantitative measure of the probability that a track corresponds to hypothesis i:

$$\mathcal{P}_i = \frac{\mathcal{L}_i}{\mathcal{L}_i + \sum_{j \neq i} \mathcal{L}_j} \, . \tag{12}$$

For a global PID, all remaining particle hypotheses serve as alternatives j, while for a binary PID, only a single alternative hypothesis is used.

Given a momentum measurement, the particle's mass - and thus its identity - can be inferred using an additional observable. These secondary observables include the particle's velocity, ionization energy loss, and total deposited energy. Velocity-based PID information can be accessed at Belle II through two measurements: the time of flight in the TOP detector and the Cherenkov angle in the ARICH. As discussed in Section 2.2.2, these methods provide the strongest separation power for high-momentum tracks. At lower momenta, PID is dominated by the ionization energy loss measured in the CDC. Finally, measurements of total electromagnetic energy in the ECL provide powerful discrimination for electrons, whose characteristic shower profiles distinguish them from hadrons.

2.3 Boosted Decision Trees

Boosted decision trees (BDTs) are a widely used machine learning technique in high-energy physics for classification tasks, such as discriminating between signal and background events. Conceptually, a binary decision tree is a recursive search for the best selection (decision boundary) over the set of input features. Optimal selection, determined from a figure of merit (FOM), splits the original input sample into two nodes, each containing a subset of events more likely to be either signal or background. The tree continues to recursively split each node until the FOM no longer improves significantly or a maximum depth is reached.

Each individual binary tree is a "weak classifier": it provides limited separation between signal and background and is prone to overfitting if used alone. Boosting overcomes this limitation by iteratively training new trees, in which it increases the relative weight of misclassified events in the previous tree. As the weight of each event determines its contribution to the FOM at each node, higher weighted events are more likely to be correctly classified in subsequent splits. The final output of the BDT is the weighted average of resulting predictions of each independent tree, thereby providing a strong, robust separation of events between classes. [14]

Training a BDT involves three related concepts. The objective function defines the overall task the model is solving (e.g. binary or multiclass classification) and specifies how errors are minimized across all trees. The FOM, mentioned above, guides the optimal split at each node during tree construction. Finally, the evaluation metric, such as the micro-averaged area under the receiver operating curve (AUC), quantifies classifier performance on the entire training set and is used for hyperparameter tuning and early stopping. Use of the micro-averaged AUC ensures that all samples, including those from smaller classes, are treated equally in the overall performance measure.

Adding multiple classes increases the number of decision boundaries the classifier must learn, introducing additional complexity. Properly handled, multiclass BDTs simultaneously exploit the unique kinematics of each class, yielding a customized discrimination variable that can separate multiple backgrounds from the signal in a single classifier output.

2.4 Maximum Likelihood Estimation

In a majority of particle physics analyses, a physics parameter of interest (POI) is extracted from a model which encodes the expected distribution of an observable x given a set of parameters μ. The corresponding model is described by the probability density function (PDF), mathematically denoted as $f(x;\mu)$. Modifications to the normalization and shape of the PDF are encoded as nuisance parameters (NP) α each of which represents a quantity that can alter the relation between x and μ.

Given a set of observed data points $\vec{x}$, the PDF evaluated at the set of fixed observations is known as the likelihood:

$$\mathcal{L}(\mu,\alpha) = \prod_i f(x_i;\mu,\alpha)\,, \tag{13}$$

which provides a systematic measure of the agreement between the observed data and the probability model. In a binned model, the likelihood is constructed based on each bin-center i weighted by the number of events in that bin n_i:

$$\mathcal{L}(\mu,\alpha) = \prod_i f(x_i;\mu,\alpha)^{n_i}\,. \tag{14}$$

To extract the best estimate of the POI $\hat{\mu}$, it is convenient to work with a "regular" (quadratic) function, such that it can be described using only the curvature and the extrema. $\mathcal{L}(\mu)$, itself, is not quadratic but the logarithm $\ell(\mu) = \log\,\mathcal{L}(\mu)$ is. Thus, optimization of the likelihood is done through finding the minimum value of the combined negative log-likelihood:

$$-\ell(\mu,\alpha) = -\sum n_i \log f(\mu,\alpha)\,. \tag{15}$$

where the negative sign is used by convention. [14]

In typical analyses, the observed events comprise both signal (S) and background (B) contributions. Each of which is encoded as templates with known shapes $f_S(x)$ and $f_B(x)$. Combined with the Poisson probability of observing n_i events in a bin when $\mu S + B$ events are expected, the signal-plus-background PDF is:

$$\mathcal{P}(x|\mu,\alpha) = \prod_i \mathrm{Pois}(n_i \mid \mu S_i(\alpha) + B_i(\alpha)) \prod_{j=1}^{n_i} \frac{\mu S_i(\alpha) f_S(x_{ij}) + B_i(\alpha) f_B(x_{ij})}{\mu S_i(\alpha) + B_i(\alpha)}\,. \tag{16}$$

The previously mentioned NP, may be encoded through the use of constraint terms (Equation 26). Each of these constraints is described using auxiliary data distributed in accordance with a prior, usually of a Gaussian nature.

For robust estimation, analyses are performed simultaneously across disjointed channels (regions), which improves constraints on background templates while avoiding double-counting. Each template is required to have non-zero entries in all bins across all regions, thereby ensuring smooth derivatives and stability of the estimator. The full likelihood is then the product over all bins and regions, including both Poisson terms and NP constraints. [27]

3 Event Reconstruction and Signal Selection

As will be discussed in Section 7, while results are presented only for the muon channel, both electron and muon channels are presented in the reconstruction of the B_{sig} candidate.

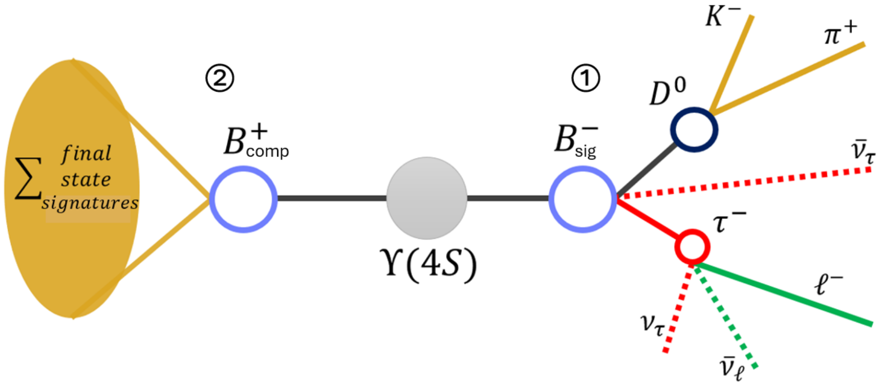


Figure 3.1: Simple diagram of reconstructed event topology.

Conventional tagged analyses, discussed in Section 1.3, select out potential signal events by first reconstructing the B_{tag}, then filtering them based on reconstruction of the B_{sig} candidate. To enhance signal efficiency, an inclusive tagging method is employed. As illustrated in Figure 3.1, signal events are selected through first reconstructing the B_{sig} (Section 3.2.1), then grouping together all remaining tracks and clusters into the rest-of-event (Section 3.2.2). The approach yields high signal efficiency but low purity; therefore, following the event reconstruction, two successive boosted decision trees are used for both background suppression and production of highly discriminating variables (Section 3.3).

3.1 Reconstruction Modes

Reconstructed channels proceed through a D^0 intermediate which is permitted to decay through one of two clean decay modes: $K\pi$ or $K3\pi$. The resulting D^0 candidate is further combined with a muon or electron candidate to complete the reconstruction of the B_{sig} meson. Considering neutrinos are undetected, reconstruction of the B_{sig} is completed simultaneously for both the signal $B \to D^0\tau\nu$ and the normalization $B \to D^0\ell\nu$ modes. The signal-side reconstruction, therefore, proceeds via:

$$B^{\pm}_{\text{sig}} \to (D^0 \to K^-\pi^+_A,\ K^-\pi^+_A 2\pi^{\pm}_B)\,\ell^{\pm}\nu\,, \quad \ell = e, \mu\,. \tag{17}$$

Explicit reconstruction of the D channel provides not only measurement of $R(D)$ but $R(D^*)$ as there is substantial feed-down resulting from $B \to (D^* \to D^0\,\pi/\gamma)\,\tau/\ell\nu$ where there is a missed intermediate state.

The respective selections made in the reconstruction of Equation 17 are outlined below. In particular, the reconstructed D^0 channels have pion daughters that are subdivided into two classifications: π_A and π_B. The first pion, π_A, occurring in both the two-body and four-body D^0 decay is required to be well-defined through application of particle identification (PID) with momenta $p > 0.3$ GeV. The remaining two pions in the four-body decay are permitted to span into the low momentum ($p < 0.3$ GeV) region where no requirement is placed on PID, as described below. These pions are denoted from now on as π_B.

3.2 Reconstruction of Event Candidates

Though only the reconstructed muon channel is investigated in the background suppression boosted decision trees and fit, the selection of the electron candidate (to be merged with the muon candidate) is outlined here.

3.2.1 Signal B Reconstruction

Signal event candidates are first selected through reconstruction of the B_{sig} from the interaction of final state particles with the Belle II detector (Section 2.2.2). A table summary of the selection criteria is presented in Appendix A.1.

All charged tracks are required to originate close to the interaction point (IP) and have detectable trajectories. These requirements include having a polar angle within CDC acceptance, a transverse distance from the IP of less than 1 cm, and a point of closest approach to the IP of less than 3 cm. Tracks with these selections are referred to as "good tracks" in the remainder of this section.

When particle identification (PID) selection is placed on the hadron signatures, the signal pion and kaon candidates are additionally required to have more than 20 CDC hits and a lab frame momentum greater than 0.3 GeV. The lower momentum threshold is placed to ensure that there are sufficient statistics available for PID corrections (Appendix B). Hadron PID is performed with neural network-based (NN) identification, which is trained on information such as individual subdetector likelihoods, track momentum, and charge [6]. Kaons are required to exceed a binary NN likelihood of 0.5, where the alternative hypothesis in Equation 12 is a pion. To increase the purity of the reconstructed D^0, one pion (π_A) is required to be well-defined through selecting candidates that have an NN PID greater than 0.1. The remaining two pions (π_B) in the four-body D^0 decay are permitted to have momenta spanning below 0.3 GeV. If the π_B candidate momentum is greater than 0.3 GeV, π_A selections are applied. Otherwise, no PID is applied and only minimal (greater than 0) CDC hits are required to reduce observed excess in data resulting from low-momentum tracks that do not reach the CDC.

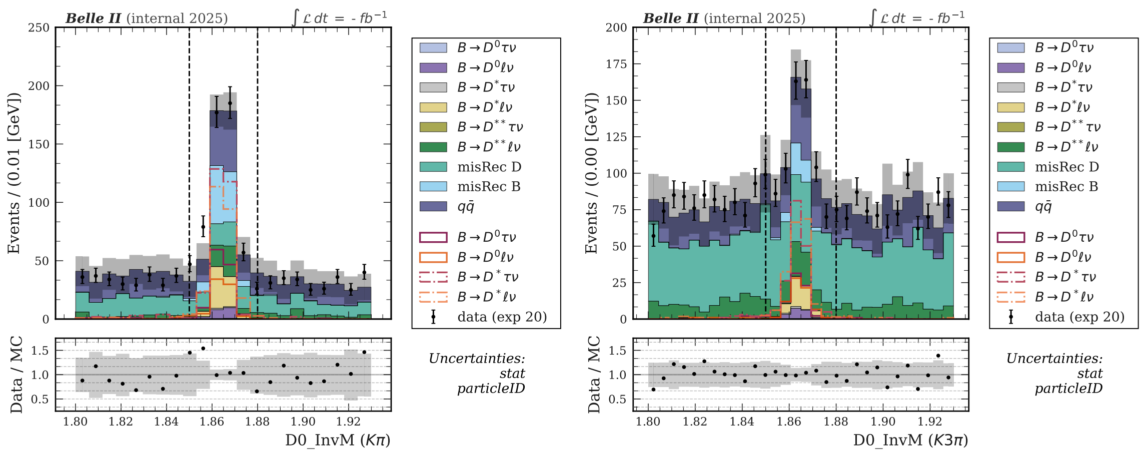


Figure 3.2: D^0 invariant mass selection for the $K\pi$ channel (left) and the $K3\pi$ channel (right). 3σ selection widow displayed by dashed lines. The displayed data and MC are generated from 5 randomly selected runs of experiment 20 and a small subset of dedicated signal.

Once the signal pion and kaon candidates are identified, all potential unique combinations of the corresponding four-momenta are summed together to reconstruct the D^0 meson candidates. If more than one unique combination occurs, multiple D^0 candidates are assigned to the event. Figure 3.2 shows the invariant mass spectrum of the D^0 candidates which is required to be within approximately 3σ of the nominal 1.86 GeV value. D^0 candidates passing this selection, along with an upper center-of-mass momentum threshold of 2.5 GeV, are combined with a lepton candidate to reconstruct the B_{sig}.

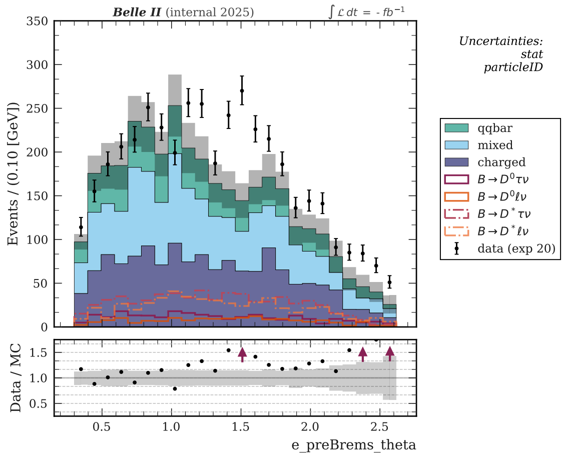


Figure 3.3: Polar angle distribution for electron candidates with a lower momentum threshold of 0.2 GeV. The displayed data and MC are generated from 5 randomly selected runs of experiment 20 and a small subset of dedicated signal.

Lepton candidates must satisfy flavour-specific selections. Electrons are required to not only be “good tracks” but have a momentum greater than 0.4 GeV, such that a majority of the electrons reach the ECL. Angle dependent selections were investigated where the tighter 0.4 GeV selection was placed in the transverse plane with $\theta \in [1.13, 1.88]$ radians while the remainder of the angular distribution was required to have 0.2 GeV. Placement of this angle dependent momentum threshold provides a reduction in the misidentification rate as well as removes data excess from low momentum curlers that peak around a polar angle of $\theta = \pi/2$ (Figure 3.3). However, due to poor modeling of variables which would aid in further discrimination of electron fakes, the tighter 0.4 GeV lower threshold was applied to the entire spectrum to obtain a reduction in the misidentification rate. The respective relative efficiencies compared to the angle dependent selection displayed approximately 90% signal $B \to D^{(*)}\tau/\ell\nu$ efficiency and 15-20% background efficiency.

Electron identification is performed with a BDT-based likelihood, where the primary features such as E/p come from the ECL. In addition to requiring an electron likelihood of greater than 0.9, a loose proton veto is placed with protonID < 0.99 to remove the observed small excess of mis-identified protons. Post PID selections, radiative Bremsstrahlung photons found within a cone of angle 0.05 radians are added to the uncorrected electron candidate if they pass an additional energy constraint of $0.01 < E < 1.0$ GeV (barrel) and $0.1 < E < 1.0$ GeV (endcaps). The resulting addition of the radiative photons provides an increase in energy resolution in the resulting corrected electron candidate. Muons, defined primarily in the KLM, are selected to originate near the IP and have an angular acceptance aligned with the KLM. Additionally, a lower momentum threshold is placed at 0.7 GeV to reduce the number of pion fakes and ensure sufficient statistics in PID corrections (Appendix B). Simple likelihood-based muon identification is placed where muon candidates are required to exceed a muon likelihood of 0.9.

The B_{sig} is reconstructed through summing the four-momentum of the D^0 and lepton candidates. The reconstructed B_{sig} undergoes a simultaneous vertex fit to all levels in the reconstructed decay chain (Equation 17). In the vertex fit, the daughter four-momenta are converged onto a common point where final state masses are constrained to the PDG values. Fitting is performed through varying the momenta of the respective daughters, within measurement uncertainties, such that further constraints (e.g. a common vertex point) are met exactly. The resulting best values of the varied momentum are used to compute the corresponding χ^2, which encodes the quality of the fit. A confidence level of greater than 0.1% is placed on the vertex fit, which maintains high signal efficiency while providing a relative efficiency of around 30% for continuum background and misreconstructed D^0 candidates and around 60% for misreconstructed B_{sig} candidates.

3.2.2 Companion B Reconstruction with Rest-of-Event

After reconstruction and selection of the B_{sig} candidate, all remaining tracks and clusters are grouped into the rest-of-event (ROE) and associated with the companion B_{comp} meson. Similar to the reconstructions described for the B_{sig} above, the ROE proceeds to sum the four-momenta of all potential unique combinations of the unused tracks and clusters to reconstruct the B_{comp}. If more than one unique combinations exist, multiple B_{comp} candidates are associated with the one B_{sig} candidate.

To further clean up the list of potential B_{comp} candidates, a simple mask is placed to ensure the ROE is associated with well-defined tracks and clusters (table summary in Appendix A.1). "Good tracks" are selected and additionally required to have more than zero CDC hits. ROE photon clusters are required to have trajectories in CDC acceptance and energies greater than 0.1 GeV in the endcaps and 0.05 GeV in the barrel region. As described in Section 2.2.2, there are several backgrounds that can result in misidentified cluster candidates. Neutral clusters resulting from hadronic split-off, are suppressed by selecting ROE photons with a distance exceeding 20 centimeters from the nearest track. Beam-background clusters are typically out of time with the event and thus have cluster times far from the time of the event t_0, whereas true photons from the interaction point are expected to have times consistent with t_0. To suppress these out-of-time backgrounds, we require the cluster timing (photon time minus t_0) to be under 200 nanoseconds.

3.3 Multivariate Classification

Due to the low purity resulting from the inclusive tagged approach, high-performance background suppression on the reconstructed $B \to D^0\mu\nu$ sample is achieved through two successive gradient boosted decision trees (BDTs). Each BDT is constructed as a multiclass XGBoost classifier, where hyperparameters are optimized with a Bayesian search [7]. BDT1 aims to remove poorly reconstructed B_{sig} candidates such that there is a significant reduction in sample size, thereby permitting refined studies into signal-like backgrounds (Section 4). BDT2 aims to provide not only suppression of the identified signal-like backgrounds but also production of highly discriminating fit variables.

As described in Section 2.3, multiclass BDTs display increased complexity due to the need to decipher between multiple decision boundaries in the input features. This increased complexity requires careful optimization of performance. For both BDT1 and BDT2, high performance is achieved through selecting the validation signal AUC as the evaluation metric. Rather than selecting the best model based on the evaluation metric, robust classification and regularization are achieved through selecting the best model based on a penalized AUC:

$$\begin{aligned} & AUC_{\text{sig, test}} - 0.5|AUC_{\text{sig, train}} - AUC_{\text{sig, test}}| \\ & + 0.5 * (AUC_{\text{micro, test}} - 0.5|AUC_{\text{micro, train}} - AUC_{\text{micro, test}}|) \,, \end{aligned} \tag{18}$$

where the penalty factor of 0.5 is determined experimentally. The *micro-averaged* AUC in Equation 18 provides proper averaging for highly-imbalanced training class populations by combining the weights of all events before averaging (Section 2.3).

Unless explicitly stated, the commonly used "BDT1 output" and "BDT2 output" refer to the signal classifier output, which provides one-vs-rest separation between the respective signal class and all background classes.

3.3.1 BDT1

A multiclass BDT architecture is adopted to enable the simultaneous suppression of both $B\bar{B}$ background and continuum events within a single unified classifier. This approach exploits the intrinsically different topologies of the two background categories: continuum events exhibit a characteristic jet-like structure, whereas $B\bar{B}$ events exhibit a more spherical event shape. A multiclass classifier, therefore, provides a natural framework for learning these distinct patterns without forcing them into a single background label. The D^* and D^{**} feed-down modes are excluded from the training to avoid contaminating the signal and background class with events that exhibit signal-like kinematic properties.

BDT1 is trained and optimized on approximately 3.65 fb^{-1} of generic Monte-Carlo (MC) and around 115,200 truth-matched signal events reconstructed from a dedicated $B \to D^{(*)}\tau\nu$ sample. As will be discussed in Section 7, only a subsample of the full dedicated sample was reconstructed. Therefore, the dedicated signal is reweighted based on detector and modeling effects (Appendix B); instead of reweighting by luminosity, the sample is reweighted by the $B \to D^0\tau\nu$ branching fraction then subbed into the generic sample with a normalization factor of $\mathcal{O}(0.0001)$. This normalization factor essentially smears the B_{comp} and signal D^0 branching fraction corrections over the dedicated signal events. With this substitution, signal events have weights on the order of 10^{-4} while background events from the generic sample have weights on the order of 10^{-1}. So, to ensure the signal events are not overlooked in the boosting, the signal training weights are multiplied by a factor of 1000, such that they are on the same order of magnitude as the background. The merged sample is then classified into three training classes:

1. *Signal*: properly reconstructed $B \to D^0\tau\nu$ and $B \to D^0\mu\nu$ events
2. *$B\bar{B}$ background*
3. *Continuum*

and split into equal-sized samples used to train/test the BDT (Figure 3.4) and provide independent optimization of selection criteria. To ensure proper physics event ratios in both samples, the splitting is performed with stratification of the event type (e.g. signal, charged B background, neutral B background, etc.).

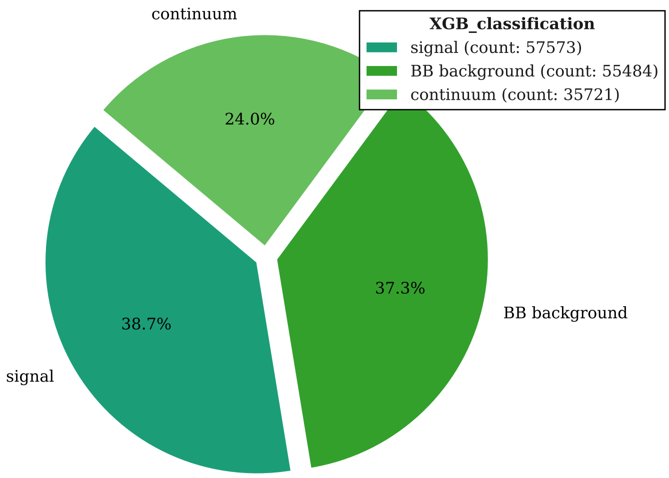


Figure 3.4: Original $\sum$ candidates of input BDT1 training sample.

Input Features

BDT1 is primarily designed to separate signal from background which is easily separable, such as continuum. Therefore, the most important of the fifteen selected well-modeled, highly discriminating input features are associated with event shape. A key example is the thrust, defined as:

$$T = \frac{\sum |\mathbf{T} \cdot \mathbf{p}_i|}{\sum |\mathbf{p}_i|}, \tag{19}$$

where $\mathbf{p}_i$ is the momentum of particle i and $\mathbf{T}$ is the thrust axis, defined as the unit vector along which the total momentum projection is maximized. Additional event shape variables used include the rotationally invariant Fox-Wolfram moments:

$$H_k = \sum |\mathbf{p}_i||\mathbf{p}_j| P_k(\cos\theta_{ij}) \,, \tag{20}$$

where, $\mathbf{p}_{i,j}$ are the momenta of particles i and j, θ_{ij} is the angle between the momentum vectors, and P_k is the k-th order Legendre polynomial. There are a total of 17 available Fox-Wolfram moments, subdivided into variables measuring the two-particle correlation between a particle from the reconstructed signal B_{sig} and a particle from the ROE B_{comp}, these correlations are denoted as "so". Variables describing the correlation between two ROE particles are denoted as "oo".

Other variables that aid in discriminating not just poorly reconstructed continuum events, but poorly reconstructed $B\bar{B}$ events include the ROE ΔE, measured as:

$$\Delta E_{\text{ROE}} = \frac{E_{\text{cms}}}{2} - \sum E_{\text{ROE particles}} \,, \tag{21}$$

and the beam-constrain mass associated with the ROE particles:

$$Mbc_{\text{ROE}} = \sqrt{E^2_{\text{beam}} - \sum |\mathbf{p}_{\mathbf{i},\text{ROE}}|} \,. \tag{22}$$

For properly reconstructed events, the beam-constrained mass provides an excellent, sharp peak around 5.28 GeV, the nominal B mass. In the case of poorly reconstructed B_{comp} mesons, values tend to be smeared toward lower masses.

As shown in Figure 3.5, XGBoost tracks the gain contributed by each feature, which represents the improvement in the evaluation metric when a node is split on that feature. This gain is weighted by the number of observations the node affects and averaged over all trees in the ensemble, providing a quantitative measure of each feature's separation power.

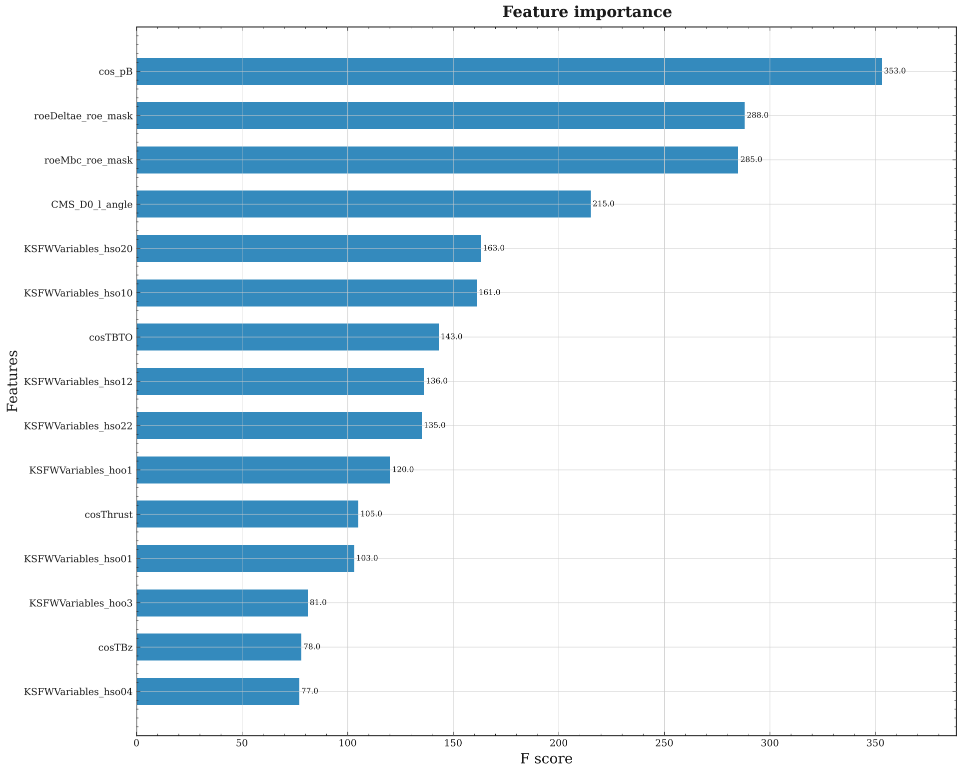


Figure 3.5: BDT1 input feature importance ranking based on XGBoost computed gain. Modeling and alias descriptions of the input features are given in Appendix D.

The feature that exhibits the largest gain is the cosine of the angle between the reconstructed B_{sig} candidate and the expected trajectory of a nominal B. Properly reconstructed B candidates have trajectories in alignment with the nominal trajectory, while misreconstructed B candidates exhibit smeared distributions (Figure D.1).

Training and Performance

As displayed in Figure 3.6, the multiclass classifier produces one output per training class, each encoding the selected objective function which displays the probability that each event is associated with the respective classification. As a result of selecting the simple set of input features described above, BDT1 displays strong discrimination between signal and strongly collimated continuum background. In contrast, the resulting separation power between signal and $B\bar{B}$ background events, which share similar event shapes, is significantly weaker.

The selected model exhibits closely matching classifier–output distributions for both the training and testing samples, indicating that the hyperparameter optimization and model-selection procedure - guided by the penalized AUC (Equation 18) - successfully regularized the classifier and yielded a robust final model. This is essential to ensure that similar performance is expected and observed on independent samples.

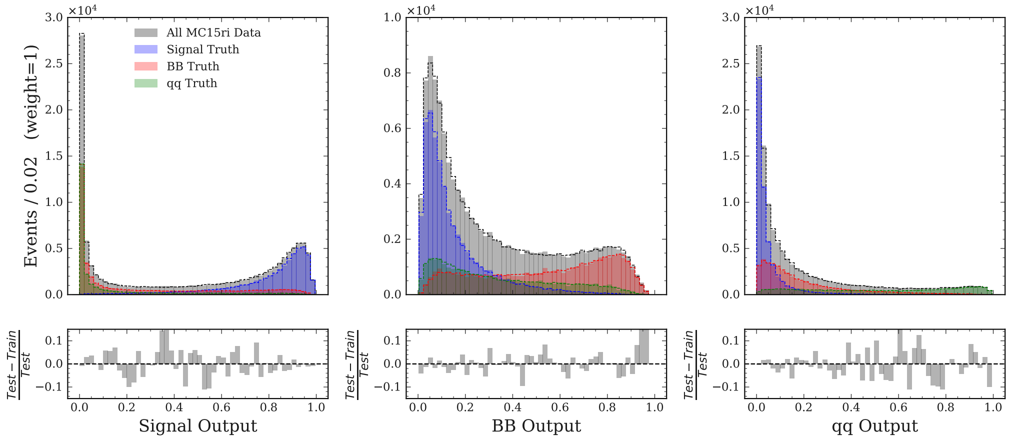


Figure 3.6: The resulting BDT1 classifier outputs for the optimal model resulting from training projected into the validation subsample.

Selection via Figure of Merit

The optimized selection criterion is found through maximizing the figure of merit (FOM),

$$\frac{S}{\sqrt{S+B}}, \tag{23}$$

which quantifies the statistical significance with which the signal could be distinguished from background under Poisson counting statistics. Here, S and B denote the expected signal and background yields after applying a given classifier cut. Therefore, placing a requirement at the maximum FOM selects the best signal sensitivity in the presence of background through balancing signal efficiency with background suppression.

To achieve optimal statistical significance of the signal candidates, the signal classifier output is scanned in steps of 0.05 and the FOM is computed at each value where $S = B \to D^0\tau\nu$ and $B \to D^0\mu\nu$ is excluded from the background B in Equation 23. The resulting optimal selection on the signal classifier output is found to be greater than 0.5. The relative efficiencies listed in Table 3.1 demonstrate substantial suppression of continuum ($q\bar{q}$) background, achieving a relative efficiency of 12%, while preserving high $B \to D^{(*)}\tau\nu$ signal efficiencies of approximately 90%. The lower efficiencies observed for the $B \to D^{(*)}\mu\nu$ normalization modes are acceptable given their larger branching fraction. As shown in Section 4, despite the reduced relative efficiency, the normalization channels still dominate the sample composition.

While general classifiers typically target background reduction to below 10% efficiency, BDT1 functions primarily as a first-level filter to reduce the sample and provide discriminating input features for BDT2, which performs suppression of signal-like background; therefore, the observed performance is sufficient for this thesis.

BDT1 Selection Statistics (Signal Classifier > 0.5)		
Classification	Relative Efficiency [%]	Purity [%]
$B \to D^0\tau\nu$	92	$0.09 \to 0.29$
$B \to D^0\mu\nu$	63	$3.47 \to 7.24$
$B \to D^*\tau\nu$	89	$0.29 \to 0.87$
$B \to D^*\mu\nu$	73	$12.70 \to 30.69$
$B \to D^{**}\tau\nu$	69	$0.02 \to 0.05$
$B \to D^{**}\mu\nu$	58	$3.52 \to 6.74$
misreconstructed D	26	$37.85 \to 32.90$
misreconstructed B	24	$11.34 \to 8.82$
$q\bar{q}$ Background	12	$30.72 \to 12.43$

Table 3.1: BDT1 efficiencies and respective purities for each classification discussed in Section 4. The reported purity values ("$X \to Y$") show the purity prior to selection ("X") on the signal classifier output and the purity after this selection ("Y")

.

BDT1 Sideband

Post application of BDT1 selection, the samples are close enough to the signal region (SR) that care must be taken when viewing on-resonance data. To investigate the modeling of BDT2 features (Appendix E) and complete additional modeling studies, a control region is defined as moderate BDT1 scores (0.2 < BDT1 < 0.4). The selected BDT1 SR and sideband are checked to have similar event compositions in terms of track and cluster multiplicity.

As shown in Figure 3.7, consistent modeling between the selected sideband and SR is checked through the use of off-resonance data and MC. Considering the well-modeled on-resonance distribution in the sideband and the observed consistency between off-resonance distributions in the sideband and SR, the BDT1 signal output is expected to be well-modeled in the on-resonance SR.

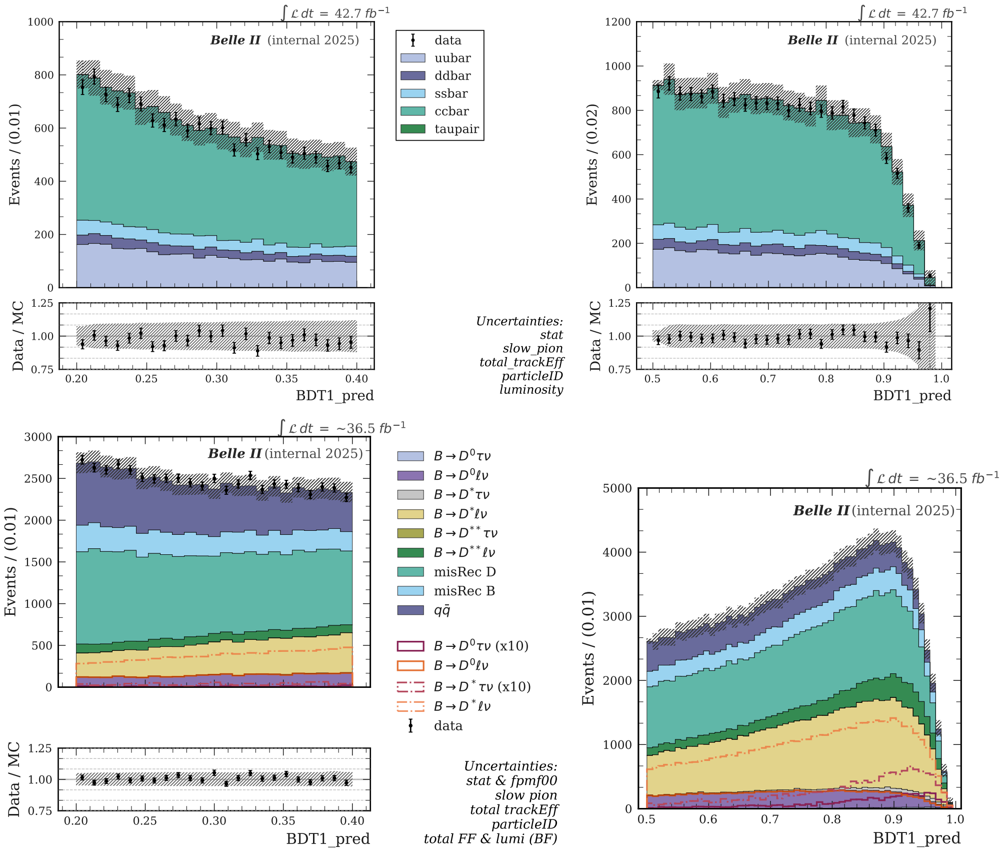


Figure 3.7: Consistent off-resonance modeling between the sideband (top left) and signal region (top right). The lower plot displays similar modeling of the on-resonance BDT1 sideband (bottom left) and signal region distribution (bottom right).

3.3.2 BDT2

Similar to BDT1, a multiclass BDT is used to exploit the unique topologies through simultaneous discrimination between the main backgrounds, discussed in Section 4. A total of five training classes:

1. *Signal*: properly reconstructed $B \to D^{(*)}\tau\nu$

2. D^{**} *Background*: missed intermediate decays of resonant/gap/non-resonant D^{**} contributions paired with a reconstructed lepton

3. *Misreconstructed* D^0: improper reconstruction of $K\pi$ or $K3\pi$ channel

4. *Misreconstructed* B: properly reconstructed D^0 candidate but improper reconstruction of $B \to D^0\ell\nu$

5. *Continuum*

are used to train the model, permitting construction of the background-dominant control regions used in the fit (Section 5.3). Though the dominating $B \to D^{(*)}\mu\nu$ are excluded from training there are still substantial class imbalances (Figure 3.8). To mitigate this, the best model is selected through the penalized AUC (Equation 18), which gives preference to signal classification but does not overlook the remaining classes, regardless of their size.

BDT2 is trained in the SR of BDT1 using approximately 36.5 fb^{-1} generic Monte Carlo (MC) and around 153600 properly reconstructed signal events from the dedicated $B \to D^{(*)}\tau\nu$ sample. Where, as was done in BDT1, the dedicated sample is reweighted to reflect the $B \to D^0\tau\nu$ branching fraction then subbed into the generic sample through scaling the dedicated weights by $\sum \omega_{\text{generic sig}} / \sum \omega_{\text{dedicated sig}} = \mathcal{O}(10^{-3})$. To ensure the signal events are not overlooked and to increase the relative signal importance in the determination of the FOM (Section 2.3), the initial signal training weights are multiplied by a factor of 1000. The sample is then split into equal-sized train/test and independent samples based on event type. The training sample composition is displayed in Figure 3.8.

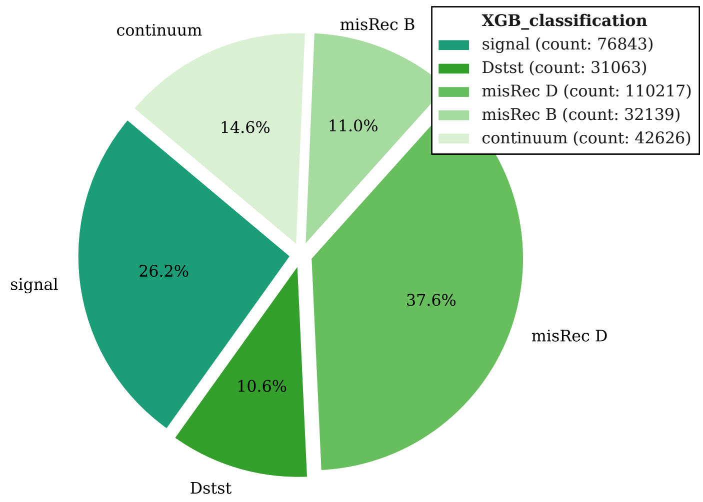


Figure 3.8: Original $\sum$ candidates of input BDT2 training sample.

Input Features

BDT2 input features are selected based on modeling quality, relative importance, and correlation. As the BDT2 signal output will be used in the two-dimensional fit with the momentum transfer q^2 (Section 5), all input features are additionally checked to have low correlation with q^2. The resulting minimal correlation between $q^2 = (p^*_\mu + p^*_\nu)^2$ and the center-of-mass momentum of the muon p^*_μ is relatively surprising. The current assumption is that the low correlation results from updating the q^2 value, computed based on the B_{sig} energy and momentum, with the direction of the ROE. However, post submission of this thesis, detailed studies on the respective correlation will be performed (Section 6.1).

The flight distance, computed based on the fitted vertex, plays a key role in discriminating misreconstructed D mesons. Likewise, the invariant mass associated with the summed four-momenta of the D daughters is particularly important in distinguishing peaking signal from flat misreconstructed D backgrounds.

Event variables such as the missing mass squared:

$$m^2_{miss} = (E^*_{beam} - P^*_{B\,sig} - P^*_{roe})^2 \tag{24}$$
$$= m^2_{ROE} + m^2_{B\,sig} - E_{beam}E_{B\,sig} - 2(\mathbf{p}^*_{ROE} \cdot \mathbf{p}^*_{B\,sig})\,, \tag{25}$$

are included to provide further separation between signal and misreconstructed B background. The respective impact each feature has in training the BDT is displayed in Figure 3.9.

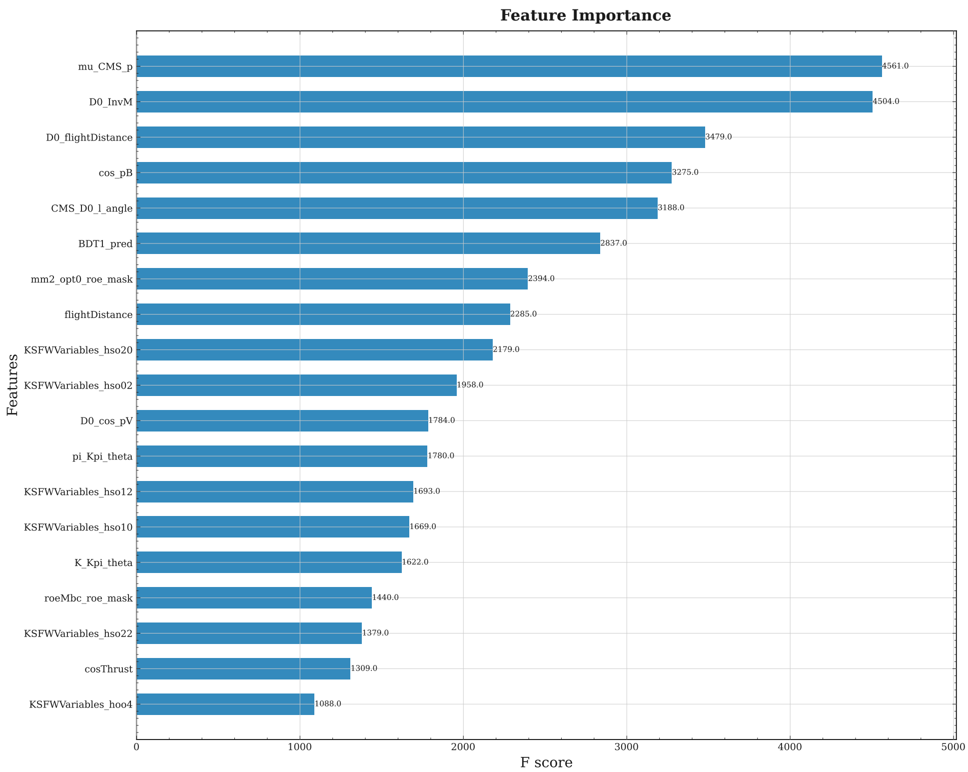


Figure 3.9: BDT2 input feature importance ranking based on XGBoost computed gain. Modeling and alias descriptions of the input features are given in Appendix E.

Training and SR definition

Training proceeds analogously to BDT1. Figure 3.10 shows the probability distributions for each class in the train and test samples, where minimal overtraining is observed primarily in low-multiplicity regions. Clear separation is achieved between signal, misreconstructed D^0, and continuum backgrounds. However, the kinematic similarity of $B \to D^{(*)}\tau\nu$, misreconstructed B, and D^{**} backgrounds results in relatively low discrimination. The resulting low separation power is expected as variables, such as those associated with vertex fits, are poorly modeled and therefore discarded to avoid potential bias in the resulting BDT output.

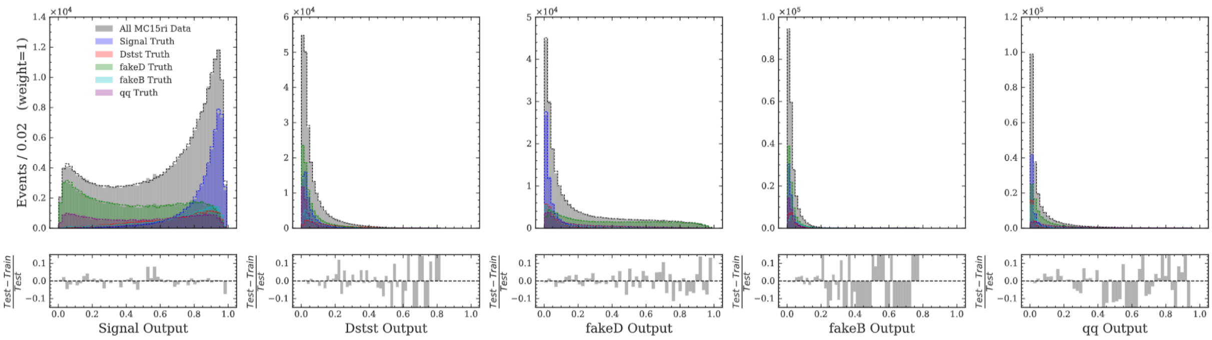


Figure 3.10: The resulting BDT1 classifier outputs for the optimal model resulting from training projected into the validation subsample.

The signal region (SR) is defined by scanning the signal classifier output in steps of 0.0001, maximizing the FOM in Equation 23, where S represents the sum of properly reconstructed $B \to D\tau\nu$ and $B \to D^{*}\tau\nu$ contributions. Normalization $B \to D^{(*)}\mu\nu$ are excluded from the background contribution B. The optimized SR, containing ∼68% of the $B \to D\tau\nu$ signal, is set at 0.8. This selection is used to define the SR in the direct extraction of $R(D^{(*)})$ (Section 5), which is completed in bins of the BDT2 signal output and q^2. A normalization region (NR) spanning 0.6–0.8 is used to constrain normalization and background shapes. To further constrain background yields and shapes, background-dominant control regions (CR) are defined and fitted in the remaining BDT2 classifier outputs. As discussed in Section 5, the addition of each individual background CR strongly constrains the background yield when the corresponding normalization is floated in the fit.

4 Background Analysis

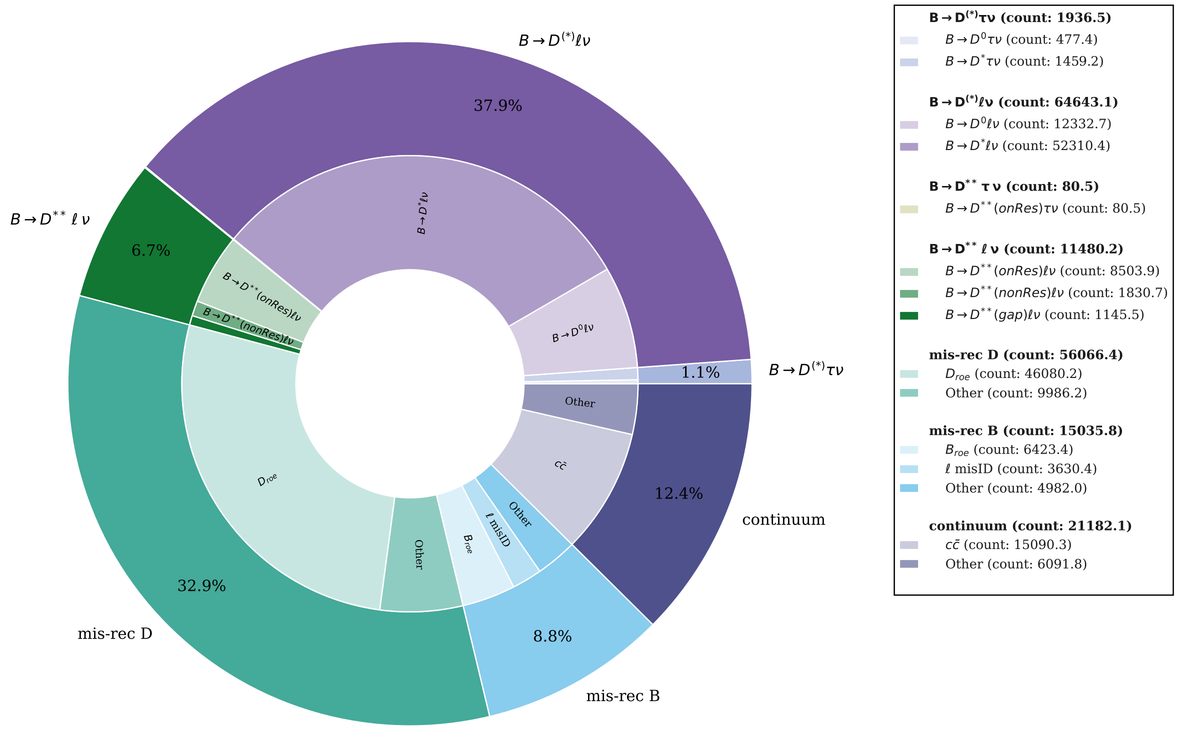


Figure 4.1: Generic sample composition post BDT1.

Although this thesis focuses on the muon channel, both the electron and muon channels share the same dominant background signatures and exhibit similar background compositions. To characterize these components, approximately 36.5 fb^{-1} of generic events passing BDT1 selections are investigated. The principal background categories are the following:

1. $\mathbf{B \to D^{**}\ell\nu}$

   Semi-leptonic decays involving excited D^{**} mesons constitute approximately 6.7% of the reconstructed muon-channel. In these events, the parent B meson decays to a higher-mass D^{**} intermediate resonance. The subsequent decay to the reconstructed D^0 produces additional final-state particles that are frequently absorbed into the rest-of-event (ROE). To provide consistent kinematics between background classifications, no additional truth-matching is required. Although recent progress has improved our understanding of these channels, available information remains insufficient to treat them as part of the signal. They are therefore handled as signal-like backgrounds: suppressed in BDT2 and extracted separately in the final fit.

2. **Misreconstructed D Background**
   The inclusive reconstruction of the companion B_{comp} introduces significant combinatorial effects, particularly in the reconstruction of the D^0 candidate. Events are designated as D_{roe} when one of the reconstructed D daughters originates from the B_{comp}. The remaining improperly reconstructed candidates — most often originating from semi-leptonic D^0 decays — are grouped into the “Other” category.

   Highly discriminating variables include the invariant mass of the reconstructed D^0 which displays a peak around 1.86 GeV for properly reconstructed candidates, whereas misreconstructed candidates have a smeared distribution.

3. **Misreconstructed B Background**
   Misreconstructed B combinatorics, denoted as B_{roe}, arise primarily from incorrect associations between the signal-side lepton and a companion-side D^0 candidate. Another subcategory is misidentification of the lepton candidate. In general, lepton misidentification is especially problematic in the reconstructed muon channel (Section 2.2.2). In the analyzed sample, muon misidentification is the result of misreconstructed B mesons around 24% of the time. Events that fail to fall into one of these subclassifications — largely composed of $B \to D_s^{(*)} D$ decays — are assigned to the “Other” B background category.

   Variables that provide strong discrimination against this background include the angular difference between the projected trajectory of a nominal B and the reconstructed candidate, as well as the center-of-mass momentum of the reconstructed lepton.

4. **Continuum ($\mathbf{q\bar{q}}$)**
   Continuum events arise from non-resonant $e^+e^- \to q\bar{q},\ \tau^+\tau^-$ production, whose hadronization can yield final states mimicking signal topologies. Although these events share the same $\sqrt{s}$ as $B\bar{B}$ events, their lower invariant mass produces a more collimated topology.

   Variables that provide strong discrimination include those related to thrust and event shape, such as the Fox-Wolfram moments (Section 3.3.1).

5 Signal Extraction

Following the logic presented in Section 2.4, the likelihood may be written as:

$$\mathcal{L} = \prod_c \prod_i \text{Pois}(n_{c,i}|N_{c,i}(\mu,\theta)) \prod_j C(\theta_j) \tag{26}$$

where c denotes the regions in which a simultaneous fit is performed and i indexes the bins within each region. The selected regions used to extract each parameter of interest (POI) μ are detailed in Section 5.3. The first term in Equation 26 describes the probability of observing $n_{c,i}$ events given the expected yield $N_{c,i}$ using a Poisson probability density function (PDF). This term inherently captures statistical fluctuations in the observed data, commonly referred to as *statistical uncertainties*. The second term consists of constraint factors $C(\theta_j)$, one for each nuisance parameter (NP), which penalize deviations of the NP from its nominal value according to prior knowledge. These NPs encode the various systematic uncertainties discussed in Section 5.4.

5.1 Statistical Model

As formulated in Equation 3, $R(D^{(*)})$ is theoretically defined as:

$$R(D^{(*)}) = \frac{\mathcal{B}(B \to D^{(*)}\tau\nu)}{\mathcal{B}(B \to D^{(*)}\ell\nu)} \tag{27}$$

where in this thesis, $\ell = \mu$. In this theoretical formulation, both the ratio and the denominator contain contributions from the neutral and charged B mesons. Assuming isospin symmetry, the only difference between the charged and neutral branching fractions lies in the lifetime, τ_{B^+,B^0}:

$$\Gamma(B^\pm \to D^{(*)}\ell/\tau\nu) \approx \Gamma(B^0 \to D^{(*)}\ell/\tau\nu) \tag{28}$$

$$\tau_{0+}\mathcal{B}(B \to D^{(*)}\ell/\tau\nu) = \mathcal{B}(B \to D^{(*)}\ell/\tau\nu) \tag{29}$$

where $\tau_{0+} = \tau_{B^0}/\tau_{B^+}$. Applying this to Equation 27, $R(D^{(*)})$ can be reformulated as only a function of the charged or neutral B meson:

$$\begin{aligned} R(D^{(*)}) &= \frac{\mathcal{B}(B^\pm \to D^{(*)}\tau\nu) + \mathcal{B}(B^0 \to D^{(*)}\tau\nu)}{\mathcal{B}(B^\pm \to D^{(*)}\ell\nu) + \mathcal{B}(B^0 \to D^{(*)}\ell\nu)} && (30)\\ &= \frac{(1+\tau_{0+})\mathcal{B}(B^\pm \to D^{(*)}\tau\nu)}{(1+\tau_{0+})\mathcal{B}(B \to D^{(*)}\ell\nu)} && (31)\\ &= \frac{\mathcal{B}(B^\pm \to D^{(*)}\tau\nu)}{\mathcal{B}(B^\pm \to D^{(*)}\ell\nu)}\,. && (32) \end{aligned}$$

Thus, providing a valid extraction of the nominal value of $R(D^{(*)})$ through the use of only the charged $B^{\pm}$ while providing proper parameterization of the $B^0 \to D^*\tau/\ell\nu$ feed-down included in the D^* signal/normalization templates. Unfortunately, extracting $R(D^{(*)})$ based on only the reconstruction of the charged $B^{\pm} \to D^0\mu^{\pm}\nu$ channel results in significantly lower statistical sensitivity then if a combined $B \to D^{(*)}\ell\nu$ analysis were to be performed.

Experimentally, $R(D^{(*)})$ may be expressed as:

$$R(D^{(*)}) = \frac{\nu_s\,\varepsilon_n}{\nu_n\,\varepsilon_s} \prod \frac{\xi_s^i}{\xi_n^i} \tag{33}$$

where ν_s and ν_n are the fitted yields for the signal $B \to D^{(*)}\tau\nu$ and normalization $B \to D^{(*)}\mu\nu$ channels, $\varepsilon_{s,n}$ are the corresponding selection efficiencies, and $\xi_{s,n}^i$ encode the efficiency impact on $R(D^{(*)})$ from the applied correction factors (Appendix B).

To unfold these effects directly in the fit, the statistical model (PDF is described as a function of $R(D^{(*)})$ by writing the corresponding fitted yields in terms of the normalization channel, ν_n:

$$\nu_n = \nu_n \tag{34}$$

$$\nu_s = R(D^{(*)})\,\nu_n\,\frac{\varepsilon_s}{\varepsilon_n} \prod \frac{\xi_s^i}{\xi_n^i}\,. \tag{35}$$

5.2 Fit Variable Modeling

Modeling of the BDT2 signal output and momentum transfer q^2 distributions is validated using a combined BDT1+BDT2 sideband, defined as $0.2 < \mathrm{BDT1} < 0.4$ and $0.6 < \mathrm{BDT2} < 0.8$. The selected sideband contains a negligible contribution from signal $B \to D^{(*)}\tau\nu$ events and is thus insensitive to characteristics of the signal distributions. On-resonance modeling is examined exclusively in the sideband region. Consistency between the selected sideband and signal region (SR) is verified through comparisons of track and cluster multiplicities, as well as modeling in off-resonance data.

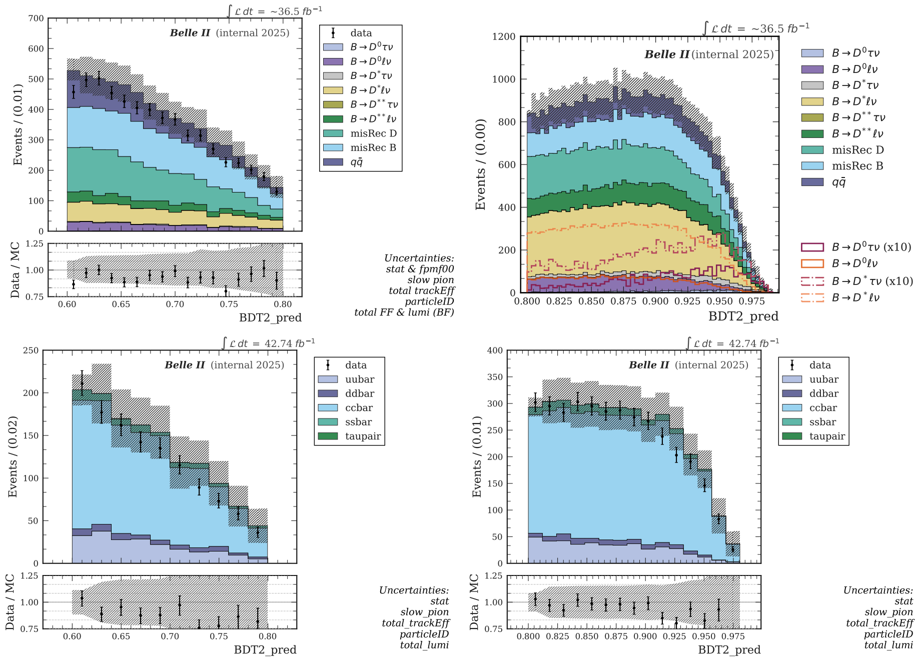


Figure 5.1: Resulting modeling/MC distribution of the BDT2 signal output in on-resonance samples (top) and off-resonance samples (bottom) for both the BDT1+BDT2 sideband region (left) and the BDT2 SR (right).

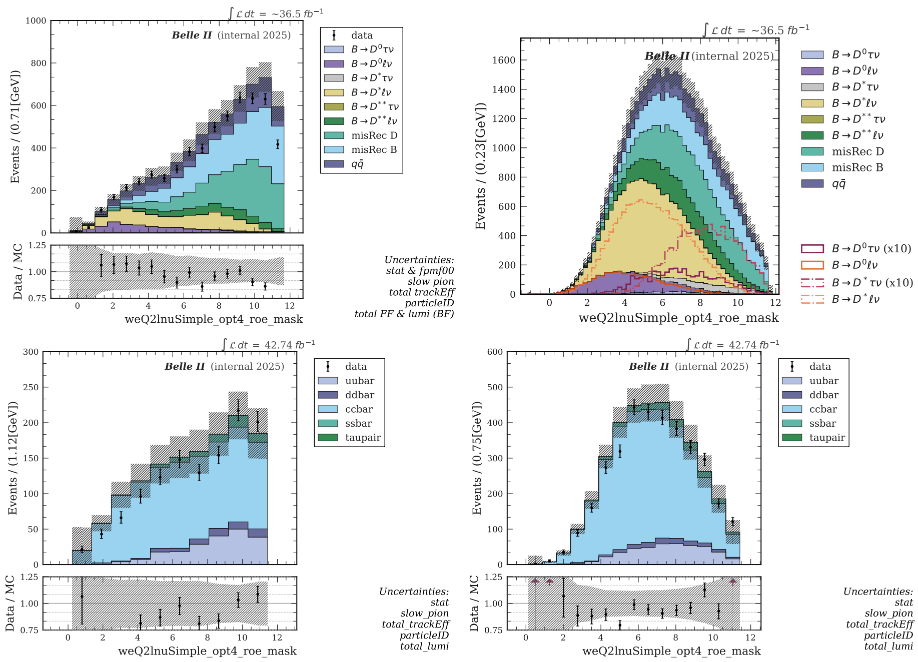


Figure 5.2: Resulting modeling/MC distribution of the q^2 in on-resonance samples (top) and off-resonance samples (bottom) for both the BDT1+BDT2 sideband region (left) and the BDT2 SR (right).

5.3 Fitted Regions

2D Regions

As described above, the signal region (SR) is determined by optimizing the figure of merit (Equation 23) on the BDT2 signal output and corresponds to BDT2 > 0.8. To constrain the normalization and background templates, a normalization region (NR) is defined using the intermediate BDT2 range $0.6 <$ BDT2 < 0.8 and is required to be mutually exclusive from each of the following background-dominant control regions (CR).

In both the SR and NR, events are binned into six BDT2 intervals and sub-binned into eight q^2 intervals which span the full kinematic range. Figure 5.3 shows the two-dimensional distribution and one-dimensional projections. Figure 5.4 shows the resulting Asimov fit to the two-dimensional BDT2 x q^2 distributions. Where a symmetric binning scheme is adopted, with indices increasing from high to low values of BDT2 and q^2, as shown in Appendix A.

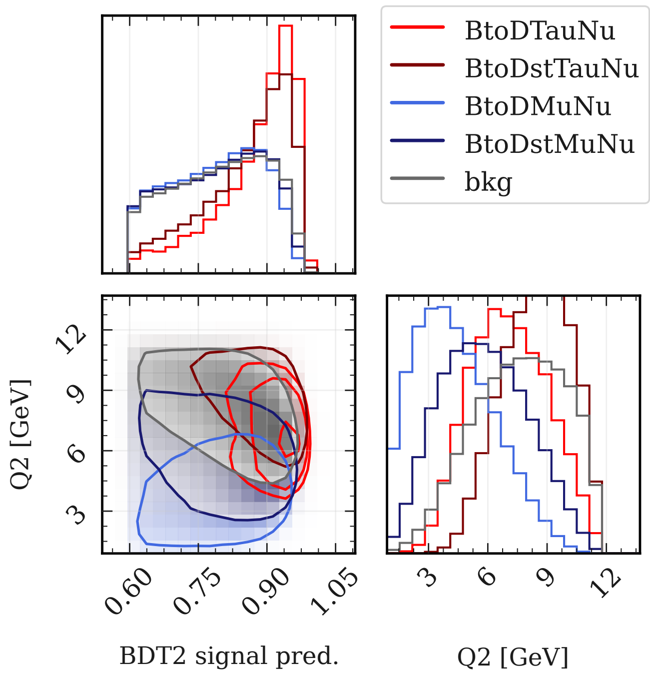


Figure 5.3: Two-dimensional contour of the fitted BDT2 signal output x q^2.

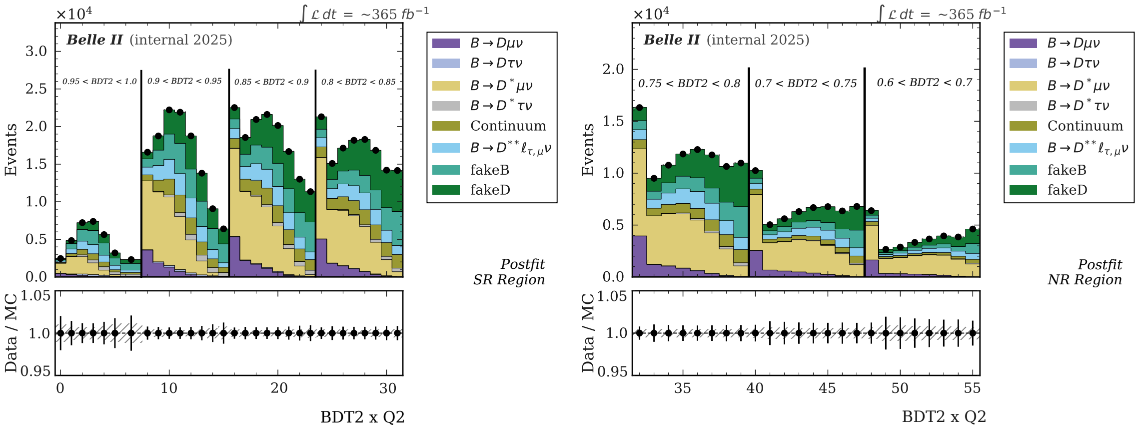


Figure 5.4: Two-dimensional Asimov fit to BDT2 slices in 8 bins of q^2 in the SR (left) and NR (right). The NR is defined to be mutually exclusive from each background-dominant CR presented below.

1D Background Fit Regions

Each background-dominated control region (CR) is defined using the BDT2 output classifiers to be mutually exclusive with both the two-dimensional SR and NR as well as all other background CRs (Appendix A). The binning choices for these one-dimensional distributions are optimized to provide sensitivity to both the normalization and the shape of each background component across the SR+NR and the respective CR.

For each newly introduced background CR, the associated background normalization is floated in the fit and verified to be well constrained by the available data.

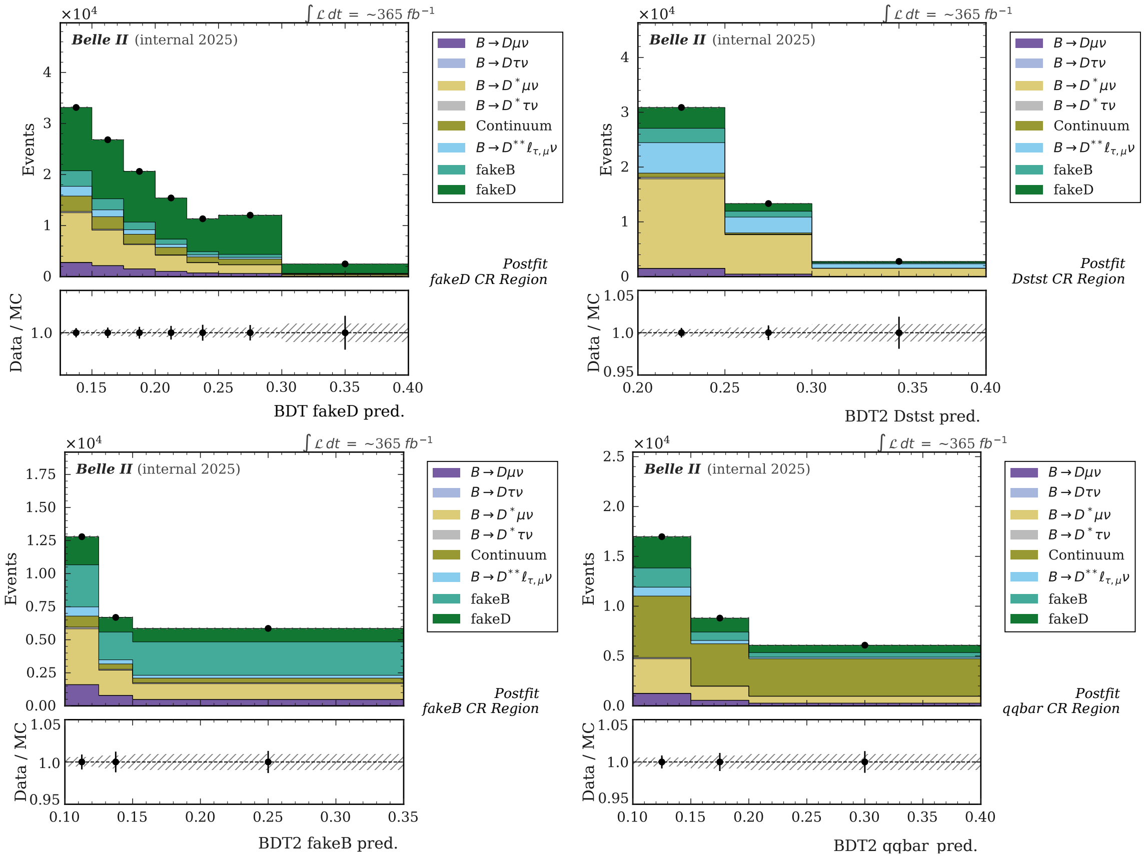


Figure 5.5: One-dimensional Asimov fit to background-dominant CRs defined from BDT2 background output distributions. The misreconstructed D^0 region is split into 7 bins (top left) and the D^{**} (top right), misreconstructed B (bottom left), and continuum (bottom right) fits are completed in 3 bins (top right).

5.4 Systematics

In high-energy physics analyses, systematic uncertainties arise from incomplete knowledge of the experimental apparatus, detector response, and theoretical modeling of the underlying physics processes. These inherent limitations are encoded as NP, each accompanied by an appropriate constraint term in the likelihood (Equation 26). Properly accounting for these uncertainties is essential to produce reliable measurements and meaningful comparisons with theoretical predictions. Systematic uncertainties generally fall into two broad categories: uncorrelated and correlated.

Uncorrelated systematics are those for which variations in one source do not affect other measurements or processes, such as branching fractions. Each source is treated as an independent NP in the PDF, with the corresponding constraint term.

Correlated systematics, by contrast, arise from sources that simultaneously affect multiple observables or events. Examples include uncertainties in particle identification efficiencies and slow track reconstruction. The respective uncertainties are represented using covariance matrices, which encode the correlations between the selected correction bins. The covariance matrix C can be decomposed into a set of orthogonal eigenvectors V, each of which represents an independent source of uncertainty:

$$C = V \Lambda V^{-1} \, , \tag{36}$$

where Λ is the diagonal matrix of the eigenvalues, λ_i. Each eigenvariation is then the product of each eigenvector and the square root of the corresponding eigenvalue, $\xi_i = v_i \sqrt{\lambda_i}$. The resulting eigenvariations are then treated as a separate NPs in the PDF with the appropriate constraint term.

In general, NP are encoded through the construction of additional $\pm 1\sigma$ templates. As detailed in Section 6.1, only a subset of the systematic uncertainties are encoded in the PDF at this time, each of which is outlined below. The remaining systematics detailed in Appendix B, including particle identification, on-resonance $B \to D^{**} \mu\nu$ branching fractions, and uncertainties pertaining to continuum reweighting, as detailed in Appendix B, will be added post submission of this thesis. In addition, the systematics will be switched to the *SysVar* package [38], such that results may be easily combined with other analyses and experiments (Section 6.1).

5.4.1 Simulation Statistics

Uncertainties arising from the finite number of simulated events in each fitted bin are governed by Poisson statistics. Consequently, each bin has an associated uncertainty:

$$\sigma_{\text{bin}} = \sqrt{\sum_i \omega_i^2}, \tag{37}$$

where ω_i is the weight of the i-th event in the bin. Statistical uncertainties are determined through the use of 2500 pseudo experiments where Gaussian-based fluctuations of each bin are randomly sampled. In the current configuration of the fit model, the background yields are encoded as Gaussian-constrained NP, with the 1σ prior set to 50% of the nominal MC yields and would thus be considered systematics. However, prior analyses show sensitivity to free-floating background yields. Therefore, this study allows the background yields to float so that a consistent sensitivity definition can be used for comparison.

5.4.2 Branching Fractions

Considering the non-zero probability of reconstructing either B candidate in the $\Upsilon(4S)$ decay, both sides are reweighted based solely on the truth-level information of the generated MC decay. Considering each B is treated as an independent random variable, each normalization+shape NP encodes the $\pm 1\sigma$ variations for both sides of the $\Upsilon(4S)$ decay.

Currently, the included branching fraction (BF) systematics correspond to the following decays:

- $B \to D^{(*)} e\nu$,
- $B \to D^{**} e/\tau\nu$ (on-resonance/non-resonance/gap),
- $B \to D^{**} \mu\nu$ (non-resonance/gap),
- and $D^0 \to K\pi, K3\pi$,

of which the values and corresponding technical details are outlined in Appendix B.1. As the leptonic $\tau^{\pm}$ branching fractions are sufficiently accurate in MC, they are neither corrected nor treated as a source of systematic uncertainty in this analysis.

Additional BFs for dominant backgrounds will be added as additional correction factors and NP post submission of this thesis (Section 6.1).

5.4.3 Luminosity

To account for the uncertainty in the integrated luminosity, a global normalization NP is introduced in the fit, allowing all samples to vary coherently by $\sigma = \pm 0.0007$ relative to the nominal yields.

5.4.4 Form Factors

Modeling of the hadronic current, briefly mentioned in section 1, is encapsulated into a single function per decay mode, known as a form factor (FF). Although effects from the hadronic current largely cancel in the branching-fraction ratio $R(D^{(*)})$, the form factors depend on q^2. As a result, corrections and associated systematic uncertainties are included to account for the kinematic differences between the $B \to X_c\tau\nu$ and $B \to X_c\ell\nu$ decays.

Each $b \to c\tau/\ell\nu$ transition has different kinematics and transition matrix elements, and thus are represented as independent FF model. One exception occurs for the updated $B \to D^{(*)}\tau/\ell\nu$ model, BLPRXP [11], as it provides coherent treatment of the $\tau^{\pm}$ and $\ell^{\pm}$ as well as the D and D^* decay rates through a combined fit to data. Technical details on the correction factors, implemented in the `HAMMER` framework [22], are given in Appendix B.2.

`HAMMER` likewise internally computes each variation using the correlated systematic approach discussed above, and one shape+normalization NP is encoded for each eigenvariation given for each model. To account for additional mismodeling in the broad D^{**} widths (Appendix B.7), a 2.5σ variation is used instead of the 1σ variation.

Systematics associated the form factors of the on-resonance $B \to D^{**}\mu\nu$ decays lead to instability in model and are thus omitted at this time.

5.4.5 Slow Pions

To properly account for discrepancies in slow pion tracking efficiencies, correction factors are derived from fits, performed by [21], in three bins of momentum (Appendix B.6). Disentanglement of efficiency correlations are completed through the construction of a 3x3 covariance matrix. The appropriate corrections and systematic uncertainties are then extracted through eigendecomposition as described above. Each of the three 1σ eigenvarations is encoded into the PDF as individual normalization+shape NP.

5.5 Preliminary Results

As discussed in the previous sections, the correct implementation of systematic uncertainties is essential for a reliable extraction of $R(D)$ and $R(D^*)$. At present, the analysis still exhibits inaccuracies in the treatment of several systematics—both those already included in the fit model and those not yet incorporated. Dedicated, comprehensive studies of the systematic encoding will be carried out following the submission of this thesis (Section 6.1).

To provide preliminary results and to validate the overall analysis strategy, Asimov fits are performed using the model and fit regions introduced above. Systematic uncertainties are added incrementally, and after each inclusion, the stability of the fit is verified through the observation of a parabolic likelihood scan. Systematics are excluded if they fail the nominal fit or any fit in the log likelihood scan. In this preliminary configuration, where only a subset of systematics is active (Section 5.4), the statistical behavior of the model appears consistent with expectations.

Due to high correlations between $R(D)$ and $R(D^*)$, results are displayed for two configurations: independent fits to either $R(D)$ or $R(D^*)$ and a simultaneous fit to $R(D^{(*)})$. Focusing first on the independent fits, statistical uncertainties are extracted through a purely statistical model where the only configured NPs are those associated with the background yields. From the resulting best fit model, 2,500 pseudo-experiments are drawn to extract the corresponding statistical uncertainty (Figure 5.6). The respective Monte Carlo (MC) systematic contribution is extracted by adding in the NP constraints described in Section 5.4 and subtracting in quadrature from the HESSE error. Finally, a fit is performed to the full model, with all induced systematics, to get the "total" uncertainty on each POI.

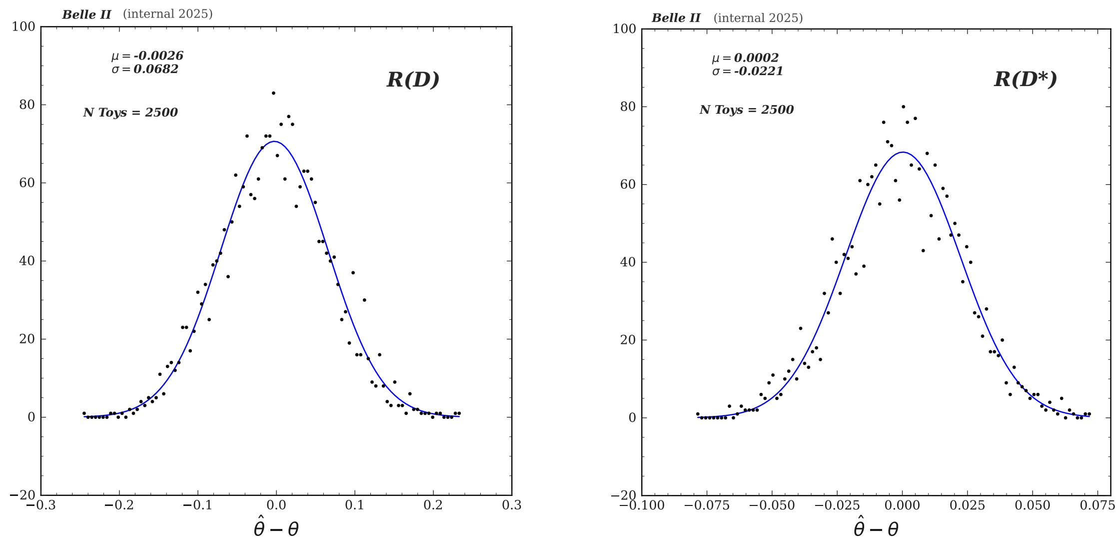


Figure 5.6: Distribution of the best-fit $R(D)$ value minus $R(D)_{SM}$ (left) and the best-fit $R(D^*)$ value minus $R(D^*)_{SM}$ (right) obtained from 2,500 pseudo-experiments (black points) used to estimate the statistical uncertainty. $\hat{\theta}$ represents the best-fit value of $R(D)$ or $R(D^*)$ and θ represents the corresponding SM prediction. The 1σ width of the overlaid Gaussian fit (blue line) provides the statistical uncertainty.

In these independent fits, the statistical precision achieved with respect to the hadronic and semi-leptonic results is notable, particularly given that only $B^{\pm} \to D^0\mu^{\pm}\nu$ decays are reconstructed. The values obtained from the Asimov fit are:

$$R(D) = 0.299 \pm 0.0682 \text{ (stat)} \pm 0.033 \text{ (MC stat)} \pm 0.0328 \text{ (syst)} \ , \tag{38}$$

$$R(D^*) = 0.257 \pm 0.0221 \text{ (stat)} \pm 0.011 \text{ (MC stat)} \pm 0.0136 \text{ (syst)} \ . \tag{39}$$

Using the respective correlation from the simultaneous fit described below, the statistical (black dashed) and total (black solid) uncertainty ellipses are plotted in Figure 5.7 with respect to the SM prediction.

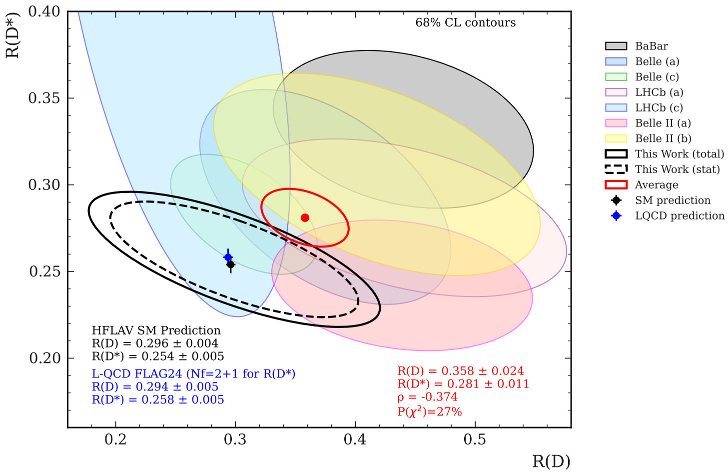


Figure 5.7: Respective ellipse projections for independent fits to either $R(D)$ or $R(D^*)$ in the case of a statistics-only model (dashed black) and full model, encoding statistics and discussed subset of systematics (solid black). The displayed uncertainty bands correspond to the symmetric HESSE uncertainties.

In the simultaneous fit, an inflation in both the statistical and systematic uncertainties is observed. The current hypothesis is that the inflation results from the high level of anti-correlation ($\rho = -0.72$) between $R(D)$ and $R(D^*)$. This is in part inherited from the significant anti-correlation ($\rho = -0.80$) between the $B \to D^0\mu\nu$ and $B \to D^*\mu\nu$ normalizations. The respective statistical uncertainty is obtained using 2,500 pseudo-experiments drawn from the nominal model where both $R(D)$ and $R(D^*)$ float freely (Appendix C). The best-fit values on Asimov data are:

$$R(D) = 0.299 \pm 0.0951 \text{ (stat)} \pm 0.045 \text{ (MC stat)} \pm 0.0368 \text{ (syst)} \,, \tag{40}$$

$$R(D^*) = 0.257 \pm 0.0308 \text{ (stat)} \pm 0.014 \text{ (MC stat)} \pm 0.018 \text{ (syst)} \,, \tag{41}$$

$$\rho = -0.72 \,. \tag{42}$$

The statistical (black dashed) and total (black solid) uncertainty ellipses are plotted in Figure 5.8 with respect to the SM prediction.

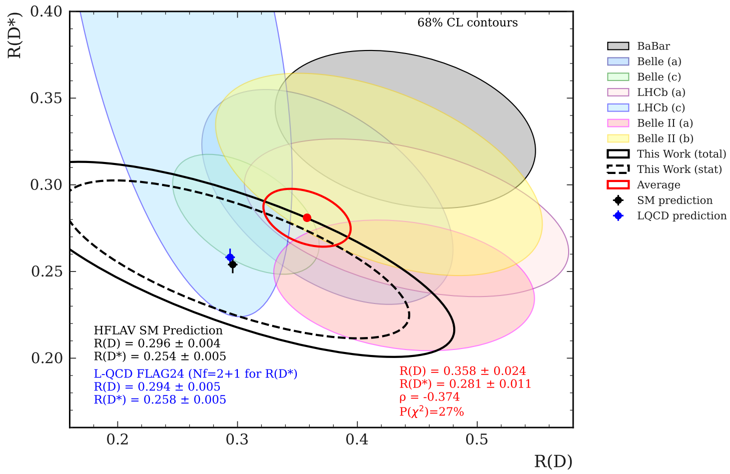


Figure 5.8: Respective ellipse projection for simultaneous fit to $R(D)$ and $R(D^*)$ in the case of a statistics-only model (black dashed) and full model, encoding statistics and discussed subset of systematics (black solid). The displayed uncertainty bands correspond to the symmetric HESSE uncertainties.

6 Conclusions

This thesis presents the first simultaneous analysis of $R(D)$ and $R(D^*)$ using inclusive tagging. Although further work is needed to finalize the fit construction, the current results demonstrate the statistical validity of extracting $R(D^{(*)})$ with an inclusive tagging. The preliminary statistical sensitivity achieved in the independent fits described above - 23% for $R(D)$ and 9% for $R(D^*)$ - is slightly less sensitive than the achieved statistical precision in the hadronic tagged analysis (Table 6). It should be noted, however, that the hadronic tagged analysis has a complete reconstruction of $B \to D^{(*)}\ell\nu$ where a large number of D and D^* decays are considered.

Provided the limited reconstruction of $B^{\pm} \to D^0\mu^{\pm}\nu$, the results presented in this thesis display the first predictions of the lower bounds for the potential statistical sensitivity on $R(D^{(*)})$ using an inclusive tag approach.

$R(D)/R(D^*)$	% Stat. Uncertainty (wrt SM Prediction)	B Tagging	τ Decays	References
$R(D) = 0.418 \pm 0.074 \pm 0.051$	18% (25%)			
$R(D^*) = 0.306 \pm 0.034 \pm 0.018$	10% (12%)	SL	$\tau \to \ell\nu\bar{\nu}$	[2]
$\rho = -0.22$				
$R(D) = 0.439 \pm 0.055 \pm 0.046$	12% (21%)			
$R(D^*) = 0.242 \pm 0.019 \pm 0.016$	8% (6%)	Hadronic	$\tau \to \ell\nu\bar{\nu}$	[39]
$\rho = -0.22$				
$R(D) = 0.299 \pm 0.0682 \pm 0.0465$	23%	Inclusive	$\tau \to \ell\nu\bar{\nu}$	This Work
$R(D^*) = 0.257 \pm 0.0221 \pm 0.0175$	9%			

Table 6.1: Statistical sensitivity of independent fits in this work and recent Belle II results for simultaneous $R(D^{(*)})$ analyses.

6.1 Future Research

The results and work presented in this thesis represent an ongoing effort. There are a few items that require updates and/or detailed studies post-submission of this thesis. In no particular order, these items include:

- **Update simulation sample**
  Improved modeling is expected when the samples are updated to the recent release and detailed modeling studies will be performed, including investigations into SVD timing. Likewise, the dedicated sample will be used to update generic signal events in the fit.

- **Addition of dominant background BF corrections/systematics**
  Branching fraction (BF) corrections and systematics are presently applied only to the $B \to X_c\tau/\ell\nu$ and subsequent signal decay channels; extending them to the primary background modes will provide a more accurate description of the dominant backgrounds.

- **Investigation of mismodeling of p^*_μ and $E^{\mathbf{roe}}_{\mathbf{extra}}$**
  A minor error in the BDTc weight application (Appendix B.8) was found after the presented results were finalized; correcting it improves low-momentum modeling of the muon's center-of-mass momentum but significant high-momentum mismodeling remains, which requires further study. A previously observed positive shift in the extra ECL energy of the rest-of-event (ROE) will also be investigated.

- **Investigation into systematic encoding in fit model**
  The systematic uncertainties have recently been migrated to the new *SysVar* package [38]. Updating this analysis to that framework will be a priority after submission, followed by checks to resolve any remaining encoding issues and ensure a stable fit with proper cancellation of uncertainties in Equation 27. Because the current encoding results in inconsistencies, no systematic impact breakdown is given; however, comparison with Belle II's published $R(D^{(*)})$ analyses (Table A.7) and the inclusive tag $B \to K\nu\bar{\nu}$ analysis [20] indicates that the leading uncertainties should stem from limited Monte Carlo (MC) statistics, D^{**} gap branching fractions, background normalizations, and lepton identification, with most others expected to be at the $\mathcal{O}(1\%)$ level.

- **Validation of full fit model**
  The complete fit model, including statistical and systematic uncertainties, will be validated using a pull test with 2,500 pseudo-experiments. Linearity tests will be performed by scaling the yields of $B \to D^{(*)}\tau\nu$ and $B \to D^{(*)}\ell\nu$ to ensure the estimator is unbiases and robust.

- **Reconstruction of $B^{\pm} \to D^0 e^{\pm}\nu$**
  A more complete reconstruction of $B \to D\ell\nu$ will be completed by including the electron final state, described in Section 3.2.1. Independent studies of the electron channel will be conducted before merging it with the muon channel into a combined measurement.

7 Notes on Analysis and Acknowledgments

This thesis presents results for the analysis of

$$R(D^{(*)}) = \frac{\mathcal{B}(B^{\pm} \to D^{(*)}\tau^{\pm}\nu)}{\mathcal{B}(B^{\pm} \to D^{(*)}\mu^{\pm}\nu)} \,. \tag{43}$$

Studies of the neutral B^0 modes are ongoing by Boyang Zhang and Dr. Thomas Browder (U. Hawaii), and are therefore not presented in this work. While event reconstruction is described for both electron and muon final states, the boosted decision trees (BDTs) and subsequent fits are restricted to the muon channel, providing a conservative estimate of the statistical uncertainties.

The analysis employs both generic and dedicated Monte Carlo (MC) samples. Generic MC, used in the fit and BDTs, provides a broad spectrum of $B\bar{B}$ and continuum events, whereas dedicated MC, used in the BDTs, is produced specifically for signal $B \to D^{(*)}\tau\nu$ decays. Due to computational constraints and partial reconstruction of the available dedicated samples, a random best-candidate selection is applied to ensure proper reweighting, combination of generic and dedicated MC, and valid data/MC checks.

Not all systematic uncertainties have been included in the current fit; notable exclusions with potentially significant impact include PID corrections and on-resonance branching fractions for $B \to D^{**}\mu\nu$. As discussed in Section 6.1, these will be incorporated, along with the electron channel, in post-submission studies.

Recent collaborative efforts between DESY (author) and LMU have unified the systematic treatment. Fabio Novissi (LMU) converted the systematics to the *SysVar* package [38], allowing for consistent encoding across analyses and facilitating future combined results. While his work is not presented here, regular discussions with Fabio, Dr. Thomas Lueck (LMU), and Dr. Thomas Kuhr (LMU) were essential to the development of this analysis. Additional key contributions came from regular discussions with Dr. Markus Prim (Bonn), Dr. Florian Bernlochner (Bonn), Dr. Slavomira Stefkova (Bonn), Ilias Tsaklidis (Bonn), and others, whose guidance helped shape the presented analysis.

This analysis would not have been possible without the support and guidance of my main supervisor, Dr. Alexander Glazov (DESY), who not only contributed to the analysis direction but also to troubleshooting logic and encoding errors. Gratitude is extended to Ana Luisa Moreira de Carvalho (DESY), Marcel Hohmann (DESY), Fabio Novissi (LMU), and Tommy Martinov (Trieste) who entertained and participated in discussions on the presented work. In particular, to Marcel who proofread an early draft of this thesis along with Dr. Alexander Glazov, Luisa, and Fabio who offered support and patience on a daily basis.

References

[1] T. Abe and et al. Belle II Technical Design Report. Technical report, KEK, 2010.

[2] I. Adachi and et. al. Test of lepton flavor universality with measurements of $R(D^+)$ and $R(D^{*+})$ using semileptonic B tagging at the Belle II experiment. *arXiv preprint*, 2025.

[3] Y. Amhis and et. al. Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016. *Eur. Phys. J. C*, 2017.

[4] G. Barr, R. Devenish, R. Walczak, and T. Weidberg. Accelerators. In *Particle Physics in the LHC Era*. Oxford University Press, 01 2016.

[5] Belle II Collaboration. Systematic Corrections Framework, 2024.

[6] Belle II Collaboration. Belle II Software Documentation, 2025.

[7] James Bergstra, Daniel Yamins, and David D. Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*, pages I–115–I–123. JMLR Workshop and Conference Proceedings, June 2013.

[8] Florian U. Bernlochner and Zoltan Ligeti. Semileptonic $B_{(s)}$ decays to excited charmed mesons with e, μ, τ and searching for new physics with $R(D^{**})$. *arXiv preprint*, 2016.

[9] Florian U. Bernlochner and Zoltán Ligeti. Semileptonic $B_{(s)}$ decays to excited charmed mesons with e, μ, τ and searching for new physics with $R(D^{**})$. 2016.

[10] Florian U. Bernlochner and Zoltan Ligeti. Semileptonic $B_{(s)}$ decays to excited charmed mesons with e, μ, τ and searching for new physics with $R(D^{**})$. *Phys. Rev. D*, 2017.

[11] Florian U. Bernlochner, Zoltan Ligeti, Michele Papucci, Markus T. Prim, Dean J. Robinson, and Chenglu Xiong. Constrained second-order power corrections in HQET: $R(D^{(*)})$, $|V_{cb}|$, and new physics. *Phys. Rev. D*, 2022.

[12] Florian U. Bernlochner, Zoltan Ligeti, and Dean J. Robinson. Model-independent analysis of semileptonic B decays to D^{**} for arbitrary new physics. *Phys. Rev. D*, 2018.

[13] Alessandro Bettini. *Introduction to Elementary Particle Physics*. Cambridge University Press, Cambridge, UK, 2 edition, 2008.

[14] A. J. Bevan and et al. *The Physics of the B Factories*. 2015.

[15] Alejandro Celis, Martin Jung, Xin-Qiang Li, and Antonio Pich. Combined analysis of semileptonic B decays to D and D*: R(D*), —V$_{cb}|andnewphysics.$ 2017.

[16] Chuan-Hung Chen and Takaaki Nomura. Charged-Higgs on $R_{D^{(*)}}$, τ polarization, and FBA. *The European Physical Journal C*, 2017.

[17] Tianqi Chen and et. al. *XGBoost: Extreme Gradient Boosting*, 2025. Open-source software, Apache License 2.0.

[18] Belle II Collaboration. Top view of the Belle II detector, 2022.

[19] Andreas Crivellin, David Müller, and Toshihiko Ota. Simultaneous explanation of $R(D^{(*)})$ and $b \to s\mu^+\mu^-$: the last scalar leptoquarks standing. *Journal of High Energy Physics*, 2017.

[20] Filippo Dattola and II Collaboration Belle˙Search for $B^+ \to K^+\nu\bar{\nu}$ decays with an inclusive tagging method at the Belle II experiment. *arXiv preprint*, 2021.

[21] Daniel Dorner, Petar Rados, and Christoph Schwanda. Relative tracking efficiency study of charged, low momentum pions, 2024. Iternal Belle II Note.

[22] Stephan Duell, Florian U. Bernlochner, Zoltan Ligeti, Michele Papucci, and Dean J. Robinson. *An Introduction to Hammer: Helicity Amplitude Module for Matrix Element Reweighting.*

[23] Daniel Ferlwezicz, Phillip Urquijo, and Eiasha Waheed. Revisiting fits to $B^0 \to D^{*-}\ell^+\nu_\ell$ to measure $|V_{cb}|$ with novel methods and preliminary LQCD data at non-zero recoil. *Phys. Rev. D*, 2021.

[24] R. Glattauer and et al. Measurement of the decay $B \to D\ell\nu$ in fully reconstructed events and determination of the Cabibbo–Kobayashi–Maskawa matrix element $|V_{cb}|$. *Phys. Rev. D*, 2016.

[25] David Griffiths. *Introduction to Elementary Particles.* Wiley-VCH, Weinheim, Germany, 2nd revised edition edition, 2009.

[26] Heavy Flavor Averaging Group (HFLAV). Average of $R(D)$ and $R(D^*)$ for CKM 2025. `https://hflav-eos.web.cern.ch/hflav-eos/semi/ckm25/html/RDsDsstar/RDRDs.html`, 2025. Last updated 28 Sep 2025.

[27] Lukas Heinrich, Matthew Feickert, and Giordon Stark. pyhf: v0.7.6.

[28] Henrik Alexander Junkerkalefeld. *Tests of lepton universality in the rates of inclusive semileptonic B-meson decays at Belle II.* PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, July 2024.

[29] Ilias Tsaklidis and et. al. Measurement of the $R(D^*)$ ratio and $\mathcal{P}_\tau inhadronic1 - prong\tau$ decays with the hadronic FEI. https://docs.belle2.org/files/4547/BELLE2-NOTE-PH-2025-031/8/BELLE2-NOTE-PH-2025-031.pdf, 2025. Internal Belle II Note.

[30] Q. P. Ji, S. Jia, and C. P. Shen. Energy resolution and bias with $E_\gamma < 2.0$ GeV using symmetric decays of $\pi^0 \to \gamma\gamma$ and $\eta \to \gamma\gamma$ at Belle II, 2021. Internal Belle II Note.

[31] Zoltán Ligeti and Frank J. Tackmann. Precise predictions for $b \to x_c \tau \bar{\nu}$ decay distributions. *Physical Review D*, 90(3):034021, 2014.

[32] Michael Eliachevitch. PIDvar: Offline Particle ID efficiency and fake rate corrections for Belle II. `https://gitlab.desy.de/meliache/pidvar`, 2023.

[33] S. Navas and Particle Data Group. Review of particle physics. *Phys. Rev. D*, 110:030001, 2024. Available online at `https://pdg.lbl.gov/`.

[34] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Westview Press, 1995.

[35] Markus Prim. Update of mc branching fractions. Internal presentation / note, 2023. Based on previous work by Florian Bernlochner.

[36] M. Tanabashi and et al. (Particle Data Group). Review of Particle Physics. *Phys. Rev. D*, 2018.

[37] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[38] Ilias Tsaklidis. Combining $R(D^{(*)})$ measurements at BelleII. Talk at Challenges in Semileptonic B Decays, 2024.

[39] Ilias Tsaklidis. Lepton flavor universality at belle ii, 2025. Talk presented at CKM 2025.

[40] Shoji Uno. Status of CDC (Central Drift Chamber) at Belle II, 2024. Talk presented at The 3rd DRD1 meeting.

A Table Summaries and Signal Extraction Binning

A.1 Event Selection

Event Selection	
$5 \leq$ nTracks ≤ 15 $3 \leq$ nTracks(goodTrack) ≤ 11 $4.0 < E_{vis} < 10.0$	
$K^{\pm}$ goodTrack $0.3 <$ p < 3.0 # CDC Hits > 20 & KID > 0.5	
$\pi_A^{\pm}$ goodTrack $0.3 <$ p < 3.0 # CDC Hits > 20 & πID > 0.1	
$\pi_B^{\pm}$ goodTrack ($0.3 <$ p < 3.0 & πID > 0.1 & # CDC Hits > 20) OR (p < 0.3 & # CDC Hits > 0)	
$e^{\pm}$ goodTrack $0.4 < p_{e^{\pm}} < 3.0$ eID > 0.9	$\mu^{\pm}$ goodTrack $0.7 < p_{\mu^{\pm}} < 3.0$ μID > 0.9
$1.85 < D^0$ InvM < 1.88 $p_D^* < 2.5$	
B treeFit (conf_level=0.001)	

Table A.1: Summary of event selection criteria. Selections are explained in the following text. "goodTrack" refers to $\theta \in$ CDC Acceptance , $dr < 1.0$ cm, and $|dz| < 3.0$ cm. "nTracks(goodTrack)" corresponds to tracks loaded in as pions with the described "goodTrack" selections. E_{vis} refers to the visible energy of the event. All energy selections are in GeV and all distance selections are in cm.

ROE Selection		
	track	**cluster**
loose mask	$\theta \in$ CDC Acceptance & $\lvert$dz$\rvert < 20$ & dr < 10	goodGamma
tight mask	$\theta \in$ CDC Acceptance & $\lvert$dz$\rvert < 4$ & dr < 2 & nCDCHits > 0	goodGamma & clusterTiming < 200 & $\theta \in$ CDC Acceptance & minC2TDist > 30

Table A.2: Simple ROE mask placed to clean up the tag-side and ensure a general "good" B_{tag} candidate. "goodGamma" refers to photons selected with $E > 0.1$ GeV, $E > 0.05$ GeV in the endcaps and barrel regions, respectively. *minC2TDist* refers to the minimum distance between the cluster and the nearest track candidate. All energy selections are in GeV, all distance selections are in cm, and all timing selections are in ns.

A.2 BDT1

XGBoost Hyper-Parameters			
Hyper-Parameter	Description	Parameter Space	Optimized Value
max depth	maximum depth of tree - increases complexity (generates over-fitting) [default=6]	log uniform integers: [1,8]	6
max bin	maximum number of discrete bins of continuous features [default=256]	uniform integers: [256, 310]	270
min child weight	minimum sum of instance weight for a child - decreases partitioning (more conservative) [default=1]	uniform integers: [200, 300]	270
eta	step size shrinkage (learning rate) - prevents over-fitting [default=0.3]	log uniform: [0.001, 0.5]	0.11
subsample	subsample ratio of each boosting iteration - prevents over-fitting [default=1]	uniform: [0.5, 1]	0.94
gamma	minimum loss required to make a partition on the node (more conservative) [default=0]	uniform: [0,5]	0

Table A.3: Bayesian optimized XGBoost hyper-parameters for the BDT1 classifier (non-integers rounded to 2 significant figures). Max depth and eta (learning rate) have parameter spaces associated with a log uniform distribution to prioritize the selection of lower values without explicitly excluding the region of larger values in the search. [7] [17] The respective parameter space ranges are determined through knowledge of the variable impact as well as from manual grid searching completed in previous iterations of this analysis.

A.3 BDT2

XGBoost Hyper-Parameters			
Hyper-Parameter	Description	Parameter Space	Optimized Value
max depth	maximum depth of tree - increases complexity (generates over-fitting) [default=6]	log uniform integers: [1,8]	7
max bin	maximum number of discrete bins of continuous features [default=256]	uniform integers: [256, 310]	271
min child weight	minimum sum of instance weight for a child - decreases partitioning (more conservative) [default=1]	uniform integers: [200, 300]	212
eta	step size shrinkage (learning rate) - prevents over-fitting [default=0.3]	log uniform: [0.001, 0.5]	0.08
subsample	subsample ratio of each boosting iteration - prevents over-fitting [default=1]	uniform: [0.5, 1]	0.96
gamma	minimum loss required to make a partition on the node (more conservative) [default=0]	uniform: [0,5]	0

Table A.4: Bayesian optimized XGBoost hyper-parameters for the BDT2 classifier (non-integers rounded to 2 significant figures). Max depth and eta (learning rate) have parameter spaces associated with a log uniform distribution to prioritize the selection of lower values without explicitly excluding the region of larger values in the search. [7] [17] The respective parameter space ranges are determined through knowledge of the variable impact as well as from manual grid searching completed in previous iterations of this analysis.

A.4 Fit Region Definitions

Signal Region (SR)	$\mathcal{O}_{BDT2,\,sig} > 0.8$
Normalization Region (NR)	$\mathcal{O}_{\mathrm{BDT2,\,sig}} < 0.8$ $\mathcal{O}_{\mathrm{BDT2},\,q\bar{q}} < 0.1$ & $\mathcal{O}_{\mathrm{BDT2,\ \ fakeD}} < 0.125$ $\mathcal{O}_{\mathrm{BDT2},\,D^{**}} < 0.2$ & $\mathcal{O}_{\mathrm{BDT2,\ \ fakeB}} < 0.1$
Mis-reconstructed $\mathbf{D^0}$ CR	$\mathcal{O}_{\mathrm{BDT2,\,fakeD}} > 0.125$ $\mathcal{O}_{\mathrm{BDT2},\,q\bar{q}} < 0.1$ & $\mathcal{O}_{\mathrm{BDT2,\ \ sig}} < 0.8$ $\mathcal{O}_{\mathrm{BDT2},\,D^{**}} < 0.2$ & $\mathcal{O}_{\mathrm{BDT2,\ \ fakeB}} < 0.1$
Continuum CR	$\mathcal{O}_{\mathrm{BDT2},\,q\bar{q}} > 0.1$ $\mathcal{O}_{\mathrm{BDT2,\,sig}} < 0.8$ & $\mathcal{O}_{\mathrm{BDT2,\ \ fakeD}} < 0.125$ $\mathcal{O}_{\mathrm{BDT2},\,D^{**}} < 0.2$ & $\mathcal{O}_{\mathrm{BDT2,\ \ fakeB}} < 0.1$
$\mathbf{D^{**}}$ CR	$\mathcal{O}_{\mathrm{BDT2},\,D^{**}} > 0.2$ $\mathcal{O}_{\mathrm{BDT2},\,q\bar{q}} < 0.1$ & $\mathcal{O}_{\mathrm{BDT2,\ \ fakeD}} < 0.125$ $\mathcal{O}_{\mathrm{BDT2,\,sig}} < 0.8$ & $\mathcal{O}_{\mathrm{BDT2,\ \ fakeB}} < 0.1$
Mis-reconstructed B CR	$\mathcal{O}_{\mathrm{BDT2,\,fakeB}} > 0.1$ $\mathcal{O}_{\mathrm{BDT2},\,q\bar{q}} < 0.1$ & $\mathcal{O}_{\mathrm{BDT2,\ \ fakeD}} < 0.125$ $\mathcal{O}_{\mathrm{BDT2},\,D^{**}} < 0.2$ & $\mathcal{O}_{\mathrm{BDT2,\ \ sig}} < 0.8$

Table A.5: Selection criteria for the two-dimensional SR and NR and the 4 one-dimensional background dominant control regions. The mutually exclusive regions are defined based on the respective BDT2 output classifiers $\mathcal{O}_{\mathrm{BDT2,\,i}}$.

A.5 Fit Binning

Bin Edges	
Signal Region (SR)	q^2: [1, 4, 5, 6, 7, 8, 9, 10, 12] $\mathcal{O}_{\text{BDT2, sig}}$: [0.8, 0.85, 0.9, 0.95, 1.0]
Normalization Region (NR)	q^2: [1, 4, 5, 6, 7, 8, 9, 10, 12] $\mathcal{O}_{\text{BDT2, sig}}$: [0.6, 0.65, 0.7, 0.75]
Mis-reconstructed D^0 CR	[0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.3, 0.4]
Continuum CR	[0.1, 0.15, 0.2, 0.4]
D^{**} CR	[0.2, 0.25, 0.3, 0.4]
Mis-reconstructed B CR	[0.1,0.125, 0.15, 0.35]

Table A.6: Configured bin edges for the flattened $\mathcal{O}_{\text{BDT2, sig}}$ x q^2 histogram, see Figure A.3, of the BDT2 signal classifier output x q^2 in the SR and NR and the one-dimensional background dominant control regions (CR). To provide ease in configuring the fit, the bins (Figure A.3) are mapped such that the signal bins go from 0-31 and the normalization bins span. It should be noted that the pre-mapped bins 14 and 15 in Figure A.3 are merged to ensure sufficient statistics of each template in each bin.

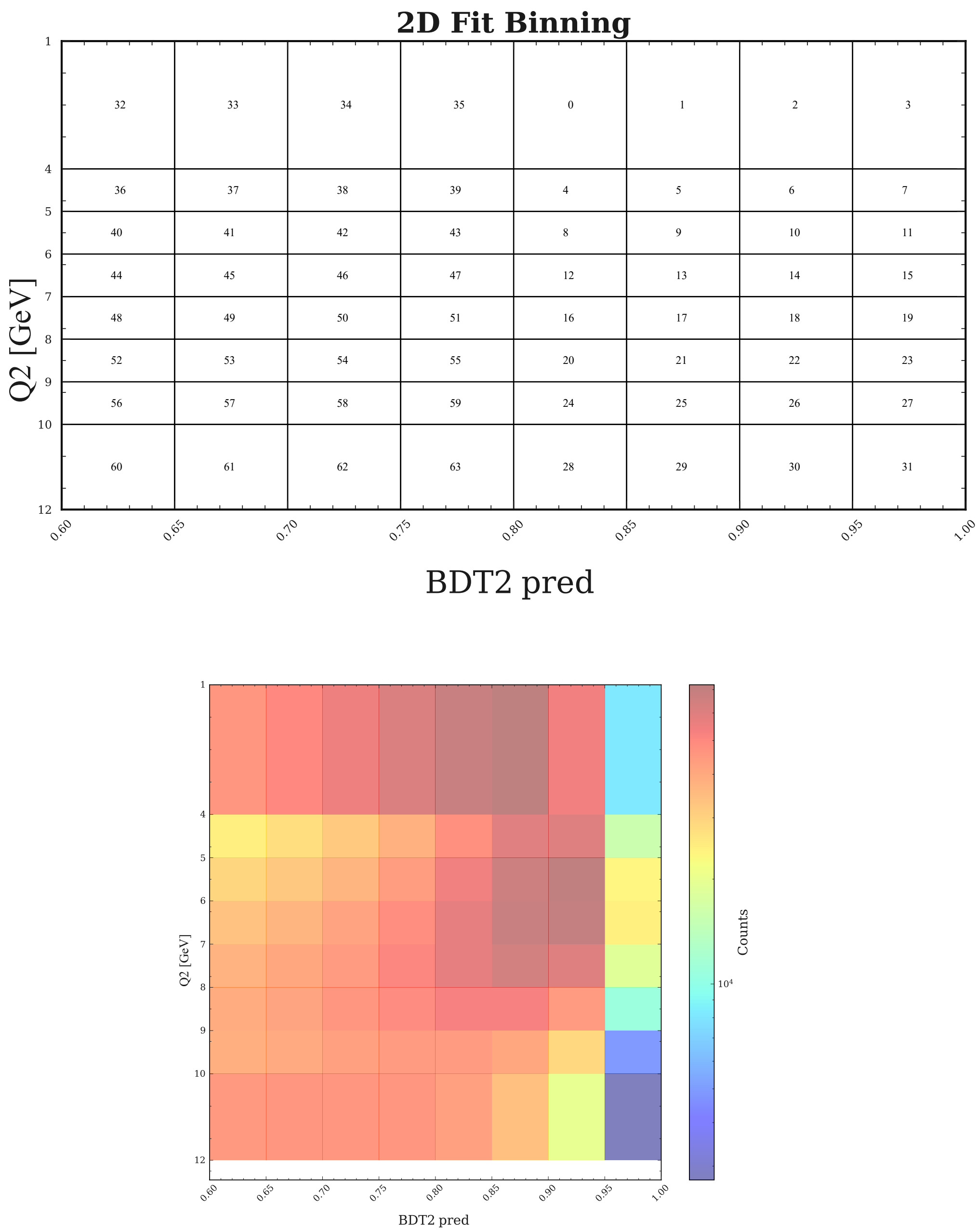


Figure A.1: Encoding of flattened $\mathcal{O}_{\text{BDT2, sig}}$ x q^2 bin indices (top) and full sample bin density (bottom).

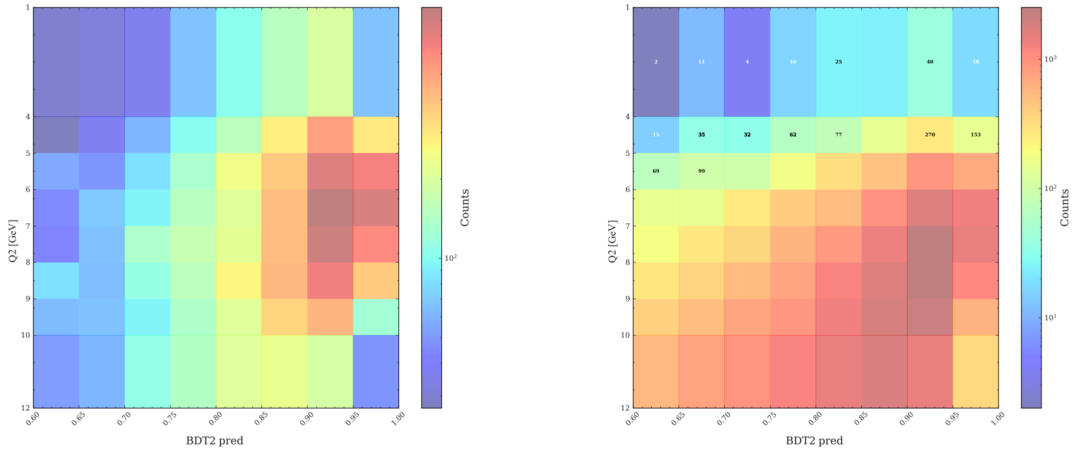


Figure A.2: Bin density of $\mathcal{O}_{\mathrm{BDT2,\ sig}}$ x q^2 for $B \to D\tau\nu$ (left) and $B \to D^*\tau\nu$ (right).

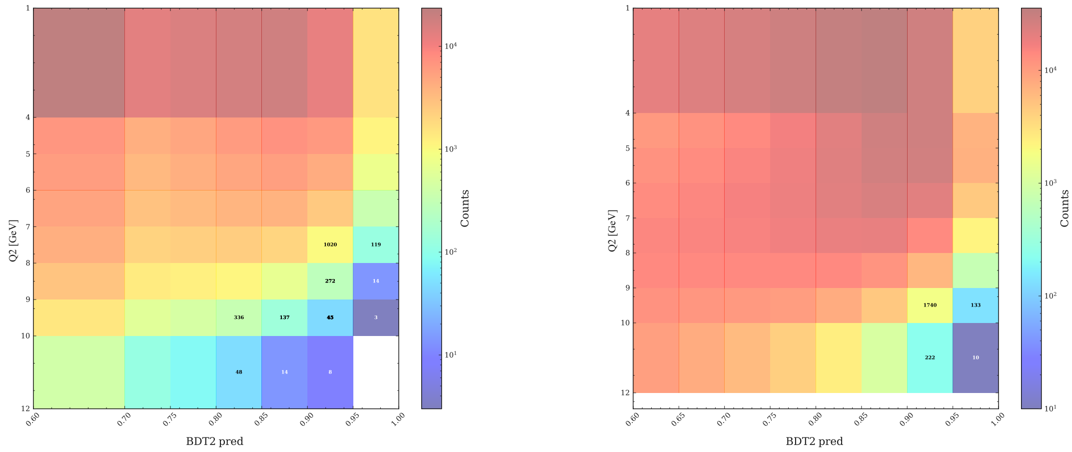


Figure A.3: Bin density of $\mathcal{O}_{\mathrm{BDT2,\ sig}}$ x q^2 for $B \to D\ell\nu$ (left) and $B \to D^*\ell\nu$ (right).

A.6 Uncertainty Projections

Published Belle II $R(D^{(*)})$ Uncertainties				
Source	**Semi-leptonic Tag** [2]		**Hadronic Tag** [39]	
	$R(D)$	$R(D^*)$	$R(D)$	$R(D^*)$
MC Sample Size	8.1%	4.1%	4.8%	8.4%
$B \to D^{**}\tau/\ell\nu$ (gap) BF	9.1%	0.1%	2.7%	3.1%
$B \to D^{**}\tau/\ell\nu$ BF	0.5%	0.1%	0.3%	1.3%
$\tau \to \ell\nu\bar{\nu}$ BF	0.2%	0.2%	-	-
Hadronic B Decay BF	-	-	1.6%	1.5%
Continuum Background	0.7%	0.2%	2.4%	2.1%
μID Efficiency	6.0%	0.5%	-	-
μID Fake Rate	0.1%	0.1%	-	-
eID Efficiency	0.3%	0.2%	-	-
eID Fake Rate	3.7%	0.7%	-	-
Form Factors	2.3%	3.6%	0.5%	0.9%
Low Momentum Tracking Efficiency	0.7%	0.5%	2.2%	2.4%
Tagging Efficiency	-	-	0.9%+0.1%	1.8%+1.8%
M^2_{miss} Resolution	-	-	0.5%	0.8%
Fit Biases	-	-	0.9%	1.2%
Misreconstructed $D^{(*)}$	-	-	0.5%	1.2%
Other	0.3%	0.5%	0.7%	1.4%
Total Systematic Unc.	13.4%	5.8%	6.7%	16.3%
Total Statistic Unc.	17.9%	11.4%	8.3%	16.3%

Table A.7: Reported impact of systematic uncertainties on $R(D)$ and $R(D^*)$. In the case of the semi-leptonic tagged analysis, the relative uncertainty percentages are combined through $\sqrt{\sigma_{add.} + \sigma_{mult.}}$. The relative impacts from this analysis are merged in the same manner. All reported impacts for the hadronic tagged analysis are independently based on toy studies.

B Corrections and Systematics

To account for intrinsic discrepancies between data and Monte Carlo (MC) simulation, corrections are applied based on recommendations from the corresponding performance groups.

Data-MC agreement is generally evaluated in low-signal regions. Prior to BDT1 selections, the low purity ensures that signal effects are negligible. After BDT1, dedicated sidebands are used to validate modeling, and agreement with signal-enhanced regions is verified through comparisons with off-resonance data and track/cluster multiplicities.

B.1 B and D Branching Ratios

Simulating the correct relative branching ratios in MC is non-trivial due to interference between decays. As MC generators sum decays independently to ensure the total branching fraction is normalized to unity, interference is neglected. To correct for this, event-by-event weights are applied for the signal and normalization channels (D^0, D^*) as well as background D^{**} channels. All relevant B (Table B.1) and $D^{(*)}$ (Table B.2) decays are corrected.

Reconstruction bias is removed through reweighting based solely on the generator-level information. Because the large normalization branching fractions allow signal decays to occur on either or both sides of the $\Upsilon(4S)$, corrections are applied to both sides to reflect the correct event probability. In dedicated samples, the *0th daughter* corresponds to the fixed signal decay, while the *1st daughter* decays generically.

The 0th daughter is corrected through an equivalent luminosity:

$$\mathcal{L}_{equiv} = \frac{N_{gen}\mathcal{B}_{fixed}}{\sigma_{B\bar{B}}(2\mathcal{B}_{true} - \mathcal{B}_{true}^2)}\,, \tag{44}$$

where $N_{gen}\mathcal{B}_{fixed}$ is the number of generated decays for the signal channel, $\sigma_{B\bar{B}}$ is the $\Upsilon(4S)$ cross-section, and $\mathcal{B}_{true}$ is the isospin-averaged branching fraction.

The 1st daughter is reweighted according to $\mathcal{B}_{true}/\mathcal{B}_{generic\,MC}$. To preserve normalization of the 0th daughter, background events are rescaled by:

$$\frac{\sum \omega_L - \sum \omega_{\mathcal{B}=\mathrm{sig}}}{\sum \omega_{\mathcal{B}\neq\mathrm{sig}}}\,, \tag{45}$$

where ω_i is the weight corresponding to event i.

Generic samples are reweighted to the desired luminosity, $\mathcal{L} = N_i/\sigma_i$, per data type i. The $B \to X_c\tau/\ell\nu$ events not explicitly generated in the dedicated samples are reweighted according to branching fraction and the same background renormalization, done for dedicated sample, is applied to both sides of the decay.

After reweighting, the dedicated signal events replace the corresponding signal decays in the generic MC, improving the smoothness of distributions and reducing MC statistical uncertainty, which scales as $\sqrt{\sum \omega_i}$.

The same procedure is applied to properly reweight generated $D^0 \to K\pi$ and $D^0 \to K3\pi$ decays.

Systematic uncertainties for each branching fraction are evaluated by varying the nominal value up and down by 1σ. The resulting nuisance parameters encode the joint variation of both B mesons for each independent branching fraction.

Branching Fraction Corrections for B decays			
B Decay	$\mathcal{B}$ MC $[10^{-2}]$	$\mathcal{B}$ Averaged $[10^{-2}]$	AVG/MC
$B^- \to D^0 \ell\nu$	2.31	2.27±0.06	0.983
$\bar{B}^0 \to D^+ \ell\nu$	2.14	2.11±0.05	0.986
$B^- \to D^{*0} \ell\nu$	5.49	5.27±0.12	0.960
$\bar{B}^0 \to D^{*+} \ell\nu$	5.11	4.90±0.11	0.959
$B^- \to D^0 \tau\nu$	0.69	0.68±0.02	0.983
$\bar{B}^0 \to D^+ \tau\nu$	0.64	0.63±0.02	0.986
$B^- \to D^{*0} \tau\nu$	1.42	1.35±0.03	0.956
$\bar{B}^0 \to D^{*+} \tau\nu$	1.32	1.26±0.03	0.955

Table B.1: Isospin averaged corrections for branching fractions of B mesons [35]. $B \to D^{(*)}\tau\nu$ branching fractions are estimated as $R(D^{(*)}) * \mathcal{B}_{\text{avg}}(B \to D^{(*)}\ell\nu)$ where $R(D^{(*)})$ is taken from [15].

Branching Fraction Corrections for $D^{(*)}$ decays			
D^0 Decay	$\mathcal{B}$ MC $[10^{-2}]$	$\mathcal{B}$ Averaged $[10^{-2}]$	AVG/MC
$D^0 \to K^-\pi^+$	3.95	3.945±0.03	0.9987
$D^0 \to K^-3\pi^\pm$	8.07	8.22±0.14	1.0186
D^* Decay	$\mathcal{B}$ MC $[10^{-2}]$	$\mathcal{B}$ Averaged $[10^{-2}]$	AVG/MC
$D^{*0} \to D^0\pi^0$	64.70	64.70±0.90	1.0
$D^{*0} \to D^0\gamma$	35.30	35.30±0.90	1.0
$D^{*+} \to D^0\pi^+$	67.70	67.70±0.50	1.0

Table B.2: Corrections for branching fractions of $D^{(*)}$ mesons. [33]

B.1.1 D^{**} Branching Fraction Corrections

One of the primary background channels is the heavy 1P state of the D-meson, $B \to D^{**}\ell\nu$, which has contributions from the two broad resonances, D_0^* and D_1' (D_1^*), the two narrow resonances D_1 and D_2^*, $B \to D^{(*)}\pi(\pi)/K^{(*)}$, and the gap modes $B \to D^{(*)}\eta$. Due to the complicated nature of measuring the heavier D^{**} resonances, from overlapping decay products, only partial branching fractions have been observed. Furthermore, the respective branching fractions are computed through partial branching fractions assuming isospin symmetry (Tables B.3, B.4).

Post summation of the resonant branching fractions (Table B.3) and the non-resonant branching fractions (Table B.4), there is still a difference between the inclusive branching fraction and the summation of the exclusive branching fractions which is filled with the gap modes: $B \to D\eta\ell\nu$ and $B \to D^*\eta\ell\nu$, each of which are assigned 100% uncertainty (Table B.4).

Branching Fraction Corrections for $B \to D^{**}\ell\nu$ decays			
B Decay	$\mathcal{B}$ MC $[10^{-2}]$	$\mathcal{B}$ Averaged $[10^{-2}]$	AVG/MC
$B^- \to D_1^0\ell^-\bar{\nu}_\ell$	0.76	0.64±0.10	0.845
$B^- \to D_0^{*0}\ell^-\bar{\nu}_\ell$	0.39	0.13±0.03	0.324
$B^- \to D_1^{\prime 0}\ell^-\bar{\nu}_\ell$	0.43	0.28±0.04	0.65
$B^- \to D_2^{*0}\ell^-\bar{\nu}_\ell$	0.37	0.32±0.03	0.861
$\bar{B}^0 \to D_1^+\ell^-\bar{\nu}_\ell$	0.70	0.59±0.1	0.838
$\bar{B}^0 \to D_0^{*+}\ell^-\bar{\nu}_\ell$	0.36	0.12±0.03	0.323
$\bar{B}^0 \to D_1^{\prime +}\ell^-\bar{\nu}_\ell$	0.40	0.26±0.04	0.648
$\bar{B}^0 \to D_2^{*+}\ell^-\bar{\nu}_\ell$	0.35	0.30±0.03	0.862

Table B.3: Isospin averaged corrections for branching fractions of background $B \to D^{**}\ell\nu$ decays. Corrections are only applied based on the first-level decay. [35]

Branching Fraction Corrections for $B \to D^{(*)}\pi(\pi)/\eta\,\ell\nu$ decays			
B Decay	$\mathcal{B}$ MC $[10^{-2}]$	$\mathcal{B}$ Averaged $[10^{-2}]$	AVG/MC
$B^- \to D\pi\ell^-\nu$	0.15	0.0	0.0
$B^- \to D^*\pi\ell^-\nu$	0.15	0.0	0.0
$B^- \to D\pi\pi\ell^-\nu$	0.05	0.07±0.09	1.175
$B^- \to D^*\pi\pi\ell^-\nu$	0.26	0.22±0.10	0.837
$B^- \to D^0\eta\ell^-\nu$	0.20	0.90±0.15	4.478
$B^- \to D^*\eta\ell^-\nu$	0.20	0.90±0.15	4.478
$B^+ \to D_sK\ell^+\nu$	0.03	0.03±0.01	1.0
$B^+ \to D_s^*K\ell^+\nu$	0.03	0.03±0.02	0.967
$\bar{B}^0 \to D\pi\ell^-\nu$	0.14	0.0	0.0
$\bar{B}^0 \to D^*\pi\ell^-\nu$	0.14	0.0	0.0
$\bar{B}^0 \to D\pi\pi\ell^-\nu$	0.05	0.07±0.08	1.429
$\bar{B}^0 \to D^*\pi\pi\ell^-\nu$	0.24	0.2±0.1	0.816
$\bar{B}^0 \to D\eta\ell^-\nu$	0.22	0.86±0.14	3.963
$\bar{B}^0 \to D^*\eta\ell^-\nu$	0.22	0.86±0.14	3.963

Table B.4: Isospin average corrections for branching fractions of background $B \to D^{**}\ell\nu$ non-resonance and gap decays. [35]

The tauonic D^{**} branching fractions are not directly stated. Therefore, the respective branching fractions are determined by multiplying the isospin-averaged $\mathcal{B}(B \to D^{**}\ell\nu)$ with the conservative D^{**} branching ratios stated in [10]:

$$R(D_2^*) = 0.06 \pm 0.01 \tag{46}$$

$$R(D_1) = 0.10 \pm 0.02 \tag{47}$$

$$R(D_1') = 0.06 \pm 0.02 \tag{48}$$

$$R(D_0^*) = 0.08 \pm 0.04 \tag{49}$$

Branching Fraction Corrections for $B \to D^{**}\tau\nu$ decays			
B Decay	$\mathcal{B}$ MC $[10^{-2}]$	$\mathcal{B}$ Averaged $[10^{-2}]$	AVG/MC
$B^- \to D_1^0 \tau^- \bar{\nu}_\tau$	0.13	0.06±0.02	0.454
$B^- \to D_0^{*0} \tau^- \bar{\nu}_\tau$	0.13	0.01±0.01	0.072
$B^- \to D_1^{\prime 0} \tau^- \bar{\nu}_\tau$	0.20	0.02±0.01	0.126
$B^- \to D_2^{*0} \tau^- \bar{\nu}_\tau$	0.20	0.02±0.0	0.105
$\bar{B}^0 \to D_1^+ \tau^- \bar{\nu}_\tau$	0.13	0.06±0.02	0.454
$\bar{B}^0 \to D_0^{*+} \tau^- \bar{\nu}_\tau$	0.13	0.01±0.01	0.072
$\bar{B}^0 \to D_1^{\prime +} \tau^- \bar{\nu}_\tau$	0.20	0.02±0.01	0.078
$\bar{B}^0 \to D_2^{*+} \tau^- \bar{\nu}_\tau$	0.20	0.02±0.0	0.105

Table B.5: Isospin average corrections for branching fractions of background $B \to D^{**}\tau\nu$ decays [35]. $B \to D^{(*)}\tau\nu$ branching fractions are estimated as $R(D^{(*)}) * \mathcal{B}_{\text{avg}}(B \to D^{(*)}\ell\nu)$ where $R(D^{(*)})$ is taken from [9].

The tauonic gap modes are assumed to fill the discrepancy between the inclusive isospin average prediction and the isospin averages of the simulated channels:

$$\mathcal{B}(B \to X\tau\nu)_i = \left(\frac{\mathcal{B}(B \to X\ell\nu)_i}{\sum \mathcal{B}(B \to X\ell\nu)}\right) * \left[R(X_c)\sum \mathcal{B}(B \to X_c\ell\nu) - \sum \mathcal{B}(B \to X_c\tau\nu)\right] \tag{50}$$

where the relative gap mode composition is assumed to be the same as in the light lepton gap channels and $R(X_c) = 0.223 \pm 0.005$ as computed in [31].

Branching Fraction Corrections for $B \to D^{(*)}\pi(\pi)/\eta\,\tau\nu$ decays			
B Decay	$\mathcal{B}$ MC $[10^{-3}]$	$\mathcal{B}$ Averaged $[10^{-3}]$	AVG/MC
$B^- \to D\pi\pi\tau^-\nu$	-	0.01±0.01	1.0
$B^- \to D^*\pi\pi\tau^-\nu$	-	0.03±0.02	1.0
$B^- \to D^0\eta\tau^-\nu$	-	0.13±0.13	1.0
$B^- \to D^*\eta\tau^-\nu$	-	0.13±0.13	1.0
$\bar{B}^0 \to D\pi\pi\tau^-\nu$	-	0.01±0.01	1.0
$\bar{B}^0 \to D^*\pi\pi\tau^-\nu$	-	0.03±0.02	1.0
$\bar{B}^0 \to D\eta\tau^-\nu$	-	0.13±0.13	1.0
$\bar{B}^0 \to D^*\eta\tau^-\nu$	-	0.13±0.13	1.0

Table B.6: Corrections for branching fractions of background $B \to D^{**}\tau\nu$ non-resonance and gap decays. The MC branching fractions for $D^{(*)}\pi\pi$ and $D^{(*)}\pi$ are the summation of all pion configurations for the respective $D^{(*)}$ contributions.

Systematic uncertainties for each branching fraction are evaluated by varying the nominal value up and down by 1σ. The resulting nuisance parameters encode the joint variation of both B mesons for each independent branching fraction.

B.2 Form Factor Modeling in $B \to X\ell\nu$

The semi-leptonic B decays in the generic MC are modeled with EVTGEN. However, the form-factor parameterizations in the current simulation are not up-to-date with current values. To account for this, the MC simulation is corrected with event-by-event weights:

$$\omega_i = \frac{\Gamma_{MC}}{\Gamma_{corr}} \frac{d\Gamma_{corr}}{d\Gamma_{MC}} \tag{51}$$

via the software package `HAMMER` (Helicity Amplitude Module for Matrix Element Reweighting) version 1.4.0.

Form Factor Corrections			
Decay	Form Factor Model	Parameters	References
$B \to D\ell\nu$	BGLB2→BLPRXP	8→9	[24], [11]
$B \to D^*\ell\nu$	BGL→BLPRXP	6→9	[23], [11]
$B \to (D_0^*, D_1')\ell\nu$	LLSW (BLR)	4	[8], [12]
$B \to (D_1, D_2^*)\ell\nu$	LLSW (BLR)	4	[8], [12]
$B \to D\tau\nu$	CLN (HQET3) → BLPRXP	3→9	[3], [11]
$B \to D^*\tau\nu$	CLN (HQET3) → BLPRXP	5→9	[3], [11]
$B \to (D_0^*, D_1')\tau\nu$	LLSW (BLR)	4	[8], [12]
$B \to (D_1, D_2^*)\tau\nu$	LLSW (BLR)	4	[8], [12]

Table B.7: Form factor parameterization for semi-leptonic $B \to X_c\ell\nu$ transitions. The corrected form-factor parameterizations for $B \to D^{(*)}\ell/\tau\nu$ incorporate in the HQET relation to $\mathcal{O}(M_{QCD}/m_{c,b})$ based on a combined fit of D and D^* with the BLPRXP model [11].

To properly account for the collinear radiative photons appended to the vertices with PHOTOS, the MC truth is extracted from the root level of the generated MC before the QED radiative corrections. In doing so, the respective four-momentum does not need to undergo kinematic rebalancing and does not require inclusion of additional higher multiplicity vertices in HAMMER.

One nuisance parameter is encoded for each systematic shape uncertainty, evaluated in HAMMER through rotating the parameters into an uncorrelated eigenbasis. The respective variations are found by varying the central values by $\pm 1\sigma$ ($\pm 1.5\sigma$ for the broad D^{**} mesons) where σ is the eigenvalue of the respective rotated eigenbasis. [28]

B.3 Track Momentum Scale

To take into account potential shifts in momenta from imperfections in the magnetic field map, each D-meson mass window is given a loose cut of 3σ. Doing so removes potential biases that may result from signal falling outside the selection window. To further account for the discrepancy, corrections are applied online based on recommendations from the tracking performance group.

Systematics will be extracted through reprocessing the sample with weight tables corresponding to the up and down variations and will be encoded as one nuisance parameter.

B.4 Photon Energy Bias

Based on recommendations from Neutral Performance Group in [30], discrepancies in ECL cluster energies are accounted for through applying the respective correction tables during reconstruction.

Systematics will be extracted through reprocessing the sample with weight tables corresponding to the up and down variations and will be encoded as one nuisance parameter.

B.5 Particle Identification

Deviations from the applied particle efficiency and fake rate correction factors are taken into account through applying event-by-event correction factors with the Systematics Framework (SF) [5]. SF provides reconstruction of clean calibration channels through simple momentum selections. Momentum and theta-dependent correction factors are then determined through an unbinned fit on the mass in bins of p and θ. Application of the associated weighted is completed through *PIDVar* [32], which applies the generated weight table to corresponding events.

Systematics will be determined through *SysVar* [38] via the generation of toys. Each toy is pulled from a multidimensional Gaussian distribution where the standard deviation is the combined total uncertainty. These toy variations are used to construct a covariance matrix in dimensions corresponding to (n samples x fit bins) on which eigen-decomposition is performed. The resulting large set of eigenvariations will be turnicated, resulting in a small number of independent uncorrelated nuisance parameters encoded in the fit. The remaining eigenvariations are approximated to be uncorrelated and are thus diagonalized and summed, such that they can be encoded as one additional nuisance parameter, providing a better estimation of the full covariance matrix.

B.5.1 Lepton ID

SF only has the J/Ψ model implemented for leptonic final states which requires a lower momentum threshold of 1.5 GeV. Version 0.7.0 of SF implemented a J/Ψ model that decorrelates the lepton momenta through completing an angular transformation, thereby allowing the momentum region to go below 1.5. Unfortunately, this did not produce significant effects. Considering further models such as $ee \to \gamma$ and $ee \to \ell\ell$ are still being workshopped in SF, the central efficiency tables are implemented.

Fake rates, on the other hand, use the SF hadronic models with the binning determined based on the lepton efficiency tables. Therefore, these are generated through the SF framework where the standard constraints are removed, such that no requirements are placed on the number of CDC hits. Pre-cuts for determination of the resulting corrections in each bin are defined as $dr < 1$ cm and $|dz| < 3$ cm which have a minimal effect on the corrections. Most notable is the case where pions are misIDed as electrons, where the tight PID selection and low fake rates result is relatively significant statistical fluctuations between bins, resulting in deviations between the charge-independent nominal correction values.

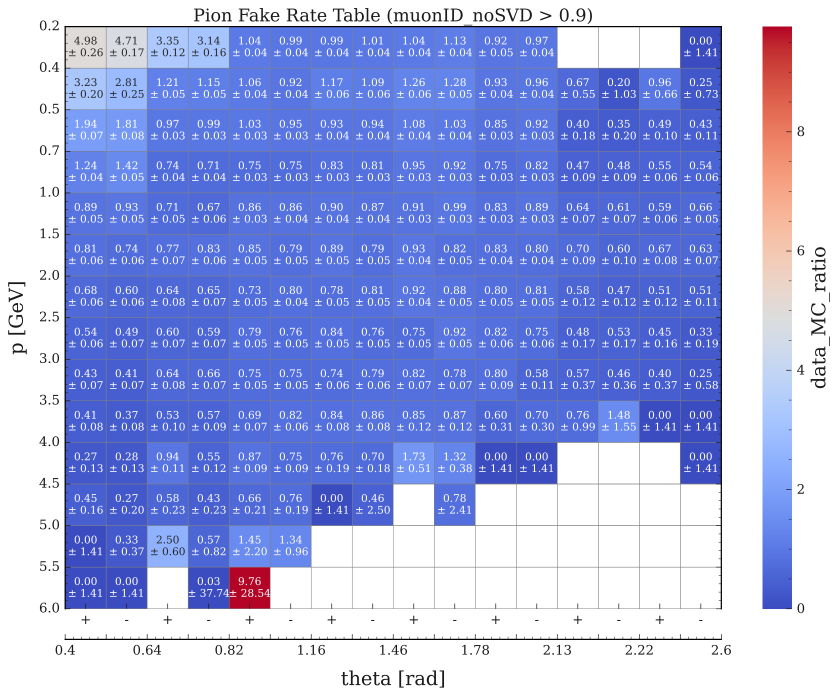


Figure B.1: Fake rate corrections of pion mis-identification in the muon candidate. Relatively large deviations occur between bins due to low statistics (both at the edges of the distribution and internally), the transition between the barrel and endcap.

B.5.2 Kaon and Pion ID

Per recommendations by Alessandro Gaz, pion fake rates and efficiencies are determined based on the $D^{*+} \to (D^0 \to K^-\pi^+)\pi^+$ as it is cleaner and provides smaller systematic uncertainties. Kaon fake rates and efficiencies are likewise determined based on the D^* model. Below $p = 0.3$ GeV, there is very little statistics for the determination of fake rates corrections so signal kaons and pions are required to have $p > 0.3$ (Section 3.2.1) such that proper PID corrections can be applied. The resulting efficiency and fake rate table, computed with pre-cuts of $dr < 1$ cm, $|dz| < 3$ cm, and # CDC Hits> 20 are done in the standard 14 p bins, 9 θ bins, and 2 charge bins, as displayed in Figure B.4, for example.

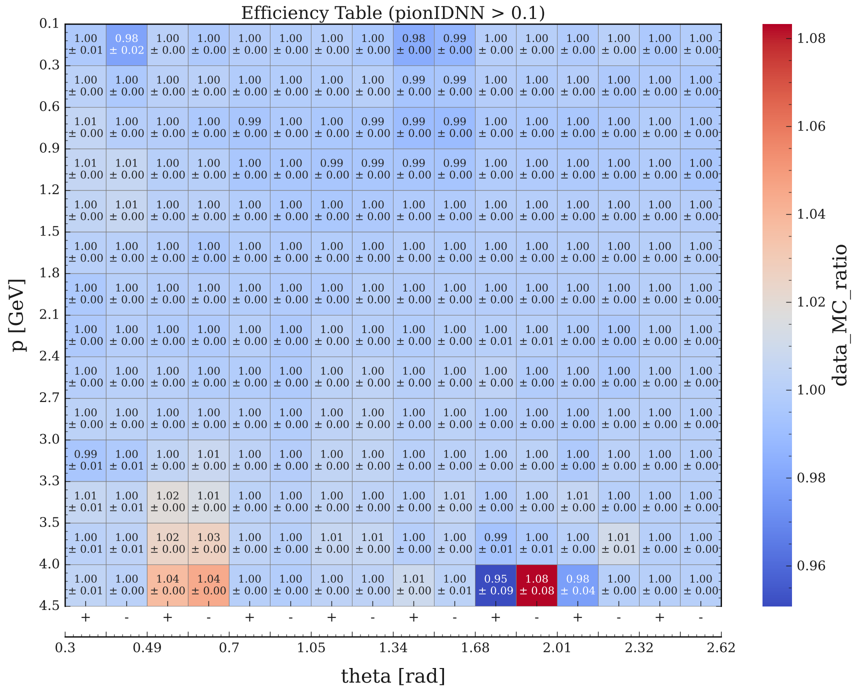


Figure B.2: Efficiency corrections of the pion candidate.

B.6 Slow Pion Tracking

Deviations in the relative tracking efficiency of slow pions lead to further discrepancies in data-MC. These are corrected for through reweighting the respective slow pion signal candidate momentum in a mode-dependent manner: $\prod \omega_i^{N(slow\,\pi_{sig})}$ where $N(slow\,\pi_{sig}) = (1, 2, 3)$ and i corresponds to the appropriate momentum bin. The correction values are applied in bins of momentum from 50 to 200 MeV in accordance with recommendations from [21]:

Charged Slow Pion Relative Efficiency Corrections	
Momentum Region [GeV]	**weight**
0.05-0.12	0.996 ± 0.020 (uncorr.) ± 0.013 (corr.) ± 0.003 (sys.)
0.12-0.16	0.990 ± 0.015 (uncorr.) ± 0.013 (corr.) ± 0.003 (sys.)
0.16-0.20	0.987 ± 0.017 (uncorr.) ± 0.013 (corr.) ± 0.003 (sys.)

Table B.8: Corrections and uncertainties for relative tracking relative tracking efficiency of charged π_s events normalized to the high momentum bin.

The three independent systematic variations, each encoded as uncorrelated nuisance parameters, are found by using the covariance matrix to compute the $\pm 1\sigma$ variation for each parameter in its corresponding eigendirection.

B.7 Broad D^{**} Resonances

As mentioned in B.1, the D^{**} resonances are poorly understood and in particular, the broad resonances D_0^* and $D_1^{'}$ which decay through the D-wave have overlapping final states leading to non-converging properties of the Breit-Wigner. The resulting large widths cause the simulated D^{**} mass to be significantly increased with respect to the nominal value causing an unphysical enhancement of the hadronic recoil:

$$\omega = \frac{m_B^2 + m_{X_c}^2 - q^2}{2 m_B m_{X_c}} \tag{52}$$

where q^2 is the invariant mass of the lepton and neutrino system squared. The enhancement is postulated to be due to the BLR model [8] implemented in the simulation that assumes a narrow width approximation. To correct for these unphysical contributions, events corresponding to the questionable phase space are removed. [28]

Broad D^{**} Resonance Corrections			
	Nominal Mass [GeV]	**Nominal Width [GeV]**	**Threshold [GeV]**
D_0^*	2.343	0.229	3.030
$D_1^{'}$	2.412	0.314	3.197
D_1	2.422	0.0313	-
D_2^*	2.461	0.0473	-

Table B.9: Central values for D^{**} resonances [36], corresponding upper selection thresholds placed on the broad resonances of $m + 3\sigma$ for D_0^* and $m + 2.5\sigma$ for $D_1^{'}$, and resulting truth matched efficiencies.

To account for the alternation of the total number of events, the remaining events are reweighted by a universal factor such that the sum of the weights is equal to the sum prior to the removal. Unfortunately, this leads to further uncertainty in the modeling and simulated yield of D_0^* and D_1' which is accounted for by varying the respective efficiencies and inflating the form factor uncertainties for the BLR models as described above. [28]

At the moment, this procedure is done based only on generic MC. In practice, however, the reweighting procedure outlined above will be done based on dedicated $B \to D^{**}\tau/\ell\nu$ samples (Section 6.1).

B.8 BDT Based Off-Resonance Correction

Due to a limited understanding of the composition of the off-resonance data, there are significant levels of mismodeling in continuum backgrounds. To reduce the level of data-MC discrepancy while maintaining statistics, a binary BDT is trained with classes corresponding to the full 42.84 fb^{-1} off-resonance dataset (class 1) and on-resonance MC (class 0). The resulting classifier output, C_i, is used to reweight on-resonance events by:

$$\omega_i = \frac{C_i}{1 - C_i} \, . \tag{53}$$

Between the acquisition of the preliminary results discussed in this thesis and the submission of the thesis, Fabio found a small error in the application of BDTc weights to the on-resonance MC. Rather than properly applying the weights to only on-resonance continuum events, the weights were applied to the entire on-resonance sample. Checks performed on the impact of this error showed a slight shift in low-momentum muon candidates, leading to better data-MC agreement. Post submission of this thesis, the procedure outlined above will be updated such that BDTc weights are only applied to on-resonance continuum events.

A corresponding nuisance parameter will encode 100% uncertainty on C_i.

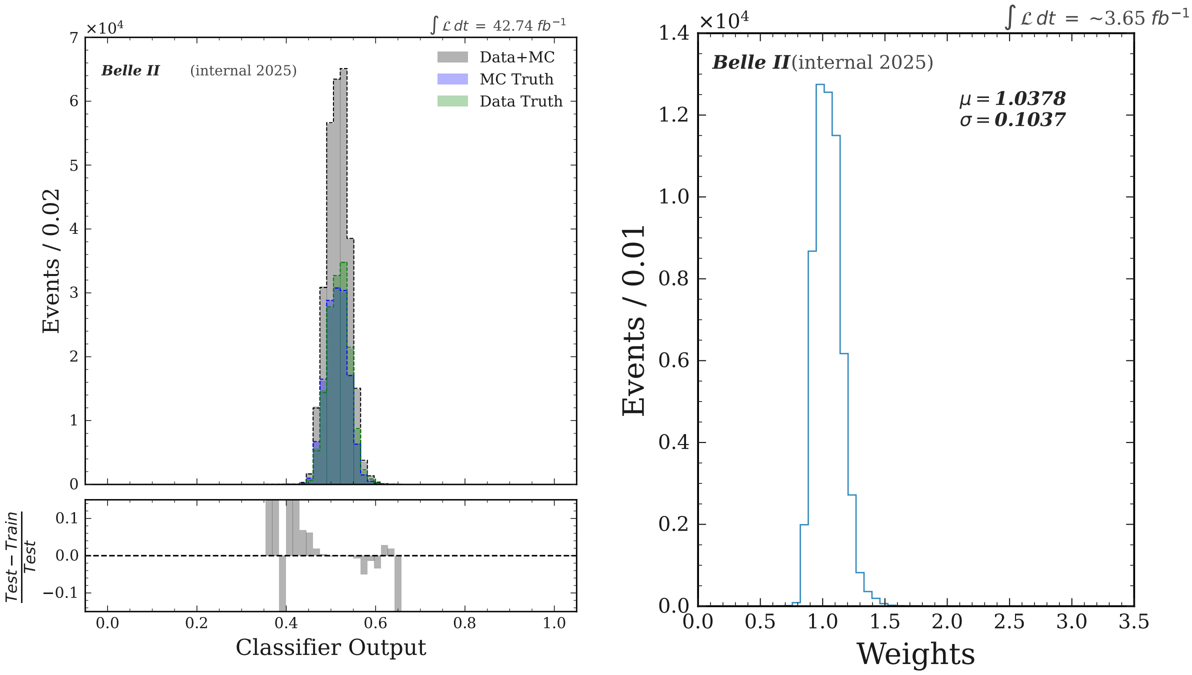


Figure B.3: Classifier output (left) and weights applied to $\sim 1\%$ on-resonance generic MC (right).

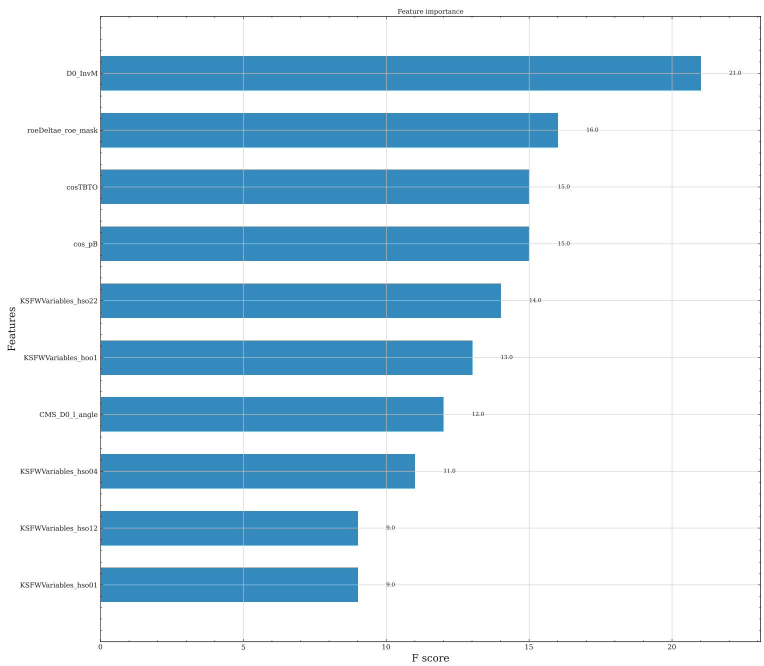


Figure B.4: BDTc input feature ranking based on XGBoost computed gain.

C Statistical Sensitivity of Simultaneous $R(D^{(*)})$ Fit

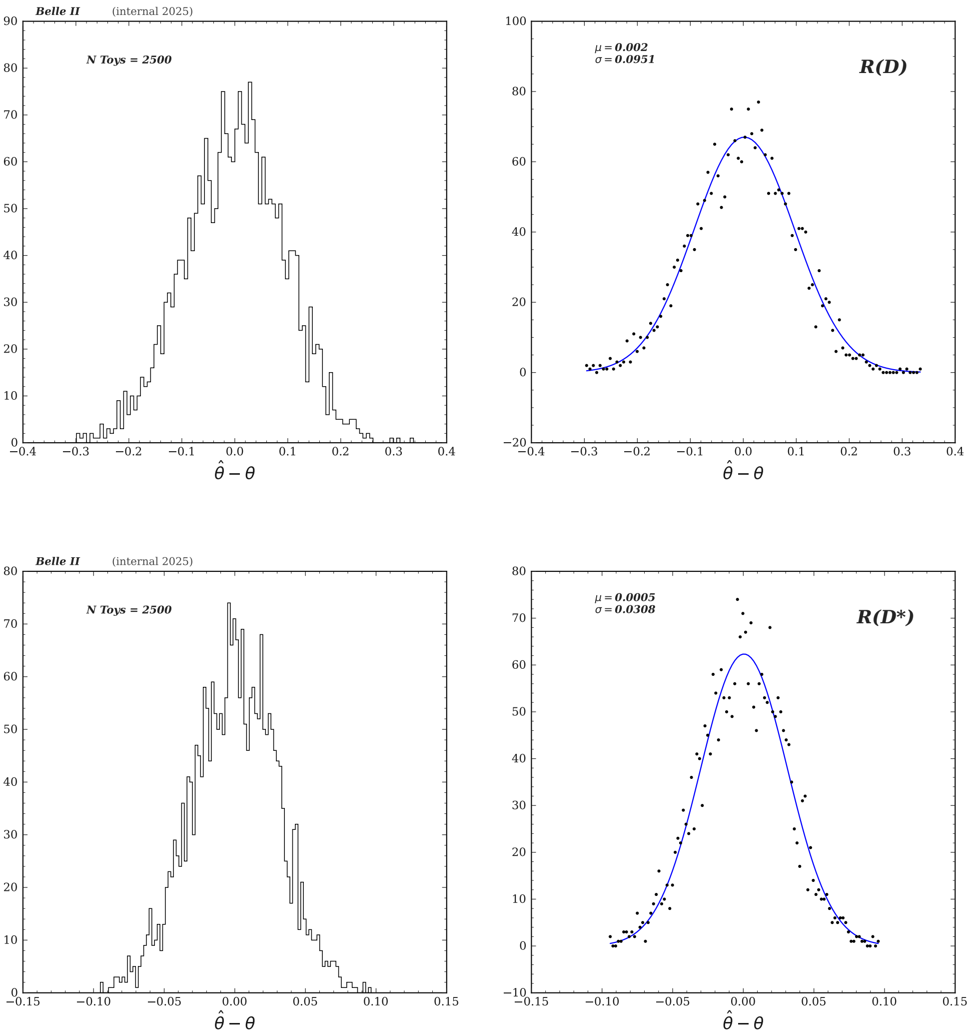


Figure C.1: Distribution of the simultaneous best-fit $R(D)$ value minus $R(D)_{SM}$ (top) and the $R(D^*)$ value minus $R(D^*)_{SM}$ (bottom) obtained from 2,500 pseudo-experiments (black points) used to estimate the statistical uncertainty. $\hat{\theta}$ represents the best-fit value of $R(D)$ or $R(D^*)$ and θ represents the corresponding SM prediction. The 1σ width of the overlaid Gaussian fit (blue line) provides the statistical uncertainty.

D BDT1 Input Features

Fifteen well-modeled features associated with event shape, thrust, and the ROE are used to provide separation between the three training classes.

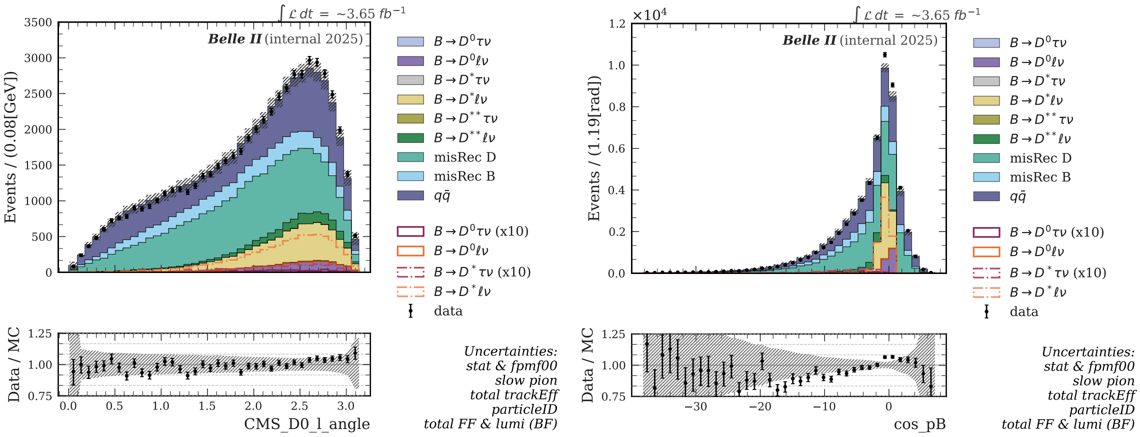


Figure D.1: Center-of-mass angle between the reconstructed D^0 and lepton candidate (left). The cosine of the angle between the reconstructed B and a nominal B (right), displays relatively poor modeling but exhibits high discrimination power in BDT1 training.

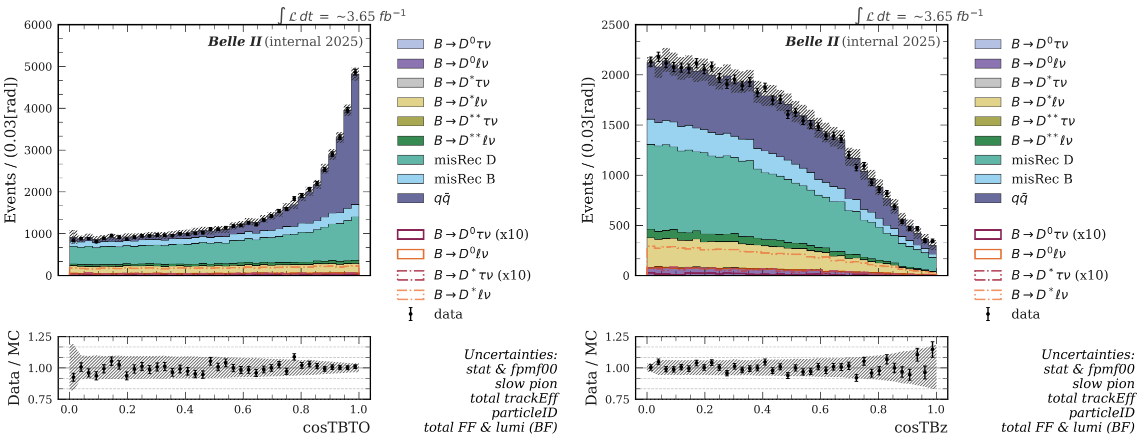


Figure D.2: Cosine of the angle between the reconstructed B thrust axis and the ROE thrust axis (left). Cosine between the reconstructed B thrust axis and the z-axis (right).

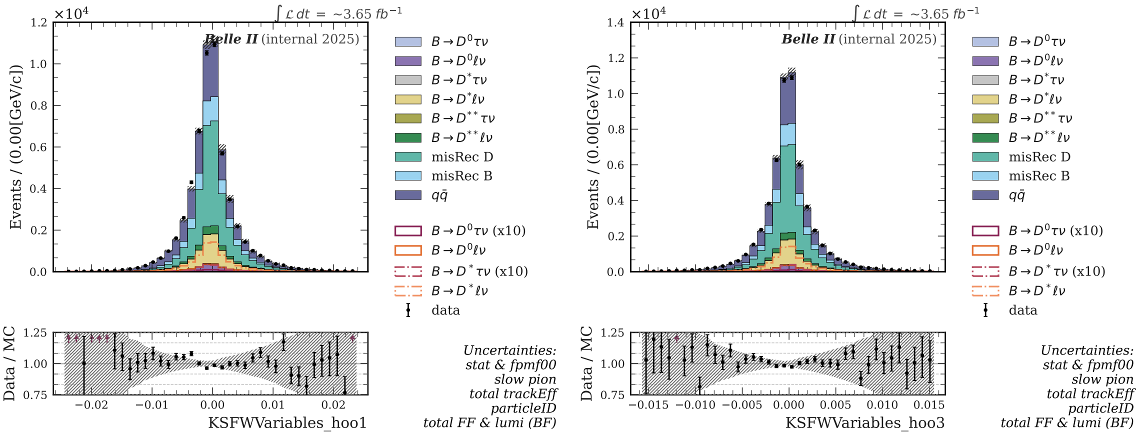
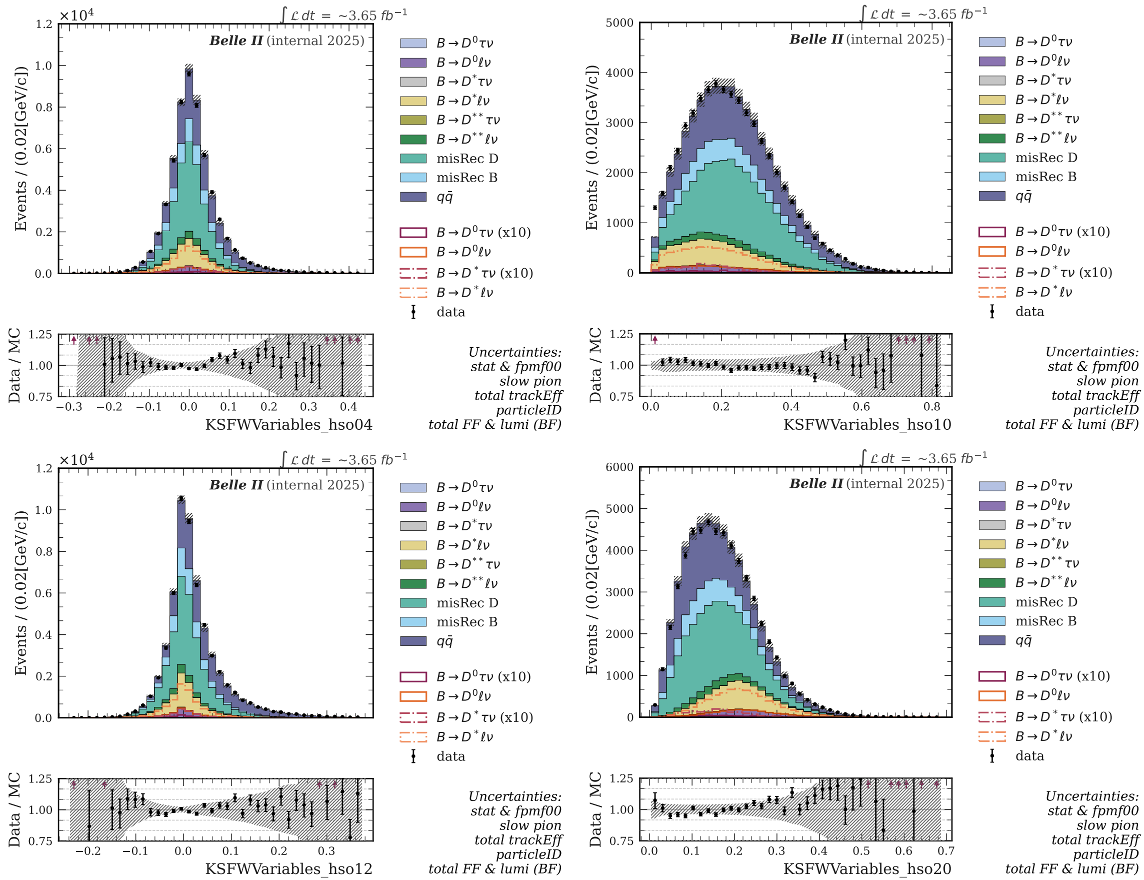


Figure D.3: kth-order super Fox-Wolfram moment using the cosine between ROE particles.



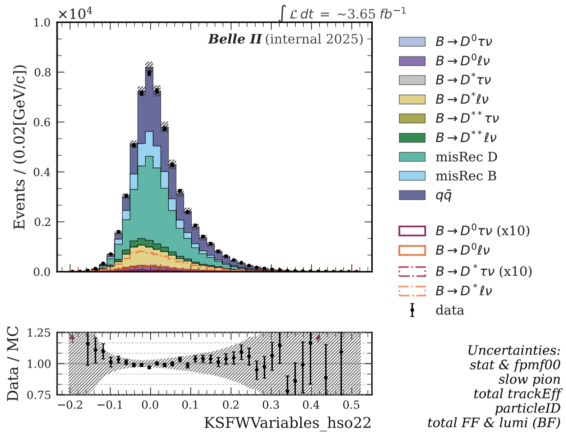


Figure D.4: kth-order super Fox-Wolfram moment using the cosine of the angle between a signal B daughter and an ROE particle.

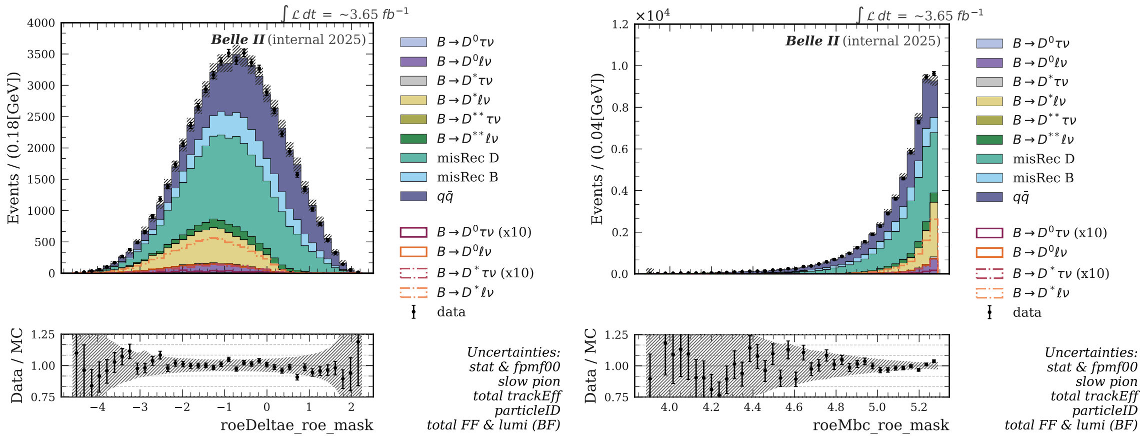


Figure D.5: ΔE associated with the difference between the summed ROE signatures and $E_{cms}/2$ (left). Beam constrain mass of the summed ROE signatures (right).

E BDT2 Input Features

After the application of BDT1, the samples are close enough to the signal region that care must be taken when viewing data. On-resonance modeling is checked using approximately 10% of the generic MC, in the BDT1 sideband. To ensure consistent modeling between the sideband and signal region, off-resonance modeling is checked in both regions and the respective track/cluster multiplicities are compared.

BDT2 input features, are selected to discriminate between signal D and D^* tauonic modes and each of the main backgrounds discussed in Section 4.

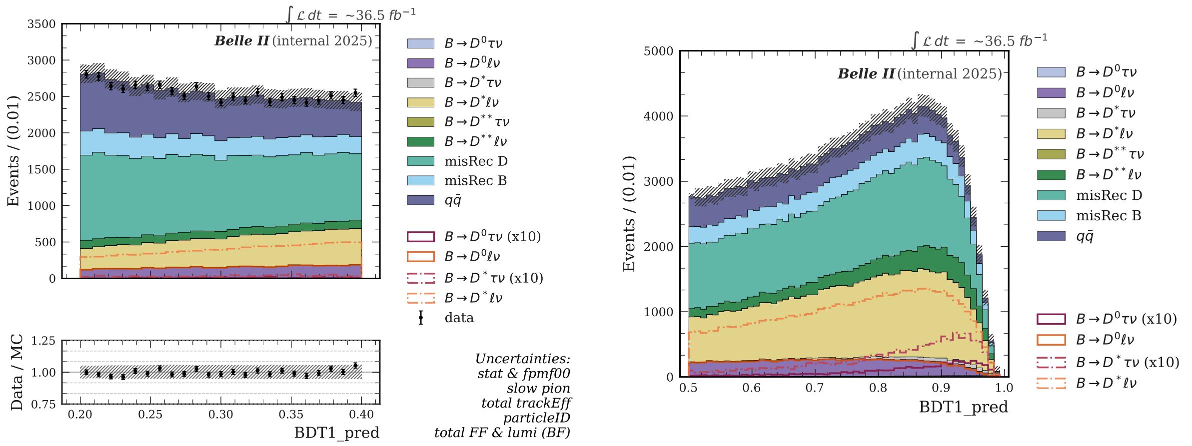


Figure E.1: BDT1 signal classifier output in the BDT1 sideband (left) and SR (right).

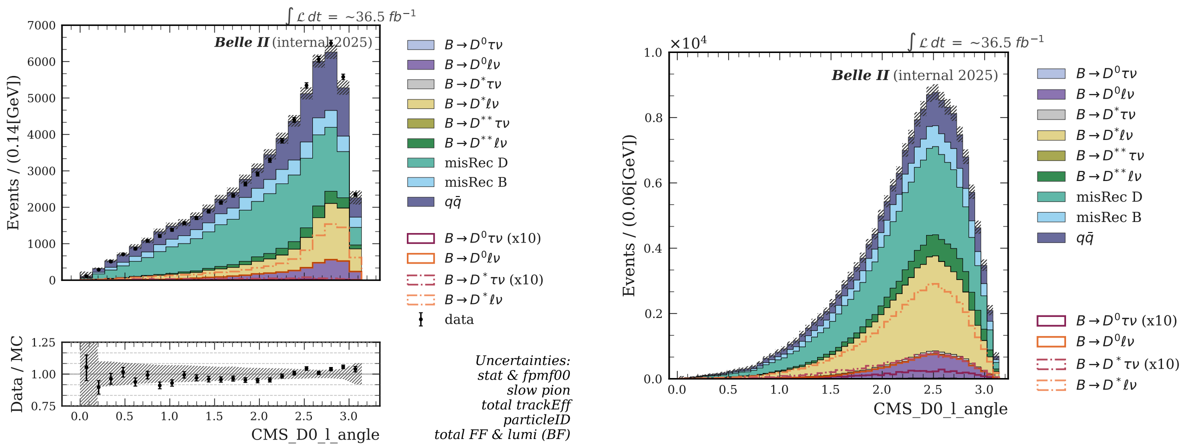


Figure E.2: Angle between the reconstructed lepton and D candidates, boosted into the center-of-mass frame, in the BDT1 sideband (left) and SR (right).

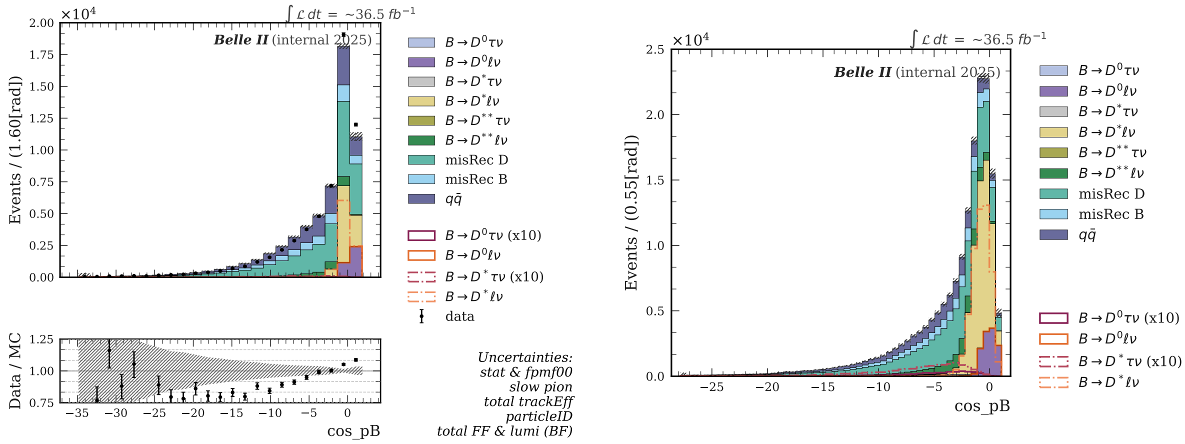


Figure E.3: Cosine of the angle between a nominal B and the reconstructed signal B in the BDT1 sideband (left) and SR (right). Significant mismodeling is present; however, detailed studies (Section 7) in current mismodeling are being conducted. So, due to the high level of discrimination power, the feature is kept at this time.

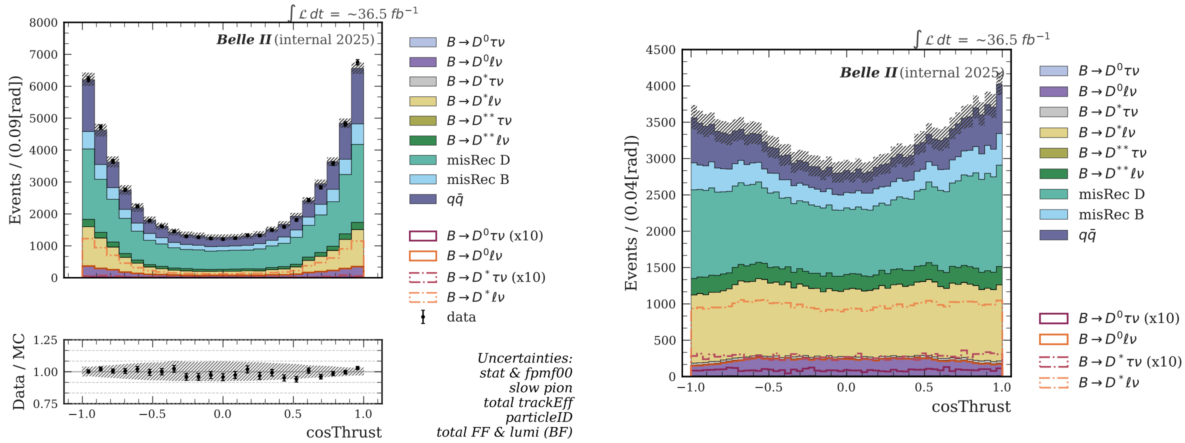


Figure E.4: Cosine of the reconstructed signal B to the thrust axis in the BDT1 sideband (left) and SR (right).

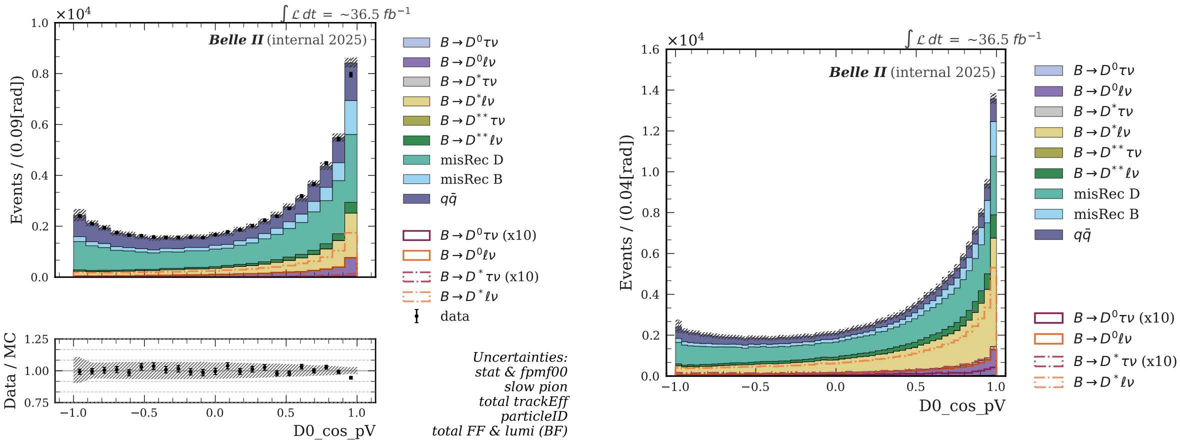


Figure E.5: Cosine of the angle between the reconstructed signal B momentum and the vertex vector (connecting the IP and fitted B vertex) in the BDT1 sideband (left) and SR (right).

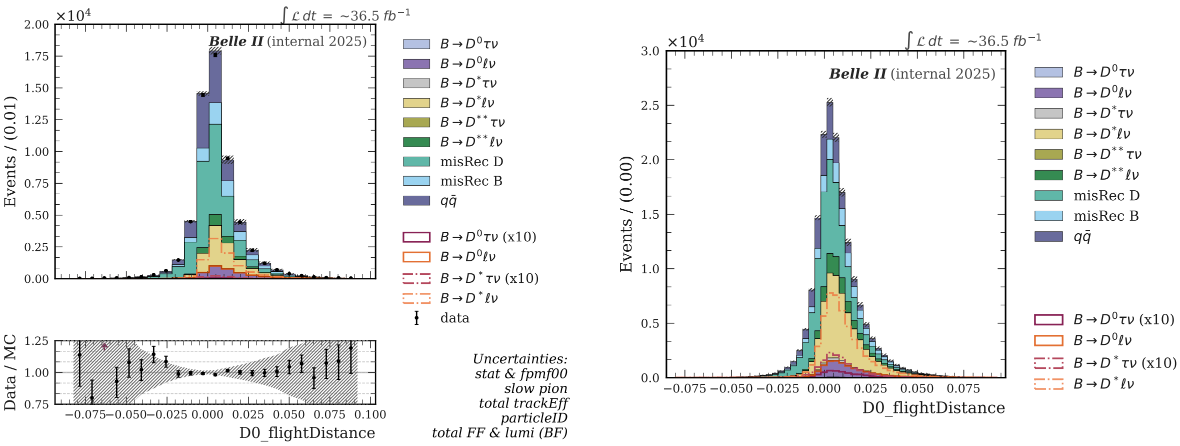


Figure E.6: Flight distance of the reconstructed D^0 candidate in the BDT1 sideband (left) and SR (right).

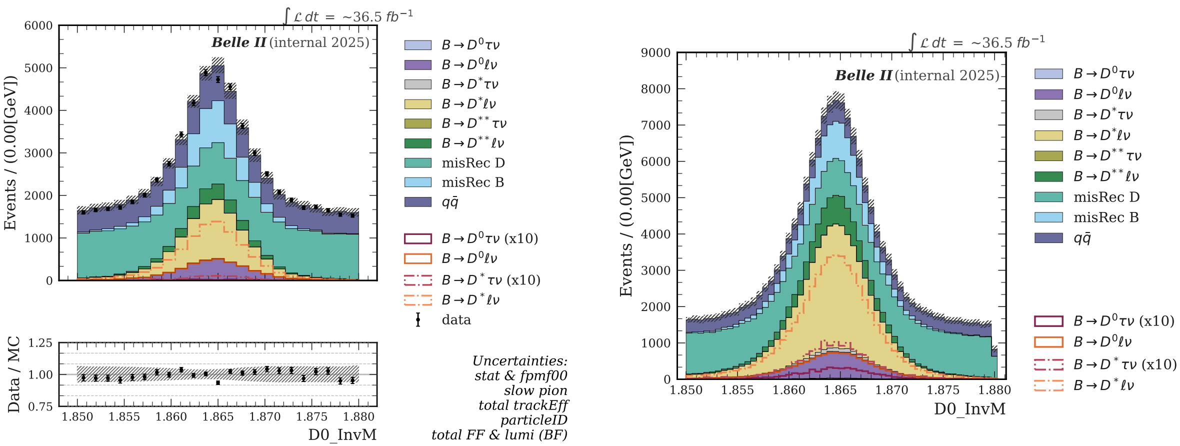


Figure E.7: Invariant mass of the reconstructed D^0 candidate in the BDT1 sideband (left) and SR (right).

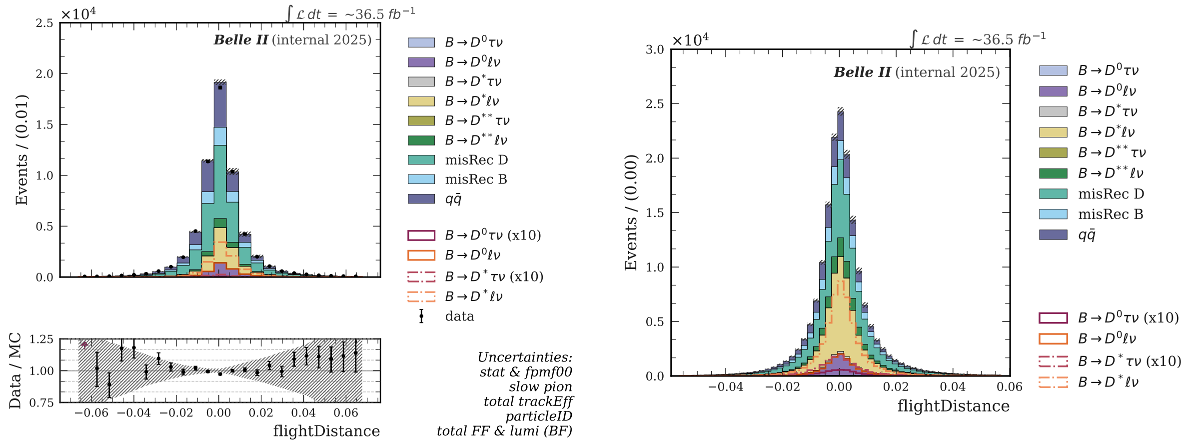


Figure E.8: Flight distance of the reconstructed signal B candiate in the BDT1 sideband (left) and SR (right).

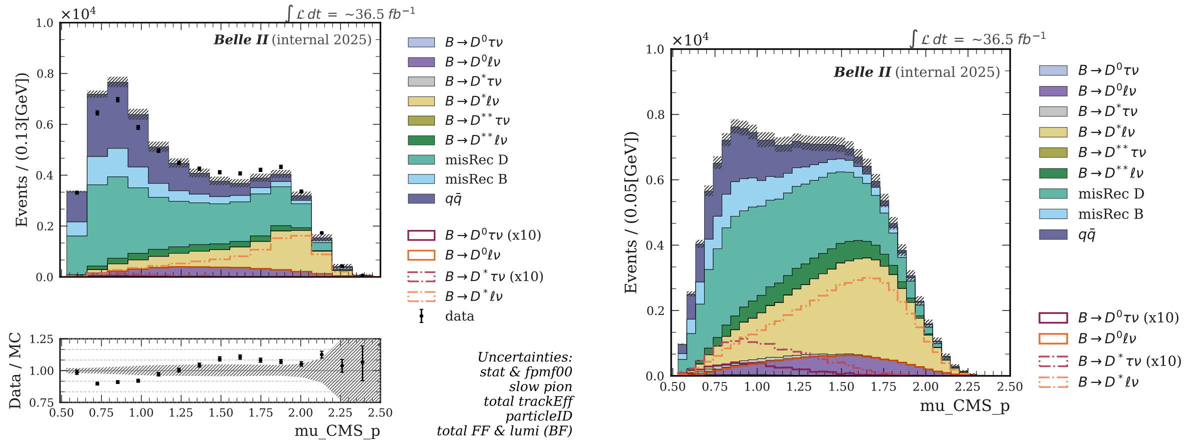


Figure E.9: Center-of-mass momentum of the muon candidate in the BDT1 sideband (left) and SR (right). Mismodeling is observed; however, studies are being conducted (Section 7), so the highly discriminating variable is kept for the current study.

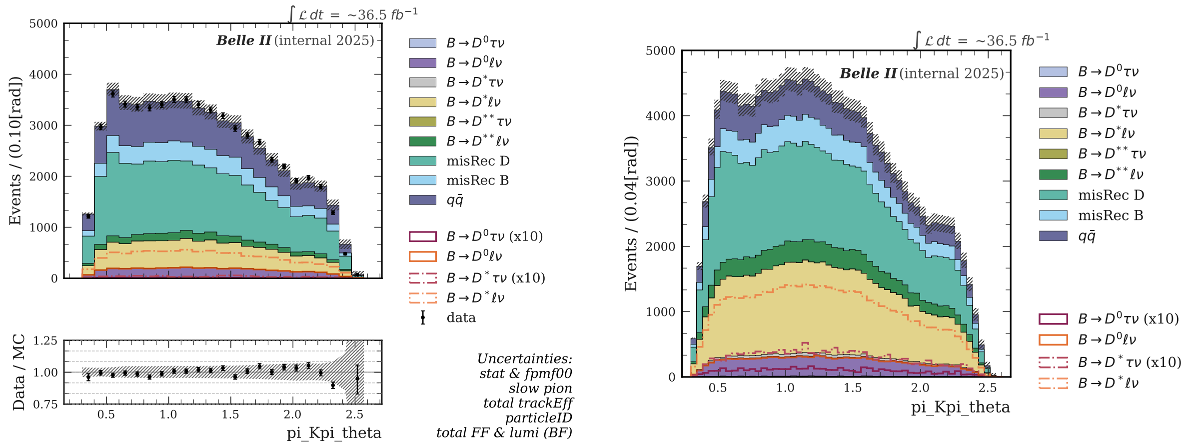


Figure E.10: Polar angle of the first pion daughter in the reconstructed D^0 decay in the BDT1 sideband (left) and SR (right).

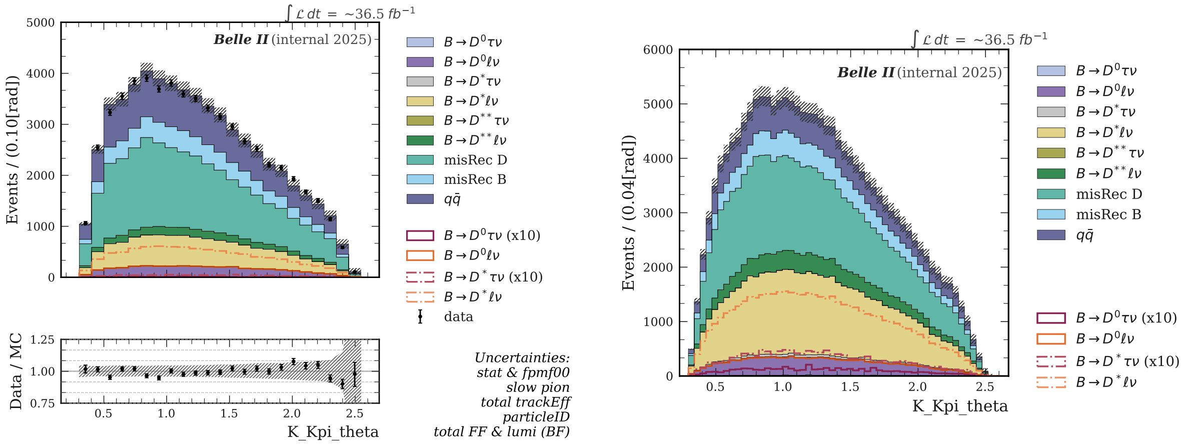


Figure E.11: Polar angle of the kaon daughter in the reconstructed D^0 decay in the BDT1 sideband (left) and SR (right).

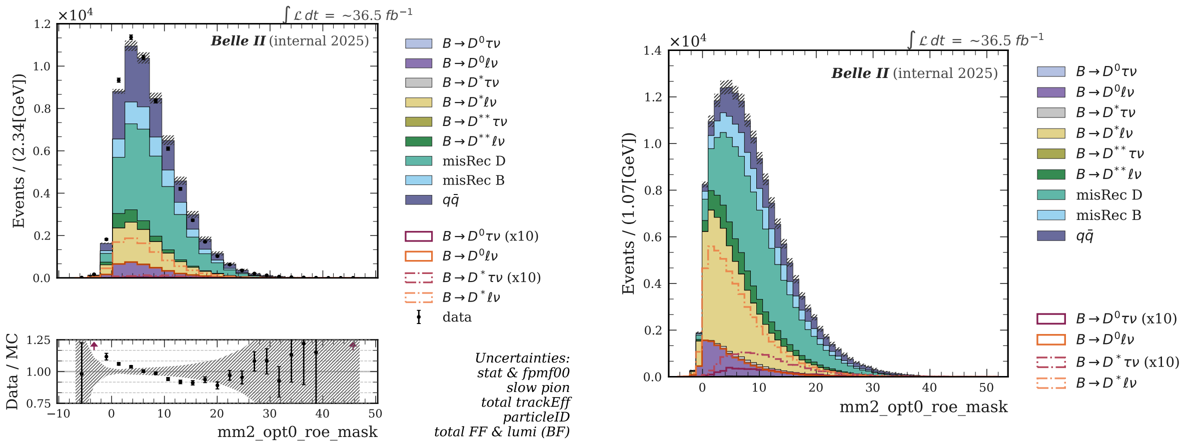


Figure E.12: Missing mass squared using the center-of-mass energy and momentum of charged particles and photons in the BDT1 sideband (left) and SR (right).

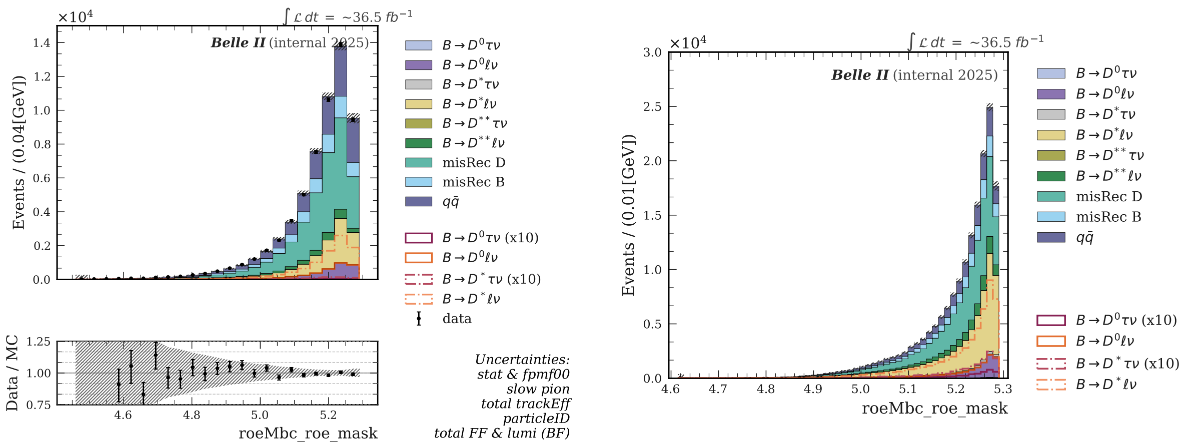


Figure E.13: Beam constrain mass associated with the summed ROE signatures in the BDT1 sideband (left) and SR (right).

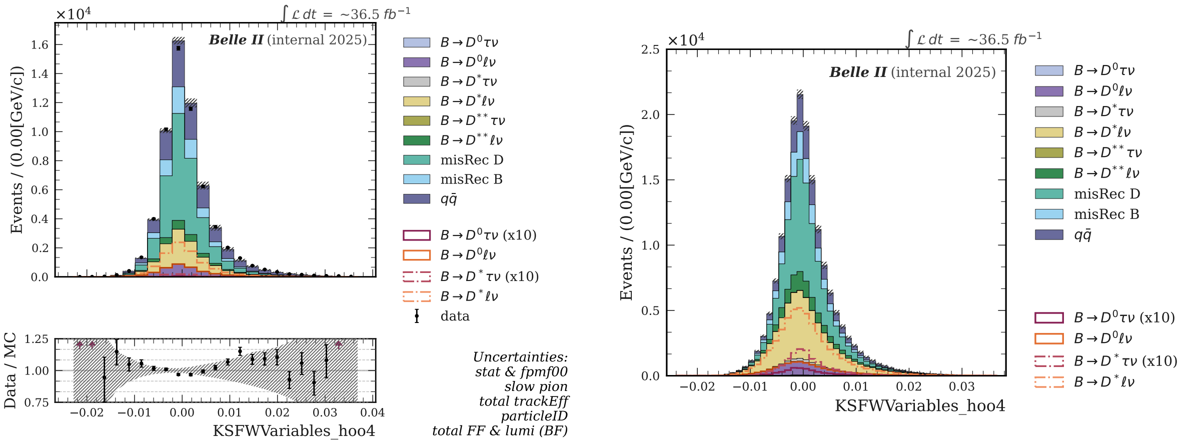
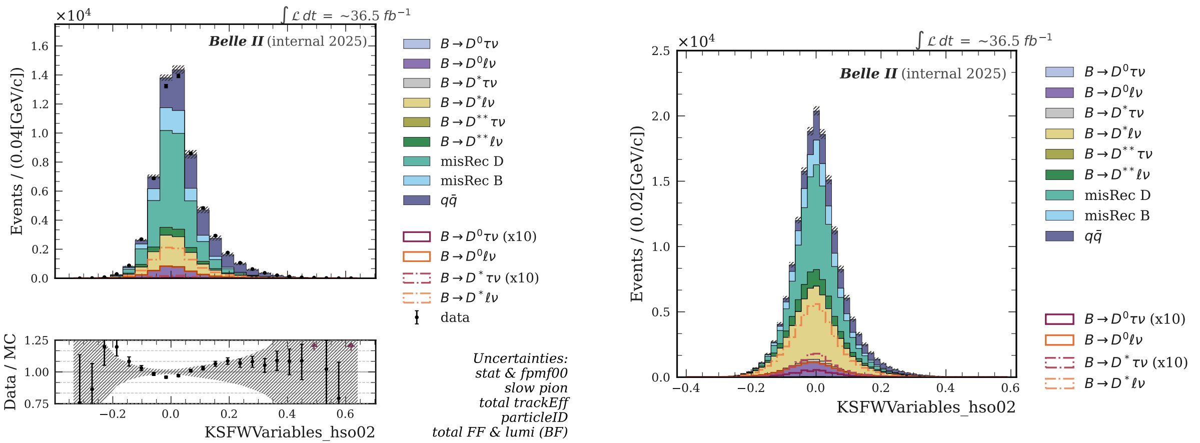
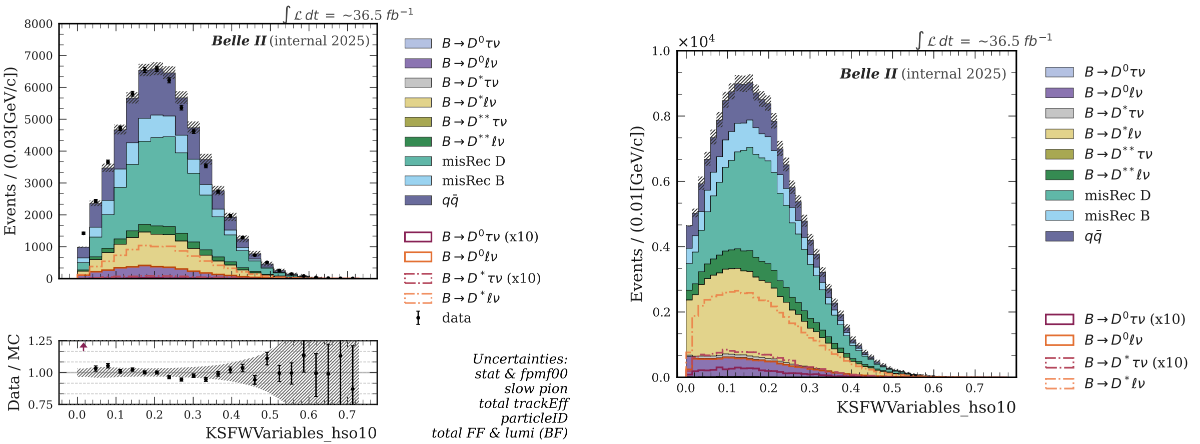


Figure E.14: 4th-order super Fox-Wolfram moment using the cosine of the angle between two ROE particles in the BDT1 sideband (left) and SR (right).





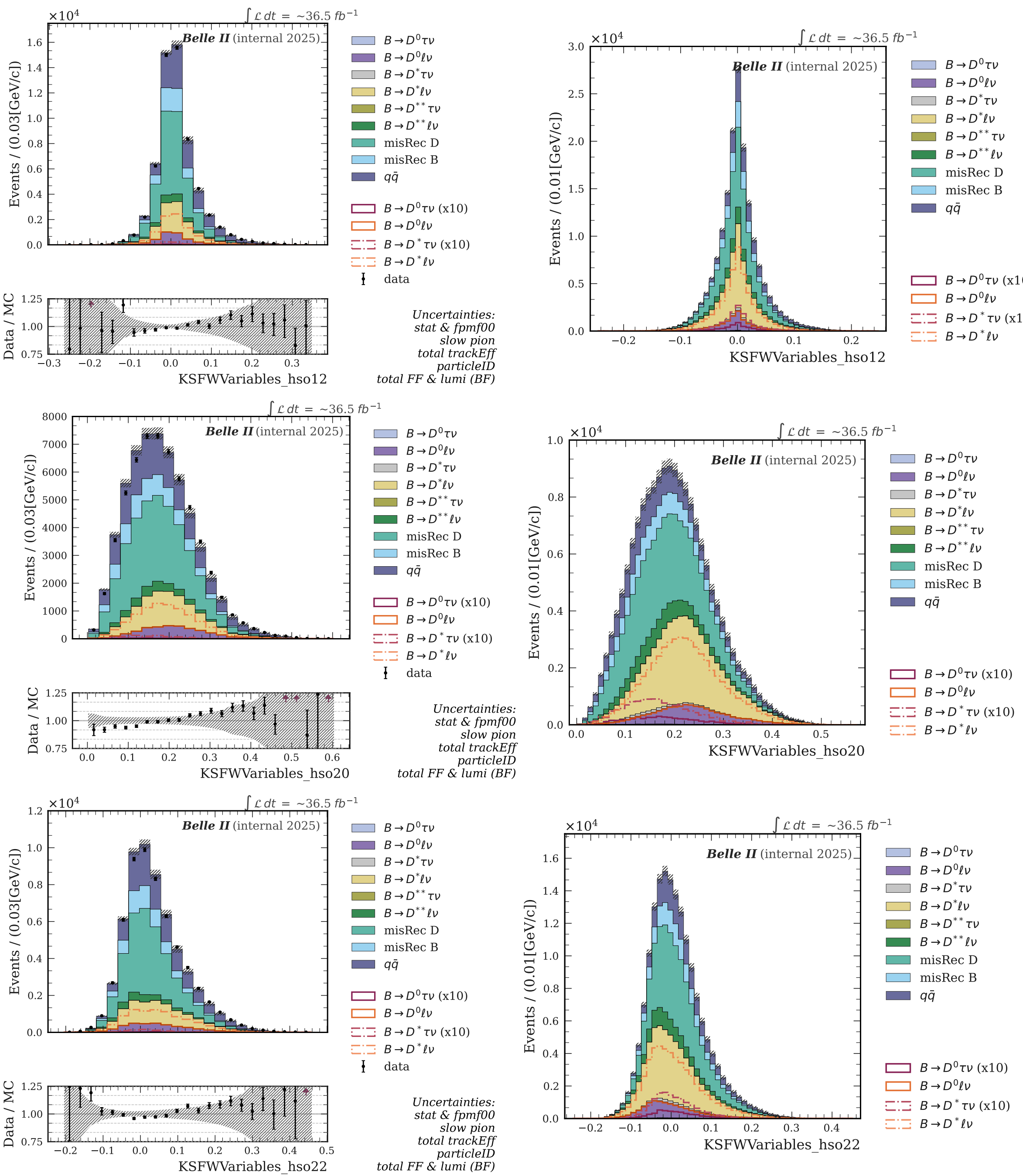


Figure E.15: kth-order super Fox-Wolfram moment using the cosine of the angle between a signal B daughter and an ROE particle in the BDT1 sideband (left) and SR (right). Significant mismodeling is present in hso20; however, studies are being conducted (Section 7), so the highly discriminating variable is kept for the current study.