Measurement of $R(D^*)$ with Inclusive B Meson Tagging at Belle II

Master's Thesis
Experimental Flavour Physics Group

submitted by

Stephanie Steinmetz

supervised by

Prof. Dr. Thomas Kuhr
Dr. Thomas Lück

Munich, 15 September 2023

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
FAKULTÄT FÜR PHYSIK

$R(D^*)$-Messung mit inklusivem B-Mesonen-Tagging bei Belle II

Masterarbeit

in der Arbeitsgruppe Experimentelle Flavourphysik

vorgelegt von

Stephanie Steinmetz

betreut von

Prof. Dr. Thomas Kuhr
Dr. Thomas Lück

München, den 15. September 2023

Contents

Chapter 1

Introduction and Motivation

The Standard Model (SM) of particle physics describes nature to an astounding precision [1], and a little over 10 years ago, it was completed by the discovery of the Higgs boson in July 2012 [2] at CERN, the European Organisation for Nuclear Research. However, some observed phenomena cannot be explained in the SM, such as the existence of dark matter deduced from the circular velocities of galaxies [3]. There is also the observed baryon asymmetry in the Universe, even though the SM does not allow for such a large imbalance between matter and antimatter [4]. The non-zero neutrino masses needed to explain neutrino flavour oscillations [5] are impossible to implement in the SM, and it should be noted that neutrino masses could only explain up to 1.6% of the dark matter content of the Universe [4]. There are also other shortcomings of the SM which do not pose an immediate inconsistency but do not have a satisfactory explanation either, such as the large number of free parameters in the SM as well as the strong CP problem. Furthermore, the SM is generally understood to be incompatible with the equally well established theory of General Relativity.

As illustrated above, there is a multitude of reasons to believe that the SM can only be the low-energy limit of a more fundamental theory yet to be uncovered. Experiments at the Large Hadron Collider (LHC) at CERN aim to probe new physics (NP) beyond the SM directly by achieving higher and higher centre-of-momentum system (CMS) energies. However, these analyses tend to suffer from low statistics and the busy environment containing debris from the colliding protons, even if a NP process were recorded. It is also possible to probe new physics *indirectly*, by the corrections they would inflict on known processes – this is the approach that B factories, such as Belle II, take.

At a B factory, an electron and a positron are collided at a CMS energy of 10.58 GeV, corresponding to the mass of the $\Upsilon(4S)$ resonance of two b quarks. The $\Upsilon(4S)$ is just heavy enough to decay into an entangled $B\bar{B}$ pair, and does so with a very large branching fraction. In the CMS frame, the B mesons are almost at rest, and their relatively long lifetime is exploited by using asymmetric beam energies, allowing to resolve the secondary decay vertices. B Factories offer a clean environment, thanks to the precisely known initial state, with record luminosities yielding unprecedented statistics.

To uncover hints on how the SM needs to be expanded, every deviation from its predictions is worth investigating. A quantity that has consistently shown an excess over the SM expectation is the measured ratio of branching fractions

$$R(D^*) = \frac{\mathcal{B}(B \to D^* \tau \nu)}{\mathcal{B}(B \to D^* \ell \nu)} \tag{1.1}$$

where ℓ denotes an electron or a muon and charge conjugated modes are implied here and throughout the entire thesis. Furthermore, $R(D^*)$ has the advantage of both a very clean

theoretical prediction and being a similarly clean experimental measurement, as most uncertainties such as hadronic form factors, the dependence on the quark-mixing matrix element $|V_{cb}|$ and reconstruction efficiencies cancel in the ratio [6], giving even more credibility to the discrepancy. The current deviation of the combined measurements from the SM prediction is 2.15σ. The analogously defined ratio $R(D)$ for semileptonic B decays to a ground state D meson shows a discrepancy of 1.98σ, resulting in a combined deviation of 3.2σ [7]. A plot showing a summary of measurements of $R(D)$ and $R(D^*)$ done so far can be found in Figure 1.1.
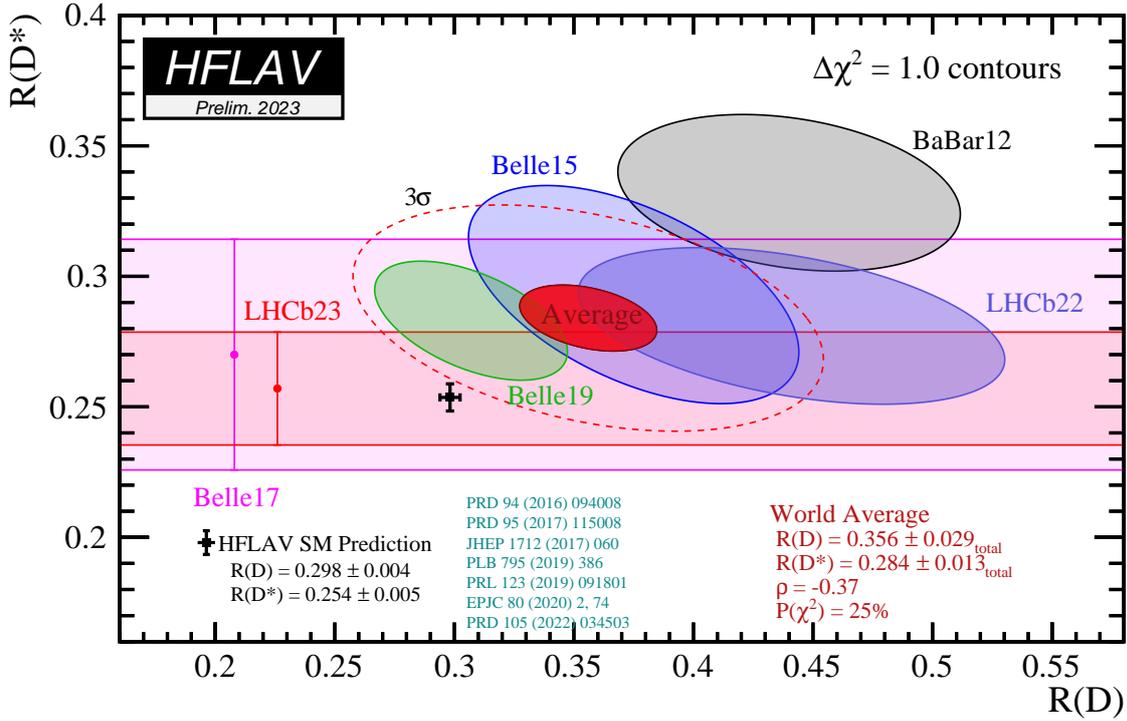

Figure 1.1: Summary of previous $R(D^{(*)})$ measurements as well as the SM prediction created by the Heavy Flavor Averaging Group, HFLAV [7].

As B Factories, BaBar and Belle have two B mesons available per event. While one is reconstructed in $B \to D^{(*)}\tau\nu$ or $B \to D^{(*)}\ell\nu$ to measure $R(D^{(*)})$, the choice on channels in which to reconstruct the second B meson (referred to as the tag B) is an open one. The BaBar analysis from 2012 reconstructed the second B hadronically [8], as did the Belle analyses from 2015 [9] and 2017 [6]. This has the advantage of avoiding additional missing particles in the event, giving access to the entire kinematics of the tag B decay. The Belle result from 2019 used a semileptonic tag instead [10]. The first measurement of $R(D^*)$ by Belle II has recently been performed with hadronic tagging [11].

In this analysis, the tag B is reconstructed *inclusively*, rather than in predefined decay modes, in order to avoid a decrease in statistics. The goal of this thesis is to determine the feasibility of extracting a competitive result for $R(D^*)$ from the currently available Belle II data set of $362(2)\,\text{fb}^{-1}$, and continues the work of Sofia Palacios Schweitzer for her master's thesis [12]. The analysis is done using the *Belle II Analysis Software Framework, basf2* [13].

As this is a blinded measurement, the analysis is developed using simulated events before doing the actual measurement on real data (the latter unfortunately lies outside the scope of this thesis), in order to avoid a bias in the method. However, a partial unblinding was done to check if distributions are modelled well within the simulation, affecting the choice of variables used in the analysis. The unblinded sideband data at low momentum transfer squared contains

mostly $B \to D^* \ell^+ \nu_\ell$ events. Though these decays are relevant to the measurement, they have been measured to a high precision in the past, and their branching fractions are well-established.

The thesis is structured as follows: first, Chapter 2 discusses the theory behind $R(D^*)$, followed by a quick introduction of the Belle II detector and the used data sets in Chapter 3. Events are reconstructed in $B^0 \to D^{*-}(\bar{D}^0 \pi^-_{\mathrm{slow}})\tau^+(\ell^+ \nu_\ell \bar{\nu}_\tau)\nu_\tau$, simultaneously picking up the decays in both the numerator and the denominator of $R(D^*)$ due to the neutrinos not being detected, with the tag B being reconstructed inclusively as mentioned above; this is discussed in detail in Chapter 4. Chapter 5 documents attempts at using a multiclass classifier to reject background, as well as the reason it was not used in the end. A description of the two-dimensional fit in the missing mass squared and the CMS lepton momentum used to extract event yields can be found in Chapter 6, along with the obtained results. Chapter 7 covers systematic uncertainties, and Chapter 8 serves as a conclusion.

Chapter 2

Theory

Lepton Flavour Universality (LFU) is the term given to the assumption that the electroweak force couples to all lepton flavours with the same strength, and is a fundamental postulate of the SM. Within the SM, differences only arise from the smaller phase space factor for the heavier τ lepton. An observation of LFU violation beyond this effect could only be explained by NP and is therefore the topic of many analyses. Precise measurements of semileptonic $B \to D^{(*)}\ell\nu_\ell$ decays where $\ell = e, \mu$ have been performed and were found to agree well with the SM. Semitauonic $B \to D^{(*)}\tau\nu_\tau$ decays on the other hand are expected to be sensitive to NP effects due to the large τ mass, and have consistently shown deviations from the SM expectations. Among these is the branching fraction ratio $R(D^*)$, defined in Eq. 1.1, which has the additional advantage of small theoretical uncertainties due to cancellations in the ratio. In particular, it is independent of the Cabibbo-Kobayashi-Maskawa (CKM) matrix element $|V_{cb}|$, allowing for several different methods to calculate the $R(D^*)$ theory expectation, even without experimental input using only lattice quantum chromodynamics (lattice-QCD), or through different combinations of lattice-QCD form factors and/or those from light cone sum rules (LCSR) with constraints from heavy quark effective theory (HQET) and fits to experimental data. Furthermore, many uncertainties from hadronic form factors as well as reconstruction efficiencies also cancel. [4, 14]

2.1 Signal and normalisation processes

In this thesis, only neutral B meson decays are considered. There are two relevant decay modes to determine $R(D^*)$: semitauonic $B^0 \to D^{*-}\tau^+\nu_\tau$ decays are referred to as signal events, and $B^0 \to D^{*-}\ell^+\nu_\ell$ with $\ell = e, \mu$ is called the normalisation mode. A full signal mode $\Upsilon(4S)$ decay is illustrated in Figure 2.1, and a quark-level Feynman diagram of a signal decay can be found in Figure 2.2.

Throughout most of this thesis, the Figure of Merit used to optimise the event selection (described in Chapter 4) is the significance of the signal mode event yield N_{sig},

$$Z(N_{\mathrm{sig}}) = \frac{N_{\mathrm{sig}}}{\sqrt{N_{\mathrm{sig}} + N_{\mathrm{bkg}}}} \tag{2.1}$$

where N_{bkg} is the yield of background events. The number of normalisation mode events, N_{norm}, is ignored in this optimisation, as they are abundantly present.
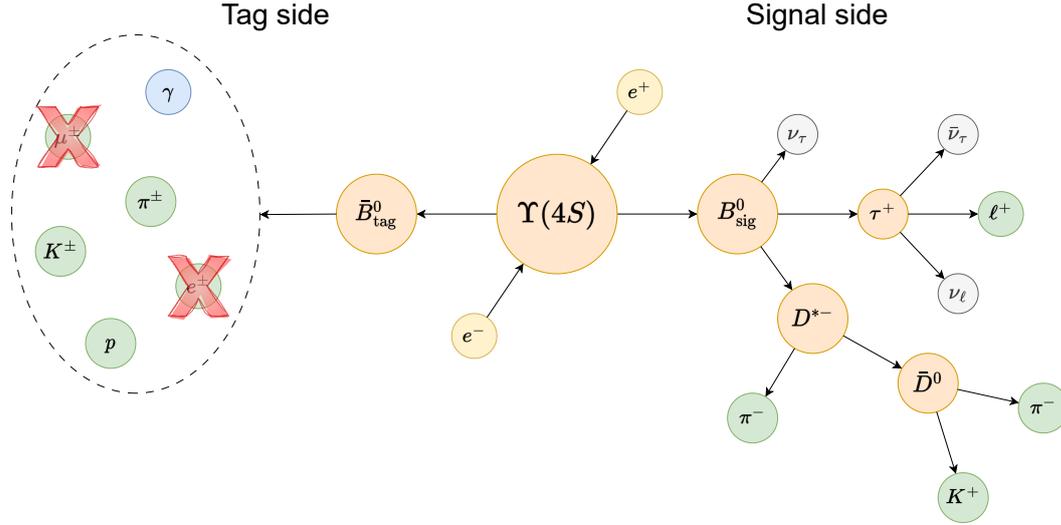
Figure 2.1: Exemplary signal $\Upsilon(4S)$ decay. The tag B is reconstructed inclusively, i.e. directly from all final state tracks and photon candidates, the different types of which are plotted here as an example. Final state electrons or muons on the tag side are vetoed to prevent additional neutrinos.
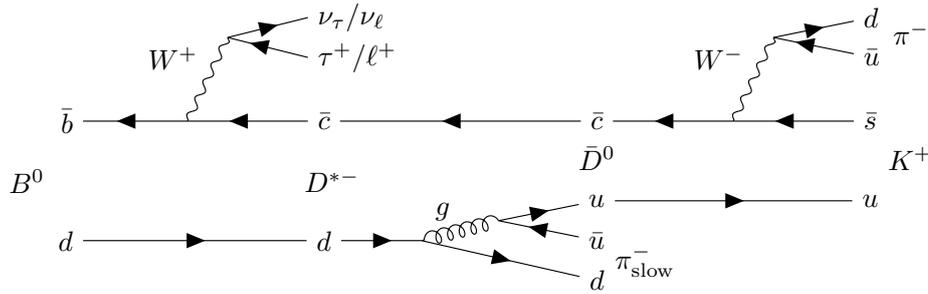
Figure 2.2: Quark-level Feynman diagram of $B^0 \to D^{*-}\tau^+\nu_\tau$ / $B^0 \to D^{*-}\ell^+\nu_\ell$, where $D^{*-} \to \bar{D}^0(K^+\pi^-)\pi^-_{\text{slow}}$.

2.2   Background processes

$B \to D^{**}\ell\nu_\ell$ decays to orbitally excited D mesons $(D_1, D_0^*, D_1', D_2^*)$, denoted D^{**}, are a problematic background to $R(D^*)$ analyses due to their kinematic similarity to signal events in case of a missed daughter, while not being well-understood themselves. As such, their yield is a free parameter in the fit (see Chapter 6) of this analysis, and the associated uncertainties are discussed in Section 7.4. Since only neutral B meson decays are considered as normalisation (or indeed signal), charged $B^+ \to \bar{D}^{*0}\ell^+\nu_\ell$ decays are given their own fit component as well, but made to scale with the normalisation mode yield, due to their peaking nature in the missing mass squared (Section 4.3.2) and being indistinguishable from normalisation to the fit. Continuum $e^+e^- \to q\bar{q}$ events with $q = u, d, s, c$ were found to be rarely selected, so no dedicated continuum suppression is employed. Out of the remaining background, no event types are particularly frequent, though the so-called gap mode (also discussed in Section 7.4) was found to cause large systematic uncertainties.

Chapter 3

Experimental Setup – Belle II

The Belle II detector is situated at the SuperKEKB collider complex in Tsukuba, Japan. At SuperKEKB, electrons and positrons are collided with asymmetric beam energies of $7\,\mathrm{GeV}$ and $4\,\mathrm{GeV}$ respectively, so that $\Upsilon(4S)$ mesons are resonantly produced. The $\Upsilon(4S)$ decays to a $B\bar{B}$ pair with a large branching fraction of at least 96% [4], hence the term B factory. The beams are collided inside Belle II, a general purpose detector covering almost the entire 4π solid angle. Starting from the innermost component closest to the beam Interaction Point (IP), it consists of a *pixel vertex detector* (PXD) and a *silicon vertex detector* (SVD) to precisely determine the position of decay vertices, followed by the *central drift chamber* (CDC). These three detector systems track the trajectories of charged particles passing through them. Outside the CDC there are the *time-of-propagation* counter (TOP) in the barrel region as well as an *aerogel ring imaging Cherenkov* detector (ARICH) in forward direction, which allow the identification of different particles, mainly in order to reliably distinguish between kaons and pions. Beyond the two particle identification subdetectors lies the *electromagnetic calorimeter* (ECL) measuring the energy and position of electromagnetic showers produced by charged particles and photons, where electrons and photons deposit their entire energy. The ECL is surrounded by a superconducting solenoid creating a longitudinal magnetic field of $1.5\,\mathrm{T}$ along the beam axis within the previously mentioned detector components, allowing to use the curvature of tracks to determine the momentum and charge of the corresponding charged particle. As the outermost part, a K_L^0 and *muon* detector (KLM) completes Belle II. [11, 15, 16]

A detailed description can be found in the *Belle II Technical Design Report* [16], and an annotated schematic of Belle II is shown in Figure 3.1.

In the Belle II coordinate system, the z-axis follows the direction of the electron beam, with the origin at the IP. The y-axis points vertically upwards, and the x-axis of the right-handed coordinate system points horizontally away from the centre of the SuperKEKB storage rings. Furthermore, one defines the azimuthal angle ϕ around the z-axis such that $\phi = 0$ for the x-axis, as well as the polar/zenith angle θ with respect to the z-axis. [17]

3.1   Data sets

The analysis is developed using (and performed on) events simulated for the Belle II detector using Monte Carlo (MC) techniques. Both *run-dependent* and *run-independent* MC has been used at different stages throughout the analysis, with the former samples generated specifically using the same conditions as were present during the actual corresponding data taking periods, while the run-independent samples aim to use a good middle ground that should be

approximately valid for all data taking, including that of the future. As a further difference, run-dependent MC includes real beam background data, while the beam background in the run-independent samples is simulated. Mostly, so-called *generic* MC is used, where different decays occur as often as would be expected in real data. On some occasions, to improve the statistics of relatively rare processes, *signal MC* samples are considered, in which only certain decays of interest are simulated.

The used MC samples include $e^+e^- \to B^0\bar{B}^0$, $e^+e^- \to B^+B^-$, and $e^+e^- \to q\bar{q}$ events where $q = u, d, s, c$. The $B\bar{B}$ events are generated with EVTGEN [18], using a HQET model for the $B \to D^*l^+\nu_l$ ($l = e, \mu, \tau$) decays, namely EVTGEN's HQET3. PYTHIA [19] was used for the $q\bar{q}$ continuum events, including their subsequent fragmentation. Final state radiation (FSR) is modelled using PHOTOS [20], and the detector response is simulated in GEANT4 [21].

In total, $1444\,\mathrm{fb}^{-1}$ of generic run-dependent MC was used, which mimic the conditions during the Exp. 7–26 data taking.

In the signal MC used, one B meson always decays as $B^0 \to D^{*-}\tau^+\nu_\tau$ or $B^0 \to D^-\tau^+\nu_\tau$ with the first decay occurring about twice as often as the second, while the other B meson decays generically. They are generated using EVTGEN with the HQET3 decay model for $B^0 \to D^{(*)-}\tau^+\nu_\tau$, use PHOTOS for FSR, and simulate the detector response in GEANT4. The run-dependent signal MC contains $1\,981\,012$ such events and again mimics the conditions during Exp. 7–26.

The real data used in this analysis is the LS1 data set recorded before the first long shutdown, comprising Exp. 7–26. The integrated luminosity recorded at the $\Upsilon(4S)$ energy is $362(2)\,\mathrm{fb}^{-1}$.

Several corrections need to be applied to the MC samples to adjust for small modelling errors. Among these are corrections to the particle identification efficiencies and fake rates (discussed in Section 7.2), the efficiency with which the slow pion is detected (Section 7.5), the $B \to D^{**}\ell^+\nu_\ell$ and τ branching fractions (Sections 7.4 and 7.6), and tracking efficiencies (Section 7.7).

Uniquely, Belle II also applies a correction to data, slightly rescaling track momenta (see Section 4.4), to account for an observed shift in mass resonances attributed to the calibration of the magnetic field.

Before switching to the mainly used run-dependent MC samples, generic run-independent MC was also used, containing $3.000\,71\,\mathrm{ab}^{-1}$ of $B^0\bar{B}^0$ events, $3\,\mathrm{ab}^{-1}$ of B^+B^- and $1\,\mathrm{ab}^{-1}$ of $q\bar{q}$.

KL and muon detector
Resistive Plate Counter (barrel outer layers)
Scintillator + WLSF + MPPC
(end-caps , inner 2 barrel layers)

EM Calorimeter
CsI(Tl), waveform sampling electronics

electrons (7 GeV)

Particle Identification
Time-of-Propagation counter (barrel)
Prox. focusing Aerogel RICH (forward)

Vertex Detector
2 layers Si Pixels (DEPFET) +
4 layers Si double sided strip DSSD

positrons (4 GeV)

Central Drift Chamber
Smaller cell size, long lever arm
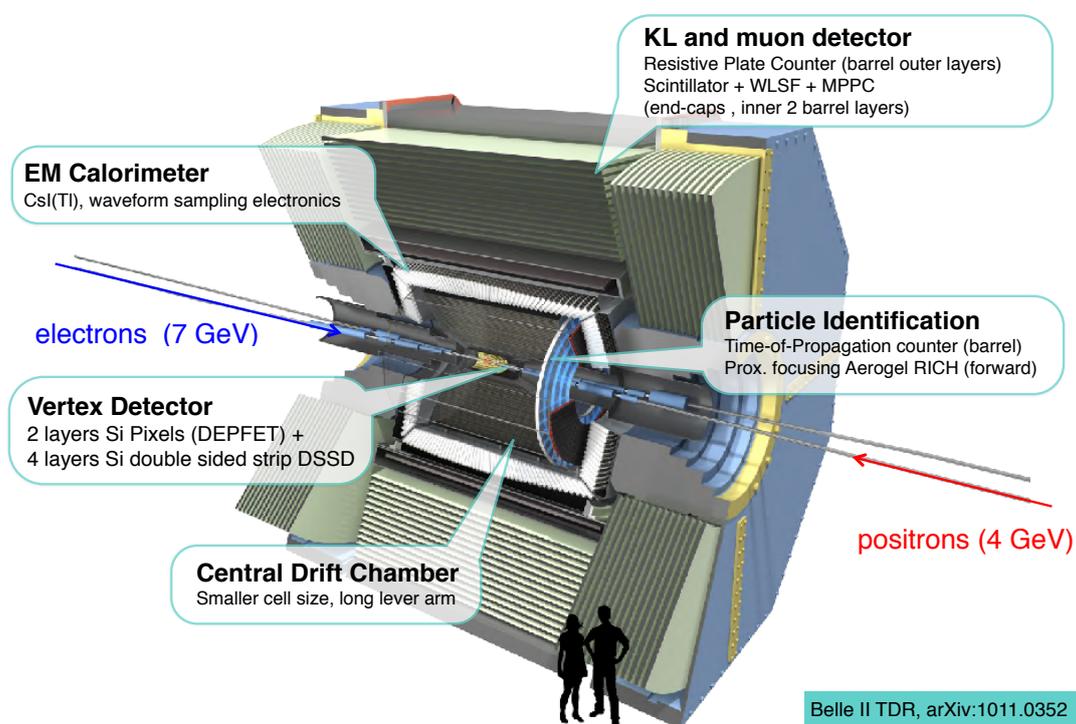
Belle II TDR, arXiv:1011.0352

Figure 3.1: Annotated schematic of the Belle II detector [16].

Chapter 4

Reconstruction and Selection

4.1 Signal side selection

The signal B is reconstructed as $B^0_{\text{sig}} \to D^{*-}\tau^+(\ell^+\nu_\ell\bar{\nu}_\tau)\nu_\tau$ with $\ell = e, \mu$. The neutrinos go undetected, so technically the reconstruction looks like $B^0_{\text{sig}} \to D^{*-}\ell^+$. As signal and normalisation mode events share the same final state, the normalisation mode $B^0_{\text{sig}} \to D^{*-}\ell^+\nu_\ell$ is implicitly included in the reconstruction. Properties such as the missing mass can be used to distinguish between the two.

All charged final state daughters must fulfil the tracking criteria $dr < 0.5\,\text{cm}$ and $|dz| < 2\,\text{cm}$, where dr and dz refer to the distance between the point-of-closest-approach (POCA) and the measured IP in transverse and z-direction, respectively. Furthermore, the polar angle θ has to lie within $17° < \theta < 150°$, corresponding to the CDC's angular acceptance.

4.1.1 Formal τ "reconstruction"

A τ is formally reconstructed from a light lepton and invisible neutrinos. This "reconstruction" therefore simply takes the final state light lepton and calls it a τ. As a result, signal and normalisation mode events are covered at the same time, depending on whether the light lepton actually came from an intermediate τ or not. The light leptons must fulfil the tracking criteria mentioned above and their CMS momentum p_ℓ^{cms} has to lie above $0.4\,\text{GeV/c}$. The electrons are brems-corrected, meaning that Bremsstrahlung photons are recovered by adding low-energetic photons found in a cone around the electron, the opening angle of which depends on the electron's transverse momentum as detailed in Table 4.1.

Electron p_{T} [GeV/c]	E_γ [GeV]	Angle threshold [rad]
≤ 0.6	< 0.09	0.1368
$\in (0.6, 1.0]$	< 0.90	0.0737
> 1.0	< 1.20	0.0632

Table 4.1: The criteria used to select Bremsstrahlung photons, depending on the e candidate's transverse momentum. Adapted from internal working group recommendations [22].

In addition, cuts on so-called particle ID (PID) variables are applied. These variables are the output of a classifier reflecting the probability of such a particle truly being of the assigned type, which is obtained by combining information from different subdetectors. For electrons, the recommended electron ID is based on a Boosted Decision Tree (BDT) due to reduced fake rates (see Figure 4.1), and a cut of `electronID` > 0.9 is applied.

For muons, the chosen muon ID is likelihood-based, defined as $\texttt{muonID} = \mathcal{L}_\mu / \sum_P \mathcal{L}_P$ where $P = e, \mu, \pi, K, p, d$ are the considered particle hypotheses and $\mathcal{L}_P$ is the likelihood for type P, obtained by combining individual likelihoods from different subdetectors [23]. This muon ID is recommended due to smaller systematic uncertainties on data-MC corrections than for the BDT approach, and the same threshold of $\texttt{muonID} > 0.9$ is employed. In both cases, the *global* lepton IDs are used, which discriminate between electron (muon) and all other charged particle hypotheses. Neither uses information from the SVD due to problems in the simulation.
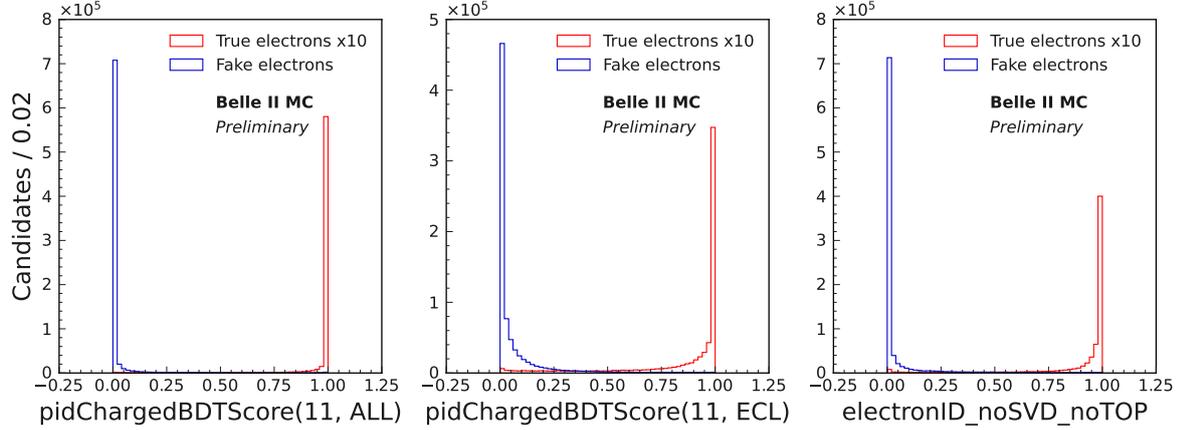

Figure 4.1: Distributions of different electron IDs for true and fake electron candidates. `pidChargedBDTScore(11, ALL)` and `pidChargedBDTScore(11, ECL)` are global BDT-based electron IDs using information from all particle identification detectors (excluding the SVD) or only the ECL respectively. They were compared to `electronID_noSVD_noTOP` which is likelihood-based and excludes information from the SVD and TOP subdetectors due to issues in the MC simulation. Plotted are electron candidates from $182\,500$ run-independent neutral $B\bar{B}$ events. In this analysis, `pidChargedBDTScore(11, ALL)` is used, referred to in the text as `electronID`. Note that the true electron component is scaled up by a factor of 10.

4.1.2 π^0 Reconstruction

The π^0 is reconstructed as $\pi^0 \to \gamma\gamma$ with the pion's invariant mass between $0.115\,\text{GeV/c}^2$ and $0.155\,\text{GeV/c}^2$. This cut was chosen by fitting a Gaussian to a sample of true π^0 candidates as shown in Figure 4.2 and selecting the range corresponding to 2σ around the peak, where σ is the standard deviation of the fitted Gaussian. The full distribution can be seen in Figure 4.3.
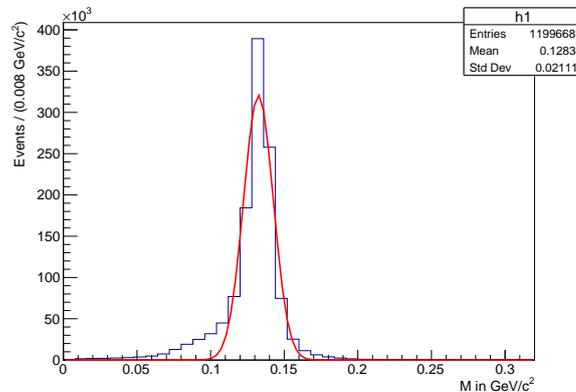

Figure 4.2: Peak in the π^0 invariant mass and fitted Gaussian. Plotted are true π^0 candidates from $378\,000$ neutral and charged $B\bar{B}$ events ($189\,000$ each).
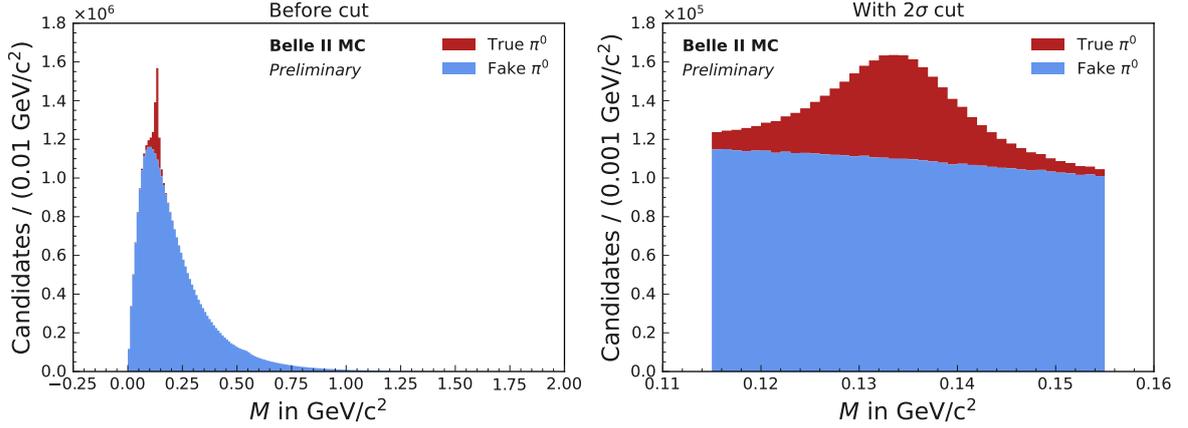
Figure 4.3: Stacked distribution of the π^0 invariant mass before and after the 2σ cut. Plotted are the $32\,560\,308$ π^0 candidates from $378\,000$ neutral and charged $B\bar{B}$ events ($189\,000$ each).

The selection of the daughter photons was optimised using the output of two MVA classifier variables: `beamBackgroundSuppression` and `hadronicSplitOffSuppression`, which distinguish real photons coming from the $\Upsilon(4S)$ decay from beam background and fake photons, respectively. The training method is explained in [24], and the two variables are plotted in Figure 4.4. They were varied independently in steps of 0.05 and the optimal cut combination was determined by calculating the significance of the true neutral pion yield,

$$Z(N_{\text{true}}) = \frac{N_{\text{true}}}{\sqrt{N_{\text{true}} + N_{\text{fake}}}} \tag{4.1}$$

where N_{true} (N_{fake}) is the number of true (fake) neutral pions. The best combination is

$$\texttt{beamBackgroundSuppression} > 0.35 \text{ and } \texttt{hadronicSplitOffSuppression} > 0.35 \tag{4.2}$$

which yields $778\,920$ true neutral pions alongside $2\,174\,193$ fake π^0 candidates out of an original sample of $1\,057\,643$ true and $14\,040\,852$ fake neutral pions from $378\,000$ neutral and charged $B\bar{B}$ events ($189\,000$ each) with the 2σ cut applied to the π^0 candidates.
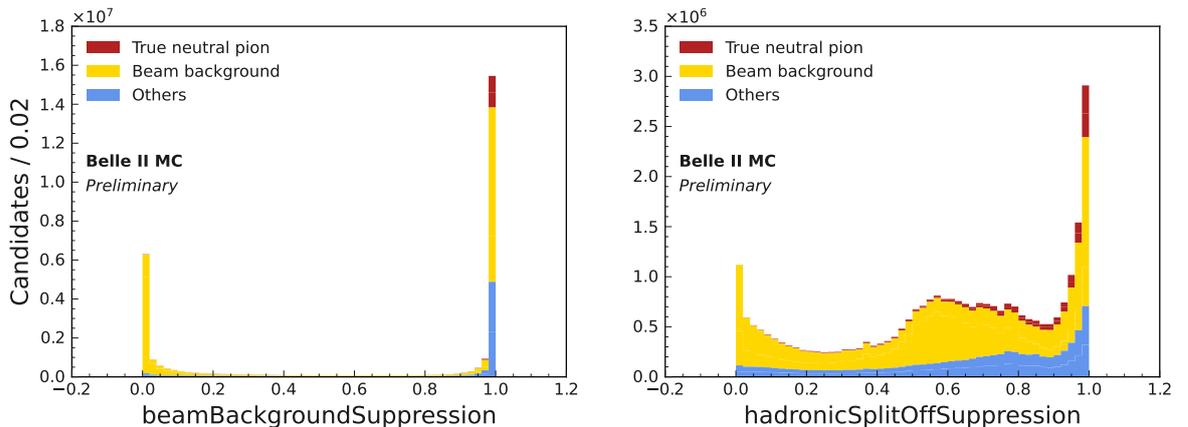
Figure 4.4: Stacked distributions of the photon MVA variables `beamBackgroundSuppression` and `hadronicSplitOffSuppression` for both daughters of the $15\,098\,495$ π^0 candidates from $378\,000$ neutral and charged $B\bar{B}$ events with the 2σ cut applied, separated into true neutral pions, beam background (particles not coming from the $\Upsilon(4S)$ decay that therefore have no MC equivalent) and other π^0 candidates reconstructed from at least one fake photon.

4.1.3 K_S^0 reconstruction

For the K_S^0 selection, both the standard K_S^0 and the `goodBelleKshort` [25] particle lists were considered as a basis. They were compared by reconstructing the D^* as usual (described in Section 4.1.5, though at this stage it was also reconstructed in $D^{*+} \rightarrow D^+\pi^0$ and not just $D^{*+} \rightarrow D^0\pi^+$, which was later confirmed to not affect the results and only improve the statistics) but not applying any further cuts or best candidate selection. A Gaussian was fitted to the invariant mass distributions of the true K_S^0 mesons of each list (see Figure 4.5) and the corresponding 2σ ranges of the peaks were determined. The full distributions as well as the determined invariant mass ranges can be seen in Figure 4.6.
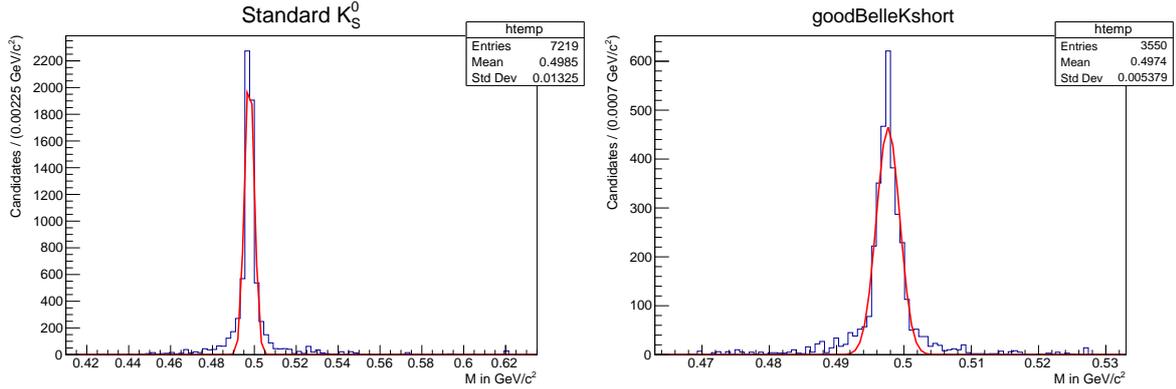
Figure 4.5: Fit to the peaks in the K_S^0 invariant mass for the standard list (left) and `goodBelleKshort` (right). Plotted are correctly reconstructed, true K_S^0 candidates from 40 000 neutral and charged $B\bar{B}$ events (i.e. 20 000 each) that were reconstructed up to the D^*.

The resulting 2σ ranges in the K_S^0 invariant mass are 0.493 to 0.502 GeV/c^2 for the standard list and the slightly narrower 0.494 to 0.501 GeV/c^2 for `goodBelleKshort`.
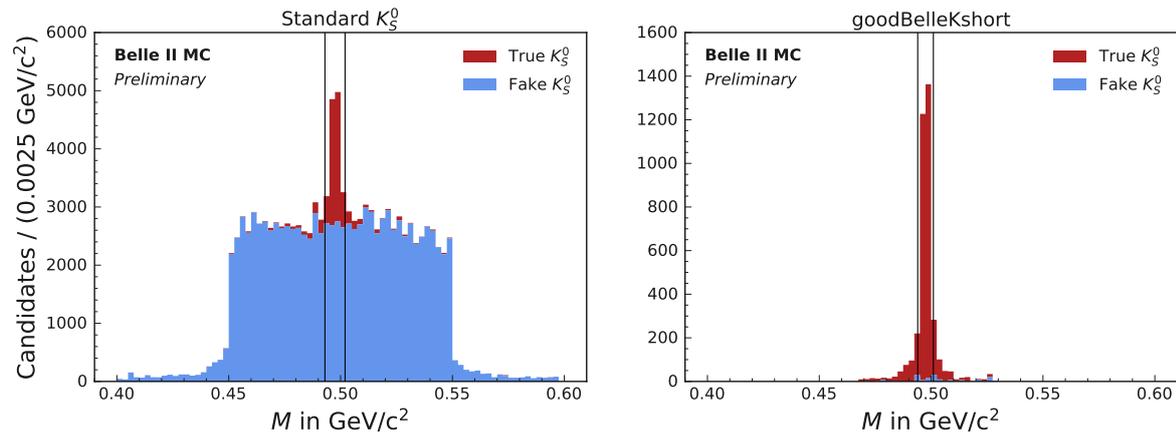
Figure 4.6: Distributions of the invariant mass of K_S^0 candidates for the standard vs. `goodBelleKshort` particle lists. The true distributions are stacked on top of the fake ones, and the vertical black lines indicate the 2σ ranges determined by fitting a Gaussian to the true K_S^0 peaks. Plotted are candidates from 40 000 neutral and charged $B\bar{B}$ events that were reconstructed up to the D^*.

Next, the best D^* candidate selection was applied (see Section 4.3) after restricting the K_S^0 candidates to their 2σ region. The distributions of surviving K_S^0 candidates are shown in Figure 4.7.
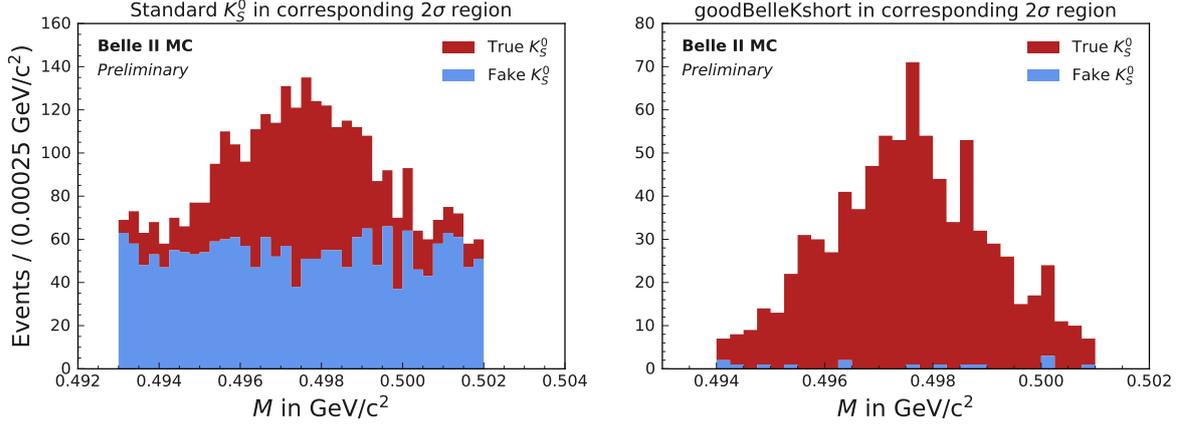
Figure 4.7: Comparison of the two K_S^0 lists after selecting the best D^* candidate according to the usual best candidate selection (see Section 4.3). Each list is restricted to its 2σ invariant mass range. The stacked plots contain K_S^0 candidates from $40\,000$ neutral and charged $B\bar{B}$ events that were reconstructed up to the D^*.

As one can easily see, `goodBelleKshort` yields much better results. This was backed up by calculating the significance of the true kaon yield (K_S^0 or charged, compare Table 4.3) over events where the kaon candidate is misreconstructed, analogously to Eq. 4.1. As such, `goodBelleKshort` is chosen as the basis, and the corresponding 2σ cut of

$$0.494\,\mathrm{GeV/c^2} < M < 0.501\,\mathrm{GeV/c^2} \tag{4.3}$$

is applied to the K_S^0 candidate's invariant mass.

The `goodBelleKshort` selection itself [25] works as follows:

- First, reconstruct $K_S^0 \to \pi^+\pi^-$ with $0.3\,\mathrm{GeV/c^2} < M < 0.7\,\mathrm{GeV/c^2}$.

- Next, perform a vertex fit (`kFit`) on K_S^0 and discard candidates with a failed fit.

- Then limit the mass range further to $0.468\,\mathrm{GeV/c^2} < M < 0.528\,\mathrm{GeV/c^2}$ and apply a momentum-dependent selection as detailed in Table 4.2.

p [GeV/c]	$\min(dr_{\pi^+}, dr_{\pi^-})$ [cm]	$d\phi$ [rad]	z-dist [cm]	fl [cm]
< 0.5	> 0.05	< 0.3	< 0.8	$-$
$\in (0.5, 1.5)$	> 0.03	< 0.1	< 1.8	> 0.08
> 1.5	> 0.02	< 0.03	< 2.4	> 0.22

Table 4.2: Momentum-dependent `goodBelleKshort` criteria [25].

The variables used for the `goodBelleKshort` selection are

- $\min(dr_{\pi^+}, dr_{\pi^-})$, the smaller one of the two daughters' dr values,

- $d\phi$, the azimuthal angle between the momentum vector of a K_S^0 candidate and the vector pointing from the IP to the reconstructed decay vertex,

- z-dist, the distance in z-direction between the two daughter tracks at their closest point, and

- fl, the K_S^0 candidate's flight length travelled in the x-y plane.

4.1.4 D^0 reconstruction

In addition to the tracking criteria, the D meson's final state daughters are required to pass the following selection using their respective (global, likelihood-based) particle IDs:

- $\pi^{\pm}$: `pionID` > 0.1

- $K^{\pm}$: `kaonID` > 0.5

From this the neutral D meson is reconstructed in the following decay modes:

Decay mode	Branching fraction [%]	Invariant mass range [GeV/c²]
$D^0 \to K^- \pi^+$	3.947 ± 0.030	$1.86 - 1.87$
$D^0 \to K_S^0 \pi^0$	1.240 ± 0.022	$1.83 - 1.90$
$D^0 \to K^- \pi^+ \pi^0$	14.4 ± 0.5	$1.84 - 1.89$
$D^0 \to K^- \pi^+ \pi^- \pi^+$	8.22 ± 0.14	$1.86 - 1.87$
$D^0 \to K_S^0 \pi^+ \pi^-$	2.80 ± 0.18	$1.86 - 1.87$

Table 4.3: Reconstructed D meson decay modes with cuts on invariant mass corresponding to a 2σ range around the peak. The respective branching fractions are taken from [4].

The invariant mass ranges correspond once again to a 2σ region around the nominal mass of $1.864\,84\,\text{GeV/c}^2$ [4]; the plots can be found in Figure 4.9. A vertex fit (`treeFit`) is performed, where the masses of the π^0 daughters are constrained, and candidates with a failed fit are discarded.

Also considered was the decay mode $D^0 \to K_S^0 \pi^+ \pi^- \pi^0$ with a 2σ invariant mass range of 1.84 to $1.89\,\text{GeV/c}^2$. The impact of excluding different decay modes and combinations thereof on the Figure of Merit (Eq. 2.1) was studied using $1444\,\text{fb}^{-1}$ of generic run-dependent MC[1]. The best combination was found to be the case where $D^0 \to K_S^0 \pi^+ \pi^- \pi^0$ is excluded while all other decay channels, listed in Table 4.3, are kept. A comparison between the D decay channels used for the reconstruction (including $D^0 \to K_S^0 \pi^+ \pi^- \pi^0$) and the true decay modes can be found in Figure 4.8.
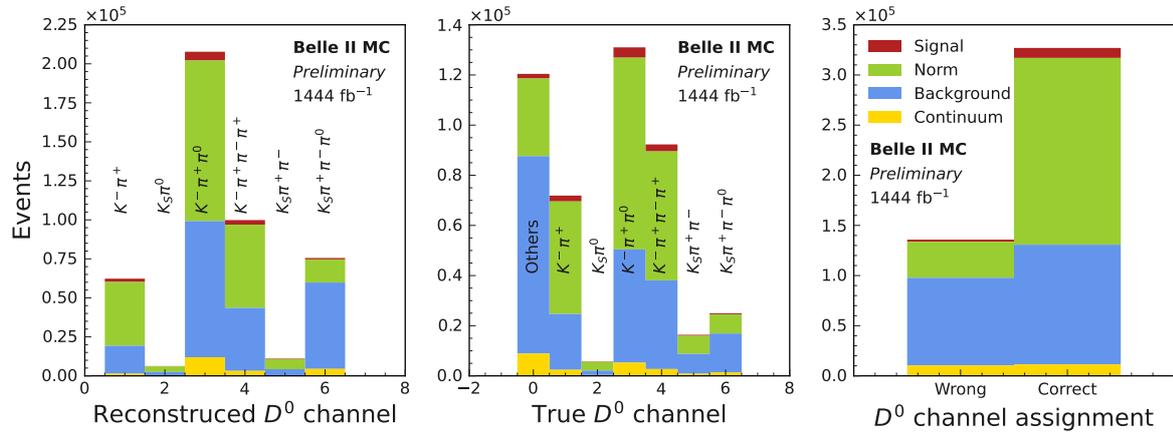
Figure 4.8: Reconstructed and true D meson decay modes for $1444\,\text{fb}^{-1}$ of generic run-dependent MC. The plots share the same legend, found in the rightmost plot.

---

[1]A different integrated luminosity only leads to a constant factor in the Figure of Merit. For optimisation purposes, it therefore only affects the statistics but not the optimal selection inferred.
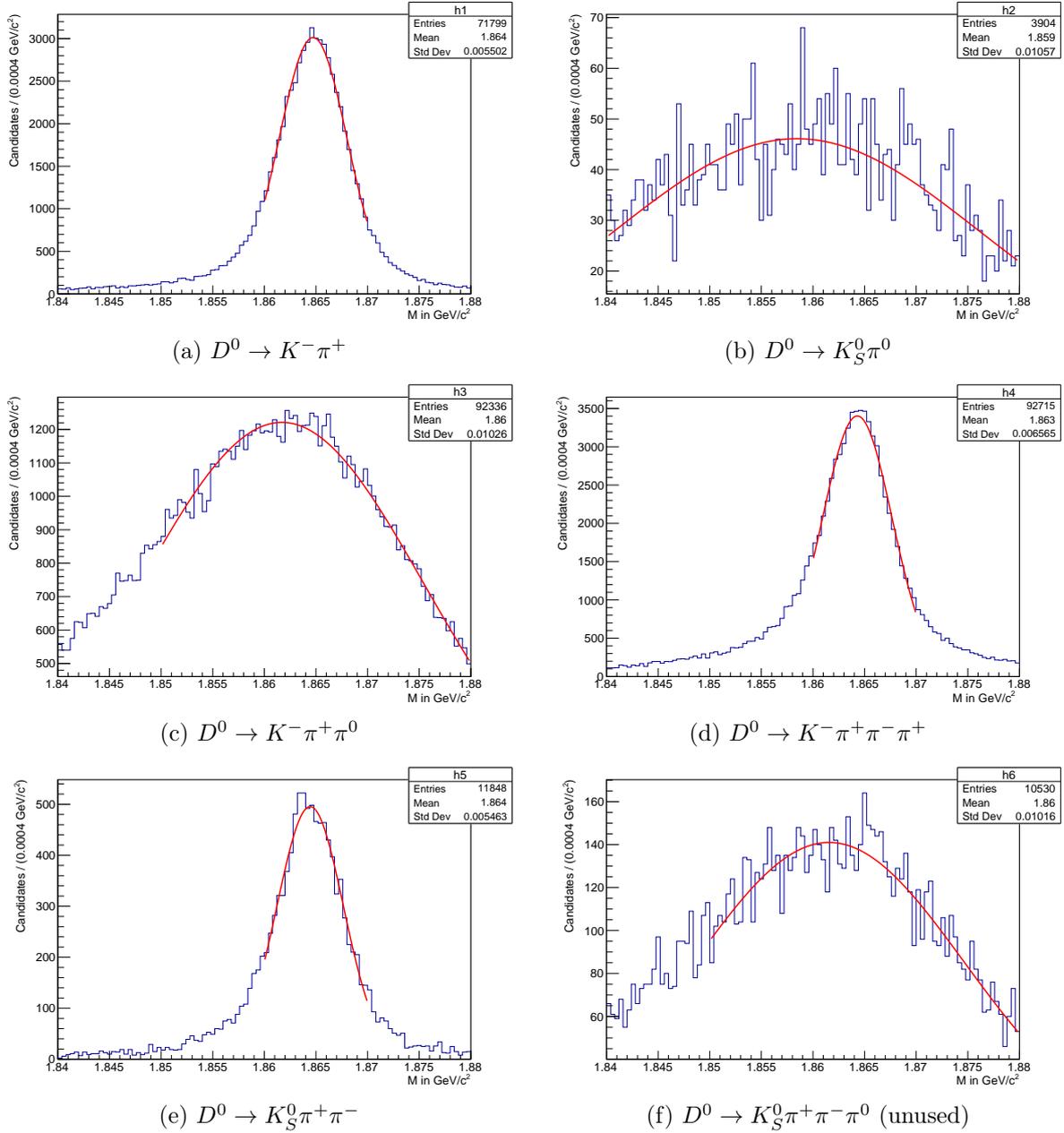
Figure 4.9: Peaks in the D^0 invariant mass and fitted Gaussians for different decay modes. The fit regions were chosen to capture the peaks well, ranging from $1.86\,\text{GeV/c}^2$ to $1.87\,\text{GeV/c}^2$ for the narrow peaks, and being wider for the broad peaks, for which the distributions seem more flat on the left side. Plotted are correctly reconstructed, true D^0 candidates from $7\,560\,000$ mixed and charged events each where the reconstruction went up to the D^* without a best candidate selection.

4.1.5 Signal B reconstruction

The D^0 can now be combined with a slow pion to form a $D^*(2010)^+$ (simply referred to as D^{*+} throughout this thesis). The slow pion is required to have a CMS momentum below $0.4\,\mathrm{GeV/c}$. The D^{*+} must have a CMS momentum of $p^{\mathrm{cms}} < 2.45\,\mathrm{GeV/c}$ in order to suppress highly energetic D^{*+} mesons from the $c\bar{c}$ continuum background, and its invariant mass must lie 0.14 to $0.15\,\mathrm{GeV/c^2}$ above the D^0's invariant mass. The D^* momentum is plotted in Figure 4.10. A vertex fit (`treeFit` with constrained π^0 mass) is performed on the D^* and candidates with a failed fit are discarded.

Originally, the D^* was also reconstructed as $D^{*+} \to D^+\pi^0$, though this was quickly dropped again due to the very low signal yield compared to the additional background it introduced; this can also be seen in Figure 4.10.
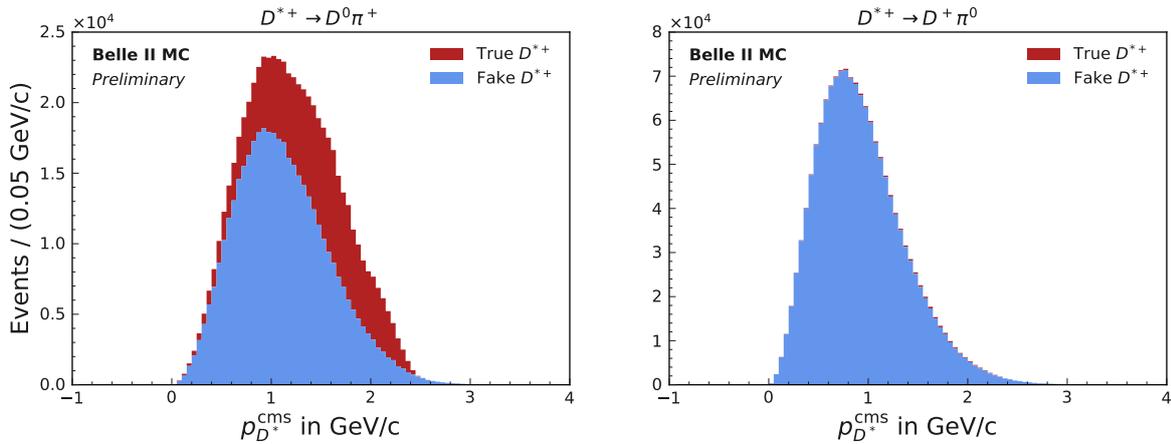

Figure 4.10: Distribution of the D^* CMS momentum, on the left for $D^{*+} \to D^0\pi^+$ and on the right for the unused channel $D^{*+} \to D^+\pi^0$. Plotted are $D^{*\pm}$ candidates from $7\,560\,000$ mixed and charged events each with the usual best D^* candidate selection (Section 4.3) applied.

Finally the D^* is combined with the τ candidate to get the B^0_{sig} candidate. Another vertex fit (`treeFit`) is applied and signal B candidates with failed fits are discarded.

4.2 Tag side selection

In order to do an inclusive tag B reconstruction, the so-called "Rest of Event" (ROE) object corresponding to the signal B candidate is built. It includes all tracks and ECL clusters of the event that are not used in the signal B's reconstruction (all KLM clusters are rejected via a cut of $E < 0$). In order to not discard relatively long-lived particles, only loose cuts are applied at first that require tracks to have $dr < 10\,\mathrm{cm}$, $|dz| < 20\,\mathrm{cm}$ and $E < 5.5\,\mathrm{GeV}$ as well as for the polar angle θ to lie within the CDC's angular acceptance of $17° < \theta < 150°$.

4.2.1 Cleaning up the tag side

To improve the track selection, first V0 particles are reconstructed, i.e. neutral particles decaying into two tracks, namely $\gamma \to e^+e^-$, $K^0_S \to \pi^+\pi^-$ and $\Lambda^0 \to p^+\pi^-$. The V0 decay vertex from which the tracks originate is typically displaced from the IP due to the relatively long lifetime of the V0s. The difference between the reconstructed V0 and its nominal mass is required to be less than $0.1\,\mathrm{GeV/c^2}$, and its daughters' particle IDs should be > 0.2. A mass fit (`treeFit` with constrained V0 mass) is applied and if successful, the two daughter tracks

in the ROE are replaced by the reconstructed/fitted V0. Now the track selection is tightened to only include final state tracks and V0s with $dr < 2\,\mathrm{cm}$ and $|dz| < 4\,\mathrm{cm}$.

The ROE can be further cleaned up by discarding beam background and fake photons. This is achieved using the photon MVA variables and requiring `beamBackgroundSuppression` > 0.1 and `hadronicSplitOffSuppression` > 0.1. These thresholds were taken from the previous analysis [12] and have been confirmed to work well. As a quick test, the effect of applying the signal side thresholds of 0.35 was studied, which not only lead to a slight decrease in the number of signal events, but also a significant decrease in the number of normalisation mode events, while the background levels increased greatly.

4.2.2 Lepton veto

Next, the goal is to veto candidates with a light lepton on the tag side since additional neutrinos in the event would make it more difficult to distinguish signal from normalisation. To achieve this, signal B candidates are discarded if there is an electron or muon candidate in the ROE with BDT-based `electronID` or likelihood-based `muonID` (ignoring the SVD) above 0.9 respectively. Muon candidates must also have a transverse momentum of $p_\mathrm{T} > 0.3\,\mathrm{GeV/c}$ in order to not discard too many misidentified pions.

4.2.3 Selection and tag B reconstruction

Now tag side selection criteria are applied. Specifically, the beam-constrained mass of the ROE

$$M_\mathrm{bc} \equiv \sqrt{\left(\frac{E_\mathrm{tot}^\mathrm{cms}}{2c^2}\right)^2 - \left(\frac{\vec{p}_\mathrm{tag}^{\,\mathrm{cms}}}{c}\right)^2} \tag{4.4}$$

which makes use of the fact that the total energy on the tag side is expected to be equal to half of the precisely known total CMS energy $E_\mathrm{tot}^\mathrm{cms}$, with $\vec{p}_\mathrm{tag}^{\,\mathrm{cms}}$ simply being the combined reconstructed tag side CMS momentum, should be at least $5.27\,\mathrm{GeV/c^2}$ and therefore large enough to accommodate the second B meson (mass $5.279\,66(12)\,\mathrm{GeV/c^2}$ [4]). Furthermore, the energy difference ΔE between the reconstructed tag side CMS energy $E_\mathrm{tag}^\mathrm{cms}$ and the expected value of $E_\mathrm{tot}^\mathrm{cms}/2$,

$$\Delta E \equiv E_\mathrm{tag}^\mathrm{cms} - \frac{E_\mathrm{tot}^\mathrm{cms}}{2}, \tag{4.5}$$

should fulfil $|\Delta E| < 0.5\,\mathrm{GeV}$. The tag B is then reconstructed directly from the remaining content of the ROE. However, it is in fact advantageous to only use the information on the direction of the reconstructed tag B momentum, while setting the energy and momentum magnitude to the expected values as shown in Sections 4.3.1 and 4.3.2. This is also the approach taken in this analysis.

4.3 $\Upsilon(4S)$ reconstruction and selection

Finally, the $\Upsilon(4S)$ meson is reconstructed from the signal and tag B in both the unmixed $\Upsilon(4S) \rightarrow B_\mathrm{sig}^0 \bar{B}_\mathrm{tag}^0$ as well as mixed $\Upsilon(4S) \rightarrow B_\mathrm{sig}^0 B_\mathrm{tag}^0$ case.

So far, there is no guarantee that each event only contains one $\Upsilon(4S)$ candidate, although around 84% do. The multiplicity that does exist arises from several D^* candidates present in an event, so $\Upsilon(4S)$ candidates are ranked, first by choosing the ones with the smallest difference between the nominal and reconstructed D^* masses and then by choosing the one with the smallest difference between the D^0 nominal and reconstructed masses.

An ambiguity in the lepton candidate was found to be extremely rare, since the lepton veto removes most events with several lepton candidates that pass the lepton selections. However, the cuts on signal and veto leptons are not quite the same, leading to a small fraction (0.03% after best D^*/D^0 selection) of events having both an electron and a muon candidate. In those cases, the electron candidate is correct more frequently and is chosen over the muon candidate.

4.3.1   Event cut on q^2

Due to the large τ mass, $B^0_{\mathrm{sig}} \to D^{*-}\tau^+\nu_\tau$ signal events have a minimum momentum transfer squared, which corresponds to the squared τ mass. In this analysis, a slightly higher cutoff of $q^2 > 3.5\,(\mathrm{GeV/c})^2$ is applied on the momentum transfer squared q^2, which is simply the exchanged W boson's four-momentum squared. q^2 can be expressed through the (true) CMS four-momenta of other particles as

$$q^2 = (p_{B_{\mathrm{sig}}} - p_{D^*})^2 = (p_\tau + p_\nu)^2 > 3.5\,(\mathrm{GeV/c})^2 \tag{4.6}$$

where all physical quantities are considered in the CMS frame, though the superscript is omitted for readability here and in the following. For the normalisation channel, q^2 may be lower, but applying the cut regardless will remove around 70%[2] of the continuum background while still keeping two thirds of the already abundant normalisation events.

Several possible methods to obtain q^2 were compared that can be divided into two approaches.

One approach is to be as direct as possible and calculate q by adding $p_\tau + p_\nu \equiv p_\ell + p_{\mathrm{miss}}$. Since the missing momentum p_{miss} cannot be accessed directly, different approximations are used to infer it. In all of these the missing energy is fixed as

$$E_{\mathrm{miss}} = \frac{E_{\mathrm{tot}}}{2} - E_{D^*\ell} \tag{4.7}$$

where $D^*\ell$ denotes the system of the D^* and the light lepton (e or μ) and E_{tot} refers to the total measured CMS energy corresponding to the $\Upsilon(4S)$ resonance at which Belle II operates. The missing three-momentum in the CMS frame is calculated in three different ways:

$$\vec{p}_{\mathrm{miss}} = -(\vec{p}_{D^*\ell} + \vec{p}_{\mathrm{tag}}) \tag{4.8}$$

for the first definition (A) where all measured three-momenta are used, namely those of the D^*, the lepton and the tag side, even if the reconstructed tag side momentum comes with a large uncertainty due to the inclusive nature of the analysis;

$$\vec{p}_{\mathrm{miss}} = -\vec{p}_{D^*\ell} \tag{4.9}$$

for the second definition (B) where the tag side CMS momentum is effectively set to $\vec{p}_{\mathrm{tag}} = 0$ reflecting the fact that the daughters of the $\Upsilon(4S) \to B\bar{B}$ two-body decay carry very little kinetic energy in the CMS frame; and

$$\vec{p}_{\mathrm{miss}} = -0.340\,\mathrm{GeV/c} \cdot \hat{p}_{\mathrm{tag}} - \vec{p}_{D^*\ell} \tag{4.10}$$

where $\hat{p}_{\mathrm{tag}}$ is the unit vector in $\vec{p}_{\mathrm{tag}}$ direction, so that only the direction of the reconstructed tag side momentum is used while the magnitude is fixed to the expected value fully determined by the kinematics of the $\Upsilon(4S) \to B\bar{B}$ two-body decay (Definition C).

---

[2]Assuming this value from the previous analysis [12] generalises.

Also tested were two manual[3] definitions where only the D^* and the tag side are considered, ignoring the lepton and therefore automatically combining the lepton and neutrino momenta. Both definitions start from

$$q^2 = \frac{1}{c^2}\left(\frac{E_{\text{tot}}}{2} - E_{D^*}\right)^2 - \sum_{i=x,y,z}(-p_{\text{tag},i} - p_{D^*,i})^2 \qquad (4.11)$$

where one case (Definition D) then simply uses the reconstructed tag side momentum

$$p_{\text{tag},i} = p_{\text{tag},i}^{\text{rec}} \qquad (4.12)$$

while in the other case (Definition E) the magnitude of the momentum is corrected to get the expected value

$$p_{\text{tag},i} = p_{\text{tag},i}^{\text{rec}} \cdot \frac{\sqrt{\left(\frac{E_{\text{tot}}}{2}\right)^2 - (5.279\,65\,\text{GeV})^2}}{\left|\vec{p}_{\text{tag}}^{\text{rec}}\right| c} \qquad (4.13)$$

so that only the direction of the reconstructed tag side momentum is used. Note that Eqs. 4.12 and 4.13 define the spatial *components* of p_{tag}.

In all of the above cases, the tag side energy is fixed to the expected value, i.e. half the centre-of-mass energy:

$$E_{\text{tag}} = \frac{E_{\text{tot}}}{2}. \qquad (4.14)$$

To find out which definition describes q^2 the most precisely, the difference between each of the definitions to the true q^2 is plotted. As can be seen in Figure 4.11, the last definition (E, Eqs. 4.11 and 4.13) gives the most precise results and is therefore used in this analysis.
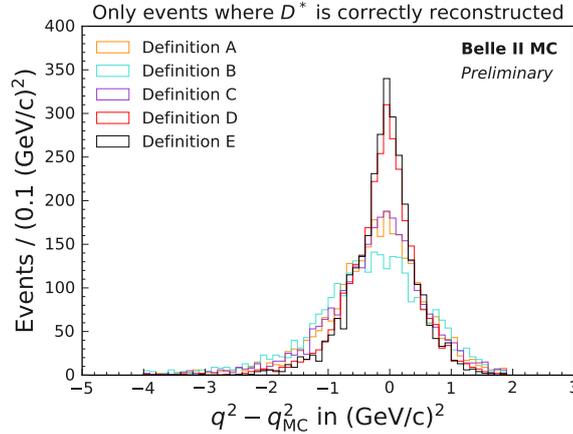
Figure 4.11: Deviation of different q^2 definitions from the true value. Only events where the D^* has been correctly reconstructed are considered, essentially leaving only normalisation mode events which have a clean signature. Definitions A-C set E_{miss} according to Eq. 4.7 while calculating $\vec{p}_{\text{miss}}$ using Eqs. 4.8, 4.9, or 4.10 respectively. Definitions D and E are based on Eq. 4.11 and get the tag side momentum components via Eqs. 4.12 and 4.13 respectively. In this analysis, Definition E is used since it shows the narrowest peak. Plotted are events from $14.96\,\text{fb}^{-1}$ of run-dependent MC where the D^* is correctly reconstructed and no q^2 cut has been applied.

---

[3]The previous definitions are already implemented in *basf2*.

4.3.2   Comparison of different definitions of M^2_{miss}

As was done for q^2, different possible definitions to obtain the missing mass squared

$$M^2_{\mathrm{miss}} = \frac{E^2_{\mathrm{miss}}}{c^4} - \frac{\vec{p}^{\,2}_{\mathrm{miss}}}{c^2} \tag{4.15}$$

were compared. Again, all quantities are taken to be in the CMS frame without explicitly writing the superscript. The missing energy E_{miss} is the difference between the known total energy E_{tot} and the observed energy in the detector. The tag side is expected to carry half of the total energy, so the missing energy can be expressed as Eq. 4.7, with $\vec{p}_{\mathrm{miss}}$ defined as was done for q^2 definitions A-C (Eqs. 4.8-4.10). As normalisation mode events contain only one neutrino on the signal side (and should have no neutrinos on the tag side), they show a nice peak at $M^2_{\mathrm{miss}} = 0$. The resolution of this peak was used as the criterion by which to choose the best definition. The histograms are shown in Figure 4.12, and the standard deviations of the distributions are listed in Table 4.4 along with their uncertainties. As one can easily see, Definition B yields by far the worst results, while for Definitions A and C the resolution is identical within its uncertainty.

To further probe the resolutions, Gaussians were fitted to the distributions in the central peak region of $|M^2_{\mathrm{miss}}| < 0.4\,\mathrm{GeV}^2/\mathrm{c}^4$ ($1\,\mathrm{GeV}^2/\mathrm{c}^4$ for Definition B to capture its peak), and the resulting standard deviations are added to Table 4.4. In the resolutions of the Gaussians, Definition C shows a tangible improvement over Definition A. Furthermore, Definition C uses a similar approach than the chosen q^2 Definition E where only the direction of the reconstructed tag side momentum is used. Hence Definition C (Eq. 4.15 with Eqs. 4.7 and 4.10) is used in this analysis.
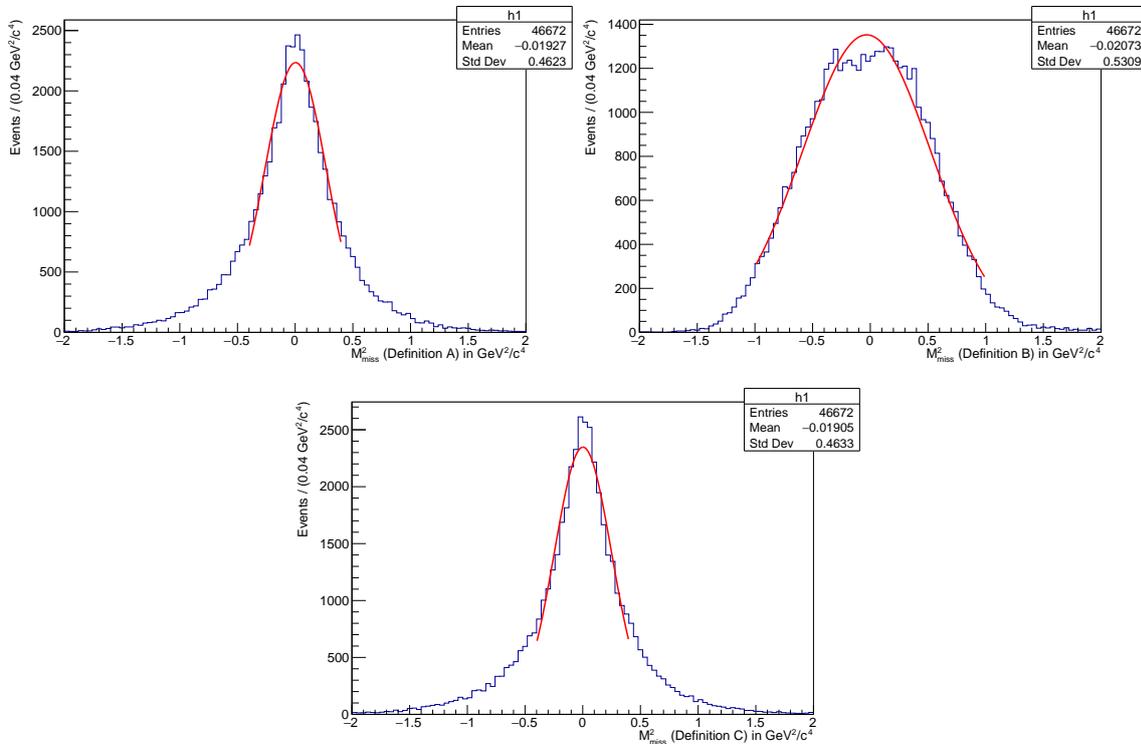

Figure 4.12: Comparison of the normalisation mode peaks in M^2_{miss} for different definitions. Plotted are the normalisation mode events from $400\,\mathrm{fb}^{-1}$ of run-independent MC.

Definition	σ_{Hist} [GeV^2/c^4]	σ_{Gaussian} [GeV^2/c^4]
A	0.4623 ± 0.0015	0.2653 ± 0.0027
B	0.5309 ± 0.0017	0.5578 ± 0.0033
C	0.4633 ± 0.0015	0.2476 ± 0.0023

Table 4.4: Resolutions of different M^2_{miss} definitions as determined by the standard deviations of the histograms and of Gaussians fitted to the normalisation mode peaks at $M^2_{\text{miss}} = 0$.

4.4 Data

When running over real data, the track momentum scale is adjusted by a small factor to correct for a bias in the map of the magnetic field used for reconstruction [26]. As the measurement is blinded, a sideband cut of $|M^2_{\text{miss}}| < 1\,\text{GeV}^2/\text{c}^4$ is applied to blind the signal region.

4.5 Selection summary

A summary of the entire selection can be found in Table 4.5.

	M [GeV/c^2]	E [GeV]	p^{cms} [GeV/c]	PID	dr [cm]	$|dz|$ [cm]	θ in CDC	γ MVAs	Vertex fit	Other
Signal side										
B^0_{sig}									treeFit	
$D^{*\pm}$	$M_{D^0} + (0.14–0.15)$								treeFit*	
D^0	see Table 4.3								treeFit*	
$K^{\pm}$				> 0.5	< 0.5	< 2	✓			
$\pi^{\pm}$				> 0.1	< 0.5	< 2	✓			
K^0_S	0.494—0.501								kFit	goodBelleKshort
π^0	0.115—0.155									
γ								> 0.35	kFit	
$\pi^{\pm}_{slow}$										
τ		< 0.4		> 0.9	< 0.5	< 2	✓			
ℓ		> 0.4	< 2.45	> 0.9	< 0.5	< 2	✓			e brems-corrected
Tag side										
B^0_{tag}/ROE	$M_{bc} > 5.27$	$|\Delta E| < 0.5$							kFit	
e_{veto}		< 5.5		> 0.9	< 2	< 4	✓			e/μ veto
μ_{veto}		< 5.5	$p_T > 0.3$	> 0.9	< 2	< 4	✓			
V0	$|\Delta M^{nom}| < 0.1$								treeFit	
Tracks		< 5.5		> 0.2	< 10	< 20	✓			
Tracks		< 5.5			< 2	< 4	✓			
γ								> 0.1		
K^0_L										
ℓ				< 0 (reject all)						

Best candidate selection: min $|M^{rec}_{D^*} - M^{nom}_{D^*}|$ (may keep several), then min $|M^{rec}_{D^0} - M^{nom}_{D^0}|$, then choose $\ell = e$ over μ (keep one)

Event cut: $q^2 > 3.5\,(\text{GeV}/c)^2$

Sideband cut for data: $|M^2_{miss}| < 1\,(\text{GeV}/c^2)^2$

Table 4.5: Selection summary. M/M_{bc} is the invariant/beam-constrained mass and ΔM^{nom} the mass difference to the particle's nominal mass, E the energy and ΔE the energy difference defined in Eq. 4.5, p^{cms}/p_T the CMS/transverse momentum, PID the respective particle ID, dr and $|dz|$ are the radial and z-distance to the IP, θ in CDC means that $17° < \theta < 150°$, and the γ MVAs refer to beamBackgroundSuppression and hadronicSplitOffSuppression. Also listed is which vertex fit method was used (candidates with a failed fit are discarded). The star denotes here that the π^0 daughter mass was constrained in the fit. Lastly, other selection criteria are indicated.

Chapter 5

Background Rejection

As an inclusive measurement, the background levels present in this analysis are increased and need to be dealt with accordingly. In order to reject background, a classifier was trained to separate events into three classes: signal, normalisation, and background. The used model was a Boosted Decision Tree (BDT), namely a *scikit-learn* `GradientBoostingClassifier` [27] with 200 boosting stages each using a subsample containing half of the training data set (see Section 5.2) for stochastic gradient boosting. To account for the situation of very unbalanced classes, the training data set was enhanced with additional signal MC, and balanced class weights were employed in the training, meaning that the loss function is modified in order to proportionally increase the punishment for misclassification of the underrepresented class. Also tested, though only superficially, were *scikit-learn*'s `HistGradientBoostingClassifier` as well as the `MLPClassifier` implementation of a neural network. The next section discusses the input variables used in the multiclass classifier, while the training data set is introduced in Section 5.2. The classifier's performance is illustrated in Section 5.3, and different approaches on how to apply the classifier are considered in Section 5.4.

5.1 Input features

As BDTs are quite robust to overtraining, the choice of input features was largely left to the classifier itself. This was done by removing all truth-based variables (including assigned event categories) from the input data, as well as variables that were found to have bad data-MC agreement. Also excluded were M_{miss}^2 and p_ℓ^{cms}, as well as all variables with an absolute correlation larger than 0.5 to the missing mass squared or CMS lepton momentum, in order to not interfere with the fit. Then a classifier of the same type was trained using all remaining variables as input features. The relative importance of each input feature was evaluated using the `feature_importances_` attribute and the ten most important features were chosen as the input for the final classifier training. The list of used features and their respective importances (given in the brackets) in the final classifier are as follows:

- The signed difference in the invariant mass of the signal B to the nominal mass in units of the invariant mass uncertainty $\mathtt{SigM}_{B_{\mathrm{sig}}} := (M_{B_{\mathrm{sig}}}^{\mathrm{rec}} - M_{B_{\mathrm{sig}}}^{\mathrm{nom}})/\sigma_{M_{B_{\mathrm{sig}}}}$ (25.16%).

- The uncertainty on the signal B's invariant mass $\sigma_{M_{B_{\mathrm{sig}}}}$ (15.74%).

- The sum of the D^* and light lepton CMS momenta $|\vec{p}_{D^*}^{\,\mathrm{cms}} + \vec{p}_\ell^{\,\mathrm{cms}}|$ which surprisingly passed the correlation requirements (10.89%).

- The Kakuno-Super-Fox-Wolfram moment H_{2m}^{so}, described below (9.71%).

- The $\Upsilon(4S)$ CMS momentum $|\vec{p}^{\,\mathrm{cms}}_{\Upsilon(4S)}|$ (8.16%).

- The momentum transfer squared q^2 (7.67%).

- The uncertainty on the signal B's flight time `flightTimeErr`$_{B_{\mathrm{sig}}}$ (6.42%).

- The D^0 decay mode ID `dID`$_{D^0}$ (6.38%).

- The mass difference between the D^* and the D^0 $M_{D^*} - M_{D^0}$ (5.89%).

- The signal B CMS momentum $|\vec{p}^{\,\mathrm{cms}}_{B_{\mathrm{sig}}}|$ (3.99%).

Notably, $|\vec{p}^{\,\mathrm{cms}}_{D^*} + \vec{p}^{\,\mathrm{cms}}_{\ell}|$ and $|\vec{p}^{\,\mathrm{cms}}_{B_{\mathrm{sig}}}|$ should be identical by definition, though the classifier deems it useful to use both. Generally, the shapes of the variable distributions are very similar for the three classes, with the signal events' shift towards higher q^2 values as the only exception. Mostly, the classifier seems to combine the variables to access helpful quantities. This was evident throughout the development of the classifier as it kept reconstructing the missing mass squared from the given variables, leading to problems in the fit (Chapter 6). For example, one can easily see how combining $|\vec{p}^{\,\mathrm{cms}}_{D^*} + \vec{p}^{\,\mathrm{cms}}_{\ell}|$ (or $|\vec{p}^{\,\mathrm{cms}}_{B_{\mathrm{sig}}}|$) with q^2 allows to draw conclusions on the missing momentum. Similarly, the $\Upsilon(4S)$ CMS momentum is expected to be equal and opposite to the combined momenta of not reconstructed particles, including neutrinos and tracks/clusters not passing the selection criteria. `SigM`$_{B_{\mathrm{sig}}}$, $\sigma_{M_{B_{\mathrm{sig}}}}$, `flightTimeErr`$_{B_{\mathrm{sig}}}$ and $M_{D^*} - M_{D^0}$ all serve as a measure of how well the particle hypotheses fit, while knowledge of the D^0 decay mode allows for decay channel specific selections.

The Kakuno-Super-Fox-Wolfram moments are designed to help with continuum suppression, with H^{so}_{2m} referring to the moment comparing *signal* side particles with *other* particles, in this case *missing* particles. The "2" refers to the use of the second-order Legendre polynomial:

$$H^{so}_{2m} = \sum_i |\vec{p}_{\mathrm{miss}}| P_2(\cos\theta_{i,\mathrm{miss}}) \tag{5.1}$$

where i iterates over all signal side particles, $\vec{p}_{\mathrm{miss}}$ is the missing momentum and $P_2(\cos\theta_{i,\mathrm{miss}})$ is the second-order Legendre polynomial where $\theta_{i,\mathrm{miss}}$ is the angle between the i-th signal side particle and the missing momentum. [28, 29]

Plots showing the data-MC agreement for these variables can be found in Figures 5.1 and 5.2.
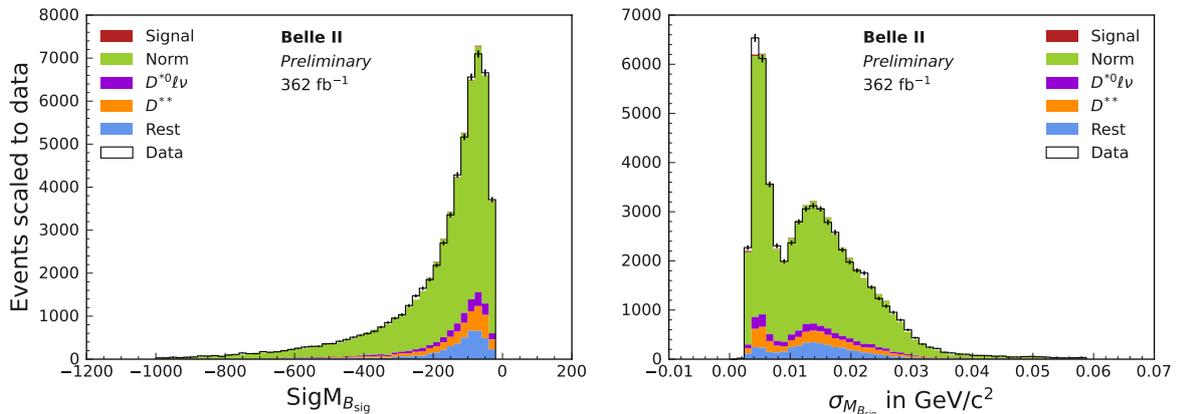

Figure 5.1: Data-MC agreement of the classifier input variables for the $400\,\mathrm{fb}^{-1}$ training data set in the $|M^2_{\mathrm{miss}}| < 1\,\mathrm{GeV}^2/\mathrm{c}^4$ sideband. Note that PID correction weights (see Section 7.2) are applied to the MC events, and that the total weighted number of MC events is scaled to that in data.

Figure 5.2: Data-MC agreement of the classifier input variables for the $400\,\text{fb}^{-1}$ training data set in the $|M^2_{\text{miss}}| < 1\,\text{GeV}^2/\text{c}^4$ sideband. Note that PID correction weights (see Section 7.2) are applied to the MC events, and that the total weighted number of MC events is scaled to that in data.
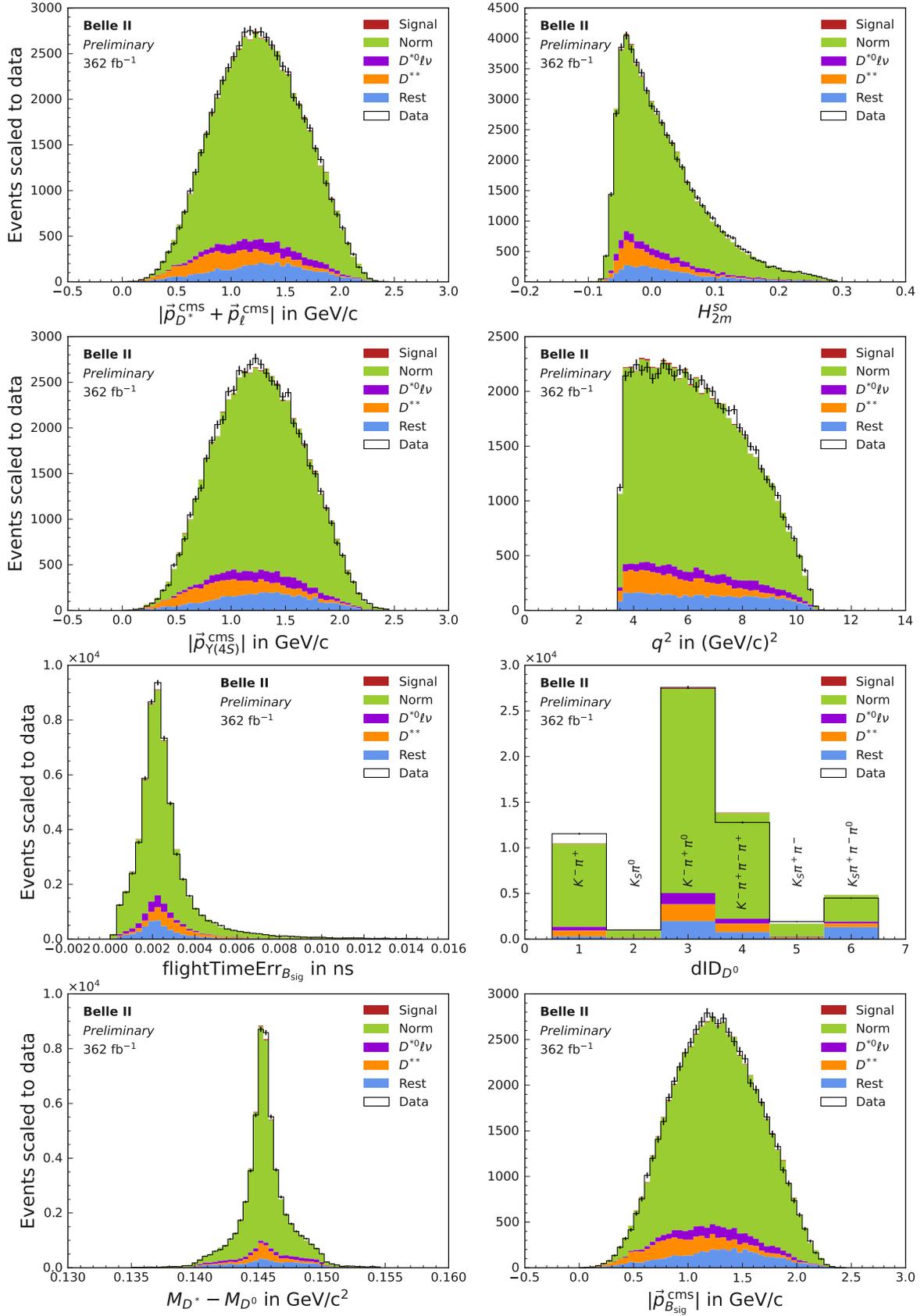
As already mentioned, the input variables were required to be well-described in the MC. Lepton ID corrections had to be applied to the MC as they affect the shapes. Since no other corrections are applied to the MC, including ones that introduce an overall scaling factor, the total number of MC events had to be scaled down to match that in the data, as the distributions would not have matched up when simply scaling by the integrated luminosity. A further difference in the scaling is expected to arise from the tag side efficiency, which is likely not particularly well-described in the MC, but should cancel in the ratio $R(D^*)$, not affecting the measurement. The only major difference visible is in the D^0 decay mode ID, $\texttt{dID}_{D^0}$. As $D^0 \to K^-\pi^+$ is the cleanest channel among those considered, the observed deviation is likely a result of the scaling, which would also explain why the more difficult decay modes involving a π^0 seem to be surprisingly well-described – the MC yields are likely shifted. Nonetheless, the agreement is deemed good enough for the variable to be used.

5.2 Training data set

The sample used for the classifier training consists of $400\,\text{fb}^{-1}$ of generic run-dependent MC as well as the available signal MC contributing an additional 1026 signal events. In total, the combined sample contains 3998+1026 signal events, 71 428 normalisation events and 60 733 background events. 90% of the sample was used for the training itself, while the remaining 10% were used for validation. As the classifier was found to profit from the additional information, the sixth D^0 decay mode, $D^0 \to K_S^0\pi^+\pi^-\pi^0$, was included in the training data set.

5.3 Performance

The classifier returns a tuple of three numbers $\texttt{sigProb}$, $\texttt{nrmProb}$, $\texttt{bkgProb}$ adding up to one for each event, corresponding to the probabilities of this event being signal, normalisation or background respectively.

The Receiver Operating Characteristic (ROC) curves of the classifier's performance for the three classes can be found in Figure 5.3, and the distributions of the output probabilities in Figure 5.4.
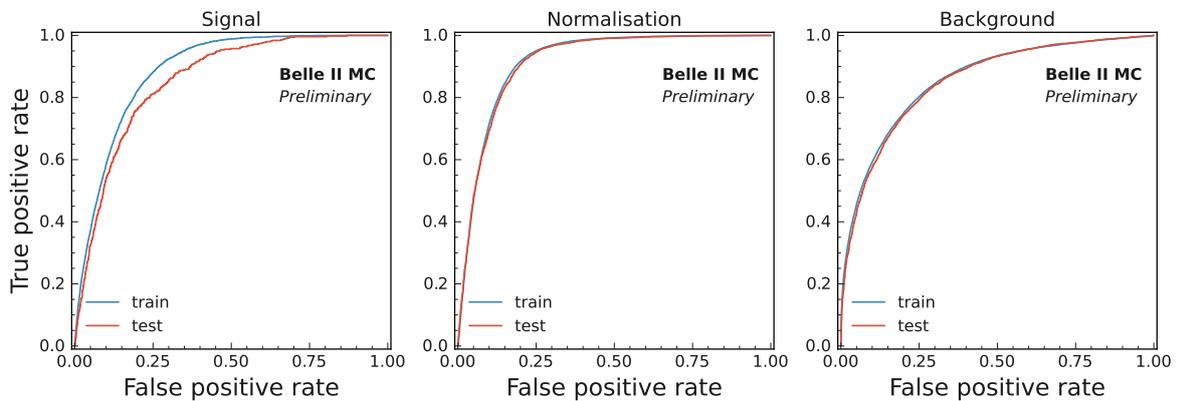

Figure 5.3: Receiver Operating Characteristic (ROC) curves of the multiclass classifier. Due to the relatively low signal statistics, there is notable overtraining in the signal class, while the classifier performs best for normalisation events thanks to their clean signature. Plotted are $400\,\text{fb}^{-1}$ generic + 1026 additional signal MC events, 90% of which were used for training.
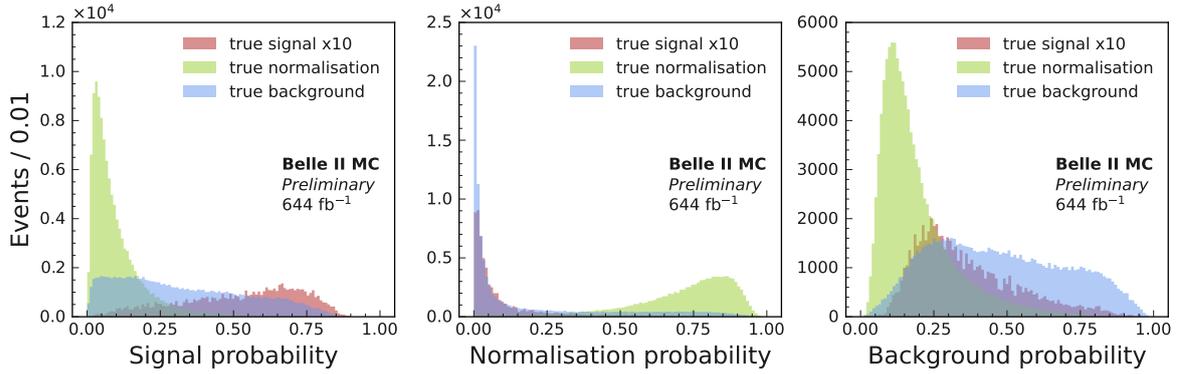
Figure 5.4: Classifier output probabilities for an independent sample of $644\,\mathrm{fb}^{-1}$ of generic, run-dependent MC. Note that the true signal component has been scaled up by a factor 10.

5.4 Application

Since the goal was to reject background, a first approach was to simply discard events which were classified as background (i.e. the background probability was larger than the signal and normalisation probabilities). However, a more nuanced selection based on the classifier output proved to be more effective and will be detailed in the following.

5.4.1 Optimising cuts by the significance of $R(D^*)$

Since the classifier is quite capable of isolating normalisation events, a new figure of merit was needed that also valued $B \to D^*\ell\nu$ events. Naturally, the significance of $R(D^*)$ was chosen, approximated by splitting events depending on their missing mass squared as

$$Z(R) = 1/\sqrt{\left(\frac{\sigma_{N_{\mathrm{sig}}}}{N_{\mathrm{sig}}}\right)^2 + \left(\frac{\sigma_{N_{\mathrm{norm}}}}{N_{\mathrm{norm}}}\right)^2} \tag{5.2}$$

where for $\sigma_{N_{\mathrm{sig}}} = \sqrt{N_{\mathrm{sig}} + N_{\mathrm{bkg}}}$, only events with $M_{\mathrm{miss}}^2 \geq 1\,\mathrm{GeV}^2/\mathrm{c}^4$ are counted, while for $\sigma_{N_{\mathrm{norm}}} = \sqrt{N_{\mathrm{norm}} + N_{\mathrm{bkg}}}$, only events with $M_{\mathrm{miss}}^2 < 1\,\mathrm{GeV}^2/\mathrm{c}^4$ contribute.

As the normalisation mode events have the cleanest signatures, they are also the easiest for the classifier to identify correctly. And as the background greatly outnumbers the signal events, most of the middle part of a ternary plot of the classifier output (Figure 5.5) is dominated by background. Therefore, when optimising the cuts on the classifier output probabilities, the form should reflect the goal of isolating the signal and normalisation corners of the triangle.

Consequently, the chosen form of the cut was (with the optimal thresholds)

$$(\texttt{sigProb} > 0.35 \ \& \ \texttt{nrmProb} < 0.20 \ \& \ \texttt{bkgProb} < 0.50)$$
$$\text{or} \tag{5.3}$$
$$(\texttt{sigProb} < 0.70 \ \& \ \texttt{nrmProb} > 0.00 \ \& \ \texttt{bkgProb} < 0.40)$$

where each bound was varied independently in steps of 0.05. As there are no events in the small triangle formed by $\texttt{sigProb} > 0.7$, $\texttt{nrmProb} > 0.2$ and the zero-background-probability axis, there is no reason to exclude the region, so the cut simplifies to

$$(\texttt{sigProb} > 0.35 \ \& \ \texttt{nrmProb} < 0.20 \ \& \ \texttt{bkgProb} < 0.50) \ \text{or} \ \texttt{bkgProb} < 0.40. \tag{5.4}$$

The resulting constraints on the classifier output are illustrated in Figure 5.5, and the effect on the M_{miss}^2 and p_ℓ^{cms} distributions can be found in Figures 5.6 and 5.7.
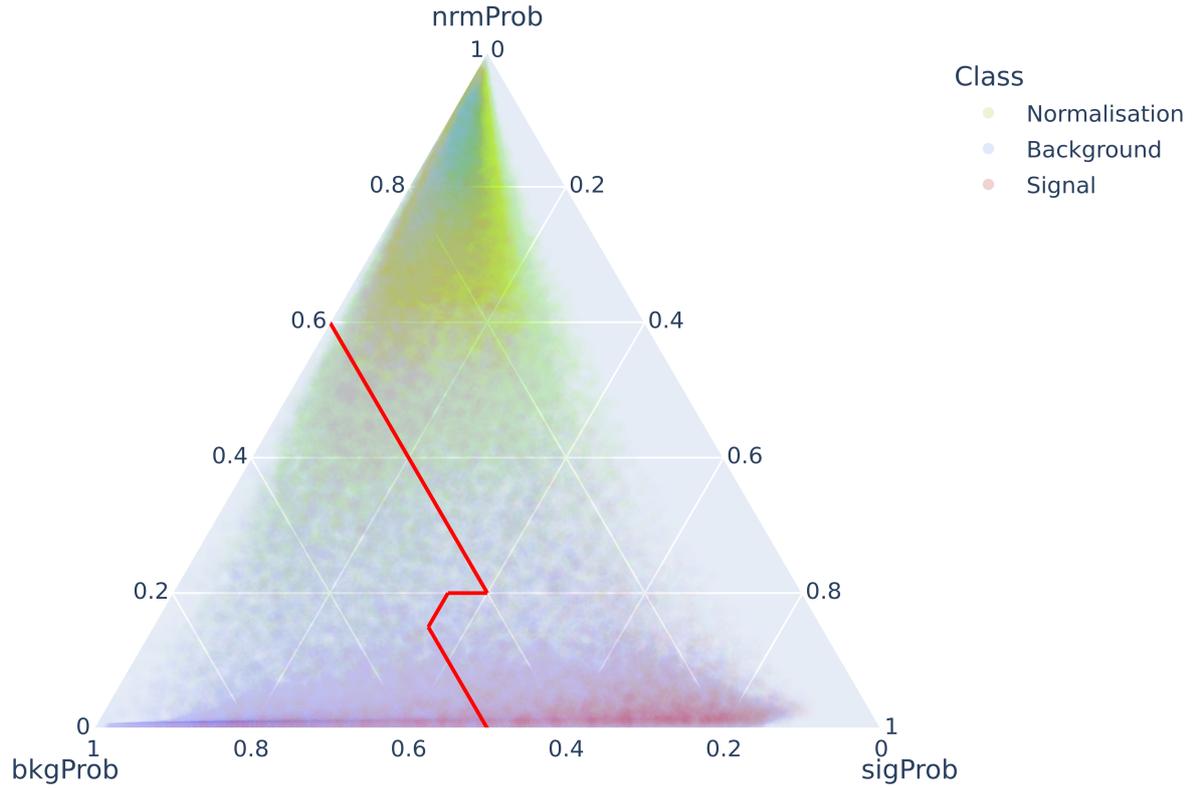
Figure 5.5: Ternary plot of the classifier output probabilities for an independent sample of $644\,\mathrm{fb}^{-1}$ of generic, run-dependent MC. The red lines indicate the optimised selections for the signal and normalisation corners. Note that the data points are drawn with opacity 0.01 in the order true normalisation > true background > true signal, meaning that e.g. background covers up normalisation events as can be seen in the high normalisation probability corner. Created with `Plotly Express` [30].
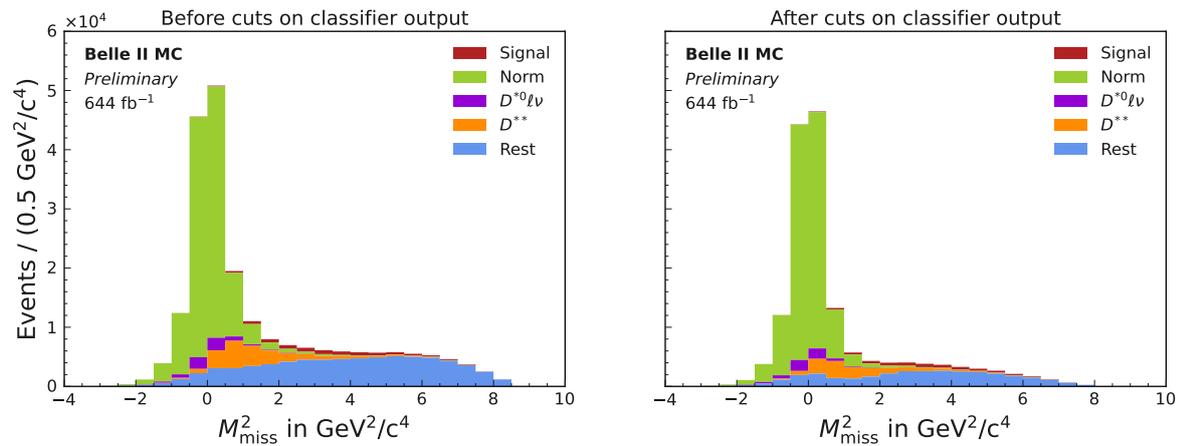

Figure 5.6: Distribution of M^2_{miss} before and after applying the optimal cut on the classifier output probabilities. Plotted are events from an independent sample of $644\,\mathrm{fb}^{-1}$ of generic, run-dependent MC.

Figure 5.7: Distribution of p_ℓ^{cms} before and after applying the optimal cut on the classifier output probabilities. Plotted are events from an independent sample of $644\,\mathrm{fb}^{-1}$ of generic, run-dependent MC.
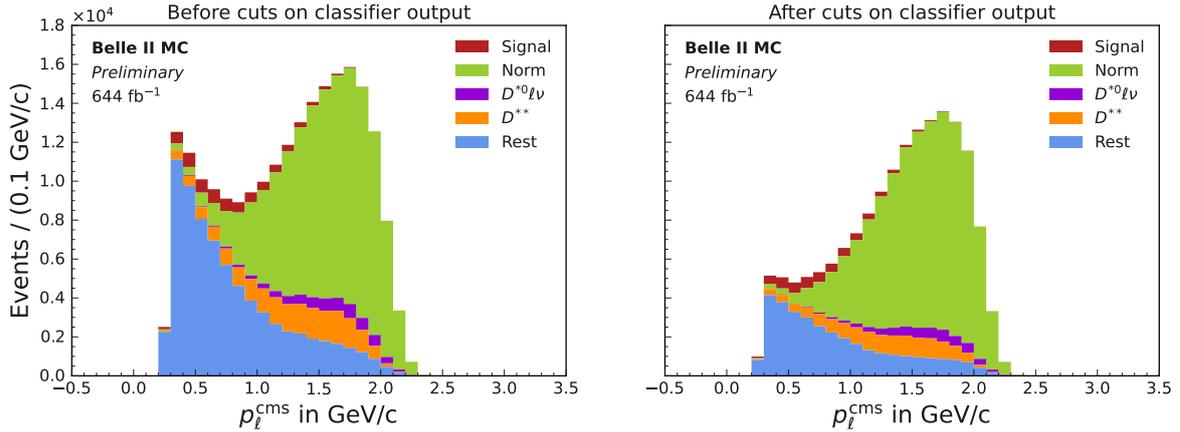
A quick test was performed to compare fit results (see Chapter 6) before and after applying these classifier cuts. Unfortunately, despite the better $R(D^*)$ significance, the statistical fit uncertainty on $R(D^*)$ turned out to be much larger (29% instead of 17%) in the case of applied cuts. A parametrisation of $Z(R)$ in terms of the uncertainties from the fit proved difficult, so a different approach was considered and will be described in the following section.

5.4.2 Optimising cuts by the statistical fit uncertainty

As the goal is to extract a result on $R(D^*)$ with as small an uncertainty as possible, the cuts on the classifier output probabilities were also optimised by the resulting statistical fit uncertainty on $R(D^*)$. This was done by applying a multitude of different cut combinations on both the PDF shape sample and the fit sample, recreating the PDF shapes and performing the fit (explained in detail in Chapter 6). It is important to note that the sample used in the classifier training was not reused for this purpose. As running the fit many times is computationally expensive, the number of possible cut combinations was reduced to the form (with the optimal threshold values)

$$\texttt{nrmProb} > 0.8 \ \texttt{or} \ (\texttt{sigProb} > 0.0 \ \texttt{\&} \ \texttt{nrmProb} < 0.7 \ \texttt{\&} \ \texttt{bkgProb} < 1.0) \qquad (5.5)$$

where the thresholds were varied in steps of 0.1 (0.2 in the high `sigProb` region that had previously shown to yield extremely bad results due to the very low statistics). Strangely, the fit prefers to discard only a small band in the high `nrmProb` corner, though this was consistently observed over different versions of the classifier and therefore unlikely to be a fluctuation. With only a slightly higher statistical fit uncertainty not even captured by the precision reported here, the second best cut combination was found to be the case of no applied cuts on the classifier output. Combined with the fact that not using the classifier allows the use of the training sample in the fit, the classifier did go unused in the end, though it seems difficult to believe that the approach of a multiclass classifier could not be useful in obtaining a more precise result. Perhaps the classifier did still affect the distributions in M_{miss}^2 and p_ℓ^{cms} too much, just not in a way that could be caught by looking at the one-dimensional projections as plotted in Figures 5.6 and 5.7. In that case, by manually picking "safe" input variables, or alternatively not fitting to M_{miss}^2, one may be able to profit from such a classifier.

5.4.3 Classifier output as fitting variables

As a final but relatively quick test, fits were also performed to classifier output probabilities. Since bimodal distributions are problematic to fit to, each probability p was transformed as

$$p \to \ln \frac{p - p_{\min} + \epsilon}{p_{\max} - p + \epsilon} \tag{5.6}$$

with a small shift $\epsilon = 0.000000001$ to avoid infinities. In one such fit, the transformed `sigProb` was fitted alongside M_{miss}^2, and in the other, `sigProb` and `bkgProb`, both transformed, were fitted. The classifiers used here were allowed to use information on (and highly correlated to) the CMS lepton momentum, with the second classifier also using the highly valuable M_{miss}^2 (and correlated information). Plots of the transformed probabilities of the second case can be found in Figure 5.8. Neither of these fits saw an improvement over the combination of M_{miss}^2 and p_{ℓ}^{cms}, with the central $R(D^*)$ values being far from the true MC value, even though the uncertainties did in fact decrease, both relatively and absolutely.
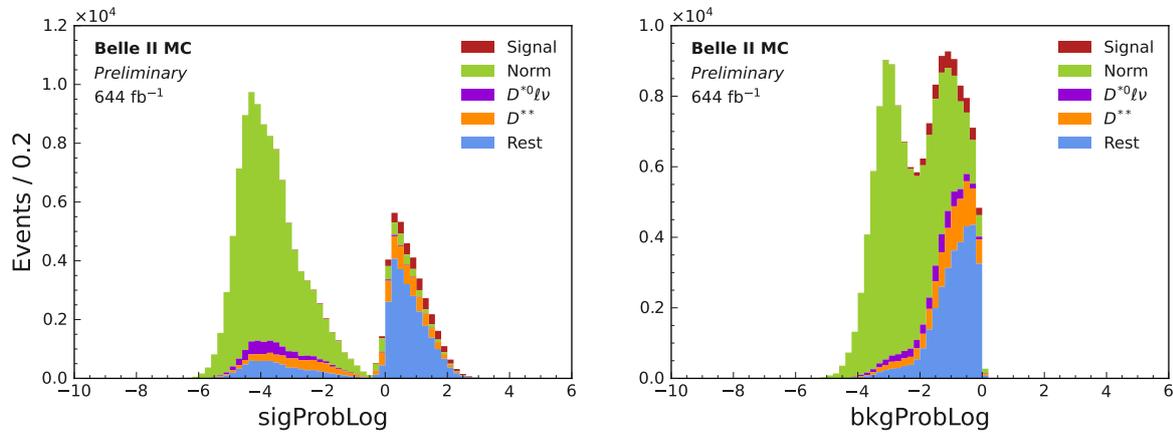
Figure 5.8: Logarithmically transformed (see Eq. 5.6) `sigProb` and `bkgProb` distributions. Plotted are events from an independent $644\,\mathrm{fb}^{-1}$ sample of generic, run-dependent MC.

With the currently unexplained, large differences to the true value, the fit results themselves are not in a presentable form. Further investigation into both the fit's behaviour and different transformations needs to be done, which was not possible due to time constraints. Certainly, this approach's uncertainties look promising. Considering the large uncertainty introduced by the so-called gap mode background (discussed in Section 7.4), suppressing backgrounds has become even more important.

Chapter 6

Fit

In order to extract the signal and normalisation mode yields, a two-dimensional unbinned maximum likelihood fit is performed using the `RooFit` [31] package. Five components are included in the fit: signal ($B^0 \to D^{*-}\tau^+(\ell^+\nu_\ell\bar{\nu}_\tau)\nu_\tau$), normalisation ($B^0 \to D^{*-}\ell^+\nu_\ell$), D^{*0} ($B^+ \to \bar{D}^{*0}\ell^+\nu_\ell$, same as normalisation mode but starting from a charged B), D^{**} ($B \to D^{**}\ell^+\nu_\ell$ with $D^{**} = D_1, D_0^*, D_1', D_2^*$), and the "rest". D^{*0} is set to scale with the normalisation component as the fit cannot distinguish between them, with the scaling factor being determined by the ratio $N_{D^{*0}}/N_{\mathrm{norm}}$ of D^{*0}/normalisation mode event yields in the $1044\,\mathrm{fb}^{-1}$ PDF shape sample (see below). The fitted variables in the 2D fit are the missing mass squared M_{miss}^2 and the CMS lepton momentum p_ℓ^{cms}.

The Probability Density Function (PDF) shapes of the fit components are obtained via a Kernel Density Estimation (KDE), which approximates the shapes through a sum of Gaussians with adaptive widths. For this, a sample containing $1044\,\mathrm{fb}^{-1}$ of generic run-dependent MC events (called the PDF shape sample) is used. The M_{miss}^2 and p_ℓ^{cms} distributions are mirrored on both sides to minimise edge effects. The combined model is fitted to a $400\,\mathrm{fb}^{-1}$ fit sample of generic, run-dependent MC to estimate the sensitivity on the currently available Belle II $\Upsilon(4S)$ data set of $362(2)\,\mathrm{fb}^{-1}$. The minimiser used is `Minuit2` [32].

Plots of the PDF shapes can be found in Figures 6.1 and 6.2. The data-MC agreement of the fitting variables in the $|M_{\mathrm{miss}}^2| < 1\,\mathrm{GeV}^2/\mathrm{c}^4$ sideband region is plotted in Figure 6.3. Note that since no corrections are applied to the MC other than PID corrections (see Section 7.2), the total number of MC events is scaled to match the event yield in data rather than scaling by the integrated luminosity L_{int}.

From the fitted signal and normalisation mode yields, $R(D^*)$ is calculated as

$$R(D^*) = \frac{2 \cdot N_{\mathrm{sig}}^{\mathrm{fit}}/\epsilon_{\mathrm{sig}}}{N_{\mathrm{norm}}^{\mathrm{fit}}/\epsilon_{\mathrm{norm}}} \tag{6.1}$$

with the factor 2 reflecting that $B^0 \to D^{*-}e^+\nu_e$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ are measured together in this analysis, while in the definition of $R(D^*)$ (Eq. 1.1), either of the two is used. The efficiencies ϵ are calculated from the number of selected divided by the number of generated events $N_{\mathrm{gen}}^l = 2\sigma_{B^0\bar{B}^0}L_{\mathrm{int}} \cdot \mathcal{B}(B^0 \to D^{*-}l^+\nu_l) \cdot \mathcal{B}(D^{*+} \to D^0\pi^+) \cdot \sum_{\mathtt{dID}_{D^0}} \mathcal{B}(D^0 \to ...)$ with $B^0\bar{B}^0$ production cross section $\sigma_{B^0\bar{B}^0}$ and MC branching fractions $\mathcal{B}$ in the PDF shape sample.

The statistical uncertainty on the fitted $R(D^*)$ value is calculated from the statistical fit errors on the yields via the Gaussian error propagation:

$$\sigma_R = \sqrt{\left(\frac{2/\epsilon_{\mathrm{sig}}}{N_{\mathrm{norm}}^{\mathrm{fit}}/\epsilon_{\mathrm{norm}}}\right)^2 \sigma_{N_{\mathrm{sig}}^{\mathrm{fit}}}^2 + \left(\frac{2N_{\mathrm{sig}}^{\mathrm{fit}}/\epsilon_{\mathrm{sig}}}{(N_{\mathrm{norm}}^{\mathrm{fit}})^2/\epsilon_{\mathrm{norm}}}\right)^2 \sigma_{N_{\mathrm{norm}}^{\mathrm{fit}}}^2}. \tag{6.2}$$
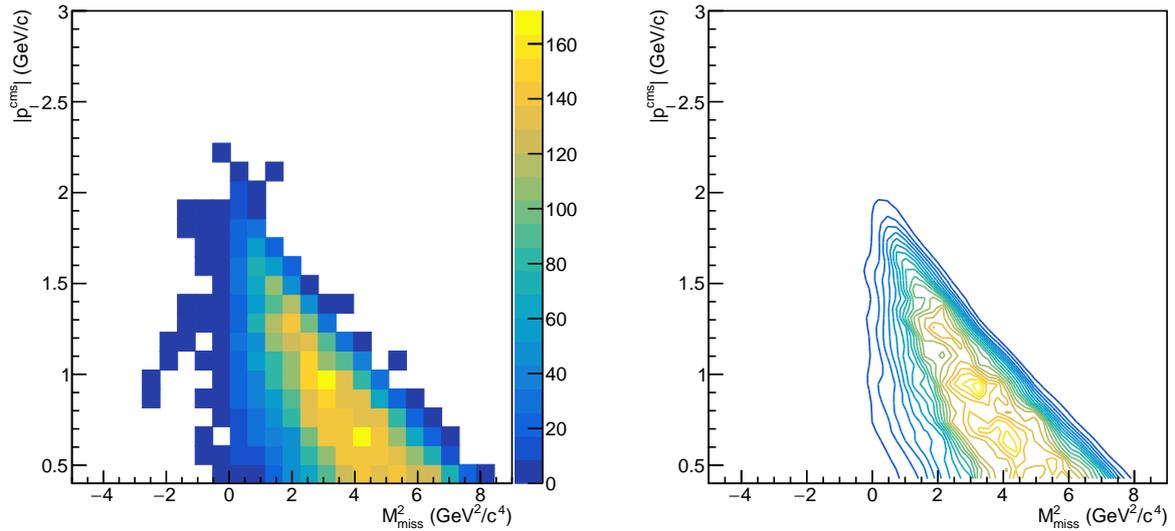
(a) Signal component
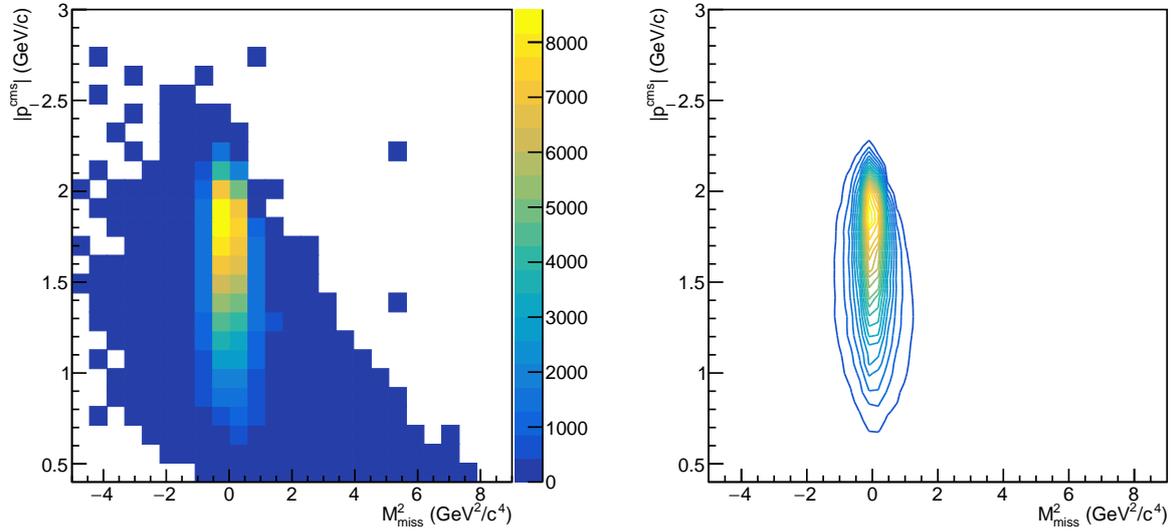
(b) Normalisation component
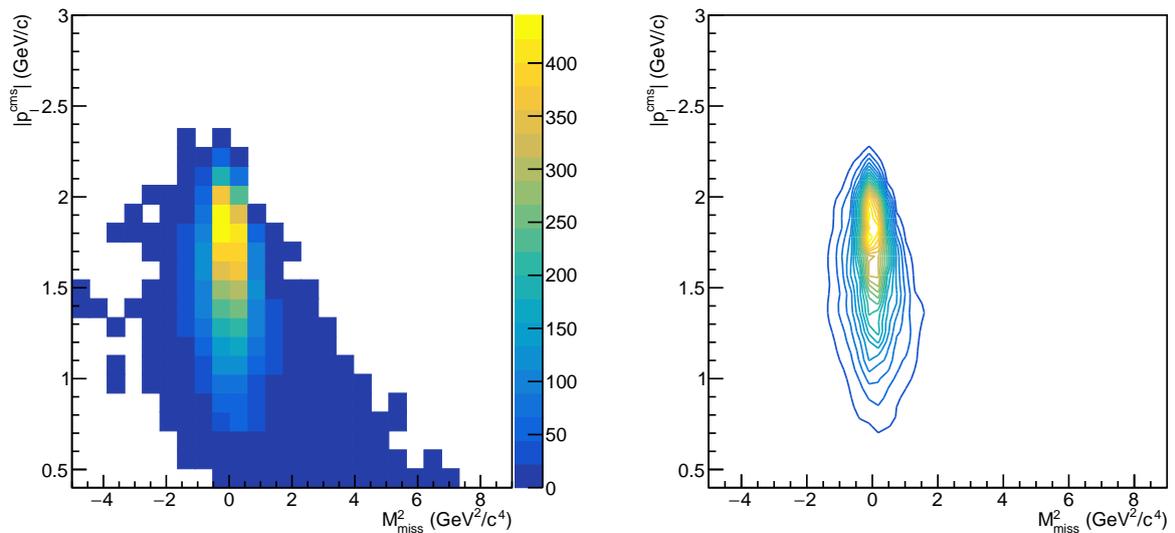
(c) D^{*0} component

Figure 6.1: Number of events in the $1044\,\mathrm{fb}^{-1}$ PDF shape sample (left) and corresponding PDF shape generated, separated by fit component. Since PDFs are normalised, their scales are arbitrary.
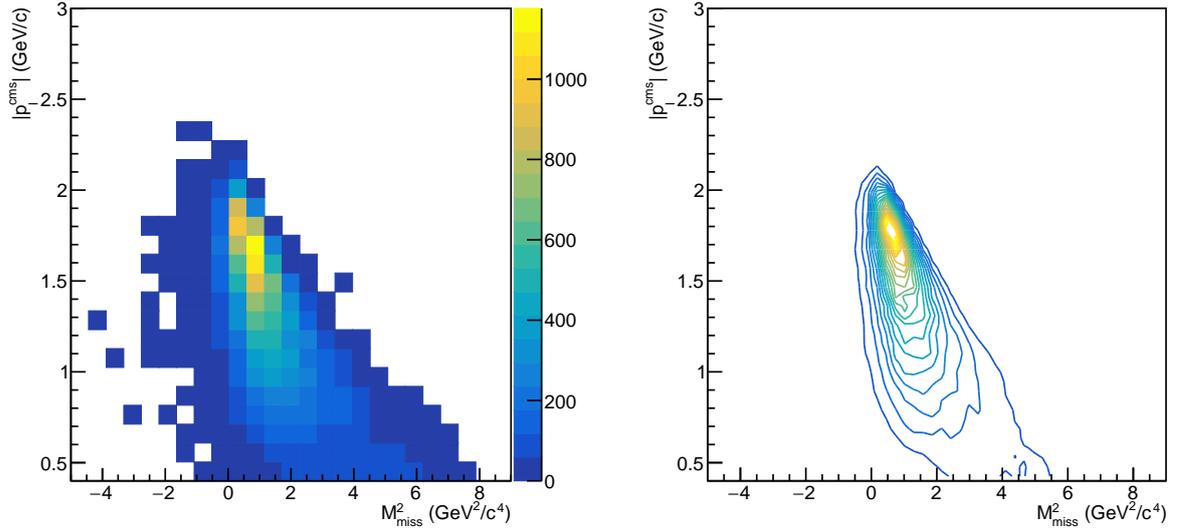
(a) D^{**} component
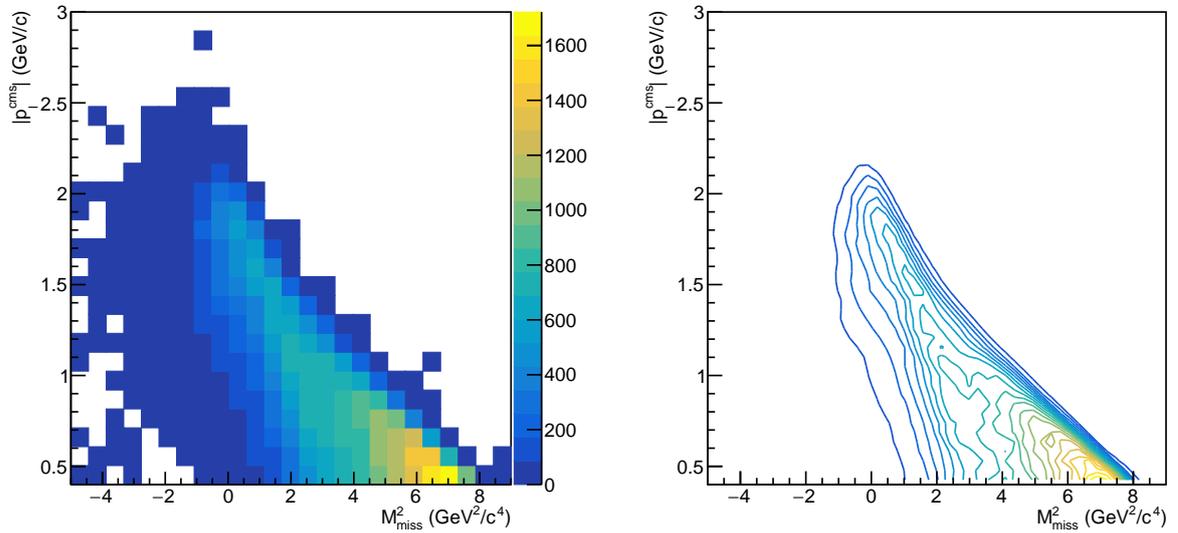
(b) "Rest" component

Figure 6.2: Number of events in the $1044\,\mathrm{fb}^{-1}$ PDF shape sample (left) and corresponding PDF shape generated, separated by fit component. Since PDFs are normalised, their scales are arbitrary.

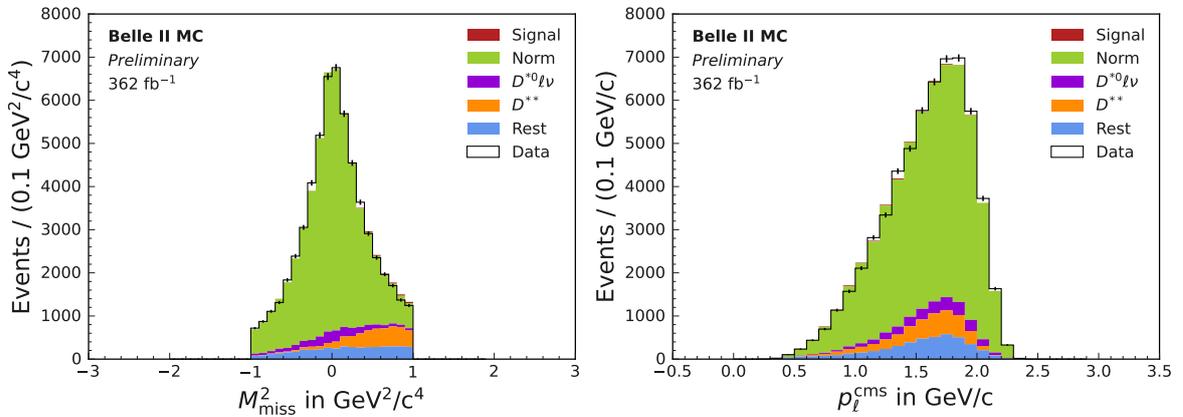

Figure 6.3: Data-MC agreement of M^2_miss and p^cms_ℓ in the $|M^2_\mathrm{miss}| < 1\,\mathrm{GeV}^2/\mathrm{c}^4$ sideband region after PID corrections. The total number of MC events is scaled to match the one in data.

6.1 Main fit result

The "main" fit is performed purely on MC, which does not need any corrections. The N_{D*0} scaling factor is 0.052779 and the fitted yields are

Component	Fitted yield	True yield
Signal	2662 ± 477	3293
Normalisation	66325 ± 302	66615
D^{*0}	3501	3484
D^{**}	11101 ± 294	10225
Rest	25840 ± 499	25810

Table 6.1: Fitted and the true yields in the $400 \, \mathrm{fb}^{-1}$ fit sample for the main fit result.

with the covariance matrix given in Table 6.2 and the correlation matrix in Table 6.3.

	N_{D**}	N_{norm}	N_{rest}	N_{sig}
N_{D**}	86300	-35450	15680	-53560
N_{norm}	-35450	90990	-29250	35240
N_{rest}	15680	-29250	249100	-208200
N_{sig}	-53560	35240	-208200	227400

Table 6.2: The covariance matrix of the main fit result.

	N_{D**}	N_{norm}	N_{rest}	N_{sig}
N_{D**}	1	-0.4001	0.1070	-0.3823
N_{norm}	-0.4001	1	-0.1943	0.2450
N_{rest}	0.1070	-0.1943	1	-0.8749
N_{sig}	-0.3823	0.2450	-0.8749	1

Table 6.3: The correlation matrix of the main fit result.

This yields

$$R(D^*)^{\mathrm{fit}}_{\mathrm{main}} = 0.2095 \pm 0.0375 \tag{6.3}$$

which lies 1.29σ below the true MC value of 0.2580. Noticeably, the fitted signal yield is off by almost 20%, and the true number lies outside the statistical fit uncertainty, while the D^{**} component is increased. However, from the correlation matrix one can see that the signal yield has a strong correlation to the number of "rest" events, which is estimated by the fit to good accuracy, being off by only 30 events with a statistical uncertainty of almost 500 events. The fit's attribution of events to the five components is therefore more involved. Interestingly, a toy study (described and discussed in Section 6.4) found the fit to be unbiased, with the signal and "rest" components being described particularly well, hinting at the deviation from the true yields being rooted in an unfortunate fluctuation. Projections of the fitted M^2_{miss} and p^{cms}_ℓ distributions and their pulls

$$g(x_i) = \frac{N^{\mathrm{true}}(x_i) - N^{\mathrm{fit}}(x_i)}{\sqrt{N^{\mathrm{true}}(x_i)}} \tag{6.4}$$
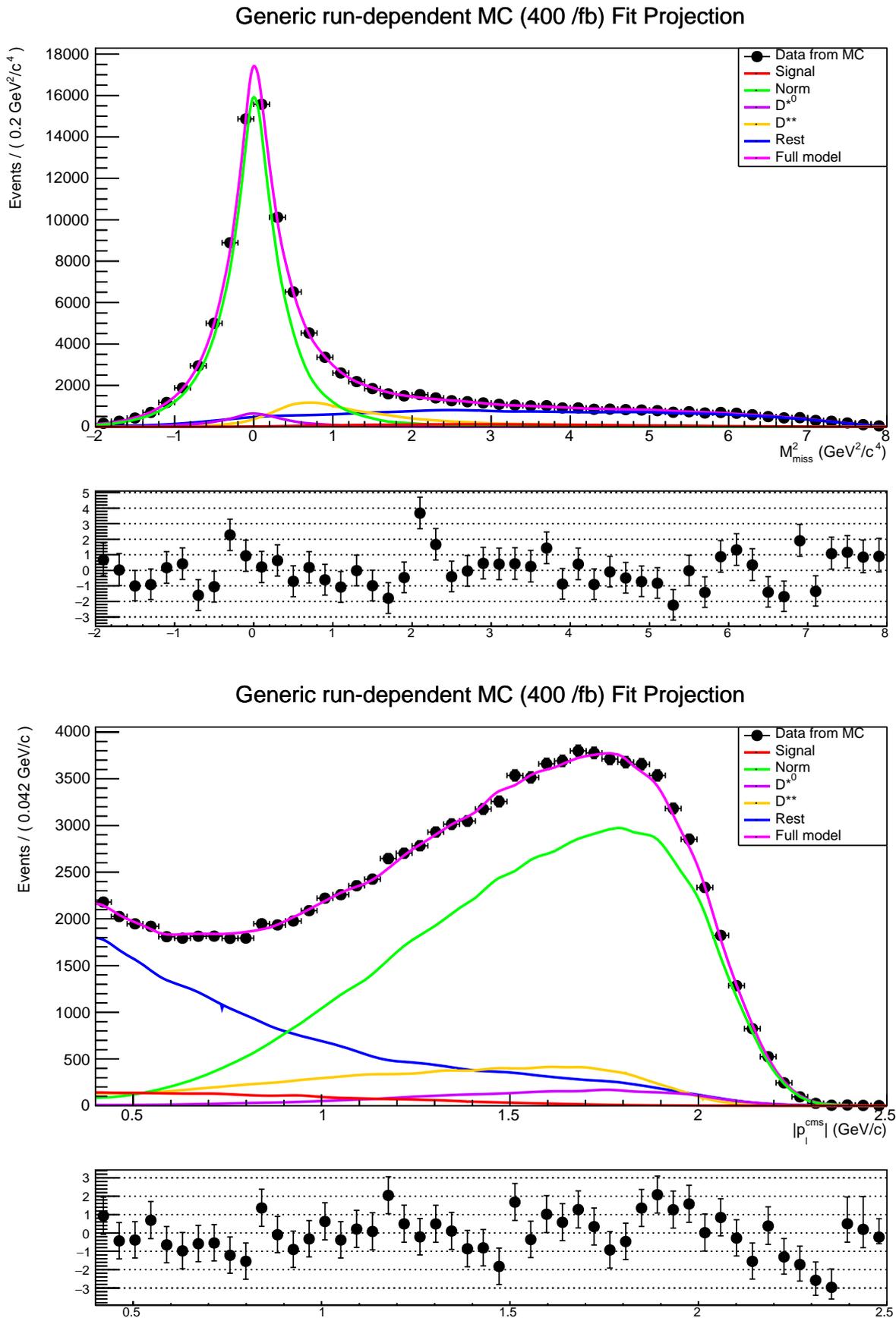
for bin x_i are plotted in Figure 6.4.

Figure 6.4: Fitted M_{miss}^2 and p_ℓ^{cms} distributions and their pulls for the main fit result.

6.2   Effect of scaling the kernel bandwidth

A smoothing factor ρ can be passed to the constructor of `RooNDKeysPdf`, used to create the PDF shapes, modifying the bandwidth of the kernels in the KDE in order to better follow rapid changes or smooth out fluctuations by applying an overall scaling factor. The default value is $\rho = 1.0$. As the main fit result does not reproduce the true underlying branching fraction ratio used in the MC generation very well, it was checked whether varying ρ could improve the modelling. To gauge the effect, the fit was performed two more times with ρ set to 0.8 and 1.2, respectively. For $\rho = 1.2$, the fit result gets worse as expected: $R(D^*) = 0.1974 \pm 0.0379$, while for $\rho = 0.8$ it does in fact get slightly better: $R(D^*) = 0.2207 \pm 0.0368$. However, as one can see from the plots in Figure 6.5, while the PDF now describes the high p_ℓ^{cms} tail better, it also picks up many fluctuations, for example at p_ℓ^{cms} right above $1.5\,\mathrm{GeV/c}$. Since tuning ρ comes at the risk of introducing a systematic error [33], the improvement was deemed too small to change the chosen approach. However, a careful optimisation of ρ for the generation of the 2D PDF shapes may well be worth investigating in the future.

(a) Fitted M_{miss}^2 for $\rho = 0.8$.

(b) Fitted p_ℓ^{cms} for $\rho = 0.8$.

(c) Fitted M_{miss}^2 for $\rho = 1.2$.

(d) Fitted p_ℓ^{cms} for $\rho = 1.2$.
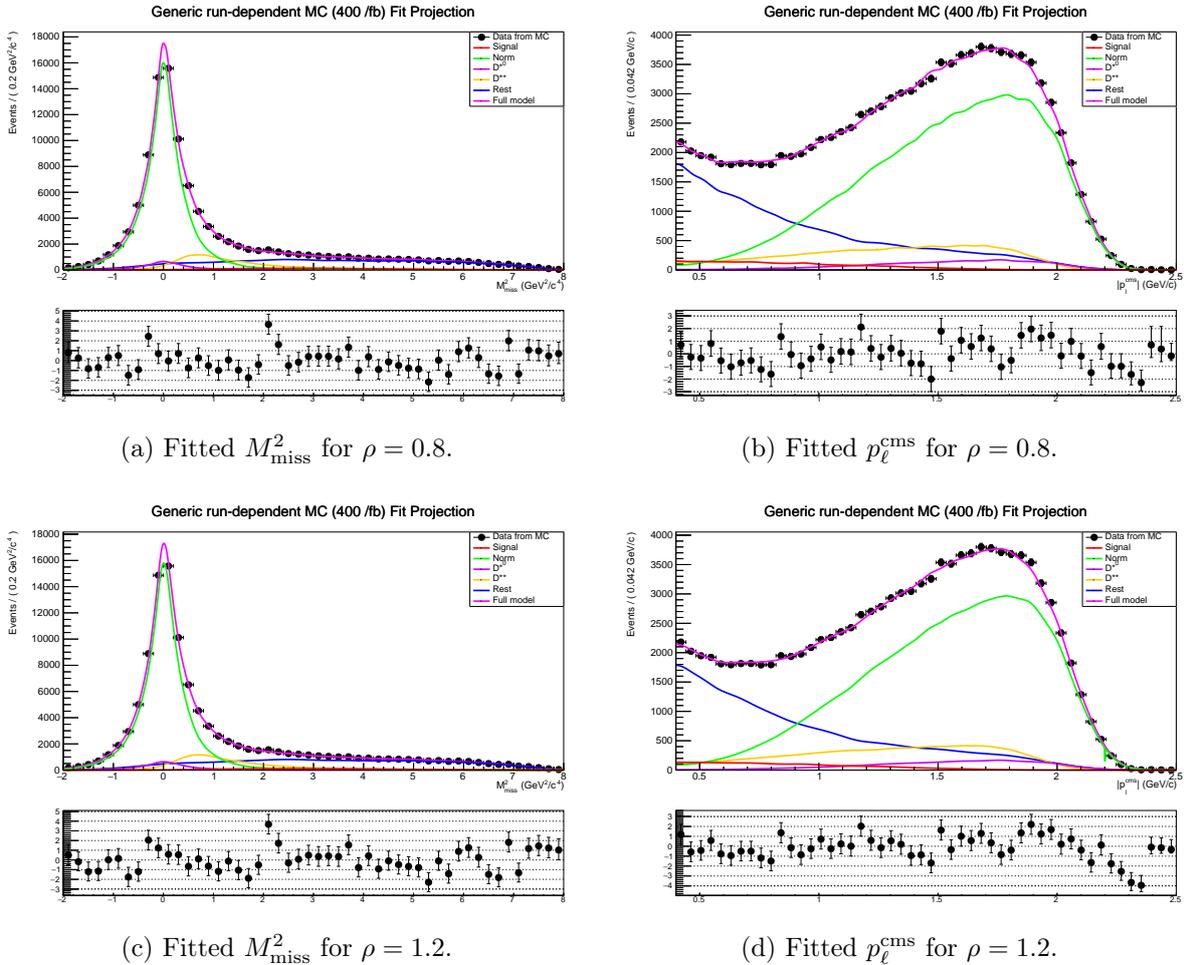
Figure 6.5: Fitted M_{miss}^2 and p_ℓ^{cms} distributions for modified kernel bandwidth scaling factor ρ affecting the smoothness of the PDF shapes.

6.3 Fully corrected fit result

Naturally, the goal is to use real data for the fit sample in the end, though this is not covered in this analysis. The PDF shape sample always consists of MC, in order to be able to separate events into the fit components. When using real data and MC together like this, corrections need to be applied to the MC events. The relevant corrections are introduced in Chapter 7 when their systematic uncertainties are estimated. Even though the estimation of systematic uncertainties was done on the previous main fit result, the fit was also performed with all corrections applied at once to both samples. This was done to better estimate how the fit would perform on actual data, the MC modelling of which is improved by the corrections, and to see the effect these corrections have. The fully corrected fit's N_{D*0} scaling factor is 0.052819 and the yields are

Component	Fitted yield	True yield
Signal	2822 ± 475	3197
Normalisation	64169 ± 291	64415
D^{*0}	3389	3372
D^{**}	9590 ± 282	8848
Rest	25658 ± 504	25792

Table 6.4: Fitted and the true yields in the $400\,\text{fb}^{-1}$ fit sample for the fully corrected fit result.

while the covariance matrix can be found in Table 6.5, with the correlation matrix in Table 6.6.

	N_{D**}	N_{norm}	N_{rest}	N_{sig}
N_{D**}	79760	-32820	11560	-47900
N_{norm}	-32820	84780	-27400	32690
N_{rest}	11560	-27400	253800	-209200
N_{sig}	-47900	32690	-209200	225900

Table 6.5: The covariance matrix of the fully corrected fit result.

	N_{D**}	N_{norm}	N_{rest}	N_{sig}
N_{D**}	1	-0.3994	0.0813	-0.3567
N_{norm}	-0.3994	1	-0.1877	0.2369
N_{rest}	0.0813	-0.1877	1	-0.8734
N_{sig}	-0.3567	0.2369	-0.8734	1

Table 6.6: The correlation matrix of the fully corrected fit result.

The resulting value of $R(D^*)$ is

$$R(D^*)_{\text{full}}^{\text{fit}} = 0.2291 \pm 0.0386 \tag{6.5}$$

which lies only 0.75σ below the MC truth, and the corresponding plots are shown in Figure 6.6. Both the signal and normalisation yield are now compatible with the true yields within their statistical uncertainties, as is the "rest" component. The number of D^{**} is again overestimated by the fit by almost 1000 events, though this does not affect the $R(D^*)$ result, which has moved closer to the true MC value.
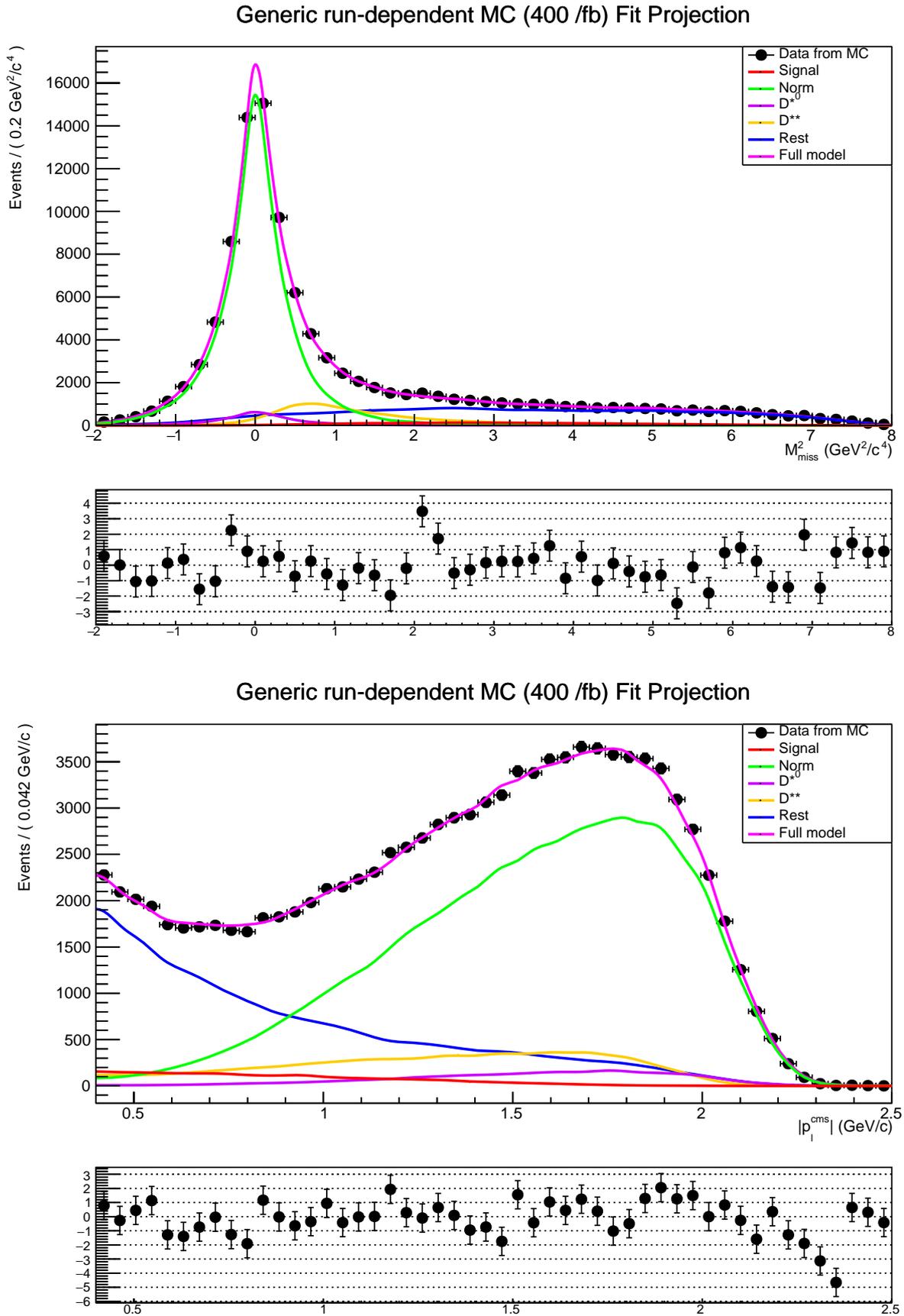
Figure 6.6: Fitted M^2_{miss} and p^{cms}_ℓ distributions and their pulls for the fully corrected fit result.

6.4 Toy study

To validate the fit, 1000 toy samples were generated from the PDF shapes and fitted using the original main fit's combined PDF shape model. The total number of events is modelled to follow a Poisson distribution with the mean corresponding to the total number of fitted events in Section 6.1, and therefore varies accordingly across the toy samples. The fitted yields of each toy sample are plotted in Figure 6.7, and their pulls are plotted in Figure 6.8. Both are approximately normally distributed, and a Gaussian is fitted to the pull distributions. For an ideal fit, the means of the pulls, which reveal the fit's bias, should be zero. Similarly, the standard deviations of the pulls should correspond to one, which would indicate that the statistical uncertainties determined by the fit are accurate. As one can see in Figure 6.8, these values look quite good, with the fit's pull means being compatible with zero within their respective uncertainty for both the signal and "rest" (here "Other background") fit components. However, the pull mean of the normalisation mode yield is underestimated by 4.3σ, and that of the D^{**} component is overestimated by 6.4σ, where σ is the uncertainty of the fitted Gaussians' means. The standard deviations of the fitted Gaussians look much better, with the signal and "rest" component again being compatible with the expected value within their uncertainty, while the standard deviation of the normalisation yield pull deviates from one by $+1.2\sigma$, and that of the D^{**} component lies just outside the $+1\sigma$ range, with σ being the uncertainty on the fitted Gaussians' standard deviations.

$R(D^*)$ only depends on the signal and normalisation yields, the first of which is described well, while the second is suppressed by the fact that the normalisation component's relative uncertainty is small to begin with. Hence the effect of the fit's bias on the result is negligible, especially compared to most of the uncertainties discussed in Chapter 7.
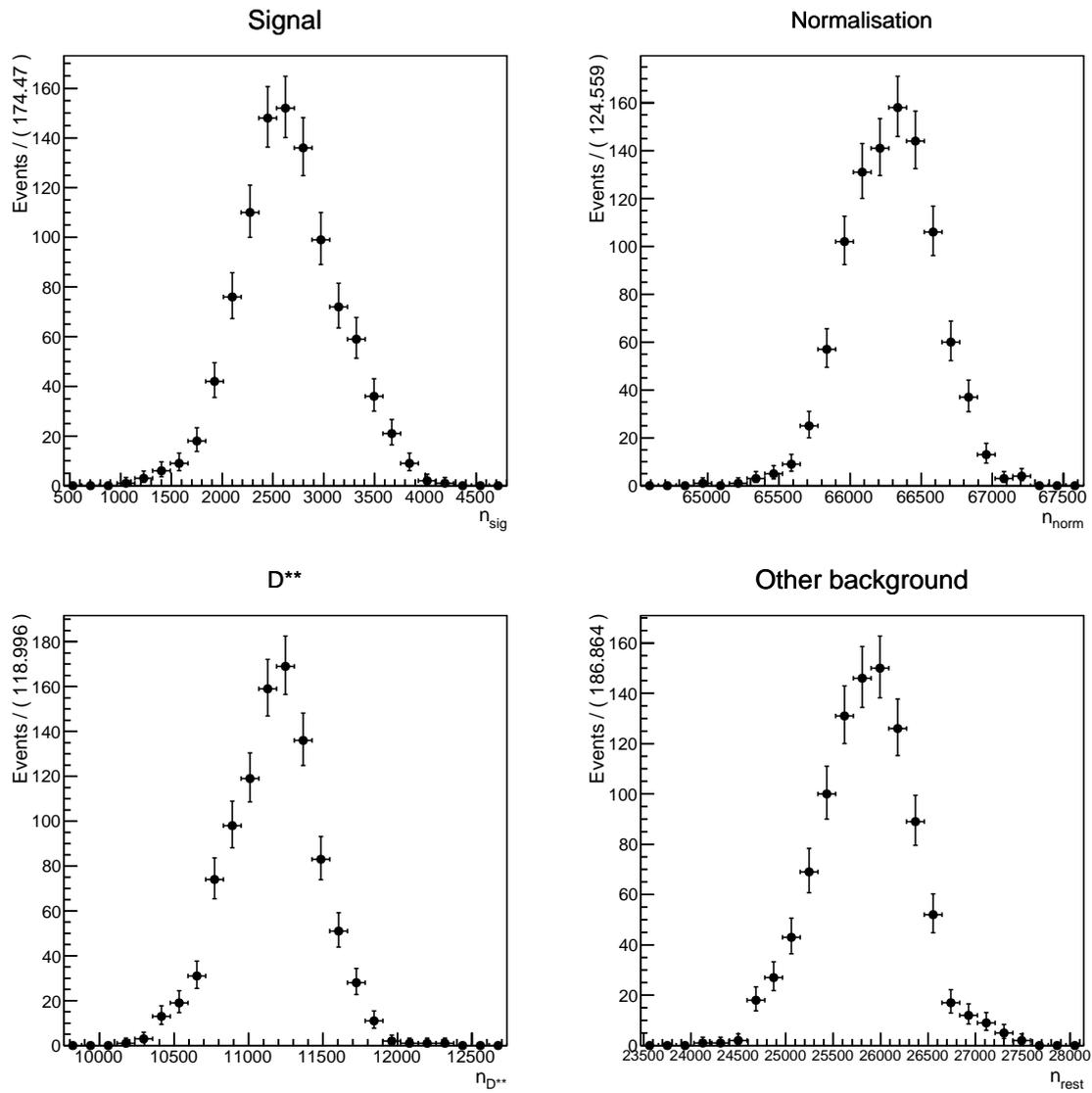
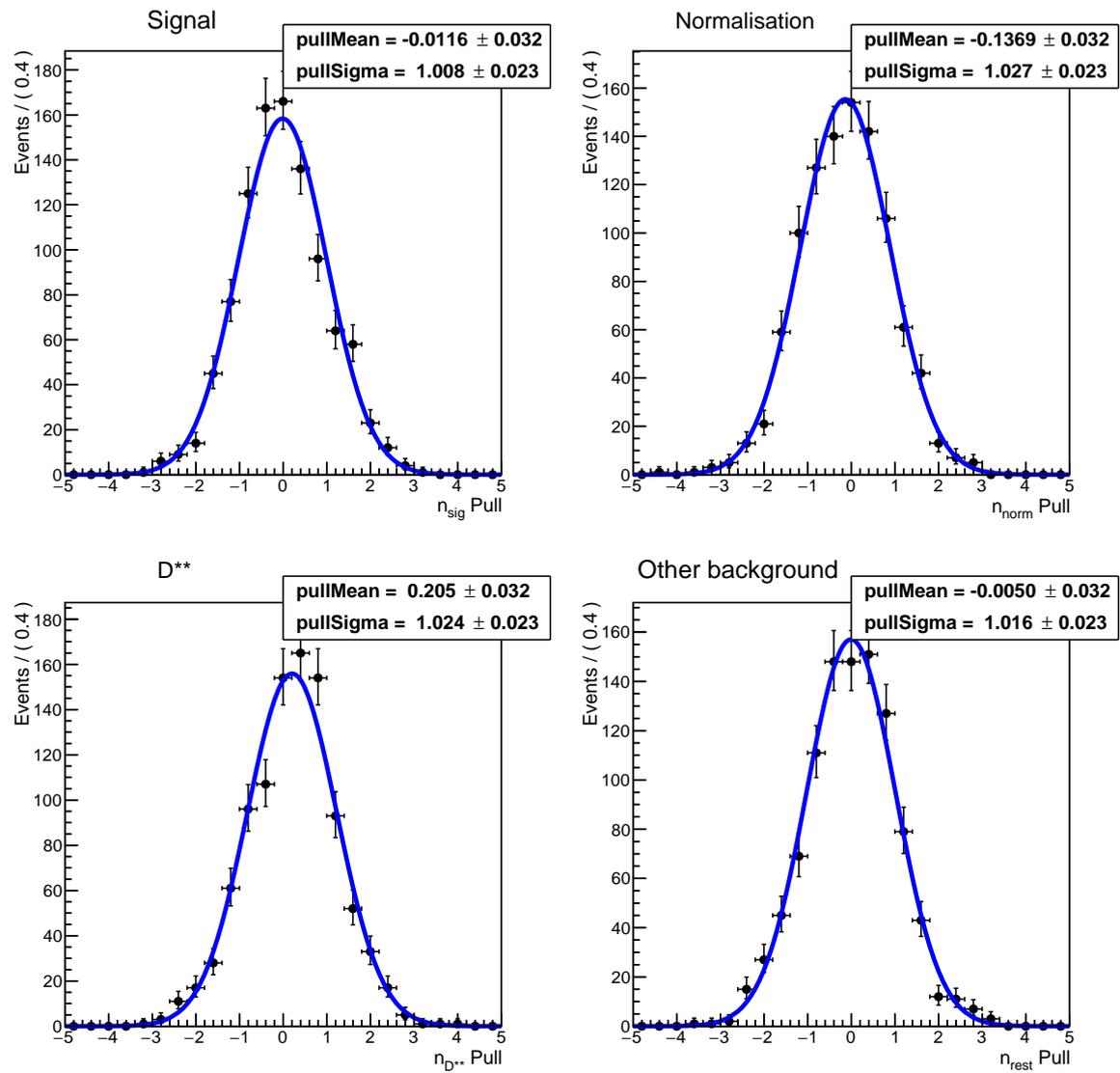Figure 6.7: Distributions of the fitted yields from 1000 toy samples generated from the PDF shapes.

Figure 6.8: Pull distributions of the fitted yields from 1000 toy samples generated from the PDF shapes.

Chapter 7

Uncertainty Estimation

Due to time constraints, corrections have been applied individually and their systematic uncertainties are determined by comparing variations to the respective nominal fit result where only that correction has been applied. To keep things clear, the following nomenclature is used: the fit result obtained in Section 6.1 is referred to as the main fit result, the reference fit result of each of the following studies is called the nominal fit result, and the fit result from Section 6.3 is called the fully corrected fit result.

7.1 PDF shapes

The uncertainty on the shapes of the fit components is estimated via resampling with replacement. This is done 1000 times by randomly selecting events from the original PDF shape sample until the so-called bootstrapped sample contains the same number of events – the difference arising from the fact that the same event can be selected multiple times. Then the unmodified fit sample is fitted using the PDF shapes generated from the bootstrapped samples. The results of the 1000 fits are plotted in Figure 7.1.
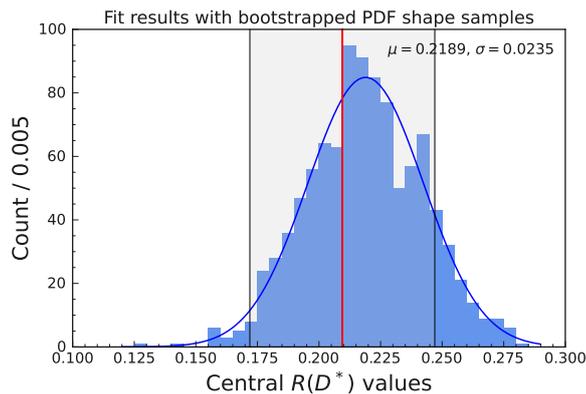
Figure 7.1: Fit results where bootstrapped PDF shape samples have been used for the PDF shape generation. The red line denotes the main (nominal) fit result, and the grey band the main result's statistical uncertainty range. A Gaussian is fitted to the fit results and the resulting standard deviation is used as the systematic uncertainty.

The distribution exhibits some features deviating from a Gaussian that may need further investigation in the future. Still, a Gaussian is fitted to the results and its standard deviation of 0.0235 is used as the systematic uncertainty from the PDF shape creation, corresponding to a relative uncertainty of 11.22%.

7.2   Lepton ID

Cutting on PID variables leads to a reduction in the selection efficiency which is not necessarily well-modelled in the MC and thus needs to be corrected. Similar corrections need to be applied on the efficiency with which an unwanted particle passes the PID cut – the misidentification probability. The fraction of misidentified particles among the candidates is called a fake rate [34]. A correction weight hence depends on both the true particle type and that as which it was reconstructed, as well as both the laboratory frame momentum and polar angle of the particle in question which determine the performance of the PID cut. Correction weights are calculated centrally within the collaboration (see [35]), using high purity samples obtained from very clean decays that do not need cuts on PID variables.

To estimate the systematic uncertainty arising from the lepton ID efficiency and fake rate corrections, 1000 variations of the corresponding weights were generated for both the electron and muon ID cuts at the same time using the `PIDvar` package [36]. Correlations between bins are automatically taken care of by `PIDvar`. Each weight was applied on the sample used for the PDF shape generation (also affecting the efficiencies used in the $R(D^*)$ calculation, see Eq. 6.1), while the nominal lepton ID weight was used for the fit sample. The results of the 1000 fits with varied PDF sample weights are plotted in Figure 7.2. Using the nominal lepton ID weight on both samples yields $R(D^*) = 0.2225 \pm 0.0382$ as indicated by the red line and grey band in the plot below.

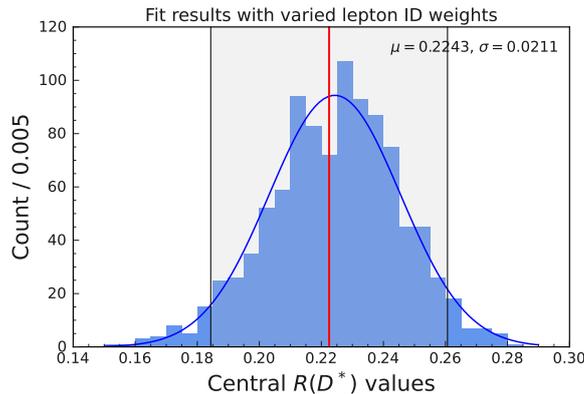
Figure 7.2: Fit results where varied lepton ID weights were applied to the sample used for PDF shape generation. The red line denotes the fit result obtained when using the nominal lepton ID weights, and the grey range indicates the statistical uncertainty on the nominal fit. A Gaussian is fitted to the fit results and the resulting standard deviation is used as the systematic uncertainty. The statistical fit uncertainty fully covers the systematic uncertainty introduced by the lepton ID corrections.

The standard deviation of the Gaussian fitted to the 1000 fit results is taken to be the systematic uncertainty introduced by the lepton ID corrections. As such, the uncertainty is determined to be 0.0211, which corresponds to a relative uncertainty of 9.50%.

Figures 7.3 and 7.4 contain the coverage plots showing for which regions in the laboratory frame lepton momentum and polar angle lepton ID corrections are available, overlaid on top of the distributions of events in the PDF shape sample. There are only a few events outside the covered regions – these are not assigned a correction weight and instead continue to be counted with a weight of one.

(a) Coverage plot for e^+ efficiency.

(b) Coverage plot for e^- efficiency.

(c) Coverage plot for μ^+ efficiency.

(d) Coverage plot for μ^- efficiency.

(e) Coverage plot for π^+ faking e^+.

(f) Coverage plot for π^- faking e^-.
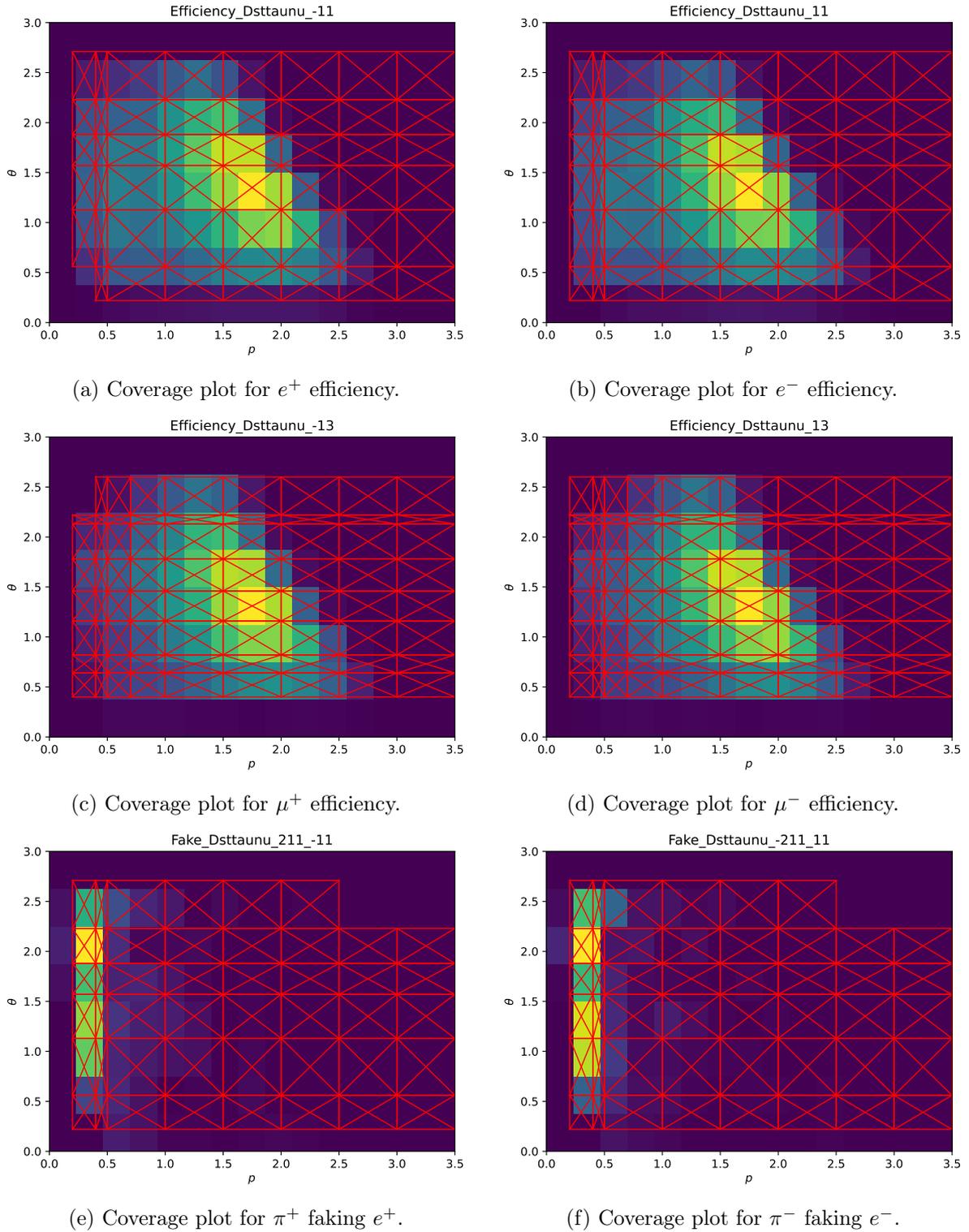
Figure 7.3: Lepton ID coverage plots showing events from the $1044\,\mathrm{fb}^{-1}$ PDF shape sample. p and θ are the laboratory frame lepton candidate momentum and polar angle in GeV/c and rad, respectively. The red boxes mark the covered regions for which lepton ID corrections are available. Created using the `PIDvar` package [36].

(a) Coverage plot for π^+ faking μ^+.

(b) Coverage plot for π^- faking μ^-.

(c) Coverage plot for K^+ faking e^+.

(d) Coverage plot for K^- faking e^-.

(e) Coverage plot for K^+ faking μ^+.

(f) Coverage plot for K^- faking μ^-.

Figure 7.4: Lepton ID coverage plots showing events from the $1044\,\mathrm{fb}^{-1}$ PDF shape sample. p and θ are the laboratory frame lepton candidate momentum and polar angle in GeV/c and rad, respectively. The red boxes mark the covered regions for which lepton ID corrections are available. Created using the `PIDvar` package [36].
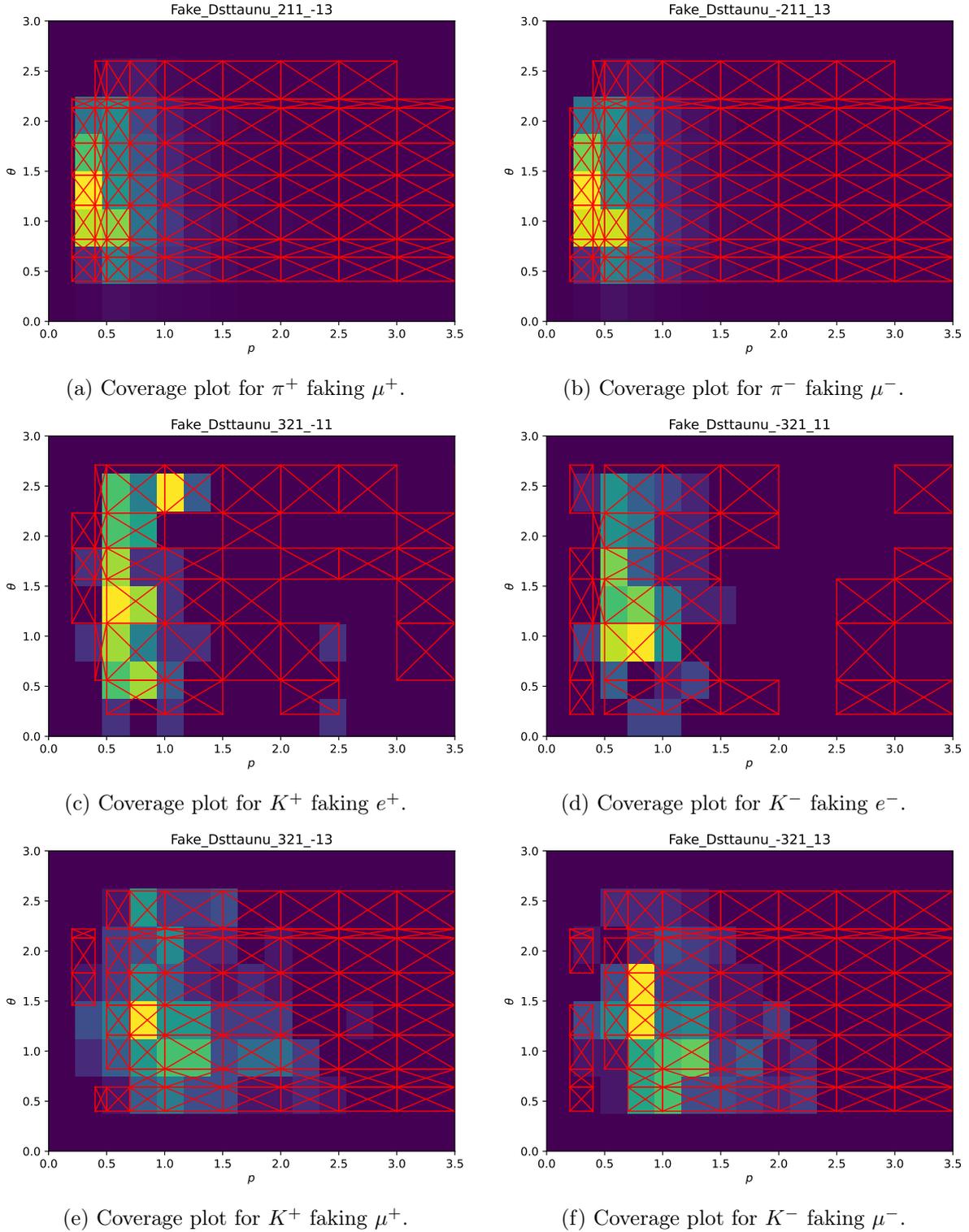
7.3 Slow pion efficiency

In analogy to the method used for the lepton ID systematic uncertainty, 1000 slow pion efficiency weight variations were created by sampling from a multinomial Gaussian distribution described by the statistical (uncorrelated and correlated) and systematic uncertainties of the weights. The used weights and their uncertainties are listed in Table 7.1. The correlation of the correlated statistical uncertainties was assumed to be 100%, the same was assumed for the systematic uncertainties. The varied weights were again applied to the sample used for PDF shape generation, also affecting the efficiencies, while the nominal correction weights were applied to the fit sample. The results can be seen in Figure 7.5. Using the nominal weight on both samples yields $R(D^*) = 0.2096 \pm 0.0375$.

p [GeV/c]	Correction	Stat. (uncorr.)	Stat. (corr.)	Syst.
0.05–0.12	0.996	0.022	0.015	0.003
0.12–0.16	0.990	0.017	0.015	0.003
0.16–0.20	0.987	0.019	0.015	0.003

Table 7.1: Momentum-dependent slow pion correction weights and statistical (uncorrelated and correlated) and systematic uncertainties [37]. p is the laboratory frame π_{slow}^- momentum. Events outside the given momentum ranges are given a weight of 1.
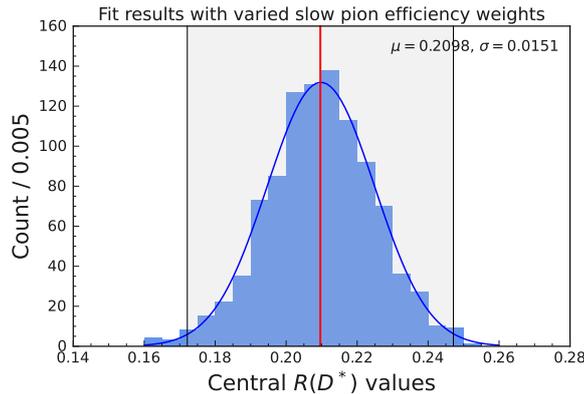

Figure 7.5: Fit results where varied slow pion efficiency weights were applied to the sample used for PDF shape generation. The red line denotes the fit result obtained when using the nominal slow pion efficiency weights, and the grey range indicates the statistical uncertainty on the nominal fit. A Gaussian is fitted to the fit results and the resulting standard deviation is used as the systematic uncertainty. Again the statistical fit uncertainty fully covers the systematic uncertainty introduced by the slow pion efficiency.

The standard deviation of the fitted Gaussian is used as the systematic uncertainty which is found to be 0.0151, corresponding to 7.21%.

7.4 $B \to D^{**}\ell^+\nu_\ell$ and gap mode branching fractions

As the four $B \to D^{**}\ell^+\nu_\ell$ branching fractions used in the Monte Carlo simulation are not the most recent, they should be corrected to the newer values provided by the *Heavy Flavor Averaging Group* (HFLAV). Notably, different weights for the four decay modes will affect the PDF shape of the D^{**} component. It is helpful to assume isospin symmetry between neutral

and charged B meson decays, relating their branching fractions via their lifetimes τ_B as

$$\mathcal{B}(B^0 \to X^0) \cdot \frac{\tau_{B^+}}{\tau_{B^0}} = \mathcal{B}(B^+ \to X^+) \tag{7.1}$$

to increase the statistics of the available branching fraction measurements [38]. Since these branching fractions are difficult to measure, they come with large uncertainties.

There are also the so-called *gap modes*, which are introduced to fill the "gap" between the measured inclusive semileptonic B decay branching fraction and the sum of all exclusive branching fractions measured so far. As such, they have not been measured yet and are merely a guess as to what further kinds of B decays might have contributed to the inclusive measurement. In the nominal Belle II MC, the gap is filled with (unmeasured) $B \to D^{(*)}\eta\ell^+\nu_\ell$ decays. Due to their guess-based nature, the gap mode branching fractions are assigned an uncertainty of 100%.

The most up-to-date $B \to D^{**}\ell^+\nu_\ell$ and gap mode branching fractions can be found in Table 7.2 along with their uncertainties.

Decay	$\mathcal{B}(B^+)$ [10^{-3}]	$\mathcal{B}(B^0)$ [10^{-3}]
$B \to D_1\ell^+\nu_\ell$	6.6322 ± 1.0894	6.1638 ± 1.0127
$B \to D_0^*\ell^+\nu_\ell$	4.2000 ± 0.7500	3.9033 ± 0.6972
$B \to D_1'\ell^+\nu_\ell$	4.2000 ± 0.9000	3.9033 ± 0.8366
$B \to D_2^*\ell^+\nu_\ell$	2.9337 ± 0.3248	2.7265 ± 0.3020
$B \to D\eta\ell^+\nu_\ell$	3.7700 ± 3.7700	4.0920 ± 4.0920
$B \to D^*\eta\ell^+\nu_\ell$	3.7700 ± 3.7700	4.0920 ± 4.0920

Table 7.2: Isospin averaged $B \to D^{**}\ell^+\nu_\ell$ and gap mode branching fractions [38].

The systematic uncertainty introduced by the uncertainty on the $B \to D^{**}\ell^+\nu_\ell$ and gap mode branching fractions is estimated separately for each decay mode by first applying a correction weight on both the PDF shape and fit sample and performing the fit, and then increasing/reducing the weight by its uncertainty, applying the new weight on the PDF shape sample while still using the nominal branching fraction weight on the fit sample and again performing the fits. In order to obtain symmetric uncertainties, the varied fit result with the larger absolute deviation is chosen to determine the systematic uncertainty. The gap modes are varied simultaneously as one decay mode. The results can be found in Table 7.3.

Decay	$R(D^*)^{\text{fit}}_{\text{nominal}}$	$R(D^*)^{\text{fit}}_{\text{down}}$	$R(D^*)^{\text{fit}}_{\text{up}}$	Abs. uncert.	Rel. uncert. [%]
$B \to D_1\ell^+\nu_\ell$	0.21049	0.19831	0.22097	0.01218	5.79
$B \to D_0^*\ell^+\nu_\ell$	0.20951	0.21016	0.20885	0.00066	0.31
$B \to D_1'\ell^+\nu_\ell$	0.20972	0.21624	0.20416	0.00653	3.11
$B \to D_2^*\ell^+\nu_\ell$	0.21251	0.21621	0.20889	0.00370	1.74
$B \to D^{(*)}\eta\ell^+\nu_\ell$	0.21140	0.32352	0.11339	0.11212	53.03
$\tau^+ \to \ell^+\nu_\ell\bar{\nu}_\tau$	0.20958	0.21006	0.20913	0.00047	0.23

Table 7.3: Fit results for the corrected and varied $B \to D^{**}\ell^+\nu_\ell$ and gap mode branching fractions. The absolute uncertainty is simply the difference between the respective nominal fit result and the varied fit result that deviates further from it. The relative uncertainty is calculated relative to the main fit result. Also contains the τ decay systematic uncertainty discussed in Section 7.6.

To get the combined systematic uncertainty from the four $B \to D^{**}\ell^+\nu_\ell$ branching fractions excluding the gap modes, the individual uncertainties are added in quadrature, yielding a rel-

ative uncertainty of 6.80%. The systematic uncertainty coming from the gap mode branching fraction is unreasonably large, especially compared to the previous analysis [12], and needs further investigation. The effect of the varied branching fractions on the PDF shapes of the fit components can be seen in Figure 7.6.
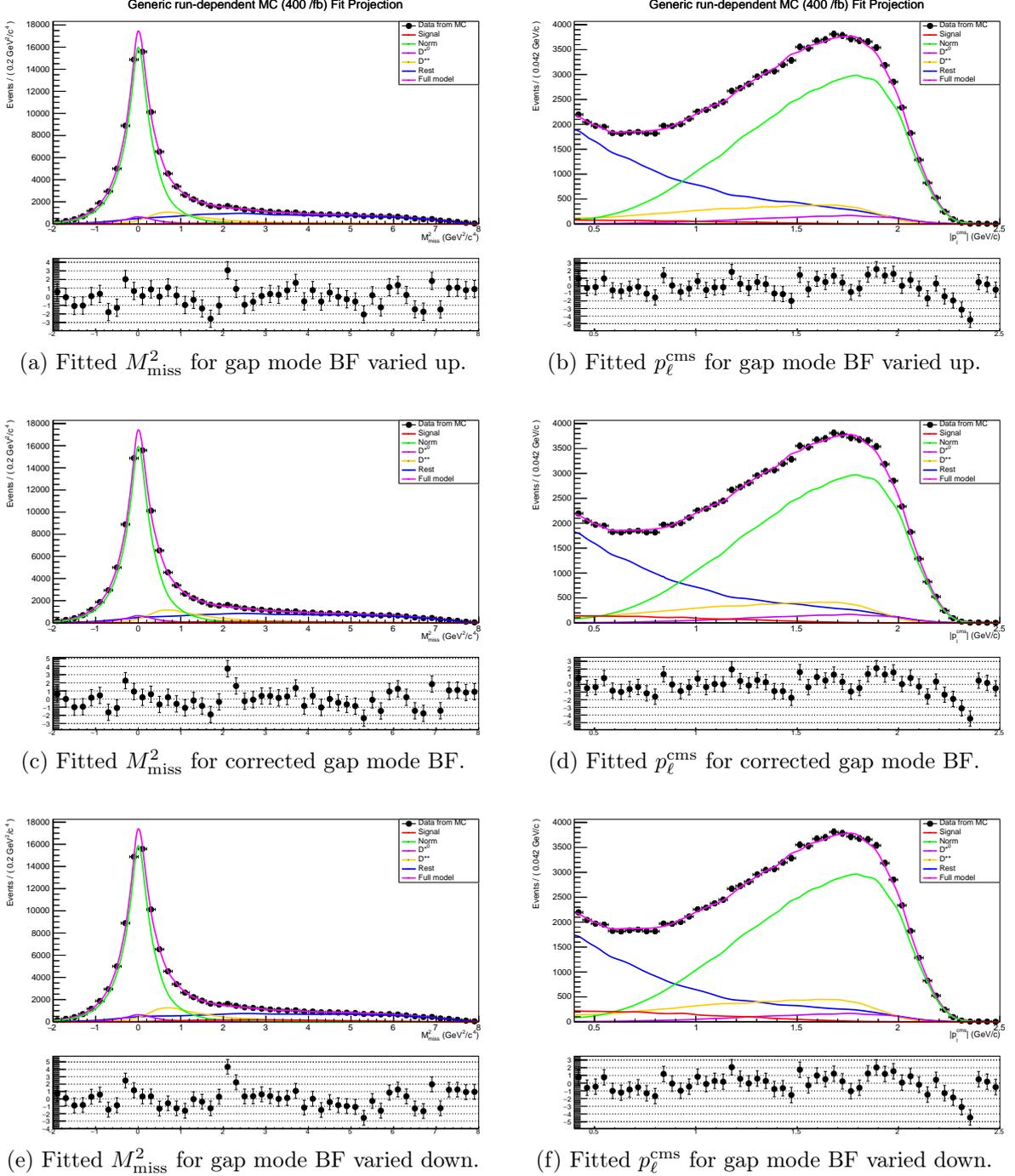

(a) Fitted M_{miss}^2 for gap mode BF varied up.

(b) Fitted p_ℓ^{cms} for gap mode BF varied up.

(c) Fitted M_{miss}^2 for corrected gap mode BF.

(d) Fitted p_ℓ^{cms} for corrected gap mode BF.

(e) Fitted M_{miss}^2 for gap mode BF varied down.

(f) Fitted p_ℓ^{cms} for gap mode BF varied down.

Figure 7.6: Fitted M_{miss}^2 and p_ℓ^{cms} distributions for corrected and varied gap mode branching fractions (BFs).

As the large deviation of the gap mode branching fraction-varied fit results came from the fit shuffling events between signal, D^{**} and "rest" (the gap mode M_{miss}^2 and p_ℓ^{cms} distributions can be modelled through a combination of signal and D^{**}), and the used gap modes are semileptonic B decays as are the D^{**} modes, an attempt was made to add the gap modes to the D^{**} component and doing the variation. However, this did worsen the results further.

7.5 Statistical uncertainty on the efficiency

Since events are selected with a probability $\epsilon = N_{\text{sel}}/N_{\text{gen}}$ out of a fixed number of generated events N_{gen}, the number of selected events N_{sel} is expected to follow a binomial distribution. The statistical uncertainty is therefore described by the binomial error:

$$\sigma_{N_{\text{sel}}} = \sqrt{N_{\text{gen}}\epsilon(1-\epsilon)}. \tag{7.2}$$

It follows for the efficiency ϵ:

$$\sigma_\epsilon = \frac{\sigma_{N_{\text{sel}}}}{N_{\text{gen}}} = \sqrt{\frac{\epsilon(1-\epsilon)}{N_{\text{gen}}}}. \tag{7.3}$$

The efficiencies and corresponding uncertainties obtained by inserting the numbers for the $1044\,\text{fb}^{-1}$ PDF shape sample used in the efficiency calculation are $\epsilon_{\text{sig}} = 0.003877 \pm 0.000042$ for signal and $\epsilon_{\text{norm}} = 0.010118 \pm 0.000024$ for normalisation mode events.

$R(D^*)$ is shifted the most if the two efficiencies are varied in opposite directions. This is done in both ways, and the larger deviation from the main fit result is used as the absolute systematic uncertainty arising from the statistical uncertainty on the efficiencies. The result is 0.00279, which corresponds to a relative uncertainty of 1.33%.

7.6 τ Decay branching fraction

As with the $B \to D^{**}\ell^+\nu_\ell$ branching fractions, the values for the τ decays used in the simulation are not the most up-to-date and should be corrected. The most recent results are $(17.82 \pm 0.04)\%$ for $\tau^+ \to e^+\nu_e\bar{\nu}_\tau$ and $(17.39 \pm 0.04)\%$ for $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$ [4]. The correction and uncertainty estimation is done in the exact same way as was done for $B \to D^{**}\ell^+\nu_\ell$ and the results have been added to Table 7.3 for convenience. To summarise however, the τ decay branching fraction systematic has been found to be 0.00047, corresponding to a relative uncertainty slightly below 0.23% where the effect on the efficiencies is included.

7.7 Tracking

As the distributions of the number of tracks on the signal/tag side are very similar for signal and normalisation mode events (see Figure 7.7), the systematic uncertainty arising from the tracking efficiency is expected to cancel in the ratio $R(D^*)$.

7.8 Summary of uncertainties

A summary of all uncertainties that have been considered can be found in Table 7.4. Compared to the previous analysis [12], one can see that the systematic uncertainty arising from the PDF shape generation was almost halved, coming down from 20% to 11% due to the larger sample used $(700\,\text{fb}^{-1} \to 1044\,\text{fb}^{-1})$. The biggest difference however is the gap mode, which previously led to a systematic uncertainty of only 5% in combination with $B \to D^{**}\ell^+\nu_\ell$ – this discrepancy is as yet unexplained. Furthermore, the uncertainty on the lepton ID corrections increased almost tenfold, from 1% to 9.50%, likely due to the way it was determined in the previous analysis. The statistical uncertainty on the efficiency was reduced from 2.7% to 1.33%, again profiting from the increased statistics of the PDF shape sample, while the τ
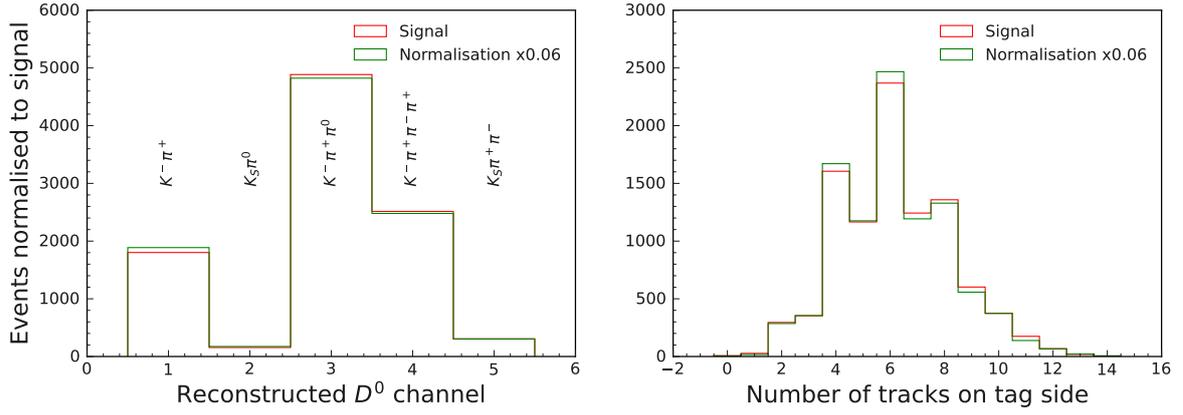
Figure 7.7: Comparison of the D^0 decay channels and the number of tracks on the tag side for signal and normalisation mode events. The number of tracks on the signal side is the decay mode-dependent number of charged D^0 daughters plus two (the lepton and the slow pion).

Source	Relative uncertainty [%]
Gap mode branching fractions	53.03
PDF shapes	11.22
Lepton ID	9.50
Slow pion efficiency	7.21
$B \to D^{**}\ell^+\nu_\ell$ branching fractions	6.80
Stat. uncertainty on the efficiency	1.33
τ Decay branching fraction	0.23
Tracking efficiency	−
Total systematic uncertainty (without gap mode)	17.78
Total systematic uncertainty (with gap mode)	55.94
Statistical uncertainty	17.92

Table 7.4: Summary of considered systematic uncertainties on $R(D^*)$. The total uncertainty is obtained by adding the individual uncertainties in quadrature. Also added is the relative statistical uncertainty on the main fit result.

decay branching fraction systematic remained unchanged. Newly included is the slow pion efficiency correction. The statistical uncertainty stayed approximately the same, despite more decay modes being included.

Chapter 8

Conclusion

To summarise, the $R(D^*)$ result measured in this analysis, using simulated data as input, is

$$R(D^*) = 0.2095 \pm 0.0375\,(\text{stat}) \pm 0.1172\,(\text{syst}) \qquad (8.1)$$

while the true value in the MC is 0.2580. A comparison with previous measurements can be found in Table 8.1.

Measurement	Result	Stat. uncert.	Syst. uncert.	Rescaled stat. uncert.
BaBar (2012)	0.332	0.024	0.018	0.052
Belle (2015)	0.293	0.038	0.015	0.106
Belle (2017)	0.270	0.035	0.028	0.097
Belle (2019)	0.283	0.018	0.014	0.050
LHCb (2022)	0.281	0.018	0.024	
LHCb (2023)	0.257	0.012	0.018	
Belle II (2023)	0.267	0.041	0.033	0.081
This analysis	0.210	0.038	0.117	0.075

Table 8.1: Statistical and systematic uncertainties of previous $R(D^*)$ measurements [7, 11] (in case of asymmetric uncertainties the larger value is listed) compared to the main fit result (Section 6.1) of this analysis. The rescaled statistical uncertainty is obtained by multiplication with $\sqrt{N_{B\bar{B}}/10^8}$, corresponding to a projection on a sample containing 100 million $B\bar{B}$ pairs.

Table 8.1 also lists the statistical uncertainties scaled by $\sqrt{N_{B\bar{B}}/10^8}$ to allow for a comparison independent of the sample size. These correspond to the statistical uncertainty projected on a sample size of 100 million $B\bar{B}$ events. LHCb is not really comparable in this regard as it is not a B factory operating at the $\Upsilon(4S)$ energy. It should also be noted that the Belle analysis from 2019 uses a semileptonic tag, leading to higher statistics and hence the smallest rescaled statistical uncertainty – this is also not directly comparable to this analysis due to the lepton veto employed in the tag side reconstruction.

Mainly due to the gap modes discussed in Section 7.4, this analysis suffers from a large systematic uncertainty compared to exclusive measurements. The uncertainty should be investigated in more detail in the future, especially since it is much larger than determined in the previous analysis done by Sofia Palacios Schweitzer [12]. However, numerous checks were performed to confirm the weighted fits were working as intended, and all D^0 decay modes listed in Table 4.3 (of which only the first was used in [12]) introduce the same fraction of gap mode events. Consequently, there is also a very real possibility that the gap mode branching fractions do indeed create such a large uncertainty. However, as the gap modes used are

merely a guess to explain the difference between inclusive and exclusive measurements, there is no guarantee that they accurately describe the real, as yet unobserved decays in the gap. In fact, this is unlikely as they are simulated using only the PHSP (phase space) decay model, which is not expected to accurately model the chosen $B \to D^{(*)}\eta\ell^+\nu_\ell$ decays. Using a more realistic model might decrease the sensitivity to the gap modes (as for example the systematic uncertainty arising from the four $B \to D^{**}\ell^+\nu_\ell$ modes, to which $B \to D^{(*)}\eta\ell^+\nu_\ell$ decays are expected to be similar, is small despite the relatively large uncertainties on the branching fractions), though ideally, the true content of the gap should be measured.

Even without the gap mode, the systematic uncertainty is still the largest at 0.0372 when compared to the previous analyses. The statistical uncertainty on the other hand is in fact better than in the majority of previous analyses, thanks to the inclusive reconstruction. In addition, a combined cut optimisation has not been performed yet, and as the multiclass classifier worked well on its own, it may be worth investigating further if e.g. a transformation of the output can be found that can reliably be fitted, providing further ways to reduce uncertainties. Further possible improvements could be the inclusion of more decay modes, as well as the already mentioned optimisation of the smoothness parameter ρ in the PDF shape generation. As such, the analysis is not finished yet (hence it is still blinded), and will hopefully shed more light onto the mystery of the $R(D^*)$ discrepancy in the future.

References

[1] M.J. Herrero, *The Standard Model*, arXiv:hep-ph/9812242.

[2] ATLAS Collaboration, *Observation of a new particle in the search for a Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716**, 1-29 (2012), arXiv:1207.7214 [hep-ex].
CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716**, 30-61 (2012), arXiv:1207.7235 [hep-ex].

[3] G. Bertone, D. Hooper, J. Silk, *Particle Dark Matter: Evidence, Candidates and Constraints*, Phys. Rep. **405**, 279-390 (2005), arXiv:hep-ph/0404175.

[4] R.L. Workman *et al.* (Particle Data Group), *The Review of Particle Physics*, Prog. Theor. Exp. Phys. **2022**, 083C01 (2022). For the online version see `https://pdg.lbl.gov/`.

[5] B. Aharmim *et al.* (SNO Collaboration), *Electron Energy Spectra, Fluxes and Day-Night Asymmetries of ^8B Solar Neutrinos from the 391-Day Salt Phase SNO Data Set*, Phys. Rev. **C72**, 055502 (2005), arXiv:nucl-ex/050202.

[6] S. Hirose *et al.* (Belle Collaboration), *Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^* \tau^- \bar{\nu}_\tau$ with one-prong hadronic τ decays at Belle*, Phys. Rev. D **97**, 012004 (2018), arXiv:1709.00129 [hep-ex].

[7] Y. Amhis *et al.* (Heavy Flavor Averaging Group Collaboration), *Averages of b-hadron, c-hadron and τ-lepton properties as of 2021*, Phys. Rev. D **107**, 052008 (2023), arXiv:2206.07501 [hep-ex]. Updated results taken from `https://hflav-eos.web.cern.ch/hflav-eos/semi/winter23_prel/html/RDsDsstar/RDRDs.html`, accessed 7.9.2023.

[8] J.P. Lees *et al.* (BABAR Collaboration), *Measurement of an excess of $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ decays and implications for charged Higgs bosons*, Phys. Rev. D **88**, 072012 (2013), arXiv:1303.0571 [hep-ex].

[9] M. Huschle *et al.* (Belle Collaboration), *Measurement of the branching ratio of $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)} \ell^- \bar{\nu}_\ell$ decays with hadronic tagging at Belle*, Phys. Rev. D **92**, 072014 (2015), arXiv:1507.03233 [hep-ex].

[10] G. Caria *et al.* (Belle Collaboration), *Measurement of $R(D)$ and $R(D^*)$ with a Semileptonic Tagging Method*, Phys. Rev. Lett. **124**, 161803 (2020), arXiv:1910.05864 [hep-ex].

[11] Belle II Collaboration, *A test of lepton flavour universality with a measurement of $R(D^*)$ using hadronic tagging at the Belle II experiment*, BELLE2-PUB-DRAFT-2023-015 (2023).

[12] S. Palacios Schweitzer, *Inclusive B-Meson Tagging for an $R(D^*)$ Measurement at Belle II*, Master's thesis, LMU 2022.

[13] T. Kuhr, C. Pulvermacher, M. Ritter *et al.*, *The Belle II Core Software*, Comput. Softw. Big Sci. **3**, 1 (2019), `https://doi.org/10.1007/s41781-018-0017-9`.

[14] F.U. Bernlochner, *Semitauonic b-hadron decays: A lepton flavor universality laboratory*, Rev. Mod. Phys. **94**, 015003 (2022), arXiv:2101.08326 [hep-ex].

[15] E. Kou *et al.*, *The Belle II Physics Book*, Prog. Theor. Exp. Phys. **2020**, 123C01 (2019), arXiv:1808.10567 [hep-ex].

[16] T. Abe *et al.*, *Belle II Technical Design Report*, KEK Report 2010-1 (2010), arXiv:1011.0352 [physics.ins-det].

[17] T. Hara, T. Kuhr, and Y. Ushiroda, *Belle II Coordinate System and Guideline of Belle II Numbering Scheme* (2011), accessed via `https://indico.mpp.mpg.de/event/2308/contributions/4092/attachments/3414/3799/Belle2NumberingScheme.pdf` on 8.9.2023.

[18] D.J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462**, 152-155 (2001).

[19] T. Sjöstrand *et al.*, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191**, 159 (2015), arXiv:1410.3012 [hep-ph].

[20] N. Davidson *et al.*, *PHOTOS Interface in C++; Technical and Physics Documentation*, Comput. Phys. Commun. **199**, 86 (2016), arXiv:1011.0937 [hep-ph].

[21] S. Agostinelli *et al.*, *Geant4–a simulation toolkit*, Nucl. Instrum. Meth. A **506**, 250 (2003).

[22] H. Junkerkalefeld, *Bremsstrahlung corrections: Recommendations for the `correctBremsBelle` module*, 41st Belle II General Meeting slides, Jan. 2022, `https://indico.belle2.org/event/6017/contributions/31516/attachments/14845/22307/2022.01.20_HJ_Bremsstrahlung.pdf`, accessed 1.9.2023.

[23] Belle II Software Group, *Belle 2 Software Documentation*, available at `https://software.belle2.org/light-2212-foldex/sphinx/index.html`, accessed 11.9.2023.

[24] P. Cheema, *Suppressing Beam Background and Fake Photons at Belle II using Machine Learning*, BELLE2-CONF-DRAFT-2023-021 (2023).

[25] P. Krishnan, M. Kumar, M. Nayak, K. Trabelsi, K_S^0 *reconstruction using phase 2 data*, BELLE2-NOTE-PH-2018-017.
For the cuts on the invariant mass one needs to look at the actual code: `https://stash.desy.de/projects/B2/repos/basf2/browse/analysis/scripts/stdV0s.py#72`. For completeness' sake, the `goodBelleKshort` criteria in Table 4.2 which are from the paper are implemented in `https://stash.desy.de/projects/B2/repos/basf2/browse/analysis/variables/src/BelleVariables.cc#39`. Both accessed 24.08.2023.

[26] Q. Zhou, *Correction for tracking momentum bias based on invariant mass peak studies*, BELLE2-NOTE-PH-2020-030 (2020).

[27] F. Pedregosa *et al.*, *Scikit-learn: Machine Learning in Python*, J. Mach. Learn. Res. **12**, 2825–2830 (2011).

[28] Ed. A.J. Bevan, B. Golob, Th. Mannel, S. Prell, and B.D. Yabsley, *The Physics of the B Factories*, Eur. Phys. J. C**74**, 3026 (2014), SLAC-PUB-15968, KEK Preprint 2014-3, arXiv:1406.6311 [hep-ex].

[29] F. Metzner, F. Bernlochner, P. Goldenzweig, M. Prim, I. Tsaklidis, $R(D^*)$ *with Leptonic τ and Hadronic FEI-Tag*, BN1616 (2023), BELLE2-NOTE-PH-2023-039-2.

[30] Plotly Technologies Inc., *Collaborative data science* (2015), Montréal, QC. Available at `https://plot.ly`.

[31] W. Verkerke, D. Kirkby, *The RooFit toolkit for data modeling*, arXiv:physics/0306116 [physics.data-an].

[32] `https://root.cern.ch/doc/master/Minuit2Page.html`, accessed 1.9.2023.

[33] K. Cranmer, *Kernel estimation in high-energy physics*, Comput. Phys. Commun. **136**, 198-207 (2001), arXiv:hep-ex/0011057.

[34] U. Tamponi, *From PID detectors to PID variables* [slides], Feb. 2020, `https://indico.belle2.org/event/1307/sessions/378/attachments/3070/5683/umberto_PIDTutorial_Feb2020.pdf`, accessed 13.9.2023.

[35] Belle II Collaboration, *Recommendations for Lepton ID – release 6*, DESY Confluence, `https://confluence.desy.de/display/BI/Recommendations+for+Lepton+ID+-+release+6`, accessed 13.9.2023.

[36] W. Sutcliffe *et al.*, `PIDvar` *– Offline Particle ID efficiency and fake rate corrections for Belle 2* [Software package], `https://stash.desy.de/users/sutclw/repos/pidvar/browse`.

[37] Belle II Collaboration, *Conference readiness*, DESY Confluence, `https://confluence.desy.de/display/BI/Conference+readiness`, accessed 1.9.2023.

[38] Belle II Collaboration, *Correction of BF($B \to X_c \ell \nu$) in generic MC*, DESY Confluence, `https://confluence.desy.de/pages/viewpage.action?pageId=202394372`, accessed 25.08.2023.

Selbständigkeitserklärung

Ich versichere hiermit, die vorliegende Arbeit mit dem Titel

$R(D^*)$-Messung mit inklusivem B-Mesonen-Tagging bei Belle II

selbständig verfasst zu haben und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben.

Stephanie Anna Helga Steinmetz

München, den 15. September 2023