Karlsruhe Institute of Technology

ETP-KA/2021-3

Study of $\mathrm{B} \rightarrow \mathrm{D}^{**}\ell\nu$ decays for a $\mathcal{R}(\mathrm{D}^{(*)})$ measurement at Belle

Patrick Ecker

Masterthesis

24th February 2021

Institute of Experimental Particle Physics (ETP)

Advisor:	Prof. Dr. Günter Quast
Coadvisor:	Prof. Dr. Florian Bernlochner

Editing time: 1st March 2020 – 24th February 2021

Karlsruhe Institute of Technology

ETP-KA/2021-3

Studien zu $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ Zerfällen für eine $\mathcal{R}(\mathrm{D}^{(*)})$ Messung bei Belle

Patrick Ecker

Masterarbeit

24. Februar 2021

Institut für Experimentelle Teilchenphysik (ETP)

Referent:	Prof. Dr. Günter Quast
Korreferent:	Prof. Dr. Florian Bernlochner

Bearbeitungszeit: 1. März 2020 – 24. Februar 2021

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, 24. Februar 2021

. . . .
(**Patrick Ecker**)

Contents

1. Introduction

What is physics? According to the Encyclopedia Britannica, physics is the science that deals with the structure of matter and the interactions between the fundamental constituents of the observable universe [1]. To do so, physics develops models to describe our universe. One of these models is the Standard Model of particle physics (SM), which has been developed in the second half of the 20th century. The standard model describes the observable particles in our universe, which are divided into fermions and bosons. There are three generations of fermions, containing leptons and quarks, that couple to gauge bosons, which are the force carriers of the fundamental interactions described by the standard model. But how can one prove if the standard model is correct? Simply said, one has to compare the results from an experiment to the predictions that can be calculated from the model, and like Richard Feynman once said: "If it [the model] disagrees with experiment, it is wrong." [2]

One of the assumptions the standard model makes is the lepton flavor universality. This assumes that the couplings of the gauge bosons do not depend on the flavor of the interacting lepton. Taking this assumption into account, one can calculate the branching fraction ratios of the semileptonic $\mathrm{B} \to \mathrm{D}^{(*)}\tau\nu_\tau$ decay to the $\mathrm{B} \to \mathrm{D}^{(*)}\ell\nu_\ell$ decay, with ℓ being one of the light leptons e or μ. These ratios, named $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$, are measured by multiple experiments resulting in different results [3–5]. In Fig. 1.1 a comparison of the measurements of $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ to the SM prediction is shown. The combined result show a $3.1\,\sigma$ deviation from the value expected from the SM. So is the standard model wrong and the lepton flavor universality broken? Not necessarily, since like Feynman said further:"...one has to check a little to make sure that it is wrong, because whoever did the experiment may have reported incorrectly, or there may have been some feature in the experiment that was not noticed, some dirt or something" [2]. To check whether the discrepancy reported by the previous $\mathcal{R}(\mathrm{D}^{(*)})$ measurements is actually there or caused by a systematic error in the measurement, a new $\mathcal{R}(\mathrm{D}^{(*)})$ measurement is ongoing at the time of writing this thesis. This $\mathcal{R}(\mathrm{D}^{(*)})$ analysis is based on the data set collected by the Belle experiment from 1999 to 2010.

Two of the main systematic uncertainties, which are reported for the previous measurements, are the limited knowledge about both the shapes and the branching fractions of the $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ decays. Here D^{**} denotes on of the four orbitally excited D mesons, D_0^*, D_1', D_1, and D_2^*. Most of the D^{**} decay into a D or D^* meson involving the emission of either a charged or neutral pion. To reduce the above mentioned uncertainties for the ongoing $\mathcal{R}(\mathrm{D}^{(*)})$ analysis, this thesis provides studies on the form factor modeling of the four different D^{**} states. Additionally, this thesis introduces a setup to measure the different $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$

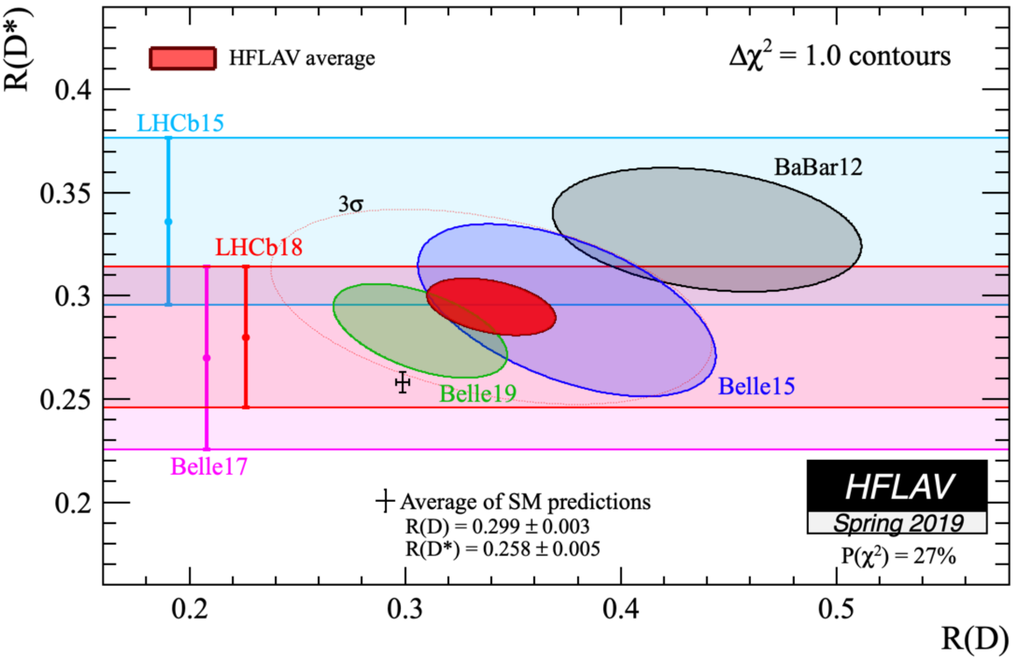


Figure 1.1.: Average of previous $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ measurements as provided by the HFLAV [7]. The current average (red circle) shows a 3.1 σ discrepancy to the prediction from the standard model (black cross).

branching fractions for each of the four different D^{**} states. Both the form factor study and the branching fraction measurement are expected to improve the understanding of the decay $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ and consequently reduce the systematic uncertainty on the underlying $\mathcal{R}(\mathrm{D}^{(*)})$ measurement. The analysis described in this thesis is based on the same data set as the underlying $\mathcal{R}(\mathrm{D}^{(*)})$ analysis.

In addition to the influence on the $\mathcal{R}(\mathrm{D}^{(*)})$ measurements, the semileptonic $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ decays also play an important role as a background component in untagged exclusive $|V_{cb}|$ matrix element measurements as well as for inclusive ones [6]. Thus, a better understanding of the $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ decay process also leads to more precise results for such measurements.

The first part of the thesis gives an overview on the experimental setup, with a description of the KEKB accelerator and the Belle experiment and a short introduction into the used software. Chapter 3 describes the available data sets, as well as the reconstruction from final state particles up to the $\Upsilon(4\mathrm{S})$ resonance, which was produced in the $\mathrm{e}^+\mathrm{e}^-$–collision, step-by-step. This chapter also covers the event selection criteria, which are applied both during the reconstruction of the events and afterwards to distinguish signal events, which are of interest for the D^{**} analysis, from background. The measured data and the Monte Carlo simulations (MC) show discrepancies, which are on the one hand caused by different reconstruction efficiencies in MC and data and on the other hand occur due to usage of outdated parameterizations for the MC production. Therefore, Monte Carlo corrections are applied as described in Chapter 4 to reduce the discrepancy between MC and data. Afterwards, the composition of the simulated MC is investigated in Chapter 5, while looking at observables which are of interest for the thesis. Chapter 6 explains the tools used for the extraction of the signal yields for both the $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ form factor study and the branching fraction measurements. The former study is explained in Chapter 7 while the latter is covered in Chapter 8. For each of the two studies the general concept is presented and a discussion of results based on Asimov data is shown. The studies are blinded at the time of writing this thesis and provide the basis for the measurements of the discussed properties based on the full Belle data set. In the final chapter a summary of the presented methods is given and also an outlook on planned improvements to the studies is provided.

2. Experimental Setup

The CP violation described by Kobayashi and Maskawa led to a whole new set of requirements for particle physics experiments. With the primary objective to study this CP violation, the Belle experiment was initiated in 1993. It was located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. During its runtime of over one decade, from 1999 to 2010, Belle was able to collect a total of $711\,\mathrm{fb}^{-1}$ of data on the $\Upsilon(4S)$ resonance [8]. The Belle experiment was in direct competition with the BarBar experiment, which had a similar physics program and was running at the same time.

The physics program, performed by both experiments, led to a lot of discoveries, with the highlight being the Nobel Prize for Kobayashi and Maskawa in 2008 for their prediction of the third quark generation.

In this chapter the main two components of the Belle experiment, the KEKB accelerator and the Belle detector, are described. A short outlook on its successor, the Belle II experiment, is also provided. The last part gives a short overview about the used software framework.

2.1. The KEKB Accelerator

In search of decays to test the idea of direct CP violation described by Kobayashi and Maskawa, B mesons seemed to be promising candidates. To provide the required large amount of B mesons, new colliders like KEKB and PEP-II, so-called B Factories, were built, which deliver B mesons in a clean environment.

A B Factory has to fulfill several requirements to make CP measurements on the $\Upsilon(4S)$ resonance possible. Due to the small branching fraction of the involved decays, a large amount of $\mathrm{B\overline{B}}$ pairs is needed ($\sim 30\,\mathrm{fb}^{-1}$). With an integrated luminosity of $771\,\mathrm{fb}^{-1}$ on the $\Upsilon(4S)$ resonance, this requirement was more than satisfied by the KEKB accelerator. To enable time-dependent CP violation measurements, the lifetime difference of the two B mesons has to be sufficiently long. Such a lifetime difference can be achieved by boosting the $\mathrm{B\overline{B}}$ pair. [9]

These requirements led to the design of the KEKB accelerator as an asymmetric $\mathrm{e^+e^-}$–collider with a circumference of about 3 km. It consists of two storage rings, which are built into the tunnel of the former TRISTAN accelerator [10]. After being accelerated in a linear accelerator (LINAC), the electrons are injected into the high energy ring (HER) with an energy of 8 GeV, while the positrons, which are produced by irridiating a tungsten plate, are discharged into the low energy ring (LER) with an energy of 3.5 GeV. The two beams collide at the interaction point (IP), where the Belle detector was located. A sketch of the

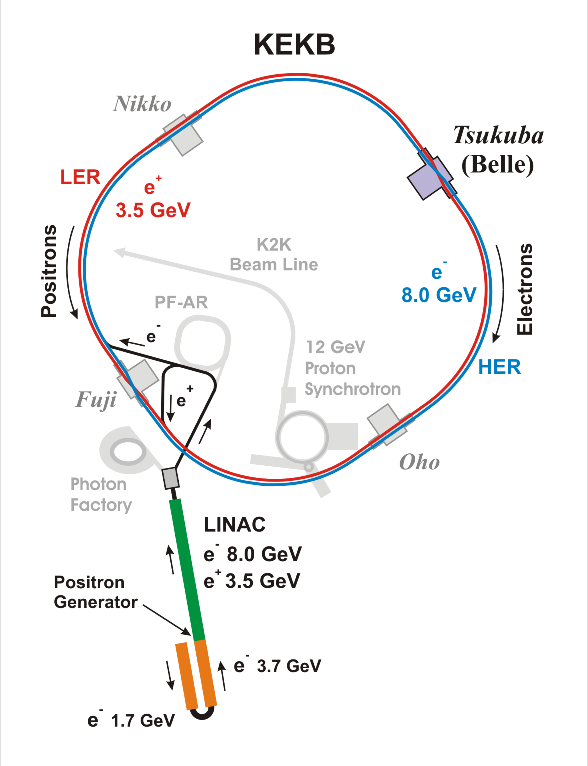


Figure 2.1.: Schematic view on the KEKB accelerator. Taken from [11].

collider can be seen in Fig. 2.1.
As the beam energies already imply, the KEKB accelerator was mainly operated on the $\Upsilon(4S)$ resonance with a center-of-mass energy of $\sqrt{s} = 10.58\,\mathrm{GeV}$. With its mass slightly higher than twice the B meson mass, the $\Upsilon(4S)$ decays in nearly a hundred percent of the cases into either a charged or neutral $\mathrm{B\overline{B}}$ pair. Furthermore, the accelerator ran also on energies corresponding to the $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ and $\Upsilon(5S)$ resonances to acquire data for additional physics studies. To investigate the non-$\mathrm{B\overline{B}}$ background, off-resonance data was also recorded at an energy 60 MeV below each of the resonances. A total list of all collected data can be found in Table 2.1.

Table 2.1.: Belle integrated luminosity taken at different center of mass energies [8].

Resonance	On-peak luminosity in fb^{-1}	Off-peak luminosity in fb^{-1}	Number of resonances
$\Upsilon(1S)$	5.7	1.8	102×10^6
$\Upsilon(2S)$	24.9	1.7	158×10^6
$\Upsilon(3S)$	2.9	0.25	11×10^6
$\Upsilon(4S)$	711.0	89.4	$772 \times 10^6 \mathrm{B\overline{B}}$
$\Upsilon(5S)$	121.4	1.7	$7.1 \times 10^6 \mathrm{B_s^0\overline{B}_s^0}$

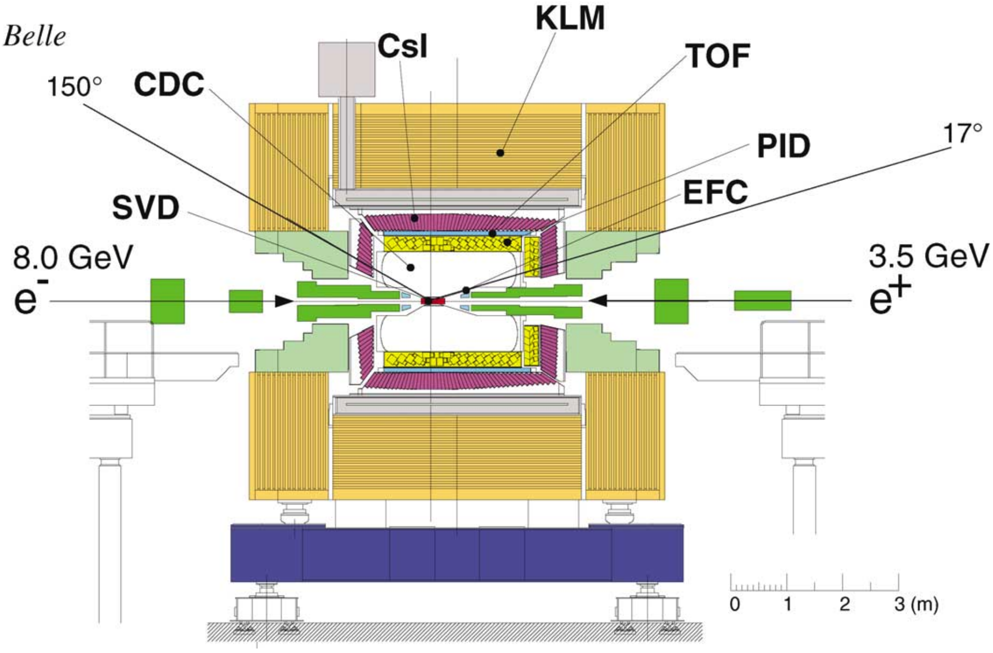


Figure 2.2.: Cross section of the Belle detector. The electromagnetic calorimeter (ECL) is named CsI in the sketch. Taken from [12].

2.2. The Belle Detector

Similar to the KEKB accelerator, the Belle detector had to fulfill several requirements to achieve high precision measurements. Generally, a B Factory requires a detector with a large coverage, a high resolution and outstanding particle identification. To comply with this, the Belle detector was built as a nearly hermetic 4π detector. This means that the there are nearly no gaps in the acceptance region of the detector. It consists of several sub-detectors, which were arranged cylindrically around the beam axis and are described below. A sketch of the Belle detector is shown in Fig. 2.2. Further information on the detector can be found in the References [8] and [12].

Silicon vertex detector (SVD)

To achieve one of the main goals of the Belle experiment, measuring time-dependent CP asymmetries, a very good vertex resolution is of uttermost importance since the difference in z-vertex positions of the B meson pair needs to be determined with a precision of $\sim 100\,\mu$m [9]. In addition to the vertexing, the information collected with the SVD also contributes to tracking.

The SVD design was strictly constrained by the characteristics of the investigated particles. In the explored energy range the vertex resolution is mainly affected by multiple scattering. To reduce multiple scattering the material budget of the detector and the beam pipe has to be kept as low as possible. Therefore, the innermost layer of the detector had to be placed as near to the beam pipe as possible, which then induced the problem that the SVD had to withstand large beam backgrounds. The readout electronics had to be placed outside of the acceptance region to reduce the material budget.

The runtime of the SVD can be split into two phases. The first version of the detector (SVD1) consisted of three double-sided strip detector (DSSD) layers, which covered an angular region of $23^\circ < \Theta < 139^\circ$. Due to radiation damage and to improve the coverage the SVD1 was later replaced with the SVD2, which consisted of four DSSD layers and covered a total region of $17^\circ < \Theta < 150^\circ$.

Central drift chamber (CDC)

In the Belle physics program, the determination of charged particle momenta played an essential role. This momentum measurement was done by reconstructing particle tracks within the CDC. Additionally, the CDC also provided information for the trigger system, as well as for particle identification by measuring the momentum loss over distance, $\mathrm{d}E/\mathrm{d}x$. A total of 8,400 drift cells were arranged in 50 cylindrical super layers, each containing three to six either axial or small-angle-stereo layers and three cathode strip layers. The chosen setup is asymmetric in z-direction and therefore provides an angular coverage of $17^\circ < \Theta < 150^\circ$. For the filling a mixture of 50% helium and 50% ethane was chosen, to on the one hand reduce multiple scattering, while on the other hand still providing a good $\mathrm{d}E/\mathrm{d}x$ resolution.

Aerogel cherenkov counter (ACC)

For analyzing the observed particles, it is crucial to identify the particle's type. In addition to the CDC and the time-of-flight system, the ACC was one of the sub-detectors used for particle identification at the Belle experiment. It makes use of the Cherenkov effect, which describes the phenomenon of the transit speed of a particle in a medium being faster than the speed of light in the same medium.

The Belle ACC consisted of 960 counter modules in the barrel region and 228 in the forward end-cap. Each module was equipped with five aerogel tiles, with photomultiplier tubes directly attached to them at the sides of the surrounding aluminum box. The refractive indices of the aerogels are chosen such that the separation power between kaons and pions is maximized.

Time-of-flight counters (TOF)

The particle identification detectors are completed by the time-of-flight counters. This system was built of 128 plastic scintillation counters, which, while taking the event time T_0 into account, measured the time of flight from the interaction point (IP) to the sub-detector. By additionally using the particle momentum given by the CDC, the mass of the particles could be calculated. The whole system was optimized for particles with $p < 1.2\,\mathrm{GeV}$, which covers most of the particles in an $\Upsilon(4\mathrm{S})$ event.

Electromagnetic calorimeter (ECL)

Photons, since they have no charge, cannot be measured with tracking detectors. However they produce electromagnetic showers when interacting with matter. This effect was used by the ECL to detect photons and measure their energy. The electromagnetic calorimeter can be divided into forward, backward, and barrel parts. A total of 8,736 CsI scintillator crystals, all approximately pointing to the interaction point, were used to detect electromagnetic showers. They do not point directly to the IP, to minimize the gaps in the acceptance region. With an angular coverage of $17^\circ < \Theta < 150^\circ$, the ECL covered the whole acceptance region also covered by the tracking devices of the Belle detector and was capable of measuring photons with an energy of up to 4 GeV.

In addition to measuring photons, the ECL was also used to distinguish electrons from charged hadrons or muons, by calculating the ratio of deposited energy and the momentum of the related track. Using the well known Bhabha signature, luminosity measurements were also performed with the electromagnetic calorimeter.

Extreme forward calorimeter (EFC)

The polar angle coverage of the ECL is extended by the EFC to $6.4^\circ < \Theta < 11.5^\circ$ in forward direction and $163.3^\circ < \Theta < 171.2^\circ$ in backward direction. This makes the detector more sensitive to some physics processes, especially $\mathrm{B} \to \tau\nu$. Monitoring the luminosity of the Belle experiment was the main purpose of the EFC, while it was also useful as a tagging device for two-photon physics.

K_L^0 and muon detection system (KLM)

Placed outside the solenoid, the KLM was designed to detect muons and K_L^0 mesons with momenta above 600 MeV with high efficiency. It was built out of alternating layers of charged particle detectors and iron plates (4.7 cm thick), providing a total of 3.9 radiation lengths in addition to the 0.8 radiation lengths of the ECL. Just as the ECL, the KLM can be divided into barrel and end-cap regions, with angular coverage of $45^\circ < \Theta < 125^\circ$ (barrel), which is extended to $20^\circ < \Theta < 155^\circ$ by the end-caps.

The KLM only allowed to determine the rough direction of the neutral K_L^0 mesons by assuming the IP as origin of the particle. Precise momentum and energy measurements were not possible, due to fluctuations in the shower size. The K_L^0 mesons and muons could be discriminated by their range and transverse scattering since muons usually travel further than the more likely interacting K_L^0.

Solenoid

A magnetic field of 1.5 T was generated by the superconducting solenoid. The coil of the solenoid had a diameter of 3.4 m and was 4.4 m in length. It was surrounded by an iron structure, which served as a return yoke for the magnetic flux, and acted as absorber material for the integrated KLM.

Trigger system

To handle the high data rates delivered by the KEKB accelerator, the Belle experiment needed a good trigger system to filter events which are of interest. The trigger system is divided into two parts, a Level-1 hardware trigger and a Level-3 software trigger. Each sub-detector sends its trigger information to a central trigger system, the Global Decision Logic (GDL). With the combined information an event is classified.

2.3. SuperKEKB and Belle II

After the shut down of the KEKB accelerator in 2010, the accelerator was upgraded to SuperKEKB, to match the requirements of the subsequent experiment Belle II. It is planned to get an about 40 times higher instantaneous luminosity with the SuperKEKB accelerator than KEKB achieved. The main changes, to achieve this goal, are different beam energies and the usage of the nano-beam scheme, which results in more focused beams at the interaction point and therefore a higher instantaneous luminosity. Table 2.2 compares the

Table 2.2.: Fundamental parameters of KEKB and SuperKEKB. Adapted from [14].

	KEKB (achieved)	SuperKEKB (planned)
Beam Energy in GeV (LER/HER)	3.5 / 8.0	4.0 / 7.0
η_y	0.129 / 0.090	0.090 / 0.088
β_y^* in mm	5.9 / 5.9	0.27 / 0.41
I in A	1.64 / 1.19	3.60 / 2.62
Luminosity in $\mathrm{cm}^{-2}\mathrm{s}^{-1}$	2.11×10^{34}	80×10^{34}

parameters of the two accelerators.

To handle the higher rates occurring from the higher luminosity the Belle detector had to be updated, accordingly. The general concept of the detector remained the same, nevertheless, there were some major changes done in most of the sub-detectors. Unlike at Belle, where the vertexing was performed by the SVD only, the vertexing detector of Belle II consists of two sub-detectors. A pixel detector (PXD), consisting of two layers of pixels, based on the DEPFET technology [13], is placed even nearer to the interaction point than the SVD at Belle was, because of the smaller beam pipe at Belle II. The PXD is surrounded by four layers of silicon strip detectors, the Belle II SVD. Due to the combination of being nearer to the interaction point, having more layers, and using pixels instead of strips in the innermost layers, the Belle II detector has a better vertex resolution than the Belle detector, which is also required due to the lower boost. The CDC got also improved by increasing the size and the number of sense wires. More detailed information about the design of the Belle II detector can be found in [14].

2.4. Analysis Software

The data which is analyzed in this thesis was collected by the Belle detector from 1999 to 2010. At this time the BELLE ANALYSIS FRAMEWORK (BASF) [15] was used for the reconstruction and analysis of the observed events. Nevertheless, this thesis uses the software, which was developed for the Belle II experiment, BASF2 [16]. This allows for the usage of many of the newly developed analysis tools, like the FULL EVENT INTERPRETATION [17]. To make the data compatible with the new framework, they are converted with the B2BII conversion tool [18].

After the offline reconstruction the Belle data was stored as PANTHER [19] tables in mDST files. These mDST files are converted into ROOT [20] based mDST files, to process them with the Belle II software framework.

3. Analysis

This chapter deals with the general analysis procedure. First, the used data samples are described. Afterwards, the reconstruction, starting from final state particles up to the $\Upsilon(4S)$ is explained step-by-step. In the end, the used selection criteria, as well as the selection algorithms, are motivated and validated.

3.1. Data Samples

During the more than one decade lasting runtime, Belle was able to collect a total of $711\,\mathrm{fb}^{-1}$ of data on the $\Upsilon(4S)$ resonance, corresponding to $(772 \pm 10) \times 10^{-6}\,\mathrm{B\overline{B}}$-pairs. To study the non-$\mathrm{B\overline{B}}$ background (continuum), an additional data set of $89.4\,\mathrm{fb}^{-1}$ was recorded at a center-of-mass energy 60 MeV below the resonance. To avoid personal biases during the analysis, the whole analysis procedure is optimized and validated on simulated events, before looking into the signal region on recorded data. Therefore a large amount of Monte Carlo (MC) events is needed, to keep the statistic uncertainties low during the analysis.

The official Belle MC is divided into streams, where each stream of simulated events corresponds to the number of events in the recorded data sample for a given decay type. Creating the MC can be divided into two steps: event generation and simulation. For the generation of the $\mathrm{B\overline{B}}$ events the package EvtGen [21], which was developed and is therefore optimized for B physics, is used. The continuum events are generated using Pythia [22]. After the events were generated the trajectories of the particles through the detector and their interaction with the detector material are simulated with Geant3 [23]. Since the official Belle MC is now about ten years old and therefore outdated regarding some particle properties as well as the modeling of some decays it is supplemented by private MC, which exchanges parts of the official MC and adds additional decays, for this analysis. For this in the Belle MC the events containing $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ decays are replaced with newly generated and simulated events of this decay type. Also, additional MC, to fill the gap between the exclusive and inclusive measured branching fractions of semi-leptonic $\mathrm{b} \to \mathrm{c}\ell\nu$ transitions, is added to the MC cocktail.

The following part of this section describes the individual MC samples used for this analysis in detail.

Official Belle MC

Generic

The generic sample consists of a total of ten streams and contains $\mathrm{b} \to \mathrm{c}$ transitions.

It is split into samples containing events of the form $\Upsilon(4S) \rightarrow B^+B^-$ (*charged*) and ones consisting of $\Upsilon(4S) \rightarrow B^0\overline{B}^0$ decays (*mixed*). The samples contain also the $B \rightarrow D^{**}\ell\nu$ events, which are removed from the generic samples and replaced by the newly generated ones described below.

Continuum

$e^+e^- \rightarrow q\bar{q}$ $(q = u, d, s, c)$ processes are provided in the continuum sample. The light quarks hadronize and form events which make up the dominant of the Belle data. The samples are also split into two types, one containing charm quarks (*charm*) and one containing the other available quark types on the given energy scale (*uds*). A total of six streams for each type is available.

$b \rightarrow u\ell\nu$ Sample

Instead of transitting from a bottom-quark to a charm-quark the b can also decay into an up-quark. Decays like this are included in the $b \rightarrow u\ell\nu$ sample and are used for this analysis since the other B meson is still able to decay via a $b \rightarrow c$ transition on the signal-side, which may be of interest. The sample consists of 20 streams.

Rare

To complete the official Belle MC that is used for this analysis, a sample that consists of decays with a very small branching fractions, e.g. $B \rightarrow \ell\nu\gamma$, is also included. The 50 streams of rare decays are taken into account for completeness.

$B \rightarrow D^{**}\ell\nu$ Sample

Masses and widths of the D^{**} particles are better known in the present days than they were when the generic Belle MC was produced. Therefore the $B \rightarrow D^{**}\ell\nu$ decays are produced with updated mass and width values taken from the *pdg* [24]. Old and updated values for the masses and widths are given in Table 3.1. These newly generated events are used to replace the ones contained in the official Belle MC. The generation is performed for each lepton type $(\ell = e, \mu, \tau)$ for decays of the form $B \rightarrow D^{**}(\rightarrow D^{(*)}\pi)\ell\nu$, where D^{**} describes one of the following orbitally excited D mesons: D_1, D_2^*, D_1', and D_0^*. The decay of the other B meson is not specified and is therefore generated, as in the generic Belle MC, via a $b \rightarrow c$ transition. For the subsequent D^{**} decay into $D^{(*)}$ with an either charged or neutral pion a ratio of 2 : 1 is used $({}^2\!/\!{}_3\pi^{\pm}, {}^1\!/\!{}_3\pi^0)$.

Gap Sample

When comparing the sum of all exclusively measured branching fractions of $B \rightarrow D^{(*,**)}\ell\nu$ decays, to the inclusive branching fraction measurement of $B \rightarrow X_c\ell\nu$, a difference of the order of about 10 % is seen. This gets often denoted as the gap problem in semi-leptonic B decays and is described in a more detailed fashion in [25].

One possible solution for this puzzle is filling the gap with a mixture of already measured decays, which are not already covered by any other sample, and additional decays, that are expected from theory predictions. At the same time some decays, which were expected back then, are removed from the generic MC. For this analysis, the gap is filled with a mix of three different decay types. $B \rightarrow D_1(\rightarrow D\pi\pi)\ell\nu$ decays form the first component, while the branching fractions for this decay are considered in the D_1 modes in the $B \rightarrow D^{**}\ell\nu$

Table 3.1.: Masses and width of the D^{**} states as used for the production of the official Belle MC. The updated values used for the production of the $B \to D^{**}\ell\nu$ sample are taken from the *pdg* [24].

	Mass in GeV		Width in MeV	
	Old value (Belle)	Updated Value	Old Value (Belle)	Updated Value
D_0^{*+}	2.308	2.349	276	221
D_0^{*0}	2.308	2.300	276	274
$D_1'^{+}$	2.422	2.427	412	384
$D_1'^{0}$	2.422	2.427	412	384
D_1^{+}	2.427	2.423	28	25
D_1^{0}	2.427	2.421	28	25
D_2^{*+}	2.459	2.465	25	47
D_2^{*0}	2.459	2.461	23	48

sample. Additionally non-resonant $B \to D^{(*)}\pi\pi\ell\nu$ decays are taken into account. The decay $B \to D^{(*)}\eta\ell\nu$ was not measured yet but is a promising candidate to fill the remaining gap and is therefore also included in the gap MC sample.

3.2. Event Reconstruction

Operating an e^+e^-–collider on the $\Upsilon(4S)$ resonance delivers a very predictable event scenario. The possibilities are far greater than compared for example to the LHC, since we are looking on any particle in the event. Since the $\Upsilon(4S)$ decays nearly exclusively into two B mesons ($> 96\,\%$), one has two reconstruct a pair of B mesons and combine them to a $\Upsilon(4S)$, when trying to reconstruct the whole event. A typical semi-leptonic event in the Belle detector is shown in Fig. 3.1. The B meson reconstruction is divided into two parts: Signal-side and tag-side reconstruction. One of the B mesons gets explicitly reconstructed on the signal side as B_{sig} in various common decay channels. Firstly, final state particles are selected and combined to form D meson candidates. These candidates will then either be combined with additional pions to reconstruct candidates for D^* and D^{**} in an intermediate step or are directly recombined with a lepton to get B meson candidates. The other B meson is used to determine further properties and for the reduction of background events. For this tag-side reconstruction the FULL EVENT INTERPRETATION (FEI) [17], which was developed for the Belle II experiment and is the successor of the FULL RECONSTRUCTION (FR) [26], that was used in previous Belle analysis, is applied.

Before performing the event reconstruction, some pre-selection criteria have to be fulfilled to get rid of obvious background events, occurring from beam gas and two-photon events. To do so for this analysis the default *HadronBJ Skim* [9] for hadronic events is used. Events have to contain at least three charged tracks with a transverse momentum greater than 0.1 GeV. Additionally the point of closest approach with respect to the interaction point of these tracks has to be near the interaction point: $|\Delta r| < 2\,\text{cm}$ and $|\Delta z| < 4\,\text{cm}$, which

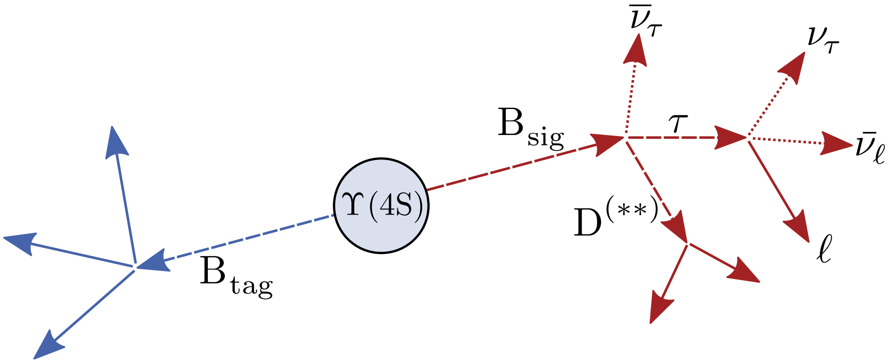


Figure 3.1.: Sketch of a typical $\Upsilon(4\mathrm{S})$ event, where the signal B meson decays semi-leptonically. The example shows a $\mathrm{B}_{\mathrm{sig}}$ decay including a τ which then decays with two additional neutrinos. The τ can be replaced by any other lepton, which would result in a more simple event topology, since the other leptons will not decay in the detector.

describe the transverse and z-direction distance to the interaction point, respectively. This skim takes also the visible energy into account. The sum of the energy of the tracks combined with the energy of all reconstruction photons has to account for at least 20 % of the center-of-mass energy in the event. More specific selection criteria are also included in the skim and are described more detailed in the reference given above.

3.2.1. Signal-Side Reconstruction

This section elaborates the recombination of the signal-side $\mathrm{B}_{\mathrm{sig}}$ meson. Each step of the reconstruction and the corresponding selection criteria are described in detail. The described reconstruction is similar to the one used in the underlying $\mathcal{R}(\mathrm{D}^{(*)})$ analysis, so the same reconstruction channels for the different particles are used and the same selection criteria are applied during the event reconstruction.

Final State Particles

The particles in the Belle detector can be divided into two main types. On the one hand particles with a short lifetime, like the D mesons, which decay within the detector, and on the other hand particles with a much longer lifetime, that do not decay inside the detector volume. This paragraph deals with the second type of particles, which are commonly called final state particles. The class of final state particles includes charged particles ($\mathrm{e}^\pm$, $\mu^\pm$, $\mathrm{K}^\pm$, and $\pi^\pm$), which were deduced from reconstructed tracks in the detector. Additionally, the information from ECL clusters is used to get photon candidates (γ). The list of final state particles is completed by the neutral K^0_S and π^0. Although these two are no final state particles by definition, since they decay within the detector, they are handled internally in the analysis like final state particles and are therefore added here.

To get rid of falsely reconstructed final state particles, particle dependent selection cuts are applied on the final state particles to improve the signal to noice ratio at this early stage of the event reconstruction. Additionally, Belle uses likelihood ratios for the separation of different particle hypotheses during the track reconstruction. These are gathered with

the information of several sub-detectors and are used to calculate particle identification (PID) values. For every charged particle type a cut of $p_{\mathrm{ID}} > 0.1$, with $p = \mathrm{e}$, μ, K, and π, is applied. The described PID cuts are very loose and only suppress obvious background, but help to reduce the number of combinations during the recombination of intermediate particles. In the later recombination to B mesons, the cuts on the electron and muon ID become stricter ($\mathrm{e}_{\mathrm{ID}} > 0.6$ and $\mu_{\mathrm{ID}} > 0.9$). At the latter stage, a cut on the momentum of these particles is applied, as well: $|\vec{p}_{\mathrm{e}}| > 300\,\mathrm{MeV}$ and $|\vec{p}_{\mu}| > 600\,\mathrm{MeV}$. During the selection of photon candidates, it is a common practice for Belle analysis to perform a selection on the so-called *goodBelleGamma* variable. Photon candidates, which are flagged with this variable, have to fulfill a requirement on their momentum based on the place of their detection. Candidates detected in the barrel region of the ECL are required to have a momentum of at least 50 MeV, while ones detected in the forward or backward region have to match the requirement of $E(\gamma) > 100\,\mathrm{MeV}$ and $E(\gamma) > 150\,\mathrm{MeV}$, respectively. The neutral K^0_S are reconstructed from V0 objects. Such a V0 object describes a displaced vertex, which occurs if a neutral particle (like K^0_S or γ (from pair-production)) decays within the detector into a pair of charged particles. Similar to the *goodBelleGamma* variable, Belle provides a flag for properly reconstructed K^0_S, the *ksnbStandard*. In addition to this, requirements on the reconstructed mass of the kaon, more specifically the difference between the reconstructed and nominal mass has to be below 100 MeV and $|(\mathrm{M}_{\mathrm{inv}} - \mathrm{M}_{\mathrm{nom}})/\sigma_{\mathrm{M}_{\mathrm{inv}}}| < 3.0$, are made. The latter cut also applies for the reconstructed π^0's, while for them their reconstructed mass has to be in the window of $104\,\mathrm{MeV} < \mathrm{M}_{\mathrm{inv}} < 165\,\mathrm{MeV}$. Since most of the neutral pions are reconstructed via $\pi^0 \to \gamma\gamma$, the photons from this decay also have to match the *goodBelleGamma* conditions described above.
A summary of all final state particles which are used for the recombination of the intermediate particles is shown in Table 3.2.

Bremsstrahlung Correction

Energy loss due to bremsstrahlung is a very crucial aspect in B physics experiments, where a good momentum resolution is required. Due to their low mass, electrons can lose large amounts of their energy along their trajectory through the detector layers, when interacting with them. This effect leads to the underestimation of the electron's momentum, which results in a tail in the squared missing mass distribution, which is described in detail in Section 5.1.1. To counteract this effect, the electrons are corrected by adding the four-momentum of bremsstrahlung photon candidates to the four-momentum of the electron. Using the B2BII conversion allows for the usage of many analysis tools, which are developed for the Belle II experiment. However, since the conversion is based on the mdst objects, and one does not have any knowledge about the tracking information anymore, the Belle II bremsstrahlung finding algorithm [27] cannot be applied here. Therefore, the search for bremsstrahlung photons is performed similarly as in previous Belle analyses, e.g. in [28].
Photon candidates are collected within a cone in the θ and ϕ plane. The cone is defined in a $2.0\,^\circ$ angle around the point of closest approach to the interaction point, with the symmetry axis pointing in the direction of the initial electron momentum at the interaction point. This leads to the restriction, that only bremsstrahlung photons, which are emitted at the innermost part of the detector can be found. Additionally, a bremsstrahlung photon candidate's momentum has to be below 0.4 GeV and shall not exceed 40 % of the electron's

Table 3.2.: List of all final state particles and their corresponding selection criteria. The values in brackets for the PID cuts show the values used during the initial selection of the candidates.

Final State Particle	Selection Criteria
Charged Tracks	$\lvert\Delta r\rvert < 2\,\mathrm{cm}$ $\lvert\Delta z\rvert < 4\,\mathrm{cm}$ $\mathrm{p}_t > 0.1\,\mathrm{GeV}$
$\mathrm{e}^{\pm}$	$\lvert\vec{p}_{\mathrm{e}}\rvert > 300\,\mathrm{MeV}$ $\mathrm{e}_{\mathrm{ID}} > 0.6 \quad (0.1)$
$\mu^{\pm}$	$\lvert\vec{p}_{\mu}\rvert > 600\,\mathrm{MeV}$ $\mu_{\mathrm{ID}} > 0.9 \quad (0.1)$ $\mu_{\mathrm{ID}}(\mathrm{BelleQuality}) = 1$
$\pi^{\pm}$	$\pi_{\mathrm{ID}} > 0.1$
$\mathrm{K}^{\pm}$	$\mathrm{K}_{\mathrm{ID}} > 0.1$
γ	*goodBelleGamma* = 1
$\mathrm{K}^0_{\mathrm{S}}$	$\lvert\Delta\mathrm{M}\rvert < 100\,\mathrm{MeV}$ $\lvert(\mathrm{M}_{\mathrm{inv}} - \mathrm{M}_{\mathrm{nom}})/\sigma_{\mathrm{M}_{\mathrm{inv}}}\rvert < 3.0$ *ksnbStandard* = 1
π^0	$104\,\mathrm{MeV} < \mathrm{M}_{\mathrm{inv}} < 165\,\mathrm{MeV}$ $\lvert(\mathrm{M}_{\mathrm{inv}} - \mathrm{M}_{\mathrm{nom}})/\sigma_{\mathrm{M}_{\mathrm{inv}}}\rvert < 3.0$ *goodBelleGamma* = 1 (for all γ)

momentum. Other parameter values, as well as methods using an arc instead of a cone and the cone not being centered around the particle's trajectory, are tested, as well. Nevertheless, the most sufficient results are achieved with the given parameters.

All found bremsstrahlung photon candidates are used to correct the electron momenta by adding their four-momenta. The electron particle list is extended by the corrected electrons, and all electron candidates are treated equally afterwards. So, the decision, if for the reconstruction of the semileptonic $\mathrm{B} \to \mathrm{D}^{(*,**)}\ell\nu$ decay the corrected or uncorrected electron is used, depends on independent selection criteria on the intermediate particle properties and the whole event.

4.74 % of all reconstructed and selected $\mathrm{B} \to \mathrm{D}^{**}\mathrm{e}\nu$ signal decays are corrected. This makes up 2.59 % of all $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ signal events, which passed the selection criteria (Section 3.3.2). Fig. 3.2 shows the impact of the bremsstrahlung correction on the squared missing mass distribution. The bremsstrahlung corrected squared missing mass distribution shows a better resolution for the signal peak, than with no correction applied.

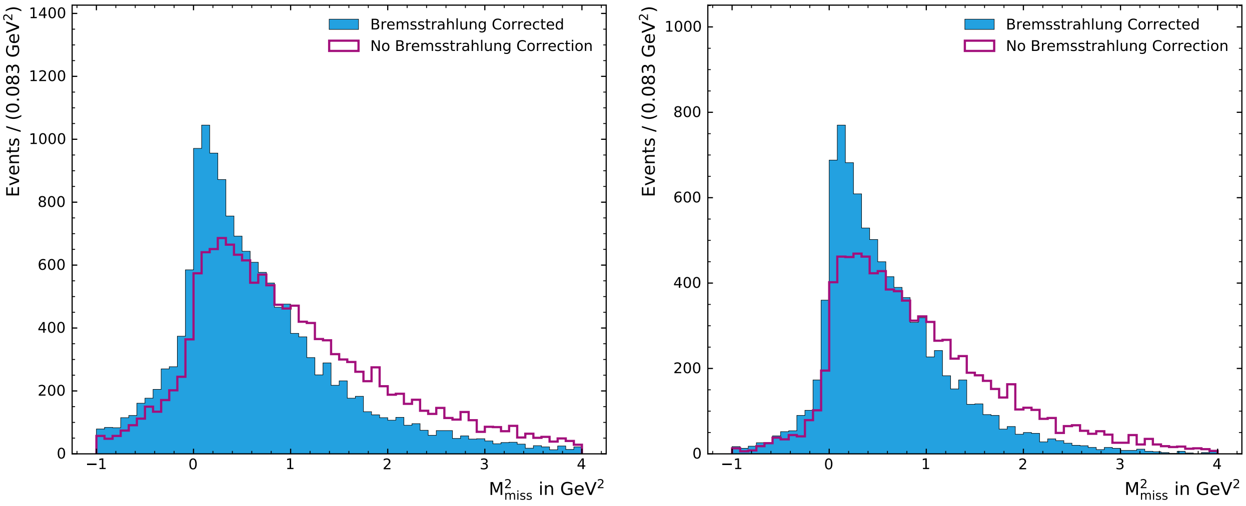


Figure 3.2.: Impact of the bremsstrahlung correction on the squared missing mass in the D^{**} reconstruction channels. All events, which were reconstructed with a bremsstrahlung corrected electron are taken into account. Left: All components in the data set are considered. Right: Only the $\mathrm{D}^{**}(\rightarrow \mathrm{D}^{(*)}\pi)\ell\nu$ signal components are considered.

$\mathrm{D}^{(*)}$ Meson Reconstruction

After the selection of the final state particles, the reconstruction starts with the combination of them to form intermediate D mesons. This is done in a total of 16 decay channels, which are equally divided into charged and neutral ones (see Table 3.3). To avoid background from falsely reconstructed D mesons, the reconstructed mesons have to pass some selection criteria. The reconstructed momentum of the D meson in the center-of-mass frame has to be below 3 GeV. Other, more channel specific, criteria are based on the involved particles. With a few minor exceptions, for D meson candidates, which were reconstructed from charged particles exclusively, a cut of $|(\mathrm{M}_{\mathrm{inv}} - \mathrm{M}_{\mathrm{nom}})/\sigma_{\mathrm{M}_{\mathrm{inv}}}| \leq 4.0$ is applied, while for decays including neutral particles, the difference between the nominal D mass and the reconstructed one has to be smaller than 40 or 50 MeV, depending on the decay channel. Decays including a neutral pion, also set a requirement on the momentum of the π^0 in the center-of-mass frame of the $\mathrm{e}^+\mathrm{e}^-$ collision ($\mathrm{p}^*(\pi^0)$), which has to be above 200 MeV. To keep the number of D meson candidates in an event reasonably small, and thereby reduce the combinatorics in further steps, a best candidates selection is applied. For this the candidates are sorted after their absolute mass difference to the nominal value ($\Delta\mathrm{M} = \mathrm{M}_{\mathrm{inv}} - \mathrm{M}_{\mathrm{nom}}$) and only the ten candidates with the smallest values are kept for further recombination.

Once the D candidates are reconstructed and selected, the procedure continues with trying to find candidates for excited D^* mesons. To get such candidates, the previously selected D mesons are combined with pions or photons. For this analysis all relevant channels (two neutral and two charged ones) are used (see Table 3.4). Background is suppressed by setting a requirement on the mass difference of the two $\mathrm{D}^{(*)}$ mesons ($\Delta\mathrm{M}(\mathrm{D}, \mathrm{D}^*) = \mathrm{M}(\mathrm{D}^*) - \mathrm{M}(\mathrm{D})$). In the decay $\mathrm{D}^{*+} \rightarrow \mathrm{D}^0\pi^+$ the momentum of the pion $\mathrm{p}^*(\pi^+)$ has to be above 400 MeV due to phase space restrictions. This helps to suppress continuum background in this channel.

Table 3.3.: List of all D decay channels and their corresponding selection criteria. For better readability the cut on the momentum of the D meson, which applies to all of the channels, is excluded in this summary.

D	Decay Channel	Selection Criteria
D^+	$\to K^-\pi^+\pi^+$	$\lvert(M_{inv} - M_{nom})/\sigma_{M_{inv}}\rvert \leq 4.0$
	$\to K^-\pi^+\pi^+\pi^0$	$\lvert\Delta M\rvert < 40\,\mathrm{MeV}$ $p^*(\pi^0) > 200\,\mathrm{MeV}$
	$\to K^-\pi^+\pi^+\pi^+\pi^-$	$\lvert(M_{inv} - M_{nom})/\sigma_{M_{inv}}\rvert \leq 4.0$
	$\to K^0_S\pi^+$	$\lvert\Delta M\rvert < 40\,\mathrm{MeV}$
	$\to K^0_S\pi^+\pi^0$	$\lvert\Delta M\rvert < 40\,\mathrm{MeV}$ $p^*(\pi^0) > 200\,\mathrm{MeV}$
	$\to K^0_S\pi^+\pi^+\pi^-$	$\lvert\Delta M\rvert < 40\,\mathrm{MeV}$
	$\to K^0_S K^+$	$\lvert\Delta M\rvert < 40\,\mathrm{MeV}$
	$\to K^+K^-\pi^+$	$\lvert\Delta M\rvert < 40\,\mathrm{MeV}$
D^0	$\to K^-\pi^+$	$\lvert(M_{inv} - M_{nom})/\sigma_{M_{inv}}\rvert \leq 4.0$
	$\to K^-\pi^+\pi^0$	$\lvert\Delta M\rvert < 50\,\mathrm{MeV}$ $p^*(\pi^0) > 200\,\mathrm{MeV}$
	$\to K^-\pi^+\pi^+\pi^-$	$\lvert(M_{inv} - M_{nom})/\sigma_{M_{inv}}\rvert \leq 4.0$
	$\to K^-\pi^+\pi^+\pi^-\pi^0$	$\lvert\Delta M\rvert < 40\,\mathrm{MeV}$ $p^*(\pi^0) > 200\,\mathrm{MeV}$
	$\to K^0_S\pi^0$	$\lvert\Delta M\rvert < 50\,\mathrm{MeV}$
	$\to K^0_S\pi^+\pi^-$	$\lvert\Delta M\rvert < 40\,\mathrm{MeV}$
	$\to K^0_S\pi^+\pi^-\pi^0$	$\lvert\Delta M\rvert < 40\,\mathrm{MeV}$ $p^*(\pi^0) > 200\,\mathrm{MeV}$
	$\to K^-K^+$	$\lvert(M_{inv} - M_{nom})/\sigma_{M_{inv}}\rvert \leq 4.0$

Table 3.4.: List of all D^* decay channels and their corresponding selection criteria.

D^*	Decay Channel	Selection Criteria
D^{*+}	$\to D^0\pi^+$	$\Delta M(D, D^*) < 155\,\mathrm{MeV}$ $p^*(\pi^+) > 400\,\mathrm{MeV}$
	$\to D^+\pi^0$	$\Delta M(D, D^*) < 155\,\mathrm{MeV}$
D^{*0}	$\to D^0\pi^0$	$\Delta M(D, D^*) < 160\,\mathrm{MeV}$
	$\to D^0\gamma$	$130\,\mathrm{MeV} < \Delta M(D, D^*) < 155\,\mathrm{MeV}$

Table 3.5.: List of all B_{sig} decay channels, with $\ell = e, \mu$. The last rows show the selection criteria for each type of channels.

$D^{(*)}$-Channels	$D^{(*)}\pi^0$-Channels	$D^{(*)}\pi^{\pm}$-Channels
$B^- \to D^0\ell^-$ $B^- \to D^{*0}\ell^-$	$B^- \to D^0\pi^0\ell^-$ $B^- \to D^{*0}\pi^0\ell^-$	$B^- \to D^+\pi^-\ell^-$ $B^- \to D^{*+}\pi^-\ell^-$
$B^0 \to D^-\ell^+$ $B^0 \to D^{*-}\ell^+$	$B^0 \to D^-\pi^0\ell^+$ $B^0 \to D^{*-}\pi^0\ell^+$	$B^0 \to \overline{D}^0\pi^-\ell^+$ $B^0 \to \overline{D}^{*0}\pi^-\ell^+$
$1.0\,\mathrm{GeV} < M_{inv} < 6.0\,\mathrm{GeV}$ $R_2 < 0.6$	$1.0\,\mathrm{GeV} < M_{inv} < 6.0\,\mathrm{GeV}$ $R_2 < 0.6$ $p^*(\pi^0) > 200\,\mathrm{MeV}$	$1.0\,\mathrm{GeV} < M_{inv} < 6.0\,\mathrm{GeV}$ $R_2 < 0.6$

D^{**} Reconstruction

For this analysis, the orbitally excited D mesons of interest, which are denoted as D^{**}, are the $1P$ states D_1, D_2^*, D_1', and D_0^*. During the reconstruction, no differentiation between the four states is made, so they are all handled equally. Furthermore, while the $D^{(*)}$ mesons are reconstructed explicitly, there is no explicit reconstruction of the D^{**}. Their decays, which are of the form $D^{**} \to D^{(*)}\pi$, are included in the reconstruction of the signal-side B meson ($B_{sig} \to D^{(*)}\pi\ell$). The D^{**} could also decay in more than one pion, an η or γ, but these decays are not considered in this analysis.

Signal B Meson Reconstruction

Once the $D^{(*)}$ candidates are reconstructed, all ingredients to form B meson candidates are available. For their reconstruction three types of semi-leptonic decay channels are defined: $B \to D\ell\nu$, $B \to D^*\ell\nu$ and for this analysis the most important one, $B \to D^{**}\ell\nu$, which are getting reconstructed as $B \to D^{(*)}\pi\ell\nu$. Four channels are considered for each of the first two types, differentiating between charged and neutral B mesons as well as electron and muon. A total of 16 channels is defined for the D^{**} case (8 with charged and 8 with neutral pions). The non-D^{**} channels are included in the reconstruction to improve the fit-procedure later on.

As with any other reconstructed particle before, the signal B meson has to match some requirements in order to be considered for further analysis. Every reconstructed B meson must have an invariant mass in the range of 1 GeV to 6 GeV, to not be rejected by the reconstruction algorithm. Furthermore, some continuum background, which is explained more detailed later on in Section 3.3.1, is suppressed when reconstructing the B meson candidates by applying a loose cut of $R_2 < 0.6$. As already indicated above, for the leptons in the decay chain, cuts on the PID values and their momenta are made (the values can be found in Table 3.2). The decays including a D^{**} are either reconstructed with a charged or neutral pion. While for the case with the charged pion no additional requirements are set, the momentum of the used π^0 has to fulfill the condition $p^*(\pi^0) > 200\,\mathrm{MeV}$.

3.2.2. Tag-Side Reconstruction

To recombine a $\Upsilon(4S)$ in the end, exactly two reconstructed B mesons are needed. After the reconstruction of the B_{sig}, which was described before, the tag-side meson B_{tag} has to be found. As already stated before, this is done by using the FULL EVENT INTERPRETATION (FEI) [17], which was originally developed for the Belle II experiment, but is written in a generic enough manner so it can also be applied on Belle data since the B2BII conversion is used in this analysis.
The FEI algorithm uses a hierarchical approach, containing six stages (see Fig. 3.3). Starting with collecting the information gathered from the detector (tracks, $V0$-objects, and neutral clusters), final state particles are formed. These are, in a similar way to the signal-side reconstruction, used to build intermediate particles, like J/ψ or D mesons, according to previously defined decay chains. Using multivariate classifiers, every reconstructed particle receives a probability of being correctly reconstructed. The FEI is capable of performing either a hadronic or semi-leptonic tag. One way of validating the performance of the FEI is by comparing it to its predecessor, the FULL RECONSTRUCTION. Referring to [17], the FEI outperforms the FR, when comparing the ROC curves. This leads to the expectation of a more statistically significant result in the end when using the FEI instead of the FR.
In this analysis a purely hadronic tag-side reconstruction is used, since the signal-side already includes invisible energy, due to the neutrino in the final state. Since there are no FEI skims available for Belle data, the tag-side reconstruction is performed after the signal-side reconstruction, using the rest of event (ROE), namely everything that was not already used for the reconstruction of the B_{sig}. Therefore, the FEI is applied on every found signal-side B candidate, searching for the corresponding B_{tag}.
The FEI can be trained either generic or specific. These two approaches differ in the fact that in the generic one the FEI is trained independently on the investigated signal decay, while in the specific case the FEI is trained on the ROE after a potential signal-side B meson is found. For this thesis the version FEIV4 of the FEI with the generic training FEIV4_2017_MCCONVERTED_TRACK14_2 is used.

B_{tag} Selection

As for the signal-side B meson, some requirements on the B_{tag} have to be set, to reduce the number of falsely reconstructed candidates. This is done by exploiting the well-known kinematics in e^+e^-–collisions at Belle. The initial four-momentum of the $\Upsilon(4S)$ is well known, and so are the theoretical absolute momenta of the B mesons, it decays in. Using this, two variables are defined to suppress falsely reconstructed candidates. The beam constrained mass, which is defined as

$$M_{bc} = \sqrt{E_{beam}^2 - p^*(B_{tag})^2} \tag{3.1}$$

with $E_{beam} = \sqrt{s}/2$ being the beam energy in the rest frame of the $\Upsilon(4S)$ and the momentum of the B_{tag} in the center-of-mass frame $p^*(B_{tag})$. To be accepted as a valid B_{tag}, the candidate has to fulfill the criteria $M_{bc} > 5.22\,\mathrm{GeV}$, since correctly reconstructed candidates peak at $M_{bc} = 5.28\,\mathrm{GeV}$ which equals the nominal mass of the B meson. Note that this cut will be more strict in the final selection, but at this stage is held loose to make

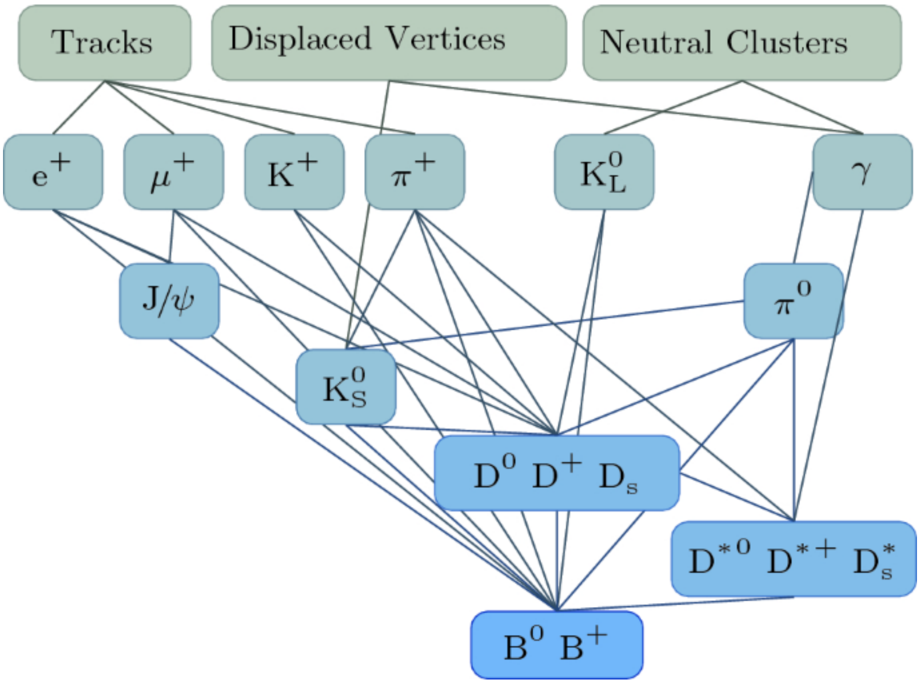


Figure 3.3.: Overview of the hierarchical structure of the FEI. Starting with detector information, final state particles are formed and combined to intermediate particles, getting B mesons in the end. Taken from [17].

sideband studies possible. Furthermore, the background is suppressed with a cut on ΔE, defined as

$$\Delta E = E_B^* - E_{beam} \tag{3.2}$$

using the energy of the B_{tag} in the center-of-mass frame E_B^*. The requirement chosen here is $-50\,\text{MeV} < \Delta E < 100\,\text{MeV}$.

3.2.3. $\Upsilon(4S)$ Reconstruction

For events, where at least one signal-side B meson and one for the tag-side are found, the reconstruction is continued with the final step, the reconstruction of the $\Upsilon(4S)$. Besides the obvious decay channels, of combining either two oppositely charged or two neutral B mesons, there are more unphysical combinations allowed in the recombination. These combine for example charged and neutral B mesons, which results in a charged $\Upsilon(4S)$ resonance. These channels are included, since for the analysis mainly the properties of the signal side B_{sig} meson are of interest. With the unphysical combinations, one avoids to dismiss events for which a slow track was missed on the tag or signal side, which could for example occur due to the limited detector acceptance. A reconstructed $\Upsilon(4S)$ must have an invariant mass within the range of $7.0\,\text{GeV} < M_{inv} < 13.0\,\text{GeV}$. Additionally, there are no additional charged tracks allowed in the event, so every charged track has to be used either for the tag or signal-side reconstruction. This check is done by constructing the rest-of-event (ROE). Tracks and clusters have to fulfill the same selection criteria already

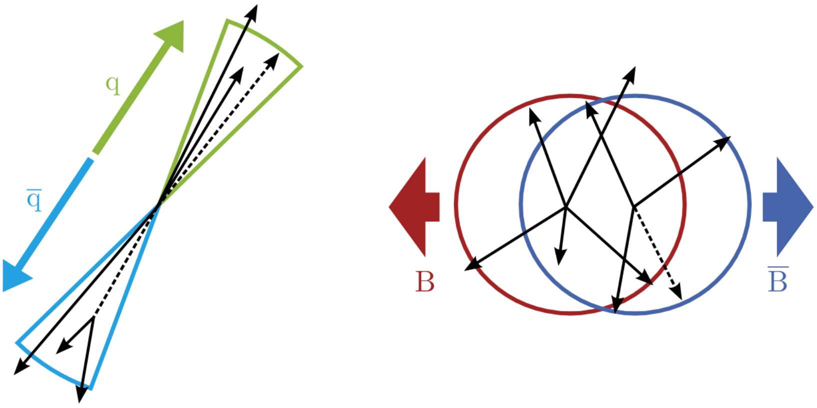


Figure 3.4.: Different event topologies, depending on the underlying process. Left: Continuum event with $e^+e^- \rightarrow q\bar{q}$. Right: $\mathrm{B\overline{B}}$ event, which is more isotropically distributed. Adapted from [29].

described for the final state particles. Clones due to curling tracks are also discarded in the construction of the ROE, by comparing the momenta and the angular information to other tracks in the event.

3.3. Event Selection

The event reconstruction delivers a mixture of $\Upsilon(4S)$ candidates, containing both signal events, which are of interest for the analysis, and falsely reconstructed events, which originated from continuum or $\mathrm{B\overline{B}}$ background. Suppressing the background events, without losing too much statistics of the signal events, is a crucial task for many analyses. In this analysis multivariate classifiers, as well as cuts on event variables are used to achieve a separation between signal and background events.

3.3.1. Continuum Suppression

The data recorded by the Belle detector not only contains events originating from $e^+e^- \rightarrow \Upsilon(4S)$, but also from the more likely process $e^+e^- \rightarrow q\bar{q}$, with q denoting one of the light quarks (u, d, s, c). Continuum events, as the latter are called, can be separated from $\mathrm{B\overline{B}}$ events by comparing the event topology. $q\bar{q}$–pairs are produced nearly back-to-back, which results in two jets. On the contrary, due to the kinematic constraints, the $\mathrm{B\overline{B}}$ pair gets produced nearly at rest in the center-of-mass frame, so the topology for such events is more isotropic. The different event shapes are shown in Fig. 3.4.
Since suppressing continuum background is necessary at most B-Factory experiments, many collaborations have dealt with the problem, resulting in a large set of predefined variables to separate continuum background.

Thrust

Originally used for quantifying jets, the thrust can also be used for the separation of continuum and $\mathrm{B\overline{B}}$-events. The thrust axis $\vec{\mathrm{T}}$ defines the axis that maximizes the sum of the longitudinal momenta of final state particles. With knowledge of the thrust axis and the particle momenta the thrust T can be calculated via

$$\mathrm{T} = \frac{\sum_{i=1}^{N} |\vec{\mathrm{T}}\vec{\mathrm{p}}_i|}{\sum_{i=1}^{N} |\vec{\mathrm{p}}_i|} \tag{3.3}$$

for N final state particles. Values for the thrust depend on the shape of an event, with $\mathrm{T} \approx 1$ for continuum events and $\mathrm{T} \approx 0.5$ for $\mathrm{B\overline{B}}$-events. The thrust (axis) can be calculated for the $\mathrm{B_{tag}}$ and the ROE including the signal side independently. The angle Θ_T between the two resulting thrust axis of the tag-side and the ROE is also a useful continuum suppression variable. Similar to this, the angle between the thrust axis of the $\mathrm{B_{tag}}$ and the beam axis Θ_B also delivers the required information to use it as as continuum suppression variable. For jet-like $q\bar{q}$ events, large Θ values are expected, whereas the angles for the $\mathrm{B\overline{B}}$-events are uniformly distributed.

Cleo Cones

The CLEO collaboration defined a total of nine Cleo cones for the suppression of continuum background [30]. They describe the scalar momentum flow around the thrust axis into concentric cones in intervals of $10\,^\circ$.

Reduced Fox-Wolfram Moment $\mathbf{R_2}$

First used for the description of event shapes in $\mathrm{e^+e^-}$–annihilation [31], the Fox-Wolfram Moments are another useful set of variables. They are calculated via

$$H_l = \sum_{i,j}^{N} |\vec{\mathrm{p}}_i||\vec{\mathrm{p}}_j| P_l(\cos\Theta_{ij}) \tag{3.4}$$

where P_l is the l-th order Legendre Polynomial and Θ_{ij} being the angle between momenta $\vec{\mathrm{p}}$ of the particles i and j. They reduced Fox-Wolfram Moments can be calculated using

$$\mathrm{R}_l = \frac{H_l}{H_0} \tag{3.5}$$

For the continuum suppression applied in this analysis, only $\mathrm{R_2}$, which was already used in the selection of $\mathrm{B_{sig}}$ candidates, is used.

Kakuno-Super-Fox-Wolfram Moments

Belle developed an improved variant of the Fox-Wolfram Moments described above, the Kakuno-Super-Fox-Wolfram Moments. A detailed description can be found in [9]. Using the Fisher discriminant, a total of 17 variables is defined.

All these variables described above are used to train a multivariate classifier, which results in one final continuum suppression variable $\mathcal{P}_\mathrm{CS}$. This is done under the use of the library FastBDT [32], which provides a speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for classification. The classifier is trained and tested on two independent samples, containing signal and continuum background events. For the background, a subset of the continuum samples, described in Section 3.1, is used, which is excluded from the generic MC sample afterwards, so the events are not longer

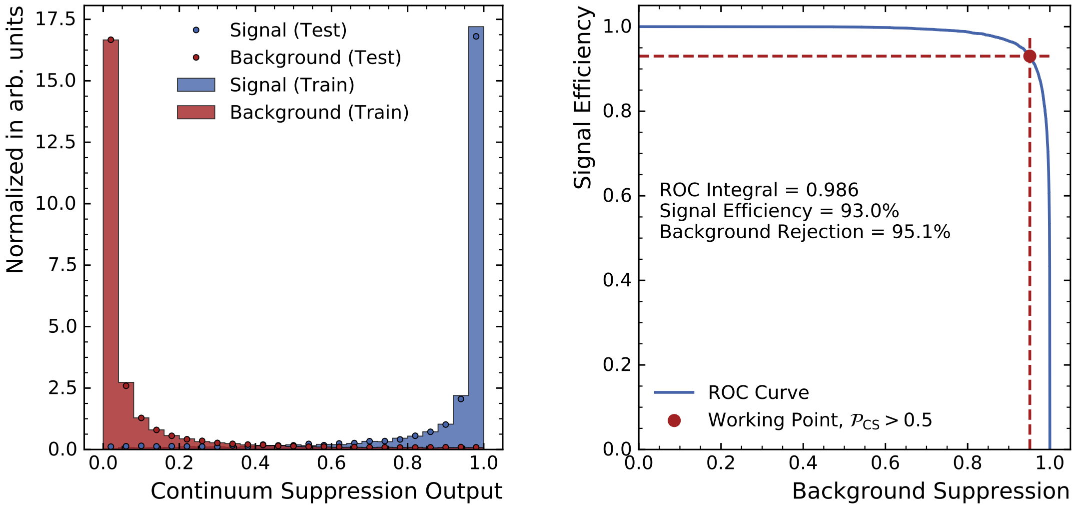


Figure 3.5.: Left: Distribution of the continuum suppression variable $\mathcal{P}_{\mathrm{CS}}$, calculated by the classifier. The output is shown for the training (bars) and the test sample (dots). Right: ROC curve for the continuum suppression BDT. The curve is evaluated on the independent test data set. Signal Efficiency and Background Rejection values are calculated for a working point of $\mathcal{P}_{\mathrm{CS}} > 0.5$.

used during the rest of the analysis. The signal samples were expressly generated for the training of multivariate classifiers, to avoid correlations with the efficiency calculation later on. Test and training samples contain the same amount of events. The classifier is trained using the default hyperparameter values for the decision trees, which are provided by the library, with only the number of trees set to $N_{\mathrm{trees}} = 200$.
Fig. 3.5 shows the validation of the continuum suppression. Looking at the continuum suppression output variable $\mathcal{P}_{\mathrm{CS}}$, one observes that a good separation between signal and continuum background is achieved by the classifier. This is also indicated by the high area-under-curve score of 0.986. Also, the distributions on test and train data are very similar, so there are no hints for overtraining of the BDT's. To suppress the most obvious background, an initial cut of $\mathcal{P}_{\mathrm{CS}} > 0.2$ is chosen. Applying this cut already rejects about 83 % of the background, while keeping over 97 % of the signal events. $\mathcal{P}_{\mathrm{CS}}$ is also used to get an even better separation between signal and background events later on.

3.3.2. Analysis Selection

Once the major part of continuum background is removed, one has to deal with the remaining background. In an ideal world, the $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ reconstruction modes would only contain events following this decay chain. However, since the reconstruction is not perfect the modes also contain a lot of other wrongly reconstructed events. To enrich the fraction of signal events ($\mathrm{B} \to \mathrm{D}^{**}\ell\nu$) compared to the background events, a combination of two different techniques is used. Cuts on discriminating event variables are applied to get rid of wrongly reconstructed events. Additionally, another multivariate classifier is trained and applied to separate signal and background events. The number of available variables for this classifier is limited by the fact that the classifier should be kept model independent and should not learn features from the MC which might not be well simulated. Also some cuts are motivated by the desired physical requirements. Both methods are described below.

Multivariate Classifier

For the training of the multivariate classifier, which is used to enrich the signal, the same setup as for the continuum suppression, only with a different target variable, is used. For this a signal event is defined as an event in which a true $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ candidate was reconstructed. A true $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ event is identified by the MC information of the lepton and the related $\mathrm{D}^{(*)}$ and pion. Searching for feature variables to use is quite difficult since the variables should on the one hand provide a separation between signal and background, but on the other hand should also not be correlated with variables, which are used for fitting at later stages of the analysis. Variables that fulfill these requirements are described in the following.

Extra energy in the calorimeter $\mathbf{E}^{\mathbf{ECL}}_{\mathbf{extra}}$

The extra energy in the calorimeter is defined as the sum of all ECL cluster energies in the ROE of the reconstructed $\Upsilon(4S)$. For correctly reconstructed events, this variable should peak at low values. Comparing data with MC shows discrepancies in the low energy region $\mathrm{E}^{\mathrm{ECL}}_{\mathrm{extra}} < 0.6\,\mathrm{GeV}$ (see Fig. 3.6). Therefore the classifier is made blind in this region by setting the value of affected events to $\mathrm{E}^{\mathrm{ECL}}_{\mathrm{extra}} = 0.6\,\mathrm{GeV}$.

$\mathbf{D}^{(*)}$ meson mass $\mathbf{M}(\mathbf{D}^{(*)})$

This variable describes the mass of the leading D meson, which can either be a D or D^{*}, is also used as a feature of the classifier. For correctly reconstructed events, this value should peak at the nominal mass of the corresponding meson, making a distinction of falsely reconstructed events possible.

$\mathbf{B}_{\mathbf{sig}}$ decay mode

This variable describes in which channel the $\mathrm{B}_{\mathrm{sig}}$ was reconstructed. It is used so the classifier is able to utilize the other variables dependent on the decay mode, without training a classifier for every decay channel separately. For the variables given above, which correlate with the decay mode, this is especially useful.

$\boldsymbol{\Upsilon}(4S)$ decay mode

Similar to the description of the $\mathrm{B}_{\mathrm{sig}}$ decay mode above, the $\Upsilon(4S)$ decay mode variable is used to account for dependencies on the reconstructed $\Upsilon(4S)$ decay.

$\mathbf{D}^{(*)}$ daughter mass $\mathbf{M}_{\mathbf{daughter}}(\mathbf{D}^{(*)})$

For decays, that have a D^{*} in leading order, this variable describes the mass of the D meson the D^{*} decays into. In other cases, where the leading particle is a D meson, the mass of the first kaon in the decay is described. A differentiation between these cases can be made by the decay mode information described above.

Continuum suppression output $\boldsymbol{\mathcal{P}}_{\mathbf{CS}}$

The output given by the continuum suppression classifier described before. Falsely reconstructed events tend to have lower values for $\mathcal{P}_{\mathrm{CS}}$.

Using all these variables, the classifier is trained and applied on data and MC, resulting in an output variable $\mathcal{P}_{\mathrm{sel}}$. Fig. 3.7 shows the distribution of the output variable, together with

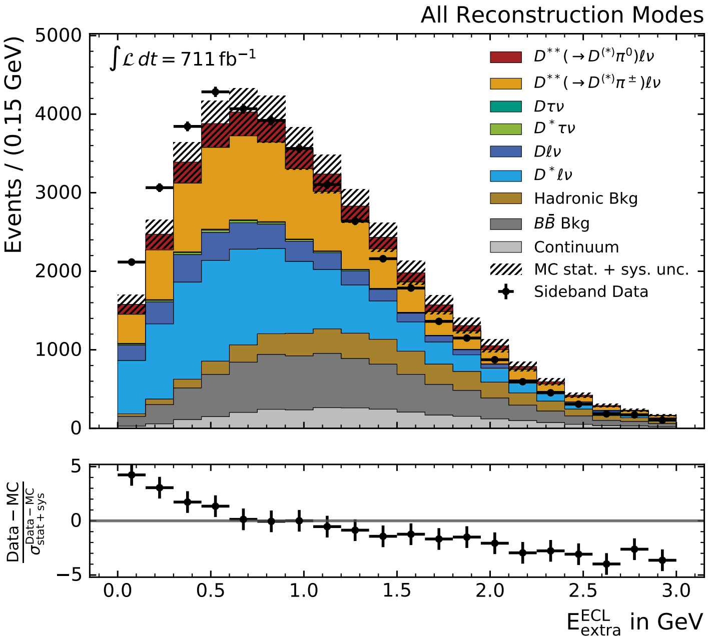


Figure 3.6.: Comparison of data and MC on the extra energy in the calorimeter $\mathrm{E}^{\mathrm{ECL}}_{\mathrm{extra}}$. The plot is based on the q^2 sideband of the underlying $\mathcal{R}(\mathrm{D}^{(*)})$ analysis. Large discrepancies can be seen in the low energy region of $\mathrm{E}^{\mathrm{ECL}}_{\mathrm{extra}} < 0.6\,\mathrm{GeV}$.

the ROC curve of the BDT. As can be seen in the figure, the signal and the background cannot be separated as well as in the continuum suppression, but the classifier still fulfilled his goal of enriching the signal decays. At the chosen working point of $\mathcal{P}_{\mathrm{sel}}$, which is later used as an analysis cut, about 44 % of the background gets rejected, while keeping over 83 % of the signal. Hence, more background is rejected than signal is lost, which should improve the significance of the fit in the end.

Analysis Cuts

The list of variables used for the separation of signal and background events is expanded by a set of variables on which cuts are applied. Some of these were already used in previous stages of the analysis, but at this stages cuts on them are introduced or tightened to get a better separation between the D^{**} events and the background. In the following the variables and their cut values are described.

FEI probability $\mathcal{P}_{\mathbf{FEI}} > 0.001$

Cutting on the FEI probability of the tag-side B meson rejects events containing an imperfectly reconstructed tag-side.

Extra energy in the calorimeter $\mathbf{E}^{\mathbf{ECL}}_{\mathbf{extra}} < 3.0\,\mathbf{GeV}$

An event, which was well reconstructed, should have used nearly all ECL clusters during the reconstruction. So the number of remaining clusters in the ROE and, consequently, the sum of their energies should be small. Large values for $\mathrm{E}^{\mathrm{ECL}}_{\mathrm{extra}}$

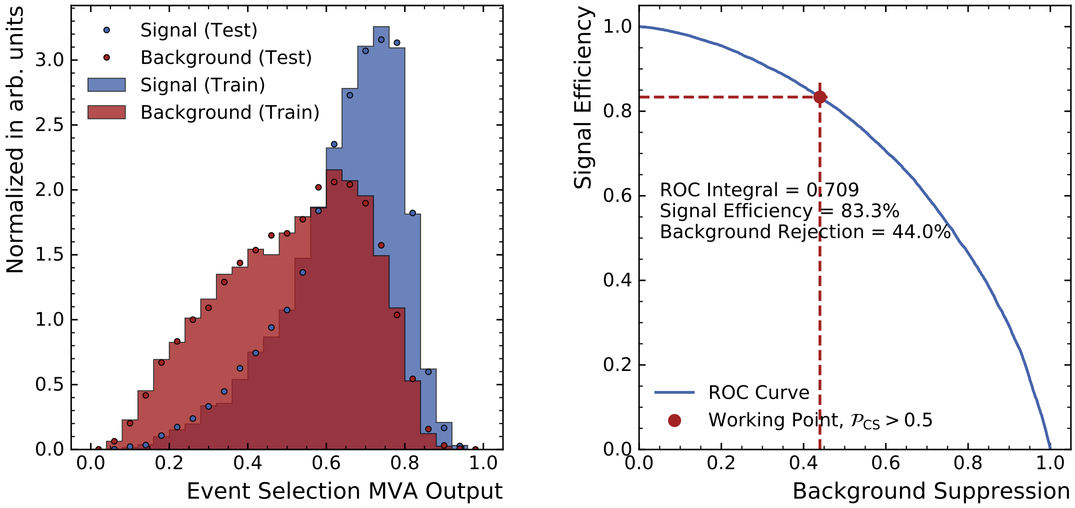


Figure 3.7.: Left: Distribution of the selection output variable $\mathcal{P}_{\mathrm{sel}}$, calculated by the multivariate classifier. The output is shown for the training (bars) and the test sample (dots). Right: ROC curve for the analysis selection BDT. The curve is evaluated on the independent test data set. Signal Efficiency and Background Rejection values are calculated for a working point of $\mathcal{P}_{\mathrm{sel}} > 0.5$.

indicate that something was left behind during the reconstruction and so the event was possibly not reconstructed correctly.

$\mathrm{D}^{(*)}$ meson mass deviation $|\Delta \mathrm{M}(\mathrm{D}^{(*)})| < 35\,\mathrm{MeV}$

This variable describes the difference between the nominal and the reconstructed mass of the $\mathrm{D}^{(*)}$ meson. For correctly reconstructed $\mathrm{D}^{(*)}$ the difference should of course be small, which motivates this cut.

Beam constrained mass of the tag-side $\mathrm{M}_{\mathrm{bc}} > 5.27\,\mathrm{GeV}$

As already described in Section 3.2.2, the beam constrained mass is a very powerful variable to suppress background caused by poorly reconstructed $\mathrm{B}_{\mathrm{tag}}$ mesons. The cut is chosen slightly below the peak value of $\mathrm{M}_{\mathrm{bc}} = 5.28\,\mathrm{GeV}$, which is the B meson mass.

Continuum suppression output $\mathcal{P}_{\mathrm{CS}} > 0.5$

A cut on $\mathcal{P}_{\mathrm{CS}}$ was already applied directly after the training to reduce the data that has to be processed. To suppress even more continuum background, a stricter cut is applied at this stage.

Lepton momentum $\mathrm{p}_{\ell} < 2.0\,\mathrm{GeV}$

Making a restriction on the lepton momentum helps to further reduce the continuum background. Additionally background from $\mathrm{e}^+\mathrm{e}^- \to \ell^+\ell^-$ events is reduced by limiting the allowed lepton momentum in the laboratory frame to 2 GeV.

Vertex fit result $\mathcal{P}_{\mathrm{Vertex}}(\mathrm{B}_{\mathrm{sig}}) > 0.0$

This cut removes candidates, where the vertex fit of the $\mathrm{B}_{\mathrm{sig}}$ candidate has not provided a successful result.

Selection MVA output $\mathcal{P}_{\mathrm{sel}} > 0.5$

Applying this cut enriches the fraction of signal to background events, as described in the previous paragraph.

3.3.3. Best Candidate Selection

One of the advantages of an $e^{+}e^{-}$–collider over a hadron collider, like LHC, is the fact that there is no pile-up in the former. This fact and the kinematic constraints of the KEKB accelerator make multiple $\Upsilon(4S)$ resonances at the same time very unlikely. Nevertheless, when reconstructing an event it is more likely to find more than one $\Upsilon(4S)$ candidate per event. Some of the candidates are rejected by the analysis selection described before, but there is still the chance of having more than one candidate per event afterwards. The probability of of this scenario is especially high for events, which contain a bremsstrahlung corrected electron, since the original and the corrected candidate are both kept during the reconstruction and the properties of the particle are not that different. For such events, and of course all other events containing more than one candidate, a best candidate selection (BCS) is applied. The selection is based on remaining energy of neutral clusters $\mathrm{nE}_{\mathrm{extra}}^{\mathrm{ECL}}$, originating from clusters not associated to a charged track, in the ROE of the $\Upsilon(4S)$. Therefore, the candidate with the lowest value of $\mathrm{nE}_{\mathrm{extra}}^{\mathrm{ECL}}$ is selected. In the example of the two candidates, with and without bremsstrahlung correction, always the corrected candidate is favored, since $\mathrm{nE}_{\mathrm{extra}}^{\mathrm{ECL}}$ is reduced since some of the energy is recovered by adding the bremstrahlungs photon to the electron.. For Candidates which share the same value of $\mathrm{nE}_{\mathrm{extra}}^{\mathrm{ECL}}$ an additional selection based on $|\Delta\mathrm{E}_{\mathrm{tag}}|$ (see Eq. (3.2)) is performed. If no distinctive candidate can be selected based on these two variables, a random candidate of the remaining ones is chosen.

4. Monte Carlo Corrections

Simulated MC will never be perfect. Therefore the production of MC is a process of continuous improvement. Since the Belle experiment stopped taking data in 2010, the last production of Belle MC took place around the same time. Science, however, did not stand still since then, and so a lot of new branching fractions measurements were done, which makes many of the parameters used to produce the Belle MC outdated. Also the models used for producing the MC are imperfect and got replaced by improved ones, as e.g. done for the Belle II MC production. For this analysis, to keep the MC as updated as possible, a lot of corrections, like form factor or branching ratio corrections, are applied.
Another issue that makes MC corrections necessary, are inconsistencies in the reconstruction efficiency between MC and data, due to an imperfect detector simulation. Examples for such corrections are the PID corrections or the tagging correction.

4.1. PID Corrections

The particle identification for leptons and hadrons at Belle shows discrepancies between MC and data. To take this into account, corrections are applied on the MC samples. The errors, associated to this correction, are treated as additional systematic uncertainties in the signal extraction.
The PID corrections used in this analysis are split into three parts, which are described below.

Lepton PID Corrections

The discrepancy in the lepton identification efficiency between data and MC has been studied in [33]. This study provides correction tables for several bins of the polar angle Θ, the lepton momentum in the lab frame $\mathrm{p_{Lab}}$, and different lepton ID cuts. The corrections have been calculated by investigating the two-photon process, $\mathrm{e^+e^-} \to \mathrm{e^+e^-}\ell^+\ell^-$. A possible influence of a hadronic environment is also investigated by comparing the two-photon process to inclusive decays $\mathrm{B} \to X\mathrm{J}/\psi(\to \ell^+\ell^-)$, which delivers an additional error on the correction.
This analysis applies the lepton PID correction on all correctly reconstructed electron and muon tracks, based on the reconstructed momentum and angle Θ.

Lepton Fake Rate

Similar to the lepton PID corrections the mis-identification rate of hadrons to leptons was studied for all Belle analysis [34]. The study provides correction factors and

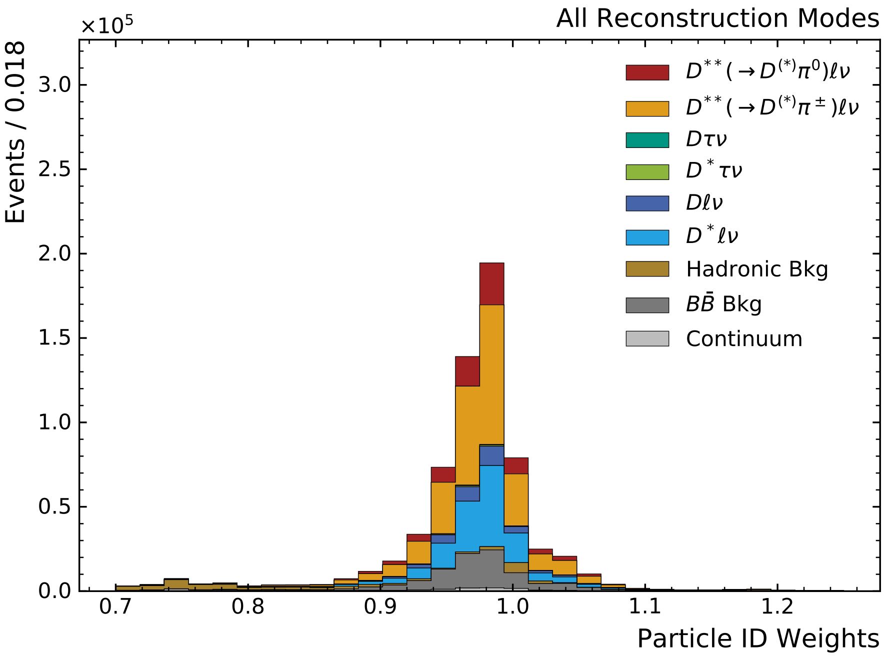


Figure 4.1.: Distribution of the combined PID weight in the selected MC sample.

uncertainties in eleven lepton momentum and eight polar angle bins. These correction factors are applied to all hadrons, which were wrongly identified as electrons or muons.

Hadron PID Corrections

Using the decay $\mathrm{D}^{*+} \to \mathrm{D}^0\pi^+$, with $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$, correction factors for the efficiency and mis-identification rate of kaons and pions are estimated in [35]. The study provides correction tables for nine different PID cuts in the range from 0.1 to 0.9. These correction tables are generated by fitting the $\Delta\mathrm{M} = \mathrm{M}(\mathrm{D}^*) - \mathrm{M}(\mathrm{D})$ distribution in a total of 384 bins (32 momentum and 12 polar angle bins).

In this analysis the hadron PID correction is applied on all kaon and charged pion tracks on the signal-side, using the information from these correction tables.

All described weights are combined into a final PID correction weight, by multiplying them for each event. The resulting correction is shown in Fig. 4.1. Additionally, their uncertainties can be considered independently to vary the effect of every single one individually.

4.2. Branching Fraction Corrections

As stated before, the production of the official Belle MC lies back around ten years from the time of writing. Since then a lot of new branching fraction measurements have been performed. The most relevant part of the MC used in this analysis are events containing $\mathrm{B} \to \mathrm{D}^{(*)}\ell\nu$ and $\mathrm{B} \to \mathrm{D}^{(*)}\pi\ell\nu$ decays, with the $\mathrm{D}^{(*)}$ decaying into any combination of particles, which was used for the Belle MC. To keep the Belle MC as up to date as possible, in this analysis weights are applied to each event, depending on the contained B and $\mathrm{D}^{(*)}$ decays. The two corrections are described below.

Table 4.1.: List of branching fractions used to produce the Belle MC and updated ones used for the reweighting in dependency of the decay.

		$\mathcal{B}^{MC}(\times 10^{-2})$	$\mathcal{B}^{Belle\,II}(\times 10^{-2})$	Weight
B^0	$\rightarrow D^{*-}\ell^+\nu_\ell$	5.33	5.11	0.9587
	$\rightarrow D^{*-}\tau^+\nu_\tau$	1.42	1.25	0.8803
	$\rightarrow D^-\ell^+\nu_\ell$	2.13	2.14	1.0047
	$\rightarrow D^-\tau^+\nu_\tau$	0.71	0.64	0.9014
B^+	$\rightarrow \overline{D}^{*0}\ell^+\nu_\ell$	5.79	5.49	0.9482
	$\rightarrow \overline{D}^{*0}\tau^+\nu_\tau$	1.54	1.35	0.8766
	$\rightarrow \overline{D}^0\ell^+\nu_\ell$	2.31	2.31	1.0000
	$\rightarrow \overline{D}^0\tau^+\nu_\tau$	0.77	0.69	0.8961

$B \rightarrow D^{(*)}\ell\nu$ Branching Fraction Corrections

The relevant $B \rightarrow D^{(*)}\ell\nu$ background decays can be split into two cases. In the first case, the decays with the lepton being either an electron or a muon are included, since the decays into electrons and muons share the same branching fractions. The decays $B \rightarrow D^{(*)}\tau\nu_\tau$ have a different branching fraction due to phase space and therefore form the second case. Depending on these cases and the charge and excitation of the D meson, the branching fractions are updated and a correction weight is calculated via

$$\text{weight} = \frac{\mathcal{B}(B \rightarrow D^{(*)}\ell\nu)^{new}}{\mathcal{B}(B \rightarrow D^{(*)}\ell\nu)^{old}} \tag{4.1}$$

with the updated branching fraction $\mathcal{B}(B \rightarrow D^{(*)}\ell\nu)^{new}$ taken from the current Belle II implementation (status as of August 2020). A complete list of the $B \rightarrow D^{(*)}\ell\nu$ branching fractions used to produce the Belle MC and the corresponding updated values are shown in Table 4.1. The calculated weights are also given.

The correction is only applied on the simulated signal side B mesons, because the branching fraction of the tag side B meson should be corrected by the calibration of the FEI. To decide whether the simulated B meson was reconstructed on the tag or the signal side, matching it with the reconstructed B meson is necessary. This leads to the fact that only the ones, where the matching worked, can be corrected. One could also get rid of the matching during the reweighting of the $B \rightarrow D^{(*)}\ell\nu$ branching fractions by applying the correction to both the signal and the tag-side B meson. This would then require a new FEI calibration and is therefore not considered for this analysis.

Since the signal $B \rightarrow D^{**}\ell\nu$ decays are simulated separately with already updated branching fractions (see Section 3.1), there is no additional correction applied on the branching fraction of these decays.

$D \rightarrow X$ Branching Fraction Corrections

Similar to the branching fractions of the $B \rightarrow D^{(*)}\ell\nu$ decays, the knowledge about

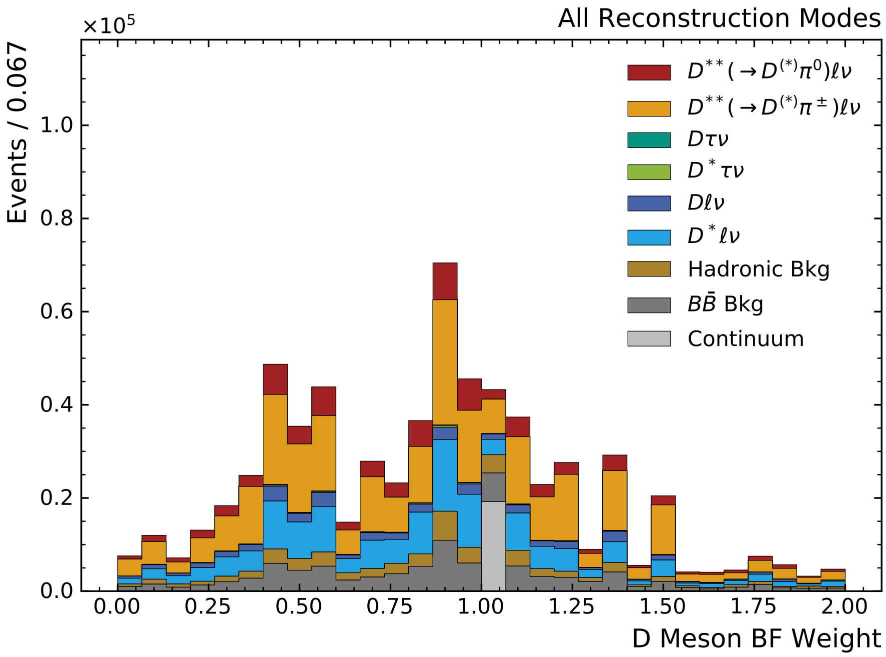


Figure 4.2.: Distribution of the $\mathcal{B}(\mathrm{D} \to ...)$ weight in the selected MC sample.

branching fractions used to simulate the D decays in the Belle MC, has also changed. Therefore an additional correction factor is needed. Although the general concept is the same as for the reweighting of the B meson branching fractions described before, the calculation of the weights is slightly different.
For calculating the correction weights in each event the following formula is used:

$$\text{weight}_i = \frac{\mathcal{B}(\mathrm{D} \to ...)_i^{new}}{\mathcal{B}(\mathrm{D} \to ...)_i^{old}} \cdot \frac{\sum \mathcal{B}(\mathrm{D} \to ...)^{old}}{\sum \mathcal{B}(\mathrm{D} \to ...)^{new}} \tag{4.2}$$

where the first part is similar to the B case, while the second part ensures that the sum of branching fractions of the considered decays remains unchanged. The updated values $\mathcal{B}(\mathrm{D} \to ...)_i^{new}$ are taken either from the current *pdg* world average [24], or, if no value can be found there, from the branching fraction assumptions made for the Belle II MC simulation. In the cases where no updated value could be found, the resulting weight was set to one, so the corresponding decay remains untouched in the MC. One could argue that such decays should be removed from the MC, by setting their weight to zero, but this would cause the problem that the occurring gap would have to be filled with new decays. This would make a reproduction of a lot of the simulated MC necessary and is therefore not achievable for this analysis.
Fig. 4.2 shows the distribution of the calculated weights in the final sample used for the fitting.

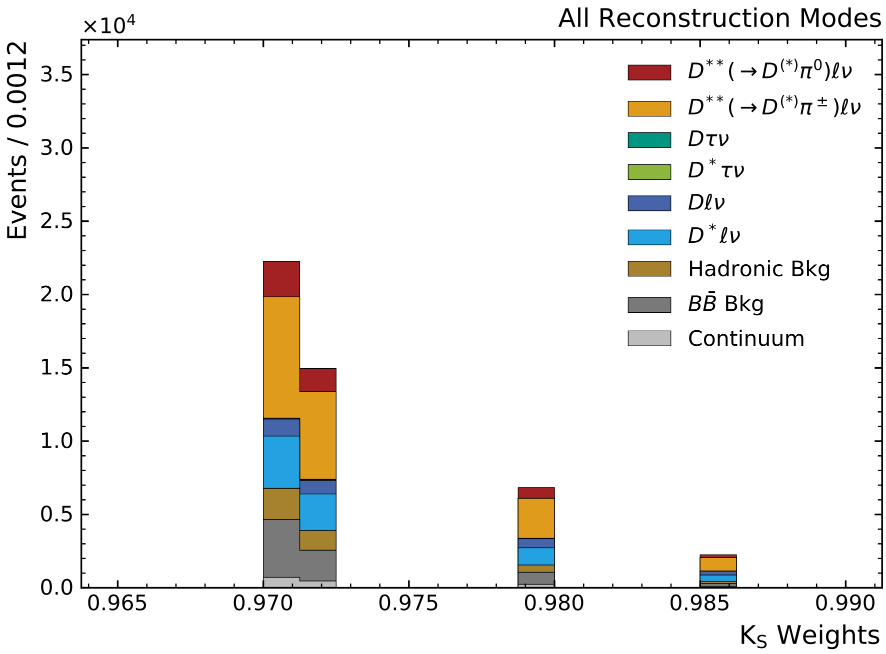


Figure 4.3.: Distribution of the K_S^0 correction weight in a subset of the selected MC sample. The subset only contains events, which were corrected.

4.3. K_S^0 Efficiency Correction

Some reconstruction modes of D meson decays used for this analysis involve K_S^0 mesons. Therefore Belle has observed differences in the reconstruction of K_S^0 mesons when analyzing MC compared to the performance on data. To consider this discrepancy, correction weights are provided by the Belle collaboration, which were obtained by a study given in [36].
A clean sample of fully reconstructed D^* decays is used to determine the K_S^0 efficiency. The study then uses the decay mode $D_s \rightarrow (D^0 \rightarrow K_S^0\pi^+\pi^-)\pi_s$, with the kaon reconstructed from two charged pion tracks: $K_S^0 \rightarrow \pi^+\pi^-$. After applying the event selection, the signal yield is extracted by fitting the K_S^0 mass distribution. An additional systematic uncertainty on the efficiency is determined by comparing data and MC. Fig. 4.3 shows the distribution of the correction weight applied to the events of this analysis. Since only some of the decay modes actually contain K_S^0, only about 10 % of the events get affected by this correction. The values of the correction weights are very discretely distributed since only seven momentum bins were used to calculate the weights.

4.4. Slow Pion Correction

As for the K_S^0, the reconstruction efficiency of slow pions ($p(\pi_s) < 200\,\mathrm{MeV}$) also has to be taken into account and therefore is represented by an additional correction factor in this analysis. Based on the full Belle data set, the discrepancy in the reconstruction efficiencies between data and MC is studied in [37]. This study reconstructs the decays $B^0 \rightarrow D^{*-}\pi^+$ and $B^+ \rightarrow \overline{D}^{*0}\pi^+$ to investigate the reconstruction efficiency of the slow pion π_s, which is

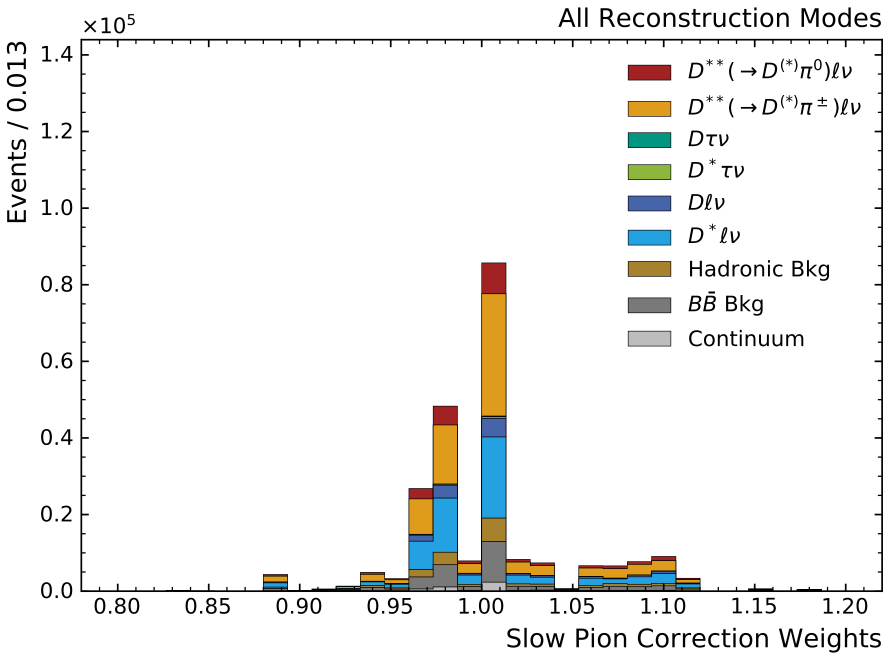


Figure 4.4.: Distribution of the slow pion correction weight. For better visualization, only events, which had a correction applied, are shown. This is satisfied by about 26 % of all selected events.

emitted in the D^* decay. The signal yields are extracted via fits in ΔE and ΔM and a ratio between the efficiencies in data and MC is calculated in bins of the slow pion's momentum. Using these ratios, correction factors depending on the experiment number and the charge of the slow pions are provided by the study. These are applied to the MC in the six provided equal bins of the pion momentum in a range from 50 to 200 MeV. The applied correction weights are shown in Fig. 4.4.

4.5. Tag Correction

For the recombination of the tag side B meson, as described in Section 3.2.2, the FEI is used. The training of the FEI is based on simulated events. A recombination efficiency can be calculated from the number of correctly reconstructed B_{tag} mesons. When comparing this efficiency on MC and data, discrepancies can be observed. These differences are caused by, among other reasons, falsely modeled MC or wrong branching fractions, as well as on imperfect detector simulation. To eliminate the effect of these differences a FEI calibration has to be performed. In this analysis the calibration described in [38] is used. The calibration delivers a overall correction factor and its uncertainties depending on the charge of the reconstructed B_{tag} meson.

$$\bar{\epsilon}_{\mathrm{charged}} = 0.810 \pm 0.012 \pm 0.054$$
$$\bar{\epsilon}_{\mathrm{neutral}} = 0.853 \pm 0.059 \pm 0.058$$

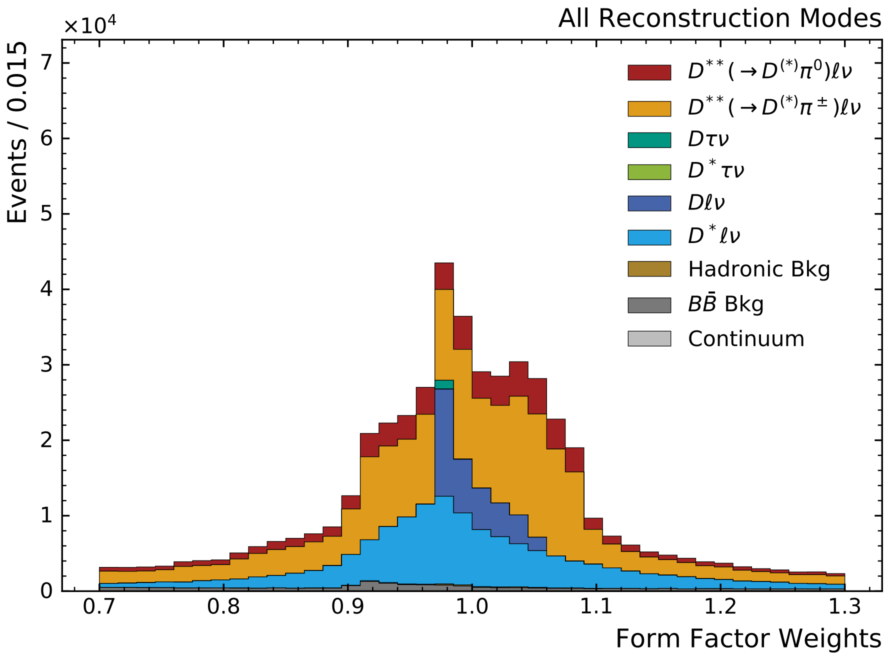


Figure 4.5.: Distribution of the combined form factor weight for all affected events.

4.6. Form Factor Corrections

For the determination of the signal and background yields, a template likelihood fit is performed. For this the templates are generated using the shapes of the Monte Carlo components. This of course includes the $B \to D\ell\nu$, $B \to D^*\ell\nu$, and $B \to D^{**}\ell\nu$ components. Therefore, an exact modeling of the MC components is required. However, just as the branching fractions, the form factor models used for the event generation of the original Belle MC are outdated. In order to cater for this, form factor weights are applied on the affected events.

In the Belle event generation, the $B \to D^{(*)}\ell\nu$ components were simulated based on a form factor parameterization of Caprini, Lellouch, and Neubert (CLN) [39]. The $B \to D^{(*)}\ell\nu$ get reweighted to a model-independent parameterization, which was proposed by Boyd, Grinstein, and Lebed (BGL) [40], using latest results taken from [41].

For the $B \to D^{**}\ell\nu$ component, a reweighting from the ISGW2 [42] to the LLSW [43] parameterization is performed. Each of the four considered D^{**} states is therefore processed individually, to be able to consider their different properties. The parameter values for the LLSW model used for the calculation of the weights, are taken from a fit performed in [6]. For the narrow states, D_1 and D_2^*, four parameters are used $(\tau(1), \tau', \tau_1, \tau_2)$, while for the broad states, D_1' and D_0^*, only three $(\zeta(1), \zeta', \zeta_1)$ are defined. For every of these parameters up and down variations are calculated. These variations can be used to determine systematic uncertainties. For more detailed information about the generation of the form factor weights see [44].

In Fig. 4.5 the resulting form factor weight, after combining each individual weight described above, is shown.

5. Data Composition

This chapter deals with the composition in Monte Carlo after applying the selection. Since the used observable differs on the scope, the variables used for fitting are described and their corresponding data composition is explained in detail. Additionally, other useful observables which were not used for fitting, but are also of importance for this analysis, are described.

5.1. Fitting Variables

For the signal extraction, it is of utmost importance to have variables that provide different shapes for the signal and background components. The choice of the variable depends also on the use case. If one only wants to separate the D^{**} components from the background components, the squared missing mass provides sufficient information, whereas, if also a differentiation between the considered D^{**} excitations (D_1, D_2^*, D_1', and D_0^*) shall also be done, another variable is needed. This differentiation for the narrow D^{**} states can be achieved using the mass difference between the D^{**} and the daughter $D^{(*)}$.

5.1.1. Squared Missing Mass

The squared missing mass M_{miss}^2 is a very powerful variable to describe the kinematics of a candidate since it includes every reconstructed particle in the event and delivers useful information for a missing energy analysis. It is defined via

$$M_{\text{miss}}^2 = (p_{\text{beam}} - p_{\text{tag}} - p_{\text{sig}})^2 \tag{5.1}$$

with $p_{\text{beam}} = p_{e^-} + p_{e^+}$ being the initial four-momentum of the collision and p_{tag} and p_{sig} being the four-momenta of the tag and signal-side B meson, respectively. Because M_{miss}^2 is a Lorentz invariant quantity, it can be calculated in a reference frame of choice. To simplify the equation the center-of-mass frame is chosen. In the center-of-mass frame, where the $\Upsilon(4S)$ is at rest, the initial momentum is consequently $p_{\text{beam}} = 0$. Breaking down p_{sig} into the momenta of the decay products of the B_{sig} then gives

$$M_{\text{miss}}^2 = (-p_{\text{tag}} - p(D^{(*)}) - p(\ell) - p(\pi))^2 \tag{5.2}$$

The missing mass was calculated and studied for three different approaches of how to calculate the four-momentum of the tag-side B meson p_{tag}. Using the four-momentum of

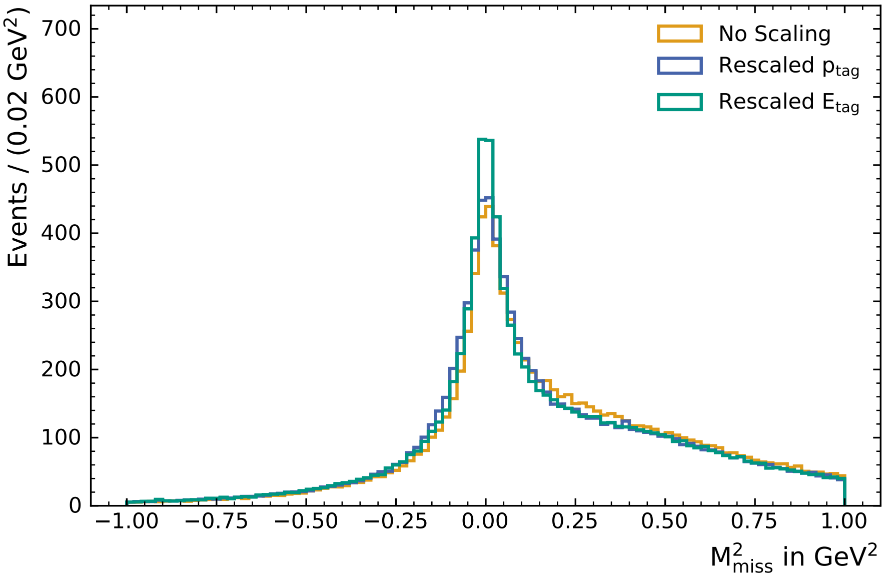


Figure 5.1.: Comparison of the missing mass distribution, depending on the choice of $\mathrm{p_{tag}}$. The curves show the signal components ($\mathrm{B} \to \mathrm{D}^{**}\ell\nu$) in the D^{**} reconstruction modes. For the orange curves, no scaling is applied on $\mathrm{p_{tag}}$, while for the blue and green curve a scaling on the whole four-momentum $\mathrm{p_{tag}}$ or only on the energy component $\mathrm{E_{tag}}$ is applied, respectively.

the $\mathrm{B_{tag}}$ given by the FEI is the most simple approach. Nevertheless, this approach can be improved by exploiting the well-known event kinematics at a B Factory. In the center-of-mass system the energy and magnitude of the momentum of the $\mathrm{B_{tag}}$ are well-known, more precisely $\mathrm{E_{tag}} = 5.29\,\mathrm{GeV}$ (half of the beam-energy) and

$$|\vec{\mathrm{p}}_{\mathrm{tag}}| = \sqrt{\mathrm{E}^2_{\mathrm{tag}} - \mathrm{m}^2_{\mathrm{B}}} \tag{5.3}$$

with $\mathrm{m_B}$ being the nominal mass of the B meson, accordingly. This knowledge is then used to adjust the $\mathrm{p_{tag}}$ which was reconstructed by the FEI. One of the approaches with an adjusted $\mathrm{p_{tag}}$ only adjusts $\mathrm{E_{tag}}$, while the other approach also scales the momentum component $\vec{\mathrm{p}}_{\mathrm{tag}}$ according to Eq. (5.3). In Fig. 5.1 the distribution of $\mathrm{M^2_{miss}}$ is shown for the three approaches for the signal components, which will be described in more detail in the next paragraph. As described in more detail later on in this section, the signal component is expected to peak at $\mathrm{M^2_{miss}} = 0\,\mathrm{GeV}^2$. Since the approach of setting only the energy of the tag-side B meson to the nominal energy shows the best resolution for the peak, this approach is chosen for the analysis.

Components in $\mathbf{M^2_{miss}}$

The events in the MC samples can be classified into several components based on generator level information. For this analysis there are in general two signal components which are components where the $\mathrm{B_{sig}}$ decayed into $\mathrm{D}^{**}\ell\nu$. At this stage we differentiate these decays only by the charge of the pion the D^{**} decays into. The identification is done by matching the reconstructed lepton to the MC lepton and evaluating the related $\mathrm{B}_{\mathrm{sig}}$ MC decay. In the following, each component is described.

$\mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi^0)\ell\nu$

In the first signal component, all events where the reconstructed lepton was matched to a $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ decay are included. The D^{**} decays into a $\mathrm{D}^{(*)}$ and a neutral pion. There is no differentiation made between the different D^{**} types since they cannot be differentiated based on their missing mass distributions. Depending on the reconstruction mode, this component is either a signal or peaking background component.

$\mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi^{\pm})\ell\nu$

Similar to the previous component described before, this component contains $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ decays and functions as a second signal component. The only difference is the charge of the pion. For D^{**} reconstruction modes, where this component is treated as signal ($\mathrm{B} \to \mathrm{D}^{(*)}\ell\pi^{\pm}$), the $\mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi^0)\ell\nu$ component represents the peaking background, and the other way around.

$\mathrm{D}^{(*)}\tau\nu$

This combines two components ($\mathrm{D}\tau\nu$ and $\mathrm{D}^*\tau\nu$), where the reconstructed lepton was matched to an electron or muon, which originated from a leptonic τ decay. For the underlying $\mathcal{R}(\mathrm{D}^{(*)})$ analysis these denote the signal components, while in this D^{**} analysis it is treated as background.

$\mathrm{D}^{(*)}\ell\nu$

The normalization components of the $\mathcal{R}(\mathrm{D}^{(*)})$ analysis provides other backgrounds with different behavior than the τ components, due to the number of involved neutrinos.

Hadronic Background

Mis-identification of charged hadrons, like kaons or pions, as leptons during the reconstruction is an additional background source. All events, that contain a reconstructed fake lepton, which was matched to such a misidentified hadron, are classified as hadronic background.

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

All the remaining events, which have not fulfilled one of the above requirements and can also not be classified as continuum background, are collected in this component. This also includes events, where the MC matching was not successful.

Continuum

The last component contains any events, which originated from $\mathrm{e}^+\mathrm{e}^- \to q\bar{q}$ continuum events and were not already rejected by the continuum suppression or the analysis selection.

All of the described components can also be seen in Fig. 5.3.

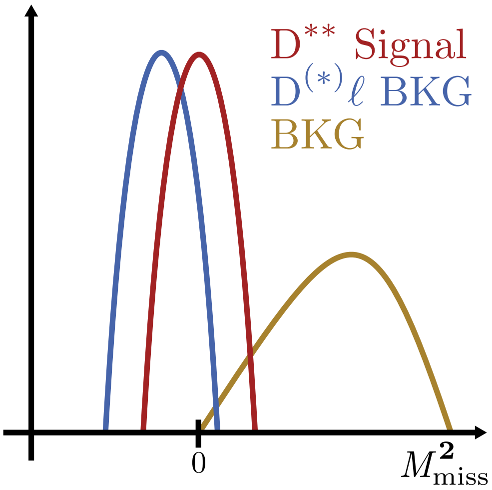


Figure 5.2.: Schematic view on the squared missing mass distribution of the different components. The signal components are drawn in red, while the background components originating from $B \to D^{(*)}\ell\nu$ are shown in blue. The BKG sums up any remaining backgrounds, including the τ components.

Shape of M^2_{miss}

The squared missing mass was chosen as a fitting variable, because the distribution of the events depends on the component. This fact is later exploited in the template likelihood fit. In the D^{**} modes, in which the B_{sig} was reconstructed via $B \to D^{(*)}\ell\pi$, the B decay of the signal component was reconstructed correctly, with only the neutrino missing. Therefore, the signal is expected to peak at $M^2_{\text{miss}} = 0\,\text{GeV}^2$. Additionally, the signal distribution shows a tail to higher squared missing mass values due to a not properly reconstructed tag-side and the limited squared missing mass resolution. The $D^{(*)}\ell\nu$ components show a different signature. In these components an additional pion was reconstructed, which was not a true daughter of the B meson. Using Eq. (5.2) this leads to a shift in the squared missing mass to negative values since the momentum of the signal B meson is overestimated. Due to the higher number of neutrinos, the missing mass in the $D^{(*)}\tau\nu$ components shows a broad spectrum at higher squared missing mass values. The remaining background components also show a broad shape. In Fig. 5.2 a schematic view of the composition of the squared missing mass distribution and its components is given. Additionally the distribution of events in M^2_{miss} after the selection and with the MC corrections applied is shown Fig. 5.3. Besides the general distribution of the MC components, they also show differences depending on the reconstruction mode. By looking at the fraction of signal to peaking background events one can see that in the D^{**} reconstruction modes containing a neutral pion (Fig. 5.4) the signal is imitated by a non-negligible amount of $D^{**}(\to D^{(*)}\pi^{\pm})\ell\nu$ peaking background events. On the contrary, the amount of $D^{**}(\to D^{(*)}\pi^0)\ell\nu$ events in the D^{**} modes containing a charged pion (Fig. 5.5) is nearly negligible. Another notable feature is that in the modes in which the D^{**} was reconstructed via a D^* the number of $D\ell\nu$ background events nearly vanishes. The $D^{(*)}\tau\nu$ components are only present in very low fractions.

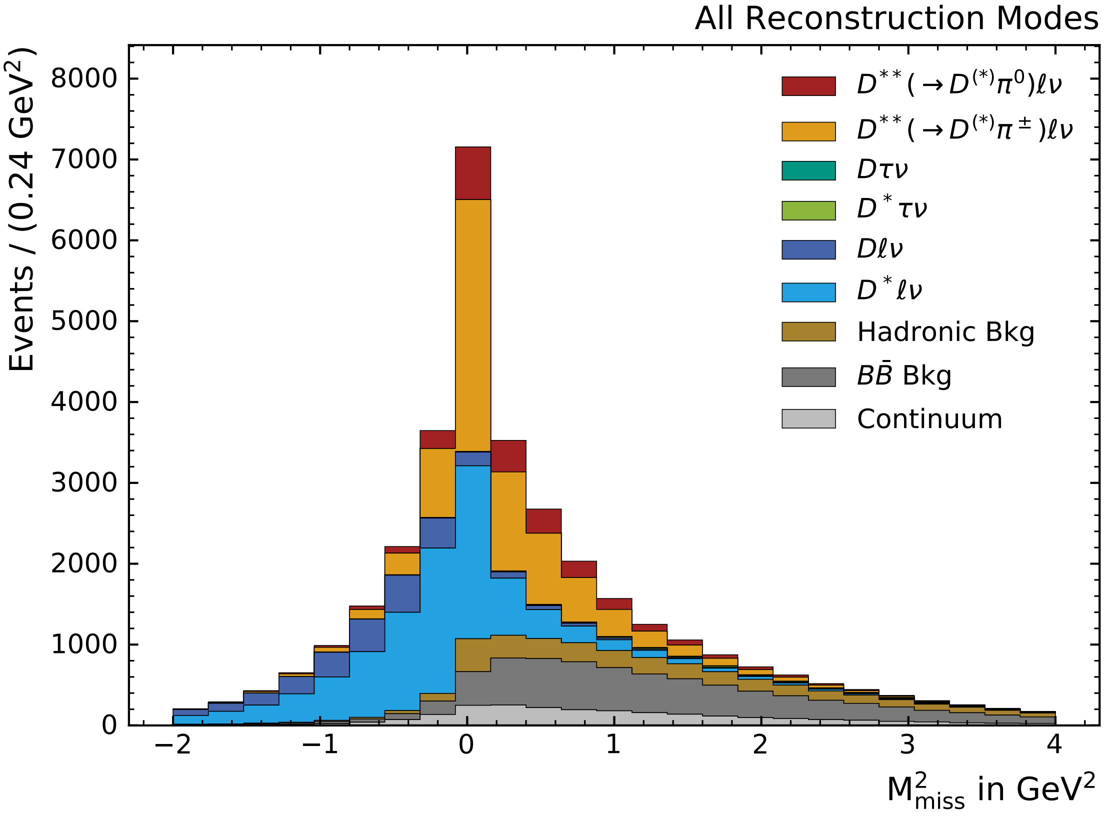


Figure 5.3.: Distribution of the MC components in the squared missing mass. The plot combines all D** reconstruction modes and all MC correction weights described in Chapter 4 are applied.

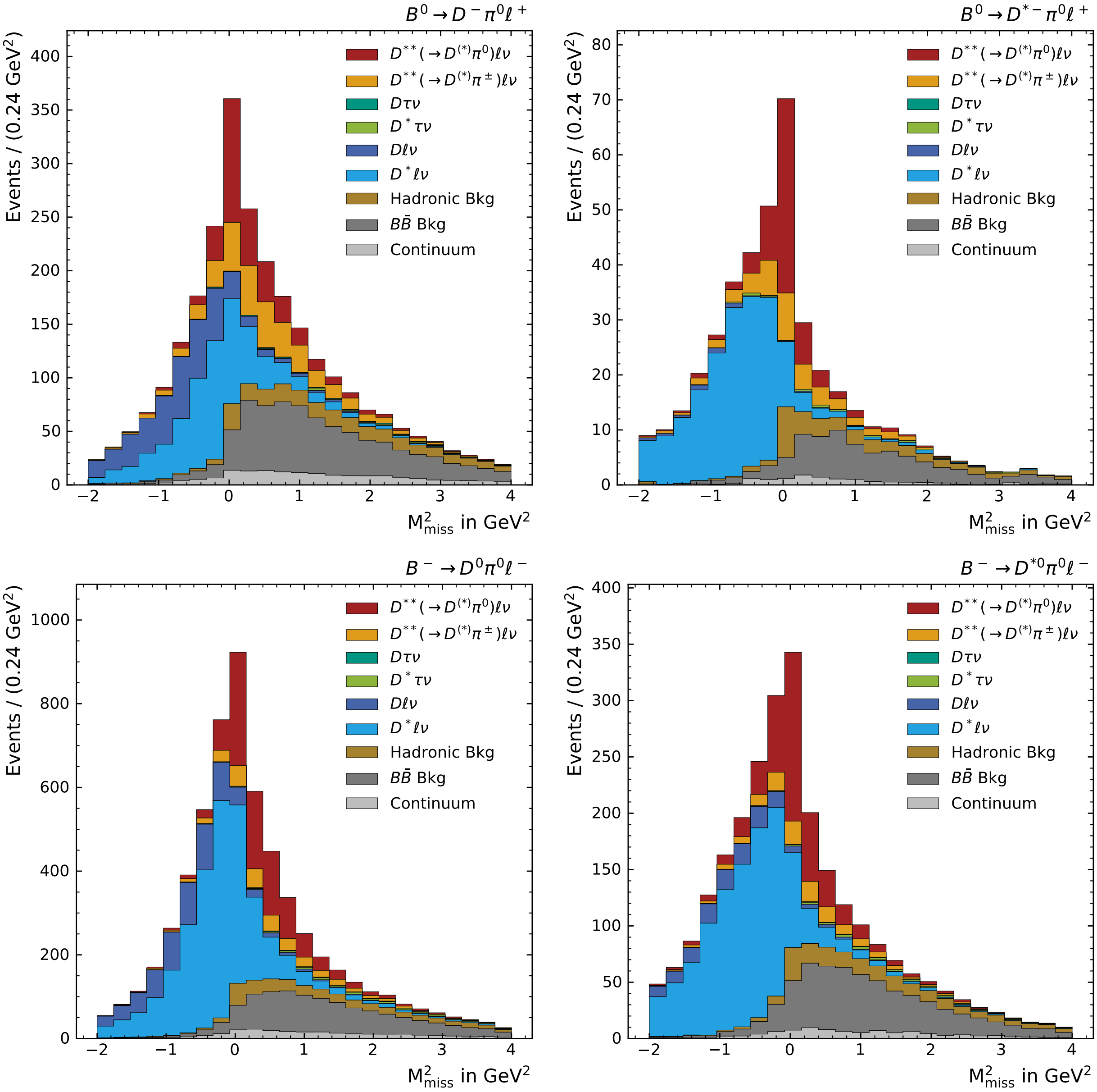


Figure 5.4.: Squared missing mass distributions of the D** reconstruction modes including a neutral pion.

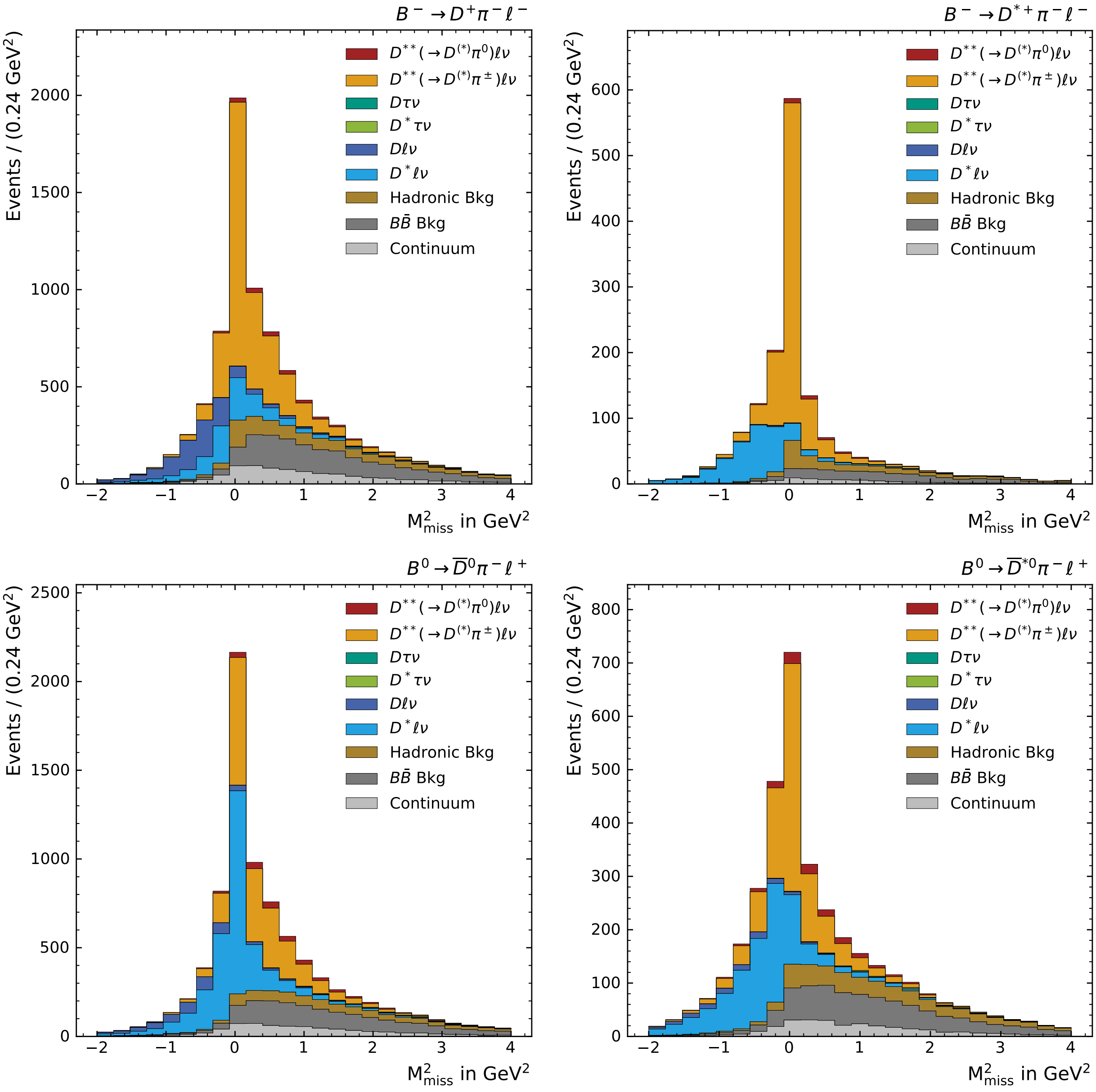


Figure 5.5.: Squared missing mass distributions of the D^{**} reconstruction modes including a charged pion.

5.1.2. D Mass Difference

If one not only wants to study all D^{**} at once, but also wants to make conclusions about the single excitations D_1, D_2^*, D_1', and D_0^*, the observable representing the mass difference between the D^{**} and its decay product $D^{(*)}$ provides the required discrimination power. The mass difference is defined as

$$\Delta M = M(D^{**}) - M(D^{(*)}) \tag{5.4}$$

using the reconstructed invariant masses of the different D mesons. Since the D^{**} excitations are not explicitly reconstructed in this analysis Eq. (5.4) changes to

$$\Delta M = M_{\text{inv}}(D^{(*)}\pi) - M(D^{(*)}) \tag{5.5}$$

with $M_{\text{inv}}(D^{(*)}\pi)$ being the invariant mass of the combination of the pion with the $D^{(*)}$, which is calculated using the reconstructed four-momenta of the particles. In the following, a detailed description of the characteristics of ΔM is given.

Components in ΔM

As the goal set for the evaluation of the ΔM observable is different than what was aimed for when studying the squared missing mass, the MC components making up the distribution are defined in a different manner. While for the squared missing mass case there were only two D^{**} components distinguished by the charge of the daughter pion, the D^{**} components are now divided into four different cases, depending on the D^{**} state ($B \to D^{**}\ell\nu$ with $D^{**} = D_1, D_2^*, D_1', D_0^*$). In this splitting no differentiation based on the charge of the daughter pion is made. $B \to D_1(\to D\pi\pi)$ decays, which are contained in the gap MC, are shown as well for comparison with the other D^{**} decays. The remaining non-D^{**} background components, which were described for the missing mass components, are combined into one background component for ΔM.

Shapes of ΔM

Before analyzing the distribution of the components in ΔM, it is worth having a look at the possible $D^{**} \to D^{(*)}\pi$ decays, which are sown in Fig. 5.6. One can separate the D^{**} states into two types based on their width. D_1 and D_2^* have a very small width and are therefore called narrow D^{**} states, while D_0^* and D_1' have a much greater width and are therefore classified as broad states. The D_0^* decays exclusively into a D meson, whereas the D_1 and D_1' are only decaying into an excited D^*, when taking only decays including one pion into account. A special role is played by the D_2^*, which can either decay into a D or D^* via $D_2^* \to D^{(*)}\pi$.

With this knowledge, the observed ΔM distributions can be explained componentwise. Since the broad states have such a great width, they occur as wide bands over the whole range in the ΔM distributions. on the contrary, the narrow D_1 and D_2^* states show nice peaks, for which the positions depend on the state. Due to the fact that the D_1 only decays via $D_1 \to D^*$ into an excited D meson, the D_1 component shows exactly one peak in the ΔM distribution. This peak is located at around $\Delta M = 0.4\,\text{GeV}$, equal to the mass difference between the D_1 and D^*. Similar to this, the D_2^* is also able to decay into a D^*, so the D_2^*

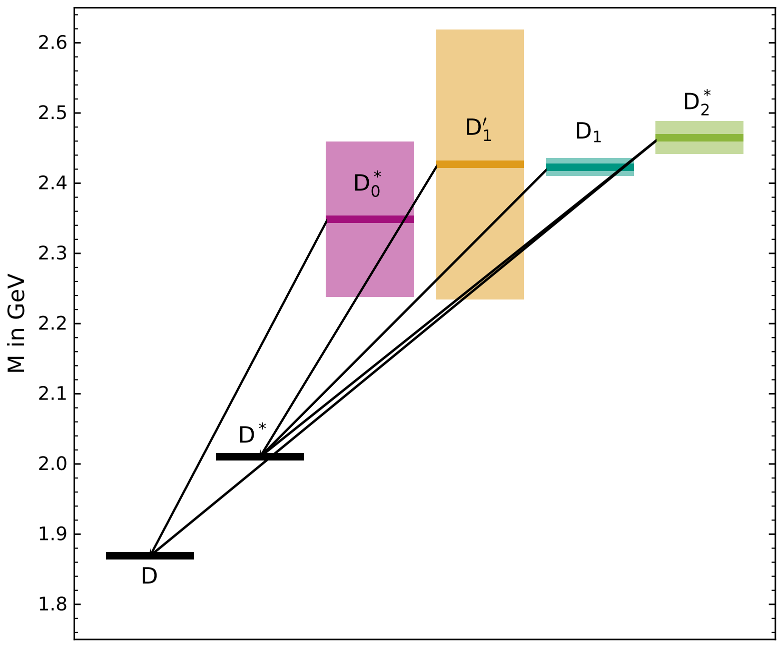


Figure 5.6.: Transitions of the four D^{**} states to $D^{(*)}$ with the emission of exactly one pion. The bands around the D^{**} masses correspond to the widths of the respective D^{**} used for the event generation. Possible transitions are indicated by black lines. Adapted from [25], with updated masses and widths.

component also shows a peak corresponding to the mass difference to the D^*. Nevertheless the peak is slightly shifted to the right, due to the slightly higher mass of the D_2^* compared to the D_1. As the D_2^* can also decay via $D_2^* \to D\pi$ this component shows a second peak, which corresponds to the mass difference to the D meson, with $\Delta M \approx 0.6\,\mathrm{GeV}$.
When looking at the MC composition after the selection, the described distribution of the D^{**} components can clearly be observed in the reconstruction modes with a charged pion (Fig. 5.7 and Fig. 5.8). However, the modes which include an excited D^* meson, show a slightly different form. Since in these modes the reconstruction requires a D^*, the second peak of the D_2^* component originating from $D_2^* \to D\pi$ is missing. For the same reason the D_0^* component is suppressed in these modes, since the D_0^* only decays directly into a D meson.
In the reconstruction modes with a neutral pion (see Fig. 5.9 and Fig. 5.10) the distributions of the components with the described peaks and bands, can also be perceived. Nevertheless, the structure is not so clear in these modes due to lesser statistics, which originates from a poorer π^0 reconstruction efficiency as compared to $\pi^\pm$.

5.2. Other Observables

Not every variable fulfills the requirements to be considered for the fitting. Nevertheless, it is worth having a closer look at some of them to get a better understanding about the MC and effects occurring during the analysis. Some of the variables, like the momentum

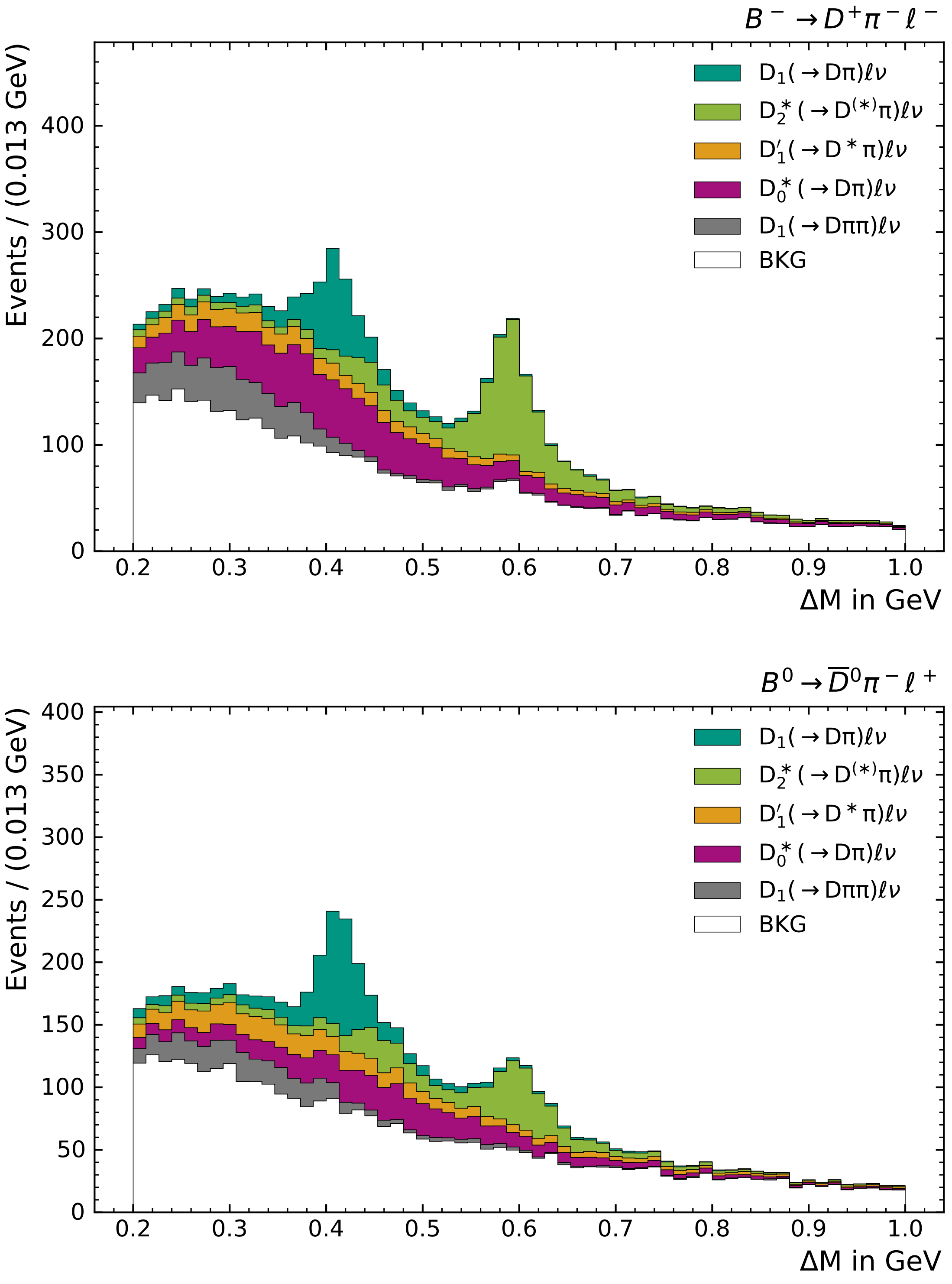


Figure 5.7.: Distribution of $\Delta M = M(D^{**}) - M(D^{(*)})$ in the $B \to D\pi\ell$ reconstruction modes, involving a charged pion.

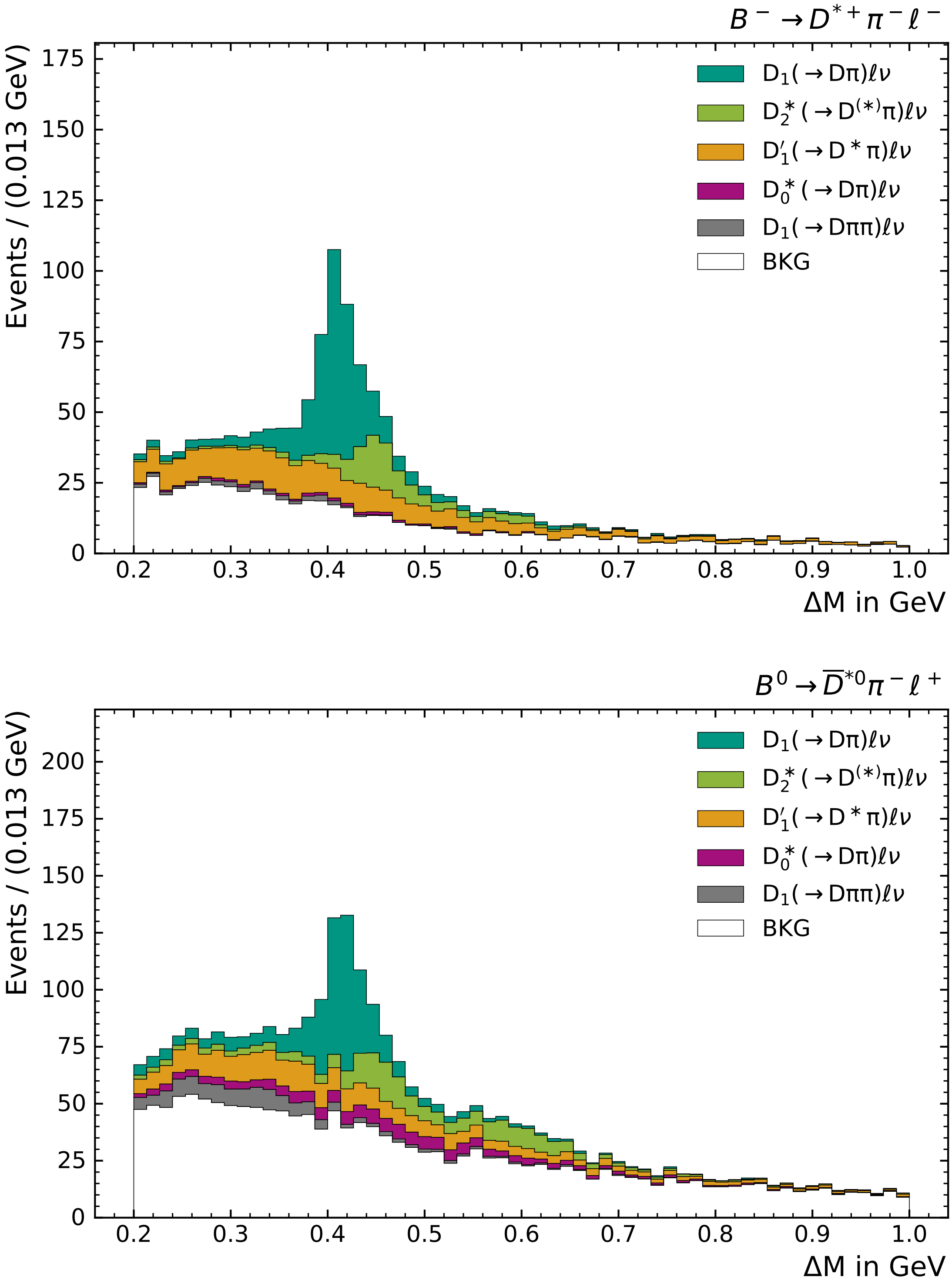


Figure 5.8.: Distribution of $\Delta M = M(D^{**}) - M(D^{(*)})$ in the $B \to D^*\pi\ell$ reconstruction modes, involving a charged pion.

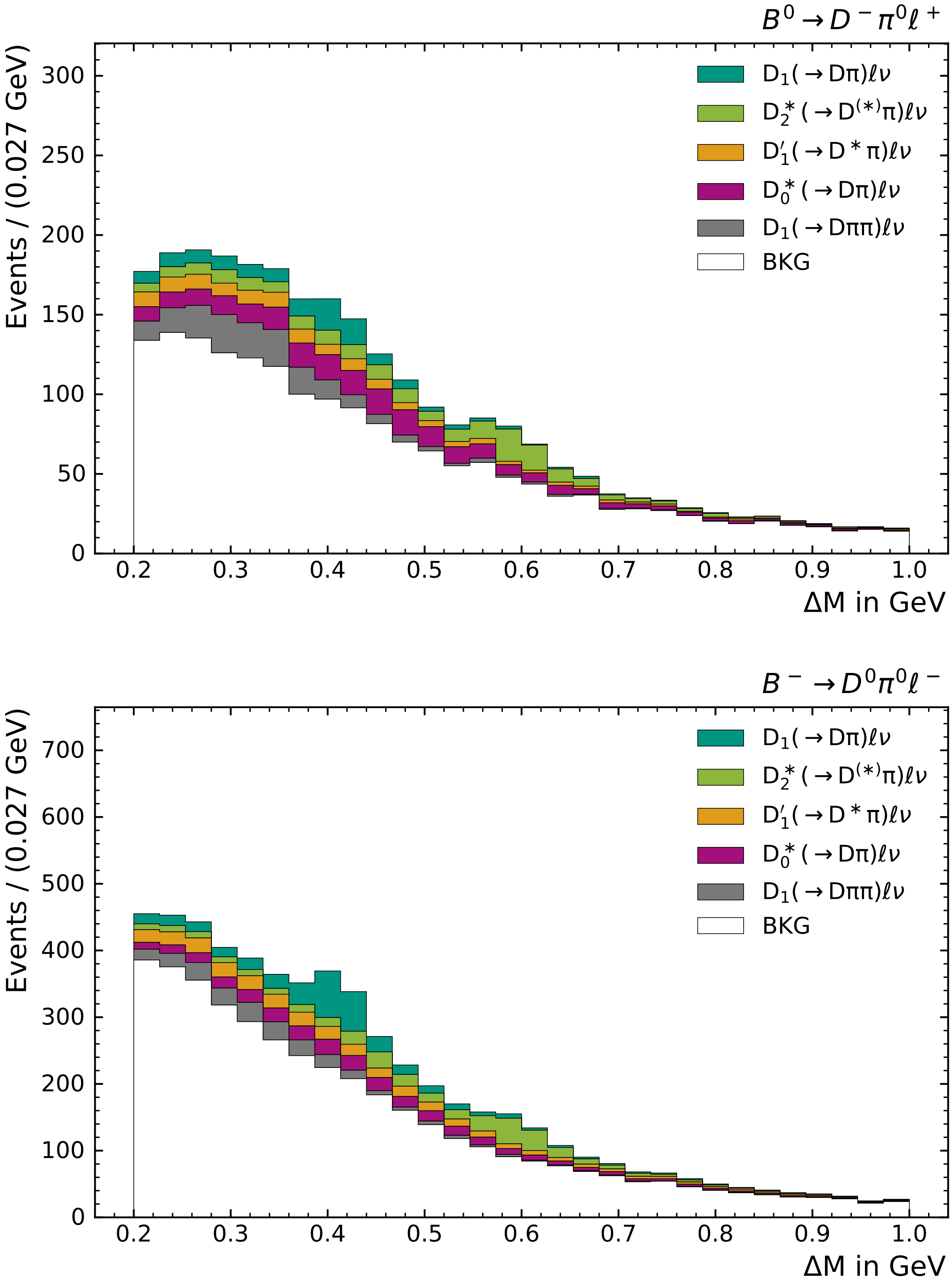


Figure 5.9.: Distribution of $\Delta M = M(D^{**}) - M(D^{(*)})$ in the $B \to D\pi\ell$ reconstruction modes, involving a neutral pion.

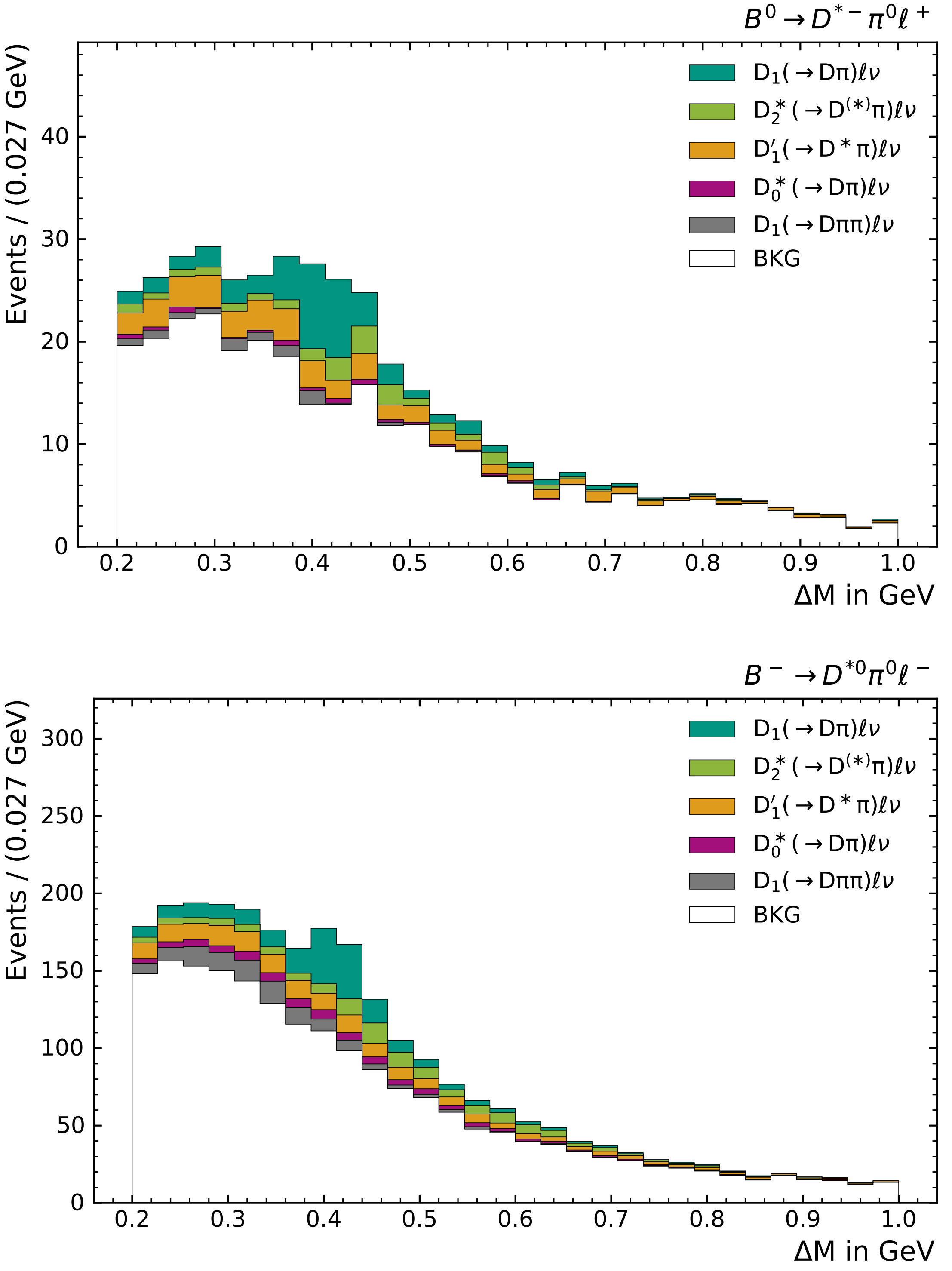


Figure 5.10.: Distribution of $\Delta M = M(D^{**}) - M(D^{(*)})$ in the $B \rightarrow D^*\pi\ell$ reconstruction modes, involving a neutral pion.

transfer, are also needed for studies later on in the analysis. Therefore, the investigated variables are described below.

Momentum Transfer $\mathbf{q^2}$

q^2 describes the momentum transfer from the signal-side B_{sig} meson to the lepton-neutrino pair. The definition of this variable is

$$q^2 = p(W^\pm)^2 = (p(B_{sig}) - p(D^{(*,**)}))^2 \tag{5.6}$$

with the four-momenta of the B_{sig} meson and the leading $D^{(*,**)}$ meson, respectively. As the momentum of the B_{sig} meson is not sufficiently well-known due to the missing neutrino, the event kinematics are again exploited and $p(B_{sig})$ is calculated via

$$\vec{p}(B_{sig}) = -\vec{p}(B_{tag}) \tag{5.7}$$

and the energy set to half of the beam energy, as for the calculation of the squared missing mass. Eq. (5.6) also shows that q^2 equals the squared four-momentum of the W boson from the semileptonic decay. q^2 also shows a strong correlation to the mass of the involved lepton. In the underlying $\mathcal{R}(D^{(*)})$ analysis this is used to favor tau decays over decays involving electrons and muons by requiring higher values for q^2. Such a cut is not performed for this analysis since there is no aim for differentiating tau decays from the lighter leptons. Taking the q^2 spectrum of the D^{**} components into account (Fig. 5.11), one would only lose a lot of statistic by applying such a cut.
Analyzing the q^2 spectrum given in Fig. 5.11 verifies the expectation that the components, which include a τ in their decay chain, can only be observed at higher values of $q^2 > 4\,\text{GeV}^2$. The spectrum of the q^2 observable does not allow for a differentiation between the D^{**} and the background components as well as among the D^{**} components themselves.
For the analysis it is also important to investigate the q^2 resolution. For this the difference of the reconstructed momentum transfer q^2_{Reco} compared to the truth value q^2_{Truth} based on MC information is shown in Fig. 5.12. The distribution of this residual shows two peaks, one at $q^2_{Reco} - q^2_{Truth} = 0\,\text{GeV}^2$, which corresponds to correctly reconstructed B_{sig} mesons. Another peak at $q^2_{Reco} - q^2_{Truth} \approx 1\,\text{GeV}^2$ originates from D^* downfeed. There the pion or photon of the D^* decay was either not found or used somewhere else, so the $D^{(*)}$ could not be reconstructed, which then results in a wrongly reconstructed momentum transfer. Additionally, a tail at negative residual values can be observed. This tail originates from $D^{(*)}\ell$ events which were falsely reconstructed as $D^{**}\ell$ events by adding additional particles which results in a higher q^2_{Reco} value. For one of the studies later on in the analysis it is necessary to divide the data into bins of q^2 Chapter 7. Since most of the events fulfill the requirement $|q^2_{Reco} - q^2_{Truth}| < 2\,\text{GeV}^2$ it is expected that most of the events will be assigned to the correct bin, if bin widths of $2\,\text{GeV}^2$ are chosen.

$\mathbf{E_{miss} - p_{miss}}$

The difference of the missing energy to the missing momentum is another observable from interest, since it is closely related to the squared missing mass via

$$M^2_{miss} = (E_{miss} - p_{miss})(E_{miss} + p_{miss}) \tag{5.8}$$

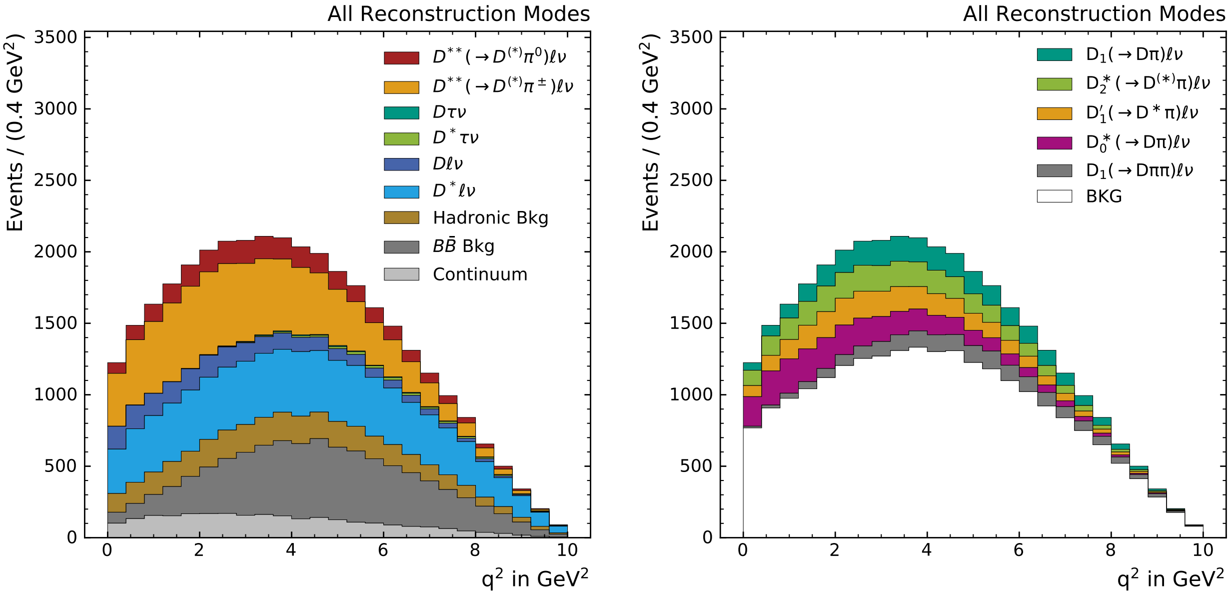


Figure 5.11.: q^2 spectrum of all D^{**} reconstruction modes combined. Left: Spectrum divided into components as used for the squared missing mass fitting. Right: Same spectrum, but split into the different D^{**} states. In both plots the components including tau leptons can only be observed at values of $q^2 > 4\,\text{GeV}^2$.

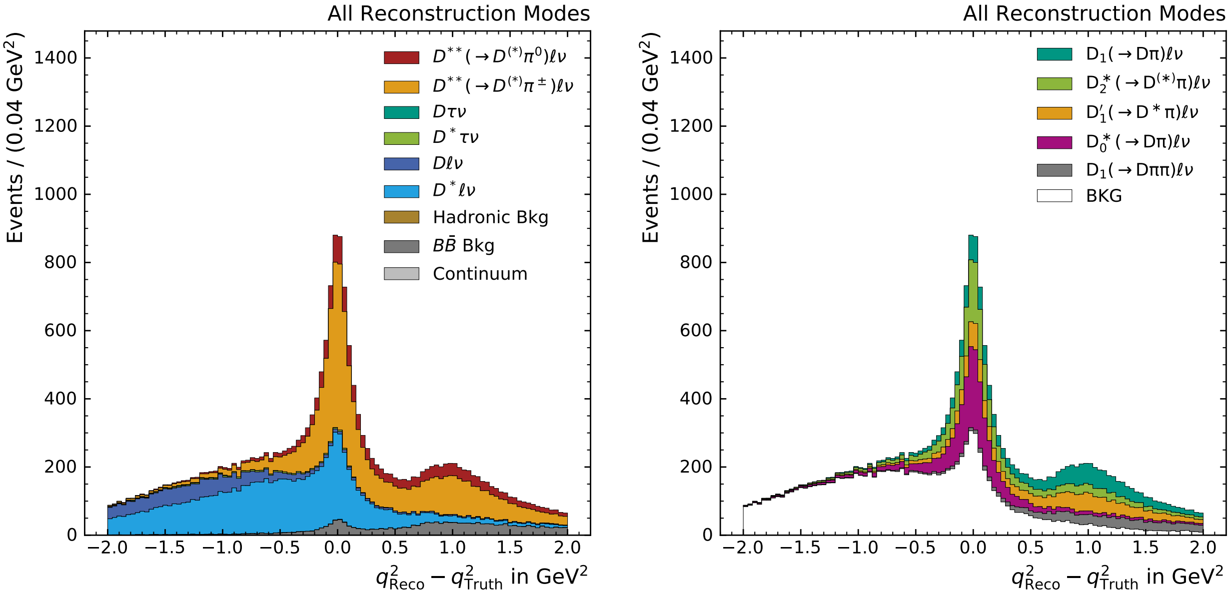


Figure 5.12.: $q^2_{\text{Reco}} - q^2_{\text{Truth}}$ residuals with different component splittings. In the plots all D^{**} reconstruction modes are combined.

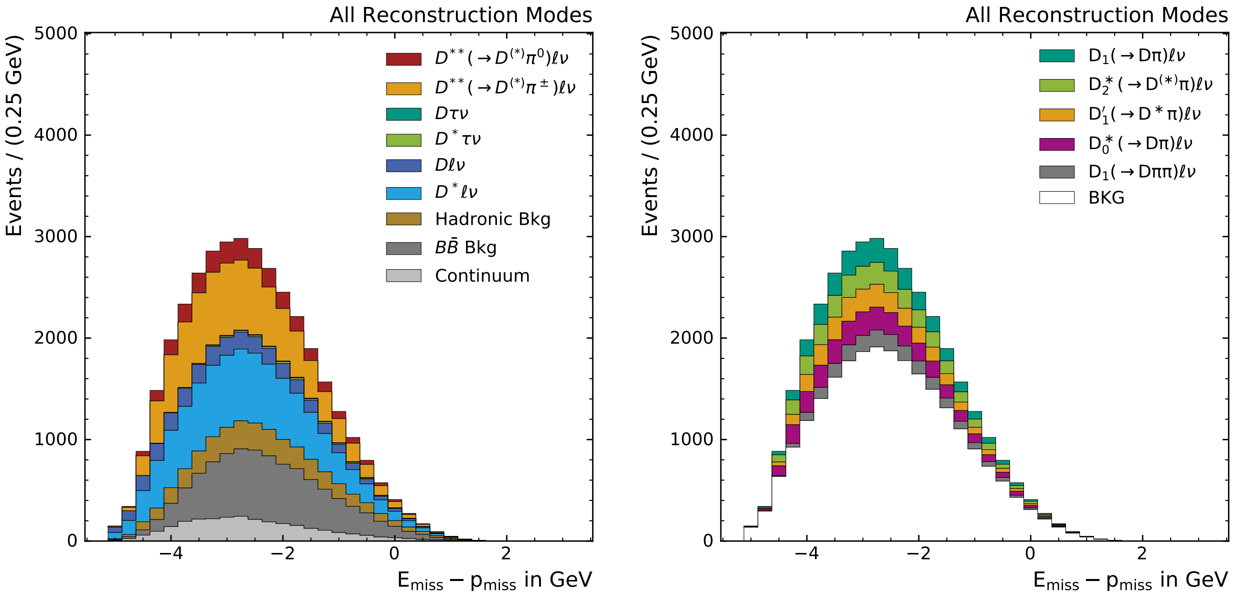


Figure 5.13.: Distributions of $E_{miss} - p_{miss}$. In the plots all D^{**} reconstruction modes are combined.

In this relation, as well as for the general calculation of the observable, E_{miss} denotes the missing energy in an event and p_{miss} the magnitude of the missing momentum vector. Fig. 5.13 shows how the events are distributed for this observable. While the related squared missing mass provides a good separation between the D^{**} signal components and the background, no differentiation can be done using $E_{miss} - p_{miss}$.

Lepton Momentum p_ℓ^*

In the underlying $\mathcal{R}(D^{(*)})$ analysis, the lepton momentum in the rest frame of the signal-side B meson p_ℓ^* will be used together with the missing mass to perform a 2D fit. These observables are chosen, because the combination of the two variables provides a very powerful separation between the $B \to D^{(*)}\tau\nu$ signal and $B \to D^{(*)}\ell\nu$ normalization components, which is needed to measure $\mathcal{R}(D^{(*)})$. Nevertheless, as shown in Fig. 5.14, the separation between the D^{**} signal components and the background (especially $D^{(*)}\ell\nu$ background) is less powerful and therefore the lepton momentum is not used for fitting in this D^{**} study. To further reduce $B\overline{B}$ and hadronic background, one could still require higher lepton momenta p_ℓ^*. For the separation of the D^{**} states among themselves, p_ℓ^* is totally unsuitable, since the spectrum does not show any differences between the single D^{**} components of interest.

Extra Energy in the Calorimeter E_{extra}^{ECL}

The extra energy in the calorimeter was already defined in Section 3.3.2, but for completeness the distribution after the selection and with applied MC reweighting are shown in Fig. 5.15 since E_{extra}^{ECL} was used in other $\mathcal{R}(D^{(*)})$ analyses. For this analysis E_{extra}^{ECL} is only used to reduce background.

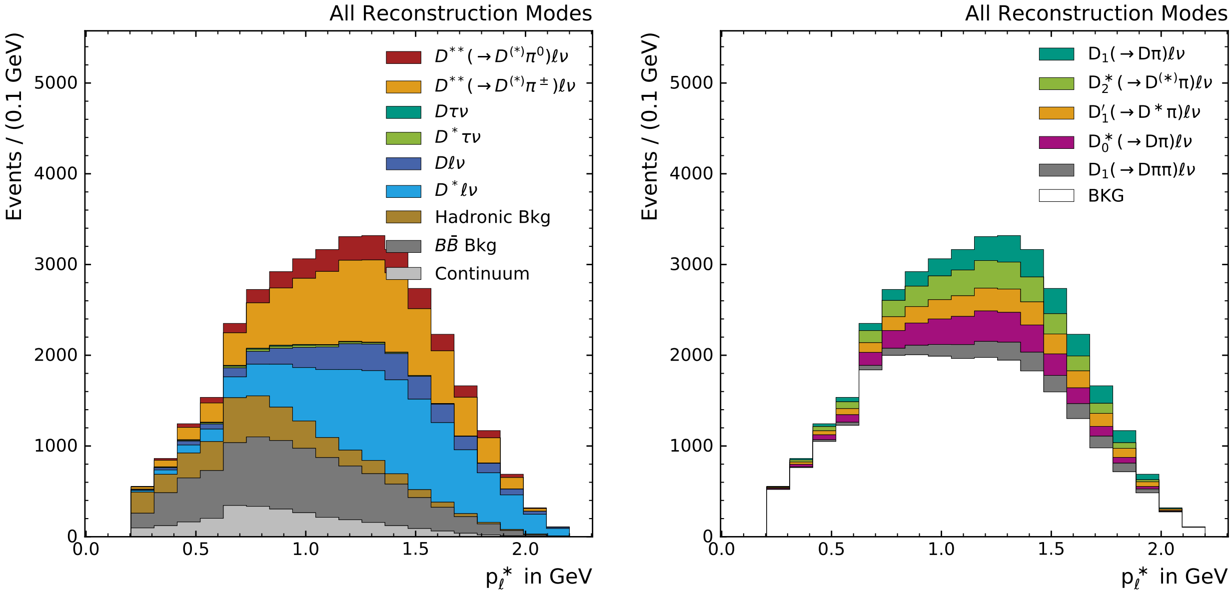


Figure 5.14.: Distributions of the lepton momentum in the B_{sig} rest frame p_ℓ^*, with all D^{**} reconstruction modes combined.

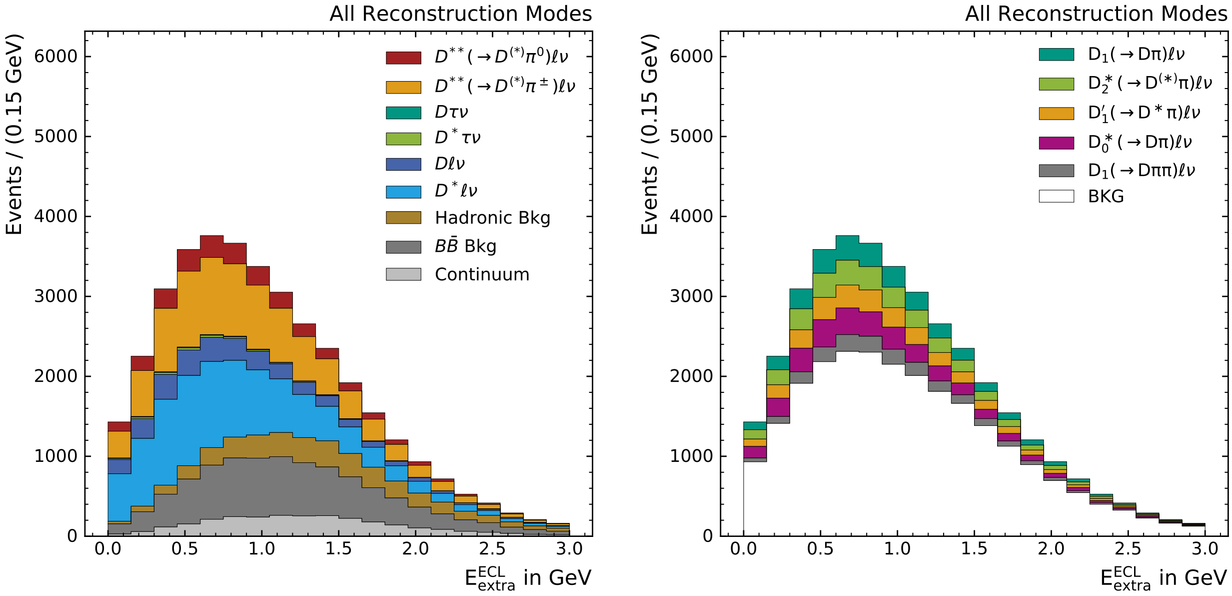


Figure 5.15.: Distributions of the extra energy in the calorimeter E_{extra}^{ECL}, with all D^{**} reconstruction modes combined.

6. Signal Extraction

For the following studies one requires knowledge about the fraction in which the investigated signal events contribute to the whole dataset. To gather the knowledge about the so-called signal yields, a template likelihood fit is performed. This chapter explains this method together with the fundamental statistical concepts the method is based on.

6.1. Maximum Likelihood Fit

As stated before, in this analysis a template likelihood fit is used for the determination of the signal yields. For the general explanations during throughout section References [9] and [45] are used.

General Concept of Maximum Likelihood

Analyzing an observable x for which N measurements $x_1, ..., x_N$ were made, is a common task in particle physics. The distribution of x can be described with a probability density function (pdf) of the form $f(x|\vec{\theta})$, where $\vec{\theta}$ describes an additional set of unknown parameters. Using the finite set of measurements, the method of maximum likelihood allows for the estimation of the unknown parameters $\vec{\theta}$. This is done by first creating a likelihood function L,

$$L = \prod_{i=1}^{N} f(x_i|\vec{\theta}) \tag{6.1}$$

which only has the unknown parameters $\vec{\theta} = \theta_1, ..., \theta_M$ as free parameters since the measurements x_i are fixed. The estimators for the unknown parameters, denoted as $\hat{\theta}_i$, are the values of θ for which the likelihood function is maximized, and therefore can be determined by solving the equation

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, ..., M \tag{6.2}$$

given that L describes a differentiable function.
In practice minimization of functions is more common and since sums are numerically easier

to calculate than products the negative log-likelihood function is used instead of Eq. (6.1) for the minimization. The negative log-likelihood function is given via

$$-\log L(\vec{\theta}) = -\sum_{i=0}^{N} \log f(x_i|\vec{\theta}) \tag{6.3}$$

To save computation time one can, instead of minimizing this function, also use the binned log-likelihood function, defined as

$$-\log L(\vec{\theta}) = -\sum_{i=0}^{N_{bins}} n_i f(y_i|\vec{\theta}) \tag{6.4}$$

with the bin-centers n_i and the number of events in each bin y_i. This is computationally more efficient for larger data sets, since the running time now no longer scales with the number of measurements N, but with the number of bins N_{bins}. Note that this may not be as precise as the unbinned method, but the difference should be negligible.

Often the number of observed events N is treated as a random variable, as well. Using this one can define the extended likelihood function by adding a multiplicative Poisson probability term to the likelihood function (for the binned case)

$$L(\nu, \vec{\theta}) = N! \frac{\nu^N}{N!} e^{-\nu} \prod_{i=1}^{N} f(x_i|\vec{\theta}) \tag{6.5}$$

with the expectation value ν, for which the estimator $\hat{\nu}$ is given by $\hat{\nu} = N$, if ν is not dependent on any other parameter.

Variance of Estimators

For measurements in physics, knowledge about the variance of the estimators is also required. There are multiple options of determining the variance, for example analytically, which is not always possible, so here only the methods used for this analysis are described.

The easiest method is to determine the variances from the inverse of the correlation matrix of the parameters, which is defined as

$$V_{ij}^{-1} = \left(\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \right)^{-1}_{\vec{\theta}=\hat{\vec{\theta}}} = H_{ij} \tag{6.6}$$

and gets evaluated for the estimated values of the parameters. H denotes the Hesse matrix, which is calculated numerically. This only works if the likelihood function can be approximated by a Gaussian distribution. If the approximation is not accurate the method of the profile likelihood is used.

Trying to calculate the uncertainty of one of the parameters in $\vec{\theta}$, denoted as ϕ, the profile negative log-likelihood ratio can be used. The definition is given by

$$\begin{aligned} -2\log\lambda(\phi) &= -2\log\frac{L(\phi, \hat{\hat{\vec{\theta}}})}{L(\hat{\phi}, \hat{\vec{\theta}})} && (6.7) \\ &= -2\log L(\phi, \hat{\hat{\vec{\theta}}}) + 2\log L(\hat{\phi}, \hat{\vec{\theta}}) && (6.8) \end{aligned}$$

with $\hat{\phi}$ being the best estimate for ϕ and the remaining parameters at the estimated values $\hat{\hat{\vec{\theta}}}$. Using this equation the uncertainties can be evaluated via

$$-\log \lambda(\phi + \sigma_{\phi}^{+}) = -\log \lambda(\phi - \sigma_{\phi}^{-}) = \frac{1}{2} \tag{6.9}$$

with the asymmetric uncertainties σ_{ϕ}^{+} and σ_{ϕ}^{-}.

Significance

When publishing a result in particle physics, a significance should be assigned to the measurement to show how good the result is. This can be done by performing a hypothesis test with the background only (H_0) and the signal plus background hypothesis (H_1). For this the Neyman-Pearson-Lemma defines the likelihood ratio as

$$\lambda(\vec{\theta}) = \frac{L(\vec{\theta}|H_1)}{L(\vec{\theta}|H_0)} \tag{6.10}$$

which is still dependent on the unknown parameters $\vec{\theta}$. To get rid of this dependency one can use the log-likelihood ratio, which is defined in Eq. (6.7) [46]. If one is only interested in measuring a positive signal, which can be expressed as $\phi \geq 0$ and would lead to the rejection of the background-only hypothesis, a variable q_0 can be defined as

$$q_0 = \begin{cases} -2\log \lambda(0) & \hat{\phi} \geq 0 \\ 0 & \hat{\phi} < 0 \end{cases} \tag{6.11}$$

This approach covers that the signal hypothesis is not tested against $\hat{\phi} < 0$, which would give evidence for a systematic error rather than for the presence of a signal. With this definition, the p-value can be calculated via

$$p_0 = \int_{q_{0,obs}}^{\infty} f(q_0|0)\mathrm{d}q_0 \tag{6.12}$$

where $f(q_0|0)$ describes the pdf for the background-only hypothesis and $q_{0,obs}$ the observed q_0. Having calculated the p-value, the significance Z of the fit can be determined using

$$Z = \Phi^{-1}(1 - p_0) \tag{6.13}$$

where Φ^{-1} is the inverse of the cumulative distribution of the standard Gaussian. The significance is usually calculated in units of standard deviations σ.
It is common to report a discovery of a signal if a significance of $Z \geq 5\,\sigma$ is observed.

χ^2-Test

To quantify how well a given model describes the measured data, it is common to perform a χ^2-test. With this χ^2-value one can determine the goodness-of-fit [45]. The χ^2 value is defined as

$$\chi^2 = \sum_{i=1}^{N_{bins}} \frac{(n_i - \hat{\nu}_i)^2}{\hat{\nu}_i} \tag{6.14}$$

with n_i being the number of measured events in a given bin and $\hat{\nu}_i$ the number of expected (fitted) ones. In the limit of large samples, this follows the χ^2 distribution with $N_{bins} - k$ degrees of freedom (k denotes the number of unconstrained parameters in the fit). This χ^2 distribution can be used to translate the χ^2 value into a p-value via

$$p = \int_{\chi^2}^{\infty} \chi^2_{N_{bins}-k}(x)\mathrm{d}x \tag{6.15}$$

with χ^2 being the value calculated in Eq. (6.14). The whole concept only holds for samples, which do not contain bins with less than about five entries. Should this be the case, the p-value has to be determined with a Monte Carlo study, where the data is sampled via a Poisson distribution. The test statistic (Eq. (6.14)) is then calculated using the sampled data points.

6.2. Template Likelihood Fit

To measure the signal yields of a given process, the contribution of this process to the whole dataset has to be determined. For this analysis this is done with the extended maximum likelihood method. Nevertheless, since the analytical descriptions of the pdfs are not known, the method has to be extended by using templates. This method exploits the shapes of different histograms, which represent the investigated physical processes, to extract the component yields in the end. MC simulations are used to determine the shapes of the histograms since they are not known in general. The histograms are normalized, so that a component can be described by multiplying the process yield with the normalized histogram. This process yield is treated as a model parameter and its best value is determined by varying all model parameters to get the best agreement between the sum of the template distributions and the measured data. The procedure is also described in [47], where the mathematical explanations are taken from.

It is assumed that the number of expected events in each bin is distributed according to a Poisson distribution. The likelihood function can then be written as

$$L(\vec{n}) = \prod_{i=1}^{N_{bins}} \frac{\nu_i(\vec{n})^{m_i}}{m_i!} e^{-\nu_i(\vec{n})} \tag{6.16}$$

with the measured number of events m_i and the expected ones $\nu_i(\vec{n})$, which depend on the process yields $\vec{n} = n_1, ..., n_M$. For a fit with M processes (templates), these $\nu_i(\vec{n})$ can be calculated using

$$\nu_i(\vec{n}) = \sum_{k=1}^{M} \nu_{ik}(n_k) \tag{6.17}$$

where the ν_{ik} denote the fractions of the process yields n_k corresponding to bin i. They can be calculated via

$$\nu_{ik}(n_k) = n_k f_{ik}(h_k) \tag{6.18}$$

Here f_{ik} describes the fraction of a normalized histogram h_k that is present in the i-th bin for the template k. This fraction is given by the simple relation

$$f_{ik}(h_k) = \frac{h_{ik}}{\sum h_{ik}} \tag{6.19}$$

where h_{ik} is the number of entries in bin i of histogram h_k. To summarize this, all the equations shown above are used to calculate the negative log-likelihood in Eq. (6.3) for different process yields n_k. So the process yields are varied, such that the negative log-likelihood is minimized.

The described situation only shows the procedure without taking systematic uncertainties, which occur due to imperfections in the event generation, simulation, the experimental setup, or the MC statistics, into account. We have to differentiate between systematic uncertainties, which affect the rate of a process and the ones which allow for changes in the template shapes. Both are incorporated into the template fit via so-called nuisance parameters $\vec{\theta}$. If there are a number of R systematic uncertainties affecting the rates of a given process, they are incorporated by the substitution of

$$n_k \rightarrow n_k \prod_{l=1}^{R} (1 + \epsilon_{kl}\theta_{kl}) \tag{6.20}$$

with the relative uncertainty ϵ_{kl} and the nuisance parameter θ_{kl}, which is constrained by a Gaussian distribution with mean $\mu = 0$ and variance $\sigma = 1$. Shape affecting uncertainties have an influence on the fractions of a process, described in Eq. (6.19), and are incorporated via

$$f_{ik} \rightarrow f'_{ik} = \frac{h_{ik}(1 + \epsilon_{ik}\theta_{ik})}{\sum h_{ik}(1 + \epsilon_{ik}\theta_{ik})} \tag{6.21}$$

where the relative uncertainty ϵ_{ik} is calculated via the total covariance matrix of the template k.

In this analysis several reconstruction channels are fitted simultaneously, which can be included in the fit by adjusting Eq. (6.16). For this case the likelihood ratio is calculated as

$$L(\vec{n}) = \prod_{j=1}^{N_c} \prod_{i=1}^{N_{bins}} \frac{\nu_{ij}(\vec{n})^{m_{ij}}}{m_{ij}!} e^{-\nu_{ij}(\vec{n})} \tag{6.22}$$

$$\nu_{ij} = \sum_{k=1}^{M} n_k f_{ijk} \epsilon_{jk} \tag{6.23}$$

so that the expected number of events in each bin and channel ν_{ij} now also depends on the fraction of each process k in the j-th channel ϵ_{jk}. Beside the process yields n_k and the nuisance parameters $\vec{\theta}$, this fraction ϵ_{jk} is another parameter, that can be varied while minimizing the negative log-likelihood function. Adding this variation highly increases the complexity of the fit.

7. Form Factor Study

In previous $\mathcal{R}(\mathrm{D}^{(*)})$ analyses, like in [3–5], one of the main systematic uncertainties arose due to the limited knowledge about the shapes of the D^{**} states. For this thesis, as well as for the underlying $\mathcal{R}(\mathrm{D}^{(*)})$ measurement, the form factors of these states were corrected, as described in Section 4.6. This chapter deals with the evaluation of how well the reweighted form factor modeling describes the measured data. The form factors are studied in dependency of the momentum transfer q^2 and the momentum of the lepton in the rest frame of the $\mathrm{B}_{\mathrm{sig}}$ meson p_ℓ^*, by extracting the signal yields from the squared missing mass $\mathrm{M}^2_{\mathrm{miss}}$.
First, the general setup is described, which is then validated with Asimov fits.

7.1. General Concept

For studying how well the measured data is described by the form factor model, a comparison of the distributions in data and Monte Carlo is performed. To do so, the samples are divided into M bins of the investigated variable and the fraction of D^{**} events in each bin with respect to the total number of D^{**} events is measured. Thus, this ratio is calculated as

$$\mathcal{R}_i = \frac{N_i^{\mathrm{D}^{**}}}{\sum_{i=1}^{M} N_i^{\mathrm{D}^{**}}} \tag{7.1}$$

where $N_i^{\mathrm{D}^{**}}$ denotes the number of expected or measured D^{**} events in the i-th bin. These measured numbers are extracted by performing a template likelihood fit (Section 6.2) on the squared missing mass. In addition to the fitted number, an uncertainty on the result σ_i is also provided for each bin by the fit. The handling of these uncertainties is described below.

Covariance

With the knowledge about the fit uncertainties one can construct a $M \times M$ covariance matrix C defined as

$$C_{ij} = \sigma_i \sigma_j \delta_{ij} \tag{7.2}$$

with δ_{ij} being the Kronecker-Delta, so the covariance matrix only has diagonal elements. Since only the yields are measured, to get the uncertainty on the ratio $\mathcal{R}$ (Eq. (7.1)) this covariance has to be transformed. This is done, as described in [48], via

$$U = ACA^T \tag{7.3}$$

with $A_{ij} = \frac{\partial \mathcal{R}_i}{\partial N_j}$, the Jacobian of the transformation. U is the resulting covariance matrix for the calculated ratio. The Jacobian for the transformation is given as

$$A_{ij} = \begin{cases} \frac{1}{N_{tot}} - \frac{N_i}{N_{tot}^2} & i = j \\ -\frac{N_i}{N_{tot}^2} & i \neq j \end{cases} \tag{7.4}$$

where $N_{tot} = \sum_{i=1}^{M} N_i$. Using this together with Eq. (7.3), the errors on the ratio can be estimated with the formula

$$\sigma_i = \sqrt{U_{ii}} \tag{7.5}$$

Form Factor Variations

As already stated in Section 4.6, several variations are provided for the form factor corrections of the D^{**} states. When calculating the ratios $\mathcal{R}$ for the Monte Carlo expectation, these variations are handled as an additional uncertainty on the ratio. In the following the procedure to calculate these uncertainties is described in detail.

For every D^{**} form factor correction weight several up and down variations are defined. If these variations are applied instead of the nominal correction, one can calculate the difference in the number of events to the nominal value in the i-th bin for variation k with the formula

$$\Delta N_i^{up,k} = N_i^{up,k} - N_i^{nom} \tag{7.6}$$

$$\Delta N_i^{down,k} = N_i^{down,k} - N_i^{nom} \tag{7.7}$$

Afterwards for every variation a covariance matrix can be defined as

$$C_{ij}^{up,k} = \Delta N_i^{up,k} \Delta N_j^{up,k} \tag{7.8}$$

and equivalently for the down variations. These covariance matrices are then used to calculate a summed covariance with

$$C^{up} = \sum_{k=1}^{K} C^{up,k} \tag{7.9}$$

$$C^{down} = \sum_{k=1}^{K} C^{down,k} \tag{7.10}$$

assuming there are a total of K up (down) variations. The total covariance for a D^{**} type can be calculated by simply averaging over the up and down variations with

$$C^{\mathrm{D}^{**}} = \frac{1}{2}(C^{up} + C^{down}) \tag{7.11}$$

In this study four D^{**} states are investigated, but since D_1 and D_2^*, just as D_0^* and D_1', share the same parameters, the corresponding variations are applied simultaneously, resulting in two covariance matrices $C^{\mathrm{D}_1,\mathrm{D}_2^*}$ and $C^{\mathrm{D}_0^*,\mathrm{D}_1'}$. The total covariance matrix for the form factor variations is then calculated as

$$C^{tot} = C^{\mathrm{D}_1,\mathrm{D}_2^*} + C^{\mathrm{D}_0^*,\mathrm{D}_1'} \tag{7.12}$$

To get the uncertainty on the ratios, this total covariance has to be transformed with Eq. (7.3) and the uncertainties can be calculated as described above.

χ^2-Test

In Section 6.1 the χ^2-test was already introduced as a test statistic for how well a given model describes the data. For this study the test is used to evaluate the agreement of the MC expectation with the measured ratio. Having theses two ratios one can calculate the difference in each bin via

$$\Delta\mathcal{R}_i = \mathcal{R}_i^{MC} - \mathcal{R}_i^{Data} \tag{7.13}$$

which results in an M-dimensional vector $\Delta\vec{\mathcal{R}}$. For the calculation of the χ^2-value this vector has to be reduced by one dimension, since if $M-1$ ratios are known one automatically has knowledge about the last ratio, so the number of degrees of freedom is given by $M-1$. The χ^2-value is then determined using the formula

$$\chi^2 = (\Delta\vec{\mathcal{R}})U^{-1}(\Delta\vec{\mathcal{R}})^T \tag{7.14}$$

where U is given by

$$U = U^{Data} + U^{MC} + U^{FF} \tag{7.15}$$

with the individual transformed covariance matrices according to Eq. (7.3), which were also reduced to $(M-1)\times(M-1)$ dimensions.

7.2. Fitting Setup

To determine the distribution of the events in dependency of the momentum transfer q^2 and the lepton momentum p_ℓ^*, the sample is divided into bins of those variables. For q^2 six non-equidistant bins in the range of $0\,\mathrm{GeV}^2 \leq \mathrm{q}^2 < 10\,\mathrm{GeV}^2$ are chosen. In the momentum variable p_ℓ^* the sample is divided into eight bins, which are likewise not equidistant, in the range from $0.1\,\mathrm{GeV} \leq \mathrm{p}_\ell^* < 2.2\,\mathrm{GeV}$. After the samples are split, a template likelihood fit, as described in Section 6.2, is performed in the squared missing mass $\mathrm{M}_{\mathrm{miss}}^2$ (Section 5.1.1) separately in every bin of the two variables to extract the signal yields. For the minimization of the negative log-likelihood the MINUIT algorithm [49] together with the corresponding python interface IMINUIT [50] is used.

With the aim to achieve independent measurements for the different decays $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}\pi^{\pm})\ell\nu$, $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^*\pi^{\pm})\ell\nu$, $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}\pi^0)\ell\nu$, and $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^*\pi^0)\ell\nu$, a total of four fits are performed in each bin of the investigated variable. The reconstruction channels

Table 7.1.: Used reconstruction channels for the extraction of the signal yields. The used number of $\mathrm{M}^2_{\mathrm{miss}}$ bins in dependency of the channel is given as well.

Signal Yield	Reconstruction Channels	$\mathrm{M}^2_{\mathrm{miss}}$ Bins
$\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}\pi^{\pm})\ell\nu$	$\mathrm{B}^0 \to \overline{\mathrm{D}}^0\pi^-\ell^+$	30
	$\mathrm{B}^- \to \mathrm{D}^+\pi^-\ell^-$	30
$\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^*\pi^{\pm})\ell\nu$	$\mathrm{B}^0 \to \overline{\mathrm{D}}^{*0}\pi^-\ell^+$	25
	$\mathrm{B}^- \to \mathrm{D}^{*+}\pi^-\ell^-$	20
$\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}\pi^0)\ell\nu$	$\mathrm{B}^- \to \mathrm{D}^0\pi^0\ell^-$	25
	$\mathrm{B}^0 \to \mathrm{D}^-\pi^0\ell^+$	20
$\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^*\pi^0)\ell\nu$	$\mathrm{B}^- \to \mathrm{D}^{*0}\pi^0\ell^-$	20
	$\mathrm{B}^0 \to \mathrm{D}^{*-}\pi^0\ell^+$	20

used for these fits depend on the signal yield which is extracted. A summary is shown in Table 7.1. Each signal yield is fitted simultaneously in the described reconstruction channels. For each reconstruction channel the number of bins in the squared missing mass is given in Table 7.1. The range of the squared missing mass is constrained to $-2\,\mathrm{GeV}^2 < \mathrm{M}^2_{\mathrm{miss}} < 4\,\mathrm{GeV}^2$. As the momentum transfer q^2 is related to the squared missing mass, the range is chosen as $-2\,\mathrm{GeV}^2 < \mathrm{M}^2_{\mathrm{miss}} < 2\,\mathrm{GeV}^2$ for the first three q^2 bins, such that there are no empty bins in the fitting histograms. For the lepton momentum a similar scenario occurs. Higher values of p^*_ℓ limit the squared missing mass and so the squared missing mass range is constrained in the last bins (for the exact values see the fit results in Appendix B). The nuisance parameters, as well as the efficiencies are kept fixed for now.

Fit Templates

The histograms for the template likelihood fit are generated with the usage of the MC expectation. Part of their definition is based on the squared missing mass components, which were described earlier in Section 5.1.1. Additionally, the template definitions depend on the extracted signal yield parameter and are given by:

- The signal component ($\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}\pi^{\pm})\ell\nu$, $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^*\pi^{\pm})\ell\nu$, $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}\pi^0)\ell\nu$, or $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^*\pi^0)\ell\nu$).
- The corresponding $\mathrm{B} \to \mathrm{D}^{(*)}\pi\ell\nu$ down- or up-feed component.
- A $\mathrm{B} \to \mathrm{D}^{(*)}\pi^{\pm}\ell\nu$ or $\mathrm{B} \to \mathrm{D}^{(*)}\pi^0\ell\nu$ depending on whether the involved pion of the D^{**} decay of the signal component is charged or not. This background is always the counterpart to the signal component with respect to this property of the pion.
- $\mathrm{D}^{(*)}\ell\nu$ background.
- Remaining background, which includes hadronic, $\mathrm{B}\overline{\mathrm{B}}$, and continuum background, as well as the $\mathrm{B} \to \mathrm{D}^{(*,**)}\tau\nu$ components.

So for the extraction of a signal yield there are a total of five templates defined. Every template is generated with all MC corrections, as described in Chapter 4, applied.

q^2 Migration

The q^2 binning, which was described above, is limited by the q^2 resolution. When dividing the events into bins of the reconstructed momentum transfer $\mathrm{q}^2_{\mathrm{Reco}}$ it is desired that a large fraction of these would be found in the same bin of the truth momentum transfer $\mathrm{q}^2_{\mathrm{Truth}}$, which is calculated via MC information. To test this, a migration matrix M can be defined as

$$M_{ij} = \frac{N_{ij}^{\mathrm{Truth}}}{N_i^{\mathrm{Reco}}} \tag{7.16}$$

where N_{ij}^{Truth} describes the number of signal events, which are reconstructed in the i-th bin, but originated from the j-th truth bin. N_i^{Reco} gives the total number of signal events reconstructed in bin i.

Since four different signal decays are measured, one can construct four independent migration matrices. The matrices are calculated using signal MC samples, in which one of the B mesons is required to decay into $\mathrm{D}^{**}\ell\nu$. For the calculation, only the reconstruction channels which correspond to the investigated decay are considered. The results are shown in Fig. 7.1. In general the matrices show the desired behavior with large values on the diagonal, but the q^2 resolution is not as perfect as one would wish. Especially for the last bin ($7\,\mathrm{GeV}^2 \leq \mathrm{q}^2 < 10\,\mathrm{GeV}^2$) only about half of the events are reconstructed in the correct bin, so this bin should be excluded in further analyses.

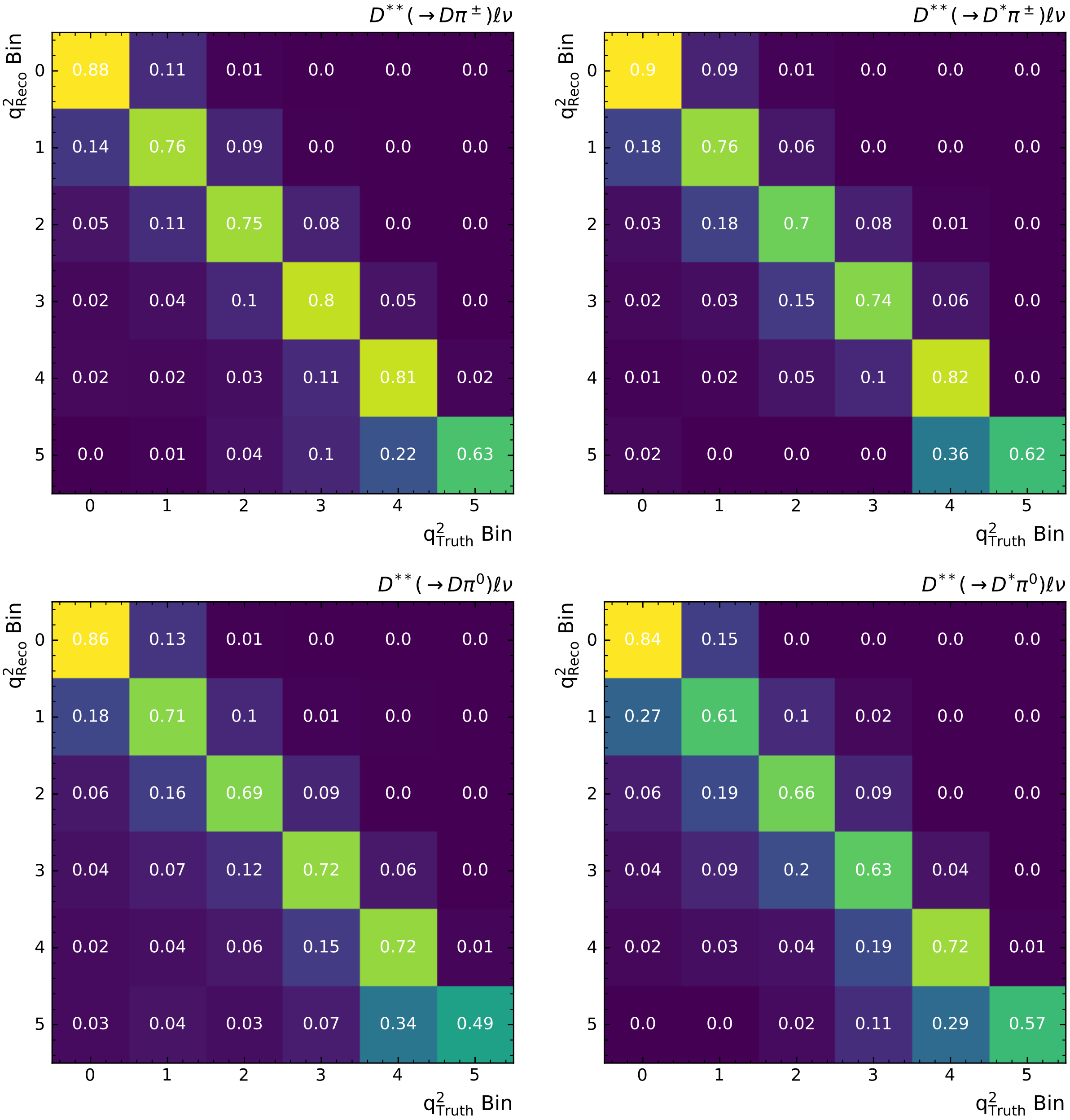
7.3. Asimov Fits

The fitting procedure is evaluated using a so-called Asimov data set, which is an artificial data set originating from the MC expectation. In this section, the properties of the fitting setup are discussed based on the results gathered with the Asimov data set, divided into results for q^2 and p^*_ℓ. Parts of the fitting setup (Section 7.2) are optimized based on these results.

7.3.1. Results for q^2

The post-fit template distributions for the Asimov fits in the third bin of q^2 ($2\,\mathrm{GeV}^2 \leq \mathrm{q}^2 < 3\,\mathrm{GeV}^2$) are shown in Fig. 7.2 and Fig. 7.3 as an example. All remaining plots can be found in Appendix A. Since the Asimov data set represents the templates per definition the fit results should exactly match the Asimov data. Since this is the case, the Asimov fits show no hints for an error in the fitting procedure, from this point of view.

For every signal yield in the q^2 bins a significance is calculated, as described in Section 6.1. Fig. 7.4 shows these significances. Large discrepancies between the components, in which the D^{**} decays into a charged pion and the neutral cases are observed. While for the charged components significances in the range from $7\,\sigma$ to $18\,\sigma$ are observed, with only the last bin being an exception, the significances for the neutral pion components are much lower ($2.4\,\sigma < Z < 5\,\sigma$). The observed significances in the last q^2 bin are much lower for all



$D^{**}(\to D\pi^{\pm})\ell\nu$

q^2_{Reco} Bin \ q^2_{Truth} Bin	0	1	2	3	4	5
0	0.88	0.11	0.01	0.0	0.0	0.0
1	0.14	0.76	0.09	0.0	0.0	0.0
2	0.05	0.11	0.75	0.08	0.0	0.0
3	0.02	0.04	0.1	0.8	0.05	0.0
4	0.02	0.02	0.03	0.11	0.81	0.02
5	0.0	0.01	0.04	0.1	0.22	0.63

$D^{**}(\to D^{*}\pi^{\pm})\ell\nu$

q^2_{Reco} Bin \ q^2_{Truth} Bin	0	1	2	3	4	5
0	0.9	0.09	0.01	0.0	0.0	0.0
1	0.18	0.76	0.06	0.0	0.0	0.0
2	0.03	0.18	0.7	0.08	0.01	0.0
3	0.02	0.03	0.15	0.74	0.06	0.0
4	0.01	0.02	0.05	0.1	0.82	0.0
5	0.02	0.0	0.0	0.0	0.36	0.62

$D^{**}(\to D\pi^{0})\ell\nu$

q^2_{Reco} Bin \ q^2_{Truth} Bin	0	1	2	3	4	5
0	0.86	0.13	0.01	0.0	0.0	0.0
1	0.18	0.71	0.1	0.01	0.0	0.0
2	0.06	0.16	0.69	0.09	0.0	0.0
3	0.04	0.07	0.12	0.72	0.06	0.0
4	0.02	0.04	0.06	0.15	0.72	0.01
5	0.03	0.04	0.03	0.07	0.34	0.49

$D^{**}(\to D^{*}\pi^{0})\ell\nu$

q^2_{Reco} Bin \ q^2_{Truth} Bin	0	1	2	3	4	5
0	0.84	0.15	0.0	0.0	0.0	0.0
1	0.27	0.61	0.1	0.02	0.0	0.0
2	0.06	0.19	0.66	0.09	0.0	0.0
3	0.04	0.09	0.2	0.63	0.04	0.0
4	0.02	0.03	0.04	0.19	0.72	0.01
5	0.0	0.0	0.02	0.11	0.29	0.57

Figure 7.1.: Migration matrices for the different $B \to D^{**}\ell\nu$ signal decays. The matrices are calculated on dedicated signal MC samples in the corresponding reconstruction channels.

measured decay types. When taking also the bad migration values described above into account, no reliable statements can be made for this q^2 bin.
The calculated ratios (Eq. (7.1)), for each of the observed signal components, are shown in Fig. 7.5. For the $B \to D^{**}(\to D^{(*)}\pi^{\pm})\ell\nu$ components, the observed statistical uncertainties are much lower than for the $B \to D^{**}(\to D^{(*)}\pi^{0})\ell\nu$ decays. This is expected due to the limited statistics in the latter reconstruction channels. For the Asimov data, the measured ratios match exactly the MC expectations ($\chi^2 = 0.0$), which is expected per definition of the data set.

7.3.2. Results for p^*_ℓ

As for the q^2 distributions, the ratios for the form factor study are also calculated in bins of the lepton momentum in the B_{sig} rest frame p^*_ℓ. In Fig. 7.6 and Fig. 7.7 the post-fit distribution of the Asimov fit in the fifth p^*_ℓ bin ($1.2\,\text{GeV} \leq p^*_\ell < 1.4\,\text{Gev}$) can be found. The plots only cover a small part of the fit results, while the remaining plots are given in Appendix B. As expected, the sum of the fitted templates matches exactly the MC expectation.
The significances of the fits are given in Fig. 7.8. Again, large discrepancies between the charged pion components and the neutral ones are observed. Double-digit significances for the $D^{(*)}\pi^{\pm}\ell\nu$ components in the central region of the p^*_ℓ distribution are measured. In the outer regions, as well as for the $D^{(*)}\pi^{0}\ell\nu$ components much lower significances are achieved. The ratios, which are shown in Fig. 7.9, show no differences between the MC expectation and the ones calculated via the measured yields on Asimov data. In comparison to the q^2 study, the form factor variations are much smaller for each bin, since the form factor weights are less dependent on the lepton momentum than on the momentum transfer q^2. Additionally, the plots show smaller statistical uncertainties, due to a smaller correlation between the bins.

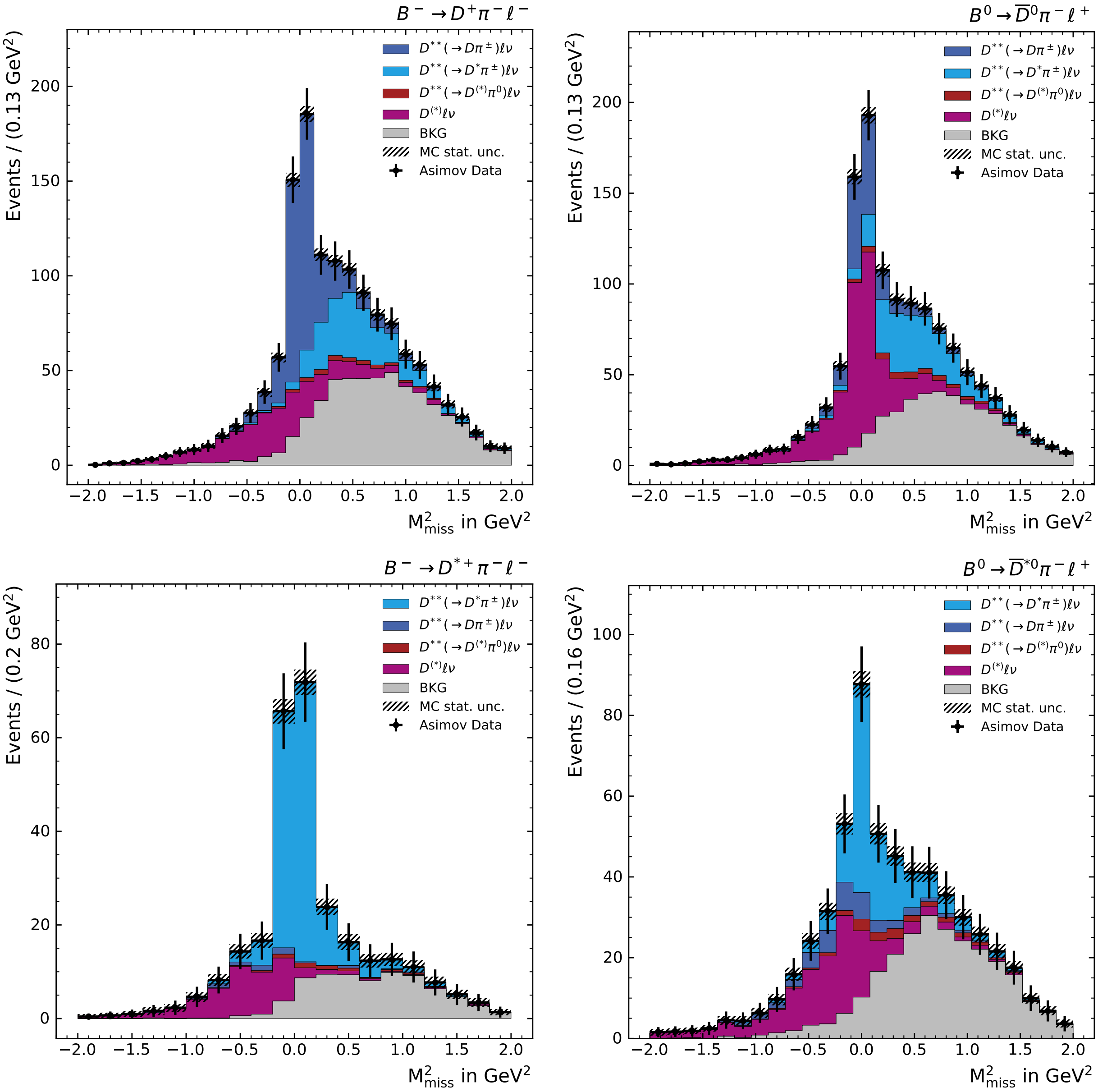


Figure 7.2.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $2\,\mathrm{GeV}^2 \leq q^2 < 3\,\mathrm{GeV}^2$.

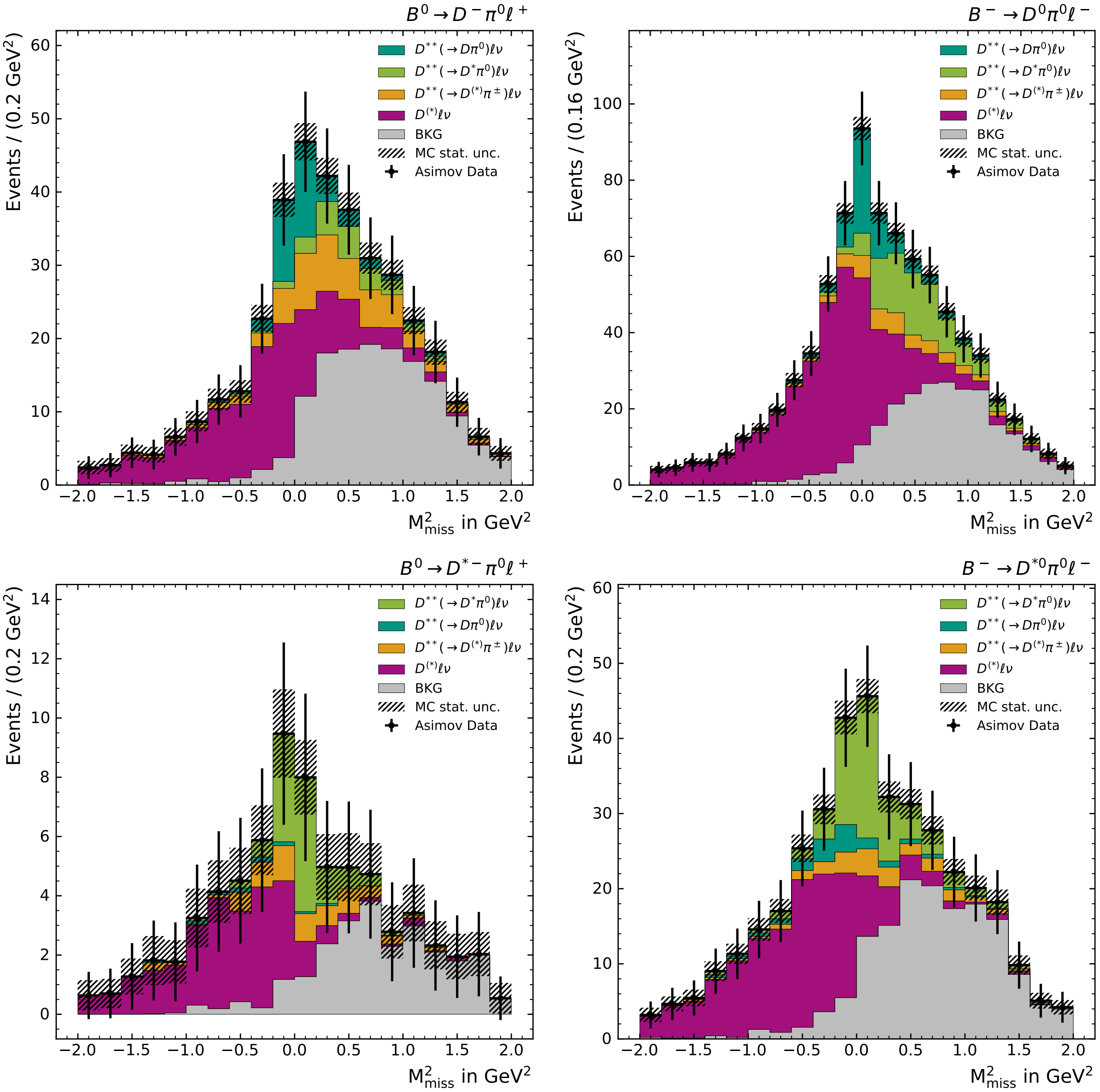


Figure 7.3.: Post-fit distributions of the template likelihood fit for the extraction of the $\mathrm{B} \to \mathrm{D}\pi^0\ell\nu$ and $\mathrm{B} \to \mathrm{D}^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $2\,\mathrm{GeV}^2 \leq \mathrm{q}^2 < 3\,\mathrm{GeV}^2$.

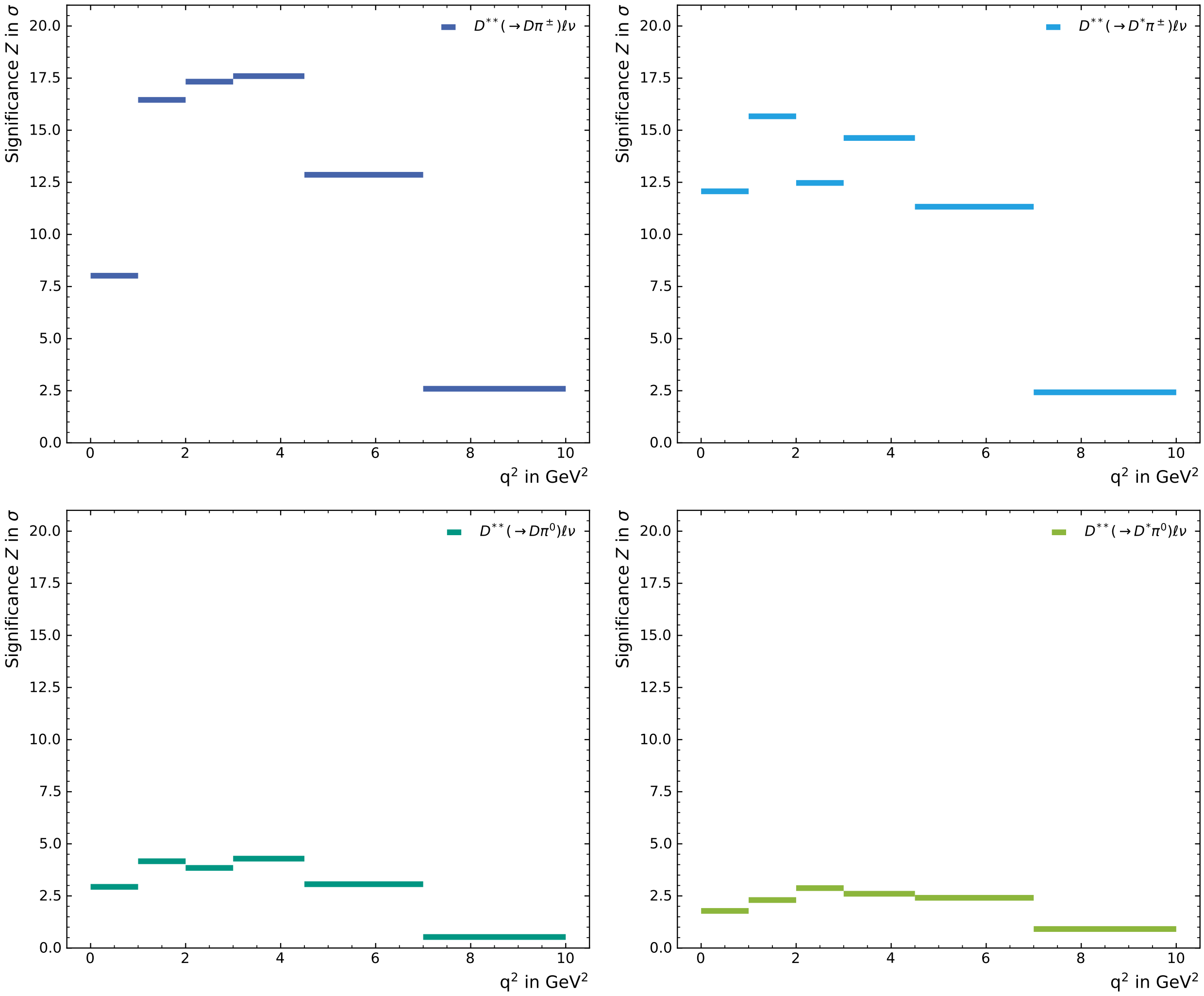


Figure 7.4.: Significances of the fit calculated on Asimov data for the different D^{**} signal components. The significance is determined in every q^2 bin independently.

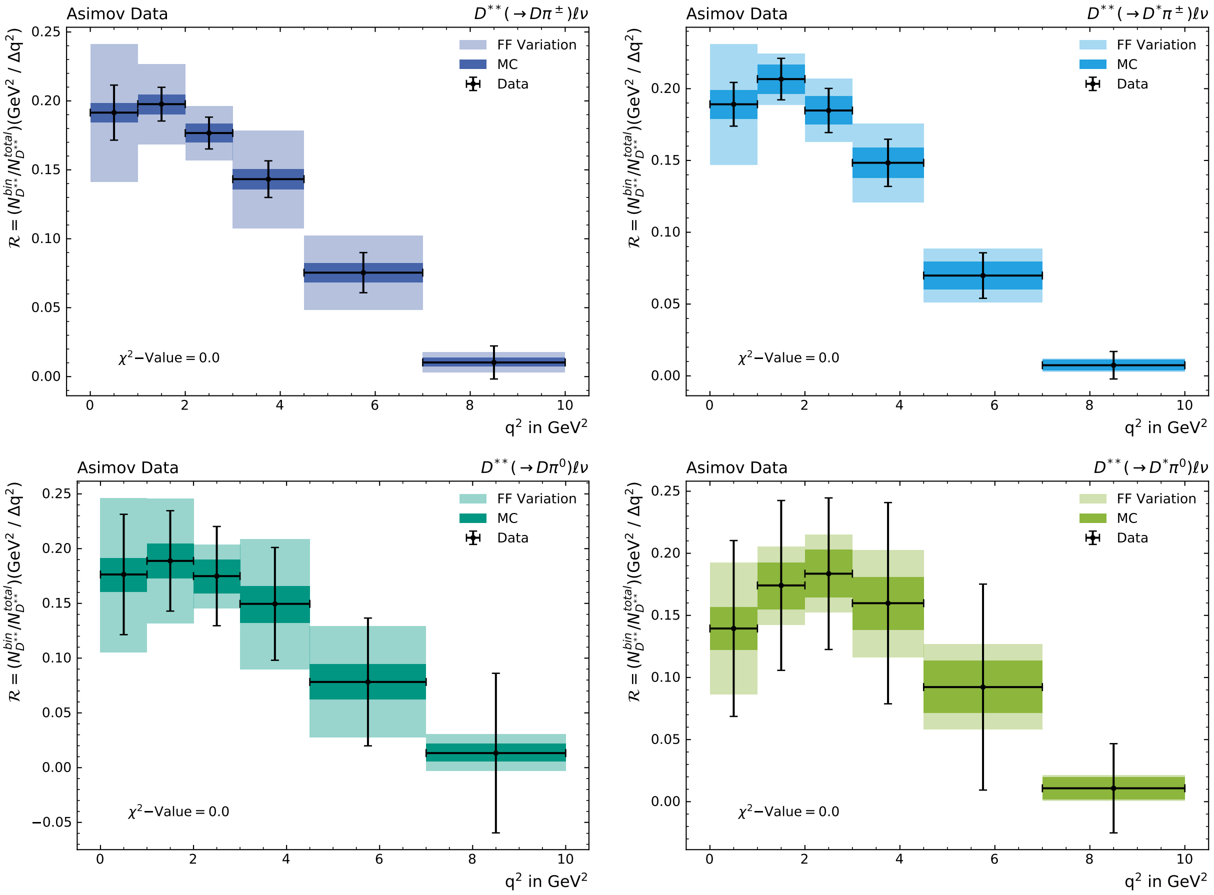


Figure 7.5.: Ratios $\mathcal{R}$ calculated on Asimov data for the different D^{**} signal components. The ratios are normalized to the width of the respective q^2 bin. The MC expectation is shown colored, with the statistical uncertainty as bands around the central values. Additional form factor variations are shown as paler bands. For the errors on the Asimov data the uncertainties are calculated from the covariance matrix of the fit results. The χ^2-values are calculated according to Eq. (7.14).

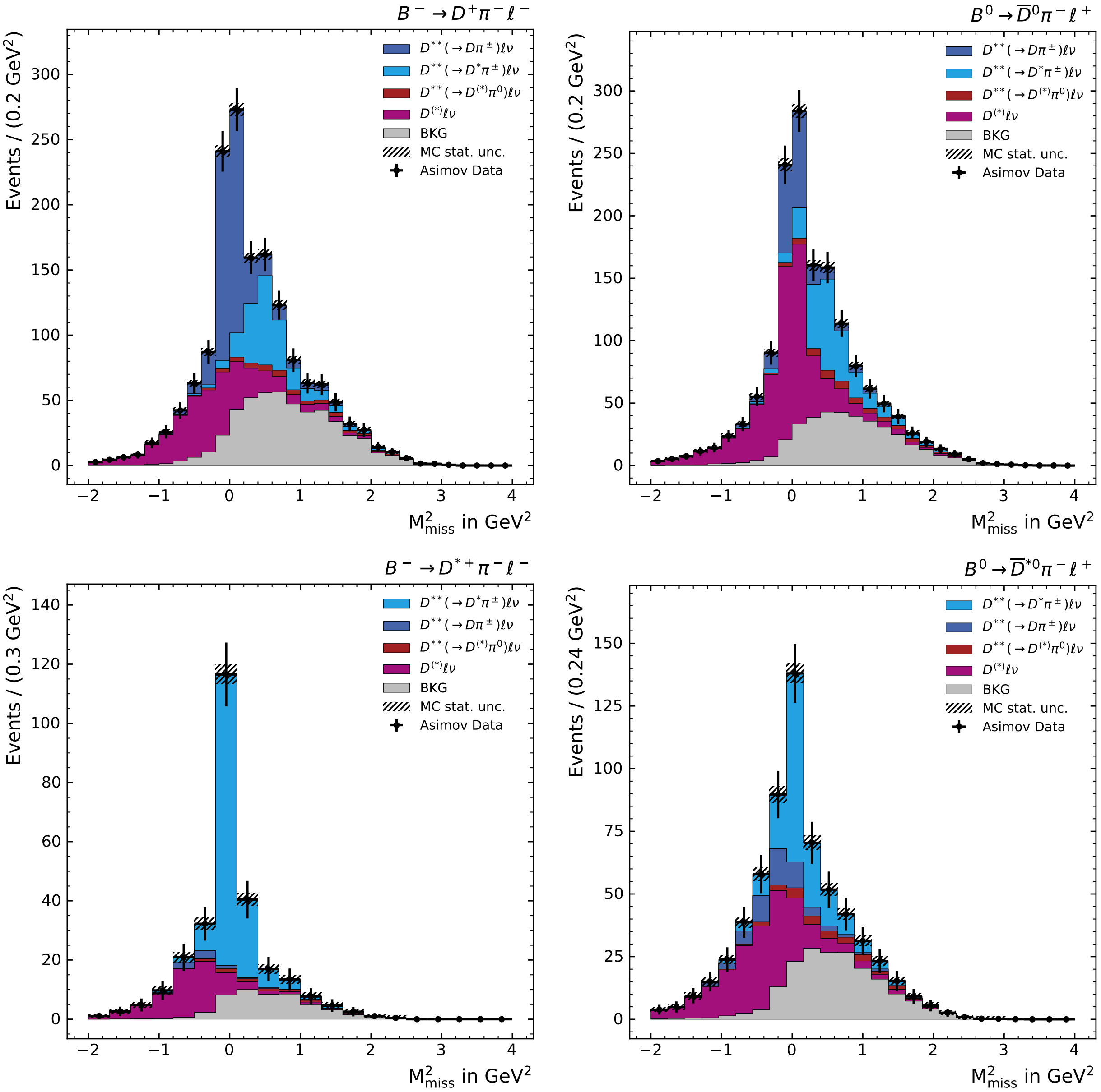


Figure 7.6.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1.2\,\text{GeV} \leq p^*_\ell < 1.4\,\text{GeV}$.

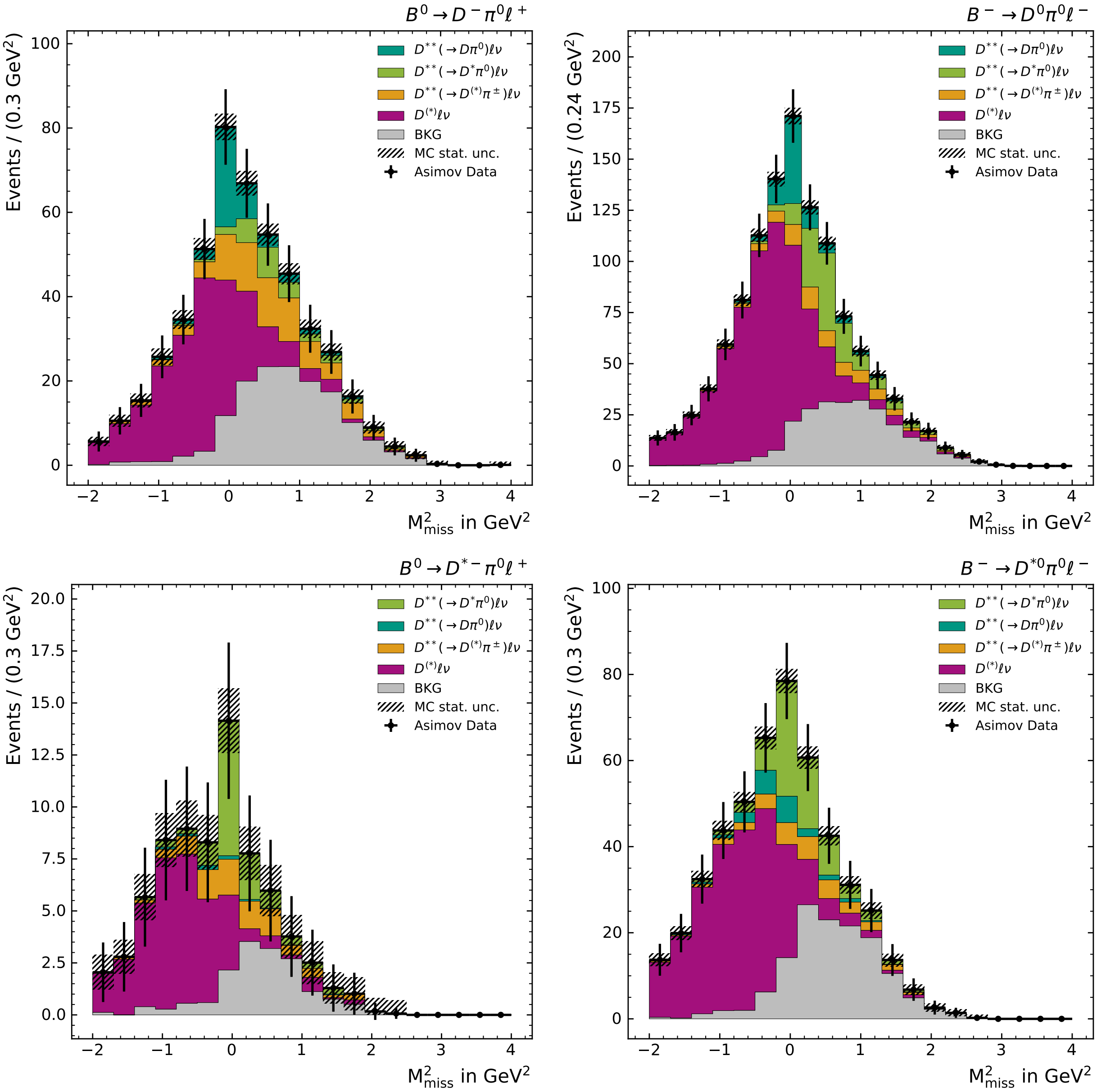


Figure 7.7.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^0\ell\nu$ and $B \to D^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1.2\,\mathrm{GeV} \leq p_\ell^* < 1.4\,\mathrm{GeV}$.

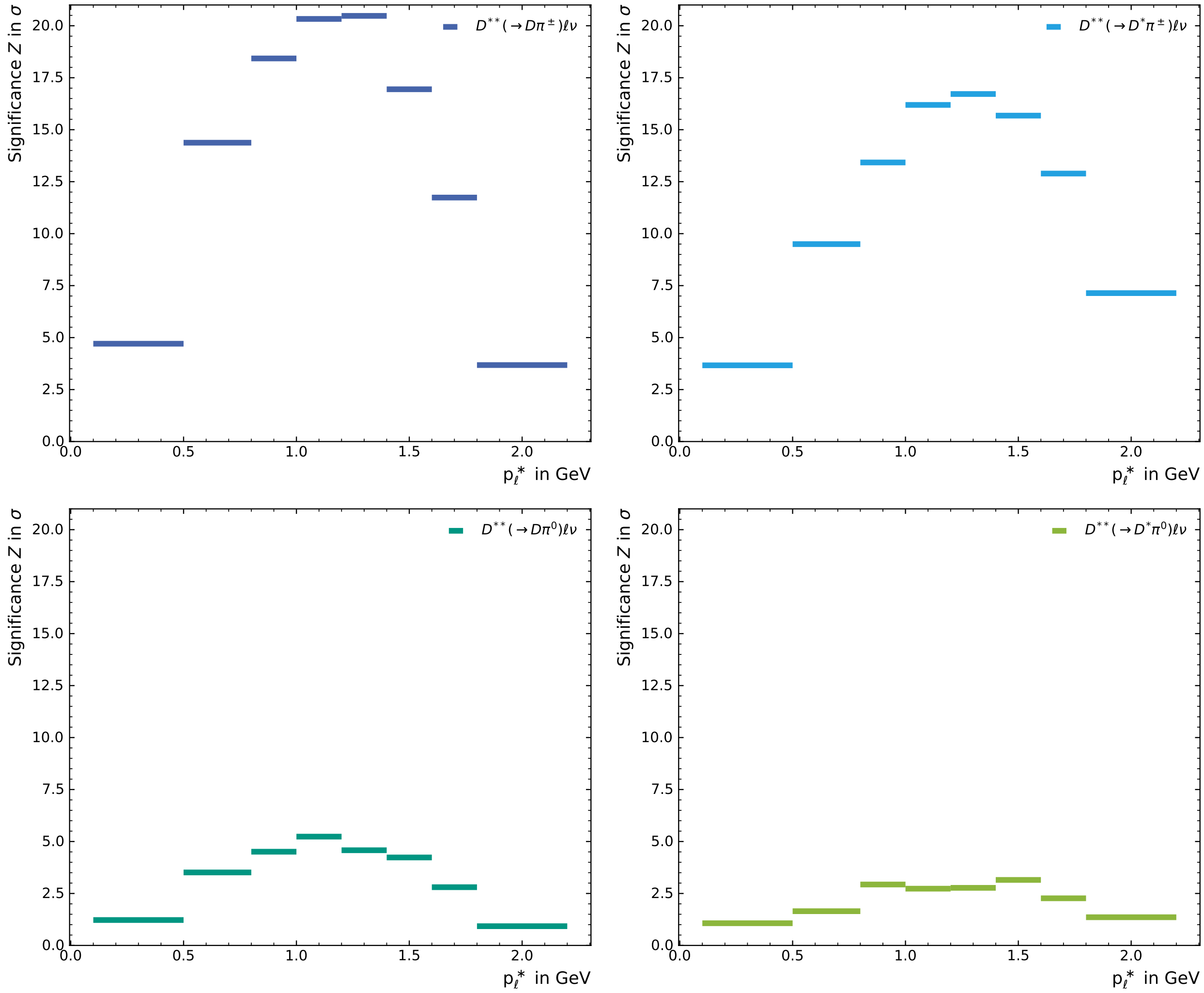


Figure 7.8.: Significancies of the fit calculated on Asimov data for the different D^{**} signal components. The significance is determined in every p_ℓ^* bin independently.

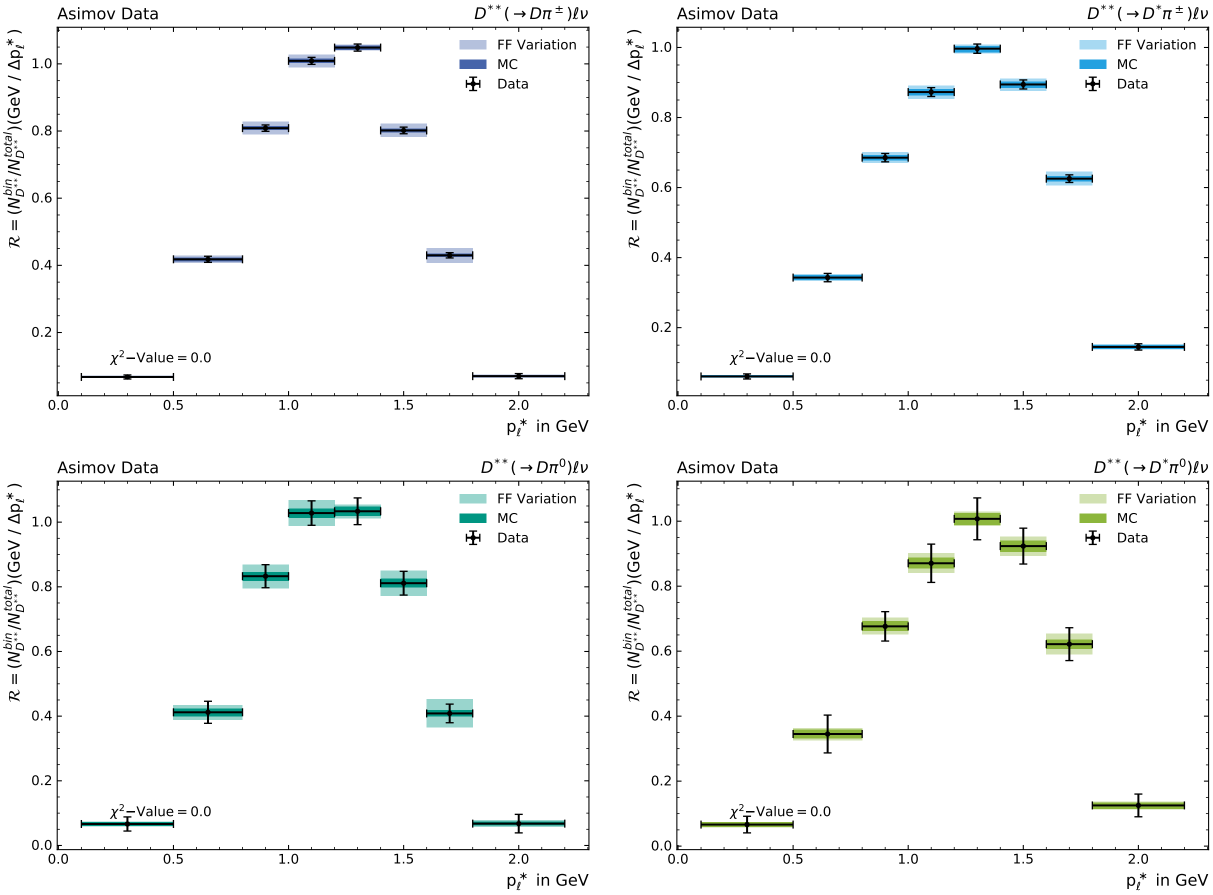


Figure 7.9.: Ratios $\mathcal{R}$ calculated on Asimov data for the different D^{**} signal components. The ratios are normalized to the width of the respective p^*_ℓ bin. The MC expectation is shown colored, with the statistical uncertainty as bands around the central values. Additional form factor variations are shown as paler bands. For the errors on the Asimov data the uncertainties are calculated from the covariance matrix of the fit results. The χ^2-values are calculated according to Eq. (7.14).

8. D^{**} Branching Fractions

In addition to the limited knowledge about the form factors of the D^{**} states, which is described in Chapter 7, another major systematic uncertainty reported for previous $\mathcal{R}(\mathrm{D}^{(*)})$ analyses [3–5] is given by the poorly measured branching fractions of the $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ decays. So, this chapter describes the fitting setup for the determination of these branching fractions for each of the four D^{**} states. As already stated in Section 5.1.2, the D^{**} states can be subdivided into narrow and broad states, based on their mass width. The narrow states, D_1 and D_2^*, show distinct peaks in the $\Delta\mathrm{M}$ distributions, while the broad states, D_0^* and D_1', show rather broad structures. This leads to the expectation that the narrow states can be handled better when trying to measure the branching fractions. The branching fractions are calculated for the cases where the D^{**} decays into a charged or neutral pion independently. The fitting procedure and the branching fraction calculation are validated on the Asimov data set since the signal region is still blinded.

8.1. Branching Fraction Calculation

In this thesis, the branching fractions of the $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi)\ell\nu$ decays are calculated via the ratio of the signal process with respect to the on of the better understood $\mathrm{B} \to \mathrm{D}^*\ell\nu$ decay. Here ℓ denotes one of the light leptons e or μ, while the τ leptons are not considered for this thesis. One could also use the $\mathrm{B} \to \mathrm{D}\ell\nu$ decay as a normalization, but this decay is statistically more limited in comparison to the D^* decay. The ratio $\mathcal{R}(\mathrm{D}^{**}, \mathrm{D}^*)$ is defined as

$$\begin{aligned}
\mathcal{R}(\mathrm{D}^{**}, \mathrm{D}^*) &= \frac{N(\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi)\ell\nu)}{N(\mathrm{B} \to \mathrm{D}^*\ell\nu)} && (8.1)\\
&= \frac{N_{\text{reco}}(\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi)\ell\nu)}{N_{\text{reco}}(\mathrm{B} \to \mathrm{D}^*\ell\nu)} \cdot \frac{\epsilon_{\text{reco}}(\mathrm{B} \to \mathrm{D}^*\ell\nu)}{\epsilon_{\text{reco}}(\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi)\ell\nu)} && (8.2)\\
&= \mathcal{R}_N(\mathrm{D}^{**}, \mathrm{D}^*) \cdot \frac{\epsilon_{\text{reco}}(\mathrm{B} \to \mathrm{D}^*\ell\nu)}{\epsilon_{\text{reco}}(\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi)\ell\nu)} && (8.3)
\end{aligned}$$

with N and N_{reco} being the number of events produced in the B decay and the measured ones, respectively. The reconstruction efficiencies ϵ_{reco}, which denote the fraction of reconstructed and selected to the number of produced events, are calculated on the basis of signal MC samples, which are independent from the samples used for the fit.

For the determination of the uncertainty on $\mathcal{R}_N(\mathrm{D}^{**}, \mathrm{D}^*)$, it is important to consider the correlation between the measured yields N since they are all fitted simultaneously. The uncertainty is then calculated analog to Eq. (7.3) and Eq. (7.5). For this transformation the Jacobian is given by

$$A_{ij} = \begin{cases} \frac{1}{N_n} & i = j \neq n \\ -\frac{N_i}{N_n^2} & i \neq j = n \\ 0 & \text{else} \end{cases} \tag{8.4}$$

where N_n denotes the yield of the normalization process $\mathrm{B} \to \mathrm{D}^* \ell\nu$. The covariance matrix C is directly taken from the fit result.
Afterwards, the branching fraction of the signal D^{**} decay is calculated via

$$\mathcal{B}(\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi)\ell\nu) = \mathcal{R}(\mathrm{D}^{**}, \mathrm{D}^*) \cdot \mathcal{B}(\mathrm{B} \to \mathrm{D}^* \ell\nu) \tag{8.5}$$

where $\mathcal{B}(\mathrm{B} \to \mathrm{D}^* \ell\nu)$ is the latest branching fraction value taken from the *pdg* [24].

8.2. Fitting Setup

To extract the signal and normalization yields for the branching fraction calculation a template likelihood fit (Section 6.2) is performed. As the fitting variable the reconstructed mass difference $\Delta\mathrm{M}$ between the D^{**} and the $\mathrm{D}^{(*)}$ (see Section 5.1.2) is chosen, since it allows for the best differentiation between the different D^{**} states D_0^*, D_1', D_1, and D_2^*.
For this thesis, a total of 16 branching fractions of the form $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi)\ell\nu$ are measured. The number 16 results from the differentiation into the four D^{**} states and a differentiation between cases where the D^{**} decays into a charged or neutral pion as well as the additional separation into B^0 and $\mathrm{B}^\pm$ branching fractions. The eight branching fractions of process including a charged pion emission are determined with a simultaneous fit in the D^{**} reconstruction modes, in which the B meson was reconstructed via $\mathrm{B} \to \mathrm{D}^{(*)}\pi^\pm\ell$ ($\mathrm{D}^{(*)}\pi^\pm$-Channels in Table 3.5). For the extraction of the $\mathrm{B}^\pm \to \mathrm{D}^{**}\ell\nu$ signal yields the two charged B modes of these channels are used, while the $\mathrm{B}^0 \to \mathrm{D}^{**}\ell\nu$ components are extracted in the neutral modes. For the processes in which a neutral pion is involved in the D^{**} decay, the same procedure is applied, while exchanging the reconstruction modes with the respective ones, which involve a neutral pion ($\mathrm{D}^{(*)}\pi^0$-Channels in Table 3.5). So in the end four independent fits, with two reconstruction channels considered in each of them, are performed to extract the 16 signal yields and the four corresponding normalization yields.
The templates for the likelihood fit are generated on the basis of MC simulations and are divided into the following seven components for the $\mathcal{B}(\mathrm{B}^\pm \to \mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi^\pm)\ell\nu)$ measurement:

$\mathrm{B}^\pm \to \mathrm{D}_0^*(\to \mathrm{D}\pi^\pm)\ell\nu$

This component includes all decays of a B meson into a D_0^*, which then decays further into a D meson and a charged pion. Due to the broad mass width of the D_0^* this component manifests itself as a broad band in the $\Delta\mathrm{M}$ distribution.

$\mathrm{B}^\pm \to \mathrm{D}_1'(\to \mathrm{D}^*\pi^\pm)\ell\nu$

Similar to the D_0^* component, the events where the B meson decays into the other broad D^{**} state D_1' show the same non-peaking ΔM distribution. In this component the D_1' decays into a excited D^* meson since this is the only allowed strong decay for this D^{**} state involving the emission of one pion.

$B^\pm \to D_1(\to D^*\pi^\pm)\ell\nu$

The B meson decays into a D_1, which then decays into a D^* and $\pi^\pm$, show a peak in the ΔM distribution at about $\Delta M \approx 0.4\,\text{GeV}$, which is the mass difference between the D_1 and the D^*.

$B^\pm \to D_2^*(\to D^{(*)}\pi^\pm)\ell\nu$

The last signal component includes all $B \to D_2^*\ell\nu$ decays, where the D_2^* decays under the emission of a charged pion into either a D meson or an excited D^*. Since both strong decays are allowed, two peaks can be observed in the ΔM distribution for this component. The peaks are at $\Delta M \approx 0.4\,\text{GeV}$ and $\Delta M \approx 0.6\,\text{GeV}$, which are the mass differences between the D_2^* and the D^* or D meson, respectively.

$B \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ Crossfeed

In this component all $B \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ decays, in which the D^{**} decayed not into a charged but into a neutral pion, are combined. The charge of the B meson is not considered in this component. Since in the charged pion reconstruction modes events of this type are wrongly reconstructed events, no structure in the ΔM variable can be seen.

$B^0 \to D^*\ell\nu$ Normalization

This component serves as the normalization component for the branching fraction calculation. It includes all events, in which the B meson decays semileptonically into an excited D^* meson. In the reconstruction modes containing a charged pion the charge of the B meson of the normalization component is the opposite of the charge of the signal components, since an additional charged pion is falsely added to the D^* meson.

Background

Everything that was not covered by one of the above components is summarized in the background component, which is distributed over the whole ΔM range. This also includes the D^{**} components, where the B meson has the wrong charge. The background component also contains the $B \to D^{**}\tau\nu$ decays, which present an additional type of semileptonic B to D^{**} decays but are not of interest for this thesis. Additional decays of the form $B \to D_1(\to D\pi\pi)\ell\nu$, which are included in the representation of the gap MC, are also covered by the background component, since a measurement of this decay is not done in the scope of this thesis.

The templates for the determination of the $B^0 \to D^{**}(\to D^{(*)}\pi^\pm)\ell\nu$ branching fractions are defined similarly and can be obtained by exchanging the charged B mesons with neutral ones and the other way around in the template definitions given above. To get the templates for the measurement of the $B \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ branching fractions, one has to exchange the charged pions in the signal components with neutral ones as well as the neutral pions in the crossfeed component with charged ones in the above definitions. Since the falsely added pions in the normalization modes are uncharged in these cases the charge of the B meson in this mode has to be similar as for the signal components. The charge of the $D^{(*)}$

and D^{**} mesons is in the signal and normalization components, where the charge of the B meson is fixed, implicitly given by the other involved particles.
The fit is performed in a ΔM range of $0.2\,\text{GeV} \leq \Delta M \leq 0.9\,\text{GeV}$, with 40 bins in the charged pion channels, and 20 bins in the neutral pion channels.

8.3. Fit on Asimov Data

The fitting setup is first tested on the Asimov data set, to validate the setup and determine the sensitivity of the fit.
The post-fit results are shown in Fig. 8.1 for the charged pion decays and in Fig. 8.2 for the neutral ones. In the $B \to D^{(*)}\pi^{\pm}\ell$ reconstruction mode the template distributions show the expected behavior and clear peaks for the D_1 and D_2^* components. On the other hand, in the $B \to D^{(*)}\pi^0\ell$ modes, which are used for the determination of the $B \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ branching fractions, the peaks can only be surmised. Additionally, the $B \to D^*\ell\nu$ normalization component dominates over the signal components in these modes. A reason for this big difference between the charged and neutral pion cases can be found in the significantly less efficient π^0 reconstruction in comparison to the reconstruction of the charged pions, which leads to lower statistics and a worse ΔM resolution in theses modes.
In Table 8.1 the extracted signal and normalization yields together with the statistical uncertainties are given. For every signal component a significance is calculated to determine the sensitivity of the fit. As expected, the narrow D_1 and D_2^* states can be measured much better than the broad D_0^* and D_1' states, since the former produce clear peaks in the ΔM spectrum. This results in lower statistical uncertainties on the yields of the narrow states and much higher significances compared to the broad D^{**} states.
To calculate the branching fractions with Eq. (8.5) one also needs the reconstruction efficiencies of the components in the considered reconstruction channels. These efficiencies are calculated on dedicated signal MC samples and are given in Table 8.2.
With all this information the measured D^{**} branching fractions on the Asimov data set can be calculated and are for the decay into broad D_0^* state given as

$$\mathcal{B}(B^{\pm} \to D_0^*(\to D\pi^{\pm})\ell\nu) = 0.52^{+0.16}_{-0.26}\,\% \tag{8.6}$$
$$\mathcal{B}(B^{0} \to D_0^*(\to D\pi^{\pm})\ell\nu) = 0.48^{+0.22}_{-0.29}\,\% \tag{8.7}$$
$$\mathcal{B}(B^{\pm} \to D_0^*(\to D\pi^{0})\ell\nu) = 0.28^{+0.41}_{-0.42}\,\% \tag{8.8}$$
$$\mathcal{B}(B^{0} \to D_0^*(\to D\pi^{0})\ell\nu) = 0.21^{+0.19}_{-0.22}\,\% \tag{8.9}$$

and for the other broad state D_1' as

$$\mathcal{B}(B^{\pm} \to D_1'(\to D^*\pi^{\pm})\ell\nu) = 0.55^{+0.29}_{-0.36}\,\% \tag{8.10}$$
$$\mathcal{B}(B^{0} \to D_1'(\to D^*\pi^{\pm})\ell\nu) = 0.57^{+0.38}_{-0.44}\,\% \tag{8.11}$$
$$\mathcal{B}(B^{\pm} \to D_1'(\to D^*\pi^{0})\ell\nu) = 0.30^{+0.48}_{-0.49}\,\% \tag{8.12}$$
$$\mathcal{B}(B^{0} \to D_1'(\to D^*\pi^{0})\ell\nu) = 0.22^{+0.48}_{-0.49}\,\% \tag{8.13}$$

The branching fractions of the B meson decays into the narrow D_1 state are calculated as

$$\mathcal{B}(\mathrm{B}^{\pm} \to \mathrm{D}_1(\to \mathrm{D}^{*}\pi^{\pm})\ell\nu) = 0.61^{+0.22}_{-0.32}\,\% \tag{8.14}$$

$$\mathcal{B}(\mathrm{B}^{0} \to \mathrm{D}_1(\to \mathrm{D}^{*}\pi^{\pm})\ell\nu) = 0.59^{+0.33}_{-0.39}\,\% \tag{8.15}$$

$$\mathcal{B}(\mathrm{B}^{\pm} \to \mathrm{D}_1(\to \mathrm{D}^{*}\pi^{0})\ell\nu) = 0.32^{+0.12}_{-0.13}\,\% \tag{8.16}$$

$$\mathcal{B}(\mathrm{B}^{0} \to \mathrm{D}_1(\to \mathrm{D}^{*}\pi^{0})\ell\nu) = 0.26^{+0.17}_{-0.20}\,\% \tag{8.17}$$

and for the decay into D_2^* mesons as

$$\mathcal{B}(\mathrm{B}^{\pm} \to \mathrm{D}_2^{*}(\to \mathrm{D}^{(*)}\pi^{\pm})\ell\nu) = 0.51^{+0.18}_{-0.27}\,\% \tag{8.18}$$

$$\mathcal{B}(\mathrm{B}^{0} \to \mathrm{D}_2^{*}(\to \mathrm{D}^{(*)}\pi^{\pm})\ell\nu) = 0.45^{+0.23}_{-0.29}\,\% \tag{8.19}$$

$$\mathcal{B}(\mathrm{B}^{\pm} \to \mathrm{D}_2^{*}(\to \mathrm{D}^{(*)}\pi^{0})\ell\nu) = 0.27^{+0.08}_{-0.10}\,\% \tag{8.20}$$

$$\mathcal{B}(\mathrm{B}^{0} \to \mathrm{D}_2^{*}(\to \mathrm{D}^{(*)}\pi^{0})\ell\nu) = 0.20^{+0.11}_{-0.14}\,\% \tag{8.21}$$

To make the results comparable to the branching fractions used for the production of these decays, the normalization branching fraction is taken as the one used for the scaling of the $\mathrm{B} \to \mathrm{D}^*\ell\nu$ events in the MC. Fig. 8.3, Fig. 8.4, Fig. 8.5, and Fig. 8.6 show the comparison to the expected value.

As expected, the setup performs better for the decays that include a charged pion emission, when compared to the neutral ones. This can be observed in the lower significance values for the neutral pion decays. Additionally, the Asimov fits prove that with the given setup the narrow D^{**} states can be measured more precisely than the broad states due to their clearer structure in the $\Delta\mathrm{M}$ distributions. Nevertheless, the uncertainties on the final results for the branching fractions are quite large. This is caused by the large uncertainty on the normalization component, which cannot be distinguished very well from other components in the fit. So, the statistical error on the branching fractions is dominated by this uncertainty. Since the normalization component can be measured more precisely in the $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi^0)\ell$ reconstruction channels, the branching fractions of the decays $\mathrm{B} \to \mathrm{D}^{**}_{\mathrm{narrow}}(\to \mathrm{D}^{(*)}\pi^0)\ell$ show a lower uncertainty as compared to the $\mathrm{B} \to \mathrm{D}^{**}_{\mathrm{narrow}}(\to \mathrm{D}^{(*)}\pi^{\pm})\ell$ branching fractions. To get more precise results for the D^{**} branching fractions one has to generally improve the determination of the normalization yields. This could be done by expanding the template fit to a two dimensional fit in $\Delta\mathrm{M}$ and the squared missing mass, since the squared missing mass provides a good separation for the $\mathrm{D}^*\ell\nu$ component.

Other than that, the fits show that the branching fractions for the charged $\mathrm{B}^{\pm}$ meson decays can be determined more precisely than the ones of the neutral B^0. The root of this lies in the higher statistics in the charged $\mathrm{B}^{\pm}$ meson reconstruction channels.

When comparing the fitted branching fraction values to the expected ones, which are used for the production of the MC, one expects a closure between the values since the fit is performed on Asimov data which represent exactly the MC composition. This is not true for this study due to the fact that the reconstruction efficiencies are calculated on independent signal MC samples and so the fitted values fluctuate within the statistical errors of these efficiencies around the expected value. Nevertheless, the fitted branching fractions agree very well with the expected ones.

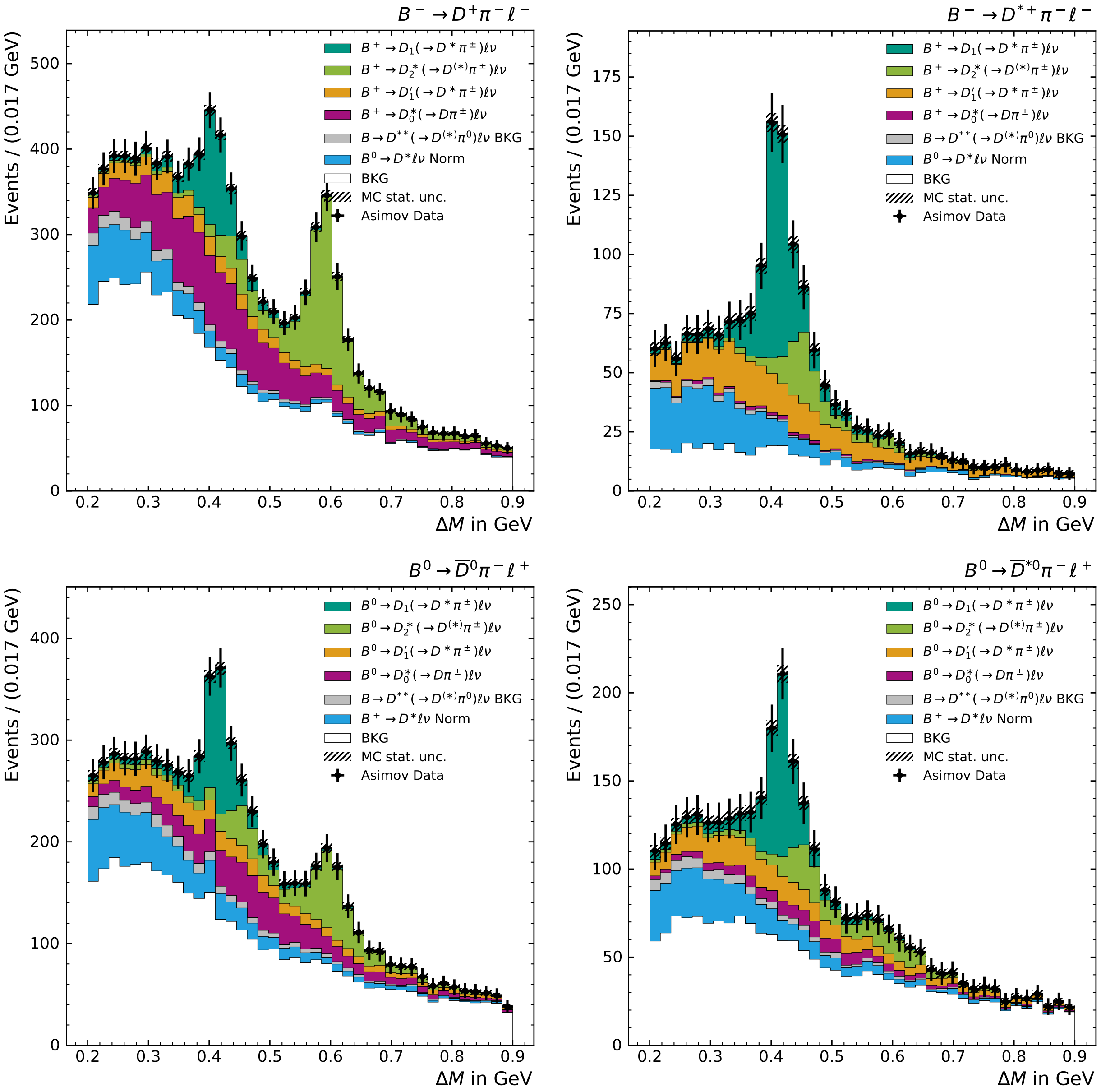


Figure 8.1.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D^{**}(\to D^{(*)}\pi^{\pm})\ell\nu$ signal and $B \to D^{*}\ell\nu$ normalization yields, which are used for the $B \to D^{**}(\to D^{(*)}\pi^{\pm})\ell\nu$ branching fraction determination. The upper two distributions are fitted simultaneously to determine the $B^{\pm} \to D^{**}(\to D^{(*)}\pi^{\pm})\ell\nu$ branching fractions and the lower ones are used for the determination of the $B^0 \to D^{**}(\to D^{(*)}\pi^{\pm})\ell\nu$ branching fractions. The fits are performed on the Asimov data set in 40 ΔM bins with $0.2\,\text{GeV} \leq \Delta M \leq 0.9\,\text{GeV}$.

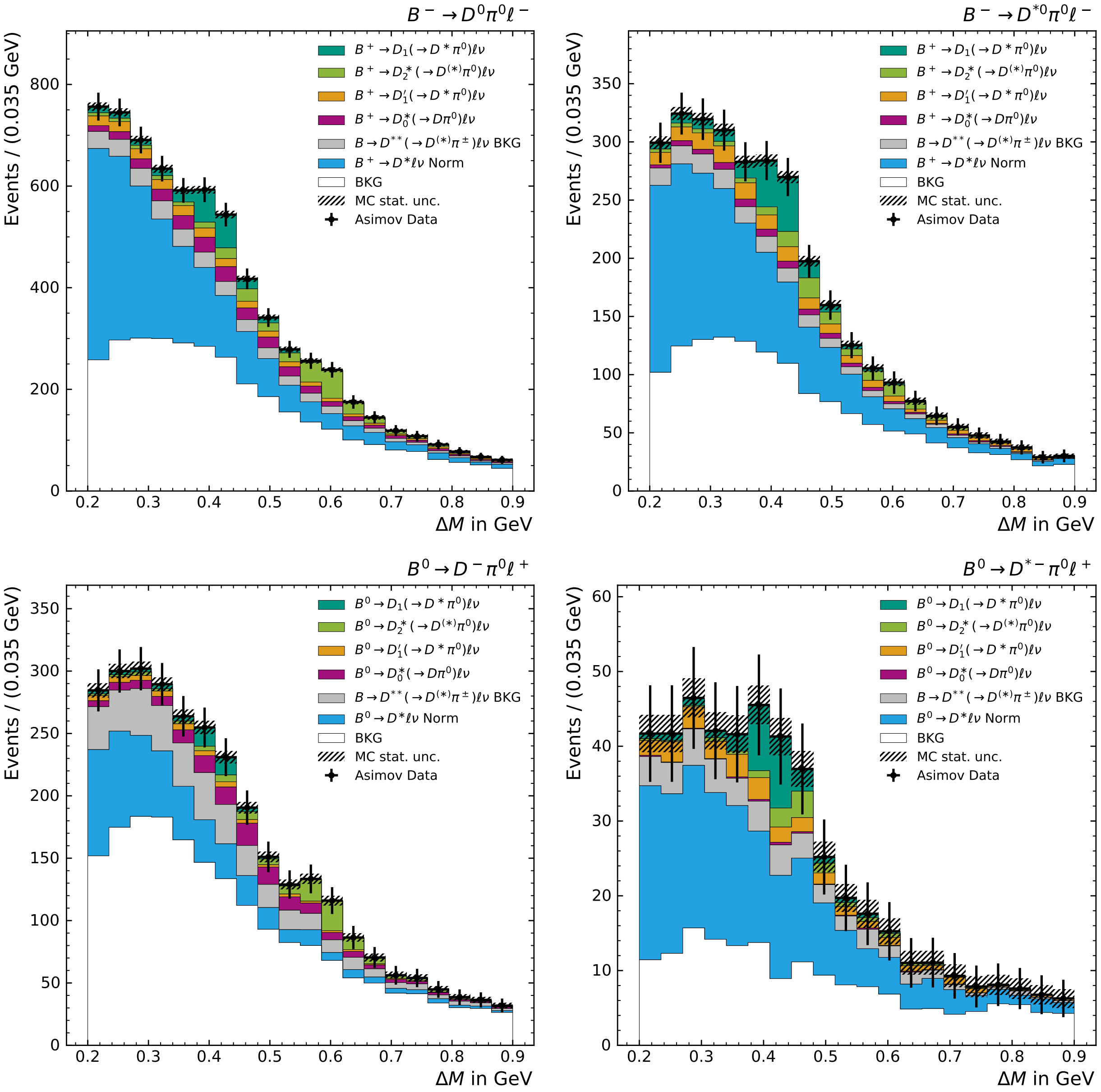


Figure 8.2.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ signal and $B \to D^*\ell\nu$ normalization yields, which are used for the $B \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ branching fraction determination. The upper two distributions are fitted simultaneously to determine the $B^\pm \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ branching fractions and the lower ones are used for the determination of the $B^0 \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ branching fractions. The fits are performed on the Asimov data set in 20 ΔM bins with $0.2\,\mathrm{GeV} \leq \Delta M \leq 0.9\,\mathrm{GeV}$.

Table 8.1.: Extracted signal and normalization yields for the D^{**} branching fraction calculation as measured with a template likelihood fit on Asimov data. The extracted yields are identical to the expected ones, which are therefore not given explicitly. The statistical uncertainty on the yields together with the significance of the fit is also given.

Component	Yield	Uncertainty	Significance Z in σ
$\mathrm{B}^\pm \to \mathrm{D}_0^*(\to \mathrm{D}\pi^\pm)\ell\nu$	1648.12	±275.18	6.03
$\mathrm{B}^\pm \to \mathrm{D}_1'(\to \mathrm{D}^*\pi^\pm)\ell\nu$	934.21	±210.95	4.61
$\mathrm{B}^\pm \to \mathrm{D}_1(\to \mathrm{D}^*\pi^\pm)\ell\nu$	962.78	±74.46	15.01
$\mathrm{B}^\pm \to \mathrm{D}_2^*(\to \mathrm{D}^{(*)}\pi^\pm)\ell\nu$	1316.09	±83.08	17.74
$\mathrm{B}^0 \to \mathrm{D}^*\ell\nu$	1283.01	±415.14	–
$\mathrm{B}^0 \to \mathrm{D}_0^*(\to \mathrm{D}\pi^\pm)\ell\nu$	879.81	±229.33	3.89
$\mathrm{B}^0 \to \mathrm{D}_1'(\to \mathrm{D}^*\pi^\pm)\ell\nu$	880.53	±393.73	2.25
$\mathrm{B}^0 \to \mathrm{D}_1(\to \mathrm{D}^*\pi^\pm)\ell\nu$	939.98	±80.02	12.73
$\mathrm{B}^0 \to \mathrm{D}_2^*(\to \mathrm{D}^{(*)}\pi^\pm)\ell\nu$	863.04	±108.58	8.48
$\mathrm{B}^\pm \to \mathrm{D}^*\ell\nu$	1421.04	±747.44	–
$\mathrm{B}^\pm \to \mathrm{D}_0^*(\to \mathrm{D}\pi^0)\ell\nu$	366.02	±577.4	0.63
$\mathrm{B}^\pm \to \mathrm{D}_1'(\to \mathrm{D}^*\pi^0)\ell\nu$	398.63	±647.04	0.61
$\mathrm{B}^\pm \to \mathrm{D}_1(\to \mathrm{D}^*\pi^0)\ell\nu$	432.71	±127.1	3.45
$\mathrm{B}^\pm \to \mathrm{D}_2^*(\to \mathrm{D}^{(*)}\pi^0)\ell\nu$	390.4	±143.62	2.74
$\mathrm{B}^\pm \to \mathrm{D}^*\ell\nu$	4074.21	±615.21	–
$\mathrm{B}^0 \to \mathrm{D}_0^*(\to \mathrm{D}\pi^0)\ell\nu$	151.64	±158.47	0.96
$\mathrm{B}^0 \to \mathrm{D}_1'(\to \mathrm{D}^*\pi^0)\ell\nu$	84.43	±162.0	0.52
$\mathrm{B}^0 \to \mathrm{D}_1(\to \mathrm{D}^*\pi^0)\ell\nu$	86.38	±51.99	1.70
$\mathrm{B}^0 \to \mathrm{D}_2^*(\to \mathrm{D}^{(*)}\pi^0)\ell\nu$	118.66	±51.75	2.37
$\mathrm{B}^0 \to \mathrm{D}^*\ell\nu$	824.94	±258.94	–

Table 8.2.: Reconstruction efficiencies of the signal and normalization components in the fitted reconstruction modes. The efficiencies are summed over all reconstruction channels which are used for the corresponding fit to extract the yields.

Component	Reconstruction Efficiency
$\mathrm{B}^{\pm} \to \mathrm{D}_0^*(\to \mathrm{D}\pi^{\pm})\ell\nu$	$(4.097^{+0.045}_{-0.320}) \times 10^{-4}$
$\mathrm{B}^{\pm} \to \mathrm{D}_1'(\to \mathrm{D}^*\pi^{\pm})\ell\nu$	$(2.197^{+0.033}_{-0.242}) \times 10^{-4}$
$\mathrm{B}^{\pm} \to \mathrm{D}_1(\to \mathrm{D}^*\pi^{\pm})\ell\nu$	$(2.016^{+0.032}_{-0.232}) \times 10^{-4}$
$\mathrm{B}^{\pm} \to \mathrm{D}_2^*(\to \mathrm{D}^{(*)}\pi^{\pm})\ell\nu$	$(3.340^{+0.040}_{-0.292}) \times 10^{-4}$
$\mathrm{B}^0 \to \mathrm{D}^*\ell\nu$	$(0.161^{+0.010}_{-0.066}) \times 10^{-4}$
$\mathrm{B}^0 \to \mathrm{D}_0^*(\to \mathrm{D}\pi^{\pm})\ell\nu$	$(2.412^{+0.035}_{-0.252}) \times 10^{-4}$
$\mathrm{B}^0 \to \mathrm{D}_1'(\to \mathrm{D}^*\pi^{\pm})\ell\nu$	$(2.026^{+0.032}_{-0.233}) \times 10^{-4}$
$\mathrm{B}^0 \to \mathrm{D}_1(\to \mathrm{D}^*\pi^{\pm})\ell\nu$	$(2.109^{+0.032}_{-0.237}) \times 10^{-4}$
$\mathrm{B}^0 \to \mathrm{D}_2^*(\to \mathrm{D}^{(*)}\pi^{\pm})\ell\nu$	$(2.546^{+0.035}_{-0.258}) \times 10^{-4}$
$\mathrm{B}^{\pm} \to \mathrm{D}^*\ell\nu$	$(0.171^{+0.010}_{-0.068}) \times 10^{-4}$
$\mathrm{B}^{\pm} \to \mathrm{D}_0^*(\to \mathrm{D}\pi^0)\ell\nu$	$(1.709^{+0.029}_{-0.215}) \times 10^{-4}$
$\mathrm{B}^{\pm} \to \mathrm{D}_1'(\to \mathrm{D}^*\pi^0)\ell\nu$	$(1.720^{+0.029}_{-0.216}) \times 10^{-4}$
$\mathrm{B}^{\pm} \to \mathrm{D}_1(\to \mathrm{D}^*\pi^0)\ell\nu$	$(1.771^{+0.030}_{-0.219}) \times 10^{-4}$
$\mathrm{B}^{\pm} \to \mathrm{D}_2^*(\to \mathrm{D}^{(*)}\pi^0)\ell\nu$	$(1.889^{+0.031}_{-0.225}) \times 10^{-4}$
$\mathrm{B}^{\pm} \to \mathrm{D}^*\ell\nu$	$(0.483^{+0.016}_{-0.117}) \times 10^{-4}$
$\mathrm{B}^0 \to \mathrm{D}_0^*(\to \mathrm{D}\pi^0)\ell\nu$	$(0.883^{+0.021}_{-0.157}) \times 10^{-4}$
$\mathrm{B}^0 \to \mathrm{D}_1'(\to \mathrm{D}^*\pi^0)\ell\nu$	$(0.472^{+0.016}_{-0.116}) \times 10^{-4}$
$\mathrm{B}^0 \to \mathrm{D}_1(\to \mathrm{D}^*\pi^0)\ell\nu$	$(0.414^{+0.015}_{-0.108}) \times 10^{-4}$
$\mathrm{B}^0 \to \mathrm{D}_2^*(\to \mathrm{D}^{(*)}\pi^0)\ell\nu$	$(0.736^{+0.020}_{-0.144}) \times 10^{-4}$
$\mathrm{B}^0 \to \mathrm{D}^*\ell\nu$	$(0.092^{+0.008}_{-0.048}) \times 10^{-4}$

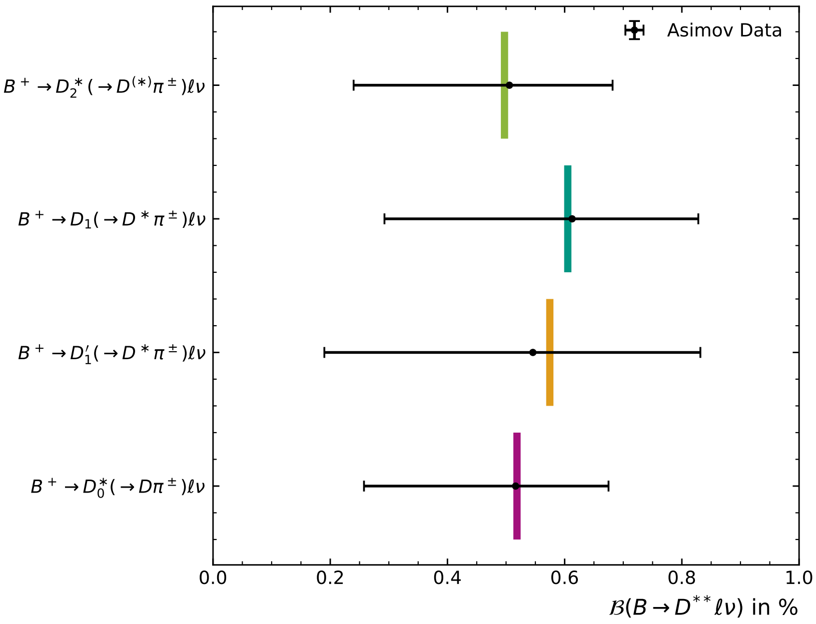


Figure 8.3.: Measured branching fractions of the $B^\pm \to D^{**}(\to D^{(*)}\pi^\pm)\ell\nu$ decays, compared to the nominal values used for the production of the MC events shown as colored vertical bars. The results are based on the Asimov data set and only include the statistical uncertainties.

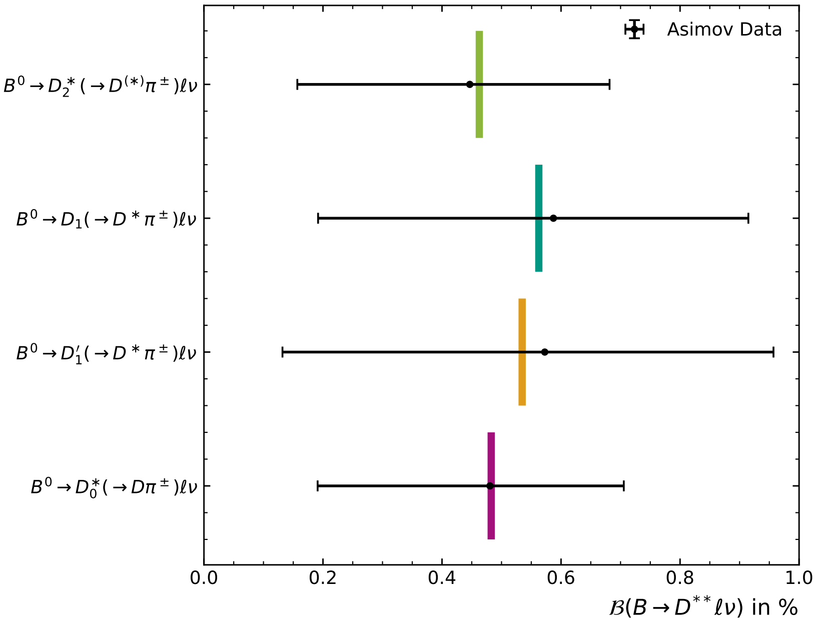


Figure 8.4.: Measured branching fractions of the $B^0 \to D^{**}(\to D^{(*)}\pi^\pm)\ell\nu$ decays, compared to the nominal values used for the production of the MC events shown as colored vertical bars. The results are based on the Asimov data set and only include the statistical uncertainties.

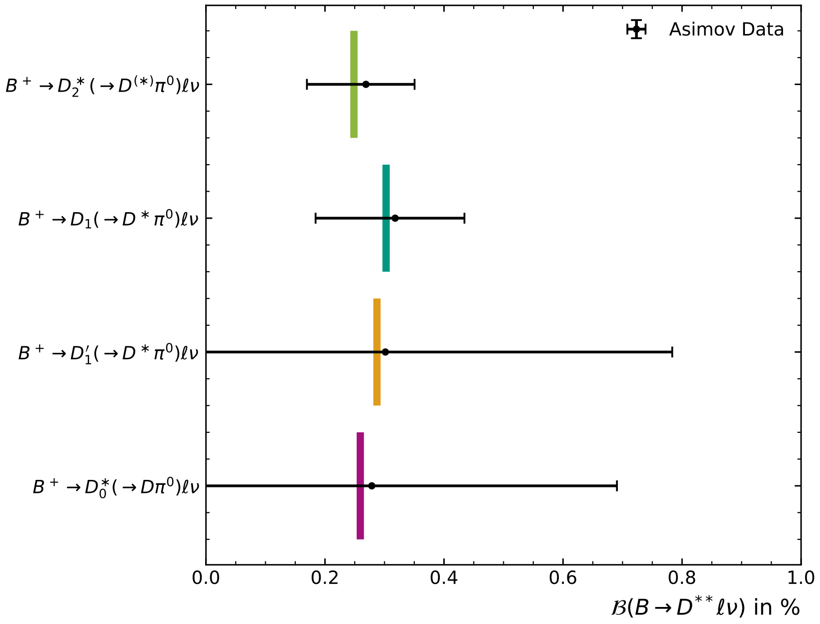


Figure 8.5.: Measured branching fractions of the $B^\pm \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ decays, compared to the nominal values used for the production of the MC events shown as colored vertical bars. The results are based on the Asimov data set and only include the statistical uncertainties.

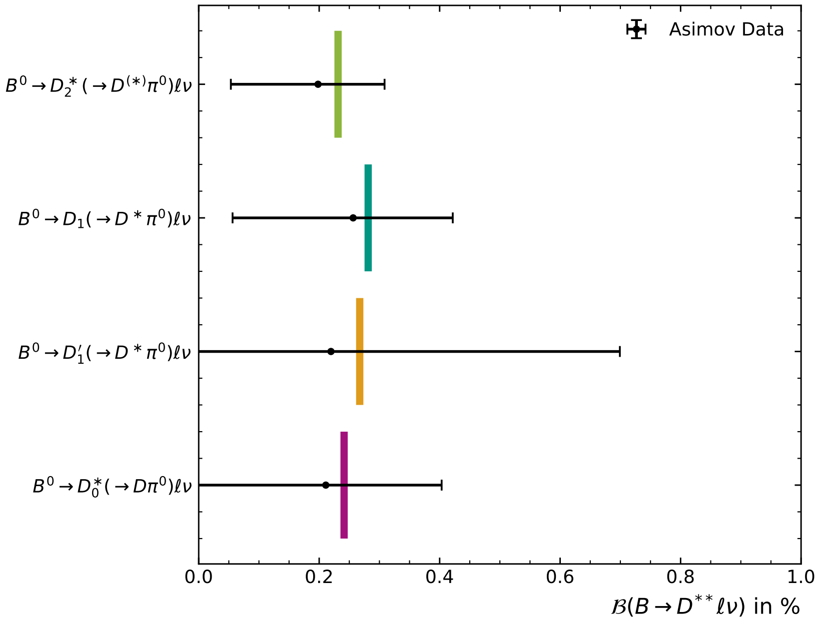


Figure 8.6.: Measured branching fractions of the $B^0 \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ decays, compared to the nominal values used for the production of the MC events shown as colored vertical bars. The results are based on the Asimov data set and only include the statistical uncertainties.

9. Summary and Outlook

This thesis deals with the properties of the semiloptonic $B \to D^{**}(\to D^{(*)}\pi)\ell\nu$ decays, which are not yet fully understood. These decays are one of the major backgrounds for $\mathcal{R}(D^{(*)})$ measurements, which provide a probe of the lepton flavor universality described by the standard model of particle physics. The analysis techniques described in this thesis are developed in the context of an ongoing $\mathcal{R}(D^{(*)})$ measurement and so the $\mathcal{R}(D^{(*)})$ analysis and the studies described herein are highly influenced by each other.

With the aim of analyzing the full Belle data set in the end, this analysis uses the B2BII conversion to make the data processable with the reconstruction software that is developed for the Belle II experiment. Thus, many newly developed algorithms, like the FULL EVENT INTERPRETATION for the hadronic tag-side reconstruction, can be used and provide higher reconstruction efficiencies, which can result in more statistically significant results as in previous measurements. The reconstructed signal candidates are separated from background events via a mixture of the application of multivariate classifiers on the reconstructed data and a cut-based approach.

To account for the quite outdated properties of some particles in the Belle Monte Carlo simulations (MC), several adjustments are made to correct the MC. Differences in the reconstruction efficiencies between data and MC as well as outdated branching fractions in the MC production are considered as correction weights in the MC events. Since the masses and widths of the four D^{**} states (D_0^*, D_1', D_1, and D_2^*) have been measured more precisely since the production of the Belle MC, events containing one of these states are replaced by newly generated and simulated MC samples. Additionally, the $B \to D^*\ell\nu$ decays are reweighted form the CLN to the BGL parameterization. The $B \to D^{**}(\to D^{(*)}\pi)\ell\nu$ signal decays are reweighted from ISGW2 to the LLSW model. The development of a procedure to evaluate how well this model describes the D^{**} decays is the aim of the first study described in this thesis.

The study on the form factors of the D^{**} decays is performed by comparing the momentum transfer q^2 and lepton momentum p_ℓ^* distributions between MC and data. This is done, independent of the involved D^{**} state, for the four decays, $B \to D^{**}(\to D\pi^\pm)\ell\nu$, $B \to D^{**}(\to D^*\pi^\pm)\ell\nu$, $B \to D^{**}(\to D\pi^0)\ell\nu$, and $B \to D^{**}(\to D^*\pi^0)\ell\nu$, separately. Since the signal region is still blinded at the time of writing this thesis, the described concept is tested on Asimov data to study the sensitivity of the approach and to test the template likelihood fit, which is used for the extraction of the signal yields. These tests show good results, especially for the decays involving charged pions in the D^{**} decay. Once the study is performed on the Belle data set, one has to check whether the MC representation describes

the data or not. A discrepancy between data and MC in the q^2 and p^*_ℓ distributions could be resolved by an additional correction factor or an additional systematic uncertainty for measurements involving such decays. The knowledge gained via this study helps to reduce the systematic uncertainty caused by the shapes of the $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ decays and makes the result of the underlying $\mathcal{R}(\mathrm{D}^{(*)})$ more precise compared to previous measurements.
To potentially reduce the systematic uncertainty on the ongoing $\mathcal{R}(\mathrm{D}^{(*)})$ measurement even more, in this thesis an additional way of measuring the branching fractions of each $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}\pi^{(0,\pm)})\ell\nu$ decay, with $\mathrm{D}^{**} = \mathrm{D}^*_0, \mathrm{D}'_1, \mathrm{D}_1, \mathrm{D}^*_2$, is introduced. The measurement of these branching fractions is based on a template likelihood fit on the reconstructed mass difference between the D^{**} meson and the $\mathrm{D}^{(*)}$ it decays into. The branching fraction is then calculated via the ratio of the respective signal decay to the $\mathrm{B} \to \mathrm{D}^*\ell\nu$ normalization process. Similar to the form factor study, the measurement is validated on Asimov data, which shows that the general approach works and gives the expected results for the branching fractions. Nevertheless, at the time of writing, the statistical uncertainties on the determined branching fractions are still quite large and have to be reduced to make the results reliable. Since the large uncertainties for the described fitting setup are mainly caused by the imprecise determination of the $\mathrm{B} \to \mathrm{D}^*\ell\nu$ normalization component, there are several approaches on how one could make the determination more accurate. Instead of fitting only the distributions in $\Delta\mathrm{M}$, the template fit can be expanded to a two dimensional fit in $\Delta\mathrm{M}$ and the squared missing mass $\mathrm{M}^2_{\mathrm{miss}}$ since the latter provides a better separation for the normalization component than $\Delta\mathrm{M}$. Another approach would be to add dedicated $\mathrm{B} \to \mathrm{D}^*\ell\nu$ reconstruction channels to the fit to improve the sensitivity on these decays. Both methods will be tested and validated.
Additionally, pull studies and linearity tests, so-called toy studies, have to be performed to further validate the fitting procedure for the $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ branching fraction measurements as well as in the form factor study. For the moment the studies only include the statistical uncertainties and so the systematic uncertainties also have to be included. Nevertheless, most of the systematic uncertainties will cancel each other out since the studies are based on ratios of fitted parameters.
Once the described studies are completely validated and the measurements are performed on the Belle data set, they will give a better understanding on semileptonic $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ decays, which then can be used in many upcoming analysis. The first measurement, which is affected by the studies provided in this thesis, is the underlying $\mathcal{R}(\mathrm{D}^{(*)})$ measurement. The systematic errors on this result, which are in large fractions caused by the $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ background, will be reduced compared to previous measurements by the better understanding of the background decays, which is gathered with the methods described in this thesis. As soon as the $\mathcal{R}(\mathrm{D}^{(*)})$ measurement is done, the results will show if the tension with the standard model prediction remains or even gets larger. This would hold the door open for theories, which describe models beyond the standard model and are able to explain the observed discrepancy.
Additionally, the findings of the described studies can be included in the upcoming MC production campaigns for the Belle II experiment.

Danksagung

Zuletzt möchte ich mich bei all denjenigen bedanken, die mich während der Anfertigung dieser Arbeit unterstützt haben.

Ich bedanke mich bei Prof. Dr. Günter Quast für die Übernahme des Referats.

Bei Prof. Dr. Florian Bernlochner möchte ich mich für die Übernahme des Koreferats, sowie die Betreuung und gebotenen Möglichkeiten während des Masterstudiums bedanken.

Ein besonderer Dank gilt Felix Metzner, für die gute Zusammenarbeit und Betreuung dieser Arbeit, sowie für das intensive Korrekturlesen.

Dr. Pablo Goldenzweig möchte ich ebenfalls für die Unterstützung dieser Arbeit danken.

Auch möchte ich mich beim gesamten ETP für die immer hilfsbereite und freundliche Arbeitsatmosphäre, die ich sehr genossen habe, bedanken.

Zuletzt möchte ich noch meiner Familie und Freunden danken, ohne die das Studium in dieser Form sicher nicht möglich gewesen wäre. Besonders hervorheben möchte ich hierbei meine Eltern, die mir während des gesamten Studiums bedingungslos zur Seite standen.

Bibliography

[1] L. M. Brown and R. T. Weidner, "Physics." `https://www.britannica.com/science/physics-science`. [Online; Accessed 18 February 2021].

[2] R. Feynman, "Seeking new Laws." `https://jamesclear.com/great-speeches/seeking-new-laws-by-richard-feynman`, 1964. [Online; Accessed 19 February 2021], lecture was held at Cornell university in November 1964.

[3] M. Huschle, *Measurement of the branching ratio of* $B \to D^{(*)}\tau\nu_\tau$ *relative to* $B \to D^{(*)}l\nu_l$ *decays with hadronic tagging at Belle*. PhD thesis, Karlsruhe Institute of Technology (KIT), 2015. `https://inspirehep.net/literature/1426741`.

[4] J. P. Lees, V. Poireau, V. Tisserand, E. Grauges, *et al.*, "Measurement of an excess of $B^- \to D^{(*)}\tau^-\nu_\tau$ decays and implications for charged Higgs bosons," *Physical Review D* **88** no. 7, (Oct, 2013) , Oct, 2013. `http://dx.doi.org/10.1103/PhysRevD.88.072012`.

[5] G. Caria, P. Urquijo, I. Adachi, H. Aihara, *et al.*, "Measurement of R (D) and R(D*) with a Semileptonic Tagging Method," *Physical Review Letters* **124** (04, 2020) , 04, 2020.

[6] F. U. Bernlochner and Z. Ligeti, "Semileptonic B(s) decays to excited charmed mesons with e, μ, τ and searching for new physics with R(D**)," *Physical Review D* **95** no. 1, (Jan, 2017) , Jan, 2017. `https://arxiv.org/abs/1606.09300`.

[7] **HFLAV**, Y. S. Amhis *et al.*, "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018," `arXiv:1909.12524 [hep-ex]`, 2019. updated results and plots available at `https://hflav.web.cern.ch/`.

[8] J. Brodzicka, T. Browder, P. Chang, S. Eidelman, *et al.*, "Physics Achievements from the Belle Experiment," 2012. `https://arxiv.org/abs/1212.5342`.

[9] A. J. Bevan, B. Golob, T. Mannel, S. Prell, *et al.*, "The Physics of the B Factories," *The European Physical Journal C* **74** no. 11, (Nov, 2014) , Nov, 2014. `http://dx.doi.org/10.1140/epjc/s10052-014-3026-9`.

[10] S. Kurokawa and E. Kikutani, "Overview of the KEKB accelerators," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **499** no. 1, (2003) 1–7, 2003.

https://www.sciencedirect.com/science/article/pii/S0168900202017710. KEK-B: The KEK B-factory.

[11] Wikipedia, "KEKB (accelerator) — Wikipedia, the free encyclopedia." http://en.wikipedia.org/w/index.php?title=KEKB%20(accelerator)&oldid=989293585, 2020. [Online; accessed 25-November-2020].

[12] A. Abashian, K. Gotow, N. Morgan, L. Piilonen, *et al.*, "The Belle detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **479** no. 1, (2002) 117 – 232, 2002. http://www.sciencedirect.com/science/article/pii/S0168900201020137. Detectors for Asymmetric B-factories.

[13] J. Schieck, "DEPFET pixels as a vertex detector for the Belle II experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **732** (Dec, 2013) 160–163, Dec, 2013. http://dx.doi.org/10.1016/j.nima.2013.05.054.

[14] T. Abe, I. Adachi, K. Adamczyk, S. Ahn, *et al.*, "Belle II Technical Design Report," 2010. https://arxiv.org/abs/1011.0352.

[15] R. Itoh, "BASF - BELLE AnalysiS Framework," in *9th International Conference on Computing in High-Energy and Nuclear Physics*. 4, 1997.

[16] A. Moll, "The Software Framework of the Belle II Experiment," *Journal of Physics: Conference Series* **331** no. 3, (Dec, 2011) 032024, Dec, 2011. https://doi.org/10.1088/1742-6596/331/3/032024.

[17] T. Keck, F. Abudinén, F. U. Bernlochner, R. Cheaib, *et al.*, "The Full Event Interpretation," *Computing and Software for Big Science* **3** no. 1, (Feb, 2019) , Feb, 2019. http://dx.doi.org/10.1007/s41781-019-0021-8.

[18] M. Gelb, T. Keck, M. Prim, H. Atmacan, *et al.*, "B2BII: Data Conversion from Belle to Belle II," *Computing and Software for Big Science* **2** no. 1, (Nov, 2018) , Nov, 2018. https://arxiv.org/abs/1810.00019.

[19] N. Katayama, R. Itoh, A. Manabe, and T. Sasaki, "Belle computing model," *Computer Physics Communications* **110** no. 1, (1998) 22–25, 1998. https://www.sciencedirect.com/science/article/pii/S0010465597001483.

[20] R. Brun and F. Rademakers, "ROOT — An object oriented data analysis framework," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **389** no. 1, (1997) 81–86, 1997. https://www.sciencedirect.com/science/article/pii/S016890029700048X. New Computing Techniques in Physics Research V.

[21] D. J. Lange, "The EvtGen particle decay simulation package," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **462** no. 1, (2001) 152 – 155, 2001. http://www.sciencedirect.com/science/article/pii/S0168900201000894. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[22] T. Sjöstrand, P. Edén, C. Friberg, L. Lönnblad, *et al.*, “High-energy-physics event generation with Pythia 6.1,” *Computer Physics Communications* **135** no. 2, (Apr, 2001) 238–259, Apr, 2001. `https://arxiv.org/abs/hep-ph/0010017`.

[23] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, *GEANT 3: user’s guide Geant 3.10, Geant 3.11; rev. version.* CERN, Geneva, 1987. `https://cds.cern.ch/record/1119728`.

[24] **Particle Data Group**, P. Zyla *et al.*, “Review of Particle Physics,” *PTEP* **2020** no. 8, (2020) 083C01, 2020.

[25] F. U. Bernlochner, Z. Ligeti, and S. Turczyk, “Proposal to solve a few puzzles in semileptonic B decays,” 2013. `https://arxiv.org/abs/1301.4448`.

[26] M. Feindt, F. Keller, M. Kreps, T. Kuhr, *et al.*, “A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **654** no. 1, (2011) 432 – 440, 2011. `http://www.sciencedirect.com/science/article/pii/S0168900211011193`.

[27] P. Ecker, “Study of Bremsstrahlung Finding at the Belle II Experiment,” Bachelor’s thesis, Karlsruhe Institute of Technology (KIT), 2018. `https://publish.etp.kit.edu/record/21390`.

[28] M. Gelb, *Search for the Rare Decay* $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ *with the Full Event Interpretation at the Belle Experiment.* PhD thesis, Karlsruhe Institute of Technology (KIT), 2018. `https://publish.etp.kit.edu/record/21546`.

[29] M. Röhrken, *Time-Dependent CP Violation Measurements in Neutral B Meson to Double-Charm Decays at the Japanese Belle Experiment.* PhD thesis, Karlsruhe Institute of Technology (KIT), 2012. `https://publish.etp.kit.edu/record/20864`.

[30] D. M. Asner, M. Athanas, D. W. Bliss, W. S. Brower, *et al.*, “Search for exclusive charmless hadronic B decays,” *Phys. Rev. D* **53** (Feb, 1996) 1039–1050, Feb, 1996. `https://link.aps.org/doi/10.1103/PhysRevD.53.1039`.

[31] G. C. Fox and S. Wolfram, “Event shapes in e+e- annihilation,” *Nuclear Physics B* **149** no. 3, (1979) 413 – 496, 1979. `http://www.sciencedirect.com/science/article/pii/0550321379900038`.

[32] T. Keck, “FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification,” 2016. `https://arxiv.org/abs/1609.06119`.

[33] L. Hinz, “Lepton ID efficiency correction and systematic error,” 2006. Belle Note 954, Internal Document.

[34] P. Urquijo, “Semi-inclusive semileptonic B decays,” 2012. Belle Note 1251, Internal Document.

[35] S. Nishida, “Study of Kaon and Pion Identification Using Inclusive D* Sample,” 2005. Belle Note 779, Internal Document.

[36] E. White, "Determination of K_s Efficiency and Systematic Uncertainty," 2011. Belle Note 1207, Internal Document.

[37] W. Dungel, "Systematic investigation of the reconstruction efficiency of low momentum $\pi^{\pm}$ and π^0," 2011. Belle Note 1176, Internal Document.

[38] J. Schwab, "Calibration of the Full Event Interpretation for the Belle and the Belle II Experiment," Master's thesis, Karlsruhe Institute of Technology (KIT), 2017. `https://publish.etp.kit.edu/record/21422`.

[39] I. Caprini, L. Lellouch, and M. Neubert, "Dispersive bounds on the shape of form factors," *Nuclear Physics B* **530** no. 1-2, (Oct, 1998) 153–181, Oct, 1998. `https://arxiv.org/abs/hep-ph/9712417`.

[40] C. G. Boyd, B. Grinstein, and R. F. Lebed, "Constraints on Form Factors for Exclusive Semileptonic Heavy to Light Meson Decays," *Physical Review Letters* **74** no. 23, (Jun, 1995) 4603–4606, Jun, 1995. `https://arxiv.org/abs/hep-ph/9412324`.

[41] R. Glattauer, C. Schwanda, A. Abdesselam, I. Adachi, *et al.*, "Measurement of the decay $B \to D\ell\nu$ in fully reconstructed events and determination of the Cabibbo-Kobayashi-Maskawa matrix element|Vcb|," *Physical Review D* **93** no. 3, (Feb, 2016) , Feb, 2016. `https://arxiv.org/abs/1510.03657`.

[42] D. Scora and N. Isgur, "Semileptonic meson decays in the quark model: An update," *Physical Review D* **52** no. 5, (Sep, 1995) 2783–2812, Sep, 1995. `https://arxiv.org/abs/hep-ph/9503486`.

[43] A. K. Leibovich, Z. Ligeti, I. W. Stewart, and M. B. Wise, "Semileptonic B decays to excited charmed mesons," *Physical Review D* **57** no. 1, (Jan, 1998) 308–330, Jan, 1998. `https://arxiv.org/abs/hep-ph/9705467`.

[44] M. Welsch, "Measurement of the branching fraction of $B \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$ in the context of a $R(D^*)$ analysis at the Belle experiment," Master's thesis, Karlsruhe Institute of Technology (KIT), 2019. `https://publish.etp.kit.edu/record/21909`.

[45] G. Cowan, "Statistical data analysis," 1998. `http://www.gbv.de/dms/goettingen/241026571.pdfhttps://ebookcentral.proquest.com/lib/kxp/detail.action?docID=49634796;http://lib.myilibrary.com/?id=81976`.

[46] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C* **71** no. 2, (Feb, 2011) , Feb, 2011. `http://dx.doi.org/10.1140/epjc/s10052-011-1554-0`.

[47] C. Hensel and K. Kröninger, *Statistical Methods Commonly Used in High Energy Physics*, ch. 10, pp. 329–356. John Wiley & Sons, Ltd. `https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527653416.ch10`.

[48] F. Bernlochner, *Determination of the CKM matrix element $|V_{cb}|$, the $B \to X_s\gamma$ decay rate, and the b-quark mass.* PhD thesis, HU Berlin, 2011. `https://www.physik.hu-berlin.de/de/eephys/babar/theses/pdfs/bernlochner.pdf`.

[49] F. James and M. Roos, "Minuit - a system for function minimization and analysis of the parameter errors and correlations," *Computer Physics Communications* **10** no. 6, (1975) 343 – 367, 1975.
`http://www.sciencedirect.com/science/article/pii/0010465575900399`.

[50] H. Dembinski and P. O. et al., "scikit-hep/iminuit," Dec, 2020.
`https://doi.org/10.5281/zenodo.4310361`.

A. Additional Fit Results for the q^2 Form Factor Study

In this appendix all remaining fit results on q^2, which are used for the determination of the ratios for the form factor study and are not shown earlier, are collected. All fits are performed on Asimov data.

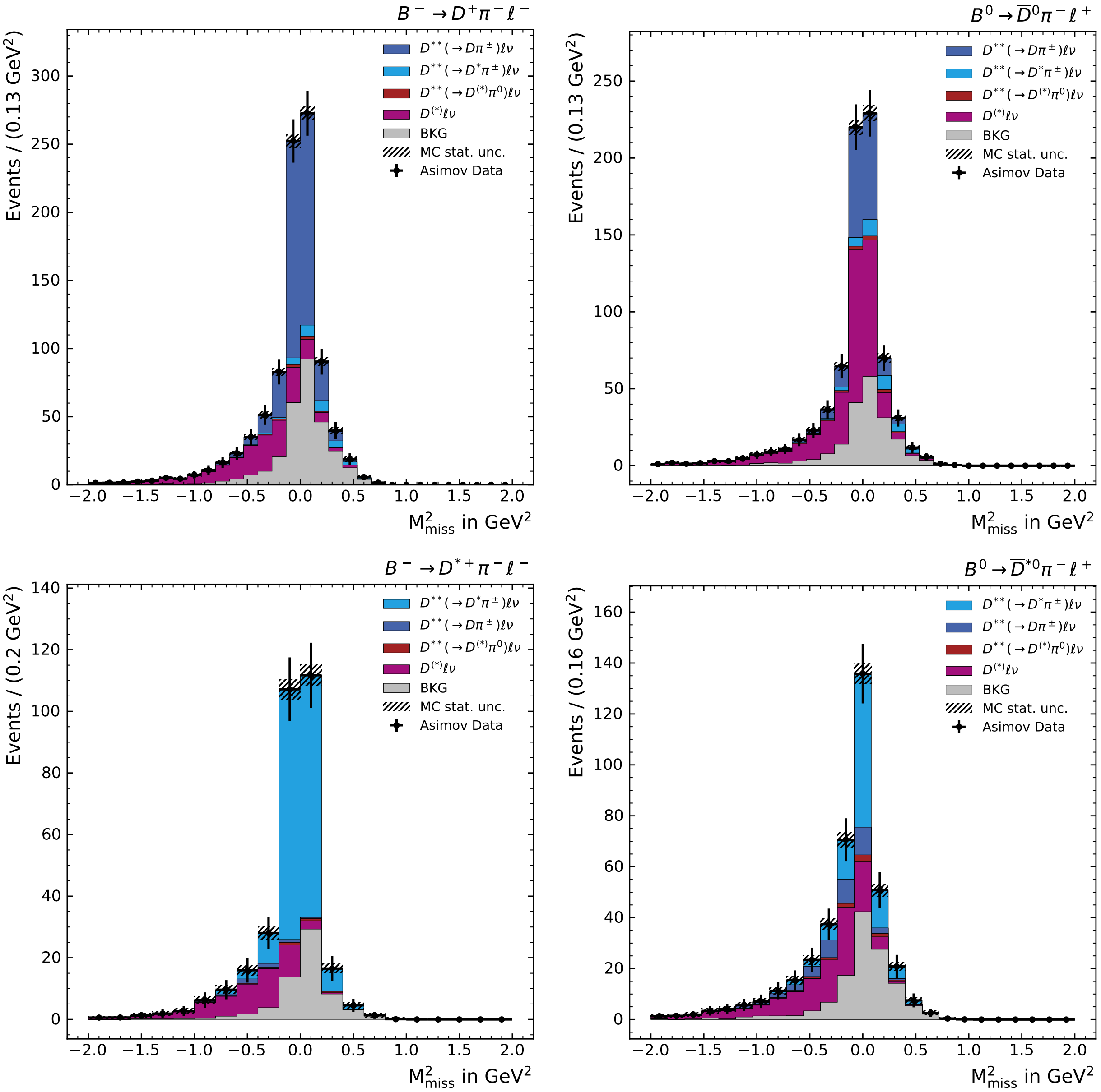


Figure A.1.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^{\pm}\ell\nu$ and $B \to D^{*}\pi^{\pm}\ell\nu$ signal yields. The fit is performed on the Asimov data set with $0\,\mathrm{GeV}^2 \leq q^2 < 1\,\mathrm{GeV}^2$.

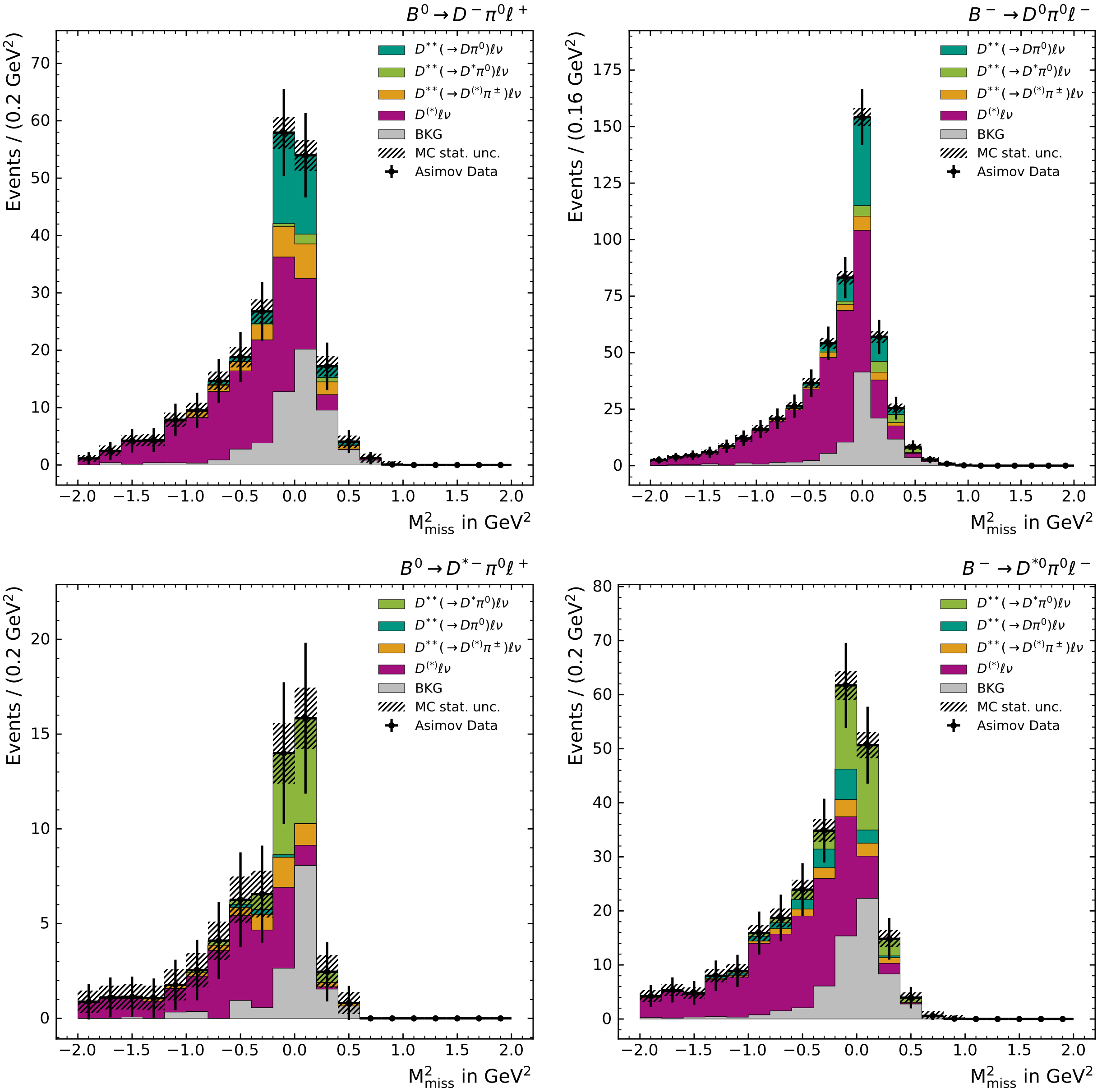


Figure A.2.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^0\ell\nu$ and $B \to D^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $0\,\mathrm{GeV}^2 \leq q^2 < 1\,\mathrm{GeV}^2$.

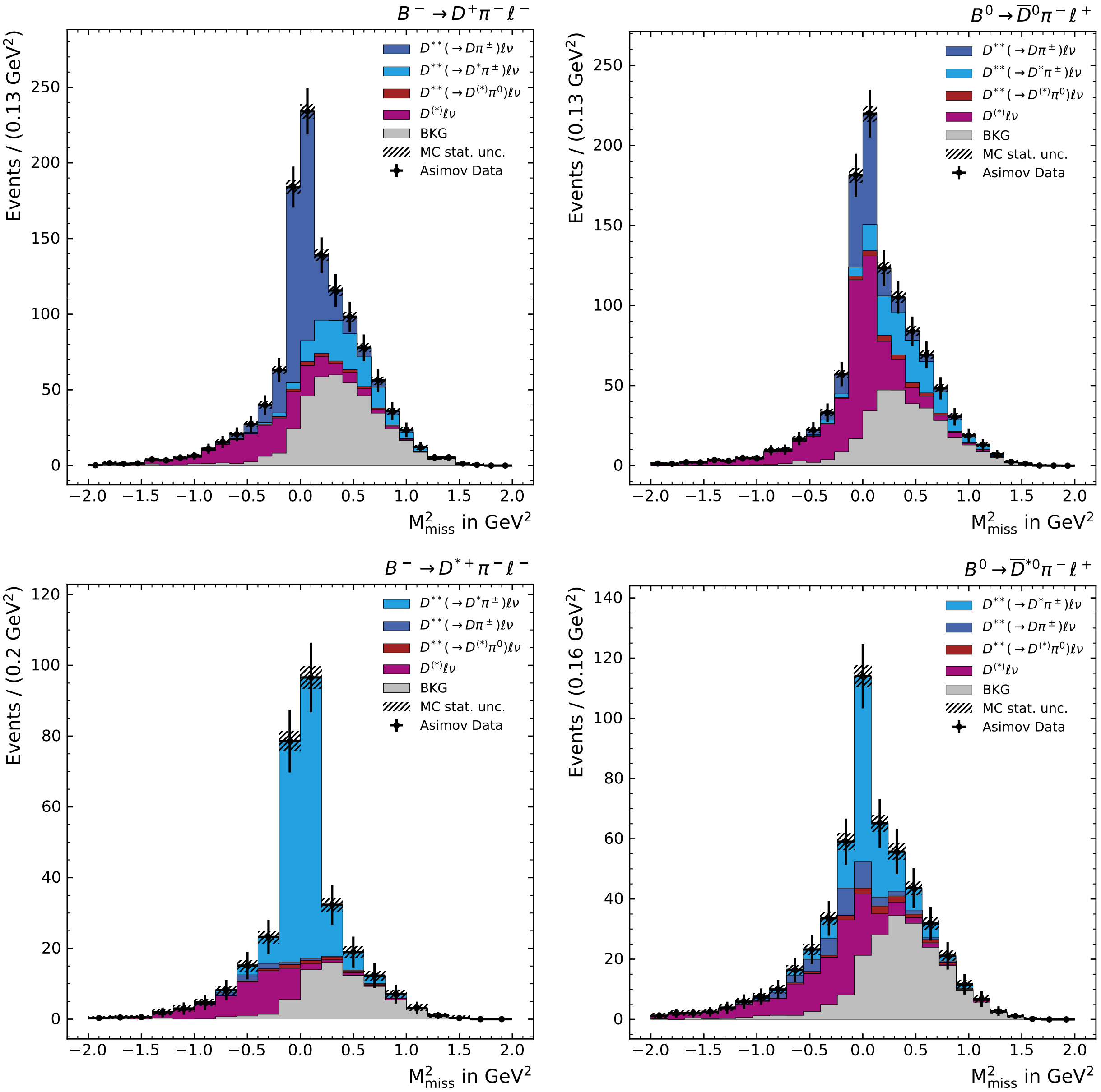


Figure A.3.: Post-fit distributions of the template likelihood fit for the extraction of the $\mathrm{B} \to \mathrm{D}\pi^\pm\ell\nu$ and $\mathrm{B} \to \mathrm{D}^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1\,\mathrm{GeV}^2 \leq \mathrm{q}^2 < 2\,\mathrm{GeV}^2$.

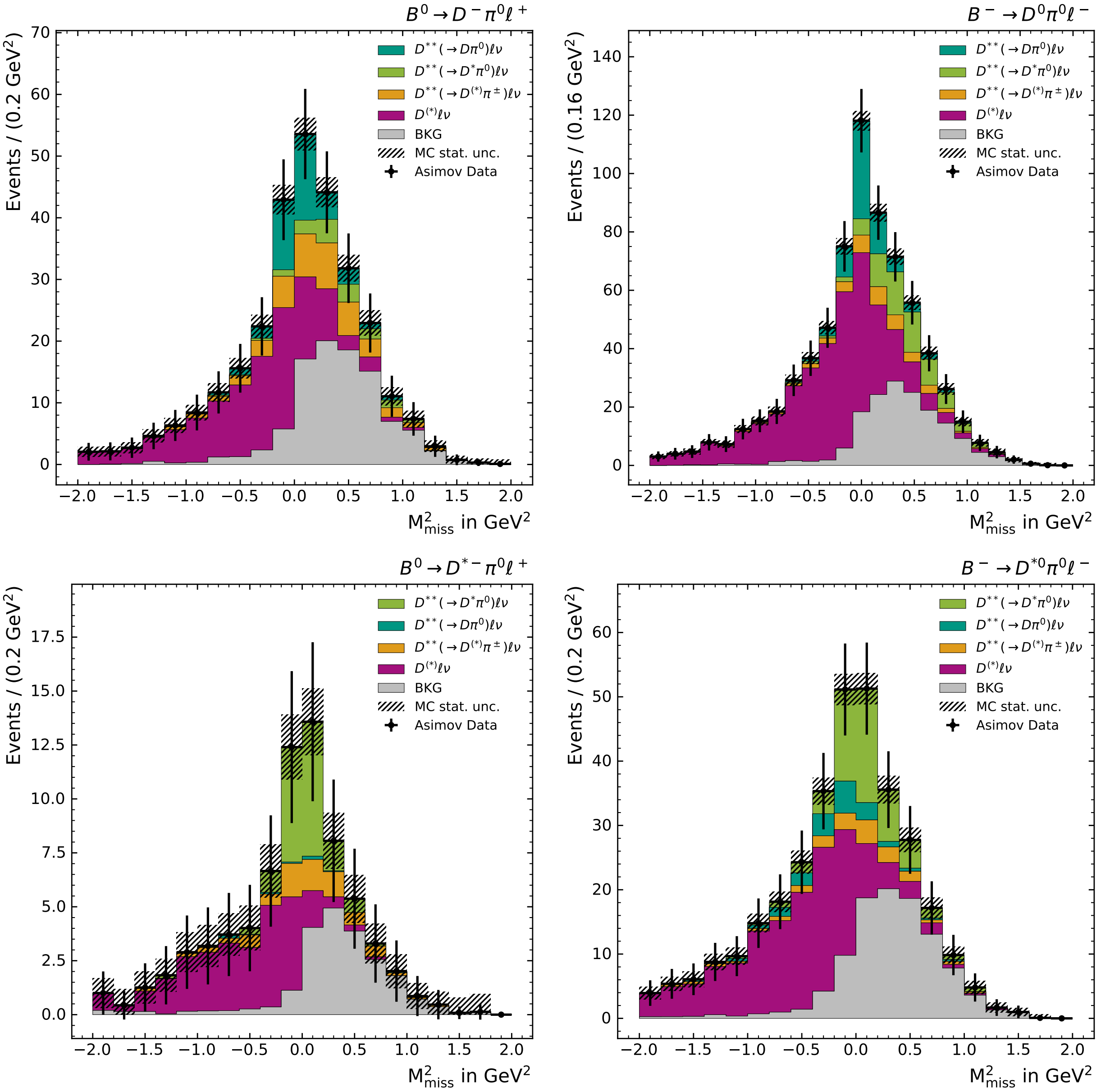


Figure A.4.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^0\ell\nu$ and $B \to D^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1\,\mathrm{GeV}^2 \leq q^2 < 2\,\mathrm{GeV}^2$.

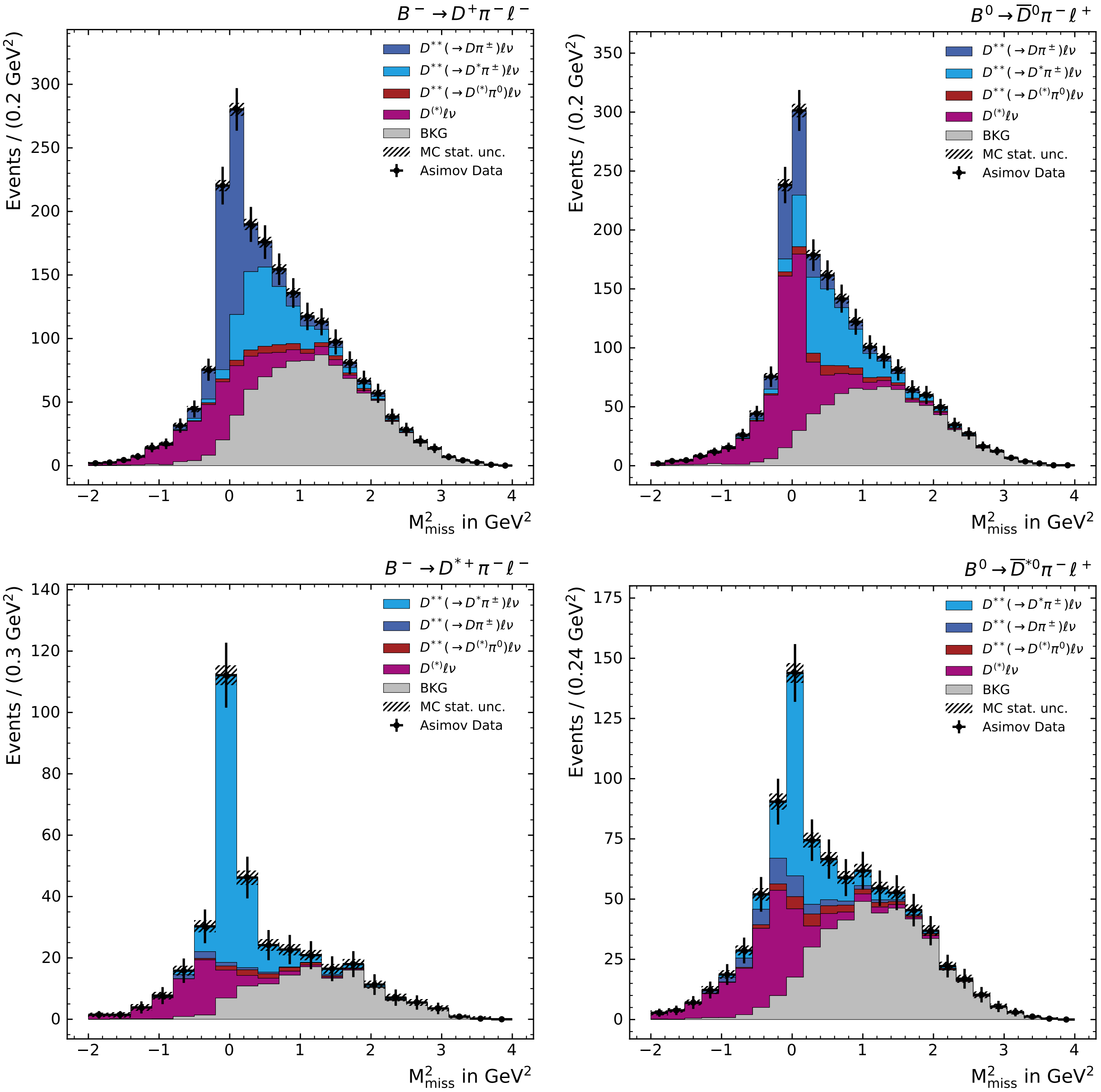


Figure A.5.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $3\,\mathrm{GeV}^2 \leq q^2 < 4.5\,\mathrm{GeV}^2$.

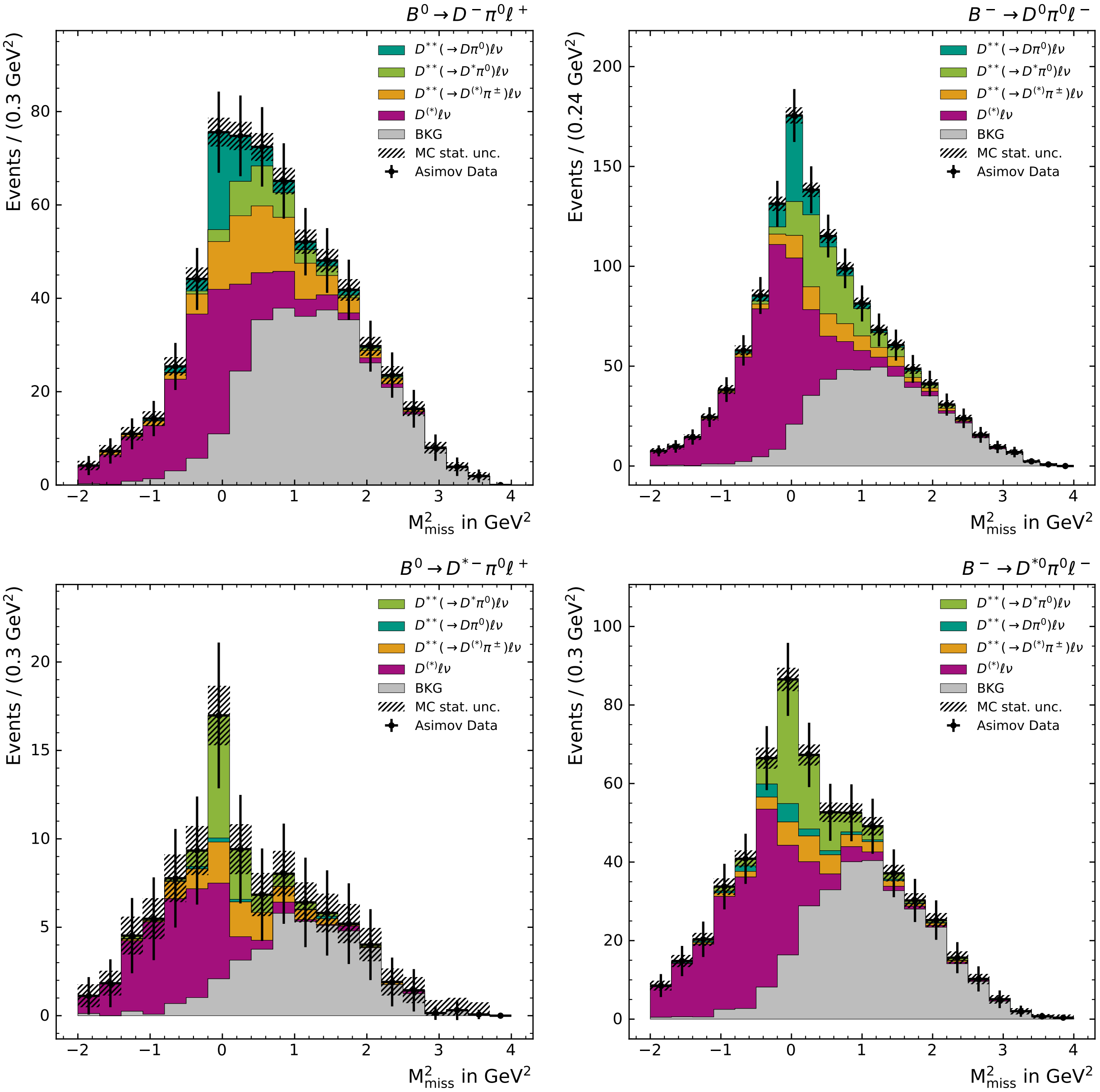


Figure A.6.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^0\ell\nu$ and $B \to D^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $3\,\text{GeV}^2 \leq q^2 < 4.5\,\text{GeV}^2$.

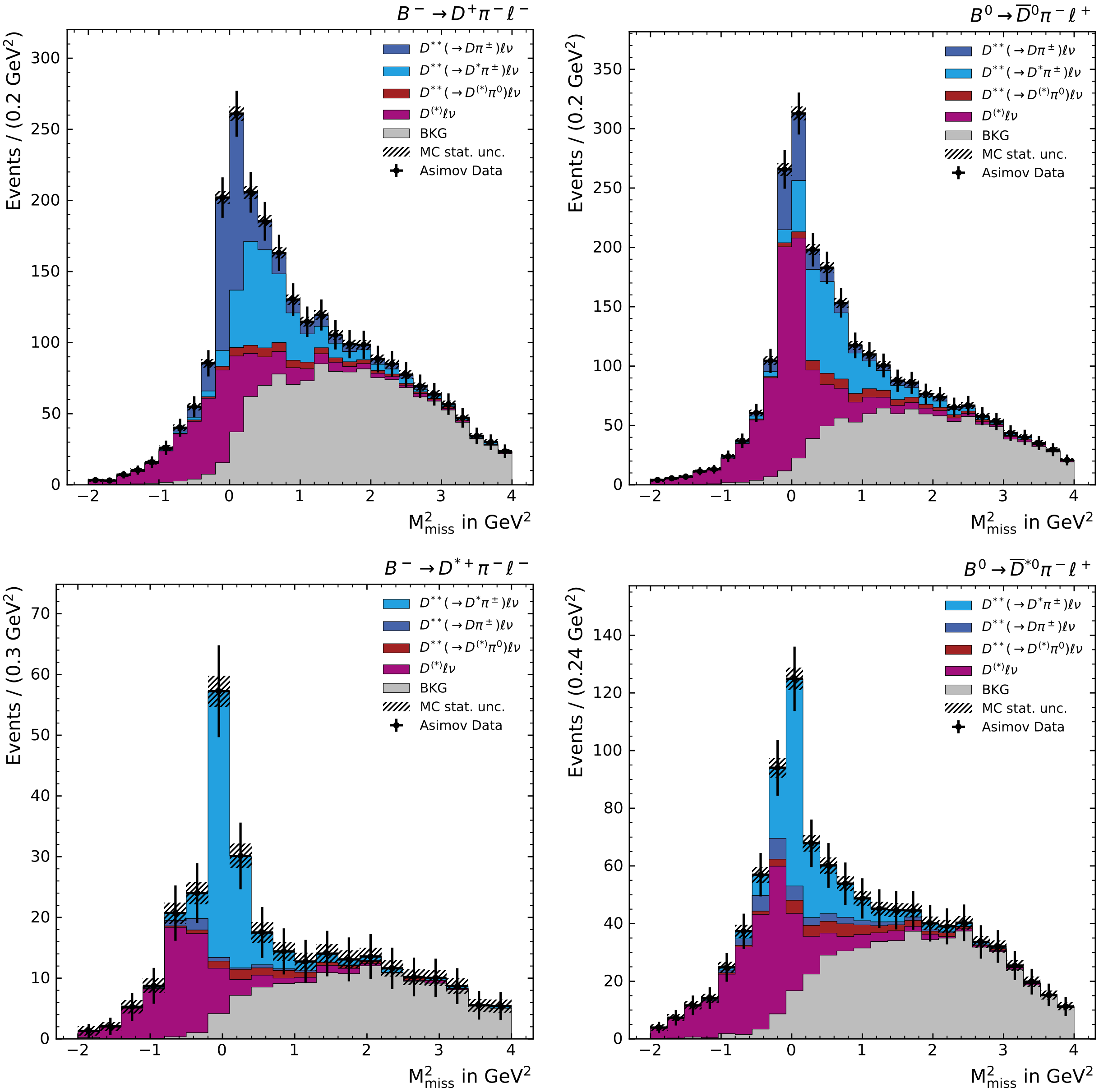


Figure A.7.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $4.5\,\mathrm{GeV}^2 \leq q^2 < 7\,\mathrm{GeV}^2$.

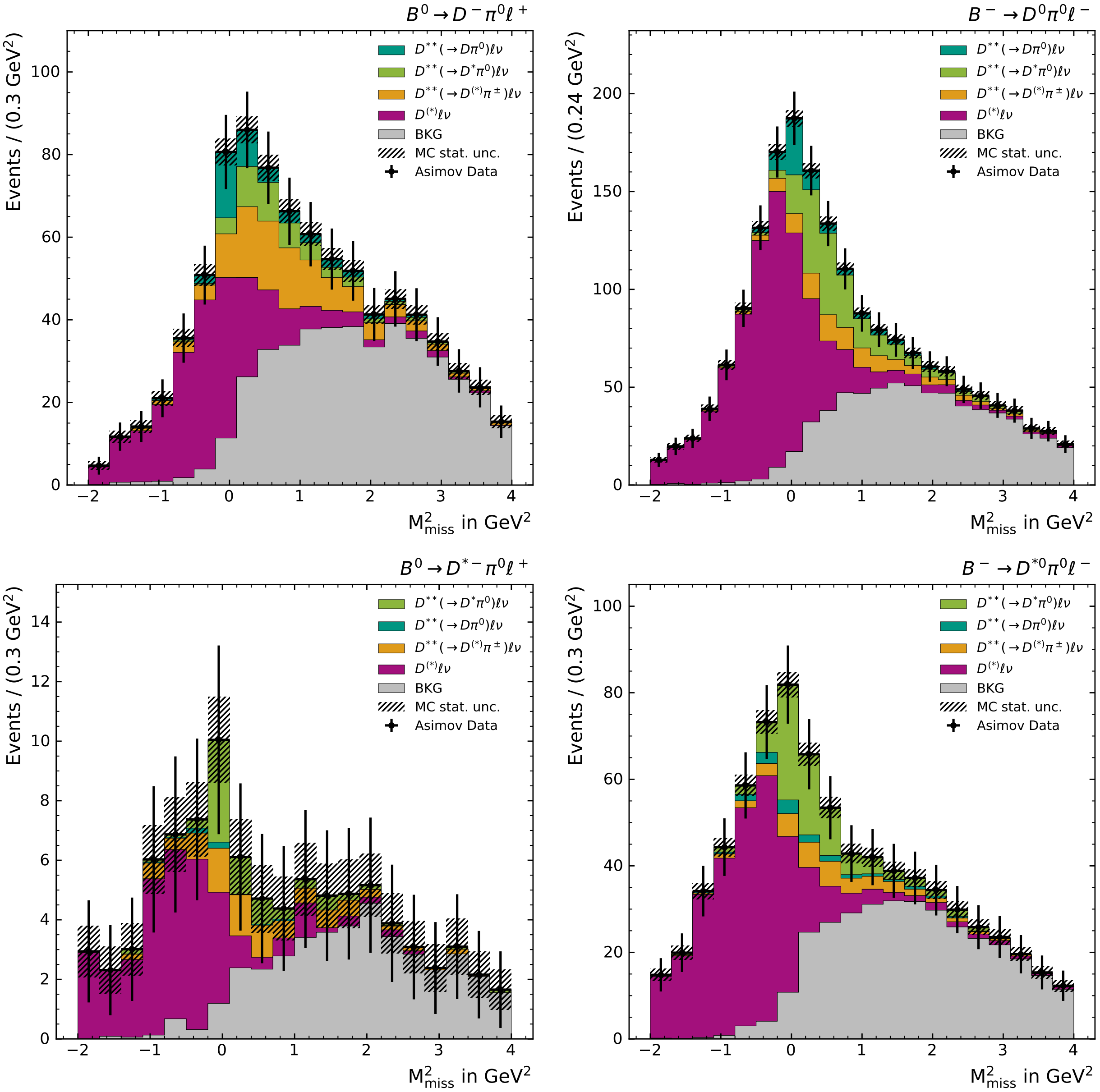


Figure A.8.: Post-fit distributions of the template likelihood fit for the extraction of the $\mathrm{B} \to \mathrm{D}\pi^0\ell\nu$ and $\mathrm{B} \to \mathrm{D}^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $4.5\,\mathrm{GeV}^2 \leq q^2 < 7\,\mathrm{GeV}^2$.

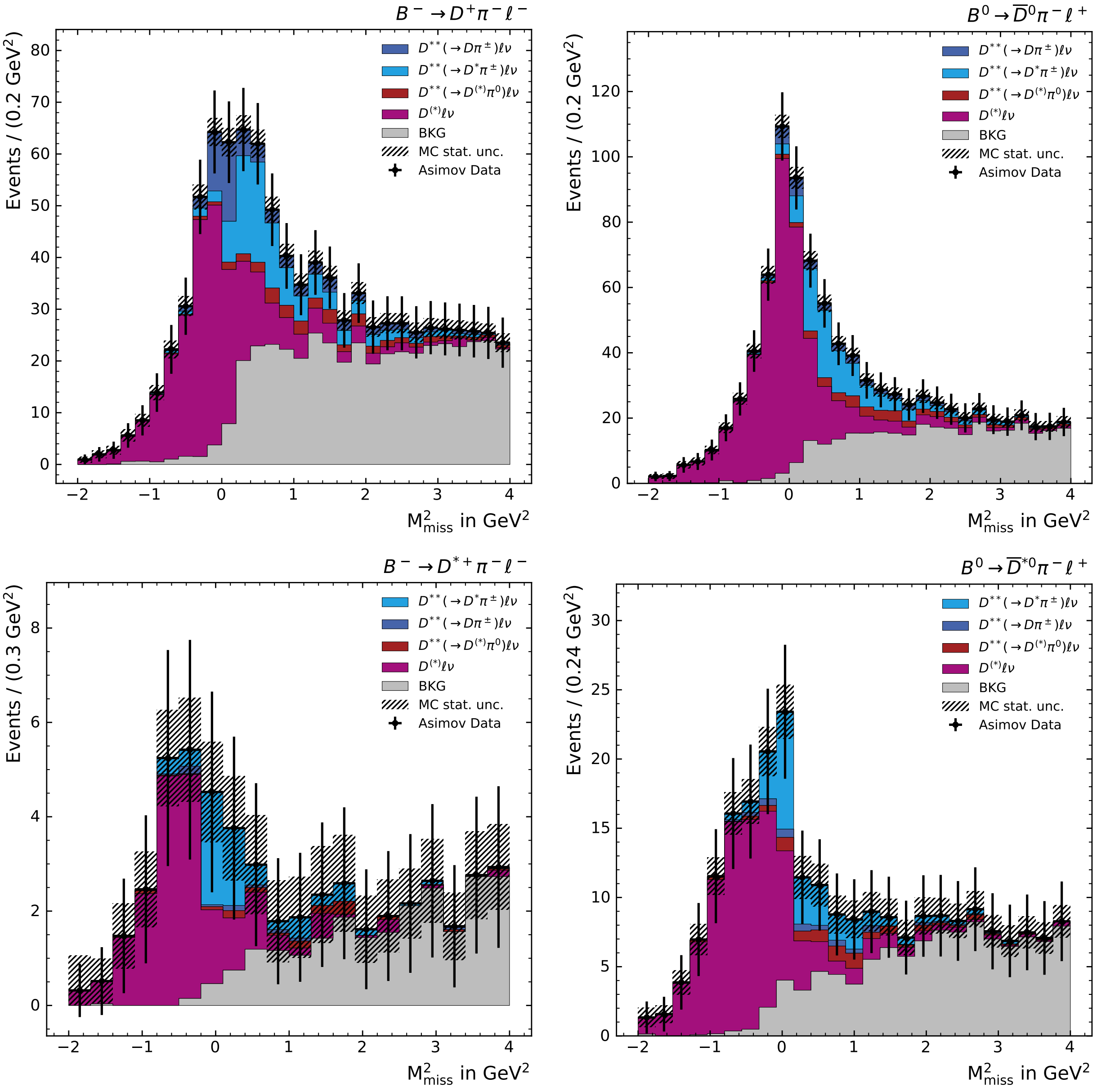


Figure A.9.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $7\,\text{GeV}^2 \leq q^2 < 10\,\text{GeV}^2$.

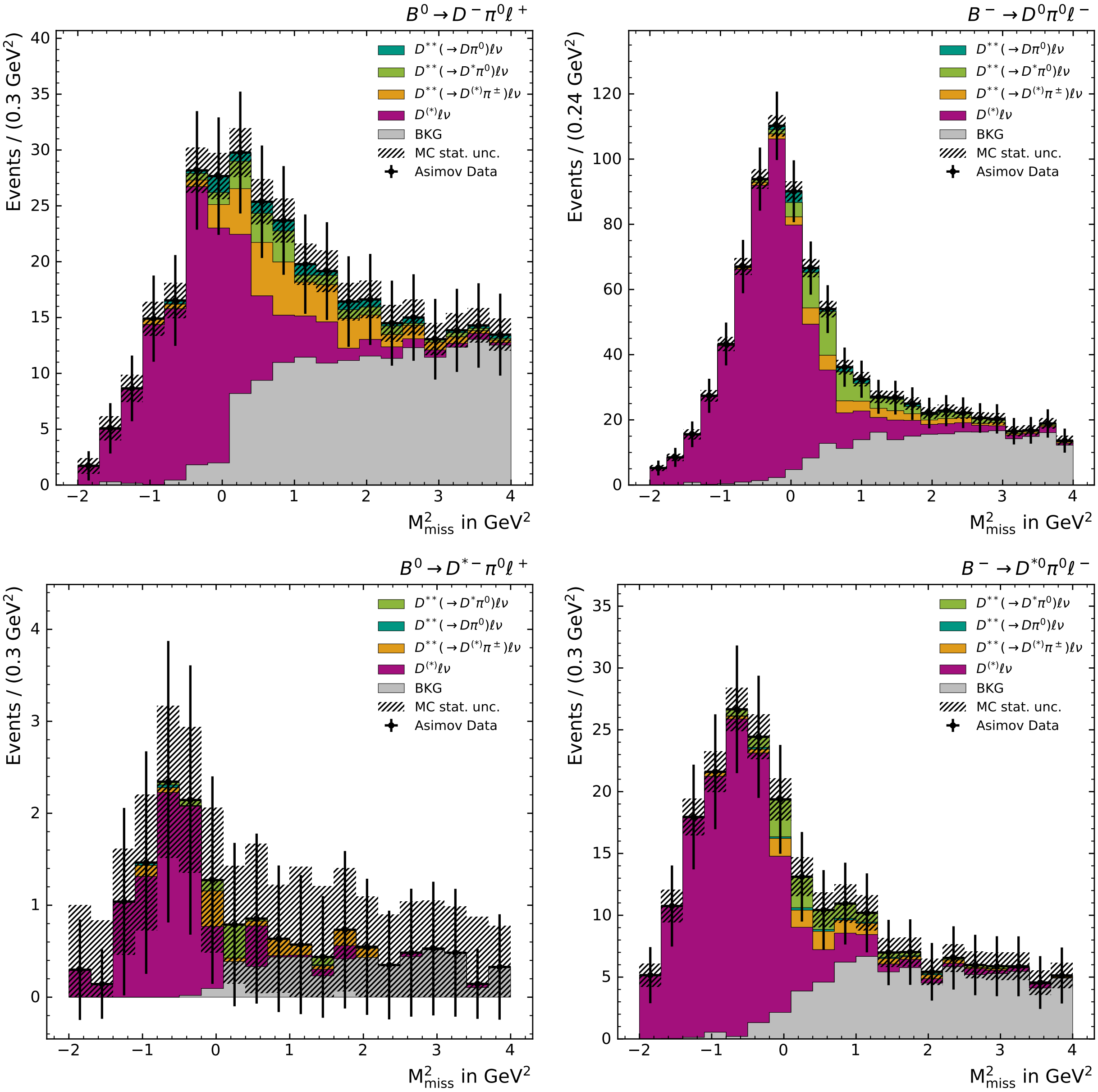


Figure A.10.: Post-fit distributions of the template likelihood fit for the extraction of the $\mathrm{B} \to \mathrm{D}\pi^0\ell\nu$ and $\mathrm{B} \to \mathrm{D}^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $7\,\mathrm{GeV}^2 \leq \mathrm{q}^2 < 10\,\mathrm{GeV}^2$.

B. Additional Fit Results for the $\mathbf{p}_\ell^*$ Form Factor Study

In this appendix all remaining fit results on p_ℓ^*, which are used for the determination of the ratios for the form factor study and are not shown earlier, are collected. All fits are performed on Asimov data.

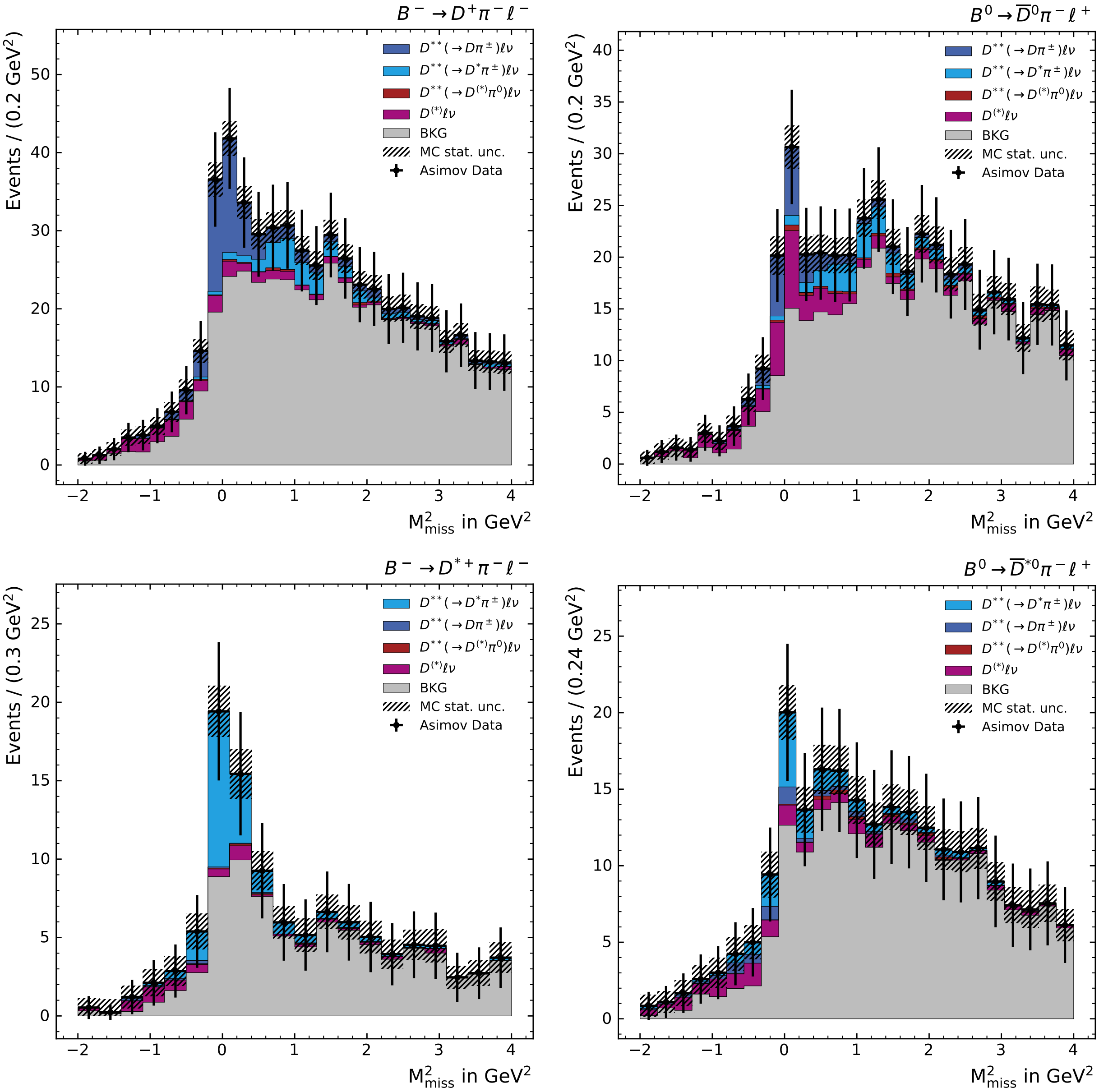


Figure B.1.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $0\,\text{GeV} \leq p_\ell^* < 0.5\,\text{GeV}$.

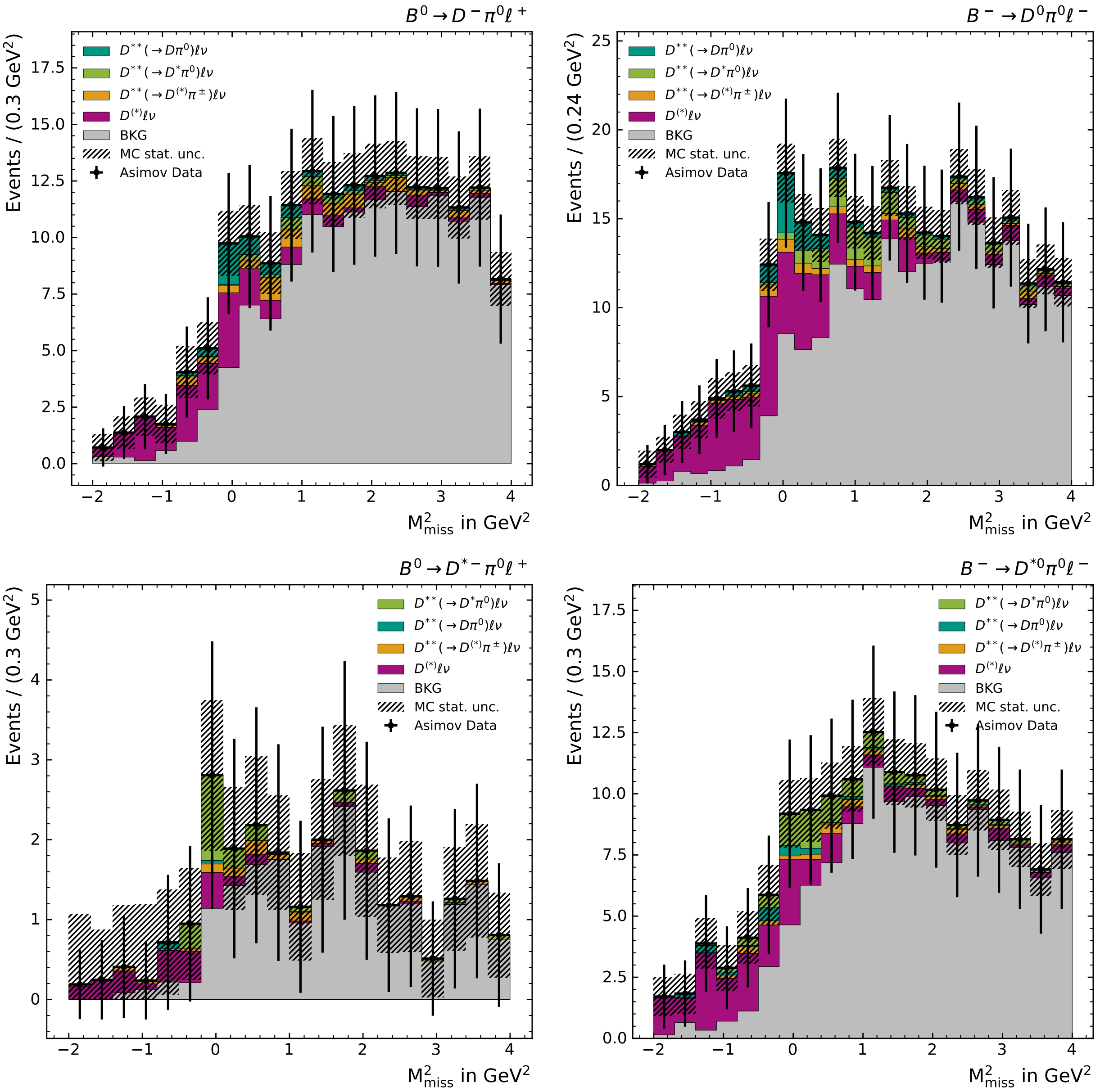


Figure B.2.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^0\ell\nu$ and $B \to D^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $0\,\mathrm{GeV} \leq p_\ell^* < 0.5\,\mathrm{GeV}$.

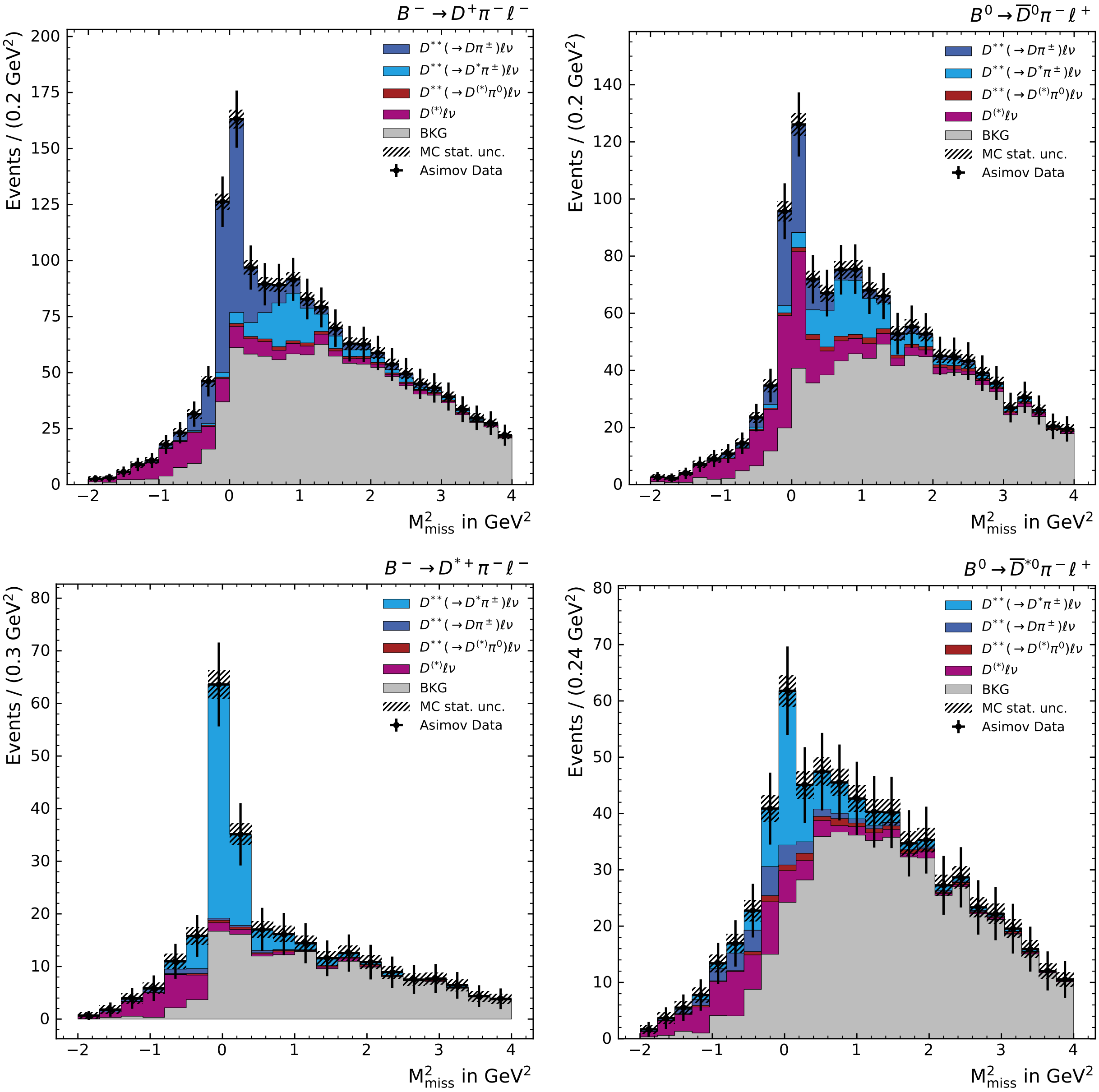


Figure B.3.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $0.5\,\text{GeV} \leq p_\ell^* < 0.8\,\text{GeV}$.

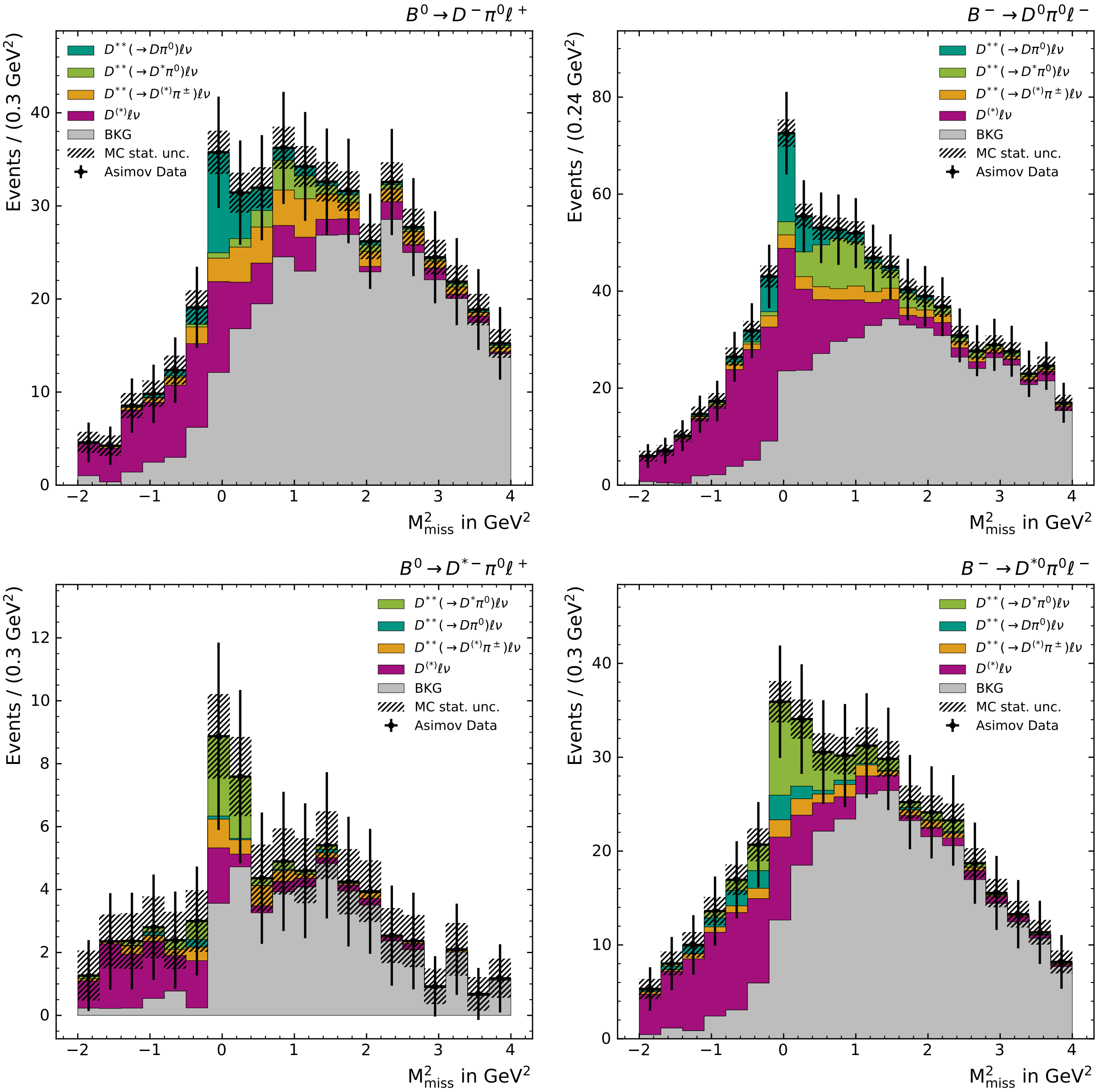


Figure B.4.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^0\ell\nu$ and $B \to D^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $0.5\,\text{GeV} \leq p^*_\ell < 0.8\,\text{GeV}$.

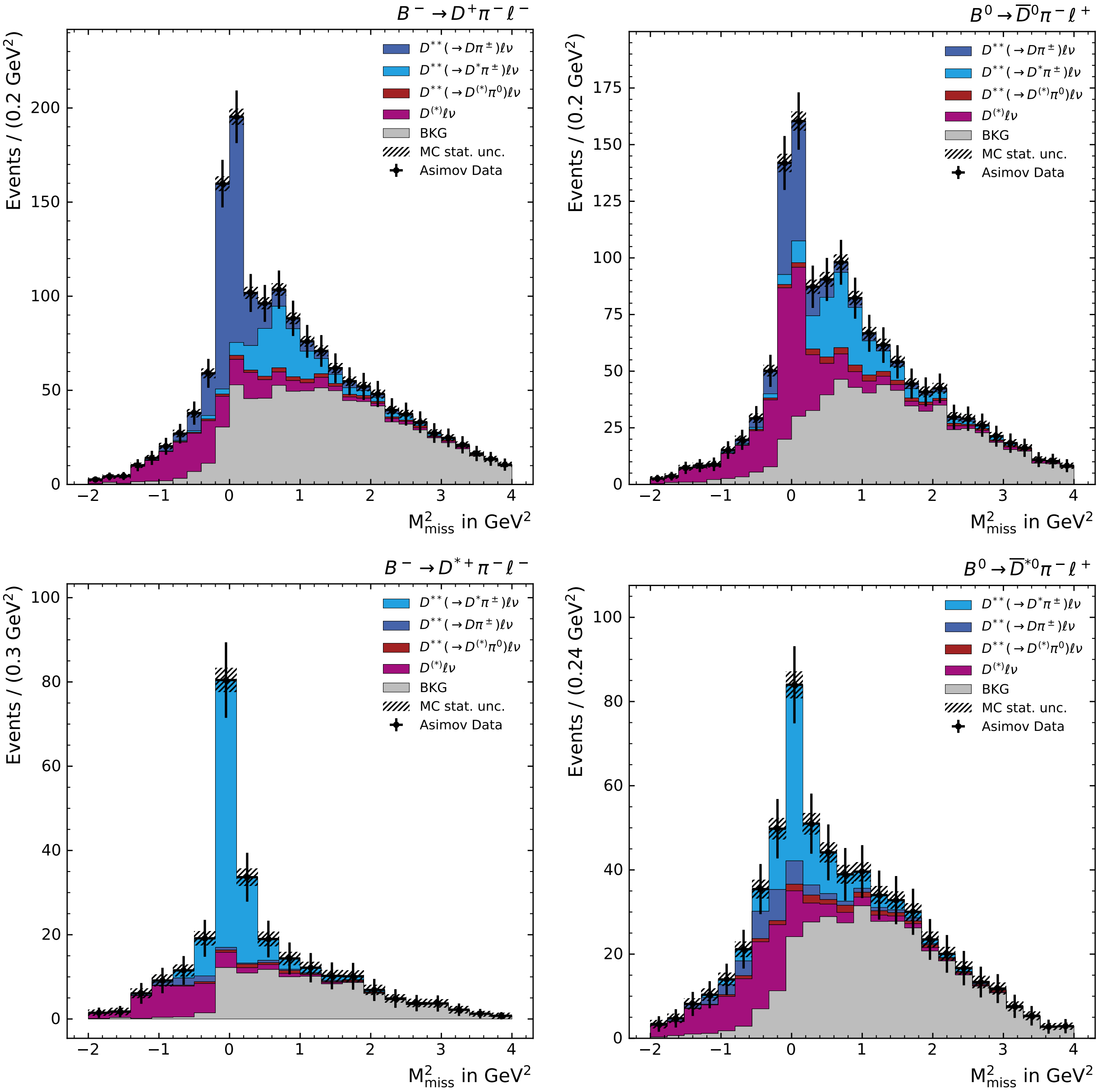


Figure B.5.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $0.8\,\text{GeV} \leq p_\ell^* < 1.0\,\text{GeV}$.

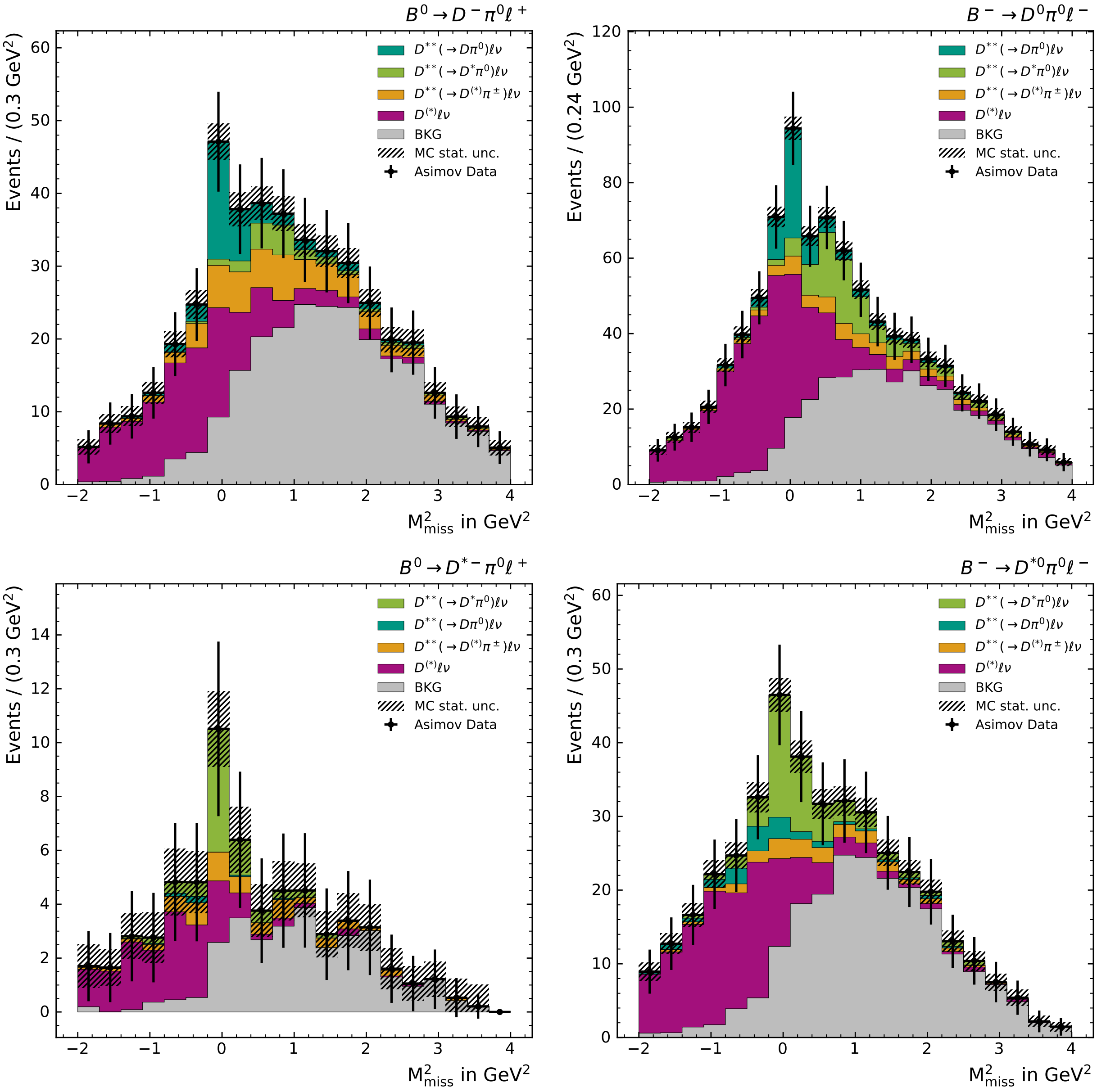


Figure B.6.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^0\ell\nu$ and $B \to D^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $0.8\,\text{GeV} \leq p^*_\ell < 1.0\,\text{GeV}$.

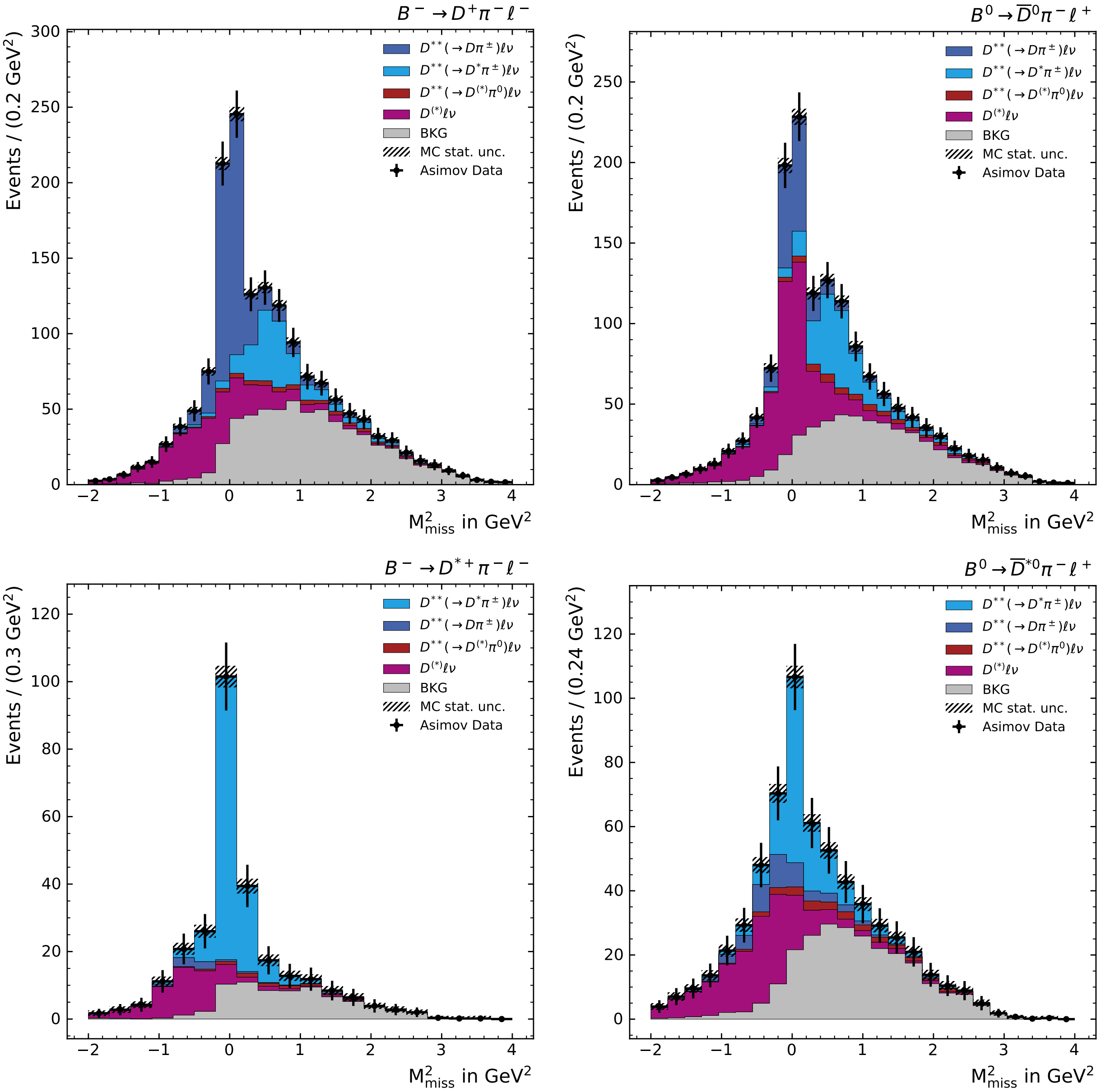


Figure B.7.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1.0\,\mathrm{GeV} \leq p_\ell^* < 1.2\,\mathrm{GeV}$.

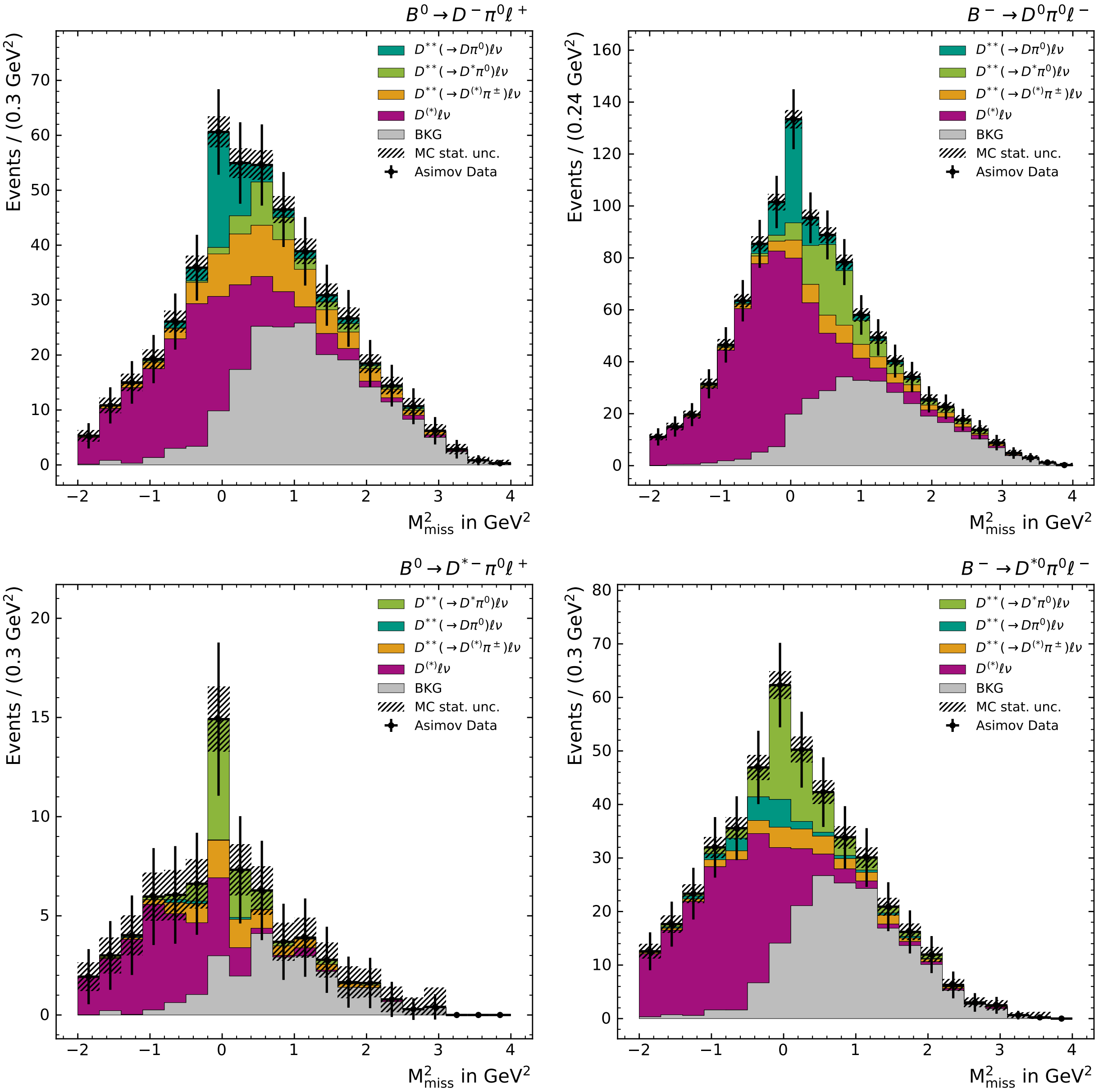


Figure B.8.: Post-fit distributions of the template likelihood fit for the extraction of the $\mathrm{B} \to \mathrm{D}\pi^0\ell\nu$ and $\mathrm{B} \to \mathrm{D}^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1.0\,\mathrm{GeV} \le \mathrm{p}_\ell^* < 1.2\,\mathrm{GeV}$.

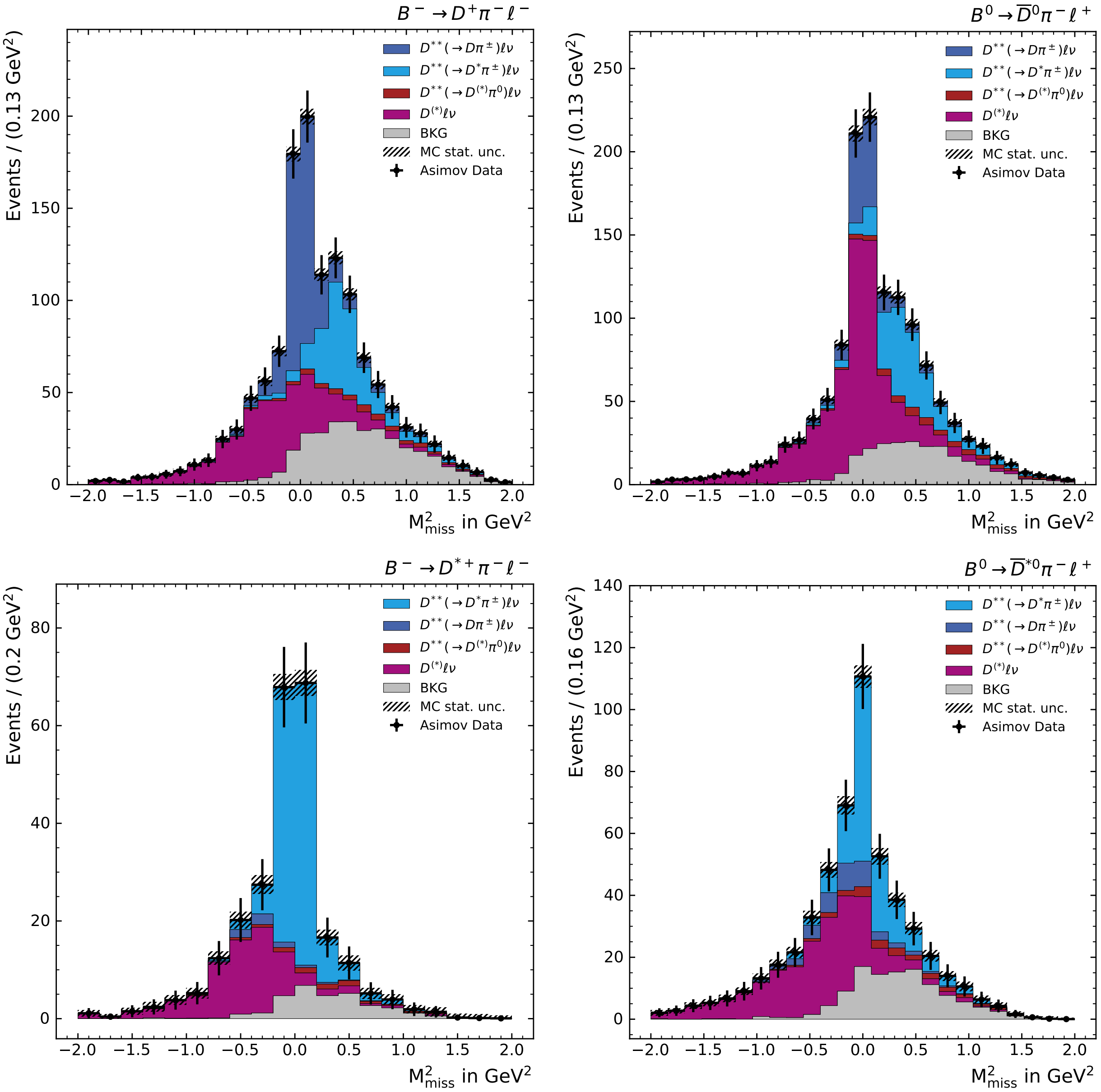


Figure B.9.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1.4\,\text{GeV} \le p_\ell^* < 1.6\,\text{GeV}$.

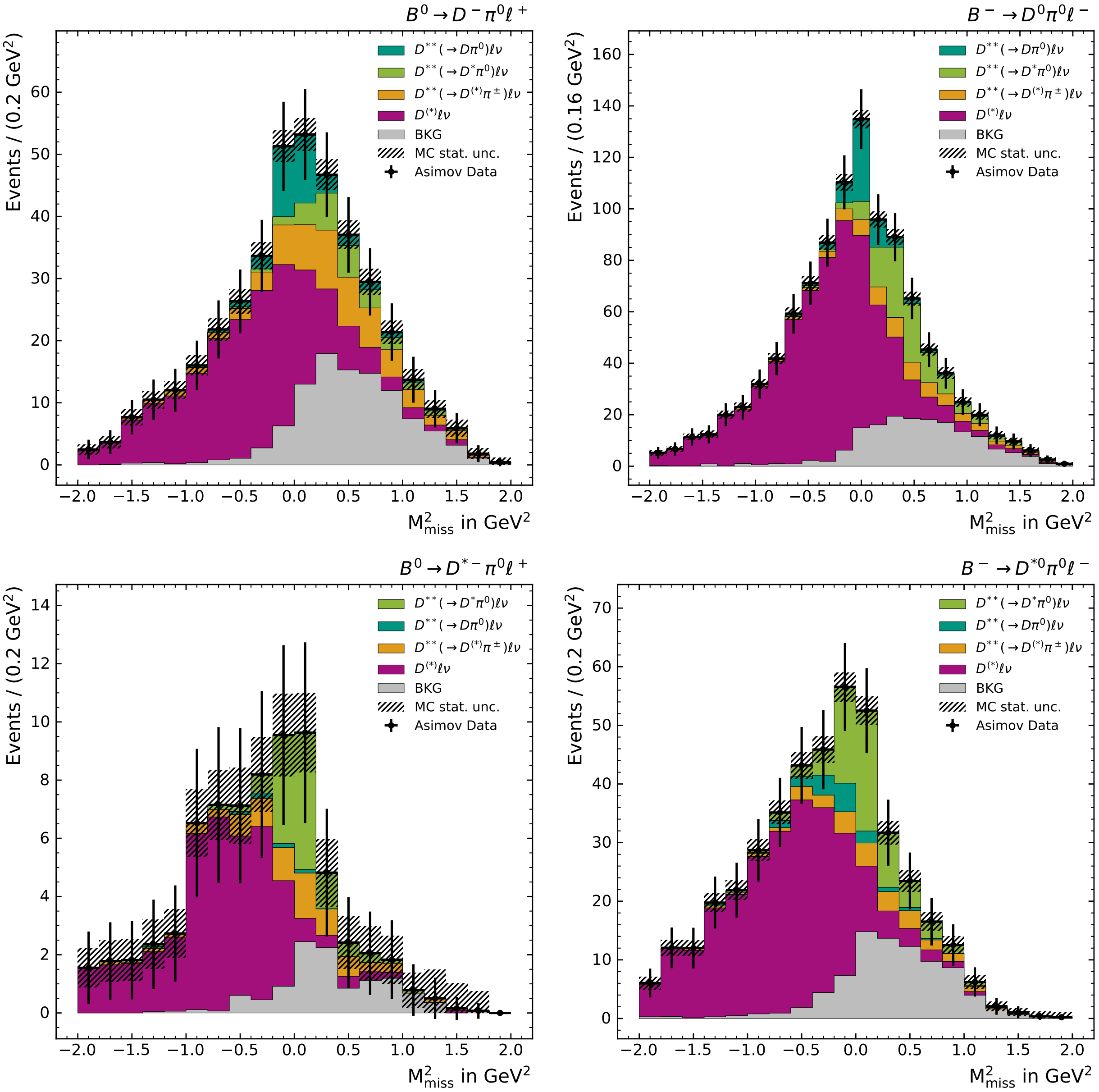


Figure B.10.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^0\ell\nu$ and $B \to D^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1.4\,\mathrm{GeV} \leq p_\ell^* < 1.6\,\mathrm{GeV}$.

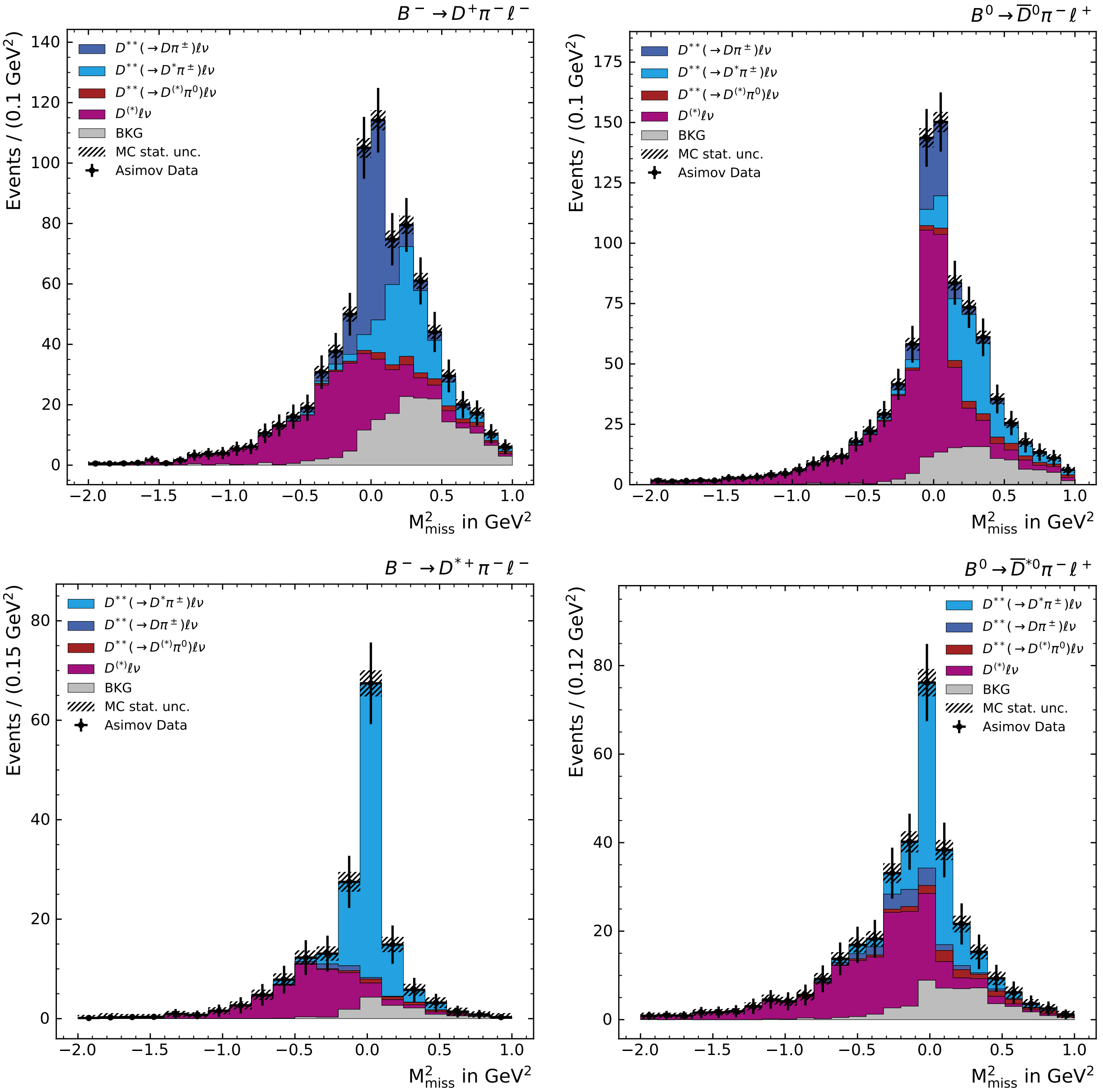


Figure B.11.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1.6\,\text{GeV} \leq p_\ell^* < 1.8\,\text{GeV}$.

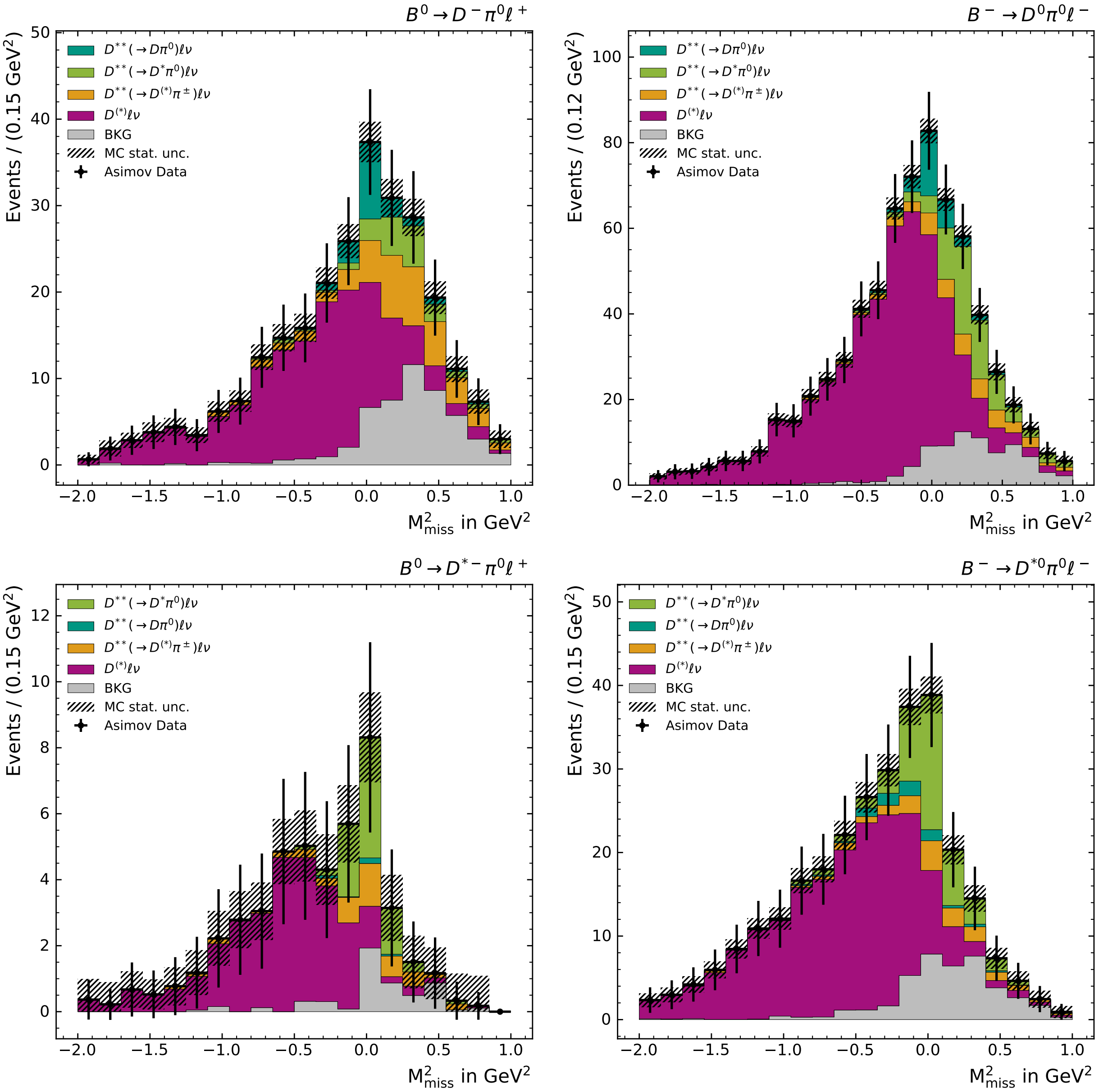


Figure B.12.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^0\ell\nu$ and $B \to D^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1.6\,\text{GeV} \leq p^*_\ell < 1.8\,\text{GeV}$.

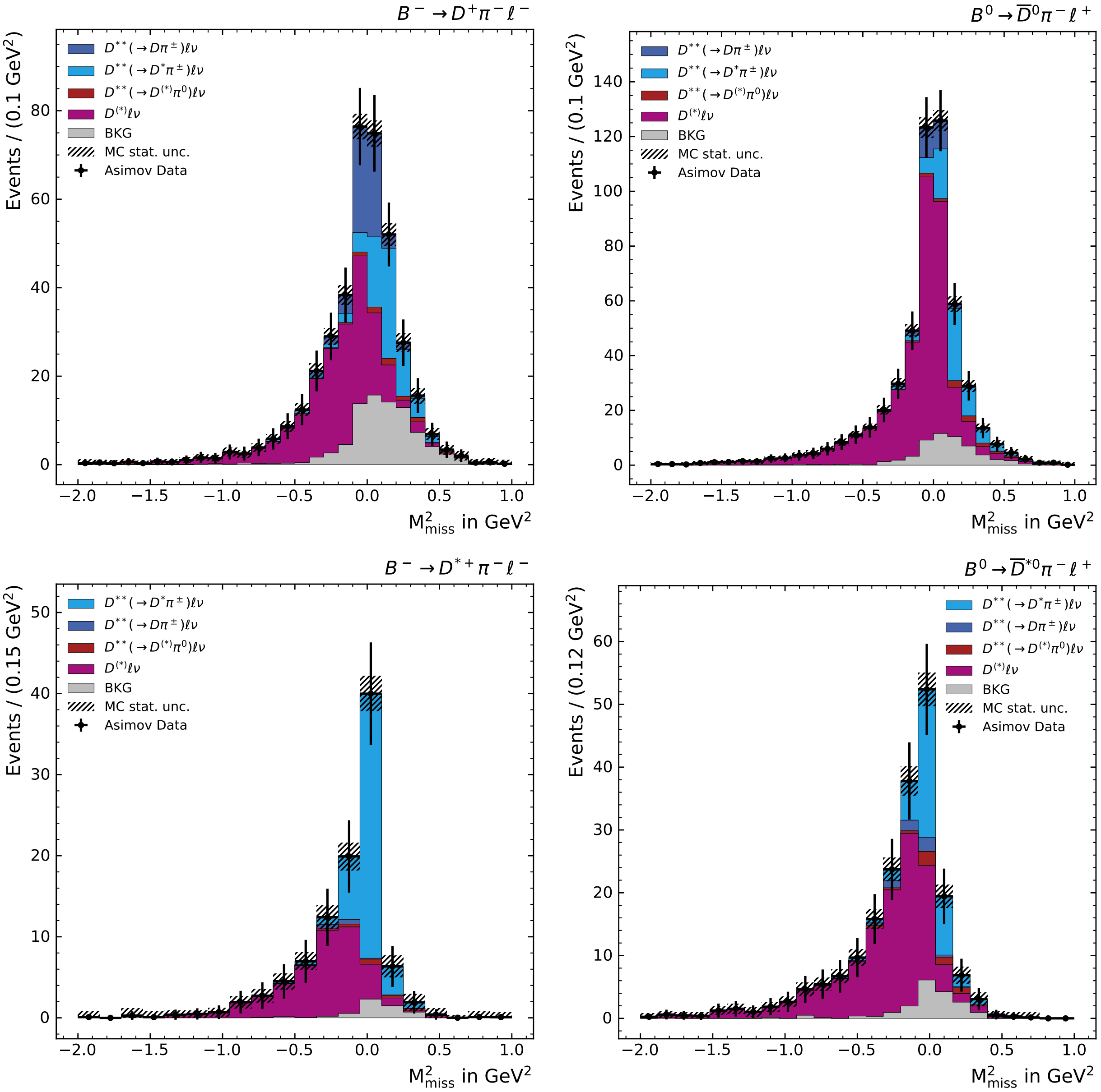


Figure B.13.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^\pm\ell\nu$ and $B \to D^*\pi^\pm\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1.8\,\mathrm{GeV} \leq p_\ell^* < 2.2\,\mathrm{GeV}$.

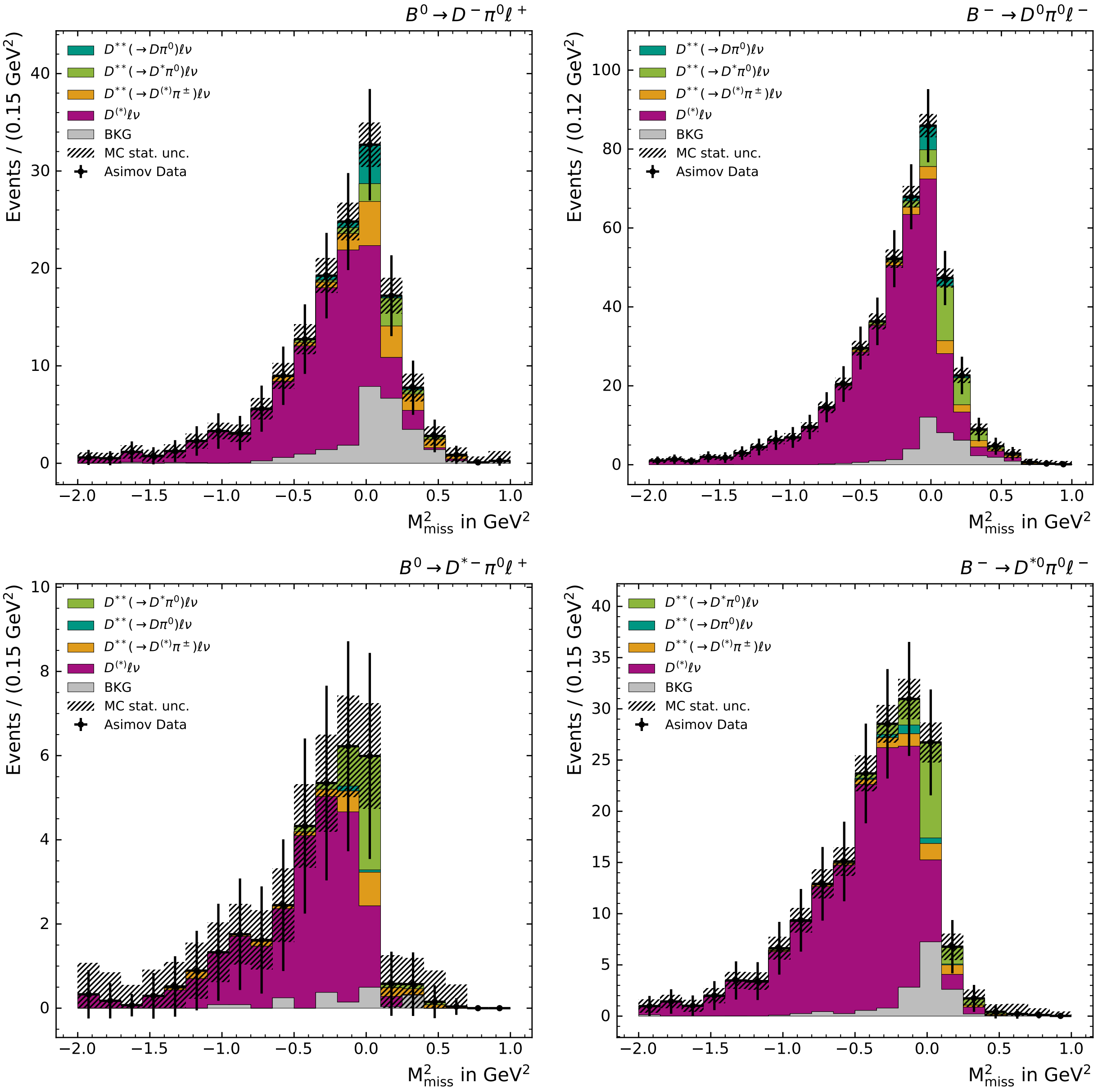


Figure B.14.: Post-fit distributions of the template likelihood fit for the extraction of the $B \to D\pi^0\ell\nu$ and $B \to D^*\pi^0\ell\nu$ signal yields. The fit is performed on the Asimov data set with $1.8\,\text{GeV} \leq p^*_\ell < 2.2\,\text{GeV}$.