Karlsruhe Institute of Technology

ETP-KA/2024-4

Search for $\mathrm{B}^+ \to \mathrm{K}^{*+}\tau^+\tau^-$ with hadronic tagging at the Belle II Experiment

Lennard Damer

Masterthesis

8th October 2024

Institute of Experimental Particle Physics (ETP)

Reviewer:	Prof. Dr. Torben Ferber
Second reviewer:	Dr. Slavomira Stefkova
Advisor:	Dr. Pablo Goldenzweig

Editing time: 10th October 2023 – 8th October 2024

Karlsruhe Institute of Technology

ETP-KA/2024-4

Suche nach $\mathrm{B}^+ \to \mathrm{K}^{*+}\tau^+\tau^-$ mit hadronischem Tagging am Belle II Experiment

Lennard Damer

Masterarbeit

8. Oktober 2024

Institut für Experimentelle Teilchenphysik (ETP)

Referent:	Prof. Dr. Torben Ferber
Korreferent:	Dr. Slavomira Stefkova
Betreuender Mitarbeiter:	Dr. Pablo Goldenzweig

Bearbeitungszeit: 10. Oktober 2023 – 8. Oktober 2024

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, 8. Oktober 2024

. . . . . .

(**Lennard Damer**)

Disclaimer

Data analyses in high-energy physics such as the measurement presented in this master thesis are a collaborative effort. The SuperKEKB particle accelerator which provides the particle beams essential for all studies at Belle II was built and is operated and maintained by the SuperKEKB accelerator group. The Belle II detector was built and is maintained and operated by the Belle II collaboration. The Belle II collaboration also creates the simulated and recorded data sets and maintains the computing infrastructure necessary to process them. The software environment necessary for studies with Belle II data plays an important role and was created and is maintained by the collaboration. I have performed all studies detailed in this thesis except for the determination of calibration factors and the associated uncertainties for:

- Tagging efficiency in Section 5.6.1
- The efficiency of π^0 reconstruction in Section 5.6.2
- The efficiency and misreconstruction rate of charged K and π mesons in Section 5.6.2
- The efficiency of lepton reconstruction in Section 5.6.3

The mentioned studies have been performed by members of the collaboration and are explicitly acknowledged whenever used.

Statement on the Employment of Techniques based on Artificial Intelligence

This thesis incorporates the use of Artificial Intelligence (AI) tools to help with grammatical or stylistic improvement of text, and program code creation:

- Grammarly[1] is utilized throughout the thesis for spell and grammar checks, as well as for paraphrasing individual, selected sentences to improve clarity and precision in academic writing. I have approved all suggested changes.
- GitHub Copilot[2] is used to aid the development of Python code, in particular class structures and string-handling code that does not constitute the core scientific work of this thesis. I have approved and tested all suggestions to provide robust and reliable results. The use of GitHub Copilot has been explicitly acknowledged whenever used.

1 `https://www.grammarly.com/` (Access date: 8th October 2024)
2 `https://github.com/features/copilot` (Access date: 8th October 2024)

Contents

1 Introduction

The Standard Model of particle physics (SM) [1–8] provides the most precise and detailed description in the domain of elementary particle interactions. Even though this theory only combines three of the four fundamental forces, it has found wide acceptance in the particle physics community. The groundbreaking success over the last 50 years stems from the accurate prediction of measurements performed by high-precision experiments. However, despite the lack of a proper description of gravity, it cannot be considered complete. Inadequacies of the SM include the inability of interpreting dark matter, the baryon asymmetry or the presence of neutrino oscillations.

Two approaches have been established for searching for phenomena beyond our current understanding in need to extend the SM: A direct search for new phenomena or high-precision measurements of SM processes to uncover inconsistencies with theoretical predictions.

Flavor physics, the study of elementary fermion interactions, is particularly sensitive to new observations. Especially interactions involving flavor-changing neutral currents (FCNCs) are highly suppressed, leading to observable effects in the presence of new phenomena. Transitions under extensive experimental investigation include $\mathrm{b} \to \mathrm{s}e^+e^-$ and $\mathrm{b} \to \mathrm{s}\mu^+\mu^-$ where several observed anomalies [9–14] prompted interpretations in terms of non-SM particles, which mainly dissolved with the latest measurements [15]. On the other hand, interesting patterns of deviations in the charged current $\mathrm{b} \to \mathrm{c}\ell^-\overline{\nu}_\ell$ have been accumulated. For this, the ratios of $\mathrm{R}(\mathrm{D}^{(*)})$ and $\mathrm{R}(\mathrm{J}/\psi)$ have been extensively studied to determine lepton flavor universality (LFU) violation by comparing the τ mode to the light lepton (e, μ) modes. Even though this charged current is a tree-level process and thus not mainly sensitive to physics beyond the SM, the measured ratios differ from their SM predictions by an average significance of $3.2\,\sigma$ [16].

Combining both hints, a large LFU violation is expected in FCNC involving τ-leptons, namely in $\mathrm{b} \to \mathrm{s}\tau^+\tau^-$. Theoretical models that allow the violation of LFU [17, 18] motivate an increase in the branching fraction by three orders of magnitude. Experiments primarily focusing on the study of B mesons are of particular interest in this regard, given their heavy quark content and experimental accessibility. Since many quark transitions are additionally observable, B mesons are a unique way to probe in this direction. The Large

Hadron Collider beauty (LHCb) experiment, as well as the B factories BaBar and Belle have successfully explored such new phenomena in the past 20 years in $B \to K\tau\tau$ transitions, setting upper limits on the branching fraction. To further advance the research, improved analysis techniques and especially a vastly larger dataset are urgently required. These enhancements are the target of the Belle II experiment, commissioned in Tsukuba, Japan. With improved reconstruction algorithms and an overall targeted dataset 50 times larger than Belle [19], the Belle II experiment provides a unique environment to study $b \to s\ell^+\ell^-$ interactions.

In this work, I will outline a first estimate of the upper limit of $B^+ \to K^{*+}\tau^+\tau^-$ decays at the Belle II experiment. This decay is of particular interest, since it has not been measured by any experiment and thus enables an exclusive viewpoint. For this analysis, I employ an advanced hadronic tagging algorithm and consider additional τ lepton modes, making it distinct from similar searches from predecessor experiments.
In the following chapters of this work, I describe the foundations of the $B^+ \to K^{*+}\tau^+\tau^-$ decay and present an outline of existing measurements (Chapter 2). Next, I introduce the Belle II experiment and focus on key components that contribute to an advanced measurement of $B \to K\tau\tau$ decays (Chapter 3). Furthermore, the necessary tools for this analysis are introduced (Chapter 4). As key part of this work, I define the reconstruction procedure and derive corrections to minimize the discrepancy between recorded and simulated data samples (Chapter 5). The validation of the reconstruction is performed on designated control samples (Chapter 6). With the gained knowledge, I define a signal extraction strategy (Chapter 7) and apply it on the simulated dataset (Chapter 8), delivering the first estimate on the upper limit of $B^+ \to K^{*+}\tau^+\tau^-$ on simulation. To further improve the result, I outline advanced techniques (Chapter 9), making the future of this analysis concise.

2 Foundations

In this chapter, I discuss the theoretical foundations required to fully describe the interaction of interest. I briefly address the mechanism of weak interactions which is fundamental for flavor physics. Furthermore, I introduce the concept of LFU and the effects of new theoretical descriptions describing $b \to s\ell^+\ell^-$ transitions. At the end of this chapter, I summarize existing measurements performed by multiple experiments in this domain.

2.1 Lepton Flavor Universality

In the SM, the electroweak unification grounds on the symmetry-breaking pattern $SU(2) \times U(1)_Y \to U(1)_{EM}$. When introducing a complex doublet H, called the Higgs multiplet, the electroweak Lagrangian reads

$$\mathcal{L} = -\frac{1}{4}(W^a_{\mu\nu})^2 - \frac{1}{4}B^2_{\mu\nu} + (D_\mu H)^\dagger (D_\mu H) + m^2 H^\dagger H - \lambda (H^\dagger H)^2 \,, \tag{2.1}$$

where B_μ is the gauge boson field affiliated with the hypercharge $U(1)_Y$ symmetry and W^a_μ the gauge boson fields associated with SU(2). The actual breaking of symmetry occurs when the Higgs multiplet obtains its vacuum expectation value ν. Expanding the Higgs multiplet around ν in the Lagrangian leads to mass terms associated with the three gauge bosons $W^\pm$ and Z.

When addressing the fermion sector, i.e. the coupling of the electroweak gauge bosons to fermions, it turns out that the SU(2) gauge bosons only couple to left-handed fermions. This is an important property since it causes the theory to be chiral and maximally parity-violating.

Left-handed leptons and quarks are represented by three generations of SU(2) doublet pairs:

$$L^i = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \quad \text{and} \quad Q^i = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}. \tag{2.2}$$

Here, the index i represents the generation. The right-handed fermions are characterized

by SU(2) singlets:

$$e_R^i = \{e_R, \mu_R, \tau_R\} \quad \nu_R^i = \{\nu_{eR}, \nu_{\mu R}, \nu_{\tau R}\}\,,$$
$$u_R^i = \{u_R, c_R, t_R\} \quad d_R^i = \{d_R, s_R, b_R\}\,. \tag{2.3}$$

Focusing on the lepton sector, the interactions between fields and gauge bosons are governed by the two coupling strengths g and g'. The corresponding Lagrangian is

$$\mathcal{L} = i\overline{L}_i(\partial\!\!\!/ - igW\!\!\!/^a T^a - ig' Y_L B\!\!\!/)L_i + i\overline{e}_R^i(\partial\!\!\!/ - ig' Y_e B\!\!\!/)e_R^i$$
$$+ i\overline{\nu}_R^i(\partial\!\!\!/ - ig' Y_\nu B\!\!\!/)\nu_R^i\,, \tag{2.4}$$

here, T^a are the generators of SU(2), Y_L defines the left-handed fields' hypercharge and Y_e, Y_ν the right-handed counterpart. The concept of hypercharge is the symmetry representation of the $\mathrm{U}(1)_Y$ symmetry group of the electroweak theory. The corresponding hypercharges in Equation (2.4) reveal itself to be the same for each generation.
The coupling strengths g and g' are considered universal, meaning they are independent of the lepton field. The gauge interaction has therefore identical couplings to different flavors of leptons. This concept is known as lepton flavor universality (LFU).
The only source of LFU violation in the SM originates from the Yukawa interaction after symmetry breaking. The Higgs doublet is hereby coupling to the fermion fields resulting in massive fermions. Focusing on the quark component, the Yukawa Lagrangian after symmetry breaking is

$$\mathcal{L}_Y = -\frac{\nu}{2} Y_{ij}^d \overline{d}_L^i d_R^j - \frac{\nu}{2} Y_{ij}^u \overline{u}_L^i u_R^j + h.c.\,, \tag{2.5}$$

where Y^d and Y^u are describing the down- or up-type couplings respectively. These two 3×3 matrices contain 18 parameters that regulate the fermion masses. To diagonalize these matrices, a unitary rotation represented by $U_{d,u}$ is performed, yielding the quark mass terms to be

$$\mathcal{L}_{\text{mass}} = -m_j^d \overline{d}_L^j d_R^j - m_j^u \overline{u}_L^j u_R^j + h.c.\,, \tag{2.6}$$

where m_j^d and m_j^u are referring to the quark masses. The kinetic term as well as interactions with Z bosons are invariant under this transformation, since these do not mix up- and down-type quarks. However, the $W_\mu^\pm$ couplings are sensitive to such flavor rotations. The Lagrangian describing the $W_\mu^\pm$ interaction is then given by

$$\mathcal{L}_W \sim \left[W_\mu^+ \overline{u}_L^i \gamma^\mu (V)^{ij} d_L^j + W_\mu^- \overline{d}_L^i \gamma^\mu (V^\dagger)^{ij} u_L^j\right]\,. \tag{2.7}$$

Here $V = U_u^\dagger U_d$ describes the flavor mixing effects and is generally known as the Cabibbo-

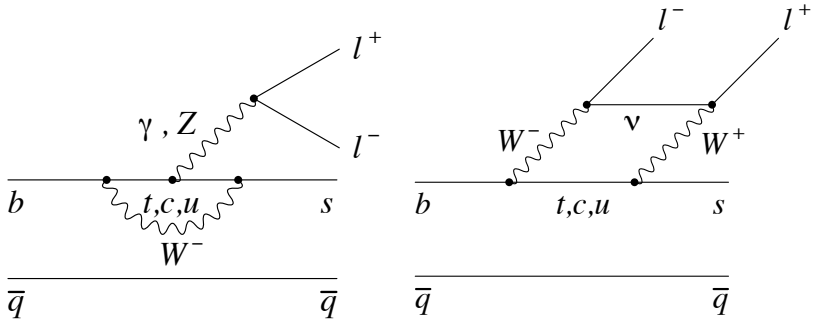


Figure 2.1: Leading order Feynman diagrams for the FCNC transition $b \to s\ell^+\ell^-$ showing the electroweak penguin (left) and box diagram contributions (right). Taken from [21].

Kobayashi-Maskawa (CKM) matrix [20].
Compared to Equation (2.4), the Yukawa interaction leads to additional terms sensitive to lepton flavor and thus LFU is in general not sustained.

2.2 Enhancement of $b \to s\tau\tau$ Transitions

A special probe for testing LFU violation is provided by FCNCs, describing flavor-changing transitions among fermions with the same electric charge. Given the unitarity of U_u and U_d as well as the CKM matrix, such transitions are heavily suppressed in the SM. This is often referred to as Glashow-Iliopoulos-Maiani (GIM) mechanism [4], historically underpinning the existence of the charm quark to fully understand the suppression of $s \to d\mu\mu$. Even though FCNC processes are forbidden at tree level, they occur in higher order diagrams.
Rare b to s quark transitions are often referred to when considering the test of LFU. The corresponding Feynman diagram of such a FCNC is depicted in Figure 2.1. First processes of this type have been observed by the Belle collaboration [22] and have been studied extensively ever since. Especially the comparison of $b \to se^+e^-$ and $b \to s\mu^+\mu^-$ in the ratios of $R(K^{(*)})$ [10, 23] have shown tensions to SM predictions. This deviation was entirely resolved by the latest LHCb measurement [15], however a binned analysis of angular components is still expected to demonstrate tensions to the SM predictions [24]. Together with the excess in $R(D^{(*)})$ and $R(J/\psi)$ concerning τ-lepton modes [16], the FCNC process $b \to s\tau^+\tau^-$ is expected to be a promising candidate to further test the SM predictions.
The branching fraction of the corresponding process $B \to K\tau^+\tau^-$ is calculated by employing unquenched lattice QCD, allowing for the first model-independent SM prediction [25]:

$$\mathcal{B}(B \to K\tau^+\tau^-) = (1.41 \pm 0.15) \times 10^{-7} \tag{2.8}$$

Given the magnitude of this result, a precise measurement of this signal with existing datasets is almost impossible since it drastically exceeds the sensitivity of current flavor experiments [19]. However, new theoretical models allow for a significant increase of the branching fraction by up to three orders of magnitude [17, 18].
A theoretical description is not trivial since the flavor-changing amplitudes involve widely separated mass scales. While perturbation theory is a solid method for calculating QCD effects associated with the scales of the bottom quark mass $m_b \approx 4.5\,\mathrm{GeV}/c^2$ and above, it breaks down at the energy scale $\Lambda_{\mathrm{QCD}} \approx 350\,\mathrm{GeV}$. A powerful tool for resolving this issue is the effective weak Hamiltonian H_{eff}. I define H_{eff} for the $\mathrm{b} \to \mathrm{s}\tau^+\tau^-$ transitions in the following way:

$$H_{\mathrm{eff}}(\mathrm{b} \to \mathrm{s}\tau^+\tau^-) = -\frac{4G_F}{\sqrt{2}}(V)^{tb}(V^*)^{ts}\sum_a C_a O_a\,. \tag{2.9}$$

Here, G_F denotes the Fermi constant and C_a the coupling coefficients of the effective operators O_a, also known as Wilson coefficients. The branching factions are sufficiently described by $C_{9,9'}$ and $C_{10,10'}$, hence the focus is set on the relevant operators

$$O^{\tau\tau}_{9(10)} = \frac{\alpha}{4\pi}\left[\overline{s}\gamma^\mu P_L b\right]\left[\overline{\tau}\gamma_\mu(\gamma^5)\tau\right] \tag{2.10}$$

$$O^{\tau\tau}_{9'(10')} = \frac{\alpha}{4\pi}\left[\overline{s}\gamma^\mu P_R b\right]\left[\overline{\tau}\gamma_\mu(\gamma^5)\tau\right] \tag{2.11}$$

where $P_{L,R} = (1 \mp \gamma_5)/2$ denotes the projector mapping onto left- and right-handed field components. In the SM, the corresponding Wilson coefficients are given by $C_9^{\mathrm{SM}} \approx 4.1$ $C_{10}^{\mathrm{SM}} \approx -4.3$ at the scale of 4.8 GeV [26]. Besides the first prediction of $\mathrm{B} \to \mathrm{K}\tau^+\tau^-$ modes in Equation (2.8), the computation of $\mathrm{B}^+ \to \mathrm{K}^{*+}\tau^+\tau^-$ is now possible. Taking the average of the charged and neutral $\mathrm{K}^*\tau^+\tau^-$ modes yields [17]

$$\mathcal{B}(\mathrm{B} \to \mathrm{K}^*\tau^+\tau^-) = (0.98 \pm 0.10) \times 10^{-7}\,. \tag{2.12}$$

Additions to the coefficients $C_{9,9'}$ and $C_{10,10'}$ are used to extend the SM prediction by accounting for New Physics (NP) effects. Interestingly, the NP effects for $\mathrm{b} \to \mathrm{s}\tau^+\tau^-$ are correlated to NP contributions that provide a solution to the $\sim 3.2\,\sigma$ anomaly in $\mathrm{R}(\mathrm{D}^{(*)})$ and $\mathrm{R}(\mathrm{J}/\psi)$: The deviations in $\mathrm{b} \to \mathrm{c}\tau^-\overline{\nu}_\tau$ decays originate in NP contributions from a scale significantly larger than the electroweak scale. This effectively suppresses $\mathrm{b} \to \mathrm{s}\tau^-\overline{\nu}_\tau$ transitions, while enhancing $\mathrm{b} \to \mathrm{s}\tau^+\tau^-$ interactions based on these assumptions. Given this case, the predictions on the branching fractions are almost entirely dominated by NP effects [17] and result in

$$\mathcal{B}(\mathrm{B} \to \mathrm{K}^*\tau^+\tau^-) \approx 0.008\left(\sqrt{\frac{\mathrm{R}(X)}{\mathrm{R}_{\mathrm{SM}}(X)} - 1}\right)^2\,. \tag{2.13}$$

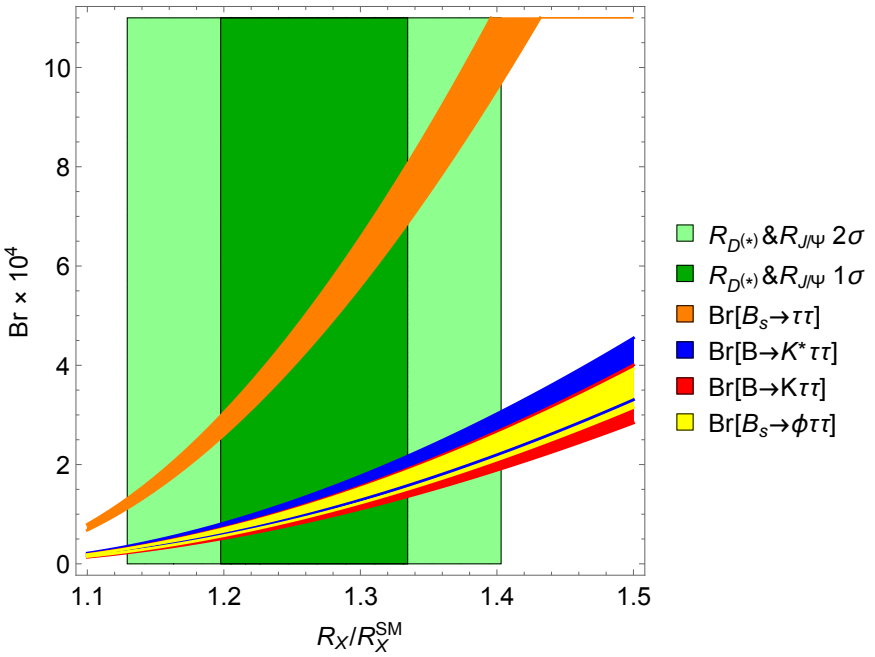


Figure 2.2: Representation of the branching ratios for multiple $b \to s\tau^+\tau^-$ transitions as a function of $R(X)/R_{SM}(X)$. The blue ribbon corresponds to the calculated branching fraction of $B \to K^*\tau^+\tau^-$ transitions, yielding an increment of 10^3 within the current tensions measured in $R(D^{(*)})$ and $R(J/\psi)$. The uncertainty on the calculation is hereby given by the width of the ribbon. Taken from [17].

As mentioned, the NP effects are correlated, making the branching fraction dependent on the tension of the measured ratios with the SM predictions. The corresponding enhancements yield an increasement of three orders of magnitude for the branching fractions as illustrated in Figure 2.2.
The resulting branching fraction of $\mathcal{B}(B^+ \to K^{*+}\tau^+\tau^-) \approx 10^{-4}$ provides a well-motivated approach in searches for violations of LFU, right at the experimental frontier of B factories such as the Belle II experiment.

2.3 Current Experimental Status

The search for $b \to s\tau^+\tau^-$ transitions have been performed by B factories in recent years. In principle, B factories are electron-positron colliders operated at the $\Upsilon(4S)$ resonance energy. The resulting $\Upsilon(4S)$ almost exclusively decays into a $B\overline{B}$ pair, making it an outstanding environment to study b transitions.
The BaBar collaboration closely studied $B^+ \to K^+\tau^+\tau^-$ transitions with the full BaBar dataset corresponding to $424\,fb^{-1}$ [21]. To set constraints on the signal B candidate, a

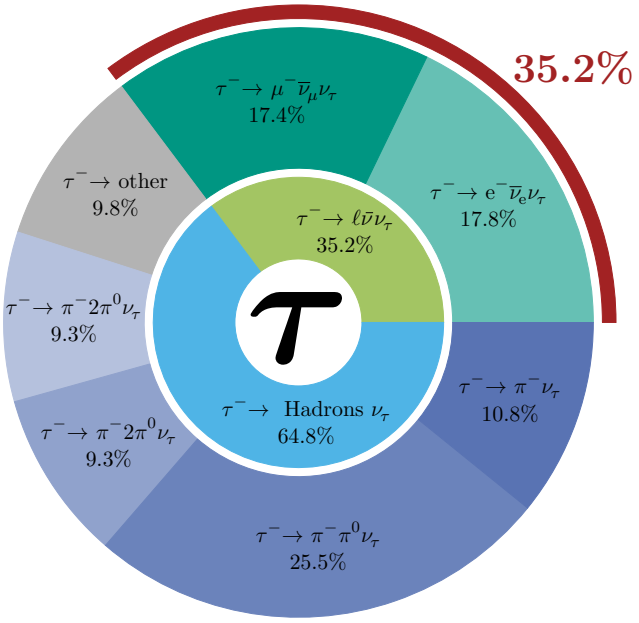


Figure 2.3: Illustration of the branching fractions of the most common τ decay processes. The two shades of green in the outer circle represent $\tau \to \ell \overline{\nu}_\ell$ modes, with their combined contribution defined by the green inner circle. The shades of blue and the other arbitrary τ decay modes are referring to decays involving hadrons. The combined contribution to the overall branching fraction is represented by the light blue fraction of the inner circle. Taken from [27].

hadronic tagging[1] algorithm was employed to increase the selection purity. By this, the partner B meson originating from the $\Upsilon(4S)$ decay is reconstructed in a variety of hadronic decay modes. On the other hand, the τ-leptons exhibit a rapid decay once they have been produced, given their short lifetime [8]. For reconstruction, the BaBar analysis solely reconstructed $\tau \to \ell \overline{\nu}_\ell$ modes with light leptons ($\ell = e, \mu$), accounting for 35.2 % of all τ decays as illustrated in Figure 2.3.

The search did not find any evidence for signal, but was able to set an upper limit of $\mathcal{B}(B^+ \to K^+ \tau^+ \tau^-) < 2.25 \times 10^{-3}$ @ 90% C.L. [21]. Compared to the predicted branching fraction, this result neither supports nor refutes the presence of NP effects. The most recent search for $b \to s \tau^+ \tau^-$ was conducted by the Belle collaboration in $B^0 \to K^{*0} \tau^- \tau^+$ decays at the KEK research facility in Japan.

In contrast to the BaBar experiment, Belle accumulated over 711 fb^{-1} of data at the $\Upsilon(4S)$ resonance. The τ-leptons are additionally reconstructed in the $\tau \to \pi \nu_\tau$ mode, covering a larger phase space of decay modes. Concerning tagging, the Belle search used the Full-Reconstruction (FR) algorithm [28] to reconstruct hadronic modes on the tag-side. Also

1 The concept of tagging and the variety of tagging methods will be introduced in more detail in Chapter 4.

in this search, no evidence of signal was found. However, the upper limit resulting in $\mathcal{B}(\mathrm{B}^0 \to \mathrm{K}^{*0}\tau^-\tau^+) < 3.1 \times 10^{-3}$ @ 90% C.L. [29] provides a good probe in the direction of $\mathrm{b} \to \mathrm{s}\ell\ell$ transitions.

3 The Belle II Experiment

This chapter is dedicated to the experimental setup of the Belle II experiment and is intended to give an overview of the fundamental concepts. I introduce the core ideas of physics at B factories before briefly describing the SuperKEKB accelerator and the Belle II detector.

3.1 Physics of B Mesons and Υ Resonances

A bound state consisting of a $b\bar{b}$ is formally described by heavy quark expansion [30] in a non-relativistic manner.
When operating an electron-positron collider far below the Z boson mass, the interaction $e^+e^- \rightarrow b\bar{b}$ is mediated by a virtual photon. In general, a variety of bound $b\bar{b}$ states corresponding to different configurations of the orbital momentum (L = S, P, D, ...) or quark spins (total spin $S = 0, 1$) are expected. However, the amount of $b\bar{b}$ pairs are limited given that the virtual photons are characterized by their quantum numbers $J^{PC} = 1^{--}$. This reduces the allowed pair states that occur in these interactions to the ones with ${}^{2S+1}L_J = {}^3S_1$ [31]. Therefore, the Υ states present in the hadronic production correspond to radial excitations of the triplet state S, named Υ(nS) where n denotes the specific excitation state. Multiple Υ resonances have been measured at energies between 9.4 and 11 GeV, illustrated in Figure 3.1. Strong interactions are suppressed below the production threshold of a $B\overline{B}$ meson pair. Therefore, electromagnetic interactions are becoming dominant, leading to radiative transitions among various excitations in the $b\bar{b}$ system. Only for energies above the threshold do strong forces become more competitive, making the decay to two B mesons the dominating decay mode for Υ(4S) with a branching fraction of $\sim 96\%$ [8]. At this resonance, the $B\overline{B}$ production threshold is just slightly exceeded by 20 MeV resulting in the B meson pair to be produced almost at rest.
In general, B mesons provide an excellent environment to study physics cases ranging from CP violation to rare decay searches. Therefore some experiments are specialized to operate at or near the Υ(4S) resonance, producing a large amount of B mesons. These experiments are usually referred to as "B factories". Examples for these kinds of experiments are the

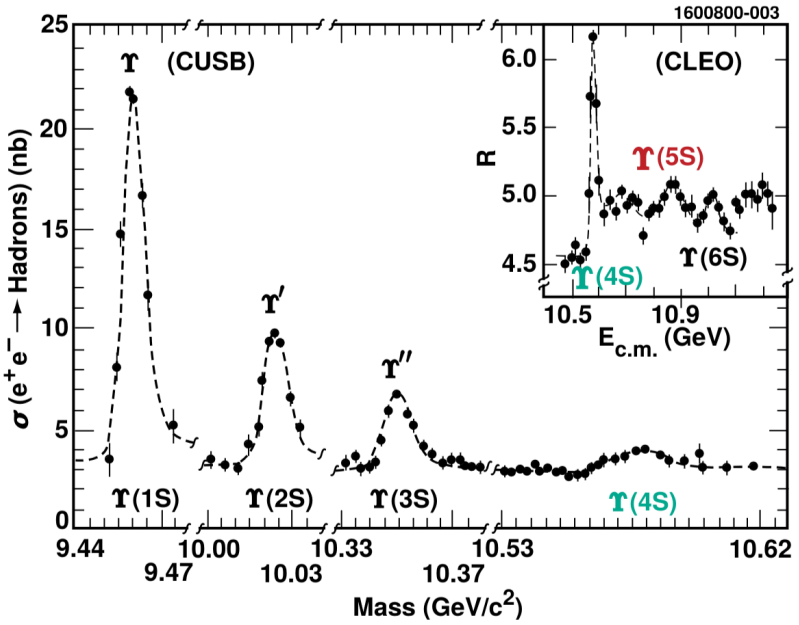


Figure 3.1: Representation of Υ(1S), Υ(2S), Υ(3S) and Υ(4S) resonances in the hadronic production cross section from e^+e^- annihilation as measured by the Columbia University-Stony Brooks (CUSB) detector [32, 33]. The insert figure depicts results of the CLEO collaboration concerning higher excitations, namely Υ(5S) and Υ(6S) [34]. This plot was initially published in Ref. [31]. This figure is taken from [35].

BaBar, Belle and the Belle II experiment. Although providing an exceptional environment, B factories are not the only type of experiments conducting investigations on B mesons.

3.2 The SuperKEKB Accelerator

In order to operate a B factory, an electron-positron accelerator is required that must meet certain criteria. For rare decays such as $B^+ \to K^{*+}\tau^+\tau^-$, a large amount of B meson interactions are needed to provide sufficient statistics. A second major component is asymmetry in beam energies. B mesons from Υ(4S) resonance are produced nearly at rest in the center-of-mass system (CMS) and decay rather quickly after traversing for $\sim 10\,\mu m$. To exceed the distance and better resolve the B decay vertices, asymmetric colliders are used to boost the $B\overline{B}$ system.

Both goals are fulfilled by the SuperKEKB accelerator (shown in Figure 3.2) at the KEK facility in Tsukuba, Japan. SuperKEKB mainly consists of two storage rings, one Low-Energy Ring (LER) for 4 GeV positrons and one High-Energy Ring (HER) for 7 GeV electrons. The two beams collide at one point along the accelerator, exactly where the Belle II detector is located.

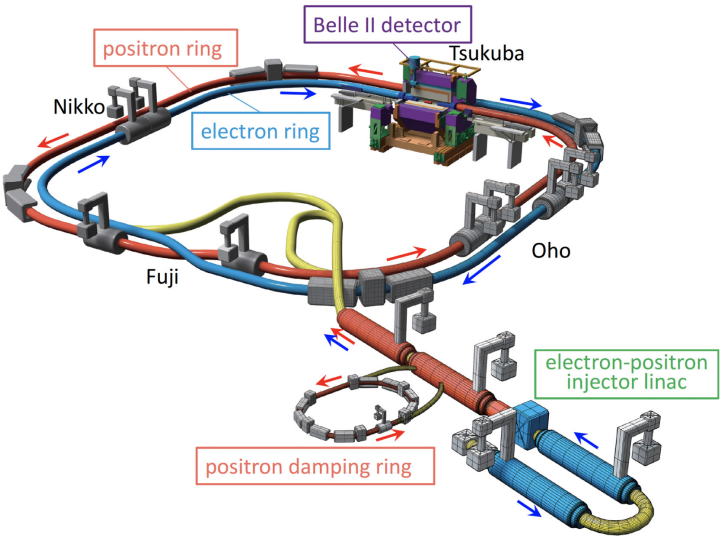


Figure 3.2: Schematic illustration of the SuperKEKB collider at the High Energy Accelerator Research Organization. The straight sections are labeled "Tsukuba", "Oho", "Fuji" and "Nikko". The Belle II detector is located in the Tsukuba area. Taken from [36].

Compared to its predecessor KEKB, the SuperKEKB accelerator plans to achieve a target luminosity 30 times higher [37]. Essential for this raise is the reduction of the beam size at the interaction point (IP) and the increase of the beam current by a factor of two. As of June 2022, SuperKEKB holds the world record for peak instantaneous luminosity with $4.71 \times 10^{34}/(\mathrm{cm}^2\,\mathrm{s})$ [38]. Concerning the boost, the increase in luminosity motivates the decrease of beam asymmetry to a Lorentz boost of $\beta\gamma = 0.28$. Given the energy in the LER, lower emittance growth due to intrabeam scattering is expected compared to KEKB [39].

3.3 The Belle II Detector

The tool to investigate collisions at the SuperKEKB accelerator is the Belle II detector. It consists of a cylindrical structure around the IP that can be subdivided into three parts: A tube-shaped component parallel to the beam axis ("barrel") and two disks positioned perpendicular and centric to the beam ("endcaps"). Regarding its predecessor Belle, the Belle II detector provides a substantial upgrade in every subdetector system. New vertex and particle identification detectors in combination with the improved main tracking detector and new electronics in the calorimeter showcase a better vertex, better track reconstruction, and an improved energy resolution. In addition, the separation between charged hadrons and suppression of particle misidentification is gradually improved.
Altogether, the Belle II detector consists of seven subdetectors as illustrated in Figure 3.3. The three innermost detectors are aiming to reconstruct the particles' trajectories ("tracks")

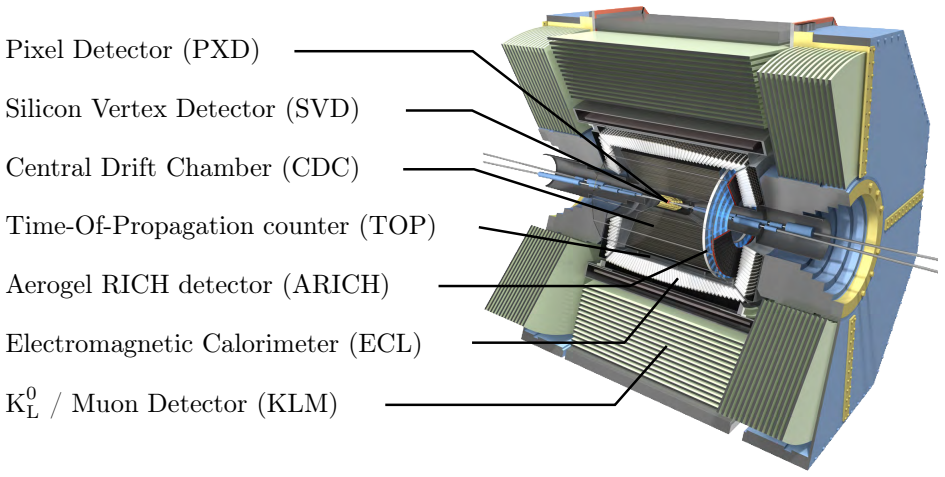


Figure 3.3: Schematic view on the Belle II detector with all subdetector systems labeled. Taken from [35].

that charged particles exhibit after collision. A superconducting solenoid magnet installed between the calorimeter and K_L^0 and muon detector produces a magnetic field of 1.5 T that bends the charged particle trajectories.
In the following, I briefly describe the different subdetectors at the Belle II experiment. An overall deeper and more technical description is available e.g. in Ref [40].

Pixel Detector The Pixel Detector (PXD) represents the first active detector layer after particles emerge from the IP. The PXD is constructed from pixelated sensors of the Depleted Field Effect Transistor (DEPFET) type [41] which are arranged in two layers at a radius of $r = 14\,\text{mm}$ and $r = 22\,\text{mm}$ around the IP [42]. The DEPFET technology allows to position the readout electronics outside of the detector itself, allowing for layer thicknesses of 75 µm. During the data taking period relevant for this work, only one half of the originally proposed 40 DEPFET modules participated in recording data [42]. However, after the first long shutdown of Belle II, a new detector with the two complete layers has been installed, expanding the capability of the detector to its full potential.

Silicon Vertex Detector Together with the Pixel Detector (PXD), the Silicon Vertex Detector (SVD) generates the supplementary component of the vertex detector [43]. Four layers equipped with double-sided silicon strip sensors build the outer layer of the vertex component.
The key objectives of the SVD are the extrapolation of trajectories from the drift chamber inward to the PXD, particle identification using the energy loss rate dE/dx and precise vertexing of K_S^0. In addition, the SVD is the only tracking component to reconstruct trajectories of particles that do not reach the near central drift chamber.

Central Drift Chamber The main detector for tracking is the Central Drift Chamber (CDC), a large-volume drift chamber with small drift cells [44]. Compared to Belle, the CDC extends to a larger radius ($r = 1130\,\text{mm}$ in contrast to $r = 880\,\text{mm}$), while covering the same polar angle $17^\circ < \theta < 150^\circ$. Furthermore, the readout electronics have been completely renewed to handle the higher trigger rate of 30 kHz with less dead time at the target luminosity. The chamber is filled with a gas mixture of 50% helium and 50% ethane, creating a low-density environment. This is especially useful since it suppresses multiple scattering, a dominant effect present at the expected charged particle momenta of 1 to $2\,\text{GeV}/c$. In total, the CDC consists of 14 336 sense wires arranged in 56 layers along the z axis. The wires are orientated either along the beam axis ("axial") or slightly skewed ("stereo"), enabling the reconstruction of three-dimensional helix trajectories. Moreover, the CDC contributes to particle identification by performing measurements of dE/dx of the charged particle.

Particle Identification Detectors Belle II consists of two subdetector systems exclusively targeting particle identification. The Aerogel Ring-Imagining Cherenkov detector (ARICH) is located in the forward endcap and consists of two 2 cm thick layers of aerogel [45]. In principle, if a charged particle of sufficient velocity hits the aerogel, photons are emitted via the Cherenkov effect. To capture the intensity of photons in each aerogel tile, 420 Hybrid Avalanche Photo-Detectors (HAPDs) arranged in 7 concentric rings are employed. The primary purpose of Aerogel Ring-imaging Cherenkov (ARICH) is to effectively separate pions and kaons in the full kinematic range of the experiment, i.e. from 0.5 to $4.0\,\text{GeV}/c$. As a second form of particle identification, the Time-Of-Propagation counter (TOP) is utilized and is located in the barrel region [46]. It comprises 16 bars of fused silica, each equipped with a mirror located at the forward end of each bar. On the backward end, a prism is located which couples to an array of micro-channel-plate photomultiplier tubes (MCP-PMTs). When a particle with a large enough velocity hits the fused silica, it emits Cherenkov radiation. Different to ARICH, the Cherenkov photons are trapped inside the bar by total internal reflection. TOP reconstructs the Cherenkov emission angle by measuring the propagation time of the Cherenkov photons from their emission to the sensor plane. The particles' initial velocity is directly connected to the reconstructed emission angle, providing an additional source of particle identification compared to ARICH.

Electromagnetic Calorimeter The Electromagnetic Calorimeter (ECL) is a high granularity calorimeter built of 8736 CsI(Tl) scintillating crystals [47]. Its main purpose is the detection of photons and neutral hadrons which produce showers within the crystals and are reconstructed from local energy depositions, referred to as "clusters". In addition, the ECL is also used to separate between electrons and charged hadrons. In the case of electrons, the particles often deposit their entire energy in the crystals, while for charged

hadrons only a small fraction of their energy is left within the ECL. Each readout channel consists of a charge-sensitive preamplifier followed by a waveform analysis realized in Field Programmable Gate Arrays (FPGAs), providing both amplitude and time reconstruction. The ECL covers the polar angle region of $12.4^\circ < \theta < 155.1^\circ$ except for two gaps of $\sim 1^\circ$ between the barrel and endcap regions [48]. While the crystals were previously used in the Belle calorimeter, the calorimeter electronics have been significantly upgraded to keep up with the higher beam background rates expected at the target luminosity of SuperKEKB.

K_L^0 and Muon Detector The outermost subdetector is the K_L^0 and muon detector (KLM), consisting of a cylindrical barrel part and two planar sections located in the endcaps. The active components of the Belle KLM were based on glass-electrode Resistive Plate Chambers (RPCs) located outside the solenoid magnet. While RPCs delivered sufficient results in the scope of the Belle experiment, the expected higher background rates exceeds the RPCs long dead time. To mitigate this problem, the KLM at Belle II additionally employs scintillator strips with wavelength-shifting fibers in the inner barrel and endcap regions.

4 Tools

The analysis presented in this work relies on multiple tools and techniques. In this chapter, I present fundamentals such as Boosted Decision Trees (BDTs), the Belle II software framework, B tagging and the datasets used in this analysis.

4.1 Boosted Decision Trees

Machine learning algorithms are widely considered when addressing classification tasks, given their advantageous handling of complex and variable data at a large scale. While many machine learning methods are suitable for classification, decision trees have established themselves as a reliable approach due to their intuitive design and good performance in modeling non-linear relationships among features. While only key concepts are summarized here, a general introduction to decision trees and the method of boosting is given in [49].
Decision tree models partition the feature space into a set of rectangles, each optimized on a Figure of Merit using a labeled dataset. Each decision is represented as a "node" and is formed in a binary split. The number of sequential decisions within a tree is referred to as the depth of the decision tree, a hyperparameter which must be optimized for each task. To set up a classifier with only a single decision tree is straightforward but is often unstable. A key reason for this is the hierarchical nature of the tree: A small change or error in the top split propagates to all nodes down below. Choosing a more stable split criterion does not remove the inherent instability of the simple estimator. To mitigate this issue, a diverse ensemble of decision trees is considered to counteract the statistical fluctuations. Two methods are very popular to create this ensemble: The method of bagging [50] and boosting [51].
In bagging (short for "bootstrap aggregating"), multiple different training sets are created by sampling with replacement from the original dataset. Each dataset, often referred to as "bootstrap" sample, is exploited to train a single decision tree. The ensemble of decision trees is then used as the classifier, drastically improving on the high variance compared to a single decision tree.
However, in boosting, each decision tree is built sequentially. During each iteration one

assigns weights to each datapoint, emphasizing those that previous models misclassified. Hence for subsequent trees, the focus is shifted to rather hard-to-classify cases, gradually improving the models' overall accuracy. Unlike bagging, the boosting approach simultaneously reduces variance and bias.
A special instance of a boosting algorithm is stochastic gradient boosting [52]. In each iteration, a subsample of the training dataset is randomly selected to fit the base learner. By adding this stochastical component, overfitting is reduced and the boosting process is more efficient. The resulting classifiers are referred to as Stochastic Gradient Boosted Decision Trees (SGBDTs).
This work uses classifiers based on decision trees in multiple instances, including B tagging, background suppression and particle identification. In two of these three applications, I employ SGBDTs from the FastBDT software package [53]. Compared to other popular frameworks used at Belle II such as TMVA [54], scikit-learn [55] and XGBoost [56], the FastBDT framework implements SGBDTs in a fast and cache-friendly manner.

4.2 The Belle II Analysis Software Framework

Modern high-energy physics (HEP) experiments such as Belle II collect a large amount of data. In order to fully explore the potential of this data, accurate and efficient algorithms concerning simulation, reconstruction and analysis are obligatory. The Belle II Analysis Software Framework (basf2) [57, 58] contains the major software components to process data at Belle II. The software is partitioned into individual modules each targeting a different purpose. The majority of components are written as modules in `C++` and build on high-energy physics (HEP) specific software such as the `ROOT` framework [59]. The modules are arranged and configured in an ordered manner using a more user-friendly `Python` front-end.
For this work, the analysis subpackage is of particular interest. The modules in this subpackage are operating on particle candidates, created by tracks and clusters associated with particle hypotheses. Given a single event, the recombination is usually ambiguous, resulting typically in multiple candidates reconstructed per event. To decrease the induced multiplicity to unity, many approaches are sufficient and are presented in this work at a later stage. A helpful tool in this regard are the tracks and clusters in an event not affiliated with the signal candidate. The Rest Of Event (ROE) objects are furthermore beneficial for the suppression of events where no B mesons are produced or missing background processes.

4.3 Full Event Interpretation

At Belle II, a huge amount of collision events containing an $\Upsilon(4S)$ resonance are recorded. These decay at least 96% of the time into a $B\overline{B}$ meson pair, gradually decaying into particles

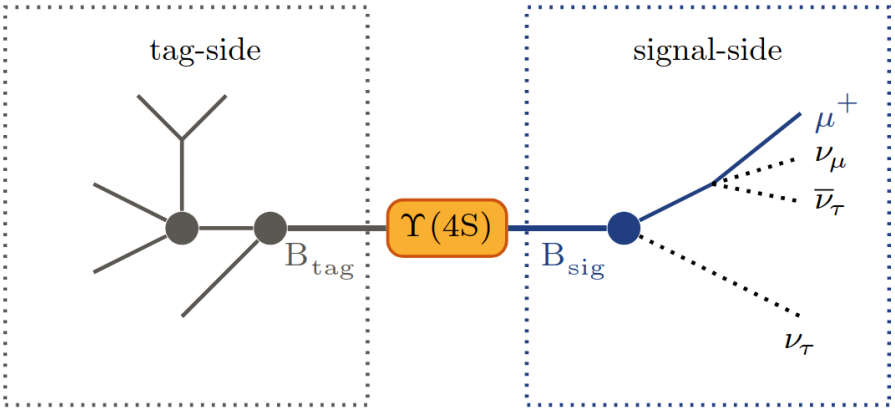


Figure 4.1: Schematic illustration of the tagging method. A typical B decay of interest $B^+ \to \tau^+(\to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$ is illustrated as signal-side B decay on the right side. The left side depicts an arbitrary reconstructed tag-side B decay. It is important to know that the two sides are overlapping spatially in the detector. The assignment of a measured track to the correct side is in general a challenging task. Taken from [60].

that are considered stable in the Belle II detector. However, the measurement of rare B decays such as $B^+ \to K^{*+}\tau^+\tau^-$ with multiple neutrinos in the final state is very challenging. An advantageous tool for this purpose is the method of tagging. For this, the $B\overline{B}$ pair is usually divided into two sides: The signal-side consists of all tracks and clusters associated with the signal B decay one is investigating. On the other hand, the tag-side is composed of remaining tracks and clusters consistent with an arbitrary B meson. This procedure is illustrated for an example decay in Figure 4.1. The tag-side can be reconstructed in both semileptonic or hadronic decay chains which distinguish themselves in several aspects. The former is easily defined due to the presence of a high-momentum lepton and in combination with the higher semileptonic branching fractions, the semileptonic reconstruction usually yields a higher tag-side efficiency. The hadronic reconstruction suffers from lower branching fractions of hadronic B decays, but in contrast, the four-momentum is well-defined given the absence of additional neutrinos. This constraint is very powerful for the considered channel of this work, hence only the hadronic reconstruction is considered in the following. The approach of tagging has been adopted by many B factories. For the Belle experiment, the concept was realized in a NeuroBayes-based Full-Reconstruction (FR) algorithm presented in Ref. [28]. In order to increase the amount of reconstructed B decay chains, a hierarchical selection procedure is employed. This approach consists of addressing each candidate with a Bayesian signal probability, which acts as an input for higher-staged neural networks. The algorithm was used previously in the measurement of $\mathcal{B}(B^0 \to K^{*0}\tau^-\tau^+)$ as presented in Chapter 2.

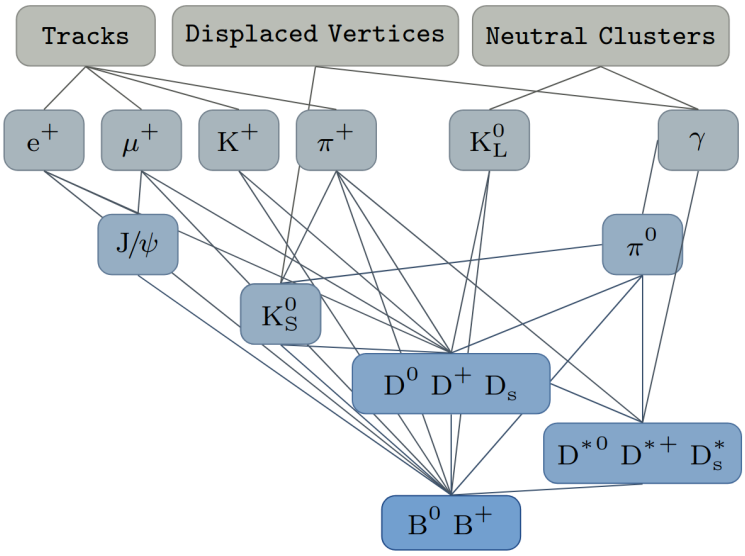
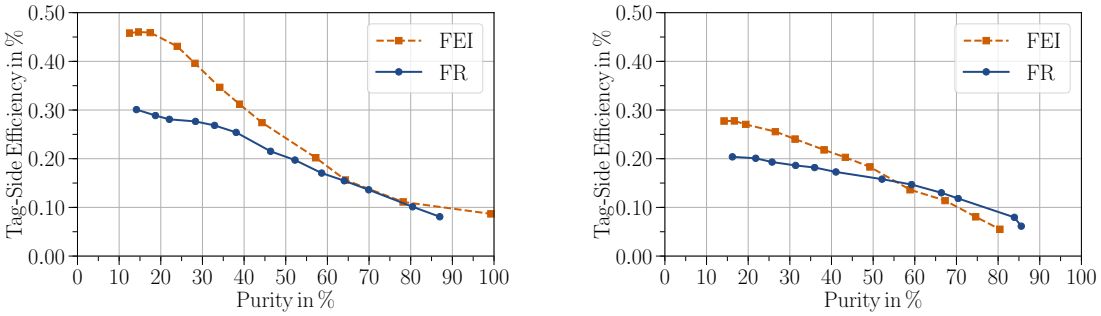


Figure 4.2: Schematic illustration of the Full Event Interpretation. B meson candidates are reconstructed in six stages. Starting from information on the detector level, long-lived particles and recursively intermediate particles are formed. At each stage a FastBDT classifier is employed to rank the reconstructed candidates. Taken from [60].

The successor of the FR algorithm for the Belle II experiment is the Full Event Interpretation (FEI) [60]. In principle, the FEI algorithm follows a hierarchical nature of six reconstruction and classification stages to forge a tag-side B candidate. A schematic overview is given in Figure 4.2. The algorithm starts by collecting available data on subdetector level such as tracks, ECL and KLM clusters as well as pre-vertexed displaced particle candidates objects. Based on this information, long-lived charged (e^-, μ^-, K^+, π^+ and p) and neutral (K_L^0 and γ) particle candidates are formed. From here, intermediate particles are recombined iteratively either from the long-lived particles or from previously reconstructed intermediate candidates. At each step, a vertexing algorithm determines the probability of the recombined constitutes stemming from a common decay vertex. A FastBDT classifier is employed for each reconstruction step. In addition to the information provided by the vertexing algorithm, case-specific features are used to determine the classifier output that is then further propagated. At the top-level, the classifier assigns a signal probability to 20 B meson candidates for the analysis. In order to keep the multiplicity to unity, a so-called Best Candidate Selection (BCS) is utilized. Typically one selects the B meson candidate with the highest signal probability, i.e. the highest classifier output.
Compared to the FR algorithm, the FEI tagging algorithm benefits from the improved classification method and additional channels leading to a larger tag-side efficiency and higher purity for the partner B meson as illustrated in Figure 4.3.



(a) Charged tag-side B mesons

(b) Neutral tag-side B mesons

Figure 4.3: Comparison of the Receiver operating characteristics of a charged (left) and a neutral (right) tag-side B meson reconstructed with the FR and FEI algorithm, both extracted from a fit on the beam-constrained mass on converted Belle data. In the charged case, the FEI outperforms the FR algorithm in the low and high purity domain. Taken from [60].

4.4 Datasets

Up until the first long shutdown in 2022, the Belle II experiment has acquired and simulated a huge amount of data which is used throughout this analysis. For the scope of this work, I refer to events measured by the Belle II detector as **recorded data** while **simulated or Monte Carlo (MC) data** refers to events generated from particle and detector interactions simulated with dedicated software[1].

Recorded Data In multiple instances, this work relies on different recorded data samples. The main dataset for B physics is recorded at the $\Upsilon(4S)$ resonance and consists of $365.37 \pm 1.70\,\text{fb}^{-1}$. An additional amount of $42.74 \pm 0.20\,\text{fb}^{-1}$ is recorded slightly below the $\text{B}\overline{\text{B}}$ pair production threshold and solely consists of $e^+e^- \to \text{q}\overline{\text{q}}$ events. The dataset is used for calibration purposes later in Section 5.6.4. A detailed overview of recorded datasets is listed in Table 4.1.

Simulated Data Simulated events are used to model the underlying composition of signal and background components with truth information provided by the event generator. Regarding the signal events, I use a simulated sample of 50 million $\text{B}^+ \to \text{K}^{*+}\tau^+\tau^-$ events at the $\Upsilon(4S)$ resonance, where one B meson is restricted to decay into a $\text{K}^{*+}\tau^+\tau^-$ product (that further decays generically) while the other decays generically. For the $\text{q}\overline{\text{q}}$ and $\text{B}\overline{\text{B}}$ backgrounds, I use an amount representing $1\,\text{ab}^{-1}$ for each type which corresponds roughly to 3 times the recorded data sample size. The collection of used datasets at the $\Upsilon(4S)$ energy is listed in Table 4.2.

1 In instances where the word "data" is used, I refer to recorded and simulated data alike.

Table 4.1: The integrated luminosity of recorded data samples collected by the Belle II experiment from 2019 to 2022. The largest share of data is obtained at the $\Upsilon(4S)$ resonance, the most relevant domain for B physics. The supporting off-$\Upsilon(4S)$ is recorded 60 MeV below the threshold. The $\Upsilon(5S)$ scan dataset contains various recorded data beyond the $\Upsilon(4S)$ threshold and is exploited for searches on exotic vector resonances. Not listed are $\Upsilon(4S)$ scans, since they make up a negligible amount of the total integrated luminosity. The uncertainty on the luminosity is composed out of a statistical and systematical component. Taken from [61].

Type	$\sqrt{s}$ (GeV)	$\mathcal{L}$ (fb^{-1})
$\Upsilon(4S)$	10.580	365.37 ± 1.70
off-$\Upsilon(4S)$	10.520	42.74 ± 0.20
$\Upsilon(5S)$ scan	10.657	3.54 ± 0.03
	10.706	1.63 ± 0.02
	10.751	9.88 ± 0.06
	10.810	4.71 ± 0.03
Total	—	427.87 ± 2.01

Table 4.2: Simulated datasets at the $\Upsilon(4S)$ energy used throughout this analysis. The background consists in each case of $1\,\mathrm{ab}^{-1}$ which calculates into generated events by considering the cross-section of the event of interest.

Process	Integrated luminosity [fb^{-1}]	Generated events $\times 10^6$
$\mathrm{B}^+ \to \mathrm{K}^{*+}\tau^+\tau^-$		50
$e^+e^- + \to \mathrm{u\bar{u}}$	1000	1586
$e^+e^- + \to \mathrm{d\bar{d}}$	1000	396
$e^+e^- + \to \mathrm{s\bar{s}}$	1000	362
$e^+e^- + \to \mathrm{c\bar{c}}$	1000	1300
$e^+e^- + \to \Upsilon(4S) \to \mathrm{B^0\overline{B}^0}$	1000	508
$e^+e^- + \to \Upsilon(4S) \to \mathrm{B^+B^-}$	1000	539

Additionally, for the off-$\Upsilon(4S)$ cases, I utilize a dedicated data sample for $\mathrm{q\bar{q}}$ events simulated at the off-$\Upsilon(4S)$ CMS energy. The integrated luminosity associated with each sample is $50\,\mathrm{fb}^{-1}$ and is only utilized for calibration purposes.

5 Reconstruction & Selection

In contrast to collision-based selections, the candidate-based selection approach allows for an independent analysis of the two B mesons originating from a $\Upsilon(4S)$ resonance. I describe in detail the reconstruction and selection criteria for the tag B characterized by the FEI algorithm (see Chapter 4) and the signal B meson. The decay topology is depicted in Figure 5.1.

5.1 Tag-Side Selection

Applying the FEI algorithm is time-consuming and computation-wise highly demanding. Therefore, all datasets used for reconstruction come with a pre-applied FEI tagging, so-called FEI skims. The algorithm operates on displaced vertices, neutral clusters and tracks affiliated with so-called final state particles (FSPs) in the detector. These are a category of particles considered stable in the relevant time domain of interactions at the Belle II experiment. A tag-side B meson is stepwise defined from FSPs through intermediate resonances and labeled with an associated probability estimated by a multivariate classifier. As a configuration for this analysis, I use the hadronic tagging method which solely reconstructs the tag B meson in hadronic decay chains. The main advantage of this procedure is the high tag-side purity stemming from the well-known four-momentum given to the absence of neutrinos. The initial energy is precisely known, hence this is a powerful constraint for selecting a signal B candidate. However, this method comes with smaller tag-side efficiency arising from the lower hadronic branching fractions of B mesons.
In order to efficiently use the algorithm, performing a calibration is necessary to counterbalance any discrepancies between Data and MC datasets. This calibration leads to a correction factor which needs to be applied when comparing data with MC samples. I describe the calibration procedure in detail in Section 5.6.
With the tag-side B meson defined and reconstructed by the FEI algorithm, the employment of further selection criteria is immensely beneficial. A more refined tag-side ensures a lower event multiplicity and further prevents the use of constituents from the tag B meson in the signal B meson reconstruction or vice versa.

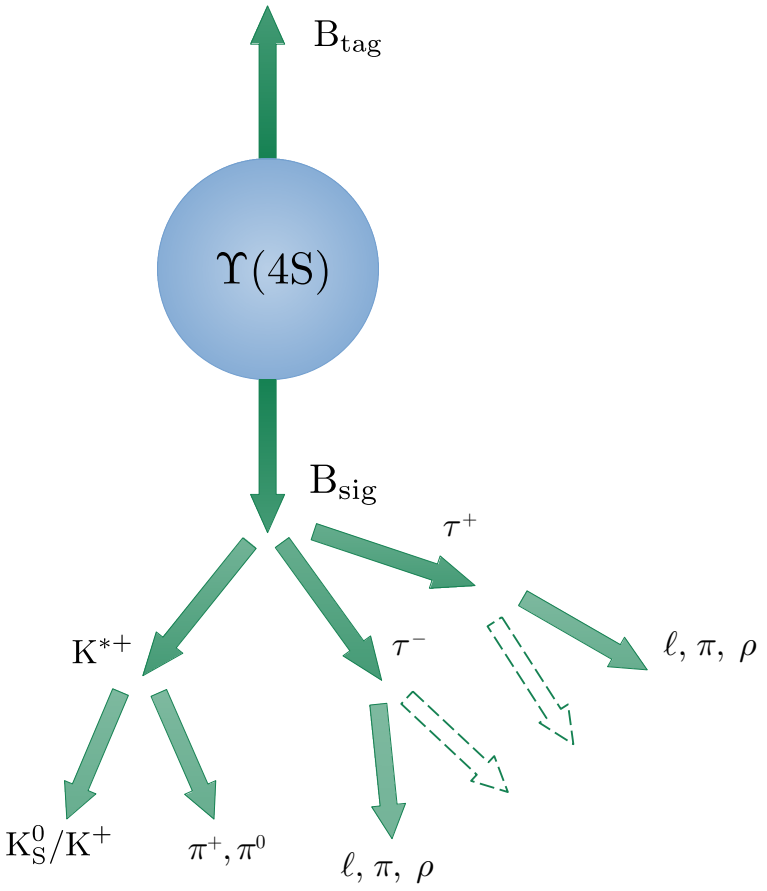


Figure 5.1: Decay of a $\Upsilon(4S)$ resonance (blue) into a signal B and an arbitrary tag-side B meson. For the signal side, only the channels used for reconstruction are illustrated. Dashed arrows denote undetected neutrinos originating from τ-lepton decays.

A good starting point to increase the ratio of correctly reconstructed tag-side B mesons to falsely corrected ones are kinematic properties. For this, it is useful to look at common properties of B mesons originating from B Factories: The invariant mass distribution of correctly reconstructed B candidates should peak around the well-known mass of B mesons and have an energy at around $\sim 50\%$ of the $\Upsilon(4S)$-resonance energy. Typically, two more engineered variables, M_{bc} and ΔE, are preferred which require precise knowledge of the beam energy in the center-of-mass system (CMS), namely $E^{\mathrm{beam}}_{\mathrm{CMS}}$. The beam constrained mass M_{bc} is defined as

$$M_{\mathrm{bc}} = \sqrt{\left(\frac{E^{\mathrm{beam}}_{\mathrm{CMS}}}{c^2}\right)^2 - \left(\frac{\mathbf{p}_{\mathrm{CMS,B}}}{c}\right)^2}, \tag{5.1}$$

with the spartial momentum of the B meson $\mathbf{p}_{\mathrm{CMS,B}}$ both in the CMS. The difference in energy is defined as

$$\Delta E = E_{\mathrm{CMS,B}} - E_{\mathrm{CMS}}^{\mathrm{beam}} \,, \tag{5.2}$$

where hereby $E_{\mathrm{CMS,B}}$ denotes the reconstructed energy of the B meson.
Since a $\mathrm{B}^+\mathrm{B}^-$ pair is produced almost at rest, these variables are only weakly correlated. In contrast to naive kinematic definitions, both variables demonstrate their advantage by exploiting the resolution properties of the detector: ΔE strongly relies on the resolution of the electromagnetic calorimeter and the particle mass hypothesis used to forge the B meson from its constituents. In case of misidentified particles in the process, the energy difference of these particles manifests in a shift in ΔE, while for correctly reconstructed events the variable is expected to peak around zero.
On the other hand M_{bc} is by design independent of the final-state particle hypothesis, making it powerful in suppressing combinatorial background (e.g. induced by a light quark pair) in B meson candidates. For correctly reconstructed B mesons, the M_{bc} distribution should peak at the nominal B mass, considering the combinatorial background is more evenly distributed within the allowed kinematic range. To ensure a high purity in the tag-side selection, a criterion of $M_{\mathrm{bc}} > 5.27\,\mathrm{GeV}/c^2$ and $-0.15 < \Delta E < 0.1\,\mathrm{GeV}$ is applied. In addition, the probability given to each reconstructed tag-side B candidate by the FEI algorithm is used to increase the tag-side purity and overall significance against background. For a more refined tag side, a requirement on $\mathcal{P}_{\mathrm{FEI}} > 0.01$ is set. The decision of this selection criteria is an important part when calibrating the FEI and correcting the simulation. For a tag-side BCS, many approaches are taken into consideration. To make as few assumptions on the tag-side B meson as possible, the candidate with the highest probability score is chosen.

5.2 Signal-Side Selection

With the tag-side B meson fully described, one is now able to move on to the key part of the reconstruction: The reconstruction of the signal B meson.
For a signal $\mathrm{B}^+ \to \mathrm{K}^{*+}\tau^+\tau^-$ reconstruction, multiple different final-state and intermediate particles are taken into consideration. I characterize the selection criteria applied on the FSPs first and the ones of the recombined particles in succession to this.
In case of the K^{*+}, this analysis focuses on the $\mathrm{K}^{*+} \to \mathrm{K}^+\pi^0$ and $\mathrm{K}^{*+} \to \mathrm{K}^0_{\mathrm{S}}\pi^+$ decay chain. Charged mesons K^+ and π^+ are considered as FSPs in the detector domain and thus are easily reconstructed (see also Section 5.6.2). The neutral components of the K^{*+} are decaying within the detector and therefore need to be recombined from their final state constituents.
Considering the τ-lepton, one final-state lepton ($\ell = e, \mu$), one final-state charged pion π^+,

or one reconstructed ρ meson is required to forge the third generation lepton candidate. The ρ meson decays via strong interaction in almost all cases into one charged and one neutral pion. Due to the very short lifetime, ρ mesons have a rather large decay width which needs to be taken into account while reconstructing it.

Final state particles Each FSP gets allocated a particle trajectory ("track") from the tracking detectors and/or an energy deposition ("cluster") in the ECL based on their underlying particle hypothesis. I use reconstructed track parameters dr and dz at the point of closest approach (POCA) with respect to the evaluated interaction point (IP).
For the charged particles, the only kind of FSPs that leave traces in the tracking detectors, a requirement for distances of $dr < 2\,\text{cm}$ and $|dz| < 4\,\text{cm}$ is set to suppress beam interaction backgrounds. In order to further increase the accuracy and its robustness in the determination of track momenta and vertices, only tracks with a transverse momentum $p_T > 100\,\text{MeV}/c$, number of hits in the CDC > 20 and the polar angle within $17^\circ < \theta < 150^\circ$ are used. The latter selection ensures a propagation within the CDC detection range for the FSPs. As a final track selection criteria, the total energy associated with the track E proves to be of high use for setting a veto on $e^+e^- \to e^+e^-$ events. While running at the $\Upsilon(4S)$-resonance energy, the final state electrons produced in Bhabha scattering are highly energetic in the domain of the Belle II experiment. Thus, a selection criterion on the track energy $E < 5.5\,\text{GeV}$ is applied to veto these kinds of events.
Along with track selection criteria, effective and efficient charged particle identification (PID) is a vital component in a physics analysis. Good particle identification is necessary to reduce multiple forms of background and isolate hadronic final states. In the case of the charged hadron FSPs, a probability calculation is performed dynamically by the particles' hypothesis. The probability of Kaon identification is hereby given as a combined likelihood function

$$\mathcal{P}(\mathrm{K}) = \frac{\mathcal{L}_\mathrm{K}}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_\mathrm{K} + \mathcal{L}_\mathrm{p} + \mathcal{L}_\mathrm{d}}, \tag{5.3}$$

where $\mathcal{L}_X$ describes the likelihood for each of the six long-lived charged particle hypotheses: electron, muon, pion, kaon, proton, and deuteron. A likelihood value is determined for each particle hypothesis using information from all available detector systems. To compute the probability of pion identification, one has to adjust the corresponding likelihood function in Equation (5.3) with $\mathcal{L}_\pi$ in order to obtain $\mathcal{P}(\pi)$. For both charged kaon and pion, a tight selection criteria of $\mathcal{P}(\mathrm{K}) > 0.9$ and respectively $\mathcal{P}(\pi) > 0.9$ is enforced to ensure hadronic final states of high purity.
Regarding the final state leptons, a pre-trained BDT classifier assigns a global probability to each lepton candidate within the reconstruction. During evaluation of this step, around $\sim 25\%$ of candidates are observed to yield `NaN` values. By default, the track associated with the lepton particle candidate is required to have a corresponding cluster within the

ECL. In some occasions it is possible that if the matched cluster lies in between the ECL gaps or its energy deposit is below 55 MeV, the probability gives `NaN` in return. For these cases, the reconstruction is insufficient and thus these particle candidates are removed. The remaining ones end up with a classifier output ranging from 0 to 1 which can be translated into a probability. A selection criterion of $\mathcal{P}_{\mathrm{BDT}}(e) > 0.9$ and $\mathcal{P}_{\mathrm{BDT}}(\mu) > 0.9$ is imposed, retaining $\sim 42\%$ of electron and $\sim 45\%$ of muon candidates.

Recombined particles In contrast to FSPs described in the previous paragraph, recombined particles cannot be considered stable as they decay within the detector. Products of the decays (except neutrinos) interact with the different detector systems leading to tracks and allocated clusters in the calorimeter. When all decay products are measured, the mother particle is reconstructed by summing up the four momenta of its constituents. Additional selection criteria are applied for each particle type individually.

Neutral $\mathbf{K_S^0}$ mesons For the $\mathrm{K_S^0}$ originating from a K^{*+} decay, the charged pion decay channel is taken into consideration. By nature, the $\mathrm{K_S^0}$ decays in $\sim 30\%$ of the time into a pair of neutral pions which subsequently each decaying into a photon pair. Due to the energy scale of the experiment, the four resulting photons are of low energy making a correct reconstruction a challenging task. Furthermore, the corresponding mass resolutions are very different which would propagate through the whole analysis, hence such events are not reconstructed.
The $\mathrm{K_S^0}$ is recombined in two ways separately: First, the recombination is managed by forging it from charged particle pairs that originate from a common neutral mother particle. These exhibit a characteristic V-shape signature emerging from the decay far off the IP. Additionally, the $\mathrm{K_S^0}$ is built from all charged pions within the observed event. Both approaches perform a vertex fit first and apply a loose mass criterion on the $\mathrm{K_S^0}$ of $0.45 < \mathrm{M} < 0.55\,\mathrm{MeV}/c^2$. The resulting list of particle candidates is ultimately merged while prioritizing $\mathrm{K_S^0}$ from a displaced vertex source in events with multiple candidates. Moreover to enrich the purity, momentum binned selection criteria from Table 5.1 on the $\mathrm{K_S^0}$ candidates are imposed. This collection of criteria was originally proposed for the Belle experiment but its use in Belle II is justified (e.g. see Ref. [62]). For a $\mathrm{K_S^0}$ candidate, a final tighter mass window is enforced to account for the comparably sharp peak in the invariant mass. To boost the sensitivity, the mass window is set to be $0.48 < \mathrm{M} < 0.52\,\mathrm{MeV}/c^2$.

Neutral $\boldsymbol{\pi^0}$ mesons Photons that qualify to create π^0 meson candidates are formed from allocated clusters in the ECL. As the neutral pions decay in a very brief moment after being formed, the efficiency and quality of the π^0 reconstruction is strongly determined by the ECL. Core restrictions in reconstructing the π^0 are photons escaping the ECL, as well as two photons merged in a single cluster making them inseparable. To increase the

Table 5.1: Momentum-binned selection criteria applied to each K_S^0 candidate. The IP parameters dr and $d\phi$ denote the vertex distance and the vertex azimuthal angle, respectively, while z_{dist} indicates the distance of the daughter tracks at POCA, and FL denotes the flight length. The selection criteria follow those presented in Ref. [63].

Momentum range (GeV/c)	dr(cm)	$d\phi$(rad)	z_{dist}(cm)	FL (cm)
< 0.5	> 0.05	< 0.3	< 0.8	-
0.5-1.5	> 0.03	< 0.1	< 1.8	> 0.08
> 1.5	> 0.02	< 0.03	< 2.4	> 0.22

photon resolution, a requirement of the number of weighted crystals > 1.5 and energy thresholds are set, the latter depending on the cluster region within the ECL. These specify to having at least 80 MeV in the forward region, 30 MeV in the barrel and 60 MeV in the backward endcap region. In addition, the cluster angle has to be in CDC acceptance region ($17^\circ < \theta < 150^\circ$) to ensure the allocated cluster stems from a neutral energy deposit.
Neutral π^0 candidates are now formed by combining two photons that surpass the previously mentioned selection criteria by imposing a mass window of $120 < \mathrm{M} < 145\,\mathrm{MeV}/c^2$. The π^0 are boosted within the lab frame, hence the photons should be emitted in a collimated manner. Thus only π^0 candidates are taken into consideration where the angle between two photons is $\phi_{\gamma\gamma} < 1.5\,\mathrm{rad}$, while the difference of the azimuthal angle between both fulfills $|\Delta\varphi| < 1.5\,\mathrm{rad}$.

Charged ρ mesons A neutral and a charged π candidate surpassing the aforementioned criteria are qualified to be recombined into a ρ meson. A broader mass window is enforced by limiting the invariant mass to $0.65 < M < 0.9\,\mathrm{GeV}/c^2$ to account for the large decay width of the particle, following the work in Ref. [64].

Charged K^{*+} mesons The final piece missing is the recombination of the excited K meson state. Again, the K^{*+} is reconstructed in $\mathrm{K}^{*+} \to \mathrm{K}_S^0\,\pi^+$ and $\mathrm{K}^{*+} \to \mathrm{K}^+\pi^0$ decays which covers about $\sim 66\%$ of all possible decays. A first mass window of $0.6 < M < 1.3\,\mathrm{GeV}/c^2$ is defined to suppress the badly reconstructed K^{*+} mesons. In the following, an additional mass cut $0.790 < M < 0.994\,\mathrm{GeV}/c^2$ is enforced that defines the signal region. This corresponds precisely to a $2\,\sigma$ window around the nominal K^{*+} mass given in Ref. [8].
The remaining mass spectrum ($0.6 < M < 0.79\,\mathrm{GeV}/c^2$ and $0.994 < M < 1.3\,\mathrm{GeV}/c^2$) is referred to as control region ("sideband") and is exploited to safely test the signal reconstruction on a large dataset without operating in the signal sensitive region (see later in Chapter 6).

With these criteria, one is now able to forge a B meson candidate that decays into a K^{*+}

and a τ-lepton pair covering all 1-prong decay modes. The reconstruction of the signal side is complete. However, to further increase the purity, further selection criteria will be applied to the Υ(4S) candidate that is formed after combining a signal B with its tag side counterpart.

5.3 Selections on the Υ(4S) Candidate

After reconstructing B_{tag} as well as B_{sig} candidates from their hypothesis, they are combined to form a Υ(4S) candidate. At this stage, one is given a powerful set of tools: constraints on the Rest Of Event (ROE). Correctly reconstructed Υ(4S) candidates should by definition not have a remaining unassigned track in the ROE, otherwise the reconstruction must be flawed since not all particles could be covered in the candidates' hypothesis. This is referred to as completeness constraint.
To ensure this criterion does not arbitrarily sort out events, a well-reconstructed track in the ROE is defined by surpassing $dr < 2\,\text{cm}$ and $|dz| < 4\,\text{cm}$ as well as $17^\circ < \theta < 150^\circ$ and a minimum transverse momentum of $p_T > 100\,\text{MeV}/c$. The multiplicity after this step ranges from 1.33 to 2.66 (see cut-flow tables in Appendix A.1) depending on the signal channel portrayed. This indicates that on average, multiple candidates deliver an appropriate description of the signal channel. In said case, it is necessary to introduce an indicator denoting the quality of multiple candidates within one event that picks the most suitable one. Which variable to use as the indicator cannot be generally determined, but is rather analysis-specific. Regarding this work, the Υ(4S) candidate with the K^{*+} mass closest to the nominal mass given by PDG [8] is selected. While this might be sufficient for some events to reduce the multiplicity to one, for some it does not and thus an additional indicator is necessary. For the sake of simplicity, an arbitrary Υ(4S) candidate from the remaining candidates is chosen, ultimately retaining a single candidate per event.
On simulated samples, it is rather easy to evaluate the BCS. Choosing a K^{*+} mass closest to its nominal one trivially favors candidates closest to the mass peak at $892\,\text{MeV}/c^2$ [8]. For truth matched Υ(4S) candidates, one observes a $\sim 10\%$ purity gain with this selection. However, the BCS does not always pick the correct candidate. It occurs in 8.2% of the time that a non-truth matched candidate survives the procedure, while the true candidate gets sorted out. Reasons for this range from a candidate stemming from combinatorial background to detector effects that lead to an incorrect mass close to the nominal one.
Choosing a Υ(4S) candidate randomly decreases the signal purity by 5.6%. Due to the very low multiplicity in this regime (see Table A.2), a large number of events solely accommodate a single candidate, refusing a potential choice in the first place.

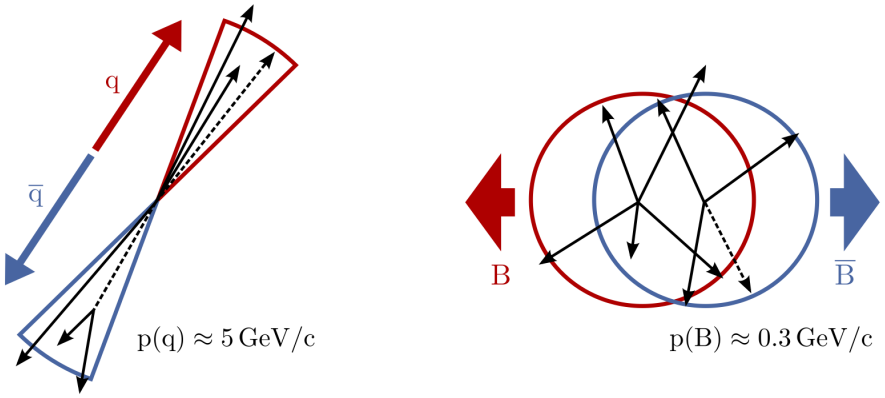


Figure 5.2: Depiction of the distinct event shapes for a continuum event (left) and a $B\overline{B}$ event (right) in the CMS. Given the high amount of energy remaining from e^+e^- collisions forming a $q\overline{q}$ pair, the back-to-back cone-like structure is observed. These events can be distinguished from the spherical topology of B meson decays produced in $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ interactions. Taken from [65].

5.4 Suppression of Light-Quark Pair Background

Most e^+e^- interactions at the $\Upsilon(4S)$ CMS energy do not result in a $\Upsilon(4S)$ resonance which then decays into a B meson pair. A large portion of these non-$\Upsilon(4S)$ events are usually unproblematic in B decay analyses as the trigger already rejects them.
Contributions from continuum background originating from $e^+e^- \to q\overline{q}$ events (where q = u, d, s, c) provide more of a challenge since B mesons reconstructed from these decays show a broad distribution in variables, making them harder to separate from the signal component. With a combined cross-section of 3.7 nb, these interactions are much more likely than the ones resulting in a $B\overline{B}$ pair corresponding to 1.1 nb.
However, the overall shape of the event is contrary: For a given $B\overline{B}$ event, both B mesons are produced almost at rest in the CMS since the $\Upsilon(4S)$ mass is barely above the $B\overline{B}$ production threshold. This results in B decay products being isotropically distributed within the $\Upsilon(4S)$ rest frame. In contrast, $e^+e^- \to q\overline{q}$ interactions produce a jet-like cone structure stemming from the large initial momentum. Both event shapes are illustrated in Figure 5.2. In pursuance of separating continuum events from signal, several powerful sets of variables are established at Belle II.

Thrust variables A fundamental difference between a signal B meson and continuum background are the distinct kinematic topologies. A useful observable in this regard is the thrust axis **T**: For a collection of particle momenta $\mathbf{p}_i$ $(i = 1, ..., N)$, the thrust axis is defined as the unit vector maximizing the sum of momenta projections onto itself, resulting

in

$$\mathbf{T} = \max_{|\mathbf{T}|=1} \left(\sum_{i=1}^{N} \mathbf{T} \cdot \mathbf{p}_i \right) . \tag{5.4}$$

This directly leads to the definition of the thrust scalar T given by

$$\mathrm{T} = \frac{\sum_{i=1}^{N} |\mathbf{T} \cdot \mathbf{p}_i|}{\sum_{i=1}^{N} |\mathbf{p}_i|} , \tag{5.5}$$

a quantity that is precisely known once the thrust axis is determined. The given variable quantifies the jet-likeness of the decay particles, taking values close to 1 for the collimated case. For B meson candidates, the isotropic distribution leads to a thrust scalar $T \sim 0.5$. A useful related variable in this context is $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ describing the cosine of the angle between the thrust axis of the B meson $\mathrm{Thrust_B} \equiv \mathbf{T_B}$ and the thrust axis of the ROE, $\mathrm{Thrust_{ROE}} \equiv \mathbf{T_{ROE}}$. Due to the isotropic distribution of the B decay products, their thrust axes are randomly scattered, leading to a uniform distribution for signal B mesons in $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ within the range $[0, 1]$. The heavily collimated manner of $\mathrm{q\overline{q}}$ events on the other hand yields a distribution strongly peaking at 1, making it possible to separate $\mathrm{B\overline{B}}$ events from it. Especially prone to continuum background are final states involving pions. Hadronic properties of the continuum events lead to a larger misidentification rate in that regime, requiring an additional selection criteria to exclude these events. As a consequence, I impose an additional selection criterion, namely $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}}) < 0.9$. This leads to a signal loss of $\sim 10\%$ while removing $\sim 34\%$ of continuum events. Furthermore, discrepancies when comparing Data to MC are mitigated given the removal of the poorly modeled candidates.
Similarly one introduces the variable $\cos(\mathrm{Thrust_B}; z)$, based on the angle between the thrust of the B meson and the beam axis (corresponding to the z-axis). The discriminating power of this variable arises from the difference in angular distributions: the spin-1 $\Upsilon(4\mathrm{S})$ is decaying into two spin-0 B mesons; in contrast to the $e^+e^- \to \mathrm{q\overline{q}}$ processes and the hadronization of the spin-1/2 quarks.

Fox-Wolfram moments Another useful parametrization of phase-space distributions of momentum and energy in an event is given by Fox-Wolfram moments [66]. For a collection of N particles with momenta p_i, the k-th order Fox-Wolfram moment H_k is defined as

$$H_k = \sum_{i.j}^{N} |p_i||p_j| P_k \left(\cos\theta_{ij} \right) , \tag{5.6}$$

where $\theta_{i,j}$ denotes the angle between each pair of particles and P_k is the k-th order Legendre polynomial. In the original work by Fox and Wolfram, an additional factor of $1/\sqrt{s}$ is used

to weight each addend. Practically, one usually makes use of the reduced Fox-Wolfram momentum R_k given by

$$R_k = \frac{H_k}{H_0}\,, \tag{5.7}$$

where the constant factor cancels out for each component. A common choice here is R_2. For events with two strongly collimated jets, R_k takes values close to zero (one) for odd (even) values of k. Sharp signatures as such provide a convenient discrimination between events exhibiting different forms of decay topologies.
Noticeable disagreement between data and current MC campaigns for $\mathrm{B\overline{B}}$ and $\mathrm{q\overline{q}}$ is observed in the R_2 variable for this analysis. This arises either from insufficiencies in the $\mathrm{q\overline{q}}$ simulation described later in Section 5.6 or effects induced by beam background. As a result, the variable R_2 will not be used to separate the continuum from signal.

To extend the continuum suppression toolbox, modified Fox Wolfram moments, referred to as Kakuno Super Fox Wolfram moments (KSFWs), are considered, first introduced at Belle [67]. Two components come into play: The H_l^{oo} ($l \in [0,4]$) variable subset focusing solely on ROE information of the B candidate and H_{xl}^{so} combining information from both signal and ROE. For H_l^{oo}, one differenciates between odd and even cases for l, defining the set of five variables as

$$H_l^{\mathrm{oo}} = \begin{cases} \sum_{i,j} |\mathbf{p}_i||\mathbf{p}_j| P_l(\cos\theta_{ij}), & \text{for even } l \\ \sum_{i,j} Q_i Q_j |\mathbf{p}_i||\mathbf{p}_j| P_l(\cos\theta_{ij}), & \text{for odd } l\,. \end{cases} \tag{5.8}$$

Hereby, i and j run over the ROE particles and $Q_{i,j}$ describe the corresponding particle charges for particles i and j. In cases where k is even, the product of Q_iQ_j is equal to $+1$, making H_k^{oo} equivalent to the associated Fox Wolfram moment for the ROE.
The H_{xl}^{so} are decomposed further into three categories, depending on whether the variable considers only charged particles ($x = 0$), neutral particles ($x = 1$) or the sum of missing particle momenta represented by a pseudo particle ($x = 2$). In the latter case, beam constraints are taken into account. For even l, H_{xl}^{so} is defined as

$$H_{xl}^{\mathrm{so}} = \begin{cases} \sum_{i,jx} |\mathbf{p}_{jx}| P_l(\cos\theta_{ij}), & \text{for even } l \\ \sum_{i,jx} Q_i Q_{jx} |\mathbf{p}_{jx}| P_l(\cos\theta_{ij}), & \text{for odd } l \end{cases} \tag{5.9}$$

following the similar variable definitions as in Equation (5.8), i is referring to the B daughters, jx indicates the ROE particle of category x with the spatial momentum $\mathbf{p}_{jx}$. Again $Q_{i,j}$ describes the charge of the particles in case of l being odd and $x = 0$, for even k the product combines once again to $+1$, while it vanishes in cases of $x = 1, 2$.

For sake of normalization, the KSFWs are scaled with $H_0^{\max}$ or $(H_0^{\max})^2$ with

$$H_0^{\max} = 2\left(E_{\mathrm{CMS}}^{\mathrm{beam}} - \Delta E\right) , \tag{5.10}$$

making these sets of variables independent of ΔE. In total 16 KSFWs are available: five stemming from H_l^{oo} and an additional eleven from H_{xl}^{so}. The choice of whether or not each variable is used later on to suppress continuum background heavily depends on their separation power and overall Data/MC agreement.

Missing mass squared and transverse energy The missing mass squared is defined as

$$M_{\mathrm{miss}}^2 = \left(E_{\Upsilon(4\mathrm{S})} - \sum_{n=1}^{N} E_n\right)^2 - \sum_{n=1}^{N} |\mathbf{p}_n|^2 , \tag{5.11}$$

where $E_{\Upsilon(4\mathrm{S})}$ is the energy of the $\Upsilon(4\mathrm{S})$ candidate, E_n and p_n the energy and momentum of the reconstructed particle n respectively, and N as the total number of FSPs. This variable represents the invariant mass of the invisible particles stemming from the signal B candidate. For well-reconstructed events with at most one massless particle, M_{miss}^2 is expected to peak at zero and to be model-independent. This makes it a good choice for a signal extraction variable (e.g. as used in Ref. [27] and Ref. [35]).
However in this search, at least two massless particles are present, essentially removing the peak at zero. Thus M_{miss}^2 is not further considered.
The transverse energy E_t is the scalar sum of the transverse four momentum for each of the N particles, yielding

$$E_T = \sum_{n=1}^{N} \left|p_{T,n}\right| . \tag{5.12}$$

For particle collisions, this quantity is used to discriminate signal-like events tending to a higher E_t against unwanted background events.

CLEO Cones The variables and tools defined above do not necessarily provide the ideal discrimination of background in all four signal channels. For channels suffering from large background rates, it is useful to introduce another set of variables. The CLEO Cone variables, named after the CLEO Collaboration, were first used in the context of charmless B decays [68]. Said variable set consists of nine quantities describing the momentum flow around the thrust axis of the signal B candidate. Binned in nine cones of 10° (see Figure 5.3), each one is merged with the corresponding counterpart from the opposite axis.
Not all nine CLEO cone variables show good agreement for the signal channels on sideband data, hence only a selection contributes for each channel individually (see Table A.5).

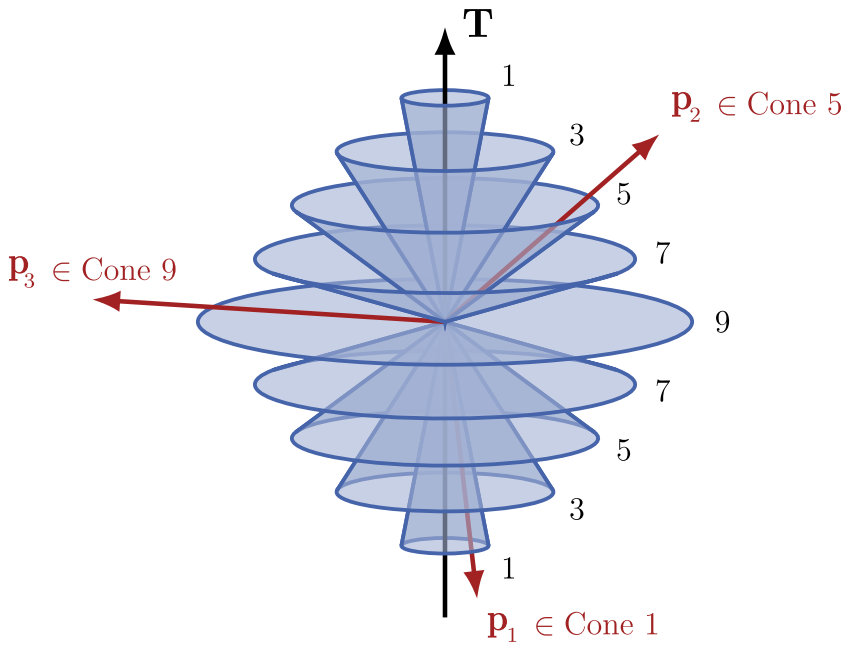


Figure 5.3: Illustration of the nine CLEO cone variables concentrically aligned around the thrust axis **T**. The cones are numbered corresponding to the related CLEO Cone index i. For simplicity, not every CLEO Cone is illustrated in this figure. In red, three arbitrary particle momenta are shown, depicting their affiliation with the momentum flow to the cone regions. Taken from [27].

5.5 Suppression of Generic B Meson Pair Background

A distinct source of background are all $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ events where neither B meson decays into a K^{*+} and a τ-lepton pair. I refer to cases where a charged B meson pair is produced as charged background, if a $B^0\overline{B}^0$ is formed it is specified to be mixed background. Both charged and mixed backgrounds provide a more ambitious challenge to separate them from the signal component. The tools and event-level variables derived in Section 5.4 are insufficient because they are sensitive to event shape topologies. The overall event shape of a charged or mixed background is closely resembling the topology of the signal event. Hence, additional quantities are required that deliver the desired discriminating power.

Variables sensitive to D mesons To find such variables it is crucial to know what types of B^+B^- or $B^0\overline{B}^0$ decays are wrongly getting reconstructed as a B_{tag} and the $K^{*+}\tau^+\tau^-$ resonance. Using simulated samples for charged and mixed background makes it possible to study what events surpass the selection criteria stated in Section 5.1 and 5.2. The important piece of information to retrieve is the decay chain determined by the event generators used to simulate the samples in the first place. Having precise knowledge of both decays of a $B\overline{B}$ pair, it is necessary to determine which of the two B mesons is used to build the

tag-side and signal-side. By extracting the PDG code [8], a unique identifier for particles, this abundance is resolved allowing to precisely identify each B meson.
The distribution of daughters of the generic B meson pair falsely reconstructed as $\mathrm{K}^{*+}\tau^+\tau^-$ is evaluated on the charged background sample only. This is justified, since it is out-dominating the amount of mixed background after reconstruction due to the signal B meson required to carry a charge. In Table 5.2 the most frequent decays misinterpreted as $\mathrm{K}^{*+}\tau^+\tau^-$ are listed, as well as the fraction relating to all decays surpassing the reconstruction algorithm. Most decays consist of at least a single (excited) D meson, ultimately leading to two charged daughters in a subsequent decay in almost all cases (see Table A.3 and A.4). The two charged daughters from a D meson as well as the additional charged particle from the B candidate leave tracks in the detector. This produces a misreconstructed signal event that consists of the same amount of tracks as a correct signal candidate would have. As a consequence of this behavior, a variable sensitive to D mesons is introduced

$$M^2\left(\mathrm{K}^{*+}; t_{1,2}\right) = \left(E_{\mathrm{K}^{*+}} + E_{t_{1,2}}\right)^2 - \left(p_{\mathrm{K}^{*+}} + p_{t_{1,2}}\right)^2 , \tag{5.13}$$

describing the squared mass of the K^{*+} and the daughter of the first or second τ-lepton t_i with $i = 1, 2$. For misidentified cases of true $\mathrm{B} \to \mathrm{D}\ell\nu$ decays on the signal side, these variables peak at the nominal D mass making them sensitive for those events.
According to Table 5.2, this retains a large portion of charged background in principle, making it not necessary to introduce further variables.

BDT classifier With the tools and variables described and estimated on their potential discriminating different types of background, BDTs are employed to classify signal from background. For this, I use BDT classifiers from the FastBDT package [53], one for each of the four signal channels. Every signal channel shows a unique and distinct decay topology and thus is prone to different sorts of backgrounds. A direct consequence of this is that some variables carry a large separating power in one channel, while signal and background distributions are too closely resembling one another in a different one. In addition, only well-modelled input features are considered for the classifier. A full list of utilized variables for each signal category is given in Appendix A.3
The dataset used to train and test the multivariate analysis (MVA) classifier consists of reconstructed events corresponding to $1\,\mathrm{ab}^{-1}$ of $\mathrm{q\overline{q}}$ and $\mathrm{B\overline{B}}$ background in total. For the signal sample, reconstructed events equivalent to 50 million generated samples are used. To train the MVA classifier, the whole dataset is shuffled and a total of 70 % is retained for training purposes. The remaining 30 % is exploited in the testing procedure.
As in many cases of classifying tasks, the input dataset should be balanced in the sense of equally representing the categories to classify for. This ensures preventing the model from largely learning patterns affiliated to only one data type and thus inducing a bias. In this

Table 5.2: Depicting the ten most frequent charged background decays falsely getting reconstructed as B_{sig}. The true signal components are in the first column, while the second describes the overall share in percent. The tool to extract these ratios is developed with the AI programming assistant GitHub Copilot.

Signal B Daughters	Fraction in %
$\overline{D}^{*0}\,\mu^+\,\nu_\mu$	12.18
$\overline{D}^{*0}\,e^+\,\nu_e$	10.59
$\overline{D}^{*0}\,e^+\,\nu_e\,\gamma$	5.45
$\overline{D}^{0}\,\mu^+\,\nu_\mu$	4.91
$\overline{D}^{0}\,e^+\,\nu_e$	4.38
$\overline{D}^{0}\,e^+\,\nu_e\,\gamma$	2.22
$\overline{D}^{*0}\,\tau^+\,\nu_\tau$	2.22
$\overline{D}^{0}\,\rho^+$	2.21
$\overline{D}^{*0}\,\mu^+\,\nu_\mu\,\gamma$	1.84
$\overline{D}^{*0}\,D_{s0}^+$	1.84
...	...

case, the background sample is significantly larger than the signal one in each individual category. As a consequence, a weight w_i is assigned to each data type

$$w_i = \frac{N_B}{N_i} \tag{5.14}$$

to account for the unequal number of events. The index i is describing the data type ($i = S, B$) whether it is signal or background and N_i the corresponding number of events. After an initial training, each input variable gets evaluated on their feature importance. Variables are deemed not important if the associated importance score is <1% of the summarized importance. Quantities that fulfill these criteria are removed for the sake of simplicity since their contributions are negligible. With the important features collected, a grid search is performed, testing out a variety of different hyperparameter settings. The best model is then defined by having the highest AUC_{test} score with the additional threshold of $(AUC_{test} - AUC_{train})\,/AUC_{test} < 1\%$, ensuring consistent performance on train and test set and to prevent overfitting. Furthermore, within the model evaluation, a k-Fold cross validation is performed: The entire training datasets is split into k subsets, training and testing models over each instance. The resulting scores should not differ by 5 % to deliver consistent and reliable predictions. After these steps, the best model is used and saved to

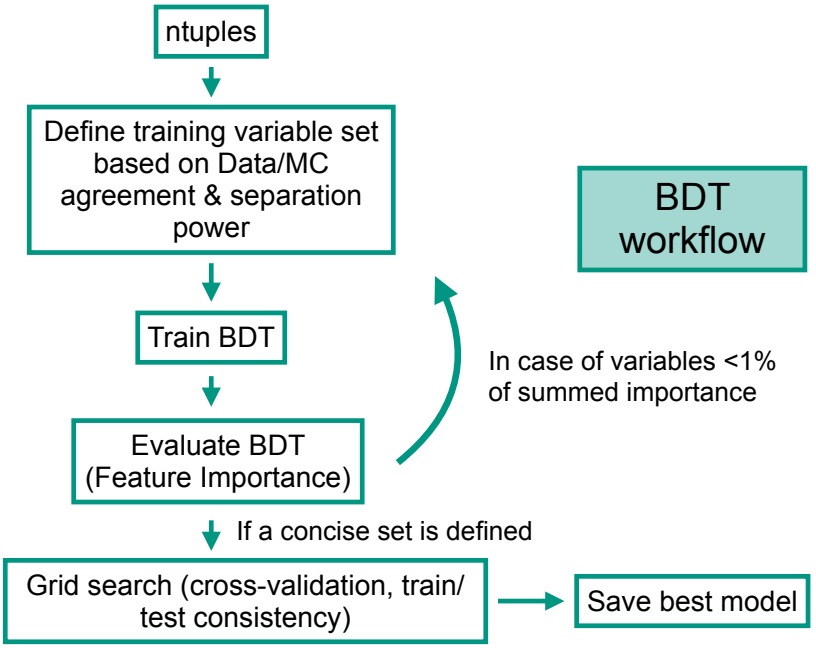


Figure 5.4: Depicting the workflow of choosing the best working point for the classifier.

ensure reproducibility. The described workflow for finding the best classifier is summarized in Figure 5.4. Evaluation metrics and the output of the classifier are illustrated in Figure 5.5 and 5.6 respectively.

A key task of the BDT classifier is background suppression by defining a signal region with the highest sensitivity. To optimize for sensitivity, a Punzi Figure of Merit (FoM) [69] is used

$$\mathrm{FoM} = \frac{\varepsilon}{3/2 + \sqrt{N_B}}, \tag{5.15}$$

with a desired significance of $3\,\sigma$. Here ε denotes the signal efficiency and N_B the total background yield at a specific threshold for the BDT output. The FoM is calculated for 50 equidistant instances of the BDT output ranging from 0 to 1, and the BDT output value producing the highest sensitivity is used to define the signal region (see Figure 5.7 and Figure 5.8).

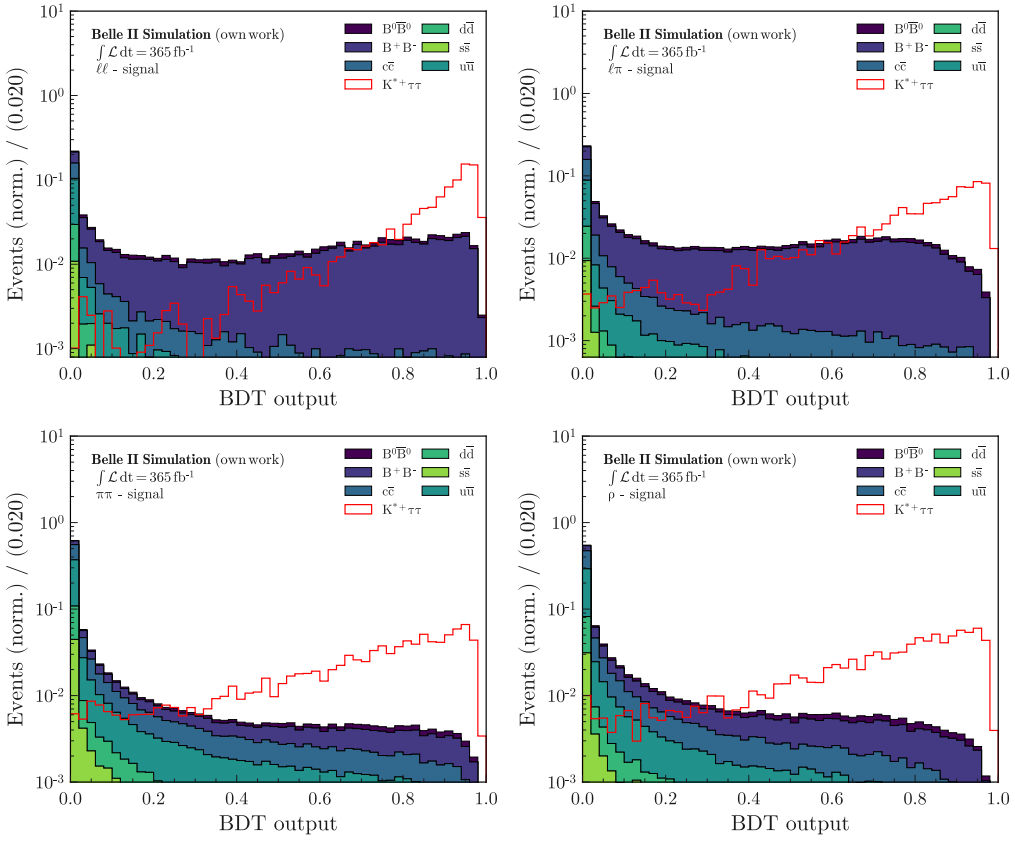


Figure 5.5: Representation of the BDT classifier output for all four signal channels on simulated signal and background samples. The different stacked components in the histogram refer to each background component, while the red histogram denotes the signal component. For visualization purposes, both distributions are normalized to unity.

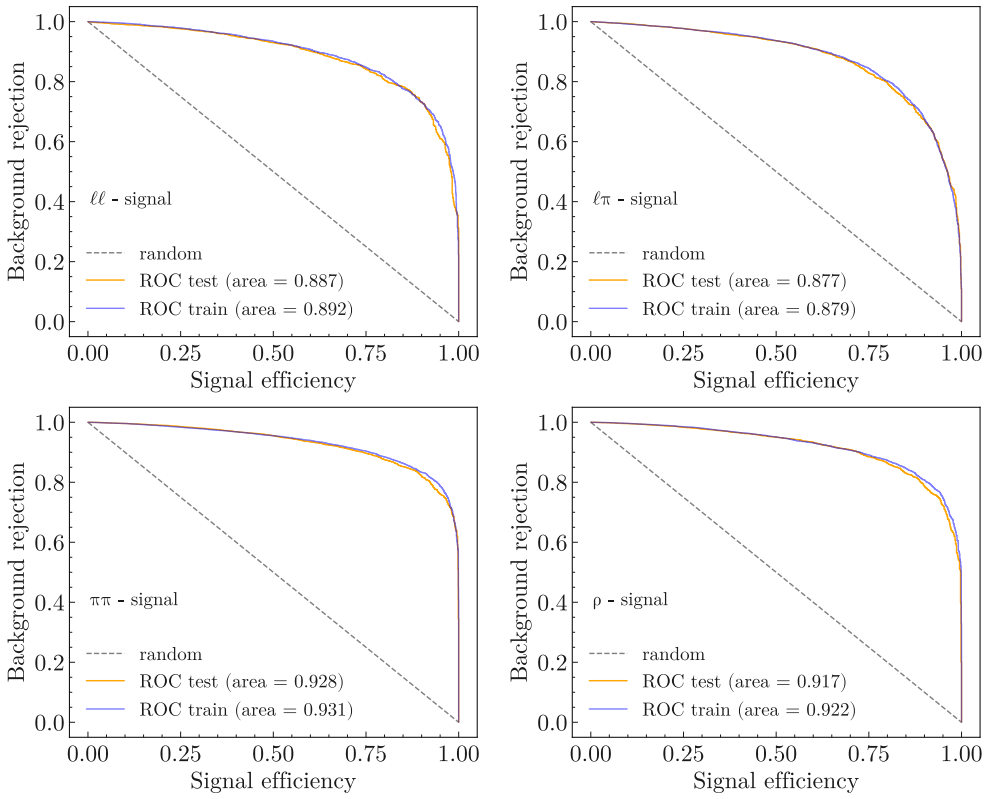


Figure 5.6: Illustrating the background rejection depending on the signal efficiency for the trained MVA classifier in each signal category. The receiver operating characteristics (ROC) curve is depicted in blue, while the ROC curve for the test set is displayed in orange. In each instance, a dashed line in gray indicates the outcome of a random classifier.

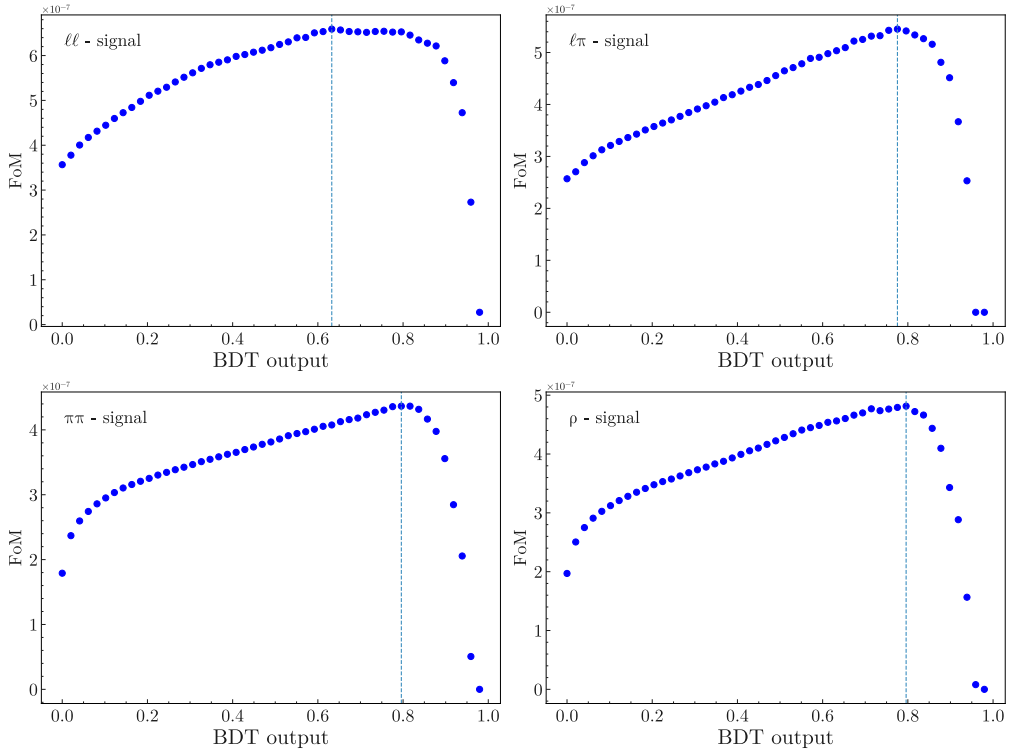


Figure 5.7: Illustration of the FoM distribution depending on the BDT output value for all four signal channels. The dashed light blue line indicates the point with the highest sensitivity and is therefore being used to define the signal region.

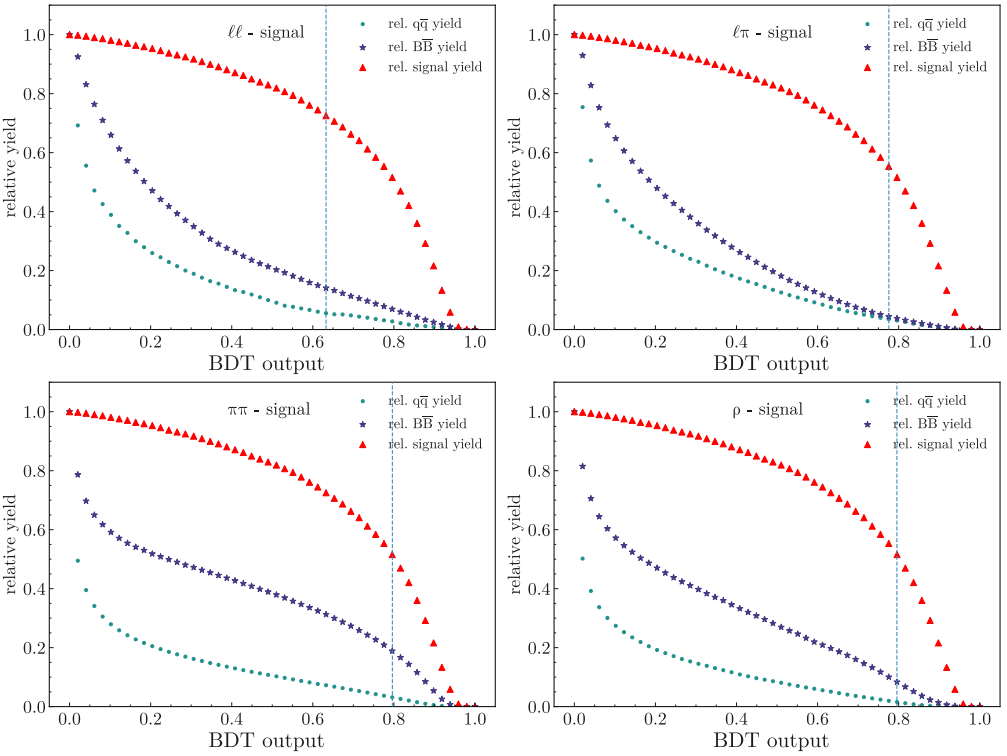


Figure 5.8: Illustration of the relative signal yield (red triangle), relative $B\overline{B}$ yield (purple star) and relative $q\overline{q}$ yield (turquoise dot) depending on the BDT output for all four signal channels. The dashed light blue line indicates the point with the highest sensitivity based on the FoM.

5.6 MC Corrections

Reconstructing events and applying selection criteria rely heavily on the employed algorithms in Belle II. The algorithms' behavior on simulated samples does not resemble the performance on real measured data in any case. To compensate for the different performance and reduce the Data/MC discrepancy, several corrections must be derived and applied.

5.6.1 FEI Tagging Efficiency Correction

The FEI tagging algorithm applied to reconstruct the tag-side B meson shows different performances on simulated and recorded events. It is possible to apply a global calibration factor ε_{FEI} to resemble the taken data with better precision. However, a more sophisticated approach is to calibrate each FEI reconstruction channel, putting more emphasis on the individual topologies. The calibration factors are then defined for each channel indexed i by

$$\varepsilon_{\text{FEI},i} = \frac{N_{\text{MC},i}}{N_{\text{Data},i}} \,. \tag{5.16}$$

To calculate each factor $\varepsilon_{\text{FEI},i}$, two independent analyses with different methods are conducted. The first one is concerning inclusive $\text{B} \to \text{X}\ell\nu$ decays performed in Ref. [70] while the second analysis examining partially reconstructed $\text{B} \to \text{D}^{(*)}\pi$ decays described in Ref. [71]. In both cases the yield of the calibration mode in data (i.e. the denominator in Equation (5.16)) is determined via fitting to a discriminating variable. To combine both calibration studies, a χ^2 parameter estimation is performed, described in detail in Ref. [72]. Resulting calibration factors from the fit are then extracted with their corresponding uncertainties, illustrated in Figure 5.9. Each reconstructed decay mode by the FEI algorithm is labeled with a unique identifier. Depending on the B_{tag} decay mode, the corresponding calibration factor is allocated to the event.

5.6.2 Hadron Efficiency and Misidentification

The reconstruction efficiency and the misidentification rate of charged kaons and pions are evaluated on specific control channels. For the prior case, the decay chain $\text{D}^{*+} \to [\text{D}^0 \to \text{K}^-\pi^+]\,\pi^+$ is exploited, while the $\text{K}^0_\text{S} \to \pi^+\pi^-$ decay is used for the latter. The difference in efficiency and misidentification rate between recorded and simulated data is determined by a straightforward comparison. Illustrated in Figure 5.10 are the resulting calibration factors for K^+ and π^+ identification, binned in terms of θ and the corresponding momentum p. A K^+ gets misidentified as a π^+ in $\sim 7\%$ of all cases which makes it far more occurring than vice versa ($< 1.5\%$). In this analysis, charged kaon and pion efficiency corrections are applied to the K^{*+} daughters as well as the hadronic τ-lepton modes. Misidentification rate corrections are solely applied for fake pions due to their relatively large occurrence.

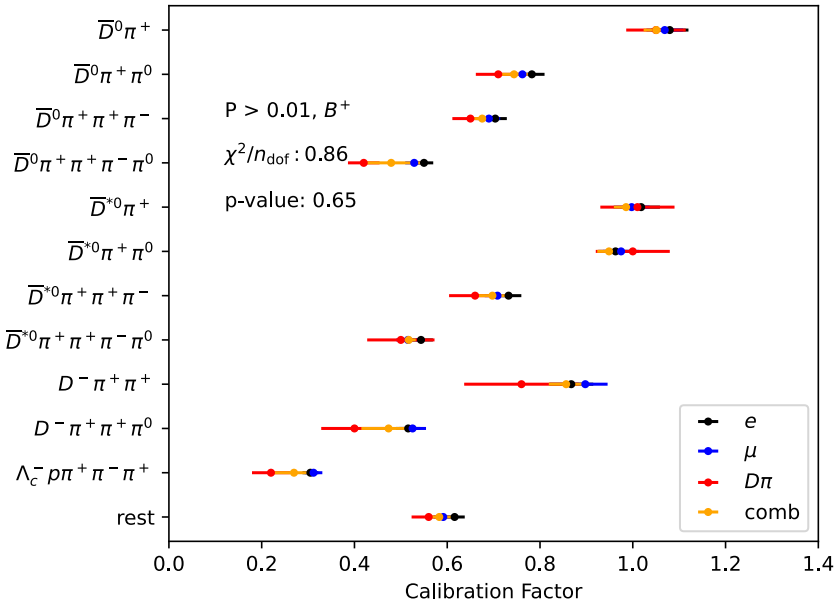


Figure 5.9: Representation of the FEI calibration factors and their uncertainties extracted from a combined χ^2 fit of $B \to X\ell\nu$ and partially reconstructed $B \to D^{(*)}\pi$ decays. The individual calibration factors are depicted in blue for μ, in black for e and red for the $D\pi$ mode. The combined result is displayed in orange. Taken from Ref. [72].

Correcting fake kaons is deemed not necessary due to the vanishing amount of observed cases. Applying weights for this induces (larger) systematics that would propagate into the final result.

For neutral π^0 candidates, the less well-defined nature of photon candidates reconstructed from electromagnetic clusters requires a proper π^0 efficiency correction. Two measurements are performed to cover a wide range of the π^0 phase space exploiting $D^0 \to K^-\pi^+\pi^0$, $D^0 \to K^-\pi^+$ and $B^+ \to D^{*0}\,[D^{*0} \to D^0(\to K^-\pi^+)\,\pi^0]\,\pi^+$ modes in Ref. [73].

A tremendous challenge in this regard is the correction for pions with a momentum below $p = 0.2\,\text{GeV}/c$ which occur in $5 - 10\%$ of events where at least one neutral pion is present. In this regard, the resulting photon pair is likely to escape detection in the ECL based on the selection criteria described in Section 5.2. Even if both photons are detected, the photon resolution is insufficient due to smearing effects in the calorimeter.

The derived corrections in Ref. [73] are used to reweight the ρ signal channel, as well as occurring $K^{*+} \to K^+\pi^0$ modes in bins of π^0 momentum from $0.2\,\text{GeV}/c$ onwards. For softer pions, no further weighting is considered due to the lack of proper corrections.

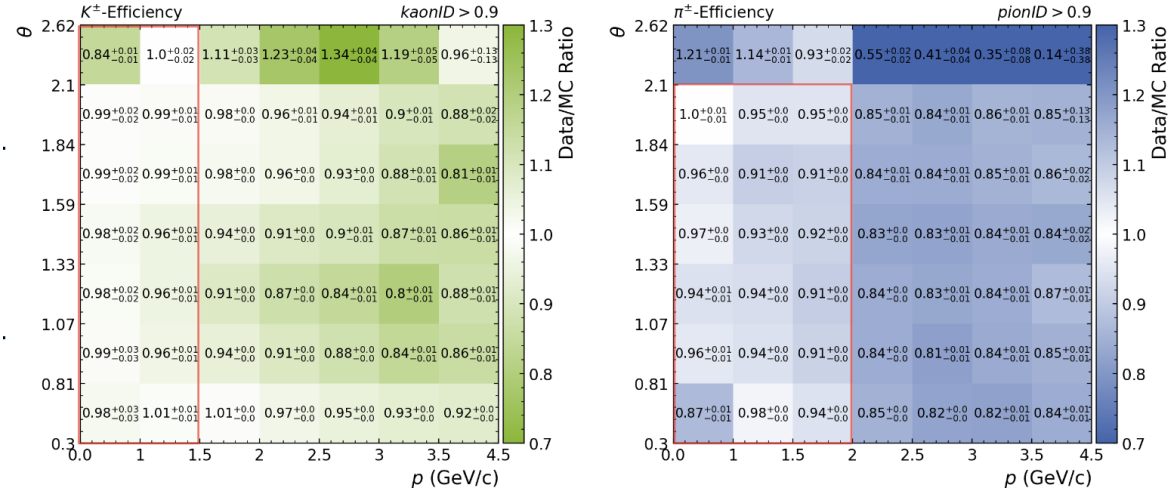


Figure 5.10: Depiction of the derived K^+ (left) and π^+ (right) corrections exploiting $D^{*+} \to \left[D^0 \to K^-\pi^+\right]\pi^+$ and $K^0_S \to \pi^+\pi^-$ decays respectively. A red box indicates where $> 99\%$ of all charged kaons and pions are located. Produced and illustrated by P. Ecker[1].

5.6.3 Lepton Efficiency Corrections

As described in Section 5.2, the final state lepton identification uses a BDT-based approach. For both electrons and muons, this identification method induces a discrepancy in efficiency and misidentification rates. In order to provide corrections for this behavior that cover a large amount of the lepton phase space, the interactions $J/\psi \to \ell^+\ell^-$, $e^+e^- \to \ell^+\ell^-(\gamma)$ and $e^+e^- \to e^+e^-\ell^+\ell^-$ are considered [74]. The efficiency corrections are applied to each identified lepton on the signal side. Misidentification of a generated charged hadron as a lepton is strongly suppressed resulting from the tight selection criteria ($\text{eID}_{\text{BDT}} < 0.9$) and thus happens only in $< 1\%$ of all cases. Even rarer occurs a misidentification based on lepton flavours which is present in $< 0.5\%$ of all events. No corrections are applied in both cases due to the rare occurrence.

5.6.4 Calibrating Continuum MC

As a large background component and module in the template fit, an accurate description of continuum background is crucial for obtaining a precise result on the signal yield. Calibrating the continuum background is a complicated task and depends heavily on the enforced selection criteria. This makes it difficult to provide central calibration and thus has to be done for each analysis individually.
Running at a lower CMS energy, several defined quantities are deviating from their $\Upsilon(4S)$

1 Electronic address: `patrick.ecker@kit.edu`

energy counterparts: The beam-constrained mass defined in Equation (5.1) depends on beam-related quantities. In order to translate the selection criteria given in Section 5.2 to the off-resonance case, one has to take the shift in beam energy into account, yielding

$$M_{\mathrm{bc}}^{\mathrm{off}} = \sqrt{\left(\frac{E_{\mathrm{CMS}}^{\mathrm{on,\,beam}}}{c^2}\right)^2 - \left(\frac{E_{\mathrm{CMS}}^{\mathrm{on,\,beam}}}{E_{\mathrm{CMS}}^{\mathrm{off,\,beam}}}\frac{\mathbf{p}_{\mathrm{CMS,B}}}{c}\right)^2} \,. \tag{5.17}$$

Here $E_{\mathrm{CMS}}^{\mathrm{on,\,beam}}$ and $E_{\mathrm{CMS}}^{\mathrm{off,\,beam}}$ denote the beam energy in the CMS in the on- or off-resonance case respectively. Compared to the quantity defined in the case of operating at $\Upsilon(4\mathrm{S})$ energy, the momentum of the B meson obtains an additional factor compensating the energy differences. For the same reason, the difference in energy ΔE defined in Equation (5.2) reads for the off-resonance case

$$\Delta E^{\mathrm{off}} = \frac{E_{\mathrm{CMS}}^{\mathrm{on,\,beam}}}{E_{\mathrm{CMS}}^{\mathrm{off,\,beam}}} E_{\mathrm{CMS,B}} - E_{\mathrm{CMS}}^{\mathrm{beam}} \,. \tag{5.18}$$

The cross-section for $e^+e^- \to \mathrm{q\overline{q}}$ interactions is depending on the initial energy of the electron pair and hence corrections have to be made in this domain. Given the energy fraction δE defined by

$$\delta E = 1 - \frac{E_{\mathrm{CMS}}^{\mathrm{off,\,beam}}}{E_{\mathrm{CMS}}^{\mathrm{on,\,beam}}} \simeq 0.6\,\% \,, \tag{5.19}$$

the cross-section is decreasing by a relative factor of $(1-\delta E)^2 \simeq 0.989$ in first order QED. For the specific case of calibration, the recorded off-resonance dataset describing $\mathrm{q\overline{q}}$ events is being used, corresponding to $43\,\mathrm{fb}^{-1}$. Regarding the simulation, two datasets are in principle suitable for deriving the calibration factors: The on-resonance continuum sample with $1\,\mathrm{ab}^{-1}$ including all energy-depending refinements, or the dedicated off-resonance sample with a size $50\,\mathrm{fb}^{-1}$. Both approaches lead to a similar result, but distinguish themselves from one another. While the former contains far more $\mathrm{q\overline{q}}$ events and therefore results in a smaller statistical error, using the dedicated sample is a much more convenient approach and proves to be indifferent against errors made in energy-calibrated quantities. Uncertainty on the quantities is given in each case solely by the statistical error originating from the underlying Poisson distribution. This is justified as contributions derived from systematic sources are presumed to have a negligible impact. The derived calibration factors along with their uncertainties for all four signal channels from both approaches are illustrated in Figure 5.11.

In direct comparison, both datasets provide a similar calibration, being nearly identical in the $\ell\pi$, $\pi\pi$ and ρ signal channel. For the $\ell\ell$ channel, the difference is significantly larger, but both results are still fairly overlapping when taking uncertainty into account. For the

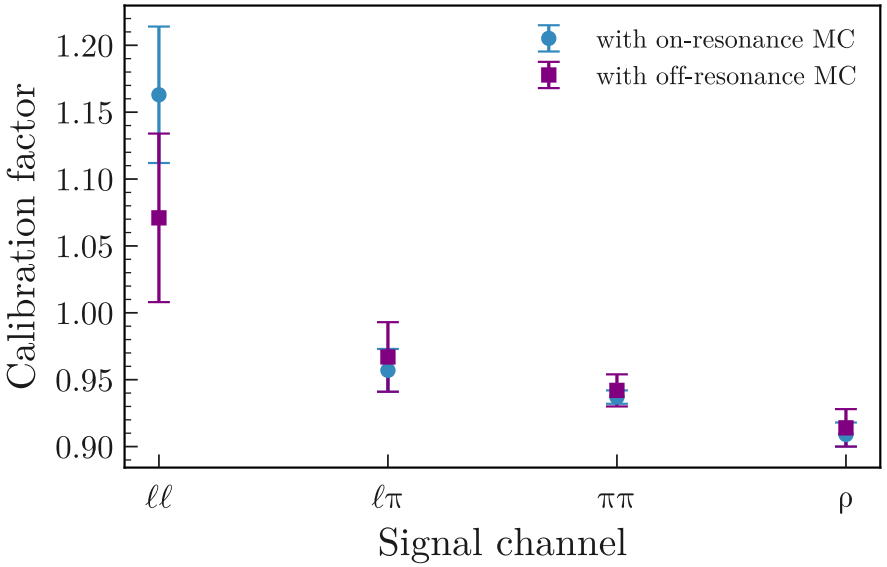


Figure 5.11: Calculated calibration factors from on-resonance MC samples (blue) and off-resonance MC samples (purple), used to model the recorded off-resonance spectrum.

Table 5.3: Depicting the derived calibration factors from the off-resonance MC samples for continuum background being used in the reconstruction. The uncertainty only consists of the statistical error modeled by a Poisson distribution.

Signal channel	Calibration factor
$\ell\ell$	1.071 ± 0.063
$\ell\pi$	0.967 ± 0.026
$\pi\pi$	0.942 ± 0.012
ρ	0.914 ± 0.014

sake of convenience and preciseness, the calibration factors used in this analysis stem from using off-resonance samples. These derived quantities are represented in Table 5.3 for each channel individually.

With this being the final correction of the simulated reconstruction, one is now able to compare the simulated template with data. To prevent operating in a signal-sensitive region, I mainly use the K^{*+} mass sideband defined in Section 5.2 to test the reconstruction algorithm as well as the applied corrections stated in this section. A special domain for testing provides the 1% dataset, where I evaluate the reconstruction in the signal region on a low amount of statistics (1% of the full $365\,\mathrm{fb}^{-1}$).

5.7 Resulting Simulated and Recorded Samples

The reconstruction efficiency as well as background yielding from $q\overline{q}$ and $B\overline{B}$ events are given in Table 5.4. All signal channels show similar efficiencies in reconstruction. The differences in background yield arise from dissimilar decay topologies and selection criteria in each mode. The final states with hadrons present are especially prone to $q\overline{q}$ background caused by misinterpretation of the hadronization. The ratio of background yield and efficiency remains almost stable for each instance, hence the resulting signal sensitivity is expected to be consistent.

Table 5.4: Resulting signal efficiency and background yield for each signal channel after BDT selection. The background yield is scaled to an integrated luminosity of $365\,\mathrm{fb}^{-1}$ to match the recorded dataset.

Signal category	$\varepsilon\ (B^+ \to K^{*+}\tau^+\tau^-)$	$q\overline{q}$ background	$B\overline{B}$ background
$\ell\ell$	1.521×10^{-5}	20.529	284.531
$\pi\ell$	1.617×10^{-5}	91.192	517.911
$\pi\pi$	1.908×10^{-5}	458.289	958.347
ρ	1.227×10^{-5}	120.772	345.528
combined	6.275×10^{-5}	690.782	2106.317

6 Data/MC Comparison

With the procedure of event reconstruction completed and all corrections applied to the simulated samples, the focus is now set on the validation based on recorded data. This is a crucial step in every analysis to compare the reconstructed events from simulation to data and spot potential discrepancies that bias the result in the fitting procedure. Especially the input variables of the BDT should properly resemble the measured data to make separating backgrounds from signal as reliable as possible. To achieve this, I use multiple datasets containing recorded data.
First, the derived calibration factors for the continuum background given in Section 5.6.4 are tested on the off-resonance dataset. Then, a validation is performed on the K^{*+} mass sideband defined in Section 5.2. This exploits the full Run 1 dataset corresponding to $365\,\mathrm{fb}^{-1}$ and therefore contains the most statistics. The actual validation is conducted in one dimension (1D), as well as in two dimensions (2D). The former approach already gives a precise insight into the Data/MC agreement, but the latter makes it possible to observe unwanted correlation patterns. With this dataset, the modeling of the $B\overline{B}$ background is additionally evaluated.
The chapter concludes with analyzing 1% of the full Run 1 dataset in the signal region, being used for testing the reconstruction on on-resonance data without operating in a signal-sensitive area. This shifts the focus more towards the actual signal region and the dataset later being used to extract the signal.

6.1 Continuum Background Validation

As stated in Section 5.6.4, a sufficient calibration is key to properly describe the underlying $q\overline{q}$ background. The calculated calibration is applied to each $q\overline{q}$ event individually and then used for reweighting purposes. As an example to analyze the result of the calibration process, the variable $\cos\left(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}}\right)$ is considered. This quantity is used in every BDT utilized in this analysis as an input feature, hence proper modeling of continuum background is essential. The distribution for data and the reweighted continuum simulation is illustrated in Figure 6.1. Due to the low amount of statistics in the $\ell\ell$ channel, no

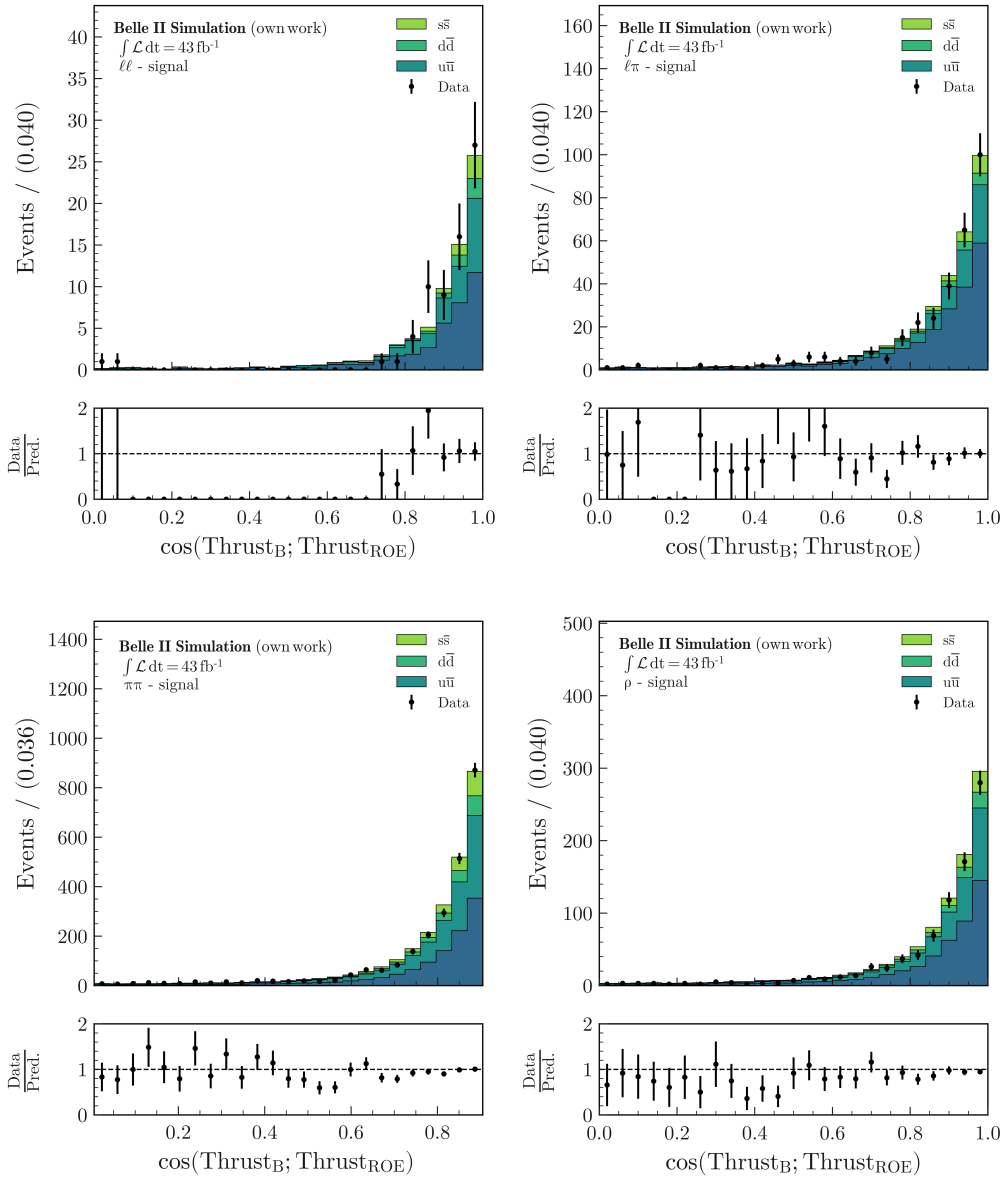


Figure 6.1: Comparison of the variable $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ for simulated continuum samples and off-resonance data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties on Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

events are observed in the spectrum between 0.1 and 0.7, distorting the ratio of data to simulated samples. As mentioned in Section 5.4, this area is usually sparsely crowded with continuum-like events, supporting this observation. However, one observes a good agreement in the densely populated area as well as in the other three channels. For further validation, examining the sideband distributions is mandatory.

6.2 Sideband Validation

By exploiting the full Run 1 data, statistical fluctuations are much more suppressed. Compared to validating continuum events, the description of $B\overline{B}$ background events is tested in addition. Both properties make the sideband validation a compelling tool in analyzing Data/MC agreement.
To further validate the input features of the BDTs, multiple distributions are taken into account: The 1D distributions are of major importance for each variable. A sufficient quantity is expected to not exhibit large deviations on data compared to simulation, especially in high-density regions. If this property is secured, 2D distributions are considered to check for large mismodeling of correlations with other training variables.

6.2.1 One-Dimensional Validation

To continue the example given in Section 6.1, a validation of $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ is necessary.
A common indicator for good agreement is the ratio of $N_{\mathrm{Data}}/N_{\mathrm{MC}}$ for each bin as in Figure 6.1. Due to the increasing statistics and decreasing statistical uncertainty, asymmetry becomes a rising factor. The asymmetry of the ratio distribution ultimately leads a skewed behaviour, making it necessary to introduce a new indicator. The Pull, defined by

$$\mathrm{Pull} = \frac{N_{\mathrm{Data}} - N_{\mathrm{MC}}}{\sqrt{\sigma^2_{\mathrm{Data}} + \sigma^2_{\mathrm{MC}}}} \tag{6.1}$$

is used, lifting the asymmetry. Here, N_{Data} and N_{MC} are referring to the counts of recorded data and MC events in each bin. The denominator describes the propagated uncertainty stemming from data and simulation. With the Pull defined, further investigations on $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ are possible. The distribution of data and simulation is pictured in Figure 6.2. No severe tendencies in the pull distribution are observed in all four signal channels. This is important since discrepancies arising at this stage will propagate into the 2D studies and lead to poorly modeled correlations between the variables. To mitigate this issue, one typically conducts a deeper investigation to identify the discrepancies source and

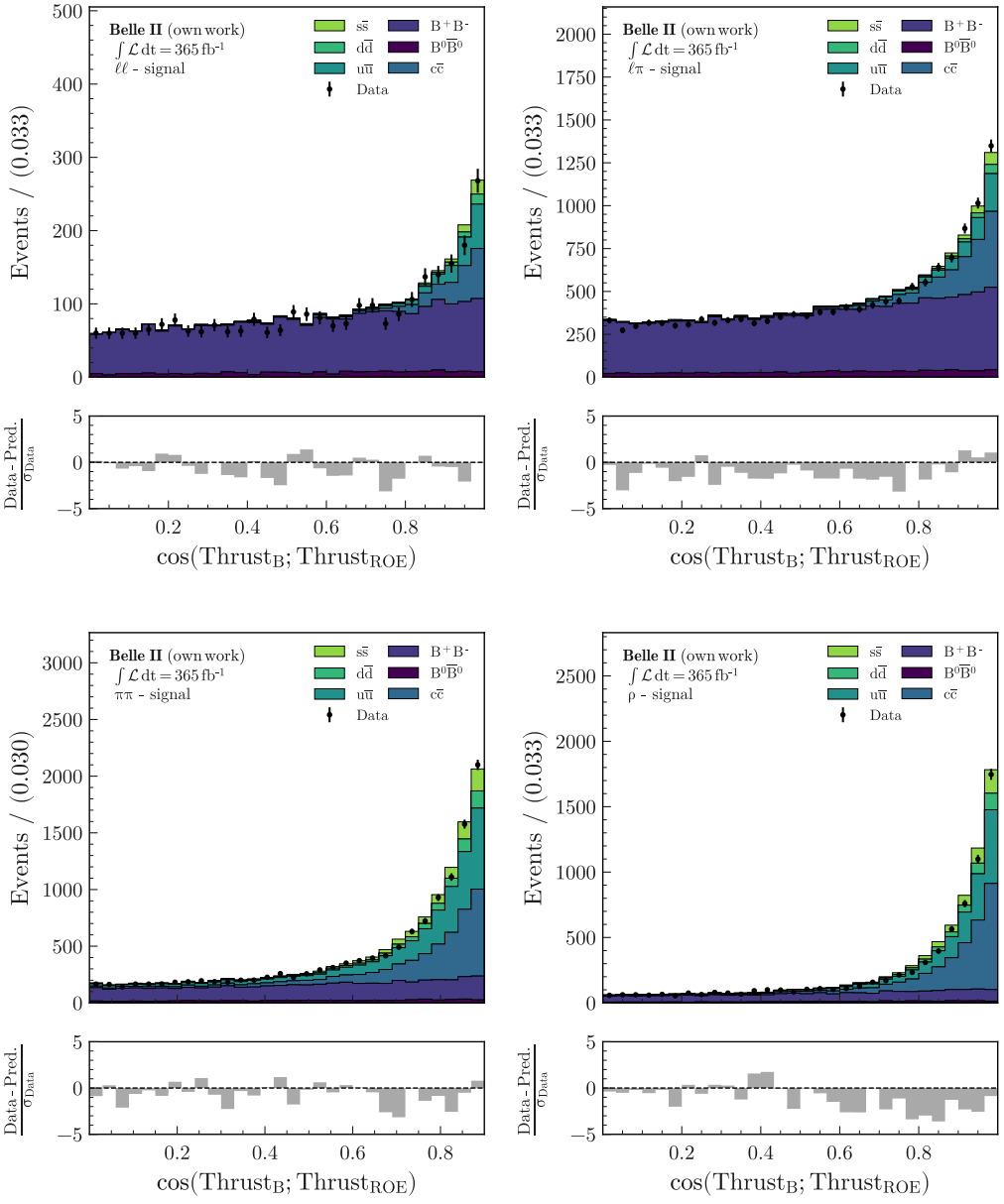


Figure 6.2: Comparison of the variable $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties on Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

potentially apply additional selection criteria to exclude those regions. The full collection of 1D validation plots is depicted in Appendix A.4.

6.2.2 Two-Dimensional Validation

A second major indicator for good Data/MC agreement is the modeling of the correlation of the input features of the BDTs. If an initial correlation is present, this will ultimately propagate through the BDT, leading to large discrepancies. To prevent this, checking the 2D distributions is obligatory.
As an indicator, the Data/MC ratio as well as the Pull is studied. The former is later being used in the template fit in a reweighting campaign and thus is of high interest. Concerning the latter, Pulls are again used to ensure a good agreement and spot a potentially mismodelled area.
As an example, the 2D Histograms of $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ and $\cos(\mathrm{Thrust_B}; z)$ are shown in Figure 6.3 and 6.4 for all four signal channels. In each instance, one is able to identify the event-dense region as a ribbon in the last two bins of $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$. Concerning the $\ell\ell$ and $\ell\pi$ channels, these ribbons are smeared out in comparison to the $\pi\pi$ and ρ channel, stemming from the reduced amount of continuum background. The 2D Data/MC ratio distribution offers a more precise insight into the modeling: In sparsely populated regions, the ratio tends towards a higher or lower value than one. This translates to the simulation either over- or underestimating the amount of recorded data in this domain. Moving to the high-density region, ratios are close to one indicating a good description. For further investigations the Pull is taken into consideration: Pull bins with no or small tendency are observed in high-density regions, supporting the results from the ratio distribution. Some outliers are observed, potentially caused by binning effects and imminent statistical fluctuations.
To summarize, the given example shows good data agreement with the simulation in this well-modeled case. Further checks of correlations with other training variables are obligatory in order to be a reliable quantity. One finds a reasonable good agreement in the majority of cases, where only a few combinations lead to observable discrepancies in areas with a large event count. Given a set of N training variables for a single signal channel, $N(N-1)/2$ combinations require evaluation. Illustrating all combinations for all four signal channels is not feasible within this thesis. However, a selection of 2D validation plots is depicted in Appendix A.5 while the entire amount has been published as a Zenodo record (see Ref. [75]).

6.3 Validation Using 1% On-Resonance Data

So far, the validation procedure solely relies on control samples instead of the actual signal region. Control samples serve as a safe region to test and evaluate the reconstruction and

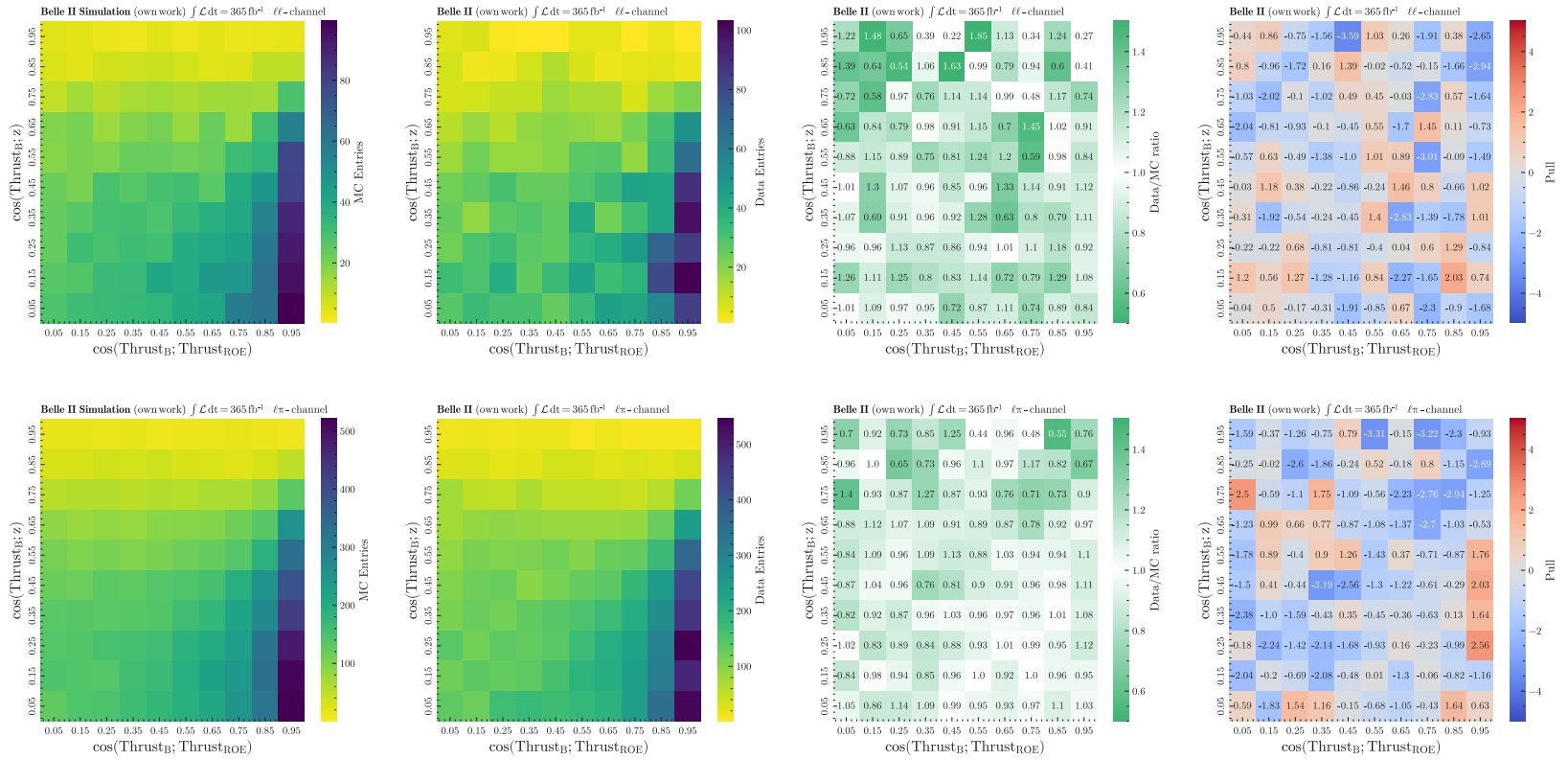


Figure 6.3: Two-dimensional comparison of recorded data and simulation in the variables $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ and $\cos(\mathrm{Thrust_B}; z)$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

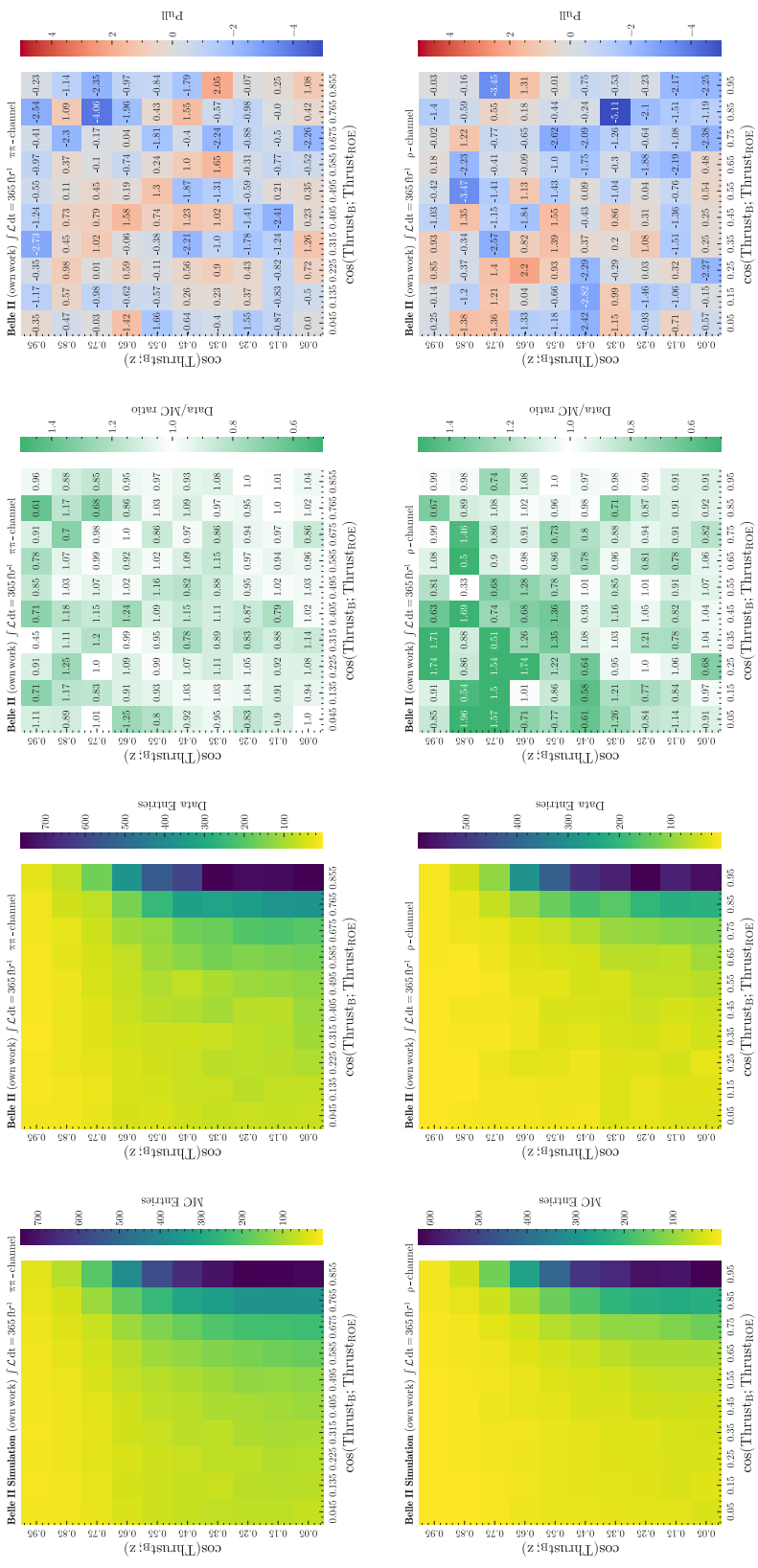


Figure 6.4: Two-dimensional comparison of recorded data and simulation in the variables $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ and $\cos(\mathrm{Thrust_B}; z)$ on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of Data, MC, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

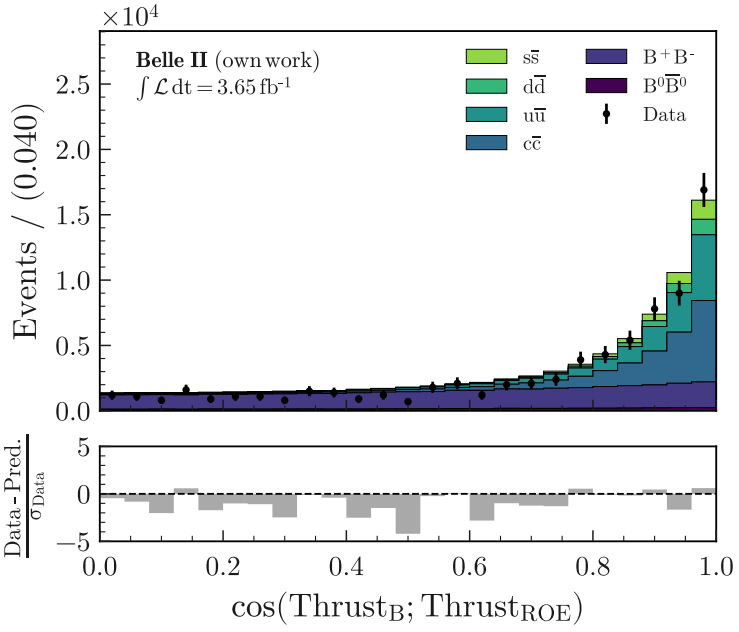


Figure 6.5: Comparison of the variable $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ on 1% of the full Run 1 dataset with simulated events for all four signal channels combined. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

correction procedure safely. However, with the signal extraction in mind, it is far more important to model the signal region properly with signal extraction in mind. Exploring this region at this stage is strongly discouraged. This would potentially induce a bias where the analyst would adapt to the outcome of the work.

It is possible to still evaluate the modeling of the signal region, without conducting studies in a signal-sensitive region. To perform studies, one takes a fraction of the full Run 1 dataset. In this work, 1% is being used, corresponding to an effective integrated luminosity of $3.65\,\mathrm{fb}^{-1}$. For this purpose, no dedicated recorded sample is available. In order to obtain the 1%, the full dataset is used and each event gets assigned a uniformly distributed parameter $O \in [0, 1]$. Before starting the reconstruction procedure, every event must surpass a selection criterion of $O < 0.01$, effectively removing the abundance of 99%. In principle, this procedure does not lead to a dataset exactly representing 1% given the random component. However, statistical fluctuations are deemed negligible here.

Once more the variable $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ is used as an example. For evaluation, all four signal channels are combined, due to the small size of the dataset. The resulting comparison is then just an average over all modes, making the identification of potential discrepancies in an individual channel a complex task. However, especially in the $\ell\ell$ channel,

this challenge would remain given the large statistical uncertainty. The result is shown in Figure 6.5. Again, this variable shows a sufficient agreement in the entire phase space, undermining the comparison results from the previous iterations.
With all validation tools now defined, the remaining training variables are evaluated on their agreement with recorded data. For this, the one- and two-dimensional representations of the K^{*+} mass sideband are examined. Furthermore, distributions on the 1% dataset are studied, providing a unique opportunity to test the reconstruction procedure safely within the signal window.

7 Signal Extraction Strategy

The events surpassing the reconstruction procedure are connected to the quantity of interest: The branching fraction of $B^+ \to K^{*+}\tau^+\tau^-$. This chapter presents two methods of estimating the desired quantity. The first approach uses the remaining simulated samples given in Section 5.7 to calculate a first sensitivity limit.
A more sophisticated approach is the estimation of the number of signal events in recorded data by fitting. This procedure is not yet implemented at the current state of this analysis. However, I describe the technique in detail, as well as the planned procedure to obtain and validate the results.

7.1 Counting Experiment

To get a first estimate of the sensitivity in an early stage of the analysis, a counting experiment is conducted. Starting with the background estimate given in Section 5.7, the number of signal events required for a $3\,\sigma$ excess is calculated. This is achieved by utilizing a Poisson distribution with an expectation value ν corresponding to the background hypothesis. With an average background estimate of N_B, an additional N_S events need to be observed to surpass the $3\,\sigma$ threshold.
Based on this result, a toy experiment is conducted, consisting of 1 million simulated observations. Each gets assigned a probability stemming from the underlying distribution. An interval search is performed to determine the one-sided confidence interval. The required confidence level (C.L.) is set to be at 90%, referring to the acceptance region. With this, an expectation value of signal events is estimated, which translates to a branching fraction (see Chapter 8).

7.2 Template Likelihood Fit

Although not explicitly accomplished in this work, fitting the recorded data to a template consisting of the simulated background components delivers a far more robust result. In this section, I describe the current fitting strategy and explain the statistical foundations.

7.2.1 Maximum Likelihood Method

The following is a short introduction to the maximum likelihood technique crucial for understanding parameter estimation. For a more detailed description, the reader is referred to Ref. [76].

Given is a random variable x, distributed according to $f(x;\boldsymbol{\theta})$ with one or more unknown parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)$. The method of maximum likelihood describes the estimation of one or multiple parameters $\boldsymbol{\theta}$ using n data samples. The likelihood function,

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}) \tag{7.1}$$

describes the joint probability density function (PDF) for the samples x_i. With the x_i treated as fixed, L solely depends on the parameter set $\boldsymbol{\theta}$.
An estimate of the true parameter $\hat{\boldsymbol{\theta}}$ can be obtained by maximizing the likelihood function by parameter variation. In general, analytic descriptions of the likelihood maxima are not always available, meaning computations usually rely on numerical methods. Rather than maximizing the likelihood function L straight away, it is instead more convenient to minimize the negative logarithm of L, referred to as "negative log-likelihood". Applying the logarithm to L converts the product to a sum, benefiting iterative algorithms used to determine the minimum.

Extended Maximum Likelihood In contrast to the previous case, the number of observations n in the sample $x_1, ..., x_n$ is often itself given by a Poisson distribution with mean value ν. Holding the result of the sample then not only consists of information about the individual observation x_i, but also of the total number n. To account for this, the likelihood function is extended with a Poisson distribution, resulting in

$$L(\nu, \boldsymbol{\theta}) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta})\,. \tag{7.2}$$

Two distinguished cases need to be considered: If there is no functional relation between ν and $\boldsymbol{\theta}$, one simply obtains the same estimators $\hat{\boldsymbol{\theta}}$ as by minimizing the logarithm of L in Equation (7.1). The other situation of interest is that ν is given as a function of $\boldsymbol{\theta}$. The resulting estimators $\hat{\boldsymbol{\theta}}$ following the extended likelihood exploit information from n and the variables x_i. As a result, $\hat{\boldsymbol{\theta}}$ typically carries a smaller variance than in cases where it exclusively relies on the x_i values.

Binned Maximum Likelihood In principle, the samples x_i do not have to follow a continuous distribution by default. Independent of the scenario, gathering data in bins that form a histogram proves to be useful. Binning the data drastically simplifies the computation of $\log f(x_i;\boldsymbol{\theta})$ since one does not have to consider each sample x_i individually. Considering the number of entries $\boldsymbol{n} = (n_1, ..., n_N)$ in N bins, the expectation value ν_i for the number of entries in each bin is given by

$$\nu_i(\boldsymbol{\theta}) = \sum_j^N n_j \times \int_{x_i^{\min}}^{x_i^{\max}} f(x_i;\boldsymbol{\theta})\, dx\,. \tag{7.3}$$

Here, $x_i^{\min}$ and $x_i^{\max}$ are indicating the bin limits. If one refers to the histogram as a single measurement of a random vector with dimension N, a joint PDF represents the underlying distribution. The log-likelihood function then yields

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^N n_i \log \nu_i(\boldsymbol{\theta})\,. \tag{7.4}$$

Considering the case where N is very large (i.e. the bin size becomes very small), the likelihood function becomes the same as in Equation 7.1. This prevents any difficulties that may occur with empty or sparsely populated bins.

7.2.2 Parameter Estimation With Templates

The determination of the branching fraction requires a precise extraction of the signal estimate. To obtain these estimate, the joint PDF is defined as the linear combination of the individual PDFs $f_i(x)$ scaled by the unknown normalization parameter set $\boldsymbol{\theta}$:

$$f(x;\boldsymbol{\theta}) = \sum_i \theta_i f_i(x)\,. \tag{7.5}$$

The key task is the estimation of the parameter set $\boldsymbol{\theta}$, giving insight into the signal estimate in the recorded data. Describing the PDFs $f_i(x)$ analytically is not feasible since it has to take into account the complex interactions inside the detector. To approximate the PDFs in each bin, approximations in the form of simulation are considered. These templates (or binned approximations) are generated from MC simulations, precisely describing the physics processes and detector properties. Since the templates are already in discrete form, the binned likelihood is described by a product of Poisson distributions $\mathcal{P}$, yielding

$$L = \prod_{i=1}^N \mathcal{P}(n_i|\nu_i(\boldsymbol{\theta}))\,. \tag{7.6}$$

With the total number of templates M, the expectation value in each bin is referred to as the sum of template contributions

$$\nu_i(\boldsymbol{\theta}) = \sum_j^M f_{ij} k_j(\theta_j) \tag{7.7}$$

where f_{ij} denotes the fraction of events in bin i for template j and k_j describes the total number of events from template j[1]. An accurate knowledge of k_j for each template leads to a precise estimate of the normalization parameter set $\boldsymbol{\theta}$.
The estimated parameter set is subject to uncertainties mainly from two sources: Derived template distributions are affected by uncertainties stemming from the limited size of the simulated sample or reconstruction effects. Furthermore, the estimators $\hat{\boldsymbol{\theta}}$ obtained from maximizing the likelihood function are carrying additional uncertainties. Discussion of methods to derive the uncertainties is beyond the scope of this work, but the interested reader is referred to Ref. [77] or Ref. [78].

7.2.3 Implementation Strategy

Having defined the concepts of likelihood and template-based parameter estimation, it is now necessary to establish a procedure for the signal extraction. To perform the template fit, I consider the neutral extra energy in the ECL, as it indicates the quality of reconstruction. This quantity describes the additional energy deposits from neutral ECL clusters in the ROE. The distribution of the quantity for simulated signal and background is illustrated in Figure 7.1. For a well-reconstructed signal event, no additional energy depositions are expected. This translates to a large signal sensitivity in the first bins of the distribution, making it a crucial region for signal extraction.
A reliable signal extraction variable should possess not only a discriminative description of the event, but also an adequate Data/MC agreement, minimizing the systematic uncertainty. To ensure no potential discrepancy in data, the 1% on-resonance dataset is considered. The comparison between the simulated background template and the recorded data is depicted in Figure 7.2, showing overall good agreement.

7.2.4 Validation Strategy

A potential validation procedure should aim to establish the reliability and model dependence of the signal extraction strategy. To obtain a first estimate of the signal sensitivity and fluctuations around it, one typically considers the Asimov dataset [79]. This approach can be conducted blind, i.e. without knowing the distribution in the signal region of the full

1 In general, the f_{ij} also depends on the parameters θ_j. For simplicity, the shape of the template is considered as fixed, making it independent of the estimator.

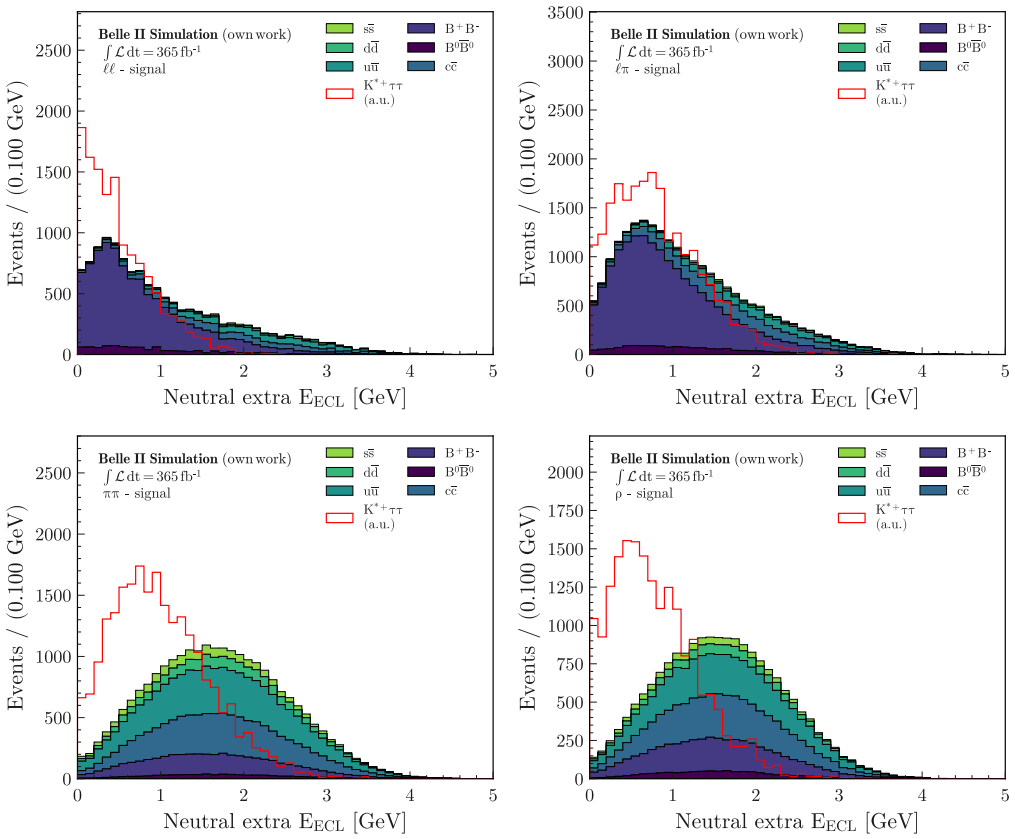


Figure 7.1: Distribution of the simulated background and signal sample for the Neutral extra energy E_{ECL} for all four signal channels. The simulation of the background is scaled to match the integrated luminosity of the recorded data. All corrections described in Section 5.6 are applied to the simulation.

dataset.

For validation, toy studies can be performed, allowing to test the reproducibility of the estimated $\hat{\boldsymbol{\theta}}$. To derive the statistical uncertainty, these toys can be sampled from a Poisson distribution $\mathcal{P}(n|\nu)$ with mean ν equal to the expected number of events in each bin of neutral extra energy E_{ECL}. By fitting the resulting toy samples, one obtains the uncertainty estimate σ_y, which can be validated using the Pull. If the fit adequately covers the data, the resulting Pull should follow a standard Gaussian distribution.

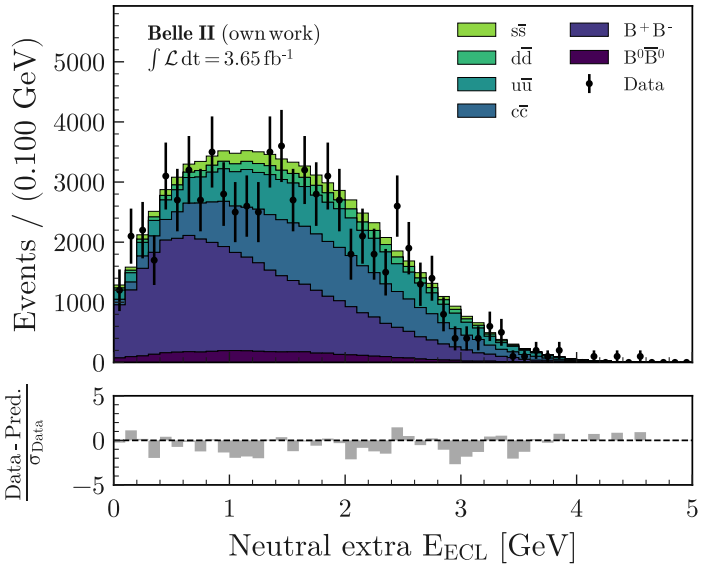


Figure 7.2: Comparison of the variable neutral extra energy E_{ECL} on 1% of the full Run 1 dataset with simulated events for all four signal channels combined. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

8 Results

To obtain an estimate of the branching fraction of $B^+ \to K^{*+}\tau^+\tau^-$, I conduct a counting experiment as described in Section 7.1.
The template based on simulated events contains the total background yield, serving as the average for the Poisson distribution. Based on Section 5.7, the background estimate results in $\nu = 2797$. The underlying Poisson distribution is now fully described, allowing the calculation of the $3\,\sigma$ threshold. To exceed it, one must measure at least 149 additional events to claim evidence. The associated probability for this observation given the background estimate is 2.51×10^{-3}.
The rate of the signal sensitivity is equivalent to the limit on signal events to observe $\nu + 149$ with a significance level of $\alpha = 0.1$. This results in an upper limit for the estimated signal component:

$$\nu_{\text{Signal}} < 218.89 \;\; @\,90\%\text{ C.L.} \tag{8.1}$$

This upper limit is further used to determine the corresponding branching fraction. Given the estimate ν_{Signal}, the calculation of the branching fraction follows Ref. [80] and calculates to

$$\mathcal{B} = \frac{\nu_{\text{Signal}}}{\varepsilon_{\text{Signal}} \cdot 2 \cdot \mathcal{B}(\Upsilon(4S) \to B^+B^-) \cdot N_{B\overline{B}}}\,, \tag{8.2}$$

where $\varepsilon_{\text{Signal}}$ denotes the combined signal efficiency for all four modes[1] and $N_{B\overline{B}}$ corresponding to the number of $B\overline{B}$ events in the entire dataset. This quantity is obtained using the cross section given by the PDG [8] and the total size of the dataset. Inserting the signal efficiency from Section 5.7 and the number of $B\overline{B}$ events represented in the full dataset, one obtains the first estimate of the signal sensitivity to be

$$\mathcal{B}(B^+ \to K^{*+}\tau^+\tau^-) < 8.38 \times 10^{-3} \;\; @\,90\%\text{ C.L.} \tag{8.3}$$

1 In other representations of this formula, the denominator consists of an additional factor representing the branching fractions of the reconstructed modes. In simulation of the signal events, the resulting K^{*+} $\tau^+\tau^-$ decays generically, leading to the factor being absorbed by $\varepsilon_{\text{Signal}}$.

The estimated branching fraction is located in the same order of magnitude as the results given by previous searches of BaBar and Belle (see Chapter 2). Compared to the latter searches, the result presented in this work does not directly rely on recorded data for signal extraction. For the definition of the signal region and validation of the MVA analysis components, detailed Data/MC studies are conducted, making this result not entirely simulation-based. However, it is still a good first estimate and the fitting procedure will help to ensure competitiveness with the ultimate goal of delivering an unprecedented sensitivity in K$\tau\tau$ modes.

9 Outlook

The branching fraction estimate from the previous chapter already serves as a good estimate of the signal sensitivity. However, much more work is required to deliver the most precise and validated measurement of the signal channel. In this chapter, I explain deeper validation strategies and uncertainty techniques, making the future of this work concise.

Combinatorial $B\overline{B}$ background Despite the applied selection criteria in Section 5.1, a misreconstructed tag-side B meson is still a key reason behind falsely determined B signal candidates. Investigating the total contribution of this category yields an additional calibration that needs to be taken into account.
For calculation, the same-flavor control sample is considered, inverting the opposite B flavor requirement at the preselection level. By construction, this sample is orthogonal to the events surpassing the standard preselection. The signal efficiency for this nonphysical sample is expected to be significantly lower, while the background should have a similar composition. Applying all corrections from Section 5.6 makes it possible to derive the calibration factor Λ for the same-flavor control sample, resulting in

$$\Lambda = \frac{N_{\text{data}} - N_{\text{MC}}^{\text{q}\overline{\text{q}}} - N_{\text{MC}}^{\text{peak}}}{N_{\text{MC}}^{\text{comb}}} \,. \tag{9.1}$$

Here, N_{data} describes the number of observed data events, $N_{\text{MC}}^{\text{q}\overline{\text{q}}}$ and $N_{\text{MC}}^{\text{peak}}$ the expected amount of continuum and peaking $B\overline{B}$ background respectively. The remaining number of events gets divided by the number of combinatorial events $N_{\text{MC}}^{\text{comb}}$ given by the same-flavor control sample. The uncertainty on Λ is entirely defined by the statistical nature of the recorded data used in the off-resonance calibration. This control sample is usually used for further validation of comparing the shapes of the BDT input variables for the simulated background.

Extra photon multiplicity The same-flavor control sample can additionally be referred to for a check on potential mismodeling of the neutral extra energy E_{ECL}. To retrieve the most precise and unbiased result from the fit, an adequate description of this variable is

essential.
If potential discrepancies occur, one typically considers the photon multiplicity showing more or less reconstructed photons than expected. Correction weights can now be obtained for each bin separately, allowing for a precise calibration of E_{ECL}. Exploring two- or multidimensional distributions can provide a more accurate description, allowing a higher validation potential.

Signal validation The reconstruction of the signal component can also suffer from Data/MC discrepancies. To investigate this, one considers a specific control sample, referred to as the embedded sample. It is based on the abundant decay $B^+ \to K^{*+}J/\psi$, exhibiting an unambiguous signature. The corresponding branching fraction is given as $\mathcal{B}(B^+ \to K^{*+}J/\psi) \sim 1 \times 10^{-3}$ [8]. For the embedding procedure, one artificially removes the signal component $B^+ \to K^{*+}J/\psi$ and replaces it with simulated samples of $B^+ \to K^{*+}\tau^+\tau^-$. The ROE and tag-side components remain unaltered. Repeating the procedure on recorded data allows for calculating the Data/MC efficiency ratio for the embedded sample.
With the corrected sample, the input variables of the BDT can be evaluated. It is important to mention that some corrections cannot be applied for the embedded sample, stemming from the loss of the true particle identity during the embedding procedure.

Systematic uncertainties Investigating and understanding all sources of uncertainties is vital for determining the error boundaries in which the true result is located. In this work, only uncertainties of statistical nature are considered. Nevertheless, systematic uncertainties play a crucial role that requires a deeper understanding of individual contributions.
These originate from underlying physics approximations, reconstruction effects, fit validation, or limitations along the signal extraction procedure. The precise estimation of the individual contributions and the uncertainty propagation is the key task in this domain.

10 Conclusion

In this work, I present the first-ever estimate of the branching fraction of the process $B^+ \to K^{*+}\tau^+\tau^-$ at the Belle II experiment.
For this analysis, I consider the full Run 1 dataset at the $\Upsilon(4S)$ resonance corresponding to an integrated luminosity of $365\,\text{fb}^{-1}$.
In the interaction, at least two neutrinos are escaping the experiment without being detected. For each event, the tag-side is reconstructed solely through hadronic decay channels, allowing for pure tag-side B mesons and setting constraints on the signal B meson.
Considering background suppression, a MVA-based classifier is employed to boost the signal sensitivity. Several calibrations are performed to counterbalance discrepancies originating from algorithm inefficiencies. Each input feature is then evaluated in terms of discriminating power and Data/MC agreement. The resulting signal efficiency and background yield allow the implementation of a counting experiment. This statistical approach permits the estimation of an upper limit on the signal sensitivity resulting in

$$\mathcal{B}(B^+ \to K^{*+}\tau^+\tau^-) < 8.38 \times 10^{-3} \;\; @\,90\%\ \text{C.L.}$$

given the underlying background hypothesis. The derived upper limit is in good agreement with recent results targeting similar $b \to s\ell\ell$ modes with data stemming from predecessor experiments.
Compared to previous searches, this analysis employs the state-of-the-art FEI tagging algorithm and reconstructs all 1-prong τ decays, resulting in the coverage of a larger phase space and a significant sensitivity gain. Furthermore, two-dimensional Data/MC distributions are considered, optimizing and validating the classifiers' input features even further.
Given the rare occurrence of this interaction, exploiting a larger dataset is a huge benefit in limiting the statistical uncertainty. With the full target dataset of $50\,\text{ab}^{-1}$, the predicted upper limit on the branching fraction is expected to decrease by a factor of two compared to existing measurements even in the baseline scenario [19].
With this in mind, the Belle II experiment provides a unique environment to precisely study

these transitions. Together with the concise outlook of this work, further steps to advance the current state of this analysis are highly motivated to offer unprecedented sensitivity in this domain.

Bibliography

[1] S. L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[2] A. Salam and J. C. Ward. "Weak and electromagnetic interactions". In: *Il Nuovo Cimento* 11 (1959), pp. 568–577. DOI: 10.1007/BF02726525.

[3] S. Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[4] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". In: *Phys. Rev. D* 2 (1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.

[5] P. W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". In: *Phys. Rev.* 145 (1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.

[6] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[8] S. Navas et al. "Review of Particle Physics". In: *Phys. Rev. D* 110 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[9] R. Aaij et al. "Search for Lepton-Universality Violation in $B^+ \to K^+\ell^+\ell^-$ Decays". In: *Phys. Rev. Lett.* 122 (2019). DOI: 10.1103/PhysRevLett.122.191801.

[10] R. Aaij et al. "Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays". In: *Journal of High Energy Physics* 2017 (2017). DOI: 10.1007/jhep08(2017)055.

[11] R. Aaij et al. "Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$". In: *Phys. Rev. Lett.* 111 (2013), p. 191801. DOI: 10.1103/PhysRevLett.111.191801.

[12] R. Aaij et al. "Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays". In: *Journal of High Energy Physics* 2014 (2014). DOI: 10.1007/jhep06(2014)133.

[13] R. Aaij et al. "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using $3\,\text{fb}^{-1}$ of integrated luminosity". In: *Journal of High Energy Physics* 2016 (2016). DOI: 10.1007/jhep02(2016)104.

[14] R. Aaij et al. "Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay". In: *Phys. Rev. Lett.* 125 (2020), p. 011802. DOI: 10.1103/PhysRevLett.125.011802.

[15] R. Aaij et al. "Test of Lepton Universality in $b \to s\ell^+\ell^-$ Decays". In: *Phys. Rev. Lett.* 131 (2023), p. 051803. DOI: 10.1103/PhysRevLett.131.051803.

[16] Y. Amhis et al. "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021". In: *Phys. Rev. D* 107 (2023), p. 052008. DOI: 10.1103/PhysRevD.107.052008.

[17] B. Capdevila et al. "Searching for New Physics with $b \to s\tau^+\tau^-$ Processes". In: *Phys. Rev. Lett.* 120 (2018), p. 181802. DOI: 10.1103/PhysRevLett.120.181802.

[18] R. Alonso, B. Grinstein, and J.M. Camalich. *Lepton universality violation and lepton flavor conservation in B-meson decays*. 2015. DOI: 10.1007/jhep10(2015)184.

[19] L. Aggarwal et al. *Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond*. 2022. arXiv: 2207.06307 [hep-ex].

[20] M. Kobayashi and T. Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[21] J. P. Lees et al. "Search for $B^+ \to K^+\tau^+\tau^-$ at the BaBar Experiment". In: *Phys. Rev. Lett.* 118 (2017), p. 031802. DOI: 10.1103/PhysRevLett.118.031802.

[22] K. Abe et al. "Observation of the Decay $B \to Kl^+l^-$". In: *Phys. Rev. Lett.* 88 (2001), p. 021801. DOI: 10.1103/PhysRevLett.88.021801.

[23] R. Aaij et al. "Test of Lepton Universality Using $B^+ \to K^+\ell^+\ell^-$ Decays". In: *Phys. Rev. Lett.* 113 (2014), p. 151601. DOI: 10.1103/PhysRevLett.113.151601.

[24] S. Descotes-Genon et al. "Implications from clean observables for the binned analysis of $B \to K^*\mu^+\mu^-$ at large recoil". In: *Journal of High Energy Physics* 2013 (2013). DOI: 10.1007/jhep01(2013)048.

[25] C. Bouchard et al. "Standard Model Predictions for $B \to K\ell^+\ell^-$ with Form Factors from Lattice QCD". In: *Phys. Rev. Lett.* 111 (2013), p. 162002. DOI: 10.1103/PhysRevLett.111.162002.

[26] T. Huber et al. "Electromagnetic logarithms in $\overline{B} \to X_s \ell^+ \ell^-$". In: *Nuclear Physics B* 740 (2006), pp. 105–137. DOI: 10.1016/j.nuclphysb.2006.01.037.

[27] F. Metzner. "Preparation of a Measurement of $\mathcal{R}(D^{(*)})$ with Leptonic τ and Hadronic FEI Tag at the Belle Experiment". PhD thesis. Karlsruher Institut für Technologie (KIT), [2022]. DOI: 10.5445/IR/1000148812.

[28] M. Feindt et al. "A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654 (2011), pp. 432–440. DOI: 10.1016/j.nima.2011.06.008.

[29] T. V. Dong et al. "Search for the decay $B^0 \to K^{*0}\tau^+\tau^-$ at the Belle experiment". In: *Phys. Rev. D* 108 (2023), p. L011102. DOI: 10.1103/PhysRevD.108.L011102.

[30] N. Uraltsev. *Heavy quark expansion in beauty and its decays*. Proceedings of the International School of Physics "Enrico Fermi". 1998. DOI: 10.3254/978-1-61499-222-6-329.

[31] D. Besson and T. Skwarnicki. "Upsilon Spectroscopy: Transitions in the Bottomonium System". In: *Annual Review of Nuclear and Particle Science* 43 (1993), pp. 333–378. DOI: 10.1146/annurev.ns.43.120193.002001.

[32] T. Böhringer et al. "Observation of Υ, Υ', and Υ'' at the Cornell Electron Storage Ring". In: *Phys. Rev. Lett.* 44 (1980), pp. 1111–1114. DOI: 10.1103/PhysRevLett.44.1111.

[33] G. Finocchiaro et al. "Observation of the Υ''' at the Cornell Electron Storage Ring". In: *Phys. Rev. Lett.* 45 (1980), pp. 222–225. DOI: 10.1103/PhysRevLett.45.222.

[34] D. Besson et al. "Observation of New Structure in the e^+e^- Cross Section above the $\Upsilon(4S)$". In: *Phys. Rev. Lett.* 54 (1985), pp. 381–384. DOI: 10.1103/PhysRevLett.54.381.

[35] M. Bauer. "Measuring the Branching Fraction of $B \to \rho \ell \nu_\ell$ Decays with the Belle II Experiment". PhD thesis. Karlsruher Institut für Technologie (KIT), 2023. DOI: 10.5445/IR/1000165627.

[36] K. Akai, K. Furukawa, and H. Koiso. "SuperKEKB collider". In: *Nuclear Instruments and Methods in Physics Research Section A* 907 (2018). Advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn), pp. 188–199. DOI: 10.1016/j.nima.2018.08.017.

[37] H. Aihara et al. *The Belle II Detector Upgrades Framework Conceptual Design Report*. 2024. arXiv: 2406.19421 [hep-ex].

[38] D. Zhou et al. *Luminosity performance of SuperKEKB*. 2023. DOI: 10.1088/1748-0221/19/02/T02002.

[39] Y. Ohnishi et al. "Accelerator design at SuperKEKB". In: *Progress of Theoretical and Experimental Physics* 2013 (2013), 03A011. DOI: 10.1093/ptep/pts083.

[40] T. Abe et al. *Belle II Technical Design Report*. 2010. DOI: 10.48550/arXiv.1011.0352.

[41] P. Fischer et al. "Progress towards a large area, thin DEPFET detector module". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 582 (2007). Proceedings of the 15th International Workshop on Vertex Detectors, pp. 843–848. DOI: 10.1016/j.nima.2007.07.108.

[42] H. Ye et al. "Commissioning and performance of the Belle II pixel detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 987 (2021), p. 164875. DOI: 10.1016/j.nima.2020.164875.

[43] K. H. Kang et al. "The Silicon Vertex Detector of the Belle II Experiment". In: *Proceedings of the 31st International Workshop on Vertex Detectors*. DOI: 10.7566/JPSCP.42.011003.

[44] N. Taniguchi. "Central Drift Chamber for Belle-II". In: *Journal of Instrumentation* 12 (2017), p. C06014. DOI: 10.1088/1748-0221/12/06/C06014.

[45] L. Šantelj et al. "Recent developments in data reconstruction for aerogel RICH at Belle II". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1055 (2023), p. 168502. DOI: 10.1016/j.nima.2023.168502.

[46] U. Tamponi. "The TOP counter of Belle II: Status and first results". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 952 (2020). 10th International Workshop on Ring Imaging Cherenkov Detectors (RICH 2018), p. 162208. DOI: 10.1016/j.nima.2019.05.049.

[47] A. Kuzmin. "Electromagnetic calorimeter of Belle II". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 958 (2020). Proceedings of the Vienna Conference on Instrumentation 2019, p. 162235. DOI: 10.1016/j.nima.2019.05.076.

[48] E. Manoni et al. "The upgrade of the Belle II forward calorimeter". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 845 (2017). Proceedings of the Vienna Conference on Instrumentation 2016, pp. 524–527. DOI: 10.1016/j.nima.2016.06.074.

[49] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning : Data Mining, Inference, and Prediction*. Second edition. Springer series in statistics. New York: Springer, 2017. DOI: 10.1007/978-0-387-84858-7.

[50] L. Breiman. "Bagging predictors". In: *Machine learning* 24 (1996), pp. 123–140. DOI: 10.1007/BF00058655.

[51] Y. Freund and R.E. Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55 (1997), pp. 119–139. DOI: 10.1006/jcss.1997.1504.

[52] J. Friedman. "Stochastic gradient boosting". In: *Computational Statistics & Data Analysis* 38 (2002). Nonlinear Methods and Data Mining, pp. 367–378. DOI: 10.1016/S0167-9473(01)00065-2.

[53] T. Keck. "FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification". In: *Computing and Software for Big Science* 1 (2017). DOI: 10.1007/s41781-017-0002-8.

[54] J. Therhaag. "TMVA - Toolkit for Multivariate Data Analysis in ROOT". In: *PoS* ICHEP 2010 (2011), p. 510. DOI: 10.22323/1.120.0510.

[55] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830. DOI: 10.5555/1953048.2078195.

[56] T. Chen and C. Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. ACM, 2016. DOI: 10.1145/2939672.2939785.

[57] The Belle II Collaboration. *Belle II Analysis Software Framework (basf2)*. Version release-06-00-09. 2022. DOI: 10.5281/zenodo.6949513.

[58] T. Kuhr et al. "The Belle II Core Software: Belle II Framework Software Group". In: *Computing and Software for Big Science* 3 (2018). DOI: 10.1007/s41781-018-0017-9.

[59] R. Brun and F. Rademakers. "ROOT — An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389 (1997). New Computing Techniques in Physics Research V, pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

[60] T. Keck et al. "The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment". In: *Computing and Software for Big Science* 3 (2019). DOI: 10.1007/s41781-019-0021-8.

[61] I. Adachi et al. *Measurement of the integrated luminosity of data samples collected during 2019-2022 by the Belle II experiment*. 2024. DOI: 10.1088/1674-1137/ad806c.

[62] P. Krishnan et al. *K^0_S reconstruction using phase 2 data*. Internal Note. 2018. URL: https://docs.belle2.org/record/957.

[63] F. Fang. *Study of Ks $\to \pi^+\pi^-$ Selection*. Internal Note. 2000. URL: https://belle.kek.jp/secured/belle_note/gn323/.

[64] S. Moneta, E. Manoni, and C. Cecchi. *Search for $B^+ \to K^{*0}\tau^+\tau^-$ decay with an hadronic tagged approach*. Internal Note. 2024. URL: https://docs.belle2.org/record/3928.

[65] M. Röhrken. "Time-Dependent CP Violation Measurements in Neutral B Meson to Double-Charm Decays at the Japanese Belle Experiment". PhD thesis. Karlsruher Institut für Technologie (KIT), 2012. URL: https://digbib.bibliothek.kit.edu/volltexte/1000028856.

[66] G. C. Fox and S. Wolfram. "Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes". In: *Phys. Rev. Lett.* 41 (1978), pp. 1581–1585. DOI: 10.1103/PhysRevLett.41.1581.

[67] A. J. Bevan et al. "The Physics of the B Factories". In: *The European Physical Journal C* 74 (2014). DOI: 10.1140/epjc/s10052-014-3026-9.

[68] D. M. Asner et al. "Search for exclusive charmless hadronic B decays". In: *Physical Review D* 53 (1996), pp. 1039–1050. DOI: 10.1103/physrevd.53.1039.

[69] G. Punzi. *Sensitivity of searches for new signals and its optimization*. 2003. arXiv: physics/0308063 [physics.data-an].

[70] W. Sutcliffe and F. Bernlochner. "Calibration of hadronic tagging with $B \to X\ell\nu$ decays with the LS1 Dataset". In: (2023). 'Internal note'. URL: https://docs.belle2.org/record/3407.

[71] M. Liu et al. "Hadronic FEI calibration with $D^{(*)}\pi$ samples". In: (2023). 'Internal note'. URL: https://docs.belle2.org/record/3373.

[72] W. Sutcliffe et al. "Combination of calibration factors for hadronic tagging". In: (2023). 'Internal note'. URL: https://docs.belle2.org/record/3642.

[73] T.Koga. *Measurement of momentum dependent π^0 reconstruction efficiency with* D*decays*. 'Internal note'. 2020. URL: https://docs.belle2.org/record/2096.

[74] Belle II Lepton ID Group. *Muon and electron identification performance with* 189 fb^{-1} *of Belle II data*. Internal note. 2021. URL: https://docs.belle2.org/record/2340.

[75] L. Damer. *Supplementation material for "Search for* $\mathrm{B}^+ \to \mathrm{K}^{*+}\tau^+\tau^-$ *with hadronic tagging at the Belle II Experiment"*. Version 1.0. 2024. DOI: 10.5281/zenodo.13891141.

[76] G. Cowan. *Statistical Data Analysis*. Oxford science publications. Clarendon Press, 1998. ISBN: 978-0-19-850155-8.

[77] O. Behnke and L. Moneta. "Parameter Estimation". In: *Data Analysis in High Energy Physics*. John Wiley & Sons, Ltd, 2013, pp. 27–73. DOI: 10.1002/9783527653416.ch2.

[78] D. Britzger. "The Linear Template Fit". In: *The European Physical Journal C* 82 (2022). DOI: 10.1140/epjc/s10052-022-10581-w.

[79] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71 (2011). DOI: 10.1140/epjc/s10052-011-1554-0.

[80] M. Gelb. "Search for the Rare Decay $B^+ \to \ell^+ \nu_\ell \gamma$ with the Full Event Interpretation at the Belle Experiment". PhD thesis. Karlsruher Institut für Technologie (KIT), 2018. DOI: 10.5445/IR/1000085417.

A Appendix

A.1 Cut-Flow Tables

In this section of the appendix I present the cut-flow tables of the reconstruction procedure described in Chapter 5. The cut-flow tables consist of efficiency and multiplicity information after each selection criterion. I present the cut-flow table for the truth matched and non-truth matched cases. The efficiency is hereby defined as the number of unique events reconstructed divided by the number of generated events. In the truth matched case, only reconstructed events are considered that hold one correct reconstructed event candidate. Truth matching information is only available after the $\Upsilon(4S)$ resonance is reconstructed.

Table A.1: Non-truth matched cut-flow table describing the efficiency and multiplicity for each selection criteria introduced in chapter 5. For reconstruction, the signal MC sample for $B^+ \to K^{*+}\tau^+\tau^-$ has been used. A dash denotes that the corresponding selection criteria is not applied in the specific instance or the information is unavailable.

Selection Criterion/Channels	$\ell\ell$		$\ell\pi$		$\pi\pi$		ρ	
	Eff. (in %)	Multipl.	Eff. (in %)	Multipl.	Eff. (in %)	Multipl.	Eff. (in %)	Multipl.
Generated	100.00	-	100.00	-	100.00	-	100.00	-
FEI Skim (hadr.)	5.1419	-	6.5930	-	9.9620	-	9.1036	-
Reconstructing $\Upsilon(4S) \to B_{tag}\, B_{sig}$	2.5868	161.48	3.5397	138.09	5.5717	101.43	3.7525	616.64
B_{tag} $M_{bc} > 5.27\,\mathrm{GeV}/c^2$	1.4592	116.13	1.6495	102.60	2.0621	63.81	1.6641	409.79
B_{tag} $-0.15 < \Delta E < 0.1\,\mathrm{GeV}$	1.3039	104.70	1.4814	91.83	1.8341	61.66	1.4766	393.78
B_{tag} $P_{FEI} > 0.01$	0.7358	69.79	0.7852	59.21	0.8218	35.92	0.6499	250.34
Track Cuts	0.6442	26.10	0.6685	18.95	0.7043	12.39	0.5698	170.51
GBKS criteria	0.5969	22.83	0.6078	15.75	0.6624	11.01	0.5134	119.21
K^0_S mass cut ($0.48 < M < 0.52\,\mathrm{GeV}/c^2$)	0.5933	22.81	0.6032	15.79	0.6594	11.01	0.5134	85.55
K^{*+} mass cut ($0.790 < M < 0.994\,\mathrm{GeV}/c^2$)	0.4658	11.01	0.4615	7.87	0.5231	5.73	0.4824	61.23
ρ mass cut ($0.65 < M < 0.90\,\mathrm{GeV}/c^2$)	-	-	-	-	-	-	0.2767	20.69
electron PID (BDT > 0.9)	0.1578	6.07	0.1915	4.27	-	-	-	-
muon PID (BDT > 0.9)	0.1023	2.39	0.0581	4.21	-	-	-	-
pionID > 0.9	0.0911	2.44	0.0251	3.92	0.1783	2.92	0.0714	5.73
kaonID > 0.9	0.0318	1.95	0.0079	3.16	0.0567	1.71	0.0201	4.12
$\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}}) < 0.9$	-	-	-	-	0.0518	1.71	-	-
N_{tracks} in ROE	0.0293	1.85	0.0053	1.87	0.0497	1.70	0.0182	3.06
BCS: K^{*+} mass closest to PDG	0.0293	1.36	0.0053	1.43	0.0497	1.07	0.0182	1.97
BCS: Choose random $\Upsilon(4S)$	0.0293	1.00	0.0053	1.00	0.0497	1.00	0.0182	1.00

Table A.2: Truth matched cut-flow table describing the efficiency and multiplicity for each selection criteria introduced in chapter 5. For reconstruction, the signal MC sample for $B^+ \to K^{*+}\tau^+\tau^-$ has been used. A dash denotes that the corresponding selection criteria is not applied in the specific instance or the information is unavailable. In addition to the non-truth matched table, efficiency losses are observed in the Best Candidate Selection due to wrongly chosen candidates in this procedure.

Selection Criterion/Channels	$\ell\ell$		$\ell\pi$		$\pi\pi$		ρ	
	Eff. (in %)	Multipl.	Eff. (in %)	Multipl.	Eff. (in %)	Multipl.	Eff. (in %)	Multipl.
Generated	100.00	-	100.00	-	100.00	-	100.00	-
FEI Skim (hadr.)	5.1419	-	6.5930	-	9.9620	-	9.1036	-
Reconstructing $\Upsilon(4S) \to B_{tag}\, B_{sig}$	1.6671	69.21	2.3866	64.51	3.9748	49.83	2.1939	263.75
B_{tag} $M_{bc} > 5.27\,\mathrm{GeV}/c^2$	0.8962	49.04	1.0430	43.22	1.3488	34.21	0.8994	189.09
B_{tag} $-0.15 < \Delta E < 0.1\,\mathrm{GeV}$	0.7743	45.30	0.9112	39.48	1.1685	33.45	0.7719	178.56
B_{tag} $P_{FEI} > 0.01$	0.4566	28.49	0.4845	24.59	0.5031	17.87	0.3422	112.41
Track Cuts	0.3991	14.44	0.4173	10.25	0.4313	7.58	0.2913	75.65
GBKS criteria	0.3532	11.61	0.3645	7.656	0.3885	6.35	0.2479	67.10
K^0_S mass cut ($0.48 < M < 0.52\,\mathrm{GeV}/c^2$)	0.3496	11.43	0.3619	7.632	0.3855	6.33	0.2471	64.52
K^{*+} mass cut ($0.790 < M < 0.994\,\mathrm{GeV}/c^2$)	0.2668	6.63	0.2709	4.26	0.2968	3.53	0.2312	31.59
ρ mass cut ($0.65 < M < 0.90\,\mathrm{GeV}/c^2$)	-	-	-	-	-	-	0.1438	10.68
electron PID (BDT > 0.9)	0.0956	3.58	0.1472	2.36	-	-	0.1345	7.58
muon PID (BDT > 0.9)	0.0627	1.41	0.0382	2.21	-	-	0.1189	4.22
pionID > 0.9	0.0467	1.37	0.0145	2.11	0.0886	2.02	0.0345	3.05
kaonID > 0.9	0.0236	1.33	0.0065	1.89	0.0358	1.39	0.0109	2.66
$\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}}) < 0.9$	-	-	-	-	0.0329	1.32	-	-
N_{tracks} in ROE	0.0221	1.21	0.0046	1.59	0.0317	1.28	0.0091	3.06
BCS: K^{*+} mass closest to PDG	0.0202	1.14	0.0039	1.21	0.0298	1.16	0.0088	1.24
BCS: Choose random $\Upsilon(4S)$	0.0185	1.00	0.0037	1.00	0.0278	1.00	0.0081	1.00

A.2 D Meson Daughters From B^+B^- Background

In this part of the appendix, I show the ten most frequent D^+ and D^0 decays which are part of the charged background that is falsely reconstructed as signal event (see Table 5.2).

Table A.3: Depicting the ten most frequent D^+ decays originating from falsely reconstructed B_{sig} candidates. The D^+ daughters are in the first column while the second describes the overall share in percent. The tool to extract these ratios is developed with the AI programming assistant GitHub Copilot.

D^+ Daughters	Fraction in %
$K^- \pi^+ \pi^-$	24.16
$K^0_S \pi^+ \pi^0$	10.61
$K^0 \mu^+ \nu_\mu$	5.67
$\overline{K}^0 e^+ \nu_e$	5.61
$K^0_L \pi^+ \pi^0$	5.18
$\overline{K}^{*0} \pi^+ \pi^0$	5.14
$K^0_S \pi^+$	4.32
$K^- \pi^+ \pi^+ \gamma$	3.46
$K^0_S a_1^+$	2.84
$\overline{K}^0_1 \pi^+$	2.31
...	...

Table A.4: Depicting the ten most frequent D^0 decays originating from falsely reconstructed B_{sig} candidates. The D^0 daughters are in the first column while the second describes the overall share in %. The tool to extract these ratios is developed with the AI programming assistant GitHub Copilot.

D^0 Daughters	Fraction in %
$K^- \pi^- \pi^0$	25.55
$K^- a_1^+$	11.27
$K^- \pi^+$	9.96
$K^{*-} \rho^+$	6.83
$K^- \pi^+ \pi^0 \pi^0$	3.92
$K^- \pi^+ \pi^0 \gamma$	3.03
$\overline{K}^{*0} \rho^0$	1.84
$K^- \pi^+ \gamma$	1.83
$K_S^0 \pi^+ \pi^-$	1.63
$K^- \mu^+ \nu_\mu$	1.61
...	...

A.3 Training Variables for the Multivariate Analysis

Table A.5: Training variables used as input features in the BDT classifier to suppress different background topologies against the signal component.

Training Variable	$\ell\ell$	$\ell\pi$	$\pi\pi$	ρ
$p_{\text{miss}}^{\text{CMS}}$	✓	✓	✓	✓
$E_{\text{miss}}^{\text{CMS}}$	✓	✓	✓	✓
E_T		✓	✓	✓
Invariant mass $M_{\text{K}^{*+}}$	✓		✓	✓
$M(\text{K}^{*+}; t_i)$	✓	✓	✓	✓
$\cos(\text{Thrust}_{\text{B}}; \text{Thrust}_{\text{ROE}})$	✓	✓	✓	✓
$\cos(\text{Thrust}_{\text{B}}; \text{Thrust}_{z})$	✓		✓	✓
Thrust_{B}	✓	✓	✓	✓
$\text{Thrust}_{\text{ROE}}$				✓
CLEO Cone 0		✓	✓	✓
CLEO Cone 1	✓	✓	✓	
CLEO Cone 2		✓	✓	✓
CLEO Cone 3		✓	✓	✓
CLEO Cone 4		✓	✓	
CLEO Cone 5				✓
CLEO Cone 6			✓	✓
CLEO Cone 7		✓	✓	
CLEO Cone 8		✓	✓	
H_{00}^{so}		✓		✓
H_{02}^{so}		✓	✓	
H_{03}^{so}	✓		✓	✓
H_{04}^{so}		✓	✓	✓
H_{12}^{so}		✓	✓	
H_{14}^{so}		✓	✓	✓
H_{20}^{so}		✓	✓	✓
H_{22}^{so}	✓	✓	✓	✓
H_{24}^{so}	✓		✓	✓
H_{01}^{oo}			✓	✓
H_{02}^{oo}		✓	✓	✓
H_{03}^{oo}			✓	✓
H_{04}^{oo}			✓	✓

A.4 One-Dimensional Sideband Validation Plots

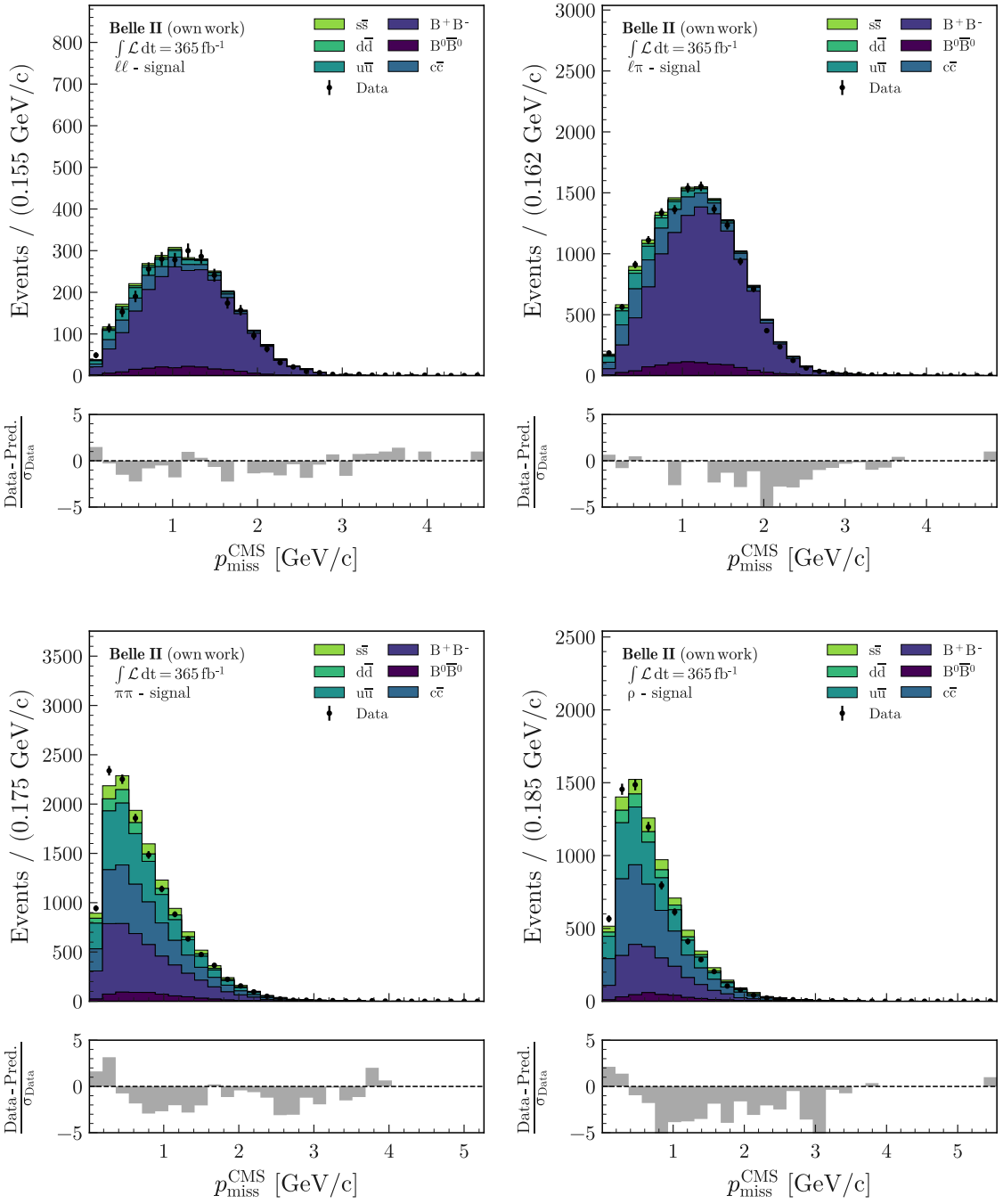


Figure A.1: Comparison of the variable $p_{\text{miss}}^{\text{CMS}}$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

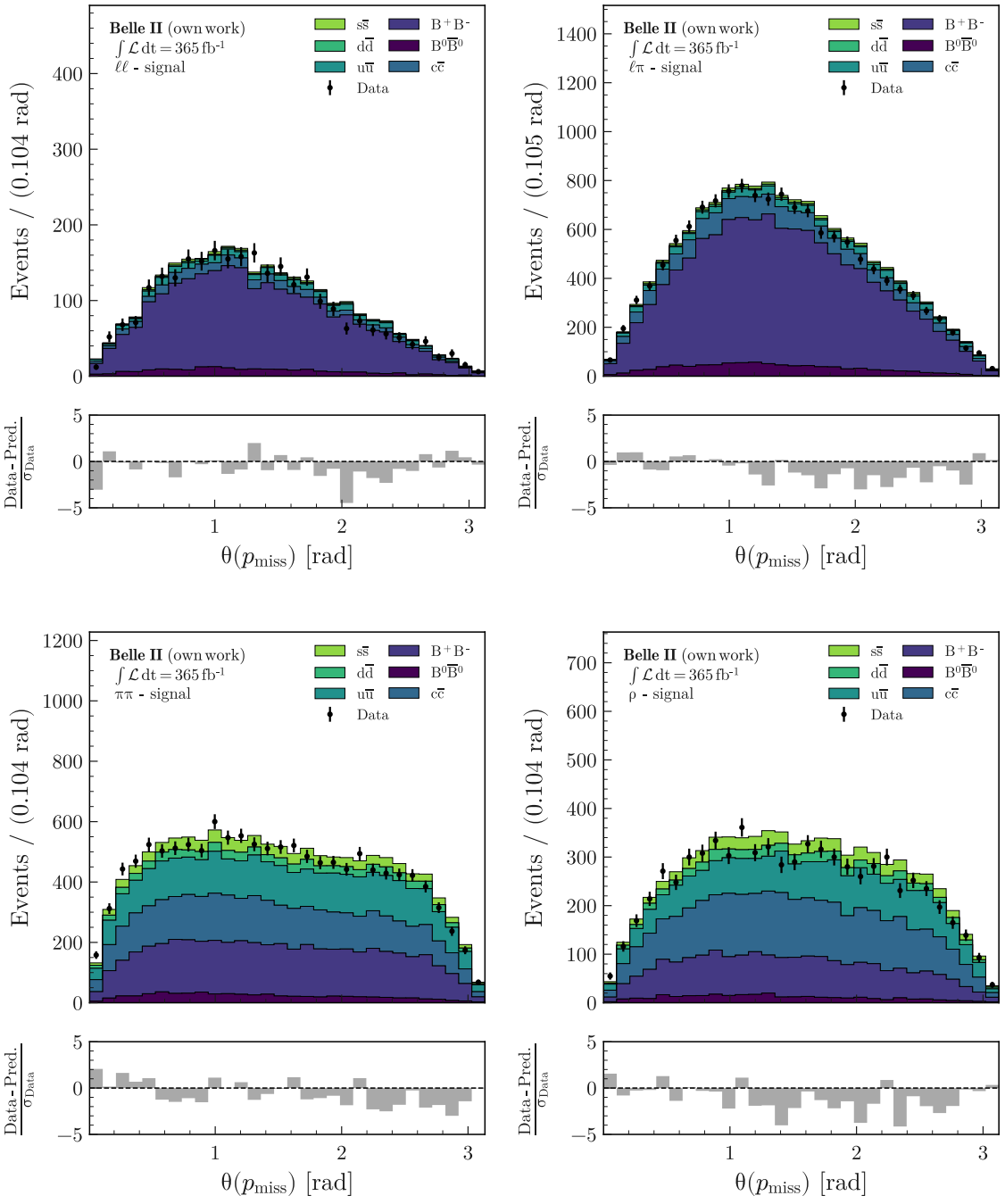


Figure A.2: Comparison of the variable $\theta(p_{\text{miss}})$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

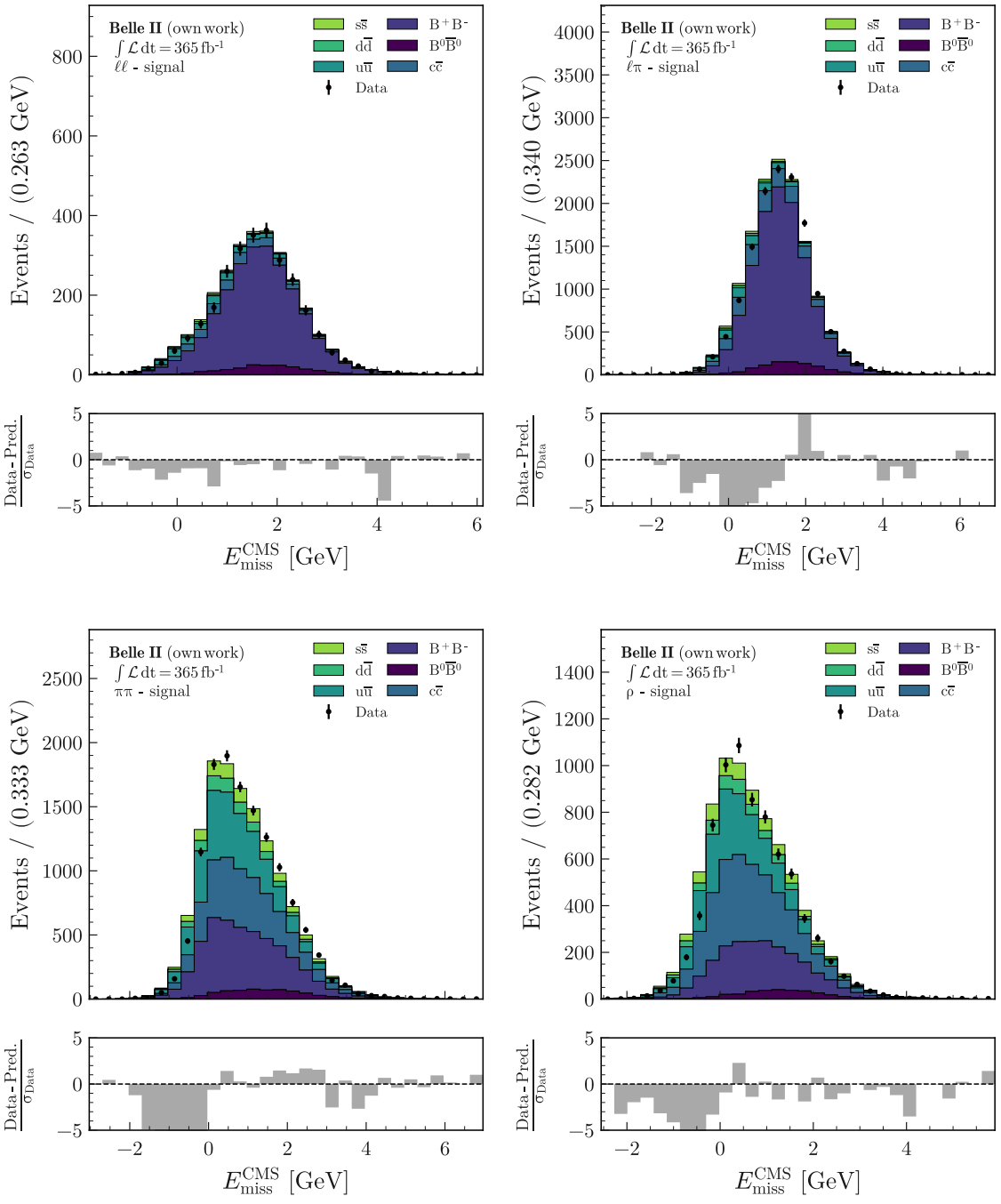


Figure A.3: Comparison of the variable $E_{\text{miss}}^{\text{CMS}}$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

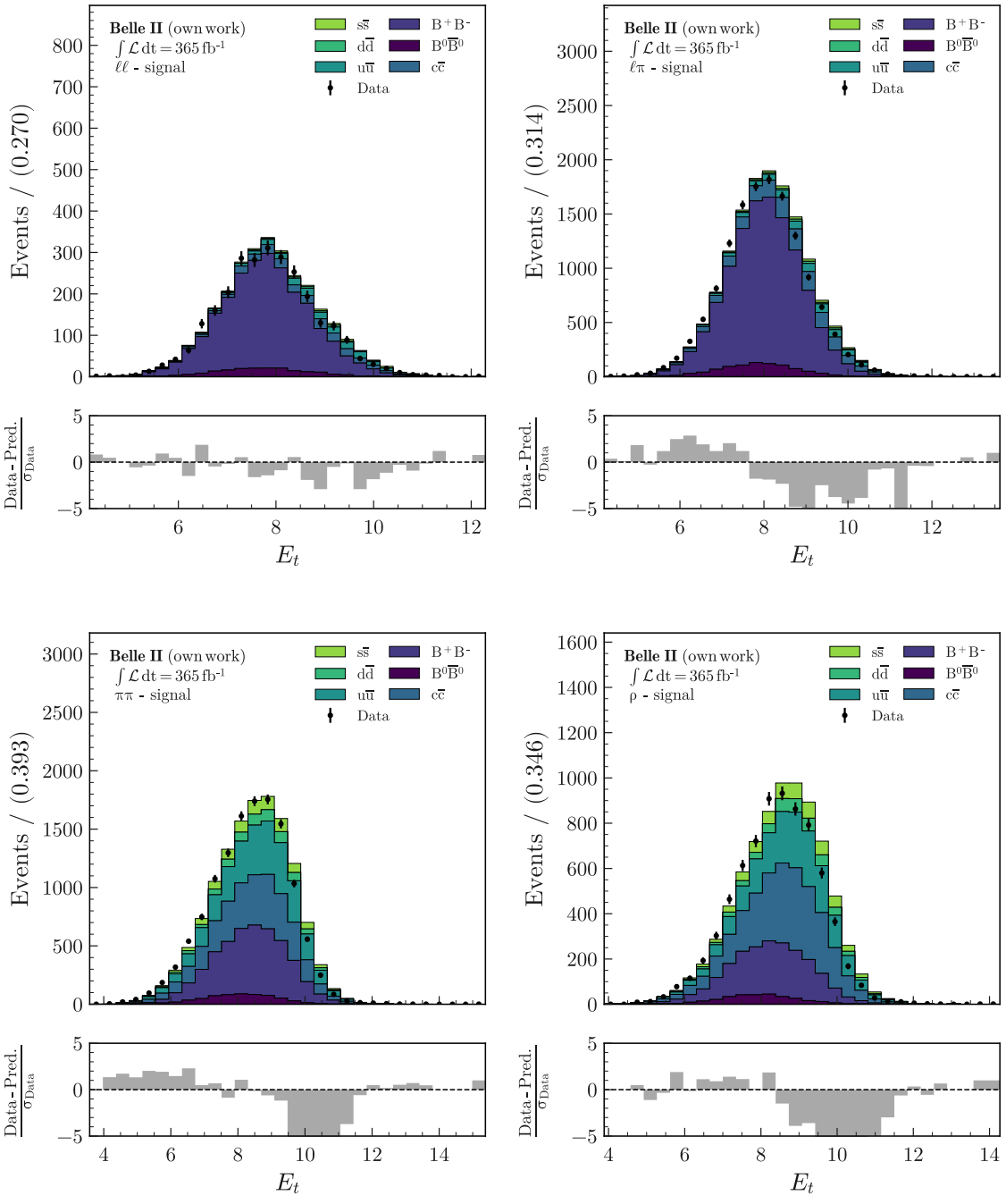


Figure A.4: Comparison of the variable E_T on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

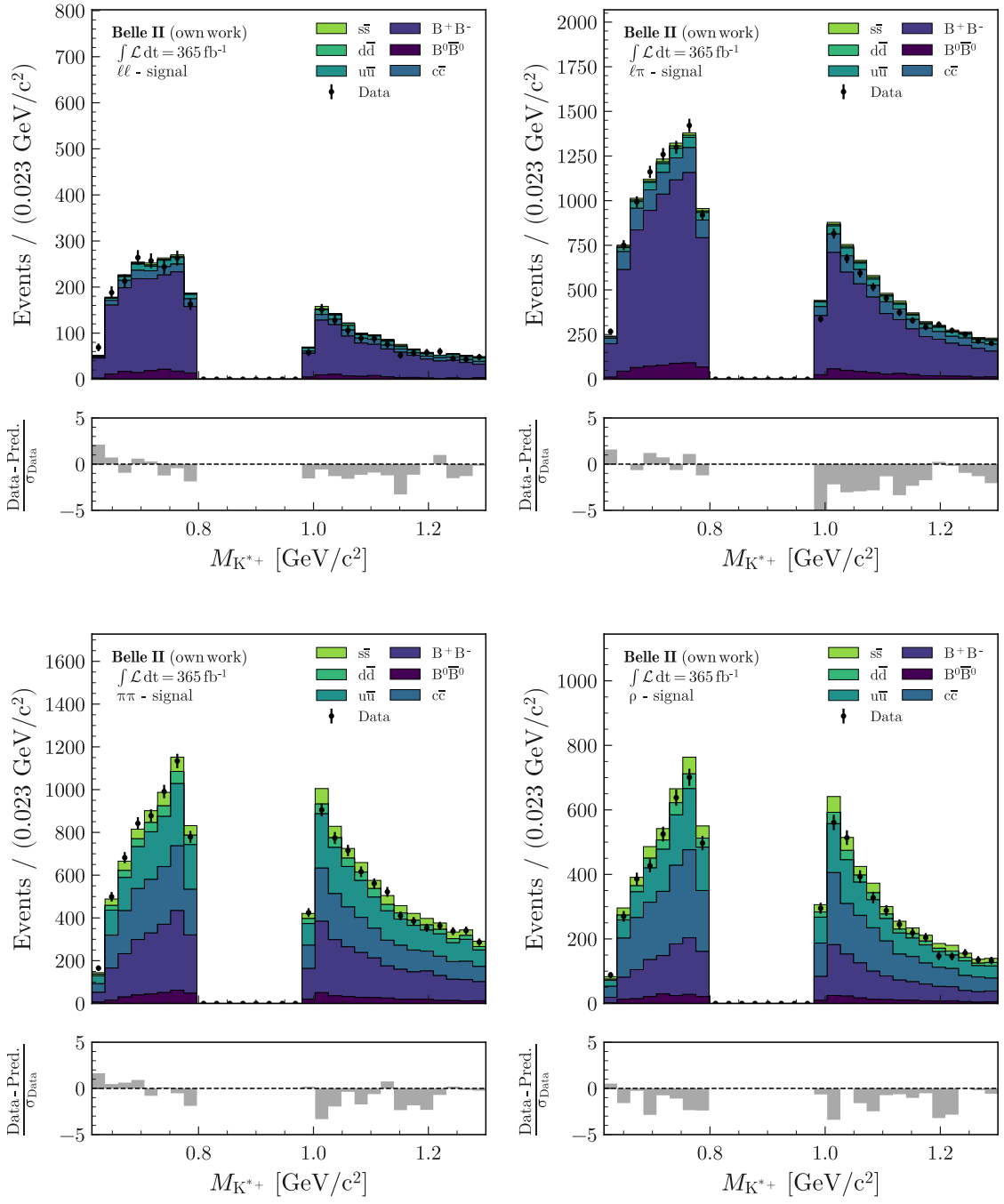


Figure A.5: Comparison of the variable $M_{K^{*+}}$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

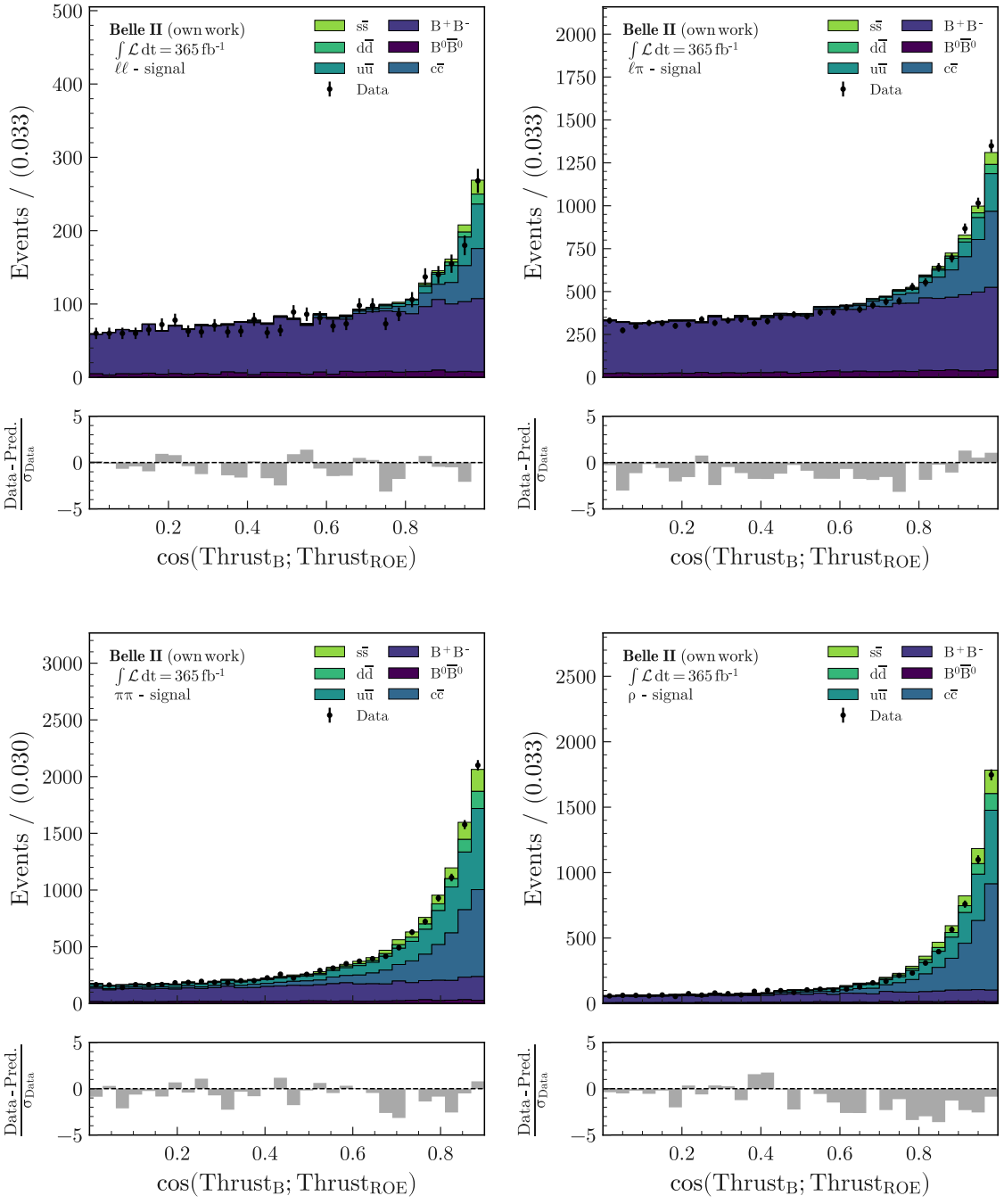


Figure A.6: Comparison of the variable $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

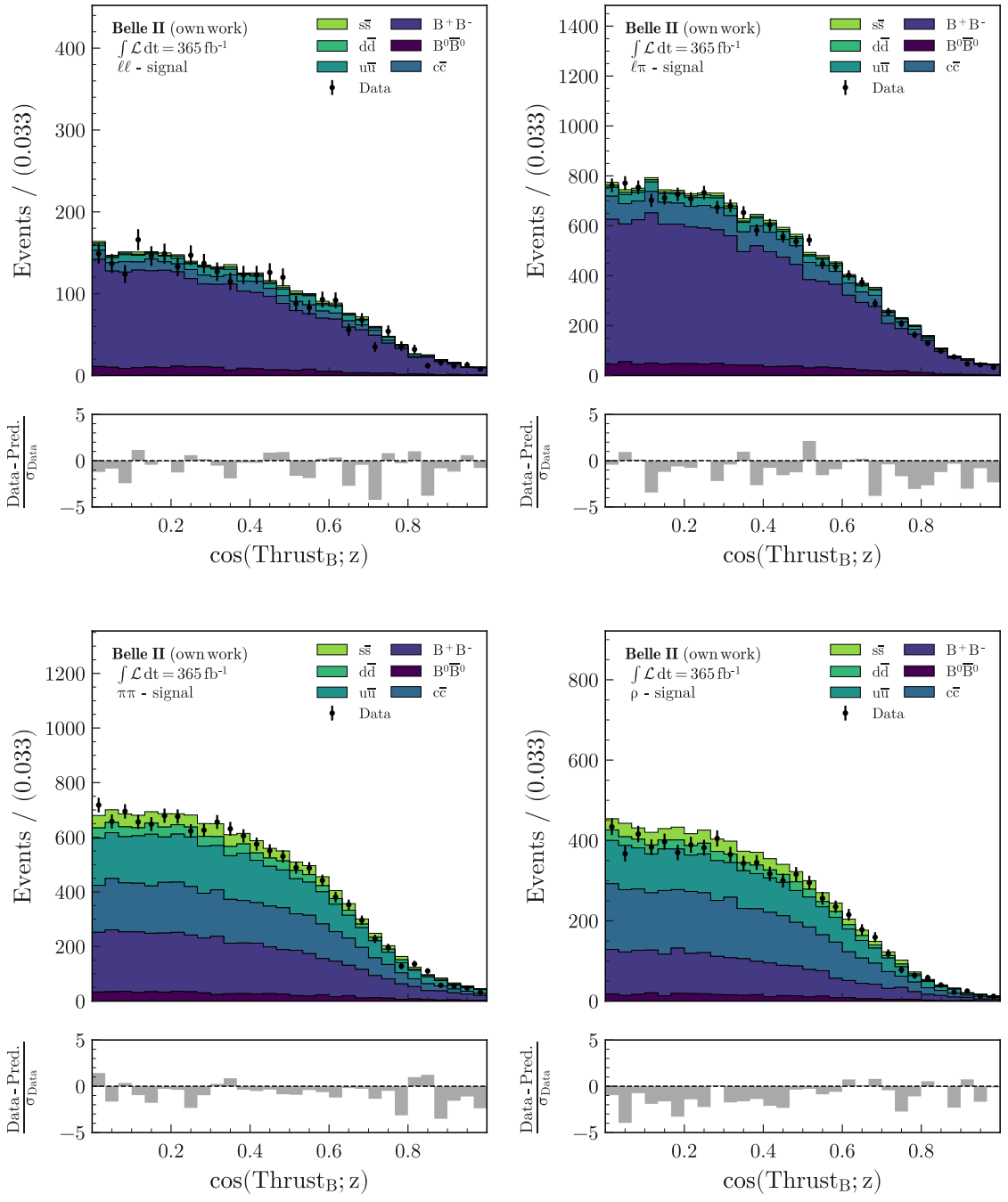


Figure A.7: Comparison of the variable $\cos(\mathrm{Thrust_B}; z)$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

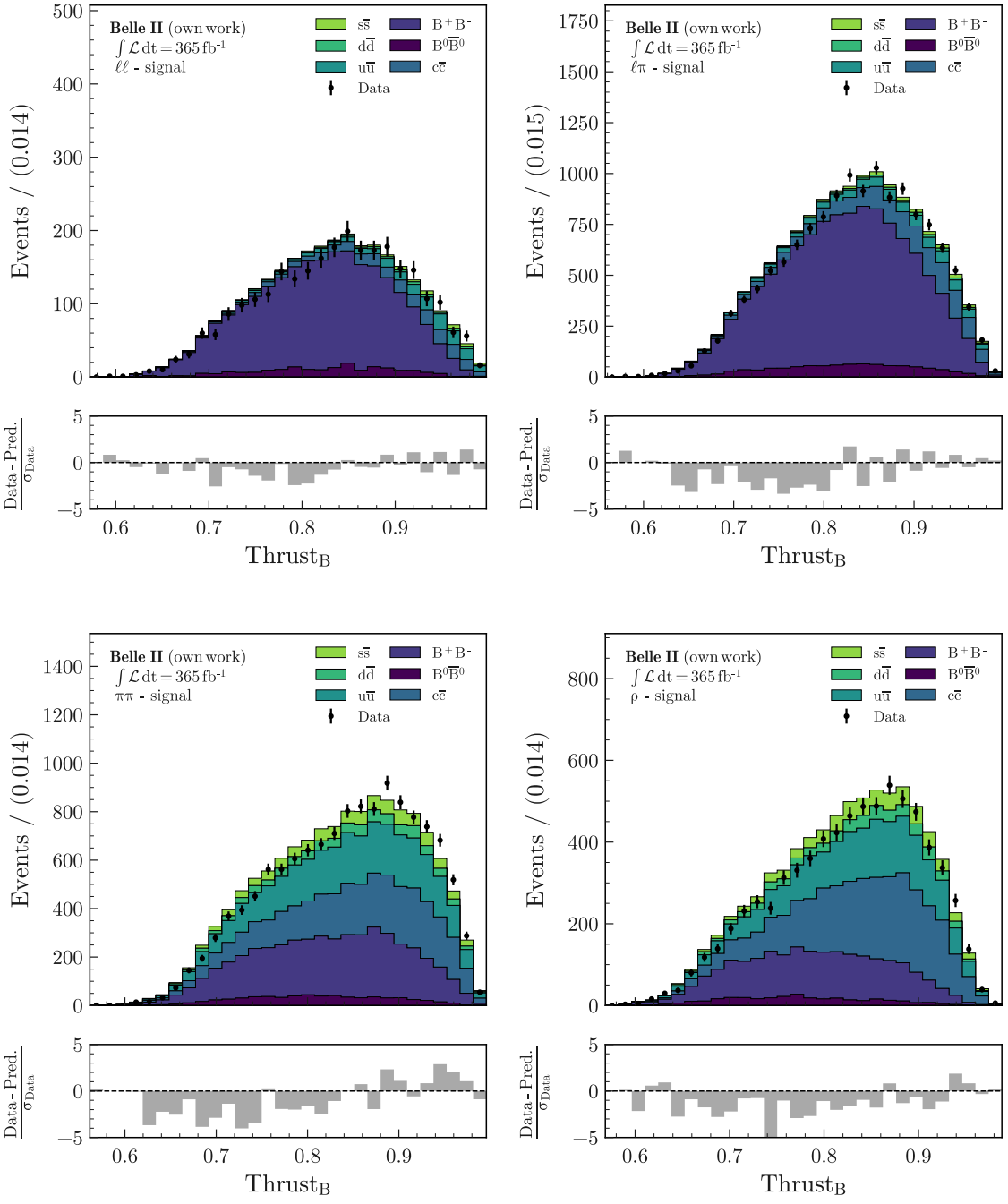


Figure A.8: Comparison of the variable $Thrust_B$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

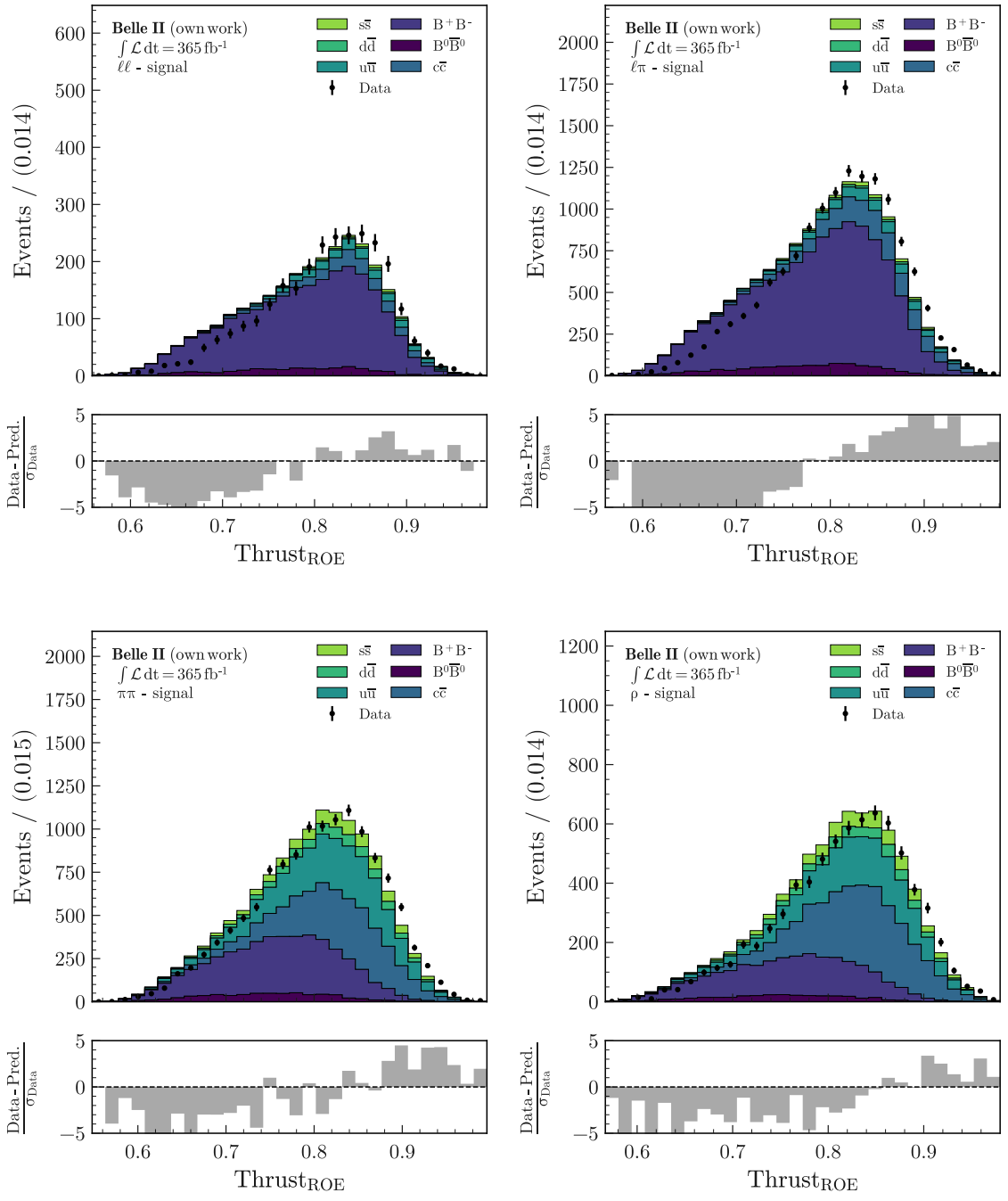


Figure A.9: Comparison of the variable $\text{Thrust}_{\text{ROE}}$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

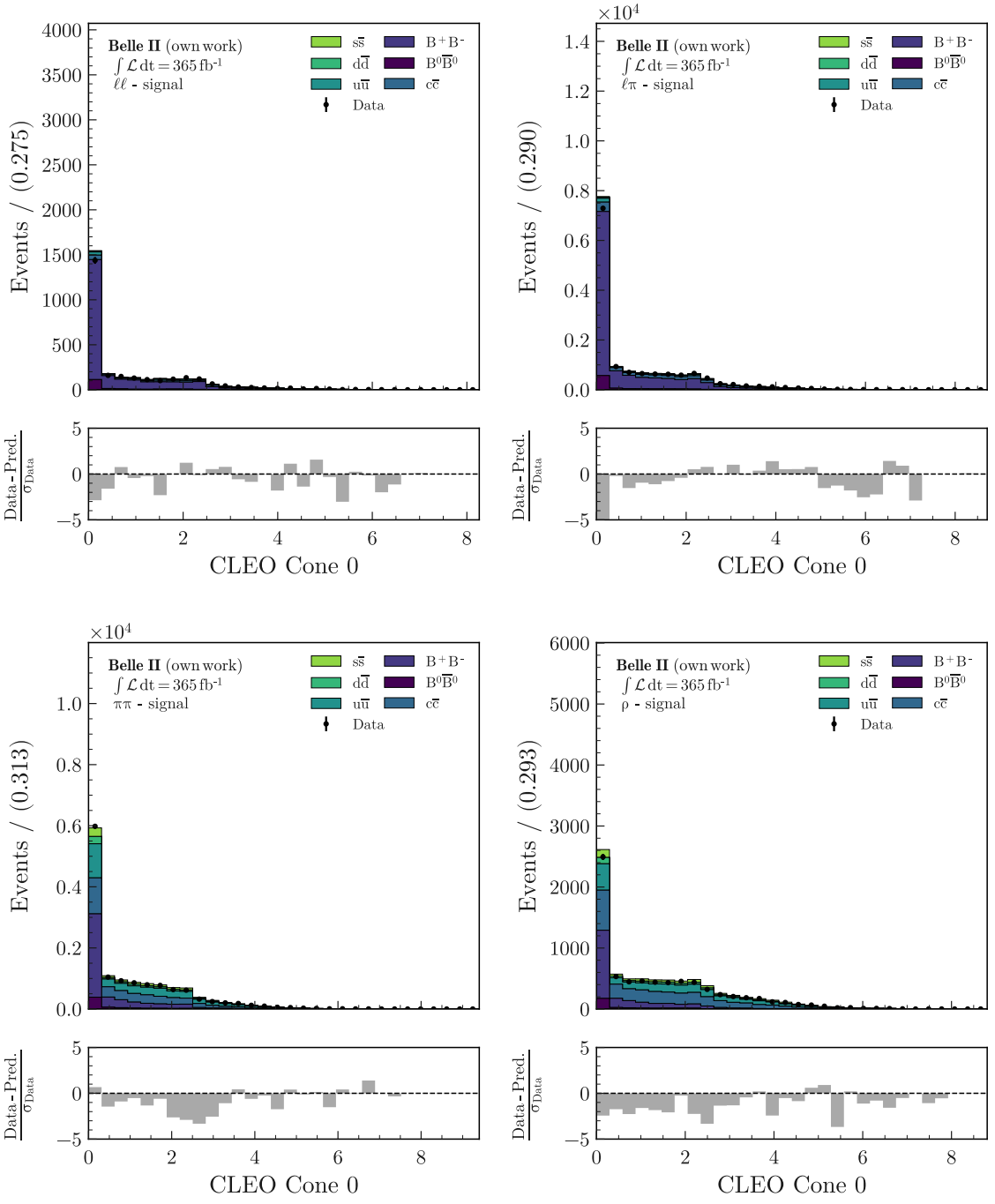


Figure A.10: Comparison of the variable CLEO Cone 0 on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

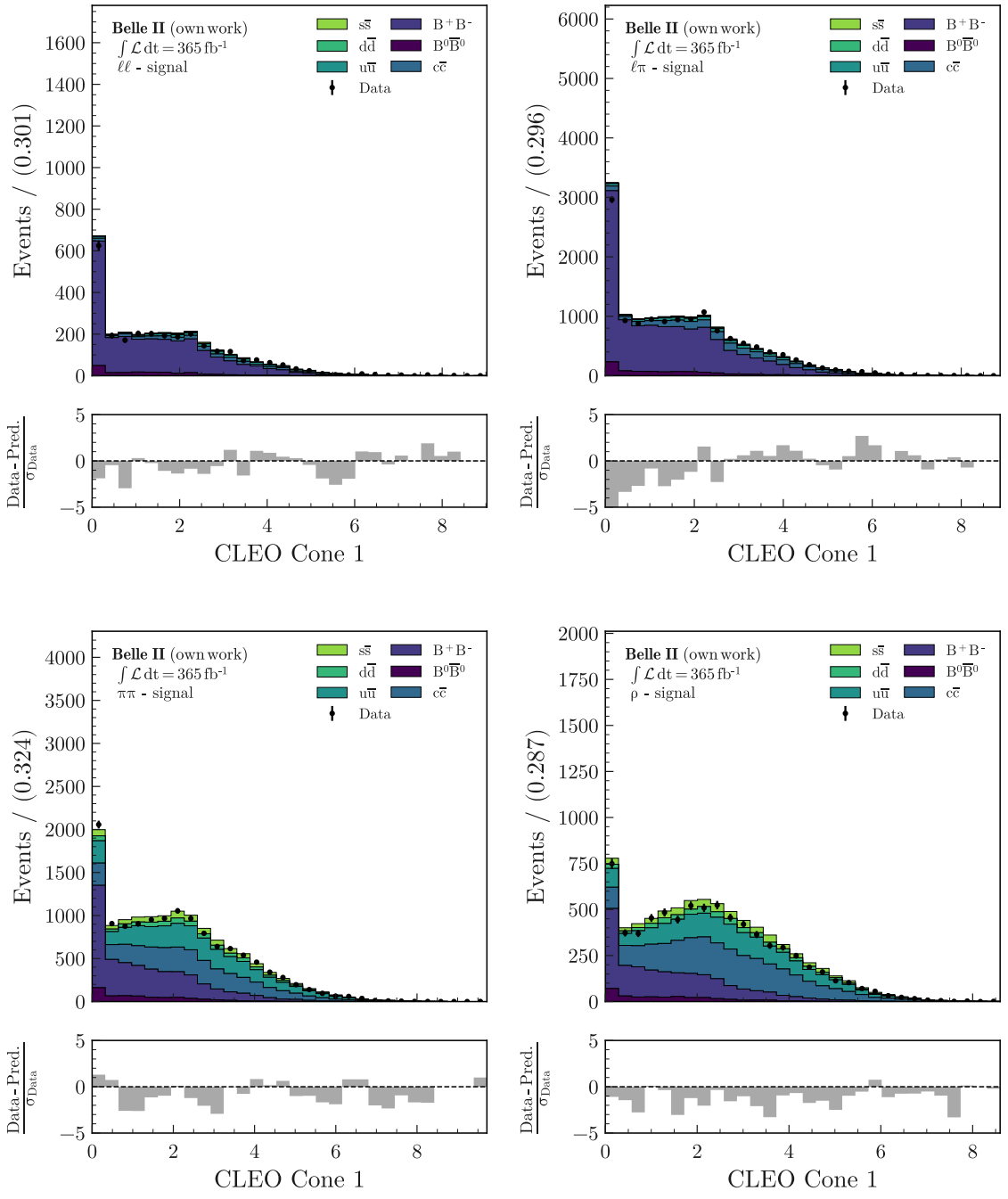


Figure A.11: Comparison of the variable CLEO Cone 1 on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

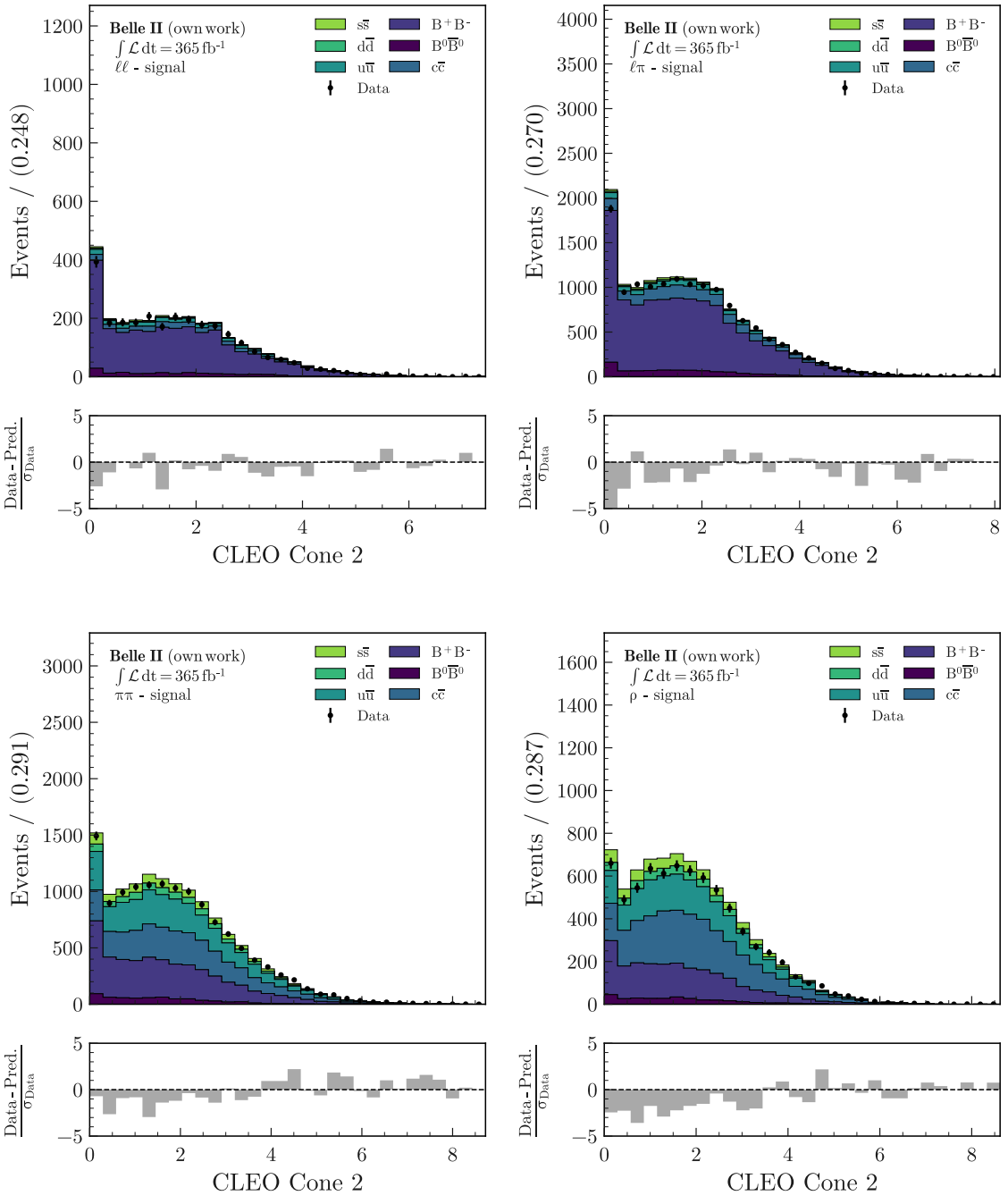


Figure A.12: Comparison of the variable CLEO Cone 2 on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

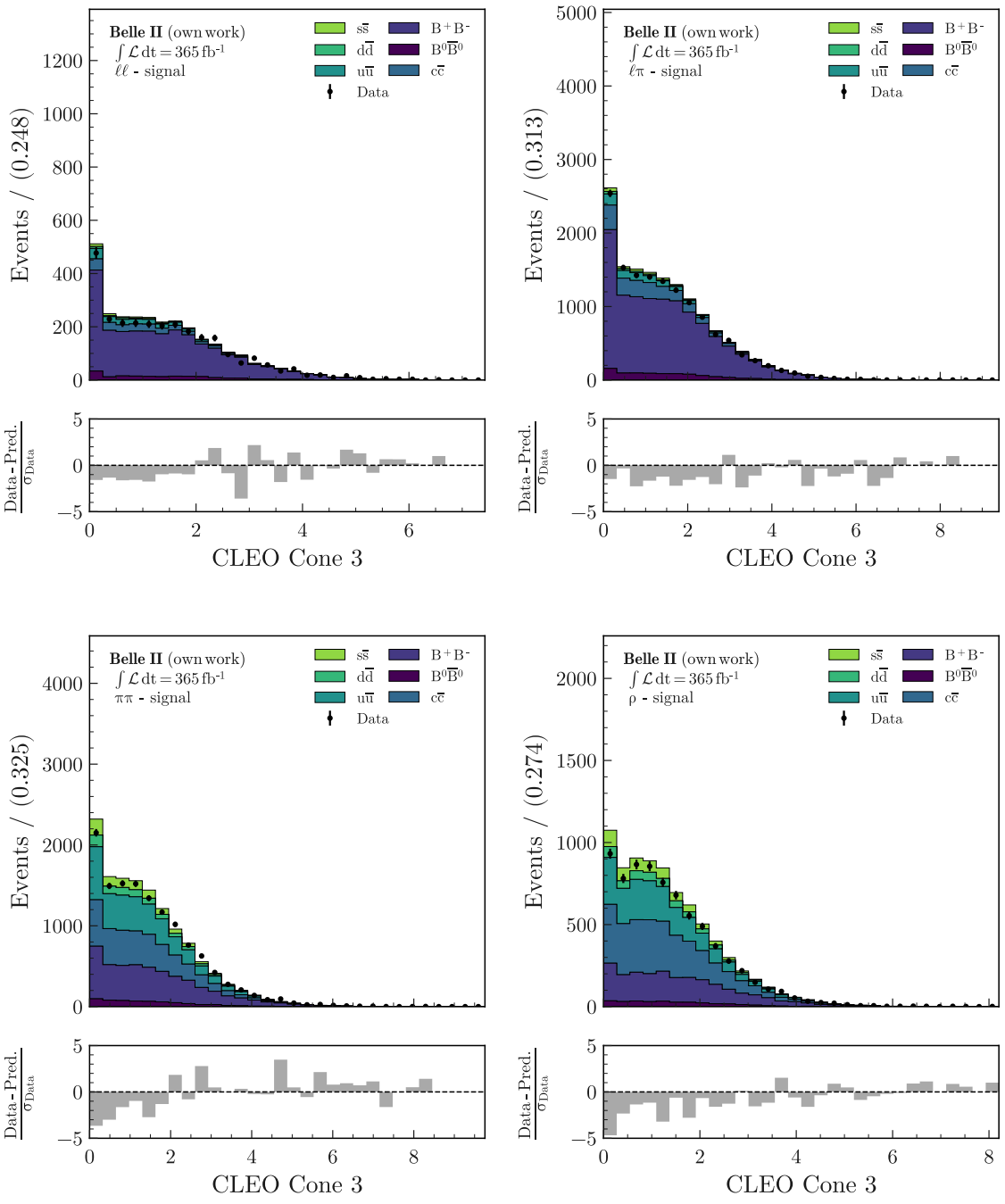


Figure A.13: Comparison of the variable CLEO Cone 3 on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

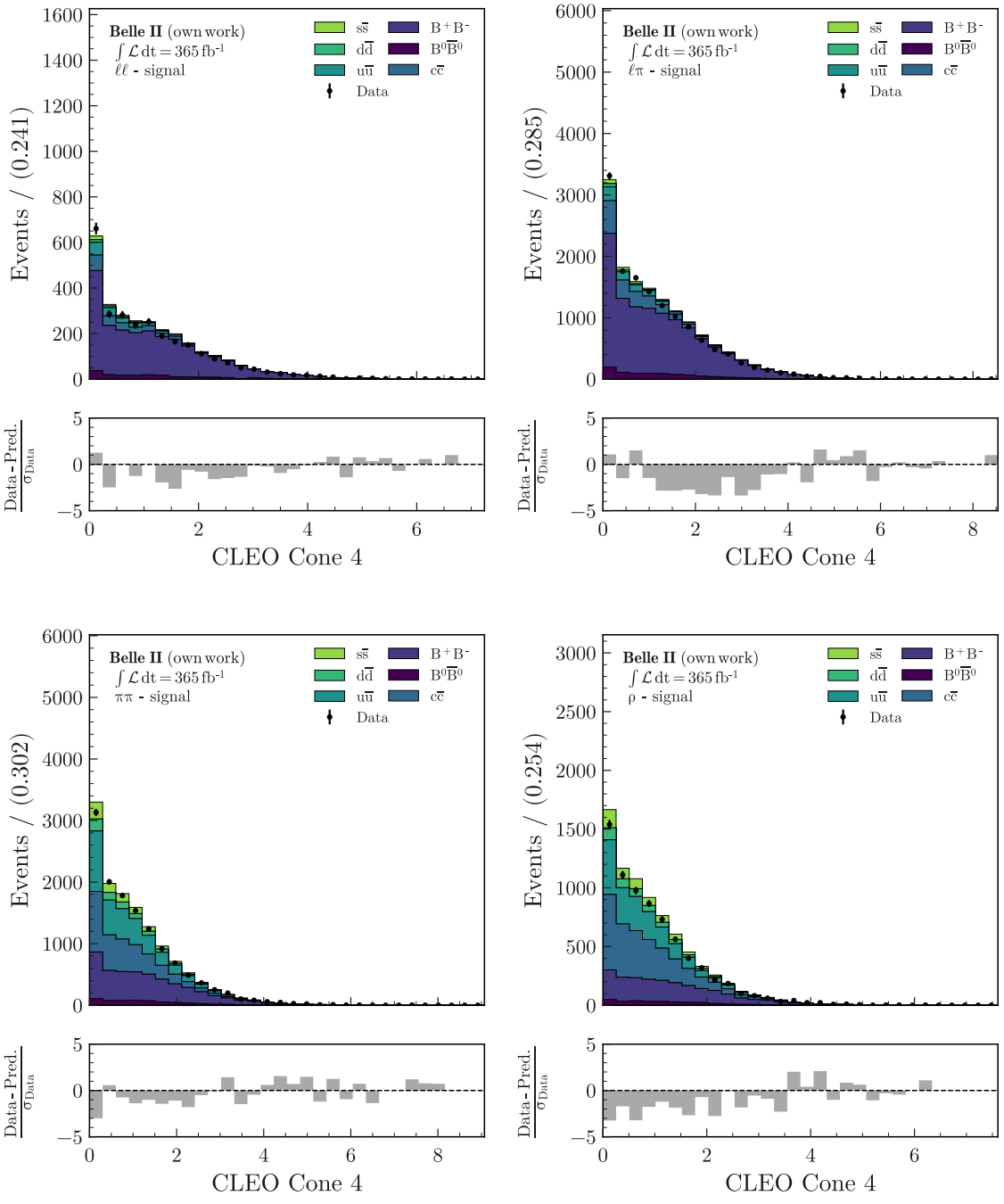


Figure A.14: Comparison of the variable CLEO Cone 4 on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

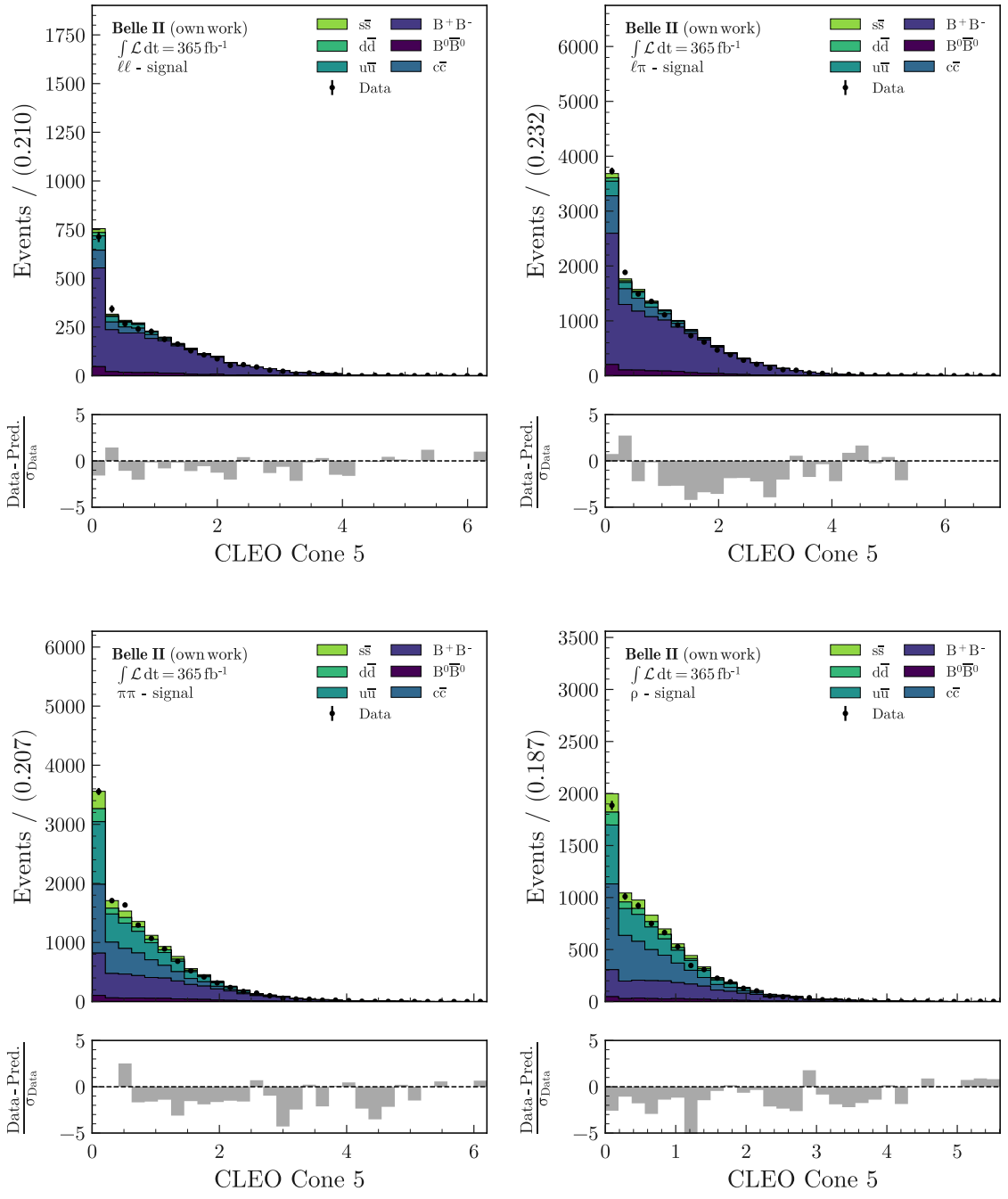


Figure A.15: Comparison of the variable CLEO Cone 5 on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

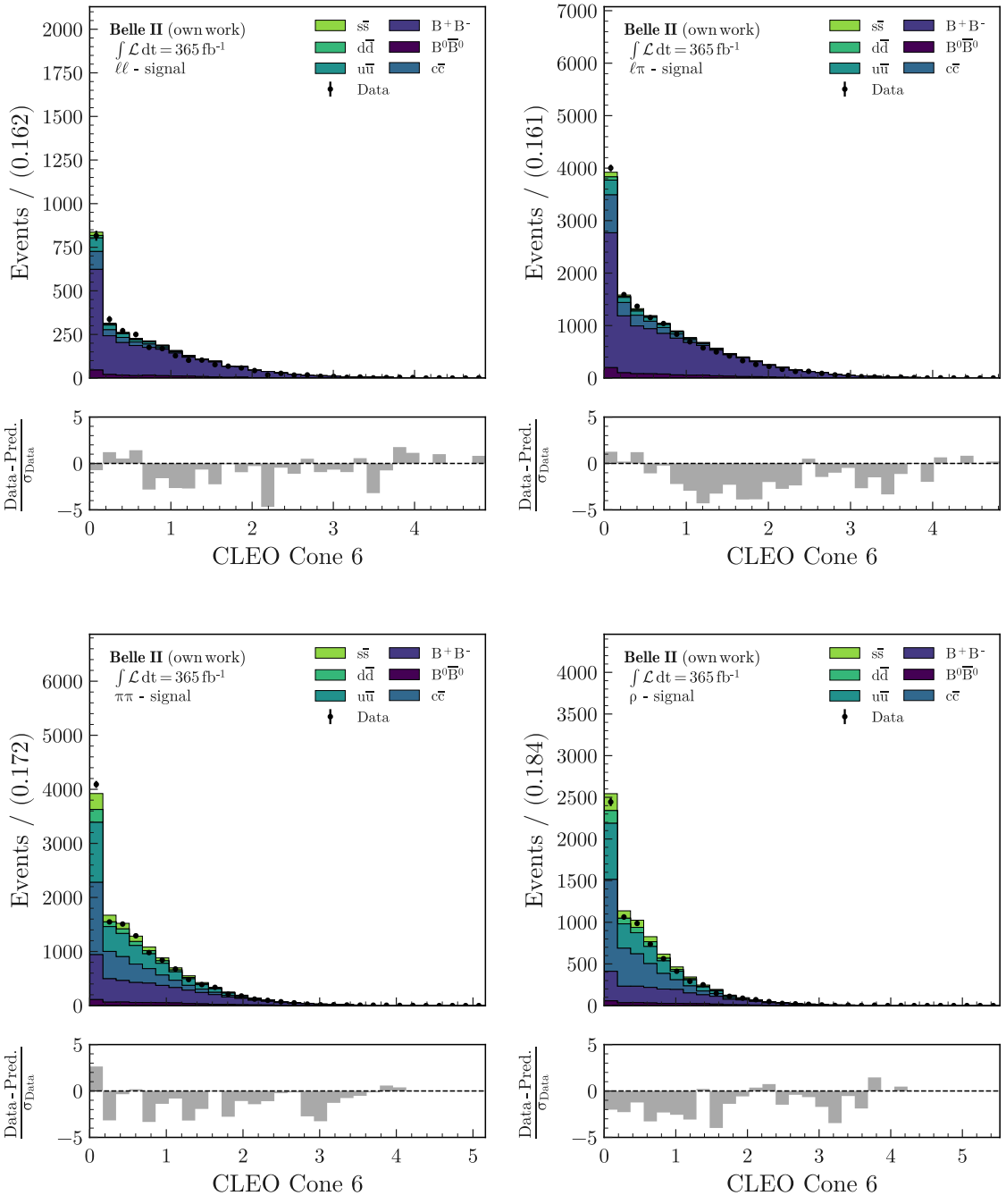


Figure A.16: Comparison of the variable CLEO Cone 6 on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

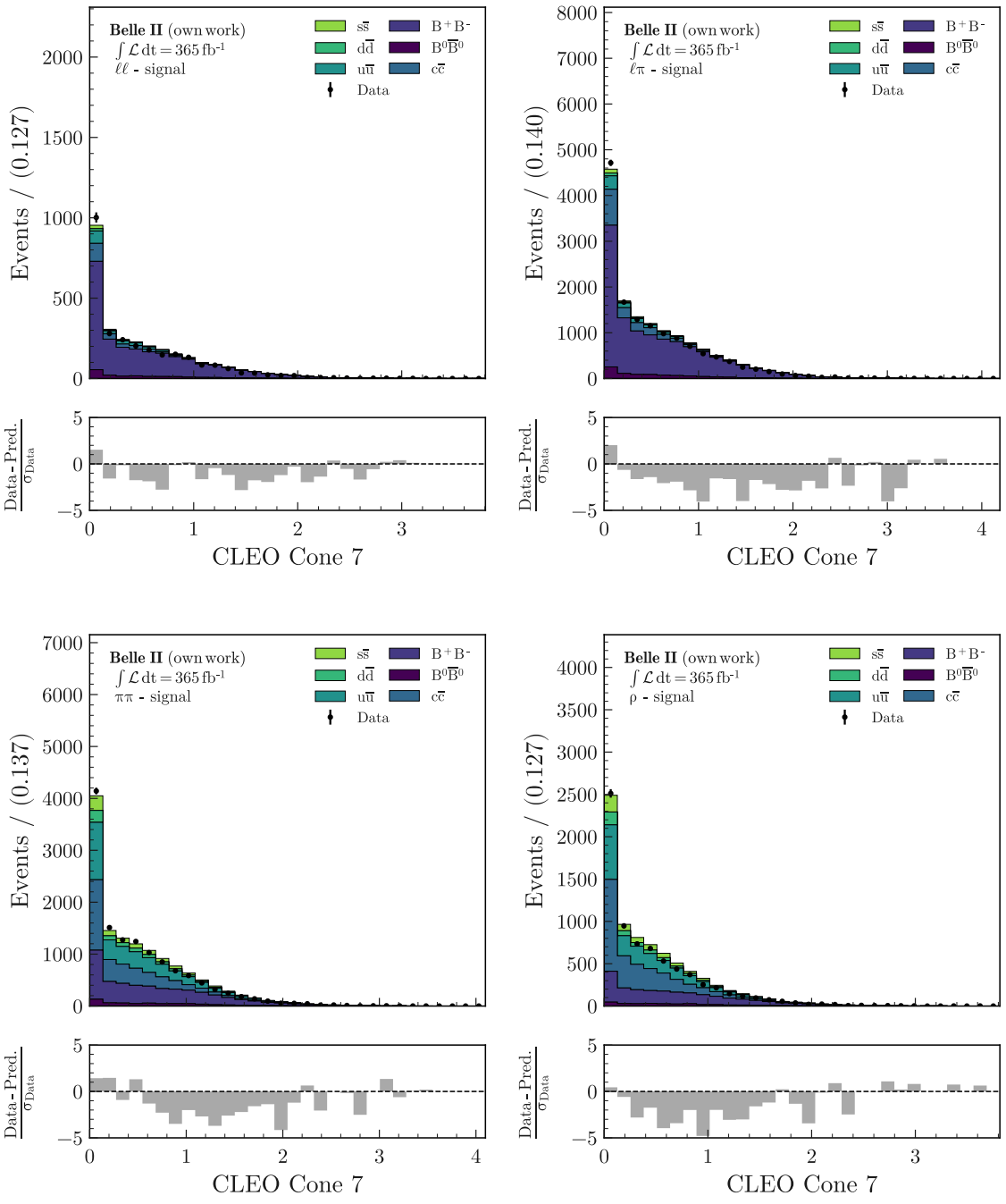


Figure A.17: Comparison of the variable CLEO Cone 7 on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

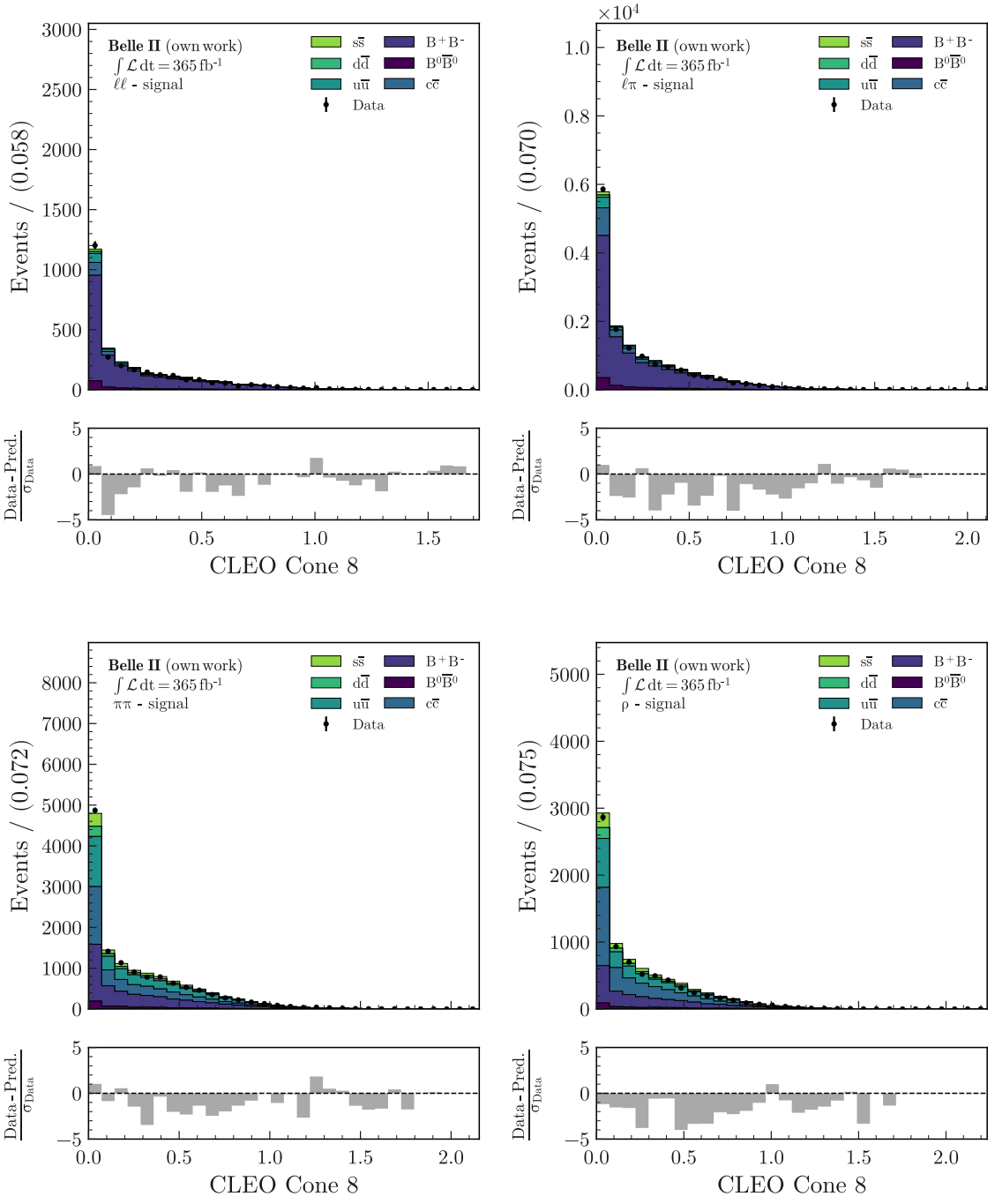


Figure A.18: Comparison of the variable CLEO Cone 8 on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

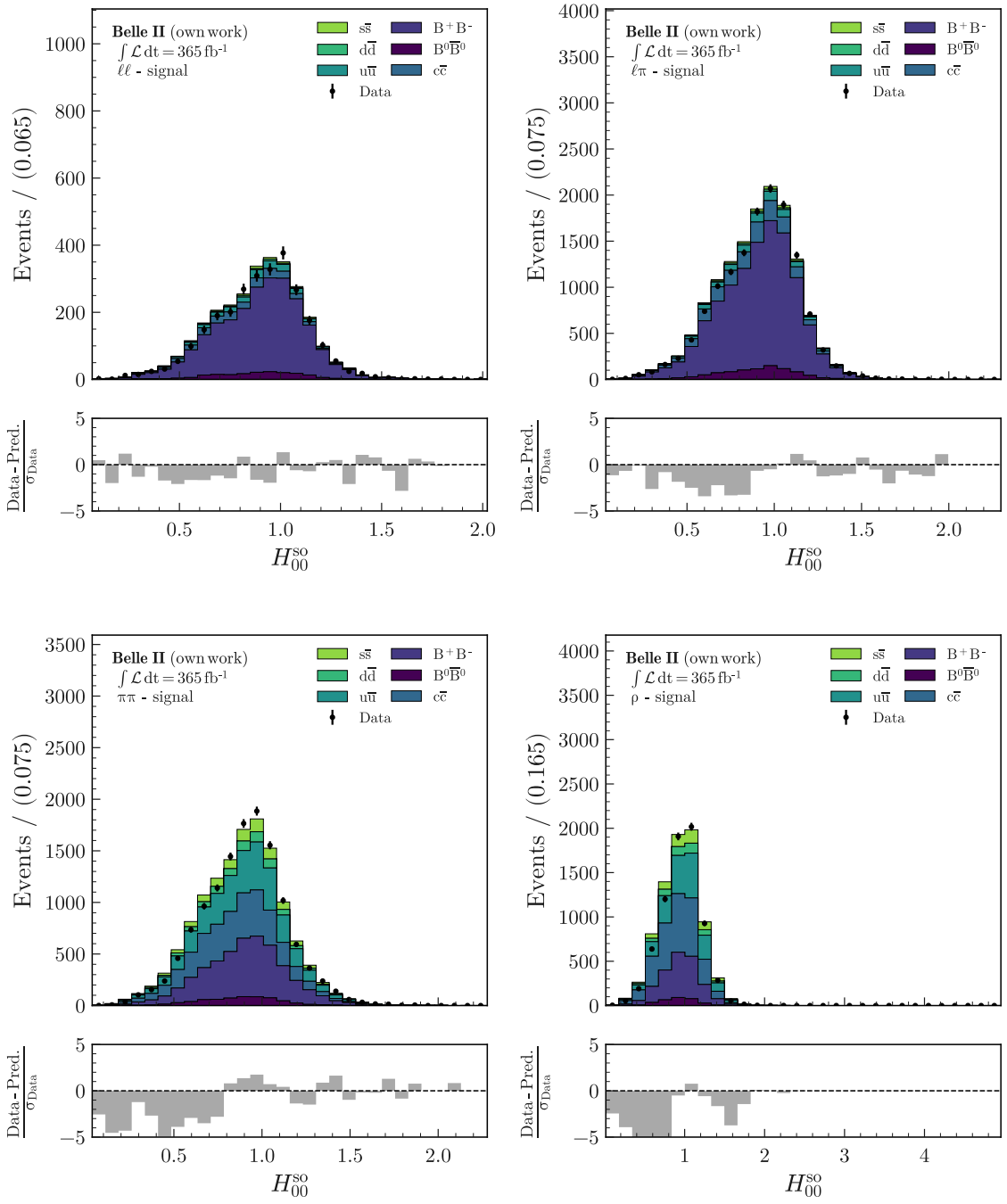


Figure A.19: Comparison of the variable H_{00}^{so} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

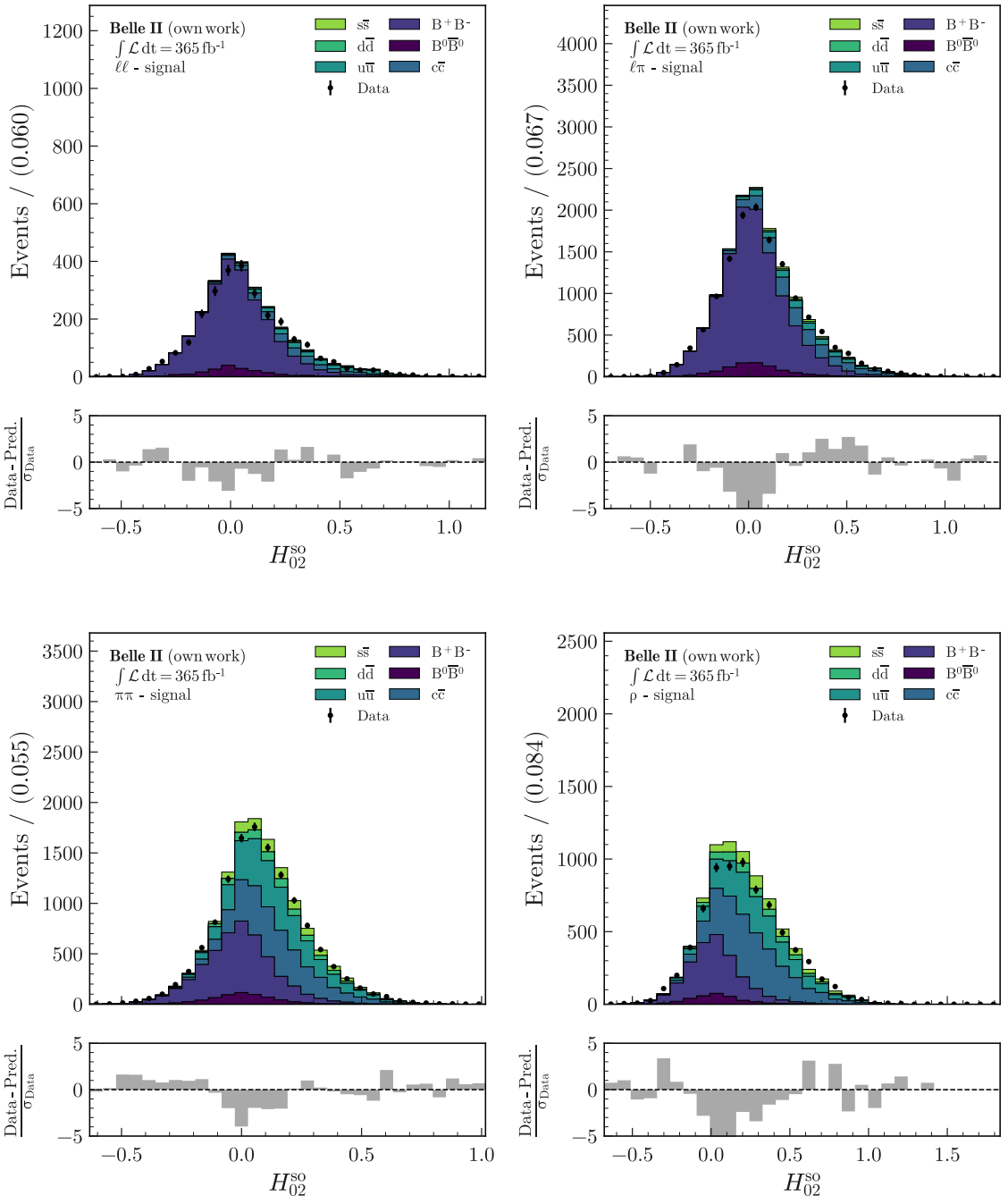


Figure A.20: Comparison of the variable H_{02}^{so} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

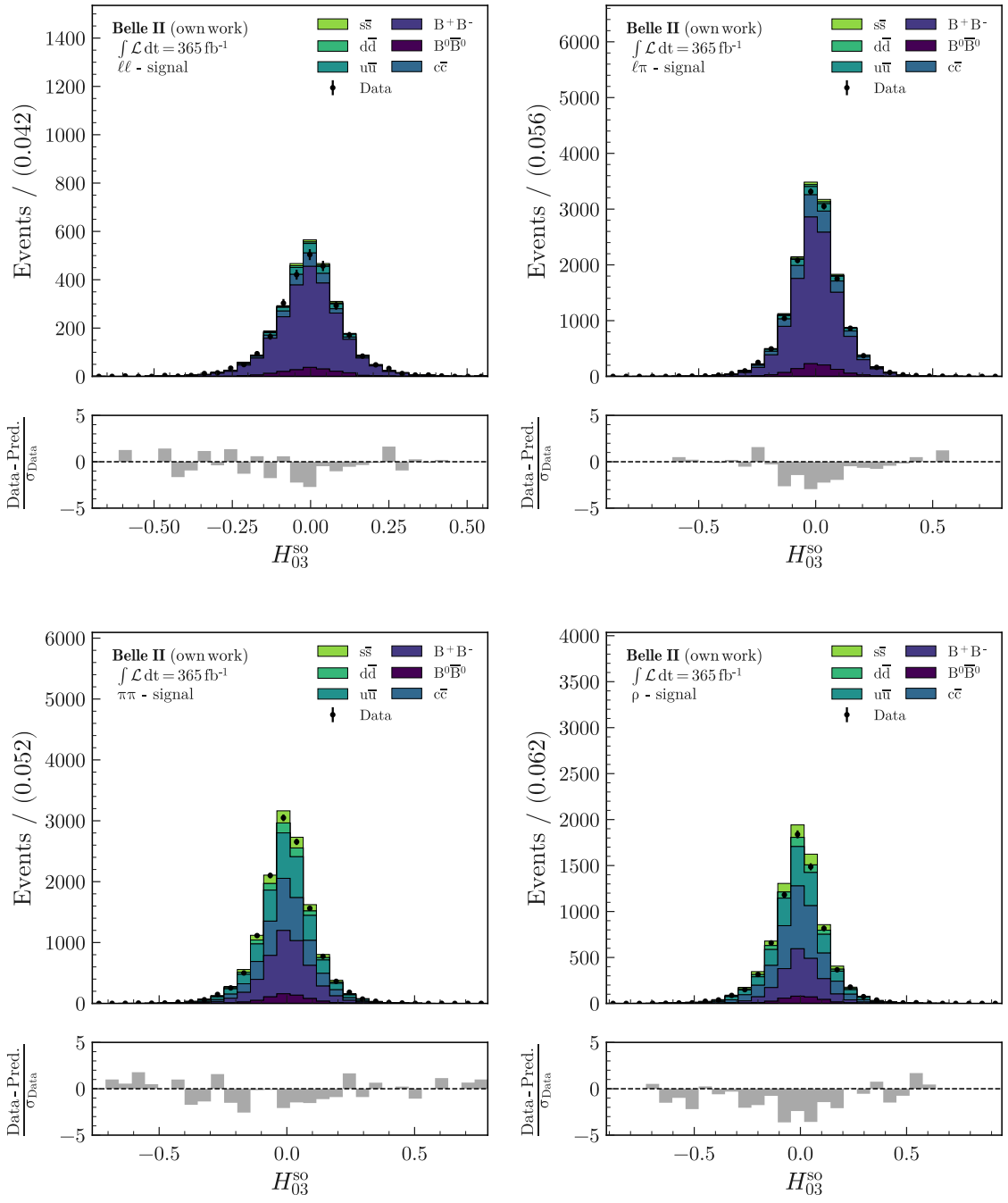


Figure A.21: Comparison of the variable H_{03}^{so} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

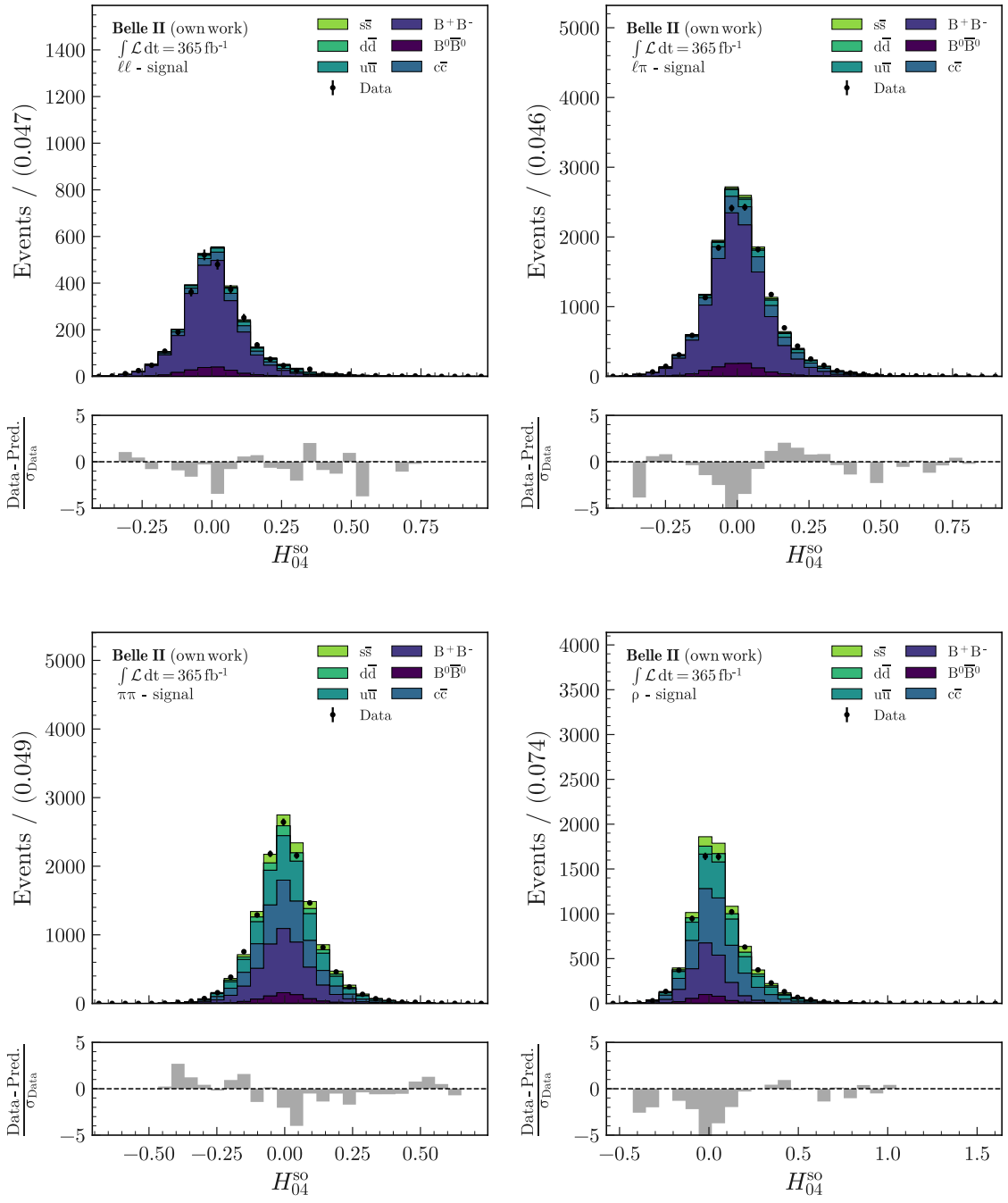


Figure A.22: Comparison of the variable H_{04}^{so} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

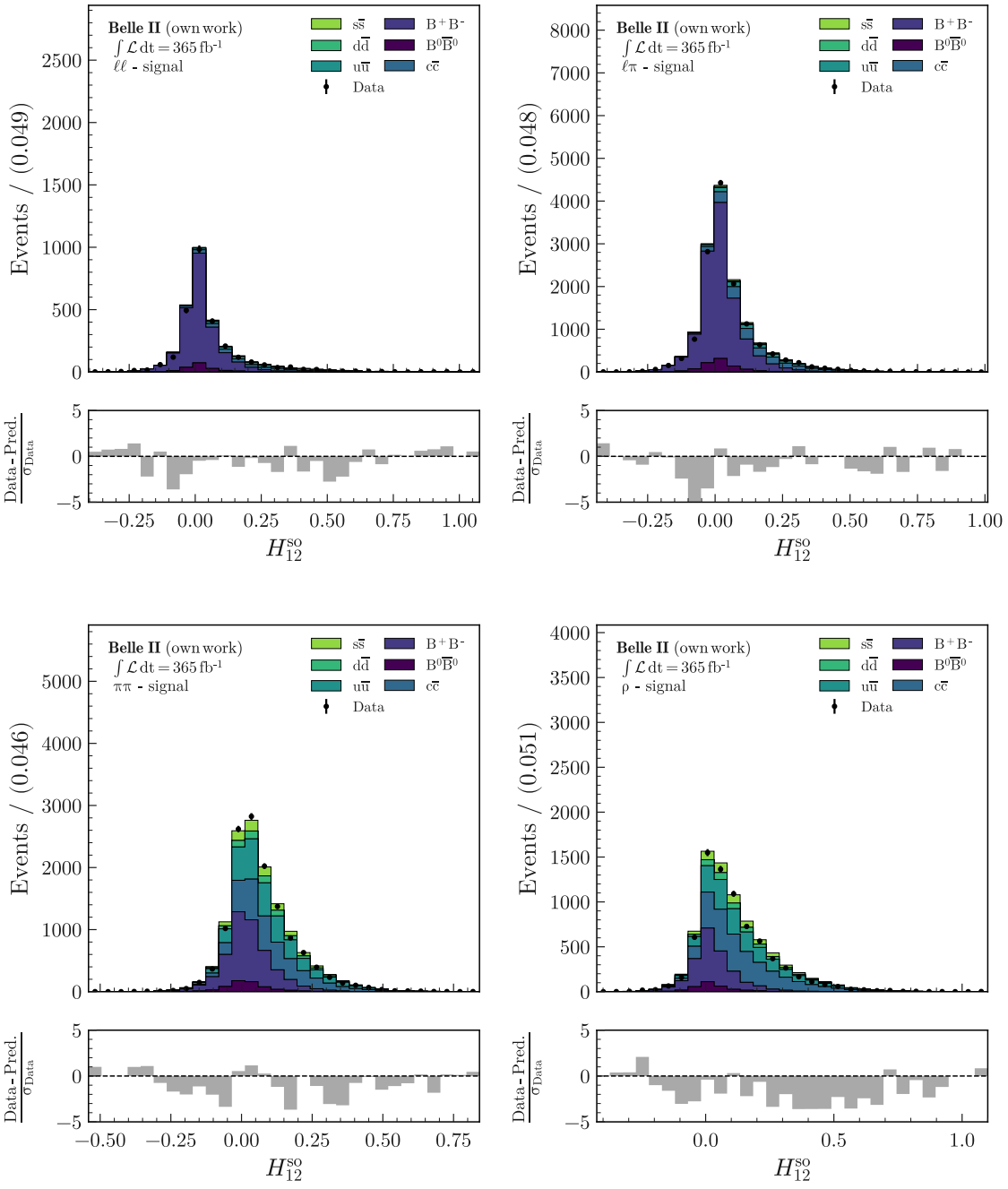


Figure A.23: Comparison of the variable H_{12}^{so} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

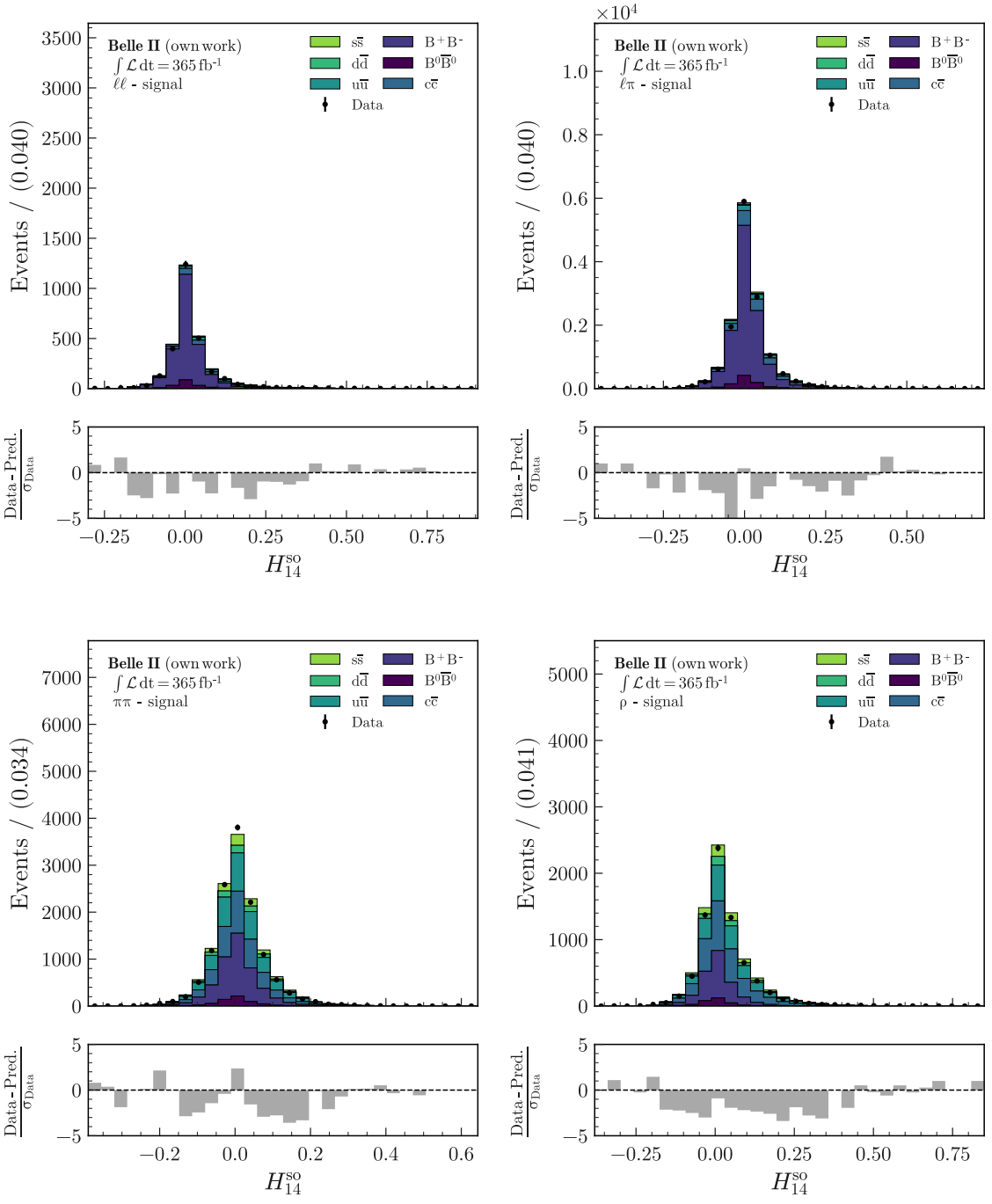


Figure A.24: Comparison of the variable H_{14}^{so} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

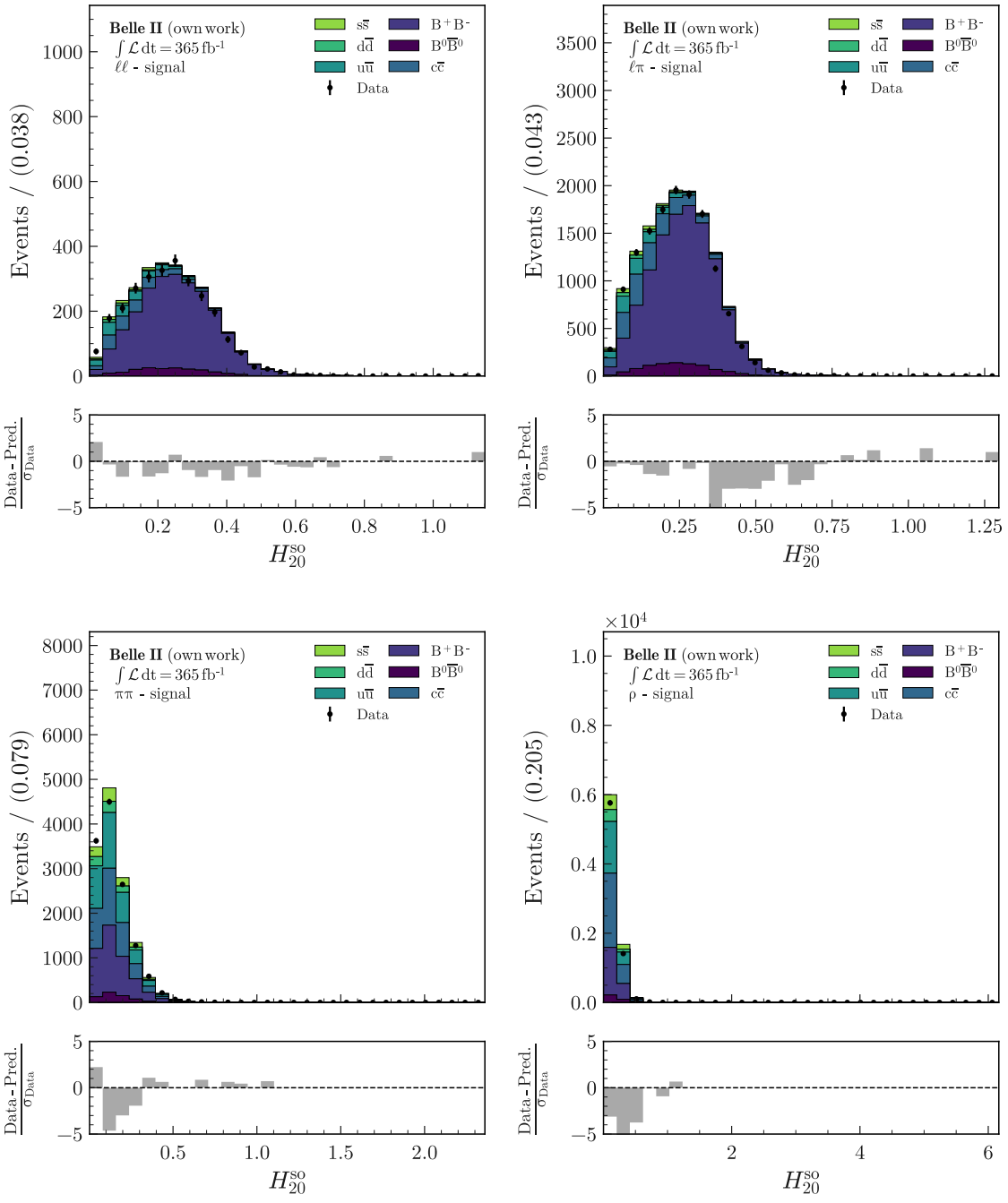


Figure A.25: Comparison of the variable H_{20}^{so} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

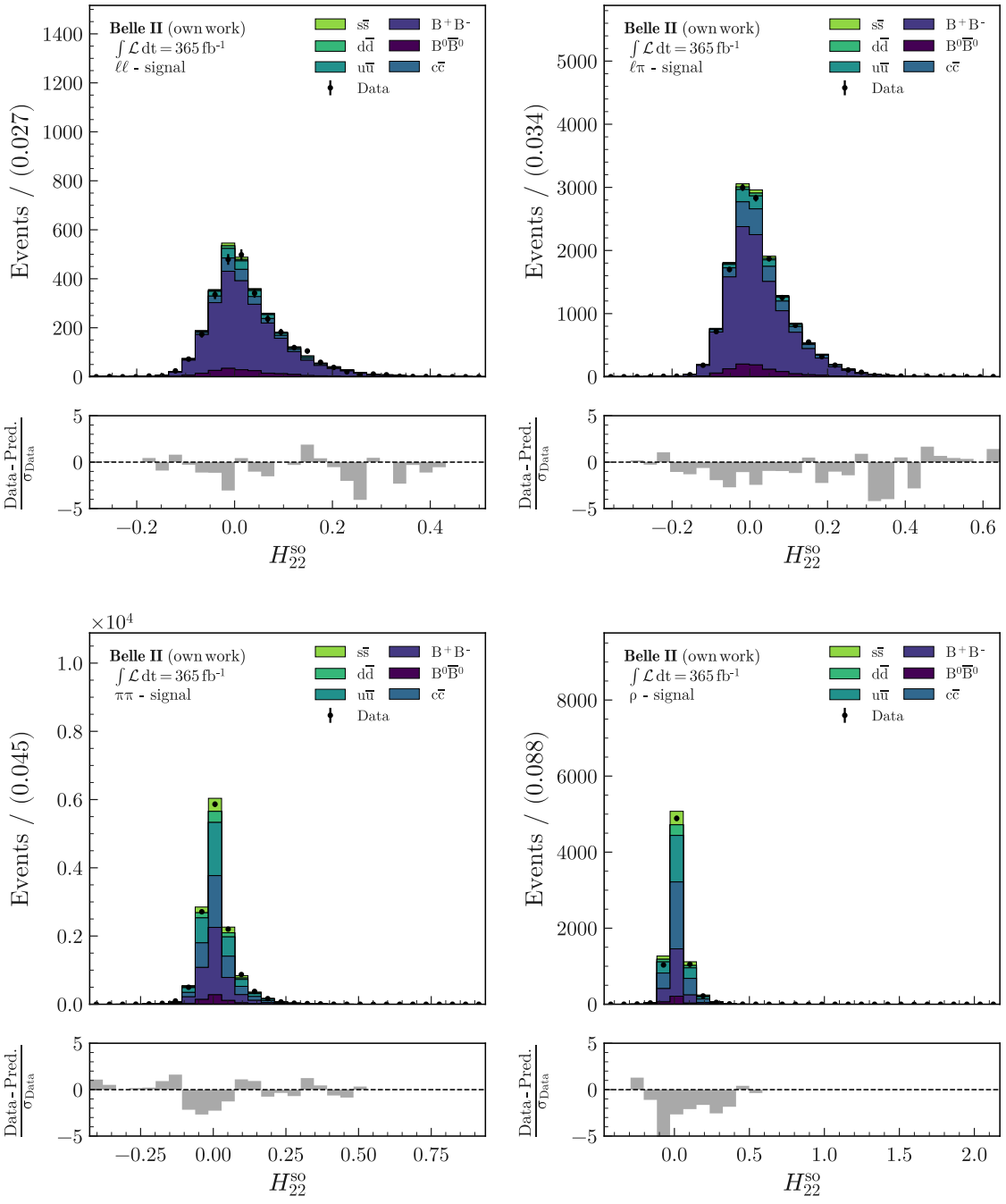


Figure A.26: Comparison of the variable H_{22}^{so} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

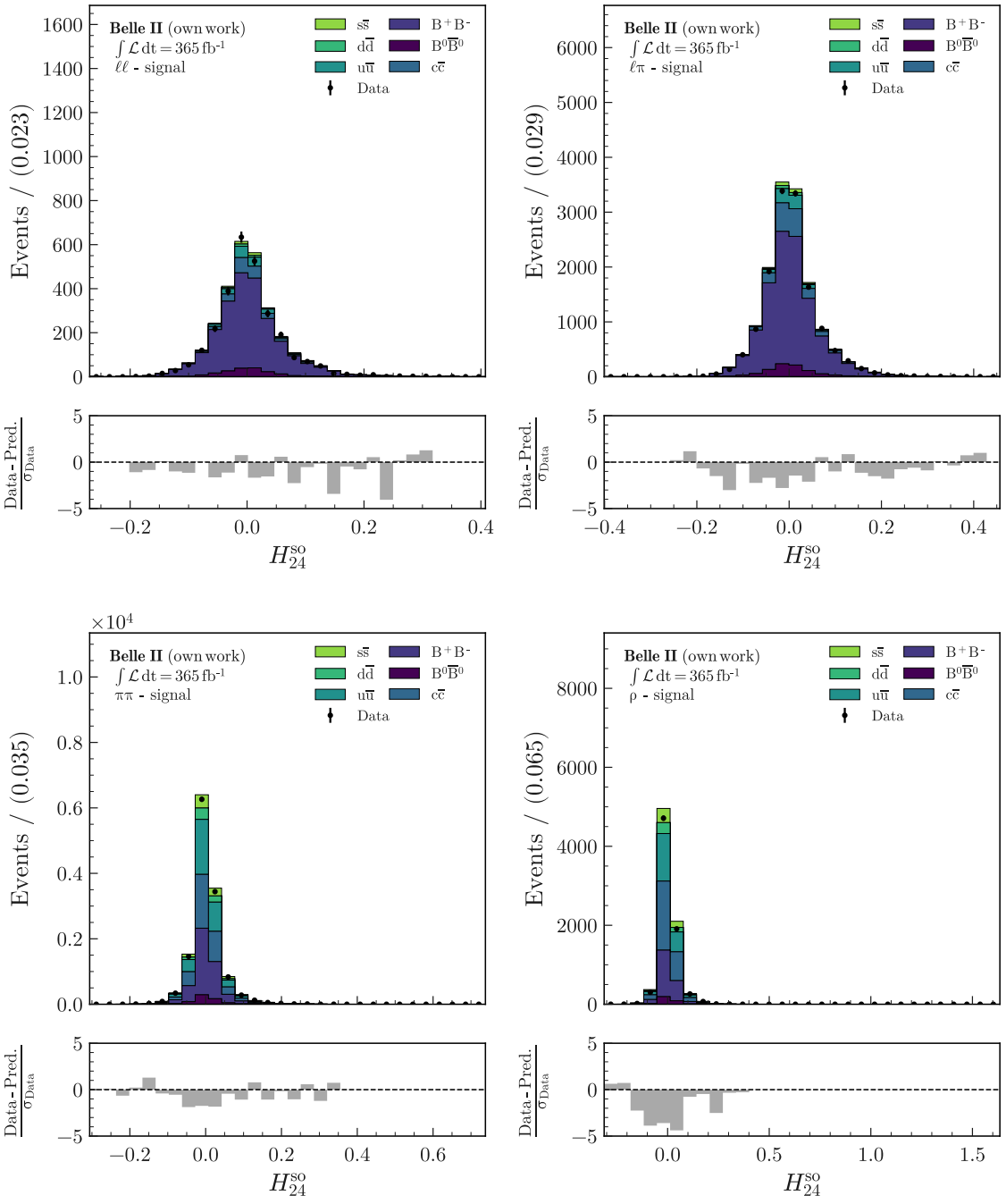


Figure A.27: Comparison of the variable H_{24}^{so} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

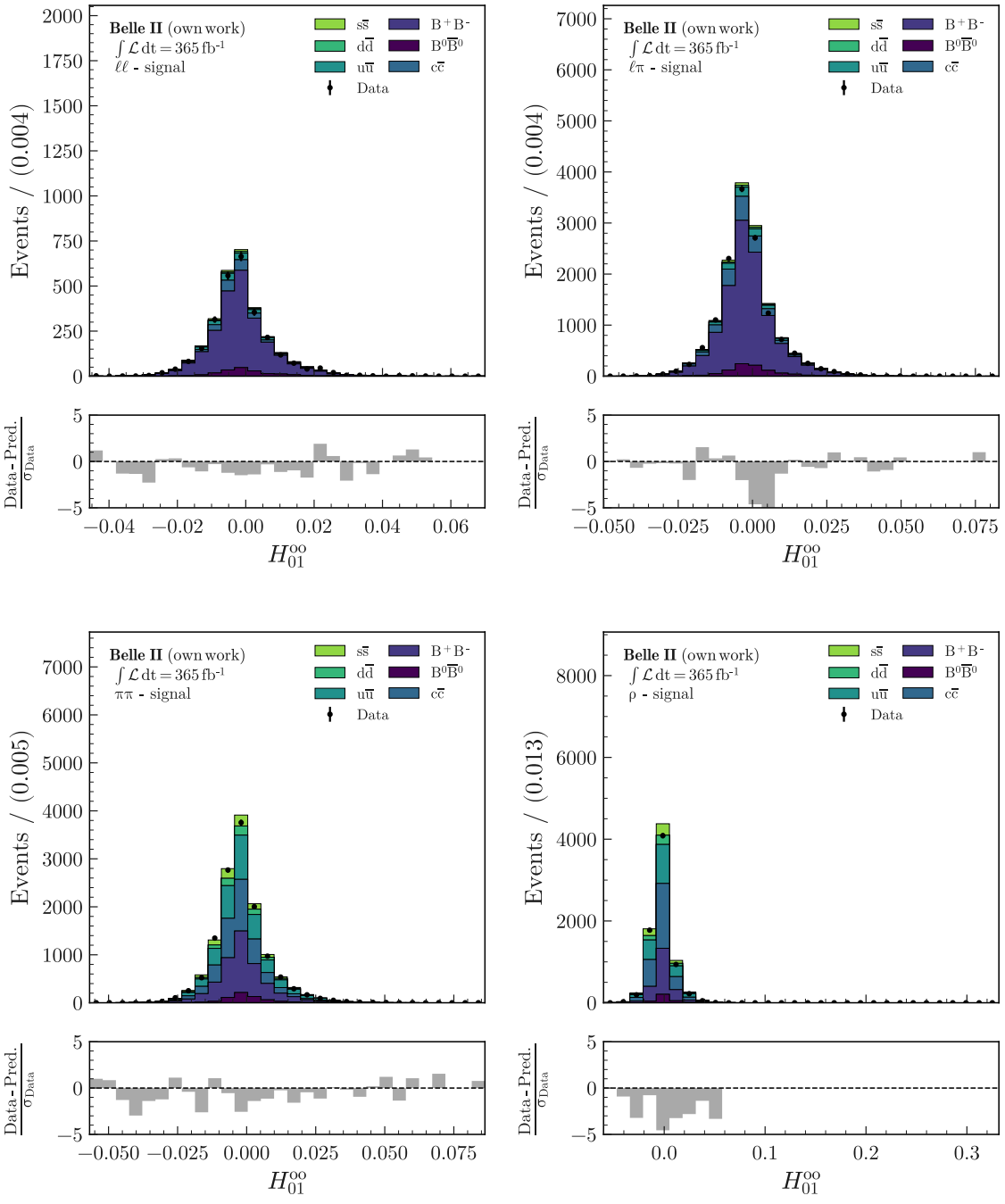


Figure A.28: Comparison of the variable H_{01}^{oo} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

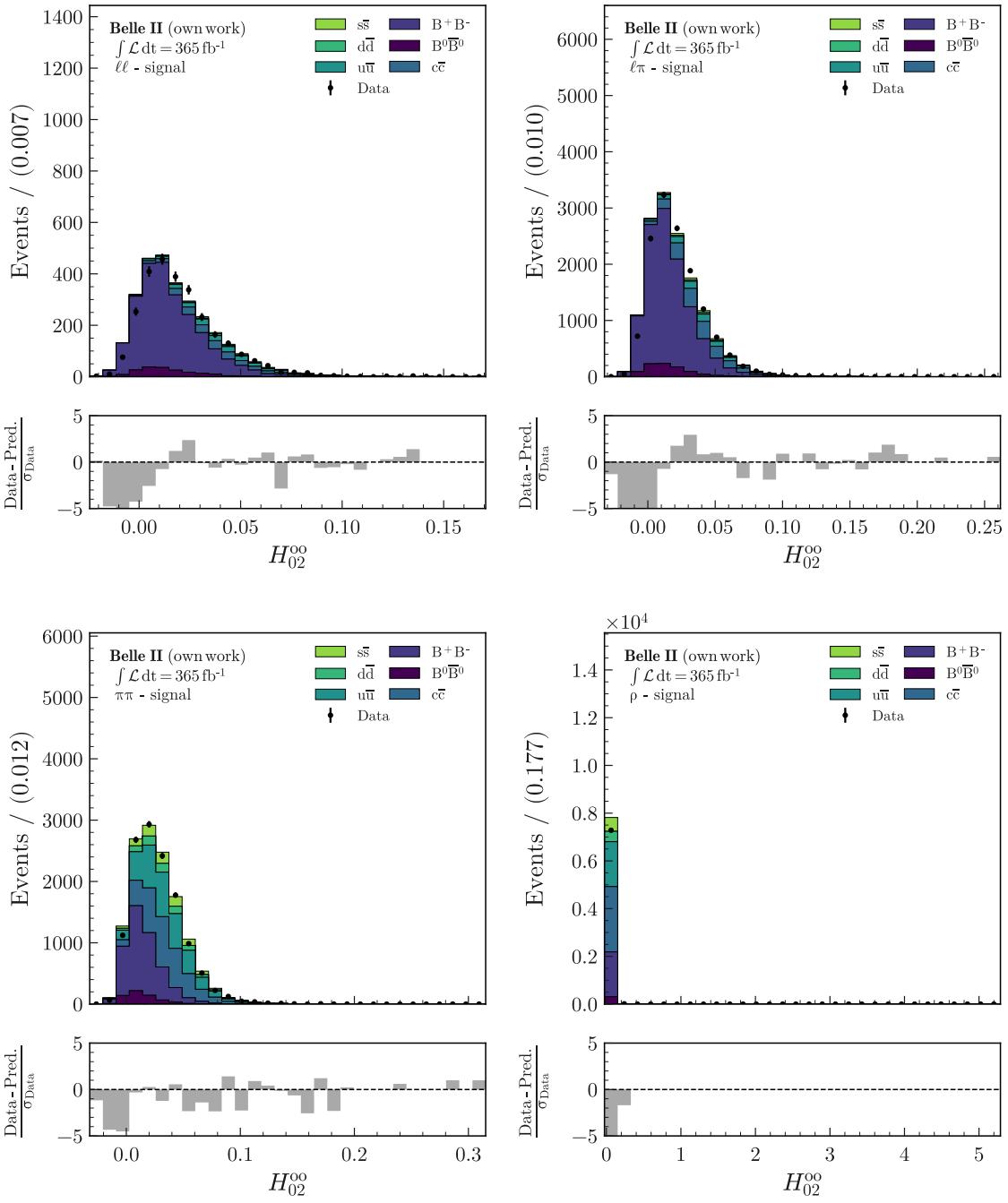


Figure A.29: Comparison of the variable H_{02}^{oo} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

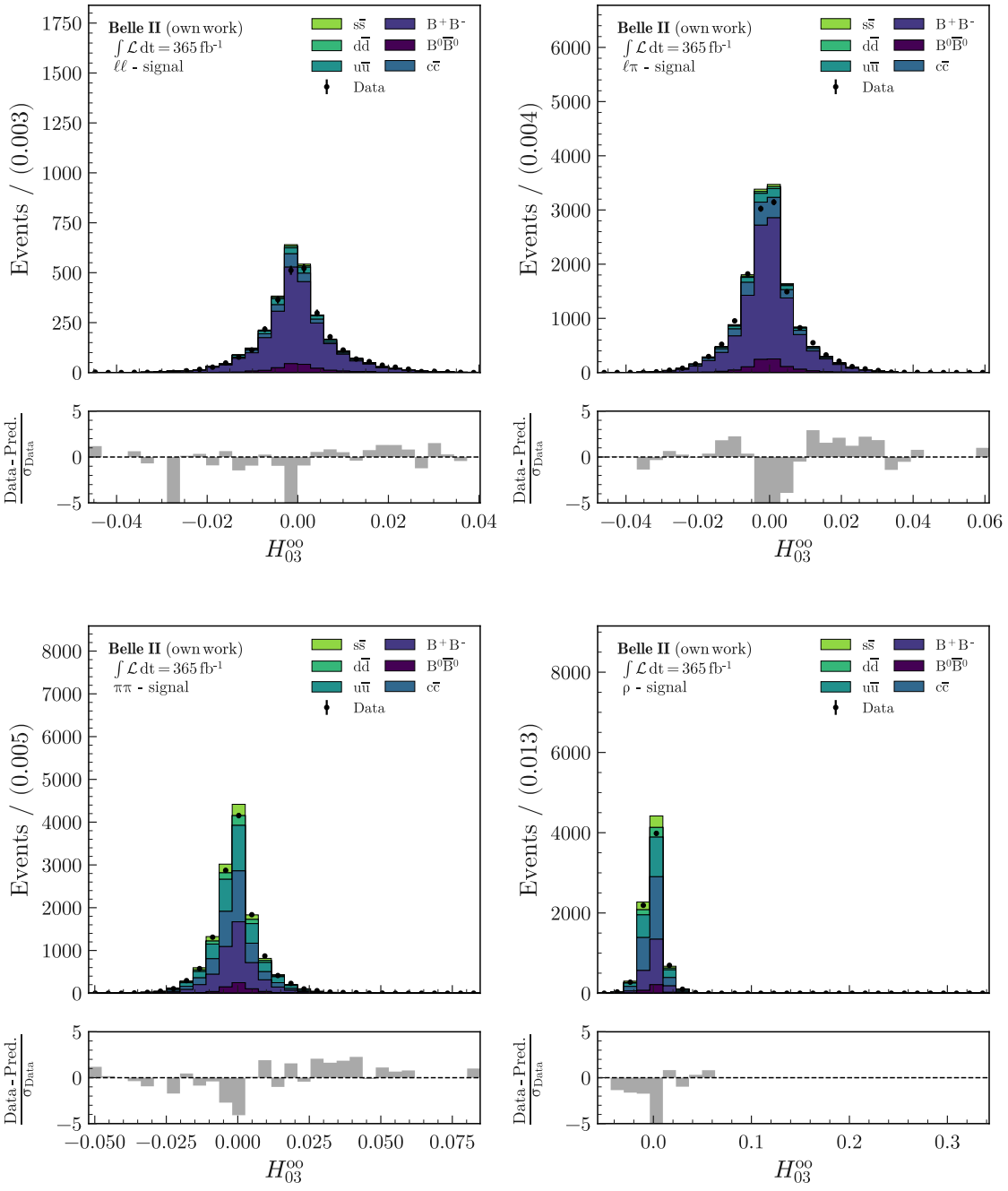


Figure A.30: Comparison of the variable H_{03}^{oo} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

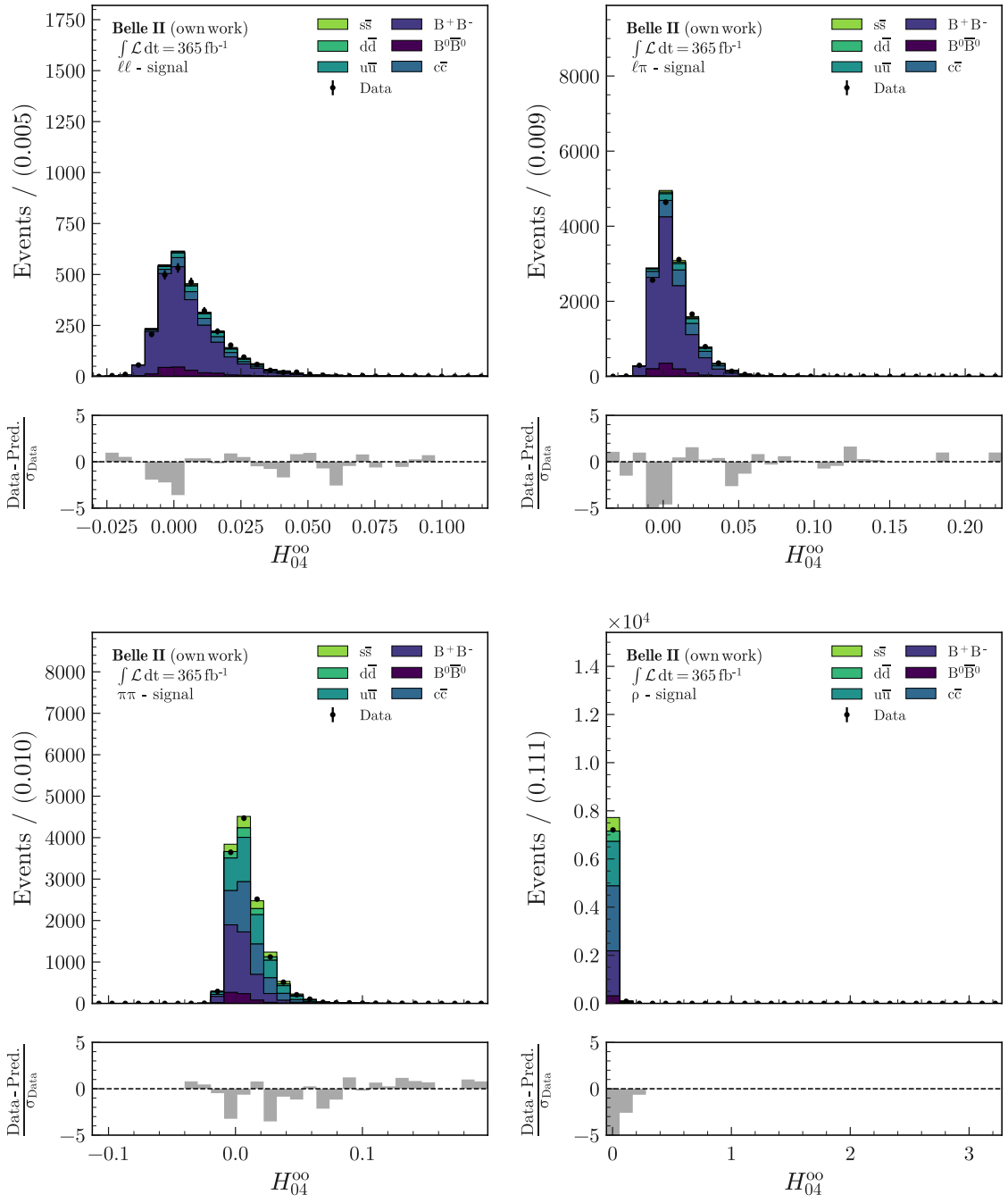


Figure A.31: Comparison of the variable H_{04}^{oo} on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

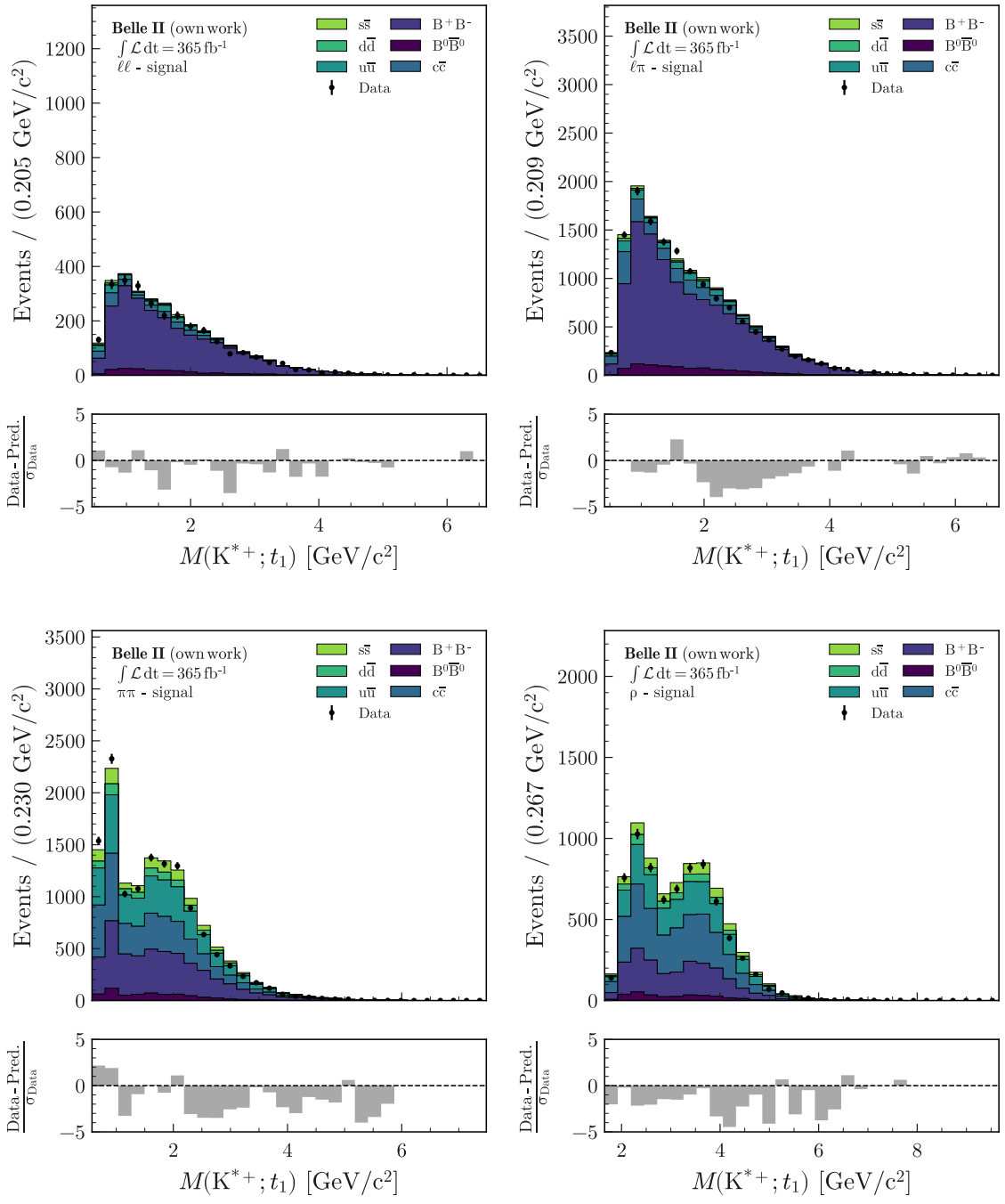


Figure A.32: Comparison of the variable $M(\mathrm{K}^{*+};t_1)$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

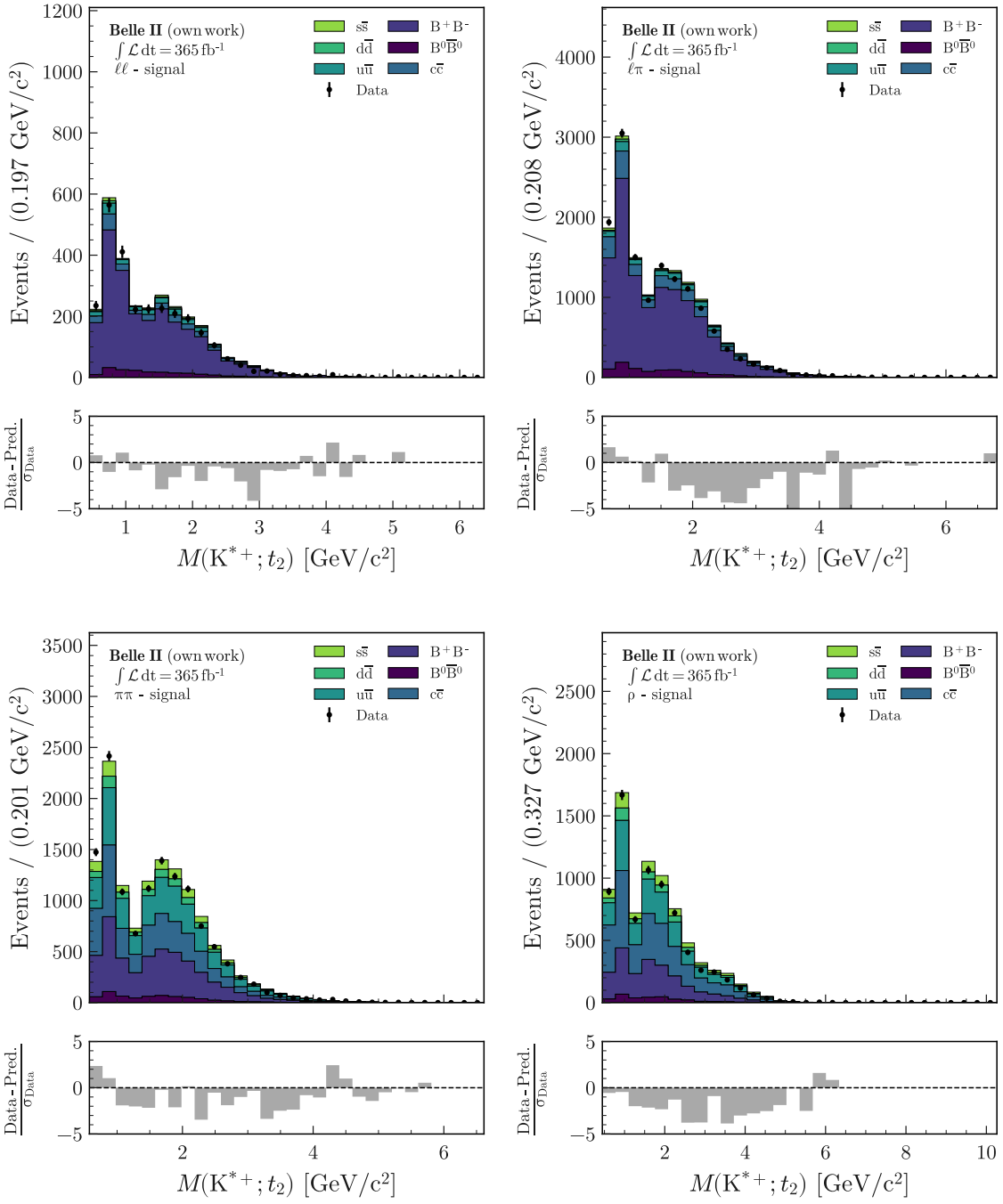


Figure A.33: Comparison of the variable $M(\mathrm{K}^{*+};t_2)$ on the K^{*+} mass sideband for simulated events and recorded data for all four signal channels. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in Data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

A.5 Two-Dimensional Sideband Validation Plots

In this section of the appendix, I present additional 2D plots of recorded and simulated data on the K^{*+} sideband used for validation purposes. Given the large number of variables used in the MVA, including all plots in this work is not feasible. However, a selection is shown here, based on the variables that are used in every signal channel: $p_{\text{miss}}^{\text{CMS}}$, $E_{\text{miss}}^{\text{CMS}}$, $\cos\left(\text{Thrust}_{\text{B}}; \text{Thrust}_{\text{ROE}}\right)$, Thrust_{B}, H_{22}^{so}, and $M(K^{*+}; t_1)$. All combinations for all channels are depicted on the following pages. The entire collection of 2D validation plots is available as a Zenodo record (see Ref. [75]).

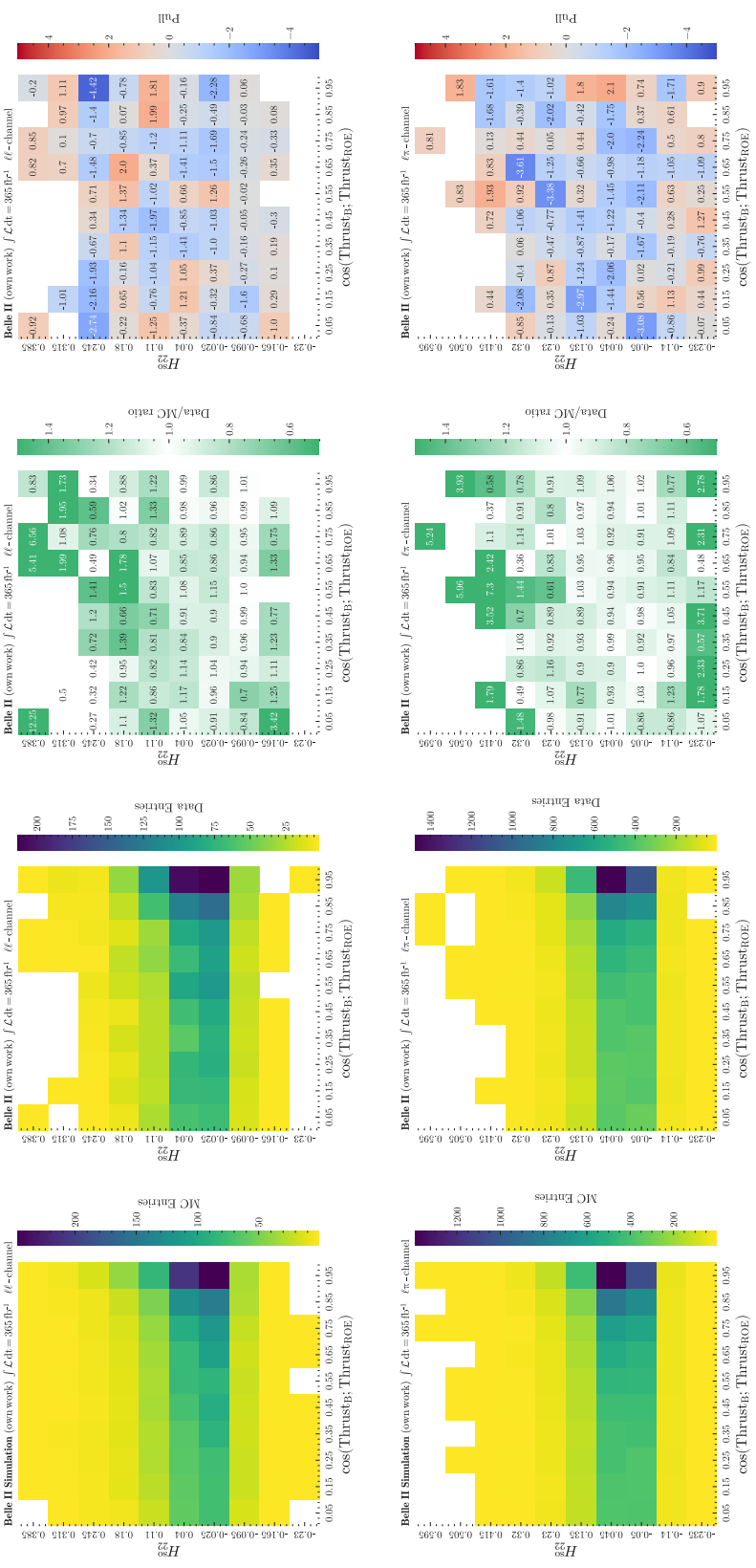


Figure A.34: Two-dimensional comparison of recorded data and simulation in the variables $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ and H_{22}^{so} on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

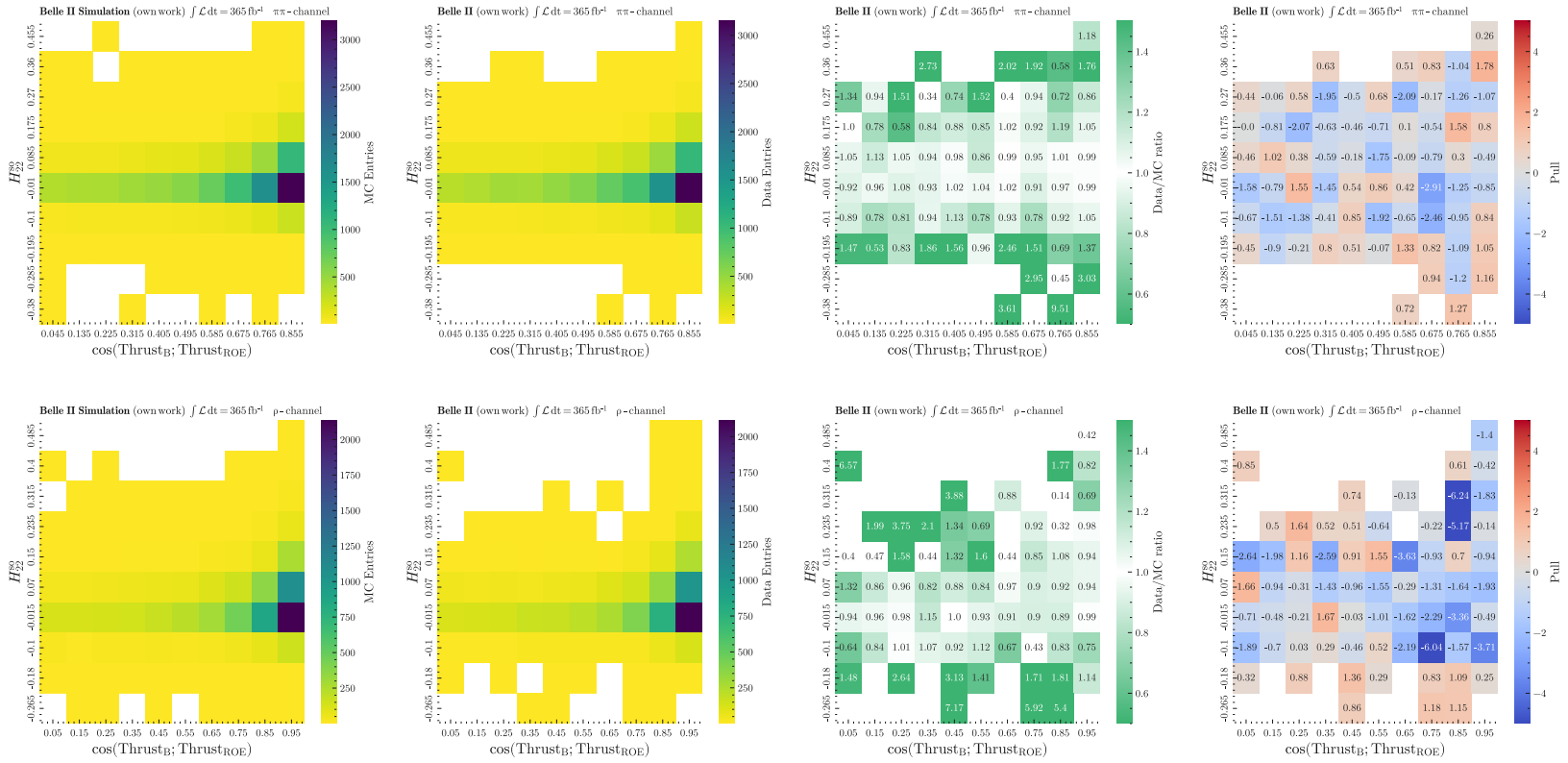


Figure A.35: Two-dimensional comparison of recorded data and simulation in the variables $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ and H_{22}^{so} on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

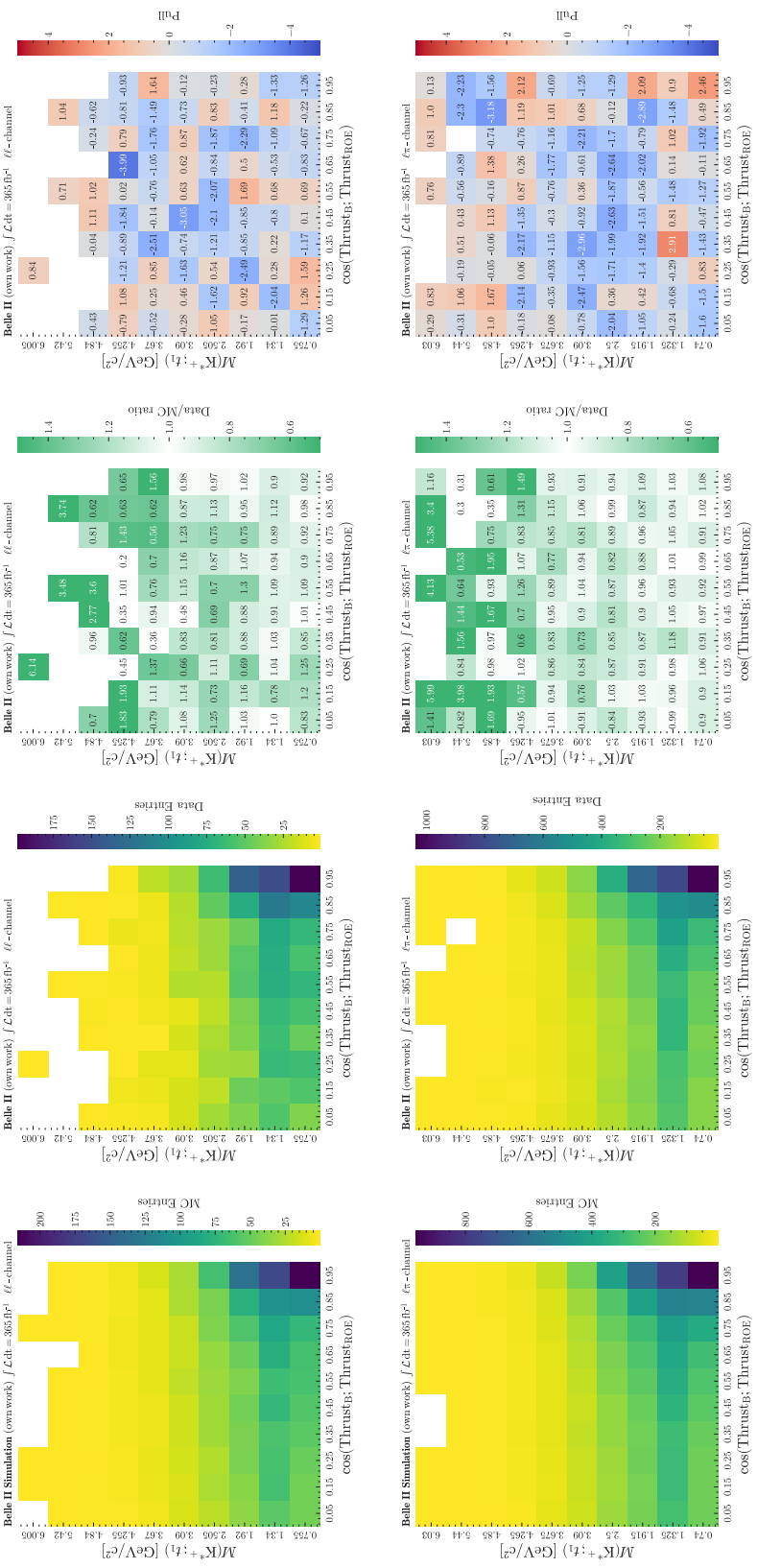


Figure A.36: Two-dimensional comparison of recorded data and simulation in the variables cos (Thrust$_B$; Thrust$_{ROE}$) and $M(K^{*+};t_1)$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

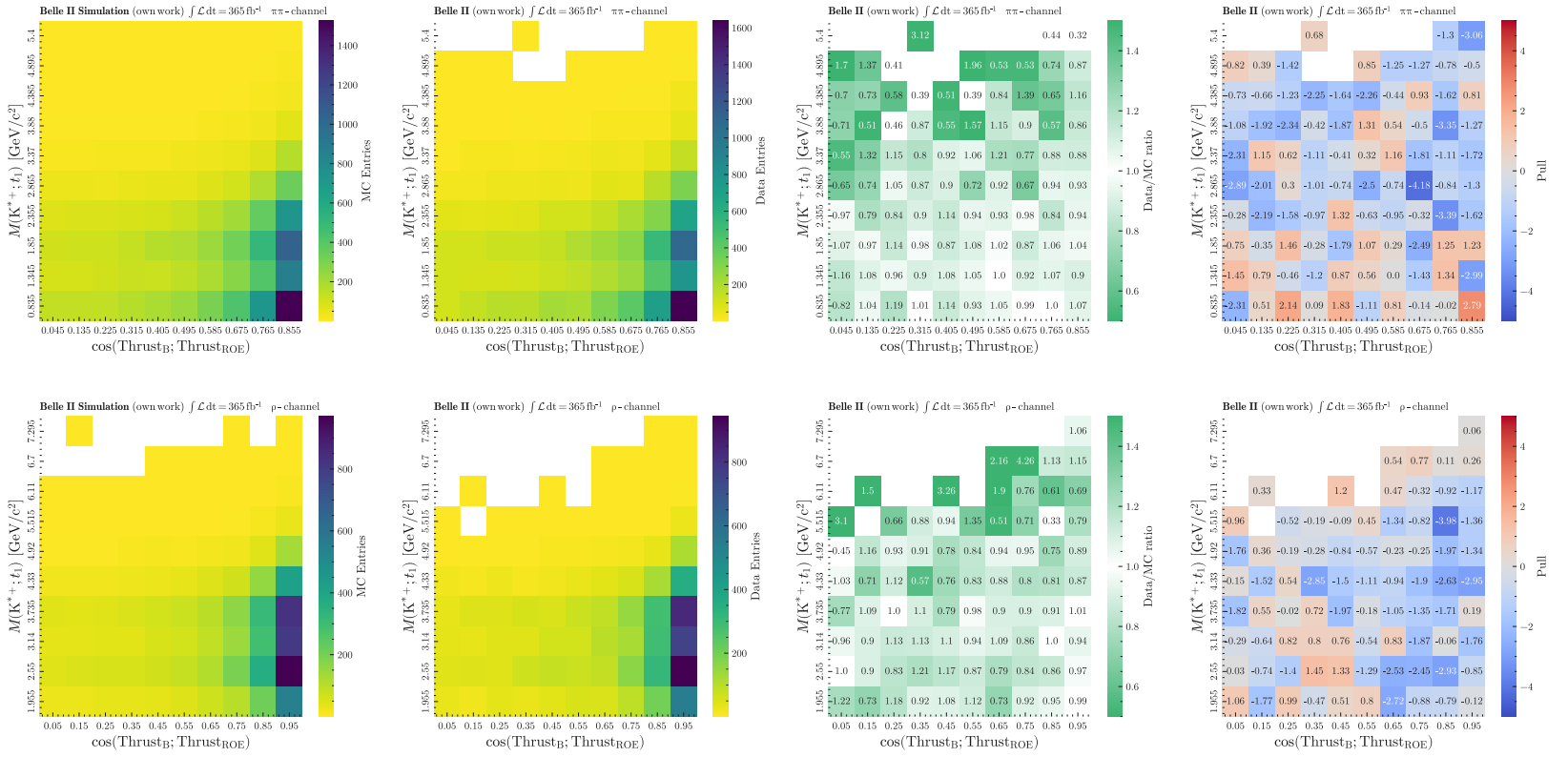


Figure A.37: Two-dimensional comparison of recorded data and simulation in the variables $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ and $M(\mathrm{K}^{*+}; t_1)$ on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

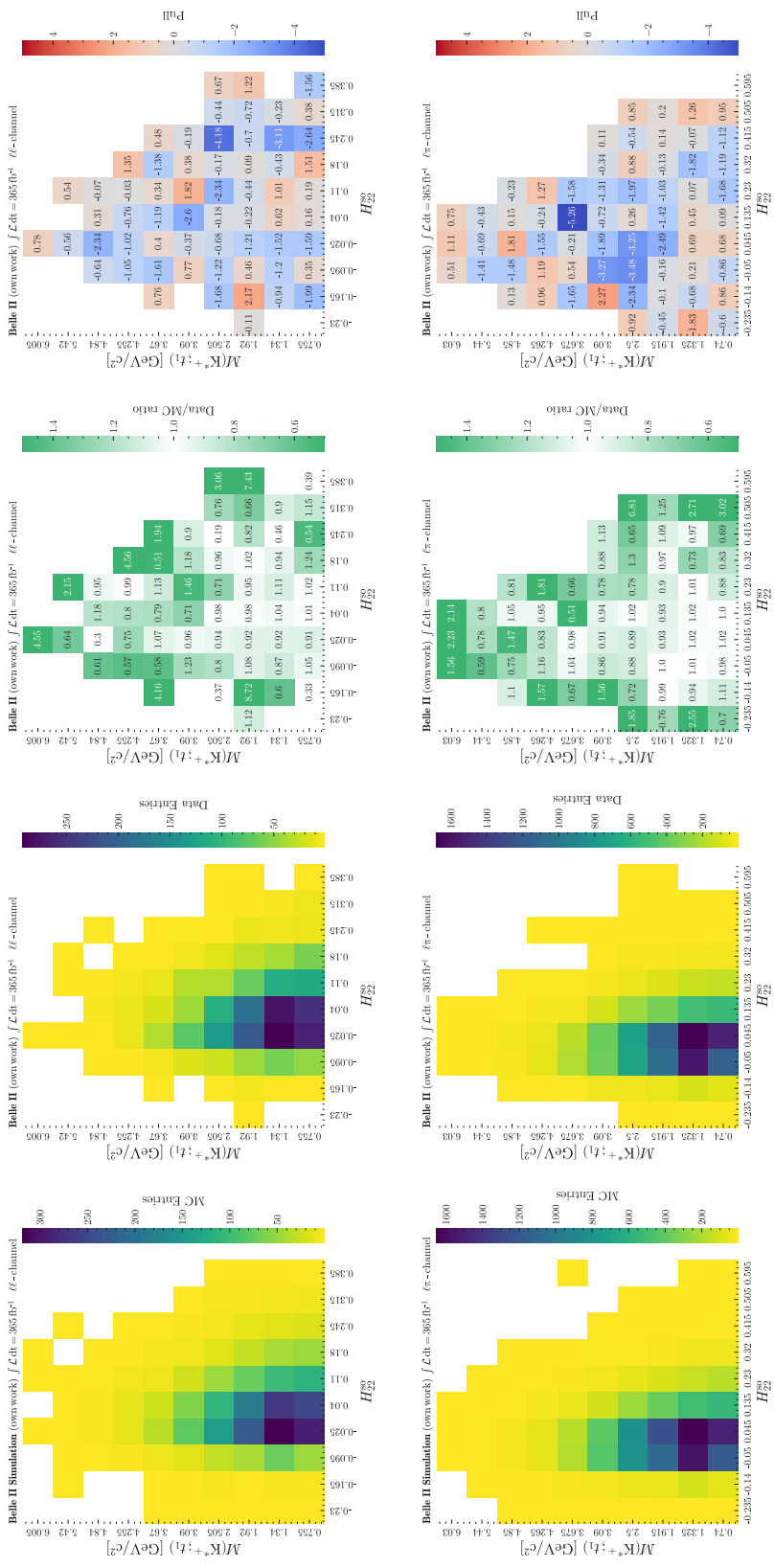


Figure A.38: Two-dimensional comparison of recorded data and simulation in the variables H_{22}^{so} and $M(K^{*+};t_1)$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

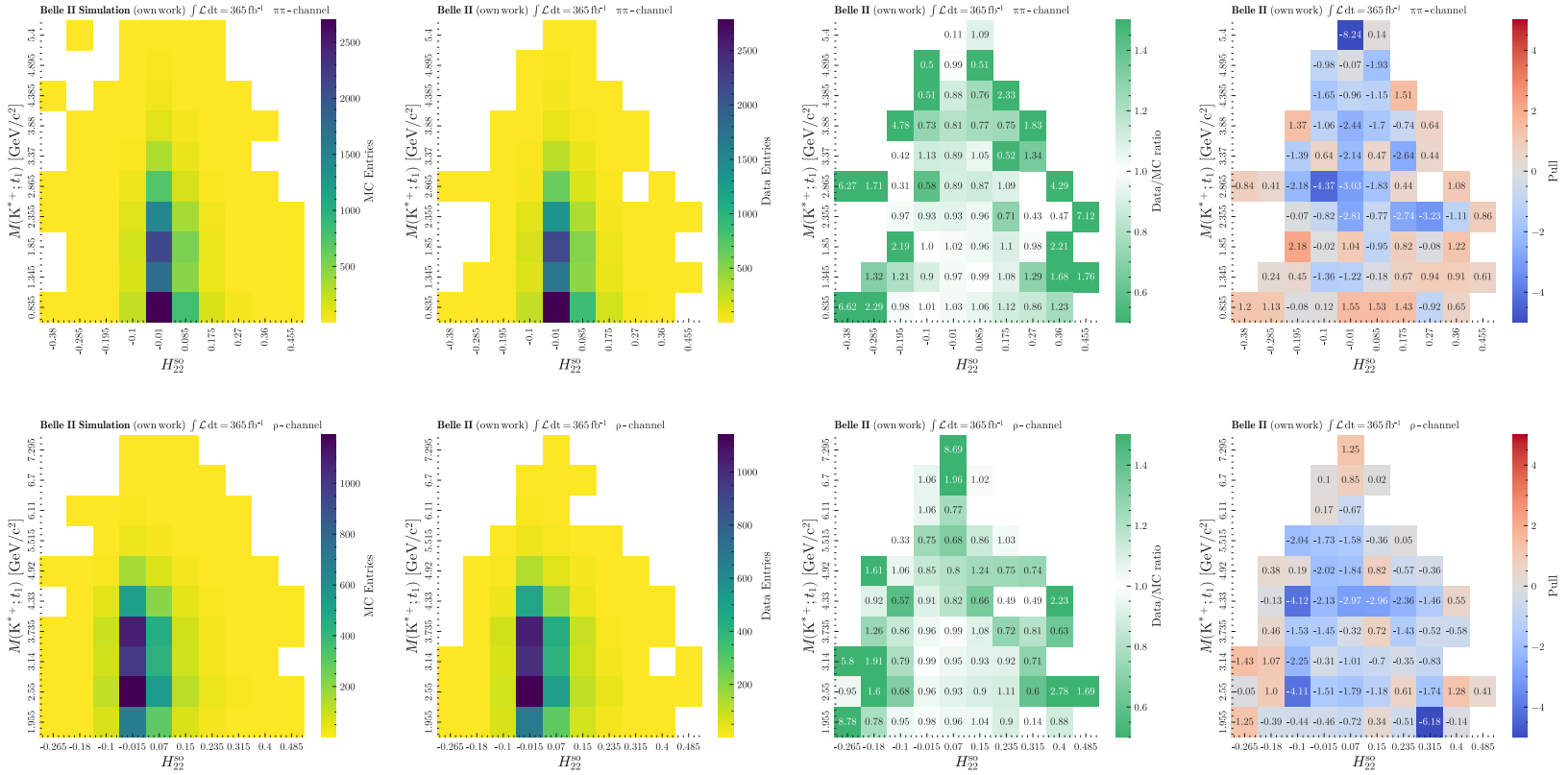


Figure A.39: Two-dimensional comparison of recorded data and simulation in the variables H_{22}^{so} and $M(K^{*+};t_1)$ on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

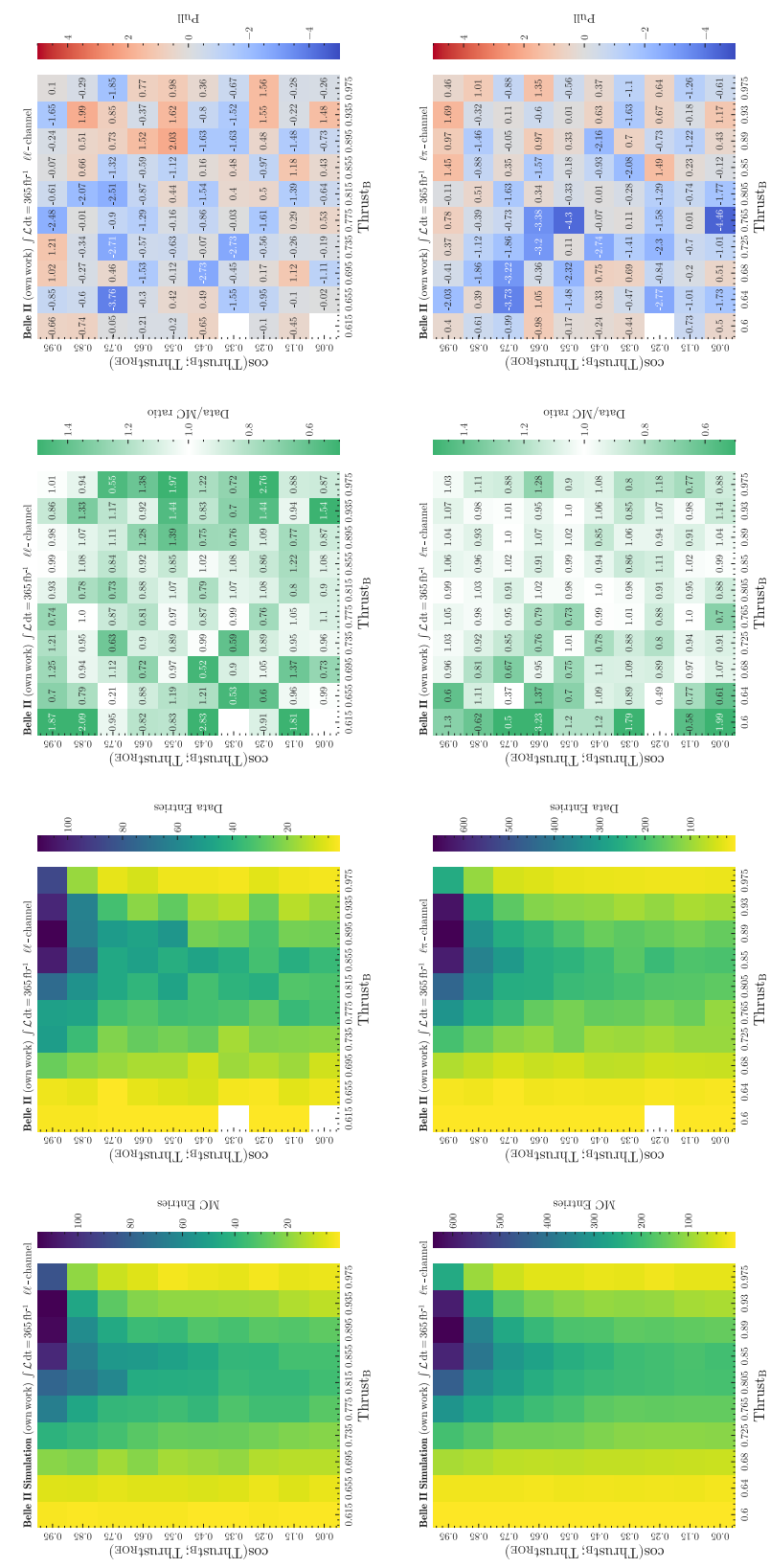


Figure A.40: Two-dimensional comparison of recorded data and simulation in the variables $\mathrm{Thrust_B}$ and $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

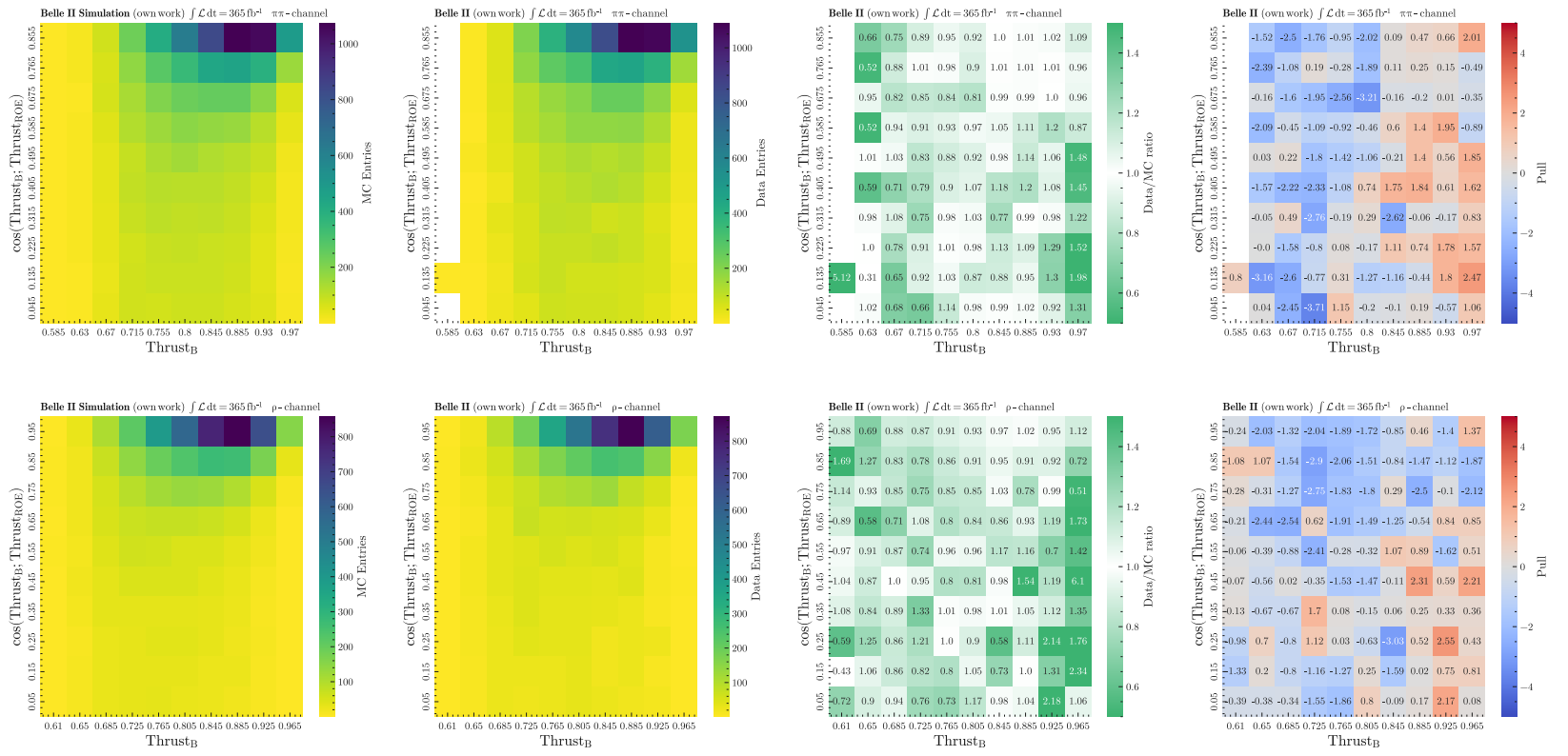


Figure A.41: Two-dimensional comparison of recorded data and simulation in the variables $\mathrm{Thrust_B}$ and $\cos\left(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}}\right)$ on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

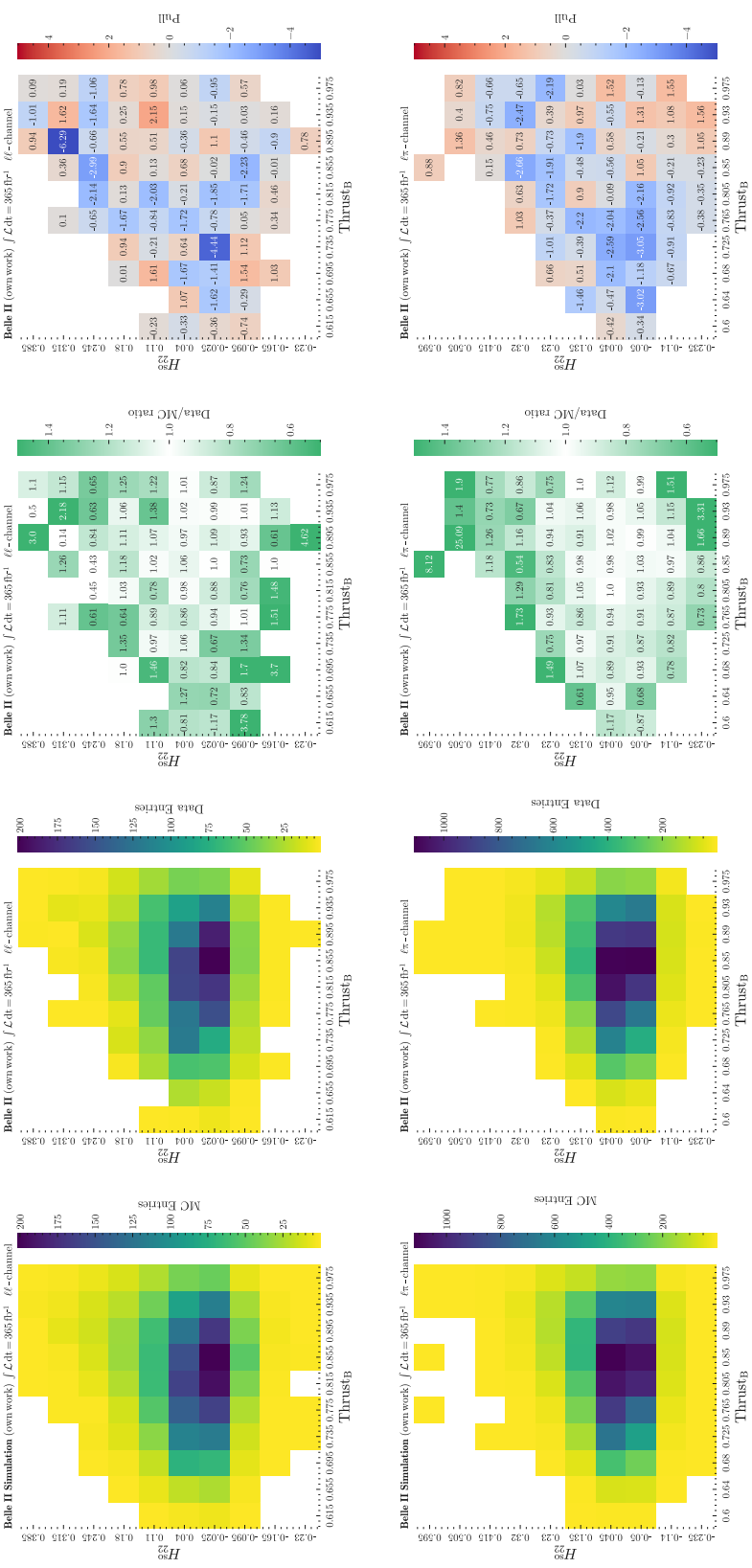


Figure A.42: Two-dimensional comparison of recorded data and simulation in the variables $\mathrm{Thrust_B}$ and H_{22}^{so} on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

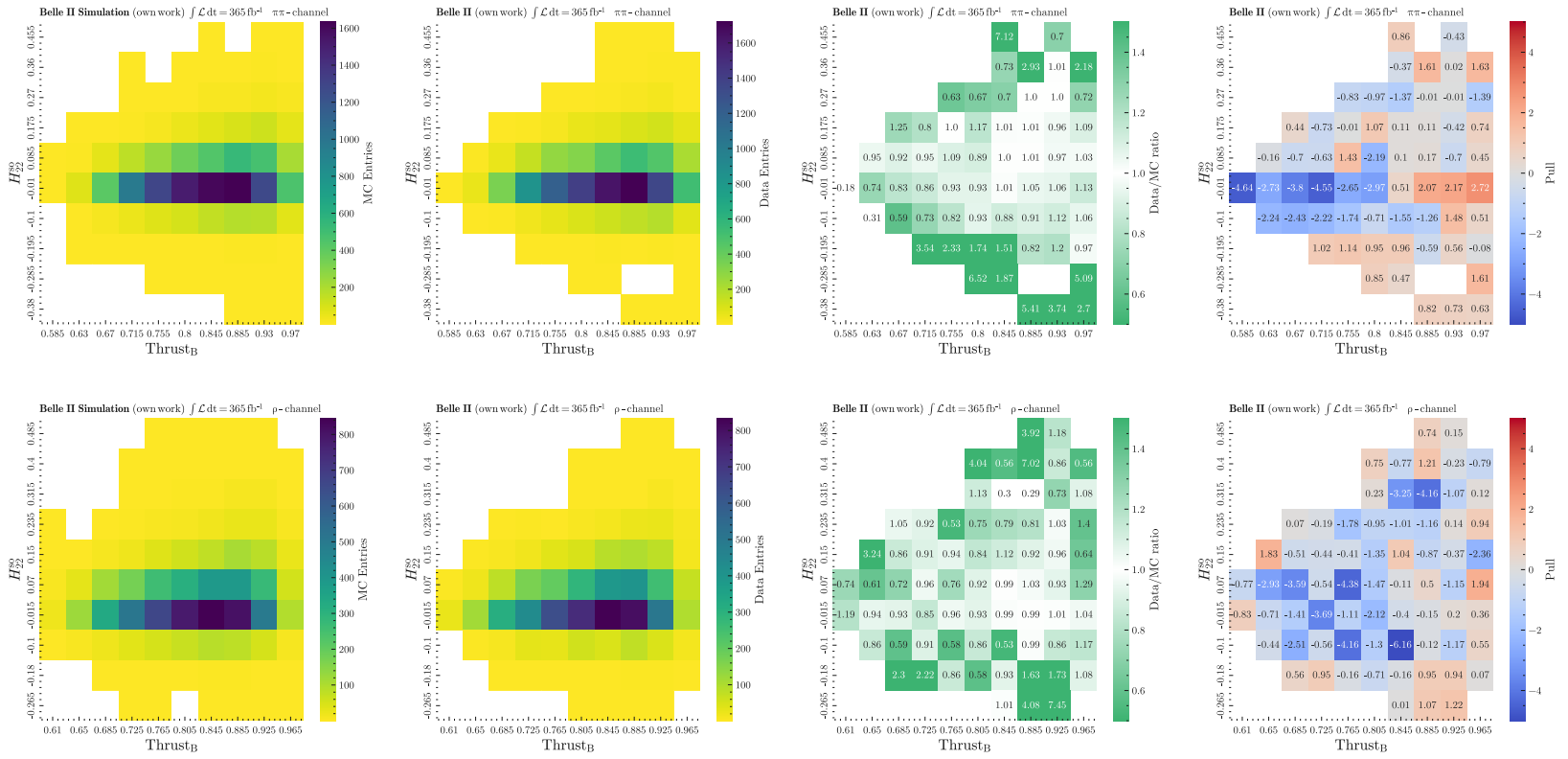


Figure A.43: Two-dimensional comparison of recorded data and simulation in the variables $\mathrm{Thrust_B}$ and H_{22}^{so} on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

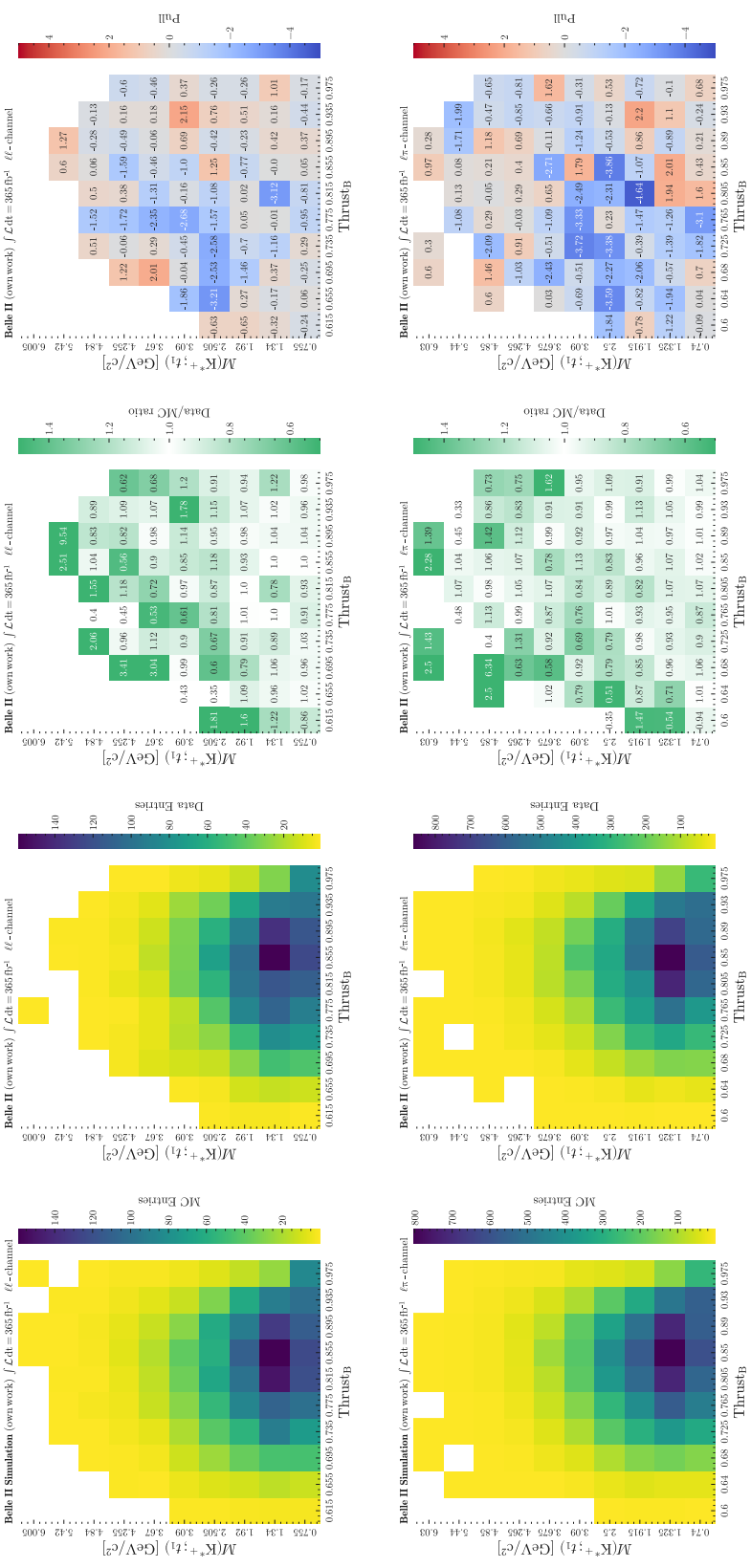


Figure A.44: Two-dimensional comparison of recorded data and simulation in the variables $\mathrm{Thrust_B}$ and $M(\mathrm{K}^{*+};t_1)$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

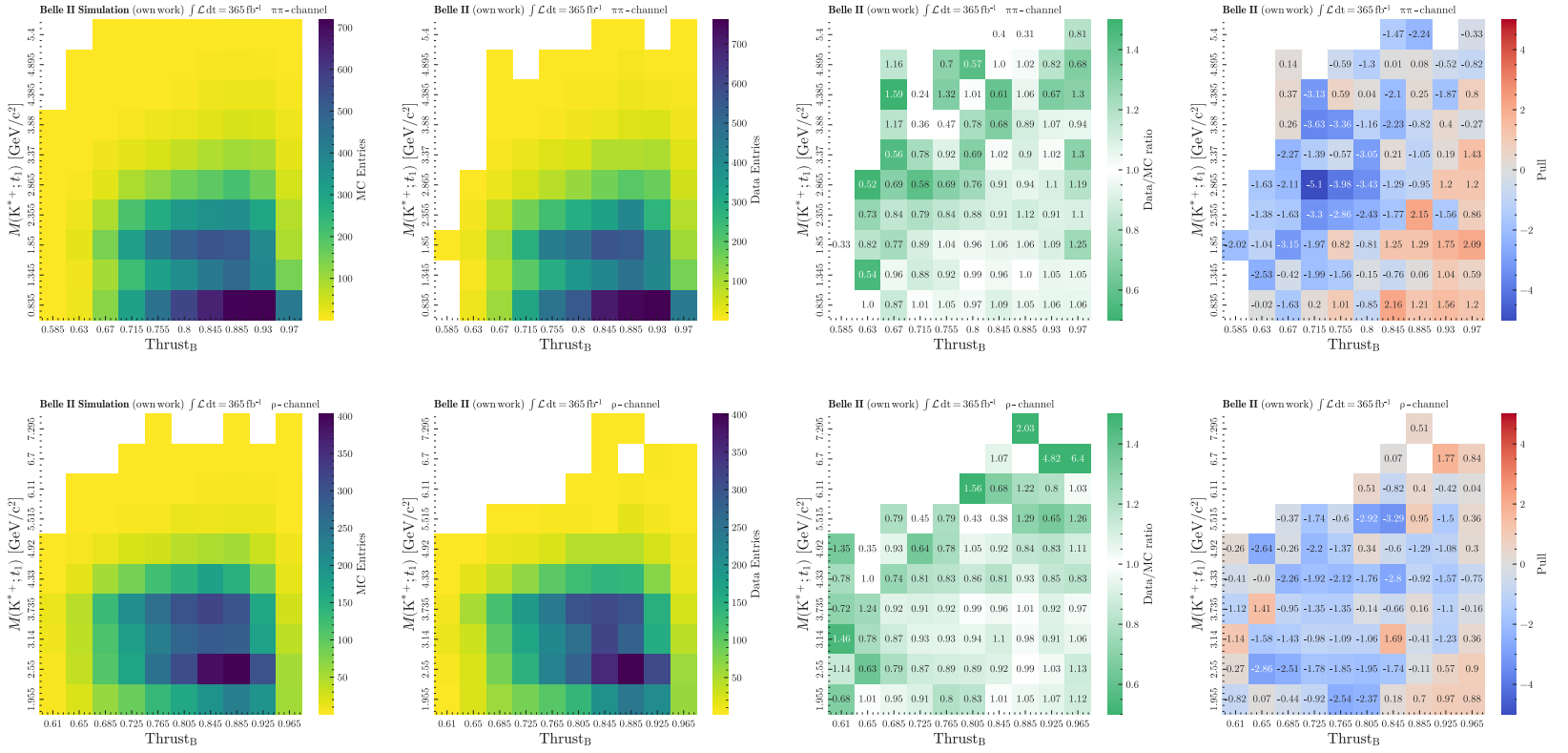


Figure A.45: Two-dimensional comparison of recorded data and simulation in the variables $\mathrm{Thrust_B}$ and $M(\mathrm{K}^{*+};t_1)$ on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

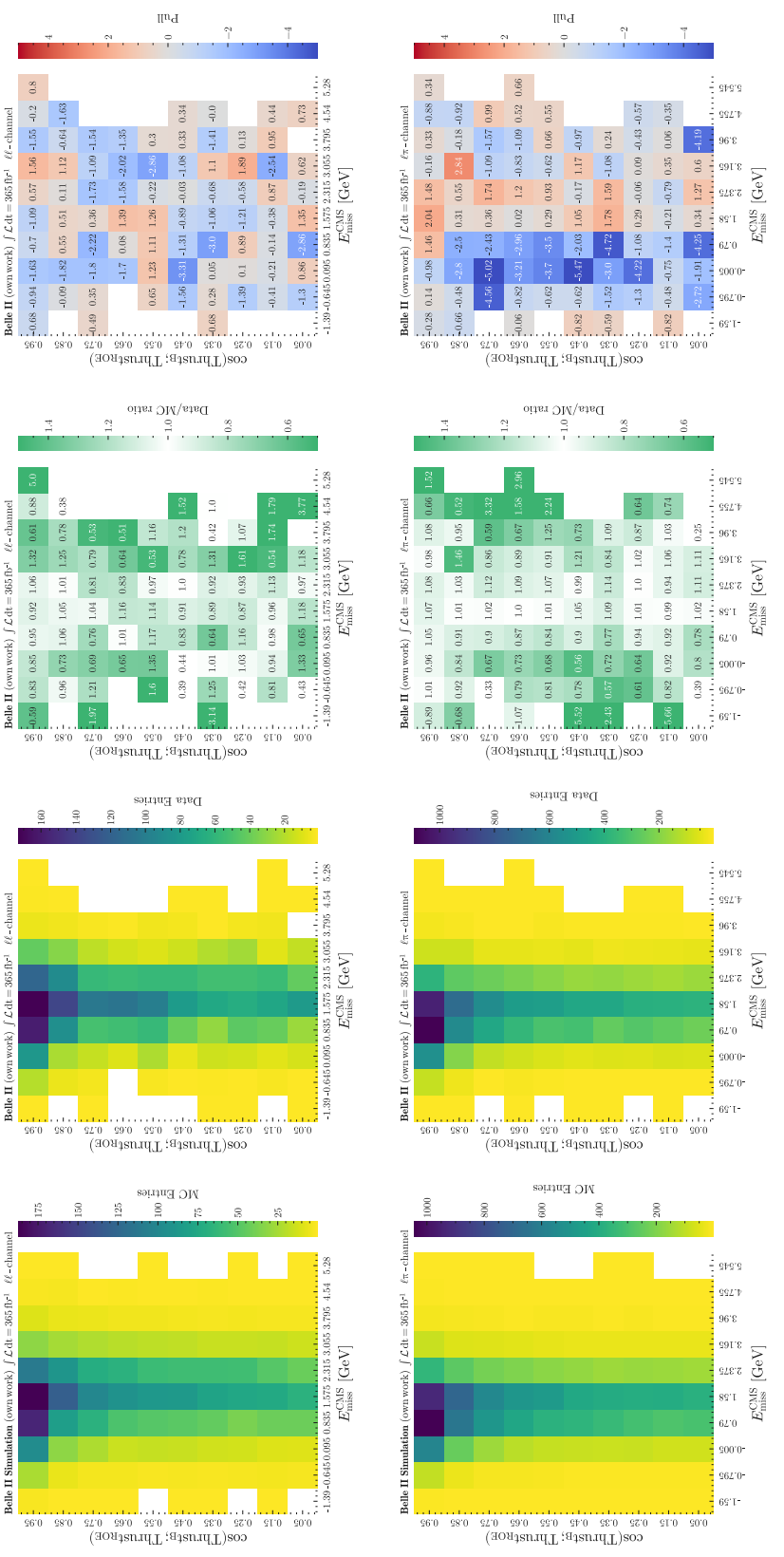


Figure A.46: Two-dimensional comparison of recorded data and simulation in the variables $p_{\mathrm{miss}}^{\mathrm{CMS}}$ and $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

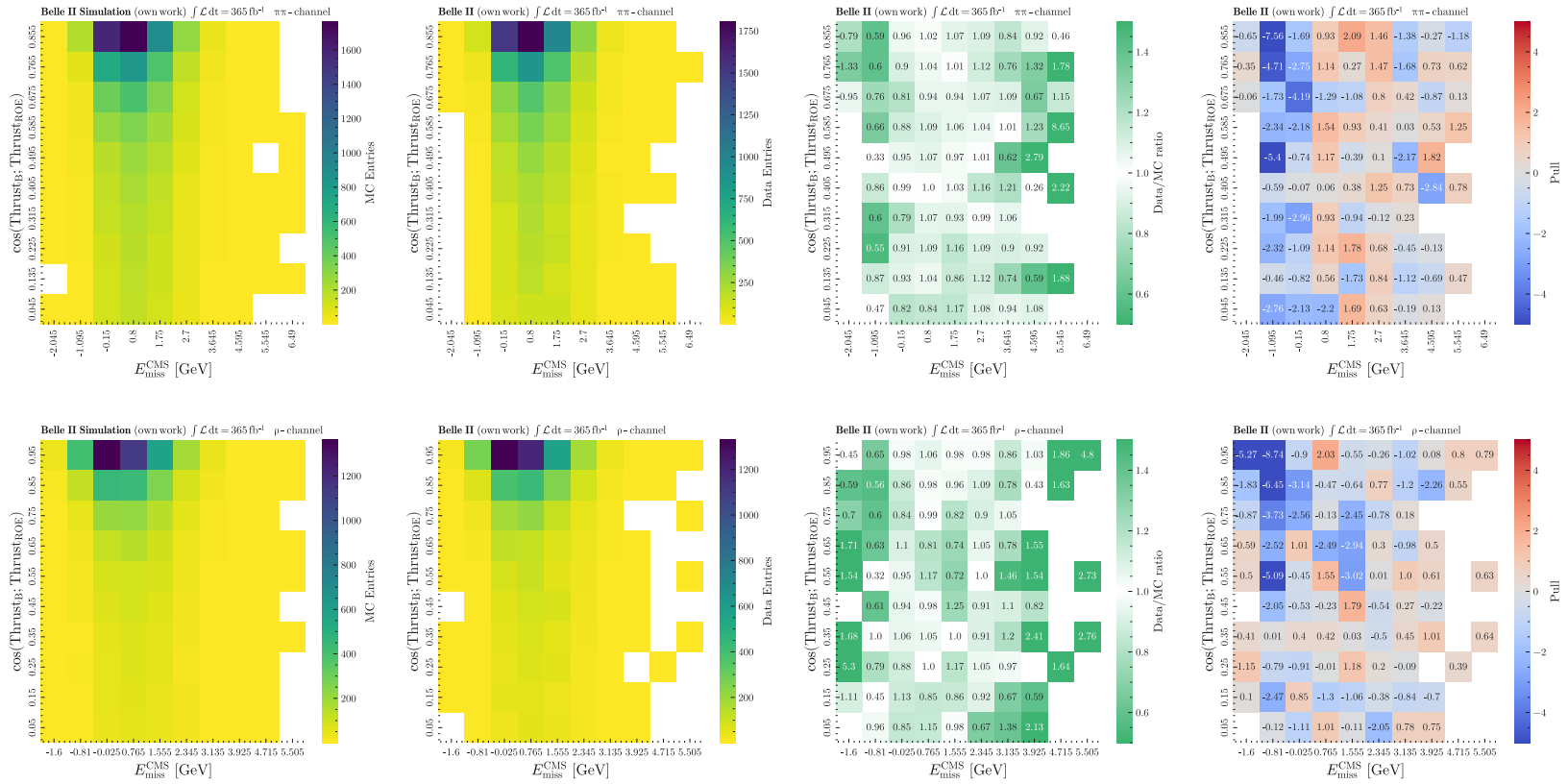


Figure A.47: Two-dimensional comparison of recorded data and simulation in the variables $p_{\text{miss}}^{\text{CMS}}$ and $\cos\left(\text{Thrust}_{\text{B}};\text{Thrust}_{\text{ROE}}\right)$ on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

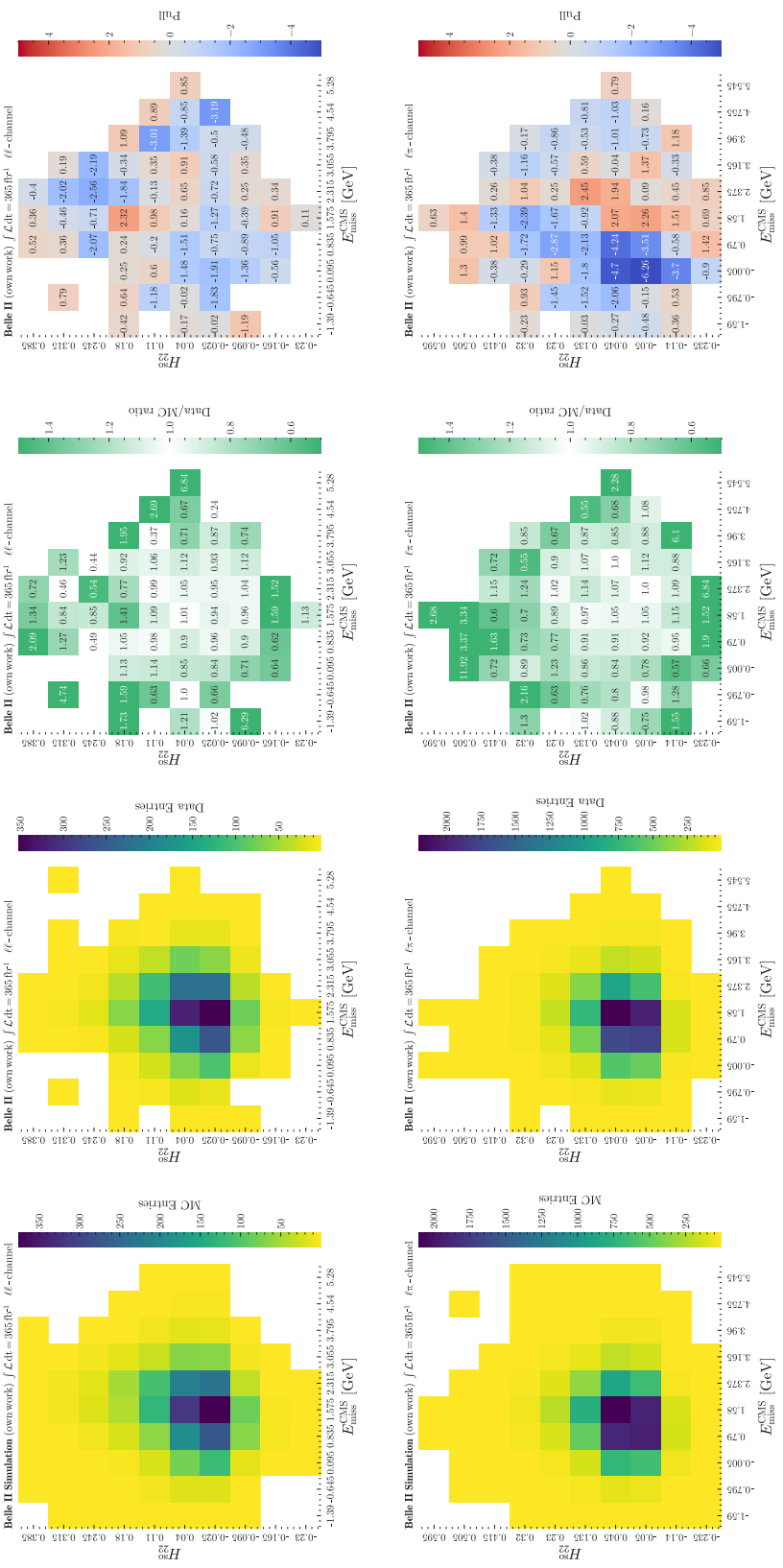


Figure A.48: Two-dimensional comparison of recorded data and simulation in the variables $E_{\text{miss}}^{\text{CMS}}$ and H_{22}^{so} on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

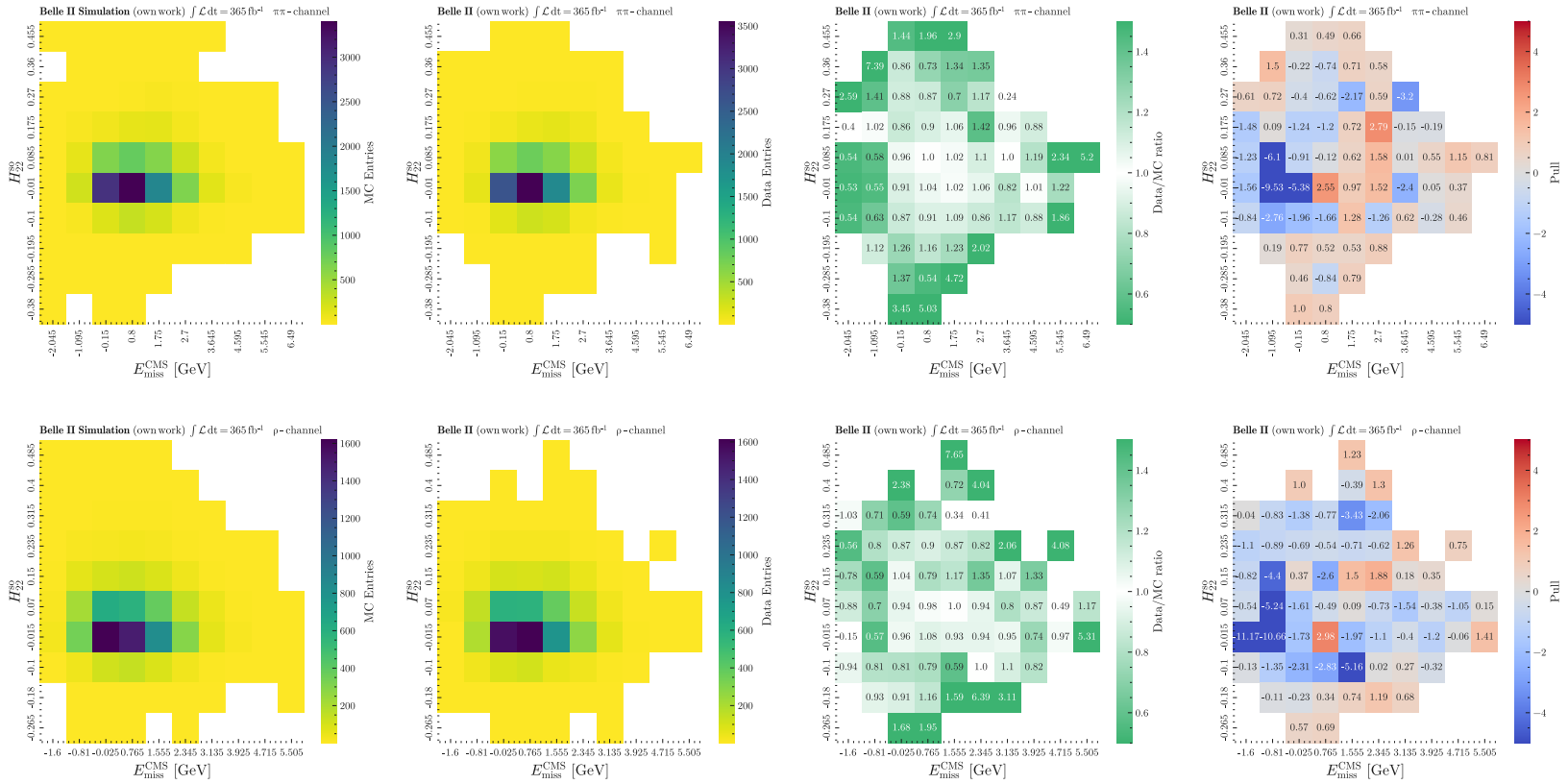


Figure A.49: Two-dimensional comparison of recorded data and simulation in the variables $E_{\text{miss}}^{\text{CMS}}$ and H_{22}^{so} on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

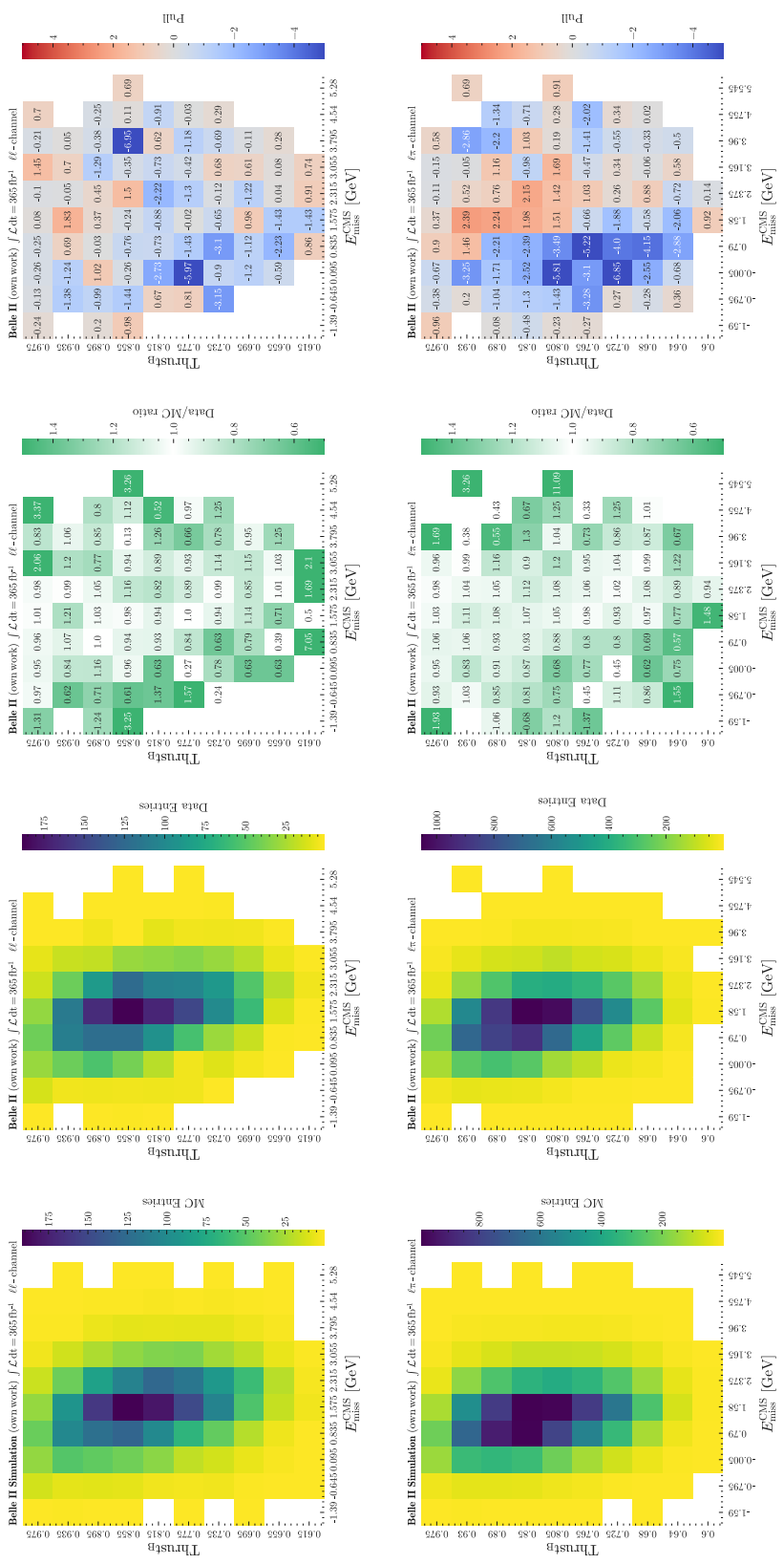


Figure A.50: Two-dimensional comparison of recorded data and simulation in the variables $E_{\mathrm{miss}}^{\mathrm{CMS}}$ and $\mathrm{Thrust_B}$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

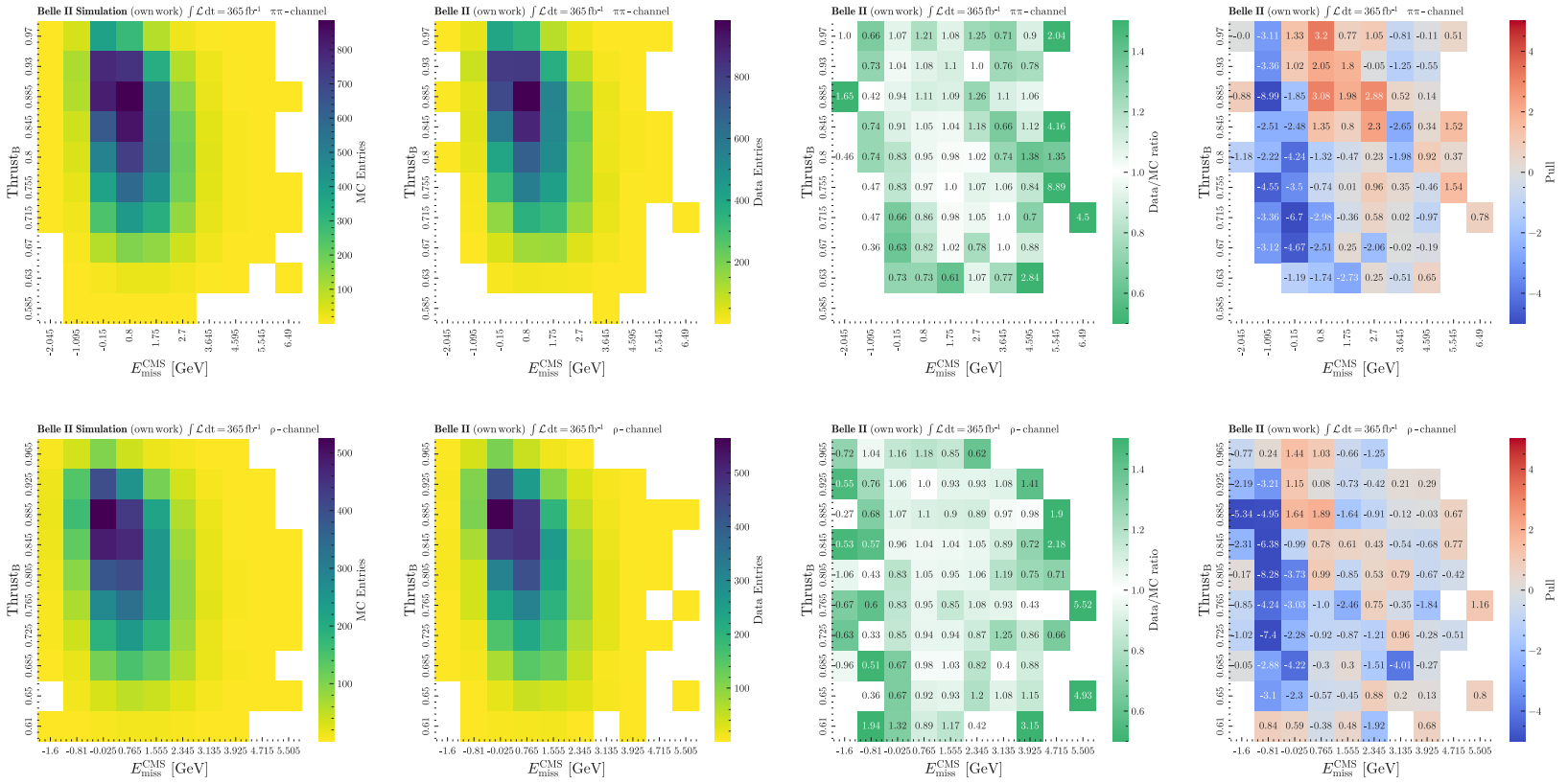


Figure A.51: Two-dimensional comparison of recorded data and simulation in the variables $E_{\text{miss}}^{\text{CMS}}$ and Thrust_{B} on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

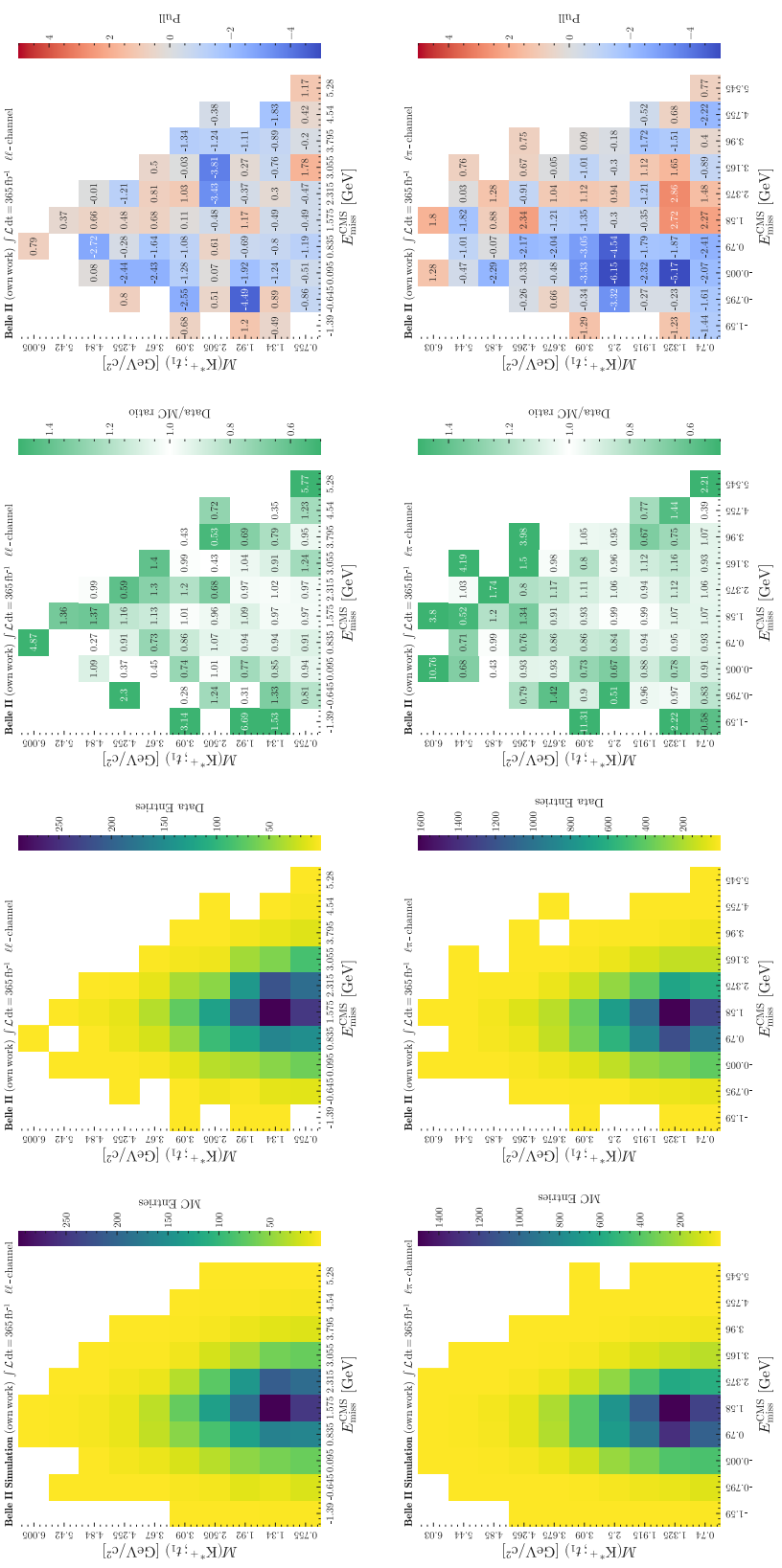


Figure A.52: Two-dimensional comparison of recorded data and simulation in the variables $E_{\text{miss}}^{\text{CMS}}$ and $M(\text{K}^{*+};t_1)$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

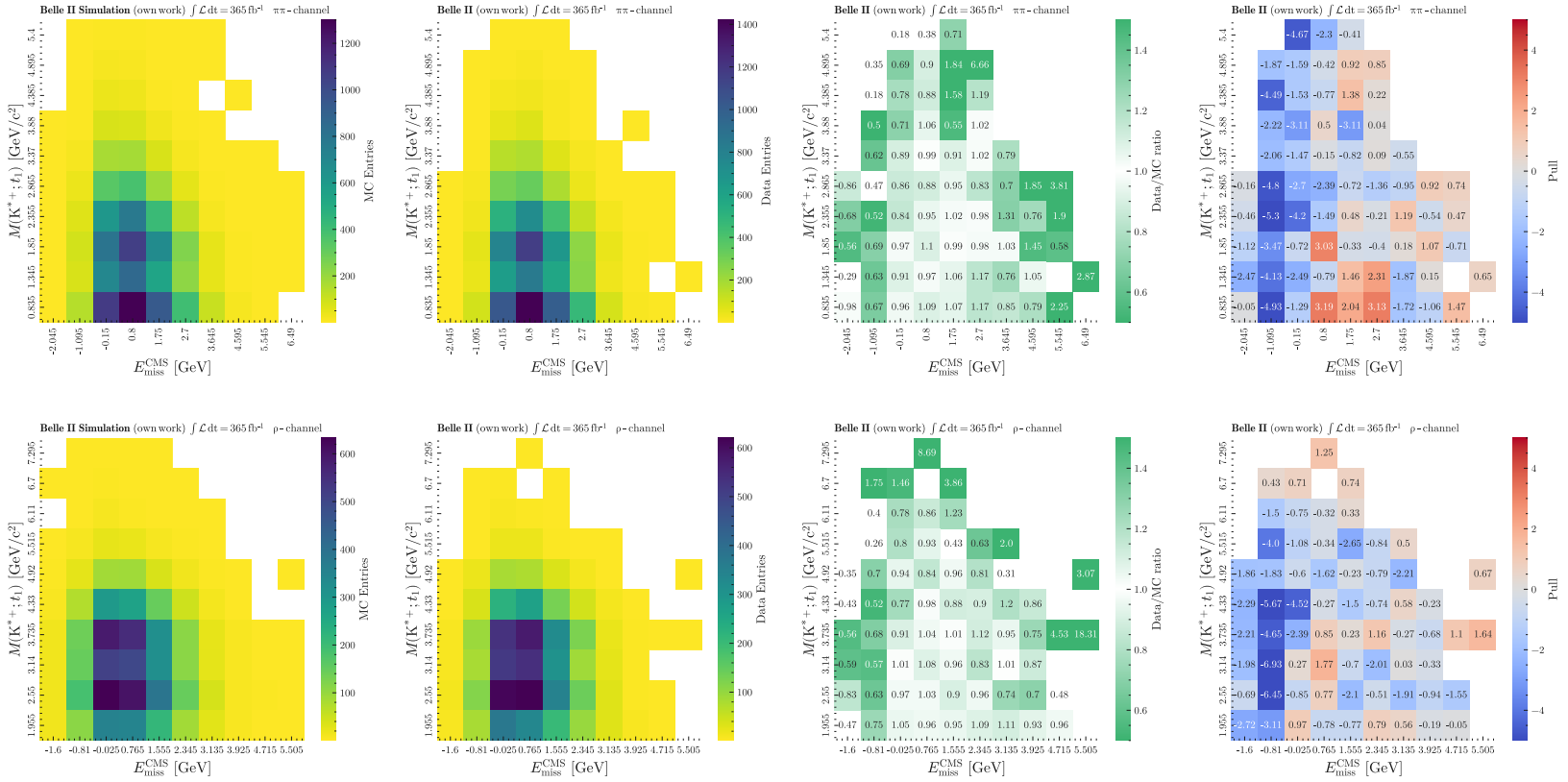


Figure A.53: Two-dimensional comparison of recorded data and simulation in the variables $E_{\text{miss}}^{\text{CMS}}$ and $M(\text{K}^{*+};t_1)$ on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

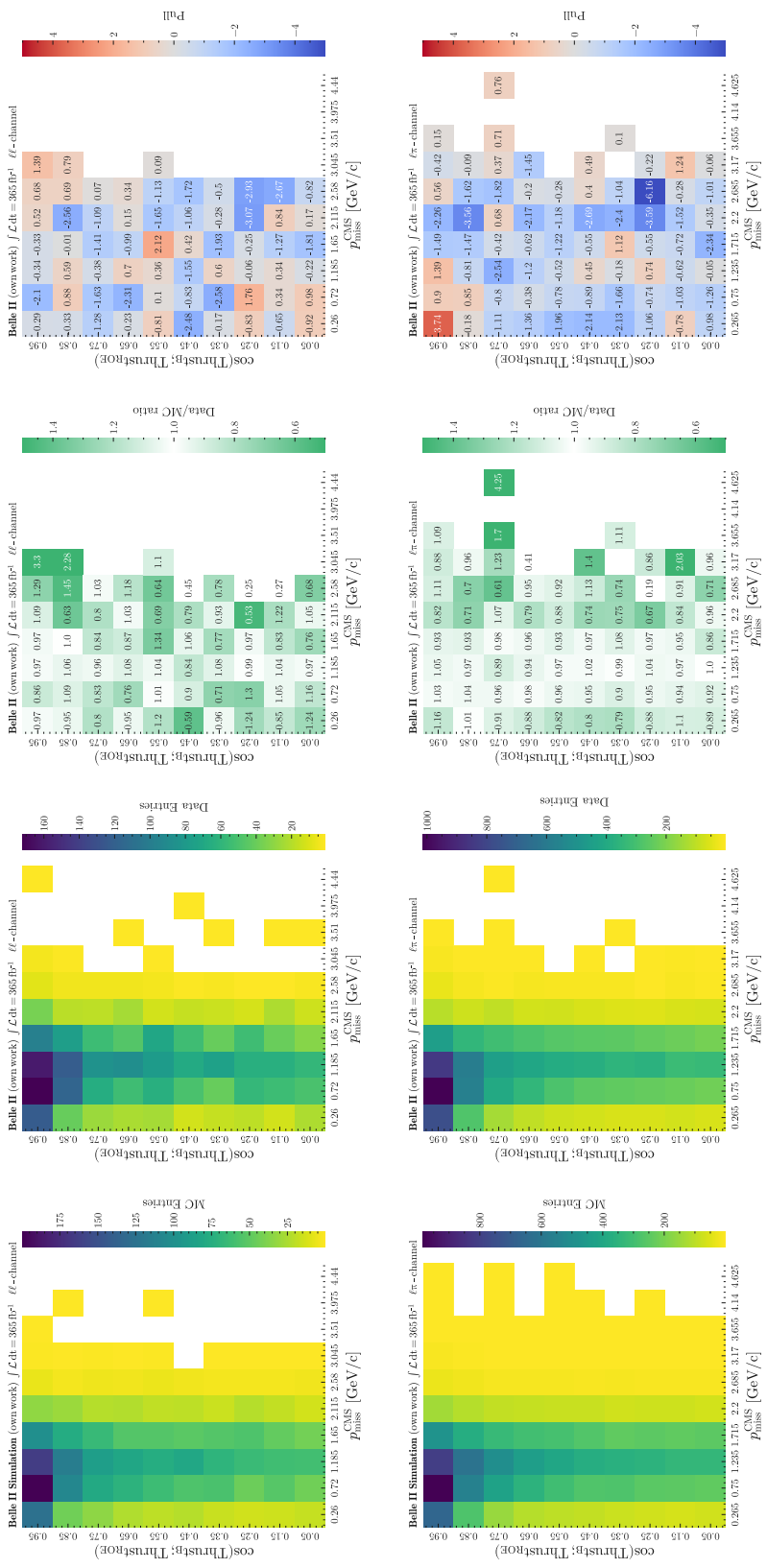


Figure A.54: Two-dimensional comparison of recorded data and simulation in the variables $p_{\mathrm{miss}}^{\mathrm{CMS}}$ and $\cos(\mathrm{Thrust_B}; \mathrm{Thrust_{ROE}})$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

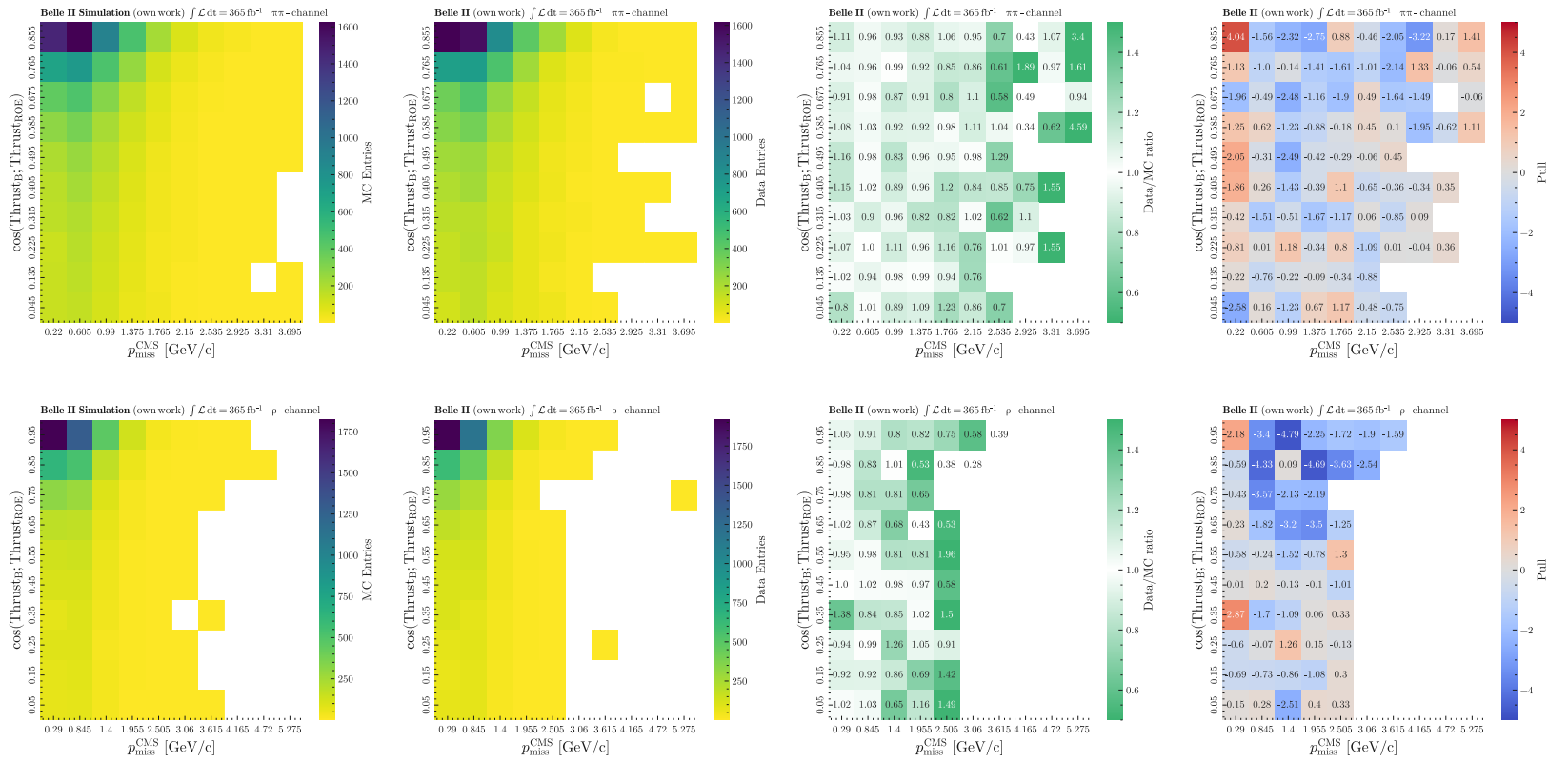


Figure A.55: Two-dimensional comparison of recorded data and simulation in the variables $p_{\text{miss}}^{\text{CMS}}$ and $\cos\left(\text{Thrust}_{\text{B}};\text{Thrust}_{\text{ROE}}\right)$ on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

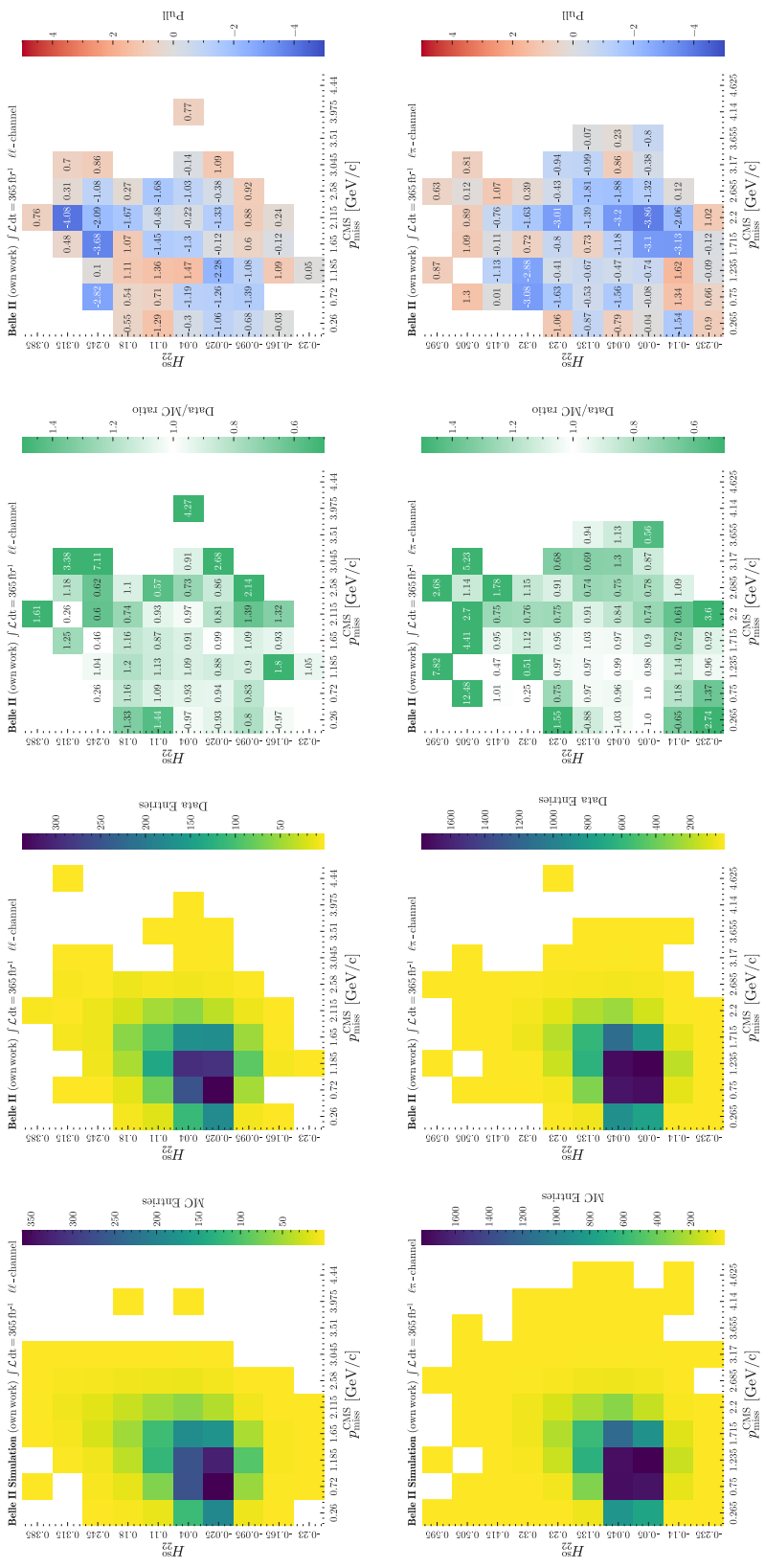


Figure A.56: Two-dimensional comparison of recorded data and simulation in the variables $p_{\text{miss}}^{\text{CMS}}$ and H_{22}^{so} on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

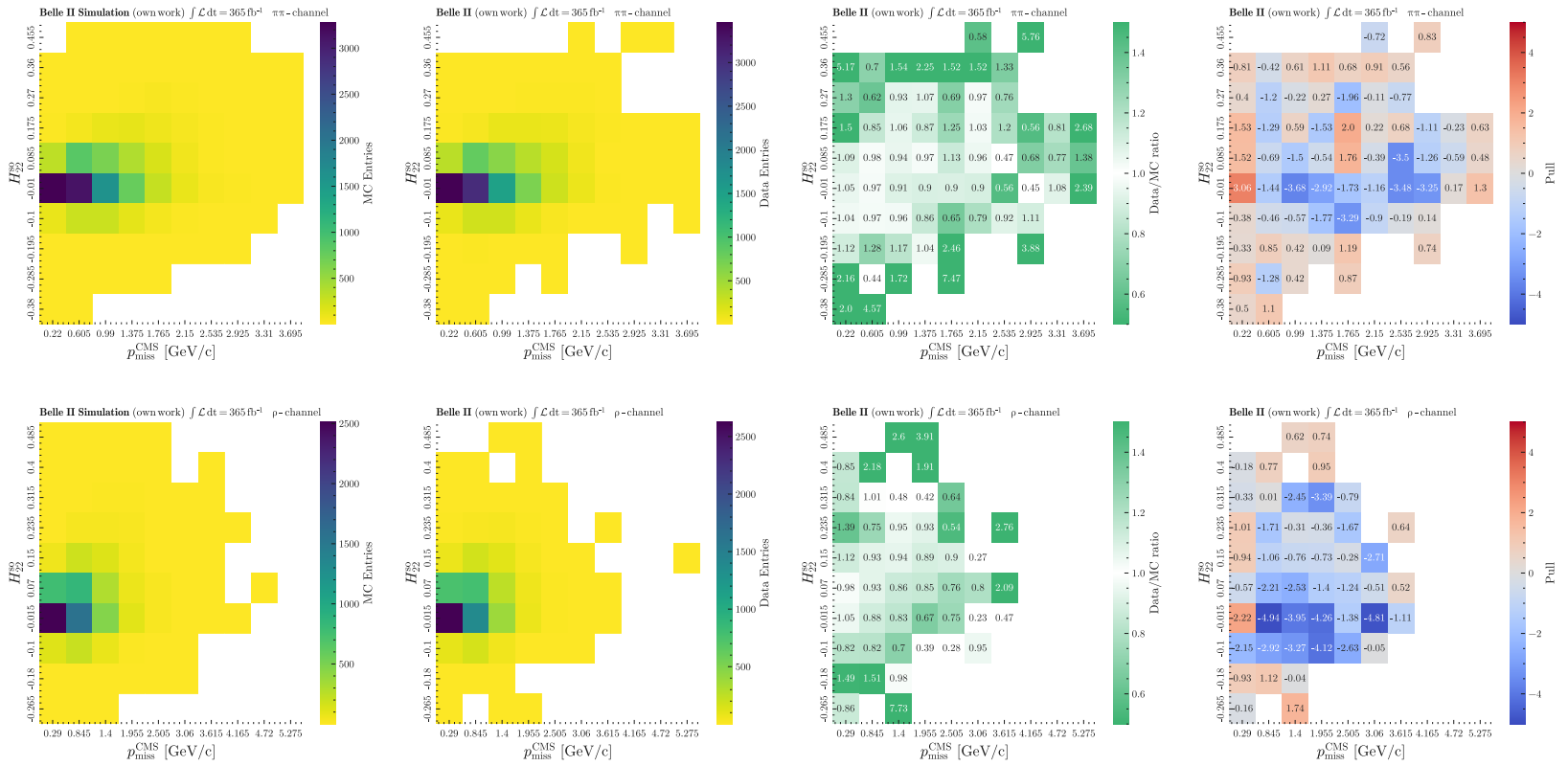


Figure A.57: Two-dimensional comparison of recorded data and simulation in the variables $p_{\text{miss}}^{\text{CMS}}$ and H_{22}^{so} on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

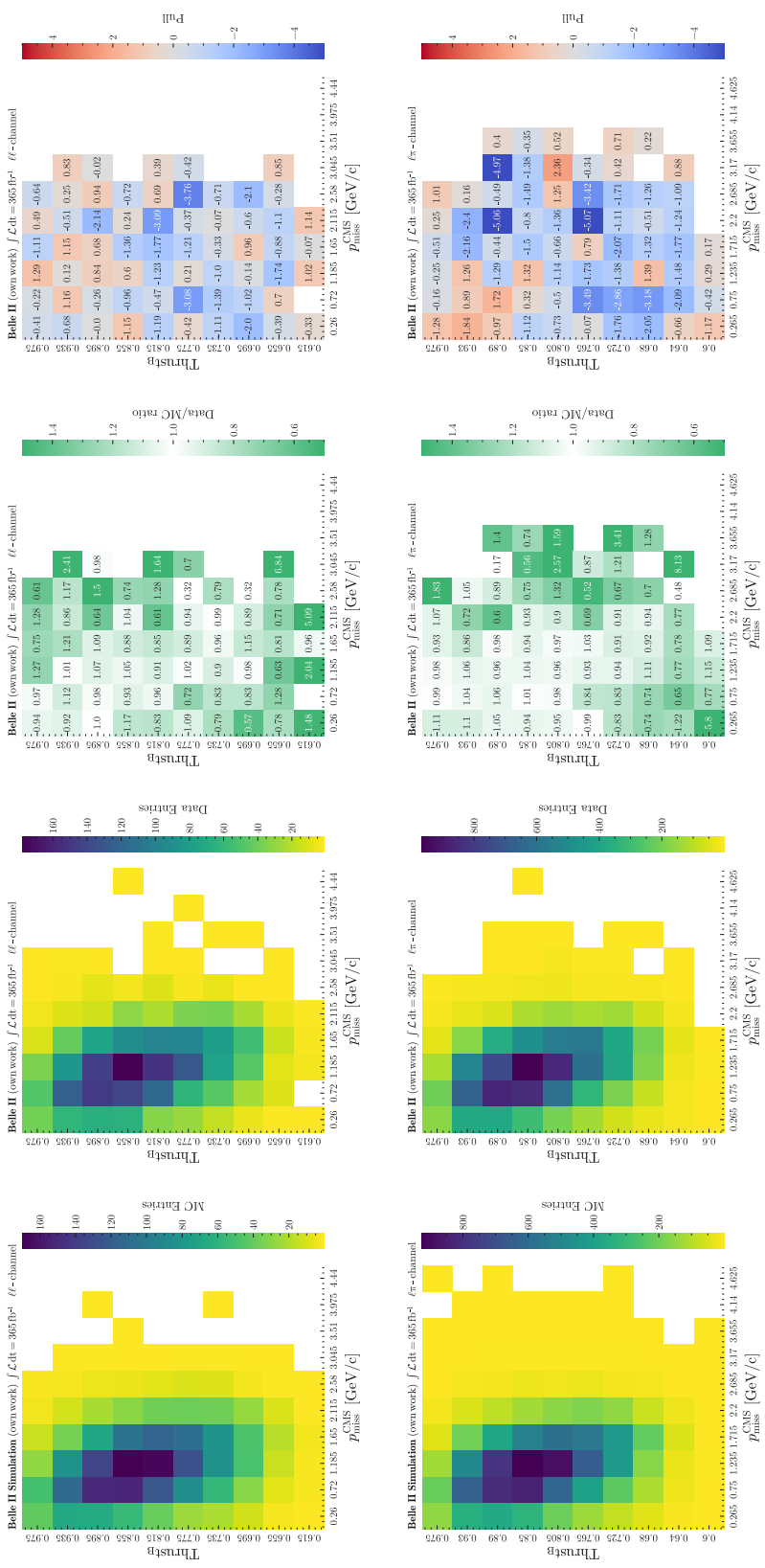


Figure A.58: Two-dimensional comparison of recorded data and simulation in the variables $p_{\text{miss}}^{\text{CMS}}$ and Thrust_{B} on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

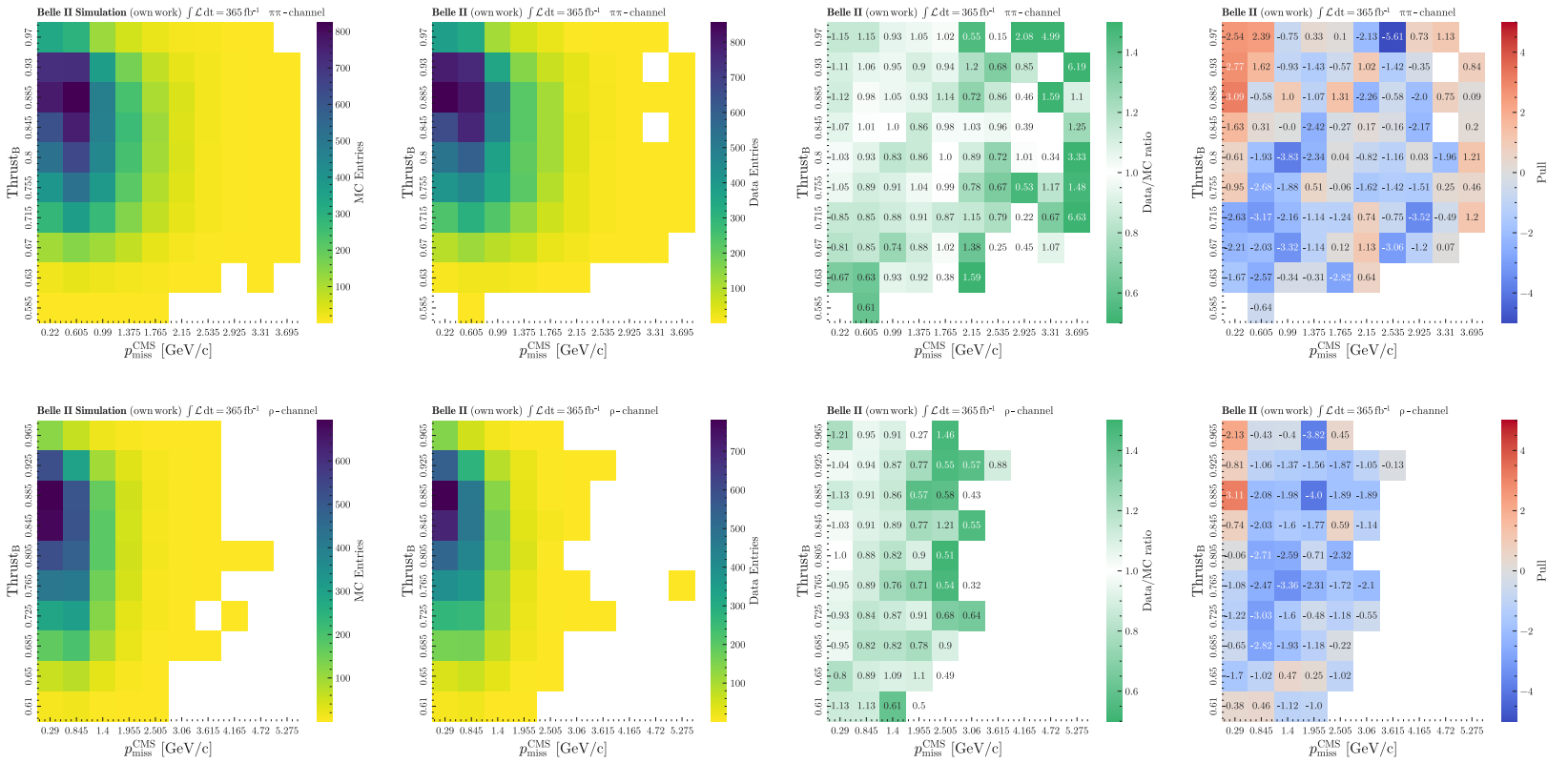


Figure A.59: Two-dimensional comparison of recorded data and simulation in the variables $p_{\text{miss}}^{\text{CMS}}$ and Thrust_{B} on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

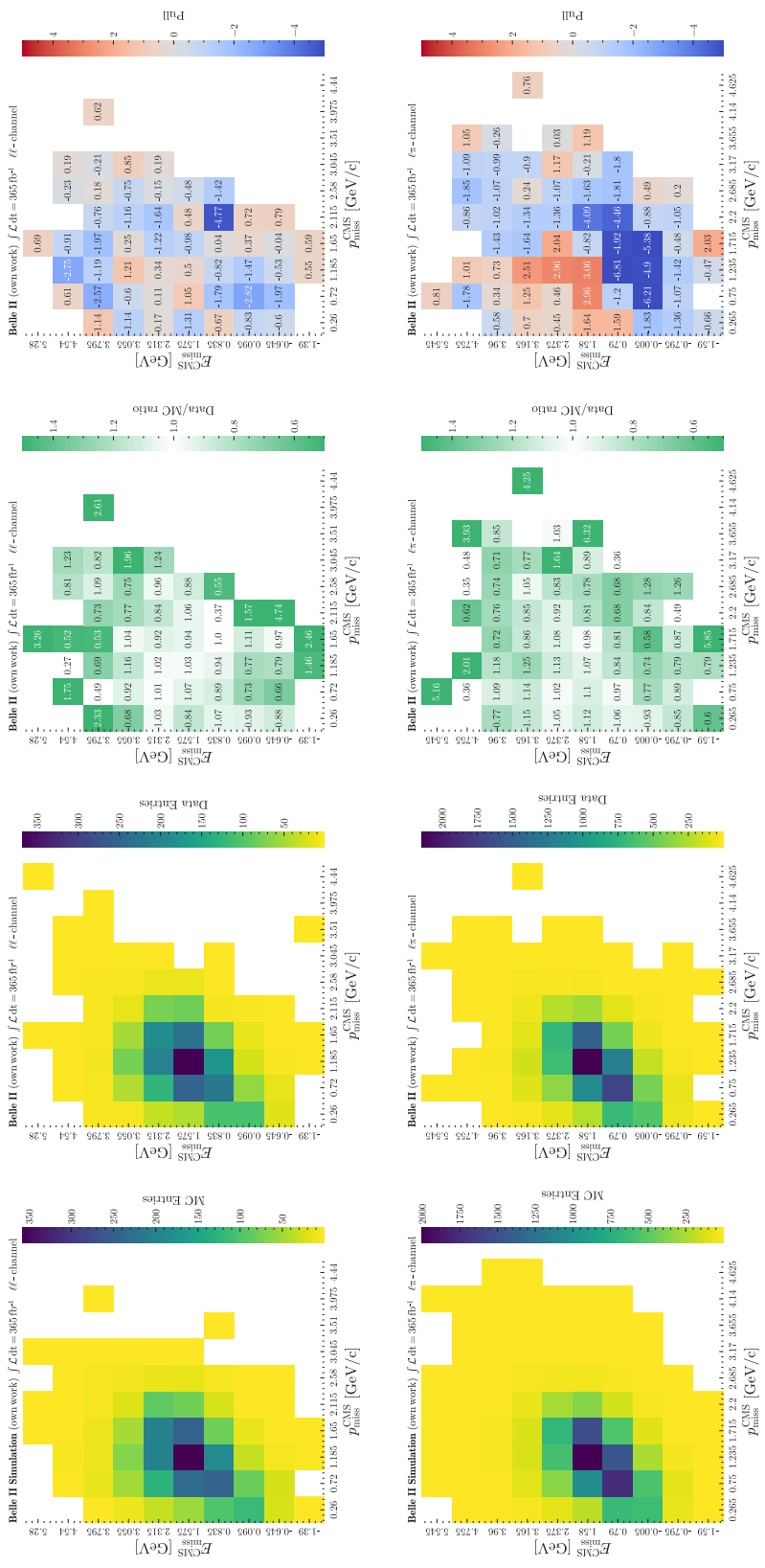


Figure A.60: Two-dimensional comparison of recorded data and simulation in the variables $p_{\text{miss}}^{\text{CMS}}$ and $E_{\text{miss}}^{\text{CMS}}$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

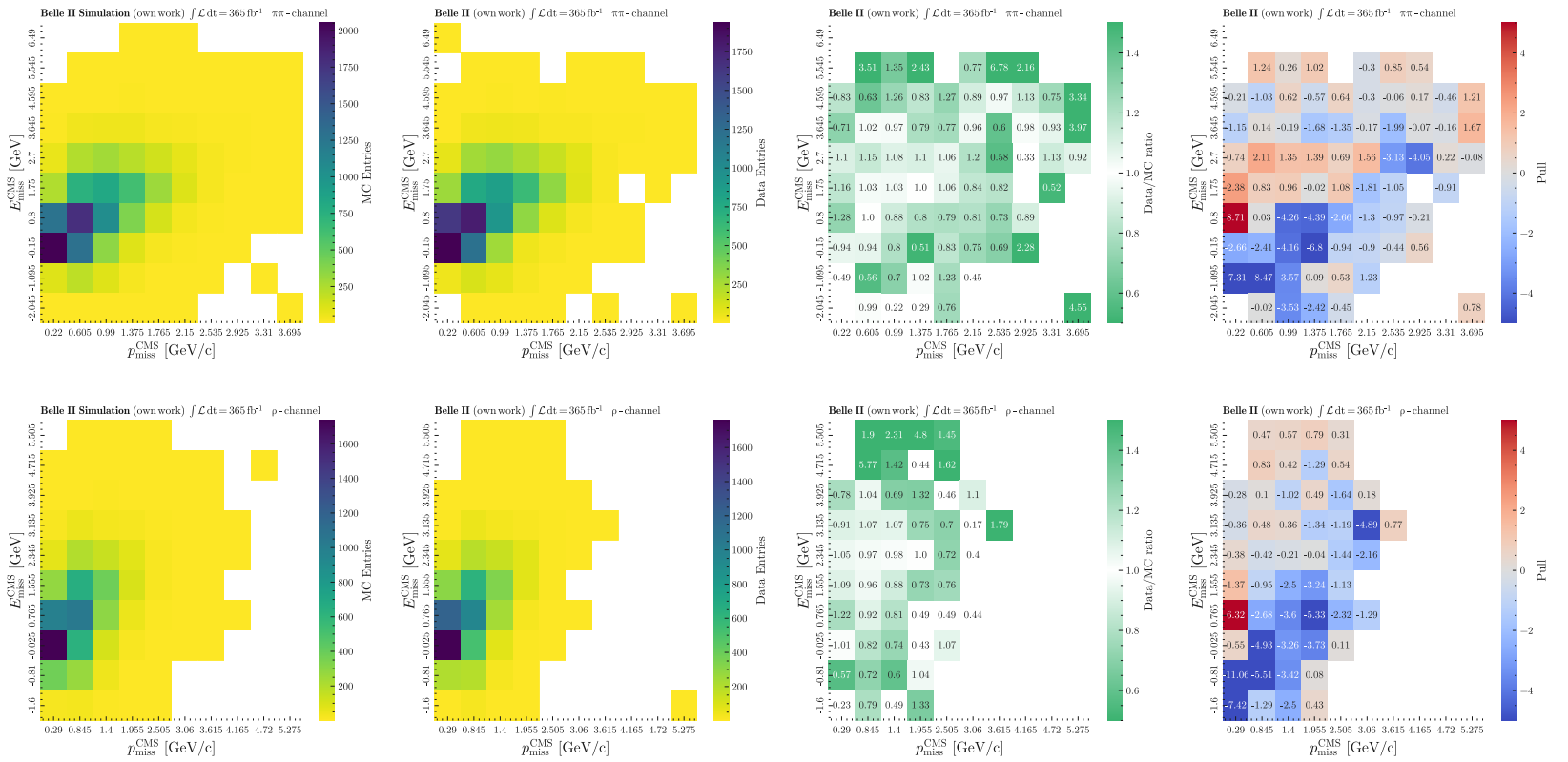


Figure A.61: Two-dimensional comparison of recorded data and simulation in the variables $p_{\text{miss}}^{\text{CMS}}$ and $E_{\text{miss}}^{\text{CMS}}$ on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

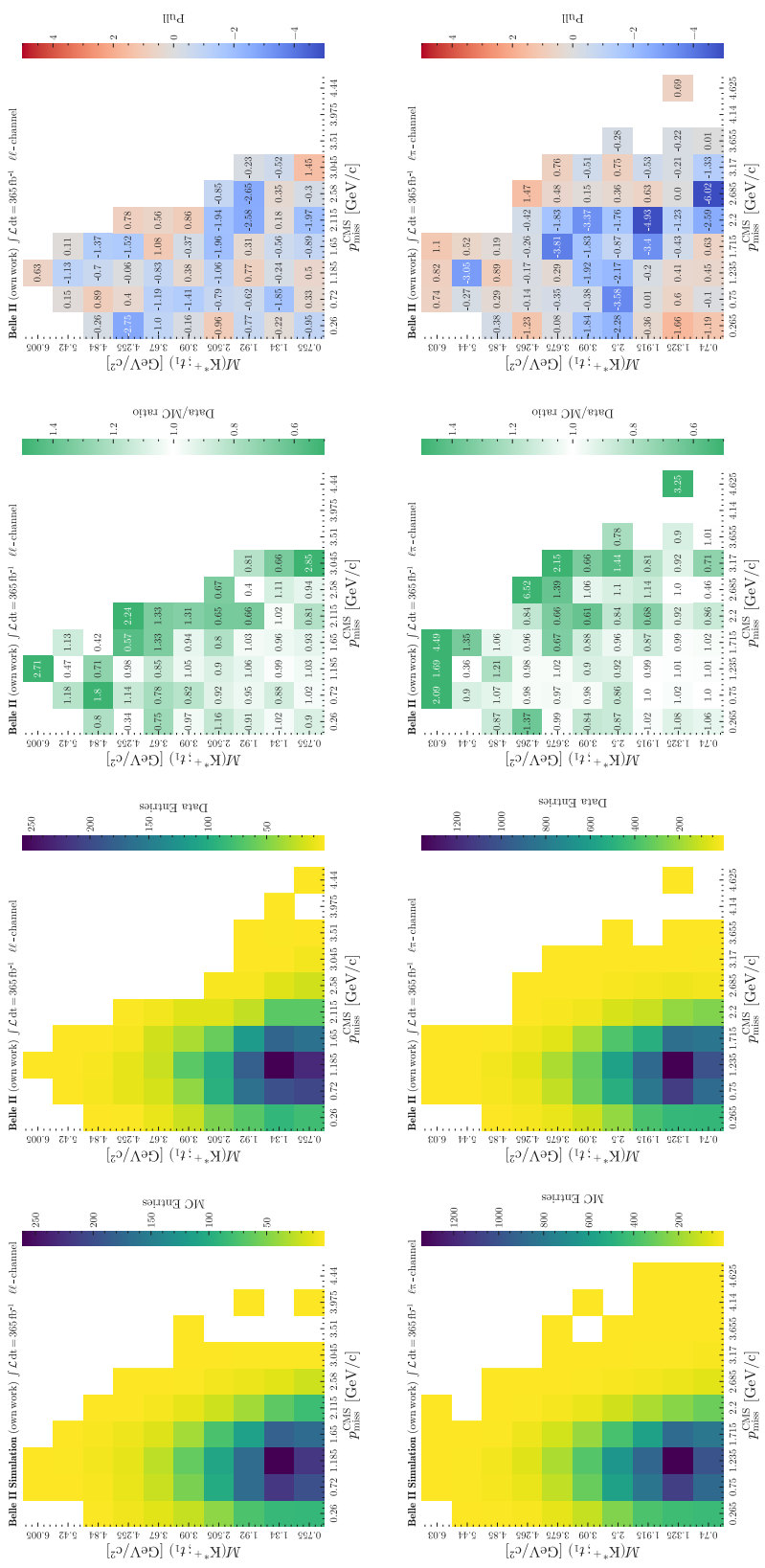


Figure A.62: Two-dimensional comparison of recorded data and simulation in the variables $p_{\text{miss}}^{\text{CMS}}$ and $M(\text{K}^{*+};t_1)$ on the K^{*+} mass sideband for the $\ell\ell$ and $\ell\pi$ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

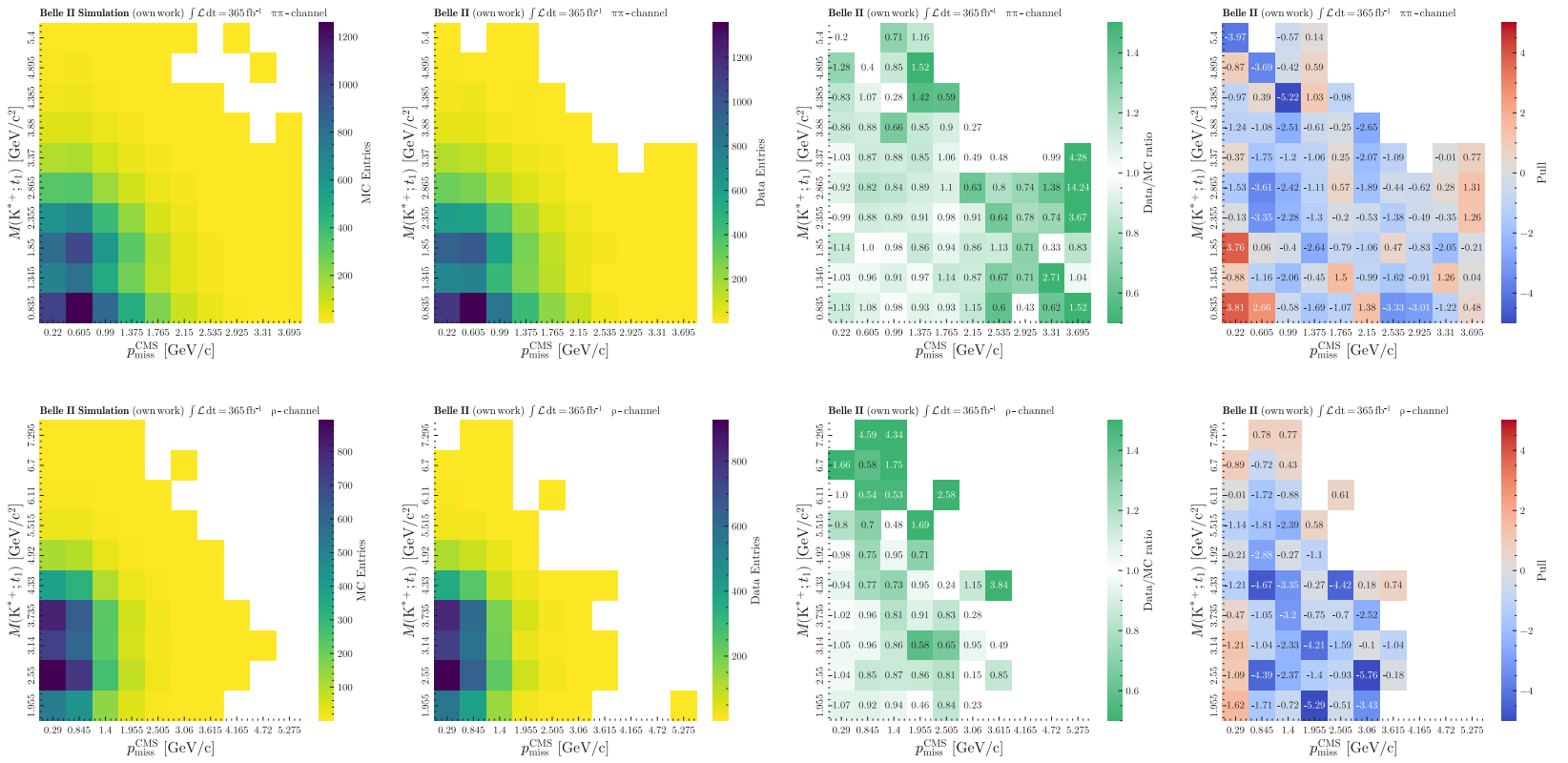


Figure A.63: Two-dimensional comparison of recorded data and simulation in the variables $p_{\mathrm{miss}}^{\mathrm{CMS}}$ and $M(\mathrm{K}^{*+};t_1)$ on the K^{*+} mass sideband for the $\pi\pi$ and ρ channel (from top to bottom). From left to right: Distribution of MC, Data, Data/MC ratio, and Pull. The simulation is scaled to match the integrated luminosity of the recorded data. The uncertainties in data follow a Poisson distribution. All corrections described in Section 5.6 are applied to the simulation.

Glossary

ARICH Aerogel Ring-imaging Cherenkov. 15

basf2 Belle II Analysis Software Framework. 18

BCS Best Candidate Selection. 81

BCS Best Candidate Selection. 20, 25, 29

BDT Boosted Decision Tree. 17, 26, 35, 44, 47, 49, 51, 53, 67, 68, 84

C.L. confidence level. 59

CDC Central Drift Chamber. 15, 26, 28

CKM Cabibbo-Kobayashi-Maskawa. 4, 5

CMS center-of-mass system. 12, 22, 24, 25, 30, 44, 45

CUSB Columbia University-Stony Brooks. 12

DEPFET Depleted Field Effect Transistor. 14

ECL Electromagnetic Calorimeter. 15, 16, 20, 26–28, 43, 62

FCNC flavor-changing neutral current. 1, 5

FEI Full Event Interpretation. 20, 21, 23, 25, 42, 43, 69

FoM Figure of Merit. 17, 37, 40, 41

FPGA Field Programmable Gate Array. 16

FR Full-Reconstruction. 8, 19–21

FSP final state particle. 23, 25–27, 33

GIM Glashow-Iliopoulos-Maiani. 5

HAPD Hybrid Avalanche Photo-Detector. 15

HEP high-energy physics. 18

HER High-Energy Ring. 12

IP interaction point. 13, 14, 26–28

KEK High Energy Accelerator Research Organization. 12, 13

KLM K_L^0 and Muon Detector. 16, 20

KSFWs Kakuno Super Fox Wolfram moments. 32, 33

LER Low-Energy Ring. 12, 13

LFU lepton flavor universality. 1, 3–5, 7

LHCb Large Hadron Collider beauty. 1

MC Monte Carlo. 21, 23, 31–33, 42, 44, 46, 51, 62, XI

MCP-PMT micro-channel-plate photomultiplier tube. 15

MVA Multivariate Analysis. 35, 39, 66, 69, 84, 118, XII

NP New Physics. 6–8

PDF probability density function. 60, 61

PID Particle Identification. 26

POCA point of closest approach. 26, 28

PXD Pixel Detector. 14

QCD Quantum chromodynamics. 5, 6

QED Quantum electrodynamics. 45

ROC Reciever Operating Characteristic. 39

ROE Rest Of Event. 18, 29, 31, 32, 62, 68

RPC Resistive Plate Chamber. 16

SB sideband. 28, 33, 49, 51

SGBDT Stochastic Gradient Boosted Decision Tree. 18

SM Standard Model of particle physics. 1, 3–7

SVD Silicon Vertex Detector. 14

TOP Time-of-Propagation. 15