
Angular analysis of $B^0 \rightarrow K^{*0} e^+ e^-$ decays

M. Atzeni¹, J. Eschle¹, A. Mauri², N. Serra¹, R. Silva Coutinho¹, Z. Wang^{1†}

¹*Physik-Institut, Universität Zürich, Zürich, Switzerland*

²*Nikhef National Institute for Subatomic Physics, Amsterdam, The Netherlands*

Abstract

An angular analysis of the $B^0 \rightarrow K^{*0} e^+ e^-$ decay is presented using the data samples collected by the LHCb experiment at centre-of-mass energies of 7, 8 and 13 TeV, corresponding to an integrated luminosity of 5.0 fb^{-1} . The analysis is performed by the first time at LHCb in the central region of the dilepton invariant mass, q^2 , *i.e.* $1.1\text{--}6.0 \text{ GeV}^2/c^4$ and $1.1\text{--}7.0 \text{ GeV}^2/c^4$.

[†]Contact author: zhenzi.wang@cern.ch

Version history

0.1 TODO list

- Minor: update the L0E threshold for 2015 from 2700 to 3000 (align with other LFU analyses)
- Most of the systematic uncertainties on the measurement still need to be evaluated. These will be populated in a later stage of the review.
- Some misalignments have been seen for the Meerkat response PIDCorr for Run-I ProbNNpi in pions (observed by M. Schubiger). Further studies are being currently performed and these will be added in the next version of the note.

The following versions of this document have been published:

Table 1: Overview of the version history of this document

Version	Date	Status/comments
1r0	25.06.2021	First draft of the analysis note - Sent out to WG conveners.
1r1	30.07.2021	Comments from WG conveners.
1r2	05.08.2021	Second set of comments from WG conveners (minor).

Contents

0.1	TODO list	3
1	Introduction	1
2	Analysis strategy	4
2.1	Dilepton invariant mass regions	5
2.2	Angular basis	6
2.3	Blinding procedure	7
3	Dataset, trigger and stripping	9
3.1	Data	9
3.2	Simulation	9
3.3	Trigger	11
3.3.1	L0 threshold issue	12
3.3.2	TISTOS bug	12
3.3.3	AALLSAMEBPV bug	12
3.3.4	Sim09[d,e,f] multiplicity bug	13
3.4	Stripping	13
3.5	Truth matching	13
3.6	Corrections to simulation	14
3.6.1	Particle identification correction	15
3.6.2	TCK trigger requirements	17
3.6.3	Reconstruction, kinematic and multiplicity corrections	22
3.6.4	Trigger correction	25
3.6.5	Total corrections	35
4	Selection	44
4.1	Preselection	46
4.2	Exclusive backgrounds	47
4.2.1	$B_s^0 \rightarrow \phi e^+ e^-$ background	47
4.2.2	$B^+ \rightarrow K^+ e^+ e^-$ background	47
4.2.3	$\Lambda_b^0 \rightarrow p K J/\psi (\rightarrow e^+ e^-)$ and $\Lambda_b^0 \rightarrow p K e^+ e^-$ backgrounds	47
4.2.4	Charmonium electron–hadron and hadron–hadron swaps	49
4.2.5	$\pi \rightarrow e$ mis-identification	50
4.2.6	Partially reconstructed backgrounds	51
4.3	Charmonium contributions	51
4.4	Multivariate classifier	57
4.4.1	Signal and background samples	57
4.4.2	Selection of discriminating variables	58
4.4.3	Selection of classifier	59
4.4.4	Uniform classifiers	61
4.4.5	k-folding	63

4.5	MVA optimisation	65
4.5.1	Optimisation results	66
4.6	Multiple candidates	70
5	Signal amplitude model	71
5.1	Mass signal model	71
5.2	Angular signal model	80
5.2.1	S-wave interference	81
6	Acceptance	84
6.1	Preliminaries	84
6.2	Parametrisation	87
6.3	Sample choice	90
6.4	Nominal acceptance	92
6.5	Alternative acceptance	93
7	Background model	105
7.1	Combinatorial	105
7.2	Double semi-leptonic cascade decays	108
7.3	Combinatorial/DSL combined modelling	111
7.4	Partially reconstructed $B^+ \rightarrow K^+\pi^+\pi^-e^+e^-$ decays	115
7.5	Control channel	118
8	Angular fit	119
8.1	Weighted unbinned maximum likelihood fit	119
8.2	Rare and control mode background components	120
8.3	Full angular fit	121
8.4	Weighted fit uncertainty correction	122
8.5	Sensitivity studies	122
8.5.1	Component yields	134
8.5.2	Background models	137
9	Control mode fit validation	139
9.1	Acceptance	141
9.2	Angular data fit	144
9.3	Systematics	149
9.3.1	Coefficient uncertainties	149
9.3.2	Higher orders	150
9.3.3	Correction weights	151
9.4	Cross checks	157
9.4.1	Subsamples results	157
9.4.2	Similar configurations	161
9.4.3	Additional backgrounds	165

10 Systematic uncertainties	179
10.1 Preliminary systematics studies	179
10.2 Signal mass shape	181
10.3 Background modelling	181
10.3.1 Fit uncertainties	181
10.3.2 DSL model	182
10.3.3 Combinatorial model	182
10.3.4 Partially reconstructed background model	183
10.4 Acceptance function	183
10.4.1 Negative regions	183
10.4.2 Higher order description	184
10.4.3 Coefficient uncertainties	184
10.4.4 Resolution model dependence	184
10.4.5 Excluding BKGCAT 60	185
10.5 J/ψ leakage	185
10.6 Neglecting S-wave	185
10.7 Control mode mass fit	185
10.8 Agreement between data and simulation	186
10.8.1 Kinematic correction	186
10.8.2 PID systematic	187
10.8.3 Trigger corrections	187
10.9 $B^+ \rightarrow K^+ e^+ e^-$ veto	188
10.10 Fit bias	189
11 Cross checks	190
11.1 Modifications to the nominal fit	190
11.1.1 Double semi-leptonic modelling	190
11.1.2 q^2 selection	190
11.2 Pre-unblinding cross-checks	190
12 Results	192
12.1 Blinded data fits	192
12.2 Unblinded fit results	198
A PID response using Meerkat	199
A.1 Correction of PID response	199
A.2 Unbinned description of calibration PDF	200
A.3 Implementation in PIDCalib	206
A.4 Additional PID response projections	207
B BDT re-weighting procedure	216
B.1 Performance	216
B.1.1 FoM_1 - simple discriminator	216
B.1.2 FoM_2 - data labelling	216

B.1.3	FoM_3 - data counting	217
B.2	$sPlot$ fits for calibration samples	217
B.3	BDT-re-weighting results	218
C	Trigger correction	221
C.1	The effect of the L0 weights	221
C.2	Compatibility between L0E and L0E HLT corrections	221
C.3	HLT weighted samples	221
C.4	Correlation between $p_T(B)$ and $\cos\theta_K$	222
D	Additional information on the MVA	235
D.1	Preliminary preselection for MVA	235
D.2	Classification performance	236
D.3	Uniform classifiers performance	236
D.4	Performance of new classifiers	241
D.5	MC/data comparison for MVA variables	246
D.6	Additional info on the MVA optimisation	249
D.6.1	Mass fits used in the MVA optimization procedure	249
D.6.2	P'_5 Figure of Merit	253
E	Selection for partially reconstructed backgrounds	256
E.1	HOP approach	256
E.2	Multivariate classifier approach	259
F	Folding technique	262
F.1	Folding for P'_4	262
F.2	Folding for P'_5	263
F.3	Folding for P'_6	263
F.4	Folding for P'_8	264
F.5	Other observables	264
G	Additional acceptance/resolution information	266
G.1	Resolution	266
H	Weighted fit features	272
H.1	Large weights	272
H.2	Weighted fractions	274
I	Double-semileptonic cascade background	279
I.1	A rough comparison of DSL in $K\pi e\mu$ and $K\pi ee$ data	279
I.2	A rough comparison of the DSL in simulation and in $K\pi e\mu$ data	280

J Control mode validation	282
J.1 Main validation fit	282
J.2 Subsamples results	283
J.2.1 Separating by Run	283
J.2.2 Separating by Run and trigger category	286
J.2.3 Separating by bremsstrahlung category	289
J.3 Similar configurations	291
J.3.1 FLATQ2 acceptance	291
J.3.2 Higher order acceptance	294
J.3.3 Resolution correction	300
J.3.4 No B^+ veto	303
J.4 Additional backgrounds	306
J.4.1 Unconstrained q^2 range	325
J.4.2 Constrained q^2 without constrained B^0 mass cut	325
J.4.3 Constrained q^2 with constrained B^0 mass cut	325
K Reweighting of the hadronic system Dalitz plots	327
References	330

¹ 1 Introduction

² Decays involving a $b \rightarrow s\ell^+\ell^-$ transition are suppressed in the Standard Model (SM) due
³ to the absence of flavour changing neutral currents (FCNC) at tree level [1]. These FCNC
⁴ processes are crucial probes for the SM, since as-yet undiscovered particles may contribute
⁵ to the decay process and cause observables to deviate from their predicted values [2–5].
⁶ The decay mode $\bar{B} \rightarrow \bar{K}^*\ell^+\ell^-$ is a prime example, which offers a rich framework to study
⁷ from differential decay widths to angular observables. In particular, these are sensitive to
⁸ a large number of Wilson coefficients as a function of the di-lepton invariant mass squared,
⁹ q^2 : $C_7^{(\prime)}$ in the very low- q^2 region; and $C_9^{(\prime)}$ and $C_{10}^{(\prime)}$ dominantly in the central- q^2 and high- q^2
¹⁰ regions (see Fig. 1).

¹¹ A full angular analysis of the decay channel $B^0 \rightarrow K^{*0}\mu^+\mu^-$ with the Run-I dataset [6]
¹² and more recently updated with 2016 dataset [7], has been performed and an anomalous
¹³ behaviour is reported notably in one of the observables with reduced form-factor uncer-
¹⁴ tainties, P'_5 [8, 9]. This deviation has been recently confirmed by the Belle collaboration,
¹⁵ including in addition the mode $B^0 \rightarrow K^{*0}e^+e^-$ and the charged B meson counterparts [10].
¹⁶ Figure 2 summarises the experimental status of these experiments. In fact, global analysis
¹⁷ fits using these results report deviations from SM predictions at the level of four to six
¹⁸ standard deviations [11–13]. These anomalies received immediate attention in the flavour
¹⁹ community, since it could be explained by a shift in the coefficient C_9 only, or C_9 and C_{10}
²⁰ simultaneously. Note that branching ratios measurements of $b \rightarrow s\ell^+\ell^-$ transitions are
²¹ also consistent with a shift in C_9/C_{10} [14–17]. Several approaches have been suggested in
²² order to interpret these results as NP signatures: existence of a new neutral gauge boson
²³ Z' with a TeV-scale mass with particular couplings to fermions [18–23]; bosonic particles
²⁴ such as leptoquarks [24–36]. Nonetheless, due to the vector-like nature of this pattern,
²⁵ this could also be explained by large hadronic contributions from $b \rightarrow sc\bar{c}$ operators (*i.e.*

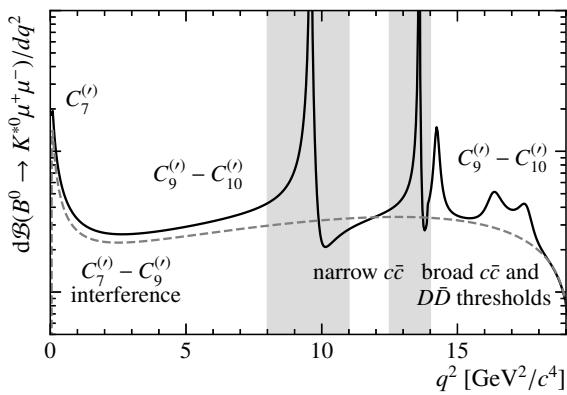


Figure 1: Schematic view of the $\bar{B} \rightarrow \bar{K}^*\ell^+\ell^-$ spectrum as a function of the q^2 region, with the dominant Wilson coefficients displayed and possible $c\bar{c}$ contributions.

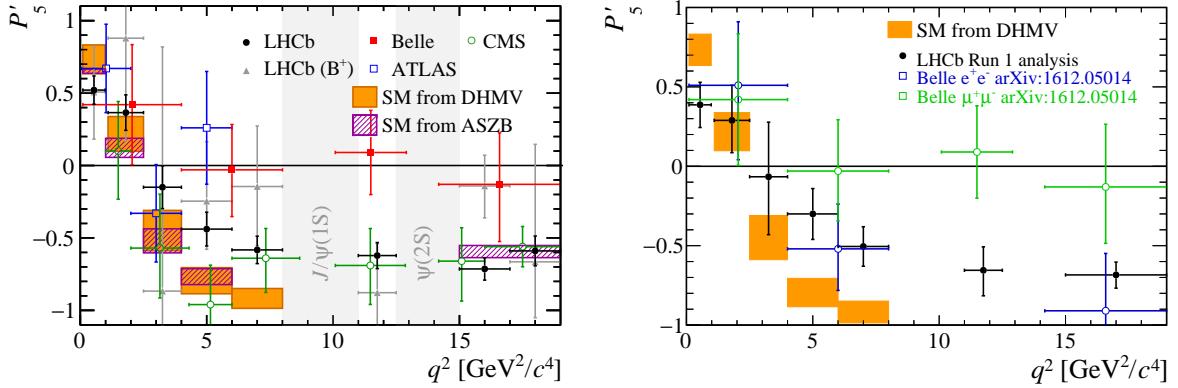


Figure 2: Fit results for (left) the P'_5 observable for muons in bins of q^2 [7, 10, 46–48] and (right) the Belle results separated for muons and electrons [10].

charm loops) that could be able to either mimic or camouflage NP effects [37–45]. In other words, even in the optimised P'_5 observable there are potentially sizeable uncertainties that are difficult to estimate, which in turn limits the strength of NP constraints.

An important feature of the SM is the property that particles couple equally to different flavours of leptons, hereafter referred to as Lepton Flavour Universality (LFU). For instance, the ratio of the branching fractions $B^+ \rightarrow K^+ \mu^+ \mu^-$ to $B^+ \rightarrow K^+ e^+ e^-$ decays, R_K , is predicted to be unity within an uncertainty of $\mathcal{O}(10^{-3})$ in the SM [49, 50]. LHCb has measured this ratio in the range of $1 < q^2 < 6 \text{ GeV}^2/c^4$ with the full dataset and found a value compatible with the SM prediction with a p-value of 0.10% [51], *i.e.* the first experimental evidence for the violation of lepton universality in these decays. The same ansatz has been verified in $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ decays in two q^2 regions, $0.045 - 1.1$ and $1.1 - 6.0 \text{ GeV}^2/c^4$, and are found to be compatible with the SM at the level of $2.1 - 2.5$ standard deviations [52]. Whilst these levels of significance are inconclusive on their own, these results are compatible with an interpretation in terms of NP involving \mathcal{C}_9 that also explains the P'_5 anomaly.

In light of these anomalous measurements, complementary approaches that can further assist in distinguishing the NP hypothesis are paramount. One interesting possibility is to examine LFU breaking in angular observables of $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ decays [53–55]. In particular, the proposed observables are optimised so that charm-induced long-distance contributions can be kinematically suppressed. A first attempt on this type of measurement has been performed by Belle [10], and the results are compatible with the SM predictions – with large statistical uncertainties. As a first stage towards this physics programme at LHCb, a preliminary measurement of the q^2 dependent CP -averaged angular observables, S_i , of both $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ to $B^0 \rightarrow K^{*0} e^+ e^-$ decays, based on 3.0 fb^{-1} of data collected by LHCb during 2011 and 2012, has been performed. These preliminary results are documented in Dr. F. Lionetto's thesis [56] and provided the baseline studies for the analysis discussed in this note.

53 This note describes the first measurement of the CP -averaged and form-factor indepen-
54 dent angular observables of the decay $B^0 \rightarrow K^{*0}e^+e^-$ at LHCb. The analysis is based on a
55 data sample corresponding to an integrated luminosity of 5.0 fb^{-1} , collected by the LHCb
56 experiment in pp collisions at a centre-of-mass energy of 7, 8 and 13 TeV. The remainder
57 of the note is arranged as follows. Section 2 briefly introduces the analysis strategy, Sec. 3
58 presents the samples of data and simulated events and Sec. 4. summarises the selection.
59 The mass and angular model for the signal and backgrounds are described in Secs. 5,
60 6 and 7, whilst the full unbinned fit (including sensitivity studies) and the angular fit
61 validation for the resonant mode are shown in in Secs. 8 and 9. Systematic uncertainties
62 are determined in Sec. 10, cross-checks/pre-unblinding test reported in Sec. 11 and the
63 results and conclusions are gathered in Sec. 12.

64 2 Analysis strategy

65 While global fit analyses of nonstandard measurements in ratios of $b \rightarrow s\ell^+\ell^-$ processes
66 - such as R_{K^*} - and angular observables in $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decays suggest a nontrivial
67 consistent pattern, an experimental connection between these two sets of results is yet to be
68 demonstrated. A series of LFU breaking observables have been proposed in order to validate
69 this hypothesis, from (weighted) difference in CP -averaged angular observables [53–55]
70 to the direct determination of observables that encode potential non-equal couplings of
71 muons and electrons [57]. As initial stage in this physics programme, the CP -averaged
72 (*e.g.* F_L and S_3 – S_9) and so-called *optimised* observables ($P_i^{(t)}$ series [58]) are measured
73 for the first time at LHCb in $B^0 \rightarrow K^{*0}e^+e^-$ decays, using RUN-I (2011 and 2012) and
74 RUN-IIa (2015 and 2016) datasets, corresponding to approximately 3.0 fb^{-1} and 2.0 fb^{-1} ,
75 respectively. The strategy used in this analysis is as follows

- 76 i. The data and simulation samples used in this analysis are introduced in Sec. 3.
77 In particular, the simulation is corrected to account for some known discrepancies
78 between simulation and data, which originate from the mismodelling of the kinematics/multiplicity of the event and the detector response (*e.g.* particle identification
79 and trigger efficiency). Note that while these corrections are implemented in a similar
80 fashion as the standard analyses of LFU ratios of branching fractions, such effects
81 have a more limited impact on the angular observables.
- 83 ii. The selection requirements are designed with similar features to those used in the R_{K^*}
84 RUN-I analysis [52] and in the current update, but optimised to reduce possible non-
85 uniformities in the phase space of the event, as shown in Sec. 4. These criteria result
86 in a sample of approximately $\mathcal{O}(300)$ $B^0 \rightarrow K^{*0}e^+e^-$ candidates in the dielectron
87 invariant mass region described in the following. The remaining background is
88 modelled in the fit.
- 89 iii. A detailed description of the theoretical framework underlying the study of the signal
90 decay $B^0 \rightarrow K^{*0}e^+e^-$ is given in Sec. 5. This focusses mainly on the formalism and
91 convention for the angular amplitude but also discusses the signal parametrisation
92 of the $K\pi e^+e^-$ invariant mass distribution. Unlike the muonic counter channel, a
93 stable and unbiased determination of the signal yield within our selection can only
94 be obtained with the full angular fit, and thus, it is reported in the results.
- 95 iv. The reconstruction and selection result in a distortion of the angular distributions of
96 the final state particles. These effects, known as acceptance effects, are parametrised
97 with Legendre polynomials in a similar fashion to Ref. [7]. Due to non-negligible
98 acceptance variations within the nominal q^2 interval (*e.g.* $[1.1, 7.0]\text{ GeV}^2/c^4$), a
99 q^2 -dependent per-event weighting approach is chosen for this analysis.
- 100 v. The mass and angular parametrisation of the different sources of background is
101 discussed in Sec. 7. Both simulated samples and data-driven approaches (*e.g.* the
102 LFV $B \rightarrow K^*\mu e$ sample) are used in this task.

- 103 vi. The unbinned maximum likelihood is finally introduced in Sec. 8, including a series
 104 of pseudo-experiments that are generated to examine the expected sensitivity of this
 105 measurement for the SM scenario (with all possible experimental aspects involved in
 106 the analysis) and a discussion on the statistical treatment. At this stage, the data
 107 fit to $B^0 \rightarrow K^{*0}e^+e^-$ decays and the corresponding signal/background yields are
 108 reported for the first time - note that the angular observables are still kept blinded.
 109 This is an important step for this analysis since it provides the relevant information
 110 to commission the angular model.
- 111 vii. A comprehensive full validation of the framework in simulation and data using the
 112 control channel $B \rightarrow K^*J/\psi(e^+e^-)$ is presented in Section 9.
- 113 viii. Systematic uncertainties are described in Sec 10. While this measurement is expected
 114 to be limited by statistical uncertainties, these studies include all contributions typi-
 115 cally involved in angular analyses. In addition, cross checks and the pre-unblinding
 116 procedure are reported in Section 11.
- 117 ix. The fit results section will be filled after the unblinding (see Sec. 12). The results
 118 we aim to provide in this section and in the paper consist in the CP -averaged
 119 observables F_L , A_{FB} , $S_{3,4,5,7,8,9}$ and corresponding $P_i^{(\prime)}$ basis for both q^2 regions
 120 discussed in Sec. 2.1. Moreover, the hypothesis of LFU breaking in the angular
 121 distributions can be examining by computing the difference between muons and
 122 electrons, *e.g.* $\Delta S_i = S_i^\mu - S_i^e$ or $Q_i = P'_\mu - P'_e$ [54], which are observables free from
 123 hadronic uncertainties and long-distance charm-loop contributions. Notice that other
 124 weighted difference observables such as D_i [55] are not quoted in this measurement.

125 2.1 Dilepton invariant mass regions

126 Two q^2 regions are mainly studied in this analysis, *i.e.* the commonly referred central q^2 bin
 127 for the rare mode $B^0 \rightarrow K^{*0}e^+e^-$ and the q^2 bin for the resonant mode $B^0 \rightarrow K^{*0}J/\psi(\rightarrow$
 128 $e^+e^-)$. In order to improve the resolution and the precision of this analysis, an alternative
 129 definition of the dilepton invariant mass is used, *i.e.* the so-called constrained q^2 , hereafter
 130 referred to as q_c^2 . This variable is computed by constraining the signal candidates to
 131 originate from the primary vertex and to have an invariant mass corresponding to the
 132 nominal mass of the B^0 meson. This requirement is expected to reduce the migration
 133 of events whilst imposing a good rejection power on background sources for which the
 134 assumption of invariant mass equal to the nominal mass of the B^0 does not hold. A more
 135 comprehensive discussion on this strategy is given in Sec. 4.

136 Two alternative intervals of the central region are investigated in this analysis for the
 137 signal candidates of the rare mode; a q_c^2 bin between 1.1 and $7.0\text{ GeV}^2/c^4$ or within 1.1
 138 and $6.0\text{ GeV}^2/c^4$. The resonant mode is instead required to have a q_c^2 between 7.0 and
 139 $11.0\text{ GeV}^2/c^4$. The lower bound in the central q_c^2 bin, which is set to $1.1\text{ GeV}^2/c^4$, is such
 140 that the background contribution due to $\phi \rightarrow e^+e^-$ decays is not expected to contribute.
 141 This choice is also consistent with previous LFU measurements at LHCb. The largest

bound between rare and resonant mode, set to $7.0 \text{ GeV}^2/c^4$, is due to the fact that the contamination from the radiative tail of the J/ψ resonance becomes too large above this value. Previous LFU measurements at LHCb considered a q^2 bound of $6.0 \text{ GeV}^2/c^4$ between rare and resonant mode. However, the introduction of the requirement on q_c^2 allows to move this threshold up to $7.0 \text{ GeV}^2/c^4$, which increases the statistics available in the rare mode; and thus is the nominal range of this measurement. In order to assess the impact of this choice as well as to guarantee the consistency with respect to the muonic channel, the measurement is also performed considering an upper q_c^2 bound of $6.0 \text{ GeV}^2/c^4$.

In summary, this analysis investigates the regions defined as

- [central- q^2]: q_c^2 in $[1.1, 7.0] \text{ GeV}^2/c^4$ or q_c^2 in $[1.1, 6.0] \text{ GeV}^2/c^4$;
- [J/ψ - q^2]: q_c^2 in $[7.0, 11.0] \text{ GeV}^2/c^4$.

Moreover, the following definition of q^2 is also adopted throughout the note

- **data and reconstructed-level simulation:** $q^2 = |p_{\ell^+} + p_{\ell^-}|^2$;
- **truth-level simulation:** $q_{\text{true}}^2 = |p_{B^0} - p_{K^{*0}}|^2$;

the latter being used to account (subtract) for final state radiation (FSR) of the leptons.

2.2 Angular basis

The final state of the decay $B^0 \rightarrow K^{*0} e^+ e^-$ can be described by q^2 , the invariant mass squared of the di-electron system, and three decay angles $\vec{\Omega} = (\theta_\ell, \theta_K, \phi)$. Different angular conventions are typically used by experimentalists and theorists; while most theorists follow the definitions in Refs. [59, 60], this analysis uses the convention adopted in previous LHCb publications [6, 7]. The angular basis used in this analysis is illustrated in Fig. 3. The angle θ_ℓ is defined as the angle between the direction of the e^+ (e^-) in the di-electron rest frame and the direction of the di-electron in the B^0 (\bar{B}^0) rest frame. The angle θ_K is defined as the angle between the direction of the kaon in the K^{*0} (\bar{K}^{*0}) rest frame and the direction of the K^{*0} (\bar{K}^{*0}) in the B^0 (\bar{B}^0) rest frame. The angle ϕ is the angle between the plane containing the e^+ and e^- and the plane containing the kaon and pion from the K^{*0} . Explicitly, $\cos \theta_\ell$ and $\cos \theta_K$ are defined as

$$\cos \theta_\ell = \left(\hat{p}_{e^+}^{(e^+ e^-)} \right) \cdot \left(\hat{p}_{e^+ e^-}^{(B^0)} \right) = \left(\hat{p}_{e^+}^{(e^+ e^-)} \right) \cdot \left(-\hat{p}_{B^0}^{(e^+ e^-)} \right), \quad (1)$$

$$\cos \theta_K = \left(\hat{p}_{K^+}^{(K^{*0})} \right) \cdot \left(\hat{p}_{K^{*0}}^{(B^0)} \right) = \left(\hat{p}_{K^+}^{(K^{*0})} \right) \cdot \left(-\hat{p}_{B^0}^{(K^{*0})} \right) \quad (2)$$

for the B^0 and

$$\cos \theta_\ell = \left(\hat{p}_{e^-}^{(e^+ e^-)} \right) \cdot \left(\hat{p}_{e^+ e^-}^{(\bar{B}^0)} \right) = \left(\hat{p}_{e^-}^{(e^+ e^-)} \right) \cdot \left(-\hat{p}_{\bar{B}^0}^{(e^+ e^-)} \right), \quad (3)$$

$$\cos \theta_K = \left(\hat{p}_{K^-}^{(K^{*0})} \right) \cdot \left(\hat{p}_{K^{*0}}^{(\bar{B}^0)} \right) = \left(\hat{p}_{K^-}^{(K^{*0})} \right) \cdot \left(-\hat{p}_{\bar{B}^0}^{(K^{*0})} \right) \quad (4)$$

¹⁷⁰ for the \bar{B}^0 decay. The definition of the angle ϕ is given by

$$\cos \phi = \left(\hat{p}_{e^+}^{(B^0)} \times \hat{p}_{e^-}^{(B^0)} \right) \cdot \left(\hat{p}_{K^+}^{(B^0)} \times \hat{p}_{\pi^-}^{(B^0)} \right) , \quad (5)$$

$$\sin \phi = \left[\left(\hat{p}_{e^+}^{(B^0)} \times \hat{p}_{e^-}^{(B^0)} \right) \times \left(\hat{p}_{K^+}^{(B^0)} \times \hat{p}_{\pi^-}^{(B^0)} \right) \right] \cdot \hat{p}_{K^{*0}}^{(B^0)} \quad (6)$$

¹⁷¹ for the B^0 and

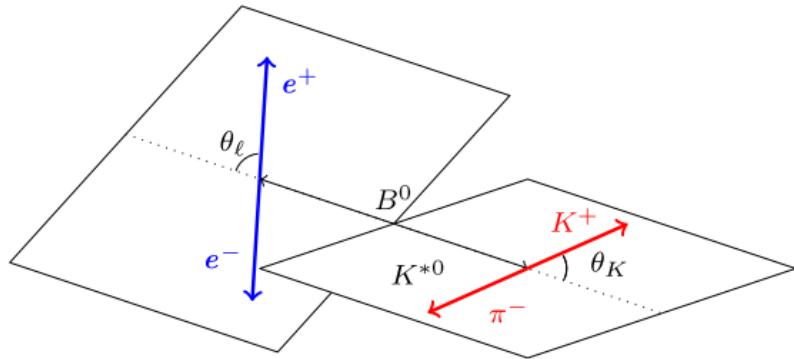
$$\cos \phi = \left(\hat{p}_{e^-}^{(\bar{B}^0)} \times \hat{p}_{e^+}^{(\bar{B}^0)} \right) \cdot \left(\hat{p}_{K^-}^{(\bar{B}^0)} \times \hat{p}_{\pi^+}^{(\bar{B}^0)} \right) , \quad (7)$$

$$\sin \phi = - \left[\left(\hat{p}_{e^-}^{(\bar{B}^0)} \times \hat{p}_{e^+}^{(\bar{B}^0)} \right) \times \left(\hat{p}_{K^-}^{(\bar{B}^0)} \times \hat{p}_{\pi^+}^{(\bar{B}^0)} \right) \right] \cdot \hat{p}_{\bar{K}^{*0}}^{(\bar{B}^0)} \quad (8)$$

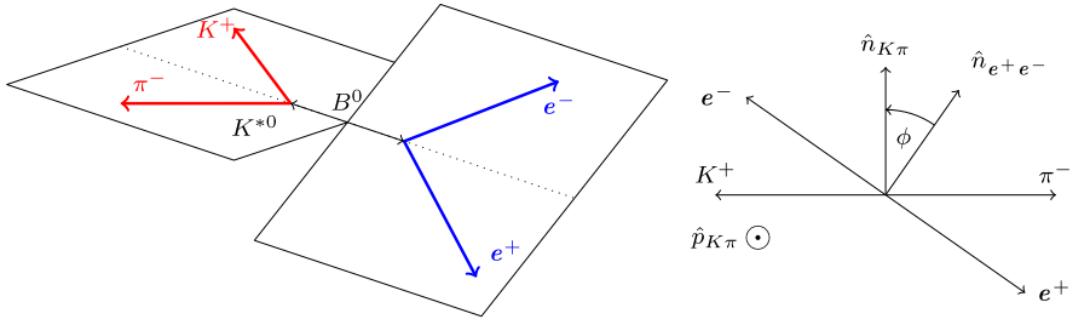
¹⁷² for the \bar{B}^0 decay. The $\hat{p}_X^{(Y)}$ are unit vectors describing the direction of a particle X in
¹⁷³ the rest frame of the system Y . In every case the particle momenta are first boosted to
¹⁷⁴ the B^0 (or \bar{B}^0) rest frame. In this basis, the angular definition for the \bar{B}^0 decay is a CP
¹⁷⁵ transformation of that for the B^0 decay.

¹⁷⁶ 2.3 Blinding procedure

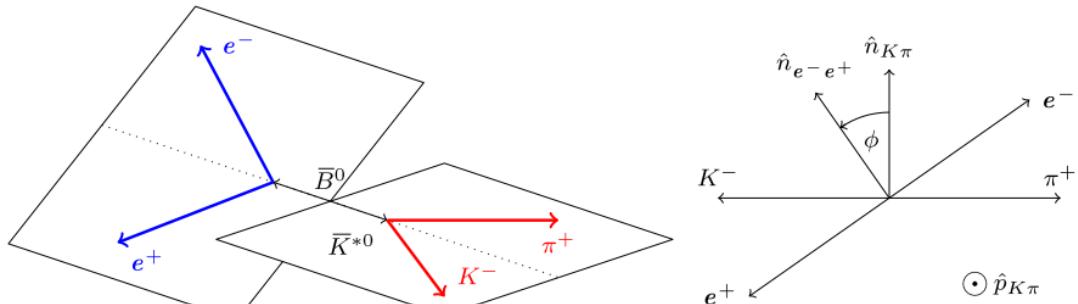
¹⁷⁷ The decay mode $B^0 \rightarrow K^{*0} e^+ e^-$ has been previously studied at LHCb for RUN-I [52], and
¹⁷⁸ more recently re-examined using the full RUN-II dataset (ongoing). In particular, the
¹⁷⁹ invariant mass and the signal/background yields have been already unblinded. Therefore,
¹⁸⁰ despite of the different q^2 definitions and selection, for this analysis these quantities are
¹⁸¹ also not preserved during the review process. Moreover, the angular distributions are also
¹⁸² examined since a proper modelling of the underlying background is not only challenging
¹⁸³ but crucial to obtain a stable and unbiased data fit. Finally, the angular parameters are
¹⁸⁴ the key blinded quantities in this measurement. Similarly to other LFU analyses, a staged
¹⁸⁵ unblinded procedure is foreseen, which aims to validate the internal consistency of our
¹⁸⁶ measurements as discussed in Sec. 11.



(a) θ_K and θ_l definitions for the B^0 decay



(b) ϕ definition for the B^0 decay



(c) ϕ definition for the \bar{B}^0 decay

Figure 3: Graphical representation of the angular basis used for $B^0 \rightarrow K^{*0}e^+e^-$ and $\bar{B}^0 \rightarrow \bar{K}^{*0}e^+e^-$ decays in this paper. The notation \hat{n}_{ab} is used to represent the normal to the plane containing particles a and b in the B^0 (or \bar{B}^0) rest frame. An explicit description of the angular basis is given in the text.

3 Dataset, trigger and stripping

The simulation and data samples used for this analysis are produced with the similar conditions as the current update of the R_{K^*} analysis, but have been produced separately. In order to further guarantee the compatibility between these analyses, a detailed validation of these samples have been performed and no difference at the pre-selection level is seen. The samples utilised in this analysis can be found in:

- /eos/lhcb/wg/RD/Bd2Kstee-Angular-Centralq2/data/

3.1 Data

The angular analysis described in this note is based on data corresponding to an integrated luminosity of approximately 5.0 fb^{-1} , collected by the LHCb experiment in pp collisions. The dataset comprises of 3 fb^{-1} of integrated luminosity collected respectively at a centre of mass energy of $\sqrt{s} = 7 \text{ TeV}$ and 8 TeV in 2011 and 2012, and 2.0 fb^{-1} collected at $\sqrt{s} = 13 \text{ TeV}$ in 2015 and 2016. Candidates are reconstructed in RUN-I with the Reco14 version of the reconstruction software, while 2015 and 2016 are produced with Reco15a and Reco16, respectively. These events have been filtered with the criteria defined in the Stripping Bu2LLKeeLine2 that is discussed in Sec. 3.4, within the campaigns 21r1p1a, 21r0p1a, 24r1 and 28r1p1. The only exception are datasets produced from the Stripping Bu2LLKmeLine that are used to study the background composition, which uses the same configuration as the nominal lines. Finally, data samples were prepared using the same version of DaVinci as the one used in the Stripping. This corresponds to v39r1p1 for RUN-I, and v42r6p1 for the 2015 and 2016 datasets.

3.2 Simulation

Simulated events are used in all stages of the analysis, including the study of the background contamination, training of the multivariate classifier and determination of the efficiencies. Samples are generated separately for 2011 (TCK 0x40760037), 2012 (TCK 0x409f0045), 2015 (TCK 0x411400a2) and 2016 (TCK 0x6138160F). These are produced using mostly Sim09 versions of the simulation software (except some backgrounds with Sim08) with PHOTOS always included to generate radiative photons. Moreover, these are all filtered productions with the corresponding Stripping campaigns, apart from particle identification criteria that are removed from the requirements. Table 2 lists the signal, control and background simulated samples used in the analysis with their event type and number of events. Generator-level simulations are also privately produced in order to normalise the angular acceptance without PHOTOS and any cuts.¹

Some considerations regarding the physics model used for the signal channels are required, in particular with respect to the form factors that parametrise the strong

¹Private samples have been generated with the same conditions as the normal LHCb production with Gauss versions of v49r8 and v49r10.

Table 2: Summary of the simulated samples of signal and background modes.

Mode	Event type	Sim	Events
$B^0 \rightarrow K^{*0} e^+ e^-$	11124002	Sim09c	42 M
$B^0 \rightarrow K^{*0} e^+ e^-$ (Flat- q^2)	11124008	Sim09h	42 M
$B^0 \rightarrow K^{*0} J/\psi(e^+ e^-)$	11154001	Sim09c	46 M
$B^0 \rightarrow K^{*0} \psi(2S)(e^+ e^-)$	11154011	Sim09c	11 M
$B^+ \rightarrow K^+ \pi^- \pi^+ e^+ e^-$	12125000	Sim09c	79 M
$B^+ \rightarrow K_1^+(K^+ \pi^- \pi^+) e^+ e^-$	12425000	Sim08g, Sim09c	16 M
$B^+ \rightarrow K_2^{*+}(K^+ \pi^- \pi^+) e^+ e^-$	12425011	Sim08h, Sim09d	9.0 M
$B^+ \rightarrow K^+ e^+ e^-$	12123003	Sim09(b, c, i)	14 M
$B_s^0 \rightarrow \phi e^+ e^-$	13124006	Sim09(b, d, f, h)	16 M
$\Lambda_b^0 \rightarrow p K^- e^+ e^-$	15124011	Sim09(b, c)	5M
$B_s^0 \rightarrow \phi J/\psi(e^+ e^-)$	13154001	Sim09(b, f, h)	37 M
$B^+ \rightarrow X J/\psi(e^+ e^-)$	12952000	Sim08(a, b, d, g), Sim09d(e, g)	60 M
$B^0 \rightarrow X J/\psi(e^+ e^-)$	11453001	Sim08(c, d, g), Sim09d	32 M
$B_s^0 \rightarrow X J/\psi(e^+ e^-)$	13454001	Sim08(c, d, g), Sim09(e, g)	60 M
$\Lambda_b^0 \rightarrow p K^- J/\psi(e^+ e^-)$	15154001	Sim08(i, h), Sim09(b, c)	9 M
$B^0 \rightarrow D^-(\rightarrow K^{*0} e^- \bar{\nu}_e) e^+ \nu_e$	11584022	Sim09d	21 M

217 interaction effects in decays of B mesons to lighter mesons (*i.e.* $K^{*0}(892)$). Different
 218 non-perturbative techniques can be used to calculate these form factors, for instance QCD
 219 Sum Rules on the Light-Cone ‘‘LCSR’’ (low q^2 , extrapolation to high q^2 necessary) and
 220 Lattice (high q^2). The nominal decay file used in this analysis is 11124002, which uses
 221 LCSR from Ref. [61] and implements in the `EvtGen` class ‘‘`EvtbTos11BallFF`’’. In addition,
 222 the underlying effective Wilson Coefficients $\mathcal{C}_{7,9}$ are based in Refs. [62, 63], which due to
 223 the break down of the perturbative expansion when q^2 approaches hadronic resonances
 224 have a different q^2 modelling for $q^2/m_b^2 > 0.25 \simeq 5.75 \text{ GeV}^2/c^4$. Effects related to the use of
 225 this simulation are considered in the systematics. Finally, a simulation sample generated
 226 with the angles and the q^2 distribution approximately flat is used to investigate possible
 227 effects related to resolution.

228 Moreover, a comprehensive study of the phase space distribution of $B^\pm \rightarrow$
 229 $[K^\pm \pi^\mp \pi^\pm] e^+ e^-$, which represent one of the main background of the analysis, is not available
 230 in the literature nor at the experimental level. Multibody decays can, in general, proceed
 231 through intermediate quasi-n-body and nonresonant amplitudes, which can significantly

232 modify its branching fraction contribution and its shape in the invariant mass of interest.
 233 For this purpose three different physics model are examined: “12425000” based exclusively
 234 on $B^+ \rightarrow K_1^\pm(1720)e^+e^-$ transitions, with the intermediate resonant states $K^{*0}(892)\pi^+$
 235 (20%), $K^+\rho^0(770)$ (40%), $K^+\omega$ (< 1%) and a non-resonant term (40%); “12125000” with
 236 the so-called LSFLAT model, generated flat in the $K^\pm\pi^\mp\pi^\pm$; and “12425011” with a similar
 237 contribution of the first model, but in this case from the $B^+ \rightarrow K_2^{*+}(1430)e^+e^-$ transition.

238 Finally, the $B^0 \rightarrow D^-(\rightarrow K^{*0}e^-\bar{\nu}_e)e^+\nu_e$ decay has a branching ratio four orders of
 239 magnitude larger than that of the signal and it is one of the most important source
 240 of background for this analysis. In order to investigate the best strategy (if any) to
 241 remove/reduce this contribution, a series of fully simulated background events has been
 242 generated with the decay files “11584021” and “11584022”. The main decay chain is
 243 common to both models, *i.e.* PHOTOS HQET2 1.185 1.081, while the D meson decay is
 244 either produced with PHOTOS ISGW2 or PHSP. Since the PHOTOS ISGW2 model is known to
 245 not provide an accurate description of the phase-space, the PHSP generation can provide
 246 an easy template to be re-weighted with more physics motivated models. Moreover, a
 247 more efficient generation can be obtained by imposing a restrict window of the visible
 248 mass at the generator level. The value is chosen to be larger than 4.5 GeV, based on
 249 toyMC ensembles produced with **RapidSim** [64].

250 3.3 Trigger

251 Events are triggered by the lines listed in Tab. 3, where the logical *and* of the L0, HLT1
 252 and HLT2 lines, and the logical *or* of lines of the same level is required. Candidates are
 253 required to be Triggered-On-Signal (TOS) at all stages, except for L0Global for which they
 254 have to be Triggered-Independently-of-Signal (TIS).

255 Since the distribution of the signal candidates in the electron mode depends on the
 256 nature of the particle that has triggered the event (lepton or hadron), the MC and data
 257 samples are split in two independent trigger categories, which are treated separately in
 258 each stage of the corrections/selections. The nominal trigger categories are defined as
 259 follows:

Table 3: Summary of the trigger lines used to select the ee channels. All lines are required to be TOS, unless otherwise stated.

Run-I	2015/16
LOElectron	
LOGlobal(TIS)	
Hlt1TrackAllL0	Hlt1TrackMVA
	Hlt1TwoTrackMVA
Hlt2Topo(E) [2,4] BodyBBDT	Hlt2Topo(E) [2,4] Body

- 260 • **L0I**: events triggered by particles not in the signal candidate, *i.e.* `B0_L0Global_TIS`,
 261 • **L0E**: events triggered by at least one of the electrons in the signal candidate and
 262 not triggered by particles that are not in the signal candidate, *i.e.*
 263 `(E1_L0Electron_TOS | E2_L0Electron_TOS) & !(B_L0Global_TIS);`

264 where the nomenclature L0I and L0E is used throughout the text. These two categories
 265 are chosen for this analysis in order to have mutually exclusive datasets with comparably
 266 equivalent statistics. Note that this definition differs from previous published results
 267 including electrons, *e.g.* R_K and R_{K^*} , in which a priority to a statistical dominant
 268 inclusive `L0Electron` decision has been made, and two other decisions (`L0Hadron` and
 269 `L0Global TIS`) with lower statistics also have been examined. However, such approach
 270 limits the available statistics to properly model data/simulation discrepancies from proxy
 271 samples, and thus it is not followed in this analysis. In particular, the exclusive efficiency
 272 of the hadronic trigger category is rather low (below $< 10\%$ of the full statistics) and such
 273 inclusion would impose again difficulties in tuning the misalignments in simulation with
 274 no significant statistical gain. Therefore, such trigger line is not considered in this analysis.
 275 Further details on this choice are given in Sec. 3.6.4.

276 3.3.1 L0 threshold issue

277 The transformation of ECAL ADC counts into the transverse energy deposited in the
 278 calorimeter, E_T , includes the usage of cell-dependent calibration constants. At the
 279 beginning of the 2016 data taking, their values were inadvertently reset to the default ones,
 280 and the data was recorded in these conditions during all year. In this sense, the simulation
 281 utilised for this analysis for the signal and control modes are based in the `Sim09c` version
 282 or above, which calibration constants are aligned with the one used during the data taking.

283 3.3.2 TISTOS bug

284 During the 2016 data taking, charged particles that were selected by both `Turbo` and `Full`
 285 streams of the HLT gave false negative TOS decision when matched to offline candidates.
 286 As a consequence this led to a wrong trigger decision for some analyses. In particular,
 287 this is known to affect Stripping lines that require TIS/TOS in their selection. Since the
 288 Stripping `Bu2LLKeeLine2` used for this analysis does not require this criteria at this level,
 289 this measurement is not affected by this issue. Moreover, the `DaVinci` software used in
 290 the tuple production were also processed with more recent versions that already included
 291 the necessary corrections.

292 3.3.3 AALLSAMEBPV bug

293 Another issue presented in the 2016 dataset is referred to as `AALLSAMEBPV` bug. The
 294 functor `ALLSAMEPV` is responsible to ensure that all tracks come from the same primary
 295 vertex (PV). However, this functor was effectively requiring a single PV in the event. This

296 bug can potentially enter in two stages of the selection: at the topological trigger decision
297 or at the Stripping level. Whilst at the HLT level the `AALLSAMEBPV` bug is essentially
298 ignored given the requirement of the minimum χ^2_{IP} of the tracks with respect to the PV to
299 be larger than 16, the Stripping line imposes only $\chi^2_{IP} < 9$ in the final state particles. This
300 issue is addressed as a possible source of systematics, which is expected to be marginal
301 since it should not significant vary as a function of the angular phase space.

302 3.3.4 `Sim09[d,e,f]` multiplicity bug

303 Due to a bug in the `Sim09[d,e,f]` conditions at generating level, the multiplicities in
304 those samples are lower compared to `Sim09c` samples. The signal and resonance MC
305 samples used in this analysis exclude those versions and thus this bug does not affect our
306 measurement.

307 3.4 Stripping

308 The stripping versions adopted in the analysis are produced within the campaigns `21r1p1a`,
309 `21r0p1a`, `24r1` and `28r1p1` of the (`Bu2LLKeeLine2`). Table 4 lists all the requirements
310 applied at this stage.

311 The `Bu2LLKeeLine2` line uses the `DiElectronMaker` to create dielectron candidates.
312 This tool improves the efficiency in the reconstruction of events containing an e^+e^- pair
313 with a low invariant mass, by improving the invariant mass resolution and avoiding
314 double-counting of the bremsstrahlung photons added. The energy of any photon whose
315 position in the ECAL is compatible with the extrapolation of an electron track from the
316 first tracking station is added to the electron energy. If the same bremsstrahlung photon
317 candidate is found for both the e^+ and the e^- tracks, the photon energy is added randomly
318 to one of the two. The corrected e^+ and e^- tracks are then extrapolated backward in order
319 to determine the dielectron vertex and a fit is performed to reconstruct the dielectron
320 momentum.

321 The `Bu2LLKmeLine` line is also used in the analysis (blinding the signal region) in
322 order to have a better understanding of the possible combinatorial and double semi-
323 leptonic background contributions. This stripping has been designed to be identical to
324 the `Bu2LLKeeLine2` line, apart from the requirements on the leptons. Therefore, the fake
325 J/ψ candidate is reconstructed by combining the electron with a muon with the same
326 kinematics criteria. The only exception are requirements on the muons, which are further
327 identified by imposing `isMuon` and `hasMuon`.

328 3.5 Truth matching

329 Reconstructed candidates of signal and control channels are always truth-matched using
330 one of the categories below

- 331 • Corrections to simulation and acceptance modelling: truth-matching performed
332 by requiring that the particle, mother and grand-mother `TRUEID` fulfil the relevant

Table 4: Summary of the requirements of the `Bu2LLKeeLine2` stripping.

Requirement	
Event	$N_{\text{PV}} \geq 1$
	$n_{\text{SPD}} \leq 600$
	$ m - m_{B^0}^{\text{PDG}} < 1000 \text{ MeV}/c^2$
	$\text{DIRA} > 0.9995$
B^0	χ_{IP}^2 (primary) < 25
	$\chi_{\text{vtx}}^2/\text{ndf} < 9$
	PV χ^2 separation > 100
	$ m - m_{K^{*0}}^{\text{PDG}} < 300 \text{ MeV}/c^2$
K^{*0}	$p_{\text{T}} > 500 \text{ MeV}$
	$\chi_{\text{vtx}}^2/\text{ndf} < 25$
K	$\text{DLL}_{K\pi} > -5$
	χ_{IP}^2 (primary) > 9
π	χ_{IP}^2 (primary) > 9
	$m < 5500 \text{ MeV}/c^2$
e^+e^-	$\chi_{\text{vtx}}^2/\text{ndf} < 9$
	PV χ^2 separation > 16
	$\text{DLL}_{e\pi} > 0$
e	$p_{\text{T}} > 300 \text{ MeV}/c$
	χ_{IP}^2 (primary) > 9

decay chain; this effectively corresponds to candidates classified as (0) *Signal*, (10) *Quasi-signal* or (50) *Low-mass background* by `TupleToolMCBackgroundInfo`.

- Mass fits: truth-matching performed by requiring that the candidate background category is *Signal*, *Quasi-signal*, *Low-mass background* or (60) *Ghost*.

3.6 Corrections to simulation

While the design of the set of data/simulation corrections follows a similar philosophy as the angular analysis of the muonic mode [65], it has been slightly modified in order to account for the different features of final states involving electrons. Moreover, in order to account for further possible missing features in this strategy, the alternative correction chain designed for the R_{K^*} is used as a source of systematic effect.

The scripts used in this section can be found in:

- `ewp-bd2ksteeangular-central-q2/scripts/corrections`

339 One of the important aspects for this analysis are the simulation sample corrections
340 developed to account for some known discrepancies between simulation and data. These
341 discrepancies originate from features such as mis-modelling of the kinematic of the event, as
342 well as from a mis-modelling of the detector response in terms of particle identification and
343 trigger efficiency. This is needed since the simulation is used as a proxy of the signal and
344 background contributions in several parts of the analysis, for example in the multivariate
345 analysis and acceptance parametrisation.

346 In particular, there are four steps on this procedure, as enumerated in the following

- 347 1. the particle identification response;
- 348 2. TCK trigger requirements;
- 349 3. the generated event multiplicity and B kinematics;
- 350 4. and the trigger response (*i.e.* efficiency).

351 The following sub-sections describe in detail how the simulated samples of signal and
352 control channels are corrected to better reproduce the data. However, for the benefit of the
353 reader some initial definitions are in place. The first stage of corrections is implemented
354 using a re-sampling procedure, where new PID responses are added to the sample. To
355 further improve the alignment of the emulated trigger in the simulation, a procedure is
356 implemented to better match the response of the TCK in data. The subsequent correction
357 makes use of a multivariate technique that essentially provides even-by-event weights,
358 w_{BDT} , that are normalised to the total number of events. In turn, the absolute efficiency at
359 this stage can be obtained simply from summing the weights of the remaining events after
360 a given selection divided by the previous step. Notice that the corrections implemented
361 up to this stage are used in the training of the multivariate classifier. Finally, the trigger
362 efficiency corrections are computed on top of all the previous steps and provide an efficiency
363 correction factor, w_{Trig} . Further details on the use of these weights in the determination of
364 the angular acceptance map is provided in Sec. 6.

365 3.6.1 Particle identification correction

The scripts used to perform the particle identification resampling as well as all validation
plots shown in this section can be found in:

- 366 ○ `ewp-bd2ksteeangular-central-q2/scripts/corrections/pid`

367 An important feature of the LHCb PID algorithm is the observed large disagreement
368 between the simulation and the actual detector response. This is in general caused
369 by several second-order effects (*e.g.* detector occupancy) and changes in the operation
370 conditions during the data taken that are not properly described in the simulation. A
371 common practice to improve this performance is obtained via a data-driven approach. In
372 particular, the PID calibration procedure (**PIDCalib**) exploits high-purity control samples

373 (i.e. $D^{*+} \rightarrow D^0\pi^+$, $\Lambda^0 \rightarrow p\pi^-$, $\Lambda_c^+ \rightarrow pK^+\pi^-$ for hadrons, $J/\psi \rightarrow \mu^+\mu^-$ for muons, and
374 $B^+ \rightarrow J/\psi(e^+e^-)K^+$ for electrons) to correct the known poorly modelled PID distributions
375 in the LHCb simulation.

376 An interesting application of this technique is the re-sampling of the PID simulation
377 response based on the kinematics of the calibration dataset (usually some combination
378 of p , p_T , η and/or number of tracks). This idea has been first implemented for the
379 $B^0 \rightarrow K^{*0}\mu^+\mu^-$ analysis and more recently made available in the **PIDCalib** package
380 via the **MCResampling** tool. The procedure consists in obtaining for each particle a
381 multidimensional THnSparse, which is used to re-sample the PID response according to
382 the corresponding distribution. Although this method is a good approximation for most
383 of the calibration samples, it has certain limitations. Firstly, the PID descriptors which
384 involve neural network discriminant (**ProbNN** variables) depend on many more parameters
385 of the track than the standard features momentum and pseudorapidity. Secondly, different
386 PID variables for the same track are strongly correlated (e.g., the kaon-like track is less
387 pion-like, therefore there has to be negative correlation between ProbNNK and ProbNNpi
388 variables). As a result, one cannot combine different PID variables after resampling
389 since the correlations will be ignored. In theory this can be overcome by resampling the
390 combination of variables itself [e.g. ProbNNK \times (1 – ProbNNpi)], but this in turn is not a
391 solution in the event that these variables are used in a multivariate classifier. The use of
392 several PID variables in the multivariate classifier requires the knowledge of correlations
393 between them, and can only be accomplished by increasing the dimensionality of the
394 PDF to be resampled. For instance, in order to use both ProbNNK and ProbNNpi in
395 the classifier one needs to describe the calibration data in five dimensions: ProbNNK,
396 ProbNNpi, p , η and number of tracks. This requires coarser binning and thus reduces the
397 precision of PID resampling.

398 Finally, the PID resampling is problematic from the point of view of evaluation of
399 systematic uncertainties. If one needs to vary the PID response (for instance, to assess
400 the effect of binning or the calibration sample size), resampling will produce a statistically
401 independent sample for each PDF variation. As a result, after applying the selection which
402 includes the PID variables, the selected samples will in turn be statistically independent,
403 and thus the determination of selection efficiency for each systematic variation will include
404 a statistical factor.

405 An alternative approach has been recently proposed within the **Meerkat** framework [66]
406 to overcome at least the main limitations of the calibration sample reweighting and PID
407 resampling techniques. The basic idea behind is to transform a PID variable in such a way
408 that its distribution matches the one seen in the calibration sample (for any kinematics
409 of the track), and that the transformed variable is strongly correlated with the output
410 of simulation. In this case, the correlations of the PID variable present in simulation
411 (with other PID variables for the same track, or other parameters of the track and the
412 event in general) will be preserved. Details of the **Meerkat** PID response correction and
413 its implementation in the LHCb software are given in Appendix A. Figure 4 shows a
414 representative set of simulated corrected-signal for the 2012 configuration. There is in
415 general a good agreement over the entire range of all the spectra, and any further residual

416 differences are accommodated as systematics. In order to obtain further insights on the
417 effect of the data/simulation corrections for the angular analysis, it is possible to examine
418 the relative impact as a function of the phase space. Figures 5 and 6 illustrate this
419 effect separately for the Run1 and Run2 conditions for the various angular distributions.
420 Note that in general non-flat corrections to the angular distributions are only observed
421 for $\cos \theta_K$ (especially for RunI), where the unbalance of the PID response correction for
422 pions and kaons induces these effects. Nevertheless, the overall very small deviation from
423 a uniform PID correction response already suggests that systematic effects related to
424 the PID corrections should be suppressed. Similar figures are found for the specific year
425 configurations and can be found in Appendix A.4.

426 **3.6.2 TCK trigger requirements**

427 The alignment of the TCK used in the simulation with respect to the data is described in
this section. The procedure is identical to the R_{K^*} methodology, with the exception
of the *prior chain* correction used to obtain the generator level corrections. In this
analysis, a different approach is used to obtain such correction (as described in Sec. 3.6.3),
and thus, this is not included in this discussion. For completeness, the TCK alignment
designed for the R_{K^*} measurement is reproduced in verbatim in the following.

428 Real data are collected using different *Trigger Configuration Keys* (TCKs), encoding the
429 set of trigger lines, reconstruction sequence properties and threshold cuts applied in the
430 trigger reconstruction and selection chain (`L0`, `HLT1`, `HLT2`). The LHCb simulation, however,
431 emulates those conditions based on a single fixed TCK only which leads to discrepancies
432 between data and simulation. In order to minimise the discrepancies introduced by
433 mismatching TCKs between data and simulation two approaches are followed:

- 434 1. When the simulated TCK is aligned with the loosest data TCK it is sometimes
435 possible to emulate the tighter data cuts in simulation offline. This emulation is
436 performed according to the fractions of TCKs used during actual data taking. This
437 approach is used to achieve a better agreement at the HLT1 selection level in 2012
438 and 2016, as detailed in the following.
- 439 2. In the case the TCKs used in the simulation is tighter than the ones used in the
440 data, an additional data selection is applied using criteria matched to simulation, as
441 described below. The same approach is applied to the resonant modes.

442 All the alignment(s) in term of TCKs and thresholds have been extracted using the `TCKsh`
443 application from the `Moore` package.

444 **HLT1 threshold alignment in 2012**

445 For the 2012 data taking period, the `HLT1AllTrackL0` trigger line selects only those events
446 which have at least one reconstructed track at HLT1 level satisfying the requirements shown

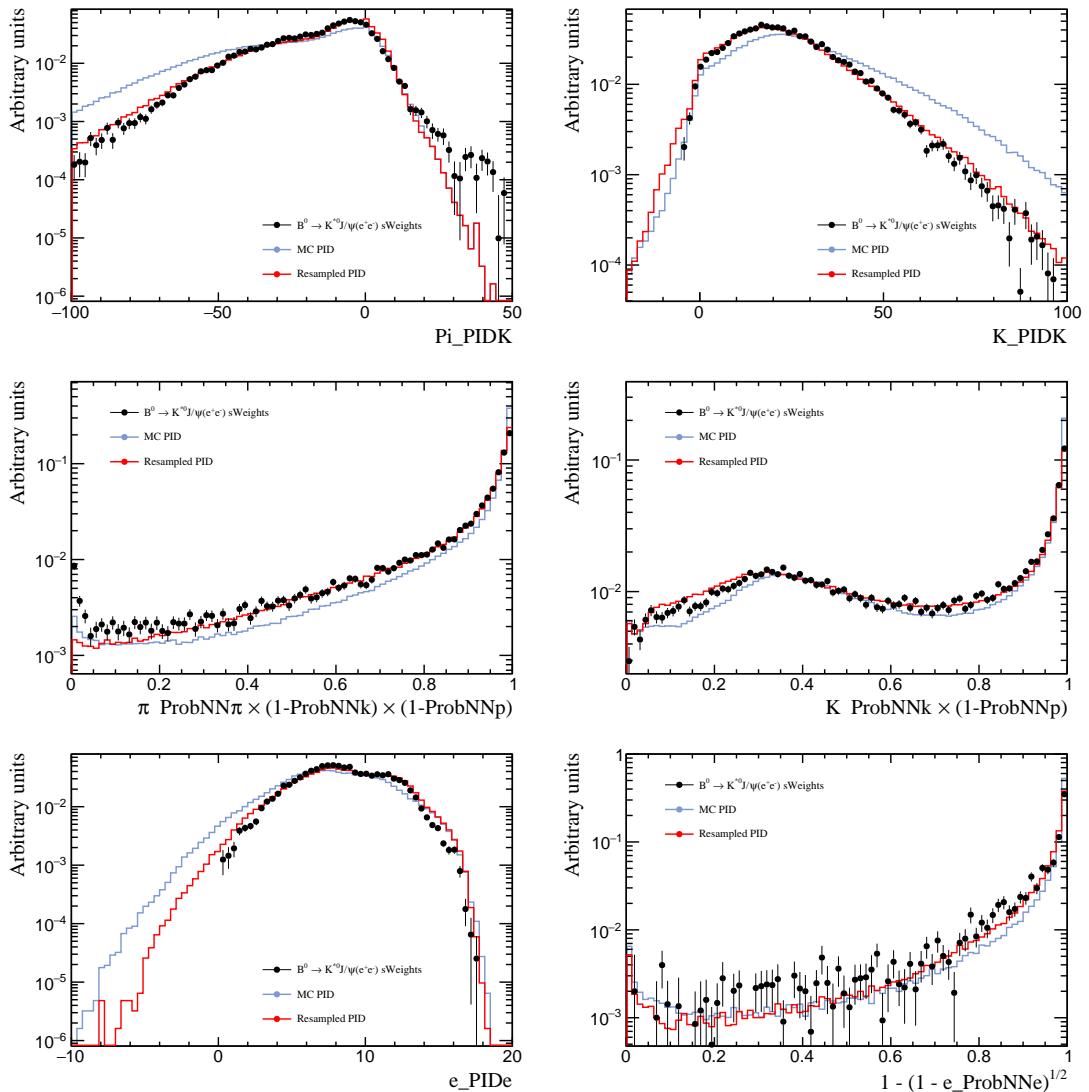


Figure 4: Particle identification distributions of the responses utilised in the analysis for sWeighted $B^0 \rightarrow K^*0 J/\psi(e^+e^-)$ data (points with the error bars), uncorrected (blue solid line) and corrected (red solid line) Monte Carlo.

in Table 5. The exact cut values depend on the TCK and only the loosest TCK is modelled in simulation. Since the TCK in simulation always has looser or equivalent cuts as the data TCKs, it is possible to align a similar fraction of the simulation to the tighter TCKs that are present in data in order to emulate the mixture of TCKs in data properly. The p_T considered for electrons is the one measured without brem-recovery, since no brems recovery is performed in HLT1.

454 HLT1 threshold alignment in 2016

455

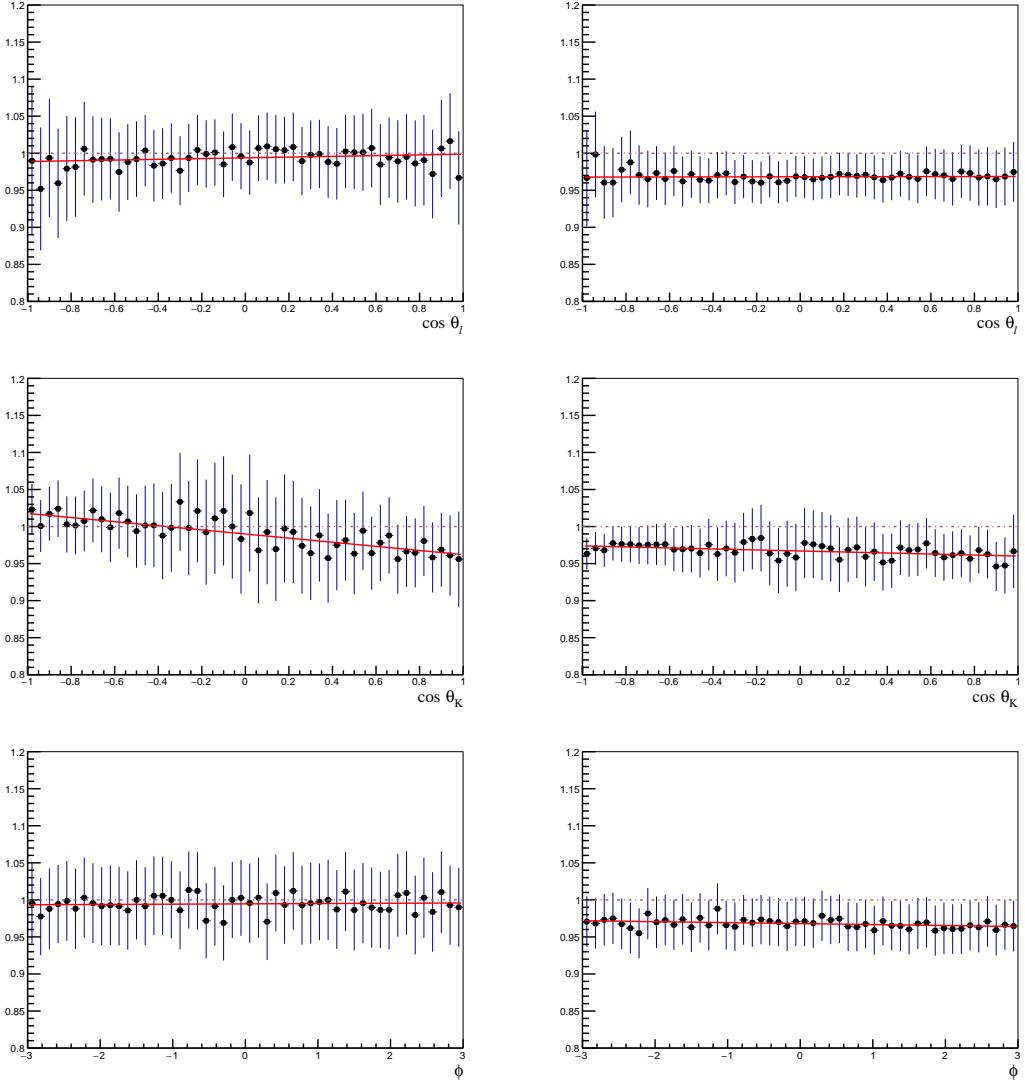


Figure 5: Ratio of the angular distributions between the PID corrected simulation and uncorrected ensembles for the truth-matched signal channel for the (left) RunI and (right) II conditions.

456 The 2016 alignment follows a similar approach, with the difference that the B candidate
 457 must be TOS with respect to the HLT1TrackMVA². Its selection obtained from the Moore
 458 TCKsh application for the HLT1TrackMVA trigger line is the following :

$$\begin{aligned}
 & (p_T(\text{GeV}) > 25 \text{ AND } \log \chi_{IP}^2 > 7.4) \\
 & \text{OR} \\
 & (\log \chi_{IP}^2 > \frac{1}{(p_T(\text{GeV}) - 1)^2} + \frac{b}{25} \cdot (25 - p_T(\text{GeV})) + \log(7.4))
 \end{aligned} \tag{9}$$

²Notice that the HLT1TwoTrackMVA criteria is also used in this analysis. However, no dedicated alignment has been performed for this line. The impact of this decision will be investigated in the systematics.

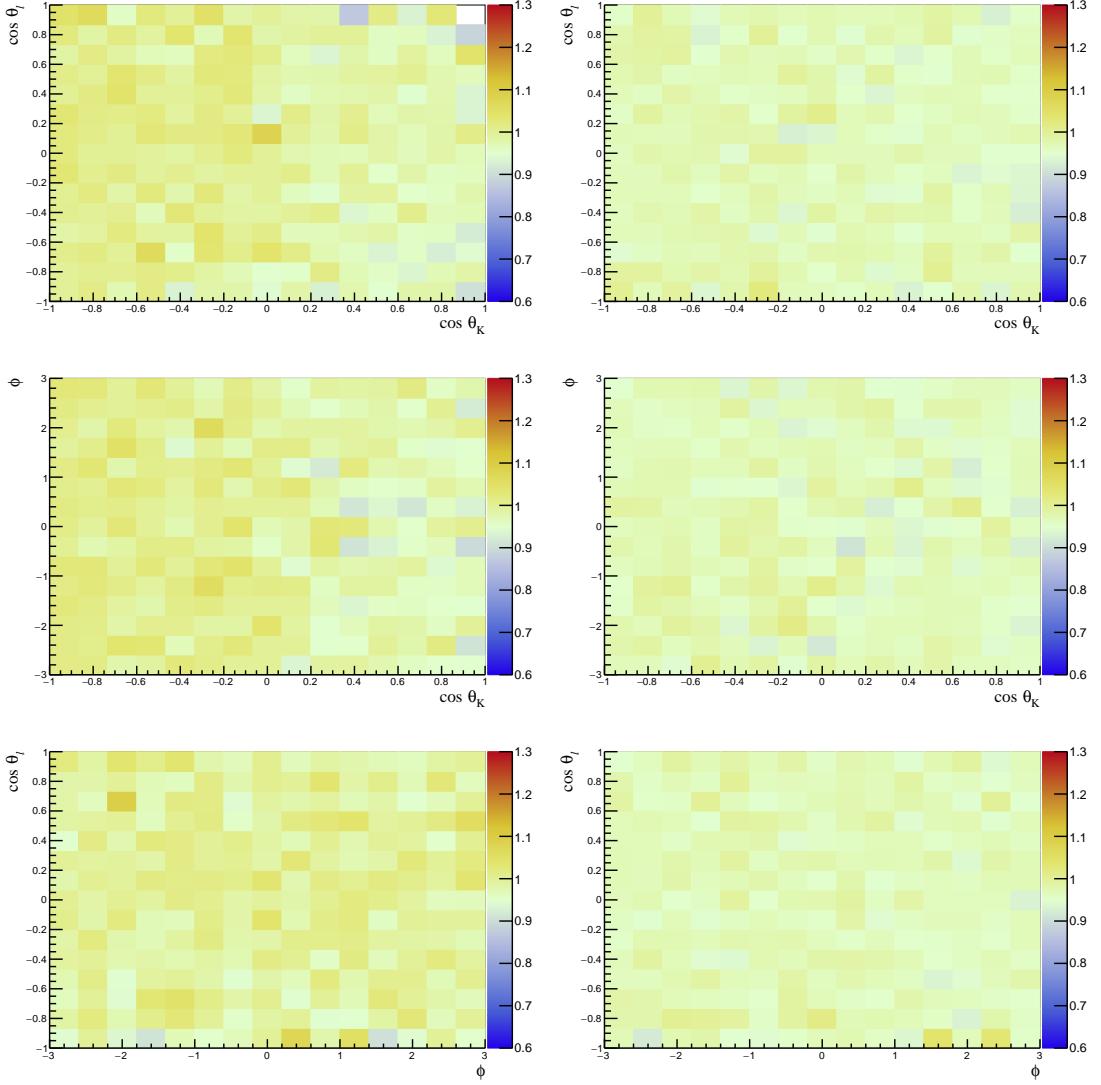


Figure 6: Two-dimensional ratio of the angular distributions between the PID corrected simulation and uncorrected ensembles for the truth-matched signal channel for the (left) RunI and (right) II conditions.

459 The differing trigger thresholds are defined in Table 6, where the variation of the
 460 thresholds in the various TCKs are encoded in the parameter b in Eq. 9. For simulation
 461 the parameter is always set to $b = 1.1$ (fixed TCK in production), while data are taken
 462 with a mixture of three different b values. The simulation is aligned by emulating the
 463 HLT1 selection according to the fraction of data taken with different b values, and aligning
 464 each fraction separately, re-defining the HLT1TrackMVA_TOS decision. The values of b and
 465 the fractions used in data that are emulated in simulation can be found in Tab. 6. The
 466 p_T considered for electrons is the one measured without brem-recovery, since no brems
 467 recovery is performed in HLT1.

HLT TCK value	% data	p [GeV]	p_T [GeV]	IP [mm]	IP χ^2	χ^2_{ndof} (track)	HLT TCK value	% data	p [GeV]	p_T [GeV]	IP [mm]	IP χ^2	χ^2_{ndof} (track)
MC	100.0	> 3.0	> 1.6	> 0.1	>16.0	<2.0	MC	100.0	>3	>1.6	>0.1	>16.0	<2.0
0x990042	19.2	> 3.0	> 1.6	> 0.1	>16.0	<2.0	0x990042	35.75	>3	>1.6	>0.1	>16.0	<2.0
0x97003d	15.19	> 10	> 1.7	> 0.1	>16.0	<1.5	0xa30044	17.52	>3	>1.6	>0.1	>16.0	<2.0
0xa30044	14.67	> 3.0	> 1.6	> 0.1	>16.0	<2.0	0x94003d	16.03	>10	>1.7	>0.1	>16.0	<1.5
0x990044	11.21	> 3.0	> 1.6	> 0.1	>16.0	<2.0	0x97003d	11.98	>10	>1.7	>0.1	>16.0	<1.5
0xac0046	10.43	> 3.0	> 1.6	> 0.1	>16.0	<2.0	0xa10045	7.19	>3	>1.6	>0.1	>16.0	<2.0
0x94003d	9.7	> 10	> 1.7	> 0.1	>16.0	<1.5	0x990044	3.03	>3	>1.6	>0.1	>16.0	<2.0
0xa90046	6.37	> 3.0	> 1.6	> 0.1	>16.0	<2.0	0xa10044	2.67	>3	>1.6	>0.1	>16.0	<2.0
0x8c0040	6.31	> 10	> 1.7	> 0.1	>16.0	<2.0	0x9f0045	2.66	>3	>1.6	>0.1	>16.0	<2.0
0xab0046	5.24	> 3.0	> 1.6	> 0.1	>16.0	<2.0	0xa30046	1.89	>3	>1.6	>0.1	>16.0	<2.0
0xa30046	1.12	> 3.0	> 1.6	> 0.1	>16.0	<2.0	0xac0046	0.88	>3	>1.6	>0.1	>16.0	<2.0
0x860040	0.45	> 1.0	> 1.7	> 0.1	>16.0	<2.25	0x990043	0.26	>3	>1.6	>0.1	>16.0	<2.0
0x95003d	0.1	> 10	> 1.7	> 0.1	>16.0	<2.0	0xa20044	0.17	>3	>1.6	>0.1	>16.0	<2.0
0x8e0040	0.05	> 10	> 1.7	> 0.1	>16.0	<2.0	0xad0046	0.01	>3	>1.6	>0.1	>16.0	<2.0
0x7f0040	0.02	> 1.0	> 1.7	> 0.1	>16.0	<2.25	0x9a0042	0.01	>3	>1.6	>0.1	>16.0	<2.0
0xad0046	0.01	> 3.0	> 1.6	> 0.1	>16.0	<2.0	0x95003d	0.01	>10	>1.7	>0.1	>16.0	<2.0
0x7e003a	0.01	> 1.0	> 1.3	> 0.1	>16.0	<2.25							

Table 5: TCK content of `HLT1AllTrackL0` decision for the 2012 Magnet down(left) and up(right) data taking period. The blue rows highlight the TCK values in data which configure the lines with a different selection w.r.t. simulation.

Table 6: Fractions of candidates with HLT threshold cuts corresponding to the parameter b on 2016 data. The parameter b is parametrizing the threshold cut given in Eq. 9.

Polarity	$b = 1.1 [\%]$	$b = 1.6 [\%]$	$b = 2.3 [\%]$
MagDown	89.8	0.0	10.2
MagUp	30.0	15.5	54.5

468 L0E trigger alignment selection

470 In addition to the obvious nominal misalignment of the TCK in the simulation, the ageing
 471 of the calorimeter causes the meaning of a data E_T cut to vary with time, which is not
 472 the case in simulation. It is therefore necessary to align the electron TCK thresholds by
 473 emulating an L0 selection which is tighter than either the data or simulation TCK criteria,
 474 and retrospectively applying it to both data and simulation.

475 The measured transverse energy at L0 level at the calorimeter for a given final state
 476 particle is not directly accessible. This information is accessed offline using the projected
 477 E_T at the calorimeter location using the `L0CaloTool` tool in `DaVinci`. An extra selection
 478 is performed in both data and simulation for the L0E trigger category:

$$E\{1, 2\} \text{--L0Electron_TOS} \rightarrow E\{1, 2\} \text{--L0Electron_TOS} \& E_T^{L0}(E\{1, 2\}) > \text{TCKthreshold} \quad (10)$$

479 The threshold values are configured in the following way: the E_T^{L0ADC} values tabulated
 480 for each data taking year are compared. They are expressed in ADC counts accounting
 481 for the conversion factors.³ The choice of these thresholds is given to align the simulation

³The conversion factors used to convert the ADC counts to transverse energy measured at the calorimeter, *i.e.* $E_T^{L0}(e)$, is 20 MeV/ADC in 2015 and 2016, and 24 MeV/ADC for 2011 and 2012.

482 and data. In particular if the data has looser thresholds than simulation, to avoid that
 483 there will be regions where there is no simulation with which to derive L0 corrections. The
 484 values used in the analysis are summarised in Tab. 7.

Year	$E_T^{L0} [MeV]$
2011	>2500.
2012	>3000.
2015	>3000.
2016	>2700.

Table 7: L0E threshold alignment selections applied offline to both data and simulation to re-define the $E\{1, 2\}_\text{L0Electron_TOS}$ trigger decision.

485 3.6.3 Reconstruction, kinematic and multiplicity corrections

486 A similar correction strategy as the angular analysis of the muonic channel is designed
 487 to this measurement, *i.e.* differences between simulation and data in reconstructed
 488 variables are corrected by comparing distributions of the control channel events in data
 489 to the corresponding distributions for simulated events. The scripts used in this section
 490 can be found in:

- `ewp-bd2ksteeangular-central-q2/scripts/corrections/reco`

491 The mis-modelling of the kinematic and multiplicity of the event as well as of various
 492 reconstruction variables is corrected by performing a multidimensional re-weighting. The
 493 variables that show the largest deviation between simulation and data are typically the
 494 number of tracks, B^0 pt , η , flight distance, impact parameter χ^2 and decay vertex χ^2 .
 495 A multidimensional re-weighting based on a Boosted Decision Tree (BDT) approach is
 496 discussed in the following. This method has several advantages compared to the standard
 497 histogram re-weighting, as the one used previously in Ref. [52]. In the BDT re-weighting
 498 approach the multidimensional space of the considered features is split in several regions
 499 by using decision trees. The optimal regions are identified by maximising the symmetrised
 500 binned χ^2 defined as

$$\chi^2 = \sum_{\text{reg}} \frac{(w_{\text{original}}^{\text{reg}} - w_{\text{target}}^{\text{reg}})^2}{w_{\text{original}}^{\text{reg}} + w_{\text{target}}^{\text{reg}}},$$

501 where $w_{\text{original}}^{\text{reg}}$ and $w_{\text{target}}^{\text{reg}}$ are the total weights of the events in the selected region of
 502 the target and original distributions. The performance of the re-weighting procedure
 503 is evaluated by training a multivariate classifier to distinguish between simulation and
 504 data before and after the BDT re-weighting is applied. If the re-weighter improves the
 505 agreement between simulation and data, in fact, the discrimination between the two
 506 becomes harder when the correction is applied. The performance of this algorithm can be

503 evaluated by comparing two ROC curves, corresponding to two different trainings of the
504 same multivariate classifier, one based on original sample and target sample and another
505 based on re-weighted sample and target sample.

506 A dedicated BDT re-weighting framework has been designed for this analysis [67], with
507 particular emphasis on the optimisation of the hyper-parameter space of the algorithm as
508 well as the stability/robustness of a re-weighting metric. Further details on this approach
509 are provided in Appendix B. This procedure consists in a single stage BDT reweighting
510 obtained from the resonant charmonium mode.⁴ Notice that in order to avoid biases
511 in the procedure and re-use of the “calibration samples”, a k-fold approach is adopted,
512 where 9/10 of the data are used to determine the correction, which is then applied to the
513 remaining 1/10.

514 The control channel $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ is used as input for the multidimensional
515 re-weightings. A clear signal is visible in data after the preselection (Sec. 4.1) and the
516 background vetoes (Sec. 4.2) are applied. Nevertheless, the *sPlot* technique [68] is used to
517 statistically subtract the remaining background by fitting the PV- and J/ψ -constrained
518 invariant mass distribution of the B^0 candidates. The fit is performed separately for each
519 year (but combined with both trigger categories) and the model chosen is similar to the
520 one used in the final mass fit (more details on the procedure are given in Appendix B).

521 The distributions of the number of tracks, $B^0 p_T$, η , flight distance, impact parameter
522 χ^2 and decay vertex χ^2 are the features used to correct the mis-modelling of the kinematic
523 of the event.⁵ The improvement in the agreement between data and simulation can be
524 seen in Fig .7, which show the comparison between data and MC distributions before
525 and after the correction is applied for the 2012 data. The distributions that undergo the
526 major changes due to the BDT re-weighting are the number of tracks and $B^0 p_T$, while the
527 changes of the other distributions are mostly marginal. The re-weighted distributions do
528 not match perfectly the target distributions since a stronger re-weighting of the selected
529 features would induce a disagreement in some other variables used in the analysis as well
530 as it would be affected by overfitting. In order to obtain further insights on the effect
531 of the data/simulation corrections for the angular analysis, it is possible to examine the
532 relative impact as a function of the phase space. Figures 8 and 9 illustrate this effect
533 separately for the Run1 and Run2 conditions for the various angular distributions. The
534 response of the BDT correction is found to be extremely uniform in all of the angular
535 distributions. Similar figures are found for the specific year configurations and can be
536 found in Appendix B.

⁴Alternative two-stage procedures to obtain a *prior* mis-modelling correction at the “generation” level have also been investigated using $B^0 \rightarrow K^{*0} J/\psi (\rightarrow \mu^+ \mu^-)$ decays (either from TOS or TIS exclusive samples). While the single stage BDT has been chosen as the nominal correction for this analysis, the two-stage alternative chain is investigated as a systematic source inherited to this decision.

⁵An alternative strategy in which the number of SPD hits is included in the BDT re-weighting is also investigated. It is observed that the agreement between the simulation and data distributions of most of the features is worse, particularly the number of tracks is significantly worse than the original after the correction is applied. For this reason, it is preferred to adopt the strategy removing this variable.

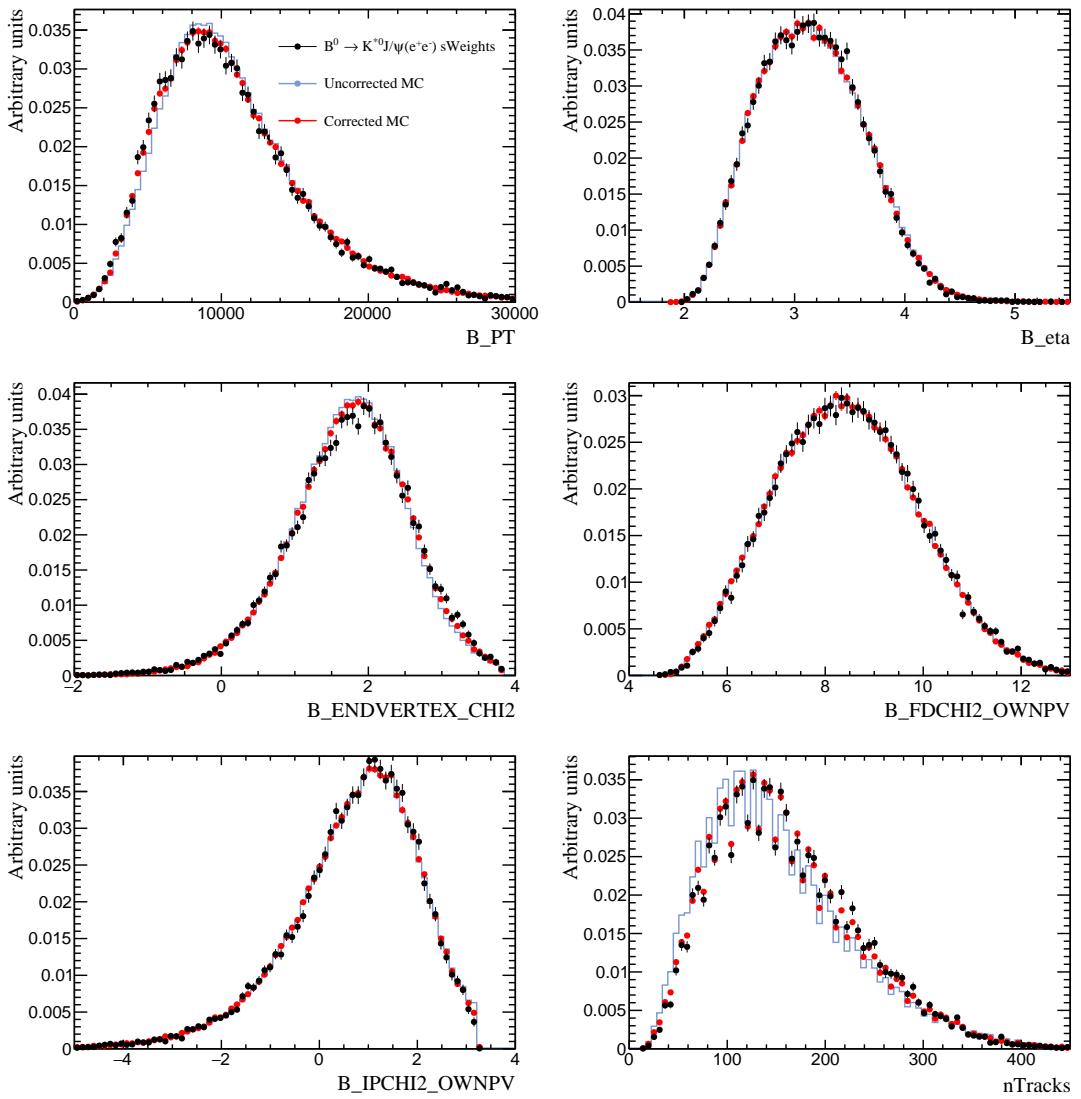


Figure 7: Distribution of (top left) the $B^0 p_T$, (top right) $B^0 \eta$, (middle left) B^0 decay vertex χ^2 , (middle right) B^0 flight distance χ^2 , (bottom left) B^0 IP χ^2 and (bottom right) number of tracks, for $B^0 \rightarrow K^{*0}J/\psi(\rightarrow e^+e^-)$ sWeighted data (black), and simulation before (blue) and after (red) the correction is applied, for the 2012 dataset.

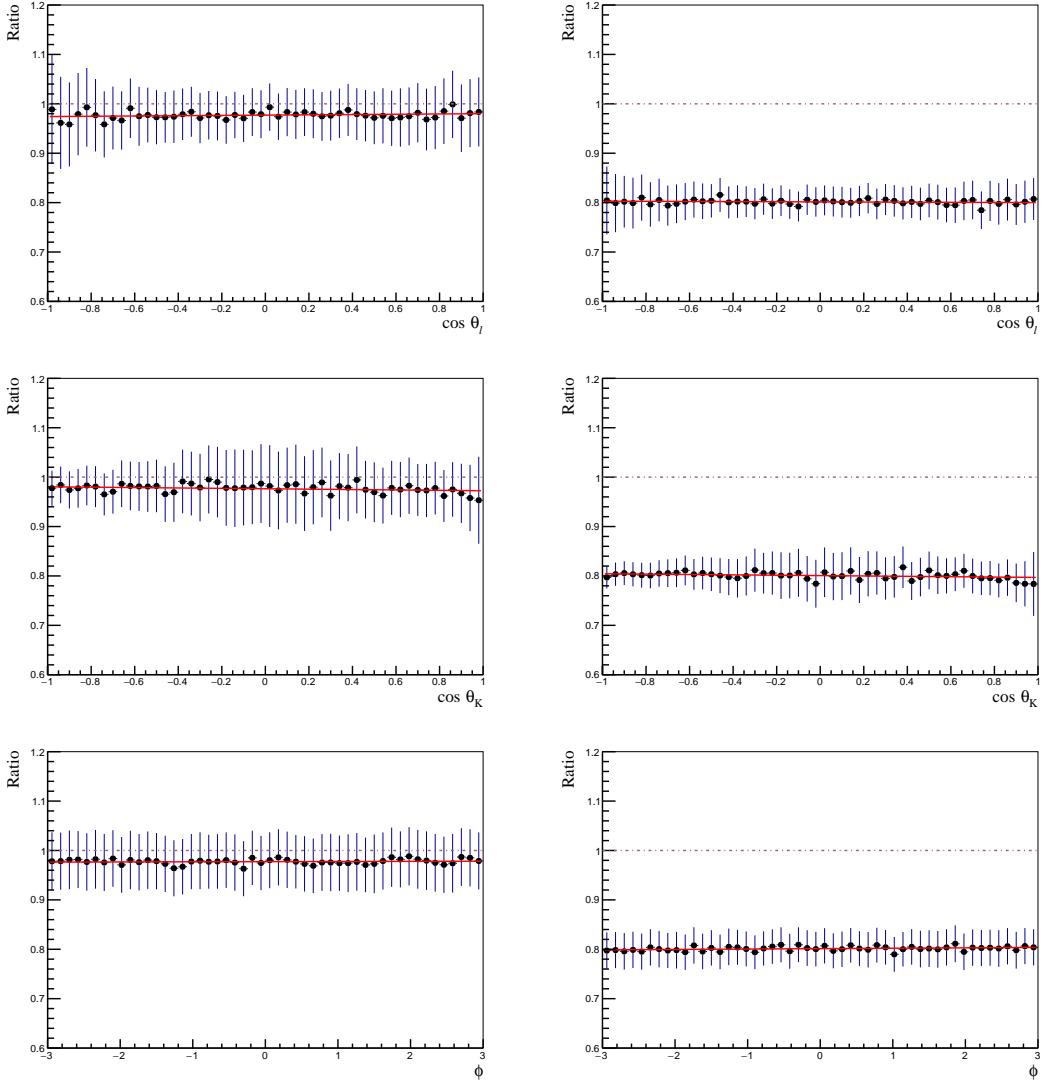


Figure 8: Ratio of the angular distributions between the reconstruction corrected simulation and uncorrected ensembles for the truth-matched signal channel for the (left) RunI and (right) II conditions.

3.6.4 Trigger correction

Differences between simulation and data trigger response have been extensively studied in LHCb for final states involving electrons. Therefore, the strategy to correct for possible misalignments follows the one employed in other lepton-flavour universality measurements and, for brevity, summarised in the following.

The scripts used in this section can be found in:

- o `ewp-bd2ksteeangular-central-q2/scripts/corrections/`

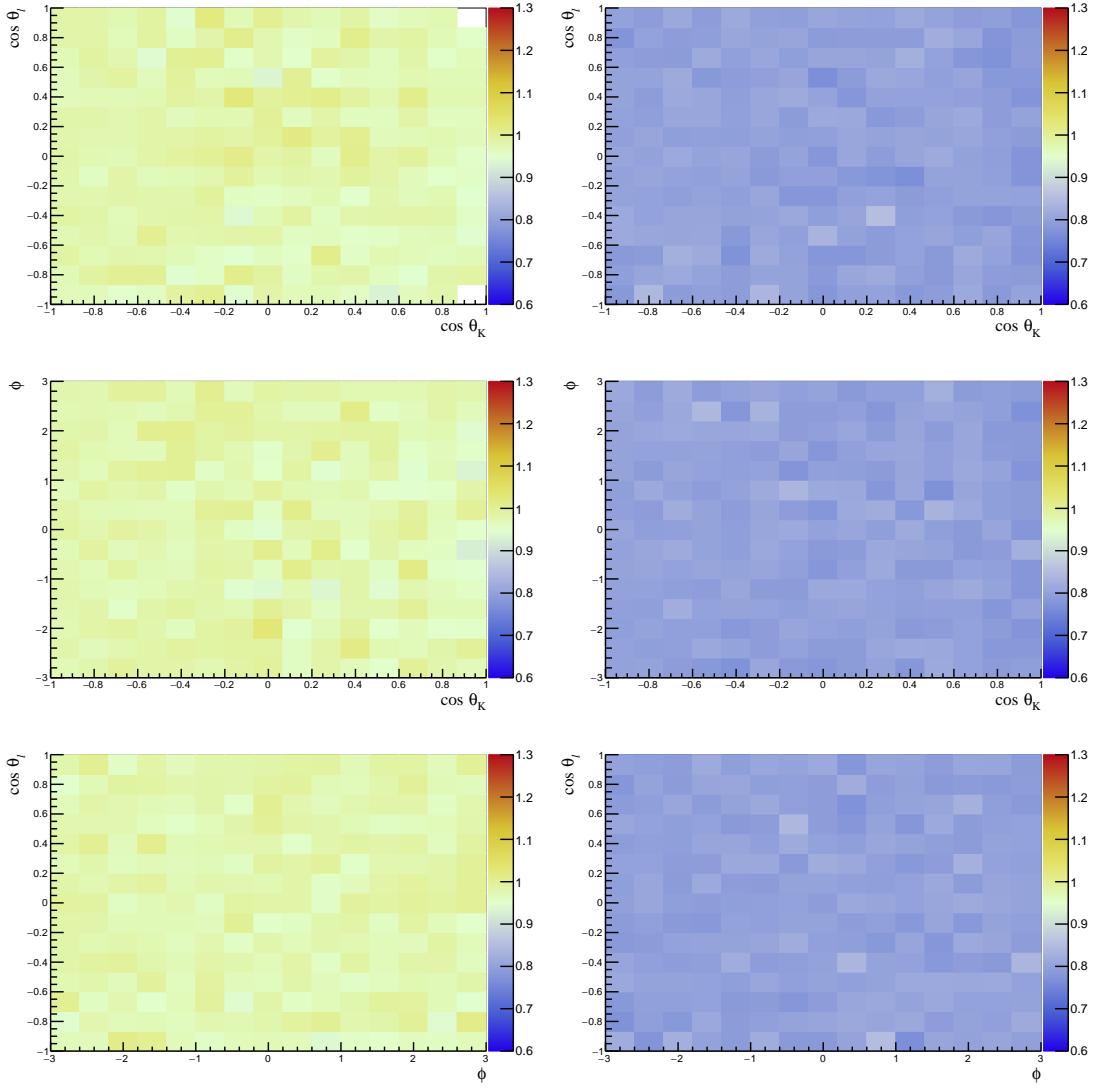


Figure 9: Two-dimensional ratio of the angular distributions between the reconstruction corrected simulation and uncorrected ensembles for the truth-matched signal channel for the (left) RunI and (right) II conditions.

539 The trigger efficiency, *i.e.* the efficiency of triggering the signal candidates that have
 540 been reconstructed by the detector, is not well described in the simulation. This is due
 541 to its dependency on the detector occupancy and the correlation between the signal and
 542 the underlying event. To reduce the differences between data and simulation, a correction

543 factor is assigned to each simulated event in the trigger categories L0I and L0E, so that
544 the efficiency of the simulated sample matches the one observed in data. This is done
545 through the use of two different weights: the first one, w_{L0} , is used to correct the L0 trigger
546 efficiency while the second one, w_{HLT} , is obtained after applying the L0 correction and is
547 used to correct the HLT trigger efficiency. In both cases the efficiencies have been obtained
548 taking into account the weights coming from the BDT reweighting procedure discussed
549 above. The trigger efficiencies ϵ_{data} and ϵ_{MC} are estimated using a tag-and-probe approach,
550 where the tag selection is chosen to be as uncorrelated as possible with respect the selection
551 of interest to minimise any bias. These calibration histograms are then used to obtain
552 event-by-event corrections to the simulation. In the following, a brief discussion on the
553 criteria chosen for these reference histograms and the impact of this in the simulation are
554 shown.

555 **L0 trigger corrections**

556
557 The L0 efficiencies are calibrated separately for each year of data-taking using the resonant
558 mode $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ after the preselection chain and the PID criteria are applied
559 (see Sec. 4). To further reduce the background contributions in the samples and to study
560 the impact of possible correlation between trigger levels, the HLT requirements are also
561 imposed in some cases. Pure signal samples can be obtained from data without the use of
562 the *sPlot* technique by applying a cut of ± 60 MeV around the J/ψ constrained B^0 mass,
563 $m(K\pi\ell\ell)_{\text{PV}-J/\psi}$. In the following are presented the calibration histograms for (i) TIS and
564 (ii) single electron TOS efficiencies obtained for each year.

565 [i] **TIS calibration histogram**

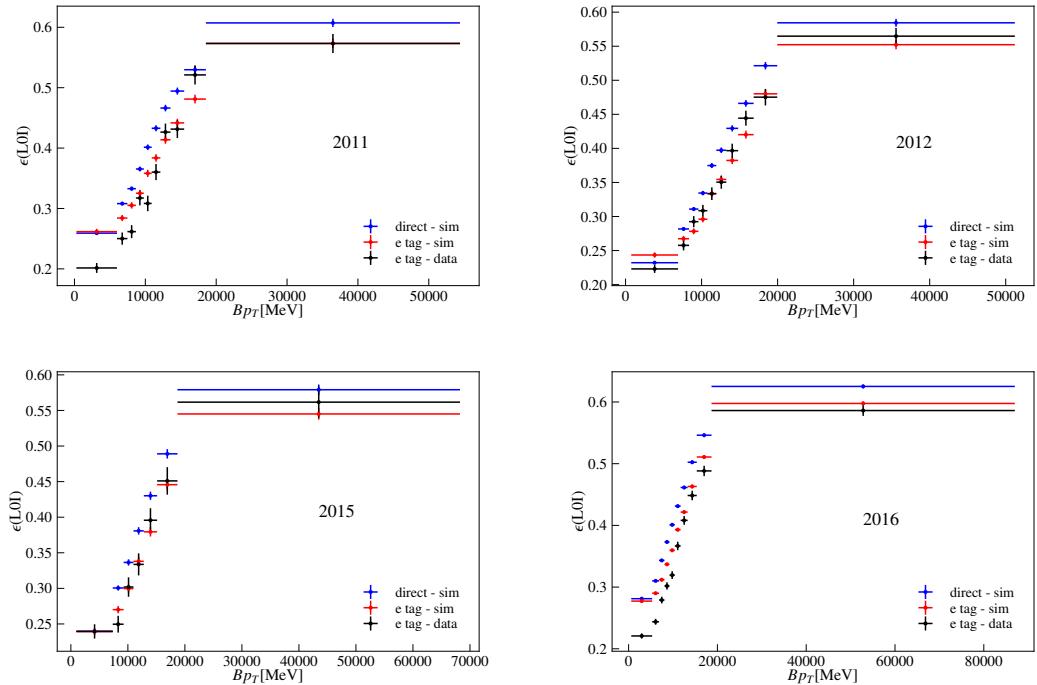
566
567 The probability that an event triggers independently of the specific decay of interest is
568 studied as a function of the transverse momentum of the reconstructed B^0 candidate. A
569 dependency on this variable is expected since the momentum of the reconstructed B^0 is
570 correlated with the transverse momentum of the hadronisation products of the opposite
571 b -quark, that influences the TIS efficiency. The efficiency is measured on events selected
572 in two ways:

- 573 • Simulated sample selected with the preselection and PID criteria applied (see Secs. 4.1
574 and 4.2). This efficiency is referred to as “*direct efficiency*” and is the quantity that
575 needs to be corrected and that the tag-and-probe method tries to approximate. This
576 quantity is only available for the simulated sample and is only used as reference.
- 577 • Both data and simulated samples retained with the preselection, PID and HLT TOS
578 requirements. The tag selection corresponds to events that have been triggered by
579 one of the two electrons in the final state. For this reason is referred to as “*e tag*”.

580 The efficiencies for data and simulation for the years 2011, 2012, 2015 and 2016 are shown
581 in Fig. 10 and the corresponding ratios between data and simulation are shown in Fig. 11.

582 Two features are of particular interest. First, a noticeable bias is observed (also called *tag*
 583 *bias*) in the tag-and-probe efficiency in the simulation with respect to the true efficiency,
 584 as can be observed in Fig. 10. The second one is the size of the correction weights that, at
 585 low $B^0 p_T$, are quite displaced from unity.

586 The size of both these effects can be anticipated to be not negligible and their impact
 on the measurement is studied in detail in Sec. 10.



587 Figure 10: Data and MC efficiencies for the L0I for $B^0 \rightarrow K^{0*} J/\psi(\rightarrow e^+e^-)$ for the years 2011,
 588 2012, 2015, 2016 as a function of $B^0 p_T$.

589

590 [ii] Single electron TOS calibration histogram

591

592 The probability that an electron triggers an event is parametrised as a function of the
 593 region of the ECAL (inner, middle or outer) and of the transverse energy deposited in
 594 the calorimeter. Since the ECAL cannot distinguish between electrons and positrons, the
 595 efficiency is computed separately for the two leptons and then averaged. The efficiency is
 measured on events selected in two ways

- 596
- 597 • The efficiency is obtained for simulated samples where the full preselection and PID
 598 is applied. Due to the fact that no tag selection is applied, this efficiency is referred

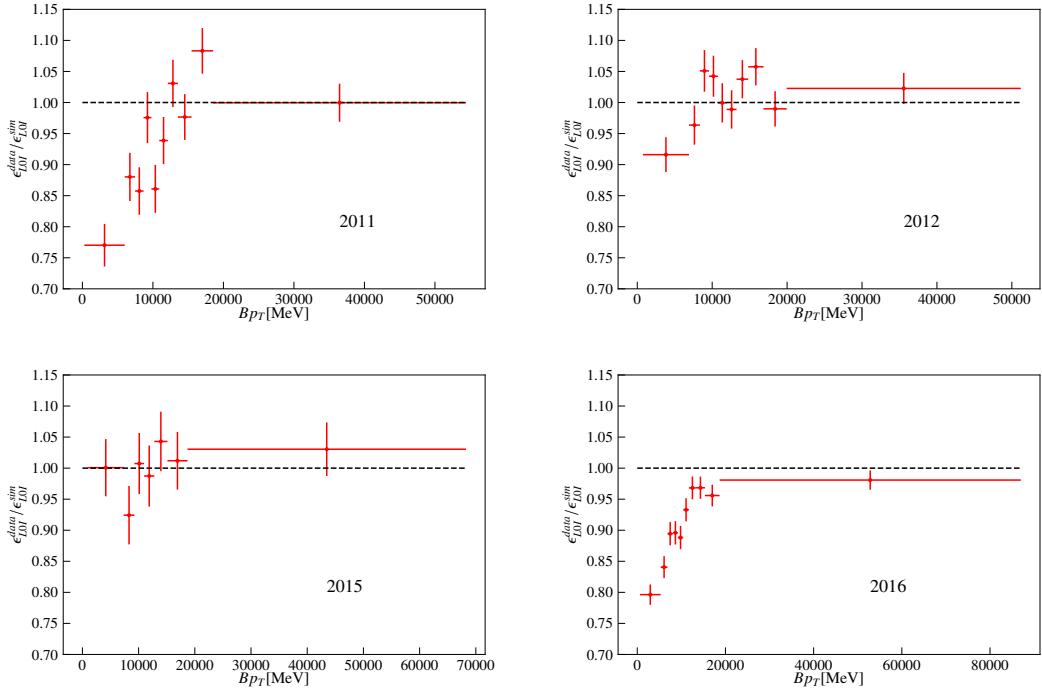


Figure 11: Data/MC ratios for the L0I tag-and-probe efficiency for $B^0 \rightarrow K^{0*} J/\psi (\rightarrow e^+e^-)$ for the years 2011, 2012, 2015, 2016 as a function of $B^0 p_T$.

597 to as “*direct*” efficiency and corresponds exactly to the efficiency that we try to
598 approximate with the tag-and-probe method.

- 599
- 600 • The efficiency is obtained for data and simulated samples where the full preselection,
601 PID and HLT are applied. The tag selection corresponds to events that have been
602 triggered independently of the signal B^0 due to the presence of a muon or an hadron.
For this reason in referred to as “*TIS tag*”.

603 The efficiencies for data and simulation for each year in Run-I and II are gather respectively
604 in Figs. 12 and 13, while the corresponding ratios between data and simulation are shown
605 in Figs. 14 and 15. Unlike the TIS calibration histograms, the “*TIS tag*” seems to allow
606 for an estimation of the efficiency with a negligible tag bias. The ratios between data and
607 simulation efficiencies are quite flat and close to unity for all years of data taking and
608 regions of ECAL when the values of E_T that are above 5000 MeV. Below this however,
609 due to a poorly simulated ECAL response, the ratio is sensibly displaced from unity.

610 **L0 correction weights**

611 Using the aforementioned calibration histograms it is possible to assign event-by-event
612 weights to the simulation samples depending on the classification of the L0 trigger response,

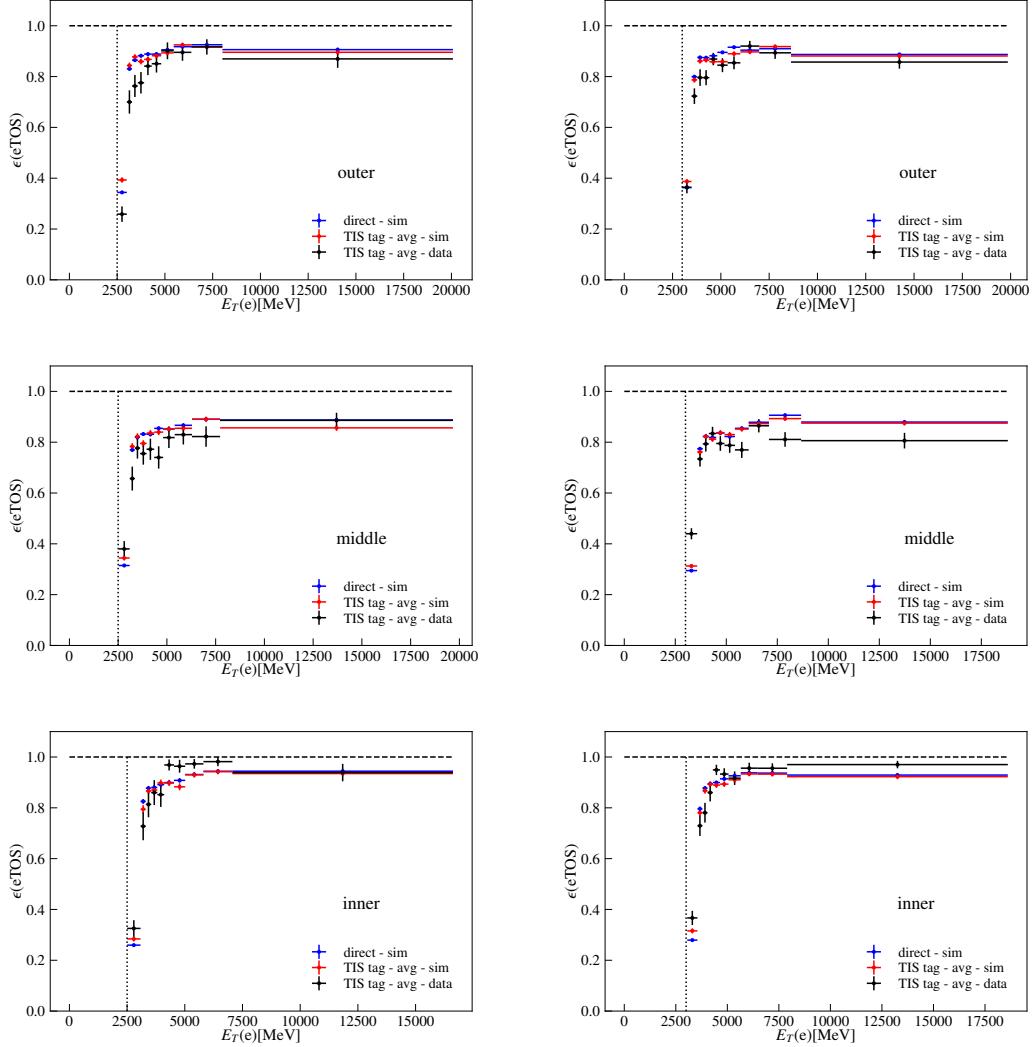


Figure 12: Data and MC efficiencies for (left) 2011 and (right) 2012 $B^0 \rightarrow K^{0*} J/\psi (\rightarrow e^+ e^-)$ decays as a function of the region of detection in the ECAL.

614 *i.e.* w_{L0}^{LOI} and w_{L0}^{LOE} , as discussed in the following

615 **[LOI weights]**

616 The LOI events, *i.e.* events triggered by particles not in the signal candidate, in the
617 simulation are corrected with the weight

$$w_{L0}^{LOI} = \frac{\epsilon_{data}^{TIS}(B^0 p_T)}{\epsilon_{sim}^{TIS}(B^0 p_T)} \quad (11)$$

618 **[LOE weights]**

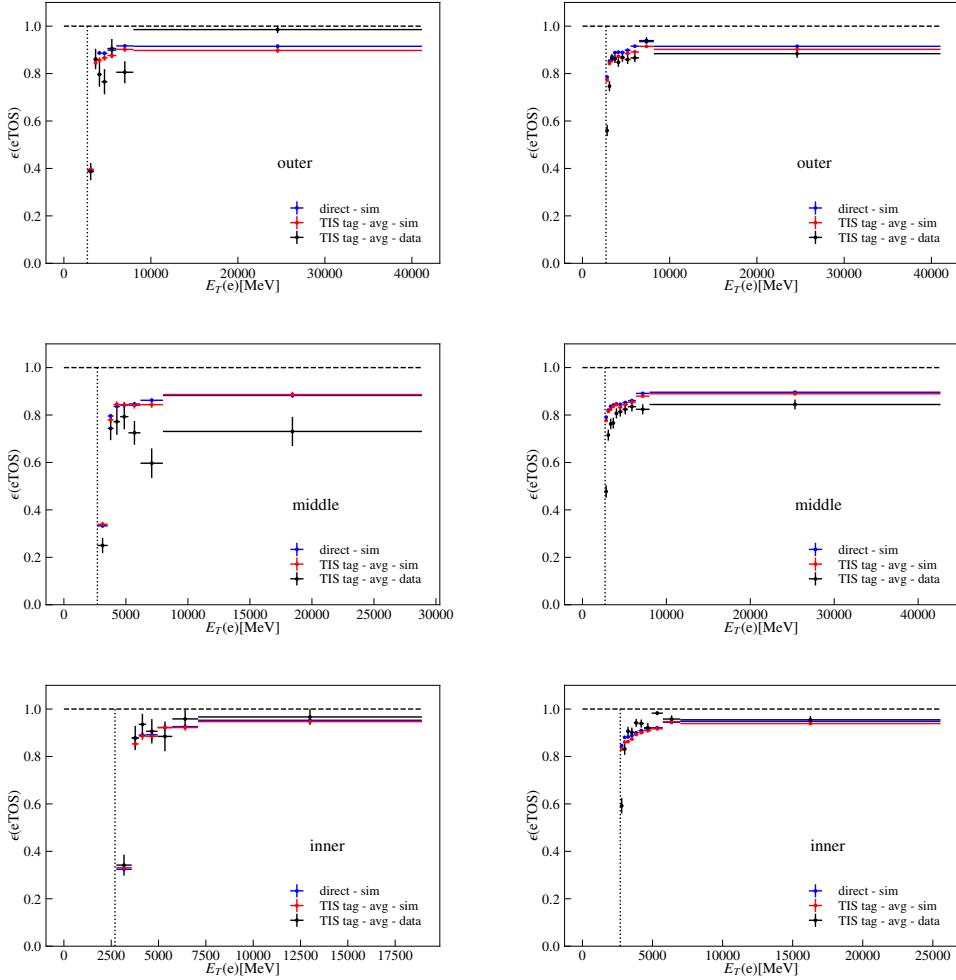


Figure 13: Data and MC efficiencies for (left) 2015 and (right) 2016 $B^0 \rightarrow K^{0*} J/\psi (\rightarrow e^+ e^-)$ decays as a function of the region of detection in the ECAL.

619 The L0E events, *i.e.* at least one of the electrons in the signal candidate and not
 620 triggered by particles that are not in the signal candidate, in the simulation are
 621 weighted by

$$w_{L0}^{L0E} = \frac{(1 - \epsilon_{data}^{TIS}(B^0 p_T))}{(1 - \epsilon_{sim}^{TIS}(B^0 p_T))} \cdot \frac{(1 - (1 - \epsilon_{data}^{eTOS}(e^+)) \cdot (1 - \epsilon_{data}^{eTOS}(e^-)))}{(1 - (1 - \epsilon_{sim}^{eTOS}(e^+)) \cdot (1 - \epsilon_{sim}^{eTOS}(e^-)))} \quad (12)$$

622 Notice that the factorisation of the weights, used here to build w_{L0}^{L0E} , is only exact
 623 if the trigger efficiencies of the two electrons are independent of each other and
 624 independent of the L0I selection.

625 HLT trigger corrections

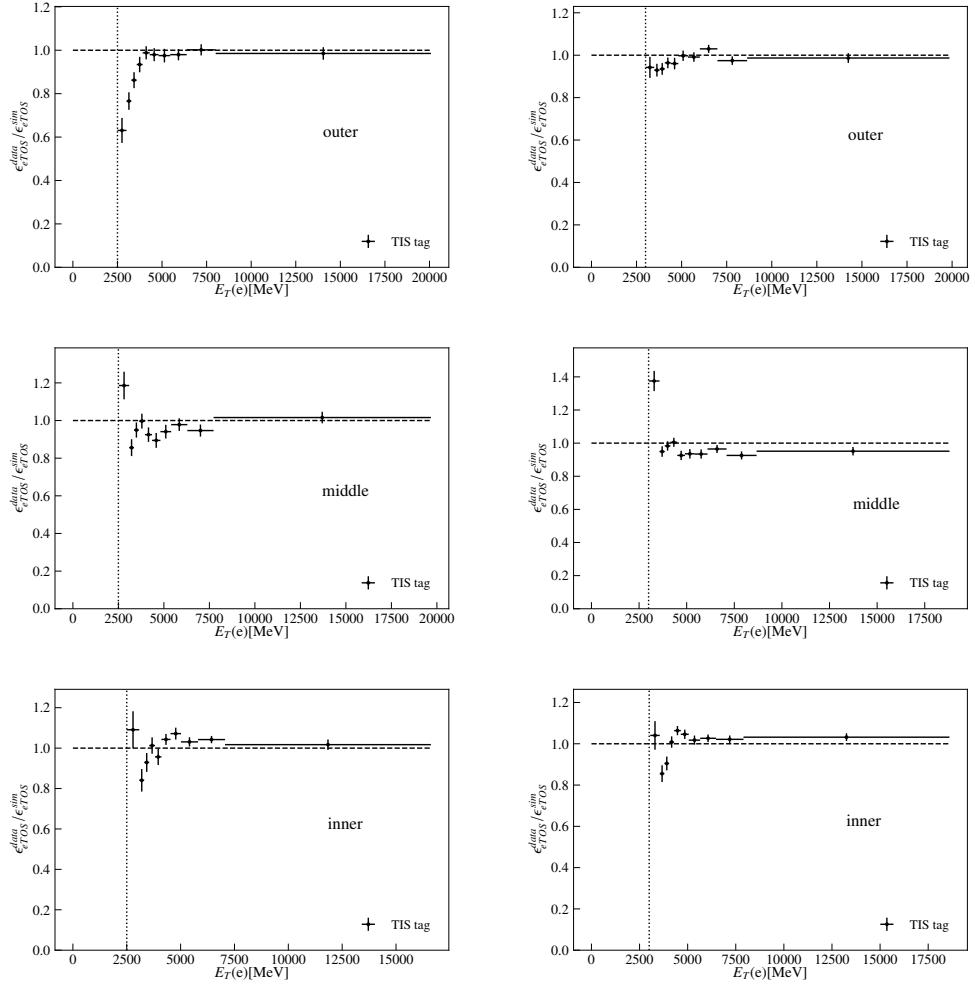


Figure 14: Data/MC tag-and-probe efficiency ratio for (left) 2011 and (right) 2012 $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+e^-)$ decays as a function of the region of detection in the ECAL.

626

627 The HLT efficiencies are calibrated separately for each year of data taking and for the
 628 two L0 trigger categories L0I and L0E, using the resonant mode $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+e^-)$
 629 after the full offline selection chain with the exclusion of the HLT trigger selections and
 630 the MVA selection. Due to a non-negligible contribution of combinatorial background
 631 under the signal, the calibration histograms for ϵ_{data} are obtained by performing separate
 632 fits to the *probe* and *tag* data samples as a function of the transverse momentum of B^0 .
 633 The efficiency histograms ϵ_{data} are then defined by the ratio of the extracted yields, with
 634 a binning scheme chosen to have a similar number of events in each bin. On the other
 635 hand, ϵ_{sim} is evaluated in the same bins of $B^0 p_T$ by looking at the ratio of the sum of
 636 (weighted) events between the *probe* and *tag* simulated samples. The tag selection used

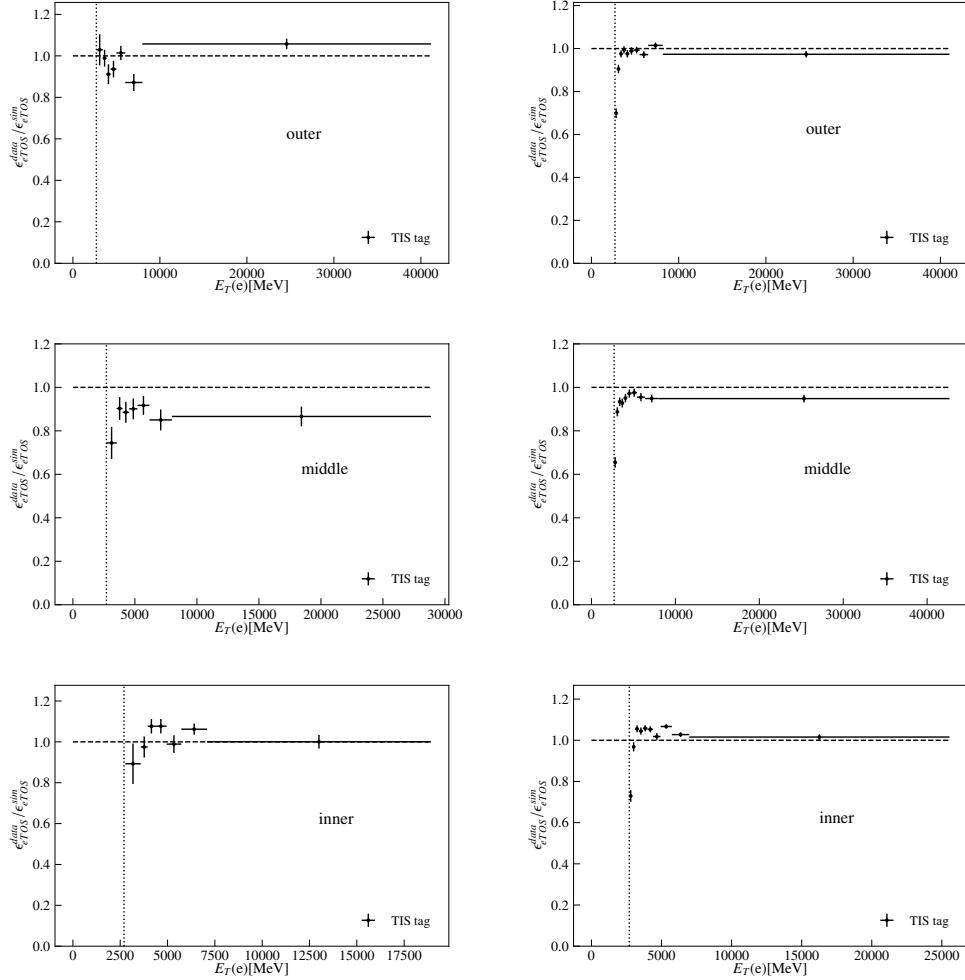


Figure 15: Data/MC tag-and-probe efficiency ratio for (left) 2015 and (right) 2016 $B^0 \rightarrow K^{0*} J/\psi (\rightarrow e^+ e^-)$ decays as a function of the region of detection in the ECAL.

637 corresponds to HLTPHYS for data and HLTOR for simulated events, as defined in Table 8.
 638 This choice is a compromise between using HLTOR selection in data, that would greatly
 639 reduce the statistics available, and using HLT{1,2}PHYS in simulation, that would include
 640 a relevant fraction of HLT lines that are ignored in data.

641
 642 The ratios between data and simulation are shown in Fig. 16 for the L0 inclusive
 643 categories, *i.e.* B_L0Global_TIS and LOElectron_TOS. Due to the limited statistics in the
 644 L0E sample, *i.e.* events triggered by at least one of the electrons in the signal candidate
 645 and not triggered by particles that are not in the signal candidate, these corrections are
 646 derived from the inclusive L0E line; a good agreement between the calibration histograms
 647 for both selections is observed (see Appendix C) which further justifies this choice.

Table 8: Tag selections used for the HLT correction for data (HLTPHYS) and simulated events (HLTOR). The HLT1OR and HLT2OR are defined as the logical OR of the listed single lines.

Label	Run-I	2015/16
HLT1PHYS	HLT1_PHYS_TIS	
HLT2PHYS	HLT2_PHYS_TIS	
HLTPHYS	HLT1_PHYS_TIS & HLT2_PHYS_TIS	
HLT1OR	B_Hlt1TrackAllL0Decision_TIS B_Hlt1TrackMuonDecision_TIS	B_Hlt1TrackMVADecision_TIS B_Hlt1TwoTrackMVADecision_TIS B_Hlt1TrackMuonDecision_TIS
HLT2OR	B_Hlt2Topo2BodyBBDTDecision_TIS B_Hlt2Topo3BodyBBDTDecision_TIS B_Hlt2Topo4BodyBBDTDecision_TIS B_Hlt2TopoMu2BodyBBDTDecision_TIS B_Hlt2TopoMu3BodyBBDTDecision_TIS B_Hlt2TopoMu4BodyBBDTDecision_TIS B_Hlt2TopoE2BodyBBDTDecision_TIS B_Hlt2TopoE3BodyBBDTDecision_TIS B_Hlt2TopoE4BodyBBDTDecision_TIS	B_Hlt2Topo2BodyDecision_TIS B_Hlt2Topo3BodyDecision_TIS B_Hlt2Topo4BodyDecision_TIS B_Hlt2TopoE2BodyDecision_TIS B_Hlt2TopoE3BodyDecision_TIS B_Hlt2TopoE4BodyDecision_TIS
HTLOR	HLT1OR & HLT2OR	

648 HLT correction weights

649

650 The LOI events in the simulation are corrected with the weight

$$w_{HLT}^{LOI} = \frac{\epsilon_{LOI,data}^{HLT}(B^0 p_T)}{\epsilon_{LOI,sim}^{HLT}(B^0 p_T)}, \quad (13)$$

651 while the LOE events are corrected with the weight

$$w_{HLT}^{LOE} = \frac{\epsilon_{LOE,data}^{HLT}(B^0 p_T)}{\epsilon_{LOE,sim}^{HLT}(B^0 p_T)}. \quad (14)$$

652 Similarly to the previous stages of the correction chain, it is possible to investigate the
 653 impact of the trigger correction on the angular distributions. Figures 17 and 18 show the
 654 effect of the trigger corrections for Run1 and 2 conditions separated respectively between
 655 LOE and LOI for the various angular distributions. Note that a mild effect is seen in the
 656 boundary of the $\cos \theta_K$ distribution in the case of the LOI category. Further details on this
 657 pattern is given in Appendix C. Two-dimensional projections are also gathered in Fig. 19
 658 and 20, where no evident correlation between the angles for this correction is seen.

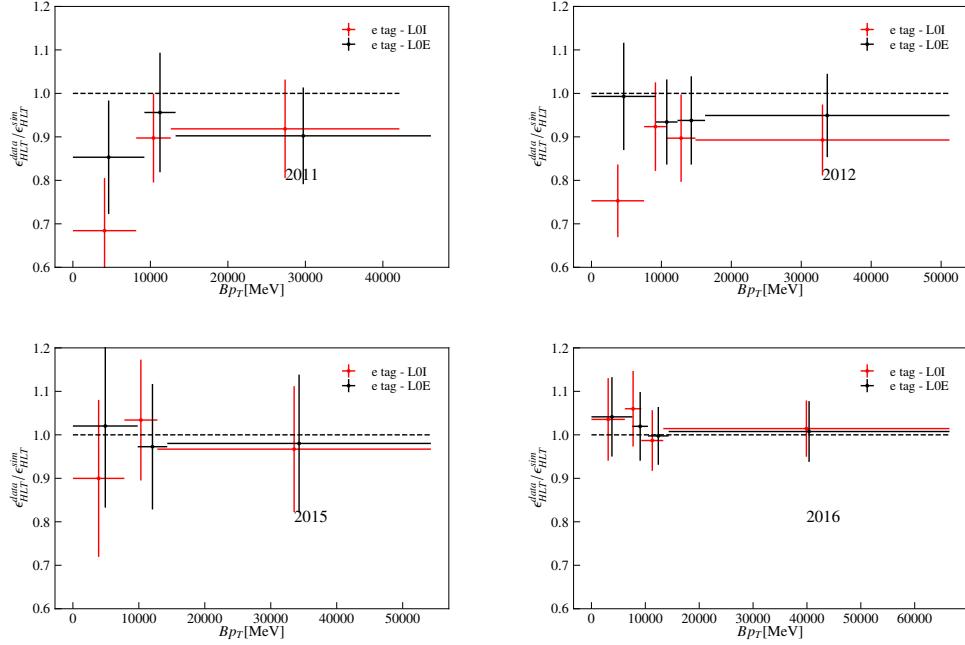


Figure 16: Data/MC ratio of the HLT tag-and-probe efficiency for $B^0 \rightarrow K^{0*} J/\psi (\rightarrow e^+e^-)$ as a function of the p_T of the B^0 for the trigger categories LOI and LOE.

659 3.6.5 Total corrections

660 While the impact of the simulation/data correction chain is discussed in the context of the
 661 acceptance parametrisation in Sec. 6, it is useful to anticipate possible non-uniformities
 662 across the phase space. Figures 21, 22, 23 and 24 illustrate the effect of all corrections
 663 combined with respect to uncorrected simulations. Despite the significant difference on
 664 the absolute efficiencies, there is in general a mild impact on the shape of the angular
 665 distributions, which in turn suggests a limited impact on the parameters of interest.

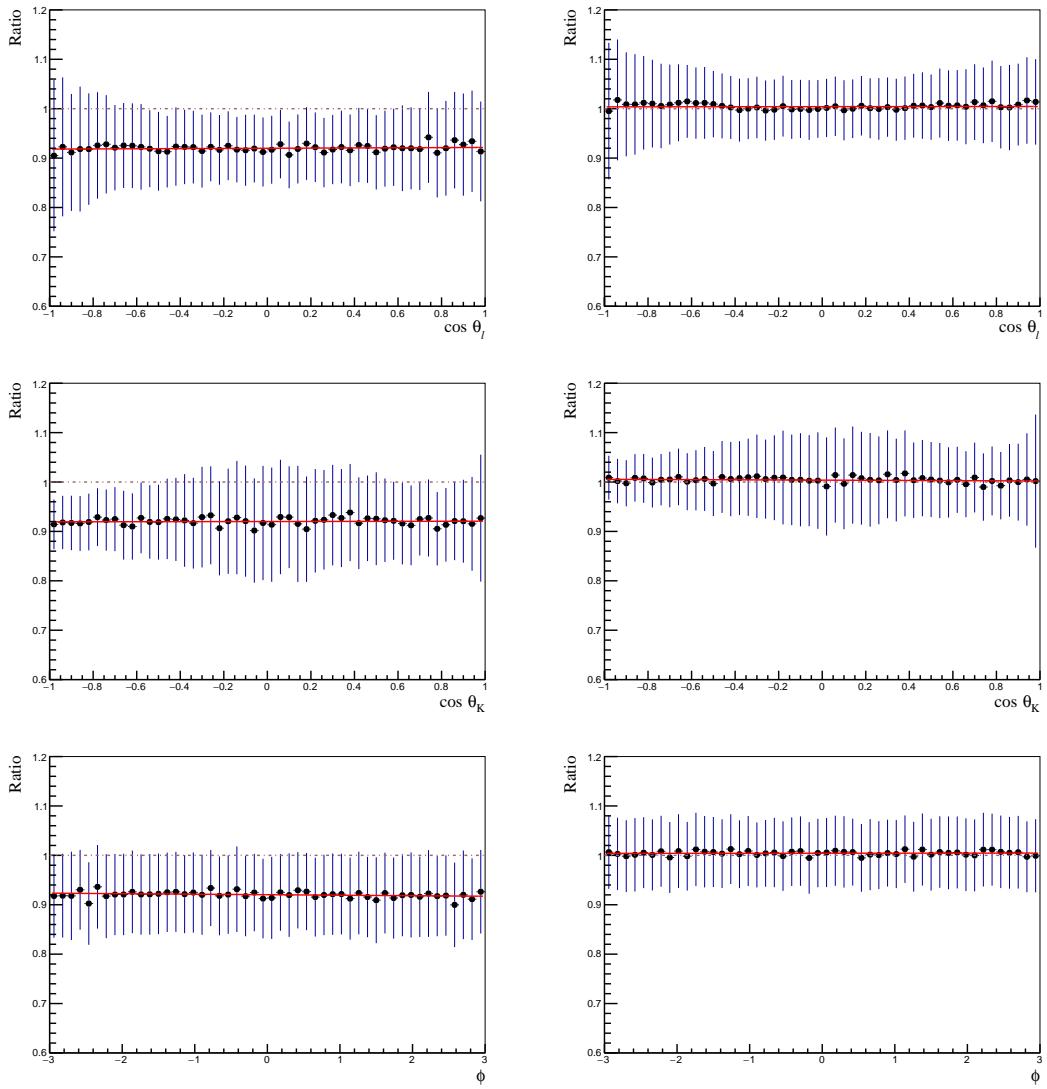


Figure 17: Ratio of the angular distributions between the trigger corrected simulation and uncorrected ensembles for the truth-matched signal channel for the LOE decision with (left) RunI and (right) II conditions.

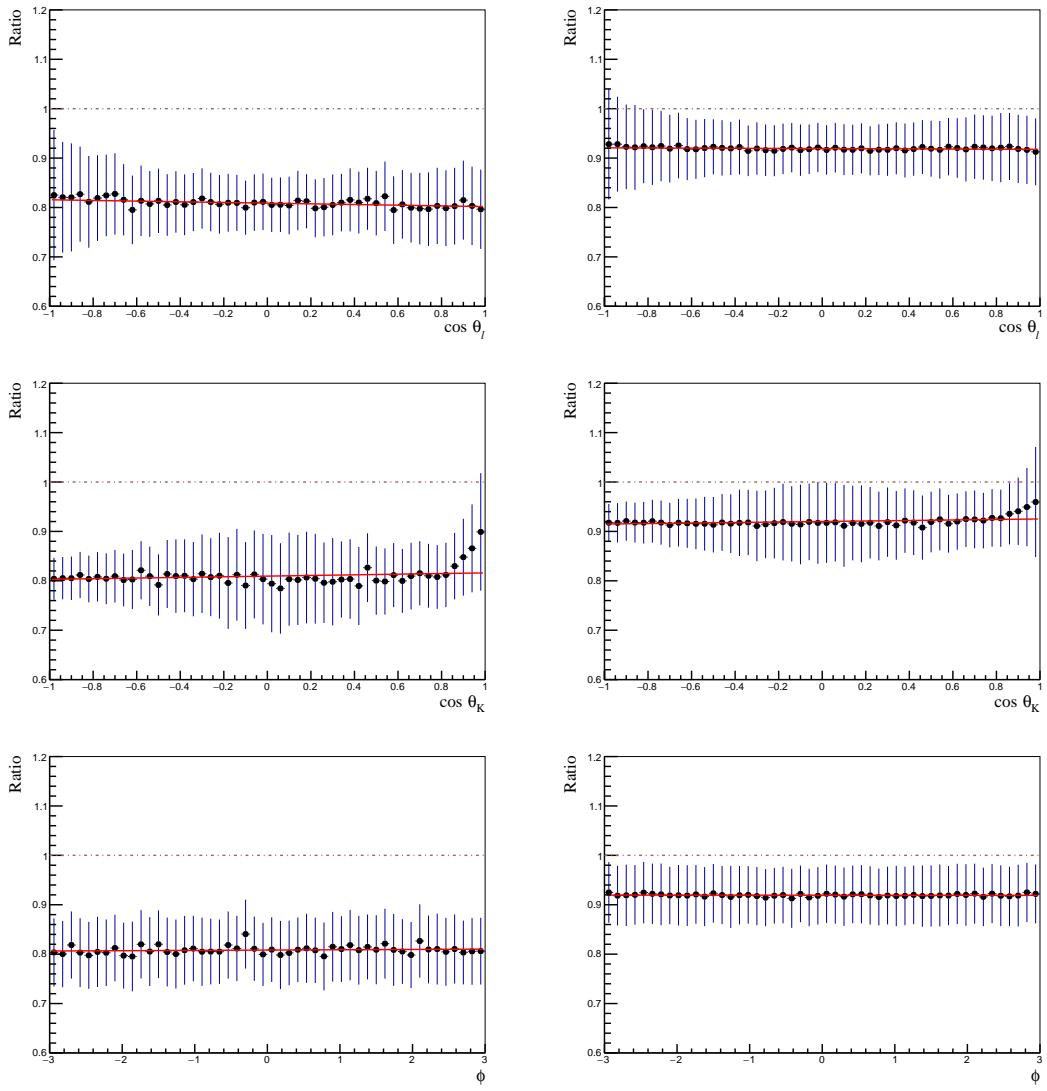


Figure 18: Ratio of the angular distributions between the trigger corrected simulation and uncorrected ensembles for the truth-matched signal channel for the LOTIS decision with (left) RunI and (right) II conditions.

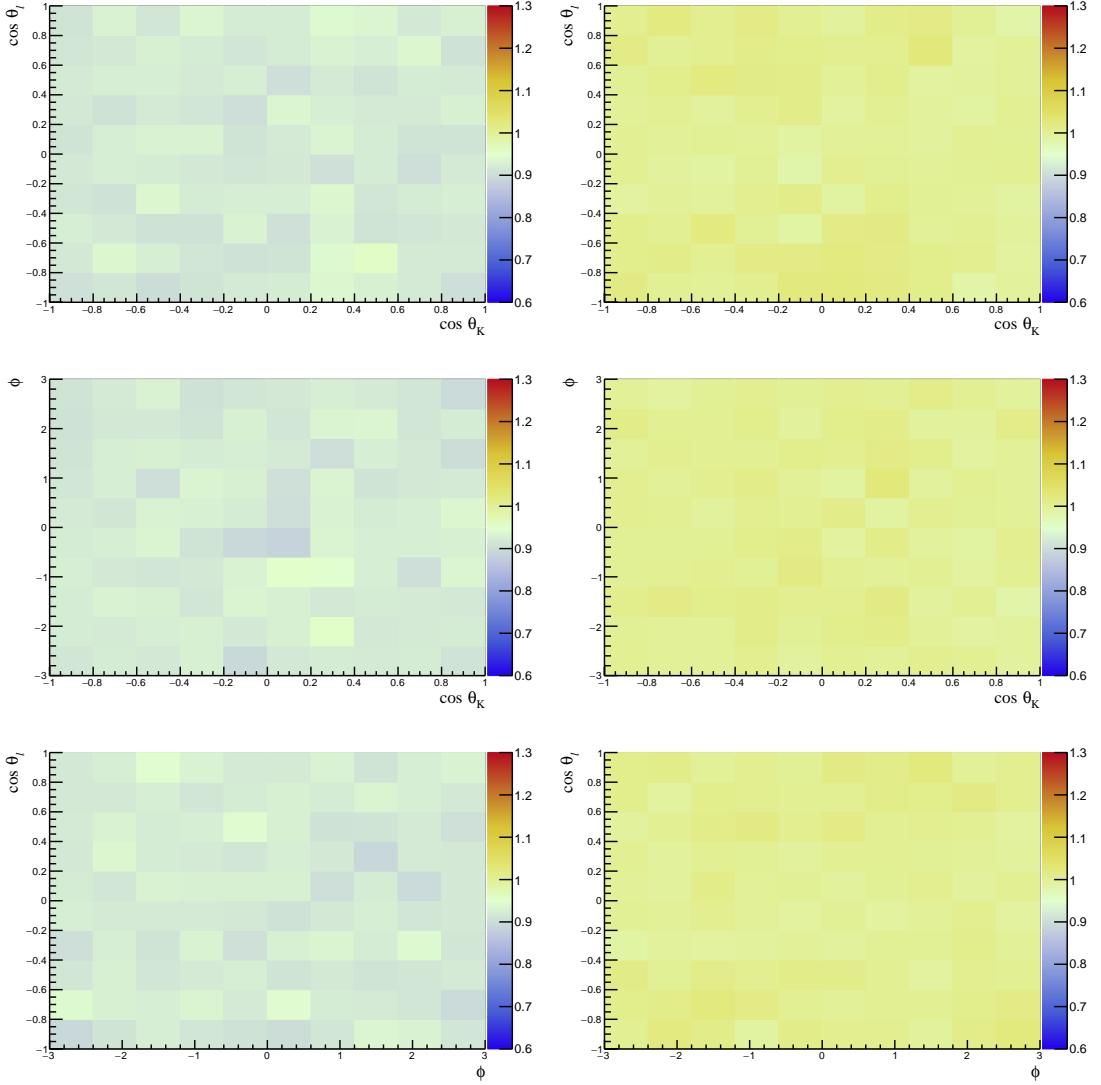


Figure 19: Two-dimensional ratio of the angular distributions between the reconstruction corrected simulation and uncorrected ensembles for the truth-matched signal channel for the LOE decision with (left) RunI and (right) II conditions.

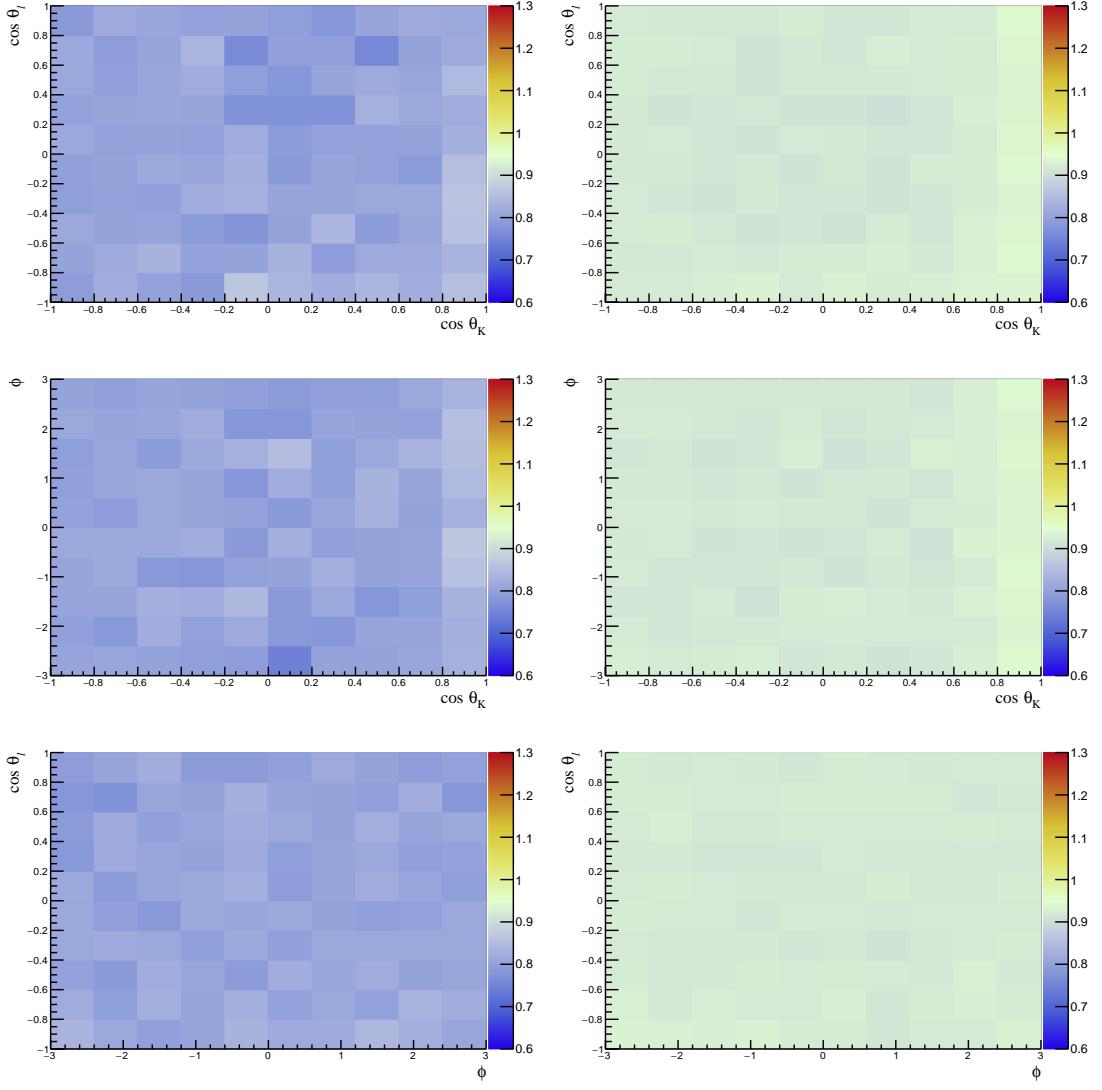


Figure 20: Two-dimensional ratio of the angular distributions between the reconstruction corrected simulation and uncorrected ensembles for the truth-matched signal channel for the LOTIS decision with (left) RunI and (right) II conditions.

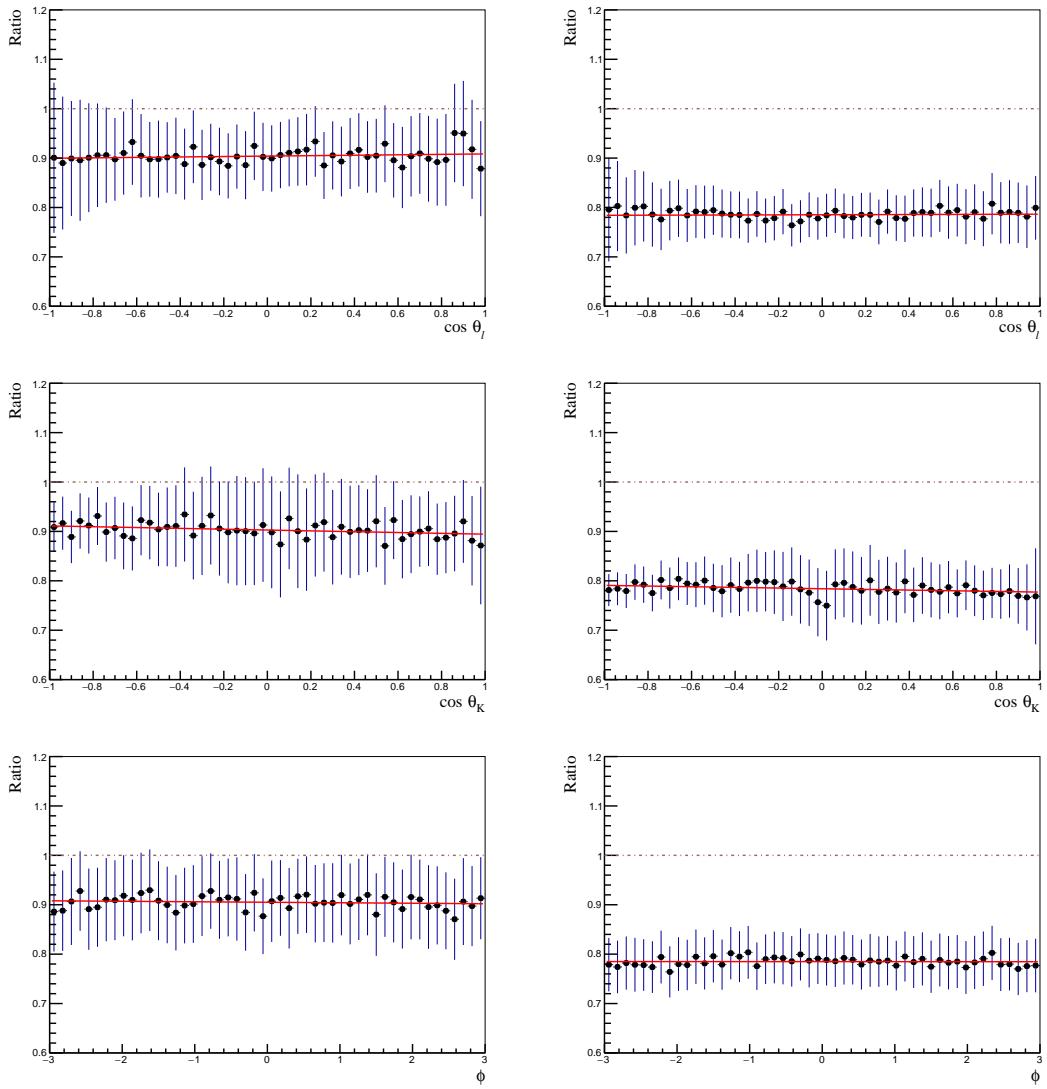


Figure 21: Ratio of the angular distributions between the all corrections to simulation and uncorrected ensembles for the truth-matched signal channel for the LOE decision with (left) RunI and (right) II conditions.

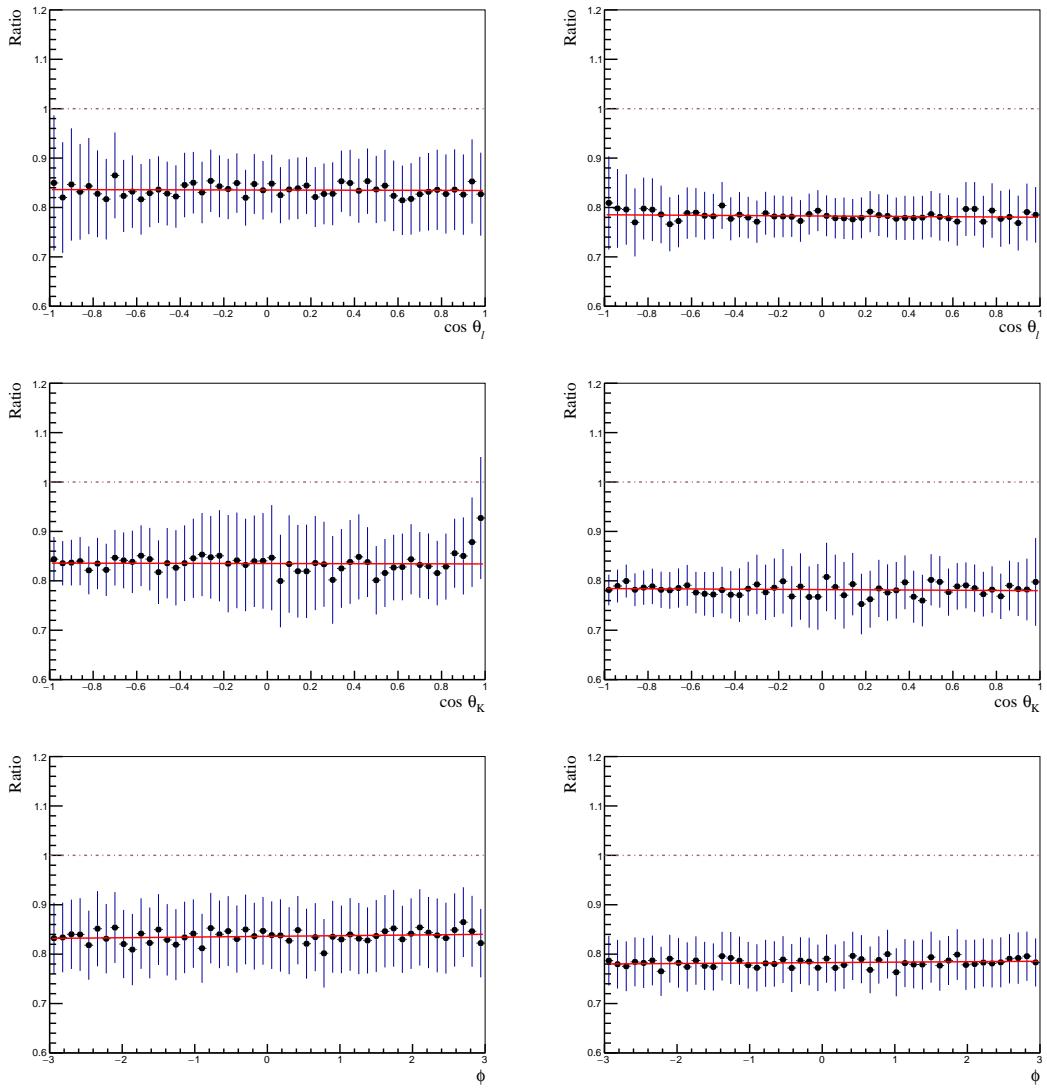


Figure 22: Ratio of the angular distributions between the all corrections to simulation and uncorrected ensembles for the truth-matched signal channel for the LOTIS decision with (left) RunI and (right) II conditions.

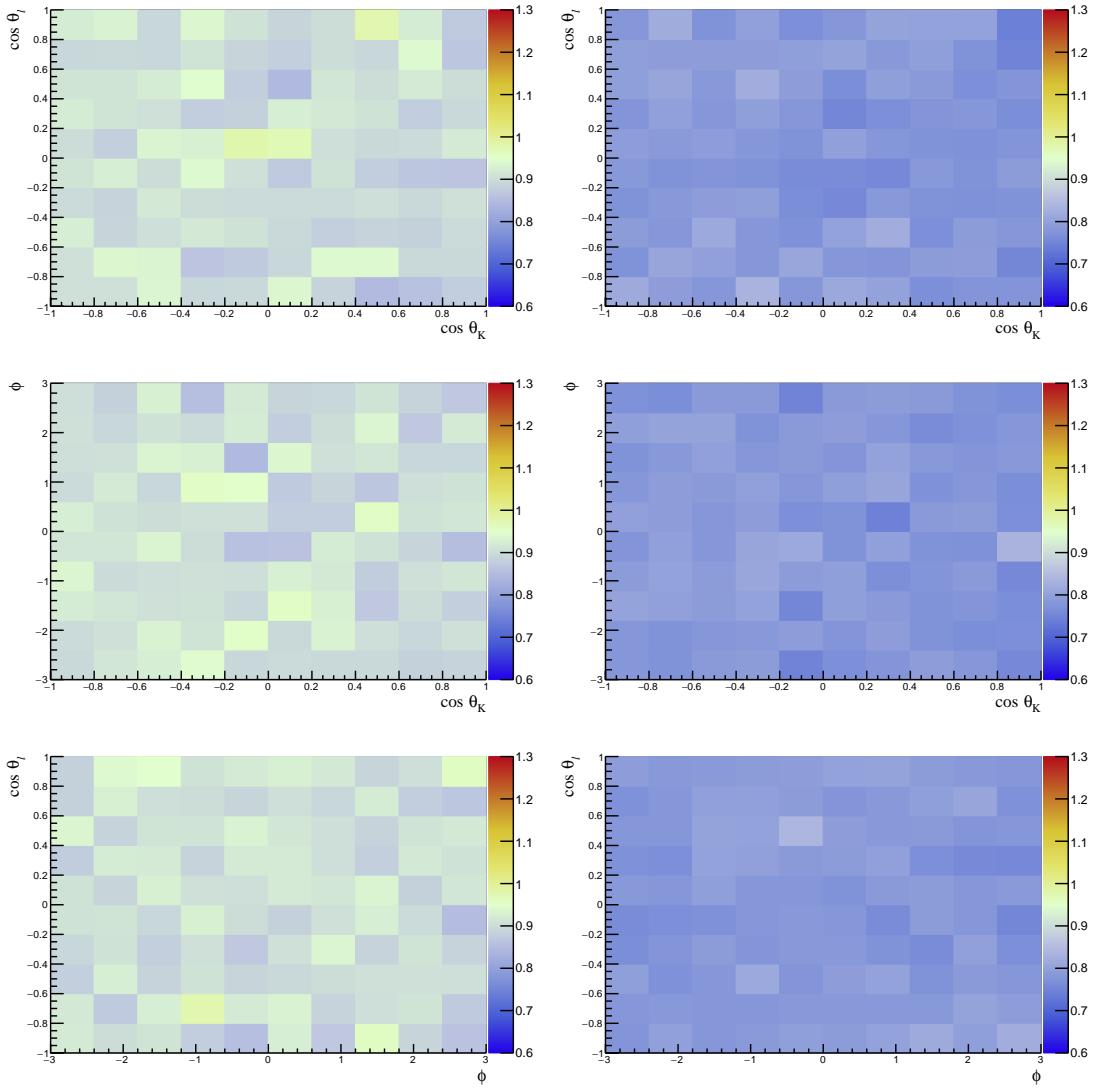


Figure 23: Two-dimensional ratio of the angular distributions between all corrections to simulation and uncorrected ensembles for the truth-matched signal channel for the LOE decision with (left) RunI and (right) II conditions.

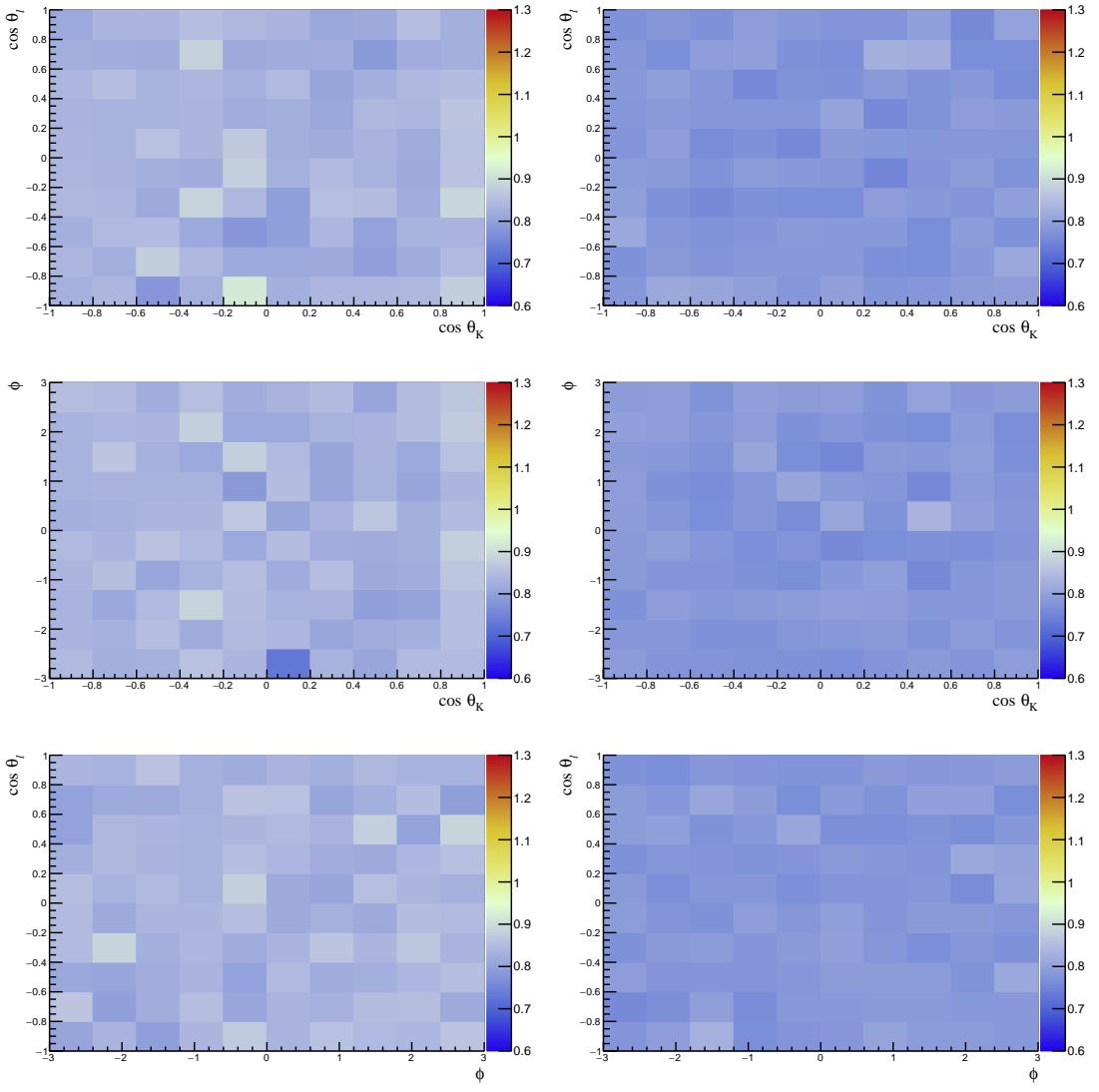


Figure 24: Two-dimensional ratio of the angular distributions between all corrections to simulation and uncorrected ensembles for the truth-matched signal channel for the **LOTIS** decision with (left) **RunI** and (right) **II** conditions.

666 **4 Selection**

The scripts used in the background studies, selection design and optimisation, and MVA training and application can be found in:

- 667 ○ `ewp-bd2ksteeangular-central-q2/scripts/selection/`

668
669 The sensitivity to the angular observables of interest are strongly associated to the ability
670 to discriminate between signal and background. In this section, the many components
671 involved in the offline selection of the rare mode are discussed. For the benefit of the
672 reader, a schematic view of the selection flow is given in the following

- 673 ● Candidates accepted by the Stripping `Bu2LLKeeLine2` lines and that have fired the
674 trigger decisions reported in Sec. 3.3 are considered in this process;
- 675 ● The surviving events pass through a simple set of pre-selection requirements to
676 ensure the quality of the data and also further reduce the level of combinatorial
677 background. This includes high-signal efficiency sanity cuts as well as particle
678 identification requirements;
- 679 ● In the following various categories of background contributions are examined, and
680 specific sources of peaking backgrounds are vetoed;
- 681 ● At this stage, a multivariate (MVA) discriminator is used to further reduce the level
682 of combinatorial background. This procedure has been fully revised from Ref. [52],
683 considering alternative approaches and features;
- 684 ● The choice of the optimal MVA cut value is made for the two L0 categories through
685 an optimisation procedure that uses as a figure-of-merit the sensitivity on P'_5 as
686 a probe. This approach takes into account simultaneously multiple factors that
687 affects the final sensitivity. Both combinatorial, and the partially reconstructed
688 backgrounds are considered in this step.

689 The selection requirements applied to identify the rare candidates and the corresponding
690 charmonium decays are gathered in Tab. 9 and discussed in the following.

Table 9: Summary of the offline selection requirements discussed in this section.

Type		Requirement
Quality	all tracks	$\chi^2/\text{ndf} < 3$ $\text{GhostProb} < 0.4$
	e	$\text{region}_{\text{ECAL}}^{\text{Local}} \geq 0$ $!(\text{xProjection}_{\text{ECAL}}^{\text{Local}} < 363.6 \text{ mm}$ $\& \text{yProjection}_{\text{ECAL}}^{\text{Local}} < 282.6 \text{ mm})$ InAccEcal $\sqrt{p_{x,\text{track}} + p_{y,\text{track}}} > 200$
Clone	K, π, e	$\theta(\pi, e) > 0.0005, \theta(K, e) > 0.0005, \theta(K, \pi) > 0.0005$
ID	K^{*0}	$ m(K\pi) - m_{K^{*0}}^{\text{PDG}} < 100 \text{ MeV}/c^2$
PID	all	hasRich
	e	hasCalo
	K, π	$p_T > 250 \text{ MeV}/c$
	e	$p_T > 500 \text{ MeV}/c, p > 3000 \text{ MeV}/c$
	K	$\text{ProbNNk} \cdot (1 - \text{ProbNNp}) > 0.05$
	π	$\text{V2ProbNNpi} \cdot (1 - \text{ProbNNk}) \cdot (1 - \text{ProbNNp}) > 0.1$
	e	$\text{ProbNNe} > 0.2$
	K	$\text{DLL}_{K\pi} > 0$
BKG	e	$\text{DLL}_{e\pi} > 2$
	charmonium $e - h$ swap $B^+ \rightarrow K^+ e^+ e^-$ $B_s^0 \rightarrow \phi e^+ e^-$ $B^0 \rightarrow \bar{D}^0 (\rightarrow K^+ \pi^-) \pi^- e^+ \nu_e$ $B^0 \rightarrow D^- (\rightarrow K^{*0} (\rightarrow K^+ \pi^-) \pi^-) e^+ \nu_e$ K - π swap part-reco (J/ψ only)	$!(m_{J\psi, \psi(2S)\text{cons.}}(K_{\rightarrow e}\pi e_{\rightarrow K}e) - m^{\text{PDG}}(B^0) < 60 \& \text{eProbNNe} < 0.8)$ $\ m_{J\psi, \psi(2S)\text{cons.}}(K\pi_{\rightarrow e}ee_{\rightarrow \pi}) - m^{\text{PDG}}(B^0) < 60 \& \text{eProbNNe} < 0.8)$ $\max(m(Ke^+e^-), m(\pi_{\rightarrow K}e^+e^-)) < 5100 \text{ MeV}/c^2$ $m(K(\pi_{\rightarrow K})) > 1040 \text{ MeV}/c^2 \& \pi \text{ProbNN} \pi < 0.8$ $!(m(Ke_{\rightarrow \pi}) - m^{\text{PDG}}(D^0) < 30 \text{ MeV} \& \text{eProbNNe} < 0.8)$ $!(m(K\pi e_{\rightarrow \pi}) - m^{\text{PDG}}(D^-) < 30 \text{ MeV} \& \text{eProbNNe} < 0.8)$ $K \text{ DLL}_{K\pi} - \pi \text{ DLL}_{K\pi} > 0$ $m(K\pi e^+e^-)_{J/\psi} > 5150 \text{ MeV}/c^2$
	comb	$\text{MVA} > 0.99$

691 **4.1 Preselection**

692 A B^0 candidate is formed from a pair of well-reconstructed oppositely-charged particles
693 identified as electrons and a pair of well-reconstructed oppositely-charged particles, one
694 identified as a kaon and the other identified as a pion. The $K^+\pi^-$ invariant mass is
695 required to be within $100 \text{ MeV}/c^2$ of the known K^{*0} mass. Particles in the final state are
696 required to have ghost probability smaller than 0.4, that is, low probability of originating
697 from fake tracks, and track χ^2/ndf smaller than 3, which allows to select good quality
698 tracks. The kaon, pion, and electron candidates are required to have associated hits in the
699 RICH detectors. Electrons are also required to have associated clusters in the calorimeters.

700 It has been shown in Ref. [52] that in the very centre of the ECAL calorimeter the cells
701 have not been properly read-out. This results in some discrepancies between the data and
702 simulated events. Therefore, in addition to require that electrons (hadrons) to be in the
703 ECAL (HCAL) acceptance, a fiducial cut is introduced to remove the region defined as
704 $|\text{xProjection}_{\text{ECAL}}^{\text{Local}^0}| < 363.6 \text{ mm} \& |\text{yProjection}_{\text{ECAL}}^{\text{Local}^0}| < 282.6 \text{ mm}$. This requirement
705 rejects around 1% of offline selected signal events.

706 The particle identification criteria for this analysis relies on the neural network-based
707 **ProbNN** approach. The **ProbNN** parameters range between 0 and 1 and can be treated as
708 probabilities. A cut on the combination, *i.e.* $\text{ProbNN}_\pi \times (1 - \text{ProbNN}_K)$, requires that the
709 track satisfies a given threshold of being identified as a pion and not as a kaon. Hereafter
710 such criteria are referred to as $\text{ProbNN}_{hh'}$, where $h^{(\prime)}$ stands for the particle hypothesis
711 under consideration.⁶ Two different tunes, V2 and V3, are used for RunI samples, as the
712 first works better for hadrons, and the second for leptons. A single tune, V1 is used for both
713 leptons and hadrons in RunII. Similarly to Ref. [52], in order to maximise the rejection
714 power of the PID requirements the following requirements are applied

- 715 • for kaon candidates, $\text{ProbNN}_K \cdot (1 - \text{ProbNN}_p) > 0.05$;
- 716 • for pion candidates, $\text{ProbNN}_\pi \cdot (1 - \text{ProbNN}_K) \cdot (1 - \text{ProbNN}_p) > 0.1$;
- 717 • for electron candidates, $\text{ProbNN}_e > 0.2$;

718 A minimum p_T of $250 \text{ MeV}/c$ is required for kaons and pions, while a threshold of $500 \text{ MeV}/c$
719 is preferred for electrons. The electrons are also required to have p greater than $3000 \text{ MeV}/c$.
720 These requirements guarantee a good coverage of the PIDCalib calibration samples.

721 A pair of tracks is defined as clones when they share at least 70% of the hits in the
722 VELO and part of the T-stations [69]. These are reduced by the Clone Killer algorithm,
723 which is ran at the end of the track reconstruction procedure. Nevertheless, some of these
724 background events remain. This is of particular concern for decays containing electrons in
725 the final state, as track quality is reduced by bremsstrahlung effects, making the creation
726 of clone tracks more likely. An efficient strategy to avoid these contribution is to remove
727 events that have angles between final state particles close to zero. The cut against clone

⁶ It is also possible to impose cuts on ProbNN_h alone, without multiplying by $(1 - \text{ProbNN}_{h'})$. However, the combination is found to be effective in this and several other analyses.

728 tracks requires the minimum angle between the kaon or pion and any of the two electrons
729 to be at least 0.5 mrad, namely

$$\theta(\pi, e) > 0.0005 \text{ \& } \theta(K, e) > 0.0005 \text{ \& } \theta(K, \pi) > 0.0005 .$$

730 This requirement is highly efficient for the signal, retaining more than 99 % of selected
731 signal candidates from simulation, and it is effective in vetoing candidates containing clone
732 tracks, which show up as narrow peaks at angles near zero as shown in Figure 25.

733 4.2 Exclusive backgrounds

734 In addition to the requirements described in the previous section, specific requirements
735 are applied to reduce the contamination from the exclusive decays, *e.g.* $B_s^0 \rightarrow \phi e^+ e^-$,
736 $B^+ \rightarrow K^+ e^+ e^-$, $\Lambda_b^0 \rightarrow p K J/\psi (\rightarrow e^+ e^-)$, $\Lambda_b^0 \rightarrow p K e^+ e^-$, and semileptonic cascade decays,
737 such as $B^0 \rightarrow \bar{D}^0 (\rightarrow K^+ \pi^-) \pi^- e^+ \nu_e$ and $B^0 \rightarrow D^- (\rightarrow K^{*0} (\rightarrow K^+ \pi^-) \pi^-) e^+ \nu_e$. Further
738 requirements are applied to remove backgrounds originating from the swapping of a kaon
739 with a pion, and contributions from decays containing J/ψ (or $\psi(2S)$) with electron–hadron
740 swap that escapes the usual q_c^2 criteria.

741 4.2.1 $B_s^0 \rightarrow \phi e^+ e^-$ background

742 If a kaon from the $\phi \rightarrow K^+ K^-$ decay is misidentified as pion, the ϕ state of a $B_s^0 \rightarrow \phi e^+ e^-$
743 decay can be reconstructed as K^{*0} and end up in the selected K^{*0} invariant mass window.
744 This source of background is reduced by requiring the invariant mass of the $K\pi$ pair
745 system, calculated assigning the kaon mass hypothesis to the pion, to be larger than
746 1040 MeV/ c^2 , and the probability of the pion being a pion is low, *i.e.*

$$m(K\pi_{\rightarrow K}) > 1040 \text{ MeV}/c^2 \text{ \& } \pi \text{ProbNN}\pi < 0.8.$$

747 4.2.2 $B^+ \rightarrow K^+ e^+ e^-$ background

748 The $B^+ \rightarrow K^+ e^+ e^-$ decay can be a source of background if a soft pion from the rest of the
749 event is associated to the kaon or if the kaon is misidentified as pion and a soft kaon from
750 the rest of the event is associated to it. In both cases, a fake K^{*0} candidate is reconstructed.
751 The invariant mass of such B^0 candidates peaks at values above 5380 MeV/ c^2 and hence
752 contaminates the upper B^0 invariant mass sideband. The contamination from this decay
753 is reduced by requiring the maximum value between the three-body invariant masses
754 $m(e^+ e^- K)$ and $m(e^+ e^- \pi_{\rightarrow K})$ to be smaller than 5100 MeV/ c^2 , that is

$$\max(m(e^+ e^- K), m(e^+ e^- \pi_{\rightarrow K})) < 5100 \text{ MeV}/c^2.$$

755 4.2.3 $\Lambda_b^0 \rightarrow p K J/\psi (\rightarrow e^+ e^-)$ and $\Lambda_b^0 \rightarrow p K e^+ e^-$ backgrounds

756 The $\Lambda_b^0 \rightarrow p K J/\psi (\rightarrow e^+ e^-)$ and $\Lambda_b^0 \rightarrow p K e^+ e^-$ decays can be a source of background if the
757 proton is misidentified as pion or if there is a double misidentification in which the proton

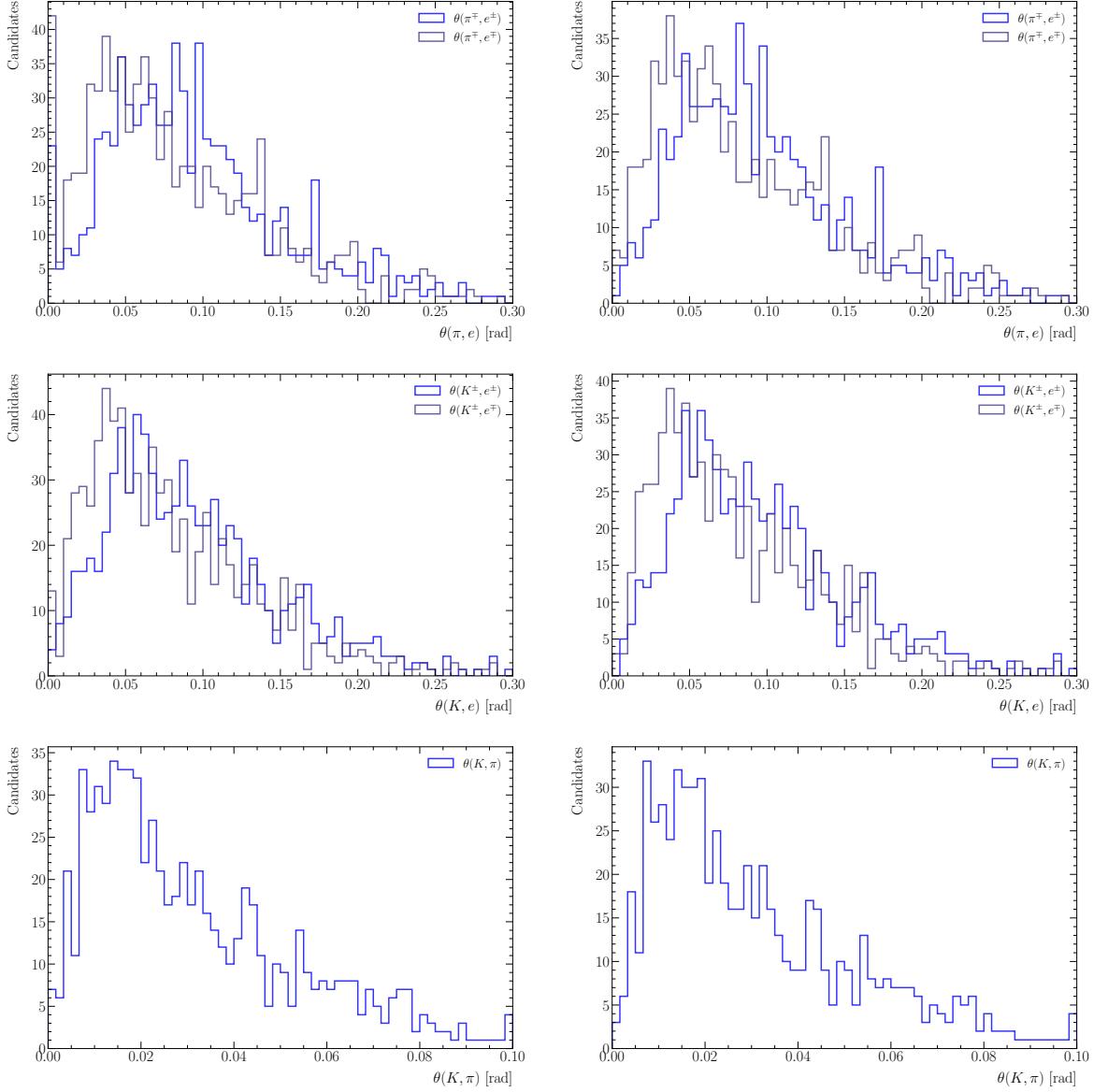


Figure 25: Distributions of the angles between the tracks of the pions and that of the electrons of signal candidates that passed all preselection except the clone veto (top left), and with the inclusion of the veto (top right), showing the removal of the pathological peak. The middle and bottom rows show the angles between the track of the kaon and electrons (middle), and pion and kaon (bottom) with (right) and without (left) the clone cut.

758 is misidentified as kaon and the kaon is misidentified as pion. This contribution can be
 759 reduced by discarding B^0 candidates for which:

$$5575 < m_{(K(\pi \rightarrow p)e^+e^-)} < 5665 \text{ MeV}/c^2$$

760 and $\text{DLL}_{p\pi}$ of the pion is larger than zero (that is, the pion is likely to be a proton), and
 761 candidates for which:

$$5575 < m_{((K \rightarrow p)(\pi \rightarrow K)e^+e^-)} < 5665 \text{ MeV}/c^2$$

762 and $\text{DLL}_{K\pi}$ of the pion is larger than zero (that is, the pion is likely to be a kaon). However,
 763 this veto is found to have limited effectiveness when tested using $\Lambda_b^0 \rightarrow pKJ/\psi (\rightarrow e^+e^-)$
 764 phase-space simulation, as it is seen to reject $32 \pm 2\%$ ($24 \pm 3\%$) of this background for
 765 **Run-I** (**Run-II**).

766 The contamination from $\Lambda_b^0 \rightarrow pKJ/\psi (\rightarrow e^+e^-)$ and $\Lambda_b^0 \rightarrow pKe^+e^-$ decays can be
 767 estimated using

$$\mathcal{N}_{\text{data}}^{\Lambda_b^0} = \mathcal{N}_{\text{data}}^{B^0} \frac{f_{\Lambda_b^0}}{f_d} \frac{\mathcal{B}(\Lambda_b^0 \rightarrow pK(J/\psi \rightarrow) e^+e^-)}{\mathcal{B}(B^0 \rightarrow K^{*0}(J/\psi \rightarrow) e^+e^-) \times \mathcal{B}(K^{*0} \rightarrow K^+\pi^-)} \frac{\epsilon_{\text{tot}}^{\Lambda_b^0}}{\epsilon_{\text{tot}}^{B^0}},$$

768 where $\mathcal{N}_{\text{data}}^{B^0}$ is the rare or control mode yield in data, $f_{\Lambda_b^0}/f_d$ is twice $f_{\Lambda_b^0}/(f_u + f_d)$, with
 769 the latter taken from Ref. [70] and ϵ_{tot} is the reconstruction, trigger, and selection efficiency
 770 calculated from simulation. For the control mode $\mathcal{N}_{\text{data}}^{B^0}$ is obtained by fitting the J/ψ
 771 constrained B^0 mass, while for the rare mode it is taken from the blinded data fit made for
 772 toy generation. The branching fractions are taken from PDG [71], with the exceptions of \mathcal{B}
 773 ($\Lambda_b^0 \rightarrow pKe^+e^-$), for which the value of 3.1×10^{-7} from [72] is used, and $\mathcal{B}(B^0 \rightarrow K^{*0}e^+e^-)$,
 774 which is approximated to 10^{-6} . Using this approach, $\mathcal{N}_{\text{data}}^{\Lambda_b^0}$ for the control mode is found
 775 to be around 83 ± 13 (37 ± 7) events for **Run-I** (**Run-II**) without the veto, and 56 ± 9
 776 (28 ± 5) events with the veto. For the rare mode, contribution from this background is
 777 found 0.6 ± 3.5 events **Run-I** and **Run-II** combined without the veto⁷.

778 The veto against Λ_b^0 decays is included as part of the nominal list of selections despite
 779 its low background rejection power, as it has relatively high signal efficiency for the rare
 780 mode of around 97 to 99%. However, as $\Lambda_b^0 \rightarrow pKJ/\psi (\rightarrow e^+e^-)$ is modelled in the control
 781 mode fit, and the contribution of $\Lambda_b^0 \rightarrow pK \rightarrow e^+e^-$ is found to be negligible for the rare
 782 mode, this veto will be removed in future iterations of this note.

783 4.2.4 Charmonium electron–hadron and hadron–hadron swaps

784 The rejection of events containing electron-hadron swap, such as $B^0 \rightarrow K^{*0}J/\psi (\rightarrow e^+e^-)$
 785 with an electron mis-identified as a kaon or pion, or vice versa, is complicated by limited
 786 resolution. Simple vetoes on $m_{J/\psi, \psi(2S)}(h_{\rightarrow e}\pi e_{\rightarrow h}e)$ tend to lack signal efficiency and/or
 787 background rejection power. One proposed method adopted in the recent update of R_{K^*} is
 788 to make use of $m_{J/\psi, \psi(2S)}(h_{\rightarrow e}\pi e_{\rightarrow h}e)$ mass calculated with $J/\psi/\psi(2S)$ mass constraints –
 789 $m_{J/\psi, \psi(2S)} \text{ cons.}(h_{\rightarrow e}\pi e_{\rightarrow h}e)$. However, currently it is not possible to apply mass constraints
 790 using the DTF to a combination of particles that do not share a common source. Thus
 791 these constrained masses are calculated offline by incorporating the constraint using the
 792 method of Lagrange multipliers in the track kinematic fit.

⁷Due to the lack of 2015 samples, all Λ_b^0 background estimations for **Run-II** are obtained using 2016 samples.

793 The cut used for the rejection of electron-hadron swap requires the exclusion of events
 794 with

$$|m_{J/\psi, \psi(2S) \text{ cons.}}(K_{\rightarrow e} \pi e_{\rightarrow K} e) - m^{PDG}(B^0)| < 60 \text{ \& } \text{eProbNNe} < 0.8$$

795 and

$$|m_{J/\psi, \psi(2S) \text{ cons.}}(K \pi_{\rightarrow e} ee_{\rightarrow \pi}) - m^{PDG}(B^0)| < 60 \text{ \& } \text{eProbNNe} < 0.8$$

796 This requirement has a high signal efficiency of around 98% on simulated, and offline
 797 selected signal events. Using simulated $B^0 \rightarrow K^{*0} J/\psi$ and $B^0 \rightarrow K^{*0} \psi(2S)$ events with
 798 loosened PID cuts to increase statistics (background of this type is already strongly
 799 suppressed by the PID requirements), and truth-matched to obtain only swap events by
 800 requiring that the kaon (pion) is an electron and vice versa, the overall swap rejection for
 801 those two modes is found to be $(63 \pm 7)\%$ ($(76 \pm 6)\%$) for swapped $B^0 \rightarrow K^{*0} J/\psi$ events
 802 of Run-I (Run-II), and $(78 \pm 10)\%$ ($(81 \pm 8)\%$) for swapped $B^0 \rightarrow K^{*0} \psi(2S)$ of Run-I
 803 (Run-II)⁸.

804 The hadron–hadron swaps are vetoed by requiring the $\text{DLL}_{K\pi}$ of the kaon to be larger
 805 than the $\text{DLL}_{K\pi}$ of the pion, that is, by requiring that the kaon candidate is more likely to
 806 be a kaon than the pion candidate. This veto is around 99% efficient for rare mode signal
 807 events, and is estimated from simulation to remove around $(69 \pm 4)\%$ and $(54 \pm 3)\%$ of swap
 808 background for Run-I and Run-II samples, respectively. Using the ratio of efficiencies
 809 from rare mode MC and signal yields taken from the blinded data fit made to obtain yield
 810 estimations for toy generation, namely,

$$N_{\text{swap, selected}}^{\text{est.}} = \frac{\epsilon^{\text{swap}}}{\epsilon^{\text{selected}}} N_{\text{selected}}^{\text{data}},$$

811 contribution from remaining swap events after the application of the veto is expected to
 812 be negligible (0.56 ± 0.06 events expected for the full data sample). Note that it is not
 813 possible to separate out this component using the mass fit alone as they peak under the
 814 signal distribution, thus any remaining swap events in the data fit would be considered
 815 signal. Their incorporation would inflate signal yield, and therefore increase the number
 816 of estimated swap events.

817 4.2.5 $\pi \rightarrow e$ mis-identification

818 Decays such as $B^0 \rightarrow \bar{D}^0(\rightarrow K^+ \pi^-) \pi^- e^+ \nu_e$ and $B^0 \rightarrow D^-(\rightarrow K^{*0}(\rightarrow K^+ \pi^-) \pi^-) e^+ \nu_e$ can
 819 be mistaken for signal with the mis-identification of a pion as an electron. These are vetoed
 820 by requiring that the invariant mass of the kaon (or kaon and pion) and the electron as
 821 pion is not within $30 \text{ MeV}/c^2$ of the D^0 (or D^-) mass, and that the electron in question
 822 does not have low probability of being an electron. That is, it excludes events with

$$|m(K e_{\rightarrow \pi}) - m^{PDG}(D^0)| < 30 \text{ MeV} \text{ \& } \text{eProbNNe} < 0.8$$

⁸The usage of constrained K^{*0} mass obtained through a similar procedure is possible, and is seen to give comparable performance. Rejection based on $m_{K^{*0} \text{ cons.}}$ veto would also be more generic than vetoes based on J/ψ or $\psi(2S)$ masses, as it would also, for example, reduce rare mode electron-hadron swaps, which are, nevertheless, not expected to be a significant source of background after full selection.

823 and

$$|m(K\pi e_{\rightarrow\pi}) - m^{PDG}(D^-)| < 30\text{MeV} \text{ & } \text{eProbNN} < 0.8 .$$

824 These requirements remove less than 1% of the selected signal events. Note that the track
825 momenta variables are used to calculate $m(K\pi e_{\rightarrow\pi})$.

826 4.2.6 Partially reconstructed backgrounds

827 Partially reconstructed decays are decays for which at least one particle in the final state has
828 not been reconstructed by the detector. These decays can contribute to the background if
829 the reconstructed particles in the final state are the same of the signal and if the resulting
830 invariant mass is within the region considered in the mass fit. In this measurement,
831 the dominant contribution to this background originates from $B^+ \rightarrow K^{*(**)} e^+ e^-$ and
832 $B^+ \rightarrow K^{*(**)} J/\psi (\rightarrow e^+ e^-)$ decays, where $K^{*(**)}$ represents a generic kaon resonance
833 heavier than K^{*0} . Typical decays of such resonances consist of a $K\pi$ pair together with
834 one or more pions and can pass the selection requirements described in the previous sections
835 if one of the pions in the final state is not reconstructed. If the decay is reconstructed as
836 $B^0 \rightarrow K^{*0} e^+ e^-$ or $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$, the resulting B^0 invariant mass is smaller than
837 the nominal value and is expected to peak at nearly $5140\text{ MeV}/c^2$ if one pion is missing.

838 For the $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ decay mode, this background contribution can be easily
839 vetoed by requiring the PV- and J/ψ -constrained invariant mass of the B^0 candidates to
840 be larger than $5150\text{ MeV}/c^2$. The residual contamination is estimated from simulated $B^+ \rightarrow K_1^+ J/\psi$ events, with $J/\psi \rightarrow e^+ e^-$ and $K_1^+ \rightarrow K^+ \pi^+ \pi^-$. The ratio f_u/f_d is approximated
841 to unity. The residual contamination is found to be below the percent level and can hence
842 be neglected in the model used in the mass fit.

843 For the $B^0 \rightarrow K^{*0} e^+ e^-$ it is difficult to remove partially reconstructed decays efficiently
844 and effectively, in particular its contribution near, and below the signal mass peak, which
845 is of greater importance to the measurement than the removal of events in the low mass
846 region with little signal. Two possible methods have been investigated, namely the usage
847 of the HOP mass and a dedicated MVA, but both show limited success (see Appendix E
848 for details). Given the difficulties associated with its removal, and the decision to move
849 to a reduced mass window of $4900\text{ MeV}/c^2$ to $5700\text{ MeV}/c^2$, the partially reconstructed
850 background will be modelled from simulation, and included as a component in the fit.

852 4.3 Charmonium contributions

853 Charmonium resonances, such as J/ψ and $\psi(2S)$, peak in q^2 , and thus an optimised choice
854 of the central- q^2 bin is paramount. In particular, one needs to study the expected size of
855 this contribution, since large bremsstrahlung radiation effects are foreseen. The strategy
856 pursued in this analysis, as well as in the low q^2 angular analysis [?], involves introducing a
857 requirement on the constrained q^2 , *i.e.* the q_c^2 . The advantage of this strategy can be seen
858 in Fig. 26. The plots show the residual background due to the leakage of the radiative
859 tail of the J/ψ in the rare mode for different selections of q^2 , separately for each trigger
860 category. By selecting the q_c^2 it is possible to have a larger bin in the central q^2 region,

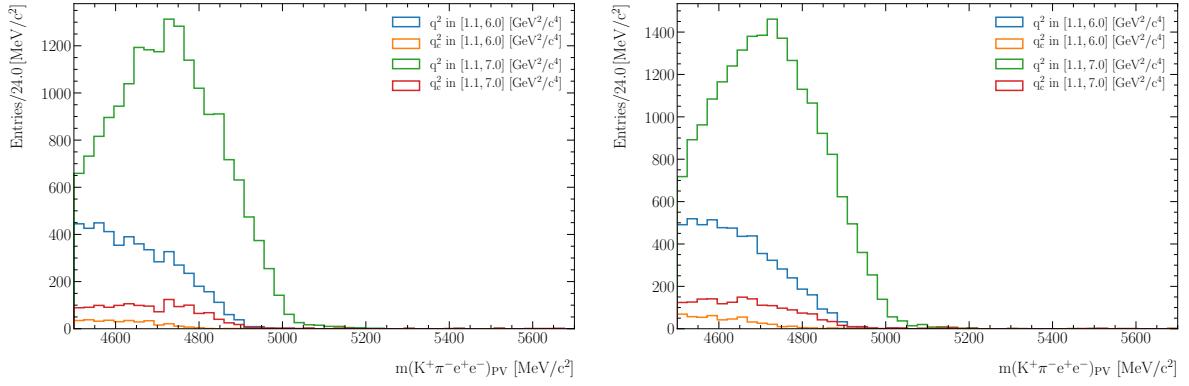


Figure 26: Distribution of the PV-constrained B^0 invariant mass of the background due to $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ candidates leaking in the central q^2 bin for the LOE (left) and LOTIS (right) trigger categories, as obtained from simulation. When the cut on the constrained q^2 is applied, the contribution from this background becomes insignificant above $4900 \text{ MeV}/c^2$.

up to $7.0 \text{ GeV}^2/c^4$ instead of $6.0 \text{ GeV}^2/c^4$, with a smaller background contamination than when considering a maximum q^2 of $6.0 \text{ GeV}^2/c^4$.⁹ This background is further reduced by moving up the lower limit of the B^0 invariant mass window from $4500 \text{ MeV}/c^2$ to $4900 \text{ MeV}/c^2$. The combined effect of the q_c^2 and adjusted B^0 mass window allows for the safe omission of this background component from the final mass and angular fit model. The remaining leakage events from J/ψ are expected to be around 13 ± 2 events for Run-I and Run-II combined, and is treated as a source of systematic uncertainty. As the cut on the q_c^2 is used for the rare mode, the constrained q^2 cut of 7 to $11 \text{ GeV}^2/c^4$ is also adopted for the resonant mode to keep the two samples independent. Note that the contamination from the J/ψ leakage for the upper limit $6.0 \text{ GeV}^2/c^4$ is even more suppressed.

The distribution of the $B^0 \rightarrow K^{*0} e^+ e^-$ and $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ data candidates as a function of PV-constrained B^0 invariant mass $m(K^+ \pi^- e^+ e^-)_{\text{PV}}$ and q^2 is shown in Fig. 27. The analogous distribution as a function of the q_c^2 is also shown in Fig. 28. The two horizontal bands in Fig. 27 correspond to the J/ψ and $\psi(2S)$ resonances, while the diagonal bands correspond to the radiative tail generated by bremsstrahlung emission. Due to the constraint on q^2 , the $B^0 \rightarrow K^{*0} e^+ e^-$ and $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ data candidates appear rotated in Fig. 28.

The usage of the constrained q^2 results in some distortion of the combinatorial background distribution. This effect, while slightly more pronounced for the resonant mode, is generally mild when the PV constrained B^0 invariant mass is used. It should be noted that the imposition of the constraint forces some pathological events, most noticeably combinatorial events that contain real J/ψ , to enter the signal q^2 region at high B^0 mass. The distribution of these events in simulation is shown on Fig. 29 and Fig. 30. In data,

⁹Note that the estimated upper bound from Ref. [52] is of 0.1–0.5% of $B^0 \rightarrow K^{*0} J/\psi (e^+ e^-)$ events to leak into the q^2 interval (depending on the trigger category).

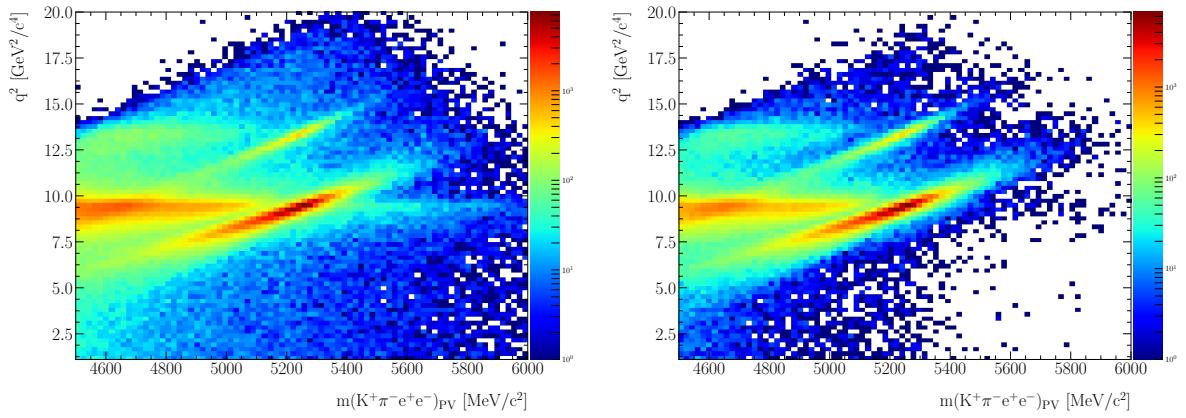


Figure 27: Distribution of the $B^0 \rightarrow K^{*0} e^+ e^-$ and $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ data candidates as a function of PV-constrained B^0 invariant mass $m(K^+ \pi^- e^+ e^-)_{\text{PV}}$ and q^2 , including the full selection (left) without and (right) with the MVA applied.

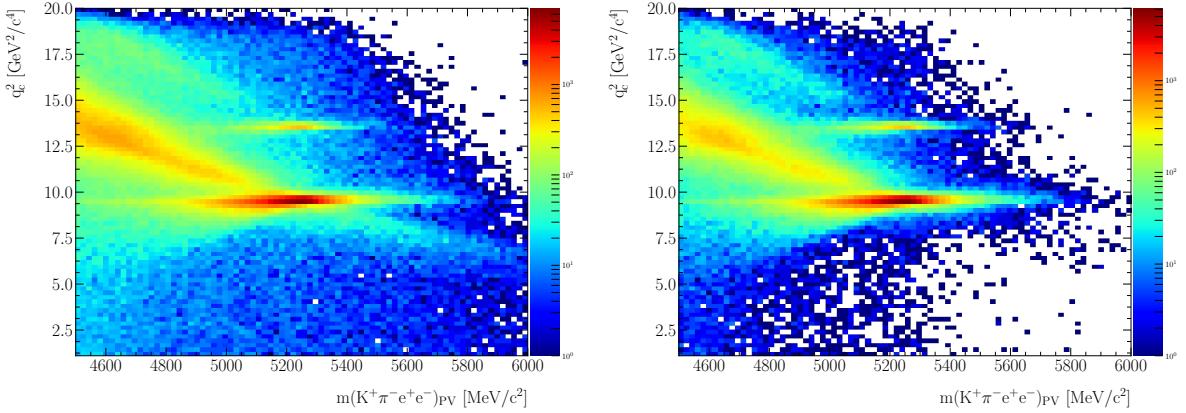


Figure 28: Distribution of the $B^0 \rightarrow K^{*0} e^+ e^-$ and $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ data candidates as a function of PV-constrained B^0 invariant mass $m(K^+ \pi^- e^+ e^-)_{\text{PV}}$ and constrained q^2 , including the full selection (left) without and (right) with the MVA applied.

they are clearly visible in the absense of the MVA cut as the horizontal band centered around the J/ψ q^2 in Fig. 27, which becomes rotated in Fig. 28, leading to part of the tail entering the signal region. The application of the MVA cut reduces this contribution significantly due to its combinatorial nature. The number of remaining events can be estimated by fitting the data in the upper mass side band above 5700 MeV/ c^2 using a exponential function to describe the combinatorial component, and a KDE based on the distribution of events with $\text{BKGAT} > 60$ in simulation. To increase statistics, the fit is performed on the combined Run-I and Run-II data sample with a looser MVA cut of 0.9. All other cuts are unchanged from their nominal settings. The resulting yield of the Jpsi

893 containing combinatorial is then scaled back to that of the nominal fit region and MVA
894 cut using simulation via

$$N_{J/\psi \text{ comb}}^{\text{nominal}} = \frac{\epsilon^{\text{nominal}}}{\epsilon^{\text{UMSB}}} N_{J/\psi \text{ comb}}^{\text{UMSB}}, \quad (15)$$

895 where $N_{J/\psi \text{ comb}}^{\text{nominal}}$ is the estimated yield in the signal region, $N_{J/\psi \text{ comb}}^{\text{UMSB}}$ is the yield from the
896 fit to the upper mass side band, and the two ϵ s give the ratio of the number of background
897 events in the nominal region over those in the upper mass side band with loosened MVA
898 threshold.

899 The result of the fit, shown in Fig. 31, leads to an estimation of 8 ± 5 events in the
900 signal region of the full data sample. The impact of this remaining background is thus
901 expected to be marginal, and will be assessed as a systematic uncertainty.

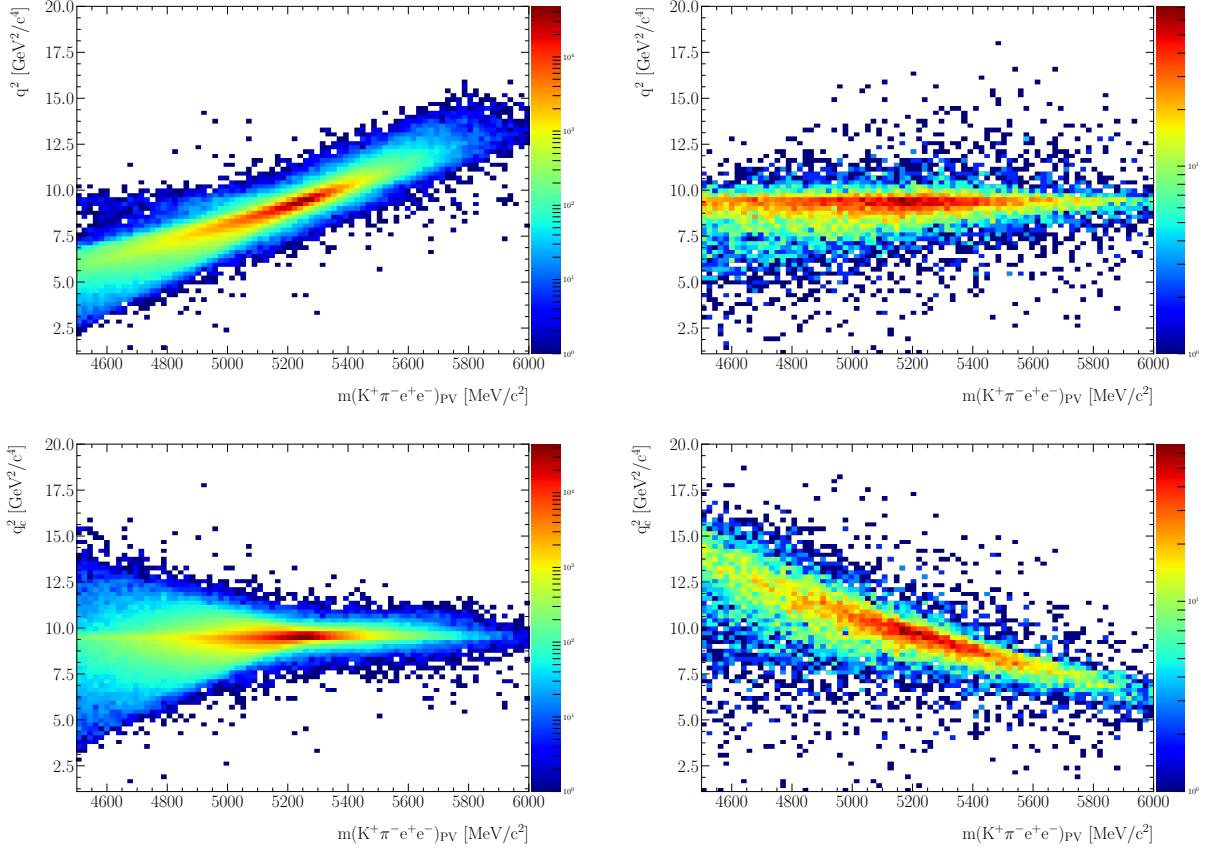


Figure 29: Distribution of $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+e^-)$ simulation candidates as a function of PV-constrained B^0 invariant mass $m(K^+\pi^-e^+e^-)_{\text{PV}}$ and q^2 (unconstrained – top row, constrained – bottom row), with all offline signal selection applied except the MVA cut. The left plots show the distribution of events belonging to background categories of 0, 10 and 50 (signal and signal-like), while the right plots show background categories above 60 (combinatorial like). The application of the constraint shifts the leakage from the resonant mode away from the signal q^2 at low mass, but brings in some J/ψ containing combinatorial events at high mass. Note that this background is significantly reduced by the combinatorial MVA, thus the MVA cut has been removed here to make its contribution visible.

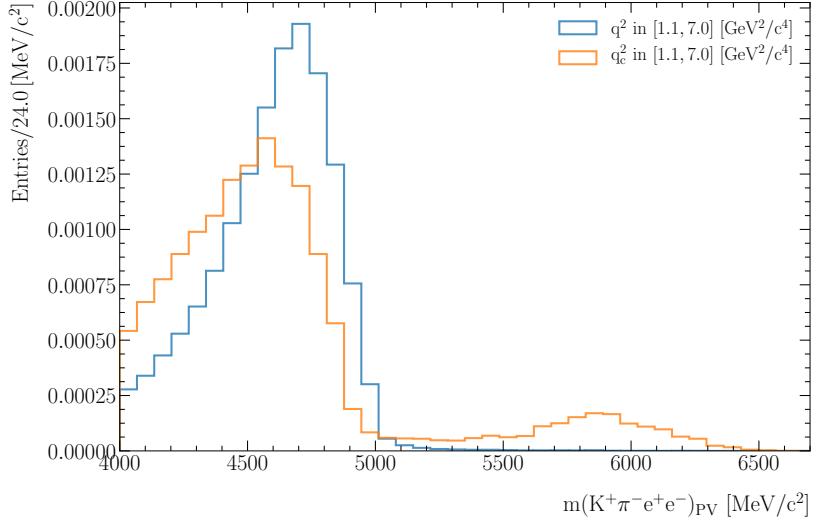


Figure 30: Distribution of $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+e^-)$ simulated candidates as a function of PV-constrained B^0 invariant mass $m(K^+\pi^-e^+e^-)_{\text{PV}}$. The offline signal selection is applied except for the MVA cut. Truth-matching has not been applied, and as a result, a bump is visible at high mass when the constrained q^2 cut is applied. These events have background categories greater than 60 and are combinatorial in nature. Therefore, after the MVA requirements are applied this contribution is reduced to a marginal level.

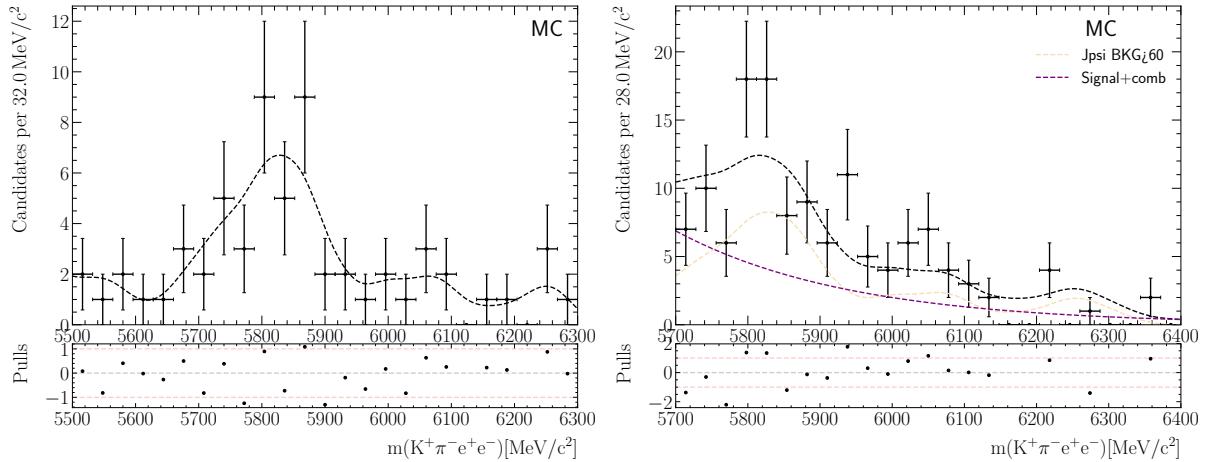


Figure 31: Plot showing the distribution of J/ψ containing combinatorial from simulation described by a KDE (left) and the fit to the upper mass sideband of the full data sample (right) used to estimate the level of this background in the signal region.

Sample	Run-I	Run-II
Signal	41733	106433
Background	3548	7650

Table 10: Table showing the sizes of the signal and background samples used in training.

902 4.4 Multivariate classifier

903 Random combinations of particles from the same pp collision can be a source of background
 904 if they have invariant masses compatible with those of the signal candidates. A large
 905 fraction of this *combinatorial background* is removed by means of a multivariate analysis.
 906 To this purpose, machine learning techniques implemented in the `scikit-learn` python
 907 framework [73] and in the Reproducible Experiment Platform (REP) and High-Energy-
 908 Physics Machine Learning (`hep_ml`) python libraries are used. A multivariate classifier
 909 is first trained and tested on two samples labelled as signal and background, and then
 910 applied to the samples from which the combinatorial background has to be rejected. The
 911 investigated classifier response is set to be between zero and unity, with values close to
 912 zero (unity) indicating a large probability of the reconstructed candidate to be background
 913 (signal). An optimisation procedure has been designed to determine the threshold on the
 914 classifier response that results in the best sensitivity in an angular analysis.

915 4.4.1 Signal and background samples

916 The signal sample used to train and test the multivariate classifier is obtained from
 917 $B^0 \rightarrow K^{*0} e^+ e^-$ simulation with corrections to account for differences between data and
 918 simulation, while B^0 candidates populating the upper sideband of the B^0 invariant mass
 919 distribution in data ($m_{B^0} > 5600 \text{ MeV}/c^2$) are used as background. Preliminary studies
 920 on the choice of MVA hyper-parameter space and input features have been performed in
 921 Ref. [56] using Run-I samples and Sim08 simulation with similar preselection (see Secs. D.1
 922 and D.2 for details). Since a similar performance is observed with the updated samples for
 923 this analysis, the results for the MVA setting and input variables previously found are not
 924 re-optimised.

925 Signal and background candidates must satisfy the same requirements described in
 926 Table 9, apart from some small modifications designed to increase the statistics available
 927 for the training. The ProbNN requirement for the electrons is loosened from greater than
 928 0.2 to greater than 0.05, and a looser set of DLL cuts is applied ($K \text{ DLL}_{K\pi}$ greater than -5
 929 instead of 0 and $e \text{ DLL}_{e\pi}$ greater than 0 instead of 2). The $K\pi$ invariant mass requirement
 930 is loosened to $200 \text{ MeV}/c^2$ around the nominal value of the K^{*0} mass. Furthermore, the
 931 lower bound of the q_c^2 bin is shifted to 0.1 during the training. Finally, the training is split
 932 by run periods, *i.e.* Run-I (2011+2012) and Run-II (2015+2016), combining both trigger
 933 categories. A summary of the available signal and upper sideband statistics can be seen in
 934 Tab. 10 showing that in every case the upper sideband statistics represents the limiting
 935 factor to the training process.

936 **4.4.2 Selection of discriminating variables**

937 As an initial exercise, a large number of features are investigated in the training process
938 and interactively removed. The variables that guarantee the best separation between
939 signal and background are obtained by starting from all possible combinations of at least
940 six of the following variables: p_T , χ_{IP}^2 , flight distance χ^2 , decay vertex χ^2/ndf , DIRA, and
941 χ^2 of the kinematic fit of the B^0 candidate; decay vertex χ^2/ndf of the K^{*0} candidate;
942 decay vertex χ^2/ndf of the dilepton pair; sum of the p_T and sum of the χ_{IP}^2 of the hadrons
943 (electrons) in the final state; minimum and maximum p_T and χ_{IP}^2 of the hadrons (electrons)
944 in the final state; ProbNN variables of the four particles in the final state. The variables
945 related to the B^0 candidates are included in all combinations since they are expected
946 to provide the best discrimination between signal and background. A *gradient boosting*
947 classifier is trained using as input the variables in each combination. The accuracy of the
948 classifier, which corresponds to the fraction of candidates in the testing sample that are
949 labelled correctly as signal or background, is then used to quantify the discriminating
950 power of a given combination. The optimal set of variables to be used in the classification
951 is given by the combination associated to the highest accuracy.

952 In this initial set of features, some considerations can be made. The PID variables
953 do not seem to provide a good discriminating power, and therefore, these are removed
954 from this procedure in the first iteration. The sum of the p_T and the sum of the χ_{IP}^2
955 are highly correlated with the minimum and maximum p_T and χ_{IP}^2 and provide a worse
956 discrimination, so these are also ignored in the following. The results for the remaining
957 variables of this iterative procedure are shown in Fig. 32, which shows the accuracy of each
958 combination as a function of the number of variables in the combination. The accuracy
959 does not have a strong dependence on the number of variables taken into account, though
960 a gradual improvement is visible when increasing the number of variables from six to 14.
961 The accuracy is maximum when 14 variables are used in the training and decreases slightly
962 if additional degrees of freedom have to be taken into account. This is probably due to
963 overfitting, *i.e.*, to the fact that, with the statistics available, the classifier is not able to
964 extract useful information when too many variables are provided as input.

965 For a given number of variables, the best accuracy is reached when including all B^0
966 variables. These variables alone are able to provide almost the same discriminating power
967 as the full set of variables described above. The 13 variables resulting in the maximum
968 accuracy are used to discriminate between signal and background and are gathered in
969 Table 11. These are the six variables related to the B^0 candidate, the decay vertex χ^2/ndf
970 of the K^{*0} candidate, the decay vertex χ^2/ndf of the electron pair, the maximum p_T and
971 the minimum χ_{IP}^2 of the hadrons, and the minimum and maximum p_T and maximum χ_{IP}^2
972 of the electrons. Their correlation matrices, as obtained from the signal and background
973 samples, are shown in Fig. 33. As expected for the signal, the χ^2 of the kinematic fit of the
974 B^0 candidate is highly correlated with the χ_{IP}^2 (63%) and decay vertex χ^2/ndf (72%) of the
975 candidate itself. This does not hold for the background, since the latter consists of random
976 combinations of particles. The K^{*0} and B^0 decay vertex χ^2 are also correlated with each
977 other (56%) in the signal sample, while no correlation appears in the background sample.

Table 11: Summary of the input features used in the MVA training.

Particle	Variables
B^0	$p_T, \chi_{\text{IP_OWNPV}}^2, \chi_{\text{FD_OWNPV}}^2, \chi_{\text{vtx}}^2/\text{ndf}, \chi_{\text{DTF}}^2/\text{ndf}, \text{DIRA}$
$K^{*0}, J/\psi$	$\chi_{\text{vtx}}^2/\text{ndf}$
h	$\min(p_{T,K}, p_{T,\pi}), \min(\chi_{\text{IP_OWNPV},K}^2, \chi_{\text{IP_OWNPV},\pi}^2)$
ℓ	$\min, \max(p_{T,e^+}, p_{T,e^-}), \min, \max(\chi_{\text{IP_OWNPV},e^+}^2, \chi_{\text{IP_OWNPV},e^-}^2)$

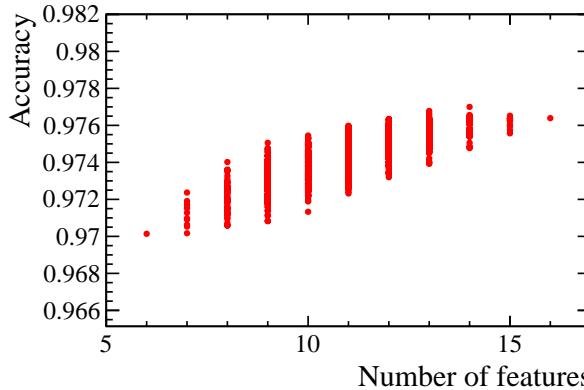


Figure 32: Accuracy of each combination of discriminating variables as a function of the number of variables in the combination.

978 The p_T of the B^0 candidate is highly correlated with the minimum and maximum p_T of
 979 the electrons (56% and 70%, respectively) and, to a smaller extent, to the maximum p_T of
 980 the hadrons (54%). Two other pairs of highly correlated variables are given by the flight
 981 distance χ^2 of the B^0 and the maximum (67% and minimum (57%) impact parameter χ^2
 982 of the electrons.

983 4.4.3 Selection of classifier

984 After choosing the optimal set of variables, several classifiers are optimised and then tested
 985 against each other. The following classifiers are considered:

- 986 ● TMVA;
- 987 ● GradientBoosting;
- 988 ● XGBoost;
- 989 ● DTC (Decay Tree Classifier);
- 990 ● EXC (Extra Trees Classifier).

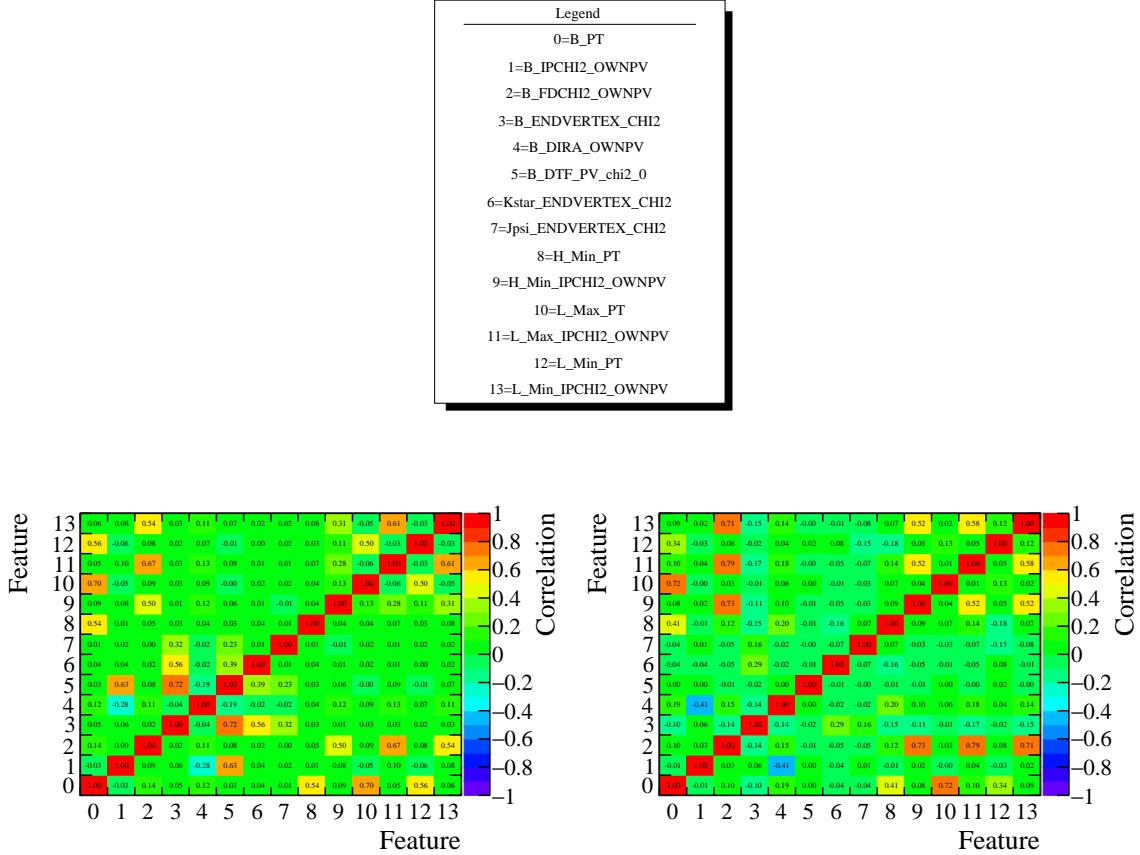


Figure 33: Correlation among the best discriminating variables for the signal (left) and the background (right) samples. The numbers on the axes refer to the variables reported in the legend.

The optimisation of a given classifier is achieved by performing a grid search in its parameter space, that is, by evaluating the performances of the classifier for different settings of the available parameters. Further details on the settings used in each algorithm is given in Appendix D.2.

Different configurations of the same classifier are compared to each other by plotting their Receiver Operating Characteristic (ROC) curves and calculating the integral of these curves (AUC). These show the dependence between true positive rate, corresponding to the fraction of signal candidates that are correctly labelled, and false positive rate, corresponding to the fraction of background candidates that are wrongly labelled, for different thresholds of the classifier response. The ideal classifier would have true positive rate equal to unity and false positive rate equal to zero. In reality, the true positive rate is equal to unity only for a threshold equal to zero, which corresponds to the maximum false positive rate, and goes from unity to zero as the threshold increases. The configuration associated to the largest ROC AUC is chosen as the optimal configuration of the classifier.

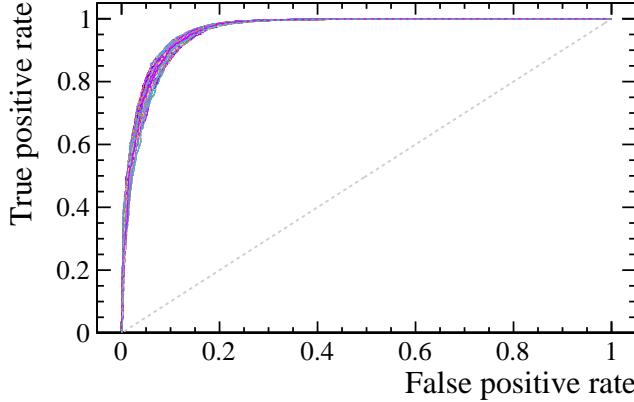


Figure 34: ROC curve of each configuration of the `XGBoost` classifier. Each color refers to a specific configuration of the classifier.

Both `GradientBoosting` and `XGBoost` algorithms provide the highest ROC AUC with similar response. The `XGBoost` is used as the nominal classifier for this analysis. The configuration corresponding to the best ROC AUC of `XGBoost` is then used as the nominal classifier in this analysis.

Figures 34, 36, 37, and 38 show the grid-search performed as ROC curves, the classifier response, the feature importance, and the efficiency as a function of the B^0 invariant mass and of the ϕ , θ_K , and θ_ℓ decay angles for both signal and background samples in several thresholds of the classifier response, respectively. The classifier is flat in the three decay angles, but shows a linear dependence of the efficiency on the B^0 invariant mass for both signal and background. Some of the BDT output plots feature a peak of the background test samples close to 1, which could indicate some small signature of overtraining. This seems to be associated to the small size of the background training sample, *e.g.* it disappears when the maximum tree depth is reduced. In any case, the impact of this feature should be irrelevant, in particular because all efficiencies are estimated effectively using the “test” scores, and no strange artefacts are seen when using tight MVA cuts.

4.4.4 Uniform classifiers

The possibility to use a uniform classifier to discriminate between signal and combinatorial background is also investigated. A uniform classifier can be trained in a way that guarantees a uniform classifier response over one or more variables. Such uniformity, in particular, can be required for either the signal or background sample used in the training. One typical situation in which uniformity in the predictions can be useful is when no prior information on the dynamics of a given phase-space is available. In this case, a uniform classifier reduces potential risks on bias the signal distribution of interest. In this analysis the invariant mass and in the decay angles θ_ℓ , θ_K , and ϕ are the features to consider. If the classifier response is not uniform in these variables, a bias might appear in some of the

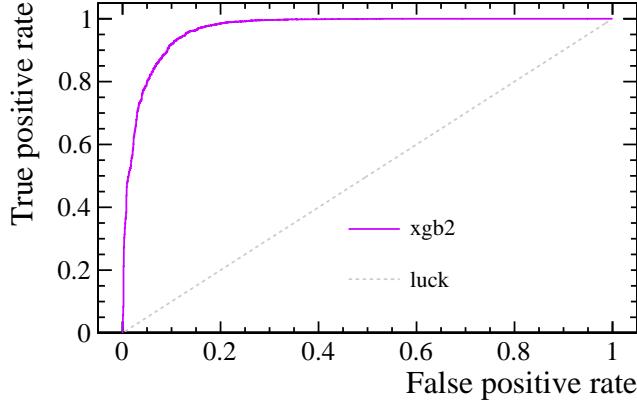


Figure 35: ROC curve, corresponding to the selected configuration of XGBoost.

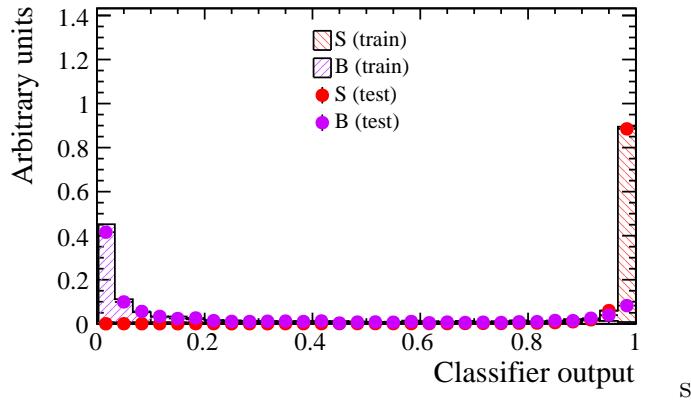


Figure 36: Comparison among the classifier responses to the training and testing samples of signal and background, corresponding to the selected configuration of XGBoost.

1030 angular observables.

1031 The performances of the **XGBoost** classifier obtained from the optimisation described
1032 above are compared to those achievable by using the following uniform classifiers:

- 1033 • **uBoost**;
- 1034 • **UGBknnAda**;
- 1035 • **UGBFatnessLoss**.

1036 For each classifier listed above, two configurations are tested, one with uniformity in the
1037 B^0 invariant mass only and one with uniformity also in the decay angles θ_ℓ , θ_K , and ϕ .
1038 The second configuration is expected to provide lower discriminating power, since more
1039 constraints have to be accommodated. It is found, however, that the performances of the

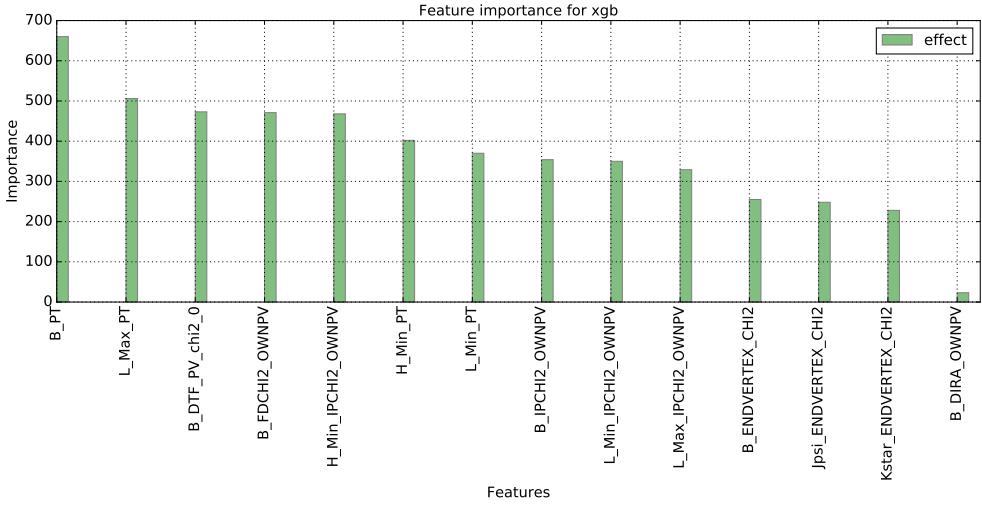


Figure 37: Feature importance, corresponding to the selected configuration of **XGBoost**.

uniform classifiers do not change significantly when imposing uniformity over additional variables. The performance of these classifiers are gathered in Appendix D.3 for brevity. The main conclusion from these studies is that the overall performances of the uniform classifiers are worse than those achievable using **XGBoost**, and given that the improvement in uniformity with respect to **XGBoost** is negligible, it is preferred to use **XGBoost**.

4.4.5 k-folding

In order to make use of the full statistics of the signal and background samples in the training, a k-folding approach is adopted. This strategy consists in splitting the input sample randomly in k subsamples s_i , with $i \in \{1 \dots k\}$. After that, k different classifiers are trained. The classifier j is trained on the union of the s_i subsamples, with $i \neq j$, and then tested on the excluded subsample s_j . This procedure is repeated k times, so that all combinations of $k-1$ subsamples out of k are used in the training phase. The k classifiers are then combined together and used to classify candidates that have not been used during the training and testing. This can be achieved by selecting one classifier randomly and using its prediction or by averaging the predictions of all classifiers. For the measurement described here, the former method has been preferred.

As k increases, more and more candidates are used in the training of each classifier. This is expected to increase the stability of the classifier response and decrease the risk of overfitting. On the other hand, the testing sample becomes smaller with increasing k , so the uncertainty on the performance assessment increases. In addition, given two classifiers c_j and c_k , the overlap between the candidates used in the training of c_j and in the training of c_k increases with k . At the extreme case in which all but one candidate out of N are used to train each of $N - 1$ classifiers, the overlap is maximum. In order to find the optimal value of k , three different possibilities have been tested: $k = 2$, $k = 5$, and $k = 10$. No difference in the results has been observed when varying the value of k . The

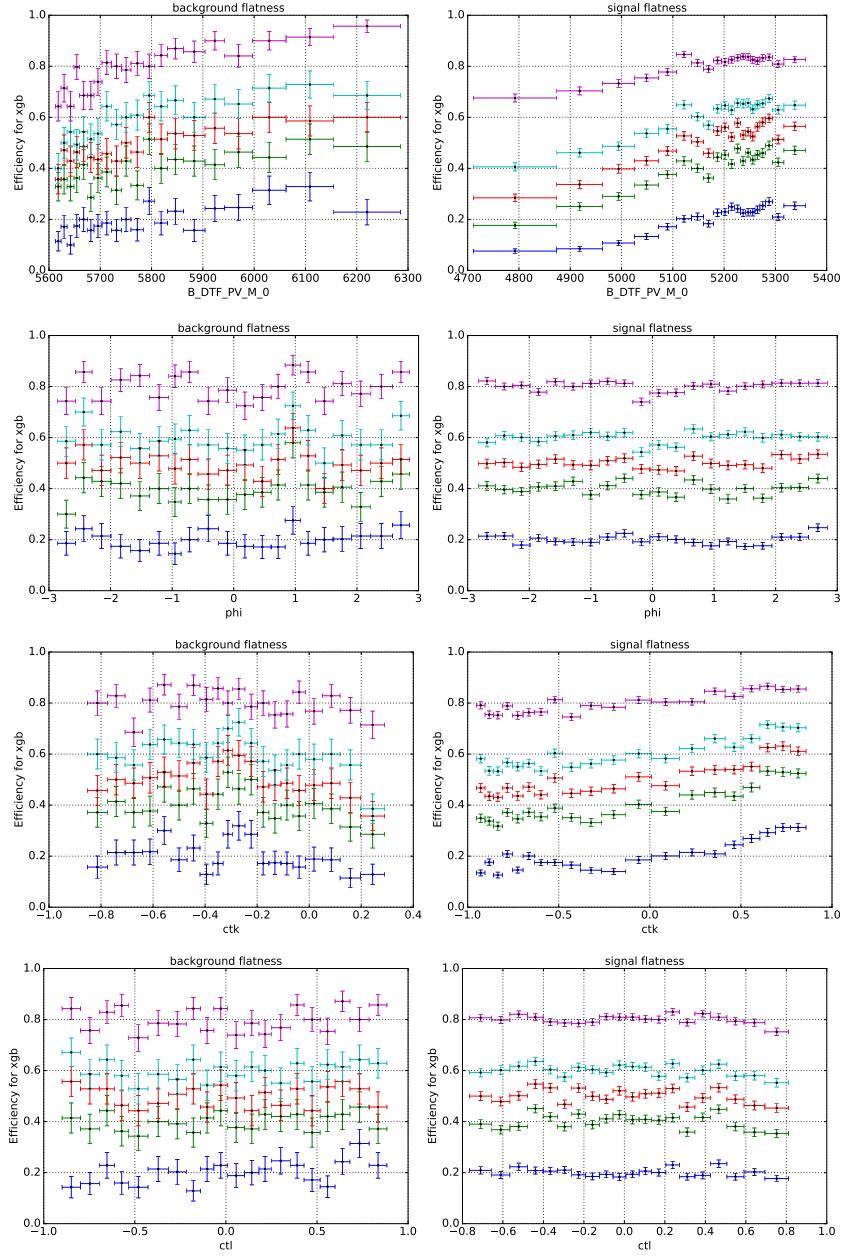


Figure 38: Efficiency as a function of the B^0 mass (a) and of the ϕ (b), θ_K (c), and θ_ℓ (d) decay angles, corresponding to the selected configuration of XGBoost. The plots on the left show the efficiency for the background sample, while those on the right show the efficiency for the signal sample.

1065 folding with $k=10$ has been chosen as default strategy since it is consistent with previous
 1066 measurements at LHCb.

1067 **4.5 MVA optimisation**

1068 The choice of the optimal cut on the MVA response can be driven by several figures-of-merit
 1069 (FoM), each appropriate for a certain situation. The statistical significance of the signal
 1070 with respect to background

$$\frac{S}{\sqrt{S+B}} , \quad (16)$$

1071 is commonly employed in many statistically limited analyses. A valid alternative would be
 1072 instead to optimize the MVA selection based on the expected uncertainty on the angular
 1073 observables. Both these FoMs are considered in this section, it is found that both provide
 1074 very similar optimal working points. The optimization procedure acts as follows

- 1075 (1) A mass fit is performed on the rare mode data with a very tight MVA cut (> 0.999).
 1076 This is done in order to have a stable estimation of the signal and partially recon-
 1077 structed yields in a configuration with heavily suppress combinatorial background.
 1078 The result of the fit to this reference point is shown in Figure 39 and Table 40.
 1079 Additional details about the implementation of the mass fit used at this stage are
 1080 given in Appendix D.6.
- 1081 (2) For each MVA cut, the expected yields of the different fit components are extracted
 1082 for the two L0 trigger categories of Run-I and Run-II separately. Depending on the
 1083 component, this estimation is done in a different way:
 - 1084 i *signal*: the expected signal yield is scaled from the tight MVA reference point
 1085 using MC ratio via

$$N_{\text{MVA}>x}^{\text{exp.}}(x) = \frac{\epsilon_{\text{MVA}>x}^{\text{MC}}}{\epsilon_{\text{MVA}>\text{ref.}}^{\text{MC}}} \times N_{\text{MVA}>\text{ref.}}^{\text{data}} , \quad (17)$$

1086 this allows an estimation that is free from possible statistical fluctuation present
 1087 in data;

1088 ii *backgrounds*: part-reco and combinatorial yield are extracted from data with a
 1089 mass fit carried out for each MVA cut under consideration. In order to help the
 1090 stability of the fit (especially in the case of large background contamination),
 1091 the partially reconstructed component is Gaussian constrained to the value
 1092 obtained by Eq. 17 (the standard deviation is taken to be half of the expected
 1093 yield). For illustration, the results of such mass fits for three different MVA
 1094 thresholds are shown in Appendix D.6.

- 1095 (3) From the estimated signal and background yields the FoM $S/\sqrt{S+B}$ can be cre-
 1096 ated. Alternatively, from the same quantities toys can be generated to access the
 1097 sensitivity to the angular observable P'_5 , which is used as a proxy for a FoM that
 1098 aims at minimizing the experimental uncertainty of the measurement. A detailed
 1099 description of the treatment of the toys generated for this specific purpose is given
 1100 in Appendix D.6.

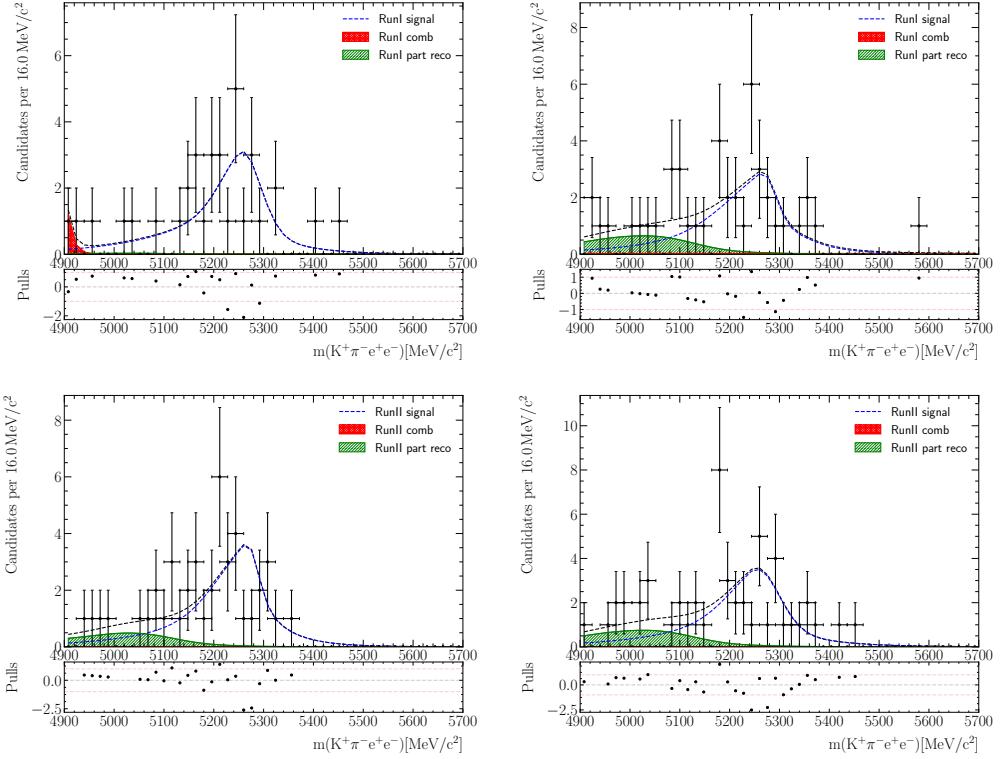


Figure 39: Results of the fit to the data sample with a tight MVA cut of 0.999. This fit is used in the estimation of partially reconstructed and signal yields in data fits and toy generation, respectively. Run-I (Run-II) results are shown in the top (bottom) rows, and two L0 categories are shown on the left (L0E) and right (L0TIS) columns.

Parameter	Run-I L0E	Run-I L0TIS	Run-II L0E	Run-II L0TIS
N_{Sig}	32 ± 6	33 ± 7	36 ± 7	41 ± 7
N_{Comb}	2 ± 2	1 ± 3	0 ± 45	0 ± 5
$N_{\text{Part. reco.}}$	0 ± 5	7 ± 4	36 ± 7	11 ± 5
$\lambda_{\text{exp.}}$	-0.09 ± 0.1	0 ± 0.1	0.0 ± 0.8	0.0 ± 0.9

Figure 40: Results obtained from the fit the the data sample with a tight MVA cut of 0.999. An extended maximum likelihood fit is made to directly obtain yields and their uncertainties.

1101 In summary, a schematic diagram illustrating the key components of the optimization
1102 procedure is shown in Figure 41.

1103 4.5.1 Optimisation results

1104 Given the four-folds split of the dataset - the two L0 trigger categories and the Run-I
1105 and Run-II data taking periods - the optimization procedure results in a four dimensional
1106 problem. A methodical grid search of the optimal solution is computationally very



Figure 41: Schematic diagram showing the steps involved in an evaluation of the P'_5 -FOM for a generic MVA cut. The fit to the sample with a high MVA threshold is performed once, while the other steps within the blue dotted line are repeated for every cut. The same estimation of the signal and background yields that enter in the generation of the toys are used to build the standard $S/\sqrt{S+B}$ FoM.

expensive, therefore some simplifications are necessary to make the dimensionality of the parameter space more accessible.

First, the considered MVA cuts are assumed to be common between two data taking periods, while only the L0 trigger categories are allowed to have separate cuts. This choice is motivated by the distribution of the classifier, which is found to be very compatible between the two runs (see Fig. 144 in App. D.4). On the other hand, the different trigger categories can be characterized by a different signal/background ratio, which may lead to different optimal MVA selection.

The resulting two-dimensional optimisation procedure is then approached as follows. First, a set of points along the diagonal of the two dimensional plan of L0E and L0TIS cut values is evaluated, to determine, in a one dimensional scenario, the region of interest. This region is found to extend from approximately 0.85 to 1.0. Subsequently it is divided into bins of varying sizes, with narrower bins closer to the limit at around 0.9964, beyond which fits tend to become unstable, and points at the centers of the bins are evaluated. Figure 42 shows the one-dimensional and two-dimensional scans obtained for the P'_5 -FoM. The results indicate no significant advantage in using different cuts for L0E and L0TIS, thus one common value will be used for both. Furthermore, the P'_5 -FoM seems to favour tight MVA cuts, however there is no clear optimal solution since all cuts above 0.9 gives similar sensitivity. On the other hand, the failure rate associated to the simplified angular

1126 fit performed in this toy study is found to increase for higher MVA thresholds (see bottom
1127 plots of Fig. 42).

1128 Finally, Figure 43 shows the $S/\sqrt{S+B}$ FoM, which suggest that the optimal point
1129 lies close to the limit of very tight MVA cuts. In conclusion, both FOMs suggest a tight
1130 or very tight working point, furthermore, considering the need to keep fit failure rate low,
1131 the threshold of 0.99 is chosen as the nominal working point.

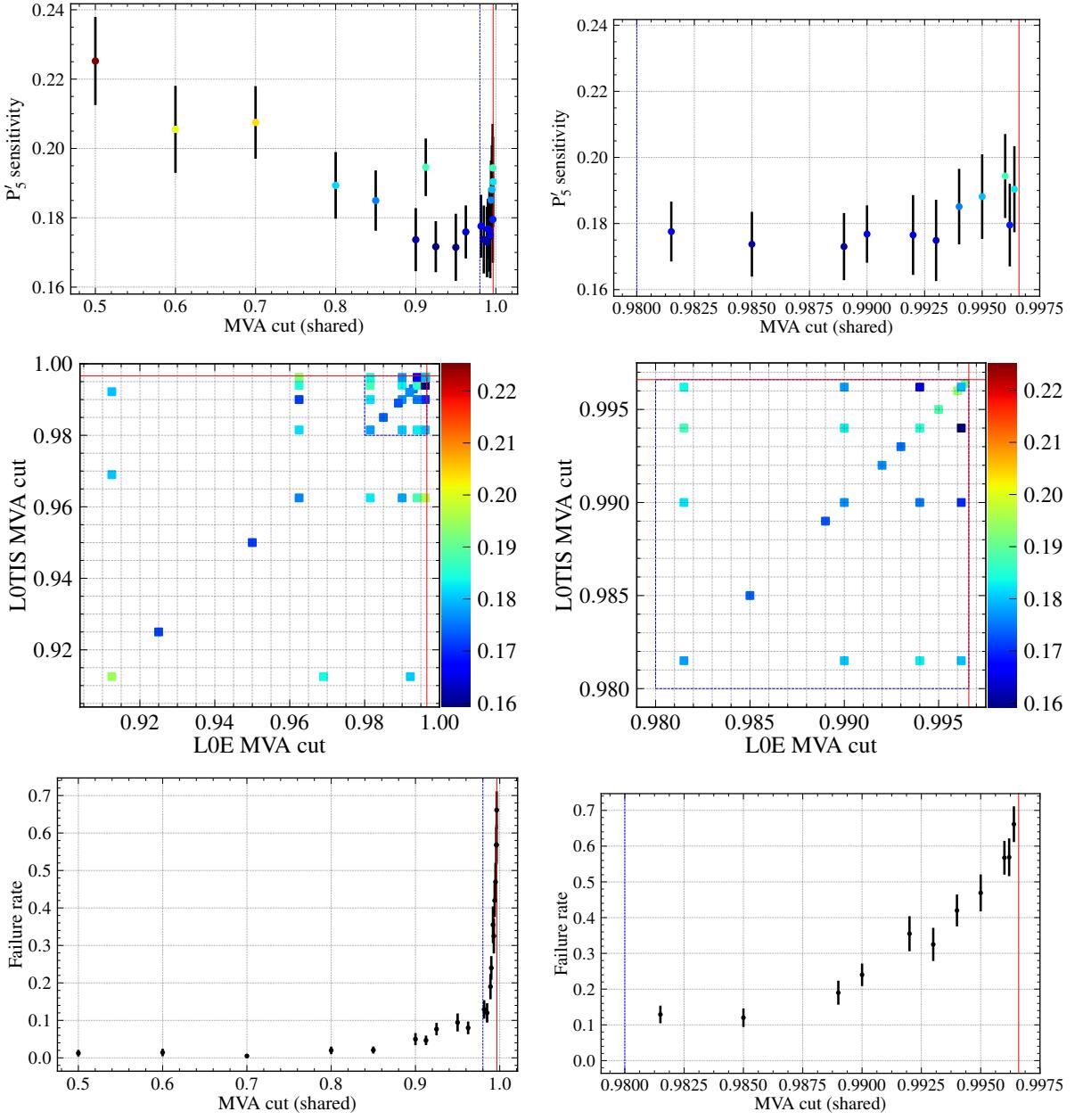


Figure 42: One dimensional scan of the P'_5 sensitivity against MVA cut thresholds shared by L0E and L0TIS categories (first row), and two dimensional scan in the region of interest (second row). In the two dimensional plots, the colors give indications of the sensitivities. The fraction of failed toys is shown for the one dimensional case in the third row. Tight MVA cuts are favoured, although the sensitivity approximately plateaus above $MVA > 0.9$. Fit failure rate rises steeply at very tight cut values as signal and background yields drop. No clear advantage is seen for the usage of different MVA cuts for the two L0 categories. Some points at very tight cut values for one of the two categories do seem to suggest improvements to sensitivity, but they are not considered as they are associated with very high fit failure rates.

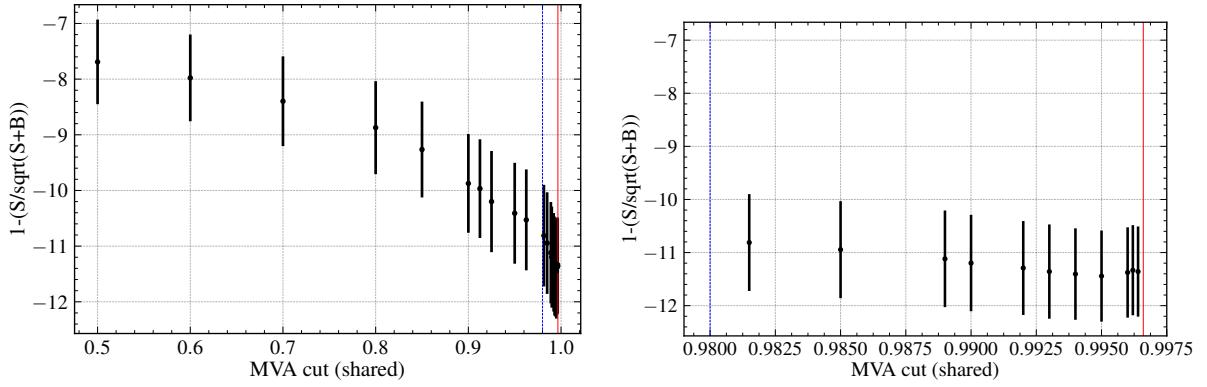


Figure 43: The common figure of merit based on signal and background yields show good agreement with the P'_5 sensitivity FOM. Here one minus the FOM value is plotted to allow for the ease of visual comparison. Here the background yield refers to the sum of the yields of the combinatorial and partially reconstructed components.

1132 4.6 Multiple candidates

1133 After the full selection described in the previous sections is applied, around 0.3% of
1134 the $B^0 \rightarrow K^{*0} e^+ e^-$ candidates and 0.5% of the $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ candidates from
1135 simulation belong to proton–proton collisions with multiple candidates. In data, 0.7% of
1136 the control mode candidates belong to proton–proton collisions with multiple candidates,
1137 while no instances are found for the rare mode. When multiple candidates are present,
1138 one is chosen randomly, while the others are discarded.

1139 **5 Signal amplitude model**

The scripts used to model the mass and angular distribution of both signal and control modes can be found in

- 1140 ○ `ewp-bd2ksteeangular-central-q2/scripts/signal/`

1141
1142 While an accurate description of the mass spectrum is crucial in the separation of
1143 the signal from the different sources of background, additional information from the
1144 angular distribution can further aid in this process, in particular in the case of $B^0 \rightarrow$
1145 $K^{*0}e^+e^-$ decays. Therefore, unlike branching ratio analyses or the muonic version of this
1146 measurement, the invariant mass fit is only discussed in combination with the angular
1147 parametrisation. This section details the signal modelling definition and corresponding
1148 features involved in the data fit.

1149 **5.1 Mass signal model**

1150 The invariant mass distribution of the signal candidates has a clear asymmetric tail pattern
1151 that result from a combination of effects, such as final state radiation, bremsstrahlung (over-)
1152 corrections and stochastic tracking imperfections. The model chosen to parametrise truth-
1153 matched $B^0 \rightarrow K^{*0}e^+e^-$ decays consist in a sum of two Crystal Ball (CB) functions [74],
1154 that share common values for the peak position and width, but have independent power
1155 law tails on opposite sides of the peak. The DCB function is defined as

$$\mathcal{P}_{\text{DCB}}(m|\mu, \sigma, \alpha, n) = \begin{cases} \frac{a_L}{\left(b_L - \frac{(m-\mu)}{\sigma}\right)^n} & \text{if } \frac{(m-\mu)}{\sigma} < \alpha_L \\ e^{\frac{-(m-\mu)^2}{2\sigma^2}} & \text{if } -\alpha_L < \frac{(m-\mu)}{\sigma} < \alpha_R \\ \frac{a_R}{\left(b_R + \frac{(m-\mu)}{\sigma}\right)^n} & \text{if } \frac{(m-\mu)}{\sigma} > \alpha_R \end{cases} \quad (18)$$

1156 with

$$a_{L,R} = \left(\frac{n_{L,R}}{|\alpha_{L,R}|} \right)^{n_{L,R}} e^{-\frac{1}{2}\alpha_{L,R}^2}, \quad (19)$$

$$b_{L,R} = \frac{n}{|\alpha_{L,R}|} - \alpha_{L,R},$$

1157 where μ and σ are the mean value and resolution of the Gaussian part of the function,
1158 The sign of the α parameter governs the right-handed or left-handed location of the tail
1159 and the parameter n drives the power law associated to the decrease of the tail.

1160 As the bremsstrahlung recovery procedure creates a significant impact on the shape of
1161 the signal mass distribution, $S(m)$, events belonging to the following three categories are
1162 parametrised separately:

- 1163 • 0γ – no photon recovery;
- 1164 • 1γ – one electron received momentum correction from one associated photon cluster;
- 1165 • 2γ – more than one electron received momentum corrections from one or more
1166 associated photon clusters.

1167 The full mass model is then

$$S(m) = f_{0\gamma}S_{0\gamma}(m) + f_{1\gamma}S_{1\gamma}(m) + (1 - f_{0\gamma} - f_{1\gamma})S_{2\gamma}(m) , \quad (20)$$

1168 where $f_{0\gamma}$ and $f_{1\gamma}$ – the fractions of events belonging to the 0γ and 1γ categories – are
1169 calculated from simulation. In the case of the control mode of $B^0 \rightarrow K^{*0}J/\psi$, due to
1170 the increased mass range, a Gaussian component is added to the DCB for the 1γ and
1171 2γ categories when fitting the constrained B^0 mass. This is done to allow for improved
1172 modelling of the tails of the mass distributions.

1173 The modelling of the rare and control mode mass shapes is carried out using kinematic
1174 and trigger correction weights, but without acceptance weights, as the correlation between
1175 mass and angles is expected to be minimal. The exclusion of acceptance weights also avoids
1176 problems associated with very large weight values that would require manual removal of
1177 events (see Section 8). However, the veto against background from B^+ decays introduces
1178 some correlation between mass and $\cos\theta_K$ by removing part of the phase space, which
1179 leads to minor non-uniformities across the mass range, namely that events located on the
1180 upper mass tail tend to receive smaller acceptance weights than those on the lower tail.
1181 The impact of neglecting acceptance weights will be assessed as a systematic associated
1182 with the signal model choice.

1183 The results of the fits to simulation for the rare and control modes are shown in Figs 44,
1184 45, 46 and 47. The corresponding parameter values obtained are given in Tables 12
1185 and 13. The sizes of the samples used for each category, and the calculated 0γ and 1γ
1186 fractions are reported in Tables 14 and 15, respectively. In fits to data, to account for
1187 residual simulation-data differences, the mean of the mass peak is allowed shift about the
1188 value determined from simulation, and likewise its width is allowed to scale. This is made
1189 possible by the redefinition of μ as $\mu + \delta_\mu$, and σ as $\sigma + \delta_\sigma$. For simplicity, a single set of
1190 shift and scaling parameters is used for all bremsstrahlung categories, and for both the
1191 DCB and Gaussian components where applicable.

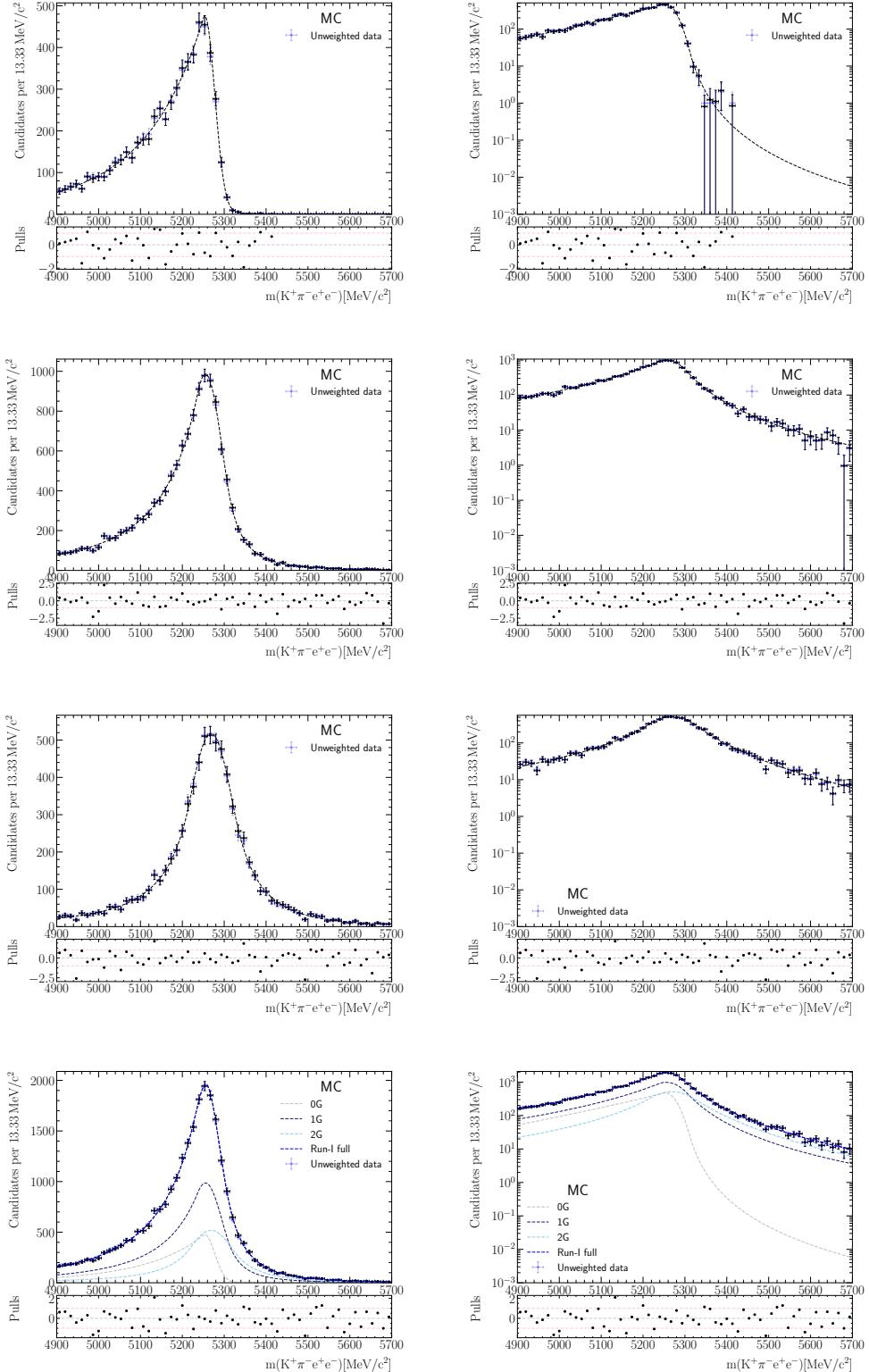


Figure 44: Invariant mass distributions of the signal for the (top) 0γ , (second row) 1γ , (third row) 2γ and (bottom) combined bremsstrahlung categories for Run-I dataset.

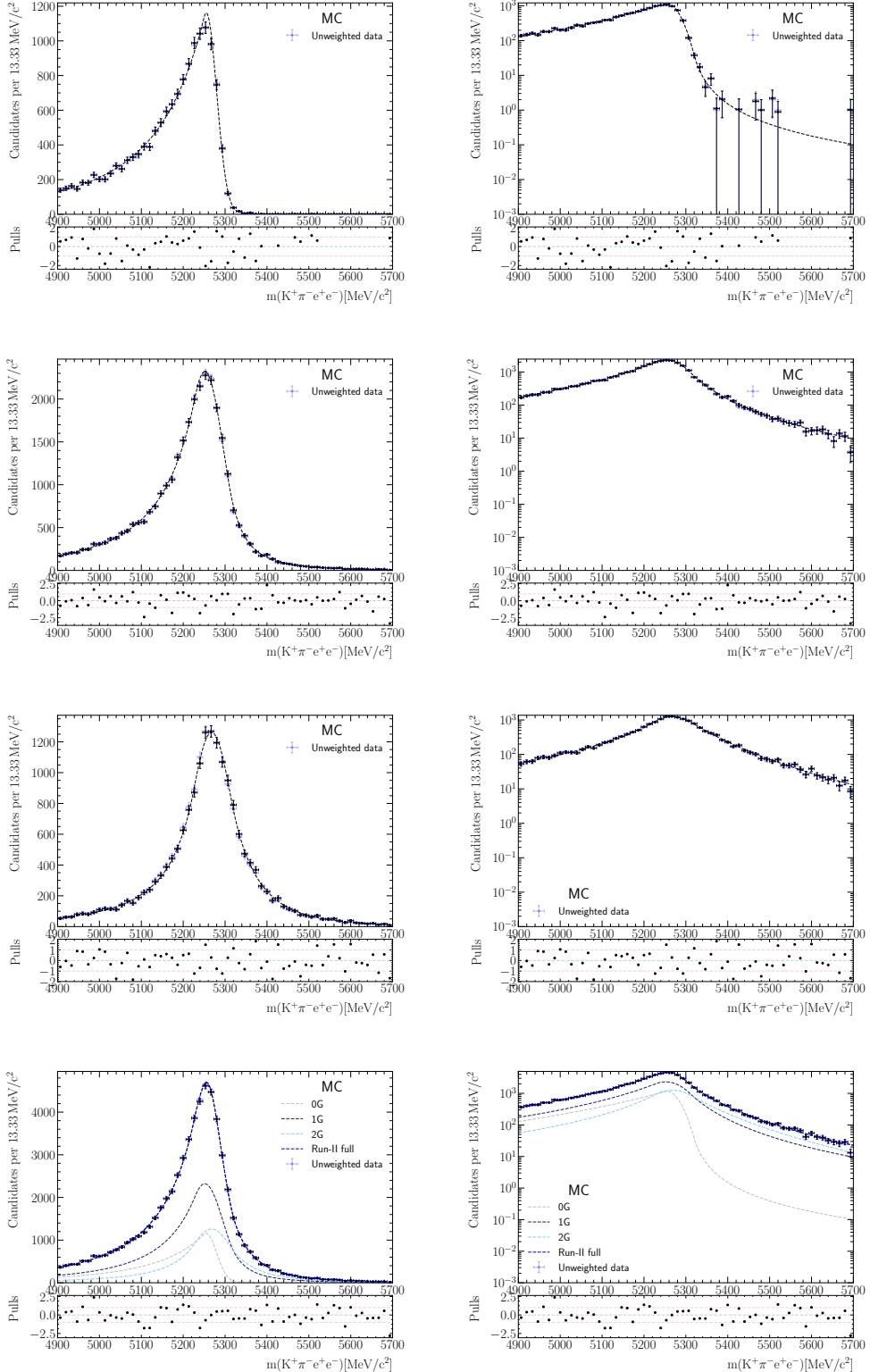


Figure 45: Invariant mass distributions of the signal for the (top) 0γ , (second row) 1γ , (third row) 2γ and (bottom) combined bremsstrahlung categories for Run-II dataset.

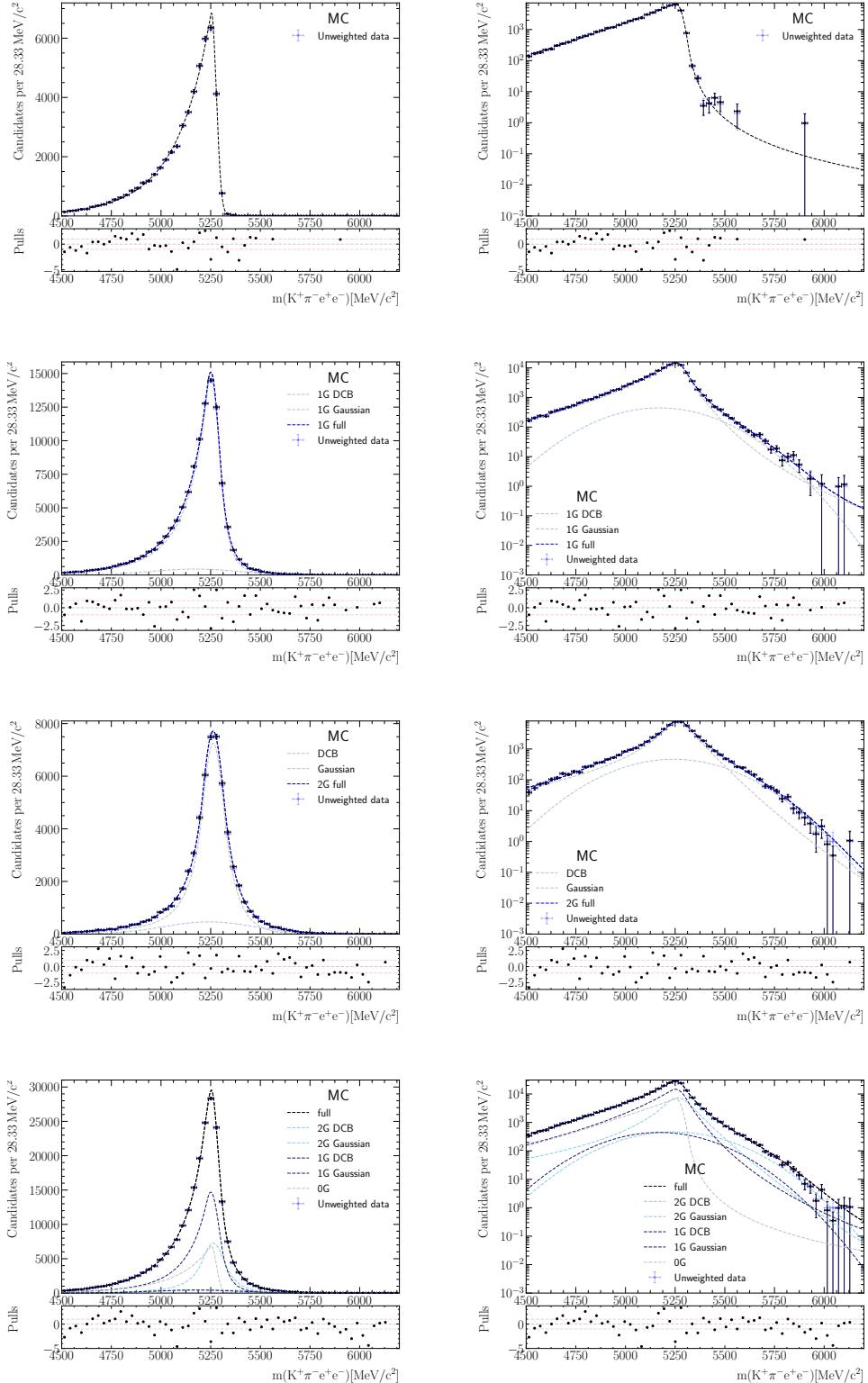


Figure 46: Invariant mass distributions of the control channel for the (top) 0γ , (second row) 1γ , (third row) 2γ and (bottom) combined bremsstrahlung categories for Run-I dataset.

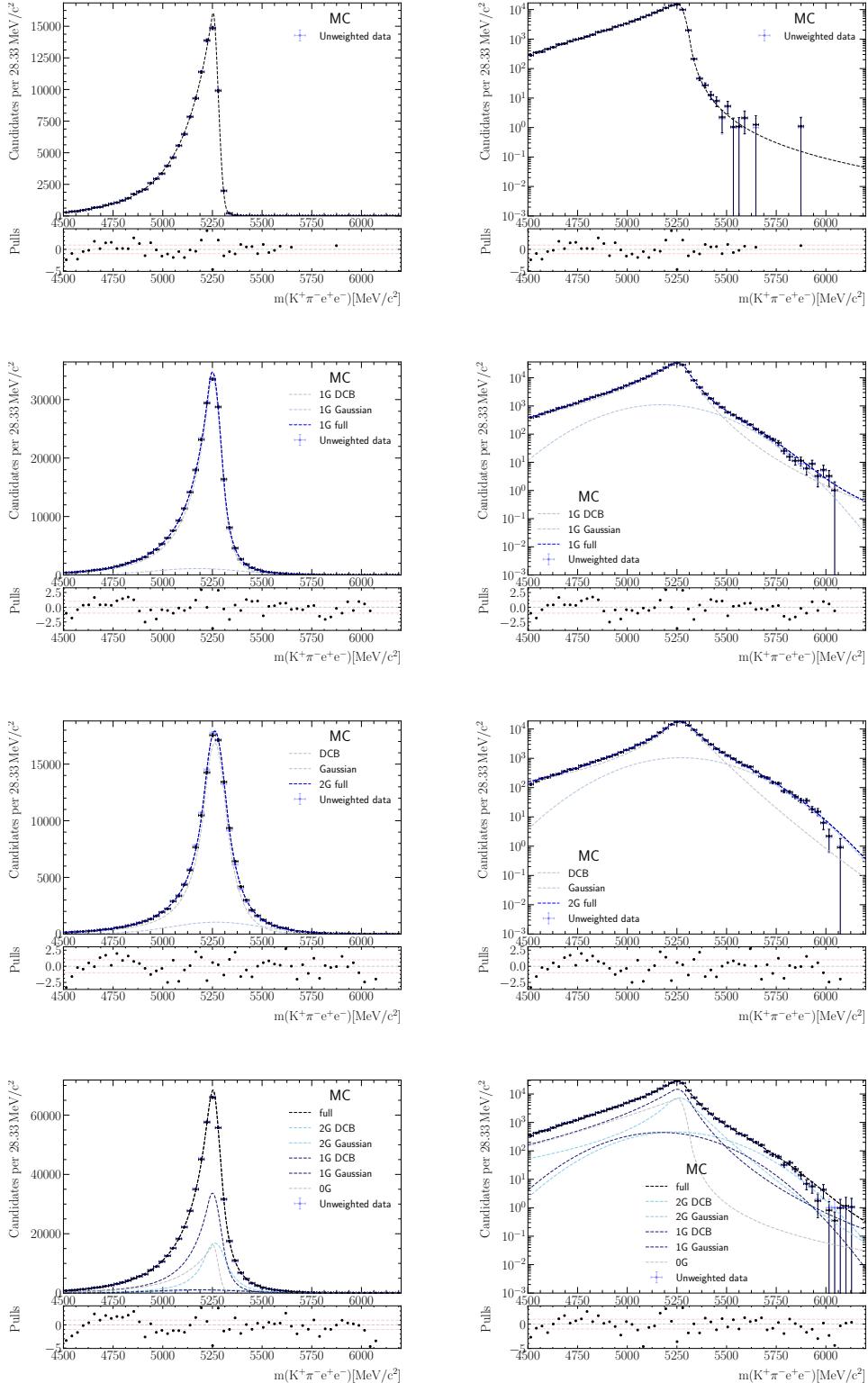


Figure 47: Invariant mass distributions of the control channel for the (top) 0γ , (second row) 1γ , (third row) 2γ and (bottom) combined bremsstrahlung categories for Run-II dataset.

Parameter	Run-I value	Run-II value
$\mu_{0\gamma}^{\text{DCB}}$	5253.8 \pm 2.3	5255.1 \pm 1.6
$\mu_{1\gamma}^{\text{DCB}}$	5255.3 \pm 1.4	5251.7 \pm 1.1
$\mu_{2\gamma}^{\text{DCB}}$	5268.5 \pm 1.6	5267.9 \pm 1.0
$\sigma_{0\gamma}^{\text{DCB}}$	23.7 \pm 1.3	24.8 \pm 1.0
$\sigma_{1\gamma}^{\text{DCB}}$	40.9 \pm 1.8	44.7 \pm 1.2
$\sigma_{2\gamma}^{\text{DCB}}$	54.8 \pm 3.4	50.8 \pm 2.4
$\alpha_{L;0\gamma}^{\text{DCB}}$	0.167 ± 0.021	0.211 ± 0.018
$\alpha_{L;1\gamma}^{\text{DCB}}$	0.471 ± 0.035	0.537 ± 0.030
$\alpha_{L;2\gamma}^{\text{DCB}}$	0.92 ± 0.10	0.81 ± 0.05
$\alpha_{R;0\gamma}^{\text{DCB}}$	-2.40 ± 0.13	-2.61 ± 0.16
$\alpha_{R;1\gamma}^{\text{DCB}}$	-1.27 ± 0.07	-1.31 ± 0.04
$\alpha_{R;2\gamma}^{\text{DCB}}$	-1.02 ± 0.10	-0.88 ± 0.06
$n_{L;0\gamma}^{\text{DCB}}$	9.0 \pm 8.0	3.7 \pm 0.7
$n_{L;1\gamma}^{\text{DCB}}$	2.8 \pm 0.4	2.78 \pm 0.30
$n_{L;2\gamma}^{\text{DCB}}$	2.2 \pm 0.4	2.43 \pm 0.27
$n_{R;0\gamma}^{\text{DCB}}$	3.3 \pm 0.9	1.9 \pm 0.5
$n_{R;1\gamma}^{\text{DCB}}$	2.92 \pm 0.30	2.92 \pm 0.19
$n_{R;2\gamma}^{\text{DCB}}$	3.7 \pm 0.7	4.8 \pm 0.7

Table 12: Parameter values obtained from fits to the PV-constrained B^0 mass of the $B^0 \rightarrow K^{*0} e^+ e^-$ simulation.

Parameter	Run-I value	Run-II value
$\mu_{0\gamma}^{\text{DCB}}$	5253.7 \pm 2.1	5253.3 \pm 0.6
$\mu_{1\gamma}^{\text{DCB}}$	5250.6 \pm 1.8	5250.4 \pm 0.6
$\mu_{2\gamma}^{\text{DCB}}$	5262.4 \pm 0.6	5261.9 \pm 1.1
$\sigma_{0\gamma}^{\text{DCB}}$	24.8 \pm 1.0	25.8 \pm 0.4
$\sigma_{1\gamma}^{\text{DCB}}$	44.3 \pm 2.1	45.4 \pm 0.7
$\sigma_{2\gamma}^{\text{DCB}}$	58.4 \pm 1.3	59.8 \pm 2.0
$\alpha_{L;0\gamma}^{\text{DCB}}$	0.146 ± 0.021	0.169 ± 0.004
$\alpha_{L;1\gamma}^{\text{DCB}}$	0.43 ± 0.07	0.456 ± 0.014
$\alpha_{L;2\gamma}^{\text{DCB}}$	0.92 ± 0.04	0.87 ± 0.14
$\alpha_{R;0\gamma}^{\text{DCB}}$	-2.48 ± 0.10	-2.48 ± 0.07
$\alpha_{R;1\gamma}^{\text{DCB}}$	-1.21 ± 0.20	-1.243 ± 0.028
$\alpha_{R;2\gamma}^{\text{DCB}}$	-0.96 ± 0.05	-0.96 ± 0.10
$n_{L;0\gamma}^{\text{DCB}}$	14.0 \pm 14.0	9.6 \pm 0.7
$n_{L;1\gamma}^{\text{DCB}}$	5.1 \pm 0.9	4.65 \pm 0.17
$n_{L;2\gamma}^{\text{DCB}}$	2.80 \pm 0.19	2.8 \pm 0.7
$n_{R;0\gamma}^{\text{DCB}}$	2.60 \pm 0.30	2.84 \pm 0.26
$n_{R;1\gamma}^{\text{DCB}}$	7.0 \pm 5.0	7.1 \pm 0.4
$n_{R;2\gamma}^{\text{DCB}}$	23.0 \pm 5.0	30.0 \pm 50.0
$\mu_{1\gamma}^{\text{G}}$	5170.0 \pm 90.0	5174.0 \pm 17.0
$\mu_{2\gamma}^{\text{G}}$	5242.0 \pm 18.0	5270.0 \pm 15.0
$\sigma_{1\gamma}^{\text{G}}$	219.0 \pm 14.0	223.0 \pm 7.0
$\sigma_{2\gamma}^{\text{G}}$	227.0 \pm 8.0	229.0 \pm 16.0
$f_{\text{DCB};1\gamma}$	0.92 ± 0.06	0.914 ± 0.008
$f_{\text{DCB};2\gamma}$	0.844 ± 0.015	0.85 ± 0.06

Table 13: Parameter values obtained from fits to the PV-constrained B^0 mass of the $B^0 \rightarrow K^{*0} J/\psi$ simulation.

Sample		Run-I	Run-II
$B^0 \rightarrow K^{*0} e^+ e^-$	0γ	4692	11168
	1γ	10247	24586
	2γ	5775	13862
$B^0 \rightarrow K^{*0} J/\psi(\rightarrow e^+ e^-)$	0γ	51000	113617
	1γ	110629	253349
	2γ	59659	143366

Table 14: Number of events in samples used to determine rare and control mode mass shapes.

Mode	Parameter	Run-I	Run-II
$B^0 \rightarrow K^{*0} e^+ e^-$	$f_{0\gamma}$	0.227 ± 0.004	0.2251 ± 0.0024
	$f_{1\gamma}$	0.495 ± 0.006	0.496 ± 0.004
$B^0 \rightarrow K^{*0} J/\psi(\rightarrow e^+ e^-)$	$f_{0\gamma}$	0.2305 ± 0.0012	0.2226 ± 0.0007
	$f_{1\gamma}$	0.4999 ± 0.0019	0.4964 ± 0.0012

Table 15: Bremsstrahlung category fractions calculated from corrected simulations for Run-I and Run-II samples of the signal and control modes.

5.2 Angular signal model

The final state of the decay $B^0 \rightarrow K^{*0} e^+ e^-$ can be described by q^2 , the invariant mass squared of the di-lepton system, and three decay angles $\vec{\Omega} = (\theta_\ell, \theta_K, \phi)$. After summing over the lepton spins and integrating over the $K\pi$ invariant mass, its differential decay rate can be written as [75]

$$\begin{aligned} \frac{d^4\Gamma[\bar{B}^0 \rightarrow \bar{K}^{*0} e^+ e^-]}{dq^2 d\vec{\Omega}} &= \frac{9}{32\pi} \sum_i I_i(q^2) f_i(\vec{\Omega}) \text{ and} \\ \frac{d^4\Gamma[B^0 \rightarrow K^{*0} e^+ e^-]}{dq^2 d\vec{\Omega}} &= \frac{9}{32\pi} \sum_i \bar{I}_i(q^2) f_i(\vec{\Omega}), \end{aligned} \quad (21)$$

where $I_i = I_i(q^2)$ are q^2 -dependent angular coefficients, which can be expressed as bilinear combinations of six complex decay amplitudes, $\mathcal{A}_{0,\parallel,\perp}^{R,L}$, which correspond to the different transversity states of the K^{*0} meson and the different (left- and right-handed) chiralities of the dimuon system. The total list of angular terms is summarised in Table 16.

Following the notation of Ref. [59], the q^2 -dependent CP -averaged observables S_i can be defined as

$$S_i = (I_i + \bar{I}_i) / \left(\frac{d\Gamma}{dq^2} + \frac{d\bar{\Gamma}}{dq^2} \right). \quad (22)$$

In the massless limit, the CP -averaged observables $S_{1(s,c)}$ and $S_{2(s,c)}$ obey the relations $S_{1s} = 3S_{2s}$, $S_{1c} = -S_{2c}$ and $\frac{3}{4}(2S_{1s} + S_{1c}) - \frac{1}{4}(2S_{2s} + S_{2c}) = 1$, as detailed in Ref. [59]. These relationships reduce the number of independent CP -averaged observables from eleven to eight. The S_{1c} observable corresponds to the fraction of longitudinal polarisation of the K^{*0} meson and is therefore more commonly referred to as F_L , with

$$F_L = S_{1c} = \frac{|\mathcal{A}_0^L|^2 + |\mathcal{A}_0^R|^2}{|\mathcal{A}_0^L|^2 + |\mathcal{A}_0^R|^2 + |\mathcal{A}_\parallel^L|^2 + |\mathcal{A}_\parallel^R|^2 + |\mathcal{A}_\perp^L|^2 + |\mathcal{A}_\perp^R|^2} \quad (23)$$

It is also conventional to replace S_{6s} by the forward-backward asymmetry of the di-lepton system A_{FB} , with $A_{FB} = \frac{3}{4}S_{6s}$. The CP -averaged angular distribution of the $B^0 \rightarrow K^{*0} e^+ e^-$ decay can then be written as

$$\begin{aligned} \frac{1}{d(\Gamma + \bar{\Gamma})/dq^2} \frac{d^4(\Gamma + \bar{\Gamma})}{dq^2 d\vec{\Omega}} &= \frac{9}{32\pi} \left[\frac{3}{4}(1 - F_L) \sin^2 \theta_K + F_L \cos^2 \theta_K \right. \\ &\quad + \frac{1}{4}(1 - F_L) \sin^2 \theta_K \cos 2\theta_\ell \\ &\quad - F_L \cos^2 \theta_K \cos 2\theta_\ell + S_3 \sin^2 \theta_K \sin^2 \theta_\ell \cos 2\phi \\ &\quad + S_4 \sin 2\theta_K \sin 2\theta_\ell \cos \phi + S_5 \sin 2\theta_K \sin \theta_\ell \cos \phi \\ &\quad + \frac{4}{3}A_{FB} \sin^2 \theta_K \cos \theta_\ell + S_7 \sin 2\theta_K \sin \theta_\ell \sin \phi \\ &\quad \left. + S_8 \sin 2\theta_K \sin 2\theta_\ell \sin \phi + S_9 \sin^2 \theta_K \sin^2 \theta_\ell \sin 2\phi \right], \end{aligned} \quad (24)$$

Additional sets of observables, for which the leading $B^0 \rightarrow K^{*0}$ form-factor uncertainties largely cancel, can be built from F_L and $S_3 - S_9$. Examples of such *optimised* observables

1213 include the transverse asymmetry $A_T^{(2)}$ [76], where $A_T^{(2)} = 2S_3/(1 - F_L)$, and the $P^{(\prime)}$ series
 1214 of observables [58]. In this analysis the notation used is

$$\begin{aligned} P_1 &= \frac{2S_3}{(1 - F_L)} = A_T^{(2)}, \\ P_2 &= \frac{2}{3} \frac{A_{FB}}{(1 - F_L)}, \\ P_3 &= \frac{-S_9}{(1 - F_L)}, \end{aligned} \tag{25}$$

$$\begin{aligned} P'_{4,5,8} &= \frac{S_{4,5,8}}{\sqrt{F_L(1 - F_L)}}, \\ P'_6 &= \frac{S_7}{\sqrt{F_L(1 - F_L)}}, \end{aligned}$$

(26)

1215 The definition of the P observables differs from that of Ref. [58], but is consistent with the
 1216 notation used in the LHCb analysis of Ref. [6, 7].

1217 5.2.1 S-wave interference

1218 Throughout this analysis the K^{*0} resonance refers to the $K^{*0}(892)$ vector meson, recon-
 1219 structed through its decay products $K^{*0} \rightarrow K^+ \pi^-$ and commonly known as P-wave state.
 1220 However, the reconstructed $K^+ \pi^-$ system can also originate either from a non-resonant
 1221 decay or from the decay of scalar resonances, generically denominated as S-wave pollution.
 1222 The contribution of the S-wave component has been measured to be small, below 10% in
 1223 the range $796 < m_{K\pi} < 996$ MeV/ c^2 [17]¹⁰. Nevertheless, events with the $K\pi$ system in
 1224 the S-wave configuration are known to modify the distributions of the decay angles [77].

1225 In this analysis, the S-wave contribution is neglected in the nominal signal pdf and
 1226 treated as a systematic uncertainty (see Sec. 10). This approach follows the one employed
 1227 in the first angular analyses of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays [78, 79], which were characterised by
 1228 a comparable statistic and differ from the more recent Ref. [6, 7] where this contribution
 1229 was explicitly added to the angular pdf of Eq. 24. For completeness the additional angular
 1230 terms necessary to describe the S-wave component are described below.

1231 The addition of an S-wave component introduces two new complex amplitudes, $\mathcal{A}_S^{L,R}$,
 1232 and results in the six additional angular terms that are given in the lower part of Table 16.
 1233 The presence of a $K\pi$ system in an S-wave configuration modifies the angular distribution

¹⁰Ref. [17] measured the S-wave fraction in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays, *i.e.* with muons instead of electrons in the final state. However, even if possible NP effect can also be present in the S-wave contribution, this is always expected to be a subleading effect and can be safely treated as a systematic given the limited statistic of this analysis.

¹²³⁴ to

$$\begin{aligned} \frac{1}{d(\Gamma + \bar{\Gamma})/dq^2} \frac{d^4(\Gamma + \bar{\Gamma})}{dq^2 d\vec{\Omega}} \Big|_{S+P} &= (1 - F_S) \frac{1}{d(\Gamma + \bar{\Gamma})/dq^2} \frac{d^4(\Gamma + \bar{\Gamma})}{dq^2 d\vec{\Omega}} \Big|_P \\ &\quad + \frac{3}{16\pi} \left[F_S \sin^2 \theta_l + S_{S1} \sin^2 \theta_l \cos \theta_K \right. \\ &\quad \quad \quad \left. + S_{S2} \sin 2\theta_l \sin \theta_K \cos \phi \right. \\ &\quad \quad \quad \left. + S_{S3} \sin \theta_l \sin \theta_K \cos \phi \right. \\ &\quad \quad \quad \left. + S_{S4} \sin \theta_l \sin \theta_K \sin \phi \right. \\ &\quad \quad \quad \left. + S_{S5} \sin 2\theta_l \sin \theta_K \sin \phi \right]. \end{aligned}$$

¹²³⁵ where F_S denotes the S-wave fraction,

$$F_S = \frac{|\mathcal{A}_S^l|^2 + |\mathcal{A}_S^R|^2}{|\mathcal{A}_S^l|^2 + |\mathcal{A}_S^R|^2 + |\mathcal{A}_0^L|^2 + |\mathcal{A}_0^R|^2 + |\mathcal{A}_{||}^L|^2 + |\mathcal{A}_{||}^R|^2 + |\mathcal{A}_{\perp}^L|^2 + |\mathcal{A}_{\perp}^R|^2} \quad (27)$$

¹²³⁶ and the terms S_{S1}, S_{S5} arise from interference between the S- and P-wave amplitudes. In
¹²³⁷ this analysis, the modified angular pdf of Eq. 27 is only used in systematic studies with
¹²³⁸ pseudoexperiments (see Sec. 10).

Table 16: Dependence of the angular coefficients $I_i(q^2)$ on the transversity amplitudes and the corresponding angular terms $f_i(\cos \theta_l, \cos \theta_K, \phi)$.

i	$I_i(q^2)$	$f_i(\vec{\Omega})$
1s	$\frac{3}{4} [\mathcal{A}_\perp^L ^2 + \mathcal{A}_\parallel^L ^2 + (L \rightarrow R)]$	$\sin^2 \theta_K$
1c	$[\mathcal{A}_0^L ^2 + \mathcal{A}_0^R ^2]$	$\cos^2 \theta_K$
2s	$\frac{1}{4} [\mathcal{A}_\perp^L ^2 + \mathcal{A}_\parallel^L ^2 + (L \rightarrow R)]$	$\sin^2 \theta_K \cos 2\theta_l$
2c	$-[\mathcal{A}_0^L ^2 + \mathcal{A}_0^R ^2]$	$\cos^2 \theta_K \cos 2\theta_l$
3	$\frac{1}{2} [\mathcal{A}_\perp^L ^2 - \mathcal{A}_\parallel^L ^2 + (L \rightarrow R)]$	$\sin^2 \theta_K \sin^2 \theta_l \cos 2\phi$
4	$\frac{1}{\sqrt{2}} \mathcal{R}e [\mathcal{A}_0^L \mathcal{A}_\parallel^{L*} + (L \rightarrow R)]$	$\sin 2\theta_K \sin 2\theta_l \cos \phi$
5	$\sqrt{2} \mathcal{R}e [\mathcal{A}_0^L \mathcal{A}_\perp^{L*} - (L \rightarrow R)]$	$\sin 2\theta_K \sin \theta_l \cos \phi$
6s	$2 \mathcal{R}e [\mathcal{A}_\parallel^L \mathcal{A}_\perp^{L*} - (L \rightarrow R)]$	$\sin 2\theta_K \cos \theta_l$
7	$\sqrt{2} \mathcal{I}m [\mathcal{A}_0^L \mathcal{A}_\parallel^{L*} - (L \rightarrow R)]$	$\sin 2\theta_K \sin \theta_l \sin \phi$
8	$\frac{1}{\sqrt{2}} \mathcal{I}m [\mathcal{A}_0^L \mathcal{A}_\perp^{L*} + (L \rightarrow R)]$	$\sin 2\theta_K \sin 2\theta_l \sin \phi$
9	$\mathcal{I}m [\mathcal{A}_\perp^L \mathcal{A}_\parallel^{L*} + (L \rightarrow R)]$	$\sin^2 \theta_K \sin^2 \theta_l \sin 2\phi$
1c ^S	$\frac{1}{3} [\mathcal{A}_{S0}^L ^2 + \mathcal{A}_{S0}^R ^2]$	1
2c ^S	$-\frac{1}{3} [\mathcal{A}_{S0}^L ^2 + \mathcal{A}_{S0}^R ^2]$	$\cos 2\theta_l$
1c	$\frac{2}{\sqrt{3}} \mathcal{R}e [\mathcal{A}_{S0}^L \mathcal{A}_0^{L*} + \mathcal{A}_{S0}^R \mathcal{A}_0^{R*}]$	$\cos \theta_K$
2c	$-\frac{2}{\sqrt{3}} \mathcal{R}e [\mathcal{A}_{S0}^L \mathcal{A}_0^{L*} + \mathcal{A}_{S0}^R \mathcal{A}_0^{R*}]$	$\cos \theta_K \cos 2\theta_l$
4	$\sqrt{\frac{2}{3}} \mathcal{R}e [\mathcal{A}_{S0}^L \mathcal{A}_\parallel^{L*} + (L \rightarrow R)]$	$\sin \theta_K \sin 2\theta_l \cos \phi$
5	$\sqrt{\frac{8}{3}} \mathcal{R}e [\mathcal{A}_{S0}^L \mathcal{A}_\perp^{L*} - (L \rightarrow R)]$	$\sin \theta_K \sin \theta_l \cos \phi$
7	$\sqrt{\frac{8}{3}} \mathcal{I}m [\mathcal{A}_{S0}^L \mathcal{A}_\parallel^{L*} - (L \rightarrow R)]$	$\sin \theta_K \sin \theta_l \sin \phi$
8	$\sqrt{\frac{2}{3}} \mathcal{I}m [\mathcal{A}_{S0}^L \mathcal{A}_\perp^{L*} + (L \rightarrow R)]$	$\sin \theta_K \sin 2\theta_l \sin \phi$

1239 **6 Acceptance**

The scripts used for the parametrisation of the acceptance as well as validation plots discussed in this section are given in

- 1240 ○ `ewp-bd2ksteeangular-central-q2/scripts/acceptance`

1241
1242 The variation of efficiency across the phase space that defines the signal decay, *i.e.* q_c^2
1243 and $\vec{\Omega} = (\cos \theta_\ell, \cos \theta_K, \phi)$, is an important feature to be included in the angular fit. While
1244 an ideal analysis would select events with equal probability from any region of the phase
1245 space, it has been already shown that several elements can distort such distributions, *e.g.*

- 1246 1. final state radiation (FSR);
1247 2. reconstruction and selection ('acceptance'); and
1248 3. resolution.

1249 Figure 48 illustrates the potential impact of these different components that distort the
1250 distributions of interest discussed below. The nominal acceptance function is obtained by
1251 parametrising reconstructed quantities in the post-selection sample with inverse weights
1252 from the generation level model (without FSR effects). This results in a function that
1253 corrects for both resolution (including FSR) and acceptance effects. It is also possible to
1254 remove the resolution correction by using true quantities in the parametrisation of the
1255 post-selection sample. This is used to study the impact of the model-dependency
1256 of resolution correction in the systematics. In the following, the definition, samples and
1257 parametrisation of the acceptance for the signal and control model are discussed.

1258 **6.1 Preliminaries**

1259 Final-state radiation (FSR) QED effects refer to the emission of photons by charged
1260 particles in the final state of the decay, and thus can naturally distort the distributions of
1261 interest. A complete evaluation of these corrections in $b \rightarrow s\ell^+\ell^-$ decay amplitudes is a
1262 non-trivial task, due to the interplay of perturbative and non-perturbative dynamics (see
1263 *e.g.* Ref. [80]). FSR QED is the part of electroweak correction that is not lepton flavour
1264 universal, *e.g.* different collinear logarithms $(\alpha/\pi) \log^2(m_B/m_\ell)$ [50, 81–83], and thus has
1265 to be properly modelled in the simulation. Bremsstrahlung emission in the decay of
1266 particles is simulated using the PHOTOS software with the default configuration [84], which
1267 has been used extensively in the community since 1989. More specifically, FSR corrections
1268 are performed at the generation level with a certain probability of an extra photon(s) to
1269 be added and the kinematics of other particles adjusted before running simulation of the
1270 detector.

1271 In addition, the reconstruction, trigger and selection criteria (see Sec. 4) also modify
1272 the distribution of the signal candidates. Finally, resolution effects refer to the smearing
1273 of different features of the decay chain due to the finite resolution of the detector. This

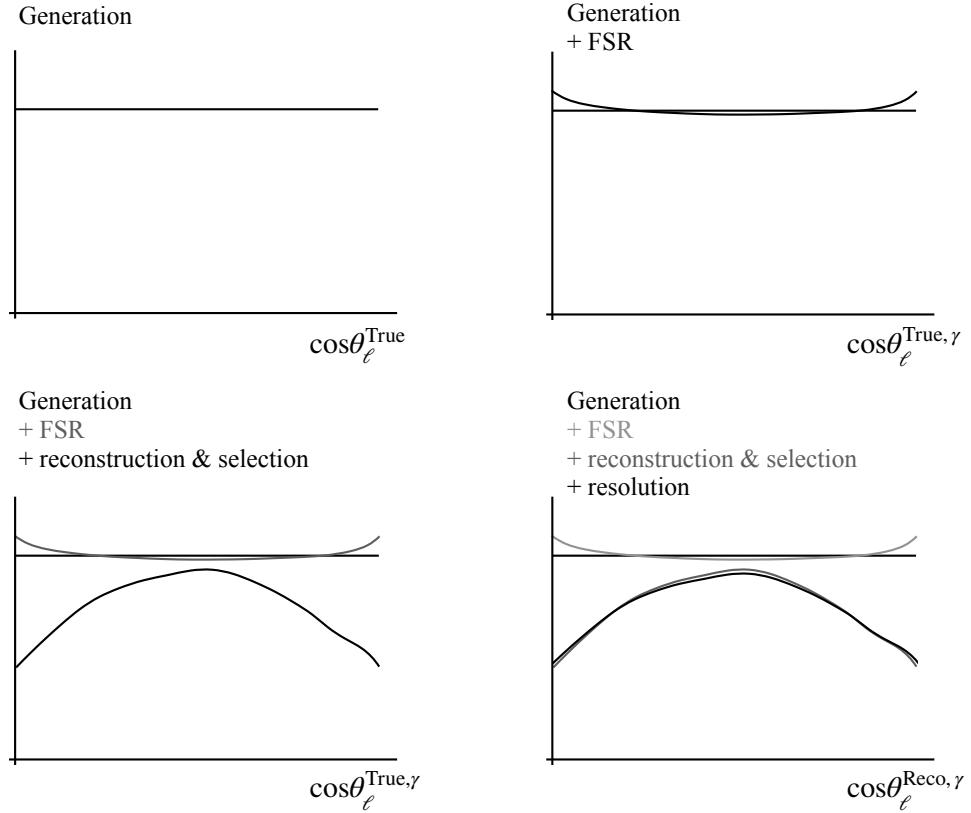


Figure 48: Illustration of key components relevant to the discussion of the acceptance strategy. The top left diagram shows the true distribution of an angle in a simulation that is uniform at generation (the situation is analogous for simulation generated with a given physics model). This is then corrected by PHOTOS to produce the modified distribution on the top right. The simulation of measurement effects (trigger, reconstruction and selection) results in a reduced number of events and the shaping of the distribution as illustrated in the lower left figure (acceptance effect), as well as a set of measured (rather than true) quantities, which are affected by resolution (lower right figure).

effect is not factorisable, and can be described by a convolution of the signal model with a kernel K that encapsulates the detector response,

$$\mathcal{P}_S^{\text{FSR, Acc, Reso}}(\Omega_{\text{Reco}}, q_{\text{Reco}}^2) = \int K(\Omega_{\text{Reco}}, q_{\text{Reco}}^2, \Omega_{\text{True}}, q_{\text{True}}^2) \times \mathcal{P}_S^{\text{FSR, Acc}}(\Omega_{\text{True}}, q_{\text{True}}^2) d\Omega_{\text{True}} dq_{\text{True}}^2. \quad (28)$$

Note that in this generalisation the resolution effects in the kernel are modelled using an eighth rank probability transition tensor. In this higher dimensional space there are several possible limitations; *e.g.* higher chance of un-populated bins due to limited statistics, matrix can become non-invertible, introduction of non-physics artefacts and naturally a significant higher computational demand in the angular fit. Therefore, it is interesting to examine the impact of the resolution effects alone and the possibility to model instead an *effective acceptance* including all elements of the detector response.

Table 17: Bin migration pattern for different data conditions and definitions of the q^2 . The percentage of events that either enter/leave the q^2 range from the left/right side of the interval, *i.e.* low/high index, is reported in the table alongside with the net effect of the samples, in which positive (negative) values indicates “net in” (“net out”). The M_{net} effect quoted in the table refers to the relative variation defined as $|M_{\text{net}} - N(\text{in} \rightarrow \text{in})|/N(\text{in} \rightarrow \text{in})$.

	Run-I				Run-II			
	q^2	q_c^2	q^2	q_c^2	q^2	q_c^2	q^2	q_c^2
$\%(in \rightarrow in)$	0.926 ± 0.011	0.927 ± 0.011	0.935 ± 0.010	0.930 ± 0.010	0.921 ± 0.006	0.921 ± 0.006	0.929 ± 0.005	0.923 ± 0.005
$\%(in \rightarrow out)_{[\text{low}]}$	0.028 ± 0.001	0.012 ± 0.001	0.022 ± 0.001	0.009 ± 0.001	0.025 ± 0.001	0.009 ± 0.000	0.020 ± 0.001	0.007 ± 0.000
$\%(in \rightarrow out)_{[\text{high}]}$	0.047 ± 0.002	0.062 ± 0.002	0.043 ± 0.002	0.060 ± 0.002	0.054 ± 0.001	0.069 ± 0.001	0.052 ± 0.001	0.069 ± 0.001
$\%(out \rightarrow in)_{[\text{low}]}$	0.007 ± 0.001	0.008 ± 0.001	0.008 ± 0.001	0.008 ± 0.001	0.007 ± 0.000	0.008 ± 0.000	0.007 ± 0.000	0.009 ± 0.000
$\%(out \rightarrow in)_{[\text{high}]}$	0.041 ± 0.001	0.009 ± 0.000	0.055 ± 0.001	0.013 ± 0.000	0.042 ± 0.000	0.010 ± 0.000	0.058 ± 0.001	0.013 ± 0.000
$M_{\text{net}}(\%)$	0.112 ± 0.003	-0.026 ± 0.001	0.118 ± 0.003	-0.024 ± 0.001	0.106 ± 0.002	-0.029 ± 0.001	0.112 ± 0.001	-0.033 ± 0.001

Figures 49 and 50 illustrate the resolution effects for q_c^2 and $\vec{\Omega}$ using 2016 conditions as reference. Additional information for the different datasets can be found in Appendix G. While in general a good resolution and no appreciable correlation is observed for the angular distributions, there is a sizeable effect on the q^2 distribution. Here is important to highlight the noticeable improvement in the q_c^2 definition, which in turn significantly reduces the migration of events. Table 17 reports the relative migration of events that are ($in \rightarrow in$) generated and reconstructed inside the considered q^2 interval, ($out \rightarrow in$) generated outside the interval but reconstructed inside and ($in \rightarrow out$) generated inside but falls outside. With these quantities is also possible to define a net effect, *i.e.* $M_{\text{net}} = N(in \rightarrow in) + N(out \rightarrow in) - N(in \rightarrow out)$, which clearly indicates a reduced migration of events with the usage of the q_c^2 ; as reference, a marginal negative net between [0.01, 0.2] % is observed for $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ analysis [7, 52]. As discussed in the following, these results already suggest that the impact of ignoring a full description of the resolution effects are expected to be small. Note that from Sec. 6.4 is evident that the impact on the angular observables is negligibly small compared to the statistical uncertainty reported in Sec. 8.5. Therefore, hereafter a simple *effective* approach is preferred for the acceptance modelling, which also includes both resolution and FSR effects as

$$\epsilon(\Omega_{\text{Reco}}, q_{\text{Reco}}^2) = \mathcal{P}_S^{\text{FSR, Acc, Reso}}(\Omega_{\text{Reco}}, q_{\text{Reco}}^2) / \mathcal{P}_S(\Omega_{\text{True}}, q_{\text{True}}^2), \quad (29)$$

where $\mathcal{P}_S^{\text{FSR, Acc, Reso}}(\Omega_{\text{Reco}}, q_{\text{Reco}}^2)$ is simply the events distribution obtained with reconstructed quantities of the post-selected simulation. While possible model-dependent effects in this *effective* treatment of the resolution may introduce a source of systematic uncertainty, considering the large expected statistical uncertainty of the measurement, the benefit of developing an accurate resolution matrix is not justified at this stage. Finally, this acceptance correction is made through the usage of per event weights obtained by taking the inverse of the output of the effective acceptance function for each event. Figure 51 depicts the several alternatives in the acceptance modelling studied in this analysis as discussed in the following sections.

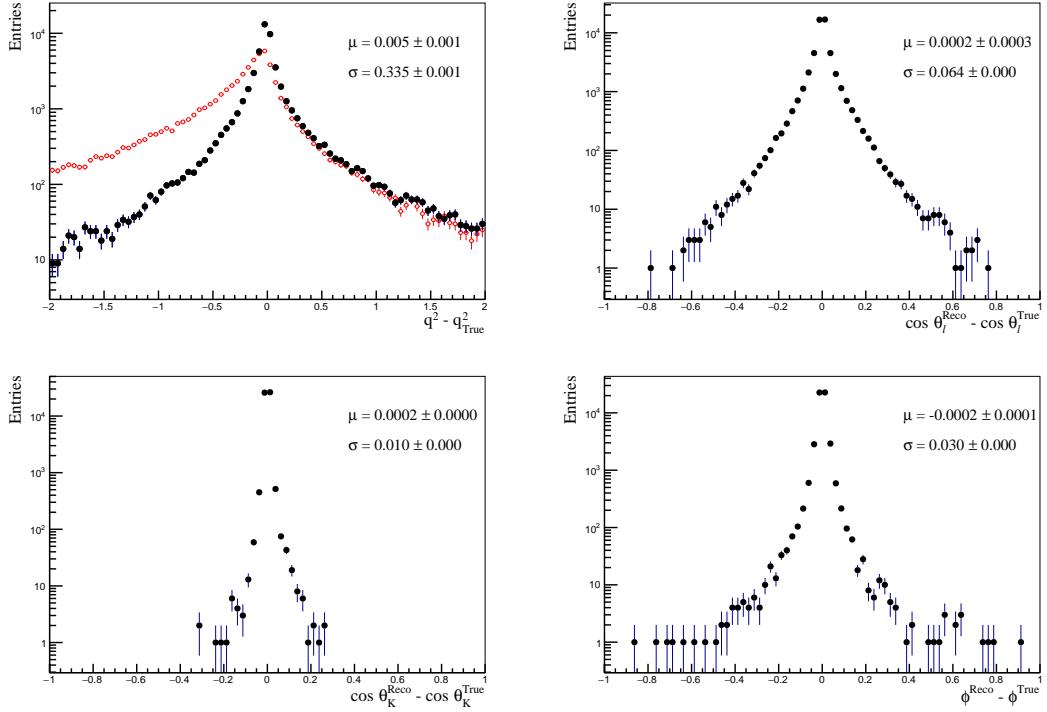


Figure 49: Resolution distributions defined as the difference between reconstructed and generator level variables for (top left) q^2 , (top right) $\cos \theta_\ell$, (bottom left) $\cos \theta_K$ and (bottom right) ϕ . Note that in the case of the q^2 distribution both (black) q_c^2 and (red) normal q^2 are overlaid for comparison.

6.2 Parametrisation

The acceptance function is parametrised using Legendre polynomials as a function of $\cos \theta_\ell, \cos \theta_K, \phi$ and q_c^2 without assuming any factorisation. Explicitly, the expression used is,

$$\epsilon(\cos \theta_\ell, \cos \theta_K, \phi, q_c^2) = \sum_{klmn} c_{klmn} L_k(\cos \theta_\ell) L_l(\cos \theta_K) L_m(\phi') L_n(q'^2), \quad (30)$$

where $L_a(x)$ is the Legendre polynomial of order a in the variable x , with x having values in the range $[-1, 1]$. The primed variables q'^2 and ϕ' in this formula highlight the fact that, while $\cos \theta_\ell$ and $\cos \theta_K$ are already in the desired range, a change of variables is needed for q_c^2 and ϕ . The following change of variables is applied to q_c^2 :

$$q_c^{2'} = \frac{2q_c^2 - q_{c(\min)}^2 - q_{c(\max)}^2}{q_{c(\max)}^2 - q_{c(\min)}^2}, \quad (31)$$

where $q_{c(\min)}^2$ and $q_{c(\max)}^2$ are 1.1 and 7.0 GeV^2/c^4 , respectively. The values of ϕ are instead transformed according to

$$\phi' = \frac{\phi}{\pi}, \quad (32)$$

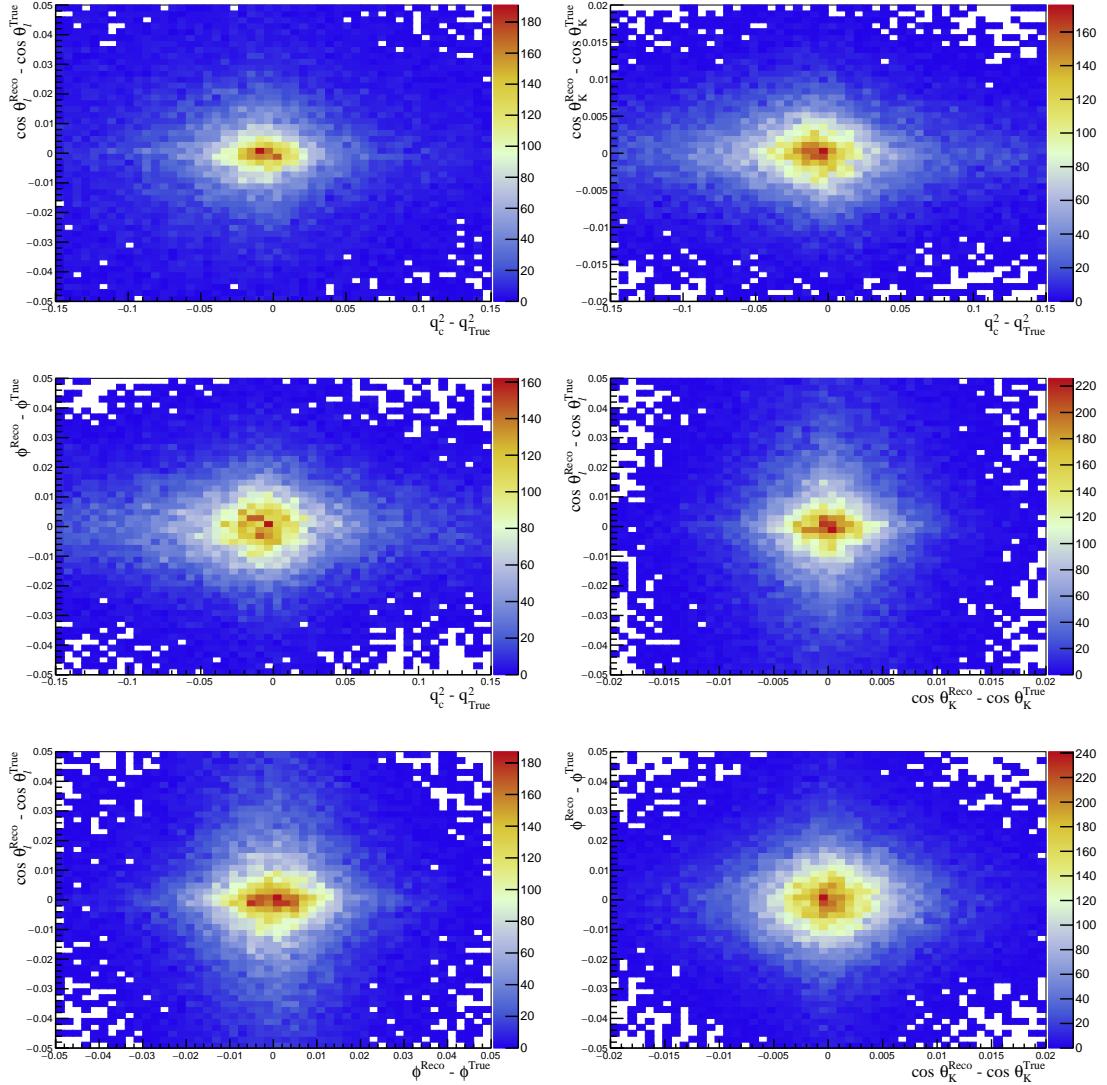


Figure 50: Two-dimensional resolution distributions defined as the difference between reconstructed and generator level variables for (top left) q^2 , (top right) $\cos \theta_\ell$, (bottom left) $\cos \theta_K$ and (bottom right) ϕ .

since ϕ is originally defined between $-\pi$ and π . The coefficients c_{klmn} are determined through a method of moments calculation. Making use of the orthogonality of the Legendre polynomials,

$$\int_{-1}^1 L_a(x) L_{a'}(x) dx = \frac{2}{2a+1} \delta_{aa'} , \quad (33)$$

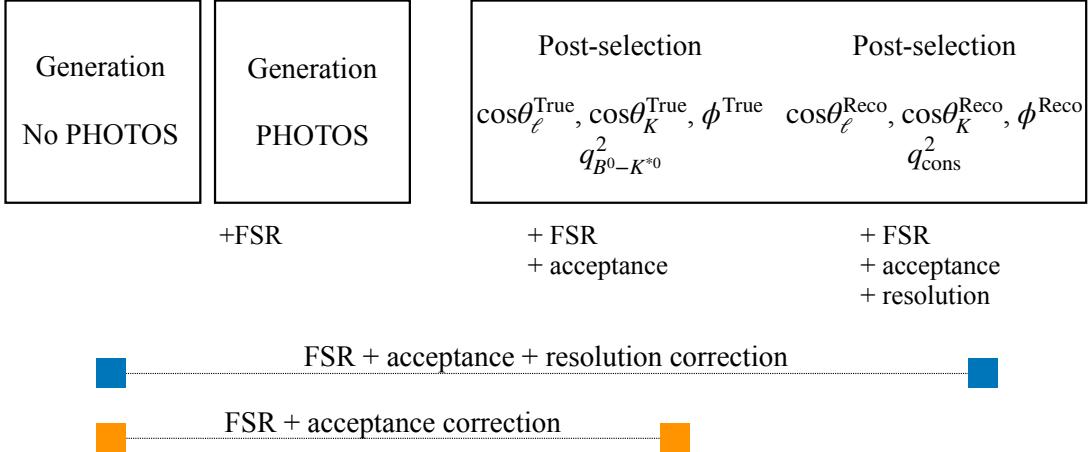


Figure 51: Schematic view of the quantities involved in the parametrisation of the effective acceptance function. Note that a comparison between blue and orange definitions can indicate the possible size of the effect in ignoring resolution effects.

1322 the equation for a specific coefficient, c_{klmn} , can be simplified to

$$c_{klmn} = \frac{1}{N^{(i)}} \sum_{i=1}^N (\omega_i) \left[L_k(\cos\theta_{li}) L_l(\cos\theta_{ki}) L_m(\phi_i) L_n(q_i^2) \right. \\ \left. \times \left(\frac{2}{2k+1} \right) \left(\frac{2}{2l+1} \right) \left(\frac{2}{2m+1} \right) \left(\frac{2}{2n+1} \right) \right],$$

1323 where the summation is over the contributions of all events, N , of the sample to be
1324 parametrised. In the absence of per-event weights, ω_i , $N' = N$. Otherwise

$$N' = \sum_{i=1}^N \omega_i . \quad (34)$$

1325 The entries of the covariance matrix associated with the determination of the c_{klmn}
1326 coefficients are given by

$$\text{cov}(c_{klmn}, c_{k'l'm'n'}) = \frac{1}{\sum_i \omega_i} \times \frac{1}{\sum_i \omega_i - 1} \left[\sum_i \omega_i (c_{klmn,i} - c_{klmn})(c_{k'l'm'n',i} - c_{k'l'm'n'}) \right], \quad (35)$$

1327 where the $1/\sum_i \omega_i$ term is multiplied to the corrected weighted sample standard deviation
1328 in order to obtain the standard error of the sample mean (the sample mean refers to the
1329 calculated c_{klmn} coefficients).

1330 The full correction is applied by weighting each event with $\cos\theta_{\ell_i}, \cos\theta_{K_i}, \phi_i, q_i^2$ by the
1331 inverse of the output of $\epsilon(\cos\theta_{\ell_i}, \cos\theta_{K_i}, \phi_i, q_i^2)$. The effectiveness of this correction can
1332 be validated by checking the consistency between angular observables obtained from a
1333 fit to the generation level distribution, and those obtained from a weighted fit to the
1334 post-selection simulation.

1335 6.3 Sample choice

1336 The acceptance function can be parametrised from two types of samples:

- 1337 1. simulation generated with physics content (**PHYS**); and
- 1338 2. simulation generated flat in the angles and q^2 (**FlatQ2**).

1339 In the case of the **PHYS** acceptance, the method of moments calculation is performed twice
1340 – first using generator-level samples to parametrise the angular and q^2 distributions without
1341 the effects of selection and reconstruction, ϵ_{gen} , and then using the post-selection samples
1342 with per event weights of $1/\epsilon_{gen}$ to produce the final acceptance function, ϵ . In the case
1343 of the **FlatQ2** acceptance, the calculation in principle needs to only be applied to the
1344 post-selection distributions because the generation distribution is flat. However, as the q^2
1345 distribution of the **FlatQ2** samples deviates from flatness slightly due to phase space limit
1346 and K^{*0} width, a one-dimensional parametrisation is performed on the q^2 distribution of
1347 the generator-level sample, and the second, final parametrisation is made with the inverse
1348 weights from this function.

1349 There are two aspects that differ the acceptances obtained from these samples. The first
1350 feature is related to the size of the available simulations and the underlying physics dynamics
1351 utilised at the generation level, as illustrated in Fig. 52. The statistics of the **FlatQ2** samples
1352 is a factor of 2-3 higher than that of the **PHYS**. Moreover, the uniform population of events
1353 across the phase space allows for a good estimation of acceptance/efficiency throughout,
1354 leading to fewer problematic regions where the parametrization of the acceptance function
1355 might become negative due to statistical fluctuation in the distribution of events. For
1356 instance, assuming a uniform sample, the fraction of events landing in a negative region
1357 of the acceptance function is at the level of 0.01 % and 0.0007 % for **PHYS** and **FlatQ2**
1358 acceptances, respectively. Another advantage of use of the **FlatQ2** acceptance as nominal
1359 choice is the reduction of the frequency of very large correction weights (see discussion in
1360 Sec. 8), which occur, for example, when events land near one of these negative regions
1361 where the acceptance function is close to zero (the inverse weight is then large). In any
1362 case, both factors suggest that a more stable parametrisation of the acceptance effects
1363 is achieved with the **FlatQ2** sample, and any remaining effect due to this behaviour is
1364 investigated in the systematics.

1365 The second key difference in the available simulation is related to effective modelling of
1366 the convolution of the resolution (and FSR) with the underlying physics model. In contrast
1367 to the pure detector acceptance, which in principle is the same for both simulations,
1368 the strategy to use an effective approach to the resolution corrections leads to model-
1369 dependences. As the **PHYS** model is more aligned to the true underlying model of the

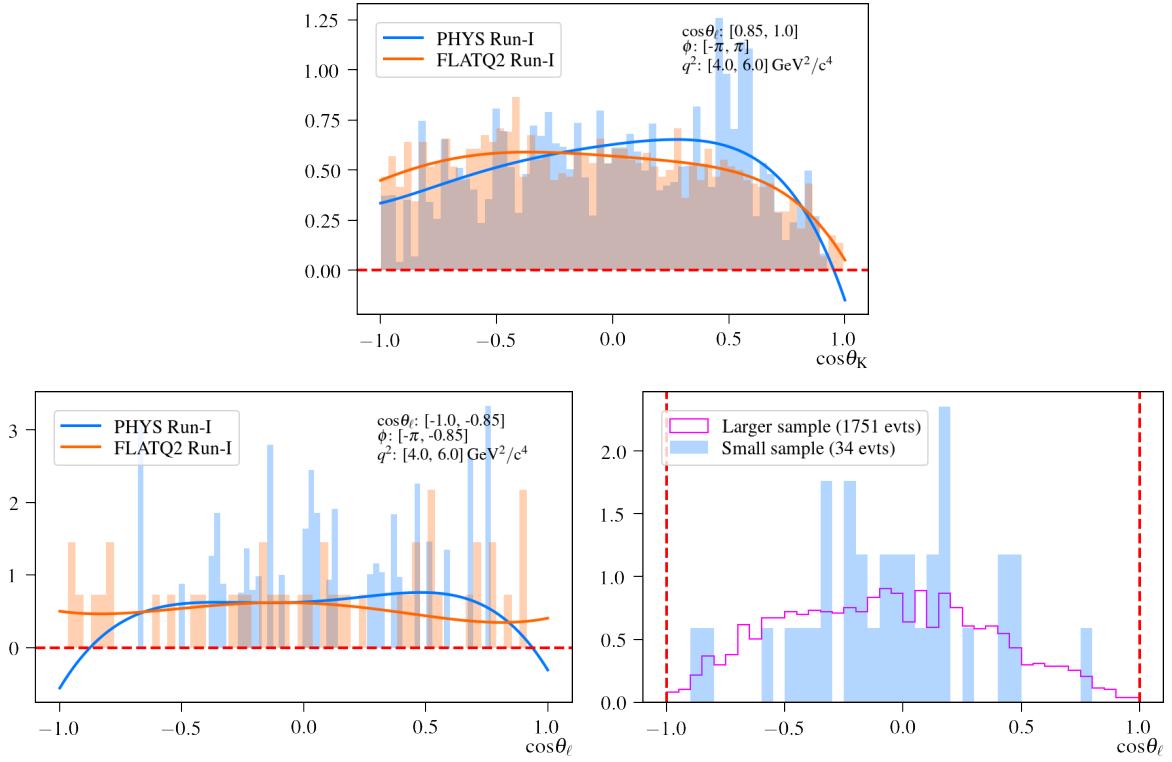


Figure 52: Phase-space that illustrate the two types of negative regions in the acceptance function. While the (top) first type corresponds to a deficit of events in a region of low efficiency (near $\cos\theta_K = 1$), the (bottom left) second type shows the PHYS acceptance with a lack of events near both edges of $\cos\theta_\ell$. Note that in this case the efficiency is not necessarily low, but the regions are sparsely populated due to the shape of the physics PDF. This is depicted in the bottom right figure, where a small sample with the same number of events (34) is drawn at random from a larger signal only sample of the same region. Both types of issue are reduced with increased statistics.

1370 data, it is expected to provide a more accurate resolution correction. This is checked
 1371 by applying two pairs of acceptances to the post-selection PHYS simulation, obtained
 1372 respectively from FlatQ2 and PHYS samples either from true or reconstructed quantities.
 1373 Both are used to correct the quantities that are used in parametrisation, that is, the
 1374 acceptances parametrised with true/reconstructed quantities are applied to the matching
 1375 true/reconstructed quantities of the sample. Tables 18 and 19 report the fit results for
 1376 each of these options for the S_i and P_i observables, respectively, indicating good retrieval
 1377 of generator level values for both acceptances parametrised with, and applied to, true
 1378 quantities. When resolution effects are included, the difference between the PHYS and
 1379 the FlatQ2 acceptances becomes more prominent as expected, with the former showing
 1380 better performance than the latter. In particular, notice that the S_5 and A_{FB} observables
 1381 are shifted by 0.01 of the generated values. Nevertheless, the difference between the two
 1382 acceptances is marginal when compared against the expected statistical uncertainty of the

1383 data fit.

1384 Both the **PHYS** and **FlatQ2** samples are possible choices for the parametrisation of the
1385 acceptance function. While the former is expected to provide more accurate resolution
1386 corrections, the latter results in smoother and more well-behaved acceptance functions.
1387 Considering that the potential systematic resulting from the **FlatQ2** correction is expected
1388 to be small and easily quantified by taking the difference between the two acceptances
1389 using pseudoexperiments, the pathological behaviours exhibited more frequently in the
1390 **PHYS** acceptance (impact of large weights) are more difficult to estimate, and thus the
1391 **FlatQ2** acceptance is chosen as the nominal acceptance. The **PHYS** acceptance is kept as
1392 an alternative for cross-check purposes and systematics evaluation.

1393 6.4 Nominal acceptance

1394 The acceptance is modelled using the lowest orders of the polynomial that shows good
1395 description of the angular and q^2 distributions. Since there are problematic regions where
1396 the acceptance function may become very small or reach negative values around the edges
1397 of the phase space, the q^2 range is extended from the nominal $q^2 \in [1.1, 7.0]$ to $q^2 \in [0.1,$
1398 $9.0] \text{GeV}^2/\text{c}^4$ in order to reduce the statistical fluctuations of the sample. Weights are used
1399 to correct deviations from flatness due to phase space limit. These are obtained by taking
1400 the inverse of the one dimensional parametrisation of the **FlatQ2** generator-level true q^2
1401 (q^2 calculated from the difference of the four-vectors of the B and that of the K^* , using
1402 true quantities) in the range of $q^2 \in [0.1, 15.0] \text{GeV}^2/\text{c}^4$, with a Legendre polynomial of
1403 order two or less, as depicted in Fig. 53. The corrected q^2 and angular distributions of
1404 the post-selection **FlatQ2** samples are then modelled using Legendre polynomials; order
1405 three or less is used for q^2 , five or less for $\cos\theta_K$, four or less for $\cos\theta_\ell$ and six or less for ϕ .
1406 A total of 840 coefficients is used for the parametrisation of the post-selection samples.
1407 The separation of the simulation into different configurations, *i.e.* splitting by year and L0
1408 categories, year only, Runs and L0 categories, and Runs only, have been studied in order to
1409 examine the possible impact of the selection and parametrisation of the acceptance on the
1410 observables of interest. Table 20 reports the fit results for the S_i and P_i basis, which shows
1411 marginal difference between the several scenario. Therefore, the acceptance separating by
1412 Runs alone have been chosen as nominal parametrisation since it has the largest statistics.
1413 As a systematic uncertainty the nominal data fit will be repeated with alternative settings.
1414 The statistics available for the one dimensional generator-level parametrisation is 1659245
1415 events. The final post-selection samples used contain 82654 and 188588 events for **Run-I**
1416 and **Run-II**, respectively. The kinematic and trigger correction weights are included in
1417 the parametrisation of acceptances used for data fits. One-dimensional projections of the
1418 final **FlatQ2** acceptances functions including all weights are shown in Figs. 54, 55 and 56
1419 respectively for **Run-I**, **Run-II** and a comparison between the two scenario, in which a
1420 good agreement is observed in each of the projections.

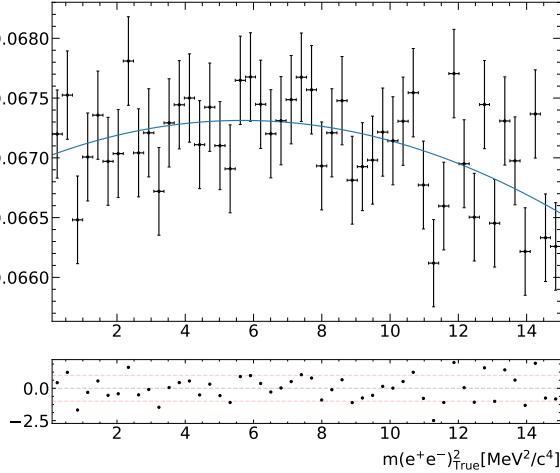


Figure 53: One-dimensional parametrisation of the signal generation level q^2 distribution for flatness correction.

6.5 Alternative acceptance

The alternative acceptance function obtained from the rare mode physics simulations (PHYS acceptance) is used for the evaluation of systematic uncertainties only. For the generator-level, Legendre polynomials of order twelve or less is used for q^2 (explicitly q^2 calculated from the difference of the four-vectors of the B and that of the K^* to avoid final state radiation effects). An extended q^2 range of $0.5 < q_{B-K^*}^2 < 10.0 \text{ GeV}^2/\text{c}^4$ is used. For the decay angles, orders of five or less is used for $\cos\theta_K$, four or less for $\cos\theta_\ell$, and eight or less or ϕ . This amounts to a total of 3510 coefficients. For the post-selection samples, the same polynomial orders are used as in the case of the nominal acceptance. However, the q^2 range is kept to the standard range of $q^2 \in [1.1, 7.0] \text{ GeV}^2/\text{c}^4$. The statistics available for the generator-level parametrisation is 3127699 events, and that which is available for the final parametrisation of the post-selection distributions are 25097 and 59681 for Run-I and Run-II, respectively. The one dimensional projections for the generator-level and post-selection parametrisations for each run, are shown in Figs. 57, 58, 59.¹¹ For brevity, a comparison between this alternative acceptance and the nominal one is shown for Run-I and II in Figs. 60 and 61, respectively. Similarly to the nominal acceptance, a good agreement is observed between the parametrisation and the weighted simulation.

¹¹There is a clear discontinuity at $q^2/m_b^2 = 0.25$, which is a known effect related to the way the effective WC $C_{7,9}$ at the NNLO (*i.e.* including hadronic corrections to the Wilson coefficients) are parametrised in EvtGen - based on the implementation reported in Ref. [63].

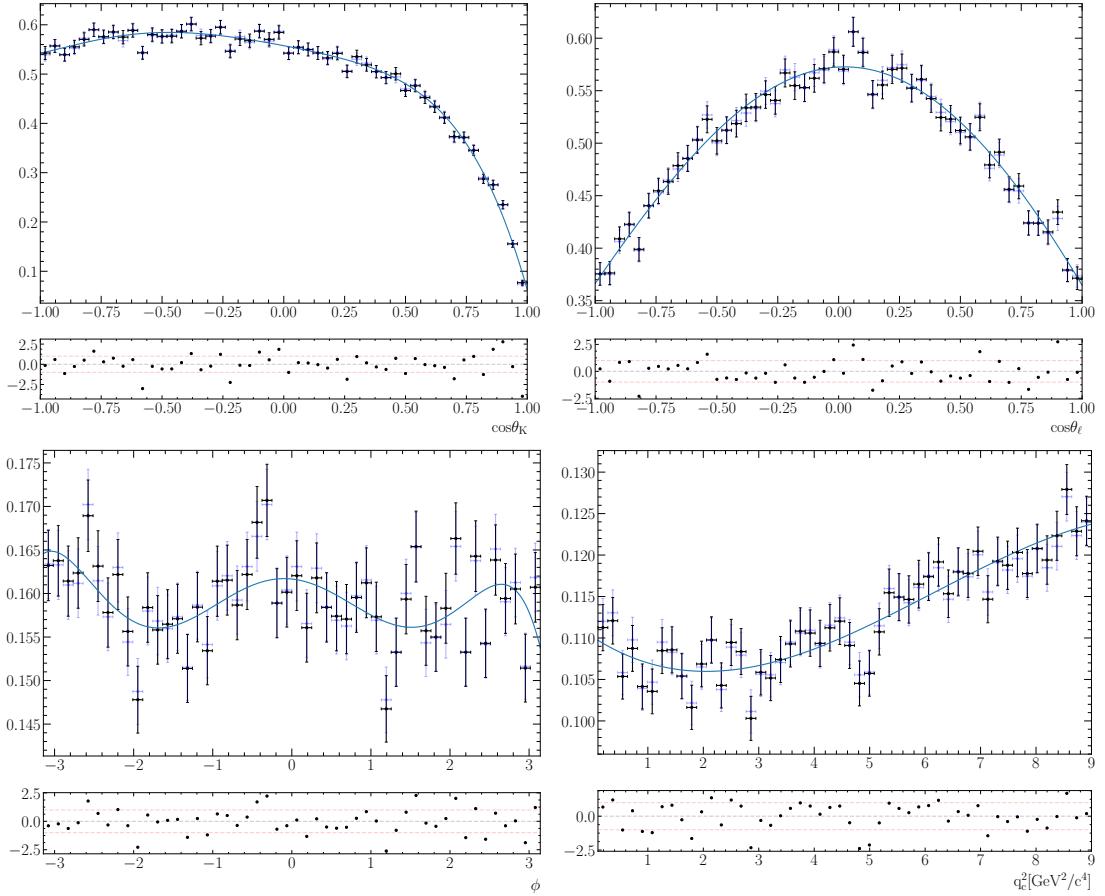


Figure 54: One-dimensional projections of the four-dimensional parametrisation of post-selection FLATQ2 simulation, with kinematic, trigger correction and q^2 flatness correction weights, for Run-I simulation. The blue-dotted points correspond to the unweighted simulation whereas the black-dotted ones refer to the weighted simulation.

	Generation			
	F_L	S_3	S_4	S_5
	0.7066 ± 0.0005	-0.0097 ± 0.0007	-0.1483 ± 0.0008	-0.2037 ± 0.0008
	A_{FB}			
	0.0098 ± 0.0005	0.0124 ± 0.0008	0.0031 ± 0.0008	0.0013 ± 0.0007

	PHYS			
	Acceptance +FSR		Acceptance +FSR +resolution	
F_L	0.7065 ± 0.0034	0.0001	0.7048 ± 0.0031	0.0018
S_3	-0.0061 ± 0.0028	-0.0036	-0.0071 ± 0.0030	-0.0026
S_4	-0.143 ± 0.006	-0.0056	-0.151 ± 0.005	0.0025
S_5	-0.206 ± 0.005	0.0021	-0.208 ± 0.004	0.0040
A_{FB}	0.0107 ± 0.0031	-0.0009	0.0150 ± 0.0027	-0.0052
S_7	0.012 ± 0.004	0.0004	0.012 ± 0.004	0.0003
S_8	0.008 ± 0.005	-0.0049	0.006 ± 0.005	-0.0034
S_9	0.0053 ± 0.0028	-0.0040	0.0051 ± 0.0028	-0.0039

	FLATQ2			
	Acceptance +FSR		Acceptance +FSR +resolution	
F_L	0.7085 ± 0.0028	-0.0018	0.7077 ± 0.0028	-0.0010
S_3	-0.0064 ± 0.0027	-0.0033	-0.0056 ± 0.0027	-0.0041
S_4	-0.145 ± 0.004	-0.0032	-0.148 ± 0.004	-0.0001
S_5	-0.208 ± 0.004	0.0044	-0.216 ± 0.004	0.0121
A_{FB}	0.0129 ± 0.0026	-0.0031	0.0216 ± 0.0025	-0.0118
S_7	0.007 ± 0.004	0.0056	0.006 ± 0.004	0.0068
S_8	0.009 ± 0.005	-0.0057	0.007 ± 0.004	-0.0042
S_9	-0.0018 ± 0.0027	0.0030	-0.0021 ± 0.0027	0.0034

Table 18: Results of acceptance validation tests for the S_i -basis observables performed with acceptance functions obtained from rare mode physics (PHYS) and flat in the angles and q^2 (FlatQ2) simulation. In each case, three different functions are parametrised, and tested. These range from the pure acceptance function, to the full nominal function that corrects for acceptance and resolution effects. The observable values (and differences with respect to generation level results), are shown for each case. Minor differences between observables retrieved using the two acceptances are seen in the case of the full correction. This is understood to be mainly due to the model-dependency of the approximate correction.

	Generation			
	F_L	P_1	P'_4	P'_5
	0.7066 ± 0.0005	-0.066 ± 0.005	-0.3256 ± 0.0017	-0.4474 ± 0.0017
	P_2		0.0224 ± 0.0010	
	P'_6		0.0273 ± 0.0018	
	P'_8		0.0068 ± 0.0018	
	P_3		-0.0043 ± 0.0023	

	PHYS			
	Acceptance +FSR	Acceptance +FSR	Acceptance +FSR +resolution	Acceptance +resolution
F_L	0.7065 ± 0.0034	0.0001	0.7049 ± 0.0031	0.0018
P_1	-0.041 ± 0.019	-0.0248	-0.048 ± 0.020	-0.0179
P'_4	-0.313 ± 0.013	-0.0123	-0.330 ± 0.010	0.0048
P'_5	-0.452 ± 0.010	0.0046	-0.455 ± 0.009	0.0080
P_2	0.024 ± 0.007	-0.0020	0.034 ± 0.006	-0.0116
P'_6	0.026 ± 0.010	0.0008	0.027 ± 0.009	0.0007
P'_8	0.018 ± 0.012	-0.0108	0.014 ± 0.011	-0.0074
P_3	-0.018 ± 0.010	0.0137	-0.017 ± 0.009	0.0130

	FLATQ2			
	Acceptance +FSR	Acceptance +FSR	Acceptance +FSR +resolution	Acceptance +resolution
F_L	0.7085 ± 0.0028	-0.0018	0.7077 ± 0.0028	-0.0010
P_1	-0.044 ± 0.018	-0.0225	-0.039 ± 0.018	-0.0277
P'_4	-0.319 ± 0.010	-0.0063	-0.326 ± 0.009	0.0002
P'_5	-0.458 ± 0.008	0.0105	-0.474 ± 0.008	0.0270
P_2	0.029 ± 0.006	-0.0071	0.049 ± 0.006	-0.0270
P'_6	0.015 ± 0.009	0.0122	0.012 ± 0.008	0.0149
P'_8	0.019 ± 0.010	-0.0125	0.016 ± 0.009	-0.0092
P_3	0.006 ± 0.009	-0.0104	0.007 ± 0.009	-0.0116

Table 19: Results of acceptance validation tests for the P_i -basis observables performed with acceptance functions obtained from rare mode physics (PHYS) and flat in the angles and q^2 (FlatQ2) simulation. In each case, three different functions are parametrised, and tested. These range from the pure acceptance function, to the full nominal function that corrects for acceptance and resolution effects. The observable values (and differences with respect to generation level results), are shown for each case. Minor differences between observables retrieved using the two acceptances are seen in the case of the full correction. This is understood to be mainly due to the model-dependency of the approximate correction.

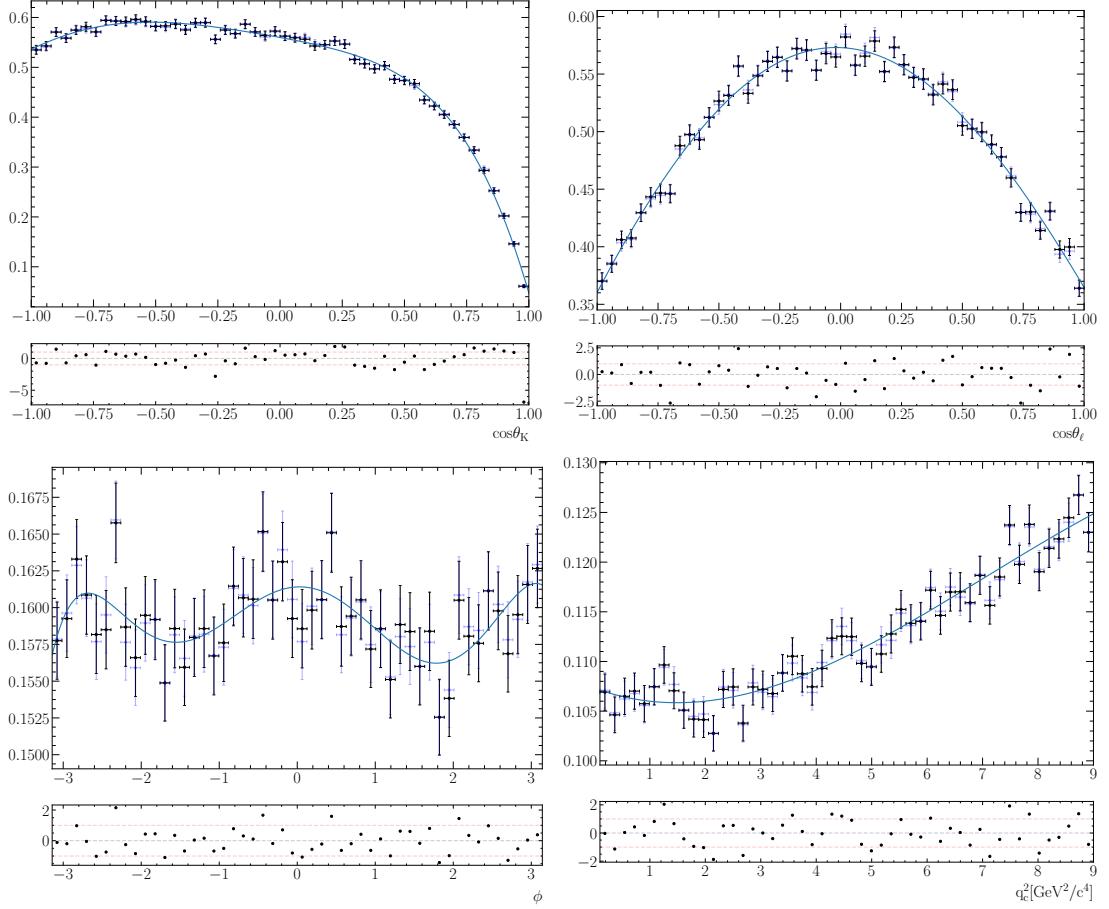


Figure 55: One-dimensional projections of the four-dimensional parametrisation of post-selection FLATQ2 simulation, with kinematic, trigger correction and q^2 flatness correction weights, for Run-II simulation. The blue-dotted points correspond to the unweighted simulation whereas the black-dotted ones refer to the weighted simulation.

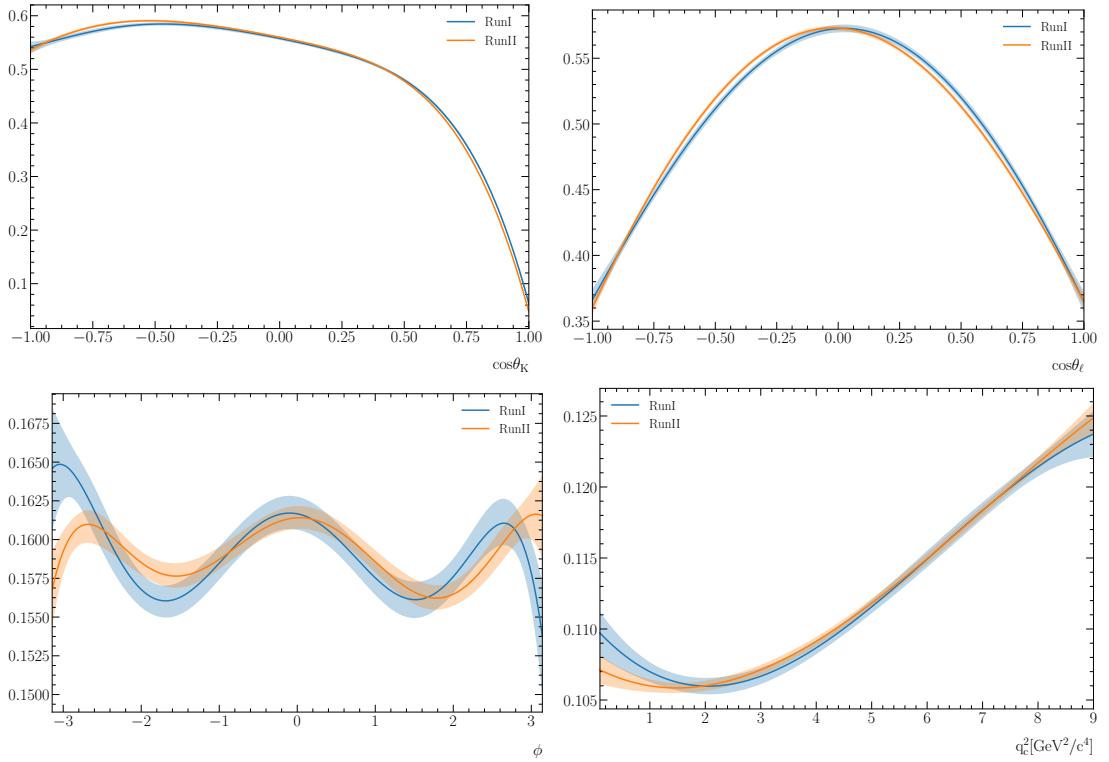


Figure 56: Projections of the acceptance function for Run-I and Run-II. The widths of the band is given by the standard deviation of the values taken from 500 alternative sets of coefficients, as obtained via the covariance matrix from the method of moments calculation.

	Year-L0		Year	
F_L	0.7117 ± 0.0030	-0.0051	0.7094 ± 0.0028	-0.0028
S_3	-0.0029 ± 0.0027	-0.0068	-0.0050 ± 0.0027	-0.0047
S_4	-0.149 ± 0.005	0.0004	-0.150 ± 0.004	0.0014
S_5	-0.213 ± 0.004	0.0098	-0.215 ± 0.004	0.0115
A_{FB}	0.0225 ± 0.0026	-0.0126	0.0216 ± 0.0024	-0.0118
S_7	0.007 ± 0.004	0.0058	0.005 ± 0.004	0.0074
S_8	0.005 ± 0.004	-0.0024	0.007 ± 0.004	-0.0038
S_9	-0.0024 ± 0.0027	0.0036	-0.0019 ± 0.0027	0.0032
	Run-L0		Run	
F_L	0.7094 ± 0.0027	-0.0027	0.7077 ± 0.0028	-0.0010
S_3	-0.0049 ± 0.0027	-0.0049	-0.0056 ± 0.0027	-0.0041
S_4	-0.150 ± 0.004	0.0020	-0.148 ± 0.004	-0.0001
S_5	-0.216 ± 0.004	0.0124	-0.216 ± 0.004	0.0121
A_{FB}	0.0228 ± 0.0023	-0.0130	0.0216 ± 0.0025	-0.0118
S_7	0.007 ± 0.004	0.0054	0.006 ± 0.004	0.0068
S_8	0.004 ± 0.004	-0.0013	0.007 ± 0.004	-0.0042
S_9	-0.0023 ± 0.0027	0.0035	-0.0021 ± 0.0027	0.0034
	Year-L0		Year	
F_L	0.7117 ± 0.0030	-0.0051	0.7094 ± 0.0028	-0.0028
P_1	-0.020 ± 0.019	-0.0461	-0.035 ± 0.018	-0.0316
P'_4	-0.328 ± 0.010	0.0026	-0.330 ± 0.009	0.0041
P'_5	-0.471 ± 0.009	0.0239	-0.474 ± 0.008	0.0266
P_2	0.052 ± 0.006	-0.0297	0.050 ± 0.006	-0.0273
P'_6	0.015 ± 0.009	0.0127	0.011 ± 0.008	0.0163
P'_8	0.012 ± 0.009	-0.0053	0.015 ± 0.009	-0.0083
P_3	0.008 ± 0.009	-0.0125	0.007 ± 0.009	-0.0110
	Run-L0		Run	
F_L	0.7094 ± 0.0027	-0.0027	0.7077 ± 0.0028	-0.0010
P_1	-0.033 ± 0.018	-0.0329	-0.039 ± 0.018	-0.0277
P'_4	-0.331 ± 0.009	0.0053	-0.326 ± 0.009	0.0002
P'_5	-0.476 ± 0.008	0.0285	-0.474 ± 0.008	0.0270
P_2	0.052 ± 0.005	-0.0300	0.049 ± 0.006	-0.0270
P'_6	0.016 ± 0.008	0.0117	0.012 ± 0.008	0.0149
P'_8	0.010 ± 0.009	-0.0030	0.016 ± 0.009	-0.0092
P_3	0.008 ± 0.009	-0.0121	0.007 ± 0.009	-0.0116

Table 20: Results of rare mode acceptance validation tests performed for four different choices: assigning correction weights based on year and L0 category; year only; Run and L0 category and Run only. While the best retrieval of generator-level values is achieved by assigning weights based on Run and L0 category, the Run split is chosen for greater stability, as the further splitting of the sample used for parametrisation increases the frequency of problematic correction weights.

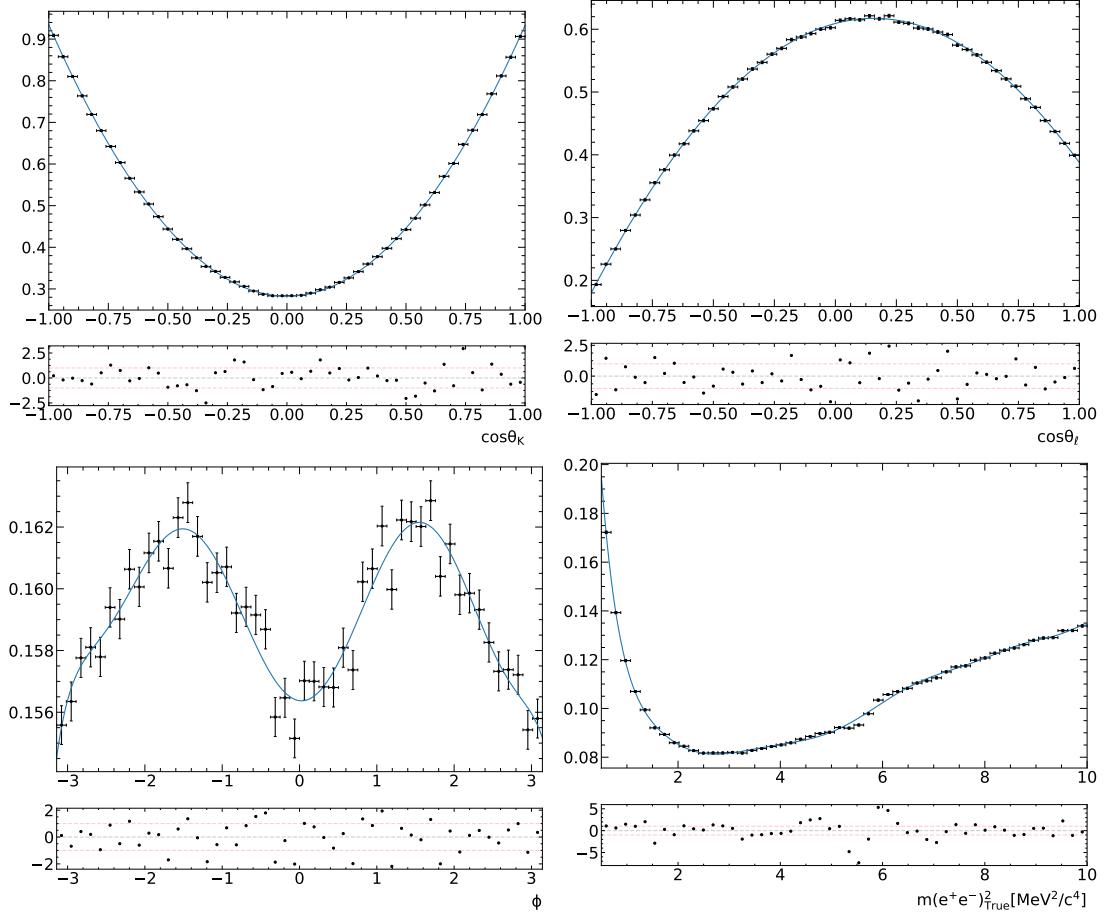


Figure 57: One-dimensional projections of the four-dimensional parametrisation of the PHYS generation level q^2 and angular distributions.

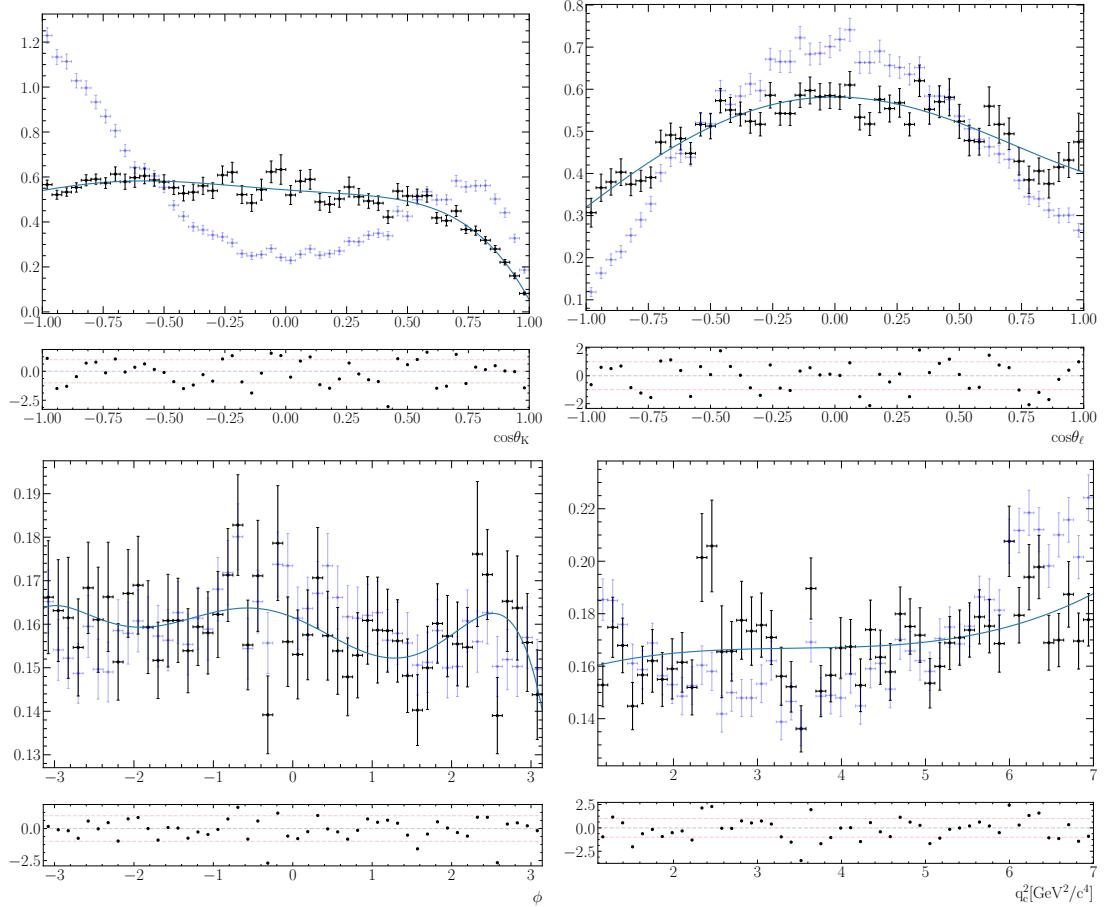


Figure 58: One-dimensional projections of the four-dimensional parametrisation of post-selection physics simulation, with per-event weights from the inverse of the function used to parametrise the generator-level sample, as well as kinematic and trigger correction weights, for Run-I simulation. The blue-dotted points correspond to the unweighted simulation whereas the black-dotted ones refer to the weighted simulation.

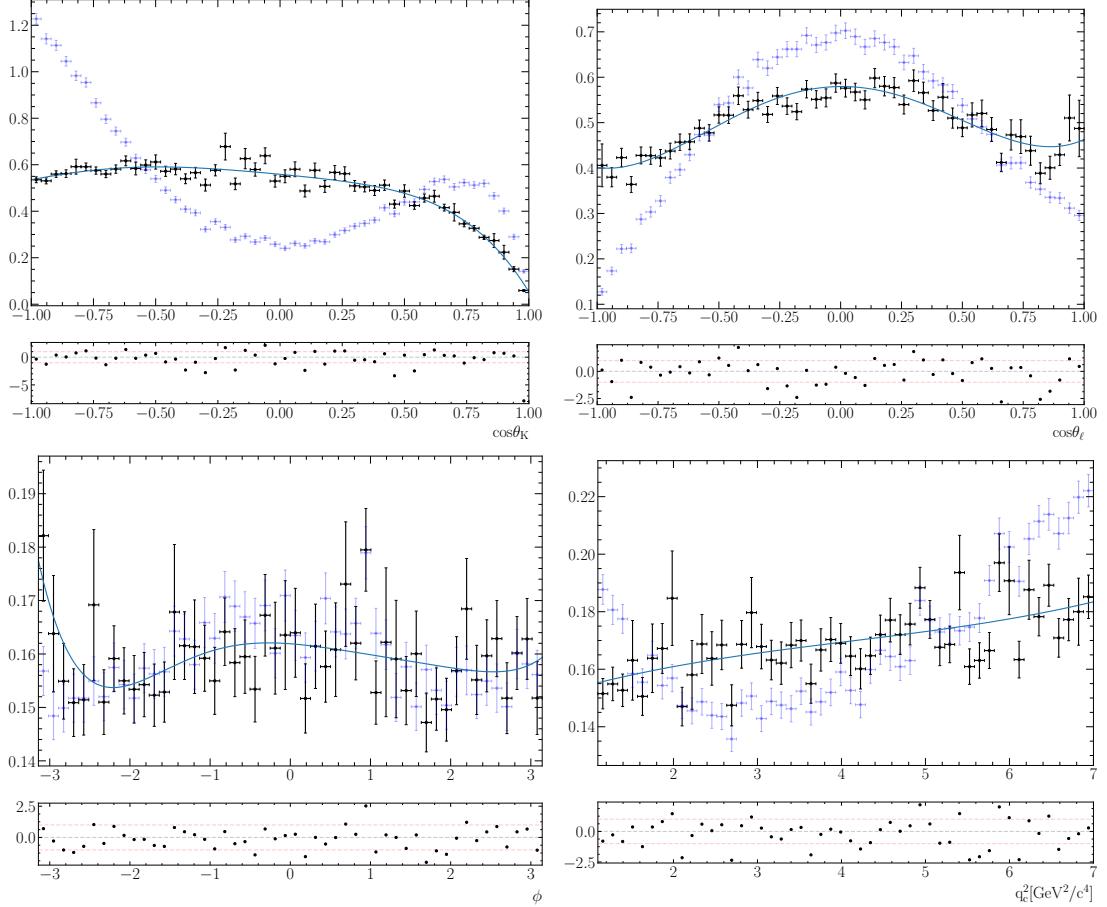


Figure 59: One-dimensional projections of the four-dimensional parametrisation of post-selection physics simulation, with per-event weights from the inverse of the function used to parametrise the generator-level sample, as well as kinematic and trigger correction weights for Run-I simulation. The blue-dotted points correspond to the unweighted simulation whereas the black-dotted ones refer to the weighted simulation.

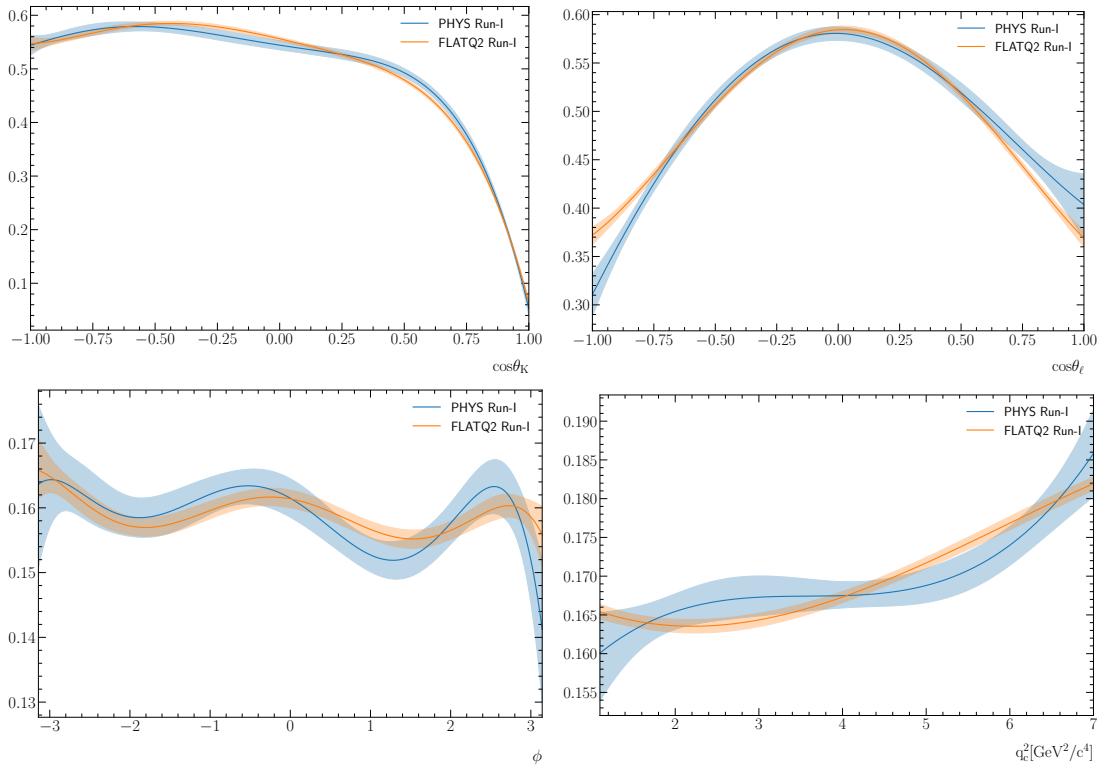


Figure 60: Projections of PHYS and FLATQ2 acceptances for Run-I simulation. For comparison, the FLATQ2 acceptance has been integrated in q^2 to match the range of the PHYS.

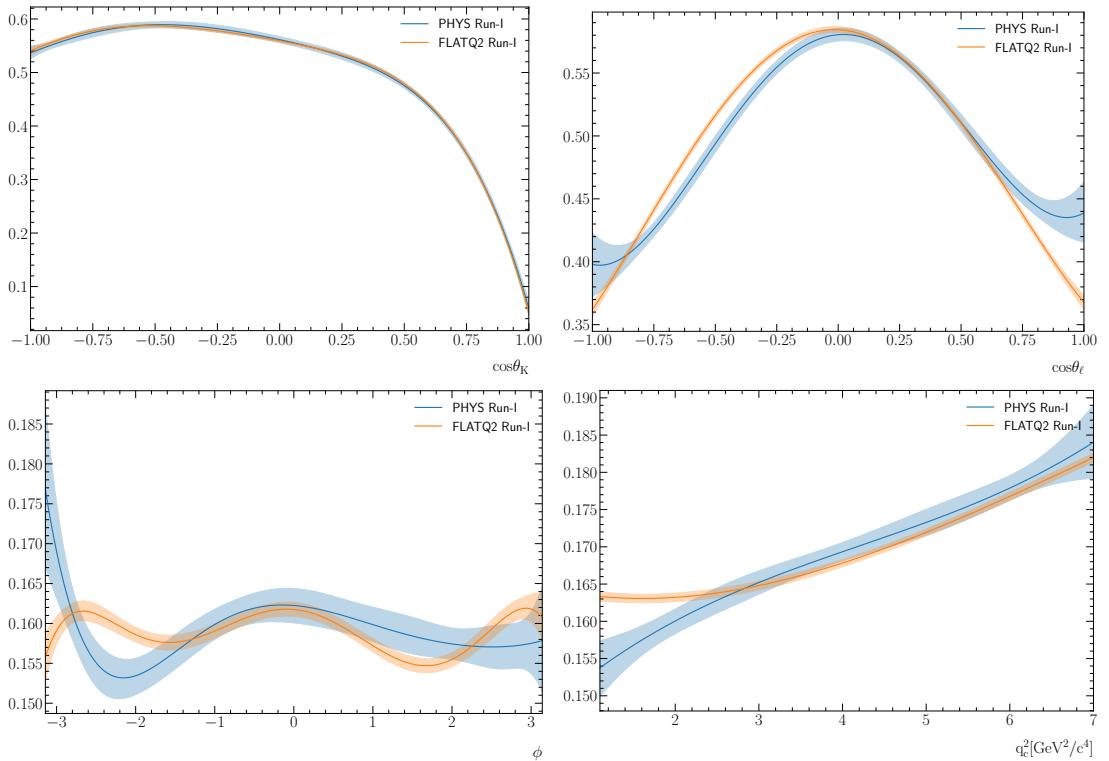


Figure 61: Projections of PHYS and FLATQ2 acceptances for Run-II simulation. For comparison, the FLATQ2 acceptance has been integrated in q^2 to match the range of the PHYS.

1438 7 Background model

The scripts used in the modelling of the different sources of backgrounds as well as validation studies can be found in:

- 1439 ○ `ewp-bd2ksteeangular-central-q2/scripts/background/`

1440
1441 The backgrounds that contribute to the rare mode spectra are modelled assuming no
1442 correlation between the B^0 reconstructed mass and decay angles as

$$\mathcal{P}_{bg}(m_{K\pi ee}, \cos\theta_\ell, \cos\theta_K, \phi) = \mathcal{P}(m_{K\pi ee}) \times \mathcal{P}(\cos\theta_K, \cos\theta_\ell, \phi), \quad (36)$$

1443 where $\mathcal{P}(m_{K\pi ee})$ may be an exponential PDF or a non-parametric kernel density estimator
1444 (KDE), and $\mathcal{P}(\cos\theta_K, \cos\theta_\ell, \phi)$ is typically given by

$$\mathcal{P}(\cos\theta_K, \cos\theta_\ell, \phi) = \left[\sum_{i=0} c_i T_i(\cos\theta_\ell) \right] \times \left[\sum_{j=0} c_j T_j(\cos\theta_K) \right] \times \left[\sum_{k=0} c_k T_k(\cos\phi) \right], \quad (37)$$

1445 where T_m denotes Chebyshev polynomial of order m , and the c_m are its associated
1446 coefficients. In non-trivial angular distributions where a parametrisation with Chebyshev
1447 polynomials is challenging, the KDE is preferred as discussed below.

1448 As the acceptance correction for the signal is carried out through the application of
1449 per-event weights, all backgrounds are unavoidably affected in the process. The changes
1450 to the shapes of the background angular distributions caused by these weights need to
1451 be taken into account. This is achieved via the inclusion of acceptance weights in the
1452 background modelling as well.

1453 In the following, the different background sources and corresponding parametrisation
1454 are discussed, *i.e.* combinatorial (Sec. 7.1), double-semileptonic (DSL) decays (Sec. 7.2)
1455 and partially-reconstructed backgrounds (Sec. 7.4). In particular, the modelling of the
1456 DSL contribution is among the most challenging aspects in this analysis. A dedicated
1457 data-driven step-wise procedure has been designed to obtain such contribution from the
1458 lepton-flavour-violating cousin channel. Since this approach cannot fully disentangle
1459 some combinatorial-like contribution to the DSL modelling, the combinatorial background
1460 lineshape is also simultaneously derived from this procedure. Further details of this
1461 methodology are presented in Sec. 7.3. Finally, the background contributions to the
1462 invariant mass of the control channel are briefly introduced in Sec. 7.5.

1463 7.1 Combinatorial

1464 The combinatorial background consists of events pieced together from fragments of multiple
1465 decays of the same pp collision. It is exponentially distributed in the $K\pi e^+e^-$ mass, and has
1466 a distributions in the angles that is typically flat or with a mild curvature. This background
1467 is modelled with an exponential function for $m_{K\pi e^+e^-}$, and Chebyshev polynomials for
1468 the $\cos\theta_\ell$, $\cos\theta_K$ and ϕ . The order of the polynomials is chosen to be up to two for all

angles, as the statistics available for the determination of its shape is expected to be limited, and order two is the minimum required to account for any non-linear behaviour. Full factorisation between mass and angles, and between the three angles is assumed. This assumption is expected to be valid to a good extent due to the random nature of this background. However, it is known that the $B^+ \rightarrow K^+ e^+ e^-$ veto introduces some correlation between mass and angles, in particular with $\cos \theta_K$. The impact of ignoring such a correlation is studied in the systematics.

Cutting on constrained q^2 alters the natural exponential mass distribution. The extent of the distortion is investigated using LFV $K\pi e\mu$ sample after standard pre-selection and additional requirements to reduce electron-muon mis-identification. At the nominal MVA cut of $MVA > 0.99$ the deviation from exponential behaviour due to the cut on q_c^2 is insignificant for the signal (Figure 62). However, it is more pronounced for the control mode (Figure 63). The B^0 mass with both PV and J/ψ mass constraint is especially strongly affected, although the impact within the fit range of 5150 to 5900 MeV/ c^2 is limited. This effect is reduced for the PV constrained B^0 mass, which is the nominal spectra used in this measurement.

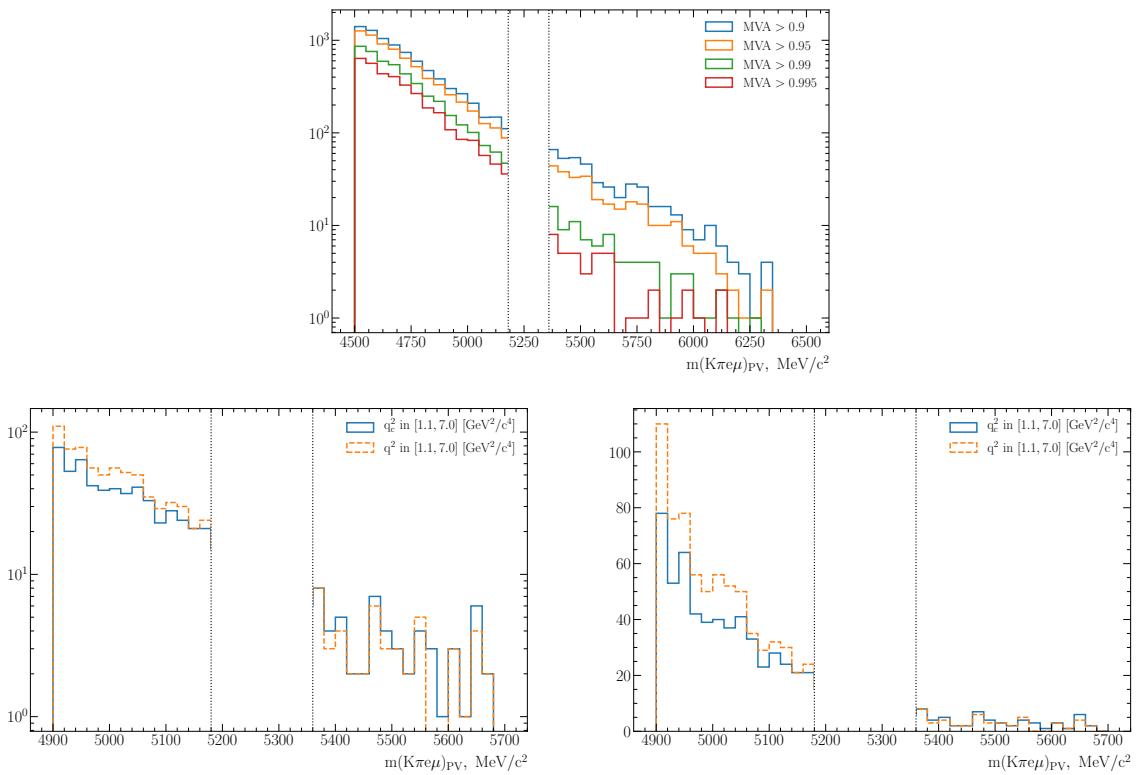


Figure 62: Distribution of PV constrained $m(K\pi e\mu)$ mass in the signal q^2 region for different MVA thresholds (top), and for the nominal working point of $MVA > 0.99$ (bottom row; log plot left, normal right). The bottom plots are cut to the fit region of 4900 to 5700 MeV/ c^2 .

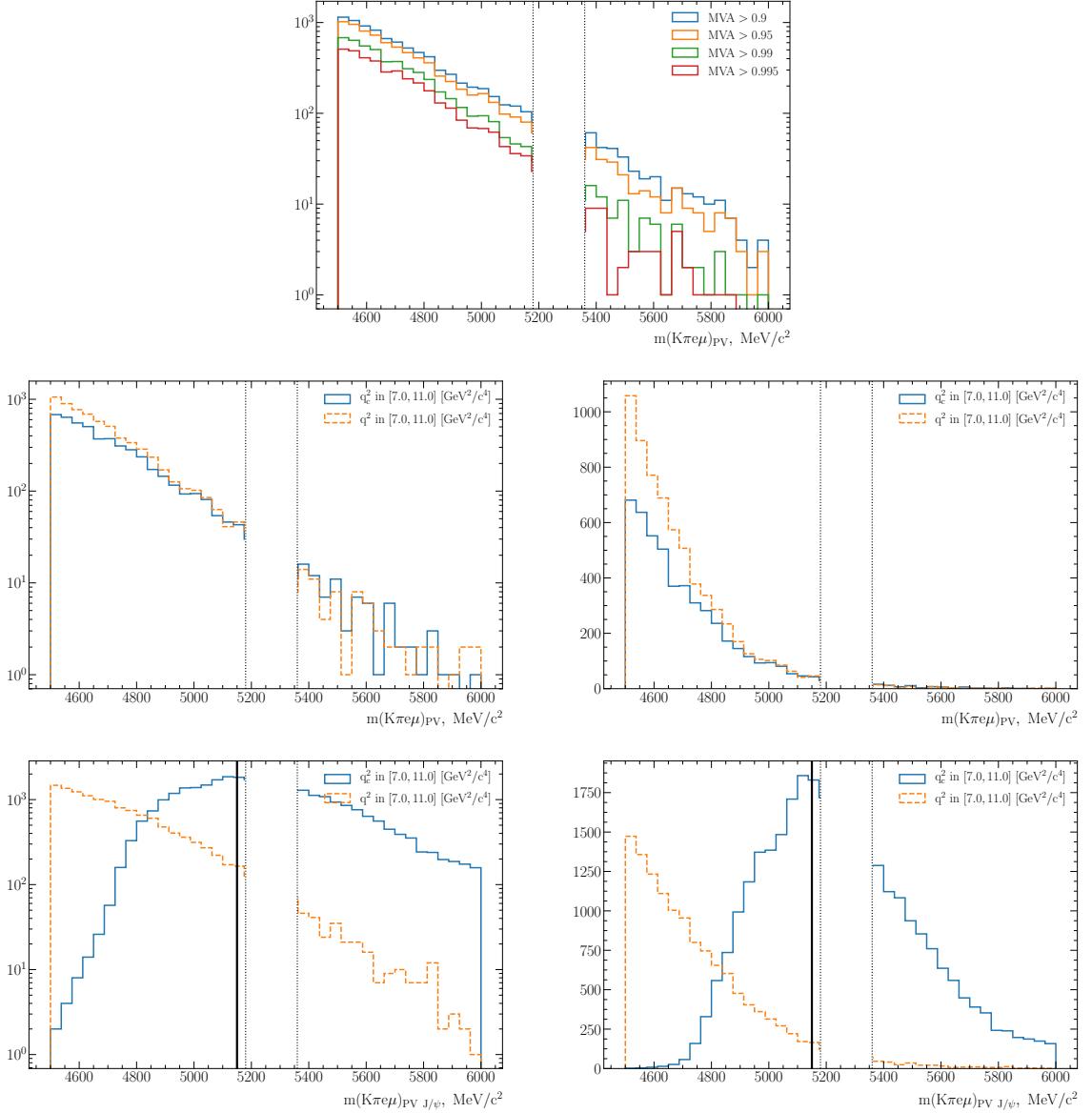


Figure 63: Distribution of PV constrained $m(K\pi e\mu)$ mass in the resonant mode q^2 region for different MVA thresholds (top), and for the nominal working point of $MVA > 0.99$ (middle row; log plot left, normal right). The corresponding distributions for the J/ψ and PV constrained $m(K\pi e\mu)$ mass is shown in the bottom row. Note that the fit region above 5150 MeV/ c^2 , indicated by the black vertical line, is not strongly affected in any case.

1485 7.2 Double semi-leptonic cascade decays

1486 The double semi-leptonic cascade background (DSL) is comprised of decays of the type
1487 $B^0 \rightarrow D^- (\rightarrow K^{*0} e^- \bar{\nu}_e) e^+ \nu_e$. These decays are particularly dangerous because they
1488 present the same visible final state particles of the signal, they pass all the PID and $m_{K\pi}$
1489 requirements and have branching fractions of $\mathcal{O}(10^{-4})$, roughly a thousand times larger
1490 than the signal. Due to the energy loss from the missing neutrinos, this background has a
1491 distribution in $K\pi e^+ e^-$ mass similar to that of the combinatorial, but with a slope that
1492 dies at the mass of the B^0 due to kinematics. The similarity is however not mirrored in the
1493 angular distribution of θ_ℓ : as the positron (electron) from the B (\bar{B}) tends to have higher
1494 momentum than the one from the D , and θ_ℓ by definition will always refer to the angle
1495 between the B and this positron/electron (Section 2.2), it will generally take on values
1496 close to zero, giving rise to a distinctive, asymmetric $\cos\theta_\ell$ distribution peaked near one.
1497 In previous analyses, two alternatives have been used to veto this contribution: a cut has
1498 been applied to the invariant mass of the K^{*0} (\bar{K}^{*0}) and the e^- (e^+) above the D^- mass
1499 at 1780 or 1900 MeV/ c^2 , assuming a very low momentum neutrino; a cut on $|\cos\theta_l| < 0.8$
1500 has been applied (angular analysis at the very low- q^2 region [85] and R_{K^*} [52]). Notice
1501 that both criteria are found to have a similar performance, though the first strategy would
1502 sculpt the angular distribution of θ_ℓ quite strongly, making the parametrisation of the
1503 acceptance after this veto quite challenging.

1504 In order to further investigate the best strategy to model this type of contribution, a
1505 series of fully simulated $B^0 \rightarrow D^- (\rightarrow K^{*0} (\rightarrow K^+ \pi^-) e^- \bar{\nu}) e^+ \nu$ events has been generated
1506 with the decay files “11584021” or “11584022”. While the main decay chain is common to
1507 both, the $D^- \rightarrow K^- \pi^+ e^+ \nu_e$ decay is either generated with the ISGW2 or from the recent
1508 analysis from BESIII of the eponymous mode [86]. Further details on this simulation can
1509 be found in Sec. 3.2. Figure 64 shows the mass and angular distributions for these two
1510 alternative models. Notice that in overall there is not a significant difference between
1511 the two samples, indicating a marginal impact due to the underlying dynamics produced.
1512 Moreover, the motivation to typically cut on the $K\pi e$ invariant mass or in the distribution
1513 of $\cos\theta_\ell$ to reduce this contribution is illustrated.

1514 The validation of this exclusive decay chain as dominant background in this spectra
1515 is paramount. A possible data proxy for this test is the lepton-flavour-violating channel
1516 $B^0 \rightarrow K^* e^\pm \mu^\mp$, where other sources of backgrounds such as partially reconstructed or
1517 J/ψ leakage are absent. Figure 65 shows a comparison between simulated events for
1518 $B^0 \rightarrow D^- (\rightarrow K^{*0} (\rightarrow K^+ \pi^-) e^- \bar{\nu}) e^+ \nu$ and $B^0 \rightarrow D^- (\rightarrow K^{*0} (\rightarrow K^+ \pi^-) \mu^- [e^-] \bar{\nu}) e^+ [\mu^+] \nu$
1519 decays, where the square brackets indicate the decay where the electron and the muon are
1520 swapped. Note a good agreement in all distributions, which suggests that the impact of the
1521 different resolutions between electron and muons in the variables of interest is negligible for
1522 this background. Therefore, it is reasonable to assume that the DSL contribution in the
1523 LFV data is a good representation of the one expected in the signal rare mode. Further
1524 compatibility checks using data-driven approaches are reported in Appendix I.

1525 Figure 66 depicts the different distributions for the $K\pi e\mu$ data and $B^0 \rightarrow D^- (\rightarrow$
1526 $K^{*0} e^- \bar{\nu}) e^+ \nu$ simulation. The LFV sample has been obtained by applying the same selection

1527 as the rare mode and in the signal region of $m_B \in [4500, 5500]$.¹² While contributions from
 1528 combinatorial backgrounds are still present in the data (though dominated by DSL events),
 1529 an immediate inspection of these distributions clearly show a significant disagreement
 1530 between both samples. In particular, the width of the peaking structure around $\cos \theta_\ell$
 1531 close to +1 seems larger for data than for simulation. In fact, this is further confirmed
 1532 by subtracting the left-hand side of the $\cos \theta_\ell$ distribution from the right-hand side (see
 1533 Appendix I for details). Some of these differences could be explained by the presence of
 1534 events from other DSL decays not included in the single-component simulation, such as
 1535 $B^0 \rightarrow D^*(2010)e^-\bar{\nu}_e$, which may modify the shape and width of the main peak. Since this
 1536 rich structure seen in the LFV sample is difficult to model in simulation (*e.g.* a cocktail
 1537 of semi-leptonic decays or even the combination of random K^* with true DSL events), a
 1538 data-driven approach to obtain a parametrisation for the DSL-like contribution is preferred
 1539 and discussed in the following.

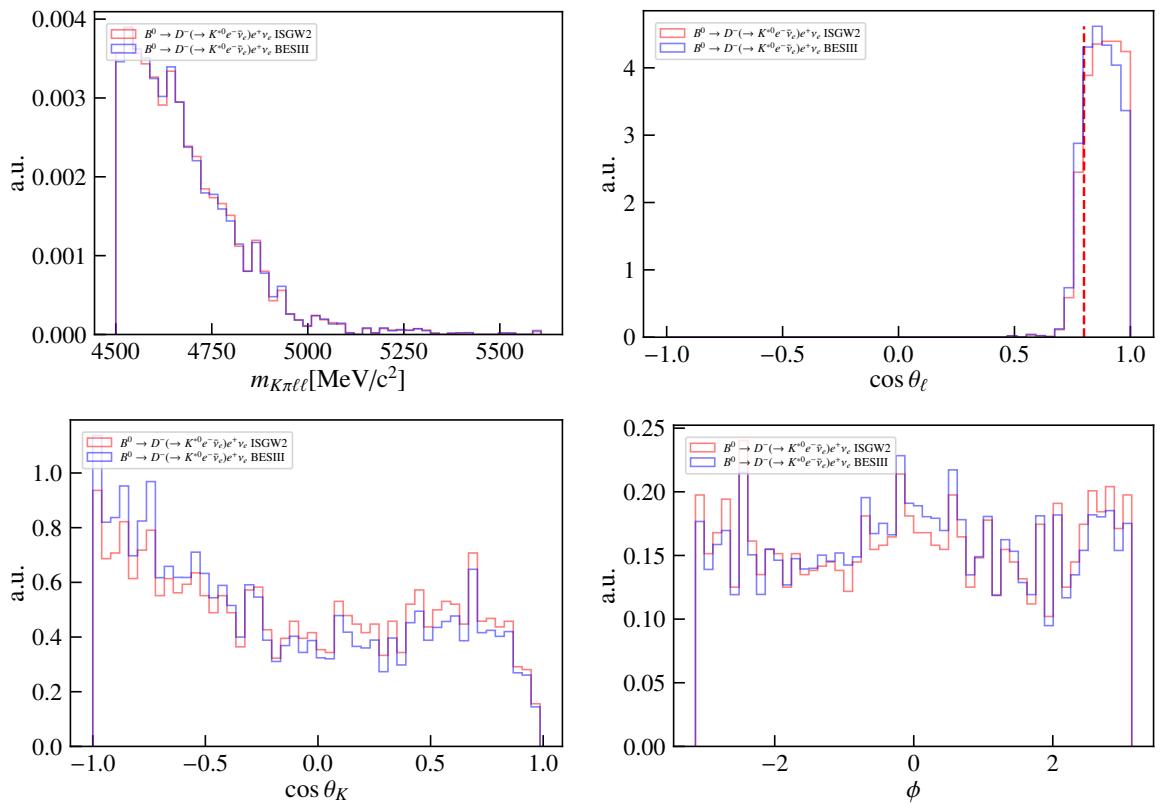


Figure 64: Comparison between the invariant mass and angular distributions of DSL simulations using different form factors parametrisations for $B^0 \rightarrow D^- (\rightarrow K^{*0} e^- \bar{\nu}_e) e^+ \nu_e$ decays.

¹²Note that while the signal region where potentially the LFV channel can appear is investigated, no inspection to the invariant mass is performed. In addition, any impact of this contribution in the angular distributions are expected to be marginal. Therefore, it is safe to assume that the relevant figures for the search for this unobserved channel are preserved.

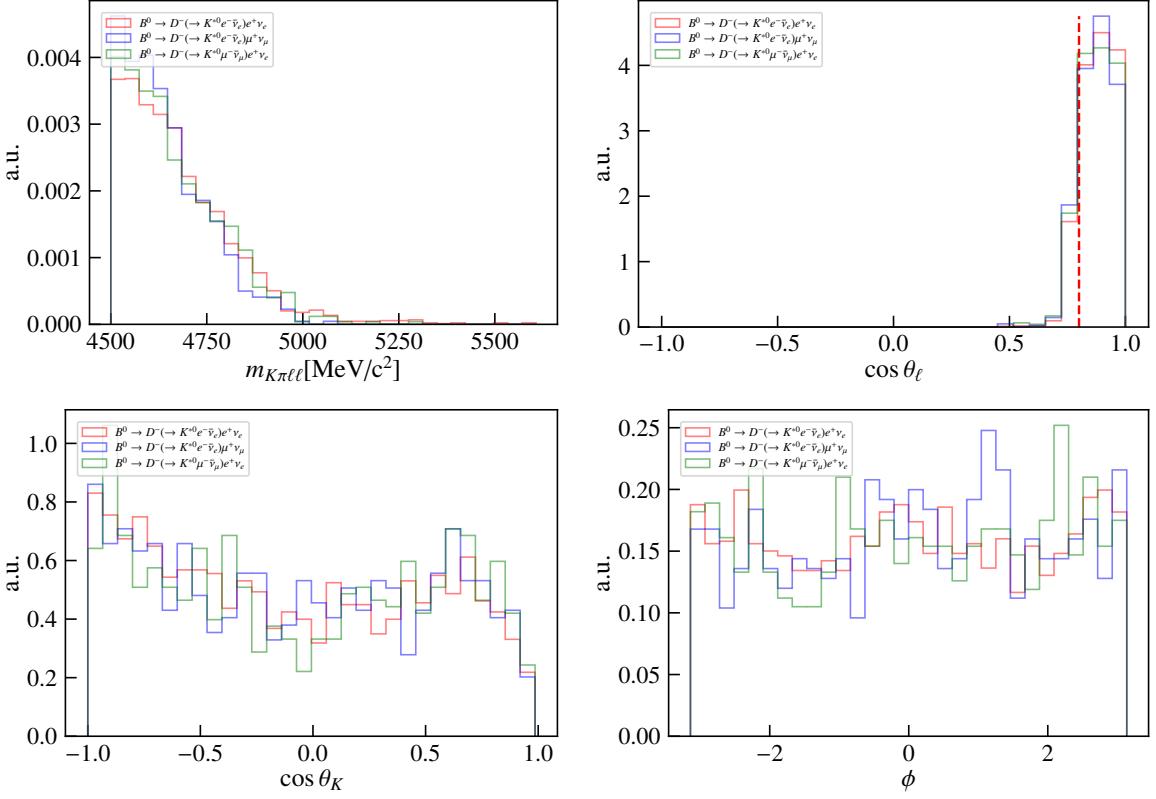


Figure 65: Comparison between the invariant mass and angular distributions of DSL simulations for (red) $B^0 \rightarrow D^- (\rightarrow K^{*0} e^- \bar{\nu}_e) e^+ \nu_e$ and (blue or green) $B^0 \rightarrow D^- (\rightarrow K^{*0} (\rightarrow K^+ \pi^-) \mu^- [e^-] \bar{\nu}) e^+ [\mu^+] \nu$ decays.

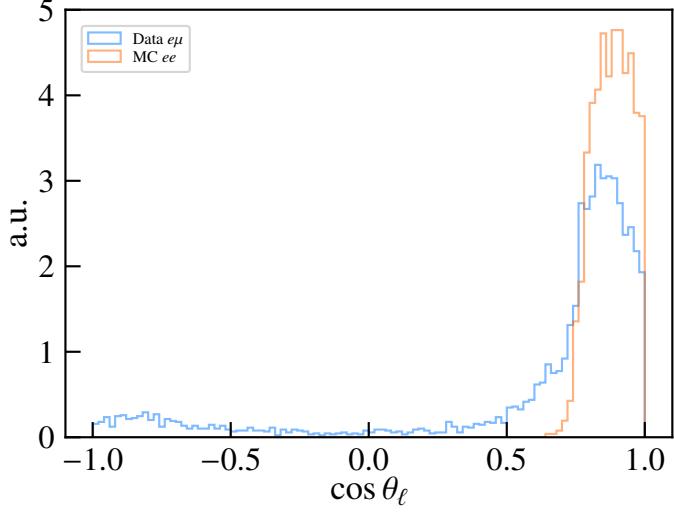


Figure 66: Comparison between $K\pi e\mu$ data and $B^0 \rightarrow D^- (\rightarrow K^{*0} e^- \bar{\nu}) e^+ \nu$ simulation in the $\cos \theta_\ell$ distribution.

1540 **7.3 Combinatorial/DSL combined modelling**

1541 A data-driven approach has been designed to obtain both combinatorial and double-
 1542 semileptonic cascade background parametrisations. This methodology is one of the key
 1543 features of this analysis, since it provides the possibility to examine the entirely $\cos \theta_\ell$
 1544 distribution of the rare mode, which in turn, enhances the sensitivity of the analysis. The
 1545 idea consists in using the $B^0 \rightarrow K^{*0} e^\pm \mu^\mp$ data in a two-stage procedure as follow

- 1546 i Isolate a sample enriched in DSL with minimal combinatorial contamination by
 1547 taking a lower B invariant mass range of [4500, 5200]MeV/c² and applying a very
 1548 tight cut to the combinatorial MVA output (MVA > 0.9985). In this sample, a
 1549 fit to angular distributions alone is performed to obtain the lineshape of for the
 1550 DSL contribution. Note that for $\cos \theta_K$ and ϕ projections a Chebychev polynomial
 1551 function is used to model while a KDE is used in the non-trivial $\cos \theta_\ell$ distribution.
 1552 Figure 67 shows the corresponding fit results to the DSL enriched sample, which
 1553 parameters are gathered in Table 21.
- 1554 ii With the parameters of the angular part fixed for the DSL, an invariant mass and an-
 1555 gular fit to an independent sample containing contributions from both combinatorial
 1556 and DSL events to obtain the slope of the DSL, and the angular shape and slope of
 1557 the combinatorial background is performed. This sample is taken from the nominal
 1558 mass range of [4900, 5700]MeV/c², with the standard MVA cut (MVA > 0.99). The
 1559 cut of MVA < 0.9985 is applied in addition to exclude events used in stage one.
 1560 Figure 68 and Table 22 report the fit results of this procedure.
- 1561 Notice that, due to the inability to achieve good DSL-combinatorial separation through
 1562 the usage of mass alone, both combinatorial and DSL shapes need to be extracted together.
 1563 Further details on the compatibility between the parametrisation obtained from the $K\pi e\mu$
 1564 sample, the $K\pi ee$ sample and the simulation can be found in Appendix I. In any case, the
 1565 impact of this model choice is also discussed in the systematics.

Parameter	Run-I value	Run-II value
$c_{1;\text{DSL}}^K$	-0.10 ± 0.08	-0.15 ± 0.08
$c_{1;\text{DSL}}^\phi$	-0.10 ± 0.07	-0.10 ± 0.07
$c_{2;\text{DSL}}^K$	0.33 ± 0.06	0.21 ± 0.07
$c_{2;\text{DSL}}^\phi$	-0.09 ± 0.07	0.01 ± 0.07

Table 21: Parameter values obtained from fits to step one of the two-step procedure used to model the DSL and combinatorial backgrounds.

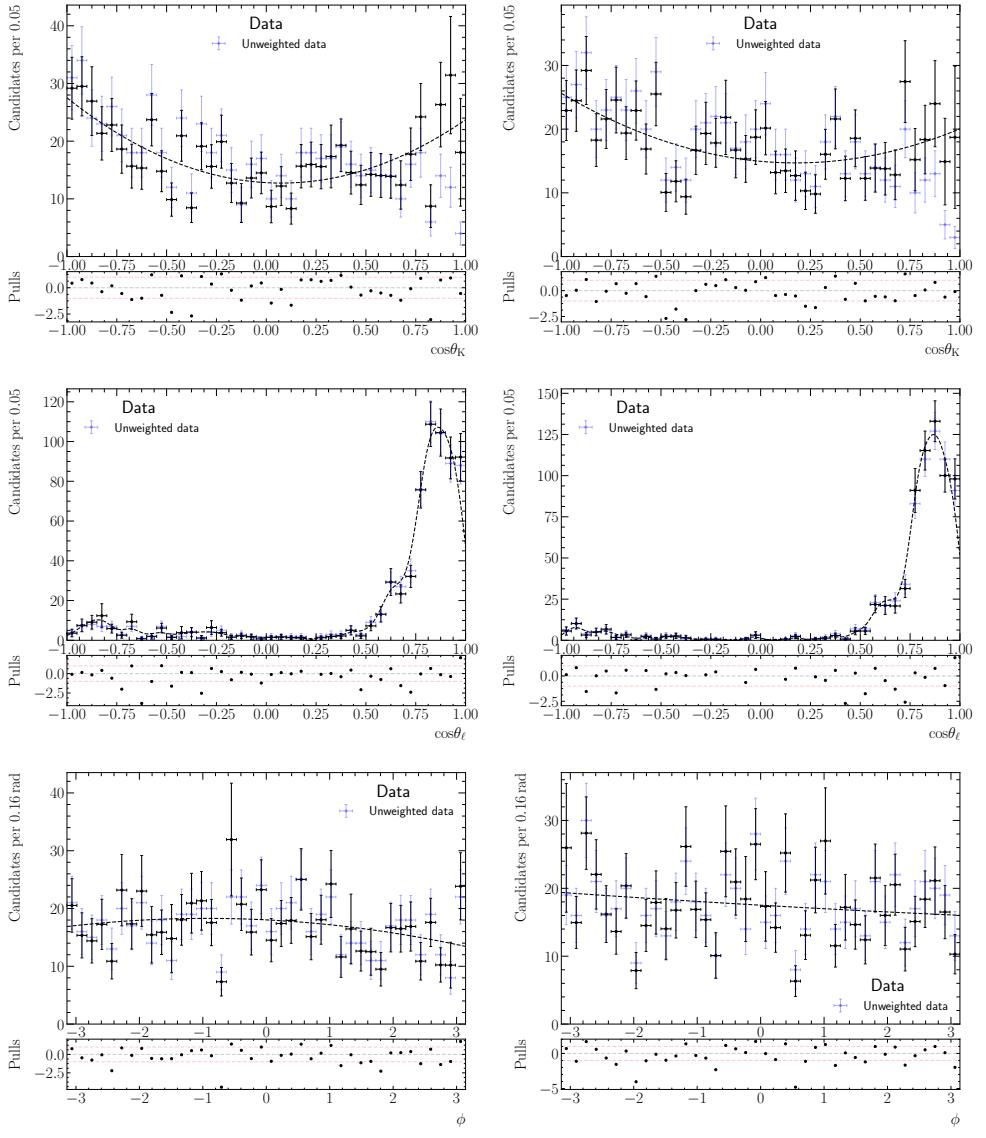


Figure 67: Data fits to enriched DSL samples (1st stage procedure) from $B^0 \rightarrow K^{*0} e^\pm \mu^\mp$ decays for (left) Run-I and (right) Run-II conditions.

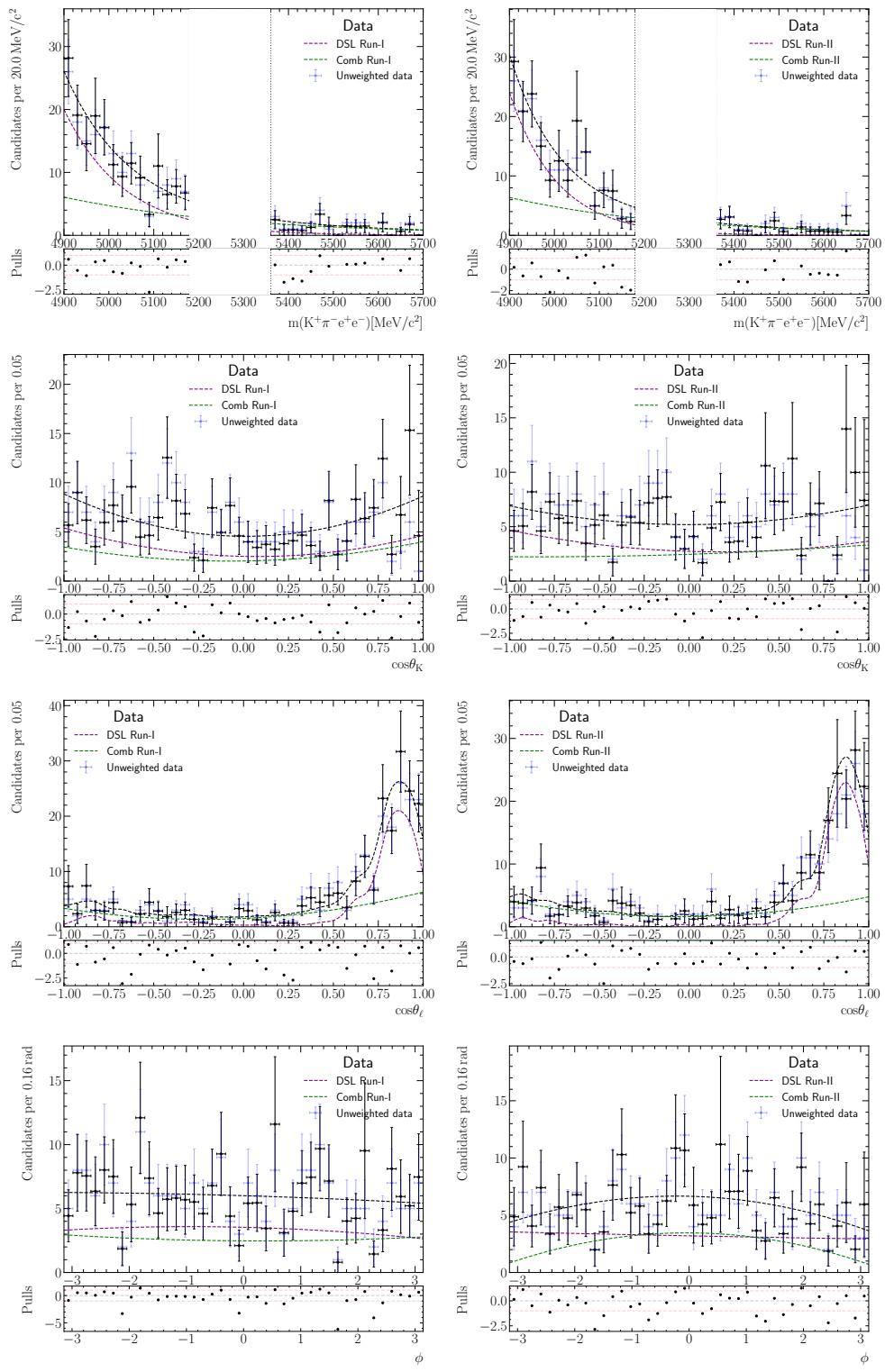


Figure 68: Data fits to $B^0 \rightarrow K^{*0} e^\pm \mu^\mp$ decays in the nominal mass range to obtain simultaneously (2nd stage procedure) the combinatorial angular shape and the DSL mass modelling for (left) Run-I and (right) Run-II conditions. Note that in these fits the DSL angular model is fixed to the parametrisation obtained in the 1st stage of the procedure.

Parameter	Run-I value		Run-II value	
f_{DSL}	0.56	± 0.11	0.56	± 0.08
$c_{1;\text{Comb}}^\ell$	0.47	± 0.21	0.10	± 0.24
$c_{1;\text{Comb}}^K$	0.09	± 0.22	0.21	± 0.22
$c_{1;\text{Comb}}^\phi$	-0.03	± 0.23	-0.03	± 0.21
$c_{2;\text{Comb}}^\ell$	0.52	± 0.23	0.47	± 0.19
$c_{2;\text{Comb}}^K$	0.29	± 0.19	0.06	± 0.21
$c_{2;\text{Comb}}^\phi$	0.07	± 0.21	-0.63	± 0.25
λ_{Comb}	-0.0025 ± 0.0009		-0.0028 ± 0.0006	
λ_{DSL}	-0.0075 ± 0.0009		-0.0093 ± 0.0013	

Table 22: Parameter values obtained from step two of the two-step procedure to model DSL and combinatorial.

1566 7.4 Partially reconstructed $B^+ \rightarrow K^+\pi^+\pi^-e^+e^-$ decays

1567 Partially reconstructed background refers to decays that are incorrectly reconstructed as
 1568 signal, with one or more particles in the final state unaccounted for. The main source
 1569 of this type of background comes from decays featuring heavier kaon resonances of the
 1570 type $B^+ \rightarrow K^{*(**)}e^+e^-$, such as $B^+ \rightarrow K_1^+(\rightarrow K^+\pi^+\pi^-)e^+e^-$ and $B^+ \rightarrow K_2(1430)^+(\rightarrow$
 1571 $K^+\pi^+\pi^-)e^+e^-$, which can be reconstructed as the signal with one or more pions missing.
 1572 Due to the missing pion(s), the shape of the partially reconstructed background in the
 1573 B^0 invariant mass is that of a broad peak with its center shifted towards the lower mass
 1574 region. The modelling strategy for this backgrounds involves using KDE for the B mass,
 1575 and factorised Chebyshev polynomials up to second order for the angles.

1576 The KDE and the angular polynomial coefficients are obtained from $B^+ \rightarrow$
 1577 $K^+\pi^+\pi^-e^+e^-$ simulation reconstructed as the signal, with data-driven correction weights
 1578 based on $B^+ \rightarrow K^+\pi^+\pi^-(J/\psi \rightarrow \mu^+\mu^-)$. The impact of these corrections can be visualised
 1579 in Fig. 69. The distributions corrected by both hadronic dynamics and acceptance effects
 1580 are shown on Figure 70.

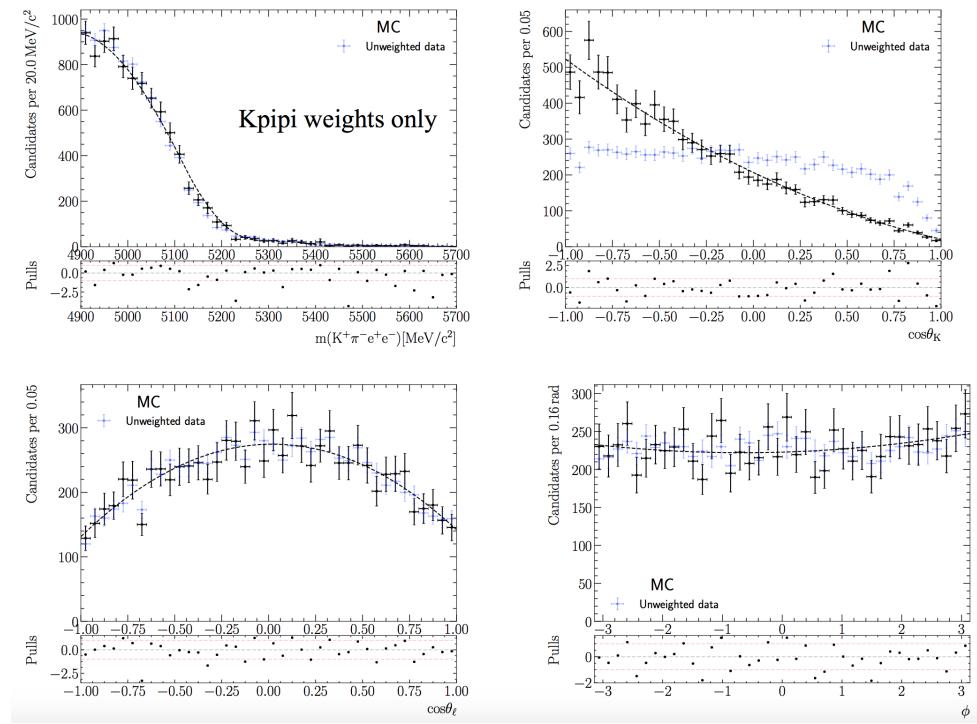


Figure 69: Results of fits to the $B^+ \rightarrow K^+\pi^-\pi^-e^+e^-$ simulation reconstructed as the signal for 2016 (only available sample) including only the data-driven corrections.

Parameter	Run-I value	Run-II value
$c_{1; \text{PR}}^\ell$	0.042 ± 0.035	0.048 ± 0.034
$c_{1; \text{PR}}^K$	-0.874 ± 0.020	-0.852 ± 0.021
$c_{1; \text{PR}}^\phi$	0.058 ± 0.032	0.043 ± 0.032
$c_{2; \text{PR}}^\ell$	-0.09 ± 0.04	-0.085 ± 0.035
$c_{2; \text{PR}}^K$	0.211 ± 0.023	0.229 ± 0.023
$c_{2; \text{PR}}^\phi$	0.047 ± 0.031	0.050 ± 0.031

Table 23: Parameter values found from weighted fits of the 2016 $B^- \rightarrow K^+ \pi^- \pi^- e^+ e^-$ simulation reconstructed as the signal (only available sample). The Run-I and Run-II labels reflect only the acceptance functions used.

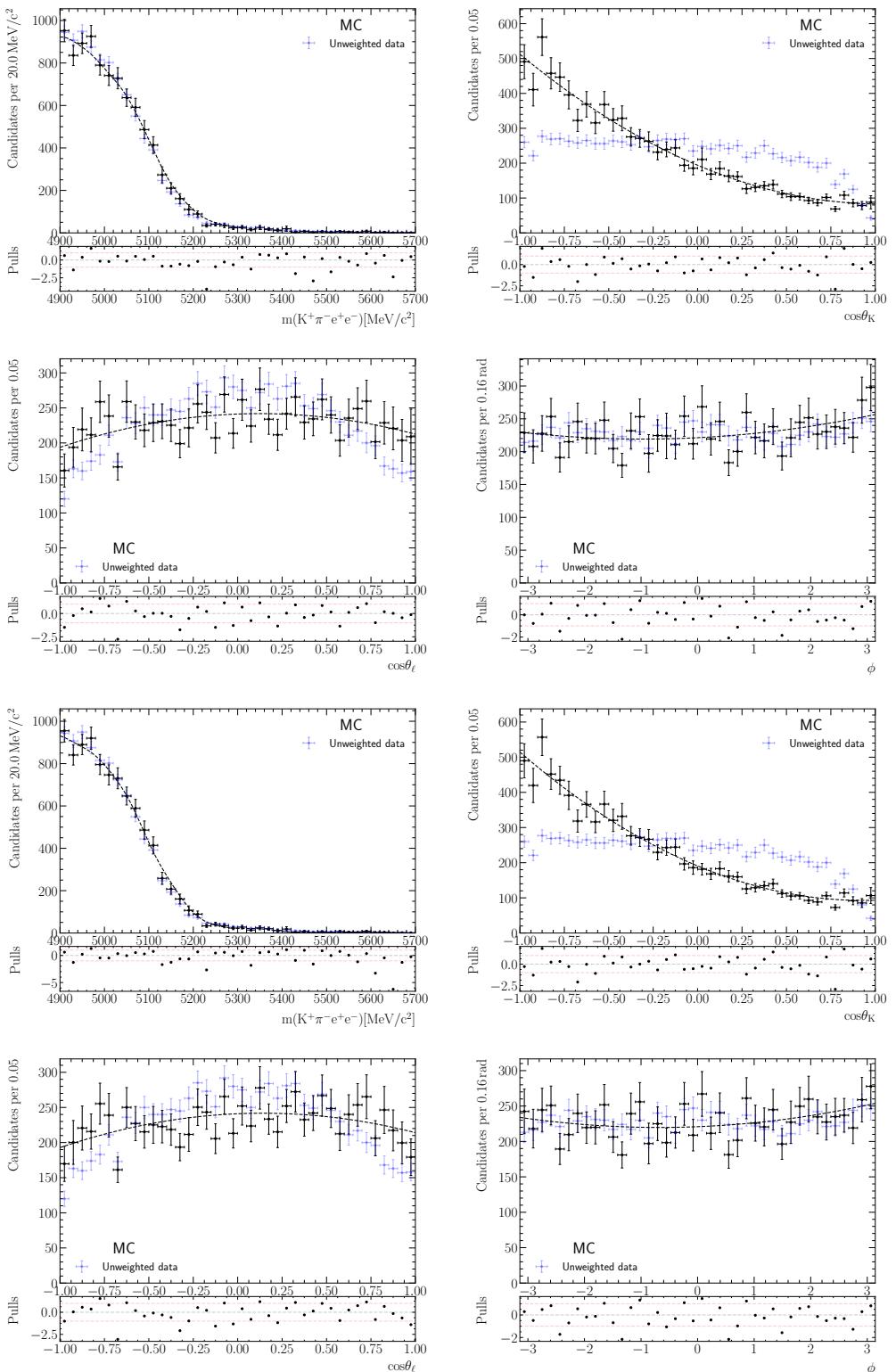


Figure 70: Results of fits to the $B^- \rightarrow K^+ \pi^- \pi^- e^+ e^-$ simulation reconstructed as the signal for 2016 (only available sample). The top (bottom) four plots show the results obtained with Run-I (Run-II) acceptance weights. The acceptance weights are applied in addition to data-driven correction weights.

1581 7.5 Control channel

1582 The $B^0 \rightarrow J/\psi(\rightarrow e^+e^-)K^{*0}$ control mode in this analysis is utilised either in the
 1583 validation of the angular fit or in the aid of the stability of the rare mode mass fit by
 1584 providing information on the shift and scaling of the width and mean position seen in
 1585 data, respectively. While a detailed discussion of the background contribution for the
 1586 angular fit cross-check is given in Sec. 9, the relevant mass fit has a simplified structure. In
 1587 particular, due to the cut on the PV- J/ψ constrained invariant mass of greater than 5150,
 1588 a significant reduction of different sources of background are expected. In the nominal
 1589 fit, only contributions from combinatorics (free in the fit) and Λ_b^0 mid-identified decays
 1590 are considered. Other sources such as DSL, $K - \pi$ swap, partially reconstructed and B_s^0
 1591 misID decays are expected to contribute to a marginal level, and are considered in the
 1592 systematics.

1593 The decay of $\Lambda_b^0 \rightarrow pKJ/\psi(\rightarrow e^+e^-)$ can become a background for the control mode if
 1594 the proton is mis-identified as a pion. The distributions of these events are modelled using
 1595 simulated $\Lambda_b^0 \rightarrow pKJ/\psi$ events reconstructed as $B^0 \rightarrow \pi KJ/\psi$. However, contributions
 1596 from multiple Λ resonances and exotic states [87] in data pose additional complications.
 1597 To take these into account, a data-driven approach is used, where per-event correction
 1598 weights are determined for the phase-space simulation of $\Lambda_b^0 \rightarrow pKJ/\psi$, based on the $K\pi$
 1599 and $J/\psi p$ invariant mass squared of $\Lambda_b^0 \rightarrow pKJ/\psi(\rightarrow \mu^+\mu^-)$ data. Figure 71 shows the
 1600 projections for the weighted distributions fitted with a KDE model.

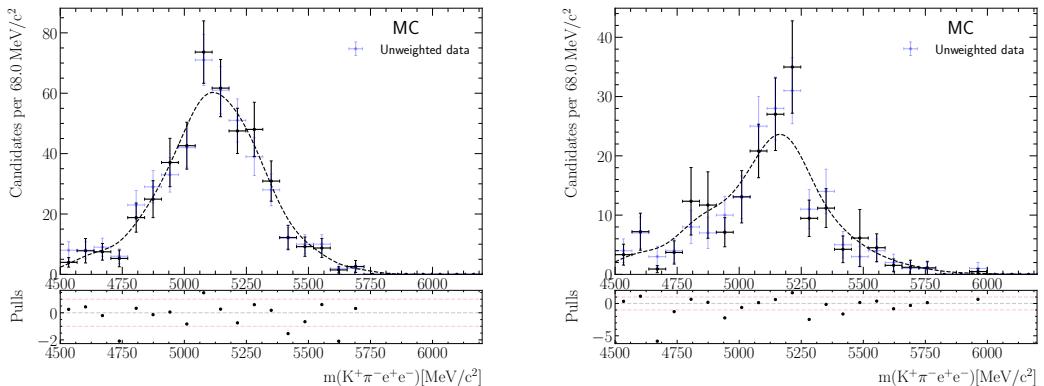


Figure 71: Non-parametric mass models for mis-identified Λ_b^0 decays obtained using $\Lambda_b^0 \rightarrow pKJ/\psi$ simulation reconstructed as $\Lambda_b^0 \rightarrow \pi KJ/\psi$ for (left) Run-I and (right) 2016 samples.

1601 8 Angular fit

1602 Details on the unbinned data fits and corresponding validations performed with simulated
 samples are discussed in this section. The scripts used in this task can be found in

- 1603 ○ `ewp-bd2ksteeangular-central-q2/scripts/angular_fit`

1604 The main measurement of this analysis is made by applying an unbinned weighted
 1605 maximum likelihood fit to selected signal candidates. Three sources of backgrounds are
 1606 considered for the rare mode – combinatorial, DSL and partially reconstructed, and two
 1607 for the control mode mass fit, which is used only for the determination of the shift and
 1608 scaling parameters of the signal mass shape, namely combinatorial and mis-identified
 1609 $\Lambda_b^0 \rightarrow pKJ/\psi(\rightarrow e^+e^-)$ decays. For the rare mode, all background angular parameters are
 1610 fixed in the fit. The decision to fix also the parameters of the combinatorial background is
 1611 made to increase fit stability at the cost of an additional source of systematic uncertainty.
 1612 Further details on studies that motivated this decision will be added in the next version
 1613 of the note in the appendix. Parameters related to the mass distributions are allowed to
 1614 vary. Detailed information regarding the nominal fit set-up and the expected sensitivity
 1615 values are given in the following sections.

1616 8.1 Weighted unbinned maximum likelihood fit

1617 The values of the angular observables are extracted from a weighted unbinned maximum
 1618 likelihood fit to the reconstructed B^0 mass with PV constraint, and the three decay angles,
 1619 $\vec{\Omega} = (\theta_\ell, \theta_K, \phi)$. Both signal and background components are described by analytically
 1620 normalised PDFs, which depend on observables of interest, λ_{phys} (the S_i or $P_i^{(\prime)}$), or
 1621 nuisance parameters, λ_{nuis} . The total PDF is of the form

$$1622 \quad \mathcal{P}_{\text{tot}}(m_{K\pi ee}, \vec{\Omega}) = f_S \mathcal{P}_S(m_{K\pi ee}, \vec{\Omega}) + \sum_i^{n-1} f_{B,i} \mathcal{P}_{B,i}(m_{K\pi ee}, \vec{\Omega}) + (1 - f_S - \sum_i^{n-1} f_{B,i}) \mathcal{P}_{B,n}(m_{K\pi ee}, \vec{\Omega}), \quad (38)$$

1623 where f_S is the signal fraction and $f_{B,i}$ are the fractions of the background components.
 1624 As the full PDF needs to be normalised to one, for convenience, the fraction of one
 1625 background – the combinatorial unless stated otherwise – is not fitted explicitly, but taken
 1626 to be the remainder after subtracting the sum of all other component fractions from unity.
 Factorisation between decay angles and B^0 mass is assumed for all components, that is,

$$1627 \quad \mathcal{P}_S(m_{K\pi ee}, \vec{\Omega}) = \mathcal{P}_S(m_{K\pi ee}) \times \mathcal{P}_S(\vec{\Omega}), \quad (39)$$

$$1627 \quad \mathcal{P}_B(m_{K\pi ee}, \vec{\Omega}) = \mathcal{P}_B(m_{K\pi ee}) \times \mathcal{P}_B(\vec{\Omega}). \quad (40)$$

1628 To determine observable values, the negative log-likelihood is minimised with respect to
 1629 the λ_{phys} , and $\lambda_{\text{nuisance}}$. It is given by

$$1627 \quad -\ln \mathcal{L} = - \sum_{\text{events } e} \omega_e \times \ln \mathcal{P}_{\text{tot}}(m_{K\pi ee,e}, \vec{\Omega}_e | \lambda_{\text{phys}}, \lambda_{\text{nuis}}), \quad (41)$$

1630 where the per-event weights, ω_e , are the inverse values of the acceptance function, ϵ ,
 1631 evaluated for each event, *i.e.*

$$\omega_e = 1/\epsilon(q^2_e, \cos\theta_{\ell e}, \cos\theta_{K e}, \phi_e) . \quad (42)$$

1632 8.2 Rare and control mode background components

1633 The backgrounds included in the full model for the rare mode consist of the double
 1634 semi-leptonic, partially reconstructed and combinatorial components. The fractions of
 1635 the first two components are allowed to vary, and the fraction of combinatorial events is
 1636 the remainder after subtracting signal, double semi-leptonic and partially reconstructed
 1637 fractions from one. For all these components, the parameters governing the shapes of
 1638 angular distributions are fixed from simulation or data fits as described in Sec. 7. The
 1639 slope of the exponential of the combinatorial background is allowed to vary, as it shows
 1640 clear correlation with the signal fraction, but is at the same time difficult to reliably extract
 1641 from simulation or data-driven methods.

1642 To take into account remaining data-simulation differences, the width and mean
 1643 position of the signal mass peak are allowed to vary. Due to the low statistics of the
 1644 rare mode, the shift and scaling parameters, δ_μ and δ_σ , are determined in a simultaneous
 1645 fit to the control mode (mass-only) for which three background components are used,
 1646 namely the combinatorial, $\Lambda_b^0 \rightarrow pKJ/\psi(\rightarrow e^+e^-)$ with $p \rightarrow \pi$ mis-identification and
 1647 $B_s \rightarrow K^{*0}J/\psi(\rightarrow e^+e^-)$. The $B_s \rightarrow K^{*0}J/\psi(\rightarrow e^+e^-)$ mass distributions is assumed to
 1648 be the same as that of $B^0 \rightarrow K^{*0}J/\psi(\rightarrow e^+e^-)$. A shift of $\delta_{B_s-B^0} = m_{B_s}^{\text{PDG}} - m_{B^0}^{\text{PDG}} =$
 1649 $(87.2 \pm 0.3)\text{MeV}/c^2$ in the mean position of the peak is added to adjust for the difference
 1650 in invariant mass between the B_s and B^0 . The B_s contribution as a fraction of the B^0
 1651 fraction is fixed to the values calculated from the branching fractions of $B_s \rightarrow K^{*0}J/\psi$
 1652 and $B^0 \rightarrow K^{*0}J/\psi$ decays [71], and the fragmentation fraction ratio f_s/f_d [88] as

$$f_{B_s} = \frac{f_s}{f_d} \times \frac{\mathcal{B}(B_s \rightarrow K^{*0}J/\psi)}{\mathcal{B}(B^0 \rightarrow K^{*0}J/\psi)} , \quad (43)$$

1653 where same overall efficiency for both B_s^0 and B^0 modes is assumed. The fractions of
 1654 0.009 ± 0.001 and 0.0079 ± 0.0009 for Run-I and Run-II, respectively, are found to be in
 1655 good agreement with fits to B^0 mass with J/ψ and PV constraints, which offers enhanced
 1656 sensitivity to this contribution.

1657 The Λ_b^0 mis-identification background fractions are estimated in a similar fashion, but
 1658 without assuming equality in efficiencies,

$$f_{\Lambda_b^0 \text{ misid}} = \frac{f_{\Lambda_b^0}}{f_d} \times \frac{\epsilon_{\Lambda_b^0 \text{ misid}}}{\epsilon_{B^0}} \frac{3}{2} \times \frac{\mathcal{B}(\Lambda_b^0 \rightarrow pKJ/\psi)}{\mathcal{B}(B^0 \rightarrow K^{*0}J/\psi)} . \quad (44)$$

1659 The efficiencies of $\epsilon_{\Lambda_b^0 \text{ misid}}$ and ϵ_{B^0} are obtained from simulation, the factor 3/2 comes from
 1660 the Clebsch-Gordon factor for $K^{*0} \rightarrow K^+\pi^-$. Using Λ_b^0/f_d fractions from Ref. [70] and
 1661 assuming $f_u = f_d$, $f_{\Lambda_b^0 \text{ misid}}$ fractions are found to be 0.0011 ± 0.0003 and 0.0004 ± 0.0001

for Run-I and Run-II, respectively. Note that the Run-I value is calculated by weighting 2011 and 2012 simulation results with the signal yields obtained from a mass fit using B^0 mass with J/ψ and PV constraints, that is,

$$f_{\Lambda_b^0}^{\text{Run-I}} = \frac{N_{J/\psi}^{2011} f_{\Lambda_b^0 \text{ misid}}^{2011} + N_{J/\psi}^{2012} f_{\Lambda_b^0 \text{ misid}}^{2012}}{N_{J/\psi}^{2011} + N_{J/\psi}^{2012}}, \quad (45)$$

while the Run-II fraction is taken to the value obtained using the 2016 simulation due to the lack of 2015 samples (which are also expected to have significantly lower statistics compared to 2016 samples).

The slope of the combinatorial component is allowed to vary freely in the fit. Although the double semi-leptonic background is still expected to be present in the control mode but with a smaller contribution, this background is not treated separately from the combinatorial, as its fraction is difficult to determine without angular information. Partially reconstructed decays are expected to be significantly reduced by the cut on the B^0 mass with J/ψ and PV constraints, and thus are also accommodated by the combinatorial component. The impact of the choice to model both backgrounds within the combinatorial lineshape on the determination of the shift and scaling parameters are expected to be marginal, and is examined as a source of systematic uncertainties.

8.3 Full angular fit

The full angular fit used to extract the nominal observable values involves a simultaneous fit to Run-I and Run-II rare mode samples, as well as that of the control mode (mass only). Note that the control mode is necessary for the unbiased determination of the shift and scaling of the signal mass shape to account for any remaining data-simulation differences. The full negative log-likelihood to be minimised is then given by the sum of the four components

$$\begin{aligned} -\ln \mathcal{L}_{tot} = & - \sum_{\text{rare events } e, \text{Run-I}} \omega_e \times \ln \mathcal{P}_{tot}(m_{K\pi ee,e}, \Omega_e | \lambda_{\text{phys}}, \lambda_{\text{nuis}}^{\text{Run-I}}) \\ & - \sum_{J/\psi \text{ events } e_{J/\psi}, \text{Run-I}} \omega_{e_{J/\psi}} \times \ln \mathcal{P}_{tot}(m_{K\pi ee,e_{J/\psi}} | \lambda_{\text{nuis}}^{\text{Run-I}}) \\ & - \sum_{\text{rare events } e, \text{Run-II}} \omega_e \times \ln \mathcal{P}_{tot}(m_{K\pi ee,e}, \Omega_e | \lambda_{\text{phys}}, \lambda_{\text{nuis}}^{\text{Run-II}}) \\ & - \sum_{J/\psi \text{ events } e_{J/\psi}, \text{Run-II}} \omega_{e_{J/\psi}} \times \ln \mathcal{P}_{tot}(m_{K\pi ee,e_{J/\psi}} | \lambda_{\text{nuis}}^{\text{Run-II}}) \end{aligned} \quad (46)$$

The observables, λ_{phys} , are shared for both Run-I and Run-II samples. The nuisance parameters allowed to vary in the fit include the shift and scaling parameters, δ_μ and δ_σ , which are shared between the rare and control modes for each Run, and the slopes of the combinatorial mass distributions that are allowed to take on different values for each Run and mode. Furthermore, the fractions of signal, DSL and partially reconstructed

1689 backgrounds are varied separately for both Runs (the combinatorial background fractions
 1690 are given by the sum of the other component fractions subtracted from one). This sim-
 1691 taneous fit approach is motivated by differences seen in the simulation. Further splitting
 1692 the dataset into data collection years and/or trigger categories are challenging for this
 1693 given statistics due to limited sample size available for background modelling (mainly
 1694 affects combinatorial and double semi-leptonic background) and acceptance parametrisa-
 1695 tion. Since per-event weights are calculated from the inverse acceptance function, some
 1696 events may receive relatively large weights. When examining the data this needs to be
 1697 validated and any instance of such behaviour is removed from the dataset and assigned a
 1698 corresponding systematic (if any). Appendix H further details the features involved in
 1699 this weighted fit. Finally, it is worth mentioning that due to the nature of the weighted fit,
 1700 the relative fraction between the different species in the fit are only nuisance parameters
 1701 in our results.

1702 8.4 Weighted fit uncertainty correction

1703 For weighted fits, the uncertainties obtained from the inverse of the matrix of second
 1704 derivatives of the negative log-likelihood are no longer guaranteed to provide correct
 1705 coverage. This is because the presence of the weights ($w \neq 1$) prevents the simplification
 1706 of the full covariance matrix expression to the inverse Hessian. The asymptotically correct
 1707 approach, documented in Ref. [89], involves the calculation of the full expression, which is
 1708 given by

$$\begin{aligned}
 V_{ij} = & \left(\sum_e^N \omega_e^2 \frac{\partial^2 \ln P(x_e | \vec{\lambda})}{\partial \lambda_i \partial \lambda_k} \Big|_{\hat{\vec{\lambda}}} \right)^{-1} \times \left(\sum_e^N \omega_e^2 \frac{\partial \ln P(x_e | \vec{\lambda})}{\partial \lambda_k} \Big|_{\hat{\vec{\lambda}}} \frac{\partial \ln P(x_e | \vec{\lambda})}{\partial \lambda_l} \Big|_{\hat{\vec{\lambda}}} \right) \\
 & \times \left(\sum_e^N \omega_e^2 \frac{\partial^2 \ln P(x_e | \vec{\lambda})}{\partial \lambda_l \partial \lambda_j} \Big|_{\hat{\vec{\lambda}}} \right)^{-1}.
 \end{aligned} \tag{47}$$

1709 In practise, the first and third terms are the Hessian matrices obtained from **HESSE**, and
 1710 the middle term is calculated by multiplying the matrix of first derivatives of the weighted
 1711 negative log-likelihood (split into single-event terms) with its transpose.

1712 8.5 Sensitivity studies

1713 Pseudoexperiments are performed to determine sensitivity to observables, check for fit
 1714 biases, and to determine whether or not uncertainties are properly estimated. To best
 1715 replicate the characteristics of the data fit, toys are generated using the products of five
 1716 dimensional PDFs (*i.e.* B mass, q^2 and Ω) and acceptance functions. The amplitude model
 1717 used for the unbinned $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ [ANA-2020-062] and ΔC_i [ongoing] analyses is used
 1718 to produce signal ensembles under the Standard Model hypothesis ¹³. The addition of q^2

¹³The signal model is generated with the standard formalism defined in Ref. [59], with the form factor basis commonly used in the literature (see Refs. [?, 61]), and are described using two complementary parametrisations [?, ?, ?, ?]. In these simulations, no S-wave or charm loop contributions are considered.

for the backgrounds – the combinatorial, partially reconstructed and double semi-leptonic components – is achieved by modelling the background q^2 distributions using second order Chebyshev polynomials, assuming factorisation with angles and B^0 mass. No other aspect of the background modelling procedure discussed in Sec. 7 is altered. The full pdf is given by

$$\begin{aligned} \mathcal{P}_{\text{toy}} = & \epsilon^{\text{RunI}}(q^2, \cos \theta_\ell, \cos \theta_K, \phi) \times \mathcal{P}_{\text{tot}}^{\text{RunI}}(m_{K\pi ee}, \Omega, q^2 | \lambda_{\text{phys}}, \lambda_{\text{nuis}}^{\text{RunI}}) + \\ & \epsilon^{\text{RunII}}(q^2, \cos \theta_\ell, \cos \theta_K, \phi) \times \mathcal{P}_{\text{tot}}^{\text{RunII}}(m_{K\pi ee}, \Omega, q^2 | \lambda_{\text{phys}}, \lambda_{\text{nuis}}^{\text{RunII}}), \end{aligned} \quad (48)$$

where $\epsilon(q^2, \cos \theta_\ell, \cos \theta_K, \phi)$ is the acceptance function discussed in Sec. 6, and

$$\begin{aligned} \mathcal{P}_{\text{tot}}(m_{K\pi ee}, \Omega, q^2 | \lambda_{\text{phys}}, \lambda_{\text{nuis}}) = & \mathcal{P}_{\text{S}}(m_{K\pi ee}) \times \mathcal{P}_{\text{S}}(\vec{\Omega}, q^2) + \\ & \mathcal{P}_{\text{B}}(m_{K\pi ee}) \times \mathcal{P}_{\text{B}}(\vec{\Omega}, q^2). \end{aligned} \quad (49)$$

The signal and background yields used in generation are obtained from a blinded data fit, and adjusted to remove the effects of weighting as discussed below. The sample size of each component is randomly varied based on poisson distributions with mean values given by the adjusted yields. The fits are then made using the default strategy, but without the inclusion of the control mode. Instead, shift and scaling parameters are fixed to generation values. Further details on the set-up are given in the sections below. Both the nominal and the reduced q^2 scenarios are considered. Table 25 shows the values of the angular observables associated with the pseudoexperiments, calculated from the amplitude model used in generation. In both cases, around 2000 pseudoexperiment fits are used to estimate the expected statistical uncertainties of the S and P-basis observables. For consistency, the reduced q^2 scenario is investigated using the toys generated in the nominal q^2 range with the removal of events that have q^2 values above $6.0 \text{ GeV}^2/c^4$. Table 26 provides the summary of sensitivity results along with the mean and widths of the pull distributions. Figures 72 to 79 show the parameter and pull distributions obtained for both q_c^2 ranges and S- and P-basis. Notice that we observe some sizeable bias in F_L and A_{FB} , with pulls biased at $10 - 15\%$, and to a less extend in the width of S_3 . Such regions (under the SM hypothesis) are close to the physics boundaries of the PDF, and therefore might suffer from undercoverage. In order to examine the behaviour of the PDF around this region, in the next version of the note a detailed study of the coverage as function of the generated values will be presented. The results of these studies will determine the size of correction expected for these observables.

	[1.1, 7.0] GeV ² /c ⁴	[1.1, 6.0] GeV ² /c ⁴
F_L	0.744	0.771
S_3	-0.019	-0.015
S_4	-0.177	-0.157
S_5	-0.251	-0.215
A_{FB}	0.066	0.028
S_7	0.0	0.0
S_8	0.0	0.0
S_9	0.0	0.0

Table 24: S-basis angular observable values associated with the pseudoexperiments.

	[1.1, 7.0] GeV ² /c ⁴	[1.1, 6.0] GeV ² /c ⁴
F_L	0.744	0.771
P_1	-0.146	-0.128
P'_4	-0.405	-0.374
P'_5	-0.576	-0.512
P_2	0.171	0.081
P'_6	0.0	0.0
P'_8	0.0	0.0
P_3	0.0	0.0

Table 25: P-basis angular observable values associated with the pseudoexperiments.

$1.1 < q_c^2 < 7.0$, SM				
	Sensitivity	Pull mean	Pull width	
F_L	0.0607 ± 0.0010	-0.155 ± 0.022	0.993 ± 0.016	
S_3	0.0635 ± 0.0010	-0.020 ± 0.024	1.091 ± 0.017	
S_4	0.0914 ± 0.0014	-0.060 ± 0.023	1.021 ± 0.016	
S_5	0.0768 ± 0.0012	-0.003 ± 0.023	1.026 ± 0.016	
A_{FB}	0.0557 ± 0.0009	0.108 ± 0.023	1.036 ± 0.016	
S_7	0.0793 ± 0.0013	0.002 ± 0.023	1.021 ± 0.016	
S_8	0.0959 ± 0.0015	0.039 ± 0.024	1.049 ± 0.017	
S_9	0.0597 ± 0.0009	0.042 ± 0.023	1.042 ± 0.017	
F_L	0.0598 ± 0.0009	-0.152 ± 0.022	0.989 ± 0.015	
P_1	0.513 ± 0.008	0.006 ± 0.022	1.013 ± 0.016	
P'_4	0.2145 ± 0.0033	-0.029 ± 0.022	0.999 ± 0.016	
P'_5	0.1862 ± 0.0029	0.069 ± 0.022	0.990 ± 0.015	
P_2	0.1500 ± 0.0023	0.065 ± 0.022	0.995 ± 0.015	
P'_6	0.1842 ± 0.0029	0.016 ± 0.022	1.011 ± 0.016	
P'_8	0.2222 ± 0.0035	0.031 ± 0.023	1.032 ± 0.016	
P_3	0.245 ± 0.004	-0.049 ± 0.022	0.999 ± 0.015	

$1.1 < q_c^2 < 6.0$, SM				
	Sensitivity	Pull mean	Pull width	
F_L	0.0681 ± 0.0011	-0.134 ± 0.023	0.986 ± 0.016	
S_3	0.0732 ± 0.0012	-0.009 ± 0.025	1.109 ± 0.018	
S_4	0.1054 ± 0.0018	-0.072 ± 0.024	1.029 ± 0.017	
S_5	0.0872 ± 0.0014	-0.016 ± 0.024	1.031 ± 0.017	
A_{FB}	0.0640 ± 0.0011	0.130 ± 0.024	1.044 ± 0.016	
S_7	0.0893 ± 0.0015	-0.010 ± 0.024	1.030 ± 0.017	
S_8	0.1095 ± 0.0018	0.035 ± 0.025	1.054 ± 0.018	
S_9	0.0679 ± 0.0011	0.033 ± 0.025	1.047 ± 0.017	
F_L	0.0665 ± 0.0011	-0.146 ± 0.022	0.981 ± 0.016	
P_1	0.736 ± 0.012	-0.013 ± 0.022	0.992 ± 0.016	
P'_4	0.266 ± 0.004	-0.021 ± 0.022	0.990 ± 0.016	
P'_5	0.230 ± 0.004	0.052 ± 0.022	0.984 ± 0.016	
P_2	0.2145 ± 0.0034	0.131 ± 0.022	0.972 ± 0.015	
P'_6	0.2190 ± 0.0035	0.011 ± 0.023	1.009 ± 0.016	
P'_8	0.271 ± 0.004	0.033 ± 0.023	1.025 ± 0.016	
P_3	0.332 ± 0.005	-0.042 ± 0.022	0.960 ± 0.015	

Table 26: Result of fits to around 2000 pseudoexperiments. The mean and widths of Gaussian fits to pull distributions for the S- and P-basis observables shown. The widths of the distribution of the parameter values from the fits are given as the expected sensitivities of the parameters.

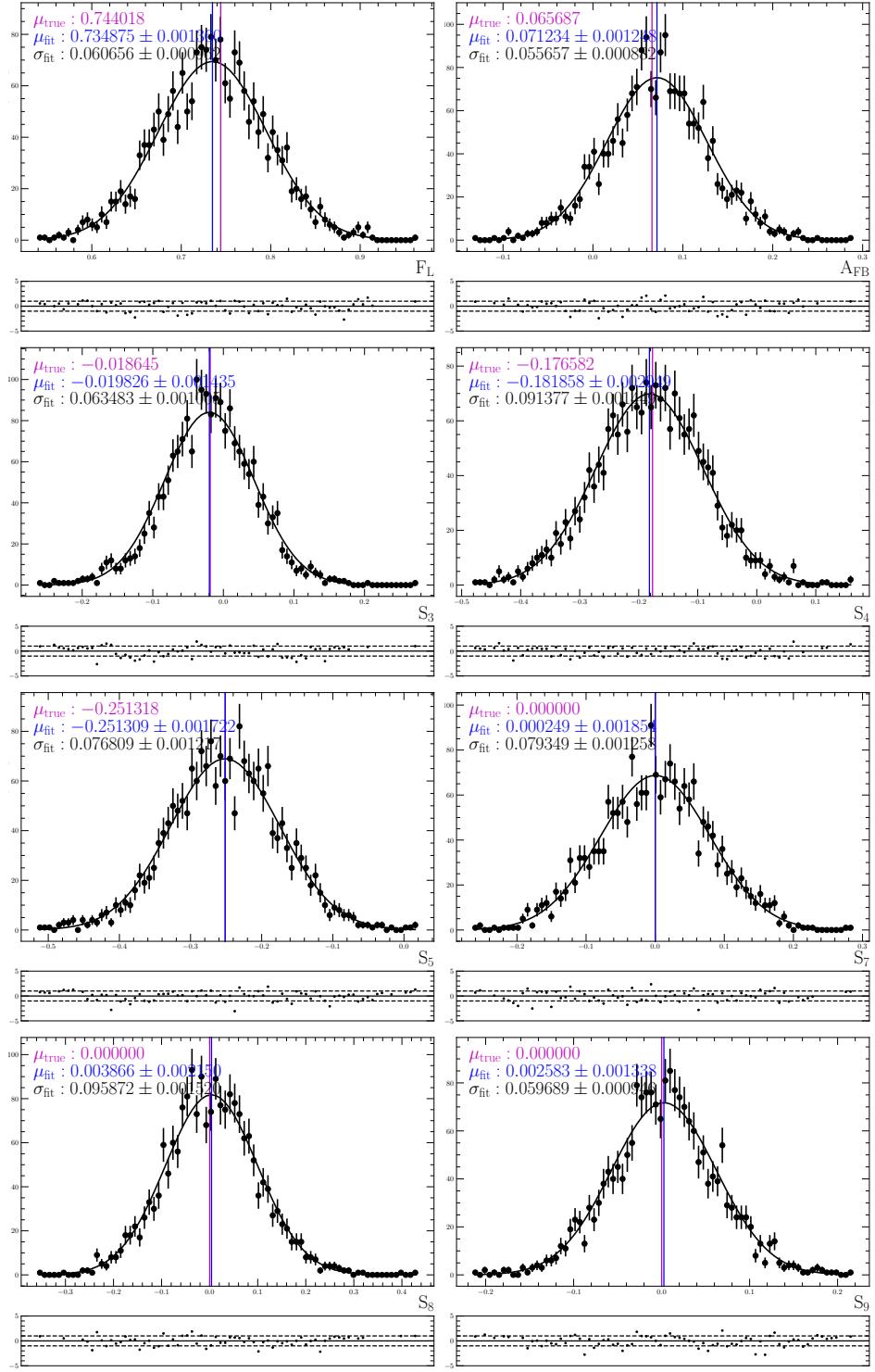


Figure 72: S-basis angular observable distributions for pseudoexperiments in the nominal q_c^2 range of $1.1 < q_c^2 < 7.0 \text{ GeV}^2/c^4$. A Gaussian fit is performed in these distributions and the difference to the generated values are reported in the plots.

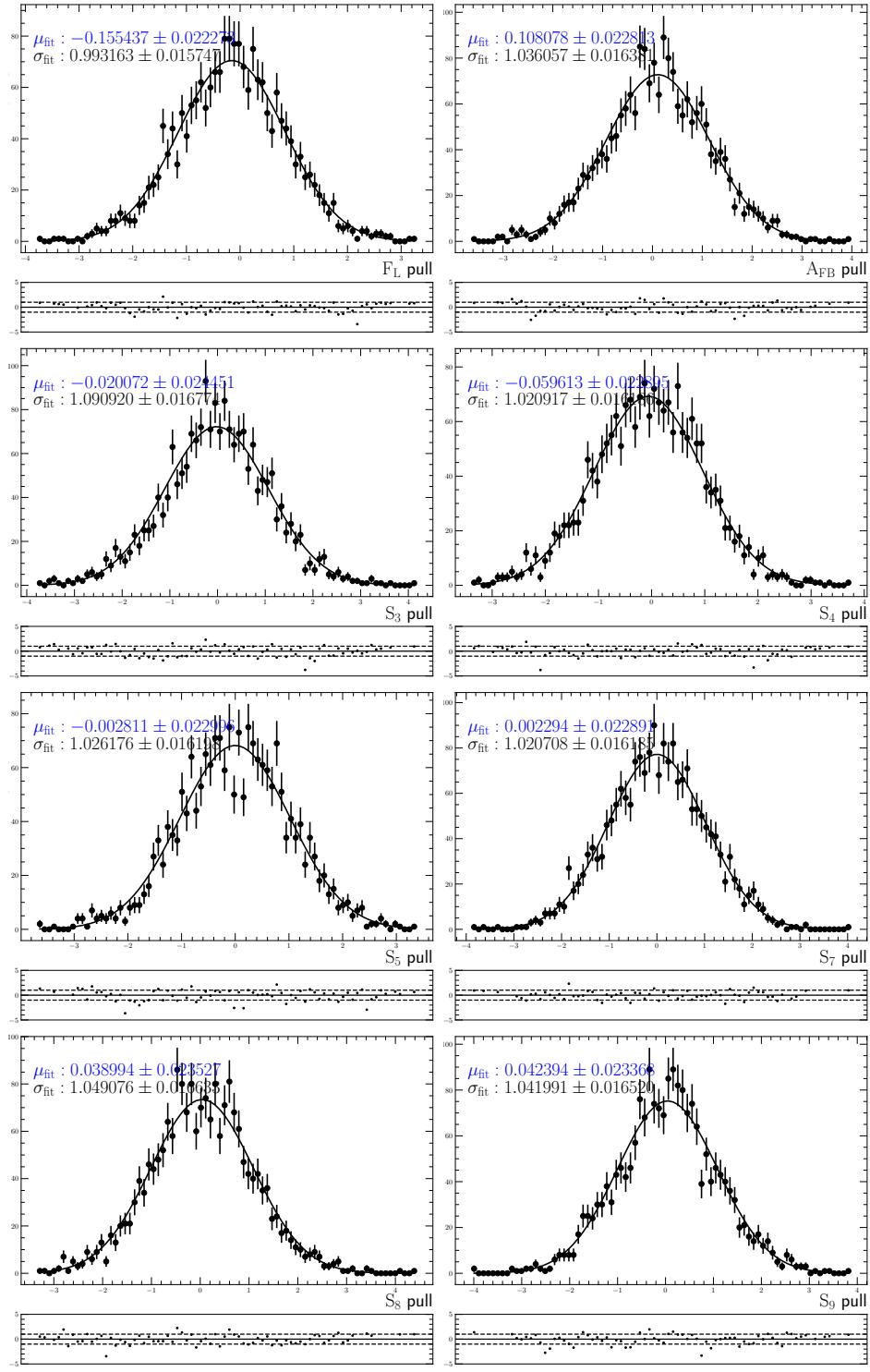


Figure 73: S-basis angular observable pull distributions for pseudoexperiments in the nominal q_c^2 range of $1.1 < q_c^2 < 7.0 \text{ GeV}^2/c^4$. A Gaussian fit is performed in these distributions and results are reported in the plots.

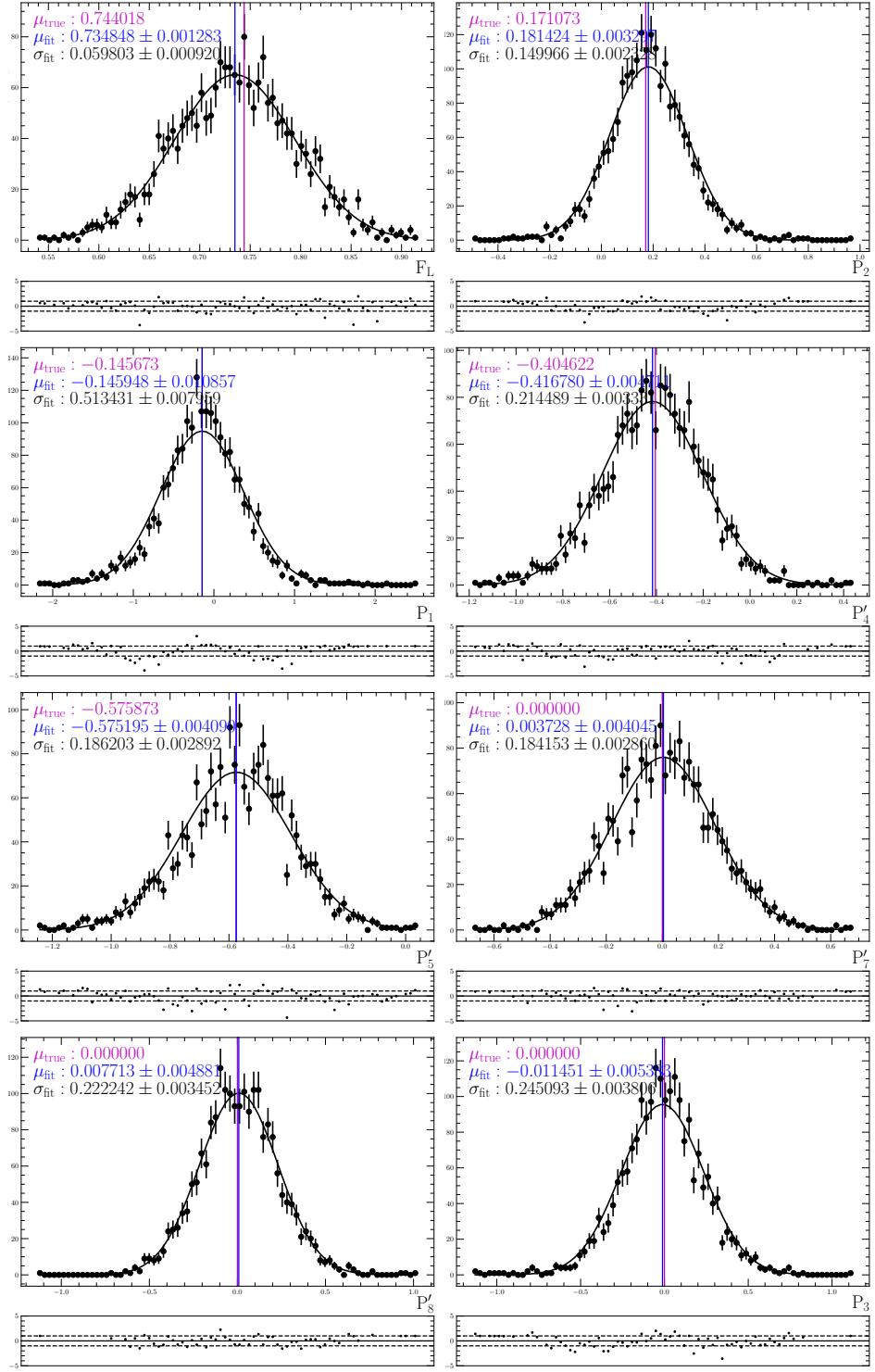


Figure 74: P-basis angular observable distributions for pseudoexperiments in the nominal q_c^2 range of $1.1 < q_c^2 < 7.0 \text{ GeV}^2/c^4$. A Gaussian fit is performed in these distributions and the difference to the generated values are reported in the plots.

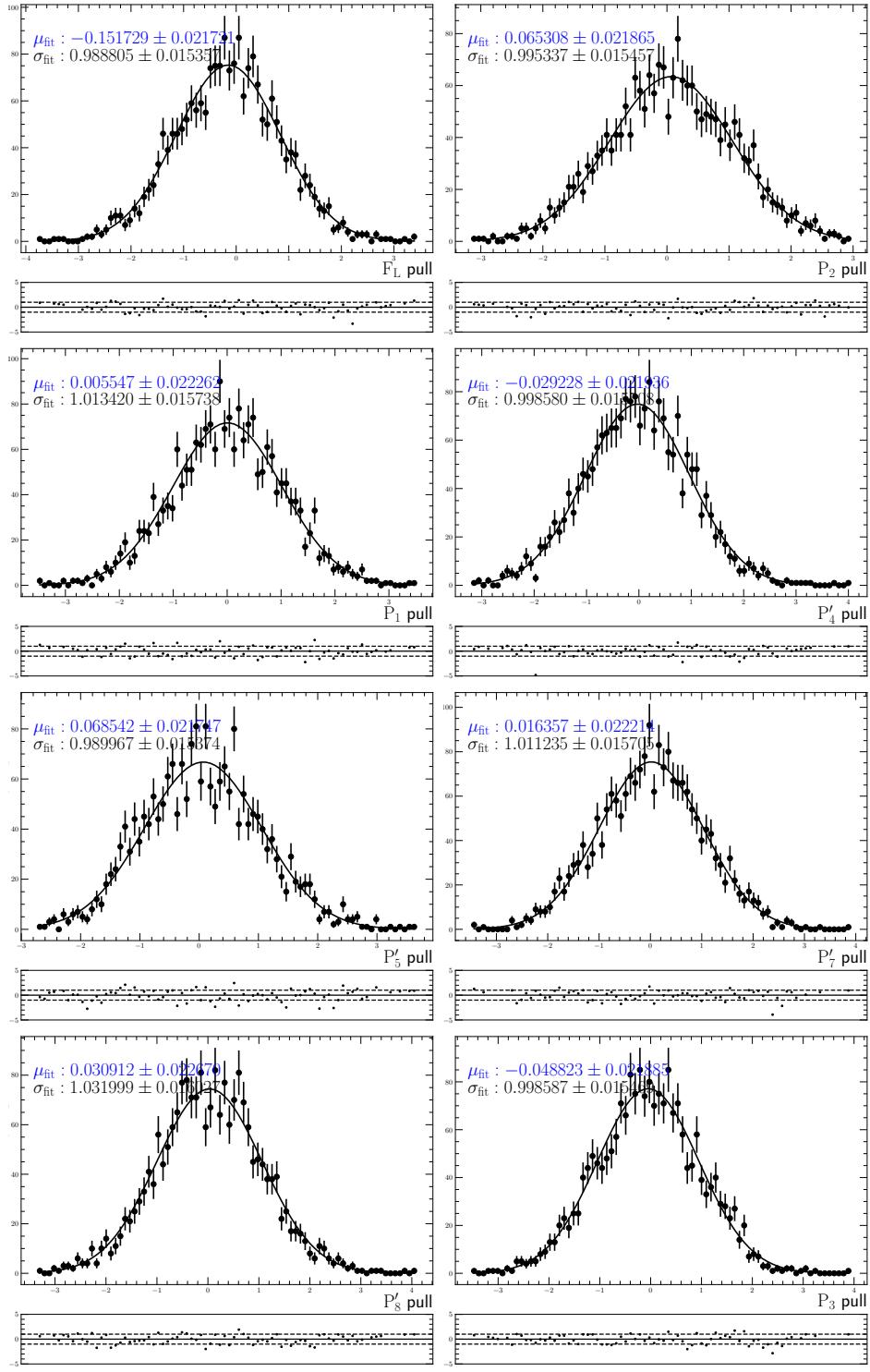


Figure 75: P-basis angular observable distributions for pseudoexperiments in the nominal q_c^2 range of $1.1 < q_c^2 < 7.0 \text{ GeV}^2/c^4$. A Gaussian fit is performed in these distributions and results are reported in the plots.

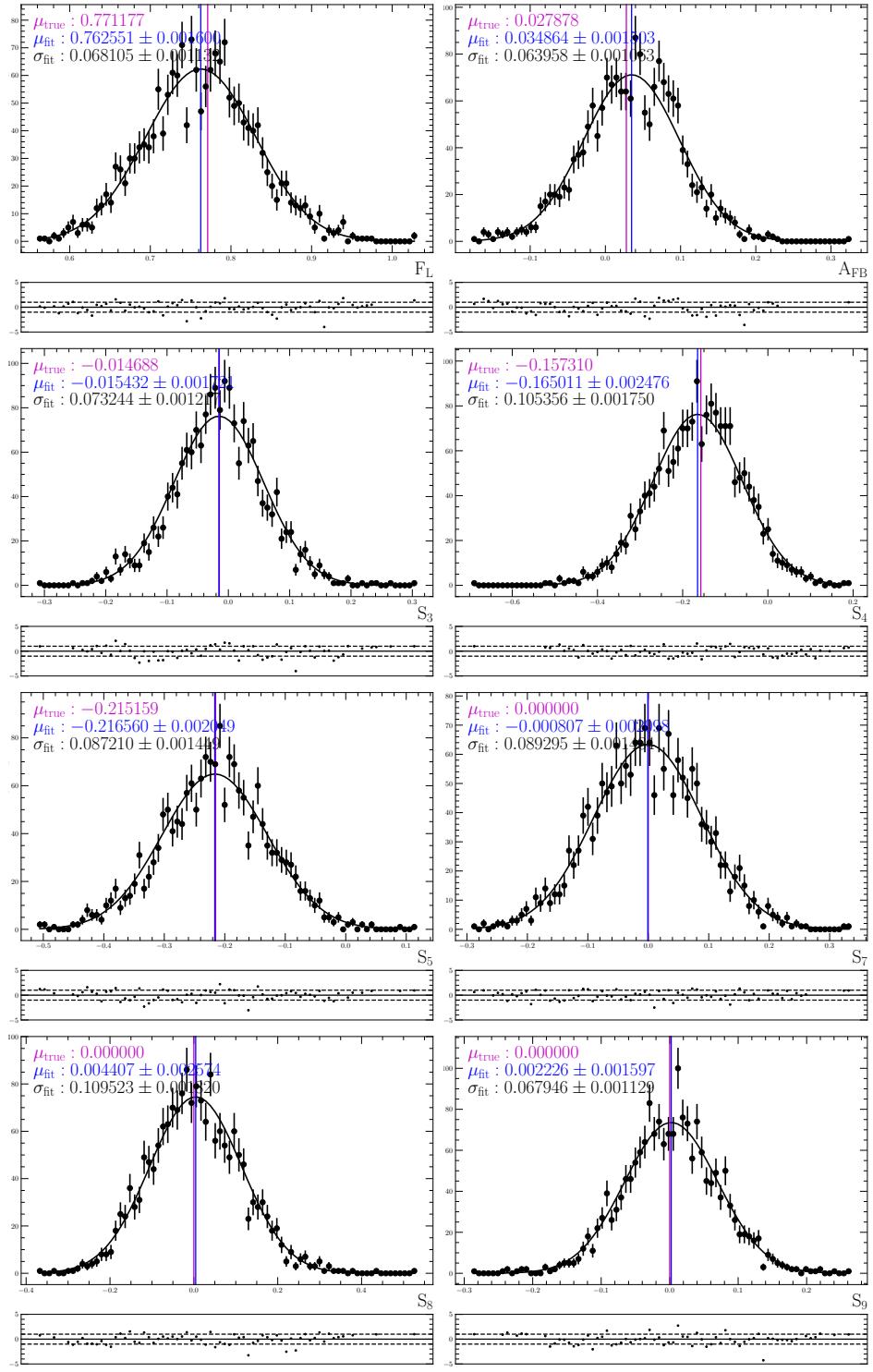


Figure 76: S-basis angular observable distributions for pseudoexperiments in the nominal q_c^2 range of $1.1 < q_c^2 < 6.0 \text{ GeV}^2/c^4$. A Gaussian fit is performed in these distributions and the difference to the generated values are reported in the plots.

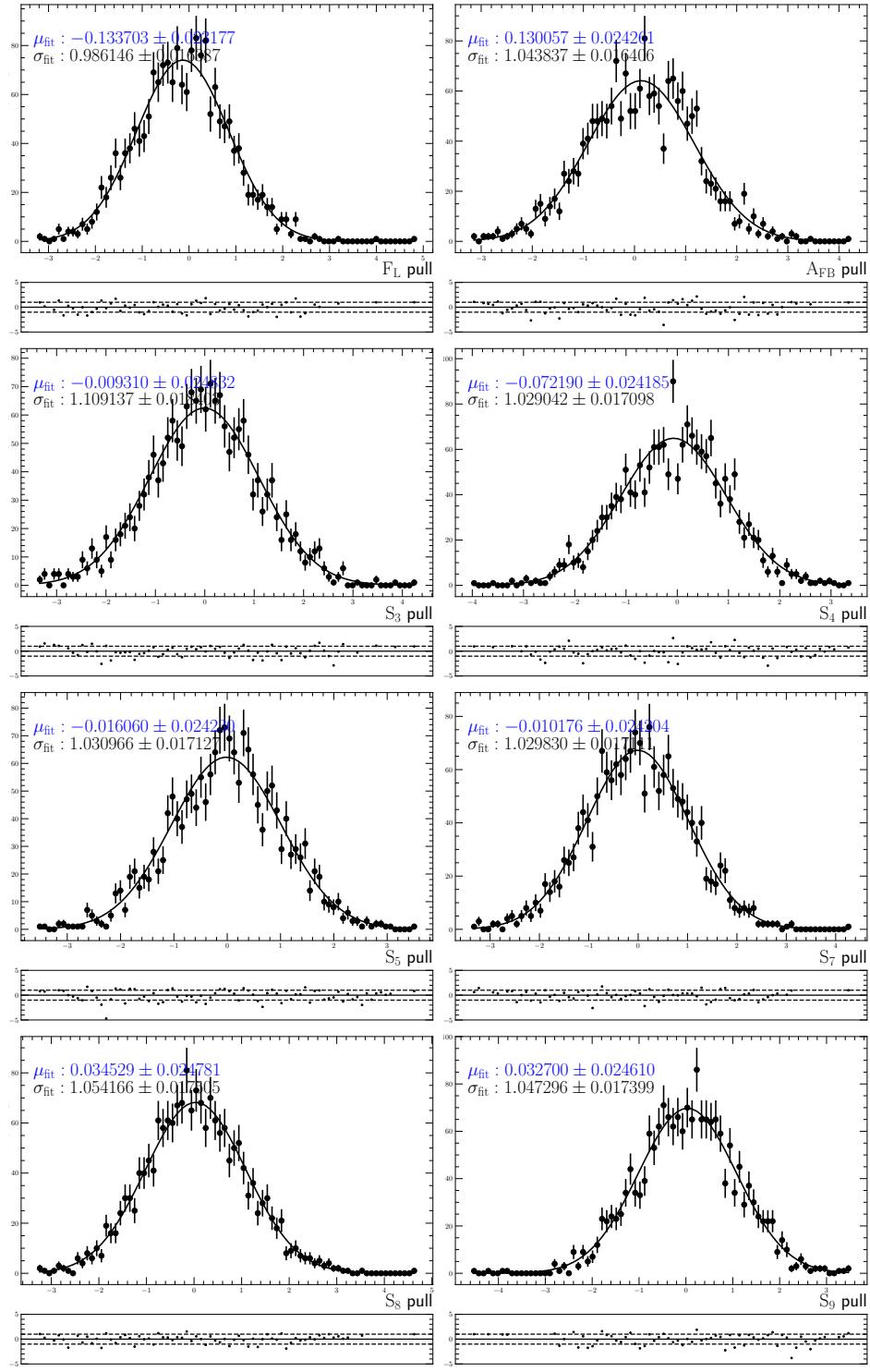


Figure 77: S-basis angular observable distributions for pseudoexperiments in the nominal q_c^2 range of $1.1 < q_c^2 < 6.0 \text{ GeV}^2/c^4$. A Gaussian fit is performed in these distributions and results are reported in the plots.

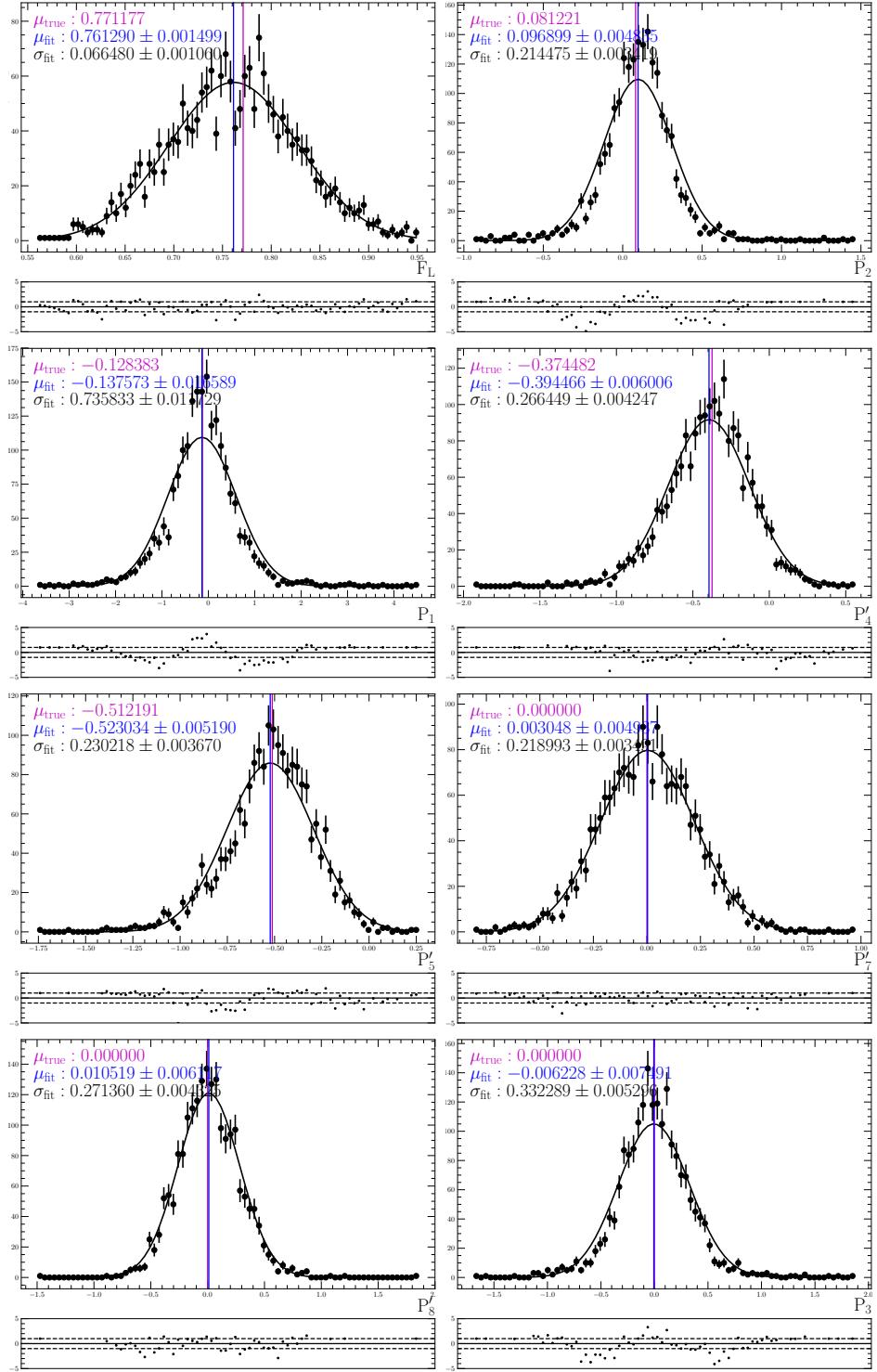


Figure 78: P-basis angular observable distributions for pseudoexperiments in the nominal q_c^2 range of $1.1 < q_c^2 < 6.0 \text{ GeV}^2/c^4$. A Gaussian fit is performed in these distributions and the difference to the generated values are reported in the plots.

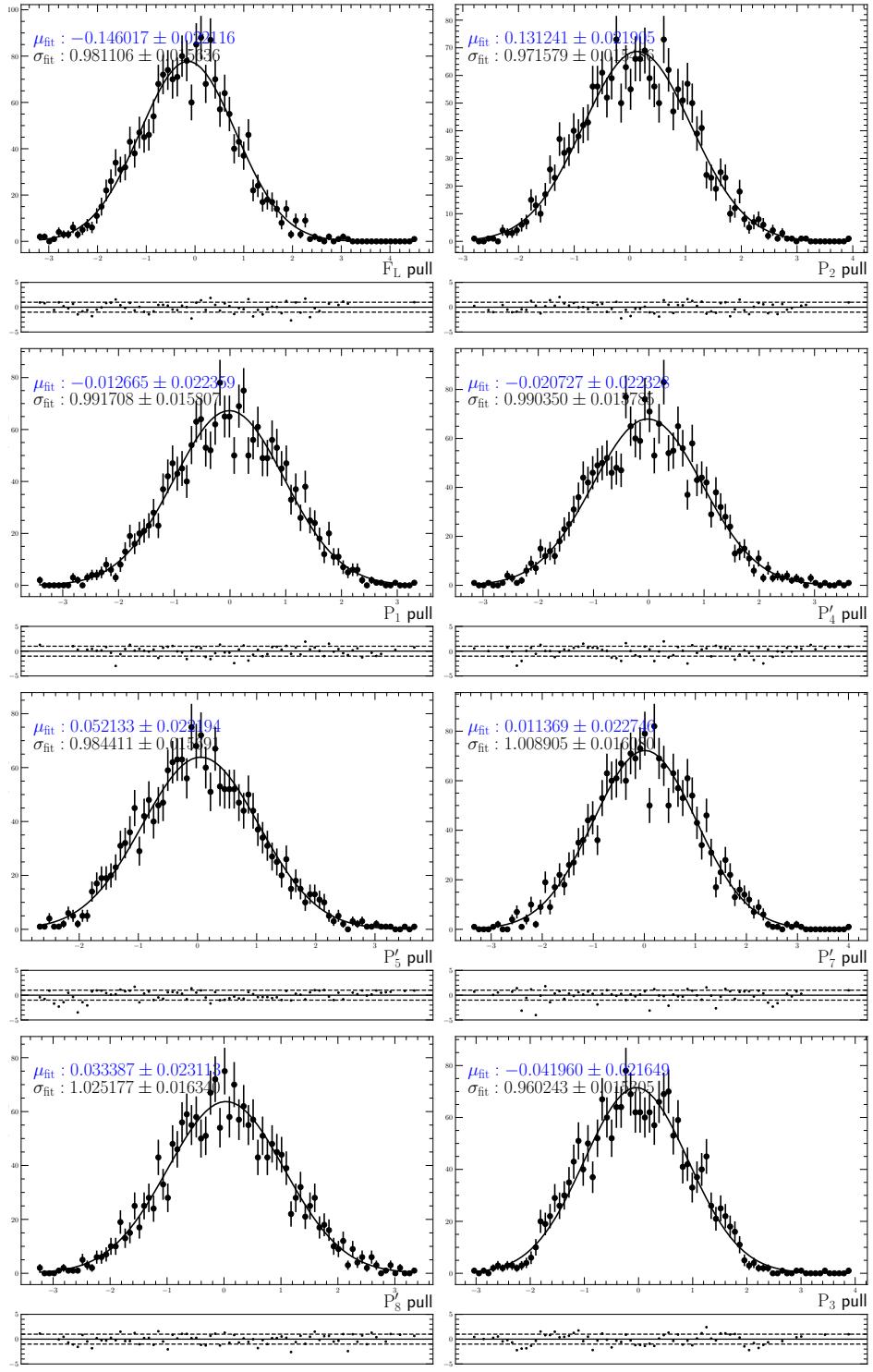


Figure 79: P-basis angular observable distributions for pseudoexperiments in the nominal q_c^2 range of $1.1 < q_c^2 < 6.0 \text{ GeV}^2/c^4$. A Gaussian fit is performed in these distributions and results are reported in the plots.

1746 **8.5.1 Component yields**

1747 To obtain estimates for signal and background yields, a simplified angular fit is performed
1748 to data where the fractions of DSL and partially reconstructed backgrounds with respect
1749 to signal are shared between Run-I and Run-II. That is, the fraction of weighted DSL/PR
1750 events in data is given by $f_S^{\text{RunI(RunII)}} \times f_{\text{DSL(PR)}}^{\text{wrt S}}$. The results of the fit, except the angular
1751 observable values which remain blinded, are given in Table 27, and the associated mass
1752 distributions are shown on Figure 80. As the measured weighted component fractions
1753 generally do not correspond to the fraction of actual events in data, as discussed in
1754 Appendix H, the expectation value of the weights for each component is estimated using
1755 toys. This is carried out by generating large statistics (200,000 events) toys with the
1756 background or signal model multiplied to the corresponding acceptance function, and
1757 then taking the average of the set of per-event correction weights obtained from the
1758 same acceptance function. The adjusted yields can then be calculated by dividing the
1759 weighted yield $f_{S(B)} \sum_{\text{all events}} w$ - which is (asymptotically) equal to $N_{S(B)} E(w_{S(B)})_{\text{True}}$ -
1760 by $E(w_{S(B)})_{\text{est}}$. The effect of this correction is small in all cases, as the expectation values
1761 of the weights are similar for all components. Both the direct and adjusted yields are
1762 summarised in Table 27. Note that the adjusted total number of events of 307 (394) for
1763 Run-I (Run-II) does not equal the actual total number of events in the samples, which
1764 are 312 and 385 for Run-I and Run-II, respectively. This is likely due to a combination of
1765 statistics (low sample size), and incorrect estimation of expectation values, e.g. due to the
1766 usage of factorised toys for the background components. However, since the differences
1767 are small compared to the statistical uncertainty of the fit, and is furthermore unlikely to
1768 cause significant changes in final sensitivity estimations, the decision is made to use the
1769 adjusted yields.

Parameter	Run-I value		Run-II value	
f_{sig}	0.36	± 0.05	0.41	± 0.04
$f_{\text{DSL}}^{\text{wrt S}}$	0.93	± 0.14		
$f_{\text{PR}}^{\text{wrt S}}$	0.18	± 0.08		
δ_μ	-9.0	± 0.5	-17.8	± 0.7
δ_σ	1.10	± 0.01	1.08	± 0.01
λ_{comb}	-0.0032 ± 0.0012		-0.0033 ± 0.0011	
$\lambda_{\text{comb}}^{J/\psi}$	-0.0040 ± 0.0002		-0.0041 ± 0.0001	

Table 27: Signal and background fractions and values of nuisance parameters obtained directly from the simplified data fits. The combinatorial fraction is calculated by subtracting all other component fractions from unity. The fractions of DSL and partially reconstructed components are fractions with respect to signal, and shared between both Runs. Those raw fractions need to be adjusted for pseudoexperiment generation.

Component	$E(w)$ estimation	Run-I		
		$f^{\text{fit}} \times \sum_{\text{evts}}(w)$	Adjusted yield	
Signal	16.88 ± 0.02	1924	± 269	114 ± 16
Combinatorial	17.69 ± 0.02	1336	± 442	75 ± 25
DSL	18.99 ± 0.03	1788	± 309	94 ± 16
PR	15.06 ± 0.01	347	± 166	23 ± 11
Estimated total			306	

Component	$E(w)$ estimation	Run-II		
		$f^{\text{fit}} \times \sum_{\text{evts}}(w)$	Adjusted yield	
Signal	16.94 ± 0.02	2878	± 296	170 ± 17
Combinatorial	17.53 ± 0.02	929	± 449	53 ± 26
DSL	19.00 ± 0.02	2675	± 249	141 ± 13
PR	15.01 ± 0.01	519	± 227	35 ± 15
Estimated total			399	

Table 28: Values used in the estimations of the unweighted component yields and the values of the adjusted yields obtained. The expectation values of the weights for the four components estimated using large statistics toys are multiplied to the fraction of weighted events in the sample (component fraction value times sum of weights of full sample) to produce the adjusted yields. The uncertainties on the adjusted yields include both uncertainties of the fit fractions and the statistical uncertainties of the expectation value estimations.

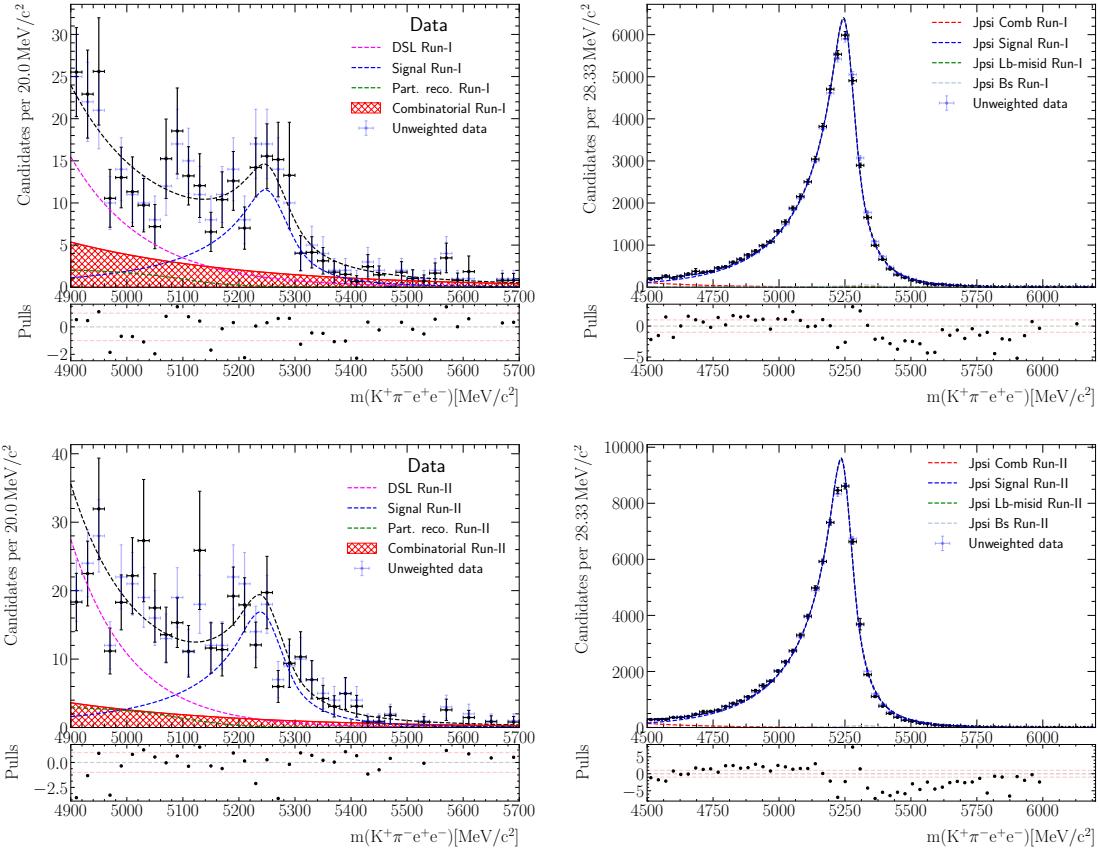


Figure 80: Result of simplified fit to data in the q^2 range of 1.1 to $7.0\text{GeV}^2/c^4$ used to obtain signal and background fractions and values of nuisance parameters for pseudoexperiments. For fit stability, the one set of fractions of the DSL and partially reconstructed components with respect to signal is used for both Runs.

1770 **8.5.2 Background models**

1771 The background models used for pseudoexperiments are obtained using the same procedure
 1772 as the nominal models, with the only difference being the addition of q^2 , which in all cases
 1773 are modelled by second order Chebyshev polynomials, and assumed to factorise with the
 1774 angles and B^0 mass. The modelling of the additional q^2 variable for the combinatorial,
 1775 DSL and partially reconstructed components is shown in Fig. 81 with the coefficients
 1776 shown in Tables 29, 30 and 31.

Parameter	Run-I value	Run-II value
$c_{1;\text{DSL}}^{q^2}$	0.19 ± 0.08	0.19 ± 0.07
$c_{2;\text{DSL}}^{q^2}$	0.02 ± 0.07	-0.05 ± 0.07

Table 29: Parameter values obtained from fits to step one of the two-step procedure to model DSL and combinatorial for the nominal q^2 range. For pseudoexperiment generation, q^2 is modelled in addition to the angles.

Parameter	Run-I value	Run-II value
$c_{1;\text{Comb}}^{q^2}$	0.175 ± 0.237	0.15 ± 2.51
$c_{2;\text{Comb}}^{q^2}$	0.188 ± 0.208	0.265 ± 0.184

Table 30: Parameter values obtained from fits to step two of the two-step procedure to model DSL and combinatorial for the nominal q^2 range. For pseudoexperiment generation, q^2 is modelled in addition to the angles.

Parameter	Run-I value	Run-II value
$c_{1;\text{PR}}^{q^2}$	-0.09 ± 0.03	-0.08 ± 0.03
$c_{2;\text{PR}}^{q^2}$	0.02 ± 0.03	0.01 ± 0.03

Table 31: Parameter values obtained from fits to $B^- \rightarrow K^+ \pi^- \pi^- e^+ e^-$ simulation reconstructed as the signal in the nominal q^2 range. For pseudoexperiment generation, q^2 is modelled in addition to the angles.

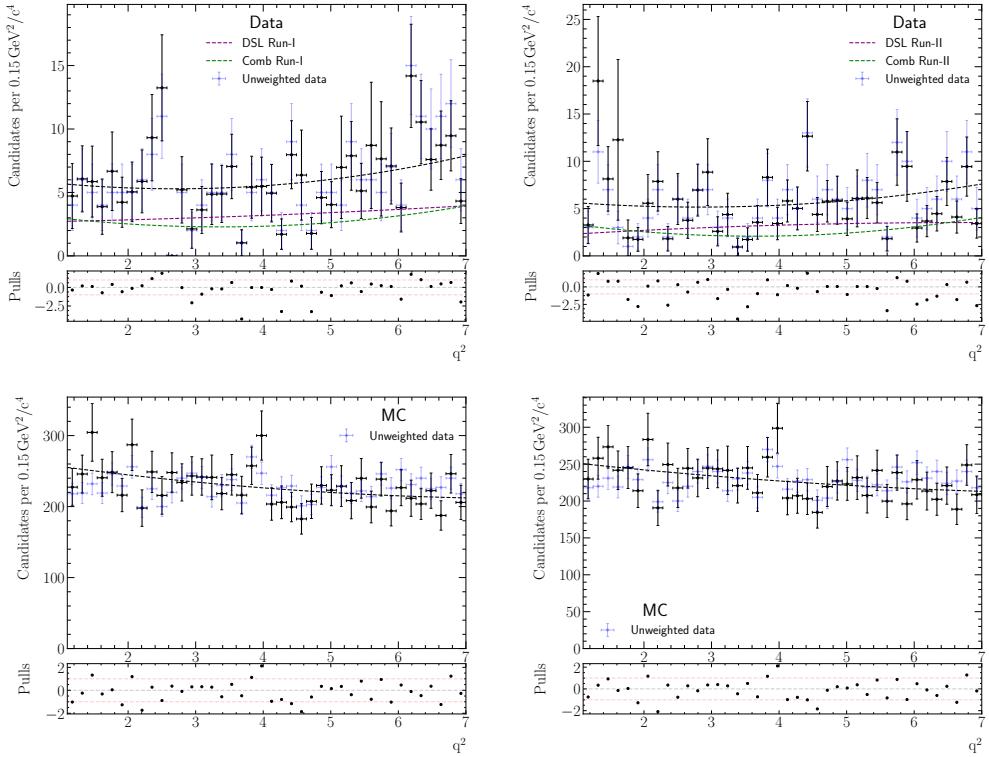


Figure 81: Top: Results of the second stage of the two-stage procedure to extract DSL and combinatorial models for the q^2 range of 1.1 to 7.0 GeV^2/c^4 for Run-I (left) and Run-II (right). Bottom: Results of fits to the $B^- \rightarrow K^+ \pi^- \pi^- e^+ e^-$ simulation reconstructed as the signal for 2016 (only available sample) in the q^2 range of 1.1 to 7.0 GeV^2/c^4 for the event distribution obtained with Run-I (left) and Run-II (right) acceptance weights. The acceptance weights are applied in addition to data-driven correction weights.

1777 9 Control mode fit validation

The scripts utilised in the study of the angular fit to the control mode channel (acceptance, data fit, systematics and cross-checks) are collected in

- 1778 ○ `ewp-bd2ksteeangular-central-q2/scripts/control/validation`

1780 The control mode channel $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ provides an important avenue for
1781 the validation of the angular fit strategy. This decay has the same final state particles as
1782 the rare mode, which leads to comparable selection and reconstruction effects. Since it
1783 features a tree-level transition, a significantly larger branching fraction in comparison to
1784 the rare mode is expected, which ensures the availability of large statistics for fit validation.
1785 The muonic counterpart $B^0 \rightarrow K^{*0} J/\psi (\rightarrow \mu^+ \mu^-)$ has been extensively studied in the past
1786 via multiple angular analyses, all of which found results consistent with the SM. These
1787 include measurements by LHCb [90], BaBar [91] and Belle [92, 93]. The LHCb angular
1788 analysis results obtained with 1 fb^{-1} of data taken in 2011, translated to the current basis,
1789 are given in Table 32. Besides published results, this mode was used for validation purposes
1790 as part of the angular analyses of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ [7]. The set of internal results from the
1791 most recent update is also given in the same Table. As LFU in the decay of the J/ψ is
1792 well established, the electron mode results can be compared directly against those of its
1793 muon counterpart.

1794 The angular distributions of the control mode final state particles can be described by
1795 the nominal signal pdf introduced in Sec. 5. Unlike the rare mode case, in which the lack
1796 of sensitivity to the S-wave contribution allows the use of a P-wave only pdf (see Eq. 24),
1797 the larger statistics of the control mode requires an explicit inclusion of the scalar terms
1798 as in the modified Eq. 27. Ideally the addition of the $K\pi$ mass as a fourth dimension to
1799 the fit would improve sensitivity to S-wave and S-P interference observables. This strategy
1800 is employed, for example, in the validation of the angular analyses of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$
1801 decays using the charmonium channel. This is nevertheless not implemented here to reduce
1802 differences between the control and the rare mode fit strategies. Another complication in
1803 data fits to $B^0 \rightarrow J/\psi K^+ \pi^-$ final states is the presence of exotic charmonium states [93],
1804 e.g. $Z_c^+(4200)$ and $Z_c^+(4430)$, which decay to the $J/\psi \pi$ pair. While this type of background
1805 significantly distort the low region of the $\cos\theta_K$ distribution, a good description of this
1806 contribution goes beyond the scope of this analysis. This is not considered to be a serious
1807 problem for the fit validation, as existing muon results are similarly affected and the
1808 impact is expected to be flavour universal.

1809 The main fit validation strategy involves making a full angular fit to control mode
1810 candidates selected in the same way as those used in the rare mode fit, with two exceptions:
1811 a cut on the constrained q^2 of $7.0 < q_c^2 < 11.0\text{ GeV}^2/c^4$ (centred around the known J/ψ
1812 invariant mass) and on the reconstructed B^0 mass with primary vertex and J/ψ mass
1813 constraints of $m_B^{J/\psi} > 5150\text{ MeV}/c^2$ to reduce partially reconstructed background. A larger
1814 mass window of $[4500, 6200]\text{ MeV}/c^2$ is also used. Due to low levels of remaining background,
1815 only two components are considered – misidentified Λ_b^0 decays and combinatorial – with

1816 the latter allowed to vary fully in the fit. The acceptance correction is made via per-event
 1817 weights obtained from three dimensional acceptance functions parametrised using the
 1818 control mode simulation. Systematic uncertainties associated with the acceptance functions,
 1819 which is expected to be a dominant source, are also evaluated. In addition, alternative
 1820 strategies related to background modelling and acceptance correction are explored to check
 1821 the sensitivity of the observables to those variations. Details and results of the main
 1822 validation fit, and those of the additional checks are provided in the following sections.

	LHCb 1fb^{-1} $m_{K\pi} \in [825.9, 965.9]\text{MeV}/c^2$	$B^0 \rightarrow K^* J/\psi(\rightarrow \mu^+ \mu^-)$		
		Simultaneous fit	Run-I only	2016 only
F_L	0.572 ± 0.008	0.5609 ± 0.0010	0.5577 ± 0.0014	0.5643 ± 0.0014
S_3	-0.013 ± 0.010	-0.0003 ± 0.0013	0.0001 ± 0.0018	-0.0008 ± 0.0018
S_4	-0.250 ± 0.006	-0.2489 ± 0.0014	-0.2518 ± 0.0020	-0.2459 ± 0.0020
S_5	0	0.0007 ± 0.0014	-0.0020 ± 0.0019	0.0035 ± 0.0020
A_{FB}	0	0.0005 ± 0.0009	0.0005 ± 0.0001	0.0005 ± 0.0012
S_7	0	0.0011 ± 0.0014	0.0005 ± 0.0020	0.0017 ± 0.0020
S_8	-0.048 ± 0.007	-0.0544 ± 0.0014	-0.0517 ± 0.0020	-0.0573 ± 0.0021
S_9	-0.084 ± 0.006	-0.0852 ± 0.0013	-0.0860 ± 0.0018	-0.0844 ± 0.0018
F_S	0.064 ± 0.010	0.0616 ± 0.0015	0.0645 ± 0.0021	0.0586 ± 0.0022
S_{S_1}		-0.2282 ± 0.0025	-0.2350 ± 0.0035	-0.2208 ± 0.0036
S_{S_2}		0.0221 ± 0.0016	0.0228 ± 0.0022	0.0214 ± 0.0023
S_{S_3}		0.0007 ± 0.0015	0.0027 ± 0.0021	-0.0013 ± 0.0021
S_{S_4}		0.0000 ± 0.0015	0.0007 ± 0.0021	-0.0007 ± 0.0022
S_{S_5}		-0.0671 ± 0.0016	-0.0687 ± 0.0023	-0.0655 ± 0.0024

Table 32: Summary table of external results discussed in this section. The first column contains the results of the amplitude fit to $B^0 \rightarrow K^{*0} J/\psi(\rightarrow \mu^+ \mu^-)$ [90], translated to the S-basis angular observables. The set of values shown correspond to the $m_{K\pi}$ range of 825.9 to 965.9 MeV/ c^2 . Note that the S-P-wave interference terms are absent as the fit over the extended $K\pi$ range is P-wave only. The F_s value is obtained by integrating over the S-wave fraction in four bins of $K\pi$ mass. The P-wave observables S_5 , A_{FB} and S_7 are zero as only tree-level amplitudes are included in the fit. The next three columns contain cross-check results obtained as part of the $B^0 \rightarrow K^* \mu^+ \mu^-$ angular analysis [7]. They include, from left to right, the simultaneous fit to the full data set (Run-I and 2016), and separate fits to Run-I and 2016 samples.

1823 **9.1 Acceptance**

1824 Similarly to the parametrisation of the rare mode acceptance, two simulation samples are
1825 available, *i.e.* **PHYS** $B^0 \rightarrow K^{*0} J/\psi(\rightarrow e^+e^-)$ and **FLATQ2** rare mode samples. Note that
1826 the **FLATQ2** acceptance functions is the same as the one used in the rare mode but extended
1827 to the q^2 region of interest. However for these validation studies the parametrisation
1828 of the acceptance based on the **PHYS** simulation is adopted whereas the **FLATQ2** is used
1829 as a systematics. This choice is made for several reasons. Firstly, there are known
1830 differences between the **FLATQ2** and **PHYS** acceptances from simulation studies, with the
1831 latter providing improved composite resolution (including FSR) and acceptance corrections.
1832 Therefore, in order to ensure comparability with existing muon mode results (which are
1833 minimally affected by resolution), it is preferable to use the **PHYS** acceptance. Secondly,
1834 the difference between **FLATQ2** and **PHYS** acceptances for the control mode may not match
1835 that of the rare mode, due to, for example, the application of an additional cut on the
1836 J/ψ constrained B^0 mass (and the usage of a different mass range), which is typically not
1837 applied in the acceptance parametrisation, and the difference in the true q^2 distribution.
1838 This *effective* correction shows some model dependence that, albeit negligible in comparison
1839 to the statistical uncertainty for the rare mode, might be of importance for this validation.
1840 Thirdly, for the control mode a three-dimensional (angles only) acceptance function should
1841 suffice, as the narrow resonance is analogous to taking a very small bin in q^2 , and this
1842 reduction in dimensionality, together with the larger statistics of the simulation samples,
1843 means that holes and very large corrections weights are not a problem when using the
1844 **PHYS** acceptances. In any case, differences between the **PHYS** and **FLATQ2** acceptances are
1845 checked in simulation and data in Sec. 9.4.2).

1846 The control mode acceptances used for the main validation fit are parametrised
1847 separately for the two **Runs** using the $B^0 \rightarrow K^{*0} J/\psi(\rightarrow e^+e^-)$ simulation with the full list
1848 of selection applied. For the generator-level parametrisation, orders of seven, six and ten
1849 are used for $\cos\theta_K$, $\cos\theta_\ell$ and ϕ , respectively, leading to a total of 616 coefficients, which
1850 are calculated using a sample of 1000030 events. For the post-selection sample, orders
1851 of five, four and six are used for $\cos\theta_K$, $\cos\theta_\ell$ and ϕ , respectively, leading to a total of
1852 210 coefficients that are determined using 210495 events for Run-I and 484453 events for
1853 Run-II. The projections of the generator and post-selection distributions are shown on
1854 Figures 82 and 83. Acceptance validation tests are performed using the control mode
1855 simulation for the (rare mode) **FLATQ2** acceptance and the control mode **PHYS** acceptance.
1856 Those two results are shown in Table 33. As expected, the rare mode acceptance correction
1857 results in worse performance compared to the control mode acceptance, in particular for
1858 F_L . Other observables are affected to a lesser extent.

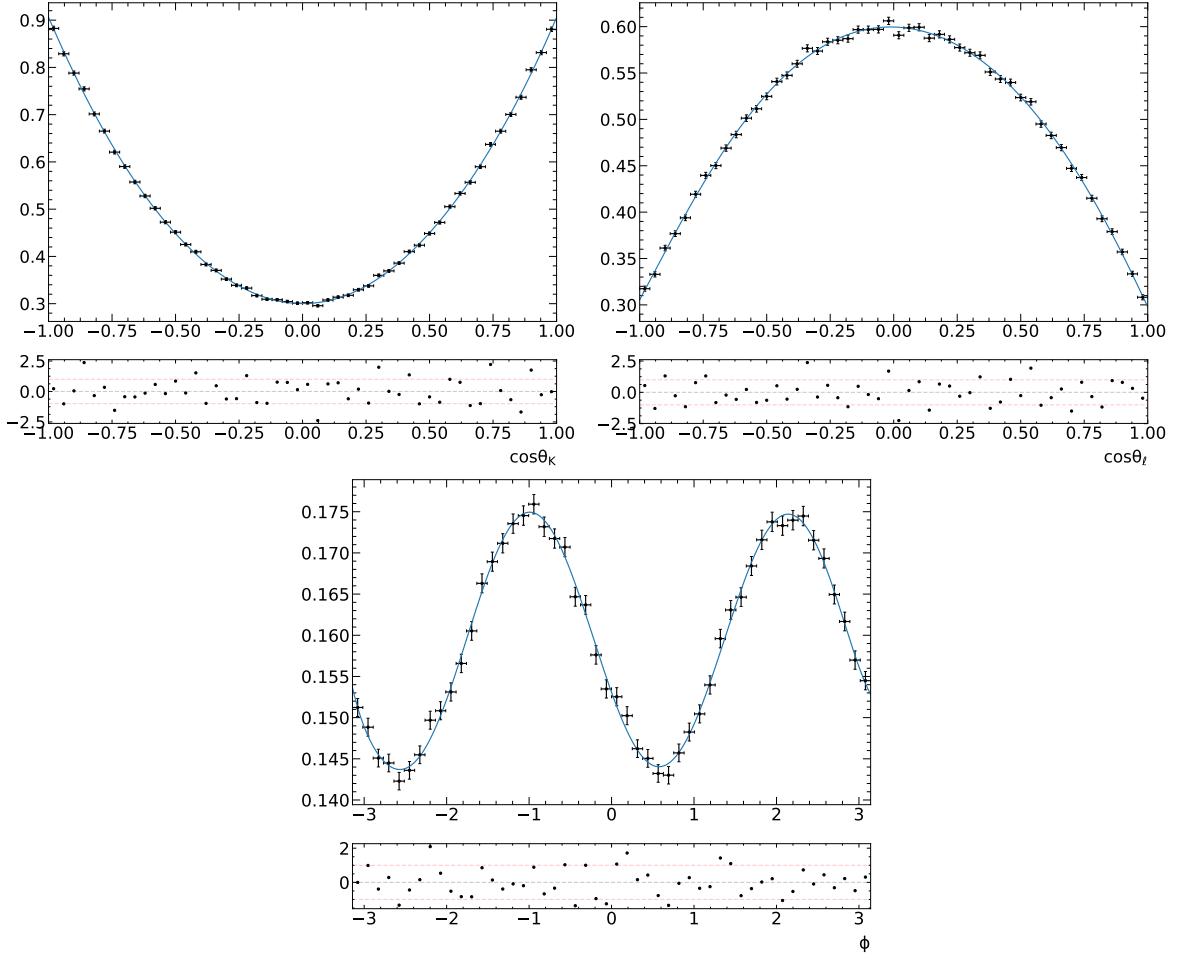


Figure 82: One-dimensional projections of the three-dimensional parametrisation of the control mode generation level.

	PHYS	$\Delta_{\text{Gen-PHYS}}$	FLATQ2	$\Delta_{\text{Gen-FLATQ2}}$	$\Delta_{\text{PHYS-FLATQ2}}$
F_L	0.5996 ± 0.0010	0.0008	0.6090 ± 0.0010	-0.0086	-0.0094
S_3	-0.0408 ± 0.0011	0.0009	-0.0397 ± 0.0011	-0.0001	-0.0010
S_4	-0.2135 ± 0.0013	-0.0015	-0.2165 ± 0.0013	0.0015	0.0030
S_5	-0.0004 ± 0.0013	0.0004	0.0032 ± 0.0013	-0.0032	-0.0036
A_{FB}	-0.0014 ± 0.0008	0.0014	-0.0005 ± 0.0007	0.0005	-0.0009
S_7	-0.0028 ± 0.0014	0.0028	0.0048 ± 0.0014	-0.0048	-0.0076
S_8	0.0361 ± 0.0014	0.0011	0.0389 ± 0.0014	-0.0017	-0.0028
S_9	-0.0883 ± 0.0011	-0.0004	-0.0860 ± 0.0010	-0.0028	-0.0023

Table 33: Results of fits to acceptance corrected post-selection control mode simulation. The acceptance correction is carried out separately for the two Runs using either FLATQ2 acceptance (for rare mode) or PHYS (parametrised from control mode simulation with matching selections). The three columns of differences show the difference between generation values and PHYS/FLATQ2 corrected fit result, and the difference between the results obtained using the two acceptances.

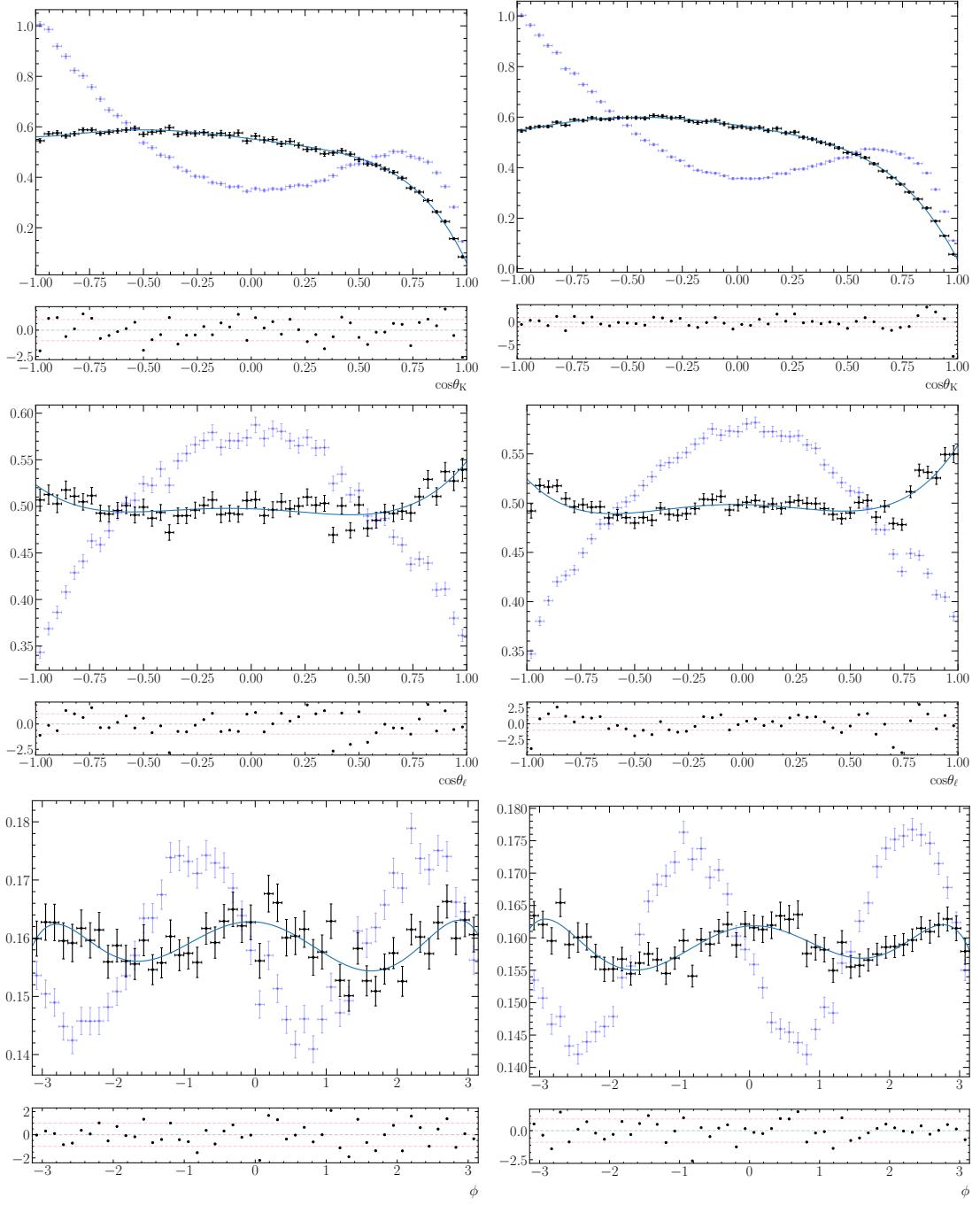


Figure 83: One-dimensional projections of the three-dimensional parametrisation of post-selection control mode physics simulation, with per-event weights obtained from the inverse of the function used to parametrise the generator-level sample, as well as kinematic and trigger correction weights. Run-I (Run-II) sample parametrisations are shown on the left (right) figures. The blue-dotted points correspond to the unweighted simulation whereas the black-dotted ones refer to the weighted simulation.

1859 **9.2 Angular data fit**

1860 A full angular fit to data is carried out on $B^0 \rightarrow K^{*0}J/\psi(\rightarrow e^+e^-)$ candidates selected
1861 in the same way as rare mode candidates, with the exceptions of the q^2 and mass ranges
1862 used ($7.0 < q_c^2 < 11.0$ and $[4500, 6200] \text{ MeV}/c^2$), and an additional cut to reduce partially
1863 reconstructed background of $m_B^{J/\psi} > 5150 \text{ MeV}/c^2$. Per-event correction weights are ob-
1864 tained from three-dimensional acceptance functions parametrised using the control mode
1865 simulation, as discussed in Section 9.1. Four components are considered in the fit – signal,
1866 $B_s \rightarrow K^{*0}J/\psi(\rightarrow e^+e^-)$, $\Lambda_b^0 \rightarrow pKJ/\psi(\rightarrow e^+e^-)$ with $p \rightarrow \pi$ misidentification, and com-
1867 binatorial. The signal mass distributions are modelled separately for each bremsstrahlung
1868 category; a double crystal ball function is used for the 0γ category while a DCB and a
1869 Gaussian for the other categories. For the angles, the modified angular signal pdf with six
1870 additional terms to describe the S-wave and the S-P-wave interference (Section 5, Eq. 27)
1871 is used. The B_s component is described by the same mass and angular pdfs as used for
1872 the signal, with the mean position of the mass peak displaced from that of the signal
1873 by $\delta_{B_s-B^0} = (87.2 \pm 0.3) \text{ MeV}/c^2$ [71]. Due to the worse resolution in the unconstrained
1874 invariant mass, this contribution is fixed to the relative expected fraction with respect
1875 to the B^0 mode obtained from simulation (Table 34). The Λ_b^0 component is modelled
1876 from phase-space $\Lambda_b^0 \rightarrow pKJ/\psi(\rightarrow e^+e^-)$ simulation with data-driven correction weights
1877 (from the pentaquark analysis). Here to reduce computational requirements, its mass
1878 distribution is modelled using a DCB rather than through a KDE, as is done for the
1879 nominal control mode mass fit. The angular distributions are modelled using first order
1880 Chebyshev polynomials (straight line) with the exception of $\cos\theta_K$, which is modelled using
1881 two Gaussians. The full Λ_b^0 model is shown on Figure 84. Similar to B_s , its contribution is
1882 fixed to the estimated relative fractions with respect to the signal in the fit. The combina-
1883 torial background is modelled using an exponential function for the mass and Chebyshev
1884 polynomials up to second order for the angles. All combinatorial shape parameters are
1885 allowed to vary. The choice of using a single combinatorial component to describe what is
1886 most likely a mixture of combinatorial, DSL and partially reconstructed events is motivated
1887 by the difficulty of separating these contributions reliably when the background level is
1888 very low. For example, the usage of separate DSL and combinatorial components makes
1889 the acquisition of good converging fits difficult unless first order Chebyshev polynomials
1890 are used for the combinatorial angular distributions. However, reducing the order is not
1891 ideal as partially reconstructed events may be present, and are expected to be non-linearly
1892 distributed. Alternative background treatments are nevertheless explored in Section 9.4.3.

1893
1894 A maximum likelihood fit is performed simultaneously for Run-I and Run-II samples.
1895 The parameters allowed to vary in the fit (besides the observables) are the shift and scaling of
1896 the mass peak, the slope of the exponential distributions, and the parameters of the angular
1897 combinatorial distributions. The results of the fit are shown in Figure 86 and Table 35. Most
1898 of the observables show good agreement with the results of the $B^0 \rightarrow K^{*0}J/\psi(\rightarrow \mu^+\mu^-)$
1899 validation fit, which are the most precise results currently available for comparison.
1900 However, a sizeable difference is seen for the longitudinal polarisation F_L . This observable

Parameters	Run-I	Run-II
f_{B_s}	0.0086	0.0079
$f_{\Lambda_b^0}$	0.0011	0.0004
$f_{\text{sig}}^{0\gamma}$	0.23	0.23
$f_{\text{sig}}^{1\gamma}$	0.50	0.50

Table 34: Parameters fixed in the data fit. The first two rows show the fixed fractions of $B_s \rightarrow K^{*0} J/\psi(\rightarrow e^+e^-)$ and mis-identified Λ_b^0 decays with respect to signal, and the last two rows give the bremsstrahlung category fractions of the signal mass model.

is particularly sensitive to the curvature of the $\cos\theta_K$ distribution, which is directly affected by acceptance corrections and backgrounds. While a full systematic assessment goes beyond the scope of these studies, some possible contributions that can impact the angular observables determination are examined in the following. Taking into account the systematic uncertainties associated with the acceptance function, notably from the simulation-data correction strategy, the discrepancy in F_L reduces to below 1.6σ . The impact of background modelling could further improve the agreement with the muonic channel. It is worth mentioning that the level of disagreement initially seen in F_L is at least a factor 5 smaller than the statistical uncertainty expected for the rare mode in this observable.

	Result	Difference (σ)
F_L	0.5452 ± 0.0026	5.6
S_3	-0.0044 ± 0.0029	1.3
S_4	-0.242 ± 0.004	-1.8
S_5	-0.0051 ± 0.0034	1.6
A_{FB}	-0.0021 ± 0.0021	1.1
S_7	-0.0087 ± 0.0034	2.6
S_8	-0.055 ± 0.004	0.2
S_9	-0.0826 ± 0.0030	-0.8
F_S	0.065 ± 0.005	-0.6
S_{S_1}	-0.226 ± 0.006	-0.4
S_{S_2}	0.030 ± 0.004	-1.9
S_{S_3}	0.0036 ± 0.0034	-0.8
S_{S_4}	-0.0052 ± 0.0035	1.4
S_{S_5}	-0.064 ± 0.004	-0.6

Table 35: Results of the control mode angular fit and differences with respect to the muon mode Run-I and 2016 simultaneous validation fit made as part of the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ angular analysis [7]. The differences refer to the number of standard deviations between corresponding results of the two sets calculated considering only statistical uncertainties.

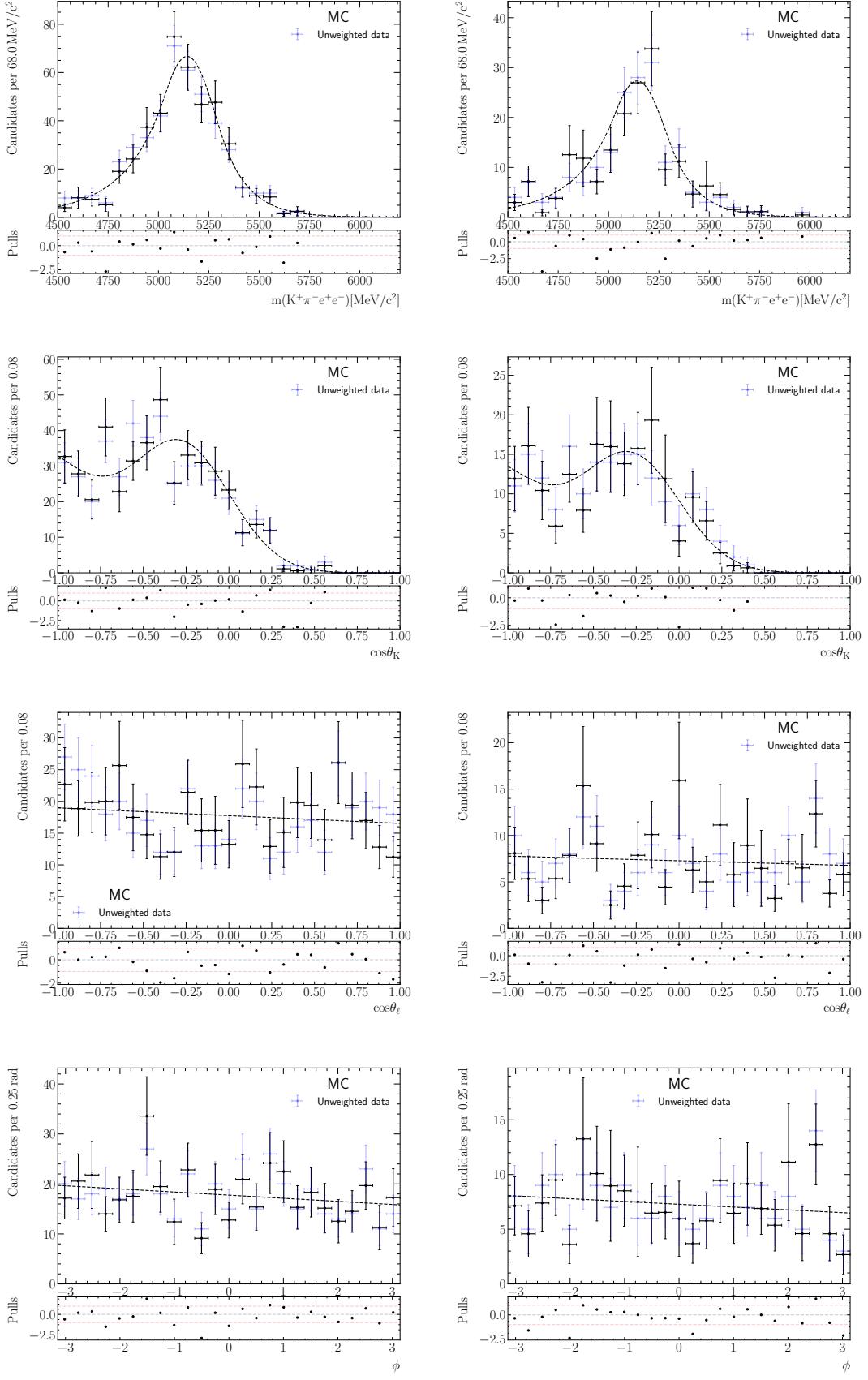


Figure 84: Mass and angular modelling used for the Λ_b^0 mis-id component. Due to low statistics, the shapes are extracted from a simultaneous fit to (left) Run-I and (right) 2016 simulation samples. Data-driven correction weights are used in addition to acceptance weights.

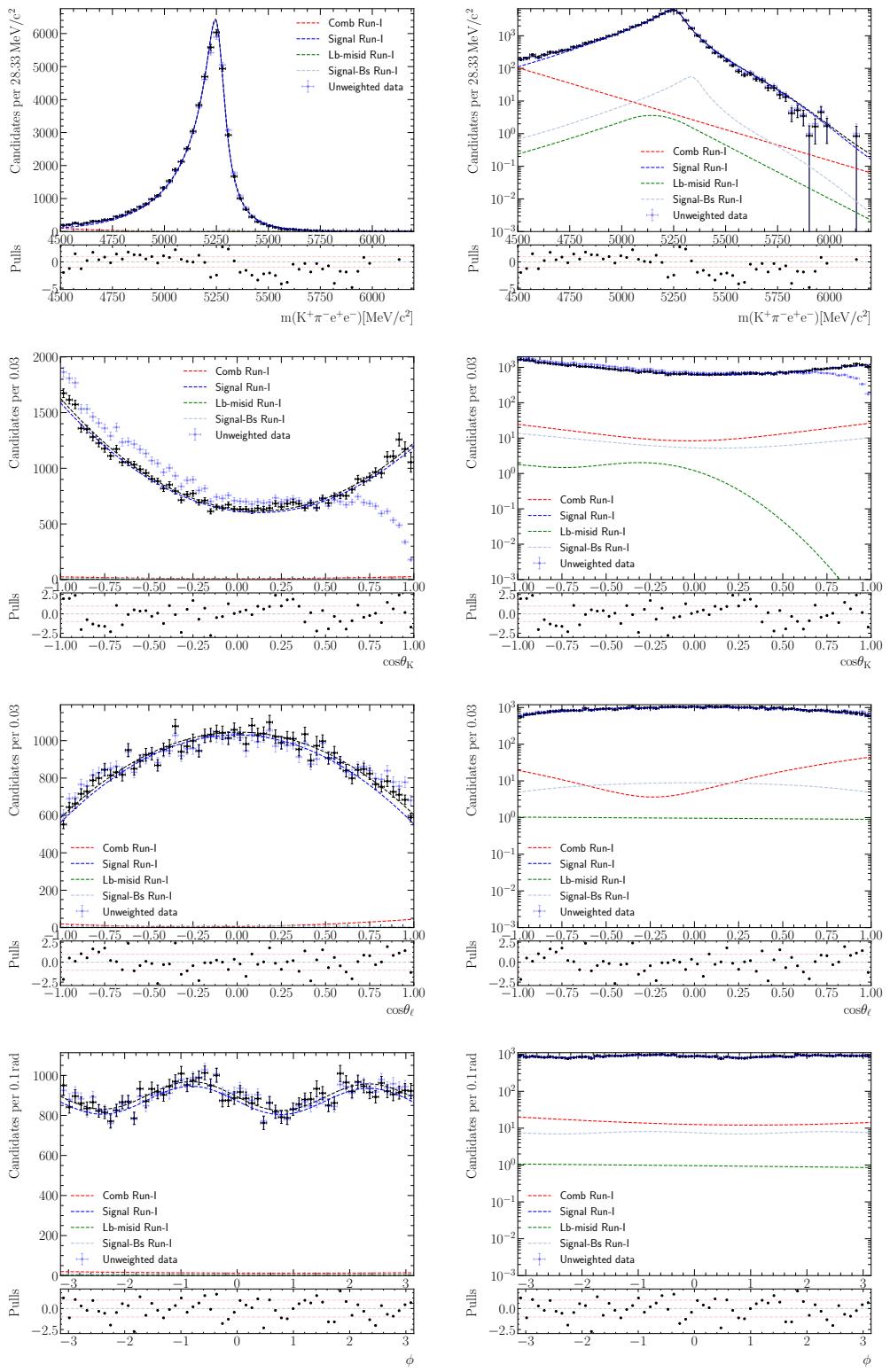


Figure 85: Results of the control mode angular fit with correction weights taken from the control mode acceptance for Run-I dataset. A single combinatorial component is used, but allowed to vary to account for remaining DSL and partially reconstructed events.

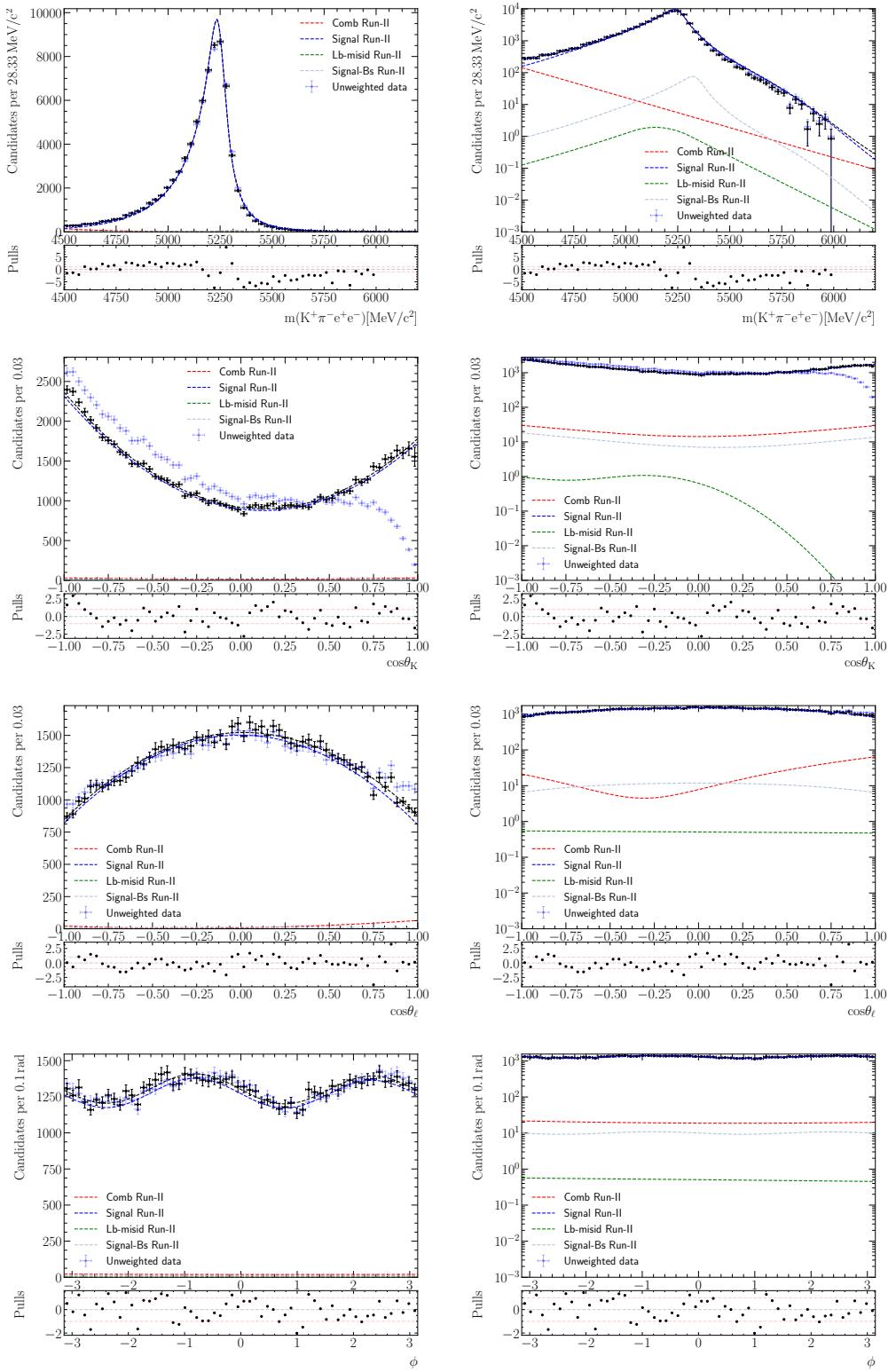


Figure 86: Results of the control mode angular fit with correction weights taken from the control mode acceptance for **Run-II** dataset. A single combinatorial component is used, but allowed to vary to account for remaining DSL and partially reconstructed events.

1911 **9.3 Systematics**

1912 Due to the large statistics of the control channel, in this section we focus the discussion on
1913 possible systematic uncertainties due to the acceptance function. A series of toy ensembles
1914 are generated in order to investigate four main sources of systematics:

- 1915 1. uncertainties of the calculated coefficients;
1916 2. usage of higher order polynomials in the acceptance parametrisation; and
1917 3. the choice of simulation-data correction strategy.

1918 Note that these checks are performed for the acceptance function derived from the
1919 $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ physics simulation, as the aim is to understand the contributing
1920 factors behind the discrepancy of some observables with existing results, rather than the
1921 difference between FLATQ2 and PHYS acceptances. Nevertheless, it is interesting to realise
1922 that the large disagreement between the two acceptances in Table 33 are also seen in the
1923 same observables.

1924 The studies are performed using signal only pseudoexperiments. Since a simultaneous fit
1925 is carried out using data from both Run-I and Run-II samples, each with its corresponding
1926 acceptance function, the ensembles are likewise generated using both acceptances with the
1927 same ratio of signal events as obtained in the weighted data fit ($N_{J/\psi}^{\text{Run-I}}/N_{J/\psi}^{\text{Run-II}} = 0.69$).
1928 Three hundred pseudoexperiments with a total of 1 M events (approximately eight times
1929 the data statistics) are generated to examine the four scenarios. Table 36 summarises
1930 the results for the different systematic uncertainties. While most of the uncertainties are
1931 marginal, the contribution from the alternative approach to the data/simulation corrections
1932 seems to have an important impact on F_L . Details on each of these sources are discussed
1933 in the following.

1934 **9.3.1 Coefficient uncertainties**

1935 To study the systematic uncertainty related to the uncertainties on the acceptance function
1936 coefficients, alternative sets of coefficients, and therefore alternative acceptances, are
1937 obtained at random based on the covariance matrices of the nominal Run-I and Run-
1938 II acceptances. That is, the coefficient values are drawn from a multivariate normal
1939 distribution, the mean of which are the coefficient values of the nominal acceptances, and
1940 the widths of which come from the covariance matrices. In total, three hundred randomised
1941 acceptances are used for each Run. Those acceptance functions are multiplied to the
1942 signal pdf to generate pseudoexperiments with acceptance effect. Fits are then made using
1943 either the matching randomised acceptances, or the nominal versions. The differences in
1944 observable values obtained using random versus nominal acceptance correction is then
1945 taken to be the systematic uncertainty associated with coefficient uncertainties. The
1946 distribution of observable differences are shown in Figure 87.

	Coefficient uncertainties	Higher orders	Alternative correction	Total
F_L	0.001 248	0.001 808	0.011 512	0.011 719
S_3	0.001 363	0.000 262	0.002 267	0.002 658
S_4	0.001 470	0.000 083	0.002 648	0.003 030
S_5	0.001 485	0.000 317	0.000 011	0.001 518
A_{FB}	0.000 848	0.000 029	0.000 022	0.000 849
S_7	0.001 261	0.000 074	0.000 233	0.001 285
S_8	0.001 520	0.000 311	0.001 541	0.002 187
S_9	0.001 365	0.001 298	0.000 180	0.001 892
F_S	0.002 431	0.002 473	0.000 673	0.003 533
S_{S_1}	0.003 346	0.004 691	0.029 508	0.030 066
S_{S_2}	0.001 764	0.000 348	0.003 213	0.003 682
S_{S_3}	0.001 567	0.000 200	0.000 071	0.001 582
S_{S_4}	0.001 565	0.000 061	0.000 336	0.001 601
S_{S_5}	0.001 580	0.000 327	0.000 263	0.001 635

Table 36: Summary of systematic uncertainty associated with the control mode acceptance functions. The systematic uncertainty corresponding to the coefficient uncertainties are taken to be the width of the distribution of differences between the results of the fit using nominal and alternative corrections. The other values are taken to be the mean of the difference distribution. The total is obtained by summing the contributions of the three sources in quadrature.

9.3.2 Higher orders

The orders of the polynomials used for the nominal acceptance are chosen as the lowest orders that show satisfactory description of the samples to be parametrised. The usage of higher orders could allow for improved parametrisation of certain regions, such as the edge of $\cos\theta_K$. However, it is avoided in the nominal procedure due to the limited statistics of the available samples, and the negligible impact of mis-modelling as tested through the retrieval of generator level values from post-acceptance simulation. Nevertheless, the choice of the order is a source of systematic uncertainty that could potentially impact the F_L determination, since it is highly sensitive to the corrections of the edge of $\cos\theta_K$. To evaluate this source of systematic uncertainty, alternative acceptances are parametrised using orders of eight, seven and nine for $\cos\theta_K$, $\cos\theta_\ell$ and ϕ , respectively. This is an increase of three from the nominal choice for all angles. Pseudoexperiments are then generated with the alternative set of acceptances and fitted with this parametrisation and the nominal one. The results are shown in Figure 88. Due to the mismatch in the acceptances, large weights can occur in regions of the phase space where the generation acceptance has larger values than that of the nominal, as illustrated for a selected region in Fig. 89. This is a problem in particular for the edge of $\cos\theta_K$, and is responsible for the non-Gaussian distribution of the differences. Excluding events within the problematic region (or vetoing events with weights larger than an arbitrary common value) results

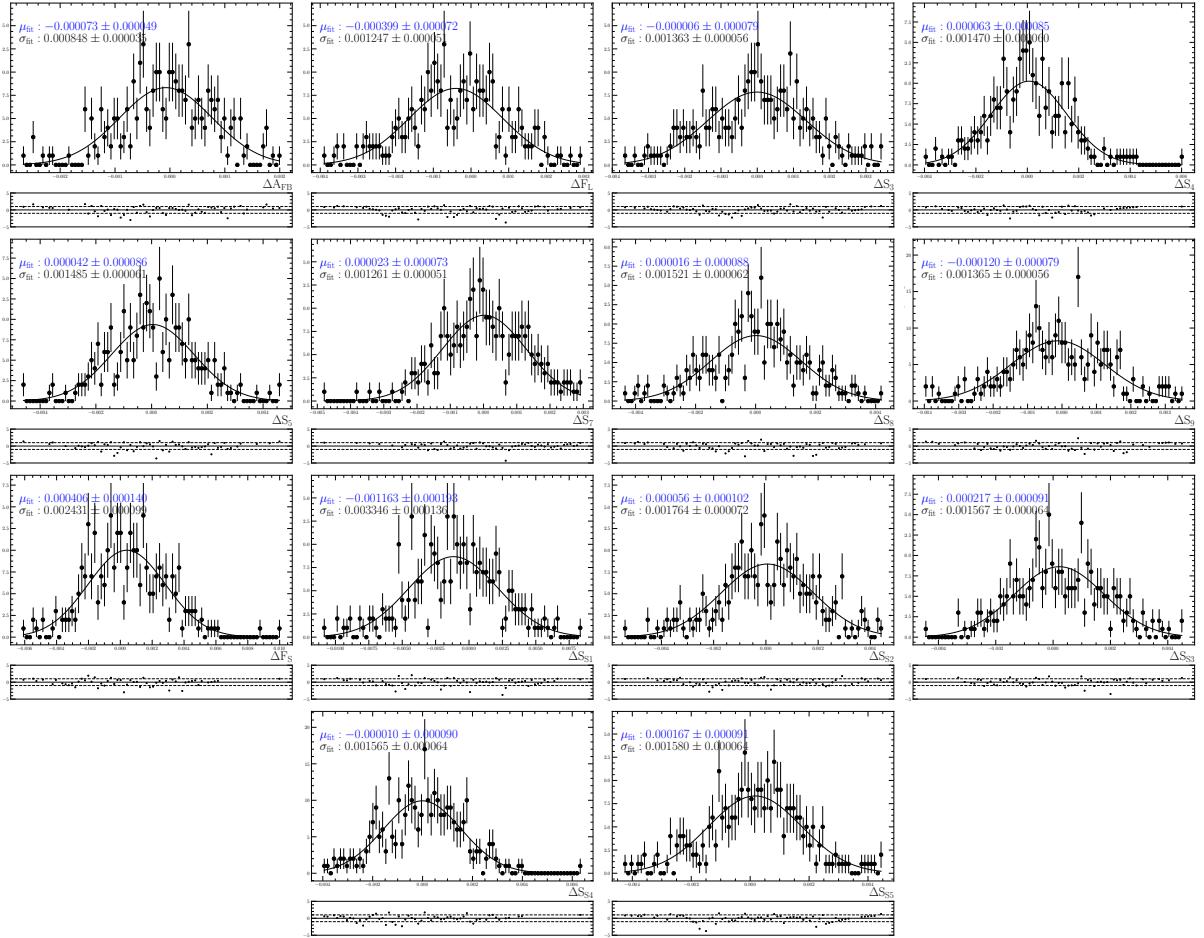


Figure 87: Gaussian fits to the difference between the nominal and alternative (randomised from the covariant matrix of the Legendre polynomials) acceptances on the angular observables returned by the toy study. The corresponding observable is labelled on each plot.

in more standard behaviour as illustrated on Figs. 90 and 91. Values given in summary table 36 correspond to the direct fit results (Figure 88) without the removal of problematic events.

9.3.3 Correction weights

Several elements in the data/simulation corrections can be examined as a source of systematic uncertainty. However, in order to broadly evaluate the impact of such corrections, an alternative methodology is verified. The nominal correction strategy has been designed as a hybrid approach combining the muonic angular analysis [7] and the LFU ratio measurements [51, 52] corrections chain. Minor differences with respect to the $R_{K,X}$ analyses are expected due to the Trigger corrections, while the PID (*e.g.* Meerkat vs PIDCalib table or fit-and-count) and kinematic corrections have a more relevant effect

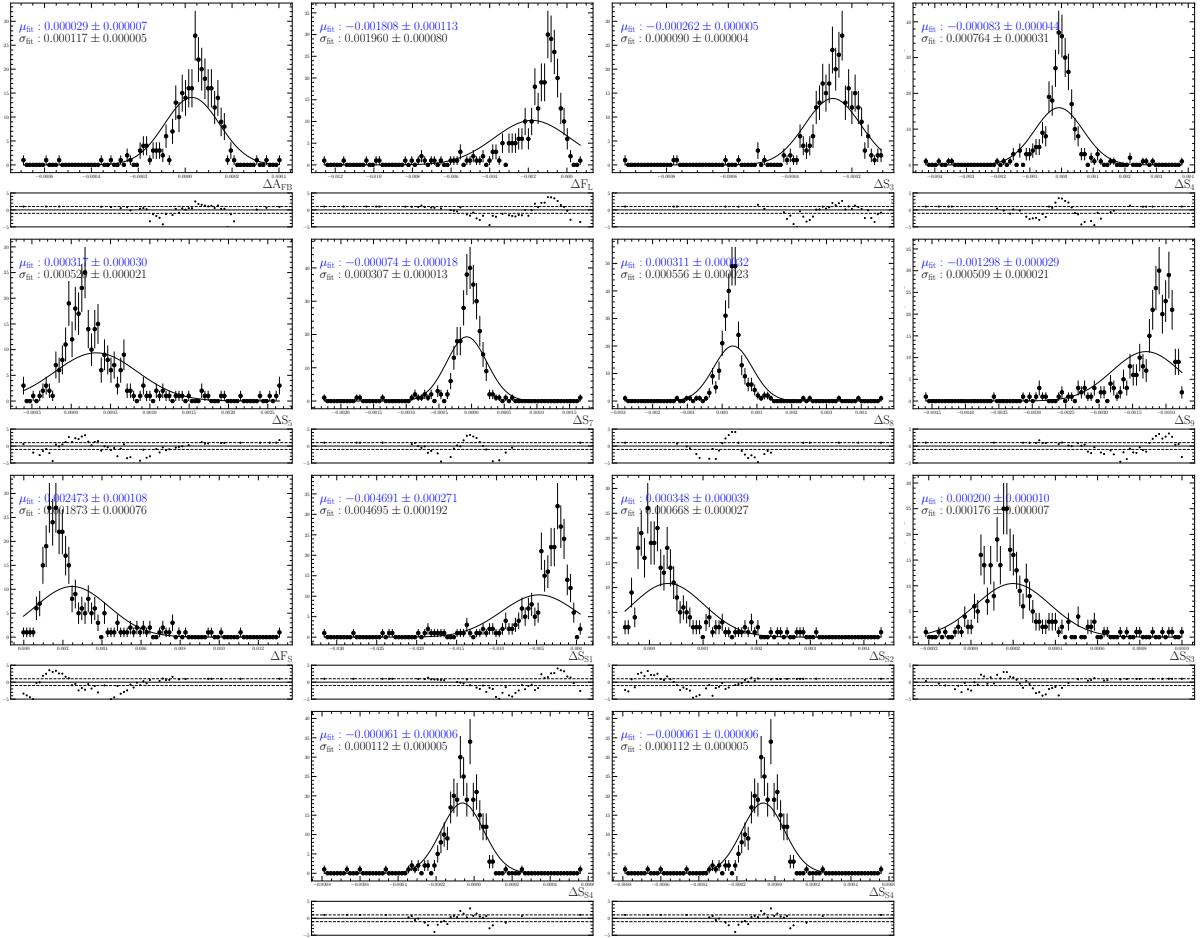


Figure 88: Gaussian fits to the difference between the nominal and alternative (using higher order polynomials) acceptances on the angular observables returned by the toy study. The corresponding observable is labelled on each plot.

in the angular distributions. Figure 92 shows the impact of the R_X alternative chain of corrections. In order to assess the systematic related to the choice of the nominal strategy, pseudoexperiments are generated with acceptance functions parametrised with the alternative weights, and fitted with either the alternative, or the nominal set. The difference between the two is taken as the systematic uncertainty. The results shown on Figure 93 indicate that this source of systematic uncertainty is especially large for F_L (and S_{s1}).

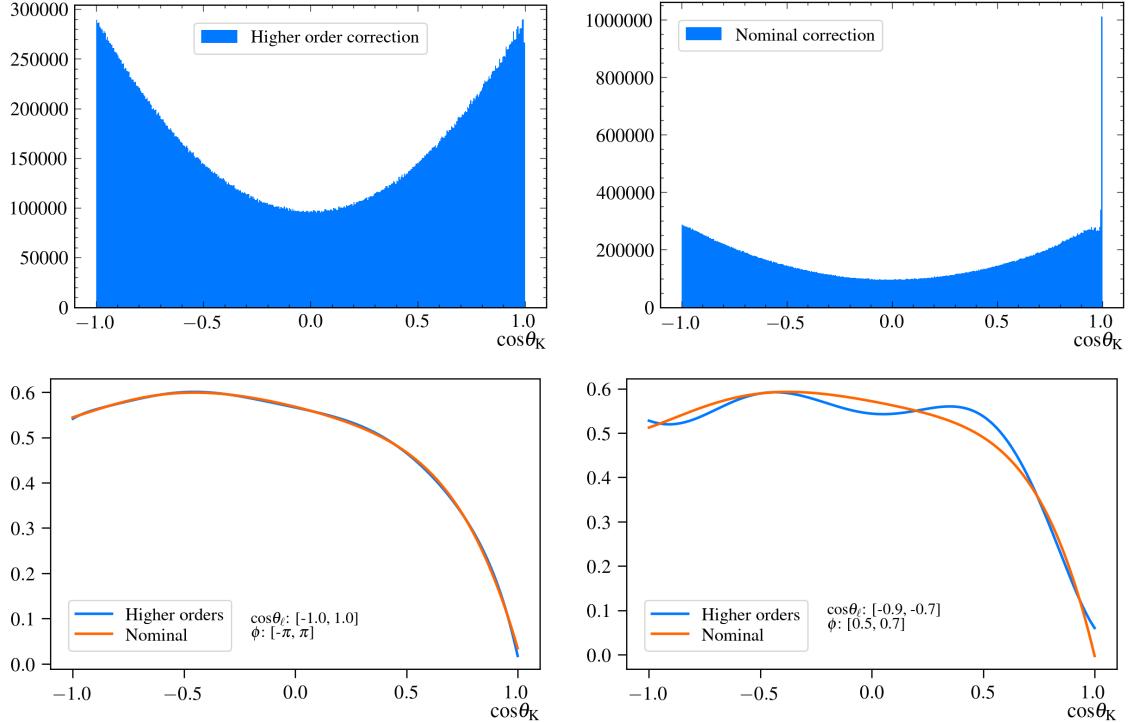


Figure 89: Illustration of the appearance of large weights in pseudoexperiments generated with acceptances parametrised with higher orders. The plots in the first row show the effect of using correction weights from the corresponding higher orders acceptance (left), and from the nominal acceptance (right). The projection of the nominal versus the higher order acceptance function (bottom left) does not show visibly the cause of the large weights. However, by integrating in $\cos\theta_\ell$ and ϕ , it is possible to see the region near $\cos\theta_K = 1$ where the higher order acceptance function has values that exceed that of the nominal acceptance, leading to events that would acquire large correction weights when the latter is used.

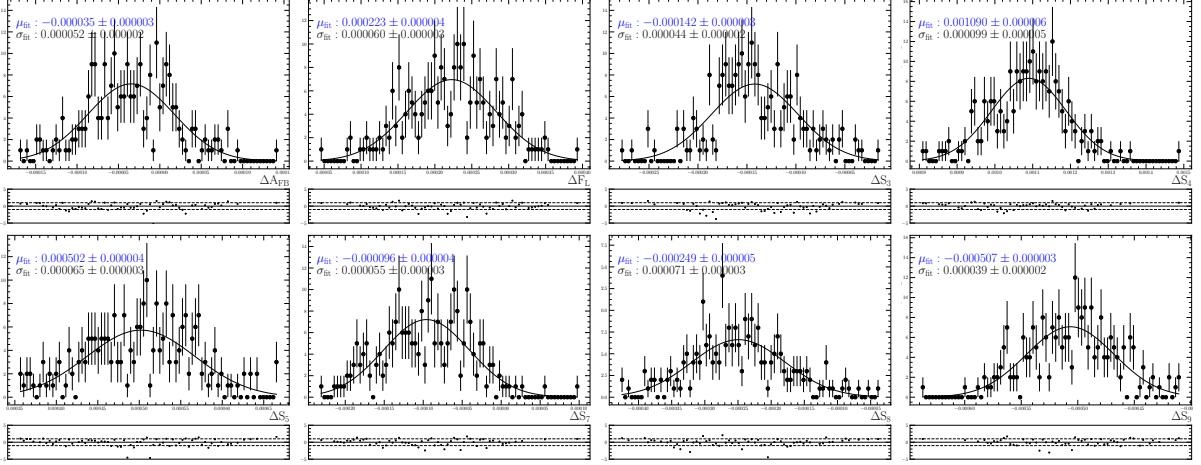


Figure 90: Observable difference distributions of pseudoexperiments fitted within the restricted range of $|\cos\theta_K| < 0.8$, which avoids the region where the difference between higher order and the nominal acceptances is large.

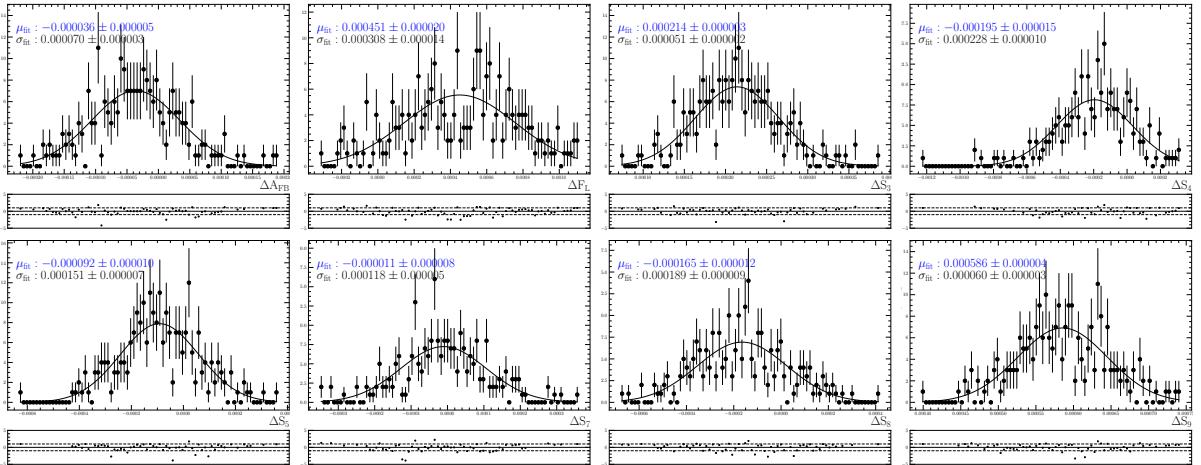


Figure 91: Observable difference distributions of pseudoexperiments fitted with the removal of events with large weights (in this case the arbitrary value of 1000 is used as a threshold for not normalised acceptance weights).

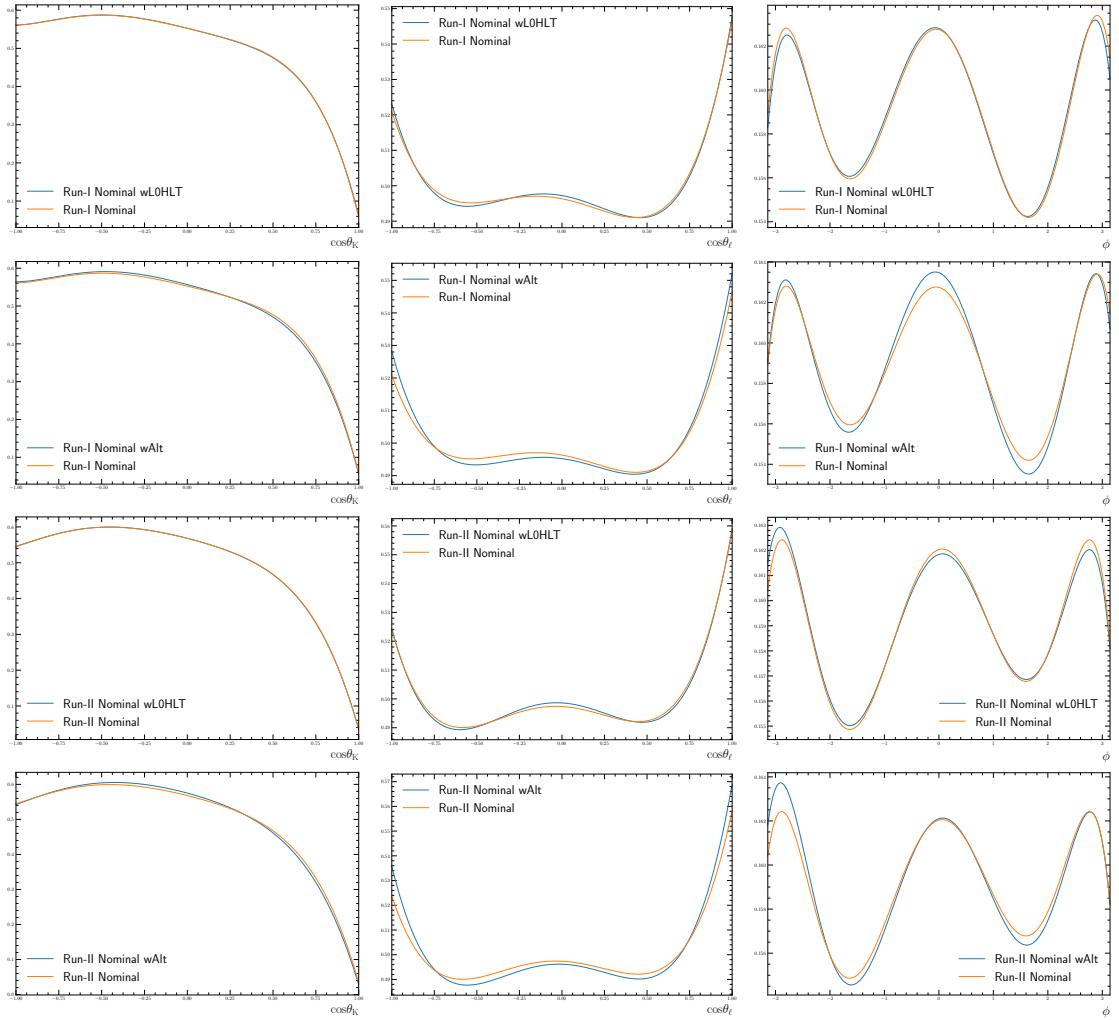


Figure 92: Projections of three dimensional control mode acceptances parametrised with and without (top) nominal correction weights (wL0HLT), and (bottom) alternative correction weights (wAlt).

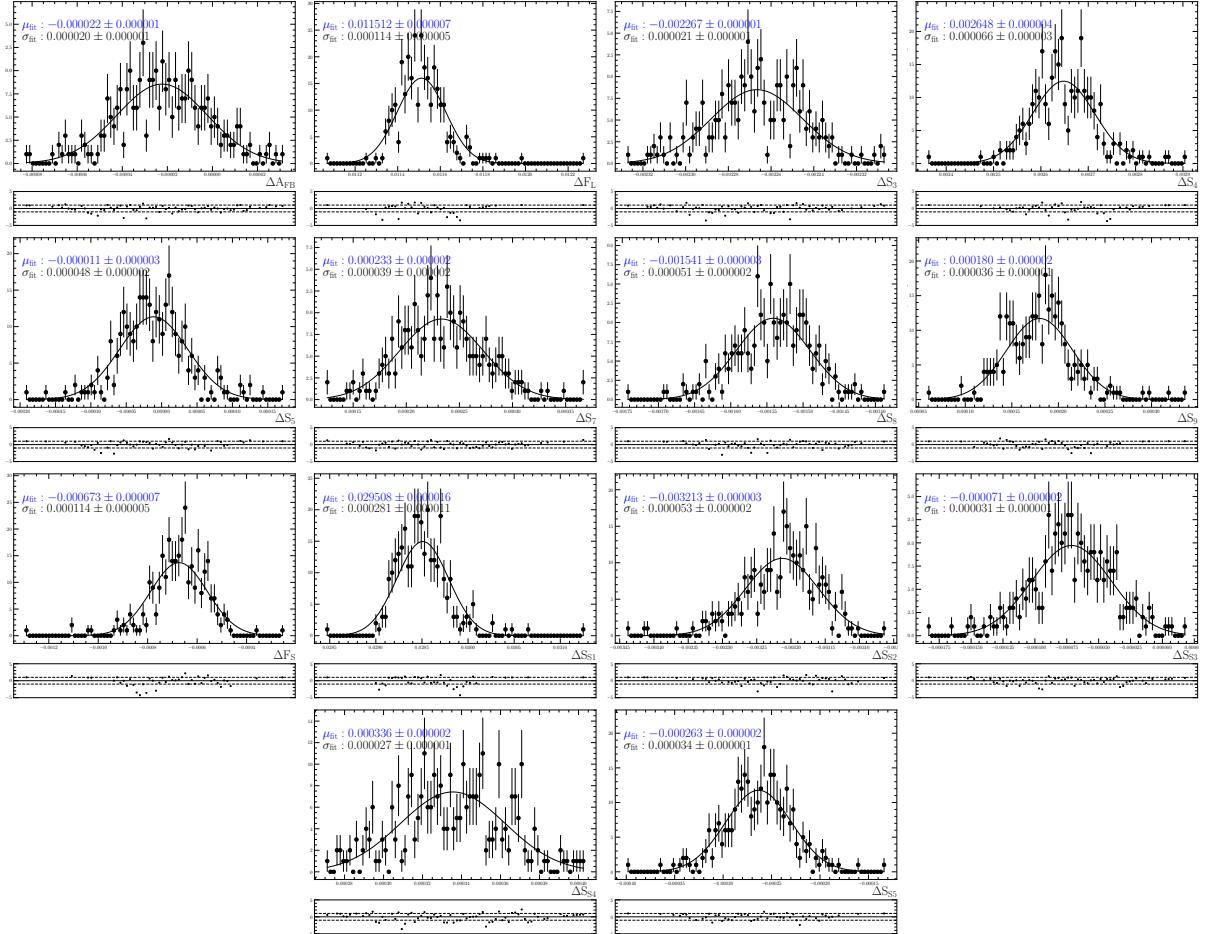


Figure 93: Three hundred pseudoexperiments are generated with alternative acceptances parametrised using simulation-data correction weights obtained following the R_X strategy. The distributions shown are the differences in observable values obtained when correcting with the nominal versus the alternative acceptances.

1984 **9.4 Cross checks**

1985 In addition to the nominal simultaneous fit to the Run-I and Run-II datasets, separate
1986 fits are made to subsamples corrected using their corresponding acceptance functions,
1987 namely individual Run, trigger, and bremsstrahlung categories, to check the consistency of
1988 the results. Furthermore, alternative choices are explored to probe the sensitivity of the
1989 observables to those variations. They include,

- 1990 • usage of FLATQ2 acceptance;
1991 • removal of veto against B^+ decays;
1992 • usage of acceptance functions parametrised with higher orders; and
1993 • changing the resolution model of the acceptance functions.

1994 For each of the listed points, a fit is made with the aforementioned set-up where only
1995 aspects of the model related to the mentioned detail are altered.

1996 To check the impact of cutting on the constrained q^2 (and the highly correlated cut
1997 on the J/ψ constrained B^0 mass) as well as background modelling on the observable
1998 values, fits are made for the following settings using a more detailed parametrisation of
1999 the background components:

- 2000 • unconstrained q^2 cut only ($6.0 < q^2 < 11.0$);
2001 • constrained q^2 cut ($7.0 < q_c^2 < 11.0$), no cut on $m_B^{J/\psi}$; and
2002 • constrained q^2 cut ($7.0 < q_c^2 < 11.0 \& m_B^{J/\psi} > 5150 \text{ MeV}/c^2$).

2003 Due to the lack of a PID variable (piPIDp) in some background samples, the veto against
2004 misidentified Λ_b^0 decays (Section 4) is not applied. This veto is currently part of the
2005 nominal selection, but it will be removed in the next iteration of this note.

2006 For display purposes, only the final numerical results and a selected number of figures
2007 are shown in the following sections. Additional information can be found in Appendix J.
2008 In any case, in overall a good stability in the observable of interest are seen in the many
2009 variations, validating the results discussed in this section.

2010 **9.4.1 Subsamples results**

2011 Separate fits are made to each Run, Run and L0 category, and bremsstrahlung category
2012 (in this case a simultaneous fit is made to both Runs). In all cases tailored acceptance
2013 functions based on the control mode simulation are used (*e.g.* for the 0γ category
2014 the corresponding acceptance functions are parametrised using 0γ events in simulation
2015 only). Due to low background contribution, for stability, modifications are made to the
2016 combinatorial component with respect to the setting used for the main result in the case
2017 of the Run-L0, and the bremsstrahlung categories fits. Background fractions (B_s and
2018 mis-identified Λ_b^0), and the bremsstrahlung categories fractions for the signal mass peak

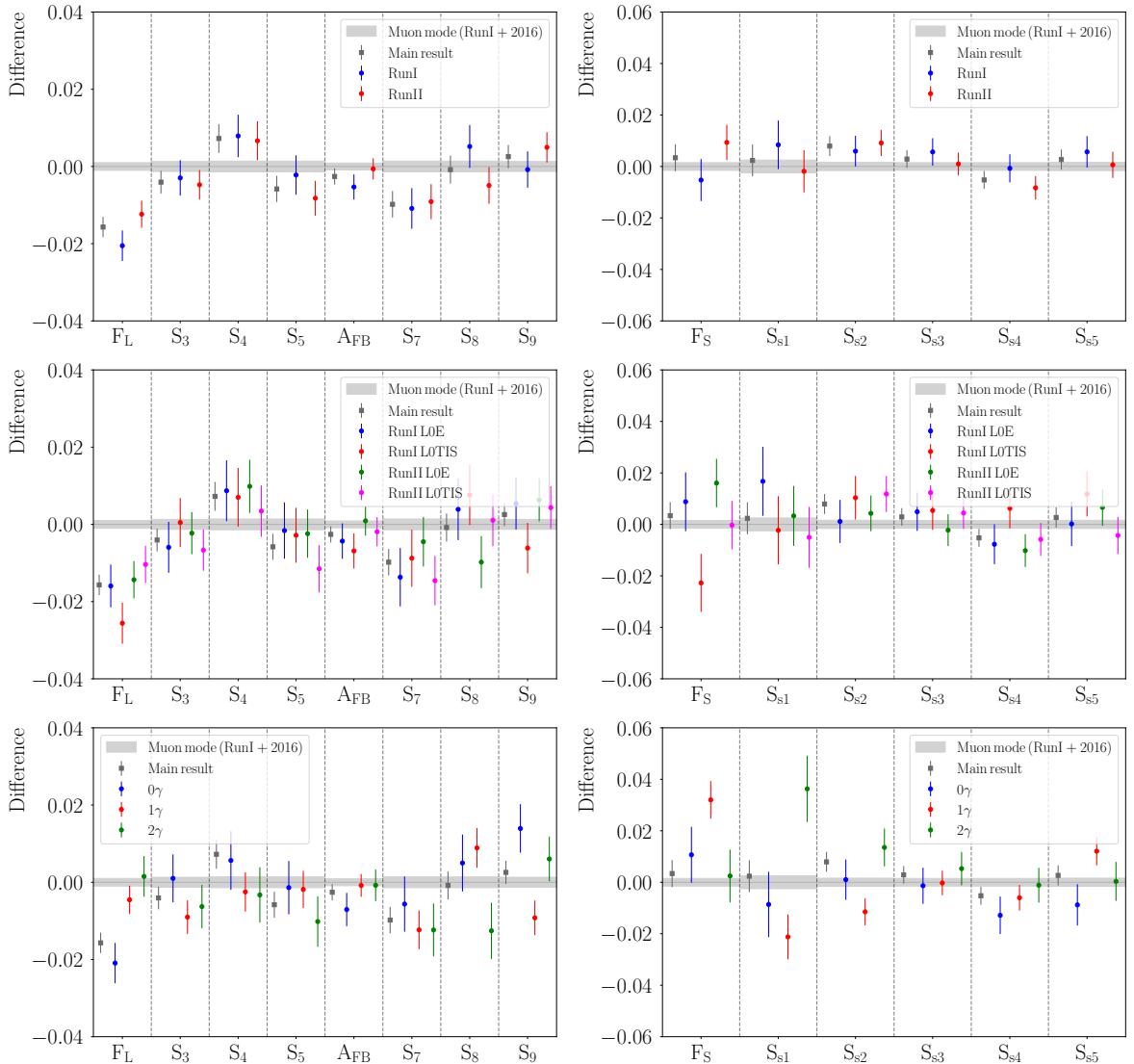


Figure 94: Summary plot showing the difference between the subsamples fit results and the muon mode simultaneous fit result (centered at zero, with uncertainty given by the gray band). The difference between the observable values of the main validation result and the muon mode result is also given (gray square).

2019 are recalculated where necessary. The observable values resulting from the fits are given
 2020 in the categories below, and more details can be found in Appendix J. The differences
 2021 between them and the muon mode simultaneous fit result (Table 32) is also displayed on
 2022 Figure 94 along with that of the main result.

2023 Separating by Run

2024 Results of separate angular fits to Run-I and Run-II samples each corrected with its
 2025 own corresponding acceptance functions are shown in Table 37. Good agreement is seen

	Run-I	Run-II	$\sigma(\text{Run-I}/\text{Run-II})$
F_L	0.540 ± 0.004	0.5485 ± 0.0035	-1.5
S_3	-0.003 ± 0.005	-0.005 ± 0.004	0.3
S_4	-0.241 ± 0.005	-0.242 ± 0.005	0.2
S_5	-0.002 ± 0.005	-0.008 ± 0.005	0.9
A_{FB}	-0.0048 ± 0.0033	-0.0001 ± 0.0028	-1.1
S_7	-0.010 ± 0.005	-0.008 ± 0.005	-0.3
S_8	-0.049 ± 0.006	-0.059 ± 0.005	1.4
S_9	-0.086 ± 0.005	-0.080 ± 0.004	-0.9
F_S	0.056 ± 0.008	0.071 ± 0.007	-1.4
S_{S_1}	-0.220 ± 0.009	-0.230 ± 0.008	0.8
S_{S_2}	0.028 ± 0.006	0.031 ± 0.005	-0.4
S_{S_3}	0.006 ± 0.005	0.002 ± 0.004	0.7
S_{S_4}	-0.001 ± 0.005	-0.008 ± 0.005	1.1
S_{S_5}	-0.061 ± 0.006	-0.066 ± 0.005	0.6

Table 37: Results of separate fits to Run-I and Run-II samples. The third column gives the differences between these two sets of results. Only statistical uncertainties of the fits are considered.

2026 between both sets of results.

2027 Separating by Run and trigger category

2028 Four separate fits are made to the Run-I LOE and LOTIS, and Run-II LOE and LOTIS
 2029 subsamples corrected with matching acceptance functions. Due to reduced statistics, in
 2030 particular for Run-I subsamples, for fit stability (to avoid the combinatorial component
 2031 from reaching negative values in some regions of the phase space), linear functions (first
 2032 order polynomials) are used to describe the angular distributions of the combinatorial
 2033 background. The results shown in Table 38 show good consistency between the two trigger
 2034 categories of each Run. Furthermore, good agreement is seen for most observables between
 2035 all four results and the results of the main fit, which suggests that the usage of the averaged
 2036 acceptance functions (averaged over the trigger categories) is sufficient.

2037 Separating by bremsstrahlung category

2038 Simultaneous fits are made to the 0γ , 1γ and 2γ sub-samples of Run-I and Run-II. Separate
 2039 acceptance functions are parametrised from the control mode simulation for each Run of
 2040 these categories. The angular fit is carried out with a fixed mis-identified Λ_b^0 component
 2041 and a combinatorial component that is allowed to vary. Due to decreased statistics, for
 2042 stability, linear functions are used for the angular distributions of the combinatorial. The
 2043 results of these three fits are given in Table 39.

2044 The agreement of the observable values between the 0γ and 1γ category, the largest of
 2045 the three, is good for most observables. The agreement between the 2γ category results

	Run-I L0E	$\Delta_{\text{Main} - \text{RunI L0E}}$	Run-I L0TIS	$\Delta_{\text{Main} - \text{RunI L0TIS}}$
F_L	0.545 ± 0.006	0.0003	0.535 ± 0.005	0.0099
S_3	-0.006 ± 0.007	0.0019	0.000 ± 0.006	-0.0045
S_4	-0.240 ± 0.008	-0.0015	-0.242 ± 0.008	0.0002
S_5	-0.001 ± 0.007	-0.0042	-0.002 ± 0.007	-0.0030
A_{FB}	-0.004 ± 0.005	0.0017	-0.006 ± 0.005	0.0043
S_7	-0.013 ± 0.008	0.0039	-0.008 ± 0.007	-0.0010
S_8	-0.050 ± 0.008	-0.0047	-0.047 ± 0.008	-0.0084
S_9	-0.080 ± 0.007	-0.0028	-0.091 ± 0.007	0.0087
F_S	0.070 ± 0.011	-0.0054	0.039 ± 0.011	0.0261
S_{S_1}	-0.211 ± 0.013	-0.0144	-0.231 ± 0.013	0.0047
S_{S_2}	0.023 ± 0.008	0.0068	0.032 ± 0.008	-0.0024
S_{S_3}	0.006 ± 0.007	-0.0020	0.006 ± 0.007	-0.0025
S_{S_4}	-0.008 ± 0.008	0.0025	0.006 ± 0.008	-0.0115
S_{S_5}	-0.067 ± 0.009	0.0025	-0.055 ± 0.009	-0.0091

	Run-II L0E	$\Delta_{\text{Main} - \text{RunII L0E}}$	Run-II L0TIS	$\Delta_{\text{Main} - \text{RunII L0TIS}}$
F_L	0.547 ± 0.005	-0.0013	0.551 ± 0.005	-0.0053
S_3	-0.003 ± 0.005	-0.0018	-0.007 ± 0.005	0.0026
S_4	-0.239 ± 0.007	-0.0026	-0.245 ± 0.007	0.0038
S_5	-0.002 ± 0.006	-0.0034	-0.011 ± 0.006	0.0057
A_{FB}	0.001 ± 0.004	-0.0035	-0.001 ± 0.004	-0.0007
S_7	-0.003 ± 0.006	-0.0053	-0.013 ± 0.006	0.0048
S_8	-0.064 ± 0.007	0.0090	-0.053 ± 0.007	-0.0019
S_9	-0.079 ± 0.006	-0.0038	-0.081 ± 0.006	-0.0018
F_S	0.078 ± 0.009	-0.0127	0.061 ± 0.009	0.0037
S_{S_1}	-0.225 ± 0.012	-0.0009	-0.233 ± 0.012	0.0074
S_{S_2}	0.026 ± 0.007	0.0036	0.034 ± 0.007	-0.0039
S_{S_3}	-0.002 ± 0.006	0.0051	0.005 ± 0.006	-0.0015
S_{S_4}	-0.010 ± 0.006	0.0050	-0.006 ± 0.006	0.0006
S_{S_5}	-0.061 ± 0.007	-0.0039	-0.071 ± 0.007	0.0070

Table 38: Results of fits to Run-I L0E and L0TIS samples (upper table) and Run-II L0E and L0TIS samples (lower table). The differences are given with respect to the main validation fit.

and those of 1γ is good with the exception of F_L (and S_{s1}). Possible effects such as the nominal acceptance order choice, which is selected to parametrise the full sample well, and may therefore be more suitable for the largest 1γ category, could explain such results. For example, the $\cos\theta_K$ distribution of the 0γ sample seems to show marked under-correction near $\cos\theta_K = 1$, while that of the 2γ seems to show over-correction (see Appendix X). F_L is sensitive to the edges of $\cos\theta_K$, and in general over/under-correction tend to lead to higher/lower F_L values.

	0γ	$\Delta_{\text{Main}-0\gamma}$	1γ	$\Delta_{\text{Main}-1\gamma}$	2γ	$\Delta_{\text{Main}-2\gamma}$
F_L	0.540 ± 0.005	0.0052	0.556 ± 0.004	-0.0111	0.562 ± 0.005	-0.0172
S_3	0.001 ± 0.006	-0.0051	-0.009 ± 0.004	0.0050	-0.007 ± 0.006	0.0022
S_4	-0.243 ± 0.008	0.0016	-0.251 ± 0.005	0.0098	-0.252 ± 0.007	0.0105
S_5	-0.001 ± 0.007	-0.0044	-0.001 ± 0.005	-0.0040	-0.009 ± 0.007	0.0044
A_{FB}	-0.007 ± 0.004	0.0045	-0.0003 ± 0.0029	-0.0018	-0.000 ± 0.004	-0.0018
S_7	-0.005 ± 0.007	-0.0041	-0.011 ± 0.005	0.0025	-0.011 ± 0.007	0.0026
S_8	-0.049 ± 0.007	-0.0058	-0.045 ± 0.005	-0.0097	-0.067 ± 0.007	0.0118
S_9	-0.071 ± 0.006	-0.0114	-0.094 ± 0.004	0.0118	-0.079 ± 0.006	-0.0035
F_S	0.072 ± 0.011	-0.0073	0.094 ± 0.007	-0.0286	0.064 ± 0.010	0.0009
S_{S1}	-0.237 ± 0.013	0.0110	-0.249 ± 0.009	0.0236	-0.192 ± 0.013	-0.0339
S_{S2}	0.023 ± 0.008	0.0069	0.011 ± 0.005	0.0194	0.036 ± 0.007	-0.0056
S_{S3}	-0.001 ± 0.007	0.0042	0.000 ± 0.005	0.0032	0.006 ± 0.006	-0.0024
S_{S4}	-0.013 ± 0.007	0.0076	-0.006 ± 0.005	0.0008	-0.001 ± 0.007	-0.0041
S_{S5}	-0.076 ± 0.008	0.0115	-0.055 ± 0.005	-0.0094	-0.067 ± 0.008	0.0023

Table 39: Observable values obtained from fits to the 0γ , 1γ and 2γ categories with separate acceptance functions.

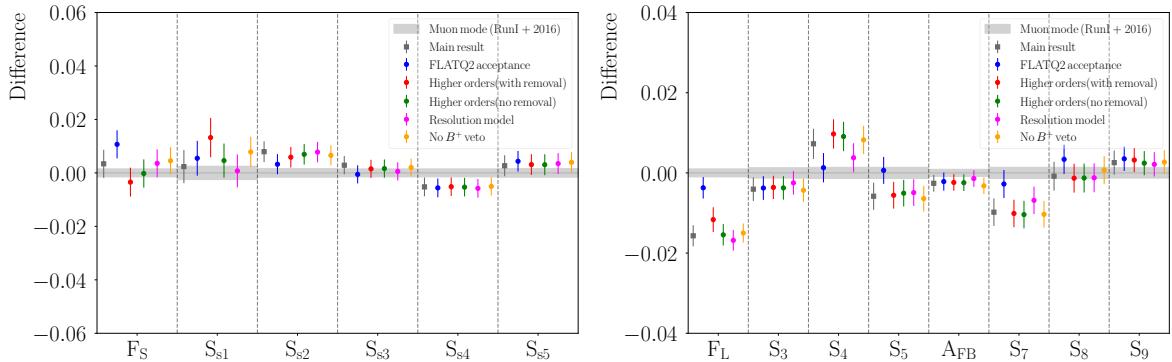


Figure 95: Summary plot showing the difference between the additional checks results and the muon mode simultaneous fit result (centered at zero, with uncertainty given by the gray band). The difference between the observable values of the main validation result and the muon mode result is also given (gray square).

9.4.2 Similar configurations

Additional checks related to the acceptance functions and the usage of the veto against B^+ decays are made using the same configuration as the main validation fit with the acceptances (and cut) under investigating changed (and minor updates to backgrounds where necessary). The numerical results are given in the following sections. Figure 95 provides a graphical display of the differences between each of those results and the muon mode simultaneous fit result.

2060 **FLATQ2 acceptance**

2061 The acceptance functions parametrised from FLATQ2 samples for the rare mode in the
 2062 extended q^2 range of $0.1 < q_c^2 < 12.0 \text{ GeV}^2/c^4$ can be used to correct the control mode
 2063 distributions. While the control mode acceptances are the preferred choice for fit validation,
 2064 as discussed in Section 9.1, a fit using the same set-up as the main validation fit with
 2065 FLATQ2 acceptances is carried out to check the the difference between FLATQ2 and PHYS
 2066 corrections in data. The results of this fit, given in Table 40, show that the differences
 2067 between the two are generally consistent with those seen in simulation studies (Table 33).

	Result	$\Delta_{\text{Main}-\text{FLATQ2}}$
F_L	0.5572 ± 0.0027	-0.0120
S_3	-0.0041 ± 0.0029	-0.0003
S_4	-0.248 ± 0.004	0.0060
S_5	0.0013 ± 0.0034	-0.0064
A_{FB}	-0.0017 ± 0.0023	-0.0004
S_7	-0.0017 ± 0.0035	-0.0070
S_8	-0.051 ± 0.004	-0.0042
S_9	-0.0817 ± 0.0030	-0.0010
F_S	0.072 ± 0.005	-0.0073
S_{S_1}	-0.223 ± 0.007	-0.0031
S_{S_2}	0.025 ± 0.004	0.0047
S_{S_3}	0.0002 ± 0.0034	0.0034
S_{S_4}	-0.0056 ± 0.0035	0.0004
S_{S_5}	-0.063 ± 0.004	-0.0017

Table 40: Values for the S-basis angular observables obtained from a fit to control mode candidates using correction weights from rare mode FLATQ2 acceptances. The differences are given with respect to the main result.

2068 **Higher order acceptance**

2069 The usage of higher orders in the acceptance parametrisation has the potential to improve
 2070 the description of the acceptance effect in certain regions, in particular near $\cos\theta_K = 1$,
 2071 to which F_L is particularly sensitive. The impact of choosing higher orders (increase by
 2072 three for all angles) is checked in the data fit as well as using pseudoexperiments (see
 2073 Section 9.3).

2074 The data fit is carried out using correction weights from higher order acceptances. It
 2075 is otherwise made with the main configuration (combinatorial only). Two fits are made
 2076 with and without the removal of six events with very large correction weights, located near
 2077 $\cos\theta_K = 1$. Both results are shown in Table 41.

2078 **Resolution correction**

2079 The resolution correction is model dependent. While the underlying model of the control

	No removal	$\Delta_{\text{Main}-\text{no removal}}$	With removal	$\Delta_{\text{Main}-\text{with removal}}$
F_L	0.5492 ± 0.0031	-0.0046	0.5455 ± 0.0027	-0.0008
S_3	-0.0039 ± 0.0029	-0.0003	-0.0040 ± 0.0029	-0.0002
S_4	-0.239 ± 0.004	-0.0037	-0.240 ± 0.004	-0.0031
S_5	-0.0049 ± 0.0033	0.0007	-0.0044 ± 0.0033	0.0002
A_{FB}	-0.0019 ± 0.0021	-0.0004	-0.0019 ± 0.0021	-0.0004
S_7	-0.0090 ± 0.0034	0.0002	-0.0093 ± 0.0034	0.0005
S_8	-0.056 ± 0.004	0.0007	-0.056 ± 0.004	0.0007
S_9	-0.0820 ± 0.0030	-0.0007	-0.0828 ± 0.0030	0.0001
F_S	0.058 ± 0.005	0.0066	0.061 ± 0.005	0.0034
S_{S_1}	-0.215 ± 0.007	-0.0115	-0.224 ± 0.006	-0.0029
S_{S_2}	0.028 ± 0.004	0.0012	0.029 ± 0.004	0.0001
S_{S_3}	0.0022 ± 0.0034	0.0013	0.0023 ± 0.0034	0.0012
S_{S_4}	-0.0052 ± 0.0035	0.0000	-0.0053 ± 0.0035	0.0002
S_{S_5}	-0.064 ± 0.004	-0.0002	-0.064 ± 0.004	-0.0001

Table 41: Observable values obtained from the fit made with corrections from acceptance functions parametrised using higher orders (all angle orders increased by three). The fit is otherwise set-up in the same way as the nominal (combinatorial only) configuration. The two columns show the fit results without and with the removal of six events with large weights.

mode simulation is expected to be similar to that of the data, some differences remain, as the generator level observables differ significantly from the results of the existing measurements and cross-checks. Furthermore the simulation does not include S-wave contribution. Increasing the alignment between simulation and data could improve the resolution correction. While the ideal approach would be to generate new simulation samples with an alternative physics model, which would also improve the correction of FSR (also model dependent), it is a complex and time consuming procedure more suited to a dedicated angular analysis of the control mode. Instead, a simpler approach is used that involves changing the physics model using truth information of the simulation. In place of the inverse generation level weights applied to the post-selection sample in the standard acceptance parametrisation,

$$w = 1/P_S^{(\text{old})}(\vec{\Omega}_{\text{Reco}}, q_{\text{Reco}}^2), \quad (50)$$

where $P_S^{(\text{old})}$ is model that describes the generator level sample, to change the physics model, a new weight, w_{new} , is applied:

$$w_{\text{new}} = (1/P_S^{\text{old}}(\vec{\Omega}_{\text{True}}, q_{\text{True}}^2)P_S^{\text{new}}(\vec{\Omega}_{\text{True}}, q_{\text{True}}^2)) \times (1/P_S^{\text{new}}(\vec{\Omega}_{\text{Reco}}, q_{\text{Reco}}^2)). \quad (51)$$

Here, P_S^{new} is the function that describes the revised underlying physics model. In this way, the old physics model is ‘removed’ from the sample, and the new model is multiplied to it. The new model used is the angular pdf containing both S and P-wave contributions,

2096 with observable values taken from the muon mode cross-check (Table 32), and the old
 2097 model is the P-wave only pdf, with observables set to generation values.

2098 Note this approach does not improve the correction of the FSR, as true information in
 2099 the simulation samples contain FSR effects. Also no ‘pre-FSR’ quantities exists for each
 2100 event to allow for the direct application of the same procedure to FSR correction.

2101 The alternative acceptance functions are used in a data fit with otherwise nominal
 2102 configuration (combinatorial only), and the results (Table 42) show minor improvement
 2103 compared to the nominal values. Interestingly S_7 seems to shift closer to the muon mode
 2104 value, suggesting that resolution correction may play a role in causing the minor tension
 2105 seen for this observable.

	Result	$\delta_{\text{Main-reso. corr.}}$
F_L	0.5441 ± 0.0026	0.0005
S_3	-0.0028 ± 0.0029	-0.0014
S_4	-0.245 ± 0.004	0.0022
S_5	-0.0042 ± 0.0033	0.0000
A_{FB}	-0.0009 ± 0.0021	-0.0014
S_7	-0.0057 ± 0.0034	-0.0031
S_8	-0.056 ± 0.004	0.0006
S_9	-0.0831 ± 0.0030	0.0004
F_S	0.065 ± 0.005	-0.0004
S_{S_1}	-0.227 ± 0.006	0.0009
S_{S_2}	0.030 ± 0.004	-0.0007
S_{S_3}	0.0013 ± 0.0034	0.0023
S_{S_4}	-0.0058 ± 0.0035	0.0006
S_{S_5}	-0.064 ± 0.004	-0.0005

Table 42: Observable values obtained from the fit with alternative acceptance obtained with an adjusted resolution model. The fit is otherwise set-up in the same way as the nominal (combinatorial only) configuration. The differences are given with respect to the muon mode simultaneous fit results.

2106 No B^+ veto

2107 The B^+ veto is known to distort the combinatorial background in the high mass region,
 2108 and introduce correlation between $\cos\theta_K$ and B^0 mass. This is illustrated by the two
 2109 dimensional histograms on Figure 96. The removal of this cut leads to the inclusion of a
 2110 limited number of B^+ events, which are, however, unlikely to cause significant impact on
 2111 the angular observables. The decreased distortion of the combinatorial background may
 2112 result in better fits, and improved separation between signal and combinatorial events.

2113 In the fit without B^+ veto, the $\cos\theta_\ell$ distribution of the combinatorial component is
 2114 seen to become negative close to $\cos\theta_\ell = -0.25$ for the smaller Run-I sample. This is likely
 2115 due to a combination of statistical fluctuation, and the inclusion of DSL events in the

modelling and is not expected to be related to the veto itself. To avoid this situation, and considering the similarity of the angular combinatorial (includes partially reconstructed and DSL events) distributions, one common set of parameters is used for both Runs. The results of this test is given in Table 43. The angular observables obtained are in agreement with that of the nominal results.

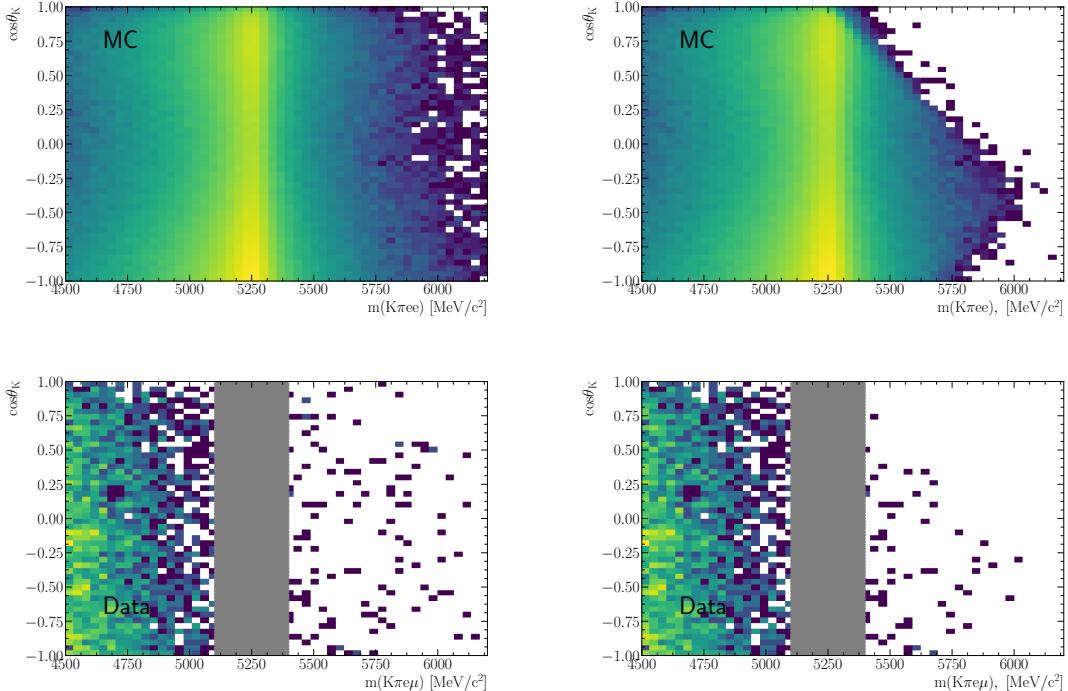


Figure 96: $B^0 \rightarrow K^* J/\psi (\rightarrow e^+ e^-)$ simulation (top row) and $K\pi e\mu$ data (bottom row) without (left) and with (right) the application of the veto against $B^+ \rightarrow K^+ e^+ e^- (+\pi)$ decays. This veto creates some correlation between $\cos\theta_K$ and B^0 mass for the signal, and shapes the combinatorial background in the high mass region.

9.4.3 Additional backgrounds

Besides variations of the main validation fit, the fit strategy is also tested in different regimes and with the inclusion of additional backgrounds, which are either neglected in the main fit, or incorporated into the ‘combinatorial’ component. For simplicity, the studies performed in the R_X analysis are used as benchmark in the estimation and modelling (including samples) of possible sources of background in the mass fit. To reduce computational requirements, the usage of KDE is avoided for all components. The results of the fits are discussed in the following sections, and additional information on the background modelling can be found in Appendix J. The results of those fits and given in the following sections. Figure 97 provides a visual summary of those results, showing the differences for all observables of each result with that of the muon mode simultaneous fit.

	Result	$\Delta_{\text{Main-no veto}}$
F_L	0.5459 ± 0.0024	-0.0013
S_3	-0.0046 ± 0.0029	0.0004
S_4	-0.2407 ± 0.0035	-0.0022
S_5	-0.0057 ± 0.0032	0.0015
A_{FB}	-0.0027 ± 0.0021	0.0004
S_7	-0.0092 ± 0.0033	0.0004
S_8	-0.0537 ± 0.0035	-0.0013
S_9	-0.0825 ± 0.0030	-0.0002
F_S	0.066 ± 0.005	-0.0014
S_{S1}	-0.220 ± 0.006	-0.0062
S_{S2}	0.029 ± 0.004	0.0005
S_{S3}	0.0027 ± 0.0034	0.0008
S_{S4}	-0.0050 ± 0.0034	-0.0001
S_{S5}	-0.063 ± 0.004	-0.0010

Table 43: Observable values obtained from angular fit to the control mode in the constrained q^2 range of 7.0 to 11.0 GeV^2/c^4 , without the veto against $B^+ \rightarrow K^+ e^+ e^- (+\pi)$ decays, for the nominal and reduced mass ranges. The differences are given with respect to the muon mode simultaneous fit result.

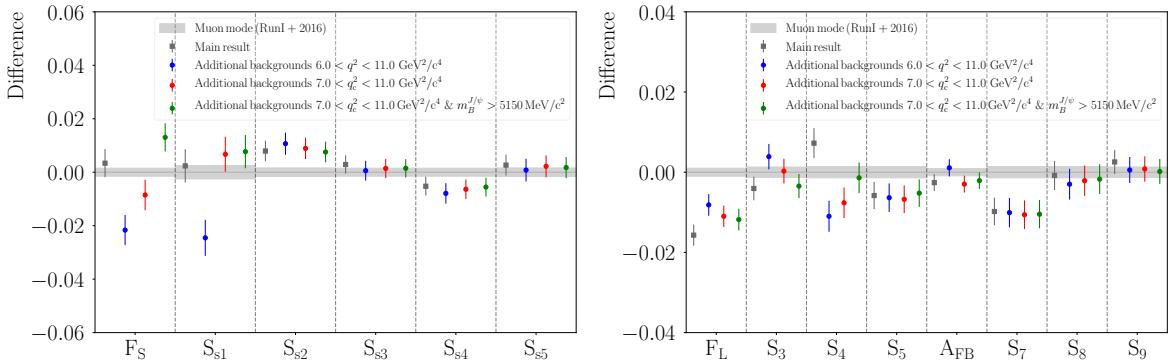


Figure 97: Summary plot showing the difference between the fit results of configurations with additional backgrounds and the muon mode simultaneous fit result (centered at zero, with uncertainty given by the gray band). The difference between the observable values of the main validation result and the muon mode result is also given (gray square).

Unconstrained q^2 range

In this configuration, the cut on the unconstrained q^2 of 6.0 to 11.0 GeV^2/c^4 is used to isolate $B^0 \rightarrow K^* J/\psi (\rightarrow e^+ e^-)$ candidates, and the mass window is extended to 4400 MeV/c^2 to increase the information available for the separation of multiple background components. This matches most closely the strategy employed by the R_X analysis. However, the aim of this fit is not to reproduce the R_X result, but rather to check the

2138 impact of background modelling on the angular observables in a more background rich
2139 regime. The advantage of cutting on the unconstrained q^2 is that it does not tend to
2140 warp the combinatorial mass distribution. Furthermore, the increase in background yield
2141 could lead to improved control of the level of partially reconstructed and combinatorial
2142 backgrounds in the signal region, which is difficult to achieve for the nominal configuration.
2143 One drawback of this test is that the angular distributions of the cocktail simulation for the
2144 partially reconstructed components are only a simplified representation of the underlying
2145 physics, which can be significantly different than the true distribution in data.

2146 The full fit is carried out simultaneously for Run-I and Run-II samples using acceptance
2147 functions parametrised from the control mode simulation with matching q^2 and mass
2148 cuts. Eight components are included in the fit besides the signal – hadronic and leptonic
2149 partially reconstructed (PRH and PRL), $B_s^0 \rightarrow J/\psi e^+ e^-$, $\Lambda_b^0 \rightarrow p K J/\psi$ with $p - \pi$
2150 mis-identification, $B_s^0 \rightarrow \phi J/\psi$ with $K - \pi$ mis-identification, DSL, signal with double
2151 hadron swap, and combinatorial. Among these the fractions with respect to signal of
2152 backgrounds that contribute at low levels (to which the fit has no sensitivity) are fixed to
2153 values calculated from simulation. These are given in Table 44. Parameters allowed to vary
2154 in the fit besides the angular observables and signal fractions include the fractions of DSL,
2155 and partially reconstructed (both types) with respect to signal, the shift and scale of the
2156 signal mass peak, slope of the combinatorial background and the angular shape parameters
2157 of the combinatorial background (here up to first order only). Separate parameters are
2158 used for each Run with the exception of the angular combinatorial parameters, which are
2159 shared for the two Runs for stability.

2160 The results of the fit are given in Table 45 and displayed in Figs. 98 and 99. Good
2161 consistency is seen for most observables, except for parameters related to S-wave terms
2162 and S_4 , and to a less extend F_L and S_3 . Possible explanations for these small discrepancies
2163 may include the poor background modelling of the $\cos\theta_K$ distributions, particularly
2164 above $\cos\theta_K = 0.25$, as well as mis-modelling of the partially reconstructed components.
2165 Nevertheless, it is reassuring that despite the drastic change in strategy, the central value
2166 shifts for most observables are small compared to the fit uncertainties of the rare mode.

Parameters	Run-I	Run-II
f_{B_s}	0.0086	0.0079
$f_{\Lambda_b^0}$	0.0039	0.0015
$f_{K-\pi}$	0.00062	0.00085
$f_{B_s \rightarrow \phi J/\psi}$	0.0056	0.0030
$f_{B^0}^{PRH}$	0.56	
$f_{B^+}^{PRH}$	0.43	
$f_{B^0}^{PRL}$	0.97	
$f_{B^+}^{PRL}$	0.02	
$f_{\text{sig}}^{0\gamma}$	0.23	0.22
$f_{\text{sig}}^{1\gamma}$	0.51	0.50

Table 44: Parameters fixed in the data fit. The first four rows show the fixed fractions of minor backgrounds with respect to signal, the next four rows provide the fractions of B^0 and B^+ functions in the full PRH/PRL model, and the last two rows give the bremsstrahlung category fractions of the signal mass model.

	Result	$\Delta_{\text{Main}-\text{uncons. } q^2}$
F_L	0.5527 ± 0.0027	-0.0075
S_3	0.0036 ± 0.0032	-0.0079
S_4	-0.260 ± 0.004	0.0182
S_5	-0.006 ± 0.004	0.0005
A_{FB}	0.0016 ± 0.0022	-0.0037
S_7	-0.009 ± 0.004	0.0003
S_8	-0.057 ± 0.004	0.0022
S_9	-0.0846 ± 0.0032	0.0020
F_S	0.040 ± 0.006	0.0250
S_{S_1}	-0.253 ± 0.007	0.0269
S_{S_2}	0.033 ± 0.004	-0.0028
S_{S_3}	0.001 ± 0.004	0.0023
S_{S_4}	-0.008 ± 0.004	0.0027
S_{S_5}	-0.066 ± 0.004	0.0019

Table 45: Observable values obtained from angular fit to the control mode in the unconstrained q^2 range of 6.0 to 11.0 GeV $^2/c^4$ in comparison to the nominal fit.

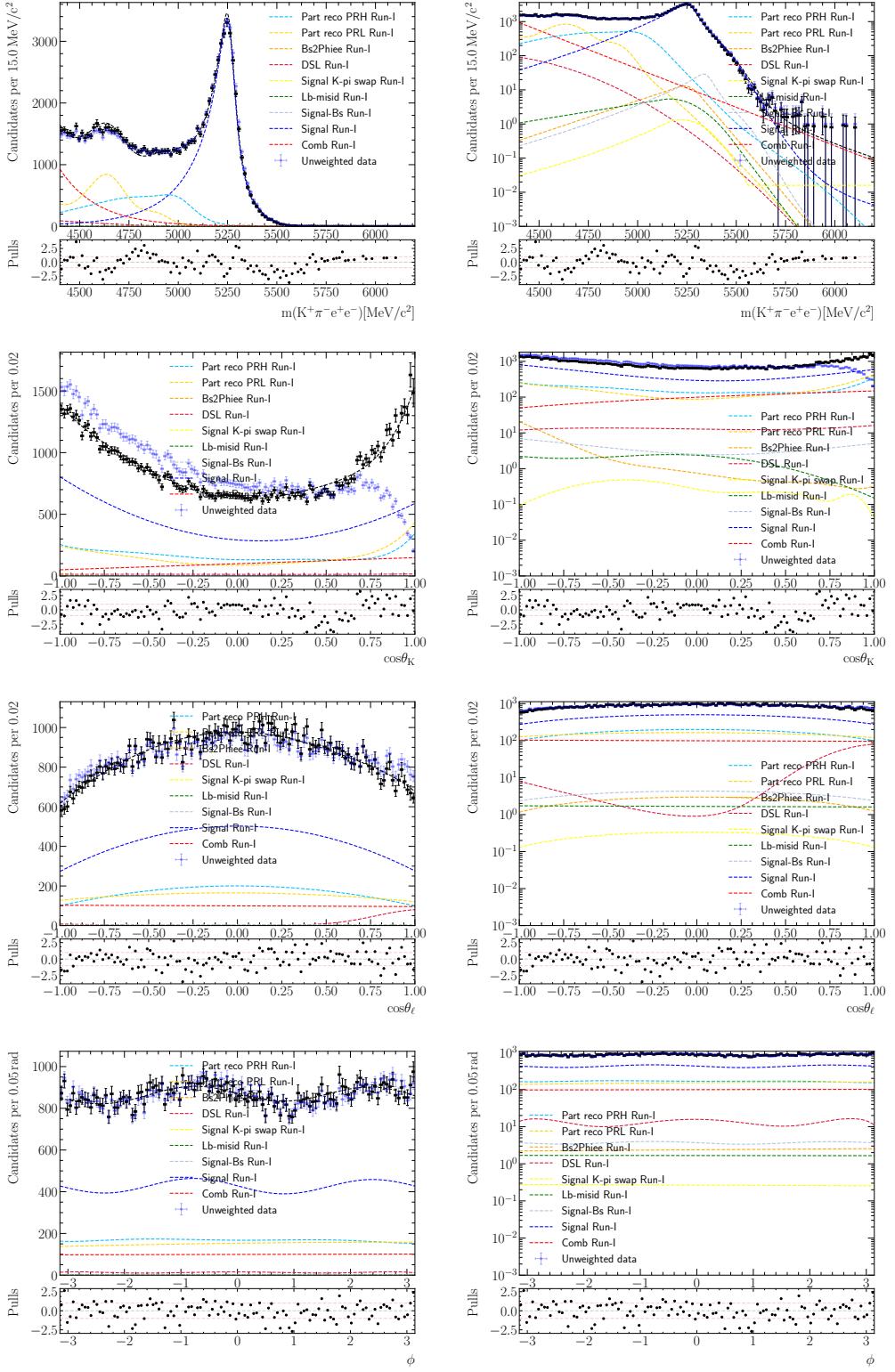


Figure 98: Result of the fit to the unconstrained q^2 range of 6.0 to 11.0 GeV^2/c^4 for Run-I dataset. There is an overall good agreement in all projections.

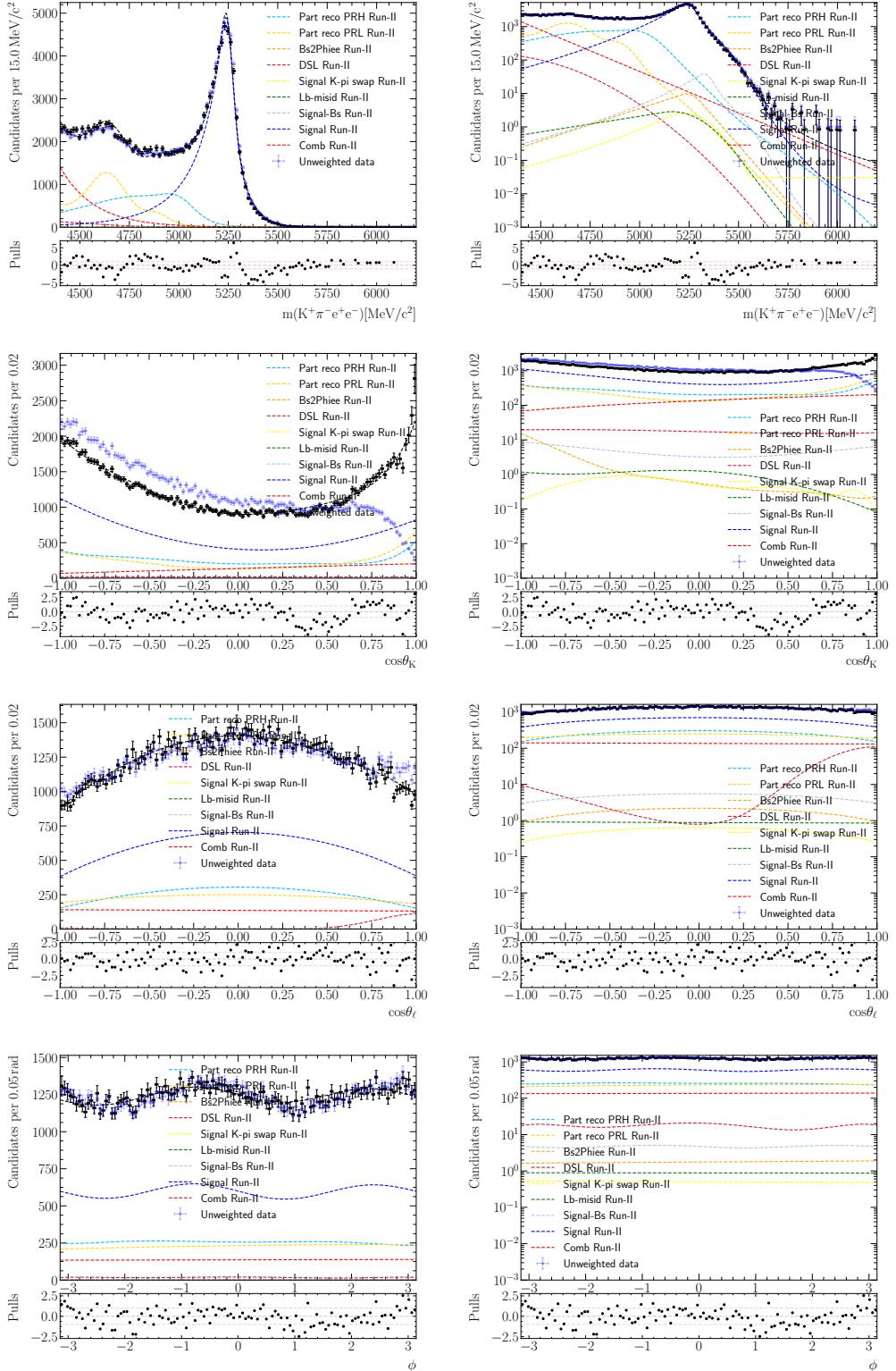


Figure 99: Result of the fit to the unconstrained q^2 range of 6.0 to 11.0 GeV^2/c^4 for Run-II dataset. There is an overall good agreement in all projections.

2167 **Constrained q^2 without constrained B^0 mass cut**

2168 For the nominal fit to the control mode the cut of $m_{B_s, J/\psi} > 5150 \text{ MeV}/c^2$ is applied
 2169 to reduce low mass backgrounds. Loosening this criteria results in a visible increase in
 2170 partially reconstructed events, and can lead to improved signal background separation.
 2171 This configuration is also an intermediate step between the fit to the unconstrained q^2 ,
 2172 and the nominal selection.

2173 To improve fit stability, some simplifications are made with respect to the unconstrained
 2174 q^2 set-up, namely that one single fraction (with respect to signal) is used for the PRL
 2175 components of both Runs, and one set of shared parameters is used for the combinatorial
 2176 angular background (up to second order for all angles), which is allowed to vary. The
 2177 fractions fixed in the fit are given in Table 46. No separate DSL component is used, and it
 2178 is assumed that the combinatorial component will incorporate DSL events. The results of
 2179 the fit given in Figs. 100 and 101. Table 47 shows good agreement with the main validation
 2180 results for most observables with the small differences in S_4 and F_S . As before, difference
 2181 at this size can be expected given the lack of knowledge on the precise distribution of
 2182 the partially reconstructed contribution. Hence, these studies further confirm the good
 2183 understanding of the angular fit for the control channel.

Parameters	Run-I	Run-II
f_{B_s}	0.0086	0.0079
$f_{\Lambda_b^0}$	0.0024	0.0008
$f_{K-\pi}$	0.00060	0.00083
$f_{B_s \rightarrow \Phi J/\psi}$	0.0052	0.0028
$f_{B^0}^{PRH}$	0.60	
$f_{B^+}^{PRH}$	0.38	
$f_{B^0}^{PRL}$	0.94	
$f_{B^+}^{PRL}$	0.0	
$f_{\text{sig}}^{0\gamma}$	0.23	0.22
$f_{\text{sig}}^{1\gamma}$	0.50	0.50

Table 46: Parameters fixed in the data fit. The first four rows show the fixed fractions of minor backgrounds with respect to signal, the next four rows provide the fractions of Bd and Bu functions in the full PRH/PRL model, and the last two rows give the bremsstrahlung category fractions of the signal mass model.

	Result	$\delta_{\text{Main-no 5150}}$
F_L	0.5499 ± 0.0027	-0.0047
S_3	-0.0000 ± 0.0031	-0.0043
S_4	-0.257 ± 0.004	0.0149
S_5	-0.0061 ± 0.0035	0.0009
A_{FB}	-0.0025 ± 0.0021	0.0004
S_7	-0.010 ± 0.004	0.0008
S_8	-0.057 ± 0.004	0.0013
S_9	-0.0844 ± 0.0032	0.0017
F_S	0.053 ± 0.006	0.0119
S_{S_1}	-0.221 ± 0.007	-0.0044
S_{S_2}	0.031 ± 0.004	-0.0010
S_{S_3}	0.002 ± 0.004	0.0015
S_{S_4}	-0.006 ± 0.004	0.0012
S_{S_5}	-0.065 ± 0.004	0.0005

Table 47: Observable values obtained from angular fit to the control mode in the constrained q^2 range of 7.0 to 11.0 GeV $^2/c^4$ with cut on the B^0 mass with B^0 primary vertex and J/ψ mass constraints.

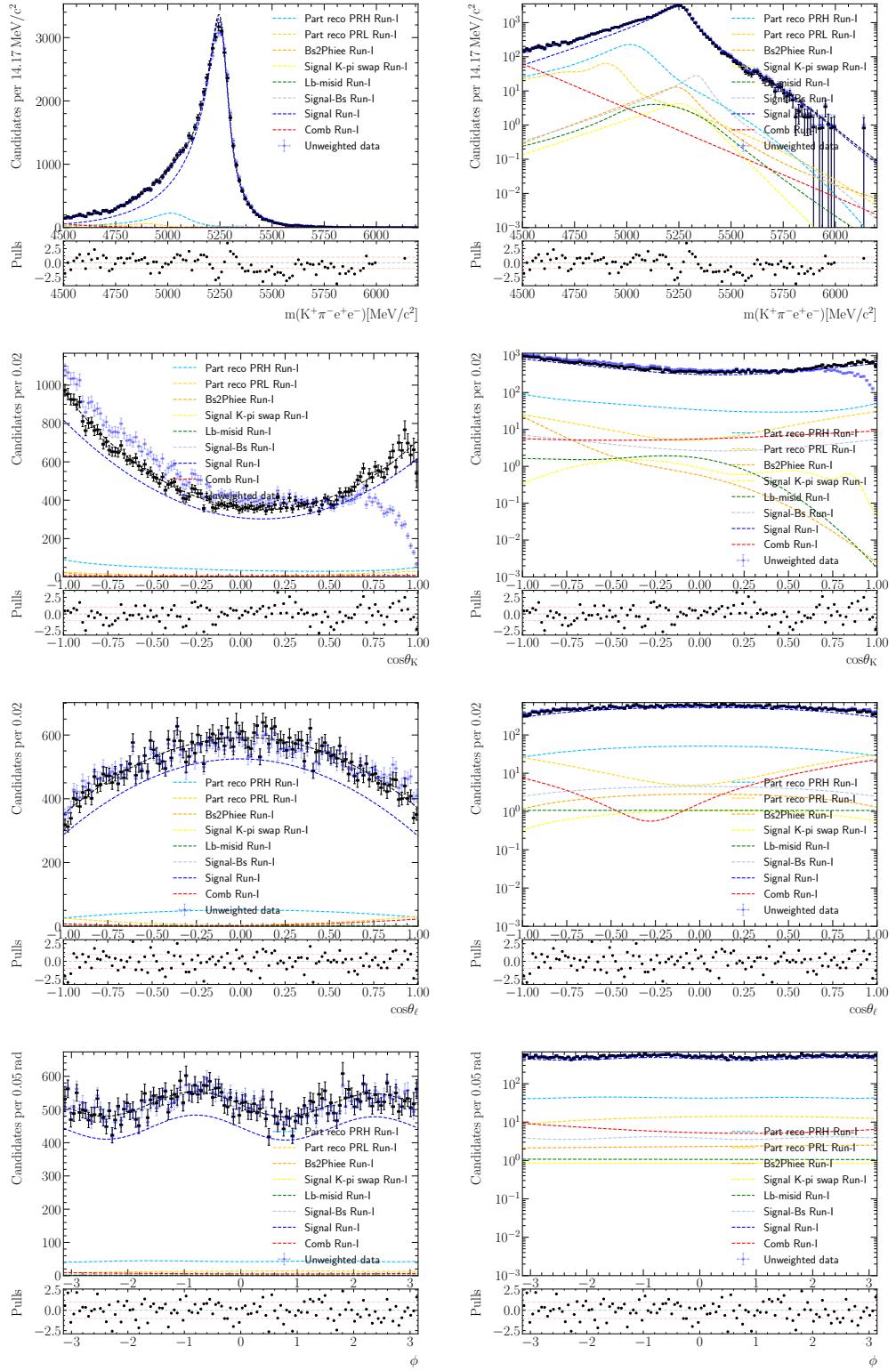


Figure 100: Result of the fit to the constrained q^2 range of 7.0 to 11.0 GeV^2/c^4 for Run-I without the cut on the B^0 mass with the primary vertex and J/ψ mass constraints. All peaking backgrounds are fixed to fractions (with respect to signal) calculated from simulation. The contributions of the DSL, partially reconstructed and combinatorial backgrounds are allowed to vary in the fit.

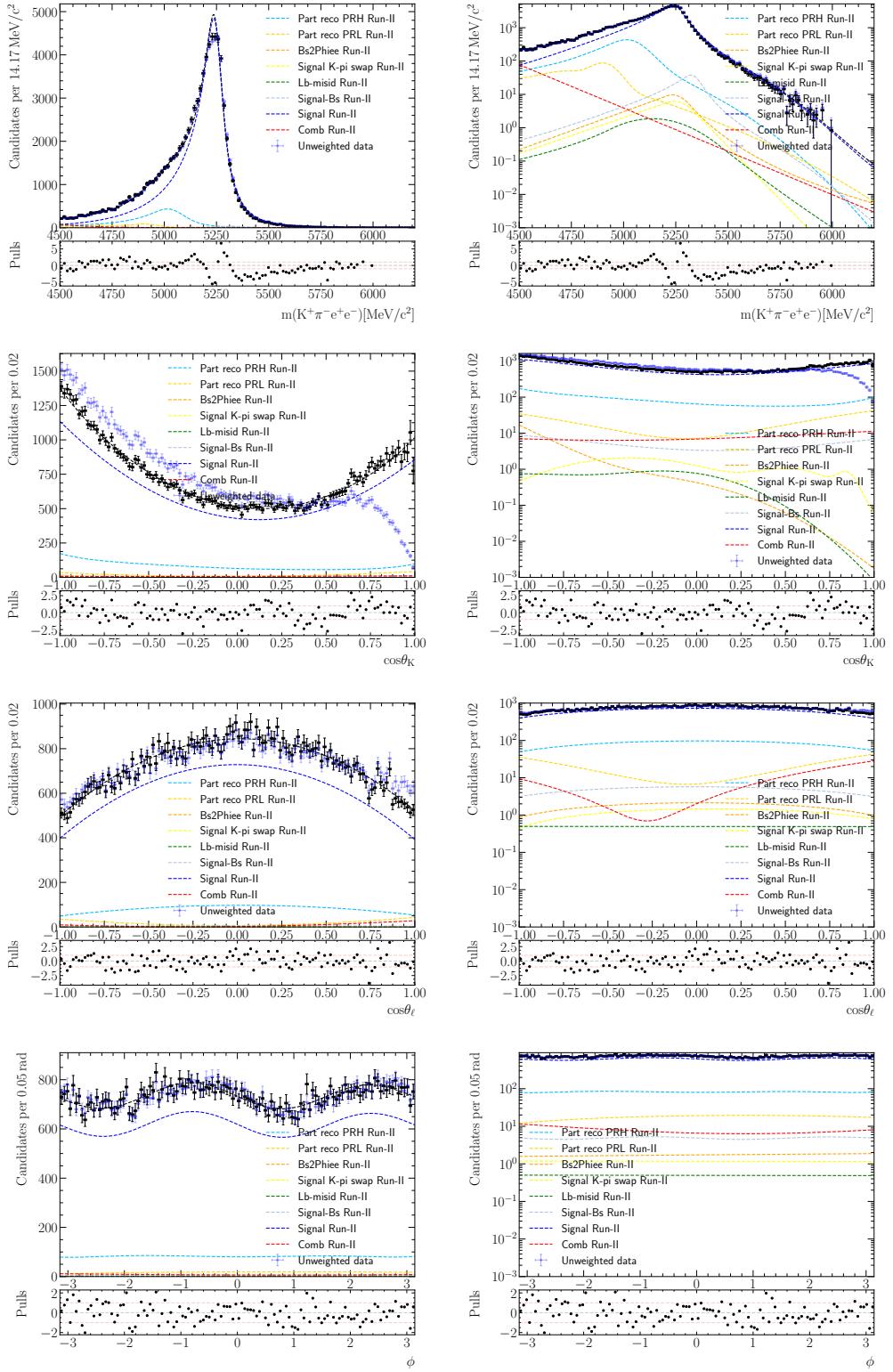


Figure 101: Result of the fit to the constrained q^2 range of 7.0 to 11.0 GeV^2/c^4 for Run-II without the cut on the B^0 mass with the primary vertex and J/ψ mass constraints. All peaking backgrounds are fixed to fractions (with respect to signal) calculated from simulation. The contributions of the DSL, partially reconstructed and combinatorial backgrounds are allowed to vary in the fit.

2184 **Constrained q^2 with constrained B^0 mass cut**

2185 A similar fit as the nominal constrained q^2 range with the $m_{B,J/\psi} > 5150\text{MeV}/c^2$ cut
 2186 can be attempted with the inclusion of all possible additional peaking backgrounds. This
 2187 is identical to the selection used for the main fit with the exception of the veto against
 2188 Λ_b^0 decays. The idea is to examine if the inclusion of additional sources (although with
 2189 marginal contribution) can distort the observables of interest.

2190 Ideally the three backgrounds with similar (exponential-like) distributions in mass – the
 2191 partially reconstructed (mostly hadronic partially reconstructed as the leptonic component
 2192 is greatly suppressed), DSL and combinatorial components – should also be included as
 2193 distinct components in the fit. However, due to low statistics, they are difficult to separate.
 2194 The angular distribution of the combinatorial background when the shape parameters
 2195 are allowed to vary suggests that a portion of the ‘combinatorial’ component may be
 2196 comprised of partially reconstructed and DSL events. See for instance a comparison of
 2197 the combinatorial angular shapes in Figure 86 with simulation for partially reconstructed
 2198 components, and DSL step two fit results. Therefore for this check, the decision is made
 2199 to test a strategy that is significantly different from the one employed in the main fit
 2200 – instead of using a combinatorial component to take into account remaining partially
 2201 reconstructed and DSL events, only the DSL and the (hadronic) partially reconstructed
 2202 components are used. The contribution from leptonic partially reconstructed component
 2203 is expected to be lower than the remaining hadronic component by about a factor of ten.
 2204 As too few events remain from simulation for modelling, and the contribution from this
 2205 background is low, this component is not included in the fit. The fixed fraction for this
 2206 scenario are given in Table 48. The results of the fit, shown in Figs. 102 and 103, and
 2207 Table 49, show good consistency with the main results. Only three observables, F_L , S_4
 2208 and F_S , demonstrate noteworthy levels of sensitivity to the background modelling strategy
 2209 in this case. Nevertheless, these results further confirm the good parametrisation of the
 control mode.

Parameters	Run-I	Run-II
f_{B_s}	0.0086	0.0079
$f_{\Lambda_b^0}$	0.0016	0.0001
$f_{K-\pi}$	0.00058	0.00077
$f_{B_s \rightarrow \phi J/\psi}$	0.0048	0.0025
$f_{B^0}^{PRH}$	0.54	
$f_{B^+}^{PRH}$	0.34	
$f_{\text{sig}}^{0\gamma}$	0.23	0.22
$f_{\text{sig}}^{1\gamma}$	0.50	0.50

Table 48: Parameters fixed in the data fit. The first four rows show the fixed fractions of minor backgrounds with respect to signal, the next four rows provide the fractions of Bd and Bu functions in the full PRH/PRL model, and the last two rows give the bremsstrahlung category fractions of the signal mass model.

2210

	Result	$\delta_{\text{Main-with 5150}}$
F_L	0.5491 ± 0.0027	-0.0039
S_3	-0.0038 ± 0.0030	-0.0006
S_4	-0.250 ± 0.004	0.0086
S_5	-0.0045 ± 0.0034	-0.0006
A_{FB}	-0.0016 ± 0.0021	-0.0005
S_7	-0.0094 ± 0.0035	0.0007
S_8	-0.056 ± 0.004	0.0009
S_9	-0.0850 ± 0.0031	0.0024
F_S	0.075 ± 0.005	-0.0096
S_{S_1}	-0.221 ± 0.006	-0.0053
S_{S_2}	0.030 ± 0.004	0.0004
S_{S_3}	0.0022 ± 0.0034	0.0014
S_{S_4}	-0.0056 ± 0.0035	0.0003
S_{S_5}	-0.065 ± 0.004	0.0009

Table 49: Observable values obtained from angular fit to the control mode in the constrained q^2 range of 7.0 to 11.0 GeV $^2/c^4$ with cut on the B^0 mass with B^0 primary vertex and J/ψ mass constraints.

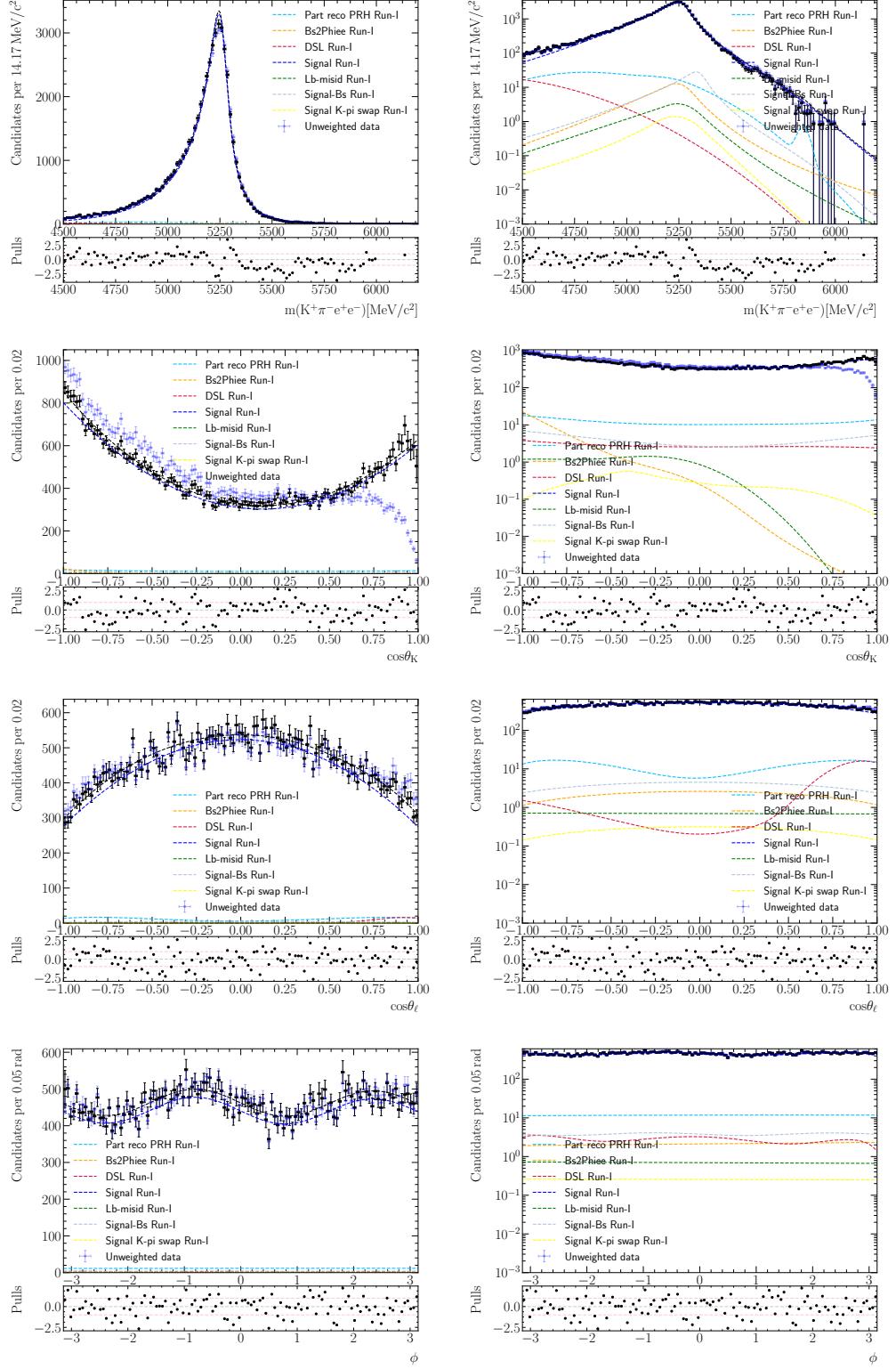


Figure 102: Result of the fit to the constrained q^2 range of 7.0 to 11.0 GeV^2/c^4 for Run-I with cut on the B^0 mass with primary vertex and J/ψ mass constraints. All peaking backgrounds are fixed to fractions (with respect to signal) calculated from simulation.

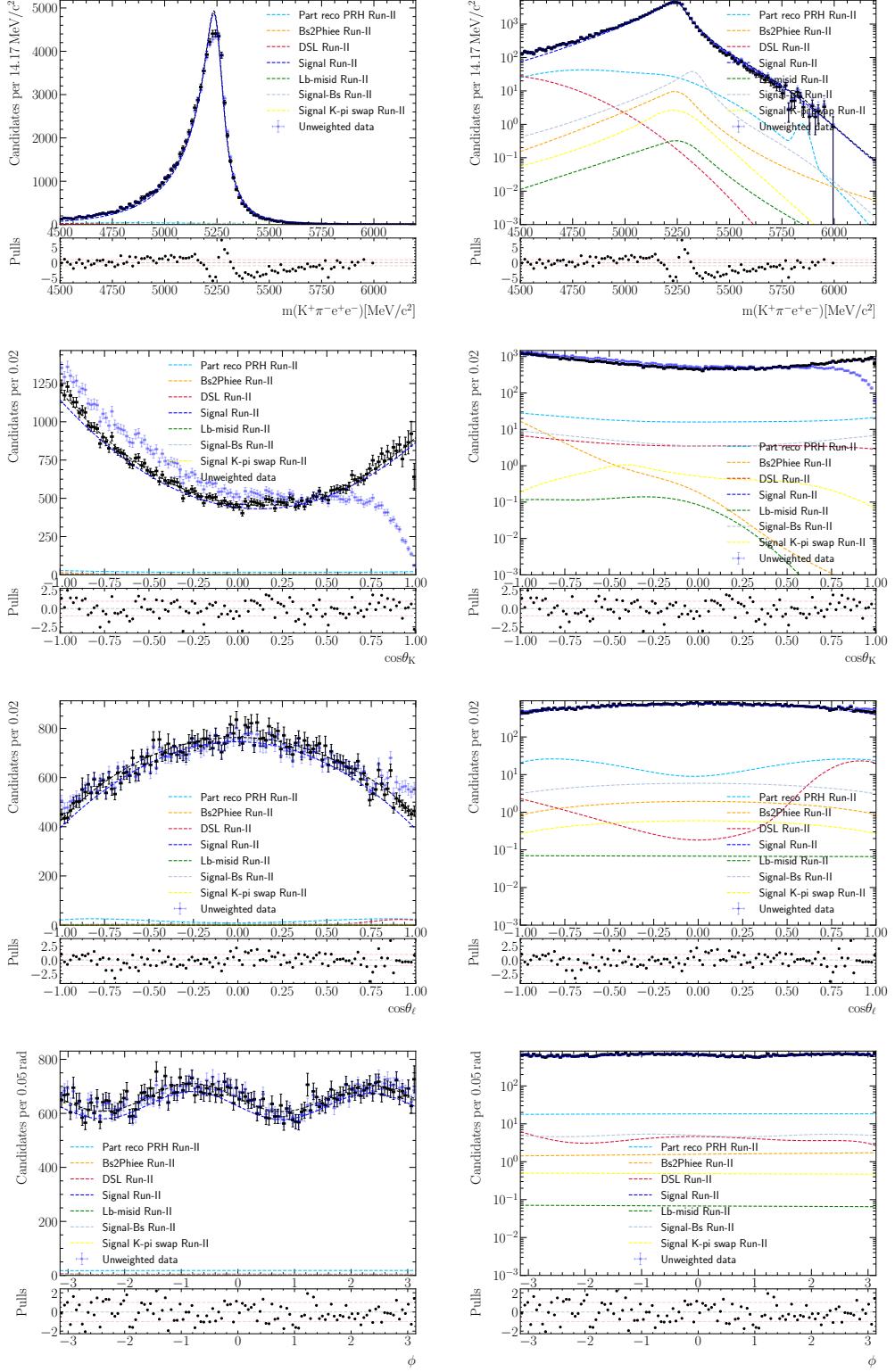


Figure 103: Result of the fit to the constrained q^2 range of 7.0 to 11.0 GeV^2/c^4 for Run-II with cut on the B^0 mass with primary vertex and J/ψ mass constraints. All peaking backgrounds are fixed to fractions (with respect to signal) calculated from simulation.

2211 10 Systematic uncertainties

2212 Sources of systematic uncertainties include the modelling of the signal and background
2213 mass and angular distributions, the parametrisation of the acceptance and resolution
2214 effects, residual backgrounds, data-simulation differences and the corrections thereof, and
2215 fit bias. The main method used to evaluate the sizes of those systematic uncertainties is the
2216 generation of large statistics pseudoexperiments. Typically one thousand large-statistics
2217 pseudoexperiments are generated with the alternative choice, and evaluated using the
2218 nominal set-up. The Gaussian mean (in some cases the width) of the difference between
2219 the observable values found using the alternative model and the nominal model is taken
2220 as the systematic uncertainty.

2221 The standard pseudoexperiment has the same settings, *i.e.* component fractions, mass
2222 and angular shapes, as those used for the sensitivity studies, with the total yield scaled up
2223 by a factor of a thousand. Modifications are made as necessary depending on the nature
2224 of the source of uncertainty under investigation. For example, the shape of the signal mass
2225 distribution may be changed from the nominal model to study the uncertainties associated
2226 with the mass modelling, while all other settings are kept unchanged. The different sources
2227 of systematic uncertainties discussed in this section are assumed to be independent, and
2228 the total uncertainty is calculated as the quadratic sum of the individual uncertainties.
2229 While the uncertainties obtained for each contribution (discussed in the following) will be
2230 added in a later stage of the review, preliminary studies are performed to examine the
2231 size of some of the dominate uncertainties and provide confidence on the measurement.
2232 These results are summarised in Section 10.1 and will be removed once all systematics are
2233 properly estimated.

2234 10.1 Preliminary systematics studies

2235 Preliminary investigations are carried out in the extended q_c^2 to assess the possible impact
2236 of three important sources of systematic uncertainties; (i) acceptance resolution corrections,
2237 (ii) DSL and (iii) combinatorial angular fixed templates. Table 50 summarises the results
2238 for the above studies. Note that these checks are not designed to replace the appropriate
2239 systematic uncertainties but instead are performed to anticipate the order of magnitude of
2240 such contributions. For comparison, the expected sensitivities to the angular observables
2241 are given in Table ??.

2242 Due to the model-dependency on the resolution corrections, acceptance functions
2243 derived from FLATQ2 samples are expected to underperform compared to PHYS accep-
2244 tance functions in this regard, *i.e.* the PHYS acceptance function is naturally a better
2245 parametrisation of the underlying physics. This is seen in simulation tests, where the PHYS
2246 acceptance correction allows for better retrieval of generator level values than that of the
2247 FLATQ2 (Section 6). To investigate the impact of this source of systematic uncertainty,
2248 around 300 signal only pseudoexperiments with the same ratio of Run-I to Run-II events
2249 as seen in data (used adjusted yields given in Section 8.5) but multiplied by one thousand
2250 (114000 events for Run-I and 170000 for Run-II), are generated with PHYS acceptances,

and fitted with corrections from either the matching PHYS acceptances, or the FLATQ2 ones. Table 50 reports the difference between the two acceptance functions. While sizeable differences are seen for the F_L and P_2 observables, these are still sub-dominant compared to the expected statistical uncertainty of the measurement.

The modelling of the DSL background from $K\pi e\mu$ data is a choice that is made based on the poor agreement between the DSL MC data and studies that suggest that the impact of using $K\pi e\mu$ rather than $K\pi ee$ is small (see Appendix I). The strategy for the assessment of the systematic uncertainty due to the modelling strategy is still under investigation. The approach taken for this preliminary check involves the usage of the DSL MC. Around 300 pseudoexperiments are generated with nominal models with the exception of DSL, which is taken from MC, and fitted with the matching DSL model and with the model extracted from $K\pi e\mu$ data. Note that here the model used differs from the nominal model only events with $\cos\theta_\ell > 0$ are considered as ‘DSL’. This quasi-nominal model is used as the DSL MC does not include events of combinatorial nature which are expected to peak approximately symmetrically at $\cos\theta_\ell = \pm 1$. Some DSL-combinatorial is nevertheless expected to remain beneath the main DSL peak, and this may inflate the systematic uncertainties for some observables. The weak point in assessing systematic using the DSL MC is that it does not include the aforementioned contribution of combinatorial nature that is included in the definition of ‘DSL’ in the data fit. The results of this investigation is given in Table 50.

Finally the impact of fixing the combinatorial angular parameters to values obtained from $K\pi e\mu$ samples is investigated using same-sign (*i.e.* $K\pi e^\pm e^\pm$) data samples. Alternative combinatorial models are made by fitting same-sign events that passed the nominal selection (with the exception of vetoes against mis-identified Λ_b^0 decays and charmonium electron-hadron/hadron-hadron events), and used to generated around 300 pseudoexperiments, which are then fitted with the nominal model. The distribution of the differences in the observable values are given in Table 50. Note that this is only intended to be a first test to probe the expected size of the systematic uncertainty. More thorough studies of the same-sign sample will be carried out to determine the optimal strategy for the final sensitivity assessments.

	DSL (MC)	Combinatorial (same-sign)	FLATQ2
F_L	0.013	0.009	0.013
P_1	0.001	0.034	0.033
P'_4	0.003	0.016	0.023
P'_5	0.009	0.014	0.039
P_2	0.039	0.041	0.018
P'_6	0.001	0.001	0.006
P'_8	0.002	0.000	0.004
P_3	0.005	0.002	0.023

Table 50: Summary of the results of preliminary systematic studies.

2281 **10.2 Signal mass shape**

2282 The nominal signal mass distribution is modelled with a single double Crystal Ball function
2283 within the nominal fit range of 4900 to 5700 MeV/ c^2 . One viable alternative is to use the
2284 combination of a double CB and a Gaussian function. In order to obtain a stable and
2285 good description of the simulated distribution it is necessary to extend the mass window
2286 to 4500 to 6200 MeV/ c^2 , which provides further flexibility to the Gaussian component as
2287 well as the modelling of the tail parameters. The impact of choosing the nominal, rather
2288 than this alternative model is evaluated using standard pseudoexperiments generated with
2289 the alternative mass shape and without acceptance, and fitted with the nominal model.

2290 The signal mass parametrisation is obtained without including the acceptance weights.
2291 While *a priori* no relevant impact is expected from this choice, the veto against B^+
2292 contributions removes events on the right-hand side of the invariant mass for events close
2293 to the upper bound of the $\cos\theta_K$ distribution. Since this region receives typically large
2294 corrections, a mild asymmetry in the weighted mass distribution could appear. This
2295 source of systematic uncertainty is assessed using pseudoexperiments generated with the
2296 alternative, weighted shapes as determined from fits to weighted simulation samples, and
2297 fitted with the nominal model. As in the data fit, the shift and scaling parameters of the
2298 nominal model are allowed to vary to adjust for the difference in shape.

2299 **10.3 Background modelling**

2300 Sources of systematic uncertainties related to background modelling include uncertainties
2301 in the determination of shape parameters from fits to simulation/data samples, and the
2302 choice of the modelling strategies. Further details are given in the sections below.

2303 **10.3.1 Fit uncertainties**

2304 In the nominal fit, all background mass and angular shapes are fixed to those determined
2305 using procedures detailed in Sec. 7, with the exception of the slope of the combinatorial
2306 background. The systematic uncertainties introduced by uncertainties in the determination
2307 of shape parameters, as well as modelling choices, namely the orders of the Chebyshev
2308 polynomials and the width of the KDE used, are evaluated using large statistics pseudoex-
2309 periments. Specifically the following factors are considered separately:

- 2310 • DSL mass slope uncertainty;
- 2311 • Uncertainties in the determination of the Chebyshev polynomial coefficients for all
2312 background angular shapes (DSL, combinatorial, partially reconstructed);
- 2313 • Alternative (higher) polynomial orders (all up to order four) for double semi-leptonic,
2314 and combinatorial angular descriptions;
- 2315 • Alternative (higher) polynomial orders (to be determined) for the angular distribu-
2316 tions of the partially reconstructed component;

- 2317 • Alternative kernel widths for the mass distribution of the partially reconstructed
2318 component.

2319 For each listed factor, a thousand pseudoexperiments are generated with the alterna-
2320 tive model, and fitted with the nominal one. Where fit uncertainties are involved, the
2321 alternative models used are generated with the slope or coefficients varied based on the fit
2322 uncertainty/covariance matrix.

2323 **10.3.2 DSL model**

2324 The nominal DSL model is obtained via a two-step procedure fit to $K\pi e\mu$ data (Section 7.3).
2325 This approach is taken as the available simulation does not appear to show good agreement
2326 with data (see Appendix I). As it is known that the trigger, reconstruction and selection
2327 process introduce some differences between electrons and muons, the systematic uncertainty
2328 due to the application of $K^+\pi^-e^+\mu^-$ derived models to $K^+\pi^-e^+e^-$ data needs to be
2329 evaluated. Two approaches are currently under investigation, the first of which involves
2330 using $K^+\pi^-e^+e^-$ events in the low mass region (below the lower bound of the signal region
2331 at $4900\text{MeV}/c^2$). In this case, the alternative DSL model will be extracted from a fit
2332 that includes additional background components, such as leakage from $B^0 \rightarrow K^{*0}J/\psi$ and
2333 partially reconstructed events. The second approach involves is the usage of a boosted
2334 decision tree (BDT) trained using $K^+\pi^-e^+\mu^-$ and $K^+\pi^-e^+e^-$ samples from the upper
2335 mass side-band of $m_B > 4900\text{MeV}/c^2$, to correct $e\mu$ distributions to that of the ee through
2336 per-event weights (the possibility of using same-sign $K^+\pi^-e^\pm e^\mp$ samples will also be
2337 considered). The algorithm involved is analogous to the one used to correct the kinematics
2338 of the simulation to better reproduce that of the data. The alternative DSL model can
2339 then be obtained by fitting the distributions of a large toy generated from the nominal
2340 DSL model with weights provided by the BDT. The impact on the observables will be
2341 assessed using pseudoexperiments generated with the alternative, and fitted with both the
2342 alternative and nominal models.

2343 **10.3.3 Combinatorial model**

2344 **Mass distribution**

2345 In the nominal set-up, separate exponential functions are used to describe the distribution
2346 of the combinatorial background in the B^0 reconstructed mass for Run-I and Run-II
2347 samples, and their slope are allowed to vary separately in the data fit. However, it is
2348 known that the usage of the constrained q_c^2 cut leads to some deviations from exponential
2349 shape ('warping'). To assess the impact of using exponential functions to describe warped
2350 distributions, alternative shapes are obtained by fitting the Run-I and Run-II $K^+\pi^-e^+\mu^-$
2351 data samples using Chebyshev polynomials up an order to be determined. A series of
2352 pseudoexperiments are then generated with this modification, and fitted using the nominal
2353 model.

2354 **Angular distributions**

2355 The shape parameters of the combinatorial background is fixed in the nominal fit to those
2356 obtained from the two-step procedure fit to $K\pi e\mu$ data (Section 7.3). As the composition
2357 of the combinatorial background for $K^+\pi^-e^+e^-$ events may be different from that of
2358 $K^+\pi^-e^+\mu^-$, this is a source of systematic uncertainty. Two strategies to obtain alternative
2359 combinatorial models are currently under investigation. The first involves fitting the
2360 same sign $K^+\pi^-e^\pm e^\pm$ sample after nominal selection, within the signal mass region. The
2361 second approach involves using the BDT described in Section 10.3.2 to correct the nominal
2362 combinatorial model. A thousand pseudoexperiments will be generated with the alternative
2363 model, and fitted with both the alternative, and nominal models. The differences in the
2364 angular observable values between the two will be assigned as systematic uncertainties
2365 due to ee , $e\mu$ shape differences.

2366 **10.3.4 Partially reconstructed background model**

2367 The nominal model used to describe the angular and mass distributions of the partially re-
2368 constructed background is parametrised from phase-space $B^+ \rightarrow K^+\pi^+\pi^-e^+e^-$ simulation
2369 reconstructed as signal, with the addition of data-driven weights. Alternatively models
2370 obtained from a cocktail of $B^+ \rightarrow K_1^{*+}(1270)e^+e^-$ and $B^+ \rightarrow K_2^{*+}(1430)e^+e^-$ simulation
2371 can be used. The systematic uncertainty associated with the strategy used to model this
2372 background is assigned by quantifying the impact on the angular observables of using either
2373 the nominal, or the alternative model. This is done using a thousand pseudoexperiments
2374 generated using the alternative model, and fitted with either the alternative, or the nominal
2375 model. The mean values of the observable difference distributions are taken as systematic
2376 uncertainties.

2377 **10.4 Acceptance function**

2378 Sources of systematic uncertainties associated with the combined acceptance and resolution
2379 functions include negative regions, the choice of the polynomial order, uncertainties of
2380 the calculated coefficients, and the model dependency of resolution correction. The
2381 contribution from each of these sources is discussed below.

2382 **10.4.1 Negative regions**

2383 Due to limited statistics, the acceptance functions determined through the method of
2384 moments calculation detailed in Section 6 contains regions where the functions take on
2385 negative values ('holes'). Events within these regions do not have well defined correction
2386 weights, and are therefore removed in fits. This removal of events from parts of the phase
2387 space without corresponding adjustment of the signal pdf leads to bias. The size of this
2388 bias is evaluated using a thousand large statistics, signal only pseudoexperiments generated
2389 without acceptance. Fits are made to those pseudoexperiments with, and without the
2390 removal of events that lie within the holes of the nominal acceptance functions. The
2391 differences in the observable values found is used to set the systematic uncertainty.

2392 **10.4.2 Higher order description**

2393 The nominal acceptance functions are parametrised using the lowest orders that show satis-
2394 factory modelling of the post-selection q_c^2 and angular distributions in the one dimensional
2395 projections. The maximum orders used for q_c^2 , $\cos\theta_K$, $\cos\theta_\ell$ and ϕ are three, five, four and
2396 six, respectively. This choice avoids overfitting, and reduces computational requirements.
2397 It also avoids possible oscillations near the phase space boundaries that can occur when
2398 high polynomial orders are used (Runge phenomena). To assess the systematics associated
2399 with this choice, the parametrisations are repeated with all polynomial orders increased
2400 by 3. A thousand high statistics, signal only pseudoexperiments are generated using
2401 these alternative acceptance functions, and fitted with the nominal ones. The size of the
2402 systematic uncertainty associated with the choice of the polynomial order is based on the
2403 differences between observable values found in these two cases.

2404 **10.4.3 Coefficient uncertainties**

2405 The uncertainty related to the determination of the acceptance function coefficients is
2406 regarded as a source of systematic uncertainty. To determine the size of this contribution,
2407 the coefficient values determined in the method of moments calculation, and the associated
2408 covariance matrix are used to set the mean and width of a multivariate gaussian distribution,
2409 from which new sets of coefficients can be generated. These correspond to alternative
2410 acceptances varied from the nominal one based on the coefficient uncertainties. To
2411 evaluate the systematic associated with uncertainties in the shape of the acceptance
2412 functions, a thousand pseudoexperiments are generated with a thousand pairs of alternative
2413 acceptances (Run-I and Run-II acceptances), and fitted with both the alternative, and
2414 nominal corrections. The Gaussian widths of the difference distributions (difference
2415 between alternative and nominal) of the observables are taken as systematic uncertainties.

2416 **10.4.4 Resolution model dependence**

2417 The ‘acceptance’ functions used to correct post-selection distributions back to their
2418 generator-level state takes into account the effects of both acceptance and resolution. The
2419 nominal acceptance functions are obtained by parametrising rare mode simulation generated
2420 flat in the angles and q^2 (FLATQ2). The alternative approach is to obtain acceptance
2421 functions using simulation generated with physics content (PHYS). The underlying pure
2422 acceptance functions are expected to be the same in both cases, as they share the same
2423 decay topology and kinematics. However, resolution effects, which can be thought of as the
2424 smearing of q^2 , and to a lesser extent, the angular distributions with a kernel, results in
2425 different effective functions, the shapes of which will vary depending on the generator-level
2426 distributions. This ‘model dependency’ of the resolution results in differences between the
2427 nominal acceptance functions, and the acceptance functions obtained from the simulation
2428 with physics content, which is expected to resemble that of the data more closely. The
2429 difference between the FLATQ2, and the PHYS acceptance is then taken to be the systematic
2430 uncertainty due to the model dependency of the resolution correction. The size of this

2431 contribution is evaluated using a thousand signal only pseudoexperiments, generated with
2432 PHYS acceptance functions, and fitted with corrections from either the PHYS acceptance
2433 functions, or that of the FLATQ2. The mean of the differences in observable values found
2434 in the two cases are assigned as systematic uncertainties.

2435 10.4.5 Excluding BKGCAT 60

2436 Events with BKGCAT 60 (at least one final-state particle used to form the candidate is a
2437 ghost) are not used in the acceptance parametrisation. This choice is made considering
2438 that those events are artifacts of the reconstruction process. However they are expected
2439 to be present in the data, as they are inseparable from signal using the mass distribution,
2440 and around 5% of events in the rare (and control) mode simulation after full selection
2441 have BKGCAT of 60. To assess the systematic uncertainty related to using BKGCAT of
2442 0, 10 and 50 only, alternative acceptance functions are parametrised including BKGCAT
2443 60 events. A thousand signal only pseudoexperiments are then generated with these
2444 alternative functions, and fitted with either the alternative, or the nominal correction.
2445 The mean values of the distributions of differences in fitted observables are taken to be
2446 the corresponding systematic uncertainties.

2447 10.5 J/ψ leakage

2448 Leakage from $B^0 \rightarrow K^{*0} J/\psi$ decays into the signal region is strongly suppressed by
2449 the q_c^2 cut, which allows for the omission of this component in the fit (see Section 4.3).
2450 Nevertheless, an estimated total of around 13 events (5 ± 1 for Run-I and 8 ± 1 for Run-II)
2451 is expected to remain. To assess the systematic uncertainty related to neglecting this
2452 background, a thousand pseudoexperiments are generated with the nominal model but
2453 including extra J/ψ leakage components parametrised from $B^0 \rightarrow K^{*0} J/\psi$ simulation,
2454 with yields fixed to the aforementioned values. Fits are then made to those toys with and
2455 without the inclusion of this component, and the mean values of the difference distributions
2456 for the observables are taken as systematic uncertainties.

2457 10.6 Neglecting S-wave

2458 Due to the lack of sensitivity, the angular pdf used in the nominal fit does not include
2459 S-wave and S-P interference terms (Section 5.2). The impact of neglecting S-wave related
2460 contributions on the angular observables is assessed using pseudoexperiments generated
2461 with additional terms (Equ. 27) and fitted with and without their inclusion. The mean
2462 values of the distribution of differences between the results from those two fits for the
2463 observables are taken to be the systematic uncertainties.

2464 10.7 Control mode mass fit

2465 The control mode mass distribution is fitted simultaneously with the rare mode to gain
2466 additional control over the shift and scaling parameters of the signal mass peak that

account for remaining simulation-data differences. The nominal control mode fit includes three components – $B_s \rightarrow K^{*0} J/\psi$, $\Lambda_b^0 \rightarrow pK J/\psi (\rightarrow e^+e^-)$ with $p - \pi$ mis-identification, and combinatorial. However, it is known that additional backgrounds contribute at low levels (see Section 9.4.3 and Appendix J). Neglecting them may lead to altered parameter values. To evaluate the systematic uncertainty related to the control mode mass fit, the data fit (mass only) is repeated with additional background components, namely partially reconstructed, $B_s^0 \rightarrow \phi J/\psi$, DSL, and control mode events with $K - \pi$ swap, and compared to the nominal fit results. This alternative set of parameters are then used to generate rare mode pseudoexperiments, which are fitted with those values fixed¹⁴ to either the nominal values, or this alternative set. Finally, the mean of the difference distributions of the angular observables are taken as systematic uncertainties.

10.8 Agreement between data and simulation

Sources of systematic uncertainties considered for the simulation correction strategy used in this analysis include the settings of the BDT used to correct for discrepancies in kinematic, multiplicity and reconstruction variables, statistics of the PID calibrations samples and the widths of the kernels used, and aspects of the trigger correction strategy. Alternative corrections mainly affect the observables of interest through their impact on the acceptance functions, which are parametrised using fully corrected simulation samples. Therefore quantifying these systematic uncertainties mainly consists in producing pseudoexperiment with alternative acceptance functions. Details regarding the methods used in each case are given in the sections below.

10.8.1 Kinematic correction

The nominal procedure to correct for discrepancies in selected kinematic, multiplicity and reconstruction variables that are known to show deviations from data involves the use of a Boosted Decision Tree (BDT) to carry out a multidimensional reweighting. The decision is made to use the particular set-up of the BDT described in Section 3.6.3, although viable alternative choices exist in 1) the settings of its hyperparameters, and 2) the usage of a two-stage procedure that involves mismodelling correction at the generation level. Acceptance functions parametrised using correction weights resulting from those alternative choices are used to set the systematic uncertainty of this procedure. For Case 1, a scan over the hyperparameter space is performed and 100 configurations are selected. The BDT is retrained 100 times using those settings, and the same number of alternative correction weights are obtained. These are then used in the parametrisation of acceptance functions. Large statistics, signal only pseudoexperiments are generated with those alternative functions, and fitted with either the alternative, or the nominal set. The Gaussian mean and width of the resulting parameters differences distributions will be

¹⁴Note that since the inclusion of the statistically dominant control mode mass distribution is expected to be negligibly different from fixing those values in the rare mode fit, it is not included to speed up computations.

2503 added in quadrature to produce the final values of the associated systematics. For Case 2,
2504 a single set of alternative correction weights is obtained using the two-stage procedure,
2505 and used to parametrise a set of acceptance functions. Likewise, pseudoexperiments are
2506 generated with the alternative set and fitted with either the alternative or the nominal
2507 acceptance functions. The resulting mean values of the difference distributions for the
2508 observables are taken as systematic uncertainties.

2509 **10.8.2 PID systematic**

2510 For this analysis, the **Meerkat** framework is used to correct the simulated PID variables.
2511 This approach involves transforming the PID distributions of the simulation to match that
2512 of the data (taken from calibration samples in the **PIDCalib** package), while preserving the
2513 correlations between them (see Section 3.6.1). Two sources of systematics are envisaged
2514 for this strategy. They include calibration sample statistics, and the widths of the kernels
2515 used to describe the PID distributions.

2516 **Calibration sample statistics**

2517 To assess the systematic uncertainty due to limited calibration sample statistics, alternative
2518 PID templates are obtained from bootstrapped samples, and used to generate variations
2519 of the PID variables. These are then used to select samples for the parametrisation of new
2520 acceptance functions. Pseudoexperiments are generated with the alternative functions,
2521 and fitted with both the matching and nominal acceptances. The widths of the usual
2522 distribution of differences in observable values are taken as the systematic uncertainties.

2523 **Kernel widths settings**

2524 The PID correction strategy requires the description of multi-dimensional PID distributions
2525 for simulation and data. This is achieved using an adaptive kernel-based technique.
2526 To assess the systematic uncertainties related to the parametrisation of these samples,
2527 alternative PID variables are made with the widths of the kernels increased by 50%. These
2528 are then used to produce new acceptance functions. The usual pseudoexperiment studies
2529 are then performed to quantify the impact on the angular observables.

2530 **10.8.3 Trigger corrections**

2531 Sources of systematic uncertainty associated with the trigger correction strategy considered
2532 include the choice of the binning approach, assumption of factorisation for the LOE category,
2533 and the biases caused by TIS calibration histograms.

2534 **Binning strategy**

2535 In the nominal strategy, L0 correction weights are assigned to each event based on binned
2536 calibration histograms. An alternative option is to fit the trigger turn-on curves, and
2537 assign weights based on the resulting functions. To assess the systematics associated with

2538 the nominal choice, the impact of those two approaches on the acceptance functions are
2539 investigated using pseudoexperiments.

2540 **Assumption of factorisation for the LOE category**

2541 Currently, correction weights are assigned to events of the LOE category assuming fac-
2542 torisation of the trigger efficiencies for the two electrons. To quantify the impact of
2543 this choice, an alternative calibration histogram will be calculated as a function of the
2544 maximum transverse energy deposited by the electrons that triggered the event , *i.e.*
2545 $\max E_T^{L0}(e_1 \cdot e_1^{\text{TOS}}, e_2 \cdot e_2^{\text{TOS}})$, and of the region of the ECAL where such electron was
2546 detected. The generated correction will be applied to the simulated sample following the
2547 same logic. By considering the two electrons together no assumption on the factorisation
2548 of the correction is made.

2549 **Biases caused by the TIS calibration histograms**

2550 As shown in Section 3.6.4, the TIS efficiency shows a tag-bias when using the triggered
2551 signal electrons as tagging selection. Hence, using an electron tag is strictly speaking not
2552 correct to compute the efficiency of the LOI category. In the case of the TIS efficiency,
2553 no alternative tag can be used since using the events triggered by the hadronic system
2554 would cause an even bigger bias. Instead, an alternative set of LOI weights is extracted
2555 by computing the ratio between the TIS efficiency computed in $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$
2556 data using the electrons as a tag, and the TIS efficiency computed on simulated events
2557 requiring the electrons not to have fired the trigger. The systematic will be computed as
2558 the difference between the values of the angular observables obtained with then nominal
2559 acceptance and the acceptance obtained using these weights. The correlation between
2560 this bias in Run 1 and in Run 2 is assumed to be 100 %. This corresponds to making
2561 the hypothesis that the electron tag bias on the TIS efficiency do not cancel at all when
2562 computing the ratio of TIS efficiencies between the data and the simulation.

2563 **10.9 $B^+ \rightarrow K^+ e^+ e^-$ veto**

2564 The veto against B^+ decays is known to distort the combinatorial distribution at high
2565 mass, and introduce correlations between mass and $\cos\theta_K$. Figure 104 shows the effect of
2566 this kinematic cut in $K\pi e\mu$ samples. While the direct impact of this cut on the signal
2567 is negligible, the distortion it causes may create problems for the determination of the
2568 signal angular shape indirectly. Note that while the nominal combinatorial model is
2569 obtained with the B^+ veto, the correlation between B^0 mass and $\cos\theta_K$ is ignored, as
2570 are distortions to the B^0 mass distribution. This results in imperfect modelling, and
2571 necessitates the evaluation of this systematic uncertainty. To quantify the associated
2572 systematic uncertainties, a model characterising the effect of the veto is first made using
2573 $K\pi e\mu$ samples after loose preselection. Specifically the sample is binned in q^2 , $\cos\theta_K$ and
2574 B^0 mass before and after the application of the veto, from which efficiency values can
2575 be calculated for each bin. One thousand pseudoexperiments are then generated with
2576 combinatorial models parametrised from data without the veto. Fits are made to those

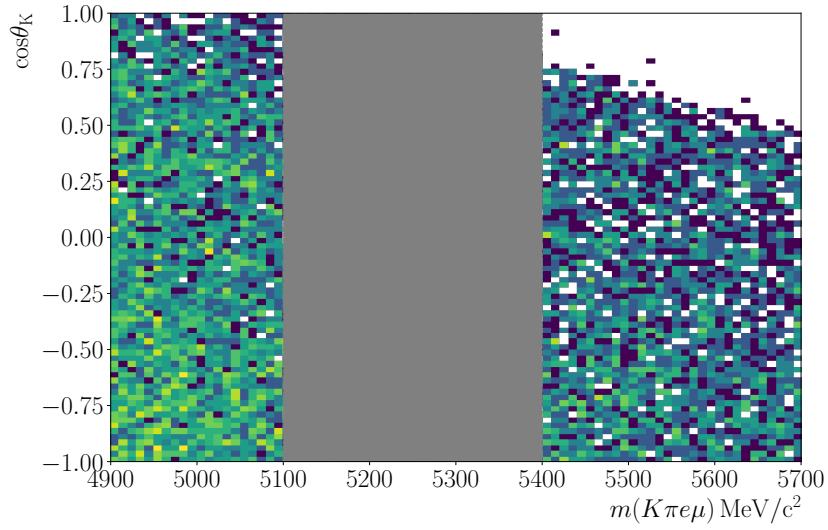


Figure 104: Effect of the B^+ veto on $K\pi e\mu$ data sample. No MVA requirements are applied. The veto removes in the upper mass region of B^0 mass, near $\cos\theta_K = 1$. The region close to the signal peak is not shown (gray band).

2577 toys with the matching configuration. The veto is then applied to the toys using the
 2578 binned model, and they are fitted a second time with the nominal model. The differences
 2579 between the mean values of the observables obtained in these two cases are taken as the
 2580 systematic uncertainties.

2581 10.10 Fit bias

2582 The results of the sensitivity studies detailed in Section 8.5 show that biases are expected
 2583 for a number of observables, in particular A_{FB} and F_L . Those biases are added to the
 2584 systematic uncertainties.

2585 **11 Cross checks**

The scripts used in the cross checks and pre-unblinding procedure are given in

- 2586 ○ `ewp-bd2ksteeangular-central-q2/scripts/xchecks(unblinding)/`

2588 This section details the studies developed to check the stability and consistency of
2589 the results. Cross checks presented in Sec. 11.1 are performed before starting the
2590 unblinding procedure and their results will be documented in next versions of the note.
2591 Cross checks discussed in Sec. 11.2 are instead part of the pre-unblinding strategy and will
2592 be completed only with the approval of the RC.

2593 **11.1 Modifications to the nominal fit**

2594 **11.1.1 Double semi-leptonic modelling**

2595 In the nominal fit, background from double semi-leptonic cascade (DSL) decays is modelled
2596 using a data-driven approach (Section 7.3). While systematic uncertainties associated to
2597 this choice are considered in the previous section, it is interesting to check the stability of
2598 the result by significantly suppressing this contribution – but also removing an important
2599 portion of the signal. One valid alternative strategy is to veto this contribution by applying
2600 a symmetric cut on $\cos\theta_\ell$, typically $|\cos\theta_\ell| < 0.8$. This cut has a relatively low signal
2601 efficiency of $(90.0 \pm 0.2)\%$ for Run-I and $(89.0 \pm 0.1)\%$ for Run-II, and rejects around 78%
2602 of DSL events from simulation (due to known simulation-data differences this value may be
2603 unreliable). If the DSL modelling in the nominal strategy is significantly flawed, reducing
2604 this background can lead to a more robust determination of the angular observables,
2605 though with reduced sensitivity. This possibility will be investigated using blinded data
2606 fits and pseudoexperiments.

2607 **11.1.2 q^2 selection**

2608 Cutting only on the constrained q_c^2 distribution can lead to the inclusion of a small number
2609 of events of combinatorial nature that peaks in the high mass region (see Section 4.3), which
2610 can potentially distort the determination of the slope of the combinatorial background.
2611 Figure 105 shows that those events are reduced by cutting on both the constrained and
2612 unconstrained q^2 , *i.e.* applying a cut such as $1.1 < q_c^2 < 7.0 \& 1.1 < q^2 < 7.0 \text{ GeV}^2/c^4$.
2613 While preliminary studies suggest that the amount of remaining spurious events in the
2614 data after the full selection is marginal, the fit will be repeated with a square cut in order
2615 to validate benchmark data fit results.

2616 **11.2 Pre-unblinding cross-checks**

2617 Two pre-unblinding checks are performed to check the consistency of the angular observables
2618 in different subsamples. These include investigations of

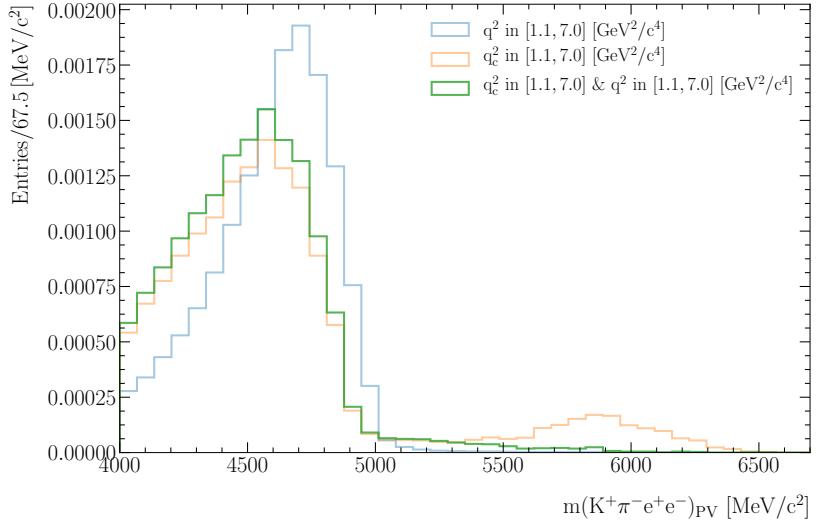


Figure 105: Distribution of $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ simulated candidates as a function of PV-constrained B^0 invariant mass $m(K^+ \pi^- e^+ e^-)_{\text{PV}}$. The offline signal selection is applied except for the MVA cut. Truth-matching has not been applied, and as a result, a bump is visible at high mass when the constrained q^2 cut is applied. These events have background categories greater than 60 and are combinatorial in nature. Therefore, after the MVA requirements are applied this contribution is reduced to a marginal level. The addition of unconstrained q^2 cut to the nominal q_c^2 cut reduces the peak to negligible levels even in the absence of tighter MVA requirements.

- 2619 1. the compatibility between data taking periods (Run-I, and Run-II up to 2016); and
 2620 2. the compatibility between L0 trigger categories (LOE and LOTIS).

2621 For these checks, the following test-statistic T_{check} is defined

$$T_{\text{check}} = (\vec{S}_1 - \vec{S}_2)(\Sigma_1 + \Sigma_2)^{-1}(\vec{S}_1 - \vec{S}_2)^T \quad (52)$$

2622 where \vec{S}_i denote the angular observables in the i -th subsample and Σ_i the corresponding
 2623 covariance matrix.

2624 In both cases, the distribution of T_{check} is studied with pseudoexperiments generated
 2625 from the (blind) full fit results. Pseudoexperiments are then fitted separately for the
 2626 Run-I and Run-II sub sets of pseudo-data for Check 1, and for LOE and LOTIS for Check
 2627 2 (where the two Runs are combined together). When fitting the two trigger categories
 2628 separately, dedicated background models and acceptance functions are determined from
 2629 LOE (LOTIS) only samples. Further splitting of the dataset becomes impossible due to the
 2630 extremely low statistics of the samples. Finally, based on the observed value of T_{check} in
 2631 data and the corresponding distribution obtained from toys, the compatibility between
 2632 each of the two pair of subsamples will be evaluated.

2633 12 Results

2634 12.1 Blinded data fits

2635 The angular observables are determined in the extended and reduced q_c^2 ranges of $1.1 < q_c^2 <$
2636 7.0 and $1.1 < q_c^2 < 6.0$ with the weighted unbinned maximum likelihood method described
2637 in Section 8.1. Four components are used for the rare mode angular fit; signal, DSL,
2638 combinatorial and partially reconstructed For the control mode mass fit three components
2639 are obtained; $B_s \rightarrow K^{*0} J/\psi(e^+e^-)$, combinatorial and mis-identified Λ_b^0 decays. Per-event
2640 weights from acceptance functions described in Section 6.4 are used. The full fit is made
2641 simultaneously to both Runs, sharing the values of the angular observables as discussed
2642 in Section 8.3. The control mode mass fit is included to enable the stable determination
2643 of parameters of the signal mass peak that take into account remaining simulation-data
2644 differences. Separate background models, acceptance functions, and shift and scaling
2645 parameters are used for Run-I and Run-II samples. In total, 24 parameters are allowed to
2646 vary in the fit – the S- or P-basis angular observables (8), mass distribution shift and scaling
2647 parameters (4), combinatorial slope parameters (4 in total, including two for the control
2648 mode), and the fractions of signal, partially reconstructed, and DSL contributions (8 in
2649 total, including two signal fractions for the control mode). The fraction of combinatorial
2650 events is given as the remainder after subtracting all other component fractions from
2651 unity. The parameter uncertainties are obtained after applying corrections detailed in
2652 Section 8.4 to the covariance matrix output of Hesse. Four fits are made in total – one
2653 for each observable basis of the two q_c^2 ranges. Figures 106, 107, 108 and 109 show the
2654 results of the fit, and Tables 51 and 52 provides the parameter values obtained. As the
2655 angular observables remain blinded, only mass distributions and nuisance parameters are
2656 provided. In general, a good agreement is observed in all projections.

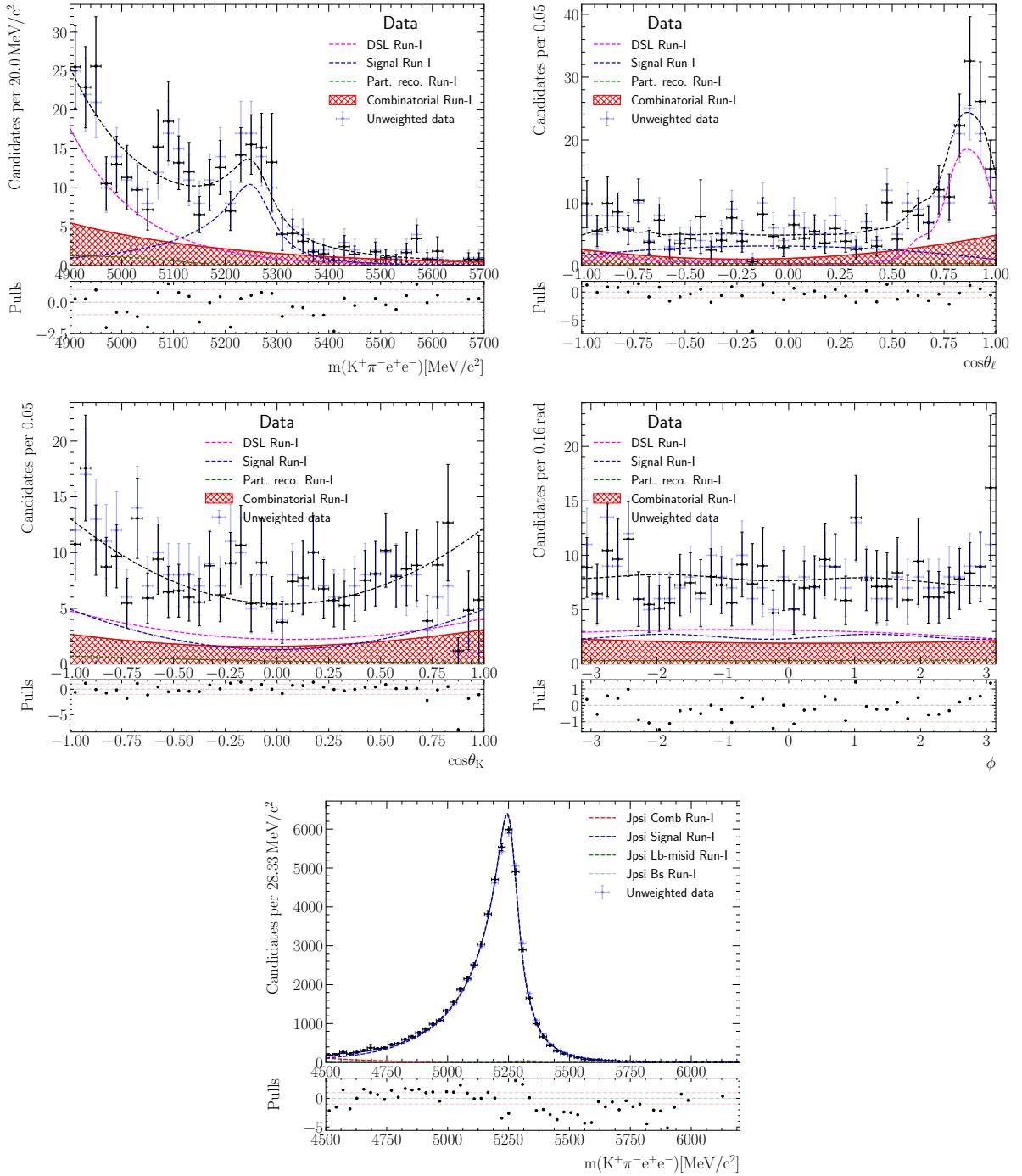


Figure 106: Data fit results with the Run-I dataset for the P-basis in the extended q_c^2 range of $1.1 < q_c^2 < 7.0$. The projections correspond to the rare mode (top left) invariant mass, (top right) $\cos\theta_\ell$, (middle left) $\cos\theta_K$, (middle right) ϕ and (bottom) control model invariant mass.

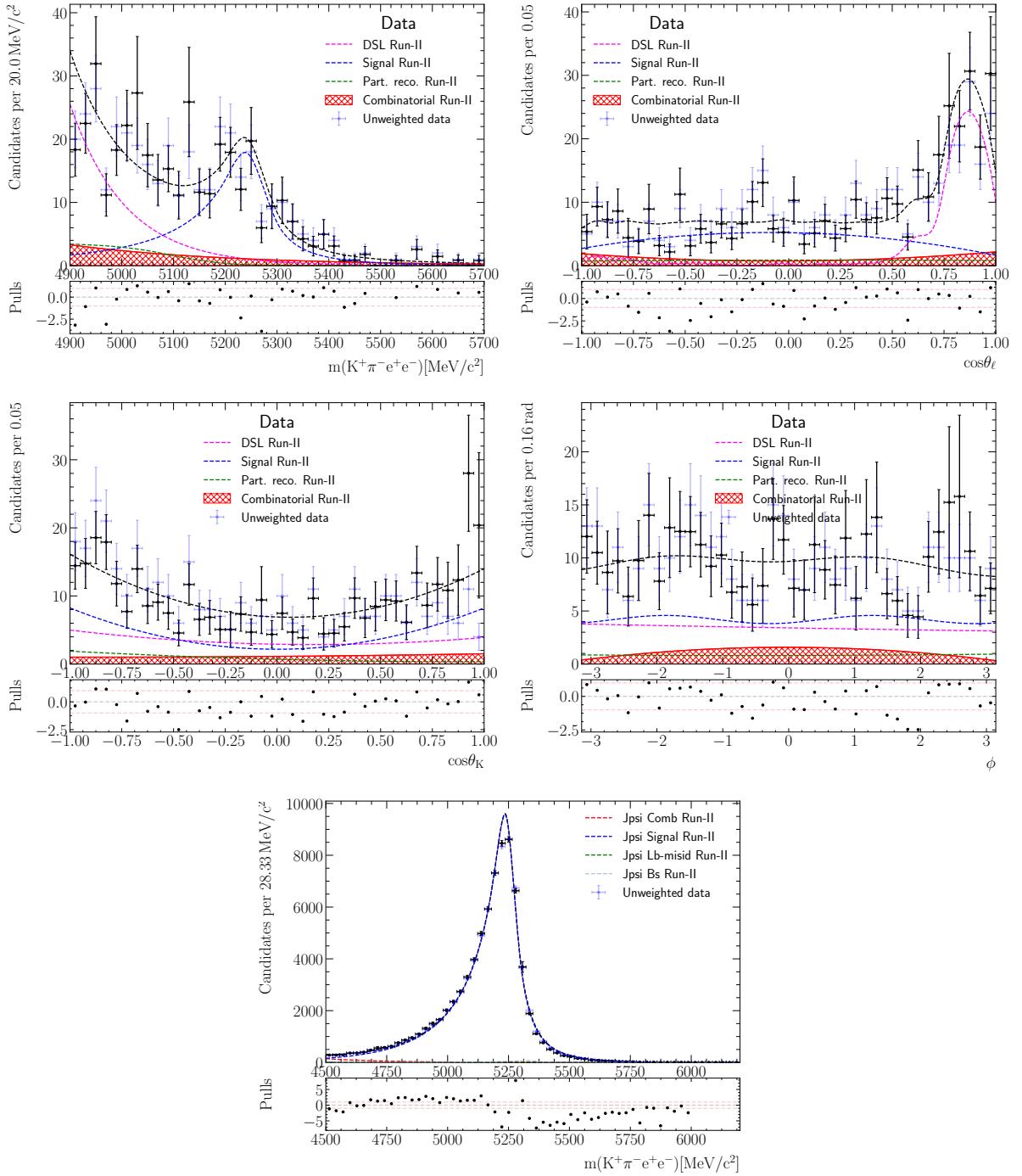


Figure 107: Data fit results with the Run-II dataset for the P-basis in the extended q_c^2 range of $1.1 < q_c^2 < 7.0$. The projections correspond to the rare mode (top left) invariant mass, (top right) $\cos \theta_\ell$, (middle left) $\cos \theta_K$, (middle right) ϕ and (bottom) control model invariant mass.

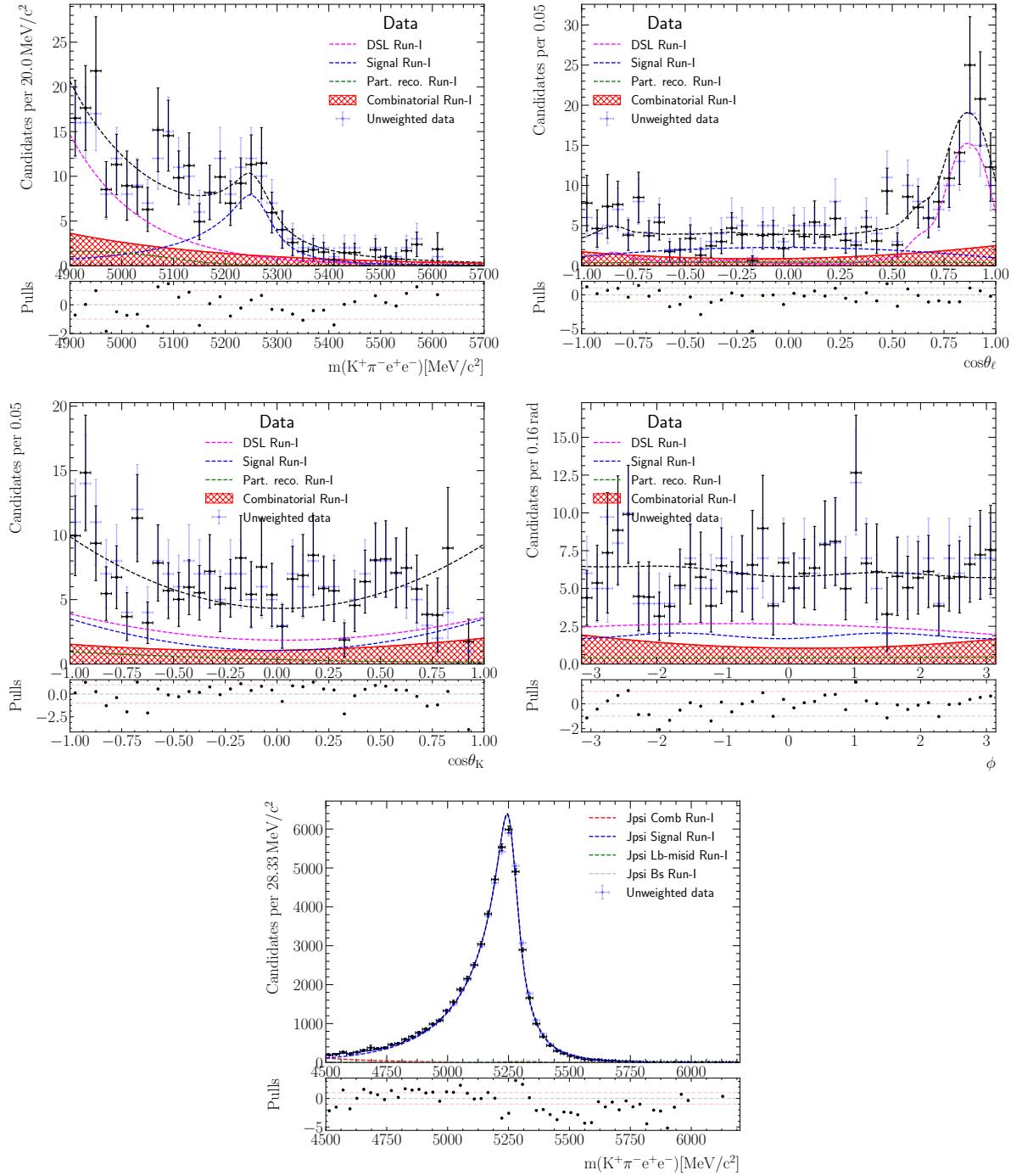


Figure 108: Data fit results with the Run-I dataset for the P-basis in the reduced q_c^2 range of $1.1 < q_c^2 < 6.0$. The projections correspond to the rare mode (top left) invariant mass, (top right) $\cos \theta_\ell$, (middle left) $\cos \theta_K$, (middle right) ϕ and (bottom) control model invariant mass.

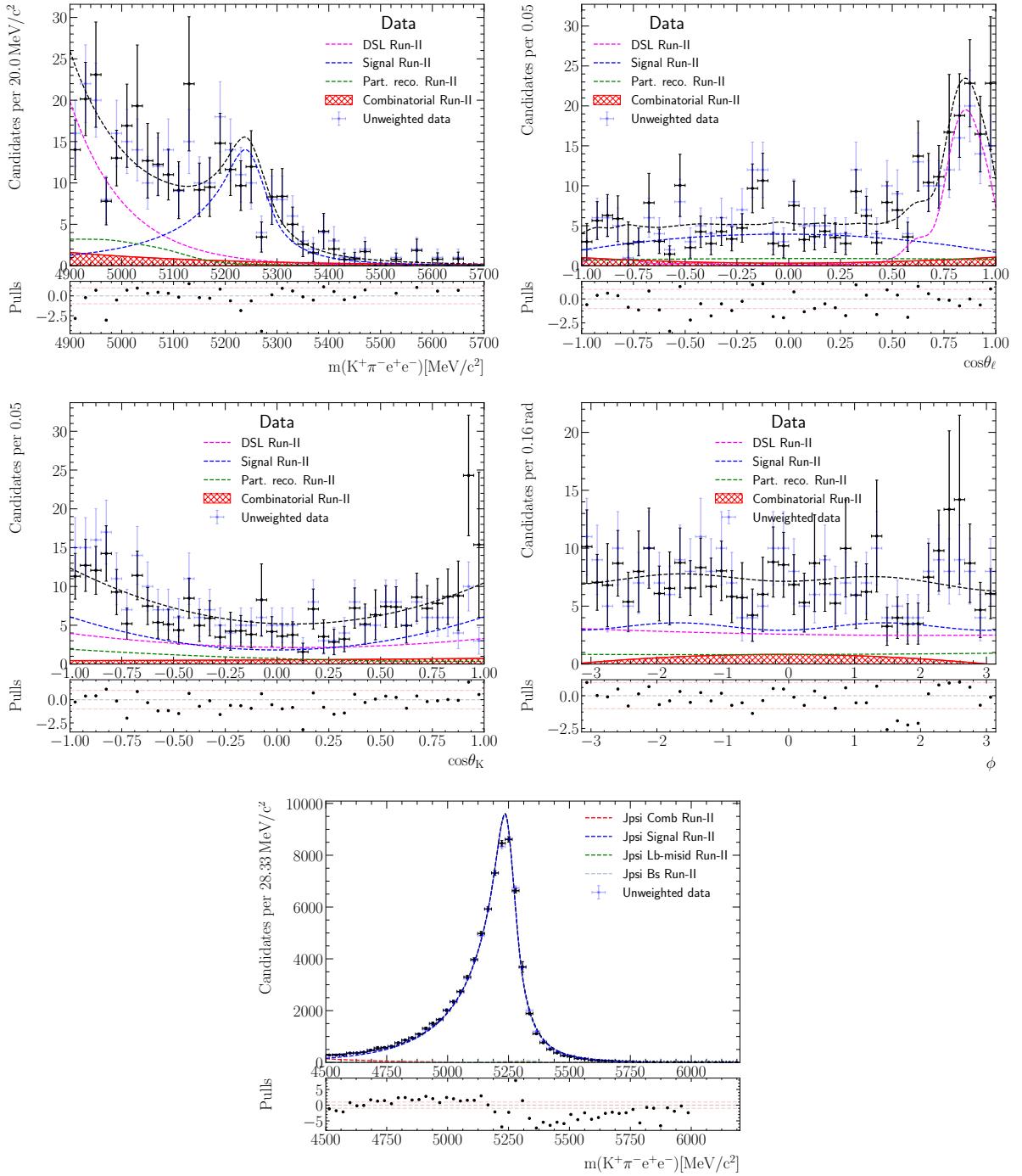


Figure 109: Data fit results with the Run-II dataset for the P-basis in the reduced q_c^2 range of $1.1 < q_c^2 < 6.0$. The projections correspond to the rare mode (top left) invariant mass, (top right) $\cos\theta_\ell$, (middle left) $\cos\theta_K$, (middle right) ϕ and (bottom) control model invariant mass.

Parameter	Result	
$f_{\text{sig}}^{\text{Run-I}}$	0.32	± 0.06
$f_{\text{sig}}^{\text{Run-II}}$	0.44	± 0.05
$f_{\text{DSL}}^{\text{Run-I}}$	0.38	± 0.06
$f_{\text{DSL}}^{\text{Run-II}}$	0.36	± 0.04
$f_{\text{PR}}^{\text{Run-I}}$	0.04	± 0.06
$f_{\text{PR}}^{\text{Run-II}}$	0.09	± 0.04
$\delta_{\mu}^{\text{Run-I}}$	-9.0	± 0.5
$\delta_{\mu}^{\text{Run-II}}$	-17.8	± 0.7
$\delta_{\sigma}^{\text{Run-I}}$	1.104	± 0.008
$\delta_{\sigma}^{\text{Run-II}}$	1.076	± 0.006
$\lambda_{\text{Comb}}^{\text{Run-I}}$	-0.0031	± 0.0013
$\lambda_{\text{Comb}}^{\text{Run-II}}$	-0.0033	± 0.0012
$f_{\text{sig}J/\psi}^{\text{Run-I}}$	0.9726	± 0.0020
$f_{\text{sig}J/\psi}^{\text{Run-II}}$	0.9761	± 0.0014
$\lambda_{\text{Comb}J/\psi}^{\text{Run-I}}$	-0.00403	± 0.00015
$\lambda_{\text{Comb}J/\psi}^{\text{Run-II}}$	-0.00413	± 0.00011

Table 51: Parameter values obtained from the P-basis fit to the extended q_c^2 range of $1.1 < q_c^2 < 7.0$. As the final result remains blinded, only nuisance parameters are shown.

Parameter	Result	
$f_{\text{sig}}^{\text{Run-I}}$	0.31	± 0.07
$f_{\text{sig}}^{\text{Run-II}}$	0.44	± 0.05
$f_{\text{DSL}}^{\text{Run-I}}$	0.41	± 0.07
$f_{\text{DSL}}^{\text{Run-II}}$	0.36	± 0.05
$f_{\text{PR}}^{\text{Run-I}}$	0.07	± 0.06
$f_{\text{PR}}^{\text{Run-II}}$	0.12	± 0.04
$\delta_{\mu}^{\text{Run-I}}$	-9.0	± 0.5
$\delta_{\mu}^{\text{Run-II}}$	-17.8	± 0.7
$\delta_{\sigma}^{\text{Run-I}}$	1.104	± 0.008
$\delta_{\sigma}^{\text{Run-II}}$	1.076	± 0.006
$\lambda_{\text{Comb}}^{\text{Run-I}}$	-0.0032	± 0.0013
$\lambda_{\text{Comb}}^{\text{Run-II}}$	-0.0034	± 0.0022
$f_{\text{sig}J/\psi}^{\text{Run-I}}$	0.9726	± 0.0020
$f_{\text{sig}J/\psi}^{\text{Run-II}}$	0.9761	± 0.0014
$\lambda_{\text{Comb}J/\psi}^{\text{Run-I}}$	-0.00403	± 0.00015
$\lambda_{\text{Comb}J/\psi}^{\text{Run-II}}$	-0.00413	± 0.00011

Table 52: Parameter values obtained from the P-basis fit to the reduced q_c^2 range of $1.1 < q_c^2 < 6.0$. As the final result remains blinded, only nuisance parameters are shown.

2657 **12.2 Unblinded fit results**

2658 This section will be filled once the analysis has received the approval to unblind by the
2659 review committee.

2660 A PID response using Meerkat

2661 A novel approach to correct the particle identification response in simulated samples
 2662 has recently become available [94]. The method is based on non-parametric unbinned
 2663 description of the PID response using kernel density estimation. Since the documentation
 2664 at this stage is not available, its current stage is reproduced below.

2665 A.1 Correction of PID response

2666 First, we remind how resampling of random variables is performed, as applied to the
 2667 problem of PID response correction. Consider that the PDF of the PID variable x is
 2668 known for any pair of kinematic variables p_T and η ¹⁵ and the number of tracks N_{tr} from
 2669 the calibration sample. We denote this PDF as $p_{\text{exp}}(x|p_T, \eta, N_{\text{tr}})$. The corresponding
 2670 cumulative distribution function in the PID variable x is

$$P_{\text{exp}}(x|p_T, \eta, N_{\text{tr}}) = \int_{-\infty}^x p_{\text{exp}}(y|p_T, \eta, N_{\text{tr}}) dy. \quad (53)$$

2671 Its inverse transformation $x = P^{-1}(\xi|p_T, \eta, N_{\text{tr}})$ provides the way to generate the vari-
 2672 able x : if ξ is distributed uniformly in the region $(0, 1)$, then the distribution of
 2673 $x = P_{\text{exp}}^{-1}(\xi|p_T, \eta, N_{\text{tr}})$ is $p_{\text{exp}}(x|p_T, \eta, N_{\text{tr}})$.

2674 In the resampling approach, the new PID variable x is drawn completely randomly
 2675 from the distribution of the calibration sample, and the original variable x_{MC} generated in
 2676 simulation is discarded. The alternative approach proposed here is to use the simulated
 2677 variable x_{MC} to obtain the sampling variable ξ from the cumulative distribution $P_{\text{MC}}(x)$
 2678 corresponding to the simulated PID response:

$$\xi = P_{\text{MC}}(x_{\text{MC}}|p_T, \eta, N_{\text{tr}}) = \int_{-\infty}^x p_{\text{MC}}(y|p_T, \eta, N_{\text{tr}}) dy, \quad (54)$$

2679 where $p_{\text{MC}}(x|p_T, \eta, N_{\text{tr}})$ is the distribution of simulated PID variable x as a function of p_T ,
 2680 η and N_{tr} (Fig. 110).

2681 As a result, we obtain the transformation of variable x_{MC} into x_{corr} that depends on
 2682 the parameters p_T, η, N_{tr} :

$$x_{\text{corr}} = P_{\text{exp}}^{-1}(P_{\text{MC}}(x_{\text{MC}}|p_T, \eta, N_{\text{tr}})|p_T, \eta, N_{\text{tr}}). \quad (55)$$

2683 An important property of the dependence $x_{\text{corr}}(x_{\text{MC}})$ is that, if the difference between
 2684 two distributions $p_{\text{exp}}(x)$ and $p_{\text{MC}}(x)$ is small, x_{MC} and x_{corr} are highly correlated. The
 2685 strong correlation of the corrected and original PID response has important consequences:

¹⁵In **PIDCalib**, the binning is usually performed in p and η , but using a pair of any two variables from p , p_T and η is obviously sufficient.

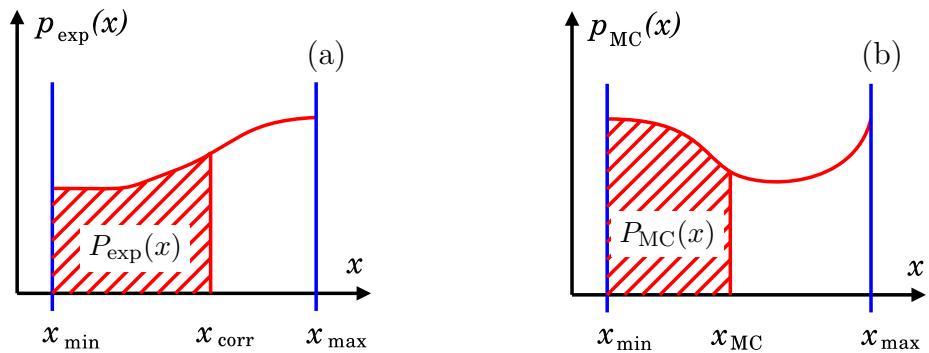


Figure 110: Illustrative graphs showing cumulative distribution function $P(x)$ for (a) experimental PDF $p_{\text{exp}}(x)$ and (b) simulation PDF $p_{\text{MC}}(x)$ of the PID variables x . The transformation of variables $x_{\text{MC}} \rightarrow x_{\text{corr}}$ is performed by assigning $P_{\text{exp}}(x_{\text{corr}}) = P_{\text{MC}}(x_{\text{MC}})$.

- Correlations of the PID variable with other properties of the event which are not accounted for by the resampling procedure are preserved via the correlations in simulation. Potentially this should provide better reproduction of PID response.
- Correlations between different PID variables for the same track are also preserved via correlations with simulated variables. As a result, the distributions of any function of PID responses for the same track can also be reproduced (although, less precisely than the “primary” variables), which allows one, in particular, to use the simulated samples after PID correction to train multivariate classifiers.
- Small variations of the calibration PDFs (for instance, in the evaluation of systematic uncertainties) will in turn result in small variations of the corrected PID variables, unlike in the case of resampling when the variables for alternative calibration PDF will be uncorrelated with the baseline ones. This helps in estimating systematic uncertainties due to PID correction.

2699 A.2 Unbinned description of calibration PDF

2700 The approach presented above implies the knowledge of the distribution of the PID variable
 2701 x as a function of p_T , η and N_{tr} , both for data, $p_{\text{exp}}(x|p_T, \eta, N_{\text{tr}})$, and for simulation,
 2702 $p_{\text{MC}}(x|p_T, \eta, N_{\text{tr}})$. Here we illustrate the approach using the variable that characterises the
 2703 probability of the track to be an electron, ProbNN_e, for the final-state electron/positron track.
 2704 The calibration data are taken typically from the sWeighted sample of $B^\pm \rightarrow J/\psi(e^+e^-)K^\pm$
 2705 decays used in PIDCalib. Descriptions of other PID variables can be obtained similarly.
 2706 Note that there are known issues related to the electron calibration samples; correlations
 2707 between the invariant mass and the track momentum and pseudorapidity as well as the
 2708 high level of background. Therefore, similar analyses have been using instead a fit and
 2709 count procedure, where fits to the calibration samples are performed in kinematic bins
 2710 before and after applying the PID cut to compute the efficiency. In addition, different

levels of backgrounds are observed depending on whether a bremsstrahlung photon has been associated to the electron track or not. As a consequence, these calibrations are further separated in the event this recovery tool has been used. This methodology can also be used in the context of an unbinned description of the calibration PDF, since in each slice of these kinematic bins a sWeighted distribution can be obtained, and thus, preserving the features required to this method.

The four-dimensional PDF $p_{\text{exp}}(\text{ProbNNp}|p_T, \eta, N_{\text{tr}})$ is obtained using kernel-based procedure with **Meerkat** package [66]. We perform the following transformation of variables to make the distributions that enter kernel estimation more uniform:

$$\text{ProbNNe}' = 1 - (1 - \text{ProbNNe})^\gamma, \quad p_T' = \ln p_T, \quad N_{\text{tr}}' = \ln N_{\text{tr}}, \quad (56)$$

where $\gamma = 1/2$. Note that the variable η is not transformed since it's already sufficiently uniform.

The kernel-based technique implemented in **Meerkat** is a modification of the common kernel density estimation approach, where the kernel-based estimator is a multiplicative correction over the other, approximation PDF. As a result, this technique allows for efficient description when the approximate PDF is known. In particular, it is suitable in the case of multidimensional PDF description when the PDF is approximately factorisable. In that case, the product of PDFs in lower dimensions can be used as the approximation PDF, and the kernel-based correction with relatively wide kernel provides an estimator of the PDF with low statistical fluctuations.

In the case of describing the four-dimensional PID PDF, the procedure used is to increase the dimensionality of the PDF in steps. These are illustrated in Fig. 111.

- 1D PDFs for each variable are estimated first (Fig. 111(a–d)).
- The pair of the most correlated variables, ProbNNe and p_T , is taken (see sWeighted binned distribution for these two variables in Fig. 111(e) for the kernel-based estimation of their 2D PDF). The approximation PDF is taken to be the product of two 1D PDFs, the resulting kernel PDF (Fig. 111(g)) is obtained with fixed kernel width, and in the second iteration the PDF with adaptive kernel width is obtained (Fig. 111(h)) where the width of the kernel is proportional to the square root of the PDF density in the point calculated at the previous step.
- Similarly, the third variable (η) is added to the PDF using the product of 2D (ProbNNe, p_T) and 1D η PDFs as approximation.
- Finally, N_{tr} is added similarly to form a 4D PDF. The slices of the resulting 4D PDF are shown in Fig. 111(i–l).

To check the quality of the resulting PDF, we split the space of $(p_T, \eta, N_{\text{tr}})$ variables into $3 \times 3 \times 3$ regions and compare the ProbNNe distributions in each of these regions for sWeighted calibration data and the result of PDF estimation for whether a bremsstrahlung photon has been associated to the electron track or not. These plots are shown in Figs. 112 and Figs. 113.

2749 Similar procedure is performed for pion and kaon tracks, for all ProbNN and CombDLL
2750 variables. The calibration samples are split by data taking year (2011 and 2012) and
2751 magnetic field polarity (up and down). All the ProbNN variables are transformed to make
2752 a more uniform distribution:

$$\text{ProbNNe}' = 1 - (1 - \text{ProbNNe})^\gamma \quad (57)$$

2753 for correct ID hypothesis (in which case the ProbNN variables are peaked at 1.0), and

$$\text{ProbNNe}' = \text{ProbNNe}^\gamma \quad (58)$$

2754 for wrong ID hypothesis (with ProbNN distribution peaked at zero). The factor γ is typi-
2755 cally chosen in the range (0.15 – 0.5). For CombDLL variables, no upfront transformation
2756 needs to be applied.

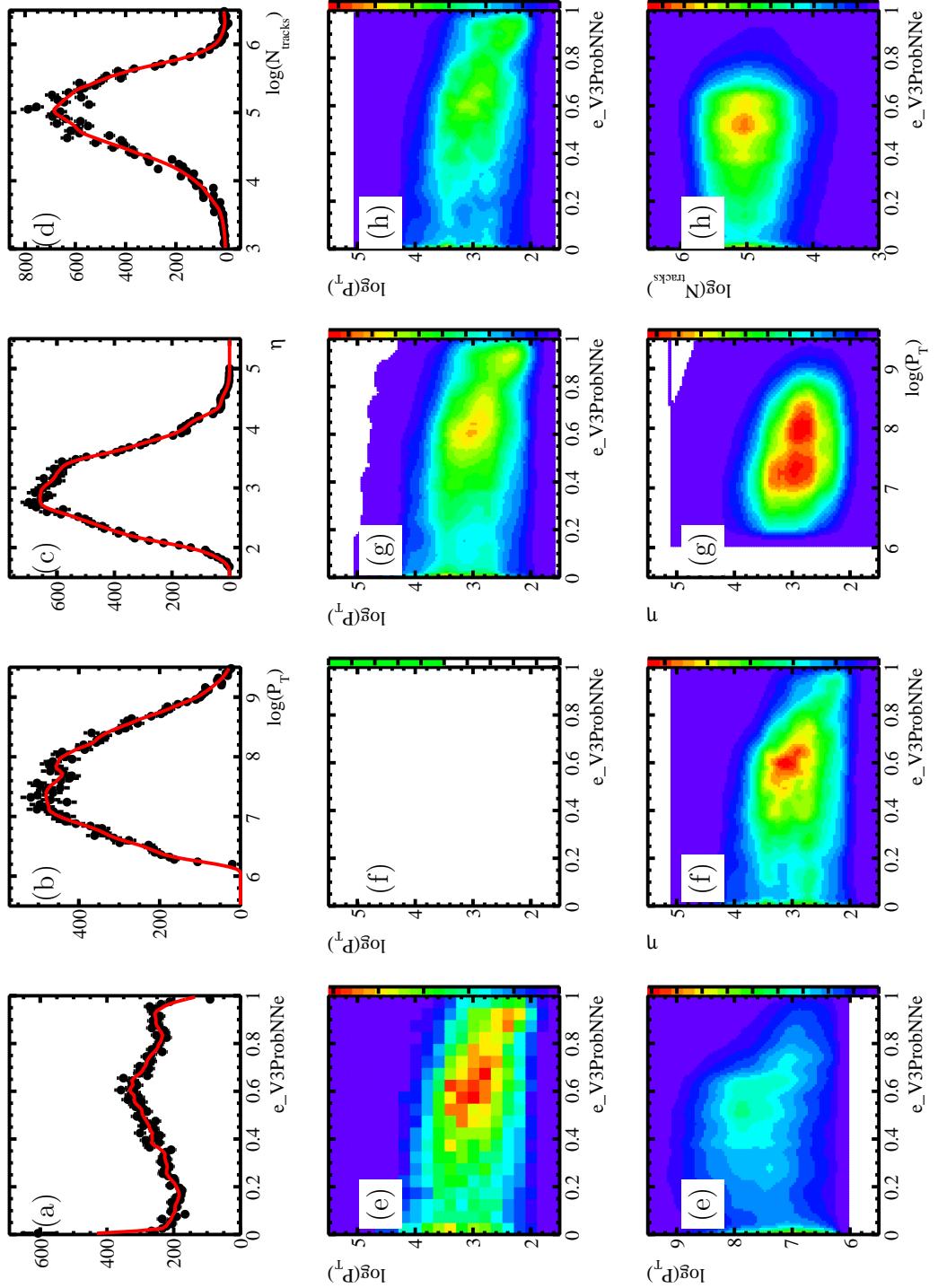


Figure 111: Illustration of the PID PDF estimation procedure. See text for details.

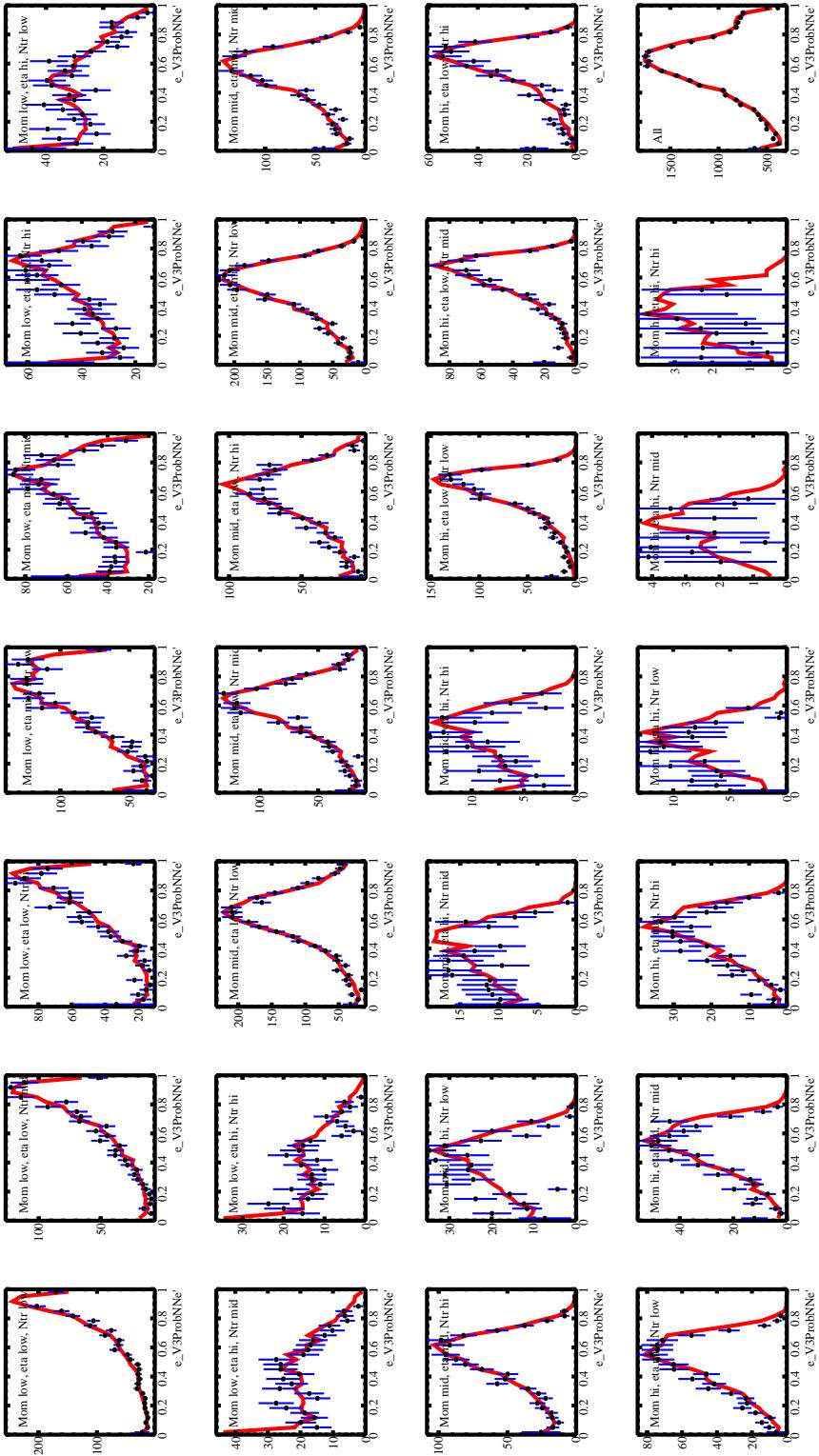


Figure 112: ProbNN' distributions in $3 \times 3 \times 3$ bins of p_T , η , N_{tracks} for electron with no bremsstrahlung photon associated with 2012 conditions. Comparison of calibration data (points with the error bars) and result of PDF estimation (solid red line).

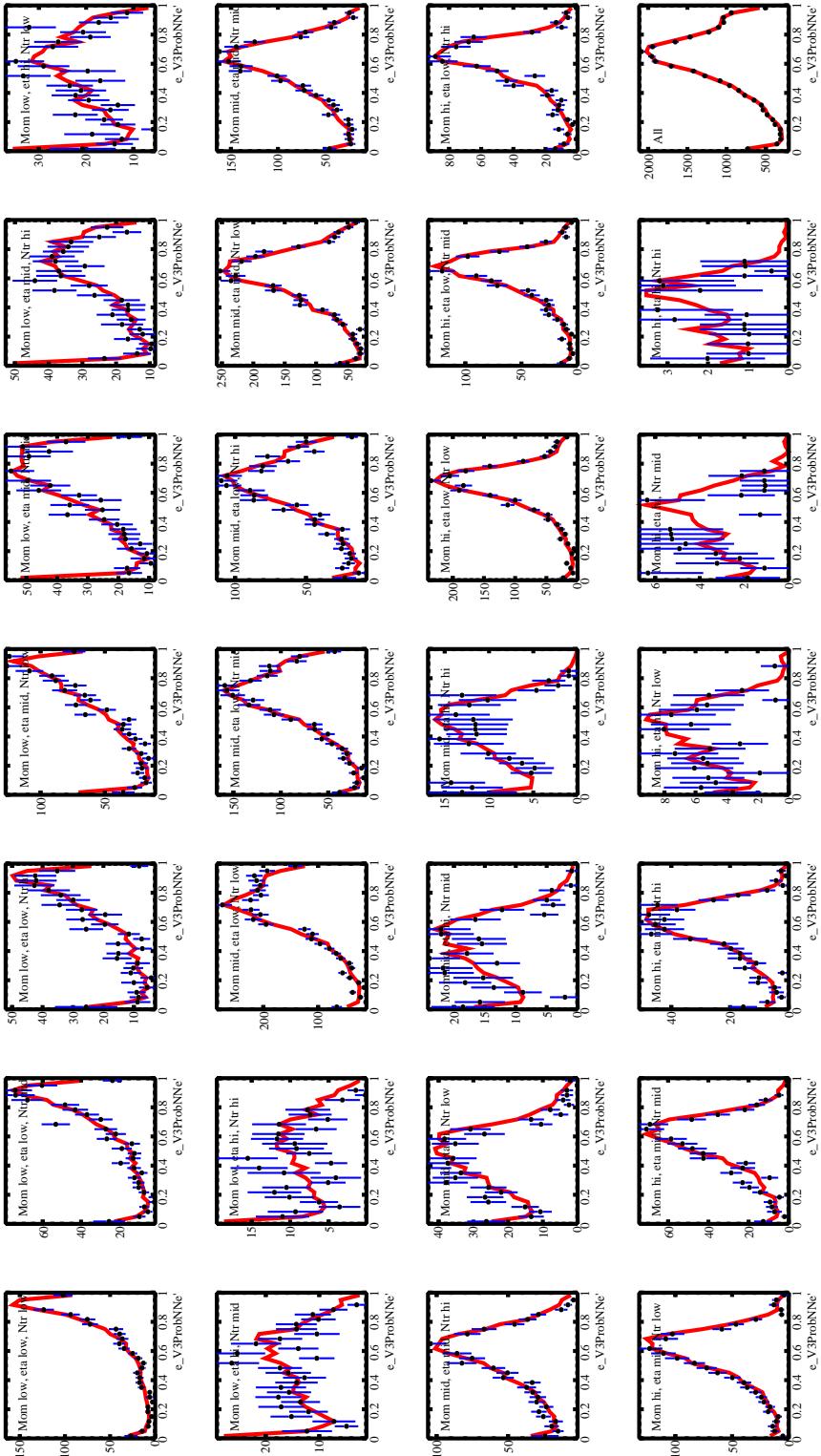


Figure 113: ProbNN' distributions in $3 \times 3 \times 3$ bins of p_T , η , N_{tracks} for electron with at least one bremsstrahlung photon associated with 2012 conditions. Comparison of calibration data (points with the error bars) and result of PDF estimation (solid red line).

Using only the PDFs from experimental calibration samples one can perform the “traditional” resampling of random variables in an unbinned way. For the approach with transformation of variables, however, one has to describe in addition the PDF of the simulated PID response. This is performed in a similar way, using the same stripping lines as for `PIDCalib` selections and the simulated samples corresponding to the decays used in calibration.

2763 A.3 Implementation in `PIDCalib`

Both techniques of PID response correction described above, resampling and transformation of variables, are implemented in LHCb software, in the `PIDCalib` package within `Urania` project. The 4D distribution of variables ($\text{PID}', \ln p_{\text{T}}, \eta, \ln N_{\text{tr}}$) is described, where PID' is the transformed PID variable (Eqs. 57, 58). The ranges of kinematic and multiplicity variables are given by

$$\begin{aligned}\ln p_{\text{T}}[\text{MeV}/c] &\in (5.5, 9.5), \\ \eta &\in (1.5, 5.5), \\ \ln N_{\text{tr}} &\in (3.0, 6.5).\end{aligned}$$

The dimensionality of the PDF which includes the PID response is increased step-by-step. The width of the multidimensional kernel is increased at each step to control statistical fluctuations of the sample. There are two scale factors which allow one to control the level of detail of the PDF description depending on the size of the calibration sample: the overall kernel width scale factor S_0 , and the kernel width scale factor for PID variables S_{PID} . The kernel widths are obtained at each step as follows. The widths for 1D estimation of PDFs for each individual variable are taken to be

$$\begin{aligned}\sigma_{\text{PID}}^{(1D)} &= s_{\text{PID}}^{(1D)} S_0 S_{\text{PID}} \cdot (x'_{\max} - x'_{\min}), \\ \sigma_{p_{\text{T}}}^{(1D)} &= s_{p_{\text{T}}}^{(1D)} S_0, \\ \sigma_{\eta}^{(1D)} &= s_{\eta}^{(1D)} S_0, \\ \sigma_{N_{\text{tr}}}^{(1D)} &= s_{N_{\text{tr}}}^{(1D)} S_0.\end{aligned}\tag{59}$$

The 2D estimation of the PDF in variables (x', p') with the product of 1D PDFs in x' and p_{T} as the approximation PDF uses the following kernel widths:

$$\begin{aligned}\sigma_{\text{PID}}^{(2D)} &= s_{\text{PID}}^{(2D)} S_0^{1/2} S_{\text{PID}} \cdot (x'_{\max} - x'_{\min}), \\ \sigma_{p_{\text{T}}}^{(2D)} &= s_{p_{\text{T}}}^{(2D)} S_0^{1/2}.\end{aligned}\tag{60}$$

The 3D estimation of the PDF in variables (x', p', η) (with the approximation PDF being the product of 2D (x', p') PDF from the last step and 1D PDF of η) uses

$$\begin{aligned}\sigma_{\text{PID}}^{(3D)} &= s_{\text{PID}}^{(3D)} S_0^{1/3} S_{\text{PID}} \cdot (x'_{\max} - x'_{\min}), \\ \sigma_{p_{\text{T}}}^{(3D)} &= s_{p_{\text{T}}}^{(3D)} S_0^{1/3}, \\ \sigma_{\eta}^{(3D)} &= s_{\eta}^{(3D)} S_0^{1/3}.\end{aligned}\tag{61}$$

Table 53: Common factors for kernel widths

Constant	$s_{\text{PID}}^{(1D)}$	$s_{p_T}^{(1D)}$	$s_\eta^{(1D)}$	$s_{N_{\text{tr}}}^{(1D)}$	$s_{\text{PID}}^{(2D)}$	$s_{p_T}^{(2D)}$	
Value	0.01	0.1	0.1	0.1	0.025	0.25	
Constant	$s_{\text{PID}}^{(3D)}$	$s_{p_T}^{(3D)}$	$s_\eta^{(3D)}$	$s_{\text{PID}}^{(4D)}$	$s_{p_T}^{(4D)}$	$s_\eta^{(4D)}$	$s_{N_{\text{tr}}}^{(4D)}$
Value	0.05	0.5	0.5	0.1	1.0	1.0	1.0

Finally, the 4D estimation of the PDF in variables $(x', p', \eta, N'_{\text{tr}})$ (with the approximation PDF being the product of 3D (x', p', η) PDF from the last step and 1D PDF of N'_{tr}) has

$$\begin{aligned}
 \sigma_{\text{PID}}^{(4D)} &= s_{\text{PID}}^{(4D)} S_0^{1/4} S_{\text{PID}} \cdot (x'_{\max} - x'_{\min}), \\
 \sigma_{p_T}^{(4D)} &= s_{p_T}^{(4D)} S_0^{1/4}, \\
 \sigma_\eta^{(4D)} &= s_\eta^{(4D)} S_0^{1/4}. \\
 \sigma_{N_{\text{tr}}}^{(4D)} &= s_{N_{\text{tr}}}^{(4D)} S_0^{1/4}.
 \end{aligned} \tag{62}$$

The variable x is the PID variable which is either a DLL-based (such as PIDK, PIDp) or a ProbNN-based (ProbNNpi, ProbNNK, etc.) variable. The “prime” in the variable definition corresponds to the upfront variable transformation as defined in Eq. (56) (for p_T , N_{tr}) and Eqs. (57) and (58) (for ProbNN-based PID variables). The upfront transformation is not applied to DLL-based PID variables. In the expressions above, the constants $s_X^{(nD)}$ are the same for all combinations of track ID and PID variable (see Table 53), while the ranges $x'_{\min, \max}$ and scale factors S_{PID} and S_0 depend on the PID response. The scale factors can also be different for data and simulation samples to account for the difference in size of these samples. For illustration, the variable ranges and scale factors for the PID responses used in the Run 1 analysis are given in Table 54 for both data and simulation.

A.4 Additional PID response projections

In addition to the illustrative validation plots for the agreement of the PID response provided in the text, it is relevant to examine the performance for different configurations. Figures 114, 115 and 116 show the agreement between the resampled simulation with respect to the sWeighted data for the resonant mode. Note that a reasonable agreement is seen in all distributions. Moreover, it is also possible to investigate the impact of the PID correction in the phase space of the event. Figures 117, 118, 119 and 120 show the ratio between corrected and uncorrected angular distributions for each year condition. Similarly to the inclusive run condition, the only non-trivial ratio is seen in the $\cos \theta_K$ distribution.

Table 54: Ranges and kernel width scale factors for various PID responses in Run 1

Track	PID variable	γ	x'_{\min}	x'_{\max}	Data		MC	
					S_0	S_{PID}	S_0	S_{PID}
π	PIDK	1	-100.	20.	0.6	1.	1.	1.
	PIDp	1	-120.	30.	0.6	1.	1.	1.
	V2ProbNNpi	-0.3	0.	1.	0.6	1.	1.	1.
	V2ProbNNK	0.15	0.	1.	0.6	1.	1.	1.
	V2ProbNNp	0.15	0.	1.	0.6	1.	1.	1.
K	PIDK	1	-20.	100.	0.6	1.	1.	1.
	PIDp	1	-100.	70.	0.6	1.	1.	1.
	V2ProbNNpi	0.2	0.	1.	0.6	1.	1.	1.
	V2ProbNNK	-0.15	0.	1.	0.6	1.	1.	1.
	V2ProbNNp	0.15	0.	1.	0.6	1.	1.	1.
e	PIDe	1	-20.	20.	0.6	1.	1.	1.
	V3ProbNNe	-0.25	0.	1.	0.6	1.	1.	1.
μ	PID μ	1	-10.	20.	0.6	1.	1.	1.
	V3ProbNN μ	-0.25	0.	1.	0.6	1.	1.	1.

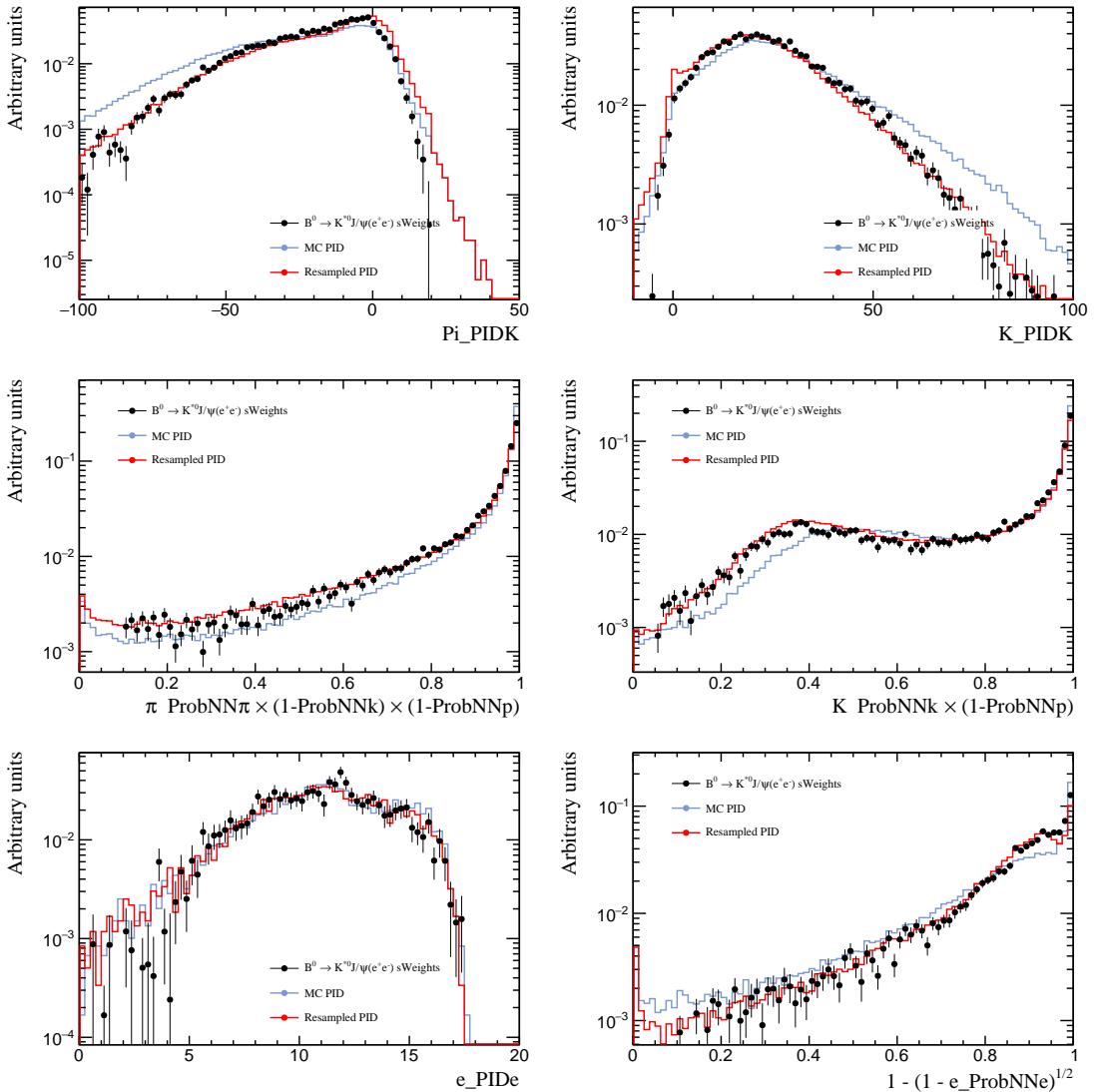


Figure 114: Particle identification distributions of the responses utilised in the analysis for sWeighted $B^0 \rightarrow K^{*0}J/\psi(e^+e^-)$ data (points with the error bars), uncorrected (blue solid line) and corrected (red solid line) Monte Carlo for the 2011 condition.

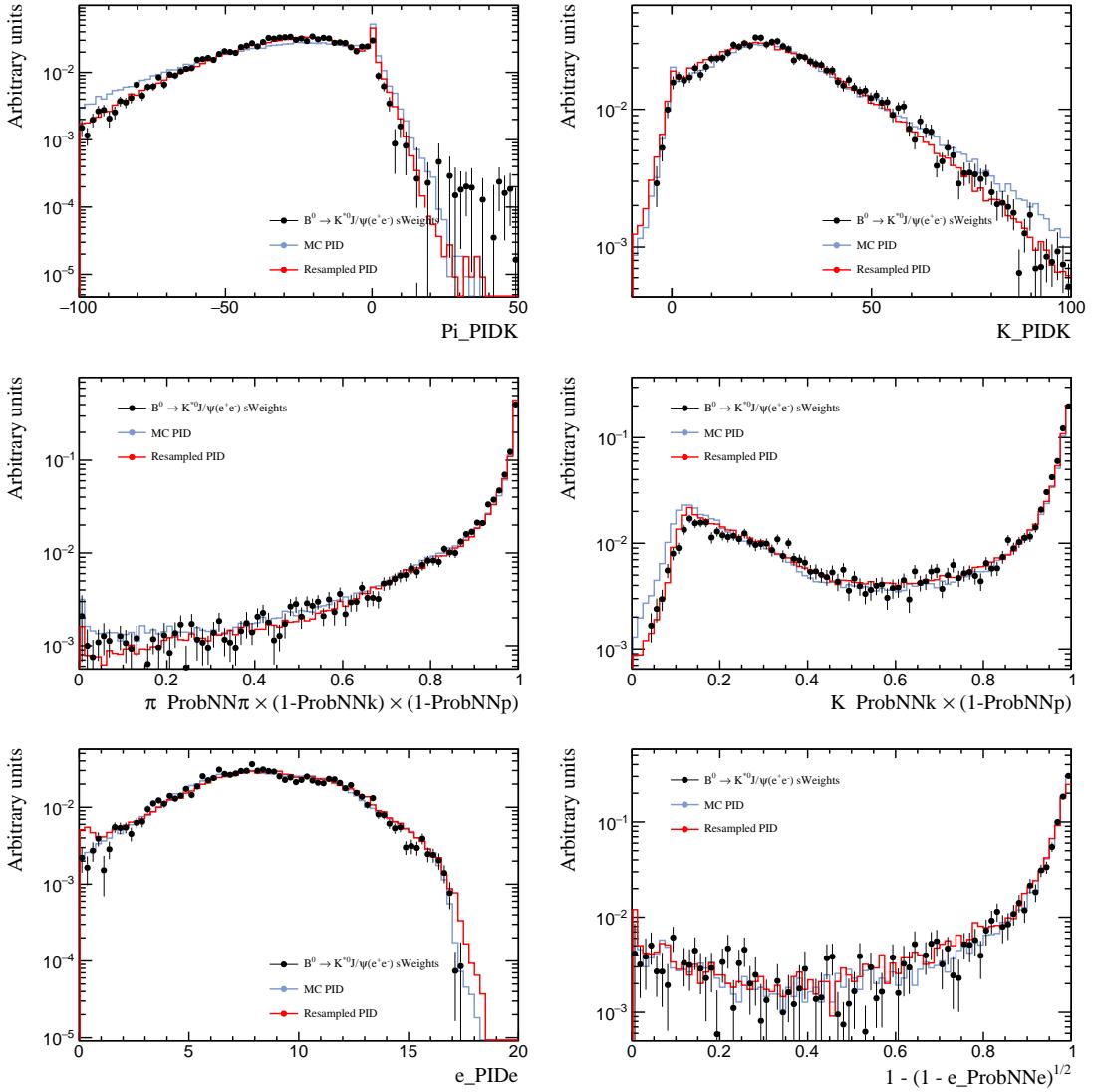


Figure 115: Particle identification distributions of the responses utilised in the analysis for sWeighted $B^0 \rightarrow K^{*0}J/\psi(e^+e^-)$ data (points with the error bars), uncorrected (blue solid line) and corrected (red solid line) Monte Carlo for the 2015 condition.

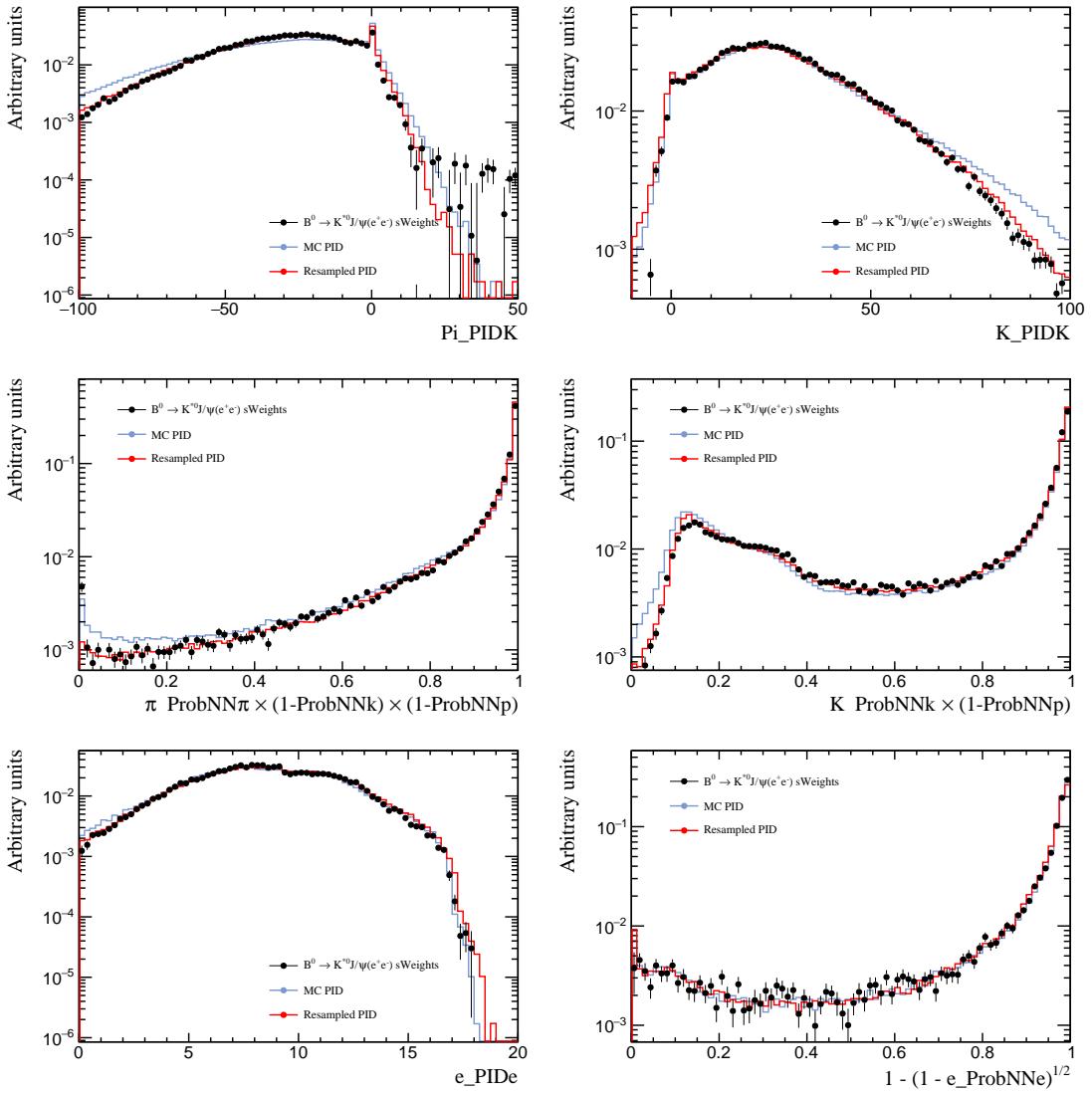


Figure 116: Particle identification distributions of the responses utilised in the analysis for sWeighted $B^0 \rightarrow K^{*0}J/\psi(e^+e^-)$ data (points with the error bars), uncorrected (blue solid line) and corrected (red solid line) Monte Carlo for the 2016 condition.

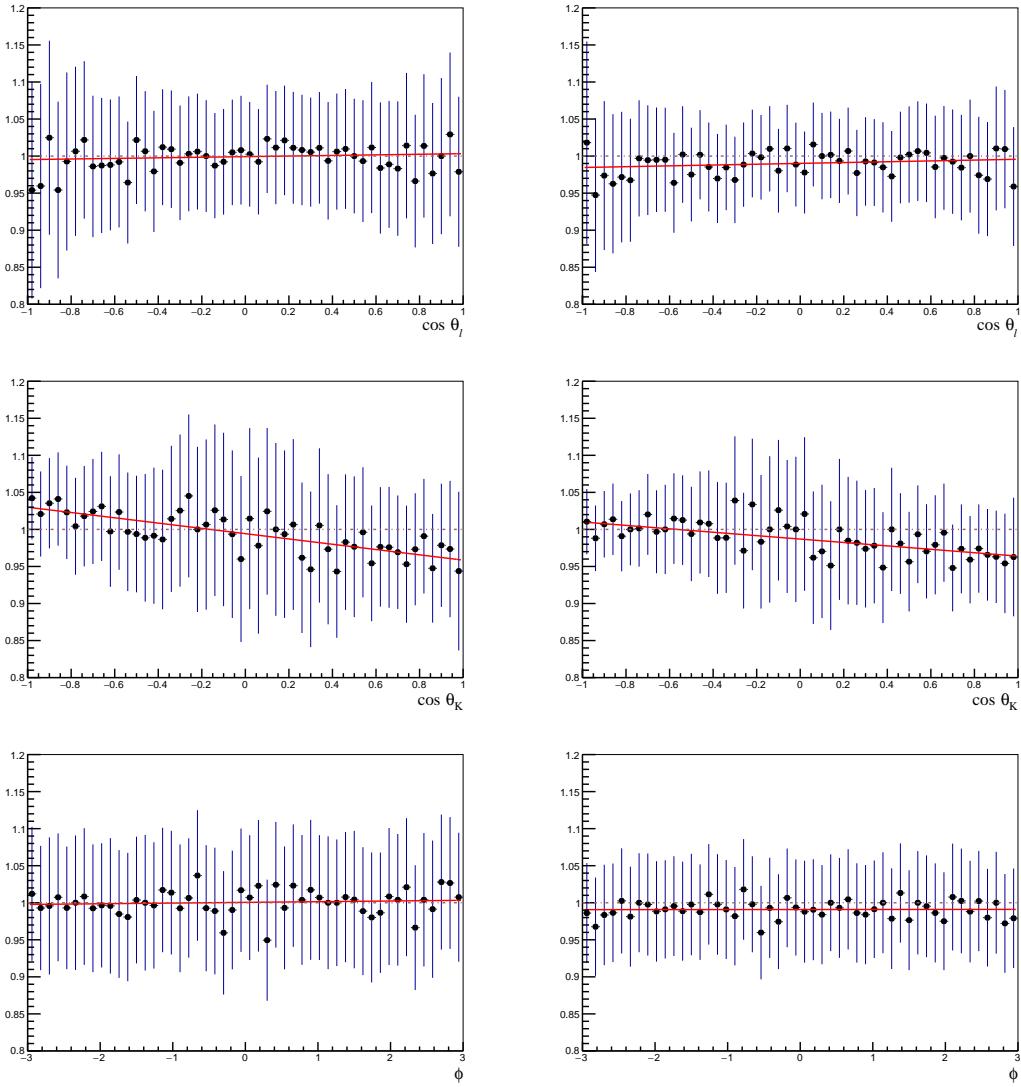


Figure 117: Ratio of the angular distributions between the PID corrected simulation and uncorrected ensembles for the truth-matched signal channel for the (left) 2011 and (right) 2012 conditions.

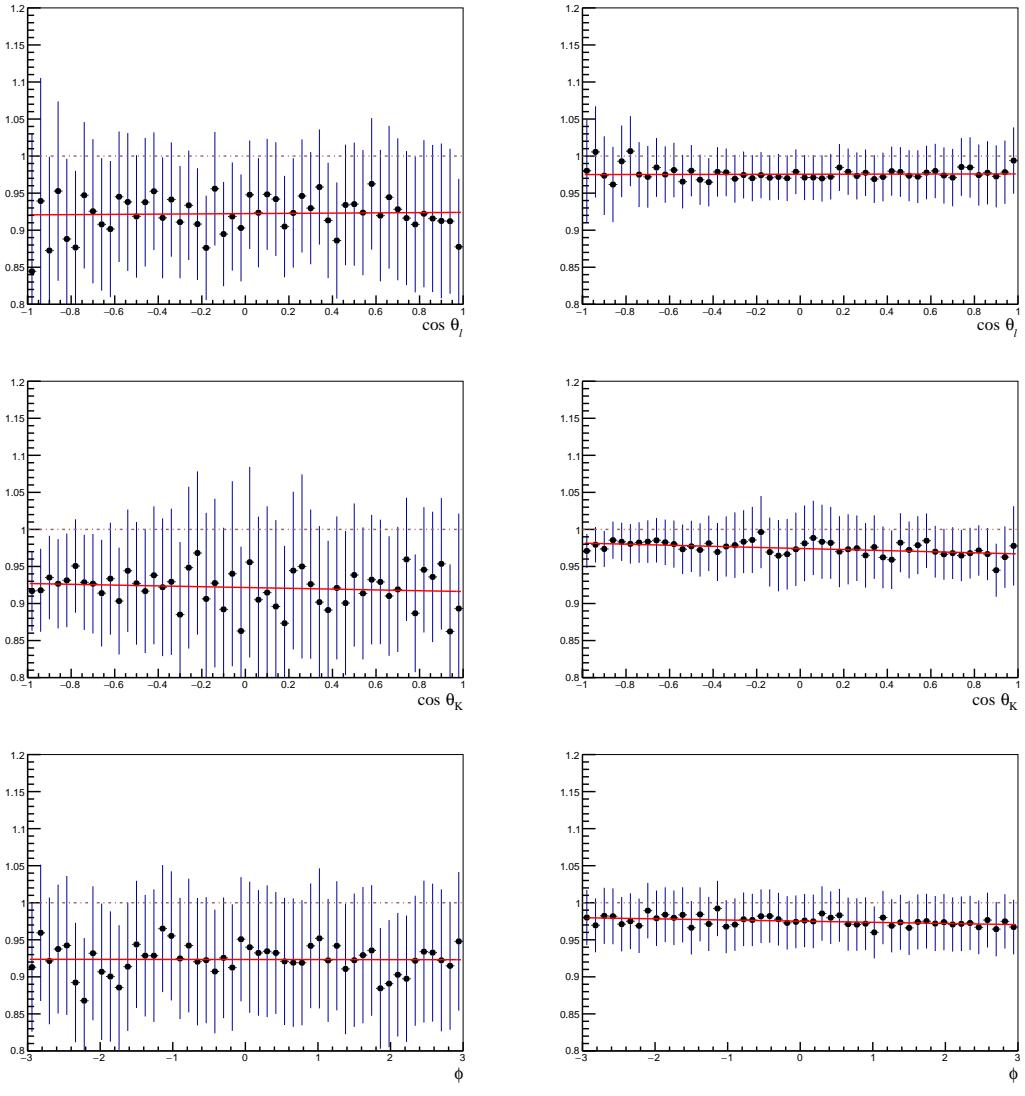


Figure 118: Ratio of the angular distributions between the PID corrected simulation and uncorrected ensembles for the truth-matched signal channel for the (left) 2015 and (right) 2016 conditions.

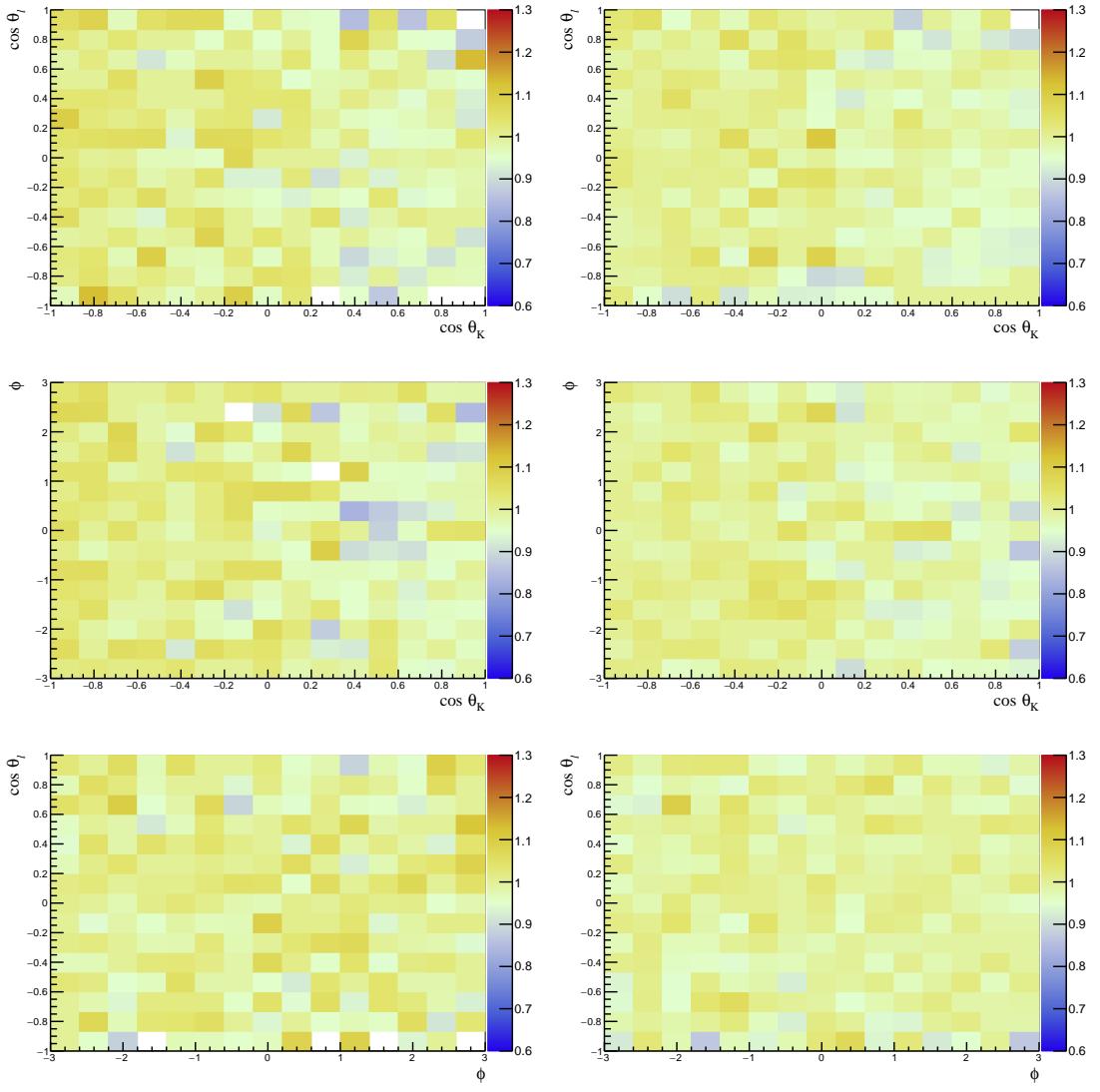


Figure 119: Two-dimensional ratio of the angular distributions between the PID corrected simulation and uncorrected ensembles for the truth-matched signal channel for the (left) 2011 and (right) 2012 conditions.

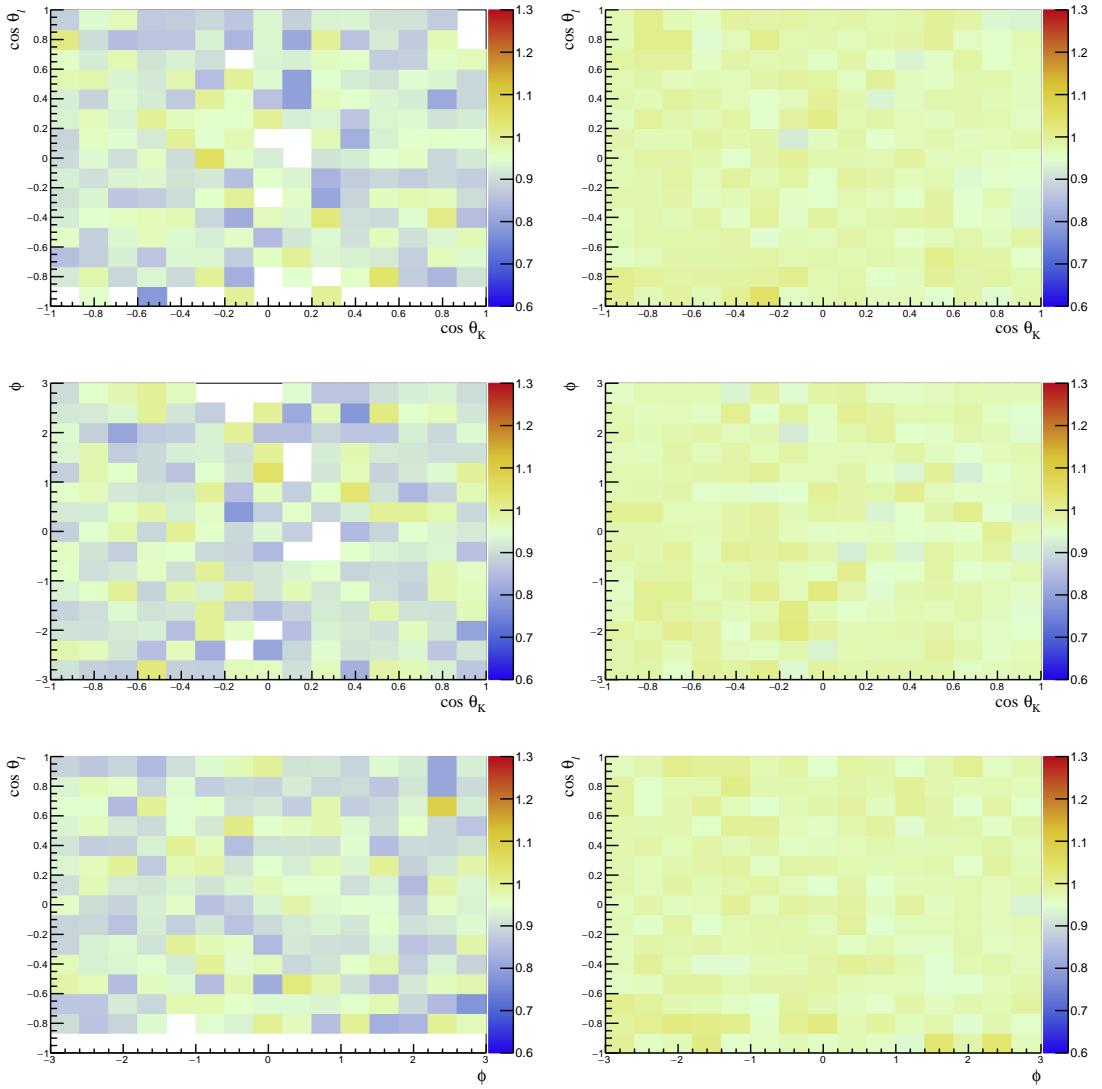


Figure 120: Two-dimensional ratio of the angular distributions between the PID corrected simulation and uncorrected ensembles for the truth-matched signal channel for the (left) 2015 and (right) 2016 conditions.

2802 B BDT re-weighting procedure

2803 B.1 Performance

2804 To find the optimal re-weighting hyper-parameters and to be able to compare the different
2805 approaches, a metric for the re-weighting quality should be established. Unfortunately,
2806 the comparison of two multi-dimensional distributions is not so simple. In contrast to one
2807 dimension, multidimensional distributions have no access to the order of each event. These
2808 are in general used in non-parametric tests such as Kolmogorov-Smirnoff and Anderson-
2809 Darling. Although certain approaches like density kernels exist for multidimensional
2810 distributions, these are not realistic for our case due to the lack of events and/or high
2811 dimensionality. On the other hand, the question that arises is not whether a certain
2812 statistics test can distinguish our samples, but if the MVA algorithm, which will be used
2813 in the MVA afterwards, can. Therefore it seems natural to rely on its predictions to find a
2814 reliable metric. In the following, three different approaches to investigate this problem are
2815 described.

2816 B.1.1 FoM₁ - simple discriminator

2817 A classifier is trained and tested on the re-weighted MC sample and on the real data
2818 using a stratified k-folding procedure and the variables that are used later in selection.
2819 To test the performance of the classifier, a single-valued metric is needed. Therefore, the
2820 receiver operation characteristic curve can be drawn and the area under the curve (AUC)
2821 calculated. Notice that the same metric is also used later in the MVA selection. Here the
2822 idea is that the lower this score is the less the classifier is able to discriminate the two
2823 distributions. Less discrimination power means that the two distributions are more similar
2824 under the assumption that an optimised MVA algorithm is used.

2825 Even though this approach yields a good idea of the similarity of the two samples, it
2826 does not prevent from overfitting. The problem arises with the event weights and the
2827 randomised training- respectively test-sample drawing. If an event a with an event weight
2828 w_a is drawn, what actually is drawn is not one event but w_a times the event a . This is then
2829 no longer randomised and uncorrelated samples, since drawing the event a implies also
2830 drawing the event a again, namely $w_a - 1$ times (for $w_a > 1$). Therefore, the paramount
2831 condition to make a statement, namely the randomised splitting, is no longer correct.
2832 The effect from this sample biasing is that the classifier makes incorrect predictions with
2833 wrongly gained strong confidence, which in turn lowers the ROC AUC more than we
2834 expect it to be. Although the effect decreases for large samples and is not expected to
2835 be too large for our case, it *can* even lead to ROC AUC values well below the 0.5 mark,
2836 which is usually assumed to be the lowest possible score.

2837 B.1.2 FoM₂ - data labelling

2838 Another metric can be obtained by training the classifier on the original MC sample
2839 without corrections as well as on the real data. This trained algorithm can be used to make

2840 predictions on three distinct samples and can be used to get hints for possible overfitting.
2841 The number of events that are predicted as real data from the following events are counted
2842 and interpreted as

- 2843 • MC: The lowest count is expected as most of the events will be predicted as MC;
2844 • re-weighted MC: a count as high as possible (but not higher than the real) is aimed
2845 for, as higher values mean more events in the sample look like a real event to the
2846 classifier;
2847 • real data: the highest count is expected.

2848 Ideally, the count of the re-weighted MC sample lies between the other two counts as
2849 close as possible to the real data. However, this score system should be used only as a
2850 guideline indication, since it only provides information about single events and not the
2851 distribution itself. Therefore, a real-like MC event with an extra large weight will wrongly
2852 dominate the score.

2853 **B.1.3 FoM_3 - data counting**

2854 In order to compare the distribution and not single events, a simple but robust approach is
2855 to train a classifier to discriminate between generated and real data. This is similar to the
2856 FoM_1 , but instead of making predictions on both the generated and the real sample, only
2857 the latter is considered. From this one can calculate the number of real events predicted
2858 as real. The more the classifier is able to learn from the distributions, the more real events
2859 will be predict correctly.¹⁶ The basic goal in this procedure is to minimise the score.
2860 Compared to the FoM_1 metric, the bias due to weights is constant and originates only
2861 from the weights of the real sample. Therefore, changing the weights of the generated
2862 samples, as a re-weighting algorithm does, does not change the bias. Although this score
2863 does not offer information at the percent level of the optimisation, it is a good indication
2864 of overfitting and complementary to the other scores.

2865 **B.2 $sPlot$ fits for calibration samples**

2866 In order to obtain a pure sample of $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ decays to be used in the BDT
2867 re-weighting procedure, a similar parametrisation of the nominal mass fit is used. The PV-
2868 and J/ψ -constrained invariant mass of the B^0 candidates¹⁷ is required to be larger than
2869 5150 MeV/ c^2 . Only B^0 candidates with successful kinematic fit when the two constraints
2870 are taken into account are retained.

¹⁶It has to be noted here that predictions are just a cut on the classifier output. To use the output for our purpose, equalised class-weights are required. Furthermore, the classifier itself has to generate probability-like predictions which XGBoost does. This is not per se the case for most algorithms.

¹⁷This is defined as the invariant mass computed by constraining the B^0 candidates to originate from the PV and the J/ψ candidates to have an invariant mass corresponding to the nominal mass of the J/ψ [71]

An unbinned extended maximum likelihood fit is performed separately for each year condition. The signal is parametrised by the sum of two Crystal Balls which share the same mean and have different values for the widths and tail parameters. The values of f_{CB} , n_1 , n_2 , α_1 , and α_2 obtained from simulation are fixed in the fit to data, while the mean and the widths are allowed to shift and scale (common to both widths) with respect to the values obtained from simulation. Moreover, $B_s^0 \rightarrow K^{*0} J/\psi (\rightarrow \ell^+ \ell^-)$ decays are also considered in the model and are described by the same model of the B^0 signal, with a mass shift corresponding to the difference between the B^0 and B_s^0 nominal masses. Finally, an exponential shape is included for accommodating the combinatorial background. In summary, the yields of the three components of the model \mathcal{N}_{B^0} , $\mathcal{N}_{B_s^0}$, and $\mathcal{N}_{\text{comb bkg}}$ are left free to vary in the fit, together with a single shift and scale, and the slope of the exponential distribution. The fit results are shown in Fig. 121 for both simulation and data. Note that the residuals of the fit are reasonably good in the whole mass range.

B.3 BDT-re-weighting results

In order to obtain the optimal configuration for the re-weighting algorithm the scores described before as well as visual inspections of the feature distributions are used. The values obtained for the different parameters are shown in Table 55. The performance of the BDT re-weighting is comparable to the standard histogram re-weighting with 30 bins in the first step and 10 bins in the second, and is found to be superior to it. All variables used in the multivariate analysis, as well as the decay angles θ_ℓ , θ_K , and ϕ and other variables relevant for the analysis are taken into account to score the performance of the BDT reweighting. Finally, the scoring is computed by using a XGBoost classifier.

Table 55: Hyper-parameter configurations for the gradient boosted reweighting. Two separated values means the first one was used for the first reweighting stage and accordingly for the second one. A single value means the same parameter was used in both stages.

Parameter	Value	Explanation
n_estimators	12(7)	Number of boosting rounds to be performed. Different configuration are used for the first and second stage of the training, respectively.
learning_rate	0.1	A factor by which the weights of each boosting stage are multiplied by. There is a trade-off between the learning_rate and n_estimators and the ratio determines (basically) how complex our model is.
max_depth	3	Maximum depth of the DT. Higher values create more complex models and are able to get higher order correlations but tend to over-fit.
min_samples_leaf	20	Determines the minimum number of events in a leaf in order to split. Larger values create more conservative models and can help to avoid overfitting.
loss_regularization	10	Adds a regularisation term to the weight inside the logarithm of the loss-function.
gb_args: subsample	0.8	The fraction of the data that is used to train each DT. Reduces overfitting.

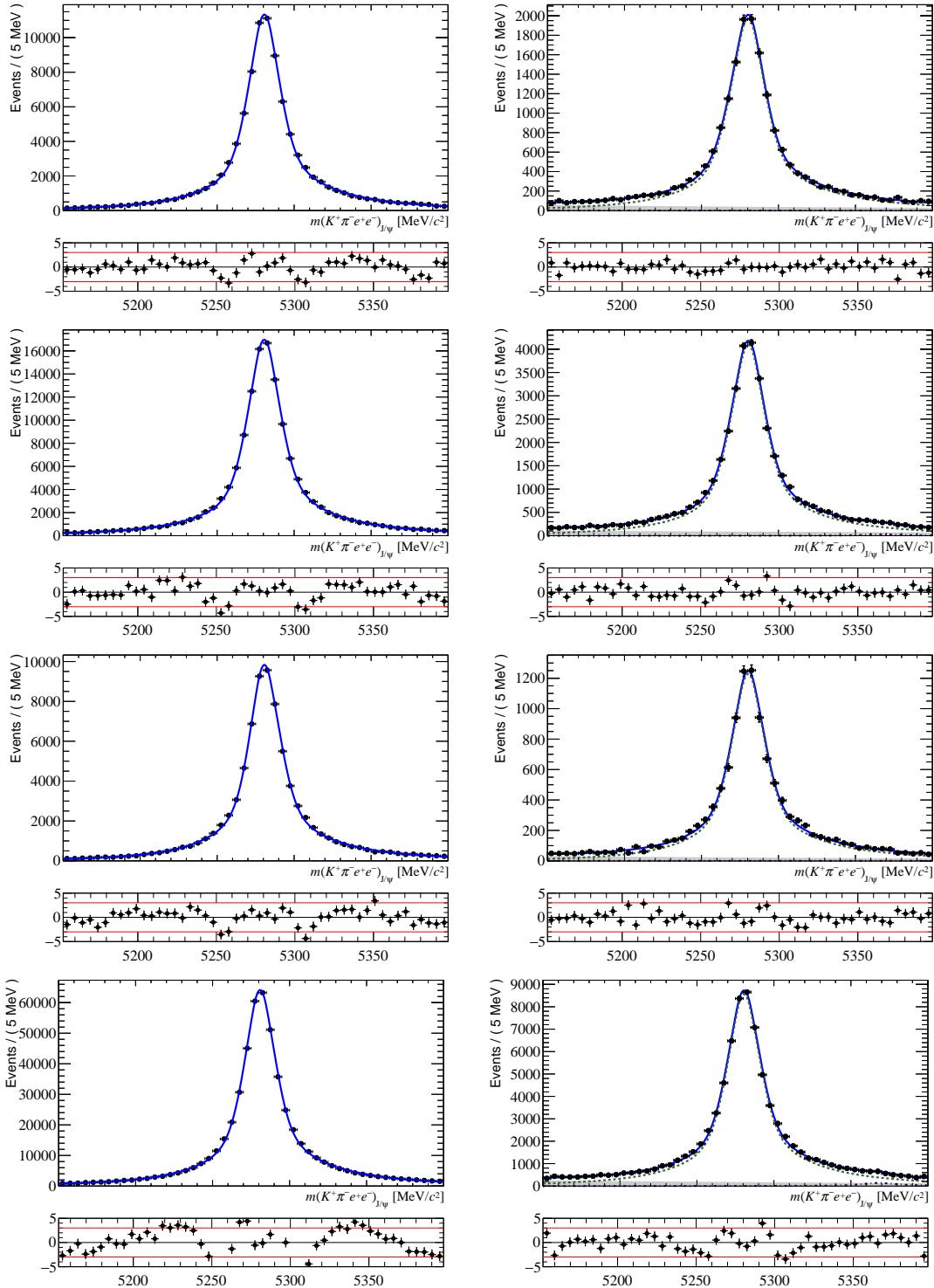


Figure 121: Fits to the mass distribution of $B^0 \rightarrow K^{*0} J/\psi \ (\rightarrow e^+e^-)$ decays for (left) simulation and (right) data candidates for (top) 2011, (second row) 2012, (third row) 2015 and (bottom) 2016. The residuals are also shown for completeness.

2893 **C Trigger correction**

2894 In the following more details on the impact of the trigger weights on the simulation is
2895 discussed.

2896 **C.1 The effect of the L0 weights**

2897 Figure 122 shows the distribution of the L0 trigger weights for the years 2011, 2012, 2015
2898 and 2016. The sample considered corresponds to simulated events of $B^0 \rightarrow K^{0*} e^+ e^-$ after
2899 the full preselection is applied and truth-matched. The weights for LOI tend to be on
2900 average below unity, as already seen in the TIS tables, suggesting that the efficiency for
2901 events triggered independently of the signal is larger in simulation than in data. The
2902 weights assigned for LOE have a more continuous distribution and are usually centred
2903 around unity, with few events that can reach 1.5 (2011) or 0.5 (2016). These last ones in
2904 particular can be greatly reduced applying the fiducial cut in E_T discussed in 3.6.2.

2905 Figures 123, 124, 125 and 126 illustrate how the L0 trigger corrections and the BDT
2906 reweighting impact the angular distributions, *i.e.* $\cos \theta_K$, $\cos \theta_L$ and ϕ , as well as the B^0
2907 invariant mass, q^2 and $B^0 p_T$ distributions for all the years considered. The distributions
2908 do not change significantly by the use of the L0 weights, suggesting that the systematic
2909 linked to this procedure should be negligible for most of the angular observables. One
2910 exception is the $\cos \theta_K$ distribution, where the ratio seems to grow towards one on the
2911 right; this issue is studied in the Sec. 10.

2912 **C.2 Compatibility between LOE and LOE HLT corrections**

2913 Figure 127 shows the comparison between inclusive and exclusive LOElectron_TOS for the
2914 HLT corrections.

2915 **C.3 HLT weighted samples**

2916 Figure 128 shows the distribution of the combination of the HLT trigger weights for
2917 the years 2011, 2012, 2015 and 2016. The sample corresponds to simulated events of
2918 $B^0 \rightarrow K^{0*} e^+ e^-$ after the full preselection is applied and truth-matched.

2919 Figure 129, 130, 131 and 132 show how the L0, HLT trigger corrections and BDT
2920 reweighting impact the angular distributions, *i.e.* $\cos \theta_K$, $\cos \theta_L$ and ϕ , as well as the B^0
2921 invariant mass, q^2 and $B^0 p_T$ distributions for all the years considered. No trend seems to
2922 be introduced in the distributions considered with the obvious exception of $B^0 p_T$.

2923 Table 56 (57) shows the variation of the number of events (relative size) of the L0
2924 trigger categories LOI and LOE when the correction weights are taken into account. Notice
2925 that such variation directly impacts the absolute efficiency of our selection.

	no weights		BDT		BDT and L0 weights		BDT, L0 and HLT weights	
	L0I	L0E	L0I	L0E	L0I	L0E	L0I	L0E
2011	5818	5535	6199.62	5625.08	5646.26	5547.06	4673.73	4972.37
2012	10003	8388	9616.07	7621.18	9471.77	7519.11	8110.98	7215.95
2015	4967	4427	4305.80	3555.13	4292.05	3360.20	4138.83	3345.96
2016	32911	28120	27082.40	21401.88	24168.33	21304.04	24759.91	21716.27

Table 56: Variation of the number of the simulated events once the BDT, L0 and HLT weights are taken into account.

	no weights		BDT		BDT and L0 weights		BDT, L0 and HLT weights	
	L0I	L0E	L0I	L0E	L0I	L0E	L0I	L0E
2011	0.51	0.49	0.55	0.50	0.50	0.49	0.41	0.44
2012	0.54	0.46	0.52	0.41	0.52	0.41	0.44	0.39
2015	0.53	0.47	0.46	0.38	0.46	0.36	0.44	0.36
2016	0.54	0.46	0.44	0.35	0.40	0.35	0.41	0.36

Table 57: Relative size of the L0 trigger categories L0I and L0E respect to the non-weighted sample when the BDT, L0 and HLT weights are taken into account.

C.4 Correlation between $p_T(B)$ and $\cos \theta_K$

As can be observed in Fig. 18 and Figs. 123, 124, 125, 126 there is an upward trend in the right region of $\cos \theta_K$ that has been caused by the reweighting in $p_T(B)$ due to the trigger corrections in L0I. This trend can be explained by the Stripping cut $\text{Pi_PT} > 250 \&& \text{K_PT} > 250$, that distorts the $\cos \theta_K$ distribution increasingly with the decrease of the $p_T(B)$, as shown in Fig. 133. As a consequence, by using a reweighting scheme that increases the importance of events with higher $p_T(B)$ as it is done for L0I, we are reducing the impact that the mentioned Stripping cut has on the integrated $\cos \theta_K$ distribution.

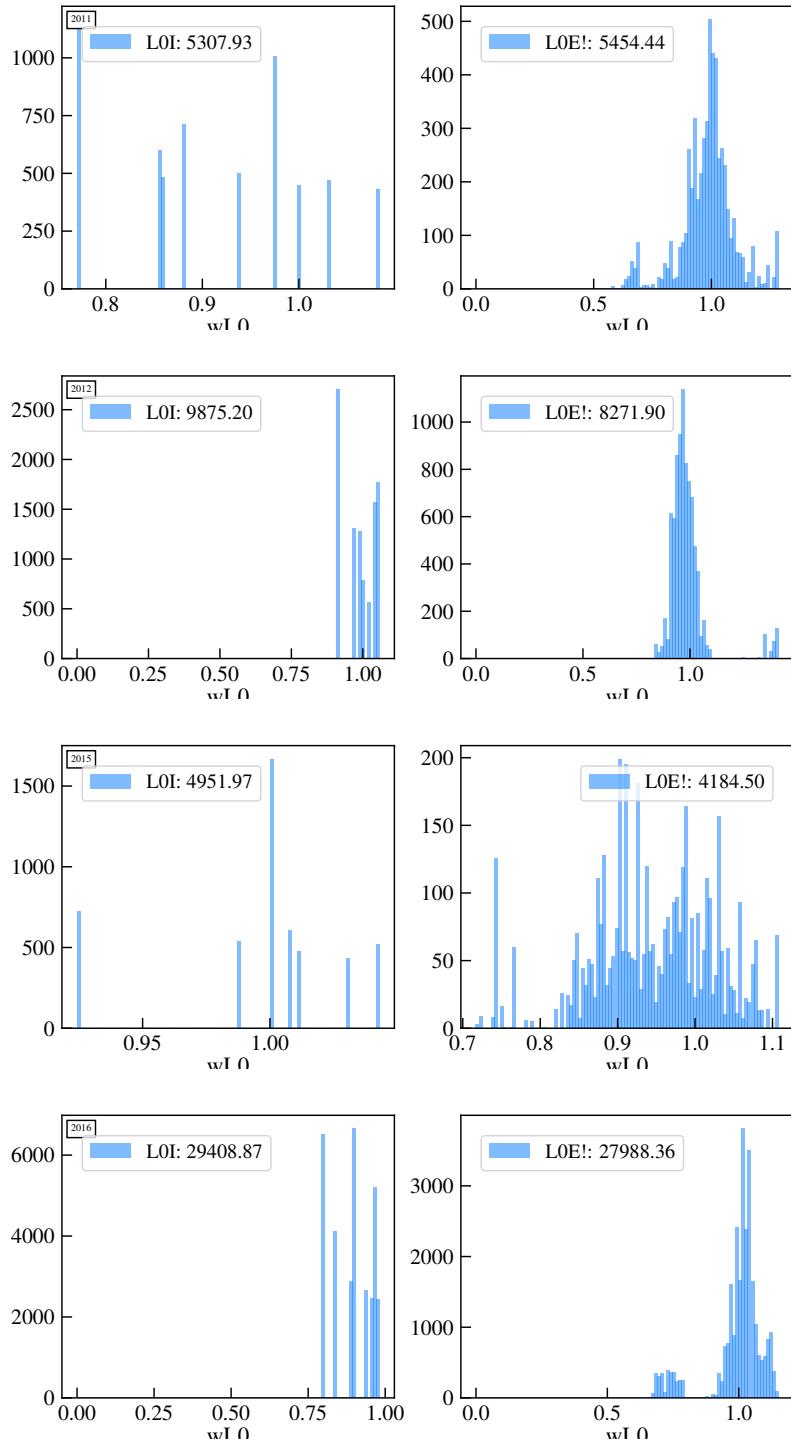


Figure 122: Distribution of the L0 weights for a simulated sample of $B^0 \rightarrow K^{*0} e^+ e^-$ for the year 2011, 2012, 2015 and 2016 in the L0 trigger categories L0I (left) and L0E (right). The numbers in the legends correspond to the sum of the L0 weights in each trigger category. No kinematic correction is considered here.

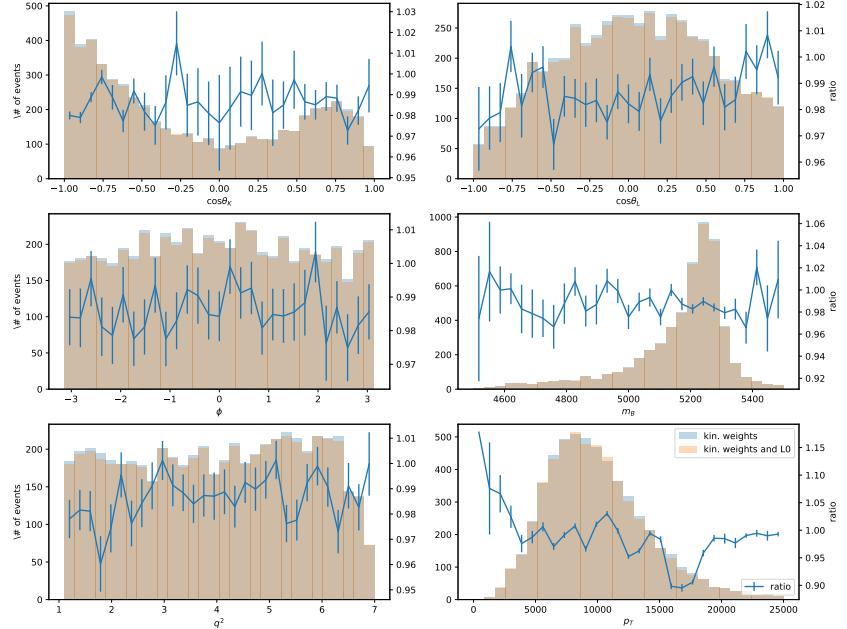
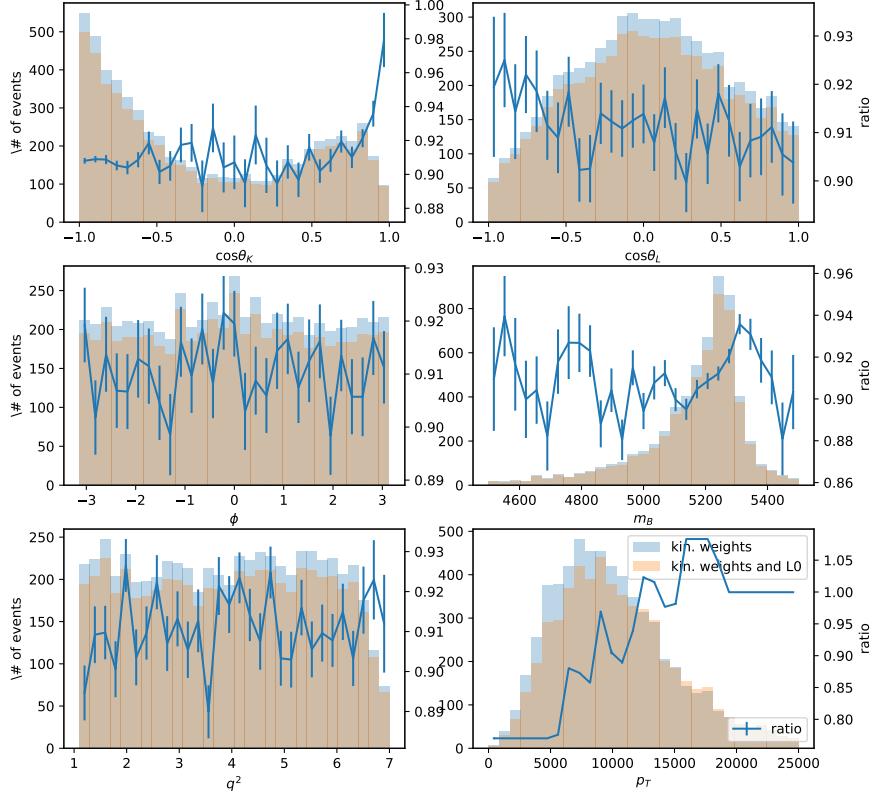


Figure 123: L0 and kinematically reweighted (blue) and only kinematically reweighted (orange) distributions of a simulated sample of $B^0 \rightarrow K^{0*} e^+ e^-$ for the year 2011 divide in the two L0 trigger categories L0I (top) and L0E (bottom). The blue points represents the ratio between the two histograms.

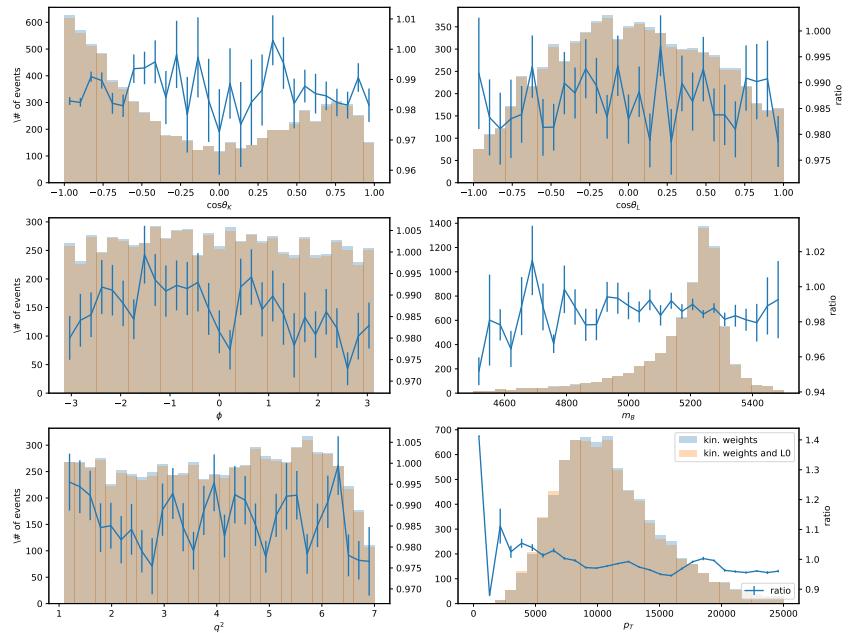
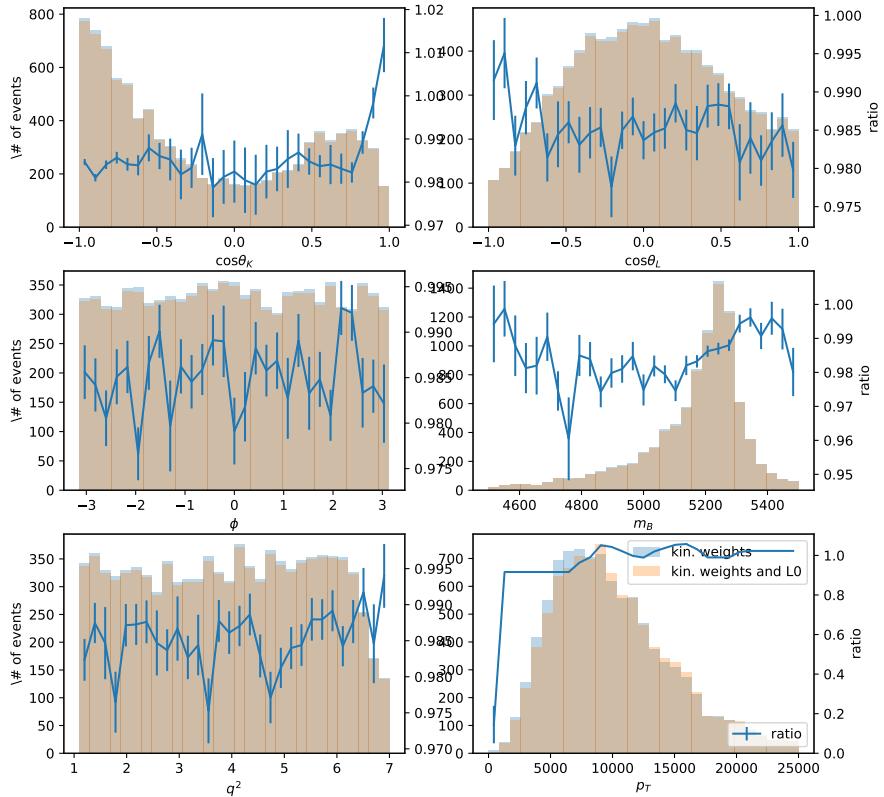


Figure 124: L0 and kinematically reweighted (blue) and only kinematically reweighted (orange) distributions of a simulated sample of $B^0 \rightarrow K^{0*} e^+ e^-$ for the year 2012 divide in the two L0 trigger categories L0I (top) and L0E (bottom). The blue points represents the ratio between the two histograms.

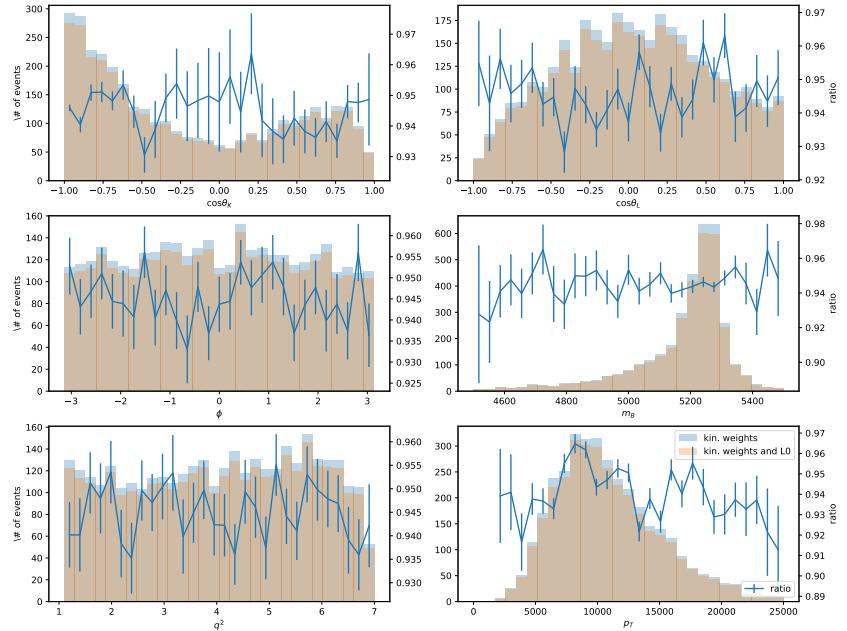
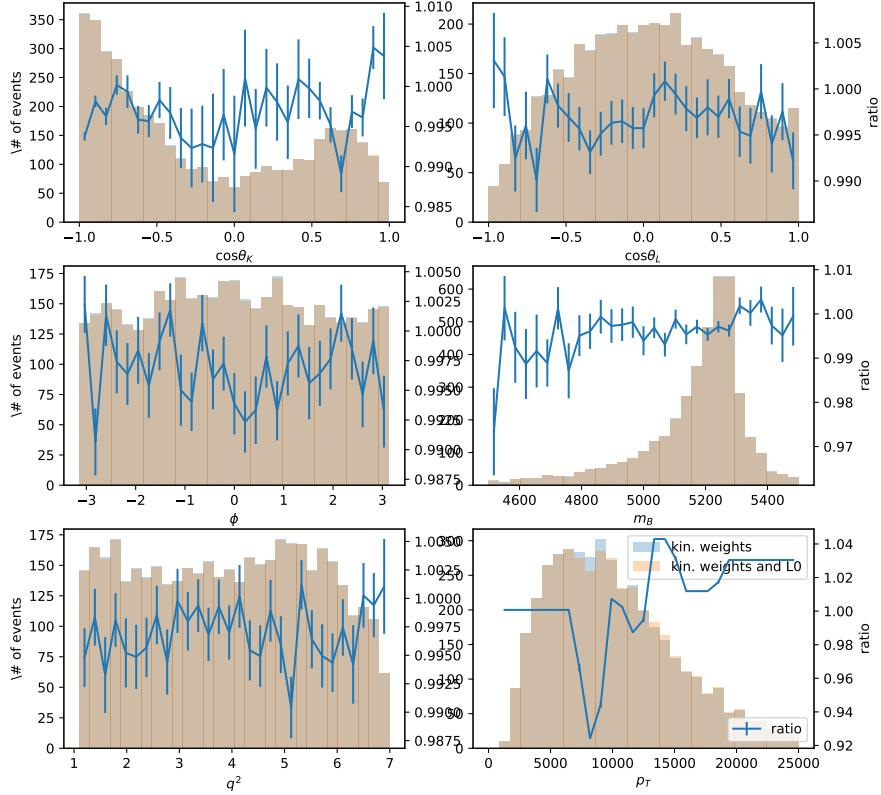


Figure 125: L0 and kinematically reweighted (blue) and only kinematically reweighted (orange) distributions of a simulated sample of $B^0 \rightarrow K^{0*} e^+ e^-$ for the year 2015 divide in the two L0 trigger categories L0I (top) and L0E (bottom). The blue points represents the ratio between the two histograms.

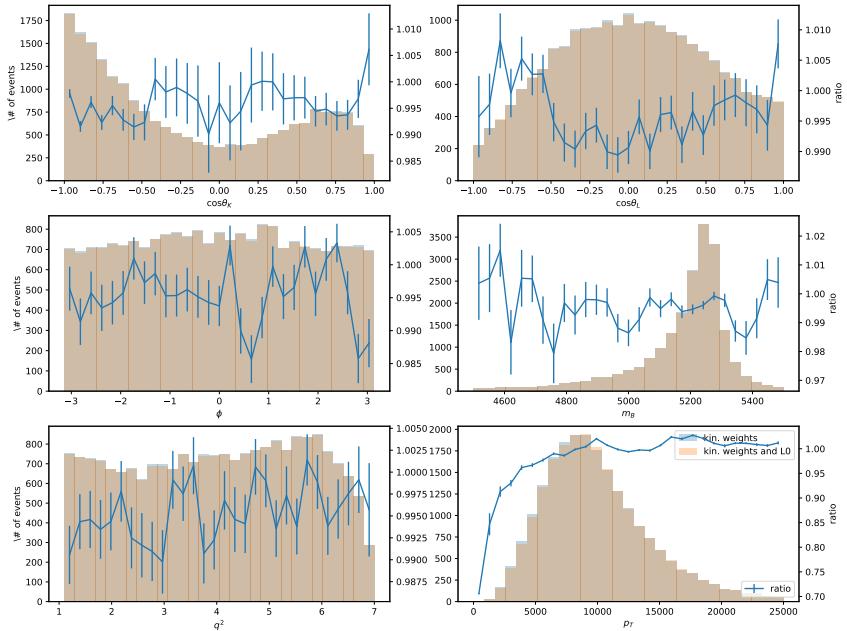
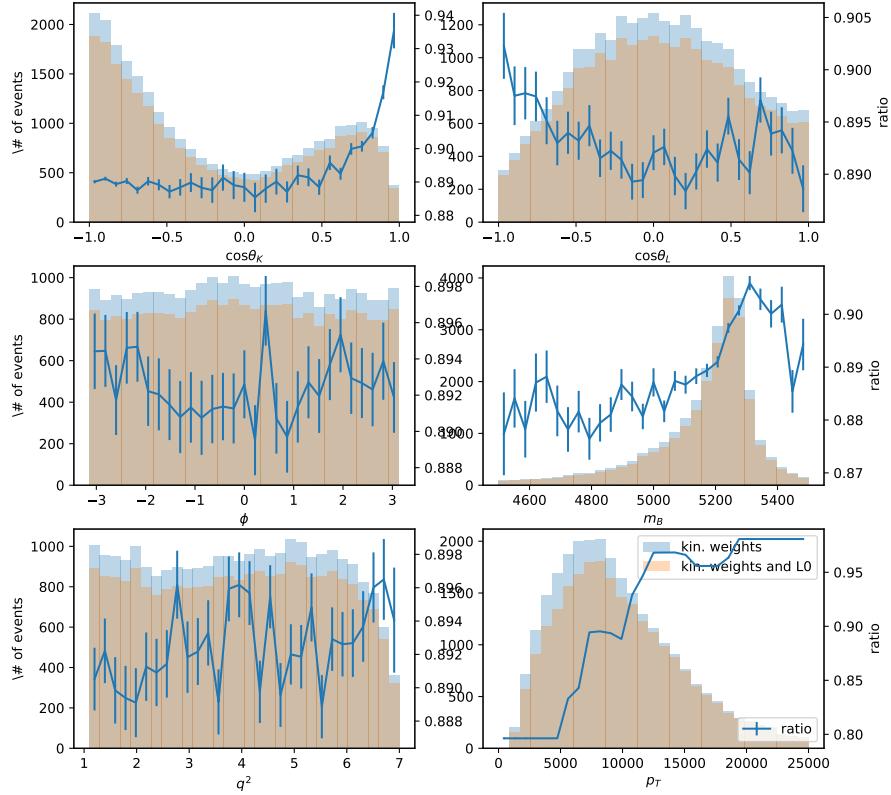


Figure 126: L0 and kinematically reweighted (blue) and only kinematically reweighted (orange) distributions of a simulated sample of $B^0 \rightarrow K^{0*} e^+ e^-$ for the year 2016 divide in the two L0 trigger categories L0I (top) and L0E (bottom). The blue points represents the ratio between the two histograms.

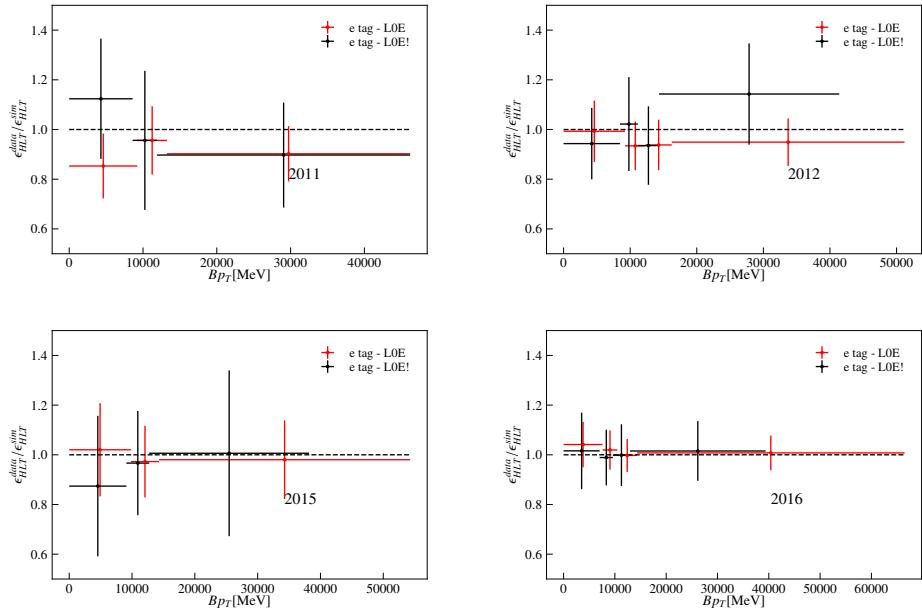


Figure 127: Data/MC ratio of the HLT tag-and-probe efficiency for $B^0 \rightarrow K^{0*} J/\psi (\rightarrow e^+e^-)$ as a function of the p_T of the B^0 for the trigger categories LOE and LOE!

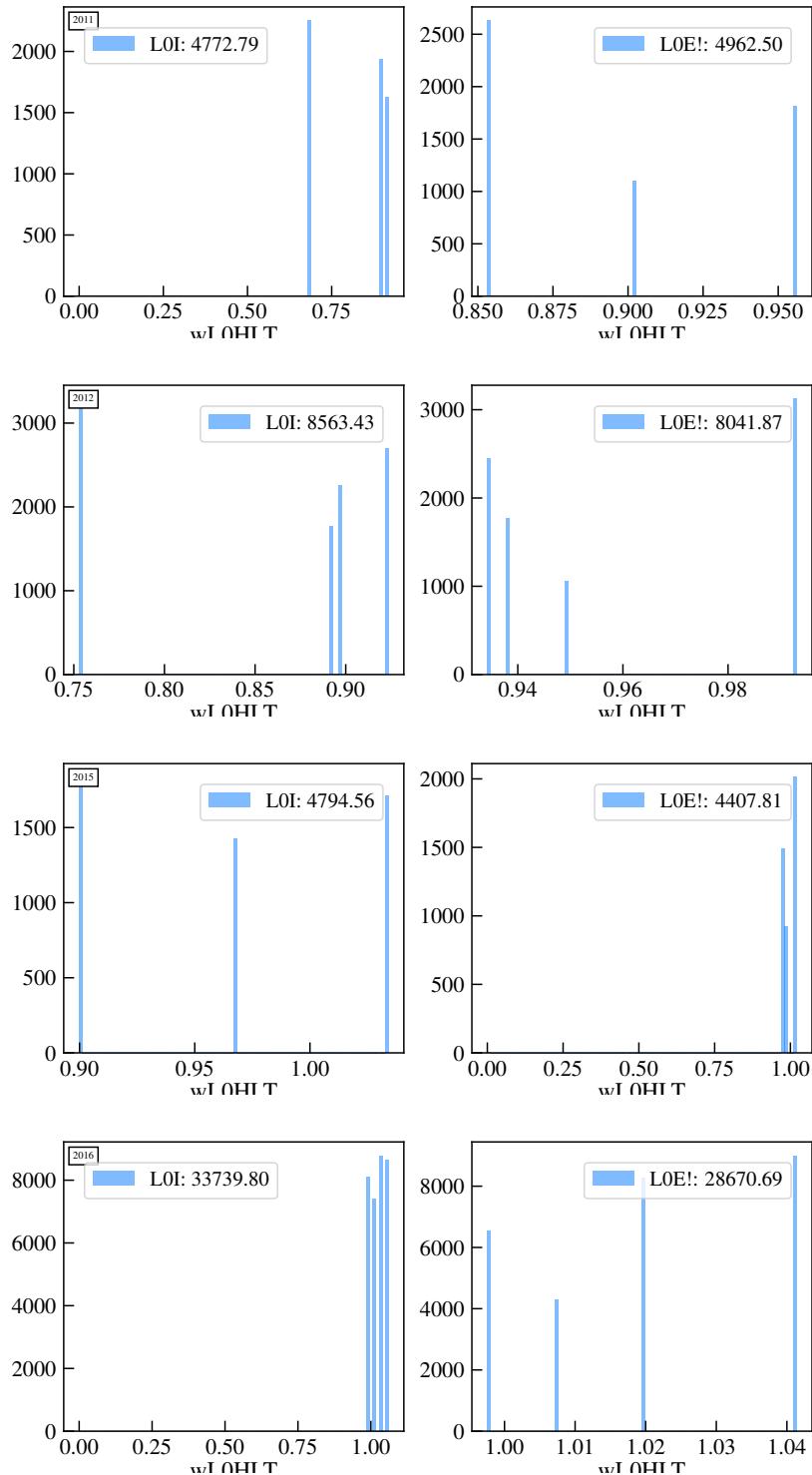


Figure 128: Distribution of the HLT weights for a simulated sample of $B^0 \rightarrow K^{0*} e^+ e^-$ for the year 2011, 2012, 2015 and 2016 in the L0 trigger categories L0I (left) and L0E (right). The numbers in the legends correspond to the sum of the HLT weights in each trigger category.

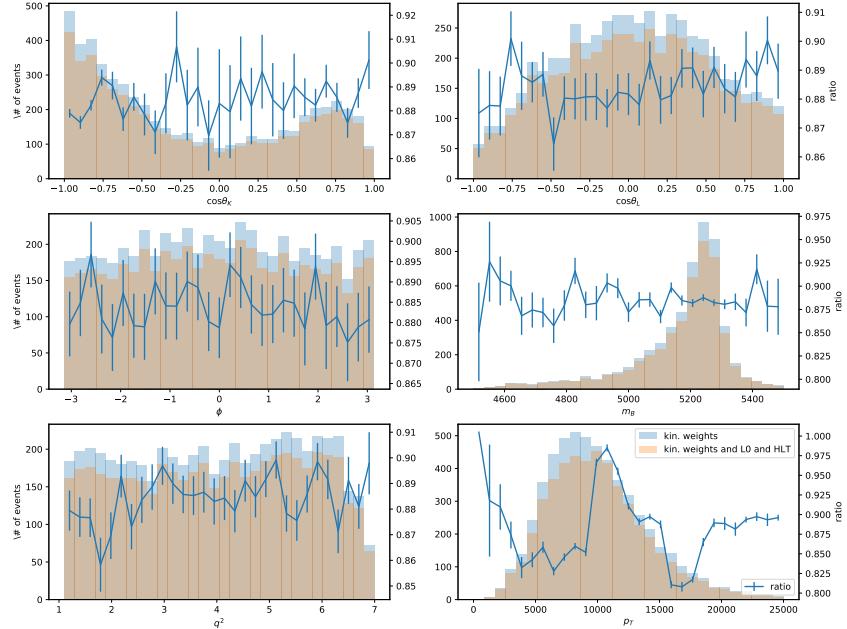
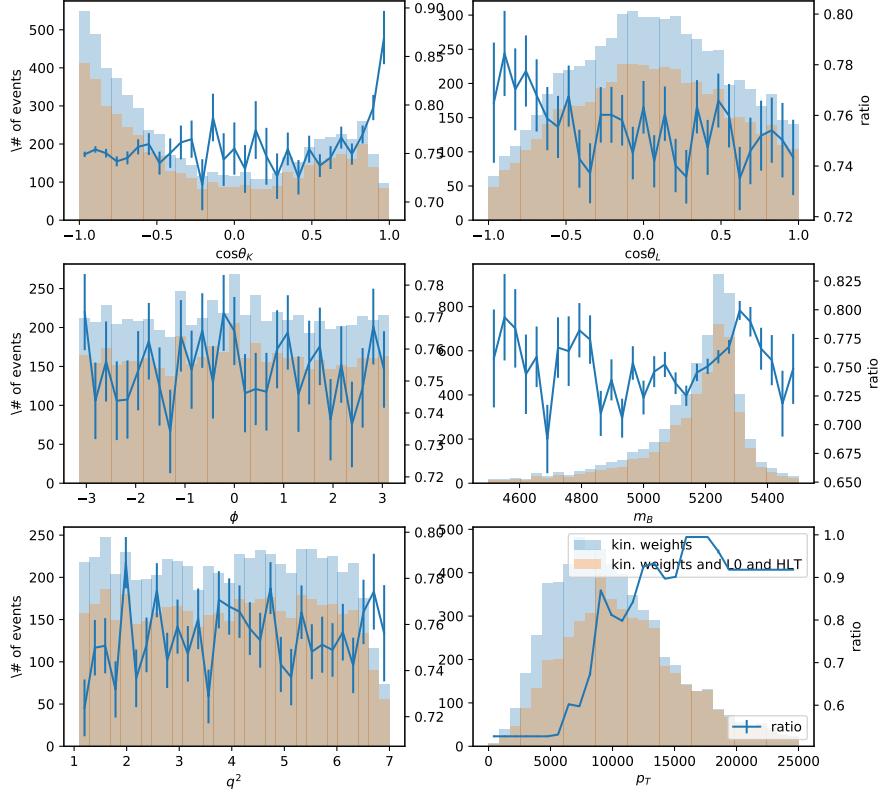


Figure 129: HLT, L0 and kinematically reweighted (blue) and only kinematically reweighted (orange) distributions of a simulated sample of $B^0 \rightarrow K^{0*} e^+ e^-$ for the year 2011 divide in the two L0 trigger categories L0I (top) and L0E (bottom). The blue points represents the ratio between the two histograms.

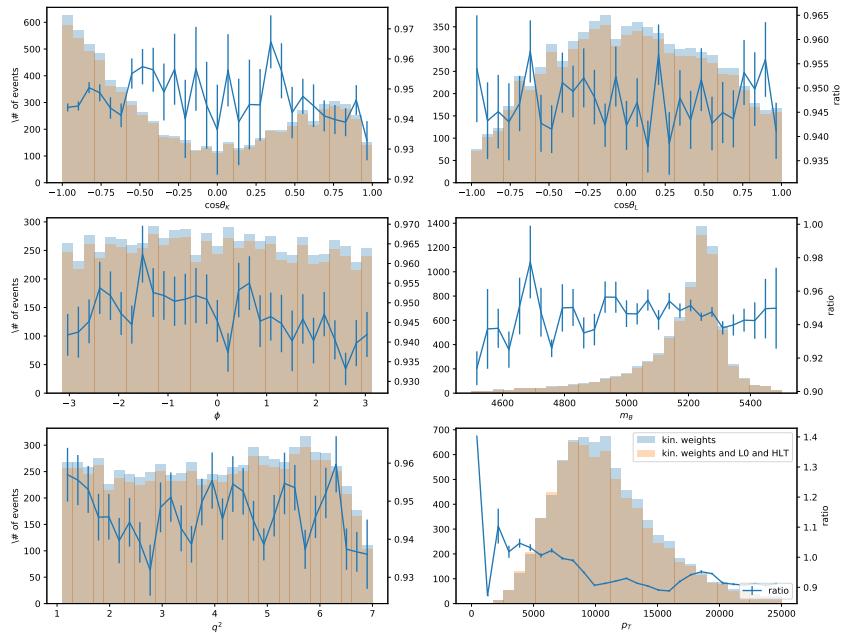
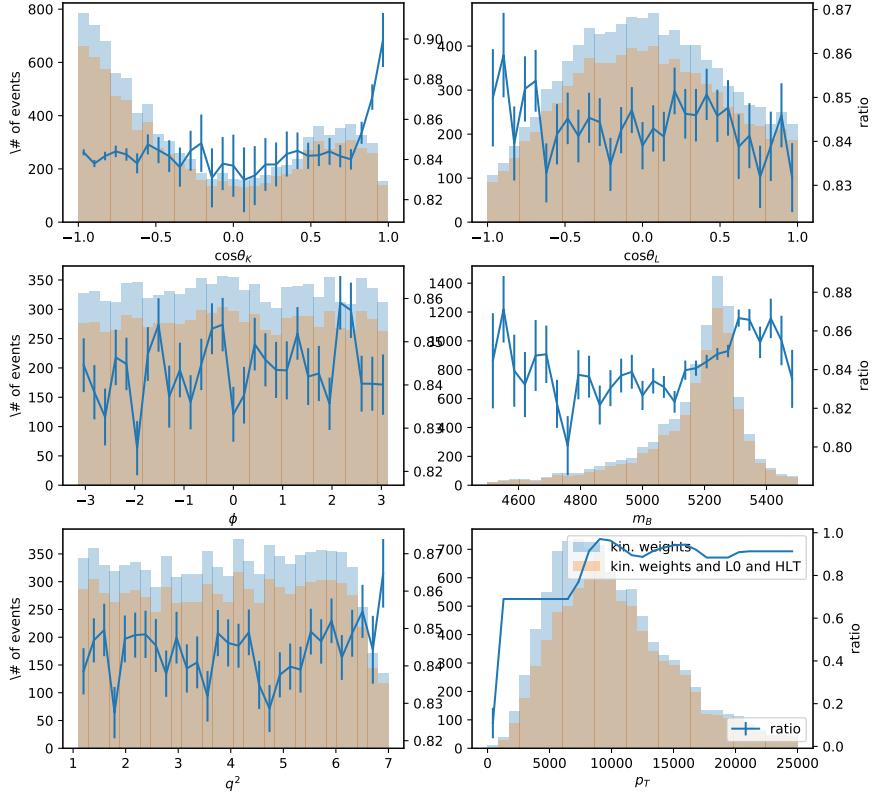


Figure 130: HLT, L0 and kinematically reweighted (blue) and only kinematically reweighted (orange) distributions of a simulated sample of $B^0 \rightarrow K^{0*} e^+ e^-$ for the year 2012 divide in the two L0 trigger categories L0I (top) and L0E (bottom). The blue points represent the ratio between the two histograms.

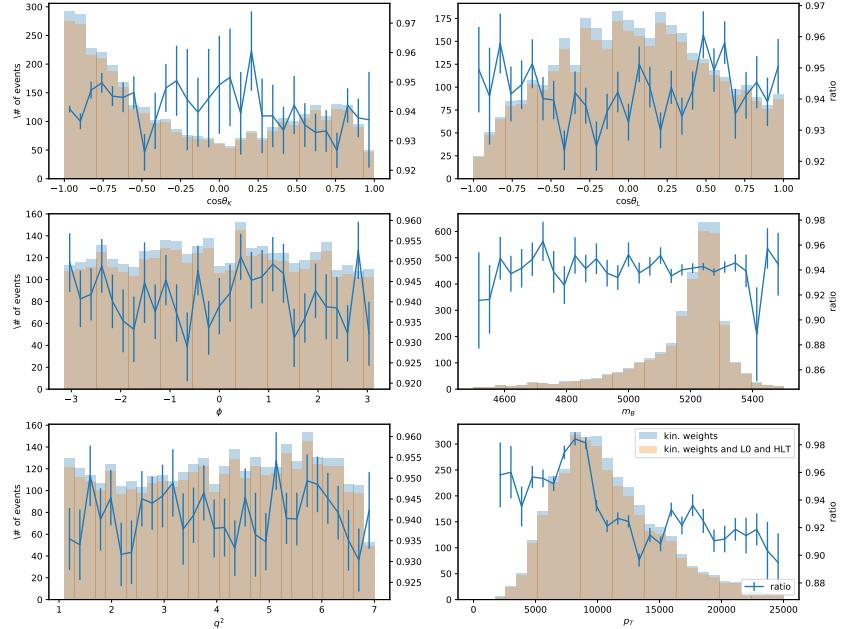
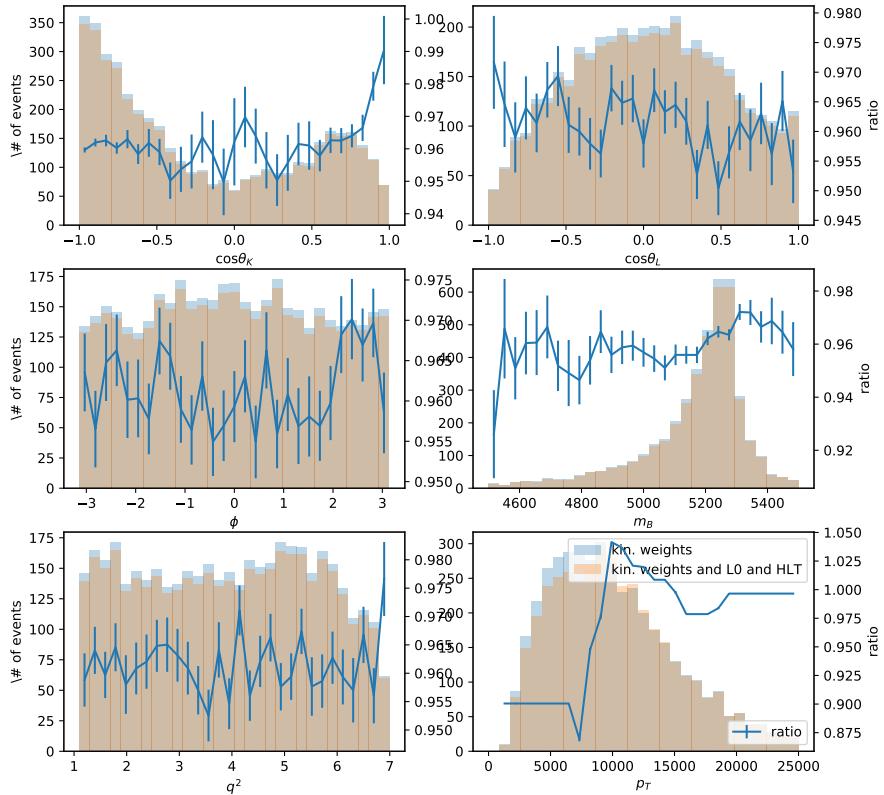


Figure 131: HLT, L0 and kinematically reweighted (blue) and only kinematically reweighted (orange) distributions of a simulated sample of $B^0 \rightarrow K^{0*} e^+ e^-$ for the year 2015 divide in the two L0 trigger categories L0I (top) and L0E (bottom). The blue points represent the ratio between the two histograms.

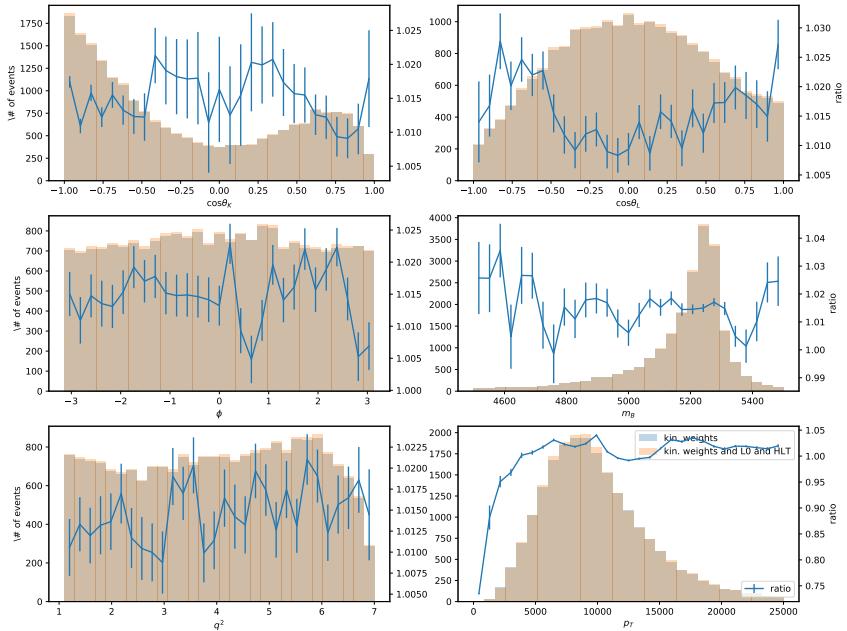
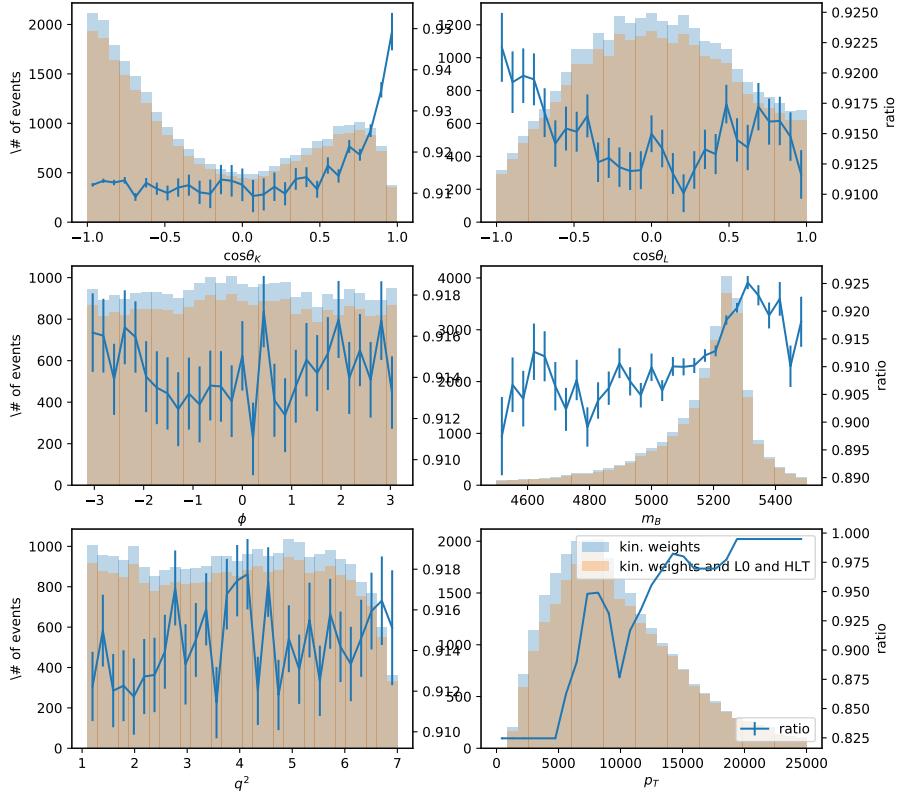


Figure 132: HLT, L0 and kinematically reweighted (blue) and only kinematically reweighted (orange) distributions of a simulated sample of $B^0 \rightarrow K^{0*} e^+ e^-$ for the year 2016 divide in the two L0 trigger categories L0I (top) and L0E (bottom). The blue points represent the ratio between the two histograms.

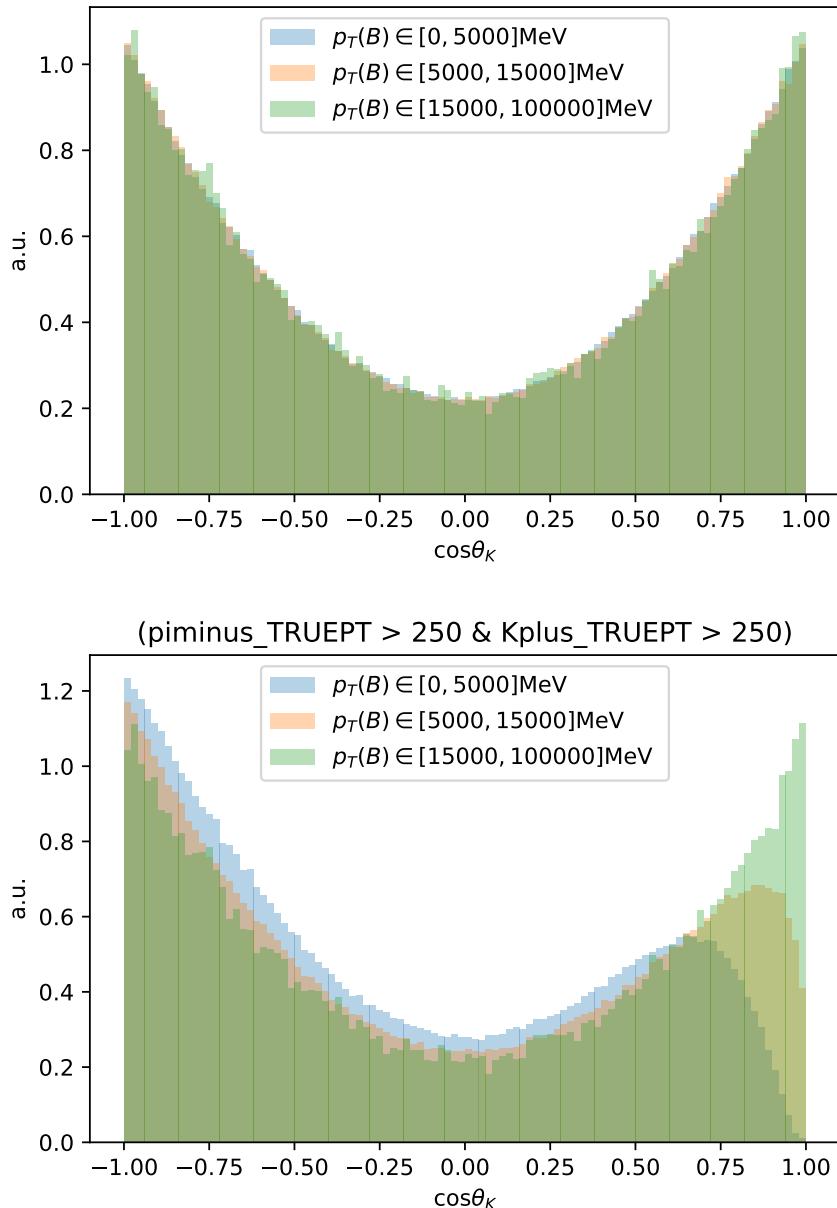


Figure 133: Variation of $\cos\theta_K$ in three bins of $p_T(B)$ with and without the Stripping cut $\text{Pi_PT} > 250 \&\& \text{K_PT} > 250$ applied. The sample used for this check corresponds to a private production of generator level MC for $B_0 \rightarrow K^{*0}e^+e^-$ in the central- q^2 region.

2934 D Additional information on the MVA

2935 D.1 Preliminary preselection for MVA

2936 The older set of offline selections given in Table Z below was used in studies that led to
 2937 the determination of the best set of features and hyper-parameters for the multivariate
 2938 classifier. These are relevant to Sections (give sections).

Table 58: Summary of the offline selection requirements applied to signal and background samples used in the determination of the best settings for the multivariate classifier. The difference between this, and the nominal selections used for the new MVAs are mainly in the addition of a few quality cuts, and modifications in selections against exclusive backgrounds.

Type		Requirement
Quality	all tracks	$\chi^2/\text{ndf} < 3$ $\text{GhostProb} < 0.4$
	K, π	$\text{region}_{\text{HCAL}}^{\text{Localo}} \geq 0$
	e	$\text{region}_{\text{ECAL}}^{\text{Localo}} \geq 0$
ID	K^{*0}	$!(\text{xProjection}_{\text{ECAL}}^{\text{Localo}} < 363.6 \text{ mm}$ $\&\& \text{yProjection}_{\text{ECAL}}^{\text{Localo}} < 282.6 \text{ mm})$
PID	all	hasRich
	e	hasCalo
	K, π	$p_T > 250 \text{ MeV}/c$
	e	$p_T > 500 \text{ MeV}/c$
	K	$\text{V2ProbNNk} \cdot (1 - \text{V2ProbNNp}) > 0.05$
BKG	π	$\text{V2ProbNNpi} \cdot (1 - \text{V2ProbNNk}) \cdot (1 - \text{V2ProbNNp}) > 0.1$
	e	$\text{V3ProbNNe} > 0.2$
	$B^+ \rightarrow K^+ e^{++} e^{--}$ $B_s^0 \rightarrow \phi e^{++} e^{--}$	$ m((h \rightarrow e)e) - m_{J/\psi,(\psi(2S))}^{\text{PDG}} > 60 \text{ MeV}/c^2$ $\max(m(Ke^+e^-), m((\pi \rightarrow K)e^+e^-)) < 5100 \text{ MeV}/c^2$ $m(K(\pi \rightarrow K)) > 1040 \text{ MeV}/c^2$
q^2	Sig	$0.1 \text{ GeV}^2/c^4 < q^2 < 7.0 \text{ GeV}^2/c^4$, $0.1 \text{ GeV}^2/c^4 < q_B^2 < 7.0 \text{ GeV}^2/c^4$
	Bkg	$0.1 \text{ GeV}^2/c^4 < q^2 < 1000.0 \text{ GeV}^2/c^4$
Mass	Bkg	$m(K\pi e^+e^-) > 5600 \text{ MeV}/c^2$

2939 **D.2 Classification performance**

2940 The optimisation of a given classifier is achieved by performing a grid search in its
2941 parameter space, that is, by evaluating the performances of the classifier for different
2942 settings of the available parameters. The values considered for each scan are given below.

2943
2944 For TMVA, the following configurations are tested:

- 2945 • `Shrinkage` = [0.01, 0.05, 0.1];
2946 • `NTrees` = [10, 50, 100];
2947 • `method` = [kBDT].

2948 For GradientBoosting and XGBoost, a finer grid search is preferred, corresponding to

- 2949 • `max_depth` = [2, 3, 4, 5, 6, 7]
2950 • `subsample` = [0.5, 0.75, 1.]
2951 • `learning_rate` [0.01, 0.05, 0.1, 0.15]
2952 • `min_samples_leaf` = [1, 3, 5, 7, 9]

2953 for GradientBoosting and to

- 2954 • `max_depth` = [5, 6, 7]
2955 • `subsample` = [0.5, 0.75, 1.]
2956 • `eta` = [0.1, 0.3, 0.5, 0.7]
2957 • `colsample` = [0.5, 0.75, 1.]
2958 • `min_child_weight` = [1, 5, 10, 20]

2959 for XGBoost. This decision is driven by the fact that these two classifiers are expected to
2960 provide a better performance compared to the others. For DTC and EXC, which share the
2961 same parameters, the following configurations are tested:

- 2962 • `max_depth` = [2, 3, 4, 5, 6, 7];
2963 • `criterion` = [gini, entropy].

2964 The grid search performed for each examined classifier is shown in Fig. 134.

2965 **D.3 Uniform classifiers performance**

2966 The ROC curve of the uniform classifiers is shown in Fig. 135, while Fig. 136 shows the
2967 classifier response for the training and testing samples of signal and background. The
2968 efficiency as a function of the uniform variables is shown in Fig. 137 for the classifiers
2969 with uniformity along the B^0 invariant mass only and in Fig. 138 for the classifiers with
2970 uniformity along the B^0 invariant mass and the decay angles θ_ℓ , θ_K , and ϕ .

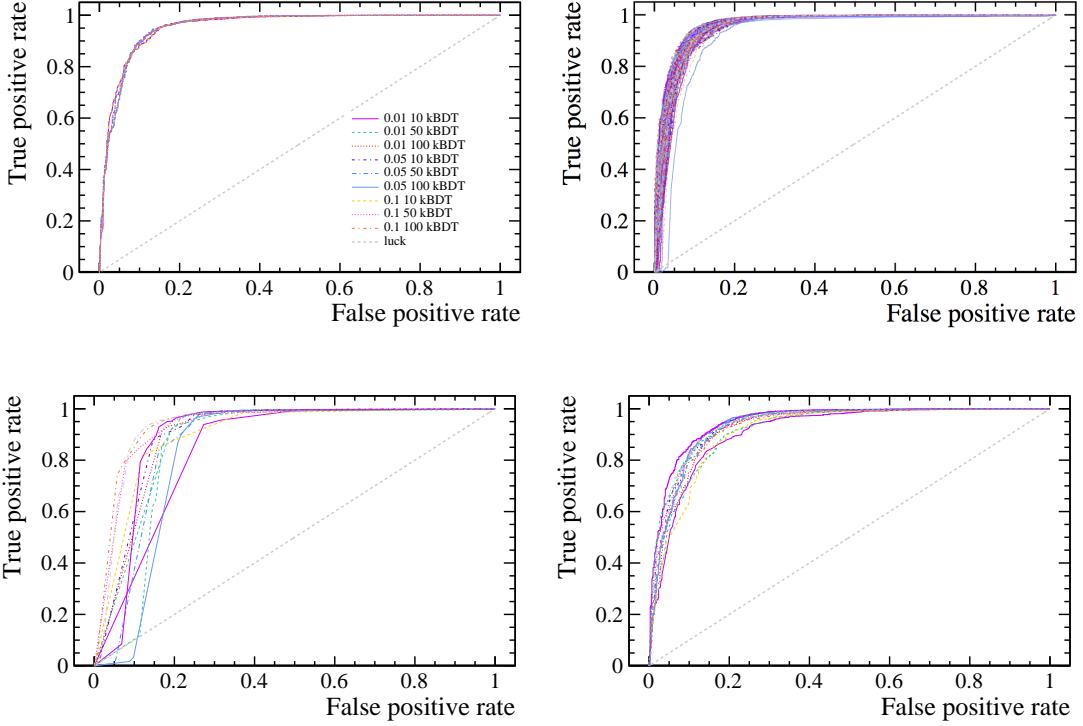


Figure 134: ROC curve of each configuration of the TMVA (top left), GradientBoosting (top right), DTC (bottom left), and EXC (bottom right) classifiers. Each color refers to a specific configuration of the classifier, as shown in the legend of the TMVA plot. The other legends are not shown, since the number of ROC curves is too high.

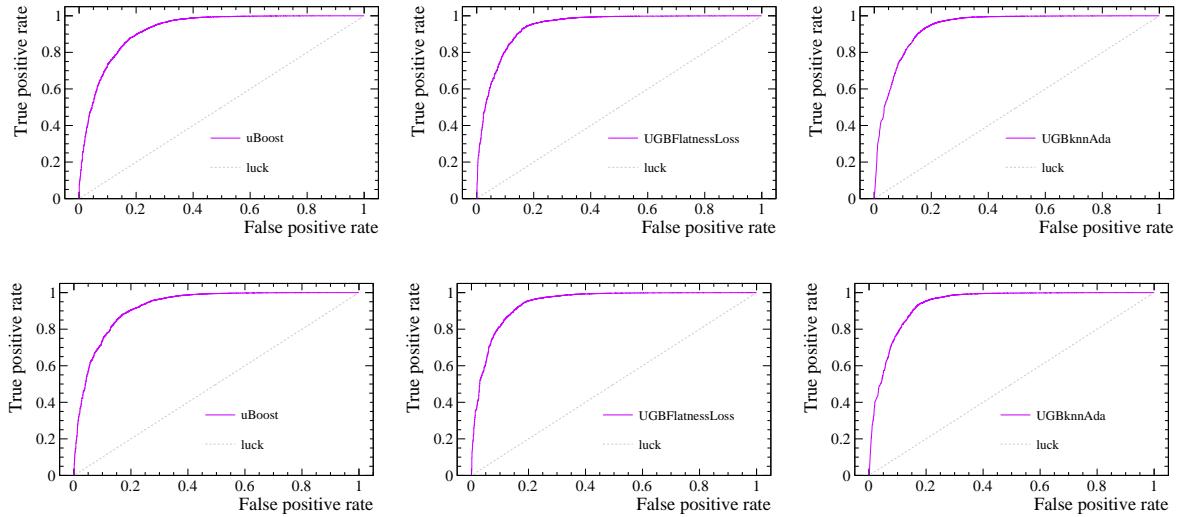


Figure 135: ROC curve of the uBoost (left), UGBknnAda (center), and UGBFlatnessLoss (right) uniform classifiers with uniformity along the B^0 invariant mass only (top) and along the B^0 invariant mass and the decay angles θ_ℓ , θ_K , and ϕ (bottom).

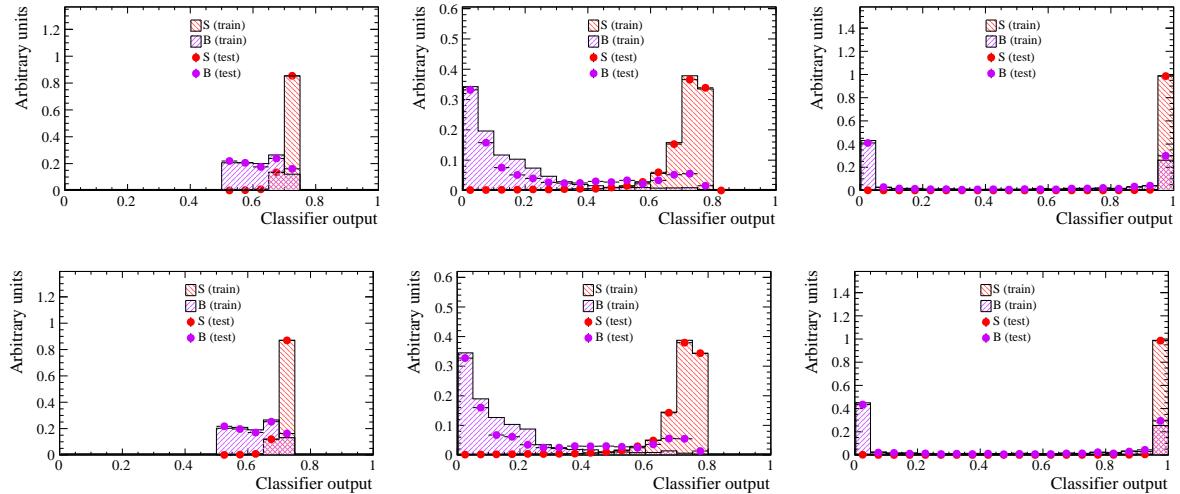


Figure 136: Classifier response for the training and testing samples of signal and background for the uBoost (left), UGBknnAda (center), and UGBFlatnessLoss (right) uniform classifiers with uniformity along the B^0 invariant mass (top) and along the B^0 invariant mass and the decay angles θ_ℓ , θ_K , and ϕ (bottom).

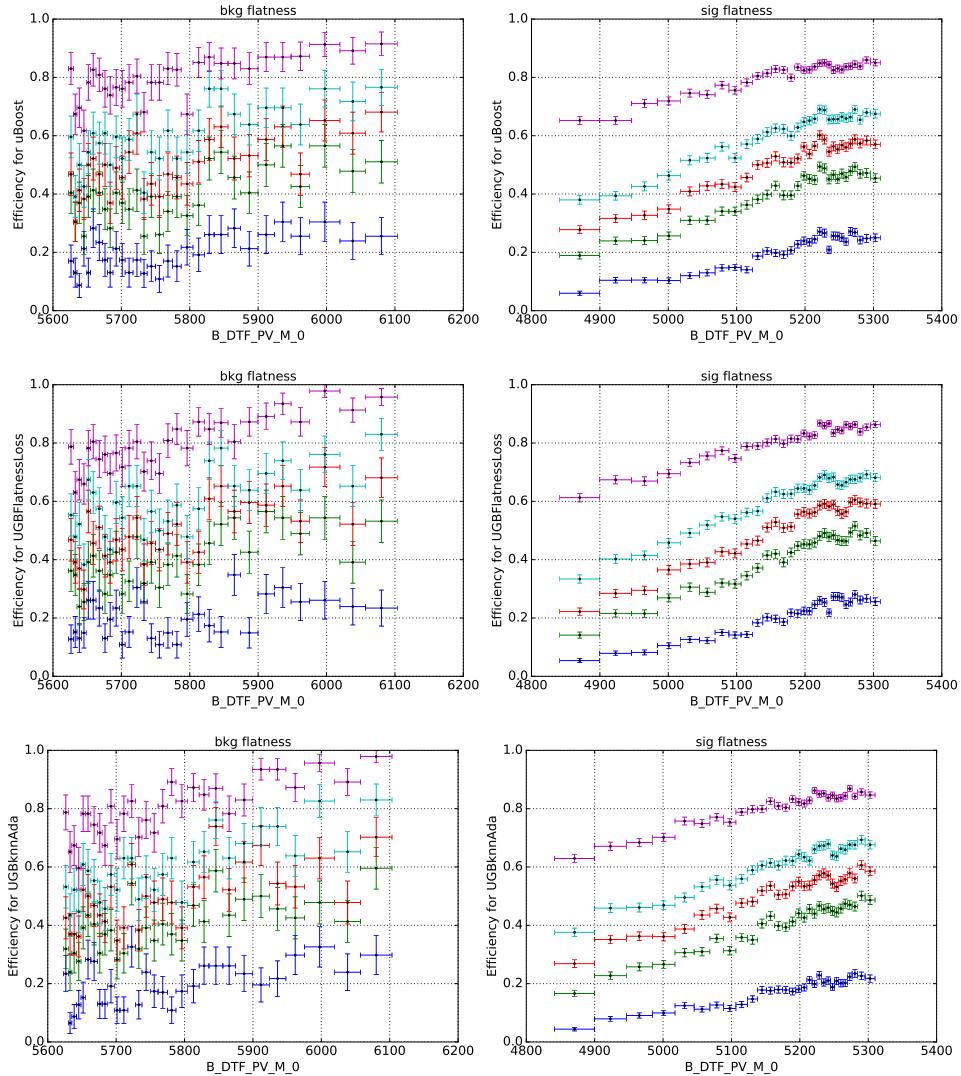


Figure 137: Efficiency as a function of the B^0 invariant mass on the background (left) and signal (right) samples for the uBoost (top), UGBknnAda (middle), and UGBFlatnessLoss (bottom) uniform classifiers with uniformity along the B^0 invariant mass.

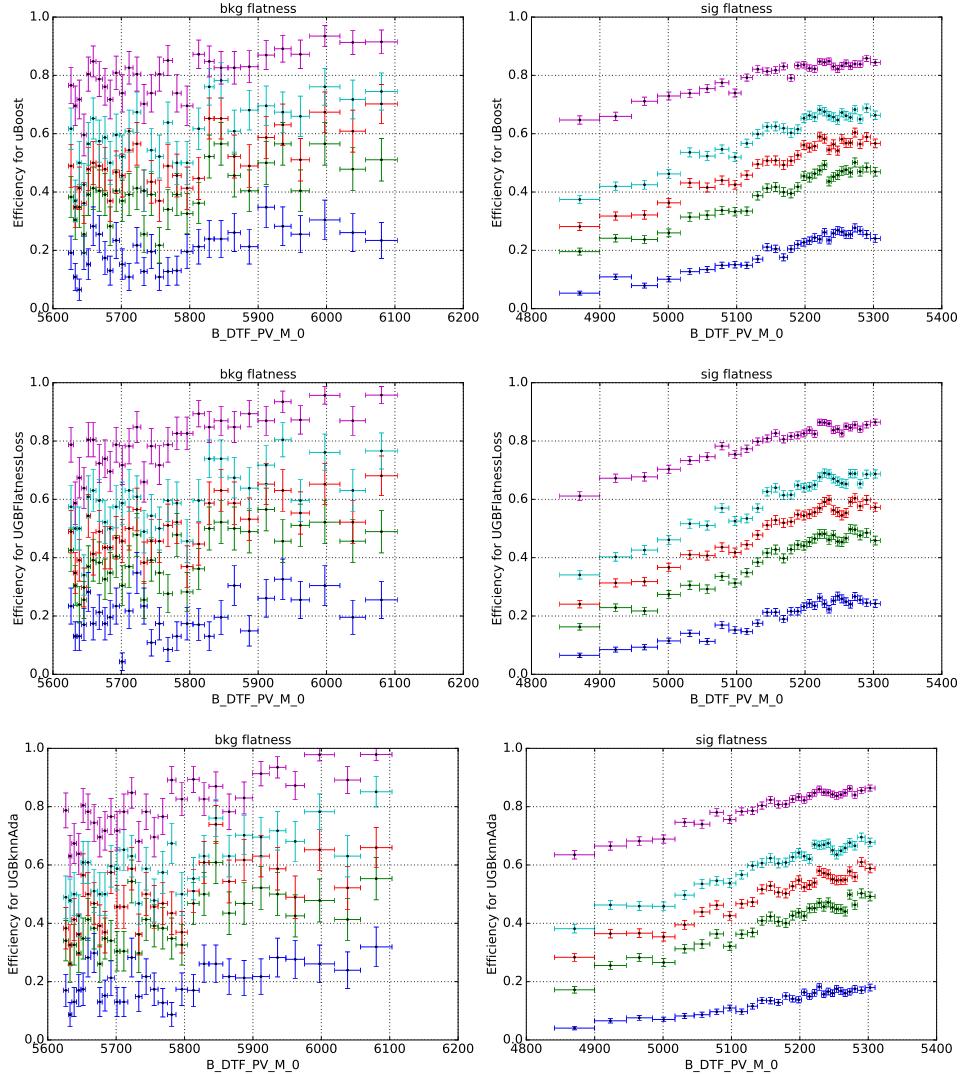


Figure 138: Efficiency as a function of the B^0 invariant mass on the background (left) and signal (right) samples for the uBoost (top), UGBknnAda (middle), and UGBFlatnessLoss (bottom) uniform classifiers with uniformity along the B^0 invariant mass and the three decay angles.

2971 **D.4 Performance of new classifiers**

2972 The multivariate classifiers used in this analysis are set up using the optimal set of features
 2973 and hyper-parameters found through studies documented in Sections X-Y, but are trained
 2974 with updated signal simulation and nominal preselection (Table X instead of Table Y).
 2975 Two separate classifiers are made for Run-I and Run-II.

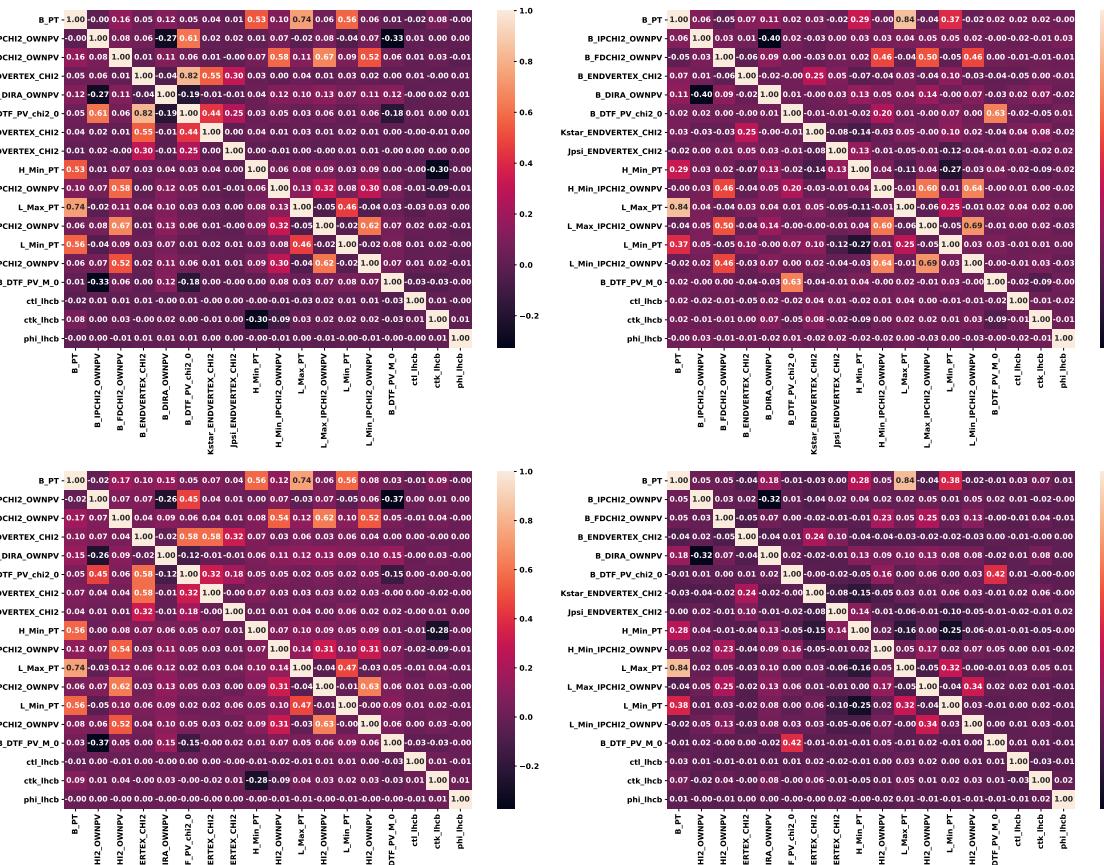


Figure 139: Correlations between variables of interest (inputs to MVA, and mass and angles used in fits) for signal and background training samples. Run-I and Run-II values are shown on the top and bottom rows, respectively.

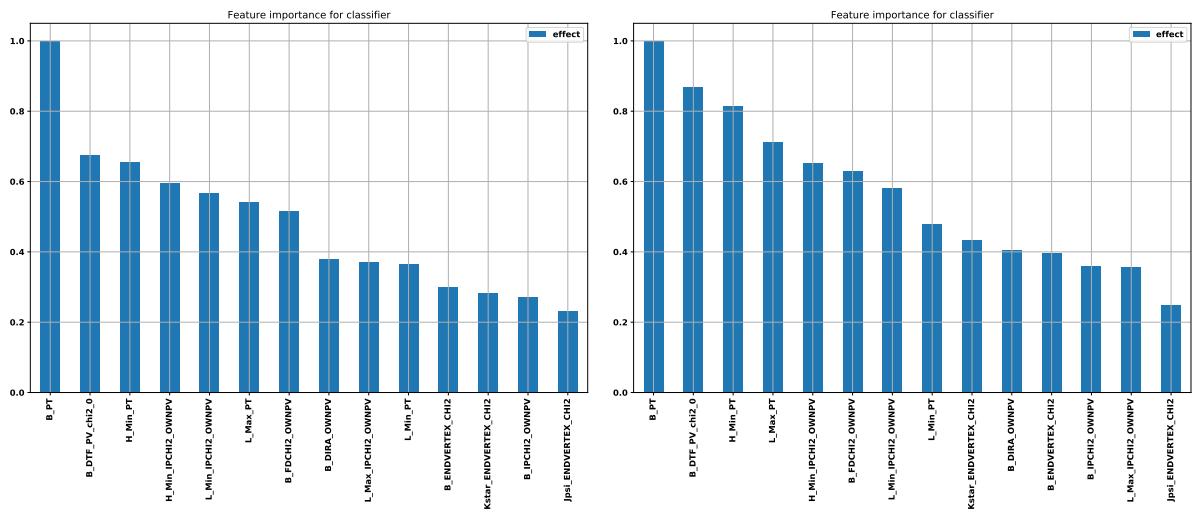


Figure 140: Importance of input features for Run-I and Run-II classifiers.

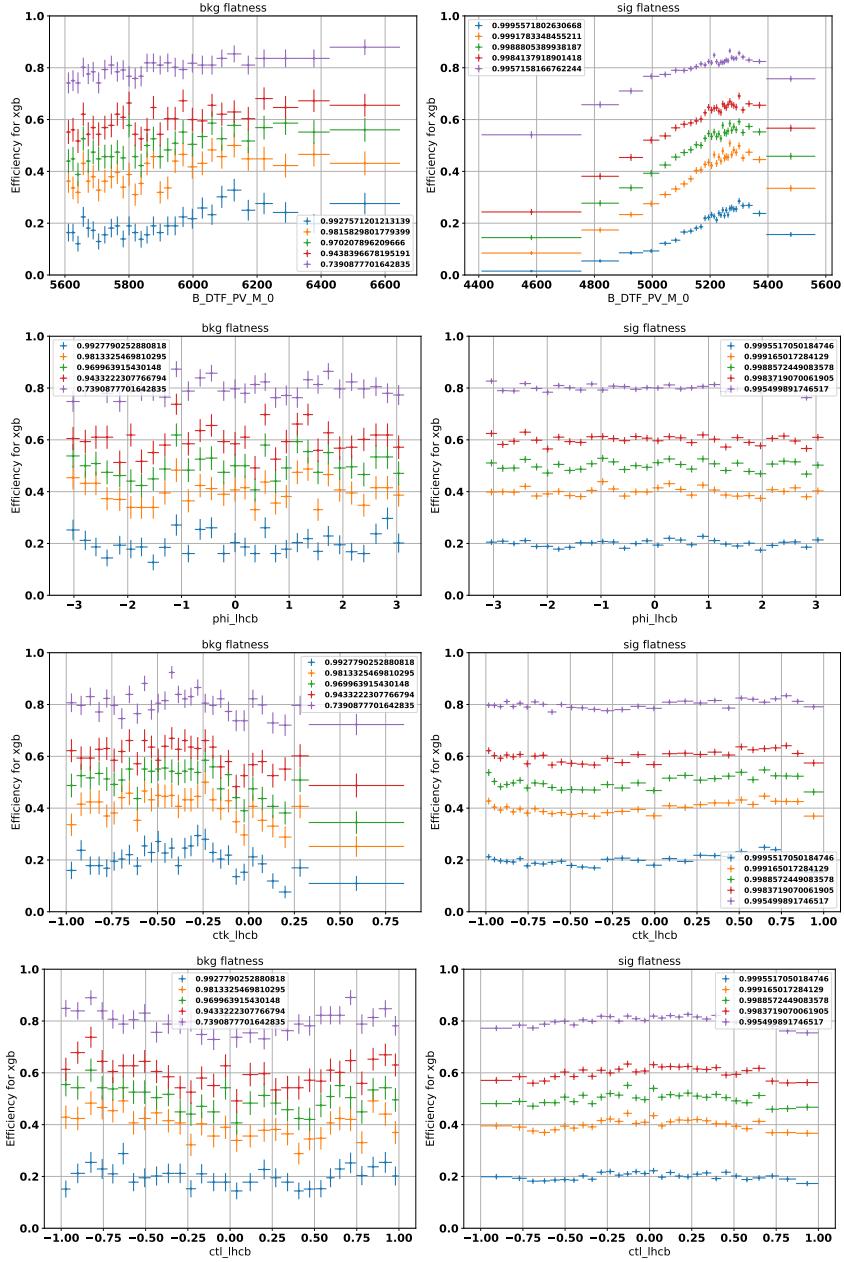


Figure 141: Efficiency as a function of the B^0 mass and angles for the Run-I classifier.

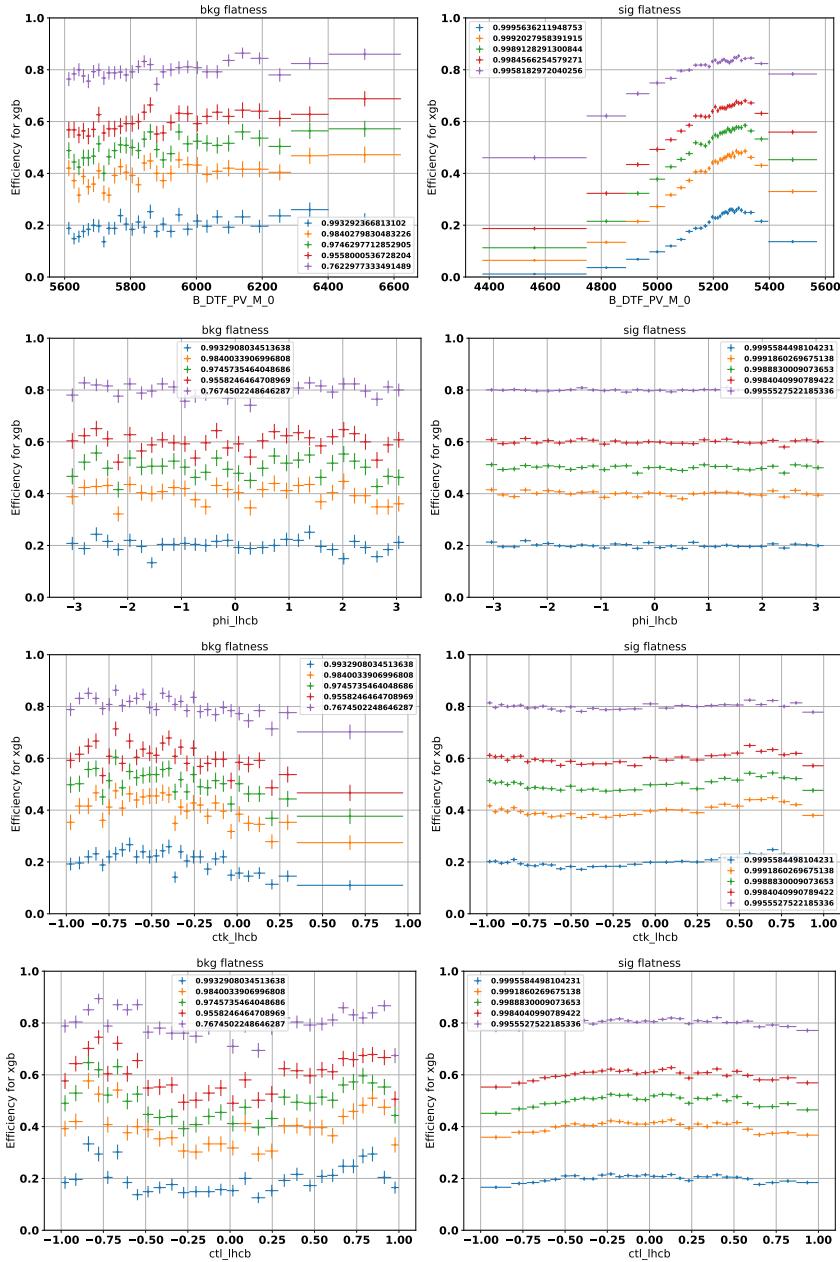


Figure 142: Efficiency as a function of the B^0 mass and angles for the Run-II classifier.

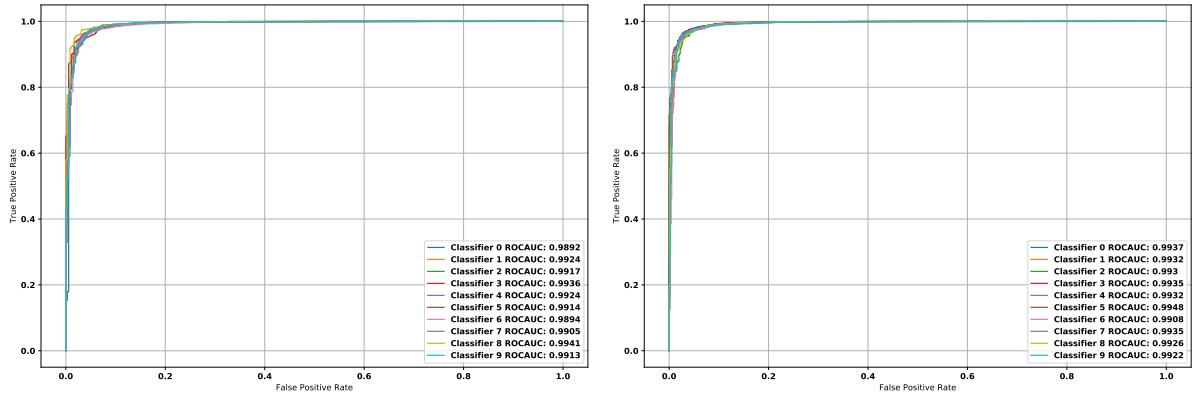


Figure 143: ROC curves corresponding to the Run-I and Run-II classifiers.

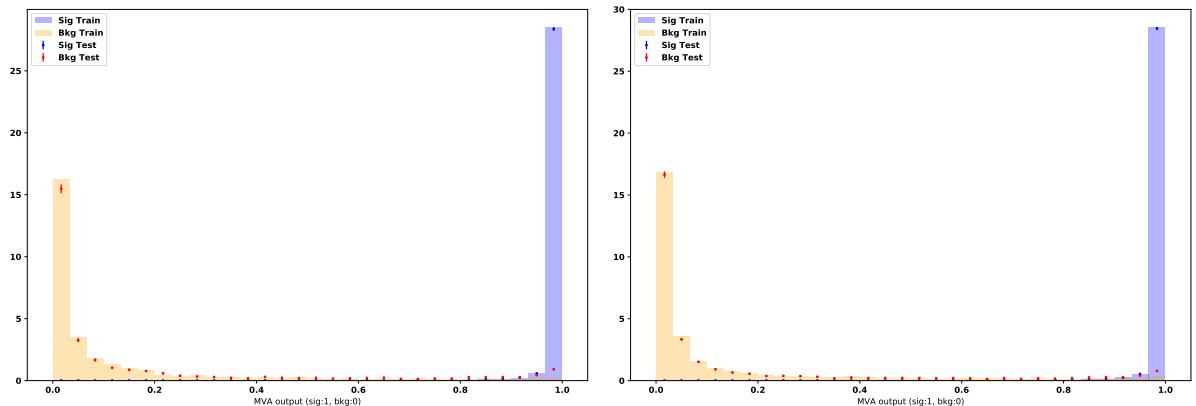


Figure 144: Comparison between classifier performance on the signal and background training and testing samples for Run-I (top) and Run-II (bottom) classifiers.

2976 **D.5 MC/data comparison for MVA variables**

2977 To check for the agreement between simulation and data for the (signal) training sample,
2978 the control mode is used for simplicity, as it is difficult to obtain s-weights for the rare mode,
2979 and the direct comparison between rare and control mode distributions is complicated
2980 by the different q_c^2 regions. S-weights are obtained from fits to J/ψ constrained B mass
2981 (with the same selection used to cut the training samples with the exceptions of q_c^2 range
2982 and mass window), and used to weight the data distributions for RunI and RunII. These
2983 are compared with simulation with kinematic correction weights only (no trigger weights
2984 are used in BDT training). The results are shown in Fig. 145, a reasonable agreement is
2985 observed in across all distributions.

2986 Additionally, distributions of RunI and RunII signal and background training samples
2987 are also shown in Fig. 146

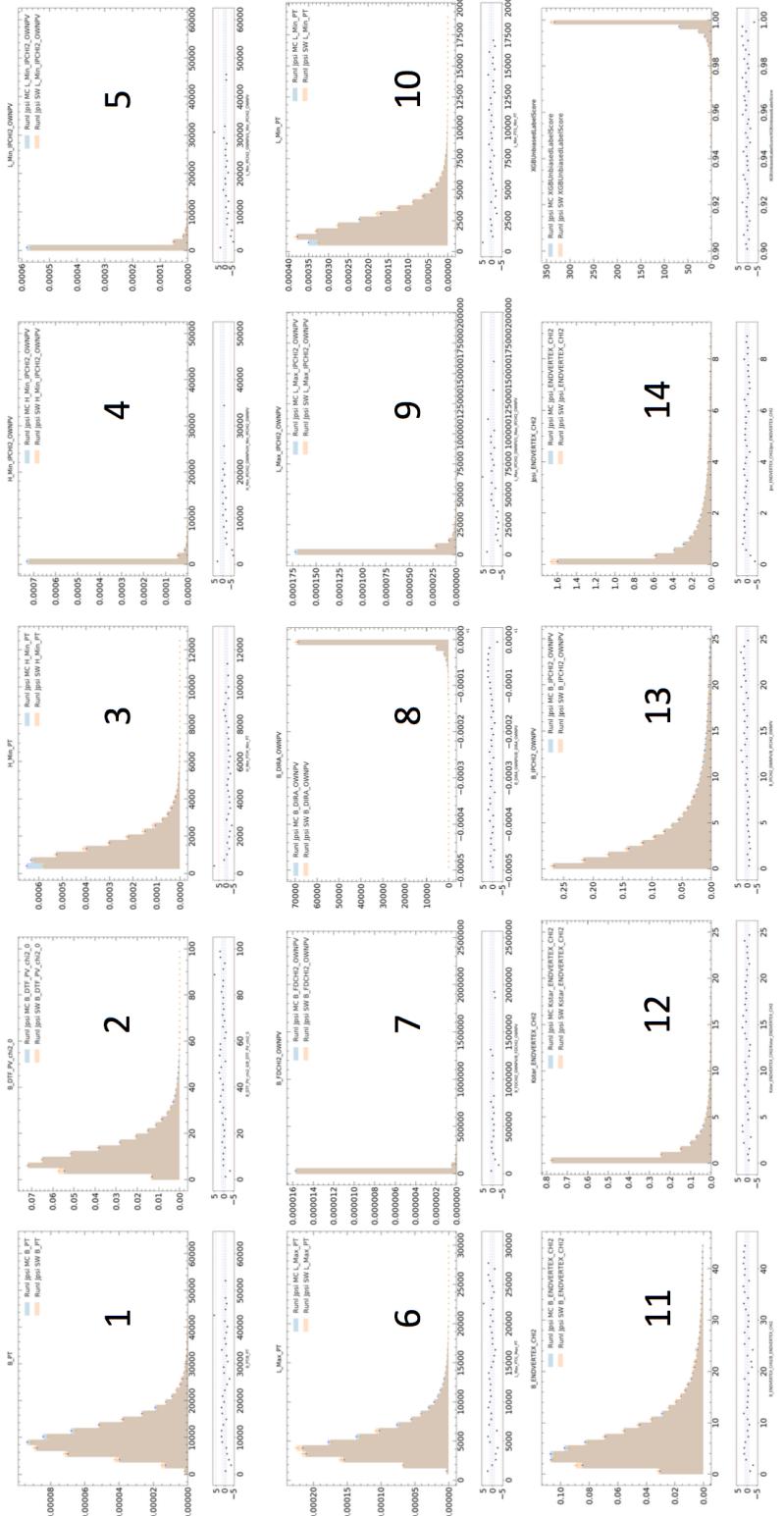


Figure 145: Comparison between signal sWeighted $B^0 \rightarrow J/\psi(\rightarrow e^+e^-)K^{*0}$ and corrected simulation for the different MVA variables used, where the 14 input variables are shown arranged in order of importance (1=most, 14=least). 247

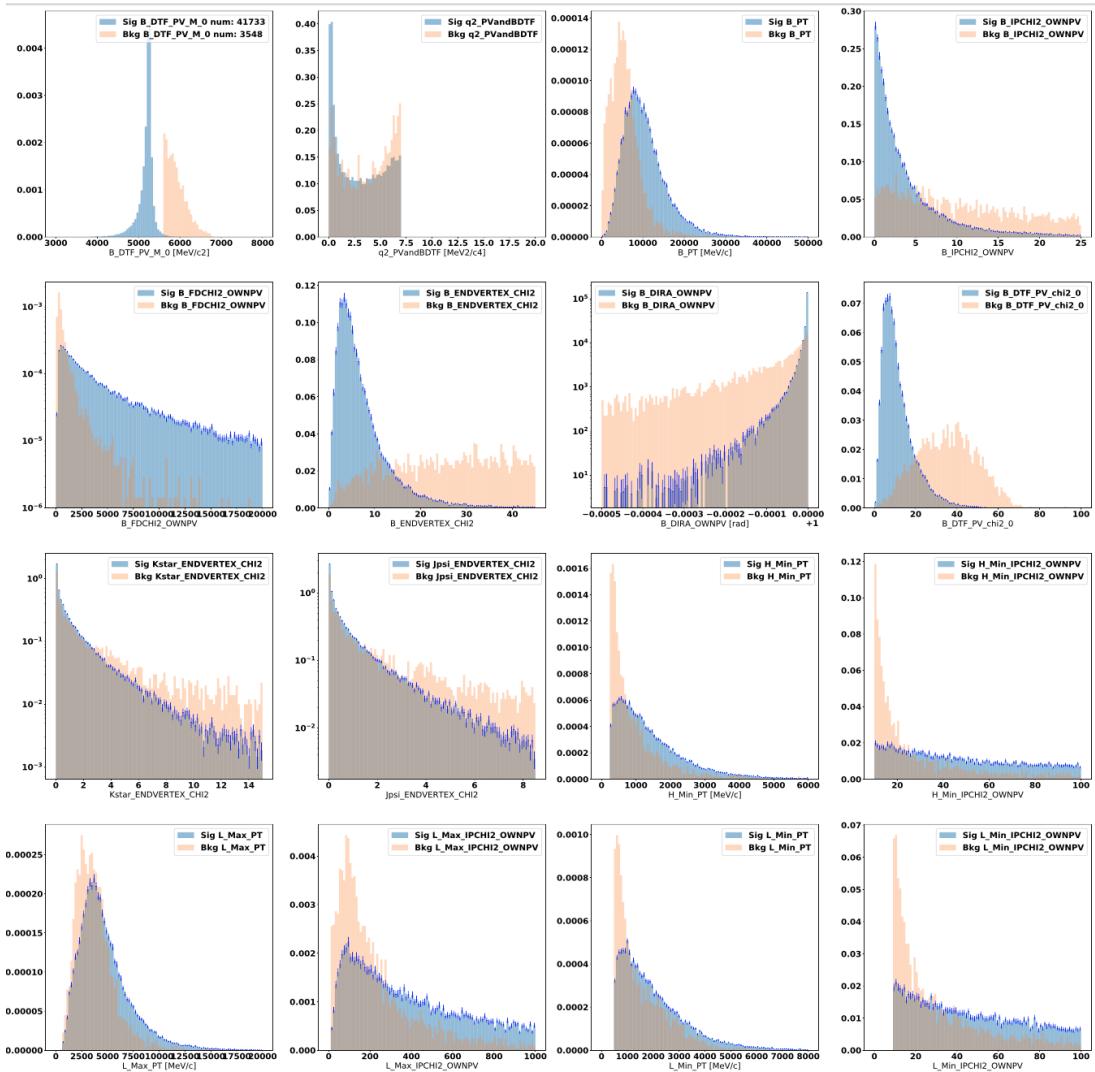


Figure 146: Signal and background training variables used as input for the MVA algorithm.

2988 **D.6 Additional info on the MVA optimisation**

2989 **D.6.1 Mass fits used in the MVA optimization procedure**

2990 Two background types are considered, namely the backgrounds due to partially recon-
2991 structed and combinatorial events. The partially reconstructed component is modelled
2992 using $B^- \rightarrow K^+ \pi^- \pi^- e^+ e^-$ simulation via kernel density estimation with data-driven cor-
2993 rection weights, while the combinatorial component is modelled with an exponential pdf,
2994 the slope of which is left free to vary in the data fit. Note that while the optimisation was
2995 carried out prior to the modelling of the DSL background, the strategy at the time – the
2996 usage of the DSL veto and mHOP cut – both reduce the contribution of this background
2997 in the sample, and its impact on the mass fit (see the summary at the end of this section
2998 for a list of the differences between the set-up used for optimisation and the nominal
2999 configuration). The mass distributions of the rare and control modes are modelled using
3000 simulation without separating bremsstrahlung categories. The control mode fit is used to
3001 determine the values of the shift and scaling parameters of the signal mass peak, which
3002 are used to take into account remaining data-simulation differences.

3003 As an example, the result of three mass fits with three very different MVA cuts (0.5,
3004 0.99, 0.9968) are shown in Figures 147, 148 and 149.

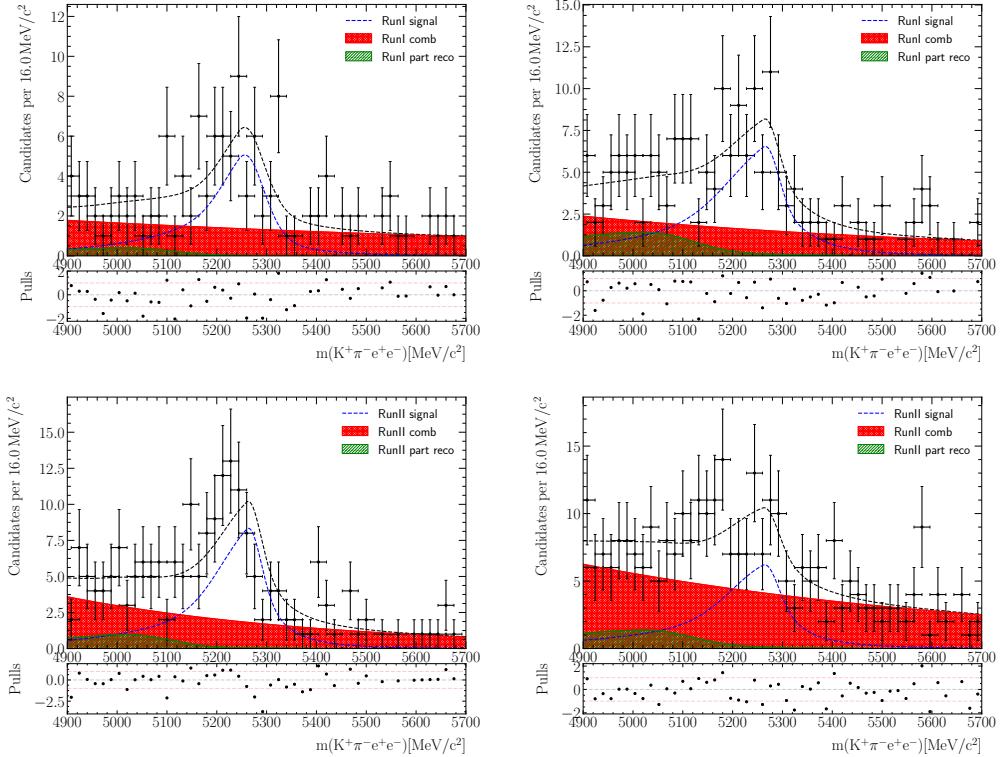


Figure 147: Result of the fit to rare mode candidates with $\text{MVA} > 0.50$ (lower end of the range considered for optimisation). For stability, the yield of the partially reconstructed component is Gaussian constrained to the value obtained by scaling the result of the fit for the tight cut of $\text{MVA} > 0.999$. Run-I results are shown on the top row, and Run-II on the bottom row. Results of the two trigger categories, LOE and LOTIS, are shown on the left and right columns, respectively.

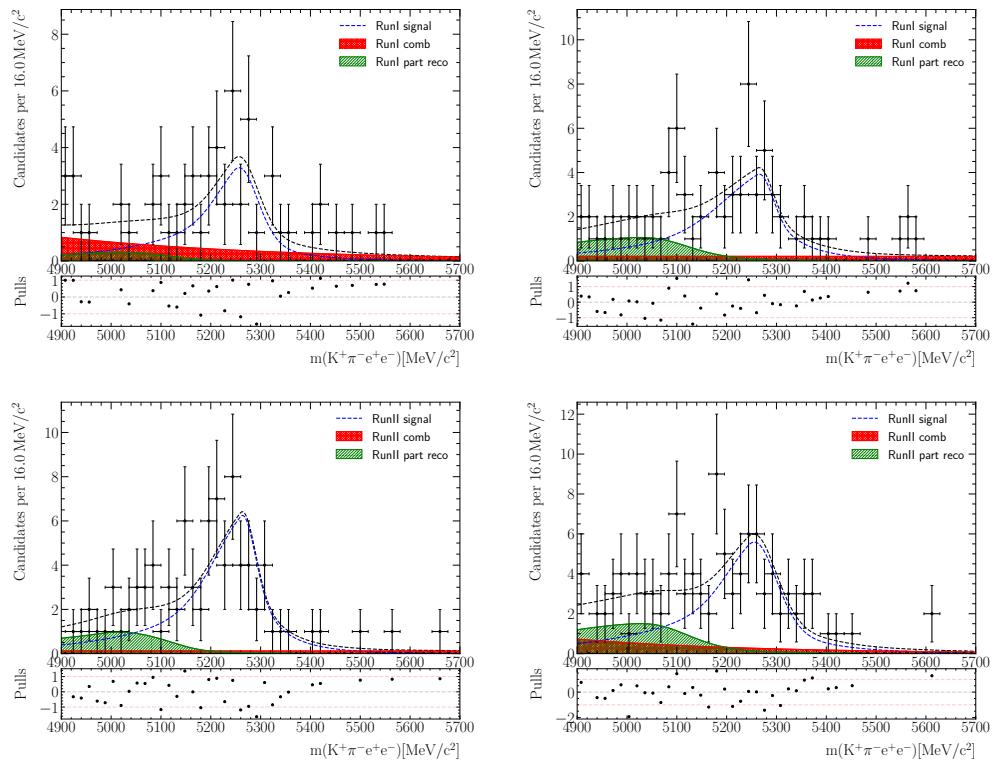


Figure 148: Result of the fit to rare mode candidates with $\text{MVA} > 0.99$. For stability, the yield of the partially reconstructed component is Gaussian constrained to the value obtained by scaling the result of the fit for the tight cut of $\text{MVA} > 0.999$. Run-I results are shown on the top row, and Run-II on the bottom row. Results of the two trigger categories, LOE and LOTIS, are shown on the left and right columns, respectively.

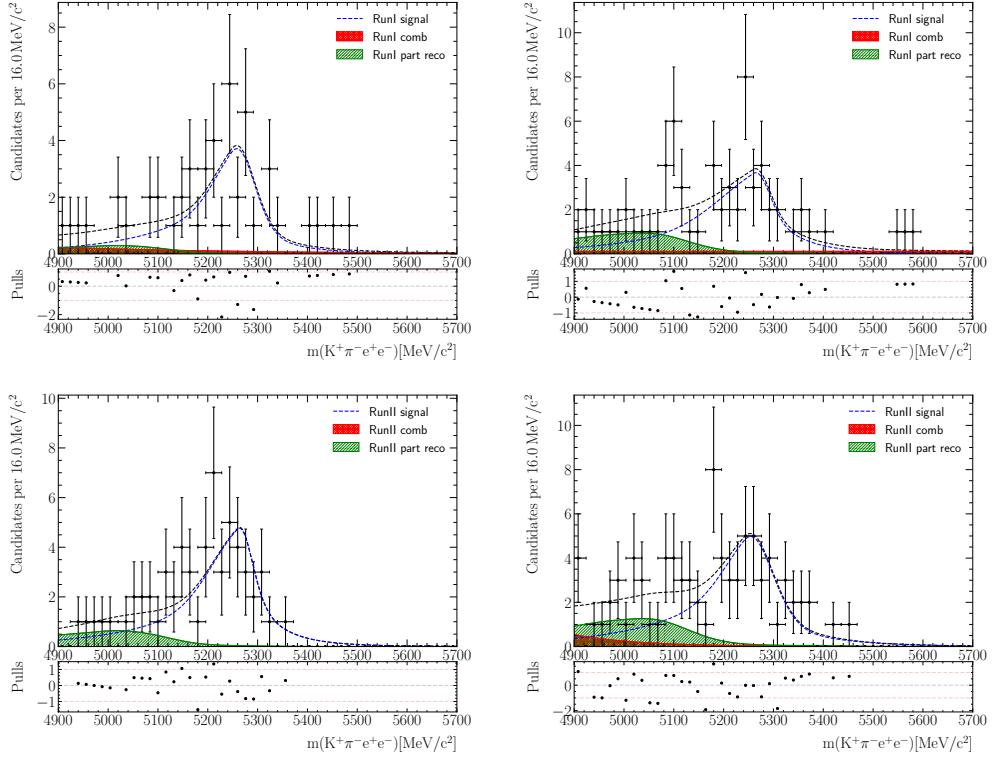


Figure 149: Result of the fit to rare mode candidates with $\text{MVA} > 0.9968$ (upper end of the range considered for optimisation). For stability, the yield of the partially reconstructed component is Gaussian constrained to the value obtained by scaling the result of the fit for the tight cut of $\text{MVA} > 0.999$. Run-I results are shown on the top row, and Run-II on the bottom row. Results of the two trigger categories, LOE and LOTIS, are shown on the left and right columns, respectively.

3005 **D.6.2 P'_5 Figure of Merit**

3006 This sections discusses additional details about the implementation of the toy study used to
3007 access the P'_5 -FoM. A simplified model has been implemented for the signal angular model
3008 utilising a folding technique that is applied to specific regions of the three-dimensional
3009 angular space to exploit the (anti)-symmetries of the differential decay rate with respect
3010 to combinations of angular variables (see Appendix F for definition). In this case, the set
3011 of transformations consists in folding the angles ϕ and θ_ℓ to remove all observables except
3012 F_L , P_1 and P'_5 . This significantly simplifies the differential decay rate and consequently
3013 the stability of these studies without losing experimental sensitivity.

3014 **Pseudoexperiment configuration** For each MVA cut, toys are generated based on
3015 the estimated signal and background yields as discussed in Sec. 4.5 and then fitted to
3016 determine P'_5 sensitivity. Standard model observable values from Flavio [95] in the q^2
3017 range of 1.1 to 7.0 GeV^2/c^4 are used for the signal angular pdf. Angular models for the
3018 partially reconstructed background are obtained by parametrising the angular distributions
3019 of the $B^- \rightarrow K^+\pi^-\pi^-e^+e^-$ MC with data-driven correction weights, and models for the
3020 combinatorial background are taken from fits to $K\pi e^+\mu^-$ data samples. For simplicity,
3021 the acceptance functions is not included in the modelling process. Separate models are
3022 made for the two L0 trigger categories of both Runs. For simplicity, the shapes of the
3023 background angular distributions are kept the same for all MVA cuts, as the MVA response
3024 is sufficiently flat for background angular distributions. Nevertheless a ‘safety cut’ of 0.2
3025 is applied to the samples used for parametrisation. Examples of the background models
3026 used are shown on Figures 150 and 151 for the Run-II L0E category. In contrast to
3027 the angular distributions, some shaping effect is seen for the mass, so the slopes of the
3028 exponential distributions for the combinatorial background are taken from data fits for each
3029 cut. Oppositely, the signal model, its width and scaling are obtained for each cut from
3030 $B^0 \rightarrow K^{*0}e^+e^-$ simulation for the former, and $B^0 \rightarrow K^{*0}J/\psi(\rightarrow e^+e^-)$ data fit for the
3031 latter.

3032 The fits are made simultaneously to the four subsamples sharing the observables of
3033 interest. The yields of all components and the four combinatorial slope parameters are
3034 allowed to vary, but the angular shape and the shift and scaling parameters are fixed to
3035 generation values. Fits are made to around two to three hundred toys to ensure that
3036 at least one hundred converged fits are obtained. The spread of the resulting P'_5 values
3037 is taken as the sensitivity. One toy fit example for the cut of $\text{MVA} > 0.99$ is shown in
3038 Figure 152 along with the distribution of P'_5 values obtained from around two hundred
3039 converged fits.

3040 **Differences with respect to nominal angular fit** The MVA optimisation is carried
3041 out early in the analysis timeline prior to the availability of additional results, findings and
3042 decisions that shaped the nominal configuration. Currently it is not considered necessary
3043 to re-perform the optimisation, as the improvements are expected to be marginal. For
3044 completeness, the list of noteworthy differences is given below:

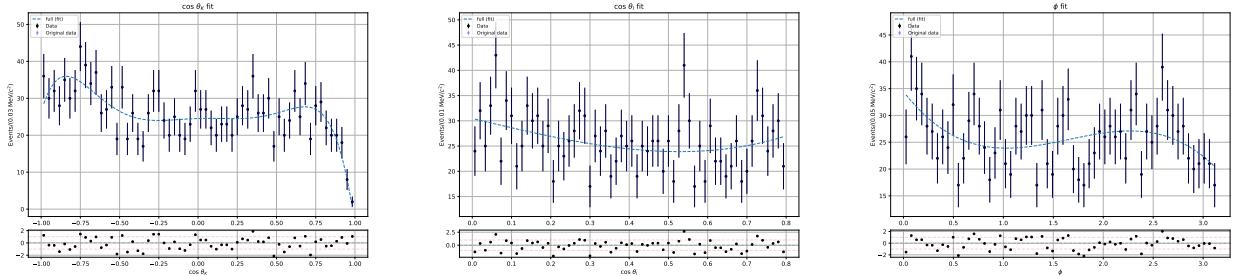


Figure 150: Angular distributions of the combinatorial component for the Run-II L0E subsample used in toy generation. Note that the signal acceptance function is not applied.

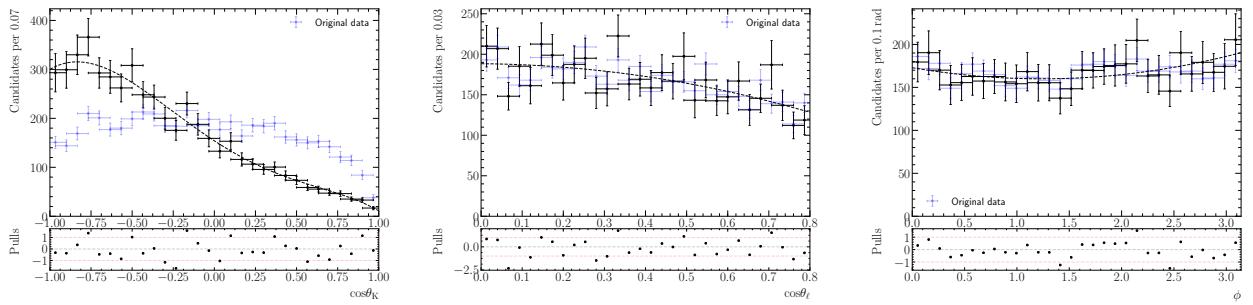


Figure 151: Angular distributions of the partially reconstructed component for the Run-II L0E subsample used in toy generation. Note that the signal acceptance function is not applied.

- 3045 • **Folding:** the optimisation is carried out using folded pdfs
- 3046 • **DSL background:** DSL background vetoed using $|\cos\theta_\ell| < 0.8$
- 3047 • **Acceptance:** no acceptance weights used in background modelling; toys are gener-
3048 ated and fitted without acceptance effects/corrections
- 3049 • **Selection:** slightly different set of selection requirements; mHOP cut of $m_{\text{corr}} >$
3050 $4926 + 10 \ln(\text{FD } \chi^2)$ applied; q^2 regions selected using square q^2 cut (e.g. $q^2 \in [1.1, 7.0] \text{GeV}^2/\text{c}^4$ & $q_c^2 \in [1.1, 7.0] \text{GeV}^2/\text{c}^4$)
- 3052 • **Mass model:** shift and scaling parameters obtained using control mode fit to PV
3053 and J/ψ constrained B^0 mass; no B_s^0 component in control mode fit
- 3054 • **Fit strategy:** extended maximum likelihood fits used in all parts of the optimisation
3055 procedure

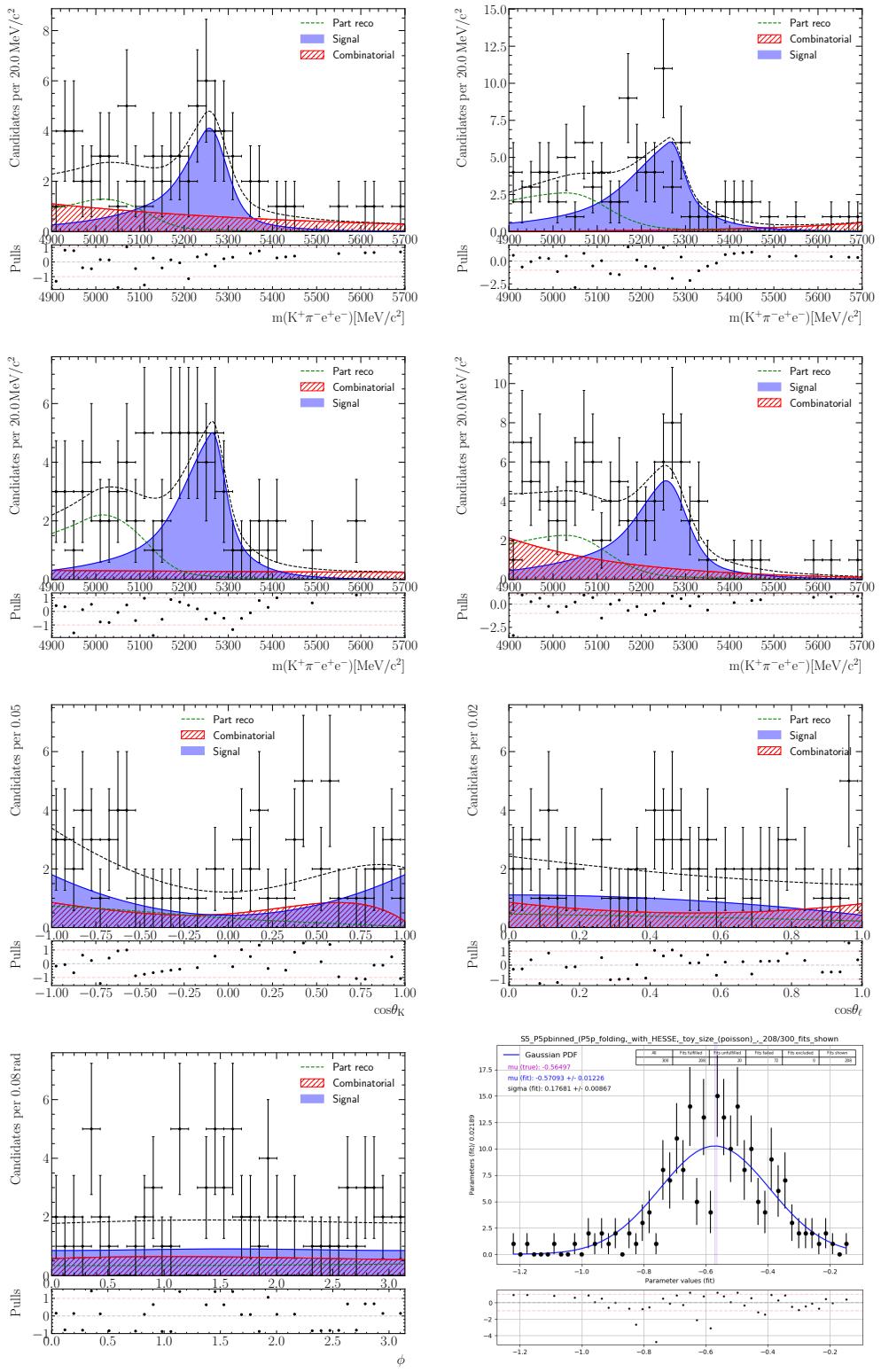


Figure 152: One example toy fit for the cut of $\text{MVA} > 0.99$. The P'_5 value from around two hundred such fits are shown in the bottom right plot. No constraints are applied to any component in the toy fit, which leads to large fluctuations in combinatorial and partially reconstructed background yields.

3056 E Selection for partially reconstructed backgrounds

3057 Partially reconstructed $B^+ \rightarrow K^{*(**)}e^+e^-$ decays are particularly complicated backgrounds
 3058 for the $B^0 \rightarrow K^{*0}e^+e^-$ channel due to the limited mass resolution. These are in general
 3059 modelled in the mass fit, however, investigations have been carried out to assess methods
 3060 to reduce possible contamination from such contributions. These studies have been
 3061 performed using Run-I Sim08 simulation and data samples using an older set of preselection
 3062 requirements, and prior to the choice of the narrower B^0 mass window of 4900 MeV/c² to
 3063 5700 MeV/c². Nevertheless, the conclusions remain valid and helped to inform the choice
 3064 of the current configuration.

3065 In this section two approaches are discussed: selection on the so-called HOP variable
 3066 together with $\ln(\text{FD } \chi^2)$, aligned to Ref. [52]; a dedicated multivariate classifier to remove
 3067 this background. The performance of the two methods is studied on simulated events of $B^+ \rightarrow K_1^+e^+e^-$, with $K_1^+ \rightarrow K^+\pi^+\pi^-$, and $B^+ \rightarrow K_2^*(1430)^+e^+e^-$, with $K_2^*(1430)^+ \rightarrow K^+\pi^+\pi^-$.

3068 The distribution of the $B^+ \rightarrow K_1^+e^+e^-$ MC candidates as a function of the PV-
 3069 constrained B^0 invariant mass $m(K^+\pi^-e^+e^-)_{\text{PV}}$ and q^2 is shown in Fig. 153 (left),
 3070 while the analogous distribution as a function of the PV-constrained B^0 invariant mass
 3071 $m(K^+\pi^-e^+e^-)_{\text{PV}}$ and constrained q^2 is shown in Fig. 153 (right). These candidates are
 3072 distributed across the whole q^2 region where the analysis is performed and, although they
 3073 are more abundant in the low mass region, can have B^0 invariant mass compatible with
 3074 that of the signal.

3076 E.1 HOP approach

3077 In the $B^0 \rightarrow K^{*0}e^+e^-$ decay, the vectorial sum of the final state particles is not expected
 3078 to have a component orthogonal to the flight direction of the B^0 meson. This property

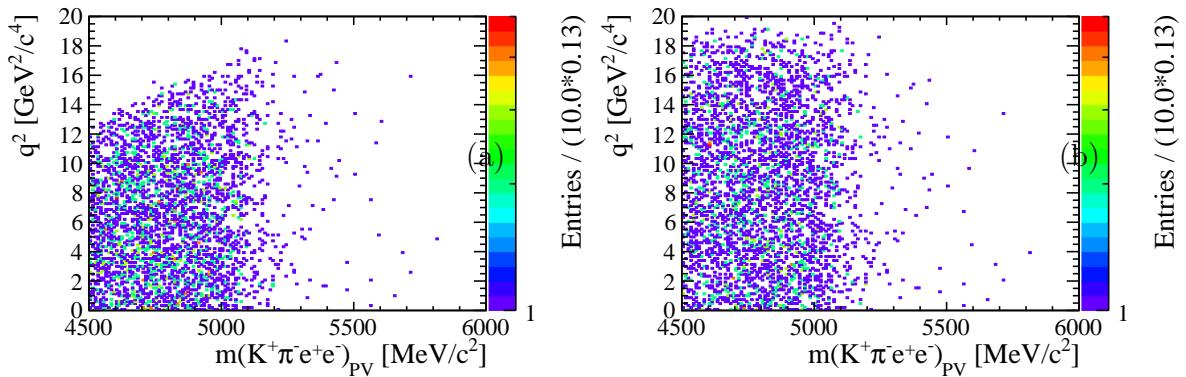


Figure 153: Distribution of the $B^+ \rightarrow K_1^+e^+e^-$ MC candidates as a function of the PV-constrained B^0 invariant mass $m(K^+\pi^-e^+e^-)_{\text{PV}}$ and q^2 (left) or constrained q^2 (right), after the selection is applied.

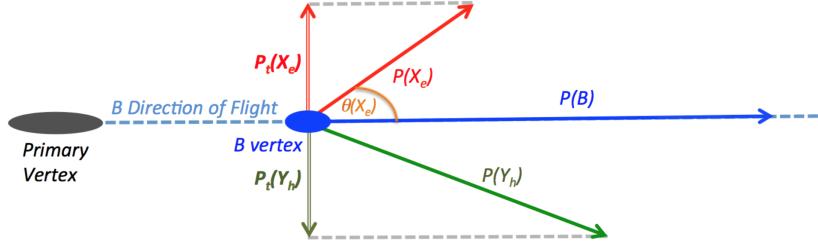


Figure 154: Sketch of the variables involved in the α_{HOP} definition.

3079 does not necessarily hold for $B^0 \rightarrow K^{*0}e^+e^-$ candidates that originate from partially
 3080 reconstructed decays or, more in general, from some kind of background. This different
 3081 behaviour can hence be exploited to improve the rejection power of the selection described
 3082 above. To this extent, two new variables, α_{HOP} and m_{corr} , are defined. The flight direction
 3083 of the B^0 meson is determined from the position of its primary and decay vertices. The
 3084 momenta of the K^{*0} meson and of the dilepton pair orthogonal to the flight direction of
 3085 the B^0 meson are used to compute the correction factor α_{HOP} , according to:

$$\alpha_{\text{HOP}} = \frac{p_{\text{T}}(K^{*0})}{p_{\text{T}}(\ell^+\ell^-)}.$$

3086 A sketch of the variables involved in the α_{HOP} definition is shown in Fig. 154. For signal
 3087 candidates, α_{HOP} is expected to be unity, with possible deviations from unity mainly due
 3088 to energy loss in the dilepton pair. The energy carried by the *bremsstrahlung* photons
 3089 can be recovered by multiplying the momentum of the dilepton pair orthogonal to the
 3090 flight direction of the B^0 meson by α_{HOP} . Since *bremsstrahlung* emission is, to a good
 3091 approximation, collinear to the direction of flight of the lepton, the same correction can
 3092 be applied to the longitudinal component, giving

$$p_{\text{corr}}(\ell^+\ell^-) = \alpha_{\text{HOP}} p(\ell^+\ell^-).$$

3093 The B^0 invariant mass can then be recalculated by using $p_{\text{corr}}(\ell^+\ell^-)$ instead of $p(\ell^+\ell^-)$.
 3094 According to the definition of α_{HOP} , the same correction is applied to both leptons when
 3095 recalculating the B^0 invariant mass. This approximation neglects the fact that each lepton
 3096 might have lost a different amount of energy via *bremsstrahlung*.

3097 The two-dimensional distribution in m_{corr} and flight distance χ^2 of the $B^0 \rightarrow K^{*0}e^+e^-$
 3098 and $B^+ \rightarrow K_1^+e^+e^-$ MC candidates after the selection is applied is shown in Fig. 155.
 3099 The analogous distribution for the $B^0 \rightarrow K^{*0}e^+e^-$ data candidates is shown at the bottom
 3100 of Fig. 155. The distributions are obtained by combining the low and central q^2 bins.
 3101 The difference between the distributions of the $B^0 \rightarrow K^{*0}e^+e^-$ and $B^0 \rightarrow K_1^+e^+e^-$ MC
 3102 candidates allows to reduce the contamination from partially reconstructed decays by
 3103 requiring

$$m_{\text{corr}} > 5072 + \ln(\text{FD } \chi^2)$$

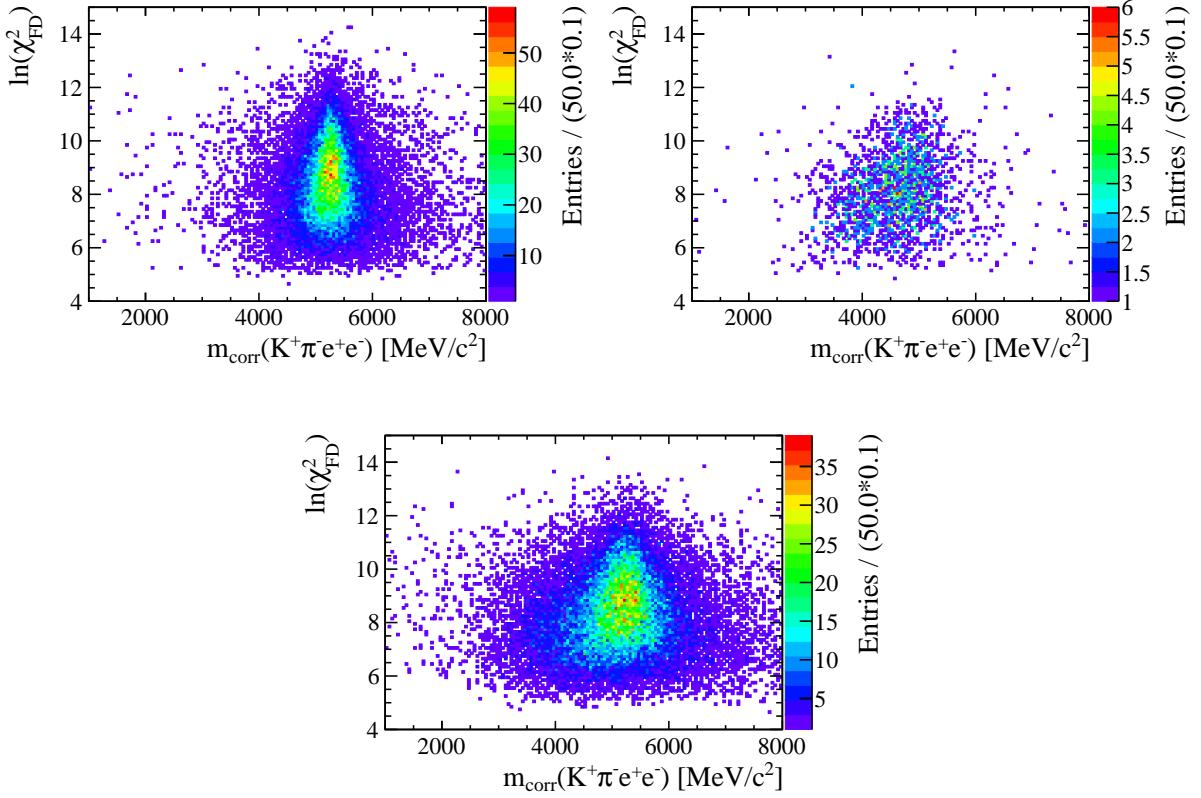


Figure 155: Number of $B^0 \rightarrow K^{*0} e^+ e^-$ (top left) and $B^+ \rightarrow K_1^+ e^+ e^-$ (top right) MC candidates as a function of m_{corr} and flight distance χ^2 , after the selection is applied. The HOP requirement corresponds to a nearly vertical straight line on the left side of the signal.

for the low q^2 bin and

$$m_{\text{corr}} > 4926 + 10 \ln(\text{FD } \chi^2)$$

for the central q^2 bin. This selection, which is optimised separately for the two q^2 bins, is 75% efficient on the signal but remove nearly 70% of the background due to partially reconstructed decays.

The invariant mass distribution of the candidates selected or discarded by the HOP requirement is shown in Fig. 156 for the $B^0 \rightarrow K^{*0} e^+ e^-$ and $B^+ \rightarrow K_1^+ e^+ e^-$ (right) MC candidates in the low and in the central q^2 bins. The plots refer to MC candidates in the LOE trigger category only, but similar results are obtained for the other trigger categories. Only candidates passing the selection are taken into account.

The benefit of adding the HOP requirement is evaluated by performing a mass fit of the $B^0 \rightarrow K^{*0} e^+ e^-$ data candidates and by computing the signal yield with and without such requirement. The fitting procedure is described in detail in Section ???. The fit results are shown in Fig. 157 for both the baseline (top) and HOP (middle) approaches. It is observed that, although the absolute number of background decays is lower when

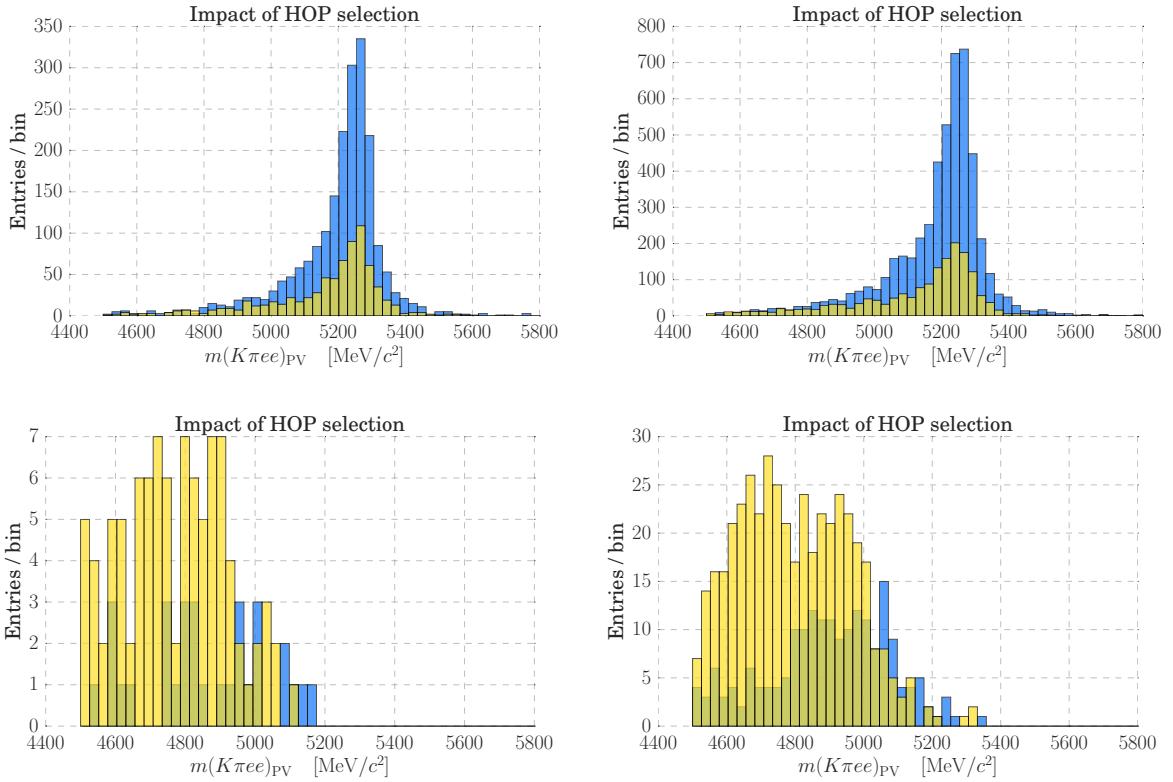


Figure 156: Invariant mass distribution of the $B^0 \rightarrow K^{*0} e^+ e^-$ (top) and $B^+ \rightarrow K_1^+ e^+ e^-$ (bottom) MC candidates in the low q^2 bin (left) and in the central q^2 bin (right) selected (blue) and discarded (yellow) by the HOP requirement.

the HOP requirement is applied, the statistical uncertainty associated to the signal yield has a negligible improvement. This is due to the fact that the HOP requirement is more efficient in removing background in the low mass region than in the region where the $B^0 \rightarrow K^{*0} e^+ e^-$ signal is, as shown in Fig. 156. For this reason, it is preferred to not include the HOP requirement in the selection.

3123 E.2 Multivariate classifier approach

Another approach that is investigated in order to reduce the contamination from partially reconstructed decays consists in using a dedicated multivariate classifier. The latter is trained to distinguish between signal decays and partially reconstructed decays. The $B^0 \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-)$ MC sample is used as signal, while the $B^+ \rightarrow K_1^+ J/\psi (\rightarrow e^+ e^-)$ MC sample is used as background. The choice of using the control mode instead of the rare mode in the training is driven by the fact that these samples have a larger statistics and are expected to show the same discrepancies between each other as the corresponding $B^0 \rightarrow K^{*0} e^+ e^-$ and $B^+ \rightarrow K_1^+ e^+ e^-$ rare modes. In addition, the resonant modes have

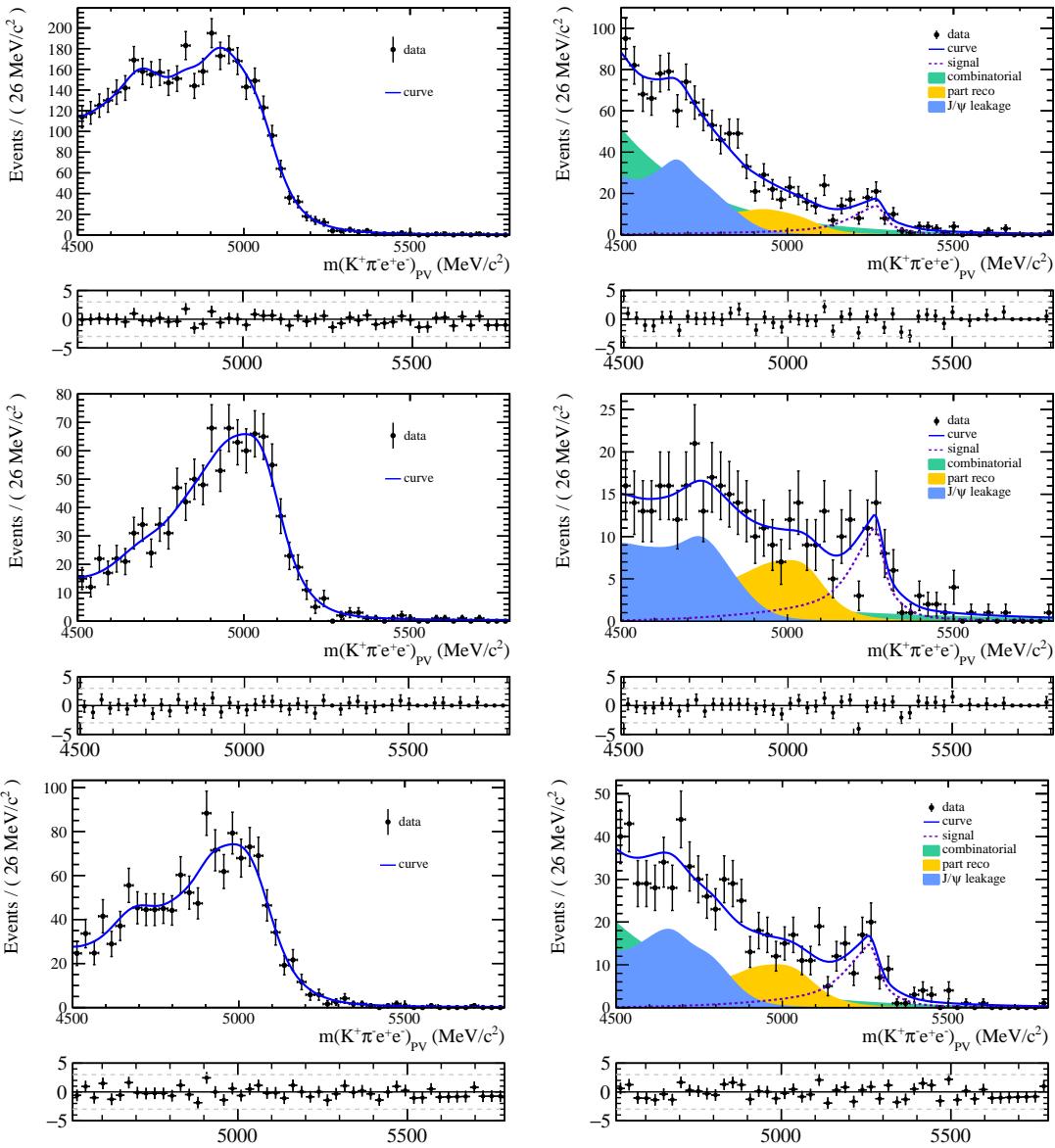


Figure 157: Invariant mass distribution of the $B^+ \rightarrow K_1^+ e^+ e^-$ MC candidates (left) and $B^0 \rightarrow K^{*0} e^+ e^-$ data candidates (right) in the central q^2 bin and in the LOE trigger category obtained when applying the baseline selection (top) and, in addition, the HOP approach (middle) or the multivariate classifier approach (bottom).

the advantage of allowing to determine the efficiency and rejection power of the trained classifier directly from the data samples.

The separation between signal and background is achieved by looking at the isolation properties of the B^0 candidates. In order to improve the rejection power of the classifier, the α_{HOP} variable defined in the previous section is also used as discriminating variable

3137 during the training. The m_{corr} variable is not used since it is largely correlated to the B^0
3138 invariant mass. In addition to these variables, the following quantities are also used:

- 3139 • p_T , χ^2_{IP} , flight distance χ^2 , and χ^2 of the kinematic fit of the B^0 candidates;
3140 • χ^2_{IP} of the leptons in the final state.

3141 As for the HOP approach, the benefit of this strategy is assessed by performing a mass
3142 fit of the $B^0 \rightarrow K^{*0} e^+ e^-$ data candidates with and without a requirement on the classifier
3143 response. The fit results are shown at the bottom and at the top of Fig. 157, respectively.
3144 Since the classifier removes background candidates mainly from the low mass region and
3145 given that the statistical uncertainty on the signal yield does not improve when removing
3146 these candidates, no selection on the classifier response is applied in the mass fit described
3147 in Section ??.

3148 F Folding technique

3149 A set of five angular transformations can be used to simplify the signal PDF, allowing for
 3150 the reduction of observables to be determined in fits without loosing any sensitivity. Four
 3151 of these angular transformations, or ‘foldings’, were used to make first measurements of
 3152 $P'_{4,5,6,8}$ in the decays of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ [8]. Folding the PDF improves fit stability, and
 3153 is most advantageous when the signal yield is low, and the number of parameters to be
 3154 determined is high, due to for example, setting background angular parameters to vary in
 3155 the fit.

3156 The usage of folded PDFs was the nominal strategy for this analysis until greater control
 3157 and understanding of the background shapes and composition was achieved following
 3158 studies of the $K\pi e\mu$ data, and the decision was made to fix all background angular
 3159 parameters. The advantage in using folded PDFs is greatly reduced in this situation, as
 3160 the number of parameters to be determined is already low. This is confirmed by simulation
 3161 studies, which showed little difference between the unfolded, and folded PDFs in the
 3162 current configuration. Nevertheless, they will still be used to cross-check the nominal
 3163 measurement results.

3164 The angular transformations, as well as the folded PDFs, are given below.

3165 F.1 Folding for P'_4

3166 Transformations applied:

$$\begin{aligned}\phi &\rightarrow -\phi & \text{if } \phi < 0 \\ \phi &\rightarrow \pi - \phi & \text{if } \theta_l > \frac{\pi}{2} \\ \theta_l &\rightarrow \pi - \theta_l & \text{if } \theta_l > \frac{\pi}{2}\end{aligned}\tag{63}$$

3167 Folded PDF:

$$\begin{aligned}\frac{1}{d(\Gamma + \bar{\Gamma})/dq^2} \frac{d^4(\Gamma + \bar{\Gamma})}{dq^2 d\vec{\Omega}} &= (1 - F_S) \frac{9}{32\pi} \left[\frac{3}{4}(1 - F_L) \sin^2 \theta_K + F_L \cos^2 \theta_K \right. \\ &\quad + \frac{1}{4}(1 - F_L) \sin^2 \theta_K \cos 2\theta_l \\ &\quad - F_L \cos^2 \theta_K \cos 2\theta_l \\ &\quad + \frac{1}{2}(1 - F_L) P_1 \sin^2 \theta_K \sin^2 \theta_l \cos 2\phi \\ &\quad \left. + \sqrt{F_L(1 - F_L)} P'_4 \sin 2\theta_K \sin 2\theta_l \cos \phi \right]\end{aligned}\tag{64}$$

3168 Ranges for the angles:

$$\begin{aligned}\theta_K (\cos \theta_K) &\in [0, \pi] ([-1, 1]) \\ \theta_\ell (\cos \theta_\ell) &\in [0, \pi/2] ([0, 1]) \\ \phi &\in [0, \pi]\end{aligned}\tag{65}$$

₃₁₆₉ **F.2 Folding for P'_5**

₃₁₇₀ Transformations applied:

$$\begin{aligned}\phi &\rightarrow -\phi \quad \text{if } \phi < 0 \\ \theta_l &\rightarrow \pi - \theta_l \quad \text{if } \theta_l > \frac{\pi}{2}\end{aligned}\tag{66}$$

₃₁₇₁ Folded PDF:

$$\begin{aligned}\frac{1}{d(\Gamma + \bar{\Gamma})/dq^2} \frac{d^4(\Gamma + \bar{\Gamma})}{dq^2 d\vec{\Omega}} = (1 - F_S) \frac{9}{32\pi} &\left[\frac{3}{4}(1 - F_L) \sin^2 \theta_K + F_L \cos^2 \theta_K \right. \\ &+ \frac{1}{4}(1 - F_L) \sin^2 \theta_K \cos 2\theta_l \\ &- F_L \cos^2 \theta_K \cos 2\theta_l \\ &+ \frac{1}{2}(1 - F_L) P_1 \sin^2 \theta_K \sin^2 \theta_l \cos 2\phi \\ &\left. + \sqrt{F_L(1 - F_L)} P'_5 \sin 2\theta_K \sin \theta_l \cos \phi \right]\end{aligned}\tag{67}$$

₃₁₇₂ Ranges for the angles:

$$\begin{aligned}\theta_K (\cos \theta_K) &\in [0, \pi] ([-1, 1]) \\ \theta_\ell (\cos \theta_\ell) &\in [0, \pi/2] ([0, 1]) \\ \phi &\in [0, \pi]\end{aligned}\tag{68}$$

₃₁₇₃ **F.3 Folding for P'_6**

₃₁₇₄ Transformations applied:

$$\begin{aligned}\phi &\rightarrow \pi - \phi \quad \text{if } \phi > \frac{\pi}{2} \\ \phi &\rightarrow -\pi - \phi \quad \text{if } \phi < -\frac{\pi}{2} \\ \theta_l &\rightarrow \pi - \theta_l \quad \text{if } \theta_l > \frac{\pi}{2}\end{aligned}\tag{69}$$

₃₁₇₅ Folded PDF:

$$\begin{aligned}\frac{1}{d(\Gamma + \bar{\Gamma})/dq^2} \frac{d^4(\Gamma + \bar{\Gamma})}{dq^2 d\vec{\Omega}} = (1 - F_S) \frac{9}{32\pi} &\left[\frac{3}{4}(1 - F_L) \sin^2 \theta_K + F_L \cos^2 \theta_K \right. \\ &+ \frac{1}{4}(1 - F_L) \sin^2 \theta_K \cos 2\theta_l \\ &- F_L \cos^2 \theta_K \cos 2\theta_l \\ &+ \frac{1}{2}(1 - F_L) P_1 \sin^2 \theta_K \sin^2 \theta_l \cos 2\phi \\ &\left. + \sqrt{F_L(1 - F_L)} P'_6 \sin 2\theta_K \sin \theta_l \sin \phi \right]\end{aligned}\tag{70}$$

₃₁₇₆ Ranges for the angles:

$$\begin{aligned}\theta_K (\cos\theta_K) &\in [0, \pi] ([-1, 1]) \\ \theta_\ell (\cos\theta_\ell) &\in [0, \pi/2] ([0, 1]) \\ \phi &\in [-\pi/2, \pi/2]\end{aligned}\tag{71}$$

₃₁₇₇ F.4 Folding for P'_8

₃₁₇₈ Transformations applied:

$$\begin{aligned}\phi &\rightarrow \pi - \phi \quad \text{if } \phi > \frac{\pi}{2} \\ \phi &\rightarrow -\pi - \phi \quad \text{if } \phi < -\frac{\pi}{2} \\ \theta_l &\rightarrow \pi - \theta_l \quad \text{if } \theta_l > \frac{\pi}{2} \\ \theta_K &\rightarrow \pi - \theta_K \quad \text{if } \theta_l > \frac{\pi}{2}\end{aligned}\tag{72}$$

₃₁₇₉ Folded PDF:

$$\begin{aligned}\frac{1}{d(\Gamma + \bar{\Gamma})/dq^2} \frac{d^4(\Gamma + \bar{\Gamma})}{dq^2 d\vec{\Omega}} = (1 - F_S) \frac{9}{32\pi} \left[\frac{3}{4}(1 - F_L) \sin^2 \theta_K + F_L \cos^2 \theta_K \right. \\ + \frac{1}{4}(1 - F_L) \sin^2 \theta_K \cos 2\theta_l \\ - F_L \cos^2 \theta_K \cos 2\theta_l \\ + \frac{1}{2}(1 - F_L) P_1 \sin^2 \theta_K \sin^2 \theta_l \cos 2\phi \\ \left. + \sqrt{F_L(1 - F_L)} P'_8 \sin 2\theta_K \sin 2\theta_l \sin \phi \right]\end{aligned}\tag{73}$$

₃₁₈₀ Ranges for the angles:

$$\begin{aligned}\theta_K (\cos\theta_K) &\in [0, \pi] ([-1, 1]) \\ \theta_\ell (\cos\theta_\ell) &\in [0, \pi/2] ([0, 1]) \\ \phi &\in [-\pi/2, \pi/2]\end{aligned}\tag{74}$$

₃₁₈₁ F.5 Other observables

₃₁₈₂ Transformation applied:

$$\phi \rightarrow \pi + \phi \quad \text{if } \phi < 0\tag{75}$$

₃₁₈₃ Folded PDF:

$$\frac{1}{d(\Gamma + \bar{\Gamma})/dq^2} \frac{d^4(\Gamma + \bar{\Gamma})}{dq^2 d\vec{\Omega}} = (1 - F_S) \frac{9}{32\pi} \left[\begin{aligned} & \frac{3}{4}(1 - F_L) \sin^2 \theta_K + F_L \cos^2 \theta_K \\ & + \frac{1}{4}(1 - F_L) \sin^2 \theta_K \cos 2\theta_l \\ & - F_L \cos^2 \theta_K \cos 2\theta_l \\ & + \frac{1}{2}(1 - F_L) P_1 \sin^2 \theta_K \sin^2 \theta_l \cos 2\phi \\ & + 2(1 - F_L) P_2 \sin^2 \theta_K \cos \theta_l \\ & - (1 - F_L) P_3 \sin^2 \theta_K \sin^2 \theta_l \sin 2\phi \end{aligned} \right] \quad (76)$$

₃₁₈₄ Ranges for the angles:

$$\begin{aligned} \theta_K (\cos \theta_K) &\in [0, \pi] ([-1, 1]) \\ \theta_\ell (\cos \theta_\ell) &\in [0, \pi] ([-1, 1]) \\ \phi &\in [0, \pi] \end{aligned} \quad (77)$$

3185 **G Additional acceptance/resolution information**

3186 **G.1 Resolution**

3187 Figures 158, 159, 160, 161, 162 and 163 depict the resolution distributions for q_c^2 and angular distributions in 2011, 2012 and 2015.

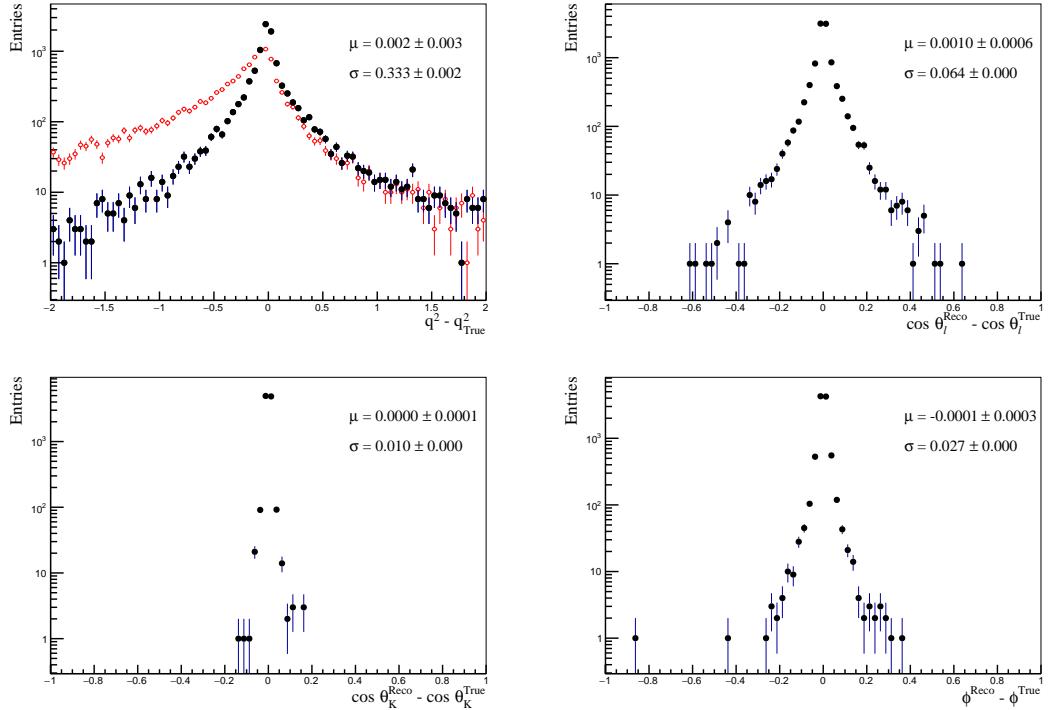


Figure 158: Resolution distributions defined as the difference between reconstructed and generator level variables in 2011 conditions for (top left) q^2 , (top right) $\cos \theta_\ell$, (bottom left) $\cos \theta_K$ and (bottom right) ϕ . Note that in the case of the q^2 distribution both (black) q_c^2 and (red) normal q^2 are overlaid for comparison.

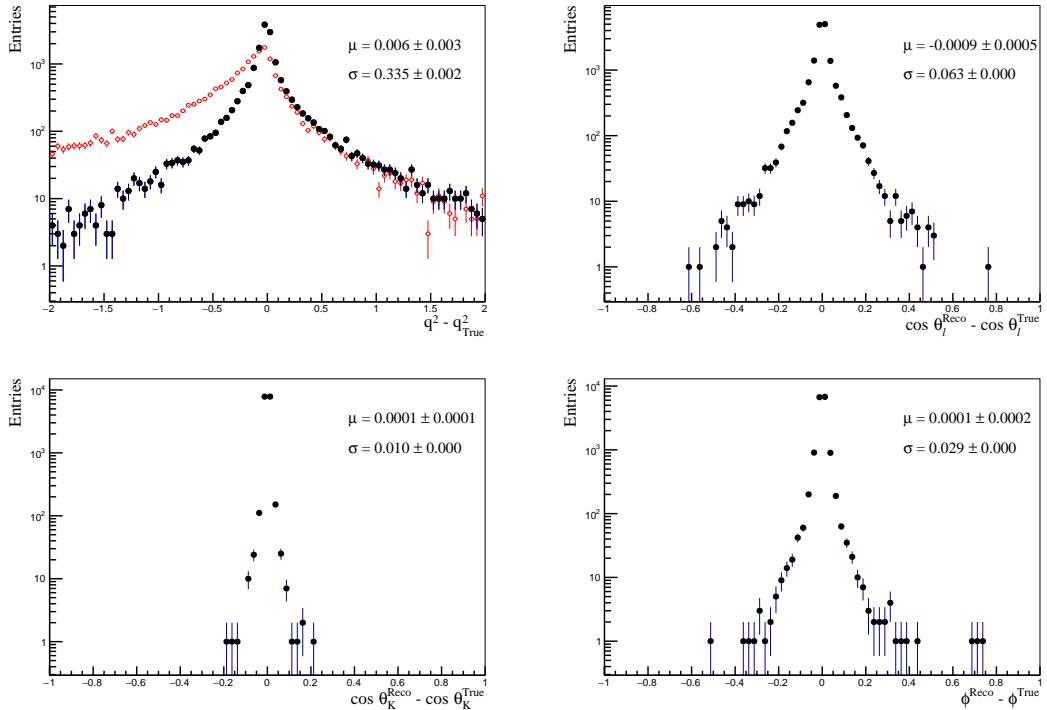


Figure 159: Resolution distributions defined as the difference between reconstructed and generator level variables in 2012 conditions for (top left) q^2 , (top right) $\cos \theta_\ell$, (bottom left) $\cos \theta_K$ and (bottom right) ϕ . Note that in the case of the q^2 distribution both (black) q_c^2 and (red) normal q^2 are overlaid for comparison.

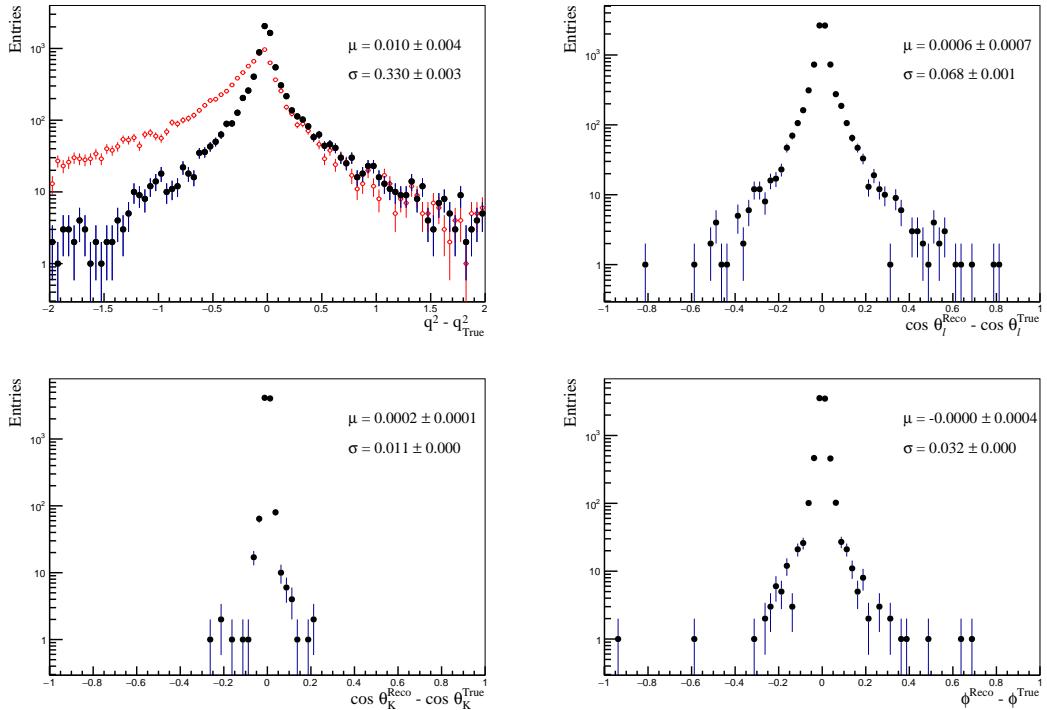


Figure 160: Resolution distributions defined as the difference between reconstructed and generator level variables in 2015 conditions for (top left) q^2 , (top right) $\cos \theta_\ell$, (bottom left) $\cos \theta_K$ and (bottom right) ϕ . Note that in the case of the q^2 distribution both (black) q_c^2 and (red) normal q^2 are overlaid for comparison.

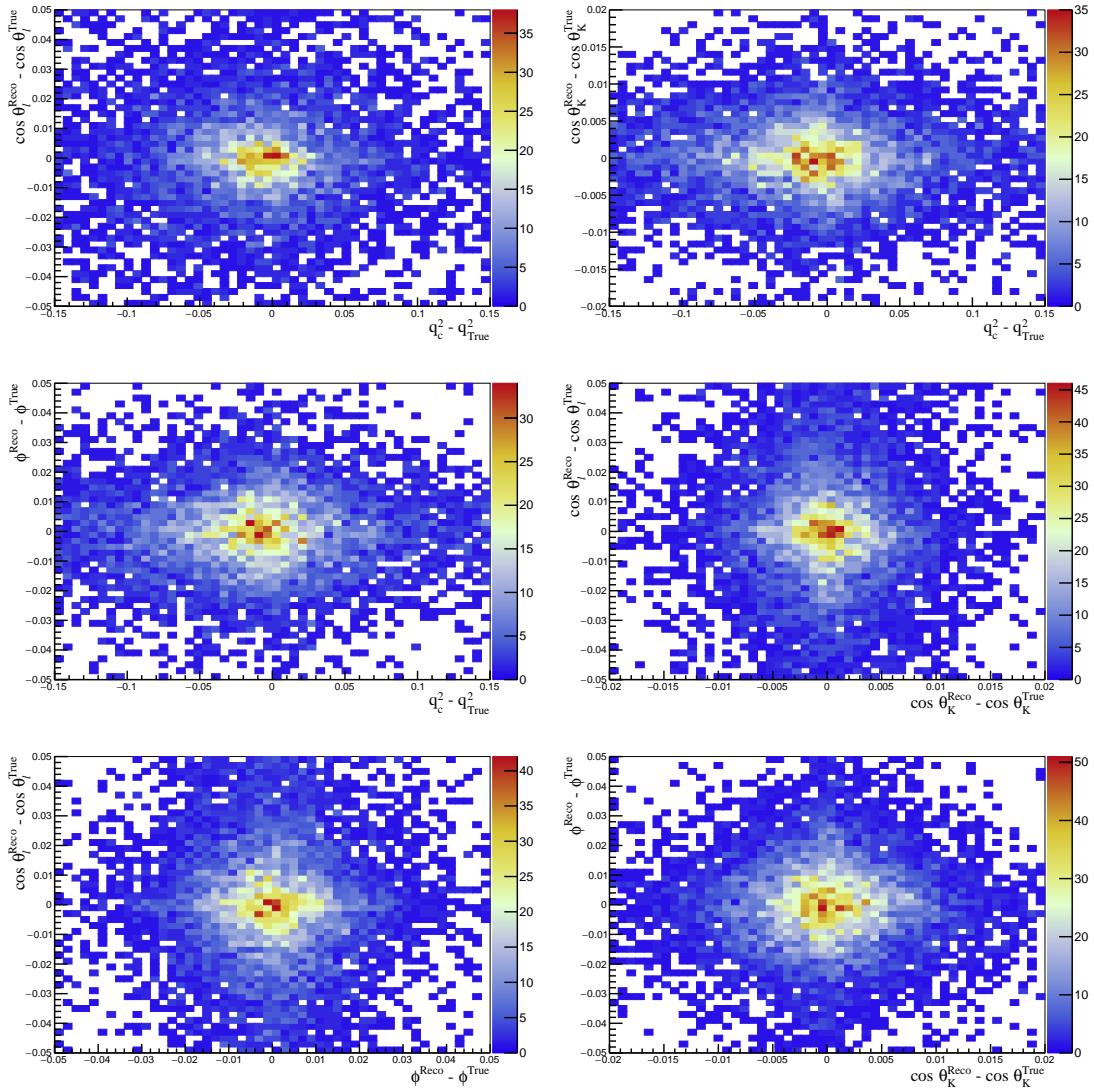


Figure 161: Two-dimensional resolution distributions with 2011 conditions defined as the difference between reconstructed and generator level variables for (top left) q^2 , (top right) $\cos \theta_\ell$, (bottom left) $\cos \theta_K$ and (bottom right) ϕ .

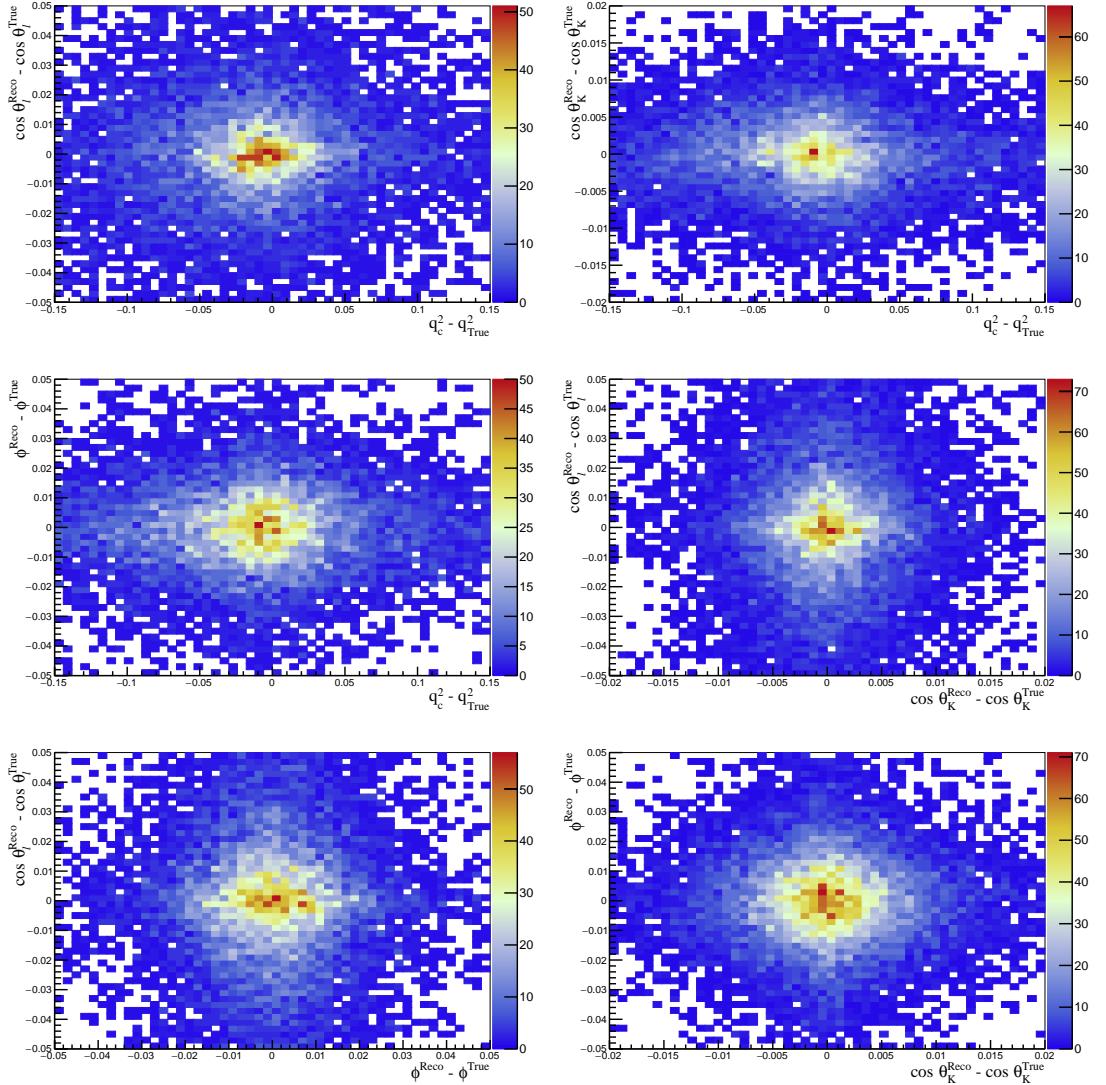


Figure 162: Two-dimensional resolution distributions with 2012 conditions defined as the difference between reconstructed and generator level variables for (top left) q^2 , (top right) $\cos \theta_\ell$, (bottom left) $\cos \theta_K$ and (bottom right) ϕ .

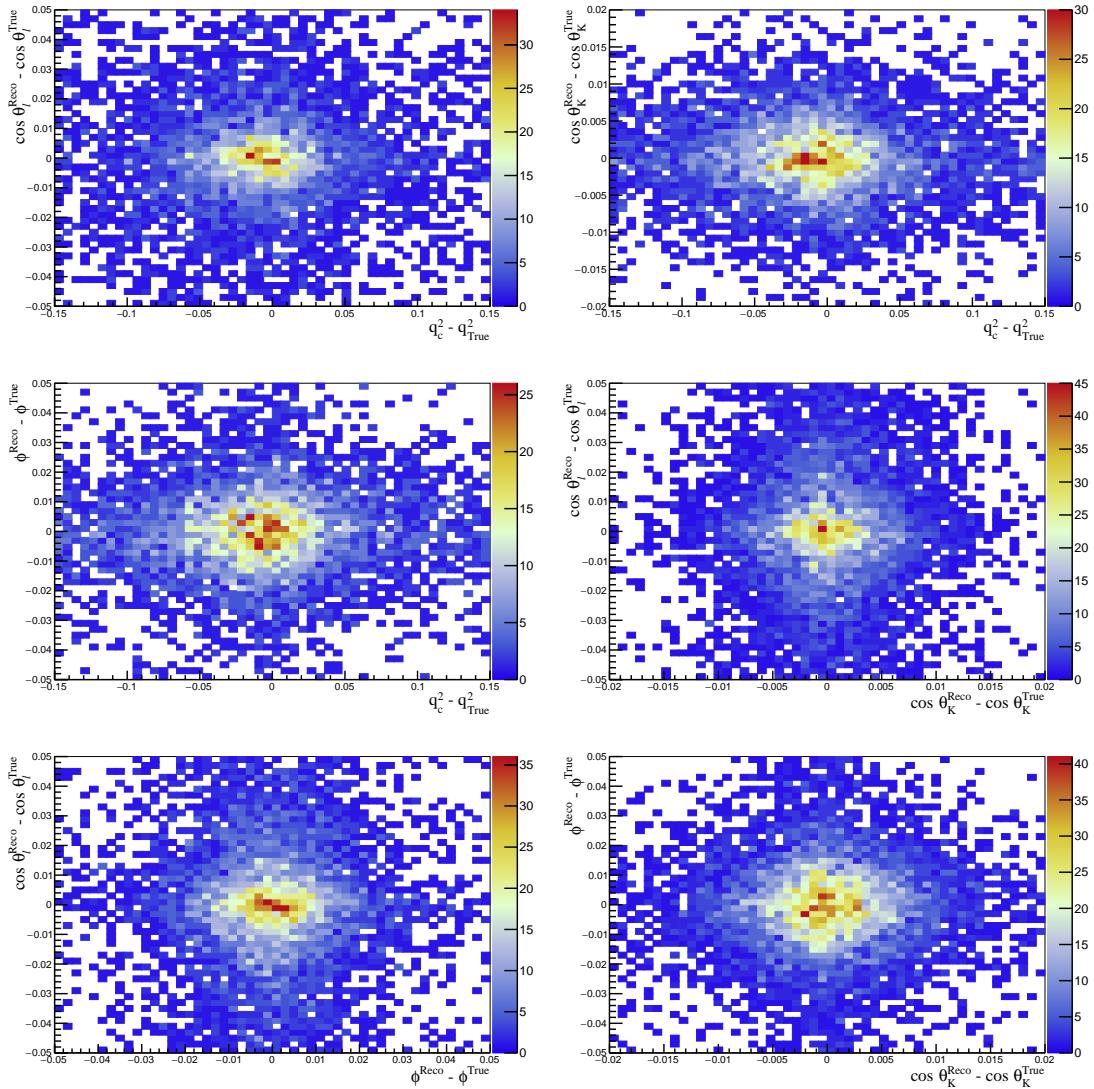


Figure 163: Two-dimensional resolution distributions with 2015 conditions defined as the difference between reconstructed and generator level variables for (top left) q^2 , (top right) $\cos \theta_\ell$, (bottom left) $\cos \theta_K$ and (bottom right) ϕ .

Table 59: S_i -basis observable fitted results and corresponding differences with respect to generator level values obtained with and without the removal of events with large weights.

	No removal		With removal	
F_L	0.728 ± 0.009	-0.0210	0.712 ± 0.003	-0.0051
S_3	-0.002 ± 0.003	-0.0075	-0.003 ± 0.003	-0.0068
S_4	-0.143 ± 0.005	-0.0057	-0.149 ± 0.005	0.0004
S_5	-0.216 ± 0.005	0.0119	-0.213 ± 0.004	0.0098
A_{FB}	0.022 ± 0.003	-0.0119	0.022 ± 0.003	-0.0126
S_7	0.009 ± 0.005	0.0038	0.007 ± 0.004	0.0058
S_8	0.004 ± 0.005	-0.0009	0.005 ± 0.004	-0.0024
S_9	-0.003 ± 0.003	0.0038	-0.002 ± 0.003	0.0036

3189 H Weighted fit features

3190 The number of events in a given sample, k , can be described by a Poisson distribution
 3191 with expectation value and variance of $E(k) = V(k) = \mu$, where μ denotes the predicted
 3192 number of events. When these events are weighted, the sum of the weighted events can
 3193 be described by a compound Poisson distribution, which, in the limit of $\mu \rightarrow \infty$, can be
 3194 described by a Gaussian distribution with [96]

$$E(k_w) \equiv E\left(\sum_i (w_i)\right) = \mu E(w) \quad (78)$$

$$V(k_w) = \mu E(w^2). \quad (79)$$

3195 The characteristics of the expectation value and variance of the sum of weighted events
 3196 lead to two features that are relevant for the current analysis. These are explained in the
 3197 following sections.

3199 H.1 Large weights

3200 When per-event weights are calculated from the inverse acceptance function (or the inverse
 3201 generator-level parametrisation), some of the events may receive relatively large weights,
 3202 as introduced in Sec. 6.3. These instances can occur purely due to low statistics and/or
 3203 a correction function that is not smooth (large corrections required in certain regions).
 3204 Figure 164 show the instance of large weights and their corresponding removal in an
 3205 extreme case where the dataset is split into 2011-2012 and L0 categories (smallest samples).
 3206 The marginal impact of the removal of these events on the angular observables can be
 3207 visualised in Table 59.

3208 Another simplified example can be useful to visualise this behaviour. Let us produce
 3209 a uniform toy distribution of a generic quantity ranging from zero to ten. These are
 3210 made non-uniform by an *acceptance effect*, where all *events* with values of zero to nine are
 3211 accepted, while only one percent of events with values between nine and ten are taken

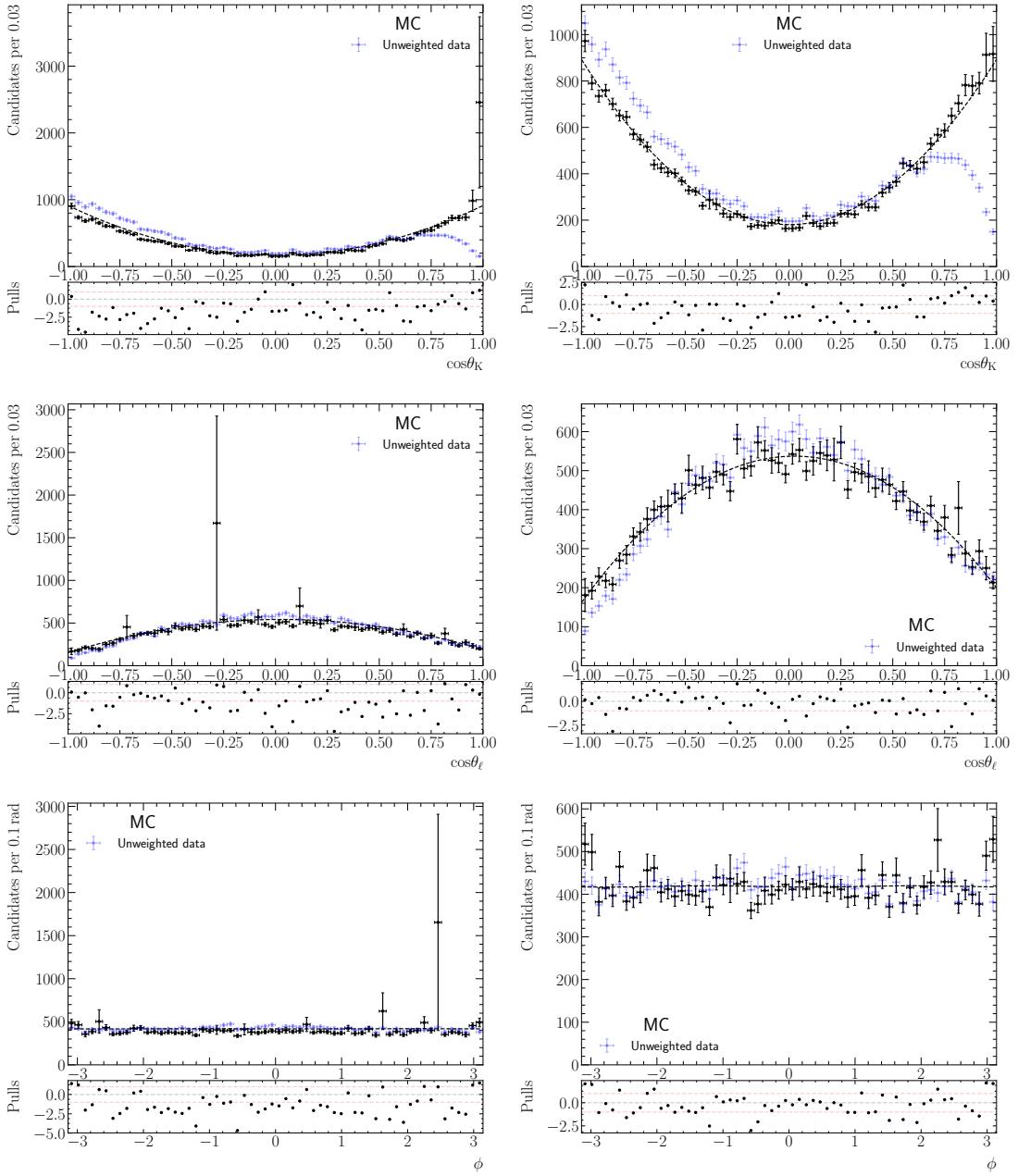


Figure 164: Angular distributions for Run-I rare mode simulation with weights taken from the FLATQ2 acceptance functions parametrised, separately for each year and L0 category, without (left) and with (right) the removal of large weights (0.03% of the full post-selection sample). The blue-dotted points correspond to the unweighted simulation whereas the black-dotted ones refer to the weighted simulation.

³²¹² ($\epsilon = 0.01$). The correction weights in this case are chosen to be unity for all events between
³²¹³ zero and nine (no correction), and a hundred ($1/\epsilon = 1/0.01$) for events with values between

3214 nine and ten. When $N_{\text{gen}} = 1000$ events are generated in total and split into ten bins
 3215 (Fig. 165), each of which spans one unit, the last bin covering the range of nine to ten
 3216 will be affected by the ‘acceptance’. The standard deviation of the sum of weights in the
 3217 last bin, σ_{bin} , is given by $\sqrt{((N_{\text{gen}}/N_{\text{bins}}) \times \epsilon)E(w^2)}$, where $E(w^2) = 100^2$. If σ_{bin} is large
 3218 compared to the expected number of events, *large weights* can randomly appear. When
 3219 the statistics is very low a threshold can be reached when the retrieval of the generator
 3220 level distribution is no longer possible, as shown on Fig. 166. Increasing statistics (increase
 3221 $N_{\text{gen}}/N_{\text{bins}}$) and/or decreasing the magnitude of the correction (ϵ value) can reduce the
 3222 occurrence of large weights (Fig. 167). The appearance of large weights at low statistics
 3223 can also be seen in pseudoexperiments generated and corrected with the same acceptance
 3224 function, as illustrated in Fig. 168. This is the reasoning behind the choice of using the
 3225 smoother FLATQ2 acceptance as the nominal acceptance.

3226 Note that large weights can also (and perhaps more frequently) occur due to the
 3227 mismatch between the correction function used and the true underlying function. An
 3228 example is if the true acceptance function has a higher value in a region where the applied
 3229 function is close to zero. The ‘mechanism’ here is the same. With increased statistics,
 3230 large weights from mismatched functions also melt away to leave smooth (but incorrect)
 3231 weighted distributions.

3232 **H.2 Weighted fractions**

3233 The expectation value of the sum of weighted events in the post-selection sample can be
 3234 written as

$$E(k_w^{\text{tot}}) = E(k_w^S + k_w^B) = \mu_S E(w_S) + \mu_B E(w_B) , \quad (80)$$

3235 where the two summations are over all signal and background events, respectively, and
 3236 $E(w_S) \neq E(w_B)$ in general, as the two components populate the phase space differently.
 3237 The signal and background fractions determined from the fit are given by $k_w^{S/B}/k_w^{\text{tot}}$. From
 3238 this it can be seen that the results of the fit does not give indications of the ‘true’ fractions of
 3239 signal and background events present in the sample ($\mu_{S/B}/\mu_{\text{tot}}$), which are in fact unknown,
 3240 but are rather some ‘post-correction’ fractions. Those can in principle correspond to the
 3241 fractions of the generator level sample if the matching correction functions are used for all
 3242 backgrounds, and the correction functions are proper efficiency functions (determined using
 3243 matching generator level and post-selection samples). This is not the case for the current
 3244 measurement. The correction functions used correct for multiple effects simultaneously, and
 3245 is only (approximately) correct for the signal (will be generally incorrect for backgrounds).
 3246 Therefore the signal and background fractions obtained should not be used to estimate
 3247 branching fractions, and care should be taken when the values are compared against those
 3248 obtained from unweighted fits.

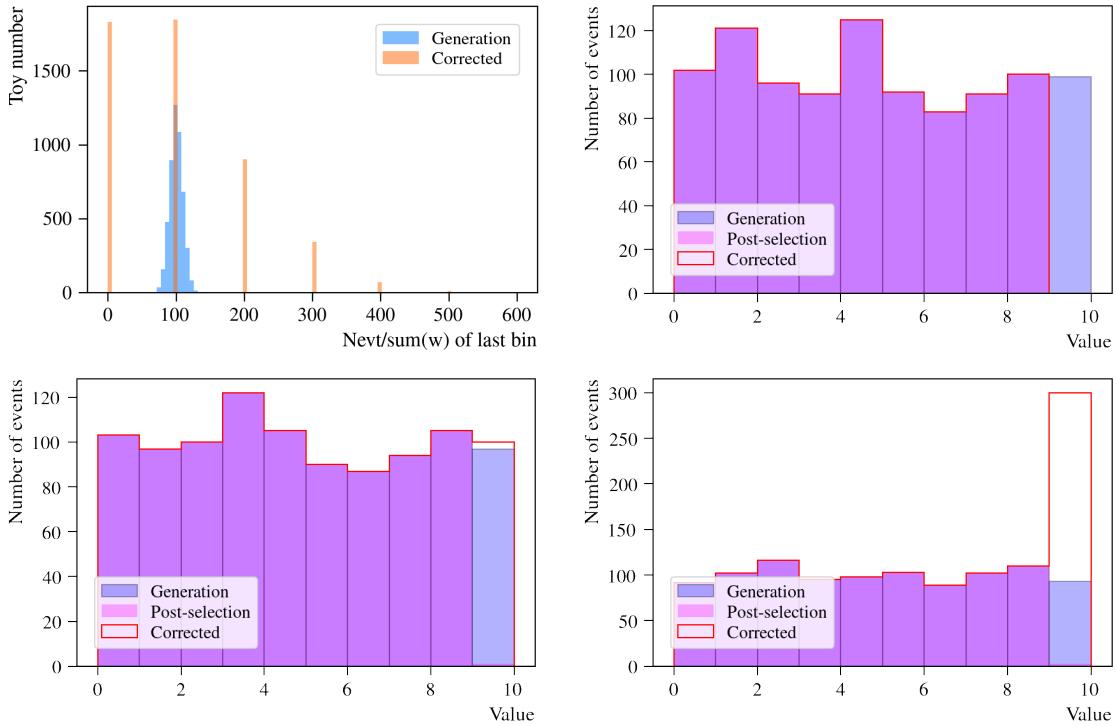


Figure 165: Simplified example featuring samples of 1000 events each drawn from a uniform distribution with values ranging from zero to ten. An acceptance effect is applied to ‘events’ with values between nine and ten (last bin). This is achieved by randomly retaining one percent of events from the initial (generation) sample. Events with other values are unaffected. Correction weights (100 for affected events, unity/no correction for the rest) are applied to the post-acceptance sample retrieve the generation distribution. The top left figure shows the distribution of the number of events for the initial sample and the distribution of the sum of weights in the last bin. The other three plots show examples of under, adequate and over-correction (‘large-weights’). Those three examples are chosen such that the sum of weights are zero, 100 and 300 (two times the standard deviation from mean), respectively.

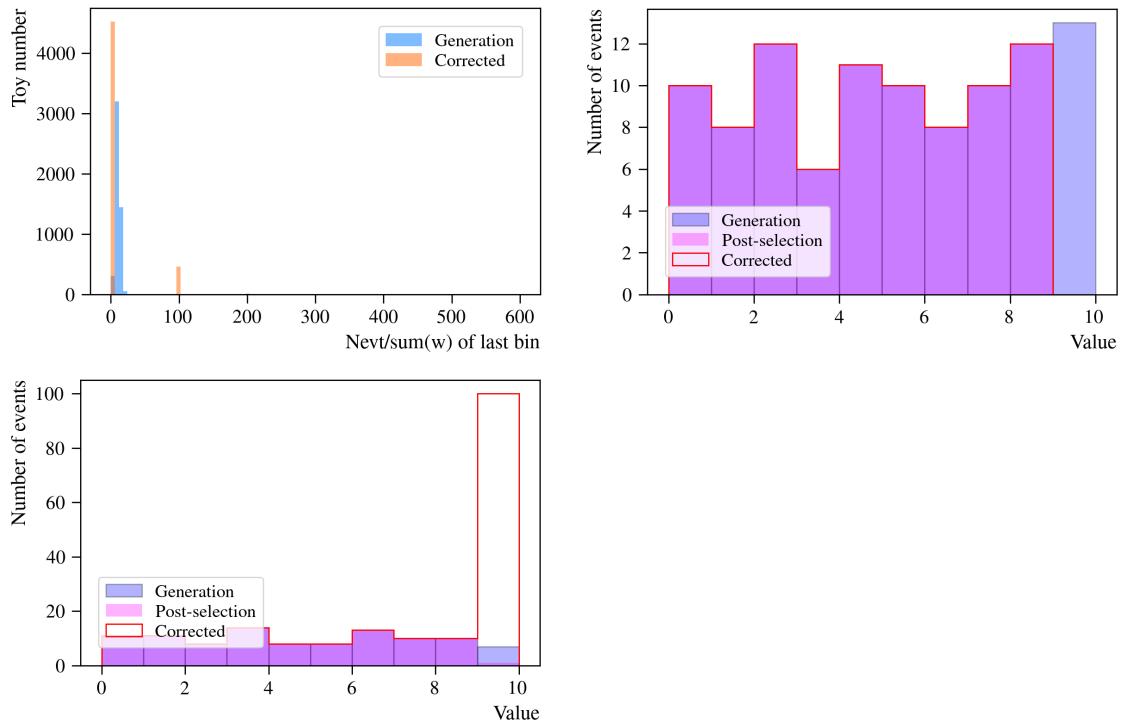


Figure 166: Simplified example featuring samples of 100 events each drawn from a uniform distribution with values ranging from zero to ten. An acceptance effect is applied to ‘events’ with values between nine and ten (last bin). The top left figure shows the distribution of the number of events for the initial sample and the distribution of the sum of weights in the last bin. In this case, the mean number of events in the last bin in the generation sample (10) is lower than the correction weight ($1/\epsilon = 100$), which means the correct retrieval of the generation distribution is not possible.

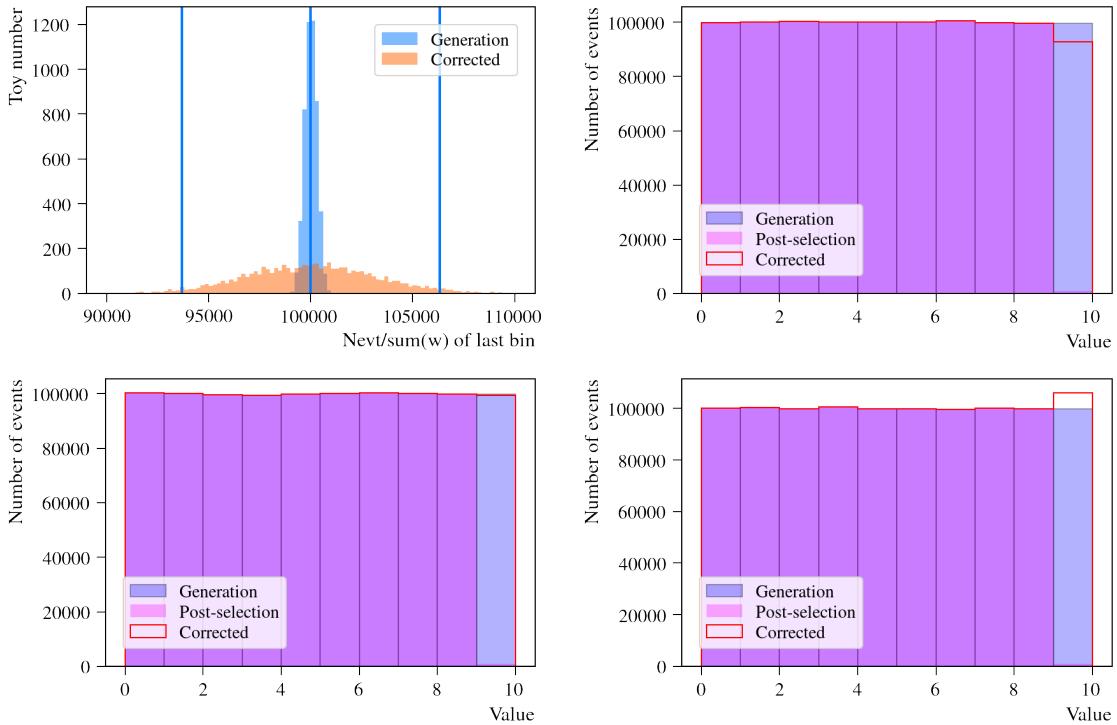


Figure 167: Simplified example featuring samples of 1000000 events each drawn from a uniform distribution with values ranging from zero to ten. An acceptance effect is applied to ‘events’ with values between nine and ten (last bin). The top left figure shows the distribution of the number of events for the initial sample and the distribution of the sum of weights in the last bin. The other three plots show examples of under, adequate and over-correction. Those three examples are chosen such that the sum of weights are approximately equal to the mean value (100000) minus two times the standard deviation (3162), the mean value, and the mean value plus two times the standard deviation, respectively.

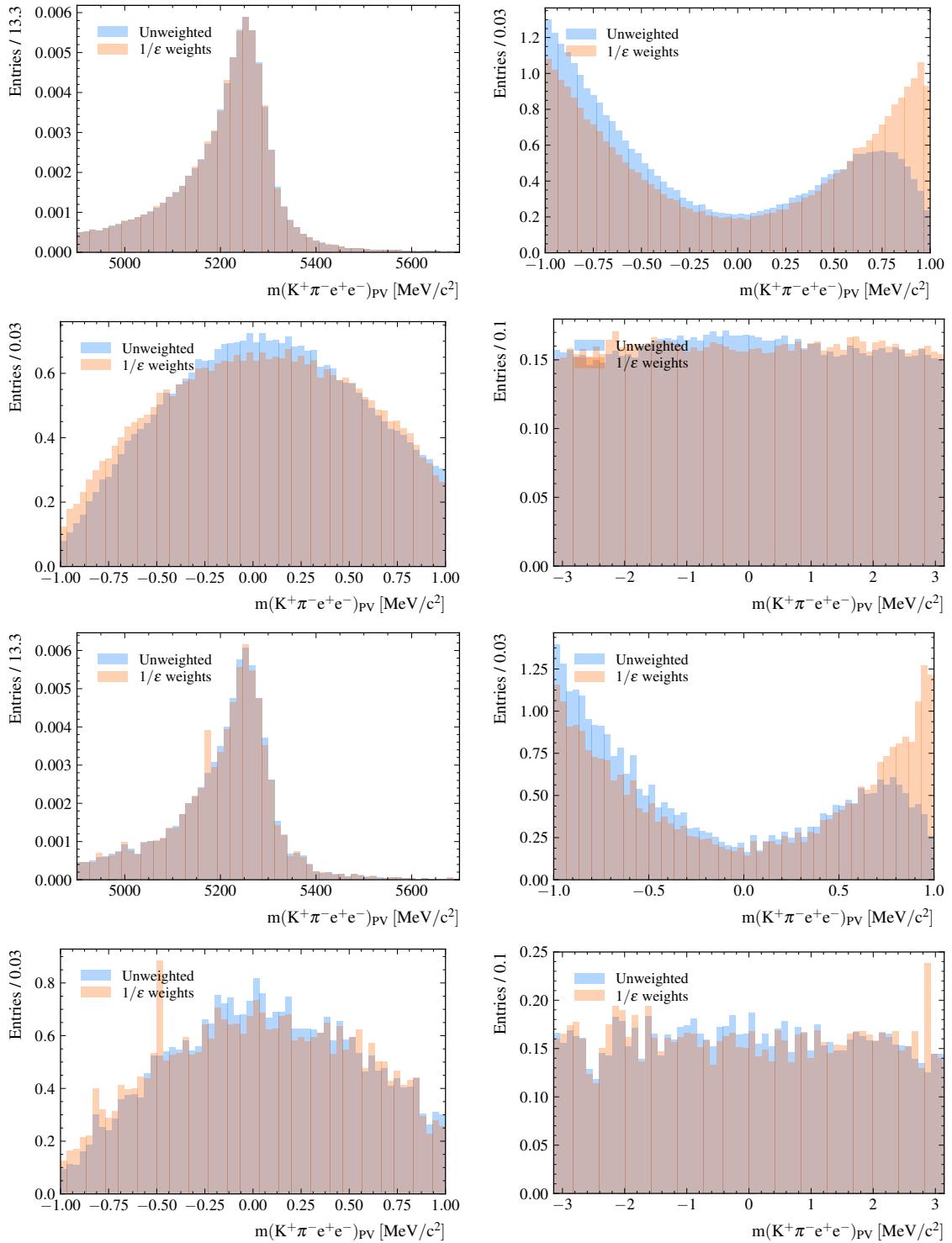


Figure 168: Mass and angular distributions of a large statistics pseudoexperiment (200,000 events) generated with acceptance, with and without correction weights calculated from the same acceptance function (top four figures), and the corresponding distributions of 10,000 events taken at random from the large sample. ‘Large weights’ (associated with two events) can be seen in subsample with lower statistics.

3249 I Double-semileptonic cascade background

3250 I.1 A rough comparison of DSL in $K\pi e\mu$ and $K\pi ee$ data

3251 Even though the simulation for $B^0 \rightarrow D^- (\rightarrow K^{*0} e^- \bar{\nu}_e) e^+ \nu_e$ and $B^0 \rightarrow D^- (\rightarrow$
3252 $K^{*0} \mu^- [e^-] \bar{\nu}) e^+ [\mu^+] \nu$ points to the fact that the angular shapes are minimally effected
3253 by the presence of an electron or a muon in the final state, it is important to check the
3254 portability of the parametrization obtained from $K\pi e\mu$ data to $K\pi ee$ data.

3255 It is possible to do so by looking in both datasets at the mass window $m_B \in$
3256 $[3800, 4300]\text{MeV}/c^2$ after a tight BDT cut ($BDT > 0.99$). Here we expect:

- 3257 • only double-semileptonic cascade decays and combinatorial backgrounds for $K\pi e\mu$
3258 data,
- 3259 • double-semileptonic cascade decays, combinatorial, $B^0 \rightarrow K^{*0} J/\psi$ leakage and
3260 partially reconstructed for $K\pi ee$ data.

3261 A comparison between the $\cos \theta_\ell$ distributions is shown in Fig. 169(a) and allows for three
3262 comments:

- 3263 • the qualitative behavior of the two distributions is similar as expected, since the
3264 dominant contribution should be given by DSL,
- 3265 • both samples present a second peak for values of $\cos \theta_\ell$ close to -1 that is not
3266 reproduced in simulation,
- 3267 • the relative abundance of this peak is not the same between the two samples, however
3268 additional sources of background in the $K\pi ee$ could be responsible for such differences
3269 (e.g. J/ψ -leakage or partially reconstructed decays).

3270 The presence of a second peak in data but not in simulation seems to point to a recon-
3271 struction effect. One possible explanation for this discrepancy is a contribution from
3272 combinatorial-like components in data, such as the combination of a random K^* with
3273 a true DSL event, which could show up as a symmetric structure with peaks at both -1
3274 and 1. This would imply two things: first such combinatorial-like contribution could be
3275 probably merged in the standard parametrization of the combinatorial, secondly, due to
3276 the fact that also the $B^0 \rightarrow K^{*0} J/\psi$ leakage and partially reconstructed are symmetric in
3277 $\cos \theta_\ell$, we could manage to compare the asymmetric part of the DSL distribution (that
3278 is the dominant) in data between the $K\pi e\mu$ and the $K\pi ee$ dataset, by subtracting the
3279 left-hand side of the $\cos \theta_\ell$ distribution from the right-hand side. This attempt is shown in
3280 Fig. 169(b), where the agreement shown between the two is very good.

3281 In conclusion it seems that, where it is possible to have a meaningful comparison, the
3282 agreement between the DSL shapes obtained from $K\pi e\mu$ data to $K\pi ee$ is such that no
3283 differences are expected when using a parametrization of the DSL background from $K\pi e\mu$
3284 as a proxy of the DSL in $K\pi ee$ data.

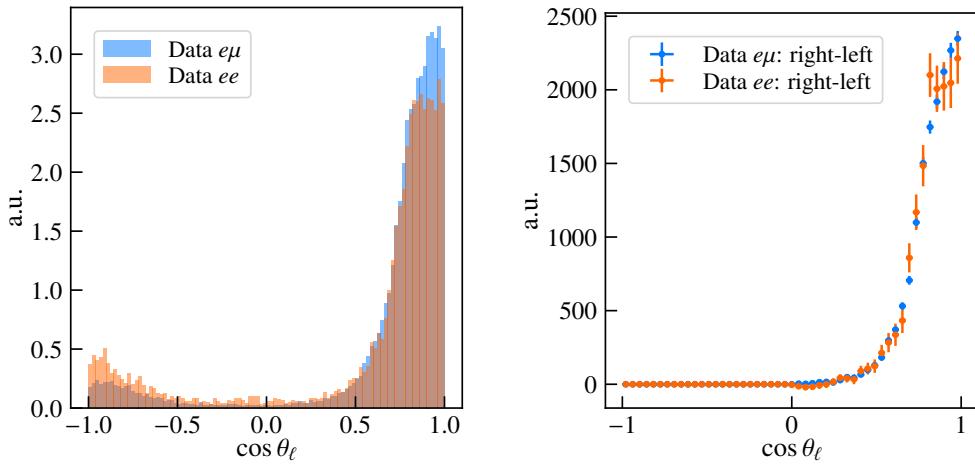


Figure 169: Comparision between the $\cos\theta_\ell$ distribution of DSL simulations and data (left) in the DSL-enriched mass region with tight cut on the combinatorial MVA after the subtraction of the symmetric component in the data (right).

I.2 A rough comparison of the DSL in simulation and in $K\pi e\mu$ data

By simply superimposing the distribution of the $B^0 \rightarrow D^- (\rightarrow K^{*0} e^- \bar{\nu}) e^+ \nu$ simulation on top of the $K\pi e\mu$ data sample in $m_B \in [4900, 5500]$ with a BDT cut $BDT > 0.99$ as shown in Fig. 170(a), is possible to notice two main differences:

- the data presents a second peak for values of $\cos\theta_\ell$ close to -1 that is not reproduced in simulation as discussed above,
- the width of the peaking structure around +1 seems bigger for data than for simulation and the agreement does not improve if only the asymmetric component is used in the comparison (see Fig. 170(b)).

A difference such as the latter between data and simulation could be explained by the presence of events from other DSL decays not included in the single-component simulation, such as $B^0 \rightarrow D^*(2010) e^- \bar{\nu}$, which may modify the shape and width of the main peak. However, since this difference between data and mc is not understood, it raises a question on how much a veto completely based on simulation can be trusted (especially in $\cos\theta_\ell$). For this reason, together with the impact that such a veto would have on the acceptance, a direct parametrization of this background is preferred to a veto. In conclusion, due to clear limitations of the current simulation, the data-driven approach is taken to be the preferred method.

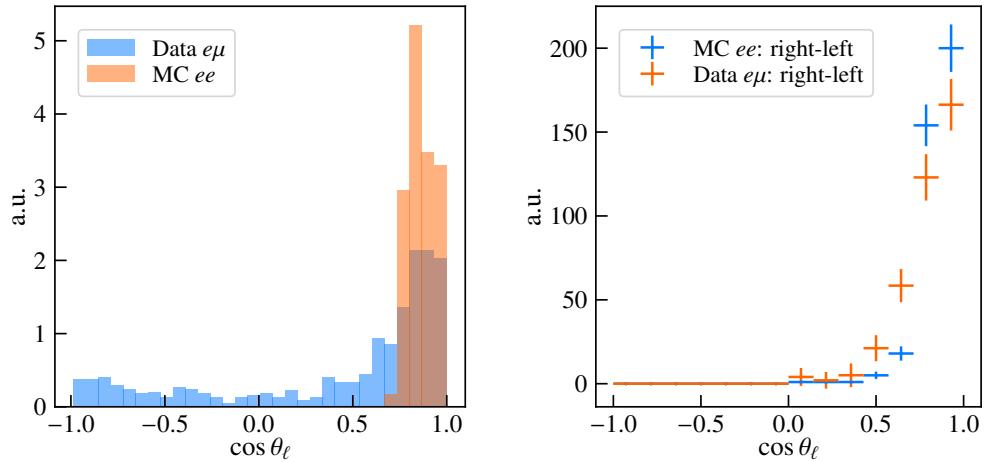


Figure 170: Comparision between the $\cos \theta_\ell$ distribution of DSL simulations (left) and data in the DSL-enriched mass region with tight cut on the combinatorial MVA (right) after the subtraction of the symmetric component in the data. The difference in the width of main peak persists.

3304 J Control mode validation

3305 This Appendix contains additional information and outputs of the control mode fits
 3306 discussed in Section 9, and is organised analogously.

3307 J.1 Main validation fit

3308 Most of the information and results associated with the main validation fit are reported in
 3309 Section 9. Table 60 shows the additional nuisance parameters of the fit.

	Result
$f_{\text{sig}}^{\text{Run-I}}$	0.9748 ± 0.0018
$f_{\text{sig}}^{\text{Run-II}}$	0.9764 ± 0.0015
$\delta_{\mu}^{\text{Run-I}}$	-8.6 ± 0.5
$\delta_{\mu}^{\text{Run-II}}$	-19.1 ± 0.4
$\delta_{\sigma}^{\text{Run-I}}$	1.101 ± 0.008
$\delta_{\sigma}^{\text{Run-II}}$	1.066 ± 0.006
$\lambda^{\text{Run-I}}$	-0.00437 ± 0.00028
$\lambda^{\text{Run-II}}$	-0.00418 ± 0.00026
$c_1^K; \text{Comb Run-I}$	0.06 ± 0.13
$c_1^K; \text{Comb Run-II}$	0.04 ± 0.14
$c_2^K; \text{Comb Run-I}$	0.50 ± 0.09
$c_2^K; \text{Comb Run-II}$	0.38 ± 0.10
$c_1^{\ell}; \text{Comb Run-I}$	0.68 ± 0.13
$c_1^{\ell}; \text{Comb Run-II}$	0.75 ± 0.17
$c_2^{\ell}; \text{Comb Run-I}$	0.73 ± 0.11
$c_2^{\ell}; \text{Comb Run-II}$	0.73 ± 0.11
$c_1^{\phi}; \text{Comb Run-I}$	-0.19 ± 0.13
$c_1^{\phi}; \text{Comb Run-II}$	0.02 ± 0.14
$c_2^{\phi}; \text{Comb Run-I}$	0.16 ± 0.13
$c_2^{\phi}; \text{Comb Run-II}$	0.13 ± 0.15

Table 60: Nuisance parameters of the main control mode angular fit.

3310 **J.2 Subsamples results**

3311 Additional information related to fits to sub-samples discussed in Section 9.4.1 are given
3312 below.

3313 **J.2.1 Separating by Run**

3314 Figure 172 shows the results of the the separate fits to Run-I and Run-II samples. Table 61
3315 contains the nuisance parameter results.

	Run-I		Run-II	
f_{sig}	0.9749	± 0.0018	0.9763	± 0.0015
δ_μ	-8.6	± 0.5	-19.1	± 0.4
δ_σ	1.101	± 0.008	1.066	± 0.006
λ_{Comb}	-0.00441	± 0.00029	-0.00415	± 0.00026
$c_{1; \text{Comb}}^K$	0.05	± 0.13	0.05	± 0.14
$c_{2; \text{Comb}}^K$	0.52	± 0.09	0.37	± 0.10
$c_{1; \text{Comb}}^\ell$	0.69	± 0.14	0.73	± 0.17
$c_{2; \text{Comb}}^\ell$	0.71	± 0.11	0.74	± 0.11
$c_{1; \text{Comb}}^\phi$	-0.21	± 0.13	0.04	± 0.14
$c_{2; \text{Comb}}^\phi$	0.16	± 0.13	0.13	± 0.15

Table 61: Nuisance parameters of the separate fits to Run-I and Run-II samples.

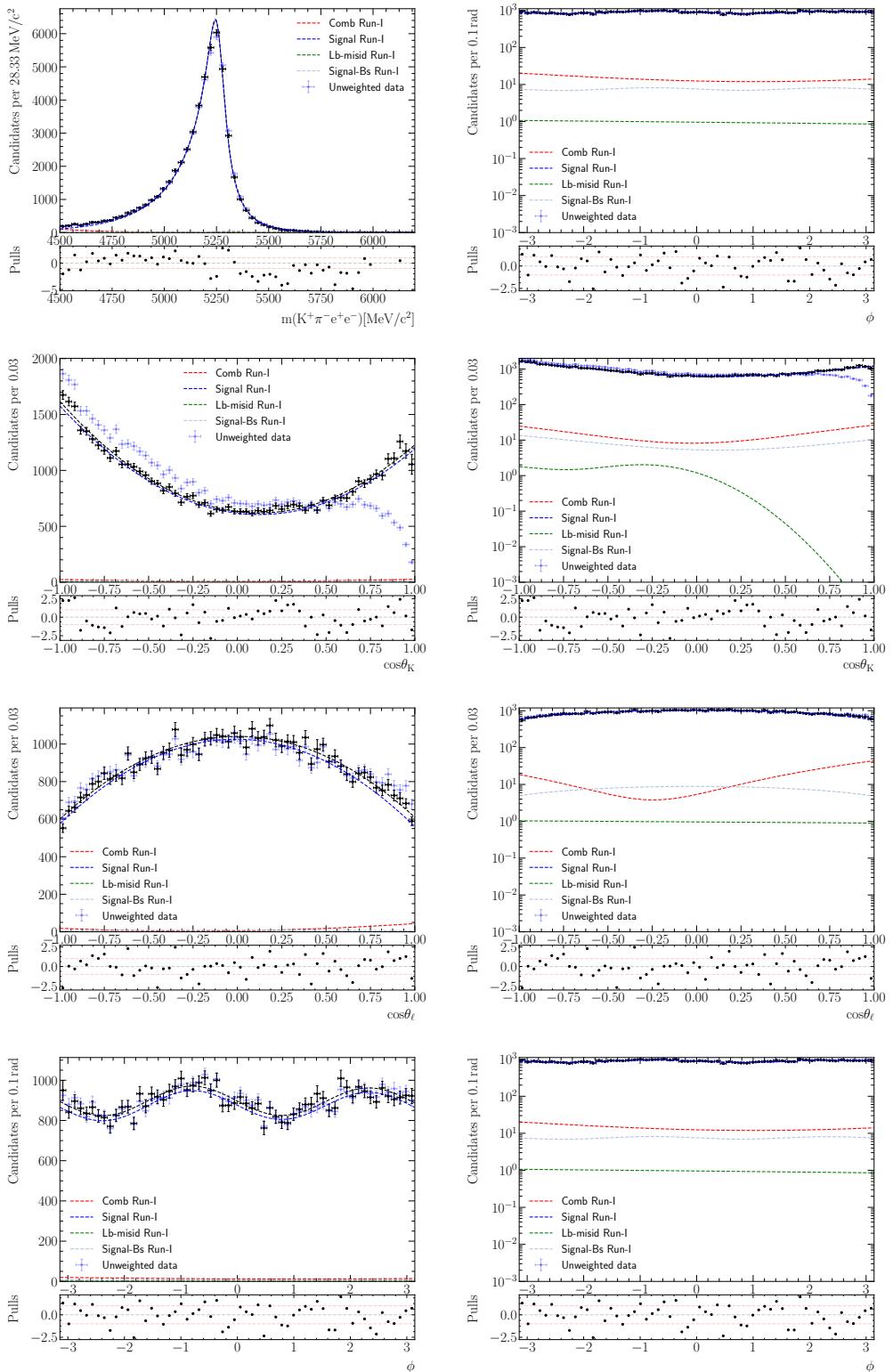


Figure 171: Result of fit to Run-I samples. The fit strategy used is the same as that one used for the nominal combinatorial only fit.

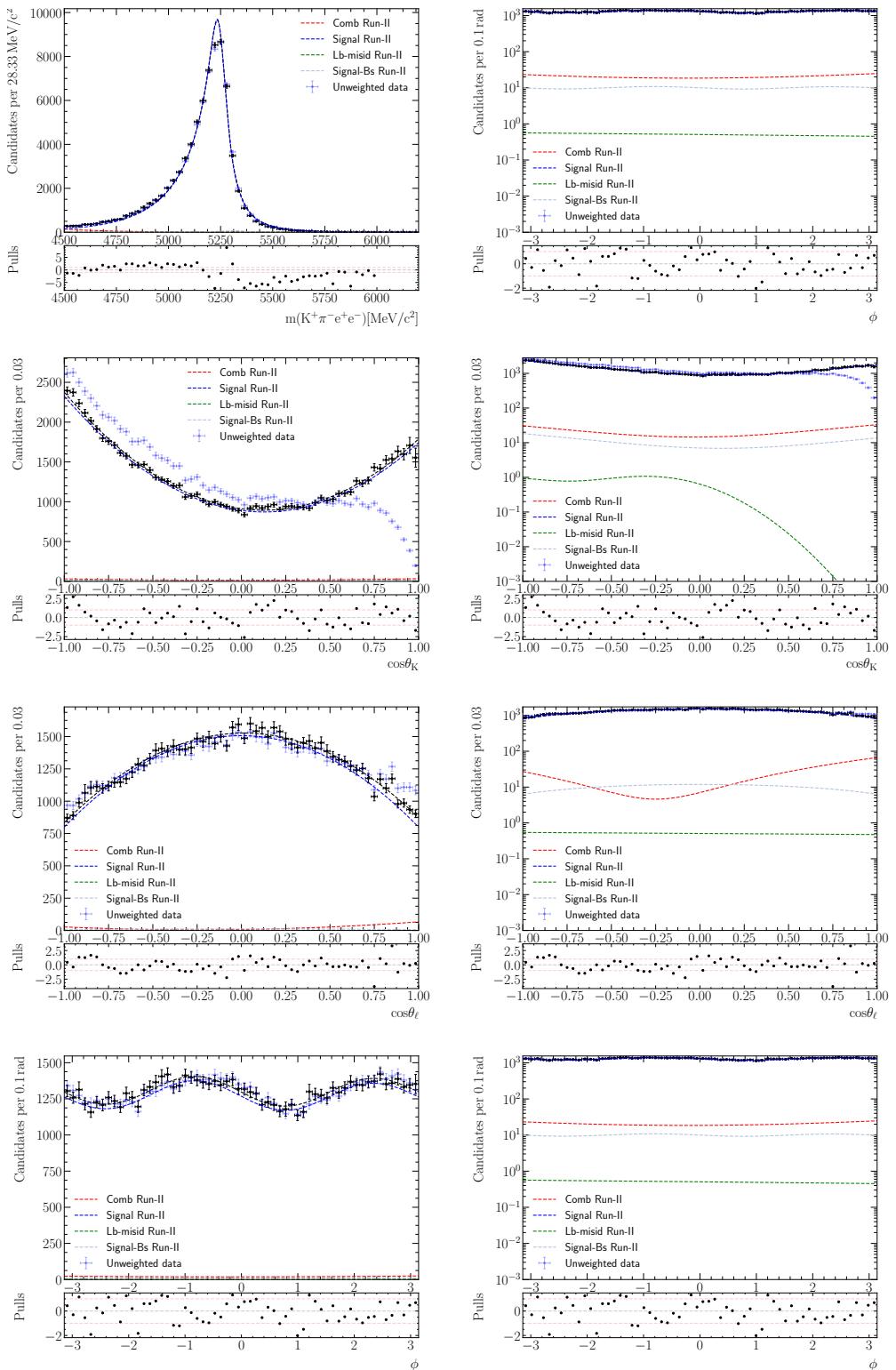


Figure 172: Result of fit to Run-II samples. The fit strategy used is the same as that one used for the nominal combinatorial only fit.

3316 **J.2.2 Separating by Run and trigger category**

3317 Figure 174 shows the results of the separate fits to the LOE and LOTIS trigger categories
3318 of Run-I and Run-II. Table 62 contains the nuisance parameter results.

	Run-I LOE		Run-II LOE		Run-I LOTIS		Run-II LOTIS	
f_{sig}	0.9776	± 0.0020	0.9816	± 0.0015	0.9696	± 0.0023	0.9732	± 0.0019
δ_μ	-7.0	± 0.7	-16.0	± 0.6	-9.9	± 0.8	-23.3	± 0.7
δ_σ	1.107	± 0.010	1.082	± 0.008	1.082	± 0.010	1.055	± 0.008
λ_{Comb}	-0.0039	± 0.0004	-0.0046	± 0.0004	-0.00452	± 0.00035	-0.00420	± 0.00023
$c_{1; \text{Comb}}^K$	-0.31	± 0.22	-0.17	± 0.23	0.25	± 0.15	0.14	± 0.13
$c_{1; \text{Comb}}^\ell$	0.62	± 0.21	1.00	± 0.26	0.76	± 0.17	0.69	± 0.15
$c_{1; \text{Comb}}^\phi$	0.04	± 0.21	-0.08	± 0.21	-0.41	± 0.16	0.00	± 0.14

Table 62: Nuisance parameters of the separate fits to Run-I LOE and LOTIS, and Run-II LOE and LOTIS samples.

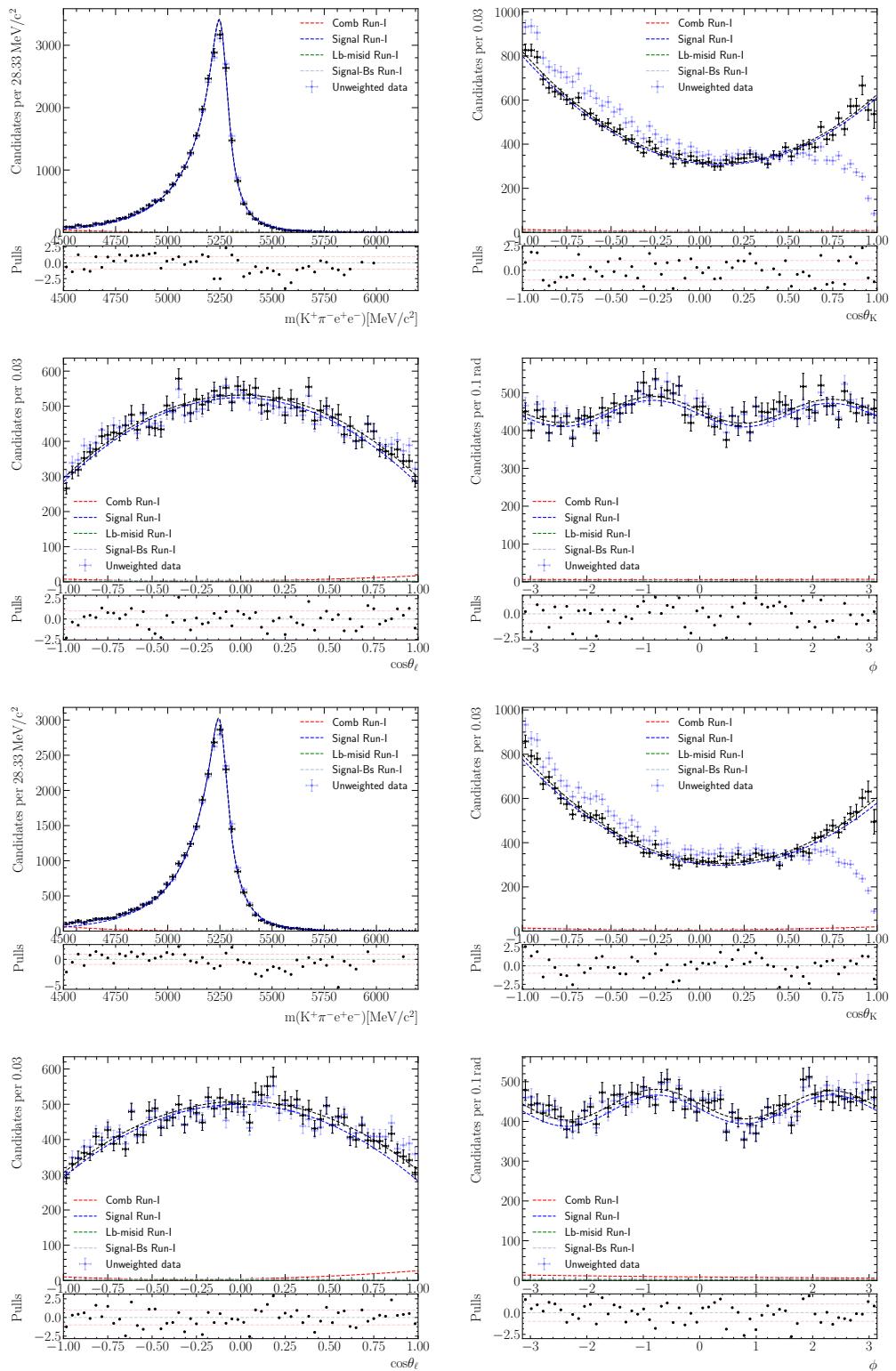


Figure 173: Result of fit to Run-I L0En and L0TIS samples. The fit strategy used is the same as that one used for the nominal combinatorial only fit with the exception that linear functions are used to model the angular distributions of the combinatorial background.

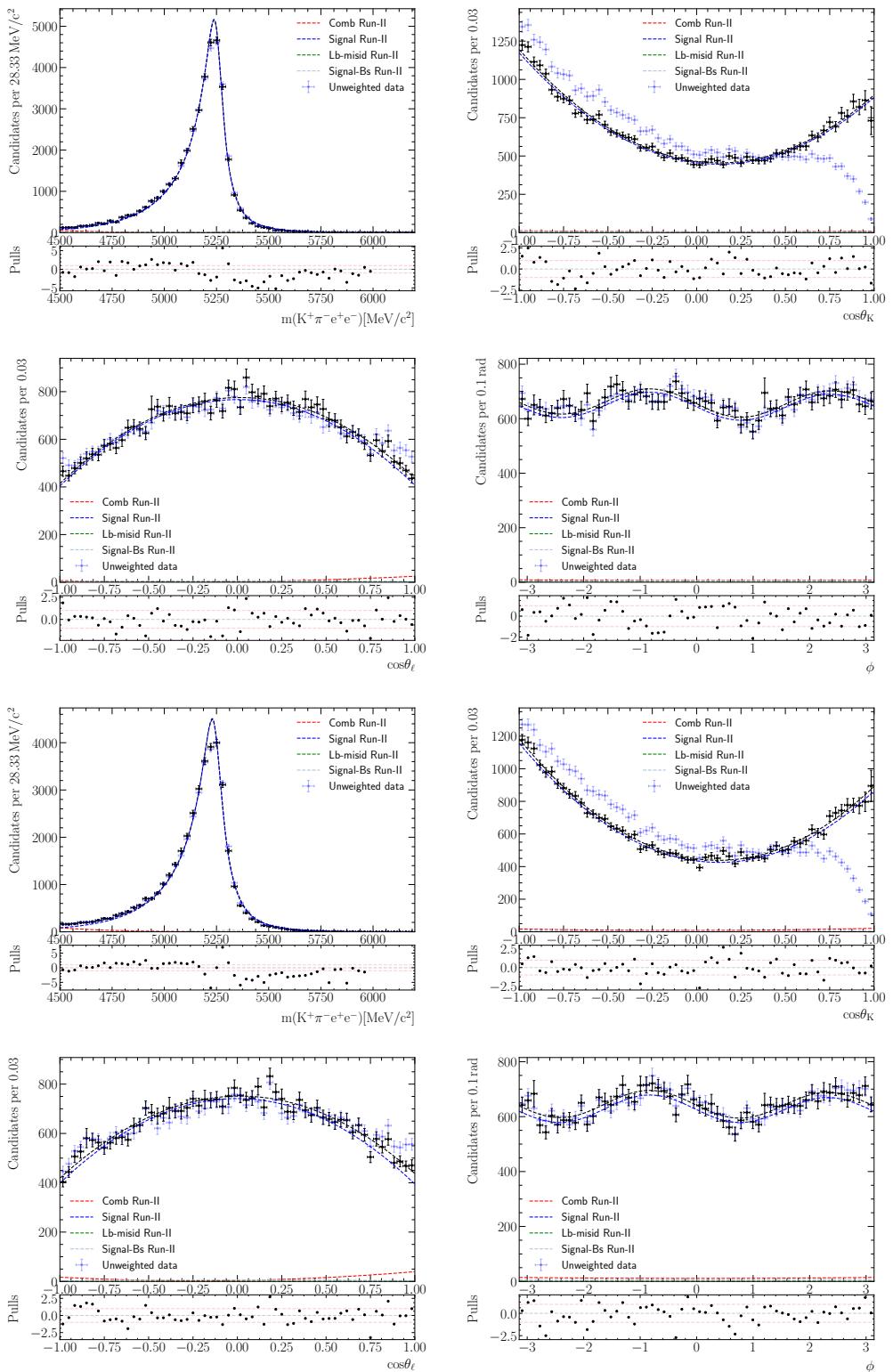


Figure 174: Result of fit to Run-II L0En and L0TIS samples. The fit strategy used is the same as that one used for the nominal combinatorial only fit with the exception that linear functions are used to model the angular distributions of the combinatorial background.

³³¹⁹ **J.2.3 Separating by bremsstrahlung category**

³³²⁰ Figure 175 shows the results of the separate fits to the 0γ , 1γ and 2γ categories. Table 63
³³²¹ contains the nuisance parameter results.

	0γ		1γ		2γ	
$f_{\text{sig}}^{\text{Run-I}}$	0.9769	± 0.0033	0.9763	± 0.0020	0.9701	± 0.0031
$f_{\text{sig}}^{\text{Run-II}}$	0.9793	± 0.0023	0.9772	± 0.0019	0.9754	± 0.0020
$\delta_{\mu}^{\text{Run-I}}$	-10.1	± 0.8	-9.3	± 0.7	-5.1	± 1.1
$\delta_{\mu}^{\text{Run-II}}$	-11.1	± 0.6	-19.4	± 0.6	-26.3	± 1.0
$\delta_{\sigma}^{\text{Run-I}}$	1.121	± 0.018	1.108	± 0.010	1.053	± 0.013
$\delta_{\sigma}^{\text{Run-II}}$	1.128	± 0.013	1.088	± 0.008	1.038	± 0.011
$\lambda_{\text{Comb}}^{\text{Run-I}}$	-0.00294 ± 0.00035		-0.0050 ± 0.0006		-0.00304 ± 0.00032	
$\lambda_{\text{Comb}}^{\text{Run-II}}$	-0.00254 ± 0.00023		-0.0045 ± 0.0004		-0.0042 ± 0.0004	
$c_{1;\text{Comb shared}}^K$	-0.85	± 0.33	0.16	± 0.14	0.28	± 0.15
$c_{1;\text{Comb shared}}^{\ell}$	1.03	± 0.20	0.77	± 0.11	0.95	± 0.17
$c_{2;\text{Comb shared}}^{\ell}$	0.25	± 0.19	0.77	± 0.09	0.75	± 0.11
$c_{1;\text{Comb shared}}^{\phi}$	0.18	± 0.21	-0.32	± 0.13	-0.10	± 0.16

Table 63: Nuisance parameters of the separate fits to the three bremsstrahlung categories.

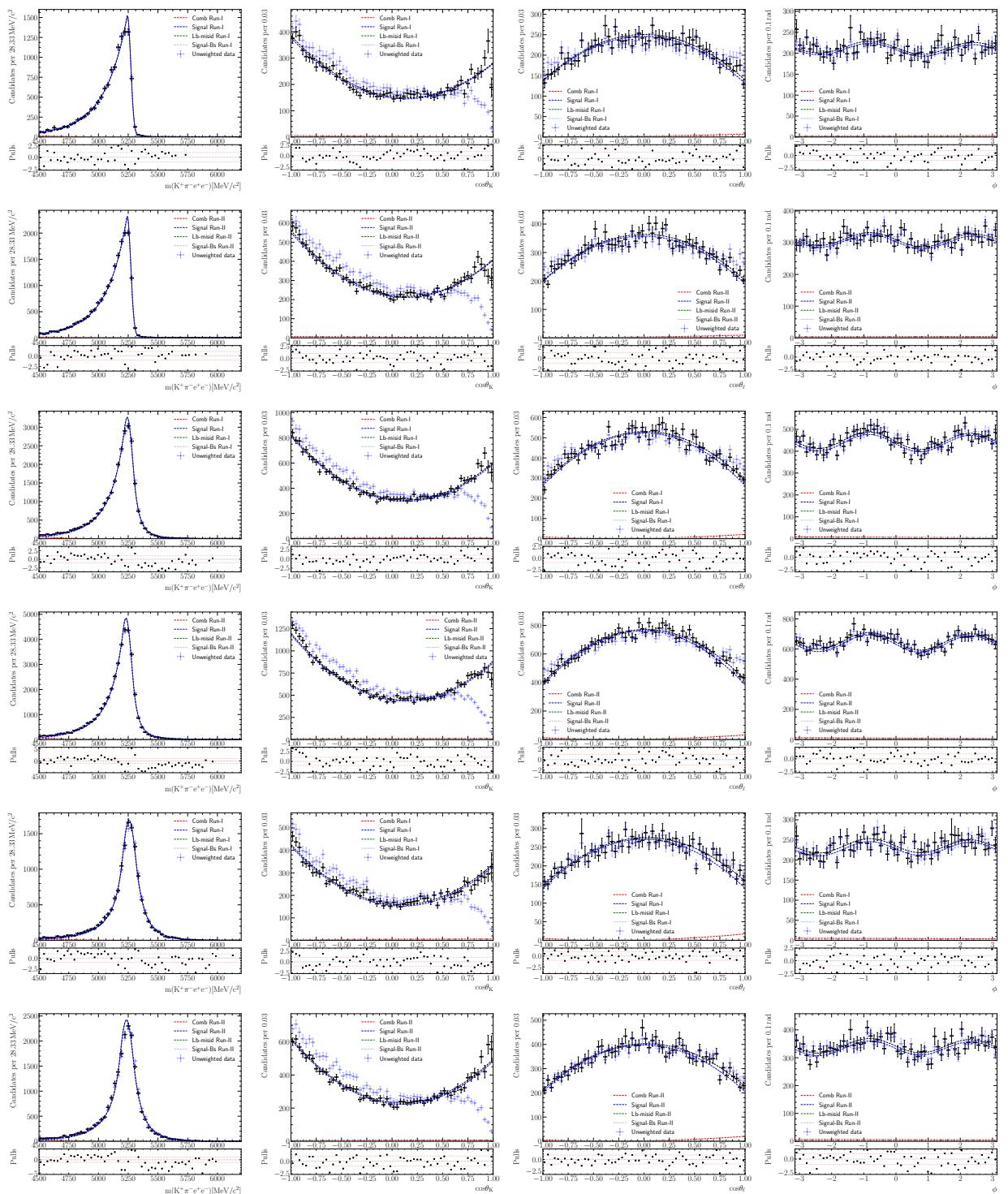


Figure 175: Result of angular fits to three separate bremsstrahlung categories. The events belonging to the three categories of 0γ , 1γ and 2γ are shown in the top, middle and bottom two rows. Separate acceptance functions and fixed models for the misidentified Λ_b^0 decay background are used.

3322 J.3 Similar configurations

3323 Additional information associated with cross-checks using variations of the main validation
 3324 fit (Section 9.4.2) are shown in the sections below.

3325 J.3.1 FLATQ2 acceptance

3326 Figure 177 shows the results of the fit made with rare mode FLATQ2 acceptance correction.
 3327 Table 64 contains the nuisance parameter results.

	Result	
$f_{\text{sig}}^{\text{Run-I}}$	0.9721	± 0.0022
$f_{\text{sig}}^{\text{Run-II}}$	0.9769	± 0.0015
$\delta_{\mu}^{\text{Run-I}}$	-9.0	± 0.5
$\delta_{\mu}^{\text{Run-II}}$	-18.8	± 0.4
$\delta_{\sigma}^{\text{Run-I}}$	1.103	± 0.008
$\delta_{\sigma}^{\text{Run-II}}$	1.070	± 0.007
$\lambda_{\text{Comb}}^{\text{Run-I}}$	-0.00402	± 0.00019
$\lambda_{\text{Comb}}^{\text{Run-II}}$	-0.00439	± 0.00028
$c_1^K; \text{Comb Run-I}$	0.17	± 0.21
$c_1^K; \text{Comb Run-II}$	0.02	± 0.14
$c_2^K; \text{Comb Run-I}$	0.67	± 0.14
$c_2^K; \text{Comb Run-II}$	0.34	± 0.10
$c_1^{\ell}; \text{Comb Run-I}$	0.35	± 0.33
$c_1^{\ell}; \text{Comb Run-II}$	0.82	± 0.14
$c_2^{\ell}; \text{Comb Run-I}$	0.77	± 0.15
$c_2^{\ell}; \text{Comb Run-II}$	0.74	± 0.11
$c_1^{\phi}; \text{Comb Run-I}$	-0.44	± 0.29
$c_1^{\phi}; \text{Comb Run-II}$	-0.04	± 0.13
$c_2^{\phi}; \text{Comb Run-I}$	0.13	± 0.13
$c_2^{\phi}; \text{Comb Run-II}$	0.11	± 0.13

Table 64: Nuisance parameters obtained from the fit made using FLATQ2 acceptance correction.

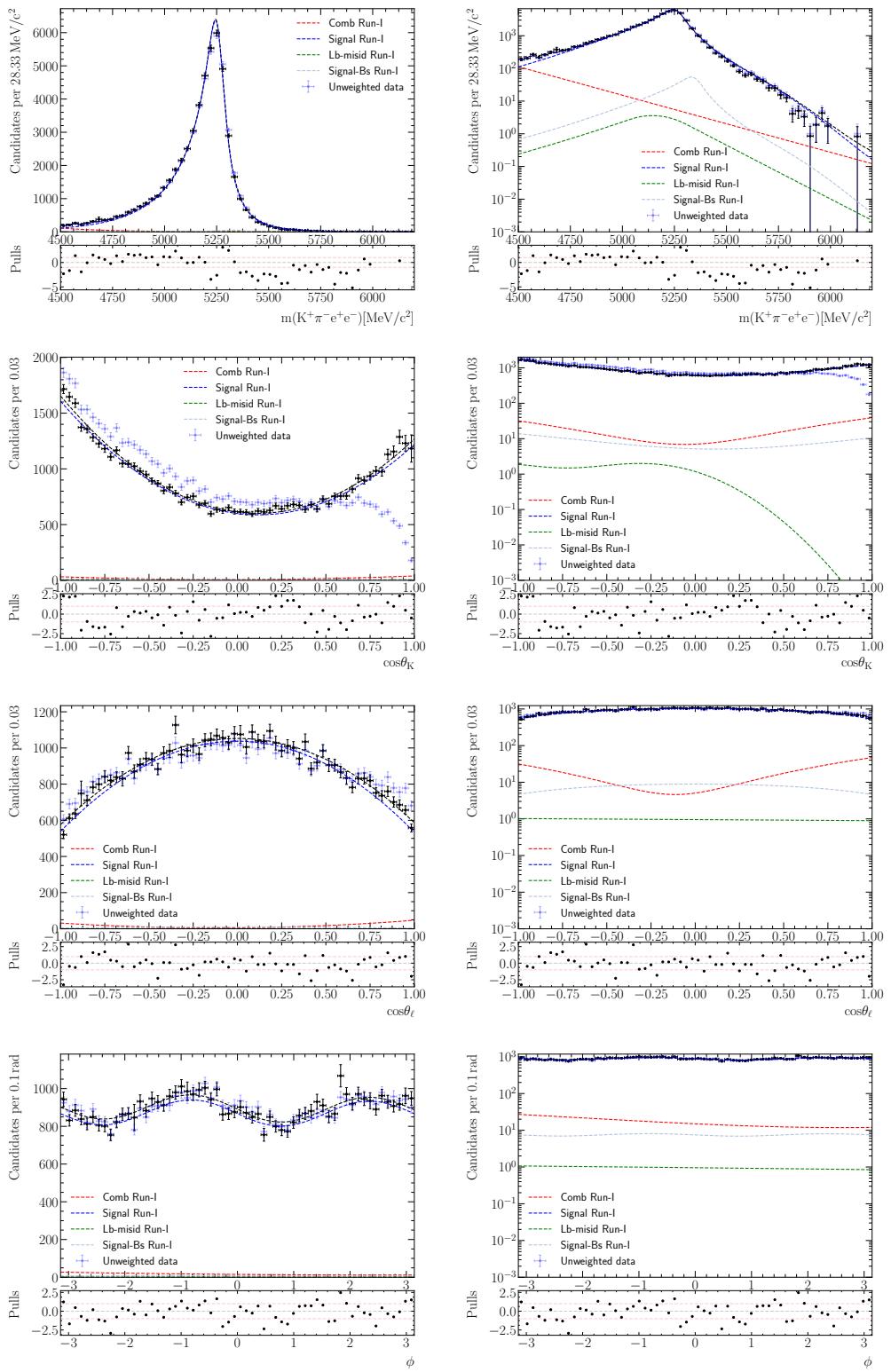


Figure 176: Results of the control mode angular fit with correction weights taken from the nominal rare mode acceptance for **Run-I**. A single combinatorial component is used, but allowed to vary to account for remaining DSL events.

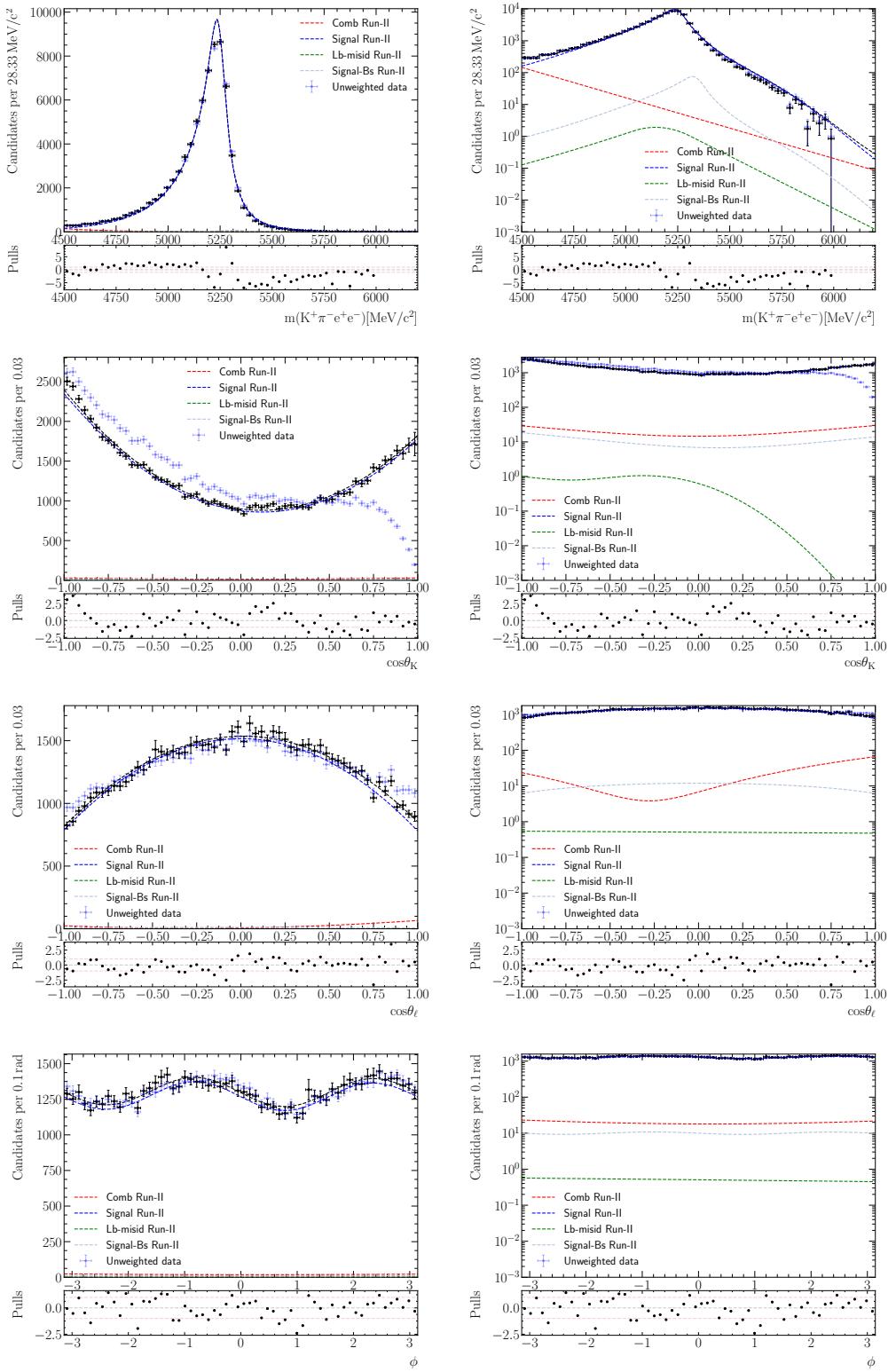


Figure 177: Results of the control mode angular fit with correction weights taken from the nominal rare mode acceptance for Run-II. A single combinatorial component is used, but allowed to vary to account for remaining DSL events.

3328 J.3.2 Higher order acceptance

3329 Figures 179 and 181 show the results of the fit made using higher order control mode PHYS
 3330 acceptance with and without the removal of six events with large weights, respectively.
 3331 Tables 65 and 66 contain the nuisance parameter results of the above two fits.

	Result	
$f_{\text{sig}}^{\text{Run-I}}$	0.9753	± 0.0020
$f_{\text{sig}}^{\text{Run-II}}$	0.9767	± 0.0014
$\delta_{\mu}^{\text{Run-I}}$	-9.1	± 0.6
$\delta_{\mu}^{\text{Run-II}}$	-19.5	± 0.5
$\delta_{\sigma}^{\text{Run-I}}$	1.101	± 0.009
$\delta_{\sigma}^{\text{Run-II}}$	1.065	± 0.006
$\lambda_{\text{Comb}}^{\text{Run-I}}$	$-0.004\ 55$	$\pm 0.000\ 32$
$\lambda_{\text{Comb}}^{\text{Run-II}}$	$-0.004\ 34$	$\pm 0.000\ 21$
$c_1^K; \text{Comb Run-I}$	0.11	± 0.13
$c_1^K; \text{Comb Run-II}$	-0.01	± 0.13
$c_2^K; \text{Comb Run-I}$	0.53	± 0.10
$c_2^K; \text{Comb Run-II}$	0.36	± 0.10
$c_1^{\ell}; \text{Comb Run-I}$	0.72	± 0.15
$c_1^{\ell}; \text{Comb Run-II}$	0.80	± 0.12
$c_2^{\ell}; \text{Comb Run-I}$	0.72	± 0.12
$c_2^{\ell}; \text{Comb Run-II}$	0.70	± 0.10
$c_1^{\phi}; \text{Comb Run-I}$	-0.26	± 0.14
$c_1^{\phi}; \text{Comb Run-II}$	-0.02	± 0.12
$c_2^{\phi}; \text{Comb Run-I}$	0.10	± 0.13
$c_2^{\phi}; \text{Comb Run-II}$	0.06	± 0.12

Table 65: Nuisance parameters obtained from the fit made using higher order control mode acceptances without the removal of events with large weights.

	Result	
$f_{\text{sig}}^{\text{Run-I}}$	0.9747	± 0.0019
$f_{\text{sig}}^{\text{Run-II}}$	0.9767	± 0.0014
$\delta_{\mu}^{\text{Run-I}}$	-8.7	± 0.5
$\delta_{\mu}^{\text{Run-II}}$	-19.3	± 0.4
$\delta_{\sigma}^{\text{Run-I}}$	1.102	± 0.008
$\delta_{\sigma}^{\text{Run-II}}$	1.067	± 0.006
$\lambda_{\text{Comb}}^{\text{Run-I}}$	-0.00449	± 0.00029
$\lambda_{\text{Comb}}^{\text{Run-II}}$	-0.00434	± 0.00021
$c_{1; \text{Comb Run-I}}^K$	0.12	± 0.13
$c_{1; \text{Comb Run-II}}^K$	0.01	± 0.12
$c_{2; \text{Comb Run-I}}^K$	0.54	± 0.09
$c_{2; \text{Comb Run-II}}^K$	0.38	± 0.10
$c_{1; \text{Comb Run-I}}^{\ell}$	0.70	± 0.15
$c_{1; \text{Comb Run-II}}^{\ell}$	0.81	± 0.13
$c_{2; \text{Comb Run-I}}^{\ell}$	0.70	± 0.12
$c_{2; \text{Comb Run-II}}^{\ell}$	0.69	± 0.10
$c_{1; \text{Comb Run-I}}^{\phi}$	-0.26	± 0.14
$c_{1; \text{Comb Run-II}}^{\phi}$	-0.02	± 0.12
$c_{2; \text{Comb Run-I}}^{\phi}$	0.09	± 0.13
$c_{2; \text{Comb Run-II}}^{\phi}$	0.06	± 0.12

Table 66: Nuisance parameters obtained from the fit made using higher order control mode acceptances with the removal of events with large weights.

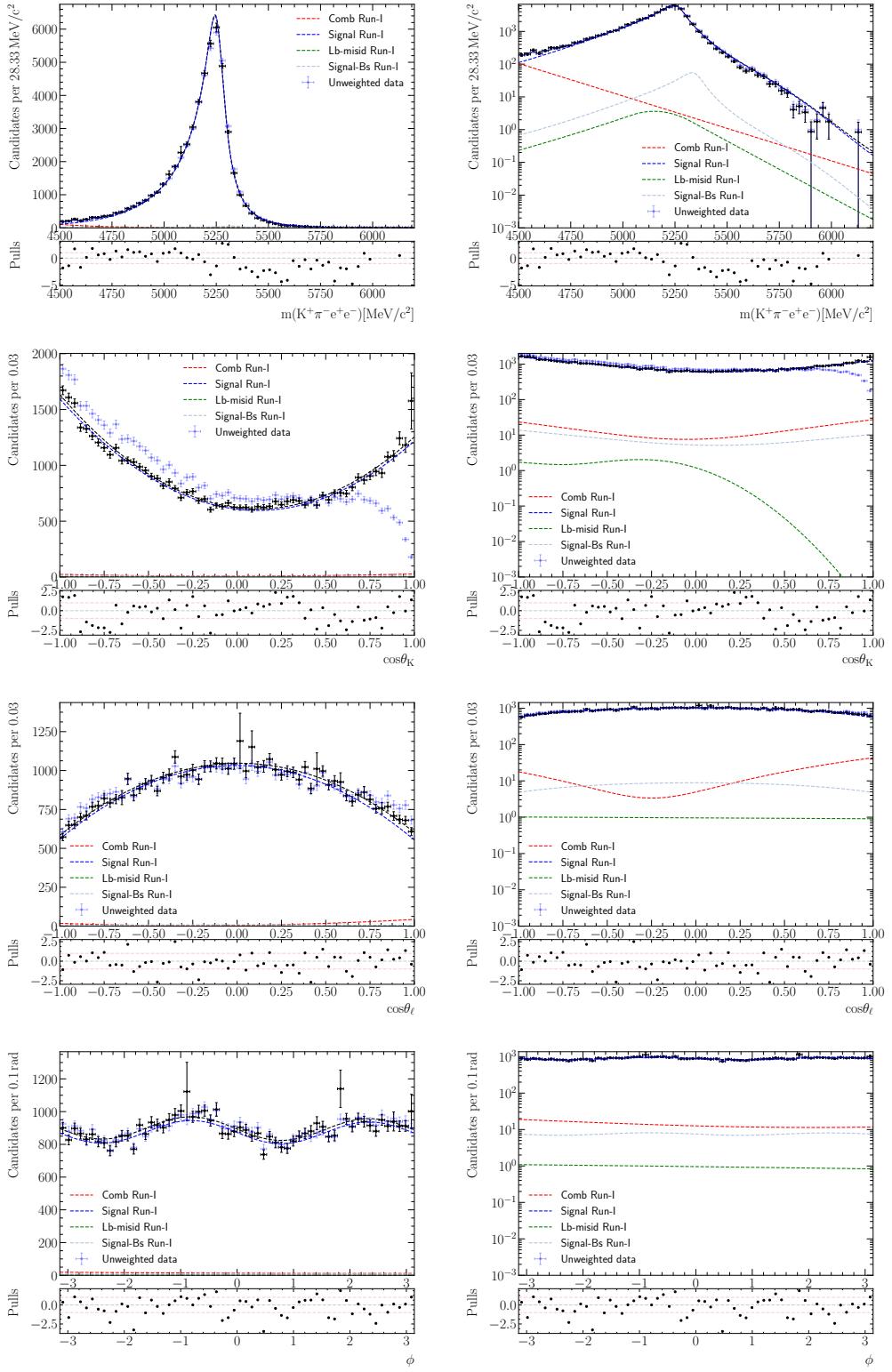


Figure 178: Result of the fit made with corrections from acceptance functions parametrised using higher orders (all angle orders increased by three) to Run-I samples. The fit is otherwise set-up in the same way as the nominal (combinatorial only) configuration. The fit attempt without removing events with large weights is shown here.

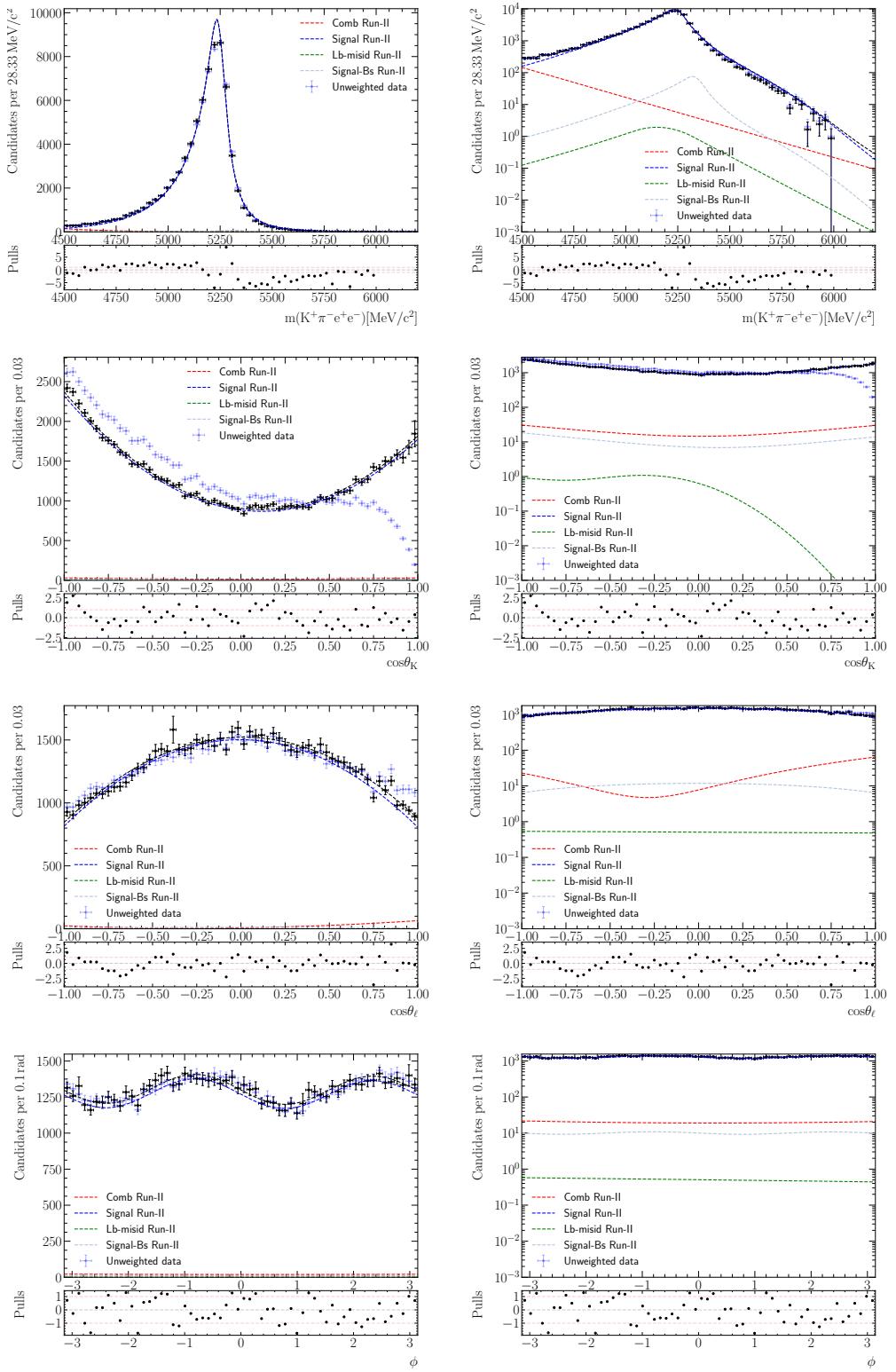


Figure 179: Result of the fit made with corrections from acceptance functions parametrised using higher orders (all angle orders increased by three) to Run-II samples. The fit is otherwise set-up in the same way as the nominal (combinatorial only) configuration. The fit attempt without removing events with large weights is shown here.

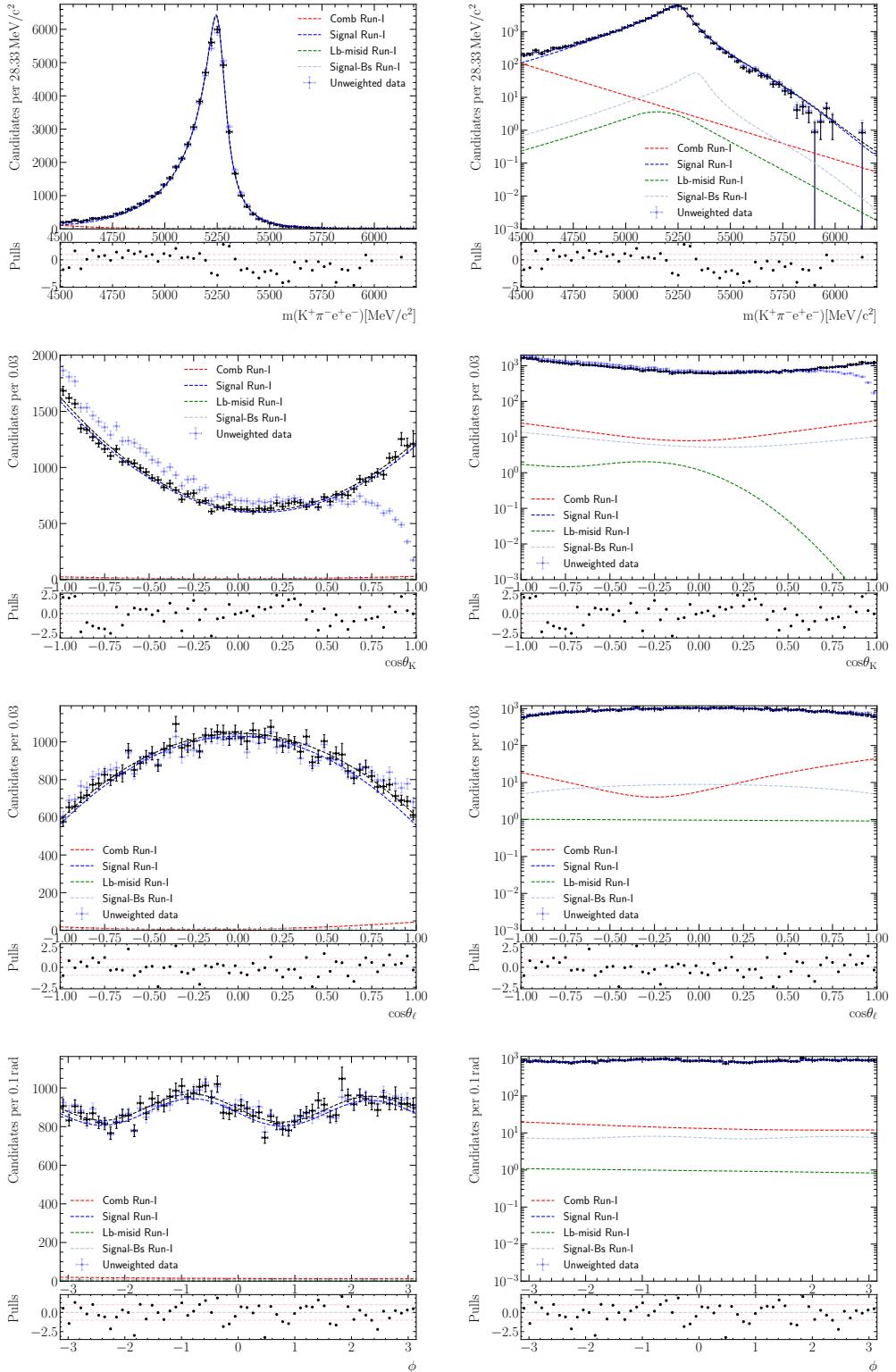


Figure 180: Result of the fit made with corrections from acceptance functions parametrised using higher orders (all angle orders increased by three) to Run-I samples. The fit is otherwise set-up in the same way as the nominal (combinatorial only) configuration. The fit attempt made with the removal of six events is shown here.

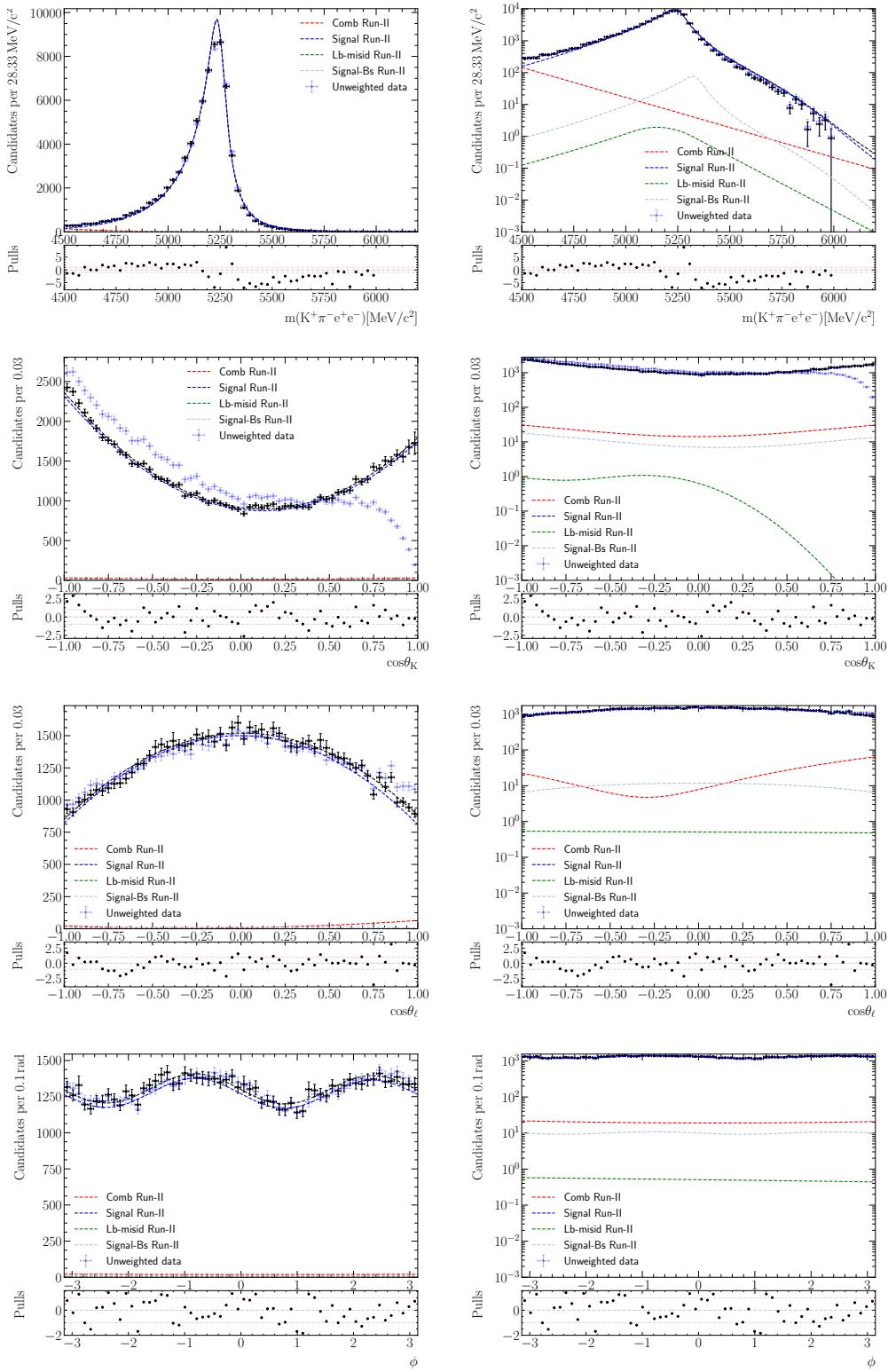


Figure 181: Result of the fit made with corrections from acceptance functions parametrised using higher orders (all angle orders increased by three) to Run-II samples. The fit is otherwise set-up in the same way as the nominal (combinatorial only) configuration. The fit attempt made with the removal of six events is shown here.

3332 **J.3.3 Resolution correction**

3333 Figure 183 shows the results of the fit made using alternative control mode acceptances
 3334 parametrised with corrections to the resolution modelling. Table 67 contains the nuisance
 3335 parameter results of the fit.

	Result	
$f_{\text{sig}}^{\text{Run-I}}$	0.9747	± 0.0018
$f_{\text{sig}}^{\text{Run-II}}$	0.9766	± 0.0014
$\delta_{\mu}^{\text{Run-I}}$	-8.5	± 0.5
$\delta_{\mu}^{\text{Run-II}}$	-19.2	± 0.4
$\delta_{\sigma}^{\text{Run-I}}$	1.102	± 0.008
$\delta_{\sigma}^{\text{Run-II}}$	1.066	± 0.006
$\lambda_{\text{Comb}}^{\text{Run-I}}$	$-0.004\,35$	$\pm 0.000\,27$
$\lambda_{\text{Comb}}^{\text{Run-II}}$	$-0.004\,27$	$\pm 0.000\,18$
$c_1^K; \text{Comb Run-I}$	0.08	± 0.13
$c_1^K; \text{Comb Run-II}$	0.04	± 0.12
$c_2^K; \text{Comb Run-I}$	0.51	± 0.09
$c_2^K; \text{Comb Run-II}$	0.39	± 0.09
$c_1^{\ell}; \text{Comb Run-I}$	0.69	± 0.14
$c_1^{\ell}; \text{Comb Run-II}$	0.84	± 0.12
$c_2^{\ell}; \text{Comb Run-I}$	0.72	± 0.12
$c_2^{\ell}; \text{Comb Run-II}$	0.71	± 0.10
$c_1^{\phi}; \text{Comb Run-I}$	-0.21	± 0.13
$c_1^{\phi}; \text{Comb Run-II}$	-0.06	± 0.12
$c_2^{\phi}; \text{Comb Run-I}$	0.15	± 0.13
$c_2^{\phi}; \text{Comb Run-II}$	0.10	± 0.12

Table 67: Nuisance parameters obtained from the fit made using acceptance functions parametrised with adjusted resolution models.

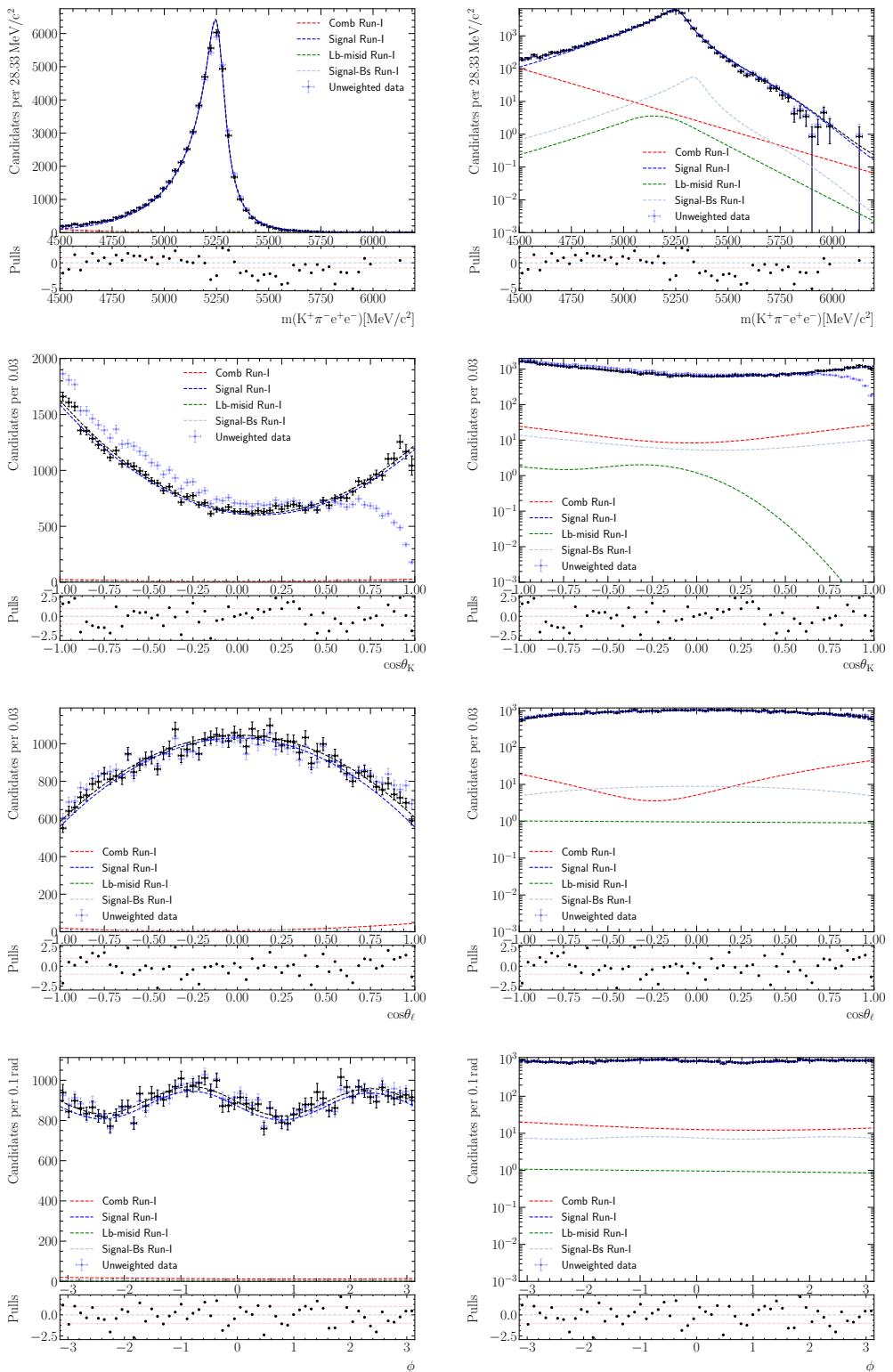


Figure 182: Result of the Run-I fit with alternative acceptance obtained with an adjusted resolution model. The fit is otherwise set-up in the same way as the nominal (combinatorial only) configuration.

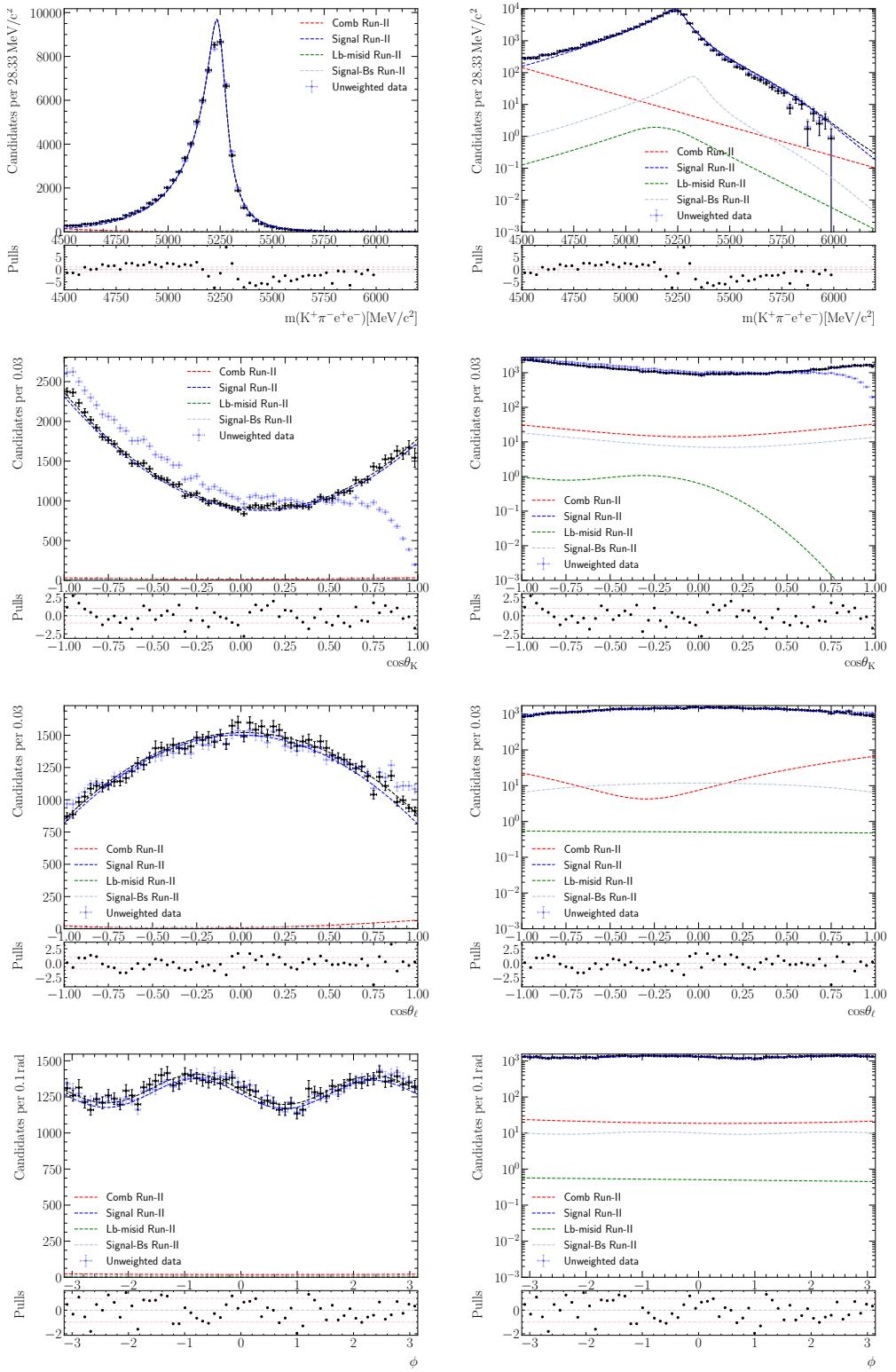


Figure 183: Result of the Run-II fit with alternative acceptance obtained with an adjusted resolution model. The fit is otherwise set-up in the same way as the nominal (combinatorial only) configuration.

3336 **J.3.4 No B^+ veto**

3337 Figure 185 shows the results of the fit without B^+ veto. Table 68 contains the nuisance
3338 parameter results.

	Result	
$f_{\text{sig}}^{\text{Run-I}}$	0.9781	± 0.0014
$f_{\text{sig}}^{\text{Run-II}}$	0.9795	± 0.0013
$\delta_{\mu}^{\text{Run-I}}$	-4.6	± 0.5
$\delta_{\mu}^{\text{Run-II}}$	-15.7	± 0.4
$\delta_{\sigma}^{\text{Run-I}}$	1.134	± 0.008
$\delta_{\sigma}^{\text{Run-II}}$	1.103	± 0.006
$\lambda_{\text{Comb}}^{\text{Run-I}}$	-0.0043	± 0.0004
$\lambda_{\text{Comb}}^{\text{Run-II}}$	-0.00400	± 0.00031
$c_1^K; \text{Comb Run-I}$	-0.12	± 0.09
$c_2^K; \text{Comb Run-I}$	0.28	± 0.08
$c_1^\ell; \text{Comb Run-I}$	0.87	± 0.09
$c_2^\ell; \text{Comb Run-I}$	0.88	± 0.08
$c_1^\phi; \text{Comb Run-I}$	-0.07	± 0.09
$c_2^\phi; \text{Comb Run-I}$	0.12	± 0.09

Table 68: Nuisance parameters obtained from the fit made without B^+ veto.

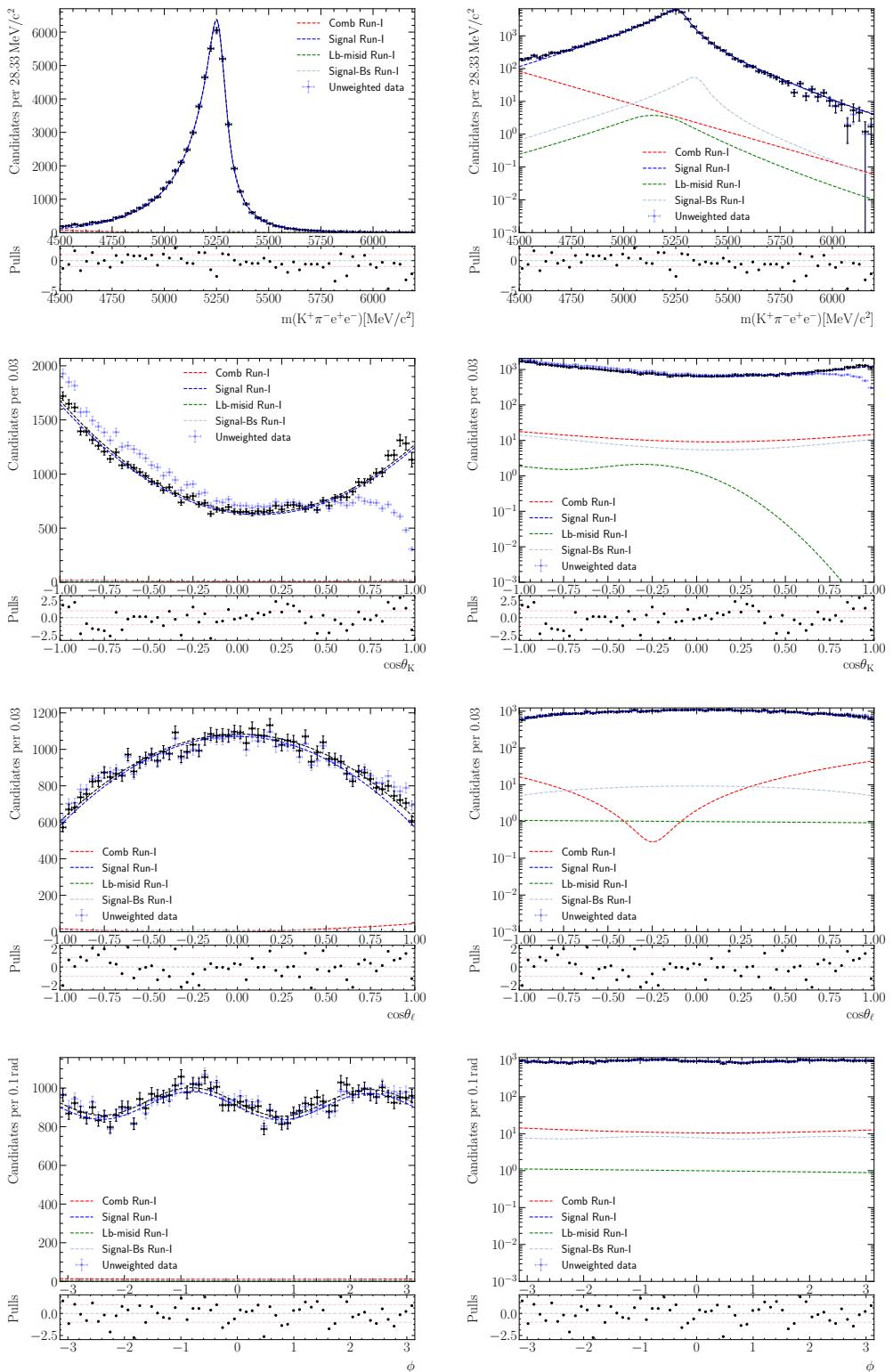


Figure 184: Result of the fit without B^+ veto to Run-I samples. The nominal selection is used (with the exception of the veto). One shared set of parameters is used for the angular distributions of the combinatorial component for both Runs.

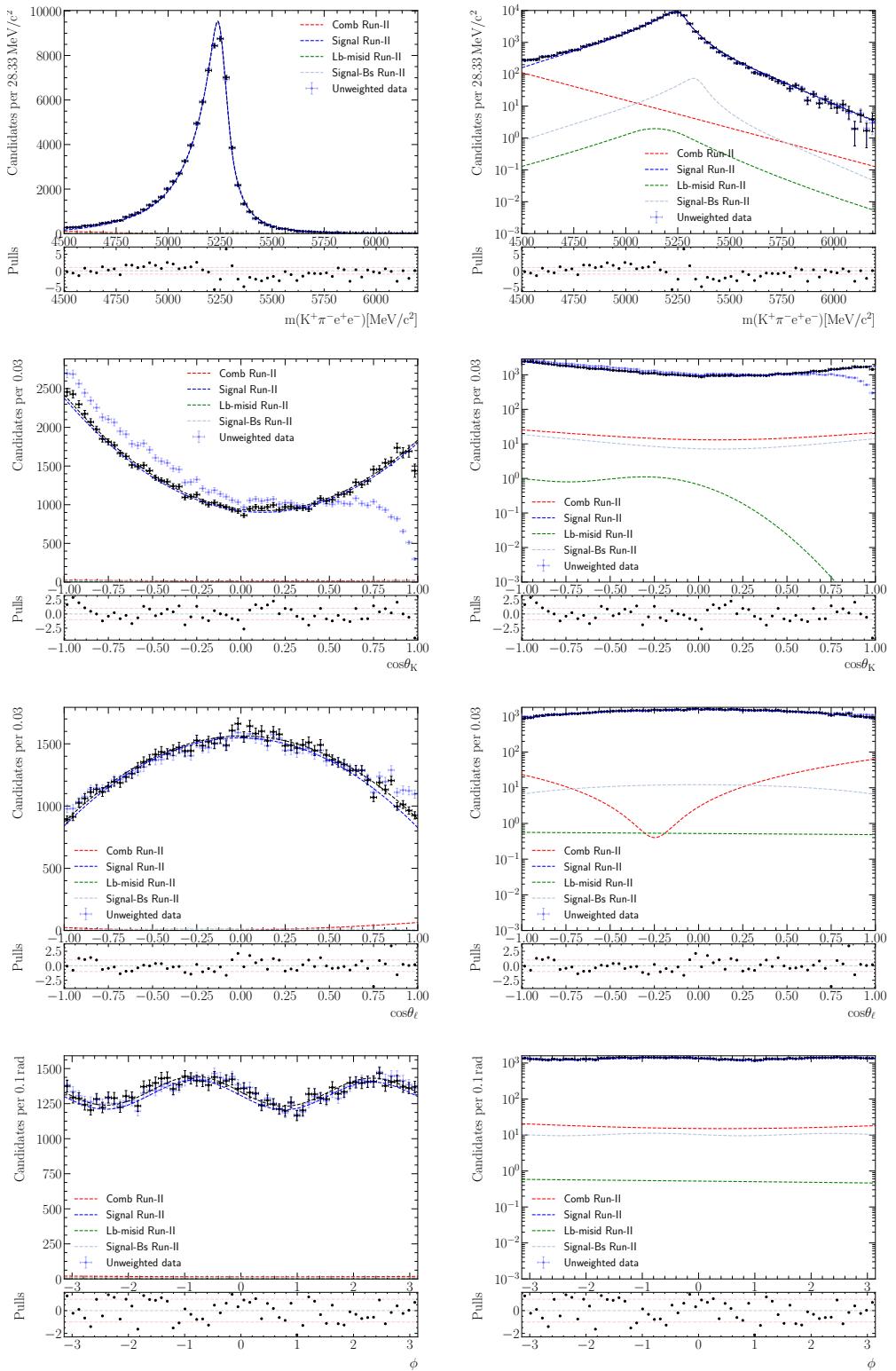


Figure 185: Result of the fit without B^+ veto to Run-I samples. The nominal selection is used (with the exception of the veto). One shared set of parameters is used for the angular distributions of the combinatorial component for both Runs.

3339 **J.4 Additional backgrounds**

3340 Additional information associated with cross-checks using additional backgrounds (Sec-
3341 tion 9.4.3) are shown in the sections below, including descriptions of the background
3342 models and fractions used in the fits.

3343 **Partially reconstructed**

3344 Partially reconstructed decays are comprised of two main types – leptonic and hadronic.
3345 The former includes decays where the J/ψ (and unreconstructed particles denoted by
3346 ‘X’) comes from the decay of some higher $c\bar{c}$ resonance, *i.e.* $B \rightarrow K^{*0}\psi(2S)(\rightarrow J/\psi(\rightarrow$
3347 $e^+e^-)X)$, and the latter includes all other partially reconstructed decays that do not fulfil
3348 the leptonic criteria, one example of which is $B^0 \rightarrow K_1(1270)(\rightarrow K^{*0}X)J/\psi(\rightarrow e^+e^-)$.
3349 The hadronic and leptonic partially reconstructed components are modelled using a mixture
3350 of inclusive simulations of the decays of B^0 , B^+ and B_s^0 , truth-matched to separate the
3351 hadronic and leptonic components, and combined using normalisation factors calculated
3352 based on the selection and reconstruction efficiencies, hadronisation fraction, and the total
3353 branching fraction of the inclusive decay channels. Note that due to the lack of availability
3354 of samples from certain years (and that the aim of the cross-check is not to make a
3355 precision measurement), where necessary, efficiency values from 2017 and 2018 samples are
3356 used for the estimation, and one common value is used for both **Runs**. Furthermore, no
3357 corrections to simulation are considered. Combinations of two to three Gaussian functions
3358 are used to describe the mass distributions of the six sub-components, and Chebyshev
3359 polynomials are used for the angles, the orders of which are chosen to achieve good
3360 description of the given distributions. The full models for the unconstrained q^2 range of
3361 $6.0 < q^2 < 11.0 \text{ GeV}^2/c^4$ are shown in Figures 188 and 187. Models used for the fit in the
3362 range of $7.0 < q_c^2 < 11.0 \text{ GeV}^2/c^4$ are shown in Figures 190 and 189, and those used for fit
3363 with the selection of $7.0 < q_c^2 < 11.0 \& m_B^{J/\psi} > 5150 \text{ MeV}/c^2$ are shown in Figure 191.

3364 **$B_s^0 \rightarrow \phi J/\psi (\rightarrow e^+e^-)$ background**

3365 The $B_s^0 \rightarrow \phi J/\psi(\rightarrow e^+e^-)$ decays can be mistaken for signal when a kaon from the
3366 $\phi \rightarrow K^+K^-$ decay is mis-identified as a pion. This type of background is suppressed by a
3367 dedicated veto (Section 4), and is therefore not modelled in the main validation fit. For
3368 cross-check fits, its mass distribution is modelled using a DCB function, and its angular
3369 distributions are modelled using Chebyshev polynomials with the exception of $\cos\theta_K$
3370 in the case of configurations that use the constrained q^2 cuts. For these situations the
3371 combination of an exponential and a Gaussian function is used. This choice is made purely
3372 for convenience. The fit to simulation for the unconstrained q^2 cut is shown in Figure 192.
3373 The model used for the fit in the range of $7.0 < q_c^2 < 11.0 \text{ GeV}^2/c^4$ is shown in Figure 193,
3374 and the one used for fit with the selection of $7.0 < q_c^2 < 11.0 \& m_B^{J/\psi} > 5150 \text{ MeV}/c^2$ is
3375 shown in Figure 194.

3376 **Double hadron swap**

3377 Control mode signal events with the kaon mis-identified as a pion and vice versa is a source

3378 of background that is suppressed by the dedicated hadron-hadron swap veto (see Section 4).
 3379 Nevertheless, a small fraction of events is expected to remain. The mass distribution of
 3380 this component is modelled using a DCB. Its angular distributions are modelled using
 3381 Chebyshev polynomials with the exception of $\cos\theta_K$, which is described by a sum of three
 3382 Gaussians. The full model for the unconstrained q^2 range is shown in Figure 195. The
 3383 model used for the fit in the range of $7.0 < q_c^2 < 11.0 \text{GeV}^2/\text{c}^4$ is shown in Figure 196, and
 3384 the one used for fit with the selection of $7.0 < q_c^2 < 11.0 \& m_B^{J/\psi} > 5150 \text{MeV}/\text{c}^2$ is shown
 3385 in Figure 197.

3386 Double semi-leptonic

3387 Background due to DSL decays for the control mode can be modelled using the same
 3388 two-step procedure developed for the rare mode. However, due to the larger mass window,
 3389 the fraction of DSL (and DSL-containing combinatorial) events far exceeds that of the
 3390 pure combinatorial in the $K\pi e\mu$ sample, and the mass distribution of the DSL component
 3391 becomes visibly non-exponential. In this case, in contrast to the rare mode, a Gaussian
 3392 distribution is chosen to describe the DSL mass distribution, as it is found to describe
 3393 the distribution well and results in stable fits with minimal additional parameters. For
 3394 the angles, to avoid using KDE (reduce memory usage), a combination of three Gaussians
 3395 is used to model the $\cos\theta_\ell$ shape – one is used to describe the main peak near $\cos\theta_\ell = 1$,
 3396 while two additional Gaussians with shared widths and means of opposite signs are used
 3397 to describe the symmetrically distributed DSL-combinatorial events that is expected to be
 3398 present. The other two angles are modelled using Chebyshev polynomials. The results
 3399 of the second step of the two-step procedure is shown for the case of the unconstrained
 3400 q^2 cut on Figure 198. The model used for the fit in the range of $7.0 < q_c^2 < 11.0 \text{GeV}^2/\text{c}^4$
 3401 with cut on constrained B^0 mass is shown in Figure 199.

3402 $\Lambda_b^0 \rightarrow pK J/\psi (\rightarrow e^+e^-)$ with mis-identification

3403 Background due to mis-identified $\Lambda_b^0 \rightarrow pK J/\psi (\rightarrow e^+e^-)$ decays is modelled using sim-
 3404 ulation with data-driven correction weights for the cross-check fits as is done for the
 3405 nominal control mode mass fit. Here, to ensure consistency, the dedicated veto is not
 3406 applied¹⁸. In contrast to the nominal model, a DCB (instead of KDE) is used for its mass
 3407 distribution, and a sum of two Gaussians is used for $\cos\theta_K$. The other two angles are
 3408 relatively featureless, and are described by Chebyshev polynomials of order one (straight
 3409 line). The result of the fit for the unconstrained q^2 cut is shown in Figure 200. The model
 3410 used for the fit in the range of $7.0 < q_c^2 < 11.0 \text{GeV}^2/\text{c}^4$ is shown in Figure 201, and the
 3411 one used for fit with the selection of $7.0 < q_c^2 < 11.0 \& m_B^{J/\psi} > 5150 \text{MeV}/\text{c}^2$ is shown in
 3412 Figure 202.

¹⁸The Λ_b^0 veto is currently part of the nominal selection. It is applied unless stated otherwise, but it will be removed in future updates.

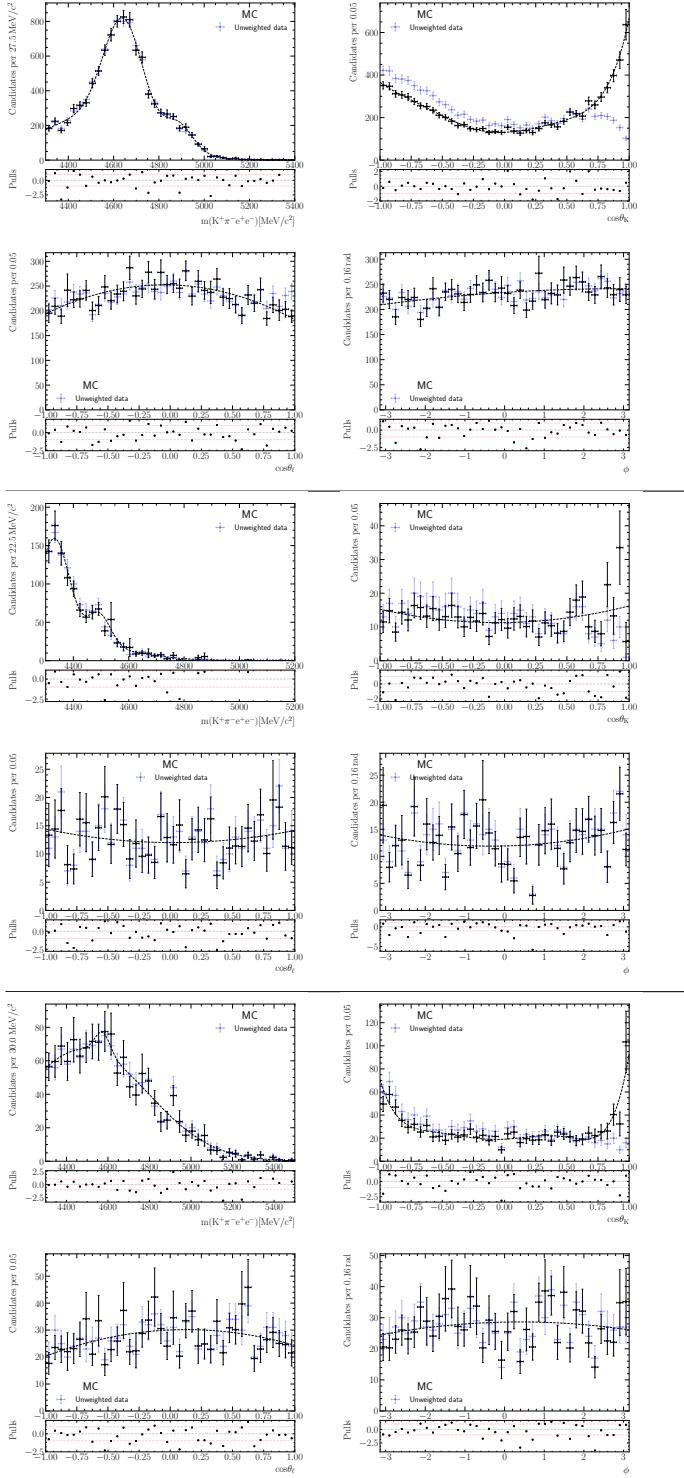


Figure 186: Mass and angular distributions of leptonic partially reconstructed decays modelled using inclusive simulation of B^0 (top), B^+ (middle) and B_s^0 (bottom) decays for the unconstrained q^2 range of $6.0 < q^2 < 11.0$. A single mass and angular model is obtained using all available samples, including simulation samples from years 2017 and 2018 (processed in the same way as 2015-16 samples). The three components are combined into one single function to describe the PRL component.

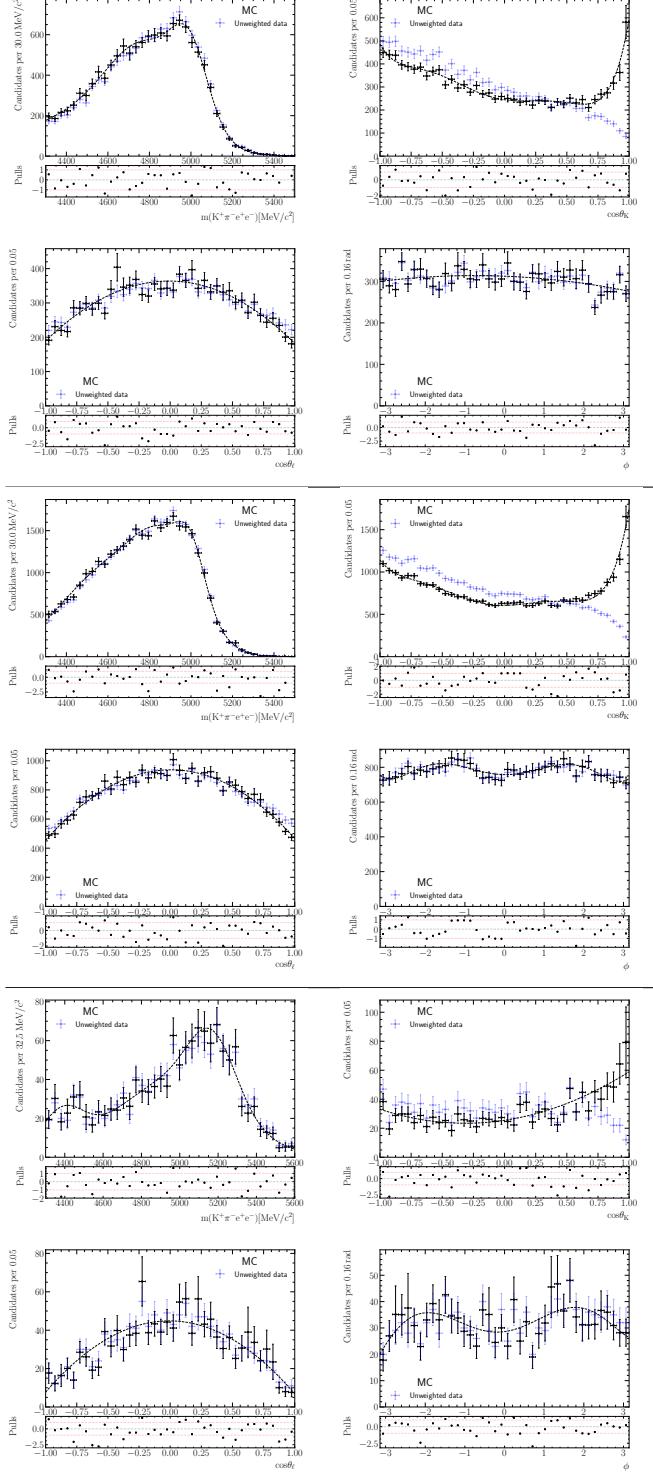


Figure 187: Mass and angular distributions of leptonic partially reconstructed decays modelled using inclusive simulation of B^0 (top), B^+ (middle) and B_s^0 (bottom) decays for the unconstrained q^2 range of $6.0 < q^2 < 11.0$. A single mass and angular model is obtained using all available samples, including simulation samples from years 2017 and 2018 (processed in the same way as Run-IIp1 samples). The three components are combined into one single function to describe the PRH component.

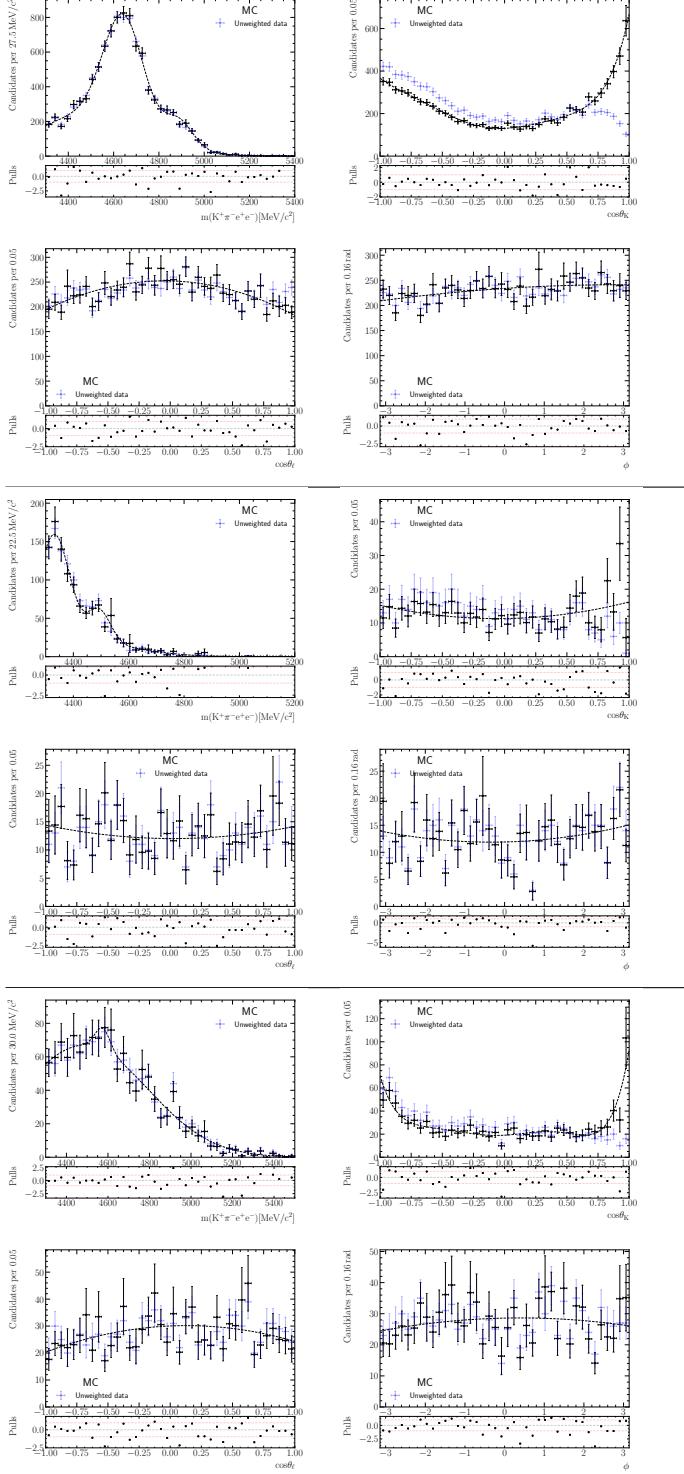


Figure 188: Mass and angular distributions of leptonic partially reconstructed decays modelled using inclusive simulation of B^0 (top), B^+ (middle) and B_s^0 (bottom) decays for the unconstrained q^2 range of $6.0 < q^2 < 11.0$. A single mass and angular model is obtained using all available samples, including simulation samples from years 2017 and 2018 (processed in the same way as 2015-16 samples). The three components are combined into one single function to describe the PRL component.

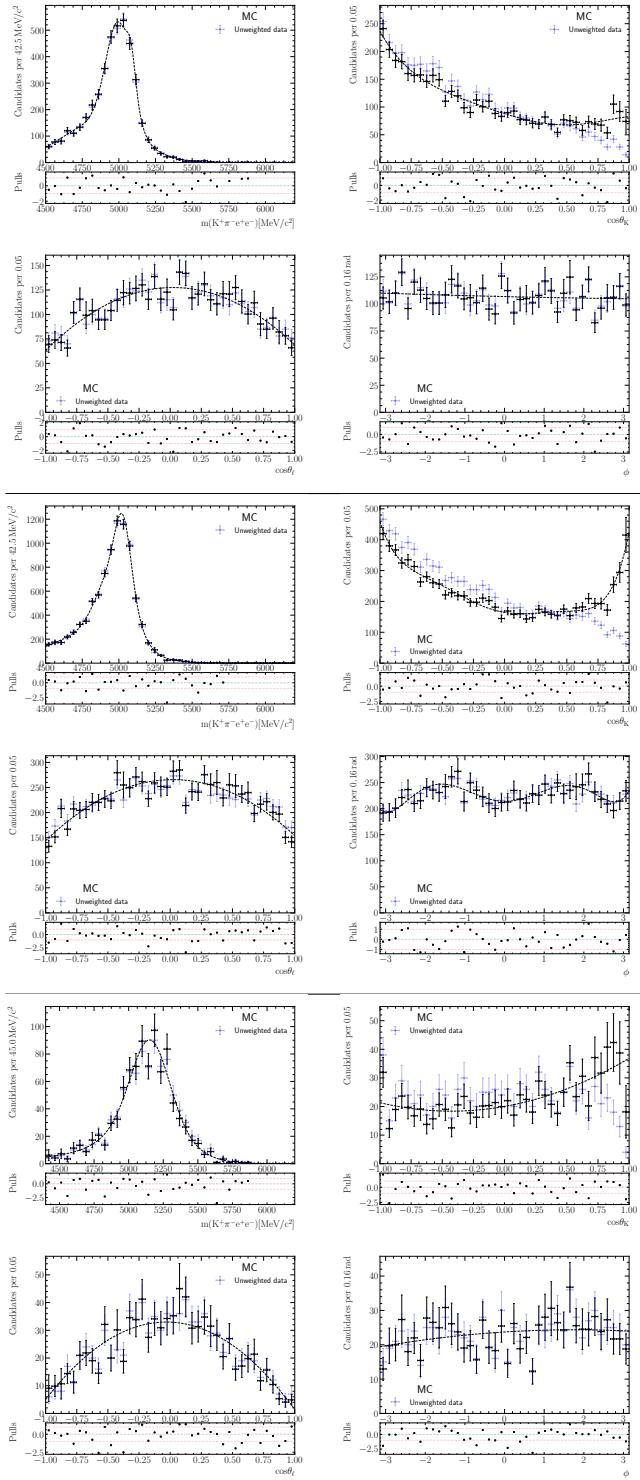


Figure 189: Mass and angular distributions of leptonic partially reconstructed decays modelled using inclusive simulation of B^0 (top), B^+ (middle) and B_s^0 (bottom) decays for the q_c^2 range of $7.0 < q_c^2 < 11.0$ (without cut on $m_{B^0}^{J/\psi}$). A single mass and angular model is obtained using all available samples, including simulation samples from years 2017 and 2018 (processed in the same way as Run-IIp1 samples). The three components are combined into one single function to describe the PRH component.

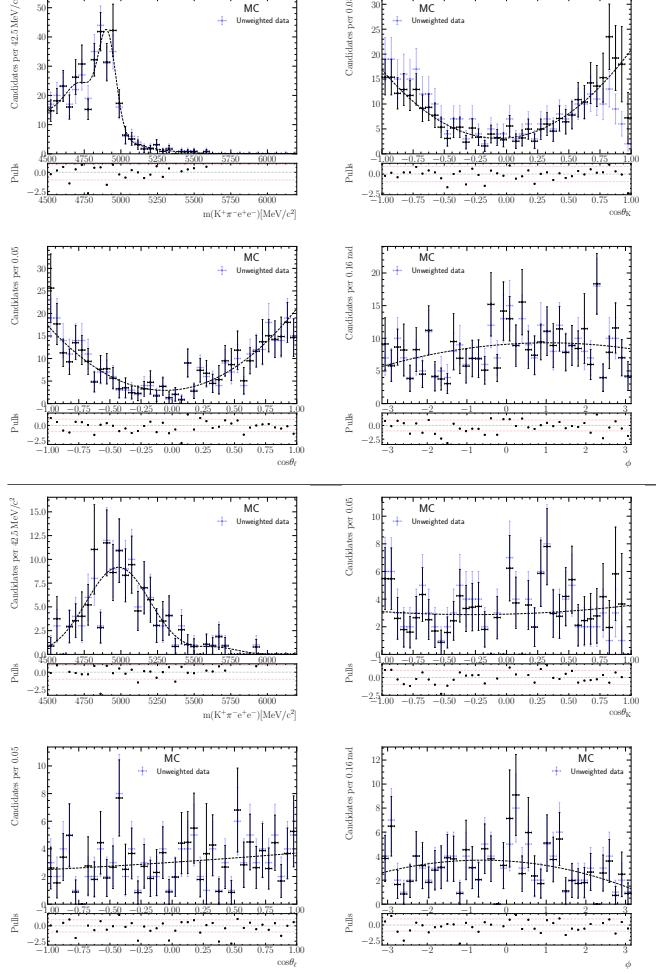


Figure 190: Mass and angular distributions of leptonic partially reconstructed decays modelled using inclusive simulation of B^0 (top), B^+ (middle) and B_s^0 (bottom) decays for the q_c^2 range of $7.0 < q_c^2 < 11.0$ (without cut on $m_B^{J/\psi}$). A single mass and angular model is obtained using all available samples, including simulation samples from years 2017 and 2018 (processed in the same way as 2015-16 samples). The three components are combined into one single function to describe the PRL component.

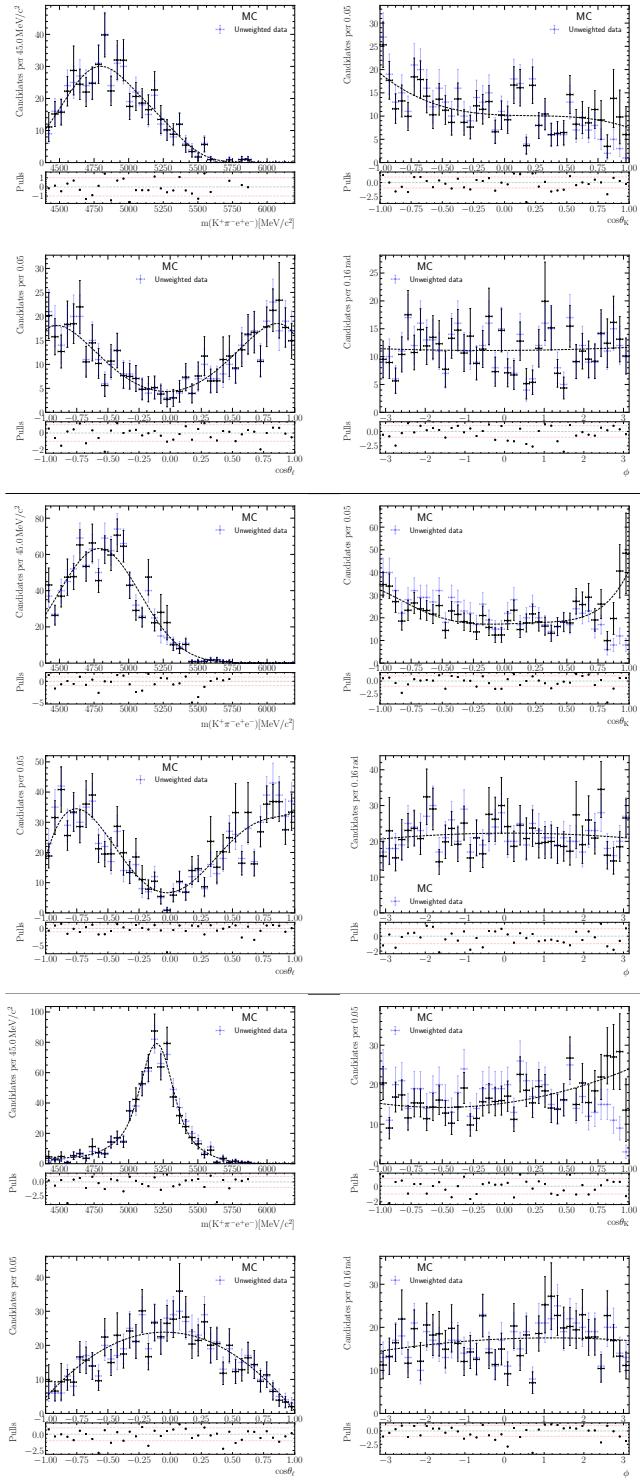


Figure 191: Mass and angular distributions of leptonic partially reconstructed decays modelled using inclusive simulation of B^0 (top), B^+ (middle) and B_s^0 (bottom) decays for the q_c^2 range of $7.0 < q_c^2 < 11.0$ (with cut on $m_{B_s}^{J/\psi}$). A single mass and angular model is obtained using all available samples, including simulation samples from years 2017 and 2018 (processed in the same way as Run-IIp1 samples). The three components are combined into one single function to describe the PRH component.

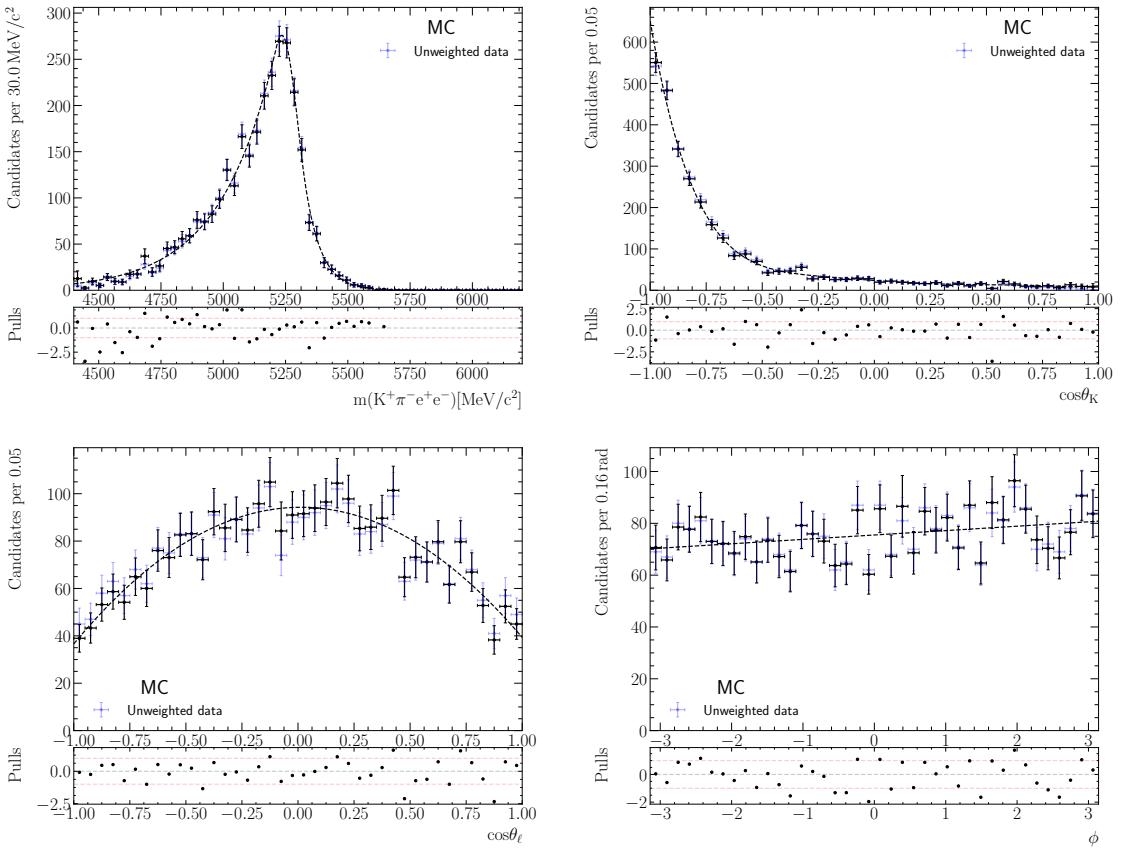


Figure 192: Result of the fit to $B_s^0 \rightarrow \phi J/\psi$ simulation for the unconstrained q^2 range of $6.0 < q^2 < 11.0$. Both Run-I and Run-II (up to 2016) samples are used to extract the shared model.

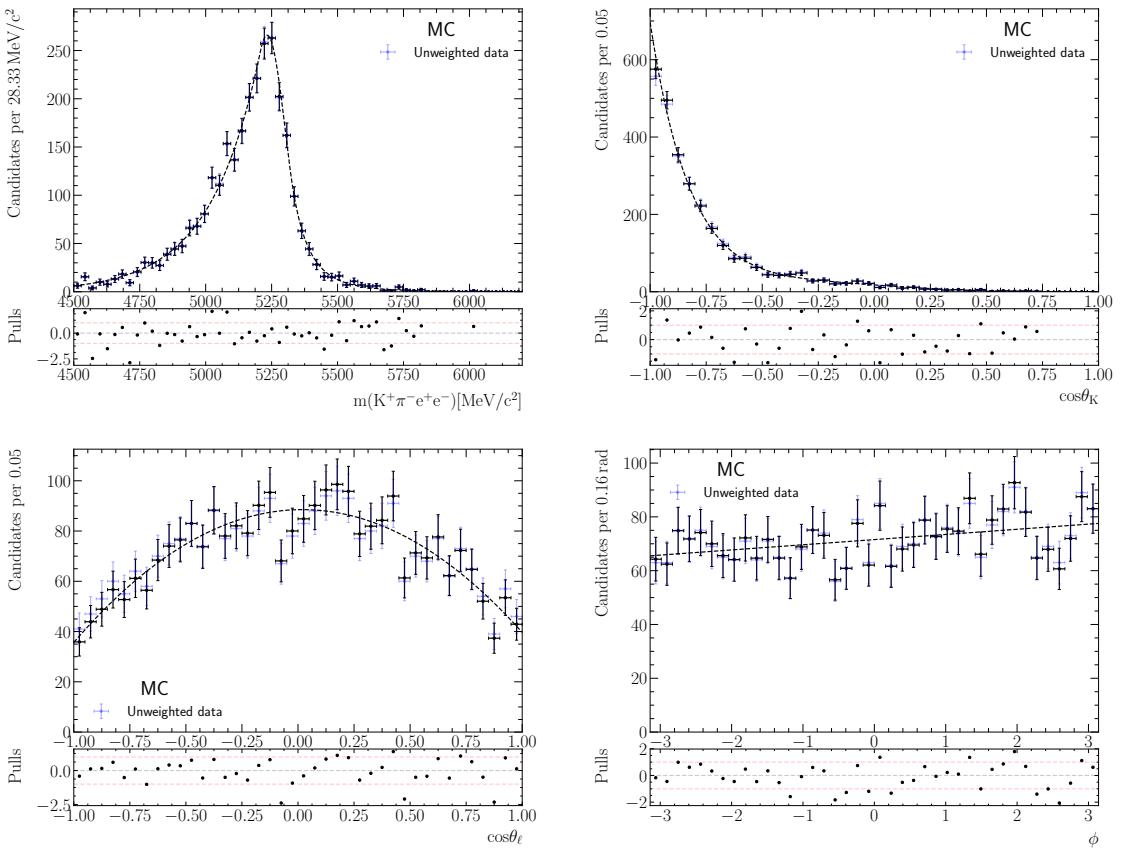


Figure 193: Result of the fit to $B_s^0 \rightarrow \phi J/\psi$ simulation for the q_c^2 range of $7.0 < q_c^2 < 11.0$ (without cut on $m_B^{J/\psi}$). Both Run-I and Run-II (up to 2016) samples are used to extract the shared model.

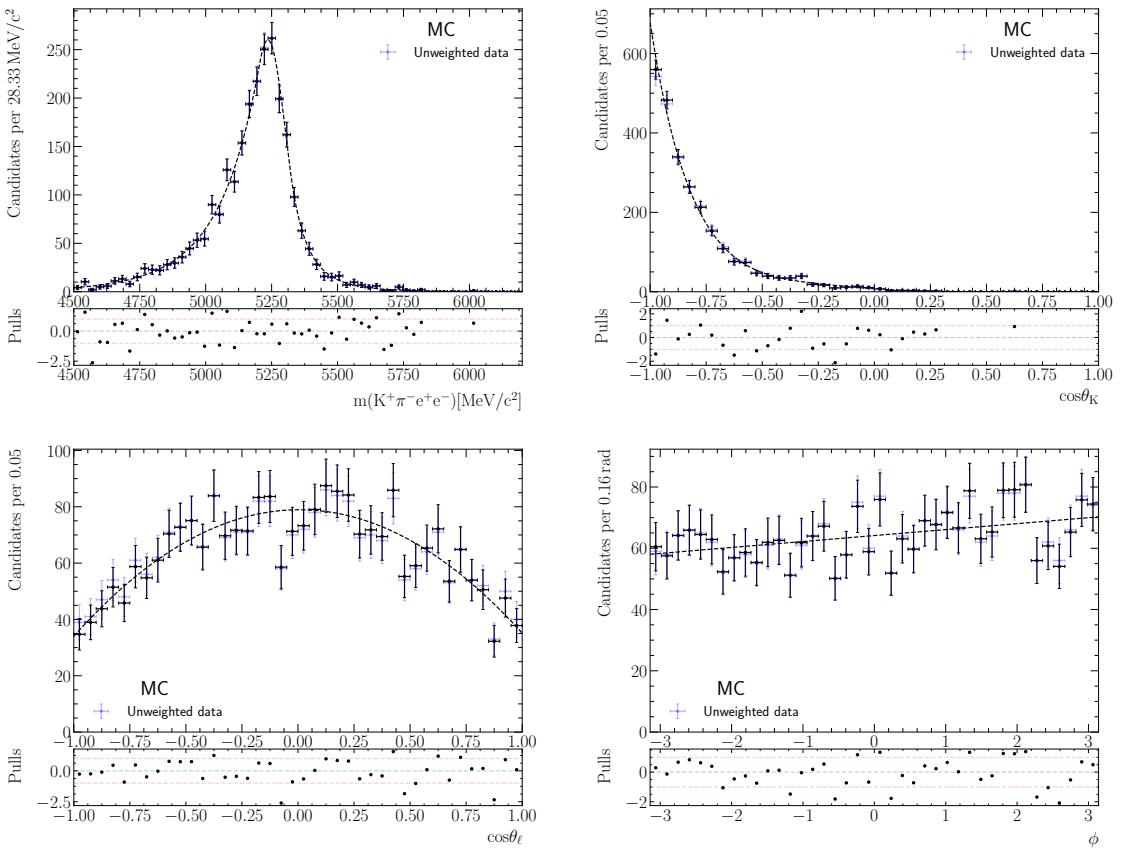


Figure 194: Result of the fit to $B_s^0 \rightarrow \phi J/\psi$ simulation for the q_c^2 range of $7.0 < q_c^2 < 11.0$ (with cut on $m_{B_s}^{J/\psi}$). Both Run-I and Run-II (up to 2016) samples are used to extract the shared model.

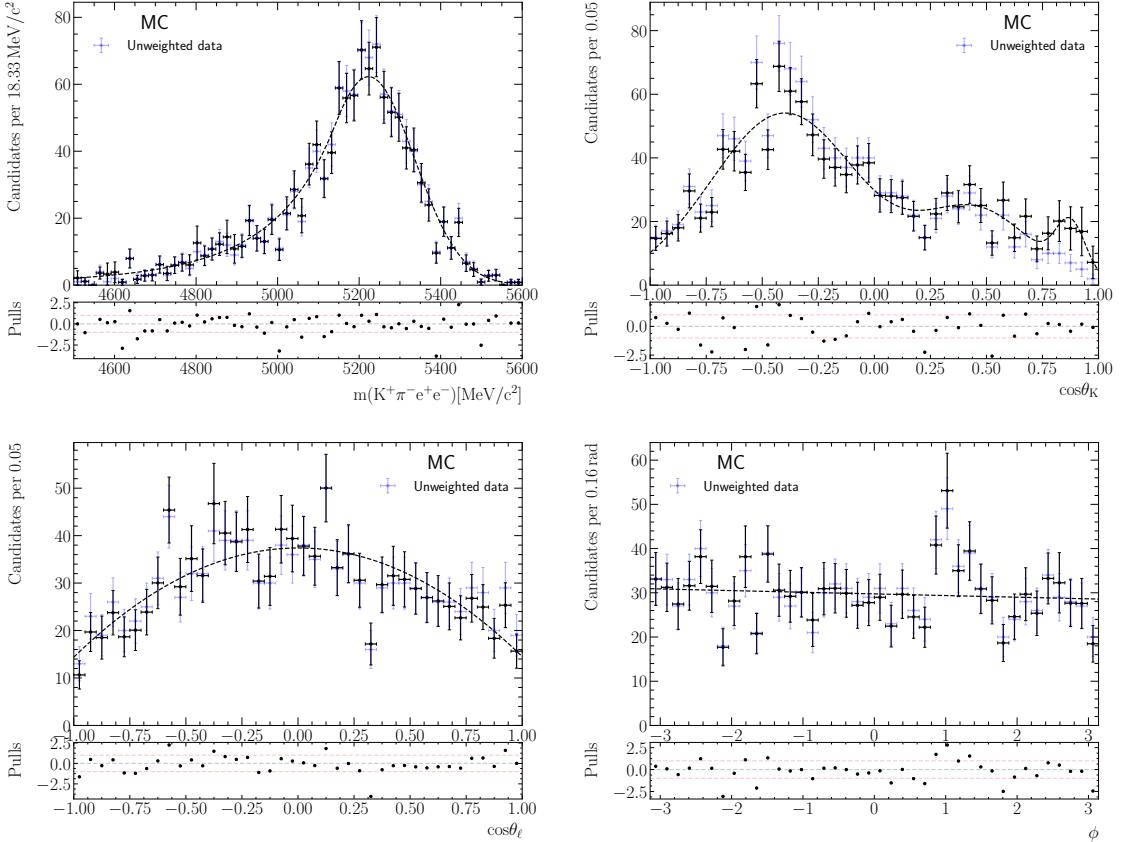


Figure 195: Mass and angular distributions of the $K \rightarrow \pi$ swap background for the unconstrained q^2 range of $6.0 < q_c^2 < 11.0$. Truth-matched control mode simulation samples are used, where the kaon is required to have the true ID of a pion and vice-versa. Due to low statistics, the fit is made to all available samples (Run-I and Run-II up to 2016), and one shape is used for both Runs.

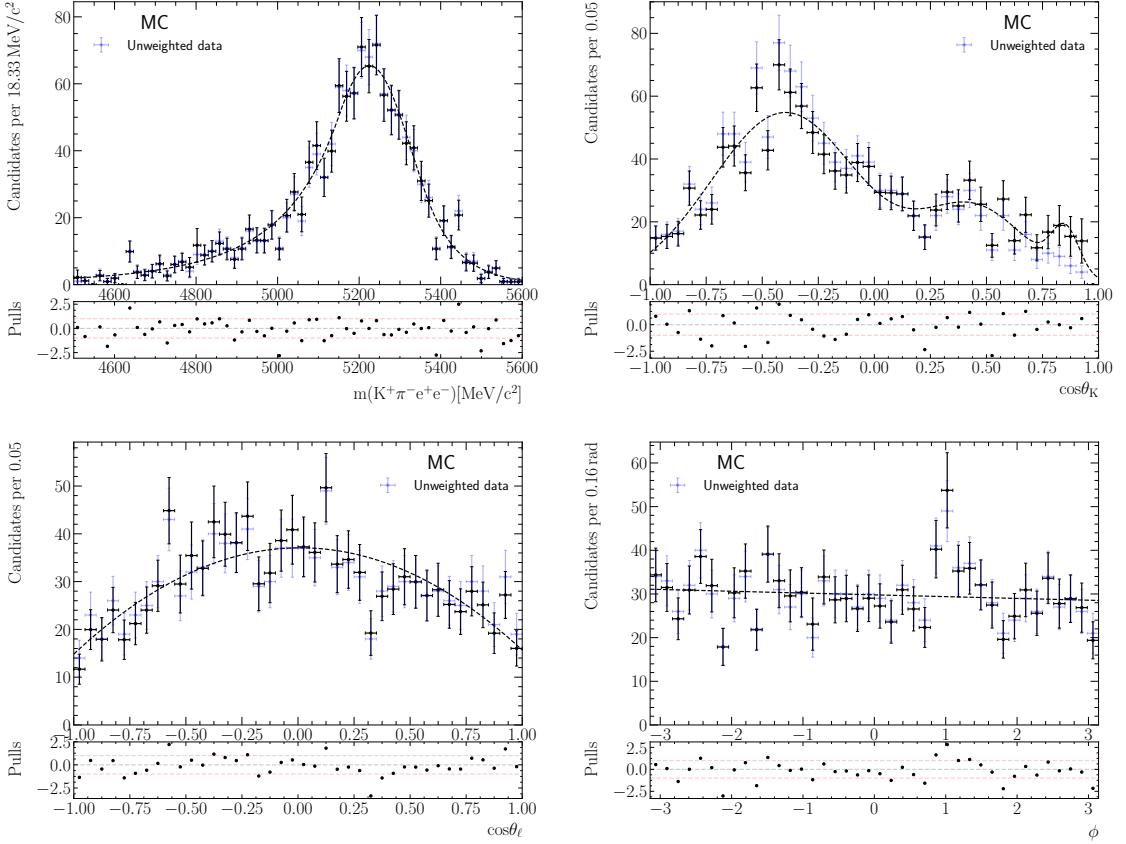


Figure 196: Mass and angular distributions of the $K \rightarrow \pi$ swap background for the q_c^2 range of $7.0 < q_c^2 < 11.0$ (without cut on $m_B^{J/\psi}$). Truth-matched control mode simulation samples are used, where the kaon is required to have the true ID of a pion and vice-versa. Due to low statistics, the fit is made to all available samples (Run-I and Run-II up to 2016), and one shape is used for both Runs.

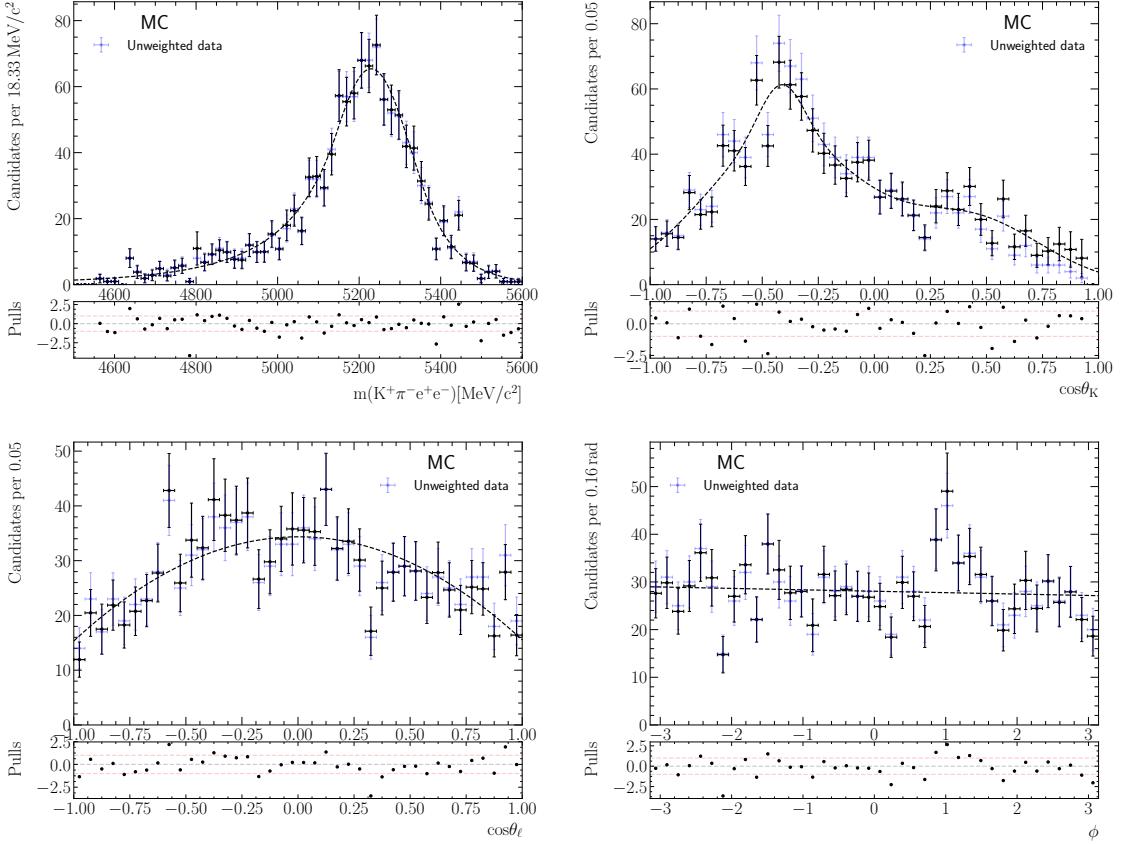


Figure 197: Mass and angular distributions of the $K \rightarrow \pi$ swap background for the q_c^2 range of $7.0 < q_c^2 < 11.0$ (with cut on $m_B^{J/\psi}$). Truth-matched control mode simulation samples are used, where the kaon is required to have the true ID of a pion and vice-versa. Due to low statistics, the fit is made to all available samples (Run-I and Run-II up to 2016), and one shape is used for both Runs.

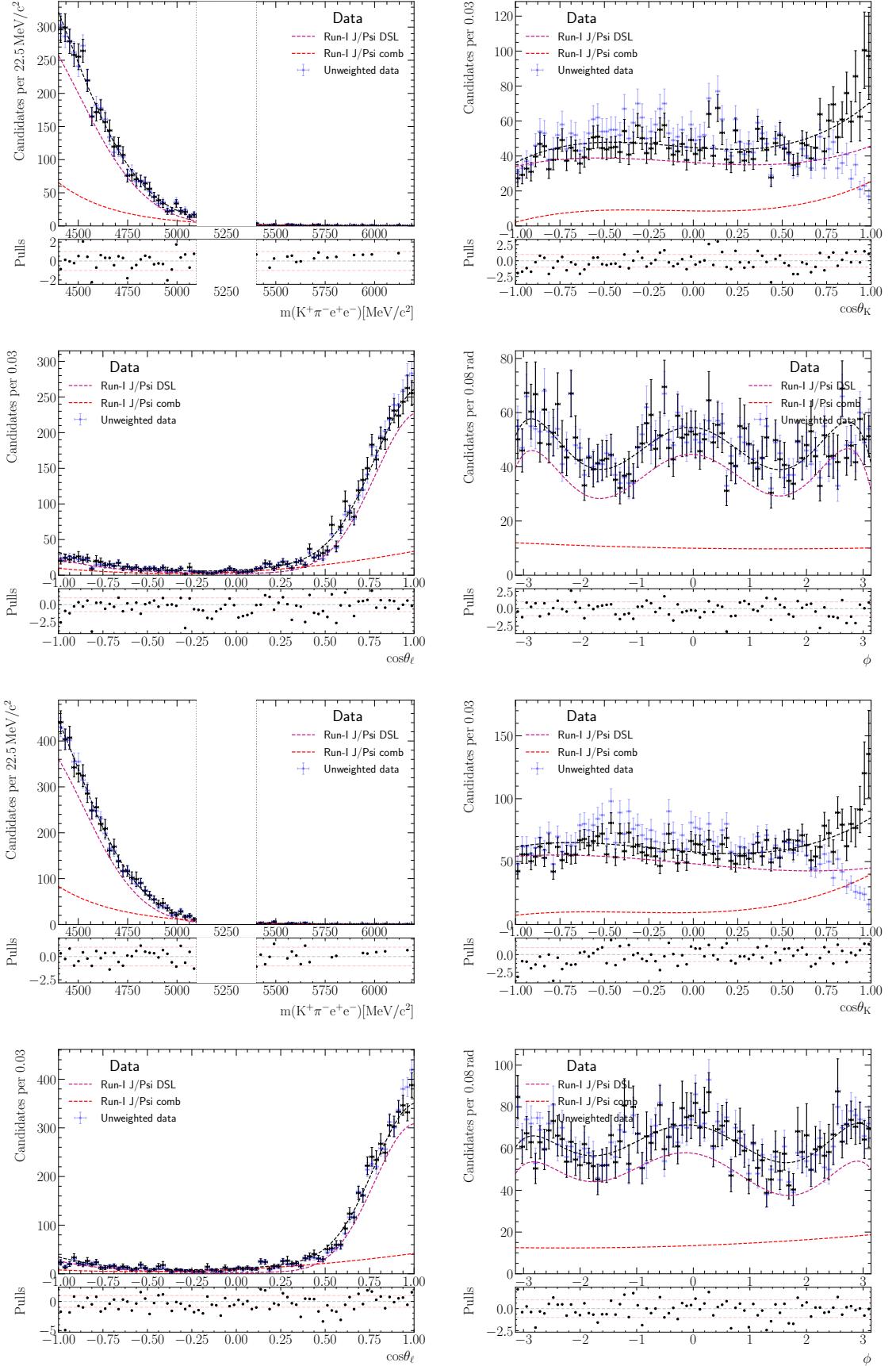


Figure 198: Fit results from step two of the two-step procedure used to extract the DSL model in the unconstrained q^2 range of 6.0 to 11.0 GeV^2/c^4 . To take into account the warping of the DSL mass distribution due to the extended mass window (closer to the peak of the distribution), a Gaussian function is used instead of an exponential for this component.

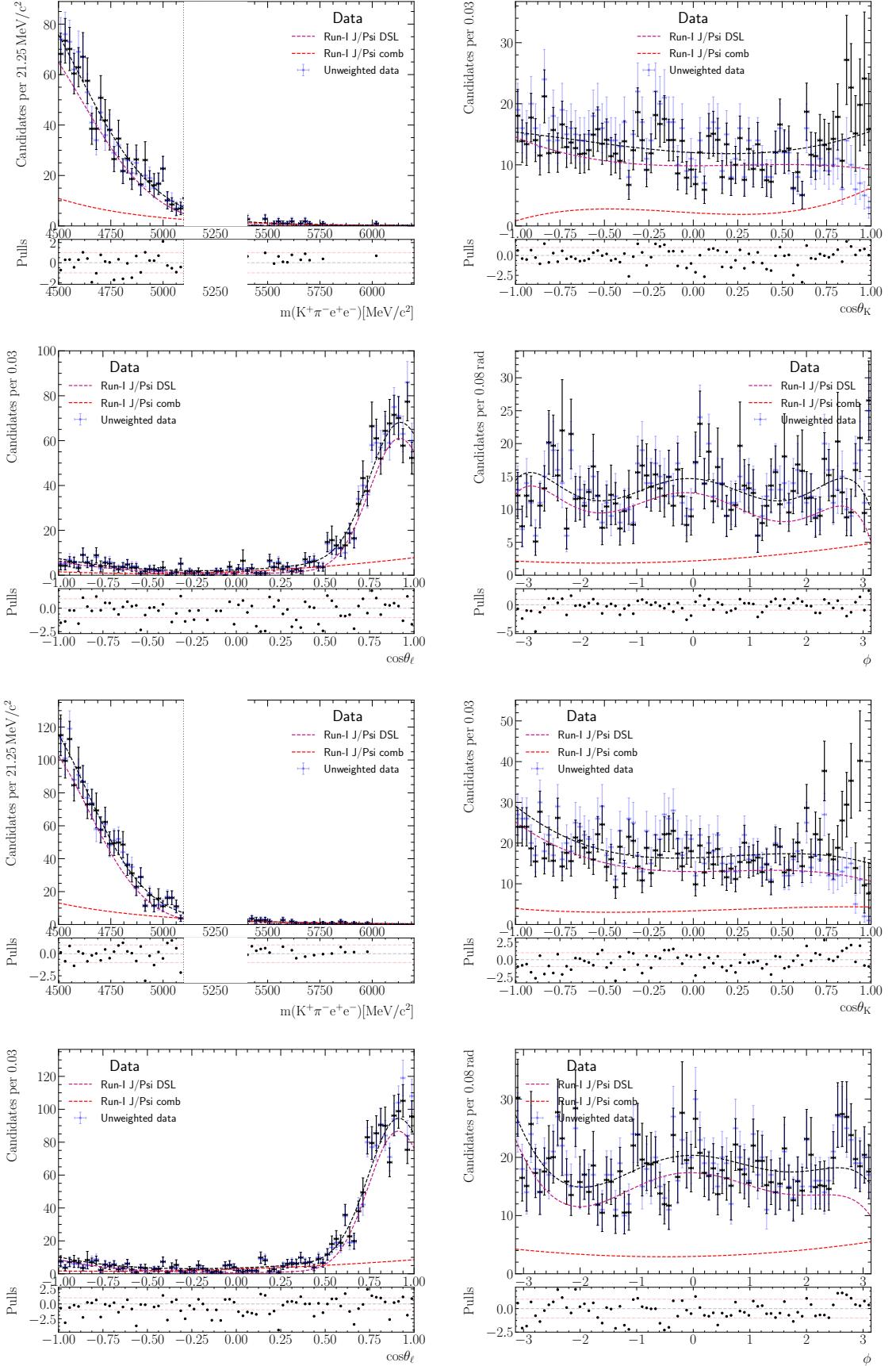


Figure 199: Fit results from step two of the two-step procedure used to extract the DSL model in the q_c^2 range of 7.0 to 11.0 GeV^2/c^4 . To take into account the warping of the DSL mass distribution due to the extended mass window (q_c^2 closer to the peak of the distribution), a Gaussian function is used instead of an exponential for this component.

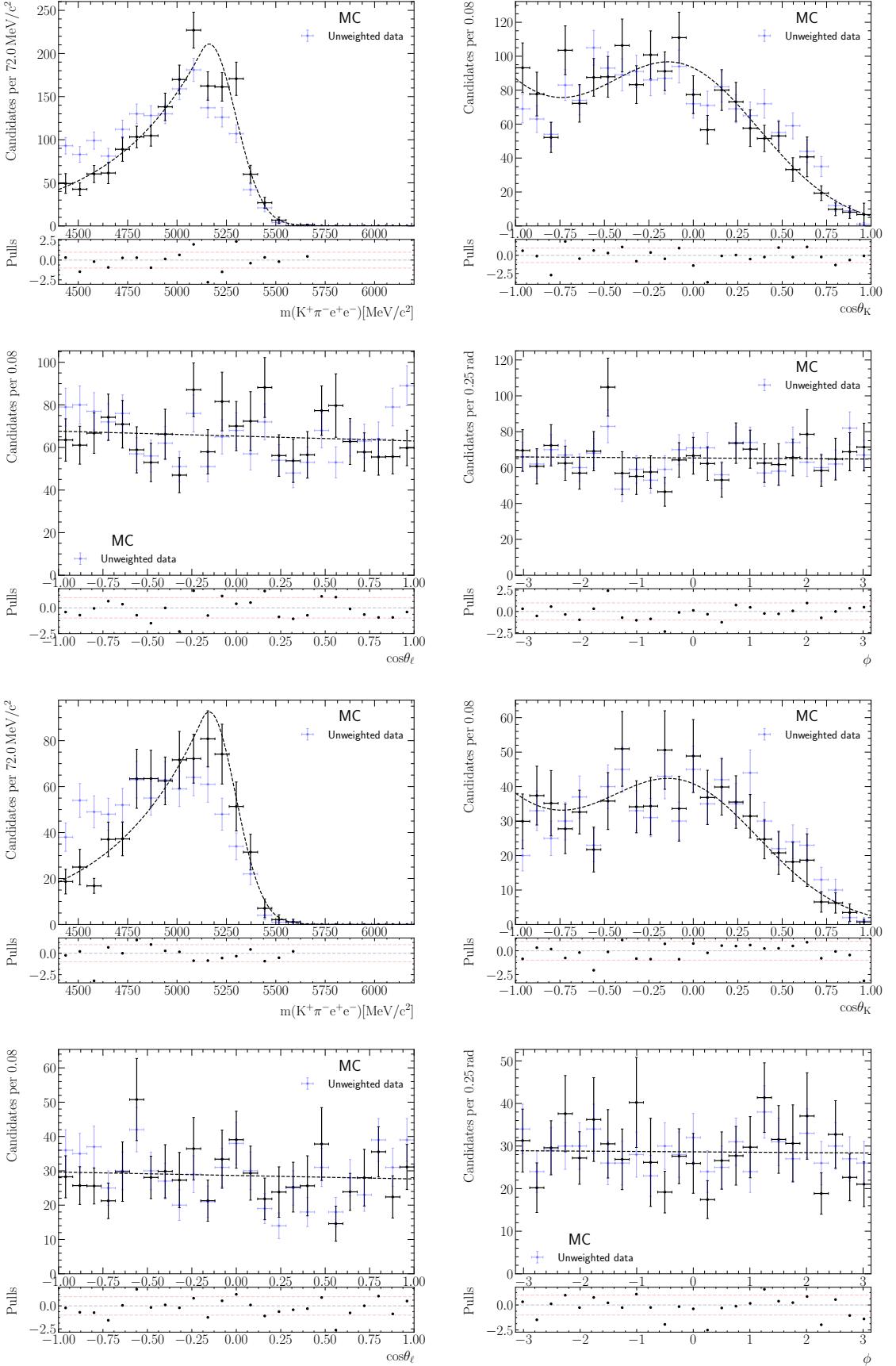


Figure 200: Mass and angular model used for the Λ_b^0 mis-id component for the unconstrained q^2 range of 6.0 to 11.0 GeV^2/c^4 . Due to low statistics, the shapes are extracted from a simultaneous fit to Run-I and 2016 simulation samples. Data-driven correction weights are used in addition to acceptance weights.

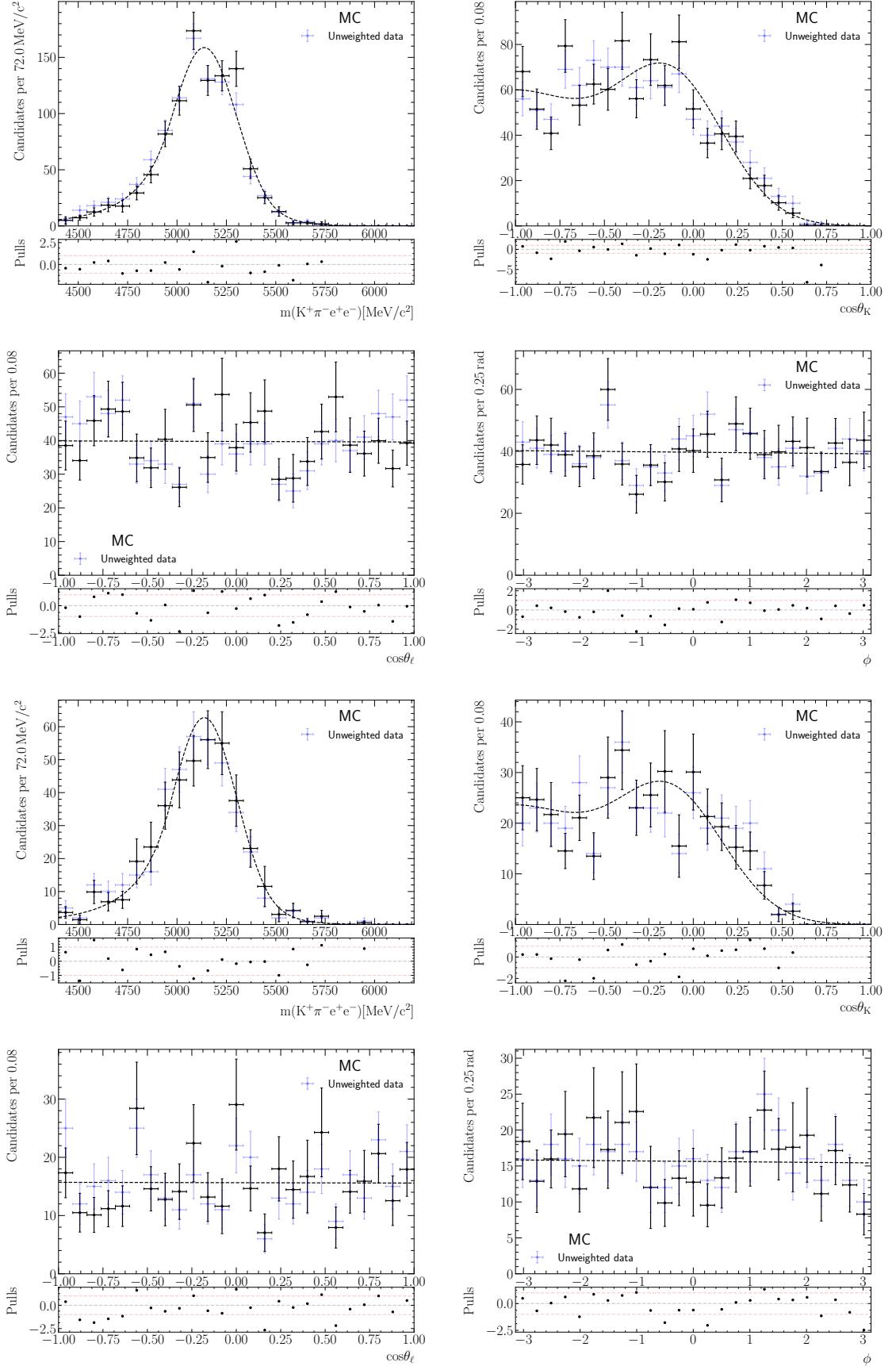


Figure 201: Mass and angular model used for the Λ_b^0 mis-id component for the q_c^2 range of 7.0 to 11.0 GeV^2/c^4 (without cut on $m_B^{J/\psi}$). Due to low statistics, the shapes are extracted from a simultaneous fit to Run-I and 2016 simulation samples. Data-driven correction weights are used in addition to acceptance weights.

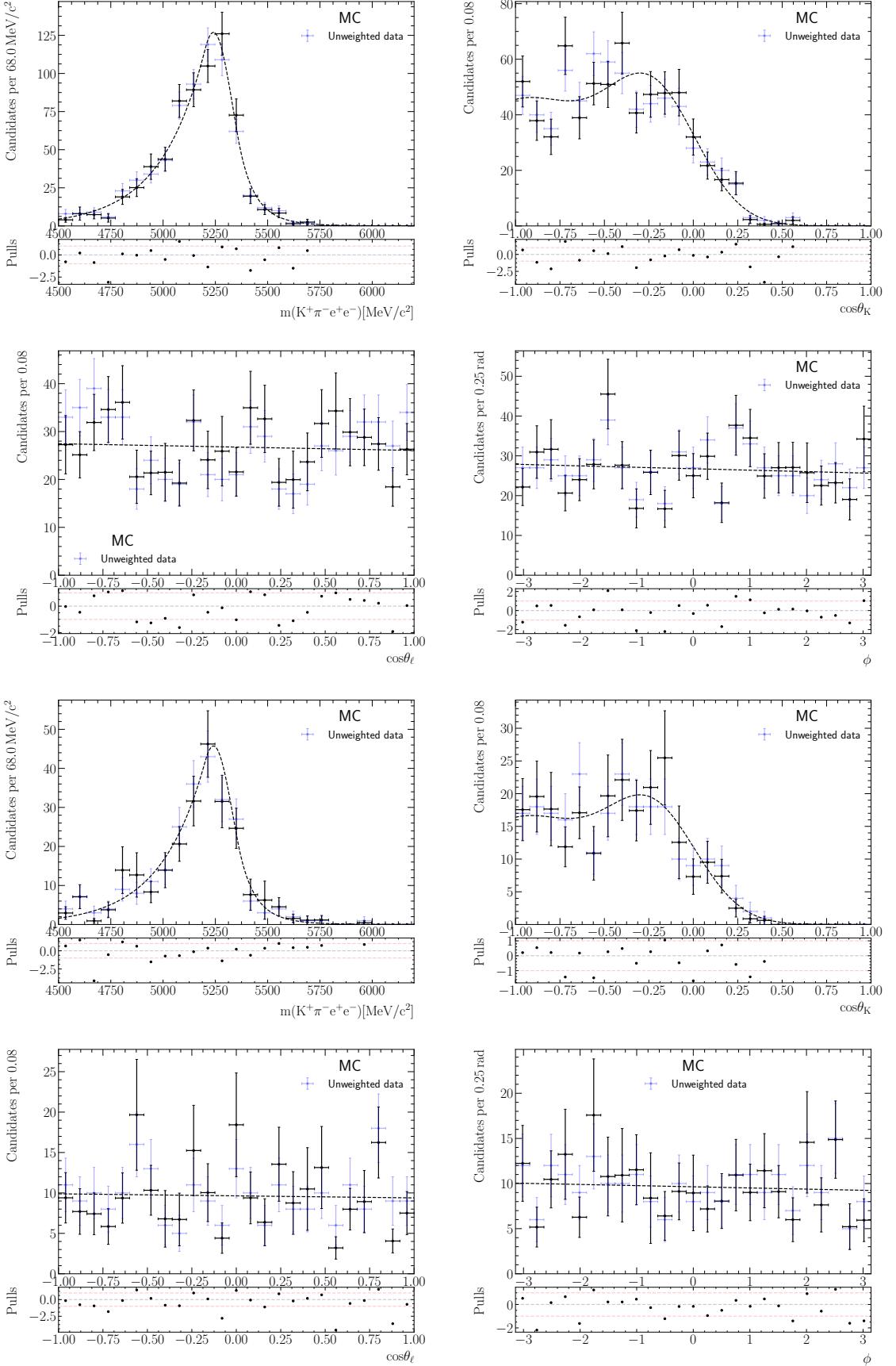


Figure 202: Mass and angular model used for the Λ_b^0 mis-id component for the q_c^2 range of 7.0 to 11.0 GeV^2/c^4 (with cut on $m_B^{J/\psi}$). Due to low statistics, the shapes are extracted from a simultaneous fit to Run-I and 2016 simulation samples. Data-driven correction weights are used in addition to acceptance weights.

³⁴¹³ **J.4.1 Unconstrained q^2 range**

³⁴¹⁴ Table 69 provides additional parameter outputs of the fit to the unconstrained q^2 range
³⁴¹⁵ with additional backgrounds.

	Result	
$f_{\text{sig}}^{\text{Run-I}}$	0.4928	± 0.0031
$f_{\text{sig}}^{\text{Run-II}}$	0.4803	± 0.0029
$f_{\text{DSL wrt sig}}^{\text{Run-I}}$	0.0310	± 0.0030
$f_{\text{DSL wrt sig}}^{\text{Run-II}}$	0.0293	± 0.0023
$f_{\text{PRH wrt sig}}^{\text{Run-I}}$	0.391	± 0.014
$f_{\text{PRH wrt sig}}^{\text{Run-II}}$	0.426	± 0.014
$f_{\text{PRL wrt sig}}^{\text{Run-I}}$	0.355	± 0.011
$f_{\text{PRL wrt sig}}^{\text{Run-II}}$	0.385	± 0.011
$\delta_{\mu}^{\text{Run-I}}$	-7.1	± 0.6
$\delta_{\mu}^{\text{Run-II}}$	-16.3	± 0.6
$\delta_{\sigma}^{\text{Run-I}}$	1.074	± 0.010
$\delta_{\sigma}^{\text{Run-II}}$	1.040	± 0.009
$\lambda_{\text{Comb}}^{\text{Run-I}}$	-0.005 19	$\pm 0.000 14$
$\lambda_{\text{Comb}}^{\text{Run-II}}$	-0.005 71	$\pm 0.000 16$
$c_1^K; \text{Comb shared}$	0.497	± 0.026
$c_1^\ell; \text{Comb shared}$	-0.035	± 0.027
$c_1^\phi; \text{Comb shared}$	0.019	± 0.021

Table 69: Nuisance parameters obtained from the fit made in the unconstrained q^2 range of $6.0 < q^2 < 11.0$.

³⁴¹⁶ **J.4.2 Constrained q^2 without constrained B^0 mass cut**

³⁴¹⁷ Table 70 provides additional parameter outputs of the fit to the constrained q_c^2 range of
³⁴¹⁸ $7.0 < q_c^2 < 11.0$ with additional backgrounds.

³⁴¹⁹ **J.4.3 Constrained q^2 with constrained B^0 mass cut**

³⁴²⁰ Table 71 provides additional parameter outputs of the fit to the constrained q_c^2 range of
³⁴²¹ $7.0 < q_c^2 < 11.0$ with cut on constrained B^0 mass of $m_B^{J/\psi} > 5150 \text{ MeV}/c^2$.

	Result
$f_{\text{sig}}^{\text{Run-I}}$	0.864 ± 0.004
$f_{\text{sig}}^{\text{Run-II}}$	0.842 ± 0.004
$f_{\text{PRL wrt sig}}^{\text{shared}}$	0.0287 ± 0.0030
$\delta_{\mu}^{\text{Run-I}}$	-7.6 ± 0.6
$\delta_{\mu}^{\text{Run-II}}$	-16.4 ± 0.5
$\delta_{\sigma}^{\text{Run-I}}$	1.077 ± 0.009
$\delta_{\sigma}^{\text{Run-II}}$	1.015 ± 0.008
$\lambda_{\text{Comb}}^{\text{Run-I}}$	-0.0059 ± 0.0006
$\lambda_{\text{Comb}}^{\text{Run-II}}$	-0.0060 ± 0.0006
c_1^K ; Comb shared	0.29 ± 0.12
c_2^K ; Comb shared	0.15 ± 0.15
c_1^{ℓ} ; Comb shared	0.90 ± 0.13
c_2^{ℓ} ; Comb shared	0.81 ± 0.12
c_1^{ϕ} ; Comb shared	-0.21 ± 0.10
c_2^{ϕ} ; Comb shared	0.20 ± 0.10

Table 70: Nuisance parameters obtained from the fit made in the constrained q_c^2 range of $7.0 < q_c^2 < 11.0$.

	Result
$f_{\text{sig}}^{\text{Run-I}}$	0.9539 ± 0.0035
$f_{\text{sig}}^{\text{Run-II}}$	0.956 ± 0.004
$f_{\text{DSL wrt sig}}^{\text{shared}}$	0.0062 ± 0.0006
$\delta_{\mu}^{\text{Run-I}}$	-8.0 ± 0.5
$\delta_{\mu}^{\text{Run-II}}$	-18.3 ± 0.5
$\delta_{\sigma}^{\text{Run-I}}$	1.080 ± 0.008
$\delta_{\sigma}^{\text{Run-II}}$	1.042 ± 0.008

Table 71: Nuisance parameters obtained from the fit made in the constrained q_c^2 range of $7.0 < q_c^2 < 11.0$ with cut on the constrained B^0 mass of $m_B^{J/\psi} > 5150 \text{ MeV}/c^2$.

3422 **K Reweighting of the hadronic system Dalitz plots**

3423 The following information is taken from the $R_{K\pi\pi}$ analysis.

3424
3425 As introduced in Section 3, the simulated samples do not correctly model the resonance
3426 structure of the hadronic system that can be observed in data. In order to reproduce
3427 this structure, the simulated samples are reweighted to resemble background subtracted
3428 data from the resonant $B^+ \rightarrow K^+\pi^+\pi^- J/\psi(\rightarrow \mu^+\mu^-)$ mode, using a boosted decision tree
3429 reweighter.

3430 To account for acceptance effects due to reconstruction, Stripping and the pre-selection,
3431 efficiency maps are calculated in bins of the $K^+\pi^-$ and $\pi^+\pi^-$ invariant masses and then
3432 these efficiencies are used to unfold the data sample. An isopopulated binning scheme for
3433 the efficiency maps is calculated using the TKDTreeBinning class included in ROOT. The
3434 efficiencies are then calculated using generator level simulation and reconstructed, selected
3435 simulation. The numerator histogram, denominator histogram and the final efficiency
3436 maps are shown in Figure 203.

3437 A Gradient Boosted Reweighting is trained on the sweighted and unfolded resonant
3438 data and a generator level sample of simulated resonant decays. For input variables the
3439 invariant masses $m_{K^+\pi^-}$, $m_{\pi^+\pi^-}$ and $m_{K^+\pi^+\pi^-}$ are used. The BDT reweighter is evaluated
3440 using cross-validation with 10 folds and when applying the trained BDT to unseen data,
3441 the average of the output of all BDTs (of each fold) is applied as per-event weight. The
3442 comparison between phase-space reweighted MC and background-subtracted collision data
3443 distributions is shown in Figure 204.

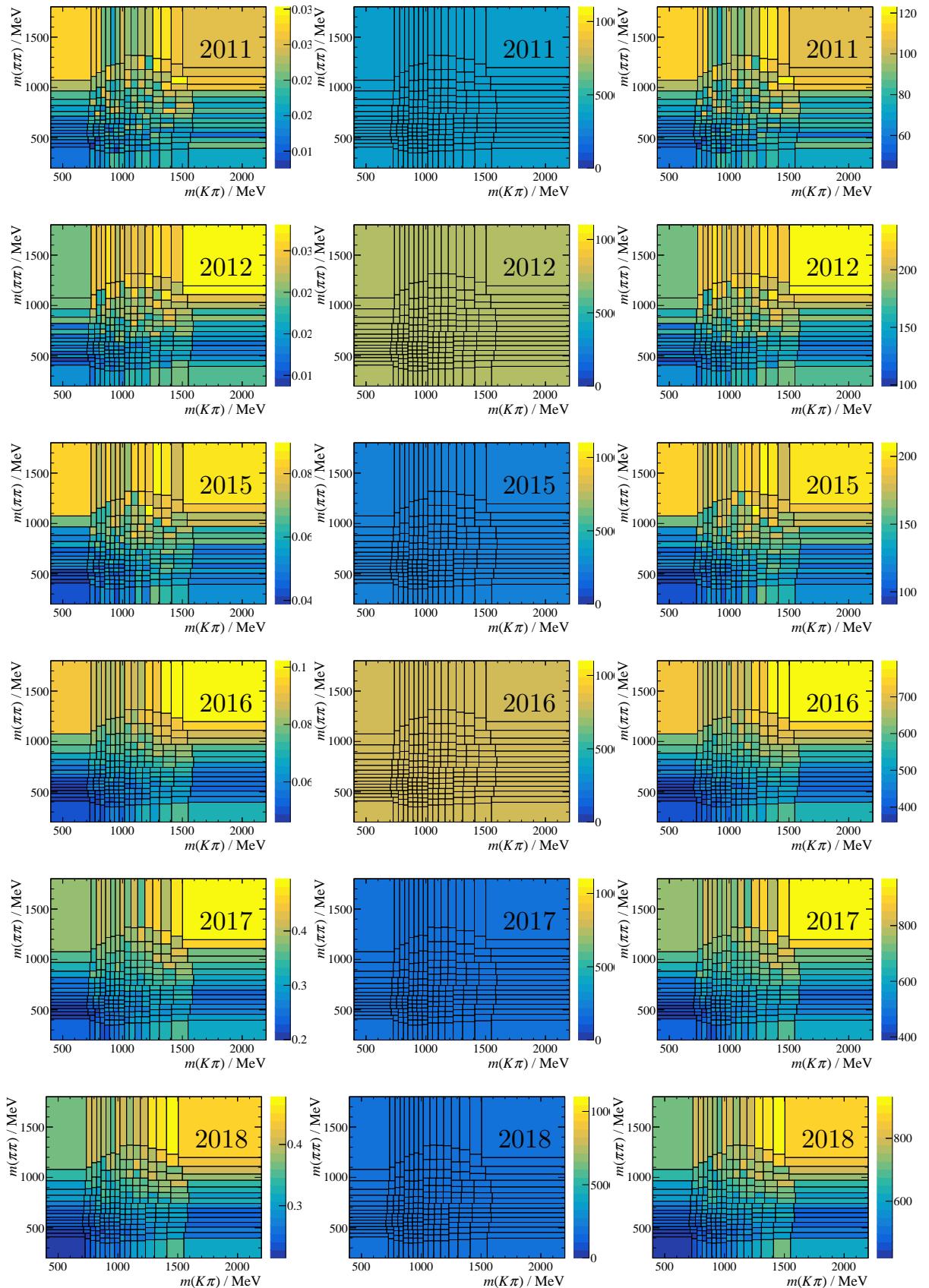


Figure 203: Efficiency maps of reconstruction, stripping and pre-selection in bins of $K^+\pi^-$ and $\pi^+\pi^-$ invariant mass.

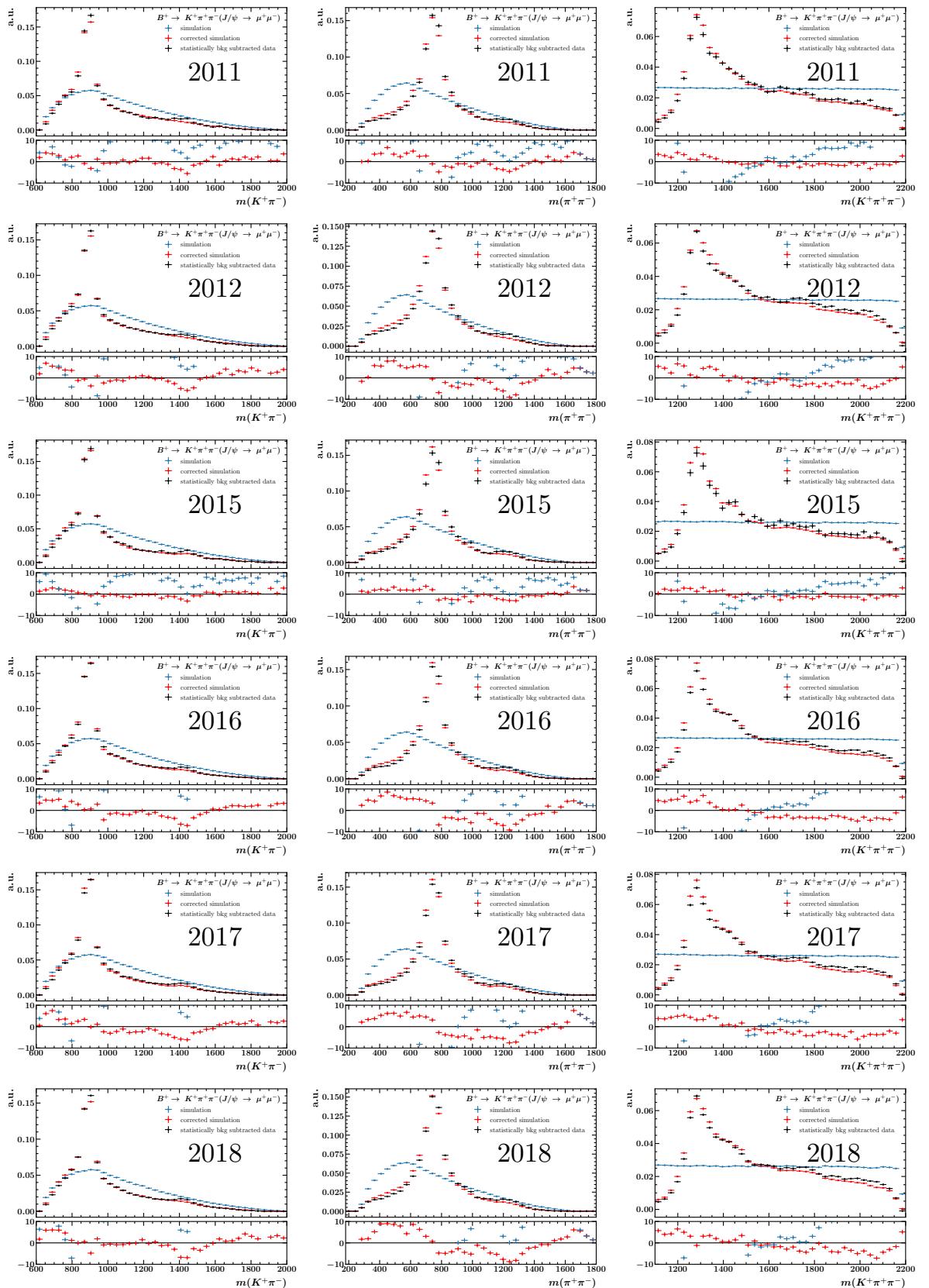


Figure 204: Comparison of the $m_{K^+\pi^-}$, $m_{\pi^+\pi^-}$ and $m_{K^+\pi^+\pi^-}$ distribution between background-subtracted collision data (black), unweighted MC simulation (blue) and reweighted MC simulation (red).²²⁹

3444 References

- 3445 [1] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron*
3446 *Symmetry*, Phys. Rev. **D2** (1970) 1285.
- 3447 [2] Y. Grossman and M. P. Worah, *CP asymmetries in B decays with new physics in*
3448 *decay amplitudes*, Phys. Lett. **B395** (1997) 241, arXiv:hep-ph/9612269.
- 3449 [3] R. Fleischer, *CP violation and the role of electroweak penguins in nonleptonic B*
3450 *decays*, Int. J. Mod. Phys. **A12** (1997) 2459, arXiv:hep-ph/9612446.
- 3451 [4] D. London and A. Soni, *Measuring the CP angle β in hadronic $b \rightarrow s$ penguin decays*,
3452 Phys. Lett. **B407** (1997) 61, arXiv:hep-ph/9704277.
- 3453 [5] M. Ciuchini *et al.*, *CP violating B decays in the standard model and supersymmetry*,
3454 Phys. Rev. Lett. **79** (1997) 978, arXiv:hep-ph/9704274.
- 3455 [6] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay*
3456 *using 3 fb^{-1} of integrated luminosity*, JHEP **02** (2016) 104, arXiv:1512.04442.
- 3457 [7] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP-averaged observables in the*
3458 *$B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay*, Phys. Rev. Lett. **125** (2020) 011802, arXiv:2003.04831.
- 3459 [8] LHCb, R. Aaij *et al.*, *Measurement of Form-Factor-Independent Observables in the*
3460 *Decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$* , Phys. Rev. Lett. **111** (2013) 191801, arXiv:1308.1707.
- 3461 [9] S. Descotes-Genon, T. Hurth, J. Matias, and J. Virto, *Optimizing the basis of $B \rightarrow$*
3462 *K^{*ll} observables in the full kinematic range*, JHEP **05** (2013) 137, arXiv:1303.5794.
- 3463 [10] Belle, S. Wehle *et al.*, *Lepton-Flavor-Dependent Angular Analysis of $B \rightarrow K^* \ell^+ \ell^-$* ,
3464 arXiv:1612.05014.
- 3465 [11] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, *Global analysis of $b \rightarrow s \ell \ell$*
3466 *anomalies*, JHEP **06** (2016) 092, arXiv:1510.04239.
- 3467 [12] J. Aebischer *et al.*, *B-decay discrepancies after Moriond 2019*, Eur. Phys. J. C **80**
3468 (2020), no. 3 252, arXiv:1903.10434.
- 3469 [13] W. Altmannshofer and P. Stangl, *New Physics in Rare B Decays after Moriond 2021*,
3470 arXiv:2103.13370.
- 3471 [14] LHCb, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the decay*
3472 *$B_s^0 \rightarrow \phi \mu^+ \mu^-$* , JHEP **07** (2013) 084, arXiv:1305.2168.
- 3473 [15] LHCb, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of*
3474 *$B \rightarrow K^{(*)} \mu^+ \mu^-$ decays*, JHEP **06** (2014) 133, arXiv:1403.8044.

- 3475 [16] LHCb, R. Aaij *et al.*, *Differential branching fraction and angular analysis of*
 3476 $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ *decays*, JHEP **06** (2015) 115, arXiv:1503.07138, [Erratum:
 3477 JHEP09,145(2018)].
- 3478 [17] LHCb, R. Aaij *et al.*, *Measurements of the S-wave fraction in $B^0 \rightarrow K^+\pi^-\mu^+\mu^-$*
 3479 *decays and the $B^0 \rightarrow K^*(892)^0\mu^+\mu^-$ differential branching fraction*, JHEP **11** (2016)
 3480 047, arXiv:1606.04731, [Erratum: JHEP04,142(2017)].
- 3481 [18] R. Gauld, F. Goertz, and U. Haisch, *An explicit Z' -boson explanation of the $B \rightarrow$*
 3482 $K^*\mu^+\mu^-$ *anomaly*, JHEP **01** (2014) 069, arXiv:1310.1082.
- 3483 [19] A. J. Buras and J. Girrbach, *Left-handed Z' and Z FCNC quark couplings facing new*
 3484 $b \rightarrow s\mu^+\mu^-$ *data*, JHEP **12** (2013) 009, arXiv:1309.2466.
- 3485 [20] W. Altmannshofer and D. M. Straub, *New Physics in $B \rightarrow K^*\mu\mu?$* , Eur. Phys. J.
 3486 **C73** (2013) 2646, arXiv:1308.1501.
- 3487 [21] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, *Quark flavor transitions in*
 3488 $L_\mu - L_\tau$ *models*, Phys. Rev. D **89** (2014) 095033, arXiv:1403.1269.
- 3489 [22] A. Crivellin *et al.*, *Lepton-flavour violating B decays in generic Z' models*, Phys. Rev.
 3490 **D92** (2015), no. 5 054013, arXiv:1504.07928.
- 3491 [23] W. Altmannshofer, J. Davighi, and M. Nardecchia, *Gauging the accidental symmetries*
 3492 *of the standard model, and implications for the flavor anomalies*, Phys. Rev. D **101**
 3493 (2020), no. 1 015004, arXiv:1909.02021.
- 3494 [24] G. Hiller and M. Schmaltz, *R_K and future $b \rightarrow s\ell\ell$ physics beyond the standard model*
 3495 *opportunities*, Phys. Rev. **D90** (2014) 054014, arXiv:1408.1627.
- 3496 [25] S. Biswas, D. Chowdhury, S. Han, and S. J. Lee, *Explaining the lepton non-*
 3497 *universality at the LHCb and CMS within a unified framework*, JHEP **02** (2015)
 3498 142, arXiv:1409.0882.
- 3499 [26] B. Gripaios, M. Nardecchia, and S. A. Renner, *Composite leptoquarks and anomalies*
 3500 *in B -meson decays*, JHEP **05** (2015) 006, arXiv:1412.1791.
- 3501 [27] M. Freytsis, Z. Ligeti, and J. T. Ruderman, *Flavor models for $\bar{B} \rightarrow D^{(*)}\tau\bar{\nu}$* , Phys.
 3502 Rev. D **92** (2015), no. 5 054018, arXiv:1506.08896.
- 3503 [28] B. Fornal, S. A. Gadam, and B. Grinstein, *Left-Right $SU(4)$ Vector Leptoquark Model*
 3504 *for Flavor Anomalies*, Phys. Rev. D **99** (2019), no. 5 055025, arXiv:1812.01603.
- 3505 [29] D. Bećirević *et al.*, *Scalar leptoquarks from grand unified theories to accommodate the*
 3506 *B -physics anomalies*, Phys. Rev. D **98** (2018), no. 5 055003, arXiv:1806.05689.

- 3507 [30] M. Blanke and A. Crivellin, *B Meson Anomalies in a Pati-Salam Model within*
 3508 *the Randall-Sundrum Background*, Phys. Rev. Lett. **121** (2018), no. 1 011801,
 3509 [arXiv:1801.07256](#).
- 3510 [31] V. Gherardi, D. Marzocca, and E. Venturini, *Low-energy phenomenology of scalar*
 3511 *leptoquarks at one-loop accuracy*, JHEP **01** (2021) 138, [arXiv:2008.09548](#).
- 3512 [32] L. Di Luzio, A. Greljo, and M. Nardecchia, *Gauge leptoquark as the origin of B-physics*
 3513 *anomalies*, Phys. Rev. D **96** (2017), no. 11 115011, [arXiv:1708.08450](#).
- 3514 [33] M. Bordone, C. Cornella, J. Fuentes-Martín, and G. Isidori, *Low-energy signatures*
 3515 *of the PS³ model: from B-physics anomalies to LFV*, JHEP **10** (2018) 148,
 3516 [arXiv:1805.09328](#).
- 3517 [34] A. Angelescu, D. Bećirević, D. A. Faroughy, and O. Sumensari, *Closing the window*
 3518 *on single leptoquark solutions to the B-physics anomalies*, JHEP **10** (2018) 183,
 3519 [arXiv:1808.08179](#).
- 3520 [35] L. Di Luzio *et al.*, *Maximal Flavour Violation: a Cabibbo mechanism for leptoquarks*,
 3521 JHEP **11** (2018) 081, [arXiv:1808.00942](#).
- 3522 [36] C. Cornella, J. Fuentes-Martin, and G. Isidori, *Revisiting the vector leptoquark*
 3523 *explanation of the B-physics anomalies*, JHEP **07** (2019) 168, [arXiv:1903.11517](#).
- 3524 [37] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y.-M. Wang, *Charm-loop effect*
 3525 *in B → K^(*)ℓ⁺ℓ⁻ and B → K^{*}γ*, JHEP **09** (2010) 089, [arXiv:1006.4945](#).
- 3526 [38] A. Khodjamirian, T. Mannel, and Y. M. Wang, *B → Kℓ⁺ℓ⁻ decay at large hadronic*
 3527 *recoil*, JHEP **02** (2013) 010, [arXiv:1211.0234](#).
- 3528 [39] S. Jäger and J. Martin Camalich, *On B → Vℓℓ at small dilepton invariant mass,*
 3529 *power corrections, and new physics*, JHEP **05** (2013) 043, [arXiv:1212.2263](#).
- 3530 [40] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, *Calculation of B⁰ → K^{*0}μ⁺μ⁻*
 3531 *and B_s⁰ → φμ⁺μ⁻ observables using form factors from lattice QCD*, Phys. Rev. Lett.
 3532 **112** (2014) 212003, [arXiv:1310.3887](#).
- 3533 [41] F. Beaujean, C. Bobeth, and D. van Dyk, *Comprehensive Bayesian analysis of*
 3534 *rare (semi)leptonic and radiative B decays*, Eur. Phys. J. C **74** (2014) 2897,
 3535 [arXiv:1310.2478](#), [Erratum: Eur.Phys.J.C 74, 3179 (2014)].
- 3536 [42] C. Hambrock, G. Hiller, S. Schacht, and R. Zwicky, *B → K^{*} form factors from flavor*
 3537 *data to QCD and back*, Phys. Rev. D **89** (2014), no. 7 074014, [arXiv:1308.4379](#).
- 3538 [43] S. Jäger and J. Martin Camalich, *Reassessing the discovery potential of the B →*
 3539 *K^{*}ℓ⁺ℓ⁻ decays in the large-recoil region: SM challenges and BSM opportunities*, Phys.
 3540 Rev. **D93** (2016), no. 1 014028, [arXiv:1412.3183](#).

- 3541 [44] J. Lyon and R. Zwicky, *Resonances gone topsy turvy - the charm of QCD or new*
 3542 *physics in $b \rightarrow s\ell^+\ell^-$?*, arXiv:1406.0566.
- 3543 [45] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, *On the impact of power*
 3544 *corrections in the prediction of $B \rightarrow K^*\mu^+\mu^-$ observables*, JHEP **12** (2014) 125,
 3545 arXiv:1407.8526.
- 3546 [46] LHCb, R. Aaij *et al.*, *Angular analysis of the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of*
 3547 *integrated luminosity*, JHEP **02** (2016) 104, arXiv:1512.04442.
- 3548 [47] ATLAS, M. Aaboud *et al.*, *Angular analysis of $B_d^0 \rightarrow K^*\mu^+\mu^-$ decays in pp collisions*
 3549 *at $\sqrt{s} = 8 \text{ TeV}$ with the ATLAS detector*, JHEP **10** (2018) 047, arXiv:1805.04000.
- 3550 [48] CMS, A. M. Sirunyan *et al.*, *Measurement of angular parameters from the decay*
 3551 *$B^0 \rightarrow K^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8 \text{ TeV}$* , Phys. Lett. B **781** (2018)
 3552 517, arXiv:1710.02846.
- 3553 [49] C. Bobeth, G. Hiller, and G. Piranishvili, *Angular distributions of $\bar{B} \rightarrow \bar{K}\ell^+\ell^-$ decays*,
 3554 JHEP **12** (2007) 040, arXiv:0709.4174.
- 3555 [50] M. Bordone, G. Isidori, and A. Patti, *On the Standard Model predictions for R_K*
 3556 *and R_{K^*}* , Eur. Phys. J. **C76** (2016), no. 8 440, arXiv:1605.07633.
- 3557 [51] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*,
 3558 arXiv:2103.11769, Submitted to Nature Physics.
- 3559 [52] LHCb, R. Aaij *et al.*, *Test of lepton universality with $B^0 \rightarrow K^{*0}\ell^+\ell^-$ decays*, JHEP
 3560 **08** (2017) 055, arXiv:1705.05802.
- 3561 [53] W. Altmannshofer and D. M. Straub, *Implications of $b \rightarrow s$ measurements*, in *Proceedings, 50th Rencontres de Moriond Electroweak Interactions and Unified Theories: La Thuile, Italy, March 14-21, 2015*, pp. 333–338, 2015. arXiv:1503.06199.
- 3564 [54] B. Capdevila, S. Descotes-Genon, J. Matias, and J. Virto, *Assessing lepton-*
 3565 *flavour non-universality from $B \rightarrow K^*\ell\ell$ angular analyses*, JHEP **10** (2016) 075,
 3566 arXiv:1605.03156.
- 3567 [55] N. Serra, R. Silva Coutinho, and D. van Dyk, *Measuring the breaking of lepton-*
 3568 *flavor universality in $B \rightarrow K^*\ell^+\ell^-$* , Phys. Rev. **D95** (2017), no. 3 035029,
 3569 arXiv:1610.08761.
- 3570 [56] F. Lionetto, *Measurements of Angular Observables of $B^0 \rightarrow K^{*0}\mu^+\mu^-$ and $B^0 \rightarrow$*
 3571 *$K^{*0}e^+e^-$ Decays and the Upgrade of LHCb*, Feb, 2018. Presented 22 Mar 2018.
- 3572 [57] A. Mauri, N. Serra, and R. Silva Coutinho, *Towards establishing lepton flavor uni-*
 3573 *versity violation in $\bar{B} \rightarrow \bar{K}^*\ell^+\ell^-$ decays*, Phys. Rev. D **99** (2019), no. 1 013007,
 3574 arXiv:1805.06401.

- 3575 [58] S. Descotes-Genon, J. Matias, M. Ramon, and J. Virto, *Implications from clean*
 3576 *observables for the binned analysis of $B^- \rightarrow K^* \mu^+ \mu^-$ at large recoil*, JHEP **01** (2013)
 3577 048, [arXiv:1207.2753](#).
- 3578 [59] W. Altmannshofer *et al.*, *Symmetries and Asymmetries of $B \rightarrow K^* \mu^+ \mu^-$ Decays in*
 3579 *the Standard Model and Beyond*, JHEP **01** (2009) 019, [arXiv:0811.1214](#).
- 3580 [60] C. Bobeth, G. Hiller, and G. Piranishvili, *CP Asymmetries in bar $B \rightarrow \bar{K}^*(\rightarrow \bar{K}\pi)\bar{\ell}\ell$*
 3581 *and Untagged \bar{B}_s , $B_s \rightarrow \phi(\rightarrow K^+ K^-)\bar{\ell}\ell$ Decays at NLO*, JHEP **07** (2008) 106,
 3582 [arXiv:0805.2525](#).
- 3583 [61] P. Ball and R. Zwicky, *$B_{d,s} \rightarrow \rho, \omega, K^*, \phi$ decay form-factors from light-cone sum*
 3584 *rules revisited*, Phys. Rev. **D71** (2005) 014029, [arXiv:hep-ph/0412079](#).
- 3585 [62] A. J. Buras and M. Munz, *Effective Hamiltonian for $B \rightarrow X(s)$ $e^+ e^-$ beyond*
 3586 *leading logarithms in the NDR and HV schemes*, Phys. Rev. D **52** (1995) 186,
 3587 [arXiv:hep-ph/9501281](#).
- 3588 [63] H. H. Asatrian, H. M. Asatrian, C. Greub, and M. Walker, *Two loop virtual corrections to $B \rightarrow X_s l^+ l^-$ in the standard model*, Phys. Lett. B **507** (2001) 162,
 3589 [arXiv:hep-ph/0103087](#).
- 3590 [64] G. A. Cowan, D. C. Craik, and M. D. Needham, *RapidSim: an application for the*
 3591 *fast simulation of heavy-quark hadron decays*, Comput. Phys. Commun. **214** (2017)
 3592 239, [arXiv:1612.07489](#).
- 3593 [65] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the $B^+ \rightarrow K^{*+} \mu^+ \mu^-$ decay*,
 3594 [arXiv:2012.13241](#), to appear in Phys. Rev. Lett.
- 3595 [66] A. Poluektov, *Kernel density estimation of a multidimensional efficiency profile*,
 3596 JINST **10** (2015), no. 02 P02011, [arXiv:1411.5528](#).
- 3597 [67] J. Eschle, *Studies of $B^+ \rightarrow K^+ \pi^- \pi^+ e^+ e^-$ decays at LHCb*, Master's thesis, Zurich
 University, 2017-06-21.
- 3598 [68] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*,
 3599 Nucl. Instrum. Meth. **A555** (2005) 356, [arXiv:physics/0402083](#).
- 3600 [69] LHCb, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. A **30** (2015),
 3601 no. 07 1530022, [arXiv:1412.6352](#).
- 3602 [70] LHCb, R. Aaij *et al.*, *Measurement of b-hadron production fractions in 7 TeV pp*
 3603 *collisions*, Phys. Rev. D **85** (2012) 032008, [arXiv:1111.2357](#).
- 3604 [71] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics*, Chin. Phys.
 3605 **C40** (2016), no. 10 100001.

- 3608 [72] R. Aaij *et al.*, *Test of lepton universality with $\lambda_b^0 \rightarrow pk^- \ell^+ \ell^-$ decays*, Journal of High
 3609 Energy Physics **2020** (2020), no. 5 40.
- 3610 [73] F. Pedregosa *et al.*, *Scikit-learn: Machine learning in Python*, Journal of Machine
 3611 Learning Research **12** (2011) 2825.
- 3612 [74] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, PhD thesis, Cracow, INP, 1986.
- 3614 [75] U. Egede *et al.*, *New observables in the decay mode $\bar{B}_d \rightarrow \bar{K}^{*0} l^+ l^-$* , JHEP **11** (2008)
 3615 032, [arXiv:0807.2589](https://arxiv.org/abs/0807.2589).
- 3616 [76] F. Kruger and J. Matias, *Probing new physics via the transverse amplitudes of $B^0 \rightarrow K^{*0}(\rightarrow K^-\pi^+)l^+l^-$ at large recoil*, Phys. Rev. D **71** (2005) 094009,
 3617 [arXiv:hep-ph/0502060](https://arxiv.org/abs/hep-ph/0502060).
- 3619 [77] T. Blake, U. Egede, and A. Shires, *The effect of S-wave interference on the $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ angular observables*, JHEP **03** (2013) 027, [arXiv:1210.5279](https://arxiv.org/abs/1210.5279).
- 3621 [78] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$* , Phys. Rev. Lett. **108** (2012) 181806, [arXiv:1112.3515](https://arxiv.org/abs/1112.3515).
- 3623 [79] LHCb collaboration, R. Aaij *et al.*, *Measurement of form-factor-independent observables in the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$* , Phys. Rev. Lett. **111** (2013) 191801,
 3624 [arXiv:1308.1707](https://arxiv.org/abs/1308.1707).
- 3626 [80] J. Gratrex, M. Hopfer, and R. Zwicky, *Generalised helicity formalism, higher moments and the $B \rightarrow K_{J_K}(\rightarrow K\pi)\bar{\ell}_1 \ell_2$ angular distributions*, Phys. Rev. D **93** (2016), no. 5 054008, [arXiv:1506.03970](https://arxiv.org/abs/1506.03970).
- 3629 [81] T. Huber, T. Hurth, and E. Lunghi, *Inclusive $\bar{B} \rightarrow X_s \ell^+ \ell^-$: complete angular analysis and a thorough study of collinear photons*, JHEP **06** (2015) 176, [arXiv:1503.04849](https://arxiv.org/abs/1503.04849).
- 3631 [82] T. Huber, T. Hurth, and E. Lunghi, *Logarithmically Enhanced Corrections to the Decay Rate and Forward Backward Asymmetry in $\bar{B} \rightarrow X_s \ell^+ \ell^-$* , Nucl. Phys. B **802**
 3632 (2008) 40, [arXiv:0712.3009](https://arxiv.org/abs/0712.3009).
- 3634 [83] T. Huber, E. Lunghi, M. Misiak, and D. Wyler, *Electromagnetic logarithms in $\bar{B} \rightarrow X_s l^+ l^-$* , Nucl. Phys. B **740** (2006) 105, [arXiv:hep-ph/0512066](https://arxiv.org/abs/hep-ph/0512066).
- 3636 [84] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and Physics Documentation*, Comput. Phys. Commun. **199** (2016) 86, [arXiv:1011.0937](https://arxiv.org/abs/1011.0937).
- 3638 [85] LHCb, R. Aaij *et al.*, *Angular analysis of the $B^0 \rightarrow K^{*0} e^+ e^-$ decay in the low- q^2 region*, JHEP **04** (2015) 064, [arXiv:1501.03038](https://arxiv.org/abs/1501.03038).
- 3640 [86] BESIII, M. Ablikim *et al.*, *Study of $D^+ \rightarrow K^-\pi^+ e^+ \nu_e$* , Phys. Rev. D **94** (2016), no. 3 032001, [arXiv:1512.08627](https://arxiv.org/abs/1512.08627).

- 3642 [87] LHCb, R. Aaij *et al.*, *Observation of $J/\psi p$ Resonances Consistent with Pen-*
 3643 *taquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays*, Phys. Rev. Lett. **115** (2015) 072001,
 3644 arXiv:1507.03414.
- 3645 [88] LHCb, R. Aaij *et al.*, *Precise measurement of the f_s/f_d ratio of fragmentation fractions*
 3646 *and of B_s^0 decay branching fractions*, arXiv:2103.06810.
- 3647 [89] C. Langenbruch, *Parameter uncertainties in weighted unbinned maximum likelihood*
 3648 *fits*, arXiv:1911.01303.
- 3649 [90] LHCb, R. Aaij *et al.*, *Measurement of the polarization amplitudes in $B^0 \rightarrow$*
 3650 *$J/\psi K^*(892)^0$ decays*, Phys. Rev. D **88** (2013) 052002, arXiv:1307.2782.
- 3651 [91] BaBar, B. Aubert *et al.*, *Measurement of decay amplitudes of $B \rightarrow J/\psi K^*, \psi(2S)K^*$,*
 3652 *and $\chi_{c1}K^*$ with an angular analysis*, Phys. Rev. D **76** (2007) 031102,
 3653 arXiv:0704.0522.
- 3654 [92] Belle, K. Chilikin *et al.*, *Experimental constraints on the spin and parity of the*
 3655 *$Z(4430)^+$* , Phys. Rev. D **88** (2013), no. 7 074026, arXiv:1306.4894.
- 3656 [93] Belle, K. Chilikin *et al.*, *Observation of a new charged charmoniumlike state in*
 3657 *$\bar{B}^0 \rightarrow J/\psi K^-\pi^+$ decays*, Phys. Rev. D **90** (2014), no. 11 112009, arXiv:1408.6457.
- 3658 [94] A. Poluektov, *Correction of simulated particle identification response in LHCb using*
 3659 *kernel density estimation*, LHCb-INT-2017-007.
- 3660 [95] D. M. Straub, *flavio: a Python package for flavour and precision phenomenology in*
 3661 *the Standard Model and beyond*, arXiv:1810.08132.
- 3662 [96] G. Bohm and G. Zech, *Statistics of weighted poisson events and its applications*,
 3663 Nuclear Instruments and Methods in Physics Research Section A: Accelerators,
 3664 Spectrometers, Detectors and Associated Equipment **748** (2014) 1.