
Test of lepton flavour universality using $B^+ \rightarrow K^+ l^+ l^-$ decays

P. Álvarez Cartelle¹, T. Humair¹, M. Patel¹

¹*Imperial College London, London, United Kingdom*

Abstract

A measurement of R_K , the ratio of branching fractions between $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^+ \rightarrow K^+ e^+ e^-$ decays, is presented. The dataset used corresponds to 5 fb^{-1} of pp collisions recorded with the LHCb detector at centre-of-mass energies of 7, 8 and 13 TeV.

Note on the analysis code

The c++ code used for this analysis is saved on CERN gitlab: <https://gitlab.cern.ch/thumair/SuperLenin>. It contains all the scripts performing the selection, fits, corrections to the simulation samples and extraction of the efficiencies, systematics and cross-checks. This code was written to run and store data on the batch system and disks at Imperial College London, and is therefore not supposed to work ‘out-of-the-box’. Most of the technical sections in this note contain a link to the corresponding script, in order to give the reader the possibility to check the code by themselves, or to re-use parts of it.

Change log

Since the “Amebelodon-R1” version, the following major changes have been made:

- Added measurement of R_K and \mathcal{B} ($B^+ \rightarrow K^+ e^+ e^-$) in Section 12;
- Added unblinded fit results in Section 9.2;
- Moved “9.2 Fits to $B^+ \rightarrow K^+ \mu^+ \mu^-$ mode” to the crosschecks in Section 11.7
- Updated preunblinding checks in Section 11.5;
- Added appendix with variation of R_K when applying the inverse of the correction weights, Section N.

Since the “Amebelodon” version, the following major changes have been made:

- Updated preunblinding checks in Section 11.5;
- Added study of fake TIS events to the evaluation of the TIS efficiency systematic to Section 10.4.4;
- Included $R_{\psi(2S)}$ in the cross-check using preliminary tag-and-probe efficiencies in Section 11.4;
- Added cross-checks for double misID backgrounds, Section 11.6;
- Added appendix with ECAL illumination plots, Appendix M;
- Added appendix with $r_{J/\psi}$ as a function of χ^2_{IP} of the leptons, Appendix K;
- Added the value of $\varepsilon^{\text{fit range}}$ in the efficiency table (and variation of R_K when turning off the correction weights) in Section 6.10;
- Added explanation on data/MC comparison for the trigger category, Section 6.6.1;
- Added appendix with the explanation of the daughters η distribution, Appendix G;
- Corrected missing $B^+ \rightarrow K^+ \psi(2S)(e^+ e^-)$ fits for the narrow q^2 selection, Section 11.1.

Since the “Elasmotherium” version, the following major changes have been made:

- Added ECal calibration constants for 2016;
- Updated study on expected compatibility with LHCb-PAPER-2014-024 to take into account the R_K value measured;
- Added tracking efficiency cross-check using preliminary tag-and-probe efficiencies ;

- Added systematics on the r_{prc}^{rt} fractions, used to constrain the amount of partially reconstructed background across the three trigger categories;
- Added results of preunblinding checks agreed with RC.

Since the “Paraceratherium” version, the following major changes have been made:

- Added fit model systematic;
- Added PID systematic from binning and trigger bias;
- Added distribution for $B^+ \rightarrow K^{*+}(\rightarrow K^+\pi^0)e^+e^-$ background;
- Added physics model systematic;
- Updates to the blinding strategy section RE: pre-unblinding cross-checks and plans for publishing $\mathcal{B}(B^+ \rightarrow K^+e^+e^-)$;
- Unblinded Run 1 $B^+ \rightarrow K^+\mu^+\mu^-$ data and compared $\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-)$ to the previously published version;
- Added values for $r_{J/\psi}$ and $R_{\psi(2S)}$ combining different trigger categories and Run 1 and Run 2 data. For $R_{\psi(2S)}$ also computed values in a narrower q^2 window in order to reduce potential systematics from charm backgrounds;
- Added $r_{J/\psi}$ determination in 2d bins of relevant kinematic variables.

Since the “Mastodonte” version, the following major changes have been made:

- The ghosts are fully taken into account when computing the efficiencies, see Section 6.2;
- Multiple candidates are removed, see Section 5.8;
- For the electron modes, the q^2 and B mass resolution is smeared to match the data better, and the nominal computation of the uncertainties take this smearing into account, see Section 6.9;
- The part reco shape has been updated to contain events from B_s decays, see Section 5.7;
- Some detail about the kinematic weights used to compute the trigger histograms have been add in Section 6.5;
- The following corrections to the L0 trigger histograms have been made and are documented in Section 6.6
 - The previous version of the note contained a mistake in the estimated statistical error on the trigger histograms, which has been corrected;

- The tag on the L0Electron histograms has been changed from `GlobalTIS` to `muonTIS` or `hadronTIS` to minimise biases;
- To extract the L0TIS histograms for the TIS! category, the kaon is explicitly required to not have fired the hadron trigger, which removes the “hidden variable” bias. In the case of the L0TIS histograms, this bias was due to the different kinematics of the kaon between the modes.
- A new scale parameter to the right-hand-side tail of the distributions is introduced in the fits to the electron mode to improve the fit quality of the normalisation mode.
- The simultaneous fit to the $B^+ \rightarrow K^+\mu^+\mu^-$ and $B^+ \rightarrow K^+e^+e^-$ samples for the extraction of R_K has been validated.
- A method to compute the compatibility of the new Run 1 measurement with the previous analysis is outlined.
- The major sources of systematics have been computed, see Section 10;
- The integrated values of $r_{J/\psi}$ and $R_{\psi(2S)}$ have been updated to take into account the ghosts and the multiple candidates, and the systematic uncertainties are shown, see Section 11.1;
- The kinematic distributions in the rare and J/ψ modes are displayed on top of the differential $r_{J/\psi}$ plot, to check if the binning is fine enough, see Section 11.2;
- Supplementary kinematic and PID plots are provided in Appendix F and in Appendix E.

Contents

1	Introduction	1
2	Commonly used variables	3
3	Analysis strategy	5
3.1	$K^+e^+e^-$ and $K^+\mu^+\mu^-$ final states at LHCb	5
3.2	Normalisation and control modes	9
3.3	Blinding strategy	12
3.4	Outline of the note	12
4	Data, control, and simulation samples	13
4.1	Data samples	13
4.2	Simulation samples	13
5	Selection	15
5.1	Types of backgrounds	15
5.1.1	Charmed and Strange Partially Reconstructed Backgrounds	15
5.1.2	Peaking backgrounds with a single mis-identified track	16
5.1.3	Peaking backgrounds with two mis-identified tracks	16
5.1.4	Cascade backgrounds	16
5.1.5	Backgrounds with converted photons	17
5.2	The Selection Chain	17
5.3	Stripping selection	17
5.4	Offline selection	19
5.4.1	Definition of q^2 and mass ranges	19
5.4.2	Fiducial cuts: acceptance of the PID system	21
5.4.3	Fiducial cuts: calorimeter acceptance	22
5.4.4	Cascade background vetoes	22
5.4.5	Particle identification	23
5.5	Trigger	25
5.5.1	TCK and trigger fiducial cuts	25
5.6	Multivariate selection	29
5.6.1	Multivariate selection training	29
5.6.2	BDT optimisation and performance	32
5.7	Fit window	36
5.8	Multiple candidates	39
6	Efficiencies and Corrections to the Simulation	40
6.1	Efficiency and branching fraction computations	40
6.2	Truth-matching, ghosts, and definition of q_{true}^2	43
6.2.1	Handling of ghosts in the simulation	43
6.2.2	Definition of q_{true}^2	44

6.3	Fit to the Pre-Selected Control Modes	46
6.4	Corrections to particle identification efficiencies	49
6.4.1	Kaon, muon and pion (mis-)identification efficiencies	49
6.4.2	Electron identification efficiencies	66
6.4.3	Combination of PID efficiencies and biases	67
6.5	Kinematic weights for trigger histograms	74
6.6	L0 efficiencies calibration	74
6.6.1	e TOS: calibration histograms	74
6.6.2	e TOS: efficiency combination	78
6.6.3	μ TOS: calibration histograms	88
6.6.4	μ TOS: efficiency combination	90
6.6.5	h TOS: calibration histograms	93
6.6.6	h TOS: efficiency combination	94
6.6.7	TIS: calibration histograms	101
6.6.8	TIS: efficiency combination	102
6.6.9	Hidden variables in L0 efficiency parametrisation	104
6.7	HLT efficiencies calibration	105
6.8	Kinematic corrections	108
6.8.1	Correction to the occupancy distribution	109
6.8.2	Kinematic distributions in simulation and data	110
6.9	q^2 and B mass resolution in the electron modes	117
6.10	Summary and list of efficiencies	121
7	Mass modelling	124
7.1	Fit to resonant modes	125
7.1.1	Fit to $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$	125
7.1.2	Fit to $B^+ \rightarrow K^+ J/\psi(e^+e^-)$	126
7.1.3	Fit to $B^+ \rightarrow K^+ \psi(2S)(\mu^+\mu^-)$	127
7.1.4	Fit to $B^+ \rightarrow K^+ \psi(2S)(e^+e^-)$	127
7.2	Rare mode fits	127
7.2.1	Fit to $B^+ \rightarrow K^+ \mu^+\mu^-$	128
7.2.2	Fit to $B^+ \rightarrow K^+ e^+e^-$	128
8	Strategy for the extraction of R_K	130
8.1	Fit validation	131
8.2	Compatibility with LHCb-PAPER-2014-024	133
9	Fit result	141
9.1	Fits to the normalisation modes	141
9.2	R_K fit results	147

10 Systematic uncertainties	150
10.1 Size of the calibration and simulation samples	151
10.2 Kinematic reweighting scheme	152
10.3 PID calibration biases	159
10.4 Trigger histograms	161
10.4.1 Biases caused by the L0Muon calibration histograms	161
10.4.2 Biases caused by the L0Electron calibration histograms	161
10.4.3 Biases caused by the L0Hadron calibration histograms	163
10.4.4 Biases caused by the LOTIS calibration histograms	164
10.5 ECAL calibration constants	168
10.6 Occupancy proxy	168
10.7 Material and rapidity	169
10.8 Momentum and q^2 resolution	173
10.8.1 Electron q^2 and $m(K^+e^+e^-)$ resolution: line shape	173
10.8.2 Electron q^2 and $m(K^+e^+e^-)$ resolution: trigger bias	174
10.8.3 Electron q^2 resolution: momentum dependence	175
10.9 Fit model	176
10.10 Partially reconstructed background constraint	176
10.11 Decay model	177
10.12 Combination of systematic uncertainties	178
11 Cross-checks	179
11.1 Integrated measurement of $r_{J/\psi}$ and $R_{\psi(2S)}$	179
11.2 Differential measurement of $r_{J/\psi}$	179
11.3 $r_{J/\psi}^{\text{bin}}$ test	217
11.4 Tracking efficiency cross-check	222
11.5 Pre-unblinding cross-checks	223
11.6 Contamination from double misID backgrounds	225
11.7 Measurement of \mathcal{B} ($B^+ \rightarrow K^+\mu^+\mu^-$)	225
12 Results	230
12.1 Measurement of R_K	230
12.2 Measurement of \mathcal{B} ($B^+ \rightarrow K^+e^+e^-$)	230
A Simulation samples	232
A.1 Run 1 samples	232
A.2 Run 2 samples	237
B BDT: supplementary plots	239
C Extra fit plots	241
C.1 $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ simulation fit	241
C.2 $B^+ \rightarrow K^+J/\psi(e^+e^-)$ simulation fit	241
C.3 $B^+ \rightarrow \pi^+J/\psi(e^+e^-)$ mis-ID mass variable transformation	249

D Fits electron calibration samples	250
D.1 2011 Fits	250
D.1.1 2011 Fits, e^- probe, before PID cut	250
D.1.2 2011 Fits, e^+ probe, before PID cut	252
D.1.3 2011 Fits, e^- probe, after PID cut	253
D.1.4 2011 Fits, e^+ probe, after PID cut	255
D.2 2012 Fits	256
D.2.1 2012 Fits, e^- probe, before PID cut	256
D.2.2 2012 Fits, e^+ probe, before PID cut	258
D.2.3 2012 Fits, e^- probe, after PID cut	260
D.2.4 2012 Fits, e^+ probe, after PID cut	262
D.3 2015 Fits	264
D.3.1 2015 Fits, e^- probe, before PID cut	264
D.3.2 2015 Fits, e^+ probe, before PID cut	266
D.3.3 2015 Fits, e^- probe, after PID cut	268
D.3.4 2015 Fits, e^+ probe, after PID cut	270
D.4 2016 Fits	272
D.4.1 2016 Fits, e^- probe, before PID cut	272
D.4.2 2016 Fits, e^+ probe, before PID cut	276
D.4.3 2016 Fits, e^- probe, after PID cut	281
D.4.4 2016 Fits, e^+ probe, after PID cut	285
E PID supplementary plots	290
F Supplementary kinematic plots	292
G Shape of the pseudo-rapidity distributions	292
H Fits to simulation	300
H.1 Signal fits	300
H.1.1 Fits to $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ simulation	300
H.1.2 Fits to $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulation	300
H.1.3 Fits to $B^+ \rightarrow K^+ \psi(2S)(\mu^+\mu^-)$ simulation	300
H.1.4 Fits to $B^+ \rightarrow K^+ \psi(2S)(e^+e^-)$ simulation	300
H.1.5 Fits to $B^+ \rightarrow K^+ \mu^+\mu^-$ simulation	301
H.1.6 Fits to $B^+ \rightarrow K^+ e^+e^-$ simulation	301
H.2 Background fits	316
H.2.1 Fits to $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ simulation	316
H.2.2 Fits to $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ simulation	316
H.2.3 Fits to $B^+ \rightarrow J/\psi(e^+e^-)X$ simulation	316
H.2.4 Fits to $B^0 \rightarrow K^{*0} e^+e^-$ simulation	317
H.2.5 Leakage from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ to $B^+ \rightarrow K^+ e^+e^-$	317
H.3 Fit model systematic	326
H.3.1 Alternative signal models	326

H.3.2 Alternative background models	326
I Fit validation	331
I.1 Fit stability and BDT cut	331
I.2 Fit validation for different R_K hypotheses	334
J Fits to data	353
K Additional $r_{J/\psi}$ differential plots	354
L $r_{J/\psi}^{\text{bin}}$ test: binning definition	355
M ECAL illumination	355
N Reweighting checks	356
O Staged unblinding results	357
P Correlation between R_K and \mathcal{B} ($B^+ \rightarrow K^+\mu^+\mu^-$)	358
References	375

¹ 1 Introduction

² In the Standard Model (SM), the couplings of the gauge bosons to the leptons are
³ independent of the lepton flavour. This symmetry, so-called lepton flavour universality
⁴ (LFU), can be violated in many New Physics (NP) scenarios [1, 2]. The LFU hypothesis can
⁵ be tested using $b \rightarrow s\ell^+\ell^-$ processes. These flavour changing neutral current transitions
⁶ are expected to be very sensitive to NP effects, as they can proceed only through loop
⁷ diagrams in the SM.

⁸ The observable R_K is defined as the ratio of branching fractions

$$R_K = \frac{\int_{q^2_{min}}^{q^2_{max}} \frac{d\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-)}{dq^2} dq^2}{\int_{q^2_{min}}^{q^2_{max}} \frac{d\mathcal{B}(B^+ \rightarrow K^+e^+e^-)}{dq^2} dq^2}, \quad (1)$$

⁹
¹⁰ where q^2 is the invariant mass of the di-lepton pair. This ratio is precisely calculable in
¹¹ the SM, given large cancellations of hadronic uncertainties, and when integrated over all
¹² q^2 it is predicted to be $R_K = 1.0000 \pm 0.0001$ [3]. When measuring R_K in a finite q^2
¹³ range, electromagnetic corrections on the order of a few percent have to be taken into
¹⁴ account [4]. These electromagnetic corrections are included in the LHCb simulation using
¹⁵ the PHOTOS software [5]. The remaining uncertainty on R_K due to possible imperfections
¹⁶ of PHOTOS is studied in Ref. [4], and is established to be 1%. The R_K observable is
¹⁷ therefore theoretically very clean, and a measured value significantly displaced from unity
¹⁸ would constitute a clear sign of new physics.

¹⁹ In a previous analysis [6], the LHCb collaboration measured R_K in the q^2 region
²⁰ between 1 and 6 GeV/c^2 . The value obtained, $R_K = 0.745^{+0.090}_{-0.074}(\text{stat}) \pm 0.036(\text{syst})$, is
²¹ consistent with the SM expectation at the level of 2.6σ , see Fig. 1. Several tensions with
²² respect to SM predictions have appeared in recent measurements of branching fractions
²³ and angular observables related to $b \rightarrow s\mu^+\mu^-$ processes. For example, the measurements
²⁴ of the differential branching fractions for $B^+ \rightarrow K^+\mu^+\mu^-$ and $B^0 \rightarrow \phi\mu^+\mu^-$ decays are
²⁵ found to be below their theoretical predictions at low q^2 [7, 8], and, in $B^0 \rightarrow K^{*0}\mu^+\mu^-$
²⁶ decays, the angular observable P'_5 shows a deviation from the SM prediction at the 3.5σ
²⁷ level [9], as shown in Fig. 2.

²⁸ A number of measurements related to $b \rightarrow s\mu^+\mu^-$ processes and to $b \rightarrow s\gamma$ processes
²⁹ are used to extract the effective coupling constants (Wilson Coefficients), $\mathcal{C}_7^{(')}$, $\mathcal{C}_9^{(')}$, $\mathcal{C}_{10}^{(')}$
³⁰ appearing in the Hamiltonian describing the underlying transitions [12–14]. Two recent
³¹ global fits [12] and [14] rely on different form factors computations, and differ slightly in
³² how they handle theoretical uncertainties. However, their conclusions are similar: they
³³ favour a shift of significance shift of \mathcal{C}_9 , or of a combination of \mathcal{C}_9 and \mathcal{C}_{10} , from their SM
³⁴ expectations. For example, the significance of the shift in \mathcal{C}_9 has a significance of 4.5σ in
³⁵ Ref. [12] and 3.7σ in Ref. [14], as illustrated in Fig. 3. Recent updates of these analyses
³⁶ have confirmed this trend [15, 16]. Care must be taken when interpreting these significances,
³⁷ as they depend on the size of theoretical uncertainties that affect both angular observables
³⁸ and branching fractions. These uncertainties stem from form factor computations, and
³⁹ interference between the short-distance contribution, mediated by the penguin decay,

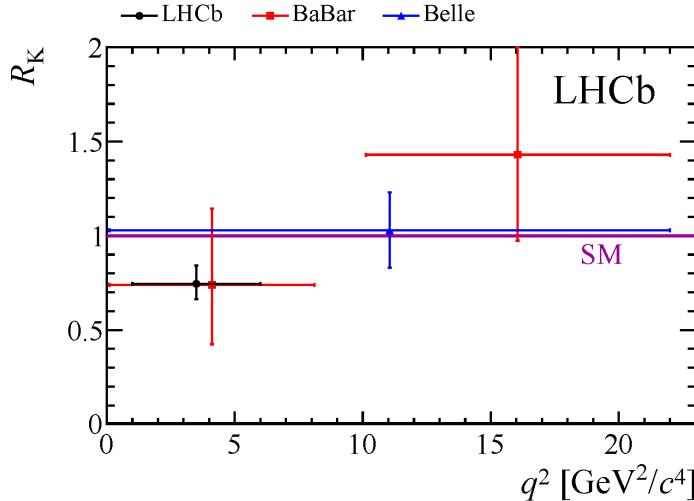


Figure 1: Measurements of R_K performed by the LHCb [6], BaBar [10] and Belle [11] collaborations.

and the long distance contribution, mediated by a charm-quark loop. However, these hadronic uncertainties do not affect the R_K observable. Moreover, the 25% shift from unity observed in the R_K measurement is consistent with the same shift in \mathcal{C}_9 , applied only to the muon operator.

Recently, the LHCb collaboration has also presented the measurement of the LFU ratio for $B^0 \rightarrow K^{*0}\ell^+\ell^-$ process [17], the so-called $R_{K^{*0}}$ observable, showing a discrepancy with unity at the level of 2.2-2.3 σ (2.4-2.5 σ) in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$ ($0.045 < q^2 < 1.1 \text{ GeV}^2$). Given this measurement, as well as the results of the global fits, improvement in the precision of the measurement of R_K is crucial to clarify whether NP is needed to explain the measurements of $b \rightarrow s\ell^+\ell^-$ processes.

In this note, an updated measurement of the R_K observable is presented, using the full Run 1 dataset, which consists of $\sim 1 \text{ fb}^{-1}$ of pp collision data taken at a centre-of-mass energy of 7 TeV and $\sim 2 \text{ fb}^{-1}$ at 8 TeV, together with the data taken by LHCb during 2015 and 2016, corresponding to an integrated luminosity of $\sim 2 \text{ fb}^{-1}$ of pp collisions at a centre-of-mass energy of 13 TeV. Given the $b\bar{b}$ production cross-section scales approximately linearly with energy [18], the amount of data in this updated analysis is expected to be approximately double that in the Run 1 R_K analysis. Taking the precision from the previous analysis and extrapolating, the expected statistical precision on the current result is expected to be $\sim 7 - 8\%$. Systematic effects at the sub-percent level then have little bearing on the overall precision.

In this updated analysis, R_K is measured in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$. In this range, the contribution from the resonant modes—the ϕ below $\sim 1 \text{ GeV}^2$ and $c\bar{c}$ resonances above $\sim 9.5 \text{ GeV}^2$ —is negligible and the decay is dominated by the penguin transition.

This q^2 range is that used in the most recent $b \rightarrow s\ell^+\ell^-$ analyses [9, 17] but a slightly wider range, $1.0 < q^2 < 6.0 \text{ GeV}^2$, was used in the previous analysis of R_K [6]. The

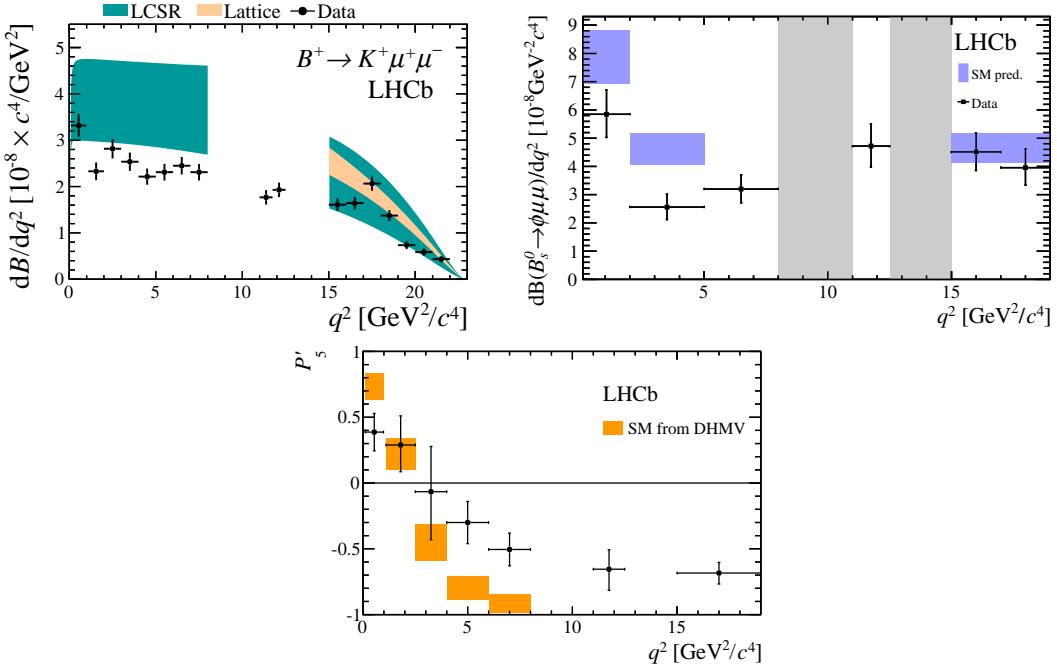


Figure 2: Recent LHCb measurements of $b \rightarrow s\mu^+\mu^-$ observables. The top left plot [7], and top right plot [8] show the differential branching fractions for $B^+ \rightarrow K^+\mu^+\mu^-$ and $B^0 \rightarrow \phi\mu^+\mu^-$ decays, respectively. The LHCb data points are shown in black and the theory predictions are shown with blue bands. The bottom plot shows the P'_5 angular observable from the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay [9], with the data points shown in black and the theory prediction in orange. All three plots show a discrepancy with respect to the SM in the low q^2 region.

reduction of the q^2 range results in a loss of $\sim 2\%$ of $B^+ \rightarrow K^+\mu^+\mu^-$ and $B^+ \rightarrow K^+e^+e^-$ events. Unlike in the recent $R_{K^{*0}}$ analysis [17], where $R_{K^{*0}}$ is measured in the region $0.045 < q^2 < 1.1 \text{ GeV}^2$, no measurement of R_K at low q^2 is performed in the present analysis. In the case of $B^0 \rightarrow K^{*0}\ell^+\ell^-$, at low q^2 the decay proceeds via a virtual photon, which significantly enhances the branching fraction. This photon-mediated contribution is forbidden by angular momentum conservation in the $B^+ \rightarrow K^+\ell^+\ell^-$ decay and there is therefore no enhancement of the $B^+ \rightarrow K^+\ell^+\ell^-$ branching fraction in the low q^2 region. Moreover, in this region, the signal decay $B^+ \rightarrow K^+e^+e^-$ interferes with the decay $B^+ \rightarrow K^+\phi(e^+e^-)$ of similar amplitude. A measurement of R_K at low q^2 would therefore be very challenging and is not attempted.

2 Commonly used variables

Variables related to quantities measured by the LHCb detector are mentioned throughout this note. These variables are defined in this section. A decay of the type $A \rightarrow b, c, \dots$ is considered, where b and c are tracks and A is a short-lived particle candidate formed by

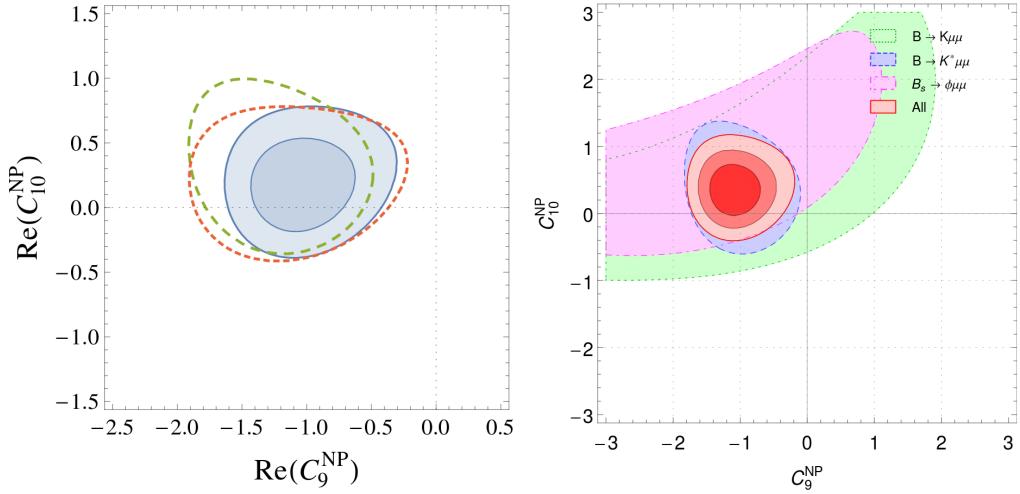


Figure 3: Results of the global fits to \mathcal{C}_9 and \mathcal{C}_{10} using $b \rightarrow s\mu^+\mu^-$ data. The left-hand plot, from Ref. [14], shows the 1 and 2 σ contours in blue, as well as the 2 σ region when form factor uncertainties (green dashed) or non-factorizable corrections (red dashed) are doubled. The right-hand plot, from [12], shows the 1, 2 and 3 σ contours of the global fit in red. The 3 σ region allowed by the $B^+ \rightarrow K^+ \mu^+ \mu^-$, $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ and $B^0 \rightarrow \phi \mu^+ \mu^-$ measurements is shown in green, blue, and purple. Both fit results are in good agreement and shows a significant deviation of \mathcal{C}_9 from its SM expectation.

79 the combination of tracks it decays into. The origin vertex of A , which is where the pp
 80 collision takes place, is referred to as the primary vertex (PV) and the decay vertex of A ,
 81 defined by the interaction point of the tracks it decays into, is referred to as secondary
 82 vertex or simply decay vertex (DV).

$x_{ECAL}(b), y_{ECAL}(b)$	x and y coordinates of track b when it reaches the electromagnetic calorimeter;
$\chi^2_{\text{DV} \leftrightarrow \text{PV}}(A)$	Difference in χ^2 obtained from two fits: either assuming that all tracks come from the same point, or assuming two vertices (the primary and decay vertex). This gives an indication of the significance of the flight distance of A ;
DIRA(A)	Directional angle: angle between the momentum vector of a particle A and the vector that links the primary vertex to the decay vertex of A ;
$\chi^2_{\text{IP}}(A)$	the difference in χ^2 of a given primary vertex when reconstructed with and without the considered candidate A ;
$\chi^2_{\text{DV}}(A)$	χ^2 of the fit of the decay vertex of A ;
$\chi^2_{\text{DV}}(A)/\text{ndf}$	χ^2 per degree of freedom of the fit of the decay vertex of A ;
PID $_{\beta}$ (b) 83	Log-likelihood difference between the hypotheses that the track b is a particle of type β , and that b is a pion ($\beta = \mu, e, K$ or p). This variable uses inputs from the calorimeters and the RICH detectors;
probNN $_{\beta}$ (b)	Probability of the track b being a particle of type β ($\beta = \mu, e, K, \pi$ or p), as estimated from information coming from all sub-detectors combined using a neural network;
isMuon(b)	Boolean variable that indicates whether b is compatible with being a muon, computed using information from the muon chambers;
prob $_{\text{ghost}}$ (b)	Probability of b being a ghost, that is a fake track coming from a random combination of hits in the tracking system;
$\chi^2_{\text{TrackFit}}(b)$	χ^2 of the b track fit;
nSPDHits	Number of hits recorded in the scintillating pad detector;
$m^{\text{track}}(b, c)$	Mass of a combination of particles using only the information from the tracking systems, meaning that, in the case of electrons, the momentum from recovered bremsstrahlung is not added.

84 3 Analysis strategy

85 In this section, the strategy of the measurement is presented, as well as the main challenges
 86 of the analysis. The origin of all the data and simulated events used in the analysis is also
 87 discussed.

88 3.1 $K^+e^+e^-$ and $K^+\mu^+\mu^-$ final states at LHCb

89 Throughout the analysis, two final states are considered and correspond to three tracks
 90 reconstructed and identified as $K^+\ell^+\ell^-$, where the lepton pair $\ell^+\ell^-$ is either a di-
 91 electron pair (e^+e^-) or a di-muon pair ($\mu^+\mu^-$). The 2D distributions $m(K^+e^+e^-), q^2$
 92 and $m(K^+\mu^+\mu^-), q^2$ are shown in Fig. 4. They correspond to the whole Run 1 and Run 2
 93 datasets, to which the pre-selection and trigger requirements discussed in Section 5.3
 94 and Section 5.5 have been applied. Various features are visible on these plots:

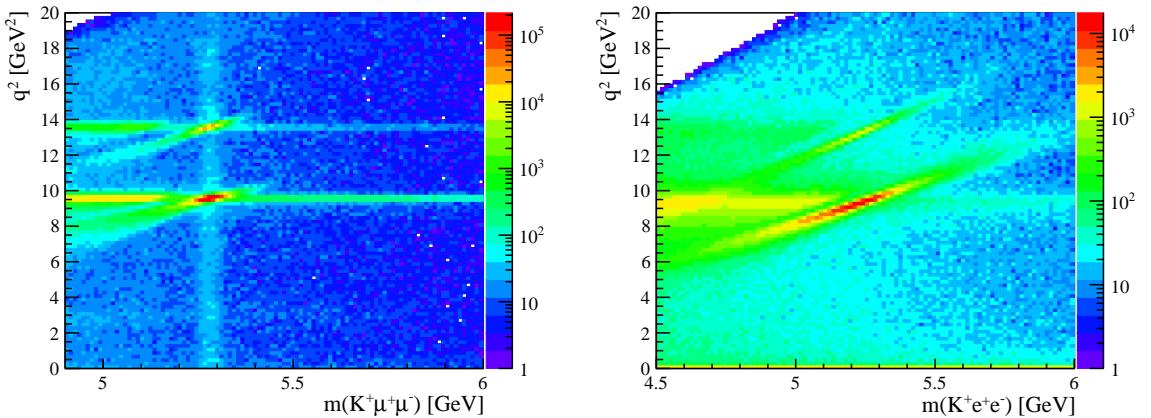


Figure 4: Two-dimensional distributions of $m(K^+\ell^+\ell^-)$, q^2 from the muon (left) and electron (right) datasets. The entire Run 1 and Run 2 datasets are shown with the pre-selection and trigger requirements discussed in Section 5.3 and Section 5.5 applied.

- A peaking structure centered at $m(K^+\ell^+\ell^-) \approx 5.28$ GeV and $q^2 \approx 9.6$ GeV^2 , corresponding to the resonant B -decay $B^+ \rightarrow K^+ J/\psi$. The radiative tails from this decay appear as a diagonal elongation;
- Another peaking structure centered at $m(K^+\ell^+\ell^-) \approx 5.28$ GeV and $q^2 \approx 13.6$ GeV^2 , corresponding to the resonant B -decay $B^+ \rightarrow K^+ \psi(2S)$;
- To the left of these peaks, elongated horizontal structures corresponding to partially-reconstructed decays;
- Finally, the non-resonant rare decay mode $B^+ \rightarrow K^+\ell^+\ell^-$, corresponding to a vertical band at $m(K^+\ell^+\ell^-) \approx 5.28$ GeV and spanning the full q^2 -range.

Comparing the $K^+e^+e^-$ and $K^+\mu^+\mu^-$ data in Fig. 4, it is clearly visible that the mass and q^2 resolutions are much worse in the di-electron case than in the di-muon case, and that the di-muon data sample contains significantly more events. This is due to fundamental differences when reconstructing final states containing a di-electron pair or a di-muon pair at LHCb. Muons are minimum ionising particles that travel across the LHCb detector losing a negligible fraction of their momentum. Electrons, being much lighter, interact with the material of the detector, emitting bremsstrahlung radiation. An electron track from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ decay loses typically $\sim 20\%$ of its momentum in bremsstrahlung before reaching the calorimeters. This significantly lowers the reconstruction efficiency and results in poorer q^2 and mass resolution. Bremsstrahlung radiation emitted before the magnet has a particularly significant impact on the momentum measurement, as it results in a lower momentum measured with the magnet spectrometer, and hence, gives rise to a large radiative tail on the lower mass side of the $B^+ \rightarrow K^+e^+e^-$ mass peak.

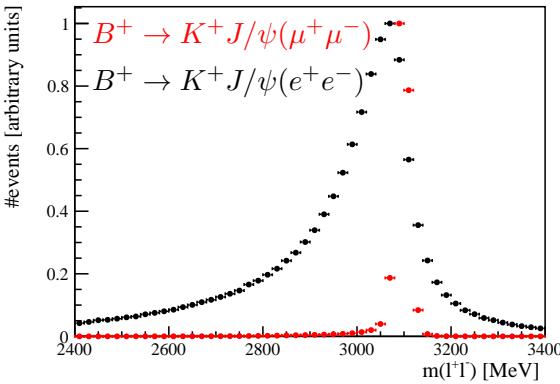


Figure 5: Distributions of $m(\ell^+\ell^-)$ for $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+J/\psi(e^+e^-)$ simulated Run 1 events. In the case of $B^+ \rightarrow K^+J/\psi(e^+e^-)$, the mass peak is much wider and shows extended tails to the left and right. The tail to the left is due to unrecovered bremsstrahlung. The tail to the right corresponds to recovered bremsstrahlung which, due to noise or the finite resolution of the calorimeter, has an over-estimated energy.

This effect is partly corrected for by the bremsstrahlung recovery process, which recovers some fraction of the energy deposited in the calorimeter by the bremsstrahlung photons and adds it back to the electron track. To avoid double counting the radiated energy, a recovered energy cluster can be added to one electron track only. More detail about the bremsstrahlung recovery algorithm can be found in Refs [19, 20]. Using the bremsstrahlung recovery, the momentum resolution for an electron track from $B^+ \rightarrow K^+J/\psi(e^+e^-)$ decay is approximately 12%. This is still very significantly worse than for a muon track, which has a momentum resolution of typically 0.6%.

The $m(\ell^+\ell^-)$ distribution for simulated $B^+ \rightarrow K^+e^+e^-$ and $B^+ \rightarrow K^+\mu^+\mu^-$ events is shown in Fig. 5, illustrating the difference in resolution.

Decays with two electrons in the final state can therefore be classified into three bremsstrahlung categories: the 0γ category, where neither of the two electrons have recovered bremsstrahlung clusters added to them, 1γ where one of the two electron has recovered bremsstrahlung energy added, and 2γ where the two electrons both have recovered bremsstrahlung energy added. The mass and momentum resolution, as well as the particle identification efficiency associated with an event, is strongly dependent on the bremsstrahlung category to which the event belongs.

The LHCb detector triggers on muons using the muon stations, and on electrons using the electromagnetic calorimeter. As discussed further in Section 6.6, the electron triggers are significantly less efficient than the muon triggers, reducing the number of recorded $B^+ \rightarrow K^+e^+e^-$ events. For electrons, the momentum resolution and tracking efficiency depend strongly on the amount of material in the detector, which must be well known to reproduce the efficiency in simulated events. Moreover, the trigger efficiency and bremsstrahlung recovery depends on the performance of the calorimeters, which is

141 affected by e.g. ageing and the occupancy of the detector, which are poorly modelled in
142 the simulation.

143 The differences in the detection of $K^+e^+e^-$ and $K^+\mu^+\mu^-$ final states have two con-
144 sequences for this analysis. First, due to the smaller electron detection efficiency, the
145 precision of the analysis is statistically limited by the yield of $B^+ \rightarrow K^+e^+e^-$ events.
146 Second, the systematic uncertainties can be expected to be dominated by effects related
147 to the detection of electrons. In fact, the muon detection efficiencies have been shown to
148 be under good control over the whole phase space in previous analyses, in particular the
149 angular analysis of $B^0 \rightarrow K^{*0}\mu^+\mu^-$ [9]. The present analysis therefore focuses primarily
150 on maximising the $B^+ \rightarrow K^+e^+e^-$ signal significance, and on demonstrating that the
151 estimate of the $B^+ \rightarrow K^+e^+e^-$ selection efficiency is well controlled. In order to capture
152 the effect of the trigger and calorimeter performance, as well as to control the electron
153 tracking efficiencies and their dependence on the amount of material in the detector,
154 control samples selected from the data are used to correct the simulation. Moreover, to
155 cancel systematic effects, the $B^+ \rightarrow K^+e^+e^-$ and $B^+ \rightarrow K^+\mu^+\mu^-$ branching fractions are
156 measured with respect to control modes, as explained in the next section.

157 3.2 Normalisation and control modes

158 The branching fractions of the decays $J/\psi \rightarrow e^+e^-$ and $J/\psi \rightarrow \mu^+\mu^-$ are measured to be
 159 equal with a sub-percent precision [21]:

$$\frac{\mathcal{B}(J/\psi \rightarrow e^+e^-)}{\mathcal{B}(J/\psi \rightarrow \mu^+\mu^-)} = 0.9983 \pm 0.0078. \quad (2)$$

160 Hence, in this analysis, the resonant decays $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow$
 161 $K^+ J/\psi(\mu^+\mu^-)$ can be used as normalisation channels and R_K is measured in a dou-
 162 ble ratio:

$$R_K = \frac{\int_{1.1 \text{ GeV}^2}^{6 \text{ GeV}^2} \frac{d\mathcal{B}(B^+ \rightarrow K^+ \mu^+\mu^-)}{dq^2} dq^2}{\int_{1.1 \text{ GeV}^2}^{6 \text{ GeV}^2} \frac{d\mathcal{B}(B^+ \rightarrow K^+ e^+e^-)}{dq^2} dq^2} \cdot \frac{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(e^+e^-))}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))} \quad (3)$$

$$= \frac{N(B^+ \rightarrow K^+ \mu^+\mu^-)}{\varepsilon(B^+ \rightarrow K^+ \mu^+\mu^-)} \cdot \frac{\varepsilon(B^+ \rightarrow K^+ e^+e^-)}{N(B^+ \rightarrow K^+ e^+e^-)} \\ \cdot \frac{\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))}{N(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))} \cdot \frac{N(B^+ \rightarrow K^+ J/\psi(e^+e^-))}{\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))}, \quad (4)$$

163 where $N(X)$ indicates the estimate of the yield of decay mode X , which is obtained
 164 from a fit to the invariant mass $m(K^+\ell^+\ell^-)$ with a suitable cut on q^2 ; and $\varepsilon(X)$ is the
 165 efficiency for selecting decay mode X , as computed from simulation samples which are
 166 corrected for various effects using data calibration samples. Hereafter, the $B^+ \rightarrow K^+\ell^+\ell^-$
 167 decay in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$ is referred to as the “rare mode”, whereas the
 168 $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ mode is referred to as the “normalisation mode”.

169 The selection applied to the normalisation modes is kept identical to that applied
 170 to the rare modes, except for the q^2 selection that differentiates between the rare and
 171 normalisation decays. In this way, many systematic uncertainties cancel in the ratio
 172 between these two modes. Indeed, the absolute size of e.g. tracking, particle identification
 173 or trigger efficiencies of one mode need not be known exactly, only the ratio of efficiencies
 174 between the rare mode and the corresponding control mode must be controlled, i.e. $\varepsilon(B^+ \rightarrow$
 175 $K^+ e^+e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+\mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))$ are
 176 the quantities that must be controlled.

177 If the kinematic distributions of variables related to the rare mode were identical to
 178 those of the normalisation mode, all efficiency ratios would be unity and the measurement
 179 would be free from any efficiency-related systematic uncertainties. Residual systematic
 180 uncertainties arise due to differences in kinematic distributions between the rare mode
 181 and the normalisation mode. Distributions for various kinematic variables for simulated
 182 $B^+ \rightarrow K^+\ell^+\ell^-$ and $B^+ \rightarrow K^+ J/\psi$ events are shown in Fig. 6. Some variables show a
 183 near-perfect agreement between the rare and normalisation mode, such as the quality
 184 of the vertex fit $\chi^2_{\text{DV}}(K^+ e^+e^-)$, the significance of the impact parameter $\chi^2_{\text{IP}}(K^+ e^+e^-)$,
 185 the pseudorapidity of all tracks η , and the fraction of an electron track’s energy emitted

via bremsstrahlung before the magnet, denoted $p^{brem}(e)/p^{tot}(e)$. For other variables, the distributions between the rare decay and the normalisation differ, and the cancellation of the efficiency, $\varepsilon(B^+ \rightarrow K^+ e^+ e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+ e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+ \mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))$, will be imperfect e.g. the transverse momenta p_T of all tracks and the dilepton angle $\alpha(\ell^+, \ell^-)$. For such variables, the dependence of the selection efficiency as a function of the variables must be controlled in order to correctly evaluate the effect of possible efficiency mis-modelling as a function of that variable, which would not fully cancel.

In order to demonstrate that the efficiencies are controlled, several cross-checks are performed. The first of these cross-checks is the measurement of the single ratio of branching fractions between the normalisation modes, $r_{J/\psi}$, which must be unity, in agreement with existing measurements (see Eqn.(2)):

$$r_{J/\psi} = \frac{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(e^+ e^-))} \quad (5)$$

$$= \frac{N(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))}{\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))} \cdot \frac{\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+ e^-))}{N(B^+ \rightarrow K^+ J/\psi(e^+ e^-))}. \quad (6)$$

As $r_{J/\psi}$ is a single ratio, the muon and electron efficiencies have to be controlled directly with respect to one another. This is therefore a stringent cross-check, as systematic effects will not cancel as they do in the double ratio R_K . Moreover, as $N(B^+ \rightarrow K^+ J/\psi(e^+ e^-))$ and $N(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))$ are both relatively large, the statistical uncertainty is small and the total uncertainty is dominated by such systematic effects.

As a further cross-check, $r_{J/\psi}$ is measured differentially, as a function of one (see Section 11.2) or several (see Section ??) kinematic variables. This checks if the dependency of the efficiencies on the kinematic variable is properly understood and will therefore cancel well in the double ratio. This is particularly important for variables that show an imperfect agreement between the normalisation and rare mode. As detailed in Section ??, the rare decay can be described using four kinematic variables, whereas the normalisation mode, by virtue of being constrained to have $m(\ell^+ \ell^-) = m(J/\psi)$, is a function of only three kinematic variables. The cross-check hence accesses only projections of the full space.

As a third cross-check, the double ratio of branching fractions $R_{\psi(2S)}$ between the decays $B^+ \rightarrow K^+ \psi(2S)(\ell^+ \ell^-)$ and $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ is measured:

$$R_{\psi(2S)} = \frac{\mathcal{B}(B^+ \rightarrow K^+ \psi(2S)(\mu^+ \mu^-))}{\mathcal{B}(B^+ \rightarrow K^+ \psi(2S)(e^+ e^-))} \cdot \frac{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(e^+ e^-))}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))} \quad (7)$$

$$= \frac{N(B^+ \rightarrow K^+ \psi(2S)(\mu^+ \mu^-))}{\varepsilon(B^+ \rightarrow K^+ \psi(2S)(\mu^+ \mu^-))} \cdot \frac{\varepsilon(B^+ \rightarrow K^+ \psi(2S)(e^+ e^-))}{N(B^+ \rightarrow K^+ \psi(2S)(e^+ e^-))} \\ \cdot \frac{\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))}{N(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))} \cdot \frac{N(B^+ \rightarrow K^+ J/\psi(e^+ e^-))}{\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+ e^-))}. \quad (8)$$

This ratio of branching fractions $\mathcal{B}(\psi(2S) \rightarrow e^+ e^-)/\mathcal{B}(\psi(2S) \rightarrow \mu^+ \mu^-)$ is expected to be unity, as no NP is expected to impact the radiative decays $\psi(2S) \rightarrow e^+ e^-$ or $\psi(2S) \rightarrow \mu^+ \mu^-$.

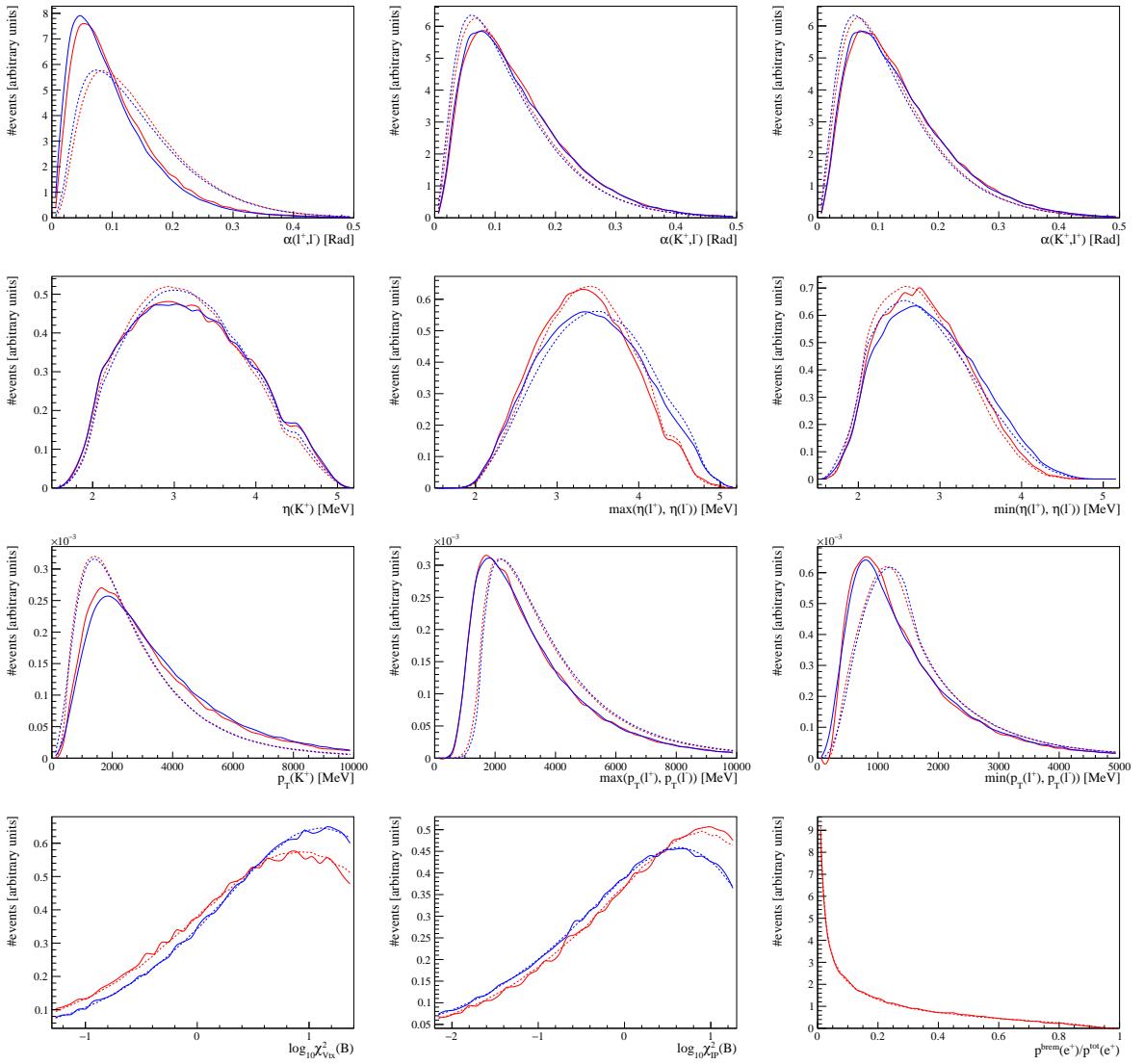


Figure 6: Distributions of kinematic variables for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode (dashed red line), $B^+ \rightarrow K^+ e^+e^-$ mode (solid red line), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode (dashed blue line) and $B^+ \rightarrow K^+ \mu^+\mu^-$ mode (solid blue line). The first line of plots from the top shows the angle between the two leptons, or one lepton and the kaon. The second line shows the rapidity distributions, and the third line the transverse momentum distributions, of all the final state particles. The bottom left plot shows the distribution for the quality of the B vertex fit. The bottom centre plot shows the significance of the B impact parameter. The bottom right plot shows the fraction of momentum radiated away by bremsstrahlung, which is shown only for the electron modes. The distributions correspond to Run 2 simulated events, to which the pre-selection discussed in Section 5.3 has been applied.

216 The precision on this ratio from previous measurements is of the order of $\sim 10\%$ [21]. As
217 it is a double ratio, $R_{\psi(2S)}$ is expected to have reduced systematic uncertainties. Hence,
218 testing that the measurement of $R_{\psi(2S)}$ is unity is a way to make sure that all systematic
219 effects cancel well in the double ratio.

220 3.3 Blinding strategy

221 The analysis procedure was developed keeping the rare decay modes blind *i.e.* without any
222 examination of $N(B^+ \rightarrow K^+ e^+ e^-)$ and $N(B^+ \rightarrow K^+ \mu^+ \mu^-)$. Following discussion with
223 the analysis referees, the Run 1 $B^+ \rightarrow K^+ \mu^+ \mu^-$ data was unblinded, $\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)$
224 determined and the compatibility with the previous LHCb result [7] checked. The results
225 are shown in Section 9 and the compatibility is excellent. Prior to further unblinding, a
226 series of cross-checks were agreed with the review committee. First, the compatibility
227 between the three different trigger categories in the Run 1 and Run 2 data was checked.
228 Second, the blinded pull plots for the mass fits were checked. Third, the consistency with
229 the previous published result was checked. The results of these pre-unblinding cross-checks
230 are detailed in Section ???. Once the full analysis is reviewed and the strategy approved,
231 further unblinding is proposed in several stages. First, the Run 1 $B^+ \rightarrow K^+ e^+ e^-$ data
232 will be unblinded, so that the value of R_K from the Run 1 data can be compared to
233 that in Ref. [6]. Second, the Run 2 $B^+ \rightarrow K^+ \mu^+ \mu^-$ data will be unblinded in order
234 to check the compatibility of the muon mode between Run 1 and Run 2. Finally, the
235 Run 2 $B^+ \rightarrow K^+ e^+ e^-$ data will be unblinded to extract R_K from the Run 2 data.
236 After unblinding, further cross checks will include comparing the estimates of partially
237 reconstructed backgrounds to the signal yields from the $R_{K^{*0}}$ analysis, and establishing
238 R_K in those events with no bremsstrahlung added. Finally, in order to prevent an incorrect
239 calculation of $\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)$ being used in future global fits, it is proposed that
240 the analysis will compute and publish $\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)$ by combining the measured
241 value of R_K with the value of $\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)$ published in Ref. [7], taking into
242 account the relevant correlations. Although the present analysis will include additional
243 $B^+ \rightarrow K^+ \mu^+ \mu^-$ candidates from the Run 2 data, the published $B^+ \rightarrow K^+ \mu^+ \mu^-$ Run 1
244 analysis is already systematically limited from the lack of knowledge of $\mathcal{B}(B^+ \rightarrow K^+ J/\psi)$
245 (which is a systematic that is not in common with the R_K analysis) and so there will be
246 no significant loss in precision on $\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)$ from using only the published Run 1
247 value of $\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)$. The present analysis is not optimised for the muon modes
248 and has $\sim 2/3$ the number of $B^+ \rightarrow K^+ \mu^+ \mu^-$ candidates of the published Run 1 analysis.

249 3.4 Outline of the note

250 The remainder of this note is organised as follows. In Section 4, all the data and simulation
251 samples used in this analysis are introduced. In Section 5, the full selection chain is
252 presented. In Section 6, the corrections applied to the simulation are reported, and the
253 efficiency computation is detailed. In Section 7 and in Section 9, the fit strategy is detailed
254 and the fit results are reported. In Section 10, the systematic uncertainties are reported.

255 In Section 8, the strategy to combine the R_K measurements performed with different
256 triggers and in different data taking conditions is presented. In Section 11, the results of
257 the cross-checks, integrated and differential $r_{J/\psi}$, $R_{\psi(2S)}$ are presented.

258 4 Data, control, and simulation samples

259 All the samples used in this analysis are introduced in this section.

260 4.1 Data samples

261 The analysis presented in this note is based on $K^+e^+e^-$ and $K^+\mu^+\mu^-$ combinations
262 taken from 1.1 fb^{-1} of collision data taken in 2011, 2.1 fb^{-1} of data from 2012, 0.3 fb^{-1}
263 of data from 2015, and 1.6 fb^{-1} of data from 2016. The 2011 and 2012 data are referred
264 to as the Run 1 dataset and the 2015 and 2016 data as the Run 2 dataset. In each
265 year, approximately half of the data is collected with the LHCb magnetic field pointing
266 downwards, and the other half with the magnetic field pointing upwards. These two
267 magnetic field conditions are respectively referred to as the MD and MU conditions.

268 In addition to the $K^+e^+e^-$ and $K^+\mu^+\mu^-$ samples, a sample of $K^+e^+\mu^-$ candidates is
269 used as a proxy for background events selected from fragments of multiple decays - so-called
270 combinatorial background. This sample is used to optimise the multivariate algorithm,
271 as detailed in Section 5.6. The sample is blinded in the region $5100 < m(K^+e^+\mu^-) <$
272 5350 MeV, to avoid any hypothetical signal from BSM $B^+ \rightarrow K^+e^+\mu^-$ decays. Outside of
273 this region, the sample is dominated by random combinations of tracks.

274 Other final states are used to measure the kaon and muon particle identification
275 efficiencies, as detailed in Section 6.4. They correspond to the standard calibration samples
276 listed in Ref. [22].

277 4.2 Simulation samples

278 To tune the signal selection and compute efficiencies, simulated events are used. All the
279 simulated samples are generated using the LHCb Gauss software [23], with an admixture
280 of sim08 and sim09 versions for Run 1, and sim09b for all Run 2 samples. Gauss is based
281 on various software: PYTHIA 6 and 8 [24, 25] are used to generate the proton-proton
282 interactions and the b hadronisation, EVTGEN is used to generate the B decays [26],
283 PHOTOS is used to generate final state radiation [5] and GEANT4 to simulate the interaction
284 of the daughter particles with the LHCb detector [27, 28].

285 The rare decay $B^+ \rightarrow K^+\mu^+\mu^-$ is generated using the DecFile 12113001, which uses
286 the transition form factors described in [29]. The rare decay $B^+ \rightarrow K^+e^+e^-$ is generated
287 using the DecFile 12123002, using the transition form factors from [30]. As explained in
288 more detail in Section 10.11, the form factors used in the decay model have a negligible
289 impact on the efficiency.

290 The simulation software in the Run 2 conditions can potentially be affected by two
291 bugs related to the trigger. The first one, the “TISTOS” bug, is a problem in the matching

292 between an offline and a trigger candidate, and is solved by using the correct version of the
293 DA VINCI software that performs this matching. The second bug is related to the ECAL
294 calibration constants, that converts the reading of the calorimeter cells into an energy.
295 The calibration constants used in the simulation do not correspond exactly to that used in
296 the data. The effect of this bug is discussed in Section 10.5.

297 Details of the sample sizes and simulation conditions for all simulation samples are
298 given in appendix A.

299 **5 Selection**

300 This section describes the selection used to isolate signal events and reduce backgrounds.
301 After the application of selection criteria, the final signal-background separation is achieved
302 and the yields are extracted by fitting the reconstructed mass $m(K^+\ell^+\ell^-)$. Two main
303 kinds of backgrounds are expected to contribute: those that accumulate at some particular
304 $m(K^+\ell^+\ell^-)$ mass, under or close to the signal peak (so-called ‘peaking’ backgrounds)
305 and those that are widely distributed in $m(K^+\ell^+\ell^-)$ (non-peaking) such as combinatorial
306 background. In this analysis, the only non-peaking background stems from combinatorial
307 events which originate from particles selected from multiple different decay modes. When
308 combined, the reconstructed particles happen to fall into the mass region of interest. Such
309 combinatorial backgrounds affect both the rare and normalisation modes in both the
310 electron and muon cases, and are reduced primarily by the use of the multivariate selection
311 described below. The remaining backgrounds discussed below are all peaking backgrounds
312 and are reconstructed with $m(K^+\ell^+\ell^-)$ below the nominal B mass, unless a particle is
313 misidentified. Where possible, these partially reconstructed backgrounds are reduced by
314 the use of specific criteria.

315 When a selection requirement is optimised, or when the fraction of a background to
316 signal yield is estimated, the branching fractions are always taken from Ref. [21]. In
317 particular, the branching fractions for the rare modes over the whole q^2 range are:

$$\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-) = (5.5 \pm 0.7) \times 10^{-7} \quad (9)$$

$$\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-) = (4.43 \pm 0.24) \times 10^{-7} \quad (10)$$

318

319 In the next subsections the different types of peaking backgrounds are first described
320 and then the different steps in the selection- the stripping, offline, trigger and multivariate
321 selections - are each detailed.

322 **5.1 Types of backgrounds**

323 In this section, the different types of backgrounds are described and the naming conventions
324 are introduced. The selection requirements used to reduce these backgrounds, as well as
325 the expected yields remaining after this selection is applied, are described in the following
326 sections.

327 **5.1.1 Charmed and Strange Partially Reconstructed Backgrounds**

328 One type of peaking background stems from partially reconstructed b -hadron decays where
329 one or several tracks are missing. The first type of partially reconstructed background is
330 a $b \rightarrow s\ell^+\ell^-$ decay where the s quark hadronises into an excited strange resonance. For
331 example, the decay $B^0 \rightarrow K^{*0}(K^+\pi^-)\ell^+\ell^-$, where the π^- is not reconstructed. This type
332 of partially reconstructed background is labelled as “strange”.

333 The second type of background is mediated by a hadronic $V_{cb} \cdot V_{cs}$ transition $b \rightarrow c\bar{c}s$,
334 where the $c\bar{c}$ system hadronises into an excited charmonium resonance, it is labelled as
335 “charmed”. For example, $B^+ \rightarrow \psi(2S)(J/\psi\pi^0)K^+$ where the π^0 is lost.

336 These backgrounds peak at a lower mass than the rare/normalisation decay modes
337 because of the missing energy carried by the particles that are missed. Due to the poorer
338 mass resolution of the electron modes, these backgrounds can leak closer to signal in the
339 case of the electron channels.

340 5.1.2 Peaking backgrounds with a single mis-identified track

341 In both the rare and normalisation modes, there is only one fully-reconstructed background
342 with a single mis-identified track: $B^+ \rightarrow J/\psi\pi_{\rightarrow K^+}^+$ or $B^+ \rightarrow \ell^+\ell^-\pi_{\rightarrow K^+}^+$, where the
343 notation $\pi_{\rightarrow K^+}^+$ indicates that the π^+ is incorrectly identified as a K^+ . This background is
344 expected to peak right under the signal, but is suppressed by a factor $|V_{cd}/V_{cs}|^2$ in the
345 case of the normalisation modes and $|V_{td}/V_{ts}|^2$ in the case of the rare decay modes.

346 5.1.3 Peaking backgrounds with two mis-identified tracks

347 The first fully reconstructed double mis-ID background stems from $B^+ \rightarrow K^+\pi_{\rightarrow\ell^+}^+\pi_{\rightarrow\ell^-}^-$
348 decays with both pions mis-identified as leptons. The branching fraction of this decay
349 is about 30 times larger than the rare decay [21]. Particle identification requirements
350 are therefore necessary to suppress it. There is no equivalent background for the control
351 modes, as $J/\psi \rightarrow \pi^+\pi^-$ and $\psi(2S) \rightarrow \pi^+\pi^-$ decays are forbidden.

352 A second kind of double mis-ID background is referred to as “swaps”. For example,
353 a $B^+ \rightarrow K_{\rightarrow\ell^+}^+ J/\psi(\ell_{\rightarrow K^+}^+ \ell^-)$ transition where the K^+ is mis-identified as the ℓ^+ and
354 vice-versa could end up in the $1.1 - 6$ GeV 2 q^2 -range and pollute the rare mode.

355 5.1.4 Cascade backgrounds

356 In this analysis, partially reconstructed semileptonic V_{cb} decays, followed by a semileptonic
357 V_{cs} decay, are referred to as “cascade decays”. They are decays of the type $H_b \rightarrow$
358 $H_c(K^+l^-\bar{\nu}_l X)l^+\nu_l Y$, where H_b is a beauty hadron (B^+ , B^0 , B_s or Λ_B), H_c an open charm
359 hadron (D^0 , D^+ , D^* , etc.), and X , Y are missing particles. It is also possible that such
360 decays can be selected with a mis-identified particle in the final state. In that case, the
361 decay can be $H_b \rightarrow H_c(K^+\pi_{\rightarrow e^-}^- X)l^+\nu_l Y$ or $H_b \rightarrow H_c(K^+l^-\bar{\nu}_l X)\pi_{\rightarrow e^+}^+ Y$, where again the
362 notation $a \rightarrow b$ indicates that the real particle a has been mis-identified as b . As at least
363 two neutrinos are missing, these decays are partially reconstructed backgrounds and peak
364 at lower masses than the fully reconstructed signal, but they can have branching fractions
365 on the order of a few percent i.e. approximately four orders of magnitude larger than the
366 rare decay signal modes.

367 Backgrounds with semileptonic V_{ub} decays have much lower branching fractions
368 ($\mathcal{O}(10^{-4})$) and contain mis-identified particles. The dominant B^+ decays have two mis-
369 identified particles in the final state. These are $B^+ \rightarrow \rho e^+\nu$, $B^+ \rightarrow \omega e^+\nu$, $B^+ \rightarrow \eta e^+\nu$,
370 where ρ , ω and η decay to a final state with two pions, one being mis-identified as the e^-

371 and the other as the K^+ . There are also B_s decays with one mis-identified particle in the
372 final state. These are decays of the form $B_s \rightarrow H_s^- e^+ \nu$, where H_s^- is a strange resonance
373 decaying to K^- and several pions, one of which must be mis-identified as e^- .

374 5.1.5 Backgrounds with converted photons

375 Another type of background to be considered is any fully or partially reconstructed decay
376 containing a photon in the final state that converts into a di-electron pair in the material
377 of the detector. This type of background is expected to lie at very low q^2 , but could
378 potentially contaminate the signal q^2 region due to resolution effects.

379 5.2 The Selection Chain

380 The selection consists of the following steps:

- 381 1. Reconstruction and stripping selection: the latter is used to reduce the size of the
382 data samples to a manageable level;
- 383 2. Offline selection:
 - 384 (a) Selection of the q^2 range that differentiate the rare and control modes from
385 each others;
 - 386 (b) Fiducial requirements to simplify the efficiency computation and vetoes to
387 reduce cascade backgrounds;
 - 388 (c) Particle identification (PID) requirements to reduce mis-ID backgrounds;
- 389 3. Trigger requirements;
- 390 4. Multivariate selection to further reduce the combinatorial background;
- 391 5. Finally, selection of a range in $m(K^+ \ell^+ \ell^-)$ mass in which a fit is made.

392 The following sections describe each of these steps in more detail.

393 5.3 Stripping selection

394 For the 2011, 2012, 2015 and 2016 datasets, stripping versions 21, 21r1, 24 and 28
395 are used, respectively. The stripping line `B2XMuMuLine` is used for the muon mode and
396 `Bu2LLKeeLine2` is used for the electron mode. Both lines make similar requirements:
397 selected tracks must satisfy loose quality and particle identification criteria and have a
398 three track combination that has a well-defined vertex that is significantly displaced from
399 any primary vertex. The stripping requirements are summarised in Tab. 1.

400 The previous R_K analysis based on Run 1 data used stripping version 20. Significant
401 changes in the reconstruction of calorimeter objects have happened since then, which
402 improve the electron reconstruction efficiency by $\sim 50\%$. Using fully selected $B^+ \rightarrow$

403 $K^+J/\psi(e^+e^-)$ data from stripping version 20 and 21, it is determined that, among the
 404 events used in the previous analysis, $\sim 80\%$ are in common with the data sample used in
 405 the present analysis.

406 The **B2XMuMuLine** used in Run 1 contains a trigger filter which requires a set of low
 407 level and high level trigger lines to have fired. As described in Section 5.5, later in the
 408 selection chain, the B candidate is required to have fired some trigger lines that are all
 409 contained in the stripping trigger filter, which therefore has no effect for this analysis. In
 410 Run 2, this filter was removed.

411 In addition to these stripping lines, the line **Bu2LLKmeLine** is used to select the $K^+e^+\mu^-$
 412 candidates that are used as a proxy for the combinatorial background (see Section 5.6.1).
 413 This line is very similar to the **Bu2LLKeeLine2**, except for the PID e requirement which is
 414 replaced by an **isMuon** requirement.

Table 1: List of stripping requirements for the two stripping selection lines that are used to select the electron and muon modes. The particle or particle combination on which the requirement is applied is indicated in the first column. The requirements in blue are tightened by the offline selection.

		Bu2LLKeeLine2	B2XMuMuLine
B^+	$\chi_{to\ PV}^2$	> 100	> 121
	$\cos(\text{DIRA})$	< 0.995	< 0.9999
	χ_{IP}^2	< 25	< 16
	χ_{DV}^2/ndof	< 9	< 8
	$m(\text{GeV})$	$> 4.28 \text{ (Run 1)}$ $> 3.78 \text{ (Run 2)}$ $< 6.28 \text{ (Run 1)}$ $< 6.78 \text{ (Run 2)}$	> 4.9
$\ell\ell$	$\chi_{DV \leftrightarrow PV}^2$	> 16	> 9
	χ_{DV}^2/ndof	< 9	< 12
ℓ	χ_{IP}^2	> 9	> 9
	$p_T(\text{MeV})$	> 300	—
	PID $_{e,\mu}$	> 0	> -3
all tracks	isMuon	—	true
	prob $_{\text{ghost}}$	—	$< 0.35 \text{ (Run 1)}$ $< 0.5 \text{ (Run 2)}$
	χ_{TrackFit}^2	$< 3 \text{ (Run 2)}$	—
K	χ_{IP}^2	> 9	> 6
	$p_T(\text{MeV})$	> 400	—
event	nSPDHits	< 600	< 600

415 **5.4 Offline selection**

416 All the selection requirements are stored as a `string` in `Super-`
 417 `Lenin/src/getPreselection.cpp`.

418 The offline selection requirements are displayed in Tab. 2 and are explained in the
 419 following sections. Apart from the q^2 cuts discussed in Section 5.4.1, the same offline
 420 requirements are applied to the signal and control modes.

Table 2: Offline selection cuts applied to the electron modes (left) and muon modes (right).

Event quality		Event quality	
nSPDHits	< 600 (Run 1)	nSPDHits	< 600 (Run 1)
	< 450 (Run 2)		< 450 (Run 2)
prob _{ghost} (K, e)	< 0.3	prob _{ghost} (K, e)	< 0.3
Cascade vetoes		Cascade & mis-ID vetoes	
$m(K^+e^-)$	> 1885 MeV	$m(K^+\mu^-)$	> 1885 MeV
$m_{e\rightarrow\pi}^{\text{track}}(K^+e^-)$	notin $m(D^0) \pm 40$ MeV	$m_{\mu\rightarrow\pi}(K^+\mu^-)$	> 1885 MeV
Fiducial cuts		$m_{K\rightarrow\mu}(K^+\mu^-)$ notin $m(J/\psi) \pm 60$ MeV	
$p_T(e)$	> 0.5 GeV	$m_{K\rightarrow\mu}(K^+\mu^-)$ notin $m(\psi(2S)) \pm 60$ MeV	
$p(e)$	> 3 GeV	Fiducial cuts	
hasRich(K, e)	= true	inMuonAcc(K, μ)	= true
hasCalo(e)	= true	hasRich(K, μ)	= true
$x_{ECAL}(e)$	> 363.6 mm	$p_T(\mu)$	> 0.8 GeV
or $y_{ECAL}(e)$	> 282.6 mm	PID cuts	
PID cuts		probNN _K (K)	> 0.2
probNN _K (K)	> 0.2	isMuon(K)	= false
PID _e (K)	< 0	PID _{μ} (μ)	> -3
PID _e (e)	> 3	isMuon(μ)	= true

421 **5.4.1 Definition of q^2 and mass ranges**

422 The q^2 ranges used to differentiate between the rare and control modes are summarised
 423 in Tab. 3.

424 As mentioned in Section 1, the lower bound of the q^2 range has been increased from
 425 1 GeV² used in the Run 1 R_K analysis to the 1.1 GeV² that is used in other $b \rightarrow s\ell^+\ell^-$
 426 analyses. This allows the contamination from $B^+ \rightarrow K^+\phi(\rightarrow \ell^+\ell^-)$ to be minimised and
 427 is particularly important for the electron mode where, because of the bremsstrahlung
 428 recovery, events can migrate to higher q^2 . The reconstructed q^2 shape for simulated
 429 $\phi(\rightarrow e^+e^-)$ events is shown in Fig. 7. Based on the $B^+ \rightarrow K^+\phi(\rightarrow \ell^+\ell^-)$ branching
 430 fraction measured in [31], and the branching fraction for $B^+ \rightarrow K^+e^+e^-$ from [21], the

Table 3: Signal and control samples and their corresponding q^2 and mass ranges. These are the only requirements that are different between the rare mode and the control mode.

	ee mode	$\mu\mu$ mode
signal mode	$1.1 < q^2 < 6.0 \text{ GeV}^2$	$1.1 < q^2 < 6.0 \text{ GeV}^2$
J/ψ mode	$4.88 < m(K^+e^+e^-) < 6.20 \text{ GeV}$ $6.00 < q^2 < 12.96 \text{ GeV}^2$ $5.08 < m_{\text{DTF}}^{J/\psi}(K^+e^+e^-) < 5.70 \text{ GeV}$	$5.18 < m(K^+\mu^+\mu^-) < 5.60 \text{ GeV}$ $8.68 < q^2 < 10.09 \text{ GeV}^2$ $5.18 < m_{\text{DTF}}^{J/\psi}(K^+e^+e^-) < 5.60 \text{ GeV}$
$\psi(2S)$ mode	$9.92 < q^2 < 16.40 \text{ GeV}^2$ $5.08 < m_{\text{DTF}}^{\psi(2S)}(K^+e^+e^-) < 5.70 \text{ GeV}$	$12.5 < q^2 < 14.2 \text{ GeV}^2$ $5.18 < m_{\text{DTF}}^{\psi(2S)}(K^+e^+e^-) < 5.60 \text{ GeV}$

⁴³¹ $B^+ \rightarrow K^+\phi(\rightarrow e^+e^-)$ yield in the range $1.0 < q^2 < 6.0 \text{ GeV}^2$ ($1.1 < q^2 < 6.0 \text{ GeV}^2$) is
⁴³² expected to be 0.5% (0.1%) of the $B^+ \rightarrow K^+e^+e^-$ yield.

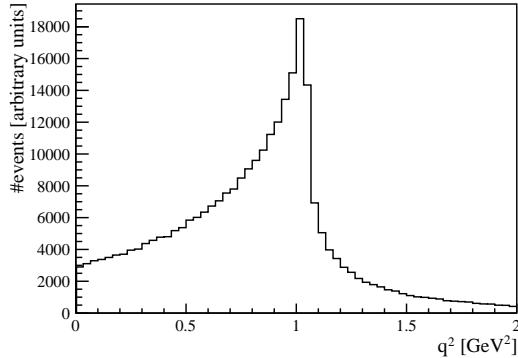


Figure 7: Reconstructed q^2 shape for simulated $B^+ \rightarrow K^+\phi(\rightarrow e^+e^-)$ events. No centrally produced simulation sample of these events is available, so the q^2 resolution model at $m(e^+e^-) = m(\phi)$ is taken from simulated $B^+ \rightarrow K^+e^+e^-$ events, selecting $0.5 < q^2 < 1.5 \text{ GeV}^2$.

⁴³³ The lower q^2 bound used is sufficient to reduce any backgrounds with a converted
⁴³⁴ photon to a completely negligible level. This is demonstrated using a sample of simulated
⁴³⁵ $B^0 \rightarrow K^{*0}\gamma$ events, where some of the photons convert to a di-electron pair. The q^2
⁴³⁶ distribution for this decay is shown in Fig. 8: out of the 2×10^6 simulated events, none is
⁴³⁷ reconstructed in the signal q^2 region.

⁴³⁸ The q^2 variable can be computed using a track and vertex fit that constrains the
⁴³⁹ $K^+e^+e^-$ mass to the nominal B mass: $m(K^+e^+e^-) = m(B^+)$ using the `DecayTreeFitter`
⁴⁴⁰ package [32]. This new q^2 variable is denoted q_{DTF}^2 and has a much better resolution than
⁴⁴¹ the standard q^2 that is computed without any constraint, as shown in Fig. 9. However,
⁴⁴² to extract the yields of the control modes, a fit is performed to the constrained mass
⁴⁴³ $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-)$ and $m_{\text{DTF}}^{\psi(2S)}(K^+e^+e^-)$, for the J/ψ and $\psi(2S)$ modes respectively. These
⁴⁴⁴ constrained masses are computed constraining $m(e^+e^-)$ to the nominal J/ψ or $\psi(2S)$

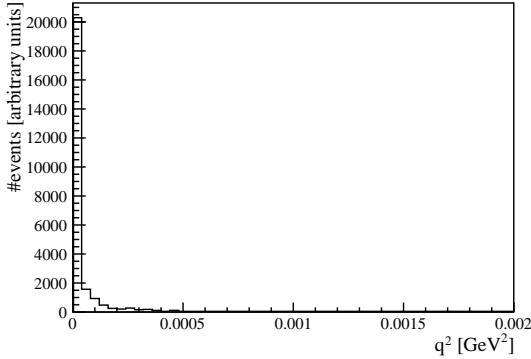


Figure 8: Reconstructed q^2 shape for simulated $B^0 \rightarrow K^{*0}\gamma$ events, where the photon has converted into a di-electron pair

masses. If a cut is imposed on q_{DTF}^2 , the background shapes in $m_{\text{DTF}}(K^+e^+e^-)$ are warped and the fit is made very difficult, as can be seen in Fig. 10. It would in principle be possible to select the rare mode using q^2 and the control modes using q_{DTF}^2 . However, in order to keep the rare and control samples mutually exclusive, and to keep the analysis simple, the unconstrained q^2 variable is the one used to select both signal and control modes, and the constrained q_{DTF}^2 is used solely for cross-checks.

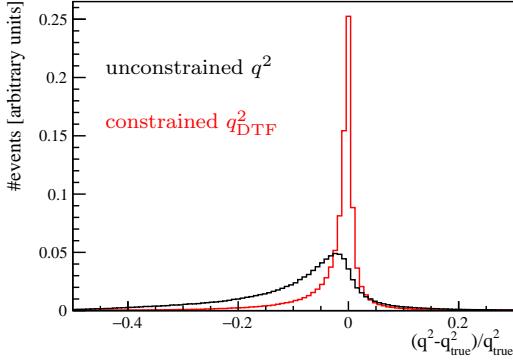


Figure 9: Resolution in q^2 in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$ from simulated $B^+ \rightarrow K^+e^+e^-$ events. In the black (red) distribution, the unconstrained (constrained) q^2 is used.

5.4.2 Fiducial cuts: acceptance of the PID system

The p and p_T cuts are applied to ensure that the signal spans the same phase space as the calibration samples used to determine the efficiency of the PID requirements (see Section 6.4.2). Similarly, tracks are required to be within the acceptance of the

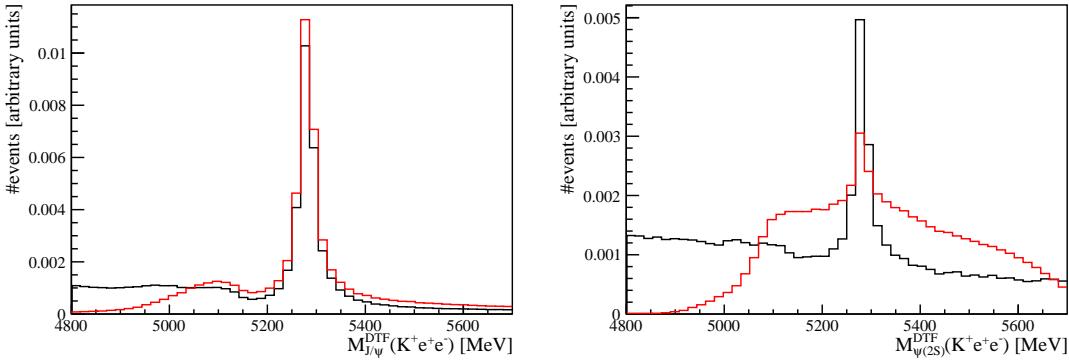


Figure 10: Constrained mass distributions for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (left) and $B^+ \rightarrow K^+ \psi(2S)(e^+e^-)$ (right) candidates from Run 1 selected data. All distributions are normalised to unit area. The black distributions correspond to events selected using the unconstrained q^2 , whereas the red distributions are obtained using the constrained q^2_{DTF} variable. It is clear from these plots that using q^2_{DTF} warps the background shapes at masses below the nominal B meson mass. Moreover, there is more background in the $m_{\text{DTF}}^{\psi(2S)}(K^+e^+e^-)$ plot due to leakage from $B^0 \rightarrow K^{*0}J/\psi(e^+e^-)$ events.

relevant sub-detectors: muons are required to be within the muon chambers (`inMuonAcc`), electrons to have associated hits in the calorimeters (`hasCalo`), and all tracks must have associated hits in the RICH detectors (`hasRich`).

5.4.3 Fiducial cuts: calorimeter acceptance

The criteria $x_{ECAL} < 363.6$ mm or $y_{ECAL} < 282.6$ mm are used to veto electron candidates which would have hit calorimeters cells which were not read out. As can be seen in Fig. 11, which shows $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ candidates selected from Run 1 data and simulated events, to which the full preselection has been applied, a drop in both particle identification and trigger efficiency is observed for electron tracks in this region. This requirement rejects less than 1% of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events. As the trigger and PID efficiencies are not corrected in fine bins of x_{ECAL} and y_{ECAL} , as detailed in Section 6.6, vetoing this region guarantees a better agreement between data and simulation.

5.4.4 Cascade background vetoes

Cascade backgrounds are suppressed by requiring $m(K^+\ell^-) > m(D^0)$. Mis-ID cascades, i.e. $H_b \rightarrow H_c(K^+\pi^-)\ell^+\nu_\ell y$, where the pion is mis-identified as a lepton, are suppressed by a similar veto, switching the lepton mass hypothesis to that of the pion. In the muon case, the veto is simply $m_{\mu \rightarrow \pi}(K^+\mu^-) > m(D^0)$. In the electron case, the D^0 mass is computed switching off the bremsstrahlung recovery to improve the mass resolution and the veto is: $|m_{e \rightarrow \pi}^{\text{track}}(K^+e^-) - m(D^0)| > 40$ MeV.

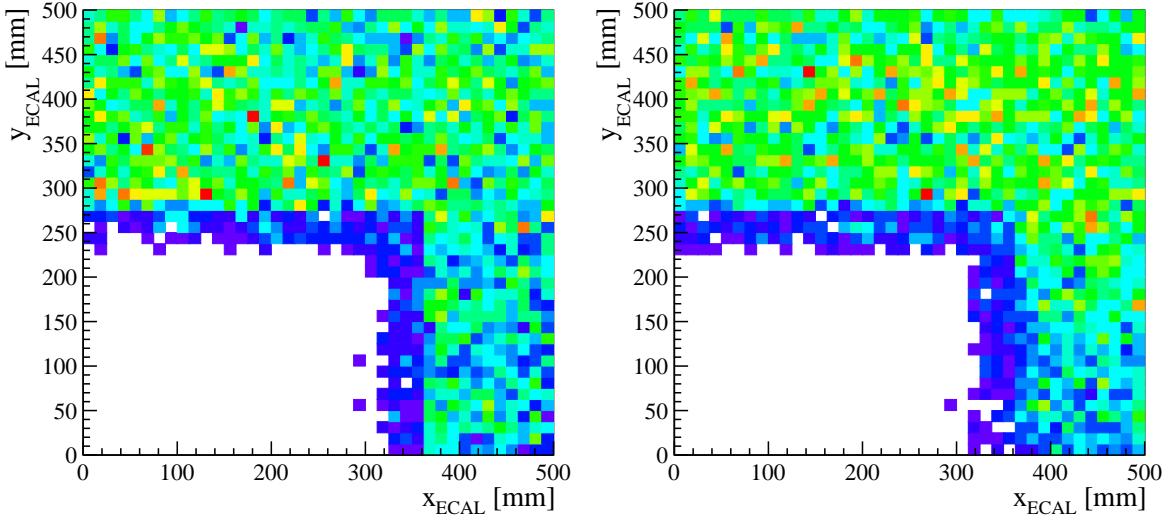


Figure 11: Distributions of electron candidates from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events as a function of x_{ECAL}, y_{ECAL} for (left) simulated events and (right) data corresponding to the Run 1 data taking conditions. The full offline selection has been applied, except the veto of the central calorimeter region. The region plotted is towards the centre of the calorimeter. The region with a lower occupancy appear blue on the plots. It corresponds to cells that are not read out, necessitating the fiducial cuts described in the text.

The cascade vetoes retain $\sim 97\%$ of offline-selected $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $\sim 95\%$ of $B^+ \rightarrow K^+ e^+ e^-$ events, and reduce the cascade backgrounds to a very small level. The expected yields, N^{exp} , for simulated cascades events, after applying the offline selection including the cascades vetoes, as well as the mass window cut described in Section 5.7 can be seen in Tab. 4. It can be seen that a small contamination from $B^+ \rightarrow \bar{D}^0 (\rightarrow K^+ \ell^- \bar{\nu}_\ell) \ell^+ \nu_\ell$ remains after applying the veto. This component is taken into account in the fit to the rare mode by including it in the partially reconstructed background shape.

5.4.5 Particle identification

To reduce mis-identified backgrounds and swaps, particle identification (PID) requirements are used. As in previous analyses of the $B^+ \rightarrow K^+ \ell^+ \ell^-$ decay [31, 33, 34], in order to identify the kaon, a $\text{probNN}_K(K) > 0.2$ requirement is applied in both muon and electron modes (using tunings MC12TuneV2 for Run 1 and MC15TuneV1 for Run 2). To reduce backgrounds containing mis-identified leptons, in the muon mode a combination of isMuon and PID_μ cuts are used; and in the electron mode a requirement on PID_e is used. These requirements reduce the peaking mis-identified backgrounds to a negligible level, as can be seen in Tab. 4, where the fraction of expected yields between the rare modes and peaking backgrounds are shown, assuming the rare mode branching fractions from [21]. In the fit to the rare modes, the peaking mis-identified backgrounds are therefore neglected. In

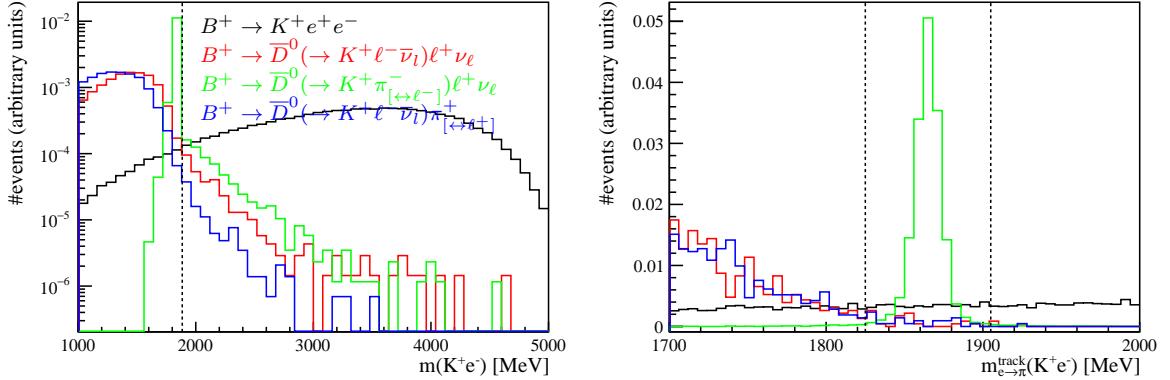


Figure 12: K^+e^- mass distributions from signal and cascade simulation samples. All the distributions are normalised to unity. In the right-hand plot, the mass is computed from the track information only, whereas, in the left-hand plot, the bremsstrahlung correction to the momentum of the electron is taken into account, resulting in a tail to the right. All plots are taken from offline-selected Run 1 simulated events.

Table 4: Expected level of cascade and mis-ID backgrounds as a fraction of the relevant $B^+ \rightarrow K^+\ell^+\ell^-$ signal yield, when the full offline selection has been applied, including the q^2 cut $1.1 < q^2 < 6.0$ GeV 2 . These estimates are computed using the branching fractions from Ref. [21] and the efficiencies are computed from the simulation samples with the PID calibrated as detailed in Section 6.4.

background type	$\frac{N^{\text{exp}}(\text{bkg})}{N^{\text{exp}}(B^+ \rightarrow K^+ e^+ e^-)}$	$\frac{N^{\text{exp}}(\text{bkg})}{N^{\text{exp}}(B^+ \rightarrow K^+ \mu^+ \mu^-)}$
cascades		
$B^+ \rightarrow \bar{D}^0 (\rightarrow K^+ \ell^- \bar{\nu}_\ell) \ell^+ \nu_\ell$	3.26%	0.00%
$B^+ \rightarrow \bar{D}^0 (\rightarrow K^+ \pi^-_{[\leftrightarrow \ell^-]} \ell^+ \nu_\ell)$	0.00%	0.00%
$B^+ \rightarrow \bar{D}^0 (\rightarrow K^+ \ell^- \bar{\nu}_\ell) \pi^+_{[\leftrightarrow \ell^+]}$	0.00%	0.00%
single and double mis-ID		
$B^+ \rightarrow \pi^+_{[\leftrightarrow K^+]} \ell^+ \ell^-$	0.41%	0.41%
$B^+ \rightarrow K^+ \pi^+_{[\leftrightarrow \ell^+]} \pi^-_{[\leftrightarrow \ell^-]}$	0.83%	0.60%
$B^+ \rightarrow K^+_{[\leftrightarrow \ell^+]} J/\psi (\ell^+_{[\leftrightarrow K^+]} \ell^-)$	0.04%	0.00%

the case of the normalisation modes, a contamination from mis-identified $B^+ \rightarrow \pi^+ J/\psi$ events remains at a level of 0.45% and is included in the fit. This component cannot be neglected due to the very high yield of the normalisation modes and are therefore included in the fit, as described in Section 7.

Compared to the Run 1 analysis of R_K [33], the PID_e requirement has been tightened

497 from $\text{PID}_e > 1$ to $\text{PID}_e > 3$. This reduces the number of combinatorial events containing a
498 pion misidentified as an electron and improves the signal purity. As shown in Section 5.6,
499 the cut $\text{PID}_e > 3$ maximises the significance of the fully selected rare mode.

500 5.5 Trigger

501 All the trigger requirements and trigger fiducial cuts are stored as `string` in *Super-*
502 *Lenin/src/getPreselection.cpp*.

503 To allow for better control of the selection efficiency, a specific set of trigger lines is
504 used that are either triggered on some element of the signal, triggered on signal (TOS), or
505 independent of the signal (TIS).

506 In the same way as in the Run 1 R_K and $R_{K^{*0}}$ analyses, the R_K ratio is extracted
507 separately, using electrons events triggered by three exclusive L0 categories, each using
508 different L0 trigger lines:

509 *e*TOS: the event is selected by the L0 trigger algorithm based on one of the two electrons
510 (`L0Electron` trigger line required to be TOS on one of the two electrons);

511 *h*TOS!: the event is selected by the L0 trigger algorithm based on the kaon (`L0Hadron` trigger
512 line required to be TOS on the kaon), excluding events from the previous category;

513 TIS!: the event is selected by the L0 trigger algorithm based on any other track in the
514 event, excluding events from the two previous categories.

515 Since the muon trigger is much more efficient than the calorimeter-based trigger, only
516 the muon trigger is used to select $K^+\mu^+\mu^-$ events (μ TOS category, requiring the `L0Muon`
517 trigger line to be TOS on one of the two muons).

518 To perform cross-checks, inclusive hadron and TIS categories are also used and are
519 denoted *h*TOS and TIS respectively.

520 The L0 and HLT requirements and lines used are all summarised in Tab. 5. The logical
521 *or* of lines at the same level, and the logical *and* of L0, HLT1 and HLT2 is required.

522 5.5.1 TCK and trigger fiducial cuts

523 This section deals with the transverse energy or momentum of tracks as measured by
524 the L0 system. These are denoted $E_T^{\text{L}0}$ and $p_T^{\text{L}0}$ and have far worse resolution than their
525 offline equivalents E_T and p_T . These quantities are measured in terms of ADC counts,
526 which can be converted into MeV using the conversion constants in Tab. 6. These values
527 are in practice provided by `TupleToolL0Data`. In the case of `L0Muon`, only the three
528 trigger candidates with the three highest values of $p_T^{\text{L}0}$ are saved, with no matching to
529 the two offline muon candidates used to reconstruct the decay $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ or
530 $B^+ \rightarrow K^+\mu^+\mu^-$. In the case of `L0Electron` or `L0Hadron`, only a unique trigger candidate
531 with the highest $E_T^{\text{L}0}$ value is saved, with no matching to the offline electron or kaon
532 candidate.

Table 5: Trigger requirements. A track or B candidate indicated in parentheses means that a TOS requirement is applied on that candidate.

Electron mode		Muon mode
Run 1		
L0	LOElectron(e) $E_T(e) > 3.0$ GeV LOHadron(K) & $E_T(K) > 3.5$ GeV LOGlobal(TIS)	L0Muon(μ) & $p_T(\mu) > 0.8$ GeV
HLT1	Hlt1TrackAllL0(B)	Hlt1TrackAllL0(B) Hlt1TrackMuon(B)
HLT2	Hlt2Topo [2,3] BodyBBDT(B) Hlt2TopoE [2,3] BodyBBDT(B)	Hlt2Topo [2,3] BodyBBDT(B) Hlt2TopoMu [2,3] BodyBBDT(B)
Run 2		
L0	LOElectron(e) & $E_T(e) > 2.7$ GeV LOHadron(K) & $E_T(K) > 3.5$ GeV LOGlobal(TIS)	L0Muon(μ) & $p_T(\mu) > 0.8$ GeV & $p_T^{L0}(\mu) \geq 37$ (2015), 57 (2016)
HLT1	Hlt1TrackMVA(B)	Hlt1TrackMVA(B)
HLT2	Hlt2Topo [2,3] BodyBBDT(B)	Hlt2Topo [2,3] BodyBBDT(B) Hlt2TopoMu [2,3] BodyBBDT(B)

Table 6: Conversion constants to transform ADC counts measured by the calorimeters (for $E_T(e)$ and $E_T(K)$) or the muon system (for $p_T(\mu)$), into MeV.

	Run 1	Run 2
muon system	40	50
calorimeters	20	24

533 The L0Muon trigger line, defining the μ TOS trigger category, requires the p_T^{L0} of a
 534 muon candidate to be above a certain threshold. Similarly, the LOElectron and LOHadron
 535 lines require E_T^{L0} of the electron or hadron candidate to be above a certain threshold.
 536 These thresholds are fixed in the simulation, whereas they vary in the data. The trigger
 537 configuration is saved and indexed by a “trigger configuration key” (TCK). All the TCK’s
 538 that have been used in this analysis are reported in Tab. 8 (Run 1) and Tab. 9 (Run 2). The
 539 different thresholds used in the data and simulation samples can give rise to inconsistencies
 540 in the L0 trigger efficiencies between these two samples. The procedure chosen to reduce
 541 these inconsistencies in each L0 category is explained below.

542 μ TOS From Tab. 8, it can be seen that the L0Muon thresholds were not changed a great
 543 deal during Run 1: they have been kept the same for 97% of the luminosity collected
 544 in 2012 and 99% of the luminosity collected in 2011. From Tab. 9, it can be seen

545 that for Run 2, however, the L0Muon thresholds change frequently, and, for both
 546 2015 and 2016, the threshold in the simulation is set to the highest value in the data.
 547 To ensure a better agreement between simulation and data, $p_T^{\text{L}0}(\mu)$ is required to be
 548 above the simulation threshold ($p_T^{\text{L}0}(\mu) \geq 57$ in 2015 and $p_T^{\text{L}0}(\mu) \geq 37$ in 2016 data
 549 taking conditions). As there is no matching between the trigger candidate from which
 550 $p_T^{\text{L}0}(\mu)$ is measured and the two offline muon candidates, the requirement of $p_T^{\text{L}0}(\mu)$
 551 could introduce a bias. In fact, consider some data taken with a L0Muon threshold t_a ,
 552 that is tightened to t_b ($t_b > t_a$). If, in a given event, one of the two muon candidates,
 553 μ_1 satisfies $t_a < p_T^{\text{L}0}(\mu_1) < t_b$ and the other, μ_2 , satisfies $p_T^{\text{L}0}\mu_2 < t_b$, and if there is
 554 a third muon in the event, independent of the signal decay, satisfying $p_T(\mu_3) > t_b$,
 555 then that event is classified as TOS when it should not, because neither of μ_1 , μ_2
 556 satisfies $p_T^{\text{L}0}(\mu) > t_b$. From simulated $B^+ \rightarrow K^+\mu^+\mu^-$ and $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$
 557 events, it is estimated that this situation occurs in less than 0.4% of events, meaning
 558 that this bias is negligible.

559 On top of the requirement on $p_T^{\text{L}0}(\mu)$, both muon candidates are required to satisfy
 560 $p_T(\mu) > 800$ MeV. This has nothing to see with the trigger but simplifies the
 561 calibration of the PID requirements, as discussed in Section 5.4.2.

562 Remaining differences between data and simulation, related to differences in resolution
 563 in $p_T^{\text{L}0}$, are corrected using the data, as shown in Section 6.6.3.

564 *e*TOS As can be seen in both Tab. 8 and Tab. 9, the L0Electron thresholds also vary in
 565 the data with different TCK’s, although not as dramatically as in the Run 2 L0Muon
 566 case discussed above. A requirement is placed on the (offline) E_T of the electron
 567 that has fired the trigger, that is tighter than the L0Electron requirement. This
 568 requirement is $E_T > 3.0$ GeV in Run 1 and $E_T > 2.7$ GeV in Run 2. This eliminates
 569 electrons that have a transverse energy far below the L0Electron threshold but
 570 are still classified as TOS because of noise in the calorimeter, and simplifies the
 571 calibration of the L0Electron efficiency, as explained in Section 6.6.1.

572 *h*TOS The situation for *h*TOS is similar to that for *e*TOS. A requirement is placed on
 573 the (offline) kaon E_T , $E_T > 3.5$ GeV, that is tighter than the L0Hadron requirement.
 574 This requirement is the same for all data taking conditions. Remaining differences
 575 between data and simulation, related to differences in resolution in $E_T^{\text{L}0}$, are corrected
 576 using the data as discussed in Section 6.6.5.

577 TIS No extra fiducial cut is placed other than requiring the L0Global line to be TIS.
 578 Remaining differences between data and simulation are correct as described in Sec-
 579 tion 6.6.8.

580 As for the L0 requirements, the HLT1 requirements also vary depending on the TCK,
 581 but are fixed in the simulation and can therefore give rise to inconsistencies. This is
 582 handled in two different ways in Run 1 and Run 2. In Run 1, the offline reconstruction is
 583 different from the HLT1 reconstruction. The efficiency of the HLT1 requirements can be
 584 expected to be different between data and simulation even at a fixed TCK, and this is

585 corrected using data calibration samples, as detailed in Section 6.7. In Run 2, however,
 586 the offline and HLT reconstructions are the same. An extra requirement is therefore made
 587 offline that aligns the HLT requirement in data and simulation, negating the need for any
 588 extra corrections at the HLT level. The HLT1 line used in this analysis, **trackMVA**, has a
 589 parameter, denoted b , that changes as a function of the TCK. This parameter appears in
 590 a 2D (p_T, χ_{IP}^2) cut that is used which is given,

$$\log(\chi_{IP}^2) > \frac{1.0}{(p_T[\text{GeV}] - 1.0)^2} + \frac{b}{25} \cdot (25 - p_T[\text{GeV}]) + \log(7.4).. \quad (11)$$

591 The effect of this 2D cut is shown in Fig. 13. In the simulation and for most of the
 592 data, $b = 1.1$, but for 11% of 2016 MD data and 55% of 2016 MU data, $b = 2.3$ is used
 593 (see Tab. 9). Therefore, for 11% of simulated 2016 MD events and 55% of 2016 MU events,
 594 the HLT1 **trackMVA** requirement is tightened by applying offline the requirement of Eq. 11
 595 with $b = 2.3$.

596 This analysis does not use the line **TwoTrackMVA** at the HLT1 level. Although adding
 597 this line enhances the HLT efficiencies for both the normalisation channels and the rare
 598 channels in Run 2 by approximately 3%, the line is not used because it contains a cut
 599 on some multivariate algorithm which is technically difficult to align between data and
 600 simulation.

Table 7: Trigger fiducial cuts. The data-taking condition for which the fiducial cut is applied is shown in the right column.

line	fiducial cut	condition
L0Muon	$p_T^{L0} > 37 \text{ ADC}$	2016
	$p_T^{L0} > 57 \text{ ADC}$	2015
L0Electron	$E_T > 3.0 \text{ GeV}$	Run 1
	$E_T > 2.7 \text{ GeV}$	Run 2
L0Hadron	$E_T > 3.5 \text{ GeV}$	all
HLT1 trackMVA	Eq. 11 with $b = 1.1$	2015
		88.8% of 2016 MD
		45.0% of 2016 MU
	Eq. 11 with $b = 2.3$	11.2% of 2016 MD 55.0% of 2016 MU

601 All the trigger fiducial requirements discussed in this section are summarised in Tab. 7.
 602 Applying these fiducial cuts greatly simplifies the evaluation of the trigger efficiencies
 603 and results in a loss of $\sim 2\%$ of eTOS $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, $\sim 2.5\%$ of eTOS
 604 $B^+ \rightarrow K^+ e^+ e^-$ events, $\sim 6\%$ of $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events, hence having a negligible
 605 impact on the R_K double ratio.

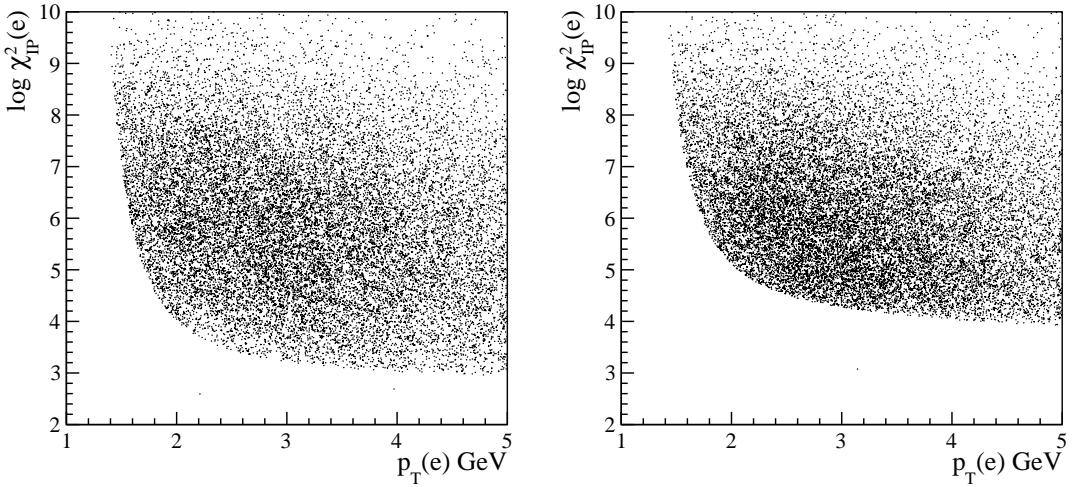


Figure 13: Distribution of $p_T(e) \times \log \chi^2_{IP}(e)$ for electrons selected from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ candidates from the Run 2 data. The electrons are required to satisfy the TrackMVA HLT1 requirement. In the left plot, the data selected correspond to TCK’s which sets $b = 1$ in Eq. 11. The right plot corresponds to $b = 2.3$.

606 5.6 Multivariate selection

607 The training of the BDT is done in *Super-*
 608 *Lenin/src/mainTrainBDTForSuperLeninMuMu.cc* and *Super-*
 609 *Lenin/src/mainTrainBDTForSuperLenin.cc* The optimisation and test of
 610 the BDT can be found in *SuperLenin/src/superMain15Run2.cc* and *Su-*
 611 *perLenin/src/superMain15.cc*, which rely mostly on functions defined in
 612 *SuperLenin/src/plotBDT.cpp*.

613 To suppress combinatorial events, TMVA Boosted Decision Trees (BDT) [35, 36] are
 614 trained. The training of the BDT is explained in the next section. The optimisation as
 615 well the expected performance of the BDT is explained in the following section.

616 5.6.1 Multivariate selection training

617 The BDT’s are trained using data events taken from the upper mass sideband
 618 ($m(K^+\ell^+\ell^-) > 5.4$ GeV) as the background sample, and simulated $B^+ \rightarrow K^+\ell^+\ell^-$
 619 events as the signal sample. Note that this is different from the previous analysis where the
 620 BDT was trained used background-subtracted $B^+ \rightarrow K^+J/\psi(\ell^+\ell^-)$ data events [33]. The
 621 BDT is however suboptimal for the rare mode in the range $1.1 < q^2 < 6.0$ GeV 2 if trained
 622 on $B^+ \rightarrow K^+J/\psi(\ell^+\ell^-)$, as the signal it learns from has leptons with a larger transverse
 623 momentum than in the rare mode. The full preselection requirements are applied to both
 624 the signal and background samples before training the classifier. All correction weights

Table 8: TCK's and L0 thresholds used in Run 1 data.

TCK hex	% lumi	$p_T^{L0}(\mu)$	$E_T^{L0}(e)$	$E_T^{L0}(K)$
2011 data				
0x360032	0.248	37	125	175
0x5a0032	12.68	37	125	175
0x5d0033	0.872	20	125	120
0x6d0032	1.647	37	125	175
0x730035	12.76	37	125	175
0x760037	15.11	37	125	175
0x790037	6.467	37	125	175
0x790038	50.07	37	125	175
0x7e003a	0.120	37	125	165
2011 MC				
		37	125	175
2012 data				
0x7f0040	0.019	37	125	175
0x860040	0.187	37	125	175
0x8c0040	2.998	37	125	175
0x94003d	10.42	44	136	181
0x97003d	3.965	44	136	181
0x990042	41.55	44	136	181
0x990044	11.39	44	148	184
0x9a0042	0.001	44	136	181
0x9f0045	1.111	44	148	187
0xa10044	2.335	44	148	184
0xa10045	3.259	44	148	187
0xa20044	0.149	44	148	184
0xa30044	9.991	44	148	184
0xa30046	2.364	44	143	187
0xa90046	1.269	44	143	187
0xab0046	2.677	44	143	187
0xac0046	6.292	44	143	187
2012 MC				
		44	148	187

discussed in Section 6 are applied to the simulated sample. The training and testing is performed using the k-folding technique [37] with ten-folds to enable the entire sample to be used to both train and test without inducing any significant statistical bias. Different BDT's are trained separately for Run 1 and Run 2 samples, and for the electron and muon modes. The total number of events used for training and testing are shown in Tab. 12.

Table 9: Summary of the HLT1 **TrackMVA** and L0 thresholds that change depending on the TCK. The luminosity recorded with each TCK is quoted in pb^{-1} . The L0 thresholds are given in ADC counts and the factor b in **TrackMVA** appears in the (χ^2_{IP}, p_T) 2D cut defined in Eq. 11

TCK (hex)	% lumi	lumi	TrackMVA		L0	
			2016 MD data	b	$E_T^{\text{L0}}(0)$	$p_T^{\text{L0}}(\mu)$
0x11291603	4.134	35.01		1.1	89	23
0x11291604	2.925	24.78		1.1	95	27
0x11291605	9.259	78.42		1.1	109	31
0x11371609	8.518	72.14		2.3	99	27
0x1137160e	2.686	22.756		2.3	99	27
0x11381609	0.810	6.86		1.1	109	31
0x1138160e	3.742	31.70		1.1	109	31
0x1138160f	67.923	575.25		1.1	101	37
2016 MU data			0.846916 fb^{-1}			
0x11321609	13.425	107.00		1.1	99	27
0x11341609	14.813	118.06		1.1	99	27
0x11351609	2.943	23.46		2.3	99	27
0x11361609	52.021	414.62		2.3	99	27
0x11381611	5.537	44.13		1.1	110	31
0x11381612	11.261	89.75		1.1	110	33
2016 simulation						
0x5138160f	—	—		1.1	101	37
2015 MD data			0.174506 fb^{-1}			
0x10600a2	55.17	96.28		1.1	76	39
0x10600a3	34.78	60.70		1.1	113	57
0x10600a6	0.028	0.049		1.1	113	57
0x10600a7	0.586	1.022		1.1	96	49
0x10700a1	3.642	6.355		1.1	71	45
0x10800a2	5.788	10.10		1.1	59	31
2015 MU data			$0.123.83 \text{ fb}^{-1}$			
0x10800a2	40.806	50.53		1.1	113	57
0x11400a8	59.194	73.30		1.1	113	57
2015 simulation				1.1	113	57
						151

630 Twelve discrimination variables are used in the classifier; they are listed in Tab. 11.
 631 The discrimination with respect to the combinatorial background relies on the transverse
 632 momenta of all the particles, the quality of the vertex fitting and the significance of the
 633 displacement of the tracks from the primary vertex. It has been checked that adding

Table 10: Total yields (all folds summed) used for training and testing in all trigger categories

	$\mu\mu$ mode		ee mode	
	signal	comb.	signal	comb.
Run 1	82106	7688	36166	4378
Run 2	68709	9998	53391	11562

extra kinematic variables to those shown in Tab. 11, such as the momenta of all the particles, the flight distance of the B^+ or the χ^2 of the decay tree fitter do not improve the discrimination power. Variables related to the PID are not included in the training, as their distributions are badly modelled in the simulation, and re-sampling the electron PID from the data is not envisaged as the statistics of the electron PID calibration sample is low, see Section 6.4.

Table 11: List of variables used by the BDT classifier.

B^+	$p_T, \log \chi_{IP}^2, \chi_{DV}^2, \text{DIRA}, \chi_{DV \leftrightarrow PV}^2$
ll	$p_T, \log \chi_{IP}^2$
K^+	$p_T, \log \chi_{IP}^2$
l	$\min, \max(p_T), \min, \max(\log \chi_{IP}^2)$

Two different boosting methods are tested: adaptive and gradient boosting. For all trigger categories and for the electron and muon mode, gradient boosting offers a better performance and is therefore the algorithm chosen in this analysis. The ROC curves illustrating the performance of the two boosting methods are shown in Fig. 14. The area under the ROC curve is smaller for the muon mode than for its electron equivalent, but the muon mode has much less background before the application of the BDT.

5.6.2 BDT optimisation and performance

To find the working point of the BDT, the expected significance \mathcal{S} :

$$\mathcal{S} = \frac{S}{\sqrt{S + B}},$$

is used as the figure of merit, where S and B are the number of signal and background events expected in the signal region. The signal region is defined as the mass range $5210 < M(K^+\mu^+\mu^-) < 5350$ for the muon mode and $5000 < M(K^+e^+e^-) < 5380$ for the electron mode.

In the case of the muon mode, the expected number of background events in the signal region is extrapolated from the sideband using a fit with an exponential function. In the

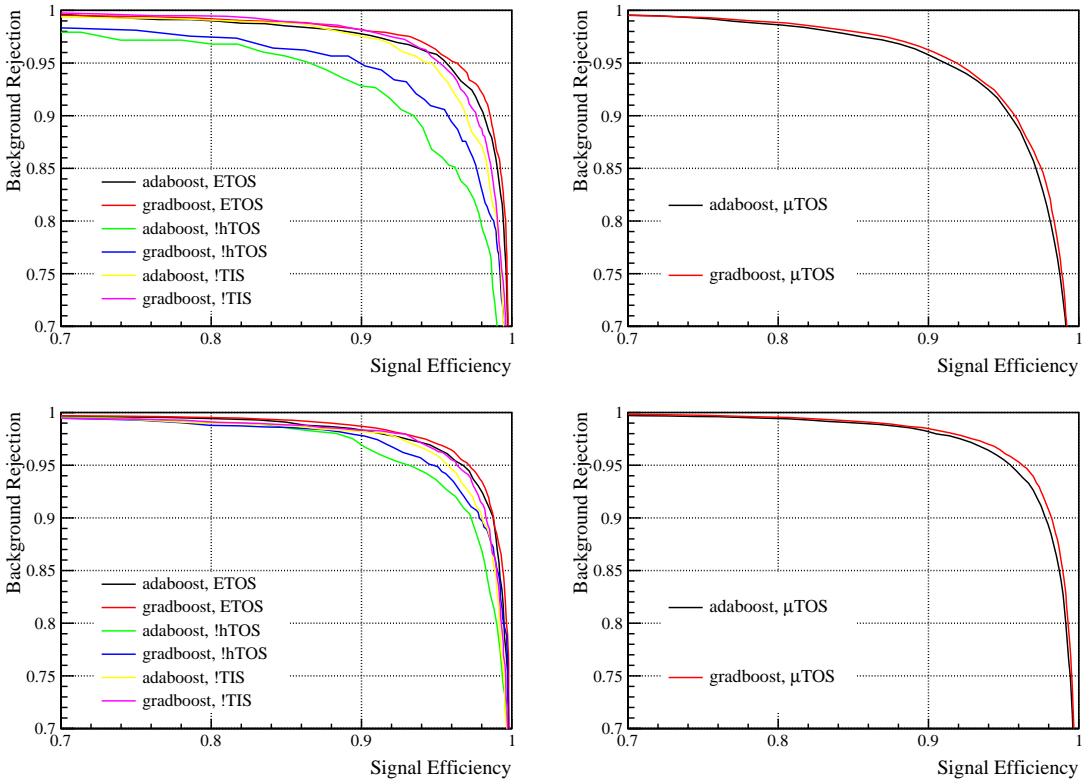


Figure 14: ROC curves for the rare modes in different trigger and data taking conditions: the top (bottom) plots correspond to Run 1 (Run 2), the plots on the left hand-side correspond to the electron mode in the three exclusive trigger categories, and those on the right hand-side to the muon mode. The curves corresponding to the BDT trained with the gradient boosting, and adaptive boosting methods are shown.

case of the electron mode, using such a fit is less precise, because the number of events in the sideband is small and the signal region wider. Instead, the $K^+e^+\mu^-$ sample is used to compute:

$$N_{\text{comb}}^{\text{SR}} = N_{\text{comb}}^{\text{SB}} \cdot \frac{N_{K^+e^+\mu^-}^{\text{SR}}}{N_{K^+e^+\mu^-}^{\text{SB}}},$$

where $N_{\text{comb}}^{\text{SR}}$ ($N_{\text{comb}}^{\text{SB}}$) is the number of combinatorial events in the signal region (sideband). $\frac{N_{K^+e^+\mu^-}^{\text{SR}}}{N_{K^+e^+\mu^-}^{\text{SB}}}$ is extracted with an exponential fit to $m(K^+e^+\mu^-)$. The fit is performed in the range $4880 < m(K^+e^+\mu^-) < 6250$ MeV for Run 1, and $4880 < m(K^+e^+\mu^-) < 6700$ MeV for Run 2, with the region $5100 < m < 5350$ MeV blinded (and excluded from the fit) to avoid any signal from lepton flavour violating modes. This fit is illustrated in Fig. 15. The

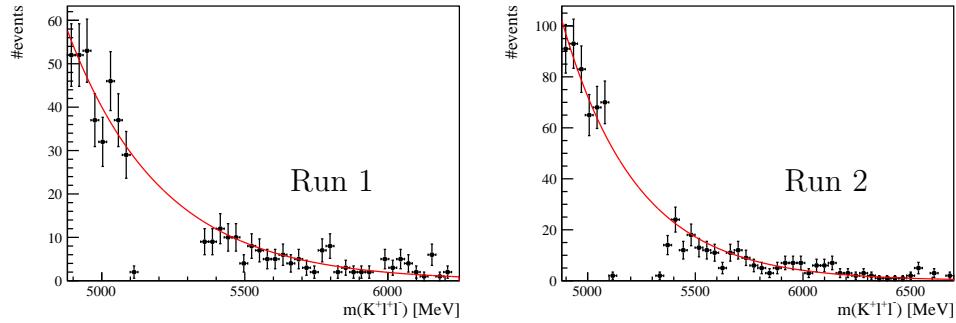


Figure 15: Illustration of the exponential fit to the $K^+e^+\mu^-$ sample to extrapolate the number of events from the sideband ($m > 5400$ MeV) to the signal region ($5000 < m(K^+e^+e^-) < 5380$ MeV). The region $5100 < m < 5350$ MeV is blinded and is excluded from the fit.

662 expected number of signal events S is computed using:

$$S = \varepsilon_{rare}^{BDT} \cdot \frac{\varepsilon_{J/\psi}^{presel}}{\varepsilon_{rare}^{presel}} \cdot \frac{\mathcal{B}(B^+ \rightarrow K^+\ell^+\ell^-)}{\mathcal{B}(B^+ \rightarrow K^+J/\psi(\ell^+\ell^-))} \cdot N^{presel}(B^+ \rightarrow K^+J/\psi), \quad (12)$$

663 where \mathcal{B} indicates a branching fraction from Ref. [21], $N(B^+ \rightarrow K^+J/\psi)$ is the number of
 664 preselected control events extracted from the fit described in Section 6.3, ε_{rare}^{BDT} , $\varepsilon_{J/\psi}^{presel}$ and
 665 $\varepsilon_{rare}^{presel}$ are the efficiencies of the BDT cut and preselection criteria on the control modes or
 666 rare modes. These efficiencies are computed using the fully-corrected simulation samples.

667 The expected significance \mathcal{S} for the electron mode is shown in Fig. 16 and 17, and in
 668 Fig. 18 for the muon mode. In the case of the electron mode, two different ways of training
 669 the BDT are tested: either training three separate BDT's, one for each trigger category,
 670 or training only one BDT common to all three trigger categories. The optimisation of
 671 \mathcal{S} shows that, for the eTOS category, using a dedicated eTOS BDT yields the largest
 672 significance, whereas for the hTOS! and TIS! categories, the BDT common to the three
 673 trigger categories is more optimal. This can be understood considering the fact that a
 674 classifier trained on hTOS! or TIS! tends to overfit, as these categories have few events.
 675 For the eTOS category, the BDT trained on eTOS is used. For the hTOS! and TIS!
 676 categories the BDT trained on all is used. To find the optimal working point of the BDT,
 677 the expected significance is fit with a fourth order polynomial, in order to smooth the
 678 curve and to mitigate statistical fluctuations.

679 Some supplementary checks of the BDT are presented in Appendix B. In particular, it
 680 is checked that the BDT working point and electron PID_e cut yield the optimal significance.
 681 The compatibility between the different folds and the overtraining plots are also shown.

682 The BDT optimal working points, as well as the expected signal and combinatorial
 683 yields in the signal window are shown in Tab. 12. In the case of the electron mode, the
 684 combinatorial yield is estimated either by using the $K^+e^+\mu^-$ sample as explained above,
 685 or, as a cross-check, by using an exponential fit to the $K^+e^+e^-$ upper mass sideband
 686 directly.

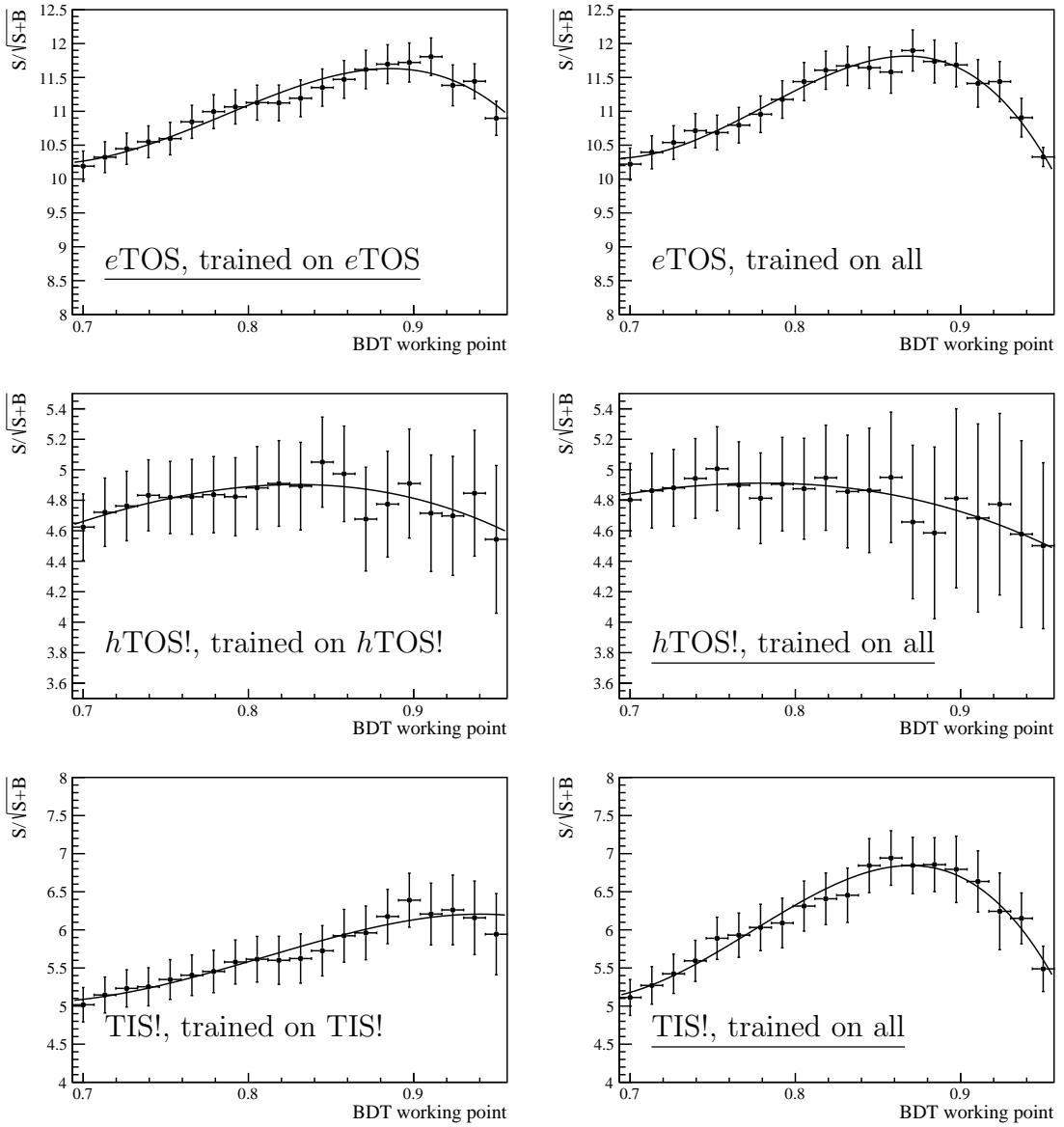


Figure 16: Expected significance as a function of the BDT working point for the $B^+ \rightarrow K^+ e^+ e^-$ decay in Run 1, for the three exclusive trigger categories. The plots on the left hand-side correspond to the BDT trained on events belonging in one trigger category only, and plots on the right hand-side corresponds to a single BDT trained on events belonging to any of the three categories. The method used as the nominal is underlined.

687 In order to keep the selection efficiency for the rare and control modes as similar as
 688 possible, the same BDT working point is used for $B^+ \rightarrow K^+ \ell^+ \ell^-$, $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$
 689 and $B^+ \rightarrow K^+ \psi(2S)(\ell^+ \ell^-)$.

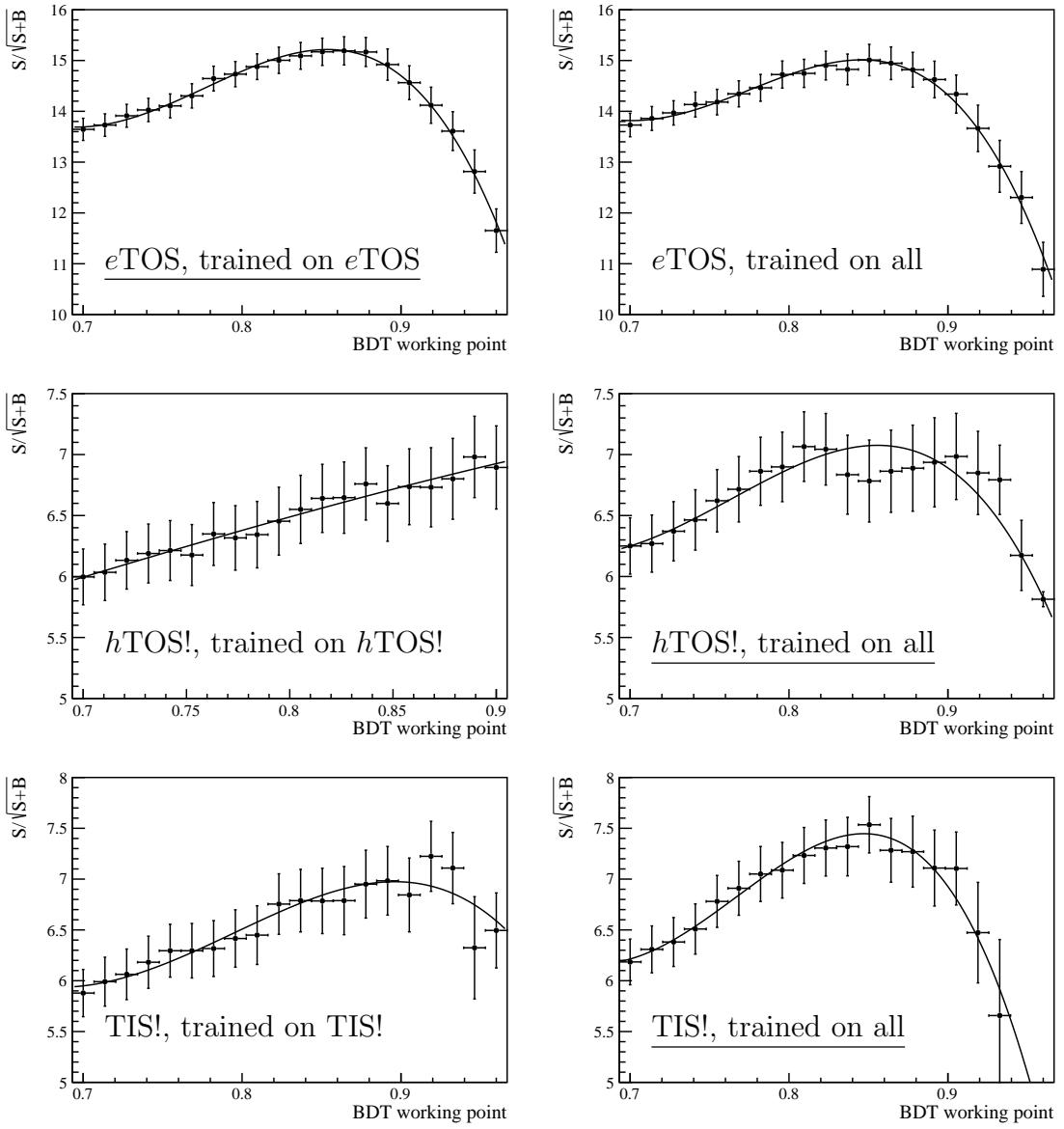


Figure 17: Expected significance as a function of the BDT working point for the $B^+ \rightarrow K^+ e^+ e^-$ decay in Run 2, for the three exclusive trigger categories. The plots on the left hand-side correspond to the BDT trained on events belonging in one trigger category only, and plots on the right hand-side corresponds to a single BDT trained on events belonging to any of the three categories. The method used as the nominal is underlined.

690 5.7 Fit window

691 The code performing the partially reconstructed background study is in *Super-*
 692 *Lenin/src/superMain2b.cc* and in *SuperLenin/src/analyseBackgrounds.cpp*

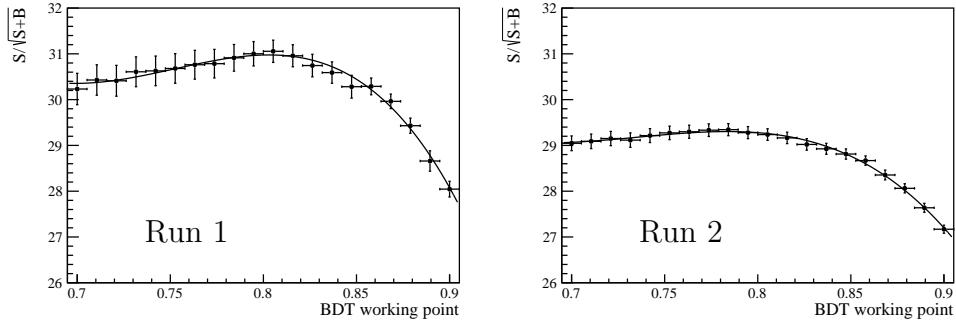


Figure 18: Expected significance as a function of the BDT working point for the $B^+ \rightarrow K^+ \mu^+ \mu^-$ decay. The left plot corresponds to Run 1 and the right plot to Run 2.

Table 12: Expected signal and combinatorial yields in the fit window for the electron and muon rare modes in different trigger and data taking conditions, for the BDT working point (wp) optimising the expected significance \mathcal{S} . In the case of the electron mode, the combinatorial level is either estimated using the $K^+ e^+ \mu^-$ sample as a proxy (comb. 1), or extrapolating the number of events from the upper mass sideband to the signal region using an exponential fit to the $K^+ e^+ e^-$ mass (comb. 2).

	wp	signal	comb. 1	comb. 2
Run 1				
$B^+ \rightarrow K^+ e^+ e^-$ eTOS	0.89	206.7 ± 2.1	115 ± 23	210 ± 110
$B^+ \rightarrow K^+ e^+ e^-$ hTOS!	0.79	70.9 ± 1.4	171 ± 34	160 ± 60
$B^+ \rightarrow K^+ e^+ e^-$ TIS!	0.87	76.2 ± 1.3	48 ± 19	60 ± 60
$B^+ \rightarrow K^+ \mu^+ \mu^-$ μ TOS	0.81	1003 ± 6		244 ± 30
Run 2				
$B^+ \rightarrow K^+ e^+ e^-$ eTOS	0.86	345.6 ± 0.9	199 ± 29	170 ± 50
$B^+ \rightarrow K^+ e^+ e^-$ hTOS!	0.86	90.7 ± 1.8	97 ± 24	120 ± 60
$B^+ \rightarrow K^+ e^+ e^-$ TIS!	0.85	98.8 ± 1.6	86 ± 18	100 ± 40
$B^+ \rightarrow K^+ \mu^+ \mu^-$ μ TOS	0.79	886 ± 5		192 ± 26

693 All the selection requirements described so far are used to suppress mis-identified,
 694 combinatorial and cascade backgrounds. However, the partially reconstructed strange and
 695 charmed partially reconstructed backgrounds, defined in Section 5.1.1, are not suppressed
 696 by any dedicated selection. The discrimination power comes from the fact that these
 697 backgrounds peak at a lower $K^+ \ell^+ \ell^-$ mass than the signal, and a fit to the mass is able
 698 to separate them. As described in Section 7, in the case of the normalisation modes, the
 699 mass used in the fit is that given by the DecayTreeFitter [32], constraining the invariant
 700 mass of the dilepton system to the mass of the J/ψ $m(\ell^+ \ell^-) = m(J/\psi)$. This gives a

701 better $K^+\ell^+\ell^-$ mass resolution.

702 In the muon case, the mass resolution is good and placing a cut at $m(K^+\mu^+\mu^-) >$
 703 5.18 GeV suppresses all the partially reconstructed backgrounds, while keeping more than
 704 99% of the signal. Indeed, partially reconstructed backgrounds are expected to peak below
 705 $m(B) - m(\pi) \approx 5.14$ GeV. This strategy is used in several $b \rightarrow s\mu^+\mu^-$ analyses [7–9].

706 In the electron case, bremsstrahlung results in the signal mass peak having a long
 707 tail to the left and it is not possible to place a tight cut on $m(K^+e^+e^-)$ to suppress all
 708 the partially reconstructed backgrounds while keeping most of the signal. Instead, it is
 709 necessary to include these backgrounds in the fit.

710 The partially reconstructed backgrounds to the $B^+ \rightarrow K^+J/\psi(e^+e^-)$ control mode are
 711 studied using inclusive simulation samples of $B^+ \rightarrow J/\psi X$, $B^0 \rightarrow J/\psi X$, $B_s \rightarrow J/\psi X$
 712 and $\Lambda_b \rightarrow J/\psi X$ offline-selected events. The J/ψ -constrained mass shape from these
 713 four samples is shown in Fig. 19. It can be seen that at high mass, the most significant
 714 components belong to the strange partially reconstructed background category, whereas
 715 at low mass, charmed components dominate. Whereas the strange components all have an
 716 equivalent in the rare mode (substituting the J/ψ for a di-electron e^+e^-), the charmed
 717 components have no such equivalent. A small fraction of mis-identified cascade backgrounds
 718 is visible in cyan in Fig. 19. In particular, a flat component in the $B_s \rightarrow J/\psi X$ plot,
 719 which correspond to $B_s \rightarrow J/\psi\rho^0(\pi^+\pi^-)$ or $B_s \rightarrow J/\psi\omega^0(\pi^+\pi^-)$, with one of the pions
 720 misidentified as a kaon. In the $B^+ \rightarrow J/\psi X$ sample, the $B^+ \rightarrow J/\psi(\rightarrow e^+e^-\gamma)K^+$ and
 721 $B^+ \rightarrow J/\psi(\rightarrow e^+e^-\gamma)\pi^+$ are removed and are taken into account in the fit as the signal
 722 and peaking mis-ID background components respectively, as explained in Section 7. For
 723 the fit, the proportion of backgrounds from B^+ , B^0 and B_s are set using the branching
 724 fractions times fragmentation fractions from [21] and [38].

725 The partially reconstructed backgrounds to the $B^+ \rightarrow K^+e^+e^-$ rare mode are studied
 726 using samples of offline-selected events from simulated $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)e^+e^-$ (also
 727 used as a proxy for $B^+ \rightarrow K^{*+}(\rightarrow K^+\pi^0)e^+e^-$), $B^+ \rightarrow K_1^+(\rightarrow K^+\pi^+\pi^-)e^+e^-$ and
 728 $B^+ \rightarrow K_2^+(\rightarrow K^+\pi^+\pi^-)e^+e^-$ decays. Simulated $B^+ \rightarrow K^+J/\psi(e^+e^-)$ decays are also
 729 used since, due to bremsstrahlung, these events can leak into the signal q^2 region and
 730 peak at low $m(K^+e^+e^-)$ mass. In addition, an inclusive sample of $B \rightarrow J/\psi X$ decays
 731 is used to extract the contribution from B decays involving charm resonances other
 732 than $B^+ \rightarrow K^+J/\psi(e^+e^-)$, such as $B \rightarrow H_c(\rightarrow J/\psi X)K^+$ or $B \rightarrow J/\psi H_s(\rightarrow KX)$,
 733 as well as to extract a proxy for the shape of $B^0 \rightarrow K_2^{*0}(\rightarrow K^+\pi^-)e^+e^-$ decays. The
 734 mass distributions for all of these background samples are shown in Fig. 20. Although,
 735 at high mass, the partially reconstructed background is completely dominated by the
 736 $B \rightarrow K^*(\rightarrow K^+\pi)e^+e^-$ component, the leakage from $B^+ \rightarrow K^+J/\psi(e^+e^-)$ is large
 737 in the region with $m(K^+e^+e^-) \lesssim 4.85$ GeV. The $B^+ \rightarrow K_1^+(\rightarrow K^+\pi^+\pi^-)e^+e^-$ and
 738 $B^0 \rightarrow K_2^{*0}(\rightarrow K^+\pi^-)e^+e^-$ components also start to be significant at low mass, which
 739 is potentially problematic in the fit, as the branching fractions for these decays are not
 740 known¹. Fig. 20 also demonstrates that charmed partially reconstructed backgrounds are

¹For the plot, they have been set to

$$\mathcal{B}(B \rightarrow K^{**}e^+e^-) = \mathcal{B}(B^0 \rightarrow K^{*0}e^+e^-) \cdot \mathcal{B}(B \rightarrow K^{**}J/\psi) / \mathcal{B}(B^+ \rightarrow K^{*0}J/\psi)$$

741 irrelevant for the rare mode. The decay $B_s^0 \rightarrow K^{*0}e^+e^-$ will be suppressed by a factor
 742 ~ 100 (~ 4 from the production and ~ 25 from the CKM suppression) compared to
 743 $B^0 \rightarrow K^{*0}e^+e^-$ decays and is neglected.

744 The use of $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)e^+e^-$ decays as a proxy for $B^+ \rightarrow K^{*+}(\rightarrow K^+\pi^0)e^+e^-$
 745 decays is justified by the distributions shown in Fig. 21. The phase space simulation of
 746 $B^+ \rightarrow K^{*+}(\rightarrow K^+\pi^0)e^+e^-$ decays is weighted to have the same angular and q^2 distribution
 747 at generator level as the $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)e^+e^-$ decay. The latter is simulated using
 748 a physics model with the Ball-Zwicky form factors, rather than just phase space. With
 749 this weighting to make the physics models the same, the distributions are in excellent
 750 agreement.

751 Considering Fig. 19 and Fig. 20, the lower mass cut is chosen to be ~ 300 GeV below
 752 the nominal B^+ mass, that is $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-) > 4.88$ GeV for the normalisation mode
 753 $B^+ \rightarrow K^+J/\psi(e^+e^-)$ and $m(K^+e^+e^-) > 4.88$ GeV for the rare mode $B^+ \rightarrow K^+e^+e^-$. In
 754 the case of the normalisation mode, this cut is more than 99% efficient on signal, and
 755 removes most of the charmed partially reconstructed backgrounds. In the case of the rare
 756 mode, the cut $m(K^+e^+e^-) > 4.88$ GeV is 85% efficient on signal and eliminates most of the
 757 leakage from $B^+ \rightarrow K^+J/\psi(e^+e^-)$ and the background from $B^+ \rightarrow K_1^+(\rightarrow K^+\pi^+\pi^-)e^+e^-$
 758 events.

759 The mass windows used for all modes are summarised in Tab. 3.

760 5.8 Multiple candidates

761 After the whole selection, the fraction of events containing several candidates is at the
 762 sub-permil level for all modes. The fractions of events having multiple candidates for the
 763 $B^+ \rightarrow K^+J/\psi$ mode in both simulation and data are reported in Tab. 13. When multiple
 764 candidates are found for the same event, only the one with the highest BDT output is
 765 retained.

Table 13: Fraction of events for which more than one $B^+ \rightarrow K^+J/\psi$ candidate are reconstructed, when the full selection chain is applied.

	simulation				data	
	rare ee	rare $\mu\mu$	J/ψ ee	J/ψ $\mu\mu$	J/ψ ee	J/ψ $\mu\mu$
Run 1	0.30%	0.01%	0.20%	0.04%	0.10%	0.01%
Run 2	0.47%	0.03%	0.32%	0.03%	0.17%	0.03%

where $K^{**} \equiv \{K_1, K_2^*\}$.

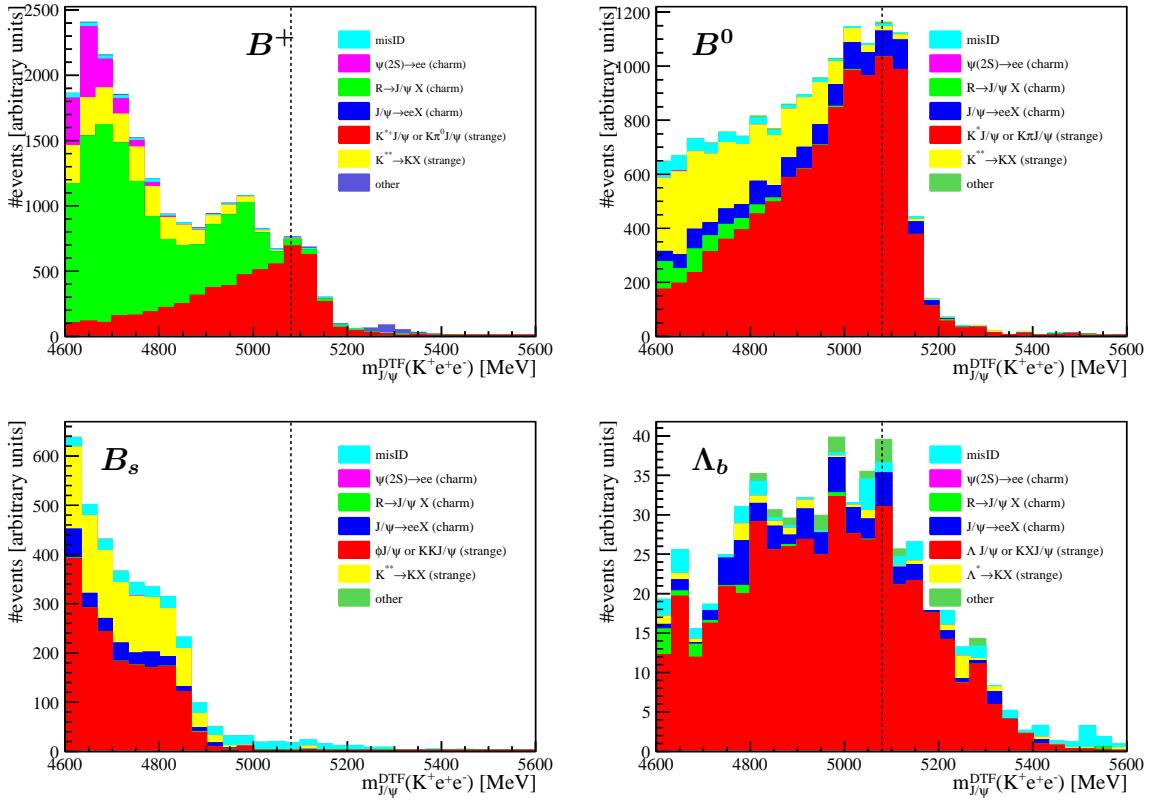


Figure 19: Mass distribution for different components in the $H_b \rightarrow J/\psi X$ inclusive simulation samples, for $H_b = B^+$ (top left), $H_b = B^0$ (top right). $H_b = B_s$ (bottom left) and $H_b = \Lambda_b$ (bottom right). The normalisation mode offline preselection, minus the PID cuts, is applied to all samples. The dark blue and green distributions are the charm partially reconstructed background components. The latter component is, at high mass, most prominent in the case of the B^+ , due to decays of the type $B^+ \rightarrow K^+ \chi_{c1}(J/\psi \gamma)$, and $B^+ \rightarrow K^+ \chi_{c2}(J/\psi \gamma)$. The lower mass cut (i.e. lower limit of the fit window) used in the present analysis is shown with a dotted black line. The mis-ID component, in light blue, is very much reduced thank to the PID requirements.

766 6 Efficiencies and Corrections to the Simulation

767 6.1 Efficiency and branching fraction computations

768 The total efficiency of the selection summarised in Section 5.2 can be factorised into several
769 terms:

$$\varepsilon^{\text{tot}} = \varepsilon^{\text{geom}} \cdot \varepsilon^{\text{rec,strip}} \cdot \varepsilon^{\text{presel}} \cdot \varepsilon^{\text{PID}} \cdot \varepsilon^{\text{trig}} \cdot \varepsilon^{\text{BDT}} \cdot \varepsilon^{\text{fit range}} \quad (13)$$

770 where:

- 771 • $\varepsilon^{\text{geom}}$ is the efficiency of the polar angle of all tracks being within 10 mrad and
772 400 mrad (DecProdCut);

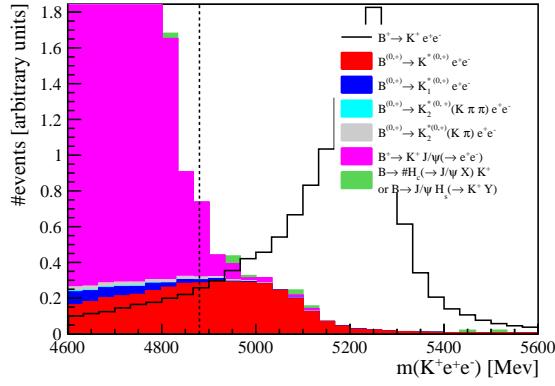


Figure 20: Mass distribution for partially reconstructed backgrounds to the rare $K^+e^+e^-$ mode, and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ leakage. The offline preselection, including the cut $1.1 < q^2 < 6.0 \text{ GeV}^2$, is applied to all samples. The lower mass cut (i.e. lower limit of the fit window) used in the present analysis is shown with a dotted black line.

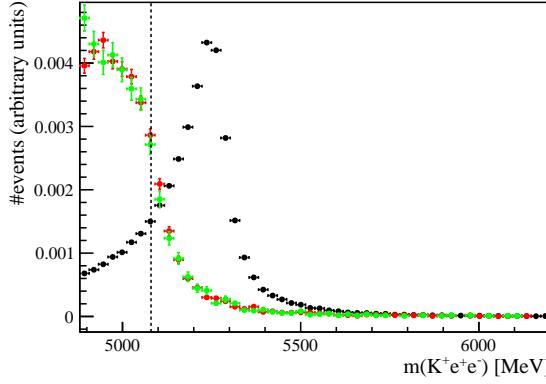


Figure 21: The $K^+e^+e^-$ mass distribution for (red) $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)e^+e^-$ decays, (green) $B^+ \rightarrow K^{*+}(\rightarrow K^+\pi^0)e^+e^-$ decays and (black) $B^+ \rightarrow K^+e^+e^-$ decays. The $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)e^+e^-$ decays are an excellent proxy for the distribution of $B^+ \rightarrow K^{*+}(\rightarrow K^+\pi^0)e^+e^-$ decays. No corrections are applied to the simulation in this comparison.

- 773 • $\varepsilon^{\text{rec,strip}}$ is the efficiency of the tracking, vertex reconstruction and stripping cuts
774 (omitting the stripping PID cuts), and the global event cut on nSPDHits;
- 775 • $\varepsilon^{\text{presel}}$ is the efficiency of the offline preselection, including the q^2 cut that is used to
776 separate the signal and normalisation modes (see Tab. 3), but omitting the offline
777 PID cuts;
- 778 • ε^{PID} is the efficiency of all PID cuts;

- $\varepsilon^{\text{trig}}$ is the efficiency of the L0 and HLT trigger requirements;
- ε^{BDT} is the efficiency of the BDT selection;
- $\varepsilon^{\text{fit range}}$ is the efficiency of the requirement on the mass being within the fit range (see Tab. 3).

The total efficiency ε^{tot} is computed using simulated samples which are corrected using control channels selected from the data. These corrections result in three types of weights which are computed on an event-by-event basis. Firstly, the PID distributions are mis-modelled in the simulation. Therefore, the PID requirements are not applied to simulated events (this includes removing the PID requirements from the stripping and `StandardParticles`). Instead, events are assigned a weight $w^{\text{PID}} = \varepsilon_{K^+}^{\text{PID}} \cdot \varepsilon_{\ell^-}^{\text{PID}} \cdot \varepsilon_{\ell^+}^{\text{PID}}$, where $\varepsilon_{K^+}^{\text{PID}}$ and $\varepsilon_{\ell^\pm}^{\text{PID}}$ are PID efficiencies extracted from control samples. The independence of the electron weights is established in Section 6.4.3. Secondly, the trigger efficiencies have to be corrected. This is done using a weight w^{trig} corresponding to a data/simulation ratio of trigger efficiencies. Thirdly, after applying the trigger and PID corrections, the kinematics of simulated B^+ have to be corrected to match the data, resulting in a set of kinematic weights labelled w^{kin} . There are actually two kinds of kinematic weights. The first is labelled $w_{\text{gen}}^{\text{kin}}$ and corrects the momentum and rapidity distributions of generator-level B^+ . The second is labelled $w_{\text{rec}}^{\text{kin}}$ and correct for the momentum and rapidity distributions, as well as variables related to the vertex quality which are not defined at generator level.

The expression for the total efficiency is then,

$$\varepsilon^{\text{tot}} = \underbrace{\frac{\sum_{\text{rec}} w_{\text{gen}}^{\text{kin}}}{\sum_{\text{gen}} w_{\text{gen}}^{\text{kin}}}}_{\varepsilon^{\text{geom}} \cdot \varepsilon^{\text{rec,strip}}} \cdot \underbrace{\frac{\sum_{\text{sel}} w_{\text{rec}}^{\text{kin}} \cdot w^{\text{PID}} \cdot w^{\text{trig}}}{\sum_{\text{rec}} w_{\text{rec}}^{\text{kin}}}}_{\varepsilon^{\text{presel}} \cdot \varepsilon^{\text{PID}} \cdot \varepsilon^{\text{trig}} \cdot \varepsilon^{\text{BDT}} \cdot \varepsilon^{\text{fit range}}}, \quad (14)$$

where:

- \sum_{gen} indicates a sum over generator level events;
- \sum_{rec} indicates a sum over reconstructed events to which the stripping requirements have been applied, including the global event cut on nSPDHits, but no PID cut;
- \sum_{sel} indicates a sum over fully selected events, to which the full selection chain — except the PID requirements — has been applied;

For the rare modes, the branching fraction of interest is defined in a range of the “true” di-lepton invariant mass, $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$. To extract it, the efficiency-corrected yield of $B^+ \rightarrow K^+ \ell^+ \ell^-$ events in $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$, $\mathcal{N}_{\text{in}}(B^+ \rightarrow K^+ \ell^+ \ell^-)$ must be computed, that is the total yield of $B^+ \rightarrow K^+ \ell^+ \ell^-$ events before any detector effect. This efficiency-corrected yield is related to the yield of fully selected rare decay events in the

810 range $1.1 < q^2 < 6.0 \text{ GeV}^2$, $N^{\text{sel}}(B^+ \rightarrow K^+\ell^+\ell^-)$, extracted with a fit to the rare mode,
 811 through the following equation:

$$N^{\text{sel}}(B^+ \rightarrow K^+\ell^+\ell^-) = \mathcal{N}_{\text{all}}(B^+ \rightarrow K^+\ell^+\ell^-) \cdot \varepsilon^{\text{tot}}(B^+ \rightarrow K^+\ell^+\ell^-) \quad (15)$$

$$= \mathcal{N}_{\text{in}}(B^+ \rightarrow K^+\ell^+\ell^-) \cdot \frac{1}{f^{q^2}} \cdot \varepsilon^{\text{tot}}(B^+ \rightarrow K^+\ell^+\ell^-), \quad (16)$$

812 where \mathcal{N}_{all} is the total number of events in the whole q^2 range, and f^{q^2} is the fraction of
 813 $B^+ \rightarrow K^+\ell^+\ell^-$ events in the range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$. In practice, f^{q^2} is computed
 814 from simulated events at generator level:

$$f^{q^2} = \frac{\sum_{\text{gen}, 1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2} w_{\text{gen}}^{\text{kin}}}{\sum_{\text{gen, all } q^2} w_{\text{gen}}^{\text{kin}}}. \quad (17)$$

815 Note that while this ratio is dependent on the generated q^2 distribution, and hence the
 816 form factors assumed in the model, the selection efficiency has a component that cancels
 817 this dependence, at least in the limit of perfect q^2 resolution.

818 Hence, the ratio of branching fractions between the rare mode, in the range
 819 $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$, and the $B^+ \rightarrow K^+J/\psi$ control mode is expressed as follows:

$$\begin{aligned} \frac{\mathcal{B}(B^+ \rightarrow K^+\ell^+\ell^-)_{\text{in}}}{\mathcal{B}(B^+ \rightarrow K^+J/\psi)} &= \frac{\mathcal{N}_{\text{in}}(B^+ \rightarrow K^+\ell^+\ell^-)}{\mathcal{N}(B^+ \rightarrow K^+J/\psi)} \\ &= \frac{\mathcal{N}_{\text{all}}(B^+ \rightarrow K^+\ell^+\ell^-)}{\mathcal{N}(B^+ \rightarrow K^+J/\psi)} \cdot f^{q^2} \\ &= \frac{N^{\text{sel}}(B^+ \rightarrow K^+\ell^+\ell^-)}{N^{\text{sel}}(B^+ \rightarrow K^+J/\psi)} \cdot \frac{\varepsilon^{\text{tot}}(B^+ \rightarrow K^+J/\psi)}{\varepsilon^{\text{tot}}(B^+ \rightarrow K^+\ell^+\ell^-)} \cdot f^{q^2} \end{aligned}$$

820 where the subscript *in* means in the range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$ and *all* means at all q^2 .

821 The following sections give more detail on how each efficiency is computed and each
 822 correction weight is extracted.

823 6.2 Truth-matching, ghosts, and definition of q_{true}^2

824 Reconstructed signal and control simulated events are truth-matched using
 825 `TupleToolMCBackgroundInfo`. To extract the trigger and kinematic weights discussed
 826 in Section 6.6 and Section 6.8, the background categories 0, 10 and 50 are selected. A
 827 significant fraction of signal events from the electron modes end up in background category
 828 50, corresponding to partially reconstructed events. This is due to final state radiation
 829 (FSR).

830 6.2.1 Handling of ghosts in the simulation

831 For a small fraction of simulated signal events, the truth-matching algorithm fails and
 832 these events end up being mis-classified in background category 60, which is supposed to

833 correspond to fake tracks, also referred to as ghosts. The fraction of ghosts is higher for
 834 the electron mode than for the muon modes, as summarised in Tab. 14. For simulated
 835 events to which only the stripping selection is applied, background category 60 does not
 836 only contain mis-classified signal events, but also a significant fraction of combinatorial
 837 events. These combinatorial events must be suppressed to avoid biasing the efficiencies
 838 and this is done in three ways. First, only events with one track classified as a ghost, and
 839 the two others correctly matched to a lepton or a kaon, are retained. Second, applying the
 840 $\text{prob}_{\text{ghost}}$ cut (see in Tab. 2) also helps reduce the fraction of combinatorial events. Third,
 841 applying the BDT requirement eliminates the majority of remaining combinatorial events.
 842 The $m_{\text{DTF}}^{J/\psi}$ distribution is shown in Fig. 22 for signal events classified as ghosts, before and
 843 after applying the full selection, clearly showing a high fraction of combinatorial events
 844 before the selection is applied. This means that $\varepsilon^{\text{rec,strip}}$ and $\varepsilon^{\text{presel}}$ cannot be computed
 845 taking signal events mis-classified as ghosts into account, only the total efficiency ε^{tot} ,
 846 including the BDT cut, can be. Hence, background category 60 is excluded to extract all
 847 the correction weights, but is included to compute the final total efficiency.

848 To compute the kinematic weights and to study the q^2 resolution (see Section 6.9), the
 849 true momenta of the mother B and the daughter particles must be accessed. For events
 850 classified as ghosts, they are not provided automatically by the truth-matching algorithm,
 851 since this algorithm fails for these candidates. They are instead recovered by matching
 852 each event in the reconstructed dataset to the corresponding event in the generated dataset
 853 using the event number.

Table 14: Relative fractions of simulated events being classified as ghosts in fully selected $B^+ \rightarrow K^+\ell^+\ell^-$ and $B^+ \rightarrow K^+J/\psi(\ell^+\ell^-)$ events, not counting multiple candidates.

	Kee	$KJ/\psi(ee)$	$K\mu\mu$	$KJ/\psi(\mu\mu)$
Run 1	4.2%	3.7%	1.5%	1.6%
Run 2	5.1%	4.2%	2.0%	1.9%

854 6.2.2 Definition of q_{true}^2

855 When computing the branching fraction for the rare mode $\mathcal{B}(B^+ \rightarrow K^+\ell^+\ell^-)_{in}$, the range
 856 $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$ corresponds to a range in the true q^2 , q_{true}^2 , that is prior to any
 857 resolution effect, and also prior to FSR. As electrons are relatively light, they can lose
 858 a significant amount of energy via FSR. For kaons and muons, the effect is negligible,
 859 at the level of $\sim 0.1\%$ [39]. Hence, the true K^+ and B^+ momenta are used to compute
 860 a proxy for q_{true}^2 , denoted $\tilde{q}_{\text{true}}^2$: $\tilde{q}_{\text{true}}^2 = |p_{B^+}^{\text{true}} - p_{K^+}^{\text{true}}|^2$. The simulation of FSR, which
 861 is handled by PHOTOS, has been tested and validated in Ref. [39]. The distributions
 862 of the reconstructed $q^2 = |p_{e^+} + p_{e^-}|^2$, as well as pre-FSR true $\tilde{q}_{\text{true}}^2 = |p_{B^+}^{\text{true}} - p_{K^+}^{\text{true}}|^2$,
 863 and the proxy for the post-FSR true $q_{\text{true,FSR}}^2 = |p_{e^+}^{\text{true}} + p_{e^-}^{\text{true}}|^2$ are shown in Fig. 23 for

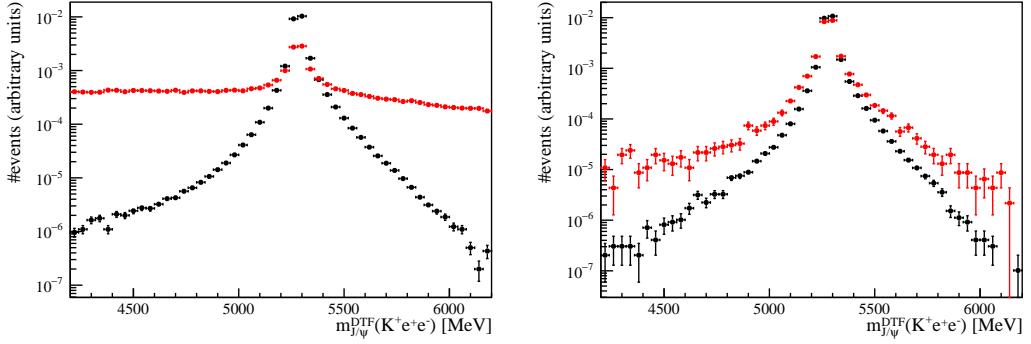


Figure 22: Distributions of $m_{\text{DTF}}^{J/\psi}$ for simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events classified as ghosts (category 60, red) and correctly truth-matched (category 0, 10 and 50, black). The plot on the left corresponds to events that satisfy the stripping requirements only, the plot on the right corresponds to events that pass the full selection chain, and that contain at most one track classified as a ghost track by the truth-matching algorithm.

simulated $B^+ \rightarrow K^+ e^+ e^-$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events. The $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ distribution post-FSR has a large tail to the left due to radiation from the electrons. The pre-FSR proxy in $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ shows a very small tail to the right, which is due to radiation coming from the kaon. It is clear from the plot that FSR coming from the kaon is negligible compared to that coming from electrons.

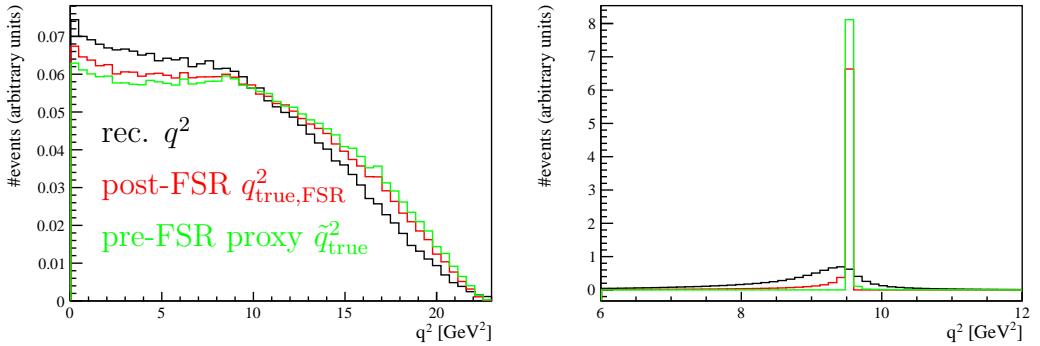


Figure 23: Distribution of the reconstructed q^2 (black), true post-FSR q^2 (red) and proxy for the true pre-FSR q^2 (green) for simulated $B^+ \rightarrow K^+ e^+ e^-$ events (left) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events (right).

869 **6.3 Fit to the Pre-Selected Control Modes**

870 The fit to $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ J/ψ -
871 constrained invariant mass is written as two classes described
872 in *superLenin/include/fitAndSplotKeeDataForTraining.h* and *super-*
873 *Lenin/include/fitAndSplotKmumuDataDTFForTraining.h*. The mains using
874 these classes to perform the actual fit to the pre-selected data are *super-*
875 *Lenin/src/superMain3.cc* and *superLenin/src/superMain3Run2.cc*.

876 A fit to the control modes $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ is performed
877 at the pre-selection stage (that is, without applying the BDT cut). This is done, firstly,
878 to obtain the yield of $B^+ \rightarrow K^+ J/\psi$ events, which is used in Section 5.6.2 to optimise
879 the BDT, and, secondly, to obtain background-subtracted data samples that are used to
880 extract the data/simulation correction weights. To statistically subtract the background,
881 the *sPlot* method is used [40].

882 The fits to $m_{\text{DTF}}^{J/\psi}$ for pre-selected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ data events in Run 1 and
883 Run 2, as well as their pulls, are shown in Fig. 24. Figure 25 shows the fits to $m_{\text{DTF}}^{J/\psi}$ for
884 pre-selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events in Run 1 and Run 2 for the three exclusive
885 trigger categories. A description of the models describing the different components in
886 these fits can be found in Section 7.

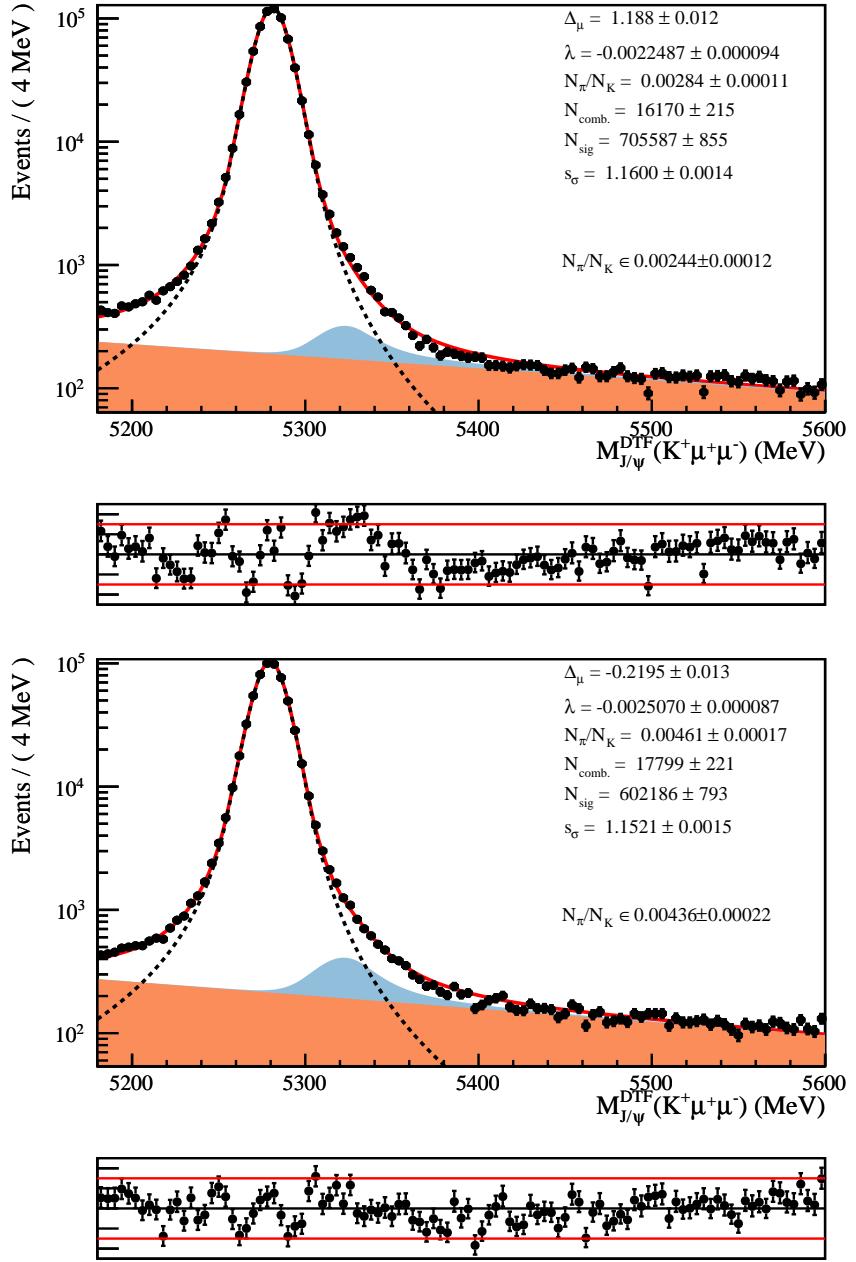


Figure 24: Fit to preselected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ in the data in Run 1 (top) and Run 2 (bottom). The combinatorial is filled in orange, the $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ in blue, the signal component is shown with a black dotted line and the total fit model with a solid red line. The values and error of the floating parameters are tabulated on the top right of the figure, and the value of the constraint of the $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ yield is indicated below that. The fit quality, while not perfect, is adequate for the purpose of extracting the control channel yield with an error that is irrelevant compared to that of the $B^+ \rightarrow K^+ e^+ e^-$ yield, which is two orders of magnitude smaller.

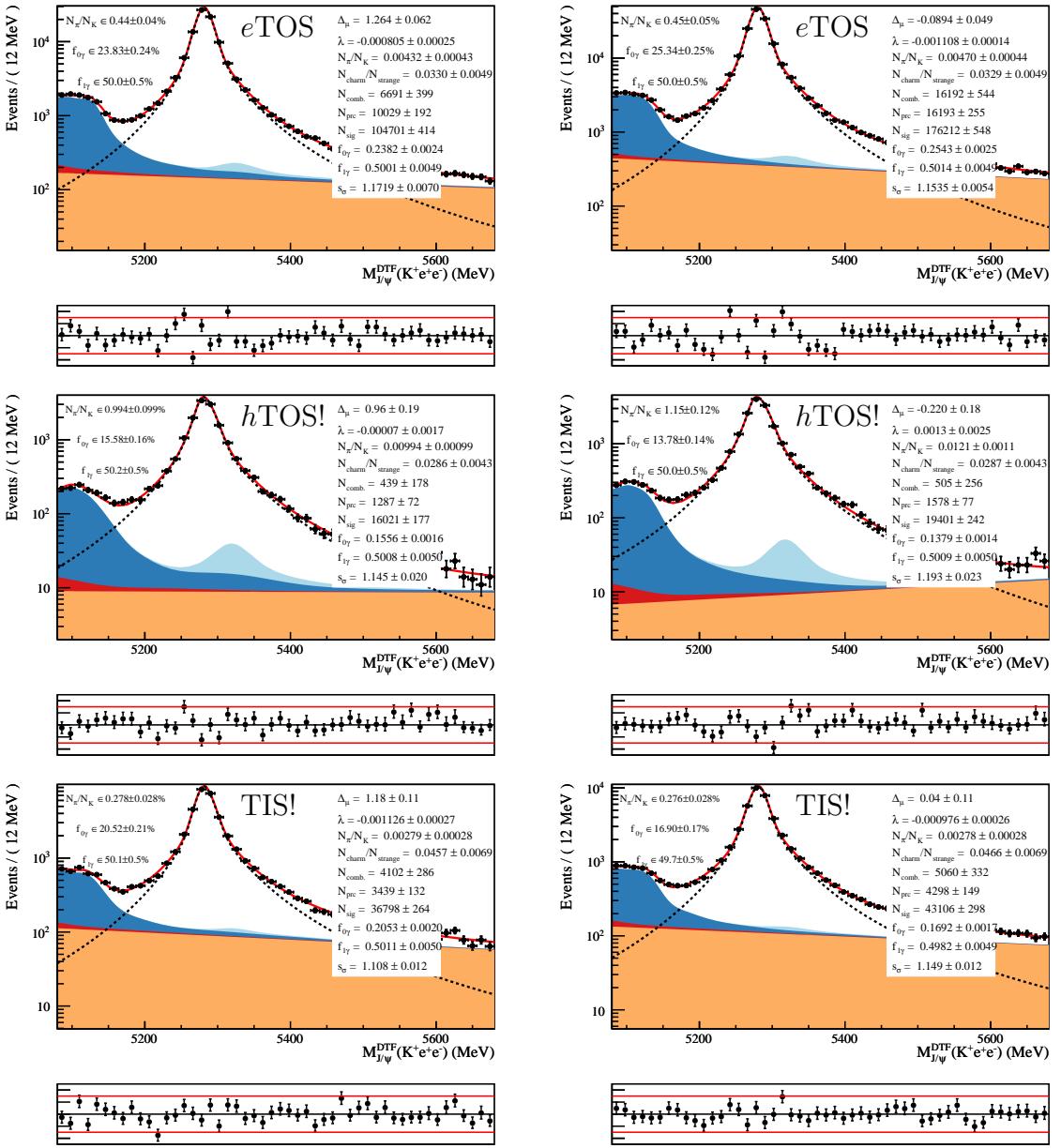


Figure 25: Fit to preselected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ in the data in Run 1 and Run 2 (right), for the three exclusive trigger categories: *eTOS* (top), *hTOS!* (middle) and *TIS!* (bottom). The combinatorial PDF is filled in orange, the $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ in light blue, the strange partially reconstructed background in dark blue and the charmed partially reconstructed background in red. The signal component is shown with a black dotted line and the total fit model with a solid red line. The values and errors of the floating parameters are tabulated on the right of the figure. The central values and standard deviations of the constraints are shown on the left.

887 **6.4 Corrections to particle identification efficiencies**

888 The performance of the RICH detector is not accurately simulated by the LHCb simulation
889 software. As a result, the efficiencies of the PID requirements summarised in Section 5.4.5
890 cannot be predicted using the simulation samples. Instead, calibration samples from the
891 data are used to compute the PID efficiencies as a function of kinematic variables. The
892 computation for the efficiencies of all particle species are presented below.

893 **6.4.1 Kaon, muon and pion (mis-)identification efficiencies**

894 The rebinning algorithm is coded in *superLenin/src/pidReweightingTools.cpp* and is
895 used in *superLenin/src/superMain2a(Run2).cc*.

896 The PID efficiencies for the kaon and muons, as well as the pion mis-identification
897 efficiencies to a kaon (which are different in the muon and electron modes), are measured
898 using data calibration samples of $D^0 \rightarrow K^+\pi^-$ decays and $J/\psi \rightarrow \mu^+\mu^-$ decays. In each
899 calibration sample, a fit to the invariant mass of the mother particle is performed to
900 extract *sWeight*'s [40]. In each kinematic bin, the efficiency ε^{PID} of a PID requirement is
901 computed using:

$$\varepsilon^{\text{PID}} = \frac{\sum_{\text{pass}} sW}{\sum_{\text{tot}} sW},$$

902 where sW is an *sWeight*, and \sum_{tot} runs over all the events in a kinematic bin, \sum_{pass} runs
903 over all the events in the bin that satisfy the PID requirement. The fits to the control
904 modes and the extraction of the efficiencies are done using the **PIDCalib** package [22],
905 making calibration histograms for each year and magnet polarity.

906 In the case of the muons and kaons from the muon modes, calibration histograms are
907 produced separately for the *isMuon* cuts and for the DLL (or probNN) cut, in order to be
908 able to study them separately. The *isMuon* efficiency is expected to be better modelled in
909 the simulation.

910 The pion to kaon mis-identification efficiency is needed for the pion from the $B^+ \rightarrow$
911 $\pi^+ J/\psi(\mu^+\mu^-)$, in order to constrain its yield in the fit to the control mode (see Section 9).
912 In this case, the efficiency of the *isMuon = 0* cut is computed from the simulation, as the
913 error due to the mis-modelling in the simulation is irrelevant compared to the uncertainty
914 on the $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ to $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ branching fraction ratio.

915 The cuts applied to the calibration samples are summarised in Tab. 15. The fiducial
916 cuts are applied to ensure that the particles are within the acceptance of the relevant
917 detector (RICH, calorimeters or muon chambers). The occupancy cut on nSPDHits is also
918 applied to the calibration samples. These cuts are the same as those applied to the signal
919 modes. As recommended in Ref. [22], to compute the efficiency of the *isMuon = 0* cut on
920 the kaon, a trigger requirement is applied, which requires some HLT lines that contain a
921 cut on *isMuon* to be TIS.

922 The tuning for the probNN variable is MC12TuneV2 for Run 1 and MC15TuneV1 for
923 Run 2. For Run 2, the BRUNEL version of all the variables is used. This means that all
924 variables, which in the PID calibration samples are at first computed online using the

925 turbo, are re-computed offline, using the BRUNEL software, to make sure they match those
 926 used for the signal data.

Table 15: List of all seven PID calibration histograms made with the `PIDCalib` package. The second column shows the PID cut for which the efficiency is computed. The third column shows the fiducial cuts applied to the sample. The term “nSPDHits cut” is short for $n\text{SPDHits} < 450$ in Run 2 and $n\text{SPDHits} < 600$ in Run 1. The term “HLT1 unbias” is a set of TIS requirements applied to a set of HLT lines in order to prevent the kaon mis-identification to muon to be biased. The last column indicates the figure showing the projections of the calibration histograms.

	PID cut	Fiducial cut	Fig.
For muon samples			
K	isMuon = 0	<code>hasRich = 1 & inMuonAcc = 1 & HLT1 unbias & nSPDHits cut</code>	26, 27
	$\text{probNN}_K > 0.2$	<code>hasRich = 1 & isMuon = 0 & inMuonAcc = 1 & nSPDHits cut</code>	28, 29
μ	isMuon == 1	<code>hasRich = 1 & inMuonAcc = 1 & nSPDHits cut</code>	30, 31
	$\text{DLL}_\mu > -3$	<code>hasRich = 1 & inMuonAcc = 1 & isMuon = 1 & nSPDHits cut</code>	32, 33
π	$\text{probNN}_K > 0.2$	<code>hasRich = 1 & isMuon = 0 & inMuonAcc = 1 & nSPDHits cut</code>	34, 35
For electron samples			
K	$\text{probNN}_K > 0.2$	<code>hasRich = 1 & nSPDHits cut</code>	36, 37
π	$\text{probNN}_K > 0.2$ $\& \text{DLL}_e < 0$	<code>hasRich = 1 & nSPDHits cut</code>	38, 39

927 All the PID efficiencies are broken down into bins of year, magnet polarity, momentum
 928 and rapidity. Efficiencies are expected to depend on the occupancy. However, they are
 929 not binned in any occupancy proxy (like nSPDHits or nTracks) because these are badly
 930 described by the simulation, and because the occupancy of the calibration samples is very
 931 similar to that of the signal samples. Kinematic and occupancy distribution comparing
 932 the rare and control modes with the PID calibration data are shown in Appendix E.

933 To find the optimal binning in momentum and rapidity, the following procedure is
 934 used, inspired from [41]:

- 935 1. An initial binning is chosen such that events are equally-distributed in ten rapidity
 936 bins and 100 momentum bins;
- 937 2. The 2D histograms are projected into two 1D histograms (one for rapidity, one for
 938 momentum);

939 3. In each of the two projections, if the efficiency values in two adjacent bins are less
940 than 2.5σ apart, the bins are merged together.

941 This method allows a compromise to be found between having enough bins to cover the
942 efficiency variation, and having enough events in each bin to be able to compute the
943 efficiency precisely.

944 The projections of all the calibration histograms are shown in Fig. 26-Fig. 39.

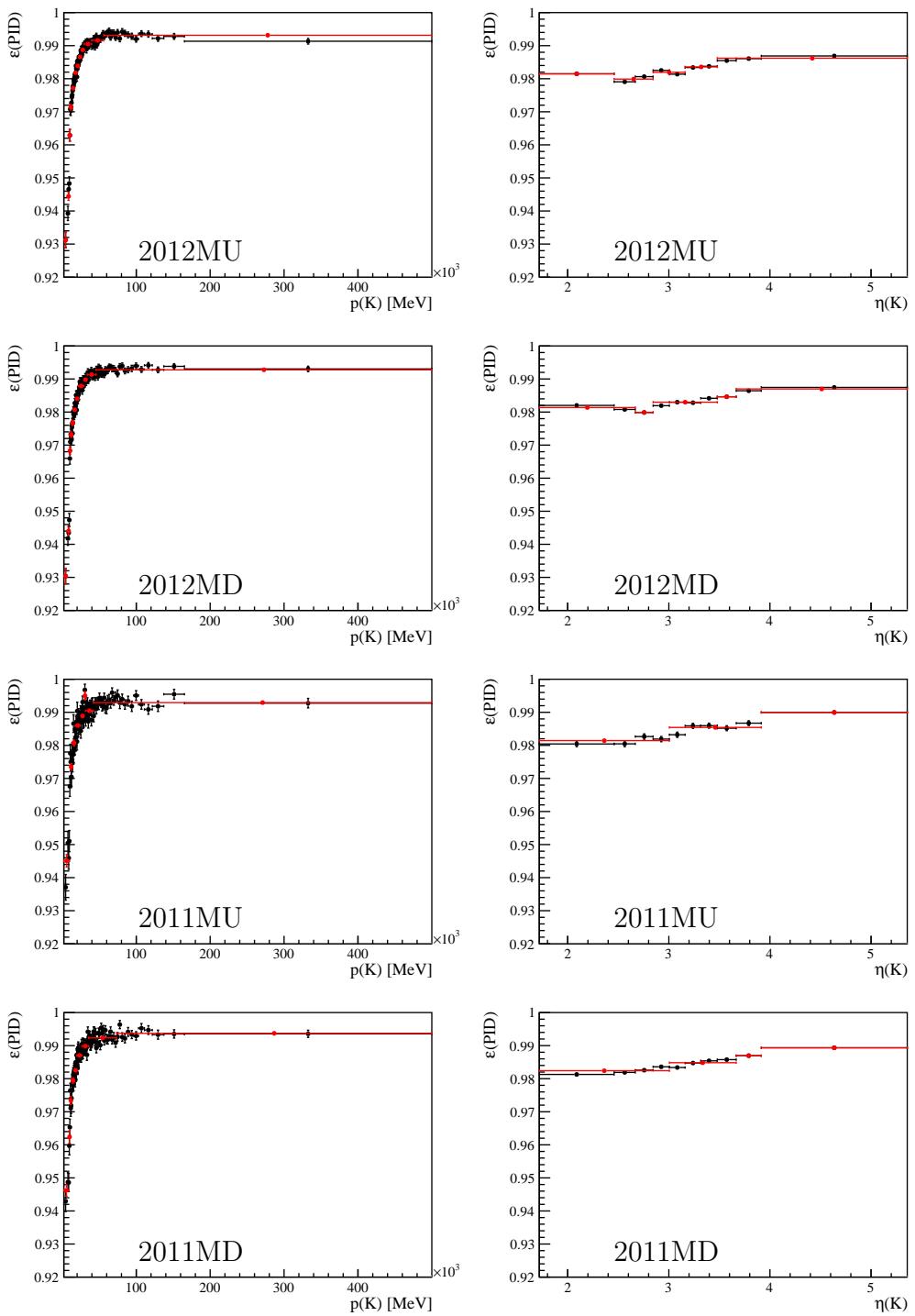


Figure 26: Particle identification calibration histograms for the cut $\text{isMuon} = 0$ for kaons in Run 1 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

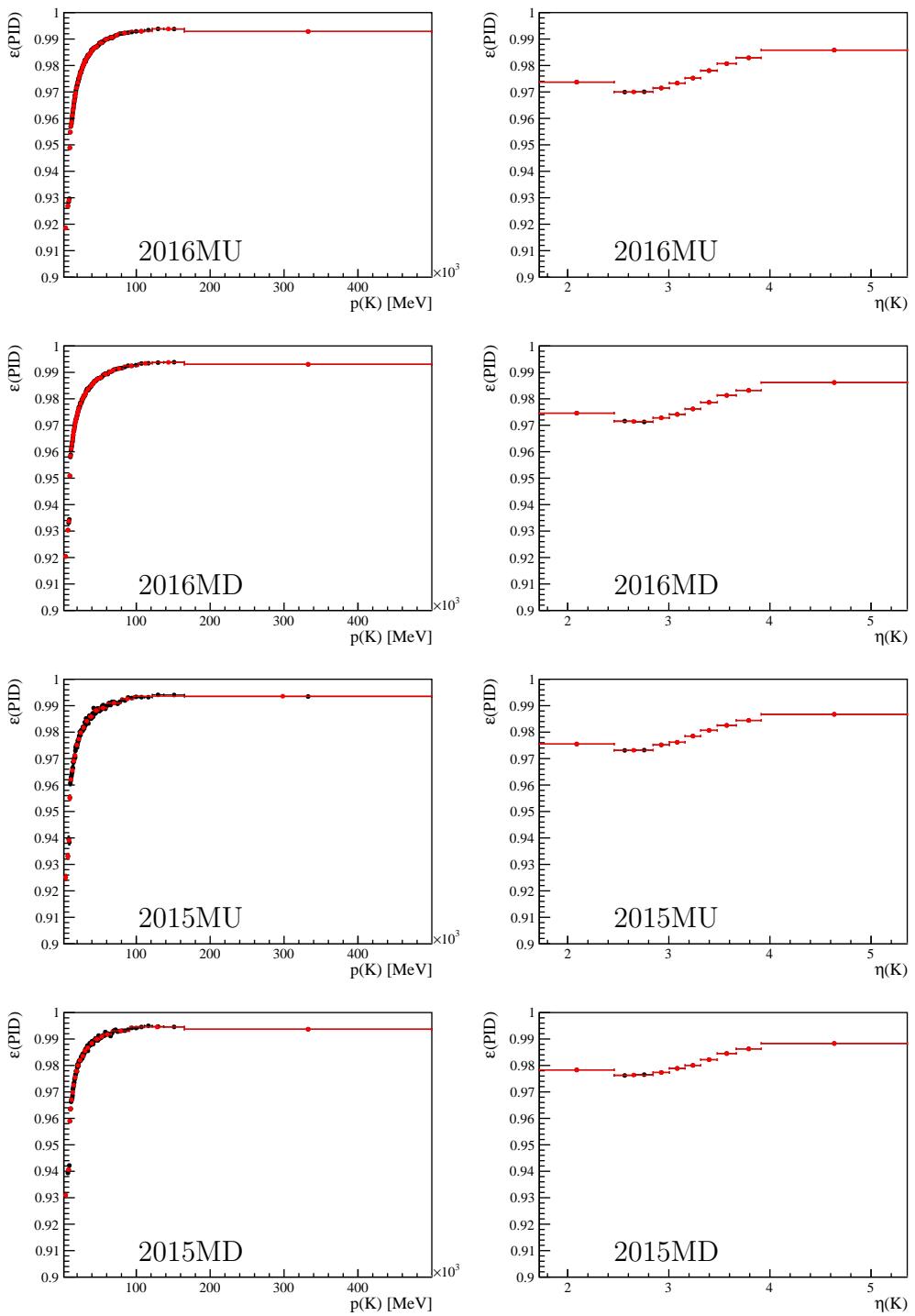


Figure 27: Particle identification calibration histograms for the cut $\text{isMuon} = 0$ for kaons in Run 2 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

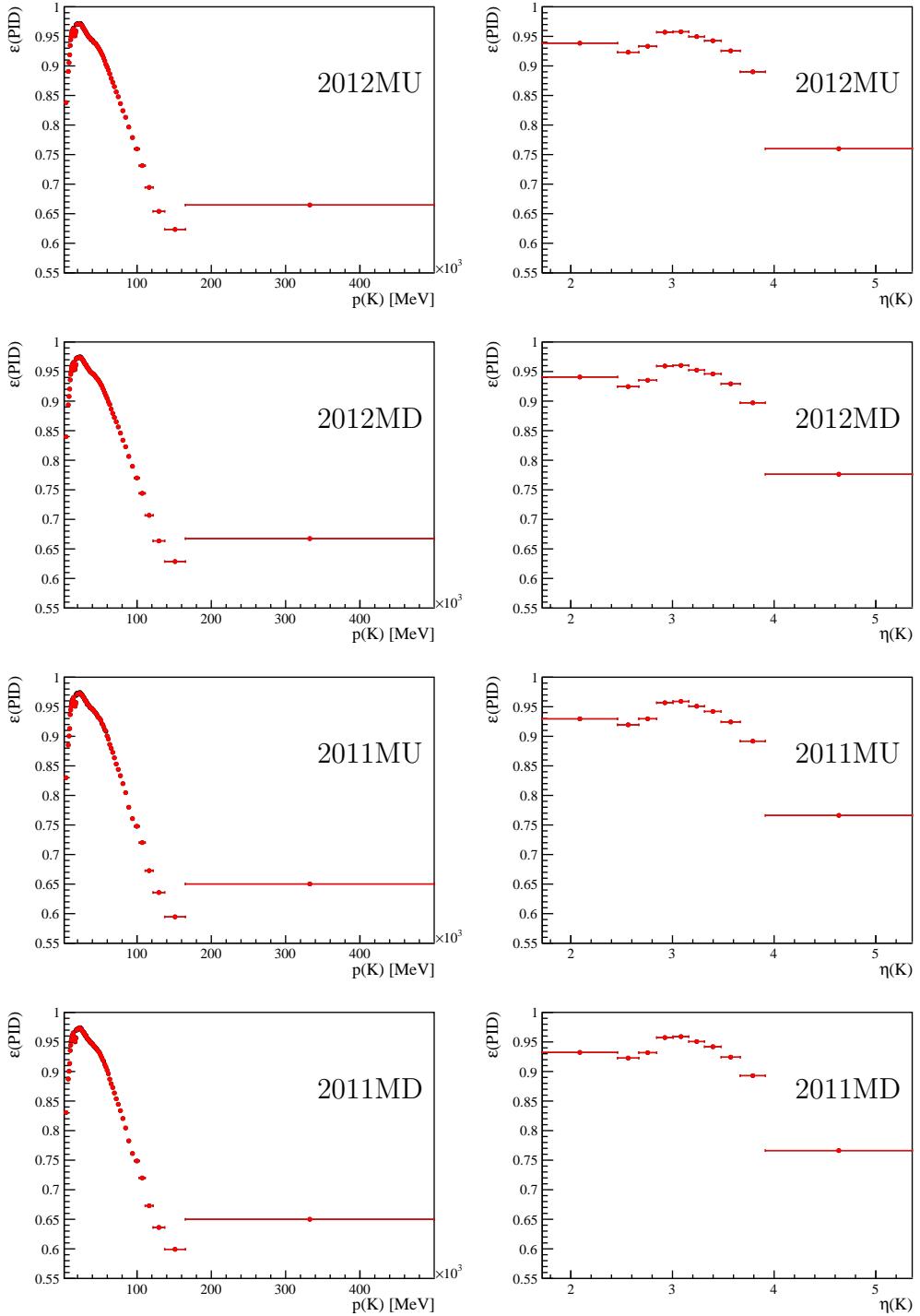


Figure 28: Particle identification calibration histograms for the cut $\text{probNN}_K > 0.2$, on top of $\text{isMuon} = 0$, for kaons in Run 1 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

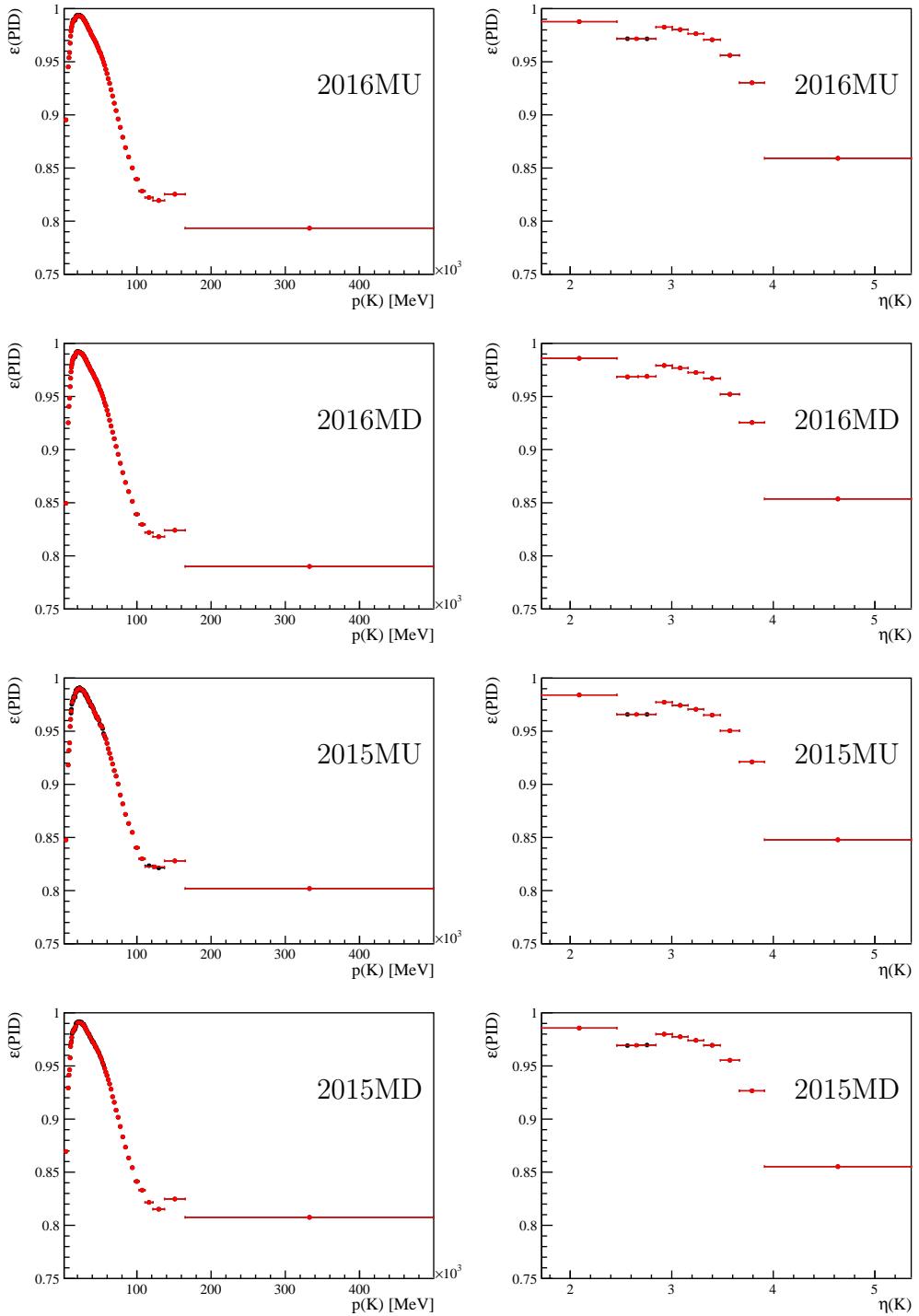


Figure 29: Particle identification calibration histograms for the cut $\text{probNN}_K > 0.2$, on top of $\text{isMuon} = 0$, for kaons in Run 2 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

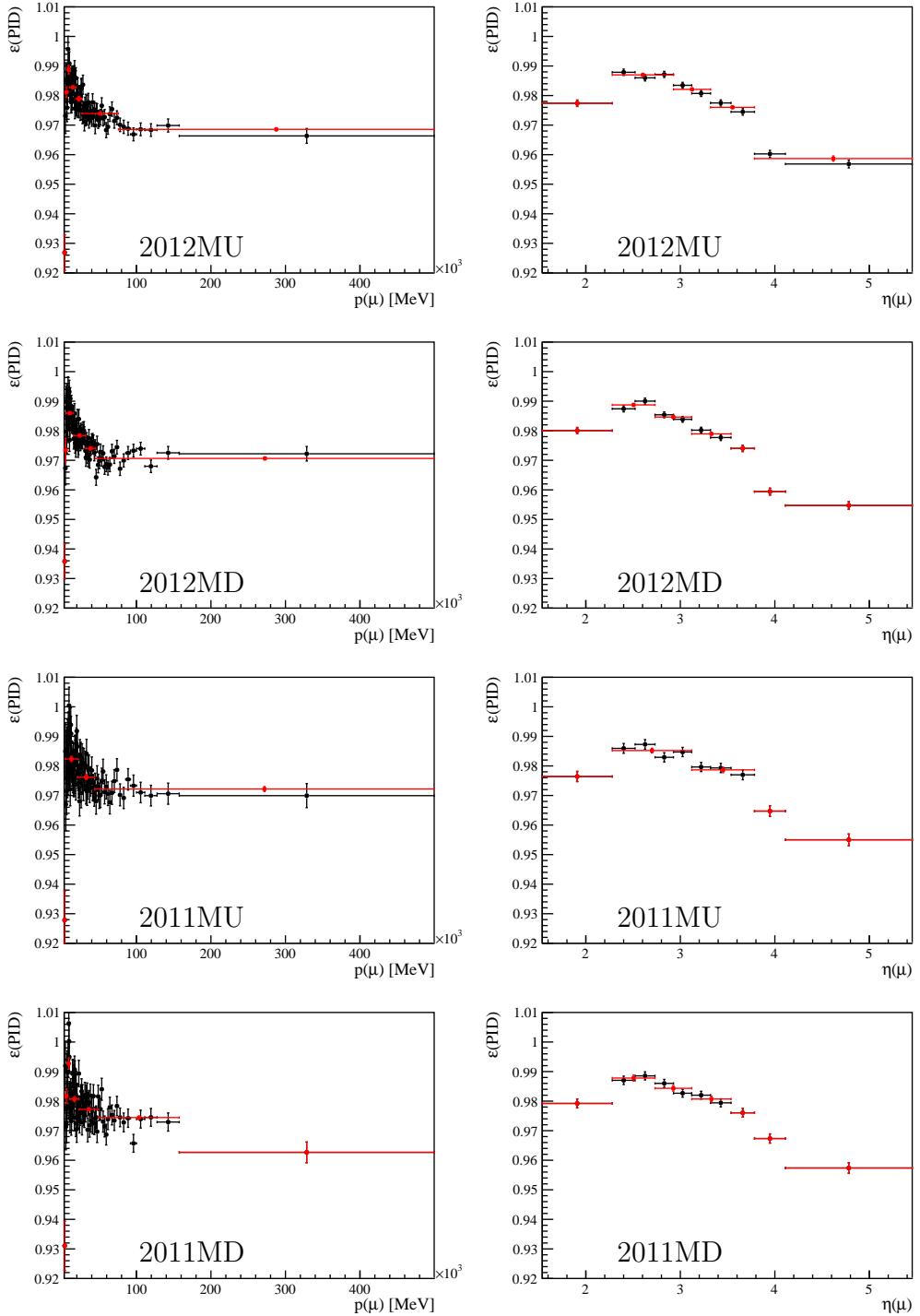


Figure 30: Particle identification calibration histograms for the cut $\text{isMuon} = 1$ for muons in Run 1 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

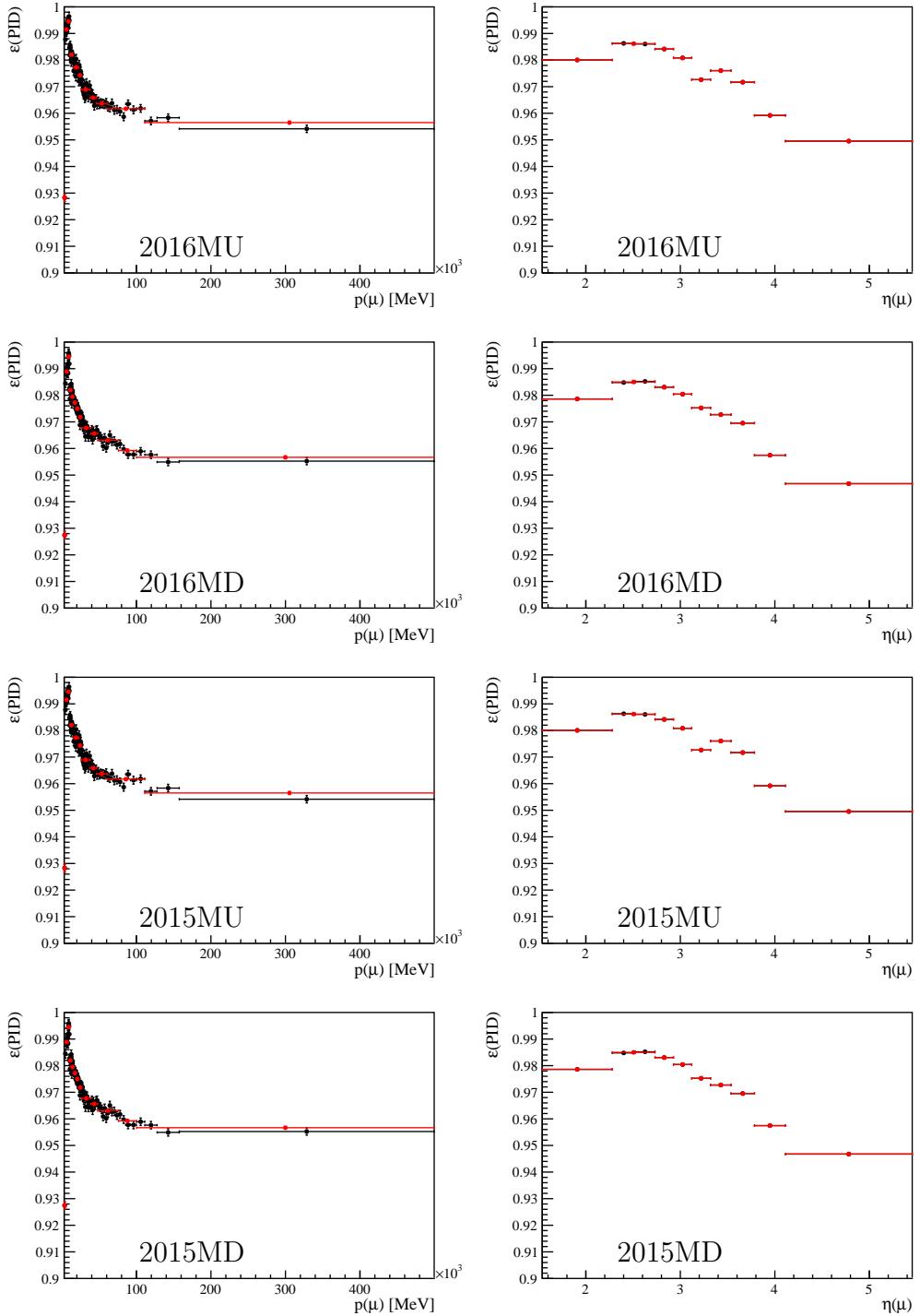


Figure 31: Particle identification calibration histograms for the cut $\text{isMuon} = 1$ for muons in Run 1 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

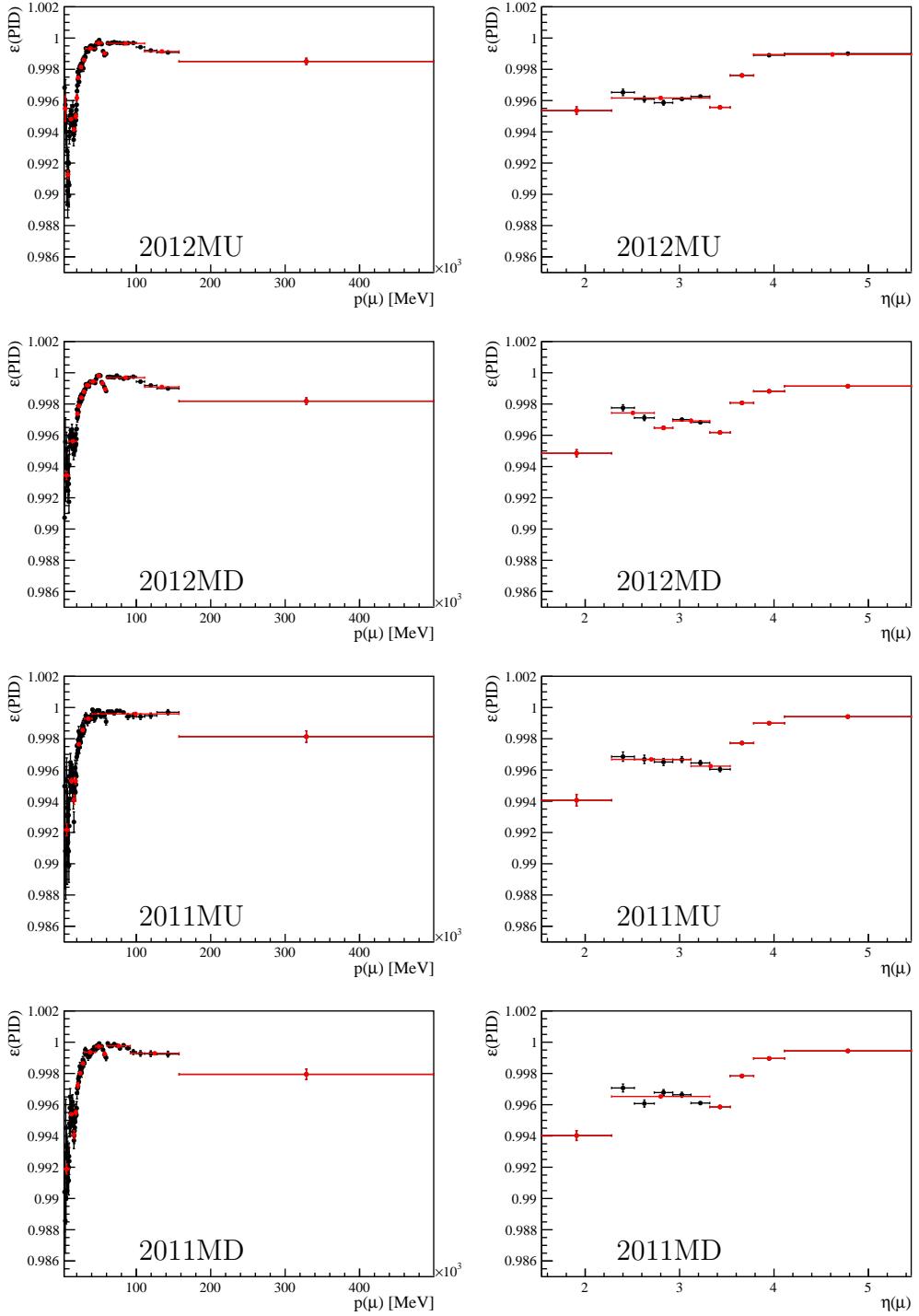


Figure 32: Particle identification calibration histograms for the cut $\text{DLL}_\mu > -3$, on top of $\text{isMuon} == 1$, for muons in Run 1 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

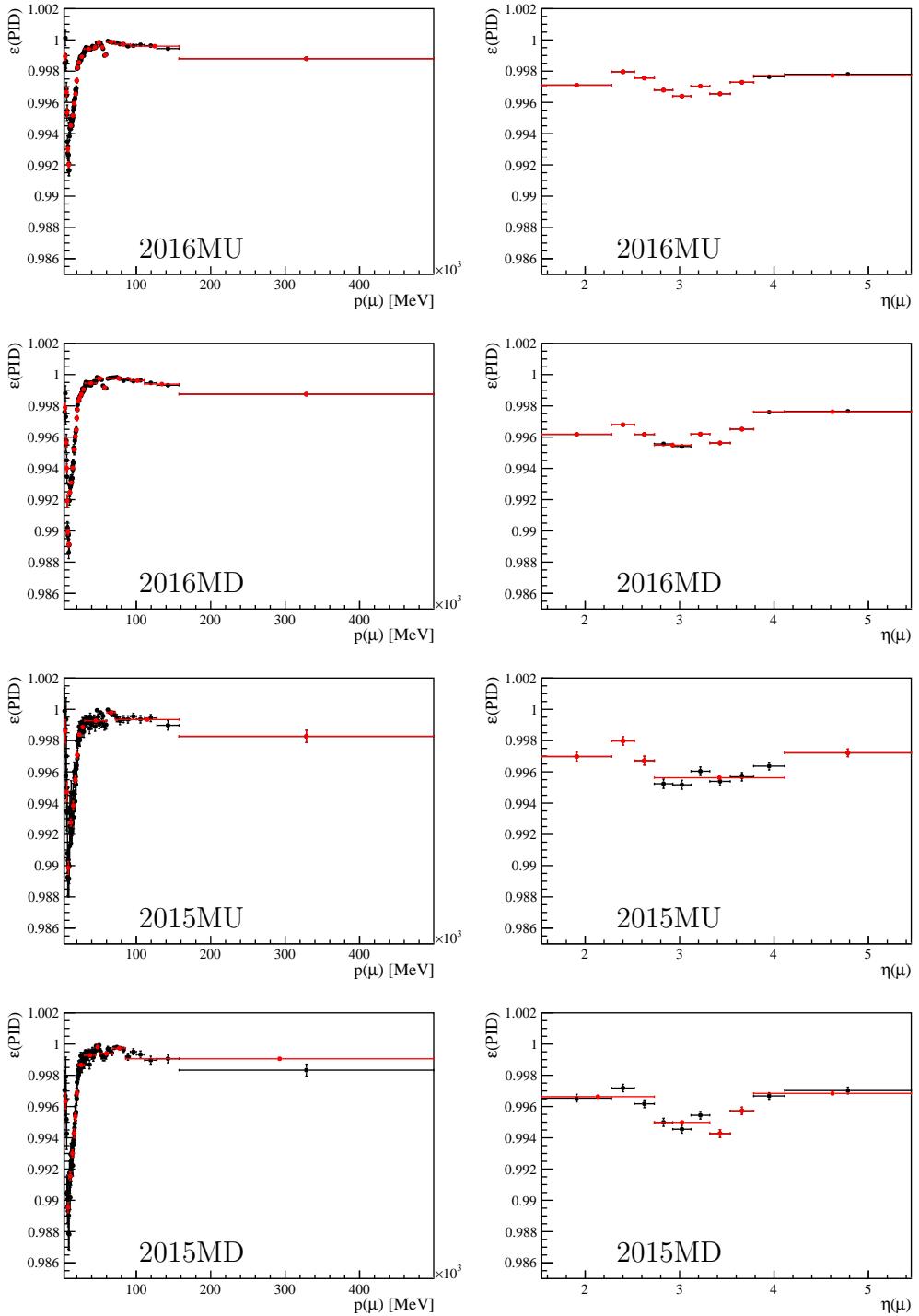


Figure 33: Particle identification calibration histograms for the cut $\text{DLL}_\mu > -3$, on top of $\text{isMuon} == 1$, for muons in Run 1 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

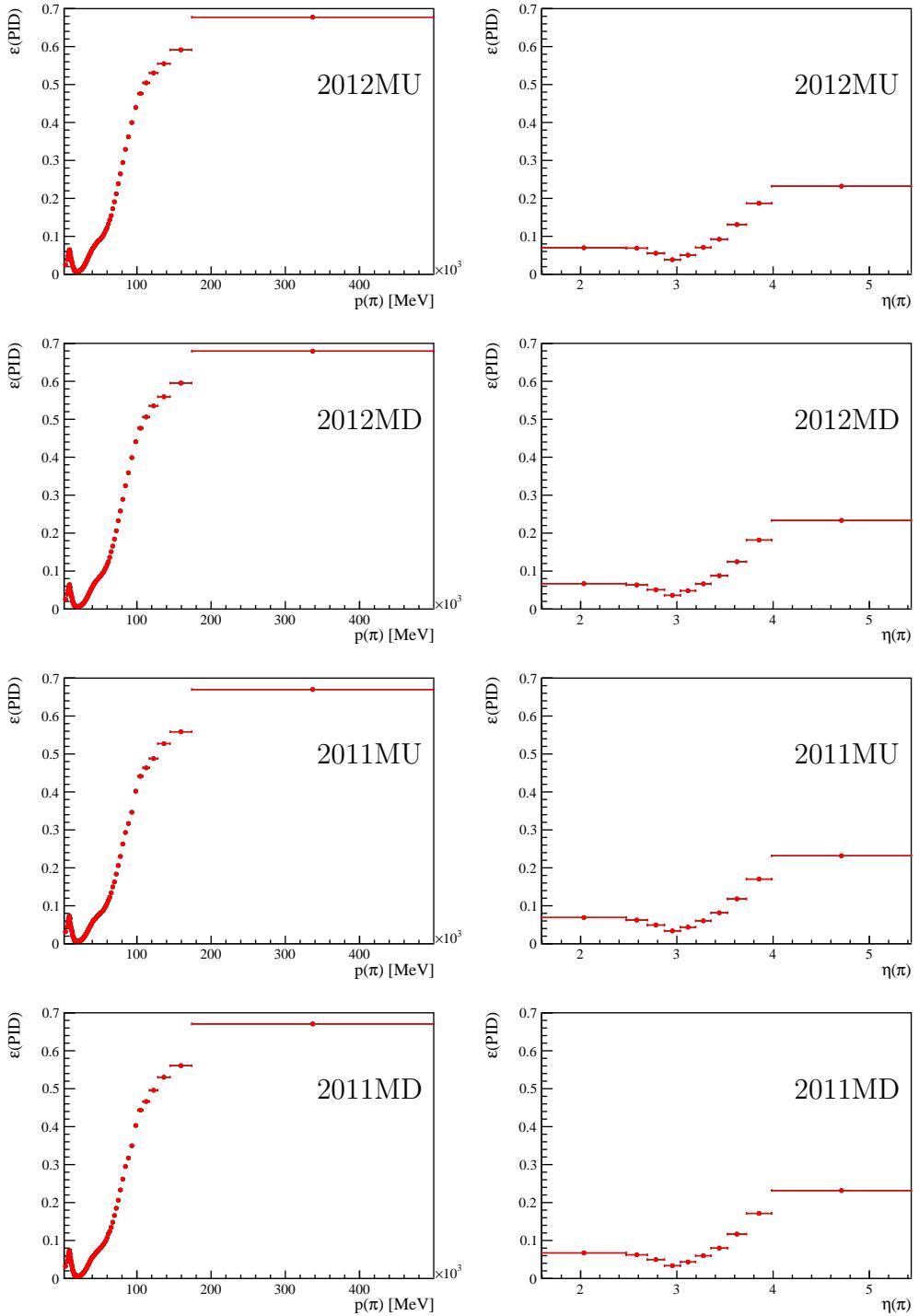


Figure 34: Particle (mis-)identification calibration histograms for the cut $\text{probNN}_K > 0.2$ for pions in Run 1 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

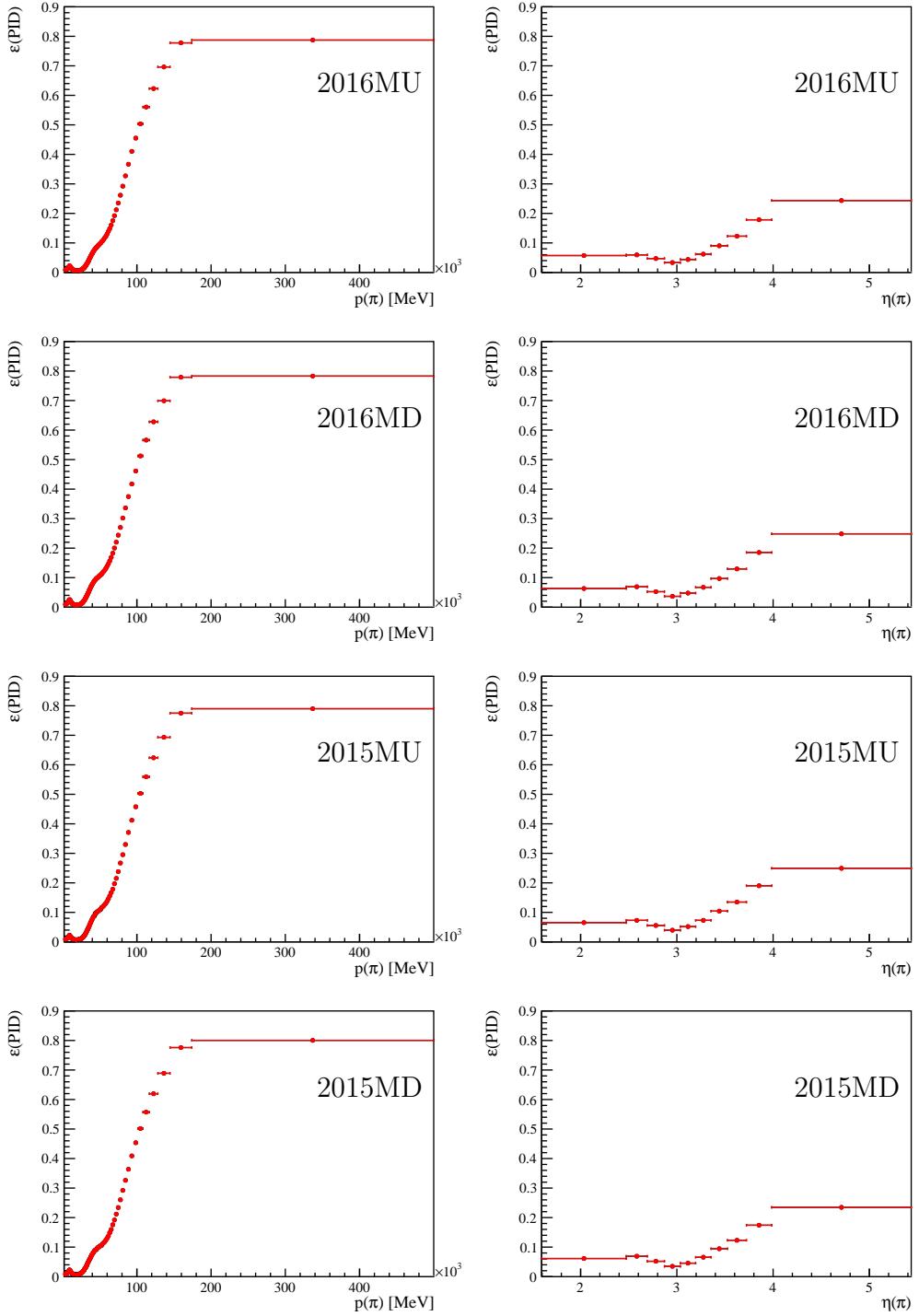


Figure 35: Particle (mis-)identification calibration histograms for the cut $\text{probNN}_K > 0.2$ for pions in Run 2 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

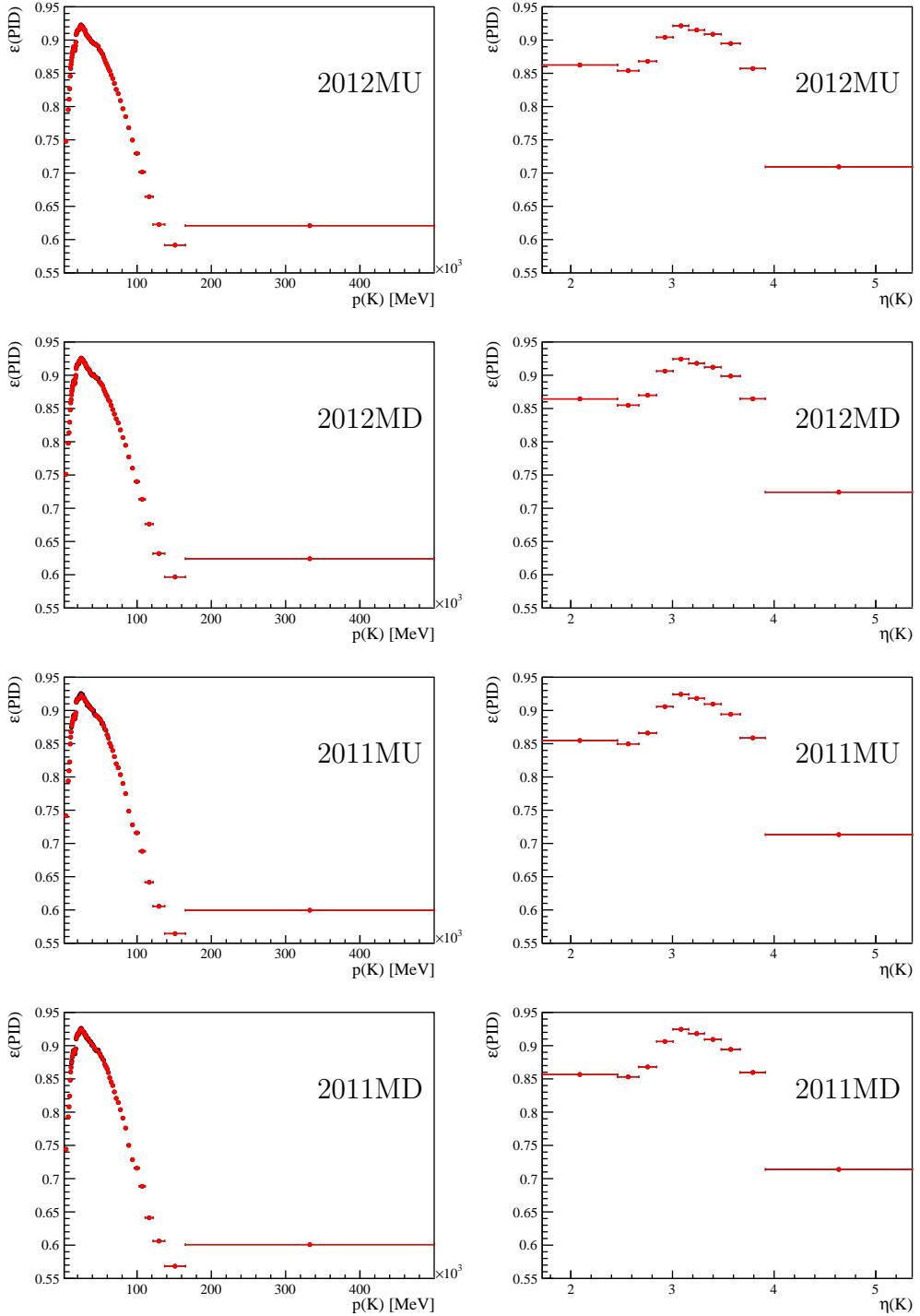


Figure 36: Particle identification calibration histograms for the cut $\text{probNN}_K > 0.2 \& \text{DLL}_e < 0$, for kaons in Run 1 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

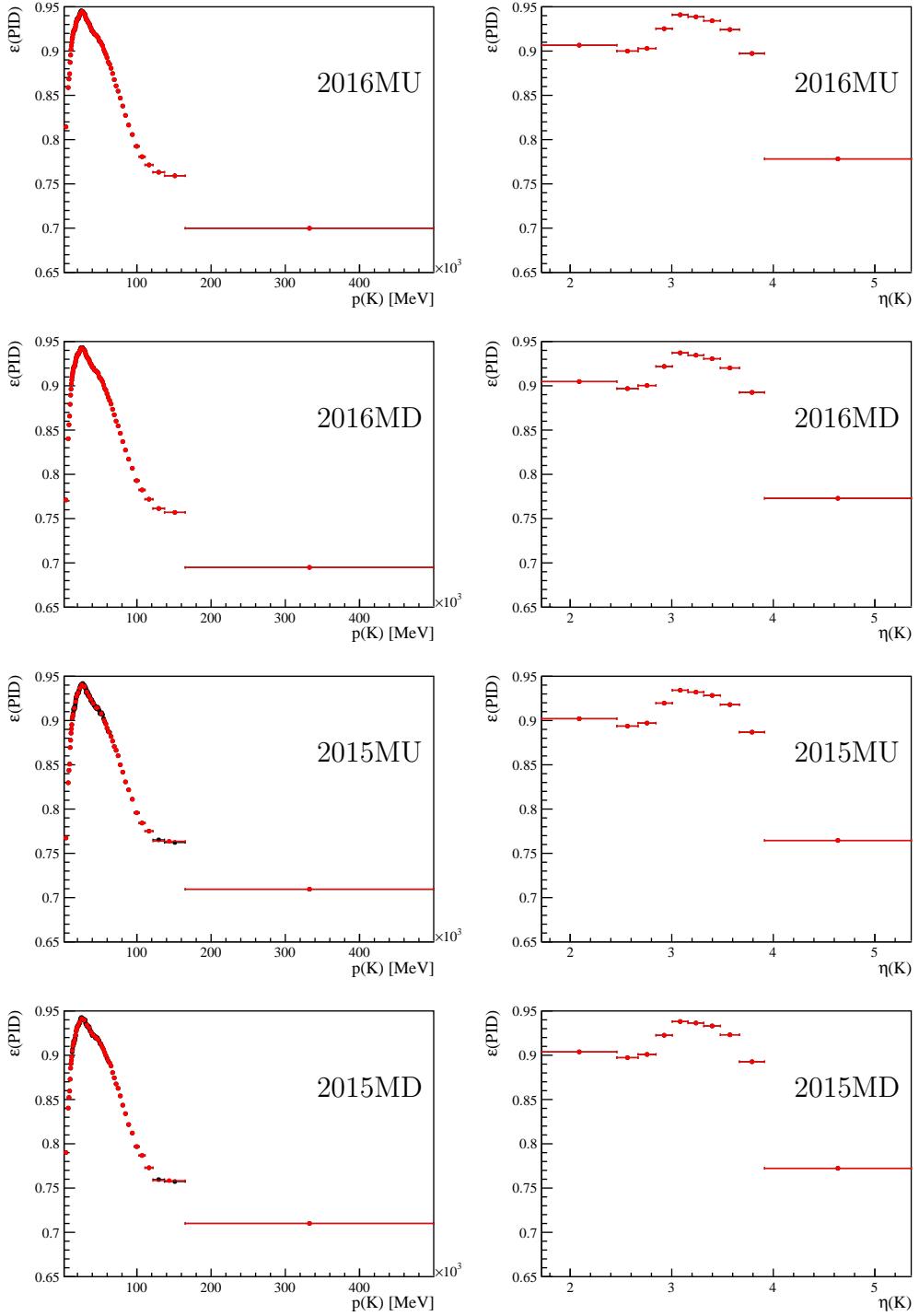


Figure 37: Particle identification calibration histograms for the cut $\text{probNN}_K > 0.2 \& \text{DLL}_e < 0$, for kaons in Run 2 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

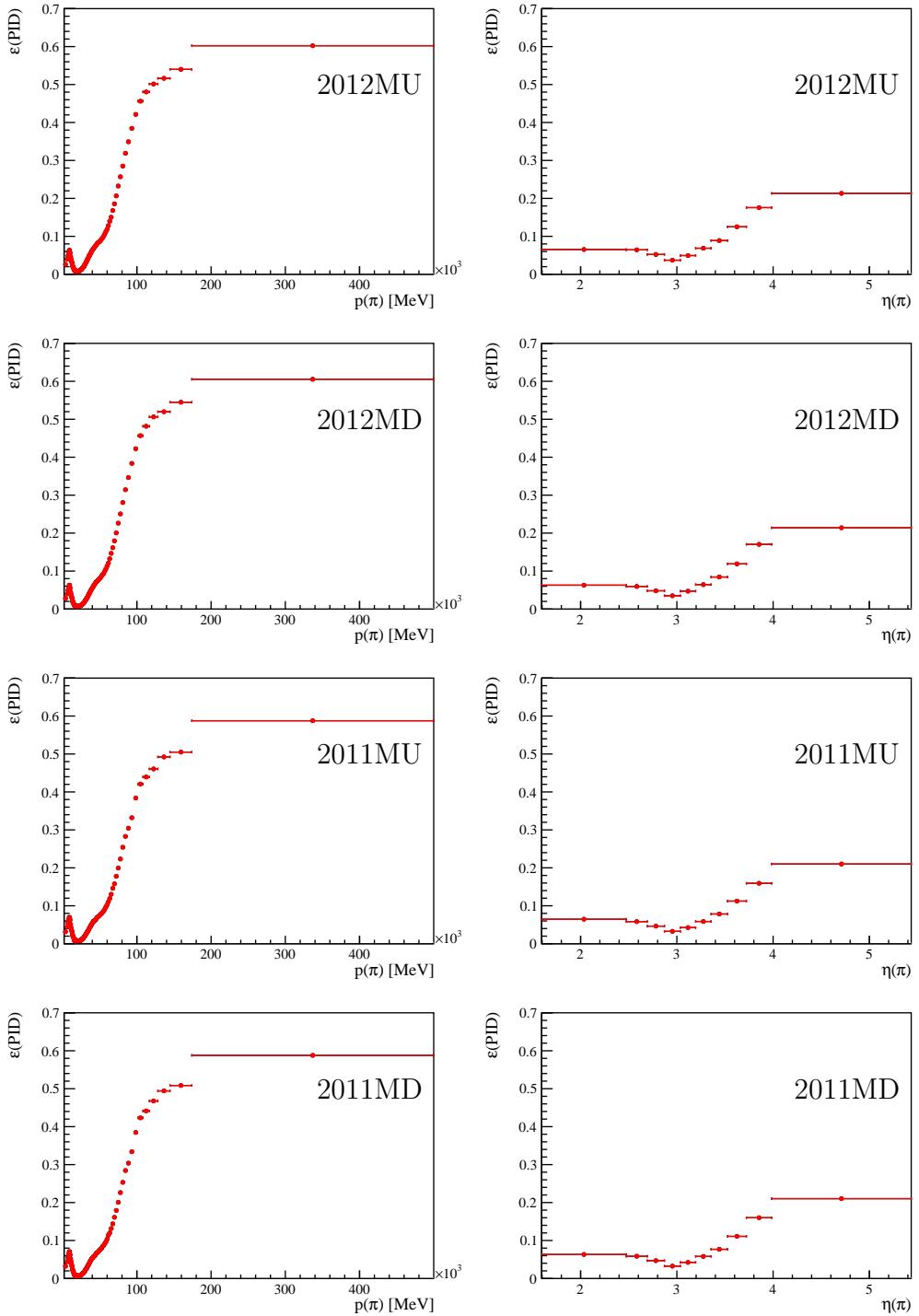


Figure 38: Particle (mis-)identification calibration histograms for the cut $\text{probNN}_K > 0.2 \& \text{DLL}_e < 0$ for pions in Run 1 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

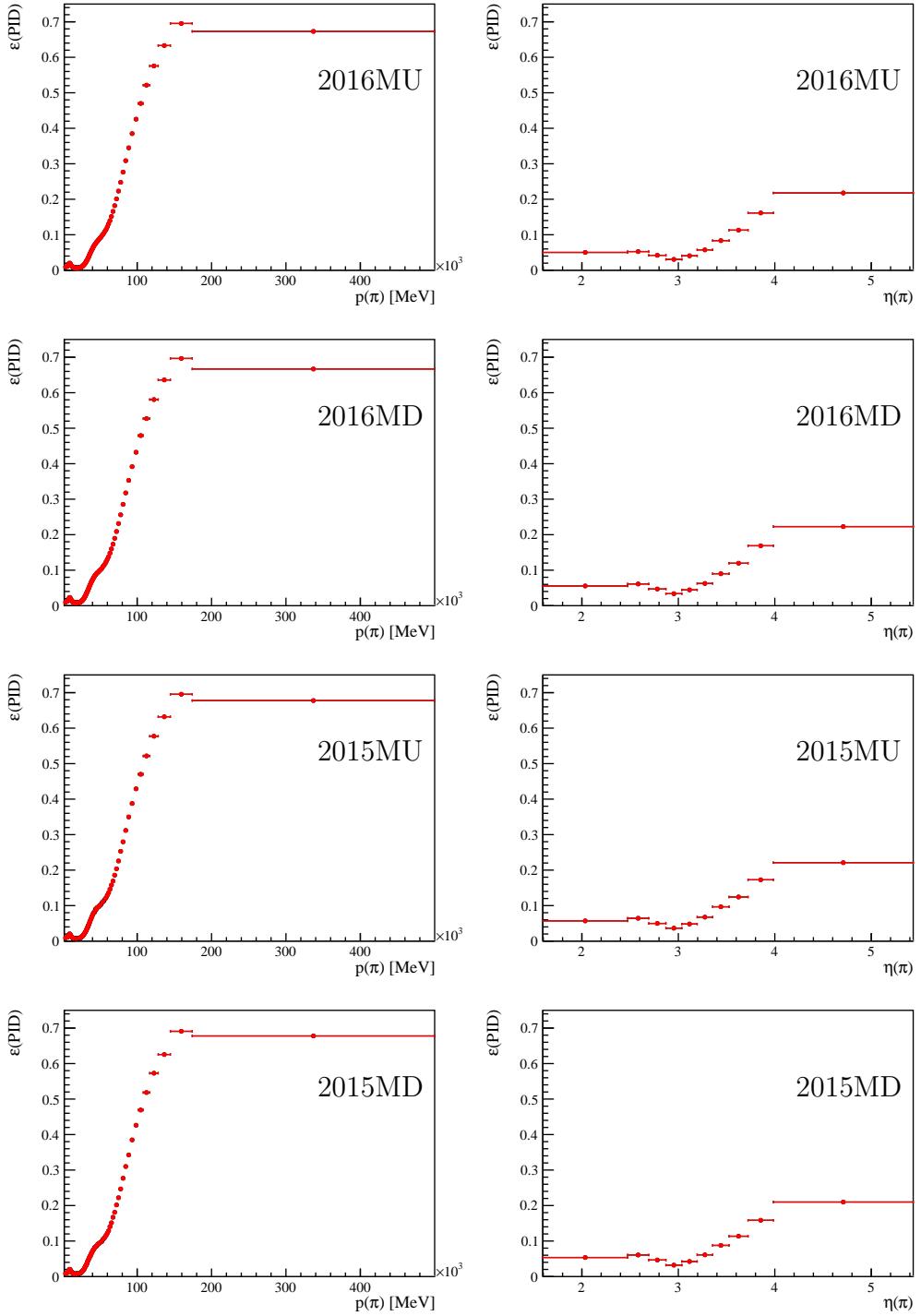


Figure 39: Particle (mis-)identification calibration histograms for the cut $\text{probNN}_K > 0.2 \& \text{DLL}_e < 0$ for pions in Run 2 data. The black (red) dots show the efficiency before (after) the rebinning procedure.

945 **6.4.2 Electron identification efficiencies**

946 The selection of the electron calibration tuples is applied in *super-*
 947 *Lenin/src/superMain5a(Run2).cc*, and the efficiencies are extracted in
 948 *superLenin/src/superMain5b(Run2).cc*, which use the functions in *super-*
 949 *Lenin/src/pidECalibHistoTools.cpp*.

950 The calibration sample for the electron PID efficiencies is the decay of the normalisation
 951 mode $B^+ \rightarrow K^+ J/\psi(e^+e^-)$. To extract the calibration histograms, the events are selected
 952 using a selection that does not apply any PID cut on the electrons. The stripping line used
 953 is the `ElectronIDBu2JpsieeK` line. On top of the requirements from the stripping, some
 954 extra requirements are placed offline to align the selection to that of the signal modes and
 955 to reduce the combinatorial background. All the cuts are summarised in Tab. 16.

Table 16: Stripping cuts from `ElectronIDBu2JpsieeK` line, and offline cuts.

Stripping cuts		Offline cuts	
<i>e</i>	$p_T > 0.5 \text{ GeV}$ $\chi_{\text{TrackFit}}^2 < 5$ $\chi_{\text{IP}}^2 > 9$	<i>e</i>	$\text{prob}_{\text{ghost}} < 0.3$ $p > 3 \text{ GeV}$
<i>K</i>	$p_T > 1.0 \text{ GeV}$ $\chi_{\text{TrackFit}}^2 < 4$ $\text{PID}_K > 0$ $\chi_{\text{IP}}^2 > 9$	<i>K</i>	$\text{probNN}_K > 0.2$
e^+e^-	$m > 2.2 \text{ GeV}$ $< 4.2 \text{ GeV}$ $\chi_{\text{DV}/\text{ndof}}^2 < 9$ $\chi_{\text{DV} \leftrightarrow \text{PV}}^2 > 5$	e^+e^-	$m > 2.45 \text{ GeV}$ $< 3.60 \text{ GeV}$
<i>B</i>	$m > 4.2 \text{ GeV}$ $< 6.0 \text{ GeV}$ $\chi_{\text{DV}/\text{ndof}}^2 < 9$	K^+e^-	$m > 1.885 \text{ GeV}$ $m_{e \rightarrow \pi} > 1.885 \text{ GeV}$
		<i>B</i>	$\chi_{\text{IP}}^2 < 9$
			$\text{nSPDHits} < 600 \text{ (Run 1)}$ $< 450 \text{ (Run 2)}$
		<i>e</i> tag	$p > 6.0 \text{ GeV}$ $p_T > 1.5 \text{ GeV}$ $\text{PID}_e > 5$

956 In the same way as for muon and kaon PID, electron PID efficiencies are computed
 957 using a tag & probe method. This consists of applying a tight PID requirement to the
 958 kaon and one of the two electrons (the tag) and using the other electron as a probe to
 959 measure the efficiency of a given PID cut. Both electrons are alternatively used as a tag
 960 or as a probe, and the two efficiencies are averaged. The extra requirements applied to
 961 the electron tag are reported in Tab. 16.

962 The electron PID efficiency is binned in four dimensions: year of data taking, momen-
 963 tum, rapidity, and two bins for whether the probe electron has a bremsstrahlung photon
 964 associated with it or not. The latter variable is denoted `hasBrem`. It is checked that the
 965 electron PID efficiencies in the two magnet polarities are compatible within their errors.

966 This electron PID efficiency depends on the hasBrem variable because the DLL_e variable
967 uses as an input the likelihood of a bremsstrahlung cluster belonging to the electron
968 candidate track.

969 The *sWeight* method that was used to extract muon and kaon PID efficiencies cannot
970 be used to precisely extract electron PID efficiencies, because the correlations between the
971 mass and the momentum and the high level of background prevent the *sWeight* method
972 from working properly. The *sWeight*'s are therefore used only to find the optimal binning
973 in p and η using the algorithm described in Section 6.4.1. To compute the efficiencies, a
974 fit to the J/ψ -constrained B mass is performed in each kinematic bin before and after
975 applying the PID cut to compute the efficiency. This method is referred to as the fit &
976 count method.

977 The electron PID efficiencies projected onto p , η and hasBrem are shown in Fig. 40-
978 Fig. 43 for each year of data taking. The efficiencies computed with the *sWeight* method,
979 as well as the default ones computed with the fit & count are shown, and the difference
980 illustrates that using the *sWeight* method would result in a biased efficiency at the level of
981 a few percent. All the fits to the electron calibration sample in each bin of p , η , hasBrem
982 are shown in Appendix D.

983 6.4.3 Combination of PID efficiencies and biases

984 The weight w^{PID} in Eq. 14 that is used to calculate the efficiency of the PID requirements
985 is computed as the product of the PID efficiencies for each track, extracted from the PID
986 calibration histograms:

$$w^{\text{PID}} = \varepsilon^{\text{PID}}(K^+) \cdot \varepsilon^{\text{PID}}(\ell^+) \cdot \varepsilon^{\text{PID}}(\ell^-). \quad (18)$$

987 Moreover, as stated in Eq. 13, the PID efficiencies ε^{PID} are defined on preselected
988 events, before applying trigger requirements.

989 The way in which the PID efficiencies are computed could result in two potential biases:

- 990 1. Factorisation bias: The assumption that the PID efficiencies of the three tracks are
991 independent might not hold perfectly. This has to be checked for the electron mode,
992 as requirements on DLL_e require inputs from the calorimeters, which have a low
993 granularity (i.e. two electron clusters can be superimposed in the calorimeter and
994 might correlate their DLL responses);
- 995 2. Trigger bias: the trigger requirements used in the calibration samples could also bias
996 the PID response in two cases. For the muon mode, the L0Muon trigger might bias
997 the efficiency of the isMuon requirement. For the electron mode, the L0Electron
998 trigger might bias the efficiency of the requirement on DLL_e .

999 These potential biases are studied on $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$
1000 simulated events, as detailed below.

1001 The efficiency of PID cut $\text{isMuon} = 1$ & $\text{DLL}_\mu > -2$ as a function of momentum for
1002 simulated muons is shown in Fig. 44. In black, the efficiency is shown for events without

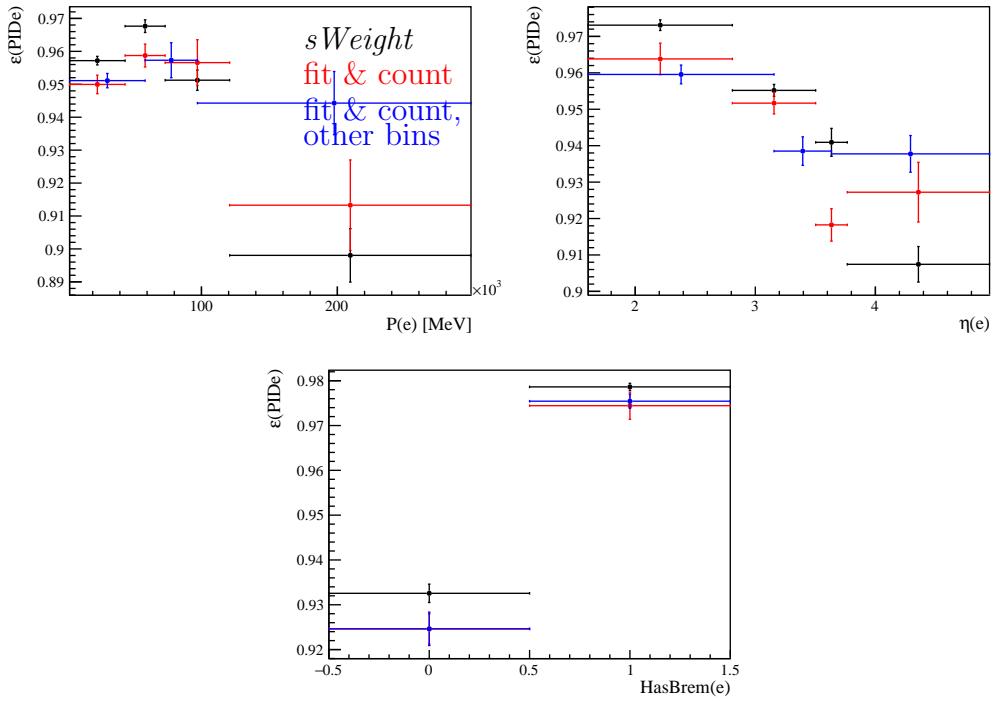


Figure 40: Efficiency of the PID requirement $\text{DLL}_e > 3$ measured from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ calibration sample in 2011. Efficiencies are shown as a function of momentum (top left), rapidity (top right) and whether the electron track has a bremsstrahlung cluster associated. The black dots correspond to the efficiencies computed with the *sWeight* method, and the red dots the efficiencies computed with the (nominal) fit & count method. The blue dots correspond to an alternative binning scheme used to compute systematic uncertainties.

any trigger or PID requirement. In red, the PID cut is applied on the other muon. The perfect agreement between the black and red curves demonstrates that the PID cuts on the two muons factorise. In blue, the efficiency is shown for events where any trigger line is fired. The bias with respect to the black line is due to the correlation between the L0Muon trigger line and the `isMuon = 1` requirement. In magenta, the muon probe is required to be TIS at the L0 and HLT level. This requirement is made in the muon PID calibration sample [22], and the good agreement between the magenta and black shows that potential trigger biases are negligible in the case of muons.

The efficiency of the PID cut $\text{probNN}_K > 0.2 \& \text{DLL}_e < 0$ as a function of momentum for simulated kaons in the decay $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ is shown in Fig. 45. The colour code is similar to that used in the muon case. In black, only the offline preselection is applied. In red, the electron PID requirements are applied. In blue, any trigger line is required to have fired. The agreement between the red and black shows that the efficiency of the kaon PID requirement is independent of the electron PID requirement. There is a slight disagreement between black and blue, showing that there is a small trigger bias.

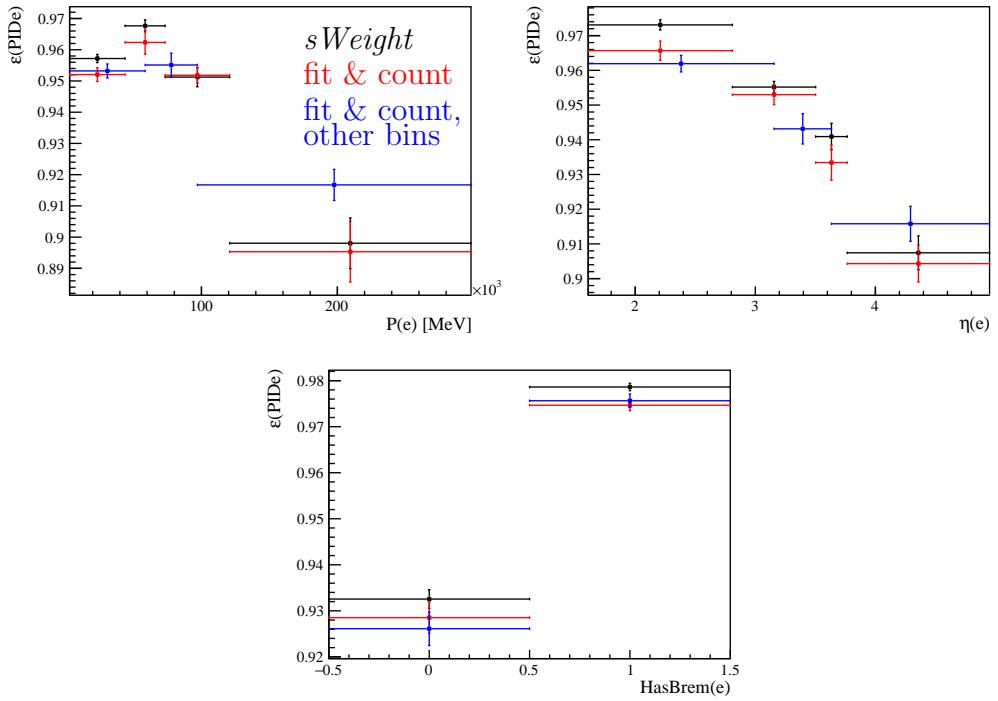


Figure 41: Efficiency of the PID requirement $\text{DLL}_e > 3$ measured from the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ calibration sample in 2012. Efficiencies are shown as a function of momentum (top left), rapidity (top right) and whether the electron track has a bremsstrahlung cluster associated. The black dots correspond to the efficiencies computed with the *sWeight* method, and the red dots the efficiencies computed with the (nominal) *fit & count* method. The blue dots correspond to an alternative binning scheme used to compute systematic uncertainties.

1018 The efficiency of the PID cut $\text{DLL}_e > 3$ applied on one of the electrons in simulated
 1019 $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events is shown in Fig. 46. In black, only the offline preselection is
 1020 applied. In red, the PID cut is applied to the other electron in the event. In blue, any
 1021 trigger line is required to have fired. In magenta, the probe electron is required to be TIS
 1022 at the L0 and HLT level. The agreement between red and black shows that there the two
 1023 electron PID efficiencies factorise. There is, however, a small trigger bias. Unlike in the
 1024 muon case, where this could be corrected with the TIS requirement on the probe, this is
 1025 not the case with electrons: there is still a disagreement between magenta and black.

1026 To summarise, the PID efficiencies factorise well between the three particles in the decay.
 1027 There is however a small trigger bias when computing the electron and kaon PID efficiencies.
 1028 The systematic uncertainty associated with this bias is discussed in Section 10.3.

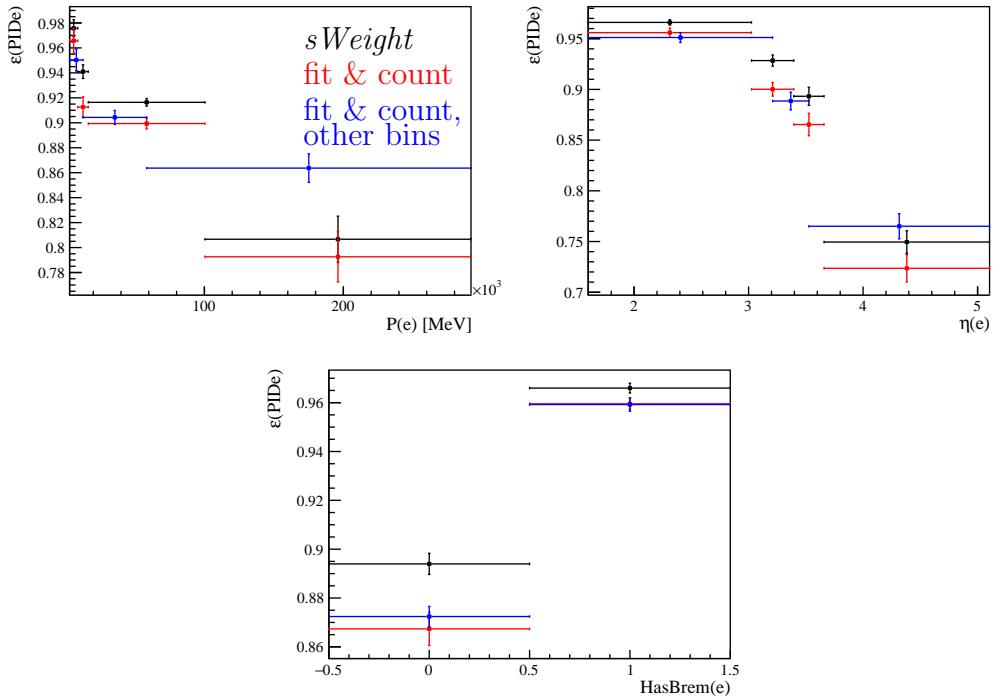


Figure 42: Efficiency of the PID requirement $\text{DLL}_e > 3$ measured from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ calibration sample in 2015. Efficiencies are shown as a function of momentum (top left), rapidity (top right) and whether the electron track has a bremsstrahlung cluster associated. The black dots correspond to the efficiencies computed with the *sWeight* method, and the red dots the efficiencies computed with the (nominal) fit & count method. The blue dots correspond to an alternative binning scheme used to compute systematic uncertainties.

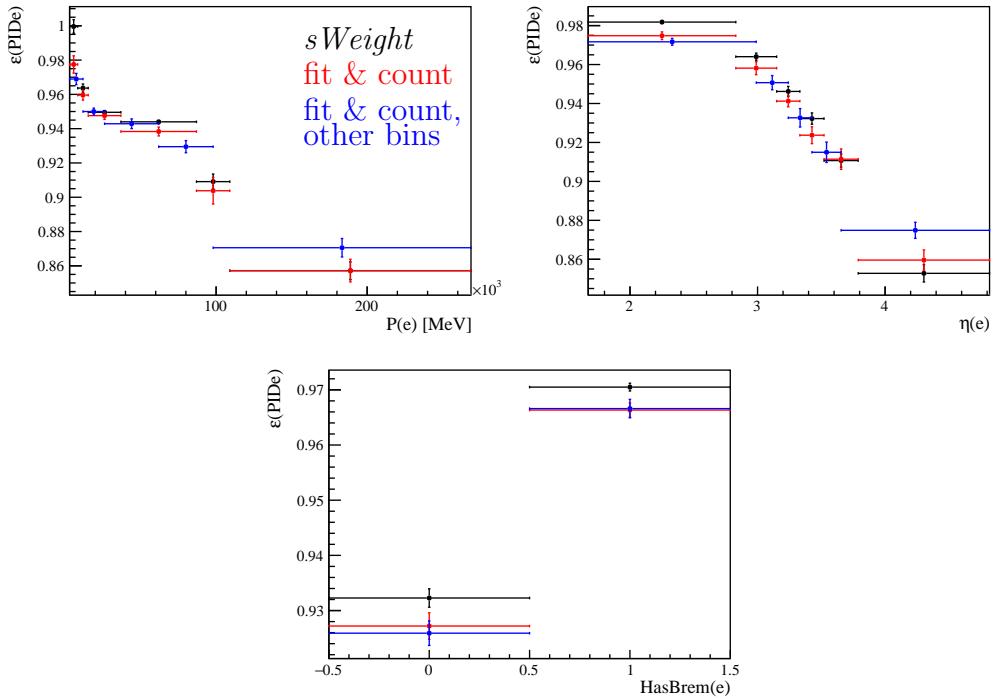


Figure 43: Efficiency of the PID requirement $\text{DLL}_e > 3$ measured from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ calibration sample in 2016. Efficiencies are shown as a function of momentum (top left), rapidity (top right) and whether the electron track has a bremsstrahlung cluster associated. The black dots correspond to the efficiencies computed with the *sWeight* method, and the red dots the efficiencies computed with the *fit & count* method (nominal). The blue dots correspond to an alternative binning scheme used to compute systematic uncertainties.

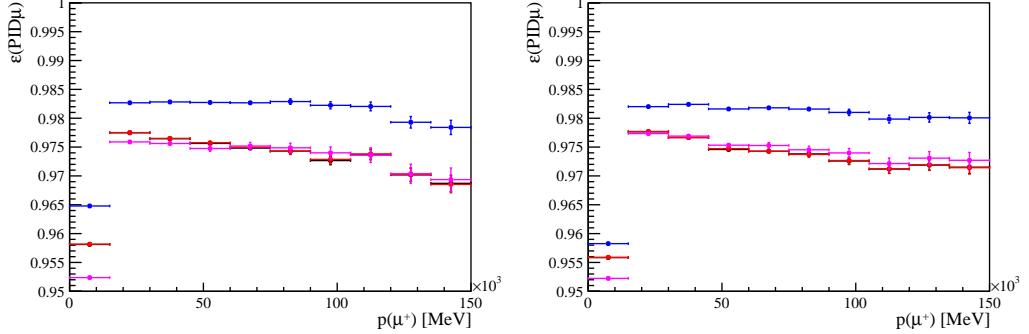


Figure 44: Efficiency of the muon PID cut as a function of momentum, for one of the two muons from simulated $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events. The left hand plot corresponds to Run 1, the right hand plot to Run 2. In black, only the offline preselection is applied. In red, the PID cut is applied to the other muon. In blue, the requirement that the event has been triggered by any line is applied. In magenta, the probe muon is required to be TIS with respect to any trigger (which is what is used in the calibration sample).

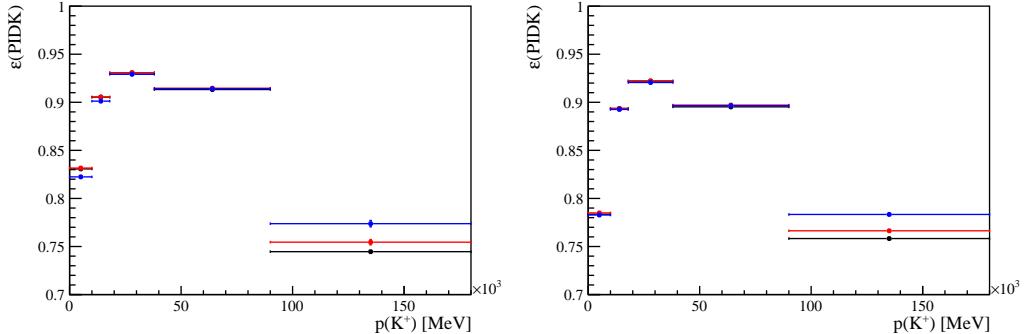


Figure 45: Efficiency of the kaon PID cut as a function of momentum, from simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events. The left hand plot corresponds to Run 1, the right hand plot to Run 2. In black, only the offline preselection is applied. In red, the PID cut is applied to the two electrons. In blue, the requirement that the event has been triggered by any line is applied.

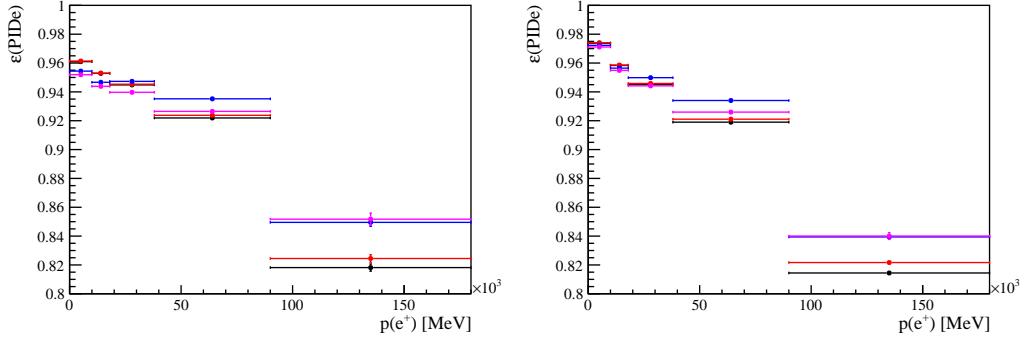


Figure 46: Efficiency of the electron PID cut as a function of momentum, for one of the two electrons from simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events. The left hand plot corresponds to Run 1, the right hand plot to Run 2. In black, only the offline preselection is applied. In red, the PID cut is applied to the other electron. In blue, the requirement that the event has been triggered by any line is applied. In magenta, the probe electron is required to be TIS with respect to any trigger.

1029 **6.5 Kinematic weights for trigger histograms**

1030 To extract the trigger weights as explained in the following two sections, a first set
1031 of kinematic weights is computed that correct the B kinematics as well as the χ^2_{DV}
1032 distributions in the simulation to match the data better. The way in which these kinematic
1033 weights are computed is exactly the same as described in Section 6.8, except for the fact
1034 that the μTOS corrections weights are not applied to the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ sample
1035 to extract the kinematic weights. Applying these kinematic weights to the simulation or
1036 not prior to extracting the trigger correction weights has next-to-no effect.

1037 **6.6 L0 efficiencies calibration**

1038 The trigger calibration histograms are computed in *super-*
1039 *Lenin/src/superMain8(Run2).cc*, using functions from *super-*
1040 *Lenin/src/triggerTablesTools(Run2).cpp* and *superLenin/src/triggerTablesFitter.cpp*.

1041 The efficiencies of the four L0 trigger categories defined in Section 5.5 are known to
1042 be described imperfectly in the simulation. Therefore, they are calibrated from the data
1043 using the tag & probe method on the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$
1044 samples, to which the whole preselection chain, including the PID cuts, is applied. To
1045 obtain statistically clean data samples, the *sWeight* method is used. Indeed, after applying
1046 the PID cuts, the normalisation samples are very clean and, unlike in the case of the
1047 electron PID calibration histograms, the bias related to the *sWeight* is negligible.

1048 The weights w^{trig} applied to the simulation are defined as a data over simulation ratio
1049 of trigger efficiencies. This is different from the PID weights, which are defined as PID
1050 efficiencies measured from the data. This has to be done because the HLT requirement
1051 is very dependent on the L0 response, and ignoring this correlation would introduce a
1052 significant bias (of the order of 10% on the total trigger efficiency).

1053 In the following sections, the calibration of the trigger efficiencies in each of the four
1054 L0 categories is presented. When several calibration histograms are shown in a figure, that
1055 used for the nominal efficiency corrections is always drawn in red.

1056 **6.6.1 $e\text{TOS}$: calibration histograms**

1057 The efficiency of an electron e satisfying the **L0Electron** line requirement, $\varepsilon^{\text{L0}e}(e)$, is
1058 measured on simulation and on data. It is parametrised as a function of two variables:
1059 the transverse energy and the three ECAL regions. The transverse energy E_T is measured
1060 with a low resolution by the ECAL and the **L0Electron** algorithm is triggered if this
1061 measured energy is above a threshold, as explained in Section 5.5. The size of the ECAL
1062 cells depends on the ECAL region (smallest in the inner and largest in the outer part),
1063 and therefore the response of the **L0Electron** trigger is expected to depend on the ECAL
1064 region.

1065 The **L0Electron** efficiency measured on simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events is
1066 shown in Fig. 47-Fig. 50 for each year of data taking, as a function of E_T and for each

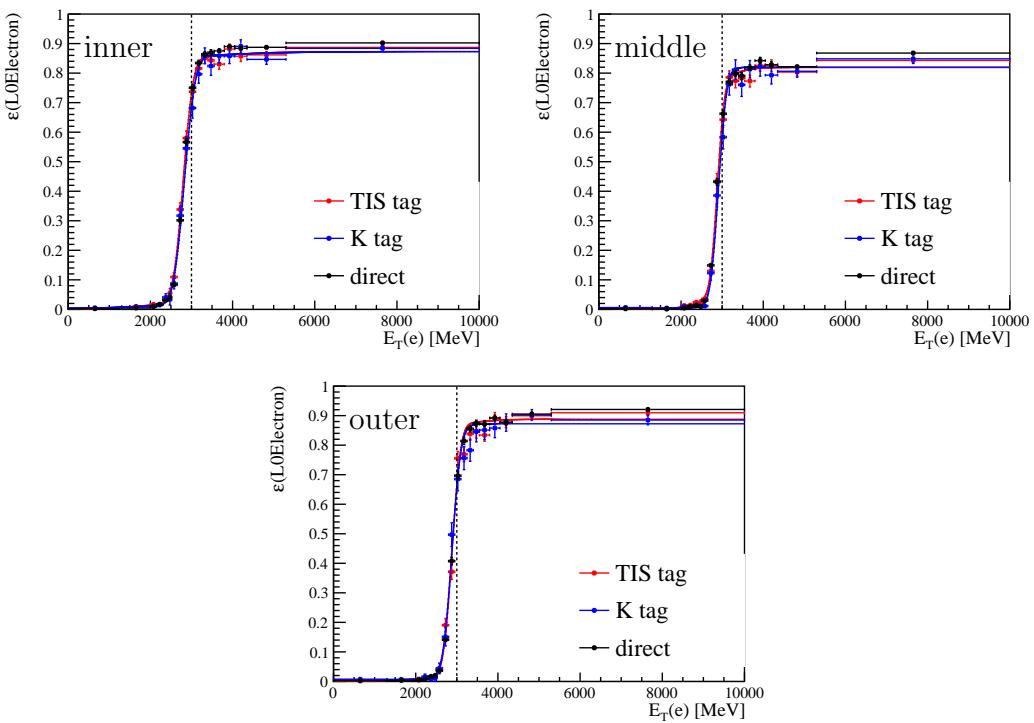


Figure 47: Efficiency of the L0Electron line as a function of the transverse energy of the electron candidate for simulated $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events, in the three ECAL regions and in the 2011 data taking conditions. The efficiencies are shown for three different tags: TIS tag, K tag or no tag (direct). The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

1067 ECAL region. The trigger fiducial cut introduced in Tab. 7 is shown with a dotted black
1068 line. The efficiency is measured on events selected in three different ways:

- 1069 • First, the efficiency is measured on events to which the offline preselection and PID
1070 requirements are applied. This is labelled as “direct” in Fig. 47-Fig. 50. This is
1071 precisely the efficiency that needs to be measured, as the L0 efficiency is defined on
1072 top of the offline preselection (see Eq. 14). However, this efficiency is not accessible
1073 in data, as data events need to have satisfied trigger requirements at the L0 and
1074 HLT level;
- 1075 • Second, the efficiency is measured on events that satisfy the offline preselection,
1076 PID requirements, HLT requirements and that have been triggered on the L0 level
1077 independently of the signal B by a hadron or a muon (in practice, by requiring
1078 B_L0Hadron_TIS or B_L0Muon_TIS). This is labelled as “TIS tag” in Fig. 47-Fig. 50;
- 1079 • Third, the efficiency is measured on events that satisfy the offline preselection, PID
1080 requirements, HLT requirements and that have been triggered by the hadron firing

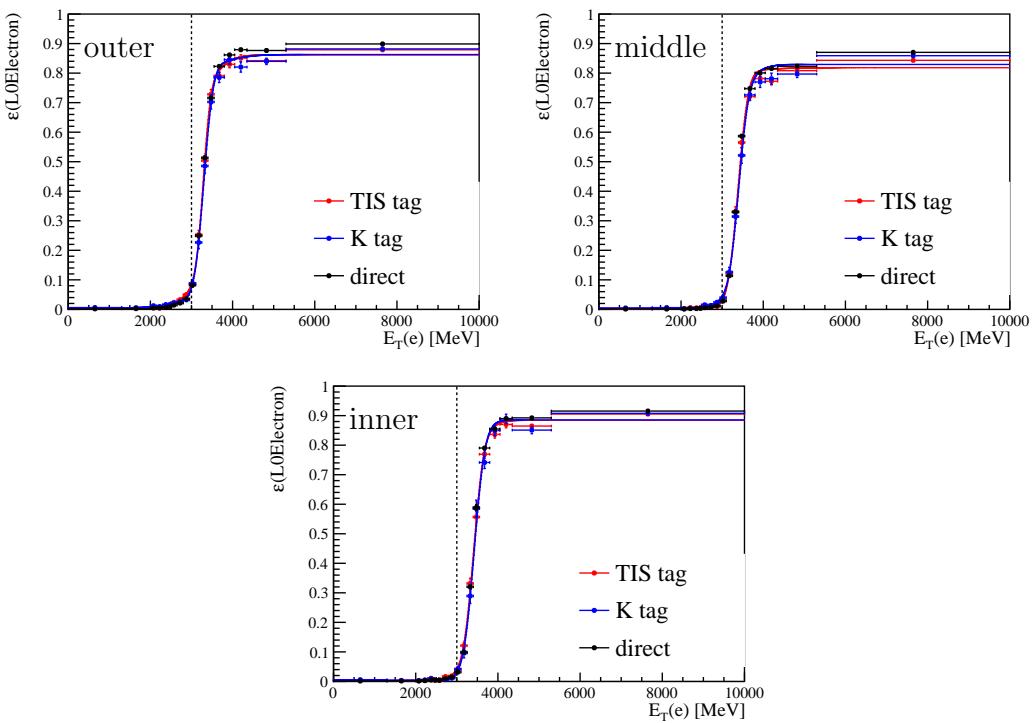


Figure 48: Efficiency of the L0Electron line as a function of the transverse energy of the electron candidate for simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, in the three ECAL regions and in the 2012 data taking conditions. The efficiencies are shown for three different tags: TIS tag, K tag or no tag (direct). The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

1081 the L0Hadron line. This is labelled as “ K tag” in Fig. 47-Fig. 50, and has poorer
1082 statistics than the previous category.

1083 The TIS tag is used for the nominal trigger corrections. The bias this tag introduces,
1084 referred to as “tag bias”, is small, as demonstrated by the fact that the efficiencies measured
1085 with the TIS tag are close to the “direct” efficiencies. The systematic uncertainties that
1086 tag bias introduce are discussed in Section 10.4.

1087 The L0Electron efficiency measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events is shown
1088 in Fig. 51-Fig. 54 for each year of data taking, for the TIS tag and K tag.

1089 The data and simulation plot are fit with an error function with three free parameters:

$$\frac{1}{2} \cdot \left(1 + \text{erf} \left(\frac{E_T - t}{\sqrt{2}\sigma} \right) \right) + a, \quad (19)$$

1090 where t is the threshold, σ the resolution and a the efficiency at $E_T = 0$ which is non zero
1091 because noise or hits in the ECAL from other tracks in the event can fire the trigger, and
1092 be falsely matched to the candidate, independently of its E_T value. The fit is performed in

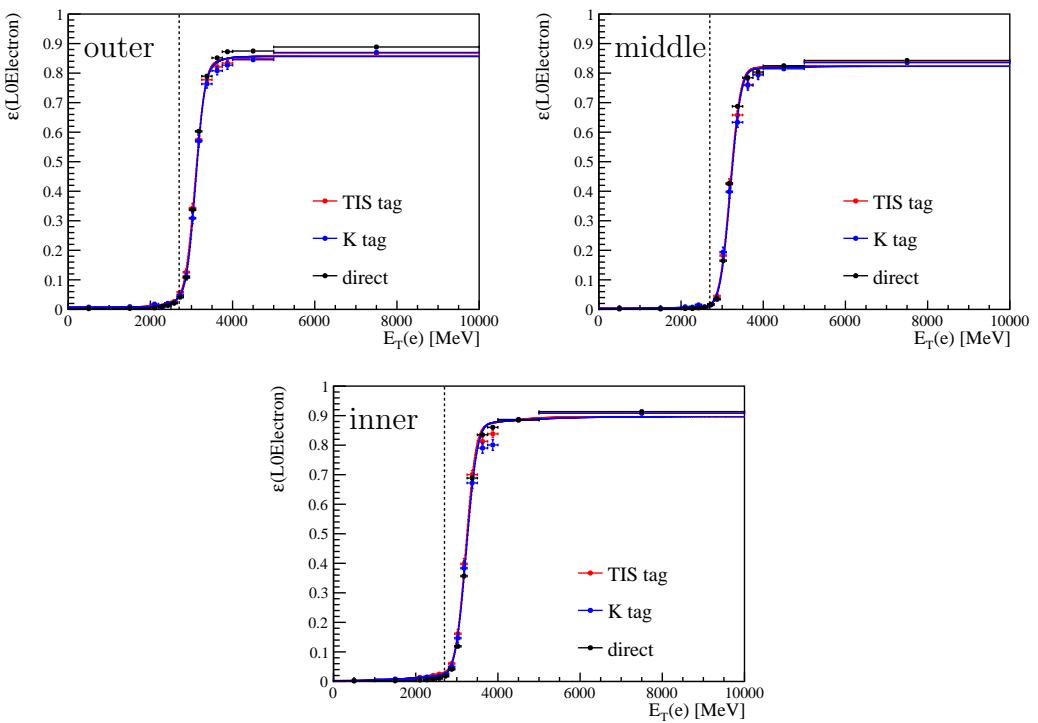


Figure 49: Efficiency of the L0Electron line as a function of the transverse energy of the electron candidate for simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, in the three ECAL regions and in the 2015 data taking conditions. The efficiencies are shown for three different tags: TIS tag, K tag or no tag (direct). The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

order to reduce the uncertainty due to the limited size of the trigger calibration samples.
The fit function is shown as a solid line on top of all trigger calibration histograms.

The ratio of data over simulation L0Electron efficiencies are shown in Fig. 55-Fig. 58. It can be seen that far above the threshold, the simulation agrees well with the data. As expected, the ratio is not equal to unity close to the threshold. In particular, the disagreement becomes very significant under the threshold, resulting in weights with very large values. This justifies the use of trigger fiducial cuts, again shown with a black dotted line, that veto the regions where the disagreement is highest. In addition, it appears that in the inner region of the ECAL, the trigger efficiency plateaus at an efficiency that is 3-5% higher in the data than in simulation. However, this is understood to be related to the fact that no PID requirements are applied to the simulation in this comparison, as PID efficiencies are extracted directly from data. Figure 59 shows how the ratio between data and simulation plateaus at 1 when the PID selection is included in the simulation. Note that the two corrections would be applied to simulation selected in two different ways, so the total efficiency would be the same regardless of which of these correction were used. The trigger histograms extracted with simulated samples in which the pid cuts are

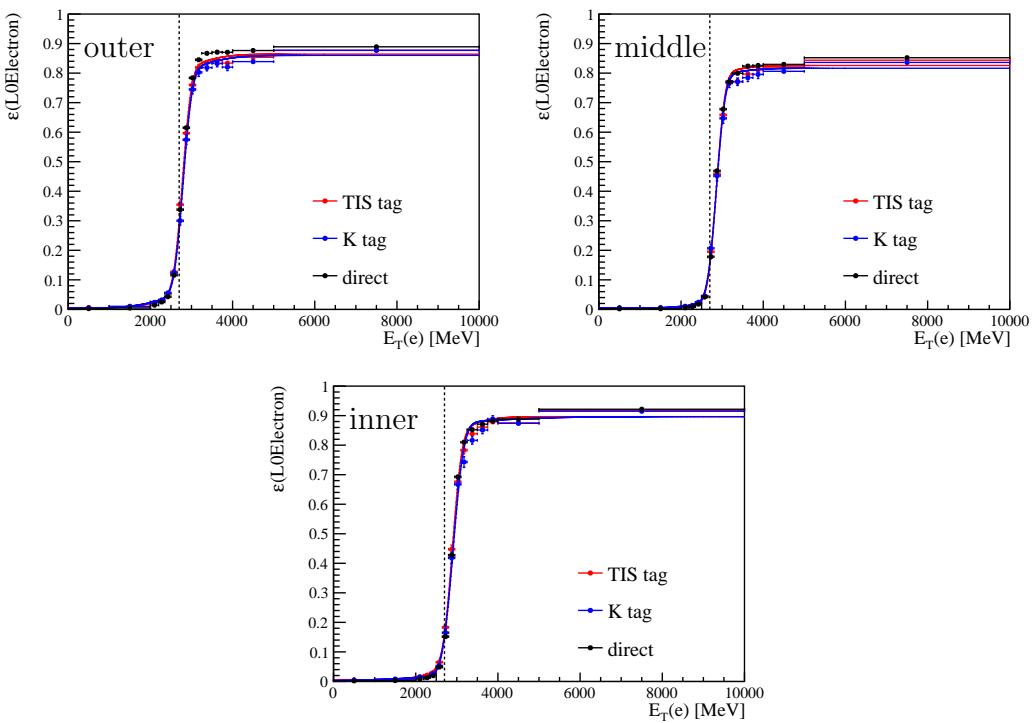


Figure 50: Efficiency of the L0Electron line as a function of the transverse energy of the electron candidate for simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, in the three ECAL regions and in the 2016 data taking conditions. The efficiencies are shown for three different tags: TIS tag, K tag or no tag (direct). The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

not applied are used as the nominal histograms in order to compute the efficiency product $\varepsilon^{\text{PID}} \cdot \varepsilon^{\text{trig}}$ consistently.

6.6.2 e TOS: efficiency combination

Using the L0Electron efficiencies measured on simulation and data, the e TOS correction weights are given

$$w^{e\text{TOS}} = \frac{1 - (1 - \varepsilon_{\text{data}}^{\text{L0e}}(e^+)) \cdot (1 - \varepsilon_{\text{data}}^{\text{L0e}}(e^-))}{1 - (1 - \varepsilon_{\text{sim}}^{\text{L0e}}(e^+)) \cdot (1 - \varepsilon_{\text{sim}}^{\text{L0e}}(e^-))}. \quad (20)$$

This formula is only exact, however, if the trigger efficiencies from the two electrons are independent of each other. This is not exactly the case, as can be seen in Fig. 60. This shows the L0Electron efficiency measured on simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events. The efficiency of one of the two electrons being TOS depends on whether the other electron is TOS, going down if the other electron is TOS. However, this bias is much reduced if the distance between the two electrons in the calorimeter is required to be above 1 m (this cut is used only as a cross-check and is not applied as part of the selection). This can

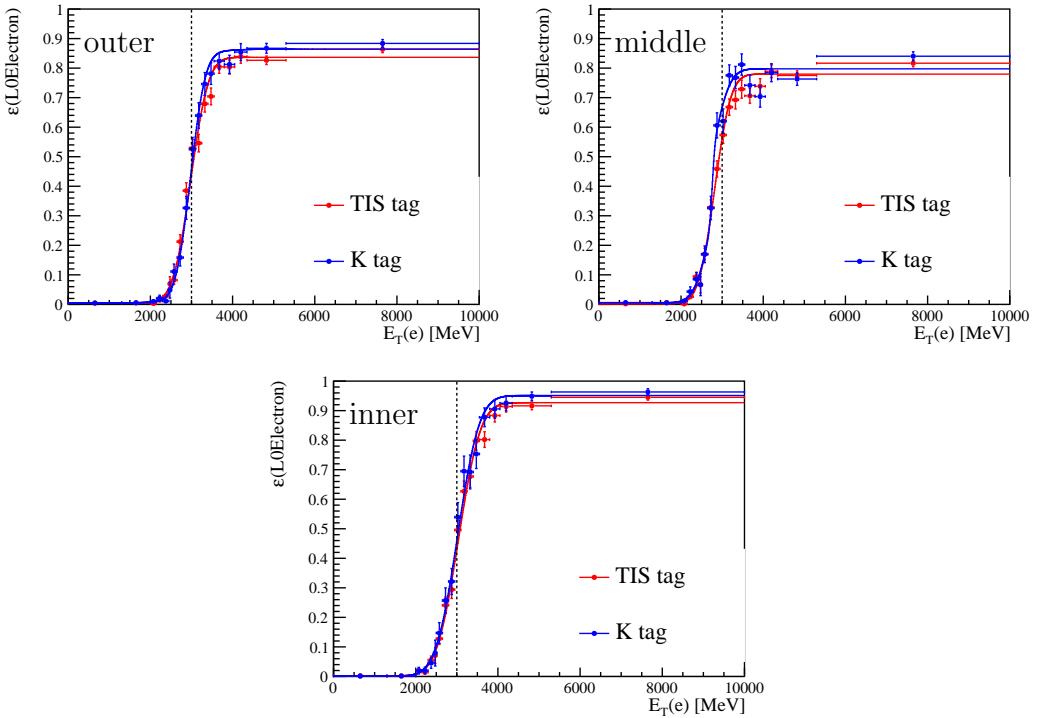


Figure 51: Efficiency of the L0Electron line as a function of the transverse energy of the electron candidate for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events, in the three ECAL regions and in the 2011 data taking conditions. The efficiencies are shown for two different tags: TIS tag, K tag. The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

be related to the fact that the front-end-boards in the ECAL only store the highest- E_T candidate on each trigger validation board (i.e. one candidate for four front-end boards). The systematic uncertainty related to the imperfect factorisation of the trigger efficiencies of the two electrons is discussed in Section 10.4.

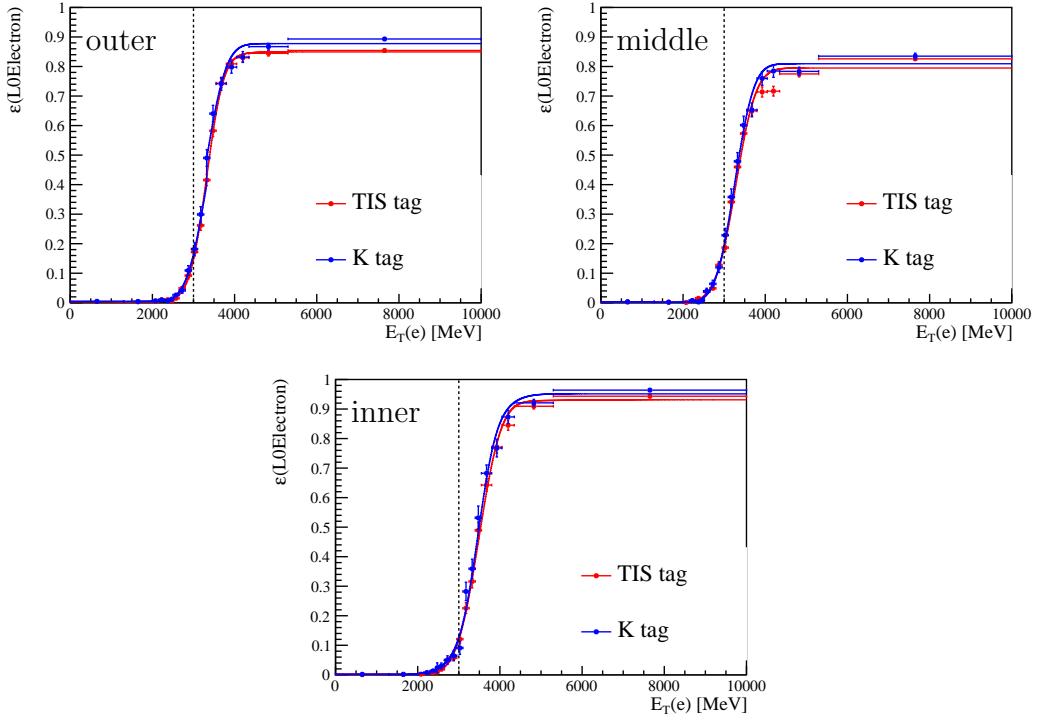


Figure 52: Efficiency of the L0Electron line as a function of the transverse energy of the electron candidate for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events, in the three ECAL regions and in the 2012 data taking conditions. The efficiencies are shown for two different tags: TIS tag, K tag. The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

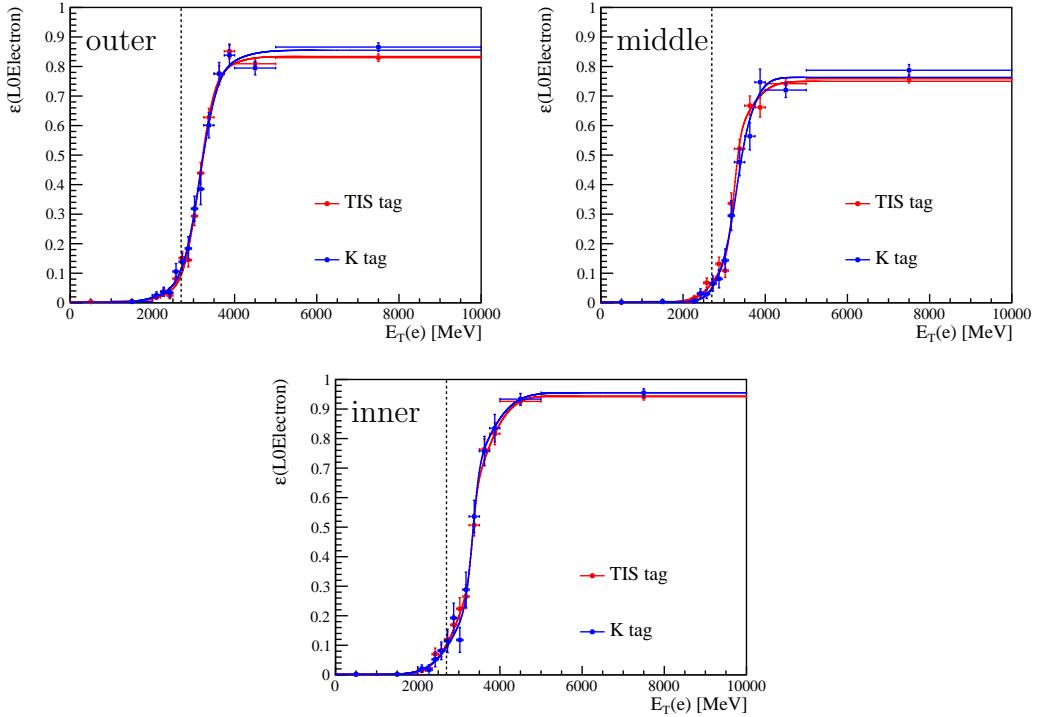


Figure 53: Efficiency of the L0Electron line as a function of the transverse energy of the electron candidate for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events, in the three ECAL regions and in the 2015 data taking conditions. The efficiencies are shown for two different tags: TIS tag, K tag. The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

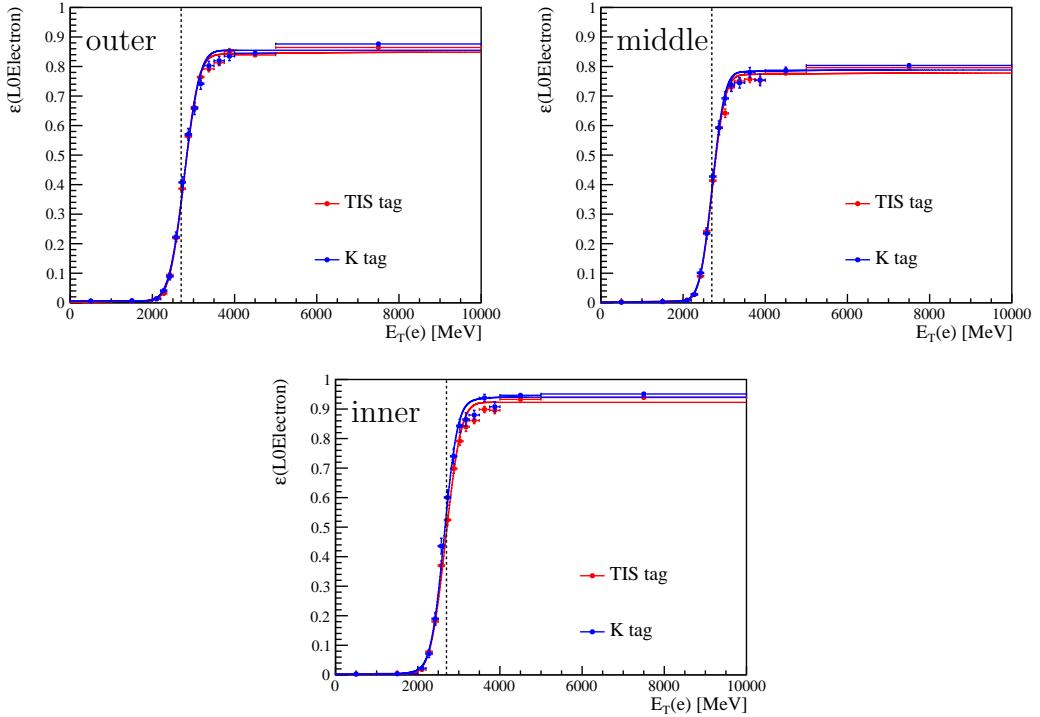


Figure 54: Efficiency of the L0Electron line as a function of the transverse energy of the electron candidate for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events, in the three ECAL regions and in the 2016 data taking conditions. The efficiencies are shown for two different tags: TIS tag, K tag. The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

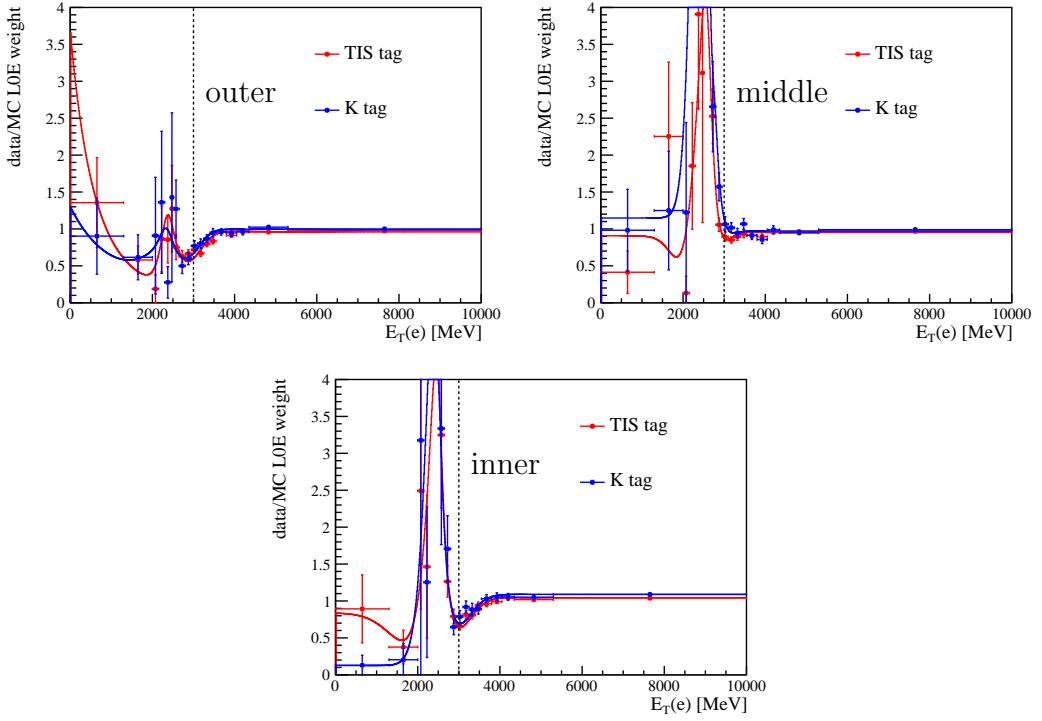


Figure 55: Data over simulation L0Electron efficiency ratio as a function of the transverse energy of the electron candidate for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, in the three ECAL regions and in the 2011 data taking conditions. The ratio is shown for two different tags: TIS tag, K tag. The ratio of the functions that fit the efficiencies are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

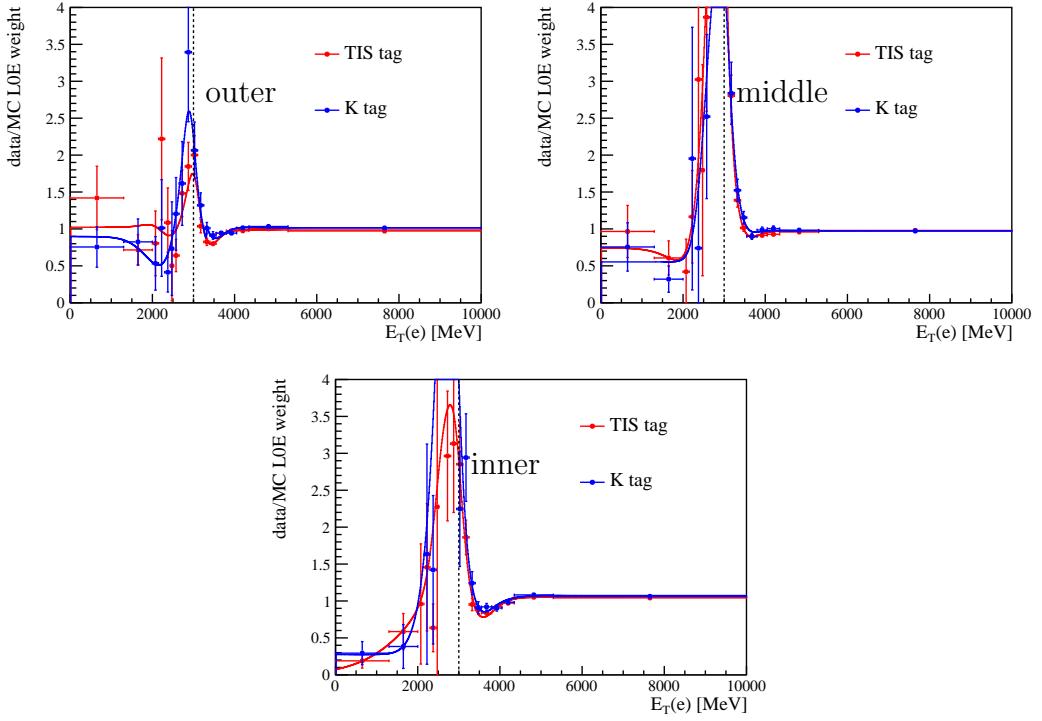


Figure 56: Data over simulation L0Electron efficiency ratio as a function of the transverse energy of the electron candidate for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, in the three ECAL regions and in the 2012 data taking conditions. The ratio is shown for two different tags: TIS tag, K tag. The ratio of the functions that fit the efficiencies are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

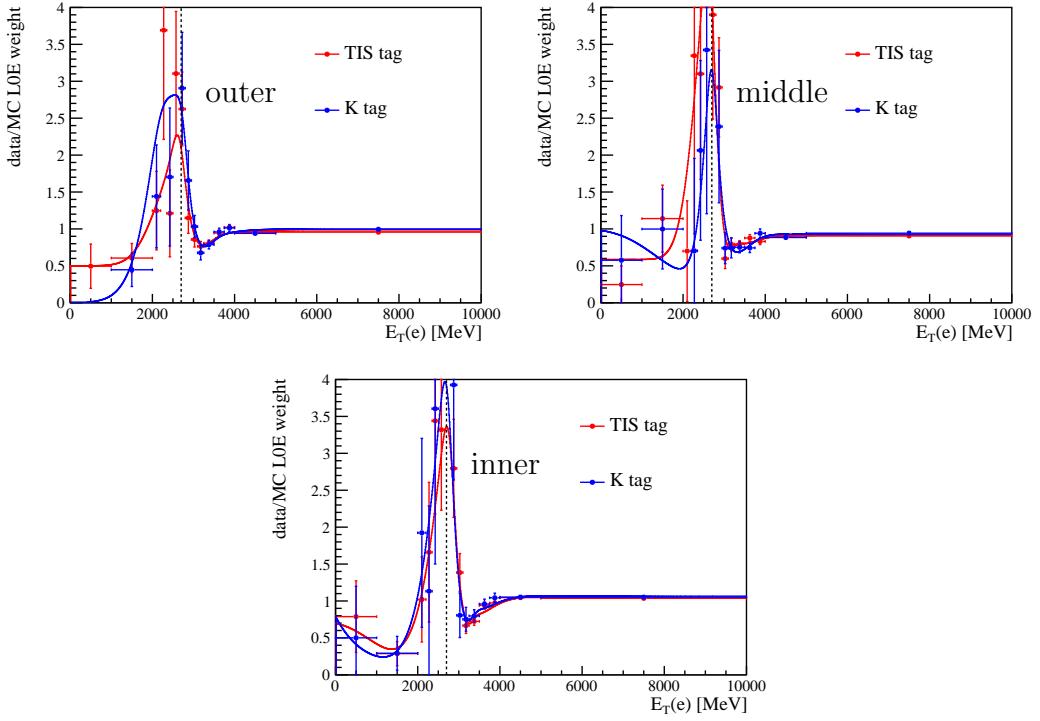


Figure 57: Data over simulation L0Electron efficiency ratio as a function of the transverse energy of the electron candidate for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, in the three ECAL regions and in the 2015 data taking conditions. The ratio is shown for two different tags: TIS tag, K tag. The ratio of the functions that fit the efficiencies are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

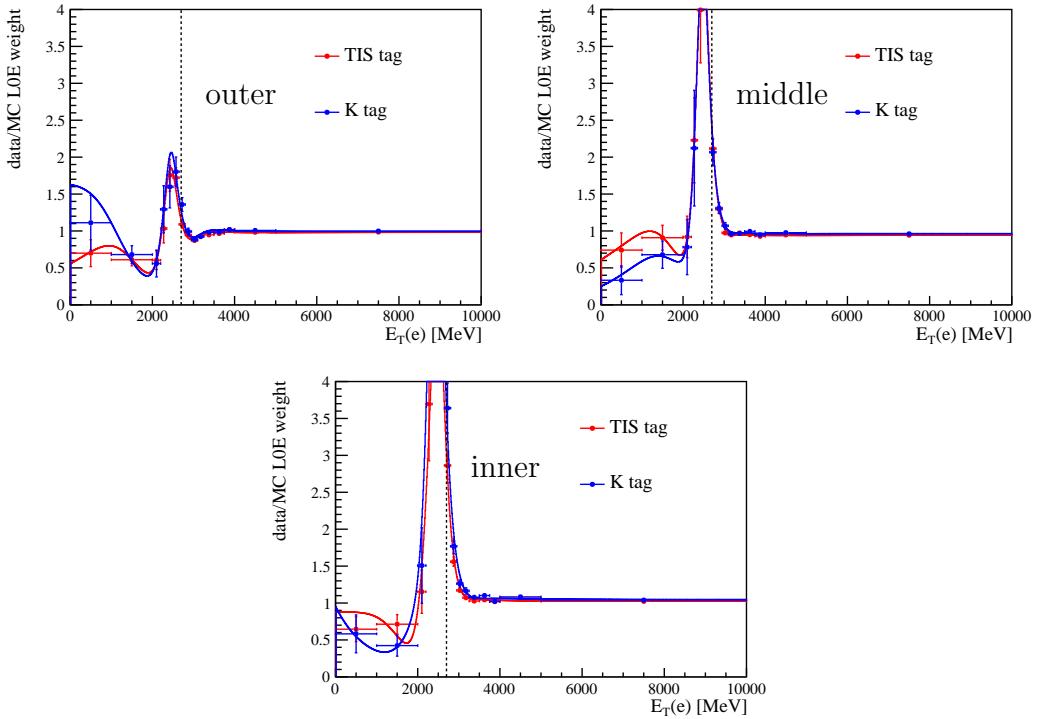


Figure 58: Data over simulation L0Electron efficiency ratio as a function of the transverse energy of the electron candidate for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, in the three ECAL regions and in the 2016 data taking conditions. The ratio is shown for two different tags: TIS tag, K tag. The ratio of the functions that fit the efficiencies are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

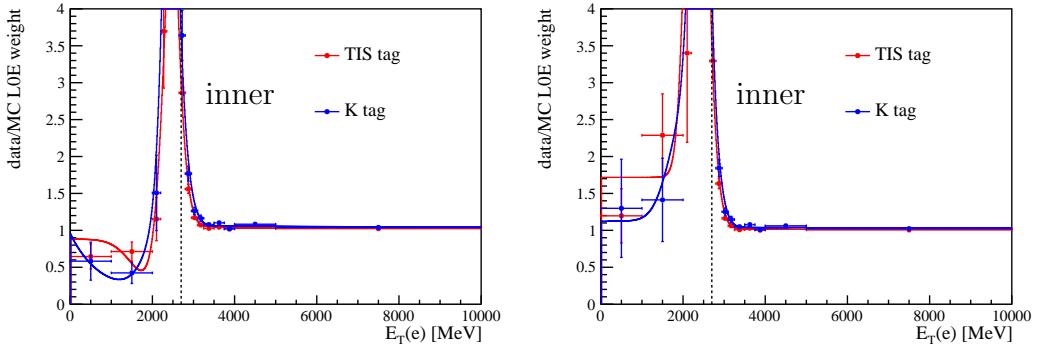


Figure 59: Data over simulation L0Electron efficiency ratio as a function of the transverse energy of the electron candidate for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, in the inner ECAL region, for (left) the default calculation and (right) after including the PID selection in the simulated sample. The ratio is shown for two different tags: TIS tag, K tag. The ratio of the functions that fit the efficiencies are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

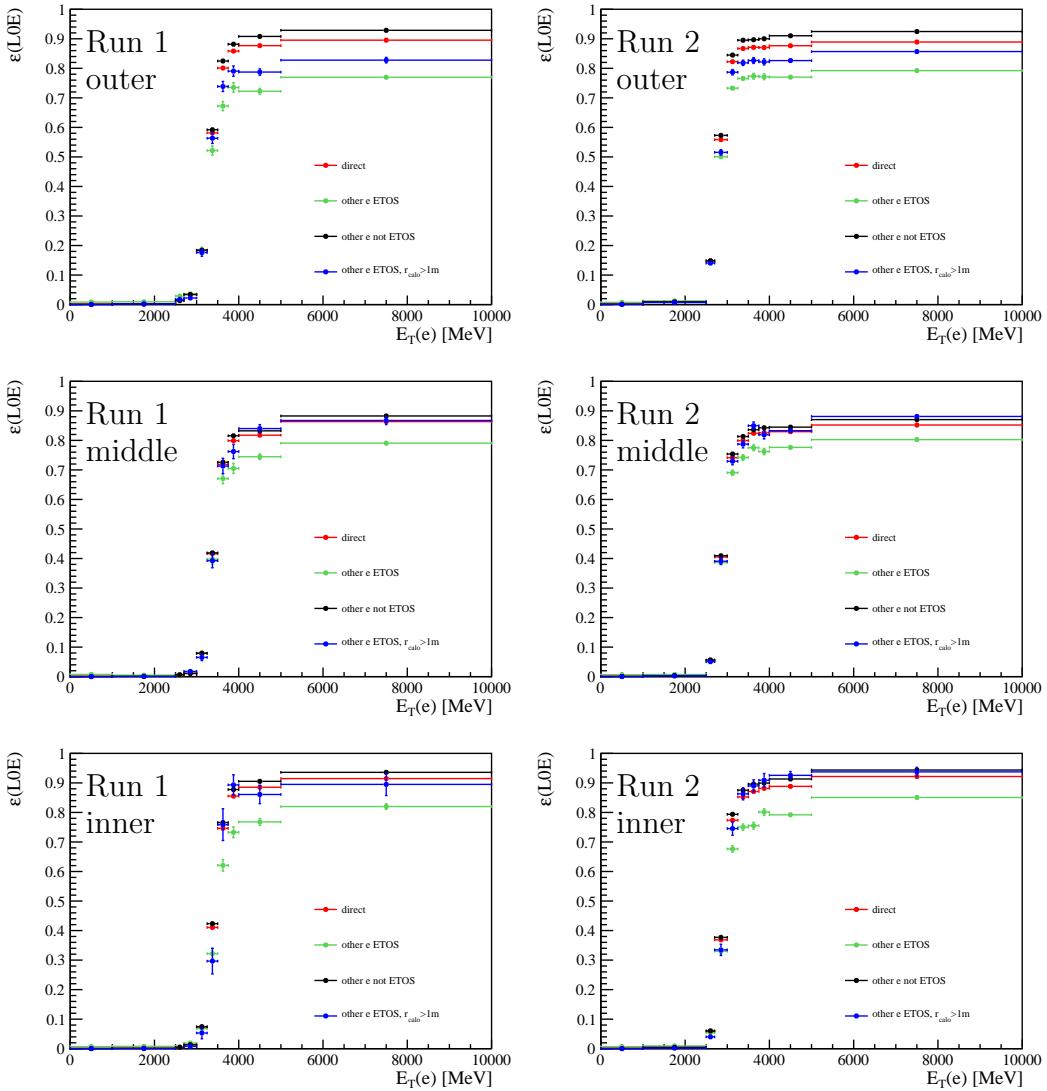


Figure 60: Efficiency of the L0Electron line as a function of the transverse energy of one of the two electron candidates for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events, in the three ECAL regions and in the Run 1 (left) and Run 2 (right) data taking conditions. The efficiency is shown in the case where no requirement is placed on the other electron (red), when the other electron is required to have fired the L0Electron trigger (green), when the other electron is required not to have fired the L0Electron trigger (black) and when the other electron has fired but is more than 1 m away from the electron probe.

1125 **6.6.3 μ TOS: calibration histograms**

1126 In a similar way to the L0Electron efficiency, the efficiency of a muon μ satisfying the
 1127 L0Muon line requirement, $\varepsilon^{\text{L}0\mu}(\mu)$ is measured on simulation and on data. The efficiency is
 1128 parameterised as a function of the transverse momentum of the muon.

1129 The L0Muon efficiency measured on simulated $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events is shown
 1130 in Fig. 61 for each year of data taking, as a function of p_{T} . The trigger fiducial cut
 1131 introduced in Tab. 7 is shown with a dotted black line. The efficiency is measured on
 1132 events selected in four different ways:

- 1133 • First, the efficiency is measured on events to which the offline preselection and PID
 1134 requirements are applied. This is labelled as “direct” in Fig. 61.
- 1135 • Second, the efficiency is measured on events that satisfy the offline preselection,
 1136 PID requirements, HLT requirements and that have been triggered on the L0 level
 1137 independently of the signal B (in practice, by requiring `B_L0Global_TIS`). This is
 1138 labelled as “TIS tag” in Fig. 61;
- 1139 • Third, the efficiency is measured on events that satisfy the offline preselection, PID
 1140 requirements, HLT requirements and that have been triggered by the other muon in
 1141 the decay. This is labelled as “ μ tag” in Fig. 61;
- 1142 • Fourth and last, the efficiency is measured on events that satisfy the offline prese-
 1143 lection, PID requirements, HLT requirements and that have been triggered by the
 1144 hadron firing the L0Hadron line. This is labelled as “ K tag” in Fig. 61.

1145 The TIS tag is used for the nominal trigger corrections. The tag bias, is very small
 1146 as demonstrated by the fact that the efficiencies measured with any tag are very close to
 1147 the “direct” efficiencies. The systematic uncertainties that tag bias introduces is discussed
 1148 in Section 10.4.

1149 The L0Muon efficiency measured on $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ data events is shown in Fig. 62
 1150 for each year of data taking, for the TIS, μ and K tags.

1151 The data and simulation plots are each fitted with a sum of two error functions with
 1152 four free parameters:

$$f \cdot \frac{1}{2} \cdot \left(1 + \text{erf} \left(\frac{E_{\text{T}} - t}{\sqrt{2}\sigma_1} \right) \right) + (1 - f) \cdot \frac{1}{2} \cdot \left(1 + \text{erf} \left(\frac{E_{\text{T}} - t}{\sqrt{2}\sigma_2} \right) \right), \quad (21)$$

1153 where t is the threshold, σ_1 and σ_2 the resolution of each error function and f the fraction
 1154 of each error function. Using a single error function with one resolution parameter only
 1155 does not yield a satisfactory fit. This might be due to the fact that the resolution of the
 1156 muon system depends on the transverse coordinate, i.e. better closer to the beam pipe
 1157 where the GEM detectors are used. The fit function is shown as a solid line on top of all
 1158 trigger calibration histograms. The fit is not made below $p_{\text{T}}(\mu) < 800$ MeV because there
 1159 are no events there due to the offline selection.

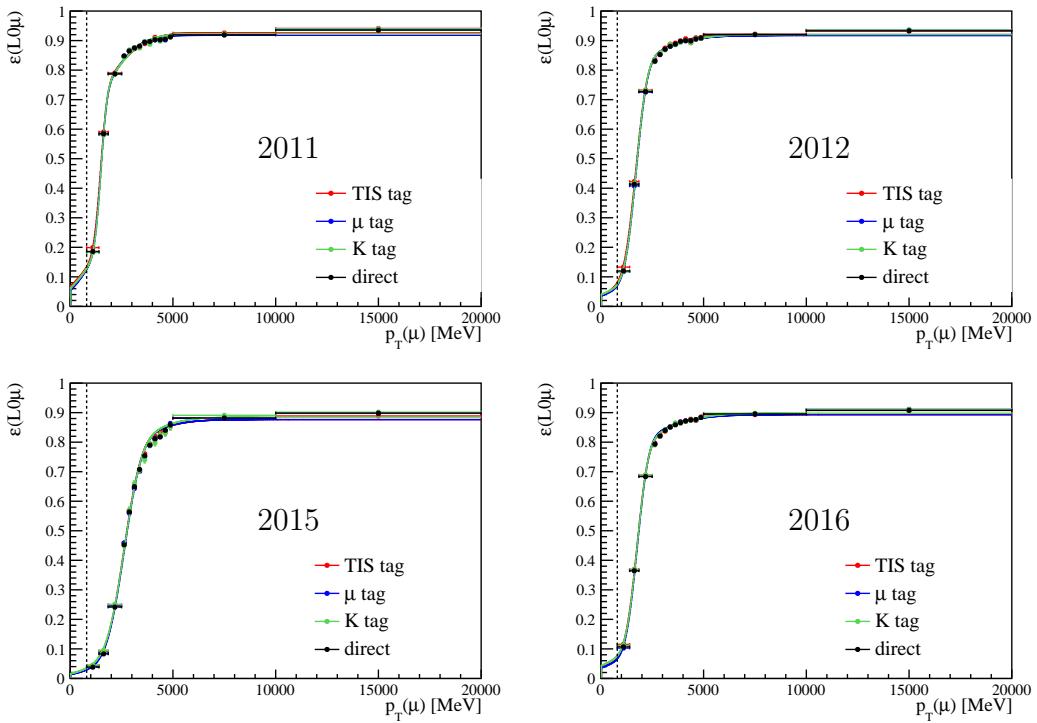


Figure 61: Efficiency of the LOMuon line as a function of the transverse momentum of the muon candidate for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ simulated events, for each year of data taking. The efficiencies are shown for four different tags: direct (no tag) TIS tag, other μ tag and K tag. The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

The ratio of data over simulation LOMuon efficiencies are shown in Fig. 63. The simulation agrees with the data within a few percent.

When computing the LOMuon efficiencies in Run 2, there is an extra possible source of bias, hereafter referred to as TCK bias, which is related to the requirement placed on p_T^{L0} , introduced in Tab. 7. Indeed, as there is no matching between the p_T^{L0} measured by the trigger system and the muon candidates, it must be ensured, when computing the trigger histograms, that the requirements placed on p_T^{L0} relates to the transverse momentum of the probe and not to any other muon in the event. There are two possible ways to do this:

- Either using the subsets of TCK's where the p_T^{L0} threshold does not need to be tightened (the TCK's requiring $p_T^{L0} > 37$ ADC in 2016 and $p_T^{L0} > 57$ ADC in 2015, see Tab. 9);
- Or using events for which there is only a single muon, the probe, that fires the LOMuon trigger, hence ensuring that the p_T^{L0} recorded by the trigger is related to the probe.

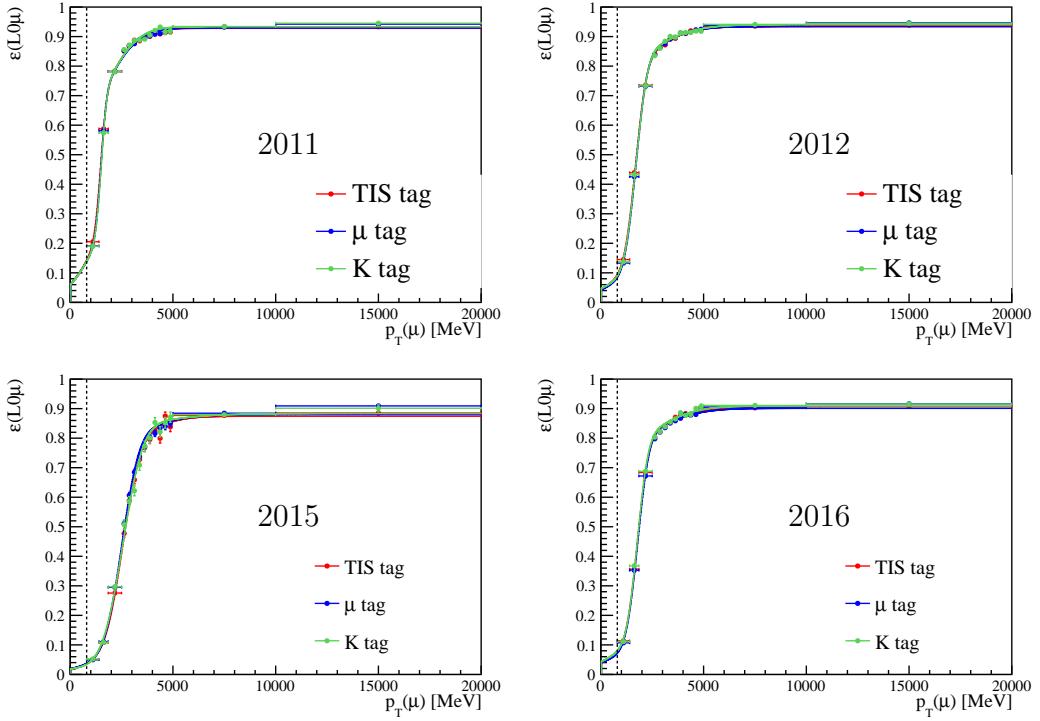


Figure 62: Efficiency of the L0Muon line as a function of the transverse momentum of the muon candidate for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ data events, for each year of data taking. The efficiencies are shown for four different tags: direct (no tag) TIS tag, other μ tag and K tag. The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

1174 The first way is chosen as the nominal method. The L0Muon weights computed both
 1175 ways are shown in Fig. 64, showing a good agreement. The systematic uncertainties related
 1176 to the TCK bias is discussed in Section 10.4.

1177 6.6.4 μ TOS: efficiency combination

1178 Using the L0Muon efficiencies measured on simulation and data, the μ TOS correction
 1179 weights read:

$$w^{\mu\text{TOS}} = \frac{1 - (1 - \varepsilon_{\text{data}}^{\text{L}0\mu}(\mu^+)) \cdot (1 - \varepsilon_{\text{data}}^{\text{L}0\mu}(\mu^-))}{1 - (1 - \varepsilon_{\text{sim}}^{\text{L}0\mu}(\mu^+)) \cdot (1 - \varepsilon_{\text{sim}}^{\text{L}0\mu}(\mu^-))}. \quad (22)$$

1180 As in the electron case, this formula holds only if $\varepsilon_{\text{data}}^{\text{L}0\mu}(\mu^+)$ and $\varepsilon_{\text{data}}^{\text{L}0\mu}(\mu^-)$ are independent
 1181 of each other. This is checked by measuring the L0Muon efficiency on one of the two muons
 1182 from simulated $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events requiring the other muon to be L0Muon TOS
 1183 or not. The result can be seen in Fig. 65 and shows that the L0Muon efficiency of the probe
 1184 is not dependent on whether the other muon fired the trigger or not.

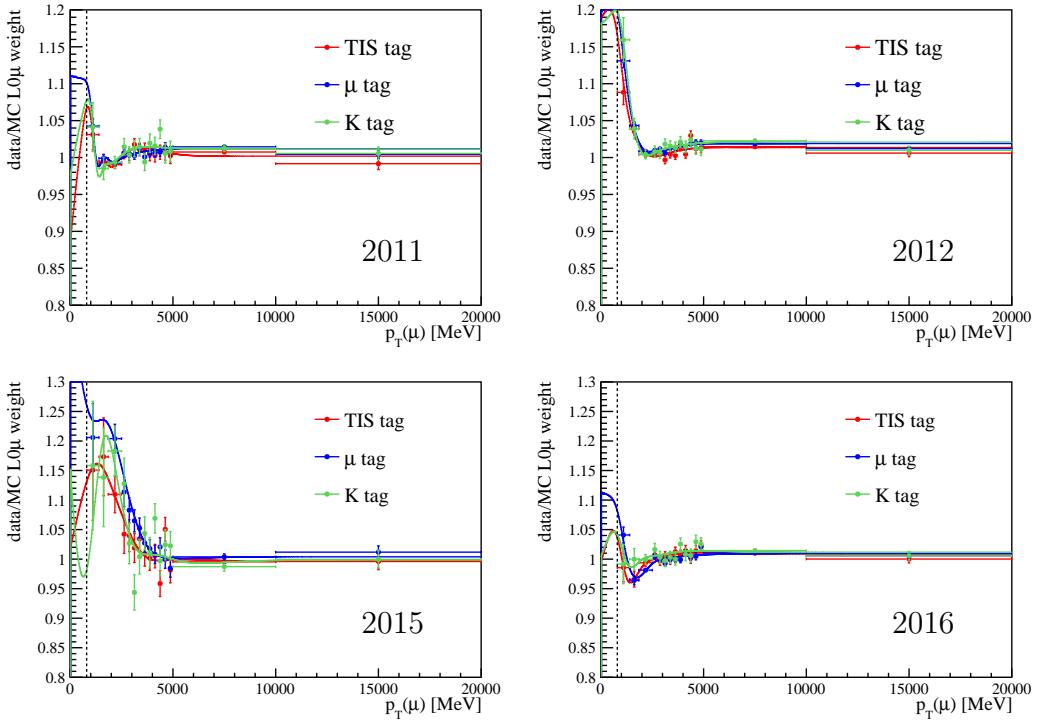


Figure 63: Ratios of data over simulation for L0Muon efficiencies as a function of the transverse momentum of the muon candidate for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events, for each year of data taking. The ratios are shown for efficiencies computed with three different tags: TIS tag, other μ tag and K tag. The fits to the data points are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

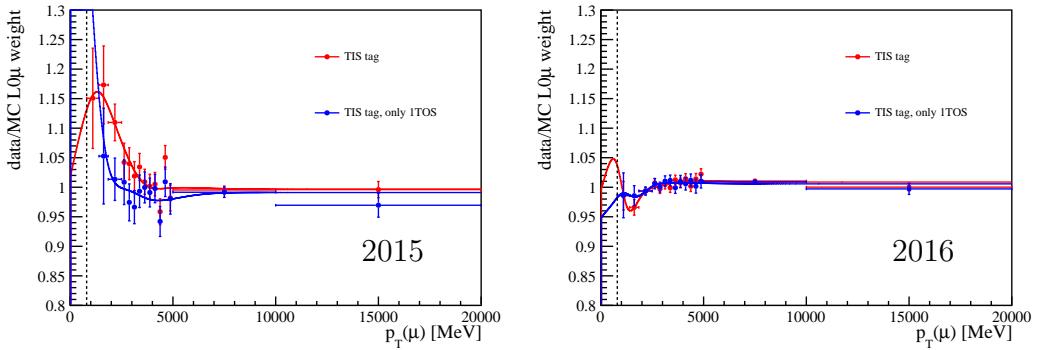


Figure 64: Ratios of data over simulation for L0Muon efficiencies as a function of the transverse momentum of the muon candidate for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events, for 2015 and 2016. The ratios are shown for efficiencies computed in two different ways: either requiring that the muon probe is the only muon firing the trigger (blue) or selecting TCK's where the L0Muon is set to 37 ADC (2016) and 57 ADC (2015). The dotted black line indicates the trigger fiducial cut.

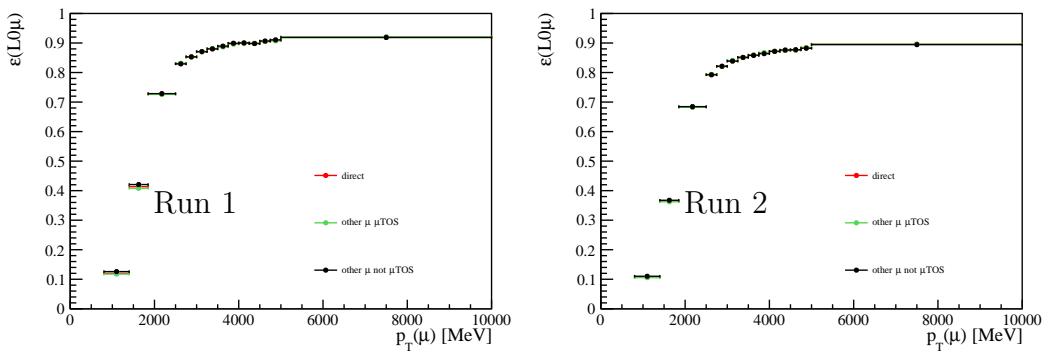


Figure 65: Efficiency of the L0Muon line as a function of the transverse energy of one of the two muon candidates for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ simulated in the Run 1 (left) and Run 2 (right) data taking conditions. The efficiency is shown in the case where no requirement is placed on the other muon (red), when the other electron is required to have fired the L0Muon trigger (green) and when the other electron is required not to have fired the L0Muon trigger (black).

1185 **6.6.5 *hTOS*: calibration histograms**

1186 In a similar way to the **L0Electron** efficiency, the efficiency of the kaon K satisfying the
1187 **L0Hadron** line requirement, $\varepsilon^{\text{L0h}}(K)$ is measured on simulation and on data. Following the
1188 same arguments as for the **L0Electron** efficiency, the **L0Hadron** efficiency is parametrised
1189 as a function of the transverse energy of the kaon and of the two HCAL regions.

1190 The **L0Hadron** efficiency measured on simulated $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow$
1191 $K^+ J/\psi(e^+e^-)$ events is shown in Fig. 66 and Fig. 67 for each year of data taking, as a
1192 function of p_T . The trigger fiducial cut introduced in Tab. 7 is shown with a dotted black
1193 line. The efficiency is measured on events selected in five different ways:

- 1194 • First, the efficiency is measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events to which the offline
1195 preselection and PID requirements are applied, labelled as “direct” in Fig. 66 and
1196 Fig. 67;
- 1197 • Second, the efficiency is measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events that satisfy the
1198 offline preselection, PID and HLT requirements and that have been triggered on the
1199 **L0** level by one of the two electrons. This is labelled as “ e tag”;
- 1200 • Third, same strategy as the previous one but using $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events
1201 and the **L0Muon** trigger. This is labelled “ μ tag”;
- 1202 • Fourth, the efficiency is measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events that satisfy the
1203 offline preselection, PID and HLT requirements and that have been triggered on the
1204 **L0** level independently of the signal B (in practice, by requiring `B_L0Global_TIS`).
1205 This is labelled “ ee TIS tag”;
- 1206 • Fifth, same strategy as the previous ones but using $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events.
1207 This is labelled as “ $\mu\mu$ TIS tag”.

1208 The μ tag is used as the nominal tag for the hadron trigger corrections, because it is
1209 the one with the highest yield. The efficiencies given by each tag are very close to each
1210 other, and the tag bias is therefore expected to be small. The systematic uncertainty
1211 related to it is discussed in Section 10.4.

1212 The **L0Hadron** efficiencies measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$
1213 data events are shown in Fig. 68 and Fig. 69 for each year of data taking and for the four
1214 different tags.

1215 The data and simulation histograms are fit with the same function that was used for
1216 the **L0Electron** efficiency, Eq. 19. The fit is performed in order to reduce the uncertainty
1217 due to the limited size of the trigger calibration samples. The fit function is shown as a
1218 solid line on top of all trigger calibration histograms.

1219 The ratio of data over simulation **L0Hadron** efficiencies are shown in Fig. 70 and Fig. 71.
1220 As is the case for the **L0Electron** efficiencies, the trigger fiducial cut, shown with a black
1221 dotted line, vetoes the low transverse energy region where the statistics are low and the
1222 agreement between data and simulation is worst.

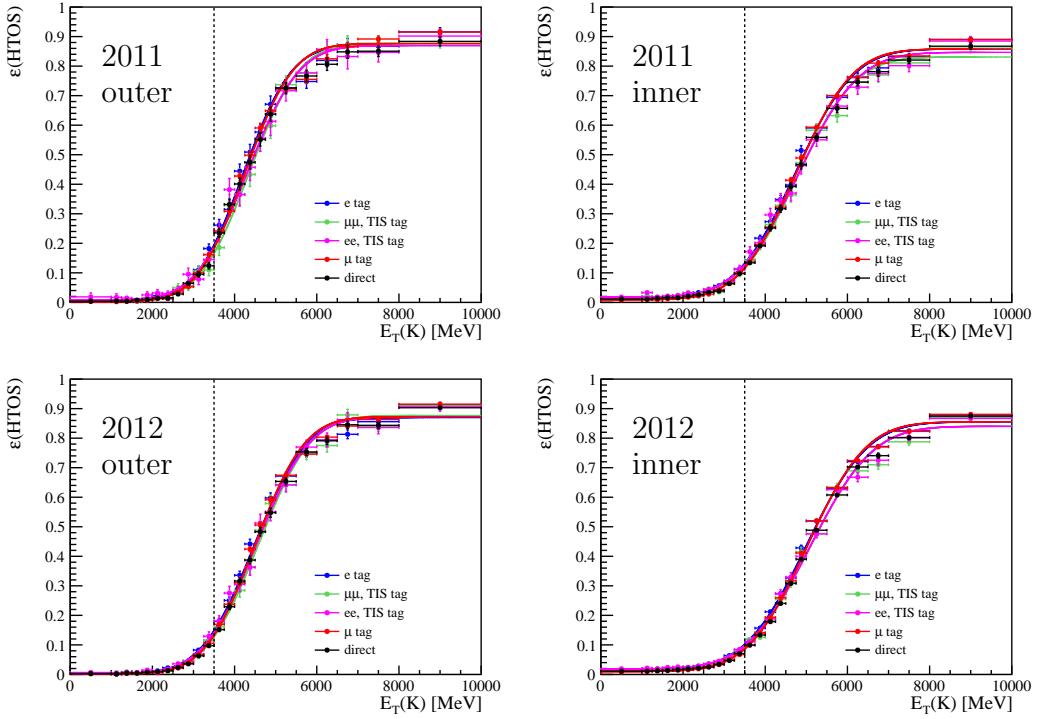


Figure 66: Efficiency of the L0Hadron trigger as a function of the transverse energy of the kaon candidate for $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ simulated events, in the two HCAL regions (left and right) and in the 2011 and 2012 data taking conditions (top and bottom). The efficiency is shown using five different tags: TIS tag and lepton tag on the electron and muon modes, and without tag on the electron mode. The fit functions are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

6.6.6 $h\text{TOS}$: efficiency combination

The weight applied to simulated events to correct the trigger L0 efficiency in the $h\text{TOS}$ category is computed:

$$w^{h\text{TOS}} = \frac{\varepsilon_{\text{data}}^{\text{L0h}}(K^+)}{\varepsilon_{\text{sim}}^{\text{L0h}}(K^+)},$$

and the weight to correct the L0 efficiency in the $h\text{TOS!}$ category is computed as:

$$w^{h\text{TOS!}} = \frac{\varepsilon_{\text{data}}^{\text{L0h}}(K^+)}{\varepsilon_{\text{sim}}^{\text{L0h}}(K^+)} \cdot \frac{(1 - \varepsilon_{\text{data}}^{\text{L0e}}(e^+)) \cdot (1 - \varepsilon_{\text{data}}^{\text{L0e}}(e^-))}{(1 - \varepsilon_{\text{sim}}^{\text{L0e}}(e^+)) \cdot (1 - \varepsilon_{\text{sim}}^{\text{L0e}}(e^-))}.$$

It is checked on simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events that the efficiency of the L0Hadron trigger line does not depend on whether one of the electrons has passed the L0Electron requirement, as can be seen in Fig. 72

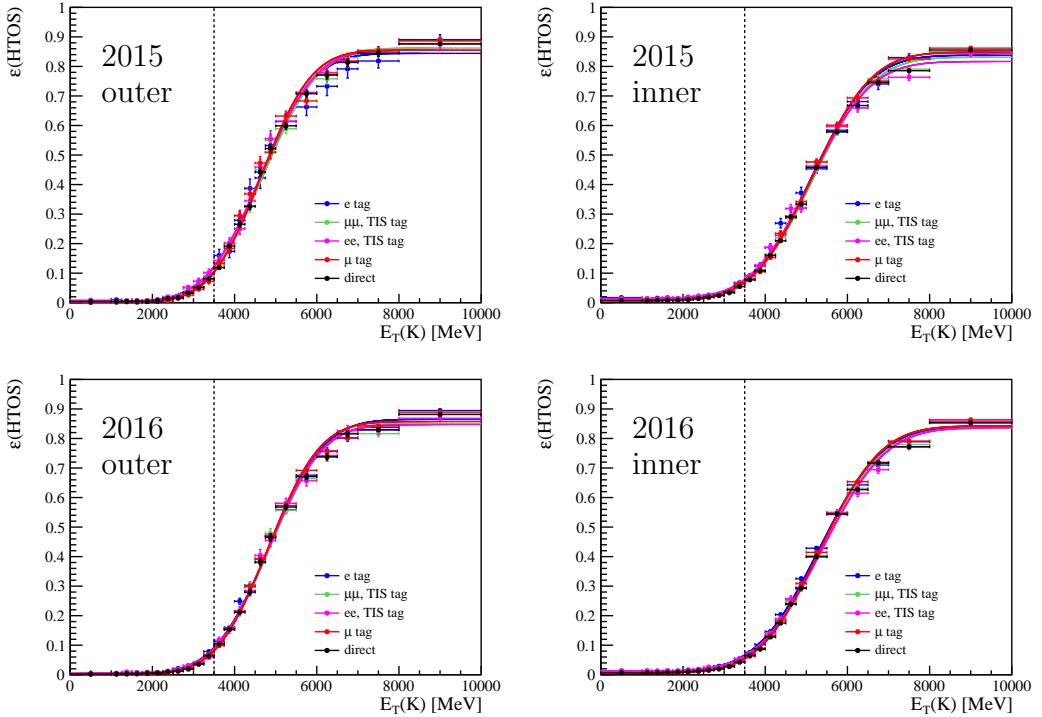


Figure 67: Efficiency of the L0Hadron trigger as a function of the transverse energy of the kaon candidate for $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ simulated events, in the two HCAL regions (left and right) and in the 2015 and 2016 data taking conditions (top and bottom). The efficiency is shown using five different tags: TIS tag and lepton tag on the electron and muon modes, and without tag on the electron mode. The fit functions are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

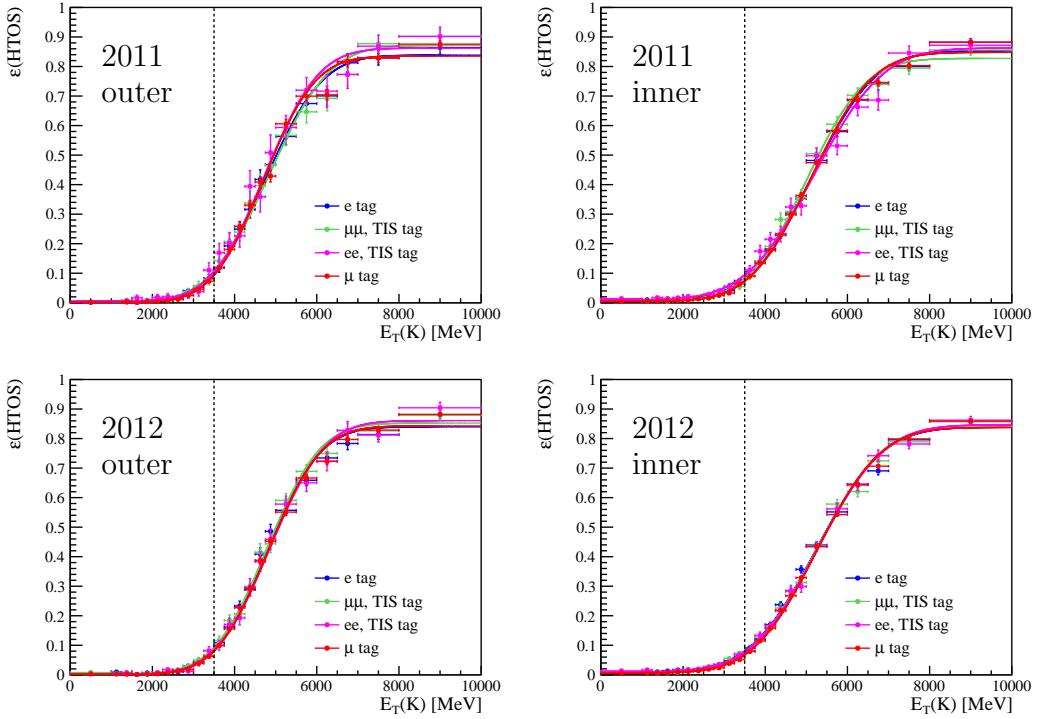


Figure 68: Efficiency of the L0Hadron trigger as a function of the transverse energy of the kaon candidate for $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ data events, in the two HCAL regions (left and right) and in the 2011 and 2012 data taking conditions (top and bottom). The efficiency is shown using four different tags: TIS tag and lepton tag on the electron and muon modes. The fit functions are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

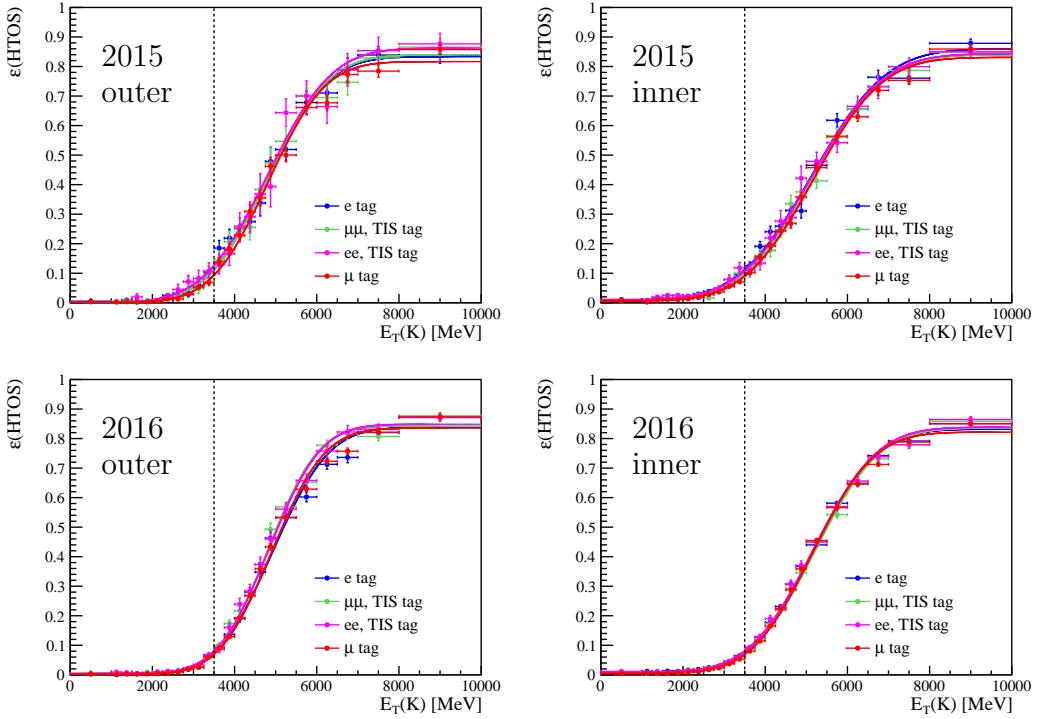


Figure 69: Efficiency of the L0Hadron trigger as a function of the transverse energy of the kaon candidate for $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ data events, in the two HCAL regions (left and right) and in the 2015 and 2016 data taking conditions (top and bottom). The efficiency is shown using four different tags: TIS tag and lepton tag on the electron and muon modes. The fit functions are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

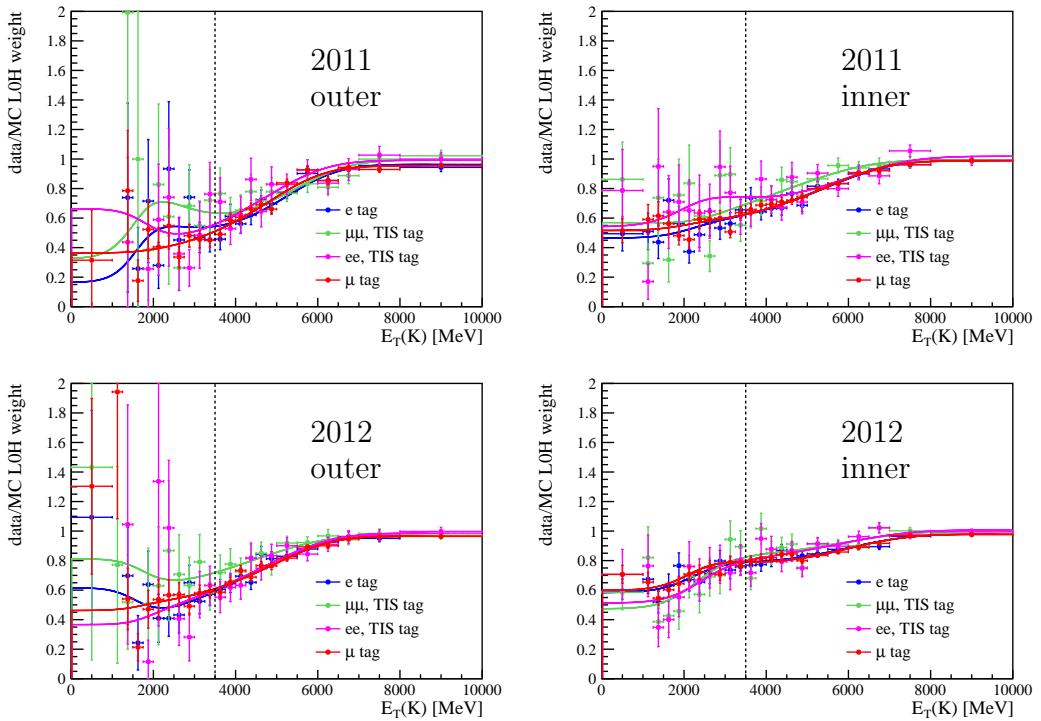


Figure 70: Data over simulation ratios of efficiencies of the L0Hadron trigger, as a function of the transverse energy of the kaon candidate from $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ events, in the two HCAL regions (left and right) and in the 2011 and 2012 data taking conditions (top and bottom). The ratios are shown using four different tags: TIS tag and lepton tag on the electron and muon modes. The fit functions to the efficiencies are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

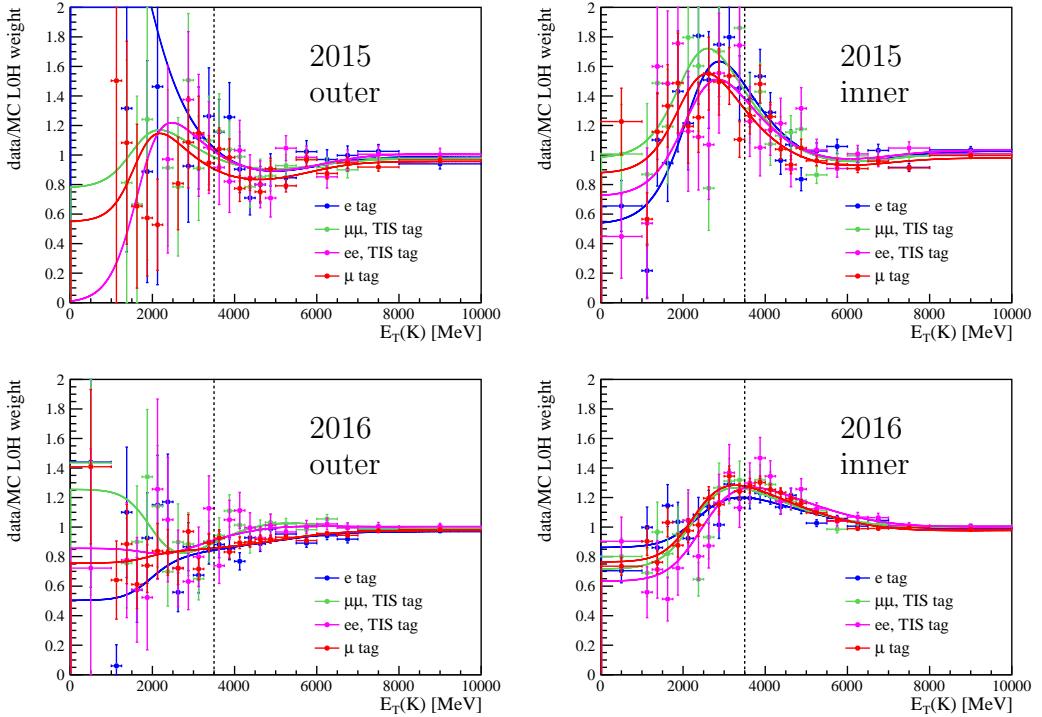


Figure 71: Data over simulation ratios of efficiencies of the L0Hadron trigger, as a function of the transverse energy of the kaon candidate from $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ events, in the two HCAL regions (left and right) and in the 2015 and 2016 data taking conditions (top and bottom). The ratios are shown using four different tags: TIS tag and lepton tag on the electron and muon modes. The fit functions to the efficiencies are shown with a solid line. The dotted black line indicates the trigger fiducial cut.

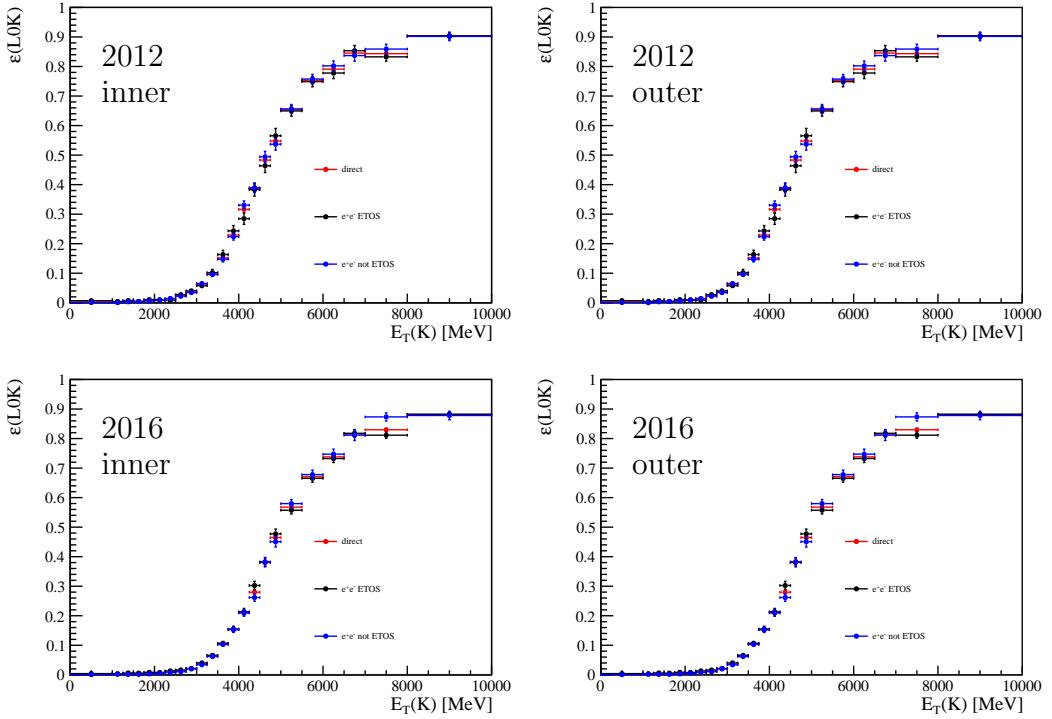


Figure 72: Efficiency of the L0Hadron line as a function of the transverse energy of the kaon for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events, in the two HCAL regions and in the Run 1 (top) and Run 2 (bottom) data taking conditions. The efficiency is shown in the case where no requirement is placed on the electrons (red), when one of the electron is required to have fired the L0Electron trigger (black) and when none of the two electron has fired the L0Electron trigger (blue).

1230 **6.6.7 TIS: calibration histograms**

1231 The efficiency of the decay satisfying the TIS requirement is measured on simulation and
1232 on data as a function of the transverse momentum of the B , $p_T(B)$, as it is expected to be
1233 correlated to the transverse momentum of the system resulting from the hadronisation of
1234 the opposite b -quark, which directly impacts on the TIS efficiency.

1235 Most of the TIS events are triggered by a hadron (approximately 55% of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and 60% of $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ TIS events), and the TIS efficiency
1236 depends on whether the kaon from the signal decay is TOS or not. As one of the trigger
1237 categories used to extract R_K is the exclusive TIS! category, the TIS calibration histograms
1238 are extracted only for events where the kaon did not fire the hadron trigger.

1240 The L0TIS efficiency measured on simulated $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow$
1241 $K^+ J/\psi(e^+e^-)$ events is shown in Fig. 73 for each year of data taking.

1242 The efficiency is measured on events selected in four different ways:

- 1243 • First, the efficiency is measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events to which the offline
1244 preselection and PID requirements are applied, and where the kaon is required to
1245 not have fired the hadron trigger, labelled as “ee, direct” in Fig. 73;
- 1246 • Second, same as the previous point but using $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events, labelled
1247 as “ $\mu\mu$, direct” in Fig. 73;
- 1248 • Third, the efficiency is measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events that satisfy the
1249 offline preselection, PID and HLT requirements, where the kaon is required to not
1250 have fired the hadron trigger, and that have been triggered on the L0 level by one of
1251 the two electrons, labelled as “ee, e tag” in Fig. 73;
- 1252 • Fourth, same strategy as the previous one but using $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events
1253 and the L0Muon trigger. This is labelled “ $\mu\mu$, μ tag”.

1254 The e tag is used as the nominal tag for the TIS trigger corrections in the electron
1255 mode. However, it is clear from Fig. 73 that the tag bias in the electron mode is large,
1256 and that the muon mode cannot be used to correct the TIS efficiencies in the electron
1257 mode. It has been checked that this bias does not disappear if the TIS trigger efficiency is
1258 measured in bins of nSPDHits. Hence, the systematic uncertainty associated to the TIS
1259 trigger efficiency is expected to be large and is discussed in Section 10.4.

1260 The ratios of data over simulation TIS efficiencies are shown in Fig. 74. They are
1261 significantly displaced from unity, especially at low p_T , resulting in large correction weights.
1262 Moreover, the weights obtained from the muon modes and those obtained from the
1263 electron modes are very different. The difference comes from events that have been
1264 triggered by the L0 photon or electron trigger lines; they constitute approximately 35% of
1265 $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events TIS events, and 45% of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events. This
1266 seems to indicate that the hits in the ECAL that triggered these events actually originate
1267 from the signal electrons.

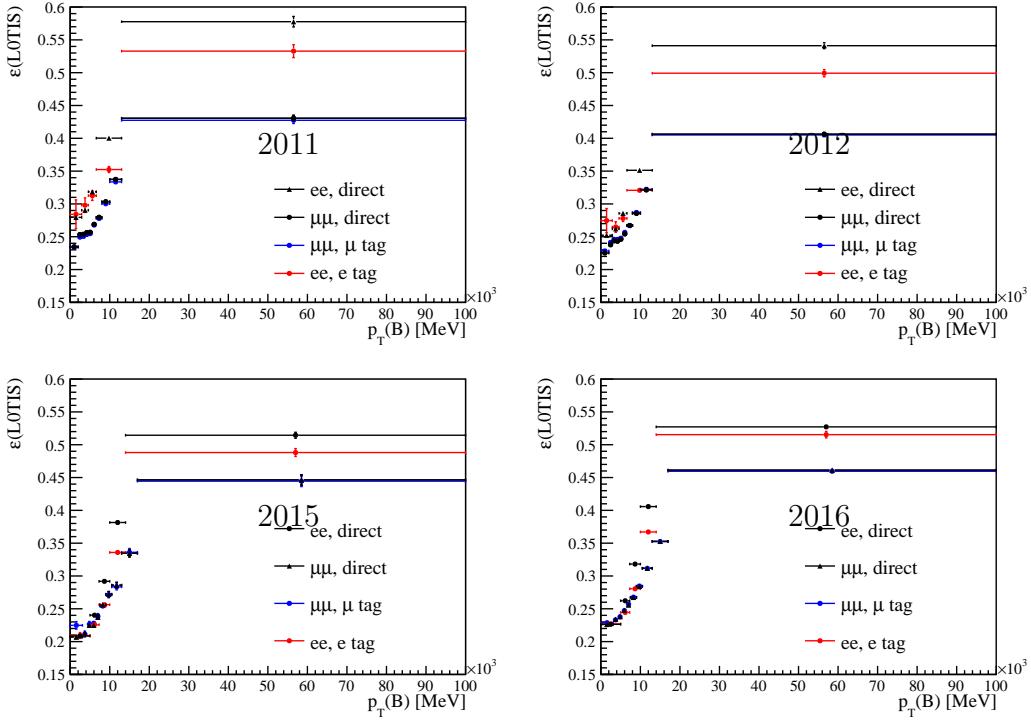


Figure 73: Efficiency of the L0 TIS requirement as a function of the transverse momentum of the B^+ candidate for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events, for each year of data taking. The efficiencies are shown for four different categories: direct (no tag) on both the electron and muon modes, *e* L0Electron tag and *mu* L0Muon tag.

1268 6.6.8 TIS: efficiency combination

1269 The weight applied to simulated events to correct the trigger L0 efficiency in the TIS!
 1270 category is computed as:

$$w^{\text{TIS!}} = \frac{\varepsilon_{\text{data}}^{\text{L0TIS}}(B^+)}{\varepsilon_{\text{sim}}^{\text{L0TIS}}(B^+)} \cdot \frac{(1 - \varepsilon_{\text{data}}^{\text{L0e}}(e^+)) \cdot (1 - \varepsilon_{\text{data}}^{\text{L0e}}(e^-))}{(1 - \varepsilon_{\text{sim}}^{\text{L0e}}(e^+)) \cdot (1 - \varepsilon_{\text{sim}}^{\text{L0e}}(e^-))} \cdot \frac{1 - \varepsilon_{\text{data}}^{\text{L0h}}(K^+)}{1 - \varepsilon_{\text{sim}}^{\text{L0h}}(K^+)}$$

1271 As the trigger calibration histograms are computed after requiring the kaon to not
 1272 have fired the hadron trigger, the above formula correctly takes into account the bias that
 1273 the hadron TOS requirement induces on the TIS efficiency. However, this formula also
 1274 relies on the fact that the TIS efficiency is independent of whether one of the electrons
 1275 has fired the trigger. As can be expected from the tag bias already discussed this is not
 1276 exactly the case, as can be seen in Fig. 75. The systematic uncertainty related to this is
 1277 discussed in Section 10.4.

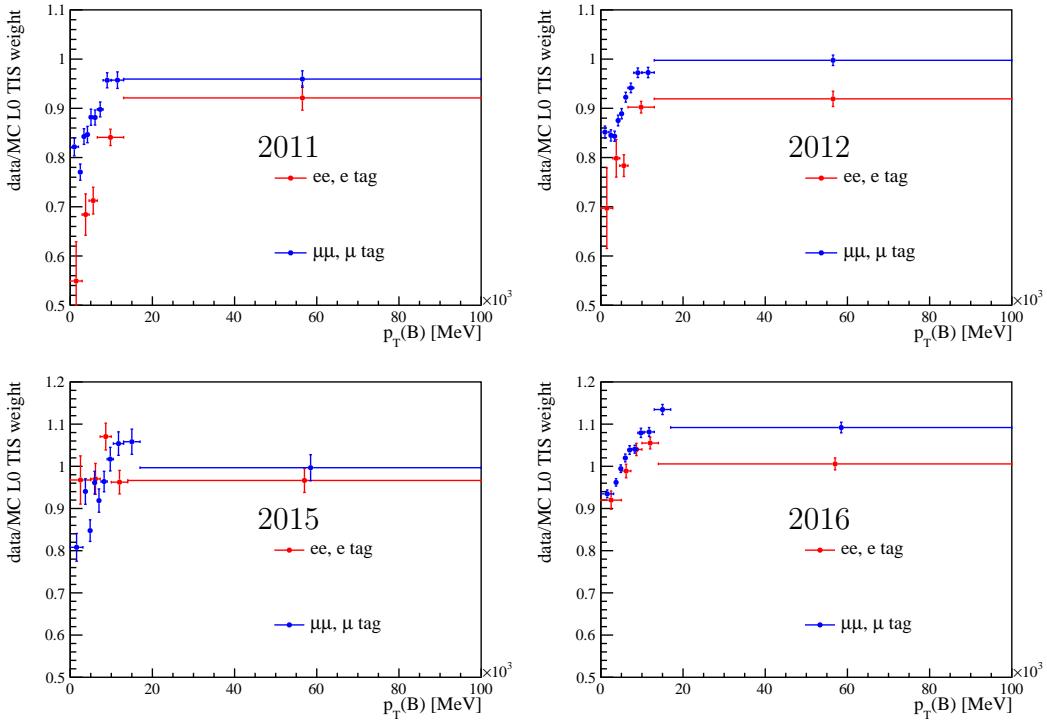


Figure 74: Ratios of data over simulation for L0 TIS efficiencies as a function of the transverse momentum of the B^+ candidate for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, for each year of data taking. The ratios are shown for efficiencies computed in two different ways: $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events using a e L0Electron tag and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events using a μ L0Muon tag.

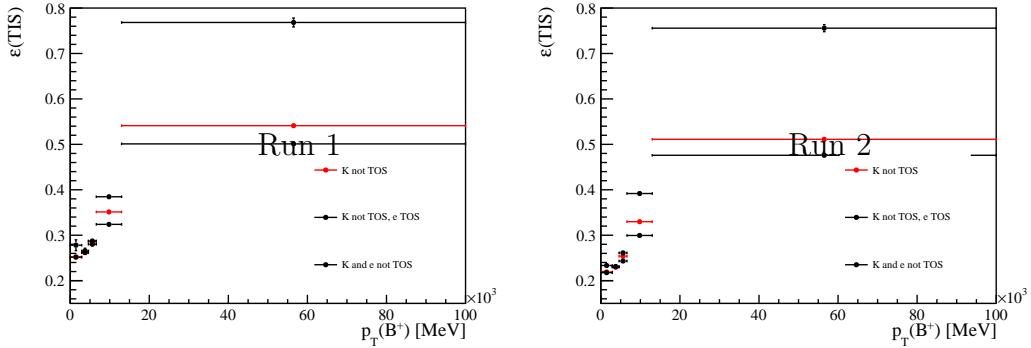


Figure 75: Efficiency of the L0 TIS requirement as a function of the transverse energy of the B^+ candidates for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ simulated in the Run 1 (left) and Run 2 (right) data taking conditions. The efficiency is shown in the case where no requirement is placed on the electrons (red), when one of the electrons is required to have fired the L0 trigger (blue) and when the electrons are required not to have fired the L0 trigger (black).

1278 **6.6.9 Hidden variables in L0 efficiency parametrisation**

1279 All the trigger correction weights have been parametrised as a function of one or two
 1280 variables. It could however happen that the trigger efficiencies depend on one or several
 1281 extra variables. If the distributions of one of these hypothetical extra variables differ
 1282 between the control modes and the rare modes, the ratio of trigger efficiencies $\varepsilon^{\text{trig}}(B^+ \rightarrow$
 1283 $K^+\ell^+\ell^-)/\varepsilon^{\text{trig}}(B^+ \rightarrow K^+J/\psi(\ell^+\ell^-))$ will be biased. The size of the bias is extracted
 1284 by computing the L0 trigger efficiency histograms using $B^+ \rightarrow K^+J/\psi(\ell^+\ell^-)$, $B^+ \rightarrow$
 1285 $K^+\psi(2S)(\ell^+\ell^-)$ and $B^+ \rightarrow K^+\ell^+\ell^-$ in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$. For all categories,
 1286 the trigger histograms computed with the three channels are not significantly different,
 1287 showing that the number of variables in the parametrisation is sufficient and that there
 1288 is no significant bias in these cases. For each L0 trigger line, one efficiency histogram is
 1289 shown in Fig. 76, in the Run 2 data taking conditions and for the three different channels.

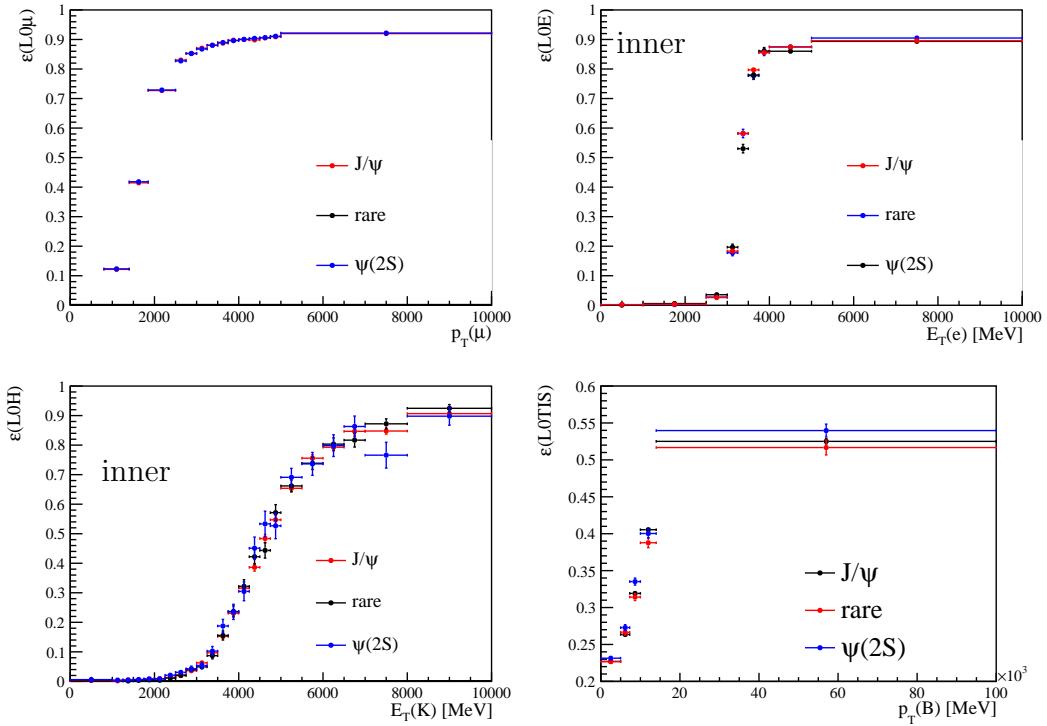


Figure 76: Efficiency of the L0Muon (top left), L0Electron (top right), L0Hadron (bottom left) and L0TIS requirements for offline selected $B^+ \rightarrow K^+J/\psi(\ell^+\ell^-)$, $B^+ \rightarrow K^+\psi(2S)(\ell^+\ell^-)$ and $B^+ \rightarrow K^+\ell^+\ell^-$ simulated events in the Run 2 data taking conditions. The efficiencies computed in the three channels are statistically compatible in the three TOS cases but not in the TIS case.

1290 **6.7 HLT efficiencies calibration**

1291 The HLT calibration histograms are computed in *super-*
1292 *Lenin/src/superMain10b(Run2).cc*, which mostly uses the class in *super-*
1293 *Lenin/src/src/hltCorrector.cpp*.

1294 The efficiency of the HLT1 and HLT2 requirements is measured on $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$
1295 data and simulated events as a function of the transverse momentum of the B^+ candidate.
1296 The HLT efficiencies are computed using offline selected events that are TIS on both
1297 the HLT1 and HLT2 levels. At the L0 level, in the muon case, the events are required
1298 to pass the μ TOS requirements, and in the electron case, HLT efficiencies are extracted
1299 in three different cases depending on whether the events pass the e TOS, h TOS or TIS
1300 requirements.

1301 To extract the HLT efficiencies on data, the *sWeight* method cannot be used because
1302 the level of background is too high for events selected with an HLT1 and HLT2 TIS
1303 requirement. Instead, the HLT efficiencies are computed using the fit & count method.

1304 The ratio of HLT efficiencies between data and simulation, for each year of data
1305 taking, is shown in Fig. 77 for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ μ TOS events and in Fig. 78 for
1306 $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ e TOS, h TOS and TIS events.

1307 As expected, because of the Run 2 HLT fiducial cuts discussed in Section 5.5.1, the
1308 agreement between data and simulation regarding the HLT efficiencies is better in Run 2
1309 than in Run 1. The ratio of efficiencies shown in Fig. 77 and in Fig. 78 are therefore used
1310 as a nominal correction in the Run 1 efficiency computation, but not in Run 2.

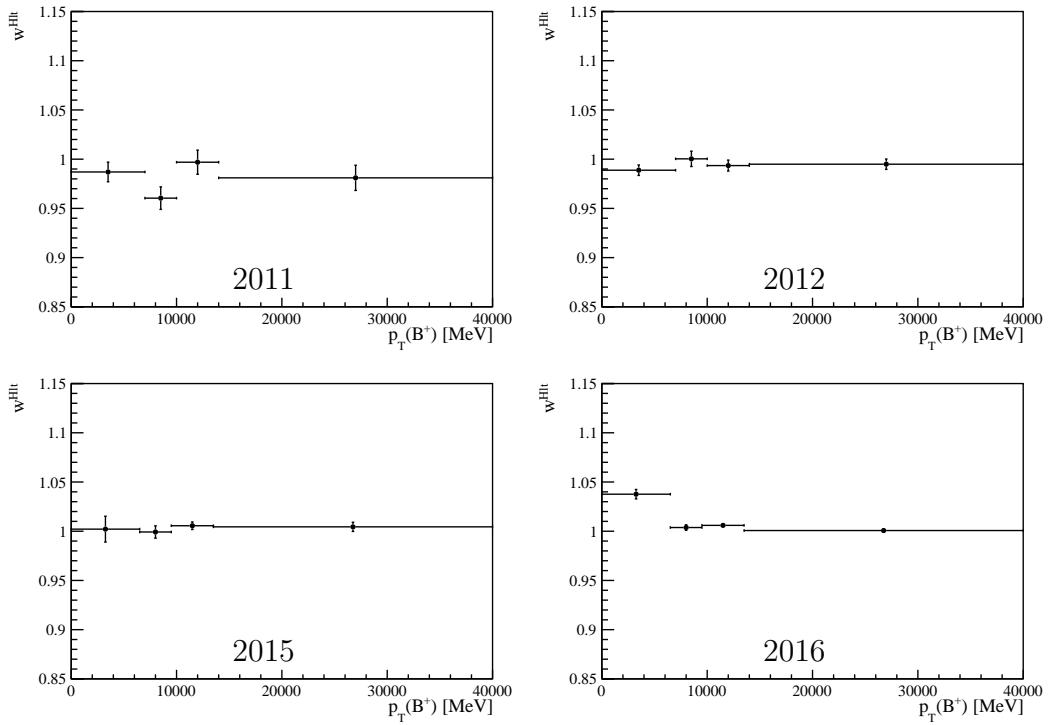


Figure 77: Data over simulation ratios of HLT trigger efficiencies as a function of the transverse momentum of the B^+ candidate. The efficiencies are measured on $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events that pass the offline selection requirements, that pass the μTOS requirement at the L0 level and that are TIS on the HLT1 and HLT2 levels.

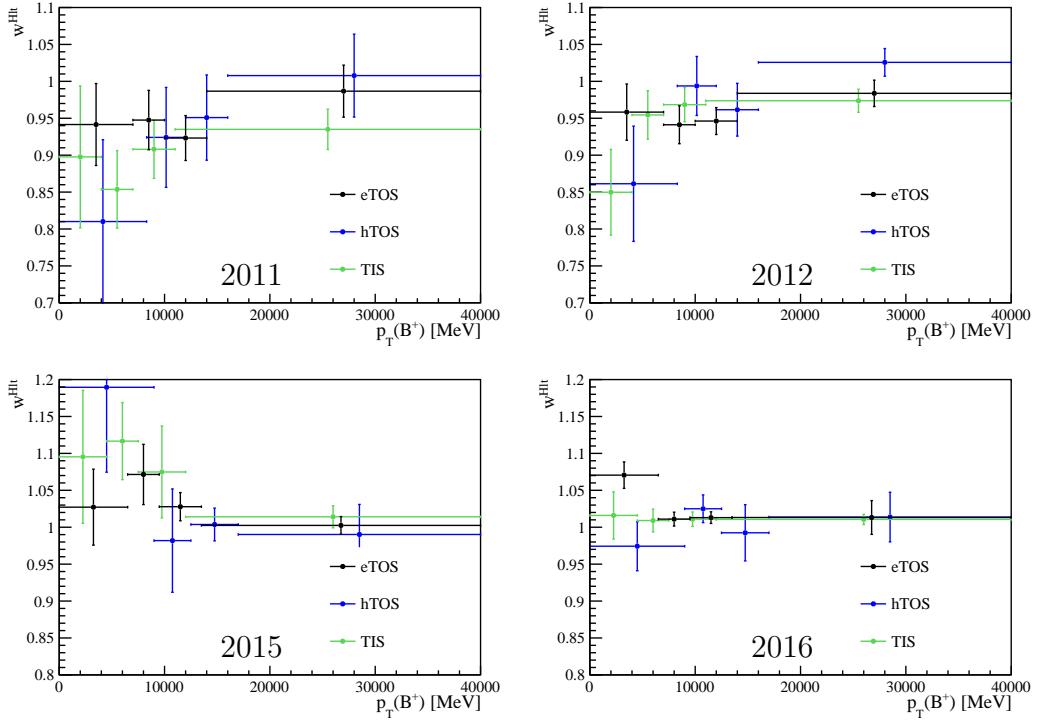


Figure 78: Data over simulation ratios of HLT trigger efficiencies as a function of the transverse momentum of the B^+ candidate. The efficiencies are measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events that pass the offline selection requirements, that are TIS on the HLT1 and HLT2 levels, and that belong to either $e\text{TOS}$ (black) $h\text{TOS}$ (blue) or TIS (green) L0 categories.

1311 **6.8 Kinematic corrections**

1312 The kinematic corrections are computed and applied to the simulation samples
1313 in *superLenin/src/superMain12(Run2).12..* The functions that perform 1D and
1314 2D iterative reweighting based on ratios of histograms are written in *super-*
1315 *Lenin/src/reweightingTools.cpp*.

1316 The B^+ momentum and rapidity spectra are generated imperfectly in the simulation.
1317 Moreover, the reconstructed χ_{DV}^2 and $\chi_{\text{IP}}^2(B^+)$ distributions are also badly modelled in
1318 the simulation. The simulation is therefore corrected using kinematic weights, which
1319 correspond to data over simulation ratios of kinematic distributions stored in histograms
1320 Two sets of kinematic weights are used in the efficiency computation:

- 1321 1. A first set that corrects the two-dimensional ($p_{\text{T}}(B^+); \eta(B^+)$) distribution to match
1322 that in the data, in order to compute the geometric and reconstruction efficiencies,
1323 $\varepsilon^{\text{geom}}$ and $\varepsilon^{\text{rec,strip}}$;
- 1324 2. A second set that corrects the two-dimensional ($p_{\text{T}}(B^+); \eta(B^+)$) distribution, followed
1325 by the one dimensional χ_{DV}^2 and $\chi_{\text{IP}}^2(B^+)$ distributions. This is done iteratively in
1326 four loops to handle the small correlations between the three distributions. These
1327 weights are used in the computation of the rest of the efficiencies, $\varepsilon^{\text{presel}}$, ε^{PID} , $\varepsilon^{\text{trig}}$
1328 and ε^{BDT} .

1329 The samples used to extract the corrections weights are the data and simulated samples
1330 of offline selected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events in the μTOS category. This particular
1331 sample is used because it has the largest number of events, and the largest signal-to-
1332 background ratio. Moreover, it is the mode for which the PID and trigger efficiencies are
1333 best under control, as the former does not show any trigger bias (see Section 6.4.3) and the
1334 latter does not suffer from any significant tag or factorisation bias (see Section 6.6.3 and
1335 Section 6.6.4). Another advantage of using the muon mode to compute the ($p_{\text{T}}(B^+); \eta(B^+)$)
1336 correction weights is that the momentum resolution is much better for muons than for
1337 electrons. Kinematic weights computed from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode are only
1338 used for cross-checks and to compute systematics, as described in Section 10.2.

1339 The PID and L0 corrections are used to correct the simulation sample, and the *sWeight's*
1340 are used to subtract the background in the data sample. The two dimensional and one
1341 dimensional histograms contain respectively 20×10 and 150 iso-populated bins.

1342 The kinematic weights are computed separately for each year of data taking and, in
1343 Run 1, separately for simulated events generated with PYTHIA 6 and PYTHIA 8, as the
1344 B^+ kinematics differ slightly in the two versions of the software (in Run 2, all simulated
1345 samples are generated with PYTHIA 8).

1346 When the weights are applied on an electron mode, the true ($p_{\text{T}}(B^+); \eta(B^+)$) is used,
1347 as the reconstructed ($p_{\text{T}}(B^+); \eta(B^+)$) has a poor resolution due to bremsstrahlung.

1348 **6.8.1 Correction to the occupancy distribution**

1349 The occupancy of the detector is known to be poorly described in the simulation. The
 1350 resolution, tracking or trigger efficiencies are all mildly dependent on the occupancy.
 1351 Some occupancy proxy, such has nSPDHits, could be considered to be corrected as
 1352 well. However, as it can be seen in Fig. 79, correcting the nSPDHits distributions in
 1353 simulated $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ to obtain a good agreement with the data results in
 1354 poorer agreements for some other occupancy proxies. In particular, the occupancy of
 1355 the ECAL, both described in terms of number of reconstructed photons and in terms of
 1356 the sum of the transverse momenta of the reconstructed photons, shows a much better
 1357 agreement before correcting the nSPDHits distribution. This is particularly important
 1358 because the L0Electron efficiency and the bremsstrahlung recovery (and therefore the
 1359 electron momentum and q^2 resolution) depend strongly on the occupancy of the ECAL.
 1360 The efficiency of the cut on nSPDHits, i.e. nSPDHits < 450 in Run 2 and < 600 in Run 1,
 1361 cancel perfectly in the efficiency ratio $\varepsilon(B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-))/\varepsilon(B^+ \rightarrow K^+\ell^+\ell^-)$ and
 1362 does therefore not constitute an issue in the present analysis.

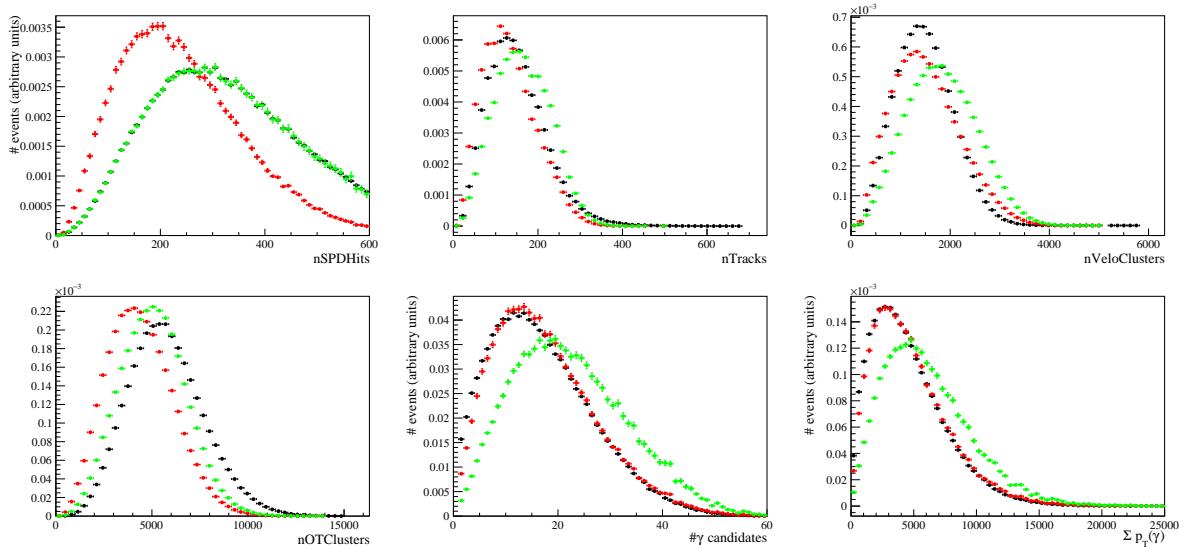


Figure 79: Distributions for various occupancy proxies for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events. The black histograms correspond to background-subtracted data, red to uncorrected simulation and green simulation in which the nSPDHits distribution is corrected to match that in the data (as it is clearly visible on the top left plot). The top distributions correspond to, from left to right, the occupancy of the SPD detector, the number of tracks reconstructed in the events and the number of clusters in the VELO. The bottom plots correspond to, from left to right, the number of clusters in the outer tracker, the number of photons reconstructed in the ECAL, and the sum of the transverse momenta of the photons reconstructed in the ECAL.

1363 The reason why different occupancy proxies show different levels of agreement between
 1364 simulation and data can be attributed to various causes. Firstly, different sub-detectors

1365 can be more or less sensitive to various badly modelled effects, such as back-splash or
1366 secondary interactions. In particular, the SPD detector is expected to be more sensitive
1367 than the ECAL to low energy, badly modelled effects. Secondly, the energy or rapidity
1368 distributions of the particles emerging from the primary proton-proton interaction is also
1369 expected to be badly modelled, impacting the various occupancy proxies in different ways.

1370 In summary, it is impossible to satisfactorily correct the occupancy of the detector
1371 by simply reweighting a single proxy. In particular, reweighting nSPDHits would lead to
1372 a strong disagreement between the ECAL occupancy in the simulation and in data in
1373 a way that would impact on the L0Electron efficiency. Therefore, in this analysis, the
1374 nominal efficiencies are computed without correcting any efficiency proxy. The systematic
1375 uncertainty related to the occupancy is discussed in Section 10.6.

1376 6.8.2 Kinematic distributions in simulation and data

1377 The distributions of several kinematic variables for offline selected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$
1378 events in the μ TOS category are shown in Fig. 80 for the Run 1 conditions and in Fig. 82 for
1379 Run 2 conditions. Similar distributions are shown for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$
1380 events in the eTOS category in Fig. 81 for the Run 1 conditions and in Fig. 83 for Run 2
1381 conditions. For each mode and each kinematic variable, three distributions are shown: the
1382 first corresponds to the background-subtracted data, the second to the simulation with the
1383 trigger and PID correction weights but not the kinematic correction weights, and the third
1384 corresponds to the simulation with all the weights including the kinematic corrections.
1385 The distributions for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ modes in the hTOS! and TIS! categories
1386 are shown in Appendix F.

1387 In the case of the muon mode, the agreement between the data and the simulation
1388 with corrected kinematics is perfect for all the distributions. In the case of the electron
1389 mode, the agreement between the data and the simulation is much improved by applying
1390 the kinematic weights. However, some discrepancies persist in the rapidity distributions²
1391 and in the angle between the two leptons. These are further discussed in Section 10.7.
1392 Some discrepancies are also visible in variables related to the alignment between the B^+
1393 momentum and the B^+ line of flight, such as the $\chi^2_{\text{IP}}(K^+e^+e^-)$, the DIRA or the
1394 missing transverse momentum p_T^{miss} . These discrepancies can however be, at least partly,
1395 attributed to the fact that these variables are correlated to the mass fit variable $m_{\text{DTF}}^{J/\psi}$, and
1396 the background subtraction, which relies on the *sWeight* method, is therefore imperfect.
1397 For these variables, the agreement between data and simulation is instead probed with
1398 the differential $r_{J/\psi}$ test discussed in Section 11.2.

²The small shoulder present in the η distribution of the daughter particles is attributed to a larger amount of material at high η , as explained in Appendix G.

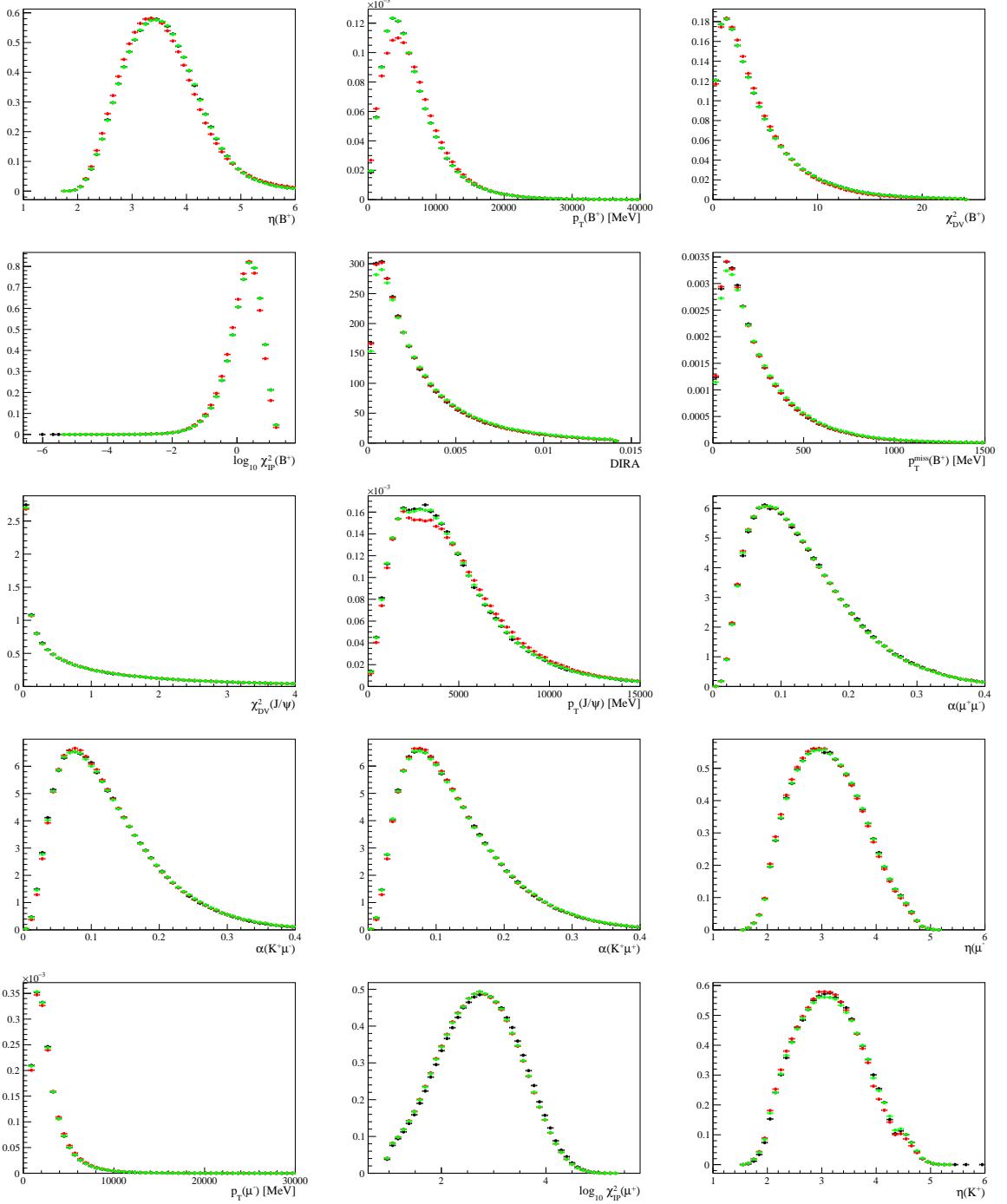


Figure 80: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events in the μ TOS category and in the Run 1 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

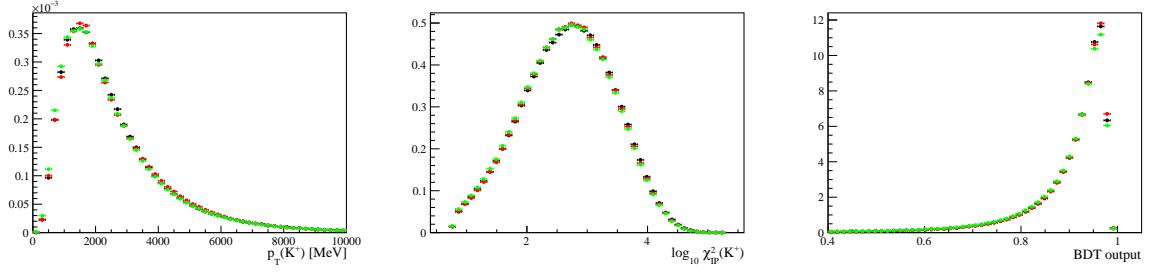


Figure 80: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events in the μ TOS category and in the Run 1 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

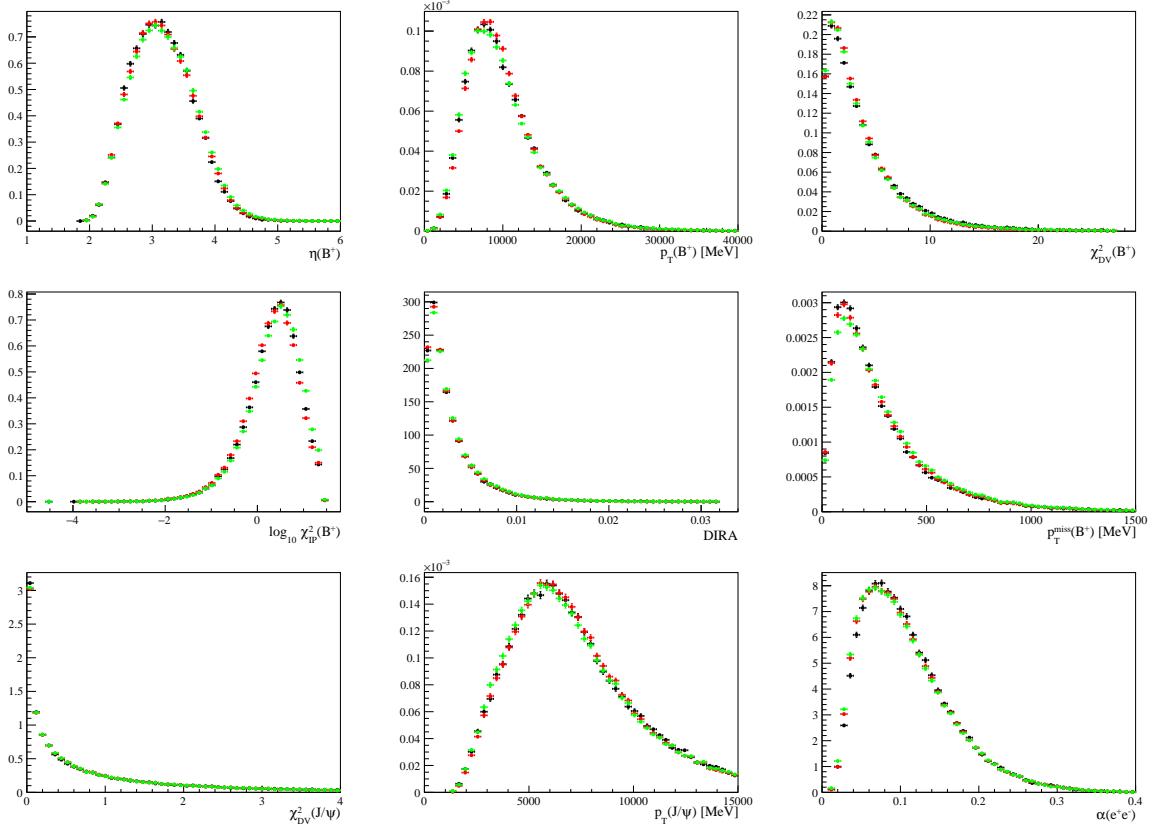


Figure 81: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the e TOS category and in the Run 1 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

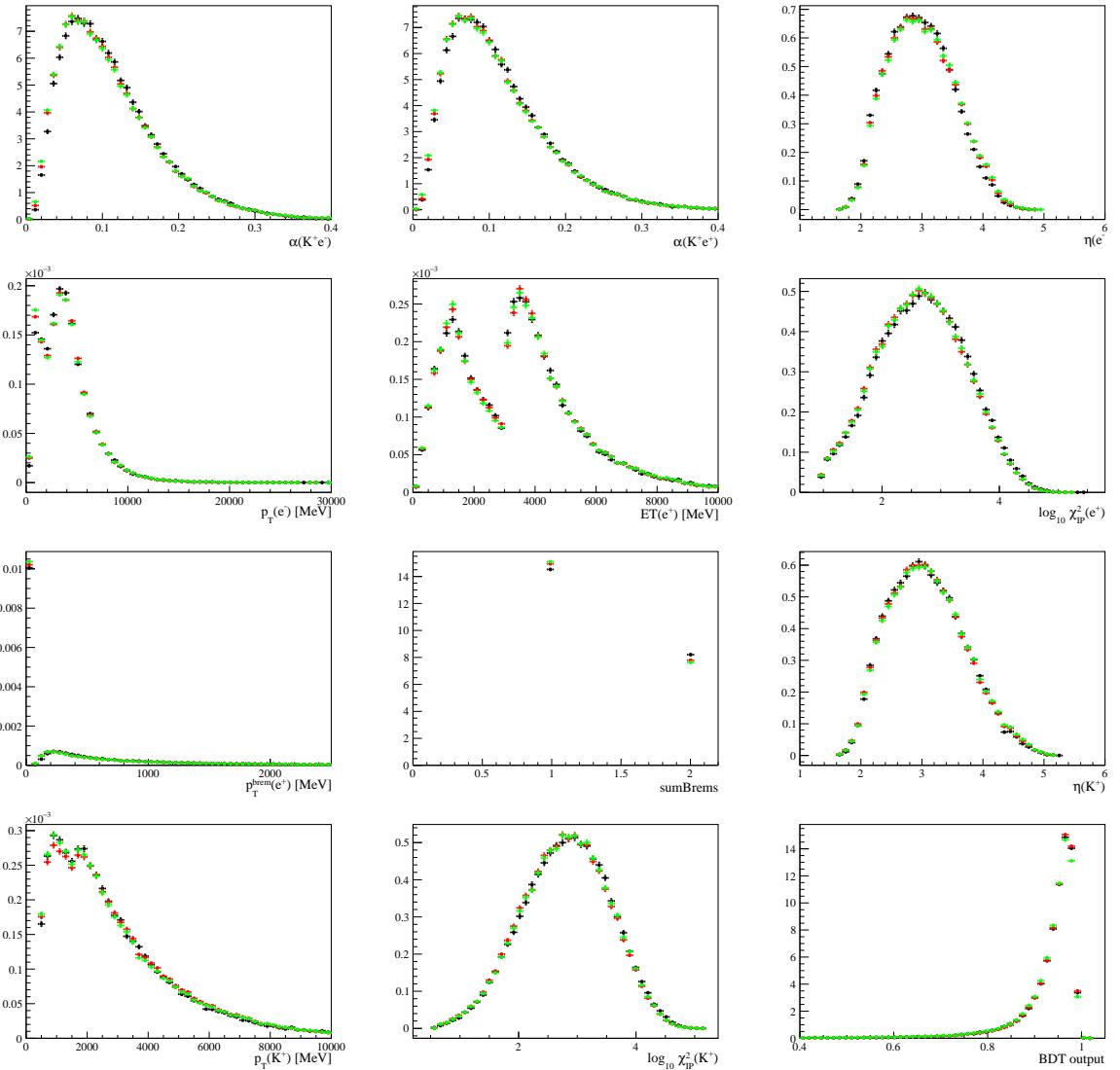


Figure 81: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the e TOS category and in the Run 1 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

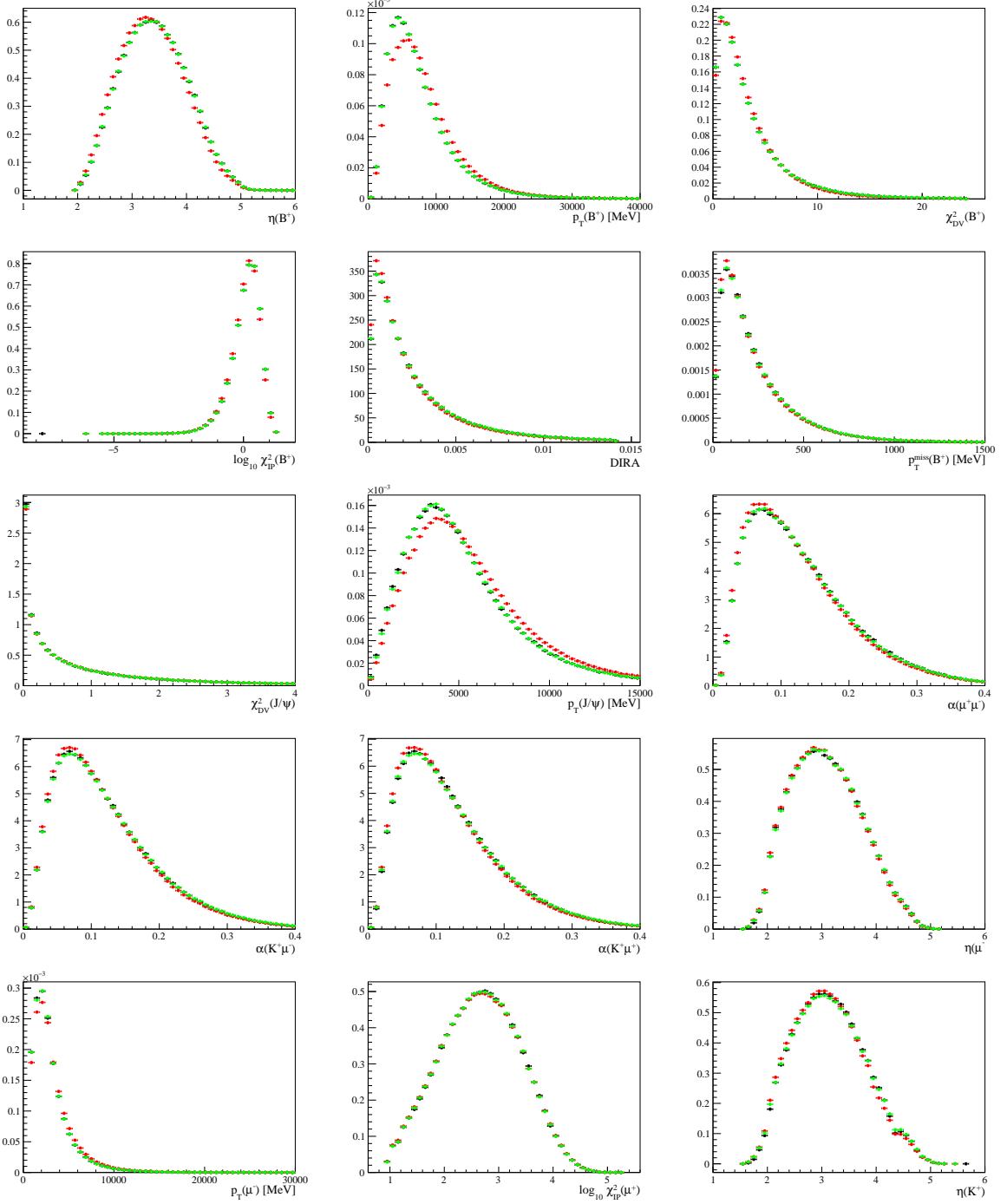


Figure 82: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events in the μTOS category and in the Run 2 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

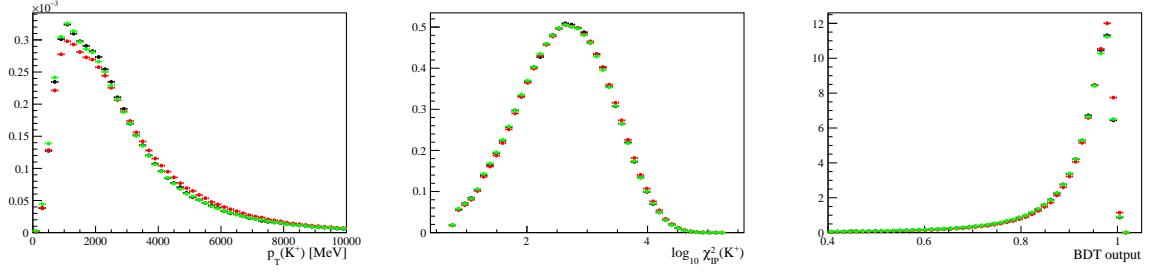


Figure 82: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events in the μ TOS category and in the Run 2 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

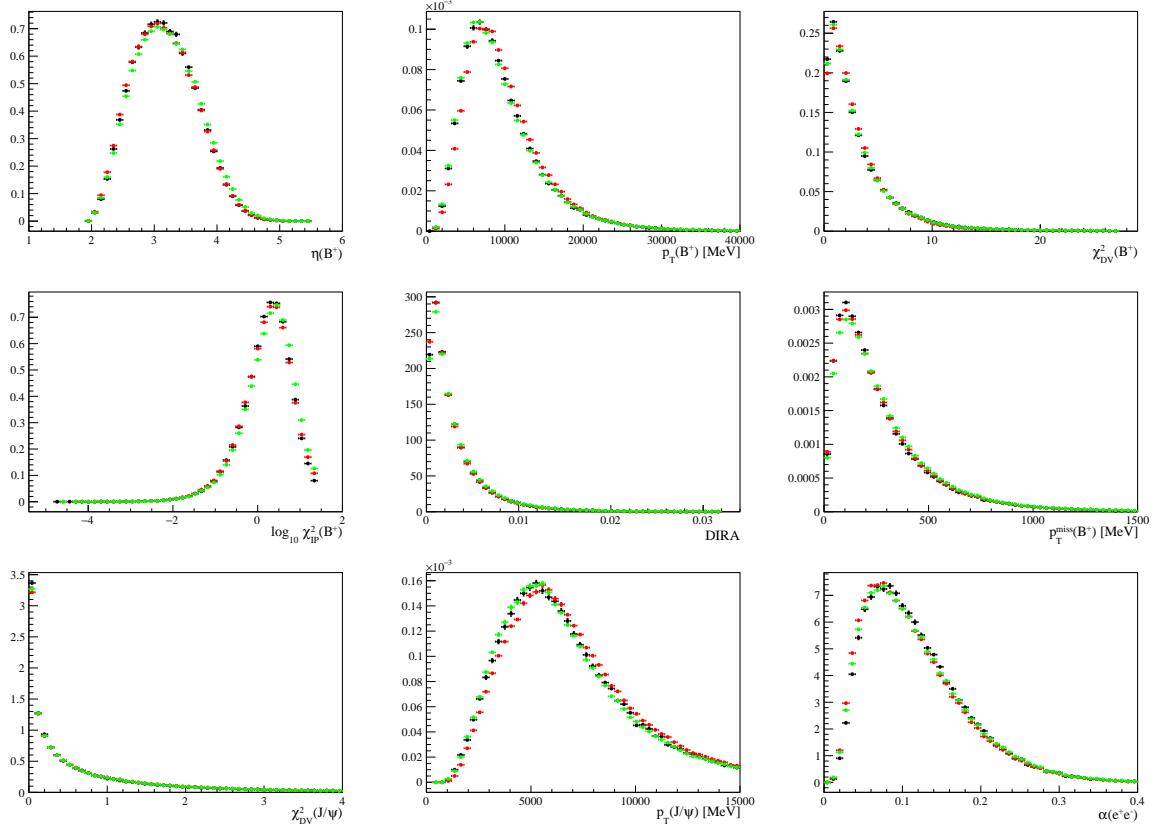


Figure 83: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the e TOS category and in the Run 2 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

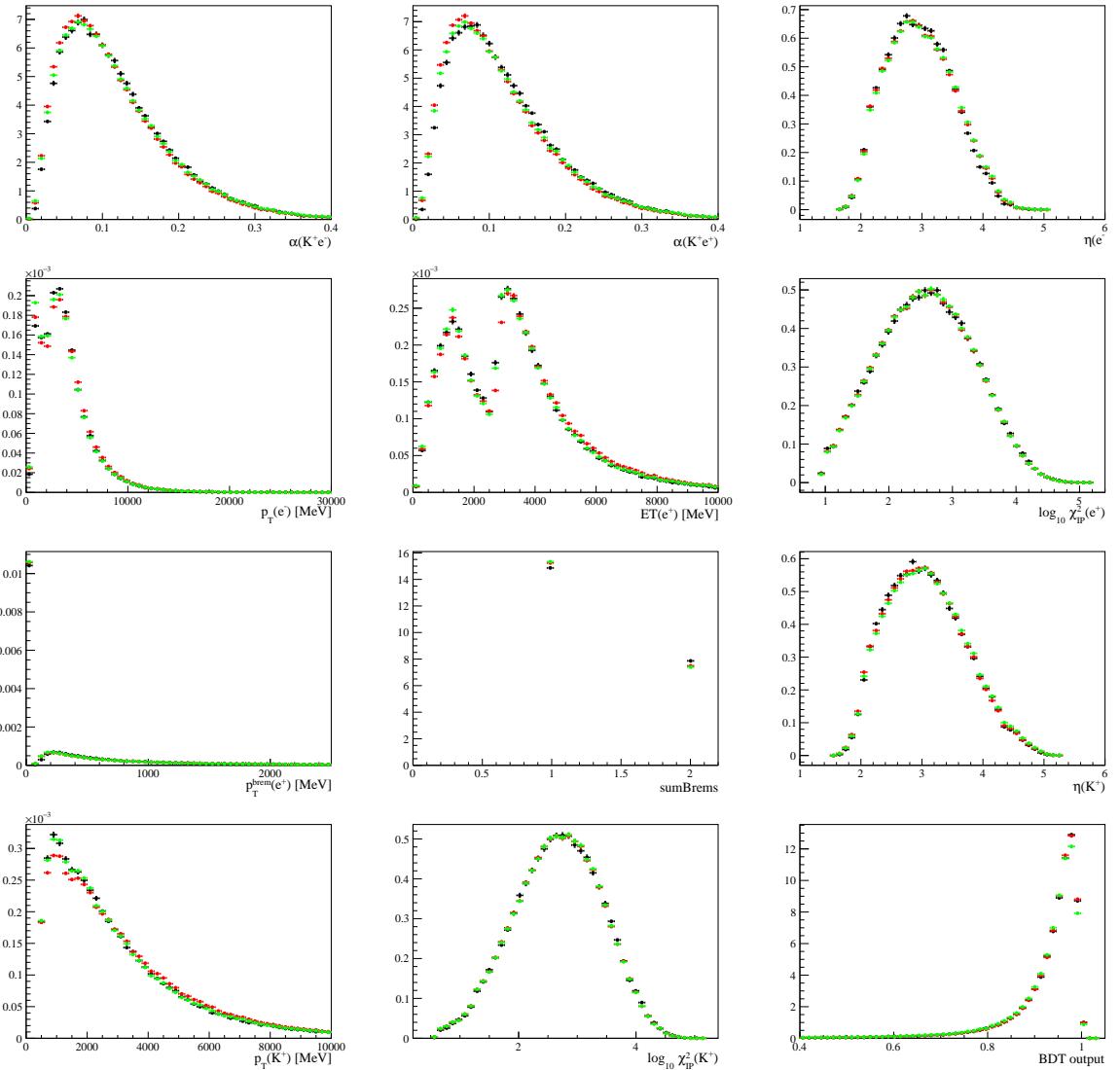


Figure 83: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the eTOS category and in the Run 1 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

1399 6.9 q^2 and B mass resolution in the electron modes

1400 In the simulation, the dielectron invariant mass has a better resolution than in the data.
 1401 For the rare mode, the fraction of events migrating in and out of the true q^2 range
 1402 $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$ is then underestimated in the simulation sample and this impacts
 1403 on the B mass lineshape, meaning that the efficiency of the mass window cut might be
 1404 overestimated if computed from simulated events. This difference in resolution is expected
 1405 for two main reasons: First, differences in the material budget of the detector between
 1406 data and simulation impacts on the amount of bremsstrahlung emitted, which in turn
 1407 impacts on the momentum resolution. Second, the poor description of the occupancy and
 1408 of the calorimeter response impacts on the bremsstrahlung recovery, which influences the
 1409 mass shape.

1410 To study the resolution in simulation and data, the $m(e^+e^-)$ distribution is fitted in the
 1411 $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ control mode both in data and in simulation, in each of the trigger
 1412 and bremsstrahlung categories separately. The full selection is applied to the samples used
 1413 in the fit, adding a cut on the constrained B mass, $5.20 < m_{\text{DTF}}^{J/\psi}(K^+e^+e^-) < 5.68 \text{ GeV}$,
 1414 to reduce the partially-reconstructed backgrounds in the data. In the simulation, $m(e^+e^-)$
 1415 is fitted using a sum of two Crystal Ball functions with opposite tails. This shape is then
 1416 used to fit the data, allowing three parameters to float. The width and mean are allowed
 1417 to float using a width scale factor s_σ and a mean shift parameter $\Delta\mu$, which relates the
 1418 width and mean of the distribution observed in the data to that of the simulation,

$$\mu^{\text{data}} = \mu^{\text{MC}} + \Delta\mu, \quad \sigma^{\text{data}} = s_\sigma \cdot \sigma^{\text{MC}}.$$

1419 In the bremsstrahlung categories 1 and 2, the upper power law tail is also allowed to float
 1420 using one parameter s_α , which is used to scale the tail parameters n and α :

$$\alpha^{\text{data}} = s_\alpha \cdot \alpha^{\text{MC}}, \quad n^{\text{data}} = \frac{1}{s_\alpha} \cdot n^{\text{MC}}.$$

1421 In the data, the residual combinatorial component is modelled using an exponential
 1422 function. The fits in each bremsstrahlung and trigger category are shown in Fig. 84 for
 1423 Run 1 and Fig. 85 for Run 2. The scale and shift factors, as well as their averages across
 1424 all trigger categories, are reported in Tab. 17.

1425 In each bremsstrahlung category, the averages of s_σ , $\Delta\mu$ and μ^{MC} among the three
 1426 trigger categories, denoted $\overline{s_\sigma}$, $\overline{\Delta\mu}$ and $\overline{\mu^{\text{MC}}}$, are computed. In all the simulation samples,
 1427 $m(e^+e^-)$ — and therefore q^2 — is smeared according to the following formula:

$$m^{\text{smeared}} = m^{\text{true}} + \overline{s_\sigma} \cdot (m - m^{\text{true}}) + \overline{\Delta\mu} + (1 - \overline{s_\sigma}) \cdot (\overline{\mu^{\text{MC}}} - m(J/\psi)), \quad (23)$$

1428 where the true and reconstructed dilepton invariant masses, m^{true} and m , are computed
 1429 for each event, $\overline{s_\sigma}$, $\overline{\Delta\mu}$ and $\overline{\mu^{\text{MC}}}$ are extracted from the fit, and $m(J/\psi)$ is the known
 1430 J/ψ mass from Ref. [21]. This formula is found by imposing that if m follows a Gaussian
 1431 distribution of width σ and mean μ , m^{smeared} will follow a Gaussian distribution of width
 1432 $s_\sigma \cdot \sigma$ and mean $\mu + \Delta\mu$. Fig. 86 top shows the fitted shapes for $m^{\text{MC}}(e^+e^-)$, $m^{\text{smeared}}(e^+e^-)$

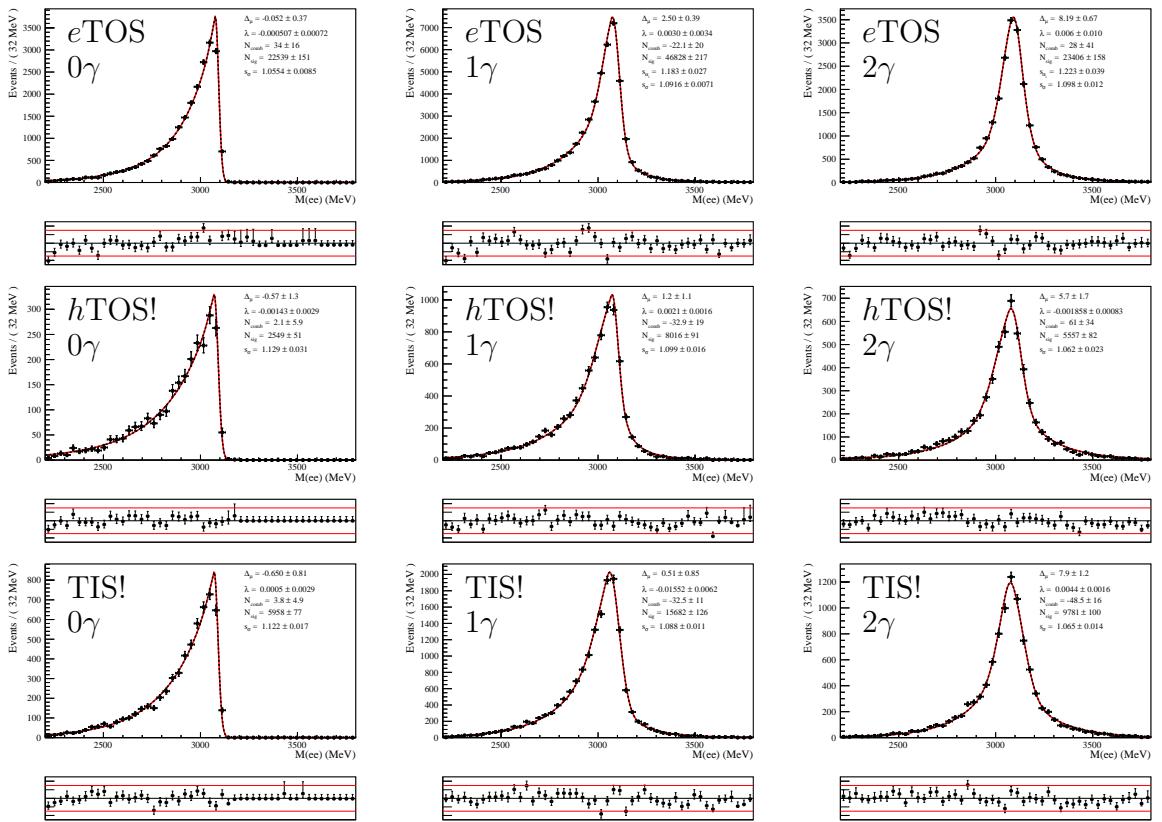


Figure 84: Fit to the $m(e^+e^-)$ invariant mass for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events in the Run 1 data taking conditions, in all trigger and bremsstrahlung categories.

and $\mu^{\text{data}}(e^+e^-)$ on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, showing the improved agreement between simulation and data after applying the smearing. As expected, some small discrepancy remains in the description of the right tail, as Eq. 6.9 does not impact the tail parameters. The systematic related to this discrepancy is computed in Section 10.8.

The mass $m(K^+e^+e^-)$, the resolution of which is dominated by the electron momentum resolution, is also smeared using Eq. 6.9. Fig. 86 bottom shows the fitted shapes for $m^{\text{MC}}(K^+e^+e^-)$, $m^{\text{smeared}}(K^+e^+e^-)$ and $\mu^{\text{data}}(K^+e^+e^-)$. The fit used to obtain these shapes is very similar to the $m(e^+e^-)$ shape and is described in detail in Section 7. The discrepancy in the upper tail is also visible.

It is expected that smearing the q^2 distribution has an impact on the fraction of rare events that migrate in and out of the range $1.1 < q^2 < 6.0 \text{ GeV}^2$. These fractions, before and after applying the smearing, are shown in Tab. 18. It can be seen that although the fraction of events migrating inside the q^2 range is not negligible, the impact of applying the smearing to the simulation is limited, changing the number of events migrating in by only $\sim 1\%$.

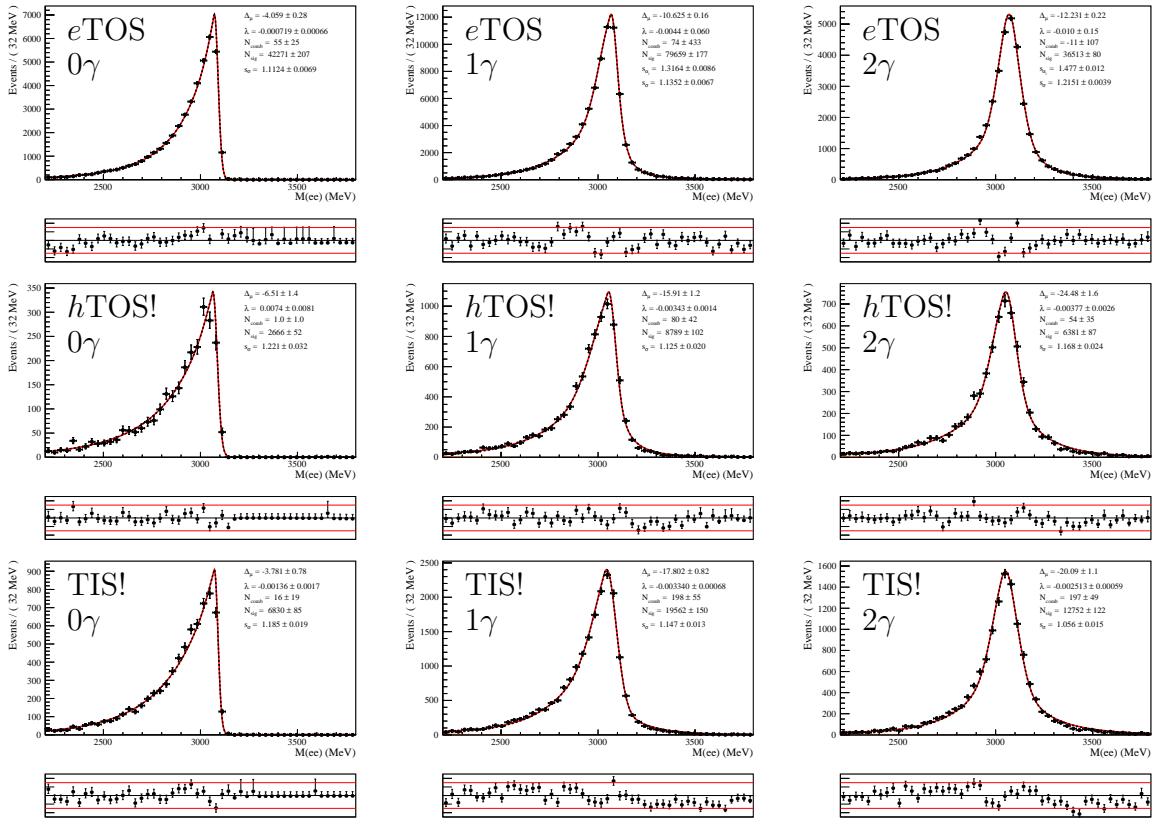


Figure 85: Fit to the $m(e^+e^-)$ invariant mass for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events in the Run 1 data taking conditions, in all trigger and bremsstrahlung categories.

Table 17: Width scale factor and mean shift parameters from the fit to $m(e^+e^-)$ in $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data. Values are reported for each trigger and bremsstrahlung category together with the average among the three trigger categories. The error cited for the average corresponds to the standard deviation between the three trigger categories.

trigger	0 γ		1 γ		2 γ	
	s_σ	$\Delta\mu$	s_σ	$\Delta\mu$	s_σ	$\Delta\mu$
eTOS	1.055 ± 0.008	1.1 ± 0.5	1.092 ± 0.007	4.8 ± 0.4	1.098 ± 0.012	9.6 ± 0.7
hTOS!	1.106 ± 0.030	2.6 ± 1.5	1.136 ± 0.021	3.7 ± 1.5	1.140 ± 0.032	9.1 ± 2.1
TIS!	1.108 ± 0.017	2.8 ± 0.9	1.089 ± 0.012	2.3 ± 1.0	1.103 ± 0.021	10.4 ± 1.5
average	1.068 ± 0.023	1.5 ± 0.7	1.094 ± 0.012	4.3 ± 0.9	1.103 ± 0.012	9.7 ± 0.4
Run 2						
eTOS	1.112 ± 0.007	-1.70 ± 0.33	1.135 ± 0.007	-8.05 ± 0.24	1.215 ± 0.004	-8.83 ± 0.34
hTOS!	1.178 ± 0.031	-1.3 ± 1.6	1.110 ± 0.021	-15.1 ± 1.3	1.29 ± 0.05	-18.3 ± 2.4
TIS!	1.143 ± 0.020	-0.8 ± 0.9	1.178 ± 0.015	-13.8 ± 1.0	1.183 ± 0.024	-15.7 ± 1.4
average	1.118 ± 0.016	-1.59 ± 0.29	1.140 ± 0.018	-8.5 ± 1.7	1.215 ± 0.008	-9.4 ± 2.0

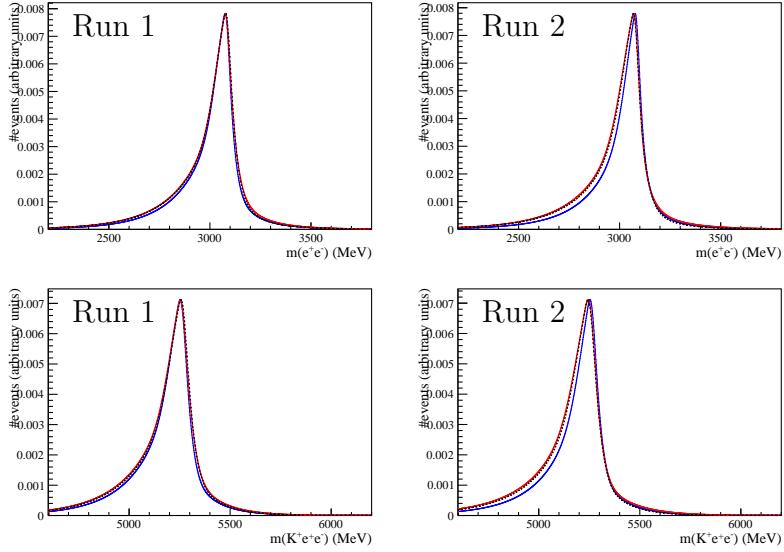


Figure 86: Fit functions to the $m(e^+e^-)$ (top) and $m(K^+e^+e^-)$ distributions in data (dotted black), simulation with no smearing (blue) and simulation after smearing (red). All functions are extracted from a fit to fully selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ eTOS events, in all bremsstrahlung categories superimposed and in the Run 1 (left) and Run 2 (right) data taking conditions.

Table 18: Number of $B^+ \rightarrow K^+e^+e^-$ events migrating in and out of the range $1.1 < q^2 < 6.0 \text{ GeV}^2$, expressed as a percentage of the total number of events in the true range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$. These fractions are estimated using simulated events in both Run 1 and Run 2 data taking conditions, with and without applying the smearing. The first two columns show the fraction of events from the true range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$, that migrate inside the reconstructed range $1.1 < q^2 < 6.0 \text{ GeV}^2$. The third column show the fraction of events inside the true range that stay in the reconstructed range, and the last two columns show the fraction of events that migrate from inside the true range out of the reconstructed range.

	up→in [%]	down→in [%]	in→in [%]	in→up [%]	in→down [%]
Run 1					
No smearing	8.00 ± 0.23	0.34 ± 0.05	96.94 ± 0.14	1.31 ± 0.10	1.75 ± 0.11
smearing	7.90 ± 0.23	0.43 ± 0.06	96.83 ± 0.15	1.53 ± 0.10	1.65 ± 0.11
Run 2					
No smearing	7.77 ± 0.18	0.48 ± 0.05	96.86 ± 0.12	1.40 ± 0.08	1.74 ± 0.09
smearing	8.78 ± 0.20	0.36 ± 0.04	96.50 ± 0.13	1.39 ± 0.08	2.11 ± 0.10

1448 6.10 Summary and list of efficiencies

1449 The functions that compute the efficiencies given a simulation sample and a set of
 1450 weight histograms are written in *superLenin/src/holyEfficiencyComputer.cpp*. The ta-
 1451 bles shown in this section are printed in *superLenin/src/mainPrintEfficiencyTable.cc*.

1452 The total efficiencies related to the $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ and $B^+ \rightarrow K^+ \ell^+\ell^-$ modes, as
 1453 well as the fractions f^{q^2} , computed taking into account all corrections as well as the ghosts,
 1454 are listed in Tab. 19. The individual efficiencies defined in Eq. 13 are listed in Tab. 20 for
 1455 Run 1 and Tab. 21 for Run 2. They are computed without taking the ghosts into account,
 1456 for the reasons explained in Section 6.2.1. The rare mode efficiencies are very similar to the
 1457 J/ψ mode efficiencies, except for the preselection efficiency, where the rare mode efficiency
 1458 is only approximately 34% of that of the J/ψ mode. This is because the preselection
 1459 contains the q^2 cut $1.1 < q^2 < 6.0 \text{ GeV}^2$ and, at generator level, the proportion of events
 1460 in the range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$ is $\sim 32.5\%$.

1461 In Tab. 20 and Tab. 21, the efficiencies are computed in three ways: first using no
 1462 correction simulation, second by including the trigger and PID corrections, but not the
 1463 kinematic correction weights, and third using all corrections including the kinematic
 1464 weights. The latter is used as the nominal way to compute the efficiencies. Particularly
 1465 in the case of the electron mode, the trigger and PID efficiencies vary quite significantly
 1466 depending on whether or not the correction weights are applied. However, the efficiency
 1467 ratio between the rare and control mode $\varepsilon(B^+ \rightarrow K^+ e^+ e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+ e^-))$ do
 1468 not vary by more than $\sim 2\%$ for the *eTOS* and *TIS!* categories, and by 6% for the *hTOS!*
 1469 category. This highlights the robustness of the R_K measurement performed as a double
 1470 ratio.

Table 19: Total efficiencies for for $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$, $B^+ \rightarrow K^+ e^+ e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ decays, as well as efficiency ratios $\varepsilon(B^+ \rightarrow K^+ e^+ e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+ e^-))$ and fraction of events in the true range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$ f^{q^2} , for the Run 1 and Run 2 data taking conditions. All efficiencies are display in percents. All corrections are taken into account, and the events classified as ghosts are also accounted.

[%]	electron modes			muon modes		
	J/ψ	rare	ratio	J/ψ	rare	ratio
Run 1						
<i>eTOS</i>	0.2167 ± 0.0009	0.0544 ± 0.0005	25.11 ± 0.23	1.3964 ± 0.0027	0.3165 ± 0.0016	22.67 ± 0.12
<i>hTOS!</i>	0.03503 ± 0.00029	0.01776 ± 0.00024	50.7 ± 0.8			
<i>TIS!</i>	0.0696 ± 0.0004	0.01925 ± 0.00026	27.7 ± 0.4			
f^{q^2}		29.72 ± 0.05				26.75 ± 0.04
Run 2						
<i>eTOS</i>	0.4188 ± 0.0013	0.1041 ± 0.0007	24.84 ± 0.19	1.4322 ± 0.0028	0.3236 ± 0.0015	22.59 ± 0.11
<i>hTOS!</i>	0.04661 ± 0.00033	0.02644 ± 0.00035	56.7 ± 0.9			
<i>TIS!</i>	0.1095 ± 0.0006	0.0316 ± 0.0004	28.9 ± 0.4			
f^{q^2}		29.67 ± 0.04				26.85 ± 0.04

Table 20: Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+\mu^-$ decays, as well as efficiency ratios $\varepsilon(B^+ \rightarrow K^+ e^+e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+\mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))$, in the Run 1 data taking conditions. All efficiencies and ratios are displayed in percent, and the uncertainty quoted is the statistical uncertainty related to the finite size of the simulation samples. The efficiencies are computed in three ways: first using no correction whatsoever to the simulation, second by including the L0, HLT and PID corrections, but not the kinematic correction weights, and third using all correction including the kinematic weights, which is the nominal way to compute the efficiencies.

$\varepsilon[\%]$	electron modes			muon modes		
	J/ψ	rare	ratio	J/ψ	rare	ratio
No corrections						
$\varepsilon_{\text{geom}}$	16.06 ± 0.03	16.32 ± 0.03	101.57 ± 0.22	16.14 ± 0.03	16.42 ± 0.03	101.72 ± 0.22
$\varepsilon_{\text{rec,strip}}$	13.33 ± 0.02	12.82 ± 0.02	96.18 ± 0.19	23.98 ± 0.01	23.67 ± 0.03	98.70 ± 0.14
$\varepsilon_{\text{presel}}$	76.96 ± 0.06	26.00 ± 0.07	33.79 ± 0.09	60.22 ± 0.03	14.66 ± 0.05	24.35 ± 0.09
ε_{PID}	81.95 ± 0.11	80.00 ± 0.33	97.63 ± 0.42	94.68 ± 0.08	93.54 ± 0.47	98.79 ± 0.50
$\varepsilon^{\ell\text{TOS}}$	23.47 ± 0.07	19.51 ± 0.14	83.15 ± 0.64	76.49 ± 0.04	72.86 ± 0.17	95.26 ± 0.23
ε^{BDT}	87.89 ± 0.12	86.86 ± 0.27	98.82 ± 0.33	87.78 ± 0.03	89.48 ± 0.14	101.94 ± 0.17
$\varepsilon^{\text{fit range}}$	99.39 ± 0.54	92.35 ± 1.13	92.92 ± 1.24	99.57 ± 0.15	97.27 ± 0.67	97.69 ± 0.69
ε^{tot}	0.277 ± 0.002	0.068 ± 0.001	24.61 ± 0.40	1.475 ± 0.004	0.338 ± 0.003	22.91 ± 0.20
$\varepsilon^{h\text{TOS!}}$	4.16 ± 0.03	7.13 ± 0.09	171.41 ± 2.59			
ε^{BDT}	96.49 ± 0.16	94.28 ± 0.30	97.71 ± 0.35			
$\varepsilon^{\text{fit range}}$	99.22 ± 1.21	88.99 ± 1.75	89.69 ± 2.08			
ε^{tot}	0.054 ± 0.001	0.026 ± 0.001	48.41 ± 1.37			
$\varepsilon^{\text{TIS!}}$	8.44 ± 0.05	8.22 ± 0.10	97.38 ± 1.27			
ε^{BDT}	82.60 ± 0.22	78.61 ± 0.50	95.17 ± 0.66			
$\varepsilon^{\text{fit range}}$	99.32 ± 0.92	90.74 ± 1.81	91.36 ± 2.01			
ε^{tot}	0.093 ± 0.001	0.026 ± 0.001	27.29 ± 0.73			
PID and trigger corrections only						
$\varepsilon_{\text{geom}}$	16.06 ± 0.03	16.32 ± 0.03	101.57 ± 0.22	16.14 ± 0.03	16.42 ± 0.03	101.72 ± 0.22
$\varepsilon_{\text{rec,strip}}$	13.33 ± 0.02	12.82 ± 0.02	96.18 ± 0.19	23.98 ± 0.01	23.67 ± 0.03	98.70 ± 0.14
$\varepsilon_{\text{presel}}$	76.04 ± 0.06	26.15 ± 0.07	34.40 ± 0.10	60.22 ± 0.03	14.66 ± 0.05	24.35 ± 0.09
ε_{PID}	76.52 ± 0.08	76.61 ± 0.16	100.12 ± 0.24	91.20 ± 0.03	90.37 ± 0.15	99.09 ± 0.16
$\varepsilon^{\ell\text{TOS}}$	21.92 ± 0.06	18.07 ± 0.12	82.47 ± 0.60	75.81 ± 0.04	72.29 ± 0.17	95.36 ± 0.23
ε^{BDT}	88.08 ± 0.11	86.95 ± 0.24	98.71 ± 0.30	87.81 ± 0.03	89.49 ± 0.14	101.92 ± 0.16
$\varepsilon^{\text{fit range}}$	99.31 ± 0.49	90.07 ± 1.02	90.69 ± 1.12	98.33 ± 0.15	97.24 ± 0.65	98.89 ± 0.68
ε^{tot}	0.240 ± 0.001	0.059 ± 0.001	24.67 ± 0.35	1.391 ± 0.003	0.324 ± 0.003	23.28 ± 0.20
$\varepsilon^{h\text{TOS!}}$	3.24 ± 0.03	5.66 ± 0.07	174.71 ± 2.47			
ε^{BDT}	96.85 ± 0.13	94.64 ± 0.26	97.72 ± 0.30			
$\varepsilon^{\text{fit range}}$	99.23 ± 1.13	87.47 ± 1.57	88.15 ± 1.88			
ε^{tot}	0.039 ± 0.000	0.019 ± 0.000	49.04 ± 1.15			
$\varepsilon^{\text{TIS!}}$	6.56 ± 0.04	6.35 ± 0.07	96.83 ± 1.20			
ε^{BDT}	84.35 ± 0.19	80.88 ± 0.43	95.89 ± 0.55			
$\varepsilon^{\text{fit range}}$	99.29 ± 0.86	89.43 ± 1.66	90.07 ± 1.85			
ε^{tot}	0.070 ± 0.000	0.019 ± 0.000	27.50 ± 0.65			
PID, trigger and kinematic corrections (nominal)						
$\varepsilon_{\text{geom}}$	16.09 ± 0.03	16.35 ± 0.03	101.59 ± 0.23	16.17 ± 0.03	16.46 ± 0.03	101.76 ± 0.23
$\varepsilon_{\text{rec,strip}}$	13.12 ± 0.02	12.68 ± 0.02	96.67 ± 0.19	23.86 ± 0.01	23.46 ± 0.03	98.34 ± 0.14
$\varepsilon_{\text{presel}}$	75.46 ± 0.06	26.18 ± 0.07	34.70 ± 0.10	60.01 ± 0.03	14.52 ± 0.05	24.20 ± 0.09
ε_{PID}	76.49 ± 0.08	76.61 ± 0.17	100.16 ± 0.24	91.32 ± 0.03	90.61 ± 0.14	99.22 ± 0.16
$\varepsilon^{\ell\text{TOS}}$	19.91 ± 0.06	16.63 ± 0.11	83.54 ± 0.63	75.43 ± 0.04	71.59 ± 0.18	94.91 ± 0.24
ε^{BDT}	86.80 ± 0.12	85.59 ± 0.27	98.61 ± 0.34	86.74 ± 0.04	88.38 ± 0.15	101.89 ± 0.18
$\varepsilon^{\text{fit range}}$	99.24 ± 0.50	89.13 ± 1.02	89.81 ± 1.12	99.51 ± 0.15	97.07 ± 0.66	97.55 ± 0.68
ε^{tot}	0.209 ± 0.001	0.053 ± 0.001	25.25 ± 0.39	1.376 ± 0.003	0.312 ± 0.003	22.67 ± 0.20
$\varepsilon^{h\text{TOS!}}$	2.88 ± 0.02	5.20 ± 0.06	180.93 ± 2.59			
ε^{BDT}	96.43 ± 0.15	93.91 ± 0.29	97.40 ± 0.34			
$\varepsilon^{\text{fit range}}$	99.15 ± 1.14	86.22 ± 1.57	86.96 ± 1.87			
ε^{tot}	0.034 ± 0.000	0.017 ± 0.000	51.04 ± 1.34			
			122			
$\varepsilon^{\text{TIS!}}$	6.58 ± 0.04	6.39 ± 0.07	97.08 ± 1.20			
ε^{BDT}	82.69 ± 0.21	79.14 ± 0.46	95.71 ± 0.61			
$\varepsilon^{\text{fit range}}$	99.24 ± 0.86	88.62 ± 1.65	89.30 ± 1.83			
ε^{tot}	0.067 ± 0.001	0.019 ± 0.000	27.88 ± 0.70			

Table 21: Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+\mu^-$ decays, as well as efficiency ratios $\varepsilon(B^+ \rightarrow K^+ e^+e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+\mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))$, in the Run 2 data taking conditions. All efficiencies and ratios are displayed in percent, and the uncertainty quoted is the statistical uncertainty related to the finite size of the simulation samples. The efficiencies are computed in three ways: first using no correction whatsoever to the simulation, second by including the L0 and PID corrections, but not the kinematic correction weights, and third using all correction including the kinematic weights, which is the nominal way to compute the efficiencies.

$\varepsilon[\%]$	electron modes			muon modes		
	J/ψ	rare	ratio	J/ψ	rare	ratio
No corrections						
$\varepsilon_{\text{geom}}$	17.30 ± 0.03	17.58 ± 0.03	101.60 ± 0.24	17.38 ± 0.03	17.65 ± 0.03	101.58 ± 0.24
$\varepsilon_{\text{rec,strip}}$	14.19 ± 0.01	13.70 ± 0.02	96.58 ± 0.18	24.10 ± 0.01	23.50 ± 0.03	97.48 ± 0.12
$\varepsilon_{\text{presel}}$	75.39 ± 0.05	25.94 ± 0.07	34.40 ± 0.10	59.64 ± 0.03	14.71 ± 0.05	24.66 ± 0.08
ε_{PID}	82.12 ± 0.09	80.87 ± 0.35	98.48 ± 0.44	95.34 ± 0.07	94.99 ± 0.42	99.63 ± 0.44
$\varepsilon^{\ell\text{TOS}}$	38.37 ± 0.07	31.98 ± 0.17	83.34 ± 0.47	72.24 ± 0.04	68.93 ± 0.16	95.41 ± 0.22
ε^{BDT}	88.34 ± 0.08	87.47 ± 0.22	99.02 ± 0.26	91.38 ± 0.03	93.33 ± 0.10	102.13 ± 0.12
$\varepsilon^{\text{fit range}}$	99.38 ± 0.35	91.15 ± 0.93	91.72 ± 0.99	99.52 ± 0.14	97.03 ± 0.59	97.50 ± 0.61
ε^{tot}	0.512 ± 0.002	0.129 ± 0.002	25.16 ± 0.33	1.565 ± 0.004	0.362 ± 0.003	23.12 ± 0.18
$\varepsilon^{h\text{TOS!}}$	4.25 ± 0.03	9.13 ± 0.10	214.86 ± 2.86			
ε^{BDT}	90.81 ± 0.20	82.43 ± 0.46	90.76 ± 0.54			
$\varepsilon^{\text{fit range}}$	99.19 ± 1.02	89.47 ± 1.75	90.20 ± 1.99			
ε^{tot}	0.058 ± 0.001	0.034 ± 0.001	58.48 ± 1.57			
$\varepsilon^{\text{TIS!}}$	8.62 ± 0.04	8.88 ± 0.10	102.99 ± 1.30			
ε^{BDT}	88.65 ± 0.16	84.56 ± 0.44	95.39 ± 0.53			
$\varepsilon^{\text{fit range}}$	99.10 ± 0.73	88.52 ± 1.73	89.32 ± 1.87			
ε^{tot}	0.115 ± 0.001	0.034 ± 0.001	29.17 ± 0.74			
PID and trigger corrections only						
$\varepsilon_{\text{geom}}$	17.30 ± 0.03	17.58 ± 0.03	101.60 ± 0.24	17.38 ± 0.03	17.65 ± 0.03	101.58 ± 0.24
$\varepsilon_{\text{rec,strip}}$	14.19 ± 0.01	13.70 ± 0.02	96.58 ± 0.18	24.10 ± 0.01	23.50 ± 0.03	97.48 ± 0.12
$\varepsilon_{\text{presel}}$	73.75 ± 0.05	26.48 ± 0.07	35.90 ± 0.10	59.64 ± 0.03	14.71 ± 0.05	24.66 ± 0.08
ε_{PID}	77.51 ± 0.07	77.50 ± 0.16	99.99 ± 0.23	93.05 ± 0.03	92.67 ± 0.11	99.59 ± 0.13
$\varepsilon^{\ell\text{TOS}}$	38.80 ± 0.07	32.04 ± 0.16	82.56 ± 0.42	72.46 ± 0.04	69.07 ± 0.15	95.32 ± 0.22
ε^{BDT}	88.54 ± 0.07	87.36 ± 0.20	98.66 ± 0.23	91.31 ± 0.03	93.31 ± 0.10	102.20 ± 0.11
$\varepsilon^{\text{fit range}}$	99.28 ± 0.33	88.03 ± 0.82	88.67 ± 0.88	99.47 ± 0.14	96.97 ± 0.58	97.48 ± 0.60
ε^{tot}	0.482 ± 0.001	0.122 ± 0.001	25.25 ± 0.30	1.530 ± 0.004	0.353 ± 0.003	23.09 ± 0.18
$\varepsilon^{h\text{TOS!}}$	4.30 ± 0.03	8.87 ± 0.09	206.35 ± 2.56			
ε^{BDT}	91.06 ± 0.19	82.79 ± 0.42	90.92 ± 0.49			
$\varepsilon^{\text{fit range}}$	99.13 ± 0.96	85.98 ± 1.56	86.74 ± 1.78			
ε^{tot}	0.055 ± 0.000	0.031 ± 0.001	56.01 ± 1.29			
$\varepsilon^{\text{TIS!}}$	8.66 ± 0.04	8.89 ± 0.09	102.66 ± 1.18			
ε^{BDT}	88.99 ± 0.14	84.63 ± 0.40	95.09 ± 0.47			
$\varepsilon^{\text{fit range}}$	99.07 ± 0.68	85.43 ± 1.54	86.23 ± 1.66			
ε^{tot}	0.110 ± 0.001	0.032 ± 0.001	29.04 ± 0.65			
PID, trigger and kinematic corrections (nominal)						
$\varepsilon_{\text{geom}}$	17.58 ± 0.03	17.87 ± 0.03	101.64 ± 0.24	17.66 ± 0.03	17.94 ± 0.03	101.60 ± 0.24
$\varepsilon_{\text{rec,strip}}$	13.66 ± 0.01	13.17 ± 0.02	96.37 ± 0.18	23.32 ± 0.01	22.70 ± 0.03	97.35 ± 0.12
$\varepsilon_{\text{presel}}$	72.99 ± 0.05	26.38 ± 0.07	36.15 ± 0.11	59.39 ± 0.03	14.37 ± 0.05	24.20 ± 0.08
ε_{PID}	77.39 ± 0.07	77.49 ± 0.17	100.13 ± 0.23	93.18 ± 0.03	92.81 ± 0.11	99.61 ± 0.13
$\varepsilon^{\ell\text{TOS}}$	34.30 ± 0.06	28.01 ± 0.15	81.67 ± 0.45	69.00 ± 0.04	65.52 ± 0.16	94.97 ± 0.24
ε^{BDT}	87.17 ± 0.08	86.42 ± 0.21	99.14 ± 0.26	90.00 ± 0.03	92.19 ± 0.12	102.43 ± 0.14
$\varepsilon^{\text{fit range}}$	99.25 ± 0.33	87.51 ± 0.83	88.18 ± 0.89	99.45 ± 0.15	96.84 ± 0.59	97.38 ± 0.61
ε^{tot}	0.403 ± 0.002	0.102 ± 0.001	25.31 ± 0.32	1.407 ± 0.003	0.318 ± 0.002	22.58 ± 0.18
$\varepsilon^{h\text{TOS!}}$	3.65 ± 0.02	7.72 ± 0.08	211.32 ± 2.68			
ε^{BDT}	89.96 ± 0.21	81.99 ± 0.44	91.15 ± 0.53			
$\varepsilon^{\text{fit range}}$	99.08 ± 0.97	85.24 ± 1.56	86.03 ± 1.79			
ε^{tot}	0.045 ± 0.001	0.026 ± 0.001	57.86 ± 1.46			
			123			
$\varepsilon^{\text{TIS!}}$	8.75 ± 0.04	8.97 ± 0.10	102.53 ± 1.18			
ε^{BDT}	87.63 ± 0.16	83.73 ± 0.42	95.55 ± 0.50			
$\varepsilon^{\text{fit range}}$	99.02 ± 0.68	84.71 ± 1.53	85.55 ± 1.66			
ε^{tot}	0.105 ± 0.001	0.031 ± 0.001	29.38 ± 0.69			

1471 7 Mass modelling

1472 The value of R_K is determined from a simultaneous fit to the $m(K^+\ell^+\ell^-)$ distribution of
 1473 $B^+ \rightarrow K^+e^+e^-$ and $B^+ \rightarrow K^+\mu^+\mu^-$ events. The fit procedure is described in more detail
 1474 in Section 8. The yields of $B^+ \rightarrow K^+J/\psi(\ell^+\ell^-)$ events are determined separately through
 1475 extended maximum likelihood fits to the $m_{\text{DTF}}^{J/\psi}(K^+\ell^+\ell^-)$ distributions, and incorporated
 1476 to the simultaneous fit as external constraints.

1477 The probability density functions (PDF) used to describe the data can be generically
 1478 expressed as

$$F^{\text{trig}}(m) \propto N_{\text{sig}}^{\text{trig}} S^{\text{trig}}(m) + N_{\text{bkg}}^{\text{trig}} B^{\text{trig}}(m) \quad (24)$$

1479 where S^{trig} and B^{trig} are the PDFs describing the signal and background distributions
 1480 in each of the trigger categories, $N_{\text{sig}}^{\text{trig}}$ and $N_{\text{bkg}}^{\text{trig}}$ are their associated yields and $m \equiv$
 1481 $\{m(K^+\ell^+\ell^-), m_{\text{DTF}}^{J/\psi}(K^+\ell^+\ell^-)\}$. Note that in the fit to the modes with muons in the final
 1482 state only one trigger category is defined (μTOS), in contrast to the three trigger categories
 1483 defined for those modes with electrons ($e\text{TOS}$, $h\text{TOS!}$ and TIS!).

1484 Additionally, for electron channels, one needs to take into account that the shape of
 1485 the signal distribution, $S^{\text{trig}}(m)$, depends strongly on whether the bremsstrahlung recovery
 1486 procedure added a photon cluster to none, either or both of the electron candidates.
 1487 Therefore, a different signal distribution model is used for events belonging to each of the
 1488 following three photon categories,

- 1489 • 0γ : No photon is recovered;
- 1490 • 1γ : The momentum of only one electron was corrected by the bremsstrahlung
 1491 recovery;
- 1492 • 2γ : The momenta of both electrons were corrected by the bremsstrahlung recovery.

1493 The signal PDF in a given trigger category is then given by

$$S^{\text{trig}}(m) = f_{0\gamma}^{\text{trig}} S_{0\gamma}^{\text{trig}}(m) + f_{1\gamma}^{\text{trig}} S_{1\gamma}^{\text{trig}}(m) + (1 - f_{0\gamma}^{\text{trig}} - f_{1\gamma}^{\text{trig}}) S_{2\gamma}^{\text{trig}}(m). \quad (25)$$

1494 In the fit to the data, the values of $f_{0\gamma}$ and $f_{1\gamma}$, the fractions of events in the 0γ and 1γ
 1495 photon categories, are constrained to the values observed in simulation, with an uncertainty
 1496 of 1% coming from the limited statistics of the simulation sample. The comparison between
 1497 the fractions of each of the individual photon categories obtained in simulation with those
 1498 extracted from fits to $B^+ \rightarrow K^+J/\psi(e^+e^-)$ candidates in data is shown in Tab. 22.

1499 The fitting ranges for electron and muon channels are also chosen to be different, see
 1500 Tab. 3. This is due to the different mass resolution between electrons and muons, which
 1501 impacts on the ability to separate the tails from the signal from the various background
 1502 components.

1503 In the following, the models used to describe the various signal and background
 1504 components entering the fits for the rare, normalisation and control modes are described.
 1505 Systematic variation of the models assumed are considered in Section 10.

Table 22: Fraction of events falling in each of the photon categories as calculated from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulation and from a fit to $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ candidates in data.

	eTOS		hTOS!		TIS!	
	Simulation	Data	Simulation	Data	Simulation	Data
$f_{0\gamma}$	0.2383 ± 0.0015	0.2485 ± 0.0020	0.1558 ± 0.0030	0.1511 ± 0.0041	0.2052 ± 0.0024	0.2049 ± 0.0032
$f_{1\gamma}$	0.4962 ± 0.0018	0.4992 ± 0.0029	0.4982 ± 0.0043	0.4970 ± 0.0082	0.4957 ± 0.0030	0.4949 ± 0.0057
$f_{2\gamma}$	0.2655 ± 0.0016	0.2536 ± 0.0018	0.3460 ± 0.0042	0.3519 ± 0.0054	0.2991 ± 0.0027	0.3014 ± 0.0034

1506 7.1 Fit to resonant modes

1507 In order to obtain the yield of the resonant modes $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ and $B^+ \rightarrow$
1508 $K^+ \psi(2S)(\ell^+\ell^-)$, a fit is performed to the masses $m_{\text{DTF}}^{J/\psi}(K^+ \ell^+\ell^-)$ and $m_{\text{DTF}}^{\psi(2S)}(K^+ \ell^+\ell^-)$
1509 where the decay fit is performed by constraining the dilepton mass to that of the J/ψ and
1510 $\psi(2S)$ mesons respectively (see Section 5.4). xsxc

1511 7.1.1 Fit to $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$

1512 The distribution of simulated $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events surviving the entire selection
1513 process is shown in Fig. 156. This distribution is modeled using a Hypatia function [42]
1514 with one power tail to the left and one to the right. The different parameters characterising
1515 the PDF, are obtained from the fit to these simulated $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events.
1516 Discrepancies in the resolution and absolute mass scale between data and simulation need
1517 to be taken into account. This is implemented by introducing a mass shift parameter,
1518 $\Delta\mu = m_B^{\text{data}} - m_B^{\text{MC}}$, and a width scale factor, $s_\sigma = \sigma^{\text{data}}/\sigma^{\text{MC}}$, which modify the mean
1519 and width of the distribution and are allowed to float in the fit to the data.

1520 In the considered mass range, the only contributions that are expected to contaminate
1521 the sample are those coming from the Cabibbo-suppressed decay $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ and
1522 from random combinations of tracks in the event, see Section 5. The $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$
1523 decay is described with a sum of two Crystal Ball (CB) functions [43], with common
1524 mean and width, one having its tail to the left and the other to the right. The tail
1525 parameters, as well as the fraction between the two CBs, are fixed from a fit to simulated
1526 $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ events, see Appendix H.1.1. The same corrective parameters $\Delta\mu$ and
1527 s_σ , are introduced in this parameterisation to correct the mean and width of the distribution
1528 in the same way as for the signal. The ratio of yields of the $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ component
1529 to the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ signal is Gaussian-constrained to the ratio of branching
1530 fractions [21] times the ratio of efficiencies,

$$\frac{N(B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-))}{N(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))} = \frac{\mathcal{B}(B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-))}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))} \cdot \frac{\varepsilon_{J/\psi K}}{\varepsilon_{J/\psi \pi}} \quad (26)$$

1531 The width of this constraint is set to the error derived from the branching fraction
1532 measurements, which is 10%. This is much larger than the error expected on the ratio of
1533 efficiencies.

1534 Finally, the combinatorial background is described by an exponential function, whose
 1535 yield and coefficient, λ , are left free to float in the fit.

1536 **7.1.2 Fit to $B^+ \rightarrow K^+ J/\psi(e^+e^-)$**

1537 In the case of the electron resonant mode, $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, the signal shape is
 1538 described as the sum of three distributions, each modelling the mass shape of one of the
 1539 three bremsstrahlung categories. Each of these three shapes is described by a sum of two
 1540 CBs with opposite tails, with a common mean and two width parameters. The parameters
 1541 of these shapes are fixed from a fit to the simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ sample, as
 1542 shown in Appendix C.

1543 Due to the worse resolution for electrons, partially reconstructed background, coming
 1544 predominantly from $B^0 \rightarrow K^{*0} J/\psi(e^+e^-)$ decays, becomes relevant in this sample. This
 1545 component is modelled by a kernel estimation function (`RooKeysPDF`), determined from a
 1546 fit to the $B^+ \rightarrow J/\psi(e^+e^-)X$ simulation samples, see Appendix H.2.3. The fraction of
 1547 strange to charmed partially reconstructed events is Gaussian-constrained to the fraction
 1548 in simulation, with a conservative error estimate of 20%. This constraint could in principle
 1549 be tightened by considering all the branching fractions of all the decays constituting
 1550 the $B^+ \rightarrow J/\psi X$ simulation samples and using their propagated errors as the standard
 1551 deviation of the constraint. This is however not necessary, as it would not improve
 1552 the precision on the control mode yield by a significant amount, which is a quantity
 1553 that in any case does not limit the precision of the R_K measurement. Misidentified
 1554 $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ events also contribute in this fitting range. However, no simulated
 1555 sample for these kind of decays is available, and so simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events
 1556 are used to extract the shape of this contribution. The method employed estimates the
 1557 shift in mass induced by a $\pi \rightarrow K$ mis identification for each candidate. The mis-ID mass
 1558 is computed,

$$m_{\text{misID}}^{J/\psi} = \left((m_{\text{DTF}}^{J/\psi}(K^+ e^+ e^-))^2 + \frac{E_B}{E_K} \cdot (m_K^2 - m_\pi^2)^2 \right)^{\frac{1}{2}} \quad (27)$$

1559 for each $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ event, where E_B and E_K are the energy of the B and
 1560 K candidates, and m_K and m_π are the mass of the kaon and pion, respectively. Given
 1561 the very similar decay kinematics of $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$,
 1562 the distribution of $m_{\text{misID}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events is then equivalent to that of
 1563 misidentified $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ events, up to small corrections induced by kaon-pion
 1564 mass differences. More detail on this method is given in Appendix C.3. The shape
 1565 of the distribution obtained is fitted using a sum of two CBs with opposite tails, as
 1566 shown in Fig. 174. As before, the yield of misidentified $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ events is
 1567 Gaussian-constrained to the ratio of branching fractions [21] times efficiencies.

1568 The combinatorial shape is again described with an exponential function. Its yield and
 1569 coefficient, λ , are allowed to float in the data fit.

1570 In the fit to the data, the means and widths of the PDFs describing the $B^+ \rightarrow$
 1571 $K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ components are left free to float but use common
 1572 scale and width parameters, $\Delta\mu$ and s_σ .

1573 **7.1.3 Fit to $B^+ \rightarrow K^+\psi(2S)(\mu^+\mu^-)$**

1574 The model used to fit candidates selected in the q^2 region around the $\psi(2S)$ contains a signal
1575 component, describing $B^+ \rightarrow K^+\psi(2S)(\mu^+\mu^-)$ decays and a combinatorial background
1576 component, to account for random combinations of tracks.

1577 The signal PDF is composed of two CB functions and its parameters are obtained
1578 from a fit to $B^+ \rightarrow K^+\psi(2S)(\mu^+\mu^-)$ simulated events, see Appendix H.1.3. The mean
1579 and width of the distribution are allowed to float in the fit to the data, through the free
1580 parameters $\Delta\mu$ and s_σ . The combinatorial background is described by an exponential
1581 function.

1582 **7.1.4 Fit to $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$**

1583 For the $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ mode, contributions from partially reconstructed back-
1584 grounds, dominated by $B^0 \rightarrow K^{*0}\psi(2S)(ee)$, and leakage from $B^+ \rightarrow K^+J/\psi(e^+e^-)$ are
1585 also relevant. These components are described using `RooKeysPdf` functions extracted from
1586 simulation samples. The shape for partially reconstructed $B^0 \rightarrow K^{*0}\psi(2S)(ee)$ events is
1587 extracted from a dedicated simulated sample of this specific decay, see Appendix H.1.4.
1588 Additional contributions are expected from partially reconstructed $B^+ \rightarrow J/\psi(e^+e^-)X$
1589 events, and so contributions from charm and strange components are extracted from
1590 the $B^+ \rightarrow J/\psi(e^+e^-)X$ simulated sample, see Appendix H.2.3. Candidates from the
1591 $B^+ \rightarrow K^+J/\psi(e^+e^-)$ decay, leaking in to the $\psi(2S)$ q^2 window due to the electron
1592 resolution, are also modelled using the $B^+ \rightarrow K^+J/\psi(e^+e^-)$ simulated sample, see Ap-
1593 pendix H.2.5.

1594 Concerning the signal component, the $m_{\text{DTF}}^{\psi(2S)}(K^+e^+e^-)$ distribution in each of the
1595 trigger and photon categories is modelled by a double CB, for which the parameters are
1596 fixed to those obtained from a fit to simulated $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ events. As for the
1597 other modes, the mean and width of these distributions are modulated by the parameters
1598 $\Delta\mu$ and s_σ , which are allowed to float in the fit to the data. The fits to the simulation for
1599 the signal components can be found in Appendix H.1.4.

1600 Finally, combinatorial background is again described by an exponential function, the
1601 shape and normalisation of which are left free in the data fit.

1602 **7.2 Rare mode fits**

1603 The value of R_K is obtained from a simultaneous fit to $B^+ \rightarrow K^+\mu^+\mu^-$ and $B^+ \rightarrow K^+e^+e^-$
1604 decays in the invariant mass of the kaon and dielectron pair, $m(K^+\ell^+\ell^-)$.

1605 As the shape of the signal and background components is derived from simulated events,
1606 discrepancies in the mass resolution and absolute scale between data and simulation need to
1607 be taken into account. As explained in the previous section, for the normalisation channels
1608 the additional parameters $\Delta\mu$ and s_σ are introduced in the fit to the data to accommodate
1609 these discrepancies. However, due to the limited statistics, it is not possible to extract the
1610 values of these parameters in the fit to the rare decays. Therefore, the parameters $\Delta\mu$
1611 and s_σ are constrained using a fit to the $m(K^+\ell^+\ell^-)$ distributions of $B^+ \rightarrow K^+J/\psi(\ell^+\ell^-)$

events. An additional parameter, s_α , that modifies the parameters describing the righthand side tail of the distribution is also introduced. This parameter accounts for discrepancies in the energy distribution of the recovered bremsstrahlung photons between data and simulation. The extra requirement $m_{\text{DTF}}^{J/\psi}(K^+\ell^+\ell^-) > 5.2 \text{ GeV}/c^2$ is used to suppress partially reconstructed backgrounds in the $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ fits.

Systematic effects induced by possible correlation between mass resolution and q^2 are investigated in Section 10.8.

7.2.1 Fit to $B^+ \rightarrow K^+ \mu^+ \mu^-$

The $m(K^+ \mu^+ \mu^-)$ distribution for both $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events is described by a combination of a double CB and a Gaussian function. The parameters of this functions are extracted from fits to $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ simulated events. The results of these fits are shown in Appendix H.1.5 and Appendix H.1.1.

For the normalisation mode, the fit components are identical to those described in Section 7.1.1 above. For the rare decay, in addition to the signal component, the combinatorial background component is described by an exponential function with freely floating coefficient and normalisation. As explained above, no partially reconstructed component is expected in the fitted region for the $B^+ \rightarrow K^+ \mu^+ \mu^-$ decay.

7.2.2 Fit to $B^+ \rightarrow K^+ e^+ e^-$

The $m(K^+ e^+ e^-)$ distribution of $B^+ \rightarrow K^+ e^+ e^-$ events depends strongly on the photon category. The shape for 0γ category is characterised by a large tail at lower masses, due to the bremsstrahlung radiation from the electrons. The mass shape is parameterised using a double CB function. For the case of the 1γ and 2γ categories, where a fraction of the bremsstrahlung energy is recovered, the lower mass tail of the distribution is reduced and the width is enlarged given the broader energy resolution in the calorimeter. Signal candidates in these categories are described with a combination of a double Crystal Ball and a Gaussian function. The parameters of these PDFs are determined from fits to $B^+ \rightarrow K^+ e^+ e^-$ simulated events in the different trigger and photon categories, see Appendix H.1.6. The same parameterisation is used to describe the $m(K^+ e^+ e^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events, but the parameters are in this case extracted from fits to $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events, see Appendix H.1.2. The means, widths and righthand-side-tail parameters of all these distributions are modified by the parameters $\Delta\mu$, s_σ and s_α , which are common to the $B^+ \rightarrow K^+ e^+ e^-$ and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ models and are allowed to float in the fit to the data.

Concerning the background model, in the fit to the rare mode, the dominant backgrounds come from partially reconstructed $B^0 \rightarrow K^{*0} e^+ e^-$ decays, leakage from the resonant mode $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ and combinatorial background.

The fit range is optimised to suppress contributions from higher-mass $K^{**} \equiv K^{*0}_0(1430), K^{*0}_2(1430), K_1$ resonances, see Section 5.1. Therefore, only $B^0 \rightarrow K^{*0} e^+ e^-$ decays are considered to contribute to the partially reconstructed background. Fig. 180

1651 shows the distribution of partially reconstructed $B^0 \rightarrow K^{*0}e^+e^-$ events obtained from
1652 simulation, as well as the `RooKeysPdf` function that is used in the fit to the data.

1653 Due to the bin migration of events induced by the bremsstrahlung radiation of the
1654 electrons, some events from the tail of the resonant decay contaminate the lower part
1655 of the rare decay window. This contamination is described by a `RooKeysPdf` function
1656 obtained from simulated $B^+ \rightarrow K^+J/\psi(e^+e^-)$ events that survive the $B^+ \rightarrow K^+e^+e^-$
1657 selection, see Appendix H.2.5. The yield of this component is given by

$$N_{\text{charm}} = f_{\text{leak}} \times N_{B^+ \rightarrow K^+J/\psi(e^+e^-)} \quad (28)$$

1658 where f_{leak} is the ratio of the efficiency of the rare mode selection relative to that of the
1659 $B^+ \rightarrow K^+J/\psi(e^+e^-)$ selection. The value of f_{leak} is Gaussian-constrained to the value
1660 estimated from the simulation.

1661 In the fit to the resonant $B^+ \rightarrow K^+J/\psi(e^+e^-)$ mode, the $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-) > 5.2 \text{ GeV}/c^2$
1662 cut suppresses the contamination from partially reconstructed backgrounds. The contribu-
1663 tions that are included in the fit are then those related to misidentified $B^+ \rightarrow \pi^+J/\psi(e^+e^-)$
1664 events and to combinatorial background. These components are modelled as described in
1665 Section 7.1.2 above.

1666 8 Strategy for the extraction of R_K

1667 In this section, the strategy to combine the observed yields of $B^+ \rightarrow K^+\mu^+\mu^-$ and
 1668 $B^+ \rightarrow K^+e^+e^-$ decays into a measurement of R_K is presented. As explained in the
 1669 previous sections, the number of $B^+ \rightarrow K^+e^+e^-$ decays can be extracted from three
 1670 exclusive samples corresponding to three different trigger selections. Furthermore, the
 1671 analysis is performed separately for the Run 1 and Run 2 samples. The value of R_K can
 1672 then be computed in each of this exclusive samples as

$$R_K^{rt} = \frac{N_{K\mu\mu}^r}{N_{Kee}^{rt}} \cdot \frac{N_{J/\psi ee}^{rt}}{N_{J/\psi\mu\mu}^r} \cdot \frac{\varepsilon_{Kee}^r}{\varepsilon_{K\mu\mu}^r} \cdot \frac{\varepsilon_{J/\psi\mu\mu}^r}{\varepsilon_{J/\psi ee}^{rt}} \quad (29)$$

$$= \frac{N_{K\mu\mu}^r}{N_{Kee}^{rt}} \cdot c_K^{rt} \quad (30)$$

1673 where $N^{r(t)}$ is the number of events from each B^+ decay observed in the sample corre-
 1674 sponding to a specific run $r = 1, 2$ and trigger category, $t = e\text{TOS}, h\text{TOS!}, \text{TIS!}$. The
 1675 symbol ε^{rt} denotes the total efficiency for each decay to be selected as part of one of these
 1676 samples. The efficiencies are calculated using simulation that has been corrected to better
 1677 match the effects seen in data, see Section 6. As most of the corrections are derived from
 1678 the normalisation channels $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+J/\psi(e^+e^-)$, there will be
 1679 correlations between the efficiencies and the normalisation channel yields. These are taken
 1680 into account when computing the factors c_K^{rt} by following the bootstrapping procedure
 1681 described in Section 10.1.

1682 All the uncertainties that impact c_K^{rt} are referred to as systematic uncertainties, in
 1683 contrast to the statistical uncertainties impacting the ratio of yields. All systematic
 1684 uncertainties, as well as their correlations, are presented in Section 10. A similar quantity
 1685 as c_K^{rt} used for the double ratio $R_{\psi(2S)}$ is denoted $c_{\psi(2S)}^{rt}$ and is defined in a similar way as
 1686 c_K^{rt} , substituting the efficiencies related to the rare mode by those related to the $\psi(2S)$
 1687 mode.

1688 Let $\log L_e^{rt}$ and $\log L_\mu^r$ be the log-likelihoods obtained from the invariant mass fits for
 1689 the electron and muon rare modes,

$$\log L_e^{rt} = \sum_i \log f_e^{rt}(m_{rt}^i | N_{Kee}^{rt}) \quad (31)$$

$$\log L_\mu^r = \sum_i \log f_\mu^{rt}(m_r^i | N_{K\mu\mu}^r) \quad (32)$$

1690 where the sum runs over all the events in each $r(t)$ sample, $m_{r(t)}^i$ is the mass associated to
 1691 one event, and $f^{e,\mu}$ are the PDFs used to model them. By performing a simultaneous fit to
 1692 $B^+ \rightarrow K^+\mu^+\mu^-$ and $B^+ \rightarrow K^+e^+e^-$ in all the categories, the log-likelihood is formulated
 1693 as a function of the parameter of interest, R_K , as

$$\sum_{rt} \sum_i \log f_e^{rt} \left(m_i^{rt} \middle| \frac{N_{K\mu\mu}^r \cdot c_K^{rt}}{R_K} \right) + \sum_r \sum_i \log f_\mu^{rt}(m_i^{rt} | N_{K\mu\mu}^r) + \sum_{rt} \log G(c_K^{rt}, \mu_{rt}, \sigma_{rt}), \quad (33)$$

for the case of uncorrelated systematics in the the c_K^{rt} factors. Taking into account correlations across categories, this log-likelihood takes the form

$$\sum_{rt} \sum_i \log f_e^{rt} \left(m_i^{rt} \middle| \frac{N_{K\mu\mu}^r \cdot c_K^{rt}}{R_K} \right) + \sum_r \sum_i \log f_\mu^{rt}(m_i^{rt} | N_{K\mu\mu}^r) + \log G(c_K^{rt}, \boldsymbol{\mu}, \boldsymbol{\sigma}) \quad (34)$$

where the last term represents a multidimensional Gaussian constraint that takes as an input the array of c_K^{rt} central values, $\boldsymbol{\mu}$, and the 6×6 covariance matrix, $\boldsymbol{\sigma}$. This means that the uncertainty obtained for R_K by performing such a fit contains not only the statistical uncertainty associated to the rare mode fits but also the systematic uncertainties coming from the ratio of efficiencies and $r_{J/\psi}$.

An additional advantage of this simultaneous procedure is that it allows the introduction of an additional constraint relating the amount of partially reconstructed background across the three trigger categories by requiring

$$\frac{N_{prc}^{rt}}{N_{prc}^{r0}} = \frac{\varepsilon_{trig,mass}^{rt}(K^*ee)}{\varepsilon_{trig,mass}^{r0}(K^*ee)} = r_{prc}^{rt}, \quad (35)$$

where r_{prc}^{rt} is determined from simulation. The value of this constraint is derived assuming that, in the considered mass window, the partially reconstructed background is composed mainly of $B^0 \rightarrow K^{*0}e^+e^-$ decays. As shown in Section 5.7, the mass distributions of $B \rightarrow K^{*0}(\rightarrow K^+\pi^-)e^+e^-$ and $B^+ \rightarrow K^{*+}(\rightarrow K^+\pi^0)e^+e^-$ events are very similar, and thus no attempt is made in the fit to separate both components. However, the presence of a π^- or a π^0 in the decay can induce differences in the trigger efficiency of the two processes. These differences are taken into account as a systematic uncertainty in the ratios r_{prc}^{rt} , as described in Section 10.10.

As described in Section 7.2, the parameters $\Delta\mu$, s_σ and s_α that account for discrepancies between data and simulation in the signal shape are constrained to those measured in the fit to the $m(K^+\ell^+\ell^-)$ distribution of $B^+ \rightarrow K^+J/\psi(\ell^+\ell^-)$ events. Their correlations with the c_K^{rt} parameters are expected to be small given the low background level in the fit to the normalisation channel and are ignored in the fit. The possible systematic effect associated with this assumption is investigated in Section 10.1.

8.1 Fit validation

To check for possible bias and for the coverage of the fit errors a toy study is performed. Using the model for the mass fit, ~ 5000 samples are generated, with the expected number of signal and background events for each of the trigger categories. These yields are calculated by assuming the value for the branching fraction of $B^+ \rightarrow K^+\mu^+\mu^-$ previously measured by LHCb [44] and taking $R_K = 1$. The simultaneous fit is performed to each of these samples and the distribution of the fitted parameters is studied. As the final set of systematic uncertainties were not available at the time of conducting the validation studies, we assume a 4% systematic uncertainty associated to each c_K^{rt} factor and a correlation of 50% among them (this can be seen to be conservative, see Tab. 48). For the r_{prc}^{rt} constants

we assume a 10% systematic uncertainty, and also a 50% correlation. The systematics in this case need to be finalised but this is again thought to be conservative. Figure 87 shows the fit to an individual toy experiment as an example. Out of the ~ 5000 toy experiments, all of the fits report back as converged (MIGRAD and HESSE status = 0). Asymmetric errors are computed with MINOS, in 6% of the cases these error are not computed and the experiments are discarded from the validation study.

The residual and pull distributions of the signal parameters obtained from the toy study for Run 1, Run 2 and the combination of both are shown in Fig. 88, Fig. 89 and Fig. 90 respectively. In all cases, it can be seen that the R_K distribution is asymmetric but is centred around the generated value. The mean of the R_K pull is slightly shifted to smaller values, indicating a small bias at the level of -4% of the statistical uncertainty. The width of the pull distribution for R_K is however compatible with unity, indicating a good coverage of the asymmetric error obtained from the fit. The toy results for the rest of the parameters in the fit are summarised in Tab. 23, Tab. 24 and Tab. 25 for Run 1, Run 2 and the combination of both respectively, and their residual and pull distributions can be found in Appendix I.2.

The small bias observed for R_K is a consequence of the poor mass resolution caused by the electrons bremsstrahlung which complicates the separation of the different background components in the fit. This impacts especially the determination of the yield for partially reconstructed and combinatorial backgrounds, but also the extraction of the electron signal yield, and hence R_K . This effect can be alleviated by choosing a slightly less optimal working point in the BDT selection, which allows a better control of the combinatorial shape (see Appendix I.1), but cannot be completely eliminated without degrading the sensitivity. Various modifications to the fit procedure have been trialled to minimise this effect. A version of the fit with a fitting window extended to lower values in $m(K^+e^+e^-)$ (4740-6200 MeV/ c^2) was tested. The R_K bias in this case stays at the same level, and, although an improvement in the global precision is observed, a further systematic associated with additional partially reconstructed components would need to be taken into account. An additional check was performed where the parameter of interest in the fit was changed to the inverse of R_K . The R_K^{-1} distribution is observed to be more symmetric than that for R_K , however no difference to the bias or to the final precision is observed. The bias is observed to remain even if only the eTOS trigger category is fitted and is much worse if either of the other trigger categories is fitted alone.

Given the results of these checks, the proposed strategy is to keep the current implementation of the fit, to adjust the R_K fit result for the bias, and assign an extra systematic uncertainty to the measurement associated with how wrong this adjustment could be. Similar biases are observed when varying the generated value of R_K , see Appendix I.2.

The relative error on R_K can be estimated from the width of the distribution of the fitted values obtained from the toy study. These uncertainties, obtained for different hypothesis on the true R_K value, are given in Tab. 26. For $R_K = 1$ a 7.0% precision on R_K is expected given the estimated background levels. This is compared with the previous Run 1 analysis which had a precision of 11.0%. For $R_K = 1$, the bias discussed above would then be $4\% \times 7\% = 0.0028$.

Table 23: Summary of the Run 1 toy study results. For each parameter the generated value (Generated), the mean and sigma of the obtained distribution (Fit), the expected relative uncertainty (Error) and the mean and sigma of the pull distribution are given.

	Generated	Fit	Error (%)	PullMean	PullSigma
		1D M fit:			
RKR1	1	1.01 ± 0.11	10.8	-0.017 ± 0.014	0.985 ± 0.011
nPartRecoKeeETOSR1	95	91 ± 42	45.6	-0.055 ± 0.014	0.996 ± 0.011
nCharmLeakKeeETOSR1	8	8.00 ± 0.79	9.82	-0.006 ± 0.014	0.988 ± 0.010
nCombKeeETOSR1	210	213 ± 57	26.7	0.040 ± 0.014	0.986 ± 0.010
expoConstKeeETOSR1	-0.00295	-0.00295 ± 0.00049	-16.7	-0.040 ± 0.014	0.994 ± 0.010
nCharmLeakKeeHTOSR1	17	17.0 ± 1.7	9.91	-0.004 ± 0.014	0.994 ± 0.010
nCombKeeHTOSR1	333	335 ± 30	8.91	0.046 ± 0.014	0.995 ± 0.011
expoConstKeeHTOSR1	-0.00255	-0.00256 ± 0.00022	-8.72	-0.026 ± 0.014	0.985 ± 0.010
nCharmLeakKeeTISR1	16	16.0 ± 1.6	10.1	-0.018 ± 0.014	1.013 ± 0.010
nCombKeeTISR1	266	268 ± 31	11.6	0.046 ± 0.014	0.984 ± 0.010
expoConstKeeTISR1	-0.00249	-0.00251 ± 0.00028	-11.2	-0.066 ± 0.014	0.992 ± 0.010
nSignalKmumuR1	$1e + 03$	1003 ± 35	3.52	-0.008 ± 0.014	0.9957 ± 0.0099
nCombKmumuR1	244	244 ± 22	9.08	-0.016 ± 0.014	0.989 ± 0.010
expoConstKmumuR1	-0.003	-0.00299 ± 0.00068	-22.9	0.006 ± 0.014	0.9960 ± 0.0100

Table 24: Summary of the Run 2 toy study results. For each parameter the generated value (Generated), the mean and sigma of the obtained distribution (Fit), the expected relative uncertainty (Error) and the mean and sigma of the pull distribution are given.

	Generated	Fit	Error (%)	PullMean	PullSigma
		1D M fit:			
RKR2	1	1.000 ± 0.088	8.75	-0.042 ± 0.014	0.994 ± 0.010
nPartRecoKeeETOSR2	159	155 ± 43	27.7	-0.055 ± 0.014	0.982 ± 0.010
nCharmLeakKeeETOSR2	7	6.98 ± 0.70	10	-0.039 ± 0.014	0.9922 ± 0.0094
nCombKeeETOSR2	170	174 ± 56	32.4	0.046 ± 0.014	0.995 ± 0.011
expoConstKeeETOSR2	-0.00291	-0.00289 ± 0.00062	-21.3	-0.019 ± 0.015	1.000 ± 0.011
nCharmLeakKeeHTOSR2	6	6.00 ± 0.59	9.76	0.005 ± 0.014	0.9737 ± 0.0097
nCombKeeHTOSR2	120	121 ± 21	17.6	0.042 ± 0.014	0.987 ± 0.011
expoConstKeeHTOSR2	-0.00336	-0.00339 ± 0.00048	-14.1	-0.041 ± 0.014	0.9915 ± 0.0095
nCharmLeakKeeTISR2	2	2.00 ± 0.21	10.5	-0.011 ± 0.014	0.991 ± 0.010
nCombKeeTISR2	100	102 ± 22	21.6	0.051 ± 0.014	1.007 ± 0.011
expoConstKeeTISR2	-0.00238	-0.00240 ± 0.00051	-21	-0.077 ± 0.015	1.0110 ± 0.0100
nSignalKmumuR2	886	886 ± 33	3.76	-0.006 ± 0.014	1.002 ± 0.010
nCombKmumuR2	192	192 ± 20	10.5	-0.019 ± 0.014	0.9919 ± 0.0099
expoConstKmumuR2	-0.003	-0.00299 ± 0.00079	-26.6	-0.005 ± 0.014	0.9932 ± 0.0096

1772 8.2 Compatibility with LHCb-PAPER-2014-024

1773 The change in the stripping version for Run 1 data, together with the reoptimisation of
 1774 the selection procedure, implies that the Run 1 candidates selected for this analysis are
 1775 different from those used to derive the previous measurement of R_K published by the
 1776 LHCb collaboration [6]. The event overlap after the whole selection chain is estimated in
 1777 the following way: a sample of $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events is reconstructed and selected

Table 25: Summary of the Run 1 and Run 2 combination toy study results. For each parameter the generated value (Generated), the mean and sigma of the obtained distribution (Fit), the expected relative uncertainty (Error) and the mean and sigma of the pull distribution are given.

	Generated	Fit	Error (%)	PullMean	PullSigma
		1D M fit:			
R_K	1	1.001 ± 0.070	7	-0.040 ± 0.014	0.991 ± 0.010
nPartRecoKeeETOSR1	95	93 ± 40	42.3	-0.006 ± 0.015	0.995 ± 0.011
nPartRecoKeeETOSR2	159	154 ± 41	26.7	-0.054 ± 0.015	0.994 ± 0.011
nCharmLeakKeeETOSR1	8	8.02 ± 0.78	9.74	0.024 ± 0.014	0.976 ± 0.011
expoConstKeeETOSR1	-0.00295	-0.00294 ± 0.00049	-16.7	-0.012 ± 0.015	0.988 ± 0.011
nCombKeeETOSR1	210	211 ± 50	23.8	-0.004 ± 0.014	0.985 ± 0.011
nCharmLeakKeeETOSR2	7	6.99 ± 0.69	9.82	-0.017 ± 0.014	0.982 ± 0.010
expoConstKeeETOSR2	-0.00291	-0.00290 ± 0.00060	-20.6	-0.037 ± 0.015	0.975 ± 0.010
nCombKeeETOSR2	170	174 ± 54	31	0.042 ± 0.015	0.988 ± 0.011
nCharmLeakKeeHTOSR1	17	17.0 ± 1.7	9.78	-0.020 ± 0.014	0.978 ± 0.011
expoConstKeeHTOSR1	-0.00255	-0.00255 ± 0.00022	-8.67	-0.008 ± 0.014	0.997 ± 0.010
nCombKeeHTOSR1	333	334 ± 28	8.3	0.011 ± 0.014	0.9879 ± 0.0098
nCharmLeakKeeHTOSR2	6	5.99 ± 0.61	10.1	-0.019 ± 0.015	1.007 ± 0.011
expoConstKeeHTOSR2	-0.00336	-0.00340 ± 0.00047	-13.9	-0.047 ± 0.014	0.9732 ± 0.0091
nCombKeeHTOSR2	120	121 ± 21	17.4	0.043 ± 0.015	0.994 ± 0.010
nCharmLeakKeeTISR1	16	16.0 ± 1.5	9.66	0.006 ± 0.014	0.966 ± 0.010
expoConstKeeTISR1	-0.00249	-0.00251 ± 0.00027	-10.8	-0.053 ± 0.014	0.986 ± 0.010
nCombKeeTISR1	266	266 ± 29	10.9	0.015 ± 0.014	0.997 ± 0.011
nCharmLeakKeeTISR2	2	2.00 ± 0.20	9.83	0.024 ± 0.014	0.986 ± 0.010
expoConstKeeTISR2	-0.00238	-0.00242 ± 0.00049	-20.1	-0.109 ± 0.015	0.989 ± 0.010
nCombKeeTISR2	100	102 ± 21	20.8	0.072 ± 0.014	0.995 ± 0.010
nSignalKmumuR1	1003	1003 ± 34	3.43	-0.016 ± 0.015	0.996 ± 0.010
expoConstKmumuR1	-0.003	-0.00302 ± 0.00068	-22.6	-0.041 ± 0.014	0.9866 ± 0.0100
nCombKmumuR1	244	244 ± 22	8.86	-0.006 ± 0.014	0.9824 ± 0.0098
nSignalKmumuR2	886	886 ± 31	3.53	-0.002 ± 0.014	0.964 ± 0.010
expoConstKmumuR2	-0.003	-0.00297 ± 0.00078	-26.3	0.012 ± 0.014	0.983 ± 0.010
nCombKmumuR2	192	192 ± 20	10.3	-0.041 ± 0.014	0.9858 ± 0.0097

1778 using a selection kept as similar as possible as that used in [6] (using `stripping` 20 and a
 1779 new BDT that is trained with the same variables as in the previous analysis). Among the
 1780 events reconstructed using the strategy of the previous analysis, approximately 80% are
 1781 also present in the current analysis. This 80% overlap is assumed to be the same for the
 1782 rare mode.

1783 To estimate an interval of values that one would expect from the measurement of R_K
 1784 in the overlapping Run 1 samples, pairs of toy experiments are generated using the model
 1785 described above in such a way that they share 80% of their events. First, a sample A
 1786 is generated assuming $R_K = 0.745$, as measured in [6], but with yields corresponding to
 1787 only 80% of the integrated luminosity used for the measurement. Then two more samples,
 1788 corresponding to the remainder 20% of the luminosity, are generated: one assuming
 1789 $R_K = 0.745$ (B) and the other with a value of R_K that is varied between 0.6 and 1.1
 1790 (C). The samples obtained by combining A+B and A+C are subsequently fitted following
 1791 the nominal procedure and the difference in the measured value of R_K , ΔR_K , is studied.

Table 26: Sensitivity on the determination of R_K obtained from toy studies with different generated hypotheses. The sensitivity expected in the Run 1 and Run 2 samples, as well as the combination of the two, are provided. The results of the previous LHCb analysis [6] are given for comparison.

Sample	R_K gen	R_K fit	$\sigma(R_K)/R_K (\%)$
Run 1	1.0	1.01 ± 0.11	10.9
Run 1	0.745	0.748 ± 0.074	9.9
Run 1	1.25	1.27 ± 0.16	12.8
Run 2	1.0	1.000 ± 0.088	8.8
Run 2	0.745	0.746 ± 0.058	7.7
Run 2	1.25	1.26 ± 0.12	9.6
Run 1 + Run 2	1.0	1.001 ± 0.070	7.0
Run 1 + Run 2	0.745	0.747 ± 0.046	6.2
Run 1 + Run 2	1.25	1.255 ± 0.098	7.8
LHCb-PAPER-2014-024	-	$0.745^{+0.090}_{-0.074}$	11.0

1792 Figure 91, shows the expected spread of values from the fits performed with the old and
 1793 new samples, i.e. the standard deviation of the ΔR_K distribution, for different true values
 1794 of R_K . This estimation assumes systematic uncertainties are fully correlated between the
 1795 two measurements.

1796 Based on these curves, and assuming the true value of R_K lies in the interval (0.5,1), a
 1797 new measurement of R_K such that $|\Delta R_K| < 0.153$ would be statistically compatible with
 1798 the previous measurement ($R_K = 0.745^{+0.097}_{-0.082}$) at the level of 2.5σ .

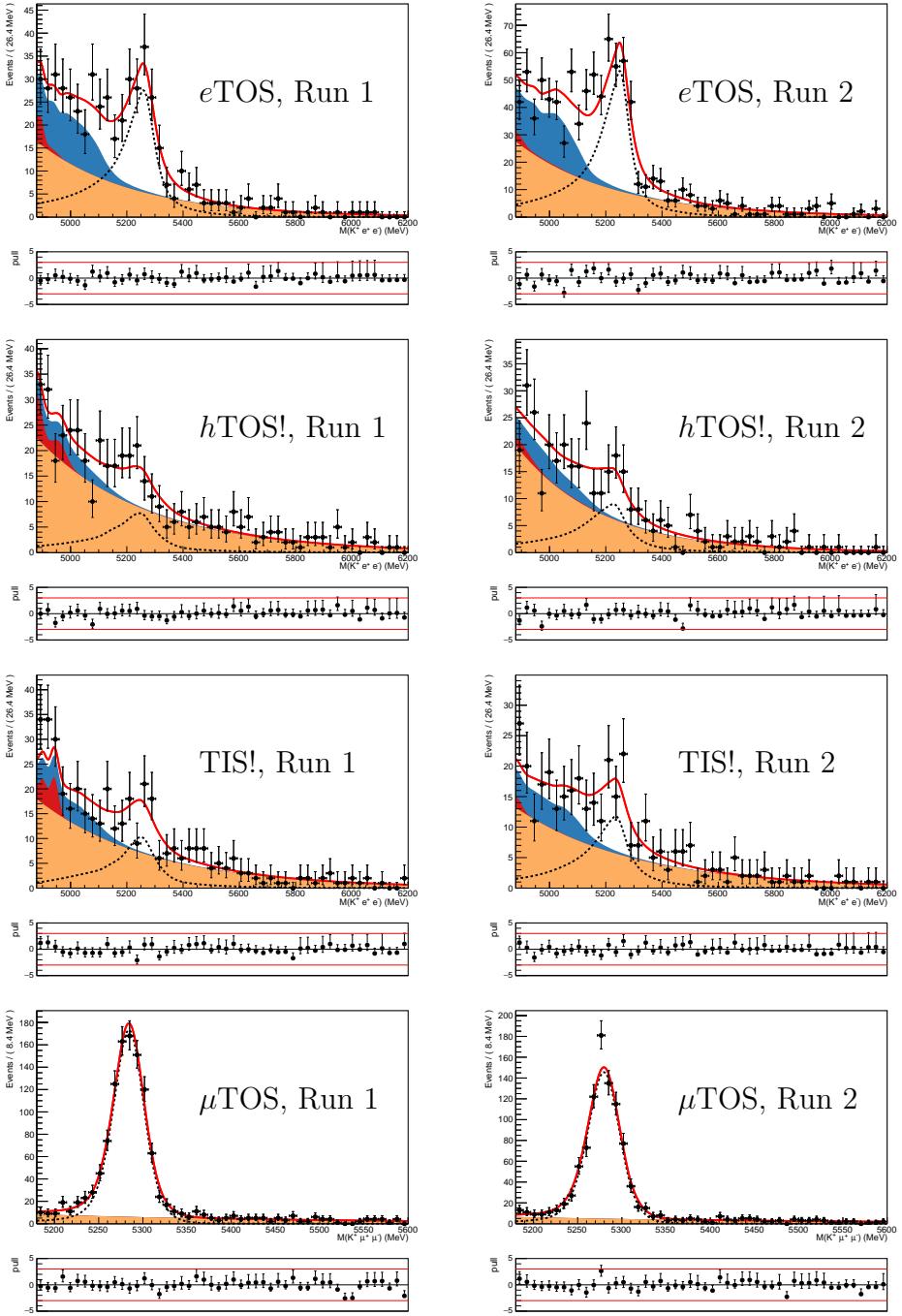


Figure 87: Example of a toy experiment result. The red solid line shows the fit model, the dotted black line is the signal component, the dark-blue filled area represents partially reconstructed background, the red filled area shows the leakage from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and the orange filled area shows the combinatorial background.

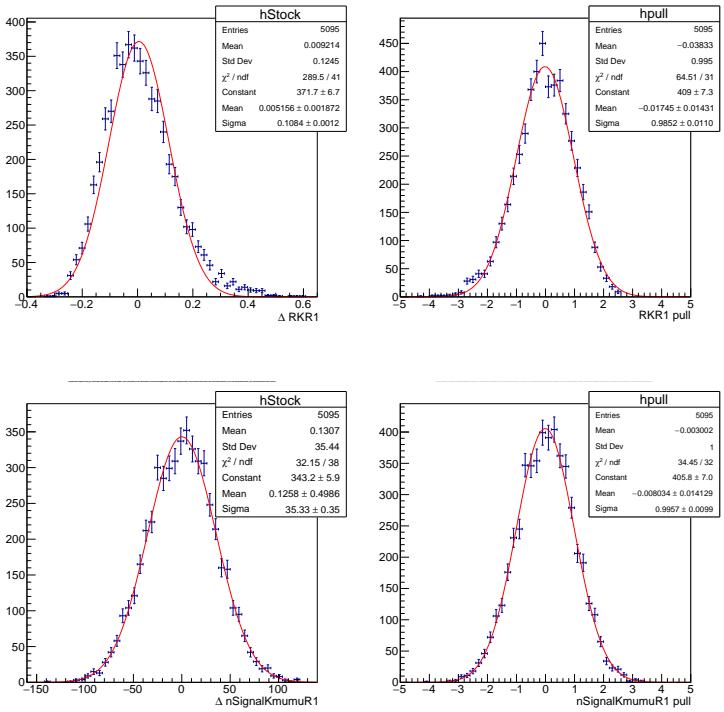


Figure 88: Distribution of the (left) residuals and (right) pulls for the fitted parameters (top) R_K and (bottom) $N_{K+}\mu^+\mu^-_R$. This toy study assumes Run 1 yields corresponding to $R_K = 1$.

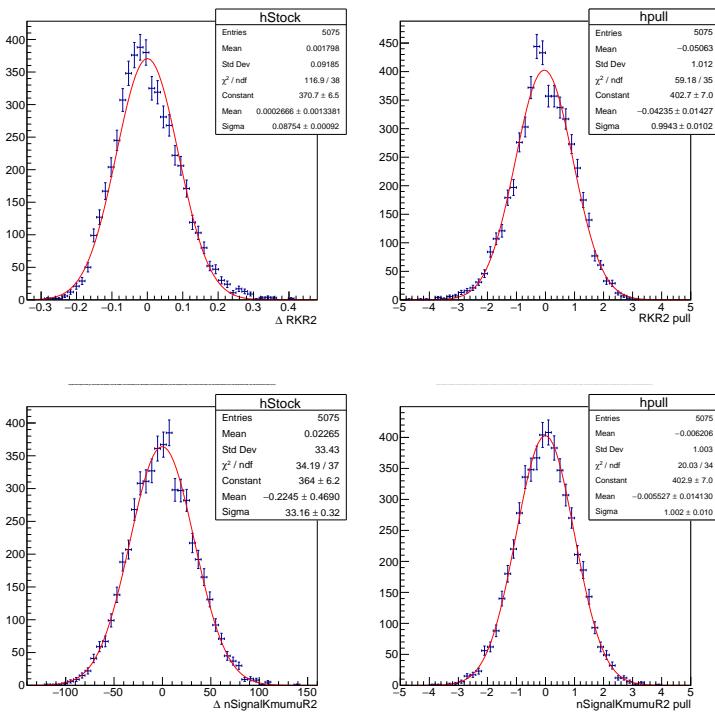


Figure 89: Distribution of the (left) residuals and (right) pulls for the fitted parameters (top) R_K and (bottom) $N_{K^+ \mu^+ \mu^-}^2$. This toy study assumes Run 2 yields corresponding to $R_K = 1$.

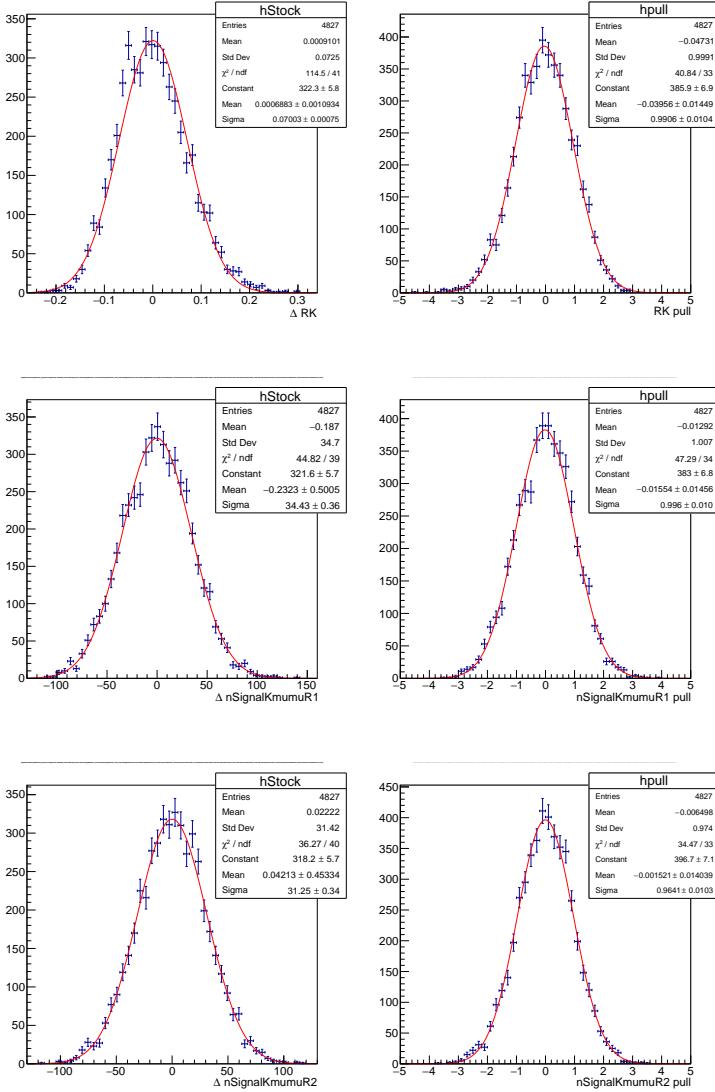


Figure 90: Distribution of the (left) residuals and (right) pulls for the fitted parameters (top) R_K , (middle) $N_{K^+\mu^+\mu^-}^1$ and (bottom) $N_{K^+\mu^+\mu^-}^2$. This toy study assumes Run 1 and Run 2 yields corresponding to $R_K = 1$.

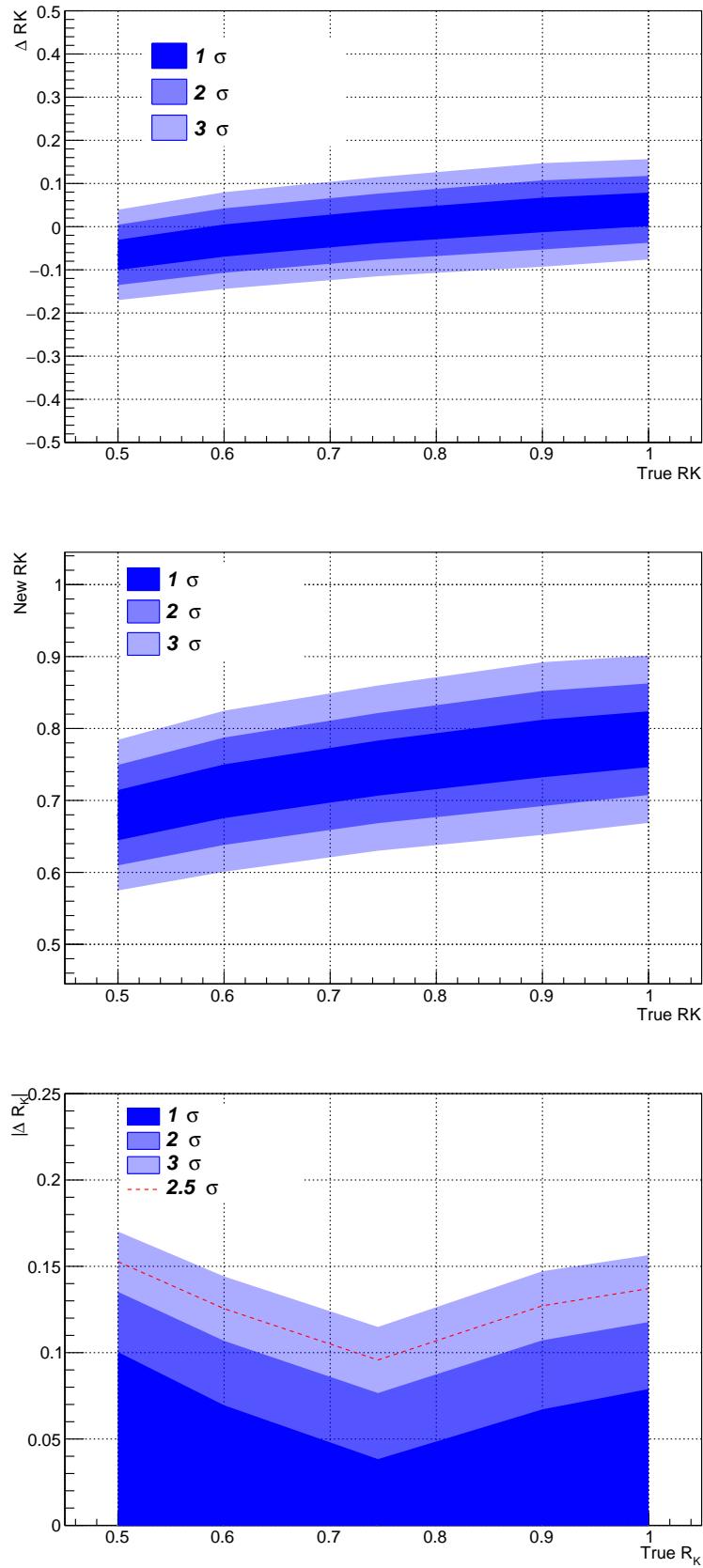


Figure 91: Top: Expected spread of values for 80% overlap samples as a function of the true value of R_K . Middle: Same figure shifted by 0.745. Bottom: Expected spread of $|\Delta R_K|$ as a function of the true value of R_K , the band corresponding to 2.5σ is defined by the dashed red line.

1799 9 Fit result

1800 9.1 Fits to the normalisation modes

1801 The result of the fits to the normalisation modes are shown in this section. Fig. 92
 1802 and Fig. 93 show the result of the fit to the $m_{\text{DTF}}^{J/\psi}(K^+\ell^+\ell^-)$ distribution of selected
 1803 $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ candidates, respectively.

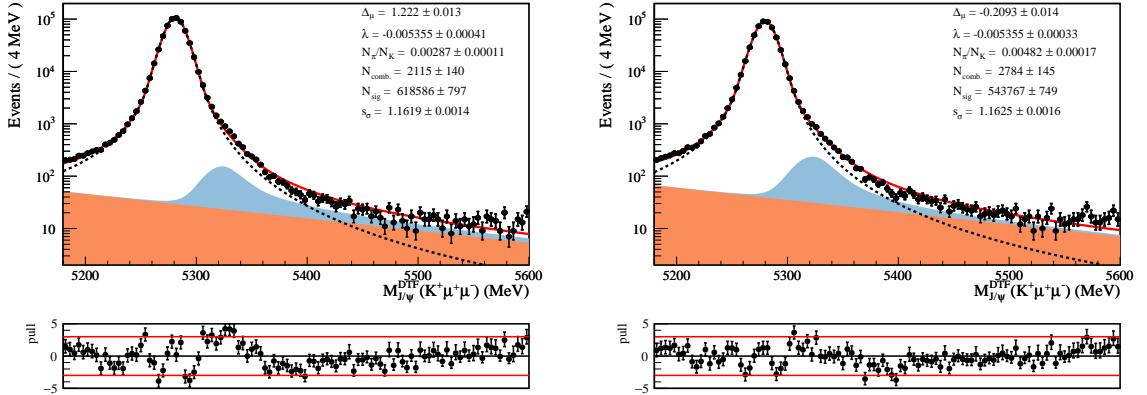


Figure 92: Fits to the $m_{\text{DTF}}^{J/\psi}(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ selected candidates in Run 1 (left) and Run 2 (right) samples. The red solid line shows the fit model, the dotted black line is the signal component, the light-blue filled area represents misidentified $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ events and the orange filled area shows the combinatorial background. Some large pulls are observed but have a completely negligible impact on the yield extracted.

1804 Additionally, the result of the fits to the $m(K^+\ell^+\ell^-)$ distributions of the normalisation
 1805 modes, which are used to constrain the parameters $\Delta\mu$ and s_σ (and s_α for the electron
 1806 mode), are shown in Fig. 94 for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and Fig. 95 for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$.
 1807 As can be seen from the fitted values of the parameters $\Delta\mu$ and s_σ , the simulation tends
 1808 to underestimate the mass resolution observed in data by $\sim 10 - 15\%$, whereas it is able
 1809 to reproduce to the position of the mass peak within $\sim 1\%$ of the value measured in
 1810 the data. The righthand side tail of the distributions are also more prominent in data,
 1811 which is reflected in the value obtained for the parameter s_α . In general, the agreement
 1812 between data and simulation is found to be worse for Run 2. The ratio between the
 1813 $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ yields obtained from the fit with and without the constraint are
 1814 found compatible with that expected from simulation, see Tab. 27.

1815 Finally, Fig. 96 and Fig. 97 show the result of the fit to the $m_{\text{DTF}}^{\psi(2S)}(K^+\ell^+\ell^-)$ distribution
 1816 of selected $B^+ \rightarrow K^+\psi(2S)(\mu^+\mu^-)$ and $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ candidates. The signal
 1817 yield found in these fits are shown in Tab. 28. These candidates are used in various
 1818 crosschecks of the analysis.

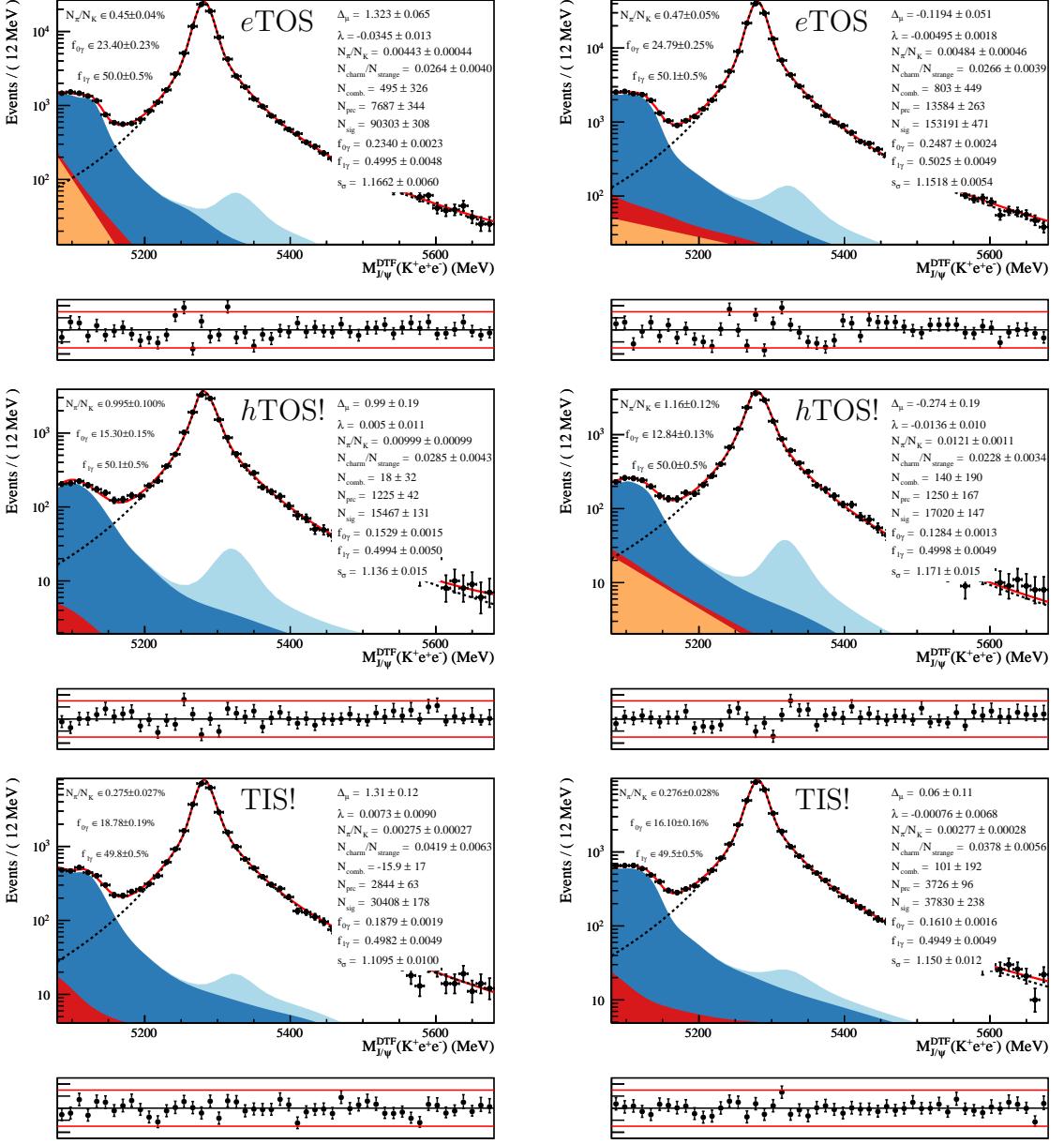


Figure 93: Fits to the $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+J/\psi(e^+e^-)$ selected candidates in Run 1 (left) and Run 2 (right) samples, for eTOS (top), hTOS! and TIS! trigger categories. The red solid line shows the fit model, the dotted black line is the signal component, the dark-blue and red filled areas show the strange and charm partially reconstructed backgrounds respectively, the light-blue filled area represents misidentified $B^+ \rightarrow \pi^+J/\psi(e^+e^-)$ events and the orange filled area shows the combinatorial background.

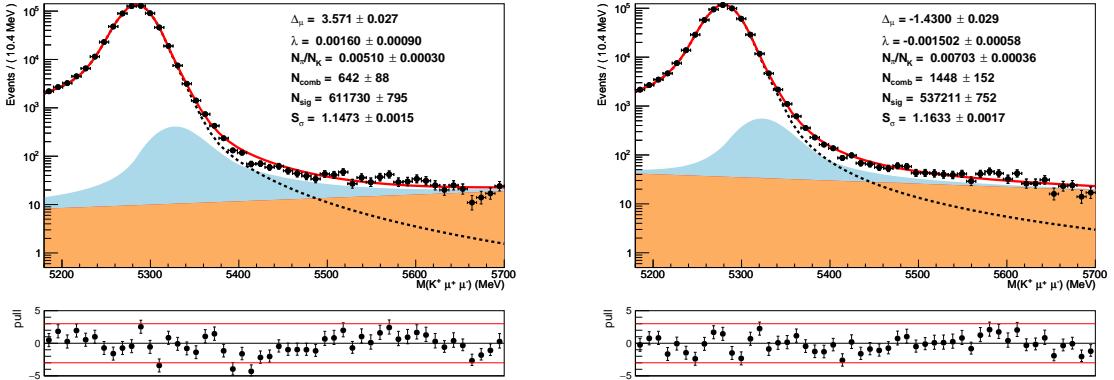


Figure 94: Fits to the $m(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ selected candidates in Run 1 (left) and Run 2 (right) samples. The red solid line shows the fit model, the dotted black line is the signal component, the light-blue filled area represents misidentified $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ events and the orange filled area shows the combinatorial background. These fits are used to constrain the parameters $\Delta\mu$ and s_σ in the fit to the rare mode.

Table 27: Ratio of the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ yield extracted from the fit to $m(K^+e^+e^-)$ over that obtained from the fit to $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-)$, and comparison with the ratio expected from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulation. Note that because of the different fitting window, mass resolution and the extra requirement of $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-) > 5.2 \text{ GeV}/c^2$ imposed for the fit to the unconstrained mass, this ratio is expected to be different from unity.

	Run 1		Run 2	
	Simulation	Data	Simulation	Data
$e\text{TOS}$	0.9743 ± 0.0007	0.9760 ± 0.0033	0.9740 ± 0.0004	0.9771 ± 0.0025
$h\text{TOS!}$	0.9636 ± 0.0018	0.9700 ± 0.0079	0.9654 ± 0.0014	0.9585 ± 0.0075
TIS!	0.9703 ± 0.0013	0.9741 ± 0.0057	0.9700 ± 0.0010	0.9699 ± 0.0051

Table 28: Signal yields obtained from the fit to the $m_{\text{DTF}}^{\psi(2S)}(K^+\ell^+\ell^-)$ distribution of the $B^+ \rightarrow K^+\psi(2S)(\ell^+\ell^-)$ selected candidates.

	Run 1	Run 2
μTOS	62617 ± 255	53708 ± 237
$e\text{TOS}$	9459 ± 129	15731 ± 167
$h\text{TOS!}$	468 ± 33	536 ± 35
TIS!	2428 ± 71	3065 ± 81

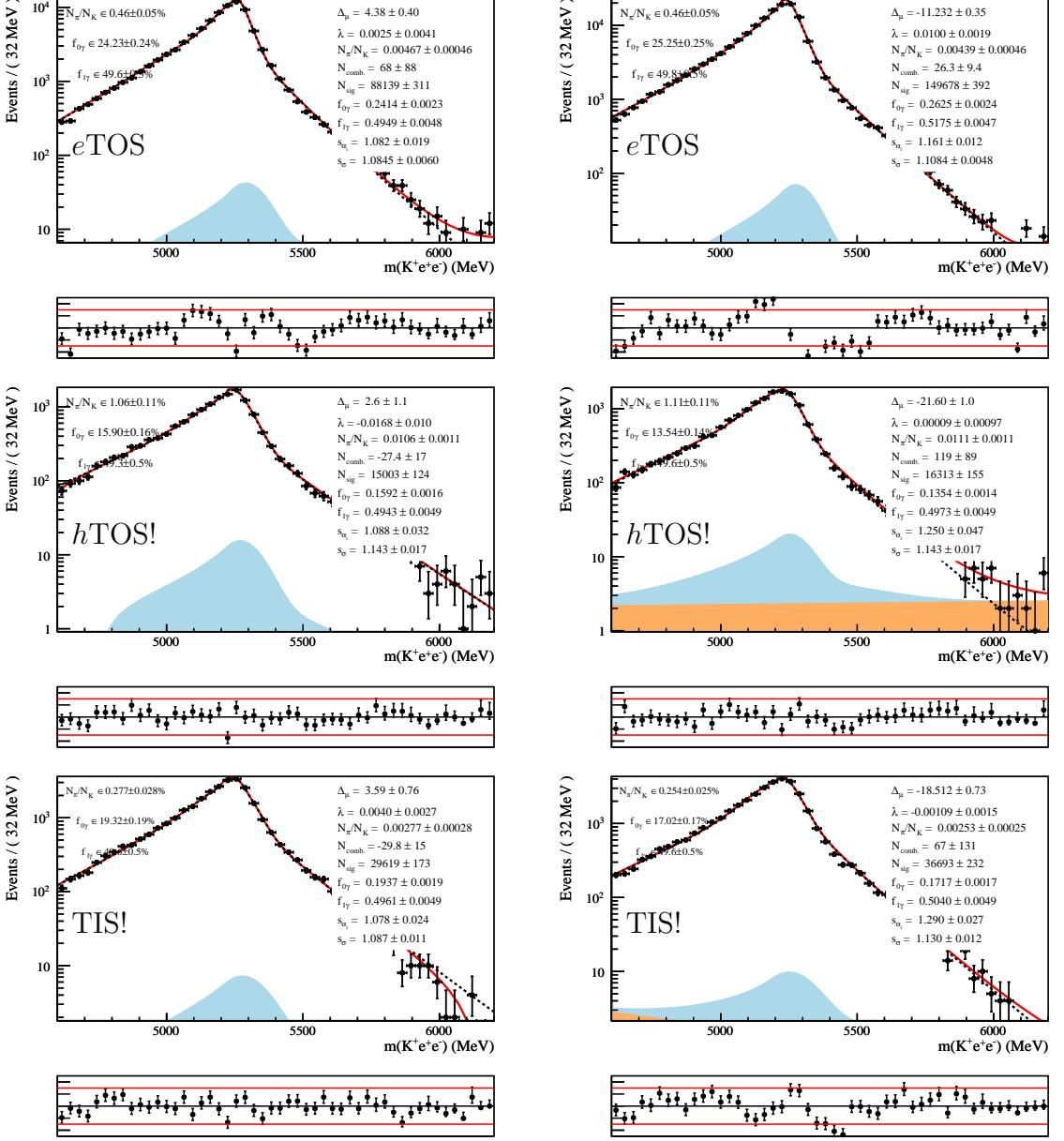


Figure 95: Fits to $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ selected candidates in Run 1 (left) and Run 2 (right) samples, for $e\text{TOS}$ (top), $h\text{TOS!}$ and $T\text{IS!}$ trigger categories. The red solid line shows the fit model, the dotted black line is the signal component, the light-blue filled area represents misidentified $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ events and the orange filled area shows the combinatorial background. These fits are used to constrain the parameters $\Delta\mu$ and s_σ in the fit to the rare mode.

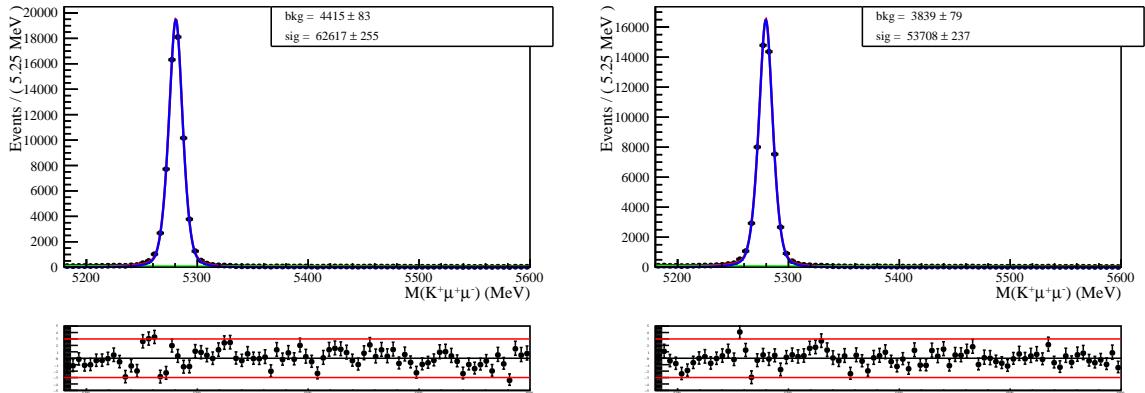


Figure 96: Fits to the $m_{DTF}^{\psi(2S)}(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+\psi(2S)(\mu^+\mu^-)$ selected candidates in Run 1 (left) and Run 2 (right) samples. The red line shows the fit model, the blue line is the signal component and the green line shows the combinatorial background.

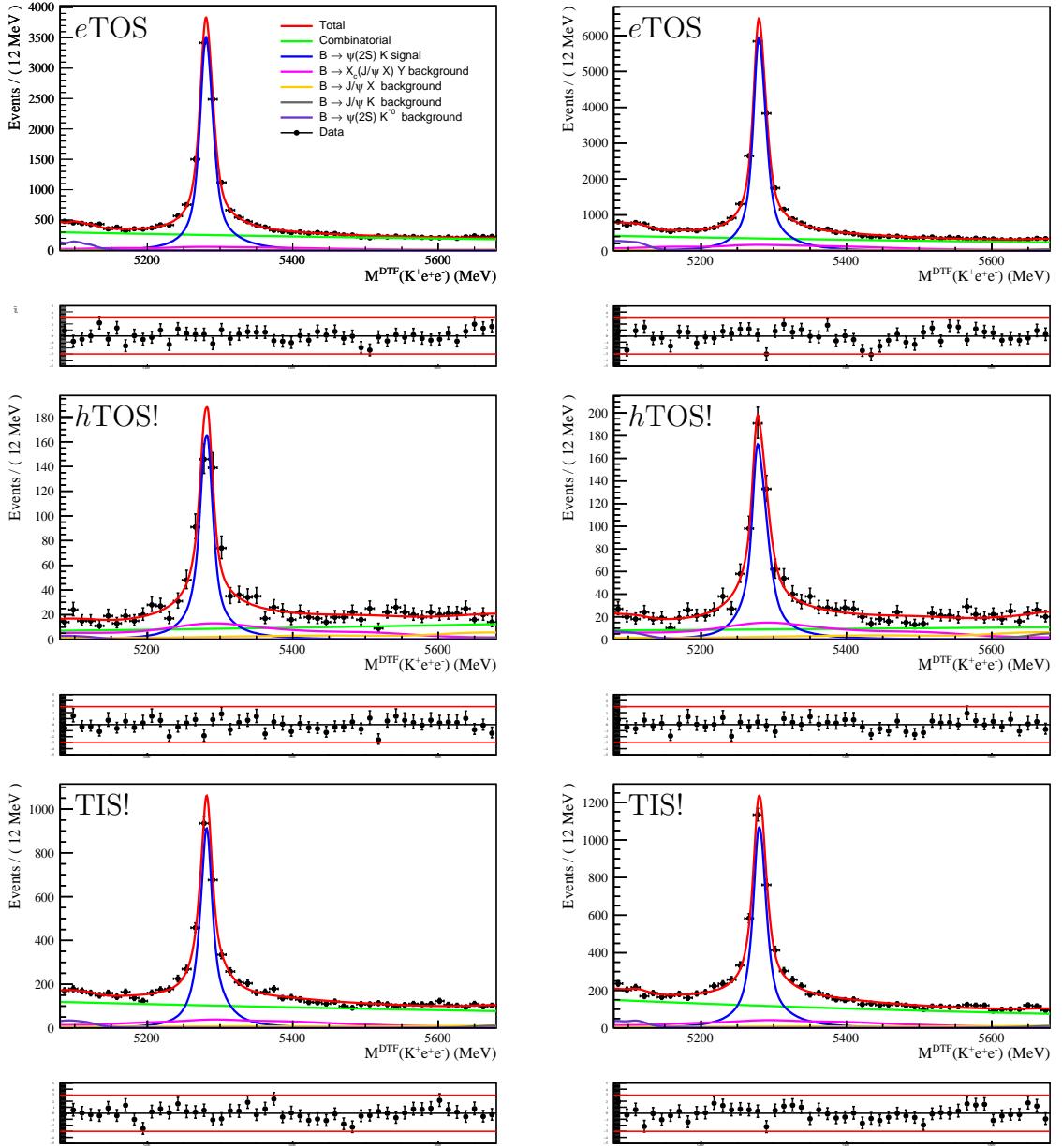


Figure 97: Fits to the $m_{DTF}(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ selected candidates in Run 1 (left) and Run 2 (right) samples, for *eTOS* (top), *hTOS!* and *TIS!* trigger categories.

1819 **9.2 R_K fit results**

1820 Following the strategy described in Section 8, a simultaneous fit to the invariant mass of
 1821 the selected $B^+ \rightarrow K^+ \ell^+ \ell^-$ candidates in the different trigger categories for the Run 1 and
 1822 Run 2 samples is used to extract the value of R_K . The c_K^{rt} and r_{prc}^{rt} factors are constrained
 1823 to the values in Tab. 29, computed from the observed yields of $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$
 1824 obtained in Section 9.1 and the ratio of efficiencies from Tab. 20 and 21. The constraint
 1825 is implemented using two multi-dimensional Gaussians with correlation matrices given
 1826 in Tab. 48 and Tab. 46, respectively.

Table 29: Values of the c_K^{rt} and r_{prc}^{rt} factors used as central values for the multidimensional constraints in the R_K fit.

$c_K^{R1,e\text{TOS}}$	0.1456	$c_K^{R2,e\text{TOS}}$	0.2839
$c_K^{R1,h\text{TOS!}}$	0.0500	$c_K^{R2,h\text{TOS!}}$	0.0715
$c_K^{R1,\text{TIS!}}$	0.0539	$c_K^{R2,\text{TIS!}}$	0.0803
$r_{prc}^{R1,h\text{TOS!}}$	0.909	$r_{prc}^{R1,h\text{TOS!}}$	0.722
$r_{prc}^{R1,\text{TIS!}}$	1.074	$r_{prc}^{R1,\text{TIS!}}$	0.933

1827 The results of the different fits performed to the data are summarised in Tab. 30.
 1828 Following the staged unblinding proposed in Section 3.3, the data were first fitted allowing
 1829 for two independent values of R_K to float (`RKFitYears`), associated to the Run 1 and Run 2
 1830 samples respectively (R_K^{R1} and R_K^{R2}). Then the combined R_K was determined from the
 1831 nominal configuration of the fit (`RKFitComb`), which allows only a common R_K parameter
 1832 to float. The result of this nominal fit is compared to the invariant mass distribution of
 1833 the selected $B^+ \rightarrow K^+ \ell^+ \ell^-$ candidates in Fig. 223.

1834 In addition, to check the compatibility between the different trigger categories, the fit
 1835 was repeated letting three independent R_K values float (`RKFitTrigger`): R_K^{eTOS} , $R_K^{\text{hTOS!}}$
 1836 and $R_K^{\text{TIS!}}$. Finally, six different values of R_K were allowed to float (`RKFitAll`), each of
 1837 them corresponding to a trigger category in a specific run. The full list of parameters
 1838 determined under each configuration, together with their invariant mass projections, can
 1839 be found in Appendix O.

Table 30: Results of the fit to the data for the different configurations of the fit described in the text.

Parameter	RKFitAll	RKFitTrigger	RKFitYears	RKFitComb
$R_K^{e\text{TOS},\text{R1}} / R_K^{e\text{TOS}} / R_K^{\text{R1}} / R_K$	$0.79_{-0.10}^{+0.11}$	$0.866_{-0.063}^{+0.071}$	$0.711_{-0.072}^{+0.084}$	$0.848_{-0.055}^{+0.061}$
$R_K^{e\text{TOS},\text{R2}} / R_K^{\text{R2}}$	$0.903_{-0.081}^{+0.094}$	—	$0.934_{-0.078}^{+0.090}$	—
$R_K^{h\text{TOS!},\text{R1}} / R_K^{h\text{TOS!}}$	$0.56_{-0.12}^{+0.19}$	$0.87_{-0.17}^{+0.24}$	—	—
$R_K^{h\text{TOS!},\text{R2}}$	$1.28_{-0.35}^{+0.65}$	—	—	—
$R_K^{\text{TIS!},\text{R1}} / R_K^{\text{TIS!}}$	$0.60_{-0.10}^{+0.13}$	$0.746_{-0.091}^{+0.128}$	—	—
$R_K^{\text{TIS!},\text{R2}}$	$0.97_{-0.21}^{+0.32}$	—	—	—
nSignalKmumuR1	1027_{-35}^{+36}	1043_{-35}^{+35}	1027_{-35}^{+36}	1044_{-35}^{+35}
nSignalKmumuR2	916_{-33}^{+34}	900_{-32}^{+33}	916_{-33}^{+34}	899_{-32}^{+33}
nPartRecoKeeETOSR1	68_{-35}^{+32}	45_{-36}^{+34}	72_{-41}^{+32}	44_{-37}^{+35}
nPartRecoKeeETOSR2	191_{-26}^{+23}	194_{-24}^{+22}	192_{-28}^{+24}	194_{-24}^{+22}
nCombKeeETOSR1	96_{-46}^{+47}	129_{-42}^{+44}	78_{-44}^{+54}	127_{-44}^{+46}
nCombKeeETOSR2	75_{-25}^{+36}	69_{-22}^{+30}	80_{-26}^{+37}	67_{-21}^{+29}
nCombKeeHTOSR1	126_{-42}^{+37}	170_{-27}^{+25}	150_{-23}^{+24}	168_{-21}^{+21}
nCombKeeHTOSR2	109_{-29}^{+28}	83_{-27}^{+25}	85_{-18}^{+18}	81_{-18}^{+18}
nCombKeeTISR1	48_{-27}^{+28}	76_{-24}^{+25}	66_{-21}^{+24}	86_{-21}^{+22}
nCombKeeTISR2	83_{-33}^{+32}	54_{-20}^{+26}	77_{-17}^{+18}	71_{-16}^{+17}
nCombKmumuR1	208_{-21}^{+22}	205_{-20}^{+22}	208_{-21}^{+22}	205_{-20}^{+22}
nCombKmumuR2	190_{-20}^{+21}	193_{-20}^{+21}	190_{-20}^{+21}	193_{-20}^{+21}

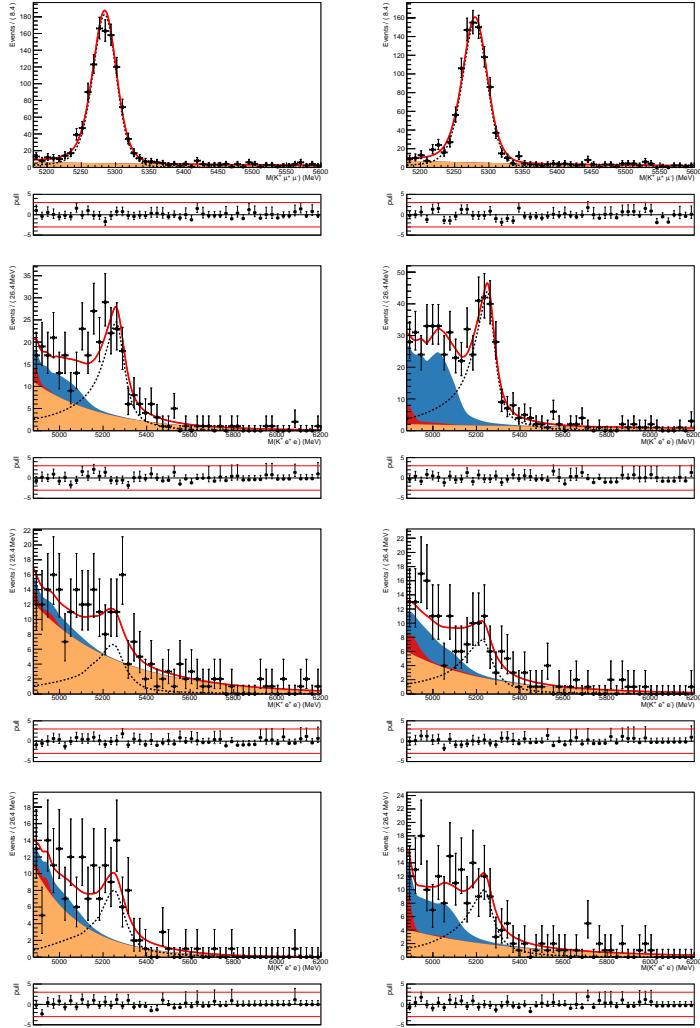


Figure 98: The fit to the $m(K^+\ell^+\ell^-)$ invariant mass distribution of selected candidates in the (left) Run 1 and (right) Run 2 data samples. The top row shows the fit to the muon modes and the subsequent rows the fits to the electron modes triggered by (second row) one of the electrons, (third row) the kaon from the B^+ decay and (last row) by other particles in the event. The points show the LHCb data, the red line the total fit, the black-dashed line the $B^+ \rightarrow K^+\ell^+\ell^-$ signal component, the orange line the combinatorial background, the blue line the background from exclusive b -hadron decays and the dark red the radiative tail of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events.

1840 10 Systematic uncertainties

1841 The systematic uncertainties and their correlations are presented in this section. The
 1842 following paragraphs introduce the methods and define the quantities common to all
 1843 systematic uncertainty calculations.

1844 In this analysis, the statistical uncertainties refer to either the uncertainty on the
 1845 yields $N_{K\mu\mu}^r$, N_{Kee}^{rt} or $N^r(B^+ \rightarrow K^+\psi(2S)(e^+e^-))$, $N^{rt}(B^+ \rightarrow K^+\psi(2S)(e^+e^-))$. All other
 1846 uncertainties are referred to as systematic uncertainties, including the uncertainties on the
 1847 yields $N^r(B^+ \rightarrow K^+J/\psi(\mu^+\mu^-))$, $N^{rt}(B^+ \rightarrow K^+J/\psi(e^+e^-))$, because these are correlated
 1848 to the statistical uncertainties, given that the majority of the calibration histograms are
 1849 extracted from the J/ψ decay modes.

1850 To illustrate how systematic uncertainties are computed, consider an efficiency com-
 1851 puted in the two run conditions. Its true value is ε_1 for Run 1 and ε_2 for Run 2, and
 1852 these efficiencies are estimated in the nominal way to be $\bar{\varepsilon}_1$ for Run 1 and $\bar{\varepsilon}_2$ for Run 2.
 1853 They are also computed in n alternative ways, taking n alternative values ε_1^i and ε_2^i ,
 1854 $1 \leq i \leq n$. Then, the systematic uncertainties on ε_1 and on ε_2 are denoted σ_1 and σ_2 and
 1855 are computed as the square root of the estimated variance on ε_1 and ε_2 :

$$\sigma_1 = \left(\frac{1}{n} \sum_{i=1}^n (\varepsilon_1^i - \bar{\varepsilon}_1)^2 \right)^{\frac{1}{2}}, \quad \sigma_2 = \left(\frac{1}{n} \sum_i (\varepsilon_2^i - \bar{\varepsilon}_2)^2 \right)^{\frac{1}{2}}. \quad (36)$$

1856 The covariance, $\text{cov}_{1,2}$, between ε_1 and ε_2 is computed using:

$$\text{cov}_{1,2} = \sum_{i=1}^n (\varepsilon_1^i - \bar{\varepsilon}_1) \cdot (\varepsilon_2^i - \bar{\varepsilon}_2), \quad (37)$$

1857 and the correlation factor, $\text{corr}_{1,2}$, between ε_1 and ε_2 is computed using:

$$\text{corr}_{1,2} = \frac{\text{cov}_{1,2}}{\sigma_1 \sigma_2}. \quad (38)$$

1858 The fractional error matrix $\boldsymbol{\sigma}$ is then defined here as a diagonal matrix containing relative
 1859 uncertainties on the diagonal. In the case of ε_1 and ε_2 , it would read:

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_1 / \bar{\varepsilon}_1 & \text{corr}_{1,2} \\ \text{corr}_{1,2} & \sigma_2 / \bar{\varepsilon}_2 \end{pmatrix} \quad (39)$$

1860 In practice, the covariance matrix is a 6×6 matrix, cov_{ij} , with entries corresponding to
 1861 the three trigger categories in the two runs. Below these covariance matrices are denoted
 1862 $\text{cov}_{ij}^{r_{J/\psi}}$, $\text{cov}_{ij}^{R_{\psi(2S)}}$, $\text{cov}_{ij}^{R_K}$ for the $r_{J/\psi}$, $R_{\psi(2S)}$ and R_K variants, respectively. The 6×6
 1863 fractional error matrices, $\boldsymbol{\sigma}$, are derived from these covariance matrices.

1864 Systematic uncertainties and their correlations are computed on three different quanti-
 1865 ties:

- 1866 1. On the single ratio $r_{J/\psi}$ in each trigger category and in each data taking condition,
 1867 yielding a six by six covariance matrix;

- 1868 2. On the double ratio R_K , systematic uncertainties impact the quantity c_K^{rt} , defined
 1869 in Eq. 8, giving rise to a six by six covariance matrix used as an input into the fit
 1870 for the rare mode;
- 1871 3. In a similar way as for R_K , systematic uncertainties on the double ratio $R_{\psi(2S)}$
 1872 impacting on $c_{\psi(2S)}^{rt}$.

1873 The following sections discuss various sources of bias and present the systematic
 1874 uncertainties they cause on these three quantities. These uncertainties are always reported
 1875 as fractional error matrices.

1876 10.1 Size of the calibration and simulation samples

1877 This systematic uncertainty is extracted using *super-*
 1878 *Lenin/src/mainSystematic1New.cc*.

1879 The statistical uncertainty related to the finite size of the normalisation channels and
 1880 the calibration histograms is classified as a systematic uncertainty and is computed in this
 1881 section. Most of the calibration histograms, for example the trigger efficiency histograms
 1882 or electron PID calibration histograms discussed in Section 6.6 and in Section 6.4.2, are
 1883 computed from the normalisation modes $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$
 1884 themselves, so that correlations have to be properly taken into account. In order to do
 1885 this, a bootstrapping method is used. To generate a bootstrapped version of the analysis,
 1886 each data or simulated event is given a weight that is distributed according to a Poisson
 1887 distribution of unit mean. All the calibration histograms are re-extracted, all efficiencies
 1888 are re-computed and all the fits to the control modes are re-performed using these weights.
 1889 This procedure is repeated 100 times, hence generating 100 versions of $r_{J/\psi}$, c_K^{rt} , and $c_{\psi(2S)}^{rt}$.
 1890 The 100 bootstrapped versions of $r_{J/\psi}$, c_K^{rt} , and $c_{\psi(2S)}^{rt}$ are used to extract correlation
 1891 matrices according to Eq. 39. The fractional error matrices for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$ are
 1892 reported in Tab. 34.

1893 The K and μ PID efficiencies are computed from samples which are very weakly
 1894 correlated to the samples used in the rest of the analysis. Therefore, they are simply
 1895 oscillated 100 times according to their errors, assumed to be Gaussian. Some other
 1896 samples, which are considered as separate in the analysis for practical reasons, are however
 1897 correlated with each other: that is the case for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events from the
 1898 ElectronIDBu2JpsieeK stripping line that is used to extract the electron PID efficiencies,
 1899 and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events from the Bu2LLKeeLine2 line. This is also the case for
 1900 the generator level simulation samples and the reconstruction level simulation samples. For
 1901 these samples, the correlation is taken into account by ensuring that an event appearing
 1902 in two separate samples is associated to the same set of 100 Poisson weights. This can be
 1903 done as each event is uniquely identifiable using its associated event number.

1904 It is checked that the errors on the yields $N(B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-))$ computed from
 1905 the bootstrapping method are compatible with those extracted from the likelihood profile
 1906 (using the HESSE algorithm from MINUIT). This comparison is reported in Tab. 31. The

1907 uncertainties are typically larger for Run 1 than for Run 2. Also, in the case of Run 1,
 1908 the uncertainties are larger on the single ratio than on the double ratios, whereas the
 1909 contrary is true for Run 2. This is because of the large statistical uncertainties of the HLT
 1910 calibration histograms, used for Run 1 but not Run 2, that cancel well in the double ratio.
 1911 All samples are completely independent between the two data taking conditions, so that
 1912 null correlations are expected between Run 1 and Run 2. Due to the limited number of
 1913 bootstrapped samples, the correlations between Run 1 and Run 2 are non zero but very
 1914 small (below 10%), and are set to 0%.

Table 31: Error on the yields from the fits to $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ in all trigger categories, estimated from the likelihood ($\sigma_{\mathcal{L}}$) and from the bootstrapping method (σ_{bt}).

trigger	Run 1		Run 2	
	σ_{bt}	$\sigma_{\mathcal{L}}$	σ_{bt}	$\sigma_{\mathcal{L}}$
μ TOS	0.12	0.13	0.12	0.13
e TOS	0.31	0.34	0.28	0.26
h TOS!	0.84	0.83	0.82	0.77
TIS!	0.63	0.59	0.53	0.53

1915 In principle, there is the possibility that the shift and scale factors are also correlated
 1916 to c_K^{rt} , and this should then be taken into account in the fit to the rare mode. However, it
 1917 is demonstrated in two ways that it is not the case. First, when fitting the unconstrained
 1918 mass of $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$, the correlation factors are extracted from the minimisation
 1919 of the likelihood, showing that the correlations of the shift and scale factors with the
 1920 yields are below 5% and hence negligible. Second, the bootstrapping method is used to
 1921 compute the correlation of the scale and shift factors with c_K^{rt} directly, yielding again small
 1922 correlation factors below the 10% level. The complete absence of correlation between the
 1923 yield and shape factors can be understood from the fact that, after the full selection chain
 1924 is applied, the control samples are very clean, and the signal yield is almost equal to the
 1925 total data sample size.

1926 10.2 Kinematic reweighting scheme

1927 This systematic uncertainty is extracted using *superLenin/src/mainSystematic2.cc*.

1928 As detailed in Section 6.8, the distribution for $p_T(B^+)$, $\eta(B^+)$, χ_{DV}^2 and $\chi_{IP}^2(B^+)$ are
 1929 imperfect in the simulation and are corrected in all simulation samples using $B^+ \rightarrow$
 1930 $K^+ J/\psi(\mu^+\mu^-)$ events from the data in the μ TOS category. However, this method is
 1931 expected to cause some biases. For example, if the L0Muon correction as a function of $p_T(\mu)$
 1932 is imperfect, this will bias the simulated $p_T(B^+)$ which will in turn bias the kinematic
 1933 weights. Another possible bias stems from the fact that the discrepancy between data
 1934 and simulation in the χ_{DV}^2 and $\chi_{IP}^2(B^+)$ distributions can be influenced by bremsstrahlung

Table 32: Fractional error matrices showing the fractional uncertainties and their correlation factors (in percent) for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$, due the statistical fluctuations of the control samples, simulation samples and calibration histograms.

$\boldsymbol{r}_{J/\psi}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
1.28	10.59	16.27	0.00	0.00	0.00
	3.04	14.24	0.00	0.00	0.00
		2.38	0.00	0.00	0.00
			0.88	26.54	21.59
				1.66	12.42
					1.43
\boldsymbol{R}_K					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
1.14	-3.82	7.07	0.00	0.00	0.00
	2.16	-6.73	0.00	0.00	0.00
		1.70	0.00	0.00	0.00
			0.95	0.01	27.65
				1.78	15.35
					1.77
$\boldsymbol{R}_{\psi(2S)}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
1.01	-1.26	6.73	0.00	0.00	0.00
	3.32	10.92	0.00	0.00	0.00
		1.76	0.00	0.00	0.00
			0.77	-2.91	2.29
				3.67	-2.83
					1.63

in the VELO for the electron mode, but not in the muon mode, so that using the muon mode to correct the electron mode might yield to inaccuracies.

To study this potential bias, kinematic distributions in the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode and in the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode are studied using various sets of kinematic weights. For the muon mode, two sets of weights are compared:

1. μ TOS weights: these are the nominal weights computed from the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ sample in the μ TOS trigger category;
2. μ TIS weights: these weights are computed using the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ sample

1943 in the inclusive TIS trigger category.

1944 For the electron mode, four sets of weights are compared:

- 1945 1. μ TOS weights: as for the muon mode, these are the nominal weights;
- 1946 2. μ TIS weights: same as the alternative weights for the muon mode;
- 1947 3. e TOS weights: these weights are computed from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ sample
1948 in the e TOS trigger category;
- 1949 4. e TIS weights: these weights are computed from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ sample
1950 in the inclusive TIS trigger category;
- 1951 5. mixed weights: the correction for $p_T(B^+)$, $\eta(B^+)$ and $\chi_{\text{IP}}^2(B^+)$ is taken from the
1952 $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ sample in the μ TOS trigger category, but the χ_{DV}^2 is taken
1953 from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ sample in the e TOS trigger category.

1954 Apart from the samples used to compute them, the procedure to subtract the background
1955 and compute the weights is exactly the same as that detailed in Section 6.8.

1956 Efficiency corrected yield plots are extracted using the various sets of weights and are
1957 shown in Fig. 99 for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode and in Fig. 100 for the muon mode.
1958 These plots correspond to a data over simulation ratio of kinematic distributions, and the
1959 procedure used to extract them is detailed in Section 11.2. The following observations can
1960 be made:

- 1961 • In both the electron and muon sample, using the μ TOS weights or μ TIS weights
1962 makes very little difference;
- 1963 • In the electron mode, using the e TOS weights improves the data versus simulation
1964 agreement with regards to the χ_{DV}^2 distributions, but worsens the $\chi_{\text{IP}}^2(B^+)$
1965 distribution. It also flattens the $\eta(B^+)$ efficiency corrected yield plot;
- 1966 • In the electron mode, using the mixed weights yields a very good agreement between
1967 data and simulation regarding the χ_{DV}^2 and $\chi_{\text{IP}}^2(B^+)$ distributions, without impacting
1968 the $\eta(B^+)$ and $p_T(B^+)$ distributions.

1969 As mentioned above, it can be expected that the agreement between data and simulation
1970 in the χ_{DV}^2 distribution is improved when using weights from the electron mode directly.
1971 The $\chi_{\text{IP}}^2(B^+)$ distribution, however, is difficult to correct from the electron mode because
1972 this variable is very correlated to the B^+ mass, meaning that subtracting the background
1973 tends to bias this distribution.

1974 The total efficiencies, as well as the efficiency ratios relevant for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$,
1975 are computed using the different kinematic weights and are reported in Tab. 33. The ratios
1976 of efficiencies reported in this table are used to extract correlation matrices according
1977 to Eq. 39. The fractional error matrices for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$ are reported in Tab. 34.

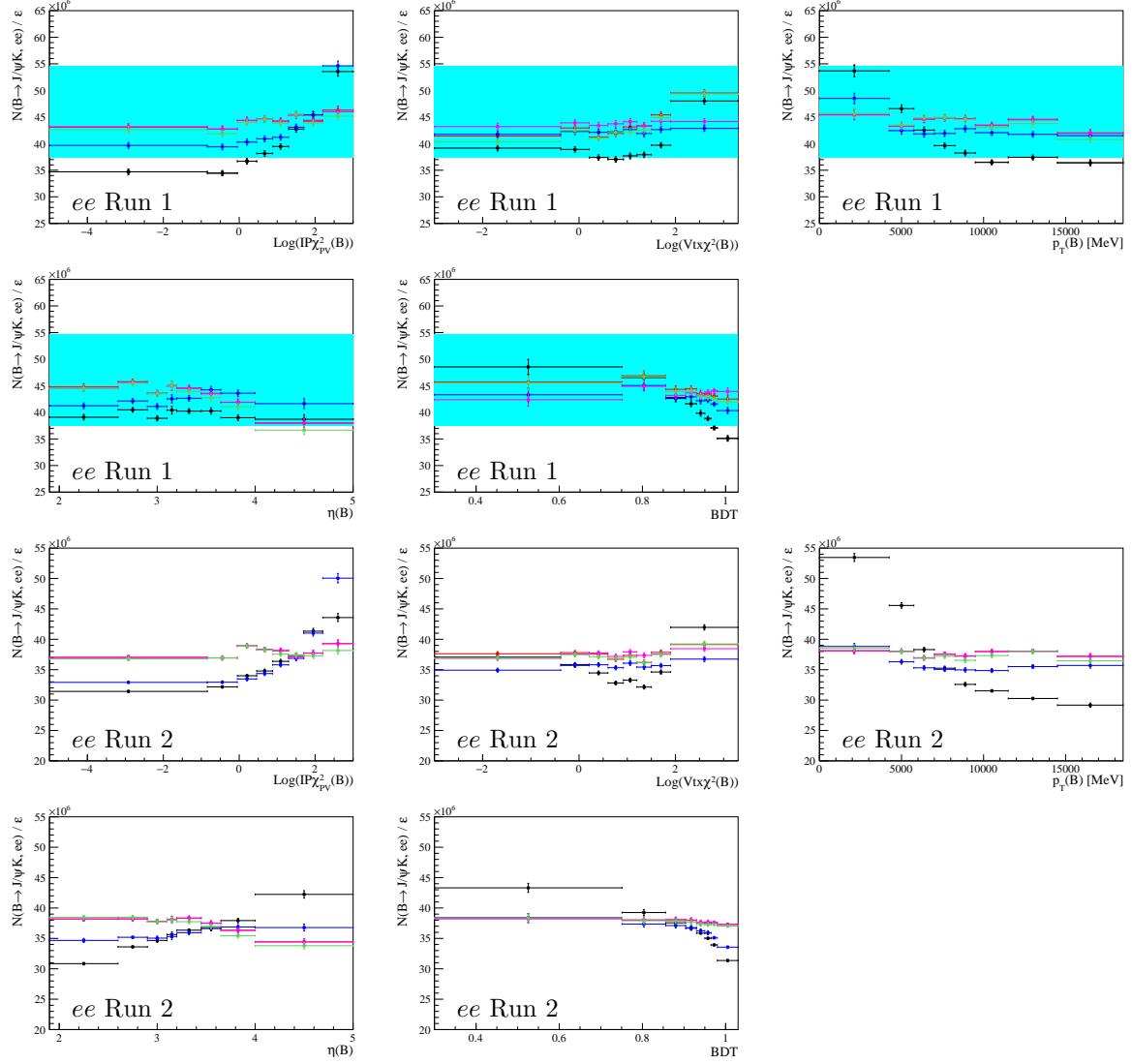


Figure 99: Efficiency corrected yield as a function of several kinematic variables for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in Run 1 (top 5 plots) and Run 2 (bottom 5 plots). To compute the efficiencies, various kinematic weights are used and correspond to the different colours: no kinematic weights at all (black), the nominal μ TOS weights (red), μ TIS weights (green), e TOS weights (blue) and the mixed weights (magenta).

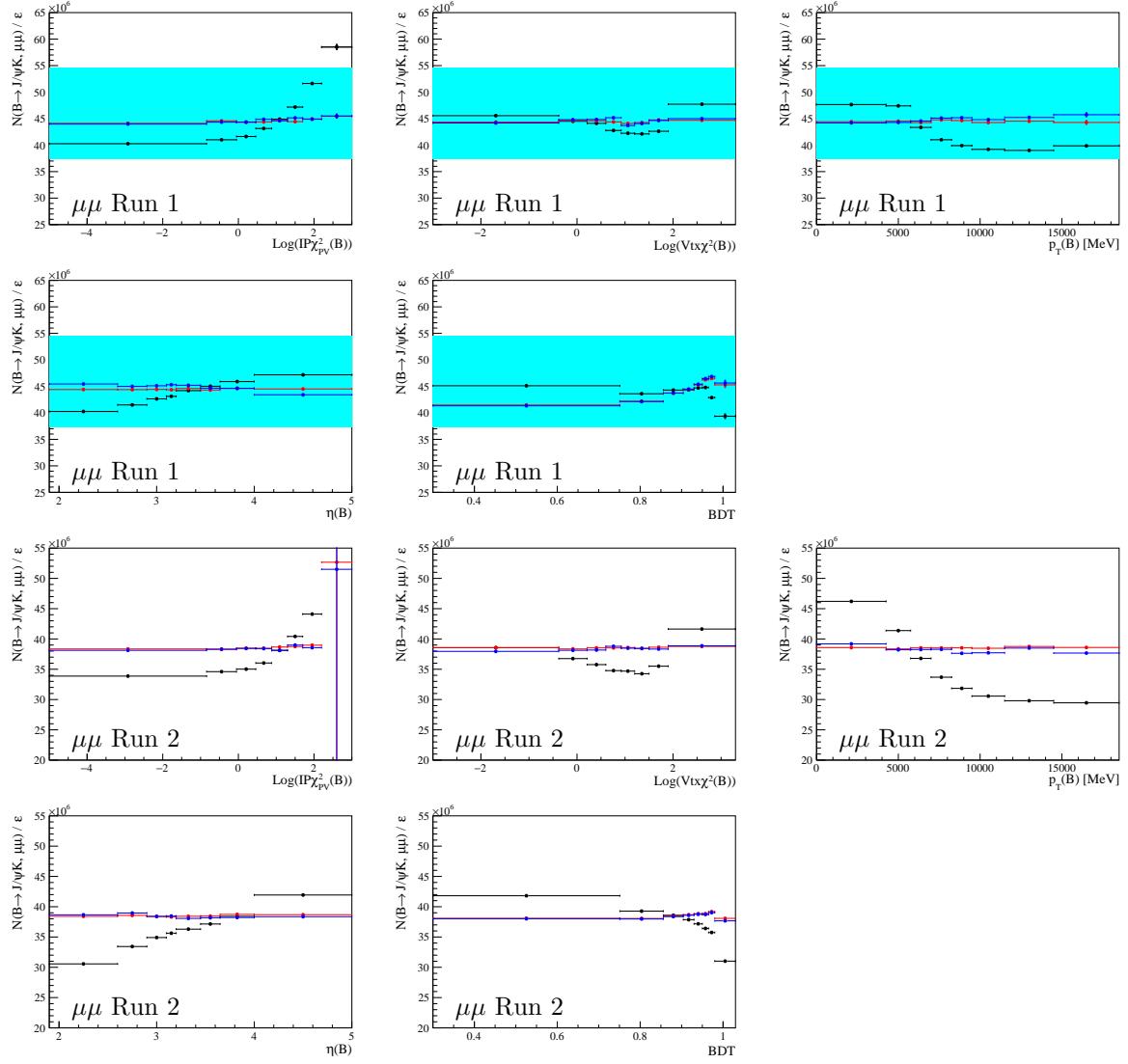


Figure 100: Efficiency corrected yield as a function of several kinematic variables for the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in Run 1 (top 5 plots) and Run 2 (bottom 5 plots). To compute the efficiencies, various kinematic weights are used and correspond to the different colours: no kinematic weights at all (black), the nominal μ TOS weights (red), and μ TIS weights (blue).

Table 33: Total efficiencies (in permille) for all modes, and ratios of efficiencies appearing in $r_{J/\psi}$, R_K and $R_{\psi(2S)}$, in all trigger categories and run conditions. Each efficiency or efficiency ratio is computed several times using kinematic weights computed in different ways. The kinematic weights used for the electron (muon) mode is indicated in the first (second) column.

e weight	μ weight	Run 1			Run 2		
		$J/\psi \varepsilon [\%]$	rare $\varepsilon [\%]$	$\psi(2S) \varepsilon [\%]$	$J/\psi \varepsilon [\%]$	rare $\varepsilon [\%]$	$\psi(2S) \varepsilon [\%]$
	μ TOS	13.966 ± 0.027	3.165 ± 0.016	15.067 ± 0.035	14.324 ± 0.028	3.236 ± 0.015	15.039 ± 0.034
	μ TIS	13.925 ± 0.027	3.162 ± 0.016	15.033 ± 0.035	14.385 ± 0.028	3.256 ± 0.015	15.099 ± 0.034
	e TOS	2.152 ± 0.009	0.541 ± 0.005	2.398 ± 0.019	4.217 ± 0.013	1.047 ± 0.007	4.302 ± 0.028
mixed		2.141 ± 0.009	0.538 ± 0.005	2.383 ± 0.019	4.212 ± 0.013	1.046 ± 0.007	4.299 ± 0.028
e TOS		2.226 ± 0.009	0.553 ± 0.005	2.509 ± 0.020	4.464 ± 0.013	1.120 ± 0.008	4.575 ± 0.029
μ TIS		2.179 ± 0.009	0.540 ± 0.005	2.438 ± 0.019	4.248 ± 0.013	1.055 ± 0.007	4.329 ± 0.028
e TIS		2.237 ± 0.009	0.555 ± 0.005	2.516 ± 0.020	4.567 ± 0.014	1.148 ± 0.008	4.671 ± 0.030
		h TOS!			TIS!		
μ TOS		0.351 ± 0.003	0.178 ± 0.002	0.120 ± 0.004	0.461 ± 0.003	0.264 ± 0.004	0.145 ± 0.005
mixed		0.349 ± 0.003	0.176 ± 0.002	0.119 ± 0.004	0.461 ± 0.003	0.264 ± 0.004	0.145 ± 0.005
e TOS		0.363 ± 0.003	0.180 ± 0.002	0.125 ± 0.004	0.490 ± 0.003	0.280 ± 0.004	0.153 ± 0.005
μ TIS		0.355 ± 0.003	0.178 ± 0.002	0.123 ± 0.004	0.465 ± 0.003	0.266 ± 0.004	0.146 ± 0.005
e TIS		0.365 ± 0.003	0.181 ± 0.002	0.127 ± 0.004	0.502 ± 0.004	0.287 ± 0.004	0.158 ± 0.005
		h TOS!			TIS!		
μ TOS		0.695 ± 0.004	0.192 ± 0.003	0.613 ± 0.009	1.096 ± 0.006	0.317 ± 0.004	0.933 ± 0.013
mixed		0.689 ± 0.004	0.190 ± 0.003	0.608 ± 0.009	1.095 ± 0.006	0.317 ± 0.004	0.932 ± 0.013
e TOS		0.702 ± 0.004	0.193 ± 0.003	0.624 ± 0.009	1.138 ± 0.006	0.332 ± 0.004	0.981 ± 0.013
μ TIS		0.694 ± 0.004	0.191 ± 0.003	0.614 ± 0.009	1.095 ± 0.006	0.317 ± 0.004	0.931 ± 0.013
e TIS		0.696 ± 0.004	0.193 ± 0.003	0.624 ± 0.009	1.138 ± 0.006	0.332 ± 0.004	0.984 ± 0.013
e weight	μ weight	$r_{J/\psi} \varepsilon$ ratio	$R_K \varepsilon$ ratio	$R_{\psi(2S)} \varepsilon$ ratio	$J/\psi \varepsilon$ ratio	$R_K \varepsilon$ ratio	$R_{\psi(2S)} \varepsilon$ ratio
		h TOS!			TIS!		
μ TOS	μ TOS	0.154 ± 0.001	1.033 ± 0.010	1.108 ± 0.012	0.294 ± 0.001	0.972 ± 0.008	1.099 ± 0.010
mixed	μ TOS	0.153 ± 0.001	1.031 ± 0.010	1.109 ± 0.012	0.294 ± 0.001	0.972 ± 0.008	1.099 ± 0.010
e TOS	μ TOS	0.159 ± 0.001	1.045 ± 0.010	1.095 ± 0.012	0.312 ± 0.001	0.976 ± 0.008	1.110 ± 0.010
μ TIS	μ TIS	0.157 ± 0.001	1.036 ± 0.010	1.092 ± 0.012	0.295 ± 0.001	0.971 ± 0.008	1.097 ± 0.010
e TIS	μ TIS	0.161 ± 0.001	1.042 ± 0.010	1.094 ± 0.012	0.317 ± 0.001	0.974 ± 0.007	1.110 ± 0.010
		h TOS!			TIS!		
μ TOS	μ TOS	0.025 ± 0.000	0.317 ± 0.010	2.238 ± 0.037	0.032 ± 0.000	0.299 ± 0.010	2.528 ± 0.040
mixed	μ TOS	0.025 ± 0.000	0.317 ± 0.010	2.228 ± 0.037	0.032 ± 0.000	0.299 ± 0.010	2.529 ± 0.040
e TOS	μ TOS	0.026 ± 0.000	0.319 ± 0.010	2.182 ± 0.036	0.034 ± 0.000	0.297 ± 0.010	2.529 ± 0.040
μ TIS	μ TIS	0.025 ± 0.000	0.322 ± 0.010	2.211 ± 0.037	0.032 ± 0.000	0.298 ± 0.010	2.525 ± 0.040
e TIS	μ TIS	0.026 ± 0.000	0.324 ± 0.011	2.184 ± 0.036	0.035 ± 0.000	0.300 ± 0.010	2.524 ± 0.040
		h TOS!			TIS!		
μ TOS	μ TOS	0.050 ± 0.000	0.818 ± 0.013	1.220 ± 0.019	0.077 ± 0.000	0.811 ± 0.012	1.281 ± 0.019
mixed	μ TOS	0.049 ± 0.000	0.818 ± 0.013	1.217 ± 0.019	0.076 ± 0.000	0.811 ± 0.012	1.281 ± 0.019
e TOS	μ TOS	0.050 ± 0.000	0.824 ± 0.013	1.215 ± 0.019	0.079 ± 0.000	0.821 ± 0.012	1.290 ± 0.019
μ TIS	μ TIS	0.050 ± 0.000	0.820 ± 0.013	1.214 ± 0.019	0.076 ± 0.000	0.810 ± 0.012	1.277 ± 0.019
e TIS	μ TIS	0.050 ± 0.000	0.830 ± 0.013	1.219 ± 0.019	0.079 ± 0.000	0.824 ± 0.012	1.288 ± 0.019

Table 34: Correlation matrices showing the fractional uncertainties and their correlation factors (in percent) for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$, due to the bias caused by the kinematic reweighting scheme.

$\mathbf{r}_{J/\psi}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
2.84	99.96	77.23	96.62	96.82	91.47
	2.89	76.59	97.15	97.34	92.38
		0.72	71.44	70.69	75.82
			4.90	99.98	97.19
				5.19	96.82
					2.59
\mathbf{R}_K					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
1.15	92.13	78.41	-66.87	80.99	-55.10
	1.85	73.48	-88.61	64.56	-81.60
		0.33	-39.65	42.81	-33.99
			0.72	-39.33	98.46
				0.11	-22.75
					0.47
$\mathbf{R}_{\psi(2S)}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.75	78.26	89.70	91.56	-14.02	93.78
	1.27	89.26	53.39	21.86	77.23
		0.84	81.34	30.23	96.84
			0.26	-7.31	93.13
				0.33	18.35
					1.05

1978 10.3 PID calibration biases

1979 There are three sources of systematic uncertainties related to the PID efficiency calibration
1980 which are as follows:

- 1981 • The first source relates to the choice of binning scheme for the muon and kaon PID
1982 calibration histograms. To estimate an uncertainty, alternative binning schemes are
1983 used, which are determined by using the rebinning algorithm described in Section 6.4
1984 using different thresholds: either 1.5, 2.0, 3.0 and 3.5σ instead of the nominal 2.5σ .
- 1985 • The second source is related to the choice of binning scheme for the electron PID
1986 histograms. In this case, due to the limited statistics in the electron PID histograms,
1987 only one alternative binning is used to assess the systematic uncertainty. This
1988 alternative binning is shown in blue (to compare with the nominal, shown in red)
1989 in Fig. 40-Fig. 43.
- 1990 • The last source is related to the trigger bias from the electron PID efficiencies.
1991 A systematic uncertainty is assessed by computing the difference in efficiencies
1992 obtained by using the nominal electron PID histogram (where the probe electron has
1993 potentially triggered any line), or by using a PID histogram extracted by requiring
1994 the probe electron to be TIS with respect to any trigger line. The nominal and
1995 alternative histograms are shown in Fig. 46.

1996 Covariance matrices corresponding to these three sources are computed and then added
1997 up to obtain the total uncertainty related to the PID calibration, which is very small. The
1998 correlation matrices related to the PID calibration are shown in Tab. 35.

Table 35: Correlation matrices showing the fractional uncertainties and their correlation factors (in percent) for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$, due to the bias caused by the binning of the PID calibration histogram, and the trigger bias affecting the electron PID histograms.

$\boldsymbol{r}_{J/\psi}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.32	88.27	89.09	64.24	64.46	65.25
	0.47	99.97	92.59	92.71	93.08
		0.53	91.99	92.11	92.50
			0.87	100.00	99.99
				1.08	99.99
					1.05
\boldsymbol{R}_K					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.13	98.94	57.68	64.53	44.13	64.54
	0.16	67.59	74.42	55.41	74.25
		0.07	98.33	97.69	98.45
			0.28	96.28	99.90
				0.03	96.71
					0.08
$\boldsymbol{R}_{\psi(2S)}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.03	87.73	91.67	73.46	92.43	75.46
	0.17	99.24	96.14	98.39	96.94
		0.14	92.27	99.76	93.43
			0.10	89.95	99.95
				0.16	91.28
					0.11

Table 36: Relative errors (in percent) on the single and double ratios due to the tag bias and the TCK bias on the L0Muon calibration histograms.

	Run 1		Run 2	
	σ_{tag} [%]	σ_{TCK} [%]	σ_{tag} [%]	σ_{TCK} [%]
$r_{J/\psi}$	0.24	—	0.31	0.07
R_K	0.03	—	0.05	0.02
$R_{\psi(2S)}$	0.02	—	0.07	0.07

10.4 Trigger histograms

The systematic uncertainties related to the L0 calibration histograms in each trigger category are discussed in the following sections.

10.4.1 Biases caused by the L0Muon calibration histograms

There are two possible causes of biases on the muon trigger histograms: the tag bias and the TCK bias.

The tag bias is evaluated by computing the total efficiency three times using three different set of L0Muon histograms: the first one computed with the nominal TIS tag, the second one using the other signal muon as a tag, and the third one using the kaon as a tag. These histograms are all displayed in Section 6.6.3.

The TCK bias only impacts the L0Muon histograms in Run 2. It is evaluated by computing the total efficiency twice using two different sets of L0Muon histograms: the nominal one that uses only one TCK, and an alternative one computed by ensuring that only a single muon, the probe, fired the muon trigger. These histograms are also shown in Section 6.6.3.

For both the tag bias and the TCK bias, the systematic uncertainties are computed from the nominal efficiencies and the alternative ones using Eq. 10. The systematic uncertainties on $r_{J/\psi}$, R_K and $R_{\psi(2S)}$ due to these two biases are shown in Tab. 36. All systematics are negligible.

10.4.2 Biases caused by the L0Electron calibration histograms

There are two possible causes of biases on the L0Electron trigger histograms: the tag bias and the factorisation bias.

The tag bias is evaluated by computing the total efficiency twice using two different sets of L0Electron histograms: the first one is computed with the nominal TIS tag, the second one uses the signal kaon as a tag. These histograms are all displayed in Section 6.6.1.

The factorisation bias is related to the fact that the L0Electron efficiencies of the two electrons are correlated, as discussed in Section 6.6.2. Therefore, an alternative way to parameterise the total electron trigger efficiency is used: the trigger efficiency is measured

in bins of the maximal electron E_T and the distance between the two electrons in the ECAL. The trigger histograms are all extracted using the nominal TIS tag and are shown in Fig. 101.

Fractional error matrices are computed according to Eq. 39 for both the tag bias and the factorisation bias, and are combined to get the total fractional error matrices due to the L0Electron histograms. The fractional error matrices related to the tag bias for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$ are shown in Tab. 37, and in Tab. 38 for the factorisation bias. The errors on the single ratios are large, at a level of a few percent, but cancel well in the double ratios.

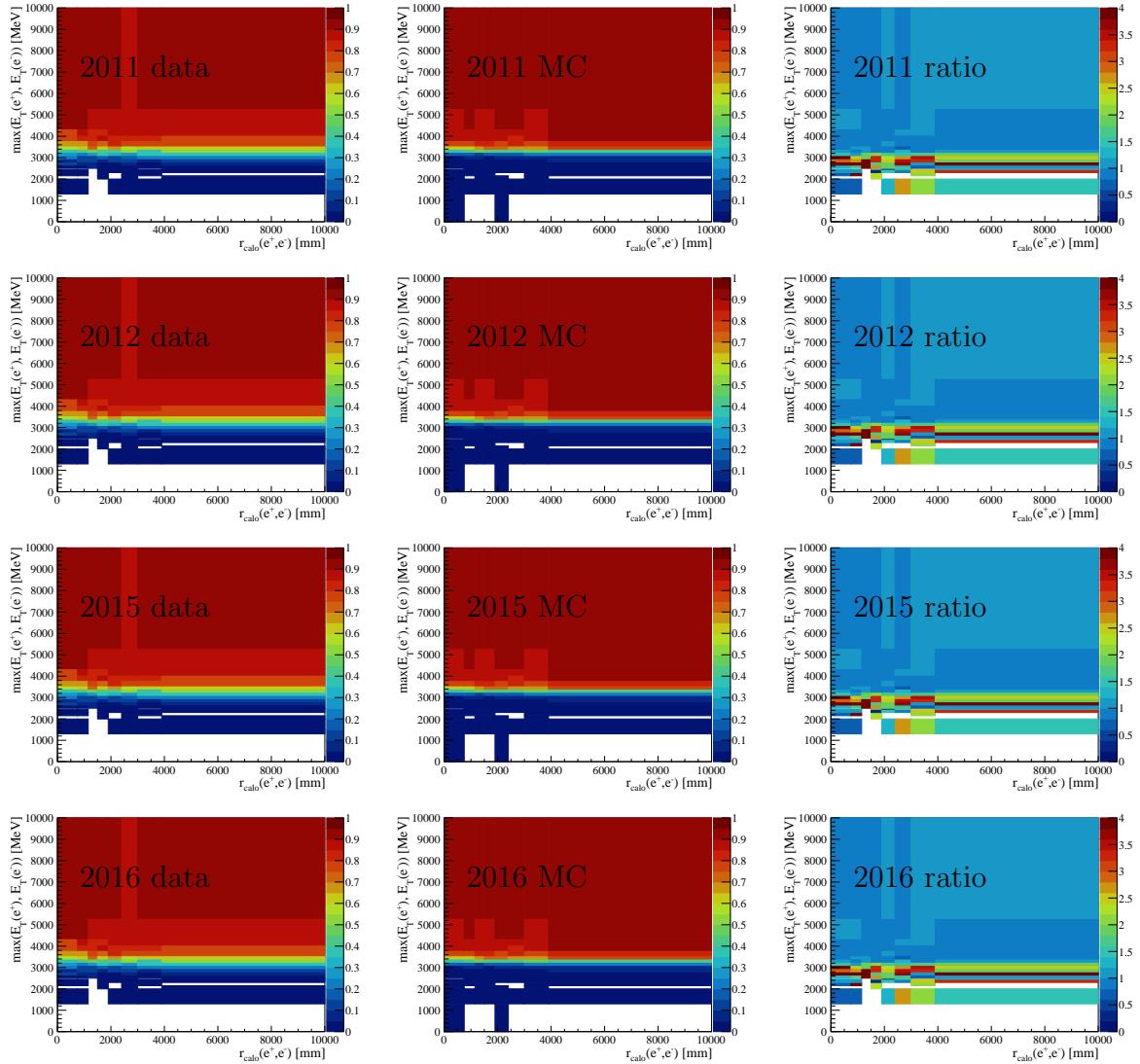


Figure 101: Efficiency of either of the two electrons in $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ triggering the L0Electron line in data (left), simulation (centre) and ratio between the two (right), in all years of data taking. All efficiencies are computed using the nominal TIS tag.

Table 37: Correlation matrices showing the fractional uncertainties and their correlation factors (in percent) for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$, due to the bias caused by the tag used to extract the L0Electron histograms.

$r_{J/\psi}$					
	Run 1		Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
3.55	-100.00	-100.00	100.00	-100.00	-100.00
	2.71	100.00	-100.00	100.00	100.00
		2.13	-100.00	100.00	100.00
			2.66	-100.00	-100.00
				2.70	100.00
					2.08
R_K					
	Run 1		Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.03	-100.00	-100.00	100.00	-100.00	-100.00
	1.07	100.00	-100.00	100.00	100.00
		0.07	-100.00	100.00	100.00
			0.04	-100.00	-100.00
				1.20	100.00
					0.25
$R_{\psi(2S)}$					
	Run 1		Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.17	100.00	100.00	100.00	100.00	100.00
	2.34	100.00	100.00	100.00	100.00
		0.54	100.00	100.00	100.00
			0.05	100.00	100.00
				2.17	100.00
					0.50

2036 10.4.3 Biases caused by the L0Hadron calibration histograms

2037 There is a source of bias on the L0Hadron trigger histograms from the tag bias. The
 2038 systematic uncertainty associated with this is evaluated by computing the total efficiency
 2039 four times using two different sets of L0Hadron histograms: the first one computed with the
 2040 nominal muon tag on the muon mode, the second one using the electron tag on the electron
 2041 mode, the third and fourth ones using the tis tag on both modes. These histograms are
 2042 all displayed in Section 6.6.5.

2043 Fractional error matrices for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$ are computed according to Eq. 39.

Table 38: Correlation matrices showing the fractional uncertainties and their correlation factors (in percent) for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$, due to the bias caused by the tag used to extract the L0Electron histograms.

$r_{J/\psi}$					
	Run 1		Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
3.18	-100.00	-100.00	-100.00	100.00	100.00
	3.01	100.00	100.00	-100.00	-100.00
		2.82	100.00	-100.00	-100.00
			1.35	-100.00	-100.00
				5.30	100.00
					4.38
R_K					
	Run 1		Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.10	-100.00	-100.00	100.00	-100.00	-100.00
	0.85	100.00	-100.00	100.00	100.00
		0.18	-100.00	100.00	100.00
			0.54	-100.00	-100.00
				0.08	100.00
					0.44
$R_{\psi(2S)}$					
	Run 1		Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.04	100.00	100.00	-100.00	-100.00	100.00
	0.62	100.00	-100.00	-100.00	100.00
		0.66	-100.00	-100.00	100.00
			0.25	100.00	-100.00
				1.08	-100.00
					0.07

2044 They are shown in Tab. 39.

2045 10.4.4 Biases caused by the LOTIS calibration histograms

2046 The LOTIS efficiency calibration is affected by the use of the electron tag to measure
 2047 the LOTIS efficiency, as discussed in this section, and owing to its dependence on the
 2048 occupancy, by the mis-modelling of the occupancy, the systematic from which is discussed
 2049 in Section 10.6.

2050 As detailed in Section 6.6.8, the LOTIS efficiency depends on whether the signal

Table 39: Correlation matrices showing the fractional uncertainties and their correlation factors (in percent) for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$, due to the bias caused by the tag used to extract the L0Electron histograms.

$r_{J/\psi}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.00	0.00	0.00	0.00	0.00	0.00
	3.72	-99.65	0.00	90.35	-86.82
		0.83	0.00	-91.14	86.89
			0.00	0.00	0.00
				2.68	-99.19
					0.39
R_K					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.00	0.00	0.00	0.00	0.00	0.00
	0.50	93.47	0.00	66.84	81.81
		0.25	0.00	84.58	95.69
			0.00	0.00	0.00
				0.17	80.37
					0.24
$R_{\psi(2S)}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
0.00	0.00	0.00	0.00	0.00	0.00
	0.45	95.48	0.00	81.94	71.50
		0.32	0.00	92.49	87.94
			0.00	0.00	0.00
				0.27	83.28
					0.18

2051 electrons have triggered or not. The most likely cause of this discrepancy is the presence of
 2052 ‘fake TIS’ events in the TIS sample, *i.e.* events triggered due to the presence of the signal
 2053 electrons but falsely classified as TIS. An example would be those events triggered by
 2054 the bremsstrahlung photons produced by the signal electrons. Such events can be found
 2055 when analysing the content of the TIS! category in the $B^+ \rightarrow K^+ e^+ e^-$ simulation. The
 2056 number of these fake TIS events depends on the kinematics of the signal electrons, and
 2057 therefore, on whether the electrons fired the L0_Electron trigger or not. Furthermore,
 2058 the parametrisation of the L0TIS efficiency correction as a function of B p_T would be
 2059 sub-optimal, and the efficiency should instead be parametrised as a function of the electron

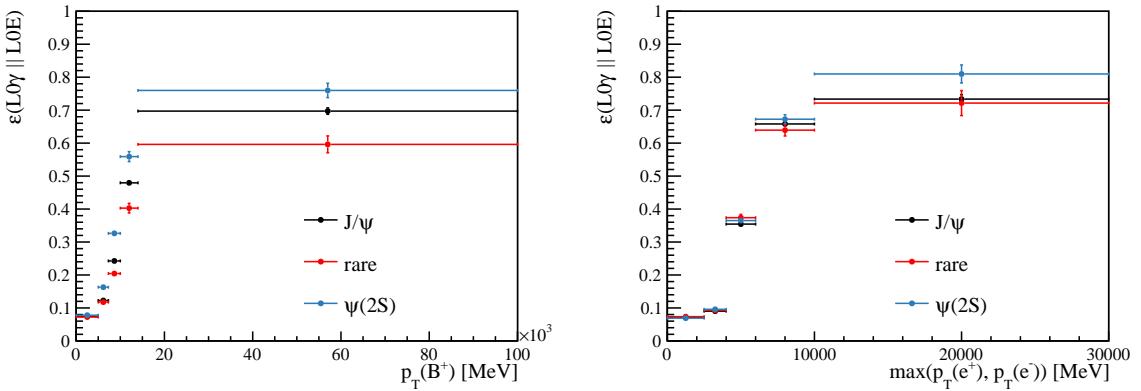


Figure 102: Efficiency $\varepsilon(\gamma\text{TIS} \parallel \text{eTIS})$ as a function of the B p_{T} (left) and the maximum p_{T} of the leptons (right) computed using different simulated samples.

kinematics in order to be insensitive to differences between the signal and normalisation channels.

To estimate the impact of fake TIS events, an alternative strategy to correct the L0TIS efficiency is used. The TIS category is further split into 2 subcategories: $\gamma\text{TIS} \parallel \text{eTIS}$, expected to contain the majority of the fake TIS events, and $\mu\text{TIS} \parallel \text{hTIS}$, expected to contain few fake TIS events. The global TIS efficiency is then calculated as

$$\varepsilon(\text{TIS}) = 1 - (1 - \varepsilon(\gamma\text{TIS} \parallel \text{eTIS})) \cdot (1 - \varepsilon(\mu\text{TIS} \parallel \text{hTIS})). \quad (40)$$

The efficiencies $\varepsilon(\gamma\text{TIS} \parallel \text{eTIS})$ and $\varepsilon(\mu\text{TIS} \parallel \text{hTIS})$, are computed as a function of the relevant quantities and using an appropriate tag for their calibration with data, as explained below.

Figure 102 (left) shows the $\varepsilon(\gamma\text{TIS} \parallel \text{eTIS})$ efficiency as a function of the p_{T} of the B , computed using $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$, $B^+ \rightarrow K^+ \psi(2S)(e^+ e^-)$ and $B^+ \rightarrow K^+ e^+ e^-$ simulated events. As expected, a discrepancy between the efficiency calculated using the different samples is found, as the fake TIS rate changes with the electron/positron momenta, indicating that B p_{T} is not the most appropriate variable to parameterise data/MC corrections. Instead these are parameterised as a function of the maximum p_{T} of the two electrons, see Fig. 102 (right). In addition, to compare the $\varepsilon(\gamma\text{TIS} \parallel \text{eTIS})$ efficiency in data and simulation, $B^+ \rightarrow J/\psi(ee)K^+$ events need to be selected in data using a tag that is independent of the electron kinematics. Figure 103 shows how employing events selected with a μ TIS tag is expected to provide an unbiased determination of the $\varepsilon(\gamma\text{TIS} \parallel \text{eTIS})$ efficiency.

On the other hand, the $\varepsilon(\mu\text{TIS} \parallel \text{hTIS})$ obtained for $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$, $B^+ \rightarrow K^+ \psi(2S)(e^+ e^-)$ and $B^+ \rightarrow K^+ e^+ e^-$ simulated events are well in agreement, indicating that this efficiency is independent of the electron kinematics, see Fig. 104 (left). In addition, the same efficiency is observed in $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ simulation independently of the tag used to select the relevant sample, see Fig. 104 (right). Therefore, the calibration of the $\varepsilon(\mu\text{TIS} \parallel \text{hTIS})$ efficiency is performed, as a function of

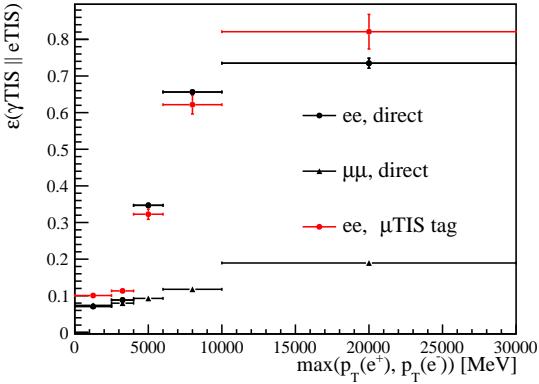


Figure 103: Efficiency $\varepsilon(\gamma\text{TIS} \parallel \text{eTIS})$ as a function of the maximum p_T of the leptons computed using $B^+ \rightarrow J/\psi(ee)K^+$ simulated events.

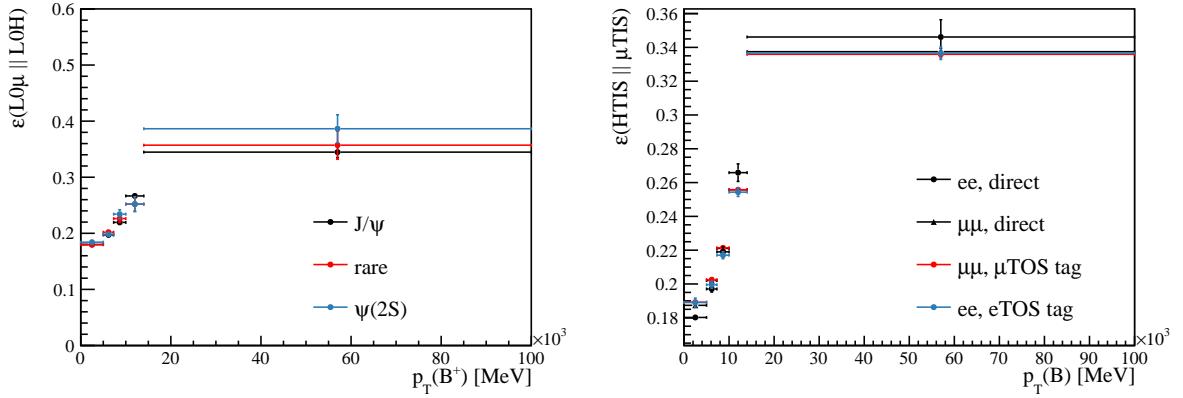


Figure 104: Efficiency $\varepsilon(\mu\text{TIS} \parallel \text{hTIS})$ as a function of the the B p_T computed using different simulation samples (left) and requiring different tags (right).

the B p_T , by comparing $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events selected using a electron TOS tag in data and simulation.

Combining $\varepsilon(\gamma\text{TIS} \parallel \text{eTIS})$ and $\varepsilon(\mu\text{TIS} \parallel \text{hTIS})$ efficiencies, an alternative value for the TIS! efficiency is computed. The difference between this value and the nominal calculation is shown in Tab. 40.

Finally, an alternative set of L0TIS weights is extracted by computing the ratio between the total L0TIS efficiency computed in $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data using the electrons as a tag, and the L0TIS efficiency computed on simulated events requiring the electrons not to have fired the trigger. The difference between the nominal total efficiency and the total efficiency obtained using these weights is also reported in Tab. 40. This corresponds to making the hypothesis that the electron tag bias on the L0TIS efficiency does not cancel at all when computing the ratio of L0TIS efficiencies between the data and the simulation. The largest of the variations computed with the two alternative corrections is then assigned

2099 as a systematic uncertainty Tab. 41. The correlation between this bias in Run 1 and in
2100 Run 2 is assumed to be 100%. Although the systematic uncertainty on the single ratio is
2101 large, it is negligible on the double ratios.

Table 40: Variation in the total efficiency on the single and double ratios in the TIS! trigger category, due to the electron tag bias on the L0TIS correction weights computed following the two alternative strategies described in the text. The correlation between Run 1 and Run 2 is assumed to be 100%.

	Run 1		Run 2	
$r_{J/\psi}$ TIS!	fake TIS	e Tag	fake TIS	e Tag
	5.2%	3.34%	0.6%	4.60%
R_K TIS!	0.29%	0.12%	0.11%	0.01%
$R_{\psi(2S)}$ TIS!	0.21%	-	0.28%	-

Table 41: Fractional systematic uncertainties on the single and double ratios in the TIS! trigger category, due to the electron tag bias on the L0TIS correction weights. The correlation between Run 1 and Run 2 is assumed to be 100%.

	Run 1	Run 2
$r_{J/\psi}$ TIS!	5.2%	4.60%
R_K TIS!	0.29%	0.11%
$R_{\psi(2S)}$ TIS!	0.21%	0.28%

2102 10.5 ECAL calibration constants

2103 The systematic uncertainty related to the mismatch between the ECAL calibration
2104 constants in data and simulation, which affects only the Run 2 data from 2016, is evaluated
2105 by recomputing the L0 decision in simulation after aligning the values of the calibration
2106 constants to those used during data taking. To do this, the tool `L0CaloFix2016` provided
2107 at Ref. [45] is used. This tool accesses the CALO RAW banks, if available, and reevaluates
2108 the L0 trigger decision for a certain TCK. If the CALO RAW banks are not available, this
2109 tool uses an approximation to correct the energy resolution at the L0Calo level.

2110 Once the trigger response is recomputed, all the trigger calibration procedure described
2111 in Section 6.6 is applied to these samples, and the trigger efficiencies for $B^+ \rightarrow K^+ e^+ e^-$
2112 and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ are recomputed. The difference between these values and those
2113 obtained through the nominal procedure is taken as a systematic uncertainty.

2114 10.6 Occupancy proxy

2115 This systematic uncertainty is extracted using `superLenin/src/mainSystematic6.cc`.

As discussed in Section 6.8.1, the occupancy of the detector is poorly described in the simulation. To study how this can impact on the total efficiency, the distributions of three occupancy proxies in the simulation are separately reweighted to match the data. This is done using the same algorithm as that described in Section 6.8, adding one occupancy proxy into the list of variables that are reweighted. The three occupancy proxies considered are:

1. nSPDHits: the number of hits in the scintillating pad detector;
2. nTracks: the number of tracks found in the detector;
3. nPVs: the number of primary vertices found in the velo.

In the case of nSPDHits, there are no events in the data above the offline cut, $nSPDHits < 600$ for Run 1 and $nSPDHits < 450$ for Run 2. In the simulation it is necessary to keep and reweight such events, in order to compute the efficiency of the nSPDHits cut. The weight assigned to the simulated generator level events above the offline cut is equal to the weight in the highest nSPDHits bin.

To compute the TIS! efficiency, the TIS calibration histograms described in Section 6.6.8, which are very sensitive to the occupancy, are also re-computed after making the occupancy correction.

The total efficiencies, as well as the single and double ratios of efficiencies computed using the various occupancy proxies are reported in Tab. 42

Using these three alternative ways to compute the total efficiencies and Eq. 39, the correlation matrix for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$ are extracted and shown in Tab. 43.

10.7 Material and rapidity

This systematic uncertainty is extracted using *superLenin/src/mainSystematic7.cc*.

In the case of the electron mode, the electron rapidity distribution does not match that in the data very well. This is somewhat expected: at the time of writing the tracking efficiency, which is known to depend on the rapidity, has not been measured for electrons at LHCb. The most likely cause of this discrepancy is a mismatch between the data and simulation material budget as a function of the rapidity: as the electron interaction cross-section with the material of the detector is relatively large, the tracking efficiency of electrons is very dependent on the material budget.

The systematic uncertainty due to this effect is computed for the double ratios R_K and $R_{\psi(2S)}$ but, as explained below, it is not possible to compute a systematic on the single ratio $r_{J/\psi}$.

The rapidity distributions for simulation and data are shown in Fig. 105. A discrepancy is visible at high rapidity, for both data taking conditions and for all trigger categories. To rectify this difference, a supplementary set of correction weights are computed. These weights are data over simulation ratios of two dimensional $\eta(e^+)$, $\eta(e^-)$ distributions,

Table 42: Total efficiencies (in permille) for all modes, and ratios of efficiencies appearing in $r_{J/\psi}$, R_K and $R_{\psi(2S)}$, in all trigger categories and run conditions. Each efficiency or efficiency ratio is computed four times by reweighting various occupancy proxies indicated in the first column.

proxy	Run 1			Run 2		
	$J/\psi \varepsilon[\%]$	rare $\varepsilon[\%]$	$\psi(2S) \varepsilon[\%]$	$J/\psi \varepsilon[\%]$	rare $\varepsilon[\%]$	$\psi(2S) \varepsilon[\%]$
μ TOS						
—	13.97 ± 0.03	3.17 ± 0.02	1.51 ± 0.00	14.324 ± 0.028	3.236 ± 0.015	1.504 ± 0.003
nSPDHits	12.959 ± 0.030	2.951 ± 0.020	1.402 ± 0.004	12.648 ± 0.026	2.869 ± 0.015	1.329 ± 0.003
nTracks	13.988 ± 0.027	3.173 ± 0.017	1.509 ± 0.004	14.364 ± 0.028	3.252 ± 0.016	1.507 ± 0.003
nPVs	14.003 ± 0.027	3.172 ± 0.016	1.512 ± 0.004	14.355 ± 0.028	3.235 ± 0.015	1.509 ± 0.003
e TOS						
—	2.15 ± 0.01	0.54 ± 0.00	0.24 ± 0.00	4.217 ± 0.013	1.047 ± 0.007	0.430 ± 0.003
nSPDHits	1.932 ± 0.010	0.477 ± 0.006	0.213 ± 0.002	3.640 ± 0.012	0.904 ± 0.007	0.372 ± 0.003
nTracks	2.155 ± 0.009	0.541 ± 0.005	0.240 ± 0.002	4.234 ± 0.013	1.048 ± 0.008	0.432 ± 0.003
nPVs	2.171 ± 0.009	0.546 ± 0.005	0.243 ± 0.002	4.260 ± 0.013	1.061 ± 0.008	0.435 ± 0.003
h TOS!						
—	0.351 ± 0.003	0.178 ± 0.002	0.012 ± 0.000	0.461 ± 0.003	0.264 ± 0.004	0.014 ± 0.000
nSPDHits	0.320 ± 0.004	0.162 ± 0.003	0.011 ± 0.000	0.404 ± 0.003	0.231 ± 0.004	0.013 ± 0.000
nTracks	0.353 ± 0.003	0.179 ± 0.002	0.012 ± 0.000	0.466 ± 0.003	0.265 ± 0.004	0.015 ± 0.001
nPVs	0.351 ± 0.003	0.179 ± 0.002	0.012 ± 0.000	0.464 ± 0.003	0.264 ± 0.004	0.015 ± 0.000
TIS!						
—	0.695 ± 0.004	0.192 ± 0.003	0.061 ± 0.001	1.096 ± 0.006	0.317 ± 0.004	0.093 ± 0.001
nSPDHits	0.600 ± 0.005	0.168 ± 0.003	0.051 ± 0.001	0.909 ± 0.005	0.263 ± 0.004	0.076 ± 0.001
nTracks	0.685 ± 0.004	0.190 ± 0.003	0.060 ± 0.001	1.073 ± 0.006	0.311 ± 0.004	0.091 ± 0.001
nPVs	0.707 ± 0.005	0.197 ± 0.003	0.063 ± 0.001	1.114 ± 0.006	0.323 ± 0.004	0.095 ± 0.001
proxy	$r_{J/\psi} \varepsilon$ ratio	$R_K \varepsilon$ ratio	$R_{\psi(2S)} \varepsilon$ ratio	$J/\psi \varepsilon$ ratio	$R_K \varepsilon$ ratio	$R_{\psi(2S)} \varepsilon$ ratio
e TOS						
—	0.154 ± 0.001	1.033 ± 0.010	1.108 ± 0.012	0.294 ± 0.001	0.972 ± 0.008	1.099 ± 0.010
nSPDHits	0.149 ± 0.001	1.018 ± 0.013	1.085 ± 0.016	0.288 ± 0.001	0.971 ± 0.009	1.095 ± 0.011
nTracks	0.154 ± 0.001	1.031 ± 0.010	1.107 ± 0.012	0.295 ± 0.001	0.972 ± 0.008	1.093 ± 0.010
nPVs	0.155 ± 0.001	1.037 ± 0.010	1.111 ± 0.012	0.297 ± 0.001	0.972 ± 0.008	1.105 ± 0.010
h TOS!						
—	0.025 ± 0.000	0.317 ± 0.010	2.238 ± 0.037	0.032 ± 0.000	0.299 ± 0.010	2.528 ± 0.040
nSPDHits	0.025 ± 0.000	0.317 ± 0.014	2.223 ± 0.051	0.032 ± 0.000	0.305 ± 0.011	2.523 ± 0.046
nTracks	0.025 ± 0.000	0.317 ± 0.011	2.239 ± 0.038	0.032 ± 0.000	0.305 ± 0.011	2.515 ± 0.042
nPVs	0.025 ± 0.000	0.316 ± 0.010	2.254 ± 0.038	0.032 ± 0.000	0.299 ± 0.010	2.525 ± 0.041
TIS!						
—	0.050 ± 0.000	0.818 ± 0.013	1.220 ± 0.019	0.077 ± 0.000	0.811 ± 0.012	1.281 ± 0.019
nSPDHits	0.046 ± 0.000	0.787 ± 0.016	1.227 ± 0.027	0.072 ± 0.000	0.800 ± 0.013	1.275 ± 0.022
nTracks	0.049 ± 0.000	0.815 ± 0.013	1.225 ± 0.020	0.075 ± 0.000	0.812 ± 0.013	1.279 ± 0.020
nPVs	0.051 ± 0.000	0.821 ± 0.013	1.231 ± 0.020	0.078 ± 0.000	0.811 ± 0.012	1.286 ± 0.020

computed in the e TOS category and in all bremsstrahlung categories separately. The weights are shown in Fig. 106.

The rapidity distributions for simulated events after applying the correction weights are shown in Fig. 105 in blue. The data versus simulation agreement is improved in all categories, except in the Run 1 TIS! category, where the rapidity correction seems to over-correct slightly. The fact that a slight discrepancy remains in this category is not surprising as the LOTIS efficiency presented in Section 6.6.8 is a function of the B^+ momentum, transverse momentum and the rapidity. The correction applied here, which is a function only of the rapidity, can therefore result in some residual effect. The impact of the rapidity correction weights on a whole range of kinematic variables is discussed further in Section 11.2. The effect of preliminary data-derived efficiency corrections is

Table 43: Correlation matrices showing the fractional uncertainties and their correlation factors (in percent) for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$, due to the bias caused by the inaccurate description of the occupancy in the simulation.

$\boldsymbol{r}_{J/\psi}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
2.35	96.65	99.96	98.82	98.72	99.88
	1.14	95.92	91.58	91.32	95.26
		4.99	99.20	99.12	99.97
			1.67	100.00	99.46
				0.57	99.39
					4.38
\boldsymbol{R}_K					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
1.52	77.61	-40.47	70.23	80.53	80.01
	0.69	26.26	99.40	25.11	99.92
		0.79	36.68	-86.81	22.47
			0.47	14.35	98.89
				0.18	28.87
					0.45
$\boldsymbol{R}_{\psi(2S)}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
1.04	21.29	98.96	91.11	-97.45	97.01
	0.32	35.12	59.67	-42.69	44.35
		2.73	96.09	-99.66	99.49
			0.04	-98.04	98.39
				1.42	-99.98
					0.97

²¹⁶⁴ cross-checked in Section 11.4.

²¹⁶⁵ The rapidity correction weights can be interpreted as a data over simulation ratio of
²¹⁶⁶ electron tracking efficiencies, times an unknown normalisation constant k :

$$w(\eta(e^+), \eta(e^-)) = k \cdot \frac{\varepsilon_{\text{data}}^{\text{track}}(\eta(e^+)) \cdot \varepsilon_{\text{data}}^{\text{track}}(\eta(e^-))}{\varepsilon_{\text{sim}}^{\text{track}}(\eta(e^+)) \cdot \varepsilon_{\text{sim}}^{\text{track}}(\eta(e^-))}.$$

²¹⁶⁷ The unknown normalisation constant k prevents the use of these weights to compute the
²¹⁶⁸ impact of the rapidity correction weights on the total efficiency $\varepsilon^{\text{tot}}(B^+ \rightarrow K^+ J/\psi(e^+e^-))$,
²¹⁶⁹ and therefore on $r_{J/\psi}$. However, in the ratio $\varepsilon^{\text{tot}}(B^+ \rightarrow K^+ e^+e^-)/\varepsilon^{\text{tot}}(B^+ \rightarrow$

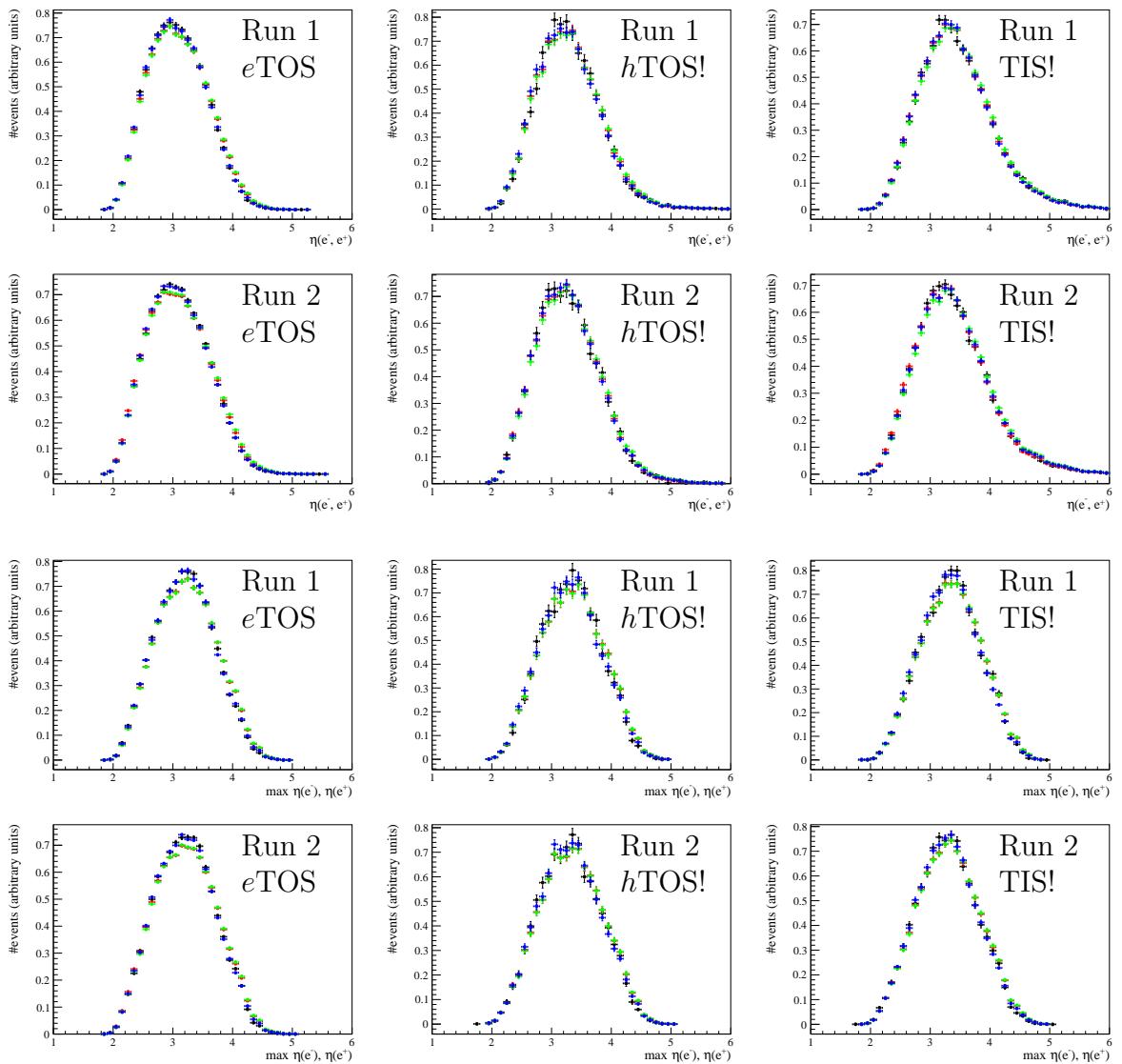


Figure 105: Distribution for the di-electron rapidity (top six plots) or maximum electron rapidity (bottom six plots) for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in all three exclusive categories and in both data taking conditions. The black points correspond to background-subtracted data, the red to simulation without kinematic corrections, the green to simulation with the nominal kinematic correction and the blue to the simulation with the extra rapidity correction.

2170 $K^+ J/\psi(e^+e^-)$), the normalisation constants cancel out. Therefore, the systematic uncer-
 2171 tainty on R_K related to the mis-modelling of the electron tracking efficiency is calculated as
 2172 the difference between the ratio of total efficiencies computed with the rapidity correction
 2173 weights, and without them. The correlation between all trigger categories and data taking
 2174 conditions is assumed to be 100%. The fractional error matrices for R_K and $R_{\psi(2S)}$ are
 2175 shown in Tab. 44. The fractional uncertainties are at the sub-percent level, which can

2176 be a priori surprising considering the large rapidity corrections. However, the rapidity
 2177 corrections cancel almost perfectly in the double ratio, due to the fact that the rapidity
 2178 distributions are very similar in the rare mode, J/ψ mode and $\psi(2S)$ mode.

Table 44: Fractional uncertainties (in percent) for R_K and $R_{\psi(2S)}$, due to the bias caused by the inaccurate description of the electron tracking efficiencies in the simulation. The correlation factors between the different data taking condition and trigger categories are all 100%.

mode	Run 1			Run 2		
	$e\text{TOS}$	$h\text{TOS!}$	TIS!	$e\text{TOS}$	$h\text{TOS!}$	TIS!
R_K	0.53	0.62	0.09	0.47	0.02	0.35
$R_{\psi(2S)}$	0.04	0.01	0.23	0.04	0.31	0.02

2179

2180 10.8 Momentum and q^2 resolution

2181 This systematic uncertainty is extracted using *superLenin/src/mainSystematic5.cc*.

2182 The electron mode q^2 and $m(K^+e^+e^-)$ smearing described in Section 6.9 is approximate.
 2183 Systematic uncertainties related to various effects are described below In this subsection,
 2184 the correlation between each trigger category and data taking condition is set to 100%.

2185 10.8.1 Electron q^2 and $m(K^+e^+e^-)$ resolution: line shape

2186 As Fig. 86 shows, the description of the q^2 and $m(K^+e^+e^-)$ line shape is imperfect even
 2187 after smearing, due to a residual discrepancy in the upper mass tails, which us related
 2188 to the bremsstrahlung recovery. To assign a systematic for this residual discrepancy,
 2189 an alternative method to correct the line shape for simulated events is used, relying on
 2190 the fact that the size of the upper tail is dependent on the occupancy. The nSPDHits
 2191 distribution is broken into three bins of uniform width and one $m(e^+e^-)$ shape is extracted
 2192 from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulation in each of these three bins, and is denoted f_{SPD}^i ,
 2193 $1 \leq i \leq 3$. Each f_{SPD}^i is normalised to the fraction of simulated events in the i th nSPDHits
 2194 bin. The J/ψ peak in $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data is then fitted using the signal model,

$$2195 N \cdot ((1 + a) \cdot f_{\text{SPD}}^1(m; s_\sigma, \Delta\mu) + f_{\text{SPD}}^2(m; s_\sigma, \Delta\mu) + (1 - a) \cdot f_{\text{SPD}}^3(m; s_\sigma, \Delta\mu)) , \quad (41)$$

2196 where N denotes the signal yield, s_σ and $\Delta\mu$ are the sigma scale and mean shift parameters,
 2197 and a is a free parameter which allows the fraction of yields between the three nSPDHits
 2198 bins to float in the data. Instead of scaling f_{SPD}^1 by $1 + a$ and f_{SPD}^3 by $1 - a$, it would
 2199 in principle be possible to give more freedom by using two independent scaling factors.
 However, this leads to convergence problems.

2200 The fit to the data obtained by using Eq. 10.8.1 as the signal shape and an exponential
 2201 for the background is shown in Fig. 107, for bremsstrahlung categories 1 and 2. The signal
 2202 shape describes the data accurately, including the upper mass tail. The fit favours a value
 2203 of a very close to 1, which is the maximal value for this parameter, and is equivalent to
 2204 eliminating all simulated events in the highest occupancy bin.

2205 Hence, $m(K^+e^+e^-)$ and $m(e^+e^-)$ are corrected in the simulation using the result of
 2206 the fit shown in Fig. 107. As for the nominal case, the distributions are smeared using
 2207 s_σ and $\Delta\mu$ (see Eq. 6.9) but the nSPDHits distribution is also reweighted: the events in
 2208 the lowest occupancy bin are reweighted by a factor two and the events in the highest
 2209 occupancy bin are cut out. The result of this procedure is illustrated in Fig. 108. The
 2210 shapes obtained from a $m(K^+e^+e^-)$ and $m(e^+e^-)$ fit to the data are shown (as in Fig. 86),
 2211 as well as the shapes obtained from a fit to the simulation using the alternative smearing.
 2212 The latter show a perfect agreement with the data.

2213 The systematic uncertainty related to the residual differences in the $m(K^+e^+e^-)$ and
 2214 $m(e^+e^-)$ shapes between the smeared simulation and the data is computed as follows.
 2215 The efficiency of the $m(K^+e^+e^-)$ mass window cut and the q^2 cut is computed using the
 2216 nominal smearing method, and the alternative smearing method described above. The
 2217 difference in the efficiency between the two methods is assigned as a systematic, which is,
 2218 for each of the three trigger categories:

		Run 1	Run 2
2219	$r_{J/\psi}$	0.30%	0.72%
	R_K	0.10%	0.14%
	$R_{\psi(2S)}$	0.27%	0.36%

2220 with a 100% correlation factor assumed between different run and trigger categories.

2221 10.8.2 Electron q^2 and $m(K^+e^+e^-)$ resolution: trigger bias

2222 Another source of systematics related to the q^2 resolution is the uncertainty on the
 2223 smearing factors applied to the simulation, which are shown in Tab. 17. The uncertainty
 2224 on these smearing factors is computed as the *RMS* between the factors computed in the
 2225 three trigger categories. To propagate the uncertainty to the efficiencies, the efficiencies
 2226 are computed using smearing factors modified by plus or minus one standard deviation.
 2227 Using Eq. 10, the relative systematic uncertainties on the ratios are then found to be

		Run 1	Run 2
2228	$r_{J/\psi}$	0.25%	0.24%
	R_K	0.13%	0.10%
	$R_{\psi(2S)}$	0.13%	0.12%

2229 with a 100% correlation factor assumed between different run and trigger categories.

2230 **10.8.3 Electron q^2 resolution: momentum dependence**

2231 It is possible that the smearing factors depend on q^2 , and that the simulation should
 2232 be smeared with a different factor at low q^2 than at $q^2 = m(J/\psi)^2$. For the electron
 2233 mode, fitting the $\psi(2S)$ peak or the ϕ peak is very challenging due to the complicated
 2234 backgrounds.

2235 The di-electron invariant mass can be written as

$$m(e^+e^-) = (q^2 = 2 \cdot p_{e^+} \cdot p_{e^-} \cdot (1 - \cos \alpha_{\ell^+\ell^-}))^{\frac{1}{2}}.$$

2236 It is checked on simulated events that the resolution on $\alpha_{\ell^+\ell^-}$ is $\sim 0.5\%$, whereas the
 2237 resolution on p is $\sim 10\%$. Therefore, the resolution on $m(e^+e^-)$ is completely dominated
 2238 by the momentum resolution itself, which is dependent on the momentum. The idea
 2239 to then assign a systematic on the q^2 dependence of the smearing factor is to use the
 2240 $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ channel in both data and simulation to measure the smearing
 2241 constants in 2D bins of $p_{\max} = \max(p_{e^+}, p_{e^-})$, $p_{\min} = \min(p_{e^+}, p_{e^-})$. To do so, in each
 2242 2D bin the standard deviation on $m(e^+e^-)$, $\sigma_{m(e^+e^-)}$ is computed. This quantity can be
 2243 approximately expressed as the propagated error on the two electron momenta which are
 2244 assumed to be independent:

$$\sigma_{m(e^+e^-)} \approx \frac{1}{\sqrt{2}} \cdot (\sigma_{p_{\max}}^2 + \sigma_{p_{\min}}^2)^{\frac{1}{2}}.$$

2245 Therefore, the error on the momentum, σ_p , can be estimated, in both data and simulation,
 2246 by minimising the following χ^2 :

$$\chi^2 = \sum \frac{\left(\sigma_{m(e^+e^-)} - \frac{1}{2} \cdot (\sigma_{p_{\max}}^2 + \sigma_{p_{\min}}^2)^{\frac{1}{2}} \right)^2}{\text{err}(\sigma_{m(e^+e^-)})^2},$$

2247 where the sum runs over all the 2D bins (the index has been dropped to simplify the
 2248 notation) and $\text{err}(\sigma_{m(e^+e^-)})$ is the estimated error on $\sigma_{m(e^+e^-)}$. The resolution σ_p , obtained
 2249 from this χ^2 , is shown in Fig. 109. Taking the ratio between data and simulation yields
 2250 a momentum smearing constant $s(p)$, which is found to have a mild dependence on the
 2251 momentum, as can be seen in Fig. 109. The smearing constant on $m(e^+e^-)$ can then be
 2252 computed in a momentum-dependent way:

$$s_\sigma(p_{e^+}, p_{e^-}) = \frac{1}{\sqrt{2}} (s(p_{e^+})^2 + s(p_{e^-}))^{\frac{1}{2}}.$$

2253 The error on R_K and $R_{\psi(2S)}$ related to the momentum dependence of the q^2 resolution
 2254 is then taken as the difference between the ratio of efficiencies computed using the nominal
 2255 momentum-independent smearing constants, and these momentum-dependent smearing
 2256 constants.

2257 The errors on the double ratios are found to be:

	Run 1	Run 2
R_K	0.39%	0.14%
$R_{\psi(2S)}$	0.36%	0.43%

2259 with a 100% correlation factor assumed between different run and trigger categories.

2260 10.9 Fit model

2261 In order to evaluate the systematic uncertainty on R_K arising from the choice of fit model,
 2262 toy experiments are generated using an alternative model and fitted twice, once with the
 2263 nominal model and once with the alternative one, to obtain R_K^{nom} and R_K^{alt} respectively.
 2264 The mean of the $\Delta R_K = R_K^{\text{alt}} - R_K^{\text{nom}}$ distribution is assigned as a systematic uncertainty.
 2265 The alternative models used for the signal and background components are:

- 2266 • Signal: The $B^+ \rightarrow K^+ e^+ e^-$ distribution is parameterised as a sum of three Gaussian
 2267 functions for the 1γ and 2γ categories. In the 0γ category a single Crystal Ball
 2268 distribution is used. The results of the fits to the signal simulation using the
 2269 alternative signal models are shown in Appendix H.3.1.
- 2270 • Partially reconstructed background: The `RooKeysPDF` is reevaluated after including
 2271 the contribution from $B \rightarrow K^{**} e^+ e^-$ decays estimated in Section 5.7, where $K^{**} \equiv$
 2272 $\{K_1, K_2^{*0(+)}\}$. The comparison with the nominal model, where only $B^0 \rightarrow K^{*0} e^+ e^-$
 2273 events are considered, can be found Appendix H.3.2.

2274 The systematic uncertainties obtained are summarised in Tab. 45. Given that these
 2275 uncertainties affect R_K directly, and not the efficiencies, they cannot be combined with
 2276 those obtained in previous sections and included in the fit. Therefore, their combination is
 2277 added separately to the final R_K result.

Table 45: Systematic uncertainties on R_K (in percent) arising from variations of the fit model.

	Systematic (%)
Signal model	1.34
Part. Reco model	0.98
Total	1.66

2278 10.10 Partially reconstructed background constraint

2279 This systematic uncertainty is extracted using `superLenin/src/superMain19.cc`.

2280 In the combined fit to extract R_K , the amount of partially reconstructed background
 2281 across the three trigger categories is constrained using the r_{prc}^{rt} ratios defined in (35).
 2282 The value of this constraint is obtained from the total efficiency of $B^0 \rightarrow K^{*0} e^+ e^-$ and
 2283 $B^+ \rightarrow K^+ e^+ e^-$ events in the three trigger categories, calculated using simulation. As they
 2284 are double electron ratios, all the systematic effects considered before cancel out in these
 2285 factors, and only differences arising from different types of backgrounds can induce a bias
 2286 in them.

2287 In particular, efficiency differences between the two largest background components,
 2288 $B \rightarrow K^{*0} (\rightarrow K^+ \pi^-) e^+ e^-$ and $B^+ \rightarrow K^{*+} (\rightarrow K^+ \pi^0) e^+ e^-$, induced by differences between
 2289 charged and neutral pions are considered. Alternative values for the r_{prc}^{rt} factors are

2290 computed using a sample of simulated $B^+ \rightarrow K^{*+}e^+e^-$ events, which had been previously
 2291 reweighted to match the physics model of $B^0 \rightarrow K^{*0}e^+e^-$. The difference between these
 2292 values and the nominal ones, obtained from $B^0 \rightarrow K^{*0}e^+e^-$ simulation, is taken as a
 2293 systematic uncertainty. The fractional error matrix for the r_{prc}^{rt} factors is shown in Tab. 46.

Table 46: Fractional error matrix on the r_{prc}^{rt} factors showing the fractional uncertainties and their correlation factors due to differences in efficiency between $B^0 \rightarrow K^{*0}e^+e^-$ and $B^+ \rightarrow K^{*+}e^+e^-$ events.

R_K			
Run 1		Run 2	
$hTOS!/eTOS$	$TIS!/eTOS$	$hTOS!/eTOS$	$TIS!/eTOS$
14.5	0.0	100	100
	1.4	0.0	100
		14.5	0.0
			1.4

2294 10.11 Decay model

2295 The Wilson coefficients, form factors and other hadronic uncertainties used in the decay
 2296 model impact on the q^2 distribution and hence on the resulting selection efficiencies. These
 2297 uncertainties are propagated to an uncertainty on R_K using the theoretical predictions from
 2298 the flavio software package (see Fig. 110). The q^2 distribution of the signal is reweighted
 2299 100 times to sample the theory uncertainties and the efficiency is recomputed in each case.
 2300 The result variation on R_K is taken as the systematic. The fractional error matrix for R_K
 2301 is shown in Tab. 47.

Table 47: Fractional error matrix on R_K showing the fractional uncertainties and their correlation factors due to the uncertainties in the decay model.

R_K					
Run 1			Run 2		
$eTOS$	$hTOS!$	$TIS!$	$eTOS$	$hTOS!$	$TIS!$
0.37	99.96	99.25	99.89	99.98	99.88
	0.22	99.50	99.78	99.99	99.80
		0.12	98.72	99.37	98.95
			0.43	99.86	99.91
				0.24	99.83
					0.37

2302 10.12 Combination of systematic uncertainties

2303 The systematic uncertainties on the efficiencies and their correlations, as discussed in the
 2304 above sections, are combined by summing all the covariance matrices corresponding to
 2305 each source of bias. The total fractional error matrices on $r_{J/\psi}$, R_K and $R_{\psi(2S)}$ are shown
 2306 in Tab. 48. Systematic uncertainties which effect R_K directly, such as those from the fit
 2307 models used (see Section 10.9), are added separately to the final uncertainty on R_K .

Table 48: Total fractional error matrices showing the fractional uncertainties and their correlation factors (in percent) for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$, obtained by propagating the uncertainty due to all the different sources of bias.

$r_{J/\psi}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
6.03	-20.14	-14.92	59.69	43.55	41.02
	7.01	36.13	29.44	26.50	4.27
		6.95	16.28	-11.11	44.32
			6.08	26.17	14.70
				8.63	60.91
					8.37
R_K					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
1.63	14.34	56.55	30.96	-4.03	41.13
	3.88	20.03	-6.13	9.68	28.83
		3.03	12.07	-6.26	36.86
			1.15	16.38	17.48
				8.56	28.12
					2.37
$R_{\psi(2S)}$					
Run 1			Run 2		
e TOS	h TOS!	TIS!	e TOS	h TOS!	TIS!
2.17	40.15	10.28	8.00	10.20	10.16
	3.89	28.72	-5.38	37.69	20.03
		2.15	2.45	14.14	12.74
			1.55	17.48	46.81
				3.60	16.58
					2.14

2308 11 Cross-checks

2309 11.1 Integrated measurement of $r_{J/\psi}$ and $R_{\psi(2S)}$

2310 The integrated $r_{J/\psi}$ and $R_{\psi(2S)}$ values are extracted in *super-*
 2311 *Lenin/src/mainHolyEfficiency(Run2).cc*.

2312 The major systematic uncertainties have been computed for the double ratio. The
 2313 systematic uncertainties for the single ratio do not yet include the material systematics.
 2314 However, even with the existing systematics the single ratios are compatible with unity.

2315 The integrated $r_{J/\psi}$ single ratio and $R_{\psi(2S)}$ double ratio, which are defined in Eq. 6
 2316 and in Eq. 8, are extracted using the fits described in Section 7 and the efficiencies
 2317 computed with the nominal trigger, PID and kinematic corrections discussed in Section 6.
 2318 These ratios are computed in Run 1 and Run 2 separately, using the five inclusive and
 2319 exclusive trigger categories for the electron mode and the μ TOS category for the muon
 2320 mode. The ratios are displayed in Tab. 49. Statistical and systematic uncertainties are
 2321 quoted separately for $R_{\psi(2S)}$, but are not separable in $r_{J/\psi}$, as explained in Section 10. The
 2322 combination of the Run 1 and Run 2 data is also shown. This average, \bar{x} , and its error,
 2323 $\sigma(x)$, are formed taking into account the relevant covariance matrix cov_{ij} (see Section 10),

$$\bar{x} = \sum_i \omega_i x_i, \quad \sigma(x) = \sum_{i,j} \omega_i \omega_j \text{cov}_{ij}$$

2324 where $i, j = e$ TOS Run 1, h TOS! Run 1, TIS! Run 1, e TOS Run 2, h TOS! Run 2, TIS! Run 2
 2325 and,

$$\omega_i = \sum_j ((\text{cov})^{-1})_{ij} / \sum_{i,j} ((\text{cov})^{-1})_{ij}.$$

2326 For $R_{\psi(2S)}$ the systematic uncertainty (derived from $\text{cov}_{ij}^{R_{\psi(2S)}}$) are combined with the
 2327 statistical uncertainty in the yields for electrons (uncorrelated) and muons (100% correlated
 2328 between categories).

2329 As shown in Tab. 50, the $R_{\psi(2S)}$ double ratio is also computed using the narrower
 2330 q^2 window $11.22 < q^2 < 16.40 \text{ GeV}/c^2$ for the electron mode, in order to reduce the
 2331 background from partially reconstructed charm resonances. For example, the background
 2332 from $B^+ \rightarrow \psi(2S)K^+$, $\psi(2S) \rightarrow J/\psi X$ decays has a branching fraction five times larger
 2333 than that of $B^+ \rightarrow K^+\psi(2S)$ decays, and has a poorly known shape. Such backgrounds are
 2334 reduced by an order of magnitude in the narrower q^2 window, as can be seen in Fig. 111. The
 2335 statistical error on $R_{\psi(2S)}$ remains the same in this narrower q^2 region, as the background
 2336 is reduced as well as the signal; while the central value of the combined Run 1 and Run 2
 2337 data is within approximately one standard deviation of unity.

2338 11.2 Differential measurement of $r_{J/\psi}$

2339 The one dimensional $r_{J/\psi}$ plots are made in *superLenin/src/mainMakeRJPsiPlots.cc*
 2340 using the functions from *superLenin/src/rJPsiTools.cpp*.

Table 49: Integrated values of $r_{J/\psi}$ and $R_{\psi(2S)}$ computed for Run 1 and Run 2 data and for the combination of Run 1 and Run 2 data in the standard q^2 window. The muon modes are measured in the μ TOS category and the electron modes are measured in the trigger category shown in the first column. The uncertainty shown is statistical only.

	$r_{J/\psi}$	$R_{\psi(2S)}$
Run 1		
e TOS	1.063 ± 0.064	$0.974 \pm 0.014 \pm 0.020$
h TOS!	1.008 ± 0.071	$0.928 \pm 0.068 \pm 0.036$
TIS!	1.015 ± 0.070	$1.018 \pm 0.031 \pm 0.022$
TIS	1.035 ± 0.072	$0.970 \pm 0.022 \pm 0.020$
h TOS	1.029 ± 0.072	$0.935 \pm 0.036 \pm 0.029$
Run 2		
e TOS	1.052 ± 0.063	$0.969 \pm 0.012 \pm 0.014$
h TOS!	1.053 ± 0.090	$0.907 \pm 0.059 \pm 0.028$
TIS!	1.112 ± 0.092	$0.979 \pm 0.026 \pm 0.020$
TIS	1.079 ± 0.089	$0.944 \pm 0.018 \pm 0.029$
h TOS	1.035 ± 0.089	$0.946 \pm 0.033 \pm 0.020$
Combination		
	1.014 ± 0.035	$0.974 \pm 0.008 \pm 0.010$

Table 50: Integrated value of $R_{\psi(2S)}$ computed for Run 1 and Run 2 data and for the combination of Run 1 and Run 2 data in the standard and narrower q^2 windows. The muon modes are measured in the μ TOS category and the electron modes are measured in the trigger category shown in the first column. The uncertainty shown is statistical only.

	$R_{\psi(2S)}$ normal	$R_{\psi(2S)}$ narrow
Run 1		
e TOS	$0.974 \pm 0.014 \pm 0.020$	$0.984 \pm 0.014 \pm 0.020$
h TOS!	$0.928 \pm 0.068 \pm 0.036$	$0.915 \pm 0.052 \pm 0.035$
TIS!	$1.018 \pm 0.031 \pm 0.022$	$1.048 \pm 0.029 \pm 0.023$
Run 2		
e TOS	$0.969 \pm 0.012 \pm 0.014$	$0.986 \pm 0.011 \pm 0.014$
h TOS!	$0.907 \pm 0.059 \pm 0.028$	$0.889 \pm 0.049 \pm 0.027$
TIS!	$0.979 \pm 0.026 \pm 0.020$	$0.971 \pm 0.023 \pm 0.020$
Combination		
	$0.974 \pm 0.008 \pm 0.010$	$0.986 \pm 0.007 \pm 0.011$

²³⁴¹ To test the data-simulation agreement of one-dimensional projections of kinematic

2342 distributions in $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ decays, the differential $r_{J/\psi}$ test is performed. This
 2343 test does not rely on the *sWeight* method, so there is no possible bias caused by correlations
 2344 between the kinematic variables and the $m_{\text{DTF}}^{J/\psi}$ variable used to perform the fit.

2345 The test proceeds in the following steps:

- 2346 1. Using simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events that pass the stripping requirements,
 2347 each kinematic variable is split into eight uniformly populated bins. This gives
 2348 similar precision in each bin, allowing any trend to be spotted more easily;
- 2349 2. For both data $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events, fits to $m_{\text{DTF}}^{J/\psi}$
 2350 are performed to extract the yields of events in each of the eight bins. Two histograms
 2351 storing the yields are obtained, one for the electron mode and one for the muon
 2352 mode. They are denoted $h_{N,e}^i$ and $h_{N,\mu}^i$ with $1 \leq i \leq 8$;
- 2353 3. Using simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events, two effi-
 2354 ciency histograms are produced. Each of the eight bins of these histograms represents
 2355 the probability of an event being reconstructed in that particular bin, denoted $h_{\varepsilon,e}^i$
 2356 and $h_{\varepsilon,\mu}^i$;
- 2357 4. Each yield histogram is divided by the corresponding efficiency histogram to obtain
 2358 two efficiency-corrected yield histograms $h_{y,e}^i = h_{N,e}^i/h_{\varepsilon,e}^i$ and $h_{y,\mu}^i = h_{N,\mu}^i/h_{\varepsilon,\mu}^i$.
 2359 Each bin of the efficiency-corrected yield histogram represents the total number
 2360 of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ or $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ decay that were generated in
 2361 the detector, before any reconstruction effect. Hence, if the data and simulated
 2362 distributions of the kinematic variable of interest agree perfectly, the efficiency-
 2363 corrected yield histograms will have uniformly populated bins;
- 2364 5. Finally, the muon efficiency-corrected yield histogram is divided by the electron
 2365 efficiency-corrected yield histogram to obtain the $r_{J/\psi}$ histogram $h_{r_{J/\psi}}^i = h_{y,\mu}^i/h_{y,e}^i$,
 2366 each bin of which should be compatible with unity assuming a perfect agreement
 2367 between data and simulation.

2368 The simulation, particularly in the case of the electron mode, is however imperfect
 2369 and some $r_{J/\psi}$ histograms are expected to show some level of variation across the bins.
 2370 To quantify these variations, the $r_{J/\psi}$ flatness parameter d_f is computed. This parameter
 2371 can be interpreted as follows: if the deviations from flatness across bins of the efficiency-
 2372 corrected yield histogram are assumed to be genuine variations of the efficiency from a
 2373 perfectly flat distribution then, given the distribution of events in the rare decay, what
 2374 would the resulting deviation in the double ratio R_K be? More explicitly, d_f is computed
 2375 as follows:

$$d_f = \frac{\sum_{i=1}^8 h_{\varepsilon,\mu}^{\text{rare},i} \cdot h_{y,\mu}^i}{\sum_{i=1}^8 h_{\varepsilon,\mu}^i \cdot h_{y,\mu}^i} \cdot \frac{\sum_{i=1}^8 h_{\varepsilon,\mu}^i}{\sum_{i=1}^8 h_{\varepsilon,\mu}^{\text{rare},i}} \sqrt{\frac{\sum_{i=1}^8 h_{\varepsilon,e}^{\text{rare},i} \cdot h_{y,e}^i}{\sum_{i=1}^8 h_{\varepsilon,e}^i \cdot h_{y,e}^i} \cdot \frac{\sum_{i=1}^8 h_{\varepsilon,e}^i}{\sum_{i=1}^8 h_{\varepsilon,e}^{\text{rare},i}} - 1}. \quad (42)$$

2376 For each $r_{J/\psi}$ histogram, d_f must be comparable to the size of the total systematic
2377 uncertainty expected on R_K . A larger d_f would indicate an extra source of systematic
2378 uncertainty that does not cancel in the ratio $\varepsilon(B^+ \rightarrow K^+\ell^+\ell^-)/\varepsilon(B^+ \rightarrow K^+J/\psi(\ell^+\ell^-))$.

2379 The $r_{J/\psi}$ histograms are shown in Fig. 112–Fig. 121, for Run 1 and Run 2, and for
2380 the electron mode selected in all five L0 categories (the muon mode is always in the
2381 μ TOS category). For each run and trigger category, the $r_{J/\psi}$ histogram is extracted using
2382 efficiencies computed in three ways. First, using the trigger and PID corrections but no
2383 kinematic corrections. Second, using the nominal kinematic correction weights. Third, the
2384 rapidity corrections discussed in Section 10.7 are added to the electron mode.

2385 The following observations can be made:

- 2386 • The transverse momentum plots are generally very flat, except in the h TOS and
2387 h TOS! categories. This is expected as the h TOS efficiency calibration is imperfect;
- 2388 • The plots related to the angles (rapidity or angles between the different tracks) show
2389 a clear trend when the efficiencies are computed in the nominal way, with the high
2390 rapidity bins being $\sim 20\%$ higher than the low bins. This impacts the momentum
2391 distributions, too. Adding the rapidity weights yield a very good improvement;
- 2392 • In Run 1, the χ^2_{DV} shows a trend, the last bin being $\sim 15\%$ below the first bin. The
2393 systematic uncertainty related to this is taken into account when computing the
2394 systematic uncertainty related to the kinematic reweighing scheme, as discussed
2395 in Section 10.2.

2396 All flatness parameters are less than a few percent, the vast majority of them are at the
2397 permille level. This again illustrates the robustness of the ratio $\varepsilon(B^+ \rightarrow K^+\ell^+\ell^-)/\varepsilon(B^+ \rightarrow$
2398 $K^+J/\psi(\ell^+\ell^-))$.

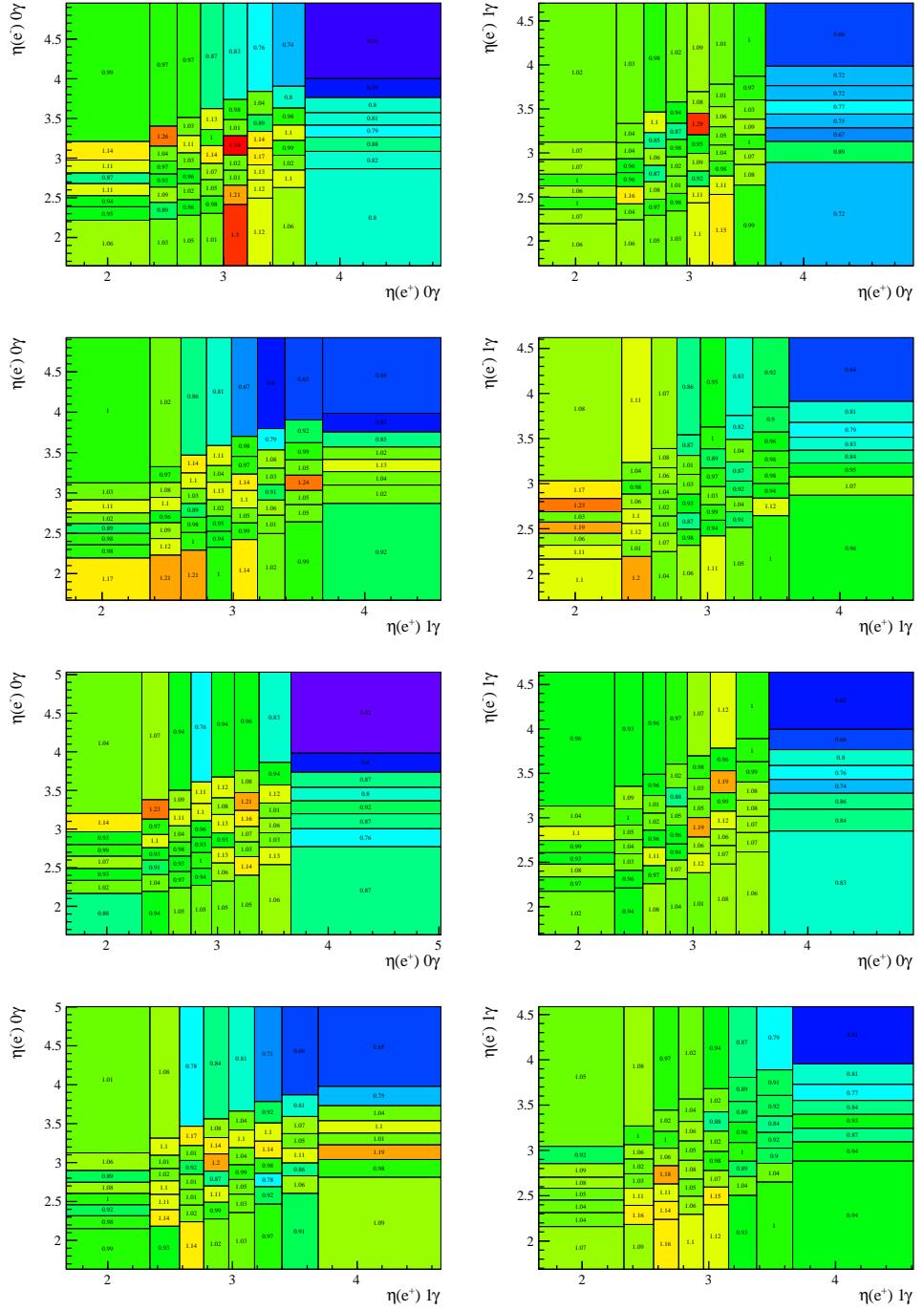


Figure 106: Data over simulation ratios of the two dimensional electron rapidity distributions for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, for Run 1 (top four plots) and Run 2 (bottom four plots).

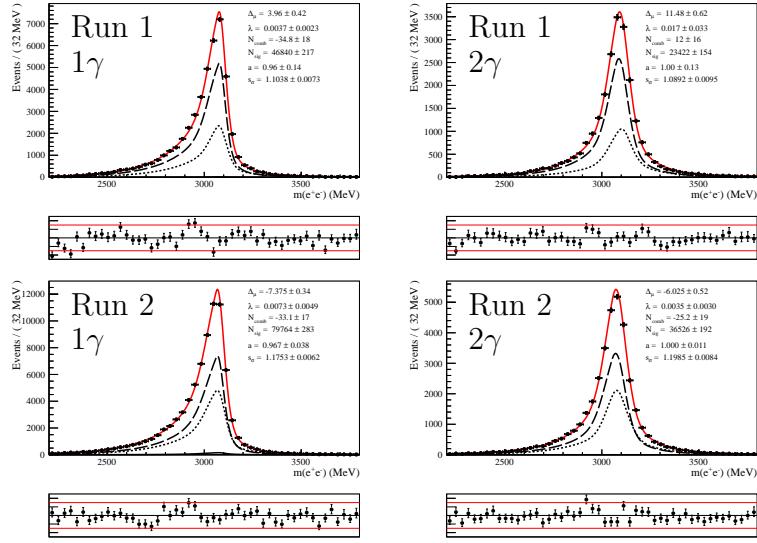


Figure 107: Fit to $m(e^+e^-)$ for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events in the Run 1 (top) and Run 2 (bottom) data taking conditions, in bremsstrahlung categories 1 and 2. The signal peak is modelled with three shapes in three nSPDHits bins: the solid black line corresponds to the lowest occupancy, the long dashed line to the central bin and the short dashed line to the upper bin.

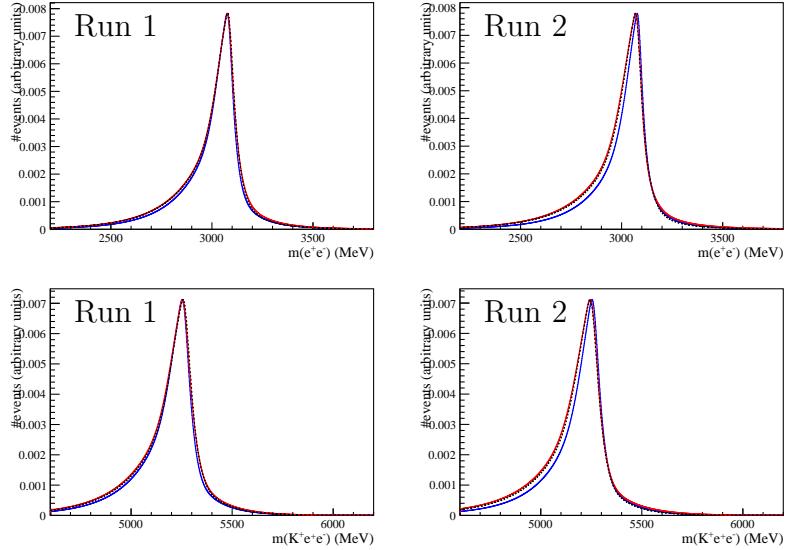


Figure 108: Fit functions to the $m(e^+e^-)$ (top) and $m(K^+e^+e^-)$ distributions in data (dotted black), simulation smeared in the nominal way (red), and simulation smeared in the alternative way (green). All functions are extracted from a fit to fully selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ eTOS events, in all bremsstrahlung categories superimposed and in the Run 1 (left) and Run 2 (right) data taking conditions.

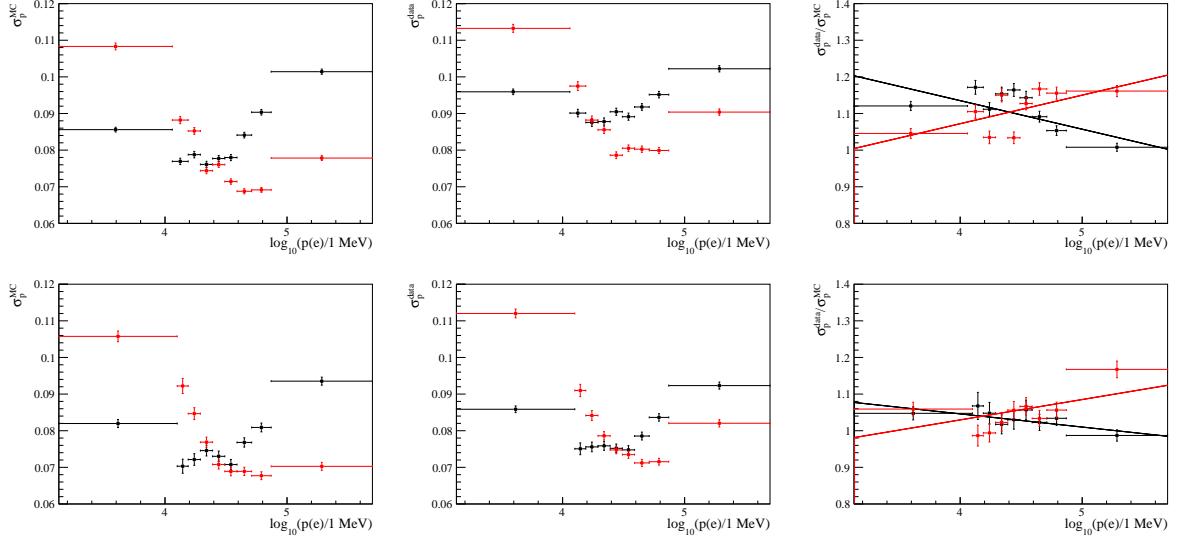


Figure 109: Momentum resolution as a function of the momentum for data (left) simulation (centre), and ratio (right) for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 1 (top) and Run 2 (bottom) data-taking conditions. Electrons with (red) and without (black) a bremsstrahlung cluster added are considered.

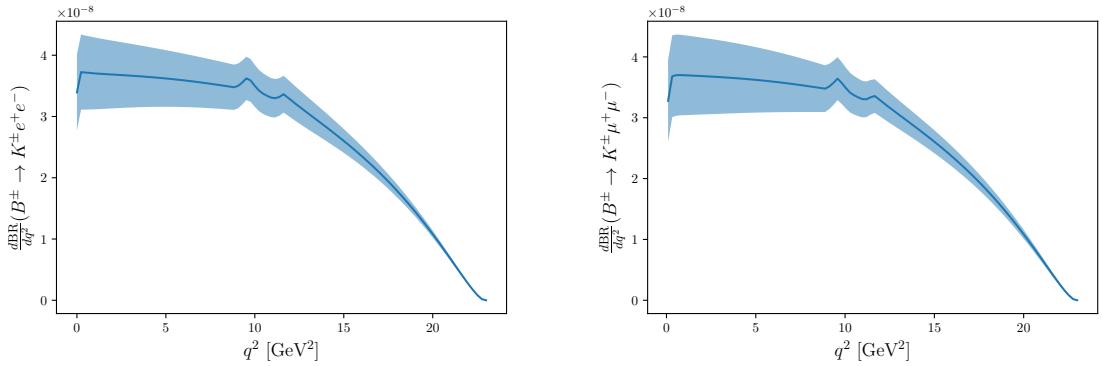


Figure 110: Differential branching fractions predictions from the flavio software for (left) $B^+ \rightarrow K^+ e^+ e^-$ and (right) $B^+ \rightarrow K^+ \mu^+ \mu^-$. The blue band shows the theoretical uncertainty that results from an extrapolation of the lattice calculation at high q^2 . The predictions in the region of narrow charmonium resonances are not meaningful.

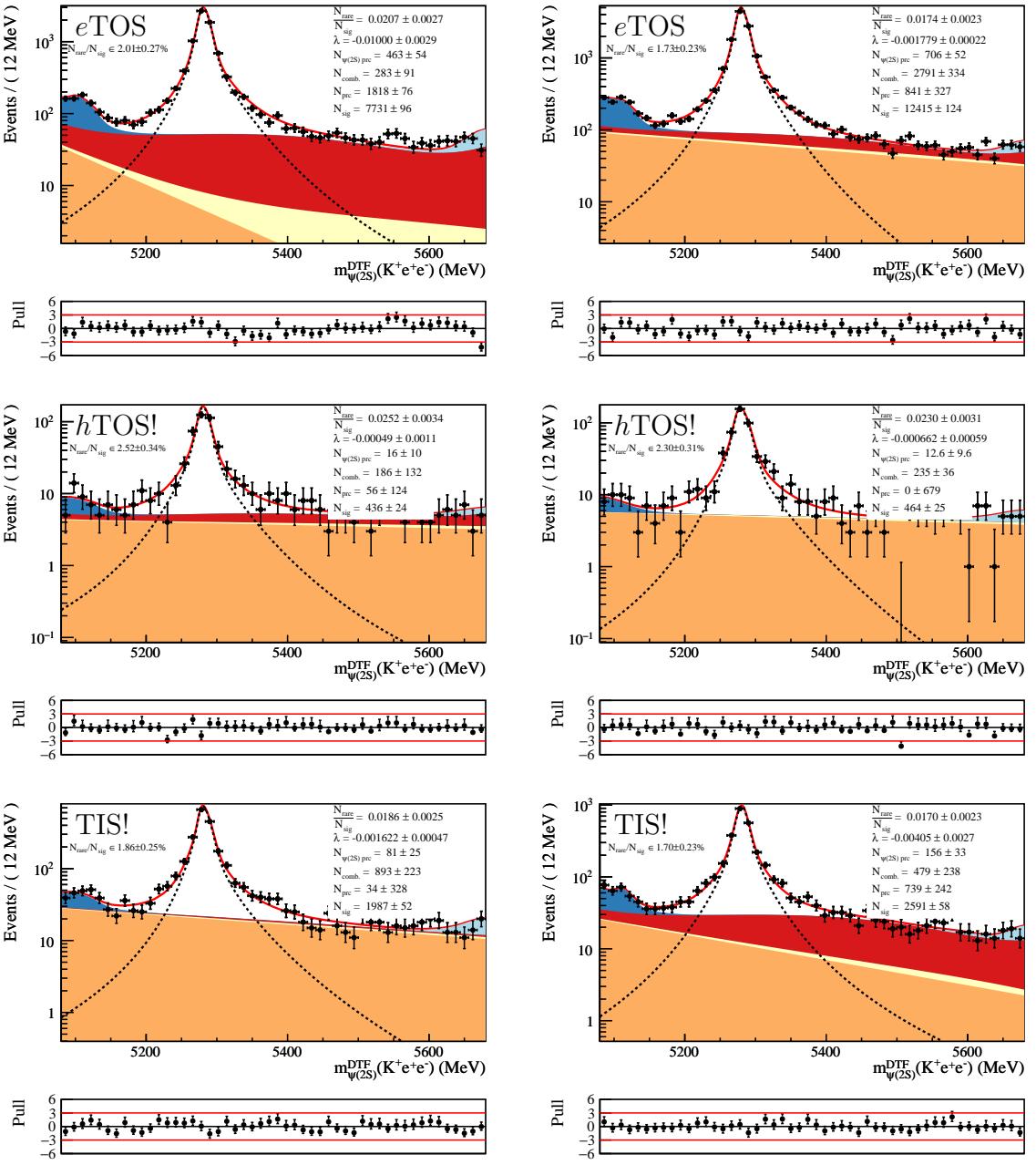


Figure 111: Fits to the $m_{\text{DTF}}^{\psi(2S)}(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ selected candidates, within a narrower q^2 window of $11.22 < q^2 < 16.40 \text{ GeV}/c^2$, in Run 1 (left) and Run 2 (right) samples, for *eTOS* (top), *hTOS!* and *TIS!* trigger categories. The $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ signal is represented as a dashed black line. Background components coming from misreconstructed $B^0 \rightarrow K^{*0}\psi(2S)(ee)$, $B^+ \rightarrow K^+\psi(2S)(J/\psi X)$ and $B^+ \rightarrow K^+J/\psi(e^+e^-)$ events are shown as a blue, red and light-blue shaded area, respectively. Contaminations from the rare decay $B^+ \rightarrow K^+e^+e^-$ and combinatorial background are depicted as yellow and orange areas respectively.

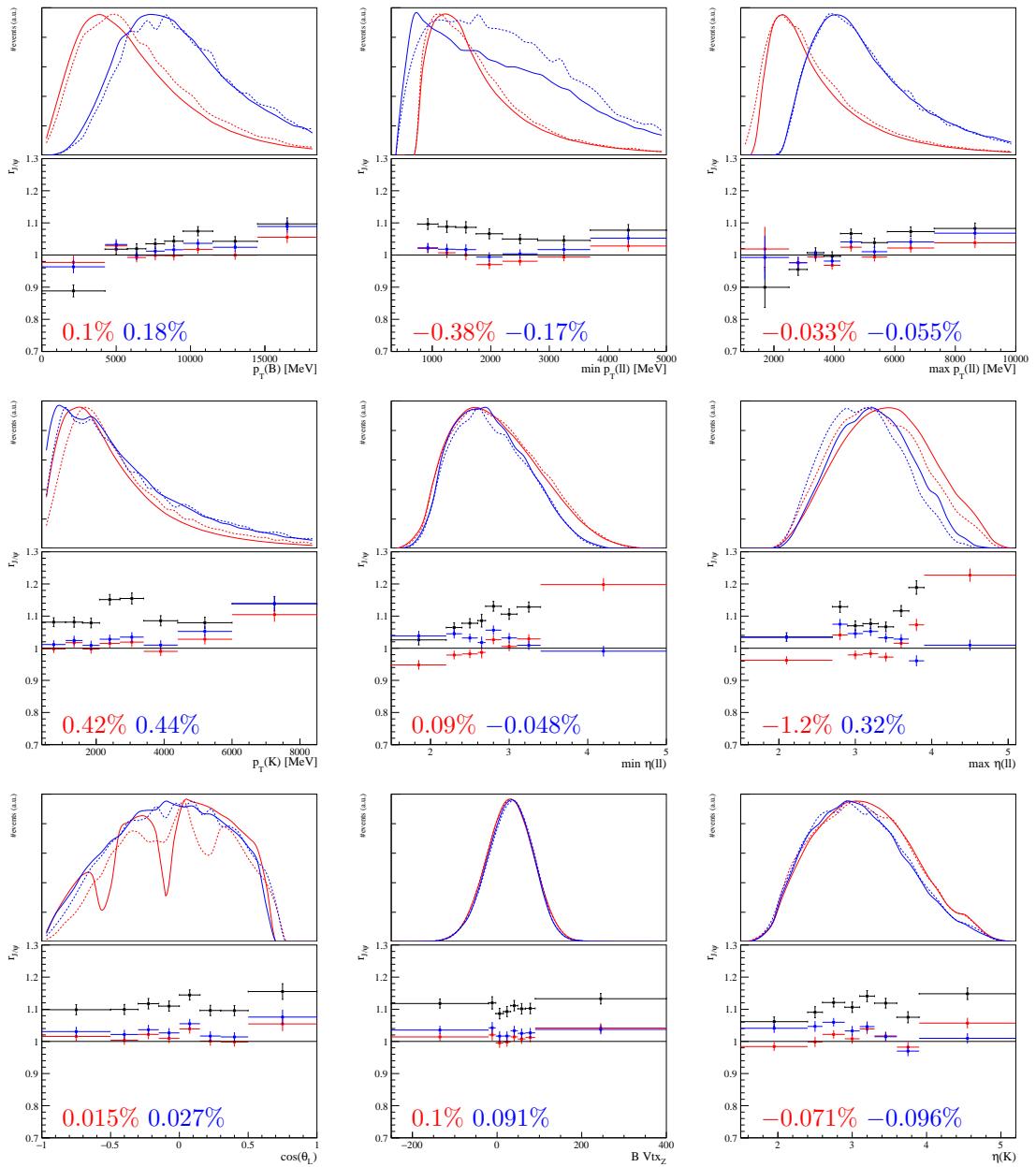


Figure 112: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the e TOS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue points are computed applying the extra rapidity correction to the electron mode. The flatness parameters d_f associated with the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

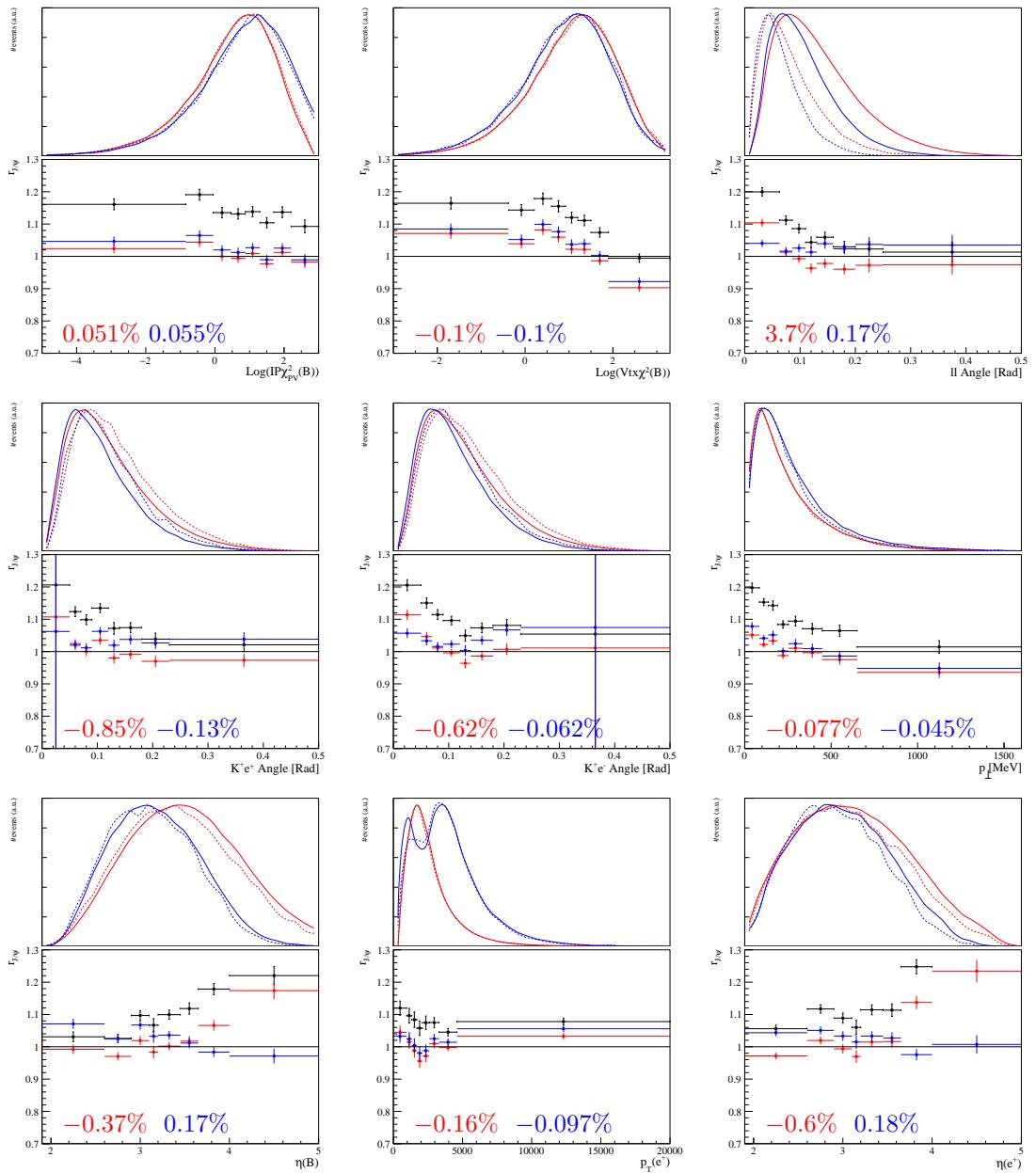


Figure 112: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode in the $e\text{TOS}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode in the μTOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue points are computed applying the extra rapidity correction to the electron mode. The flatness parameters d_f associated with the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (solid blue) are shown.

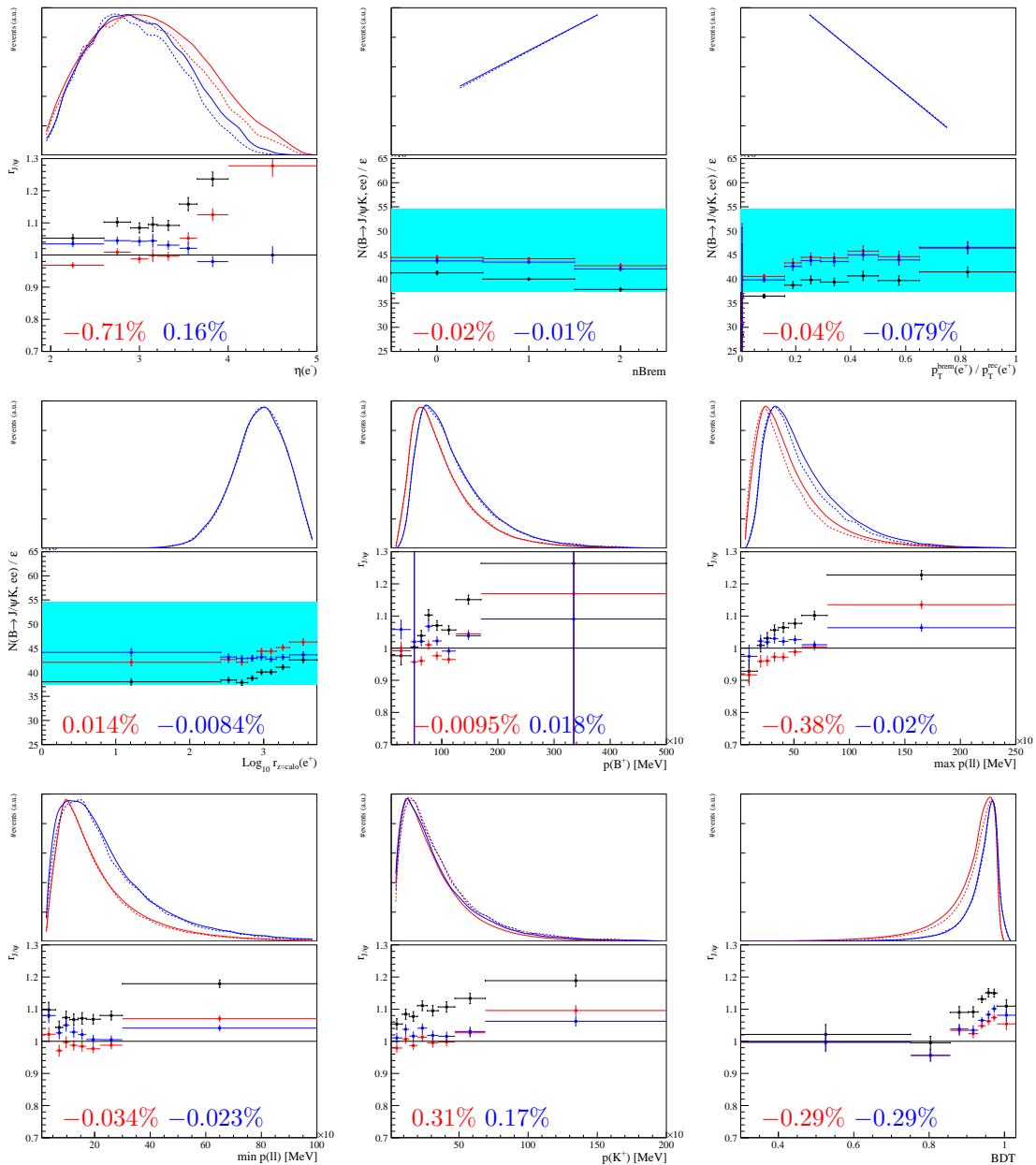


Figure 112: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the e TOS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue points are computed applying the extra rapidity correction to the electron mode. The flatness parameters d_f associated with the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

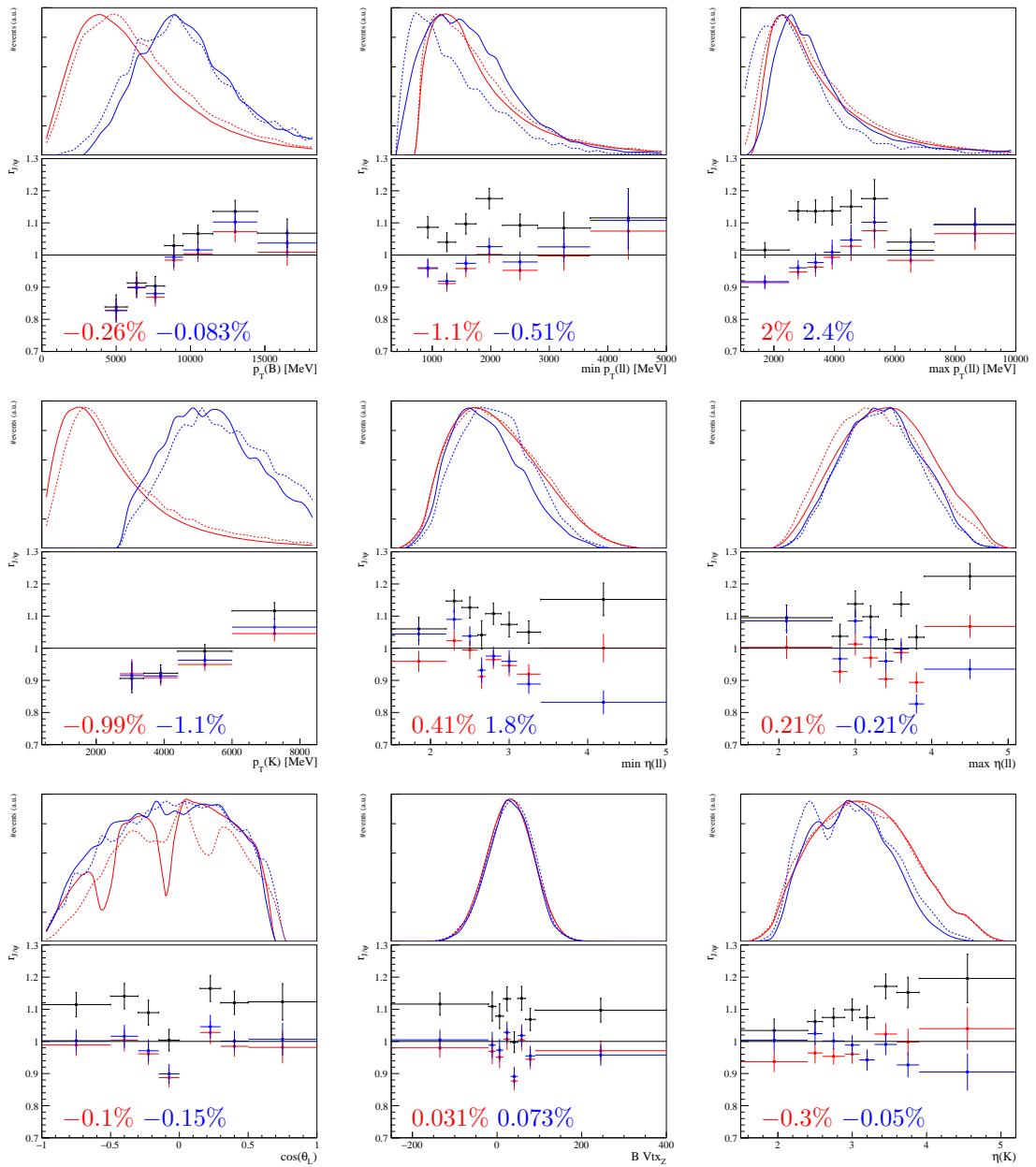


Figure 113: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $h\text{TOS!}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

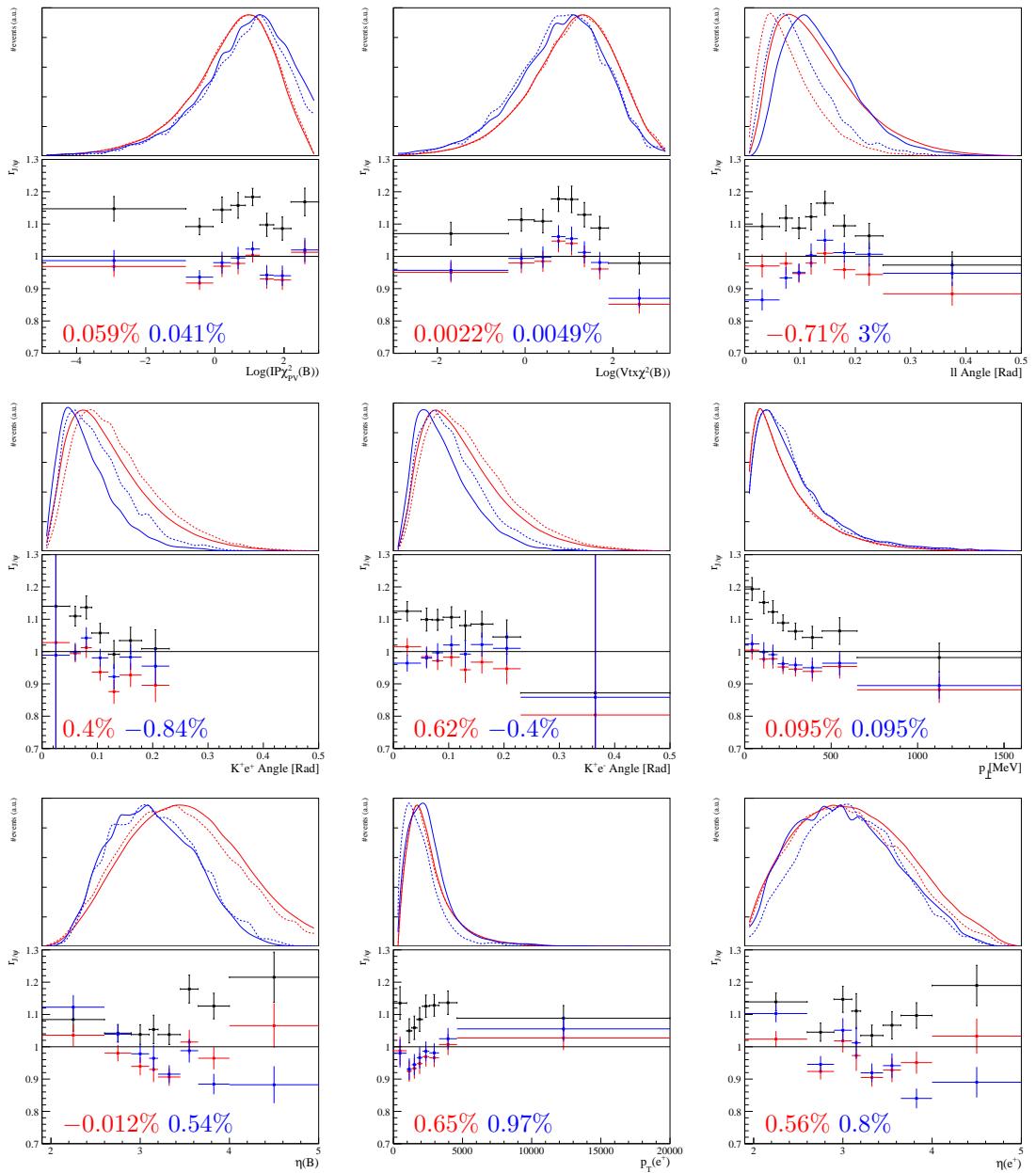


Figure 113: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $h\text{TOS!}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

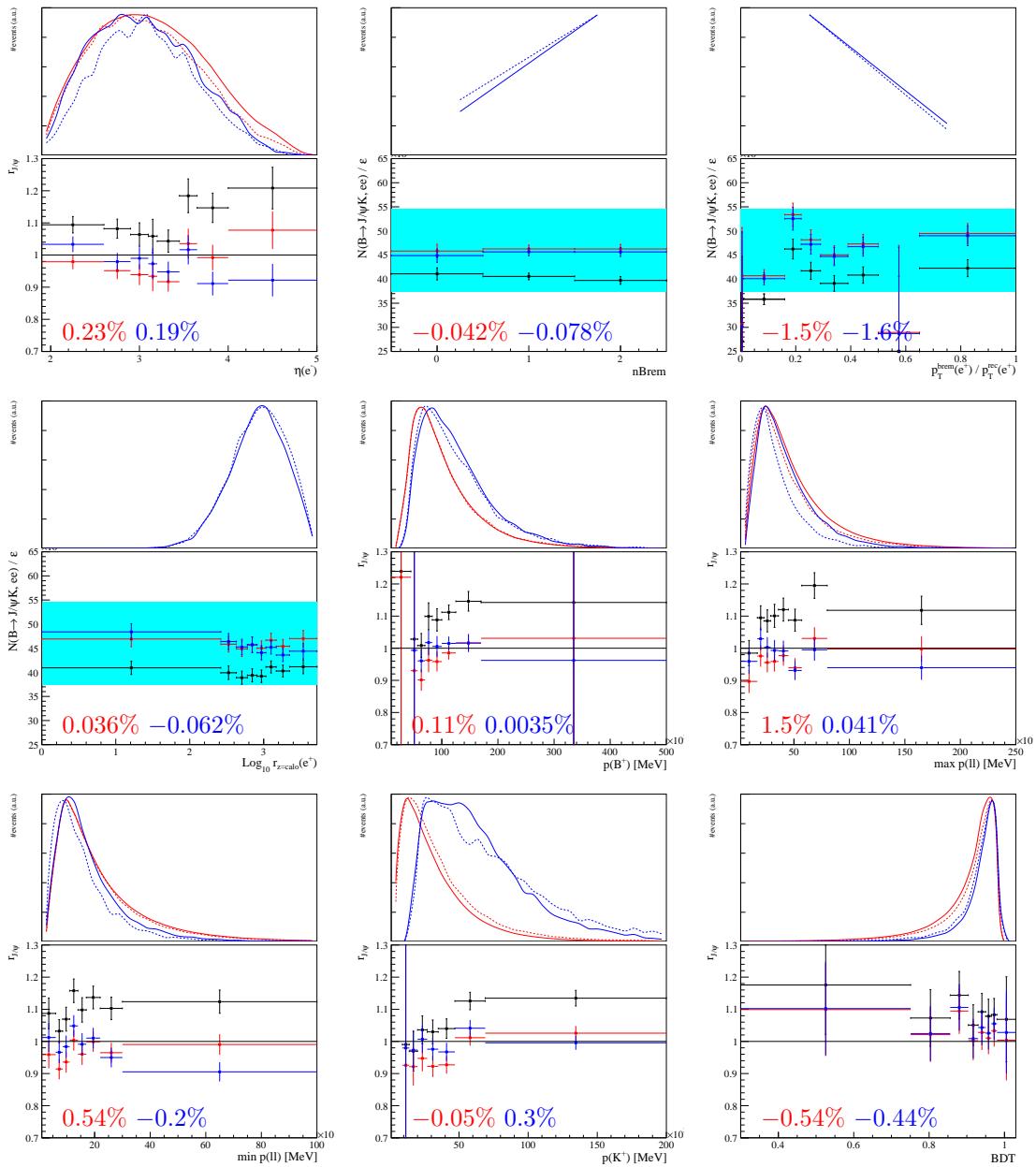


Figure 113: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $h\text{TOS!}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

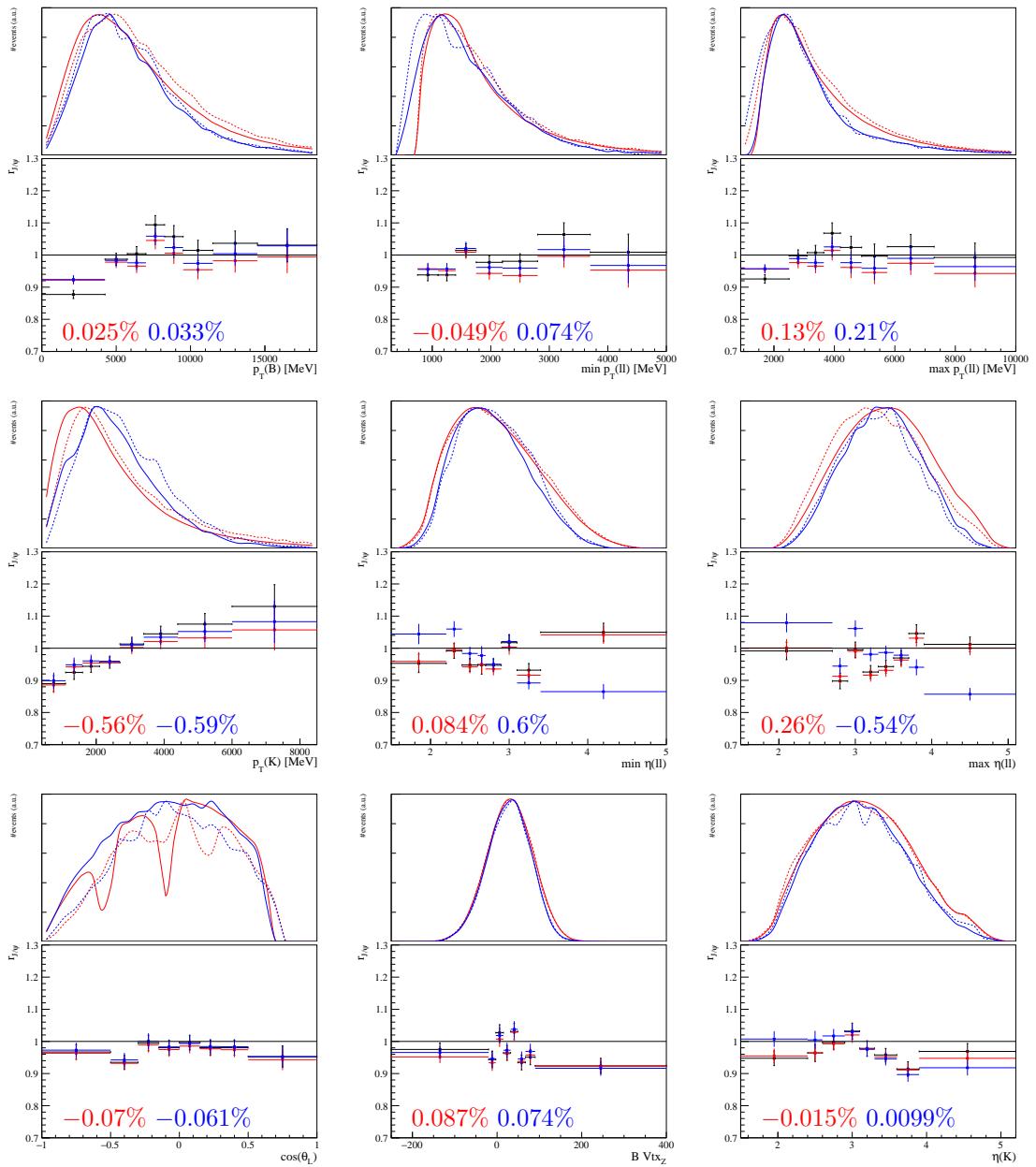


Figure 114: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS! category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

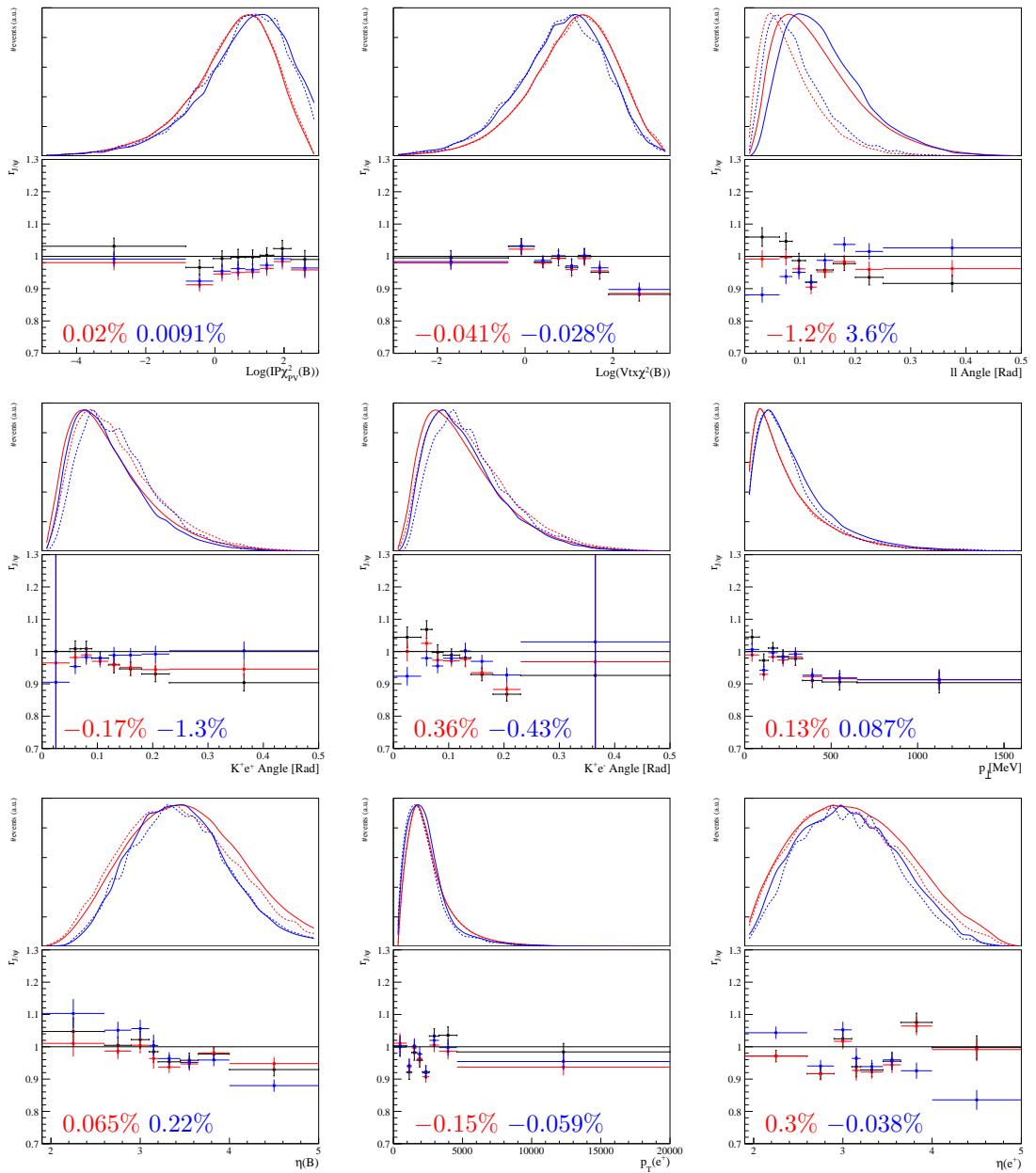


Figure 114: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+J/\psi(e^+e^-)$ mode in the TIS! category and the $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+\mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+J/\psi(e^+e^-)$ (solid blue) are shown.

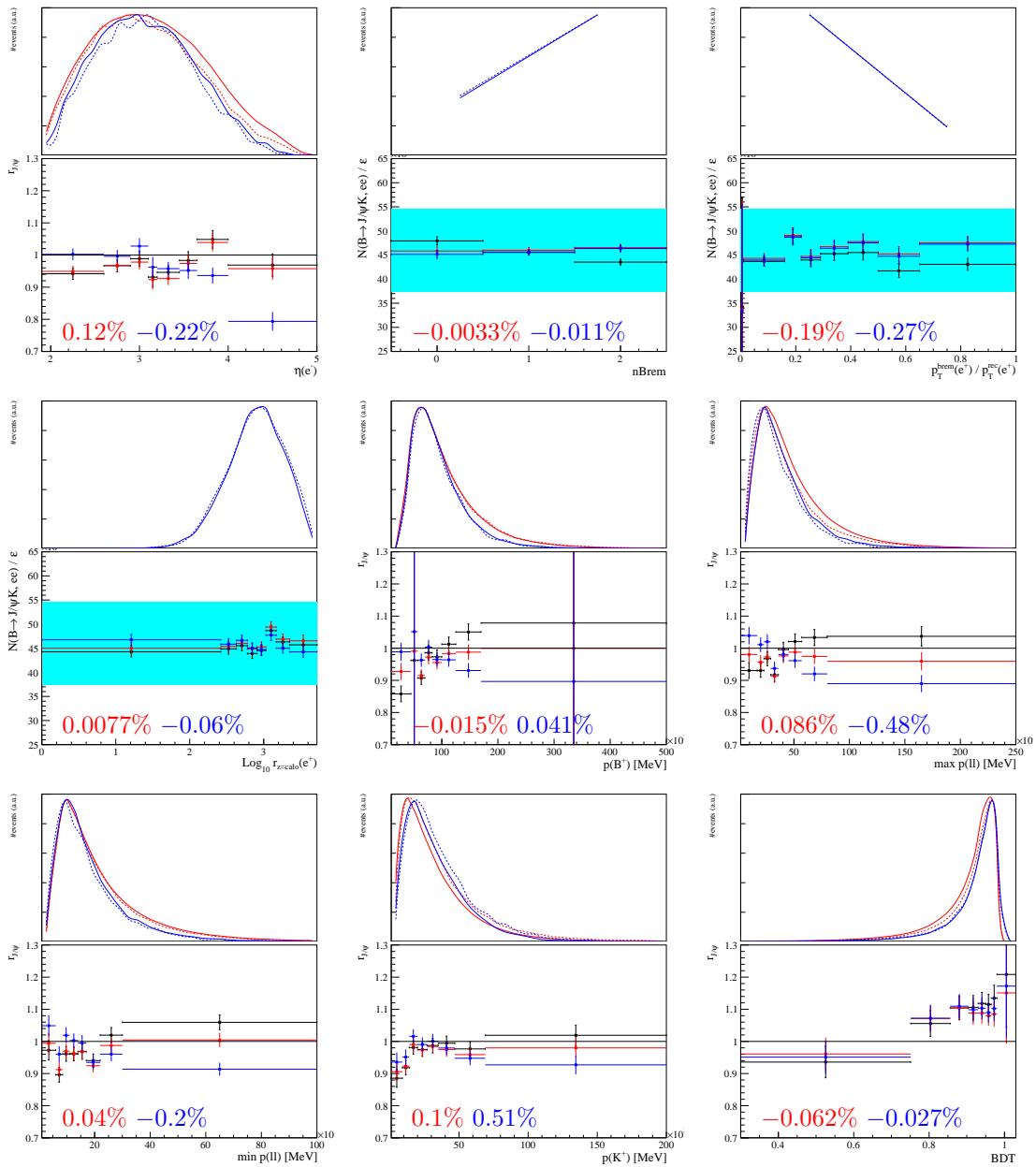


Figure 114: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS! category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

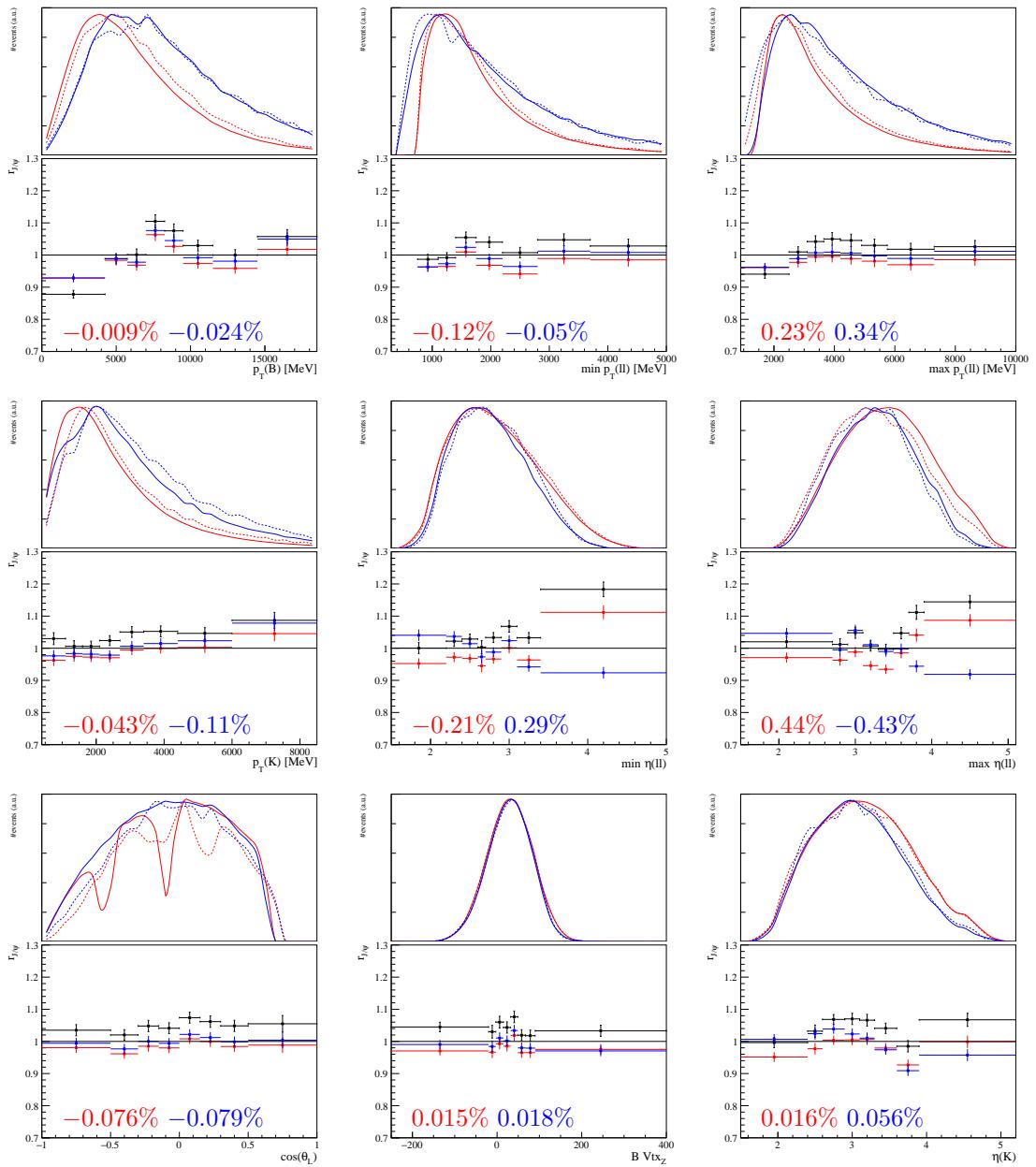


Figure 115: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{\gamma,e}^i$ is displayed instead of At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

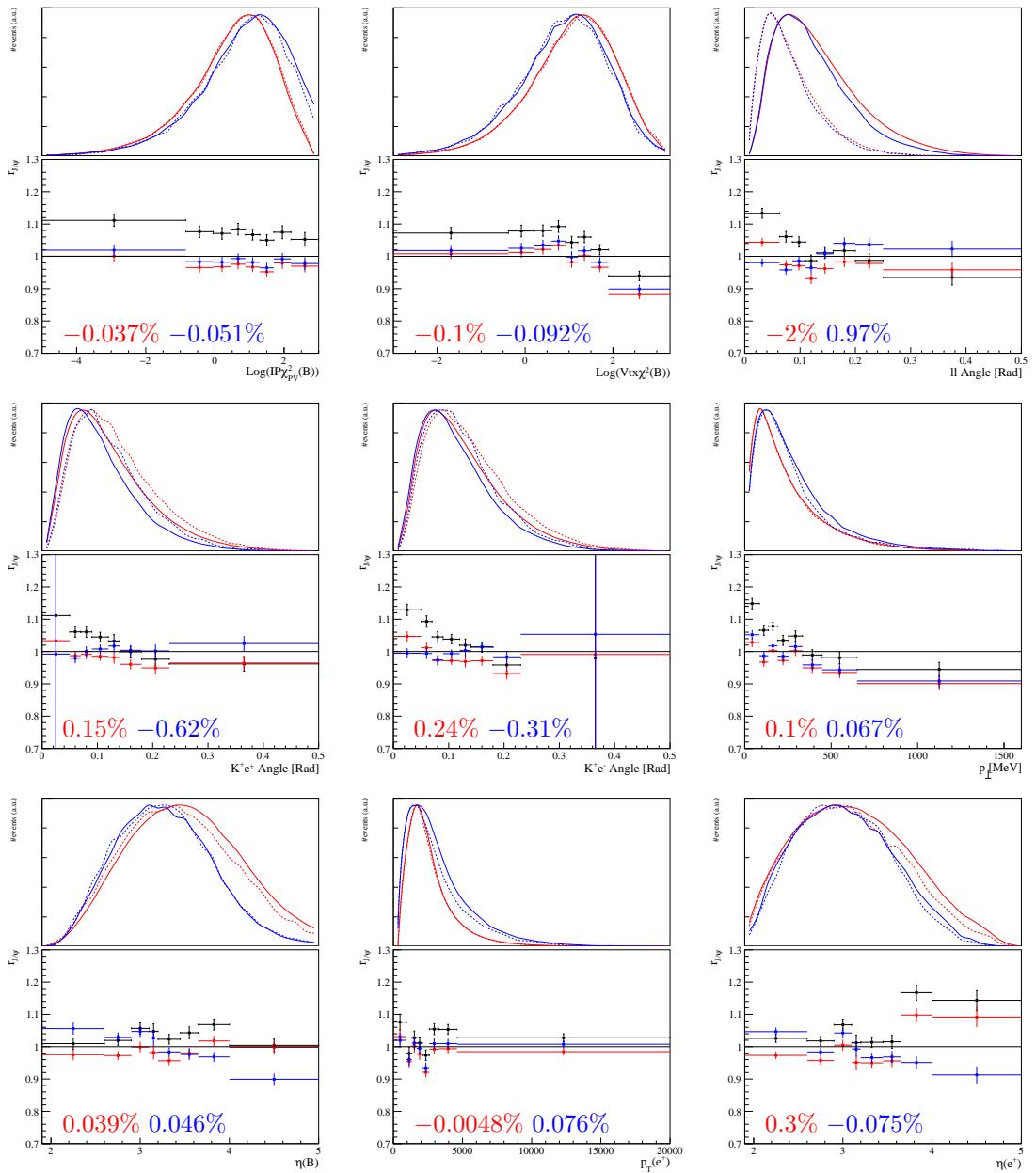


Figure 115: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

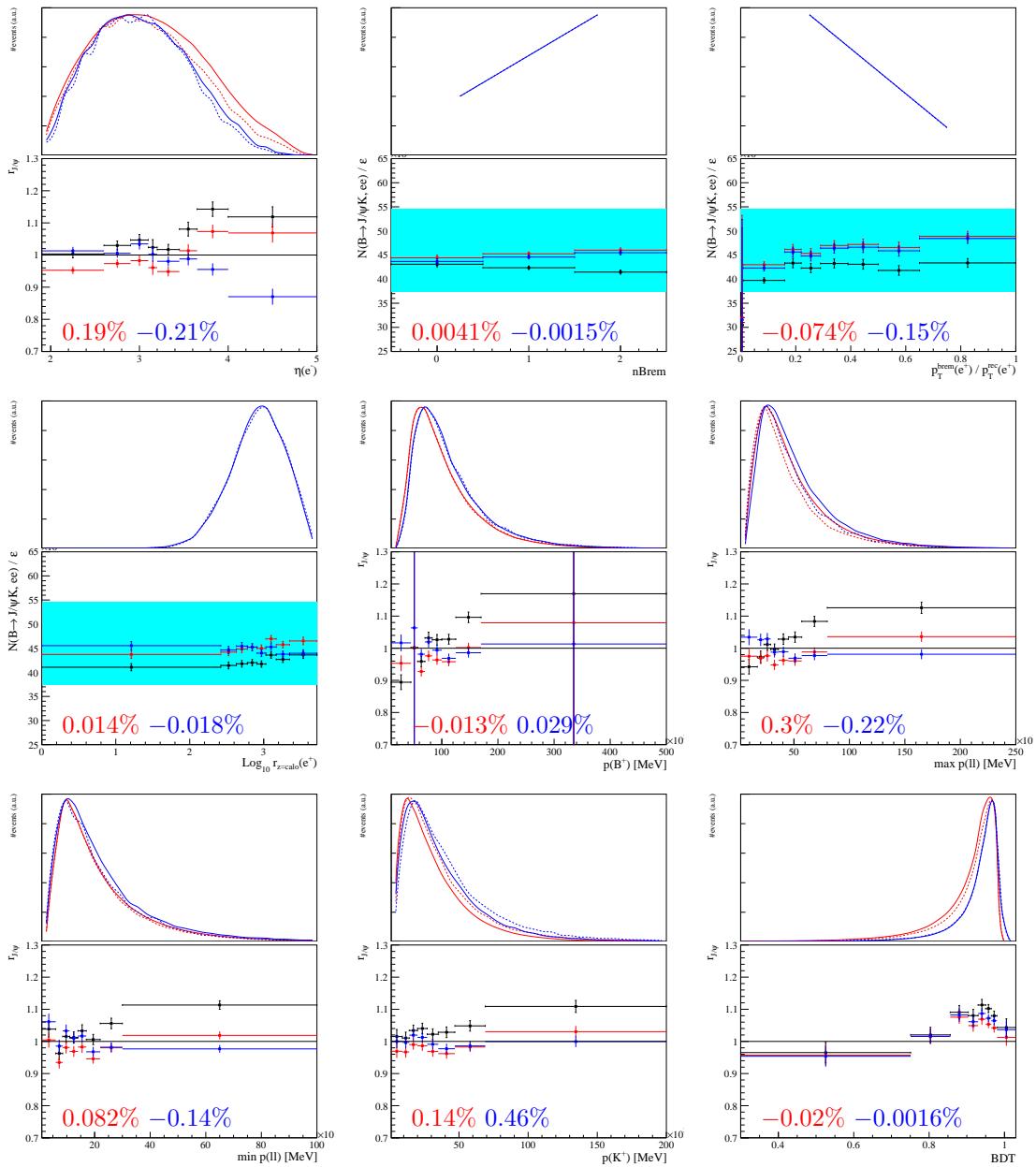


Figure 115: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

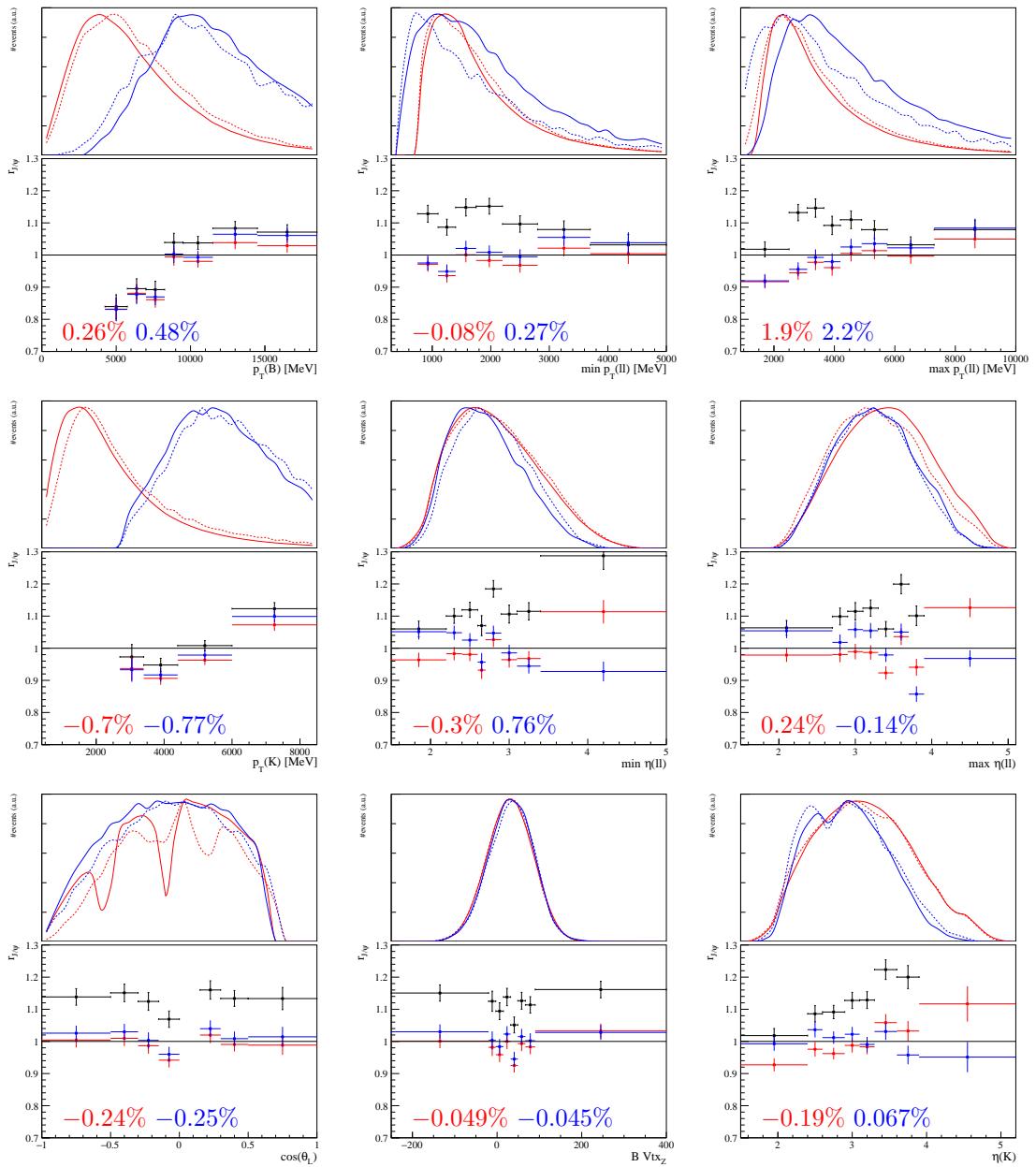


Figure 116: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $h\text{TOS}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

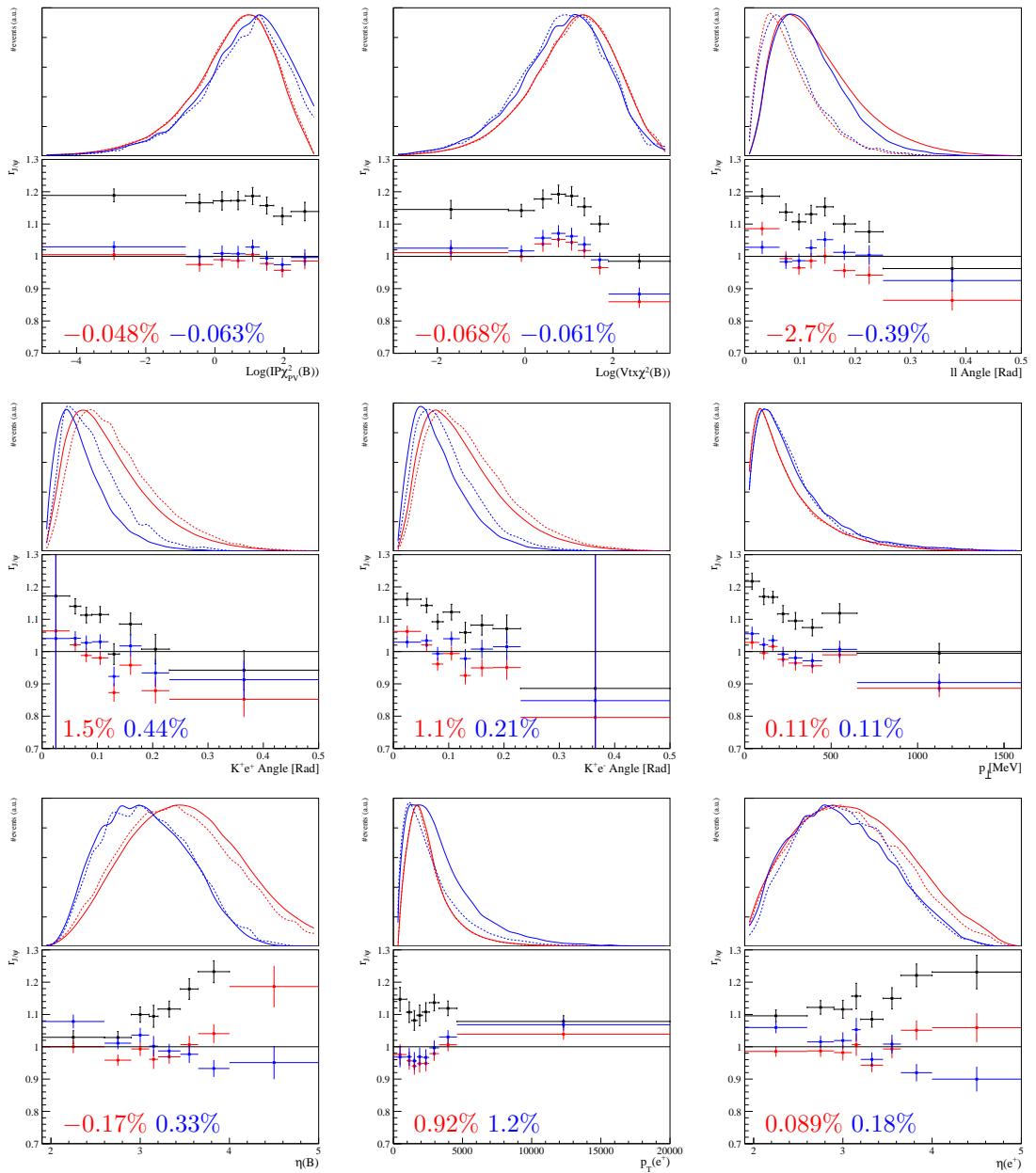


Figure 116: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $h\text{TOS}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+\mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+J/\psi(e^+e^-)$ (solid blue) are shown.

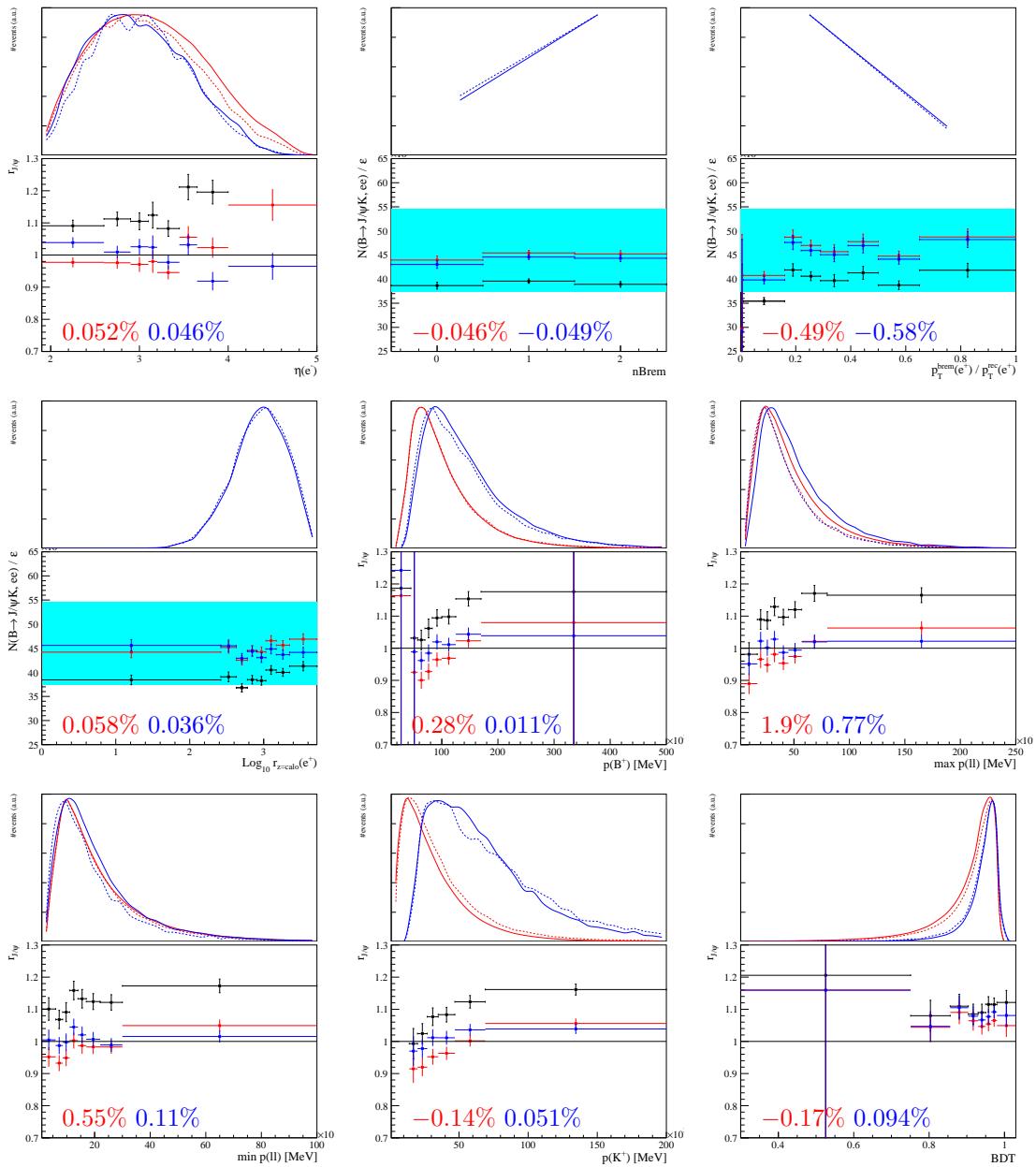


Figure 116: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode in the h TOS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode in the μ TOS category, in the Run 1 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi, e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (solid blue) are shown.

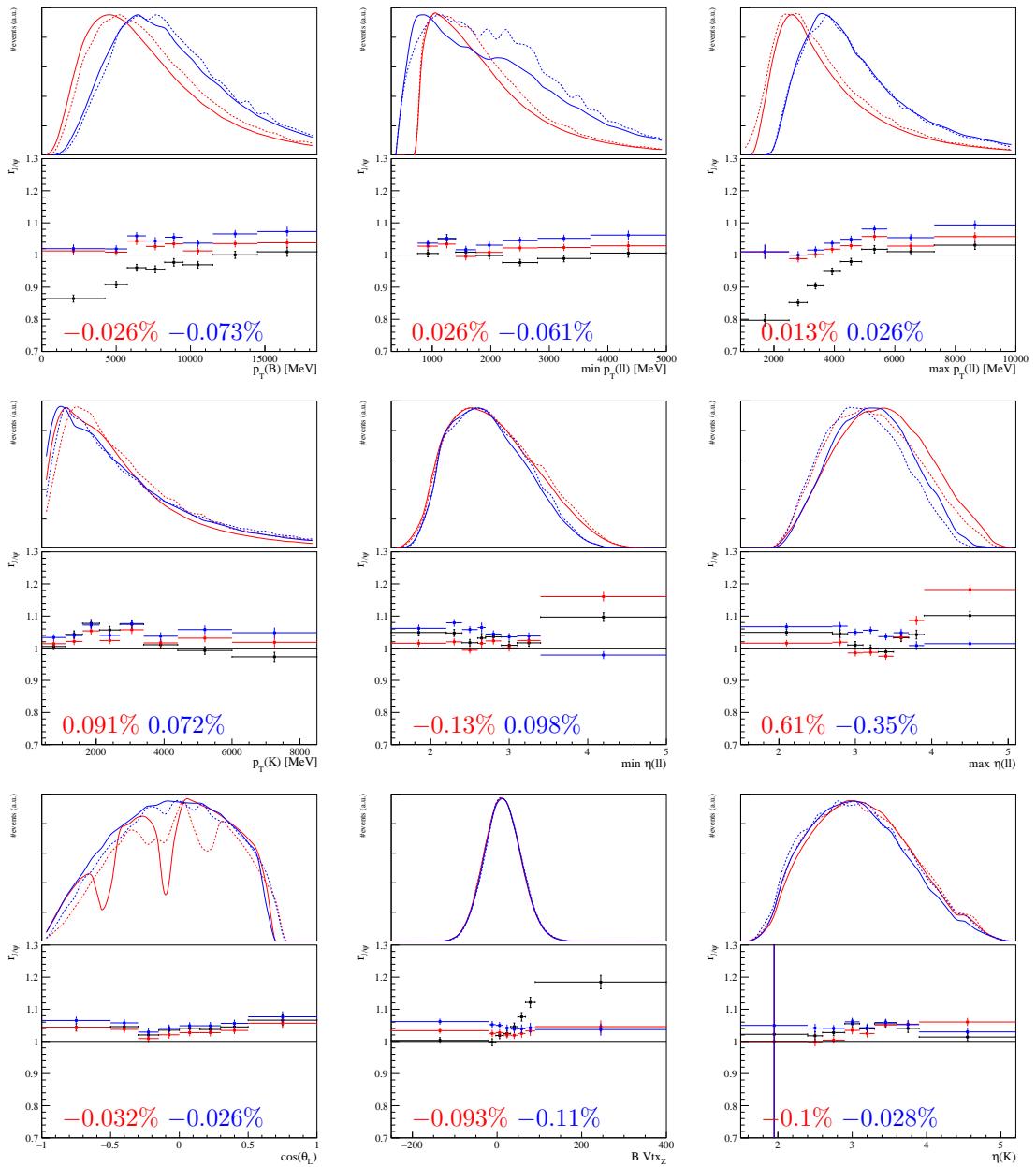


Figure 117: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode in the e TOS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode in the μ TOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (solid blue) are shown.

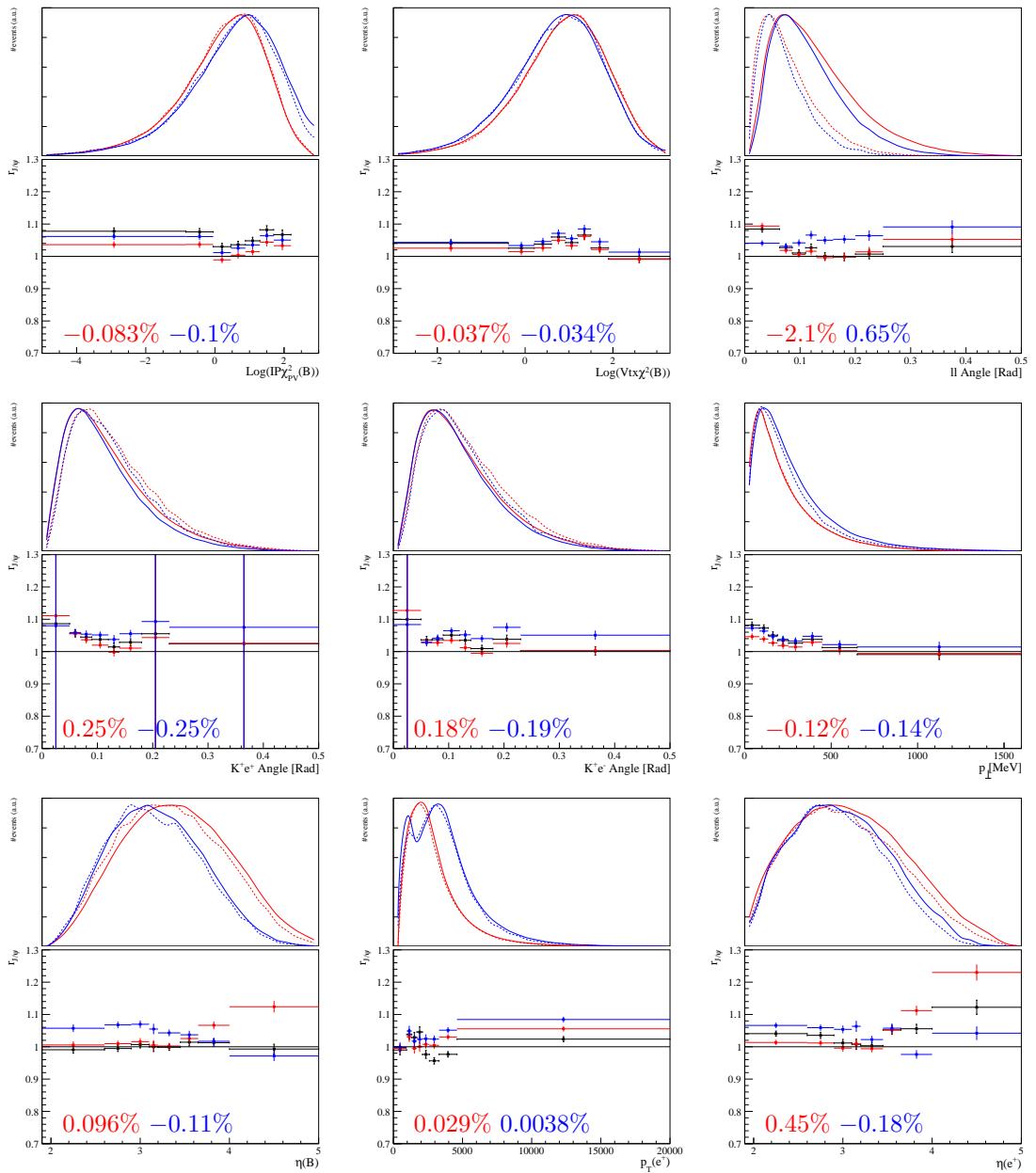


Figure 117: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the e TOS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

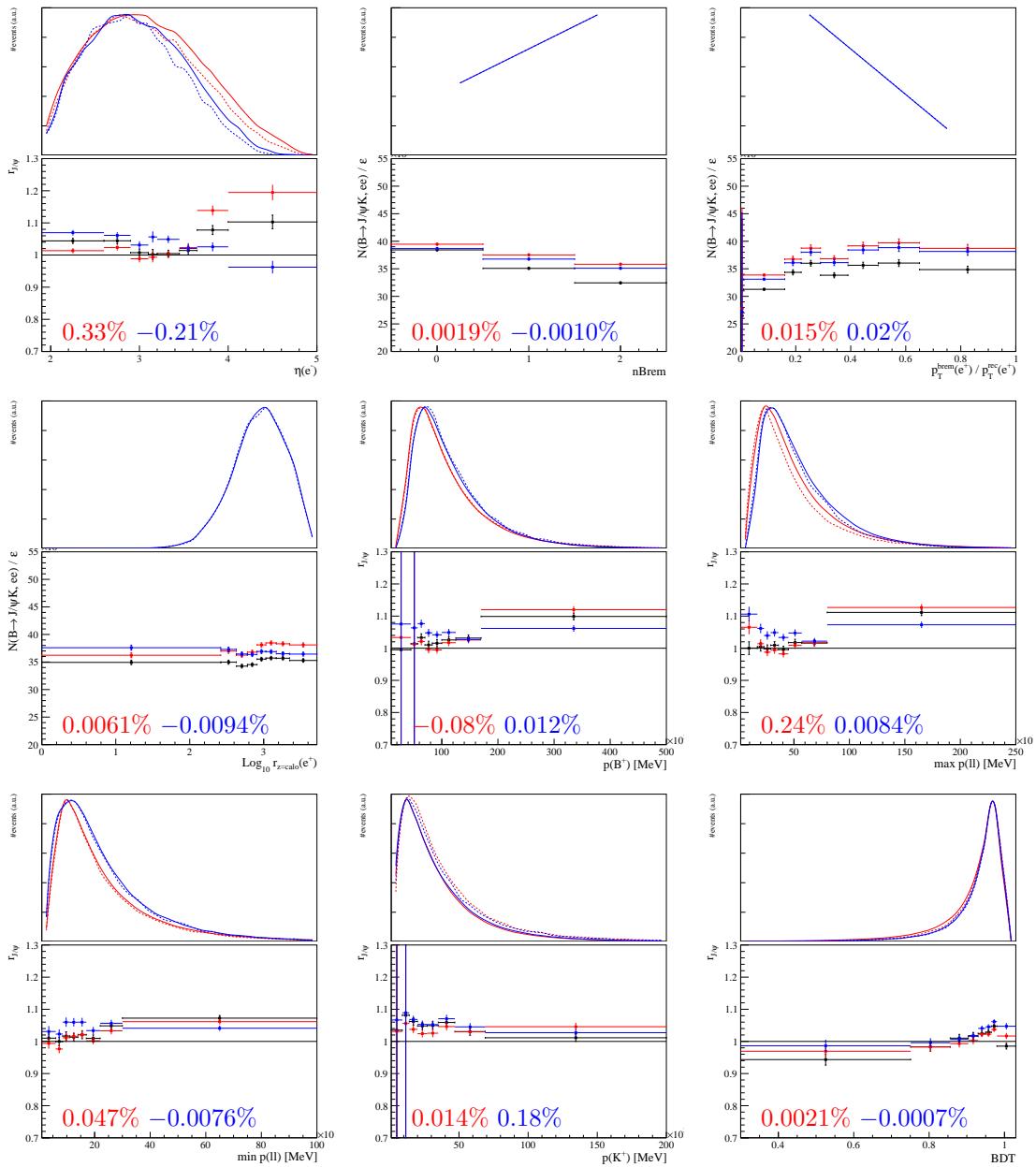


Figure 117: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $e\text{TOS}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

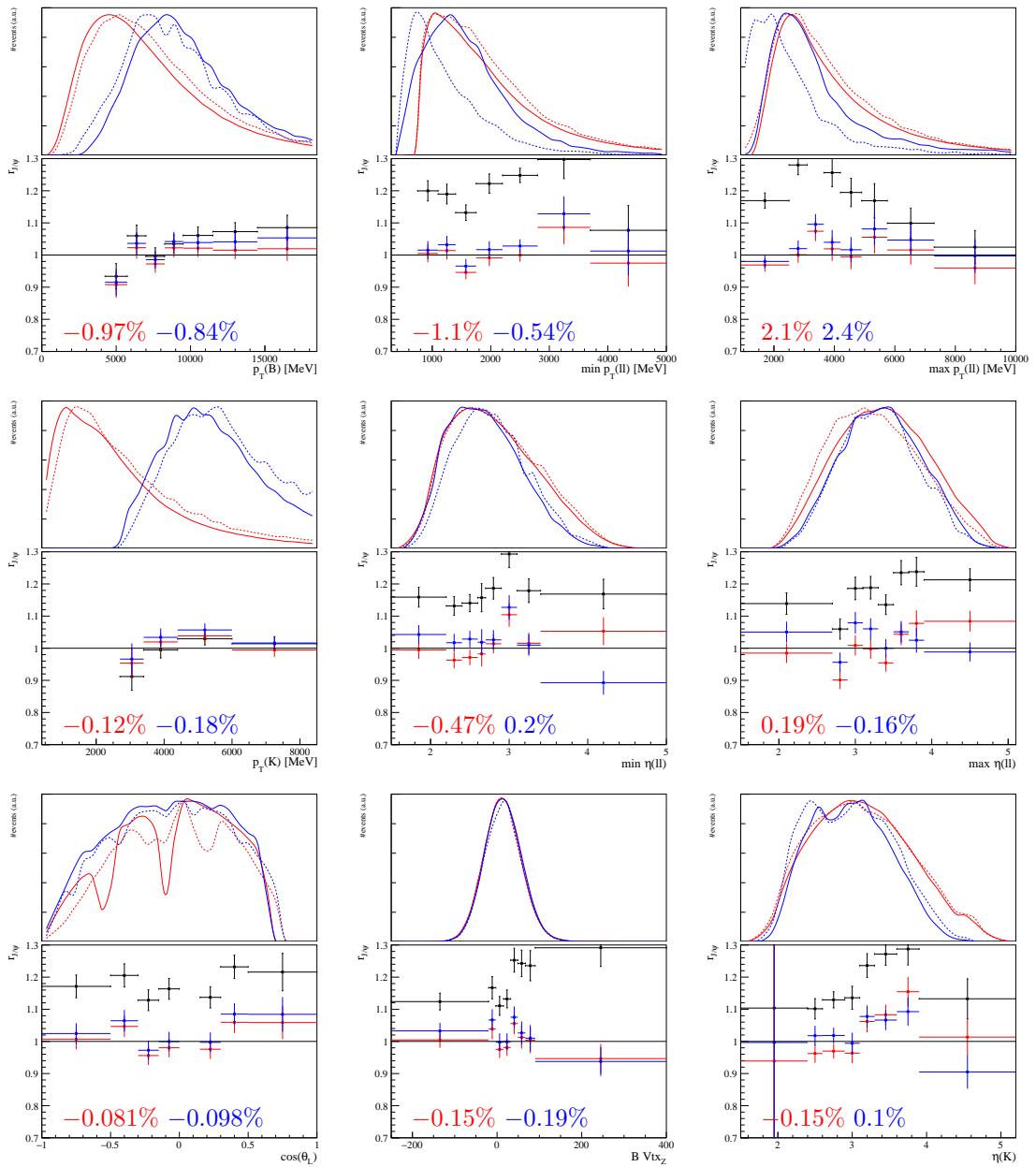


Figure 118: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $h\text{TOS!}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

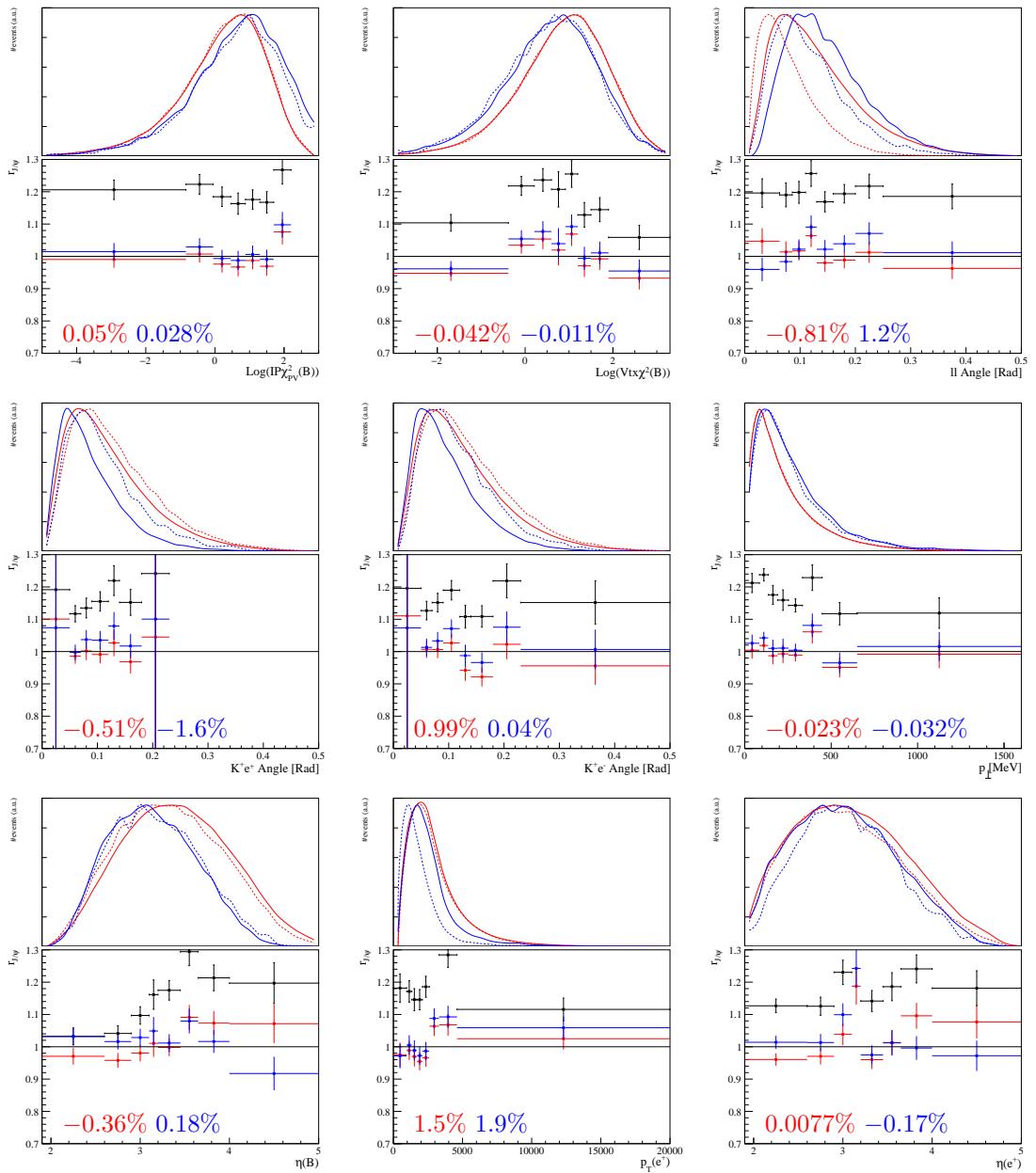


Figure 118: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $h\text{TOS!}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

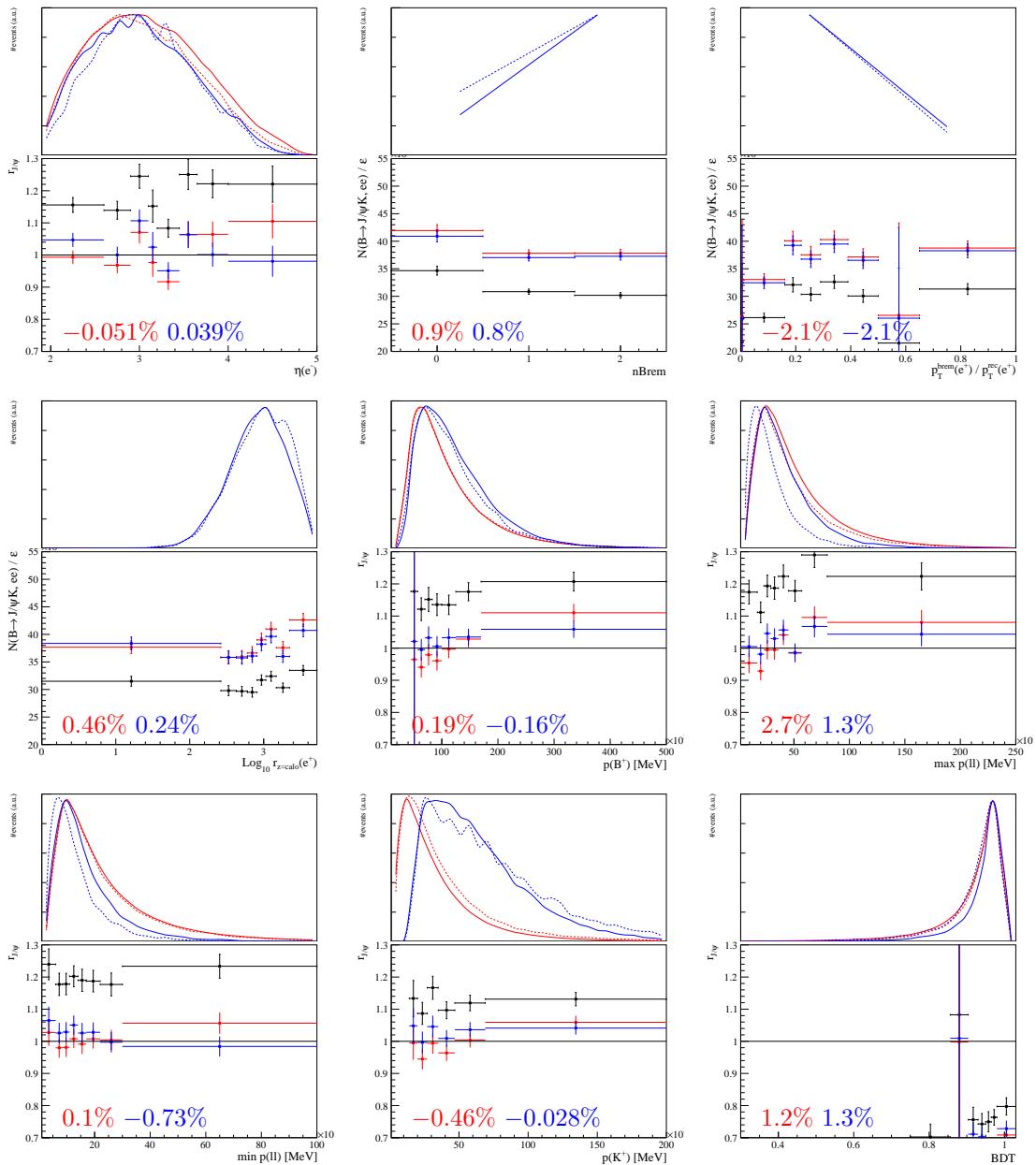


Figure 118: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $h\text{TOS!}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

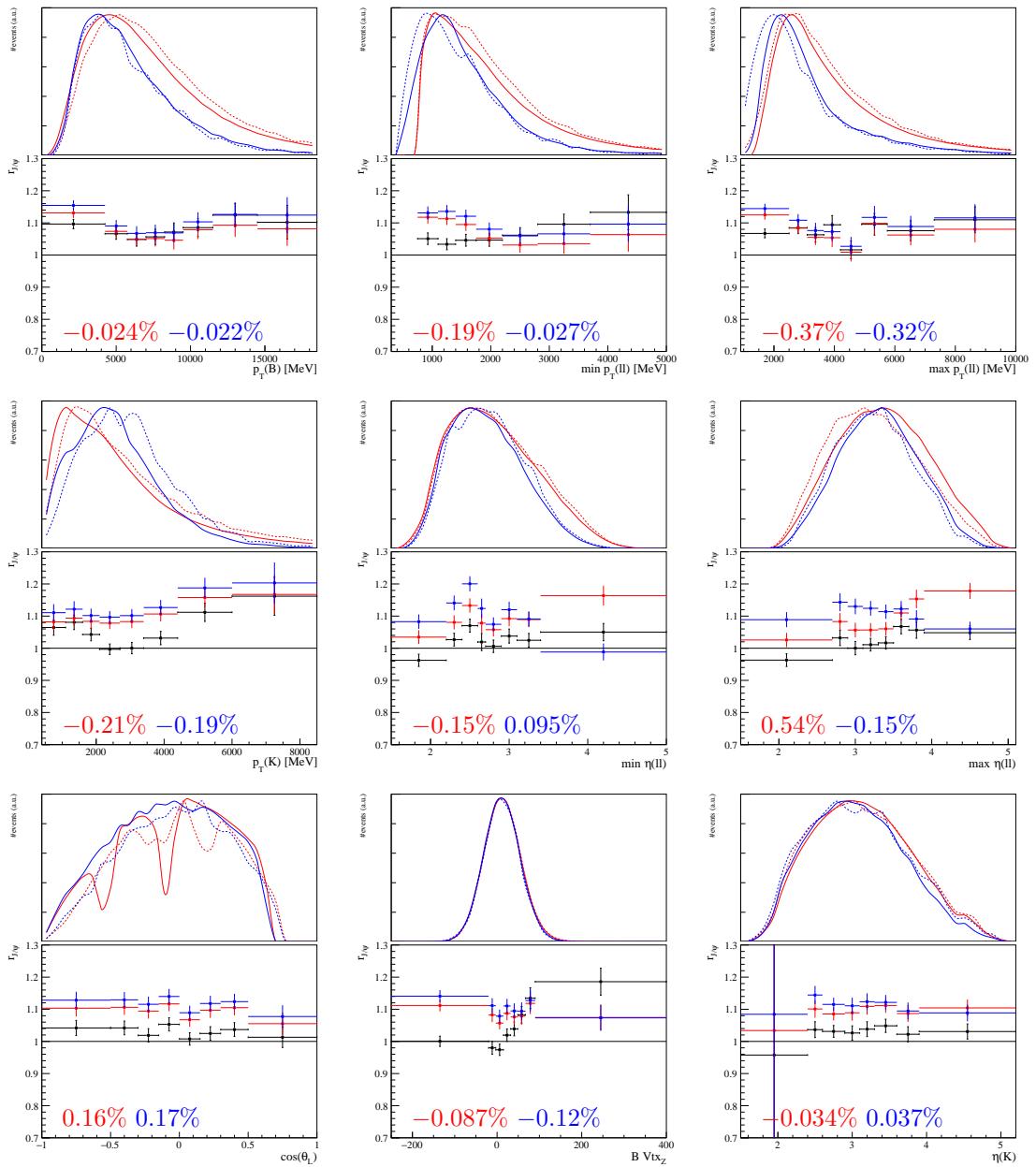


Figure 119: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS! category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

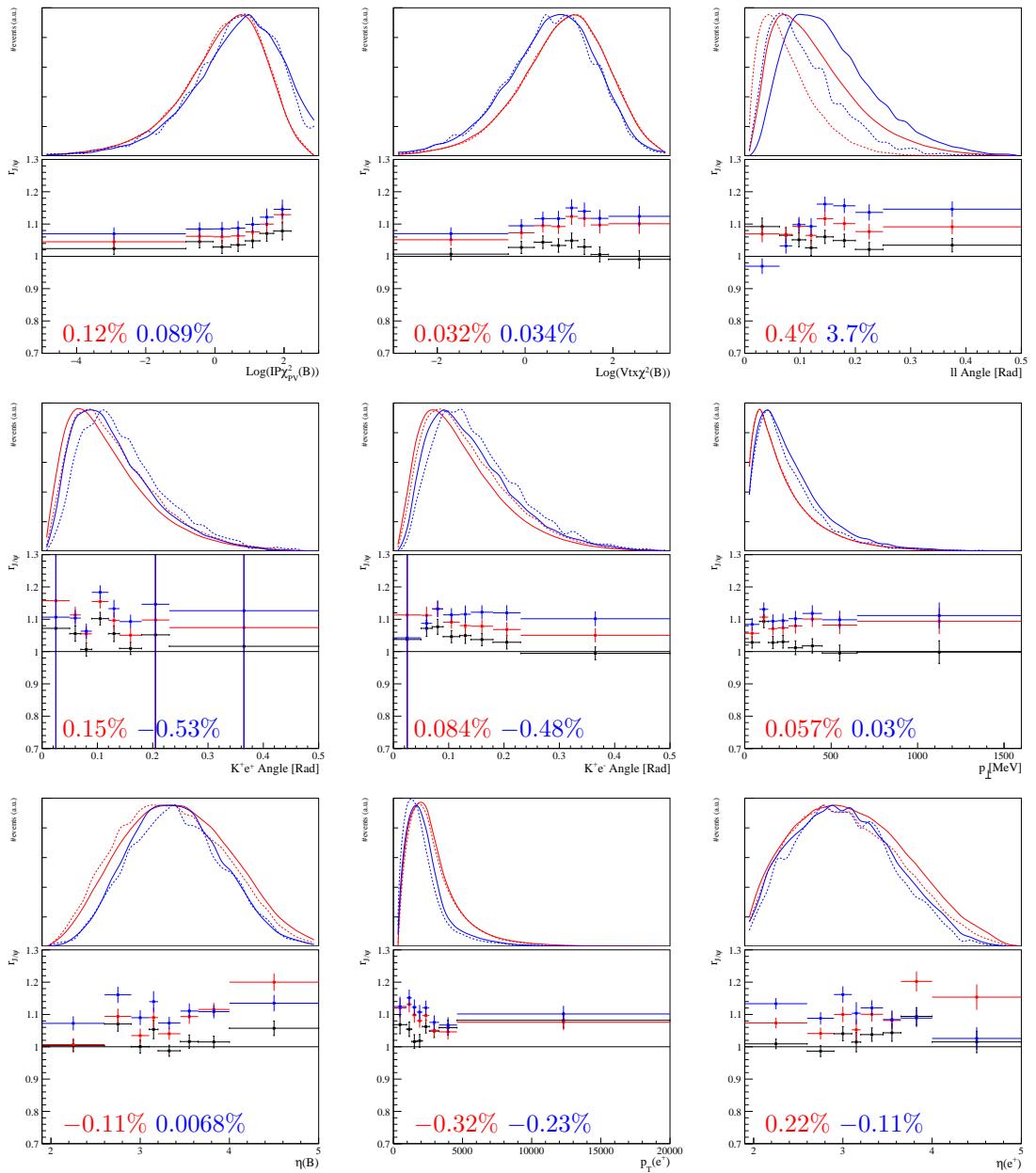


Figure 119: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode in the TIS! category and the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode in the μ TOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (solid blue) are shown.

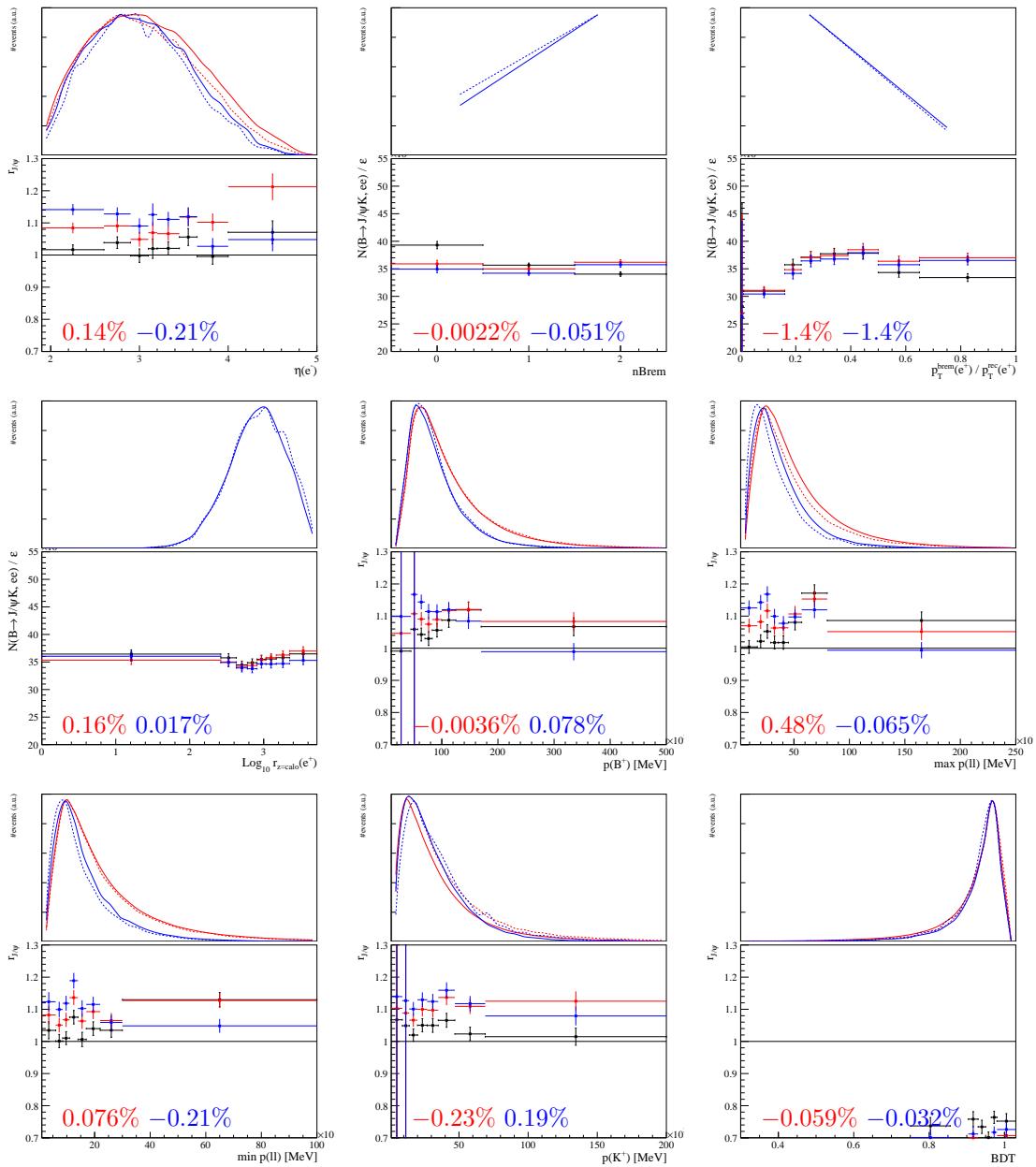


Figure 119: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS! category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (solid blue) are shown.

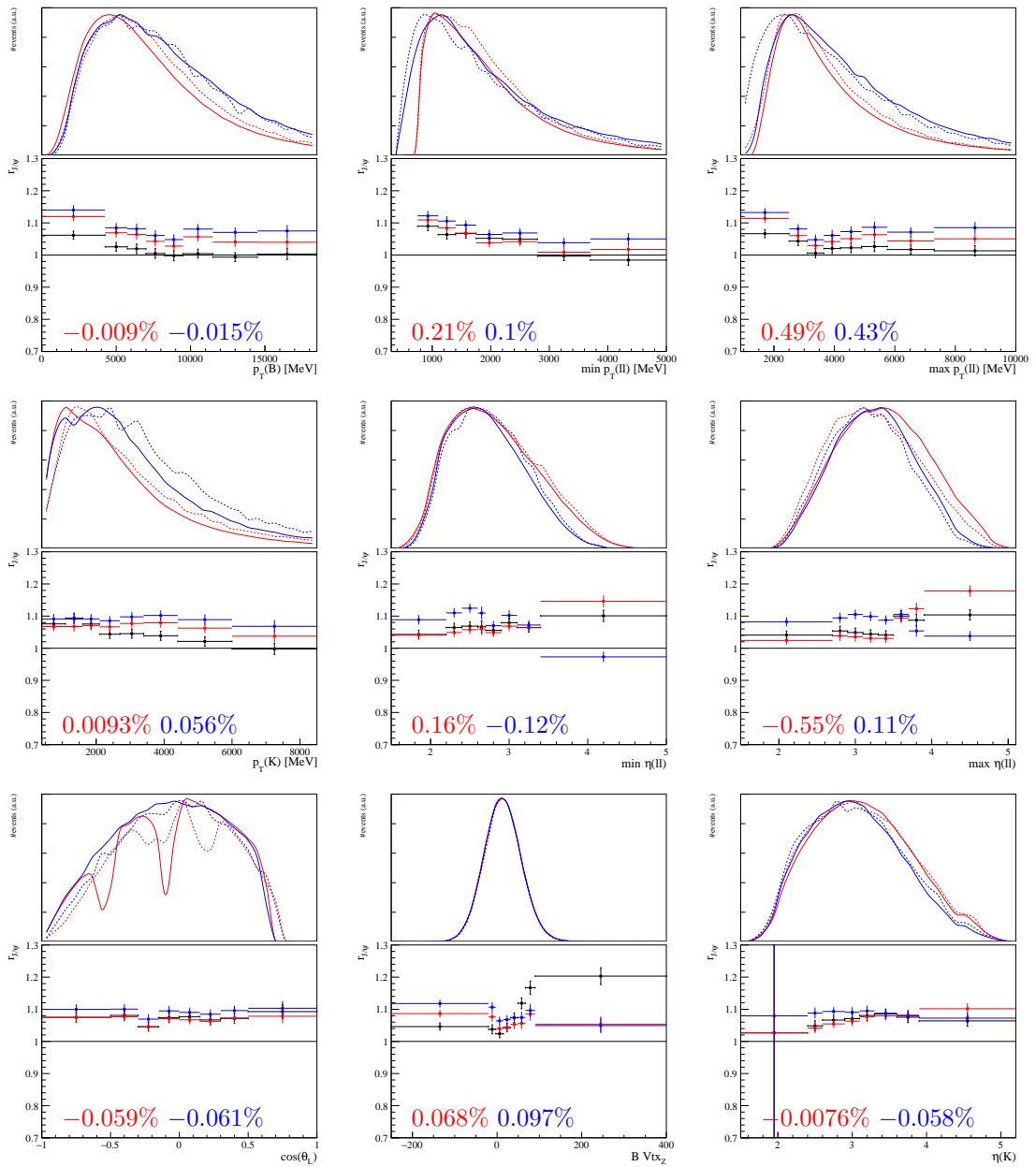


Figure 120: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

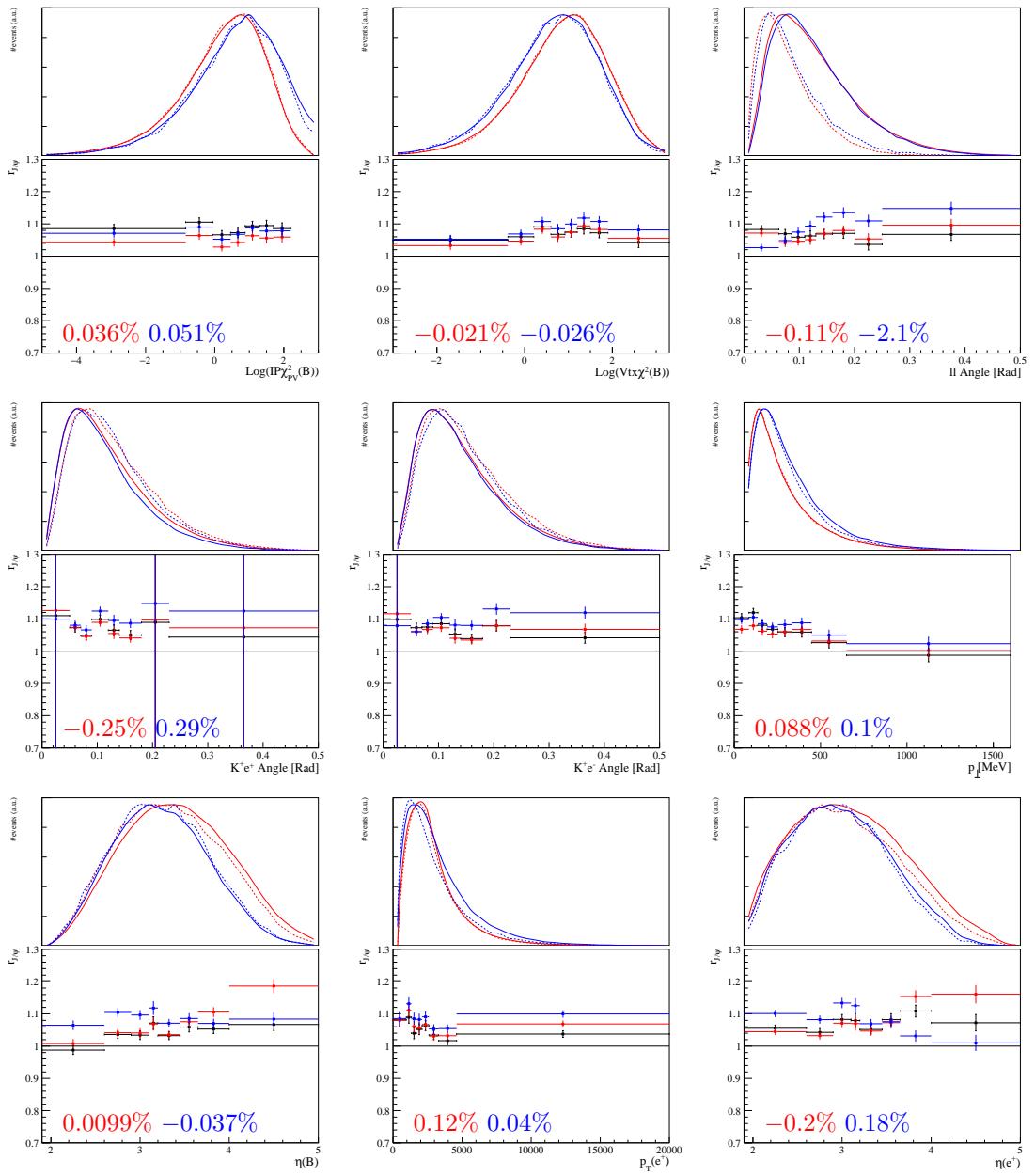


Figure 120: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+\mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+J/\psi(e^+e^-)$ (solid blue) are shown.

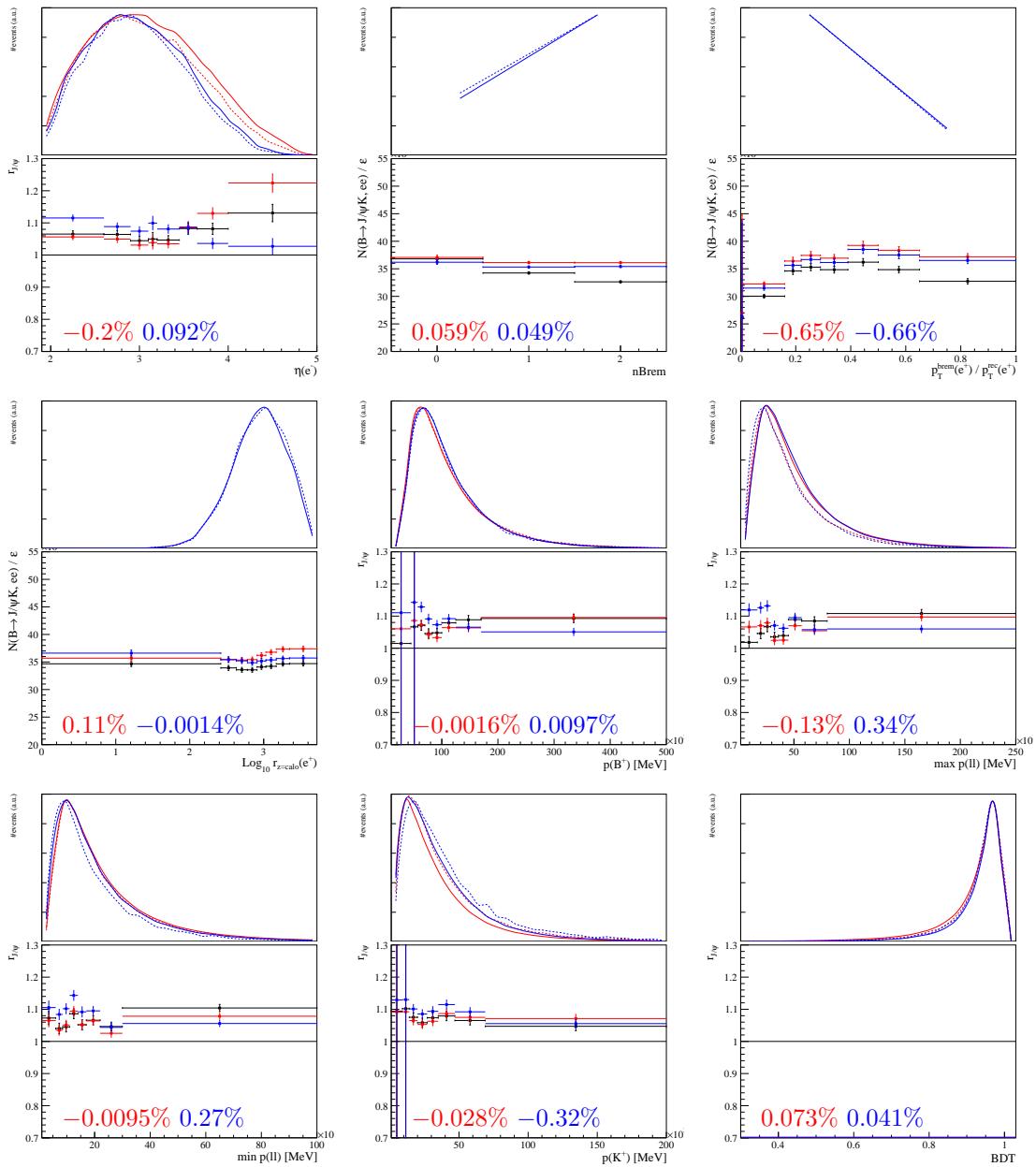


Figure 120: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

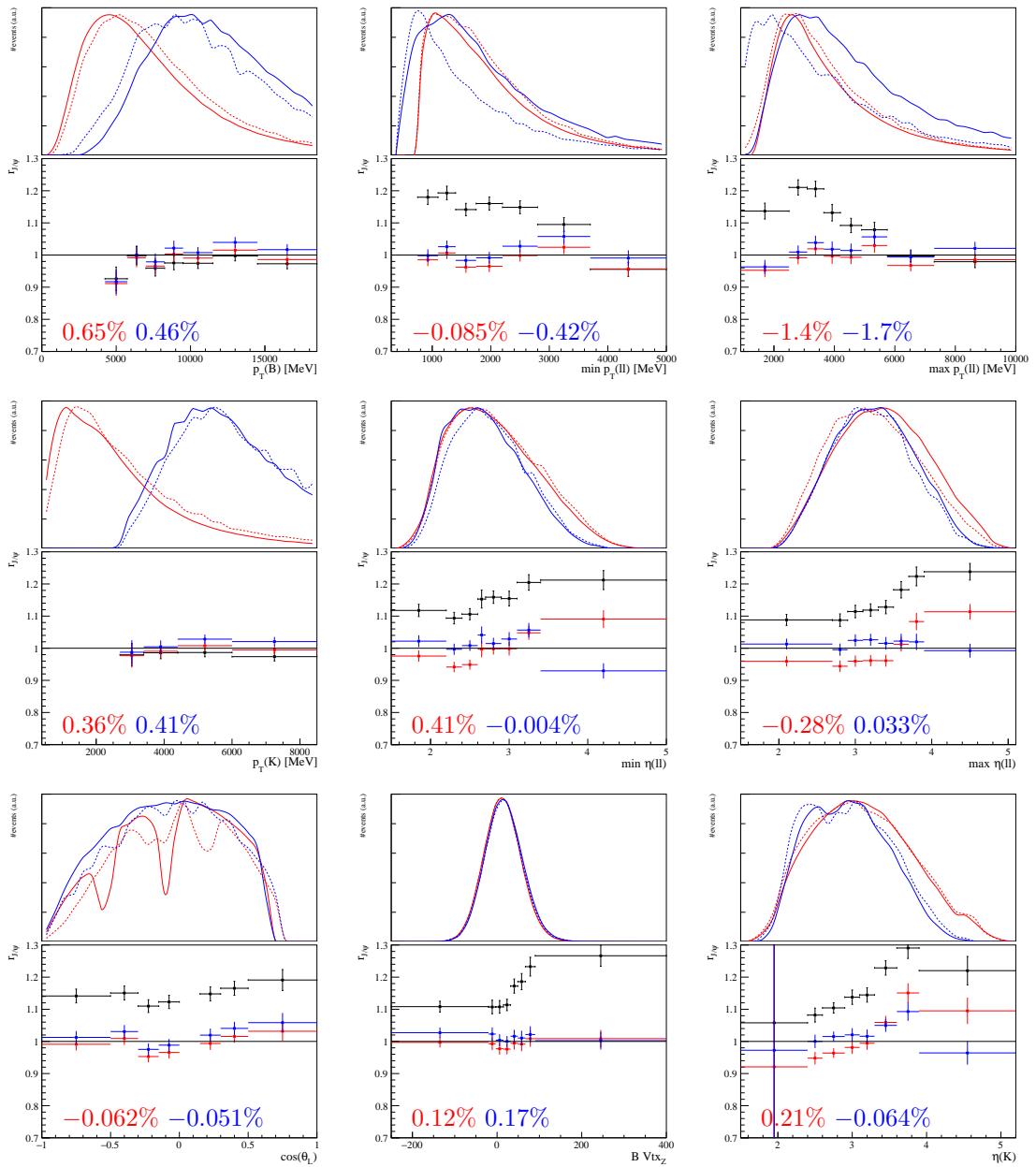


Figure 121: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $h\text{TOS}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

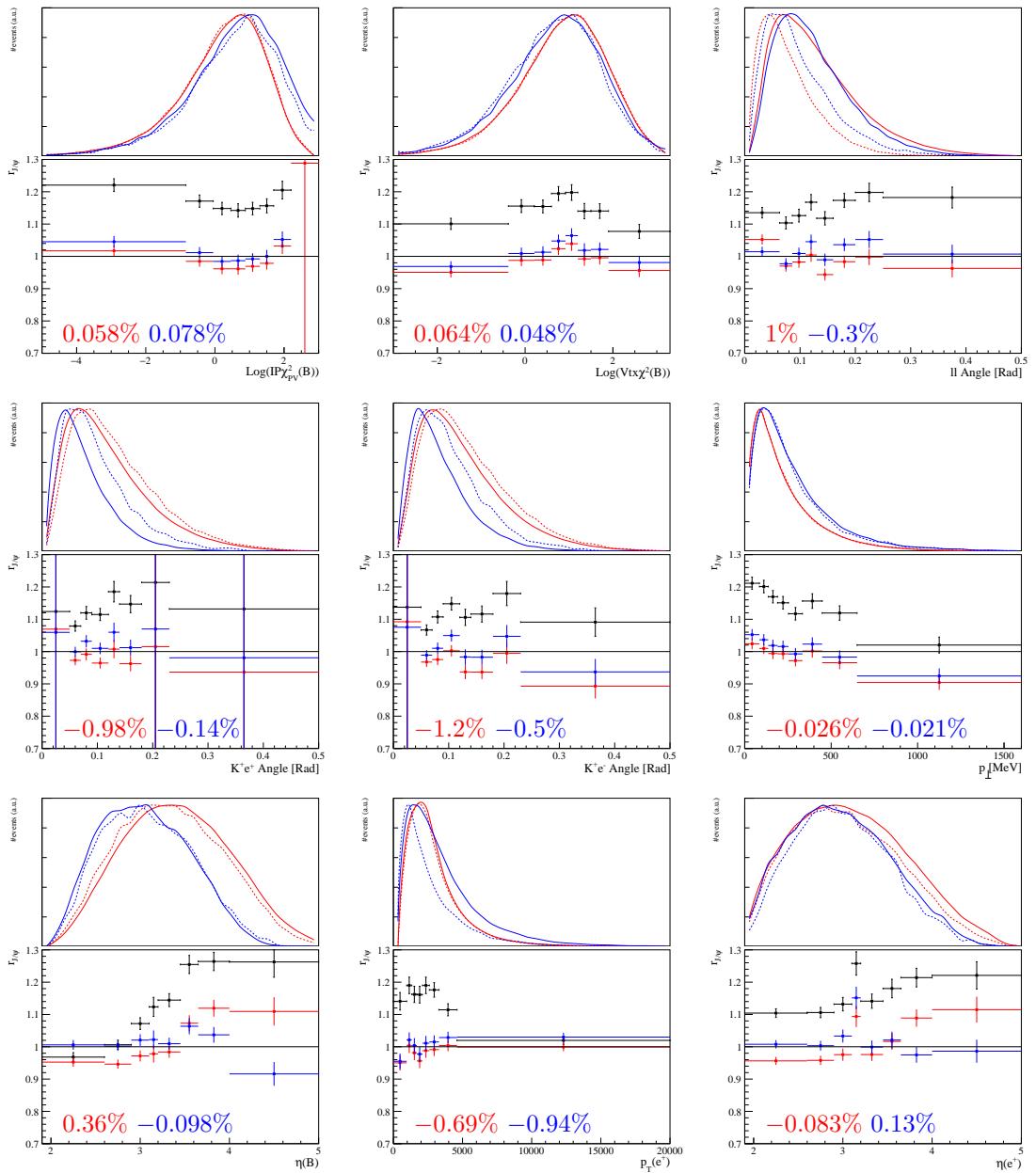


Figure 121: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode in the $h\text{TOS}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode in the μTOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (solid blue) are shown.

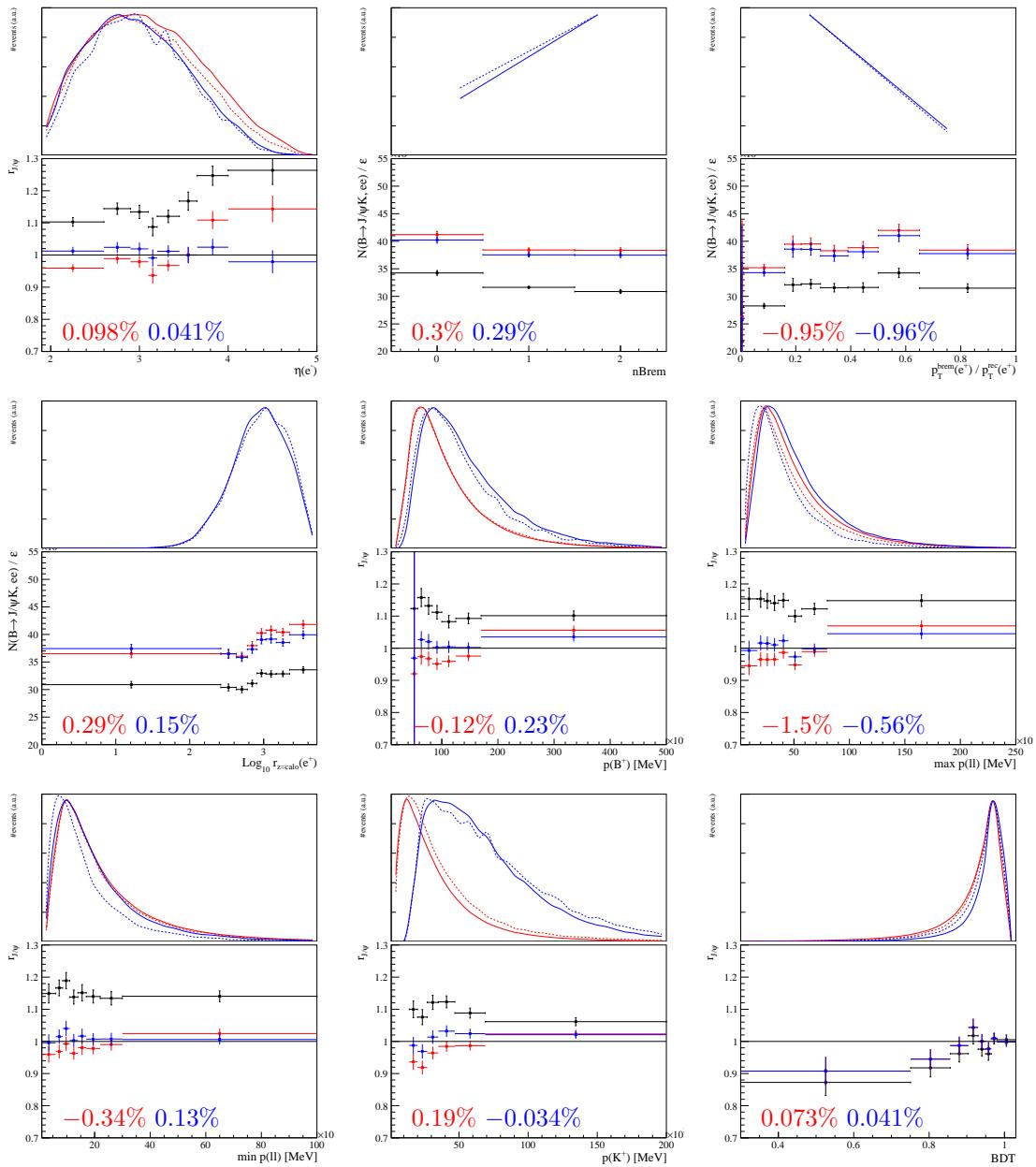


Figure 121: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $h\text{TOS}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 2 data taking condition. The L0 and PID corrections are applied to both mode. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the extra rapidity to the electron mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. For the variables related to bremsstrahlung defined only for the electron mode, the efficiency corrected yield $h_{J/\psi,e}^i$ is displayed instead of $r_{J/\psi}$. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

2399 11.3 $r_{J/\psi}^{\text{bin}}$ test

2400 While the invariant mass q^2 of the rare decay and the resonant control channels are clearly
 2401 different, the detector performance (and hence the selection efficiency) depend on lab
 2402 variables, such as the final state particles momenta and angles. As $q^2 = 2 \cdot p_{\ell^+} \cdot p_{\ell^-} \cdot (1 -$
 2403 $\cos \alpha_{\ell^+\ell^-})$, the efficiency can then be examined as a function of some but not all of the
 2404 variables p_{ℓ^+} , p_{ℓ^-} and $\alpha_{\ell^+\ell^-}$ in which there is a significant overlap between the rare and
 2405 resonant decays, even if the decays do not overlap in q^2 . In the previous section, $r_{J/\psi}$ was
 2406 examined as a function of a single variable (and with the parameter d_f , integrated over that
 2407 variable, taking into account the spectrum of the rare decay). In this section, an attempt
 2408 is made to look at $r_{J/\psi}$ in those regions of the parameter space where the control channel
 2409 has kinematics closer to that of the rare decay. Taking again the example of q^2 , while it is
 2410 not possible to look at $r_{J/\psi}$ in a bin of p_{ℓ^+} , p_{ℓ^-} and $\alpha_{\ell^+\ell^-}$, it is possible to bin in two of
 2411 these three variables. As detailed below, the decay kinematics can in fact be parameterised
 2412 as a function of four variables and $r_{J/\psi}$ can then be examined in 2D projections of this 4D
 2413 space. This test still has several shortcomings. In the 2D projection chosen the rare decay
 2414 and resonant decay will overlap only partially. The distribution in the two variables that
 2415 are not examined are integrated over and, in general, the distribution in these variables
 2416 will also differ in the rare decay and the resonant decay.

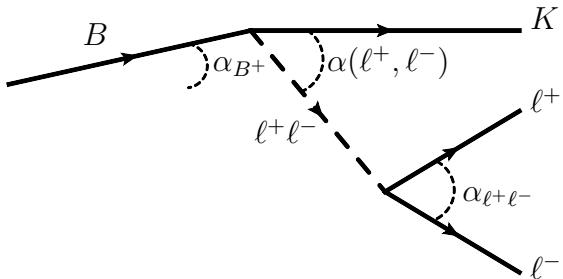


Figure 122: Illustration of the variables used to parametrise the $B^+ \rightarrow K^+ \ell^+ \ell^-$ decay

2417 To define the relevant regions, it must be first established how a $B^+ \rightarrow K^+ \ell^+ \ell^-$ decay
 2418 can be parametrised in terms of variables defined in the frame of the detector. This
 2419 is illustrated in Fig. 122. The B^+ momentum can be parametrised with three polar
 2420 coordinates: the magnitude of the momentum p_{B^+} , and a azimuthal and polar angle α_{B^+}
 2421 and ϕ_{B^+} , all defined in the frame of the detector. The B^+ then decays into a kaon and
 2422 a dilepton system. This is a two-body decay which can be fully parametrised with q^2 ,
 2423 the angle the kaon and the dilepton system $\alpha(\ell^+, \ell^-)$, and the angle polar angle of the
 2424 dilepton system $\phi_{\ell^+\ell^-}$. Finally, the dilepton system ‘decays’ into two leptons, and this
 2425 decay can be parametrised in terms of the angle between the two leptons $\alpha_{\ell^+\ell^-}$, and the
 2426 polar coordinate of one of the two leptons ϕ_ℓ . In summary, the $B^+ \rightarrow K^+ \ell^+ \ell^-$ decay can
 2427 be parametrised with 8 variables:

$$p_{B^+}, \alpha_{B^+}, \phi_{B^+}, \alpha(\ell^+, \ell^-), \phi_{\ell^+\ell^-}, q^2, \alpha_{\ell^+\ell^-}, \phi_\ell.$$

2428 Using the relation $q^2 = 2 \cdot p_{\ell+} \cdot p_{\ell-} \cdot (1 - \cos \alpha_{\ell+\ell-})$, q^2 and p_{B+} in the above list of variables
 2429 can be exchanged for $\max p_\ell$ and $\min p_\ell$. The list of degrees of freedom therefore becomes:

$$\alpha(\ell^+, \ell^-), \max p_\ell, \min p_\ell, \alpha_{\ell+\ell-}, (\phi_\ell, \phi_{\ell+\ell-}, \alpha_{B+}, \phi_{B+}).$$

2430 The last four terms have been put in parentheses because the detector is assumed to be
 2431 polar symmetric, and the distribution of α_{B+} is not related to the internal structure of the
 2432 decay. Hence, the decay $B^+ \rightarrow K^+ \ell^+ \ell^-$ can be considered to depend upon four variables.
 2433 The $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ decay has an extra constraint coming from $q^2 = m(J/\psi)^2$, which
 2434 reduces this set to three variables.

2435 To be more explicit, the $r_{J/\psi}^{\text{bin}}$ test proceeds as follows:

- 2436 1. Two variables are picked among $\alpha(\ell^+, \ell^-), \max p_\ell, \min p_\ell, \alpha_{\ell+\ell-}$;
- 2437 2. Using reconstructed $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events in the range $1.1 < q^2 <$
 2438 6.0 GeV^2 , two dimensional bins are chosen so as to span the region occupied by the
 2439 rare decay, as illustrated in Fig. 123 for $\min p_\ell, \alpha_{\ell+\ell-}$ (The binning choice for other
 2440 variable combinations can be found in Appendix L). As in the 1d $r_{J/\psi}$ test, the
 2441 bins are chosen such that they contain an approximately equal number of simulated
 2442 $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events;
- 2443 3. In each bin, the quantity $r_{J/\psi}^{\text{bin}}$ is defined as the following double ratio:

$$r_{J/\psi}^{\text{bin}} = \frac{\frac{N(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))|_{\in \text{bin}}}{\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))|_{\in \text{bin}}} \Big/ \frac{N(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))}{\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))},}{\frac{N(B^+ \rightarrow K^+ J/\psi(e^+ e^-))|_{\in \text{bin}}}{\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+ e^-))|_{\in \text{bin}}} \Big/ \frac{N(B^+ \rightarrow K^+ J/\psi(e^+ e^-))}{\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+ e^-))}},$$

2444 where the subscript $\in \text{bin}$ indicates that the event yield is measured in a given bin, and
 2445 that the efficiency corresponds to the efficiency of an event being reconstructed in that bin.
 2446 As above, a significant deviation from $r_{J/\psi}^{\text{bin}}$ from unity in any of the bins would indicate a
 2447 source of systematic uncertainty that may not cancel in the R_K double ratio. A flatness
 2448 parameter that quantifies the potential effect on R_K is again constructed based on the
 2449 deviations observed, taking into account the distribution of events in the rare decay.

2450 The $r_{J/\psi}^{\text{bin}}$ test is performed in two dimensions for $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ eTOS and $B^+ \rightarrow$
 2451 $K^+ J/\psi(\mu^+ \mu^-)$ μ TOS events. The results are shown in Fig. 124 for Run 1 and Fig. 125 for
 2452 Run 2. The efficiencies are computed in two ways. First, the nominal way of computing
 2453 the efficiencies is used. Second, the rapidity weights are added on top of the nominal
 2454 corrections to the electron mode.

2455 As for the differential $r_{J/\psi}$ test, all flatness parameters are less than a few percent and
 2456 the vast majority of them are at the permille level.

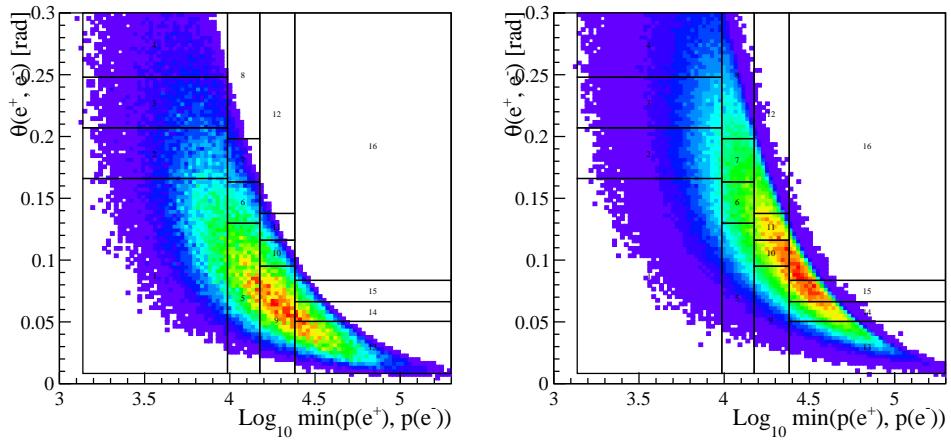


Figure 123: Two dimensional $\min p_\ell, \alpha_{\ell+\ell^-}$ distribution for reconstructed $B^+ \rightarrow K^+ e^+ e^-$ (left) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (right) simulated events. The bins used in the $r_{J/\psi}^{\text{bin}}$ test are shown on both figures.

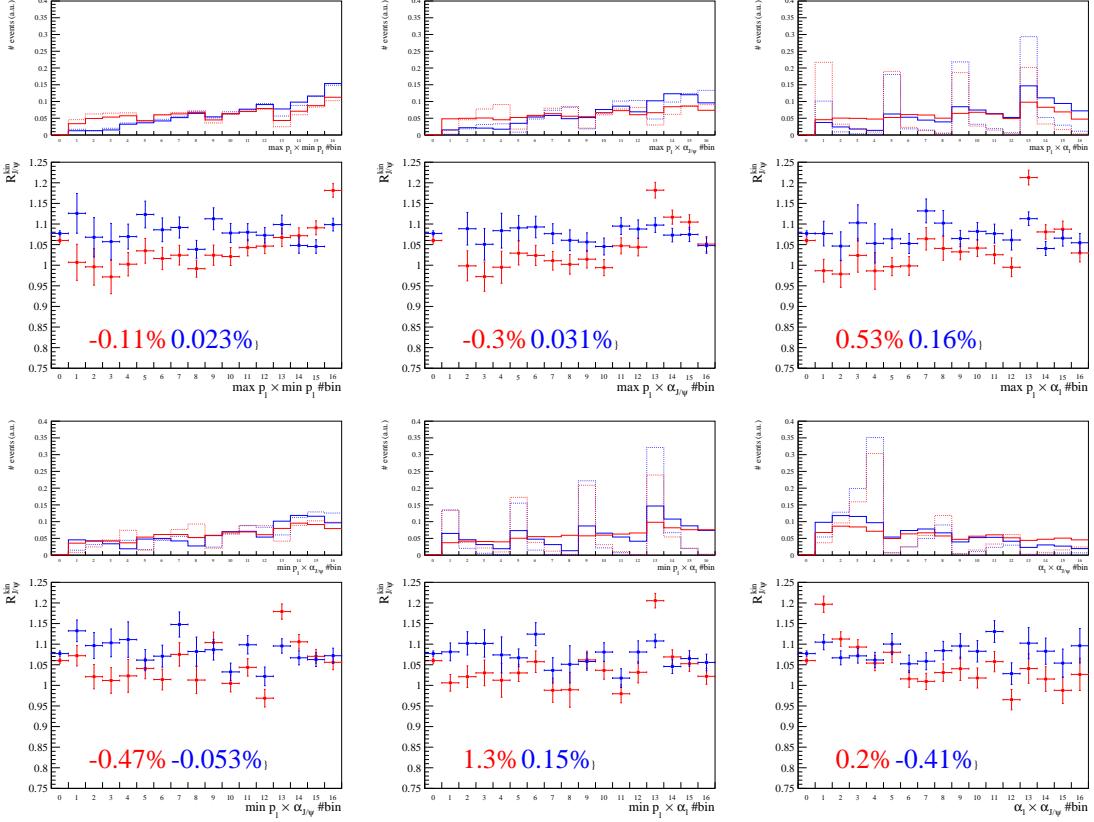


Figure 124: Result of the $r_{J/\psi}^{\text{bin}}$ test in two dimensional bins, where the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode is in the $e\text{TOS}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode is in the μTOS category, both in the Run 1 data set. The histograms in red are obtained by computing the efficiencies with the nominal trigger, PID, and kinematic corrections. For the blue ones, the η corrections are added. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. At the top of each plot, the bin distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

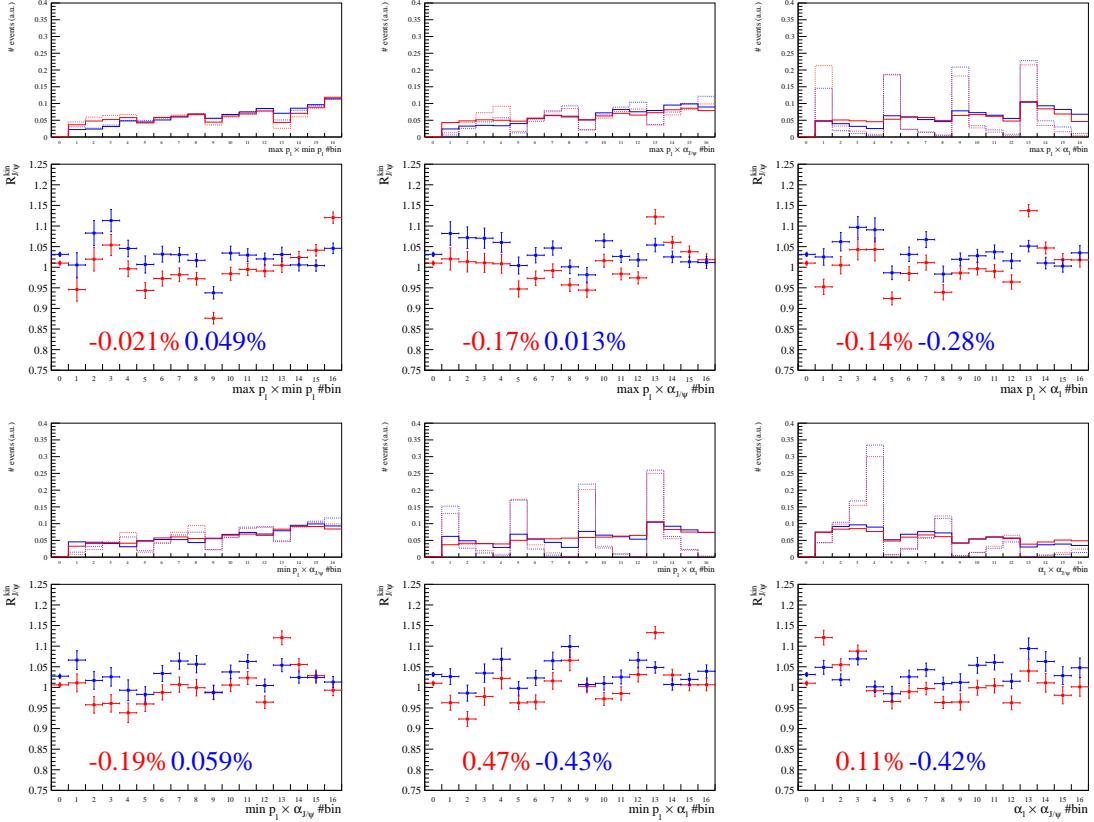


Figure 125: Result of the $r_{J/\psi}^{\text{bin}}$ test in two dimensional bins, where the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode is in the $e\text{TOS}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode is in the μTOS category, both in the Run 2 data set. The histograms in red are obtained by computing the efficiencies with the nominal trigger, PID, and kinematic corrections. For the blue ones, the η corrections are added. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. At the top of each plot, the bin distributions for $B^+ \rightarrow K^+\mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+e^+e^-$ (dotted blue) and $B^+ \rightarrow K^+J/\psi(e^+e^-)$ (solid blue) are shown.

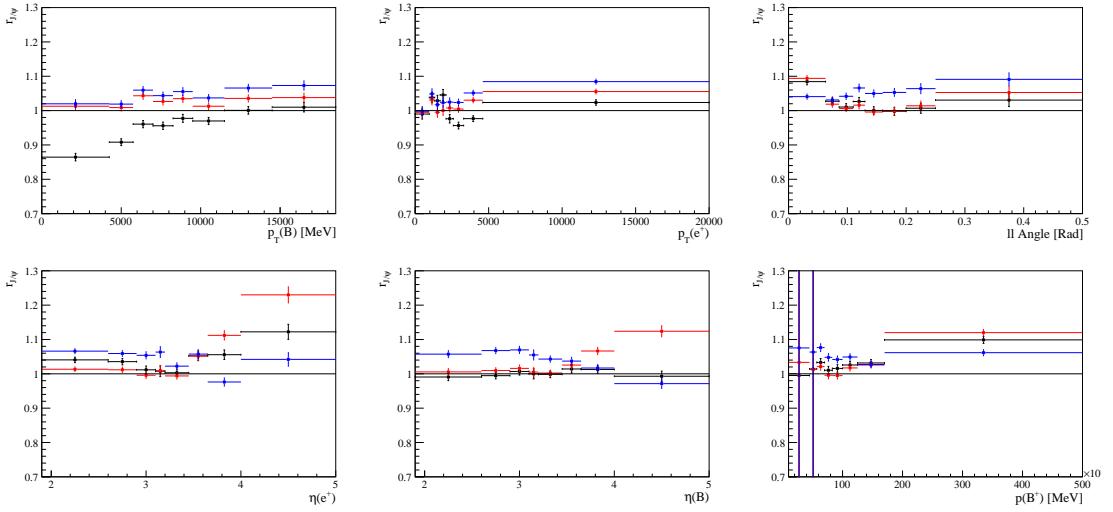


Figure 126: Plots of $r_{J/\psi}$ for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the e TOS category, in the Run 2 data taking conditions. The black points show the nominal corrections. The red points are computed applying the 2D trigger corrections and the blue points are computed applying the extra data-derived tracking corrections to the electron mode.

11.4 Tracking efficiency cross-check

An estimate of the systematic related to the tracking efficiency was detailed in Section 10.7 based on reweighting the $\eta(e^+)$ and $\eta(e^-)$ distributions in the simulation to match those observed in the data. While this analysis was in the final stages of review, a preliminary measurement of the tracking efficiency from the Run 2 data became available. The efficiency is determined using a tag-and-probe method, which measures the efficiency for finding a long e^\pm track when a VELO track has already been found [46]. As a cross-check that the existing systematic uncertainty covers for the uncertainty associated with the tracking efficiency, the effect of correcting the simulation using the weights derived from the tracking efficiency measurement was determined. The resulting $r_{J/\psi}$ differential plots are shown in Fig. 126. The application of the 2D e TOS correction used to compute the trigger systematic (see Section 10.4.2) already makes these $r_{J/\psi}$ differential plots much flatter, and adding the tracking corrections then fixes the residual discrepancies. The integrated $r_{J/\psi}$ values with and without the application of these corrections are shown in Tab. 51 and it can clearly be seen that the tracking corrections move $r_{J/\psi}$ closer to unity. Table 52 shows the effect of the tracking corrections on the R_K measurement in each of the trigger categories in the Run 2 data, together with the estimate of the tracking systematic derived in Section 10.7. The impact on R_K is small and is well covered by the existing tracking systematic uncertainties.

Table 51: The Run 2 integrated $r_{J/\psi}$ and $R_{\psi(2S)}$ values determined without and with the tracking efficiency corrections. The uncertainty shown is statistical only.

	Nominal		With track eff. corr.	
	$r_{J/\psi}$	$R_{\psi(2S)}$	$r_{J/\psi}$	$R_{\psi(2S)}$
Run 1				
e TOS	1.041 ± 0.006	0.999 ± 0.016	1.036 ± 0.006	0.985 ± 0.017
h TOS!	0.996 ± 0.012	0.904 ± 0.059	0.974 ± 0.012	0.921 ± 0.061
TIS!	0.998 ± 0.009	1.045 ± 0.032	0.983 ± 0.009	1.049 ± 0.034
Run 2				
e TOS	1.052 ± 0.005	0.986 ± 0.014	1.022 ± 0.005	0.989 ± 0.014
h TOS!	1.053 ± 0.011	0.889 ± 0.058	1.017 ± 0.011	0.895 ± 0.058
TIS!	1.112 ± 0.009	0.971 ± 0.028	1.074 ± 0.008	0.975 ± 0.028

Table 52: The impact of the tracking corrections on R_K together with the systematic uncertainty assigned to cover such corrections.

	e TOS	h TOS!	TIS!
Run 2			
Impact of tracking corrections	0.25%	0.03%	0.06%
Tracking systematic	0.47%	0.02%	0.35%

11.5 Pre-unblinding cross-checks

Rather than fitting a single R_K value simultaneously between the three different trigger categories and the Run 1 and Run 2 data, as for the nominal fit, several pre-unblinding cross-checks are used to check the consistency of the results across the different categories. The consistency between: all six different categories; the Run 1 and Run 2 data (amalgamating the three different trigger categories); and the three different trigger categories (amalgamating the Run 1 and Run 2 data) is checked using a variant of the fit which allows different R_K values to be fitted for these different categories. Using pseudoexperiments, the distribution of the difference in the log-likelihood between this variant fit and the nominal simultaneous fit is computed. Without examining the R_K values obtained, the fraction of experiments that have a delta log-likelihood greater than the likelihood difference actually observed is determined, in order to give a p-value for the compatibility of the R_K values obtained in the categories. The result of these tests are shown in Fig. 127 for all six categories (compatibility p-value 23%); in Fig. 128 for the Run 1 and Run 2 categories (p-value 6.4%); and in Fig. 129 for the three different trigger categories (p-value 82%). The different categories therefore give compatible values of R_K .

Without looking at the mass distributions themselves, the pull plots of the fits were obtained and are shown in Fig. 130. The fit to the data is observed to represent the data points well for both muons and electrons, in the Run 1 and Run 2 data and in all of the trigger categories.

Finally, the consistency with the previous published result [6] is checked by making the

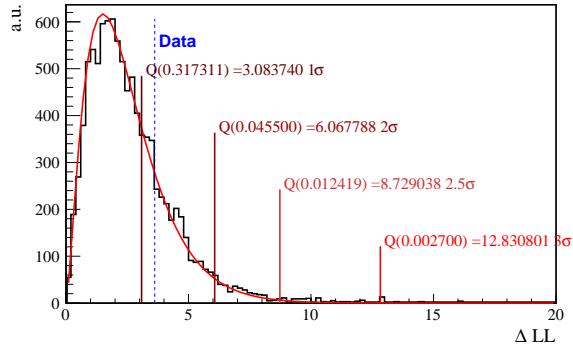


Figure 127: Difference in log-likelihood between a fit with six different R_K values for the three trigger categories and the Run 1 and Run 2 data, and the (nominal) simultaneous fit to all six categories with a single R_K value. The red curve represents a χ^2 distribution with 5 degrees of freedom. The distribution from pseudoexperiments is shown together with the difference in log-likelihood observed in the data, 23% of experiments have a ΔLL larger than that seen in the data.

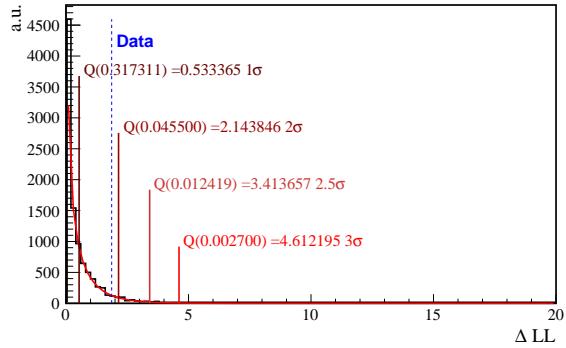


Figure 128: Difference in log-likelihood between a fit with two different R_K values for the Run 1 and Run 2 data, and the (nominal) simultaneous fit with a single R_K value. The red curve represents a χ^2 distribution with 1 degree of freedom. The distribution from pseudoexperiments is shown together with the difference in log-likelihood observed in the data, 6.4% of experiments have a ΔLL larger than that seen in the data.

2497 unsigned difference between the values, as described in Section 8.2. The results are found
 2498 to fulfil the predefined compatibility criteria $|\Delta R_K| < 0.153$, which takes into account that
 2499 80% of the data is in common. Therefore, the two measurements are considered to be
 2500 compatible at the $< 2.5\sigma$ level.

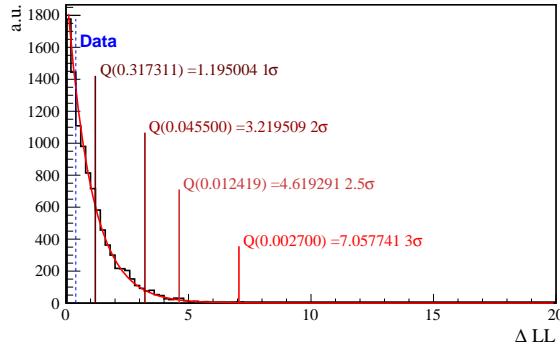


Figure 129: Difference in log-likelihood between a fit with three different R_K values for the e TOS, h TOS and TIS trigger categories, and the (nominal) simultaneous fit with a single R_K value. The red curve represents a χ^2 distribution with 2 degrees of freedom. The distribution from pseudoexperiments is shown together with the difference in log-likelihood observed in the data, 82% of experiments have a ΔLL larger than that seen in the data.

2501 11.6 Contamination from double misID backgrounds

2502 As explained in Section 5.1.3, contamination from fully reconstructed $B^+ \rightarrow K^+ \pi^+ \pi^-$
 2503 events, where both pions are mis-identified as electrons is expected to be reduced to a neg-
 2504 ligible level by the PID selection. To demonstrate that this is the case, the $B^+ \rightarrow K^+ e^+ e^-$
 2505 sample selected in data can be reconstructed ignoring the bremsstrahlung correction and
 2506 changing the mass hypothesis of each electron to that of the pion. Any contribution from
 2507 $B^+ \rightarrow K^+ \pi^+ \pi^-$ decays would appear as a narrow peak in the $m(K^+ e_{\pi^+}^+ e_{\pi^-}^-)$ mass
 2508 distribution. As shown in Fig. 131, there is no trace of $B^+ \rightarrow K^+ \pi^+ \pi^-$ misidentified
 2509 events in the final $B^+ \rightarrow K^+ e^+ e^-$ data sample.

2510 The second kind of double mis-ID background is referred to as $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$
 2511 “swaps”, where a $B^+ \rightarrow K_{\ell^+}^+ J/\psi(\ell_{K^+}^+ \ell^-)$ transition where the K^+ is mis-identified as
 2512 the ℓ^+ and vice-versa. As in the previous case, this background is expected to be negligible
 2513 after the PID selection. To demonstrate this, $B^+ \rightarrow K^+ e^+ e^-$ candidates in data are
 2514 reconstructed under the exchanged mass hypothesis $K \leftrightarrow e$, and their invariant mass
 2515 $m^{J/\psi}(K_{e^+}^+ e_{\pi^+}^+ e_{\pi^-}^-)$ is computed, where the mass $m(Ke)$ is constrained to that of the
 2516 J/ψ . Any contribution from $B^+ \rightarrow K_{\ell^+}^+ J/\psi(\ell_{K^+}^+ \ell^-)$ decays would appear as a sharp
 2517 peak at the B^+ mass. As can be seen in Fig. 132, no such contribution is found in the
 2518 final $B^+ \rightarrow K^+ e^+ e^-$ data sample.

2519 11.7 Measurement of \mathcal{B} ($B^+ \rightarrow K^+ \mu^+ \mu^-$)

2520 In order to establish if a comparable yield of $B^+ \rightarrow K^+ \mu^+ \mu^-$ candidates is observed as
 2521 in the previous Run 1 analysis [7], the $1.1 < q^2 < 6.0 \text{ GeV}/c^2$ requirement is made on
 2522 the variable $q^2 = (p(\mu^+) + p(\mu^-))^2$ (cf $q^2 = (p(B^+) - p(K^+))^2$ used for the remainder of
 2523 the present R_K analysis). The fit to the Run 1 $B^+ \rightarrow K^+ \mu^+ \mu^-$ candidates is shown in

2524 Fig. 133. A total of 1026 $B^+ \rightarrow K^+ \mu^+ \mu^-$ candidates are observed, compared to ~ 1600 in
2525 Ref. [7]. Taking the $B^+ \rightarrow K^+ J/\psi$ branching fraction to be $(0.998 \pm 0.014 \pm 0.040) \times 10^{-3}$,
2526 as in Ref. [7], the $B^+ \rightarrow K^+ \mu^+ \mu^-$ branching fraction from the Run 1 data is measured to
2527 be

$$\mathcal{B}^{\text{new}}(B^+ \rightarrow K^+ \mu^+ \mu^-) = (24.5 \pm 0.9 \text{ (stat.)}) \times 10^{-9},$$

2528 in excellent agreement with the previous Run 1 result,

$$\mathcal{B}^{\text{old}}(B^+ \rightarrow K^+ \mu^+ \mu^-) = (24.2 \pm 0.7 \text{ (stat.)} \pm 1.2 \text{ (syst.)}) \times 10^{-9}.$$

2529 The ratio between these branching fraction measurements, assuming that there are only
2530 statistical fluctuations associated with the ~ 500 candidate difference in the yields, is
2531 $\mathcal{B}^{\text{new}}(B^+ \rightarrow K^+ \mu^+ \mu^-)/\mathcal{B}^{\text{old}}(B^+ \rightarrow K^+ \mu^+ \mu^-) = 1.012 \pm 0.021$. Given that some candidates
2532 will not be in common, the uncertainty shown is a lower limit on the real statistical
2533 uncertainty. The use of the standard definition of q^2 used elsewhere in the present analysis
2534 would lower the branching fraction observed by $\sim 3\%$. The differential branching fraction
2535 measurements are shown in Fig. 134 (statistical uncertainties only) and are also in excellent
2536 agreement with the previous analysis. The individual mass fits are shown in Appendix J.

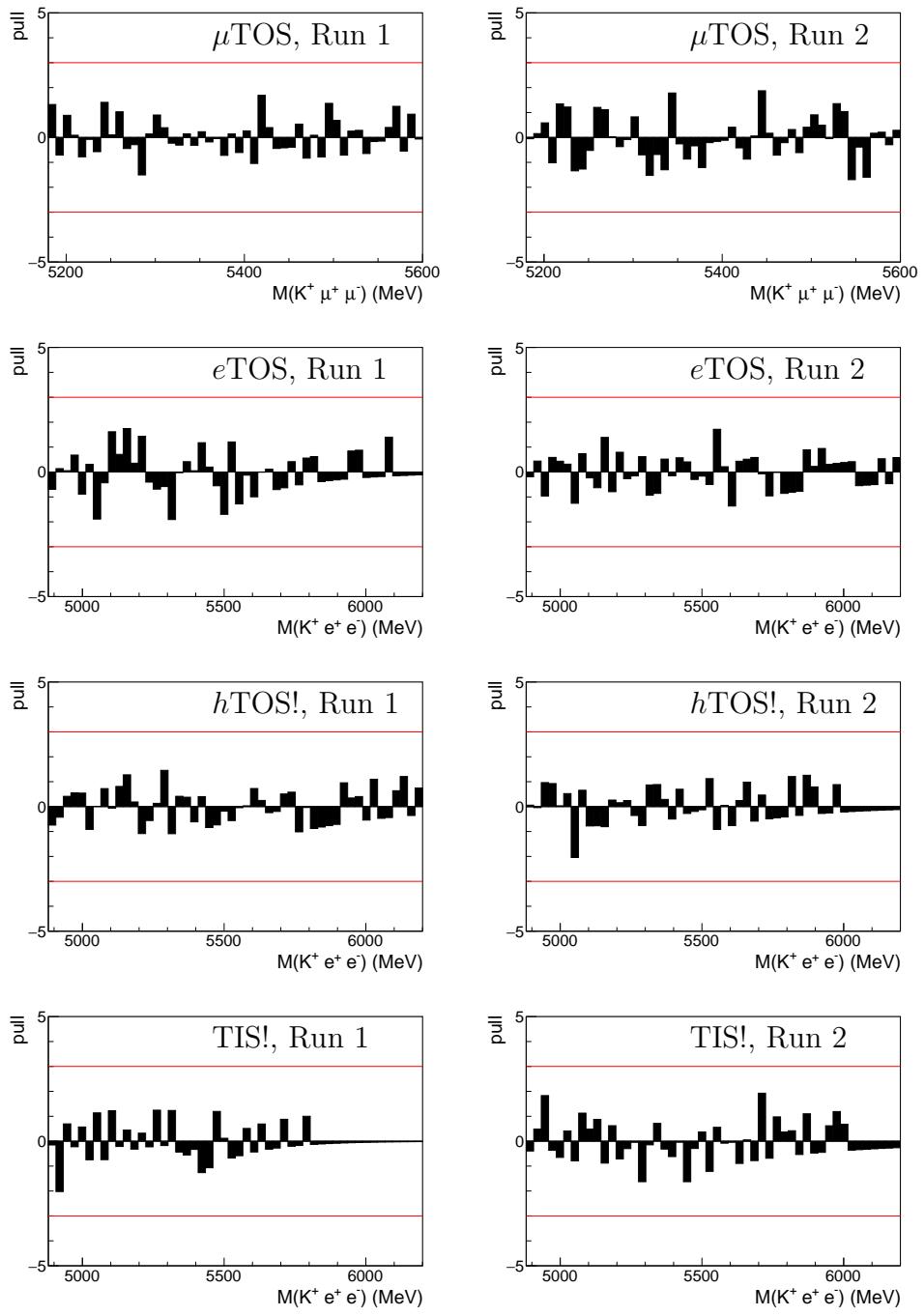


Figure 130: Pull plots from the nominal fit to the $m_{K\ell^+\ell^-}$ distribution for the rare decay modes. The top two plots show the muon modes (with the μ TOS trigger) from (left) the Run 1 data and (right) the Run 2 data. Below these are the electron modes plots in the e TOS, h TOS!, and TIS! trigger categories. These are again shown for (left) the Run 1 data and (right) the Run 2 data. Points with not negative error band, correspond to empty bins in the data.

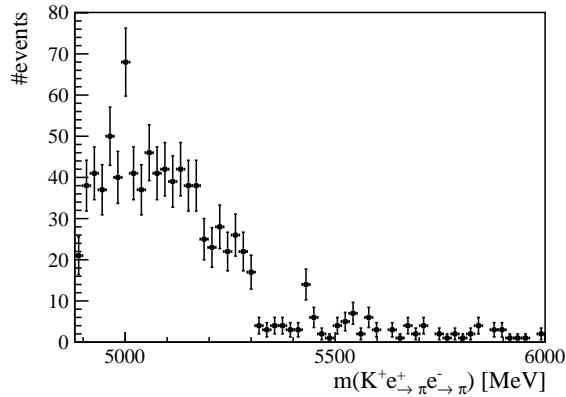


Figure 131: Distribution of the $m(K^+e^+ \rightarrow \pi^+ e^- \rightarrow \pi^-)$ for $B^+ \rightarrow K^+e^+e^-$ selected candidates in the combined Run 1 and Run 2 sample.

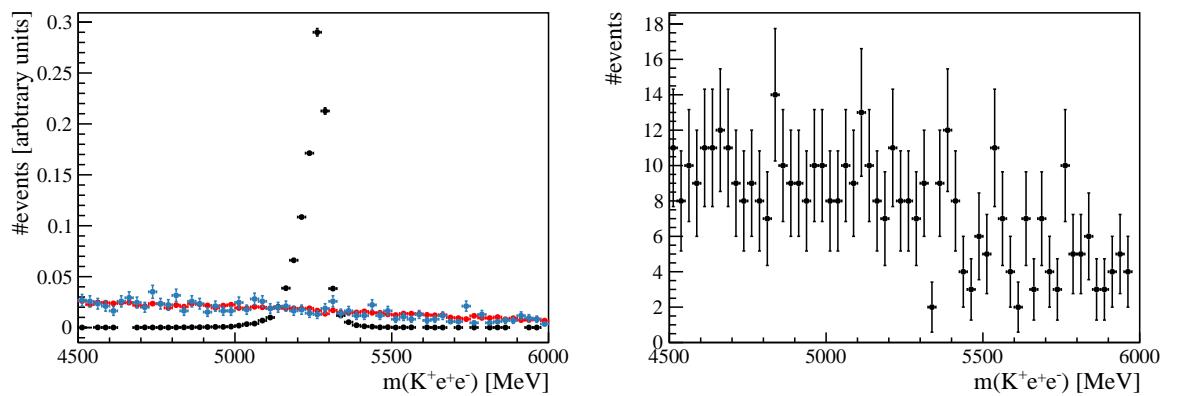


Figure 132: Left: Distribution of the $m^{J/\psi}(K^+_{\rightarrow e^+} e^+_{\rightarrow K^+} e^-_{\rightarrow \pi^-})$ invariant mass for $B^+ \rightarrow K^+_{\rightarrow \ell^+} J/\psi(\ell^+_{\rightarrow K^+} \ell^-)$ (black), $B^+ \rightarrow K^+e^+e^-$ and $B^0 \rightarrow K^*0e^+e^-$ simulated events. Right: Same distribution for $B^+ \rightarrow K^+e^+e^-$ selected candidates in the combined Run 1 and Run 2 sample.

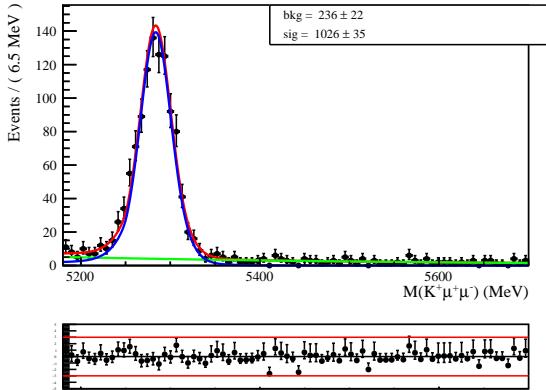


Figure 133: Fit to the $m(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+\mu^+\mu^-$ selected candidates in the Run 1 data sample. The red solid line shows the fit model, the blue line is the signal component and the green line shows the combinatorial background.

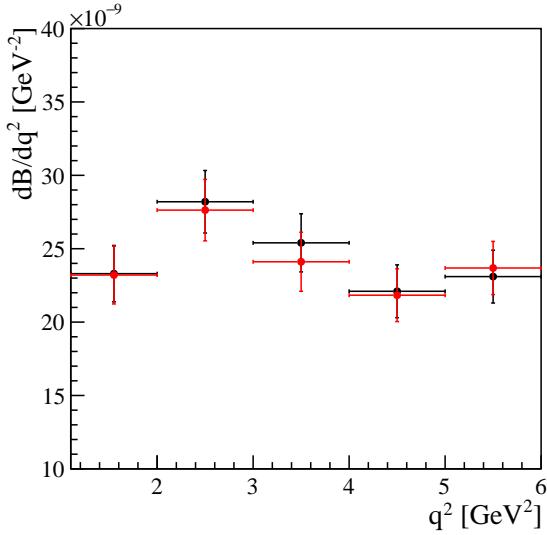


Figure 134: New measurement of $\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-)$ from the Run 1 data sample (red) compared with that from Ref. [7] (black). Only the statistical uncertainties are shown, ignoring any overlap between the two samples. The variable q^2 is computed from the μ^\pm four-momenta, rather than from $(p(B^+) - p(K^+))^2$.

2537

12 Results

2538

12.1 Measurement of R_K

2539 The results obtained in Section 9.2 need to be corrected by the fit bias. In order to
2540 establish the size of this bias in the final configuration of the fit, the toy study described
2541 in Section 8.1 is repeated using as input the measured yields for the partially reconstructed
2542 and combinatorial backgrounds measured in the data. The final bias is estimated to be
2543 $(3.5 \pm 1.0)\% \times \sigma_{\text{stat}}$. The sum in quadrature of half of the size of the bias and its own
2544 uncertainty is considered as an additional systematic uncertainty on R_K . Finally, the
2545 systematic associated with the fit model computed in Section 10.9, is also combined with
2546 the fit uncertainty by convoluting the profiled likelihood with a Gaussian of suitable width.
2547 Figure 135 shows the original likelihood profile obtained from the fit to the data, and the
2548 effect of the bias shift and the additional systematic.

2549 The final result for R_K is,

$$R_K = 0.846^{+0.062}_{-0.056} (\text{stat + syst}),$$

2550 where the uncertainties combine both statistical and systematic effects. The R_K PDF
2551 obtained from the combined log-likelihood is used to establish the level of compatibility
2552 of the result with the SM prediction. The p-value of the SM hypothesis is computed
2553 by integrating the R_K PDF above the SM value, $R_K = 1.0000 \pm 0.0001$ [3], and then
2554 converted into a Gaussian number of sigmas. The measurement of R_K is compatible with
2555 the SM at 2.5σ . Taking into account a 1% variation of the SM prediction due to possible
2556 imperfections of PHOTOS [4], the compatibility lies in the range $2.4 - 2.6\sigma$.

2557

12.2 Measurement of $\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)$

2558 Combining the measurement of R_K with the previously published value for $\mathcal{B}(B^+ \rightarrow$
2559 $K^+ \mu^+ \mu^-)$ [7] in the range $q^2 \in [1.1, 6] \text{ GeV}^2$,

$$\mathcal{B}^{\text{old}}(B^+ \rightarrow K^+ \mu^+ \mu^-) = (24.2 \pm 0.7 \text{ (stat.)} \pm 1.2 \text{ (syst.)}) \times 10^{-9}$$

2560 a measurement is made for the branching fraction $\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)$ in the same q^2 bin.
2561 Correlations between R_K and $\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)$ need to be taking into account when
2562 performing this combination, in particular that induced by the statistical uncertainty in
2563 the muon yield, since the $\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)$ measurement was performed in a sample
2564 with significant overlap with the Run 1 $B^+ \rightarrow K^+ \mu^+ \mu^-$ sample used for the present
2565 determination of R_K , as established in Section 11.7. The correlation between the statistical
2566 uncertainty of $\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)$ and R_K is estimated to be 29%, see Appendix P for more
2567 details. The systematic uncertainty of $\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)$ is dominated by the uncertainty
2568 in the branching fraction of the normalisation channel (4.2%), $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$, which
2569 is not correlated to R_K . The remaining systematics, associated with imperfections of the
2570 simulation and the physics model affecting the determination of the ratio of efficiencies,
2571 are considered to be fully correlated with the equivalent systematics in R_K .

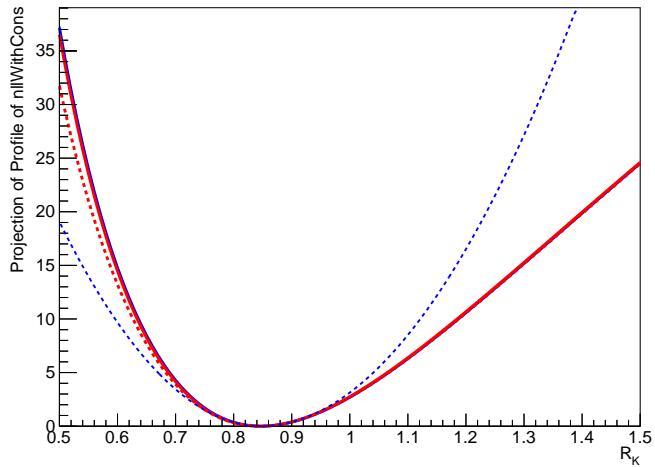


Figure 135: Profiled likelihood obtained from the fit to the data (solid blue), after the correction of the fit bias (solid red) and with the additional systematic uncertainties in R_K (dashed red). The blue dashed line depicts the expected shape of the likelihood profile if the uncertainties were gaussian.

2572 The obtained value for \mathcal{B} ($B^+ \rightarrow K^+ e^+ e^-$) in the q^2 range [1.1,6] GeV 2 is

$$\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-) = (28.6_{-1.7}^{+2.0}(\text{stat}) \pm 1.4(\text{syst})) \times 10^{-9}.$$

2573 **Appendices**

2574 **A Simulation samples**

2575 **A.1 Run 1 samples**

2576 For each simulation sample used in this analysis, the tables below report:

- 2577 • The simulation condition: year, polarity, Pythia and simulation version).
- 2578 • The “dddb” tag used to process the simulation.
- 2579 • The size of the sample at generator level (“stat” column in the tables).
- 2580 • The acceptance efficiency (ε^{geom} in the tables).
- 2581 • The data condition weight that is used to reweight each sample to the right luminosity
2582 (“dcw” column in the tables).

2583 When re-running the calo on simulation, the following `conddb` is used:

- 2584 • `sim-20141210-1-vc-md100` (`sim-20141210-1-vc-mu100`) for 2012
- 2585 • `sim-20141210-vc-md100` (`sim-20141210-vc-mu100`) for 2011

2586 **$B \rightarrow J/\psi K, ee$ 12153001**

cond	DDDB	stat	ε^{geom}	dcw
12 MD P6/sim08a	Sim08-20130503-1	255750	0.1575 ± 0.000582	0.931604
12 MD P8/sim08a	Sim08-20130503-1	272249	0.16669 ± 0.00033	0.931604
12 MD P8/sim08h	dddb-20130929-1	1009906	0.16669 ± 0.00033	0.931604
2587 12 MU P6/sim08a	Sim08-20130503-1	256000	0.1623 ± 0.000598	0.917695
12 MU P8/sim08a	Sim08-20130503-1	269999	0.1684 ± 0.000617	0.917695
12 MU P8/sim08h	dddb-20130929-1	1046869	0.1684 ± 0.000617	0.917695
11 MD	dddb-20130929	500914	0.16308 ± 0.00031	1.41327
11 MU	dddb-20130929	501774	0.16308 ± 0.00031	1.05424

²⁵⁸⁸ **$B \rightarrow Kee$ MS, 12123002**

cond	DDDB	stat	ε^{geom}	dcw
12 MD P6/sim08a	Sim08-20130503-1	253748	0.1593 ± 0.000419	1.03853
12 MD P6/sim08b	Sim08-20130503-1	257499	0.1593 ± 0.000419	1.03853
12 MD P8/sim08a	Sim08-20130503-1	270499	0.1700 ± 0.000623	1.03853
12 MD P8/sim08b	Sim08-20130503-1	254000	0.1700 ± 0.000623	1.03853
2589 12 MU P6/sim08a	Sim08-20130503-1	271749	0.1599 ± 0.000416	1.04046
12 MU P6/sim08b	Sim08-20130503-1	257998	0.1599 ± 0.000416	1.04046
12 MU P8/sim08a	Sim08-20130503-1	256250	0.1718 ± 0.000629	1.04046
12 MU P8/sim08b	Sim08-20130503-1	255497	0.1718 ± 0.000629	1.04046
11 MD P8/sim09a	MC11-20111102	502252	0.1561 ± 0.000639	1.05817
11 MU P8/sim09a	MC11-20111102	508705	0.1559 ± 0.000616	0.78068
	total	3088197		

²⁵⁹⁰ **$B \rightarrow Kee$ NO PHOTOS, 12123005**

cond	DDDB	stat	ε^{geom}	dcw
2591 12 MD	dddb-20130929-1	515546	0.16858 ± 0.00029	—
12 MU	dddb-20130929-1	554032	0.16858 ± 0.00029	—

²⁵⁹² **$B_u \rightarrow \psi(2S)K^+, ee$, 12153011**

cond	DDDB	stat	ε^{geom}	dcw
12 MD	dddb-20130929-1	511059	0.16885 ± 0.00041	1.03933
2593 12 MU	dddb-20130929-1	503111	0.16885 ± 0.00041	1.06363
11 MD	dddb-20160318-1	254619	0.16202 ± 0.00054	1.03077
11 MU	dddb-20160318-1	254619	0.16202 ± 0.00054	0.76544

²⁵⁹⁴ **$Bd \rightarrow K^*ee$, 11124001**

cond	DDDB	stat	ε^{geom}	dcw
12 MD P8/sim08a	Sim08-20130503-1	1255498	0.16710 ± 0.00080	0.890218
12 MD P8/sim08b	Sim08-20130503-1	1260991	0.16710 ± 0.00080	0.890218
12 MD P8/sim08g	dddb-20130929-1	3012770	0.16710 ± 0.00080	0.890218
2595 12 MU P8/sim08a	Sim08-20130503-1	1272496	0.16610 ± 0.00082	0.894021
12 MU P8/sim08b	Sim08-20130503-1	1265996	0.16610 ± 0.00082	0.894021
12 MU P8/sim08g	dddb-20130929-1	3008343	0.16610 ± 0.00082	0.894021
11 MD P8/sim08g	dddb-20130929	1505207	0.16710 ± 0.00080	1.61585
11 MU P8/sim08g	dddb-20130929	1550899	0.16610 ± 0.00082	1.17184
	total	14132200		

²⁵⁹⁶ **$B_u \rightarrow J/\psi X, ee, 12952000$**

cond	DDDB	stat	ε^{geom}	dcw
12 MD P6/sim08b	Sim08-20130503-1	255999	—	—
12 MD P6/sim08d	dddb-20130929-1	625499	—	—
12 MD P6/sim08g	dddb-20130929-1	2015460	—	—
12 MD P8/sim08a	Sim08-20130503-1	512500	—	—
12 MD P8/sim08b	Sim08-20130503-1	763497	—	—
12 MD P8/sim08d	dddb-20130929-1	626749	—	—
12 MD P8/sim08g	dddb-20130929-1	2030792	—	—
²⁵⁹⁷ 12 MU P6/sim08b	Sim08-20130503-1	255998	—	—
12 MU P6/sim08d	dddb-20130929-1	641495	—	—
12 MU P6/sim08g	dddb-20130929-1	2016901	—	—
12 MU P8/sim08a	Sim08-20130503-1	512999	—	—
12 MU P8/sim08b	Sim08-20130503-1	766496	—	—
12 MU P8/sim08d	dddb-20130929-1	626499	—	—
12 MU P8/sim08g	dddb-20130929-1	2027596	—	—
	total	13678480		

²⁵⁹⁸ **$B_d \rightarrow J/\psi X, ee, 11453001$**

cond	DDDB	stat	ε^{geom}	dcw
12MD P6/sim08c	dddb-20130929-1	262748	—	—
12MD P6/sim08d	dddb-20130929-1	642998	—	—
12MD P6/sim08g	dddb-20130929-1	2045185	—	—
12MD P8/sim08c	dddb-20130929-1	255750	—	—
12MD P8/sim08d	dddb-20130929-1	642998	—	—
²⁵⁹⁹ 12MD P8/sim08g	dddb-20130929-1	2064629	—	—
12MU P6/sim08c	dddb-20130929-1	265749	—	—
12MU P6/sim08d	dddb-20130929-1	643747	—	—
12MU P6/sim08g	dddb-20130929-1	2020315	—	—
12MU P8/sim08c	dddb-20130929-1	267249	—	—
12MU P8/sim08d	dddb-20130929-1	639996	—	—
12MU P8/sim08g	dddb-20130929-1	2044245	—	—
	total	11795609		

2600 $B_s \rightarrow J/\psi X, ee$, 13454001

cond	DDDB	stat	ε^{geom}	dcw
12MD P6/sim08c	dddb-20130929-1	256249	—	—
12MD P6/sim08d	dddb-20130929-1	631248	—	—
12MD P6/sim08g	dddb-20130929-1	2037454	—	—
12MD P8/sim08c	dddb-20130929-1	260499	—	—
12MD P8/sim08d	dddb-20130929-1	639500	—	—
2601 12MD P8/sim08g	dddb-20130929-1	2027193	—	—
12MU P6/sim08c	dddb-20130929-1	262250	—	—
12MU P6/sim08d	dddb-20130929-1	641999	—	—
12MU P6/sim08g	dddb-20130929-1	2017527	—	—
12MU P8/sim08c	dddb-20130929-1	262499	—	—
12MU P8/sim08d	dddb-20130929-1	639748	—	—
12MU P8/sim08g	dddb-20130929-1	2036855	—	—

2602 $\Lambda_b \rightarrow J/\psi X, ee$, 15454101

cond	DDDB	stat	ε^{geom}	dcw
12MD P6/sim08c	dddb-20130929-1	256250	—	—
12MD P6/sim08d	dddb-20130929-1	640747	—	—
12MD P6/sim08g	dddb-20130929-1	2034300	—	—
12MD P8/sim08c	dddb-20130929-1	263250	—	—
12MD P8/sim08d	dddb-20130929-1	639496	—	—
2603 12MD P8/sim08g	dddb-20130929-1	2017229	—	—
12MU P6/sim08c	dddb-20130929-1	262000	—	—
12MU P6/sim08d	dddb-20130929-1	642496	—	—
12MU P6/sim08g	dddb-20130929-1	2006349	—	—
12MU P8/sim08c	dddb-20130929-1	260000	—	—
12MU P8/sim08d	dddb-20130929-1	637248	—	—
12MU P8/sim08g	dddb-20130929-1	2020123	—	—

2604 $B_d \rightarrow J/\psi K^*, ee$, 11154001

cond	DDDB	stat	ε^{geom}	dcw
12MD P8/sim08a	Sim08-20130503-1	507000	—	—
12MD P8/sim08b	Sim08-20130503-1	507998	—	—
12MD P8/sim08g	dddb-201300929-1	4051941	—	—
2605 12MU P8/sim08a	Sim08-20130503-1	511498	—	—
12MU P8/sim08b	Sim08-20130503-1	502997	—	—
12MU P8/sim08g	dddb-201300929-1	4037069	—	—
11MD	dddb-20130929	2010541	—	—
11MU	dddb-20130929	2030405	—	—

²⁶⁰⁶ **$B_u \rightarrow J/\psi K, \mu\mu$, 12143001**

cond	DDDB	stat	ε^{geom}	dcw
12MD P6/sim08a	Sim08-20130503-1	2520993	0.15723 ± 0.00047	0.912
12MD P6/sim08e	dddb-20130929-1	2009332	0.15723 ± 0.00047	0.912
12MD P8/sim08a	Sim08-20130503-1	2516993	0.16734 ± 0.00049	0.912
12MD P8/sim08e	dddb-20130929-1	2011676	0.16734 ± 0.00049	0.912
12MU P6/sim08a	Sim08-20130503-1	2508595	0.15776 ± 0.00047	0.919
12MU P6/sim08e	dddb-20130929-1	2013916	0.15776 ± 0.00047	0.919
12MU P8/sim08a	Sim08-20130503-1	2530192	0.16712 ± 0.00049	0.919
²⁶⁰⁷ 12MU P8/sim08e	dddb-20130929-1	2008996	0.16712 ± 0.00049	0.919
11MD P6/sim08a	Sim08-20130503	378198	0.15489 ± 0.00029	1.464
11MD P6/sim08c	dddb-20130929	1026997	0.15489 ± 0.00029	1.464
11MD P8/sim08a	dddb-20130929	378499	0.16366 ± 0.00030	1.464
11MD P8/sim08c	dddb-20130929	1006497	0.16366 ± 0.00030	1.464
11MU P6/sim08a	Sim08-20130503	394399	0.15511 ± 0.00029	1.082
11MU P6/sim08c	dddb-20130929	1028996	0.15511 ± 0.00029	1.082
11MU P8/sim08a	dddb-20130929	383799	0.16379 ± 0.00030	1.082
11MU P8/sim08c	dddb-20130929	1011997	0.16379 ± 0.00030	1.082

²⁶⁰⁸ **$B_d \rightarrow K\mu\mu$, 12113001**

cond	DDDB	stat	ε^{geom}	dcw
12MD P6/sim08a	Sim08-20130503-1	271750	0.1600 ± 0.000591	2.03201
12MD P8/sim08a	Sim08-20130503-1	256999	0.1698 ± 0.000622	2.03201
²⁶⁰⁹ 12MU P6/sim08b	Sim08-20130503-1	263247	0.1603 ± 0.000591	2.02897
12MU P8/sim08b	Sim08-20130503-1	270248	0.1698 ± 0.000622	2.02897
11MD P8/sim09a	ddb-20130929	1018565	0.1571 ± 0.000409	0.52121
11MU P8/sim09a	ddb-20130929	1003962	0.1576 ± 0.000410	0.395136

²⁶¹⁰ **$B_u \rightarrow \psi(2S)K, \mu\mu$, 12143020**

cond	DDDB	stat	ε^{geom}	dcw
12MD P6/sim08e	dddb-20130929-1	1021406	0.16105 ± 0.00036	1.04734
12MD P8/sim08e	dddb-20130929-1	1063216	0.16954 ± 0.00038	1.04734
12MU P6/sim08e	dddb-20130929-1	1037708	0.16105 ± 0.00036	1.06423
²⁶¹¹ 12MU P8/sim08e	dddb-20130929-1	1029150	0.16954 ± 0.00038	1.06423
11MD P6/sim08e	dddb-20130929	515153	0.15739 ± 0.00034	1.05808
11MD P8/sim08e	dddb-20130929	504438	0.16698 ± 0.00036	1.05808
11MU P6/sim08e	dddb-20130929	536422	0.15739 ± 0.00034	0.73458
11MU P8/sim08e	dddb-20130929	560982	0.16698 ± 0.00036	0.73458

²⁶¹² **Cascade $B_u \rightarrow D_0 e\nu, K e\nu$, 12583021**

²⁶¹³ Warning: tight cuts: $p_T > 200$ MeV, $m(Kee) > 4$ GeV

	cond	DDDB	stat	ε^{geom}	dcw
2614	12MD P8/sim08h	dddb-20130929-1	1011237	0.005528 ± 0.000024	—
	12MU P8/sim08h	dddb-20130929-1	1000670	0.005471 ± 0.000023	—

2615 **Cascade $B_u \rightarrow D_0 e\nu, K\pi$, 12583013**

2616	Warning: tight cuts: $p_T > 200$ MeV, $m(Kee) > 4$ GeV	cond	DDDB	stat	ε^{geom}	dcw
2617		12MD P8/sim08h	dddb-20130929-1	1002007	0.027858 ± 0.000075	—
		12MU P8/sim08h	dddb-20130929-1	1011099	0.027732 ± 0.000078	—

2618 **Cascade $B_u \rightarrow D_0 \pi, Ke\nu$, 12183004**

2619	Warning: tight cuts: $p_T > 200$ MeV, $m(Kee) > 4$ GeV	cond	DDDB	stat	ε^{geom}	dcw
2620		12MD P8/sim08h	dddb-20130929-1	1003102	0.06207 ± 0.00015	—
		12MU P8/sim08h	dddb-20130929-1	1010150	0.06211 ± 0.00015	—

2621 **$B \rightarrow K_1 ee, K\pi\pi$, 12425000**

	cond	DDDB	stat	ε^{geom}	dcw
2622	12 MD P8/sim08g	dddb-20130929-1	1506127	0.16081 ± 0.00037	—
	12 MU P8/sim08g	dddb-20130929-1	1535079	0.16050 ± 0.00036	—

2623 **$B \rightarrow K_2 ee, K\pi\pi$, 12425011**

	cond	DDDB	stat	ε^{geom}	dcw
2624	12 MD P8/sim08h	dddb-20130929-1	722557	0.14847 ± 0.00024	—
	12 MU P8/sim08h	dddb-20130929-1	525911	0.14866 ± 0.00027	—

2625 **A.2 Run 2 samples**

2626 Information about all the Run 2 simulation samples are summarised below. All Run 2
 2627 samples are produced using sim09b. The DDDB and SIMCOND tags are the same for all
 2628 the simulation samples in run2:

- 2629 • 2016 tags: DDDB: dddb-20150724, SIMCOND: sim-20161124-2-vc-md100
- 2630 • 2015 tags: DDDB: dddb-20150724, SIMCOND: sim-20161124-vc-md100

²⁶³¹ **$B \rightarrow K^{*0} ee$ 11124001**

	cond	ProdID	stat	ε^{geom}	dcw
²⁶³²	16 MD	59827	1067005	0.17315 ± 0.00045	1.26347
	16 MU	59821	1014454	0.17334 ± 0.00044	1.2506
	15 MD	62323	503571	0.17274 ± 0.00047	0.55164
	15 MU	62315	506656	0.17220 ± 0.00056	0.38901
	total		3091686		

²⁶³³ **$B^+ \rightarrow K^+ \mu\mu$ 12113001**

	cond	ProdID	stat	ε^{geom}	dcw
²⁶³⁴	16 MD	59825	1008747	0.17729 ± 0.00045	1.30803
	16 MU	59819	1008846	0.17610 ± 0.00044	1.23082
	15 MD	62235	504547	0.17665 ± 0.00044	0.53887
	15 MU	62229	503827	0.17625 ± 0.00043	0.38288

²⁶³⁵ **$B^0 \rightarrow \psi(2S)K^+, \mu\mu$ 12143020**

	cond	ProdID	stat	ε^{geom}	dcw
²⁶³⁶	16 MD	59831	2283476	0.17634 ± 0.00042	1.06483
	16 MU	59817	2269487	0.17719 ± 0.00044	1.00824
	15 MD	62233	504529	0.17618 ± 0.00045	0.99305
	15 MU	62231	518717	0.17661 ± 0.00045	0.68531

²⁶³⁷ **$B^+ \rightarrow \psi(2S)K^+, ee$ 12153011**

	cond	ProdID	stat	ε^{geom}	dcw
²⁶³⁸	16 MD	59829	579206	0.17553 ± 0.00060	1.24937
	16 MU	59815	573446	0.17583 ± 0.00060	1.18755
	15 MD	62329	256019	0.17602 ± 0.00065	0.582417
	15 MU	62319	250872	0.17458 ± 0.00065	0.421713

²⁶³⁹ **$B \rightarrow J/\psi K^*, ee$ 11154001**

	cond	ProdID	stat	ε^{geom}	dcw
²⁶⁴⁰	16 MD	57755, 59823	2550837	0.16641 ± 0.00043	1.08366
	16 MU	57749, 59813	2731899	0.16701 ± 0.00041	0.952201
	15 MD	57733	519963	0.16659 ± 0.00042	1.09543
	15 MU	57731	536568	0.16681 ± 0.00041	0.753173

²⁶⁴¹ $B^+ \rightarrow J/\psi K^+, ee$ **12153001**

	cond	ProdID	stat	ε^{geom}	dcw
	16 MD	56239	2033236	0.17312 ± 0.00040	1.51091
²⁶⁴²	16 MU	56243	2085444	0.17261 ± 0.00039	1.38627
	15 MD	60346	1516859	0.17330 ± 0.00037	0.417315
	15 MU	60334	1409619	0.17236 ± 0.00040	0.318617

²⁶⁴³ $B^+ \rightarrow J/\psi K^+, \mu\mu$ **12143001**

	cond	ProdID	stat	ε^{geom}	dcw
	16 MD	58335, 60495	6916023	0.17381 ± 0.00047	0.99364
²⁶⁴⁴	16 MU	58331, 60491	6678686	0.17380 ± 0.00039	0.96831
	15 MD	58327	1007420	0.17377 ± 0.00047	1.40558
	15 MU	58317	1157610	0.17245 ± 0.00045	0.86790

²⁶⁴⁵ $B^+ \rightarrow K^+ ee$ **12123002**

	cond	ProdID	stat	ε^{geom}	dcw
	16 MD	62313	1005401	0.17449 ± 0.00044	1.3106
²⁶⁴⁶	16 MU	62307	1001896	0.17456 ± 0.00042	1.23767
	15 MD	62325	514308	0.17519 ± 0.00047	0.52792
	15 MU	62321	500249	0.17508 ± 0.00049	0.385095

²⁶⁴⁷ $B^0 \rightarrow K^{*0} \psi(2S), ee$ **11154011**

	cond	ProdID	stat	ε^{geom}	dcw
	16 MD	62311	500200	0.16842 ± 0.00044	1.32541
²⁶⁴⁸	16 MU	62309	508777	0.16909 ± 0.00042	1.22627
	15 MD	62327	258725	0.16963 ± 0.00046	0.528004
	15 MU	62317	252698	0.16988 ± 0.00047	0.383563

²⁶⁴⁹ $B^+ \rightarrow J/\psi \pi, \mu\mu$ **12143010**

	cond	ProdID	stat	ε^{geom}	dcw
	16 MD	57863	1042764		1.75605
²⁶⁵⁰	16 MU	57859	1077943		1.59862
	15 MD	57590	1053713		0.358081
	15 MU	57586	1024962		0.261191

²⁶⁵¹ **B BDT: supplementary plots**

²⁶⁵² This section contains some extra information about the BDT.

²⁶⁵³ The BDT output distributions for the training and test samples superimposed are
²⁶⁵⁴ shown in Fig. 136, for all folds. Due to the limited size of the combinatorial sample, some

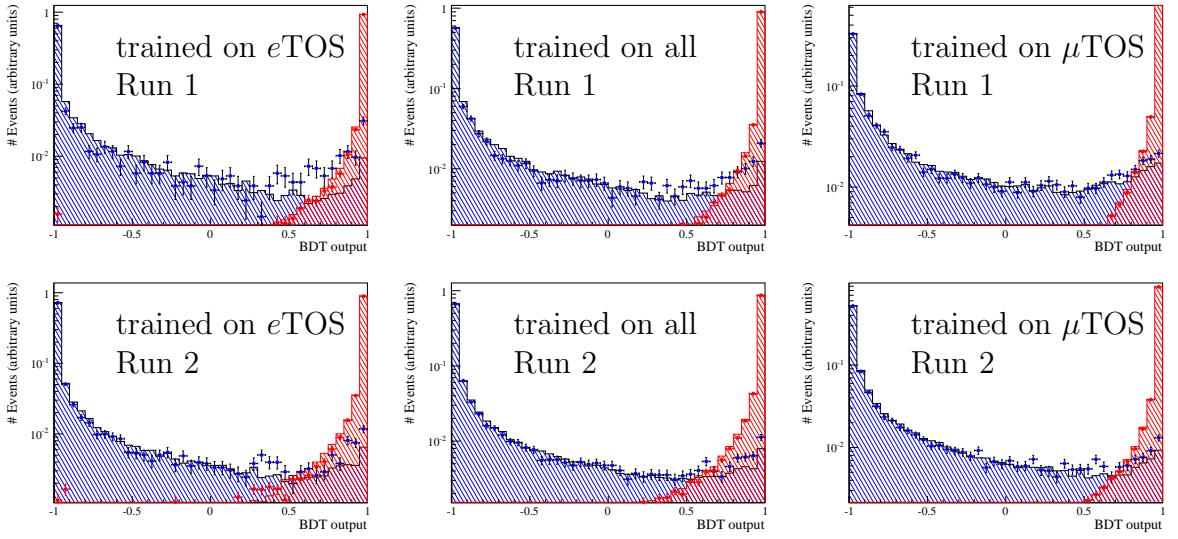


Figure 136: Overtraining plots for the various BDTs. The top plots correspond to Run 1 and the bottom to Run 2. The left plots are from the BDT trained on $e\text{TOS}$ events, the middle plots are from the BDT trained on all electron mode trigger and applied to $h\text{TOS!}$ and TIS! , the right plots are from the BDT used on the muon mode.

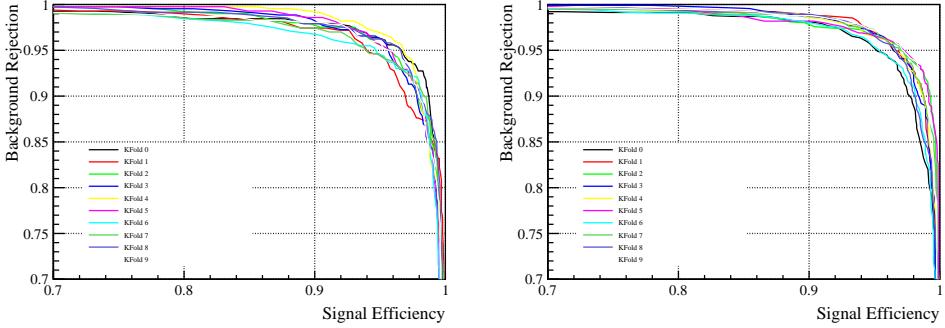


Figure 137: ROC curves for the different folds for Run 1 (left), Run 2 (right), for the electron mode.

2655 slight overtraining can be seen. However, given the folding technique, this cannot bias the
2656 efficiency computation.

2657 The ROC curves are shown separately for each fold in Fig. 137. The statistical
2658 fluctuations are large but there is good compatibility.

2659 The expected signal significance is shown in Fig. 138 as a function of the BDT and
2660 PID requirement. This shows that the requirement $PIDe > 3$ is optimal.

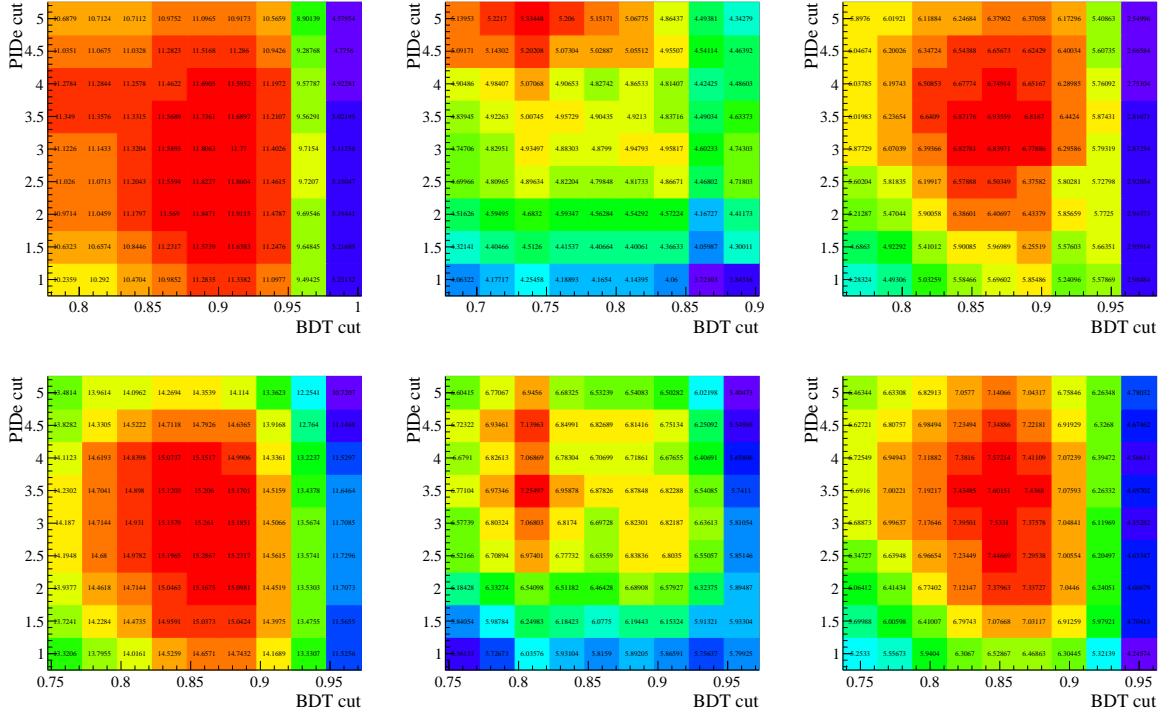


Figure 138: Expected signal $B^+ \rightarrow K^+ e^+ e^-$ significance as a function of the PID cut, for Run 1 (top) and Run 2 (bottom), and, from left to right, $e\text{TOS}$, $h\text{TOS!}$ and TIS! trigger categories. The working point corresponds to the central bin in each plot. In the case of the $e\text{TOS}$ and TIS! category, the working point is optimal within the statistical error of approximately 0.3 in each bin. For the $h\text{TOS!}$ category, tightening the PID requirement brings some improvement, but the statistical power of this trigger category is anyways limited.

C Extra fit plots

C.1 $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ simulation fit

The fits to $m_{\text{DTF}}^{J/\psi}$ for offline selected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ simulation events are shown in Fig. 139.

The fits to $m_{\text{DTF}}^{J/\psi}$ for offline selected $B^+ \rightarrow \pi_{\rightarrow K^+}^+ J/\psi(\mu^+ \mu^-)$ simulation events are shown in Fig. 140.

C.2 $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulation fit

The fits to $m_{\text{DTF}}^{J/\psi}$ for offline selected $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulation events in the three bremsstrahlung categories are shown in Fig. 141, Fig. 142, Fig. 143.

The fits to $m_{\text{DTF}}^{J/\psi}$ for offline selected $B^+ \rightarrow \pi_{\rightarrow K^+}^+ J/\psi(e^+ e^-)$ simulation events are shown in Fig. 144.

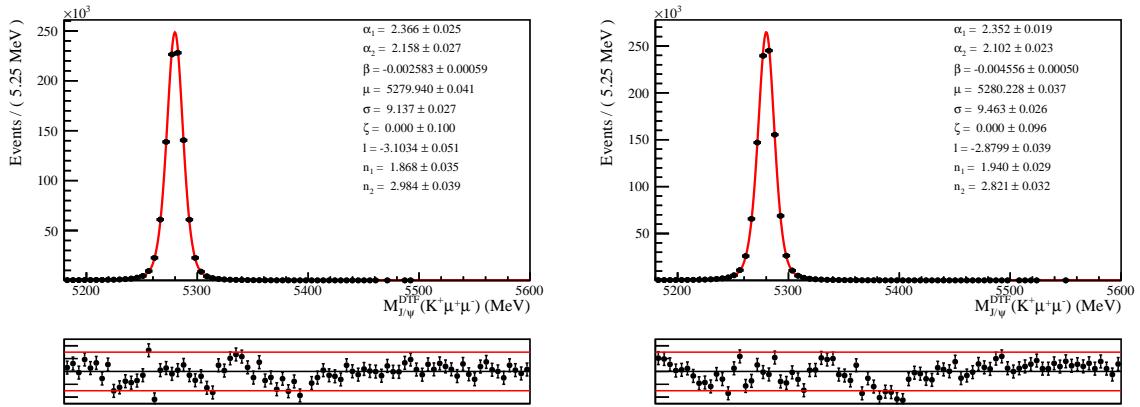


Figure 139: Fits to simulated $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events for Run 1 (left) and Run 2 (right).

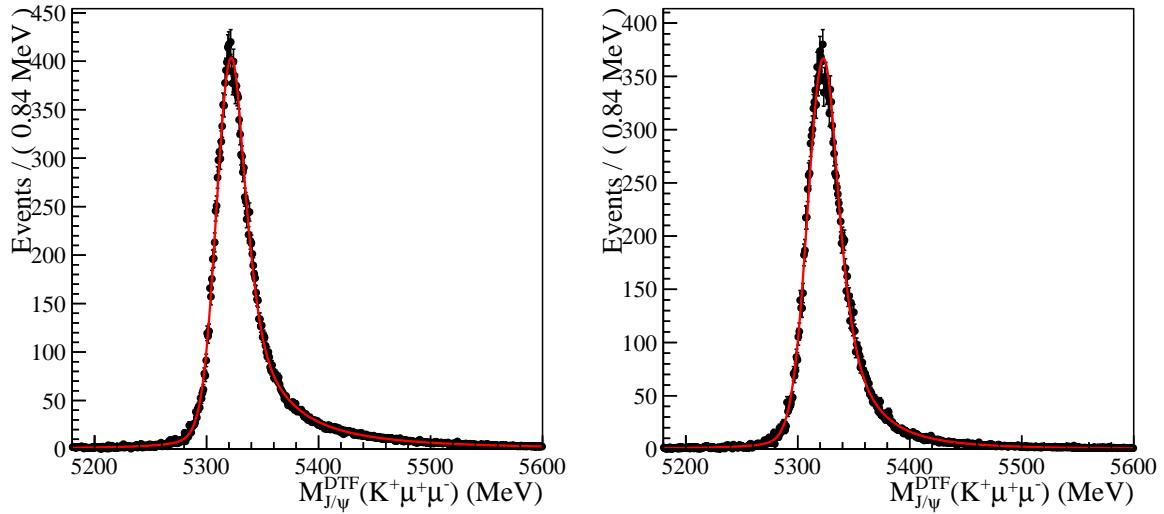


Figure 140: Fits to simulated $B^+ \rightarrow \pi_{\rightarrow K^+}^+ J/\psi(\mu^+ \mu^-)$ events for Run 1 (left) and Run 2 (right).

2672 The fits to $m_{\text{DTF}}^{J/\psi}$ for offline selected $H_b \rightarrow H_s (\rightarrow K^+ X) J/\psi$ simulation events (strange
2673 partially reconstructed background) are shown in Fig. 144.

2674 The fits to $m_{\text{DTF}}^{J/\psi}$ for offline selected $H_b \rightarrow K^+ X \Delta (\rightarrow J/\psi Y)$ simulation events
2675 (charmed partially reconstructed background) are shown in Fig. 145.

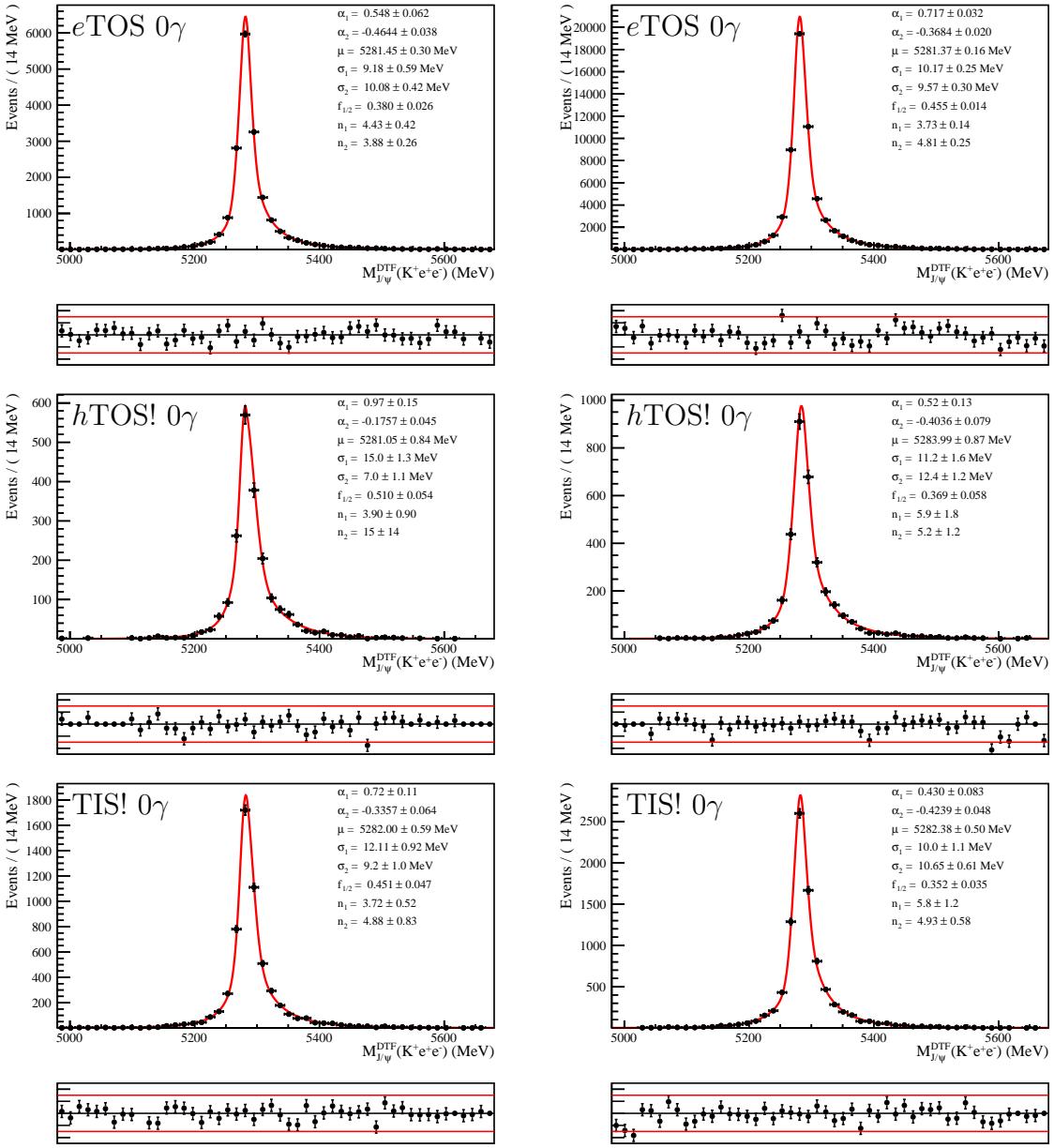


Figure 141: Fits to $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulation events in the bremsstrahlung category 0 γ for Run 1 (left) and Run 2 (right), for eTOS (top), hTOS and TIS events.

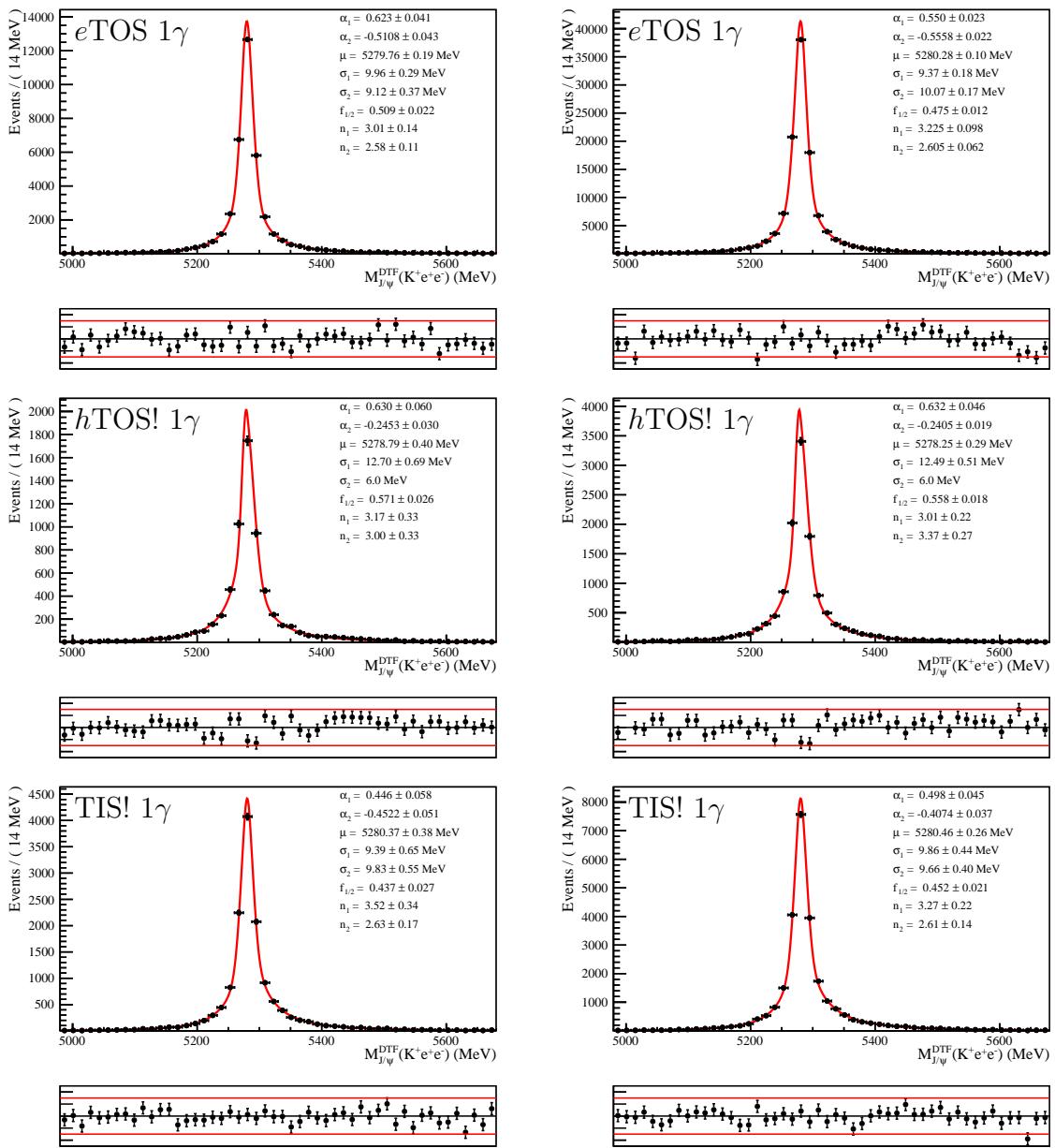


Figure 142: Fits to $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulation events in the bremsstrahlung category 1 γ for Run 1 (left) and Run 2 (right), for eTOS (top), hTOS! and TIS! events.

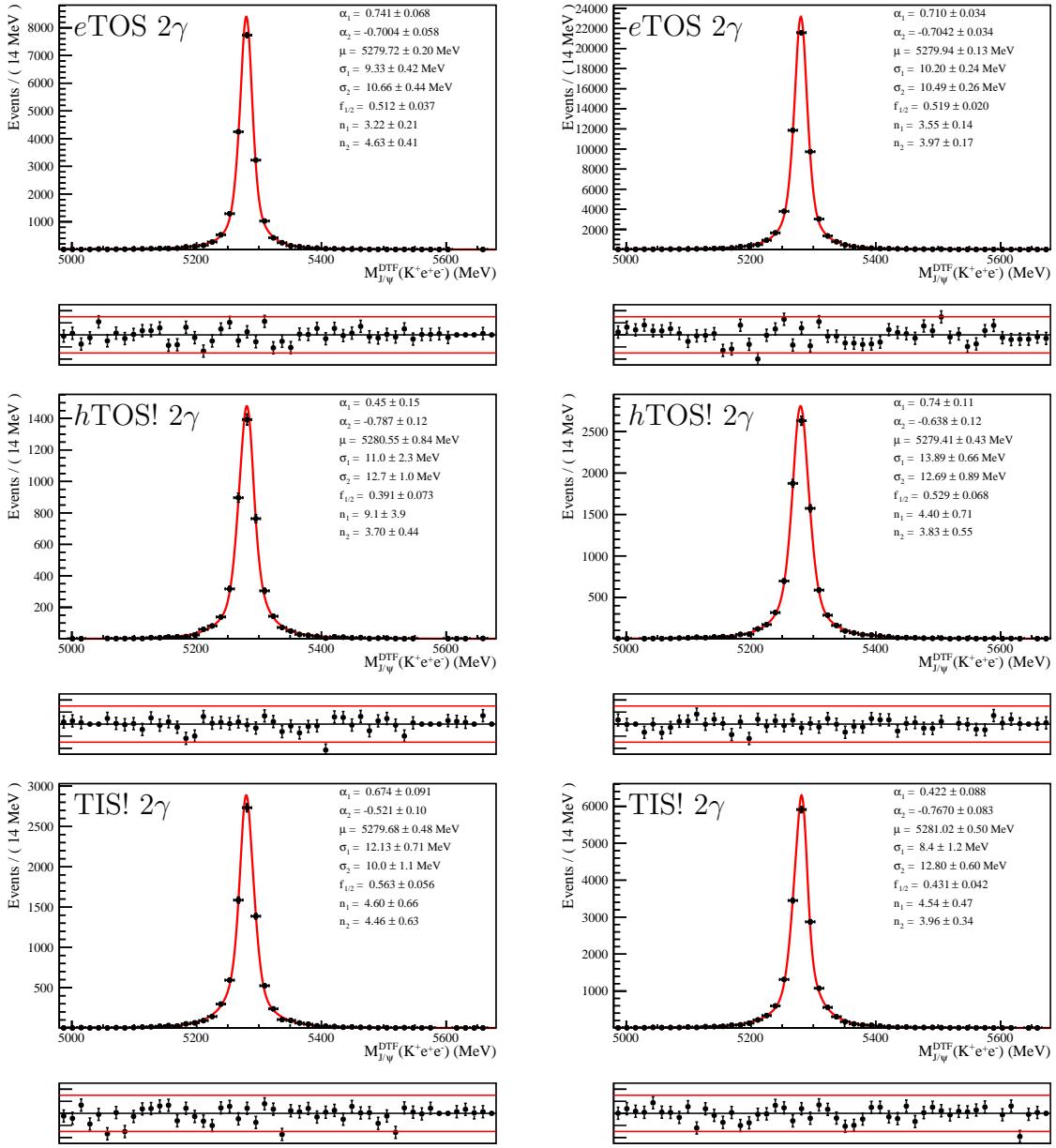


Figure 143: Fits to $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulation events in the bremsstrahlung category 2γ for Run 1 (left) and Run 2 (right), for eTOS (top), hTOS! and TIS! events.

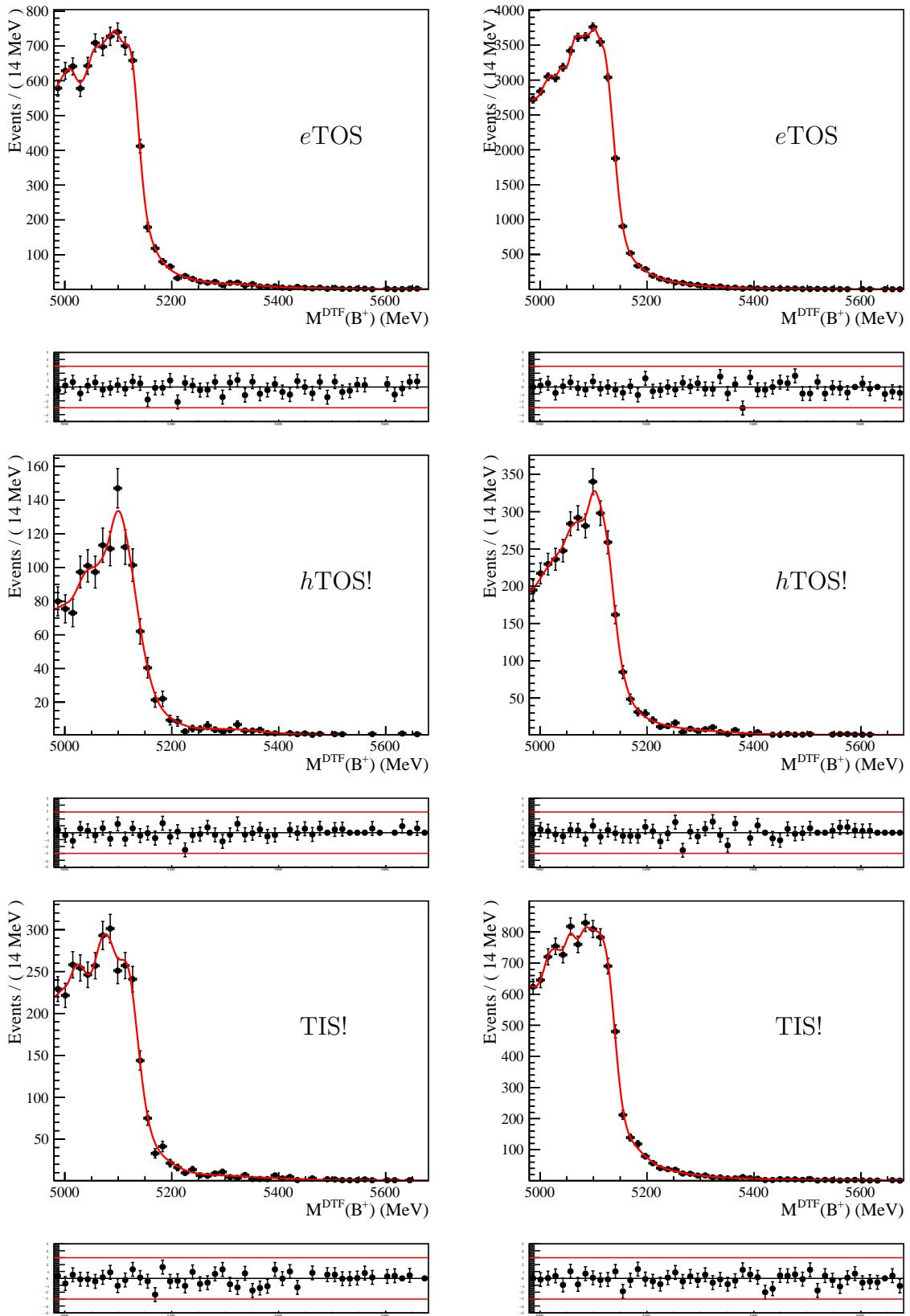


Figure 144: Fits to $H_b \rightarrow H_s(\rightarrow K^+ X) J/\psi$ simulation events (strange partially reconstructed background), for Run 1 (left) and Run 2 (right), for $e\text{TOS}$ (top), $h\text{TOS}!$ and $\text{TIS}!$ events. All three bremsstrahlung categories are summed.

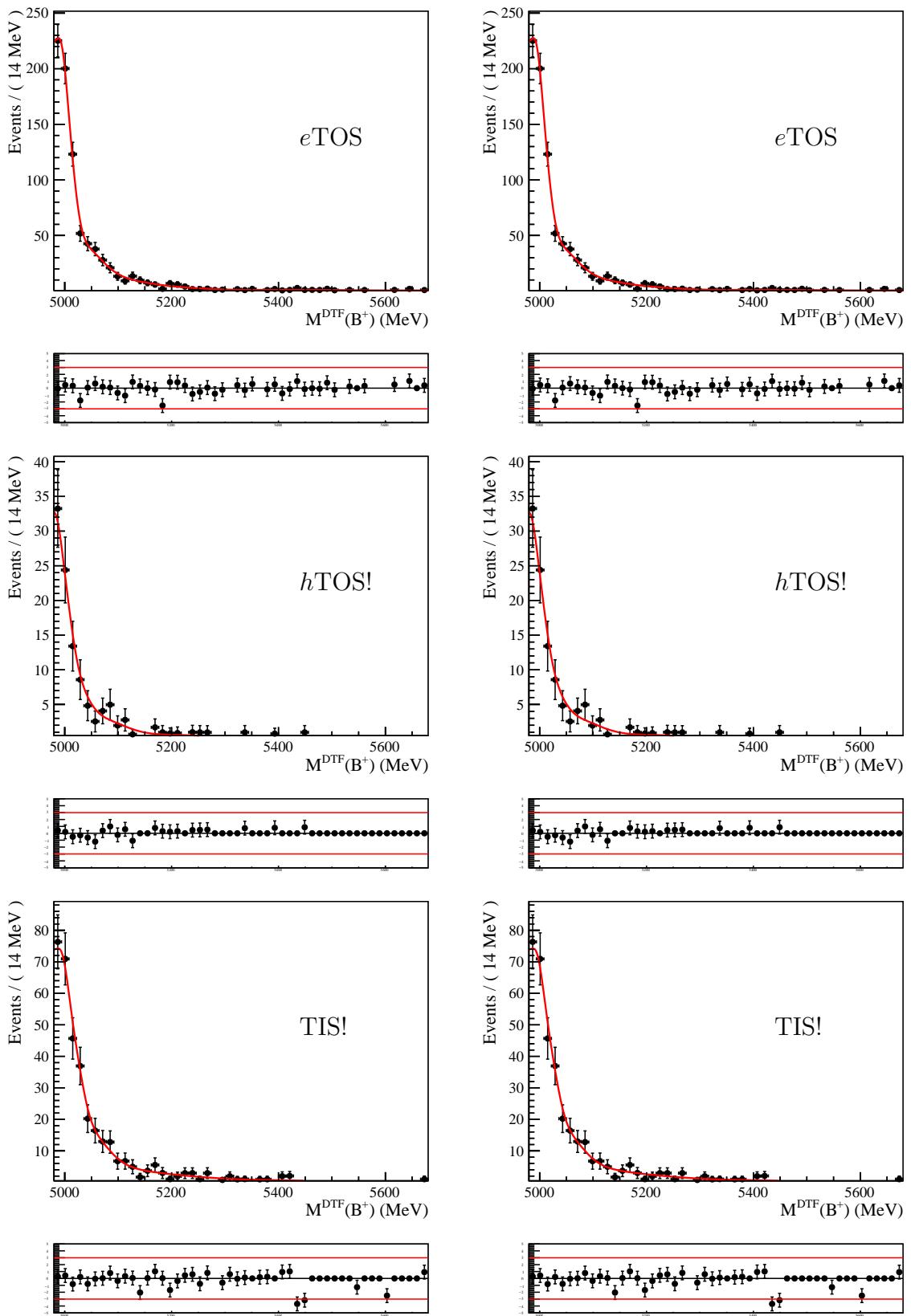


Figure 145: Fits to $H_b \rightarrow K^+ X (J/\psi \rightarrow J/247Y)$ simulation events (strange partially reconstructed background), for Run 1 (left) and Run 2 (right), for $e\text{TOS}$ (top), $h\text{TOS}!$ and $\text{TIS}!$ events. All three bremsstrahlung categories are summed.

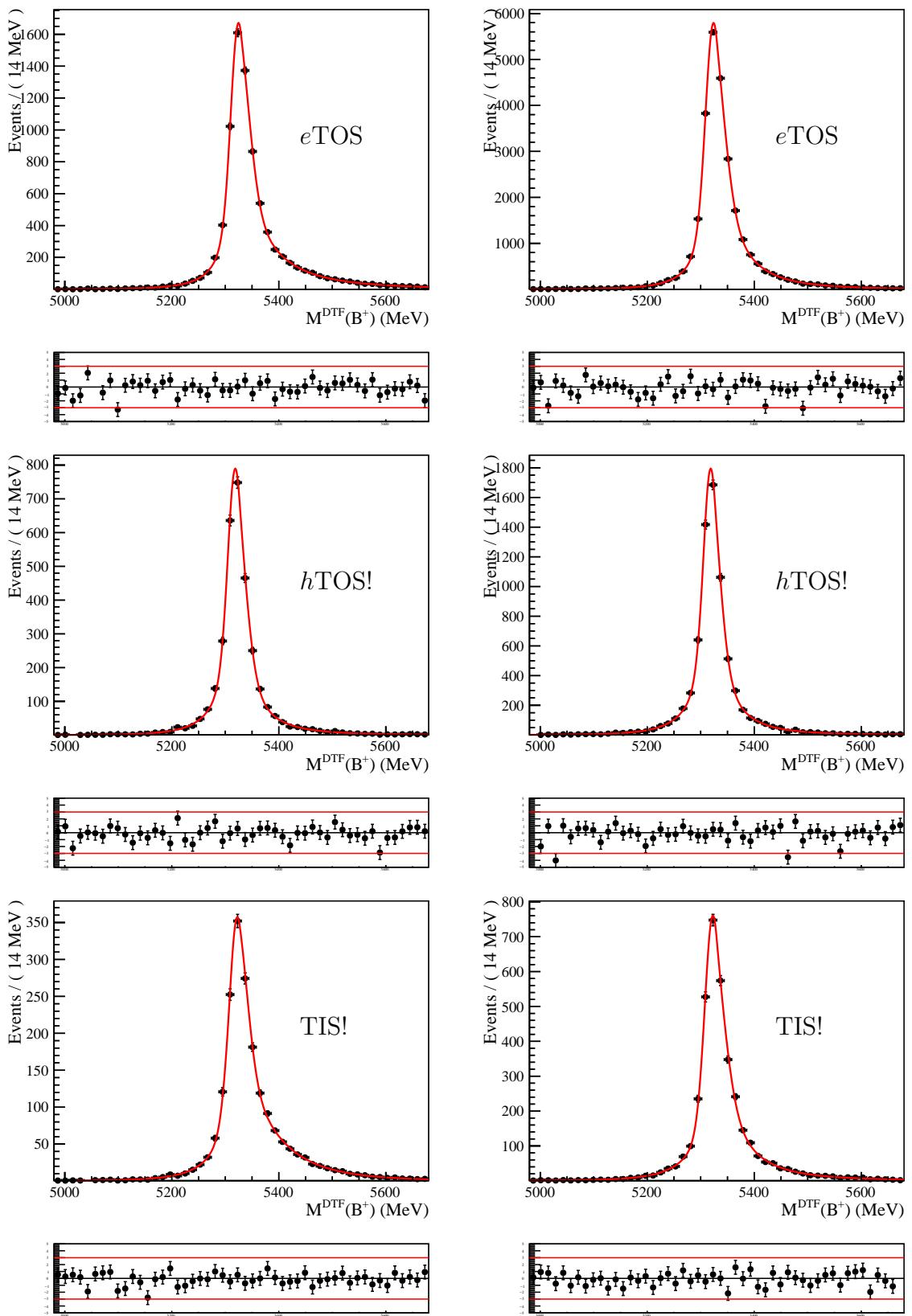


Figure 146: Fits to $B^+ \rightarrow \pi_{\rightarrow K^+}^+ J/\psi(e^+e^-)$ simulation events (charmed partially reconstructed background), for Run 1 (left) and Run 2 (right), for $e\text{TOS}$ (top), $h\text{TOS!}$ and TIS! events. All three bremsstrahlung categories are summed.

2676 **C.3 $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ mis-ID mass variable transformation**

2677 Consider a $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ event reconstructed in two ways: in one case assigning
 2678 the correct mass hypothesis to the pion track, yielding the measured energy E_π and
 2679 reconstructed B^+ invariant mass M , and in the second case assigning the kaon mass
 2680 hypothesis to the pion track, yielding the measured energy E'_π and reconstructed B^+
 2681 invariant mass M' . The difference between the squared pion and kaon masses is denoted
 2682 $\Delta^2 m = m_K^2 - m_\pi^2$. The masses M and M' are given,

$$M = (E_{J/\psi} + E_\pi)^2 - (\mathbf{p}_{J/\psi} + \mathbf{p}_\pi)^2,$$

$$M' = (E_{J/\psi} + E'_\pi)^2 - (\mathbf{p}_{J/\psi} + \mathbf{p}_\pi)^2.$$

2683 Hence:

$$\begin{aligned} M'^2 - M^2 &= 2E_{J/\psi}(E'_\pi - E_\pi) + \Delta^2 m \\ &= 2E_{J/\psi} \left(\sqrt{m_K'^2 + p_\pi^2} - \sqrt{m_\pi^2 + p_\pi^2} \right) + \Delta^2 m \\ &= 2E_{J/\psi} \left(\sqrt{m_\pi'^2 + \Delta^2 m + p_\pi^2} - \sqrt{m_\pi^2 + p_\pi^2} \right) + \Delta^2 m \\ &\approx 2E_{J/\psi} \left(\frac{\Delta^2 m}{2\sqrt{m_\pi^2 + p_\pi^2}} \right) + \Delta^2 m \\ &= \frac{E_{tot}}{E_\pi} \Delta^2 m, \end{aligned}$$

2685 where, in the fourth line, a Taylor expansion of the square roots at first order in m_K^2 is
 2686 performed. Hence, this formula holds in the realistic case of $m_K \ll p_\pi$. Therefore, the B^+
 2687 mass with the incorrect mass assignment to the pion track can be expressed as a function
 2688 of the B^+ mass with the correct mass assignment:

$$M' \approx \left(M^2 + \frac{E_{tot}}{E_\pi} \Delta^2 m \right)^{\frac{1}{2}}.$$

2689 In the case of this analysis, where no $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ simulation sample was
 2690 available, the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulation sample is used to estimate M' :

$$M' \approx \left(M^2 + \frac{E_{tot}}{E_K} \Delta^2 m \right)^{\frac{1}{2}}.$$

2691 The impact of the particle ID selection is evaluated using the PIDCalib tool. Efficiency
 2692 weights for $\pi \rightarrow K$ mis-ID as a function of the pion η and p corresponding to the particle
 2693 ID requirements in the selection are derived. Subsequently, these weights are applied to
 2694 the kaon in each $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ candidate. This procedure has been validated using
 2695 the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$, for which MC simulation is available,
 2696 see Fig. 147.

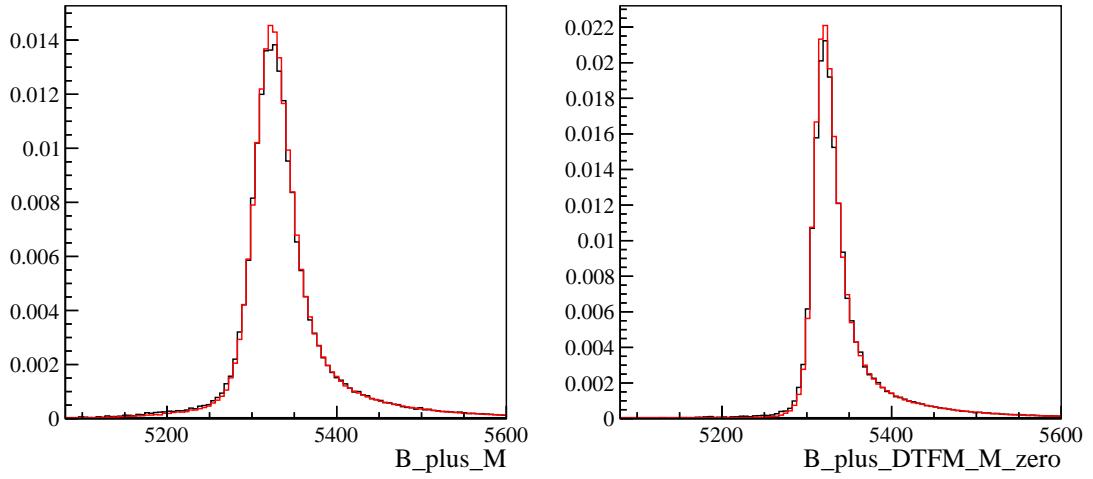
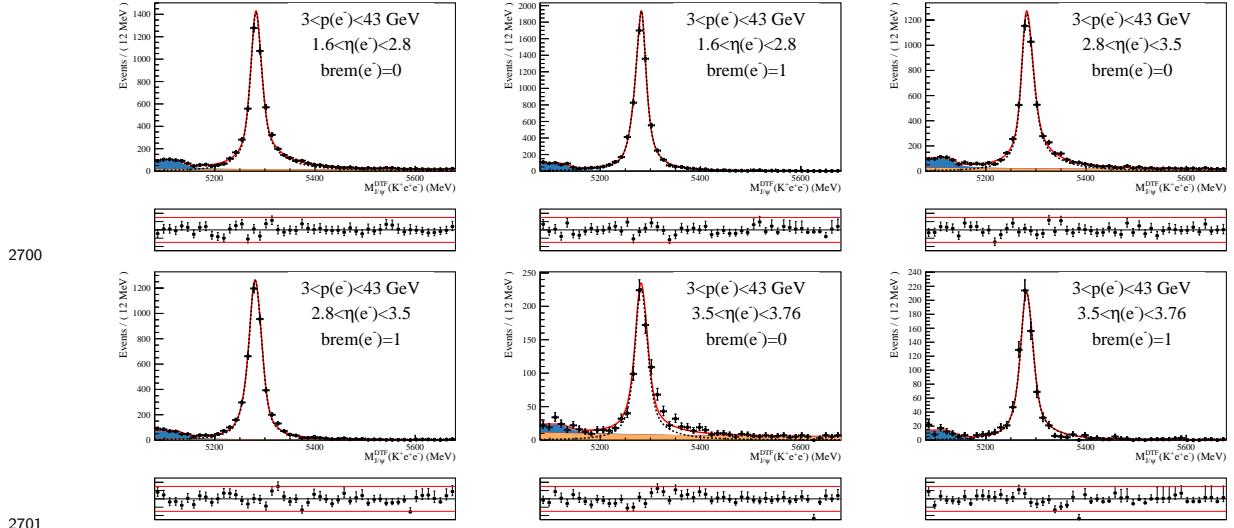


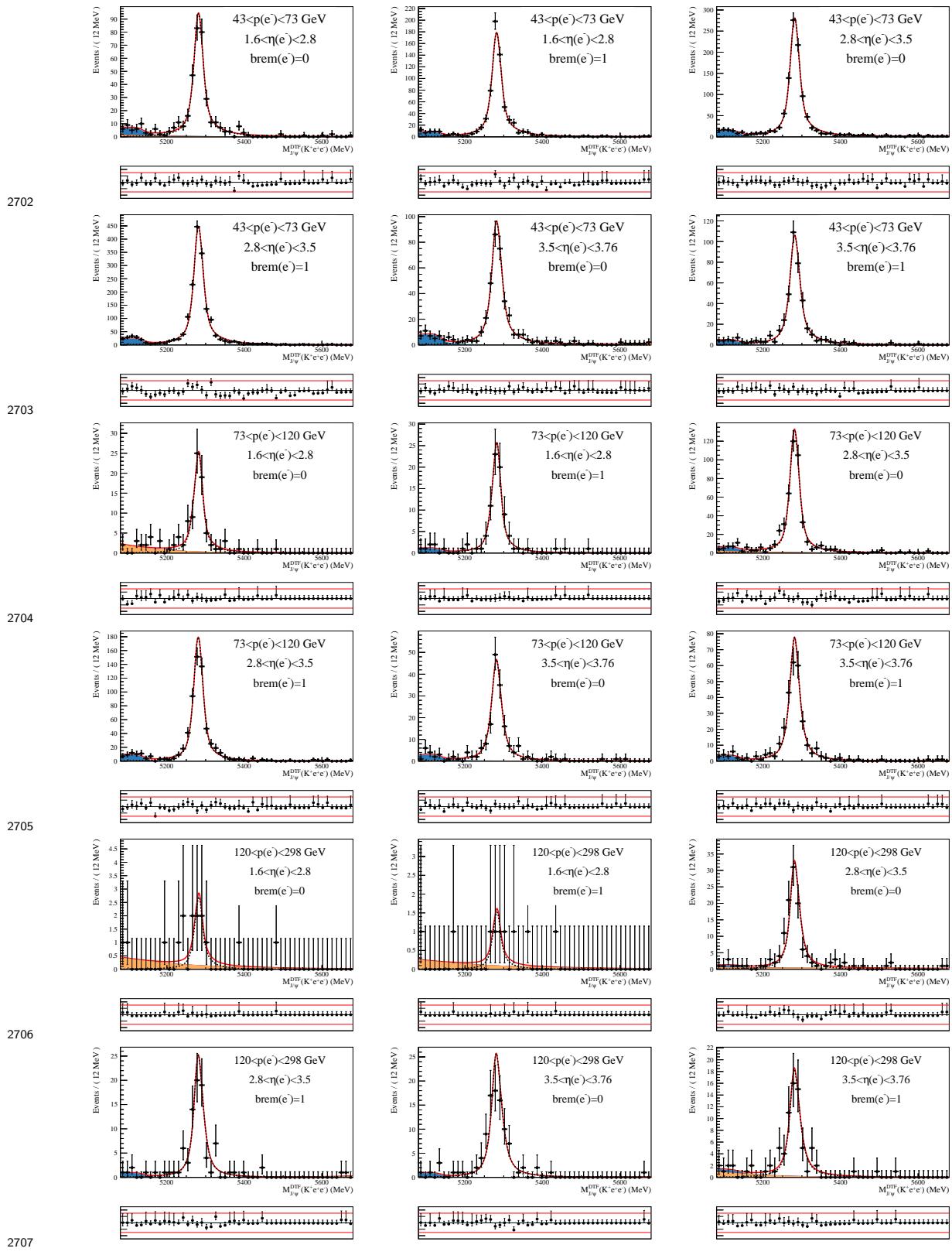
Figure 147: Comparison of the (left) $m(K^+\mu^+\mu^-)$ and (right) $m_{\text{DTF}}^{J/\psi}(K^+\mu^+\mu^-)$ distributions obtained from (black) $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ simulated events with that obtained using (red) $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ through the transformation explained in the text.

2697 D Fits electron calibration samples

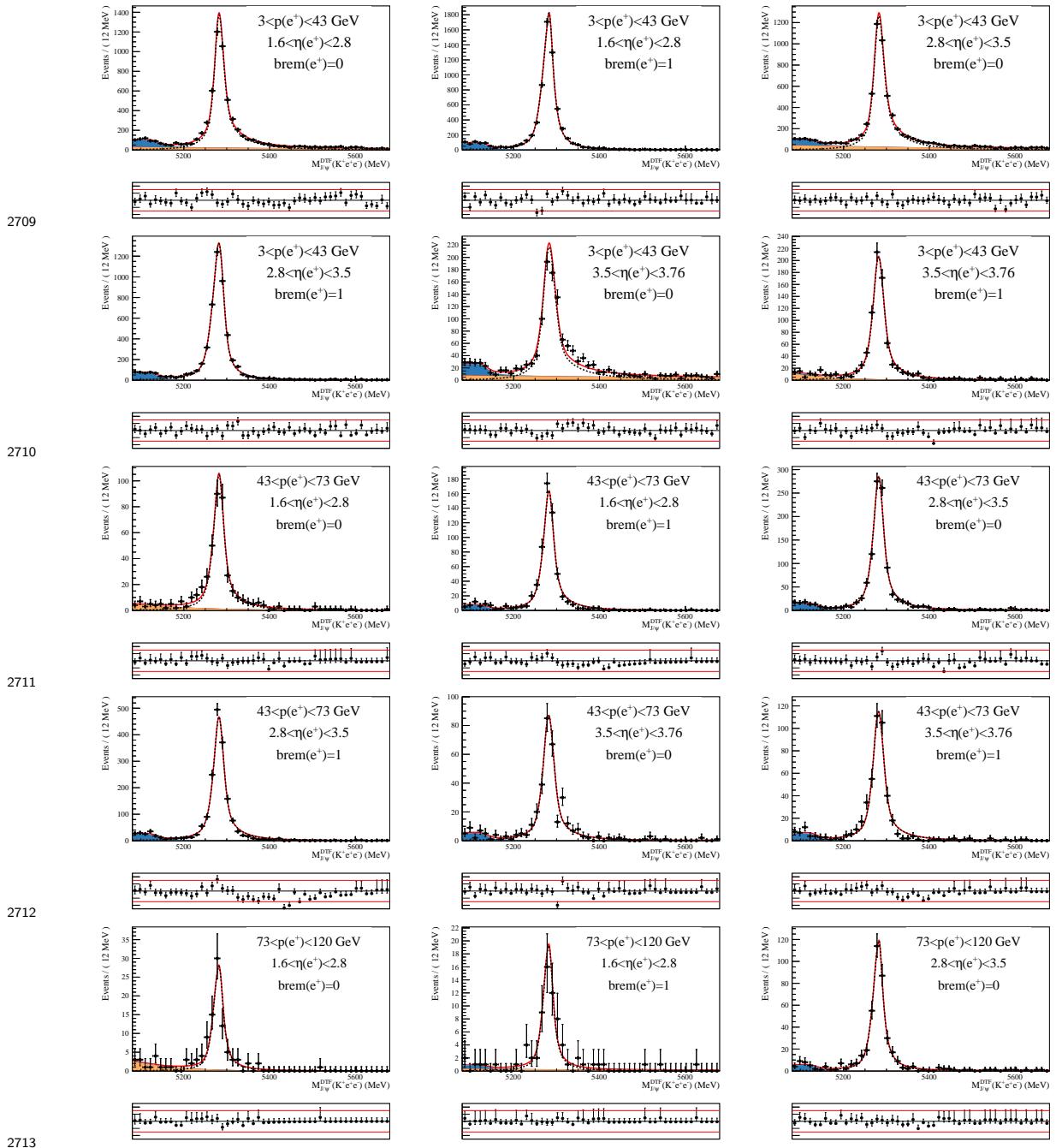
2698 D.1 2011 Fits

2699 D.1.1 2011 Fits, e^- probe, before PID cut

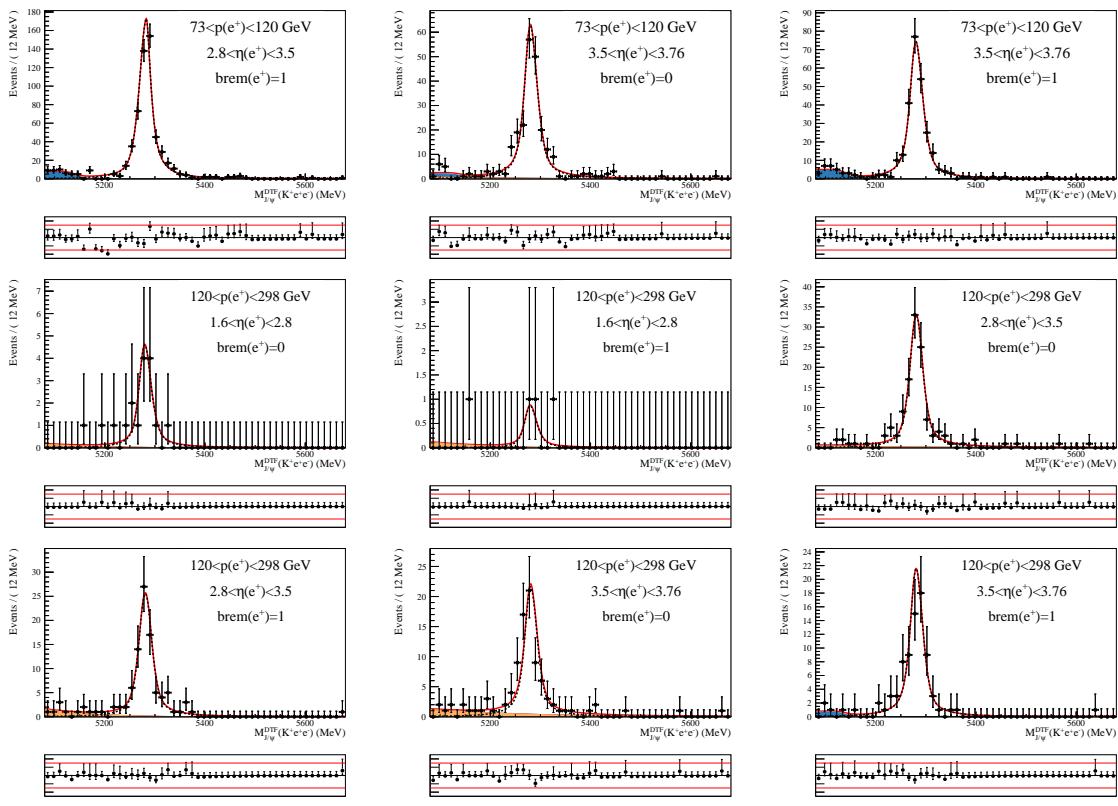




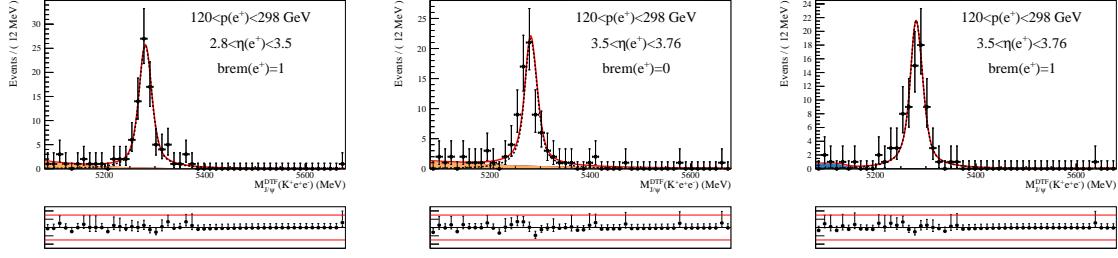
2708 D.1.2 2011 Fits, e^+ probe, before PID cut



2714



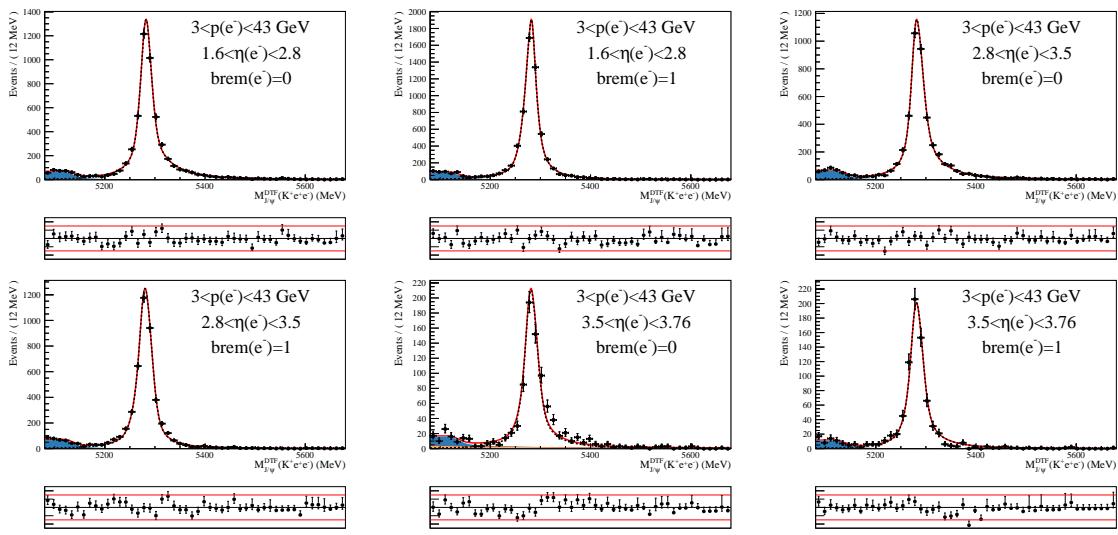
2715



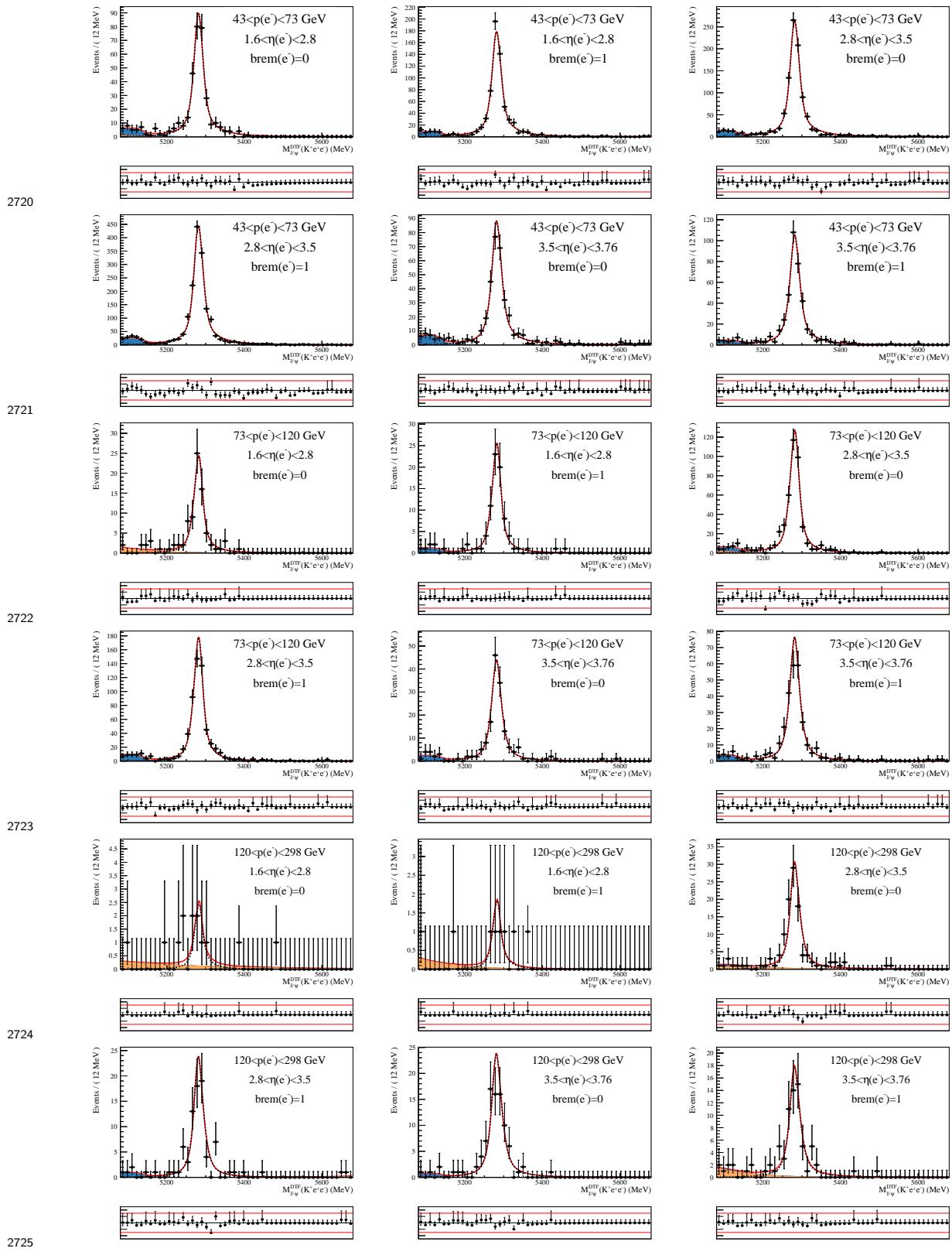
2716

2717 D.1.3 2011 Fits, e^- probe, after PID cut

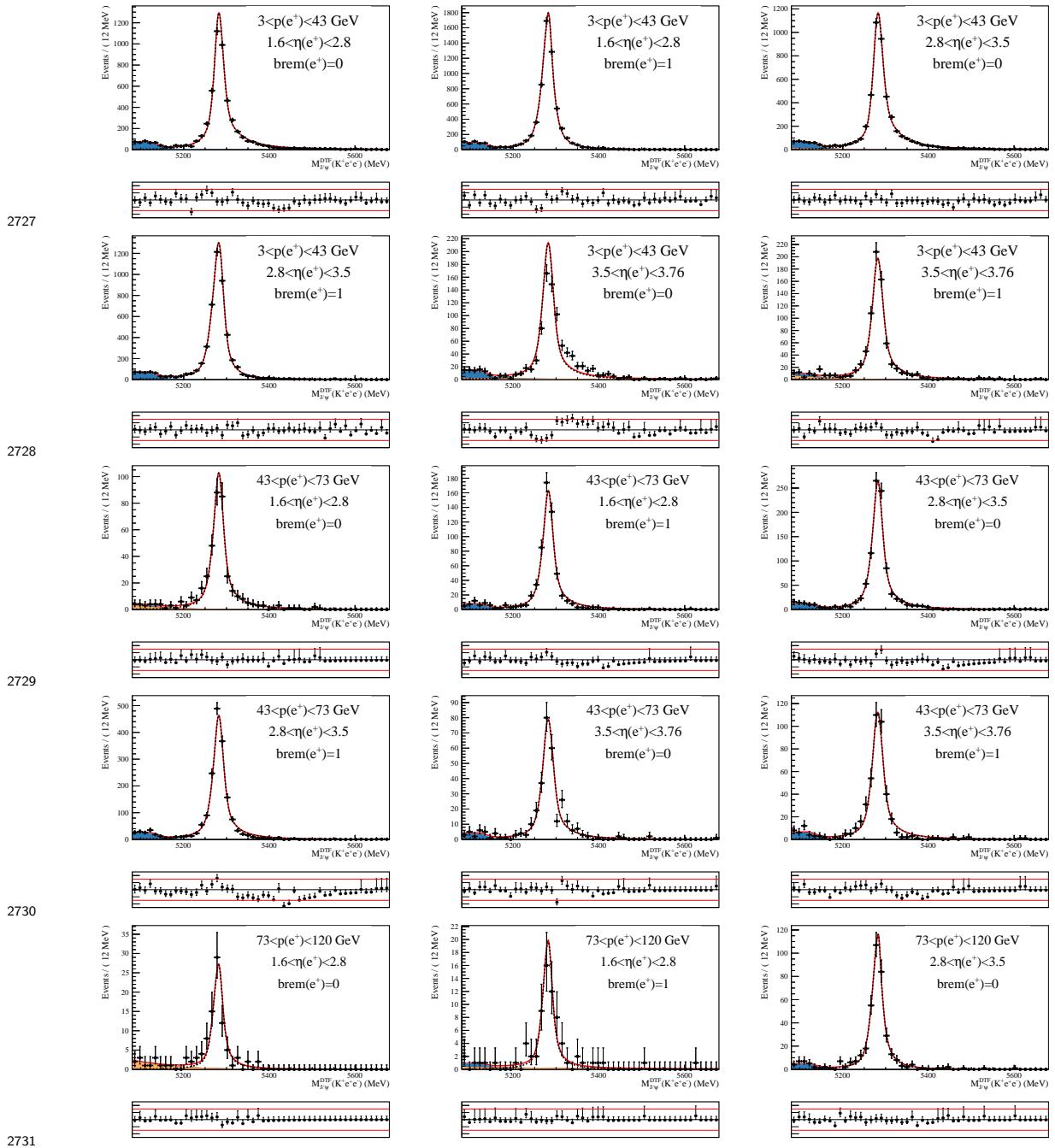
2718

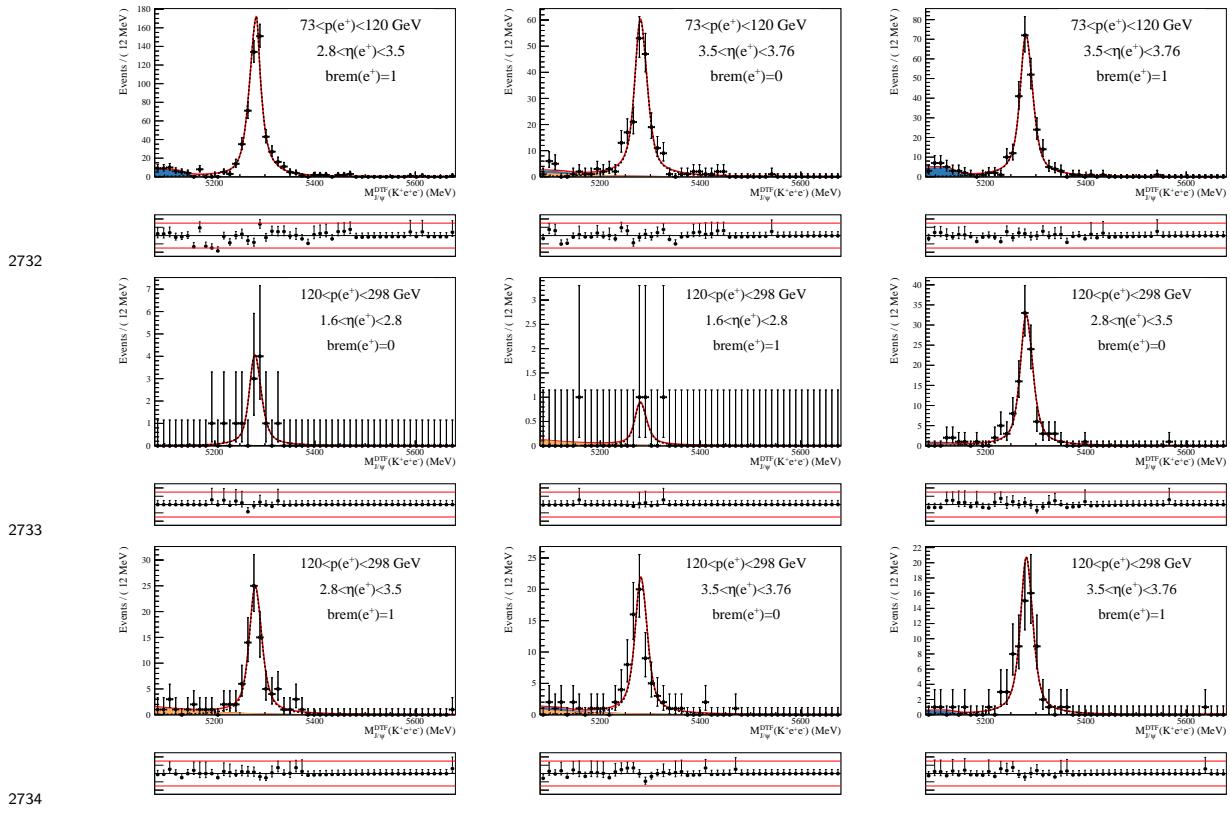


2719



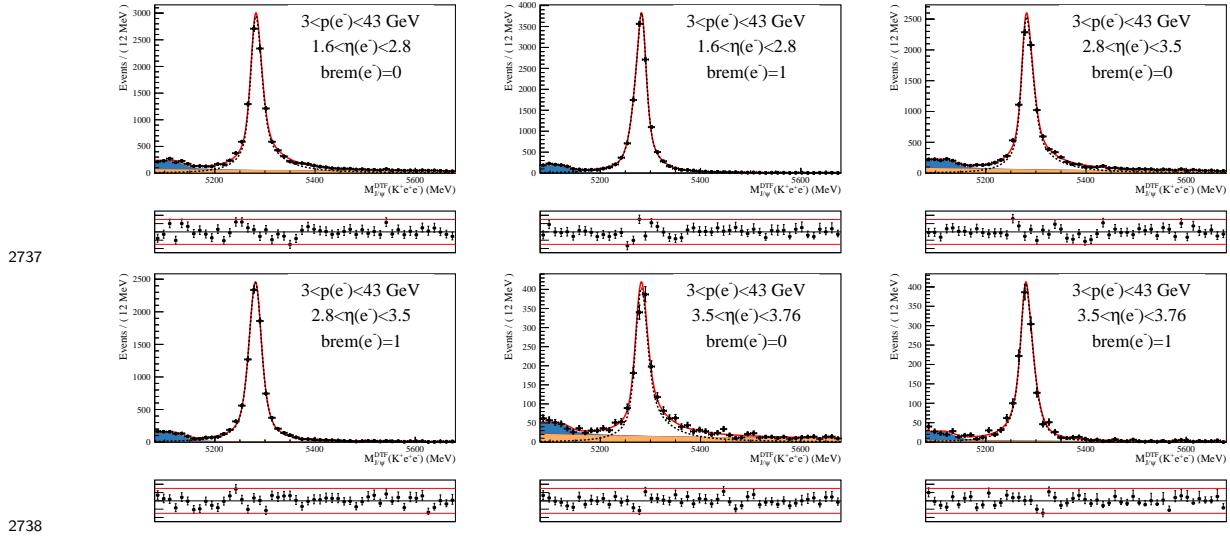
2726 D.1.4 2011 Fits, e^+ probe, after PID cut

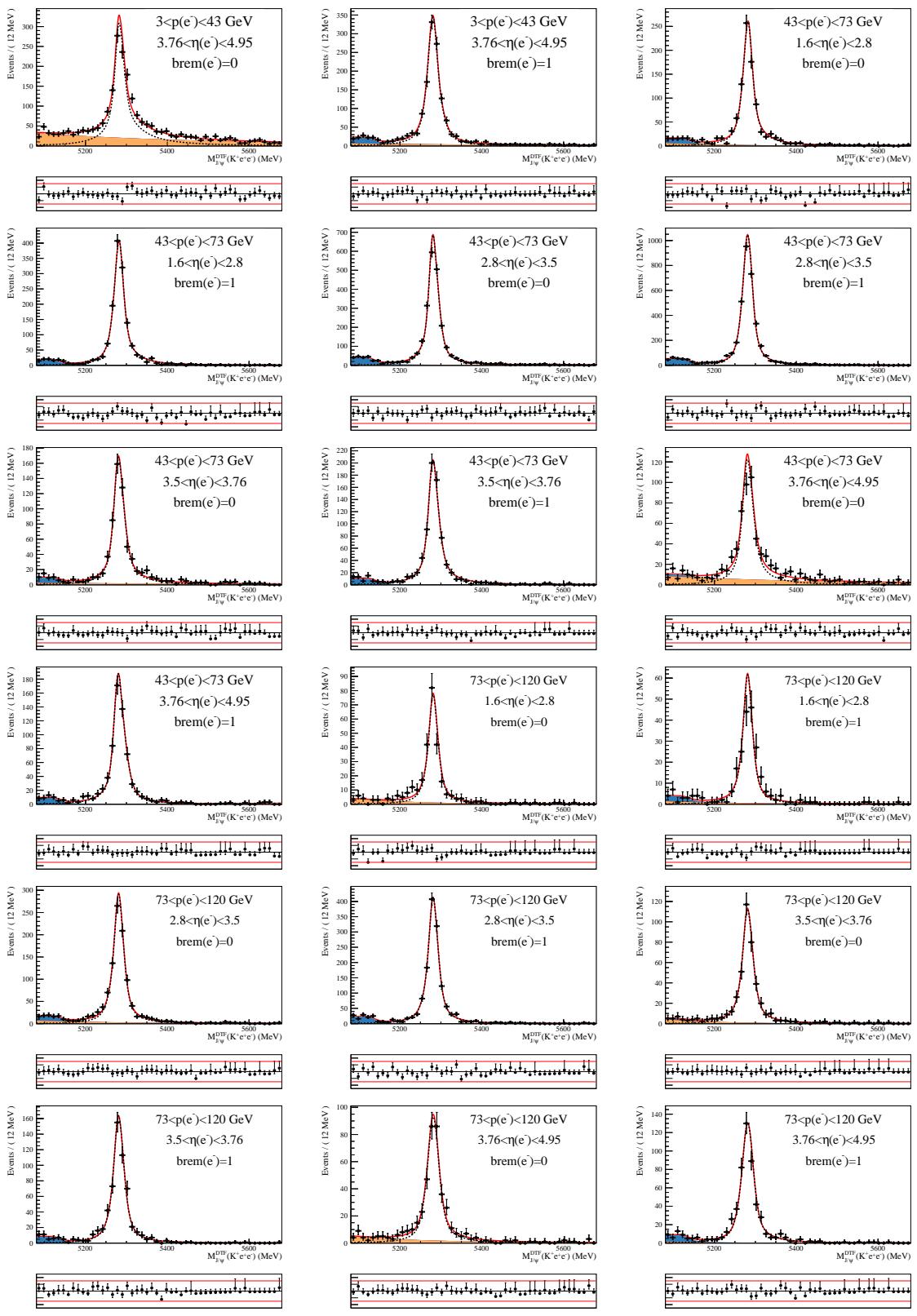


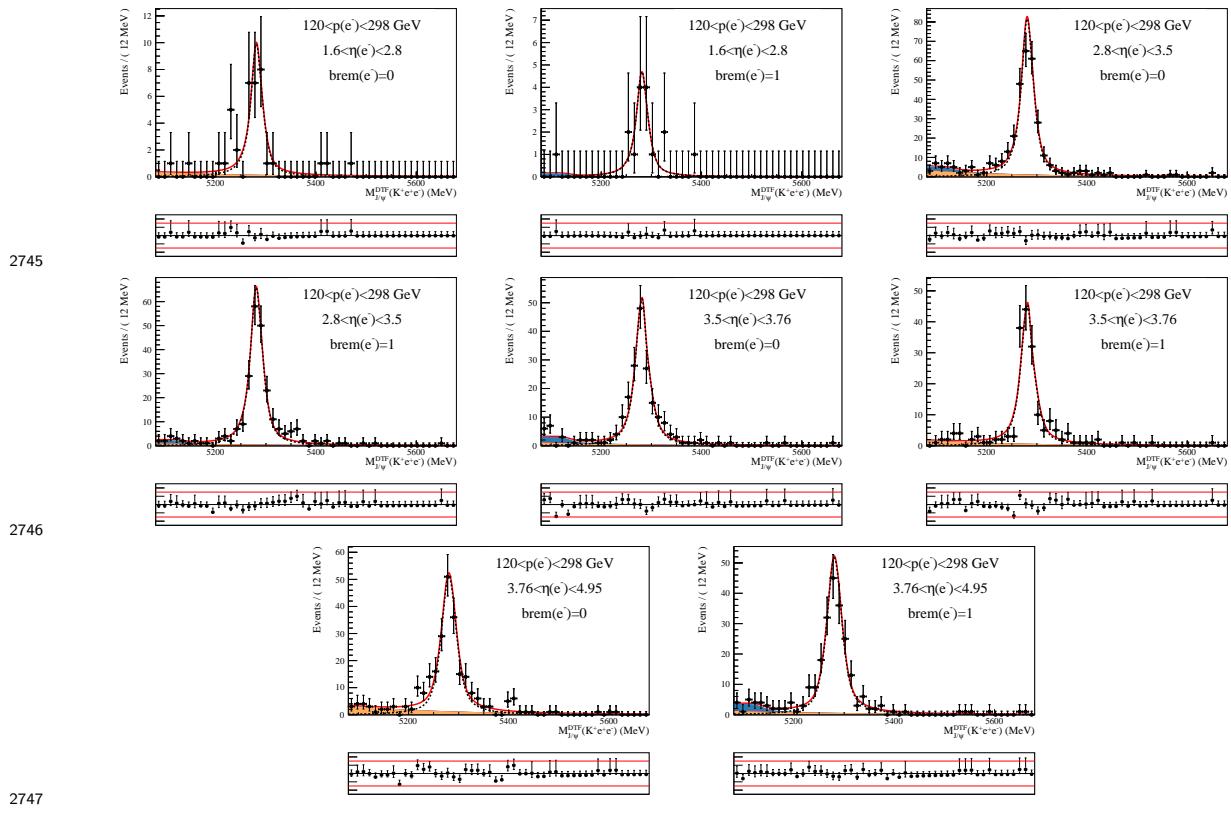


2735 D.2 2012 Fits

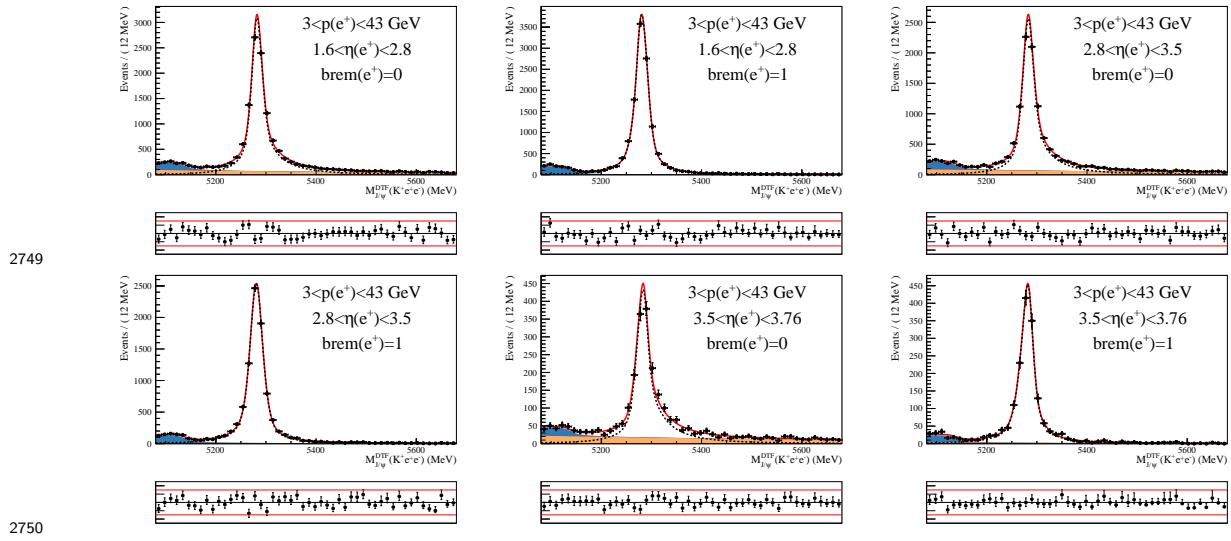
2736 D.2.1 2012 Fits, e^- probe, before PID cut

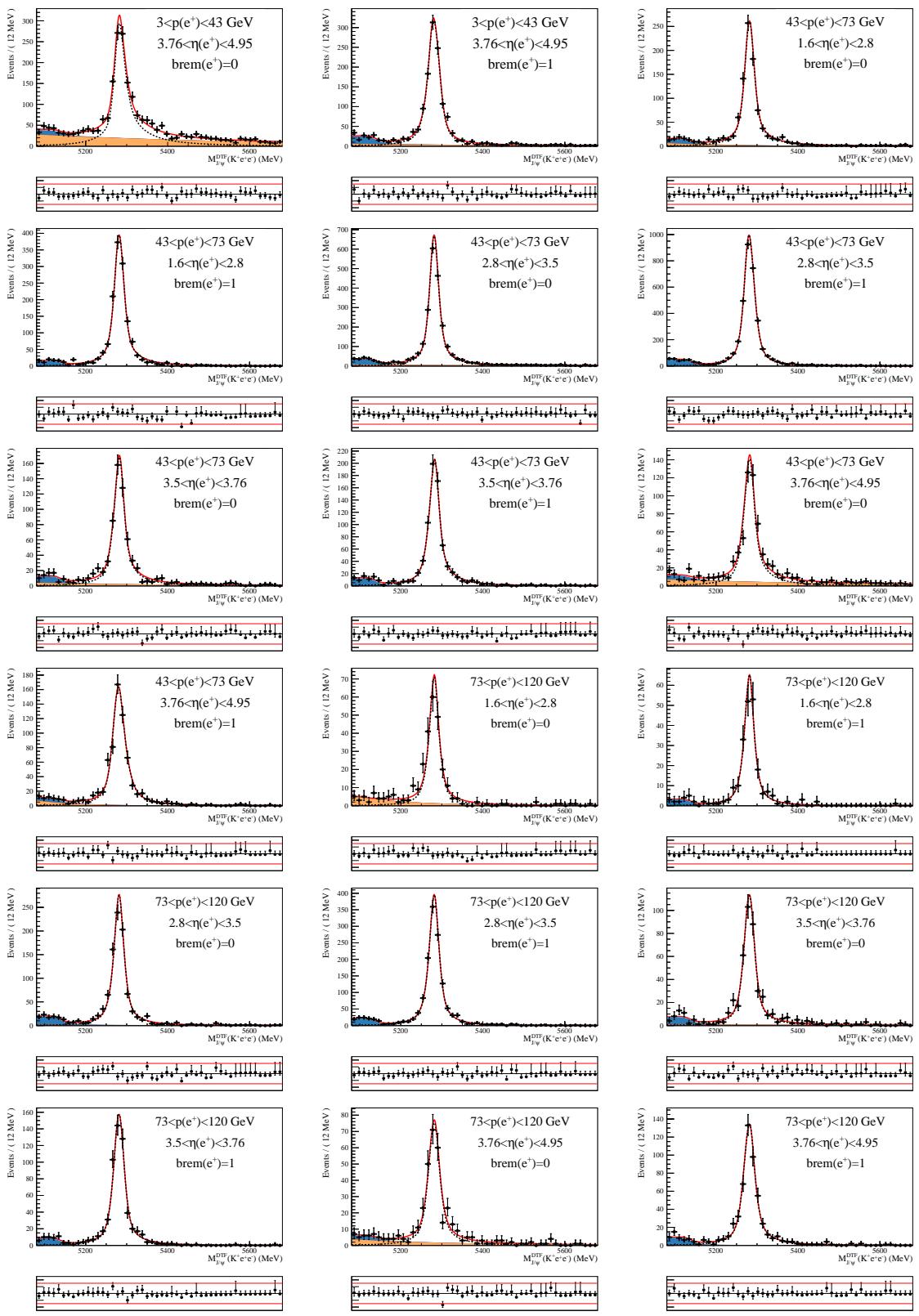


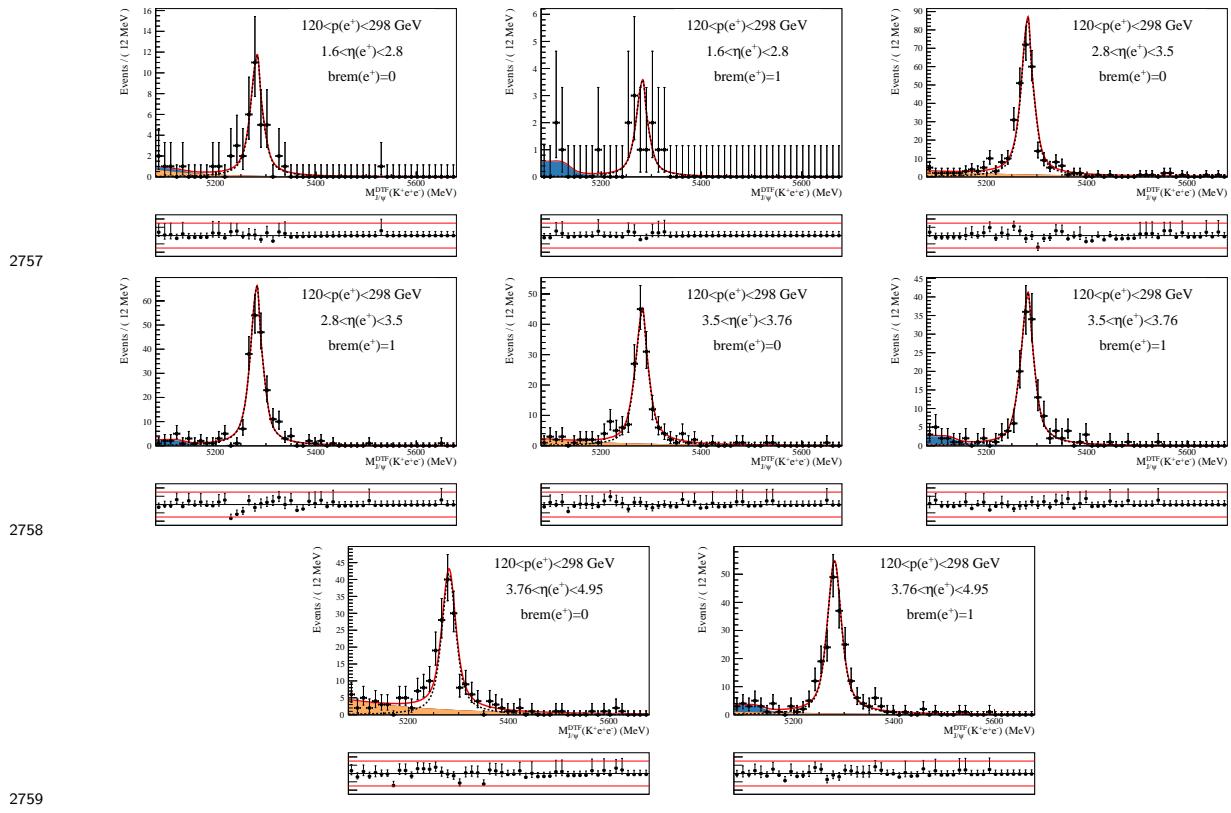




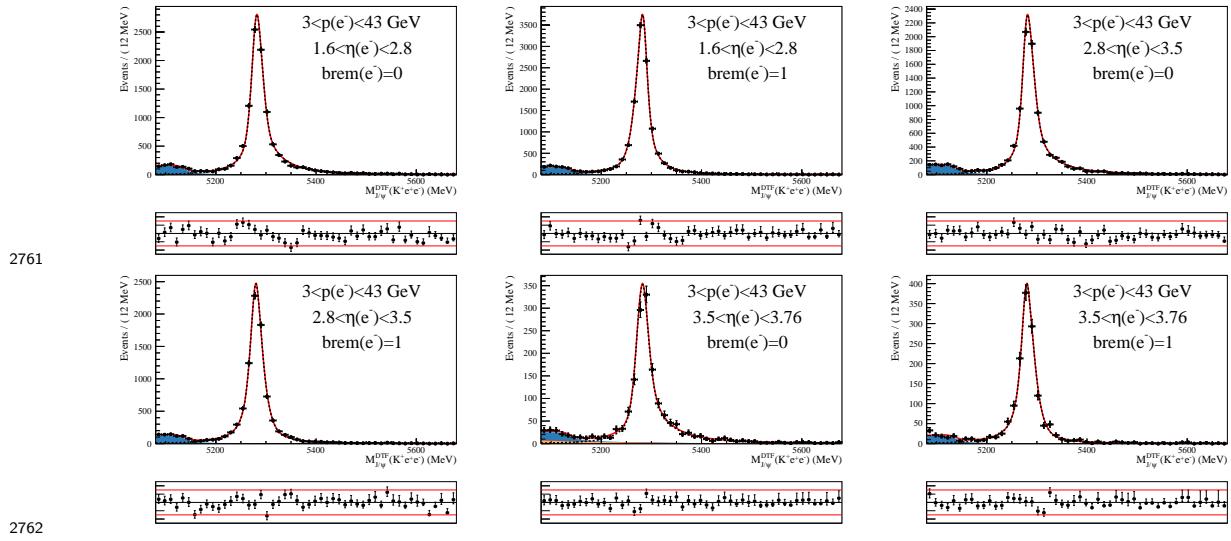
2748 D.2.2 2012 Fits, e^+ probe, before PID cut



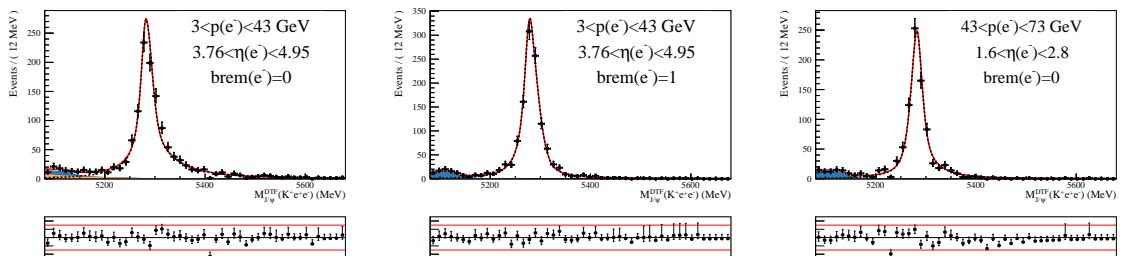




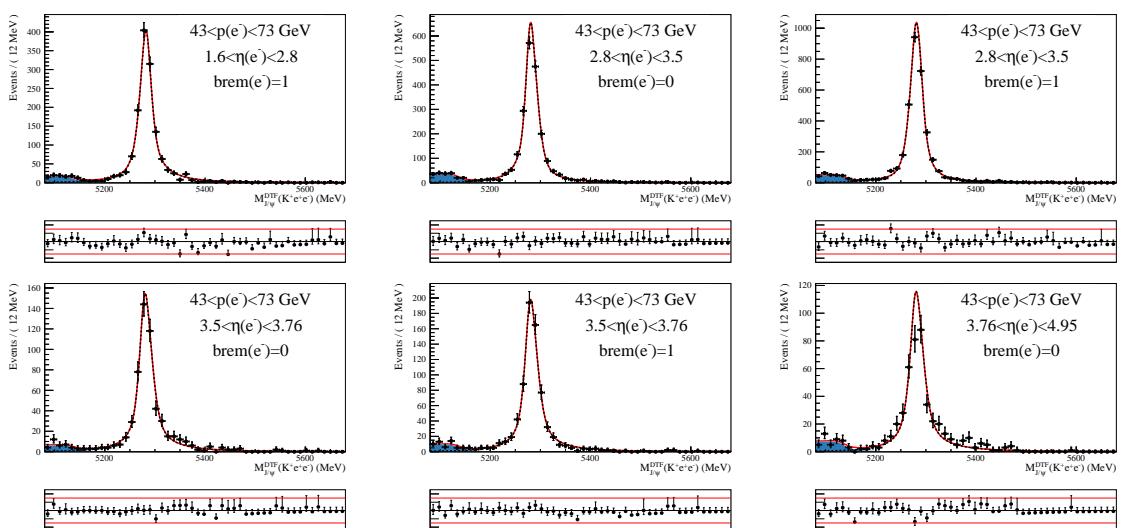
2760 D.2.3 2012 Fits, e^- probe, after PID cut



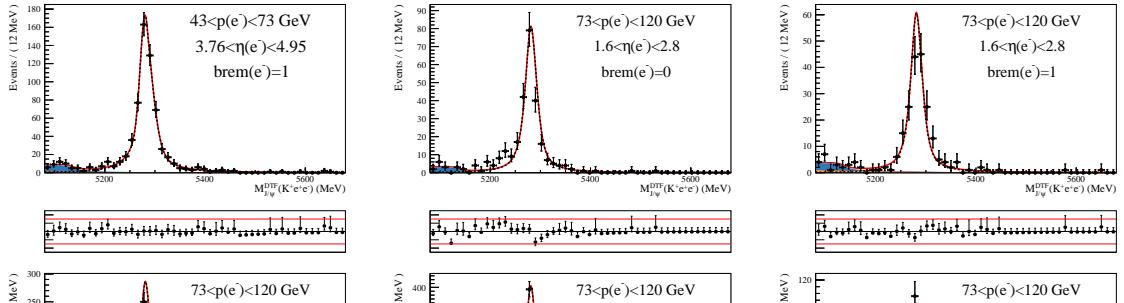
2763



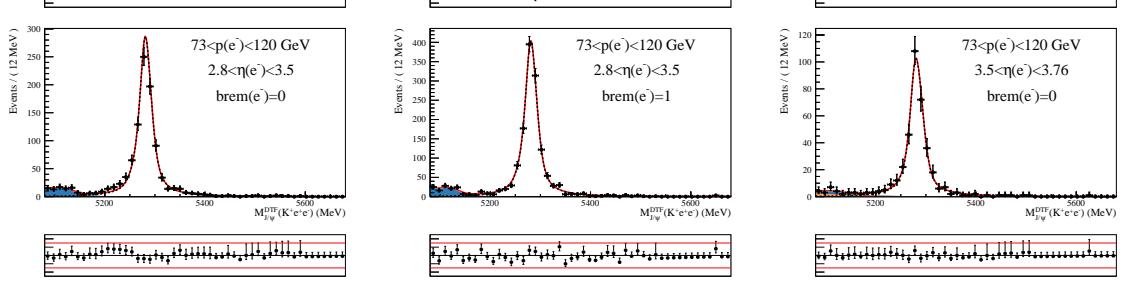
2764



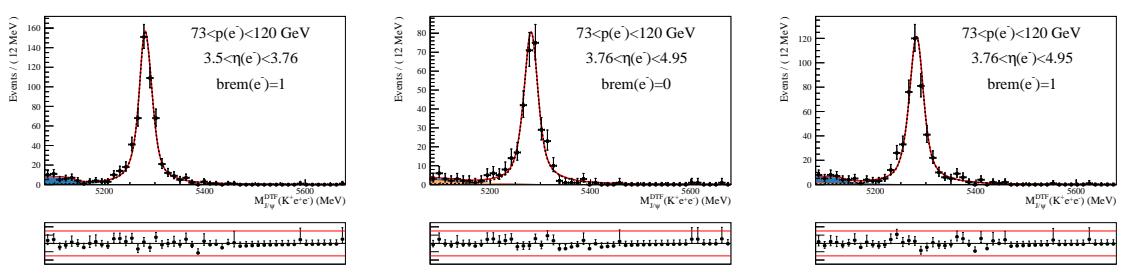
2765



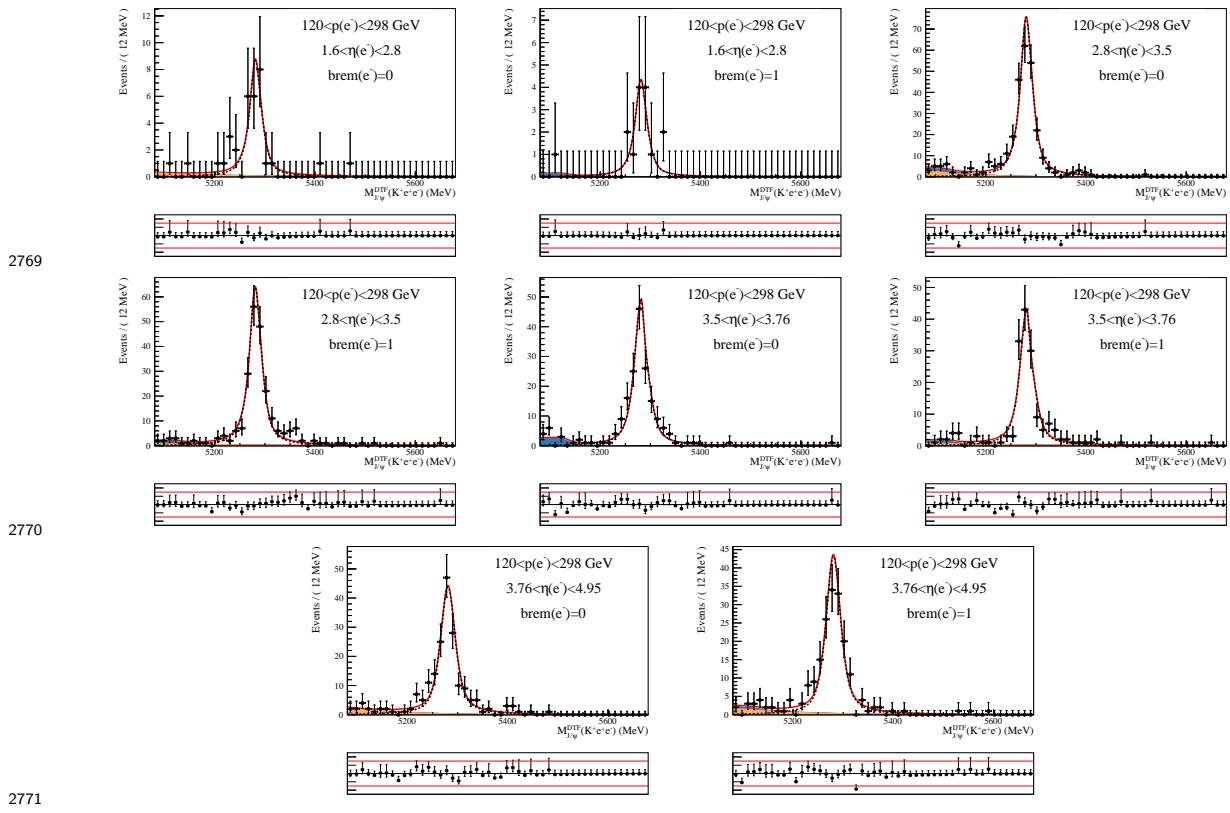
2766



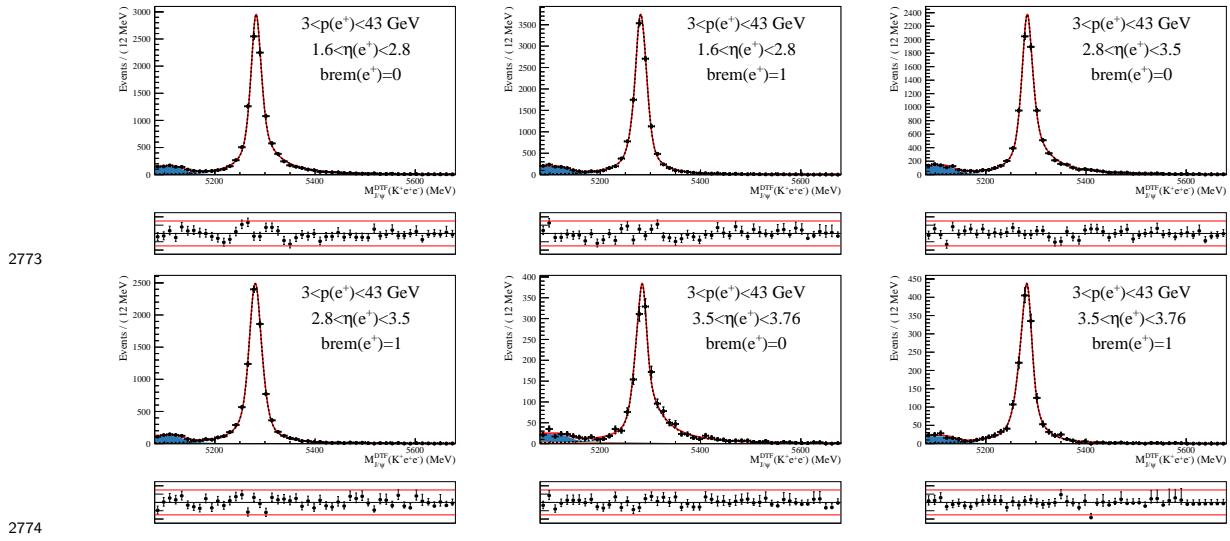
2767

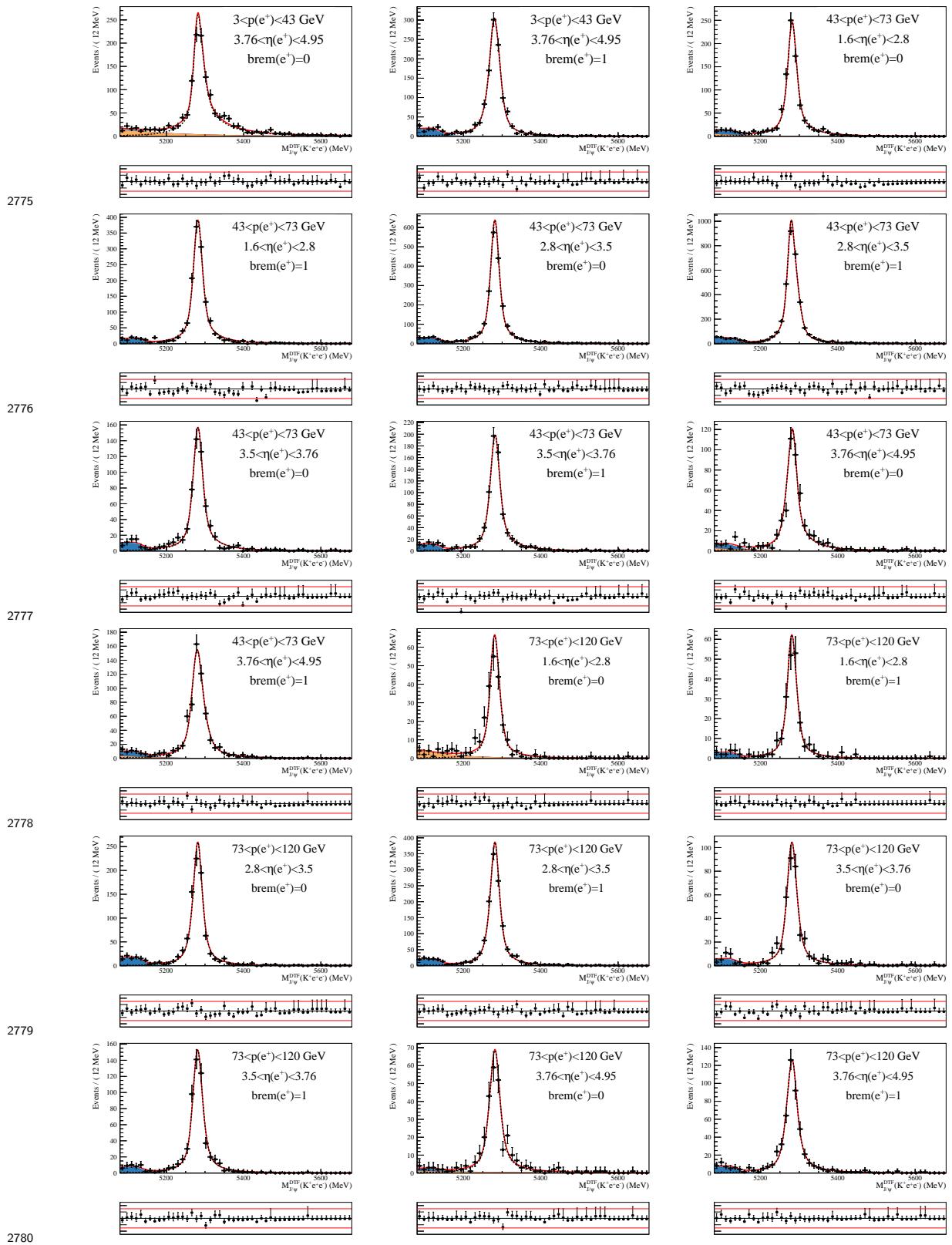


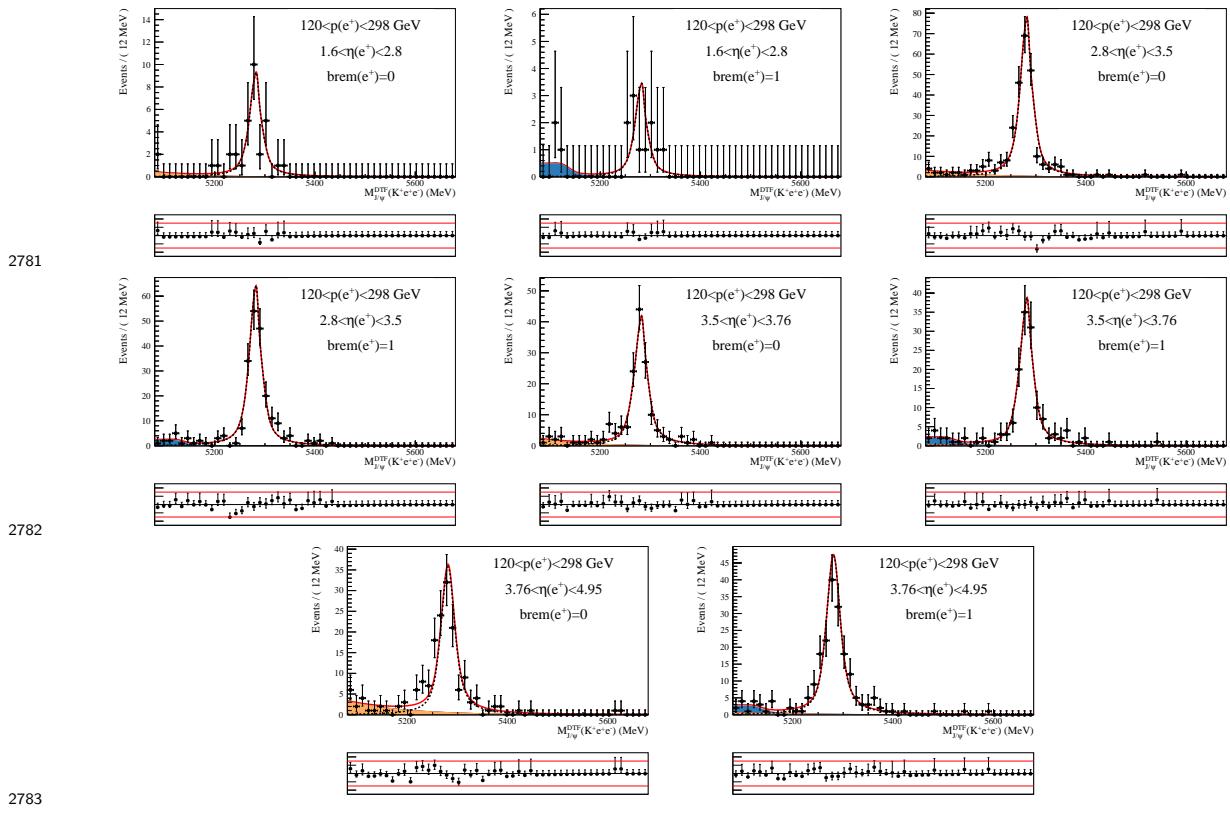
2768



2772 D.2.4 2012 Fits, e^+ probe, after PID cut

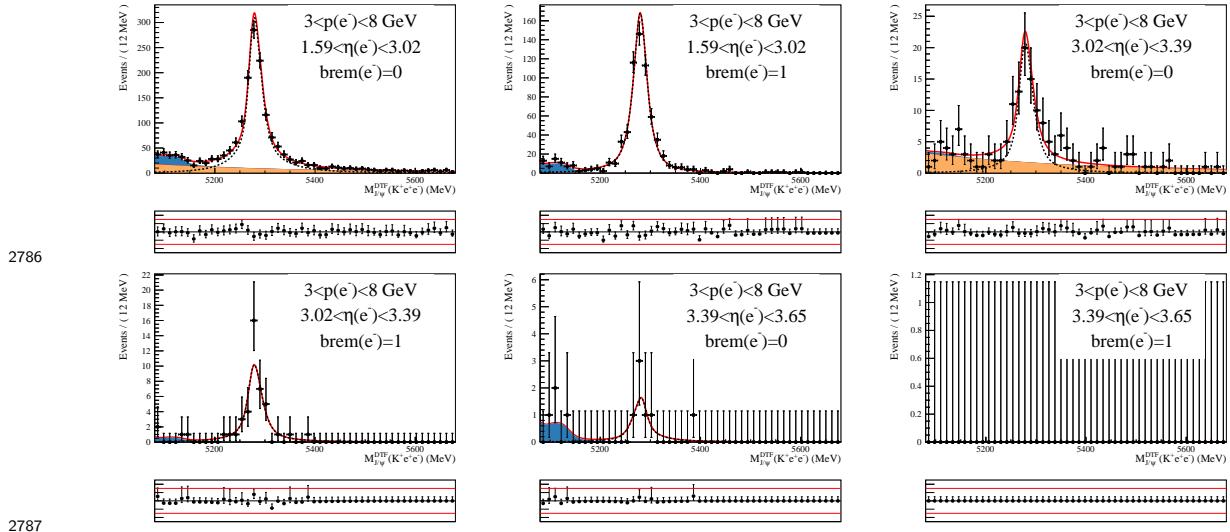


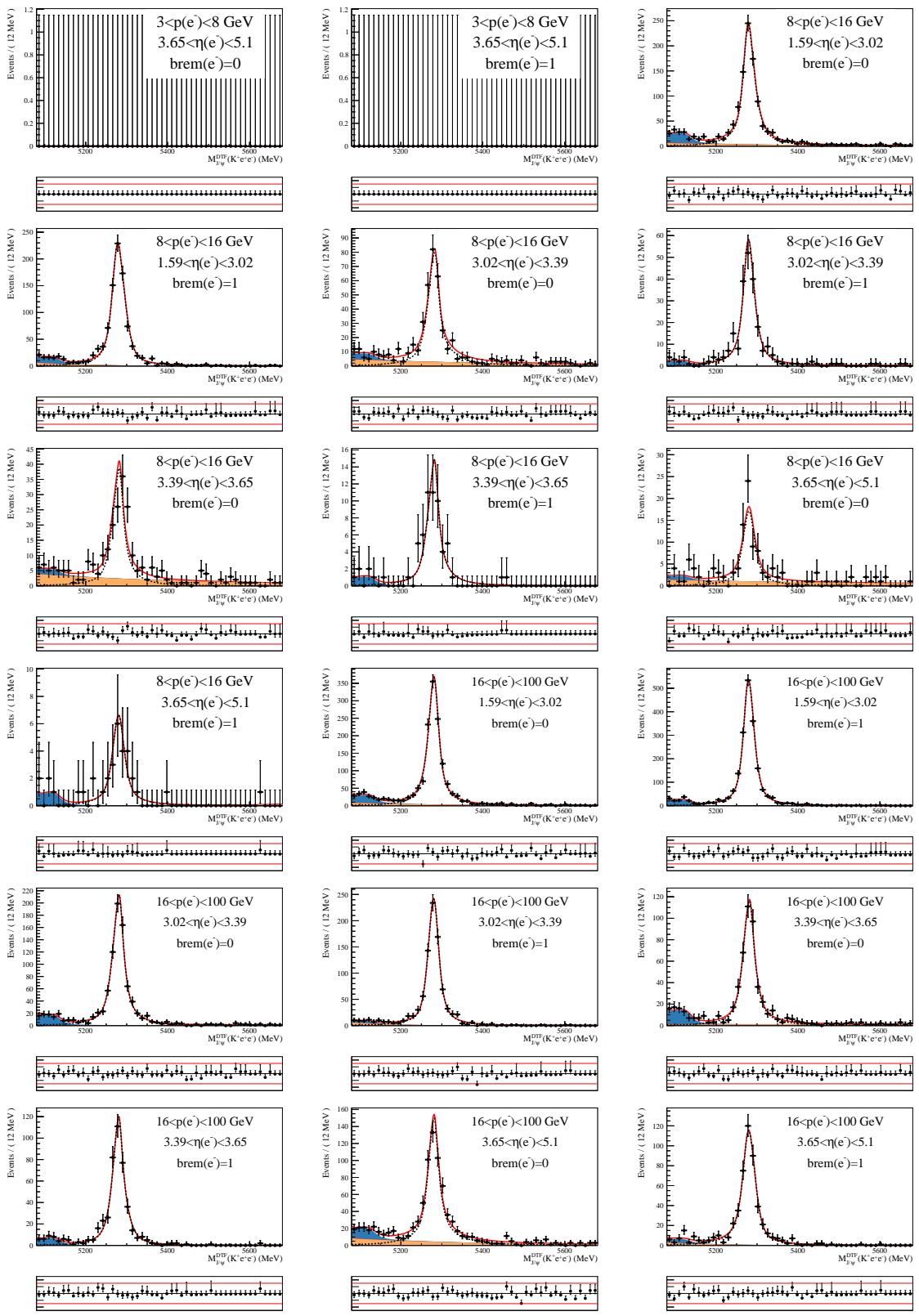


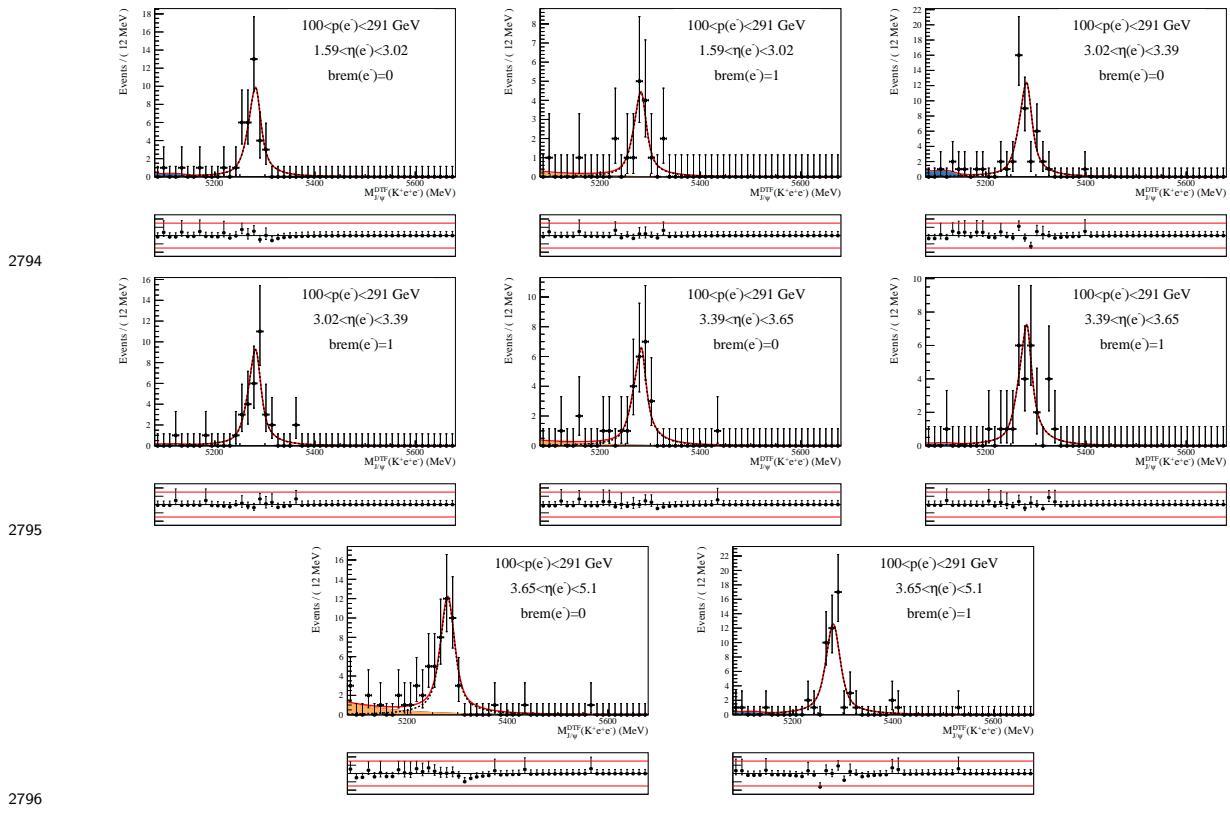


2784 D.3 2015 Fits

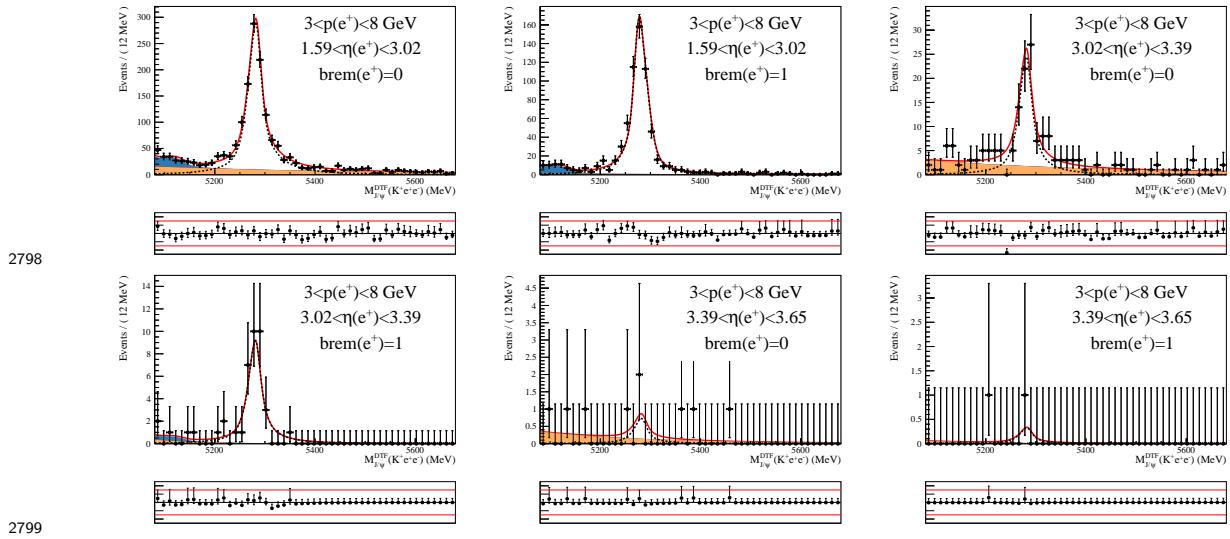
2785 D.3.1 2015 Fits, e^- probe, before PID cut



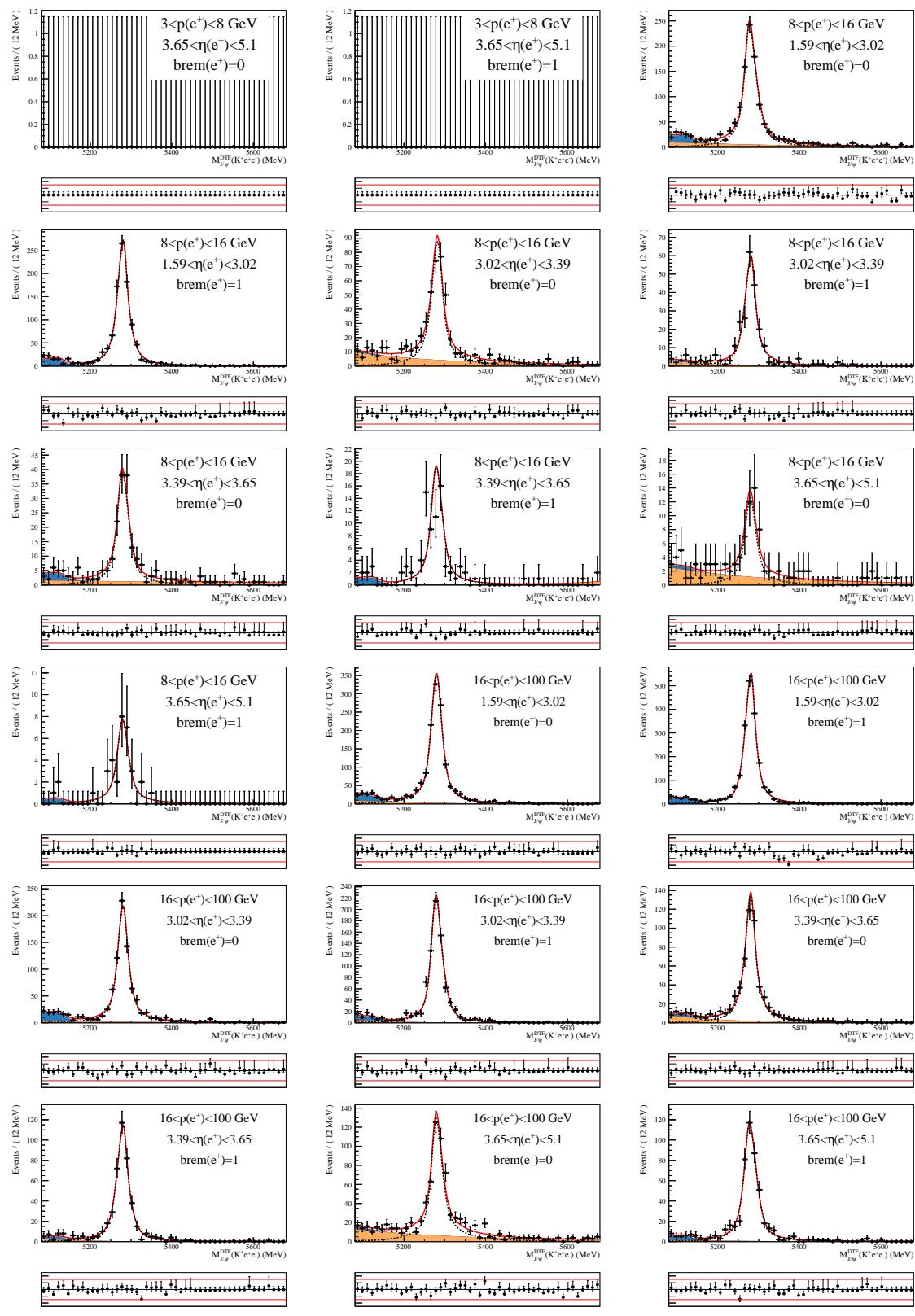




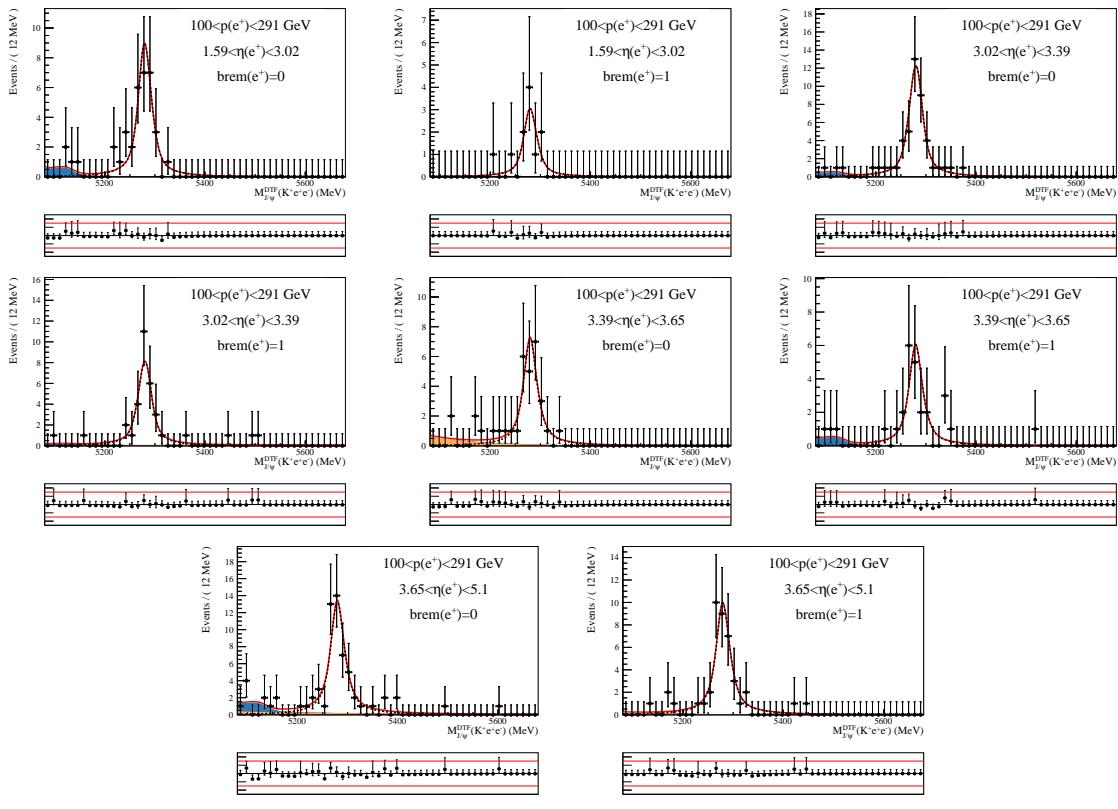
2797 D.3.2 2015 Fits, e^+ probe, before PID cut



2800



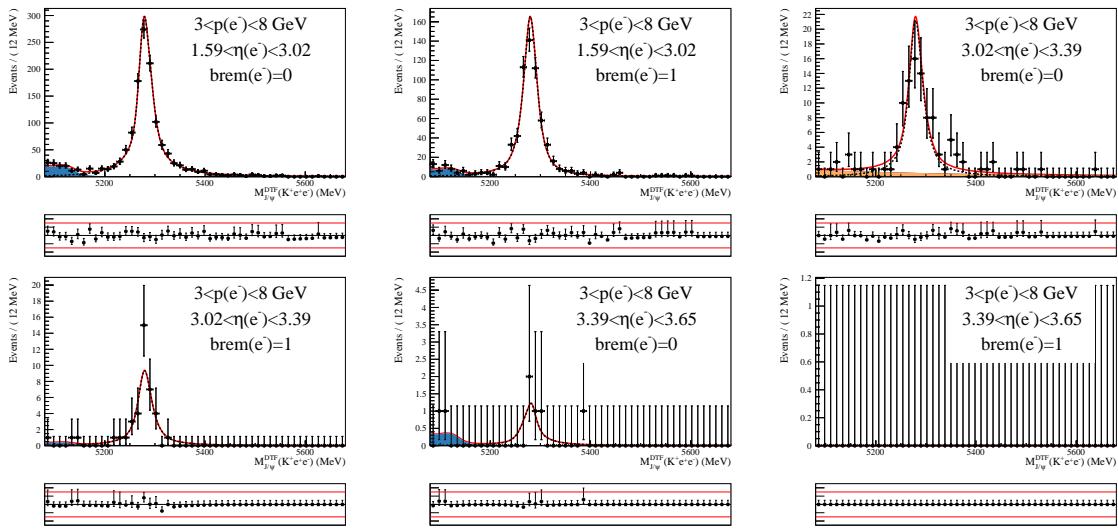
2806



2808

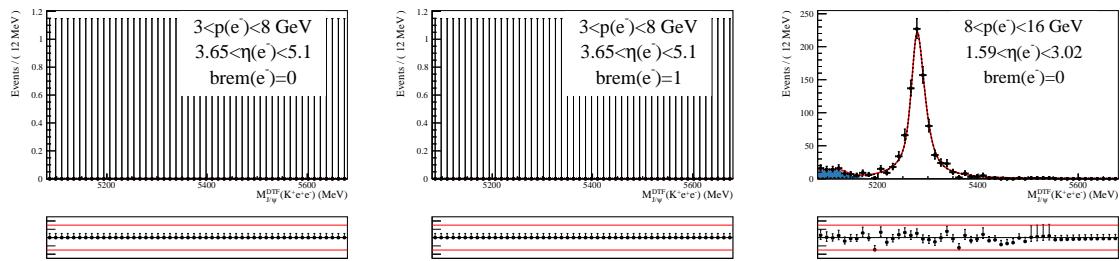
2809 D.3.3 2015 Fits, e^- probe, after PID cut

2810

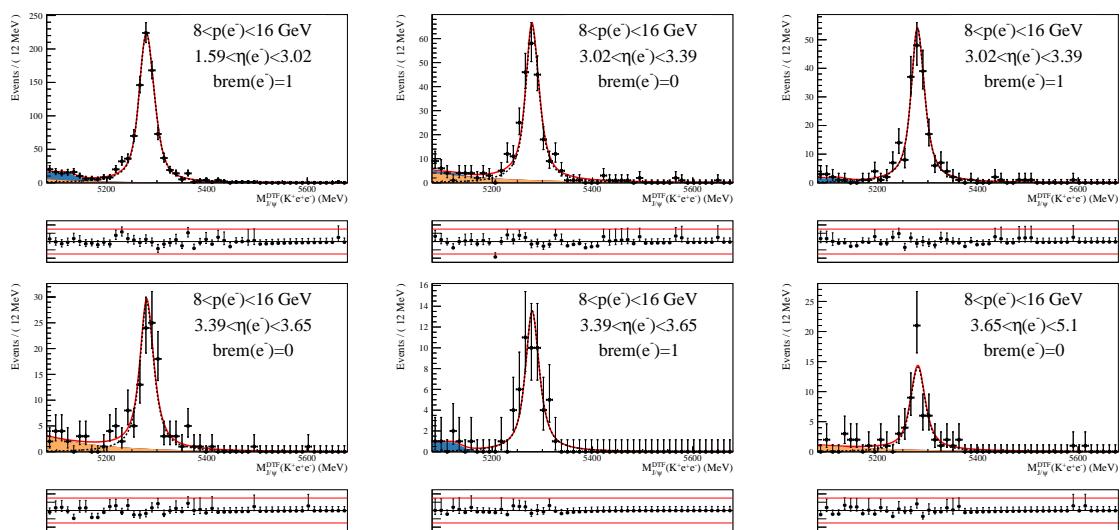


2811

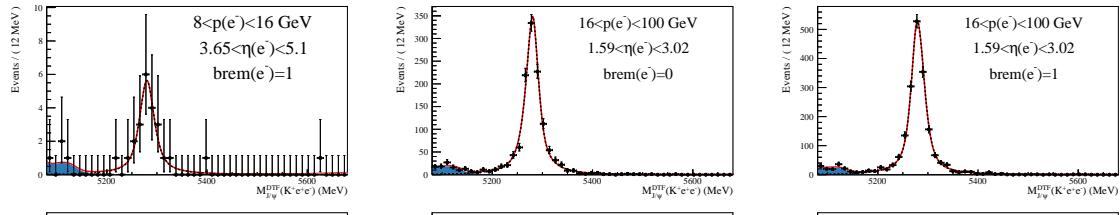
2812



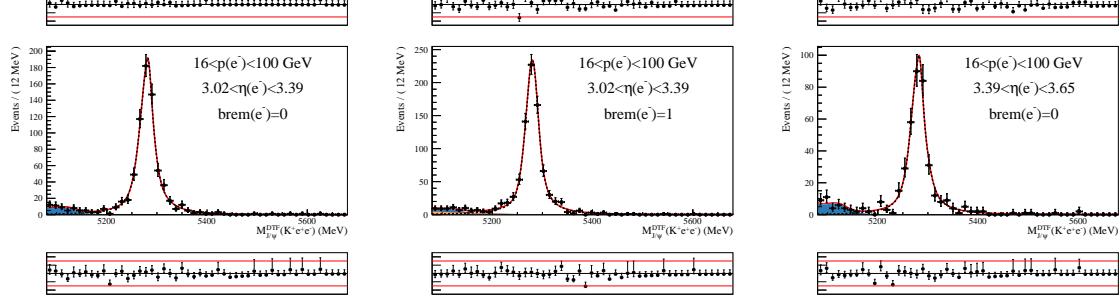
2813



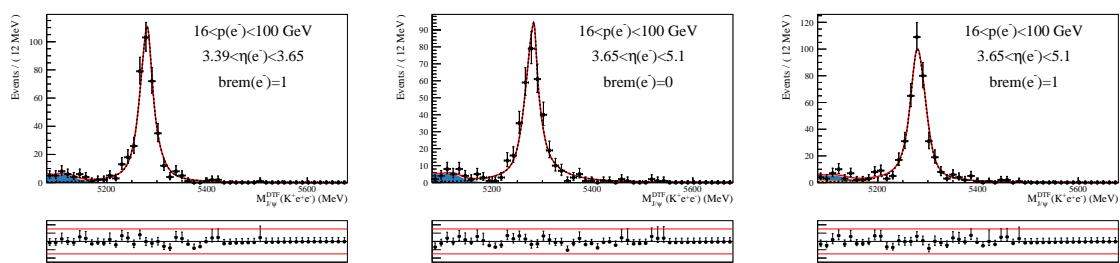
2814



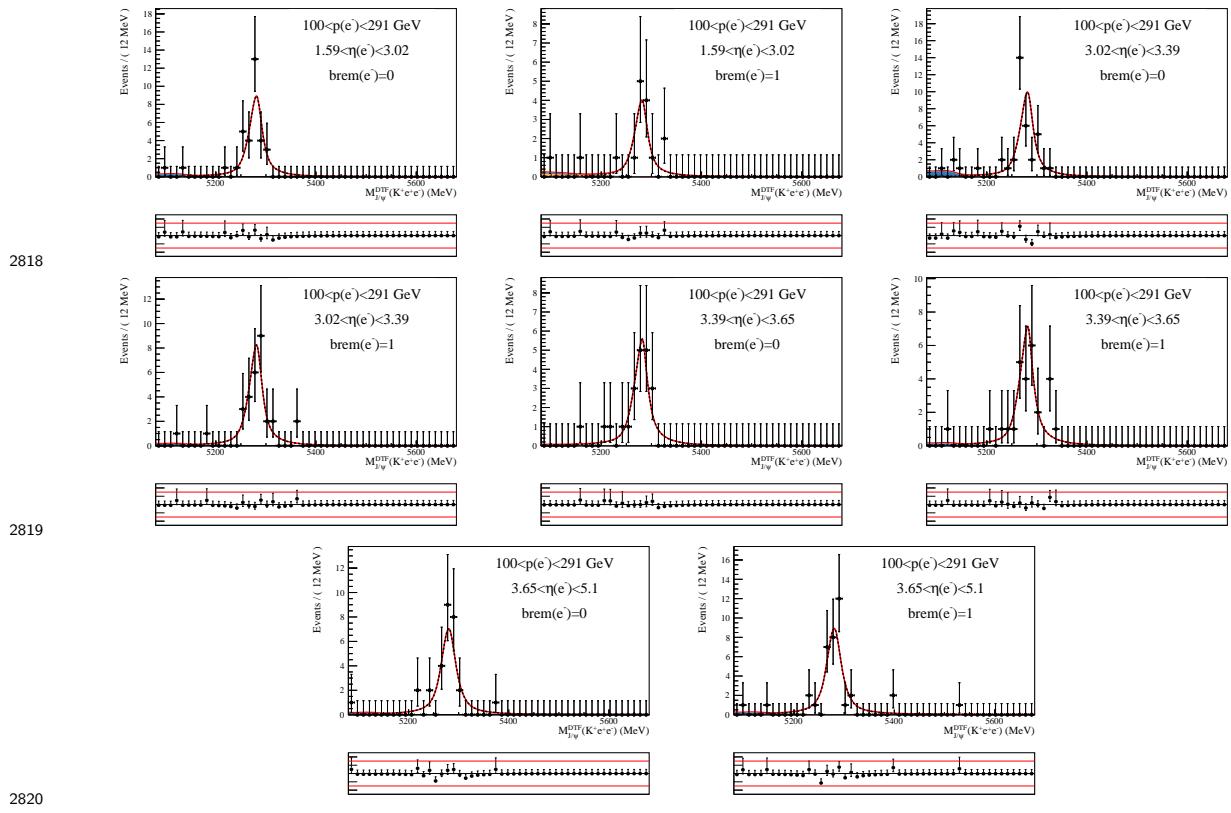
2815



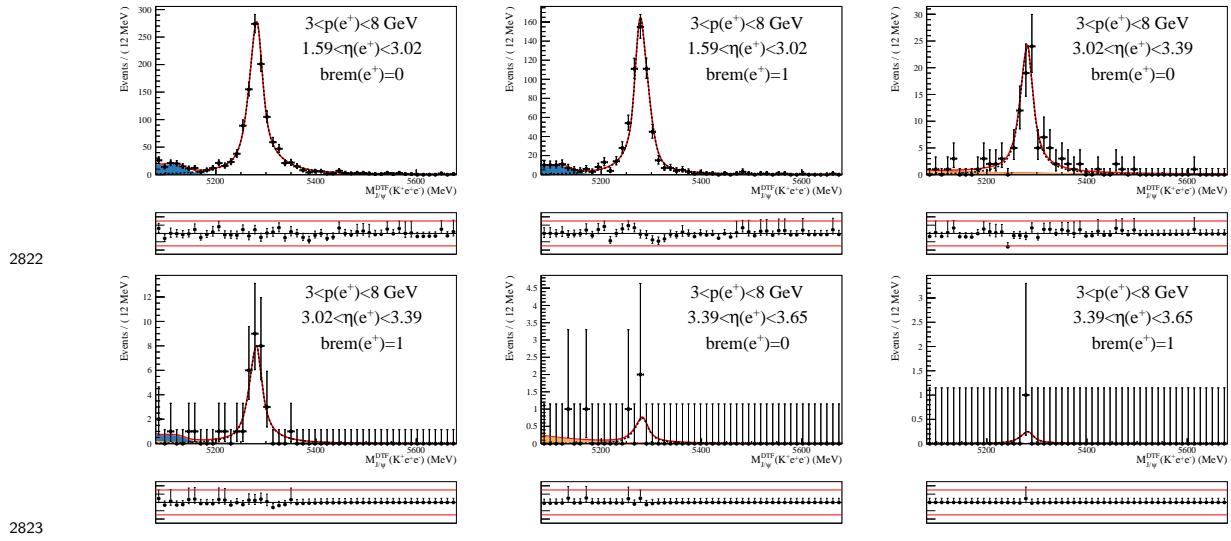
2816



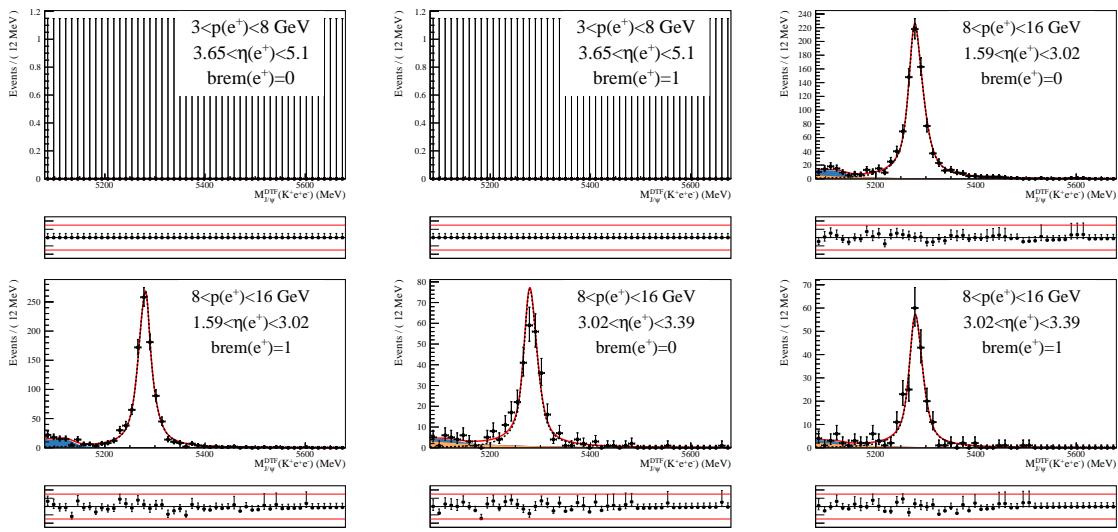
2817



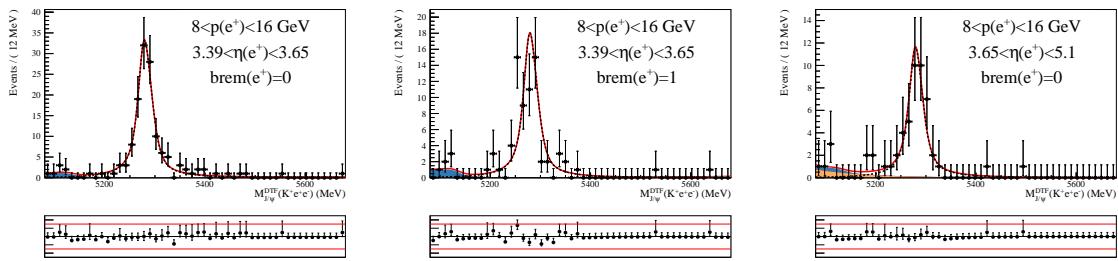
2821 D.3.4 2015 Fits, e^+ probe, after PID cut



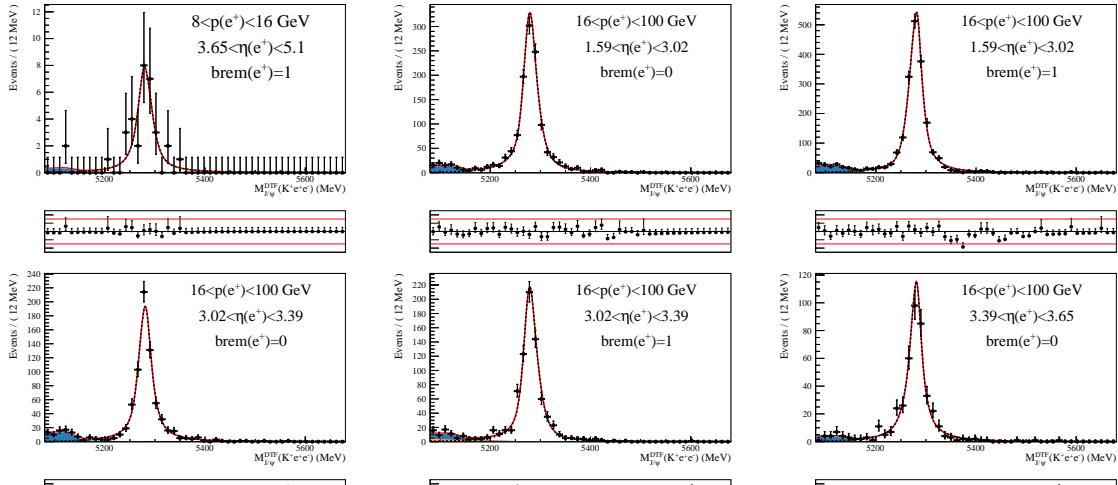
2824



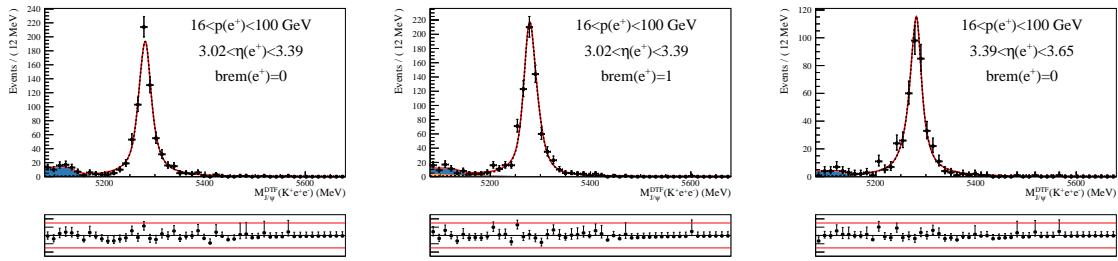
2825



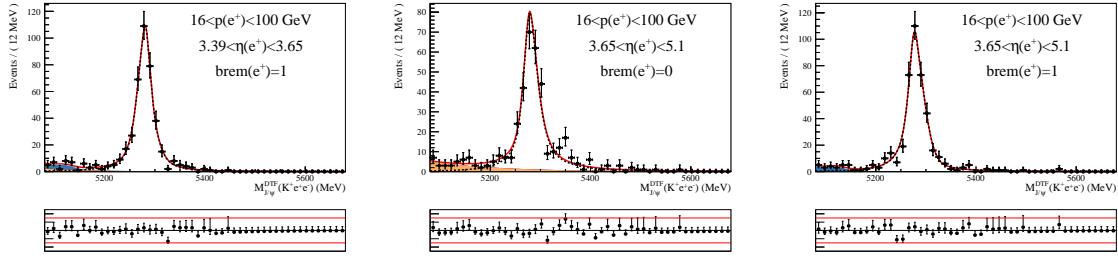
2826



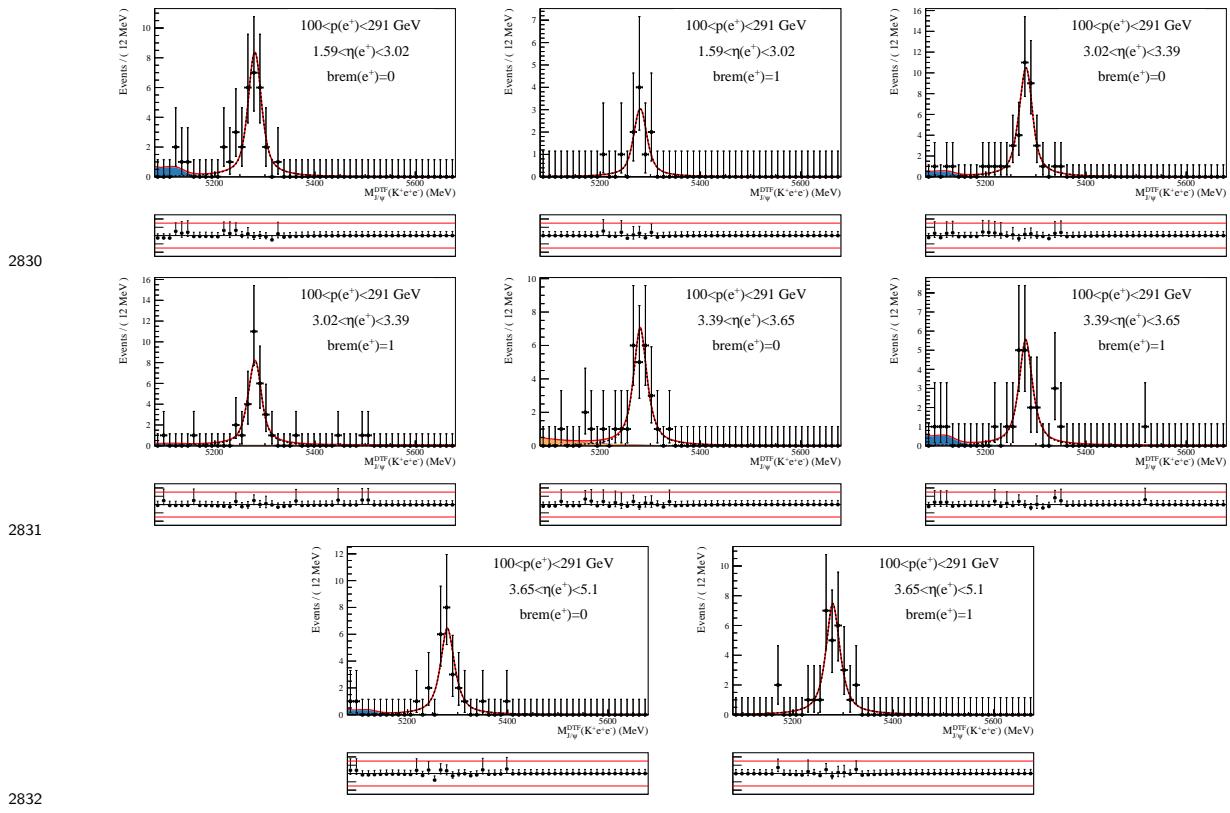
2827



2828

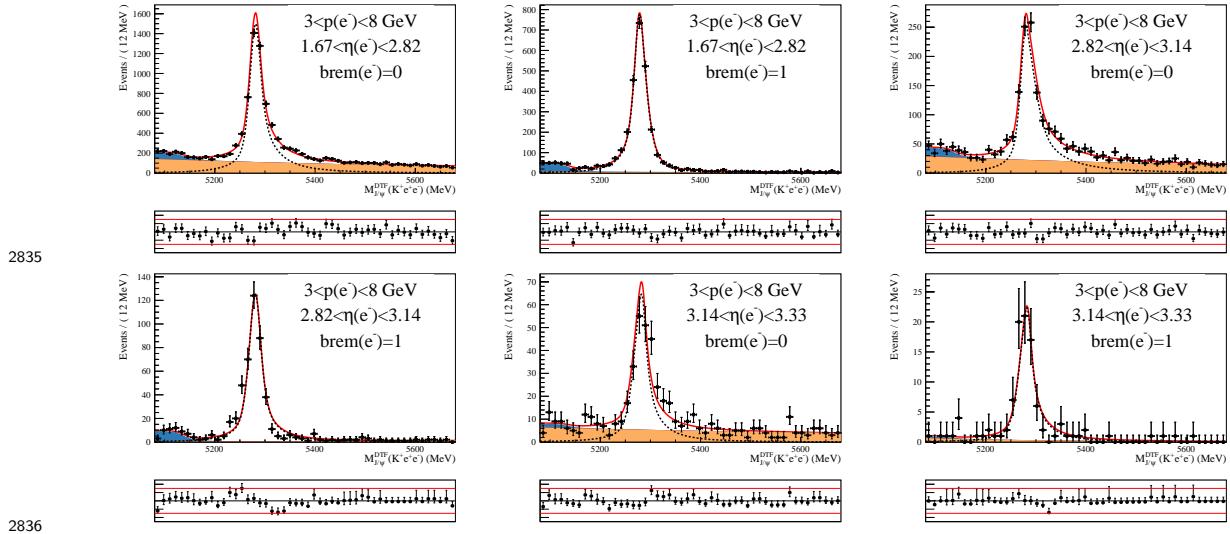


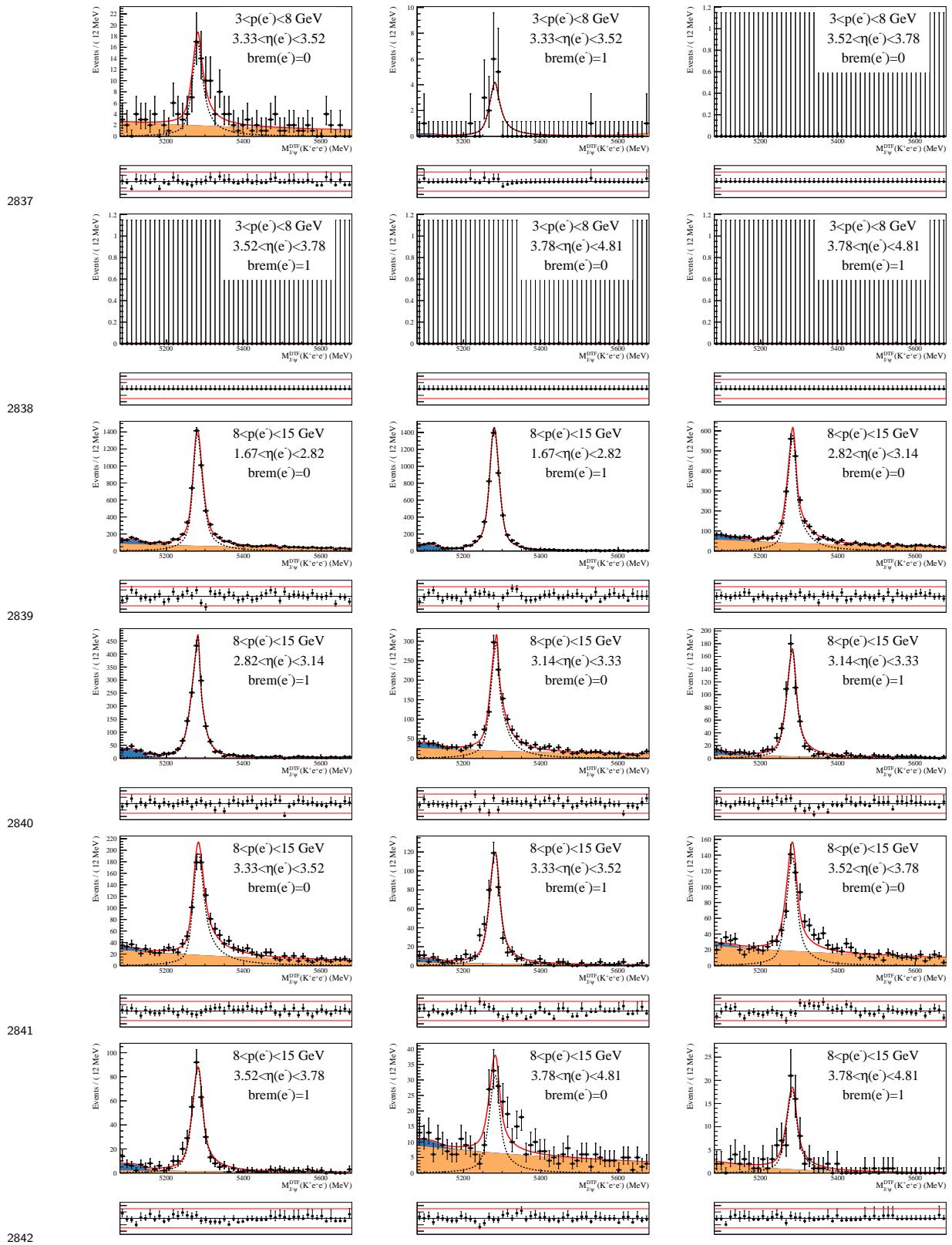
2829

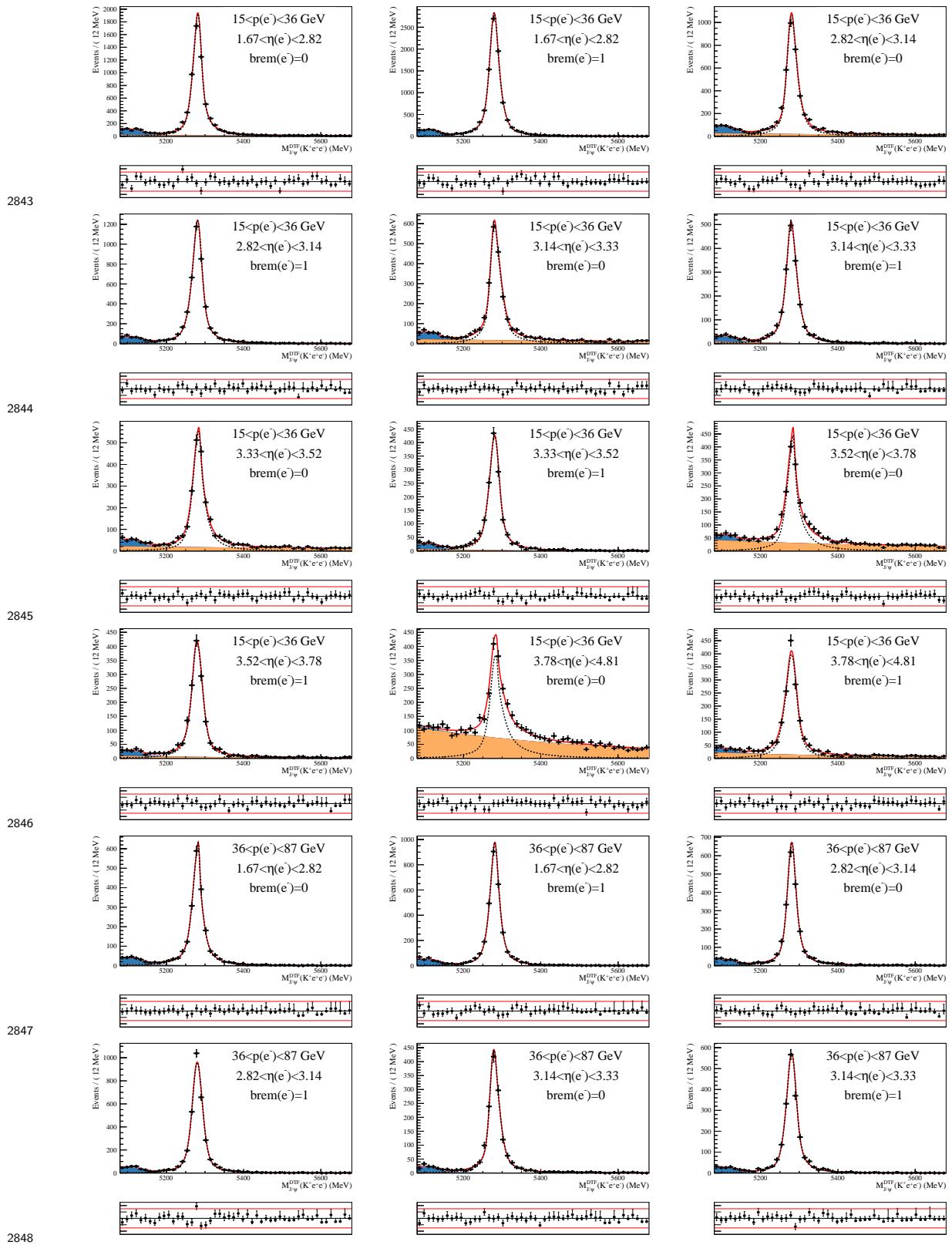


2833 D.4 2016 Fits

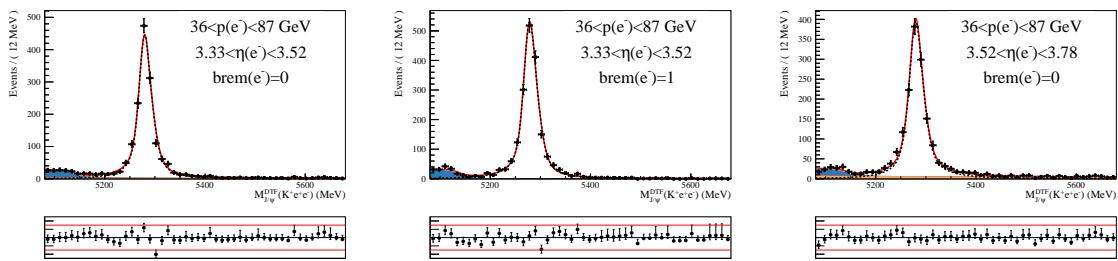
2834 D.4.1 2016 Fits, e^- probe, before PID cut



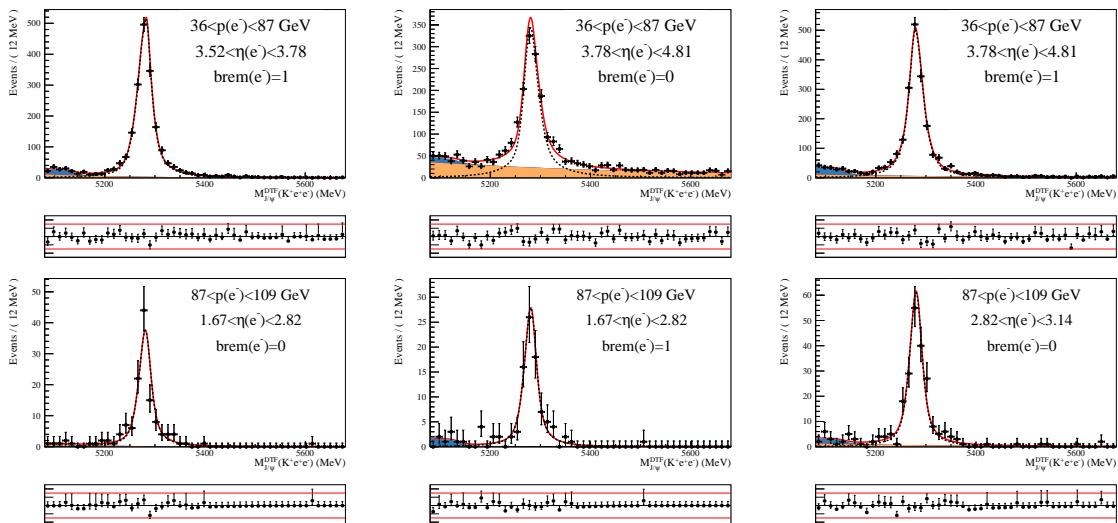




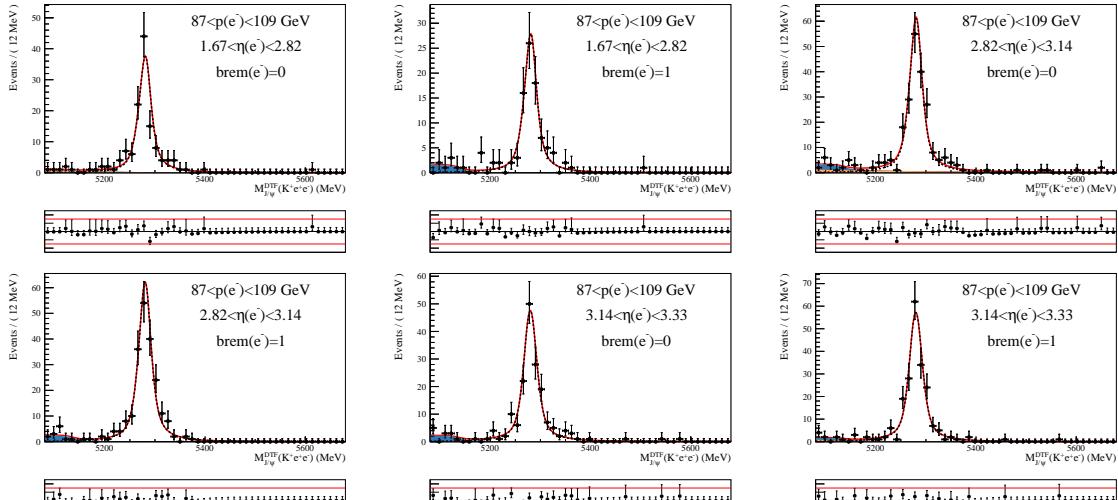
2849



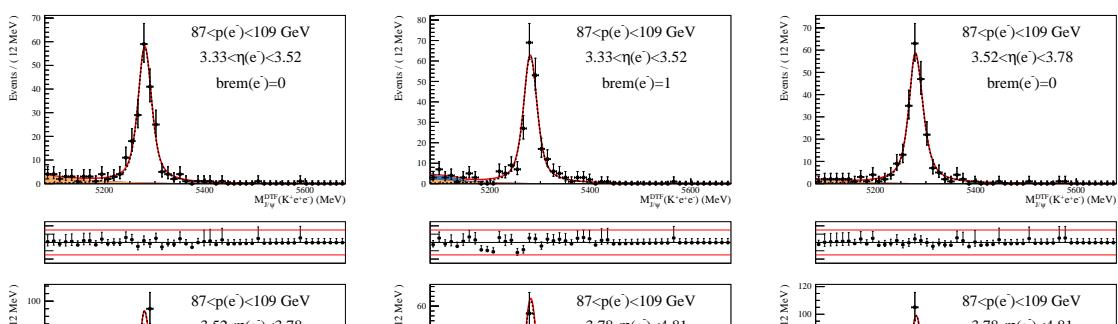
2850



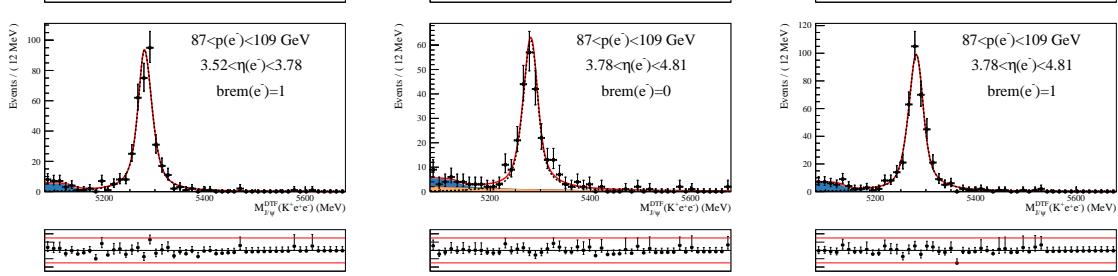
2851



2852

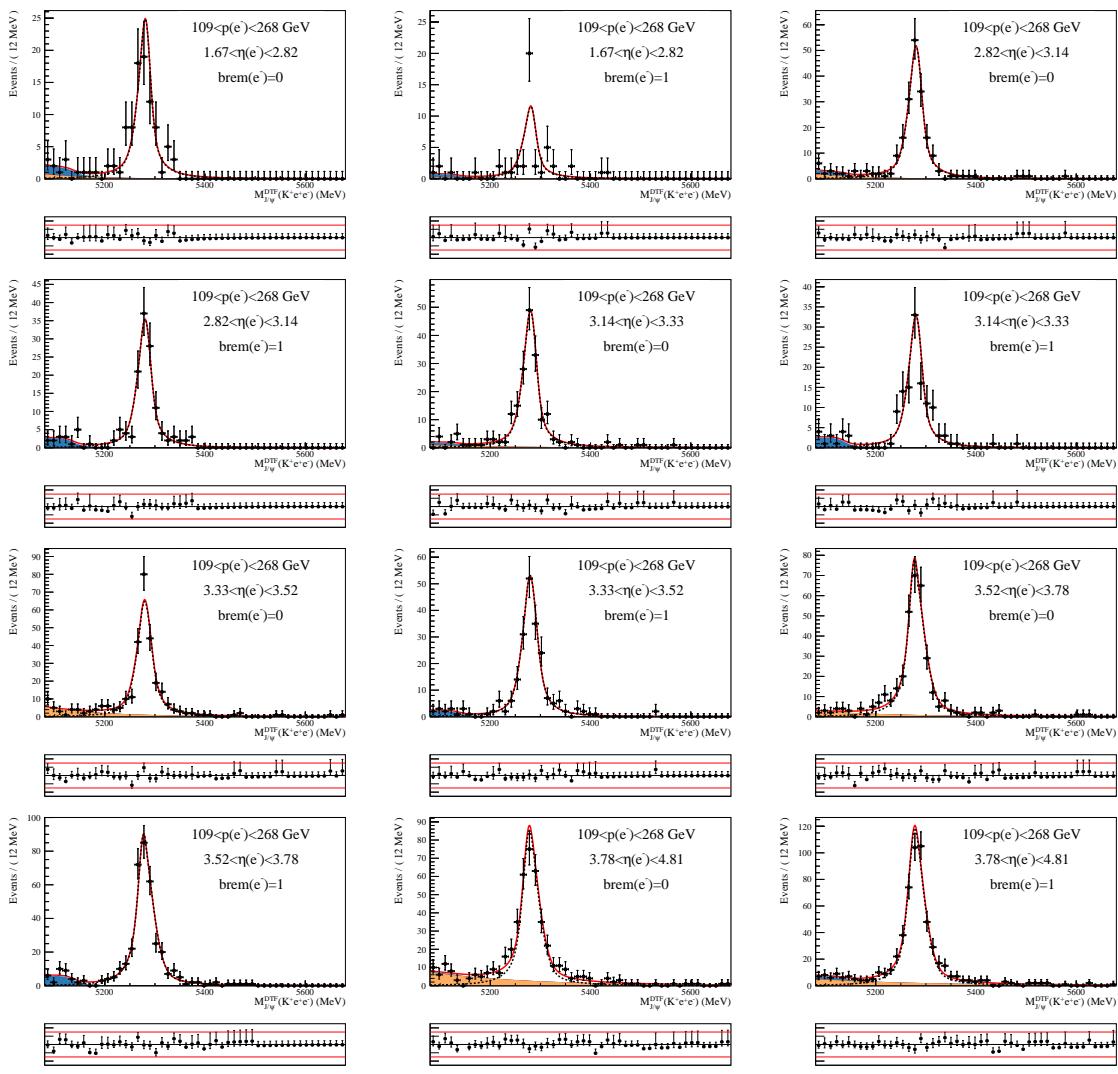


2853



2854

2855



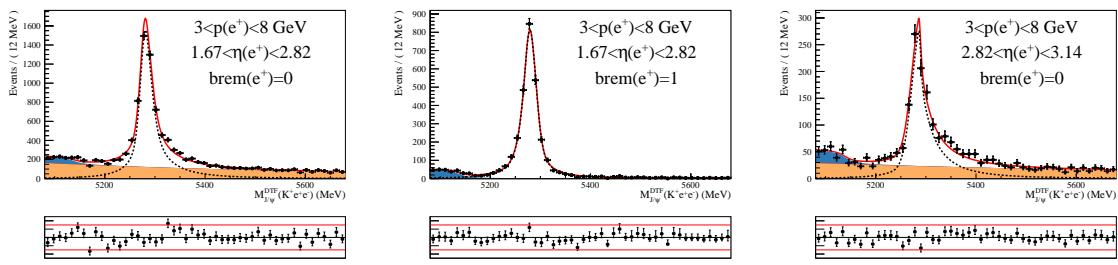
2857

2858

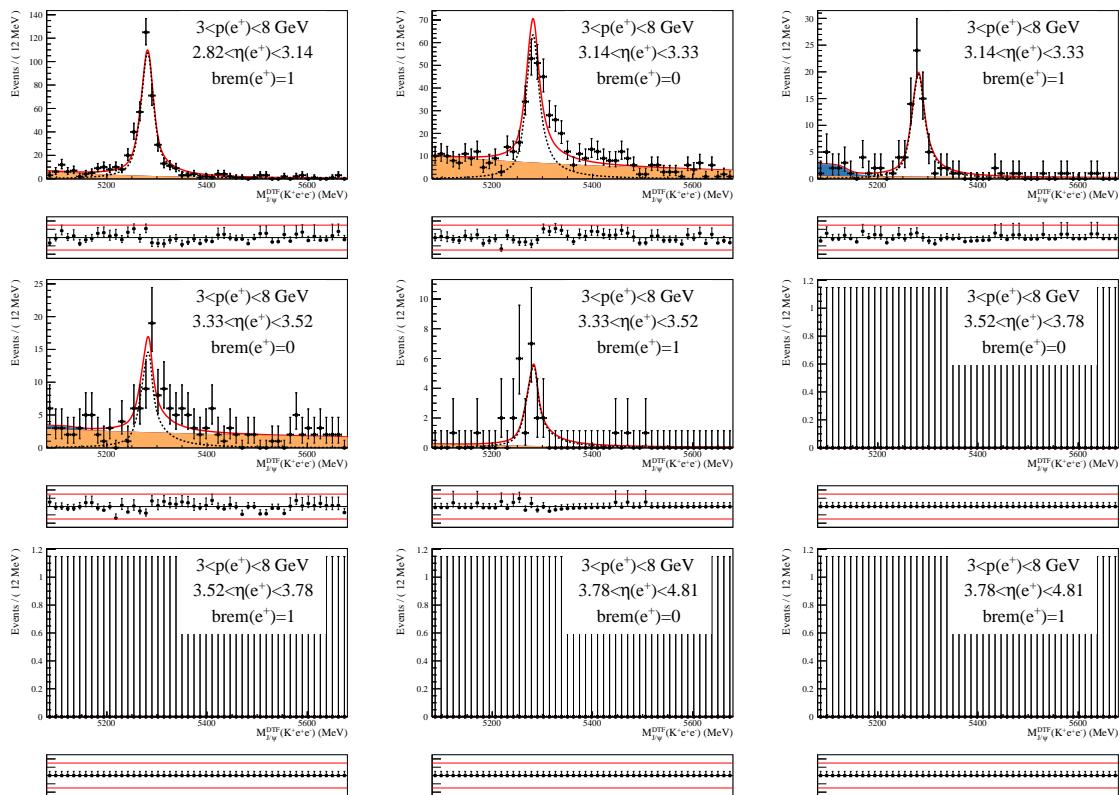
2859

D.4.2 2016 Fits, e^+ probe, before PID cut

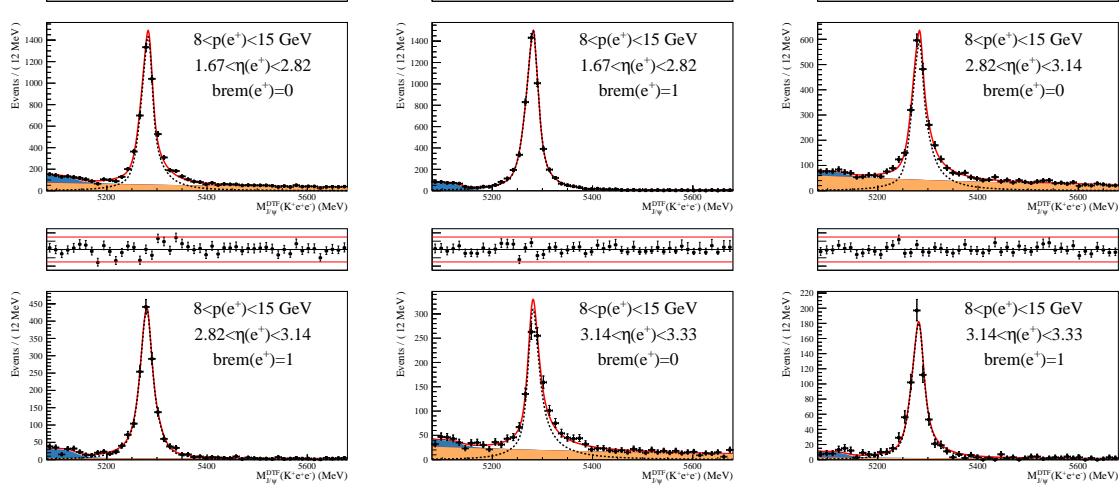
2860



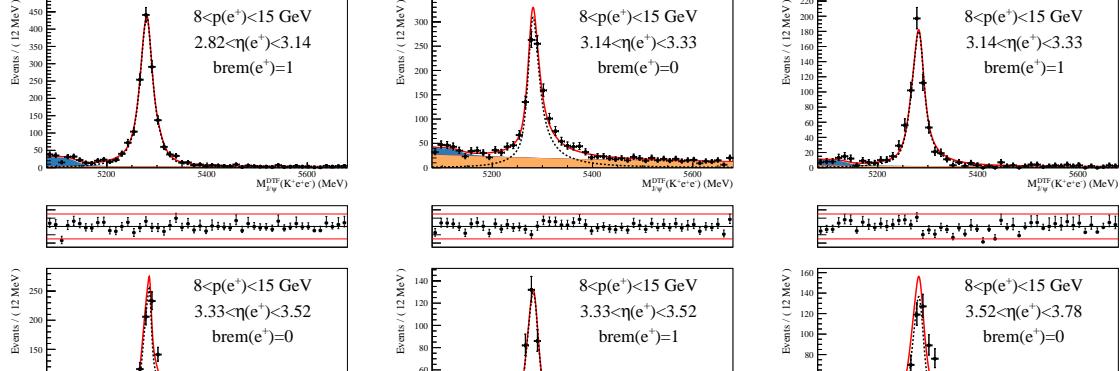
2861



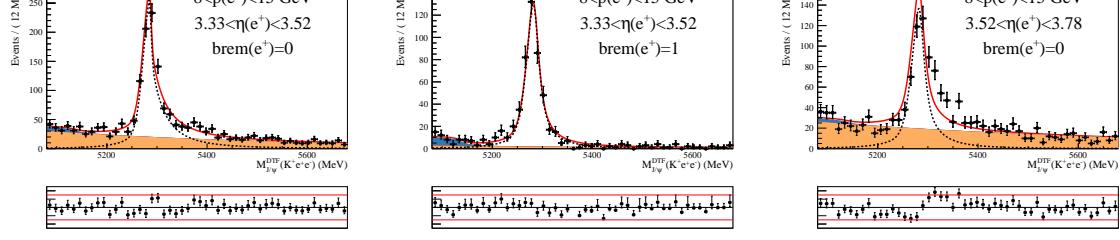
2862



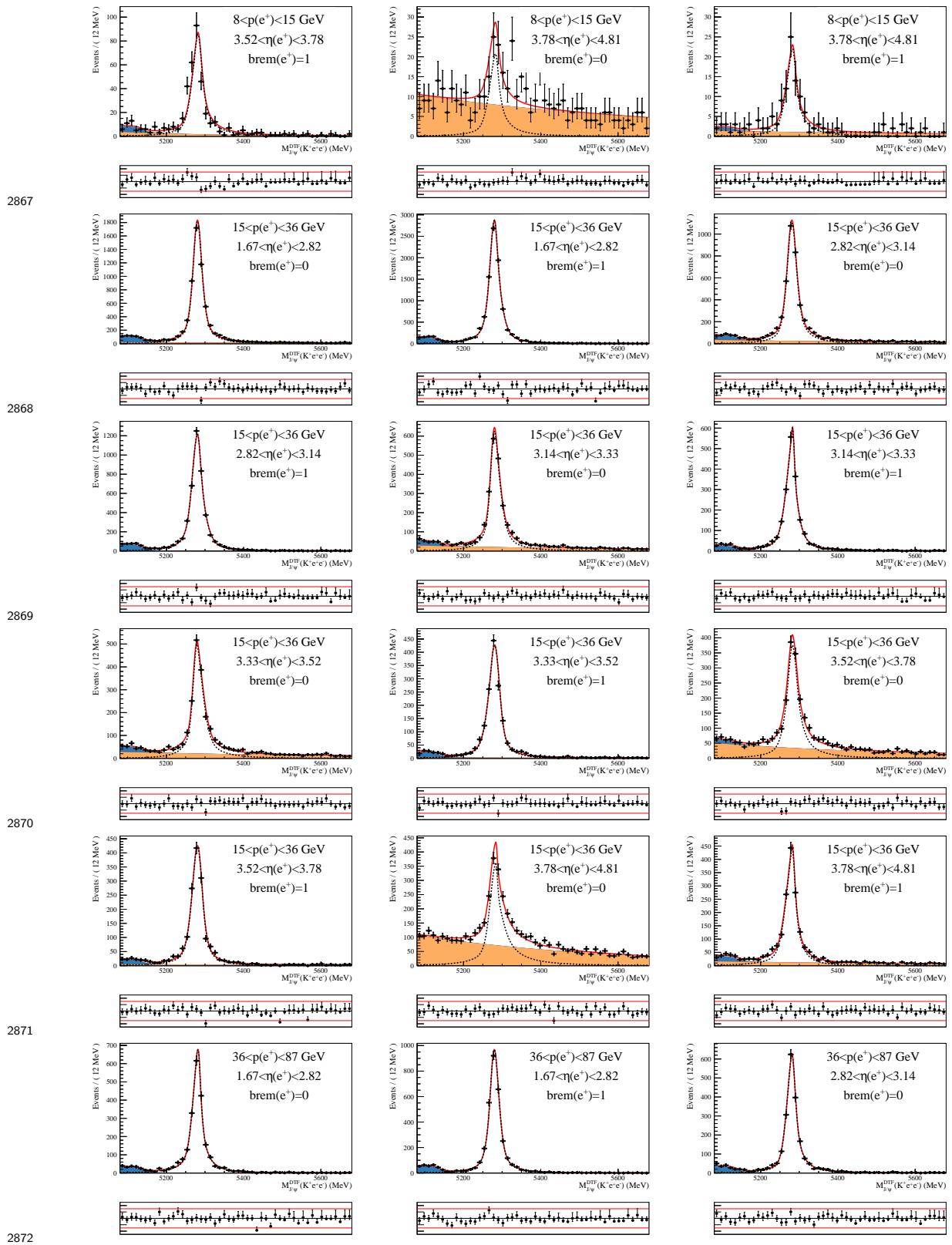
2863

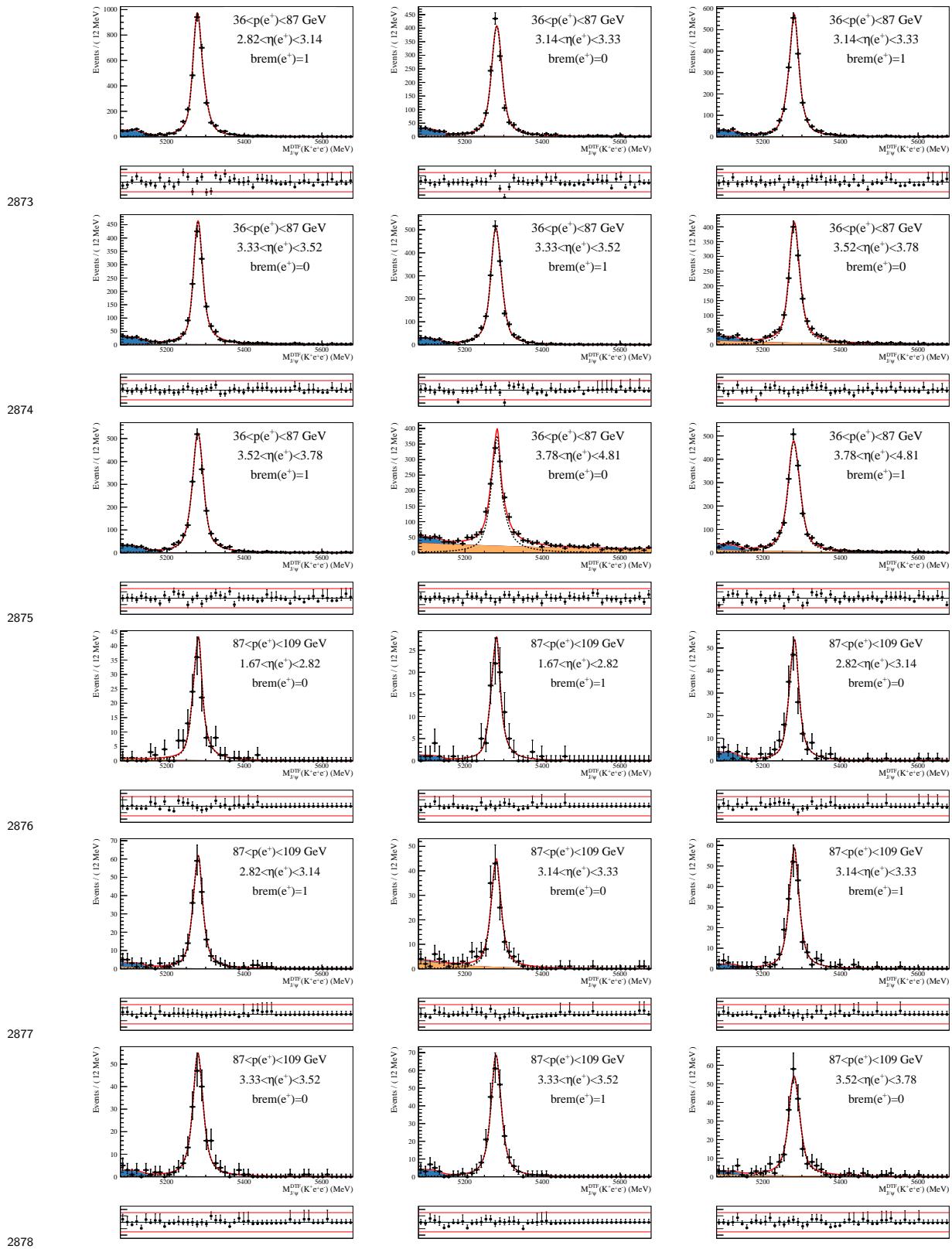


2865

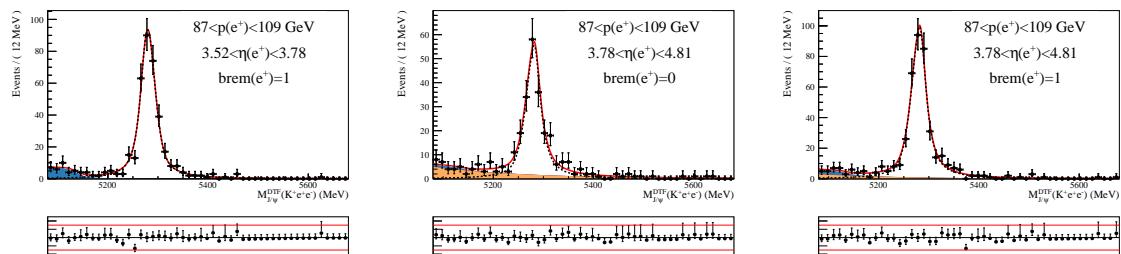


2866

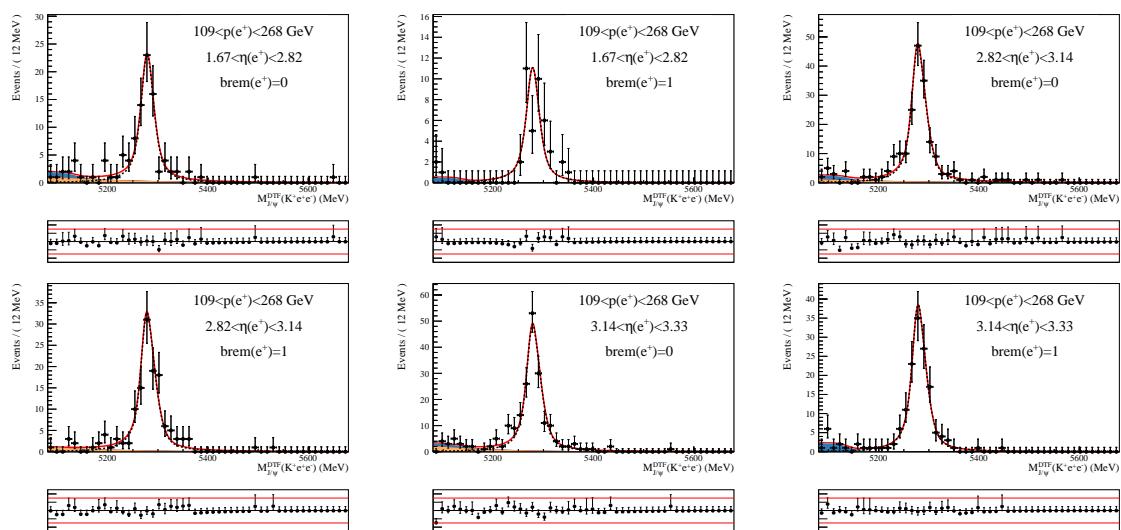




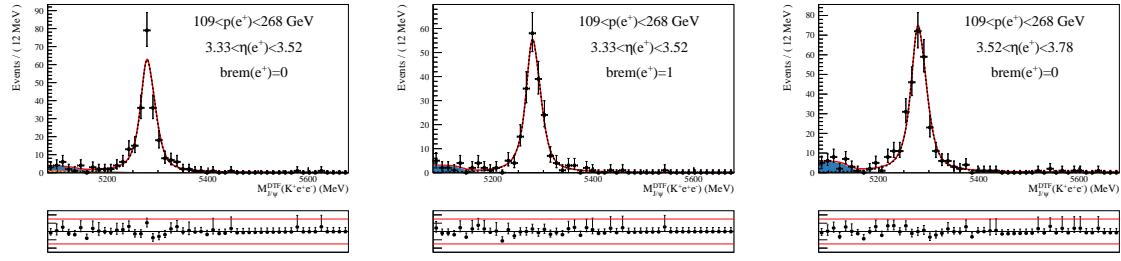
2879



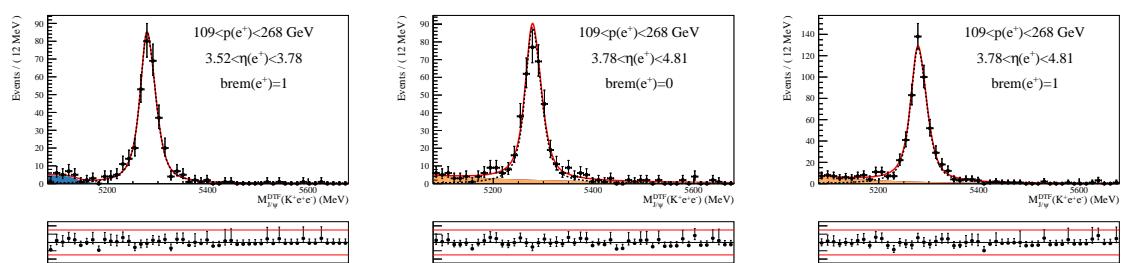
2880



2881

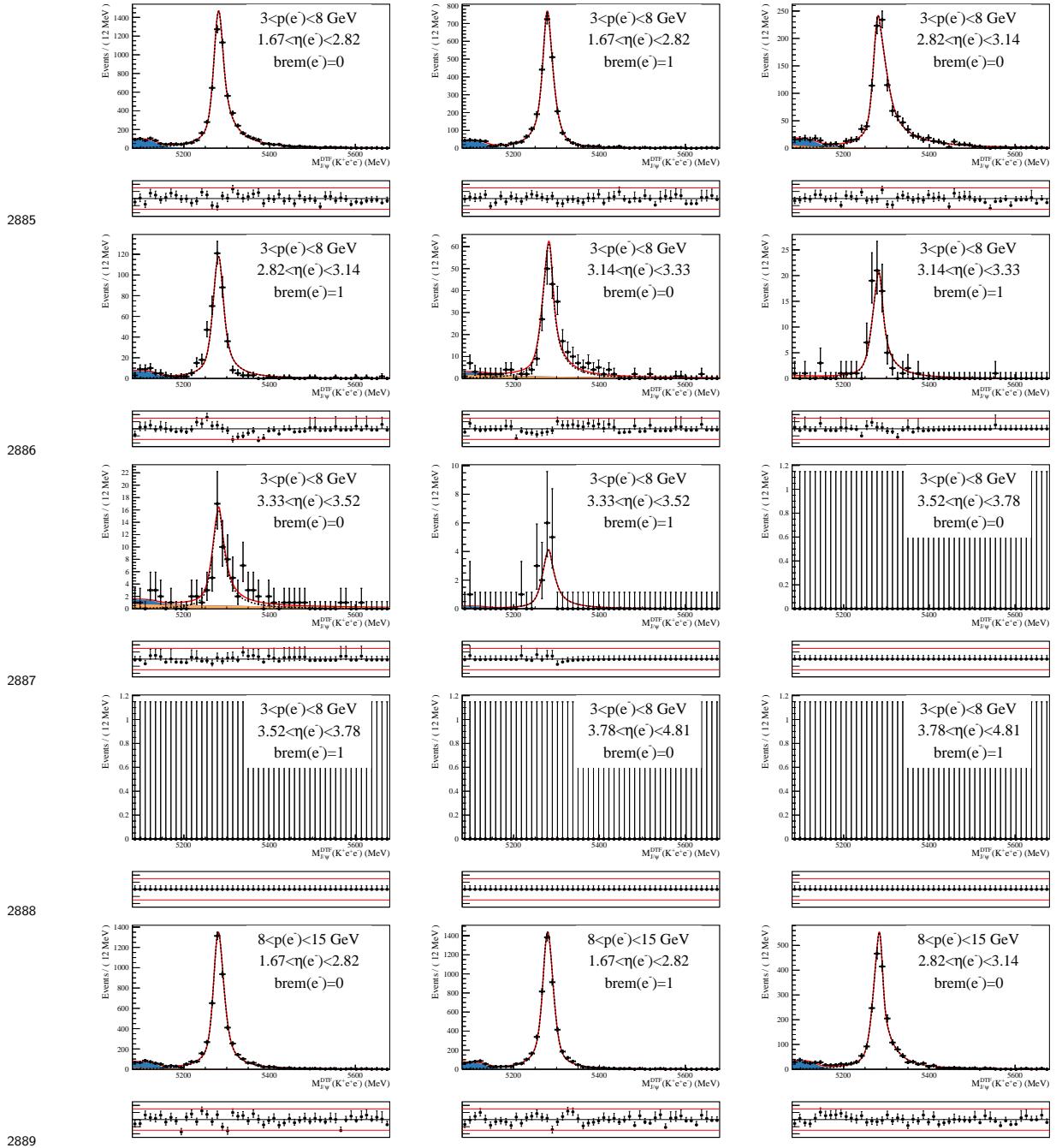


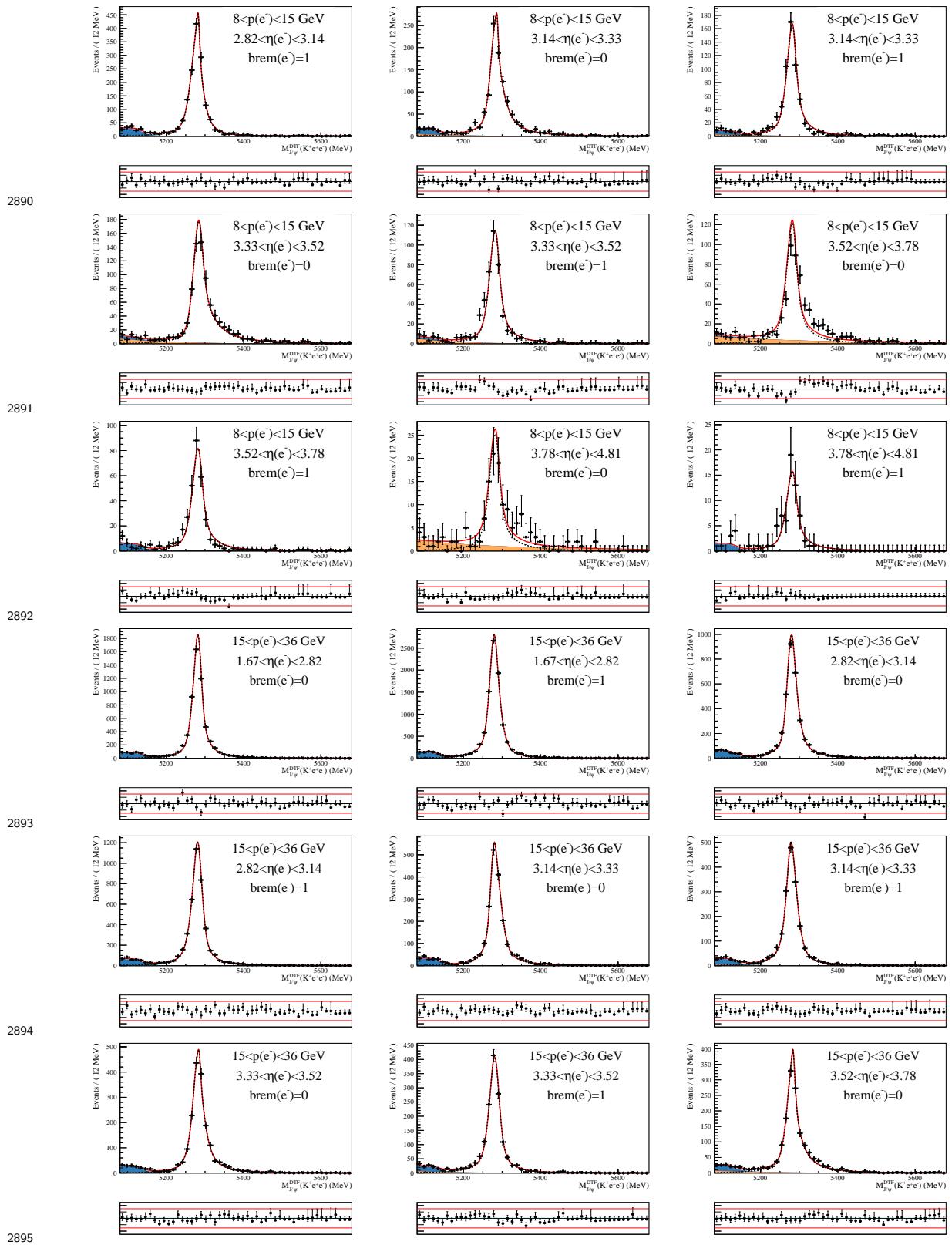
2882

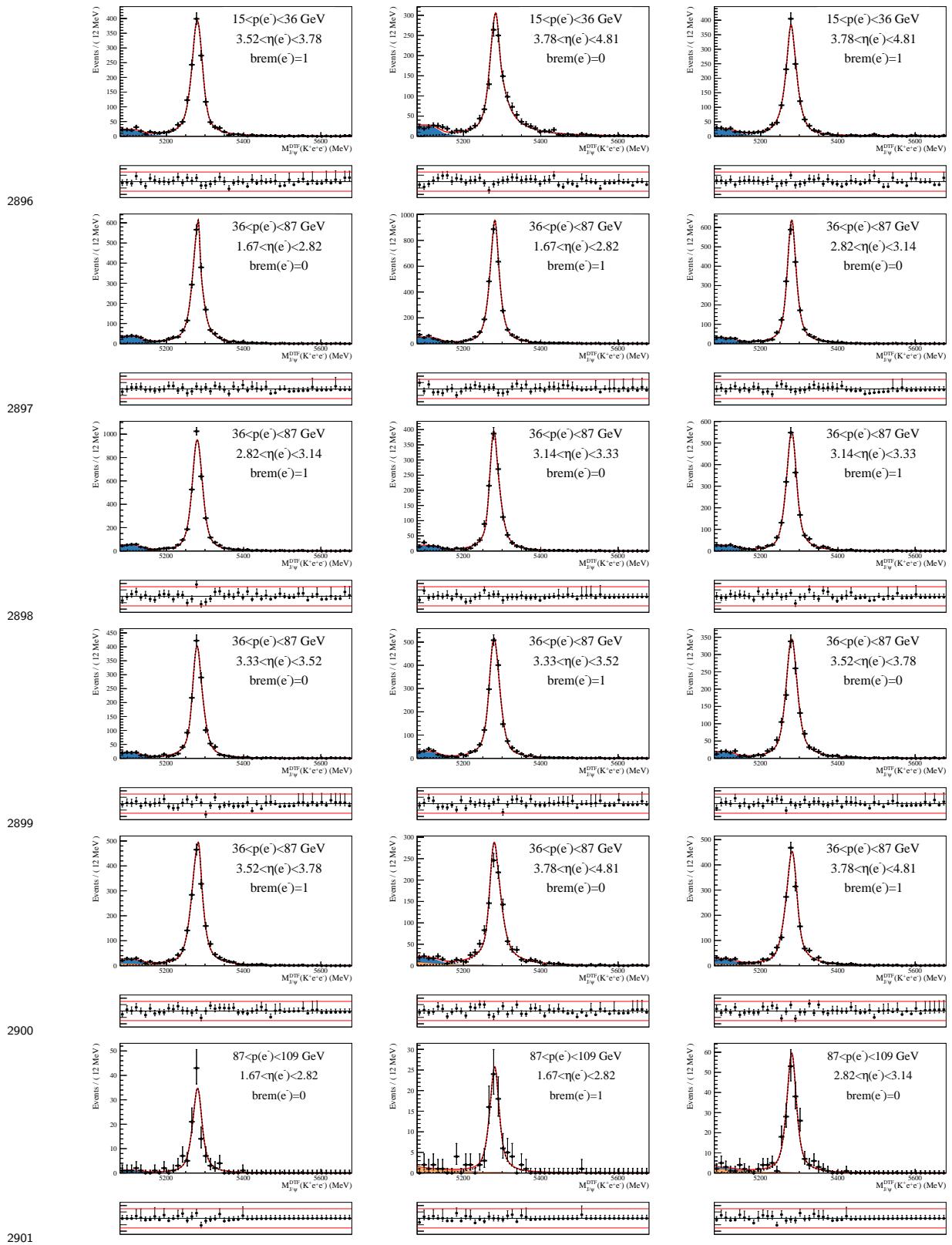


2883

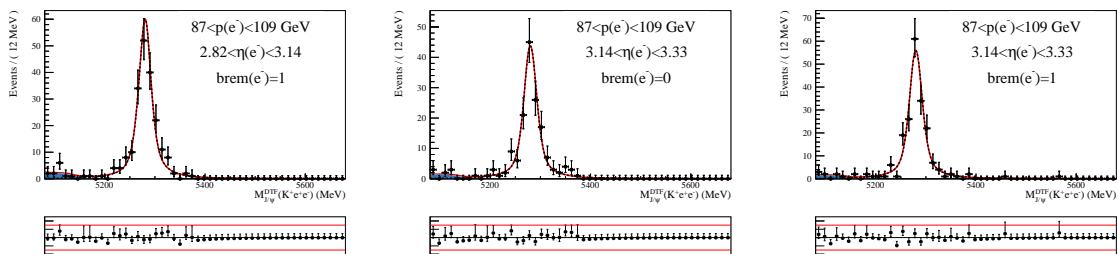
2884 D.4.3 2016 Fits, e^- probe, after PID cut



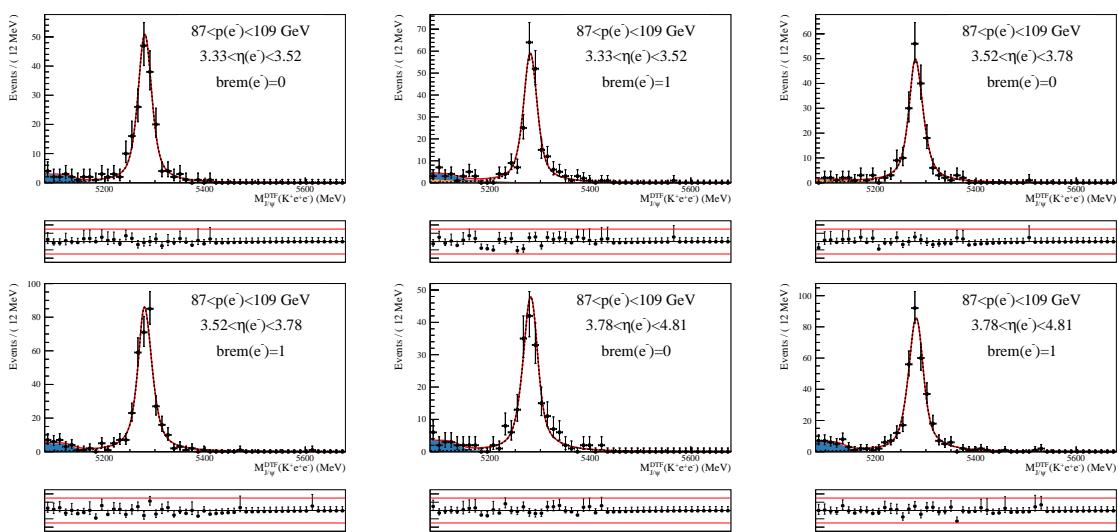




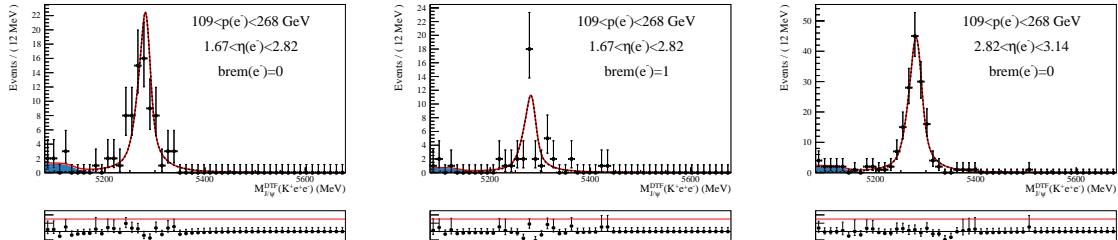
2902



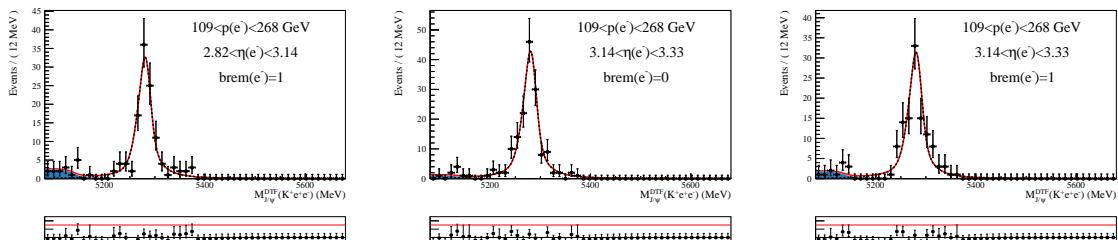
2903



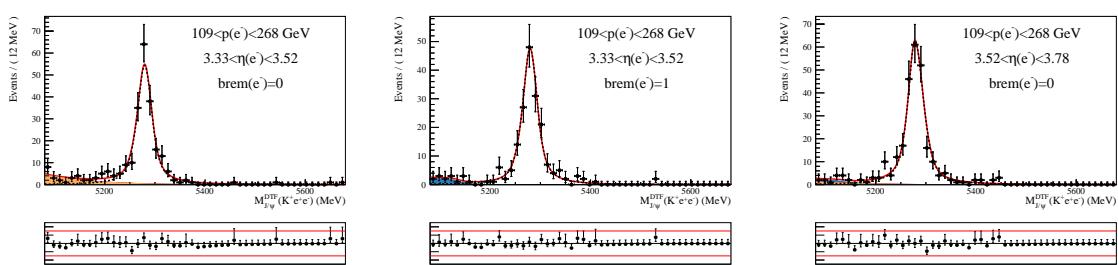
2904



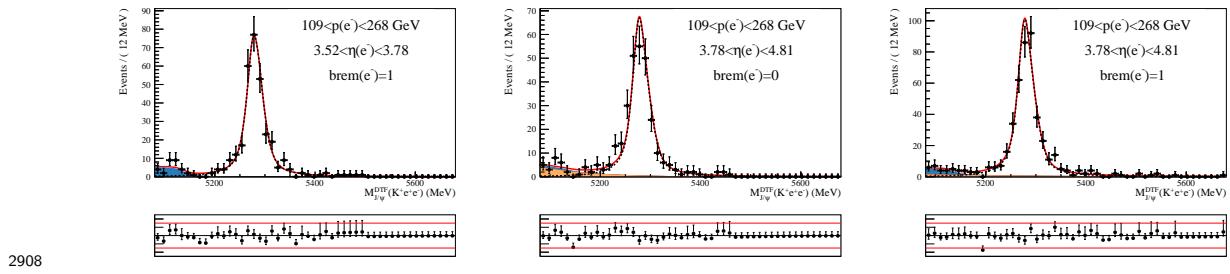
2905



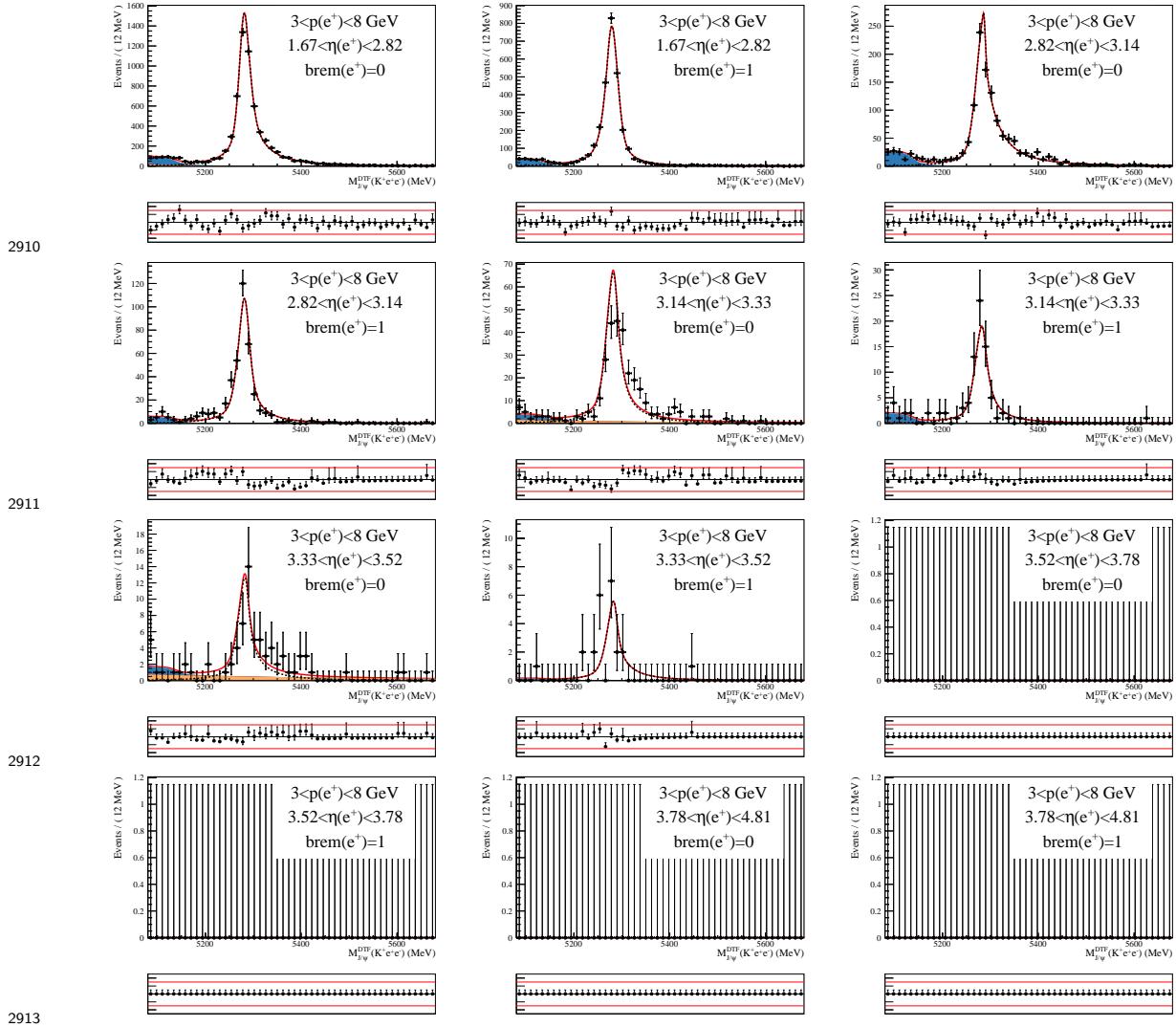
2906

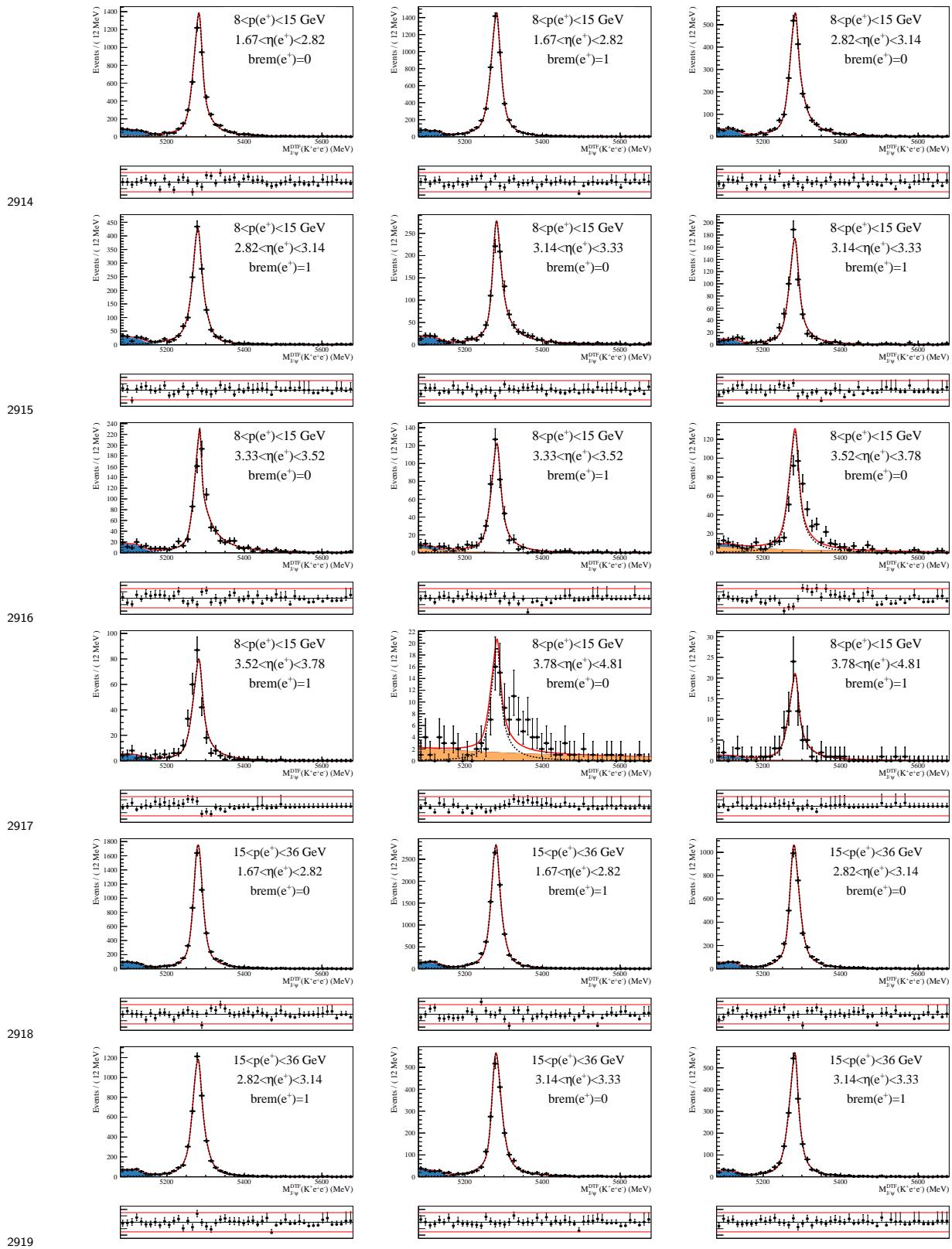


2907

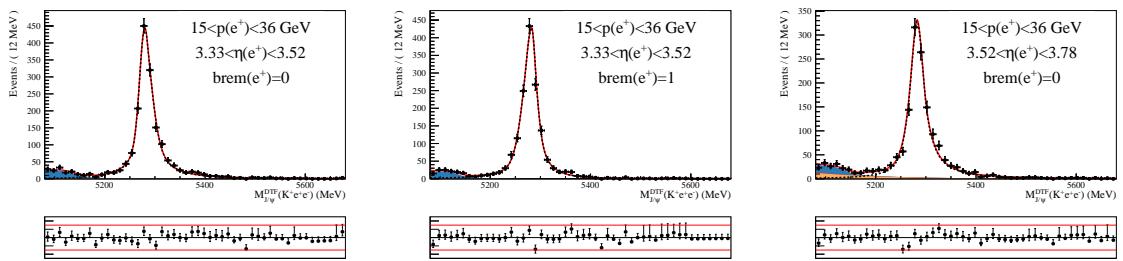


2909 D.4.4 2016 Fits, e^+ probe, after PID cut

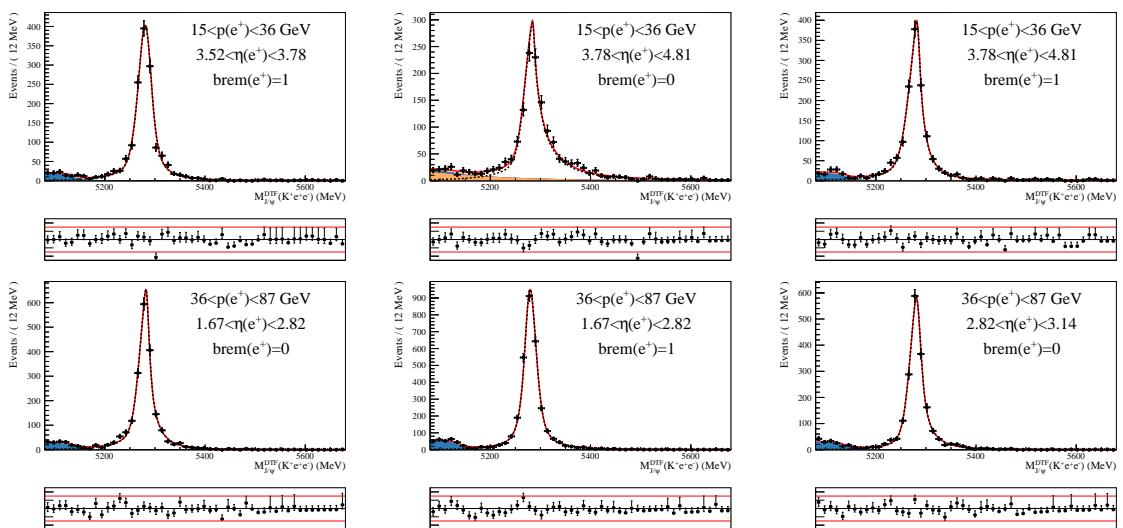




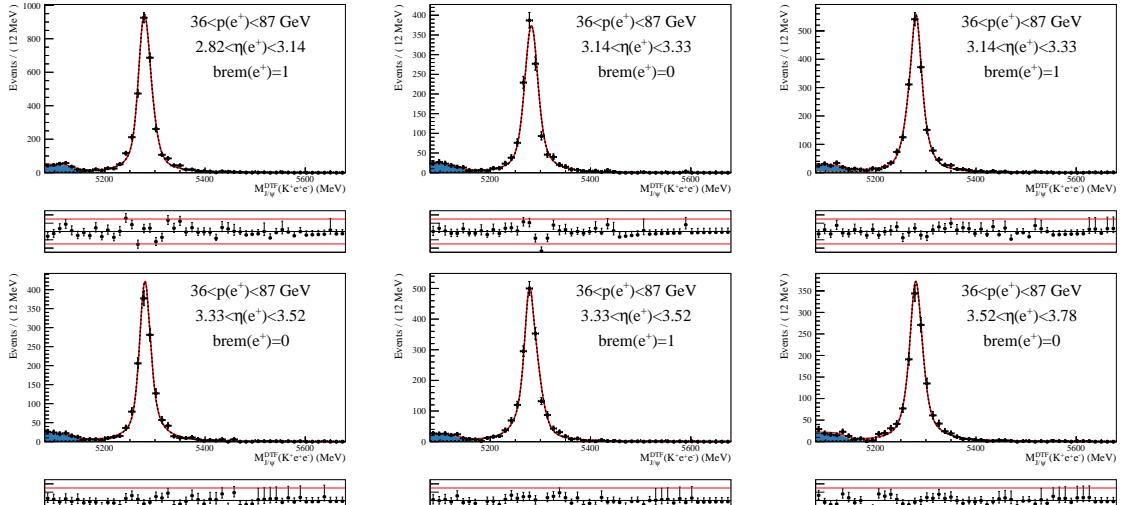
2920



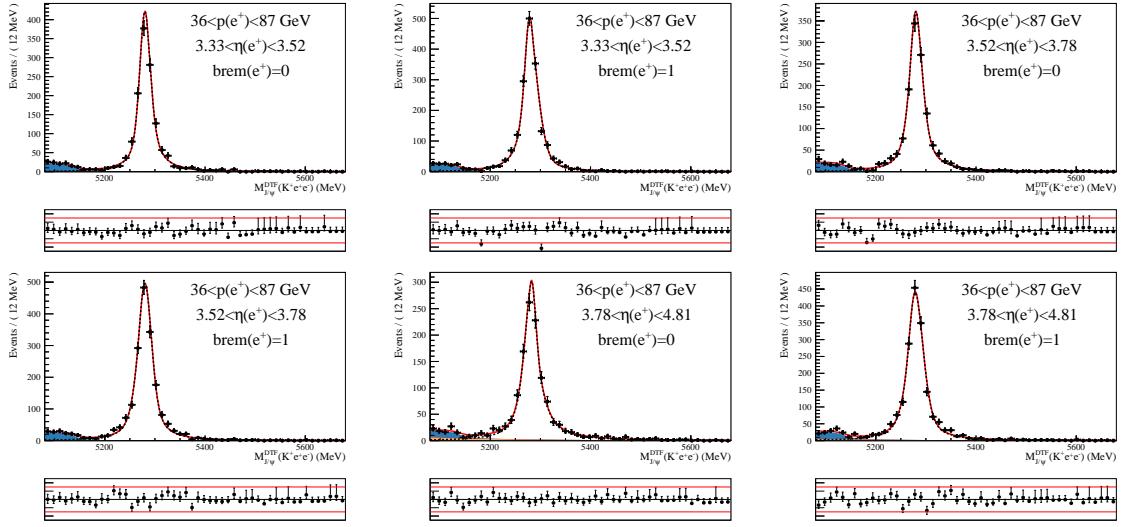
2921



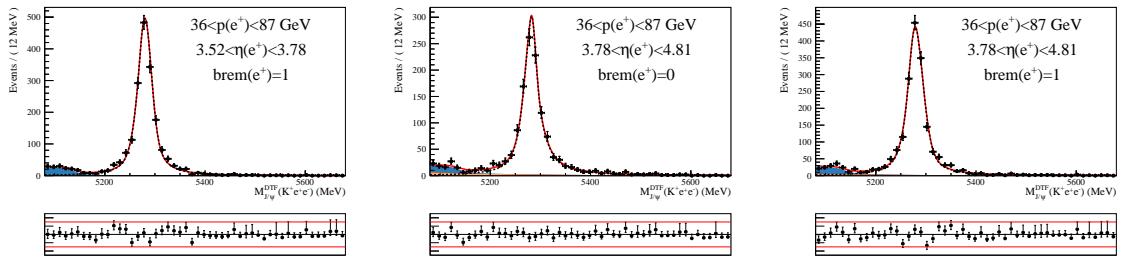
2922



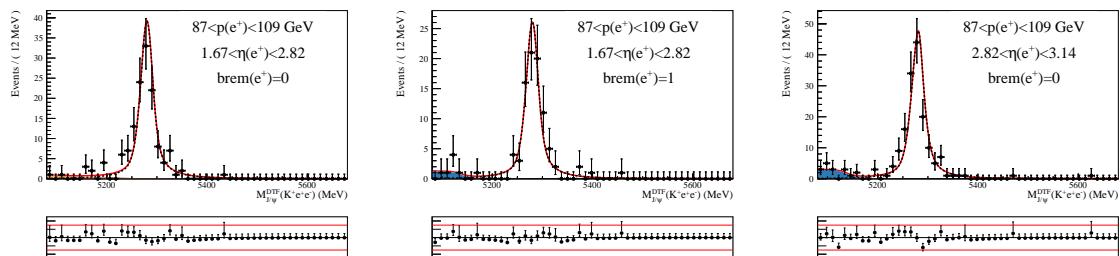
2923



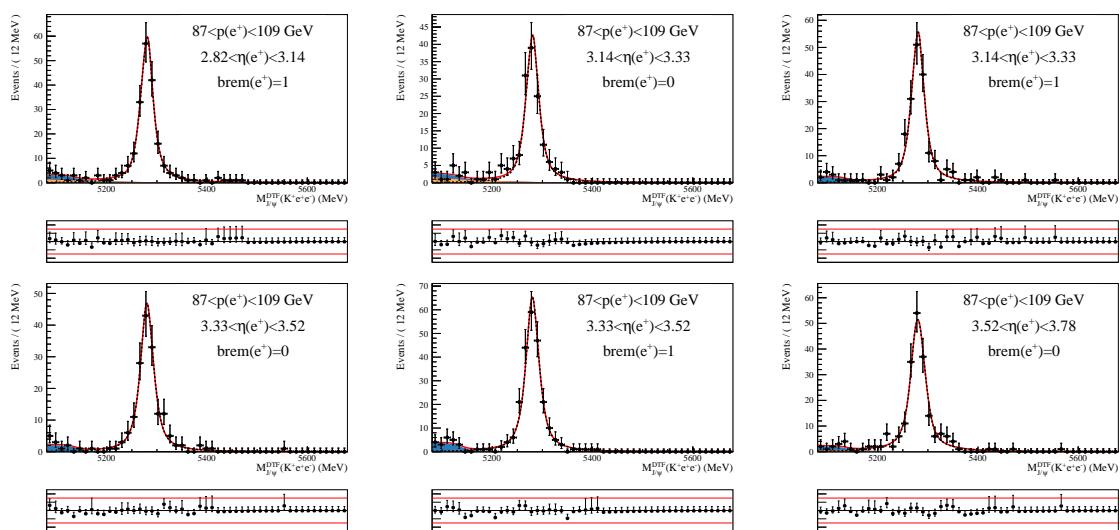
2925



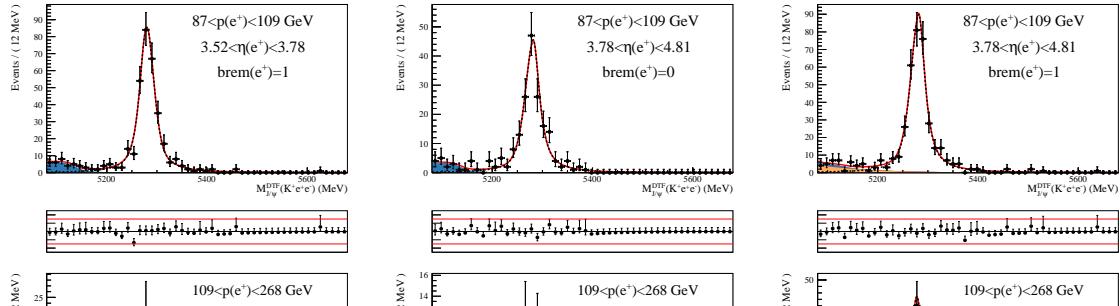
2926



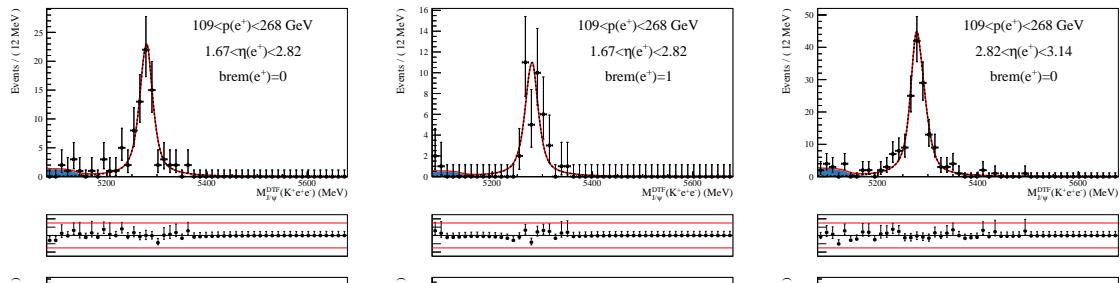
2927



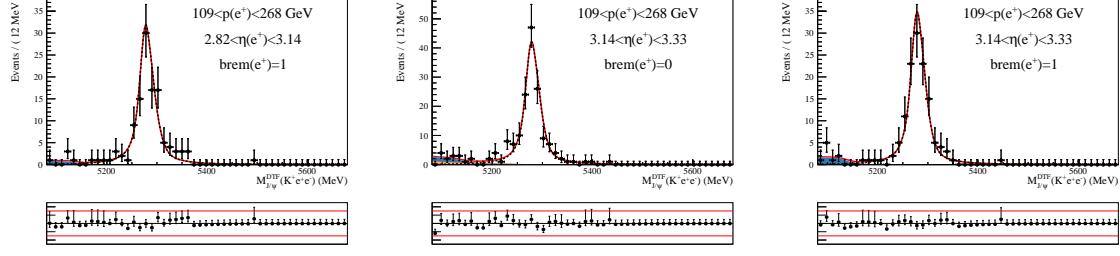
2928



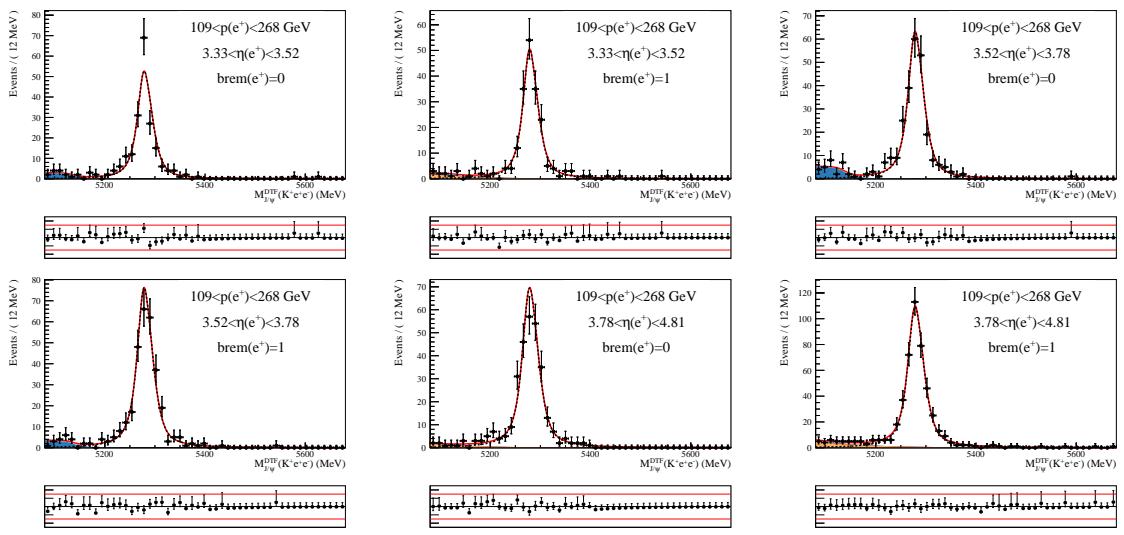
2929



2930



2931



2934 E PID supplementary plots

2935 Kinematic and occupancy distributions in the PID calibration data compared to the
 2936 control and rare mode are shown in Fig. 148-Fig. 150

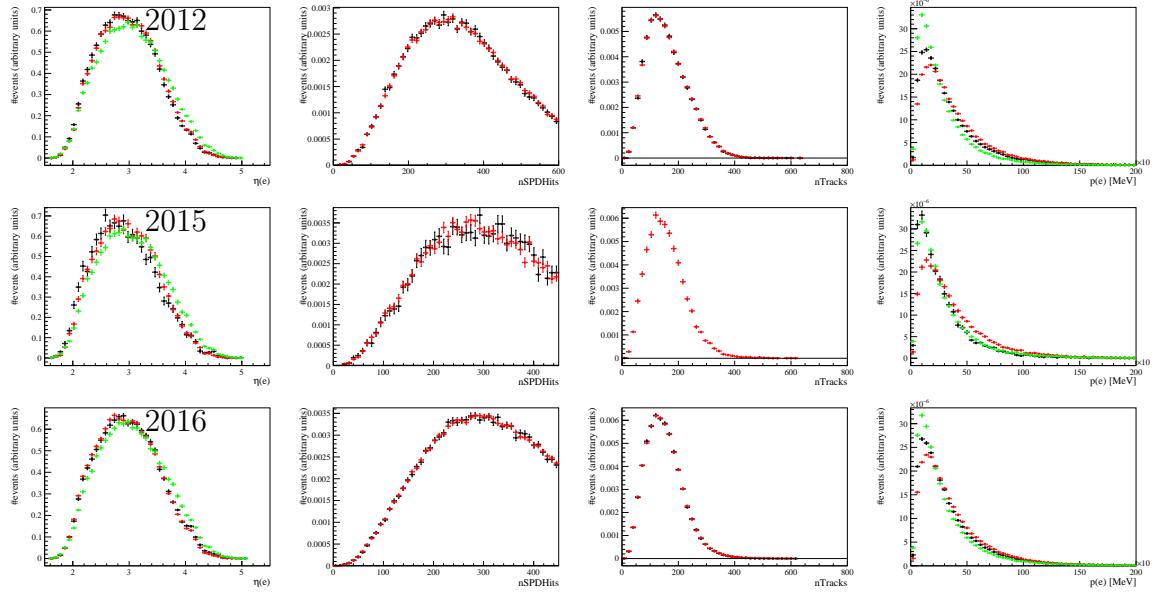


Figure 148: Electron kinematic and occupancy distributions for events taken from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ PID calibration data, $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ pre-selected data and $B^+ \rightarrow K^+ e^+e^-$ simulation in 2012, 2015 and 2016 data taking conditions.

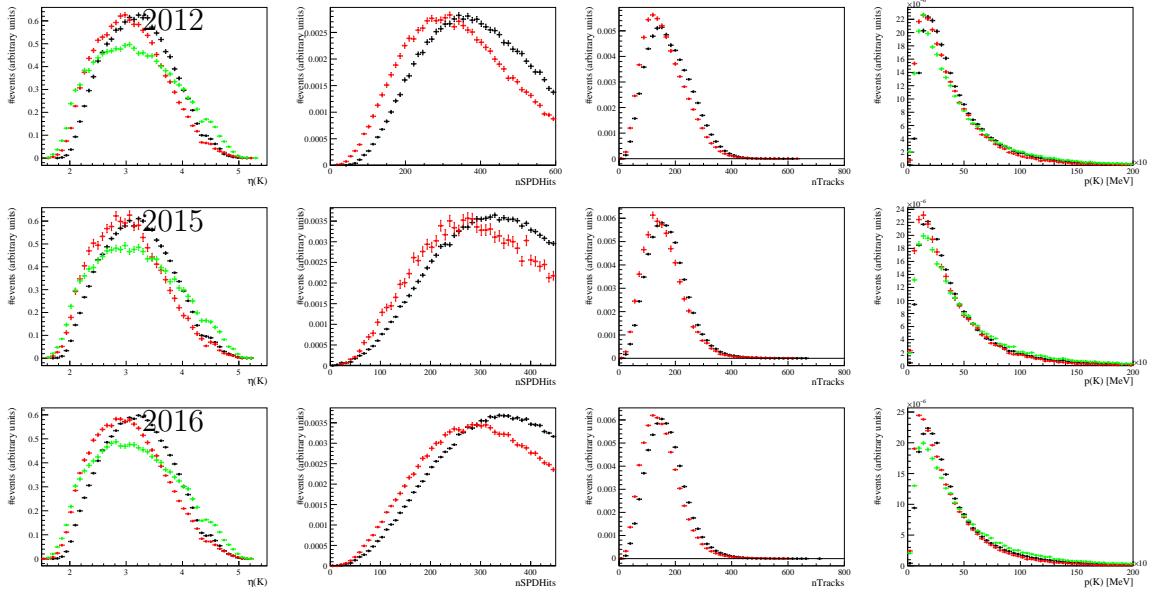


Figure 149: Kaon kinematic and occupancy distributions for events taken from $D^0 \rightarrow K^+ \pi^-$ PID calibration data, $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ pre-selected data and $B^+ \rightarrow K^+ e^+ e^-$ simulation in 2012, 2015 and 2016 data taking conditions.

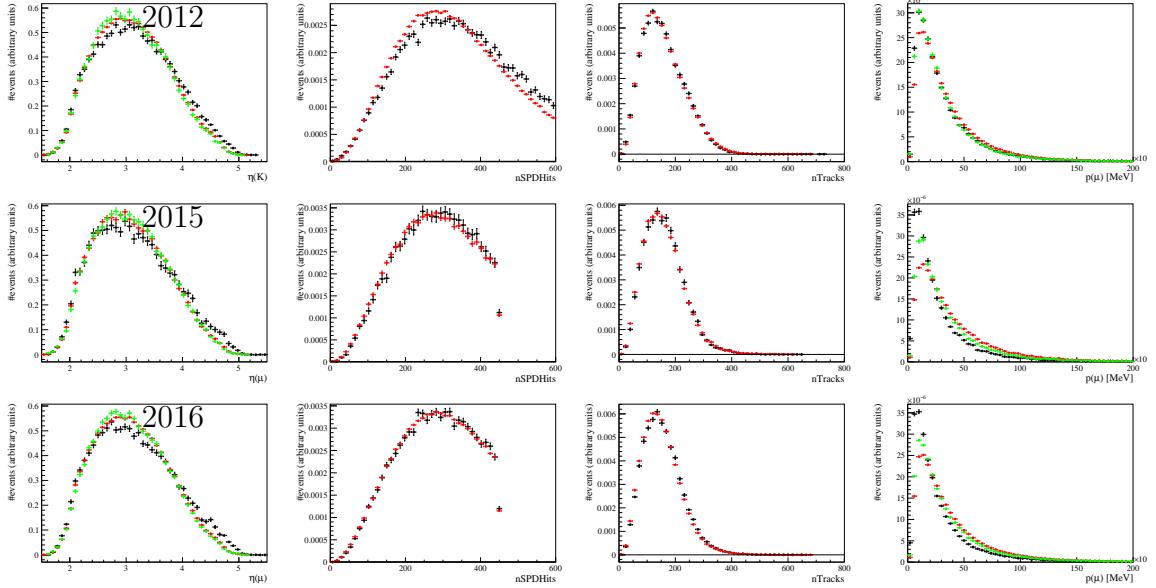


Figure 150: Muon kinematic and occupancy distributions for events taken from $J/\psi \rightarrow \mu^+ \mu^-$ PID calibration data, $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ pre-selected data and $B^+ \rightarrow K^+ \mu^+ \mu^-$ simulation in 2012, 2015 and 2016 data taking conditions.

2937 **F Supplementary kinematic plots**

2938 The distributions of several kinematic variables for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$
2939 events in simulation and background-subtracted data are shown in Fig. 151 in the *hTOS!*
2940 category in the Run 1 data taking condition, in Fig. 152 in the *TIS!* category in the
2941 Run 1 data taking condition, in Fig. 153 in the *hTOS!* category in the Run 2 data taking
2942 condition, and in Fig. 154 in the *TIS!* category in the Run 2 data taking condition.

2943 **G Shape of the pseudo-rapidity distributions**

2944 Comparing the η distributions for the daughter particles in $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ simulated
2945 events, see Fig. 155, a visible tail at high values is observed for both muons and electrons,
2946 being more pronounced in the case of muons. Moreover, a dip at $\eta \sim 4.3$ is also visible in
2947 the electron distributions, creating a shoulder at large η values. Figure 155 also shows the
2948 material distribution as a function of η present in the simulation. The difference between
2949 the muon and electron distribution at high η is therefore attributed to the increased
2950 material thickness in this region, which is expected to have a larger impact on electrons
2951 than on muons. In particular, the dip around $\eta = 4.3$ is due to the RICH 1 beam pipe.

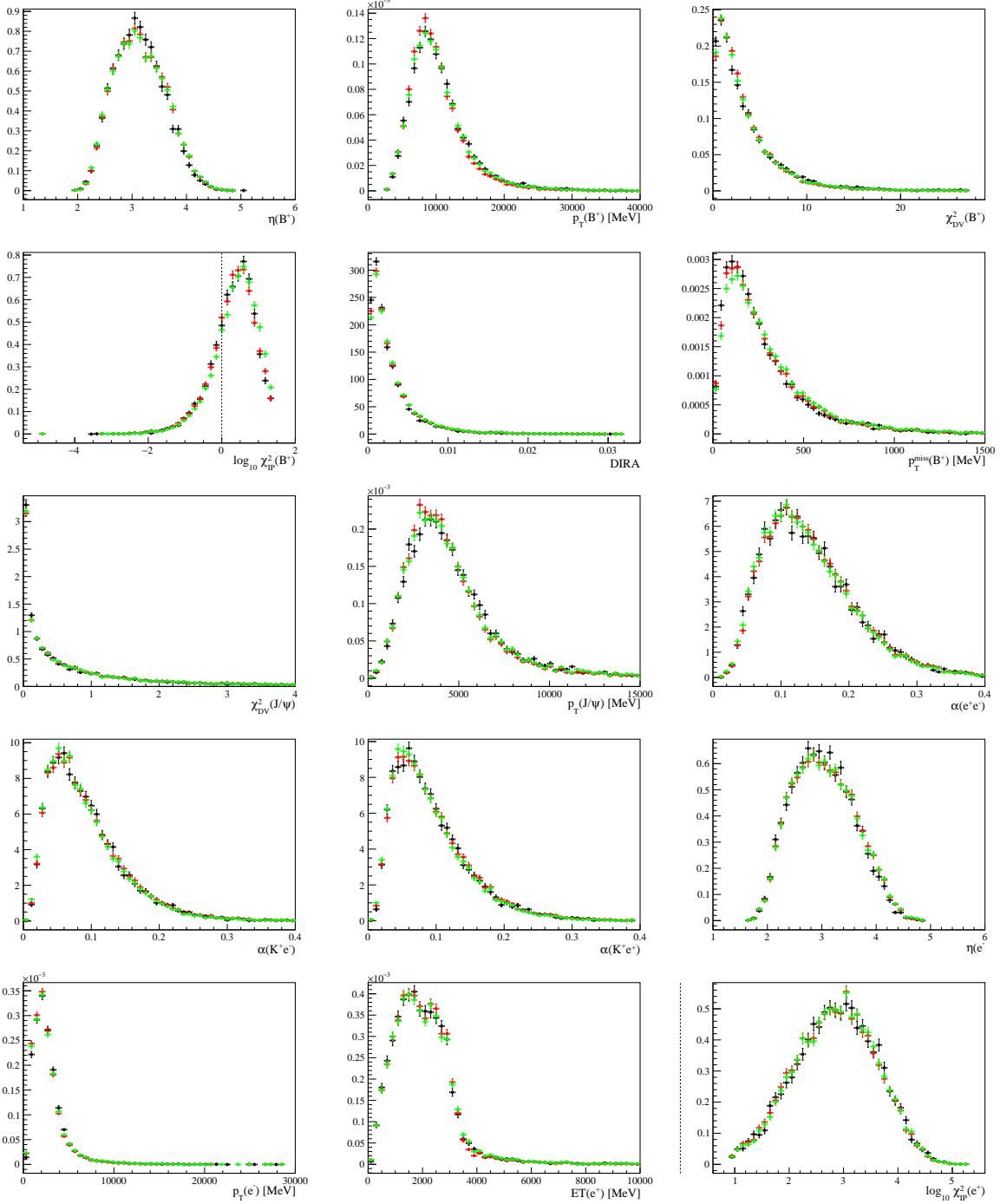


Figure 151: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events in the *hTOS!* category and in the Run 1 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

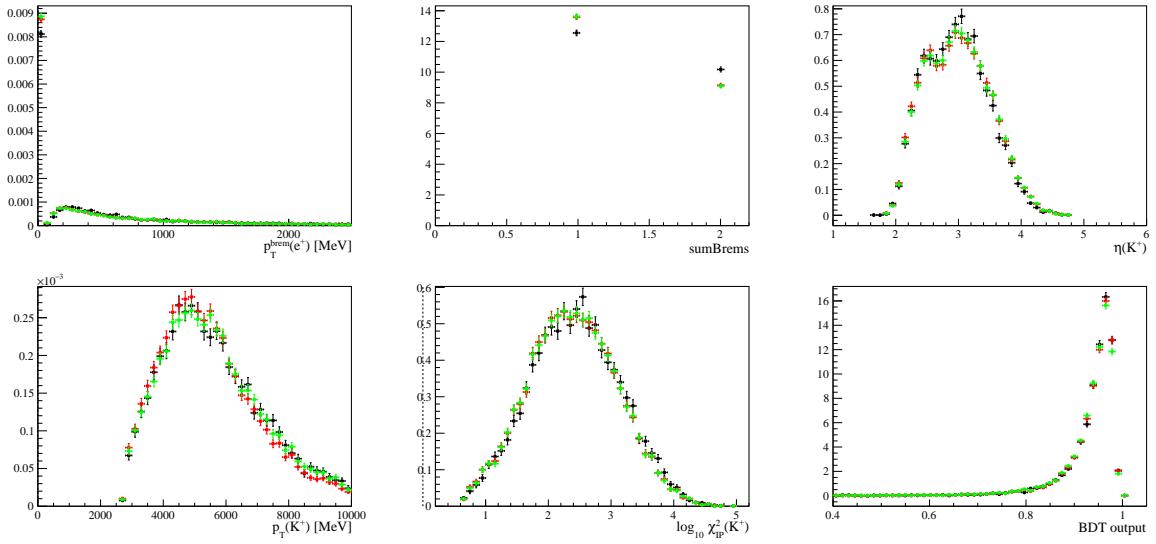


Figure 151: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the *hTOS!* category and in the Run 1 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

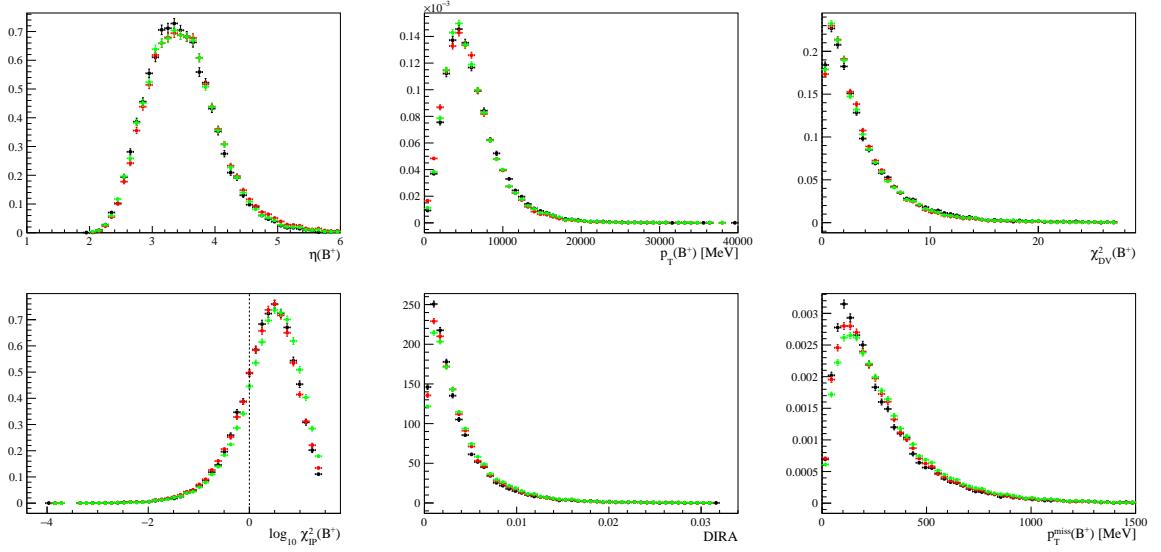


Figure 152: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the *TIS!* category and in the Run 1 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

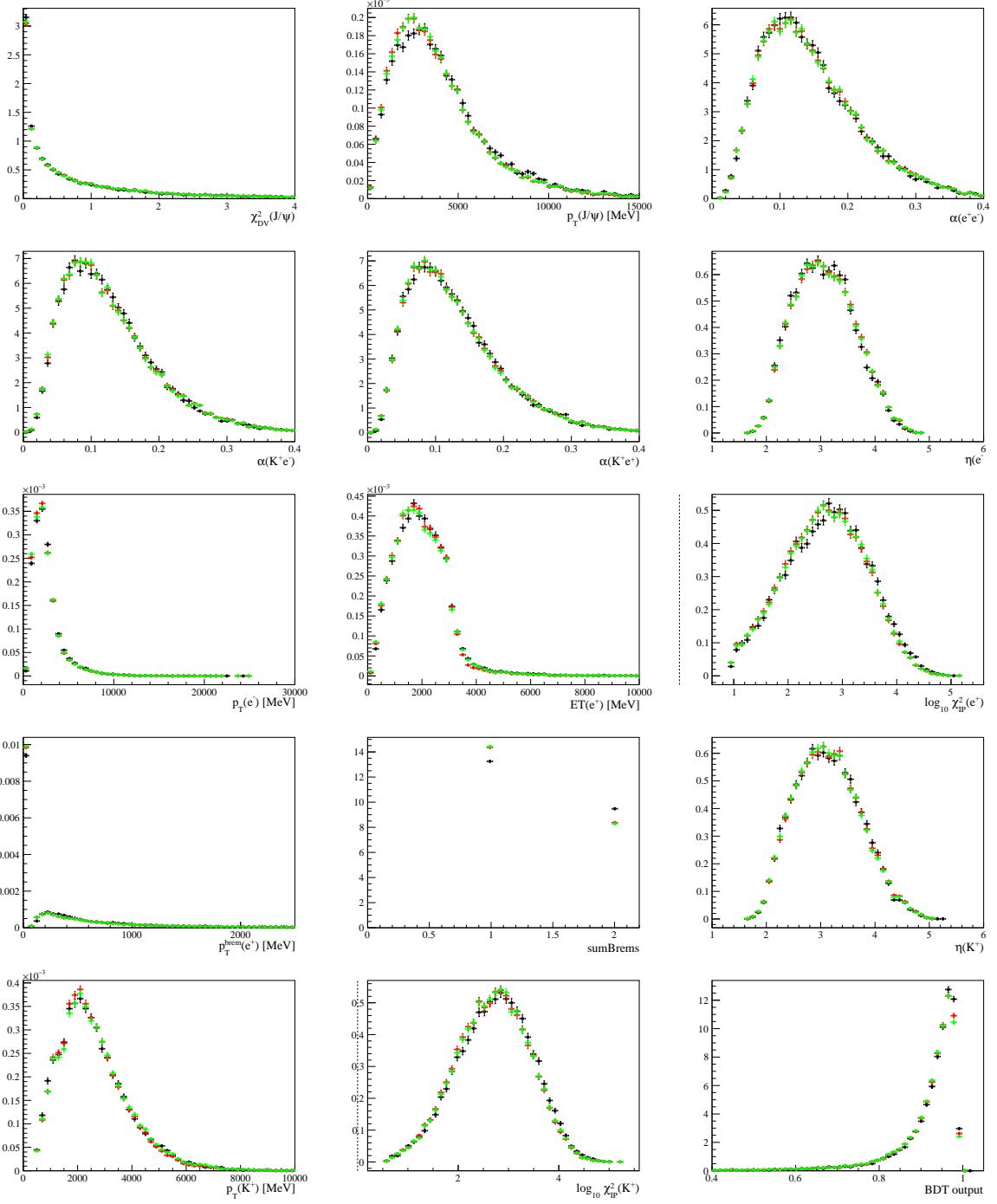


Figure 152: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the TIS! category and in the Run 1 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

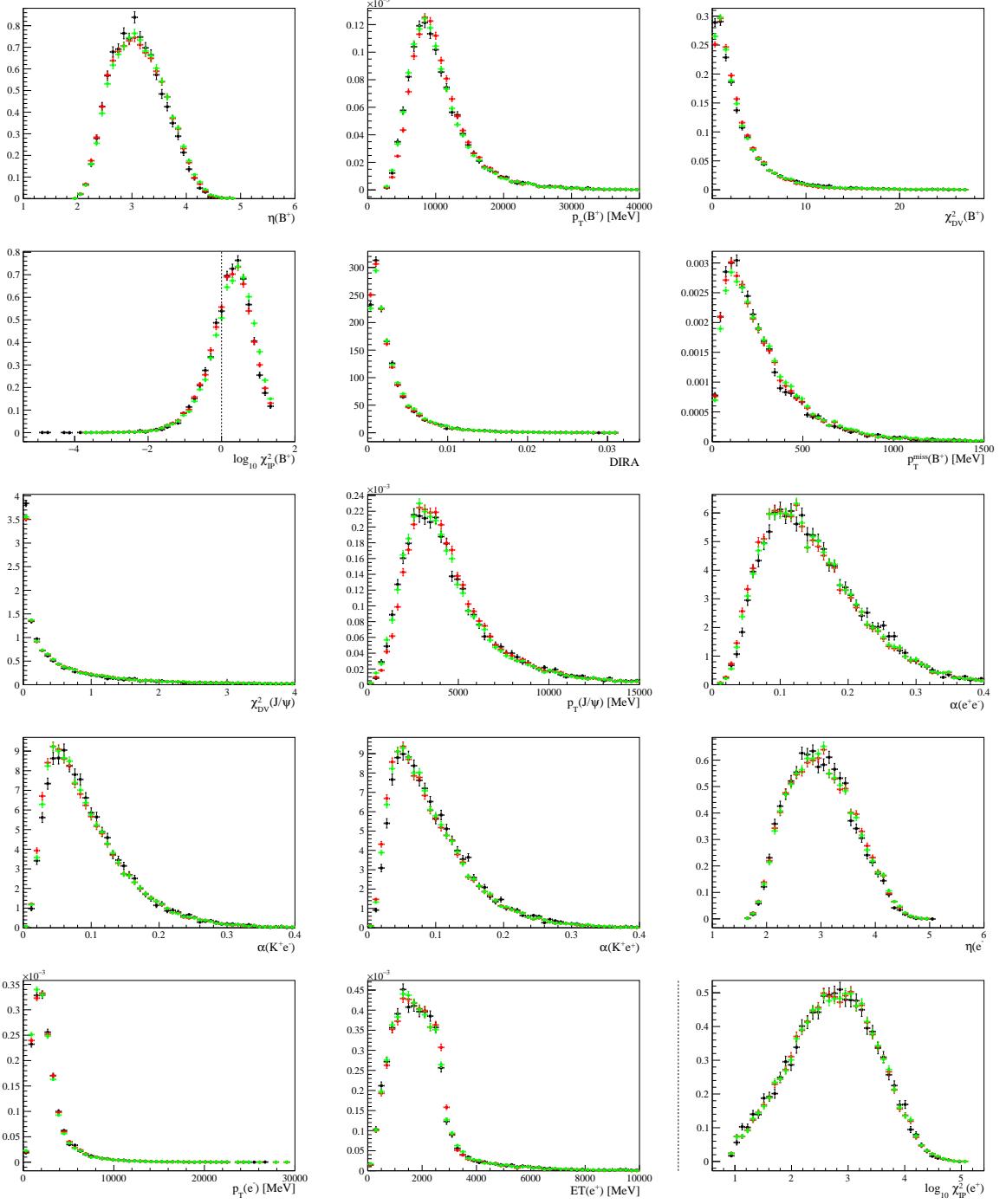


Figure 153: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi (e^+ e^-)$ events in the *hTOS!* category and in the Run 2 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

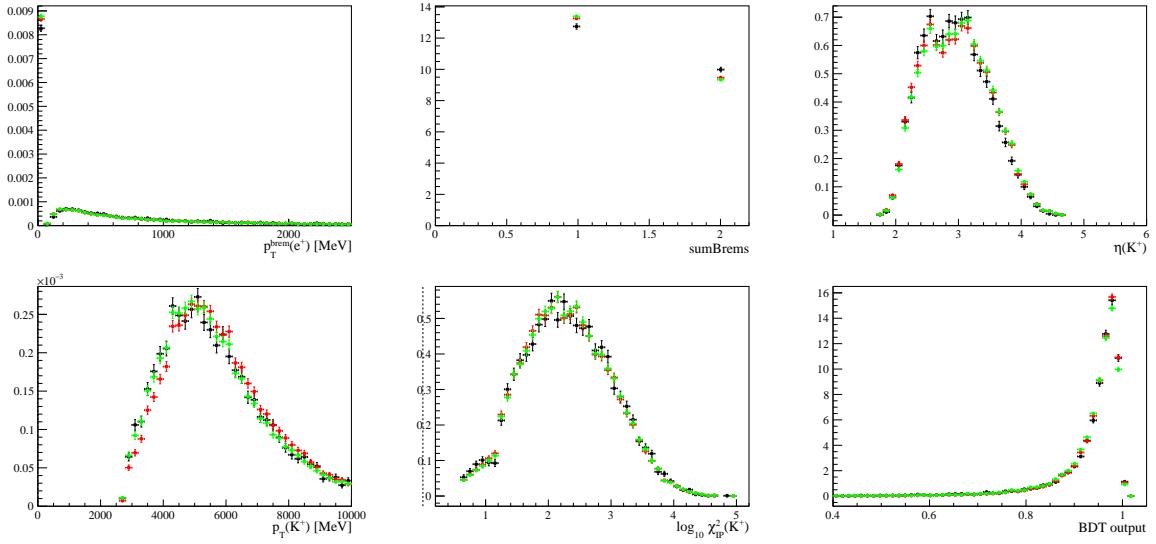


Figure 153: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the $h\text{TOS!}$ category and in the Run 2 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

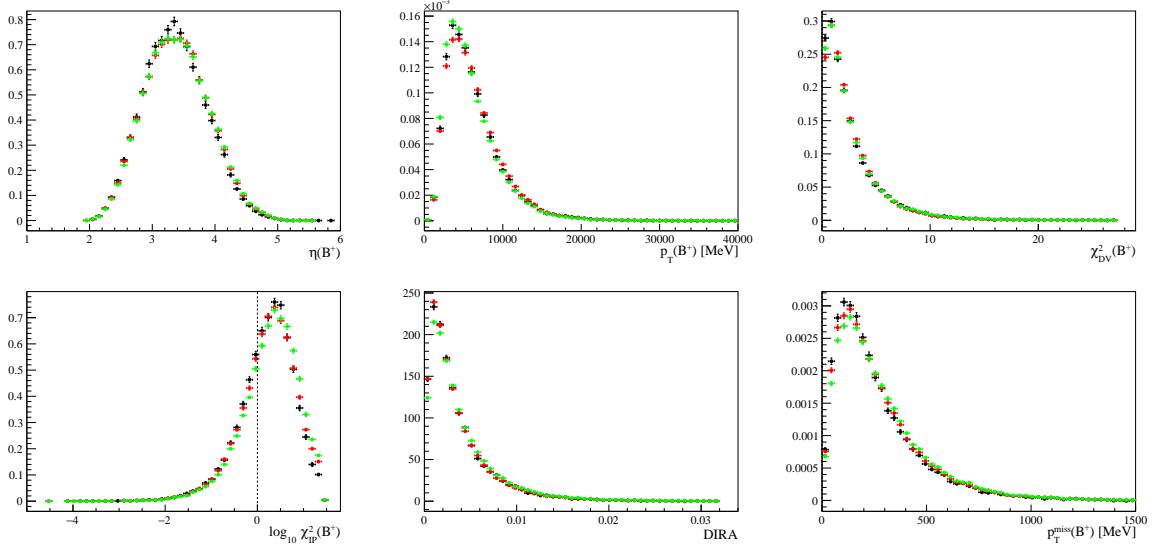


Figure 154: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the $T\text{IS!}$ category and in the Run 2 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

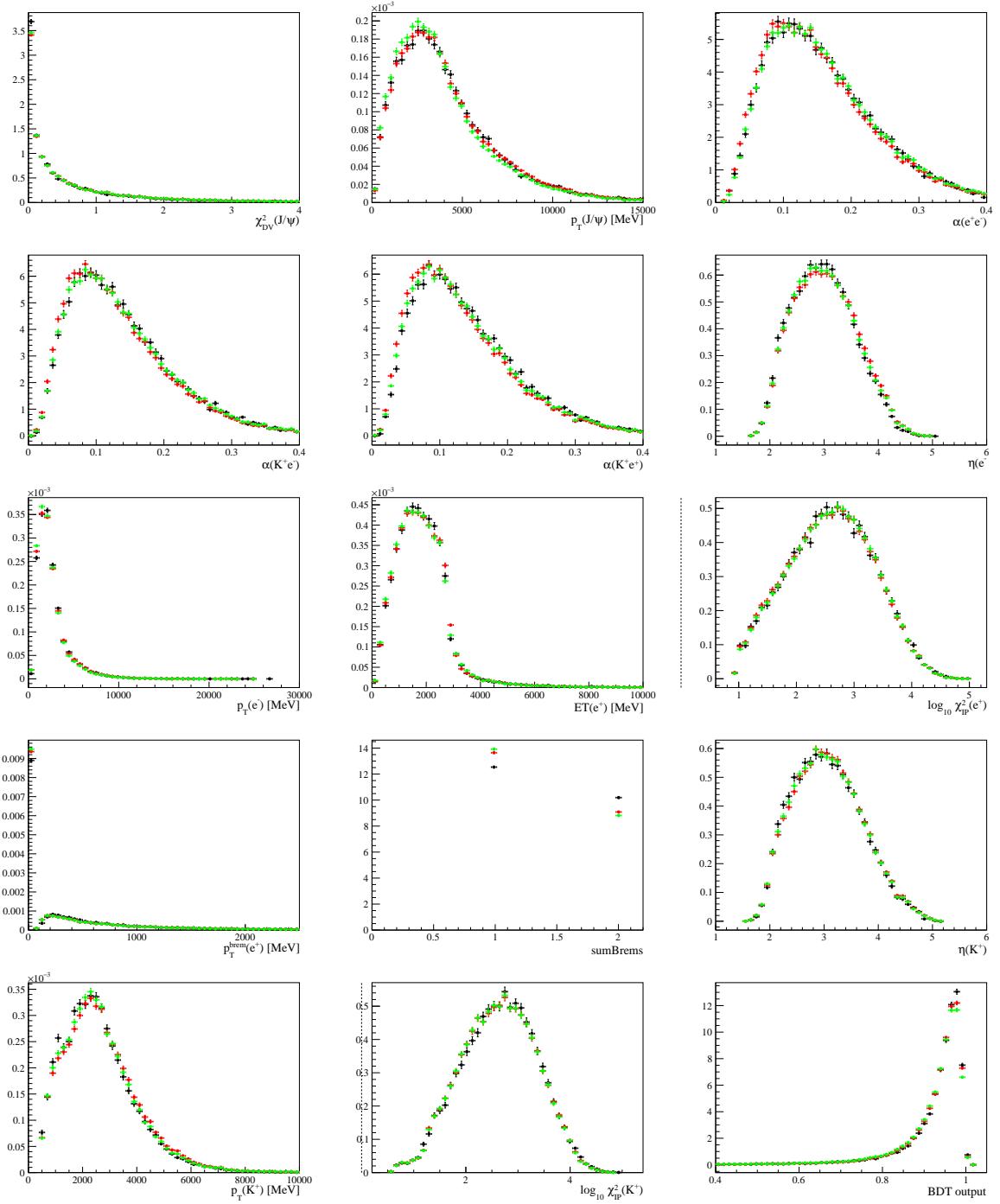


Figure 154: Kinematic distributions for offline selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the TIS! category and in the Run 2 data taking conditions. Each plots shows the background subtracted data (black), simulation (red) and simulation with kinematic correction weights (green).

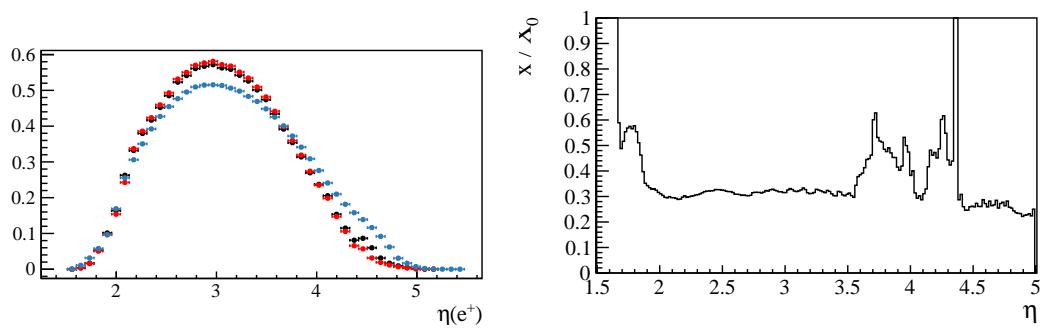


Figure 155: Left: Pseudo-rapidity distribution for reconstructed electrons (black), reconstructed electrons required to be within the CALO acceptance (red) and muons (blue), from $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ simulation. Right: Material distribution in simulation as a function of η .

2952 H Fits to simulation

2953 H.1 Signal fits

2954 H.1.1 Fits to $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ simulation

2955 Fig. 156 shows the result of the fit to the $m_{\text{DTF}}^{J/\psi}(K^+\mu^+\mu^-)$ distribution of simulated
 2956 $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events after the full selection chain. Equivalent plots for the fits to
 2957 the distribution in the unconstrained invariant mass, $m(K^+\mu^+\mu^-)$, is shown in Fig. 157.

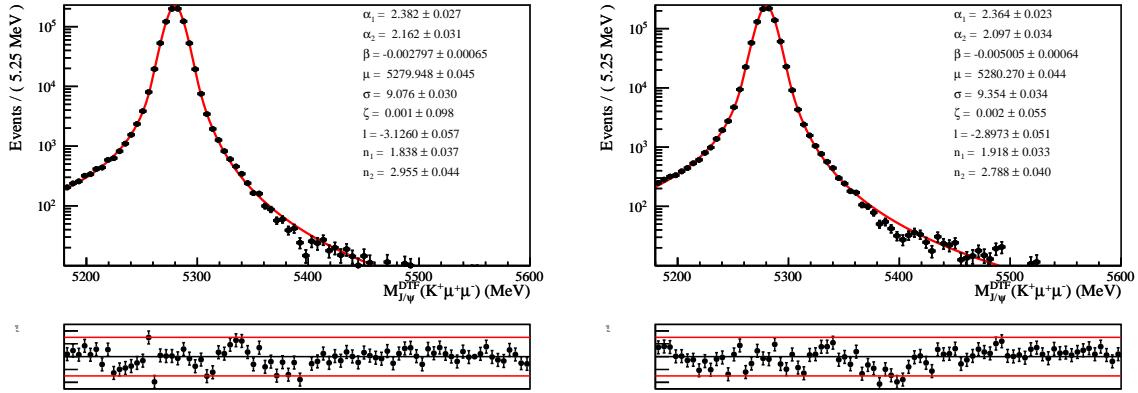


Figure 156: Fits to the $m_{\text{DTF}}^{J/\psi}(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ simulated candidates in Run 1 (left) and Run 2 (right) samples.

2958 H.1.2 Fits to $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulation

2959 Fig. 158, Fig. 159 and Fig. 160 show the result of the fits to the $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-)$ distribution
 2960 of simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the 0γ , 1γ and 2γ categories. Equivalent
 2961 plots for the fits to the distributions in the unconstrained invariant mass, $m(K^+e^+e^-)$,
 2962 are shown in Fig. 161, Fig. 162 and Fig. 163.

2963 H.1.3 Fits to $B^+ \rightarrow K^+\psi(2S)(\mu^+\mu^-)$ simulation

2964 Fig. 164 shows the result of the fit to the $m_{\text{DTF}}^{\psi(2S)}(K^+\mu^+\mu^-)$ distribution of simulated
 2965 $B^+ \rightarrow K^+\psi(2S)(\mu^+\mu^-)$ events after the full selection chain.

2966 H.1.4 Fits to $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ simulation

2967 Fig. 165, Fig. 166 and Fig. 166 show the result of the fits to the $m_{\text{DTF}}^{\psi(2S)}(K^+e^+e^-)$ distribution
 2968 of simulated $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ events in the 0γ , 1γ and 2γ categories.

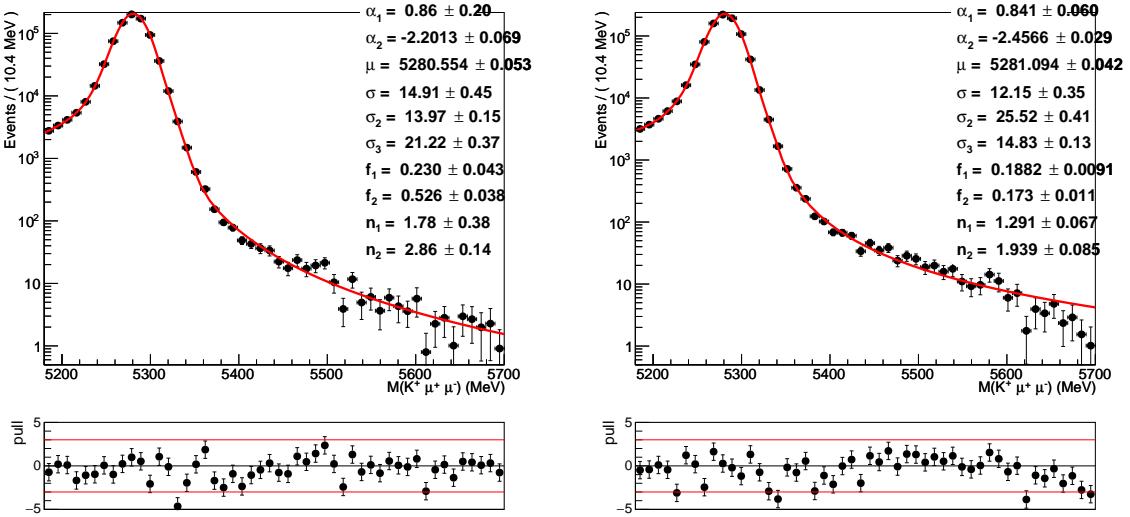


Figure 157: Fits to the $m(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ simulated candidates in Run 1 (left) and Run 2 (right) samples.

2969 H.1.5 Fits to $B^+ \rightarrow K^+\mu^+\mu^-$ simulation

2970 Fig. 168 shows the result of the fit to the $m(K^+\mu^+\mu^-)$ distribution of simulated $B^+ \rightarrow$
2971 $K^+\mu^+\mu^-$ events after the full selection chain.

2972 H.1.6 Fits to $B^+ \rightarrow K^+e^+e^-$ simulation

2973 Fig. 169, Fig. 170 and Fig. 171 show the result of the fits to the $m(K^+e^+e^-)$ distribution
2974 of simulated $B^+ \rightarrow K^+e^+e^-$ events in the 0 γ , 1 γ and 2 γ categories.

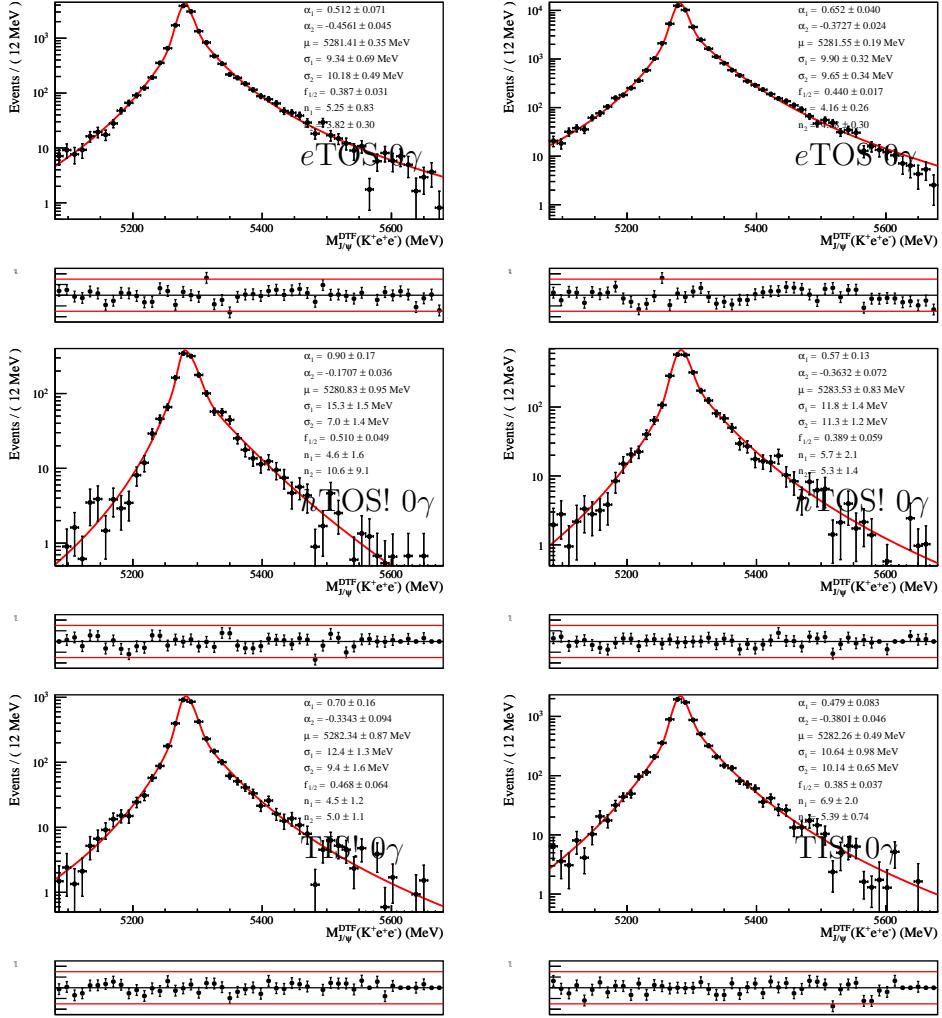


Figure 158: Fits to $m_{DTF}^{J/\psi}(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events in the bremsstrahlung category 0 γ for run 1 (left) and run 2 (right), for eTOS (top), hTOS! and TIS! events.

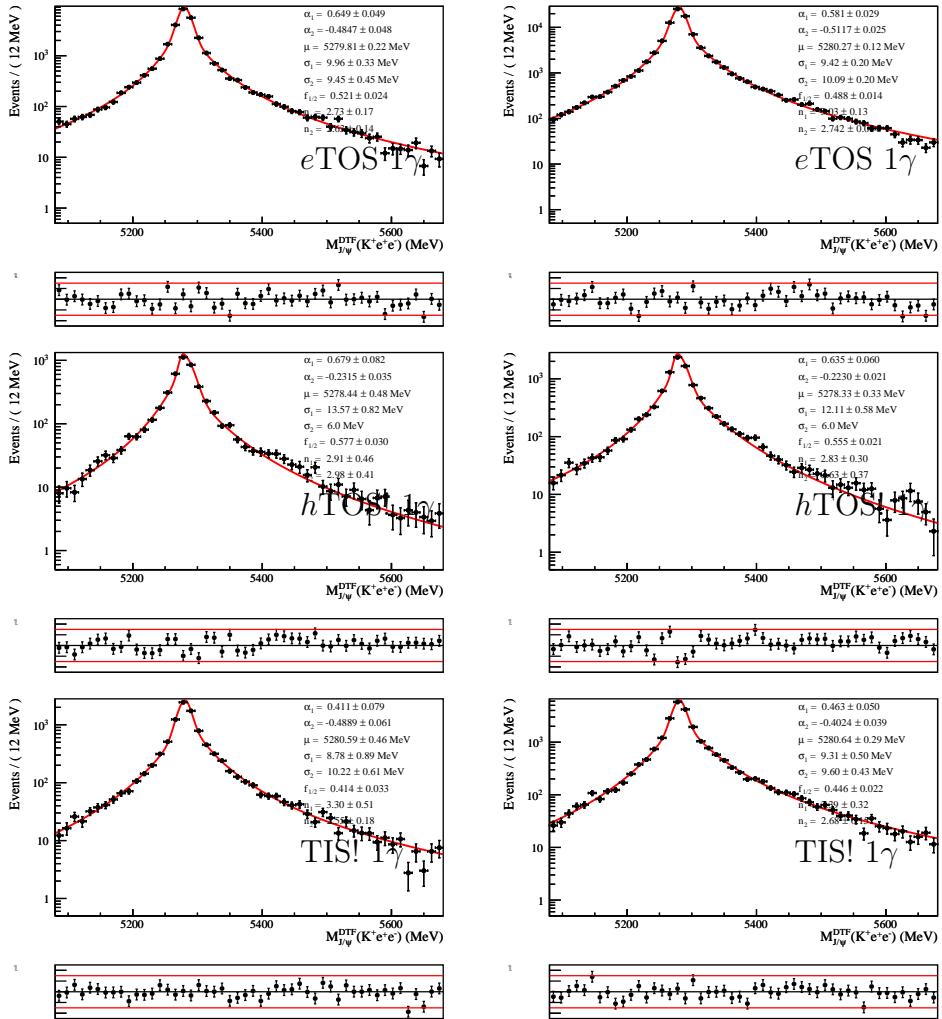


Figure 159: Fits to $m_{DTF}^{J/\psi}(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events in the bremsstrahlung category 1γ for run 1 (left) and run 2 (right), for eTOS (top), hTOS! and TIS! events.

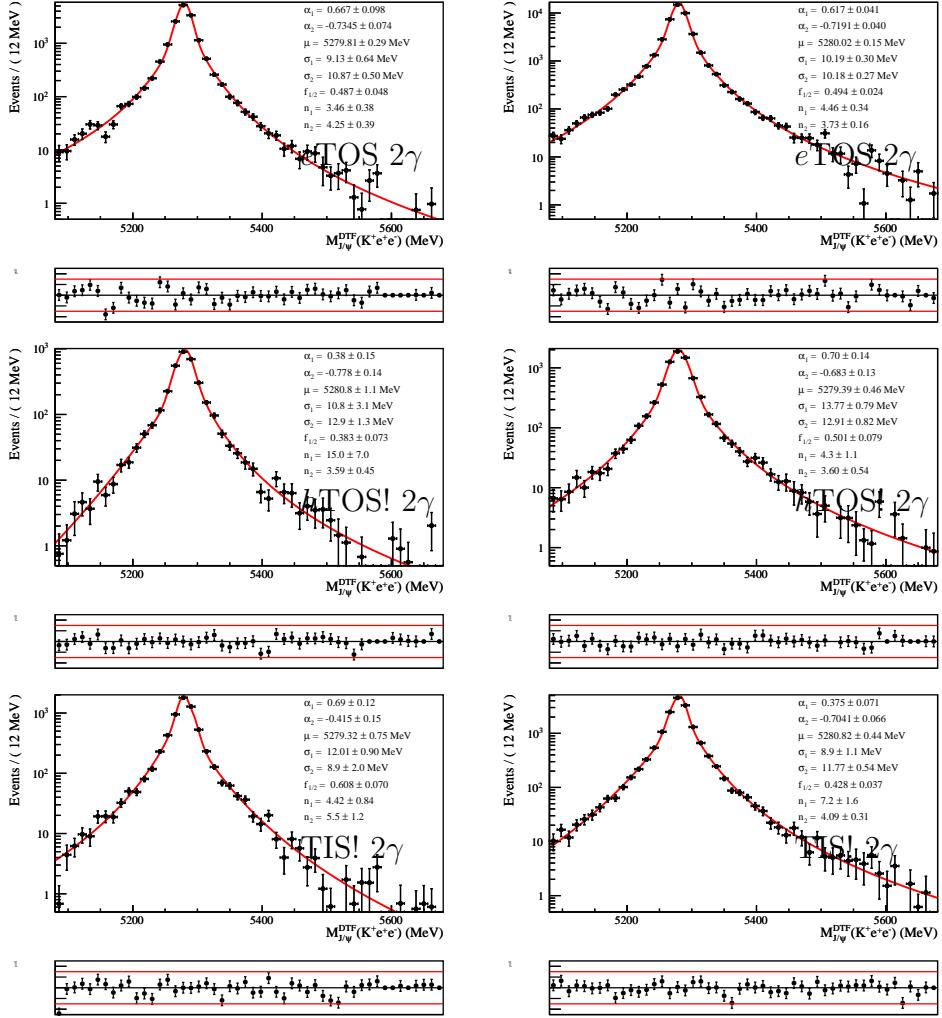


Figure 160: Fits to $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events in the bremsstrahlung category 2 γ for run 1 (left) and run 2 (right), for eTOS (top), hTOS! and TIS! events.

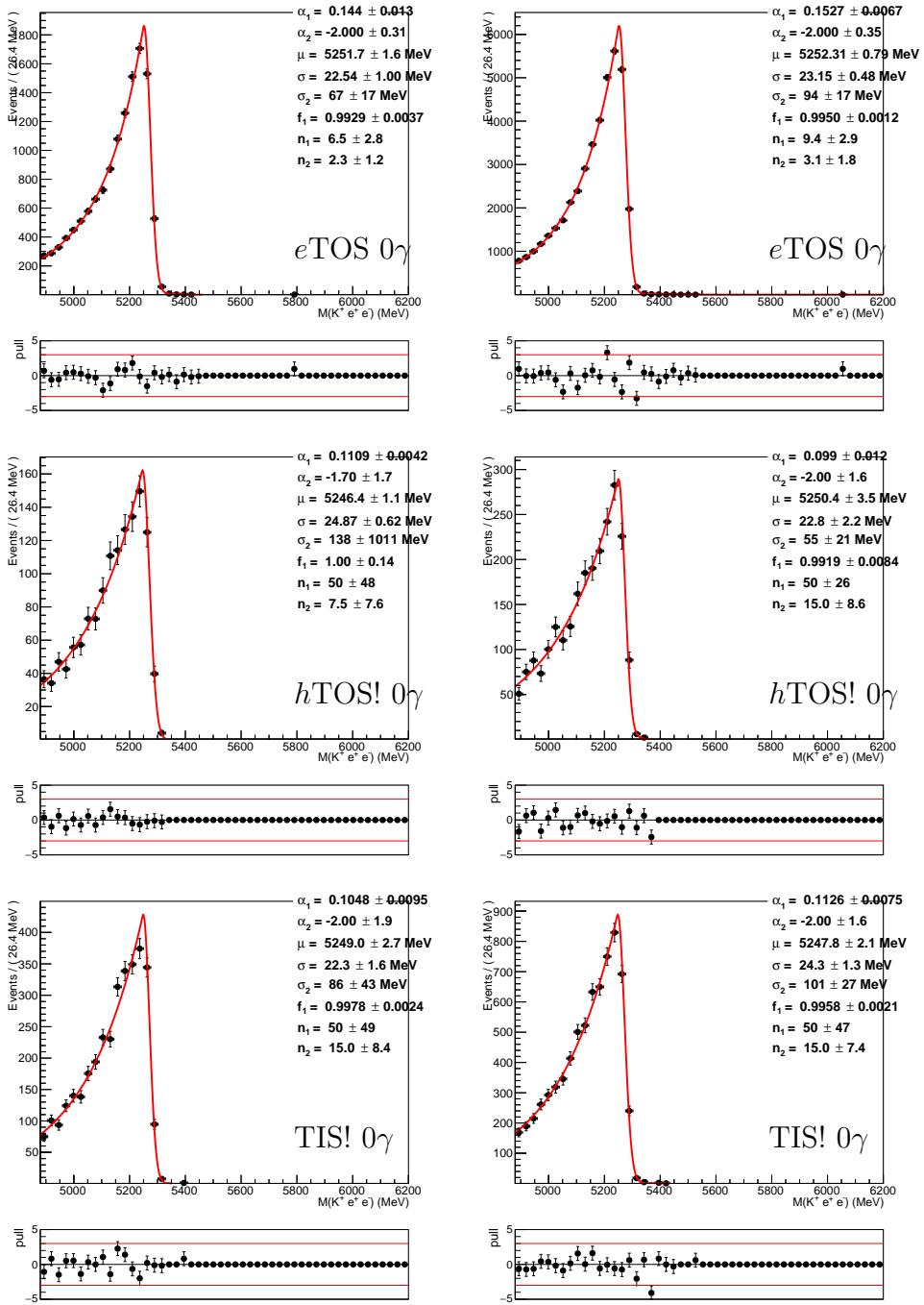


Figure 161: Fits to $m(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events in the bremsstrahlung category 0γ for run 1 (left) and run 2 (right), for $e\text{TOS}$ (top), $h\text{TOS!}$ and TIS! events.

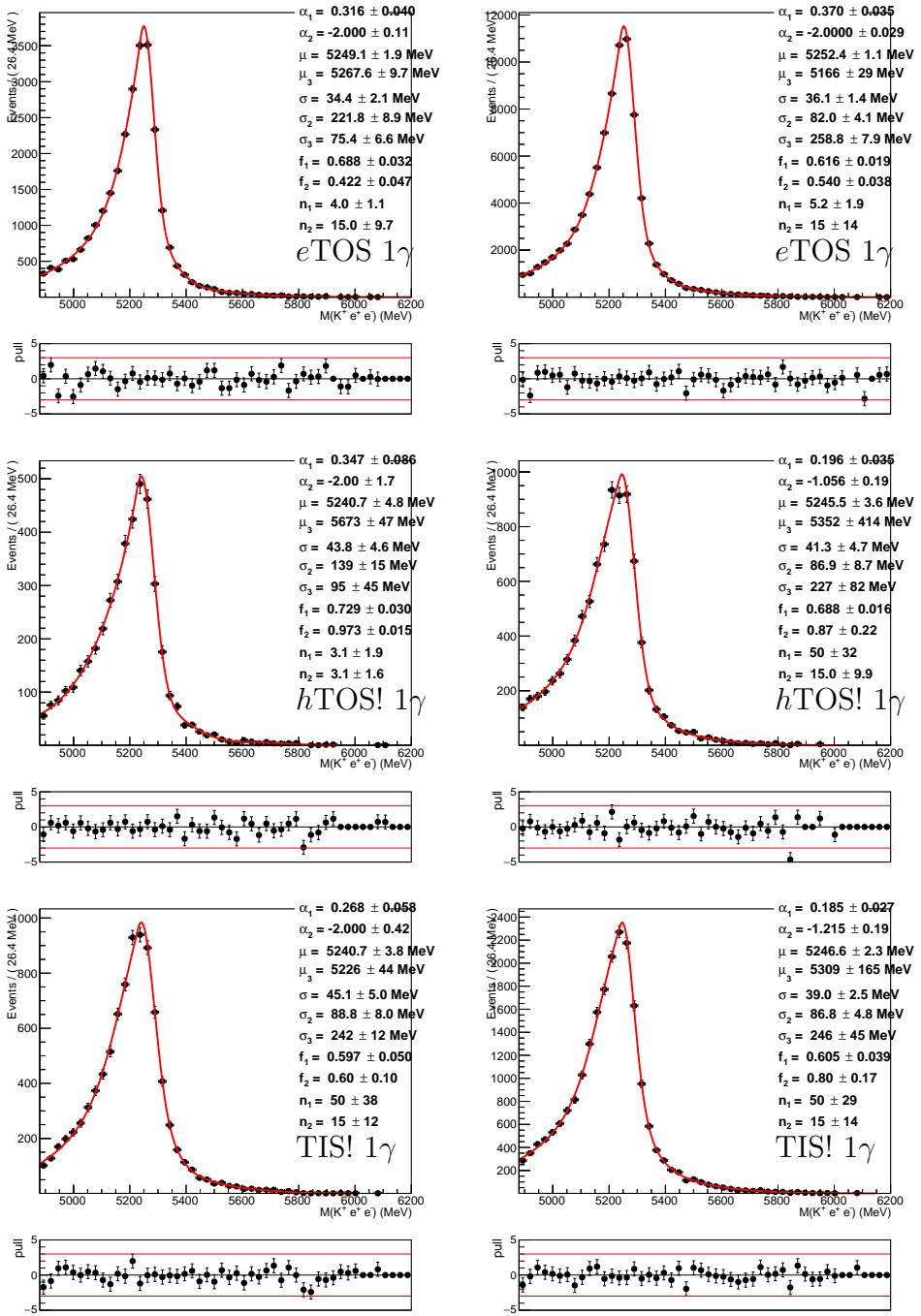


Figure 162: Fits to $m(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events in the bremsstrahlung category 1 γ for run 1 (left) and run 2 (right), for eTOS (top), hTOS! and TIS! events.

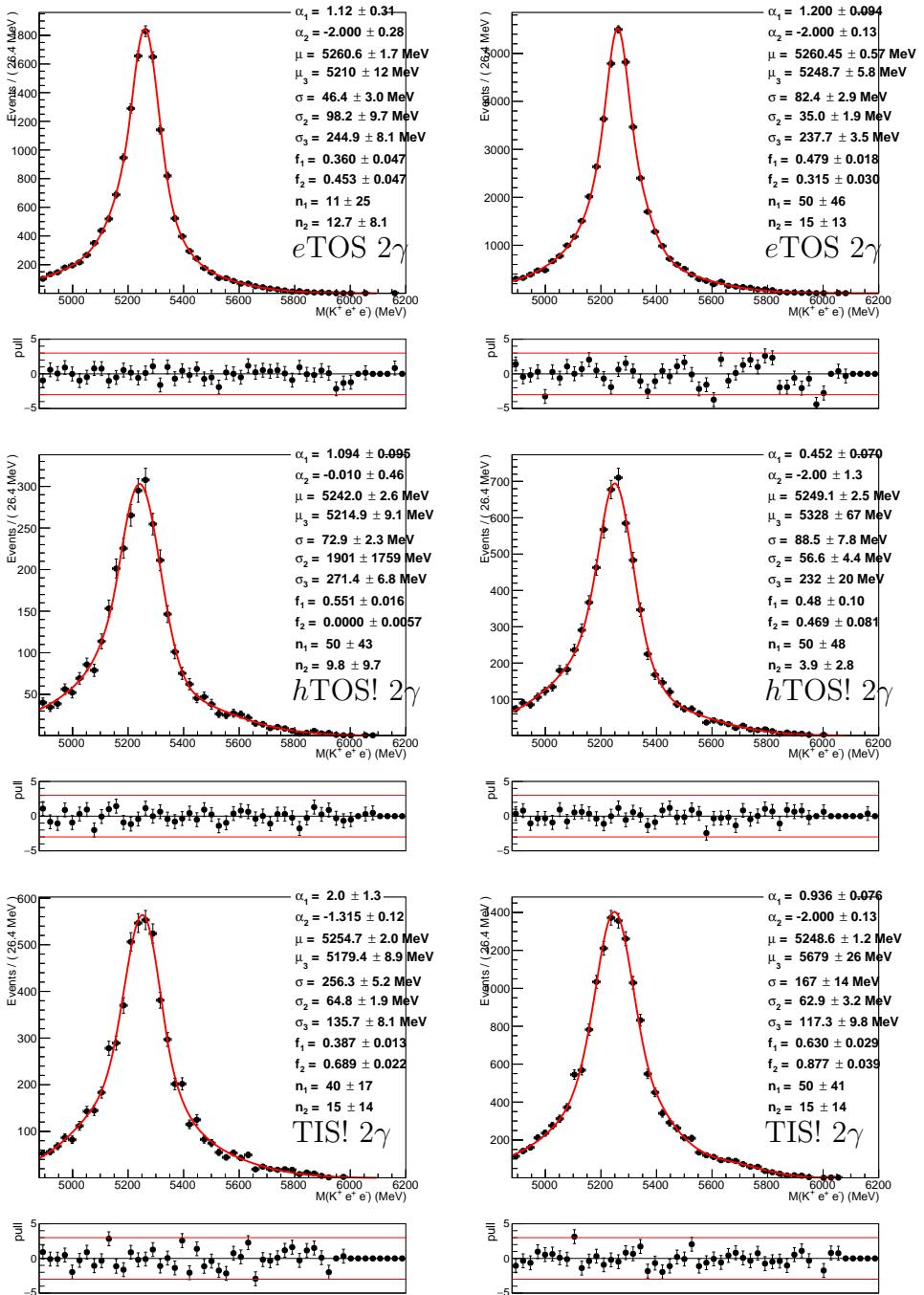


Figure 163: Fits to $m(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events in the bremsstrahlung category 2 γ for run 1 (left) and run 2 (right), for eTOS (top), hTOS! and TIS! events.

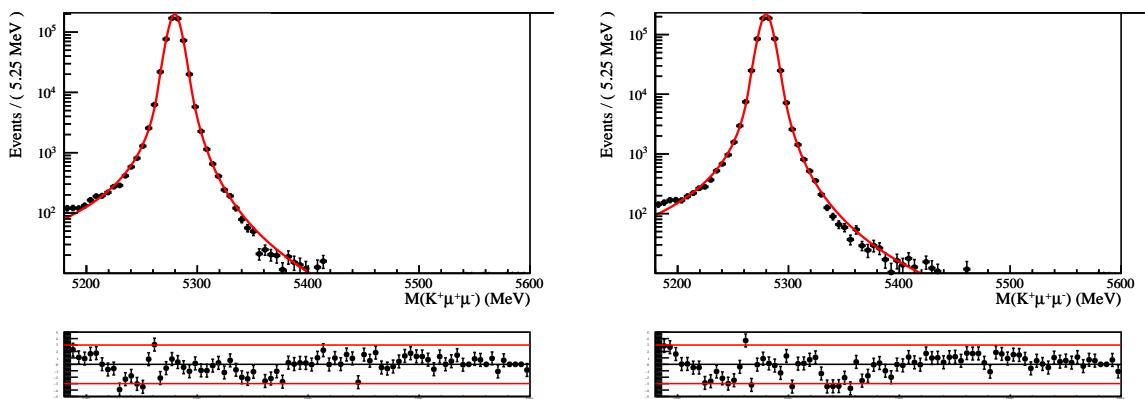


Figure 164: Fits to the $m_{\text{DTF}}^{\psi(2S)}(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+\psi(2S)(\mu^+\mu^-)$ simulated candidates in Run 1 (left) and Run 2 (right) samples.

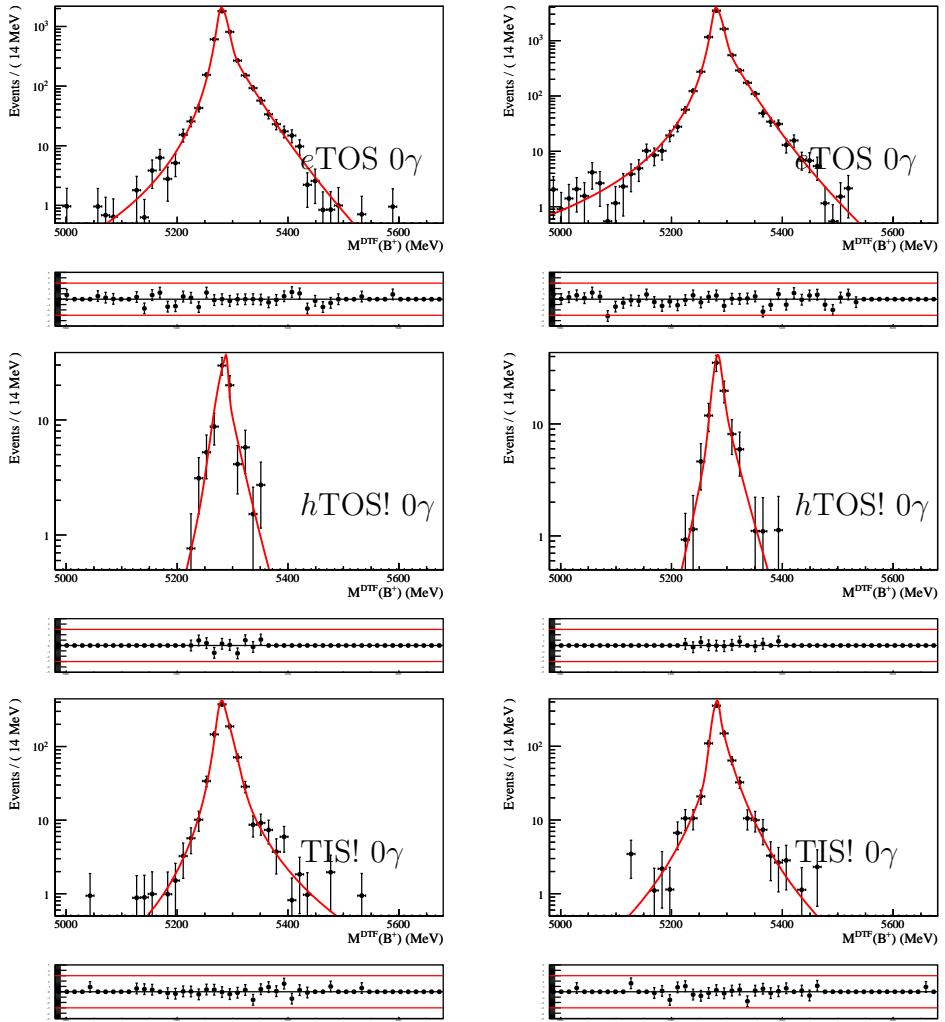


Figure 165: Fits to $m_{DTF}^{\psi(2S)}(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ simulated events in the bremsstrahlung category 0γ for run 1 (left) and run 2 (right), for $eTOS$ (top), $hTOS!$ and $TIS!$ events.

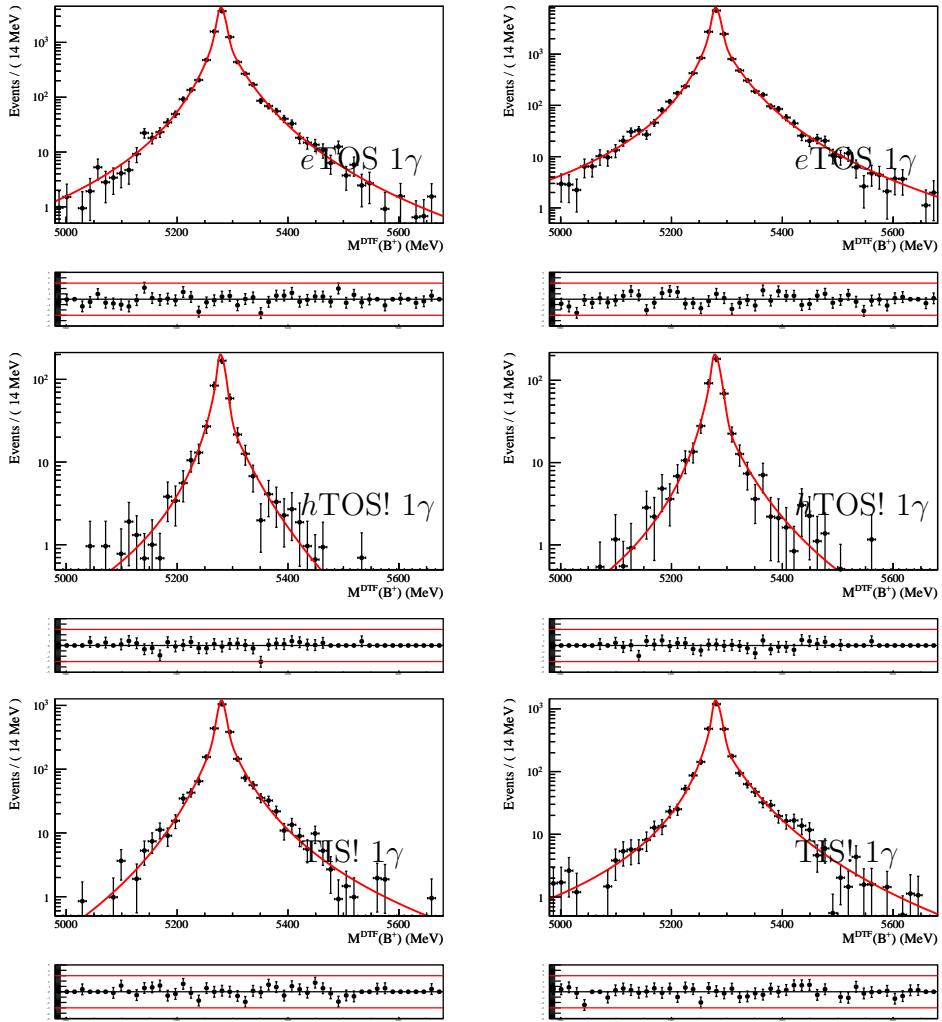


Figure 166: Fits to $m_{DTF}^{\psi(2S)}(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ simulated events in the bremsstrahlung category 1γ for run 1 (left) and run 2 (right), for $eTOS$ (top), $hTOS!$ and $TIS!$ events.

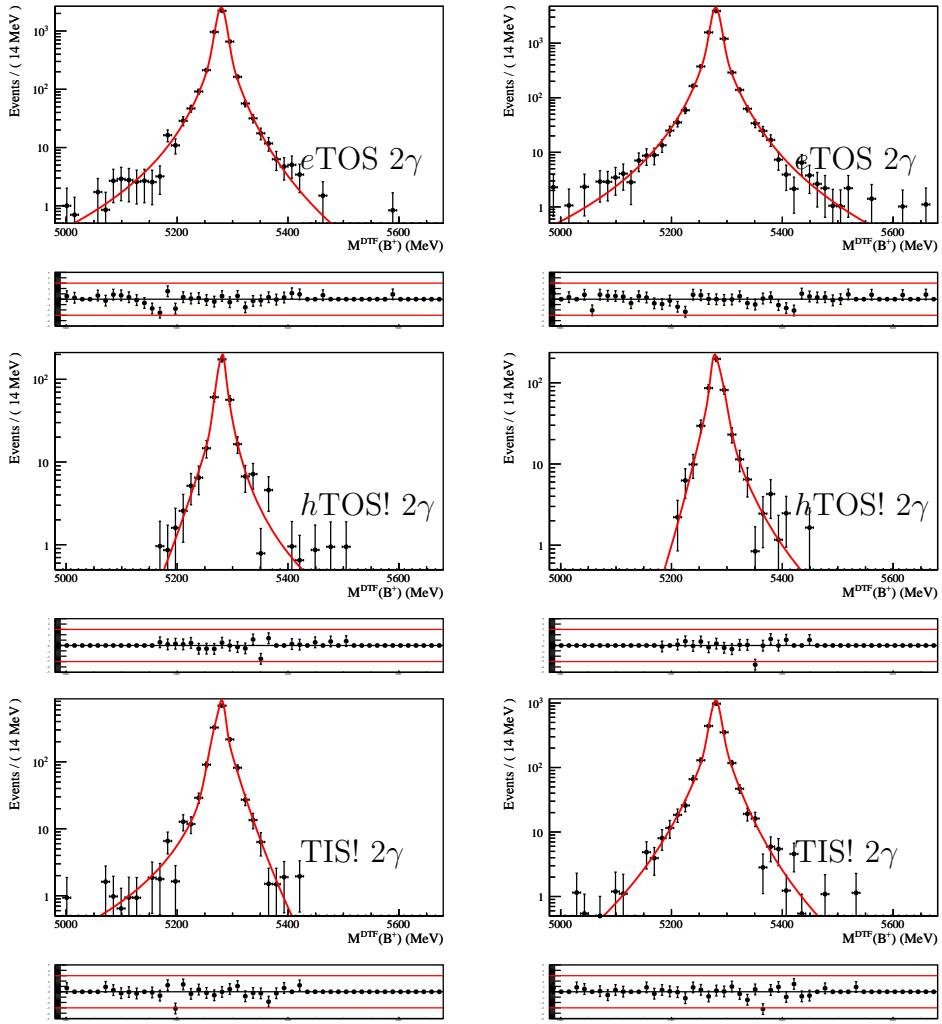


Figure 167: Fits to $m_{\text{DTF}}^{\psi(2S)}(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ simulated events in the bremsstrahlung category 2γ for run 1 (left) and run 2 (right), for $e\text{TOS}$ (top), $h\text{TOS!}$ and TIS! events.

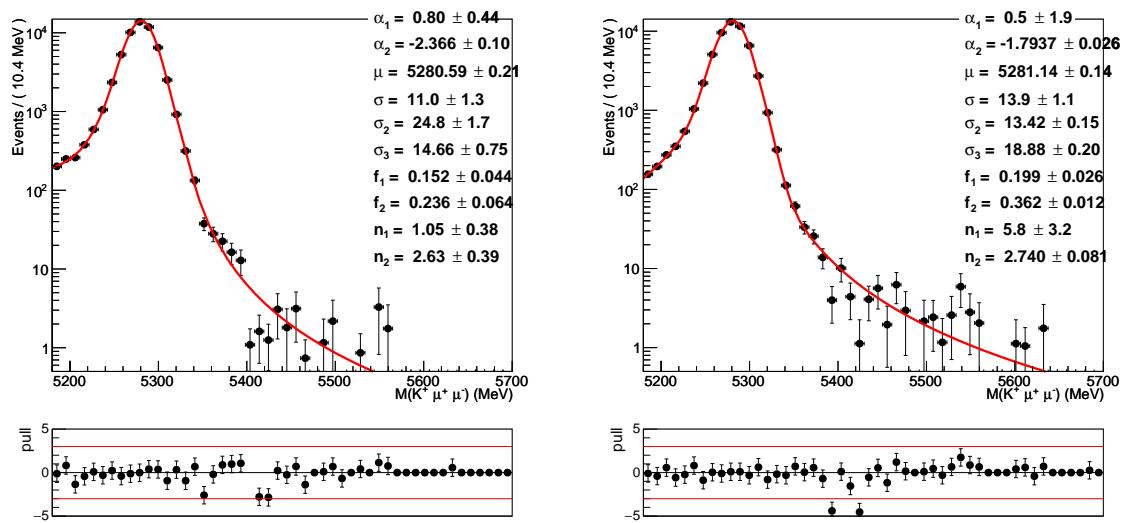


Figure 168: Fits to the $m(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+\mu^+\mu^-$ simulated candidates in Run 1 (left) and Run 2 (right) samples.

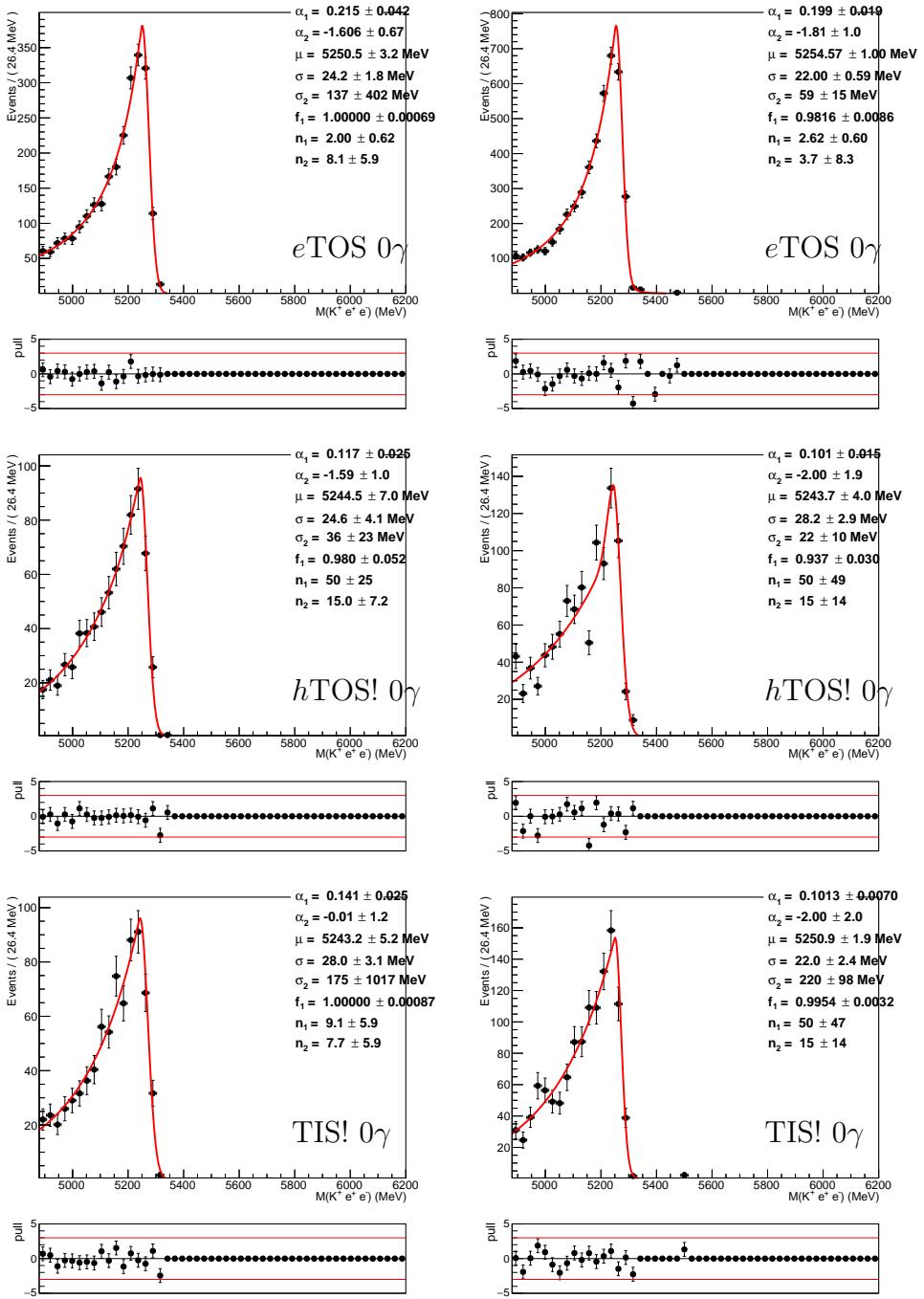


Figure 169: Fits to $m(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+e^+e^-$ simulated events in the bremsstrahlung category 0 γ for run 1 (left) and run 2 (right), for eTOS (top), hTOS! and TIS! events.

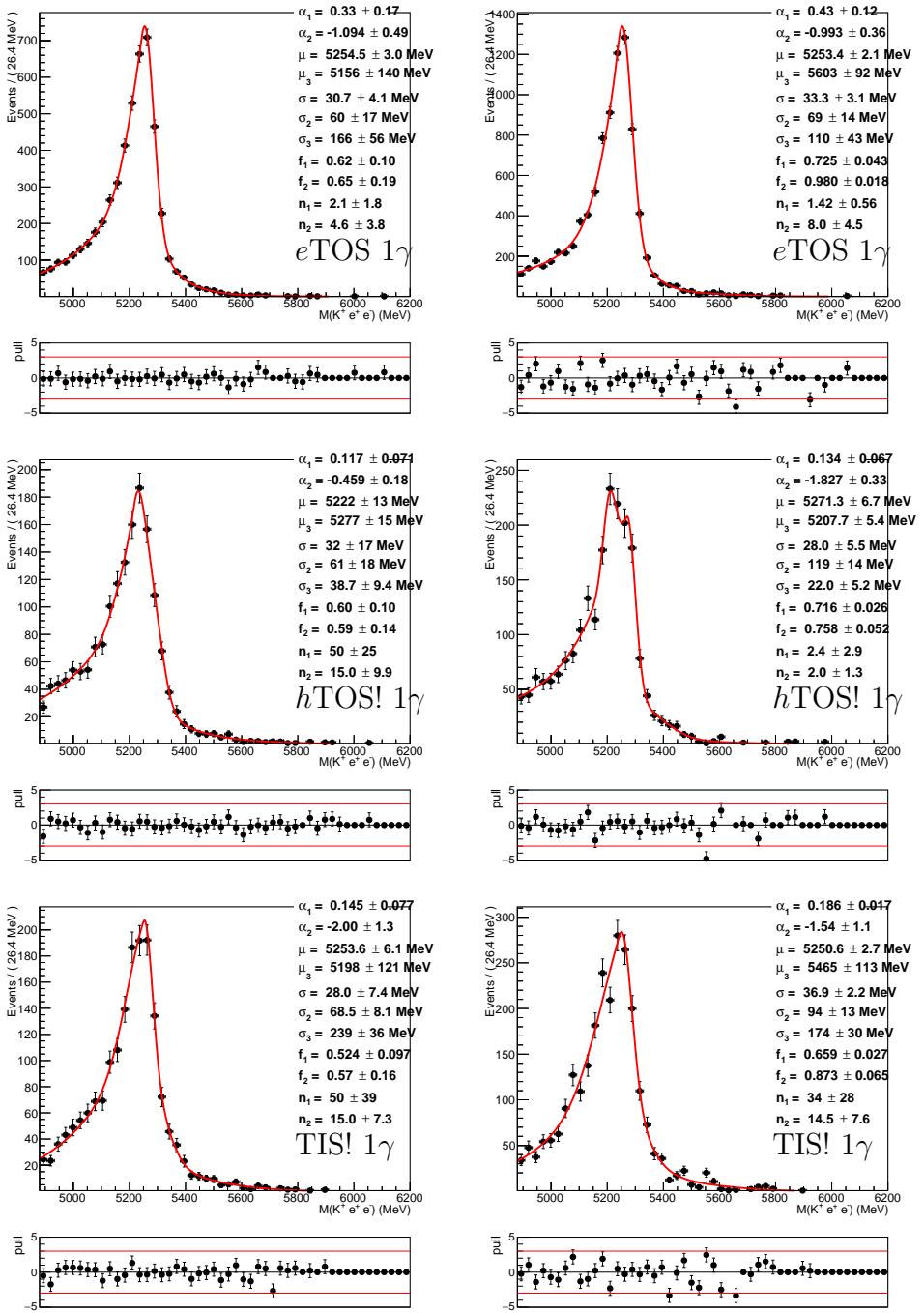


Figure 170: Fits to $m(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+e^+e^-$ simulated events in the bremsstrahlung category 1γ for run 1 (left) and run 2 (right), for $eTOS$ (top), $hTOS!$ and $TIS!$ events.

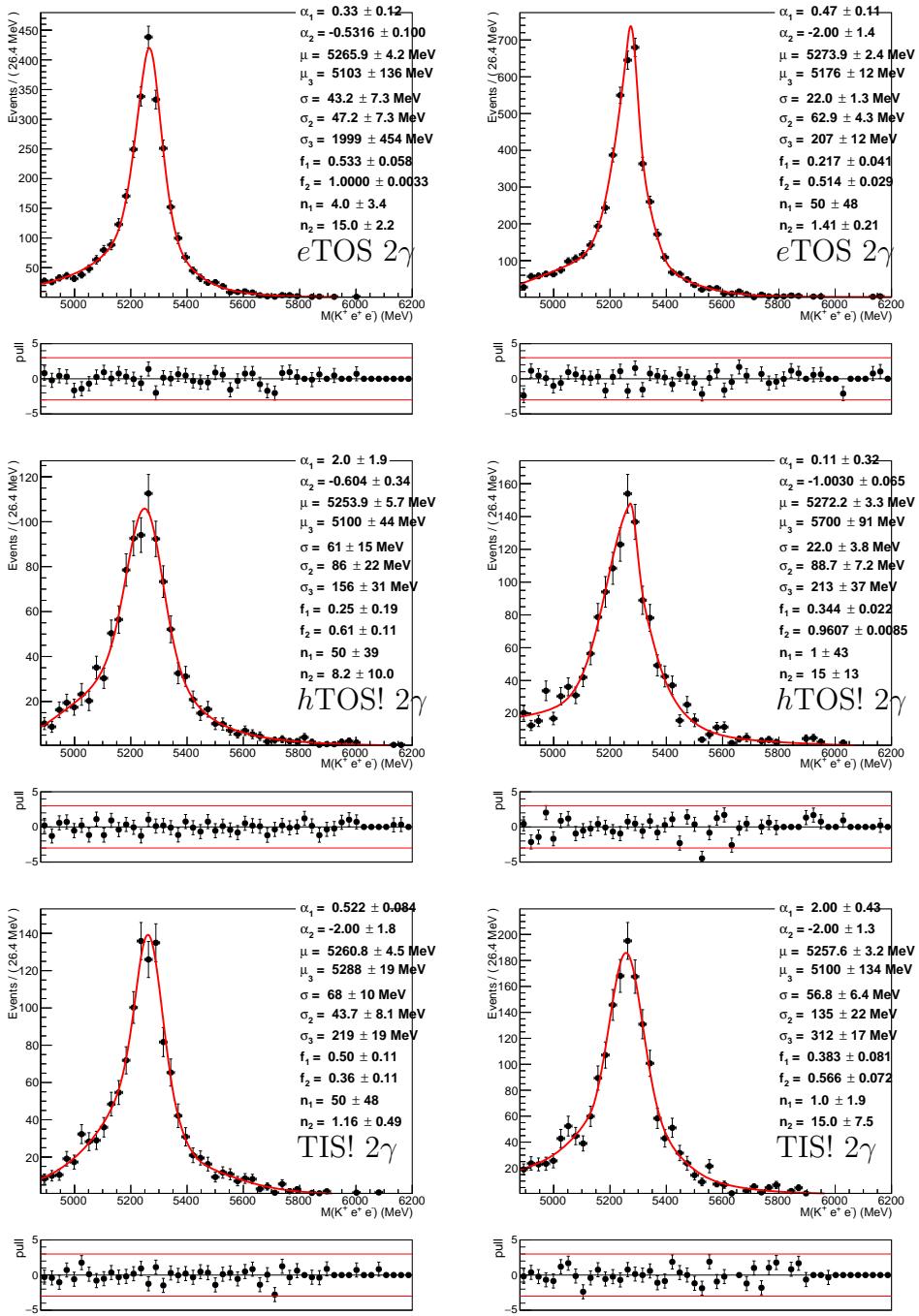


Figure 171: Fits to $m(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+e^+e^-$ simulated events in the bremsstrahlung category 2γ for run 1 (left) and run 2 (right), for eTOS (top), hTOS! and TIS! events.

2975 **H.2 Background fits**

2976 **H.2.1 Fits to $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ simulation**

2977 Fig. 172 shows the result of the fit to the $m_{\text{DTF}}^{J/\psi}(K^+\mu^+\mu^-)$ distribution of simulated
 2978 $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$ events after the full $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ selection chain. Equivalent
 2979 plots for the fits to the distribution in the unconstrained invariant mass, $m(K^+\mu^+\mu^-)$, are
 2980 shown in Fig. 173.

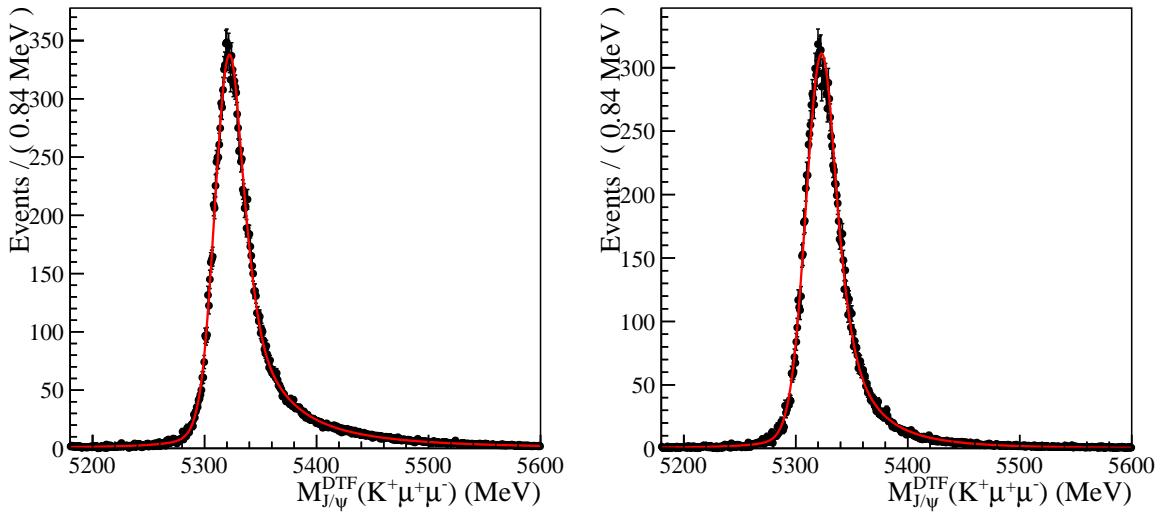


Figure 172: Fits to the $m_{\text{DTF}}^{J/\psi}(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ simulated candidates in Run 1 (left) and Run 2 (right) samples.

2981 **H.2.2 Fits to $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ simulation**

2982 As explained in Appendix C.3, the $m_{\text{misID}}^{J/\psi}$ mass distribution of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ is
 2983 used as a proxy for the shape of $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ events in the fit to the data.
 2984 Fig. 174 shows the result of the fit to this distribution. Equivalent plots for the fits to the
 2985 distributions in the unconstrained invariant mass, m_{misID} , are shown in Fig. 175.

2986 **H.2.3 Fits to $B^+ \rightarrow J/\psi(e^+e^-)X$ simulation**

2987 Partially reconstructed background from $B^+ \rightarrow J/\psi(e^+e^-)X$ is separated into two com-
 2988 ponents depending on the part of the decay that has been partially reconstructed:
 2989 higher charm resonances ($B \rightarrow KX_c(J/\psi X)$) and higher strange hadronic systems
 2990 ($B \rightarrow H_s(KX)J/\psi(ee)$). A `RooKeysPdf` function is used to describe these two com-
 2991 ponents and they are finally combined in the fit to the data as explained in the main
 2992 text. Fig. 176 shows the result of the fit to the $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-)$ distribution of the charm

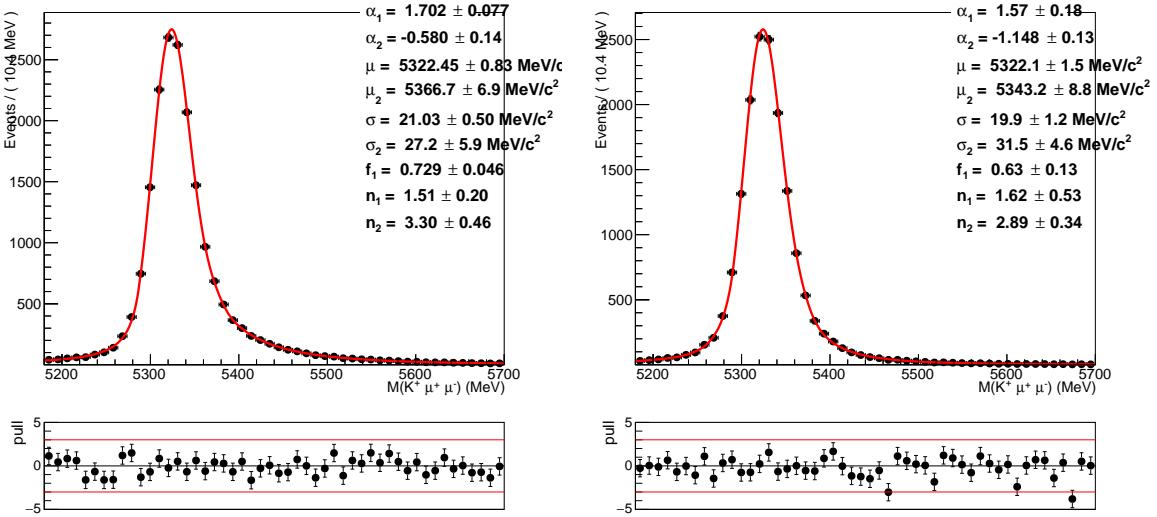


Figure 173: Fits to the $m(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ simulated candidates in Run 1 (left) and Run 2 (right) samples.

component of $B^+ \rightarrow J/\psi(e^+e^-)X$ simulated events after the full $B^+ \rightarrow K^+J/\psi(e^+e^-)$ selection chain, whereas the strange component is shown in Fig. 178. Due to limited statistics, a global distribution for the combination of all the photon categories is used. Equivalent plots for the fits to events selected using the $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ selection are shown in Fig. 177 and Fig. 179.

2998 H.2.4 Fits to $B^0 \rightarrow K^{*0}e^+e^-$ simulation

2999 Fig. 180 shows the result of the fit to the $m(K^+e^+e^-)$ distribution of simulated $B^0 \rightarrow$
3000 $K^{*0}e^+e^-$ events after the full $B^+ \rightarrow K^+e^+e^-$ selection chain. A global distribution for the
3001 combination of all the photon categories is used.

3002 H.2.5 Leakage from $B^+ \rightarrow K^+J/\psi(e^+e^-)$ to $B^+ \rightarrow K^+e^+e^-$

3003 Fig. 181 shows the result of the fit to the $m(K^+e^+e^-)$ distribution of simulated $B^+ \rightarrow$
3004 $K^+J/\psi(e^+e^-)$ events after the full $B^+ \rightarrow K^+e^+e^-$ selection chain. Only a handful of
3005 $B^+ \rightarrow K^+J/\psi(e^+e^-)$ events leak into the rare mode signal region, which reflects in the
3006 low statistics on these plots. In the fit to the rare mode, the size of this contribution is
3007 expected to be small and is constrained from the signal yield observed in the fit to the
3008 $B^+ \rightarrow K^+J/\psi(e^+e^-)$ normalisation mode. A global distribution for the combination of
3009 all the photon categories is used.

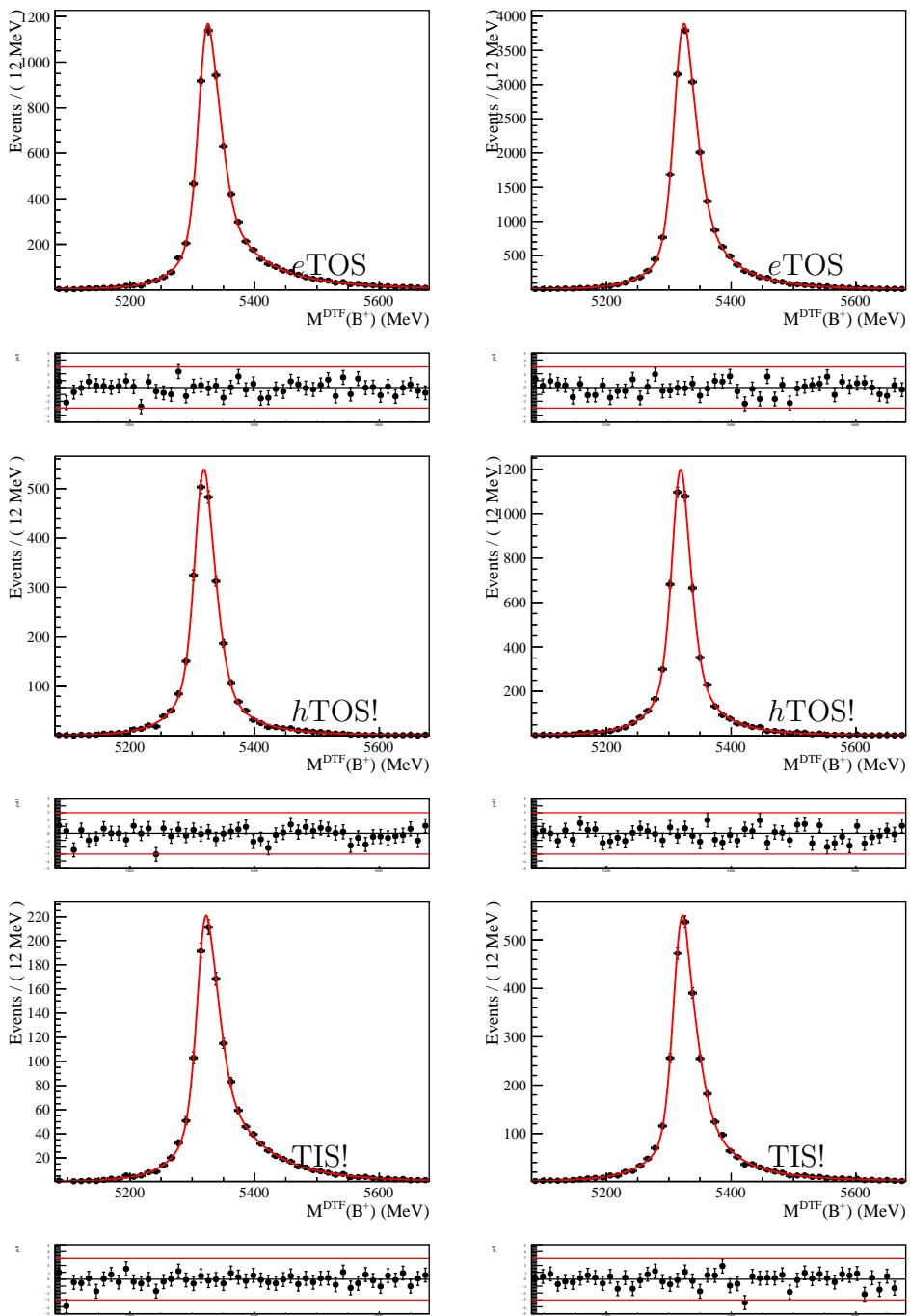


Figure 174: Distribution in $m_{\text{misID}}^{J/\psi}$ of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events after the full $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ selection chain for run 1 (left) and run 2 (right). Trigger categories $e\text{TOS}$ (top), $h\text{TOS!}$ and TIS! are shown separately.

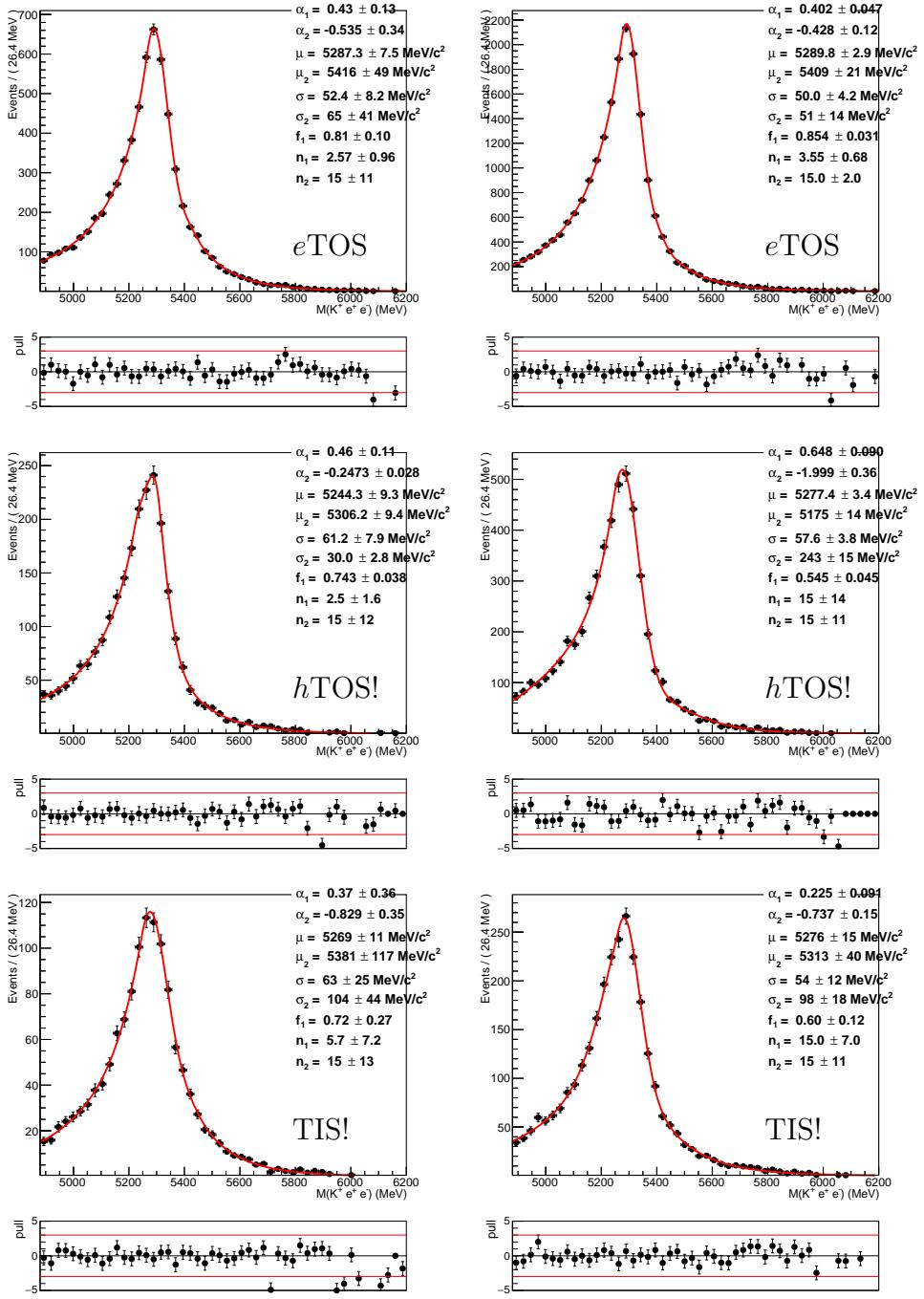


Figure 175: Distribution in m_{misID} of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events after the full $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ selection chain for run 1 (left) and run 2 (right). Trigger categories eTOS (top), hTOS! and TIS! are shown separately.

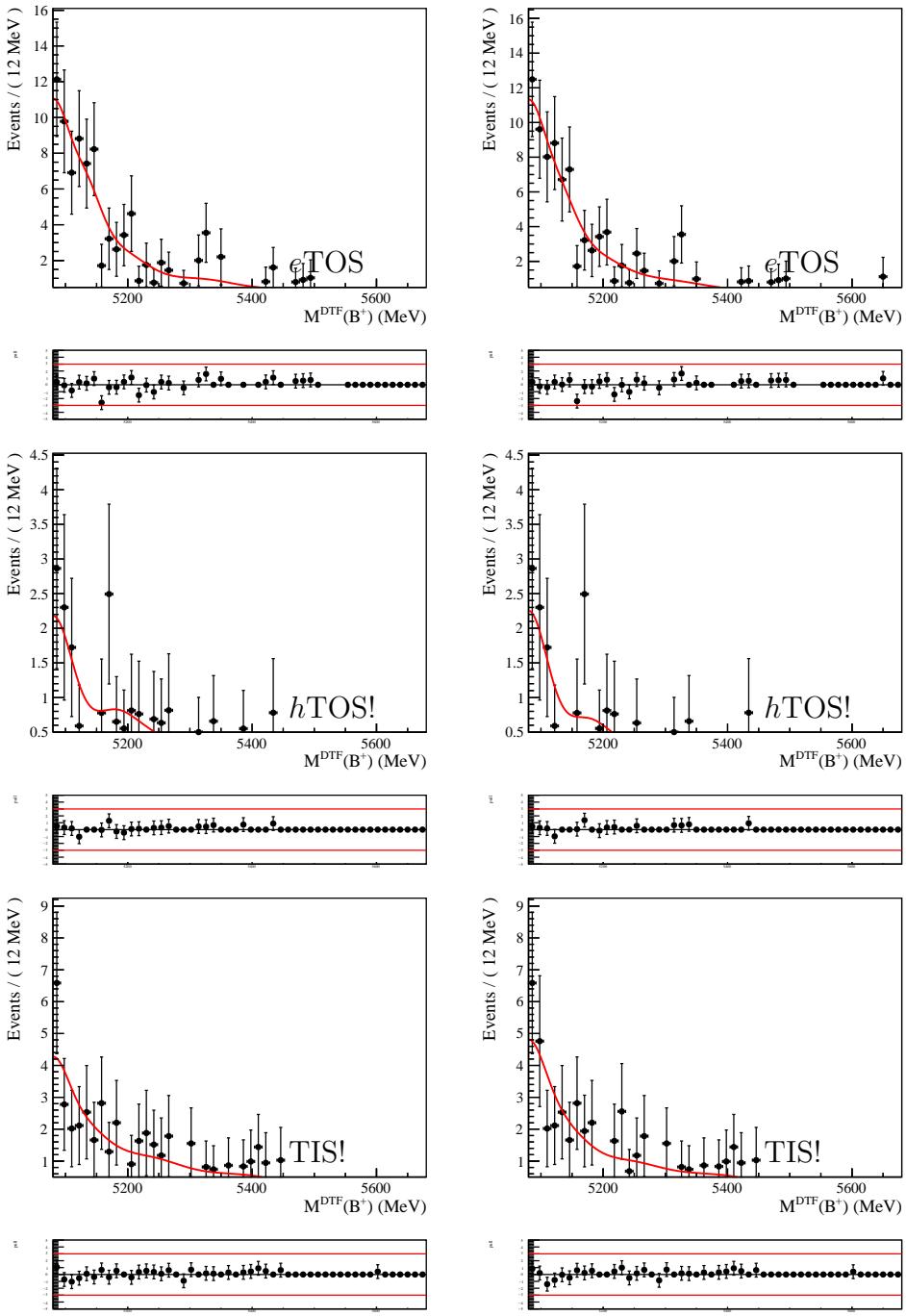


Figure 176: Distribution in $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-)$ of the charm component of $B^+ \rightarrow J/\psi(e^+e^-)X$ simulated events after the full $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ selection chain for run 1 (left) and run 2 (right). Trigger categories $e\text{TOS}$ (top), $h\text{TOS}!$ and $\text{TIS}!$ are shown separately.

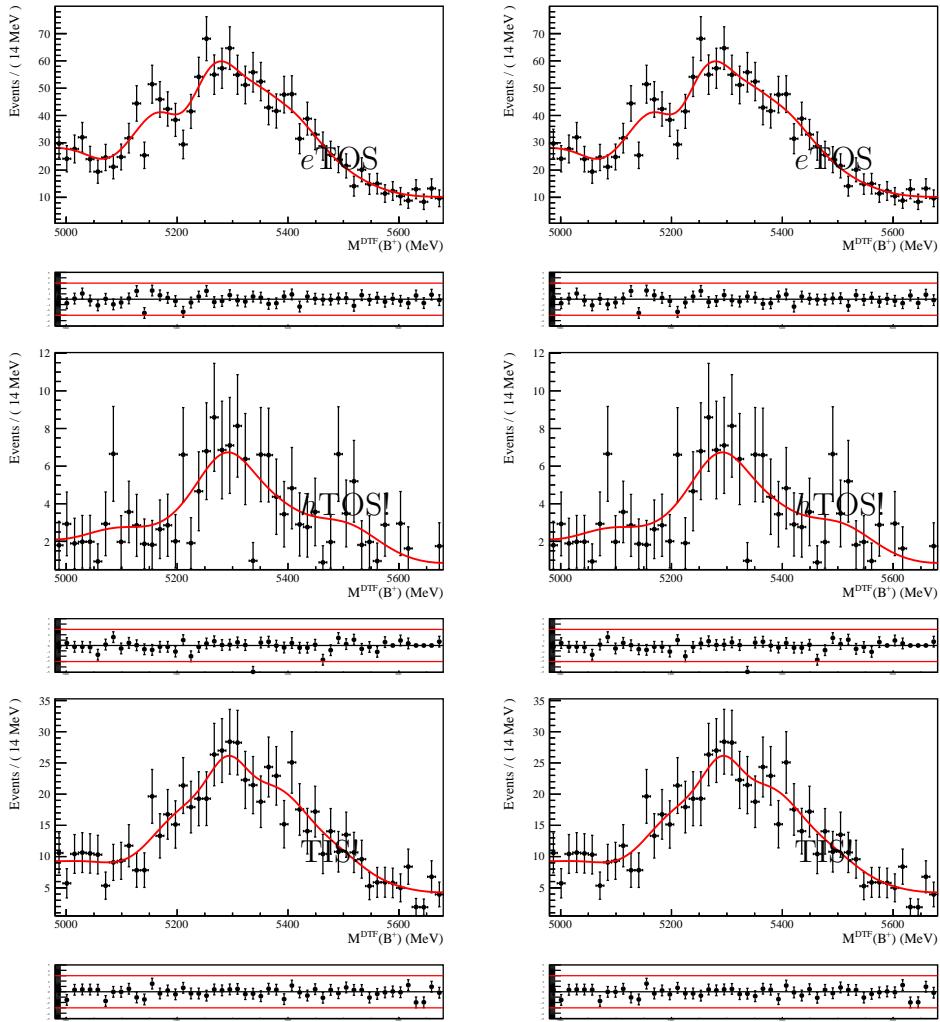


Figure 177: Distribution in $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-)$ of the charm component of $B^+ \rightarrow J/\psi(e^+e^-)X$ simulated events after the full $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ selection chain for run 1 (left) and run 2 (right). Trigger categories $e\text{TOS}$ (top), $h\text{TOS!}$ and TIS! are shown separately.

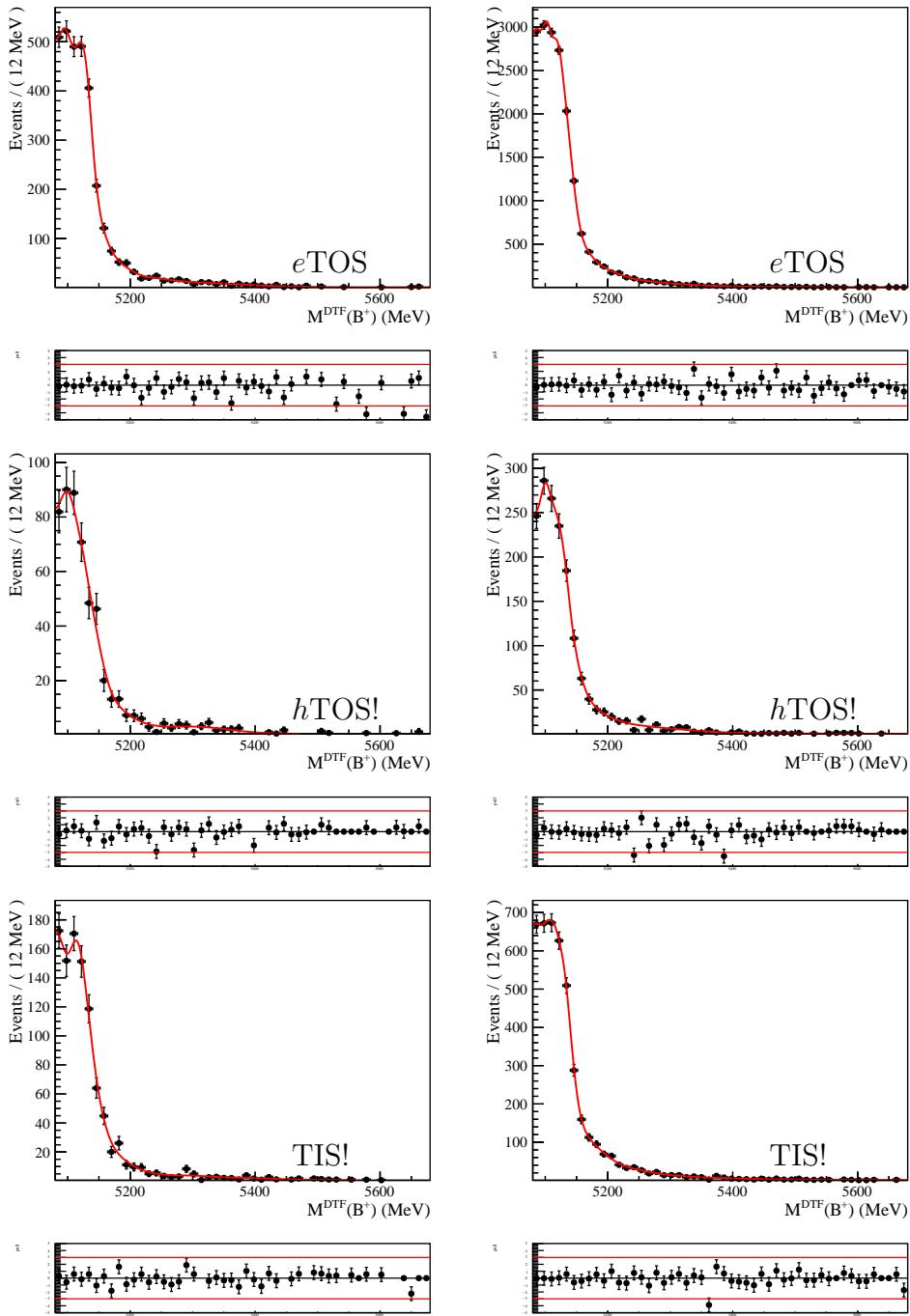


Figure 178: Distribution in $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-)$ of the strange component of $B^+ \rightarrow J/\psi(e^+e^-)X$ simulated events after the full $B^+ \rightarrow K^+J/\psi(e^+e^-)$ selection chain for run 1 (left) and run 2 (right). Trigger categories $e\text{TOS}$ (top), $h\text{TOS!}$ and TIS! are shown separately.

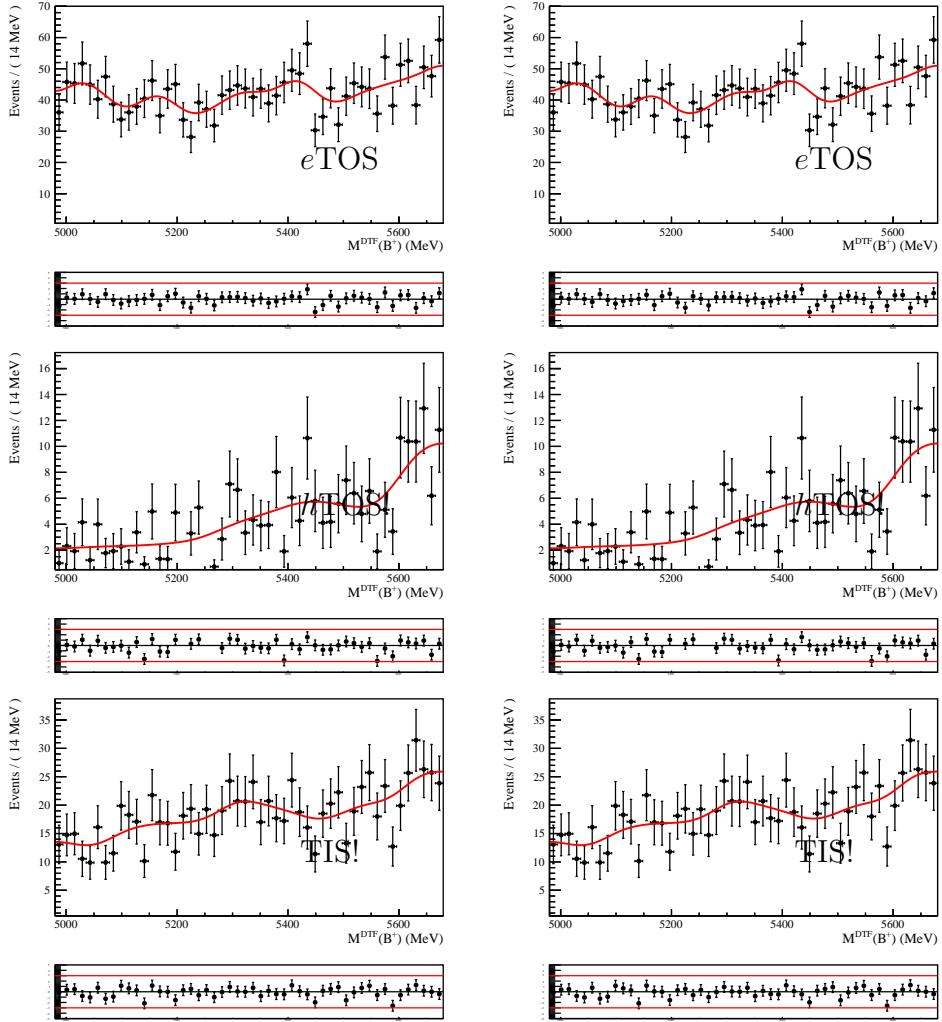


Figure 179: Distribution in $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-)$ of the strange component of $B^+ \rightarrow J/\psi(e^+e^-)X$ simulated events after the full $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ selection chain for run 1 (left) and run 2 (right). Trigger categories $e\text{TOS}$ (top), $h\text{TOS}!$ and $\text{TIS}!$ are shown separately.

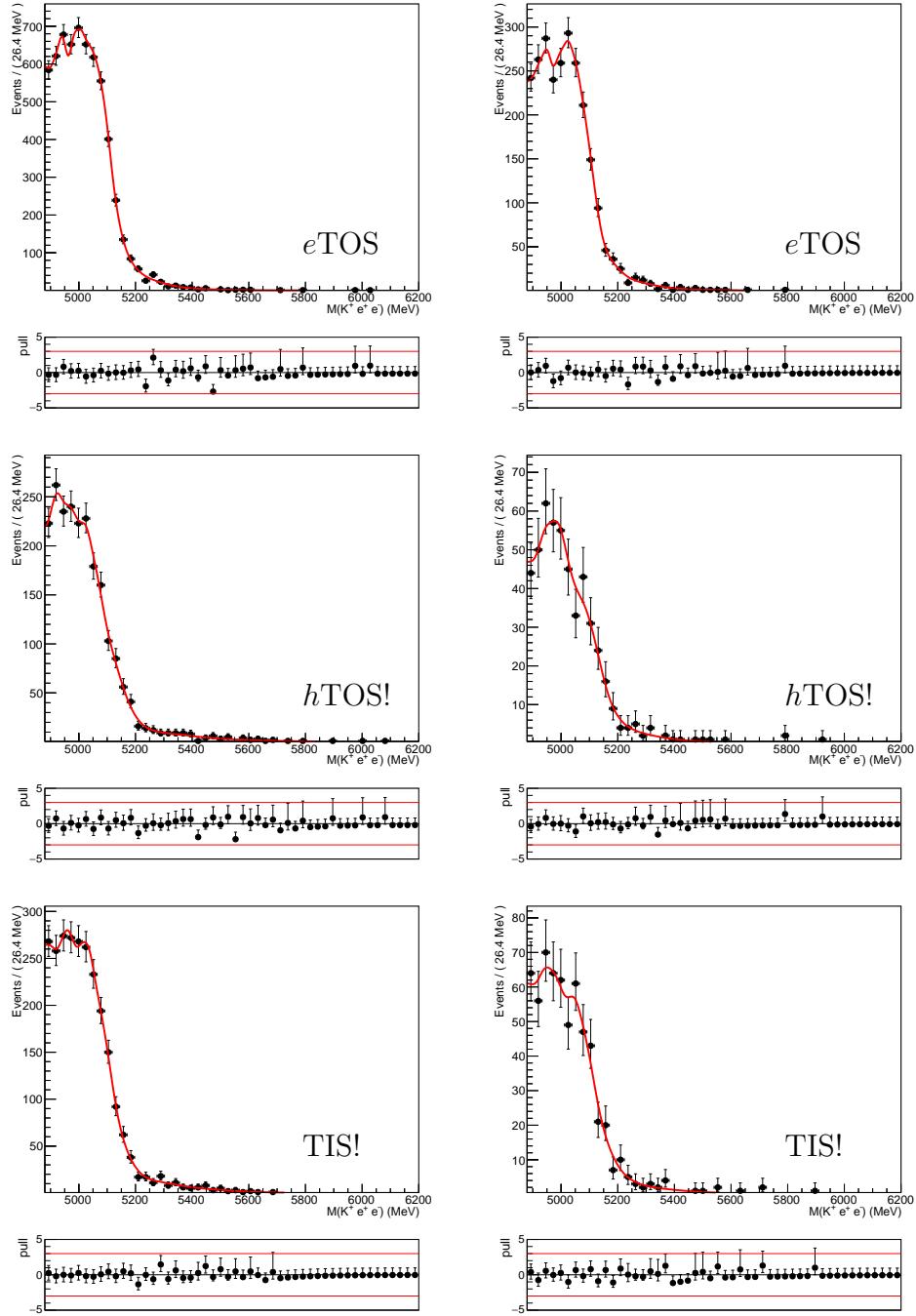


Figure 180: Distribution in $m(K^+e^+e^-)$ of $B^0 \rightarrow K^{*0}e^+e^-$ simulated events after the full $B^+ \rightarrow K^+e^+e^-$ selection chain for run 1 (left) and run 2 (right). Trigger categories eTOS (top), hTOS! and TIS! are shown separately.

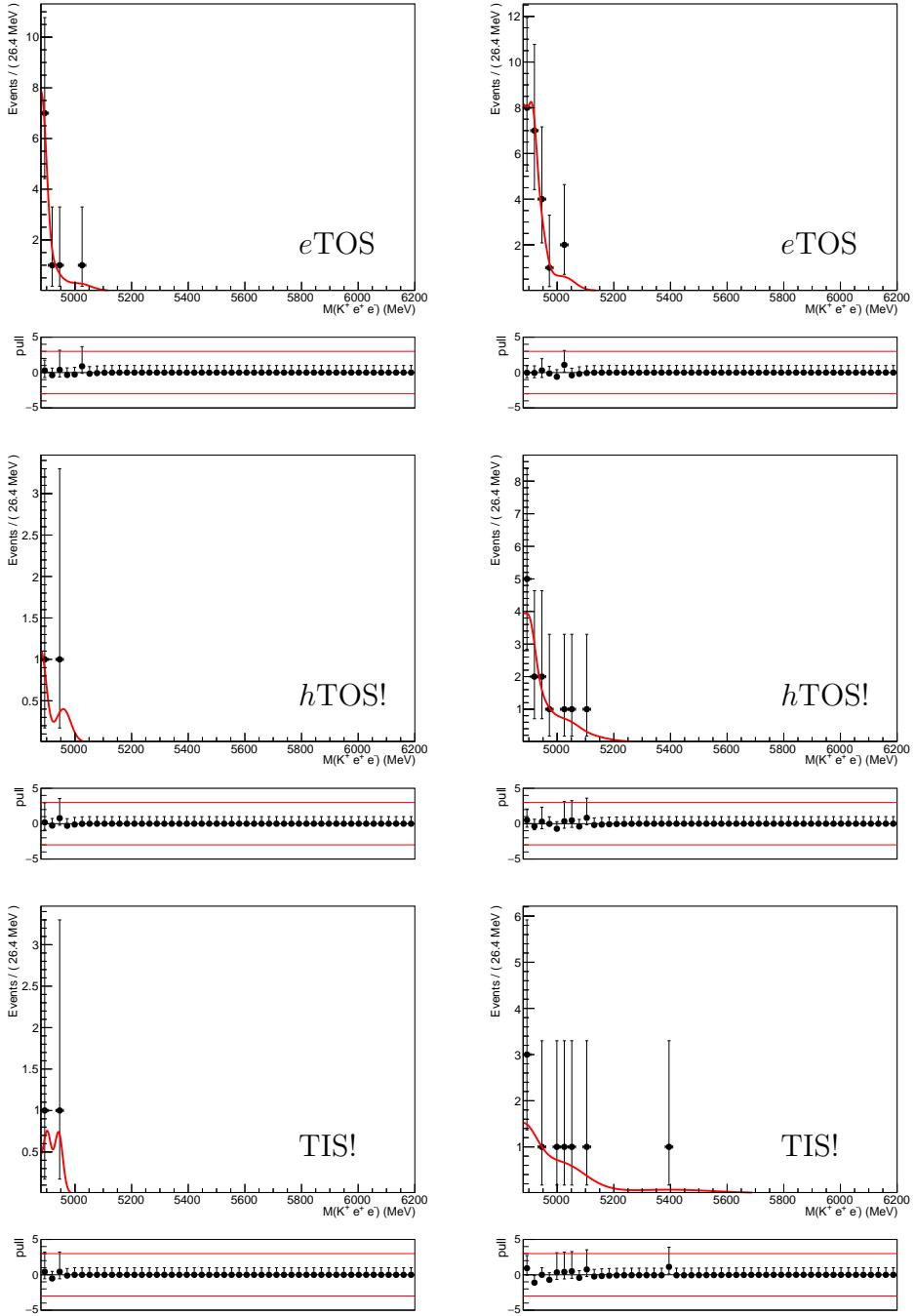


Figure 181: Distribution in $m(K^+e^+e^-)$ of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events after the full $B^+ \rightarrow K^+e^+e^-$ selection chain, with the exception of the BDT requirement, for run 1 (left) and run 2 (right). Trigger categories $e\text{TOS}$ (top), $h\text{TOS!}$ and TIS! are shown separately.

3010 **H.3 Fit model systematic**

3011 **H.3.1 Alternative signal models**

3012 The results of the fits to the $m(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+e^+e^-$ simulated events
3013 using the alternative signal models described in Section 10.9 are shown in Fig. 182, Fig. 183
3014 and Fig. 184 for events in the 0γ , 1γ and 2γ categories.

3015 **H.3.2 Alternative background models**

3016 The variation in the shape of the partially reconstructed background when the contributions
3017 from $B \rightarrow K^{**}e^+e^-$, where $K^{**} \equiv \{K_1, K_2^{*0(+)}\}$, are included is shown in Fig. 185.

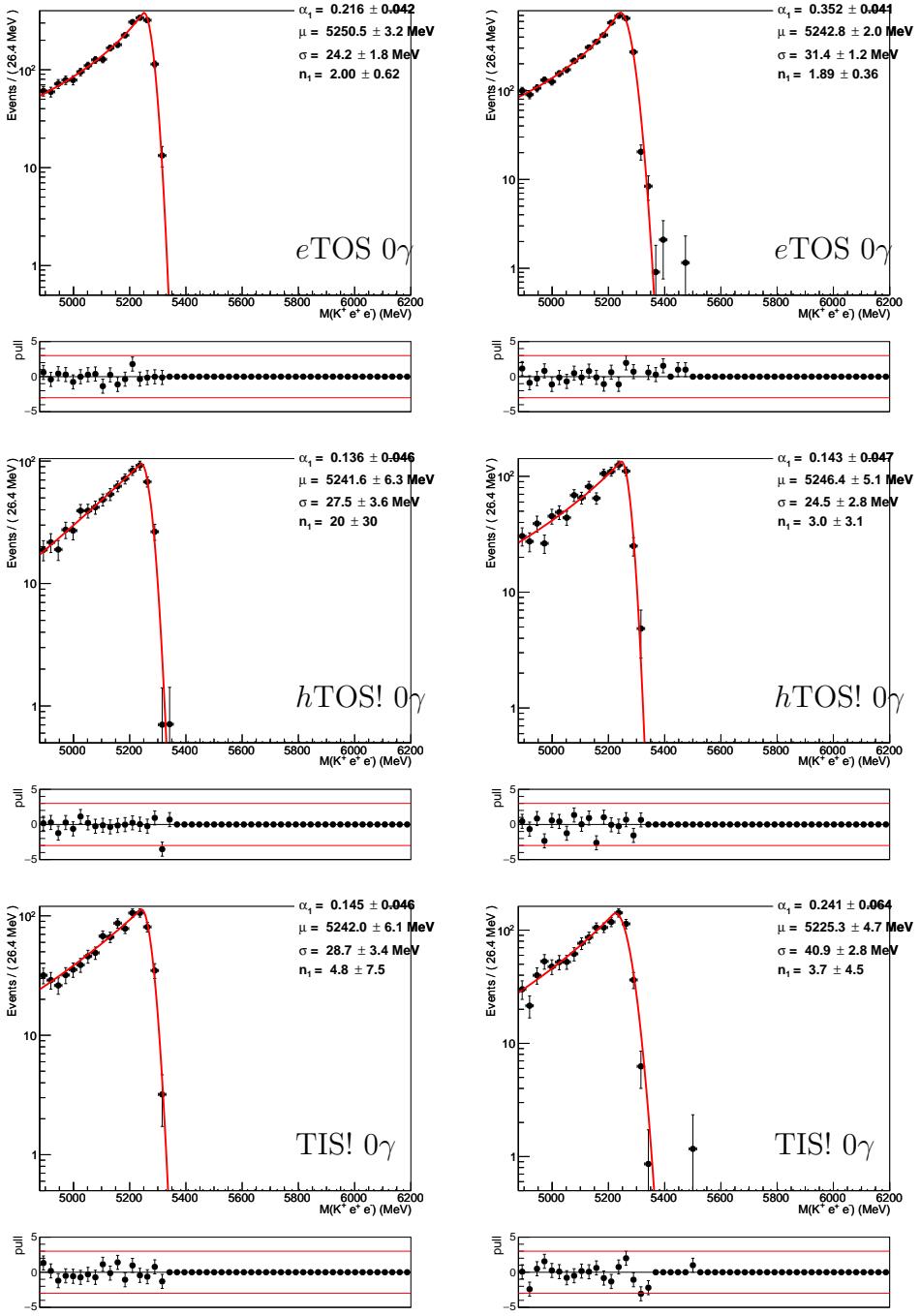


Figure 182: Fits to $m(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+e^+e^-$ simulated events in the bremsstrahlung category 0γ for run 1 (left) and run 2 (right), for $eTOS$ (top), $hTOS!$ (middle) and $TIS!$ (bottom) events.

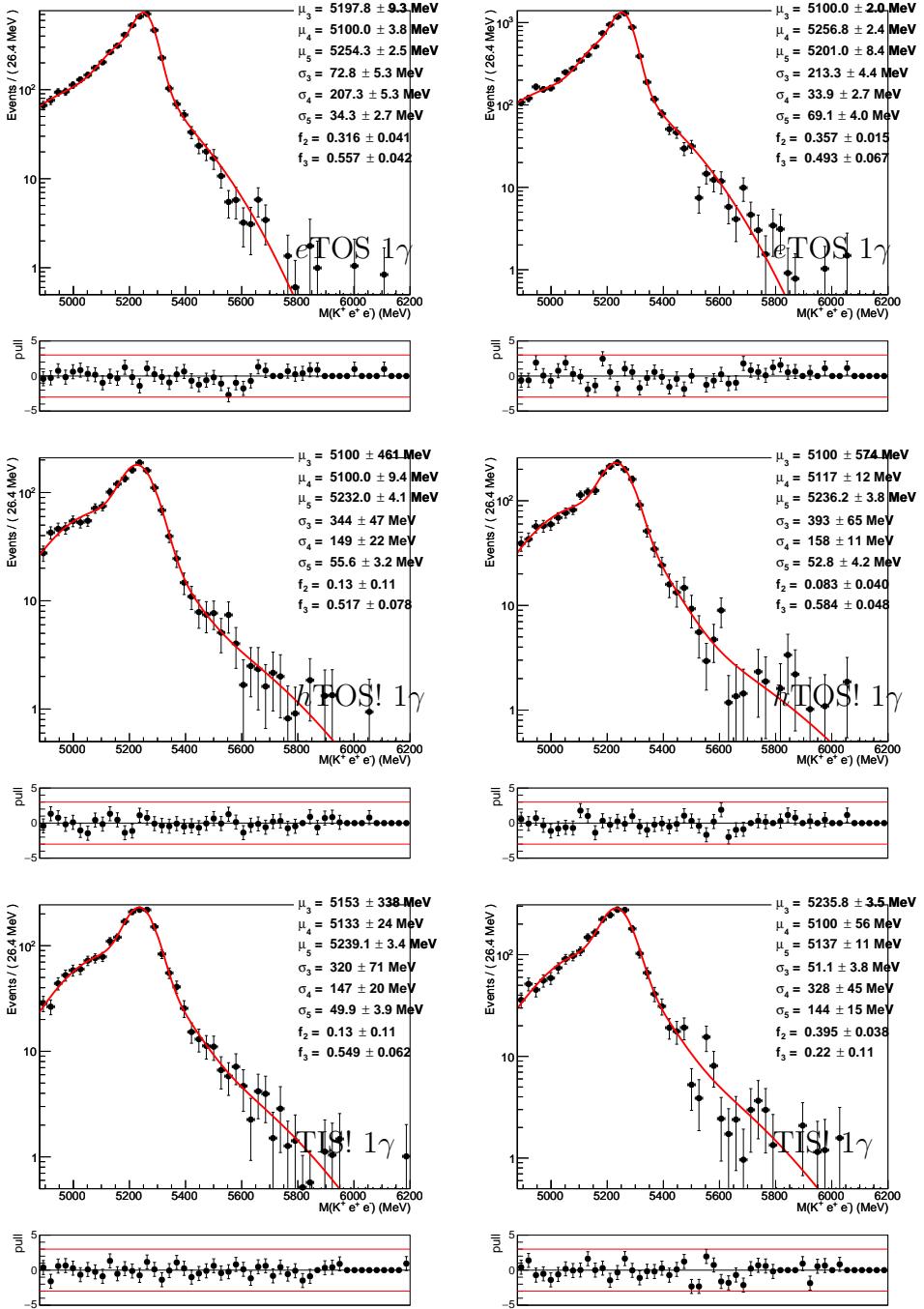


Figure 183: Fits to $m(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+e^+e^-$ simulated events in the bremsstrahlung category 1γ for run 1 (left) and run 2 (right), for $e\text{TOS}$ (top), $h\text{TOS!}$ (middle) and TIS! (bottom) events.

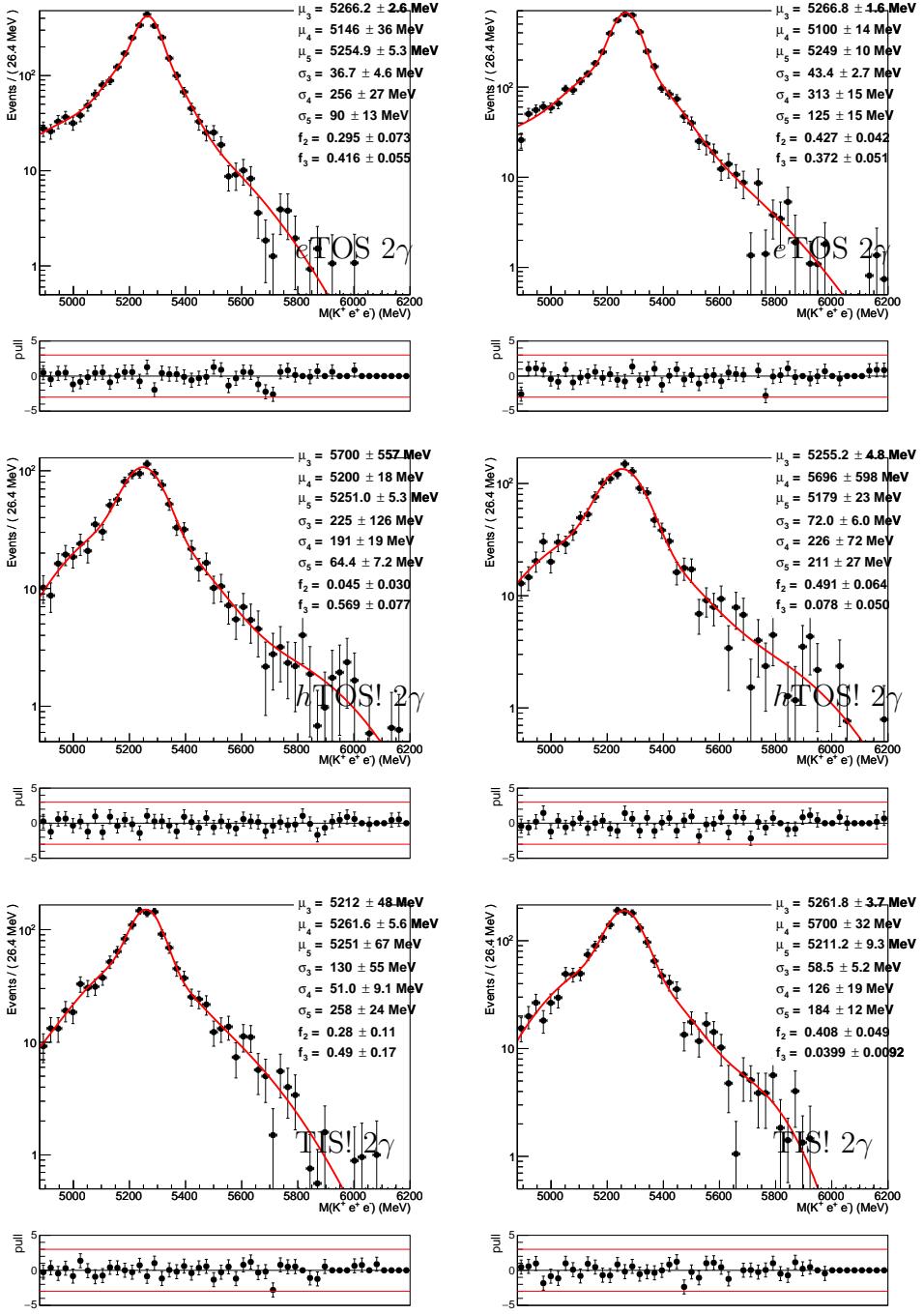


Figure 184: Fits to $m(K^+e^+e^-)$ distribution of $B^+ \rightarrow K^+e^+e^-$ simulated events in the bremsstrahlung category 2γ for run 1 (left) and run 2 (right), for $eTOS$ (top), $hTOS!$ (middle) and $TIS!$ (bottom) events.

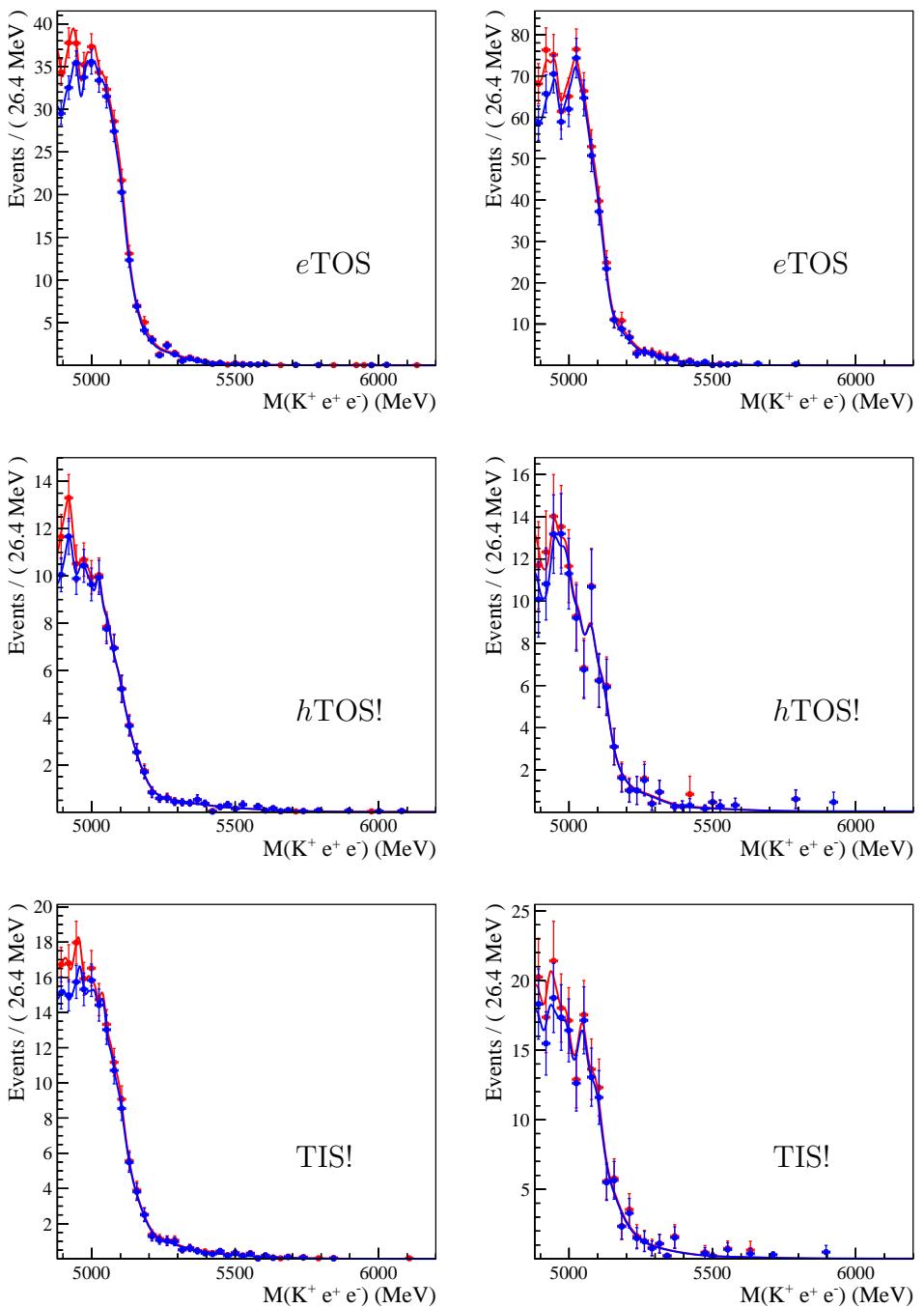


Figure 185: Distribution in $m(K^+ e^+ e^-)$ of the partially reconstructed background with (red) or without (blue) contributions from $B \rightarrow K^{**} e^+ e^-$, for run 1 (left) and run 2 (right) simulation. Trigger categories *eTOS* (top), *hTOS!* (middle) and *TIS!* (bottom), are shown separately.

3018 I Fit validation

3019 I.1 Fit stability and BDT cut

3020 The stability of the fit used to extract the $B^+ \rightarrow K^+ e^+ e^-$ yield in the different trigger
 3021 categories is found to be sensitive to the amount of background expected in the sample.
 3022 As an example, Fig. 186 shows the pull distribution obtained from a toy study of the fit
 3023 to the *hTOS!* sample, where long tails in the distributions for the background yields are
 3024 observed. These tails are also correlated, as can be seen in Fig. 187, which indicates that
 3025 the fit is not able to distinguish between these two components very well. It can also be
 3026 seen that this behaviour has an impact on the signal yield, where a small bias is observed.

3027 A way to improve the discrimination between partially reconstructed background and
 3028 combinatorial is to relax the BDT requirement, i.e. allow more combinatorial background
 3029 to contaminate the sample, so the combinatorial distribution can be better determined.
 3030 By scanning over the BDT requirement and performing a toy study with the expected
 3031 signal and background yields at each working point, the evolution of the signal bias as a
 3032 function of the BDT cut was studied. Figure 188 shows the mean and sigma of the pull
 3033 distribution for the signal yield as a function of the BDT cut. Also shown is the relative
 3034 precision in the determination of the yield, which is the main contribution to the precision
 3035 on R_K . It can be seen, that for looser BDT cuts in the *hTOS!* and *TIS!* categories in
 3036 Run 1, the bias in the signal yield disappears. The *eTOS* categories and all the categories
 3037 in the Run 2 samples show no bias at the optimal working point.

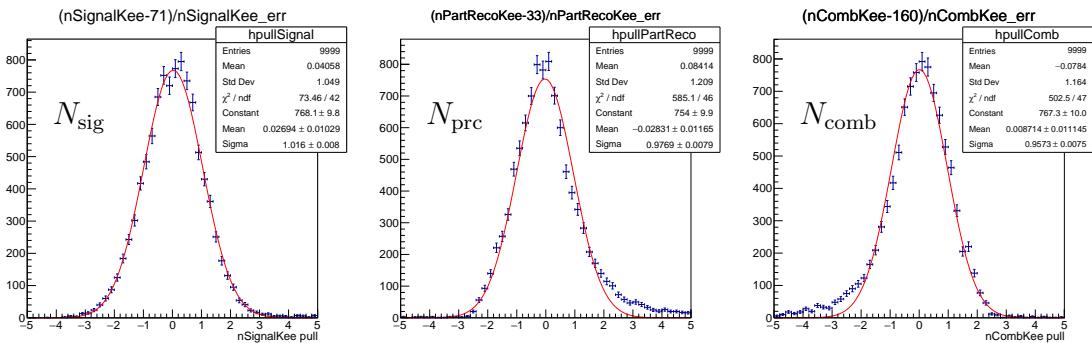


Figure 186: Pull distributions for the fitted parameters in $B^+ \rightarrow K^+ e^+ e^-$ fit for Run 1 samples in *hTOS!* trigger categories. This toy study assumes the yields corresponding to $R_K = 1$.

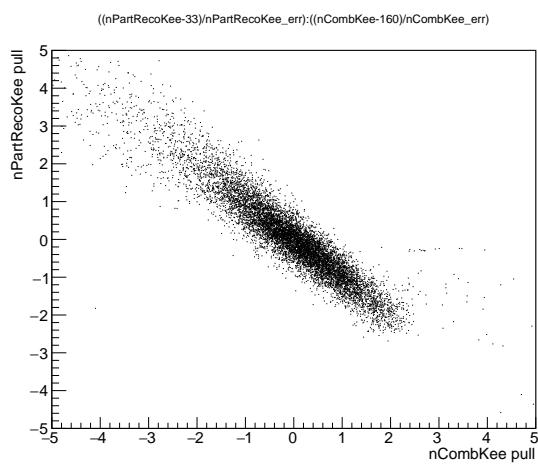


Figure 187: Scatter plot of the pull values for the fitted number of partially reconstructed and combinatorial events obtained from a toy study that assumes the yields expected from $R_K = 1$ in the *hTOS!* trigger category of the Run 1 sample.

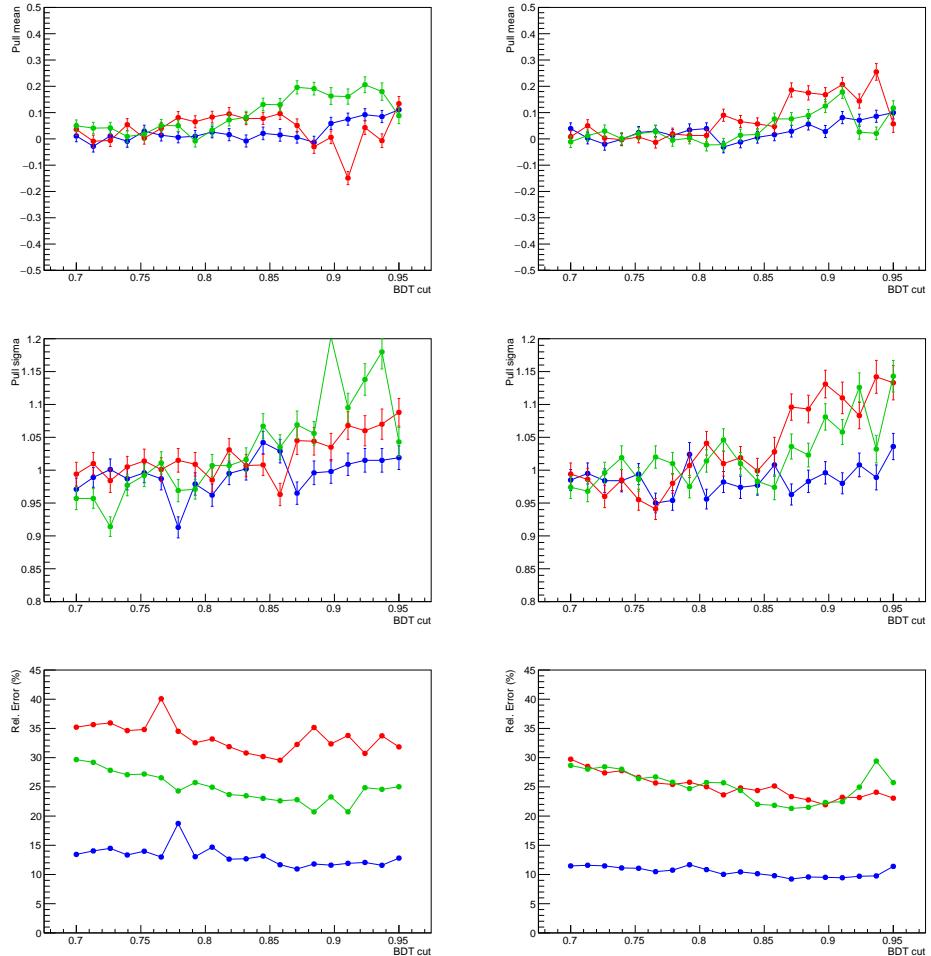


Figure 188: (Top) Mean and (middle) width of the pull distribution for the signal yield as a function of the BDT cut for (blue) $e\text{TOS}$, (red) $h\text{TOS}!$ and (green) $\text{TIS}!$ categories. (Bottom) Relative error on the determination of the signal yield as a function of the BDT cut. The optimised BDT cuts are 0.89, 0.79 and 0.87 (0.86, 0.86 and 0.85) for $e\text{TOS}$, $h\text{TOS}!$ and $\text{TIS}!$ respectively in the Run 1 (Run 2) samples.

3038 **I.2 Fit validation for different R_K hypotheses**

3039 The residuals and pull distribution for the parameters in the fit are shown in the following,
3040 for the different toy studies performed:

- 3041 • Run 1 only with $R_K = 1$, signal (background) parameters residuals Fig. 189 (Fig. 191)
3042 and pulls Fig. 190 (Fig. 192).
- 3043 • Run 2 only with $R_K = 1$, signal (background) parameters residuals Fig. 193 (Fig. 195)
3044 and pulls Fig. 194 (Fig. 196).
- 3045 • Run 1 and Run 2 combination with $R_K = 1$, signal (background) parameters
3046 residuals Fig. 197 and Fig. 198 (Fig. 201 and Fig. 202) and pulls Fig. 199 and
3047 Fig. 200 (Fig. 203 and Fig. 204).
- 3048 • Run 1 and Run 2 combination with $R_K = 0.745$, R_K , $N_{K^+\mu^+\mu^-}^{R1}$ and $N_{K^+\mu^+\mu^-}^{R2}$
3049 residuals and pulls Fig. 205.
- 3050 • Run 1 and Run 2 combination with $R_K = 1.25$, R_K , $N_{K^+\mu^+\mu^-}^{R1}$ and $N_{K^+\mu^+\mu^-}^{R2}$ residuals
3051 and pulls Fig. 205.

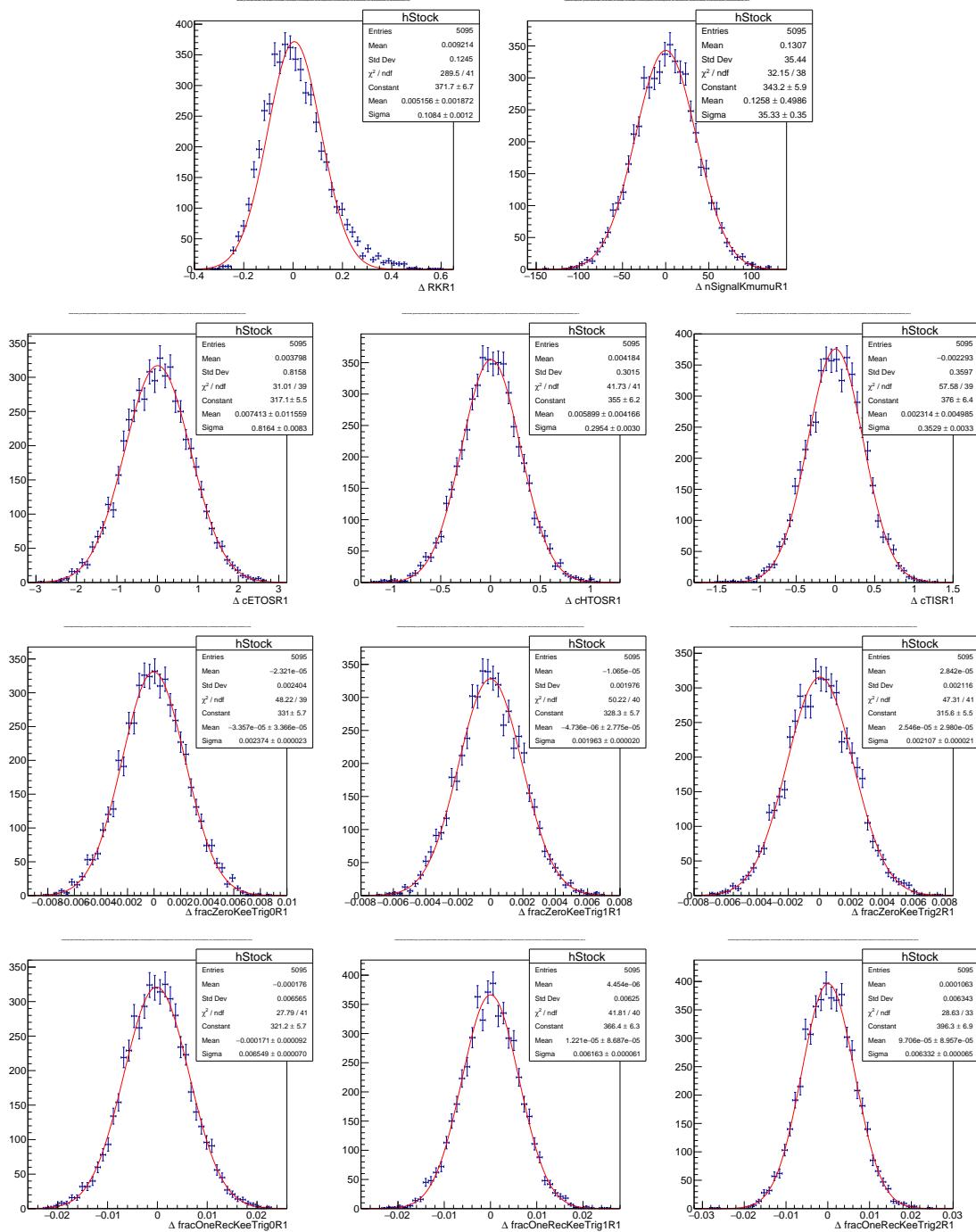


Figure 189: Residual distributions for the signal fit parameters. This toy study assumes the Run 1 yields corresponding to $R_K = 1$.

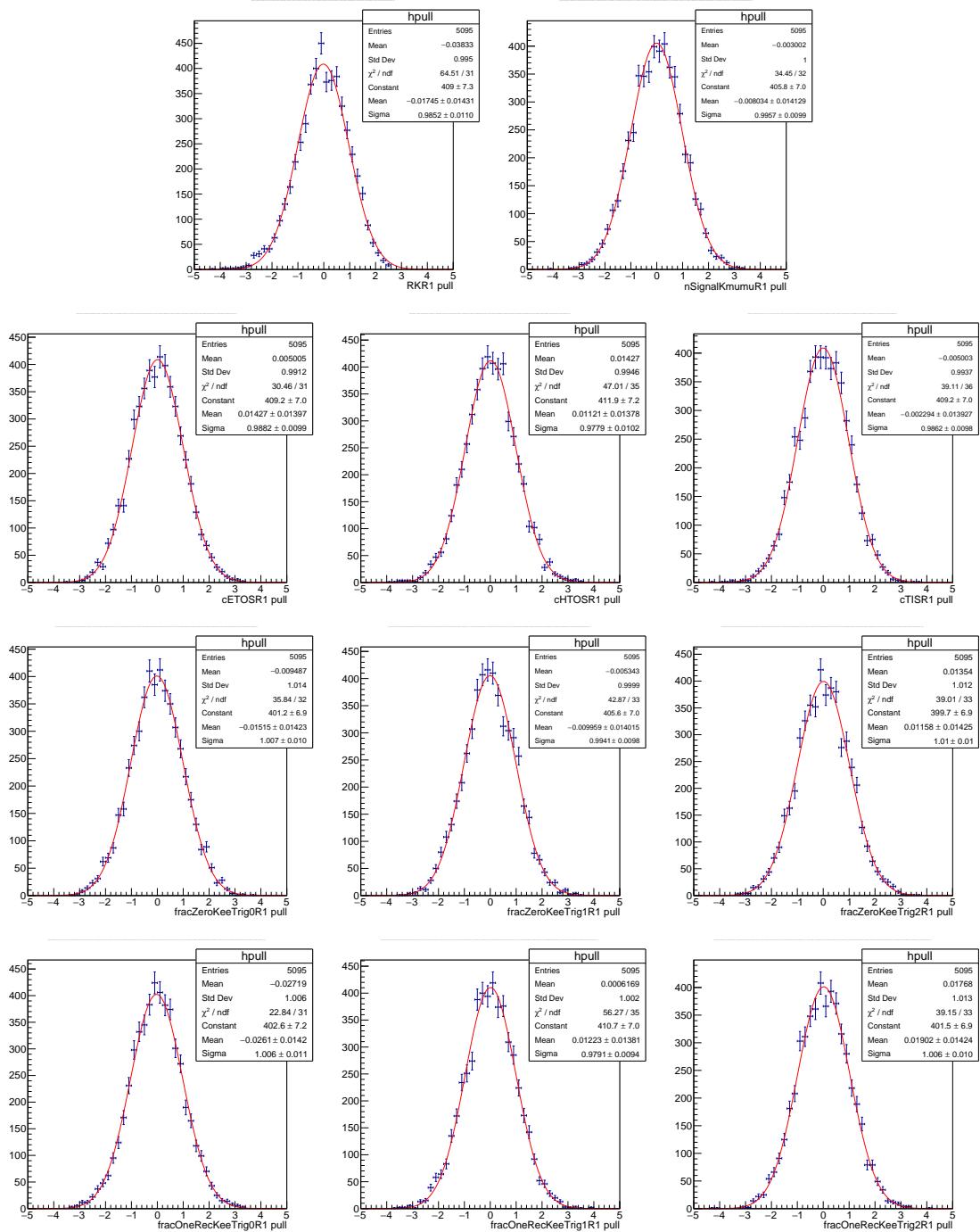


Figure 190: Pull distributions for the signal fit parameters. This toy study assumes the Run 1 yields corresponding to $R_K = 1$.

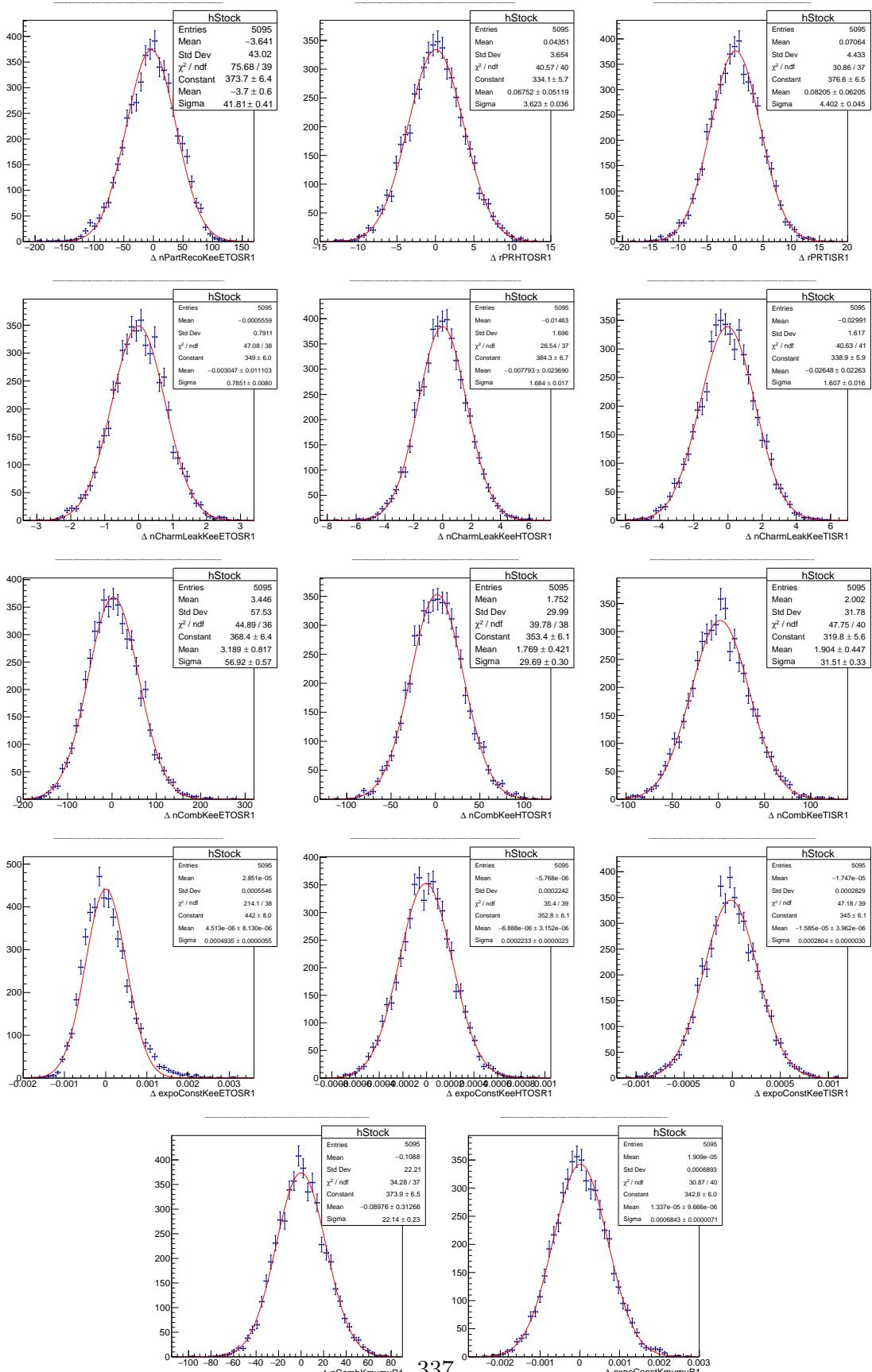


Figure 191: Residual distributions for the background fit parameters. This toy study assumes the Run 1 yields corresponding to $R_K = 1$.

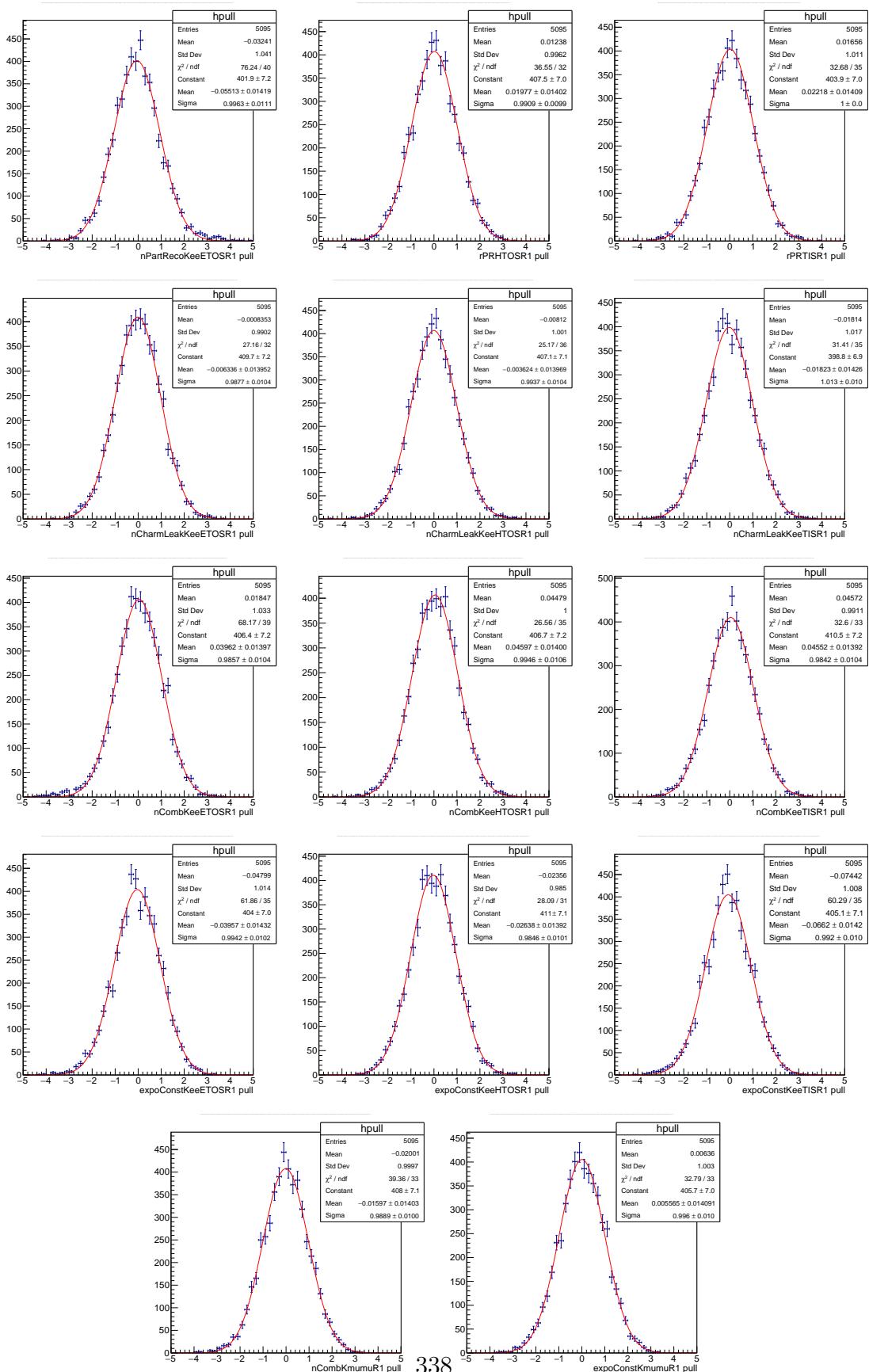


Figure 192: Pull distributions for the background fit parameters. This toy study assumes the Run 1 yields corresponding to $R_K = 1$.

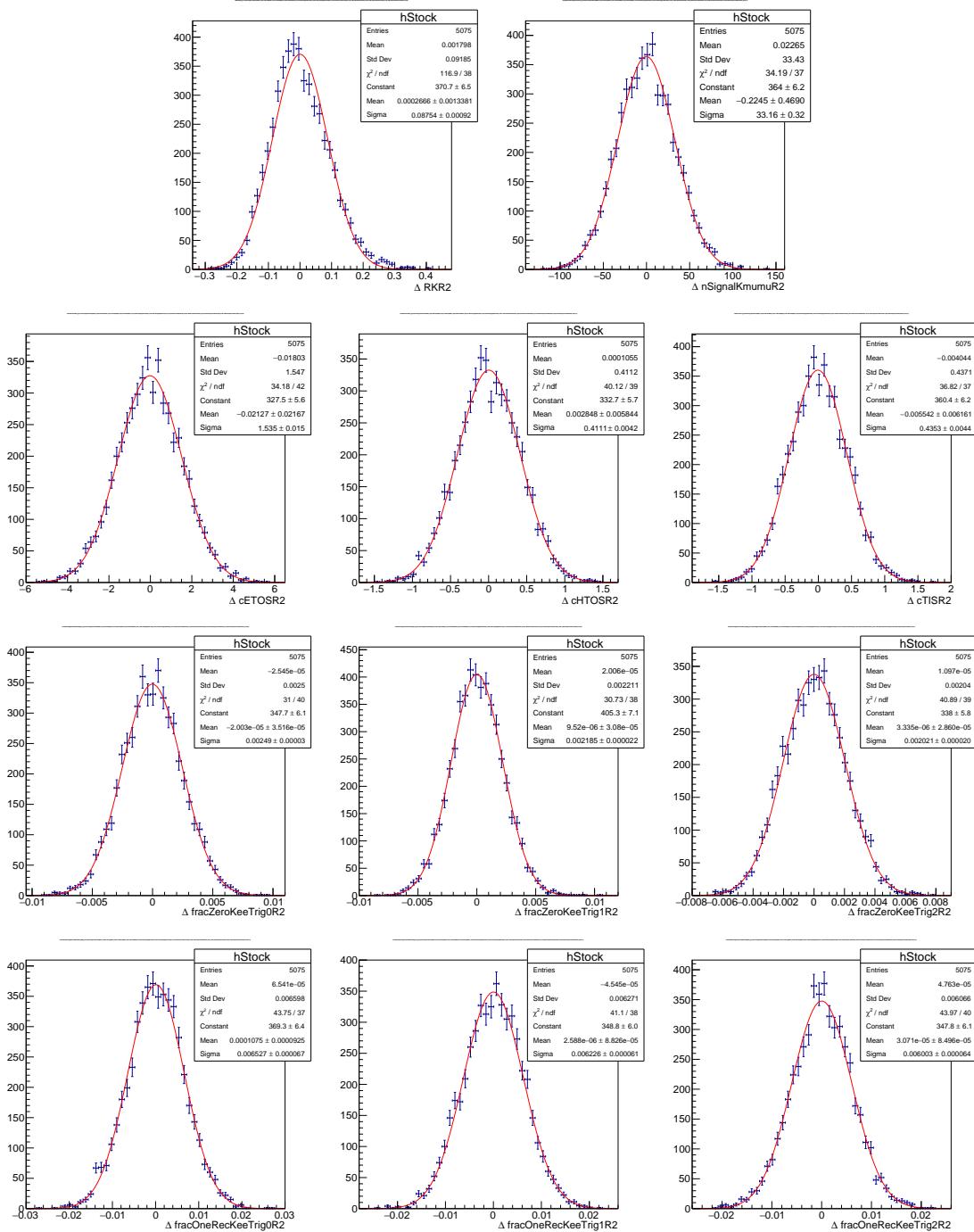


Figure 193: Residual distributions for the signal fit parameters. This toy study assumes the Run 2 yields corresponding to $R_K = 1$.

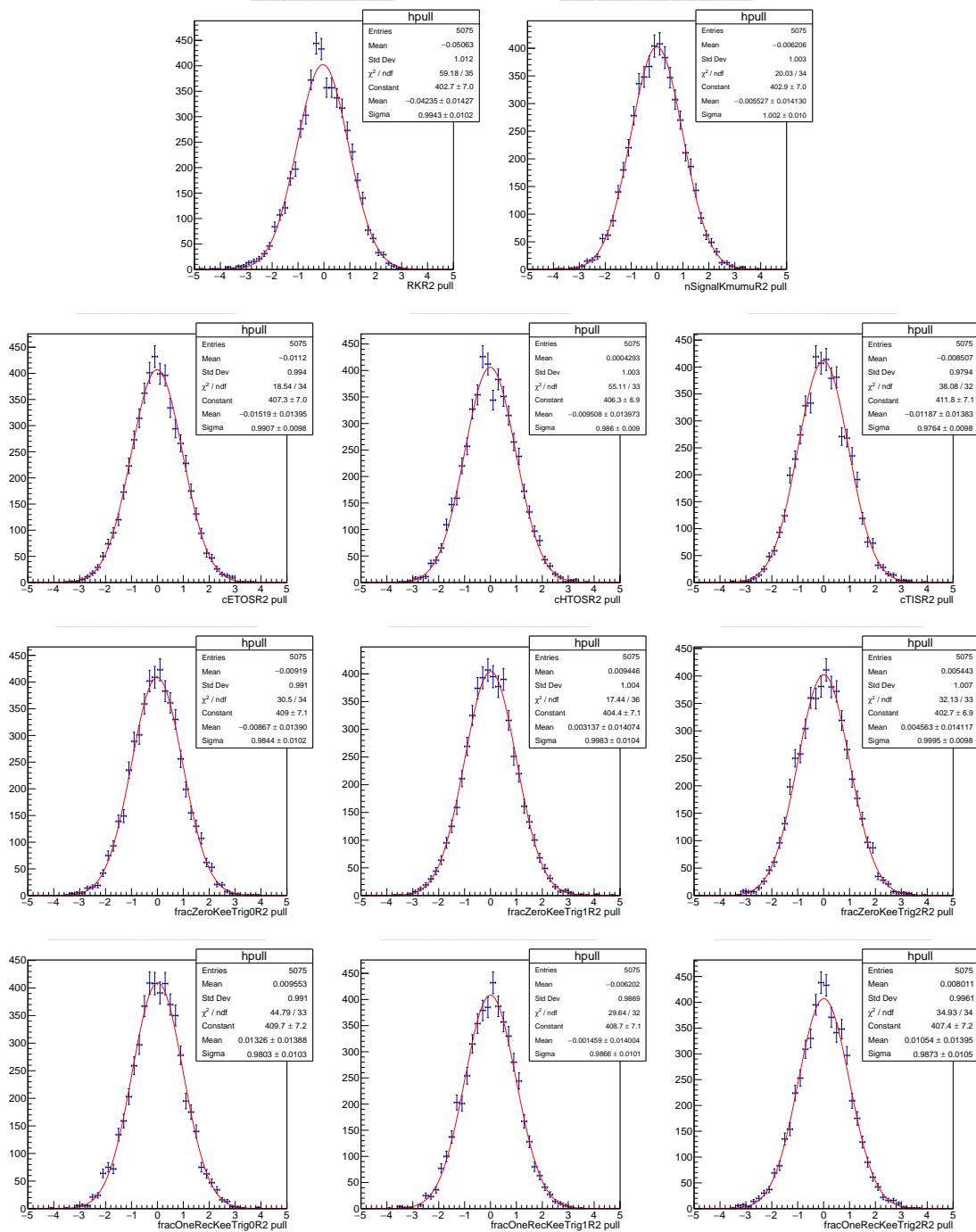


Figure 194: Pull distributions for the signal fit parameters. This toy study assumes the Run 2 yields corresponding to $R_K = 1$.

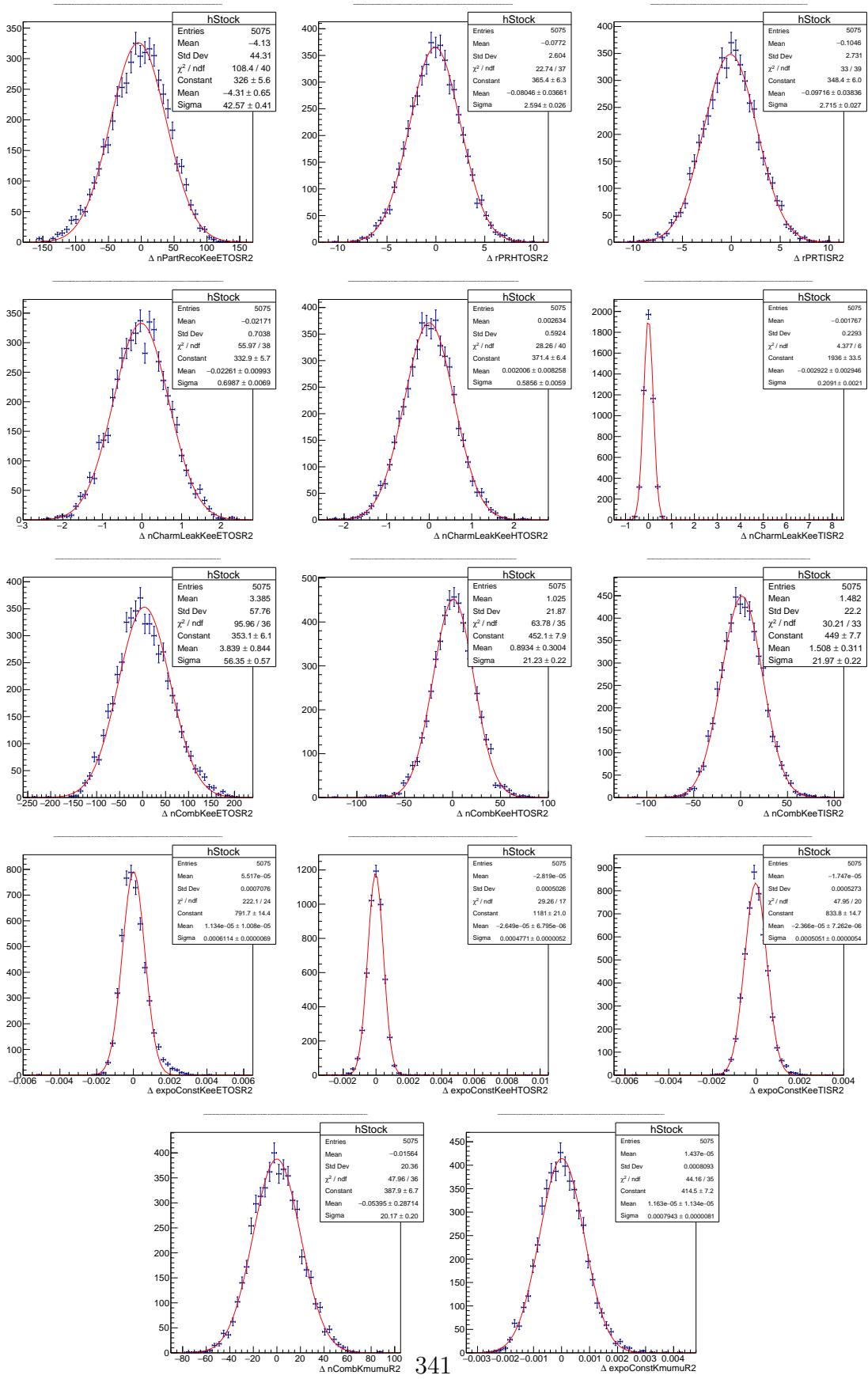


Figure 195: Residual distributions for the background fit parameters. This toy study assumes the Run 2 yields corresponding to $R_K = 1$.

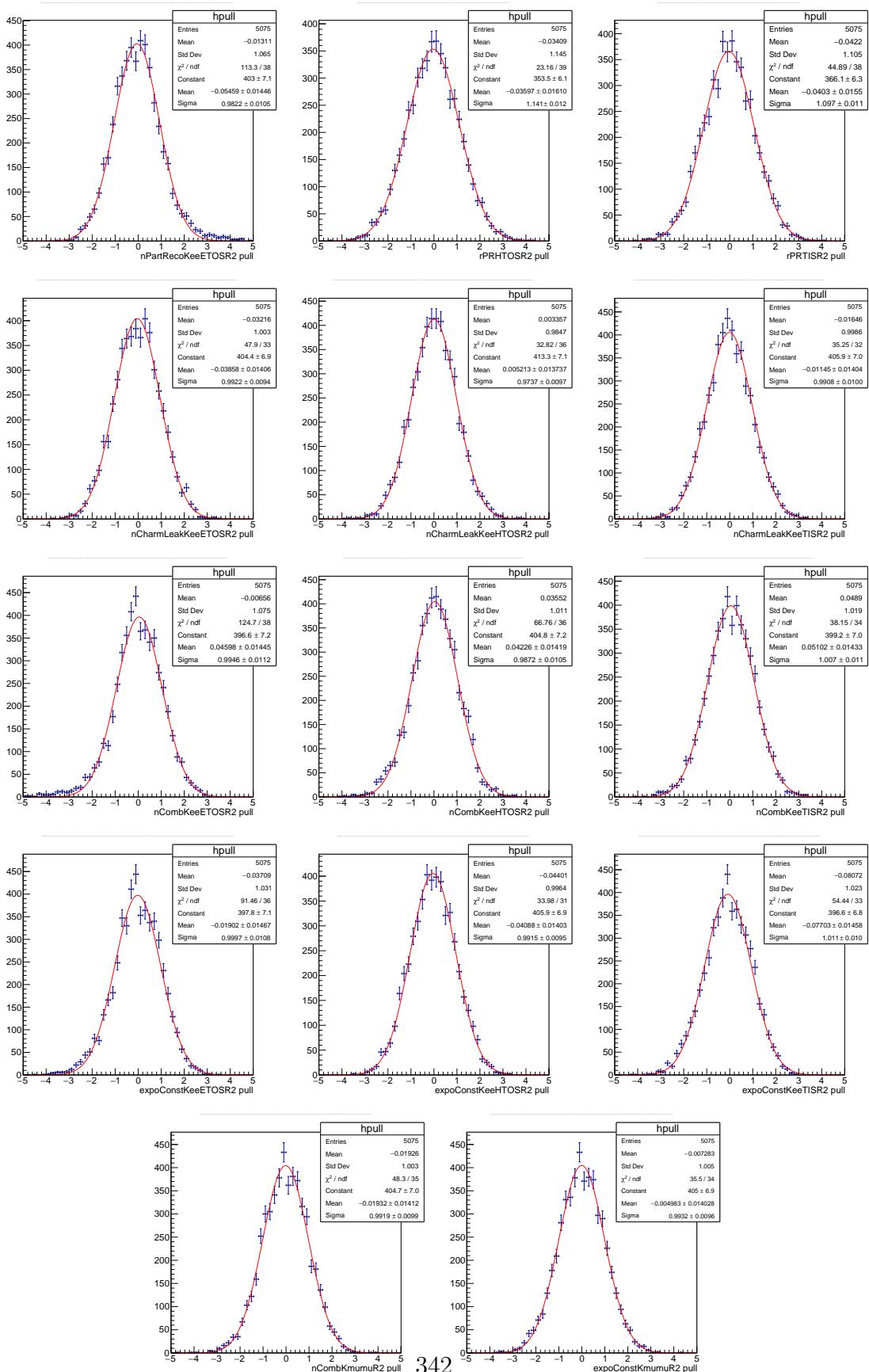


Figure 196: Pull distributions for the background fit parameters. This toy study assumes the Run 2 yields corresponding to $R_K = 1$.

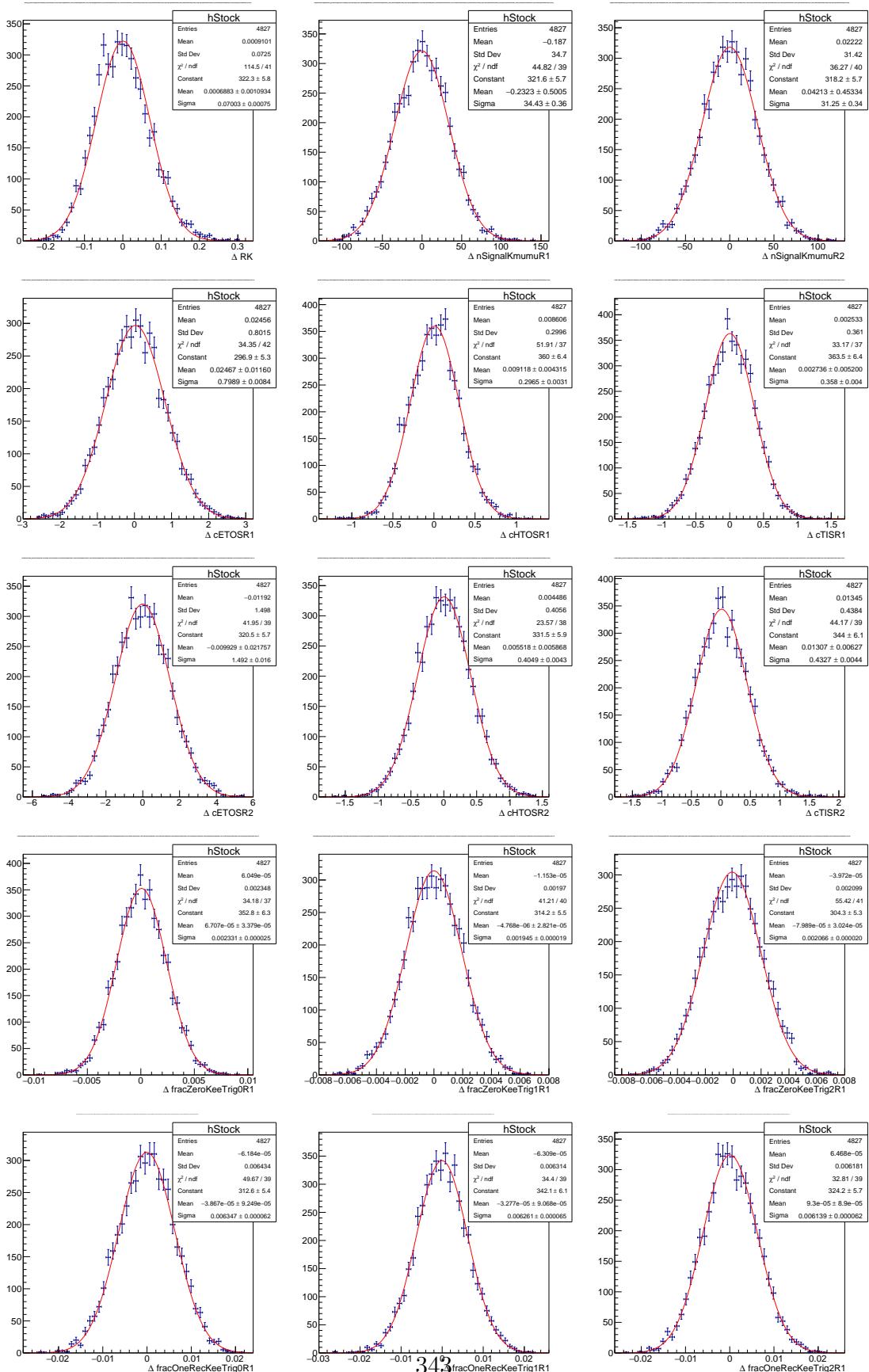


Figure 197: Residual distributions for the signal fit parameters. This toy study assumes the Run 1 and Run 2 yields corresponding to $R_K = 1$.

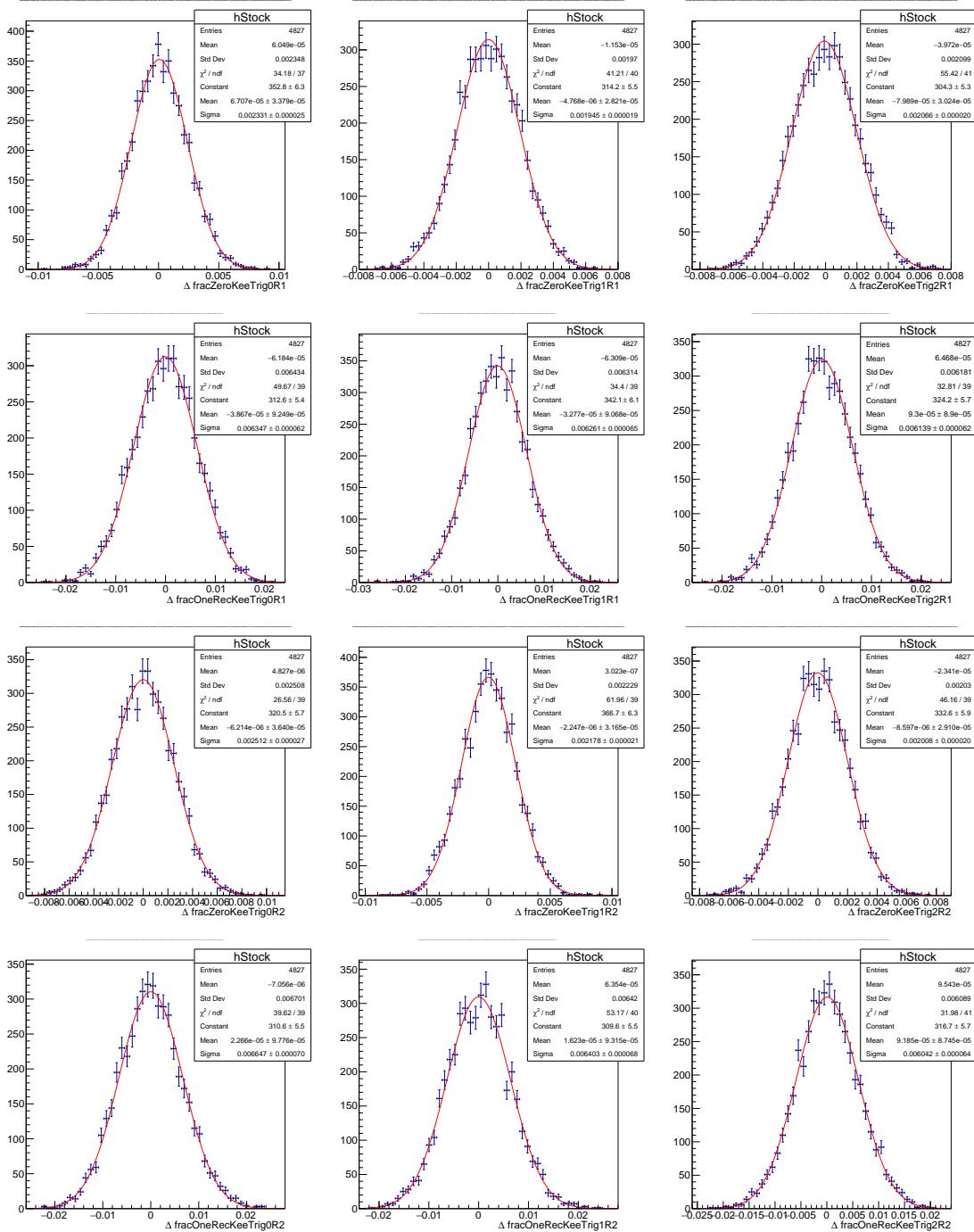


Figure 198: Residual distributions for the signal fit parameters. This toy study assumes the Run 1 and Run 2 yields corresponding to $R_K = 1$. (continuation)

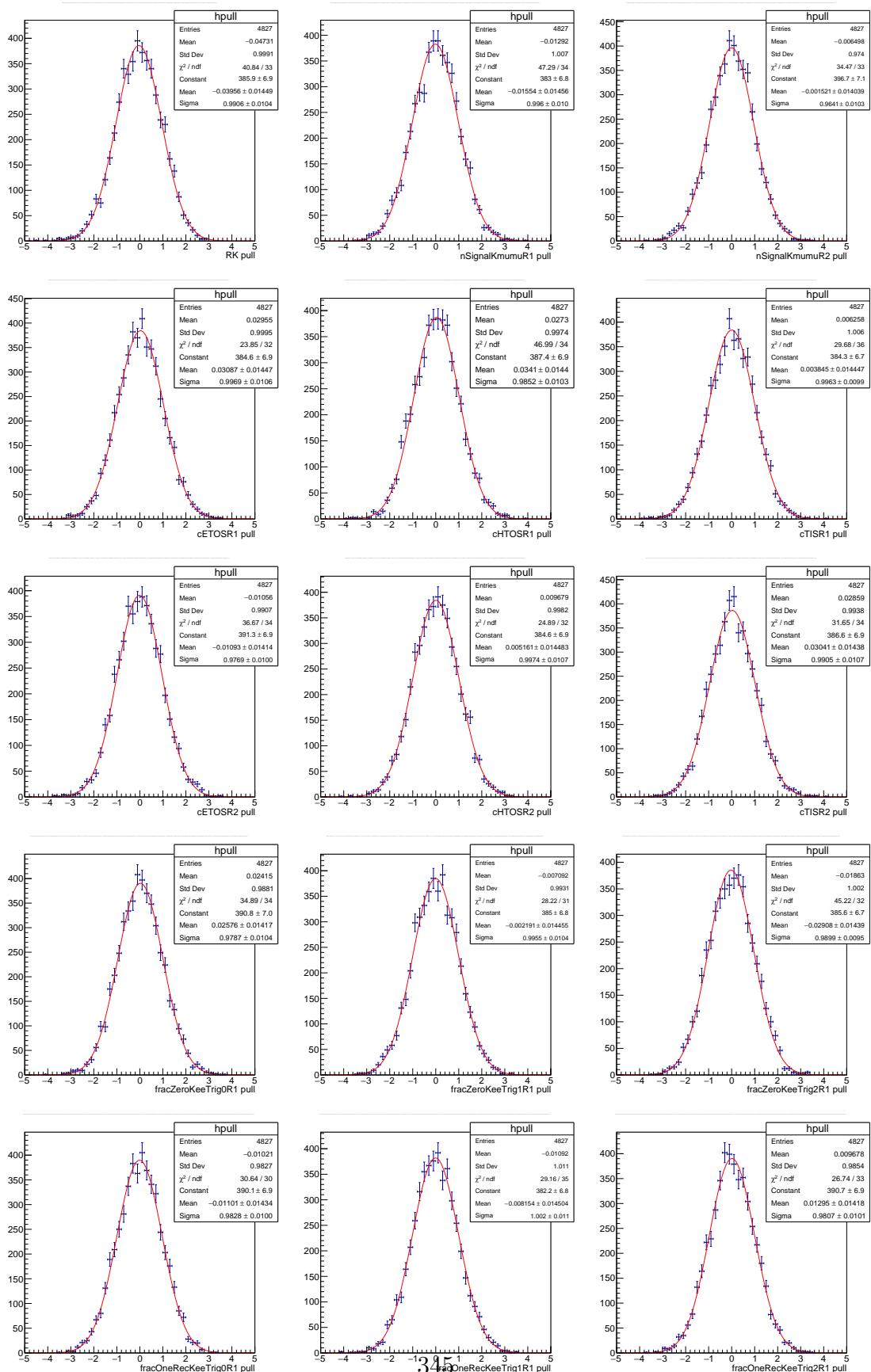


Figure 199: Pull distributions for the signal fit parameters. This toy study assumes the Run 1 and Run 2 yields corresponding to $R_K = 1$.

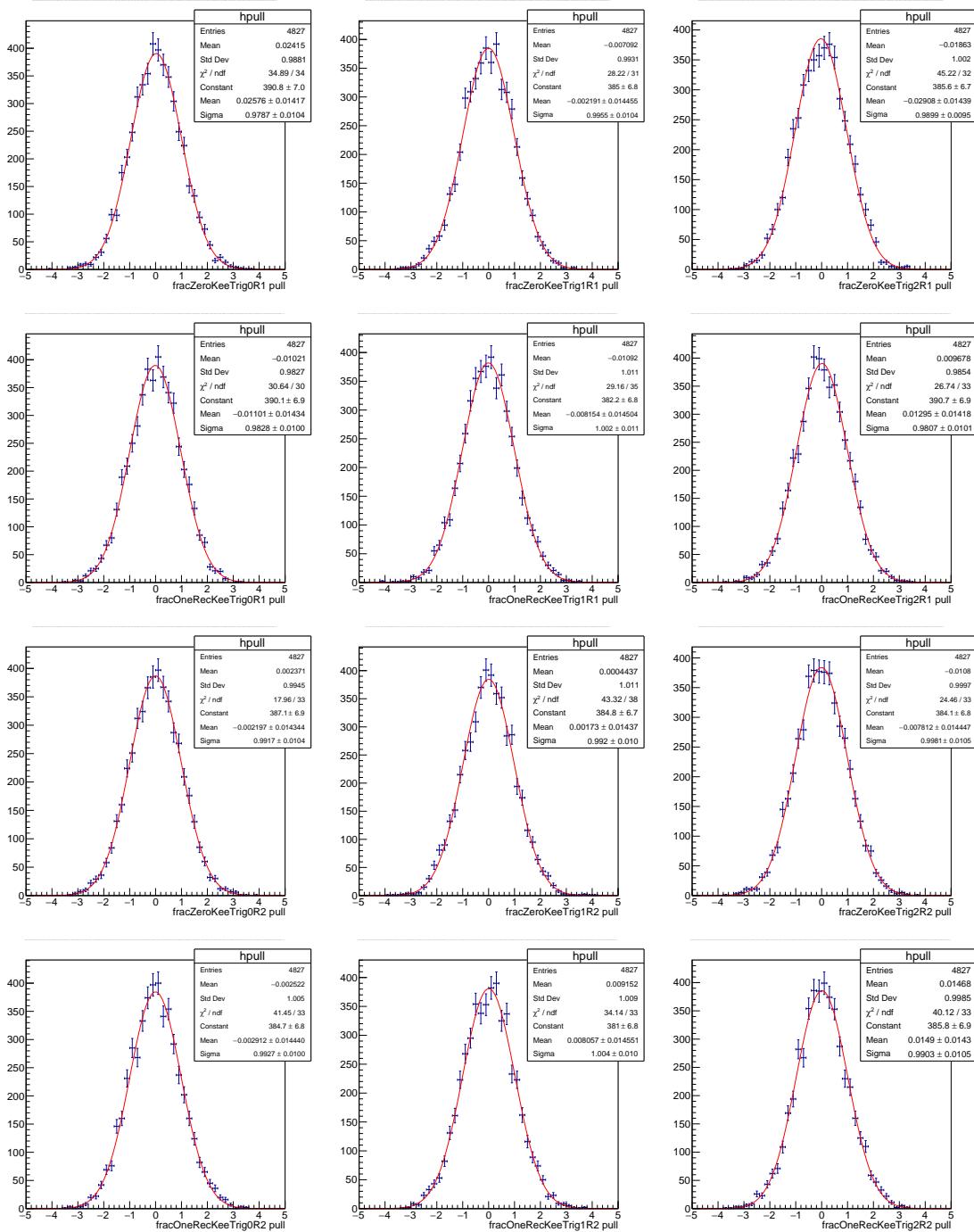


Figure 200: Pull distributions for the signal fit parameters. This toy study assumes the Run 1 and Run 2 yields corresponding to $R_K = 1$. (continuation)

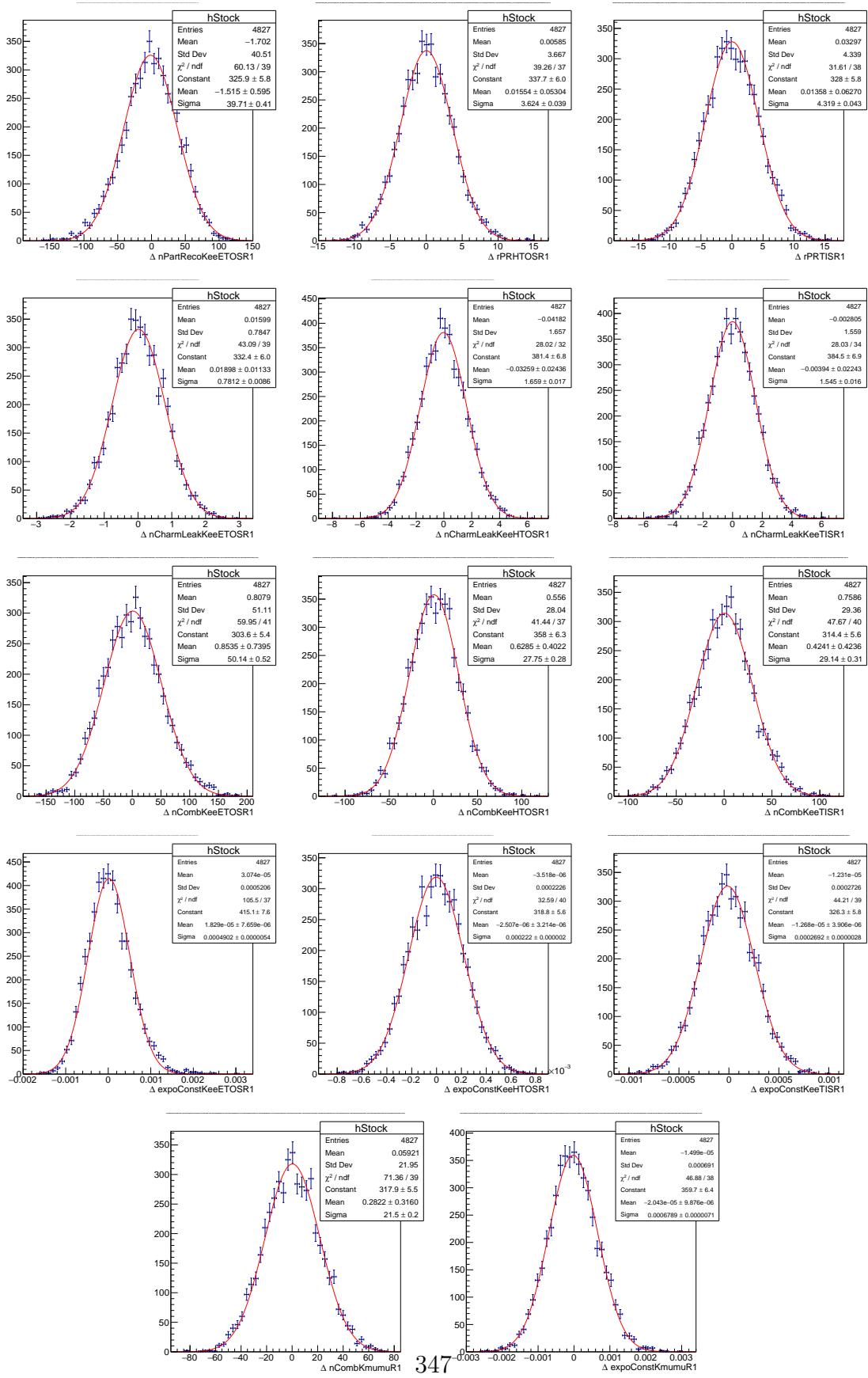


Figure 201: Residual distributions for the background fit parameters. This toy study assumes the Run 1 and Run 2 yields corresponding to $R_K = 1$.

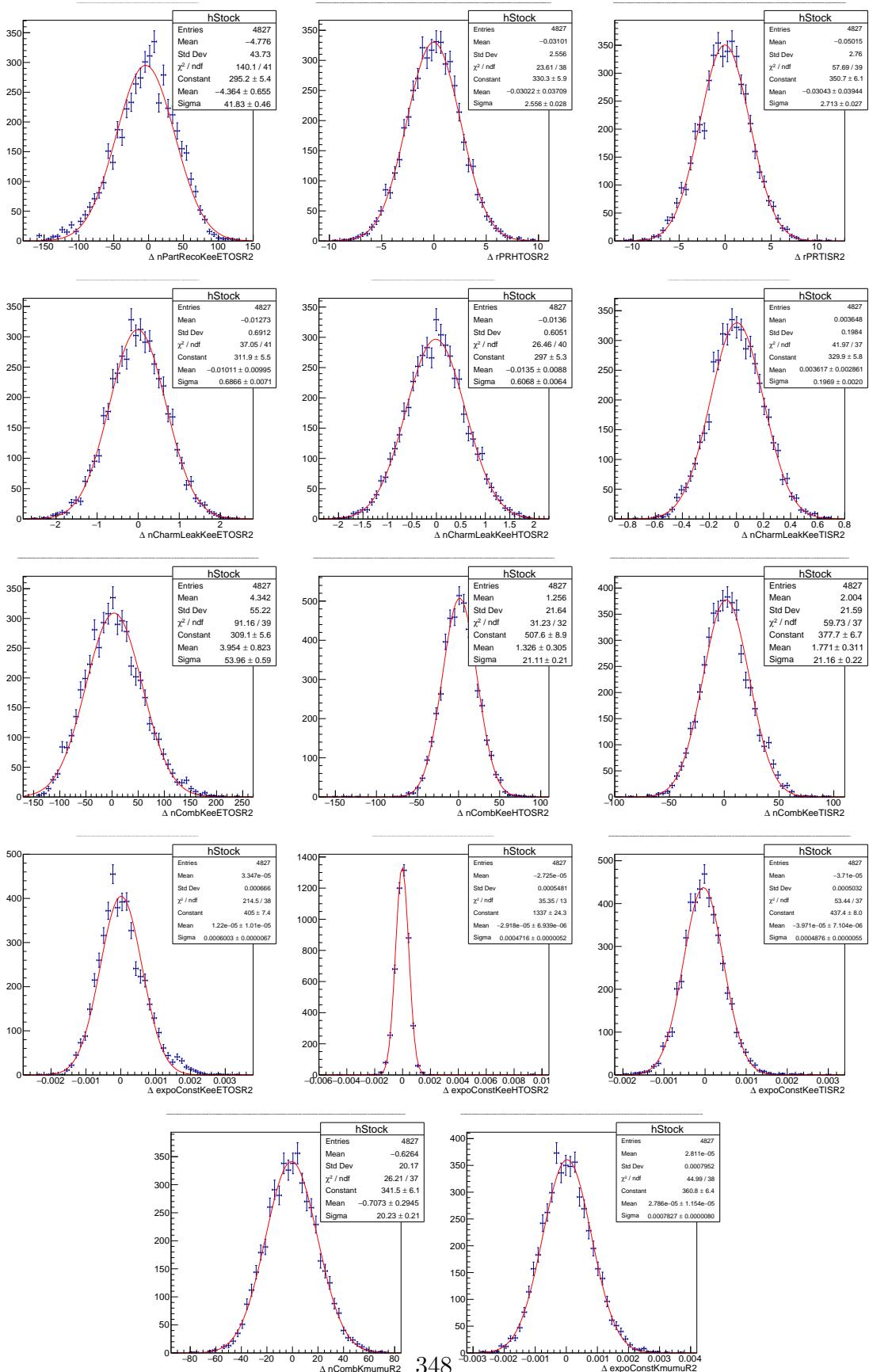


Figure 202: Residual distributions for the background fit parameters. This toy study assumes the Run 1 and Run 2 yields corresponding to $R_K = 1$. (continuation)

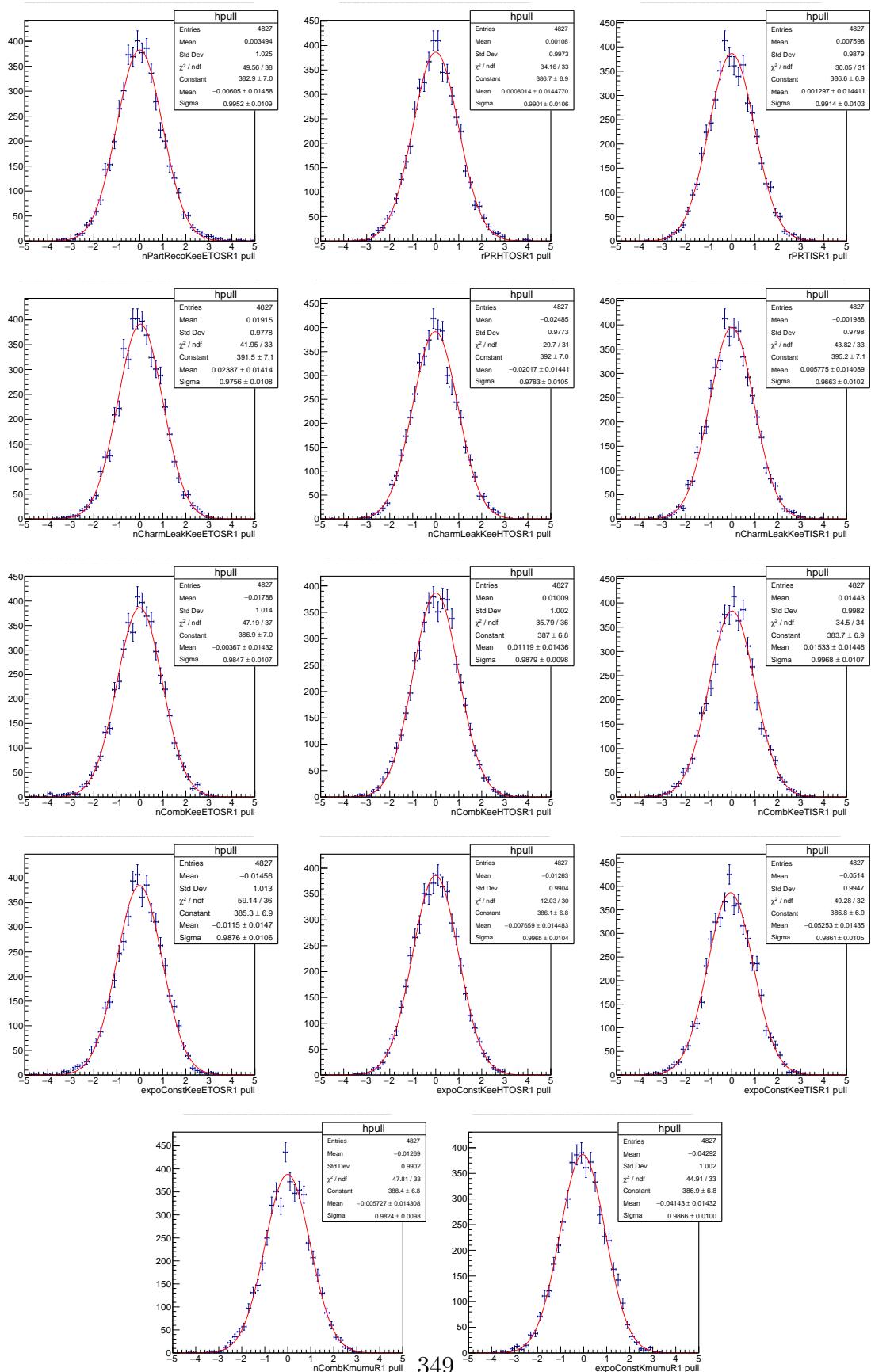


Figure 203: Pull distributions for the background fit parameters. This toy study assumes the Run 1 and Run 2 yields corresponding to $R_K = 1$.

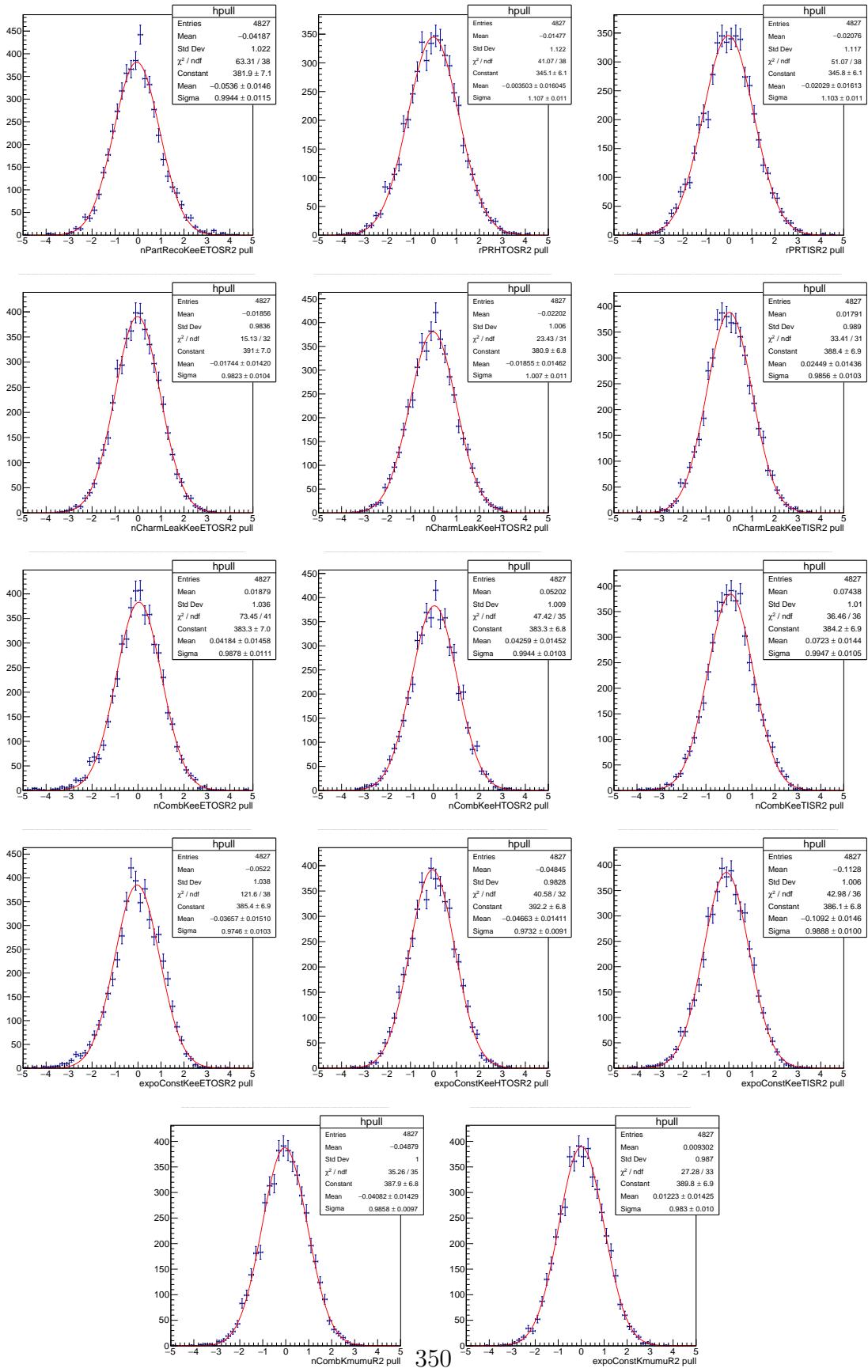


Figure 204: Pull distributions for the background fit parameters. This toy study assumes the Run 1 and Run 2 yields corresponding to $R_K = 1$. (continuation)

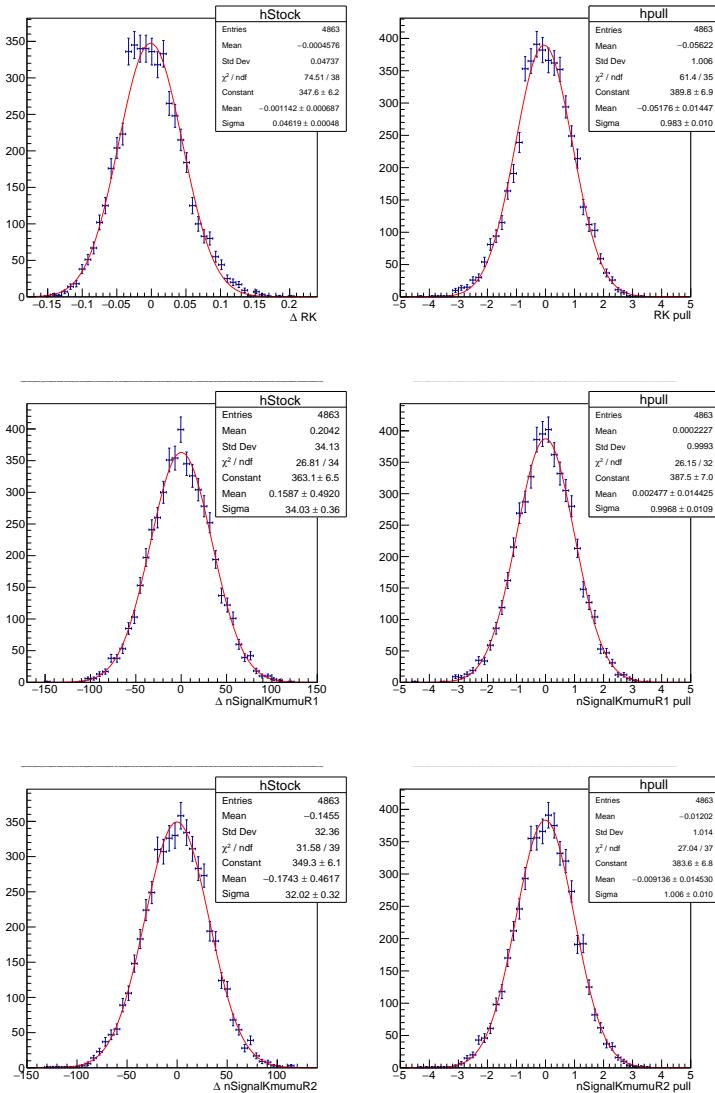


Figure 205: Distribution of the (left) residuals and (right) pulls for the fitted parameters (top) R_K , (middle) $N_{K+\mu^+\mu^-}^1$ and (bottom) $N_{K+\mu^+\mu^-}^2$. This toy study assumes Run 1 and Run 2 yields corresponding to $R_K = 0.745$.

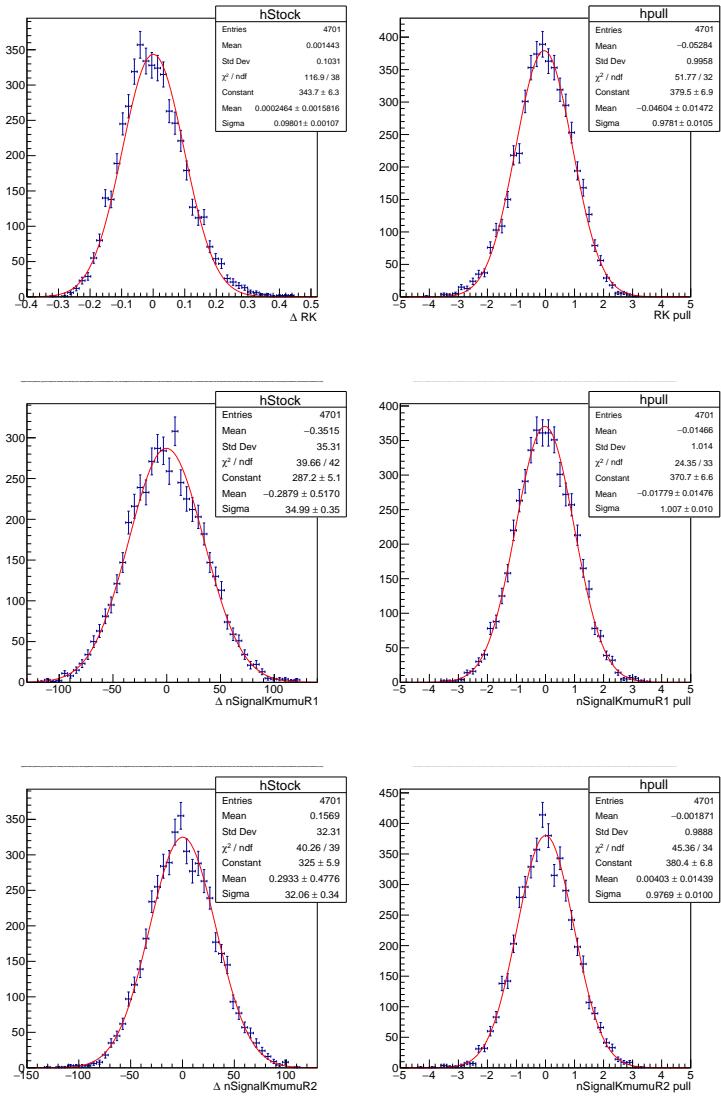


Figure 206: Distribution of the (left) residuals and (right) pulls for the fitted parameters (top) R_K , (middle) $N_{K^+\mu^+\mu^-}^1$ and (bottom) $N_{K^+\mu^+\mu^-}^2$. This toy study assumes Run 1 and Run 2 yields corresponding to $R_K = 1.25$.

3052 J Fits to data

Mass fits to $B^+ \rightarrow K^+\mu^+\mu^-$ candidates from the Run 1 data are shown in Fig. 207.

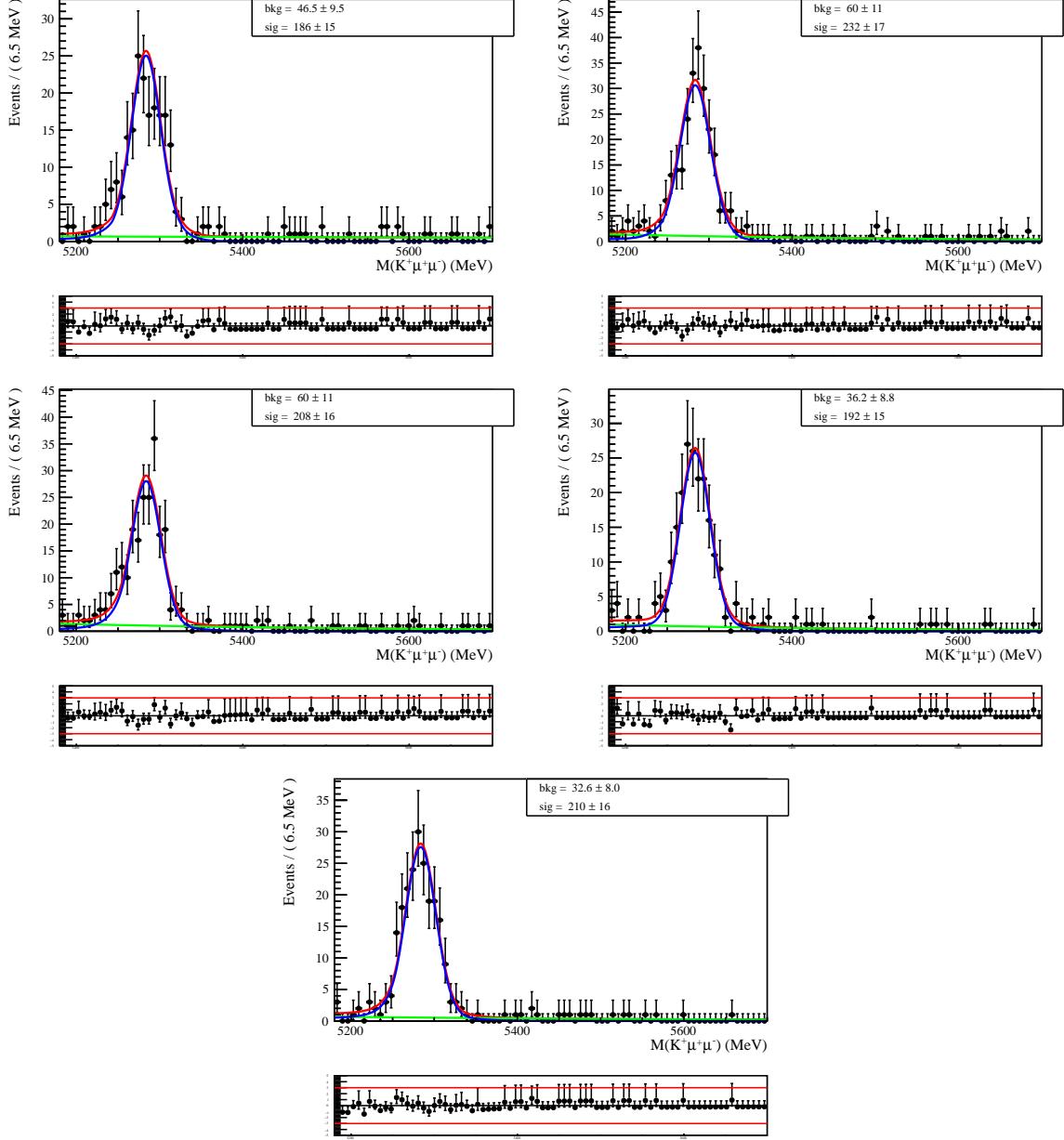


Figure 207: Fit to the $m(K^+\mu^+\mu^-)$ distribution of $B^+ \rightarrow K^+\mu^+\mu^-$ selected candidates in the Run 1 data sample in the q^2 bins $1.1 < q^2 < 2.0 \text{ GeV}/c$, $2.0 < q^2 < 3.0 \text{ GeV}/c$, $3.0 < q^2 < 4.0 \text{ GeV}/c$, $4.0 < q^2 < 5.0 \text{ GeV}/c$ and $5.0 < q^2 < 6.0 \text{ GeV}/c$. The red solid line shows the fit model, the blue line is the signal component and the green line shows the combinatorial background.

3054 K Additional $r_{J/\psi}$ differential plots

3055 The dependence of the $r_{J/\psi}$ with different variables with respect to those considered in
 3056 Section 11.2 is discussed in this section.

3057 The dependence with the χ^2_{IP} and IP of the electrons is shown in Fig. 208. A small
 3058 trend can be seen in these plots using the nominal weights computed using the muon mode,
 3059 which is expected due to the known discrepancy between data and simulation regarding
 3060 the χ^2_{IP} variable. The same pattern is observed when $r_{J/\psi}$ is computed in a 2D grid as
 3061 a function of $\max(\chi^2_{\text{IP}}(e^+, e^-))$ and $\min(\chi^2_{\text{IP}}(e^+, e^-))$, see Fig. 209 (binning definition is
 3062 shown in Fig. 218). If the data/MC discrepancy is interpreted as a genuine mis-modelling
 3063 of the efficiencies, the shift on R_K remains at the sub-percent level. Moreover, the 1D
 3064 $r_{J/\psi}$ plots extracted with kinematic weights computed using the $K^+ J/\psi(e^+ e^-)$ mode are
 3065 shown, and they do not show the same pattern as the $r_{J/\psi}$ plots extracted using the
 3066 nominal weights. This shows that the the existing systematics on the reweighting scheme
 3067 covers well for the χ^2_{IP} data/simulation mismatch.

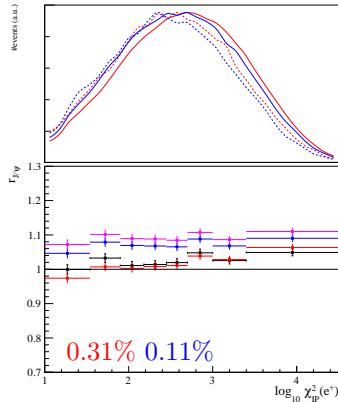


Figure 208: Plots of $r_{J/\psi}$ as a function of for the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode in the eTOS category and Run 2 data taking conditions. The L0 and PID corrections are applied to both modes. The black points are computed with no kinematic weights. The red points are computed applying the kinematic weights to both the electron and muon modes. The blue ones are computed applying the kinematic weights computed from the $K^+ J/\psi(e^+ e^-)$ eTOS mode, and the magenta ones are from the $K^+ J/\psi(e^+ e^-)$ TIS mode. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. At the top of each plot, the kinematic distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (solid blue) are shown.

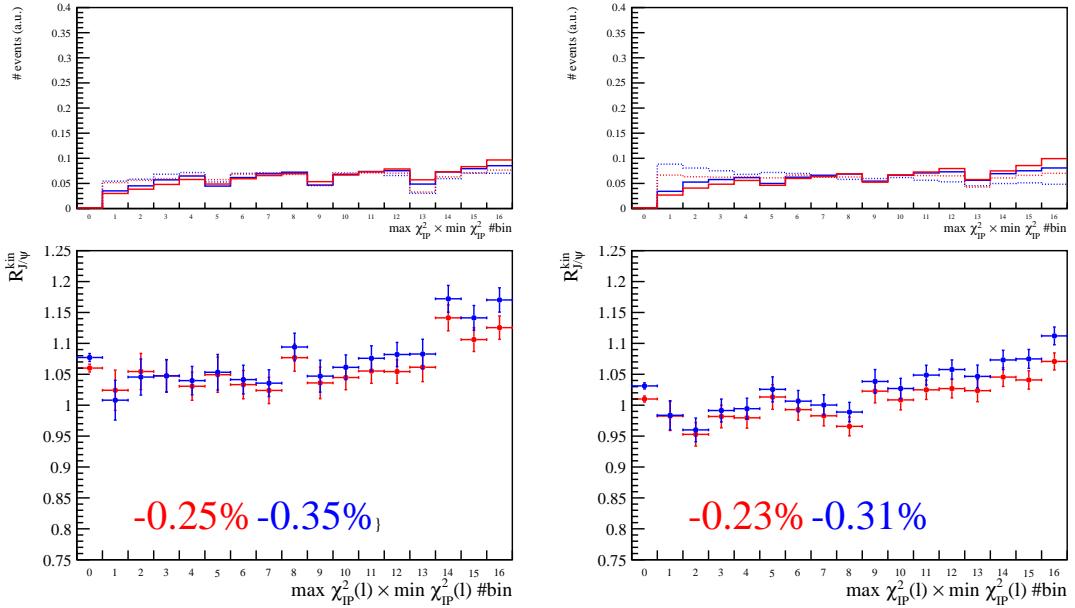


Figure 209: Result of the $r_{J/\psi}^{\text{bin}}$ test in two dimensional bins, where the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode is in the $e\text{TOS}$ category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode is in the μTOS category, both in the (left) Run 1 or (right) Run 2 data set. The histograms in red are obtained by computing the efficiencies with the nominal trigger, PID, and kinematic corrections. For the blue ones, the η corrections are added. The flatness parameters d_f associated to the red and blue histogram are displayed at the bottom of each plot in the same color. At the top of each plot, the bin distributions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted blue) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid blue) are shown.

3068 L $r_{J/\psi}^{\text{bin}}$ test: binning definition

3069 The binning scheme chosen for the $r_{J/\psi}^{\text{bin}}$ test in Run 1 is shown in Fig. 210 and Fig. 211
 3070 overlaid on the 2D distributions for $B^+ \rightarrow K^+ e^+ e^-$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the
 3071 simulation. For comparison, the same binning scheme is shown over the 2D distributions
 3072 for the muon modes, $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$, in Fig. 212 and Fig. 213.
 3073 Equivalent plots for the $r_{J/\psi}^{\text{bin}}$ test in Run 2 are shown in Fig. 214 and Fig. 215 for the
 3074 electron modes, and Fig. 216 and Fig. 217 for the muon modes.

3075 M ECAL illumination

3076 The ECAL illumination is shown in Fig. 219 for $B^+ \rightarrow K^+ e^+ e^-$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$
 3077 simulation and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ sWeighted data. Signal and normalisation modes
 3078 seem to produce a similar distributions on simulation. Some discrepancies between data
 3079 and simulation in the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode are found. These differences might come
 3080 from dead or badly performing cells in the ECAL, or extra material at high η . The impact

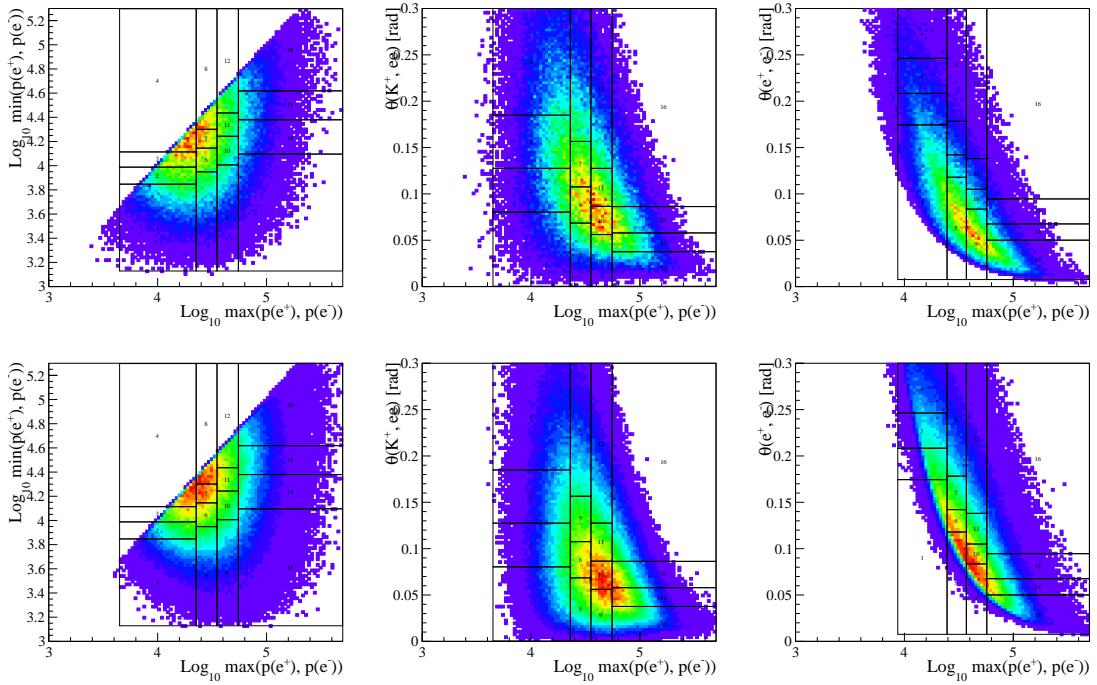


Figure 210: Two dimensional (left) $\max p_\ell$, $\min p_\ell$, (middle) $\max p_\ell$, $\alpha_{\ell+\ell^-}$ and (right) $\max p_\ell$, $\alpha(\ell^+, \ell^-)$ distributions for reconstructed (top) $B^+ \rightarrow K^+ e^+ e^-$ and (bottom) $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events. The bins used in the $r_{J/\psi}^{\text{bin}}$ test in Run 1 are shown on both figures.

of this discrepancy on R_K is evaluated by using the data/MC ratio plot as a correction weight on R_K . The relative change of R_K using these correction weights is small: 0.8%, which is well covered by our PID/trigger/tracking systematics.

N Reweighting checks

In order to check that the reweighting procedure used to correct for discrepancies between data and simulation is correct, the efficiencies for all modes, as well as the ratio of efficiencies entering R_K , are recomputed using variations of the corrections weights, as shown in Tab. 53. The rows labelled as ‘no corrections’ correspond to the efficiencies computed straight from the simulation with no corrections, except the weights that adjust the fractions of MC samples in different data taking conditions to the fractions of luminosity corresponding to these conditions. The rows labelled as ‘all corrections’ correspond to the nominal efficiency computation. In Run 2, the last three rows correspond to efficiencies computed with the inverse of all corrections weights ($1/w$). It can be seen that removing all corrections would change the R_K result by only 1 – 5%, depending on the trigger category and year of data taking. This is covered by the systematics.

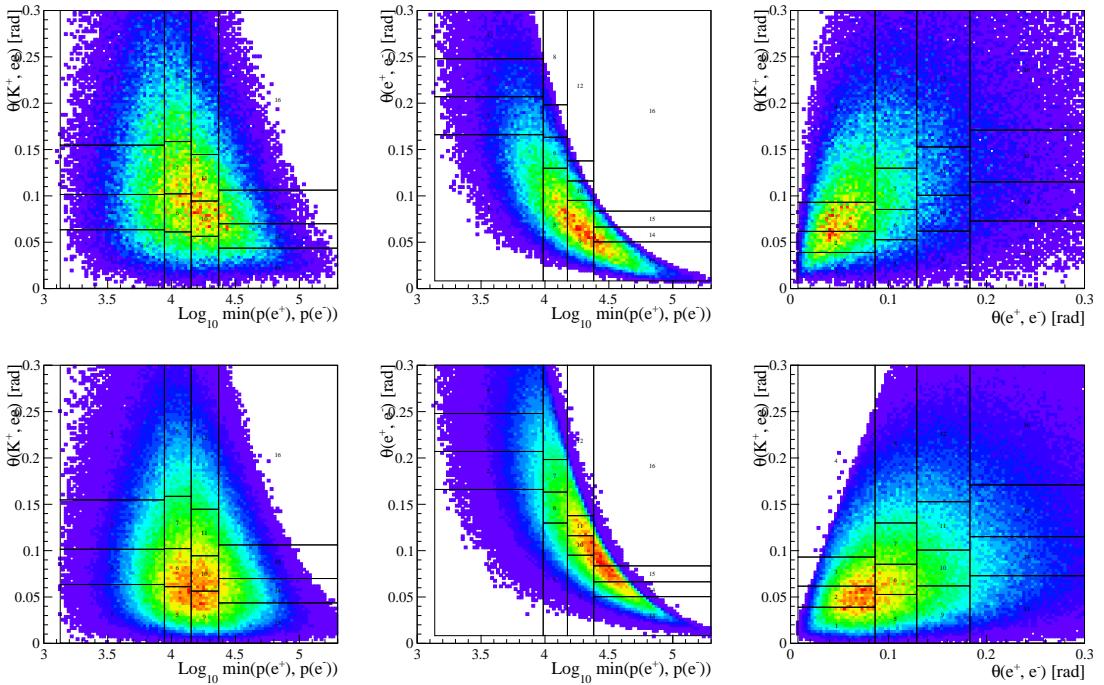


Figure 211: Two dimensional (left) $\min p_\ell$, $\alpha(\ell^+, \ell^-)$, (middle) $\min p_\ell$, $\alpha_{\ell+\ell^-}$ and (right) $\alpha_{\ell+\ell^-}$, $\alpha(\ell^+, \ell^-)$ distributions for reconstructed (top) $B^+ \rightarrow K^+ e^+ e^-$ and (bottom) $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events. The bins used in the $r_{J/\psi}^{\text{bin}}$ test in Run 1 are shown on both figures.

O Staged unblinding results

The full results of the fit to the data under different configurations are presented in this section:

- **RKFitAll:** Six independent values of R_K corresponding to each trigger category (*eTOS*, *hTOS!*, *TIS!*) and independently for the Run 1 and Run 2 samples. The results of this fit to the data are shown in Fig. 222 and the obtained parameters are listed in Tab. 55.
- **RKFitTrigger:** Three independent values of R_K corresponding to each trigger category (*eTOS*, *hTOS!*, *TIS!*). The results of this fit to the data are shown in Fig. 221 and the obtained parameters are listed in Tab. 56.
- **RKFitYears:** Two independent values of R_K corresponding to the Run 1 and Run 2 samples. The results of this fit to the data are shown in Fig. 220 and the obtained parameters are listed in Tab. 54.
- **RKFitComb:** Only one value of R_K allowed to float. The results of this fit to the data are shown in Fig. 223 and the obtained parameters are listed in Tab. 57.

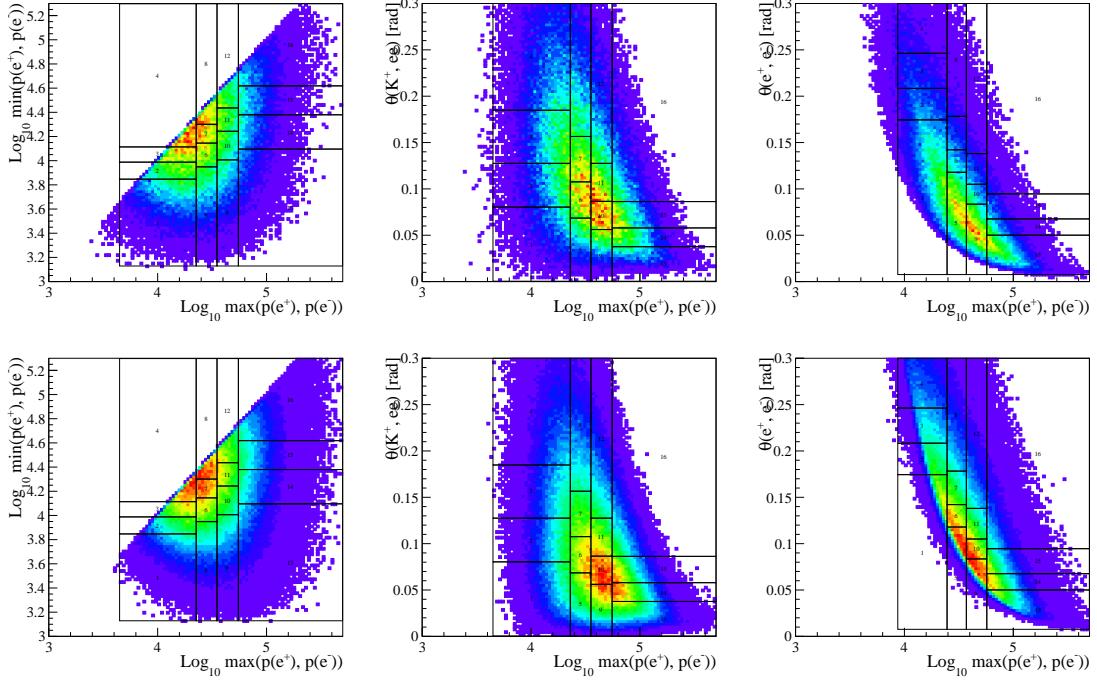


Figure 212: Two dimensional (left) $\max p_\ell$, $\min p_\ell$, (middle) $\max p_\ell$, $\alpha_{\ell^+\ell^-}$ and (right) $\max p_\ell$, $\alpha(\ell^+, \ell^-)$ distributions for reconstructed (top) $B^+ \rightarrow K^+ e^+ e^-$ and (bottom) $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events. The bins used in the $r_{J/\psi}^{\text{bin}}$ test in Run 1 are shown on both figures.

3111 P Correlation between R_K and \mathcal{B} ($B^+ \rightarrow K^+ \mu^+ \mu^-$)

3112 Assuming the correlation between R_K and \mathcal{B} ($B^+ \rightarrow K^+ \mu^+ \mu^-$) (noted as \mathcal{B}_μ) arises only
 3113 due to the number of $B^+ \rightarrow K^+ \mu^+ \mu^-$ decays observed in the Run 1 sample, $N_{\mu\mu}^{R1}$, the
 3114 correlation factor can be computed as,

$$\begin{aligned} \text{corr}(\mathcal{B}_\mu, R_K) &= \frac{\text{cov}(\mathcal{B}_\mu, R_K)}{\sigma(\mathcal{B}_\mu)\sigma(R_K)} \\ &= \frac{1}{\sigma(\mathcal{B}_\mu)\sigma(R_K)} \int (\mathcal{B}_\mu - \langle \mathcal{B}_\mu \rangle)(R_K - \langle R_K \rangle) g(N_{\mu\mu}^{R1}) dN_{\mu\mu}^{R1}, \end{aligned} \quad (43)$$

3115 where $g(N_{\mu\mu}^{R1})$ is the pdf for $N_{\mu\mu}^{R1}$, which can be described by a Gaussian centered around
 3116 the observed value, $\mu(N_{\mu\mu}^{R1})$, and with width equal to the measured uncertainty, $\sigma(N_{\mu\mu}^{R1})$.

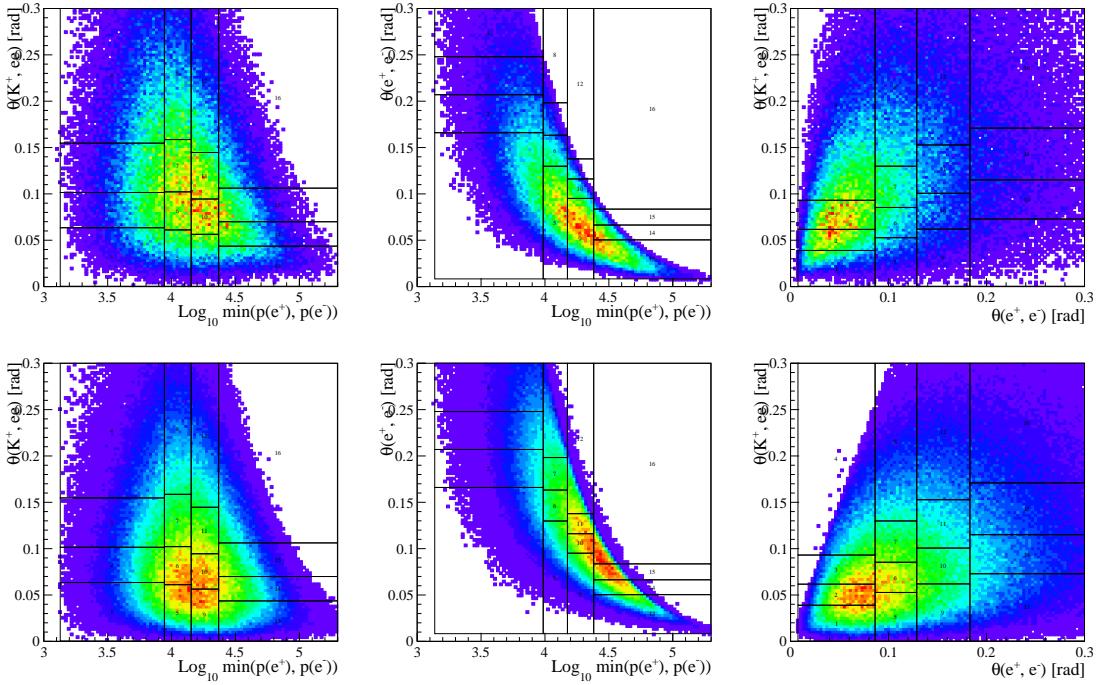


Figure 213: Two dimensional (left) $\min p_\ell$, $\alpha(\ell^+, \ell^-)$, (middle) $\min p_\ell$, $\alpha_{\ell+\ell^-}$ and (right) $\alpha_{\ell+\ell^-}$, $\alpha(\ell^+, \ell^-)$ distributions for reconstructed (top) $B^+ \rightarrow K^+ e^+ e^-$ and (bottom) $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events. The bins used in the $r_{J/\psi}^{\text{bin}}$ test in Run 1 are shown on both figures.

3117 Considering the dependences,

$$\begin{aligned} \mathcal{B}_\mu &= \frac{\epsilon(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))}{\epsilon(B^+ \rightarrow K^+ \mu^+ \mu^-)} \cdot \frac{N_{\mu\mu}^{R1}}{N_{J/\psi}^{R1}} = \alpha \cdot N_{\mu\mu}^{R1} \\ R_K &= (r_{J/\psi})^{-1} \cdot \frac{\epsilon(B^+ \rightarrow K^+ e^+ e^-)}{\epsilon(B^+ \rightarrow K^+ \mu^+ \mu^-)} \cdot \frac{N_{\mu\mu}^{R1} + N_{\mu\mu}^{R2}}{N_{ee}^{R1} + N_{ee}^{R2}} = \frac{N_{\mu\mu}^{R1} + b}{\beta} \end{aligned} \quad (44)$$

3118 the correlation factor can be computed analytically,

$$\text{corr}(\mathcal{B}_\mu, R_K) = \frac{1}{\sigma(\mathcal{B}_\mu)\sigma(R_K)} \cdot \frac{\alpha \cdot \sigma(N_{\mu\mu}^{R1})^2}{\beta}. \quad (45)$$

3119 Note that the overlap between the samples used to measure the \mathcal{B}_μ and R_K is encoded
3120 in the parameter β . Using this formula and the values given in Ref. [7] the correlation is
3121 estimated to be 29%.

3122 The statistical uncertainty in \mathcal{B} ($B^+ \rightarrow K^+ e^+ e^-$) can, therefore, be computed as

$$\sigma^2(\mathcal{B}_e) = \left| \frac{\partial \mathcal{B}_e}{\partial \mathcal{B}_\mu} \right|^2 \sigma^2(\mathcal{B}_\mu) + \left| \frac{\partial \mathcal{B}_e}{\partial R_K} \right|^2 \sigma^2(R_K) + 2 \frac{\partial \mathcal{B}_e}{\partial \mathcal{B}_\mu} \frac{\partial \mathcal{B}_e}{\partial R_K} \text{corr}(\mathcal{B}_\mu, R_K) \sigma(\mathcal{B}_\mu) \sigma(R_K). \quad (46)$$

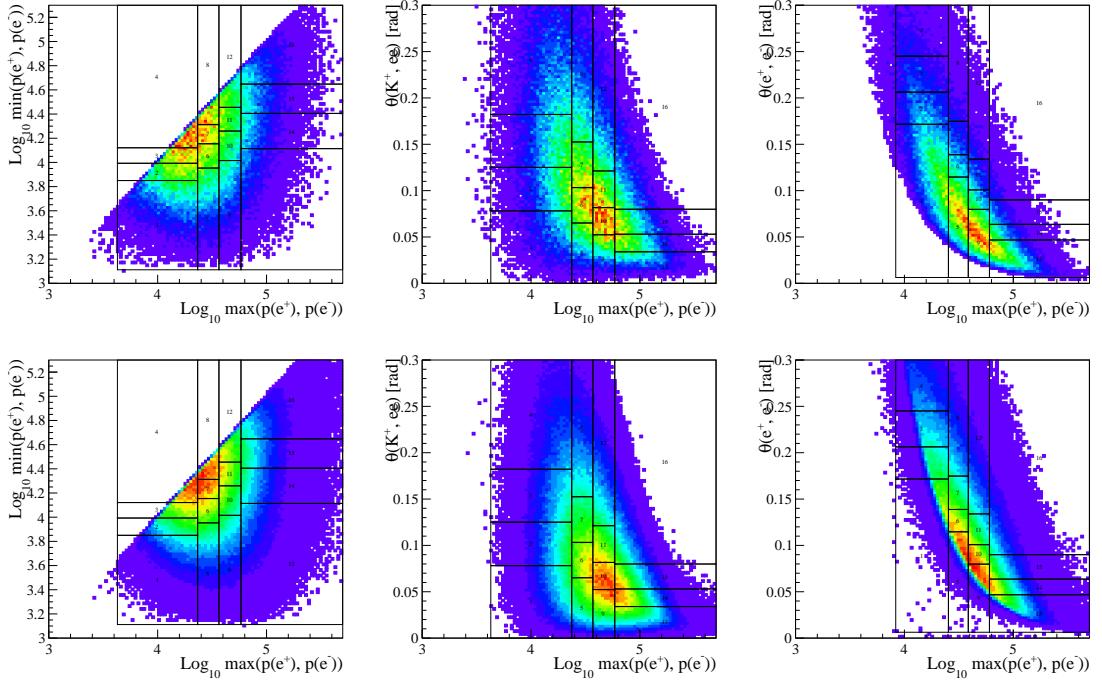


Figure 214: Two dimensional (left) $\max p_\ell$, $\min p_\ell$, (middle) $\max p_\ell$, $\alpha_{\ell^+\ell^-}$ and (right) $\max p_\ell$, $\alpha(\ell^+, \ell^-)$ distributions for reconstructed (top) $B^+ \rightarrow K^+ e^+ e^-$ and (bottom) $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events. The bins used in the $r_{J/\psi}^{\text{bin}}$ test in Run 2 are shown on both figures.

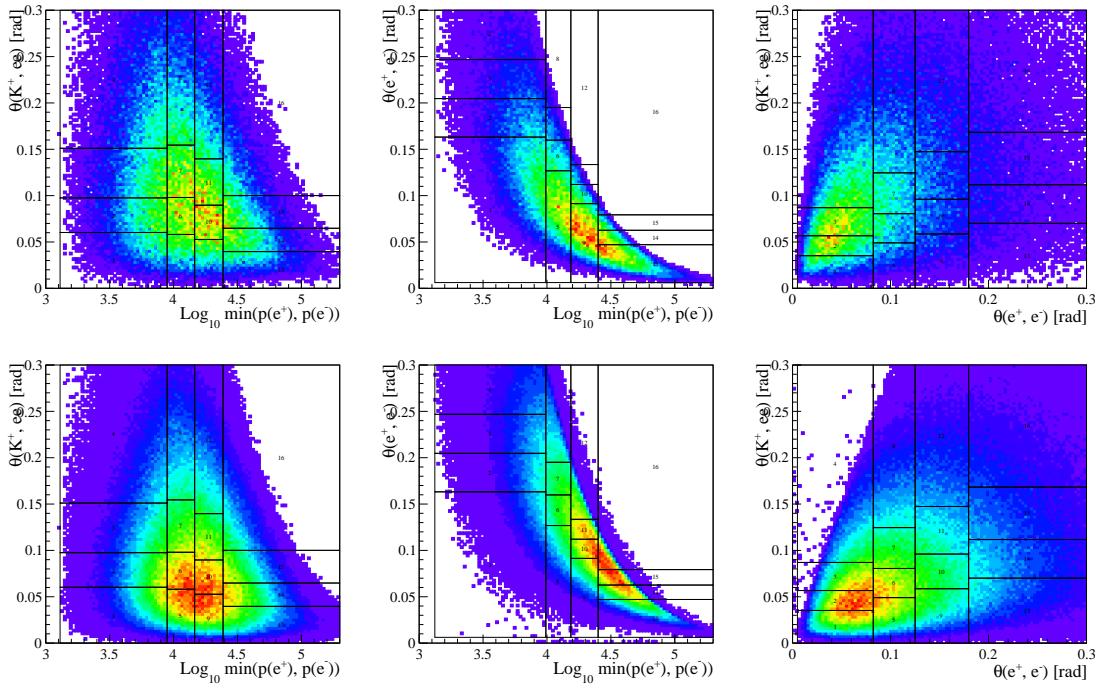


Figure 215: Two dimensional (left) $\min p_\ell$, $\alpha(\ell^+, \ell^-)$, (middle) $\min p_\ell$, $\alpha_{\ell+\ell^-}$ and (right) $\alpha_{\ell+\ell^-}$, $\alpha(\ell^+, \ell^-)$ distributions for reconstructed (top) $B^+ \rightarrow K^+ e^+ e^-$ and (bottom) $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events. The bins used in the $r_{J/\psi}^{\text{bin}}$ test in Run 2 are shown on both figures.

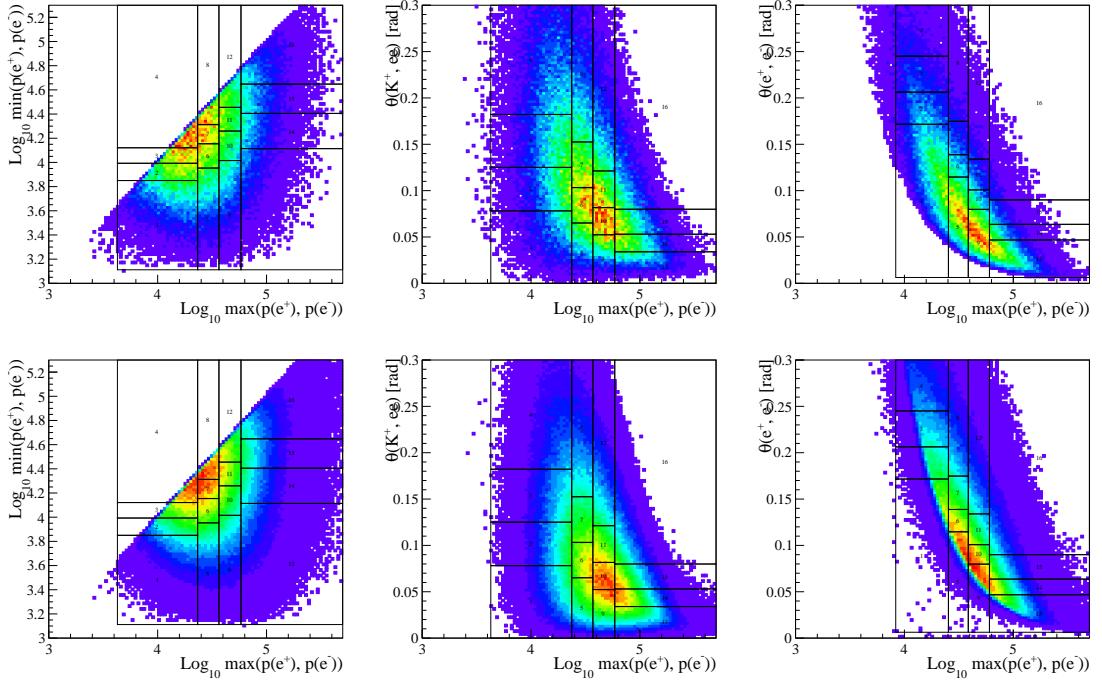


Figure 216: Two dimensional (left) $\max p_\ell$, $\min p_\ell$, (middle) $\max p_\ell$, $\alpha_{\ell^+\ell^-}$ and (right) $\max p_\ell$, $\alpha(\ell^+, \ell^-)$ distributions for reconstructed (top) $B^+ \rightarrow K^+ e^+ e^-$ and (bottom) $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events. The bins used in the $r_{J/\psi}^{\text{bin}}$ test in Run 2 are shown on both figures.

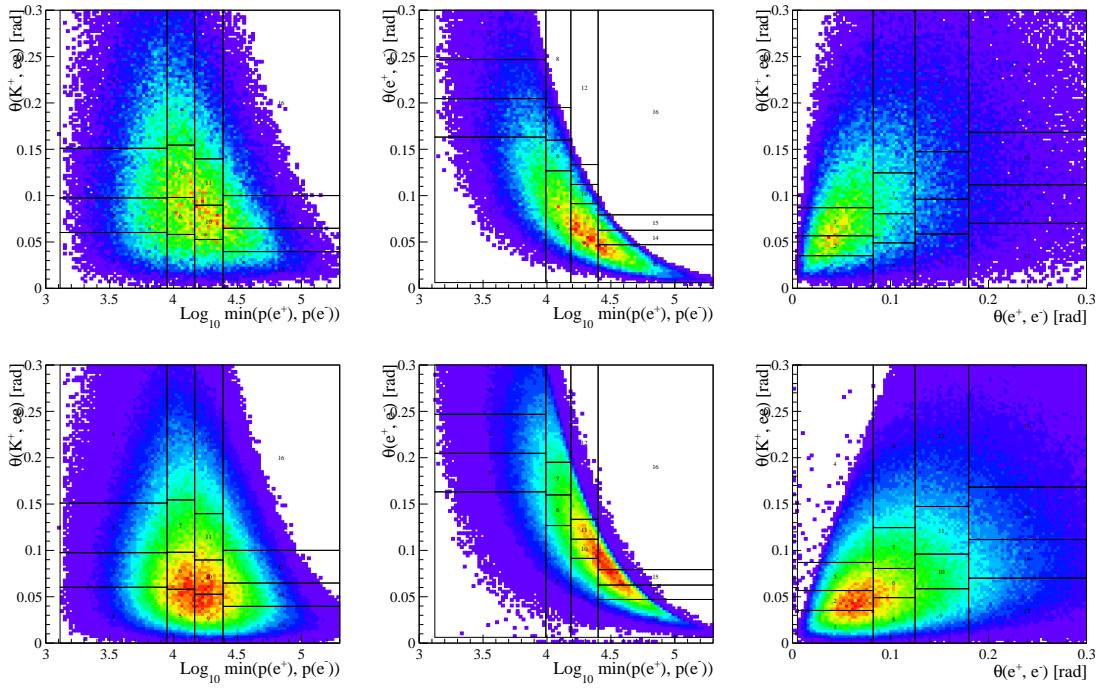


Figure 217: Two dimensional (left) $\min p_\ell$, $\alpha(\ell^+, \ell^-)$, (middle) $\min p_\ell$, $\alpha_{\ell+\ell^-}$ and (right) $\alpha_{\ell+\ell^-}$, $\alpha(\ell^+, \ell^-)$ distributions for reconstructed (top) $B^+ \rightarrow K^+ e^+ e^-$ and (bottom) $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events. The bins used in the $r_{J/\psi}^{\text{bin}}$ test in Run 2 are shown on both figures.

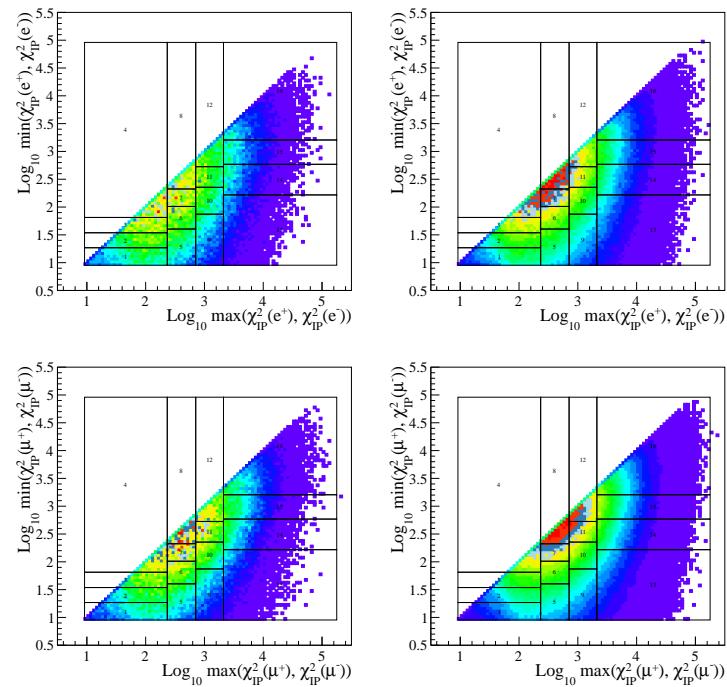
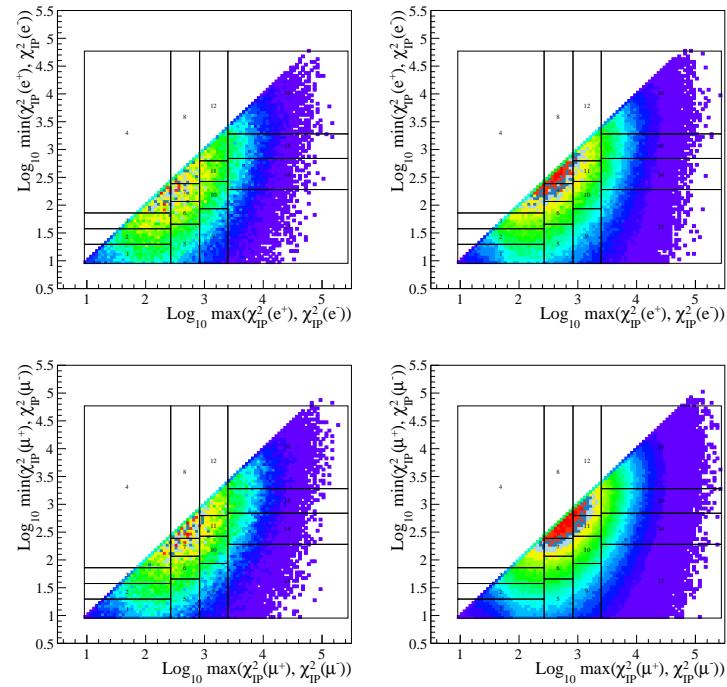


Figure 218: Two dimensional $\chi^2_{IP}(\ell^+)$, $\chi^2_{IP}(\ell^-)$ distribution for reconstructed (left) $B^+ \rightarrow K^+ \ell^+ \ell^-$ and (right) $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ simulated events. The bins used in the $r_{J/\psi}^{\text{bin}}$ test in (top) Run 1 and (bottom) Run 2 are shown on the figures.

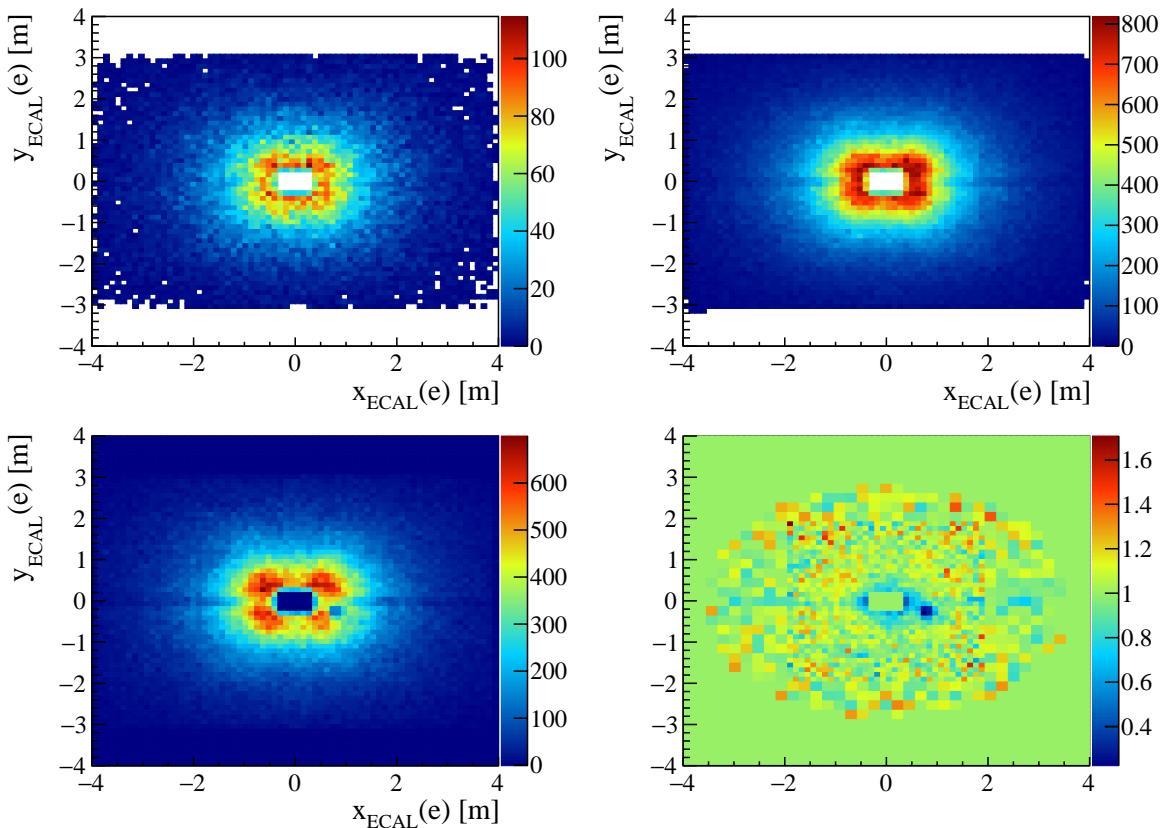


Figure 219: Illumination of the ECAL from (top left) $B^+ \rightarrow K^+ e^+ e^-$ and (top right) $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulation, and (bottom left) $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ sWeighted data, in the Run 2 eTOS category. The ratio between $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ data and simulation (bottom right) is also shown.

Table 53: Total efficiencies for rare and normalisation modes computed using simulation only (no corrections) and simulation corrected using the nominal procedure (all corrections) and the inverse of the nominal weights (1/corrections).

$\varepsilon[\%]$	muon modes		electron modes		R_K eff ratio
	J/ψ	rare	J/ψ	rare	
Run 1 no corrections					
ETOS	0.0157755	0.00366284	0.00301603	0.000747831	1.0679
HTOS!	0.0157755	0.00366284	0.000594615	0.000300272	2.17492
TIS!	0.0157755	0.00366284	0.00100641	0.000275599	1.17942
Run 1 all corrections					
ETOS	0.0139644	0.00316514	0.00216682	0.000544183	1.10803
HTOS!	0.0139644	0.00316514	0.000350145	0.000177563	2.23735
TIS!	0.0139644	0.00316514	0.000695536	0.000192505	1.2211
Run 2 no corrections					
ETOS	0.0142339	0.00338602	0.00468164	0.00117686	1.05672
HTOS!	0.0142339	0.00338602	0.000502993	0.000300688	2.51298
TIS!	0.0142339	0.00338602	0.00103467	0.000299772	1.21792
Run 2 all corrections					
ETOS	0.0143224	0.00332971	0.0041876	0.00104055	1.06882
HTOS!	0.0143224	0.00332971	0.000465853	0.000264433	2.44161
TIS!	0.0143224	0.00332971	0.00109503	0.000316485	1.24319
Run 2 1/corrections					
ETOS	0.0178365	0.00484999	0.0139854	0.00294701	0.774952
HTOS!	0.0178365	0.00484999	0.00188186	0.000828374	1.61885
TIS!	0.0178365	0.00484999	0.00281255	0.000670153	0.87628

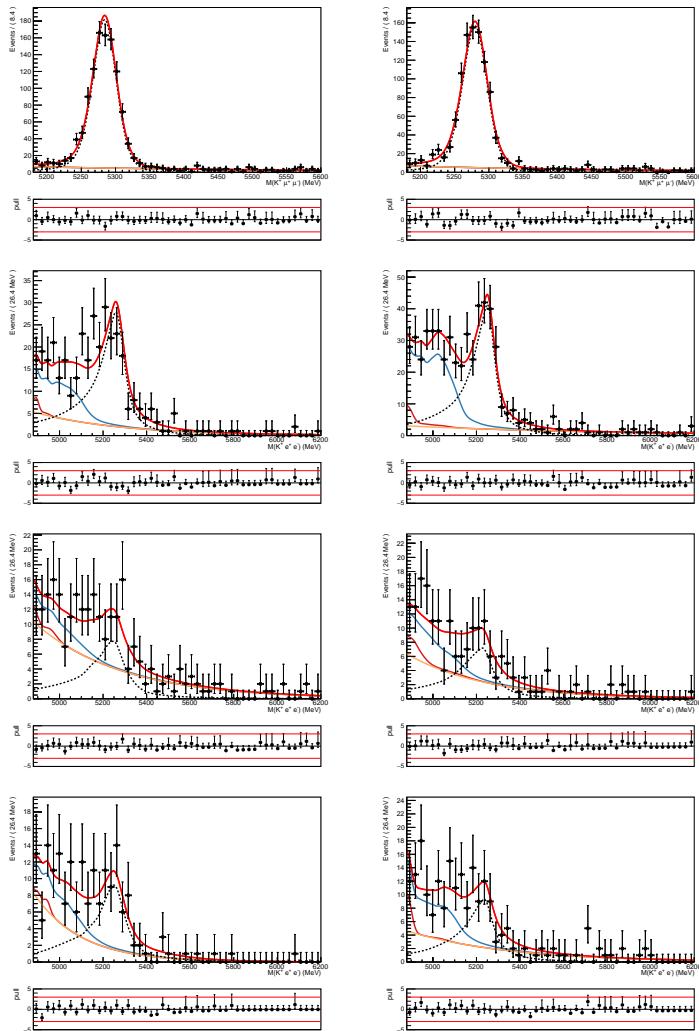


Figure 220: The fit to the $m(K^+\ell^+\ell^-)$ invariant mass distribution of selected candidates in the (left) Run 1 and (right) Run 2 data samples, when R_K^{R1} and R_K^{R2} are allowed to float independently. The top row shows the fit to the muon modes and the subsequent rows the fits to the electron modes triggered by (second row) one of the electrons, (third row) the kaon from the B^+ decay and (last row) by other particles in the event. The points show the LHCb data, the red line the total fit, the black-dashed line the $B^+ \rightarrow K^+\ell^+\ell^-$ signal component, the orange line the combinatorial background, the blue line the background from exclusive b -hadron decays and the dark red the radiative tail of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events.

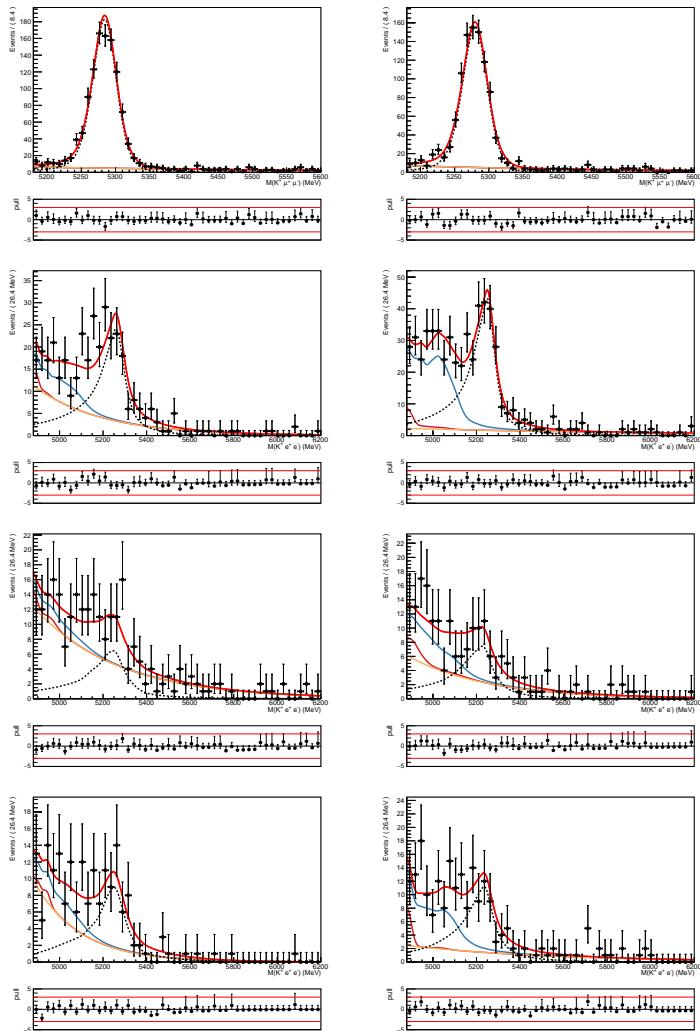


Figure 221: The fit to the $m(K^+\ell^+\ell^-)$ invariant mass distribution of selected candidates in the (left) Run 1 and (right) Run 2 data samples, when $R_K^{e\text{TOS}}$, $R_K^{h\text{TOS!}}$ and $R_K^{\text{TIS!}}$ are allowed to float independently. The top row shows the fit to the muon modes and the subsequent rows the fits to the electron modes triggered by (second row) one of the electrons, (third row) the kaon from the B^+ decay and (last row) by other particles in the event. The points show the LHCb data, the red line the total fit, the black-dashed line the $B^+ \rightarrow K^+\ell^+\ell^-$ signal component, the orange line the combinatorial background, the blue line the background from exclusive b -hadron decays and the dark red the radiative tail of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events.

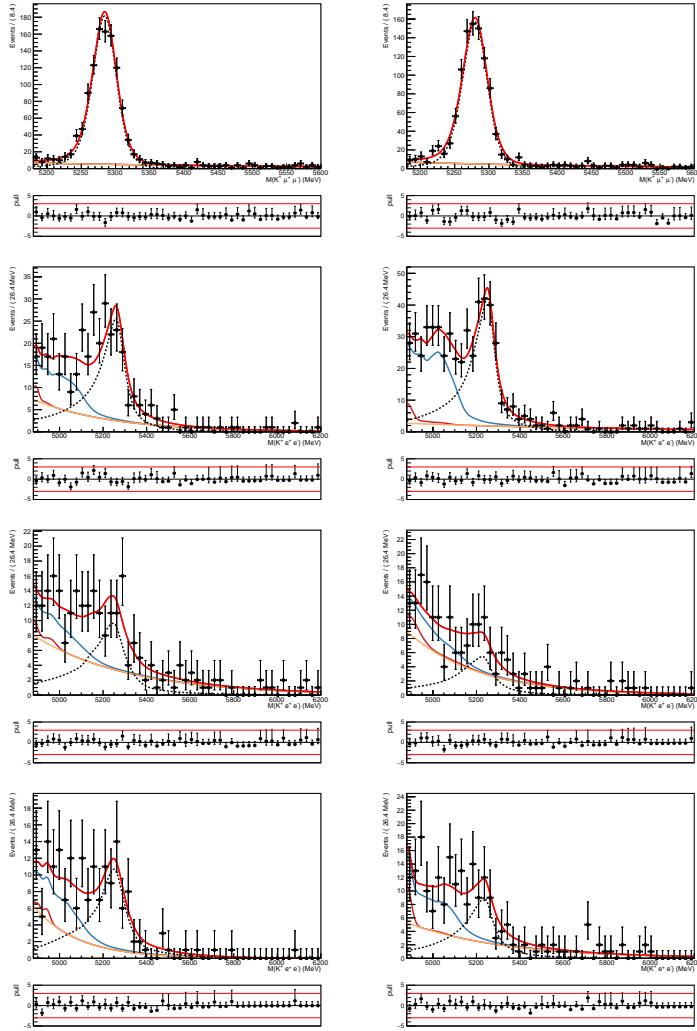


Figure 222: The fit to the $m(K^+\ell^+\ell^-)$ invariant mass distribution of selected candidates in the (left) Run 1 and (right) Run 2 data samples, when $R_K^{e\text{TOS},\text{R1}}$, $R_K^{h\text{TOS},\text{R1}}$, $R_K^{\text{TIS},\text{R1}}$, $R_K^{e\text{TOS},\text{R2}}$, $R_K^{h\text{TOS},\text{R2}}$ and $R_K^{\text{TIS},\text{R2}}$ are allowed to float independently. The top row shows the fit to the muon modes and the subsequent rows the fits to the electron modes triggered by (second row) one of the electrons, (third row) the kaon from the B^+ decay and (last row) by other particles in the event. The points show the LHCb data, the red line the total fit, the black-dashed line the $B^+ \rightarrow K^+\ell^+\ell^-$ signal component, the orange line the combinatorial background, the blue line the background from exclusive b -hadron decays and the dark red the radiative tail of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events.

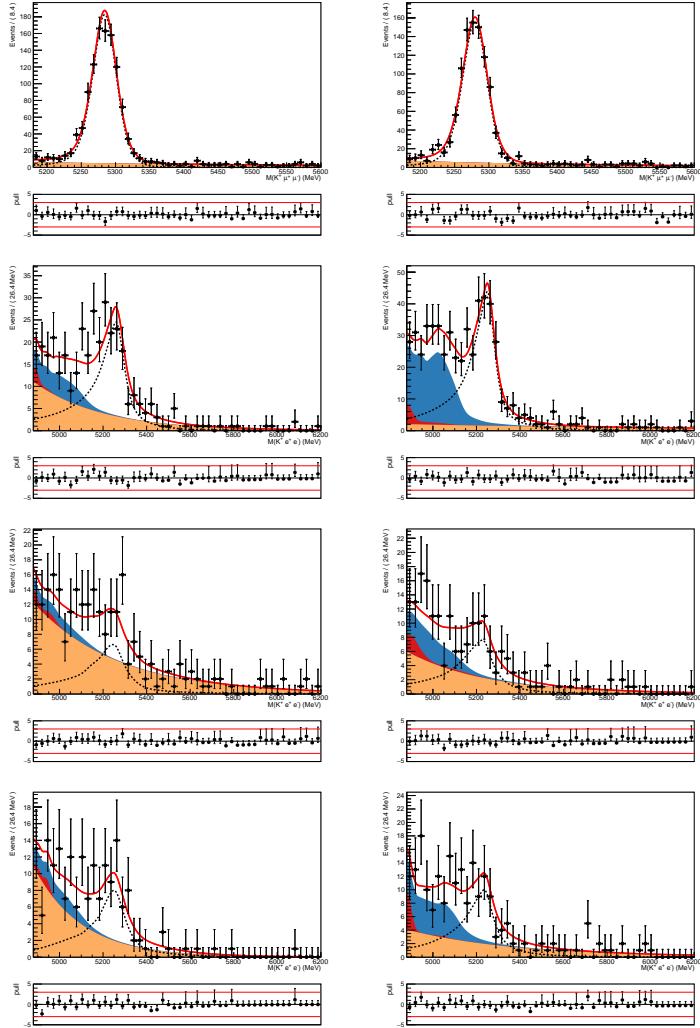


Figure 223: The fit to the $m(K^+\ell^+\ell^-)$ invariant mass distribution of selected candidates in the (left) Run 1 and (right) Run 2 data samples. The top row shows the fit to the muon modes and the subsequent rows the fits to the electron modes triggered by (second row) one of the electrons, (third row) the kaon from the B^+ decay and (last row) by other particles in the event. The points show the LHCb data, the red line the total fit, the black-dashed line the $B^+ \rightarrow K^+\ell^+\ell^-$ signal component, the orange line the combinatorial background, the blue line the background from exclusive b -hadron decays and the dark red the radiative tail of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events.

Table 54: Results of the fit to the data when R_K^{R1} and R_K^{R2} are floated independently.

Parameter	Value	Parameter	Value
R_K^{R1}	$0.711^{+0.084}_{-0.072}$	nCombKeeETOSR1	78^{+54}_{-44}
R_K^{R2}	$0.934^{+0.090}_{-0.078}$	nCombKeeETOSR2	80^{+37}_{-26}
nSignalKmumuR1	1027^{+36}_{-35}	nCombKeeHTOSR1	150^{+24}_{-23}
nSignalKmumuR2	916^{+34}_{-33}	nCombKeeHTOSR2	85^{+18}_{-18}
cETOSR1	$14.55^{+0.24}_{-0.24}$	nCombKeeTISR1	66^{+24}_{-21}
cETOSR2	$28.08^{+0.32}_{-0.32}$	nCombKeeTISR2	77^{+18}_{-17}
cHTOSR1	$5.03^{+0.19}_{-0.19}$	nCombKmumuR1	208^{+22}_{-21}
cHTOSR2	$6.93^{+0.59}_{-0.59}$	nCombKmumuR2	190^{+21}_{-20}
cTISR1	$5.42^{+0.16}_{-0.16}$	expoConstKeeETOSR1 ($\times 10^{-3}$)	$-2.54^{+1.55}_{-0.97}$
cTISR2	$8.05^{+0.19}_{-0.19}$	expoConstKeeETOSR2 ($\times 10^{-3}$)	$-1.10^{+0.88}_{-0.80}$
fracZeroKeeTrig0R1	$0.236^{+0.010}_{-0.010}$	expoConstKeeHTOSR1 ($\times 10^{-3}$)	$-2.49^{+0.36}_{-0.35}$
fracZeroKeeTrig0R2	$0.253^{+0.010}_{-0.010}$	expoConstKeeHTOSR2 ($\times 10^{-3}$)	$-2.81^{+0.54}_{-0.52}$
fracZeroKeeTrig1R1	$0.204^{+0.010}_{-0.010}$	expoConstKeeTISR1 ($\times 10^{-3}$)	$-4.65^{+0.84}_{-1.01}$
fracZeroKeeTrig1R2	$0.214^{+0.010}_{-0.010}$	expoConstKeeTISR2 ($\times 10^{-3}$)	$-2.16^{+0.56}_{-0.51}$
fracZeroKeeTrig2R1	$0.207^{+0.010}_{-0.010}$	expoConstKmumuR1 ($\times 10^{-3}$)	$-2.44^{+0.79}_{-0.75}$
fracZeroKeeTrig2R2	$0.204^{+0.010}_{-0.010}$	expoConstKmumuR2 ($\times 10^{-3}$)	$-3.21^{+0.82}_{-0.79}$
fracOneRecKeeTrig0R1	$0.652^{+0.010}_{-0.010}$	nCharmLeakKeeETOSR1	$4.5^{+2.0}_{-2.0}$
fracOneRecKeeTrig0R2	$0.667^{+0.010}_{-0.010}$	nCharmLeakKeeETOSR2	$11.9^{+3.1}_{-3.1}$
fracOneRecKeeTrig1R1	$0.632^{+0.010}_{-0.010}$	nCharmLeakKeeHTOSR1	$3.9^{+1.8}_{-1.8}$
fracOneRecKeeTrig1R2	$0.630^{+0.010}_{-0.010}$	nCharmLeakKeeHTOSR2	$8.4^{+2.6}_{-2.7}$
fracOneRecKeeTrig2R1	$0.625^{+0.010}_{-0.010}$	nCharmLeakKeeTISR1	$2.7^{+2.0}_{-2.0}$
fracOneRecKeeTrig2R2	$0.606^{+0.010}_{-0.010}$	nCharmLeakKeeTISR2	$5.8^{+2.2}_{-2.2}$
nPartRecoKeeETOSR1	72^{+32}_{-41}		
nPartRecoKeeETOSR2	192^{+24}_{-28}		
rPRHTOSR1	$0.92^{+0.13}_{-0.13}$		
rPRTISR1	$1.075^{+0.014}_{-0.015}$		

Table 55: Results of the fit to the data when $R_K^{e\text{TOS},\text{R1}}$, $R_K^{e\text{TOS},\text{R2}}$, $R_K^{\text{TIS}!,\text{R1}}$, $R_K^{e\text{TOS},\text{R2}}$, $R_K^{h\text{TOS}!,\text{R2}}$ and $R_K^{\text{TIS}!,\text{R2}}$ are floated independently.

Parameter	Value	Parameter	Value
$R_K^{e\text{TOS},\text{R1}}$	$0.79^{+0.11}_{-0.10}$	nPartRecoKeeETOSR1	68^{+32}_{-35}
$R_K^{e\text{TOS},\text{R2}}$	$0.903^{+0.094}_{-0.081}$	nPartRecoKeeETOSR2	191^{+23}_{-26}
$R_K^{h\text{TOS}!,\text{R1}}$	$0.56^{+0.19}_{-0.12}$	rPRHTOSR1	$0.93^{+0.13}_{-0.13}$
$R_K^{h\text{TOS}!,\text{R2}}$	$1.28^{+0.65}_{-0.35}$	rPRTISR1	$1.075^{+0.015}_{-0.015}$
$R_K^{\text{TIS}!,\text{R1}}$	$0.60^{+0.13}_{-0.10}$	nCombKeeETOSR1	96^{+47}_{-46}
$R_K^{\text{TIS}!,\text{R2}}$	$0.97^{+0.32}_{-0.21}$	nCombKeeETOSR2	75^{+36}_{-25}
nSignalKmumuR1	1027^{+36}_{-35}	nCombKeeHTOSR1	126^{+37}_{-42}
nSignalKmumuR2	916^{+34}_{-33}	nCombKeeHTOSR2	109^{+28}_{-29}
cETOSR1	$14.55^{+0.24}_{-0.24}$	nCombKeeTISR1	48^{+28}_{-27}
cETOSR2	$28.09^{+0.32}_{-0.32}$	nCombKeeTISR2	83^{+32}_{-33}
cHTOSR1	$5.01^{+0.19}_{-0.19}$	nCombKmumuR1	208^{+22}_{-21}
cHTOSR2	$7.10^{+0.61}_{-0.61}$	nCombKmumuR2	190^{+21}_{-20}
cTISR1	$5.40^{+0.16}_{-0.16}$	expoConstKeeETOSR1 ($\times 10^{-3}$)	$-2.79^{+1.15}_{-0.73}$
cTISR2	$8.06^{+0.19}_{-0.19}$	expoConstKeeETOSR2 ($\times 10^{-3}$)	$-0.98^{+0.91}_{-0.82}$
fracOneRecKeeTrig0R1	$0.652^{+0.010}_{-0.010}$	expoConstKeeHTOSR1 ($\times 10^{-3}$)	$-2.29^{+0.70}_{-0.46}$
fracOneRecKeeTrig0R2	$0.667^{+0.010}_{-0.010}$	expoConstKeeHTOSR2 ($\times 10^{-3}$)	$-3.09^{+0.53}_{-0.45}$
fracOneRecKeeTrig1R1	$0.632^{+0.010}_{-0.010}$	expoConstKeeTISR1 ($\times 10^{-3}$)	$-4.91^{+1.41}_{-\text{limit}}$
fracOneRecKeeTrig1R2	$0.630^{+0.010}_{-0.010}$	expoConstKeeTISR2 ($\times 10^{-3}$)	$-2.29^{+0.99}_{-0.62}$
fracOneRecKeeTrig2R1	$0.625^{+0.010}_{-0.010}$	expoConstKmumuR1 ($\times 10^{-3}$)	$-2.44^{+0.79}_{-0.75}$
fracOneRecKeeTrig2R2	$0.606^{+0.010}_{-0.010}$	expoConstKmumuR2 ($\times 10^{-3}$)	$-3.21^{+0.82}_{-0.79}$
fracZeroKeeTrig0R1	$0.236^{+0.010}_{-0.010}$	nCharmLeakKeeETOSR1	$4.3^{+2.0}_{-2.0}$
fracZeroKeeTrig0R2	$0.253^{+0.010}_{-0.010}$	nCharmLeakKeeETOSR2	$11.9^{+3.1}_{-3.1}$
fracZeroKeeTrig1R1	$0.204^{+0.010}_{-0.010}$	nCharmLeakKeeHTOSR1	$4.1^{+1.8}_{-1.8}$
fracZeroKeeTrig1R2	$0.214^{+0.010}_{-0.010}$	nCharmLeakKeeHTOSR2	$8.1^{+2.7}_{-2.7}$
fracZeroKeeTrig2R1	$0.207^{+0.010}_{-0.010}$	nCharmLeakKeeTISR1	$2.9^{+2.0}_{-2.0}$
fracZeroKeeTrig2R2	$0.204^{+0.010}_{-0.010}$	nCharmLeakKeeTISR2	$5.8^{+2.3}_{-2.2}$

Table 56: Results of the fit to the data when $R_K^{e\text{TOS}}$, $R_K^{h\text{TOS!}}$ and $R_K^{\text{TIS!}}$ are floated independently.

Parameter	Value	Parameter	Value
$R_K^{e\text{TOS}}$	$0.866^{+0.071}_{-0.063}$	nCombKeeETOSR1	129^{+44}_{-42}
$R_K^{h\text{TOS!}}$	$0.87^{+0.24}_{-0.17}$	nCombKeeETOSR2	69^{+30}_{-22}
$R_K^{\text{TIS!}}$	$0.746^{+0.128}_{-0.091}$	nCombKeeHTOSR1	170^{+25}_{-27}
nSignalKmumuR1	1043^{+35}_{-35}	nCombKeeHTOSR2	83^{+25}_{-27}
nSignalKmumuR2	900^{+33}_{-32}	nCombKeeTISR1	76^{+25}_{-24}
cETOSR1	$14.58^{+0.24}_{-0.24}$	nCombKeeTISR2	54^{+26}_{-20}
cETOSR2	$28.06^{+0.32}_{-0.32}$	nCombKmumuR1	205^{+22}_{-20}
cHTOSR1	$5.04^{+0.19}_{-0.19}$	nCombKmumuR2	193^{+21}_{-20}
cHTOSR2	$6.85^{+0.61}_{-0.61}$	expoConstKeeETOSR1 ($\times 10^{-3}$)	$-3.22^{+0.73}_{-0.58}$
cTISR1	$5.43^{+0.16}_{-0.16}$	expoConstKeeETOSR2 ($\times 10^{-3}$)	$-0.81^{+0.87}_{-0.79}$
cTISR2	$8.04^{+0.19}_{-0.19}$	expoConstKeeHTOSR1 ($\times 10^{-3}$)	$-2.63^{+0.33}_{-0.32}$
fracZeroKeeTrig0R1	$0.236^{+0.010}_{-0.010}$	expoConstKeeHTOSR2 ($\times 10^{-3}$)	$-2.79^{+0.73}_{-0.56}$
fracZeroKeeTrig0R2	$0.253^{+0.010}_{-0.010}$	expoConstKeeTISR1 ($\times 10^{-3}$)	$-4.74^{+0.78}_{-0.93}$
fracZeroKeeTrig1R1	$0.204^{+0.010}_{-0.010}$	expoConstKeeTISR2 ($\times 10^{-3}$)	$-1.53^{+1.05}_{-0.83}$
fracZeroKeeTrig1R2	$0.214^{+0.010}_{-0.010}$	expoConstKmumuR1 ($\times 10^{-3}$)	$-2.34^{+0.80}_{-0.76}$
fracZeroKeeTrig2R1	$0.207^{+0.010}_{-0.010}$	expoConstKmumuR2 ($\times 10^{-3}$)	$-3.31^{+0.81}_{-0.78}$
fracZeroKeeTrig2R2	$0.204^{+0.010}_{-0.010}$	nCharmLeakKeeETOSR1	$4.1^{+2.0}_{-2.0}$
fracOneRecKeeTrig0R1	$0.652^{+0.010}_{-0.010}$	nCharmLeakKeeETOSR2	$11.9^{+3.1}_{-3.1}$
fracOneRecKeeTrig0R2	$0.667^{+0.010}_{-0.010}$	nCharmLeakKeeHTOSR1	$3.8^{+1.8}_{-1.8}$
fracOneRecKeeTrig1R1	$0.632^{+0.010}_{-0.010}$	nCharmLeakKeeHTOSR2	$8.4^{+2.7}_{-2.7}$
fracOneRecKeeTrig1R2	$0.630^{+0.010}_{-0.010}$	nCharmLeakKeeTISR1	$2.7^{+2.0}_{-2.0}$
fracOneRecKeeTrig2R1	$0.625^{+0.010}_{-0.010}$	nCharmLeakKeeTISR2	$6.0^{+2.2}_{-2.2}$
fracOneRecKeeTrig2R2	$0.606^{+0.010}_{-0.010}$		
nPartRecoKeeETOSR1	45^{+34}_{-36}		
nPartRecoKeeETOSR2	194^{+22}_{-24}		
rPRHTOSR1	$0.92^{+0.13}_{-0.13}$		
rPRTISR1	$1.075^{+0.014}_{-0.015}$		

Table 57: Results of the fit to the data when only one value of R_K is allowed to float.

Parameter	Value	Parameter	Value
R_K	$0.848^{+0.061}_{-0.055}$	nCombKeeETOSR1	127^{+46}_{-44}
nSignalKmumuR1	1044^{+35}_{-35}	nCombKeeETOSR2	67^{+29}_{-21}
nSignalKmumuR2	899^{+33}_{-32}	nCombKeeHTOSR1	168^{+21}_{-21}
cETOSR1	$14.59^{+0.24}_{-0.24}$	nCombKeeHTOSR2	81^{+18}_{-18}
cETOSR2	$28.05^{+0.32}_{-0.32}$	nCombKeeTISR1	86^{+22}_{-21}
cHTOSR1	$5.04^{+0.19}_{-0.19}$	nCombKeeTISR2	71^{+17}_{-16}
cHTOSR2	$6.84^{+0.58}_{-0.58}$	nCombKmumuR1	205^{+22}_{-20}
cTISR1	$5.44^{+0.16}_{-0.16}$	nCombKmumuR2	193^{+21}_{-20}
cTISR2	$8.05^{+0.19}_{-0.19}$	expoConstKeeETOSR1 ($\times 10^{-3}$)	$-3.23^{+0.77}_{-0.60}$
fracZeroKeeTrig0R1	$0.236^{+0.010}_{-0.010}$	expoConstKeeETOSR2 ($\times 10^{-3}$)	$-0.73^{+0.87}_{-0.79}$
fracZeroKeeTrig0R2	$0.253^{+0.010}_{-0.010}$	expoConstKeeHTOSR1 ($\times 10^{-3}$)	$-2.63^{+0.32}_{-0.32}$
fracZeroKeeTrig1R1	$0.204^{+0.010}_{-0.010}$	expoConstKeeHTOSR2 ($\times 10^{-3}$)	$-2.77^{+0.56}_{-0.54}$
fracZeroKeeTrig1R2	$0.214^{+0.010}_{-0.010}$	expoConstKeeTISR1 ($\times 10^{-3}$)	$-4.61^{+0.66}_{-0.73}$
fracZeroKeeTrig2R1	$0.207^{+0.010}_{-0.010}$	expoConstKeeTISR2 ($\times 10^{-3}$)	$-2.04^{+0.58}_{-0.53}$
fracZeroKeeTrig2R2	$0.204^{+0.010}_{-0.010}$	expoConstKmumuR1 ($\times 10^{-3}$)	$-2.34^{+0.80}_{-0.76}$
fracOneRecKeeTrig0R1	$0.652^{+0.010}_{-0.010}$	expoConstKmumuR2 ($\times 10^{-3}$)	$-3.32^{+0.81}_{-0.78}$
fracOneRecKeeTrig0R2	$0.667^{+0.010}_{-0.010}$	nCharmLeakKeeETOSR1	$4.1^{+2.0}_{-2.0}$
fracOneRecKeeTrig1R1	$0.632^{+0.010}_{-0.010}$	nCharmLeakKeeETOSR2	$12.0^{+3.1}_{-3.1}$
fracOneRecKeeTrig1R2	$0.630^{+0.010}_{-0.010}$	nCharmLeakKeeHTOSR1	$3.8^{+1.8}_{-1.8}$
fracOneRecKeeTrig2R1	$0.625^{+0.010}_{-0.010}$	nCharmLeakKeeHTOSR2	$8.5^{+2.7}_{-2.7}$
fracOneRecKeeTrig2R2	$0.606^{+0.010}_{-0.010}$	nCharmLeakKeeTISR1	$2.6^{+2.0}_{-2.0}$
nPartRecoKeeETOSR1	44^{+35}_{-37}	nCharmLeakKeeTISR2	$5.9^{+2.2}_{-2.2}$
nPartRecoKeeETOSR2	194^{+22}_{-24}		
rPRHTOSR1	$0.92^{+0.13}_{-0.13}$		
rPRTISR1	$1.075^{+0.015}_{-0.015}$		

3123 References

- 3124 [1] G. Hiller and M. Schmaltz, *R_K and future b → sℓℓ BSM opportunities*, Phys. Rev. D
3125 **90** (2014) 054014.
- 3126 [2] M. Tanaka, *Charged Higgs Effects on Exclusive Semi-tauonic B Decays*, Z. Phys. C
3127 **67** (1995) 321.
- 3128 [3] C. Bobeth *et al.*, *Angular Distributions of B → Kll Decays*, JHEP **12** (2007) .
- 3129 [4] M. B. et al, *On the Standard Model predictions for R_K and R_{K*}*, A. Eur. Phys. J.
3130 **C76** (2016) 440.
- 3131 [5] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections*
3132 *in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, [arXiv:hep-ph/0506026](https://arxiv.org/abs/hep-ph/0506026).
- 3133 [6] LHCb collaboration, *Test of lepton universality using B⁺ → K⁺ℓ⁺ℓ decays*, Phys.
3134 Rev. Lett. **113** (2014) 151601.
- 3135 [7] LHCb, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of*
3136 *B → K^(*)μ⁺μ⁻ decays*, JHEP **06** (2014) 133, [arXiv:1403.8044](https://arxiv.org/abs/1403.8044).
- 3137 [8] LHCb Collaboration, *Angular analysis and differential branching fraction of the decay*
3138 *B_s⁰ → φμ⁺μ⁻*, JHEP **09** (2015) 179.
- 3139 [9] LHCb Collaboration, *Angular analysis of the B⁰ → K^{*0}μ⁺μ decay using 3 fb⁻¹ of*
3140 *integrated luminosity*, JHEP **02** (2016) 104.
- 3141 [10] BaBar Collaboration, *Measurement of Branching Fractions and Rate Asymmetries in*
3142 *the Rare Decays B → K^(*)l⁺l⁻*, Phys. Rev. **D86** (2012) 032012.
- 3143 [11] Belle collaboration, *Measurement of the Differential Branching Fraction and Forward-*
3144 *Backward Asymmetry for B → K^(*)l⁺l⁻*, Phys. Rev. **D103** (2009) 12.
- 3145 [12] S. D.-G. et al, *Global analysis of b → sℓℓ anomalies*, JHEP **06** (2016) 092.
- 3146 [13] T. H. et al, *On lepton non-universality in exclusive b → sℓℓ decays*, JHEP **1412**
3147 (2014) 053.
- 3148 [14] W. Altmannshofer and D. M. Straub, *New physics in b → s transitions after LHC*
3149 *run 1*, Eur. Phys. J. C **75** (2015) 382.
- 3150 [15] W. Altmannshofer, P. Stangl, and D. M. Straub, *Interpreting Hints for Lepton Flavor*
3151 *Universality Violation*, Phys. Rev. **D96** (2017), no. 5 055008, [arXiv:1704.05435](https://arxiv.org/abs/1704.05435).
- 3152 [16] B. Capdevila *et al.*, *Patterns of New Physics in b → sℓ⁺ℓ⁻ transitions in the light of*
3153 *recent data*, JHEP **01** (2018) 093, [arXiv:1704.05340](https://arxiv.org/abs/1704.05340).

- 3154 [17] LHCb collaboration, *Lepton universality test with $B^0 \rightarrow K^{*0}\ell^+\ell^-$ decays*, LHCb-
 3155 PAPER-2017-013.
- 3156 [18] LHCb, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and*
 3157 *13 TeV pp collisions*, Phys. Rev. Lett. **118** (2017), no. 5 052002, arXiv:1612.05140,
 3158 [Erratum: Phys. Rev. Lett.119,no.16,169901(2017)].
- 3159 [19] LHCb collaboration, *LHCb reoptimized detector design and performance: Technical*
 3160 *Design Report*, CERN-LHCC-2003-030. LHCb-TDR-009.
- 3161 [20] M. Needham, *Electron Reconstruction Studies*, LHCb-2001-102, CERN-LHCb-2001-
 3162 102, 2001.
- 3163 [21] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys.
 3164 **C40** (2016), no. 10 100001.
- 3165 [22] PIDCalib, <https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage>, .
- 3166 [23] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and*
 3167 *experience*, J. Phys. Conf. Ser. **331** (2011) 032023.
- 3168 [24] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05**
 3169 (2006) 026, arXiv:hep-ph/0603175.
- 3170 [25] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput.
 3171 Phys. Commun. **178** (2008) 852, arXiv:0710.3820.
- 3172 [26] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth.
 3173 **A462** (2001) 152.
- 3174 [27] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum.
 3175 Meth. **A506** (2003) 250.
- 3176 [28] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE
 3177 Trans. Nucl. Sci. **53** (2006) 270.
- 3178 [29] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, *A Comparative study of the decays*
 3179 *$B \rightarrow (K, K^*)\ell^+\ell^-$ in standard model and supersymmetric theories*, Phys. Rev. **D61**
 3180 (2000) 074024, arXiv:hep-ph/9910221.
- 3181 [30] D. Melikhov and B. Stech, *Weak form-factors for heavy meson decays: An Update*,
 3182 Phys. Rev. **D62** (2000) 014006, arXiv:hep-ph/0001113.
- 3183 [31] LHCb collaboration, T. Blake, U. Egede, P. Owen, and K. Petridis, *Measurement*
 3184 *of the phase difference between the decays $B^+ \rightarrow K^+\mu^+\mu^-$, $B^+ \rightarrow J/\psi K^+$ and*
 3185 *$B^+ \rightarrow \psi(2S)K^+$* , LHCb-ANA-2015-049.

- 3186 [32] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth.
 3187 **A552** (2005) 566, arXiv:physics/0503191.
- 3188 [33] LHCb collaboration, N. Skidmore *et al.*, *Measurement of the ratio $R_K = \mathcal{B}(B^+ \rightarrow$*
 3189 $K^+\mu^+\mu^-)/\mathcal{B}(B^+ \rightarrow K^+e^+e^-)$, LHCb-ANA-2012-112.
- 3190 [34] LHCb collaboration, S. Bifani *et al.*, *Measurement of the ratio $R_{K^{*0}} = \mathcal{B}(B^0 \rightarrow$*
 3191 $K^{*0}\mu^+\mu^-)/\mathcal{B}(B^+ \rightarrow K^{*0}e^+e^-)$, LHCb-ANA-2015-116.
- 3192 [35] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and*
 3193 *regression trees*, Wadsworth international group, Belmont, California, USA, 1984.
- 3194 [36] R. E. Schapire and Y. Freund, *A decision-theoretic generalization of on-line learning*
 3195 *and an application to boosting*, J. Comput. Syst. **55** (1997) 119.
- 3196 [37] LHCb collaboration, J. Albrecht *et al.*, *Search for the rare decay $B_q^0 \rightarrow \mu^+\mu^-\mu^+\mu^-$* ,
 3197 LHCb-ANA-2015-040.
- 3198 [38] LHCb, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its*
 3199 *dependence on B meson kinematics*, JHEP **04** (2013) 001, arXiv:1301.5286.
- 3200 [39] M. Bordone, G. Isidori, and A. Patti, *On the Standard Model predictions for R_K*
 3201 *and R_{K^*}* , Eur. Phys. J. **C76** (2016), no. 8 440, arXiv:1605.07633.
- 3202 [40] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*,
 3203 Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.
- 3204 [41] C. Burr *et al.*, *Prompt charm production in pp collisions at $\sqrt{s} = 13$ TeV in 2015*
 3205 *data*, .
- 3206 [42] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event*
 3207 *errors*, Nucl. Instrum. Meth. **A764** (2014) 150, arXiv:1312.5000.
- 3208 [43] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime*
 3209 *and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986,
 3210 DESY-F31-86-02.
- 3211 [44] LHCb collaboration, R. Aaij *et al.* JHEP **11** (2016) 047. 33 p.
- 3212 [45] P. Robbe, <https://gitlab.cern.ch/lhcb/lhcb/blob/robbep-2016l0calofix-forcarla/l0/l0calo/python/l0calo/configuration.py>, .
- 3214 [46] S. Beranek *et al.*, *Measurement of the electron detection and reconstruction efficiency*,
 3215 <https://twiki.cern.ch/twiki/pub/LHCb/Velo2longElectronTrackingEfficiency/LHCb-EleEff-ANA-v1r2.pdf>.