

CMS Draft Analysis Note

The content of this note is intended for CMS internal use and distribution only

2014/12/30

Head Id:

Archive Id: 265131:272754M

Archive Date: 2014/11/29

Archive Tag: trunk

Angular analysis and differential branching fraction of the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

Mauro Dinardo¹ and Kevin Stenson²

¹ Universita' degli Studi di Milano Bicocca and INFN - Italy

² University of Colorado at Boulder - USA

Abstract

The angular distributions and the differential branching fraction of the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ are studied using an integrated luminosity of $\sim 20.5 \text{ fb}^{-1}$ of data collected with the CMS detector at $\sqrt{s} = 8 \text{ TeV}$. The forward-backward asymmetry of the muons, A_{FB} , the fraction of $K^*(892)$ longitudinal polarization, F_L , and the differential branching fraction, $d\mathcal{B}/dq^2$, are determined as a function of the dimuon invariant mass squared (q^2). The measurements are in good agreement with the Standard Model. In the dimuon invariant mass squared range $1 < q^2 < 6 \text{ (GeV/c}^2)^2$, the results are: $A_{FB} = XXX^{+XXX}_{-XXX} \text{ (stat)} \pm XXX \text{ (syst)}$, $F_L = XXX^{+XXX}_{-XXX} \text{ (stat)} \pm XXX \text{ (syst)}$, and $d\mathcal{B}/dq^2 = XXX^{+XXX}_{-XXX} \text{ (stat)} \pm XXX \text{ (syst)}$.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor:	Mauro Dinardo
PDFTitle:	Angular analysis and differential branching fraction of the decay B0 to K*0 mu+ mu-
PDFSubject:	CMS
PDFKeywords:	CMS, physics

Please also verify that the abstract does not use any user defined symbols

1 Contents

2	1	Introduction	3
3	2	Data and Simulation Samples and Processing	5
4	3	Reconstruction and Event Selection	8
5	3.1	ϕ Particle Rejection	14
6	3.2	Optimization Procedure of the Selection Cuts	14
7	3.3	CP-State Assignment	16
8	4	Analysis Strategy	17
9	4.1	Physics Model	17
10	4.2	Correctly and Wrongly CP-Tagged Signal Events	18
11	4.3	The Differential Branching Fraction	19
12	4.4	The Probability Density Function	19
13	4.5	Feed-Through Distributions and Modeling	22
14	4.6	Fit Sequence	22
15	4.7	A_{FB} Zero Crossing Point	24
16	4.8	Angular Variables Transformations	24
17	5	Studies of Possible Background Peaking in the Signal Region	30
18	6	Data-Simulation Comparison of Kinematic and Dynamic Observables	33
19	7	Efficiency Computation	37
20	7.1	Analytical Description of Correctly Tagged Events	41
21	7.2	Description of Wrongly Tagged Events	41
22	8	Analysis Validation	55
23	8.1	Pileup Comparison Between Data and Simulation	55
24	8.2	Validation with High-Statistics, Pure Signal Simulation	55
25	8.3	Validation with High Statistics Full Simulation	58
26	8.4	Validation with Data-Like Statistics Full Simulation	60
27	8.5	Validation with Data	65
28	9	Results	68
29	9.1	Pseudo-Experiments	72
30	9.2	Comparison with Other Experiments	72
31	10	Systematic Uncertainties	76
32	10.1	Limited Number of Simulated Events	77
33	10.2	Efficiency Shape	79
34	10.3	Simulation Mismodeling	80
35	10.4	Fit Bias	80
36	10.5	Residual Correction	80
37	10.6	Wrong CP-State Assignment	80
38	10.7	$S - P$ Wave Interference	81
39	10.8	Background Mismodeling	82
40	10.9	MC Derived p.d.f. Components	82
41	10.10	Feed-Through Background	83
42	10.11	Angular Resolution	83
43	10.12	Special q^2 bin: $1 < q^2 < 6 (\text{GeV}/c^2)^2$	86

44	11	Discussion and Conclusion	87
45	A	Truth Matching	90
46	B	Results of the Selection Cut Tuning Procedure	92
47	C	Plots about Data-Simulation Comparison	98
48	D	Results of the Fits to the Simulation	101
49	E	Results of the Fits to the Data	106
50	F	Results of the Analysis Compared to SM Predictions with Lattice QCD	111
51	G	Results of the Pseudo-Experimentes	113

DRAFT

1 Introduction

The process $B^0 \rightarrow K^{*0} \mu^+ \mu^-$, sketched in Fig. 1, is a flavour changing neutral current decay. In the Standard Model (SM) such decays are suppressed, as they can only proceed via loop processes involving electroweak penguin or box diagrams as shown in Fig. 2 [1, 2]. As-yet undiscovered particles could give additional contributions with comparable amplitudes, making the decay a sensitive probe of new phenomena. A number of angular observables in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay can be theoretically predicted with a relatively good control of the relevant form factor uncertainties especially in the low and high dimuon invariant mass regions, i.e. far from the J/ψ and ψ' resonances. These include the forward-backward asymmetry of the muons, A_{FB} , and the fraction of longitudinal polarization of the $K^*(892)$, F_L , as a function of the dimuon invariant mass squared (q^2) [3–7]. These observables have previously been measured by the BaBar, Belle, CDF, and LHCb experiments [8–11]. A precise determination of A_{FB} is of particular interest as, in the $1 < q^2 < 6$ (GeV/c^2)² region, the form factor ratio used in the calculation of the zero-crossing point is almost free of hadronic uncertainties.

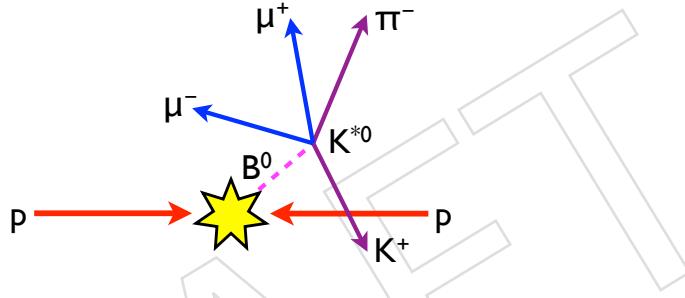


Figure 1: Topological features of the $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$ decay: B^0 detached secondary vertex decaying into a fully charged final state defined by four particle tracks.

This note presents the measurements of A_{FB} , F_L , and the differential branching fraction $d\mathcal{B}/dq^2$ performed with the Compact Muon Solenoid (CMS) experiment at the CERN LHC [12] during 2012 at the center of mass energy of 8 TeV. The amount of analyzed data corresponds to an integrated luminosity of $20.466 \pm 0.522 \text{ fb}^{-1}$ [13].

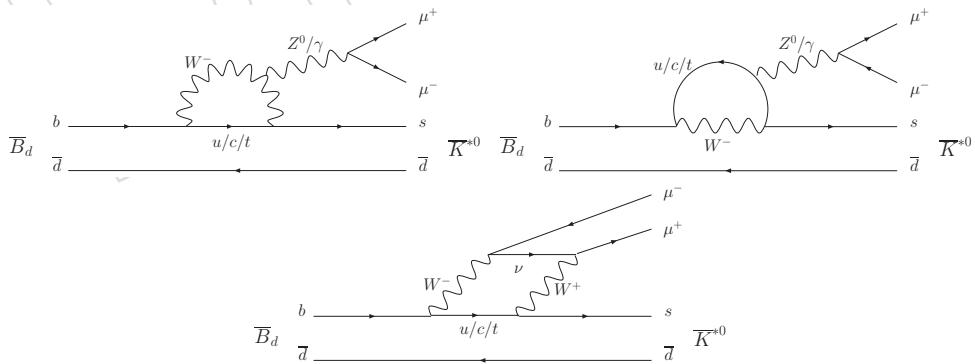


Figure 2: Dominant Standard Model Feynman diagrams for the decay $\bar{B}^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-$.

The $K^*(892)$ is reconstructed through its decay into $K^+\pi^-$ and the B^0 is reconstructed by fitting two muons and two hadron track candidates to a common vertex. A_{FB} and F_L are measured by fitting the event distribution as a function of two angular variables, the angle between the positive charged muon and the B^0 in the dimuon reference frame, and the angle between the kaon

and the B^0 in the $K^*(892)$ reference. The measurements are performed in bins of the dimuon invariant mass squared (q^2) and the trends are compared with SM predictions. Two control channels are used to validate the analysis, $B^0 \rightarrow K^{*0}J/\psi(\mu^+\mu^-)$ and $B^0 \rightarrow K^{*0}\psi'(\mu^+\mu^-)$. In particular the same measurements performed on the signal are applied on these two channels and the results are compared with known SM predictions and previous experimental values. The $B^0 \rightarrow K^{*0}J/\psi(\mu^+\mu^-)$ decay is also used as normalization channel for the measurement of the differential branching fraction.

The analysis measures both CP-conjugate states of the interesting decays, which are distinguished by means of the kaon particle charge. All reported results are derived from a mixture of B^0 and \bar{B}^0 states.

A detailed description of the detector may be found elsewhere [12]. The main detector components used in this analysis are the silicon tracker and the muon detection systems. The silicon tracker measures charged particles within the pseudorapidity range $|\eta| < 2.5$, where $\eta = -\ln[\tan(\theta/2)]$ and θ is the polar angle of the track relative to the counterclockwise beam direction. It consists of 1440 silicon pixel and 15148 silicon strip detector modules and is located in the 3.8 T field of the superconducting solenoid. It provides an impact parameter resolution of about 15 μm and a p_T resolution of about 1.5% for particles with transverse momenta up to 100 GeV. Muons are measured in the pseudorapidity range $|\eta| < 2.4$, with detection planes made using three technologies: drift tubes, cathode strip chambers, and resistive plate chambers. Events are recorded with a two-level trigger system. The first level is composed of custom hardware processors and uses information from the calorimeters and muon systems to select the most interesting events. The high-level trigger processor farm further decreases the event rate from about 100 kHz to around 350 Hz before data storage.

97 2 Data and Simulation Samples and Processing

98 The analysis is performed with the data recorded during the 2012 pp run, corresponding to an
 99 integrated luminosity of 20.466 fb^{-1} . Good runs and luminosity sections are selected with the
 100 “MuonPhysics” JSON file https://cms-service-dqm.web.cern.ch/cms-service-dqm/CAF/certification/Collisions12/8TeV/Reprocessing/Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON_MuonPhys.txt applied on the
 101 parked datasets (see Table 1):

- 104 • /DoubleMuParked/Run2012A-22Jan2013-v1/AOD;
- 105 • /MuOniaParked/Run2012B-22Jan2013-v1/AOD;
- 106 • /MuOniaParked/Run2012C-22Jan2013-v1/AOD;
- 107 • /MuOniaParked/Run2012D-22Jan2013-v1/AOD.

108 The luminosity of the datasets is computed with the officially recommended tool from the lu-
 109 minosity group, `pixelLumiCalc.py`, version V04-02-10. The luminosity shown in Table 1
 110 is per dataset after all processing including the JSON.

Table 1: Data and Monte Carlo simulation samples for signal, control channels, and background used in the analysis. The number of events in data refers to the number of decay candidates found and not to the total number of analyzed events. The numbers listed in the branching fraction column are taken from the PDG.

Sample	Events	\mathcal{E}_{filter}	Branching frac.	$\mathcal{L} (\text{fb}^{-1})$
Data				
Run2012A	564 464 (RECO)	N.A.	7.07×10^{-7}	9.24×10^{-3}
Run2012B	5 820 393 (RECO)	N.A.	7.07×10^{-7}	4.811
Run2012C	11 337 360 (RECO)	N.A.	7.07×10^{-7}	7.101
Run2012D	12 414 614 (RECO)	N.A.	7.07×10^{-7}	7.631
Signal and control channels simulation				
$B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$	38 077 566 (GEN)	1.825×10^{-4}	7.07×10^{-7}	5 951.1
$B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$	24 029 958 (GEN)	1.742×10^{-4}	530×10^{-7}	52.5
$B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$	1 902 471 (GEN)	2.096×10^{-4}	31.31×10^{-7}	58.5
Background simulation				
$B^0 \rightarrow \psi(\mu^+\mu^-)X$	50 562 450 (GEN)	1.918×10^{-4}	N.A.	9.67
$B^+ \rightarrow \psi(\mu^+\mu^-)X$	58 574 844 (GEN)	1.921×10^{-4}	N.A.	9.81
$B_s^0 \rightarrow \psi(\mu^+\mu^-)X$	14 859 324 (GEN)	6.409×10^{-4}	N.A.	12.20
$\Lambda_b \rightarrow \psi(\mu^+\mu^-)X$	2 338 904 (GEN)	3.447×10^{-4}	N.A.	6.67
$B_s^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$	4 859 324 (GEN)	6.751×10^{-4}	7.07×10^{-7}	84.75

111 Both signal ($B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$) and control samples ($B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ and
 112 $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$) are recorded with the same double muon trigger which is espe-
 113 cially designed to perform this analysis: `HLT_DoubleMu3p5_LowMass_Displaced`. This trig-
 114 ger covers the dimuon mass range from $1 \text{ GeV}/c^2$ to $4.8 \text{ GeV}/c^2$. The selections are optimized
 115 to maintain the highest possible efficiency given the rate limitations.

116 The most important selections to reduce the trigger rate are:

- 117 • single muon p_T greater than **3.5** GeV/c ;
- 118 • dimuon p_T greater than **6.9** GeV/c ;

- dimuon invariant mass window $1 \leq m(\mu\mu) \leq 4.8 \text{ GeV}/c^2$;
- the significance of the separation between dimuon vertex and beamspot in the transverse plane greater than 3;
- the dimuon χ^2 vertex probability greater than 10%;

where the p_T selection favors b -quark events from lighter quark events, the vertex probability rejects two muons from different vertices such as from a cascade decay of b to c to s or from a b -quark and anti- b -quark, and the vertex separation distinguishes dimuons from b -decays and from prompt muons.

The amount of data recorded by the dimuon trigger is computed with the officially recommended tool from the luminosity group, `pixelLumiCalc.py`, version v04-02-10, with the option `lumibyls` and with the JSON file containing only the runs and lumisections used in this analysis and obtained with the `crab report` command.

The other, less effective, pre-selection cuts applied at trigger level are:

- $|\eta^\mu| < 2.2$: muon pseudorapidity range;
- $\text{DCA}_{xy}^\mu < 2 \text{ cm}$: Distance of Closest Approach between the muon momentum and the beamspot in the transverse plane;
- $\text{DCA}^{\mu\mu} < 0.5 \text{ cm}$: DCA in three dimensions between the two muon momenta;
- $\cos(\alpha_{xy}^{\mu\mu}) > 0.9$: angle in the transverse plane between the dimuon momentum and the separation between the dimuon vertex and the beamspot.

All pre-selection cuts that depend upon a reference point are referred to the beamspot, they are therefore independent from any choice of the primary vertex.

To validate the analysis technique a detailed study based on officially produced Monte Carlo (MC) simulations is performed. The simulation is also used to determine the efficiency, the systematic uncertainties, and for a variety of systematic checks. The simulation samples are shown in Table 1, they are generated with PYTHIA MSEL = 1, using the EvtGen package to perform the decay, and allowing for radiated photons using the PHOTOS package [14]. Particles are then traced through a detailed model of the detector with GEANT4 [15], producing what from here forward is referred to as RECO MC. The samples are reconstructed within the software framework tag CMSSW_5_3. The signal and control channel samples in particular are generated with the same version of the double muon trigger used to record the data and with the run-dependent technique. This technique allows to account for changes in the data taking related to:

A: pileup;

B: muon detector configuration;

C: dimuon longitudinal distance cut configuration change after run 199 648 (see talk: http://dl.dropboxusercontent.com/u/6901242/dzfilter_inefficiency_fix.pdf for more details):

- `hltESPTrajectoryCleanerBySharedHits` ("TrajectoryCleanerESPProduc") had the following settings before (after) run 199 648: `ValidHitBonus = 5 (100); MissingHitPenalty = 20 (0)`;
- `hltDiMuonMerging("SimpleTrackListMerger")` had the following settings before (after) run 199 648: `FoundHitBonus = 5 (100); LostHitPenalty = 20 (0)`;

162 **D:** high-level trigger changes to cope with **A** and **B** changes.

163 The run-dependent feature is thoroughly validated and the results are presented at the PPD
164 meeting: <https://indico.cern.ch/event/280894/>.

165 The actual simulation dataset names are:

- 166 • /BdToKstarMuMu_EtaPtFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_-
167 RD2_START53_V19F-v1/AODSIM
- 168 • /BdToKstarJPsi_EtaPtFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_-
169 RD2_START53_V19F-v1/AODSIM
- 170 • /BdToKstarPsi2S_EtaPtFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_-
171 RD2_START53_V19F-v1/AODSIM
- 172 • /B0ToPsiMuMu_2MuPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_-
173 S10_START53_V7A-v1/AODSIM
- 174 • /BpToPsiMuMu_2MuPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_-
175 S10_START53_V7A-v1/AODSIM
- 176 • /BsToPsiMuMu_2MuPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_-
177 S10_START53_V7A-v1/AODSIM
- 178 • /LambdaBToPsiMuMu_2MuPtEtaFilter_8TeV-pythia6-evtgen/Summer12_-
179 DR53X-PU_S10_START53_V7A-v2/AODSIM
- 180 • /BsToKstarMuMu_EtaPtFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_-
181 RD2_START53_V19F-v2/AODSIM

182 Together with the aforementioned reconstructed simulation samples, other samples are pri-
183 vately generated up to only the pure generation step (GEN), without any filter or reconstruc-
184 tion step applied. Five billion events are generated for signal and control channels only, correspond-
185 ing to an integrated luminosity of 142.6, 1.9, and 32.2 fb^{-1} respectively. These samples are used
186 to compute the efficiency corrections, as detailed in Sec. 7.

187 The part of the analysis that makes the ntuples is run within the software framework tag
188 CMSSW_5_3_12_patch3. The whole analysis code, including C++, MATLAB, and Mathematica
189 codes, configuration files and cmsDriver commands to generate the MC simulation, are pub-
190 lically available at: <https://github.com/dinardo/usercode/tree/master/B0Analysis/B0KstMuMu>, moreover all simulation samples used in this analysis are pub-
191 lished on the Data Aggregation System – DAS.
192

193 3 Reconstruction and Event Selection

194 The process $B^0 \rightarrow K^{*0}\mu^+\mu^-$, together with the two control channels $B^0 \rightarrow K^{*0}J/\psi$ and $B^0 \rightarrow$
 195 $K^{*0}\psi'$, are reconstructed through their decay into the fully charged final state $K^+\pi^-\mu^+\mu^-$ (with
 196 obvious generalization to the CP-conjugate states, which are implicitly considered throughout
 197 the whole analysis unless explicitly stated). Each event is required to pass the dimuon trigger
 198 described in Sec. 2. B^0 candidates are formed by combining a candidate dimuon vertex with a
 199 candidate $K^*(892)$ meson.

200 Dimuon candidates are required to pass the trigger described in Sec. 2. Dimuon vertices are re-
 201 constructed from oppositely charged muon tracks which are selected accordingly to the soft
 202 muon selections officially recommended by the muon group and here reported for comple-
 203 ness:

- 204 • muon type: TMOneStationTight;
- 205 • tracker layers with measurements > 5 ;
- 206 • pixel layers with measurement > 1 ;
- 207 • normalized muon track $\chi^2 < 1.8$;
- 208 • distance of closest approach between the muon momentum and the primary vertex
 209 in the transverse plane < 3 ;
- 210 • distance of closest approach between the muon momentum and the primary vertex
 211 along the beam line < 30 .

212 The pre-selection cuts applied in the analysis for muon tracks and dimuon vertices have the
 213 same values as those applied at trigger level. The vertex-constrained dimuon fit allows to
 214 compute more accurate momenta of the daughter particles, which are then used to calculate all
 215 other derived quantities, i.e. invariant mass and pseudorapidity.

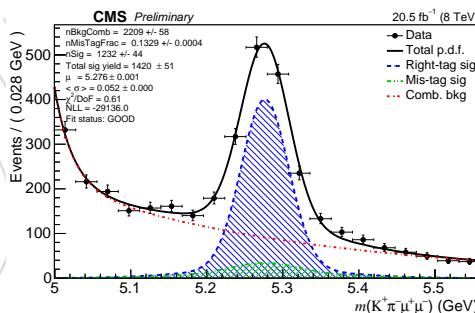


Figure 3: B^0 invariant mass from data, computed from the whole q^2 spectrum excluding the J/ψ and ψ' ranges as described in the text. Just to guide the eye the plot is fitted with a double Gaussian with unique mean to measure the signal yield (1232 ± 44 events) and with two Gaussians and a double exponential to distinguish the background.

216 Candidate $K^*(892)$ mesons are reconstructed through their decay into a kaon and a pion with
 217 opposite charge in both final states, $K^+\pi^-$ and $K^-\pi^+$. Both hadron mass hypothesis are com-
 218 puted and considered throughout the selection procedure. The hadron tracks are vetoed from
 219 being muons by rejecting all tracks which are TrackerMuonArbitrated.

220 The hadron tracks are required to pass the following selections:

- 221 • $p_T^h > 0.8$ GeV/ c : hadron transverse momentum;

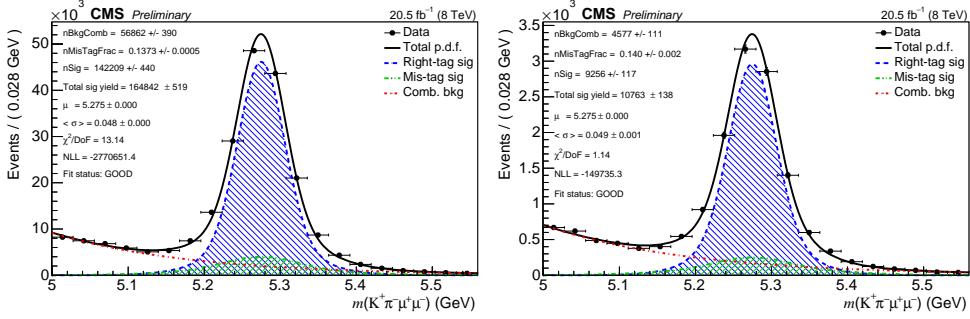


Figure 4: B^0 invariant mass for both control channels, $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ (left) and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ (right), from data. Just to guide the eye the plot is fitted with a double Gaussian with unique mean and an exponential to describe the signal and the background respectively.

- $DCA_{xy}^h/\sigma > 2$: DCA_{xy}^h is the Distance of Closest Approach between the hadron momentum and the beamspot in the transverse plane and σ is its uncertainty;
- $|m_{K^\pm\pi^\mp} - m_{K^{*0}\text{PDG}}| < 90$ MeV/ c^2 : mass window around the $K^*(892)$ PDG mass, corresponding to about 1.5 times its natural width, which is much larger than the experimental resolution of ~ 5 MeV/ c^2 , for at least one of the two hadron mass hypothesis;
- $m_{KK} > 1.035$ GeV/ c^2 : the invariant mass of the two hadron tracks with kaon mass assigned is used to reject $B^0/B_s \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ events, as detailed later in the text in Sec. 3.1.

The vertex-constrained dihadron fit allows to compute more accurate momenta of the daughter particles, which are then used to calculate all other derived quantities, i.e. invariant mass and pseudorapidity.

As a cross-check of the purity of the hadron and muon samples, invariant mass plots of opposite charged hadron and muon tracks are presented in Fig. 5. No evident peak is present which indicates that no muon, or maybe a negligible fraction, is misidentified as a hadron and vice versa.

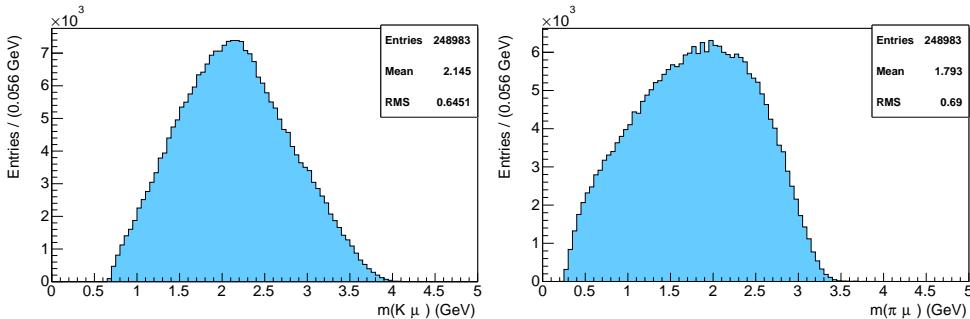


Figure 5: $K\mu$ (left) and $\pi\mu$ (right) invariant mass plots used to cross-check of the purity of the hadron and muon samples.

B^0 are then reconstructed by fitting the two muons and the two hadron track candidates to a common vertex. Likewise for the reconstruction of $K^*(892)$ meson candidates, both B^0 and \bar{B}^0 mass hypothesis are computed and considered throughout the selection procedure.

240 The B^0 candidates are required to pass the following selections:

- 241 • $p_T^{B^0} > 8$ GeV/ c : B^0 transverse momentum;
- 242 • $|\eta^{B^0}| < 2.2$: B^0 pseudorapidity range;
- 243 • $|m_{B^0}(m_{\bar{B}^0}) - m_{B^0\text{PDG}}| < 280$ MeV/ c^2 : at least one of the two B^0 mass hypothesis has
244 to lie within a window around the B^0 PDG mass;
- 245 • $\text{CL}(B^0\text{-vtx}) > 10\%$: B^0 vertex confidence level;
- 246 • $L_{xy}(B^0\text{-vtx})/\sigma > 12$: $L_{xy}(B^0\text{-vtx})$ is the transverse separation between the B^0 ver-
247 tex and the beamspot and σ is its uncertainty;
- 248 • $\cos(\alpha_{xy}^{B^0}) > 0.9994$: $\alpha_{xy}^{B^0}$ is the angle in the transverse plane between the B^0 momentum
249 and the separation between the B^0 vertex and the beamspot.

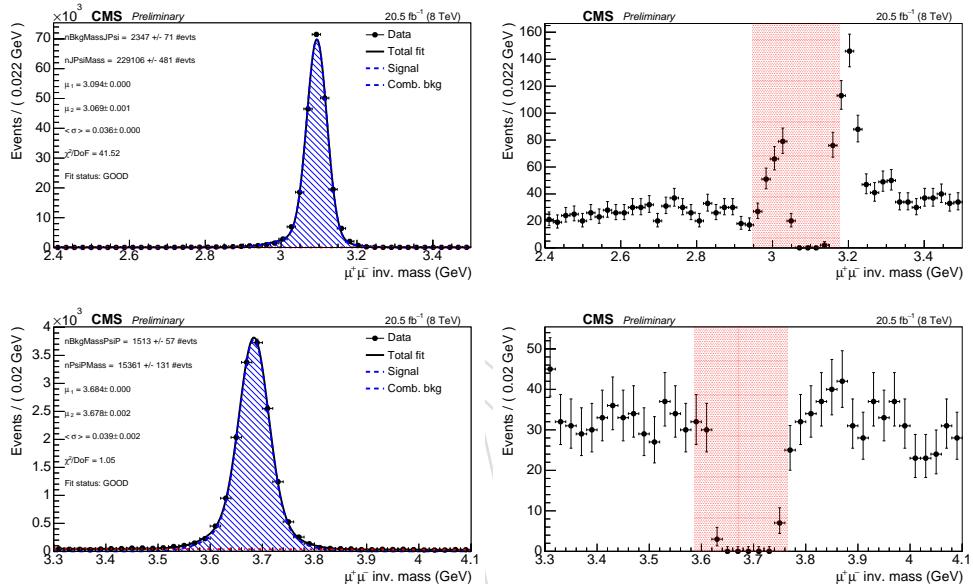


Figure 6: J/ψ (top two plots) and ψ' (bottom two plots) dimuon invariant mass regions from data shown for two different cases: before (first two plots on the left) and after (second two plots on the right) selecting the signal sample. The fits are performed with a double Gaussian with independent means and a first order polynomial to describe the signal and background respectively. The red shaded regions in the plots on the right correspond to the J/ψ and ψ' rejection bins respectively.

250 The selection procedure just described allows to have on average ~ 1.05 candidates per event.
251 Finally the choice of the best candidate is based on the best vertex fit probability, and the CP-
252 state is assigned accordingly to the criteria detailed later in the text in Sec. 3.3.

253 All selection cuts that depend upon a reference point are referred to the beamspot, they are
254 therefore independent from any choice of the primary vertex. The only exceptions are the
255 distances of closest approach computed for the muon selection, but given their loose values
256 they are likewise not influenced by the choice of the primary vertex.

257 The B^0 invariant mass plots for signal and control channels from data are shown in Fig.s 3 and 4
258 (for completeness in Fig. 7 are compared the B^0 invariant mass plots, for the whole dimuon spec-
259 trum, for correctly and wrongly tagged events from data according to the algorithm described
260 in Sec. 3.3). The invariant mass of the J/ψ and ψ' are shown in Fig. 6, while the dimuon spec-

trum and the invariant mass of the $K^*(892)$ are shown in Fig. 7. In particular, just to have a term of comparison for the distribution of the invariant mass of the $K^*(892)$, a fit with shape parameters determined from simulation is superimposed, the number of signal events within the integration range shown in the plot is 169 517, while the number of background events is 74 725. The small skewness present in the $K^*(892)$ spectrum is intrinsic to the dynamic of the decay and it can be better visualized by means of the Dalitz plot shown in Fig. 8, where the two lobes are typical for a spin-1 vector decay, like the $K^*(892)$ [16]. It's important to stress that no fit to the $K^*(892)$ mass spectrum is used in the analysis.

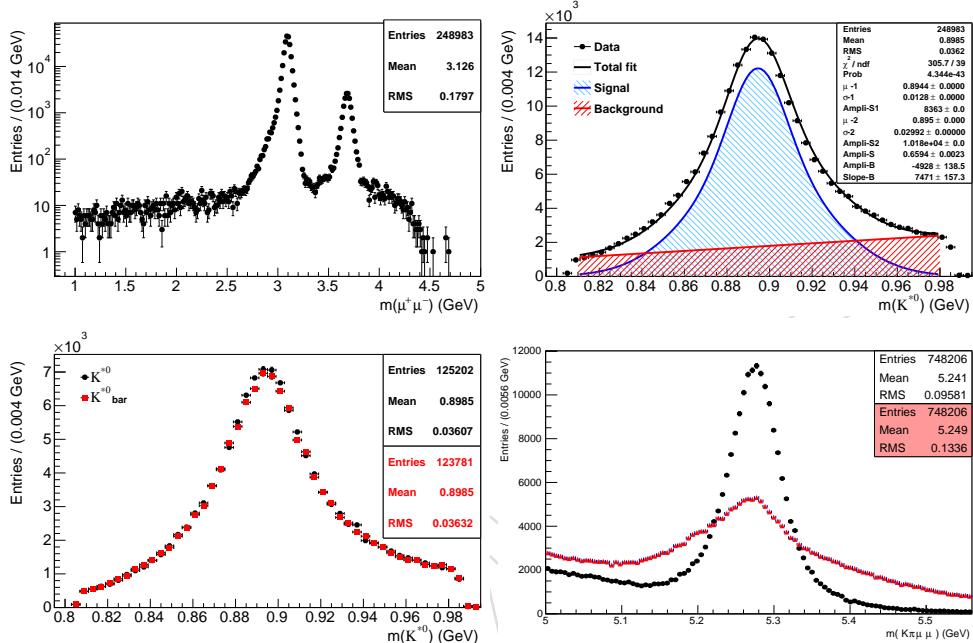


Figure 7: Dimuon (top left) and $K^*(892) + \bar{K}^*(892)$ (top right) invariant mass plots from data after applying the CP-state assignment and all selection cuts but the control channel rejections. Superposition of the $K^*(892)$ and $\bar{K}^*(892)$ invariant mass spectra from data (bottom left). The $K^*(892) + \bar{K}^*(892)$ plot is fitted with a double Gaussian with independent means (for which the shape parameters have been determined from simulation with the control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi(\mu^+\mu^-)$) and a first order polynomial to describe the signal and background respectively. B^0 invariant mass plot for correctly (black) and wrongly (red) tagged events from data (bottom right). Correctly and wrongly events are according to the algorithm described in Sec. 3.3.

The discrimination between signal, $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$, and control channels, $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi(\mu^+\mu^-)$ and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$, is based on the dimuon invariant mass. The signal sample is required to pass the selection:

- $m(\mu\mu) < m_{J/\psi\text{PDG}} - 3\sigma_{m(\mu\mu)}$ or
- $m_{J/\psi\text{PDG}} + 3\sigma_{m(\mu\mu)} < m(\mu\mu) < m_{\psi'\text{PDG}} - 3\sigma_{m(\mu\mu)}$ or
- $m(\mu\mu) > m_{\psi'\text{PDG}} + 3\sigma_{m(\mu\mu)}$;

for the control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi(\mu^+\mu^-)$ the requirement is:

- $|m(\mu\mu) - m_{J/\psi\text{PDG}}| < 3\sigma_{m(\mu\mu)}$.

while for the $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ channel is:

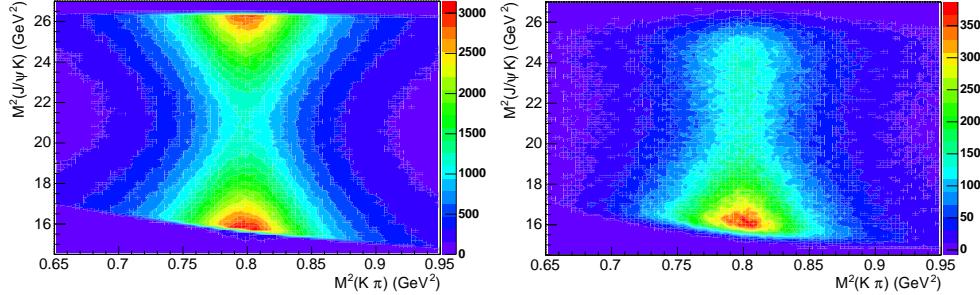


Figure 8: Dalitz plot of the decay $B^0 \rightarrow J/\psi K^* \rightarrow J/\psi K \pi$ from the simulation at pure generator level (GEN-MC left) and after reconstruction and application of all selection cuts and the CP-state assignment (RECO-MC right).

- 278 • $|m(\mu\mu) - m_{B^0\text{PDG}}| < 3\sigma_{m(\mu\mu)}$.

279 The $\sigma_{m(\mu\mu)}$ is the uncertainty on the measurement of the dimuon invariant mass. It is determined on an candidate-by-candidate basis as propagation of the tracks uncertainties, its average value in data is 26 MeV/c 2 over the whole dimuon mass spectrum. This guarantees that 280 the same fraction of events are chosen / rejected as a function of rapidity.

281 Table 2: “ $B^0\&\psi$ ” mass selection cut optimization table. The column S shows the number of signal events from simulation. The column B shows the number of control channel events from simulation. The column N shows the number of events in data. The figure of merit B_{peak}/B_{total} represents the percentage of control channels feed-through with respect to the total background events. The choice of the “ $B^0\&\psi$ ” value is made in correspondence to the figure of merit being $\leq 10\%$. All numbers are determined from a simple counting of the events within the $m(K\pi\mu\mu)$ mass window $|m(K\pi\mu\mu) - m_{B^0\text{PDG}}| < 0.28$ GeV/c 2 . The three different table colors correspond to three different $m(\mu\mu)$ regions, namely below the J/ψ , between the J/ψ and ψ' , and above the ψ' .

Δm [MeV/c 2]	S	B [data]	N [data]	B_{peak} / B_{total}	$S / (S+B)$	S	B [data]	B_{peak} / B_{total} [data]	B_{peak} / B_{total} [MC]	N [data]	B [MC]	B_{peak} / B_{total} [MC]	S	B [MC]	B_{peak} / B_{total} [MC]	N [data]	B [MC]	B_{peak} / B_{total} [MC]	
0	110520	5541	2626	62%	5.0	140075	878	1602	17%	25%	81	80889	71	435	7%	9.0	0	0	
10	-	-	-	-	-	141830	644	684	15%	18%	29%	8.6	80783	45	416	4%	0.2	-	-
20	-	-	-	-	-	141097	445	503	12%	14%	25%	9.2	80603	21	406	2%	0.3	-	-
30	-	-	-	-	-	140578	287	360	11%	19%	9.6	80256	9	400	1%	0.4	-	-	
40	110128	2895	1688	54%	6.3	139910	199	235	10.5%	8%	15%	10.0	80154	2	399	0%	0.3	-	-
50	110017	2288	1453	50%	8.7	139083	135	147	9.2%	5%	11%	10.3	79744	2	393	0%	0.4	-	-
60	109871	1773	1247	47%	7.2	137900	84	929	4%	4%	7%	10.5	79191	2	388	0%	0.4	-	-
70	109768	1302	1090	41%	7.7	136335	55	58	87%	3%	3%	10.7	79223	2	378	0%	0.4	-	-
80	109528	969	974	36%	8.2	134194	41	37	843	2%	2%	10.8	79664	2	377	0%	0.2	-	-
90	109510	713	884	30%	8.6	131439	33	29	809	2%	2%	10.8	75582	2	369	0%	0.1	-	-
100	109042	517	817	24%	8.9	127947	28	20	769	2%	1%	10.7	75391	2	360	0%	0.0	-	-
110	108670	394	764	20%	9.2	123700	23	13	731	1%	1%	10.7	71113	2	352	0%	0.8	-	-
120	108153	306	719	17%	9.4	118867	18	12	701	1%	1%	10.5	68851	1	342	0%	0.6	-	-
130	107529	245	683	15%	9.6	115334	16	11	682	1%	1%	10.3	65753	1	335	0%	0.4	-	-
140	106855	198	641	13%	9.8	107467	12	8	626	1%	1%	10.0	62731	0	326	0%	0.1	-	-
150	105514	156	615	11%	9.9	101384	9	5	597	1%	0%	9.7	59580	0	312	0%	0.9	-	-
160	104948	121	590	9%	10.0	95138	8	4	563	1%	0%	9.3	56207	0	303	0%	0.5	-	-
170	102404	87	571	7%	10.0	88376	6	2	525	0%	0%	9.0	52849	0	290	0%	0.2	-	-
180	100472	76	561	6%	10.0	81617	5	2	496	0%	0%	8.5	49488	0	280	0%	0.9	-	-
190	98507	57	527	5%	10.0	74802	4	1	452	0%	0%	8.2	49002	0	264	0%	0.6	-	-
200	96008	43	506	4%	10.0	67957	3	1	411	0%	0%	7.8	42501	0	257	0%	0.2	-	-
B0&ψ mass cut				2.450 < m($\mu\mu$) < 2.946 GeV/c2				3.177 < m($\mu\mu$) < 3.598 GeV/c2					3.798 < m($\mu\mu$) < 4 GeV/c2						

282 In order to selectively reject the control channel events from the $m(\mu\mu)$ signal region a selection 283 cut acting jointly on the $m(K\pi\mu\mu)$ and $m(\mu\mu)$ masses is applied, which it will be referred with 284 the shorthand name “ $B^0\&\psi$ ” mass cut. Events are rejected if $|(m(K\pi\mu\mu) - m_{B^0\text{PDG}}) - (m(\mu\mu) - m_{\psi\text{PDG}})| < \Delta m$, more precisely if $m(\mu\mu) < m_{J/\psi\text{PDG}}$, then:

- 285 • $|(m(K\pi\mu\mu) - m_{B^0\text{PDG}}) - (m(\mu\mu) - m_{J/\psi\text{PDG}})| < 160$ MeV/c 2 ;
- 286 • $|(m(K\pi\mu\mu) - m_{B^0\text{PDG}}) - (m(\mu\mu) - m_{\psi'\text{PDG}})| < 60$ MeV/c 2 ;
- 287 while if $m_{J/\psi\text{PDG}} < m(\mu\mu) < m_{\psi'\text{PDG}}$, then:
- 288 • $|(m(K\pi\mu\mu) - m_{B^0\text{PDG}}) - (m(\mu\mu) - m_{J/\psi\text{PDG}})| < 60$ MeV/c 2 ;

- 291 • $|m(K\pi\mu\mu) - m_{B^0\text{PDG}}| - (m(\mu\mu) - m_{\psi'\text{PDG}}) | < 60 \text{ MeV}/c^2$;

292 and if $m(\mu\mu) > m_{\psi'\text{PDG}}$, then:

- 293 • $|m(K\pi\mu\mu) - m_{B^0\text{PDG}}| - (m(\mu\mu) - m_{J/\psi\text{PDG}}) | < 60 \text{ MeV}/c^2$;

- 294 • $|m(K\pi\mu\mu) - m_{B^0\text{PDG}}| - (m(\mu\mu) - m_{\psi'\text{PDG}}) | < 30 \text{ MeV}/c^2$.

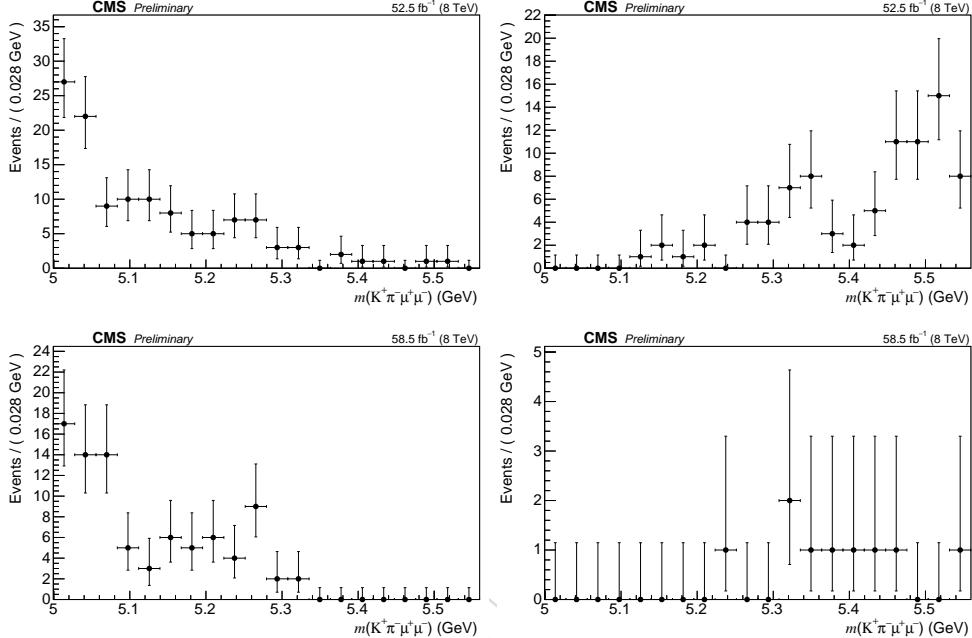


Figure 9: Invariant mass distributions from the feed-through control channel simulation in various $m(\mu\mu)$ regions after tuning the Δm selection cut. The first plot shows the remaining events for the $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ decay in the region $2.450 < m(\mu\mu) < 2.946 \text{ GeV}/c^2$. The second plot shows the remaining events for the $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ decay in the region $3.177 < m(\mu\mu) < 3.586 \text{ GeV}/c^2$. The third plot shows the remaining events for the $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ decay in the region $3.177 < m(\mu\mu) < 3.586 \text{ GeV}/c^2$. The fourth plot shows the remaining events for the $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ decay in the region $3.766 < m(\mu\mu) < 4 \text{ GeV}/c^2$. The numbering goes from left to right and from top to bottom.

295 The Δm parameter is tuned independently for the three different $m(\mu\mu)$ regions, namely below the J/ψ ($2.450 < m(\mu\mu) < 2.946 \text{ GeV}/c^2$), between the J/ψ and ψ' ($3.177 < m(\mu\mu) < 3.586 \text{ GeV}/c^2$), and above the ψ' ($3.766 < m(\mu\mu) < 4 \text{ GeV}/c^2$). The endorsement criteria is to 296 reduce the level of feed-through from the control channels to a negligible level, below the 10% 297 of the total background in data within the $m(K\pi\mu\mu)$ mass window $|m(K\pi\mu\mu) - m_{B^0\text{PDG}}| < 0.28 \text{ GeV}/c^2$ (above the ψ' a tighter cut is applied, $\leq 1\%$, in order to further reduce the residual 298 feed-through that otherwise would have peaked at $\sim 5.42 \text{ GeV}/c^2$). Table 2 shows the value of 299 the figure of merit B_{peak}/B_{total} for different Δm values and for the three $m(\mu\mu)$ regions, where 300 B_{peak} is the number of feed-through events from the control channel simulations, while B_{total} is 301 the number of events in data minus the number of events from signal simulation. All events 302 from simulation are truth-matched (see definition of truth-matching in App. A) and their num- 303 ber is rescaled by the ratio between data and simulation integrated luminosities (the latter 304 is corrected by a factor that takes into account that PYTHIA overestimates the B^0 production 305 cross-section [17]).

306 The tuning of this particular selection cut is performed after the tuning of all the other selection

310 cuts described in Sec. 3.2. The resulting mass distributions of the two feed-through control
 311 channels are separately reported in Fig. 9 for the various $m(\mu\mu)$ regions.

312 3.1 ϕ Particle Rejection

313 A possible background having an invariant mass in the B^0 region might come from B_s or B^0
 314 decaying into $J/\psi(\mu^+\mu^-)\phi(K^+K^-)$. Indeed after applying all selections, but the one on the
 315 invariant mass of the two hadron tracks with kaon mass assigned, there is evidence of the ϕ
 316 particle as shown in Fig. 10. To safely remove such a background a selection cut on the invariant
 317 mass of the two hadron tracks with kaon mass assigned is imposed by $m_{KK} > 1.035$ GeV/ c^2 .
 318 This value allows to entirely veto such a background with tiny impact on the signal selection,
 319 from which only $\sim 1.4\%$ of the events is removed.

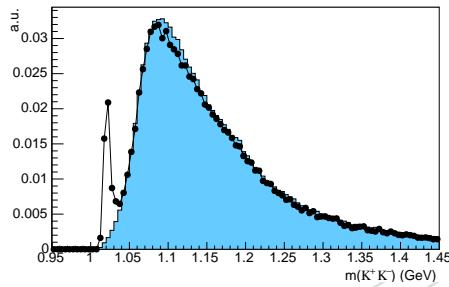


Figure 10: Invariant mass of the two hadron tracks when the kaon mass is assigned to both hadrons. The plot is obtained after applying all selections but the one on the invariant mass of the two hadron tracks with kaon mass assigned. The two superimposed plots are obtained from simulation and data, the former with the control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$, the latter is background subtracted (the plot of the sidebands is subtracted from the plot of the signal region, the definition of signal and sideband regions is detailed later in the text in Sec. 3.2), and no J/ψ nor ψ' rejections are applied, therefore the spectrum is dominated by events from the decay $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$. The first peak on the left corresponds precisely to the ϕ particle ($m(\phi) = 1020$ MeV/ c^2).

320 3.2 Optimization Procedure of the Selection Cuts

321 The optimization procedure is data-driven, based on the maximization the $S/\sqrt{(S+B)}$ figure
 322 of merit. S is the number of truth-matched events (see definition of truth-matching in App. A)
 323 computed on signal simulation whose B^0 invariant mass falls within $\pm 2.5\sigma_{mB^0}$ around the B^0
 324 PDG mass (where σ_{mB^0} is determined from signal simulation from a double Gaussian fit with
 325 unique mean: $\sigma_{mB^0} = \sqrt{f \cdot \sigma_1^2 + (1-f) \cdot \sigma_2^2} \simeq 0.045$ GeV/ c^2). B is the number of events com-
 326 puted with data whose B^0 invariant mass falls within the ranges $[(m_{B^0\text{PDG}} - 5.5\sigma_{mB^0}), (m_{B^0\text{PDG}} -$
 327 $3\sigma_{mB^0})]$ GeV/ c^2 or $[(m_{B^0\text{PDG}} + 3\sigma_{mB^0}), (m_{B^0\text{PDG}} + 5.5\sigma_{mB^0})]$ GeV/ c^2 . The number of events com-
 328 puted with the simulation (S) is rescaled by the ratio between data and simulation integrated
 329 luminosities (the latter is corrected by a factor that takes into account that PYTHIA overesti-
 330 mates the B^0 production cross-section [17]). The dimuon invariant mass region for which the
 331 selection cuts are optimized is $1 < m(\mu\mu) < 2.702$ and $4 < m(\mu\mu) < 4.359$ (GeV/ c^2) 2 . The
 332 reason for this choice is two-fold: this region is free from feed-through from the control chan-
 333 nels, see Fig.s 13, Fig. 14, and 15, and it is the region for which the theoretical predictions are
 334 the more accurate.

335 The pre-selection cuts adopted for the optimization procedure are those used at trigger level

Table 3: Selection cut optimization table. At each step of the optimization one selection cut at a time is scanned searching for a maximum of $S/\sqrt{(S+B)}$, all other selection cuts are applied according to the first column. The second and the third column report the values of the selection cuts and the number of signal events respectively at the maximum of $S/\sqrt{(S+B)}$, the fourth column reports the values of $S/\sqrt{(S+B)}$ at the maximum. In cyan is highlighted the selection cut for which $S/\sqrt{(S+B)}$ acquires the highest value. Such a selection cut is applied at the next step with a value a bit looser with respect to the one at the maximum in order to avoid false maxima.

	Step1				Step2				Step3				
	Start	Optimum	# sig	$S/\sqrt{(S+B)}$	Start	Optimum	# sig	$S/\sqrt{(S+B)}$	Start	Optimum	# sig	$S/\sqrt{(S+B)}$	
B0 vtx CL	0.01	0.29	682.7	3.30	0.01	0.29	435.5	7.95	0.01	0.29	337.7	10.69	
B0 L/ σ BS	0	21	486.2	4.13	0	19	399.6	6.58	0	16	346.3	9.89	
B0 cos(α)	0.997	0.9999	618.7	2.83	0.997	1	371.9	6.82	0.997	0.999	359.5	10.32	
had pT	0.4	1.45	364.8	7.03	0.4	1.05	352.6	9.05	0.8	-	-	-	
K*0 mass	0.15	0.045	691.7	2.39	0.15	0.04	417.8	5.87	0.15	0.045	340.9	9.83	
had DCA/ σ BS	0.8	4.5	268.8	8.11	2.0	-	-	-	2.0	3.1	314	9.41	
	Step4				Step5				Step6				
	Start	Optimum	# sig	$S/\sqrt{(S+B)}$	Start	Optimum	# sig	$S/\sqrt{(S+B)}$	Start	Optimum	# sig	$S/\sqrt{(S+B)}$	Final cuts
B0 vtx CL	0.1	-	-	-	0.1	0.21	348.4	11.45	0.1	0.19	336.3	12.18	0.1
B0 L/ σ BS	0	14	340.7	11.09	0	13	344.3	11.75	0	13	326.8	12.58	12
B0 cos(α)	0.997	0.999	334.2	11.49	0.9994	-	-	-	0.9994	0.9998	329.2	12.28	0.9994
had pT	0.8	1	339.9	10.39	0.8	0.9	352.7	11.24	0.8	0.9	334.8	12.07	0.8
K*0 mass	0.15	0.05	328.3	11.37	0.15	0.05	315.3	12.33	0.09	-	-	-	0.09
had DCA/ σ BS	2.0	3.1	291.9	10.43	2.0	2.3	349.3	11.26	2.0	2.2	340.6	12.08	2.0

and described in Sec. 2, while the selection cuts scanned for the optimization are here reported (the selection cuts that are not here reported but have been aforementioned are also applied, but the “B⁰& ψ ” mass cut):

- CL(B⁰–vtx): B⁰ vertex confidence level;
- L_{xy}(B⁰–vtx)/ σ : L_{xy}(B⁰–vtx) is the transverse separation between the B⁰ vertex and the beamspot and σ is its uncertainty;
- cos($\alpha_{xy}^{B^0}$): $\alpha_{xy}^{B^0}$ is the angle in the transverse plane between the B⁰ momentum and the separation between the B⁰ vertex and the beamspot;
- p_T^h: hadron transverse momentum;
- mass window around the K*(892) PDG mass for at least one of the two hadron mass hypothesis;
- DCA_{xy}^h/ σ : DCA_{xy}^h is the Distance of Closest Approach between the hadron momentum and the beamspot in the transverse plane and σ is its uncertainty.

In Table 3 are shown the steps of the optimization procedure. For each step are reported the starting values of the selection cuts, the maximum of the figure of merit $S/\sqrt{(S+B)}$, the values of the selection cuts, together with the number of signal events at which the maximum of $S/\sqrt{(S+B)}$ is found (all plots showing the trend of S and of $S/\sqrt{(S+B)}$ for each selection cut and for each step are reported in App. B).

The scanning procedure for each selection cut under optimization is divided in the following steps:

1. at each event the selection cuts, together with the trigger requirements, are applied to all candidates (the selection cut under optimization is first applied with its loosest value);
2. the best candidate is selected;
3. the invariant mass is computed and according to its value the counter of the signal (back-

360 ground) events is incremented, as described earlier in the text;

361 4. the selection cut under optimization is tightened and the procedure is repeated from 1.
362 until the scan is completed;

363 5. the figure of merit $S/\sqrt{(S+B)}$ is computed for each value of the selection cut under
364 optimization and its maximum value is detected.

365 The optimization sequence 1. through 5. is repeated for all selection cuts considered for the
366 optimization and until no further significant increment of $S/\sqrt{(S+B)}$ is found.

367 3.3 CP-State Assignment

368 It is of primarily importance for this analysis to make a distinction between the two CP-states
369 of the B^0 meson because the angular observables behave oppositely for each one of the two
370 CP-states. The assignment of the CP-state is based on the distance of the invariant mass of the
371 two hadrons, measured as the area under the Breit-Wigner distribution from the $K^*(892)$ PDG
372 mass (see also Fig. 11 for a sketched example). Both mass hypothesis are computed, i.e. $K^+\pi^-$
373 and $K^-\pi^+$, but only the one closer to the $K^*(892)$ PDG mass is retained, which also directly
374 determine the CP-state of the B^0 meson.

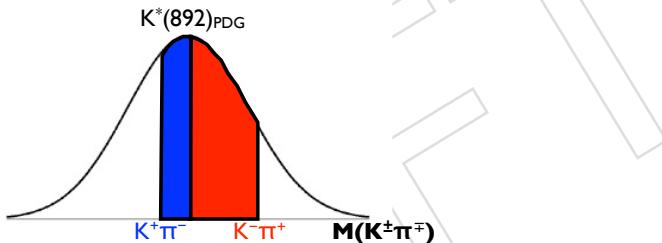


Figure 11: Sketch showing how the choice between $K^+\pi^-$ and $K^-\pi^+$ mass hypothesis is made. The integral from the $K^+\pi^-$ ($K^-\pi^+$) mass to the $K^*(892)$ PDG mass is computed with a Breit-Wigner function. The CP-state of the B^0 is assigned accordingly to the smaller integral.

375 It's important to stress that all pre-selection and selection cuts are chosen in order to not dis-
376 tinguish between CP-states, avoiding, therefore, to introduce possible biases in the analysis.
377 More specifically about the hadron track p_T , from the computation of the decay dynamics of
378 $K^{*0} \rightarrow K\pi$ one can immediately realize that there is no way to assign the CP-state based on the
379 hadron track momentum analysis due to the high q -value of the decay.

380 The CP-mistag fraction is defined as the ratio between the number of mistagged events and
381 the total number of signal events. It is determined from simulation by simply counting the
382 number of correctly tagged and wrongly tagged events, where only truth-matched events (see
383 definition of truth-matching in App. A) are considered, and no fit is involved. The CP-mistag
384 fraction goes from $\sim 12\%$ to $\sim 14\%$ depending on the $m(\mu\mu)$ region considered. More details
385 can be found in Sec. 4.2.

386 4 Analysis Strategy

387 The main goal of this analysis is to measure the forward-backward asymmetry of the muons,
 388 A_{FB} , the fraction of longitudinal polarization of the $K^*(892)$, F_L , and the differential branching
 389 fraction, $d\mathcal{B}/dq^2$, of the decay $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$ as a functions of the dimuon invariant
 390 mass squared (q^2). From the trend of A_{FB} versus q^2 the zero crossing point is also measured.
 391 The q^2 bins are defined in Table 4, they are chosen in such a way to match the measurements
 392 performed by previous experiments.

Table 4: Dimuon invariant mass bins. Both J/ψ and ψ' regions, namely q^2 bins #4 and #6, are used as control channels.

q^2 bin index	mass range (GeV/c^2) ²
0	1 – 2
1	2 – 4.3
2	4.3 – 6
3	6 – 8.68
4	8.68 – 10.09 (J/ψ region)
5	10.09 – 12.86
6	12.86 – 14.18 (ψ' region)
7	14.18 – 16
8	16 – 19

393 4.1 Physics Model

394 The forward-backward asymmetry, A_{FB} , and the fraction of longitudinal polarization, F_L , are
 395 extracted from the fit to the angular observables of the decay. In Fig. 12 are defined the angular
 396 observables for the signal and control channels, and Eq.s 1, 2, and 3 [16] describe the depen-
 397 dence of the decay-rate as a function of such variables. The angle θ_K is defined as the angle
 398 between the kaon and the opposite direction of the B^0 (\bar{B}^0) in the $K^*(892)$ ($\bar{K}^*(892)$) reference
 399 frame, while the angle θ_l is defined as the angle between the positive (negative) charged muon
 400 and the opposite direction of the B^0 (\bar{B}^0) in the dimuon reference frame.

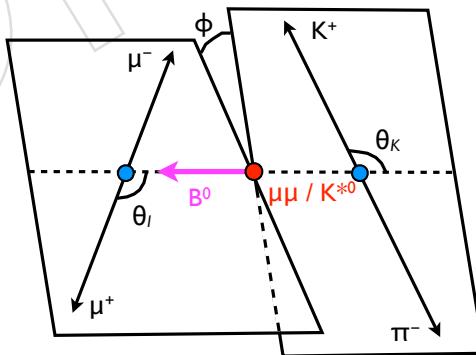


Figure 12: Sketch showing the definition of the angular observables for both signal ($B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$) and control channel decays ($B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$).

401 If the $K - \pi$ are generated from the vector meson $K^*(892)$ their angular distribution in the final
 402 state can be described with a P -wave, while if the $K - \pi$ are not generated from any resonance

their angular distribution in the final state can be described with an *S*-wave. Therefore the processes $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$ and $B^0 \rightarrow K^+\pi^-\mu^+\mu^-$ have different angular distributions and to perform the measurements that are the object of this analysis it is in principle important to disentangle *P*-wave from *S*-wave decays. Indeed the decay rates described in Eq.s 1, 2, and 3 take into account both processes as well as their interference, where the fraction of *S*-wave is described by the F_S term (the fraction of *P*-wave is simply $F_P = 1 - F_S$), while the interference between *S* and *P*-waves is described by the A_S term.

$$\frac{1}{\Gamma} \frac{d^2\Gamma}{d\cos\theta_l dq^2} = \frac{1}{8} \left(3(1 + F_L + F_S(1 - F_L)) + 8A_{FB}(1 - F_S) \cos(\theta_l) + \right. \\ \left. + 3(1 - 3F_L(1 - F_S) - 3F_S) \cos^2(\theta_l) \right), \quad (1)$$

$$\frac{1}{\Gamma} \frac{d^2\Gamma}{d\cos\theta_K dq^2} = \frac{1}{4} \left(2F_S + 4A_S \cos(\theta_K) + \right. \\ \left. + 3(1 - F_S)(1 - F_L - \cos^2(\theta_K) + 3F_L \cos^2(\theta_K)) \right), \quad (2)$$

$$\frac{1}{\Gamma} \frac{d^3\Gamma}{d\cos\theta_K d\cos\theta_l dq^2} = \frac{9}{16} \left(\left(\frac{2}{3}F_S + \frac{4}{3}A_S \cos(\theta_K) \right) (1 - \cos^2(\theta_l)) + \right. \\ \left. + (1 - F_S) \left[2F_L \cos^2(\theta_K) (1 - \cos^2(\theta_l)) + \right. \right. \\ \left. \left. + \frac{1}{2}(1 - F_L)(1 - \cos^2(\theta_K))(1 + \cos^2(\theta_l)) + \right. \right. \\ \left. \left. + \frac{4}{3}A_{FB}(1 - \cos^2(\theta_K)) \cos(\theta_l) \right] \right). \quad (3)$$

Given that the *S*-wave contribution is small with respect to the *P* one, see [16, 18] and Sec. 6, and given that the signal yield per q^2 bin is relatively small, in order to avoid correlation effects difficult to disentangle between $F_S - A_S$ and $F_L - A_{FB}$, the F_S and A_S terms are determined with the control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ with data and applied to all q^2 signal bins. The validation of this approximation is presented in the ***S-P* wave interference** Sec. 10.7.

4.2 Correctly and Wrongly CP-Tagged Signal Events

The signal and control channel are self-tagging decays. This means that in principle one can distinguish whether the mother particle is a B^0 or \bar{B}^0 by simply measuring the charges of daughter hadrons, but this in turns requires the capability of disentangling kaons and pions. Unfortunately the CMS experiment doesn't possess such a capability. To overcome such a deficit an algorithm, based on the analysis of the invariant mass of the two hadrons, is implemented as described in Sec. 3.3. The algorithm has an intrinsic percentage of failure which is referred to as mistag fraction, f_i^M , defined as the ratio # mistagged signal events / # total signal events. Thanks to the simulation one is capable to entirely describe both correctly and wrongly tagged events. Indeed in the p.d.f. in Eq. 5 both signal event categories are explicitly taken into account together with their proper efficiency, but with one common description of the physics, namely S_i^a . The mistag fraction is determined by simply counting the number of correctly and wrongly tagged events, where only truth-matched events (see definition of truth-matching in App. A) are considered, and no fit is involved (see results in Table 5).

Table 5: Number of correctly and wrongly CP-tagged events determined with signal and control channel simulation samples for each q^2 bin in Table 4.

q^2 bin index	Correctly tagged	Wrongly tagged	Mistag fraction (f_i^M)	Error
0	30 518	4 312	0.124	0.002
1	64 438	9 542	0.129	0.001
2	51 149	7 892	0.134	0.001
3	91 065	13 845	0.132	0.001
4	472 326	75 157	0.1373	0.0005
5	119 644	18 256	0.132	0.001
6	30 808	5 013	0.140	0.002
7	69 773	10 623	0.132	0.001
8	72 769	11 529	0.137	0.001

4.3 The Differential Branching Fraction

The differential branching fraction as a function of q^2 , $d\mathcal{B}/dq^2$, is measured by means of the normalization channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$:

$$\frac{d\mathcal{B}(B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-)}{dq_i^2} = \left(\frac{Y_{Si}^R}{\mathcal{E}_{Si}^R} + \frac{Y_{Si}^M}{\mathcal{E}_{Si}^M} \right) \left(\frac{Y_N^R}{\mathcal{E}_N^R} + \frac{Y_N^M}{\mathcal{E}_N^M} \right)^{-1} \frac{\mathcal{B}(B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-))}{dq_i^2}, \quad (4)$$

where the meaning of the symbols is:

- Y_{Si}^R, Y_N^R : yields of the signal and normalization channel respectively for correctly tagged signal events;
- Y_{Si}^M, Y_N^M : yields of the signal and normalization channel respectively for wrongly tagged signal events;
- $\mathcal{E}_{Si}^R, \mathcal{E}_N^R$: efficiency of the signal and normalization channel respectively for correctly tagged signal events;
- $\mathcal{E}_{Si}^M, \mathcal{E}_N^M$: efficiency of the signal and normalization channel respectively for wrongly tagged signal events;
- $\mathcal{B}(B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-))$: branching fraction of the normalization channel as reported in the PDG;

the subscript i runs over the q^2 bins listed in Table 4. It's important to highlight here that the yields Y_N^R and Y_N^M are measured without constraining the dimuon mass to lay in bin #4 (or #6), only the selection cuts described in Sec. 3 are applied. Yields and angular variables are measured in one single fit in the 3D-space: $K\pi\mu\mu$ invariant mass and angular observables.

4.4 The Probability Density Function

The interesting observables are extracted from unbinned fits using an extended maximum likelihood estimator. The p.d.f. has the following mathematical expression:

$$\begin{aligned}
\text{p.d.f.}(m, \theta_K, \theta_l) = & Y_{Si}^R \left(S_i^R(m) \cdot S_i^a(\theta_K, \theta_l) \cdot \mathcal{E}_i^R(\theta_K, \theta_l) + \right. \\
& + \frac{f_i^M}{1 - f_i^M} S_i^M(m) \cdot S_i^a(-\theta_K, -\theta_l) \cdot \mathcal{E}_i^M(\theta_K, \theta_l) \Big) + \\
& + Y_{Bi} B_i^m(m) \cdot B_i^{\theta_K}(\theta_K) \cdot B_i^{\theta_l}(\theta_l),
\end{aligned} \tag{5}$$

450 where:

- 451 • Y_{Si}^R is the yield of correctly tagged signal events;
 - 452 • f_i^M is the mistag fraction (i.e. # mistagged signal events / # total signal events);
 - 453 • Y_{Bi} is the yield of background events coming from the unavoidable fraction of mis-
 - 454 reconstructed B^0 from randomly associated dimuons to hadrons;
 - 455 • $S_i^R(m)$ describes the signal shape for correctly tagged signal events as a function of
 - 456 the $K\pi\mu\mu$ invariant mass (Fig. 16 shows this shape for the simulation);
 - 457 • $S_i^M(m)$ describes the signal shape for wrongly tagged signal events as a function of
 - 458 the $K\pi\mu\mu$ invariant mass (Fig. 17 shows this shape for the simulation, to be noticed
 - 459 the expected wider width with respect to $S_i^R(m)$ precisely due to the wrong K and π
 - 460 mass assignments to the two hadron tracks);
 - 461 • $S_i^a(\theta_K, \theta_l)$ describes the signal in the 2D-space of the angular observables;
 - 462 • $B_i^m(m) \cdot B_i^{\theta_K}(\theta_K) \cdot B_i^{\theta_l}(\theta_l)$ describes the background shape in the 3D-space: $K\pi\mu\mu$
 - 463 invariant mass and angular observables;
 - 464 • $\mathcal{E}_i^R(\theta_K, \theta_l)$ describes the efficiency for correctly tagged signal events in the 2D-space
 - 465 of the angular observables (more details on the efficiency are given in Sec. 7);
 - 466 • $\mathcal{E}_i^M(\theta_K, \theta_l)$ describes the efficiency for wrongly tagged signal events in the 2D-space
 - 467 of the angular observables (more details on the efficiency are given in Sec. 7);
- 468 the subscript i runs over the q^2 bins listed in Table 4. Unless explicitly stated in the text, every
- 469 time it is referred to $\theta_{K/l}$ it is actually meant $\cos(\theta_{K/l})$.

Table 6: Mathematical description of the background parameterization for data as a function of the q^2 bins in Table 4.

q^2 bin index	$B_i^m(m)$	$B_i^{\theta_K}(\theta_K)$	$B_i^{\theta_l}(\theta_l)$
0	1 exp.	2 nd degree	3 rd degree
1	1 exp.	2 nd degree	4 th degree
2	1 exp.	3 rd degree	4 th degree
3	1 exp.	4 th degree	2 nd degree
4	1 exp.	4 th degree	3 rd degree
5	1 exp.	4 th degree	4 th degree
6	1 exp.	4 th degree	3 rd degree
7	2 exp.	4 th degree	2 nd degree
8	1 exp.	2 nd degree	2 nd degree

470 The mathematical expression of the various terms of the p.d.f. in Eq. 5 is:

- 471 • $S_i^R(m)$: double Gaussian with unique mean (namely m_0);

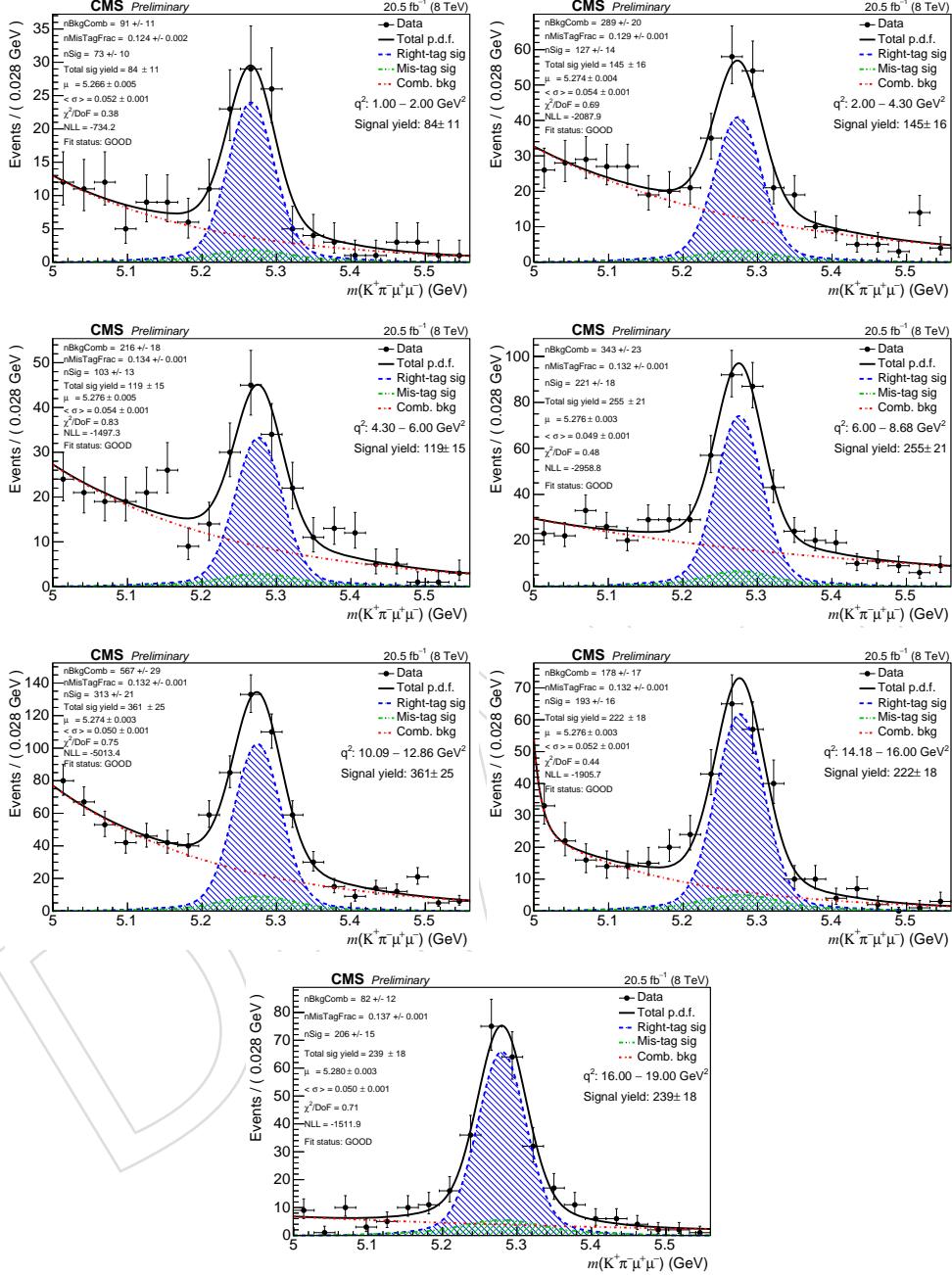


Figure 13: B^0 invariant mass of the signal sample from data for each q^2 bin in Table 4 but #4 and #6. The background is conveniently described in each bin accordingly to Table 6. The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

- $S_i^M(m)$: double Gaussian with unique mean tied to the $S_i^R(m)$ mean (i.e. m_0);
- $S_i^a(\theta_K, \theta_l)$: right-hand side term of Eq. 3;
- $B_i^m(m)$: single or double exponential, depending on the q^2 bin (see Table 6), with translation term tied to the $S_i^R(m)$ mean (i.e. $e^{-\frac{(m-m_0)}{\tau}}$);
- $B_i^{\theta_{K/l}}(\theta_{K/l})$: from second to fourth degree polynomial depending on the q^2 bin (see Table 6);
- $\mathcal{E}_i^R(\theta_K, \theta_l)$: third and fifth degree polynomials as a function of θ_K and θ_l respectively for correctly tagged signal events (see more details in Sec. 7);
- $\mathcal{E}_i^M(\theta_K, \theta_l)$: interpolated binned efficiency for wrongly tagged events (see more details in Sec. 7).

4.5 Feed-Through Distributions and Modeling

As a demonstration that the feed-through rejection cuts, in particular the “ B^0 & ψ ” mass cut described in Sec. 3, act as expected, in Fig.s 14 and 15 are reported the mass and angular distributions from control channel simulation for the adjacent q^2 bins to the J/ψ and ψ' resonances. The distributions don’t show any particular structure. This avoids to model such events with an ad-hoc component in the p.d.f., because they can be easily modeled with the mathematical functions used for the background as described in Sec. 4.4 and shown in Fig. 13.

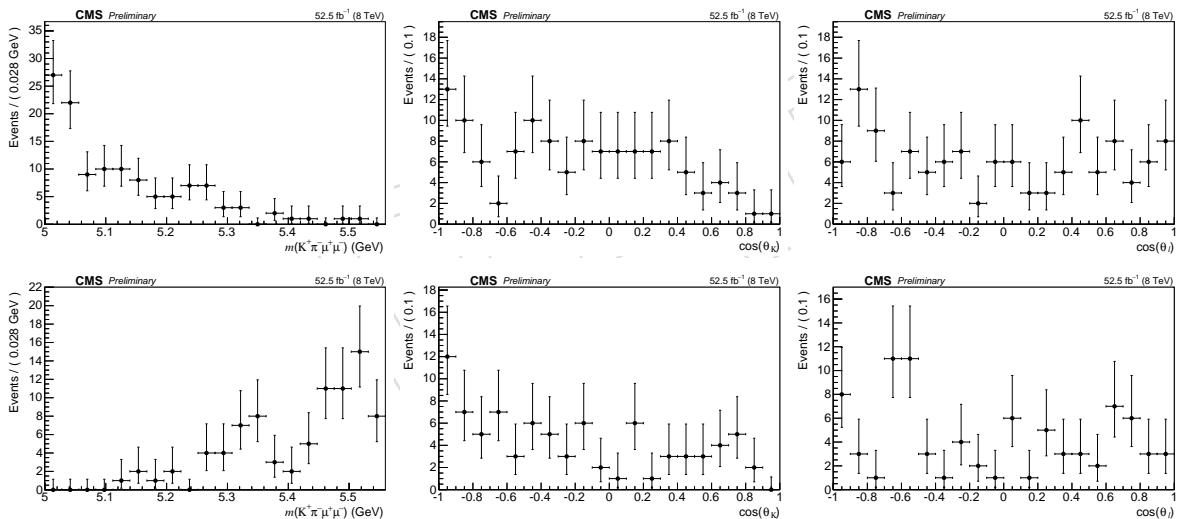


Figure 14: B^0 invariant mass and angular distributions of the $B^0 \rightarrow K^0(K^+\pi^-)J/\psi(\mu^+\mu^-)$ channel from simulation, and after truth-matching, for q^2 bin #3 (top three plots) and #5 (bottom three plots) as from Table 4. For each q^2 bin: the first plot shows the projection on the B^0 invariant mass; the second plot shows the projection on the angular observable θ_K ; the third plot shows the projection on the angular observable θ_l . The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

4.6 Fit Sequence

The p.d.f. described Eq. 5 has several parameters, some of which are determined from simulation. More specifically:

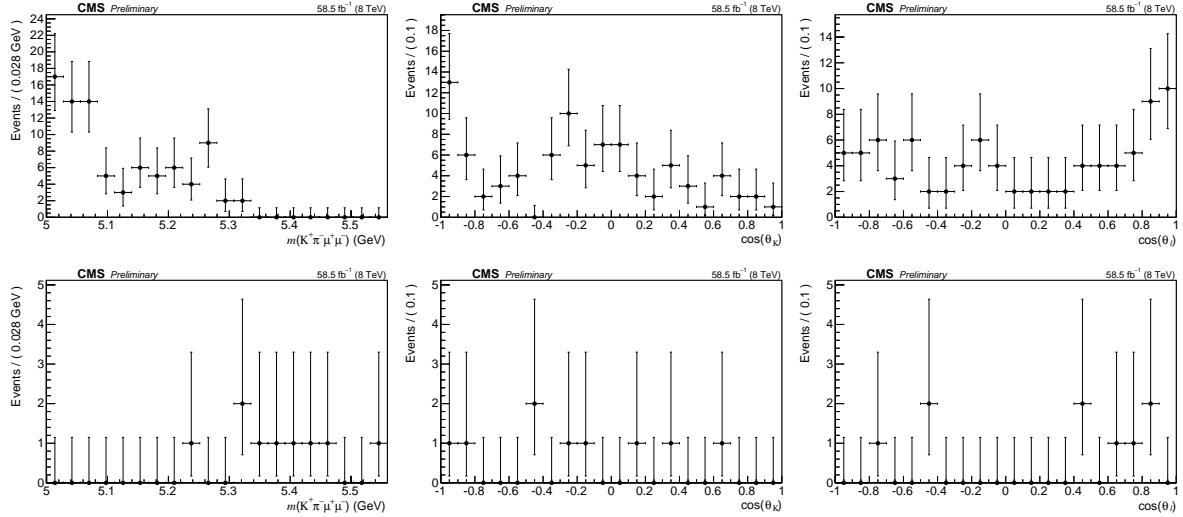


Figure 15: B^0 invariant mass and angular distributions of the $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ channel from simulation, and after truth-matching, for q^2 bin #5 (top three plots) and #7 (bottom three plots) as from Table 4. For each q^2 bin: the first plot shows the projection on the B^0 invariant mass; the second plot shows the projection on the angular observable θ_K ; the third plot shows the projection on the angular observable θ_l . The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

- $S_i^R(m)$ is determined from correctly tagged, truth-matched, events (only the two sigma and the Gaussian fraction) from a fit to the 3D-space: $K\pi\mu\mu$ invariant mass and angular observables (invariant mass plots are presented in Fig. 16);
- $S_i^M(m)$ is determined from wrongly tagged, truth-matched, events (only the two sigma and the Gaussian fraction) from a fit to the 3D-space: $K\pi\mu\mu$ invariant mass and angular observables (invariant mass plots are presented in Fig. 17);
- f_i^M is determined by simply counting the number of correctly and wrongly tagged events as described in Sec. 4.5;
- $\mathcal{E}_i^R(\theta_K, \theta_l)$ is determined from correctly tagged events in the way described in Sec. 7;
- $\mathcal{E}_i^M(\theta_K, \theta_l)$ is determined from wrongly tagged events in the way described in Sec. 7.

The correct error propagation of the variables determined with the simulation is performed in different ways:

- $S_i^R(m)$ and $S_i^M(m)$ parameter errors are propagate by multiplying the p.d.f. in Eq. 5 by Gaussian constraints whose values of mean and sigma are the fit result and error respectively;
- f_i^M errors are propagated by generating new mistag fraction values within their errors, and refitting the data as described in Sec. 10.6;
- $\mathcal{E}_i^R(\theta_K, \theta_l)$ and $\mathcal{E}_i^M(\theta_K, \theta_l)$ errors are propagated by generating new efficiency shapes within their errors and refitting the data as described in Sec. 10.1.

The remaining p.d.f. components are measured directly with data in a two-step sequence:

1. fit to the mass sideband regions (for the definition of sideband regions see Sec. 3.2) as a function of the angular observables to determine $B_i^{\theta_K}(\theta_K)$ and $B_i^{\theta_l}(\theta_l)$;

514 2. fit to the whole 3D-domain space, $K\pi\mu\mu$ invariant mass and angular observables, to de-
 515 termine the remaining p.d.f. components.

516 In summary during the second fit step the free, fixed, and constrained parameters are:

- 517 • the observables of interest, F_L and A_{FB} , are fully free to float;
- 518 • the signal mass shape parameters are free to float with Gaussian constraints;
- 519 • the mistag fraction is fixed;
- 520 • the background angular shape parameters are fixed (from the first fit step);
- 521 • the background mass shape parameters are free to float;
- 522 • the S -wave parameters, F_S and A_S , are fixed from the control channel.

523 For all fits the convergence is verified together with the positive definition of the covariance
 524 matrix, this imply that all plots of this analysis showing fit results should show the tag: Fit
 525 status: GOOD.

526 When fitting the data the starting values are determined with a recursive sequence of steps
 527 which guarantees that the absolute minimum of the likelihood is found. The steps are:

- 528 1. generation of 1000 starting values for F_L and A_{FB} , randomly extracted with a uniform dis-
 529 tribution defined over their physically-allowed domain, and fit the data (no MINOS [19]
 530 errors are computed at this stage);
- 531 2. determination of which of the angular variable starting values from step 1. give the max-
 532 imum of the likelihood and re-fit the data using the MINOS [19] error computation;
- 533 3. re-fit the data with the angular variable starting values obtained as a result of step 2.,
 534 using the MINOS [19] error computation;
- 535 4. re-fit the data with the angular variable starting values obtained as a result of either step
 536 2. or 3. depending on which one gives the highest likelihood, using the MINOS [19] error
 537 computation.

538 **4.7 A_{FB} Zero Crossing Point**

539 The measurement of the zero crossing point of A_{FB} , A_{FB}^z , is performed by simply fitting with a
 540 straight line the A_{FB} values in the first four q^2 bins, i.e. from #0 to #3 as from Table 4.

541 **4.8 Angular Variables Transformations**

542 The p.d.f. in Eq. 5 is physically meaningful, i.e. positive, over the whole angular observables
 543 domain only for values of the angular variables that lay within specific ranges. An approx-
 544 imate determination of such ranges is performed by imposing to the p.d.f. to be positive in
 545 correspondence to particular values of the angular observables, as reported in Table 7.

546 The approximate ranges of validity are therefore:

- 547 • F_L : range of validity [0,1];
- 548 • A_{FB} : range of validity $[-\frac{3}{4}(1 - F_L), +\frac{3}{4}(1 - F_L)]$;
- 549 • F_S : range of validity $[-\frac{3(1 - F_L)}{7 - 3F_L}, +\frac{3(1 - F_L)}{7 - 3F_L}]$;
- 550 • A_S : range of validity $[-\frac{1}{2}(F_S + 3F_L(1 - F_S)), +\frac{1}{2}(F_S + 3F_L(1 - F_S))]$;

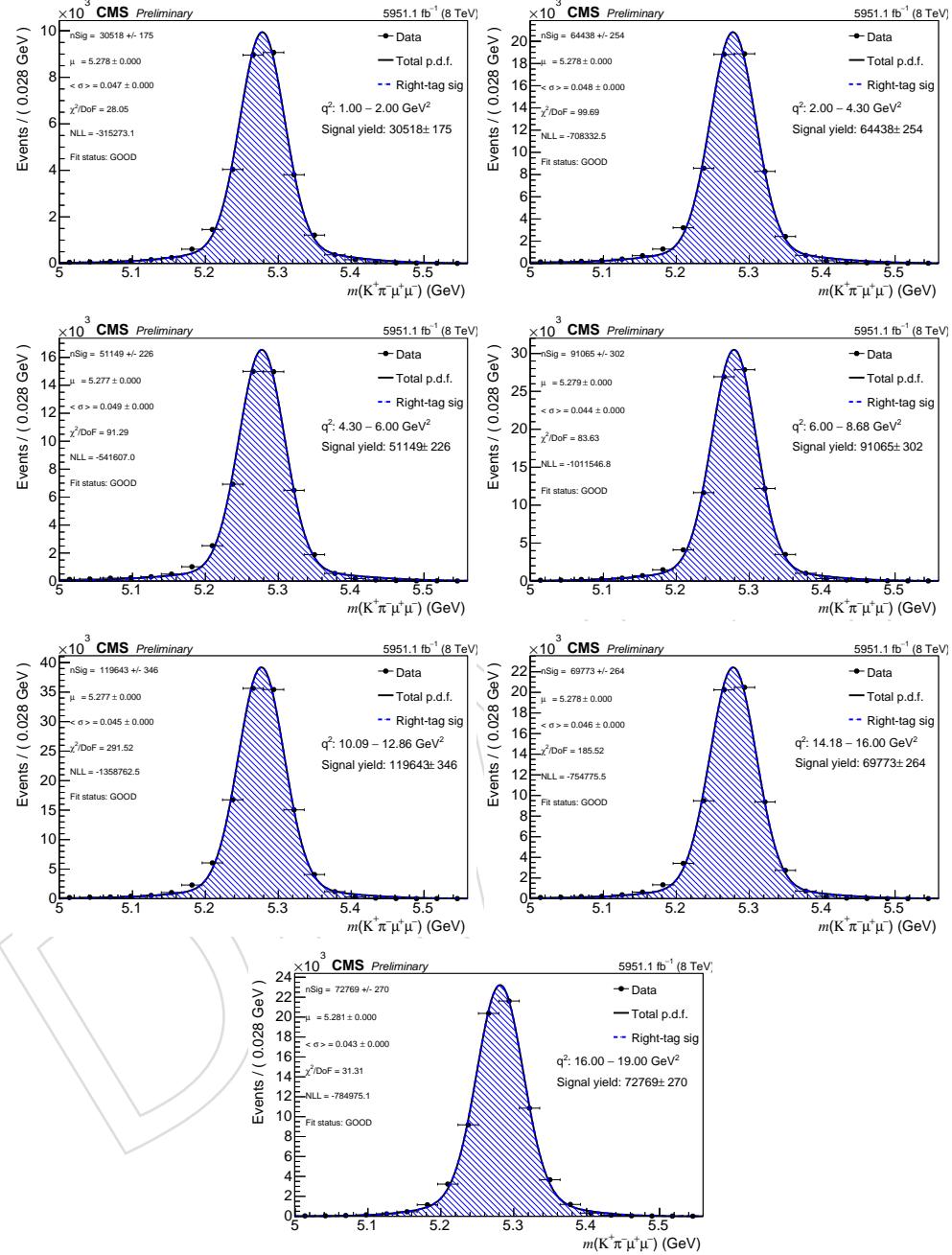


Figure 16: B^0 invariant mass of the signal sample from simulation for each q^2 bin in Table 4 but #4 and #6. The fit results are used to determine the $S_i^R(m)$ component of the p.d.f. The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

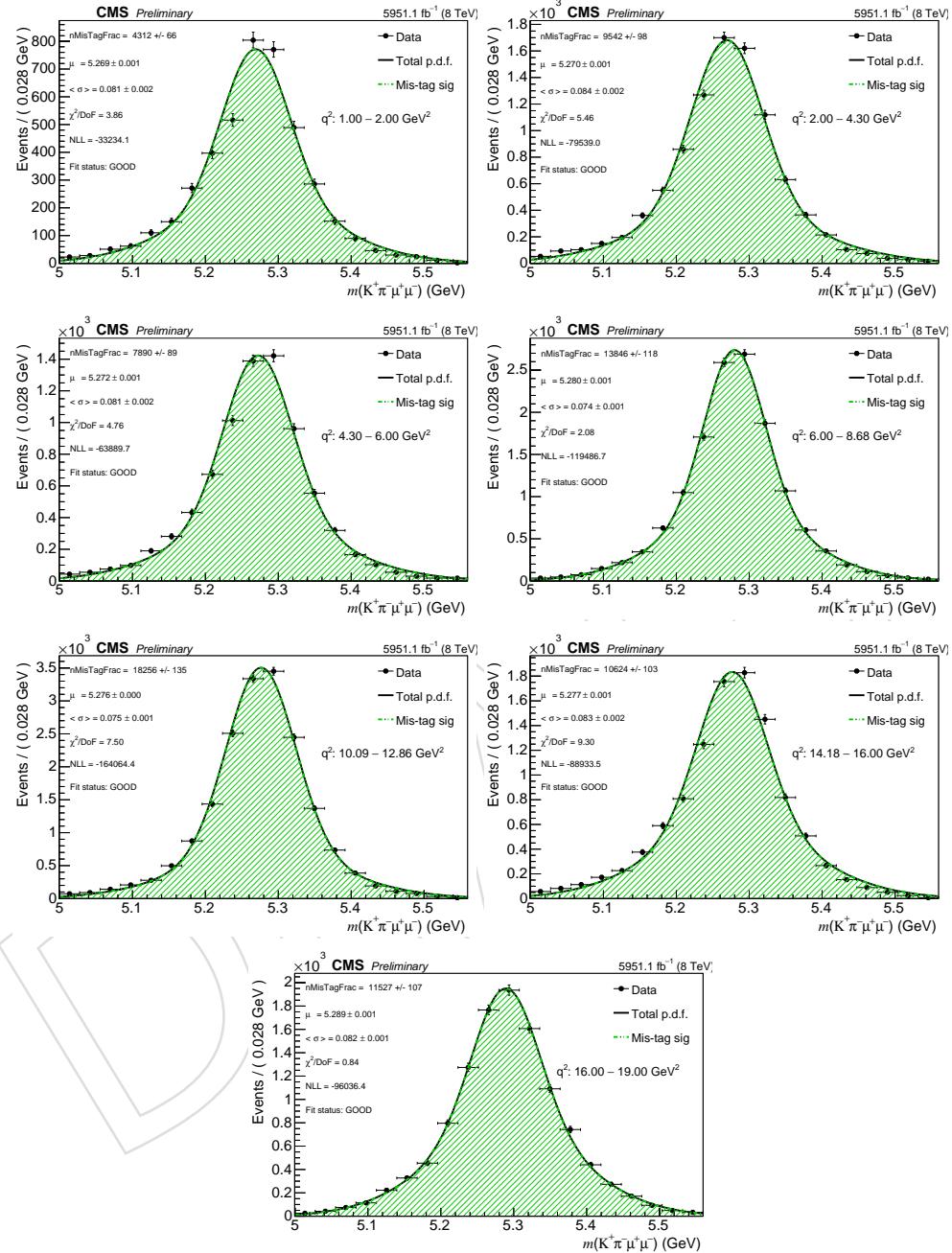


Figure 17: B^0 invariant mass of the signal sample from simulation for each q^2 bin in Table 4 but #4 and #6. The fit results are used to determine the $S_i^M(m)$ component of the p.d.f. The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

Table 7: Search for the physically-allowed domain of the angular variables. The p.d.f. in Eq. 5 is imposed to be positive for specific values of the angular observables.

$\cos(\theta_K)$	$\cos(\theta_I)$	p.d.f.
-1	-1	0
-1	0	$\frac{2}{3}F_S - \frac{4}{3}A_S + 2F_L(1 - F_S) \geq 0$
-1	+1	0
0	-1	$(1 - F_S)(1 - F_L - \frac{4}{3}A_{FB}) \geq 0$
0	0	$\frac{2}{3}F_S + \frac{1}{2}(1 - F_L)(1 - F_S) \geq 0$
0	+1	$(1 - F_S)(1 - F_L + \frac{4}{3}A_{FB}) \geq 0$
+1	-1	0
+1	0	$\frac{2}{3}F_S + \frac{4}{3}A_S + 2F_L(1 - F_S) \geq 0$
+1	+1	0

551 while the actual ranges of validity computed with the help of Mathematica is reported in
 552 Fig. 19. Since the exact expressions are too complicated for being implemented in the anal-
 553 ysis, the approximate ones are used. In particular the $F_L - A_{FB}$ physically-allowed domain is
 554 shown in Fig. 18 together with the SM expected values.

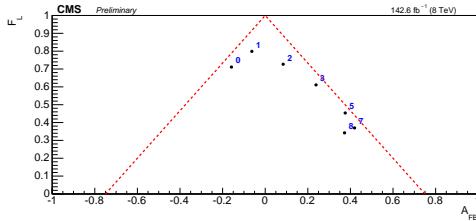


Figure 18: Approximate physically-allowed domain for F_L and A_{FB} angular variables. The numbers next to the dots correspond to the q^2 bin index as from Table 4.

555 Such a dynamic range is implemented by means of a non-linear transformation applied to the
 556 angular variables. The transformation allows MINUIT [19] to use unbounded variables, which
 557 will be referred with an hat “ $\hat{\cdot}$ ”, without the need of implementing specific p.d.f. terms to
 558 keep the physics variables, i.e. the “un-hat” variables, within the physically-allowed domain.

559 The transformations are:

- 560 • $F_L = 1/2 + \arctan(\hat{F}_L)/\pi$;
- 561 • $A_{FB} = 3/4(1 - F_L) \cdot 2 \arctan(\hat{A}_{FB})/\pi$;
- 562 • F_S is not transformed but simply limited within the range [0,1];
- 563 • A_S is not transformed but simply limited within the range [-1,1].

564 Since for F_S and A_S the dynamic domain boundaries are not implemented, the correctness
 565 of the measured values, i.e. whether they lay in the physically-allowed domain, is verified a
 566 posteriori.

567 Errors are propagated in the rigorously correct way, i.e. the finite difference is used instead of
 568 the Taylor expansion of the non-linear transformation, and covariance terms are considered
 569 when needed. In other words given $y = f(x)$, instead of computing $\sigma_y = \partial f(x)/\partial x \cdot \sigma_x$, it is
 570 used the equation $\sigma_y = f(x \pm \sigma_x) \pm y$.

- 571 • $\sigma_{F_L}^+ = 1/2 + \arctan(\hat{F}_L + \sigma_{\hat{F}_L}^+)/\pi - F_L$;

And $P_{\text{exact}}(n_1, n_2, n_3, n_4) = \frac{1}{4} \left[(2n_1 + 1)(2n_2 + 1)(2n_3 + 1)(2n_4 + 1) \right]$

$$\begin{aligned} &+ \frac{1}{2} \left[(2n_1 + 1)(2n_2 + 1)(2n_3 + 1)(-2n_4 - 1) + (2n_1 + 1)(2n_2 + 1)(-2n_3 - 1)(2n_4 + 1) \right. \\ &\quad \left. + (2n_1 + 1)(-2n_2 - 1)(2n_3 + 1)(2n_4 + 1) + (-2n_1 - 1)(2n_2 + 1)(2n_3 + 1)(2n_4 + 1) \right] \\ &+ \frac{1}{2} \left[(2n_1 + 1)(2n_2 + 1)(-2n_3 - 1)(-2n_4 - 1) + (2n_1 + 1)(-2n_2 - 1)(-2n_3 - 1)(2n_4 + 1) \right. \\ &\quad \left. + (2n_1 + 1)(-2n_2 - 1)(2n_3 + 1)(-2n_4 - 1) + (-2n_1 - 1)(-2n_2 - 1)(2n_3 + 1)(-2n_4 - 1) \right] \\ &+ \frac{1}{2} \left[(-2n_1 - 1)(2n_2 + 1)(2n_3 + 1)(-2n_4 - 1) + (-2n_1 - 1)(2n_2 + 1)(-2n_3 - 1)(2n_4 + 1) \right. \\ &\quad \left. + (-2n_1 - 1)(-2n_2 - 1)(2n_3 + 1)(2n_4 + 1) + (-2n_1 - 1)(-2n_2 - 1)(-2n_3 - 1)(2n_4 + 1) \right] \\ &+ \frac{1}{2} \left[(-2n_1 - 1)(-2n_2 - 1)(-2n_3 - 1)(-2n_4 - 1) \right] \end{aligned}$$

Figure 19: Mathematica solution of the inequality: p.d.f. Eq. 5 greater or equal than zero in the whole angular observable domain.

- 572 • $\sigma_{F_L}^- = 1/2 + \arctan(\hat{F}_L + \sigma_{\hat{F}_L}^-)/\pi - F_L;$
- 573 • $\sigma_{A_{FB}}^+ = +((3/4(1 - F_L) \cdot 2 \arctan(\hat{A}_{FB} + \sigma_{\hat{A}_{FB}}^+)/\pi - A_{FB})^2 +$
 $+ (3/4 \cdot 2 \arctan(\hat{A}_{FB})/\pi \cdot \sigma_{F_L}^+)^2 +$
 $+ 2 \cdot (3/4 \cdot 2 \arctan(\hat{A}_{FB})/\pi \cdot \sigma_{F_L}^+)/\sigma_{\hat{F}_L}^+ \cdot (3/4(1 - F_L) \cdot 2 \arctan(\hat{A}_{FB} + \sigma_{\hat{A}_{FB}}^+)/\pi - A_{FB})/\sigma_{\hat{A}_{FB}}^+ \cdot$
 $\cdot \text{cov}(\hat{F}_L, \hat{A}_{FB}))^{1/2};$
- 577 • $\sigma_{A_{FB}}^- = -((3/4(1 - F_L) \cdot 2 \arctan(\hat{A}_{FB} + \sigma_{\hat{A}_{FB}}^-)/\pi - A_{FB})^2 +$
 $+ (3/4 \cdot 2 \arctan(\hat{A}_{FB})/\pi \cdot \sigma_{F_L}^-)^2 +$
 $+ 2 \cdot (3/4 \cdot 2 \arctan(\hat{A}_{FB})/\pi \cdot \sigma_{F_L}^-)/\sigma_{\hat{F}_L}^- \cdot (3/4(1 - F_L) \cdot 2 \arctan(\hat{A}_{FB} + \sigma_{\hat{A}_{FB}}^-)/\pi - A_{FB})/\sigma_{\hat{A}_{FB}}^- \cdot$
 $\cdot \text{cov}(\hat{F}_L, \hat{A}_{FB}))^{1/2};$

581 The fit procedure described in this chapter is applied to both signal and control channels without any distinction. The whole fit procedure is thoroughly validated with simulation studies,
 582 pseudo-experiment studies, and with data by means of the control channels as described in the
 583 following chapters.



5 Studies of Possible Background Peaking in the Signal Region

In this section are presented studies of possible backgrounds peaking in the signal region. As a matter of fact there might be decays whose final state dynamics and kinematics is similar to the one of the signal or control channels, which are object of this analysis. It is of particular danger for this analysis a background that would result in a peak centered right at the B^0 mass. The possible presence of such a background is studied with the five kind of simulations listed in Table 8, from which the decay channels $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ are excluded. For completeness not only B^0 decays are considered, but also B^+ , B_s^0 , and Λ_b decays. Figures 20, 21, 22, 23, and 24 show the mother particle invariant mass plots for three different conditions:

- rejecting the J/ψ and ψ' regions;
- selecting only the J/ψ region;
- selecting only the ψ' region.

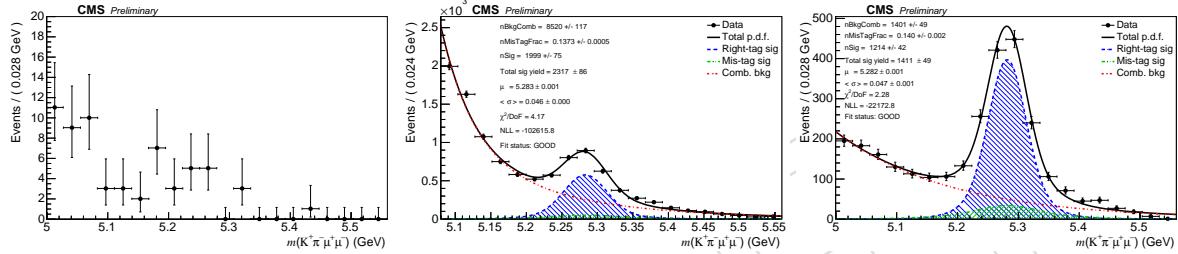


Figure 20: B^0 invariant mass plots from the simulation sample $B^0 \rightarrow \psi(\mu^+\mu^-)X$ from which the decay channels $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ are excluded. The first plot is made by rejecting the J/ψ and ψ' regions (see Sec. 3 for the rejection criteria). The second and third plots are made by selecting only the J/ψ and ψ' region respectively. The plots are fitted with a double Gaussian with unique mean (for which the shape parameters have been determined from simulation) and an exponential to describe the signal and background respectively.

For the simulation samples, B^+ , B_s^0 , and Λ_b , and for the first plot of the B^0 sample there is no evidence for a peak centered at the B^0 mass, therefore they won't be further discussed since the resulting background is already taken into account with the "combinatorial" background component of the p.d.f. (see Sec. 4).

Evidence for a peak centered at the B^0 mass is present only in the second and third plot of Fig. 20. Such peaks are likely from the decay $B^0 \rightarrow K^+\pi^-J/\psi(\mu^+\mu^-)$ where the two hadrons in the final state are not generated from the vector meson $K^*(892)$. The total number of events under the signal peak for the plot made by selecting only the J/ψ region is 3320 after rescaling to the data luminosity, which is $\sim 2.4\%$ of the total number of events of the control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ measured with data. While such a percentage is only indicative because it depends on the various branching fractions used to generate the $B^0 \rightarrow \psi(\mu^+\mu^-)X$ simulation, the direct comparison between data and simulation of the $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ channel, described in Sec. 6, allows for a more correct estimate of the fraction of $B^0 \rightarrow K^+\pi^- J/\psi(\mu^+\mu^-)$ events with respect to $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$, which has been evaluated to $\sim 4.2\%$.

The total number of events under the signal peak of the plot made by selecting only the ψ' region is 2020, which is ~ 4.5 times smaller than the total number of events of the control chan-

Table 8: Simulation samples used to study possible background peaking in the signal region. By writing “ ψ ” in the final state it is actually meant all known ψ states, e.g. J/ψ , ψ' , etc... The rescale of the luminosities to data must take into account the production cross-section difference between PYTHIA and data as reported in the referenced publications. All samples are officially produced with the following characteristics: 2MuPtEtaFilter_8TeV-pythia6-evtgen and Summer12_DR53X-PU_S10_START53_V7A-v1.

Sample	dataset	$\mathcal{L} (\text{fb}^{-1})$	ref. publication
$B^0 \rightarrow \psi(\mu^+\mu^-)X$	B0ToPsiMuMu	9.67	[17]
$B^+ \rightarrow \psi(\mu^+\mu^-)X$	BpToPsiMuMu	9.81	[20]
$B_s^0 \rightarrow \psi(\mu^+\mu^-)X$	BsToPsiMuMu	12.20	[21]
$\Lambda_b \rightarrow \psi(\mu^+\mu^-)X$	LambdaBToPsiMuMu	6.67	[22]
$B_s^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$	BsToKstarMuMu	84.75	-

615 nel $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ measured with data. Such a background might introduce a
 616 bias in the measurement of the interesting observables but since this channel is simply used as
 617 secondary control channel no particular action is taken to mitigate such a bias.

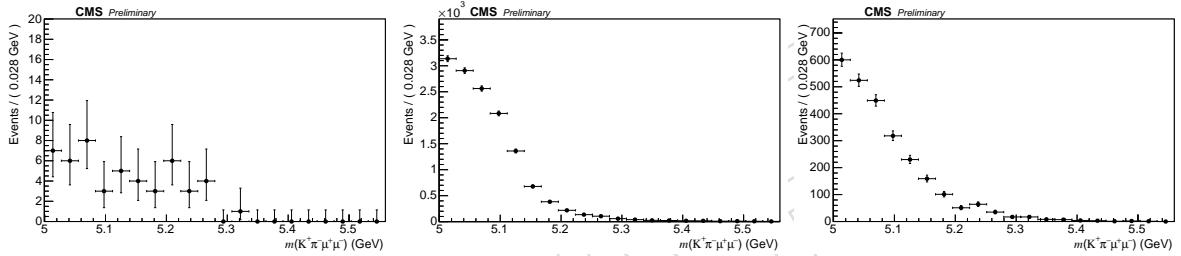


Figure 21: B^+ invariant mass plots from the simulation sample $B^+ \rightarrow \psi(\mu^+\mu^-)X$. The first plot is made by rejecting the J/ψ and ψ' regions (see Sec. 3 for the rejection criteria). The second plot is made by selecting only the J/ψ region. The third plot is made by selecting only the ψ' region. The numbering goes from left to right and from top to bottom.

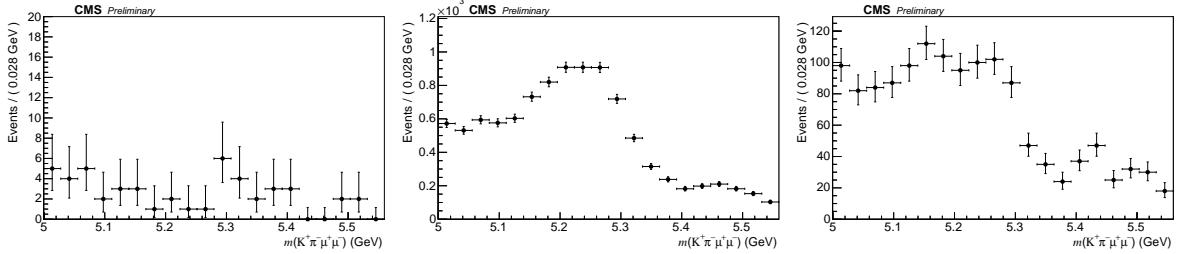


Figure 22: B_s^0 invariant mass plots from the simulation sample $B_s^0 \rightarrow \psi(\mu^+\mu^-)X$. The first plot is made by rejecting the J/ψ and ψ' regions (see Sec. 3 for the rejection criteria). The second plot is made by selecting only the J/ψ region. The third plot is made by selecting only the ψ' region. The numbering goes from left to right and from top to bottom.

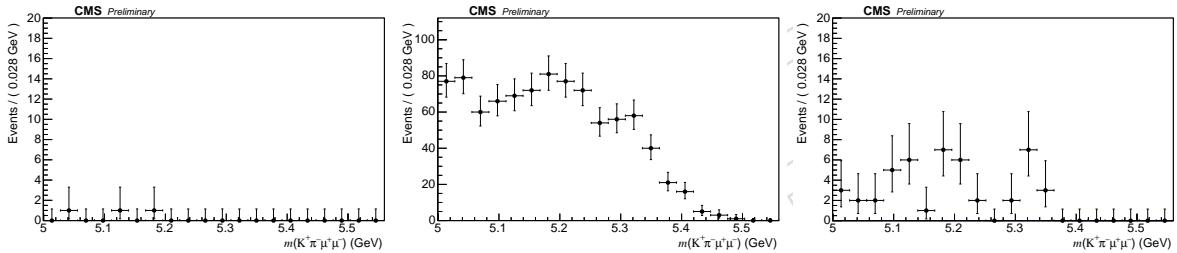


Figure 23: Λ_b invariant mass plots from the simulation sample $\Lambda_b \rightarrow \psi(\mu^+\mu^-)X$. The first plot is made by rejecting the J/ψ and ψ' regions (see Sec. 3 for the rejection criteria). The second plot is made by selecting only the J/ψ region. The third plot is made by selecting only the ψ' region. The numbering goes from left to right and from top to bottom.

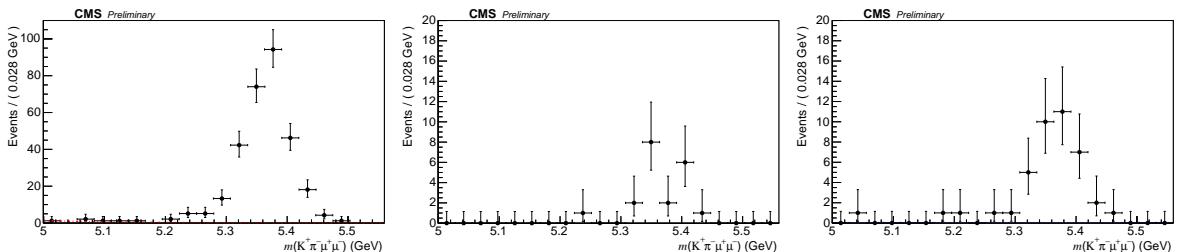


Figure 24: B_s^0 invariant mass plots from the simulation sample $B_s^0 \rightarrow K^{*0}(K^+\pi^-\mu^+\mu^-)$. The first plot is made by rejecting the J/ψ and ψ' regions (see Sec. 3 for the rejection criteria). The second plot is made by selecting only the J/ψ region. The third plot is made by selecting only the ψ' region. The numbering goes from left to right and from top to bottom.

6 Data-Simulation Comparison of Kinematic and Dynamic Observables

The main kinematic and dynamic observables, such as p_T , η , and ϕ , for B^0 and its daughter tracks, are compared between data and simulation. On data only the J/ψ and ψ' dimuon mass regions have been selected by applying the selection criteria described in Sec. 3. The simulation is a combination of both control channels weighted by their branching fractions. All plots from data are background subtracted (the plot of the sidebands is subtracted from the plot of the signal region, for the definition of signal and sideband regions see Sec. 3.2).

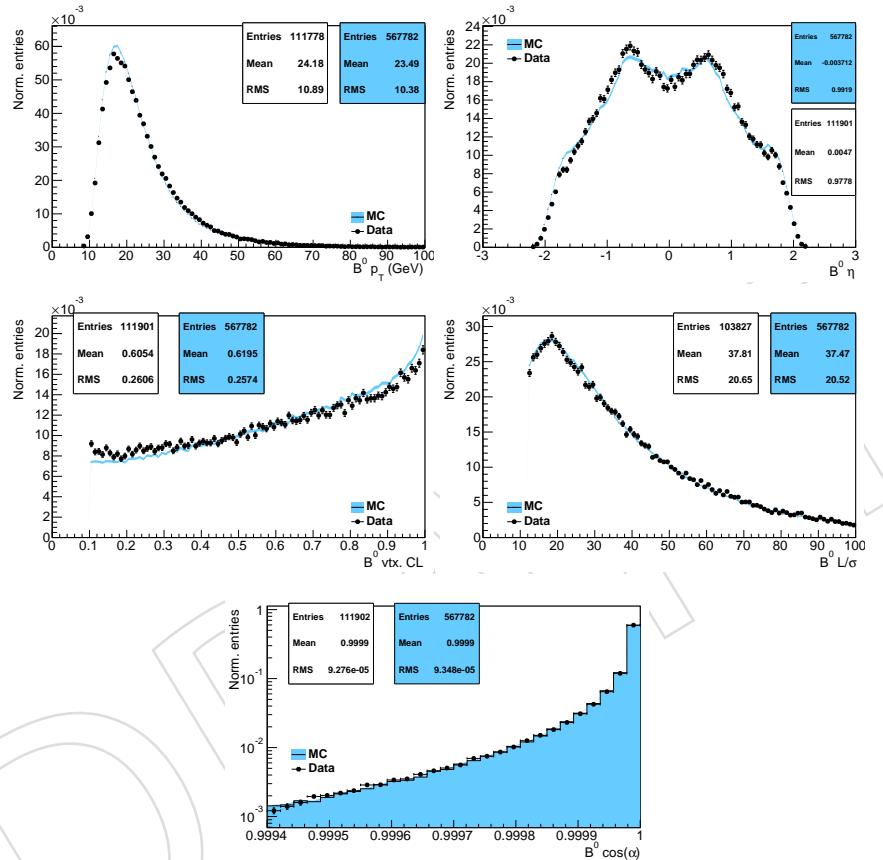


Figure 25: Comparison between data and simulation of the transverse momentum (top left), pseudorapidity (top right), vertex confidence level (middle left), vertex L_{xy}/σ (middle right), and $\cos(\alpha_{xy})$ (bottom) of the B^0 . Both data and simulation have been rescaled in order to have unit area. In data only the J/ψ and ψ' dimuon mass regions have been selected, the simulation, instead, is a combination of both control channels weighted by their branching fractions.

The compared kinematic and dynamic observables are:

- $p_T^{B^0}$: transverse momentum of the B^0 on Fig. 25;
- η^{B^0} : pseudorapidity of the B^0 on Fig. 25;
- $p_T^{h^\pm}$: transverse momentum of the hadrons on Fig. 26;
- η^{h^\pm} : pseudorapidity of the hadrons on Fig. 26;
- ϕ^{h^\pm} : azimuthal angle of the hadrons on Fig.s 89 and 88 in App. C;
- $p_T^{\mu^\pm}$: transverse momentum of the muons on Fig. 27;

- 633 • $\eta^{\mu\pm}$: pseudorapidity of the muons on Fig. 27;
 634 • $\phi^{\mu\pm}$: azimuthal angle of the muons on Fig.s 87 and 86 in App. C;
 635 • θ_K, θ_l : decay angles on Fig. 28.

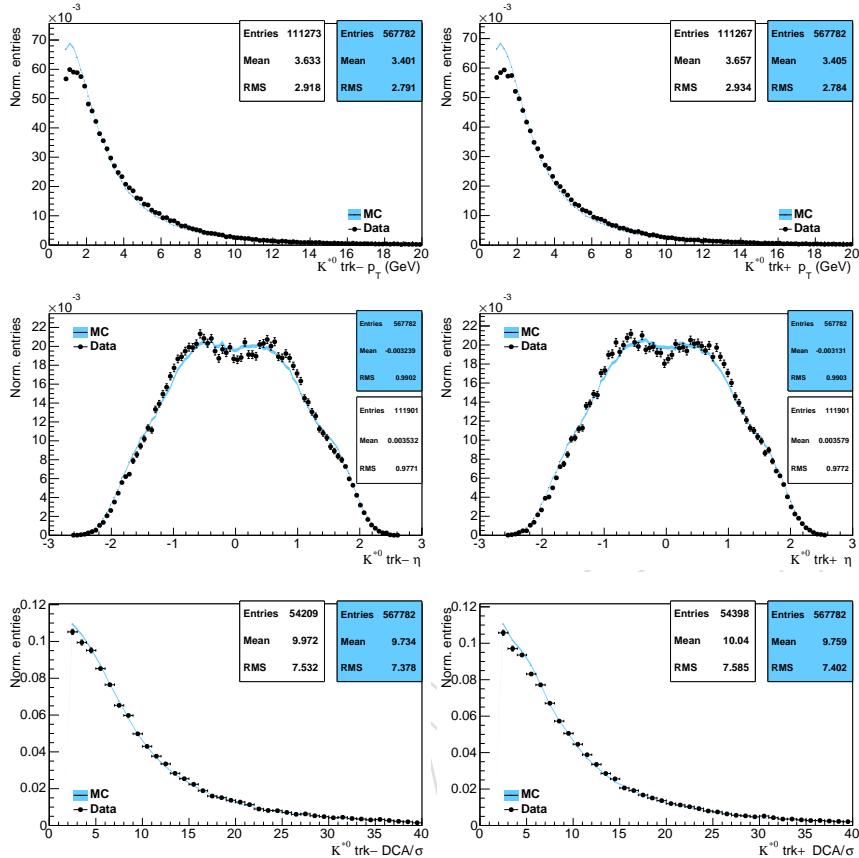


Figure 26: Comparison between data and simulation of the transverse momentum (top two plots), pseudorapidity (middle two plots), and DCA_{xy}/σ (bottom two plots) of hadron-track $^-$ (left) and hadron-track $^+$ (right). Both data and simulation have been rescaled in order to have unit area. In data only the J/ψ and ψ' dimuon mass regions have been selected, the simulation, instead, is a combination of both control channels weighted by their branching fractions.

636 The distributions of the transverse momentum of the B^0 , of the transverse momentum of the
 637 hadron-tracks, and of the angular variable θ_K , show a disagreement between data and sim-
 638 ulation. Of particular interest is the mismatch on the θ_K angular variable because it is used
 639 to measure the fraction of longitudinal polarization of the $K^*(892)$, F_L . Figure 29 shows that
 640 the θ_K distribution for the simulation control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ at gen-
 641 erator level is perfectly symmetric, as one would expect from the underlying physics describing
 642 this decay. Therefore the disagreement can only be ascribed to a known lack of the decays
 643 $B^0 \rightarrow K^+\pi^-J/\psi(\mu^+\mu^-)$ and $B^0 \rightarrow K^+\pi^-\psi'(\mu^+\mu^-)$ in the simulation. The disagreement is es-
 644 timated to be $\sim 4.2\%$, computed as half of the absolute histogram difference from Fig. 28. As
 645 it will be shown in the following, such percentage is an upper limit of the fraction of S -wave,
 646 therefore it is considered as symmetric systematic uncertainty for the measurement of the dif-
 647 ferential branching fraction (called J/ψ X MC in Sec. 10).

648 More specifically if the $K - \pi$ are generated from the vector meson $K^*(892)$ their angular dis-
 649 tribution in the final state can be described with a P -wave, while if the $K - \pi$ are not gener-

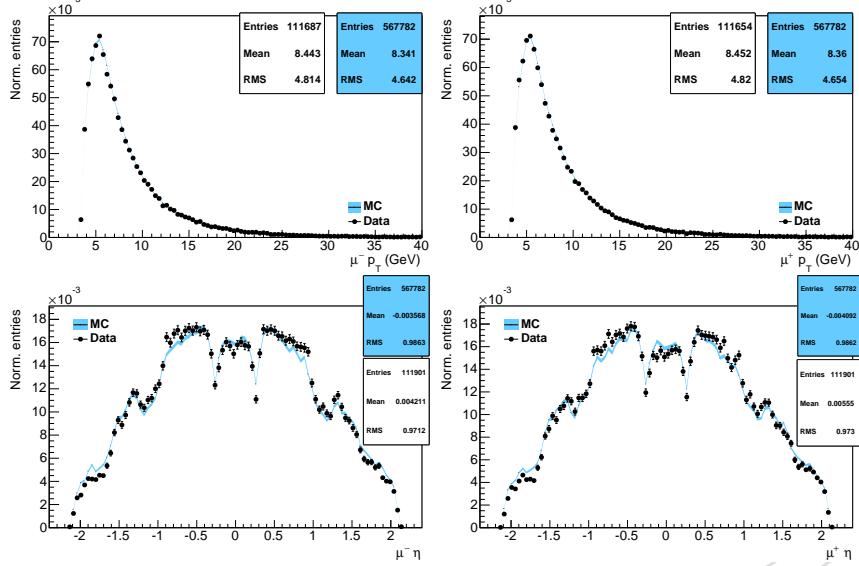


Figure 27: Comparison between data and simulation of the transverse momentum (top two plots) and pseudorapidity (bottom two plots) of μ^- (left) and μ^+ (right). Both data and simulation have been rescaled in order to have unit area. In data only the J/ψ and ψ' dimuon mass regions have been selected, the simulation, instead, is a combination of both control channels weighted by their branching fractions.

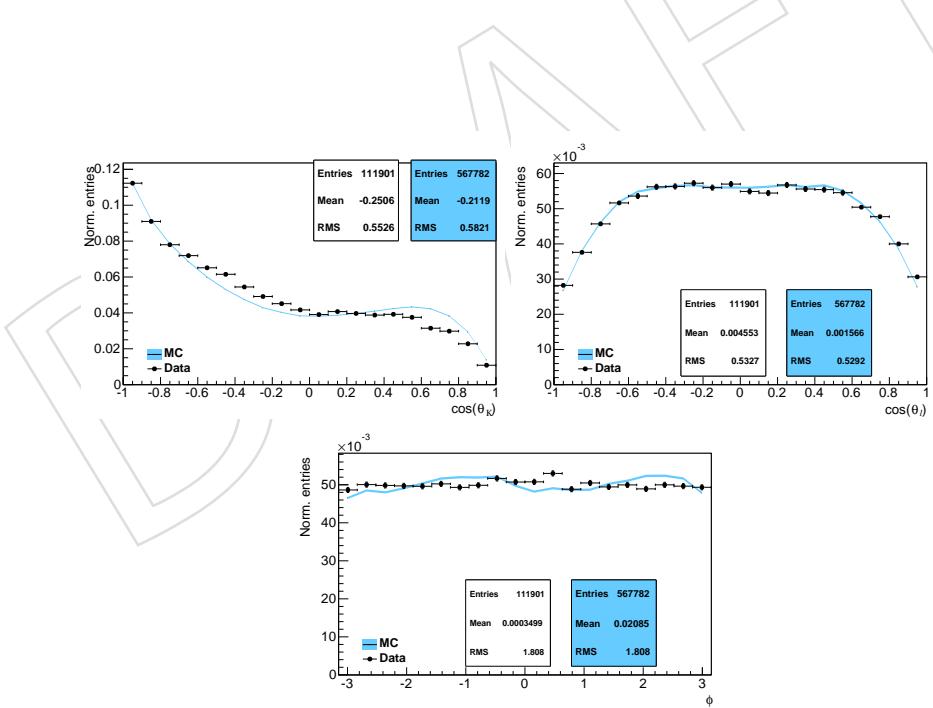


Figure 28: Comparison between data and simulation of the event distribution of the angular variables θ_K (left), θ_l (right), and ϕ (middle). Both data and simulation have been rescaled in order to have unit area. In data only the J/ψ and ψ' dimuon mass regions have been selected, the simulation, instead, is a combination of both control channels weighted by their branching fractions.

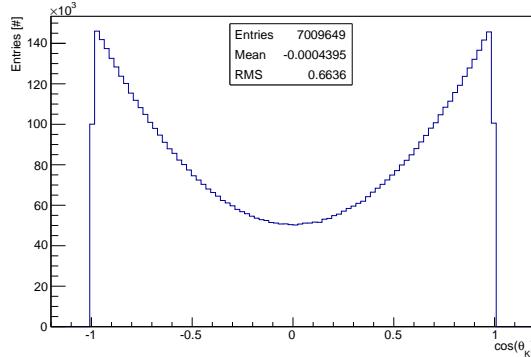


Figure 29: θ_K distribution for the simulation control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ at generator level without any filter applied. The symmetry of the distribution indicates that the simulation is correctly generated according to the physics.

ated from such a resonance, but rather they are directly originated from the B^0 , their angular distribution in the final state can be described with an S -wave. Anticipating here some considerations that will be thoroughly presented in Sec. 8, in Fig.30 are shown plots from data of events distribution as a function of θ_K superimposed to the results of the fit with a p.d.f. that completely neglects the S -wave, and with a p.d.f. that takes into account both S and P -waves together with their interference. It is clear that the latter case better describes the measured distribution of the data and the values of the parameters found by the fit justify the previous statement related to the interpretation of the data-simulation discrepancy: the major contribution to the discrepancy derives from the interference and not from the actual fraction of S -wave, indeed $F_S = 8 \times 10^{-9} \pm 2 \times 10^{-3}$, while $A_S = -0.102 \pm 0.003$.

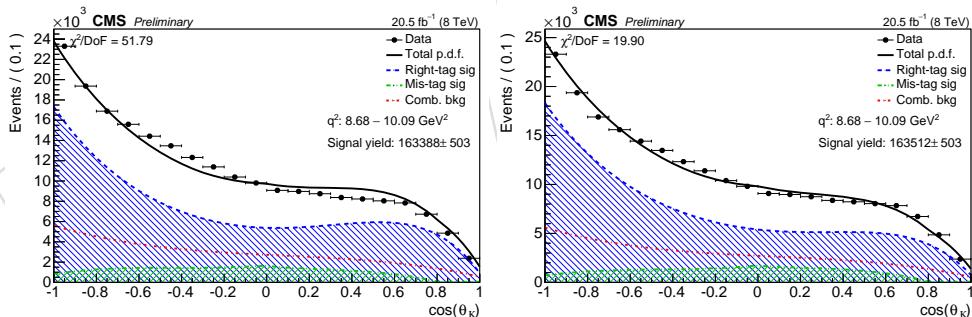


Figure 30: Fit results to the data projected along the variables that are not shown. In the plot on the left the fit is performed without S -wave, while in the plot on the right the fit is performed considering both S and P -wave together with their interference.

The lack of S -wave in the simulation is related to the ignorance of its actual fraction with respect to the P -wave, but even more important it is related to the difficulty of modelling their interference. Indeed it has been shown that the interference plays a greater role than the simple S -wave fraction. The description of the decay rate used in this analysis takes into account both S and P -wave contributions together with their interference [16], as explained in Sec. 4.4.

665 7 Efficiency Computation

666 The efficiency for reconstructing signal events is entirely computed from MC simulation. The
 667 general definition of the efficiency is the ratio between the number of reconstructed events that
 668 pass the selections over the total number of generated events.

669 The actual computation splits the efficiency into two terms as in Eq. 6:

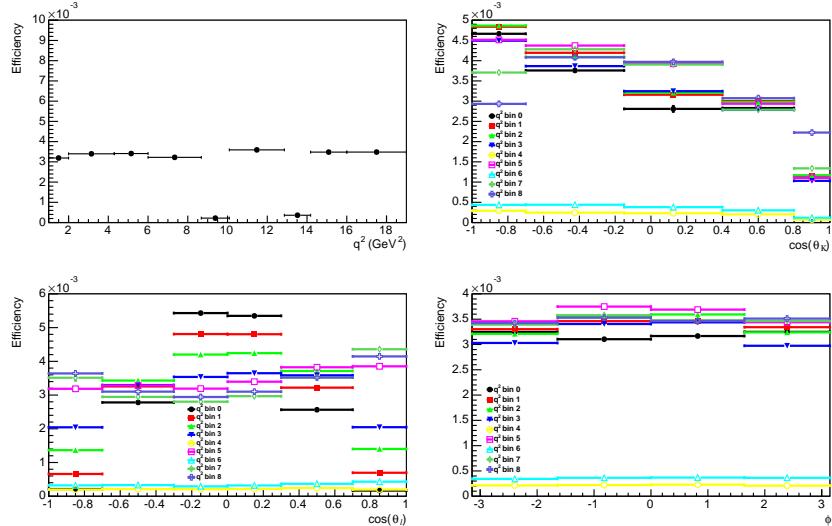


Figure 31: Efficiency computed with correctly tagged signal simulation events. The first plot shows the distribution as a function of the dimuon q^2 , integrated over all angular observables; the second plot shows the distribution as a function of θ_K for all q^2 bins, integrated over θ_l and ϕ ; the third plot shows the distribution as a function of θ_l for all q^2 bins, integrated over θ_K and ϕ ; the fourth plot shows the distribution as a function of ϕ for all q^2 bins, integrated over θ_K and θ_l . The numbering goes from left to right and from top to bottom.

$$\mathcal{E}_i^{R/M} = \mathcal{E}_{ai} \cdot \mathcal{E}_{ri}^{R/M}, \quad (6)$$

670 where:

- 671 • \mathcal{E}_{ai} is the “acceptance” efficiency defined as the ratio between the number of events
 672 that pass the single muon selections ($p_T^\mu > 3.3$ GeV/c and $|\eta^\mu| < 2.3$, this selection
 673 is performed at the GEN level of the simulation) over the total number of events
 674 generated;
- 675 • $\mathcal{E}_{ri}^{R/M}$ is the “reconstruction” efficiency defined as the ratio between the number of
 676 reconstructed events that pass the selections and are truth-matched (see definition
 677 of truth-matching in App. A) over the number of events that pass the single muon
 678 selections ($p_T^\mu > 3.3$ GeV/c and $|\eta^\mu| < 2.3$, this selection is performed at the GEN
 679 level of the simulation); $\mathcal{E}_{ri}^{R/M}$ is computed by means of the RECO-MC simulation
 680 (see Sec. 2) that also contains the GEN information, the RECO information is used to
 681 compute the numerator, while the GEN one is used to compute the denominator;
- 682 • $\mathcal{E}_i^{R/M}$, the superscript R or M refers to correctly (R) or wrongly (M) CP-tagged
 683 events respectively (see Sec. 3.3 for more details on how these categories are
 684 defined).

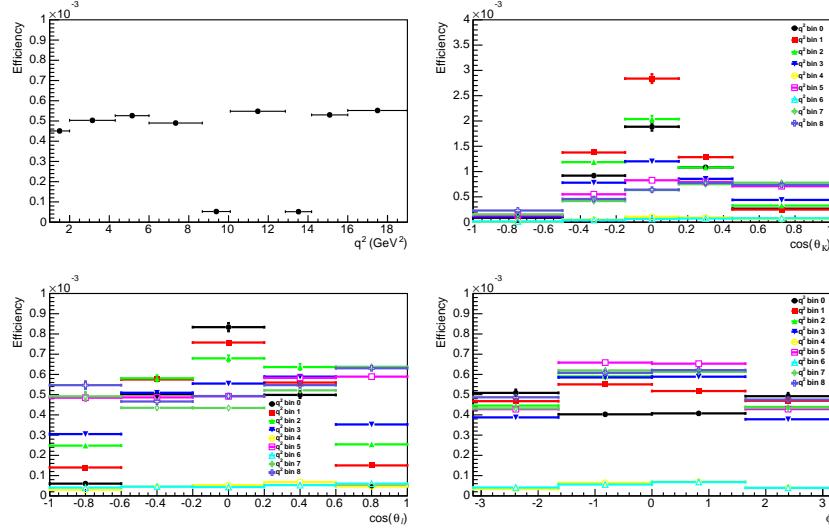


Figure 32: Efficiency computed with wrongly tagged signal simulation events. The first plot shows the distribution as a function of the dimuon q^2 , integrated over all angular observables; the second plot shows the distribution as a function of θ_K for all q^2 bins, integrated over θ_l and ϕ ; the third plot shows the distribution as a function of θ_l for all q^2 bins, integrated over θ_K and ϕ ; the fourth plot shows the distribution as a function of ϕ for all q^2 bins, integrated over θ_K and θ_l . The numbering goes from left to right and from top to bottom.

685 The subscript i runs over the q^2 bins listed in Table 4. The reason for factorizing the efficiency
 686 into two terms is to allow for the possibility of applying weights to both numerator and de-
 687 nominator of the $\mathcal{E}_{ri}^{R/M}$ term (e.g. in case of pileup reweighting, this was indeed the case for the
 688 analysis of 2011 data, while it is no longer the case for the analysis of 2012 data).

689 Kinematic and dynamic constraints are applied to all B^0 for both numerator and denominator
 690 of \mathcal{E}_{ai} and $\mathcal{E}_{ri}^{R/M}$. The constraints are described in Sec. 3 and here reported for completeness:

- 691 • $p_T^{B^0} > 8$ GeV/c: B^0 transverse momentum;
 692 • $|\eta^{B^0}| < 2.2$: B^0 pseudorapidity range.

693 For the sake of generality the efficiency is computed in bins of all the three angular observables,
 694 namely θ_K , θ_l , and ϕ (see Fig. 12). For the practical purpose of this analysis the efficiency can
 695 be considered constant along ϕ , as shown in Fig.s 31 and 32 therefore integrated along this
 696 variable. The final 2D-efficiency distributions are computed for signal simulation in each q^2
 697 bin as from Table 4, while for the control channels $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ and $B^0 \rightarrow$
 698 $K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ in the q^2 bin #4 and #6 respectively. The efficiency distributions projected
 699 over the angular variables are shown in Fig.s 31, 32 and Fig.s 33, 34, 35, 36 for signal and control
 700 channels respectively. It is important to stress that only the shape of the efficiency as a function
 701 of the angular observables and q^2 is important for this analysis, not its absolute value (see Sec. 4
 702 for more details on the analysis strategy).

703 The errors are assigned separately and independently to the two terms, \mathcal{E}_{ai} and $\mathcal{E}_{ri}^{R/M}$, according
 704 to the Binomial distribution. Unless explicitly stated in the text, every time it is referred to the
 705 efficiency it is meant the overall term $\mathcal{E}_i^{R/M}$.

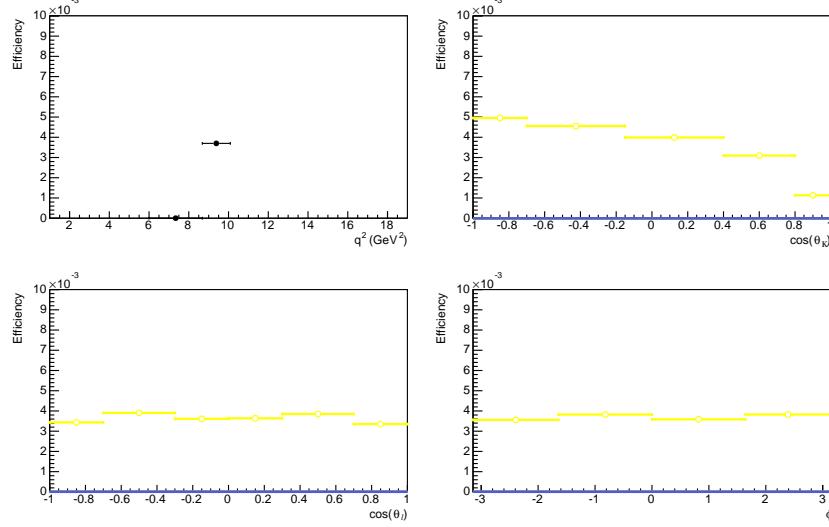


Figure 33: Efficiency computed with correctly tagged control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ simulation events. The first plot shows the distribution as a function of the dimuon q^2 , integrated over all angular observables; The second plot shows the distribution as a function of θ_K , integrated over θ_l and ϕ angular observables; the third plot shows the distribution as a function of θ_l integrated over θ_K and ϕ ; the fourth plot shows the distribution as a function of ϕ , integrated over θ_K and θ_l . The numbering goes from left to right and from top to bottom.

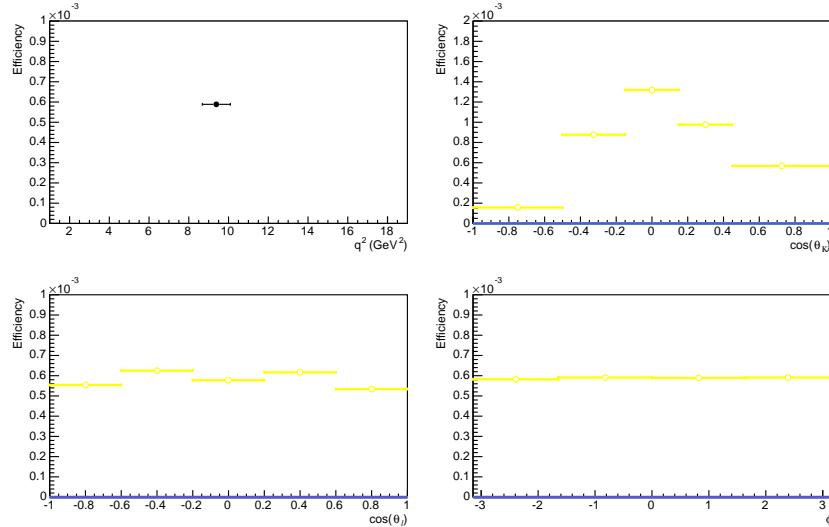


Figure 34: Efficiency computed with wrongly tagged control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ simulation events. The first plot shows the distribution as a function of the dimuon q^2 , integrated over all angular observables; The second plot shows the distribution as a function of θ_K , integrated over θ_l and ϕ angular observables; the third plot shows the distribution as a function of θ_l integrated over θ_K and ϕ ; the fourth plot shows the distribution as a function of ϕ , integrated over θ_K and θ_l . The numbering goes from left to right and from top to bottom.

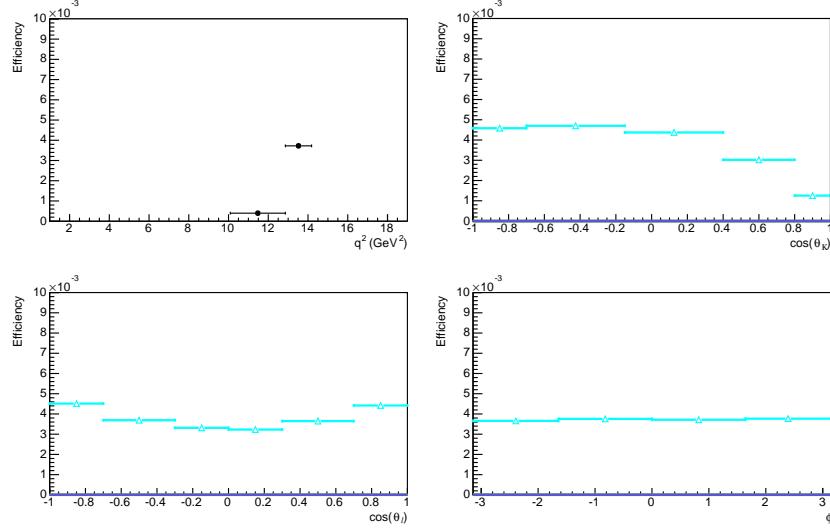


Figure 35: Efficiency computed with correctly tagged control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ simulation events. The first plot shows the distribution as a function of the dimuon q^2 , integrated over all angular observables; The secon plot shows the distribution as a function of θ_K , integrated over θ_l and ϕ angular observables; the third plot shows the distribution as a function of θ_l integrated over θ_K and ϕ ; the fourth plot shows the distribution as a function of ϕ , integrated over θ_K and θ_l . The numbering goes from left to right and from top to bottom.

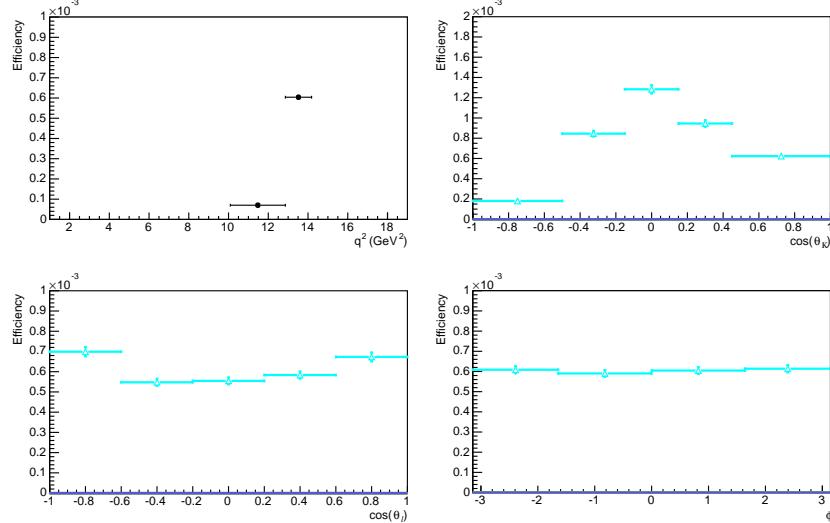


Figure 36: Efficiency computed with wrongly tagged control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ simulation events. The first plot shows the distribution as a function of the dimuon q^2 , integrated over all angular observables; The secon plot shows the distribution as a function of θ_K , integrated over θ_l and ϕ angular observables; the third plot shows the distribution as a function of θ_l integrated over θ_K and ϕ ; the fourth plot shows the distribution as a function of ϕ , integrated over θ_K and θ_l . The numbering goes from left to right and from top to bottom.

7.1 Analytical Description of Correctly Tagged Events

As described in Sec. 4 the implementation of the efficiency correction is embedded in the p.d.f. as a term that multiplies the functional form that describes the physics of the decay. Instead of directly using the binned efficiency of Fig.s 31 and 32, better fit performance can be achieved by implementing an analytical description. To this extent a 2D $3^{rd} \times 5^{th}$ -degree polynomial is conveniently used. The most general analytical form of the efficiency function $\mathcal{E}_i^R(\theta_K, \theta_l)$ is:

$$\begin{aligned}\mathcal{E}_i^R(\theta_K, \theta_l) = & (a_i + b_i x + c_i x^2 + d_i x^3) + (e_i + f_i x + g_i x^2 + h_i x^3) y + (j_i + k_i x + l_i x^2 + m_i x^3) y^2 + \\ & + (n_i + o_i x + p_i x^2 + q_i x^3) y^3 + (r_i + s_i x + t_i x^2 + u_i x^3) y^4 + (v_i + w_i x + z_i x^2 + z z_i x^3) y^5,\end{aligned}\quad (7)$$

where the subscript i runs over the q^2 bins and for the sake of readability the notation is simplified: $x \equiv \cos(\theta_K)$ and $y \equiv \cos(\theta_l)$. It's worth noting that all possible correlation terms are being considered.

In order to guarantee the convergence of the fits to the 2D binned distributions the procedure to build the analytical efficiency functions is performed in steps:

1. Fit the efficiency distributions as a function of θ_l for all θ_K and q^2 bins with a 5^{th} -degree polynomial. In order to avoid that the efficiency becomes negative constraints at -1 and/or $+1$ are set for the bins:
 - q^2 bins #0 and for all θ_K bins;
 - q^2 bins #1 and for all θ_K bins;
 - q^2 bins #2 and for all θ_K bins;
 - q^2 bins #3 and for all θ_K bins.
2. Interpolate with a 3^{rd} -degree polynomial the coefficients of the polynomials at point 1. over the θ_K variable.
3. Fit the 2D-efficiency distributions with all parameters determined at point 1. and 2. free to float. The same constraints described at point 1. are here applied. After this last step the analytical form of the efficiency function, $\mathcal{E}_i^R(\theta_K, \theta_l)$ in Eq. 5, is fully determined. As fringe benefit the full covariance matrix of the coefficients is obtained, which will be used to determine the systematic uncertainty related to the **Limited Number of Simulated Events** as described in Sec. 10.1. As a conservative approach for the computation of such a covariance matrix the errors are magnified in order to obtain a fit $\chi^2/d.o.f. \simeq 1$.

The 2D-efficiency distribution, together with the analytical function, are presented in Fig.s 37, 38, 39, and 40 for all q^2 bins for signal simulation, and in Fig.s 41 and 42 for q^2 bin #4 and #6 respectively for the control channel simulations.

7.2 Description of Wrongly Tagged Events

The correctly and wrongly tagged events have different efficiency distributions as one can see by comparing, for example, Fig.s 31 and 32. Given that the wrongly tagged events are $\sim 13\%$ of the total number of events, and given the limited amount of MC simulated events, the angular variables have been re-binned in order to avoid empty bins.

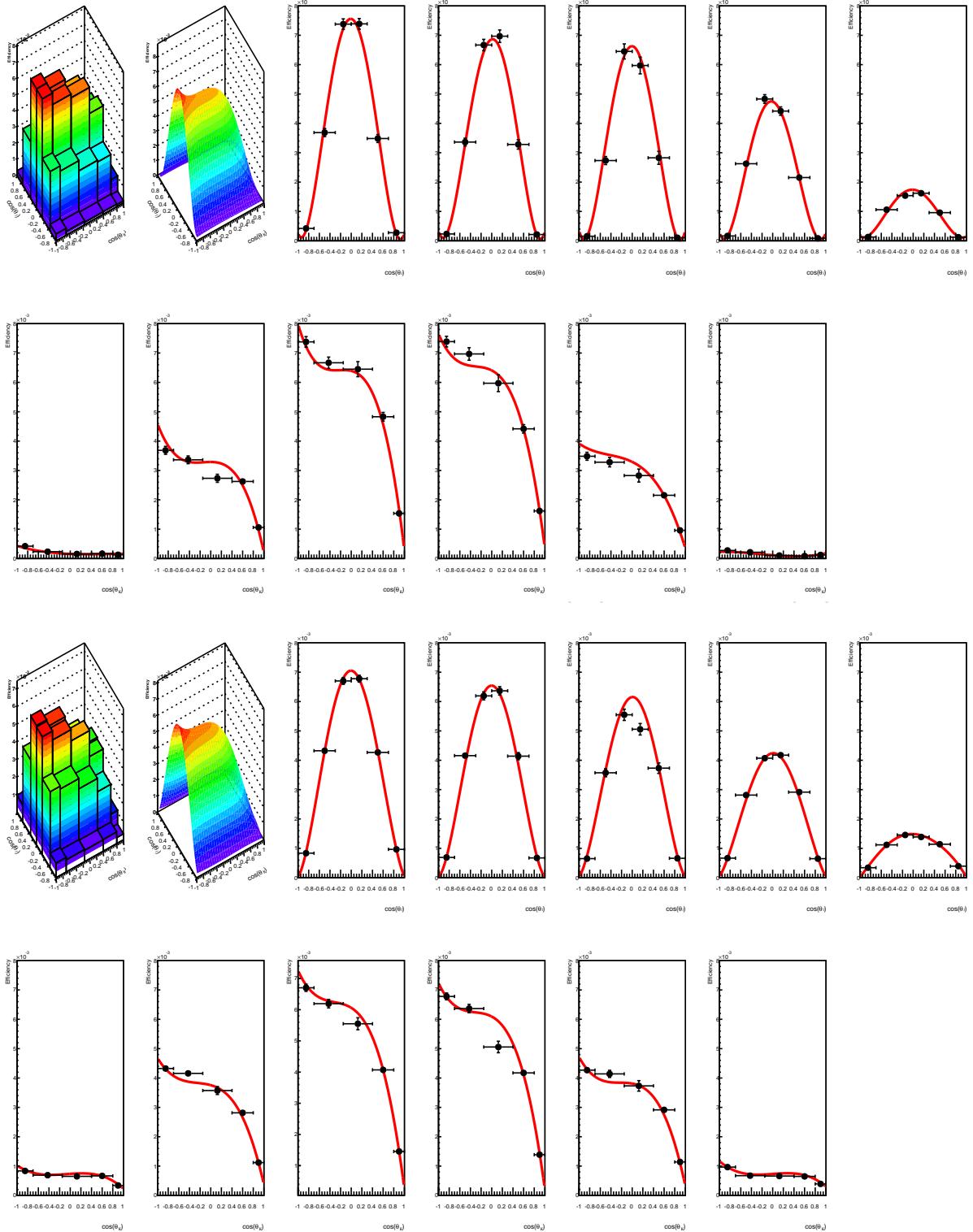


Figure 37: 2D-efficiency distributions computed with correctly tagged signal simulation and fitted with a polynomial function for the q^2 bin #0 (top thirteen plots) and #1 (bottom thirteen plots), where for each q^2 bin are shown the 2D-efficiency distribution, the analytical function, and eleven cross-sections of both the distribution and the analytical function (five along θ_l and six along θ_K).

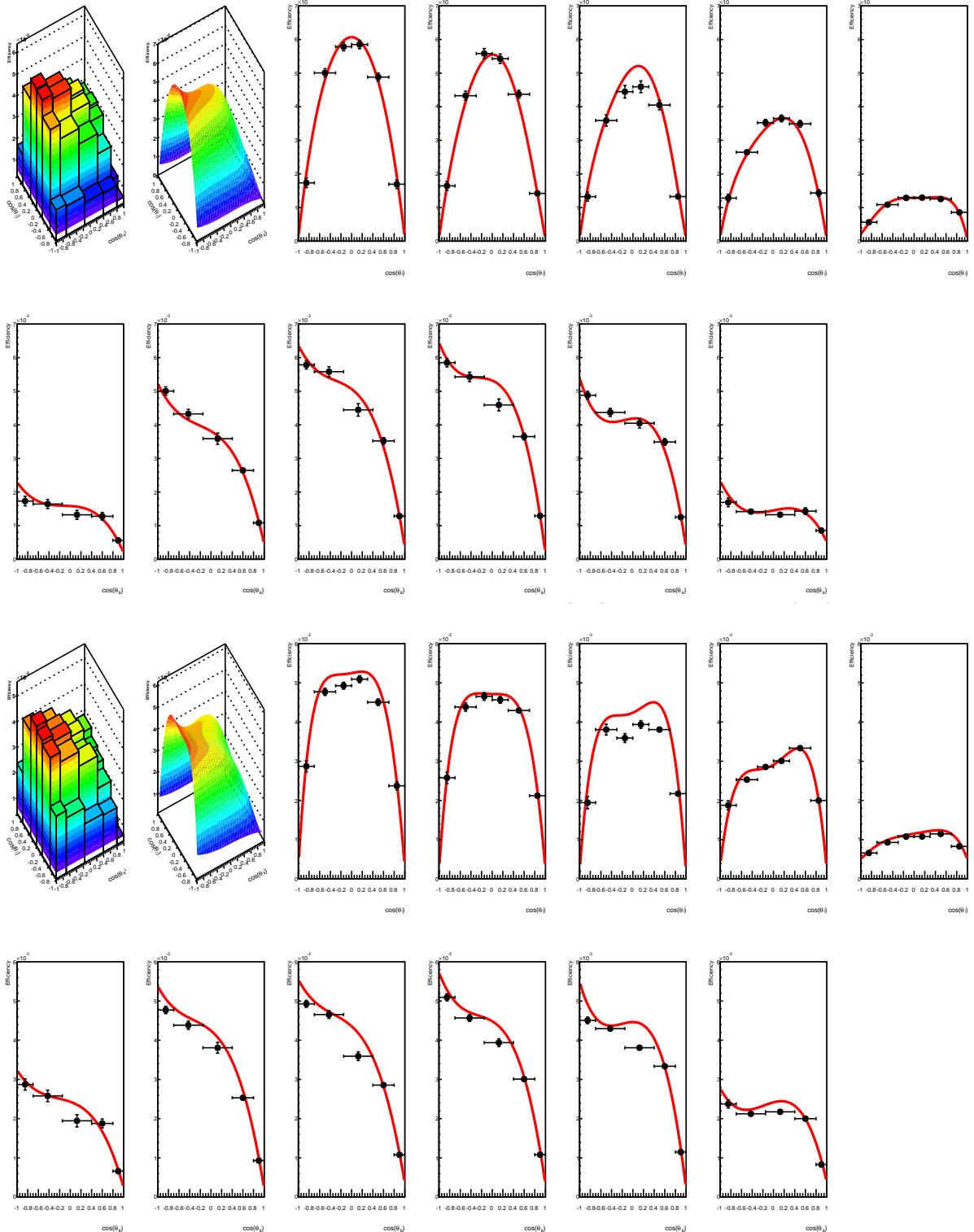


Figure 38: 2D-efficiency distributions computed with correctly tagged signal simulation and fitted with a polynomial function for the q^2 bin #2 (top thirteen plots) and #3 (bottom thirteen plots), where for each q^2 bin are shown the 2D-efficiency distribution, the analytical function, and eleven cross-sections of both the distribution and the analytical function (five along θ_l and six along θ_K).

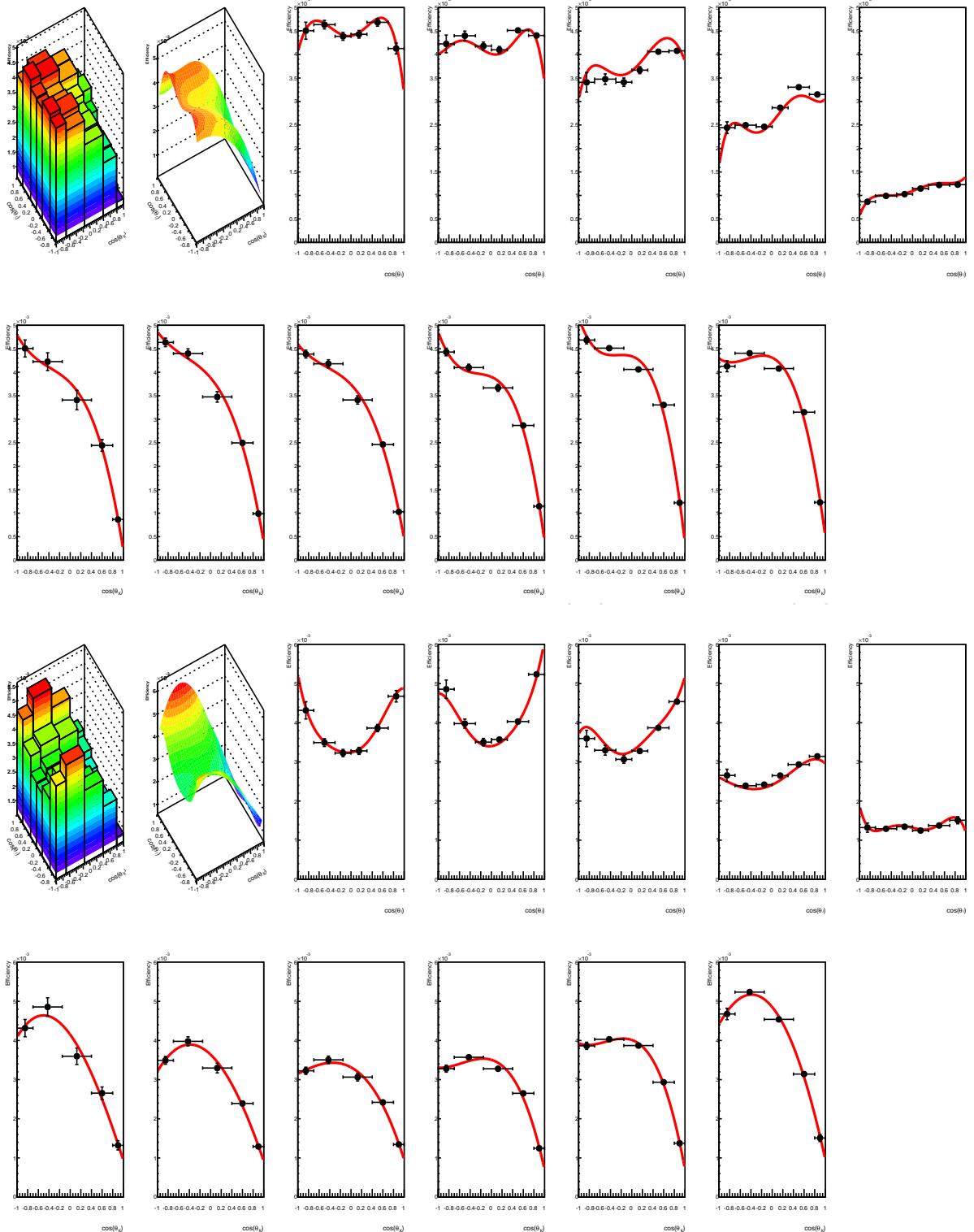


Figure 39: 2D-efficiency distributions computed with correctly tagged signal simulation and fitted with a polynomial function for the q^2 bin #5 (top thirteen plots) and #7 (bottom thirteen plots), where for each q^2 bin are shown the 2D-efficiency distribution, the analytical function, and eleven cross-sections of both the distribution and the analytical function (five along θ_l and six along θ_K).

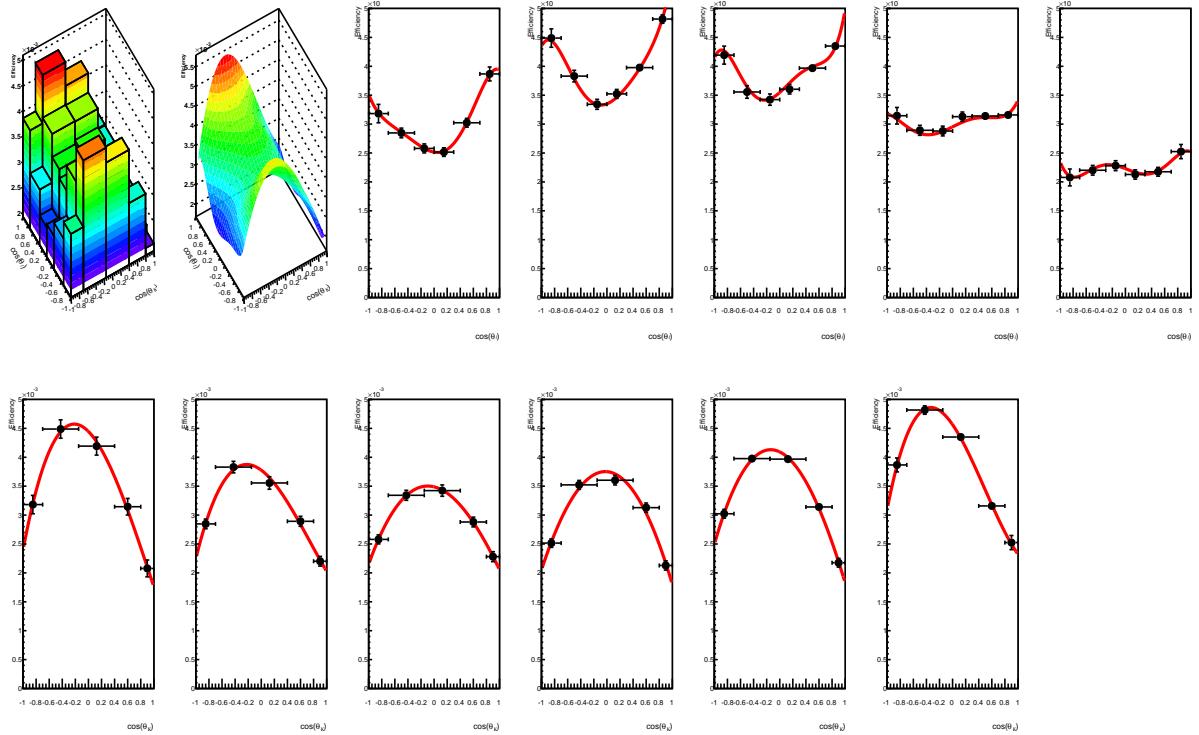


Figure 40: 2D-efficiency distributions computed with correctly tagged signal simulation and fitted with a polynomial function for the q^2 bin #8, where are shown the 2D-efficiency distribution, the analytical function, and eleven cross-sections of both the distribution and the analytical function (five along θ_l and six along θ_K).

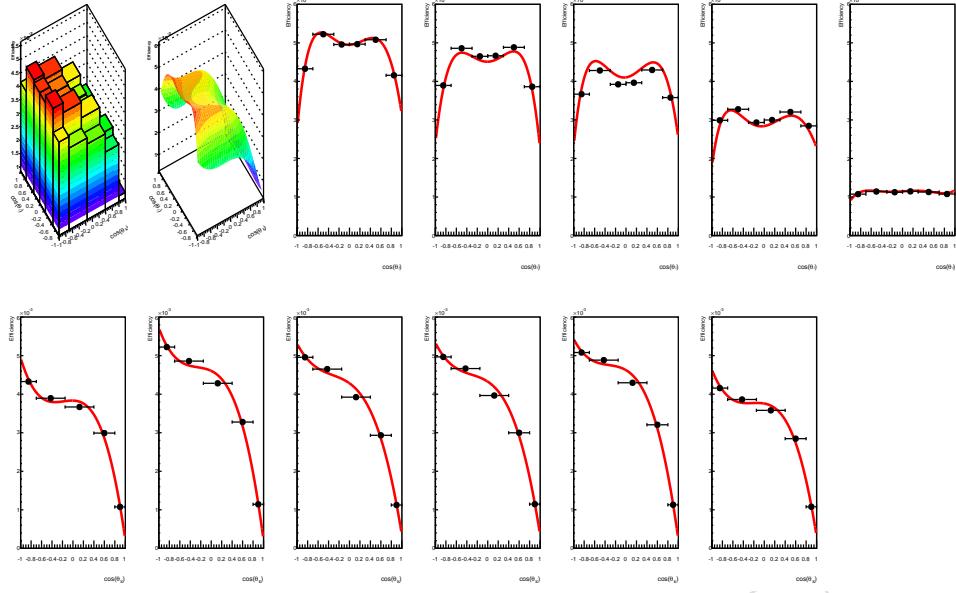


Figure 41: 2D-efficiency distributions computed with correctly tagged control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi(\mu^+\mu^-)$ simulation and fitted with a polynomial function. In the figure are shown the 2D-efficiency distribution, the analytical function, and eleven cross-sections of both the distribution and the analytical function (five along θ_l and six along θ_K).

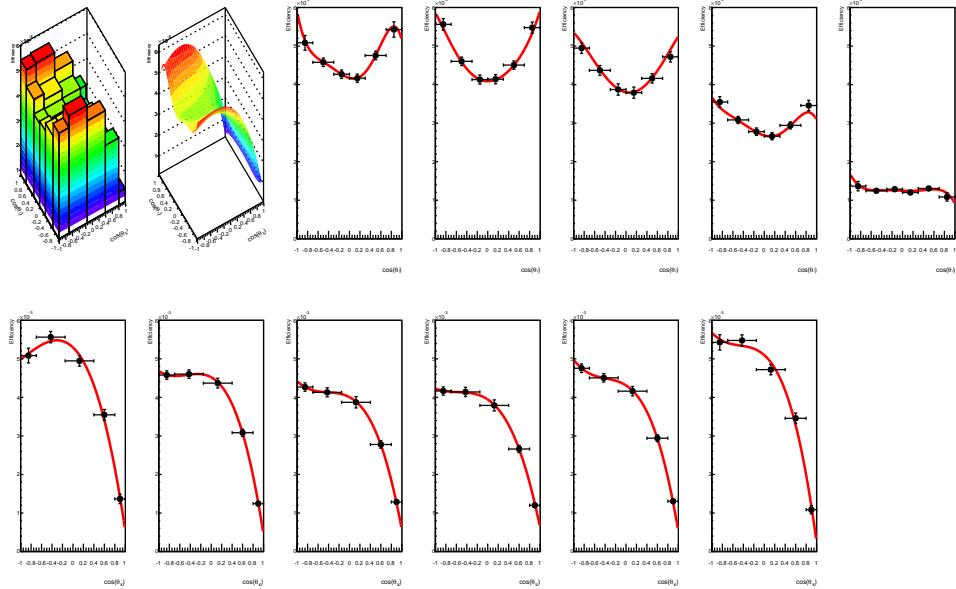


Figure 42: 2D-efficiency distributions computed with correctly tagged control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ simulation and fitted with a polynomial function. In the figure are shown the 2D-efficiency distribution, the analytical function, and eleven cross-sections of both the distribution and the analytical function (five along θ_l and six along θ_K).

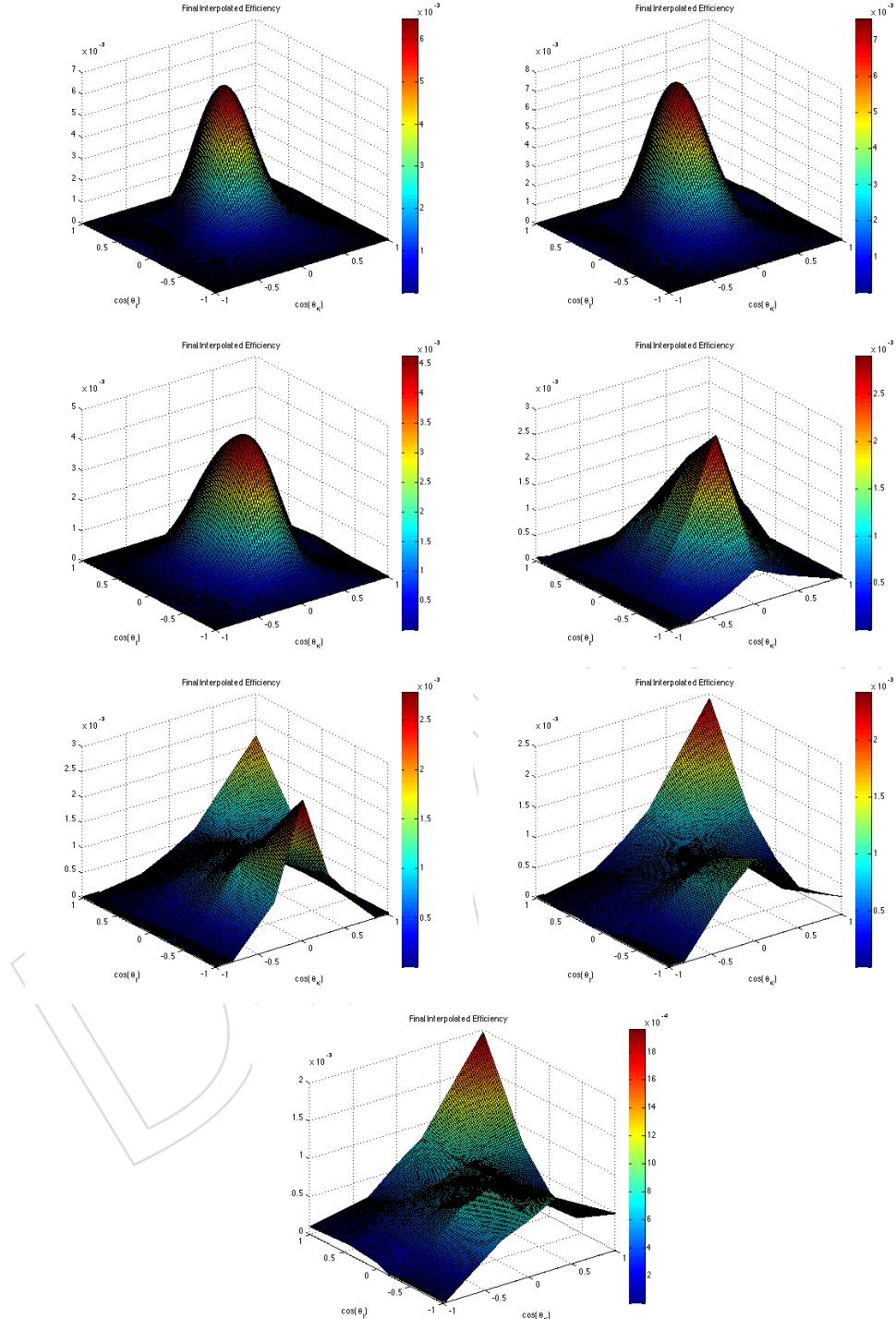


Figure 43: 2D-efficiency interpolated and re-binned over a 120×120 bin matrix. The plots refer to wrongly tagged signal events for each q^2 bin in Table 4 but #4 and #6. The numbering goes from left to right and from top to bottom.

741 It would have been preferable to apply the aforementioned procedure to compute an analytical
 742 efficiency also for the wrongly tagged events, but the lack of events together with the peculiar
 743 efficiency shape, would have required a 2D polynomial that would have saturated the available
 744 number of degrees of freedom. Therefore a different description of the efficiency is adopted:

- 745 1. Interpolate with MATLAB the coarse binned efficiency from Fig.s 32, 34, and 36.
 746 2. Finely bin the efficiency from step 1. (i.e. 120×120 bins).
 747 3. If the efficiency computed at point 2. is negative on a certain point of the domain it is
 748 then casted to $+2 \times 10^{-5}$;
 749 4. Use the binned efficiency from step 3. with a linear interpolation as $\mathcal{E}_i^M(\theta_K, \theta_l)$ in Eq. 5.

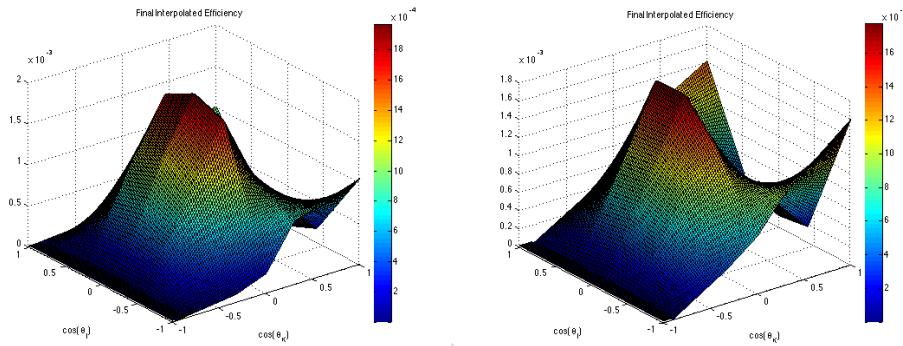


Figure 44: 2D-efficiency interpolated and re-binned over a 120×120 bin matrix. The plots refer to wrongly tagged control channel events, $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ left and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ right, for the q^2 bins #4 and #6 in Table 4 respectively.

750 The MATLAB function used to perform the interpolation is `griddedInterpolant` [23]. For
 751 the q^2 bins from #0 to #2, as from Table 4, a spline interpolation is adopted together with
 752 constraints set to $+10^{-5}$ at the boundaries of the domain in order avoid that the efficiency becomes
 753 negative. For the remaining bins a linear interpolation and a linear extrapolation is adopted,
 754 without the need of using constraints at the boundaries of the domain. The reason for adopt-
 755 ing two kind of interpolating functions, spline and linear, for two different sets of q^2 bins is
 756 because for the bins from #3 to #8 no constraints at the boundaries are needed, therefore an
 757 unconstrained spline function would have the tendency of having high curvatures on the per-
 758 ipery bins, causing a mismatch with respect to the trend of the binned efficiency. On the
 759 other hand for the bins from #0 to #2 a linear interpolation is less suited given the peculiar
 760 “narrow-bell” shape of the efficiency.

761 The MATLAB interpolation method creates interpolating functions that pass through the sam-
 762 ple data points. Therefore a query to the interpolation function at a sample location, would give
 763 back the value at that sample data point. Figures 43 and 44 show the result of the MATLAB
 764 interpolation and re-binning.

765 The 2D-efficiency distribution, together with the interpolated function, are presented in Fig.s 45,
 766 46, 47, and 48 for all q^2 bins for signal simulation, and in Fig.s 49 and 50 for q^2 bin #4 and #6
 767 respectively for the control channel simulations. It’s worth noting that in some cases in the
 768 cross-section plots the binned and the interpolated efficiencies does not precisely match. The
 769 reason is simply because the 120×120 bin matrix used for the re-binning doesn’t precisely

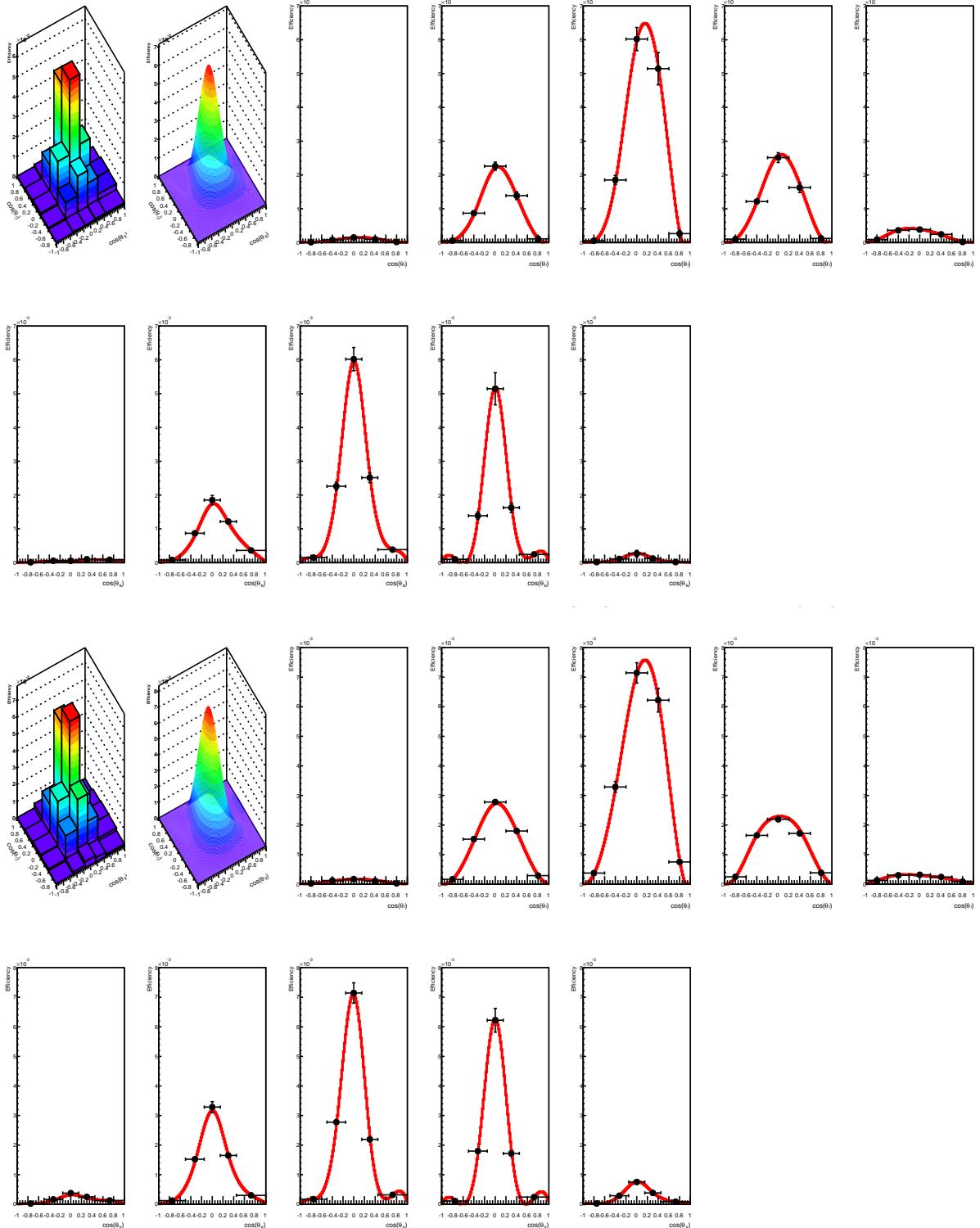


Figure 45: 2D-efficiency distributions computed with wrongly tagged signal simulation, interpolated and re-binned using MATLAB, for the q^2 bin #0 (top twelve plots) and #1 (bottom twelve plots), where for each q^2 bin are shown the 2D-efficiency distribution, the interpolated function, and ten cross-sections of both the distribution and the interpolated function (five along θ_l and five along θ_K).

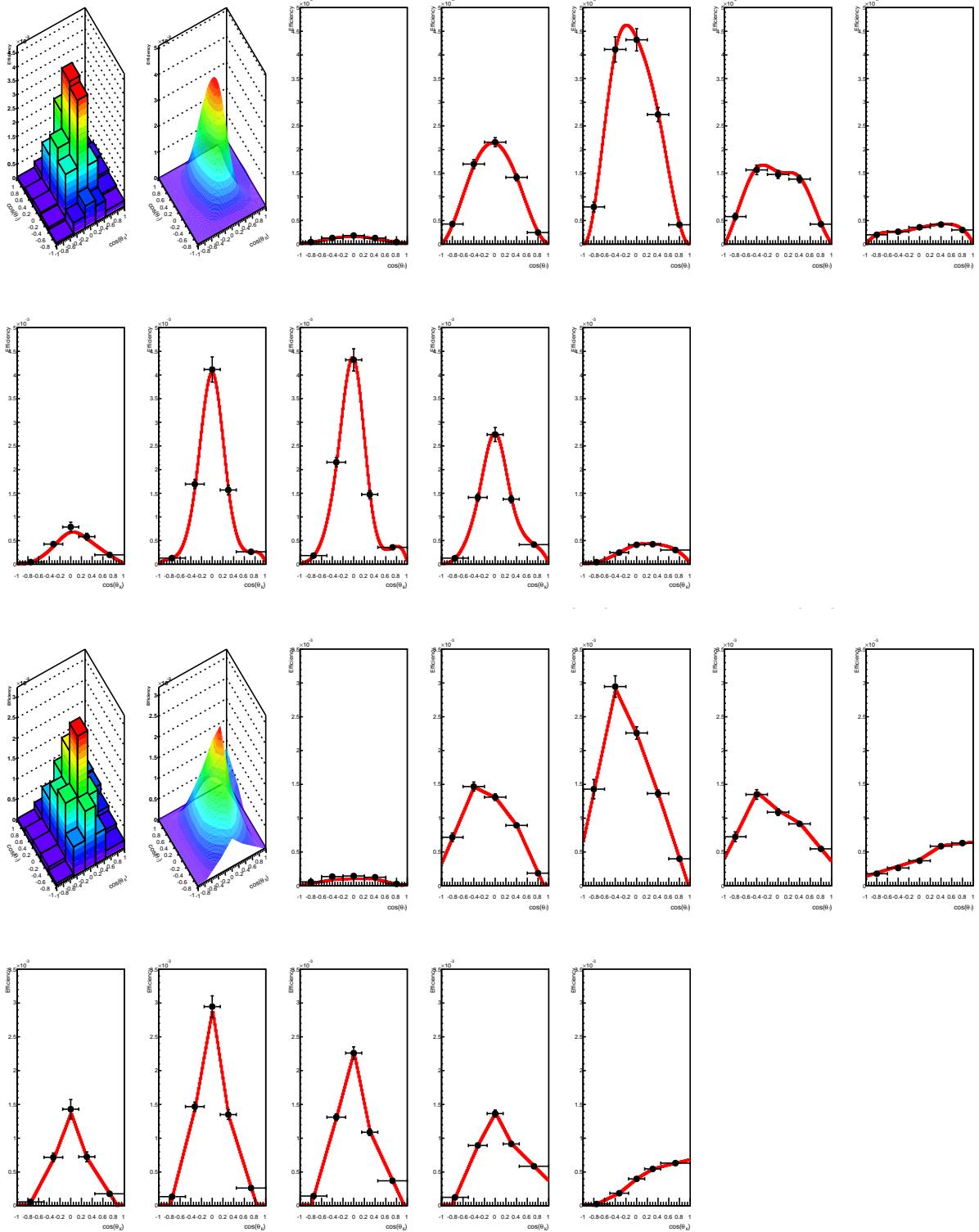


Figure 46: 2D-efficiency distributions computed with wrongly tagged signal simulation, interpolated and re-binned using MATLAB, for the q^2 bin #2 (top twelve plots) and #3 (bottom twelve plots), where for each q^2 bin are shown the 2D-efficiency distribution, the interpolated function, and ten cross-sections of both the distribution and the interpolated function (five along θ_l and five along θ_K).

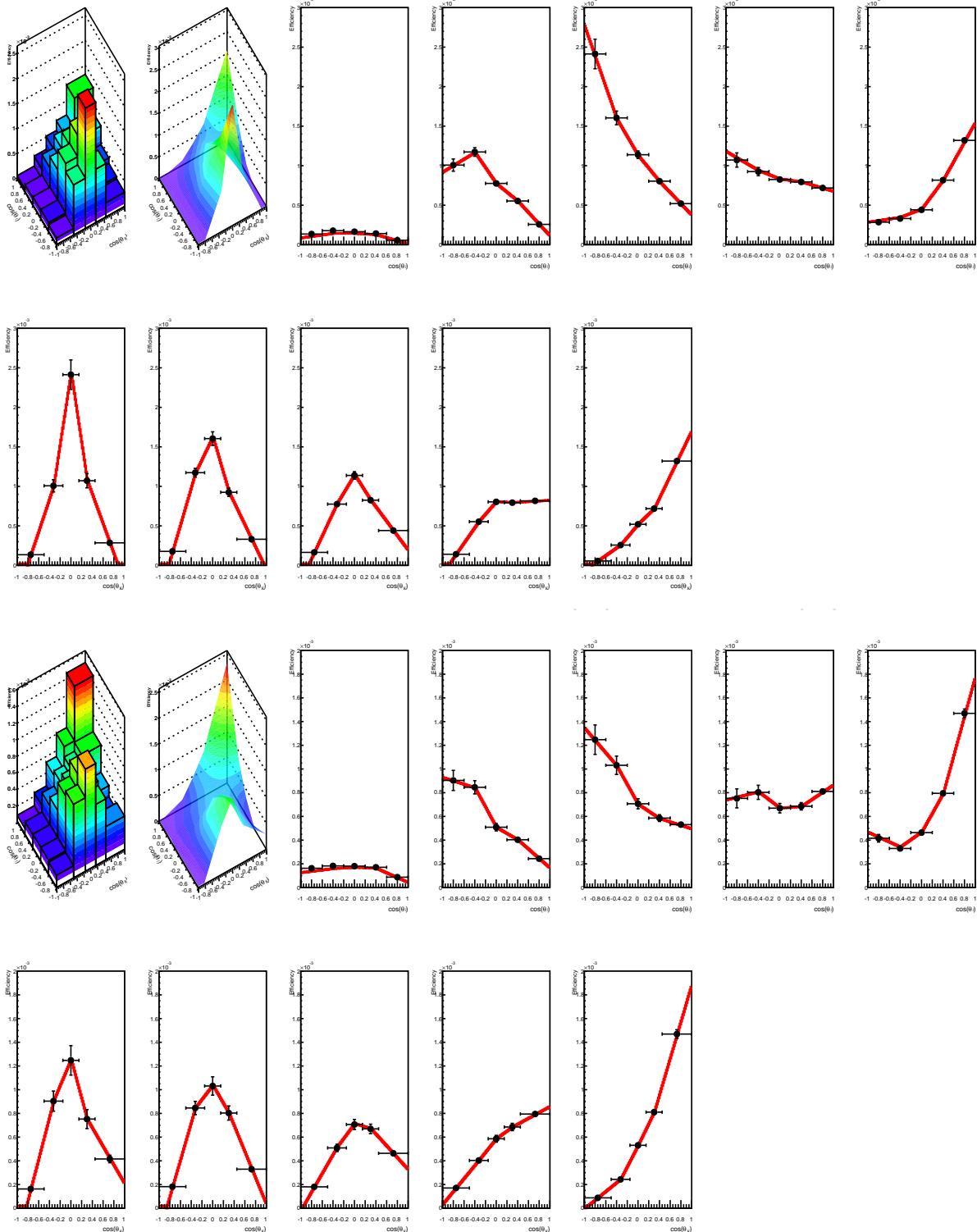


Figure 47: 2D-efficiency distributions computed with wrongly tagged signal simulation, interpolated and re-binned using MATLAB, for the q^2 bin #5 (top twelve plots) and #7 (bottom twelve plots), where for each q^2 bin are shown the 2D-efficiency distribution, the interpolated function, and ten cross-sections of both the distribution and the interpolated function (five along θ_l and five along θ_K).

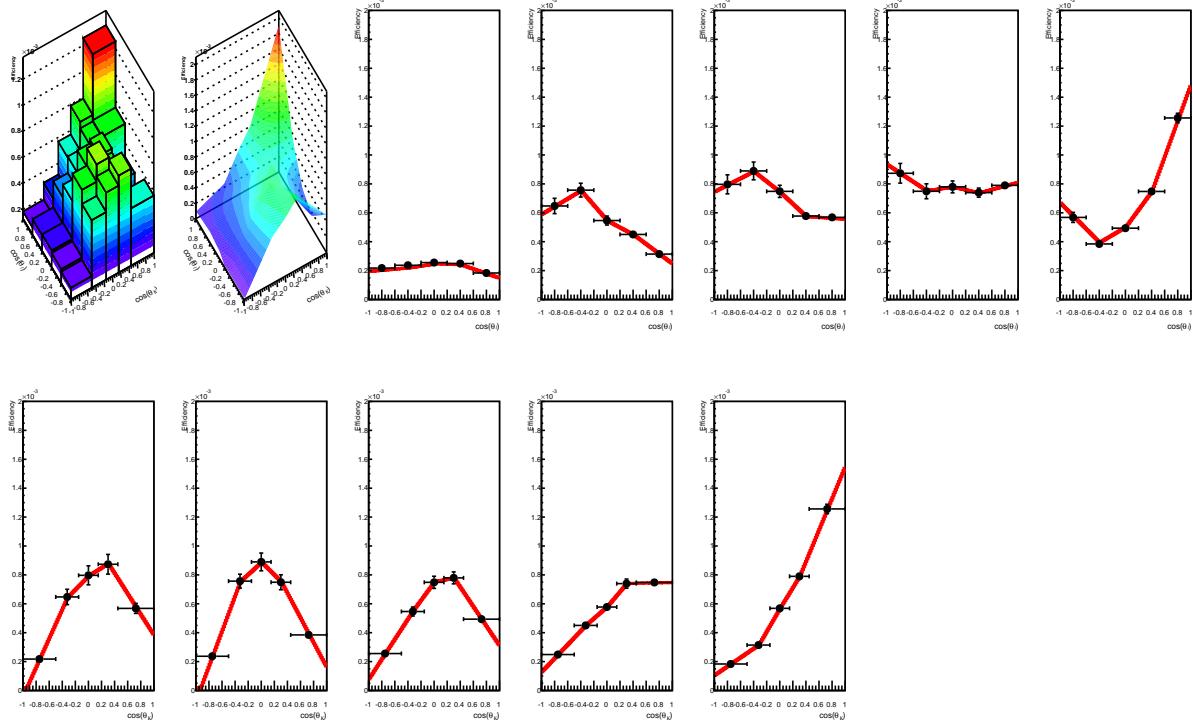


Figure 48: 2D-efficiency distributions computed with wrongly tagged signal simulation, interpolated and re-binned using MATLAB, for the q^2 bin #8, where are shown the 2D-efficiency distribution, the interpolated function, and ten cross-sections of both the distribution and the interpolated function (five along θ_l and five along θ_K).

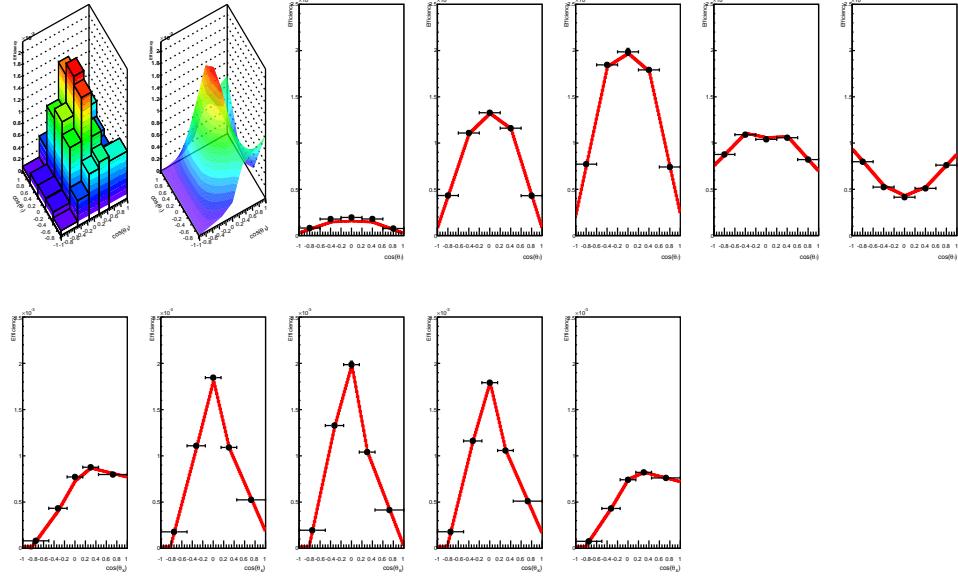


Figure 49: 2D-efficiency distributions computed with wrongly tagged control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi(\mu^+\mu^-)$ simulation, interpolated and re-binned using MATLAB, In the figure are shown the 2D-efficiency distribution, the interpolated function, and ten cross-sections of both the distribution and the interpolated function (five along θ_l and five along θ_K).

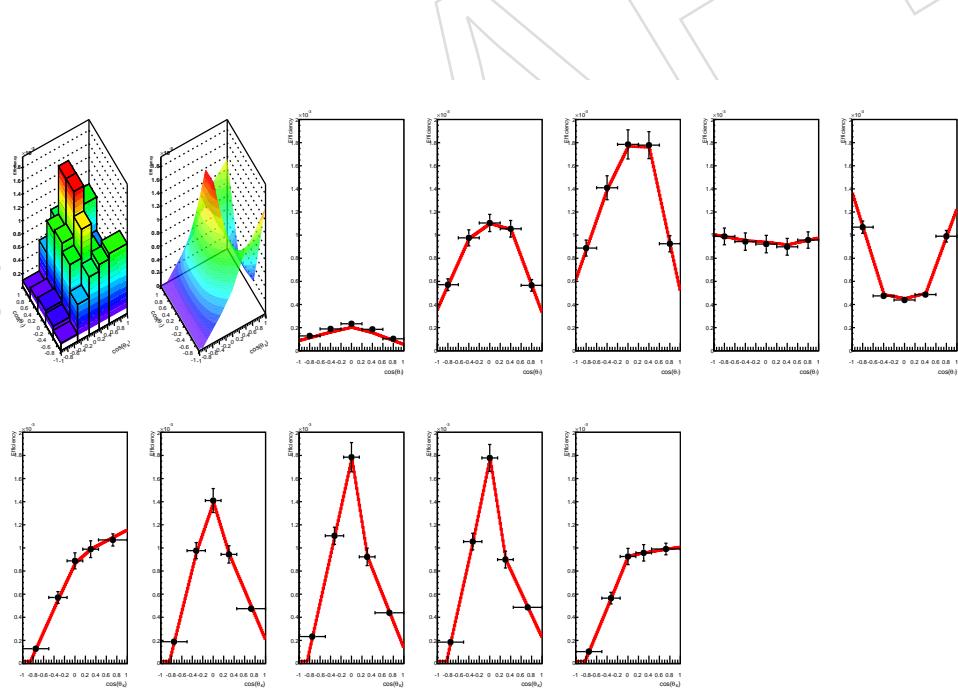


Figure 50: 2D-efficiency distributions computed with wrongly tagged control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ simulation, interpolated and re-binned using MATLAB, In the figure are shown the 2D-efficiency distribution, the interpolated function, and ten cross-sections of both the distribution and the interpolated function (five along θ_l and five along θ_K).

⁷⁷⁰ match the center of the original coarse binning. Therefore, especially in proximity of high gra-
⁷⁷¹ dient, the two efficiencies might have a slight shift.

⁷⁷² The procedures just described, both for correctly and wrongly CP-tagged events, are adopted
⁷⁷³ to compute the efficiencies both for signal and control channels. The validation of the efficiency
⁷⁷⁴ is presented in the **Efficiency Shape** Sec. 10.2.

DRAFT

775 8 Analysis Validation

776 A detailed validation of the analysis is performed with MC simulation and confirmed with
 777 data with the control channels. All analysis steps, i.e. signal selection, efficiency computation,
 778 and fit procedure, are applied on signal simulation, also called RECO-MC, just as they are ap-
 779 plied on data. The interesting observables, namely A_{FB} and F_L , are measured as a function of
 780 the dimuon mass squared, q^2 , and compared to values extracted from the simulation at pure
 781 generator level (GEN-MC). In particular on GEN-MC, due to the absence of any selection and
 782 background, only the pure angular shape is fitted to extract A_{FB} and F_L using Eq. 3. As already
 783 stated in Sec. 6, both signal and control channel simulations do not account for the decays in
 784 which the $K\pi$ particles are in the S -wave final state.

785 8.1 Pileup Comparison Between Data and Simulation

786 The MC simulation is produced with the run-dependent technique that should allow to em-
 787 ualate the same pileup distribution that is present in data (see Sec. 2). Figure 51 shows three
 788 comparison plots between data, signal, and control channels MC simulations. The distribu-
 789 tions perfectly match implying that there is no need to re-weight the events. Indeed no event
 790 re-weighting is used in this analysis.

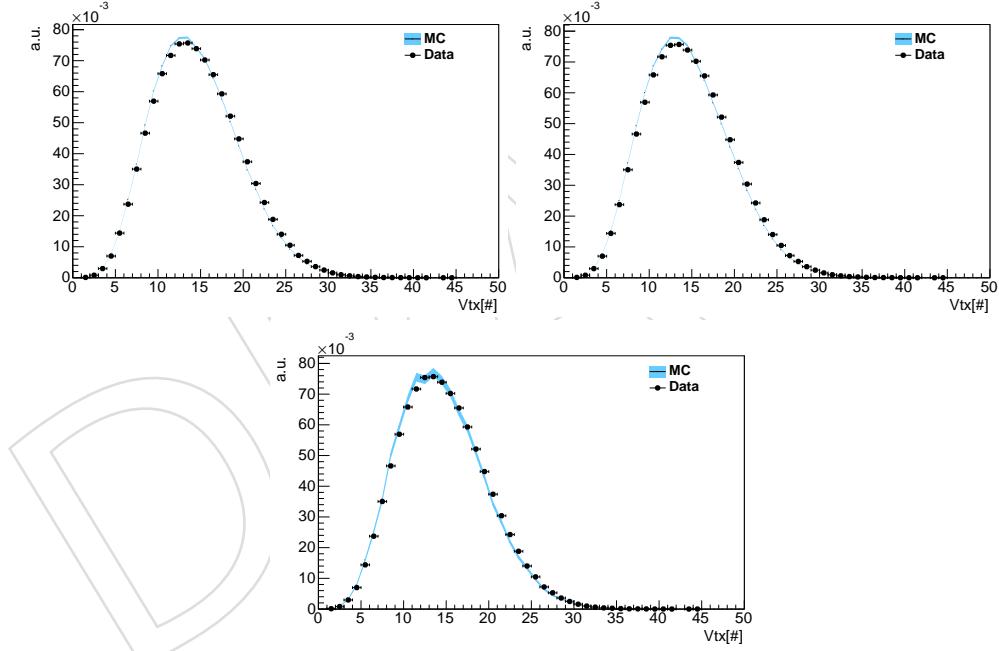


Figure 51: Primary vertex distributions. The first plot show a comparison between data and the signal MC simulation. The second plot show a comparison between data and the $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ MC simulation. The third plot show a comparison between data and the $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ MC simulation. The numbering goes from left to right and from top to bottom.

791 8.2 Validation with High-Statistics, Pure Signal Simulation

792 The analysis technique is validated with the highest number of events available signal MC
 793 simulation, without background, in order to verify whether asymptotically the analysis is able
 794 to measure the interesting observables, and to quantify possible biases in the measurements.

795 In the simulation both correctly and wrongly tagged events are present, therefore the fits are
 796 performed with a p.d.f. containing both components.

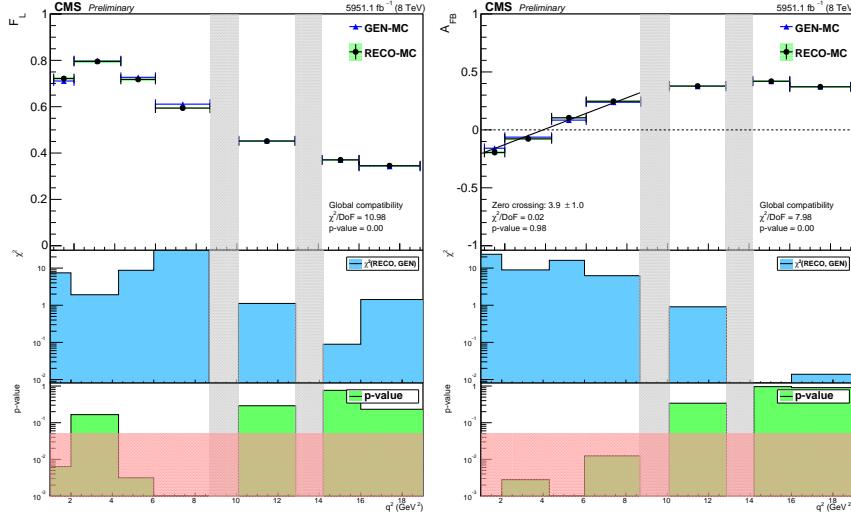


Figure 52: Validation of the analysis with truth-matched (see definition of truth-matching in App. A) signal simulation: comparison between the results of the analysis (black) and the actual values used to generate the simulation (blue) both for F_L (left) and A_{FB} (right). The first plot shows the values as a function of the dimuon q^2 . The second plot shows the χ^2 of compatibility for each bin, while the third plot shows the p-value (in red is shaded the 5% level of significance). The global χ^2 of compatibility is reported in the bottom right corner of the top plots. For A_{FB} also the zero crossing point is extracted. The numbering goes from top to bottom. The errors are purely statistical. Grey shaded regions correspond to the J/ψ and ψ' resonances.

Table 9: Discrepancies per bin between GEN-MC and RECO-MC as from Fig.s 56 and 53.

q^2 bin index	Simulation Mismodeling: F_L	Simulation Mismodeling: A_{FB}	Simulation Mismodeling: $d\mathcal{B}/dq^2$
0	± 0.0106	± 0.0369	± 0.0116
1	± 0.0037	± 0.0147	± 0.0062
2	± 0.0091	± 0.0203	± 0.0190
3	± 0.0170	± 0.0083	± 0.0151
5	± 0.0021	± 0.0027	± 0.0212
7	± 0.0008	± 0.0002	± 0.0129
8	± 0.0032	± 0.0004	± 0.0024

797 The results for A_{FB} and F_L are presented in Fig. 52. The small discrepancies, reported in Table 9,
 798 is conservatively considered as symmetric systematic uncertainty (called **Simulation Mismod-**
 799 **eling** in Sec. 10.3).

800 This validation step is also performed for the $d\mathcal{B}/dq^2$ observable. To this extent the GEN-MC
 801 yields are simply counted and rescaled by the luminosity in order to compare them with the dif-
 802 ferential branching fraction measured with RECO-MC. The results for the differential branching
 803 fraction are presented in Fig. 53. The small discrepancies, reported in Table 9, is conservatively
 804 considered as symmetric systematic uncertainty (called **Simulation Mismodeling** in Sec. 10.3).

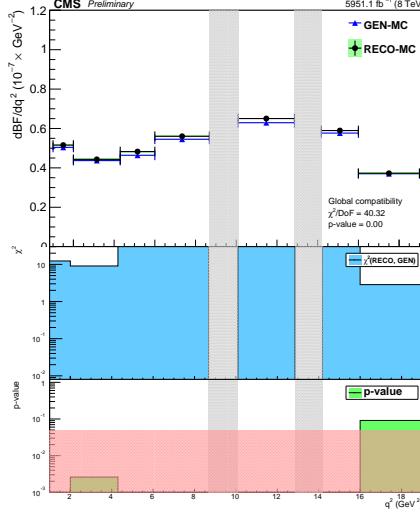


Figure 53: Validation of the analysis with truth-matched (see definition of truth-matching in App. A) signal simulation: comparison between the results of the analysis (black) and the actual values used to generate the simulation (blue) for the differential branching fraction. The first plot shows the values as a function of the dimuon q^2 . The second plot shows the χ^2 of compatibility for each bin, while the third plot shows the p-value (in red is shaded the 5% level of significance). The global χ^2 of compatibility is reported in the bottom right corner of the top plot. For A_{FB} also the zero crossing point is extracted. The numbering goes from top to bottom. The errors are purely statistical. Grey shaded regions correspond to the J/ψ and ψ' resonances.

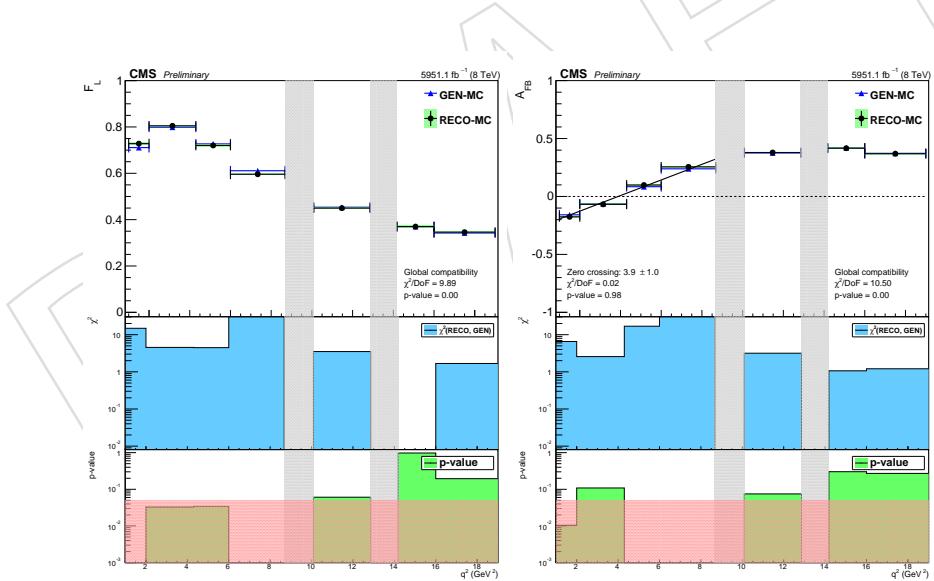


Figure 54: Validation of the analysis with correctly CP-tagged truth-matched (see definition of truth-matching in App. A) signal simulation: comparison between the results of the analysis (black) and the actual values used to generate the simulation (blue) both for F_L (left) and A_{FB} (right). The first plot shows the values as a function of the dimuon q^2 . The second plot shows the χ^2 of compatibility for each bin, while the third plot shows the p-value (in red is shaded the 5% level of significance). The global χ^2 of compatibility is reported in the bottom right corner of the top plots. For A_{FB} also the zero crossing point is extracted. The numbering goes from top to bottom. The errors are purely statistical. Grey shaded regions correspond to the J/ψ and ψ' resonances.

805 A thousand pseudo-experiments are also run with RECO-MC values. The average values of
 806 the pseudo-experiment fits agree with the generated values within 5×10^{-4} , confirming that
 807 the discrepancies are not caused by an intrinsic bias of the analysis, but rather they should be
 808 treated as systematic uncertainties.

809 8.2.1 Separate Validation for Correctly and Wrongly CP-tagged Events

810 This section is meant to go more deeply in the validation of the analysis technique. The com-
 811 parison between GEN-MC and RECO-MC is performed separately for correctly tagged, Fig. 54,
 812 and wrongly tagged, Fig. 55, events, always using the highest number of events available signal
 813 MC simulation, without background.

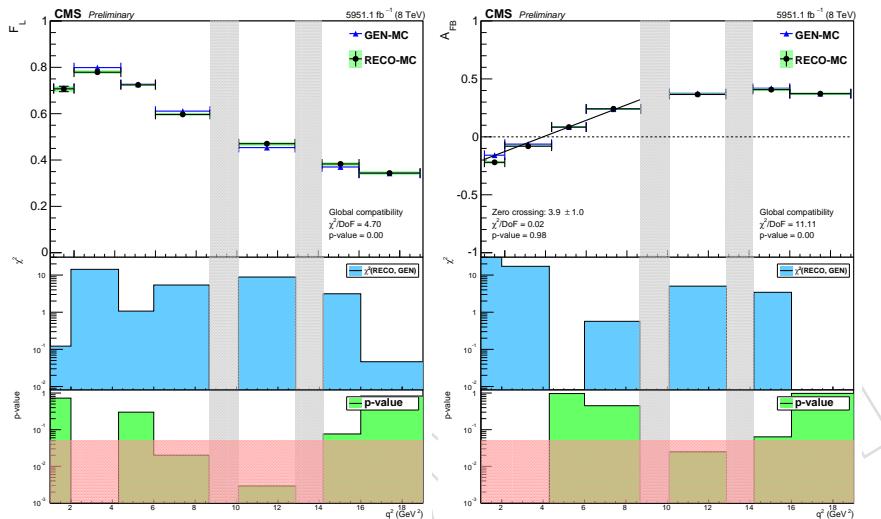


Figure 55: Validation of the analysis with wrongly CP-tagged truth-matched (see definition of truth-matching in App. A) signal simulation: comparison between the results of the analysis (black) and the actual values used to generate the simulation (blue) both for F_L (left) and A_{FB} (right). The first plot shows the values as a function of the dimuon q^2 . The second plot shows the χ^2 of compatibility for each bin, while the third plot shows the p-value (in red is shaded the 5% level of significance). The global χ^2 of compatibility is reported in the bottom right corner of the top plots. For A_{FB} also the zero crossing point is extracted. The numbering goes from top to bottom. The errors are purely statistical. Grey shaded regions correspond to the J/ψ and ψ' resonances.

814 The comparison shows that, though there are small discrepancies, all in all the two signal
 815 components are well separately modeled. The systematic uncertainty to be attributed to the
 816 validation with high-statistics, pure signal simulation is based on the GEN-MC to RECO-MC
 817 comparison having both correctly and wrongly tagged signal event components.

818 8.3 Validation with High Statistics Full Simulation

819 The analysis technique is validated with the simulation containing: signal, control channels,
 820 and background. The number of events corresponds to the full statistics of the signal MC
 821 simulation, i.e. $\sim 5951.1 \text{ fb}^{-1}$. The goal is to verify whether in data-like conditions, and with a
 822 high number of events, the analysis is able to measure the interesting observables.

823 The simulation is a mix of three datasets:

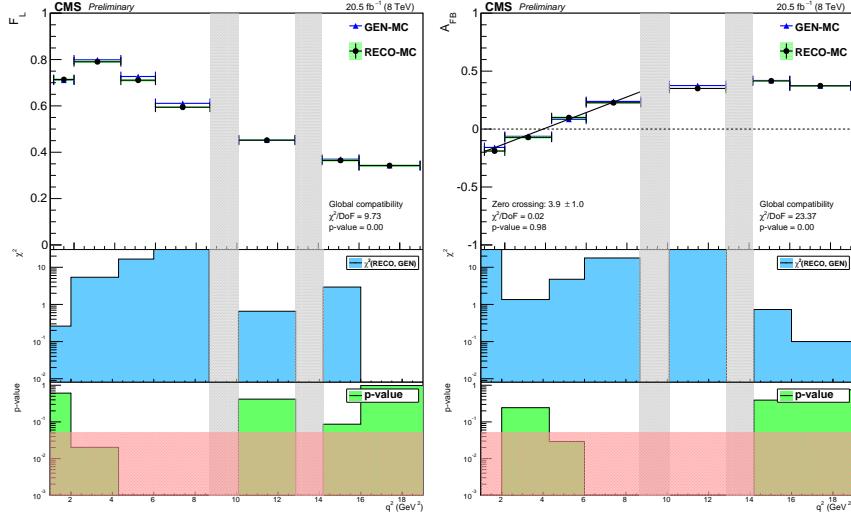


Figure 56: Validation of the analysis with the simulation containing: signal, control channels, and background. The number of events corresponds to $\sim 5951.1 \text{ fb}^{-1}$. Comparison between the results of the analysis (blue) and the actual values used to generate the simulation (black) both for F_L (left) and A_{FB} (right). The first plot shows the values as a function of the dimuon q^2 . The second plot shows the χ^2 of compatibility for each bin, while the third plot shows the p-value (in red is shaded the 5% level of significance). The numbering goes from top to bottom. The errors are purely statistical. Grey shaded regions correspond to the J/ψ and ψ' resonances.

- Signal MC simulation $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^- (\sim 5951.1 \text{ fb}^{-1})$.
- Feed-through distributions generated from the MC simulation of both $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ ($\sim 52.5 \text{ fb}^{-1}$) and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ ($\sim 58.5 \text{ fb}^{-1}$) decays. The generation is performed with pseudo-experiments whose angular shape is the result of linear interpolations of the MC's distributions. The reason for a linear interpolation is to smooth out statistical fluctuations in order to rescale these samples to a different luminosity.
- Background distribution (see Sec. 4.4 for the definition of background) generated with pseudo-experiments with p.d.f. parameters measured with data.

Both feed-through and background datasets are generated with a number of events corresponding to the same luminosity of the signal, i.e. $\sim 5951.1 \text{ fb}^{-1}$.

The results for F_L and A_{FB} are presented in Fig. 56. The small discrepancies present in all bins are comparable to those measured in Sec. 8.2. A thousand pseudo-experiments are also run with RECO-MC values. The average values of the pseudo-experiment fits disagree with the generated values by less than 2×10^{-3} , which is nonetheless greater than the errors on the average values, which are smaller than 4×10^{-4} . The disagreement has to be therefore attributed to a bias in the analysis which is more appropriately studied with data-like conditions and data-like number of events as presented in Sec. 8.4.

8.3.1 Validation of the Measurement of the A_{FB} Zero Crossing Point

The measurement of the zero crossing point of A_{FB} , A_{FB}^z , is performed by simply fitting with a straight line the A_{FB} trend in the first q^2 bins, i.e. from #0 to #3 as from Table 4.

On GEN-MC it is instead measured by simply applying the definition of forward-backward

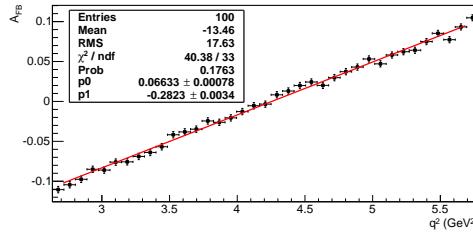


Figure 57: Extraction of the forward-backward asymmetry from GEN-MC. The values are computed by applying the definition of forward-backward asymmetry, i.e. $A_{FB} = \frac{N_f - N_b}{N_f + N_b}$ (N_f and N_b are the number of positive (negative) charged muons that fall in the forward and backward hemisphere respectively with respect to the B^0 (\bar{B}^0) direction in the dimuon reference frame), then fitted with a straight line.

asymmetry, i.e. $A_{FB} = \frac{N_f - N_b}{N_f + N_b}$, where N_f and N_b are simply the number of positive (negative) charged muons that fall in the forward and backward hemisphere respectively (with respect to the B^0 (\bar{B}^0) direction in the dimuon reference frame). The measurement is performed in several (100) q^2 bins, the result is then fit with a straight line as shown in Fig. 57, $A_{FB}^z = 4.26 \pm 0.07$. This result is compatible with the measurement performed on RECO-MC and shown in the first plot on the right-hand side of Fig. 56, $A_{FB}^z = 3.9 \pm 1.0$.

8.4 Validation with Data-Like Statistics Full Simulation

The analysis technique is validated with the simulation containing: signal, control channels, and background. The same number of events as in data, i.e. $\sim 20.5 \text{ fb}^{-1}$, is used. The goal is to verify whether in data-like conditions, and with the same number of events as in data, the analysis is able to measure the interesting observables.

The simulation is a mix of three datasets:

- Signal MC simulation $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^- (\sim 5951.1 \text{ fb}^{-1})$ divided into 400 independent sub-sets, each one having the same statistics as data (the splitting is performed taking into account the production cross-section difference between PYTHIA and data as reported in Sec. 5).
- Feed-through distributions generated from the MC simulation of both $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ ($\sim 52.5 \text{ fb}^{-1}$) and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ ($\sim 58.5 \text{ fb}^{-1}$) decays. The generation is performed with pseudo-experiments whose angular shape is the result of linear interpolations of the MC's distributions. The reason for a linear interpolation is to smooth out statistical fluctuations in order to rescale these samples to a different luminosity (the rescaling is performed taking into account the production cross-section difference between PYTHIA and data as reported in Sec. 5). The generation is performed 400 times, each one independent of each other.
- Background distribution (see Sec. 4.4 for the definition of background) generated with pseudo-experiments with p.d.f. parameters measured with data. The generation is performed 400 times, each one independent of each other.

In Fig. 58 are reported the average results of the 400 fits. The errors are simply the RMS of the measured distributions divided by the square root of the number of fits (only converged fits are considered, see definition of convergence in Sec. 4.6). The results of the fits on one dataset, out of the 400, are reported in App. D. It is worth noting that in some bins the measurements

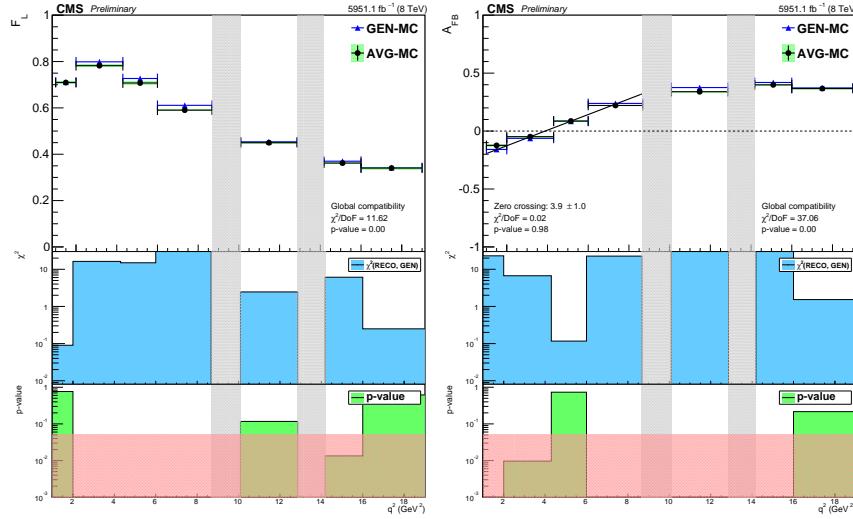


Figure 58: Validation of the analysis with the simulation containing: signal, control channels, and background. The number of events corresponds to $\sim 20.5 \text{ fb}^{-1}$. Comparison between the results of the analysis (blue) and the actual values used to generate the simulation (black) both for F_L (left) and A_{FB} (right). The first plot shows the values as a function of the dimuon q^2 . The second plot shows the χ^2 of compatibility for each bin, while the third plot shows the p-value (in red is shaded the 5% level of significance). The numbering goes from top to bottom. The errors are purely statistical. Grey shaded regions correspond to the J/ψ and ψ' resonances.

are biased, as shown in Fig.s 59 and 60 where the measured distributions are reported. By correlating the skewness of such distributions, especially for A_{FB} , with the position of the GEN-MC values in the $F_L - A_{FB}$ plane shown in Fig. 18, one can realize that the bias is caused by a two-fold effect:

- the presence of the boundaries in the $F_L - A_{FB}$ domain;
- the small number of events, which in turns corresponds to a small likelihood curvature at the minimum, which makes the minimization procedure more sensitive to the boundaries (in other words at low statistics the boundaries are “closer” to the likelihood minimum in terms of sigmas).

It is important to stress that such a bias is unavoidable with the given amount of data.

A thousand pseudo-experiments are also run with AVG-MC values as from Fig. 58. The average values of the pseudo-experiment fits disagree with the generated values as shown in Fig.s 61 and 62 on the left-hand side. The disagreement has to be therefore attributed to a bias in the analysis, which will be called **Fit Bias**.

The most clever way to deal with this bias is precisely by taking advantage of the pseudo-experiments to, in principle, correct for it. The basic idea is to run iteratively several pseudo-experiments searching for the particular F_L and A_{FB} starting values that allow to obtain average fit results within their uncertainty from the measured values with data, condition that defines the “convergence” of the method (the errors on the average fit results are simply the RMS of the measured distributions divided by the square root of the number of fits, where only converged fits are considered, see definition of convergence in Sec. 4.6).

The validation of this procedure with the data-like simulation is relatively trivial, in fact it’s already known which starting values should be set for F_L and A_{FB} , which are precisely those

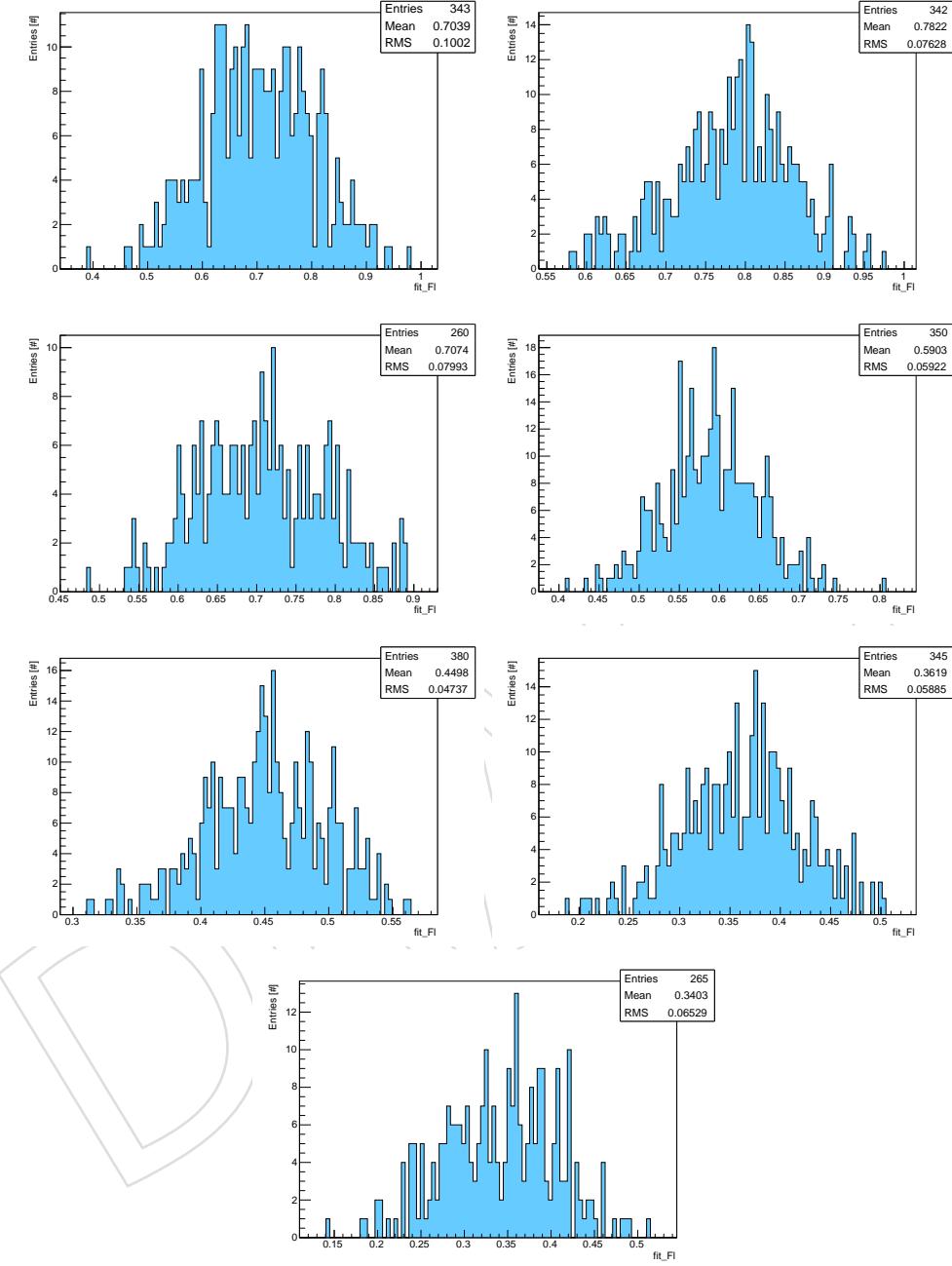


Figure 59: Distribution of the F_L results from the fit to the simulation containing: signal, control channels, and background. The number of events corresponds to $\sim 20.5 \text{ fb}^{-1}$. The number of entries corresponds to the converged fits only (see definition of convergence in Sec. 4.6) with errors greater than 0.01 and smaller than 1 to remove lone outliers related to obviously wrong fits.

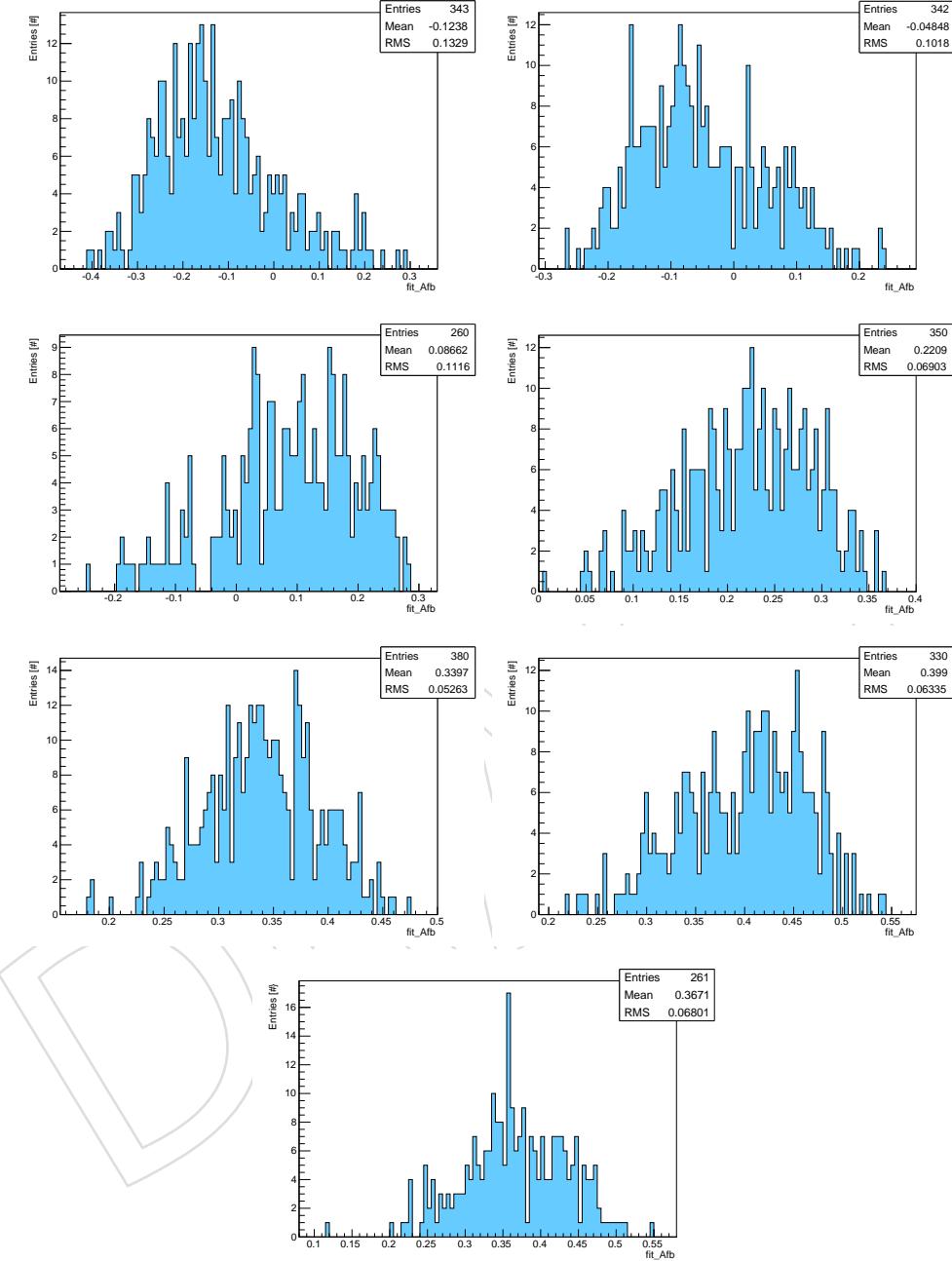


Figure 60: Distribution of the A_{FB} results from the fit to the simulation containing: signal, control channels, and background. The number of events corresponds to $\sim 20.5 \text{ fb}^{-1}$. The number of entries corresponds to the converged fits only (see definition of convergence in Sec. 4.6) with errors greater than 0.01 and smaller than 1 to remove lone outliers related to obviously wrong fits.

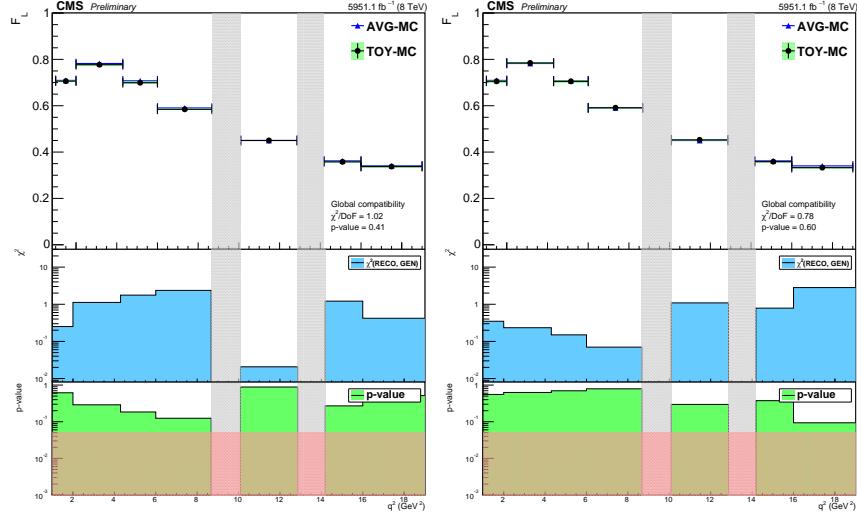


Figure 61: Validation of the procedure to correct for the F_L Fit Bias. Comparison between the measurements reported in Sec. 8.3 (AVG-MC blue) and the average results of the pseudo-experiments (TOY-MC black). The pseudo-experiments are run with AVG-MC values (left) and Sec. 8.3 values (right). The errors are computed as the RMS of the distributions divided by the square root of the number of fits (only converged fits are considered, see definition of convergence in Sec. 4.6).

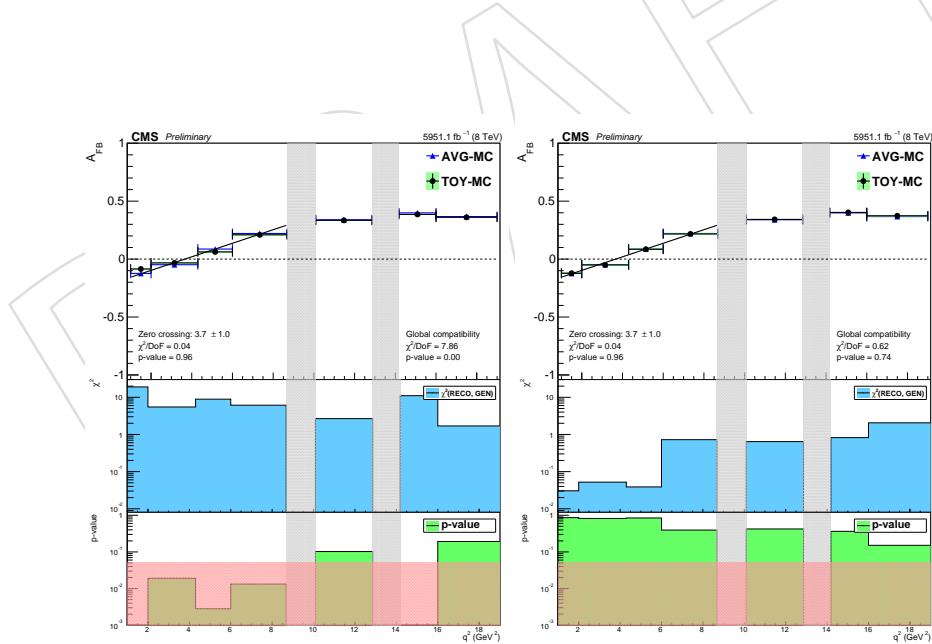


Figure 62: Validation of the procedure to correct for the A_{FB} Fit Bias. Comparison between the measurements reported in Sec. 8.3 (AVG-MC blue) and the average results of the pseudo-experiments (TOY-MC black). The pseudo-experiments are run with AVG-MC values (left) and Sec. 8.3 values (right). The errors are computed as the RMS of the distributions divided by the square root of the number of fits (only converged fits are considered, see definition of convergence in Sec. 4.6).

measured in Sec. 8.3. Indeed Fig.s 61 and 62 on the right-hand side show that the method works nicely. The systematic uncertainty that should be in principle attributed to this correction method is equal to the error with which the average pseudo-experiment fit results are determined. Finally, on data, it has been chosen to opt for simplicity and therefore to not apply any correction and to quote the **Fit Bias** as symmetric systematic uncertainty (see Sec.s 9 and 10).

8.4.1 Validation of the Error Computation

All aforementioned tests are independent from error treatment. The validation of the errors is, furthermore, not trivial due to non-linearities caused by the transformation of the angular variables, F_L and A_{FB} , described in Sec. 4.8. In fact since next to the boundaries the transformation is highly non-linear, and the average of the transformed values does not correspond to the transformed average value, in symbols: $\sum f(x_i) \neq f(\sum x_i)$. It is important to stress, though, that the final errors are computed in the rigorously correct way, which is by correctly propagating errors and correlations as described in Sec. 4.8, and by using MINOS [19] to properly take into account possible non-parabolic likelihood behaviour around the minimum. As a naive cross check of the correctness of the error computation, pulls are computed and reported in Fig. 63 for the q^2 bin #5, one of the furthest from the boundaries. The histograms are computed for both unbounded and bounded, i.e. physical, variables.

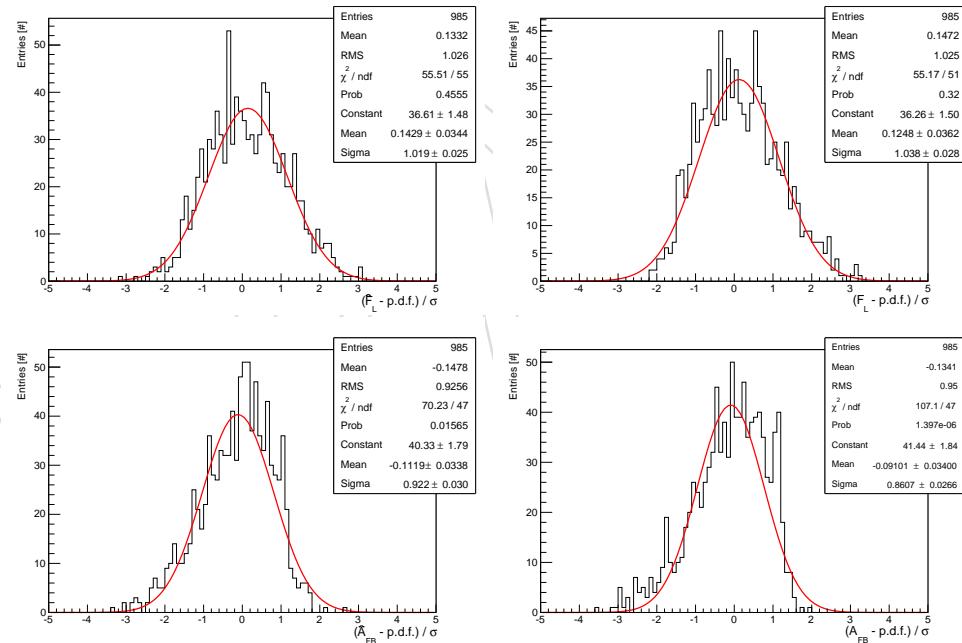


Figure 63: Pulls computed for the interesting angular observables, F_L and A_{FB} , with pseudo-experiments with starting values determined with data for the q^2 bin #5. For both angular observables both unbounded (hat, “^”, variables, left) and bounded (“un-hat” variables, right), i.e. physical, variables are presented.

8.5 Validation with Data

The analysis technique is validated with the data by means of the control channels $B^0 \rightarrow K^{*0}(K^+\pi^-)\bar{J}/\psi(\mu^+\mu^-)$ and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$, for which both the forward-backward asymmetry of the muons, and the fraction of longitudinal polarization of the $K^*(892)$, have

been measured and found in agreement with known SM predictions and previous experimental values. Together with F_L and A_{FB} also the fraction of S -wave and the S - P interference terms have been measured for each of the two control channels.

The results of the measurements for the control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ are:

- $F_L = 0.537 \pm 0.002$ (stat), to be compared with the PDG value 0.571 ± 0.007 and with the most precise measurement from BaBar [24] 0.556 ± 0.009 (stat) ± 0.010 (syst);
- $A_{FB} = 0.008 \pm 0.003$ (stat), to be compared with BaBar measurements [18];
- $F_S = 7 \times 10^{-9} \pm 0.002$ (stat), to be compared with BaBar measurements [18];
- $A_S = -0.102 \pm 0.003$ (stat), to be compared with BaBar measurements [18];

while for the control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ are:

- $F_L = 0.538 \pm 0.008$ (stat), to be compared with the PDG value $0.463^{+0.028}_{-0.040}$ and with the most precise measurement from BaBar [24] 0.48 ± 0.05 (stat) ± 0.02 (syst);
- $A_{FB} = 0.044 \pm 0.010$ (stat), to be compared with BaBar measurements [18];
- $F_S = 0.066 \pm 0.02$ (stat), to be compared with BaBar measurements [18];
- $A_S = -0.102 \pm 0.014$ (stat), to be compared with BaBar measurements [18].

To be noticed that F_S and A_S depend upon the width of the $K\pi$ mass acceptance range around the $K^*(892)$ PDG mass value, the larger the range the higher the S -wave contamination (the $K\pi$ mass range used in this analysis is defined in Sec. 3). The results of the fits are shown in Fig. 64, in which the projections on the B^0 invariant mass and on the angular observables are presented and superimposed to the data.

A systematic uncertainty related to the correct modeling of the efficiency shape, called **Efficiency Shape** in Sec. 10.2, is assigned to F_L , ± 0.034 , by comparing the results for the control channels $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ to the PDG values.

Finally the branching ratio $\mathcal{B}(B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)) / \mathcal{B}(B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-))$ is measured and compared to the PDG value. The computation is described by Eq. 8:

$$\frac{\mathcal{B}(B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-))}{\mathcal{B}(B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-))} = \left(\frac{Y_{J/\psi}^R}{\mathcal{E}_{J/\psi}^R} + \frac{Y_{J/\psi}^M}{\mathcal{E}_{J/\psi}^M} \right) \left(\frac{Y_{\psi'}^R}{\mathcal{E}_{\psi'}^R} + \frac{Y_{\psi'}^M}{\mathcal{E}_{\psi'}^M} \right)^{-1} \quad (8)$$

where $Y_{J/\psi}^R = 142209 \pm 440$, $Y_{J/\psi}^M = 22633 \pm 70$ ($Y_{\psi'}^R = 9256 \pm 117$, $Y_{\psi'}^M = 1507 \pm 19$) and $\mathcal{E}_{J/\psi}^R = 3.76 \times 10^{-3} \pm 4.44 \times 10^{-6}$, $\mathcal{E}_{J/\psi}^M = 5.66 \times 10^{-4} \pm 8.60 \times 10^{-7}$ ($\mathcal{E}_{\psi'}^R = 3.83 \times 10^{-3} \pm 1.77 \times 10^{-5}$, $\mathcal{E}_{\psi'}^M = 5.97 \times 10^{-4} \pm 4.11 \times 10^{-6}$) are the yield and the efficiency respectively of the control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ ($B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$). The measured branching ratio is 15.8 ± 0.3 , while the PDG reports is 15.6 ± 1.9 . As a conservative approach the highest relative uncertainty, i.e. the one on the PDG value, is considered as symmetric systematic uncertainty for the measurement of the differential branching fraction, i.e. 12.2% (called **Efficiency Shape** in Sec. 10.2).

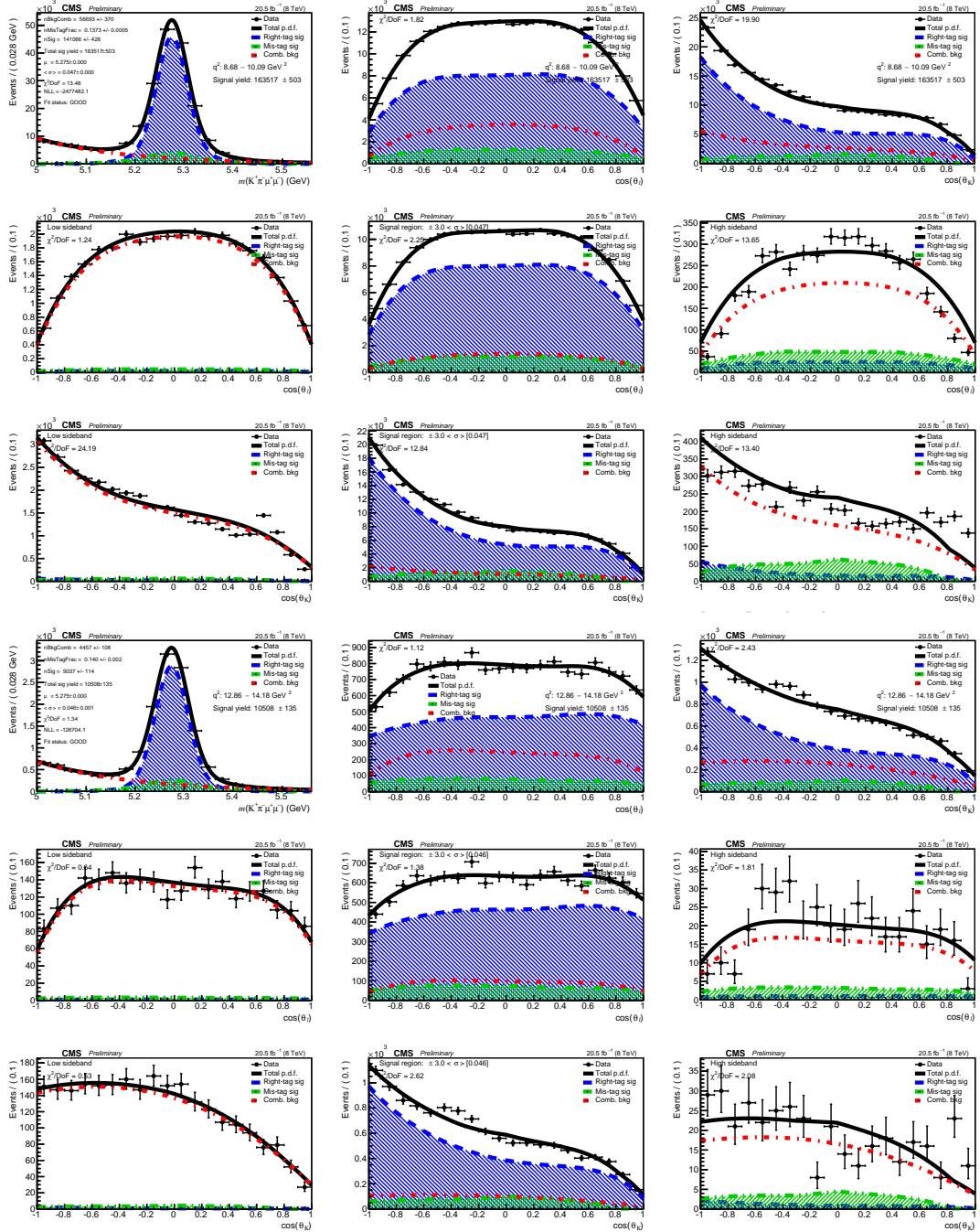


Figure 64: Fit results to measure F_L , A_{FB} , F_S , and A_S with data for the control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ (top nine plots) $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ (bottom nine plots). The different plots are, in order from left to right and from top to bottom: the projection on the B^0 invariant mass; the projection on the angular observable θ_l ; the projection on the angular observable θ_K ; the projection in the low B^0 mass sideband as a function of θ_l ; the projection in the high B^0 mass sideband as a function of θ_l ; the projection in the low B^0 mass sideband as a function of θ_K ; the projection in the high B^0 mass sideband as a function of θ_K ; for the definition of signal and sideband regions see Sec. 3.2. The degrees of freedom to compute the fit χ^2 are simply the number of bins.

9 Results

The results of the measurement of the fraction of longitudinal polarization of the $K^*(892)$, F_L , the forward-backward asymmetry of the muons, A_{FB} , and the differential branching fraction, $d\mathcal{B}/dq^2$, for the decay $B^0 \rightarrow K^{*0}\mu^+\mu^-$, are reported in Fig.s 65 and 66 and are tabulated in Tables 11, 12, and 13 together with the SM predictions. The signal and background yields are instead reported in Table 10 and the plots of the fits are reported in App. E.

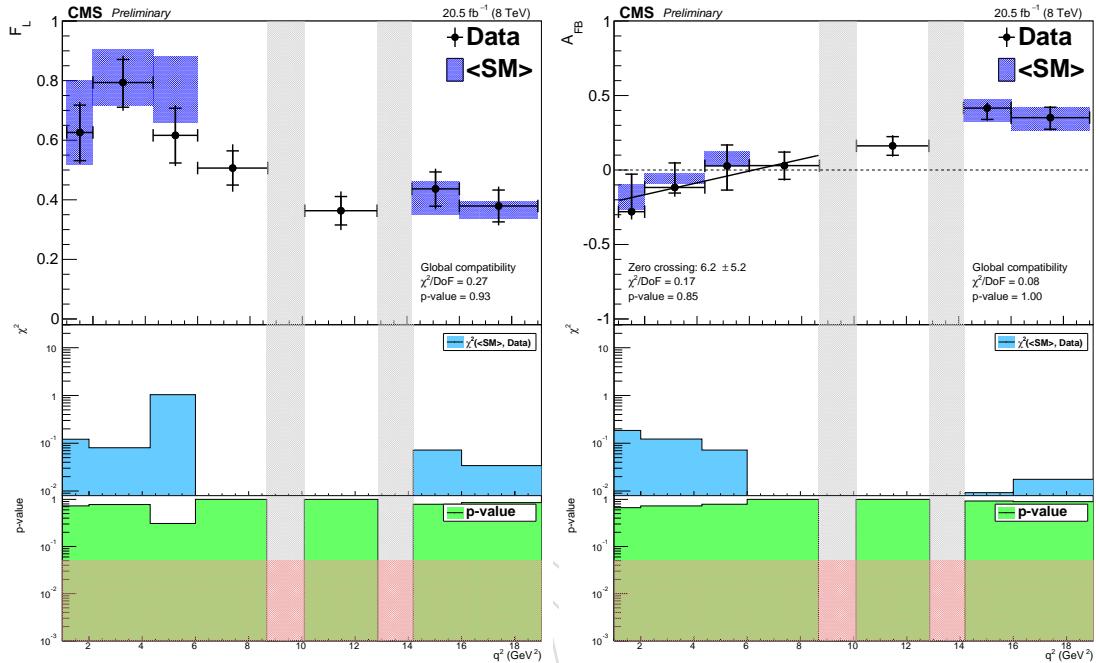


Figure 65: Results of the measurement of F_L (left three plots) and A_{FB} (right three plots). The first plot shows the values as a function of the dimuon q^2 . The black error bars with edges represent the statistical uncertainty while the edgless error bars represent both statistical and systematic uncertainties (only uncorrelated, see Sec. 10 for more details). The second plot shows the χ^2 of compatibility for each bin between data and SM predictions, while the third plot shows the p-value (in red is shaded the 5% level of significance). The numbering goes from top to bottom. Grey shaded regions correspond to the J/ψ and ψ' resonances. The SM predictions are rate-averaged across the q^2 bins as indicated by the blue regions (the form factors are computed from LCSR). No controlled SM prediction is available for the region between the two resonances and for the q^2 bin #3 (see text).

Signal candidates are observed in each q^2 bin as shown in the invariant mass plots projections in Sec. 4 and App. E. The measurements of all three quantities are in good agreement with SM predictions within the errors. The statistical uncertainties are computed with MINOS [19], while various possible sources of systematic uncertainties are thoroughly investigated and discussed in Sec. 10.

The predictions are made with the software from Ref. [6]. In the low q^2 region they rely on the factorization approach [25], which loses accuracy when approaching the J/ψ and ψ' resonances; in the high q^2 region, an operator product expansion in the inverse b-quark mass, $1/m_b$, and in $1/\sqrt{q^2}$ is used [26], which is only valid above the open charm threshold. Form factors are computed with Light Cone Sum Rules (LCSR) [5] at low q^2 and their extrapolation is used at high q^2 .

Uncertainties are computed by varying the most important parameters within prior distributions and by generating via a large sample the distributions for the observables. The smallest 68% intervals and the most probable values of these distributions are used as error-intervals and central values of the predictions. For the list of parameters and their prior distributions see Ref. [27]. Predictions for the q^2 bin #3 are not available within the model of QCD factorization.

Recent lattice results on $B \rightarrow K^*$ form factors show an impressive progress by the lattice community. In App. F are reported the results of the measurements compared with SM predictions using such lattice results [28]. These predictions show smaller uncertainties with respect to LCSR, however they are subject to the problem of a stable K^* on the lattice. The inherent systematic uncertainty of stable K^* should be further investigated before one relies on these results alone. On the other hand, the results without lattice input suffer from lack of any form factors at large q^2 giving rise to a large parametric uncertainty in the form factors.

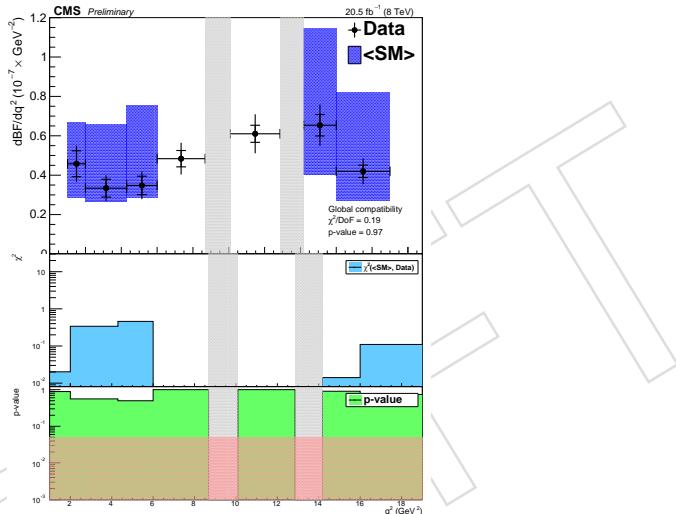


Figure 66: Result of the measurement of $d\mathcal{B}/dq^2$. The first plot shows the values as a function of the dimuon q^2 . The black error bars with edges represent the statistical uncertainty while the edgeless error bars represent both statistical and systematic uncertainties (only uncorrelated, see Sec. 10 for more details). The second plot shows the χ^2 of compatibility for each bin between data and SM predictions, while the third plot shows the p-value (in red is shaded the 5% level of significance). The numbering goes from top to bottom. Grey shaded regions correspond to the J/ψ and ψ' resonances. The SM predictions are rate-averaged across the q^2 bins as indicated by the blue regions (the form factors are computed from LCSR). No controlled SM prediction is available for the region between the two resonances and for the q^2 bin #3 (see text).

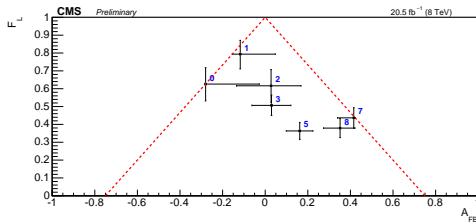
In Fig. 67 is reported the scatter plot of all measurements in the $F_L - A_{FB}$ plane. On the same figure it is also reported the region where the decay rate in Eq. 3 is always positive for every value of the angular variables. All measurements lie within the physically-allowed domain as guaranteed by construction of the p.d.f. (see Sec. 4.8 for more details).

With the measurements of A_{FB} in the first four q^2 bins the zero crossing point can be extracted (see Sec. 4.7) and compared to SM prediction and previous measurement: $A_{FB}^z = 6.2 \pm 4.3 (\text{GeV}/c^2)^2$, to be compared with SM prediction $A_{FB}^z = 4.0 \pm 0.2 (\text{GeV}/c^2)^2$ and LHCb measurement $4.9 \pm 0.9 (\text{GeV}/c^2)^2$ [11].

As stated in Sec. 1, the angular observables can be theoretically predicted with relatively good control of the relevant form factor uncertainties especially in the low dimuon invariant mass

Table 10: Yields for both signal and background. Errors are purely statistical.

q^2 bin	Y_S	Y_B
0	73.5 ± 9.6	91.2 ± 11.5
1	126.6 ± 14.2	288.7 ± 19.9
2	102.9 ± 12.7	216.2 ± 17.6
3	221.3 ± 17.8	343.1 ± 22.7
5	313.2 ± 21.4	567.2 ± 28.6
7	193.1 ± 16.1	177.6 ± 17.2
8	206.4 ± 15.2	81.8 ± 12.4

Figure 67: Scatter plot of all measurements, both from signal and control channels, in the $F_L - A_{FB}$ plane. The dashed lines mark the boundaries of the region where the decay rate is always positive for every value of the angular variables (see Sec. 4.8). The errors of the measurements are purely statistical.

region. It is therefore important to perform the measurements of the interesting observables also in the region $1 < q^2 < 6$ (GeV/c^2)². The results are here presented and compared with the most up to date SM predictions (both the plots of the fits and the description of the background p.d.f. are reported in App. E):

- $F_L = XXX^{+XXX}_{-XXX}$ (stat) $\pm XXX$ (syst, only uncorrelated, see Sec. 10.12);
- $A_{FB} = XXX^{+XXX}_{-XXX}$ (stat) $\pm XXX$ (syst, only uncorrelated, see Sec. 10.12);
- $d\mathcal{B}/dq^2 = XXX^{+XXX}_{-XXX}$ (stat) $\pm XXX$ (syst, only uncorrelated, see Sec. 10.12).

Table 11: Results of the measurement of the fraction of longitudinal polarization of the $K^*(892)$, F_L , together with the most up to date SM predictions. In the second column are reported the results of the fits, while in the third column are reported the **Fit Bias** systematic uncertainties, which are determined as described in Sec. 8.4. The first error is statistical while the second is systematic (only uncorrelated, see Sec. 10 for more details).

q^2 bin	F_L	Fit Bias	SM
0	$0.626^{+0.092}_{-0.095} \pm 0.045$	+0.000	$0.698^{+0.104}_{-0.179}$
1	$0.794^{+0.077}_{-0.083} \pm 0.046$	+0.016	$0.837^{+0.070}_{-0.122}$
2	$0.616^{+0.091}_{-0.093} \pm 0.045$	+0.000	$0.766^{+0.114}_{-0.106}$
3	$0.507^{+0.058}_{-0.057} \pm 0.055$	+0.000	N.A.
5	$0.363^{+0.048}_{-0.048} \pm 0.044$	-0.002	N.A.
7	$0.437^{+0.057}_{-0.058} \pm 0.037$	+0.005	$0.415^{+0.045}_{-0.062}$
8	$0.379^{+0.054}_{-0.053} \pm 0.037$	+0.007	$0.366^{+0.030}_{-0.030}$

¹⁰⁰¹ Most up to date theoretical predictions based on the SM for the region $1 < q^2 < 6$ (GeV/c^2)²:

- 1002 • $F_L = 0.729^{+0.077}_{-0.100}$;
- 1003 • $A_{FB} = 0.033^{+0.039}_{-0.027}$;
- 1004 • $d\mathcal{B}/dq^2 = (0.376^{+0.125}_{-0.100}) \times 10^{-7} c^4/\text{GeV}^2$;

Table 12: Results of the measurement of the forward-backward asymmetry of the muons, A_{FB} , together with the most up to date SM predictions. In the second column are reported the results of the fits, while in the third column are reported the **Fit Bias** systematic uncertainties, which are determined as described in Sec. 8.4. The first error is statistical while the second is systematic (only uncorrelated, see Sec. 10 for more details).

q^2 bin	A_{FB}	Fit Bias	SM
0	$-0.280^{+0.252}_{-0.001} \pm 0.052$	+0.000	$-0.160^{+0.103}_{-0.063}$
1	$-0.117^{+0.165}_{-0.038} \pm 0.049$	-0.024	$-5.562^{+3.767}_{-2.930} \times 10^{-2}$
2	$+0.028^{+0.141}_{-0.162} \pm 0.041$	+0.000	$-6.824^{+3.973}_{-5.463} \times 10^{-2}$
3	$+0.030^{+0.091}_{-0.092} \pm 0.040$	+0.000	N.A.
5	$+0.162^{+0.062}_{-0.063} \pm 0.036$	+0.007	N.A.
7	$+0.416^{+0.008}_{-0.076} \pm 0.036$	+0.003	$+0.405^{+0.080}_{-0.068}$
8	$+0.352^{+0.070}_{-0.078} \pm 0.035$	+0.000	$+0.336^{+0.071}_{-0.083}$

Table 13: Results of the measurement of the differential branching fraction $d\mathcal{B}/dq^2$, together with the most up to date SM predictions. The first error is statistical while the second is systematic (only uncorrelated, see Sec. 10 for more details).

q^2 bin	$d\mathcal{B}/dq^2 (\times 10^{-8} c^4/\text{GeV}^2)$	SM
0	$0.458 \pm 0.066 \pm 0.064$	$0.490^{+0.202}_{-0.179}$
1	$0.334 \pm 0.045 \pm 0.047$	$0.447^{+0.183}_{-0.209}$
2	$0.348 \pm 0.047 \pm 0.052$	$0.503^{+0.217}_{-0.253}$
3	$0.484 \pm 0.042 \pm 0.069$	N.A.
5	$0.610 \pm 0.043 \pm 0.089$	N.A.
7	$0.654 \pm 0.054 \pm 0.090$	$0.690^{+0.285}_{-0.455}$
8	$0.420 \pm 0.031 \pm 0.058$	$0.499^{+0.229}_{-0.320}$

1005 Just for completeness in Fig. 68 are reported the trend plots of the iterative **Fit Bias** method
 1006 described in Sec. 8.4. When the residual discrepancy between the average value of the 1000
 1007 pseudo-experiments is within their uncertainty from the measured values with data, the cor-
 1008 rection procedure is considered “converged”. The difference between the measurement and the
 1009 value for which the “convergence” is reached is conservatively considered as symmetric sys-
 1010 tematic uncertainty (called **Fit Bias** in Sec. 10.4). To those bins for which the convergence is not
 1011 reached, it is conservatively assigned a symmetric systematic uncertainty corresponding to the
 1012 difference between the measurement and the closest average value of the pseudo-experiments
 1013 (called **Residual Correction** in Sec. 10.5):

- 1014 • F_L **Correction** uncertainty bin #7: ± 0.004 ;
 1015 • A_{FB} **Correction** uncertainty bin #1: ± 0.019 ;
 1016 • A_{FB} **Correction** uncertainty bin #7: ± 0.007 .

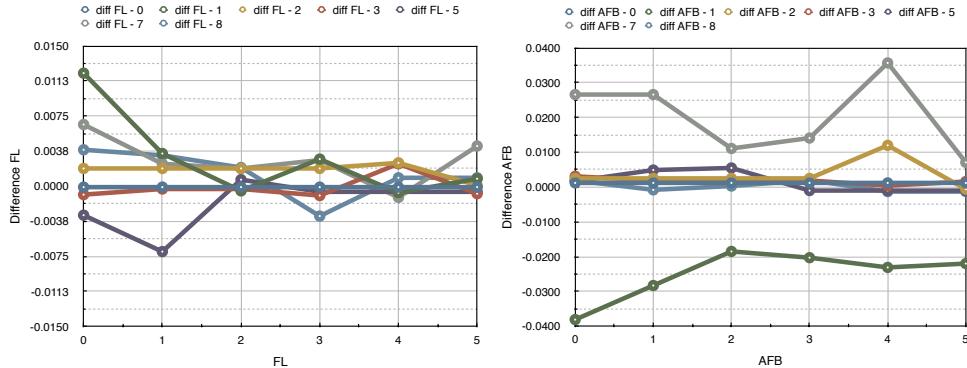


Figure 68: Trend plots of the **Fit Bias** correction. The abscissa represents the iteration step, while the ordinate of the left (right) plot corresponds to discrepancy on F_L (A_{FB}) between the average value of the 1000 pseudo-experiments with respect to the measured value with data.

1017 9.1 Pseudo-Experiments

1018 To unambiguously quantify the goodness of the fits, pseudo-experiments are generated with
 1019 p.d.f. parameters and number of events measured with data according to the fit-strategy de-
 1020 scribed in Sec. 4. The number of generated pseudo-experiments is 1000 (only converged fits are
 1021 considered, see definition of convergence in Sec. 4.6). For each pseudo-experiment the values
 1022 of the constrained variables are re-generated from the constraint distribution.

1023 The results of the fits in the $F_L - A_{FB}$ plane are reported in Fig. 70 and App. G, while in Fig. 69
 1024 are reported the p-values of the fits for the measurement of F_L , A_{FB} , and $d\mathcal{B}/dq^2$ determined
 1025 from the Negative Log-Likelihood distribution reported in App. G.

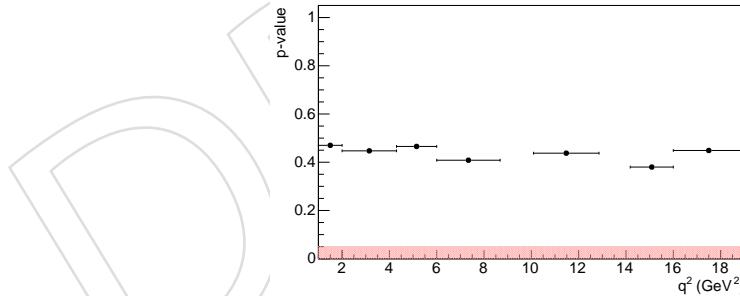


Figure 69: p-values as a function of the q^2 determined from the Negative Log-Likelihood (NLL) distributions obtained from pseudo-experiments run with p.d.f. parameters measured with data (in red is shaded the 5% level of significance). The p-values are calculated in correspondence of the NLL values determined with data and from the pseudo-experiments NLL distributions reported in App. G.

1026 9.2 Comparison with Other Experiments

1027 In this section a direct comparison of the CMS results with other experiments, namely BaBar [29,
 1028 Belle [9], CDF [10], LHCb [11], and Atlas [31], is performed for the fraction of longitudinal
 1029 polarization of the $K^*(892)$, F_L , the forward-backward asymmetry of the muons, A_{FB} , and the
 1030 differential branching fraction, $d\mathcal{B}/dq^2$, for the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$, for each q^2 bin.

1031 The comparison is shown in Fig. 71. The CMS uncertainties are better than any other experi-

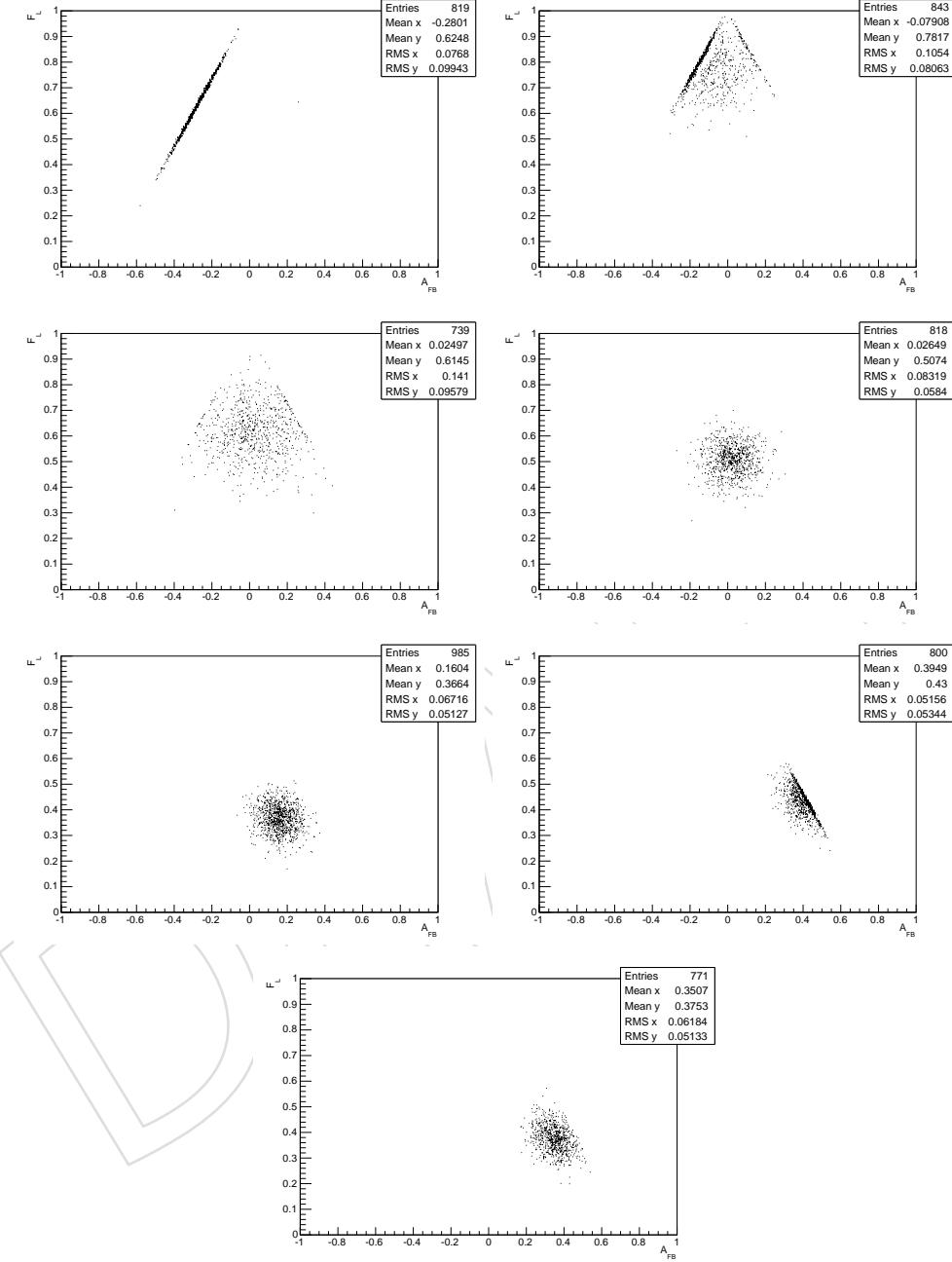


Figure 70: Scatter plots in the $F_L - A_{FB}$ plane of the pseudo-experiment results for each q^2 bin in Table 4 but #4 and #6. Only converged fits are considered, see definition of convergence in Sec. 4.6. The boundaries of the physically-allowed domain are clearly visible (see Sec. 4.8 for more details). The numbering goes from left to right and from top to bottom.

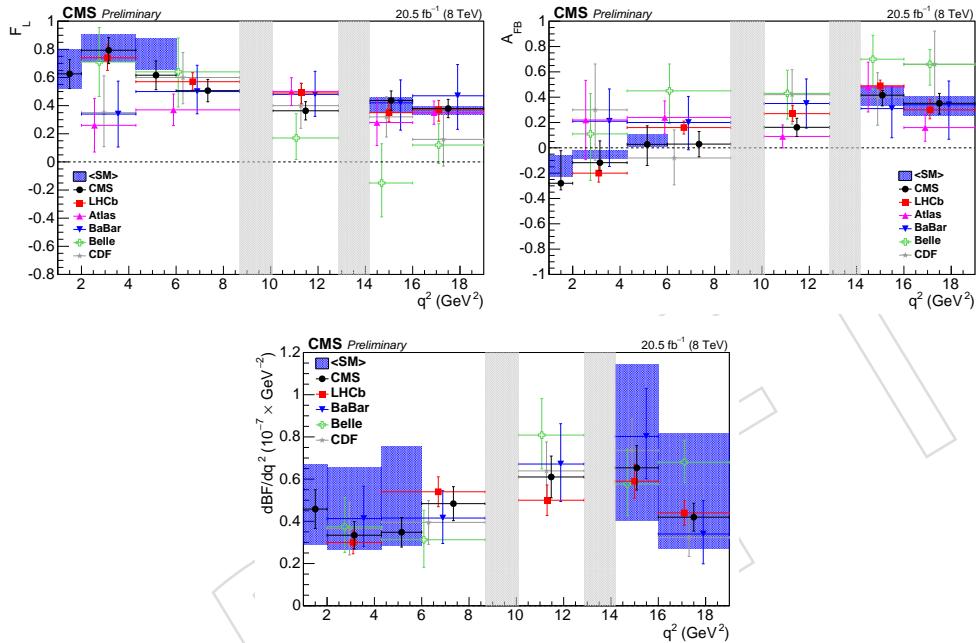


Figure 71: Comparison of the results of the measurement of F_L (first plot), A_{FB} (second plot), and $d\mathcal{B}/dq^2$ (third plot) among different experiments: BaBar [29, 30], Belle [9], CDF [10], LHCb [11], and ATLAS [31]. All errors include both statistical and systematic uncertainties. For each q^2 bin all data points, but for CMS and LHCb, are slightly shifted along the x -axis for readability purposes. The gray shaded regions correspond to the J/ψ and ψ' resonances. The SM prediction is given by the blue regions where it is rate-averaged across the q^2 bins to allow direct comparison to the data points. The numbering goes from left to right and from top to bottom.

₁₀₃₂ ment but LHCb (LHCb's analyzed data $\sim 1 \text{ fb}^{-1}$), therefore all in all the CMS measurements
₁₀₃₃ are the world second best.

DRAFT

1034 10 Systematic Uncertainties

1035 In this section are presented the main sources of systematic uncertainties. The different con-
 1036 tributions are divided into effects which are correlated and uncorrelated bin-to-bin. The cor-
 1037 related systematic uncertainties affect only the overall normalization, and not the shape of the
 1038 distributions. The contributions from the investigated sources of systematic uncertainties are
 1039 shown in Tables 14, 15, and 16. All contributions are then added in quadrature since the
 1040 hypothesis of independence is assumed to be reasonable, given the nature of the systematic
 1041 uncertainties involved and the way they are investigated.

Table 14: Systematic uncertainty contributions for each q^2 bin, as from Table 4, for the measurement of F_L . No correlated systematic uncertainties are accounted for the measurement of this observable. The bins #4 and #6 are intended for the measurement of the control channels $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$.

Syst. uncertainties \ q^2 bin	0	1	2	3	5	7	8
Uncorrelated systematic uncertainties							
limited num. sim. ev.	0.010	0.003	0.004	0.005	0.003	0.005	0.005
efficiency shape	0.034	0.034	0.034	0.034	0.034	0.034	0.034
sim. mismodeling	0.011	0.004	0.009	0.017	0.002	0.001	0.003
fit bias	0.0	0.016	0.0	0.0	0.002	0.005	0.007
residual correction	0.0	0.0	0.0	0.0	0.0	0.004	0.0
wrong CP-state	0.003	0.004	0.004	0.003	0.002	0.0	0.0
$S - P$ wave interf.	0.009	0.009	0.015	0.023	0.026	0.008	0.005
bkg. mismodeling	0.006	0.004	0.009	0.006	0.004	0.001	0.008
MC p.d.f. components	0.003	0.003	0.003	0.003	0.003	0.003	0.003
feed-through bkg.	0.0	0.0	0.0	0.027	0.0	0.0	0.0
angular resolution	0.023	0.024	0.021	0.014	0.008	0.006	0.006
Total	0.045	0.046	0.045	0.055	0.044	0.037	0.037

1042 The possible sources of systematic uncertainties investigated are:

1043 **limited number of simulated events:** the number of events of the simulation directly affects
 1044 the accuracy with which the efficiency is determined;

1045 **efficiency shape:** the shape of the efficiency is validated by measuring the interesting observ-
 1046 ables with data control channels;

1047 **simulation mismodeling:** with pure-signal simulation having a high number of events it is
 1048 validated the capability of the analysis to measure the interesting observables;

1049 **fit bias:** the presence of boundaries for the interesting observables together with the small
 1050 number of events in data makes the likelihood prone to biases;

1051 **residual correction:** the procedure to for the fit bias evaluation fails for some bins, for which a
 1052 residual correction is therefore needed and considered;

1053 **wrong CP-state assignment:** the wrong assignment of the CP-state has a dilution effect on the
 1054 measurement of the angular observables;

1055 **$S - P$ wave interference:** the $K - \pi$ in the final state can be either generated from the vector
 1056 meson $K^*(892)$ decay, they would therefore be in a P -wave state, or not generated from

Table 15: Systematic uncertainty contributions for each q^2 bin, as from Table 4, for the measurement of A_{FB} . No correlated systematic uncertainties are accounted for the measurement of this observable. The bins #4 and #6 are intended for the measurement of the control channels $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$.

Syst. uncertainties \ q^2 bin	0	1	2	3	5	7	8
Uncorrelated systematic uncertainties							
limited num. sim. ev.	0.007	0.007	0.008	0.018	0.005	0.005	0.005
efficiency shape	0.034	0.034	0.034	0.034	0.034	0.034	0.034
sim. mismodeling	0.037	0.015	0.020	0.008	0.003	0.0	0.0
fit bias	0.0	0.024	0.0	0.0	0.007	0.003	0.0
residual correction	0.0	0.019	0.0	0.0	0.0	0.007	0.0
wrong CP-state	0.002	0.002	0.002	0.002	0.0	0.003	0.006
$S - P$ wave interf.	0.006	0.007	0.005	0.002	0.004	0.004	0.002
bkg. mismodeling	0.004	0.003	0.005	0.0	0.006	0.003	0.003
MC p.d.f. components	0.001	0.001	0.001	0.001	0.001	0.001	0.001
feed-through bkg.	0.0	0.0	0.0	0.005	0.003	0.0	0.0
angular resolution	0.005	0.002	0.001	0.0	0.002	0.001	0.0
Total	0.052	0.049	0.041	0.040	0.036	0.036	0.035

such a resonance, they would be therefore in an S -wave state. The two final states have different angular distributions;

background mismodeling: the particular functional form chosen to fit the background events might affect the final results of the analysis;

MC derived p.d.f. components: some components of the p.d.f. are derived from MC, which might not describe the behaviour of the real detector with enough accuracy, affecting, in this way, the fit results;

feed-through background: residual feed-through events, which are not explicitly accounted for in the p.d.f., might affect the final results of the analysis;

angular resolution: the angular variables have finite resolution which might affect the final results of the analysis.

10.1 Limited Number of Simulated Events

The efficiency correction is entirely derived from the simulation as described in Sec. 7. The number of events of the simulation directly affects the accuracy with which the efficiency is determined. To propagate the uncertainty of the efficiency for correctly tagged events to the interesting observables, the following procedure is adopted:

1. during the last step of the determination of the efficiency function the full covariance matrix (Σ) of the parameters describing the efficiency is computed;
2. N new efficiency functions are generated, starting from the original one, with a multivariate normal distribution ($f(\mathbf{x})$) using the full covariance matrix from step 1. as in Eq 9;
3. the analysis is run N -times on data, each time in association with a new efficiency function; the RMS of the N -measurements of the interesting observables is the systematic uncertainty associated with the limited number of simulated events.

Table 16: Systematic uncertainty contributions in % for each q^2 bin, as from Table 4, for the measurement of $d\mathcal{B}/dq^2$. The bins #4 and #6 are intended for the measurement of the control channels $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$. The correlated systematic uncertainty of 4.6% arises from the uncertainty on the branching fraction of the normalization mode ($B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$) as reported in the PDG.

Syst. uncertainties \ q^2 bin	0	1	2	3	5	7	8
Uncorrelated systematic uncertainties							
limited num. sim. ev.	2.0	1.1	0.8	1.0	0.7	0.7	0.5
efficiency shape	12.2	12.2	12.2	12.2	12.2	12.2	12.2
sim. mismodeling	2.5	1.9	5.5	3.1	3.5	2.0	1.0
wrong CP-state	2.6	3.1	3.1	2.9	2.4	3.0	3.1
$S - P$ wave interf.	4.2	4.2	4.2	4.2	4.2	4.2	4.2
bkg. mismodeling	0.4	0.2	1.2	0.5	0.0	0.1	0.3
MC p.d.f. components	3.2	3.2	3.2	3.2	3.2	3.2	3.2
feed-through bkg.	0.0	0.0	0.0	1.9	4.0	0.0	0.0
angular resolution	1.5	2.1	1.4	0.7	0.4	0.2	0.2
Uncorrelated total	14.0	14.0	14.9	14.3	14.5	13.8	13.7
Correlated systematic uncertainties							
$\mathcal{B}(B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-))$	4.6	4.6	4.6	4.6	4.6	4.6	4.6
Correlated total	4.6	4.6	4.6	4.6	4.6	4.6	4.6
Total	14.7	14.7	15.6	15.0	15.2	14.6	14.5

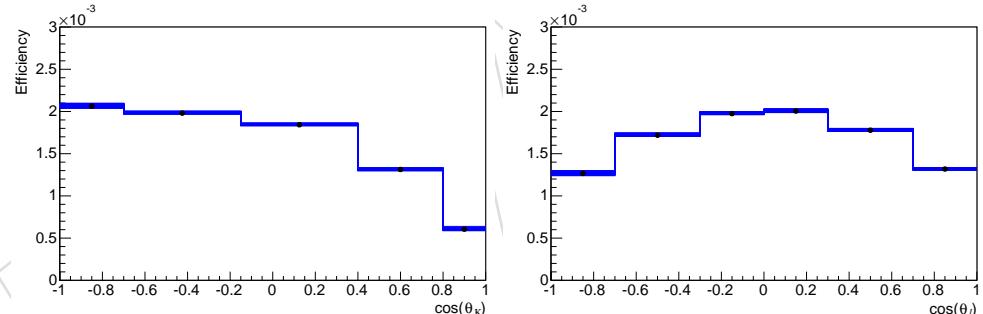


Figure 72: 100 superimposed efficiency distributions generated from the original one (black dots) with a multivariate normal distribution. They are integrated over all q^2 bins and projected on θ_K (left) and θ_l (right). The generated distributions are used to determine the systematic uncertainty associated with the limited number of simulated events.

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right), \quad (9)$$

1080 where $k = \text{rank}(\Sigma)$.

1081 The efficiency generation step is performed with MATLAB due to its simpler and faster imple-
 1082 mentation of the multivariate normal distribution function (only positively defined efficiency
 1083 functions are retained during the generation process). In Fig. 72 are presented all the N effi-
 1084 ciency functions integrated over all q^2 bins and projected on each angular variable, while in
 1085 Fig. 73 are reported the histograms of the differences between the original efficiency and the
 1086 efficiencies generated with the multivariate normal distribution.

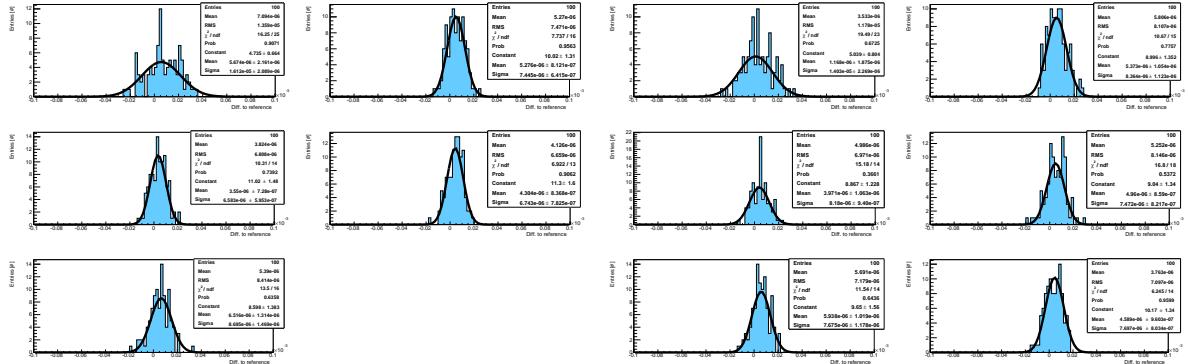


Figure 73: Histograms of the differences between the original efficiency and the efficiencies generated with the multivariate normal distribution for each bin of the angular variables, namely θ_l (five left) and θ_K (six right), superimposed to a Gaussian fit.

1087 To, instead, propagate the uncertainty of the efficiency for wrongly tagged events to the interesting observables, N new binned efficiency distributions are generated from the original one.
1088 The denominator is extracted from a Poisson distribution while the numerator is extracted from
1089 a Gaussian distribution, in both cases the momenta of the distributions are derived from the
1090 original efficiency. The bins are considered independent.
1091

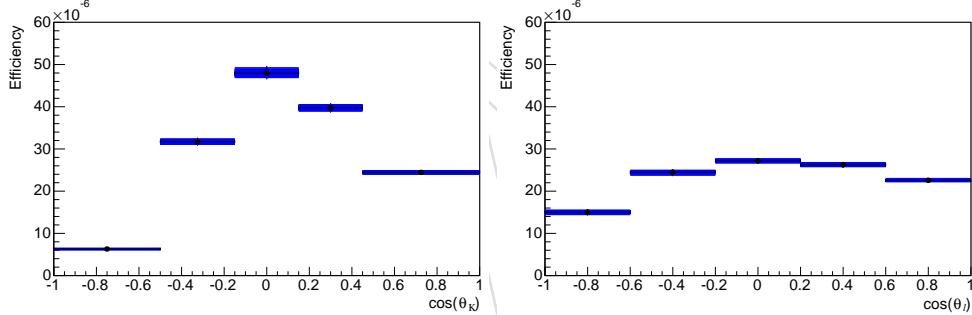


Figure 74: 100 superimposed efficiency distributions generated from the original one (black dots) with Poisson and Gauss distributions for the denominator and numerator respectively. They are integrated over all q^2 bins and projected on θ_K (left) and θ_l (right).

1092 In Fig. 74 are presented all the N efficiency distributions integrated over all q^2 bins and projected on each angular variable, while in Fig. 75 are reported the histograms of the differences
1093 between the original and the generated efficiency distributions.
1094

10.2 Efficiency Shape

1096 The possible discrepancy between the efficiency computed with the simulation and the real experimental efficiency is tested by comparing the measurements of known observables with previous measured values from other experiments and reported on the PDG, as shown in Sec. 8.5.
1097 The discrepancy in the measurements are considered as symmetric systematic uncertainty and applied to all bins.
1098

1099 Furthermore the branching ratio $\mathcal{B}(B^0 \rightarrow K^{*0}(K^+\pi^-)\psi(\mu^+\mu^-)) / \mathcal{B}(B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-))$ is also measured and compared to the PDG value, as shown in Sec. 8.5. Even if the 1100 two numbers are compatible within the errors, as a conservative approach the highest relative

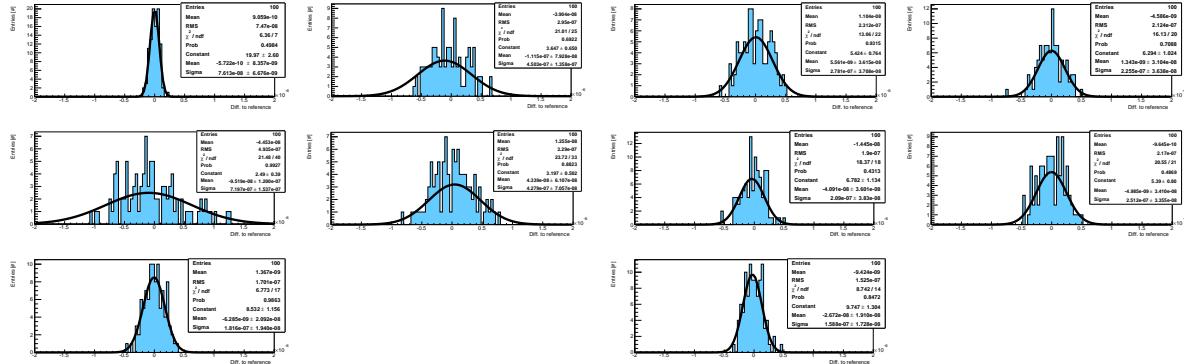


Figure 75: Histograms of the differences between the original efficiency and the efficiencies generated with Poisson and Gauss distributions for the denominator and numerator respectively. Each histogram corresponds to the bins of the angular variables, namely θ_l (five left) and θ_K (six right), superimposed to a Gaussian fit. The association plot-bin number goes from left to right and from top to bottom for each of the two sets of plots.

1104 uncertainty, i.e. the one on the PDG value, is considered as symmetric systematic uncertainty
 1105 for the measurement of the differential branching fraction, and it is applied to all bins.

1106 10.3 Simulation Mismodeling

1107 The simulation mismodeling is measured as the capability of the analysis to extract the inter-
 1108 esting observables in extremely favorable conditions, namely with signal simulation having
 1109 a high number of events. All discrepancies between the results of the analysis and the GEN-
 1110 MC values are conservatively considered as symmetric systematic uncertainties, as thoroughly
 1111 detailed in Sec. 8.3.

1112 10.4 Fit Bias

1113 The physically-allowed domain of the interesting observables is bounded. The presence of
 1114 such boundaries, together with the small number of events in data, generates biases in the
 1115 measurements as described in Sec. 8.4. In the same section it is detailed a procedure to some-
 1116 how determine the magnitude of such biases which are conservatively considered as symmetric
 1117 systematic uncertainties.

1118 10.5 Residual Correction

1119 The fit bias systematic uncertainty evaluation described in Sec. 10.4 fails for some bins. For
 1120 such bins a residual systematic uncertainty is therefore considered, as thoroughly detailed in
 1121 Sec. 9.

1122 10.6 Wrong CP-State Assignment

1123 The criteria adopted to assign the CP-state to the B^0 is described in Sec. 3.3. A wrong assign-
 1124 ment has a two-fold effect: the B^0 mass spectrum becomes wider and the measurement of the
 1125 angular observables, namely F_L and A_{FB} , becomes diluted. In Sec.s 3.3 and 4.2 it is shown
 1126 that the CP-tagging algorithm has an intrinsic mistag fraction measured on signal sample sim-
 1127 ulation. As a cross check of whether the mistag estimated from simulation well matches the
 1128 mistag on data, the total width of the B^0 measured with the control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)$

1129 $J/\psi(\mu^+\mu^-)$ is compared between the fit to the simulation and to the data with mass shape fully
 1130 free to float alternatively for correctly and wrongly tagged events.

1131 **A** simulation: $\langle\sigma\rangle = 0.0466 \pm 0.0002$;

1132 **B** data with floating correctly tagged events mass shape: $\langle\sigma\rangle = 0.0464 \pm 0.0006$;

1133 **C** data with floating wrongly tagged events mass shape: $\langle\sigma\rangle = 0.0444 \pm 0.0003$;

1134 where $\langle\sigma\rangle$ refers to the average width, between correctly and wrongly tagged events, of the
 1135 signal mass shape. The discrepancy between **A** and **C** is $\sim 4.7\%$.

1136 Another cross-check performed with data is by fitting the distributions of the control channel
 1137 $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ in the way described in Sec. 4, but simply adding the possibil-
 1138 ity to let the mistag fraction fully free to float without constraints. The final mistag fraction
 1139 measured on data is $f^M = 0.145 \pm 0.005$, to be compared with the mistag fraction obtained
 1140 with the simulation, $f^M = 0.137 \pm 0.001$. The difference is $\sim 5.8\%$, which corresponds to the
 1141 highest relative error, higher than the errors on the estimated mistag fractions in Table 5. For
 1142 the sake of completeness are also reported the values of the angular variables and the yield,
 1143 to be compared with the values reported in Sec. 8.5: $F_L = 0.540^{+0.003}_{-0.003}$, $A_{FB} = 0.008^{+0.003}_{-0.003}$, and
 1144 $Y_{J/\psi}^R = 140024 \pm 793$.

1145 In order to propagate the uncertainty of the mistag fraction the following procedure is adopted:

- 1146 1. the CP-mistag fraction is determined with an uncertainty, as described in Sec. 4.2;
- 1147 2. N CP-mistag values are generated with a Gaussian distribution centered on the nominal
 1148 CP-mistag value from step 1. and with a sigma corresponding to 5.8% of the mean value;
- 1149 3. the analysis is run N -times on data, each time in association with a new CP-mistag value;
 1150 the RMS of the N -measurements of the interesting observables is the systematic uncer-
 1151 tainty associated with the unknown true CP-mistag value.

1152 10.7 *S – P* Wave Interference

1153 Together with the decays of interest, $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$, $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$,
 1154 and $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$, for which the $K - \pi$ are generated from the vector meson
 1155 $K^*(892)$, exist also counterpart decays for which the $K - \pi$ are not generated from such a reso-
 1156 nance. The two cases differ with each other by the angular distribution of the decay products,
 1157 in the former the $K - \pi$ are in a *P*-wave state, while in the latter the $K - \pi$ are in an *S*-wave
 1158 state. To perform the measurements that are the object of this analysis, it is therefore in prin-
 1159 ciple important to disentangle *P*-wave from *S*-wave decays. On the other hand it is known
 1160 that the *S*-wave contribution is small with respect to the *P* one [16, 18], and its effect can be
 1161 neglected with the current amount of data, indeed its contribution is estimated to be at most
 1162 $\sim 4.2\%$ with respect to the $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ channel, as described in Sec. 6. The
 1163 adopted approach to deal with the *S*-wave component is to estimate it from data with the
 1164 $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ control channel and to use such an estimate to fit the signal chan-
 1165 nel.

1166 In order to evaluate the systematic uncertainty associated with this approach, the data are
 1167 fit with the *P*-wave only p.d.f. and the differences of the values of F_L and A_{FB} with respect

1168 to the standard analysis procedure, described in Sec. 4, are used as asymmetric systematic
 1169 uncertainties.

1170 Also the measurement of the differential branching fraction, $d\mathcal{B}/dq^2$, can be affected by the
 1171 $S - P$ wave interference because the actual number of events peaking at the B^0 mass includes
 1172 both kind of $K - \pi$ decays. The systematic uncertainty in this case is already estimated to be
 1173 4.2% as described in Sec.s 5 and 6, and it is applied to all bins.

1174 10.8 Background Mismodeling

1175 In order to evaluate if a systematic uncertainty is introduced by the particular angular shape
 1176 chosen to describe the background, data are fit with the degree of the polynomial increased
 1177 by one unit with respect to the nominal p.d.f. The values of F_L , A_{FB} , and $d\mathcal{B}/dq^2$ are com-
 1178 pared with the results obtained with the nominal p.d.f. and the systematic uncertainties are
 1179 evaluated as the difference on the measurements of the interesting observables and they are
 1180 conservatively considered symmetric.

1181 10.9 MC Derived p.d.f. Components

1182 Some components of the p.d.f. are derived from simulation as detailed in Sec. 4. The simulation
 1183 might not describe the behaviour of the real detector with enough accuracy, affecting, in this
 1184 way, the fit results. The components of the p.d.f. that are derived from simulation are: the
 1185 signal mass shape, but its mean value; the CP-mistag fraction, which is specifically considered
 1186 in Sec. 10.6. In Tables 17 and 18 are reported the values of the signal mass shape measured with
 1187 the simulation and with data. For the q^2 signal bins the fits are performed with constrained
 1188 mass shape parameters, while for the control channel bins, the mass shape is fully free to float
 1189 (alternatively for correctly and wrongly tagged events).

Table 17: Signal mass shape parameters comparison for correctly tagged between simulation and data measurements. The results in the columns “Data” for the signal bins, as from Table 4, are computed with constraints as described in Sec. 4.6, while for the control channel bins they are computed without constraints.

q^2 bin	Simulation				Data			
	index	μ	σ -1	σ -2	frac.	μ	σ -1	σ -2
0	5.267	0.030	0.092	0.832	5.267	0.030	0.092	0.832
1	5.274	0.031	0.098	0.844	5.274	0.031	0.098	0.844
2	5.276	0.031	0.104	0.858	5.276	0.031	0.104	0.858
3	5.276	0.030	0.089	0.856	5.276	0.030	0.089	0.856
4	5.278	0.030	0.064	0.829	5.275	0.027	0.051	0.558
	± 0.0001	± 0.0001	± 0.0004	± 0.004	± 0.0001	± 0.0004	± 0.001	± 0.023
5	5.275	0.031	0.094	0.872	5.275	0.031	0.094	0.872
6	5.278	0.033	0.062	0.910	5.275	0.029	0.054	0.637
	± 0.0002	± 0.0004	± 0.003	± 0.019	± 0.0004	± 0.002	± 0.005	± 0.095
7	5.276	0.032	0.095	0.863	5.276	0.032	0.095	0.863
8	5.280	0.032	0.077	0.828	5.280	0.032	0.077	0.828

1190 As a final check of the simulation accuracy, the values of the interesting observables are mea-
 1191 sured on data on both control channels, at first with signal mass shape for correctly tagged
 1192 events free to float:

Table 18: Signal mass shape parameters comparison for wrongly tagged between simulation and data measurements. The results in the columns “Data” for the signal bins, as from Table 4, are computed with constraints as described in Sec. 4.6, while for the control channel bins they are computed without constraints.

q^2 bin	Simulation				Data			
	index	μ	$\sigma\text{-}1$	$\sigma\text{-}2$	frac.	μ	$\sigma\text{-}1$	$\sigma\text{-}2$
0	5.267	0.108	0.046	0.458	5.267	0.108	0.046	0.456
1	5.274	0.117	0.048	0.408	5.274	0.117	0.048	0.407
2	5.276	0.111	0.047	0.431	5.276	0.111	0.047	0.433
3	5.276	0.098	0.041	0.474	5.276	0.098	0.041	0.474
4	5.278 ± 0.0001	0.103 ± 0.001	0.045 ± 0.001	0.447 ± 0.012	5.275 ± 0.0001	0.092 ± 0.003	- -	- -
5	5.275	0.110	0.047	0.346	5.275	0.110	0.047	0.347
6	5.278 ± 0.0002	0.105 ± 0.005	0.047 ± 0.003	0.461 ± 0.059	5.275 ± 0.0004	0.070 ± 0.046	0.069 ± 0.049	0.607 ± 0.527
7	5.276	0.108	0.047	0.496	5.276	0.108	0.047	0.496
8	5.280	0.102	0.045	0.569	5.280	0.102	0.045	0.571

- $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$: $F_L = 0.537^{+0.002}_{-0.002}$, $A_{FB} = 0.008^{+0.003}_{-0.003}$, $Y_{J/\psi}^R = 141043 \pm 443$;

- $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$: $F_L = 0.538^{+0.008}_{-0.008}$, $A_{FB} = 0.045^{+0.010}_{-0.010}$, $Y_{\psi'}^R = 9095 \pm 121$;

and then with signal mass shape for wrongly tagged events free to float:

- $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$: $F_L = 0.534^{+0.002}_{-0.002}$, $A_{FB} = 0.007^{+0.002}_{-0.002}$, $Y_{J/\psi}^R = 140069 \pm 443$;

- $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$: $F_L = 0.539^{+0.009}_{-0.009}$, $A_{FB} = 0.043^{+0.010}_{-0.010}$, $Y_{\psi'}^R = 8956 \pm 124$.

The highest discrepancies between these results and those reported in Sec. 8.5 are conservatively considered as symmetric systematic uncertainties and applied to all bins.

10.10 Feed-Through Background

The residual feed-through events, as in Fig. 9, which are not explicitly accounted for in the p.d.f., might affect the final results of the analysis. To the extent of evaluating the potential bias, the residual feed-through events, evaluated with $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ simulation in q^2 bins #3 and #5, are fit with a special p.d.f. component as shown in Fig. 76. Data are then fit adding such a component to the standard p.d.f. described in Sec. 4. The angular parameters of the new p.d.f. component are fixed, while all other parameters are free to float with Gaussian constraints. The results of the fits are reported in Fig. 77. Discrepancies on the interesting observables are conservatively considered as symmetric systematic uncertainties.

10.11 Angular Resolution

The angular variables, namely θ_K and θ_l , have finite resolution which might affect the fit results. To determine the impact on the analysis the measurements of the interesting observables are compared between using reconstructed and generated angular values on reconstructed signal sample simulation. The discrepancies are conservatively considered as symmetric systematic uncertainties.

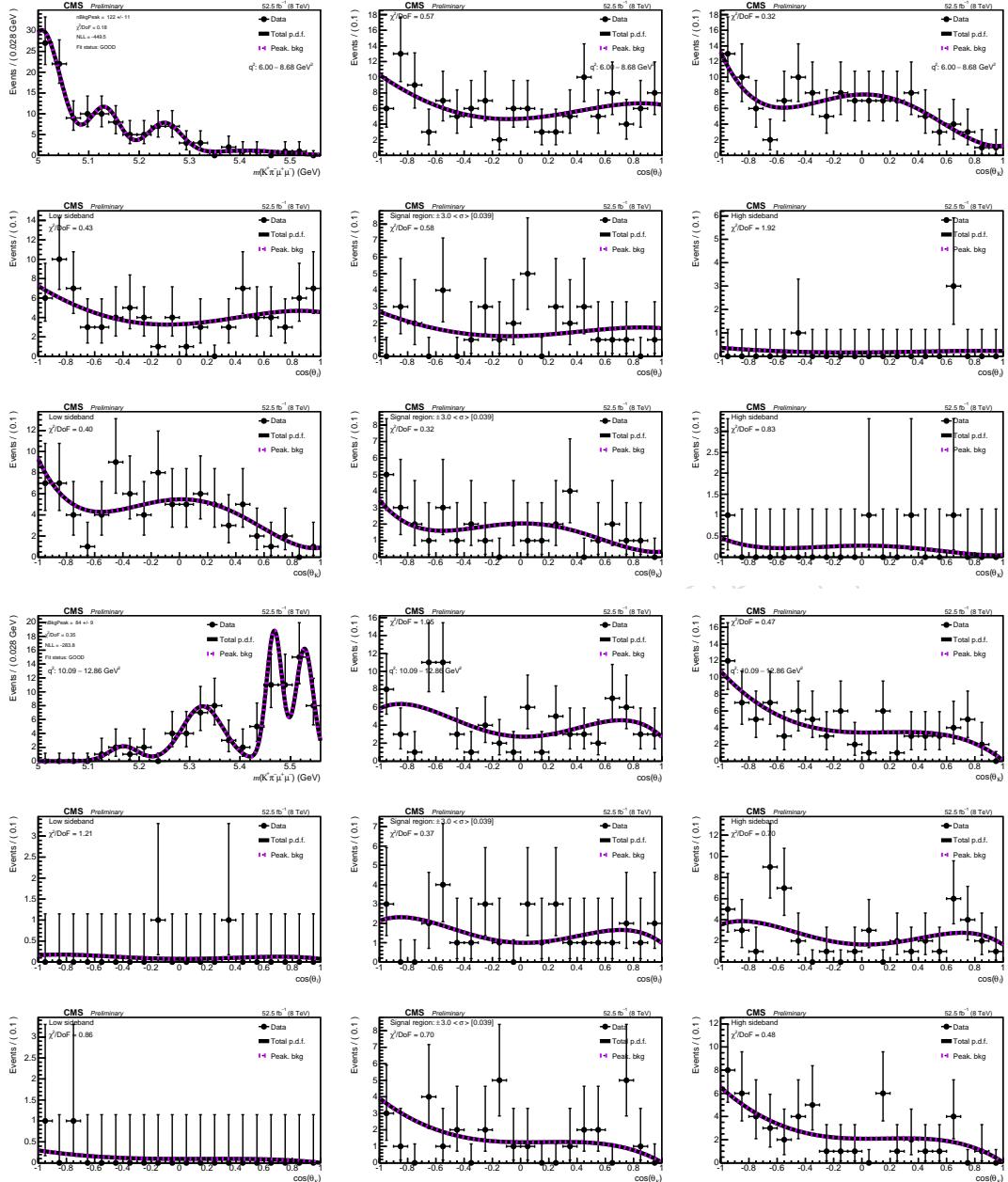


Figure 76: Fit results to $B^0 \rightarrow K^{*0}(K^+\pi^-)J/\psi(\mu^+\mu^-)$ simulation sample on q^2 bins #3 (top nine plots) and #5 (bottom nine plots). The different plots are, in order from left to right and from top to bottom: the projection on the B^0 invariant mass; the projection on the angular observable θ_l ; the projection on the angular observable θ_K ; the projection in the low B^0 mass sideband as a function of θ_l ; the projection in the signal region as a function of θ_l ; the projection in the high B^0 mass sideband as a function of θ_l ; the projection in the low B^0 mass sideband as a function of θ_K ; the projection in the signal region as a function of θ_K ; the projection in the high B^0 mass sideband as a function of θ_K ; for the definition of signal and sideband regions see Sec. 3.2. The degrees of freedom to compute the fit χ^2 are simply the number of bins.

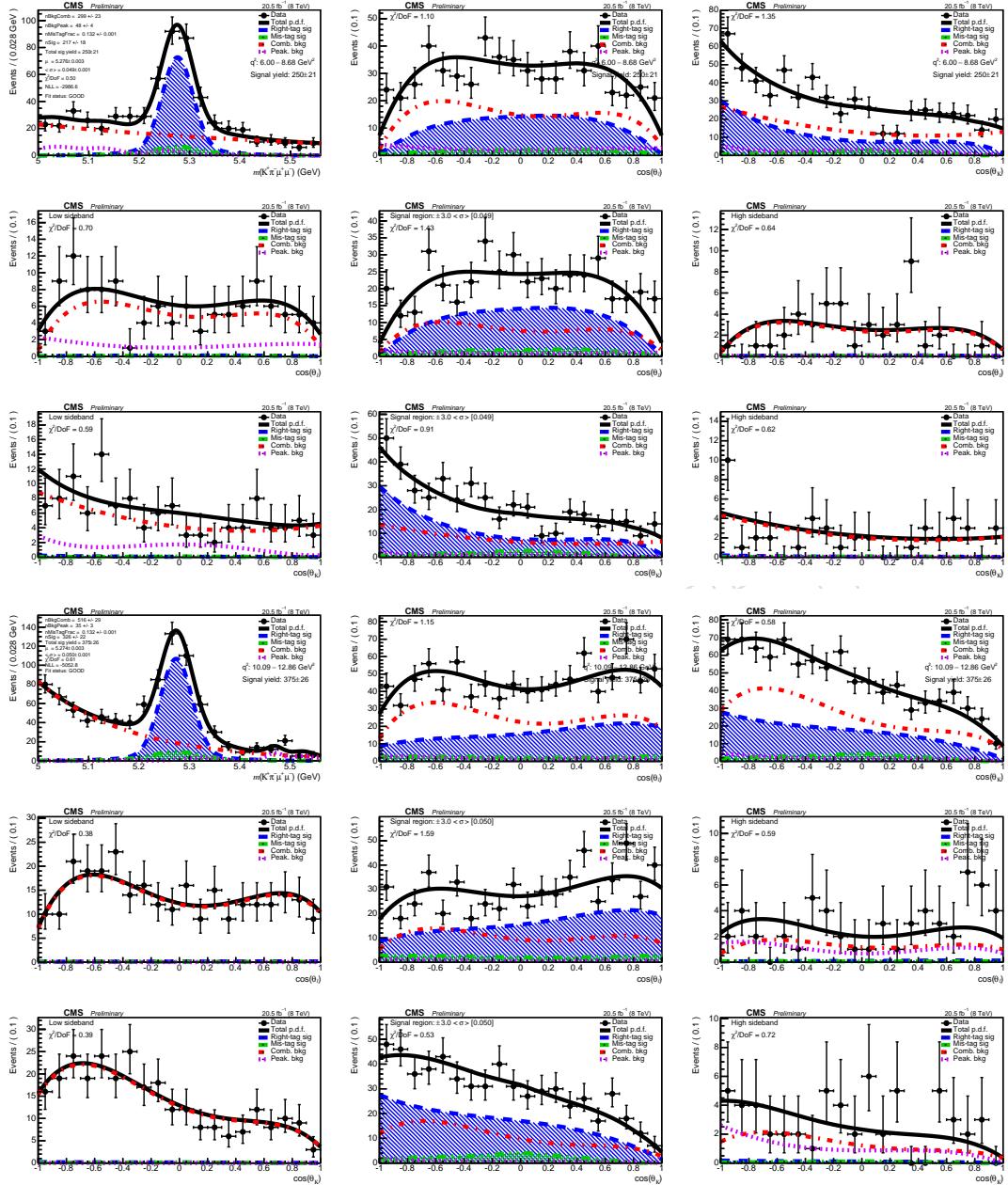


Figure 77: Fit results on data for q^2 bins #3 (top nine plots) and #5 (bottom nine plots). The different plots are, in order from left to right and from top to bottom: the projection on the B^0 invariant mass; the projection on the angular observable θ_l ; the projection on the angular observable θ_K ; the projection in the low B^0 mass sideband as a function of θ_l ; the projection in the signal region as a function of θ_l ; the projection in the high B^0 mass sideband as a function of θ_l ; the projection in the low B^0 mass sideband as a function of θ_K ; the projection in the signal region as a function of θ_K ; the projection in the high B^0 mass sideband as a function of θ_K ; for the definition of signal and sideband regions see Sec. 3.2. The degrees of freedom to compute the fit χ^2 are simply the number of bins.

₁₂₁₇ **10.12 Special q^2 bin: $1 < q^2 < 6 \text{ (GeV}/c^2)^2$**

₁₂₁₈ To do ...

DRAFT

1219 11 Discussion and Conclusion

1220 In summary, using 20.5 fb^{-1} of data recorded with the CMS detector during 2012, the angular
 1221 analysis of the decay $B^0 \rightarrow K^{*0}\mu^+\mu^-$ has been carried out measuring the forward-backward
 1222 asymmetry of the muons, A_{FB} , the fraction of longitudinal polarization of the $K^*(892)$, F_L , and
 1223 the differential branching fraction $d\mathcal{B}/dq^2$ as a function of the dimuon invariant mass squared,
 1224 q^2 . All three observables show good agreement with the SM predictions within the errors of
 1225 the measurement and, in the dimuon invariant mass squared range $1 < q^2 < 6 (\text{GeV}/c^2)^2$, the
 1226 results are: $A_{FB} = XXX^{+XXX}_{-XXX}$ (stat) $\pm XXX$ (syst), $F_L = XXX^{+XXX}_{-XXX}$ (stat) $\pm XXX$ (syst), and
 1227 $d\mathcal{B}/dq^2 = XXX^{+XXX}_{-XXX}$ (stat) $\pm XXX$ (syst).

1228 With the higher number of events available in the control channels, the measurements of the
 1229 fraction of S -wave, F_S , and the (S - P)-wave interference, A_S , have been performed. For the
 1230 control channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\text{J}/\psi(\mu^+\mu^-)$ the values of such parameters are $F_S = 7 \times$
 1231 $10^{-9} \pm 0.002$ (stat) and $A_S = -0.102 \pm 0.003$ (stat), while for $B^0 \rightarrow K^{*0}(K^+\pi^-)\psi'(\mu^+\mu^-)$ the
 1232 values are $F_S = 0.066 \pm 0.019$ (stat) and $A_S = -0.102 \pm 0.014$ (stat).



1233 References

- [1] A. Ali, T. Mannel, and T. Morozumi, "Forward-Backward asymmetry of dilepton angular distribution in the decay $b \rightarrow s\ell^+\ell^-$ ", *Phys. Lett. B* **273** (1991) 505–512.
- [2] U. Egede, "Angular correlations in the $\bar{B}^0 \rightarrow \bar{K}^{*0}\mu^+\mu^-$ decay", CERN-LHCb 057, 2007.
- [3] A. Ali, G. Kramer, and G.-h. Zhu, " $B \rightarrow K^*\ell^+\ell^-$ in soft-collinear-effective-theory", *Eur. Phys. J. C* **47** (2006) 625–641, arXiv:0601034.
- [4] C. Bobeth, G. Hiller, and G. Piranishvili, "CP asymmetries in $B^0 \rightarrow \bar{K}^*(\rightarrow \bar{K}\pi)\bar{\ell}\ell$ and untagged $\bar{B}_s, B_s \rightarrow \phi(\rightarrow K^+K^-)\bar{\ell}\ell$ ", *JHEP* **07** (2008) 106, arXiv:0805.2525.
- [5] A. Khodjamirian, T. Mannel, A.A. Pivovarov, Y.-M. Wang, "Charm-loop effect in $B \rightarrow K^{(*)}\ell^+\ell^-$ and $B \rightarrow K^* \gamma$ ", *JHEP* **1009** (2010) 089, arXiv:1006.4945.
- [6] C. Bobeth, G. Hiller, and D. van Dyk, "The benefits of $B \rightarrow K^*\ell^+\ell^-$ decays at low recoil", *JHEP* **1007** (2010) 098, arXiv:1006.5013.
- [7] C. Bobeth, G. Hiller, D. van Dyk, and C. Wacker, "The decay $B \rightarrow K\ell^+\ell^-$ at low hadronic recoil and model-independent delta $B = 1$ constraints", (2011). arXiv:1111.2558.
- [8] BaBar Collaboration, "Angular distributions in the decays $B \rightarrow K^*\ell^+\ell^-$ ", *Phys. Rev. D* **79** (2009) 031102, arXiv:0804.4412.
- [9] Belle Collaboration, "Measurement of the differential branching fraction and forward-backward asymmetry for $B \rightarrow K^{(*)}\ell^+\ell^-$ ", *Phys. Rev. Lett.* **103** (2009) 171801, arXiv:0810.0335.
- [10] CDF Collaboration, "Measurements of the angular distributions in the decays $B \rightarrow K^{(*)}\mu^+\mu^-$ at CDF", *Phys. Rev. Lett.* **108** (2012) 081807, arXiv:0810.0335.
- [11] LHCb Collaboration, "Differential branching fraction and angular analysis of the decay $B \rightarrow K^{(*)}\mu^+\mu^-$ ", *JINST* **8** (2013) 131, arXiv:1304.6325.
- [12] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [13] CMS Collaboration, "CMS luminosity based on pixel cluster counting - Summer 2013 update", CMS Physics Analysis Summary CMS-PAS-LUM 2013/001, 2013.
- [14] E. Barberio, and Z. Was, "Photos - a universal Monte Carlo for QED radiative corrections in decays", CERN-TH 7033-93, 1993.
- [15] GEANT4 Collaboration, "GEANT4 - a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.
- [16] T. Blake, U. Egede, and A. Shires, "The effect of S-wave interference on the $B^0 \rightarrow K^{*0}\ell^+\ell^-$ angular observables", *JHEP* **03** (2013) 027, doi:10.1007/JHEP03(2013)027, arXiv:1210.5279.
- [17] CMS Collaboration, "Measurement of the B^0 production cross-section in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **106** (2011) 252001, arXiv:1104.2892.
- [18] BaBar Collaboration, "Measurement of decay amplitudes of $B \rightarrow J/\psi K^*, \psi(2S)K^*$, and $\chi_{c1}K^*$ with an angular analysis", *Phys. Rev. D* **76** (2007) 031102, doi:10.1103/PhysRevD.76.031102.

- [1272] [19] F. James, and M. Winkler, "MINUIT user's guide", (2004).
[1273] <http://seal.web.cern.ch/seal/work-packages/mathlibs/minuit/doc/doc.html>.
- [1274] [20] CMS Collaboration, "Measurement of the B^+ production cross-section in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **106** (2011) 112001, arXiv:1101.0131.
- [1276] [21] CMS Collaboration, "Measurement of the B_s^0 production cross section with $B_s^0 \rightarrow J/\psi\phi$ decays in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. D* **84** (2011) 052008, arXiv:1106.4048.
- [1279] [22] CMS Collaboration, "Measurement of the Λ_b cross section and the anti- Λ_b to Λ_b ratio with $\Lambda_b \rightarrow J/\psi\Lambda$ decays in pp collisions at $\sqrt{s} = 7$ TeV", (2012). arXiv:1205.0594.
- [1281] [23] MATLAB, "MATLAB official website", (2014). <http://www.mathworks.com>.
- [1282] [24] BaBar Collaboration, "Time-integrated and time-dependent angular analyses of $B^0 \rightarrow J/\psi K\pi$: a measurement of $\cos 2\beta$ with no sign ambiguity from strong phases", *Phys. Rev. D* **71** (2005) 032005, doi:10.1103/PhysRevD.71.032005.
- [1285] [25] M. Beneke, T. Feldmann, and D. Seidel, "Systematic approach to exclusive $B \rightarrow V\ell^+\ell^-$, $V\gamma$ decays", *Nucl. Phys. B* **612** (2001) 25, arXiv:0106067.
- [1287] [26] B. Grinstein, and D. Pirjol, "Exclusive rare $B \rightarrow K^{*0}\ell^+\ell^-$ decays at low recoil: controlling the long-distance effects", *Phys. Rev. D* **70** (2004) 114005, arXiv:0404250.
- [1289] [27] F. Beaujean, C. Bobeth, and D. van Dyk, "Comprehensive Bayesian analysis of rare (semi) leptonic and radiative B decays", *Eur. Phys. J. C* **74** (2014) 2897, arXiv:1310.2478.
- [1291] [28] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, "Lattice QCD calculation of form factors describing the rare decays $B \rightarrow K^*\ell^+\ell^-$ and $B_s \rightarrow \phi\ell^+\ell^-$ ", *Phys. Rev. D* **89** (2014) 094501, arXiv:1310.3722.
- [1294] [29] BaBar Collaboration, "Measurement of branching fractions and rate asymmetries in the rare decays $B^0 \rightarrow K^{(*)}\ell^+\ell^-$ ", *Phys. Rev. D* **86** (2012) 032012, doi:10.1103/PhysRevD.86.032012.
- [1297] [30] BaBar Collaboration, "Radiative-penguin results from BaBar", (2012). Lake Louise Winter Institute conference (Canada), February 2012.
- [1299] [31] Atlas Collaboration, " $B \rightarrow K^{(*)}\mu^+\mu^-$ with Atlas", (2013). Beauty 2013 conference, Bologna (Italy), April 2013.

1301 A Truth Matching

The whole analysis validation based on MC simulation relies on the capability of associating the reconstructed events to the corresponding generated events, procedure that is also called “truth-matching”. The truth-matching is performed by means of the variable:

$$\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}, \quad (10)$$

1302 where $\Delta\eta$ and $\Delta\phi$ are the differences between the reconstructed candidate’s pseudorapidity
1303 and azimuthal angles from the true signal values.

1304 Thresholds in ΔR for matching pion, kaon and muon candidates to signal simulation events are
1305 chosen in such a way that the total number of true signal events in the region of interest ($5 \lesssim$
1306 $m(K\pi\mu\mu) \lesssim 5.56$ GeV/c 2) failing the matching is approximately equal to the total number of
1307 signal events outside of the B^0 mass peak that pass the matching in the same region. The tuning
1308 of ΔR is performed on the signal sample simulation after applying all selection cuts listed in
1309 Sec. 3. Using this criteria the threshold to match hadrons (muons) is set to $\Delta R < 0.3$ (0.004).

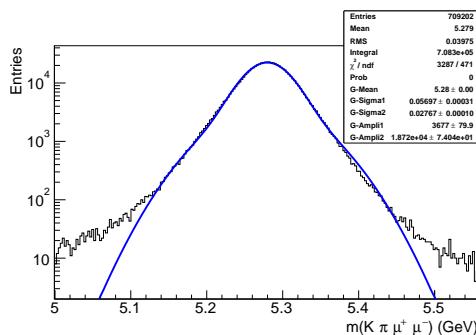


Figure 78: B^0 invariant mass distribution for truth-matched events. The distribution is fit with a double Gaussian with unique mean whose integral is 705 971 events. The total number of events falling in the region of interest ($5 \lesssim m(K\pi\mu\mu) \lesssim 5.56$ GeV/c 2) is 708 300. Events with Final State Radiation are removed.

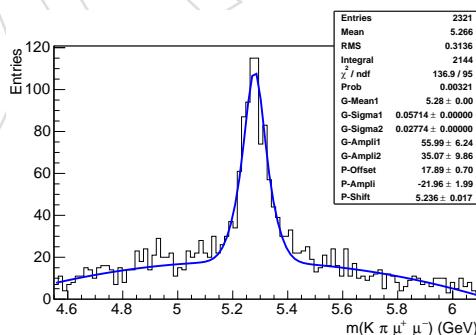


Figure 79: B^0 invariant mass distribution for events failing the truth-matching criteria. The distribution is fit with a double Gaussian with unique mean whose integral is 1299 events (the shape is fixed from the fit to the truth-matched events, see Fig. 78), and a second order polynomial describing the background. Events with Final State Radiation are removed.

1310 To consider a B^0 candidate as a truth-matched signal, all four tracks must fulfill the matching
1311 criteria. With the hadron and muon thresholds being set to 0.3 and 0.004 respectively, on a
1312 sample of 711 523 true signal events, 1299 are not matched, while 708 300 are matched, but

¹³¹³ 2329 out of which fall outside the B^0 mass peak (see Fig.s 78 and 79). Events with Final State
¹³¹⁴ Radiation are removed.

¹³¹⁵ All in all the total efficiency of the truth-matching criteria is $\sim 99.6\%$ ($= 708\,300 / 711\,523$) and
¹³¹⁶ its purity is $\sim 99.5\%$ ($= 705\,971 / 709\,202$).

DRAFT

1317 B Results of the Selection Cut Tuning Procedure

1318 In this section are reported all the plots related to the selection cut optimization procedure ex-
 1319 plained in Sec. 3. In particular are shown the trends of the signal and of the figure of merit,
 1320 $S / \sqrt{(S + B)}$, as a function of the selection cut value for all selections explored and for all opti-
 1321 mization steps.

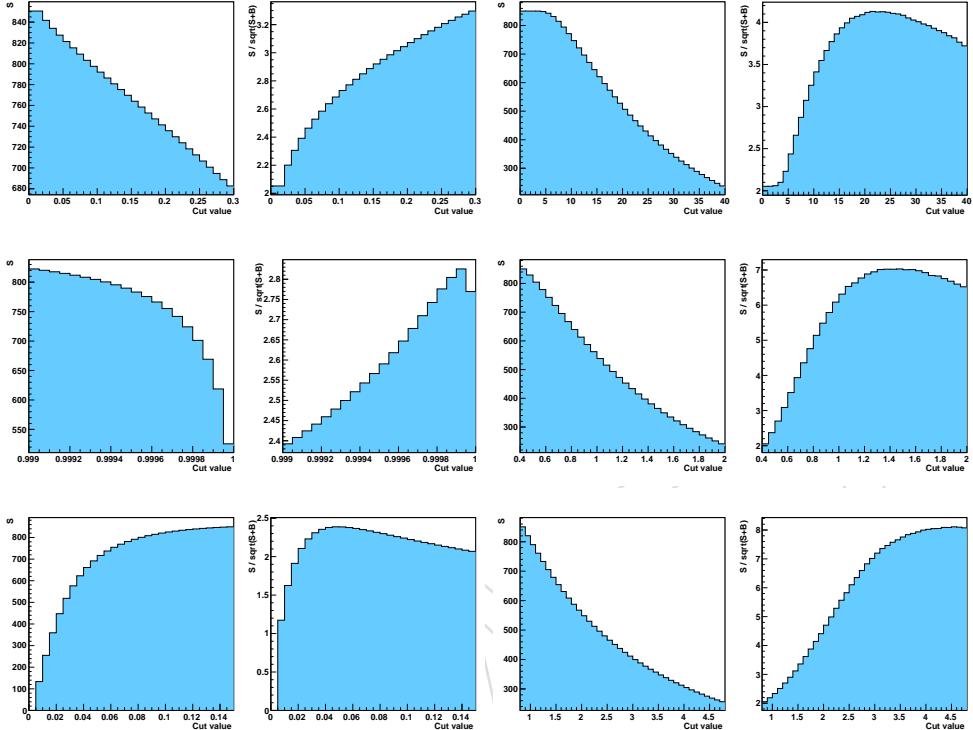


Figure 80: Step 1 of the selection cut optimization procedure. The plots are shown in pairs, the former shows the trend of the signal yield (S), while the latter shows the trend of the figure of merit $S / \sqrt{(S + B)}$:

1. shows the trends as a function of the selection cut $CL(B^0 - vtx)$;
2. shows the trends as a function of the selection cut $L_{xy}(B^0 - vtx) / \sigma$;
3. shows the trends as a function of the selection cut $\cos(\alpha_{xy}^{B^0})$;
4. shows the trends as a function of the selection cut p_T^h ;
5. shows the trends as a function of the selection cut mass window around the $K^*(892)$ mass;
6. shows the trends as a function of the selection cut DCA_{xy}^h / σ .

The numbering goes from left to right and from top to bottom.

- 1322 S is the number of truth-matched events (see definition of truth-matching in App. A) com-
 1323 puted on signal simulation whose B^0 invariant mass falls within $\pm 2.5\sigma_{mB^0}$ around the B^0 PDG
 1324 mass (where σ_{mB^0} is determined from signal simulation from a double Gaussian fit with unique
 1325 mean), while B is the number of events computed with data whose B^0 invariant mass falls
 1326 within the ranges $[(m_{B^0\text{PDG}} - 5.5\sigma_{mB^0}), (m_{B^0\text{PDG}} - 3\sigma_{mB^0})] \text{ GeV}/c^2$ or $[(m_{B^0\text{PDG}} + 3\sigma_{mB^0}), (m_{B^0\text{PDG}} + 5.5\sigma_{mB^0})] \text{ GeV}/c^2$.
 1327
- 1328 The pre-selection cuts adopted for the optimization procedure are those used at trigger level

and described in Sec. 2, while the selection cuts scanned for the optimization are reported in each figure caption (the selection cuts that are not here reported but have been aforementioned are also applied, but the “ B^0 & ψ ” mass cut). The number of events computed with the simulation (S) is rescaled by the ratio between data and simulation integrated luminosities (the latter is corrected by a factor that takes into account that PYTHIA overestimates the B^0 production cross-section [17]. The dimuon invariant mass region for which the selection cuts are optimized is $1 < m(\mu\mu) < 2.702$ and $4 < m(\mu\mu) < 4.359$ (GeV/c^2)².

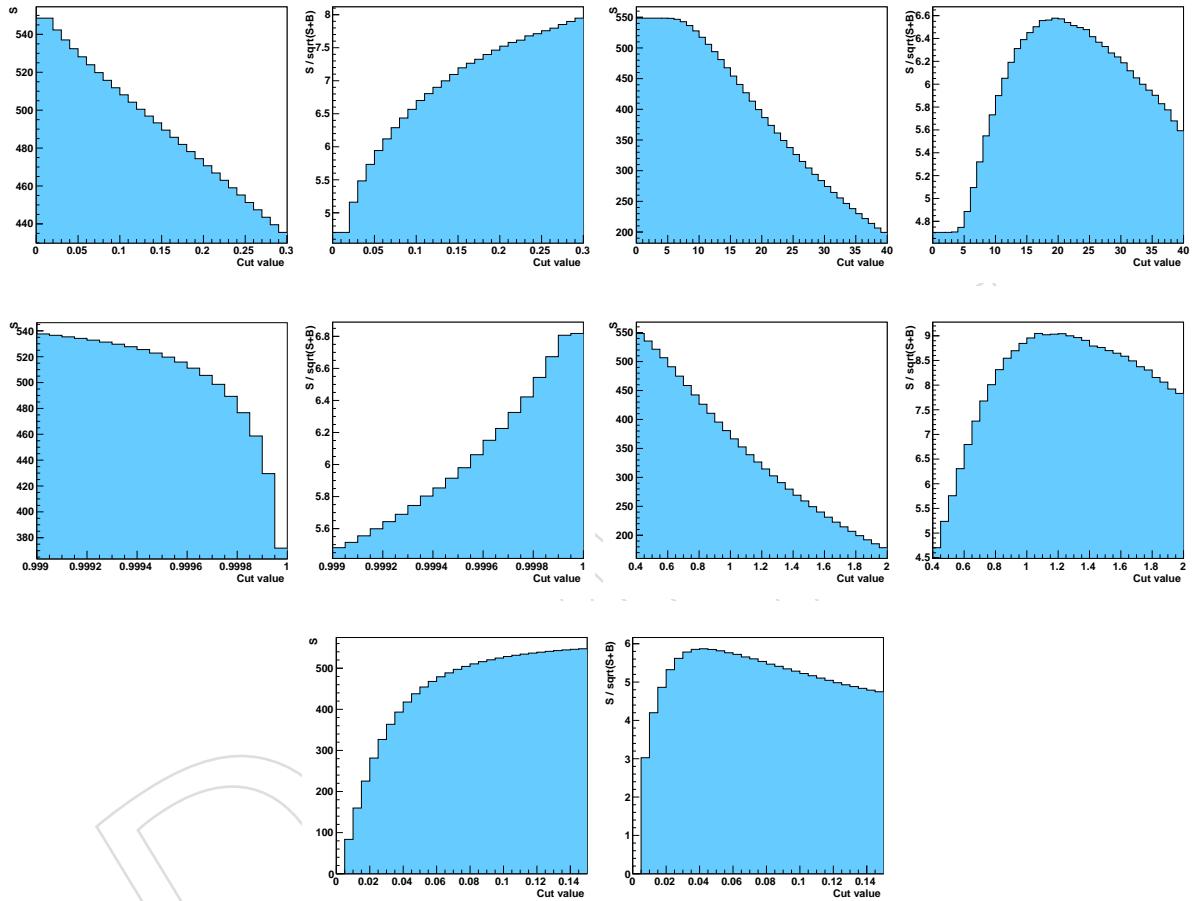


Figure 81: Step 2 of the selection cut optimization procedure. The plots are shown in pairs, the former shows the trend of the signal yield (S), while the latter shows the trend of the figure of merit $S / \sqrt{(S + B)}$:

1. shows the trends as a function of the selection cut $\text{CL}(B^0 - \text{vtx})$;
2. shows the trends as a function of the selection cut $L_{xy}(B^0 - \text{vtx}) / \sigma$;
3. shows the trends as a function of the selection cut $\cos(\alpha_{xy}^{B^0})$;
4. shows the trends as a function of the selection cut p_T^h ;
5. shows the trends as a function of the selection cut mass window around the $K^*(892)$ mass.

The numbering goes from left to right and from top to bottom.

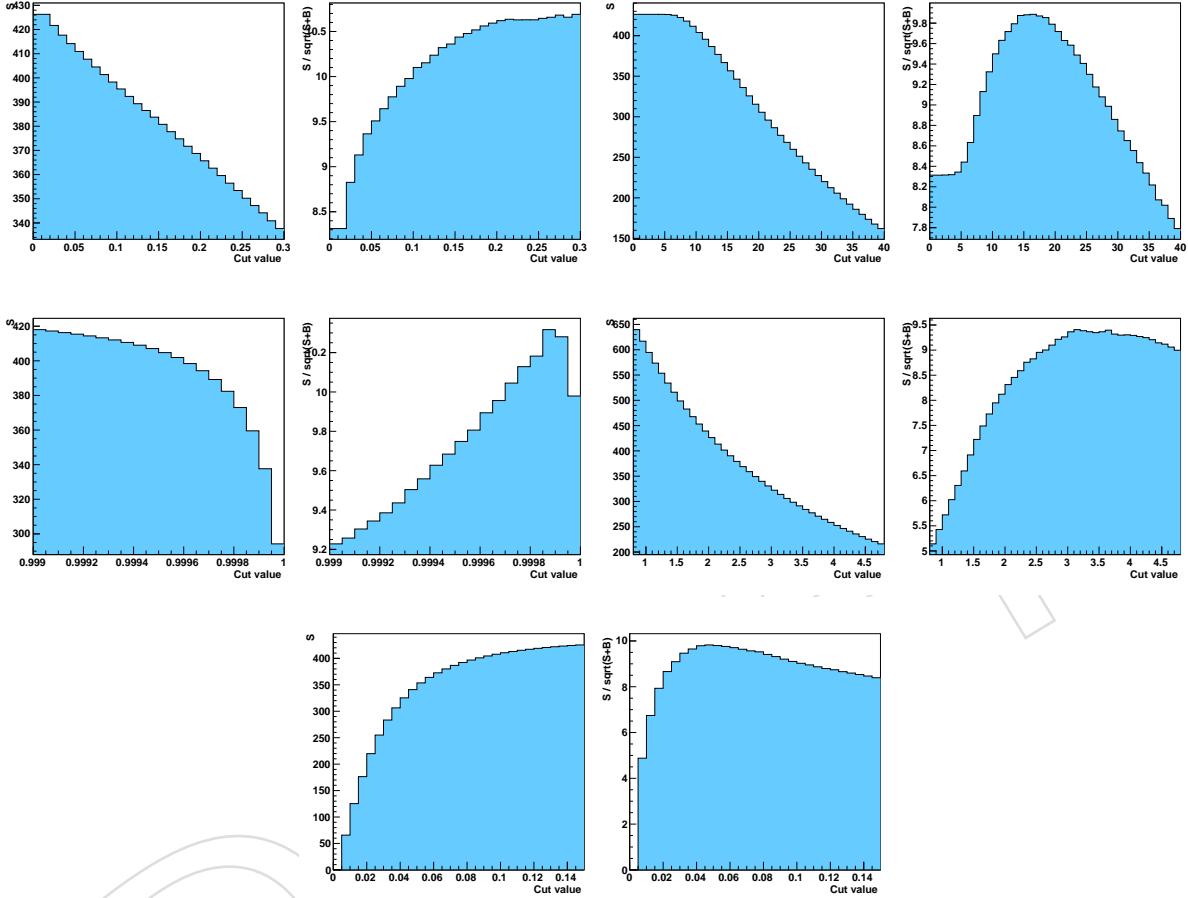


Figure 82: Step 3 of the selection cut optimization procedure. The plots are shown in pairs, the former shows the trend of the signal yield (S), while the latter shows the trend of the figure of merit $S / \sqrt{S + B}$:

1. shows the trends as a function of the selection cut $CL(B^0 - vtx)$;
2. shows the trends as a function of the selection cut $L_{xy}(B^0 - vtx) / \sigma$;
3. shows the trends as a function of the selection cut $\cos(\alpha_{xy}^{B^0})$;
4. shows the trends as a function of the selection cut DCA_{xy}^h / σ ;
5. shows the trends as a function of the selection cut mass window around the $K^*(892)$ mass.

The numbering goes from left to right and from top to bottom.

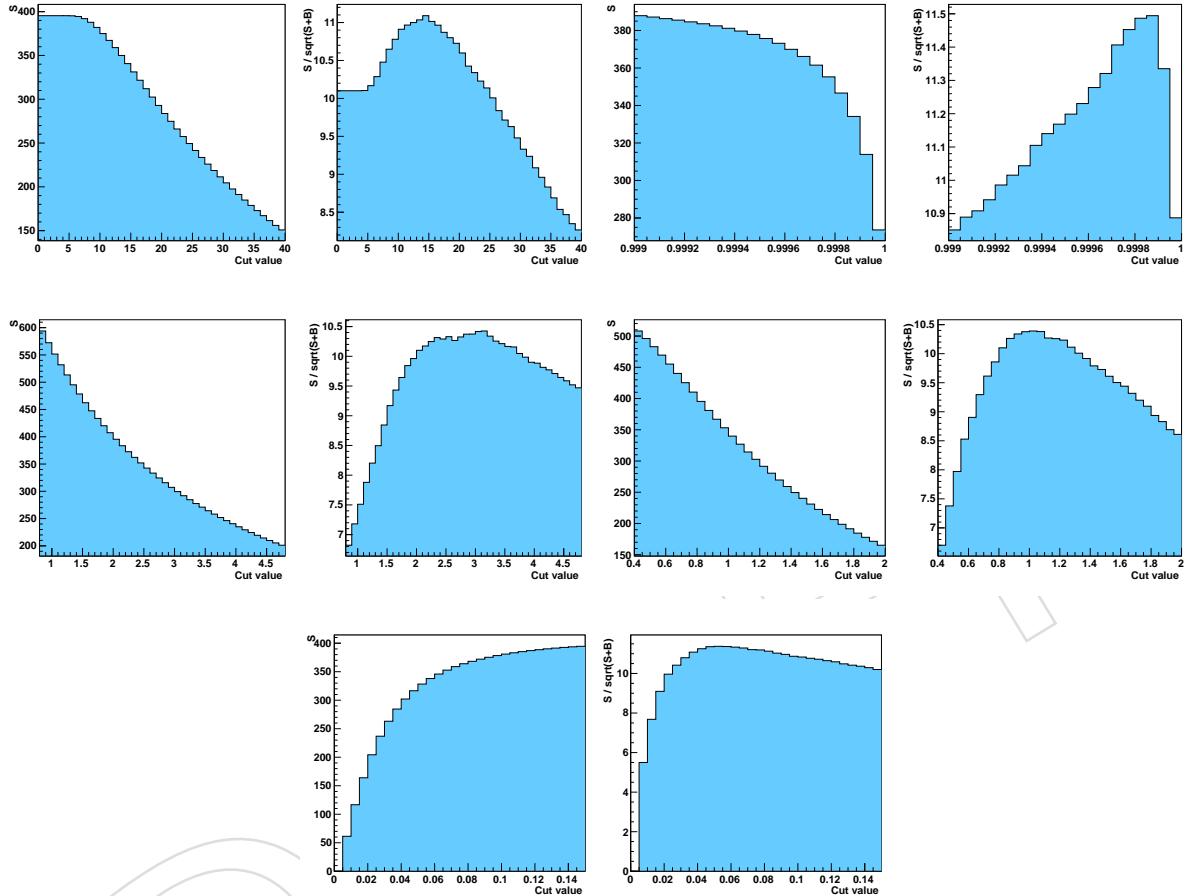


Figure 83: Step 4 of the selection cut optimization procedure. The plots are shown in pairs, the former shows the trend of the signal yield (S), while the latter shows the trend of the figure of merit $S / \sqrt{(S + B)}$:

1. shows the trends as a function of the selection cut $L_{xy}(B^0 - \text{vtx})/\sigma$;
2. shows the trends as a function of the selection cut $\cos(\alpha_{xy}^{B^0})$;
3. shows the trends as a function of the selection cut DCA_{xy}^h/σ ;
4. shows the trends as a function of the selection cut p_T^h ;
5. shows the trends as a function of the selection cut mass window around the $K^*(892)$ mass.

The numbering goes from left to right and from top to bottom.

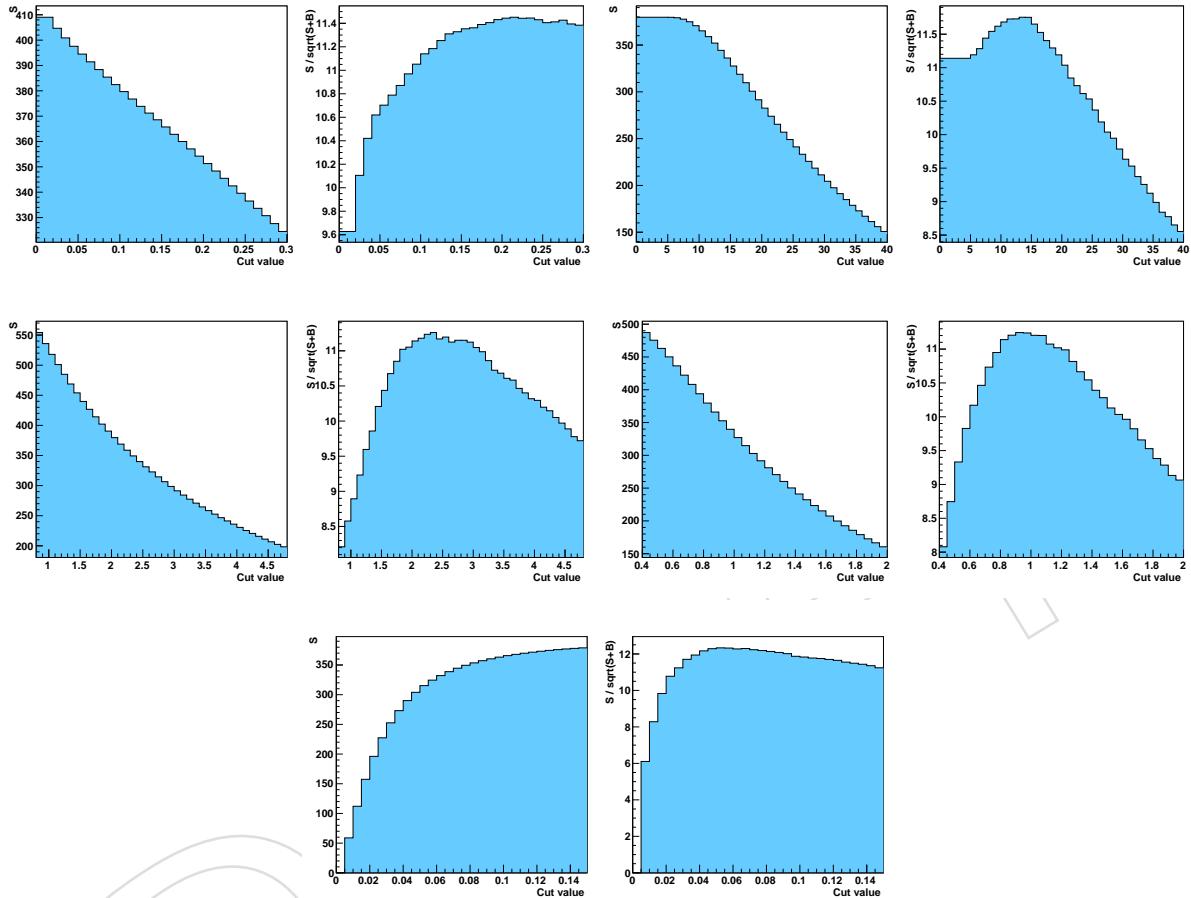


Figure 84: Step 5 of the selection cut optimization procedure. The plots are shown in pairs, the former shows the trend of the signal yield (S), while the latter shows the trend of the figure of merit $S / \sqrt{(S + B)}$:

1. shows the trends as a function of the selection cut $\text{CL}(B^0 - \text{vtx})$;
2. shows the trends as a function of the selection cut $L_{xy}(B^0 - \text{vtx})/\sigma$;
3. shows the trends as a function of the selection cut DCA_{xy}^h/σ ;
4. shows the trends as a function of the selection cut p_T^h ;
5. shows the trends as a function of the selection cut mass window around the $K^*(892)$ mass.

The numbering goes from left to right and from top to bottom.

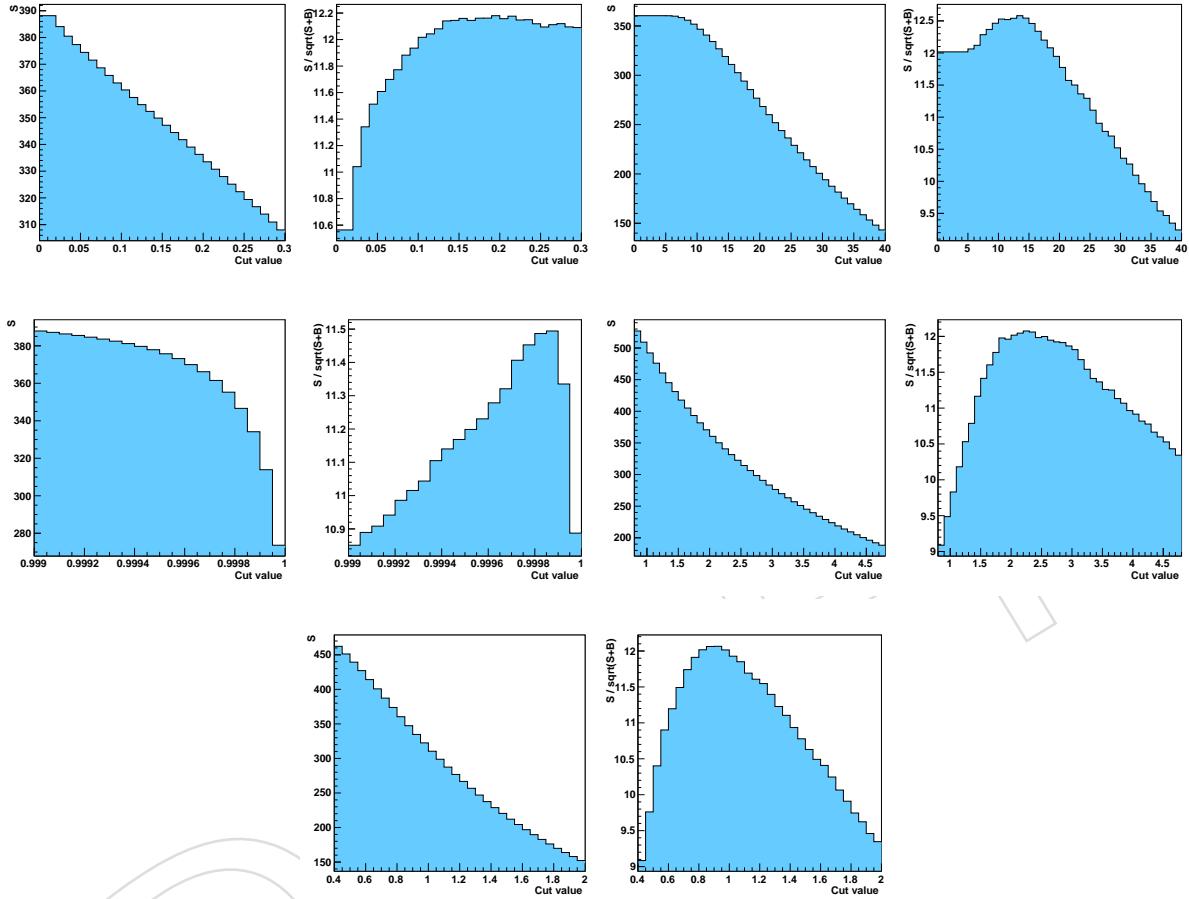


Figure 85: Step 6 of the selection cut optimization procedure. The plots are shown in pairs, the former shows the trend of the signal yield (S), while the latter shows the trend of the figure of merit $S/\sqrt{(S+B)}$:

1. shows the trends as a function of the selection cut $\text{CL}(B^0 - \text{vtx})$;
2. shows the trends as a function of the selection cut $L_{xy}(B^0 - \text{vtx})/\sigma$;
3. shows the trends as a function of the selection cut $\cos(\alpha_{xy}^{B^0})$;
4. shows the trends as a function of the selection cut DCA_{xy}^h/σ ;
5. shows the trends as a function of the selection cut p_T^h .

The numbering goes from left to right and from top to bottom.

1336 C Plots about Data-Simulation Comparison

1337 In this section are reported plots comparing data and simulation on the basic kinematic and
 1338 dynamic observables as described in Sec. 6. In data only the J/ψ and ψ' dimuon mass regions
 1339 have been selected by applying the selection criteria described in Sec. 3. The simulation is a
 1340 combination of both control channels weighted by their branching fractions. All plots from
 1341 data are background subtracted (the plot of the sidebands is subtracted from the plot of the
 1342 signal region, for the definition of signal and sideband regions see Sec. 3.2).

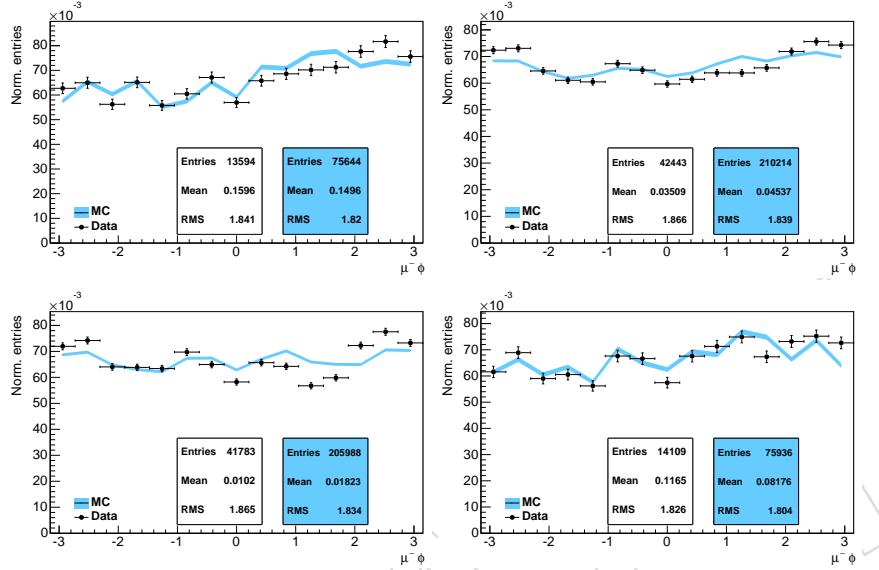


Figure 86: Azimuthal angle distributions for negative muons in different pseudorapidity ranges: $-2.4 < \eta < -1.2$ first plot, $-1.2 < \eta < 0$ second plot, $0 < \eta < 1.2$ third plot, $1.2 < \eta < 2.4$ fourth plot. In data only the J/ψ and ψ' dimuon mass regions have been selected, the simulation, instead, is a combination of both control channels weighted by their branching fractions. The numbering goes from left to right and from top to bottom.

1343 Figures 86 (88) and 87 (89) show the distributions of the azimuthal angle of the muons (hadrons)
 1344 in different pseudorapidity ranges.

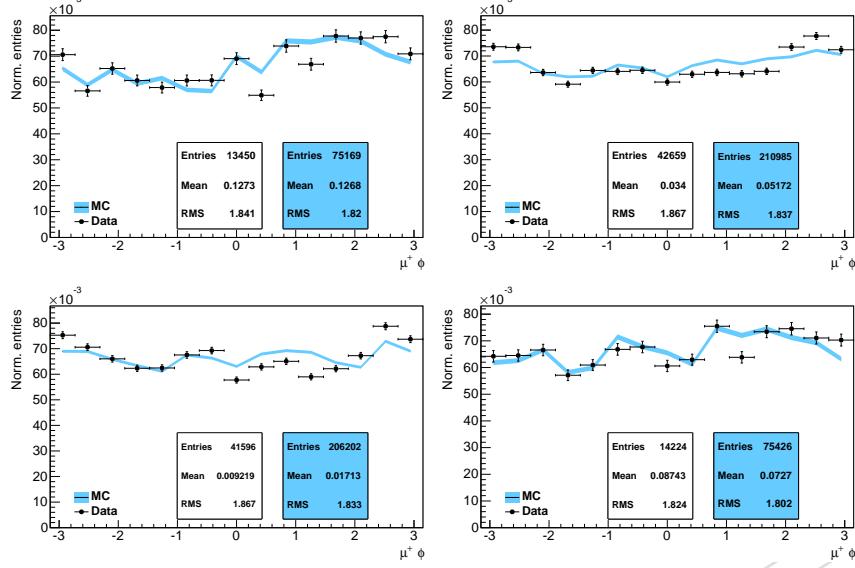


Figure 87: Azimuthal angle distributions for positive muons in different pseudorapidity ranges: $-2.4 < \eta < -1.2$ first plot, $-1.2 < \eta < 0$ second plot, $0 < \eta < 1.2$ third plot, $1.2 < \eta < 2.4$ fourth plot. In data only the J/ψ and ψ' dimuon mass regions have been selected, the simulation, instead, is a combination of both control channels weighted by their branching fractions. The numbering goes from left to right and from top to bottom.

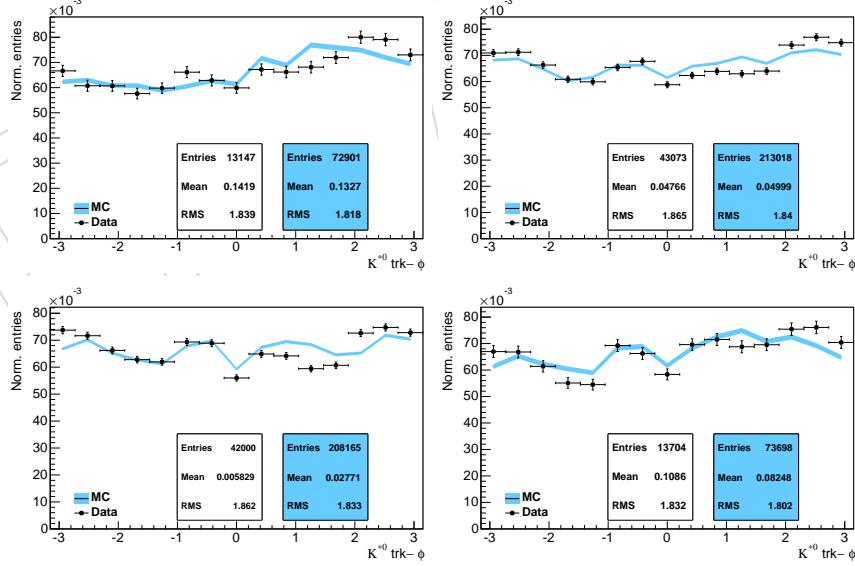


Figure 88: Azimuthal angle distributions for negative hadrons in different pseudorapidity ranges: $-3.0 < \eta < -1.2$ first plot, $-1.2 < \eta < 0$ second plot, $0 < \eta < 1.2$ third plot, $1.2 < \eta < 3.0$ fourth plot. In data only the J/ψ and ψ' dimuon mass regions have been selected, the simulation, instead, is a combination of both control channels weighted by their branching fractions. The numbering goes from left to right and from top to bottom.

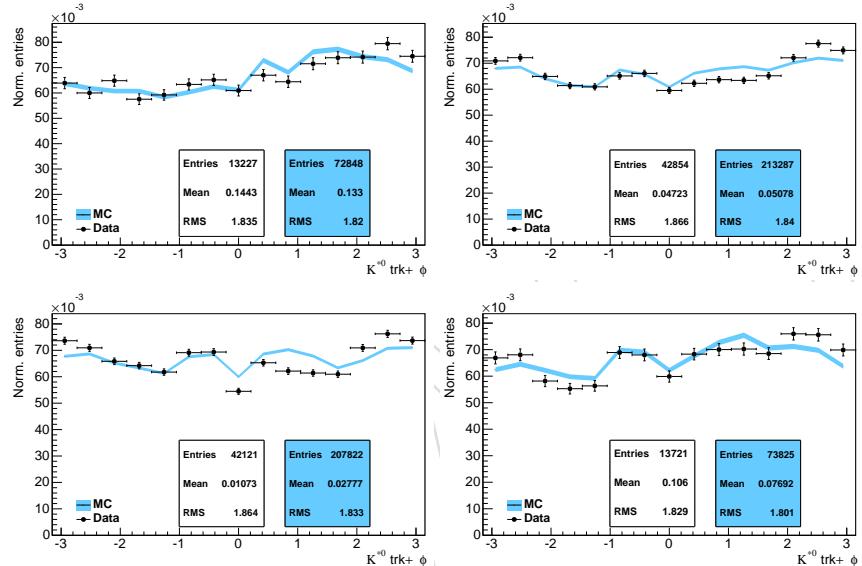


Figure 89: Azimuthal angle distributions for positive hadrons in different pseudorapidity ranges: $-3.0 < \eta < -1.2$ first plot, $-1.2 < \eta < 0$ second plot, $0 < \eta < 1.2$ third plot, $1.2 < \eta < 3.0$ fourth plot. In data only the J/ψ and ψ' dimuon mass regions have been selected, the simulation, instead, is a combination of both control channels weighted by their branching fractions. The numbering goes from left to right and from top to bottom.

1345 D Results of the Fits to the Simulation

1346 In this section are reported all fits to the data-like MC simulation containing: signal, control
 1347 channels, and background, as described in Sec. 8. The data-like MC simulation is derived from
 1348 the MC signal sample dividing it into 400 independent sub-samples. In this section are reported
 1349 the fit results of one of the 400 sub-samples (i.e. the sample of order number #0).

1350 All figures show, in order from left to right and from top to bottom:

- 1351 • projection on the B^0 invariant mass;
- 1352 • projection on the angular observable θ_l ;
- 1353 • projection on the angular observable θ_K ;
- 1354 • projection in the low B^0 mass sideband as a function of θ_l ;
- 1355 • projection in the signal region as a function of θ_l ;
- 1356 • projection in the high B^0 mass sideband as a function of θ_l ;
- 1357 • projection in the low B^0 mass sideband as a function of θ_K ;
- 1358 • projection in the signal region as a function of θ_K ;
- 1359 • projection in the high B^0 mass sideband as a function of θ_K ;

1360 where the signal and sideband regions are defined in Sec 3.2.

DRAFT

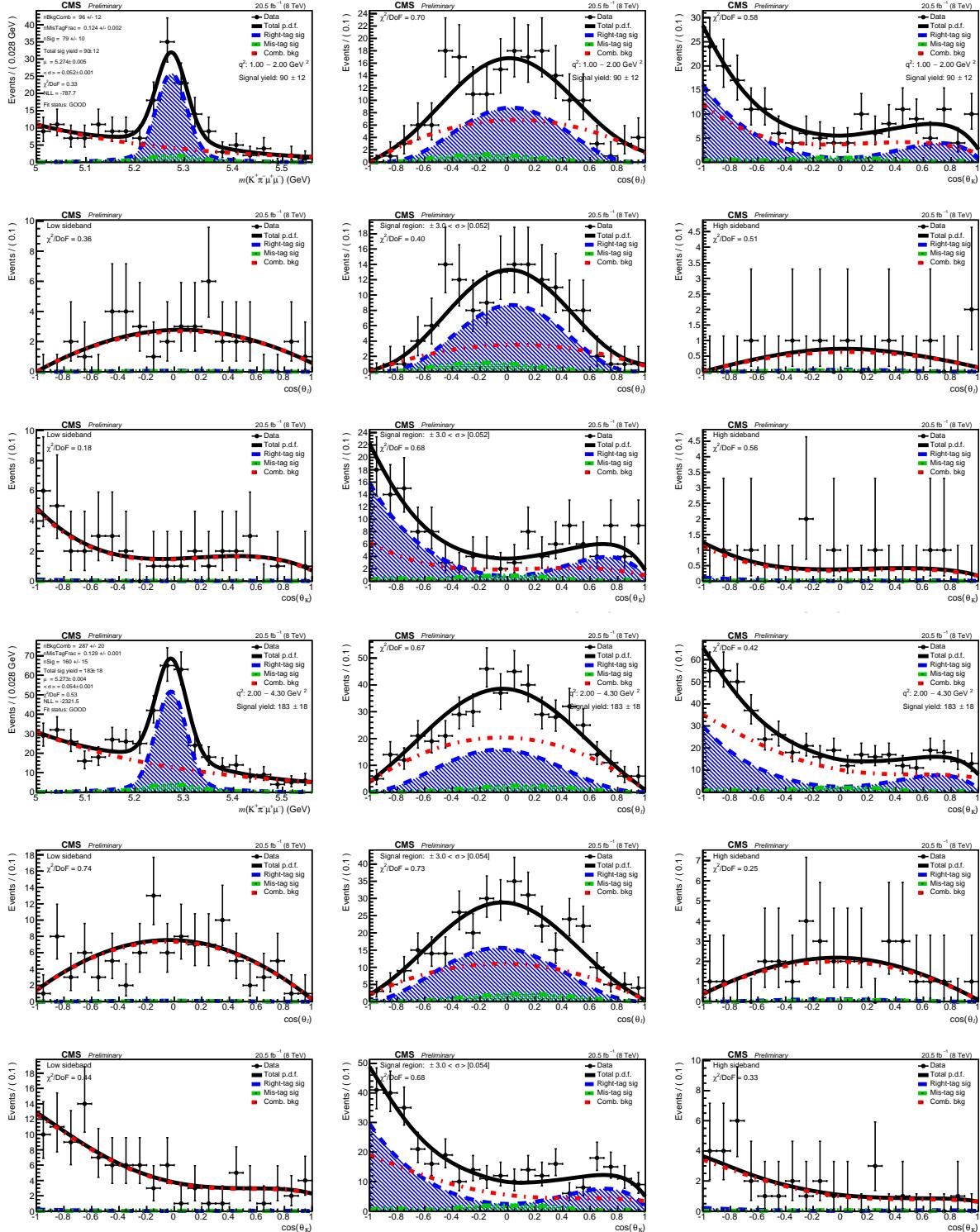


Figure 90: Fit results to measure F_L and A_{FB} on a MC simulation containing: signal, control channels, and background, as described in Sec. 8, with a data-like number of events. The figure shows the result for the dimuon q^2 bin, as from Table 4, #0 (top nine plots) and #1 (bottom nine plots). The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

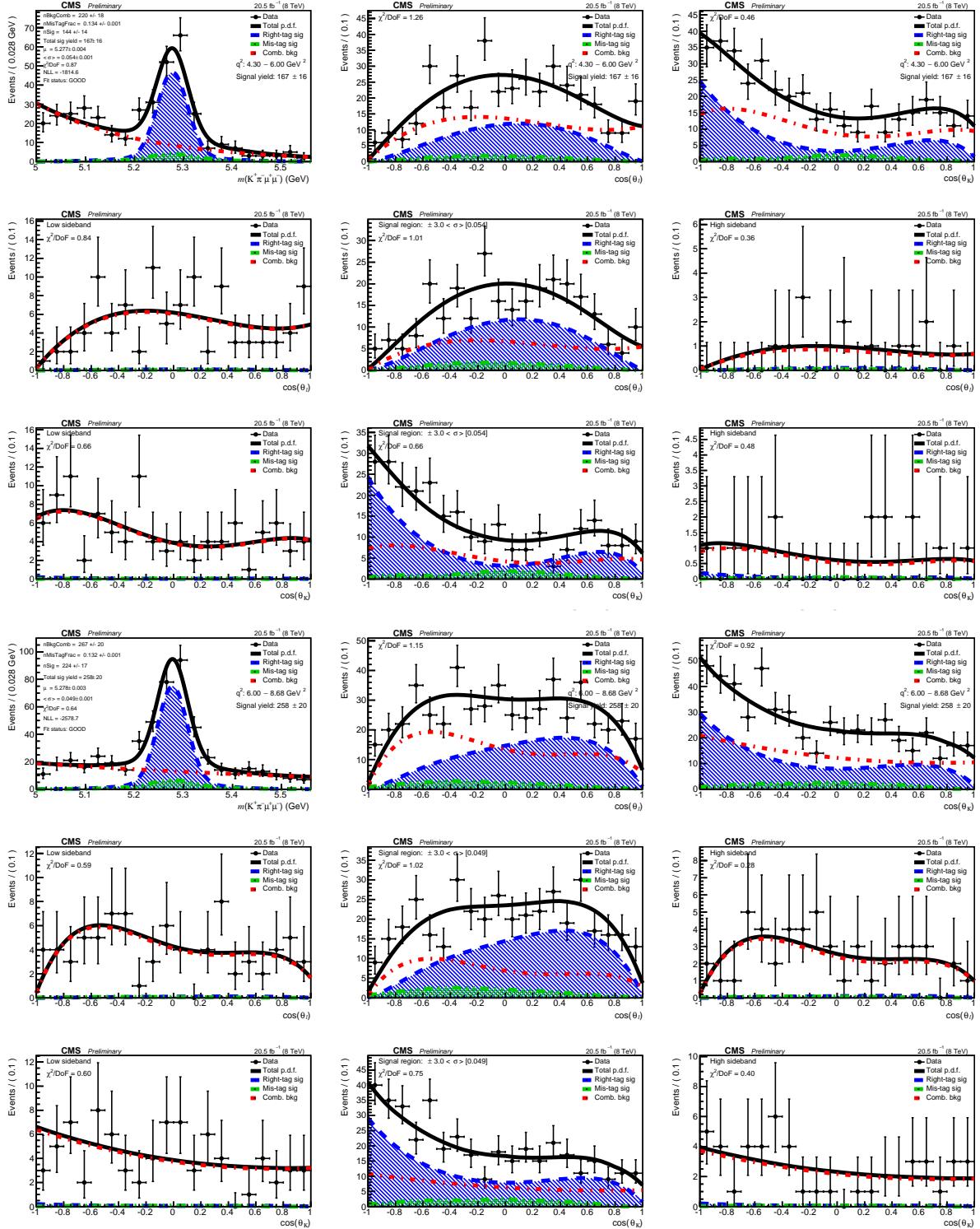


Figure 91: Fit results to measure F_L and A_{FB} on a MC simulation containing: signal, control channels, and background, as described in Sec. 8, with a data-like number of events. The figure shows the result for the dimuon q^2 bin, as from Table 4, #2 (top nine plots) and #3 (bottom nine plots). The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

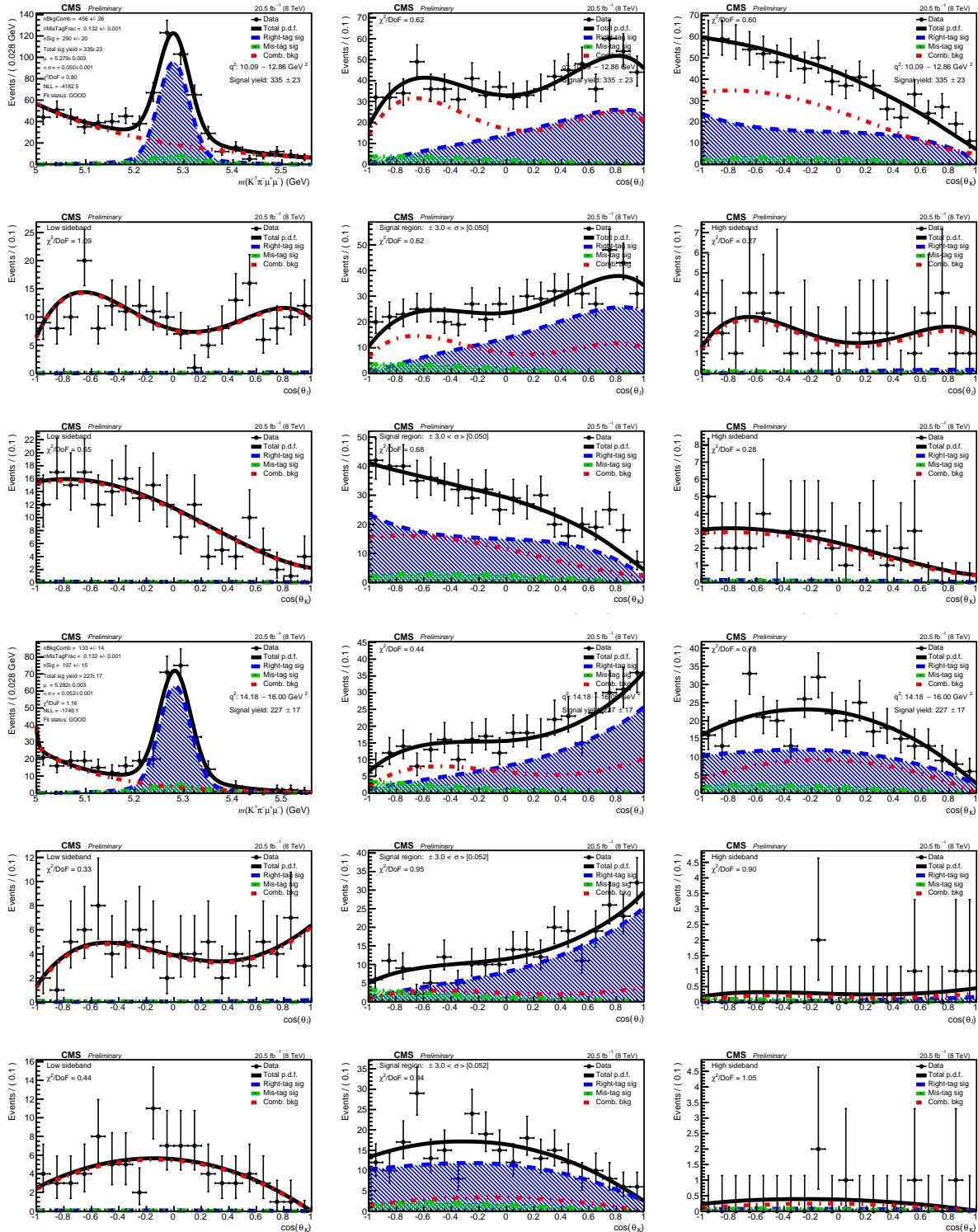


Figure 92: Fit results to measure F_L and A_{FB} on a MC simulation containing: signal, control channels, and background, as described in Sec. 8, with a data-like number of events. The figure shows the result for the dimuon q^2 bin, as from Table 4, #5 (top nine plots) and #7 (bottom nine plots). The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

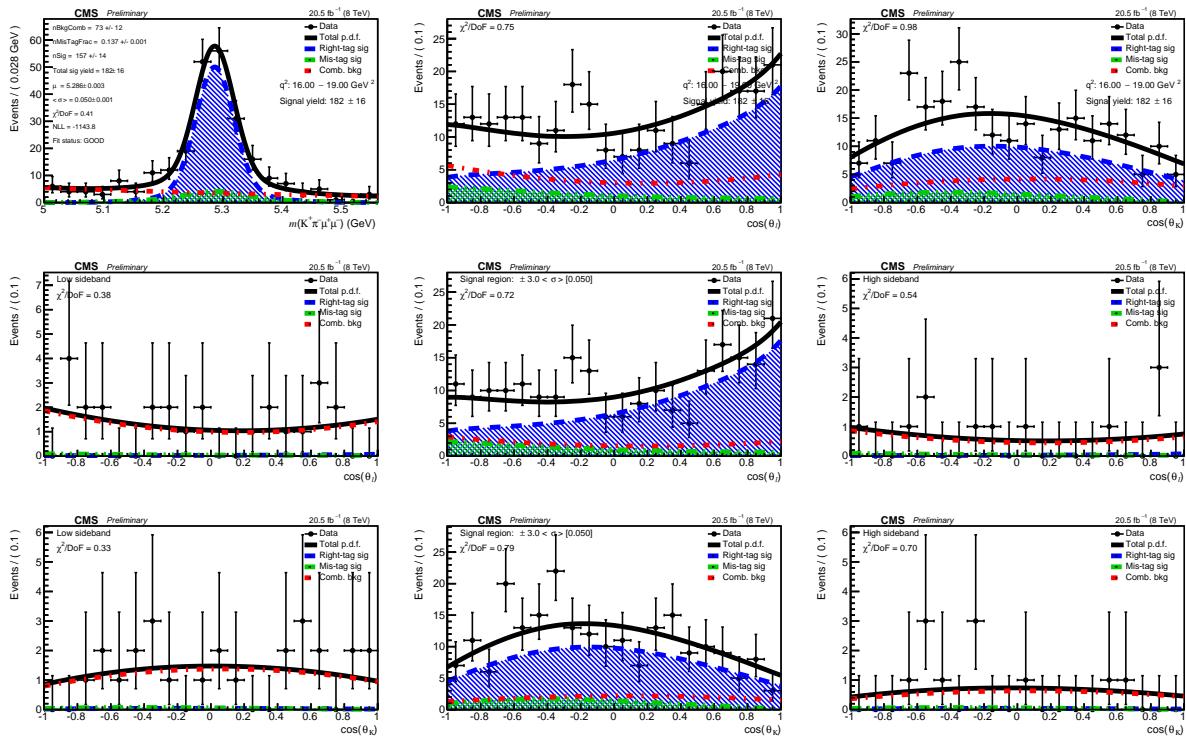


Figure 93: Fit results to measure F_L and A_{FB} on a MC simulation containing: signal, control channels, and background, as described in Sec. 8, with a data-like number of events. The figure shows the result for the dimuon q^2 bin, as from Table 4, #8. The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

1361 E Results of the Fits to the Data

1362 In this section are reported all the fits to the data to measure the fraction of longitudinal po-
 1363 larization of the $K^*(892)$, F_L , the forward-backward asymmetry of the muons, A_{FB} , and the
 1364 differential branching fraction, $d\mathcal{B}/dq^2$, as a function of the dimuon invariant mass squared,
 1365 q^2 , for the decay channel $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$. The results of the measurements are pre-
 1366 sents in Sec. 9, while the details of the p.d.f. are described in Sec. 4.

1367 All figures show, in order from left to right and from top to bottom:

- 1368 • projection on the B^0 invariant mass;
- 1369 • projection on the angular observable θ_l ;
- 1370 • projection on the angular observable θ_K ;
- 1371 • projection in the low B^0 mass sideband as a function of θ_l ;
- 1372 • projection in the signal region as a function of θ_l ;
- 1373 • projection in the high B^0 mass sideband as a function of θ_l ;
- 1374 • projection in the low B^0 mass sideband as a function of θ_K ;
- 1375 • projection in the signal region as a function of θ_K ;
- 1376 • projection in the high B^0 mass sideband as a function of θ_K ;

1377 where the signal and sideband regions are defined in Sec 3.2.

1378 The p.d.f. describing the background for the special q^2 bin $1 < q^2 < 6 (\text{GeV}/c^2)^2$ has the
 1379 components listed in Table 19 (see Sec. 4 for more details).

Table 19: Mathematical description of the background parameterization for data for the special q^2 bin $1 < q^2 < 6 (\text{GeV}/c^2)^2$.

q^2 bin index	$B_i^m(m)$	$B_i^{\theta_K}(\theta_K)$	$B_i^{\theta_l}(\theta_l)$
$1 < q^2 < 6 (\text{GeV}/c^2)^2$	1 exp.	2^{nd} degree	2^{nd} degree

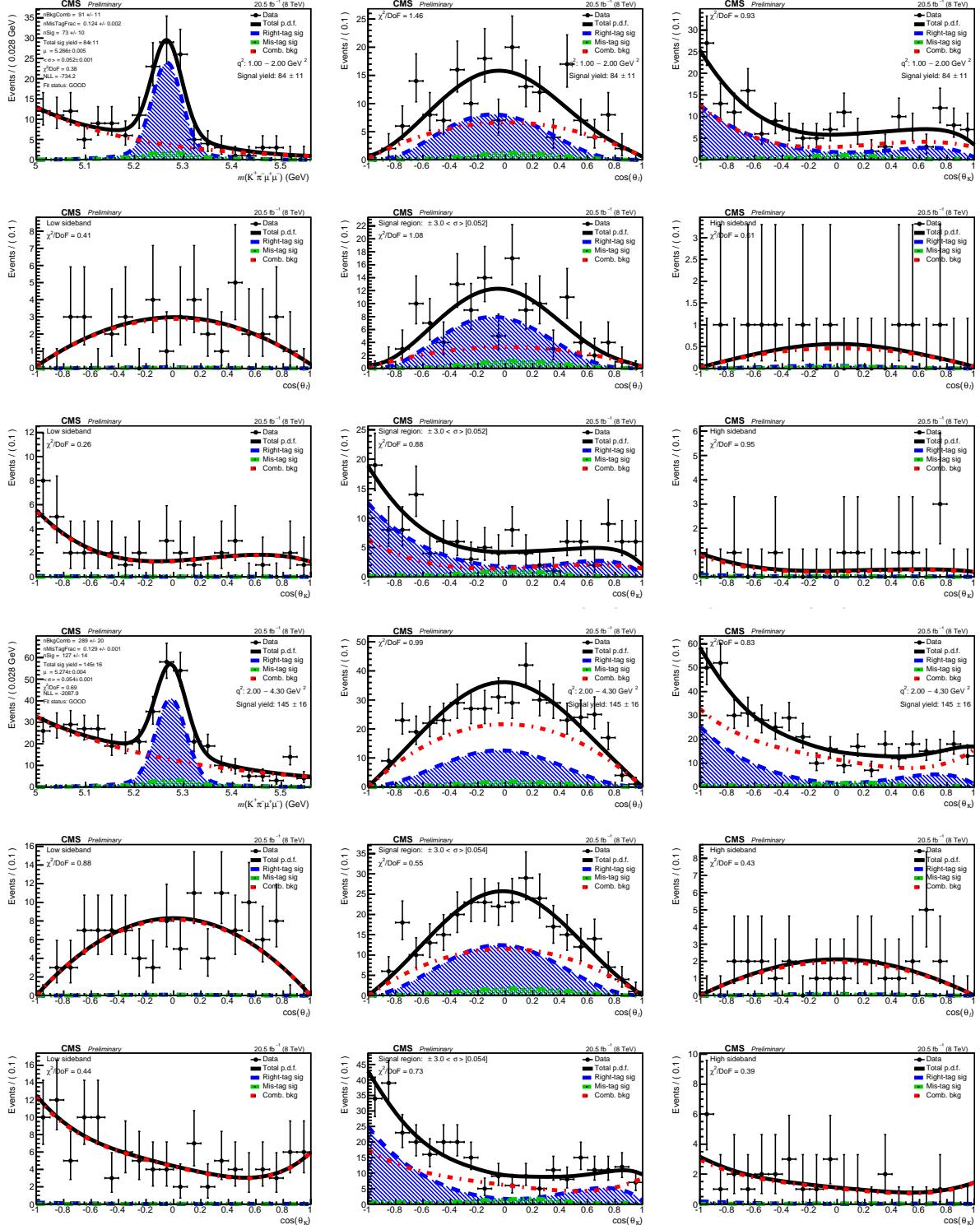


Figure 94: Fit results to measure F_L and A_{FB} with data for the decay $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$, in the dimuon q^2 bin, as from Table 4, #0 (top nine plots) and #1 (bottom nine plots). The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

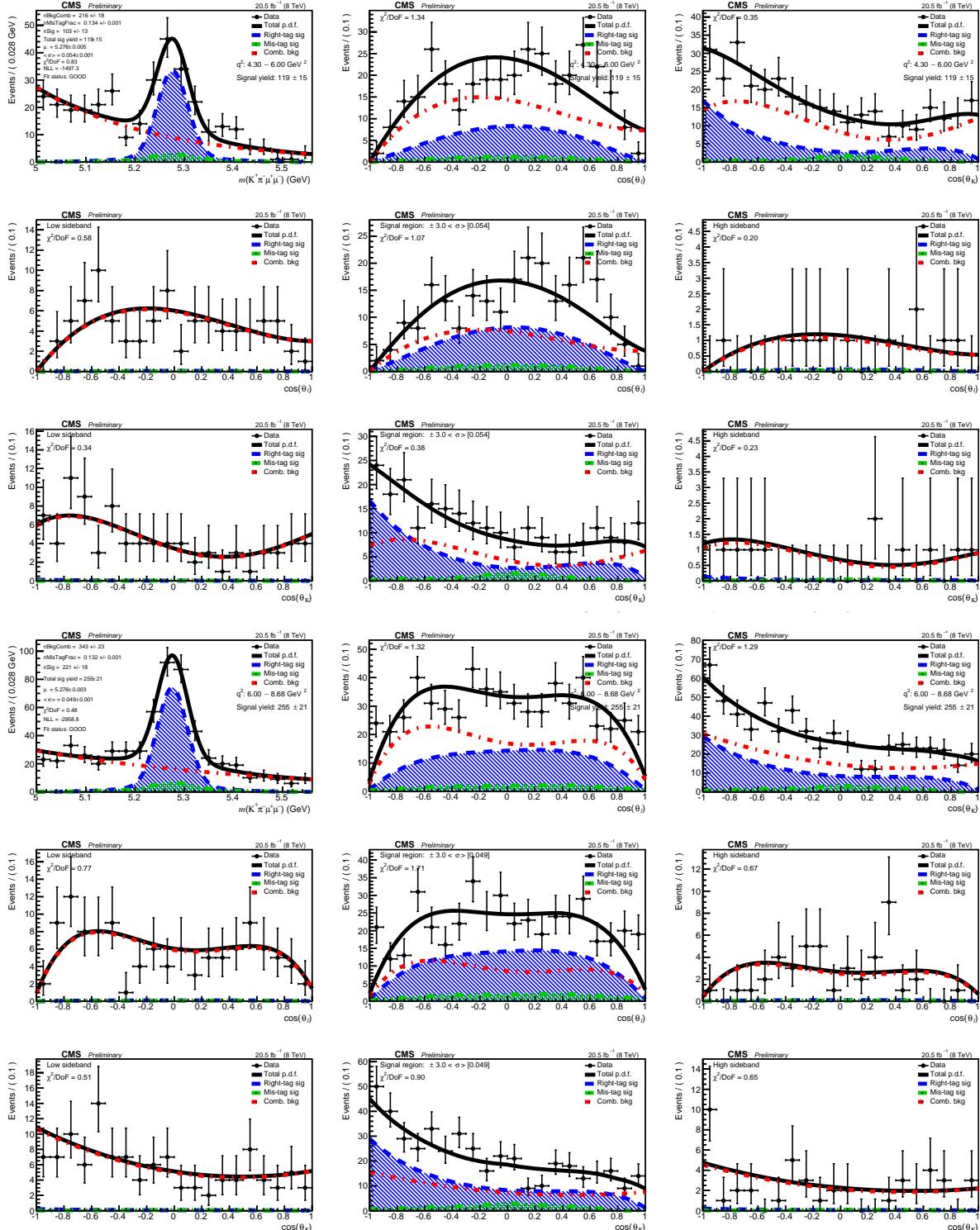


Figure 95: Fit results to measure F_L and A_{FB} with data for the decay $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$, in the dimuon q^2 bin, as from Table 4, #2 (top nine plots) and #3 (bottom nine plots). The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

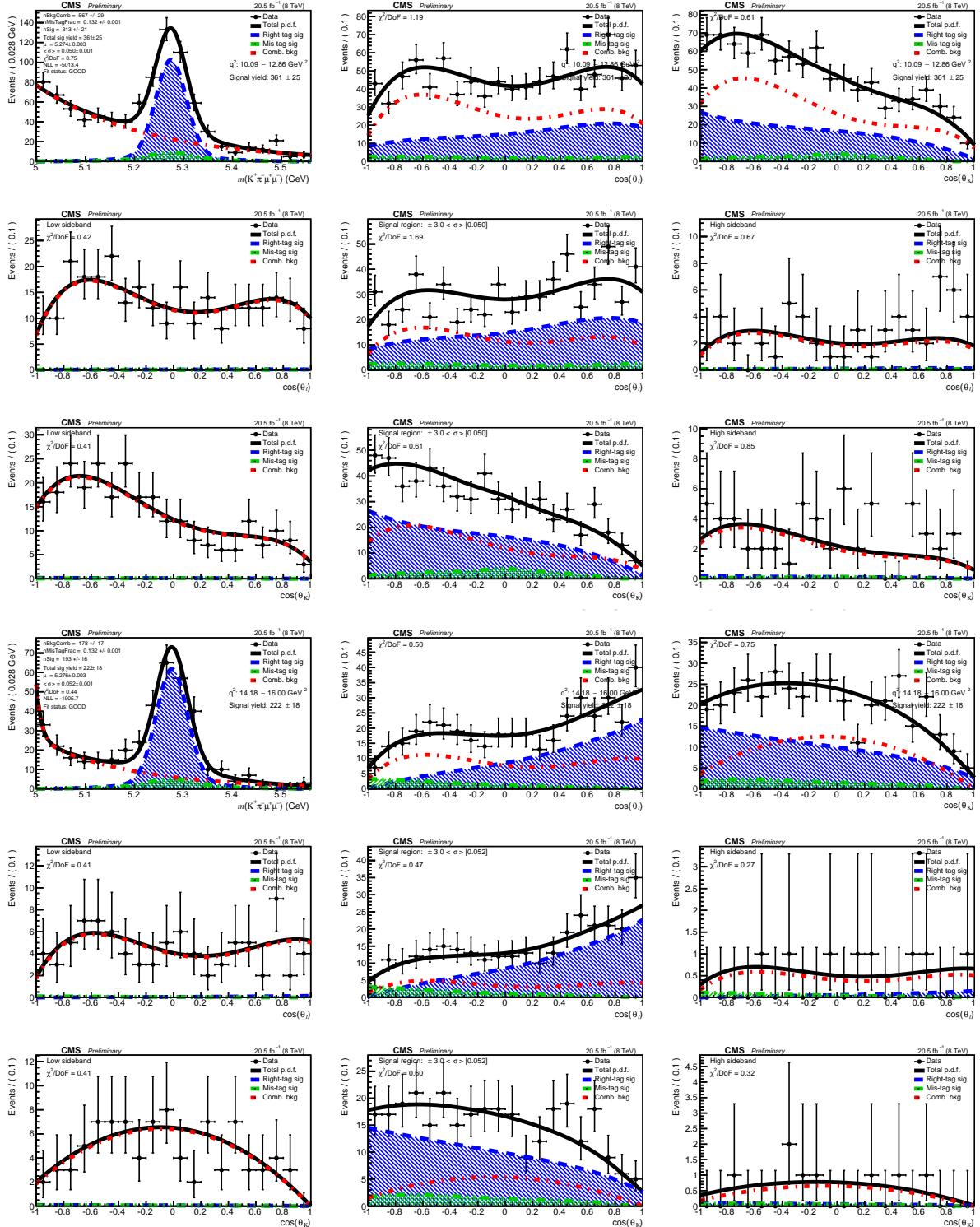


Figure 96: Fit results to measure F_L and A_{FB} with data for the decay $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$, in the dimuon q^2 bin, as from Table 4, #5 (top nine plots) and #7 (bottom nine plots). The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

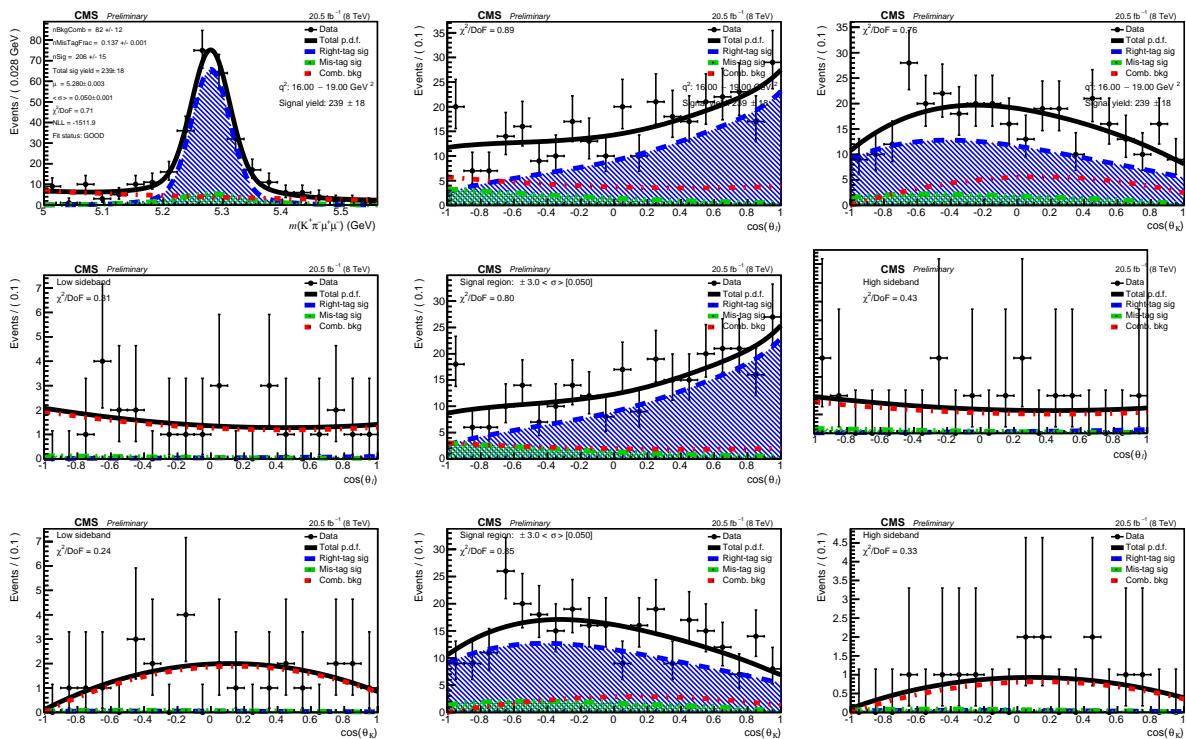


Figure 97: Fit results to measure F_L and A_{FB} with data for the decay $B^0 \rightarrow K^{*0}(K^+\pi^-)\mu^+\mu^-$, in the dimuon q^2 bin, as from Table 4, #8. The degrees of freedom to compute the fit χ^2 are simply the number of bins. The numbering goes from left to right and from top to bottom.

1380 **F Results of the Analysis Compared to SM Predictions with Lat-**
 1381 **tice QCD**

1382 Figures 98 and 99 show results of the measurements compared with SM predictions using lat-
 1383 tice results [28] on $B \rightarrow K^*$ form factors. Such predictions show smaller uncertainties with
 1384 respect to Light Cone Sum Rules (LCSR) [5], however they are subject to the problem of a
 1385 stable K^* on the lattice. The inherent systematic uncertainty of stable K^* should be further in-
 1386 vestigated before one relies on these results alone. On the other hand, the results relying on
 1387 LCSR only, i.e. without lattice input, require an extrapolation of the form factors rather than a
 1388 first principle calculation at very high q^2 , giving rise to larger parametric uncertainties.

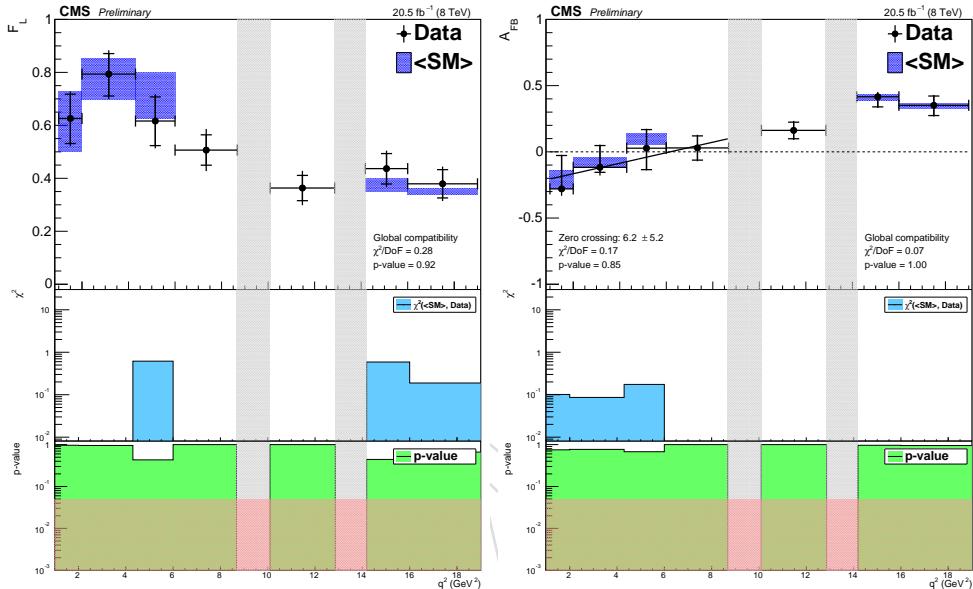


Figure 98: Results of the measurement of F_L (left three plots) and A_{FB} (right three plots). The first plot shows the values as a function of the dimuon q^2 . The black error bars with edges represent the statistical uncertainty while the edgeless error bars represent both statistical and systematic uncertainties (only uncorrelated, see Sec. 10 for more details). The second plot shows the χ^2 of compatibility for each bin between data and SM predictions, while the third plot shows the p-value (in red is shaded the 5% level of significance). The numbering goes from top to bottom. Grey shaded regions correspond to the J/ψ and ψ' resonances. The SM predictions are rate-averaged across the q^2 bins as indicated by the blue regions (the form factors are computed from lattice QCD). No controlled SM prediction is available for the region between the two resonances and for the q^2 bin #3.

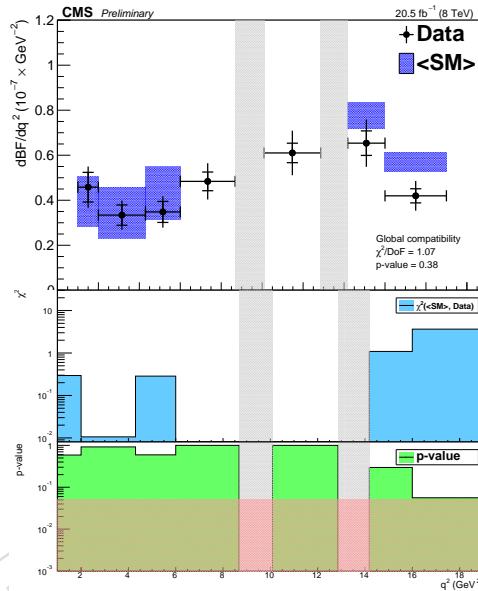


Figure 99: Result of the measurement of $d\mathcal{B}/dq^2$. The first plot shows the values as a function of the dimuon q^2 . The black error bars with edges represent the statistical uncertainty while the edgeless error bars represent both statistical and systematic uncertainties (only uncorrelated, see Sec. 10 for more details). The second plot shows the χ^2 of compatibility for each bin between data and SM predictions, while the third plot shows the p-value (in red is shaded the 5% level of significance). The numbering goes from top to bottom. Grey shaded regions correspond to the J/ψ and ψ' resonances. The SM predictions are rate-averaged across the q^2 bins as indicated by the blue regions (the form factors are computed from lattice QCD). No controlled SM prediction is available for the region between the two resonances and for the q^2 bin #3.

1389 G Results of the Pseudo-Experiments

1390 In this section are presented the results of pseudo-experiments performed in order to study the
 1391 goodness of the unbinned likelihood fits to the data.

1392 Figures 100 and 101 shows the Negative Log-Likelihood distributions, which are then used to
 1393 compute the p-value of the fits (see Sec. 9).

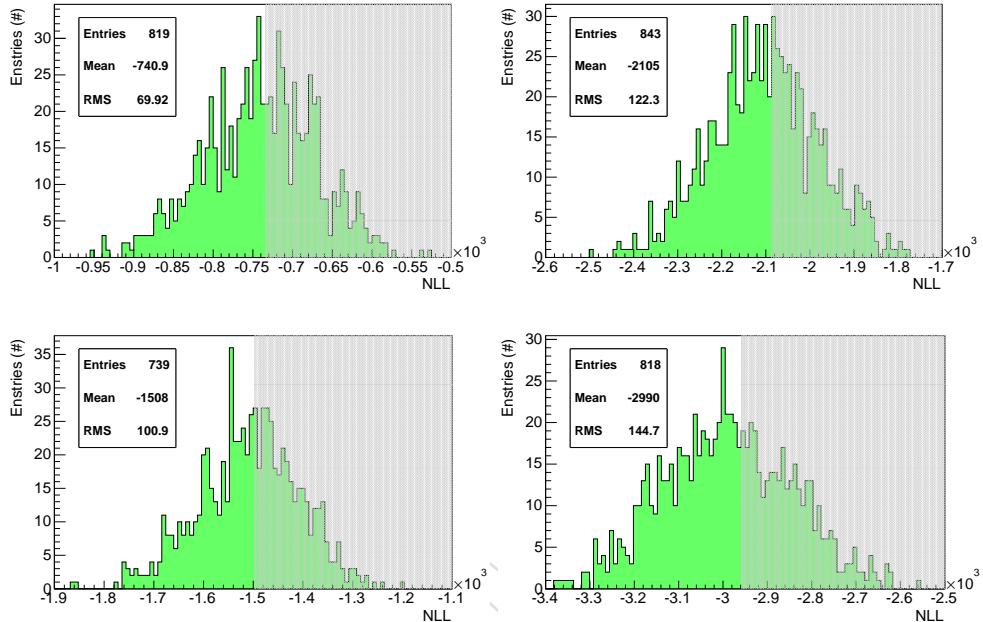


Figure 100: Results of the pseudo-experiments studies. Plots show the Negative Log-Likelihood (NLL) distributions on which it is shaded the region relative to the p-value of the NLL measured with data. They refer to the q^2 bins from #0 to #3 as from Tab. 4. The numbering goes from left to right and from top to bottom.

1394 The pseudo-experiments are generated with p.d.f. parameters and number of events mea-
 1395 sured with data according to the fit-strategy described in Sec. 4. The number of generated
 1396 pseudo-experiments is 1000 (only converged fits are considered, see definition of convergence
 1397 in Sec. 4.6). For each pseudo-experiment the values of the constrained variables are re-generated
 1398 from the constraint distribution.

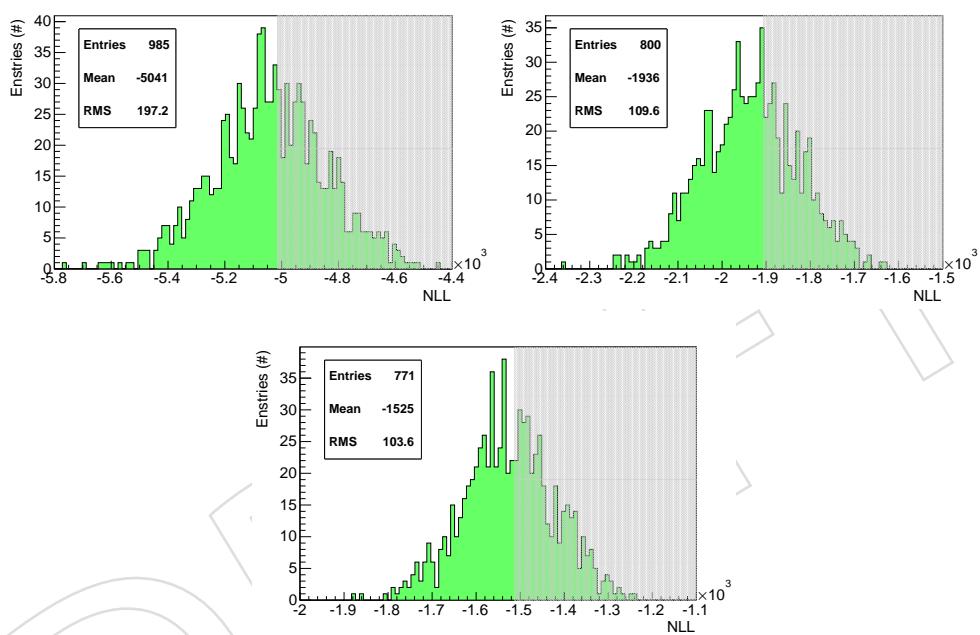


Figure 101: Results of the pseudo-experiments studies. Plots show the Negative Log-Likelihood (NLL) distributions on which it is shaded the region relative to the p-value of the NLL measured with data. They refer to the q^2 bins from #5 to #8, but #6, as from Tab. 4. The numbering goes from left to right and from top to bottom.