LHCb-ANA-2013-090
April 4, 2014

Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays

T. Blake[1], S. Cunliffe[2], U. Egede[2], P. Owen[2], C. Parkinson[3], K. Petridis[2]

[1] *Department of Physics, University of Warwick, Coventry, United Kingdom*

[2] *Imperial College London, London, United Kingdom*

[3] *Fakultät Physik, Technische Universität Dortmund, Dortmund, Germany*

Abstract

The isospin asymmetries of $B \to K\mu^+\mu^-$ and $B \to K^*\mu^+\mu^-$ decays and the partial branching fractions of the $B^0 \to K^0\mu^+\mu^-$, $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$ decays are measured as functions of the dimuon mass squared, q^2. The data used correspond to an integrated luminosity of 3 fb^{-1} from proton-proton collisions collected with the LHCb detector at centre-of-mass energies of 7 TeV and 8 TeV in 2011 and 2012, respectively. The isospin asymmetries are both consistent with the Standard Model expectations. The three measured branching fractions, while individually consistent, all favour lower values than their respective Standard Model predictions.

Contents

1 Introduction

The decays $B \to K^{(*)}\mu^+\mu^-$ are suppressed in the SM as there are no flavour changing neutral currents at tree level. They instead proceed dominantly via electroweak penguin and box diagrams which allow new particles to influence observables by introducing additional diagrams. The most obvious observable to measure is the rates of these decays, which depend on the magnitude of the Wilson coefficients $\mathcal{C}_7$ (electromagnetic), $\mathcal{C}_9$ (semi-leptonic vector) and $\mathcal{C}_{10}$ (semi-leptonic axial-vector). Unfortunately, the SM predictions for rate observables typically suffer from relatively large uncertainties from hadronic form factor calculations. However, recent lattice results [1] have improved the situation substantially meaning improved branching fraction measurements, particularly at high q^2, are highly anticipated. Despite this recent progress, cancelling the leading form factor dependence is essential to maximise the sensitivity to physics beyond the SM, which is achieved by forming ratios of observables. An example of these ratios which is largely insensitive to form factor calculations is the isospin asymmetry, defined as,

$$A_I = \frac{\Gamma(B^0 \to K^{(*)0}\mu^+\mu^-) - \Gamma(B^+ \to K^{(*)+}\mu^+\mu^-)}{\Gamma(B^0 \to K^{(*)0}\mu^+\mu^-) + \Gamma(B^+ \to K^{(*)+}\mu^+\mu^-)}, \tag{1}$$

where $\Gamma(X)$ is the partial width of a particular decay. In terms of branching fractions A_I is,

$$A_I = \frac{\mathcal{B}(B^0 \to K^{(*)0}\mu^+\mu^-) - \frac{\tau_0}{\tau_+}\mathcal{B}(B^+ \to K^{(*)+}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{(*)0}\mu^+\mu^-) + \frac{\tau_0}{\tau_+}\mathcal{B}(B^+ \to K^{(*)+}\mu^+\mu^-)}, \tag{2}$$

where $\mathcal{B}$ is the branching fraction of the decay and $\frac{\tau_0}{\tau_+}$ is the ratio of the lifetimes of the B^0 and B^+ mesons. The SM prediction for A_I is around -1% in the di-muon mass squared (q^2) region below the J/ψ resonance [2]. Although there is no precise prediction for A_I at high q^2, Ref [2] claims that it is also expected to be close to zero. The small isospin asymmetry predicted in the SM is due to initial state radiation of the spectator quark, which is different between the neutral and charged decays. Previously, A_I has been measured to be significantly below zero in the q^2 region below the J/ψ resonance [3, 4]. In particular, the combined $B \to K\mu^+\mu^-$ and $B \to K^*\mu^+\mu^-$ isospin asymmetries measured by the BaBar experiment were 3.9 σ below zero. For $B \to K^*\mu^+\mu^-$, A_I is expected to be consistent with the $B \to K^{*0}\gamma$ measurement of $5 \pm 3\%$ [5] as q^2 approaches zero. No such constraint is present for $B \to K\mu^+\mu^-$. With the 2011 dataset, the LHCb collaboration measured evidence for isospin asymmetry in $B^+ \to K^+\mu^+\mu^-$, shown in Fig. 1, a result which is yet to be understood in any physics model [6].

This analysis presents updates to the branching fraction of $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^0\mu^+\mu^-$ and $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ decays, as well as the isospin asymmetry of $B \to K\mu^+\mu^-$ and $B \to K^*\mu^+\mu^-$ decays. All results use $3\,\text{fb}^{-1}$ of integrated luminosity - the entire 2011+2012 dataset.

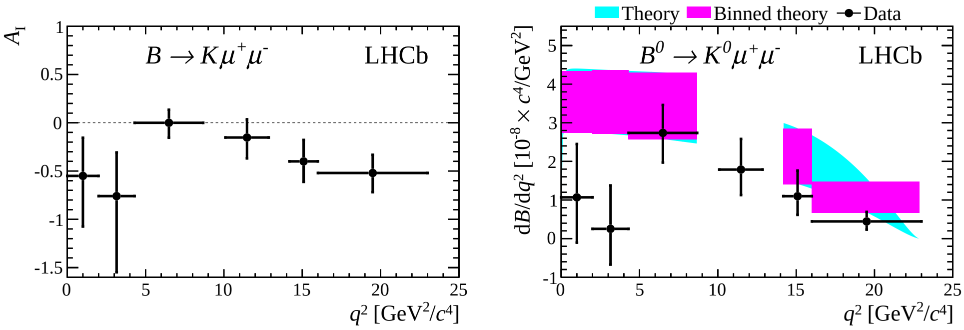


Figure 1: Isospin asymmetry of $B \to K\mu^+\mu^-$ (left) and differential branching fraction of $B^0 \to K^0\mu^+\mu^-$ (right) as measured with 2011 data in Ref. [6]. For the isospin asymmetry there is a 4 σ deviation from the naive SM expectation of zero.

2 Strategy

The goal of the analysis is to measure the branching fractions of $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ as well as the isospin asymmetry of $B \to K\mu^+\mu^-$ and $B \to K^*\mu^+\mu^-$. For the rest of this note, the $B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ decays are known as the "K^0_S channels", whereas the $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ channels are known as the "K^+ channels". As the statistics of the K^0_S channels are low, the dataset is not split into q^2 bins yet - the yields in bins of q^2 are blinded. Most sections related to the selection, backgrounds are divided into two sub sections, for $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^0\mu^+\mu^-$, $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ and $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ respectively.

For the branching fractions and isospin asymmetries, the strategy follows the previous isospin asymmetry measurement, where signal yields are converted into a branching fractions by normalising to the resonant case, $B \to J/\psi\, h$ where $J/\psi \to \mu^+\mu^-$. These $B \to J/\psi\, h$ decays have well known branching fractions and have the same final state particles as the signal decays.

For the K^0_S channels, the analysis is split between the LL and DD K^0_S reconstruction categories. The definition of a LL K^0_S is one where its daughter pions have been reconstructed inside the VELO, whereas a DD K^0_S is one where its daughters are reconstructed downstream of the VELO. LL and DD K^0_S behave very differently in efficiency and selection and so there are two separate measurements for LL and DD K^0_S which are combined for the result. What this means in practice is that there is separate selections and normalisations for the LL and DD modes.

The selection, described in Sect. 4, is based on a Boosted Decision Tree (BDT) method, where the signal sample is data corrected simulation and the background sample is the extreme upper mass sideband, which is defined as 5700-6000 MeV/c^2. A cut is placed on the BDT which optimises $S/\sqrt{(S+B)}$, where S is the expected signal yield given signal efficiency on simulation and B is the background extrapolated into the signal window from the upper mass sideband.

As for backgrounds, there is very little that can trouble these decays. The mass resolution is good enough to avoid partially reconstructed backgrounds by starting the fit range at $5170\,\mathrm{MeV}/c^2$. Fully reconstructed backgrounds, where one of more particles can be mis-identified, are described in Sect. 5, where all backgrounds apart from in the $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ case are considered to be negligible after specific vetoes have been applied.

The signal yields are determined in each q^2 bin using an unbinned extended maximum likelihood fit to the $K^{(*)}\mu^+\mu^-$ mass. Again, the J/ψ modes help here, by providing a very good signal shape proxy. The J/ψ signal shape and a small correction obtained from the simulation is used to fit the signal. The correction has a very small effect on the signal yield.

The relative efficiency is calculated using fully simulated events in each bin of q^2. Variables which are poorly represented in the simulation were corrected using data. The process of normalisation is discussed in section 7.

For the branching fractions and isospin asymmetries, the normalisation procedure cancels a large fraction of systematic uncertainties, as most variables are only weakly correlated to q^2. The systematics are described in Sect. 9.

2.1 Differences to the previous analysis

Whilst the strategy of the analysis is similar to the 2011 analysis [7], in general, effort has been devoted to make the analysis simpler. See below for specific differences between this round and last time.

- The selection for the K^0_S channels have been retrained, due to the larger dataset available. The background sample is no longer a percentage of the sideband and is rather the extreme upper sideband instead, which avoids the need to remove part of the data for the result. When this training strategy was changed there was no visible difference in the behaviour for $B^+ \to K^+\mu^+\mu^-$, apart from an increase in performance due to the larger training samples.

- There is no longer effort placed to make the K^0_S and K^+ selections the same anymore. The $B^+ \to K^+\mu^+\mu^-$ selection is taken from Ref. [8] and the $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ selection is taken from Ref. [9].

- The trigger for the K^+ channels is no longer restricted to the dimuon pair, as the trigger is very well behaved for the K^+ channels.

- Λ^0 reflections are vetoed using a mass requirement, rather than a conditional PID cut. Downstream tracks will have poor PID performance, which was not taken into account last time.

- The q^2 binning is devised to avoid all J/ψ and $\psi(2S)$ radiative and mis-reconstructed tails. This avoids having to implement a complicated charmonium veto.

- The mass shape for the signal is now fixed from normalisation channels (rather than constrained), as this turned out have a negligible effect and only served to make the fits more unreliable.

- There is no longer a component for partially reconstructed backgrounds included in the mass fit, as the background is at a much lower level this time.

- There was an issue in the previous analysis, where the treatment of the K^0_S in the "DecProdCut" efficiency was misunderstood. It was assumed that the K^0_S had the same "DecProdCut" as charged particles, but in fact DecProdCut is *not* applied to K^0_S mesons. This has been corrected this time round, however the effect is well below the statistical sensitivity of the previous analysis.

- An additional background for $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ decays was considered, from $B^+ \to J/\psi(K^{*+} \to K^0_S\pi^+)$ with a pion-muon swap.

3 Data samples

The data used for this analysis were taken in 2011 and 2012 and correspond to an integrated luminosity of about $3\,\text{fb}^{-1}$. The data samples used are stripped with stripping versions 20 and 20r1 for 2012 (Reco 14) and 2011 (Reco 14) data, respectively. All candidates are taken from the "B2XMuMu" stripping line.

The simulation samples are a mixture of MC11 (Reco 12) and MC2012 (Reco 14). For $B^+ \to K^+\mu^+\mu^-$, only MC11 is used as it agrees very nicely with both 2011 and 2012 data as from Ref. [8]. For the K^0_S channels, MC2012 samples were generated to cross-check the K^0_S reconstruction efficiency under different beam conditions and material description. The trigger for $B^+ \to K^+\mu^+\mu^-$ decays, two TCKs are simulated, 0x40760037 for MC11 and 0x40990042. These TCKs are very well representative of the data as there was hardly any changes to the muon lines throughout 2011 or 2012.

For the K^0_S channels, MC2012 is used for normalising 2012 data, which is used to demonstrate the stability of the efficiency under different reconstruction versions. For $B^0 \to K^0_S\mu^+\mu^-$, an additional TCK is simulated, 0x4097003d, introduced due to two K^0_S bugs introduced into the HLT during 2012 running. The first affects DD K^0_S during the first $\sim 0.5\,\text{fb}^{-1}$, first described in Ref. [10], where the TOS efficiency for DD K^0_S was very low. The second affects LL K^0_S in the Hlt2 Topogical trigger for the last $1.5\,\text{fb}^{-1}$ of 2012 data, and is described in Ref. [11]. The splitting between the first $0.5\,\text{fb}^{-1}$ and last $1.5\,\text{fb}^{-1}$ in 2012 for simulation is referred to as "2012 early" and "2012 late" respectively. Note that somewhat confusingly, "2012 early" simulation sample is actually MC11 for $B^0 \to K^0_S\mu^+\mu^-$ with the TCK 0x4097003d applied, whereas "2012 late" is an MC2012 sample. For $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ hardly any candidates are trigger via the K^0_S and so only the two TCKs are used (same as the K^+ channels).

The physics model used for all the decays is based on the form factors described in [12]. Although this model is quite old, the data is split into q^2 bins which reduces the dependence on the model. In any case, the physics model dependence is explored in Sect. 9.

Channel	Version	Events in sample	Physics Model
$B^+ \to K^+\mu^+\mu^-$	MC11	1M	BTOSLLBALL
$B^+ \to J/\psi\, K^+$	MC11	600K	SVS
$B^+ \to K^+\pi^+\pi^-$	MC11	1.7 M	PHSP
$B^0 \to K^0_S\mu^+\mu^-$	MC11	5M	BTOSLLBALL
$B^0 \to J/\psi\, K^0_S$	MC11	2M	SVS
$B^0 \to K^0_S\mu^+\mu^-$	MC2012	2.5M	BTOSLLBALL
$B^0 \to J/\psi\, K^0_S$	MC2012	2M	SVS
$B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$	MC11	10M	BTOSLLBALL
$B^+ \to J/\psi\,(K^{*+} \to K^0_S\pi^+)$	MC11	1M	SVV_HELAMP
$B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$	MC2012	2M	BTOSLLBALL
$B^+ \to J/\psi\,(K^{*+} \to K^0_S\pi^+)$	MC2012	600K	SVV_HELAMP
$B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$	MC11	500K	BTOSLLBALL
$B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$	MC11	500K	SVV_HELAMP
$B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$	MC2012	1M	BTOSLLBALL
$B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$	MC2012	2M	SVV_HELAMP

Table 1: Monte Carlo simulation samples used in the analysis. The physics model BTOSLLBALL is based on the form factors described in [12]. The SVV_HEPAMP model used for J/ψ modes involving a K^* takes in helicity amplitudes for the daughters. In general the physics model for the J/ψ modes makes little difference to efficiency derived from it.

4 Selection

The $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^0_S\mu^+\mu^-$ parts of this section are identical to the one in Ref. [13].

4.1 Stripping

Both decays are stripped using the B2XMuMu stripping line with version S20 and S20r1, which place the following requirements:

Candidate	Selection
B meson	IP $\chi^2 < 16$ (best PV)
B meson	$4600\,\text{MeV}/c^2 < M < 7000\,\text{MeV}/c^2$
B meson	DIRA angle < 14 mrad
B meson	Flight Distance $\chi^2 > 121$
B meson	Vertex $\chi^2/\text{ndf} < 8$
DiMuon	$M < 7100\,\text{MeV}/c^2$
DiMuon	Vertex $\chi^2/\text{ndf} < 9$
All Tracks	Clone Distance > 5000
Long Tracks	Ghost Prob < 0.4
K/π tracks from B vertex	min IP $\chi^2 > 9$
Muon	IP $\chi^2/\text{ndf} > 9$
Muon	isMuon == True
Muon	$\text{DLL}_{\mu\pi} > -3$
K^0_S	$\tau > 2\,\text{ps}$ (PDG = 90 ps)
K^0_S	$467\,\text{MeV}/c^2 < M < 527\,\text{MeV}/c^2$
GEC	SPD Mult. < 600

Table 2: The stripping selection criteria.

4.2 Trigger requirements

We place requirements associated with the trigger according to Table 3. Candidates are required to be TOS in all stages of the trigger. The dominant HLT2 lines are DiMuonDetached and MuTopo2Body, which is what one would expect as the muons dominate the trigger rate for these decays.

4.3 Pre-selection

4.3.1 Fiducial cuts

Fiducial cuts are applied to remove contamination from charmonium resonances, and partially reconstructed backgrounds. For $B^+ \to K^+\mu^+\mu^-$ there is an additional veto

Level	Requirement
L0	L0Muon
HLT1	TrackMuon or TrackAll
	DiMuonLow or DiMuonHigh
HLT2	TopoMu2BodyBDT or TopoMu3BodyBDT
	Topo2BodyBDT or Topo3BodyBDT
	SingleMuon or DiMuonDetached
	DiMuonDetachedHeavy

Table 3: Trigger requirements. For each level, candidates are required to be TOS on at least one line.

around the ϕ mass, to remove $B^+ \to (\phi \to \mu^+\mu^-)K^+$ decays. For the K^0_S channels no such veto is applied as the level of the resonance is well below the statistical sensitivity. The q^2 bins used to seperate the charmonium resonances from the signal are shown in Tab. 4. In addition to the q^2 binning, the B mass window is defined as $5170 - 5700\,\text{MeV}/c^2$.

Decay	Binning scheme (GeV/c^2)
K^0_S channels	0.1,2.0,4.0,6.0,8.0,11.00,12.50,15.00,17.00,22.00
$B^+ \to K^+\mu^+\mu^-$	0.1,0.98,1.1,2.0-8.0 in $1\,\text{GeV}/c^2$ steps,11.00,11.75,12.50,15.00-22.00 in $1\,\text{GeV}/c^2$ steps

Table 4: Binning schemes for the K^0_S channels and $B^+ \to K^+\mu^+\mu^-$, both isospin asymmetry measurements use the same binning scheme as the K^0_S channels. The $B^+ \to K^+\mu^+\mu^-$ channel has roughly three times as many q^2 bins due to the higher statistics available.

4.3.2 $B^+ \to K^+\mu^+\mu^-$

We place PID requirements to suppress peaking backgrounds (as discussed in Sect. 5). They also suppress a large amount of combinatorial background where a random pion is misidentified as the kaon in the $B^+ \to K^+\mu^+\mu^-$ decay. The PID requirements are listed in Table 5 (the criteria in red are inherited from the stripping) and have a signal efficiency of $\sim 90\%$. The PID efficiencies are calibrated from the data using two methods, described in more detail in Sect. 7.

Particle	PID requirement
Kaon	(ProbNNK – ProbNNpi) > -0.5 and ProbNNK > 0.05
Muons	isMuon $== 1$ and $\text{DLL}_{\mu\pi} > -3$ and ProbNNmu > 0.25

Table 5: PID requirements for $B^+ \to K^+\mu^+\mu^-$. Criteria in red is inherited from the stripping.

4.3.3 $B^0 \to K^0_S \mu^+ \mu^-$

Compared to $B^+ \to K^+ \mu^+ \mu^-$, PID is much less important for the $B^0 \to K^0_S \mu^+ \mu^-$ decay as the background is dominated by a real K^0_S and real muons. This is due to the tight BDT selection compared to the other modes, which are ∼90% efficient. No additional PID requirements are used after the stripping. The K^0_S mass is constrained which improves the signal resolution of the DD modes by $\sim 1\,\text{MeV}/c^2$ and the LL modes by a negligible amount.

4.3.4 $B^+ \to (K^{*+} \to K^0_S \pi^+) \mu^+ \mu^-$

Unlike $B^0 \to K^0_S \mu^+ \mu^-$, some fake muons are present in the background sample for $B^+ \to (K^{*+} \to K^0_S \pi^+) \mu^+ \mu^-$. Here we apply an additional ProbNNmu cut on both muons of 0.25 - the same as $B^+ \to K^+ \mu^+ \mu^-$. There is a $892 \pm 100\,\text{MeV}/c^2$ mass window around the $K^0_S\ \pi^+$ mass for the signal, which is the same as the previous analysis. Again, the K^0_S mass is constrained which improves the signal resolution of the DD modes by $\sim 1\,\text{MeV}/c^2$ and the LL modes by a negligible amount.

4.3.5 $B^0 \to (K^{*0} \to K^+ \pi^-) \mu^+ \mu^-$

The decay $B^0 \to (K^{*0} \to K^+ \pi^-) \mu^+ \mu^-$ suffers from a large array of peaking backgrounds, discussed in more detail in Sect. 5. To help suppress these, and combinatorial background, the PID requirements detailed in Table 6 are applied.

Particle	PID requirement
Kaon	$\text{DLL}_{K\pi} > -5$
pion	$\text{DLL}_{K\pi} < 25$
Kaon - pion	$\text{DLL}_{K\pi}$ difference > 10
muon	$\text{DLL}_{\mu\pi} > -3$

Table 6: PID requirements for $B^0 \to (K^{*0} \to K^+ \pi^-) \mu^+ \mu^-$. Criteria in red is inherited from the stripping.

4.4 Multivariate offline selection

The main selection is based on a Boosted Decision Tree (BDT) [14], with the Adaboost algorithm [15]. Only geometric and kinematic variables are included (see Tab. 10), which are generally well modelled in the simulation as shown in Sect. A in the appendix. The MVA is trained and tested on candidates with a $K\mu^+\mu^-$ mass between 5700-6000 MeV/c^2 as a background sample, which is not used for the rest of the analysis. The BDT is trained on, $B^+ \to K^+ \mu^+ \mu^-$ simulated events for a signal sample. This simulation sample is data corrected according to the procedure described in Sect. 7. This correction procedure is carried out mainly for optimisation reasons, as for the efficiency, such effects cancel with the normalisation to $B^+ \to J/\psi K^+$. The reason the BDT is trained on simulation rather than $B^+ \to J/\psi K^+$ is to minimise biases favouring dimuon masses close to the J/ψ mass. Two-thirds of the signal and background samples are used for training and one third is used for testing. The testing samples are to check overtraining and to optimise the cut placed on the BDT. The agreement of the BDTs between data and simulation is very good, see Sect. A in the appendix.

4.4.1 $B^+ \to K^+\mu^+\mu^-$

Variable
K IP χ^2 (minimum)
B^+ vertex χ^2
μ IP χ^2 (minimum)
B^+ $p_{\rm T}$
J/ψ IP χ^2 (minimum)
B^+ IP (best PV)
B^+ Flight distance χ^2
B^+ DIRA angle
μ IP χ^2 (minimum)
K^+ P
B^+ P

Table 7: Variables used in the $B^+ \to K^+\mu^+\mu^-$ BDT ordered by importance according to TMVA. There are two types of impact parameter (IP) variables used, one defined with respect to best PV (Primary Vertex) and the other defined as the minimum IP with respect to all PVs.

A cut is placed on the BDT to maximise $\mathrm{S}/\sqrt{(\mathrm{S}+\mathrm{B})}$, where S is the number of expected $B^+ \to K^+\mu^+\mu^-$ based on the number of $B^+ \to J/\psi K^+$ seen in data and B is the background extrapolated into the signal window. The assumed branching fraction is the one measured in Ref. [16]. This optimisation procedure is shown in Fig. 3 and a cut is chosen at 0.3. The efficiency of this BDT cut on $B^+ \to K^+\mu^+\mu^-$ signal is 89%, whereas the efficiency for background is 6%. After this selection, the signal is very clean (see Sect.6).

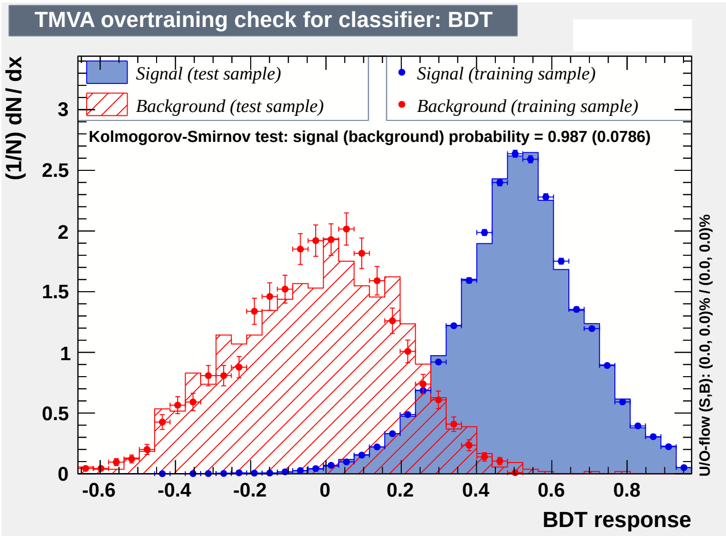


Figure 2: BDT response for $B^+ \to K^+\mu^+\mu^-$ signal (blue) and background (red). The testing and training samples are overlaid which shows a negligible amount of overtraining.

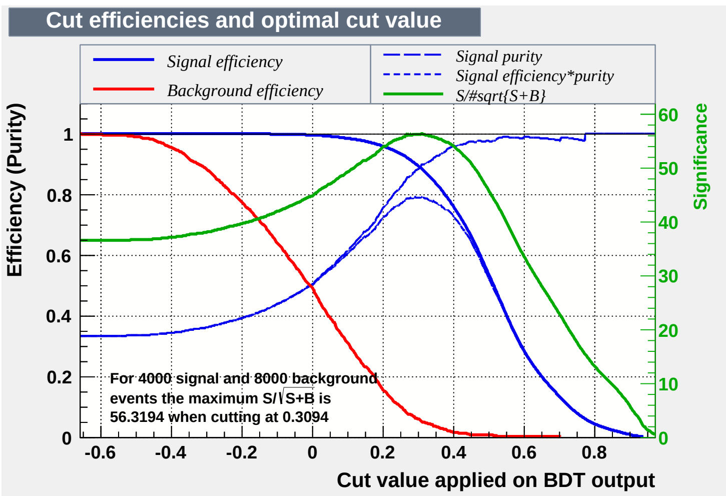


Figure 3: Metric for the signal as a function of $B^+ \to K^+\mu^+\mu^-$ BDT cut. The optimal cut is chosen to be 0.3.

4.4.2 $B^0 \to K^0_S \mu^+\mu^-$

LL	DD
K^0_S IP χ^2 (min)	K^0_S p_T
B^0 vertex χ^2	B^0 IP (best PV)
B^0 DIRA angle	B^0 vertex χ^2
B^0 p_T	μ IPχ^2 (min)
B^0 IP χ^2 (minimum)	$B^0 \tau$
B^0 P	B^0 p_T
K^0_S p_T	B^0 P
B^0 IP (best PV)	B^0 DIRA angle
B^0 τ	K^0_S P

Table 8: Variables used in the $B^0 \to K^0_S \mu^+\mu^-$ BDT ordered by importance according to TMVA. There are two types of impact parameter (IP) variables used, one defined with respect to best PV (Primary Vertex) and the other defined as the minimum IP with respect to all PVs.

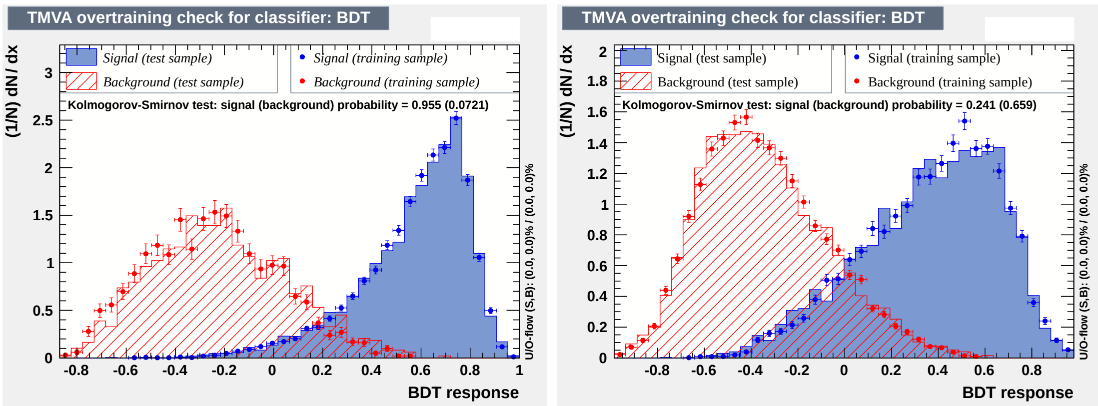


Figure 4: BDT response for the signal (blue) and background (red) in both LL(left) and DD(right) categories. The testing and training samples are overlaid which show a negligible amount of overtraining.

Cuts are placed on the BDTs to maximise $\mathrm{S}/\sqrt{(\mathrm{S}+\mathrm{B})}$, where S is the number of expected $B^0 \to K^0_S \mu^+\mu^-$ based on the number of $B^0 \to J/\psi K^0_S$ seen in data and B is the background extrapolated into the signal window. The assumed branching fraction is the one measured in Ref. [6]. The optimisation procedures are shown in Fig. 5 and a cut is chosen at 0.5(0.45) for the LL(DD) category. The efficiency of this BDT cut on $B^0 \to K^0_S \mu^+\mu^-$ signal is 66(48)%, whereas the efficiency for background is 1.3(0.2)%.

Unlike $B^+ \to K^+\mu^+\mu^-$, the selection for $B^0 \to K^0_S \mu^+\mu^-$ is not trivial. This is due to the fact that the K^0_S flies and cannot be easily distinguished from originating from the B^0

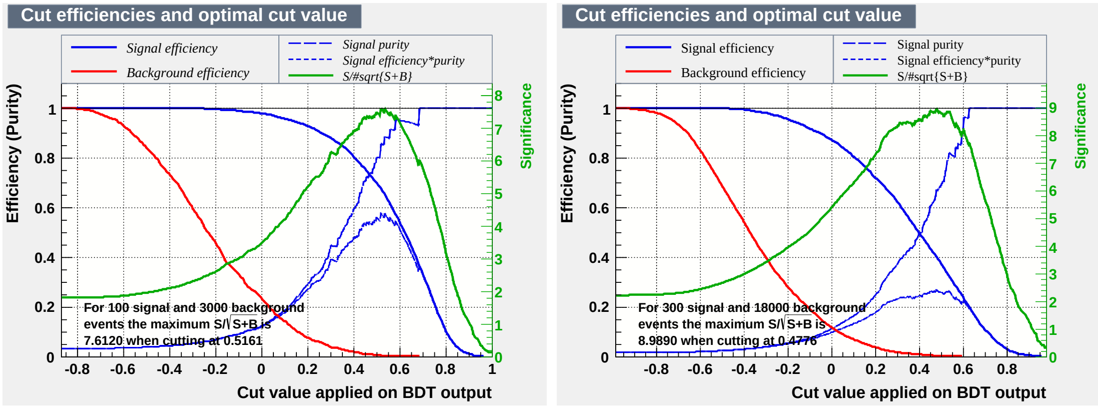


Figure 5: Among other curves, the metric in green as a function of the $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ BDT cut for the LL category (left) and DD category (right).

vertex. For the LL category, there is some separation through the K^0_S IPχ^2, which is the most important variable according to TMVA. For the DD category it is comparatively easy to randomly combine a K^0_S from the PV to form combinatorial background, which is why the BDT is less efficient for this category. The performance of the selection is slightly improved compared to the 2011, discussed in Sect. 6.

The correlation of the BDT with mass is checked by calculating the average BDT response for the sideband at different candidate B masses. The result of this study is shown in Fig. 6, where the BDT response is flat is mass. For the other channels, any possible BDT correlation is not an issue as the BDT is very efficient in those cases.

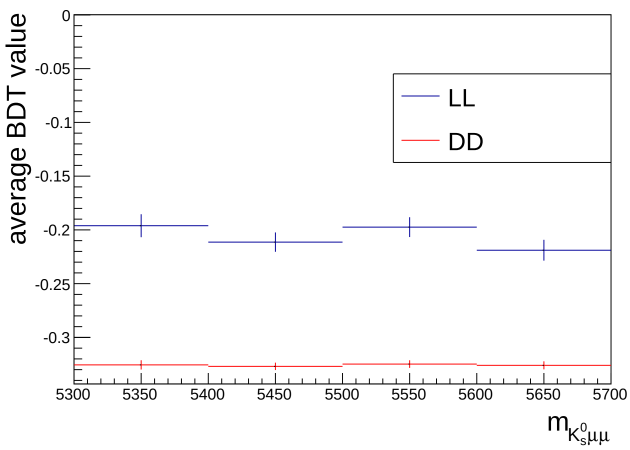


Figure 6: Average BDT response for sideband as a function of $m_{K^0_S\mu^+\mu^-}$. The BDT is uncorrelated to mass.

The candidate K^0_S mass in the signal region is shown in Fig. 7 after the full selection.

Although the statistics are small, it is clear the data is dominated by true $K_{\rm S}^0$.

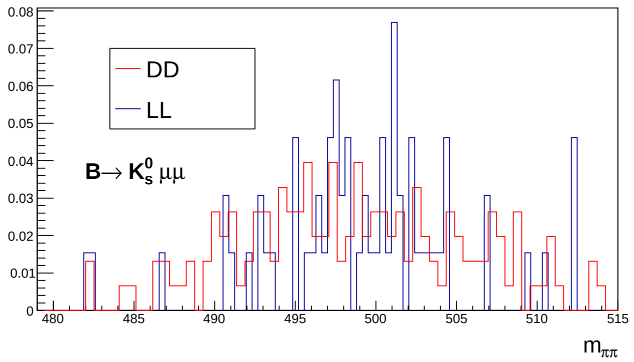


Figure 7: Candidate $K_{\rm S}^0$ mass distribution of $B^0 \to K_{\rm S}^0\mu^+\mu^-$ candidates in the signal region.

4.4.3 $B^+ \to (K^{*+} \to K^0_S \pi^+)\mu^+\mu^-$

Like $B^+ \to K^+\mu^+\mu^-$, the $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ BDT is trivial, where the efficiency is over 90% efficient for both categories. In order to boost statistics, the background sample is taken from a wider $m_{K\pi}$ mass window, around 300 MeV/c^2 of the K^{*+} PDG mass. The variables used in the $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ BDTs, ranked according to TMVA, are shown in Tab. 9, the BDT signal and background distributions are shown in Fig. 8 and the optimisation procedure is shown in Fig. 9. There is a peak at very high BDT values which is not present in the other channels. This is due to a small population of candidates which have very good pointing (DIRA ¿ 0.999999) and vertex quality (χ^2 ¡ 2) requirements. This region has no background left in it, due to the finite size of the training samples (the $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ LL mode has the smallest training sample). Cuts are placed on the BDTs to maximise $\mathrm{S}/\sqrt{(\mathrm{S}+\mathrm{B})}$, where S is the number of expected $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ based on the number of $B^+ \to J/\psi\,(K^{*+} \to K^0_S\pi^+)$ seen in data and B is the background extrapolated into the signal window. The assumed branching fraction is the one measured in Ref. [6].

LL	DD
B^+ DIRA angle	B^+ p_T
B^+ vertex χ^2	μ IP χ^2 (min)
B^+ Flight distance	B^+ DIRA angle
B^+ p_T	π IPχ^2 (min)
π IPχ^2 (min)	$K^{*+}p_\mathrm{T}$
K^0_S IPχ^2 (min)	B^+ vertex χ^2
K^0_S p_T	B^+ IP (best PV)
B^+ IP (best PV)	π p_T
π p_T	K^0_S p_T
	K^{*+} IP χ^2

Table 9: Variables used in the $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ BDT ordered by importance according to TMVA. There are two types of impact parameter (IP) variables used, one defined with respect to best PV (Primary Vertex) and the other defined as the minimum IP with respect to all PVs.

The candidate K^0_S and K^{*+} masses in the signal region are shown in Fig. 10 after the full selection. Although the statistics are small, it is clear the data is dominated by true K^0_S and K^{*+} decays.

4.4.4 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$

The $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ BDT is based on Ref. [9]. Its very efficient (∼90%), and does not bias q^2, as shown in Sect. 7. The inclusion of PID in the BDT results in complications regarding the efficiency when applied to the simulation, however given how the flat the efficiency is, this makes no difference for the analysis.

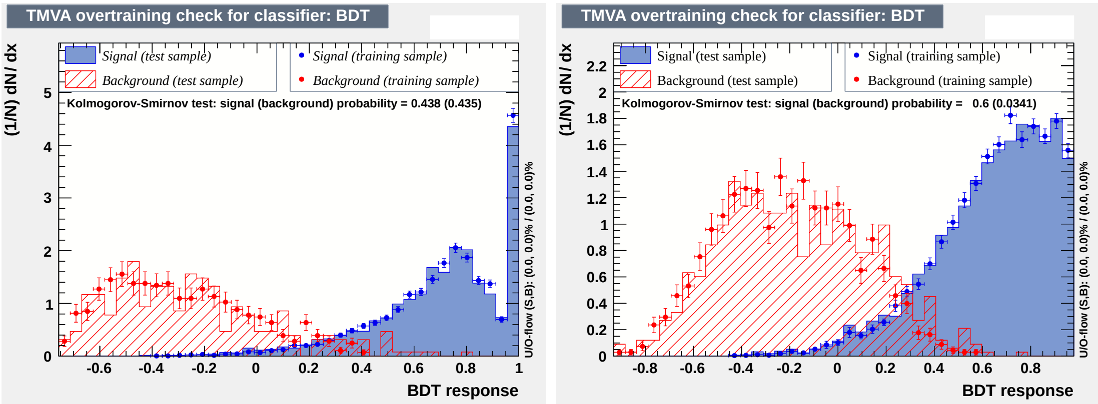


Figure 8: BDT response for the signal (blue) and background (red) in both LL(left) and DD(right) categories. The testing and training samples are overlaid which show a negligible amount of overtraining.

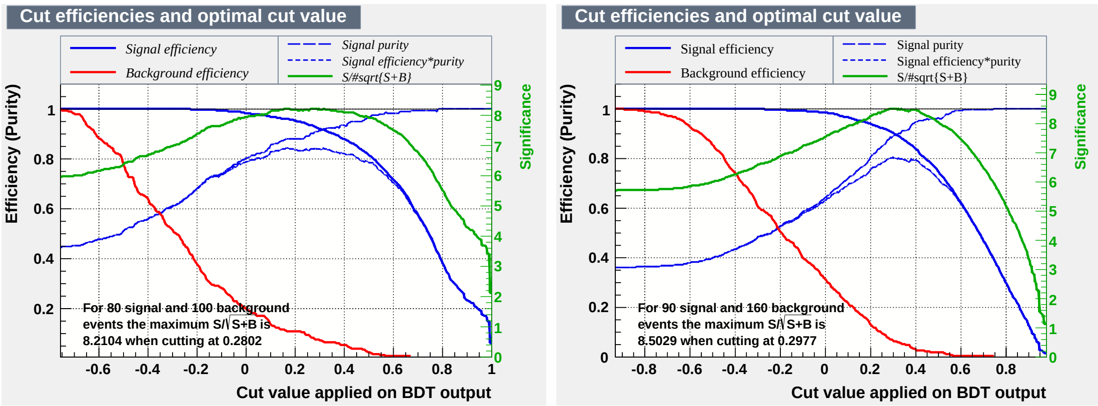


Figure 9: Among other curves, the metric in green as a function of the $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ BDT cut for the LL category (left) and DD category (right). The chosen cut for both is at 0.3.

4.5 Multiple Candidates

After the previous selection has been applied, multiple candidates are removed randomly in a reproducible way. The percentage of multiple candidates is below 0.05% at the end of the selection chain. As this is such a low fraction, no systematic uncertainty is assigned.

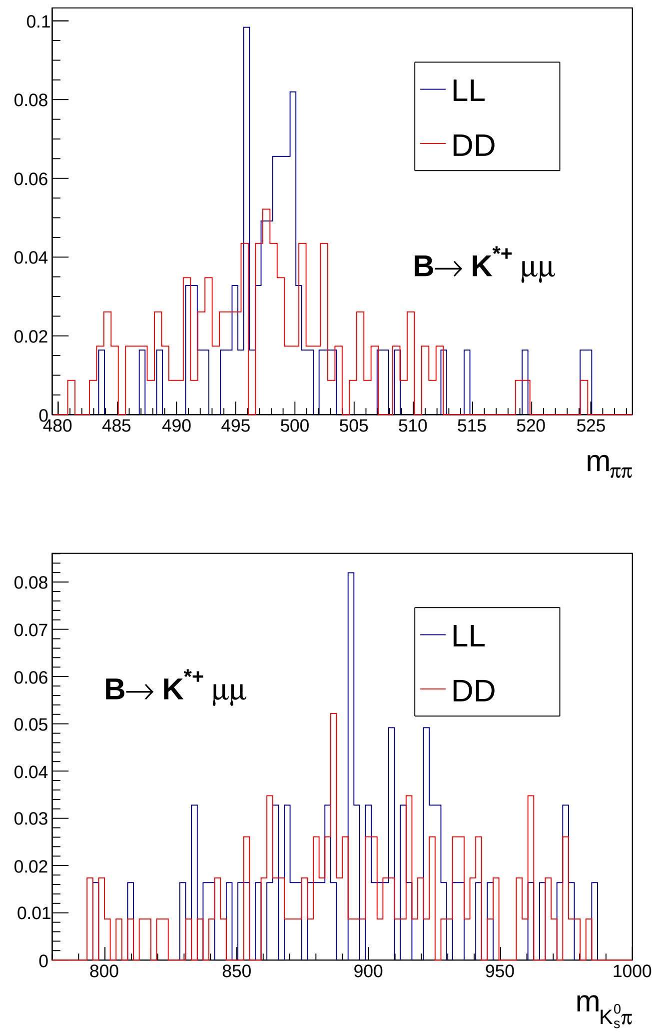


Figure 10: Candidate $K^0_{\rm S}$ (left) and K^{*+} (right) mass distributions of $B^+ \to (K^{*+} \to K^0_{\rm S}\pi^+)\mu^+\mu^-$ candidates in the signal region.

Variable
B^0 τ
B^0 vertex χ^2
B^0 $p_{\rm T}$
B^0 DIRA angle
B^0 P
K $\text{DLL}_{K\pi}$
π $\text{DLL}_{K\pi}$

Table 10: Variables used in the $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ BDT in no particular order.

5 Backgrounds

In this section peaking backgrounds are considered. All backgrounds discussed peak in the signal region, however they are assumed to be negligible after specific vetoes have been applied.

5.1 $B^+ \to K^+\mu^+\mu^-$

This section is identical to Sect. 6 in Ref. [8].

5.1.1 $B^+ \to J/\psi\, K^+$ with kaon-muon swap

The decay $B^+ \to J/\psi\, K^+$ has roughly 150 times the branching fraction of the signal and can fall out of the J/ψ veto if the kaon swaps mass hypothesis with the same-sign muon. This background is almost completely rejected by computing the $\mu^- K^+$ mass under the $\mu^-\mu^+$ hypothesis. If this mass is within $60\,\mathrm{MeV}/c^2$ of the J/ψ or $\psi(2S)$ resonances then the kaon is required to be in the muon acceptance, but fail the isMuon flag. This is over 99% efficient on signal.

There is no visible structure in the $\mu^- K^+$ mass after this veto is applied (see Fig. 11), any remaining background is assumed to be negligible.

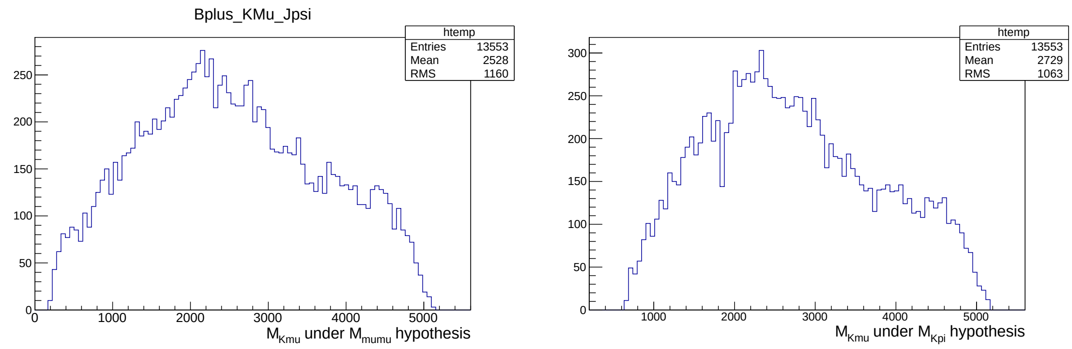


Figure 11: $m_{K\mu}$ under the $\mu^+\mu^-$ (left) and $K^+\pi^-$ (right) mass hypotheses after applying the vetoes described in the text. No structure is seen at the J/ψ, $\psi(2S)$ or D^0 masses.

5.1.2 $B^+ \to (\overline{D}^0 \to K^+\,\pi^-)\pi^+$

The decay $B^+ \to (\overline{D}^0 \to K^+\pi^-)\pi^+$ has a branching fraction of 2×10^{-4} and can fake the signal if the two pions decay in flight. This is removed by computing the $\mu^- K^+$ mass under the $\pi^- K^+$ hypothesis. Candidates which have $1850 < m_{\mu^-(\to\pi^-)K^+} < 1880\,\mathrm{MeV}/c^2$ are removed. The $\mu^- K^+$ mass under the $\pi^- K^+$ hypothesis is shown in Fig. 11. Again, this veto is over 99% efficient on signal.

5.1.3 $B^+ \to K^+ \pi^- \pi^+$

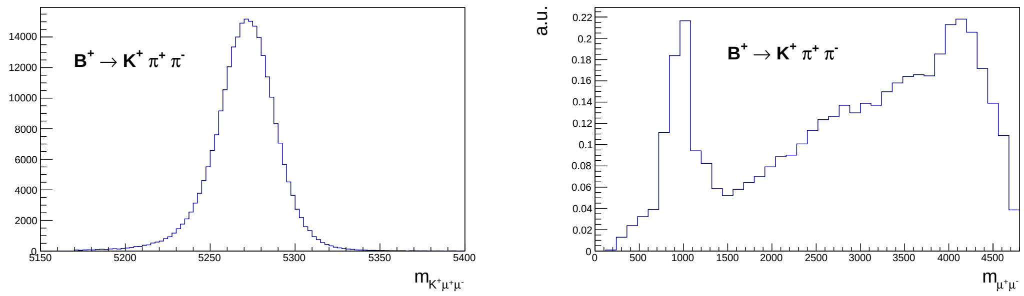


Figure 12: The K^+ μ^+ μ^- and μ^+ μ^- masses for $B^+ \to K^+ \pi^+ \pi^-$ under the $B^+ \to K^+\mu^+\mu^-$ hypothesis. The decay peaks in both mass spectra. The figure has been produced using a sample of simulated phase-space decays, without applying muon identification requirements. This sample has subsequently been re-weighted to have the correct dalitz plot structure.

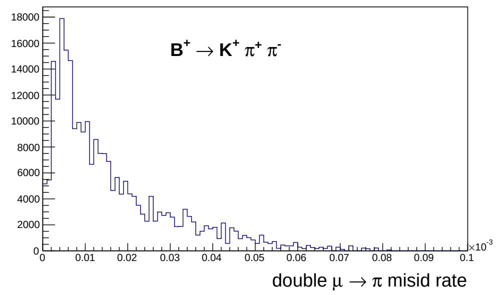


Figure 13: Fraction of $B^+ \to K^+ \pi^+ \pi^-$ candidates which are double mis-identified as $B^+ \to K^+\mu^+\mu^-$. The average double mis-ID fraction is about 1×10^{-5}.

The decay $B^+ \to K^+ \pi^- \pi^+$ has a branching fraction of 5×10^{-5} (100 times $B^+ \to K^+\mu^+\mu^-$). If both pions decay in flight it will fake signal and peak in the B mass (See Fig. 12). The $\pi \to \mu$ mis-ID rate is measured using the muon unbiased $D^0 \to K^+\pi^-$ lines from the PIDCalib package, where the π^- has no PID applied. The mis-ID rate is binned in momentum and eta, and then applied to $B^+ \to K^+ \pi^- \pi^+$ MC11 where the kinematic and geometric selection has been applied. The suppression factor both misID rates is shown in Fig. 13, with an average suppression of 1.5×10^{-5}. This mis-ID rate reduces the background to an effective branching fraction of 5×10^{-10} and hence negligible.

5.2 $B^0 \to K^0_S \mu^+ \mu^-$

5.2.1 Λ^0 reflections

One fortunate aspect of the K_s is because of its long lifetime (cτ = 2.7cm) it is difficult to misidentify its daughter pions. However the decay $\Lambda_b \to (\Lambda_0 \to p^+\pi^-)\mu^+\mu^-$ will look like $B^0 \to K^0_S \mu^+ \mu^-$ if the proton is misidentified as a pion as the Λ_0 also has a long lifetime (cτ = 7.89 cm). In order to deal with this decay and combinatorial background with a random Λ^0, a veto is applied, where the K_s mass is recomputed when one of the pions is given the proton mass. Candidates with this mass within 10(15) MeV/c^2 of the Λ_0 mass (1115 MeV/c^2) for the LL(DD) category are removed. The efficiency of the Λ^0 veto on LL(DD) signal is 96(92)%.

5.2.2 Backgrounds with real K^0_S

The decay $B^0 \to (D^- \to K^0_S\pi^-)\pi^+$, like $B^+ \to (\overline{D}^0 \to K^+\pi^-)\pi^+$ in the $B^+ \to K^+\mu^+\mu^-$ case, will peak in mass and $\cos\theta_l$. Fortunately the branching fraction for this decay is lower than in the corresponding $B^+ \to (\overline{D}^0 \to K^+\pi^-)\pi^+$ background in the $B^+ \to K^+\mu^+\mu^-$ case, at 4×10^{-5}. This, branching fraction, multiplied by a conservative double mis-id of 1×10^{-4}, results in less one event expected and hence negligible.

5.3 $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$

Similarly to $B^+ \to K^+\mu^+\mu^-$, there is background originating from $B^+ \to J/\psi\,(K^{*+} \to K^0_S\pi^+)$, where the pion and same-sign muon are swapped. This background is almost completely rejected by computing the $\mu^-\pi^+$ mass under the $\mu^-\mu^+$ hypothesis. If this mass is within 60 MeV/c^2 of the J/ψ or $\psi(2S)$ resonances then the kaon is required to be in the muon acceptance, but fail the isMuon flag. This is over 99% efficient on signal. There is no visible structure in the $\mu^-\pi^+$ mass after this veto is applied (see Fig. 14), any remaining background is assumed to be negligible.

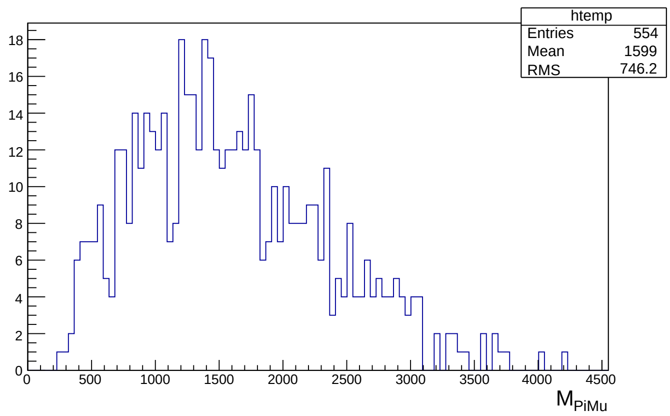


Figure 14: $m_{\pi\mu}$ under the $\mu^+\mu^-$ mass hypotheses after applying the vetoes described in the text. No structure is seen at the J/ψ or $\psi(2S)$ masses.

5.4 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$

Out of the four signal decays, $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ suffers from the largest array of peaking backgrounds. To reject these, specific vetoes based on altered mass hypotheses are employed. The peaking contribution from the decay $\Lambda_b^0 \to pK^+\mu^+\mu^-$ (inclusive of all resonances) is removed by requirements on pK invariant mass and the PID of the π candidate. Candidates are removed if:

$$(5575 < m_{(\pi\to p)K\mu\mu} < 5665)\,\mathrm{MeV}/c^2 \tag{3}$$
$$\pi\mathrm{DLL}_{p\pi} > 0 \tag{4}$$

or

$$(5575 < m_{(K\to p)(\pi\to K)\mu\mu} < 5665)\,\mathrm{MeV}/c^2 \tag{5}$$
$$\pi\mathrm{DLL}_{K\pi} > 0 \tag{6}$$

where $m_{(\pi\to p)K\mu\mu}$ is the Λ_b^0 candidate mass under the hypothesis of mis-identifying the π for the p, and $m_{(K\to p)(\pi\to K)\mu\mu}$ is the Λ_b^0 candidate mass under the hypothesis of mis-identifying the K for the p and the π for the K.

All other vetoes are described in detail in Ref. [9]. Table 11 shows the estimated yields of the peaking backgrounds. The calculations are performed before and after the vetoes and the PID requirements are required. All peaking background contributions are negligable after vetoes.

	before vetoes	
Channel	expected events	perc of signal
$\Lambda_b^0 \to \Lambda^*(1520)\mu^+\mu^-$	$(1.0 \pm 0.5) \times 10^3$	17 ± 7
$B^+ \to K^+\mu^+\mu^- + \pi$	34 ± 9	0.54 ± 0.06
$B_s^0 \to \phi\mu^+\mu^-$	$(3.1 \pm 1.3) \times 10^2$	5.0 ± 1.7
$B^0 \to K^{*0}\mu^+\mu^-$ hadron swaps	$(4.3 \pm 1.1) \times 10^2$	6.9 ± 0.6
$B^0 \to J/\psi K^{*0}$ hadron swaps	0.00022 ± 0.00006	$(3.6 \pm 0.5) \times 10^{-6}$
	after vetoes	
Channel	expected events	perc of signal
$\Lambda_b^0 \to \Lambda^*(1520)\mu^+\mu^-$	51 ± 25	0.8 ± 0.4
$B^+ \to K^+\mu^+\mu^- + \pi$	4.7 ± 1.3	0.076 ± 0.011
$B_s^0 \to \phi\mu^+\mu^-$	18 ± 7	0.29 ± 0.10
$B^0 \to K^{*0}\mu^+\mu^-$ hadron swaps	20 ± 5	0.32 ± 0.04
$B^0 \to J/\psi K^{*0}$ hadron swaps	$(8 \pm 5) \times 10^{-6}$	$(1.3 \pm 0.8) \times 10^{-7}$

Table 11: Expected yields and percenage relative to signal yield for several peaking backgrounds. The yields are estimated from MC samples before and after the vetoes are applied.

6 Mass fits

In order to determine the signal yield in each q^2 bin, an unbinned extended maximum likelihood fit is performed to the $K^+\ \mu^+\ \mu^-$ mass in the range 5170-5700 MeV/c^2. For each mass fit, the signal yield and background yield integrated within 2σ of the signal mean are shown. The signal shape is taken from a mass fit to the normalisation modes, and corrected for differences between the normalisation and signal shapes obtained from the simulation. There are no partially reconstructed backgrounds expected in the mass region we consider, as shown in Fig. 15, which is taken from Ref. [17].

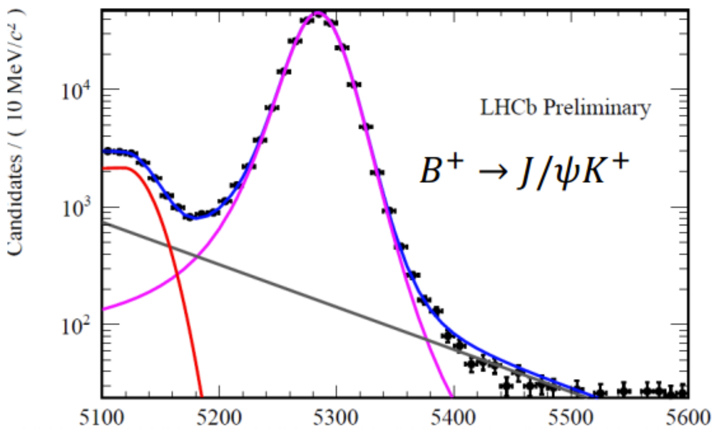


Figure 15: Mass fit to $B^+ \to J/\psi\, K^+$ decays in 2011 data, taken from Ref. [17]. The partially reconstructed background from $B^0 \to J/\psi\, (K^{*0} \to K^+\pi^-)$ does not extend beyond 5170 MeV/c^2.

6.1 $B^+ \to K^+\mu^+\mu^-$

The $B^+ \to K^+\mu^+\mu^-$ signal shape is parameterised as the sum of two Crystal Ball functions, with common tail parameters, but different widths. This shape fits $B^+ \to J/\psi\, K^+$ data OK, as shown for 2011 and 2012 in Fig. 16. The symbols μ, σ, α and n in Fig. 16 are the parameters of the Crystal Ball functions [18]. The pull distribution for $B^+ \to J/\psi\, K^+$ is not perfect due to the huge statistics here, however the effect of mis-modelling is well below the sensitivity of the analysis, and indeed the non-resonant $B^+ \to K^+\mu^+\mu^-$ decays fit well, shown in Fig. 17. The $B^+ \to K^+\mu^+\mu^-$ branching fractions obtained using equation 7,

$$\mathcal{B}(B^+ \to K^+\mu^+\mu^-) = \frac{N_{B^+ \to K^+\mu^+\mu^-}\mathcal{B}(B^+ \to J/\psi\, K^+)\mathcal{B}(J/\psi \to \mu^+\mu^-)}{\epsilon_{rel} N_{B^+ \to J/\psi\, K^+}} \tag{7}$$

The results for the 2011 and 2012 data are $(4.59 \pm 0.14) \times 10^{-7}$ and $(4.44 \pm 0.10) \times 10^{-7}$, respectively, agreeing within 1σ.

A width and mean parameters for $B^+ \to K^+\mu^+\mu^-$ decays as a function of q^2 is performed with the simulation. The absolute difference between these parameters and the

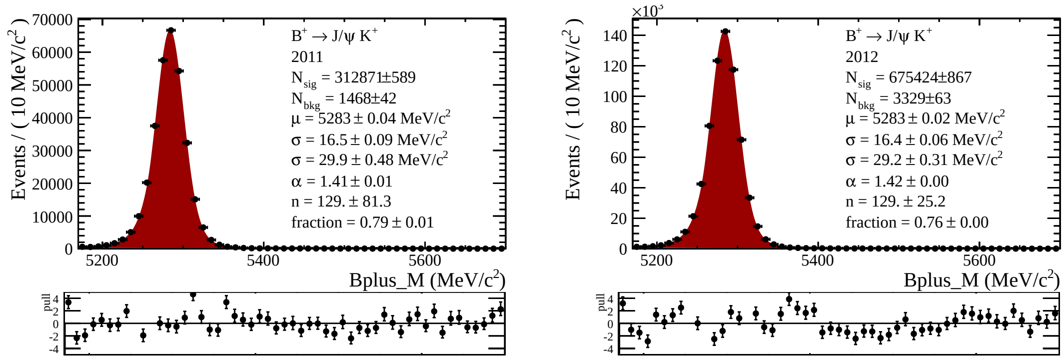


Figure 16: Mass fits to $B^+ \to J/\psi K^+$ decays in data for 2011 (left) and 2012 (right). The signal shape is very similar between the two and so the data are combined for the final fit. The fit parametrisation is a double Crystal Ball for signal and exponential as background.

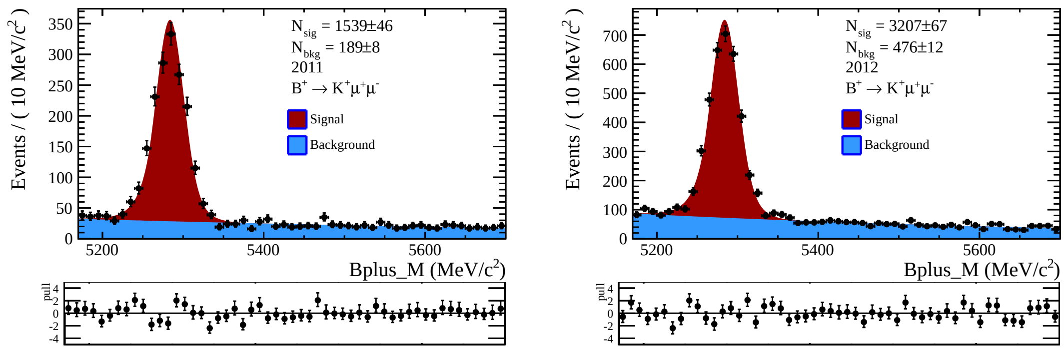


Figure 17: Mass fits to $B^+ \to K^+\mu^+\mu^-$ decays in data for 2011 (left) and 2012 (right). The signal shape is very similar between the two and so the data are combined for the final fit. The fit parametrisation is a double Crystal Ball for signal and exponential as background.

$B^+ \to J/\psi K^+$ bin (wide one) is used to correct the signal shape from $B^+ \to J/\psi K^+$ to $B^+ \to K^+\mu^+\mu^-$ decays. These corrections make very little difference to the signal yields, as shown in Fig. 19.

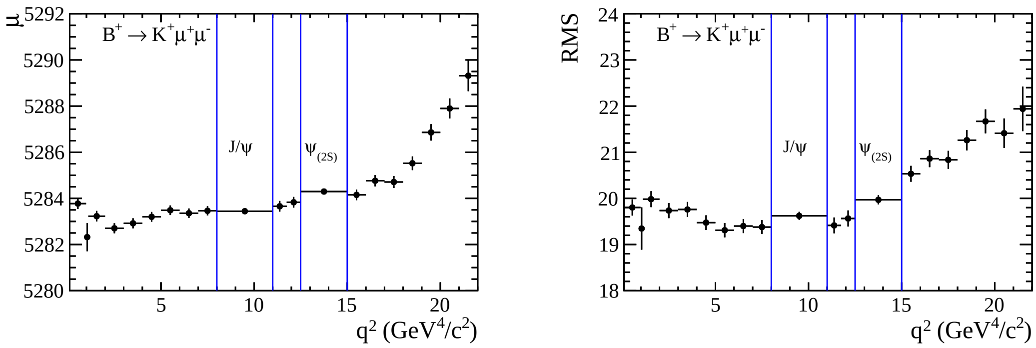


Figure 18: The mean and RMS of $B^+ \to K^+\mu^+\mu^-$ simulation as a function of q^2. The difference between the parameters at the in each q^2 bin and at the J/ψ mass are used as a correction.

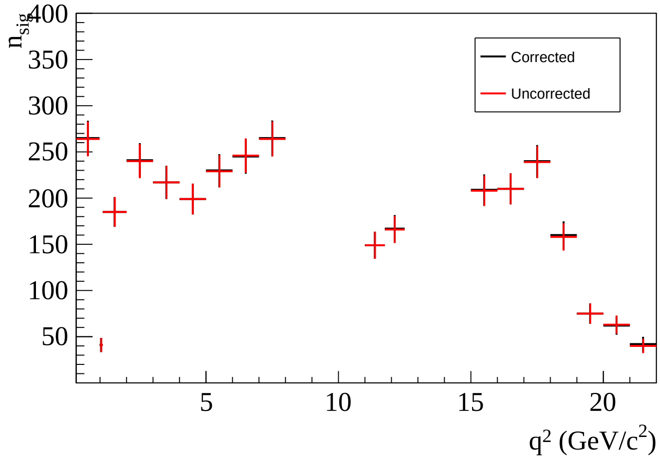


Figure 19: Yields of $B^+ \to K^+\mu^+\mu^-$ in 2012 data with and without the mass model corrections as a function of q^2. There is a negligible change in the yield.

6.2 $B^0 \to K^0_S \mu^+ \mu^-$

The fit procedure for $B^+ \to K^+ \mu^+ \mu^-$ is applied to $B^0 \to K^0_S \mu^+ \mu^-$. The only real difference is a component for $B^0_s \to J/\psi K^0_S$, where the signal shape parameters are assumed to be identical to the $B^0 \to J/\psi K^0_S$ with the exception of a mean shift obtained from difference in the B^0 and B^0_s masses. The $B^0 \to J/\psi K^0_S$ mass fits in the LL and DD categories are shown in Fig. 20.

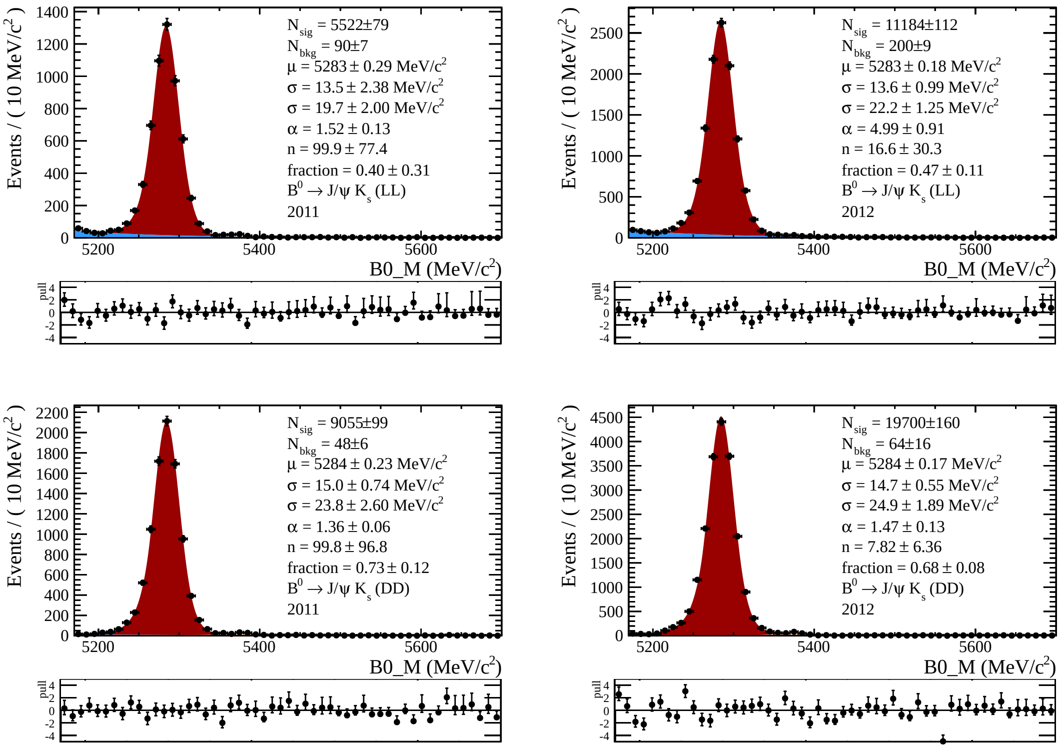


Figure 20: Mass fits to $B^0 \to J/\psi K^0_S$ decays in data for 2011 (left) and 2012 (right) in the LL (top) and DD (bottom) categories. The signal shape is very similar between the two years and so the data are combined for the final fit. The fit parametrisation is a double Crystal Ball for signal and exponential as background.

Fits to the non-resonant modes are shown in Fig. 21. There is approximately 30 times less data than $B^+ \to K^+ \mu^+ \mu^-$ due to four factors; lower visible branching fraction due to $K^0 \to (K^0_S \to \pi^+ \pi^-)$(factor 3), low reconstruction efficiency (factor 3), tighter BDT (factor 2) and lower branching fraction (factor 1.5). For the 2011 data, the signal yields are consistent with previous analysis, whereas the background is significantly reduced. To take the DD category as an example, the B/S, where B is the background extrapolated under the signal region and S is the signal, was about 1 last time, whereas now it is about 0.3.

In order to check the consistency between the categories and run periods, the branching fraction is calculated with the following equation,

$$\mathcal{B}(B^0 \to K^0\mu^+\mu^-) = \frac{N_{B^0\to K^0_S\mu^+\mu^-}\mathcal{B}(B^0 \to J/\psi\, K^0)\mathcal{B}(J/\psi \to \mu^+\mu^-)}{\epsilon_{rel} N_{B^0\to J/\psi\, K^0_S}} \quad (8)$$

where N_{signal} is the signal yield, $\mathcal{B}(B^0 \to J/\psi\, K^0)\mathcal{B}(J/\psi \to \mu^+\mu^-)$ is the branching fraction of the normalisation channel, $N_{B^0\to J/\psi\, K^0_S}$ is the number of $B^0 \to J/\psi\, K^0_S$ candidates seen in the data and the ϵ_{rel} is the relative efficiency between $B^0 \to K^0_S\mu^+\mu^-$ and $B^0 \to J/\psi\, K^0_S$ obtained using the simulation and assuming a SM distribution over q^2. This assumption is clearly broken if there is large isospin asymmetry localised in particular regions of q^2, but it is best we can do while the results are still blinded. The branching fraction results are summarised in Tab. 12 (just stat uncertainty). There is a 1.4σ fluctuation upwards of the branching fraction in 2012 compared to 2011 for the LL category, whereas the DD category fluctuates up by 0.9σ. The probability to get both these fluctuations is just under 10%. It is difficult to see any effect that could cause this fluctuation other than statistics. Both 2011 and 2012 datasets use the same reconstruction version. The only real difference being the trigger, which is dominated by the muons. Given the fact that the $B^+ \to K^+\mu^+\mu^-$ branching fractions agree very nicely between 2011 and 2012, it is unlikely that the trigger could cause this effect in $B^0 \to K^0_S\mu^+\mu^-$ and not $B^+ \to K^+\mu^+\mu^-$.

Mode	Branching fraction
LL 2011	$(2.2 \pm 0.8) \times 10^{-7}$
LL 2012	$(3.7 \pm 0.7) \times 10^{-7}$
DD 2011	$(3.3 \pm 0.7) \times 10^{-7}$
DD 2012	$(4.1 \pm 0.6) \times 10^{-7}$

Table 12: Branching fraction results obtained for the different $B^0 \to K^0_S\mu^+\mu^-$ modes. There is a combined 1.67σ fluctuation upwards for the 2012 data compared to the 2011 data.

Similarly to $B^+ \to K^+\mu^+\mu^-$, the simulation is used to calculate the variation of the mean and width parameters with q^2. This study is shown in Fig. 22. Similarly to $B^+ \to K^+\mu^+\mu^-$, the corrections make a very small difference, as shown in Fig. 23.

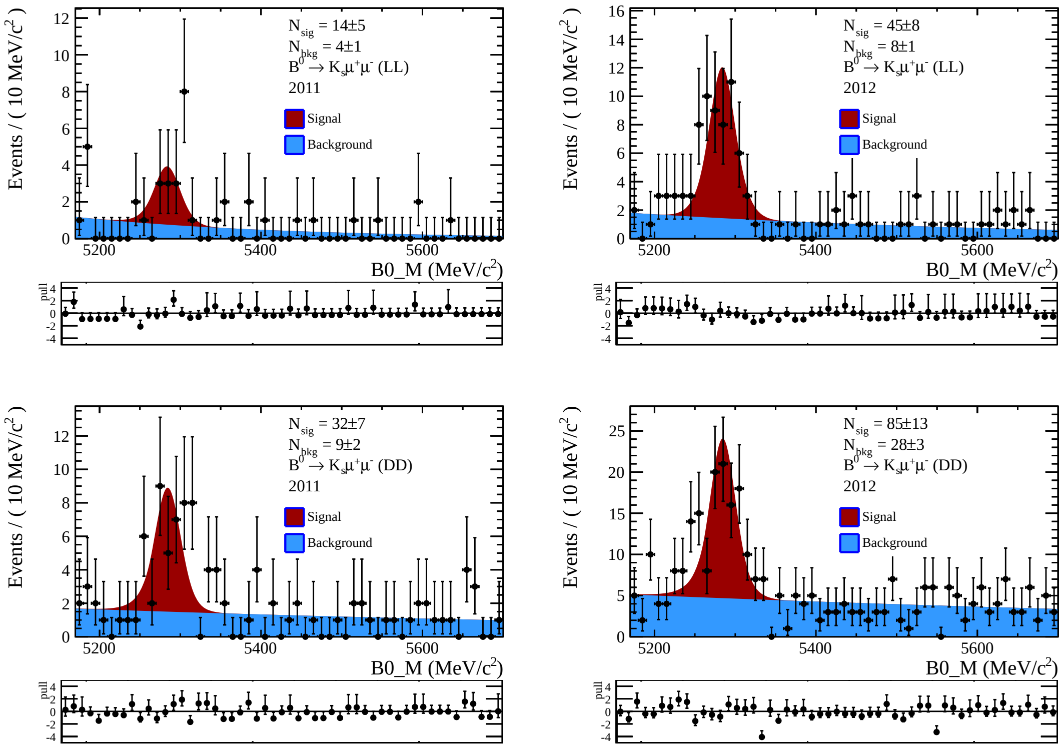


Figure 21: Mass fits to $B^0 \to K^0_S\mu^+\mu^-$ decays in data for 2011 (left) and 2012 (right) in the LL (top) and DD (bottom) categories. The signal shape is very similar between the two years and so the data are combined for the final fit. The fit parametrisation is a double Crystal Ball for signal and exponential as background.

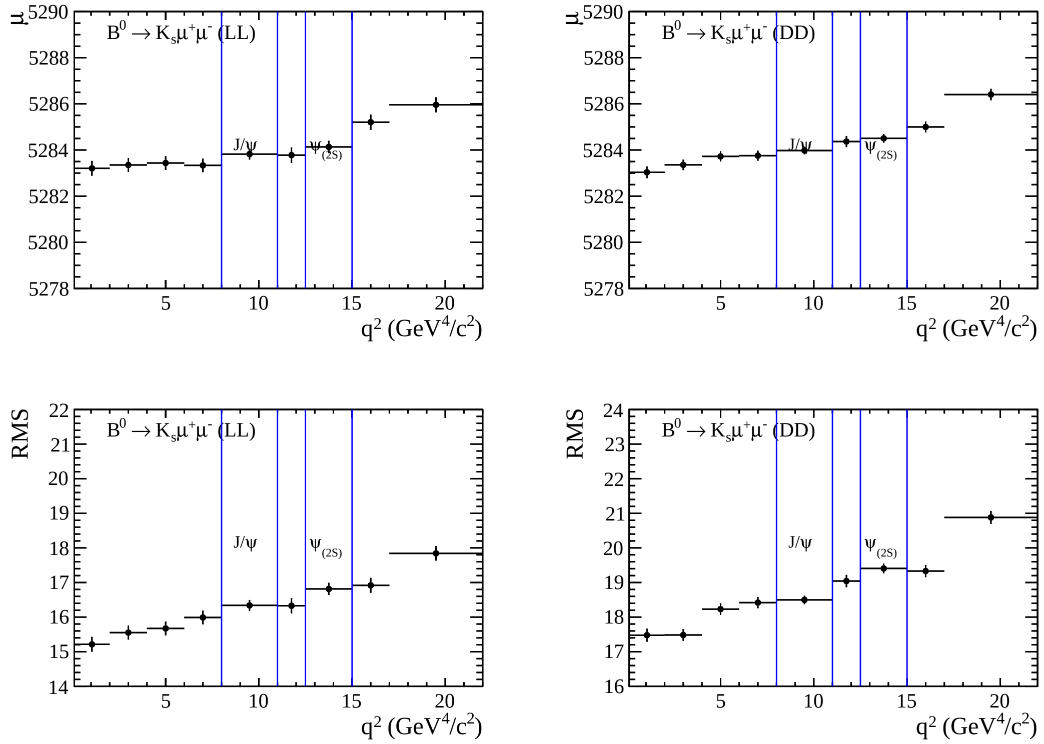


Figure 22: The mean and RMS of $B^0 \to K^0_S\mu^+\mu^-$ simulation as a function of q^2, split into the LL (left) and DD (right) categories. The absolute difference between parameters for a particular q^2 bin and the J/ψ bin are used to correct the signal shape.

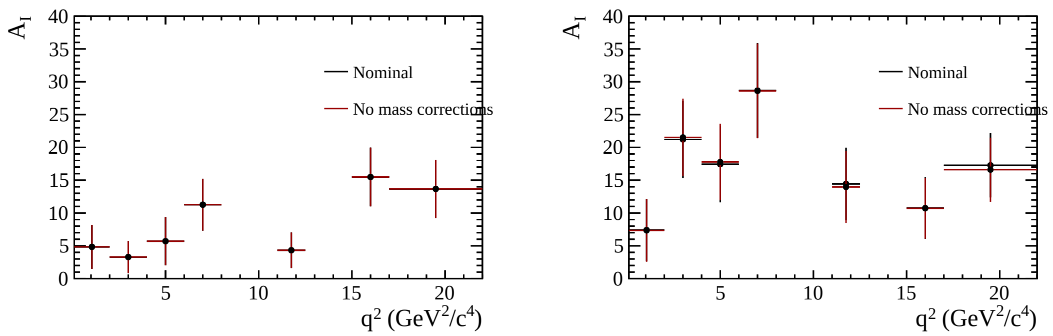


Figure 23: Signal yield of $B^0 \to K^0_S\mu^+\mu^-$ in bins of q^2 with and without the mass corrections applied.

6.3 $B^+ \to (K^{*+} \to K^0_S \pi^+)\mu^+\mu^-$

Again, using the same fitting strategy as $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^0_S\mu^+\mu^-$, the $B^+ \to J/\psi\,(K^{*+} \to K^0_S\pi^+)$ and $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ signal yields are determined, shown in Figs. 24 and 25. The signal is very clean at the end of the selection chain. In the same way to the $B^0 \to K^0_S\mu^+\mu^-$ results, the branching fractions are calculated, assuming the SM q^2 distribution, to assess the consistency between the run periods and reconstruction categories. This is shown in Tab. 13, where all branching fractions agree with each other.

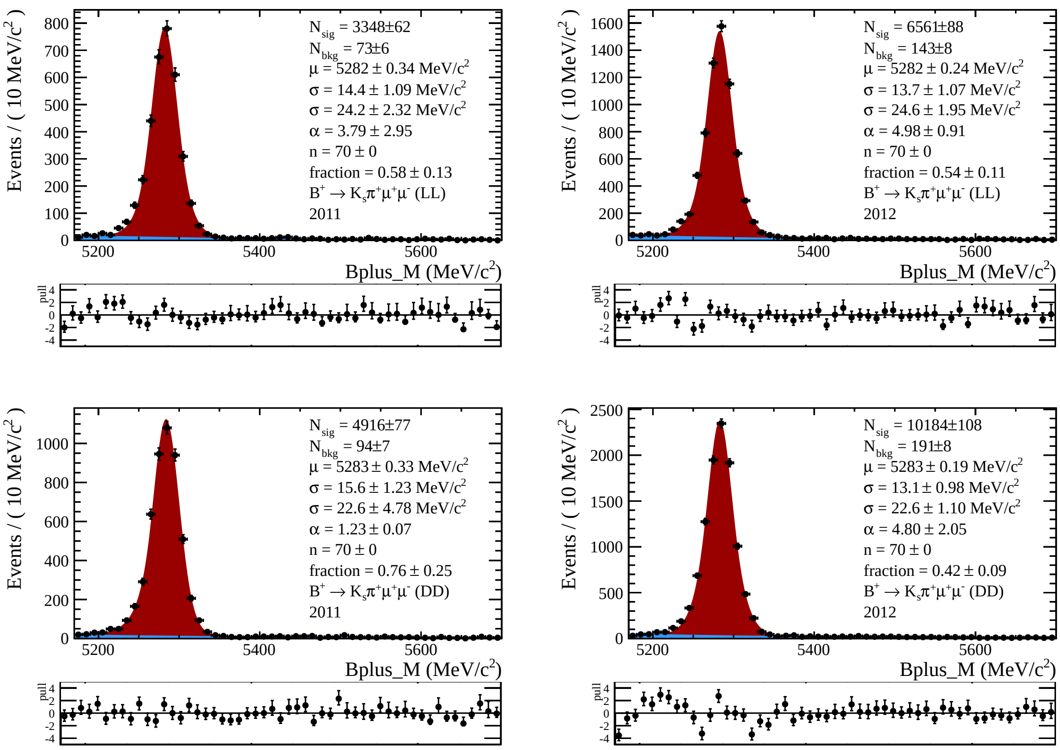


Figure 24: Mass fits to $B^+ \to J/\psi\,(K^{*+} \to K^0_S\pi^+)$ decays in data for 2011 (left) and 2012 (right) in the LL (top) and DD (bottom) categories. The signal shape is very similar between the two years and so the data are combined for the final fit. The fit parametrisation is a double Crystal Ball for signal and exponential as background.

Similarly to $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^0_S\mu^+\mu^-$, the simulation is used to calculate the variation of the mean and width parameters with q^2. This study is shown in Fig. 26. The $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ data in bins of q^2 is still blind and so the difference in yields cannot be compared yet. Similarly to $B^+ \to K^+\mu^+\mu^-$ however, it is expected that the corrections make a very small difference.

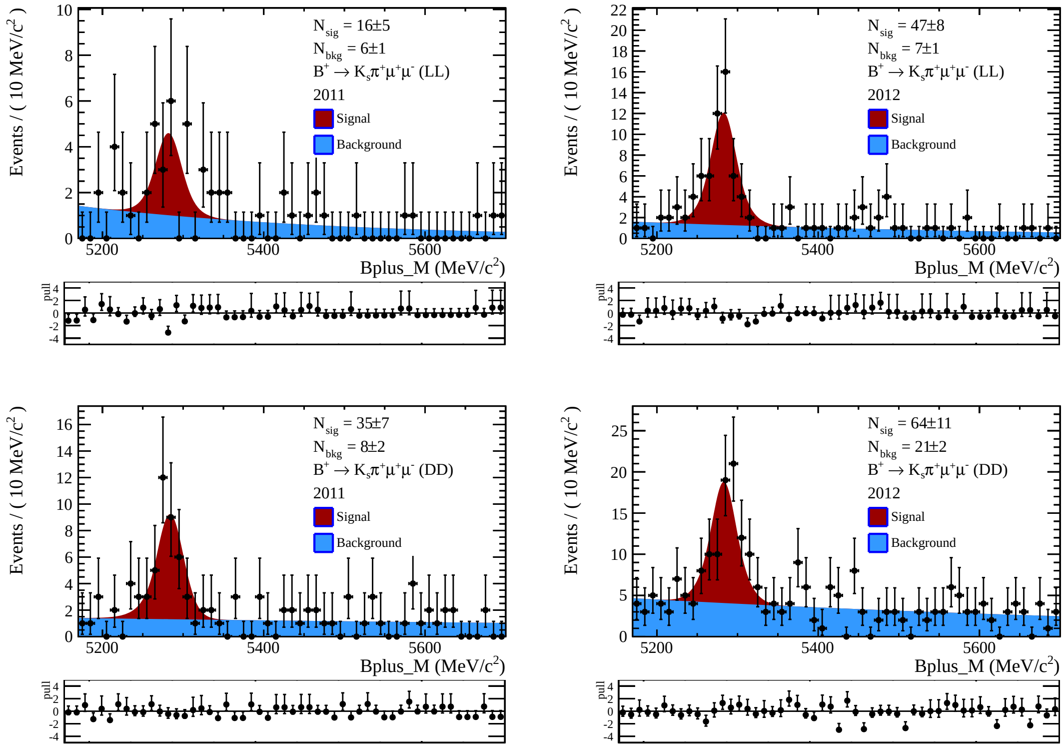


Figure 25: Mass fits to $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ decays in data for 2011 (left) and 2012 (right) in the LL (top) and DD (bottom) categories. The signal shape is very similar between the two years and so the data are combined for the final fit. The fit parametrisation is a double Crystal Ball for signal and exponential as background.

Mode	Branching fraction
LL 2011	$(7.2 \pm 2.6) \times 10^{-7}$
LL 2012	$(10.2 \pm 1.7) \times 10^{-7}$
DD 2011	$(11.2 \pm 2.2) \times 10^{-7}$
DD 2012	$(9.5 \pm 1.6) \times 10^{-7}$

Table 13: Branching fraction results obtained for the different $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ modes. All modes are compatible with each other.

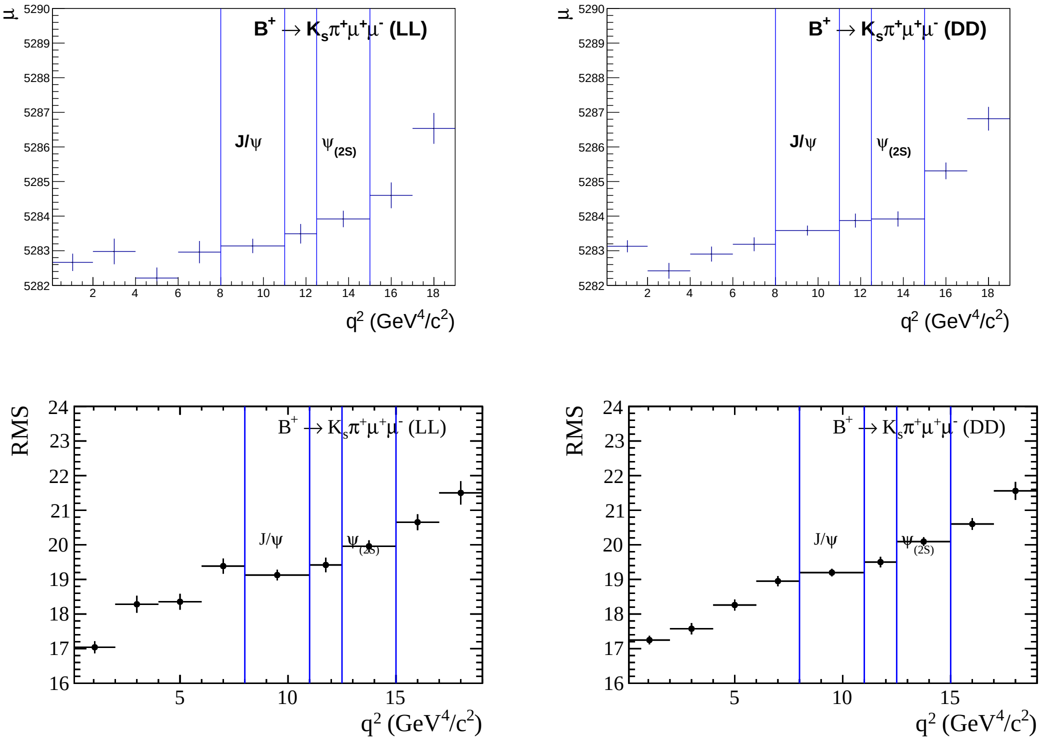


Figure 26: The mean and RMS of $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ simulation as a function of q^2, split into the LL (left) and DD (right) categories. The absolute difference between parameters for a particular q^2 bin and the J/ψ bin are used to correct the signal shape.

6.4 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$

Finally, using the same fitting strategy as above, the $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ and $B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$ signal yields are determined, shown in Figs. 27 and 28.

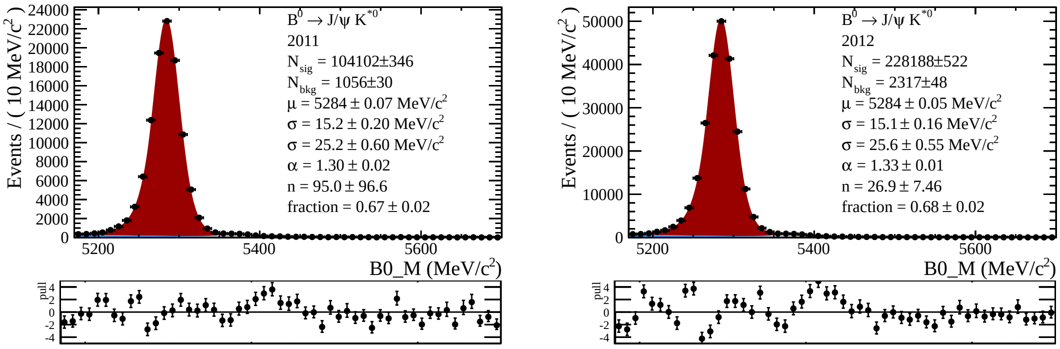


Figure 27: Mass fits to $B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$ decays in data for 2011 (left) and 2012 (right). The signal shape is very similar between the two years and so the data are combined for the final fit. The fit parametrisation is a double Crystal Ball for signal and exponential as background.

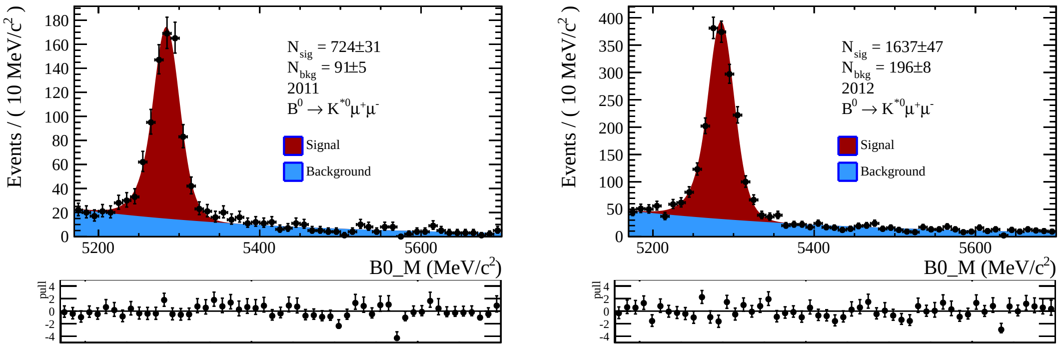


Figure 28: Mass fits to $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ decays in data for 2011 (left) and 2012 (right). The signal shape is very similar between the two years and so the data are combined for the final fit. The fit parametrisation is a double Crystal Ball for signal and exponential as background.

Similarly to the other channels, the simulation is used to calculate the variation of the mean and width parameters with q^2. This study is shown in Fig. 29.

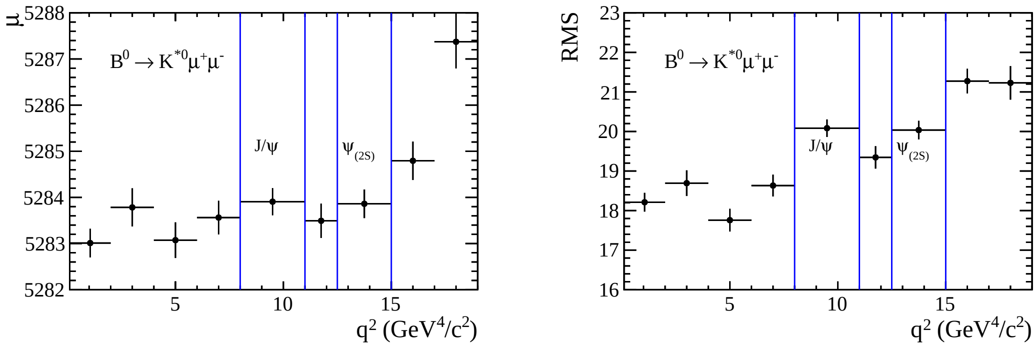


Figure 29: The mean and RMS of $B^0 \rightarrow (K^{*0} \rightarrow K^+\pi^-)\mu^+\mu^-$ simulation as a function of q^2. The absolute difference between parameters for a particular q^2 bin and the J/ψ bin are used to correct the signal shape.

7 Efficiencies

For the branching fraction measurements and isospin asymmetries, each channel is normalised to the relevant control channel in order to cancel systematic uncertainties relating to the efficiency. The control channel used for each signal channel is the resonant $B \to J/\psi\, h$ mode, where $J/\psi \to \mu^+\mu^-$ which has the identical final state and so the only difference in efficiency is due to a difference in the kinematics. In order to normalise each signal channel, the relative selection, reconstruction and trigger efficiency between the signal and control channel must be calculated. This calculation was done in each bin of q^2, ignoring any angular variables. The relative efficiency between the signal and control channels can be split up into pieces as:

$$\frac{\epsilon_{X\mu^+\mu^-}}{\epsilon_{XJ/\psi}} = \frac{\epsilon^{DPC}_{X\mu^+\mu^-}}{\epsilon^{DPC}_{XJ\psi}} \frac{\epsilon^{reco\&sel|DPC}_{X\mu^+\mu^-}}{\epsilon^{reco\&sel|DPC}_{XJ\psi}} \frac{\epsilon^{trigger|reco\&sel|DPC}_{X\mu^+\mu^-}}{\epsilon^{trigger|reco\&sel|DPC}_{XJ\psi}} \frac{\epsilon^{PID|trigger|reco\&sel|DPC}_{X\mu^+\mu^-}}{\epsilon^{PID|trigger|reco\&sel|DPC}_{XJ\psi}} \tag{9}$$

where the relative efficiency has been broken down into the DecProdCut (DPC), reconstruction & selection, PID and trigger relative efficiencies. These pieces that make up the total relative efficiency are discussed in the following sections. Each piece is calculated given the previous selection, for example the reconstruction & selection efficiency is calculated given DPC, the PID efficiency is calculated given the reconstruction & selection and DPC etc. The simulation is truth matched so that the number of simulation is the same as the number of signal candidates. The truth matching is 100% efficient above $0.1\,\mathrm{GeV}^2/c^4$ in q^2. Below this region, the truth matching removes signal due to shared muon hits. This is one of the reasons why the data below $0.1\,\mathrm{GeV}^2/c^4$ is not used in the analysis.

If one truth-matches with the $K^0_S \to \pi^+\, \pi^-$ decay descriptor, this ignores the fraction of K^0_S which undergo a material interaction with the detector. This fraction is dependent on the K^0_S momentum and hence q^2, as shown in Fig. 30. This effect is taken care of by not requiring the K^0_S decays at the truth level, but in any case the effect is small compared to the expected statistical sensitivity.

7.1 Data corrections

The simulation is corrected to match the data. With the exception of the isMuon criteria, the PID efficiencies are obtained from the data using the PIDCalib package. The isMuon criteria is taken from the simulation. The DLL cut applied at the stripping is obtained using the PIDCalib package. The effect of the ProbNN variables is estimated using $B^+ \to J/\psi\, K^+$, where no PID is applied. The sample is split into bins of momentum and pseudorapidity, the ProbNN variables are applied and the efficiency is calculated using a mass fit. The results are shown in Fig. 31, which shows the weights that are applied to the simulation. No dependence on track multiplicity is applied here, as the track multiplicity distribution is assumed to be the same as for the non-resonant modes (a good assumption as they are triggered in the same way). The reason that $B^+ \to J/\psi\, K^+$ is used rather than

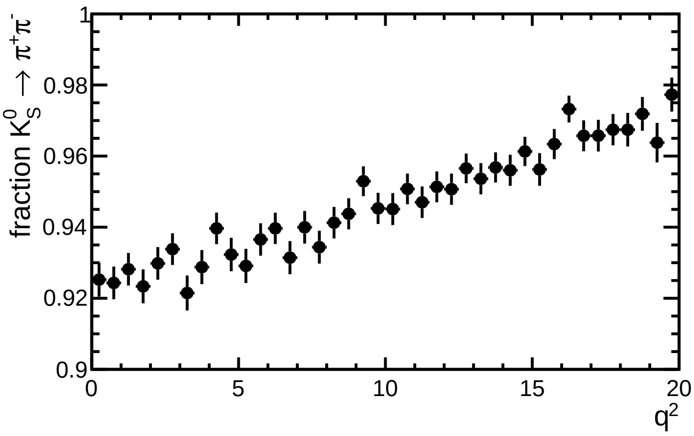


Figure 30: Fraction of K_S^0 mesons which decay before interacting with the detector as a function of q^2.

the PIDCalib package is there is a slight disagreement between the $B^+ \to J/\psi K^+$ results and weights obtained from the PIDCalib, shown in Fig. 32 for the Kaon ProbNN cut. Although the disagreement is localised in regions where there are hardly any candidates anyway, it is safer to simply use $B^+ \to J/\psi K^+$ to avoid any systematic uncertainty. For $B^0 \to K_S^0 \mu^+\mu^-$, no ProbNN variables are applied and so no correction is needed.

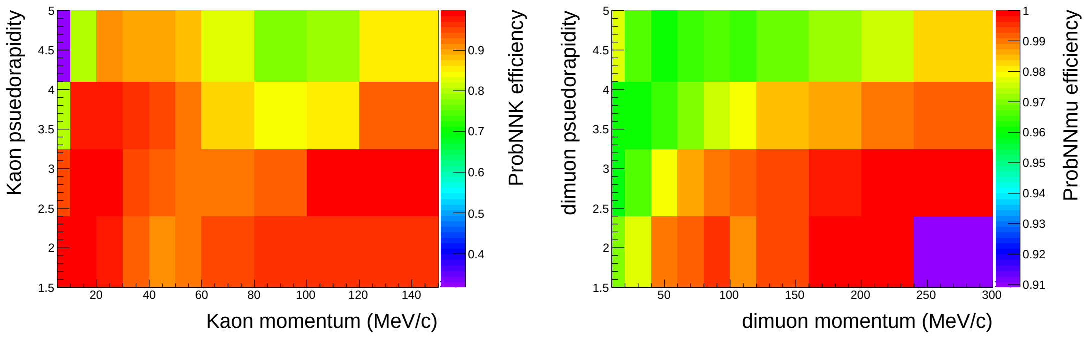


Figure 31: Efficiency of the ProbNN selection on $B^+ \to J/\psi K^+$ candidates as a function of momentum and pseudorapidtiy.

To correct for difference in IP resolution in MC11, all tracks in the best track container are smeared by approximately 20%. After which the IP distributions match nicely (see Sect. A in the appendix). Finally, the following variables are reweighed: number of tracks, B p_T and B vertex χ^2. For $B^+ \to K^+\mu^+\mu^-$, the p_T distributions of the daughters are also reweighed as there is a discrepancy for soft tracks due to the loose selection (see the mis-modelling systematic in Sect. 9 for more details). This correction procedure is done

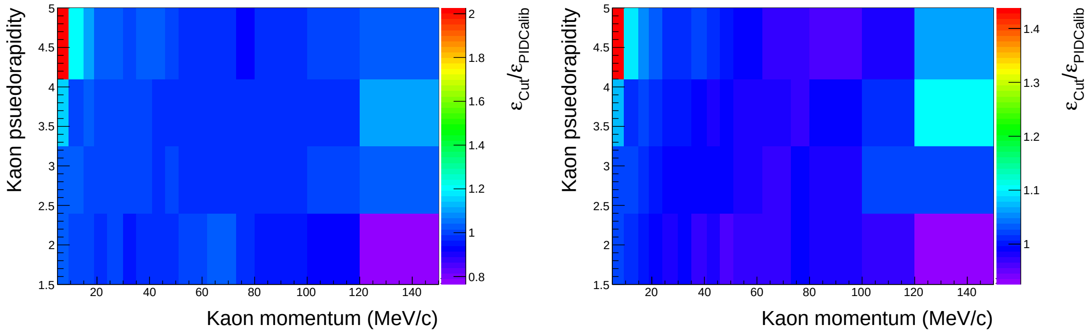


Figure 32: Ratio of ProbNNK efficiencies obtained by applying selection on $B^+ \to J/\psi K^+$ data and using the PIDCalib package. For most of the kinematic region, the methods are consistent with each other, however at the extremes there is discrepancy.

separately for 2011 and 2012 data. These corrections to the efficiency as a function of q^2, are shown in Fig: 33. For $B^+ \to K^+\mu^+\mu^-$ there is a slight decrease in efficiency at high q^2 due to the K^+ p_{T} reweighing. For $B^0 \to K^0_{\mathrm{S}}\mu^+\mu^-$ there is no significant trend, as the data/MC agreement is better out of the box in this case (due to the tighter BDT selection).

The agreement between the data and simulation can be found in the appendix. In summary, there are no variables badly modelled which are correlated to q^2. Due to the large number of variables reweighed for $B^+ \to K^+\mu^+\mu^-$, this is investigated further as a systematic uncertainty in Sect. 9.

7.2 DecProdCut

There will be a difference in the fraction of particles which enter the LHCb acceptance due to the different kinematics of the signal and control channels. These fractions are calculated by generating decays with no generator level cuts. The fraction of these decays which survive the “DecProdCut“ selection is then calculated in bins of q^2. The results are shown in Fig. 34 and 35. There is a difference between the K^0_{S} and K^+ channels as the “DecProdCut“ is not applied to the K^0_{S} candidate.

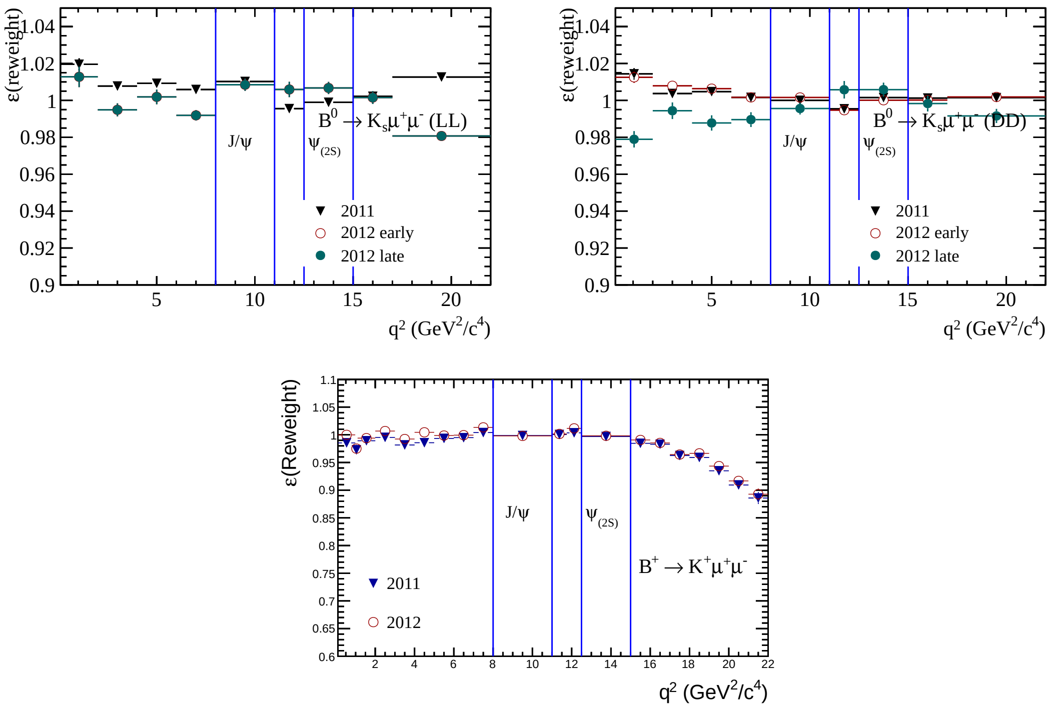


Figure 33: Effect of the reweighing on the relative efficiency as a function of q^2. For $B^0 \to K^0_S\mu^+\mu^-$ decays, there is no significant effect across q^2 whereas for $B^+ \to K^+\mu^+\mu^-$, there is a slight drop in efficiency at high q^2 due to the reweighing of the K^+ p_T.

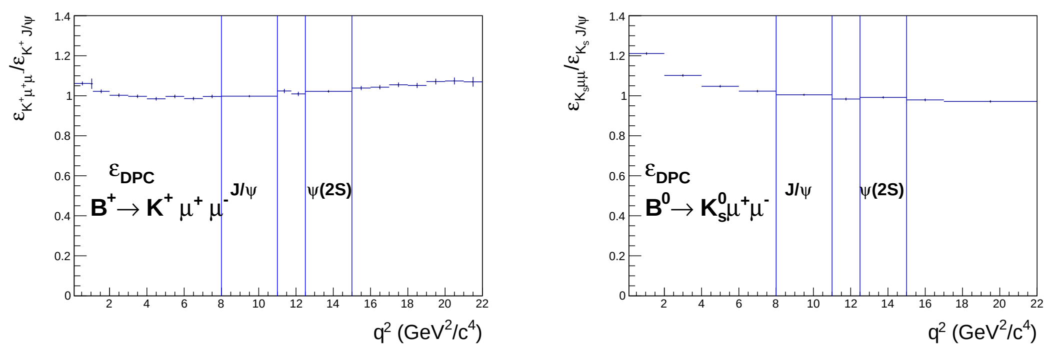


Figure 34: The relative “DecProdCut“ efficiency between the signal and normalisation channel for $B^+ \to K^+\mu^+\mu^-$ (left) and $B^0 \to K^0_S\mu^+\mu^-$ (right) decays as a function of q^2.

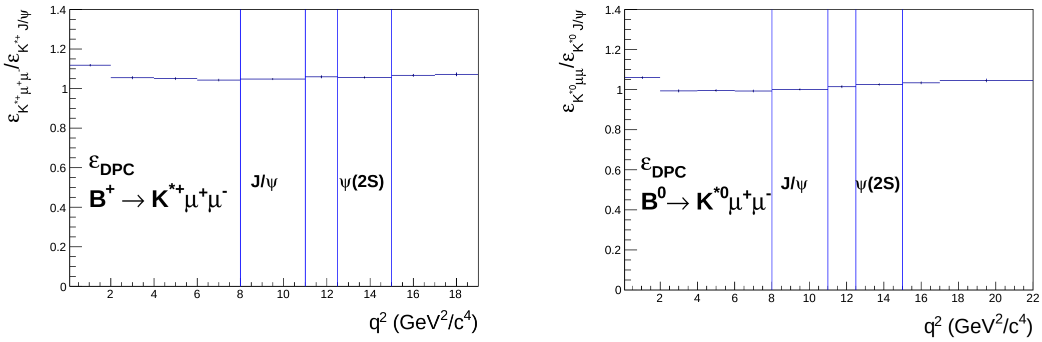


Figure 35: The relative "DecProdCut" efficiency between the signal and normalisation channel for $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ (left) and $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ decays (right) as a function of q^2.

7.3 Reconstruction & Stripping

The reconstruction and stripping efficiency is evaluated by applying the stripping selection to fully simulated signal and control events. The efficiency is defined as the number of candidates that survive this divided by the number of decays generated which survive the"DecProdCut" selection in each bin of q^2.

7.3.1 $B^+ \to K^+\mu^+\mu^-$

The reconstruction and stripping efficiency of $B^+ \to K^+\mu^+\mu^-$ relative to $B^+ \to J/\psi K^+$ is shown in Fig. 36. At high mass the K^+ starts to become collinear with the B direction and therefore IP χ^2 criteria in the stripping starts to reduce the signal efficiency. At low q^2 there is a small decrease in efficiency as the muons become soft and become less likely to reach the muon stations.

7.3.2 $B^0 \to K^0_S\mu^+\mu^-$

The reconstruction and selection efficiency of $B^0 \to K^0_S\mu^+\mu^-$ relative to $B^0 \to J/\psi K^0_S$ is shown in Fig. 37. The reconstruction efficiency of $B^0 \to K^0_S\mu^+\mu^-$ is driven by the kinematics of the K^0_S; low momentum K^0_S are more likely to decay inside the VELO and so at high q^2 there are more LL candidates reconstructed than DD candidates. Also remember that due to the absence of a "DecProdCut" on the generator level for the K^0_S means that there is a lower efficiency at low q^2 due to the implicit requirement that the K^0_S must be in acceptance. Note how similar the MC11 and MC2012 efficiency curves are, which validates the assumption that the reconstruction version is irrelevant after normalising to the J/ψ modes.

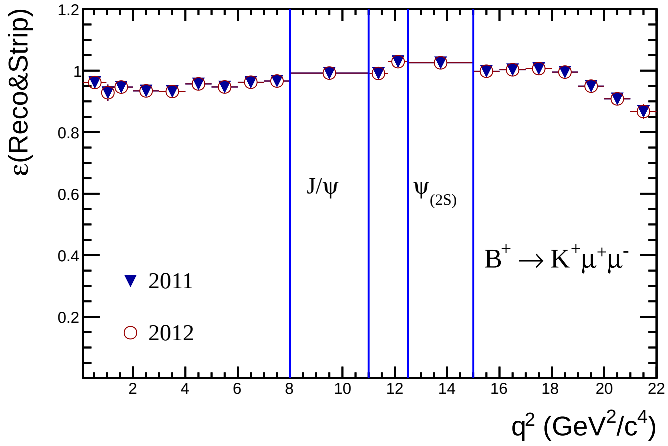


Figure 36: The relative reconstruction and stripping efficiency between $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to J/\psi\, K^+$ as a function of q^2.

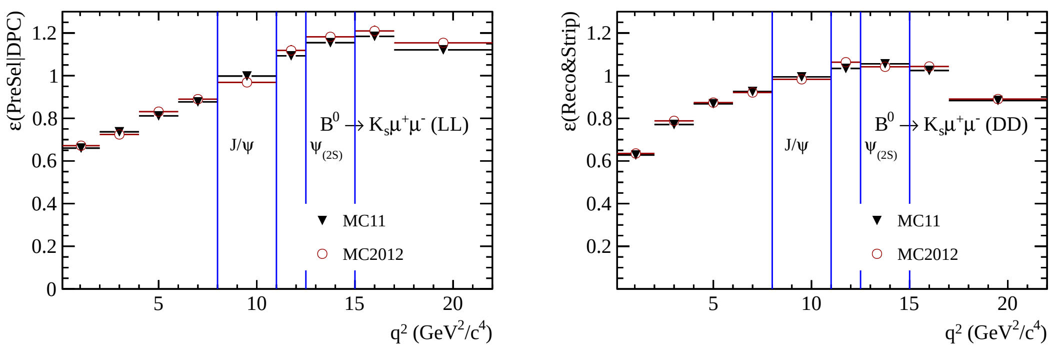


Figure 37: The relative reconstruction and stripping efficiency between $B^0 \to K^0_S\mu^+\mu^-$ and $B^0 \to J/\psi\, K^0_S$ for the LL (left) and DD (right) categories as a function of q^2.

7.3.3 $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$

The reconstruction and selection efficiency of $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ relative to $B^+ \to J/\psi\,(K^{*+} \to K^0_S\pi^+)$ is shown in Fig. 38. Unlike $B^0 \to K^0_S\mu^+\mu^-$ the efficiency curves between the LL and DD categories are quite similar. This is because the K^0_S kinematics are less correlated to q^2 due to the heavy K^* mass which is obviously constant across q^2. Note how similar the MC11 and MC2012 efficiency curves are, which validates the assumption that the reconstruction version is irrelevant after normalising to the J/ψ modes. If there is any difference it is more likely to be due to the difference in centre-of-mass energies.

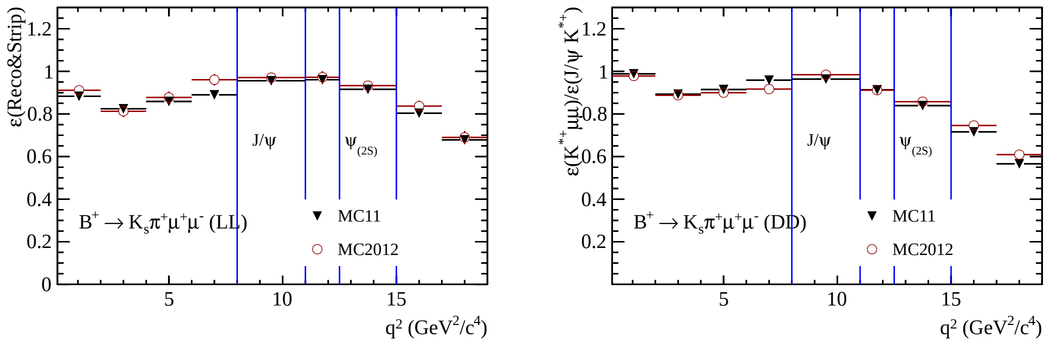


Figure 38: The relative reconstruction and stripping efficiency between $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ and $B^+ \to J/\psi\,(K^{*+} \to K^0_S\pi^+)$ for the LL (left) and DD (right) categories as a function of q^2.

7.3.4 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$

The reconstruction and selection efficiency of $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ relative to $B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$ is shown in Fig. 39. The reconstruction efficiency is similar to $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ as the K^0_S and K^+ kinematics are not so correlated q^2.

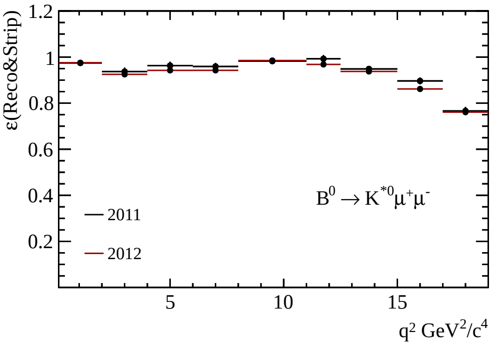


Figure 39: The relative reconstruction and stripping efficiency between $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ and $B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$ as a function of q^2.

7.4 Trigger efficiency

The trigger requirements are discussed in section 4. In general, the efficiency goes up with q^2 which is expected as the trigger decisions are dominated by the kinematics of the muons.

7.4.1 $B^+ \to K^+\mu^+\mu^-$

The relative trigger efficiency for $B^+ \to K^+\mu^+\mu^-$, is shown in Fig. 40. The efficiency split into the different levels can be found in Sect. C in the Appendix.

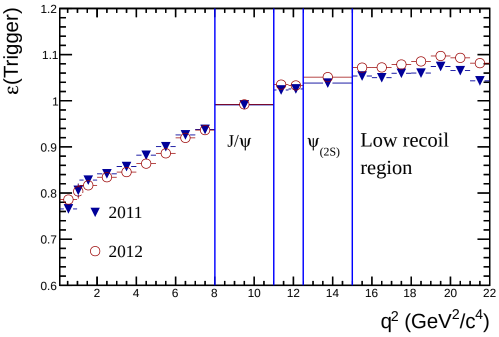


Figure 40: The relative trigger efficiency for the and selection efficiency between $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to J/\psi K^+$ as a function of q^2. The efficiency increases with q^2 as the trigger is dominated by the muons, which get harder in this region.

7.4.2 $B^0 \to K^0_S\mu^+\mu^-$

The relative trigger efficiency for $B^0 \to K^0_S\mu^+\mu^-$ is shown in Fig. 41. There is some different behaviour seen between the running periods which is due to HLT2. The efficiency split into the different levels can be found in Sect. C in the Appendix.

7.4.3 $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$

The relative trigger efficiency for $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ is shown in Fig. 42. The efficiency split into the different levels can be found in Sect. C in the Appendix. Contrary to $B^0 \to K^0_S\mu^+\mu^-$, the trigger efficiency is very similar between 2011 and 2012 as the K^0_S contributes less for this decay. There is a small difference at very low q^2 due to the HLT2DimuonDetached line, the mass threshold of which was lowered from $1.5\,\mathrm{GeV}/c^2$ to $1\,\mathrm{GeV}/c^2$.

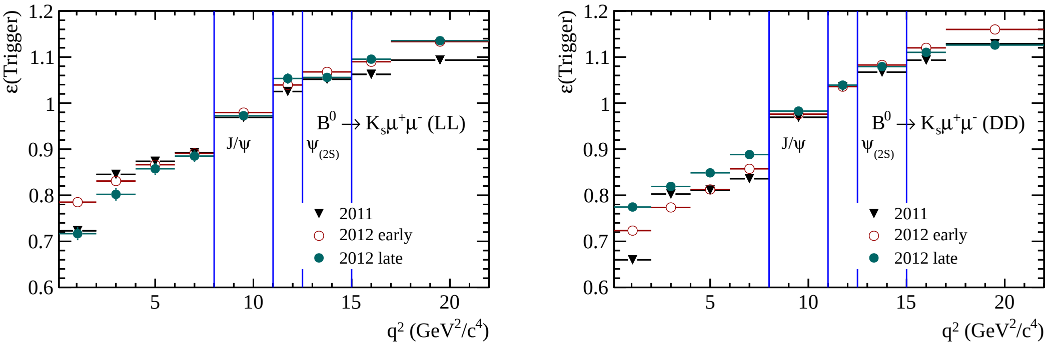


Figure 41: The relative trigger efficiency for the and selection efficiency between $B^0 \to K^0_{\rm S}\mu^+\mu^-$ and $B^0 \to J/\psi\, K^0_{\rm S}$ for the LL (left) and DD (right) categories as a function of q^2. The efficiency increases with q^2 as the trigger is dominated by the muons, which get harder in this region.

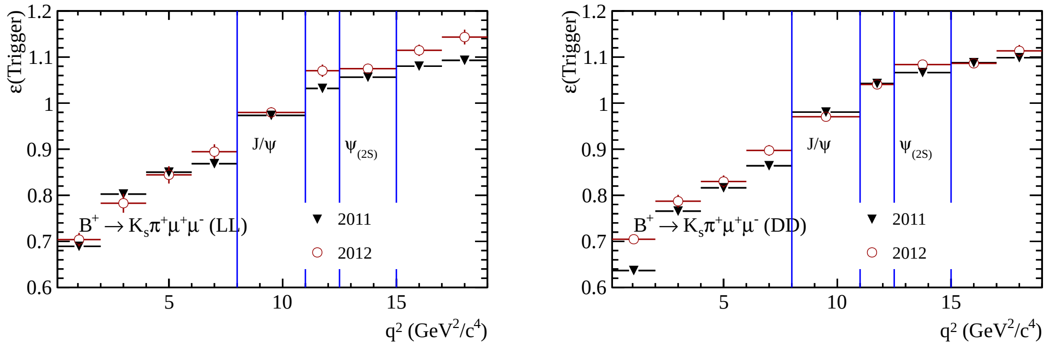


Figure 42: The relative trigger efficiency for the and selection efficiency between $B^+ \to (K^{*+} \to K^0_{\rm S}\pi^+)\mu^+\mu^-$ and $B^+ \to J/\psi\,(K^{*+} \to K^0_{\rm S}\pi^+)$ for the LL (left) and DD (right) categories as a function of q^2. The efficiency increases with q^2 as the trigger is dominated by the muons, which get harder in this region.

7.4.4 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$

Please note, the 2011 TCK for $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ has not been added, this should have a flatter dependence than 2012.

The relative trigger efficiency for $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ is shown in Fig. 43. The efficiency split into the different levels can be found in Sect. C in the Appendix. The trigger efficiency shape is driven by L0, as for the HLT the hadrons can participate which flattens the efficiency with q^2.

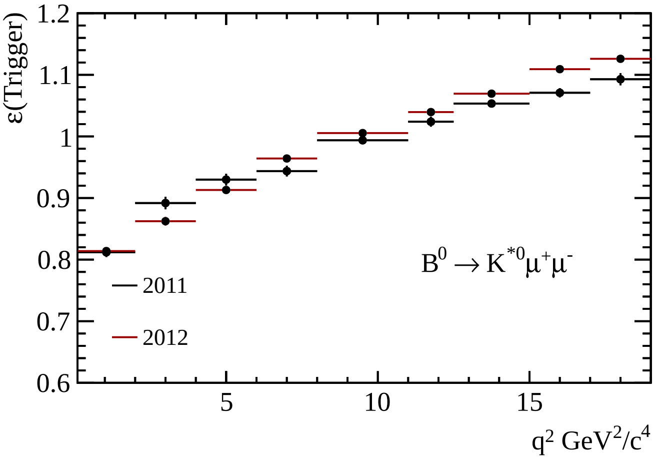


Figure 43: The relative trigger efficiency for the and selection efficiency between $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ and $B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$ as a function of q^2. The efficiency increases with q^2 as the trigger is dominated by the muons, which get harder in this region.

7.5 PID efficiency

The PID efficiency is calculated by applying the corrections obtained from the PIDCalib package for the DLL variables and using $B^+ \to J/\psi K^+$ for the ProbNN variables. In general, PID has no effect on the efficiency across q^2.

7.5.1 $B^+ \to K^+\mu^+\mu^-$

The relative PID efficiency for $B^+ \to K^+\mu^+\mu^-$, is shown in Fig. 44. It is clear from how flat these plots are that any systematic associated with the PID will be negligible.

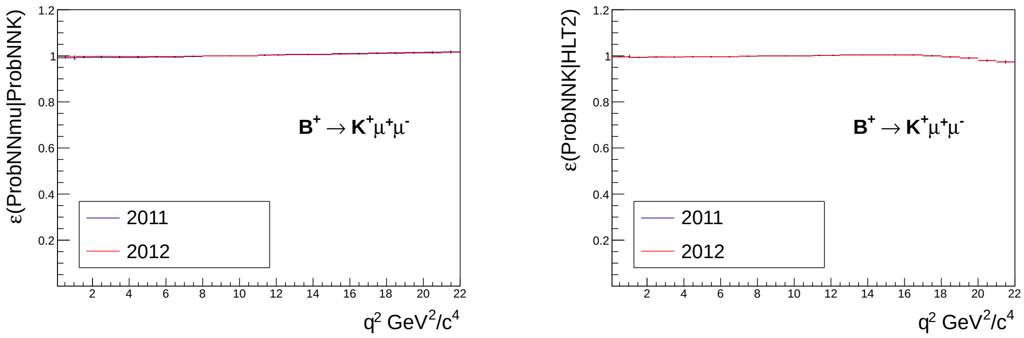


Figure 44: The relative muon (left) and hadron (right) PID efficiency between $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to J/\psi K^+$ as a function of q^2. The muon PID efficiency increases with q^2 the muons have higher momentum in this region. The kaon PID efficiency is almost flat, with a slight drop at high q^2 which is either due to the RICH kaon threshold or due to ghost rejection in the NN PID variables.

7.5.2 $B^0 \to K^0_S\mu^+\mu^-$

The muon PID efficiency for $B^0 \to K^0_S\mu^+\mu^-$ is shown in Fig. 45. The efficiency rises very slightly with q^2, as the muons have higher momentum in this region.

7.5.3 $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$

The muon PID efficiency for $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ is shown in Fig. 46. The efficiency rises very slightly with q^2, as the muons have higher momentum in this region.

7.5.4 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$

The muon PID efficiency for $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ is shown in Fig. 47. The efficiency lowers slightly for high q^2 as the hadrons get harder in this region where they PID selection is less efficient.

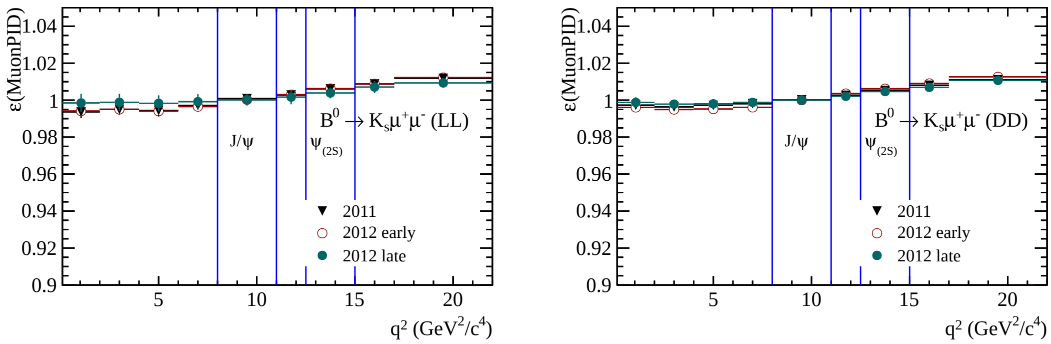


Figure 45: The relative muon PID efficiency between $B^0 \to K^0_S \mu^+\mu^-$ and $B^0 \to J/\psi\, K^0_S$ for the LL (left) and DD (right) categories as a function of q^2. The muon PID efficiency increases with q^2 as the muons have higher momentum in this region.

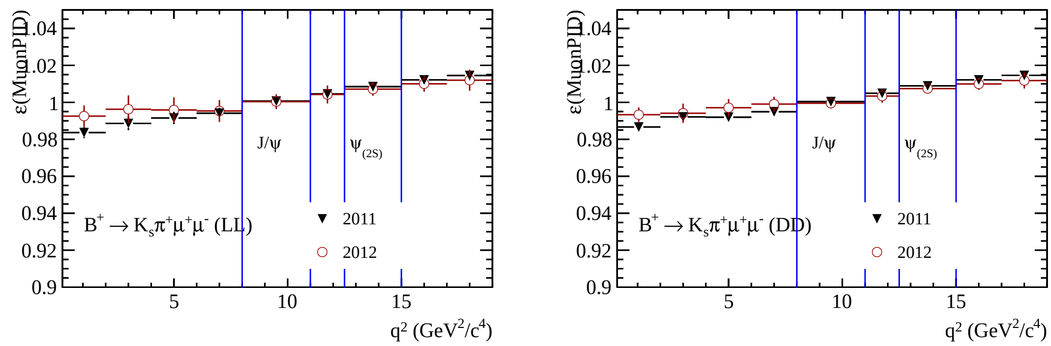


Figure 46: The relative muon PID efficiency between $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ and $B^+ \to J/\psi\,(K^{*+} \to K^0_S\pi^+)$ for the LL (left) and DD (right) categories as a function of q^2. The muon PID efficiency increases with q^2 as the muons have higher momentum in this region.

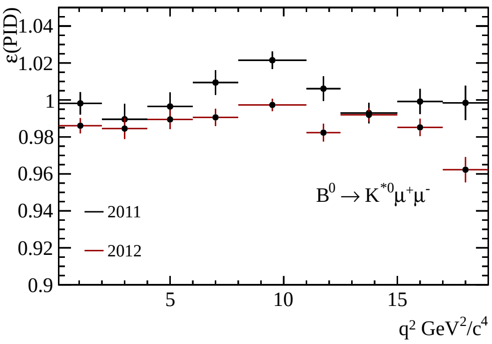


Figure 47: The relative PID efficiency between $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ and $B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$ as a function of q^2. The PID efficiency gets worse with q^2 as harder hadrons have higher momentum in this region.

7.6 BDT efficiency

The BDT efficiency as a function of q^2 for each channel is shown in this Section. In general, the BDT efficiency tends to get worse with q^2 as the most discriminating variables are associated with the hadron and so favour the lower q^2 region where the hadron has higher p_{T}, IP etc.

7.6.1 $B^+ \to K^+\mu^+\mu^-$

The relative BDT efficiency between $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to J/\psi\, K^+$ is shown in Fig. 52.

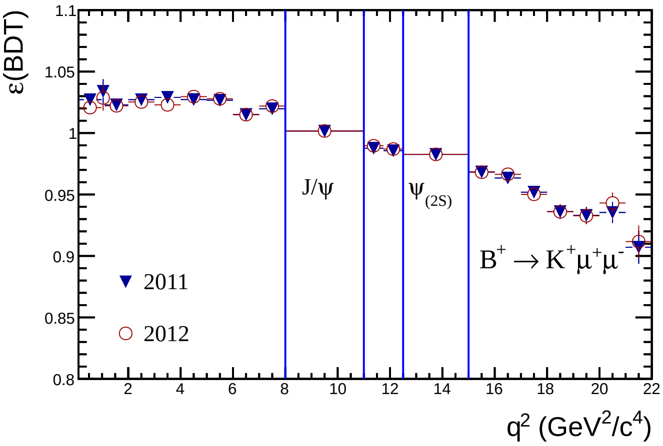


Figure 48: $B^+ \to K^+\mu^+\mu^-$ BDT efficiency as a function of q^2.

7.6.2 $B^0 \to K^0_{\mathrm{S}}\mu^+\mu^-$

The relative BDT efficiency between $B^0 \to K^0_{\mathrm{S}}\mu^+\mu^-$ and $B^0 \to J/\psi\, K^0_{\mathrm{S}}$ is shown in Fig. 53. For the LL category, there is a higher BDT efficiency at low q^2 late in 2012, which is due to the LL bug in the trigger, which required the K^0_{S} lifetime to be lower than 10 ps. These K^0_{S} which are TOS have very good IP χ^2 and thus are subsequently preferentially selected in the BDT.

7.6.3 $B^+ \to (K^{*+} \to K^0_{\mathrm{S}}\pi^+)\mu^+\mu^-$

The relative BDT efficiency between $B^+ \to (K^{*+} \to K^0_{\mathrm{S}}\pi^+)\mu^+\mu^-$ and $B^+ \to J/\psi\,(K^{*+} \to K^0_{\mathrm{S}}\pi^+)$ is shown in Fig. 55. The efficiency is fairly flat as the BDT efficiency in this channel is very efficient.

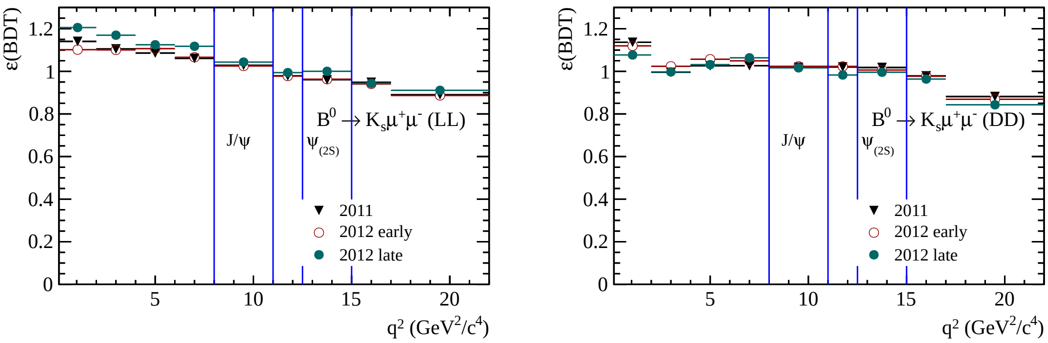


Figure 49: BDT efficiency on $B^0 \to K^0_S\mu^+\mu^-$ as a function of q^2 for the LL (left) and DD (right) categories. The efficiency gets worse with q^2 because the K^0_S p_T and IP get worse with q^2.

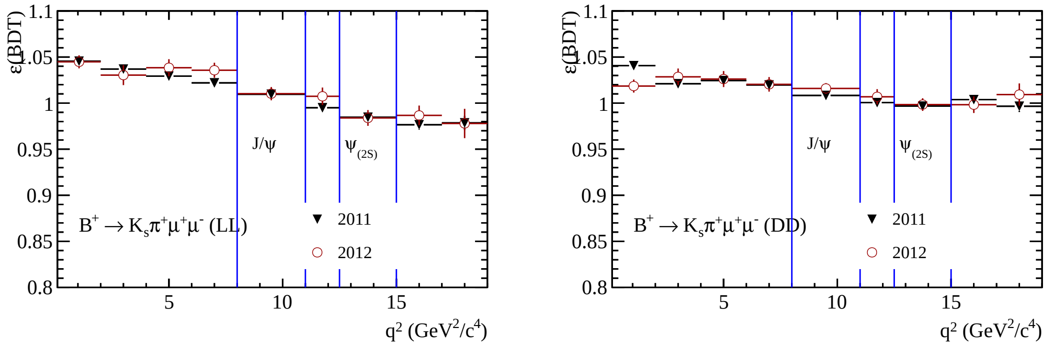


Figure 50: BDT efficiency on $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ as a function of q^2 for the LL (left) and DD (right) categories. The efficiency gets worse with q^2 because the K^0_S p_T and IP get worse with q^2.

7.6.4 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$

The relative BDT efficiency between $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ and $B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$ is shown in Fig. 51. The efficiency is flat due to the variables chosen to train the BDT.

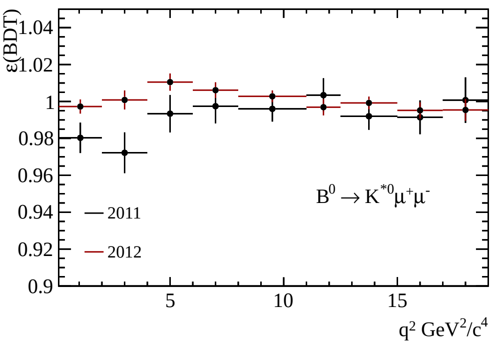


Figure 51: BDT efficiency on $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ as a function of q^2.

7.7 Total efficiency

This section describes the combination of all the previously described effects. These histograms are the ones used for the final result to convert signal yields into branching fractions and isospin asymmetries.

7.7.1 $B^+ \to K^+\mu^+\mu^-$

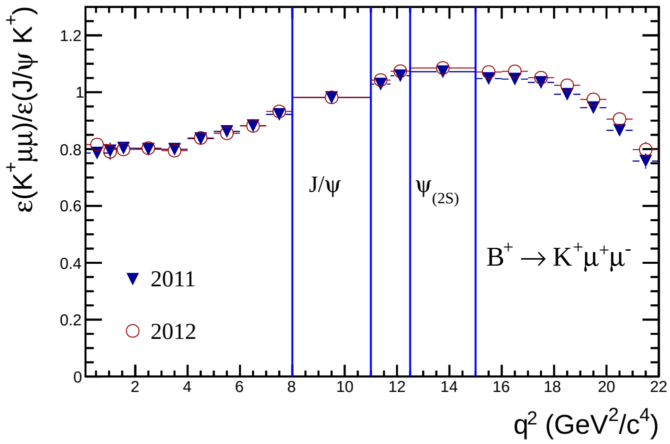


Figure 52: Relative efficiency between $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to J/\psi\, K^+$ as a function of q^2. The shape is a combination of effects described in the previous sections.

7.7.2 $B^0 \to K^0_S\mu^+\mu^-$

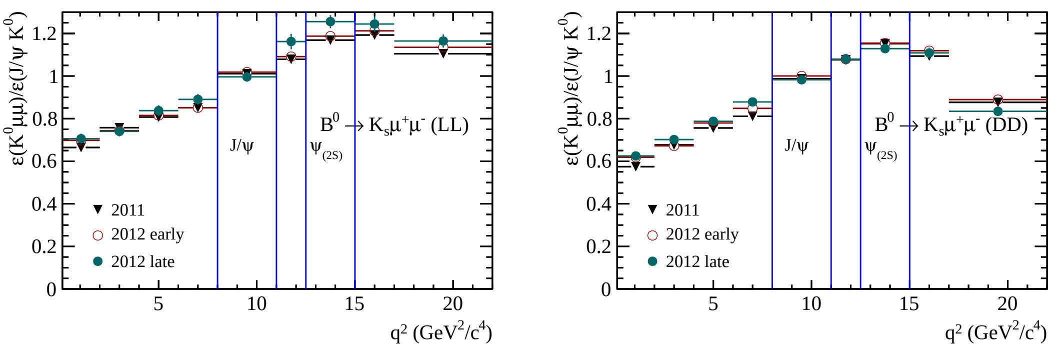


Figure 53: Relative efficiency between $B^0 \to K^0_S\mu^+\mu^-$ and $B^0 \to J/\psi\, K^0_S$ for the LL (left) and DD (right) categories as a function of q^2. The shape is a combination of effects described in the previous sections.

7.7.3 $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$

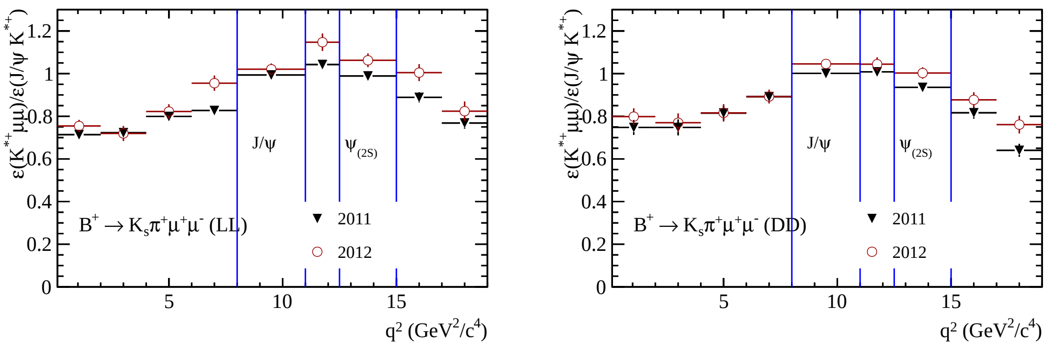


Figure 54: Relative efficiency between $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ and $B^+ \to J/\psi\,(K^{*+} \to K^0_S\pi^+)$ for the LL (left) and DD (right) categories as a function of q^2. The shape is a combination of effects described in the previous sections.

7.7.4 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$

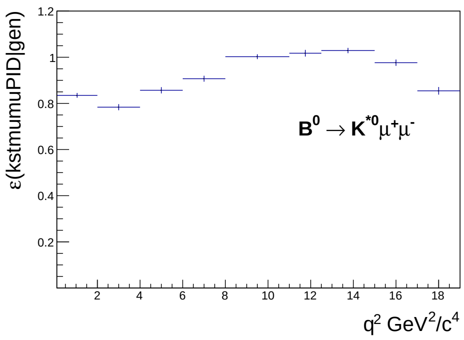


Figure 55: Relative efficiency between $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ and $B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$ as a function of q^2. The shape is a combination of effects described in the previous sections.

The raw q^2 distributions for the four channels are shown in Fig. 56 after the full offline selection. The difference between the K^+ and K^0_S distributions is small compared to the statistical sensitivity.

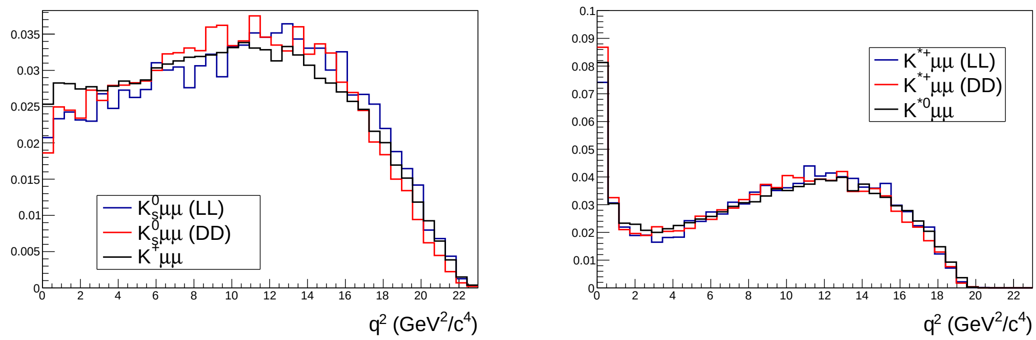


Figure 56: Distribution of q^2 for fully offline selected signal MC. Results are shown for all four signal decays.

8 Results

8.1 Branching fraction results

8.1.1 $B^+ \to K^+\mu^+\mu^-$

The differential branching fraction as a function of q^2 for $B^+ \to K^+\mu^+\mu^-$ is shown in Fig. 57. The $\psi(4160)$ is clearly visible at high q^2, and there is also a hint of enhancement at low q^2 as well, where the ρ and ω could contribute. The results are also split by year in Fig. 58, where they are compatible with each other. Finally, Tab. 14, tabulates the results, where the statistical and systematic uncertainties are shown separately. The branching fraction integrated over q^2 is obtained by extrapolating under the region removed due to charmonium resonances using the simulation. The result is $(4.42 \pm 0.07(stat) \pm 0.26(syst)) \times 10^{-7}$, which is compatible with previous results [16].

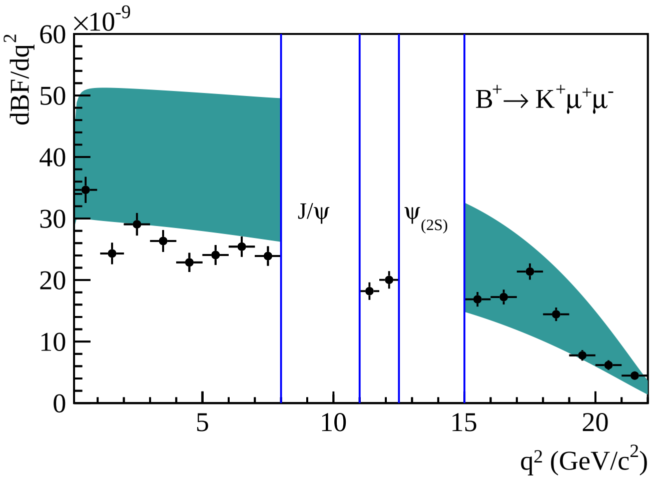


Figure 57: Differential branching fraction of $B^+ \to K^+\mu^+\mu^-$ decays as a function of q^2.

8.1.2 $B^0 \to K^0_S\mu^+\mu^-$

The differential branching fraction as a function of q^2 for $B^0 \to K^0_S\mu^+\mu^-$ is in Fig. 59. Here, the LL and DD categories are fit simultaneously, where the branching fraction is shared between the two. Note that the rate is suppressed at low q^2, which results in a negative isospin asymmetry (see later). The branching fraction integrated over q^2 is $(3.16 \pm 0.33(stat) \pm 0.16(syst)) \times 10^{-7}$, which is compatible with previous results [19].

8.1.3 $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$

The differential branching fraction as a function of q^2 for $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ is in Fig. 60. Here, the LL and DD categories are fit simultaneously, where the branching fraction is shared between the two. Tabulated results are shown in Tab 18. The branching

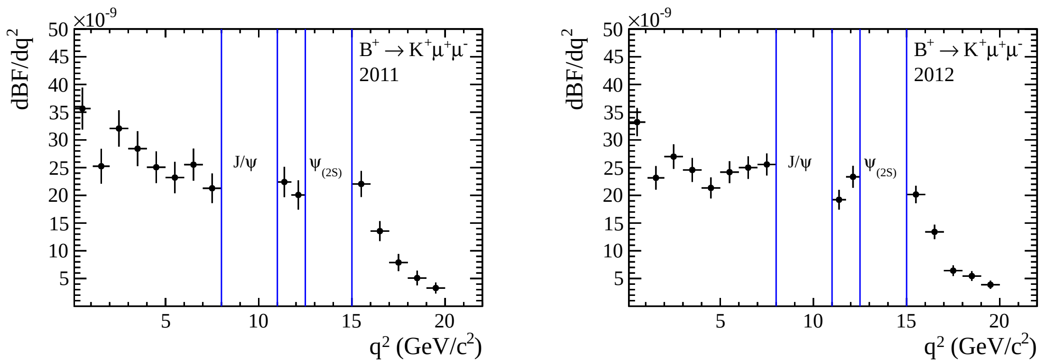


Figure 58: Differential branching fraction of $B^+ \to K^+\mu^+\mu^-$ decays as a function of q^2 split into the 2011 and 2012 results.

Table 14: Differential branching fraction results ($10^{-9} \times c^4/\mathrm{GeV}^2$) for the $B^+ \to K^+\mu^+\mu^-$ decay, including statistical and systematic uncertainties.

q^2 range (GeV^2/c^4)	central value	stat	syst
$0.1 < q^2 < 0.98$	33.2	1.8	1.7
$1.1 < q^2 < 2.0$	23.3	1.5	1.2
$2.0 < q^2 < 3.0$	28.2	1.6	1.4
$3.0 < q^2 < 4.0$	25.4	1.5	1.3
$4.0 < q^2 < 5.0$	22.1	1.4	1.1
$5.0 < q^2 < 6.0$	23.1	1.4	1.2
$6.0 < q^2 < 7.0$	24.5	1.4	1.2
$7.0 < q^2 < 8.0$	23.1	1.4	1.2
$11.0 < q^2 < 11.8$	17.7	1.3	0.9
$11.8 < q^2 < 12.5$	19.3	1.2	1.0
$15.0 < q^2 < 16.0$	16.1	1.0	0.8
$16.0 < q^2 < 17.0$	16.4	1.0	0.8
$17.0 < q^2 < 18.0$	20.6	1.1	1.0
$18.0 < q^2 < 19.0$	13.7	1.0	0.7
$19.0 < q^2 < 20.0$	7.4	0.8	0.4
$20.0 < q^2 < 21.0$	5.9	0.7	0.3
$21.0 < q^2 < 22.0$	4.3	0.7	0.2
$1.1 < q^2 < 6.0$	24.2	0.7	1.2
$15.0 < q^2 < 22.0$	12.1	0.4	0.6

fraction integrated over q^2 is $(9.11 \pm 0.92(stat) \pm 0.68(syst)) \times 10^{-7}$, which is compatible with previous results [19].

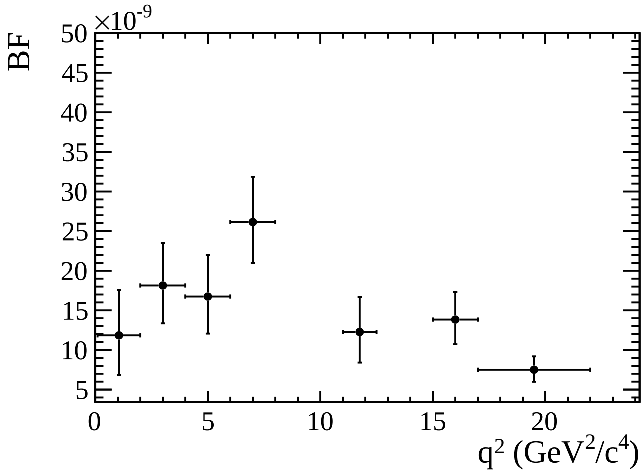


Figure 59: Differential branching fraction of $B^0 \to K^0_S \mu^+\mu^-$ decays as a function of q^2.

Table 15: Differential branching fraction results ($10^{-9} \times c^4/\mathrm{GeV}^2$) for the $B^0 \to K^0\mu^+\mu^-$ decay, including statistical and systematic uncertainties.

q^2 range (GeV^2/c^4)	central value	stat	syst
$0.1 < q^2 < 2.0$	12.2	$^{+5.9}_{-5.2}$	0.6
$2.0 < q^2 < 4.0$	18.7	$^{+5.5}_{-4.9}$	0.9
$4.0 < q^2 < 6.0$	17.3	$^{+5.3}_{-4.8}$	0.9
$6.0 < q^2 < 8.0$	27.0	$^{+5.8}_{-5.3}$	1.4
$11.0 < q^2 < 12.5$	12.7	$^{+4.5}_{-4.0}$	0.6
$15.0 < q^2 < 17.0$	14.3	$^{+3.5}_{-3.2}$	0.7
$17.0 < q^2 < 22.0$	7.8	$^{+1.7}_{-1.5}$	0.4
$1.1 < q^2 < 6.0$	18.7	$^{+3.5}_{-3.2}$	0.9
$15.0 < q^2 < 22.0$	9.5	$^{+1.6}_{-1.5}$	0.5

8.1.4 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$

Although no public results will be made for $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ decays, the branching fraction will be used for the isospin asymmetry and is shown here in Fig. 61. For the 2011 results, the angular analysis measurements are overlaid (stat only). The old angular analysis measurements agree nicely in all bins except the first.

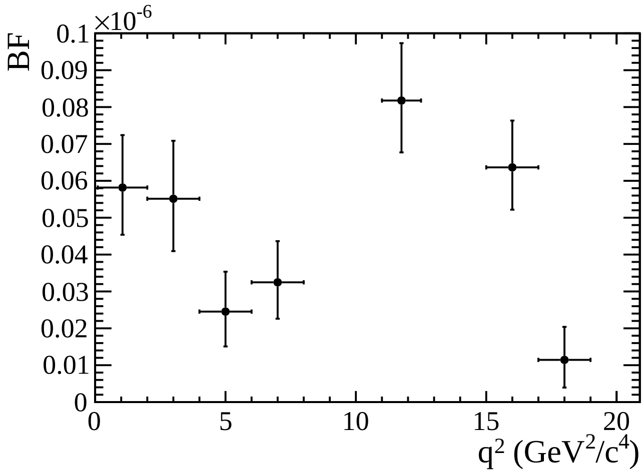


Figure 60: Differential branching fraction of $B^+ \to (K^{*+} \to K_S^0\pi^+)\mu^+\mu^-$ decays as a function of q^2.

Table 16: Differential branching fraction results ($10^{-9} \times c^4/\mathrm{GeV}^2$) for the $B^+ \to (K^{*+} \to K_S^0\pi^+)\mu^+\mu^-$ decay, including statistical and systematic uncertainties.

q^2 range (GeV^2/c^4)	central value	stat	syst
$0.1 < q^2 < 2.0$	59.2	$^{+14.4}_{-13.0}$	4.0
$2.0 < q^2 < 4.0$	55.9	$^{+15.9}_{-14.4}$	3.8
$4.0 < q^2 < 6.0$	24.9	$^{+11.0}_{-\ 9.6}$	1.7
$6.0 < q^2 < 8.0$	33.0	$^{+11.3}_{-\ 10.0}$	2.3
$11.0 < q^2 < 12.5$	82.8	$^{+15.8}_{-14.1}$	5.6
$15.0 < q^2 < 17.0$	64.4	$^{+12.9}_{-11.5}$	4.4
$17.0 < q^2 < 19.0$	11.6	$^{+\ 9.1}_{-\ 7.6}$	0.8
$1.1 < q^2 < 6.0$	36.6	$^{+\ 8.3}_{-\ 7.6}$	2.6
$15 < q^2 < 19.0$	39.5	$^{+\ 8.0}_{-\ 7.3}$	2.8

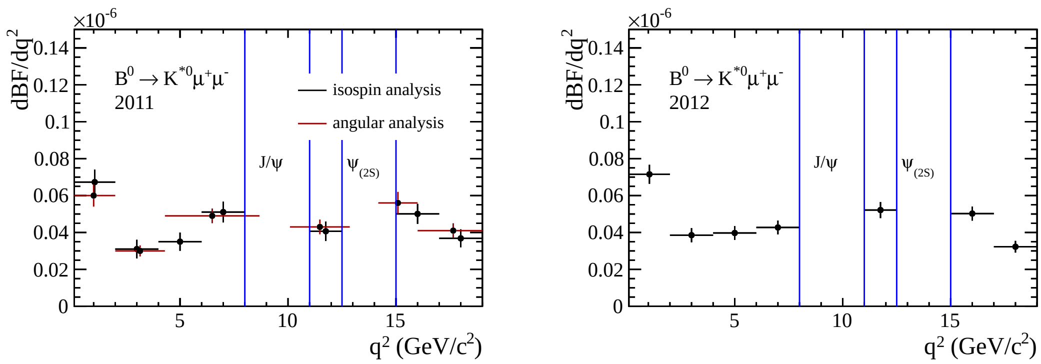


Figure 61: Differential branching fraction of $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ decays as a function of q^2 split into the 2011 and 2012 results. For 2011, the angular analysis measurements are overlaid (stat only).

8.2 Isospin asymmetry results

There are around 100 (60) candidates for each of the K^0_S DD (LL) modes, and so around 20 (10) candidates in each q^2 bin. With these low statistics, the resulting branching fraction errors will be asymmetric due to poisson statistics, and combining them is not trivial. It is therefore preferable to combine the likelihoods of each mode. This is done by rearranging the signal yields in the fit so that A_I becomes a fit parameter. The fit then automatically propagates the statistical errors to A_I by combining the likelihoods of all the signal yields. This also means systematics can be added to the fit and a significance from the SM expectation or $A_I = 0$ hypothesis can be easily obtained.

For each q^2 bin, six yields are needed; LL K^0_S signal, LL K^0_S J/ψ, DD K^0_S signal, DD K^0_S J/ψ, K^+ signal, and finallyK^+ J/ψ. Only the K^0_S signal yields are combined as the J/ψ and K^+ modes have enough statistics to be combined using error propagation. This leaves two signal likelihoods to combine with three J/ψ modes and one K^+ signal yield fit independently. Although none of the signal yields share any parameters, the two signal yields are fitted simultaneously. The K^0_S signal yields are re-expressed in terms of the K^+ yields and A_I which links all the two cateogires together. Equations 10 to 13 show this rearrangement.

Taking the isospin asymmetry of $B \to K\mu^+\mu^-$ ($B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$) as an example, equation 2 can be rearranged to give the $\mathcal{B}(K^0\mu^+\mu^-)$, the $B^0 \to K^0_S\mu^+\mu^-$ branching fraction as a function A_I and $\mathcal{B}(K^+\mu^+\mu^-)$, the $B^+ \to K^+\mu^+\mu^-$ branching fraction, shown in equation (10).

$$\mathcal{B}(K^0\mu^+\mu^-) = \frac{1 + A_I}{1 - A_I}\frac{\tau_{B^0}}{\tau_{B^+}}\mathcal{B}(K^+\mu^+\mu^-) \tag{10}$$

For example the LL $B^0 \to K^0\mu^+\mu^-$ branching fraction for the i^{th} q^2 bin is shown in equation 11,

$$\mathcal{B}(K^0\mu^+\mu^-)^i = \frac{N^i_{LL}(K^0_S\mu^+\mu^-)\mathcal{B}(J/\psi\, K^0)}{\epsilon^i_{LL}N_{LL}(J/\psi\, K^0_S)} \tag{11}$$

where $N^i_{LL}(K^0_S\mu^+\mu^-)$ is the signal yield in the i^{th} q^2 bin, $\mathcal{B}(J/\psi\, K^0)$ is the branching fraction of $B^0 \to J/\psi\, K^0$ obtained from in Sect. 9.1, $N(J/\psi\, K^0_S)$ is the control channel yield and ϵ is the relative efficiency between the signal and normalisation channels in the i^{th} q^2 bin. Substituting these branching fraction expressions into (10) yields equation (12).

$$\frac{N^i_{LL}(K^0_S\mu^+\mu^-))\mathcal{B}(J/\psi\, K^0)}{\epsilon^i_{LL}N_{LL}(J/\psi\, K^0)} = \frac{1 + A_I}{1 - A_I}\frac{\tau_{B^0}}{\tau_{B^+}}\frac{N^i(K^+\mu^+\mu^-)\mathcal{B}(J/\psi\, K^+)}{\epsilon^i_{K^+}N(J/\psi\, K^+)} \tag{12}$$

rearranging equation (12) in terms of the $B^0 \to K^0_S\mu^+\mu^-$ LL signal yield gives equation (13).

$$N^i_{LL}(K^0_S\mu^+\mu^-) = \mathrm{S^i}\frac{1 + A_I}{1 - A_I} \tag{13}$$

where

$$\mathrm{S^i} = N^i_{K^+\mu^+\mu^-}\frac{\tau_{B^0}}{\tau_{B^+}}\frac{\epsilon^i_{LL}N_{LL}(J/\psi\, K^0_\mathrm{S})}{\epsilon^i_{K^+}N(J/\psi\, K^+)}\frac{\mathcal{B}(J/\psi\, K^+)}{\mathcal{B}(J/\psi\, K^0)} \tag{14}$$

S is a combination of the B lifetimes, relative efficiencies and J/ψ yields which added in as external gaussian constraint on the fit, with the gaussian width set as the error calculated by propagating all the systematics to S beforehand. The average systematic error in each bin is around 5%, compared to the statistical error of 30%. To combine the LL category with the DD category, the same equation involving the DD yields is added, and is fit simultaneously with the LL category. A_I is then a shared parameter between the two, which enforces that the LL and DD categories give the same result. The parts of the systematic, S which are correlated between the LL and DD categories (e.g. $B^0 \to J/\psi\, K^0_\mathrm{S}$ branching fraction) are shared between the LL and DD fits.

8.2.1 Significance from $A_I = 0$ hypothesis

This section describes two statistical tests used to quote the significance of the results with respect to $A_I = 0$. As a reminder, for the 2011 result the significance was obtained by adding up the difference in likelihood (DLL) between a fit where A_I is let free and where A_I is fixed to zero for each bin and using Wilk's theorem (with one degree of freedom) to convert this into a significance. This was found to be 4.4 σ. However this was chosen after the data was seen, and relied on the fact that all bins were negative, which may not be the case this time round.

For the updated analysis, the simplest method would be to perform a χ^2 test on the A_I measurements with respect to a horizontal straight line at A_I =0. This ignores any shape information (e.g. if all bins are negative), however it is statistically well defined and is easy to describe in the paper. For the old result this comes to 2.9 σ. Although this is much more conservative than the method used previously, it has the advantage that is blind to the shape of A_I.

A more discriminating test is to assume a shape for A_I across q^2. As the shape is currently unknown, we test the simplest alternative hypothesis, which is a constant value different from zero. This fit to a constant A_I is shown in Fig. 62 as an example. This fit to a constant A_I to the seven bins can be performed for toy datasets, where the measurements are generated from $A_I = 0$ for each bin independently. The χ^2 of the fit where A_I is free is then compared to the χ^2 obtained where A_I is fixed to zero. The difference in these χ^2 values defines the test statistic. The distribution of the test statistic for 20,000 toy datasets is shown in Fig. 63. The p-value of the 2011 result is roughly 0.04% with this method, which corresponds to 3.5 σ. Compared to the simple χ^2 test, this method is much more powerful but has the disadvantage of assuming a shape for A_I.

Systematic uncertainties, in particular the J/ψ branching fraction systematic, can have a large large effect on the p-value estimation as they are 100% correlated across q^2. This systematic is roughly 7%, theoretically however, the J/ψ modes are expected to have zero isospin asymmetry at the level of roughly 1% [20]. For this reason, it is also interesting to quote a p-value assuming the J/ψ branching fractions are isospin symmetric.

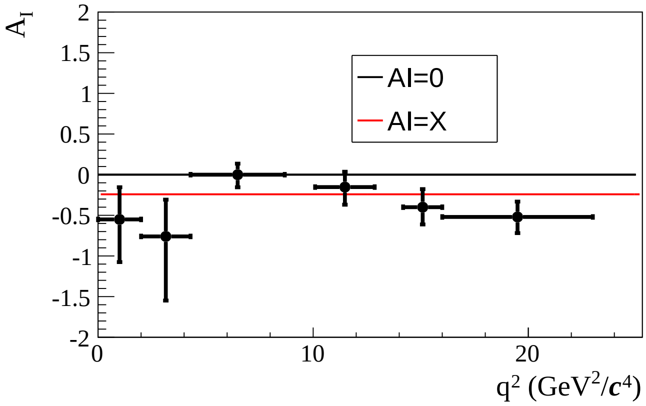


Figure 62: Fit to the 2011 data for a constant A_I. The χ^2 of the fit compared to the χ^2 with A_I fixed to zero defines the test statistic for determining the p-value from zero.

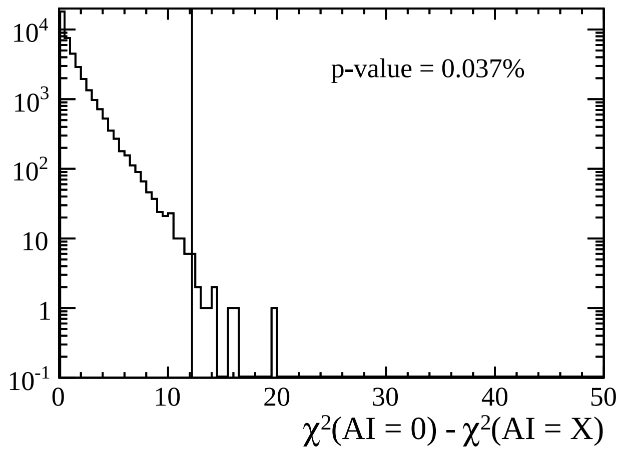


Figure 63: Distribution of the difference in χ^2 when A_I is fixed to zero and let free but constant across q^2. The vertical line shows the test statistic value for the 2011 result.

8.2.2 $B \to K\mu^+\mu^-$

The $B \to K\mu^+\mu^-$ isospin asymmetry is shown in Fig. 64. These results are obtained assuming that J/ψ isospin asymmetry is zero.

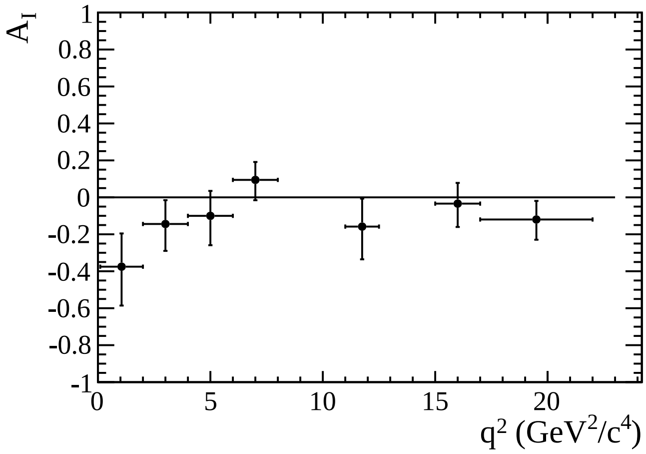


Figure 64: Isospin asymmetry of $B \to K\mu^+\mu^-$ as a function of q^2.

	Isospin asymmetry		
q^2 range	central value	stat error	syst error
$0.1 < q^2 < 2$	-0.37	$^{+0.18}_{-0.21}$	0.02
$2 < q^2 < 4$	-0.15	$^{+0.13}_{-0.15}$	0.02
$4 < q^2 < 6$	-0.10	$^{+0.13}_{-0.16}$	0.02
$6 < q^2 < 8$	0.09	$^{+0.10}_{-0.11}$	0.02
$11 < q^2 < 12.5$	-0.16	$^{+0.15}_{-0.18}$	0.03
$15 < q^2 < 17$	-0.04	$^{+0.11}_{-0.13}$	0.02
$17 < q^2 < 19$	-0.12	$^{+0.10}_{-0.11}$	0.02
$1.1 < q^2 < 6$	-0.10	$^{+0.08}_{-0.09}$	0.02
$15 < q^2 < 19$	-0.09	$^{+0.08}_{-0.08}$	0.02

Table 17: Isospin asymmetry results for $B \to K\mu^+\mu^-$

The data is well described by a zero parameter polynomial (horizontal straight line) with a χ^2 probability of 54%. This is the naive model used to obtain the p-value which will go into the paper. The observed value overlaid on the distribution of toy datasets generated from A_I =0 is shown in Fig. 65 and corresponds to a p-value of 11% ($\sim 1.5\sigma$). If one instead uses the simple χ^2 test with respect to zero, which assumes no shape in A_I, the p-value comes out to be 52%. This disparity between the two methods is expected as the simple χ^2 test ignores the fact that nearly all measurements are negative.

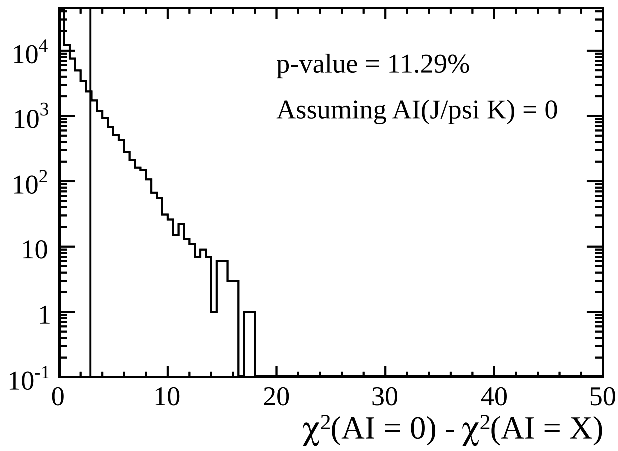


Figure 65: Distribution of the difference in χ^2 when A_I is fixed to zero and let free but constant across q^2. The vertical line shows the observed test statistic value for the result.

Additionally, one can relax the assumption that the isospin asymmetry of the J/ψ modes is zero, and take the values described in Sect. 9.1. The result is shown in Fig. 66, with the corresponding in Fig. 67. When using the PDG J/ψ branching fraction values, the observed value of the test statistic is higher as the A_I data points become more negative so that the χ^2 with respect to zero is higher. However, the distribution of the toy datasets is also wider because of the relatively large systematic (7%), which is 100% correlated across q^2. The p-value is about 6.6%, which corresponds to about 1.9σ.

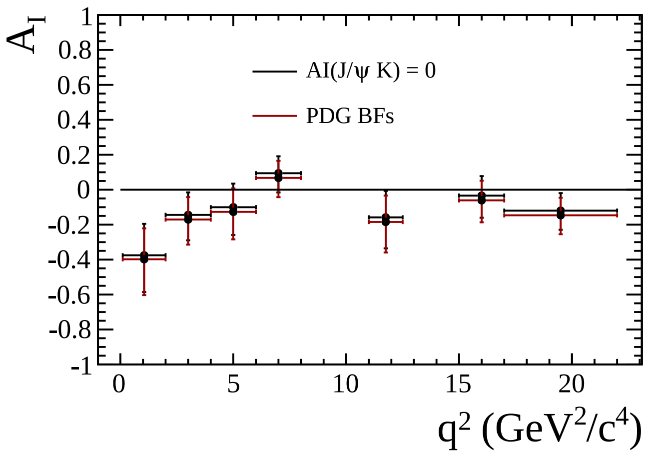


Figure 66: Isospin asymmetry of $B \to K\mu^+\mu^-$ as a function of q^2 with and without the assumption that $AI(J/\psi K^+) = 0$.

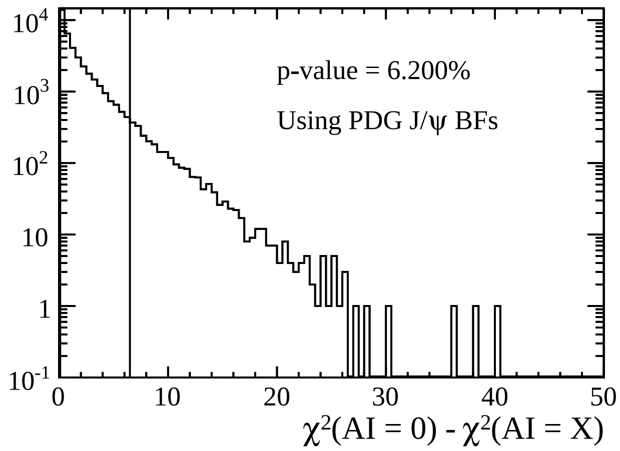


Figure 67: Distribution of the test statistic without the assumption that $AI(J/\psi K^+) = 0$.

8.2.3 $B \to K^*\mu^+\mu^-$

The $B \to K^*\mu^+\mu^-$ isospin asymmetry is shown in Fig. 68, where results are consistent with zero like last time. These results are obtained assuming that J/ψ isospin asymmetry is zero.

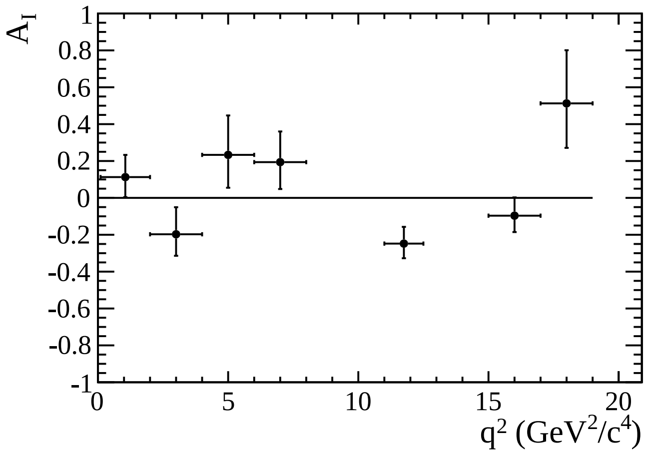


Figure 68: Isospin asymmetry of $B \to K^*\mu^+\mu^-$ as a function of q^2.

	Isospin asymmetry		
q^2 range	central value	stat error	syst error
$0.1 < q^2 < 2$	0.11	${}^{+0.12}_{-0.11}$	0.02
$2 < q^2 < 4$	-0.20	${}^{+0.15}_{-0.12}$	0.03
$4 < q^2 < 6$	0.23	${}^{+0.21}_{-0.18}$	0.02
$6 < q^2 < 8$	0.19	${}^{+0.17}_{-0.15}$	0.02
$11 < q^2 < 12.5$	-0.25	${}^{+0.09}_{-0.08}$	0.03
$15 < q^2 < 17$	-0.10	${}^{+0.10}_{-0.09}$	0.03
$17 < q^2 < 19$	0.51	${}^{+0.29}_{-0.24}$	0.02
$1.1 < q^2 < 6$	0.00	${}^{+0.12}_{-0.10}$	0.02
$15 < q^2 < 19$	0.059	${}^{+0.1}_{-0.09}$	0.02

Table 18: Isospin asymmetry results for $B \to K^{*}\mu^{+}\mu^{-}$

9 Systematic uncertainties

9.1 $B \to J/\psi\, h$ branching fraction

The branching fraction measurements of the normalisation modes from the B-factory experiments assume that the B^+ and B^0 mesons are produced with equal proportions at the $\Upsilon(4S)$ resonance [**?**, **?**, 21]. In contrast, in this paper isospin symmetry is assumed for the $B \to J/\psi\, K^{(*)}$ decays, implying that the $B^+ \to J/\psi\, K^+$ ($B^+ \to J/\psi\,(K^{*+} \to K^0_{\rm S}\pi^+)$) and $B^0 \to J/\psi\, K^0$ ($B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$) decays have the same partial width. The branching fractions used in the normalisation are obtained by: taking the most precise branching fraction results from Ref. [21] and translating them into partial widths; averaging the partial widths of the K^+, K^0 and the K^{*+}, K^{*0} modes, respectively; and finally translating the widths back to branching fractions. The calculation only requires knowledge of the ratio of B^0 and B^+ lifetimes for which we use 0.93 ± 0.01 [22]. Statistical uncertainties are treated as uncorrelated while systematical uncertainties are conservatively treated as fully correlated. The resulting branching fractions of the normalisation channels are

$$
\begin{aligned}
\mathcal{B}(B^+ \to J/\psi\, K^+) &= (0.998 \pm 0.014 \pm 0.040) \times 10^{-3},\\
\mathcal{B}(B^0 \to J/\psi\, K^0) &= (0.928 \pm 0.013 \pm 0.037) \times 10^{-3},\\
\mathcal{B}(B^+ \to J/\psi\, K^{*+}) &= (1.431 \pm 0.027 \pm 0.090) \times 10^{-3},\\
\mathcal{B}(B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)) &= (1.331 \pm 0.025 \pm 0.084) \times 10^{-3},
\end{aligned}
$$

where the first uncertainty is statistical and the second systematic. The uncertainties on the branching fractions of the normalisation modes constitute the dominant source of systematic uncertainty on the branching fraction measurements while it cancels in the isospin measurements.

9.2 Physics model

Recently, a resonance was discovered in $B^+ \to K^+\mu^+\mu^-$ decays at high q^2 [23]. This resonance alters the q^2 shape for the highest q^2 bin in $B^0 \to K^0_{\rm S}\mu^+\mu^-$ (for $B^+ \to K^+\mu^+\mu^-$ the bins too narrow for it to be an issue). The effect that this resonance has on the efficiency is estimated by reweighing the simulation at high q^2 so that it looks like the fit to the $B^+ \to K^+\mu^+\mu^-$ data (see Fig. 69). The effect that this has on the efficiency is shown in Fig. 71, where the efficiency as a function of q^2 for $B^0 \to K^0_{\rm S}\mu^+\mu^-$ is shown before and after re-weighting for the resonance. As expected, the only observed difference in efficiency is for the highest and widest q^2 bin. As we cannot be sure that the $B^0 \to K^0_{\rm S}\mu^+\mu^-$ dimuon spectrum is the same as the $B^+ \to K^+\mu^+\mu^-$ spectrum, we use the difference in efficiency curves as a systematic rather than a correction. There is also the possibility of resonances at low q^2 as well, such as the ρ, ω and ϕ. These low mass resonances are ignored as they have lower branching fractions than the $\psi(4160)$ and the q^2 bins are narrower which results in a negligible effect in the efficiency in this region. For $B^+ \to (K^{*+} \to K^0_{\rm S}\pi^+)\mu^+\mu^-$ and $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$, there is no significant evidence for the $\psi(4160)$ and the efficiency is flatter for these decays at high q^2. For these reasons no systematic is assigned for the $\psi(4160)$ for $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ and $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$.

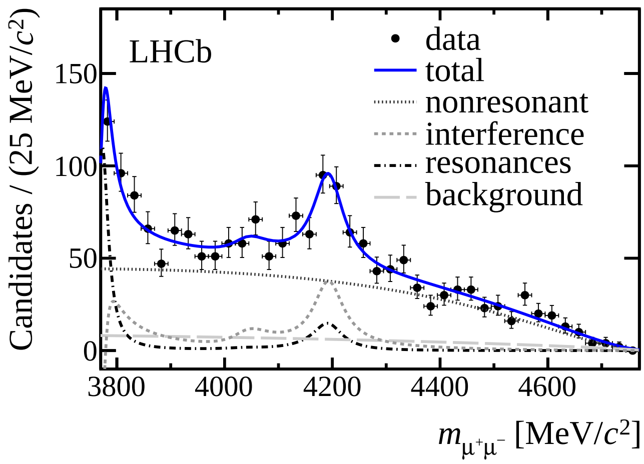


Figure 69: Fit to the dimuon spectrum of $B^+ \to K^+\mu^+\mu^-$ decays, including three charmonium resonances, the $\psi(4160)$, the $\psi(4040)$ and $\psi(3770)$. The blue curve is divided by the sum of the non-resonant curves to obtain weights for the simulation.

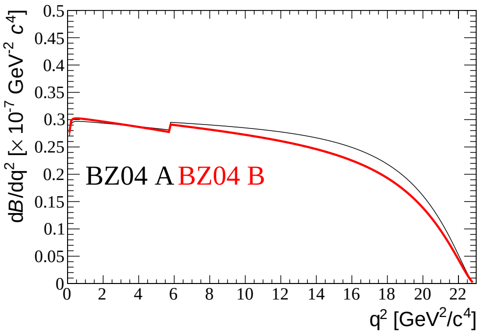


Figure 70: Distribution of q^2 after making the systematic variation as described in [?]. The slope changes at most by about 20%.

In addition to the specific correction due to the $\psi(4160)$, a more general systematic is assigned due to imperfect knowledge of form-factors and possible effects of new physics contributions to the shape within a q^2 bin. To assess this systematic, the number of q^2 bins is doubled when calculating the efficiency from simulation. These bins are then averaged under two weighting schemes, one where the lower half of each bin is given 20% the weight of the upper half and vice-versa. This 20% number is obtained by making the systematic variation of the form factor, suggested in Ref. [?], shown in Fig. 70. The difference between the weighting schemes is shown for $B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ decays

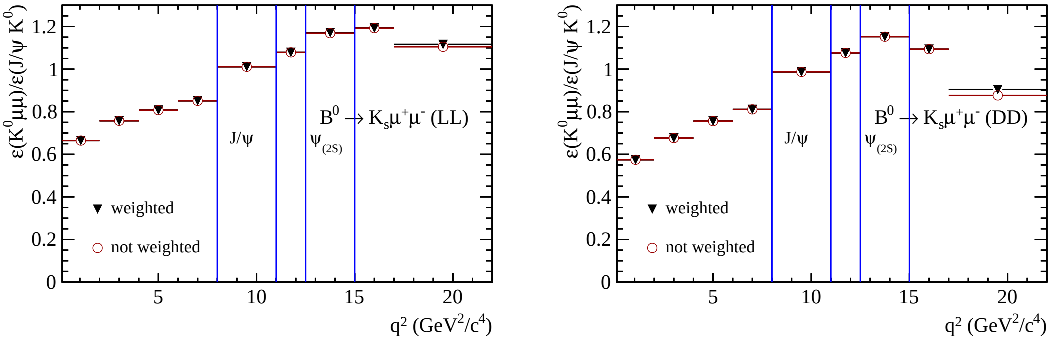


Figure 71: Relative efficiency between $B^0 \to J/\psi\, K^0_{\rm S}$ and $B^0 \to K^0_{\rm S}\mu^+\mu^-$ as a function of q^2 before and after the simulation is re-weighted due to a possible resonance at high q^2.

in Fig. 72. The difference is negligible compared to the statistical sensitively and so no systematic is assigned.

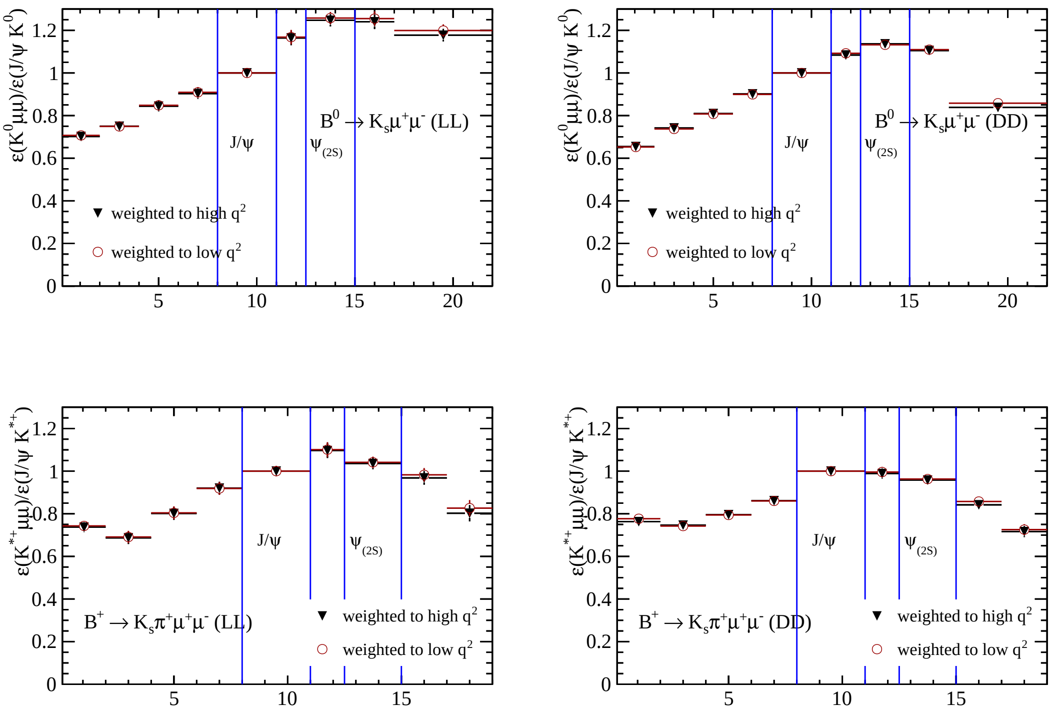


Figure 72: Relative efficiency between signal and normalisation channels as a function of q^2 when weighting simulation towards the high and low halves of each q^2 bin.

9.3 Trigger efficiency

9.3.1 K^+ channels

To assess how well the trigger efficiency for $B^+ \to K^+\mu^+\mu^-$ is re-produced in the simulation, the trigger efficiency as determined from the TISTOS method is compared to $B^+ \to J/\psi K^+$ candidates as a function of muon kinematics. This agreement is very good, even for L0, as shown in Fig. 73. Given this level of agreement no systematic is assigned for $B^+ \to K^+\mu^+\mu^-$.

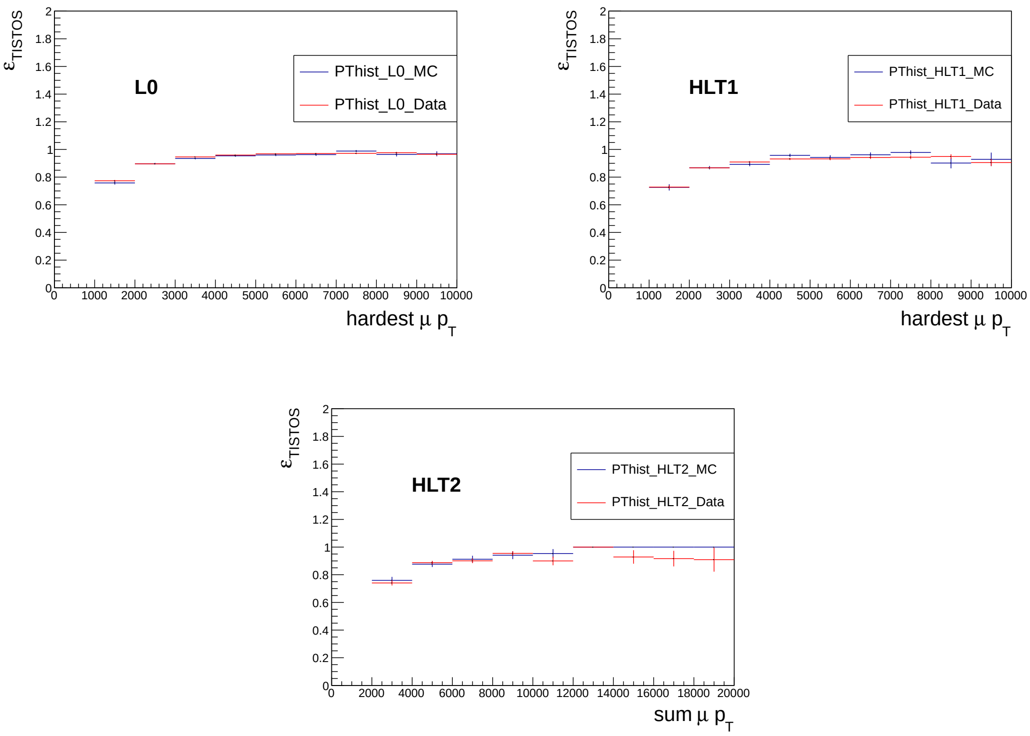


Figure 73: Trigger efficiency for $B^+ \to J/\psi K^+$ candidates as measured by the TISTOS method as a function of muon kinematics. The data/MC agreement is very good.

The TISTOS method is checked by comparing trigger efficiency obtained for $B^+ \to J/\psi K^+$ decays in simulation. The TISTOS trigger efficiency is 85% whereas the absolute efficiency is 81%. This level of disagreement is expected due to the assumption that TIS and TOS efficiency are independent of each other. When applied to both data and MC, and in bins of kinematics, this level of disagreement is expected to cancel so that the TISTOS efficiency is a good proxy for assessing the systematic uncertainty.

9.3.2 K^0_S channels

There are not enough data to make the same study for $B^0 \to J/\psi K^0_S$, however $B^0 \to K^0_S\mu^+\mu^-$ decays should be triggered mostly on the muons, which means that we can use the TISTOS method on $B^+\to J/\psi K^+$ decays, where the requirements are applied to the J/ψ to assess the systematic related to the muon triggers on $B^0\to K^0_S\mu^+\mu^-$. This study is shown in Fig. 74, where the agreement is very good. This is not a surprise given that the full $B^+\to J/\psi K^+$ candidate agrees nicely in Fig. 73.

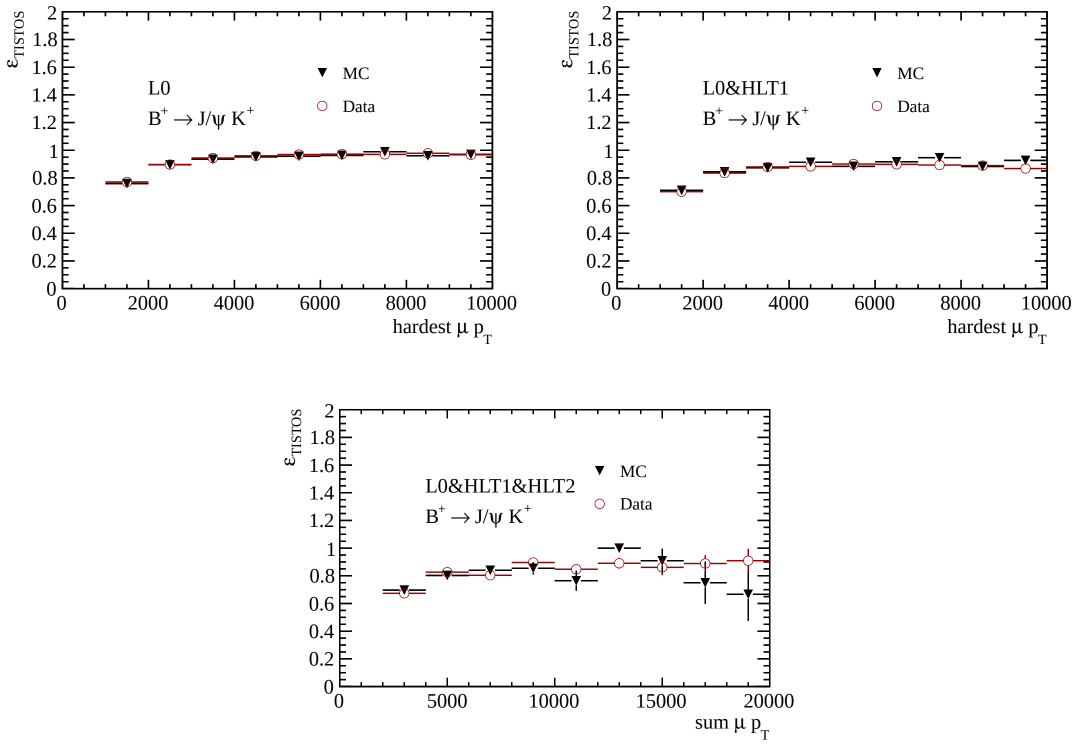


Figure 74: Trigger efficiency for J/ψ candidates in $B^+\to J/\psi K^+$ decays as measured by the TISTOS method as a function of muon kinematics. The data/MC agreement is very good.

The systematic associated with the remaining K^0_S contribution is estimated by comparing the fraction of candidates which are TOS on the 3-body topological lines, which is where the majority of K^0_S candidates participate. This fraction in data and simulation is shown in Fig. 75, where the simulation tends to overestimate the fraction of K^0_S participating. The data here is selected to avoid two K^0_S bugs in the trigger already described in Sect. 3: One corrected in June 2012 which resulted in a very low DD K^0_S trigger efficiency for the first $\sim 0.5\,\mathrm{fb}^{-1}$ of 2012. Another introduced in June 2012 which resulted in a very low LL K^0_S trigger efficiency for the last $\sim 1.5\,\mathrm{fb}^{-1}$ of 2012.

The disagreement between Data and simulation appears to be quite large, however

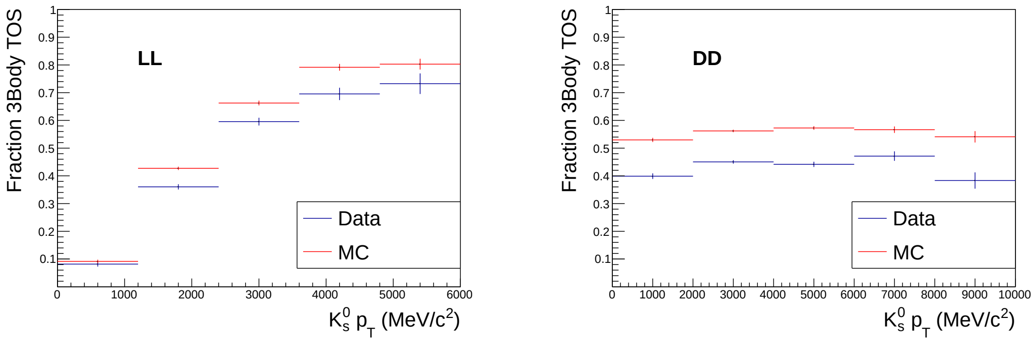


Figure 75: Fraction of offline selected candidates which fire the 3-body topological lines for data and MC as a function of K^0_S p_T. For both categories, the 3-body lines appear to be more efficient in the simulation.

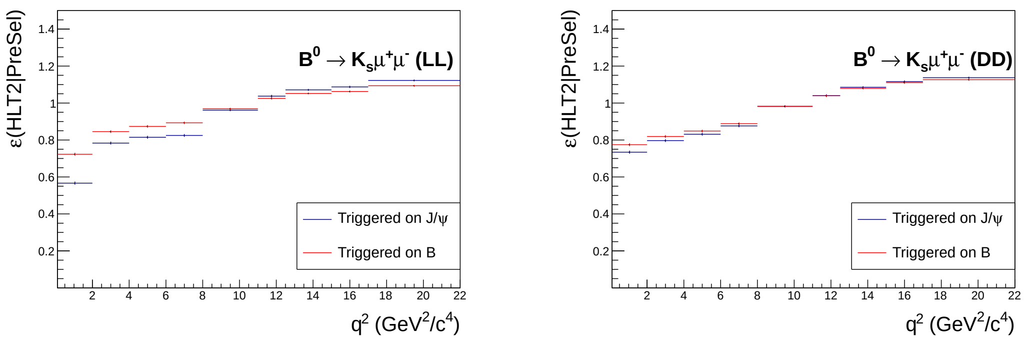


Figure 76: Comparison of the trigger efficiency on $B^0 \to K^0_S\mu^+\mu^-$ candidates when trigger requirements are placed on the B candidate and the J/ψ candidate.

most $B^0 \to K^0_S\mu^+\mu^-$ candidates would be triggered by the J/ψ anyway, as shown in Fig. 76, which compares the relative trigger efficiency between $B^0 \to K^0_S\mu^+\mu^-$ and $B^0 \to J/\psi K^0_S$ when applying requirements only on the J/ψ. The associated systematic is obtained by multiplying the fraction of candidates triggered by the K^0_S, shown in Fig. 76, by the data/MC disagreement shown in Fig. 75, which results in the systematic curve shown in Fig. 77. Here results are shown for $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ as well which are obtained in a similar way. Due to the K^0_S bugs mentioned earlier, this systematic only applies to roughly half the data in each category (DD candidates were not included in the 2011 topological trigger).

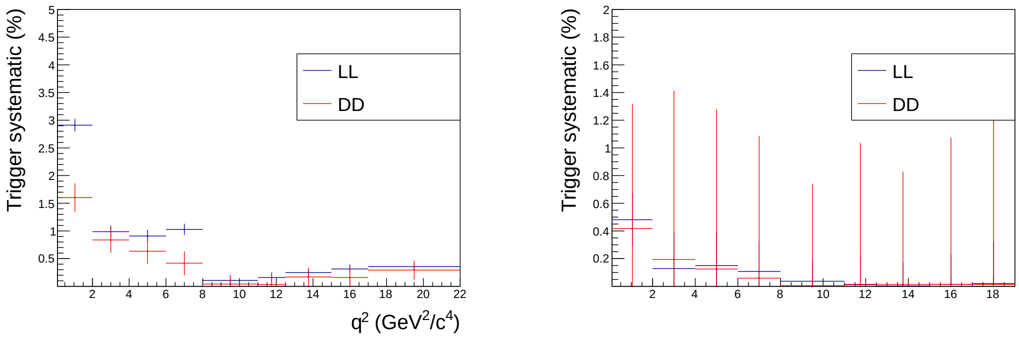


Figure 77: Systematic associated with the trigger for $B^0 \to K^0_\mathrm{S}\mu^+\mu^-$ (left) and $B^+ \to (K^{*+} \to K^0_\mathrm{S}\pi^+)\mu^+\mu^-$ (right) decays as a function of q^2.

9.4 Data/MC mis-modelling

As described in Sect. 7, the simulation is re-weighted to match the data. Due to the robust nature of the analysis via normalisation, this has a relatively small effect on the efficiency as a function of q^2.

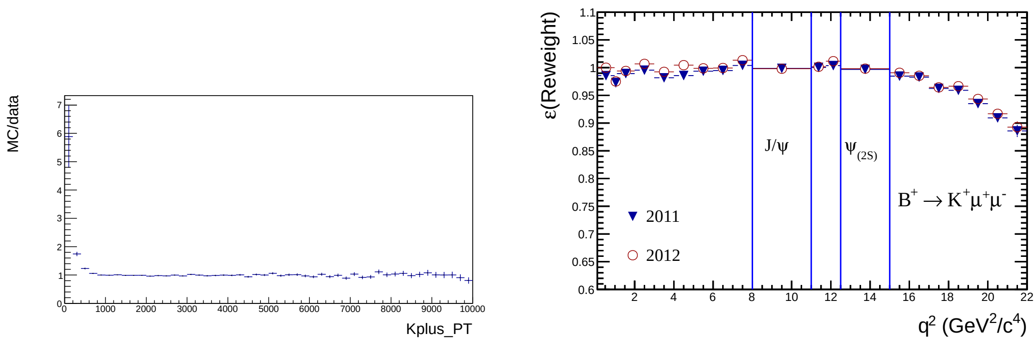


Figure 78: MC/Data distribution for K^+ p_T in $B^+ \to J/\psi\, K^+$ decays before re-weighting (left) and effect of this reweighing on the relative efficiency as a function of q^2 (right). There is a slight drop in efficiency at high q^2 due to the reweighing of the K^+ p_T.

However there is an effect for $B^+ \to K^+\mu^+\mu^-$, due to a discrepancy at low K^+ p_T, see Fig. 78. Correcting for this discrepancy removes more high q^2 $B^+ \to K^+\mu^+\mu^-$ than $B^+ \to J/\psi\, K^+$ as kaons are softer in that region. There is also a similar effect for the muon p_T, which are also re-weighted. In total, six variables are re-weighted for $B^+ \to K^+\mu^+\mu^-$, B p_T, B vertex χ^2, number of tracks and daughter p_T. These are reweighed independently of each other, and only one dimension. This means that the agreement is not perfect after re-weighting due to correlations, and other variables, such as the B IP, become badly modelled. Although the B IP is not correlated to q^2, this effect deserves a systematic, which

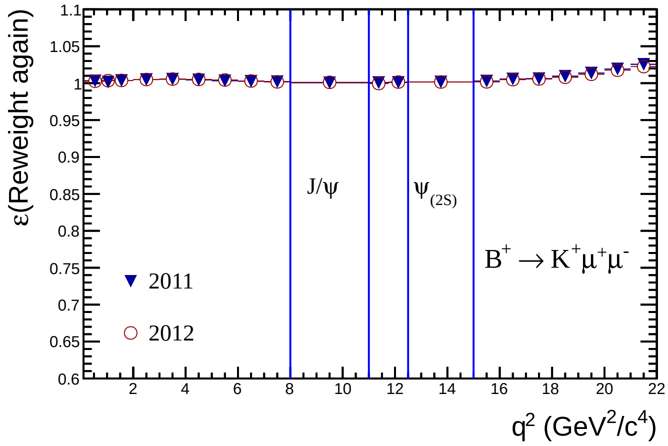


Figure 79: Effect of re-weighting $B^+ \to K^+\mu^+\mu^-$ simulation again to smooth out residual discrepancies due to correlations. This effect is used to estimate a systematic associated with remaining mis-modelling.

is obtained by re-weighting the variables again, to correct for the residual differences due to correlations. The difference in the efficiency from performing this second re-weighting, shown in Fig. 79 is used as a systematic.

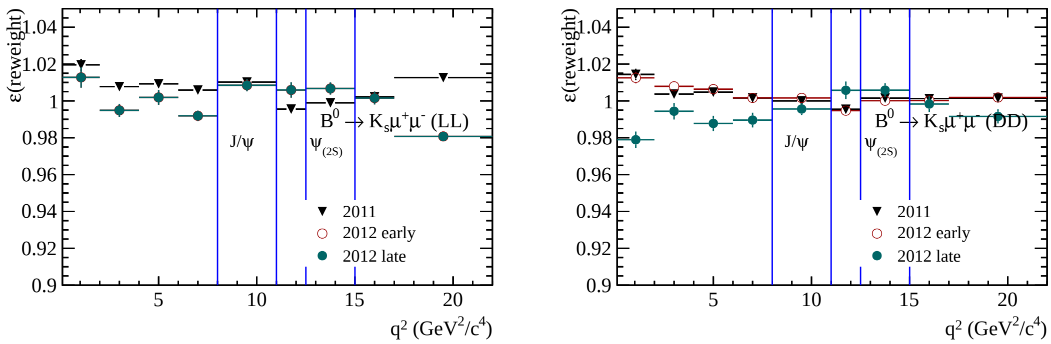


Figure 80: Effect of the reweighing on the relative efficiency as a function of q^2. There is no significant effect across q^2, the fluctuations are due to the weight uncertainties which are not propagated to this plot.

For $B^0 \to K^0_S\mu^+\mu^-$, the situation is different. Due to a much tighter selection, there are hardly any soft candidates left and so the discrepancy is not visible (see Sect. A in the appendix for more detail). The effect of applying the B and occupancy weights are shown in Fig. 80. There is no significant trend, although results do tend to fluctuate by roughly 2%, due to the weight errors which are not included. This 2% fluctuation is used as a systematic constant across q^2.

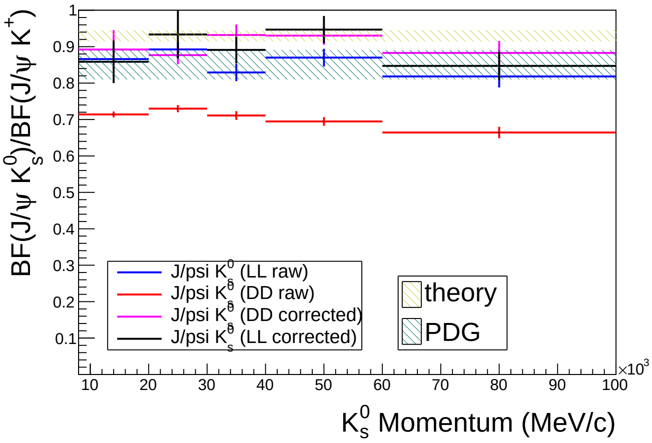


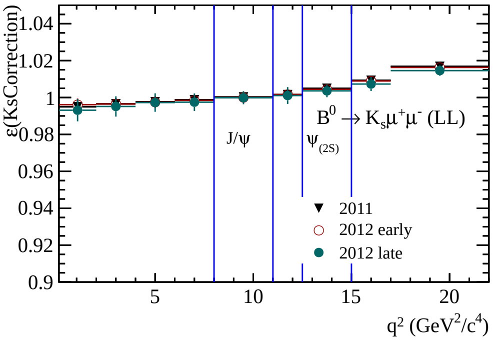
Figure 81: The branching fraction ratio, $\mathcal{B}(B^0 \to J/\psi\, K^0)/\mathcal{B}(B^+ \to J/\psi\, K^+)$, as a function of $K^0_{\rm S}$ momentum. After a correcting for the $K^0_{\rm S}$ reconstruction, the DD category agrees nicely with the theoretical predictions.

9.5 Negligible systematics

9.5.1 $K^0_{\rm S}$ reconstruction efficiency

It is known that the simulation does not reproduce the correct yield ratio between the two $K^0_{\rm S}$ reconstruction categories. The effect of possible mis-modelling of $K^0_{\rm S}$ reconstruction is estimated by measuring the $K^0_{\rm S}$ reconstruction efficiency in data using $D^0 \to \phi K^0_{\rm S}$ decays, the results of which reproduces the correct $B^0 \to J/\psi\, K^0_{\rm S}/B^+ \to J/\psi\, K^+$ branching fraction ratio nicely, see Fig. 81. More details on the $K^0_{\rm S}$ reconstruction technique can be found in Ref. [24]. As shown in Fig. 82, the effect of these corrections on the efficiency is very small, and so any associated systematic would be negligible.

The $K^0_{\rm S}$ decay Z position, psuedo-rapidity and ϕ coordinate distributions for $B^0 \to J/\psi\, K^0_{\rm S}$ and $B^+ \to J/\psi\,(K^{*+} \to K^0_{\rm S}\pi^+)$ decays are shown in Sect. D in the Appendix. The downstream tracking efficiency is also shown as a function of these variables.



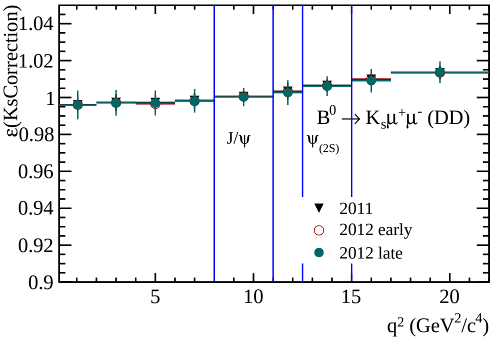


Figure 82: Effect of the K^0_S reconstruction correction on the relative efficiency between $B^0 \to J/\psi\, K^0_S$ and $B^0 \to K^0_S \mu^+\mu^-$ decays.

9.5.2 PID

The isMuon efficiency is taken from the simulation rather than derived from data, however it is well modelled in the simulation (see Fig. 83) and the corresponding systematic is negligible.

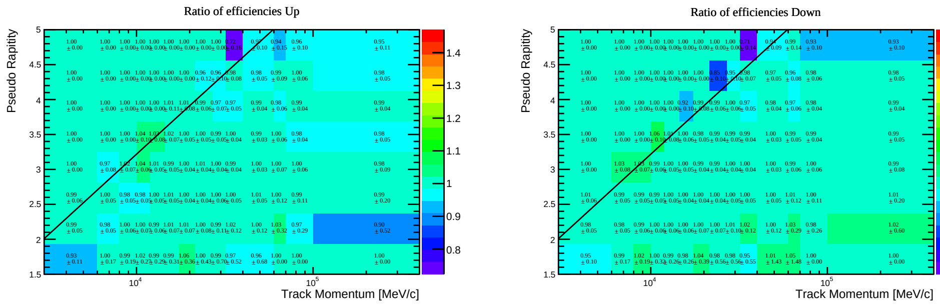


Figure 83: IsMuon efficiency ratio between data and simulation as a function of kinematics.

The remaining PID efficiency as a function of q^2 is very mild, as shown in Figs. 84 and 85. Any systematic effect would be negligible.

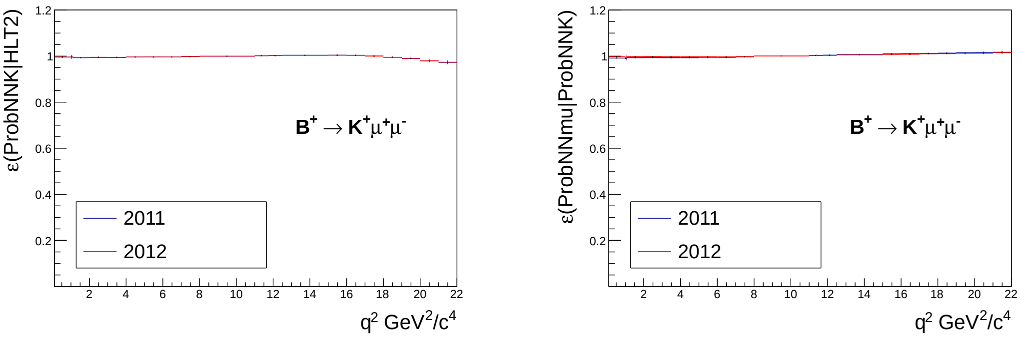


Figure 84: Relative efficiency for the muon and kaon PID selection between $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to J/\psi\, K^+$ decays as a function of q^2.

9.5.3 IP resolution

Although the daughter IP resolution is quite correlated to q^2, from Sect. A in the appendix, it can be seen that the IP variables are well described and so no systematic is assigned.

9.5.4 Mass model

Given the mass model parameterisation fits the J/ψ modes OK with a factor 200 higher statistics. At the J/ψ mass the mass model is assumed to be perfect compared to the

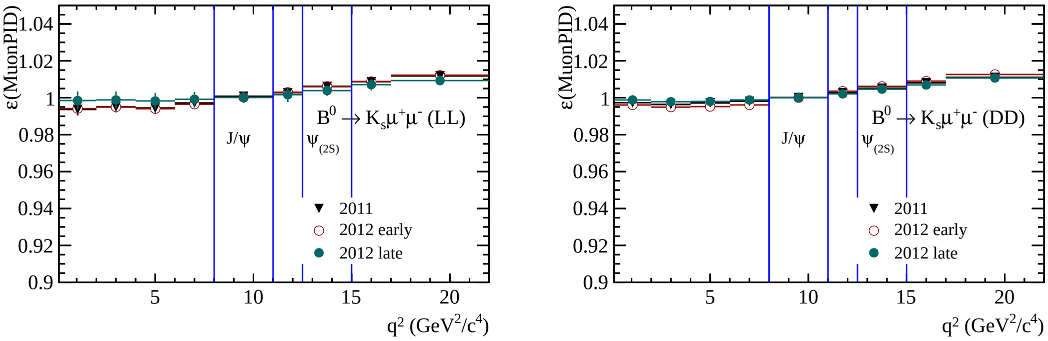


Figure 85: Relative efficiency for the muon and kaon PID selection between $B^0 \to K^0_S\mu^+\mu^-$ and $B^0 \to J/\psi\, K^0_S$ decays as a function of q^2.

statistical precision. The variation of the mass model as a function of q^2 makes very little difference to the signal yields (see Fig. 19) and so the effect of mis-modelling this correction is assumed to be negligible. The effect on $B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ will be checked when the yields as a function of q^2 are unblinded, but it is also expected to be negligible.

10 Cross checks

10.1 $B \to K\mu^+\mu^-$ cross checks

10.1.1 Old analysis vs new analysis compatibility

The old result is compared to the 2011 part of the new result to test the compatibility. To do this, the overlap of events in the signal region is calculated for each q^2 bin, shown in Fig 86. Although the q^2 binning has changed, the nearest bin is used to calculate the compatibility. The results are overlaid on Fig. 87, where the χ^2 probability, taking into account the overlap of events, is shown. The compatibility between the two is very good, with a χ^2 probability of 92%.

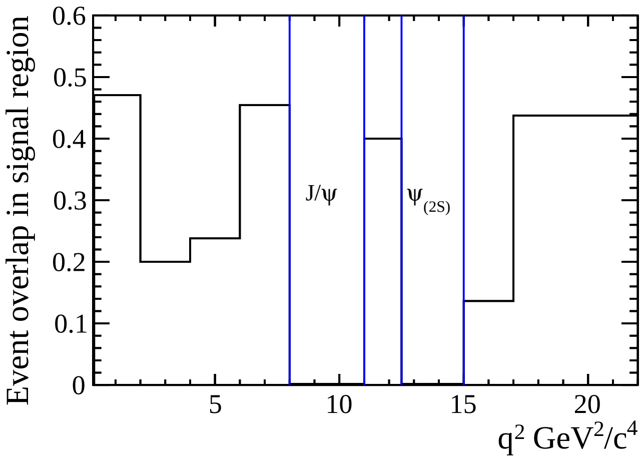


Figure 86: Overlap of events in the signal region between the old 2011 analysis and the 2011 part of the new analysis.

10.1.2 LL and DD compatibility

although this check was performed before unblinding, it is still useful and is listed here

Although the result are still blind, one can check the compatibility between the LL and DD categories already. Such a test is shown in Fig. 88, where the isospin asymmetry results are shown for the DD and LL categories separately. The DD central values have been moved to zero and the LL central values have been offset by the same amount as the DD mode, so that the LL points show the difference between the LL and DD results. The χ^2 agreement between the two categories is good, with a p-value of 15%. It is also important to note that the LL category is slightly lower than the DD category, which is the opposite situation to last time.

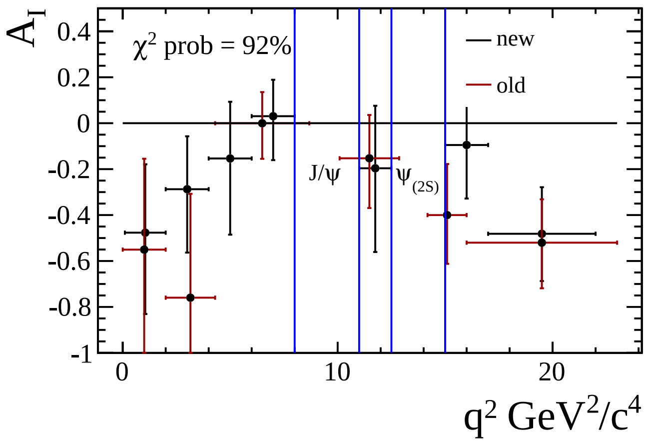


Figure 87: Isospin asymmetry of the old 2011 analysis and the 2011 part of the new analysis. The compatibility between the two is very good (92%).

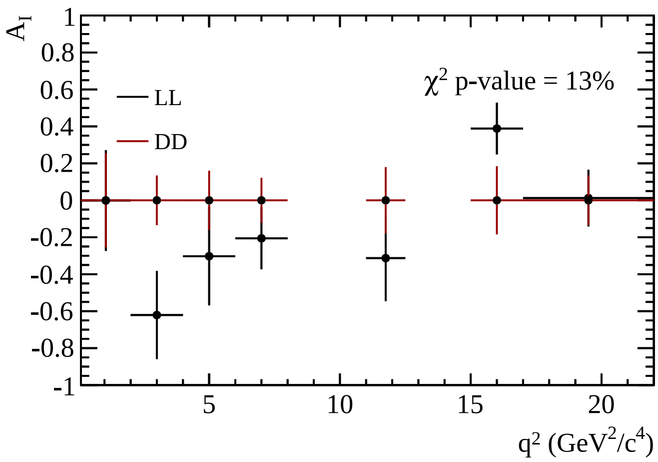


Figure 88: Blinded compatibility test between the LL and DD categories. The DD central values have been moved to zero and the LL central values have been offset by the same amount as the DD mode, so that the LL points show the difference between the LL and DD results.

10.1.3 2011 and 2012 compatibility

The isospin asymmetry results for the 2011 and 2012 datasets are shown in Fig. 89. For each bin, the 2012 dataset yields more positive central values than the 2011 dataset. The chance of this happening can be estimated by the binomial theorem (throwing 7 heads or 7 tails in a row), which comes to 1.5%, roughly 2.4σ. This is more discrepant than the 1.6σ fluctuation obtained in Sect. 6 by comparing the $B^0 \to K^0_S \mu^+ \mu^-$ branching fractions integrated over q^2. Comparisons between some kinematic variables in 2011 and 2012 data for $B^0 \to J/\psi K^0_S$ can be found in

`~powen/public/forAI/`.

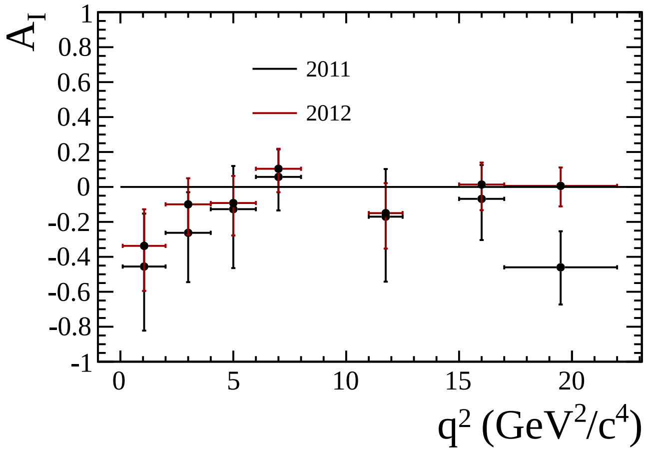


Figure 89: Isospin asymmetry for the 2011 and 2012 datasets. The 2012 results are consistently above the 2011 ones.

The data is also split into the K^0_S reconstruction categories and then compared between 2011 and 2012 for the categories separately. This is shown in Fig. 90, where there is no such evidence for a systematic shift. For the LL category there are five bins in a row where the 2012 result is higher but that is not significant compared to the seven in the combination. The DD category looks perfectly compatible between the 2011 and 2012 datasets. Given that neither category shows a clear trend, this is evidence that the systematic shift of all seven bins for the combination is indeed a statistical fluctuation as the two categories are essentially independent measurements.

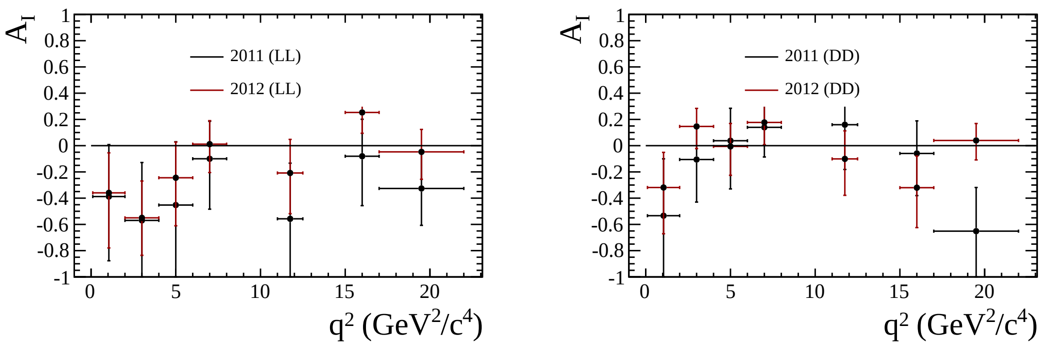


Figure 90: Isospin asymmetry for the 2011 and 2012 datasets for the LL (left) and DD (right) K^0_s reconstruction categories.

10.1.4 Stability with a mass range re-definition

The stability of the isospin result with different mass ranges is shown in Fig. 91. There is no significant change apart from if the lower mass sideband is removed for the 2-4 GeV^2/c^4 q^2 bin. The mass fit for this bin is shown here (all q^2 bins are in Sect. B in the appendix), where the background level sits slightly below the PDF. If one removes the region the background level increases and the corresponding $B^0 \to K^0_S\mu^+\mu^-$ signal yield decreases, which makes A_I more negative.

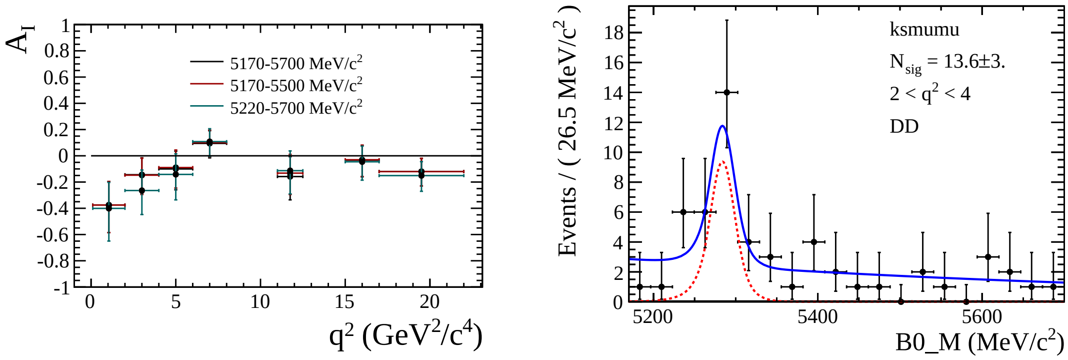


Figure 91: Isospin asymmetry under different mass ranges (5170-5700 MeV/c^2 is nominal). There is no significant change apart from if the lower mass sideband is removed for the 2-4 GeV/c^2 bin (corresponding mass fit shown on the right).

10.1.5 Apply trigger requirements on dimuon candidate

The isospin asymmetry results are compared when applying the trigger requirements on the dimuon candidate instead of the B candidate, shown in Fig. 92. This ensures that the trigger efficiency between $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^0_S\mu^+\mu^-$ cancel as they both trigger on the muons. There is only a visible effect at low q^2 as expected as that is where the K^0_S participates most. It is however a small effect and no systematic trend is observed.

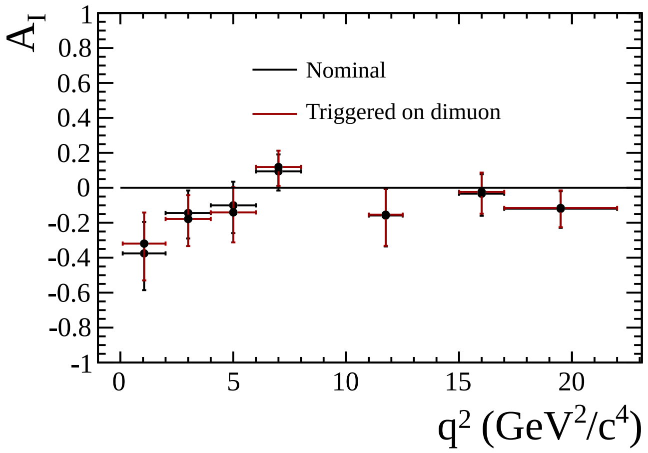


Figure 92: Isospin asymmetry results when applying the trigger requirements on the dimuon candidate.

10.1.6 Calculate $B^0 \to \psi(2S)\ K^0_{\rm S}/B \to J/\psi\ \ K^0_{\rm S}$ branching fraction

The ratio of branching fractions between $B^0 \to \psi(2S)\ K^0_{\rm S}$ and $B^0 \!\to J/\psi\, K^0_{\rm S}$ decays can be measured in the different run periods and $K^0_{\rm S}$ reconstruction categories. The relative efficiency between the two is taken from the plots in Sect. 7. The mass fits of which are shown in Fig. 93, with tabulated results in Table. 19. All four branching fractions agree with each other, the 2011 LL is slightly high, but if this were a systematic effect, it would be in the opposite direction to the non-resonant case, where the LL 2011 is low.

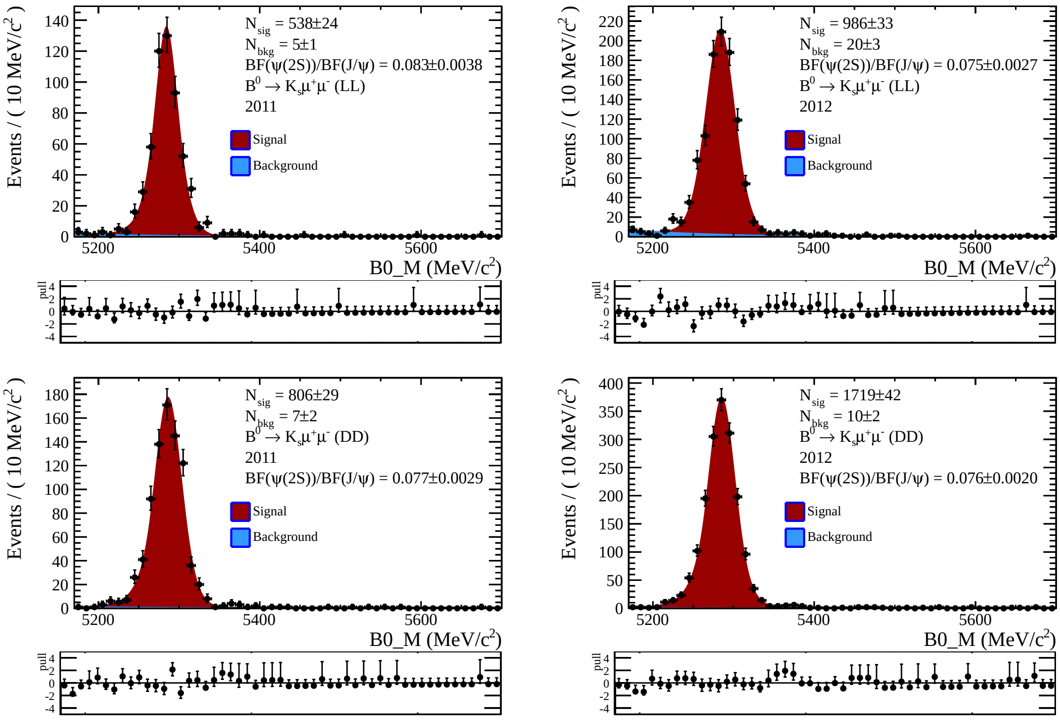


Figure 93: Mass fits to $B^0 \to \psi(2S)\ K^0_{\rm S}$ decays for the 2011, 2012 and $K^0_{\rm S}$ reconstruction categories.

Category	$\frac{\mathcal{B}B^0\to\psi(2S)K^0_{\rm S}}{\mathcal{B}B^0\to J/\psi\, K^0_{\rm S}}$
2011 LL	0.083±0.004
2012 LL	0.075±0.003
2011 DD	0.077±0.003
2012 DD	0.076±0.002

Table 19: Ratio of branching fractions, $\frac{\mathcal{B}B^0\to\psi(2S)K^0_{\rm S}}{\mathcal{B}B^0\to J/\psi\, K^0_{\rm S}}$, where the $J/\psi\ \to \mu^+\mu^-$ and $\psi(2S)$ $\to \mu^+\mu^-$ branching fractions are not taken into account.

10.2 $B \to K^*\mu^+\mu^-$ cross checks

10.2.1 LL vs DD compatibility

The results split between the LL and DD categories is shown in Fig. 94. There is no evidence for a systematic bias.

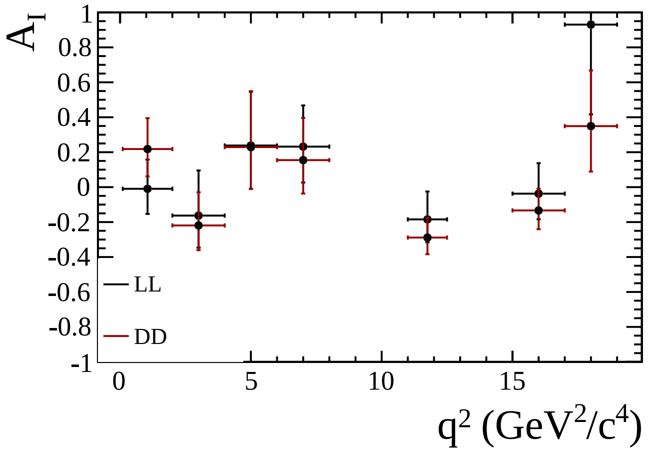


Figure 94: Compatibility test between the LL and DD categories for $B \to K^*\mu^+\mu^-$.

10.2.2 2011 vs 2012 compatibility

The results split between the 2011 and 2012 run periods is shown in Fig. 95. There is no evidence for a systematic bias.

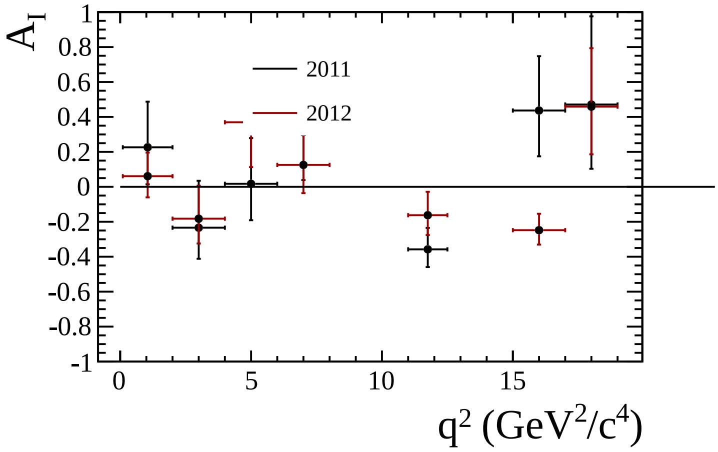


Figure 95: Compatibility test between the 2011 and 2012 run periods for $B \to K^*\mu^+\mu^-$.

Results are further split up in categories an run period in Fig. 96.

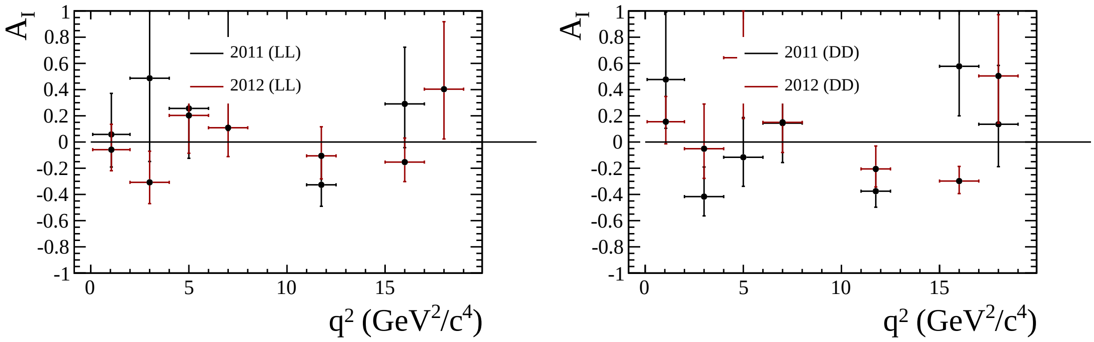


Figure 96: Isospin asymmetry for the 2011 and 2012 datasets for the LL (left) and DD (right) K^0_S reconstruction categories.

10.2.3 Error estimation

Given that the χ^2 value with zero of $B \rightarrow K^{*}\mu^+\mu^-$ A_I results is not very consistent (2%), possible underestimation of the uncertainty is checked by calculating the uncertainty on the signal yields by propagating the error from A_I. This is shown in Fig. 97, where the propagated uncertainties agree with what one would expect given that there is only a small amount of background (for mass fits see Sect. B in the appendix).

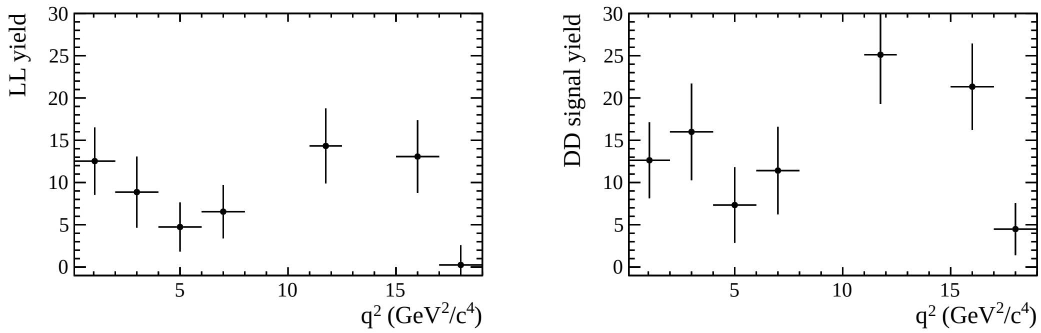


Figure 97: Signal yields of the LL and DD categories, where the uncertainties are propagated from the uncertainty on A_I.

11 Conclusions

11.1 Conclusions

The $B^+ \to K^+\mu^+\mu^-$ results are a substantial improvement over the existing measurements, and are split into which narrower q^2 bins which will help the theory community understand systematic effects due to c-$\bar{c}$ interference. In general the data is systematically below the SM prediction. Further theoretical studies are needed to ascertain whether this is due to new physics or a systematic effect relating to the theoretical predictions.

The isospin asymmetry is now consistent with zero.

References

[1] C. Bouchard *et al.*, *The rare decay $B \to K\ell^+\ell^-$ form factors from lattice QCD*, arXiv:1306.2384.

[2] J. Lyon and R. Zwicky, *Isospin asymmetries in $B \to (K^*, \rho)\gamma/l^+l^-$ and $B \to Kl^+l^-$ in and beyond the Standard Model*, Phys. Rev. **D88** (2013) 094004, arXiv:1305.4797.

[3] BaBar collaboration, B. Aubert *et al.*, *Direct CP, lepton flavor and isospin asymmetries in the decays $B \to K^{(*)}l^+l^-$*, Phys. Rev. Lett. **102** (2008) 091803, arXiv:0807.4119.

[4] Belle collaboration, J.-T. Wei *et al.*, *Measurement of the differential branching fraction and forward-backward asymmetry for $B \to K^{(*)}l^+l^-$*, Phys. Rev. Lett. **103** (2009) 171801, arXiv:0904.0770.

[5] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of early 2012*, arXiv:1207.1158, updated results and plots available at: http://www.slac.stanford.edu/xorg/hfag/.

[6] LHCb collaboration, R. Aaij *et al.*, *Measurement of the isospin asymmetry in $B \to K^{(*)}\mu^+\mu^-$ decays*, JHEP **07** (2012) 133, arXiv:1205.3422.

[7] U. Egede, P.Owen, *Measurement of the isospin asymmetry in $B \to K^{(*)}\mu^+\mu^-$ decays*, LHCb-ANA-2011-081, 2012.

[8] T. Blake, U. Egede, P. Owen and K. Petridis, *Observation of a resonance in $B^+ \to K^+\mu^+\mu^-$ decays low recoil*, LHCb-ANA-2013-039, 2013.

[9] C. Parkinson, *The angular analysis of the $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ decay at LHCb*, PhD thesis, Imperial College London, 2013.

[10] M. Vesterinen, *K^0_S π^+ π^- samples for Hlt studies*, HLT2 Piquet Report, May, 2012.

[11] D. M. Aurélien Martens, *K^0_S from B's in the 2012 HLT2Topo*, Physics Performance, Trigger and Stripping meeting, March, 2013.

[12] A. Ali, P. Ball, L. Handoko, and G. Hiller, *A Comparative study of the decays $B \to (K, K^*)\ell^+\ell^-$ in standard model and supersymmetric theories*, Phys. Rev. **D61** (2000) 074024, arXiv:hep-ph/9910221.

[13] T. Blake, U. Egede, P. Owen and K. Petridis, *Angular analysis of $B^+ \to K^+\mu^+\mu^-$ decays*, LHCb-ANA-2013-030, 2013.

[14] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neural networks for particle identification*, Nucl. Instrum. Meth. **A543** (2005) 577, arXiv:physics/0408124.

[15] R. E. Schapire and Y. Freund, *A decision-theoretic generalization of on-line learning and an application to boosting*, Jour. Comp. and Syst. Sc. **55** (1997) 119.

[16] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the $B^+ \to K^+\mu^+\mu^-$ decay*, JHEP **02** (2013) 105, `arXiv:1209.4284`.

[17] M.-O. Bettler, H. Cliff, V. Gibson, S. Wright, *Measurement of ACP in $B^+ \to K^+\mu^+\mu^-$ decays*, LHCb-ANA-2013-016, 2013.

[18] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[19] LHCb collaboration, R. Aaij *et al.*, *Measurement of relative branching fractions of B decays to $\psi(2S)$ and J/ψ mesons*, Eur. Phys. J. **C72** (2012) 2118, `arXiv:1205.0918`.

[20] R. Fleischer and T. Mannel, *General analysis of new physics in $B \to J/\psi K$*, Phys. Lett. **B506** (2001) 311, `arXiv:hep-ph/0101276`.

[21] BaBar Collaboration, B. Aubert *et al.*, *Measurement of branching fractions and charge asymmetries for exclusive B decays to charmonium*, Phys. Rev. Lett. **94** (2005) 141801, `arXiv:hep-ex/0412062`.

[22] Particle Data Group, J. Beringer *et al.*, *Review of particle physics*, Phys. Rev. **D86** (2012) 010001.

[23] LHCb collaboration, R. Aaij *et al.*, *Observation of a resonance in $B^+ \to K^+\mu^+\mu^-$ decays at low recoil*, `arXiv:1307.7595`, to appear in Phys. Rev. Lett.

[24] T. Blake and P. Owen, *Reconstruction efficiency of DD K_s^0 mesons using $D^0 \to \phi K_{\mathrm{S}}^0$ decays*, LHCb-INT-2013-051, 2013.

Appendix

A Data MC agreement

A.1 $B^+ \to K^+ \mu^+ \mu^-$

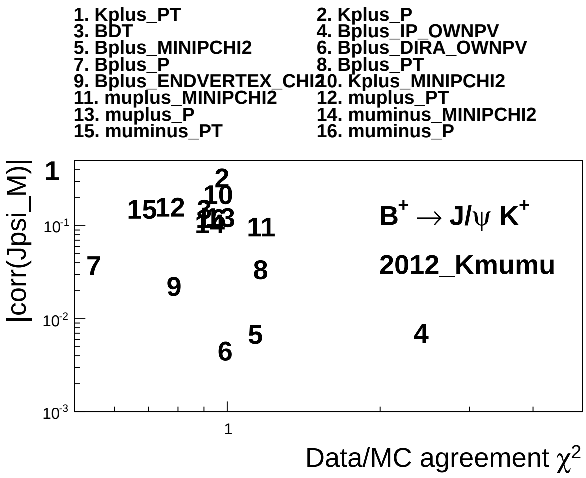


Figure 98: A summary of the data/simulation comparison for several variables in 2012. The x-axis shows the χ^2/ndf agreement between the distribution of the variable . The y-axis shows the correlation of the variable with the dimuon mass in $B^0 \to K^0_S \mu^+ \mu^-$ simulation. Variables in the top right of this plot, badly modelled and correlated to the observable, would require a systematic uncertainty.

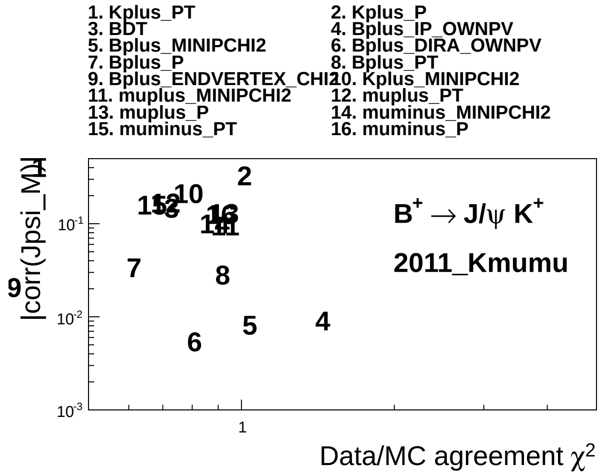


Figure 99: A summary of the data/simulation comparison for several variables in 2011. The x-axis shows the χ^2/ndf agreement between the distribution of the variable . The y-axis shows the correlation of the variable with the dimuon mass in $B^0 \to K^0_{\rm S}\mu^+\mu^-$ simulation. Variables in the top right of this plot, badly modelled and correlated to the observable, would require a systematic uncertainty.

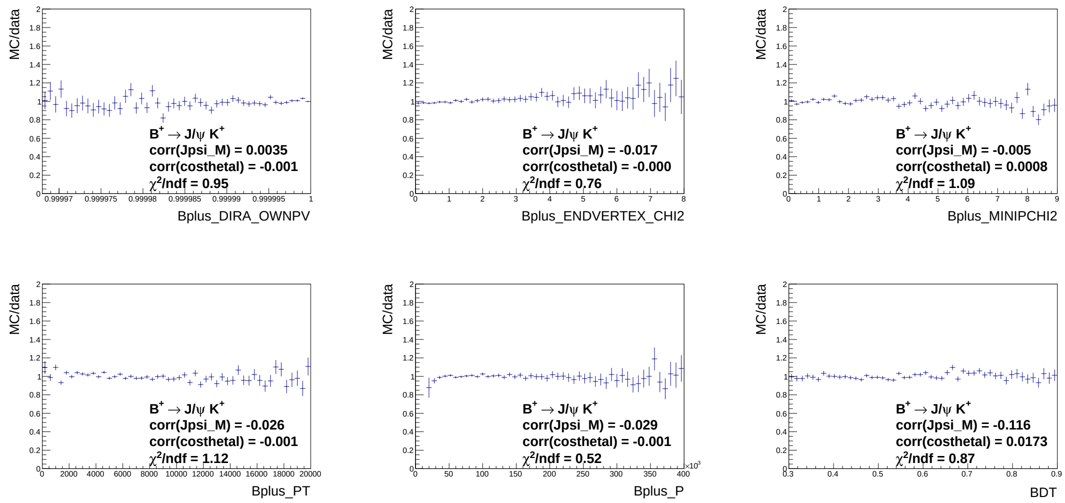


Figure 100: MC/data agreement for various B variables. For each variable the correlation to the dimuon mass in $B^+ \to K^+\mu^+\mu^-$ simulation is shown. The χ^2 of the data with a horizontal straight line at 1 is also shown.

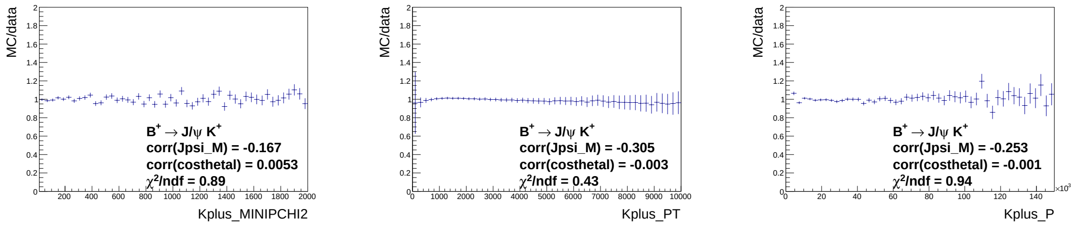


Figure 101: MC/data agreement for various kaon variables. For each variable the correlation to the dimuon mass in $B^+ \to K^+\mu^+\mu^-$ simulation is shown. The χ^2 of the data with a horizontal straight line at 1 is also shown.

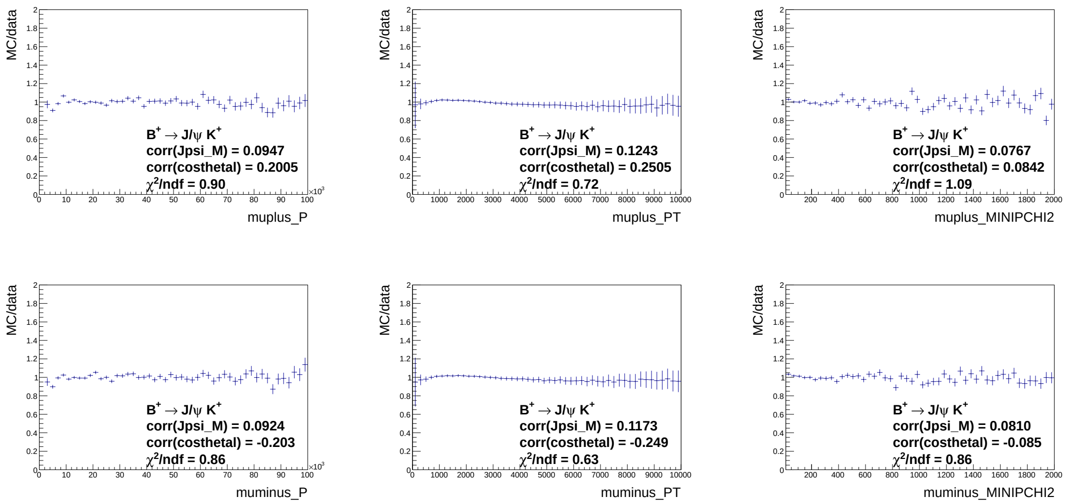


Figure 102: MC/data agreement for various muon variables. For each variable the correlation to the dimuon mass in $B^+ \to K^+\mu^+\mu^-$ simulation is shown. The χ^2 of the data with a horizontal straight line at 1 is also shown.

A.2 $B^0 \to K^0_S \mu^+ \mu^-$

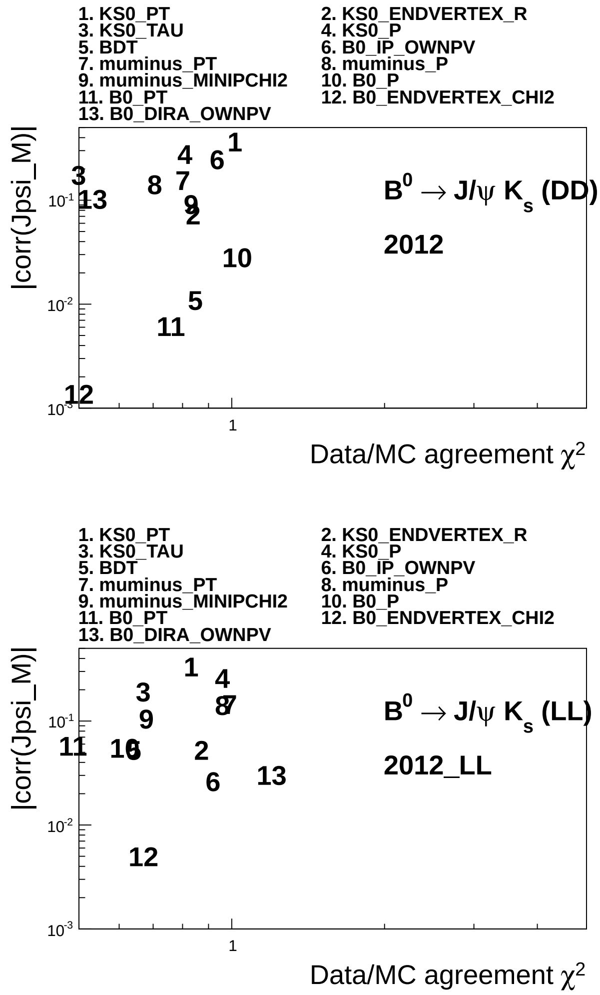
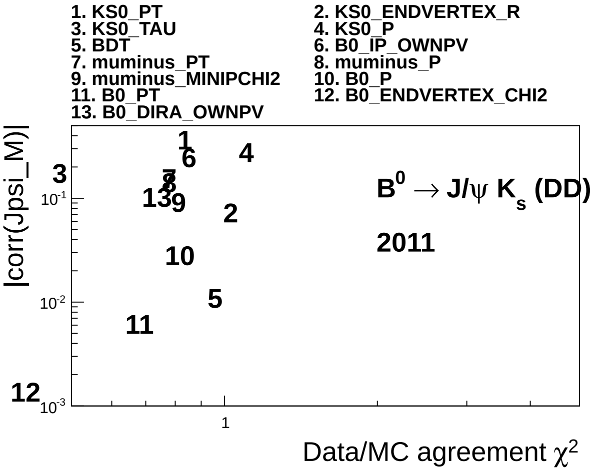


Figure 103: A summary of the data/simulation comparison for several variables in 2012. The x-axis shows the χ^2/ndf agreement between the MC/data distribution and a straight line at one. The y-axis shows the correlation of the variable with the dimuon mass in $B^0 \to K^0_S \mu^+ \mu^-$ simulation.



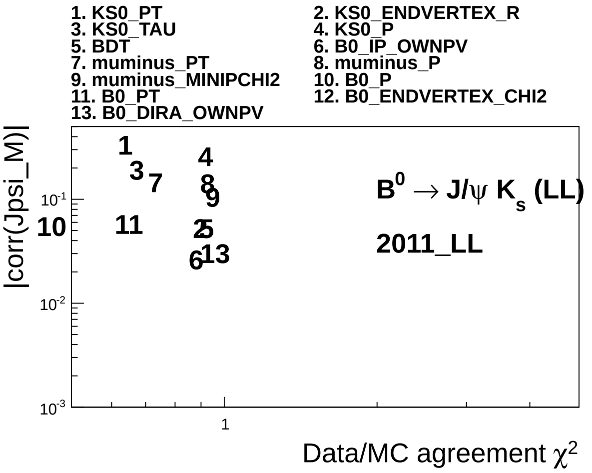


Figure 104: A summary of the data/simulation comparison for several variables in 2011. The x-axis shows the χ^2/ndf agreement between the MC/data distribution and a straight line at one. The y-axis shows the correlation of the variable with the dimuon mass in $B^0 \to K^0_s\mu^+\mu^-$ simulation.

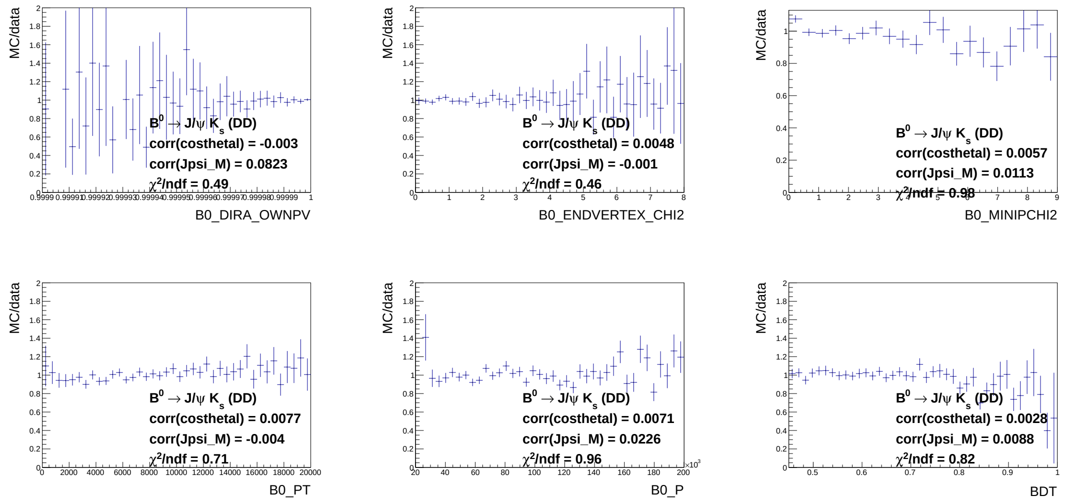


Figure 105: MC/data agreement for various B variables. For each variable the correlation to the dimuon mass in $B^0 \to K^0_S \mu^+\mu^-$ simulation is shown. The χ^2 of the data with a horizontal straight line at 1 is also shown.

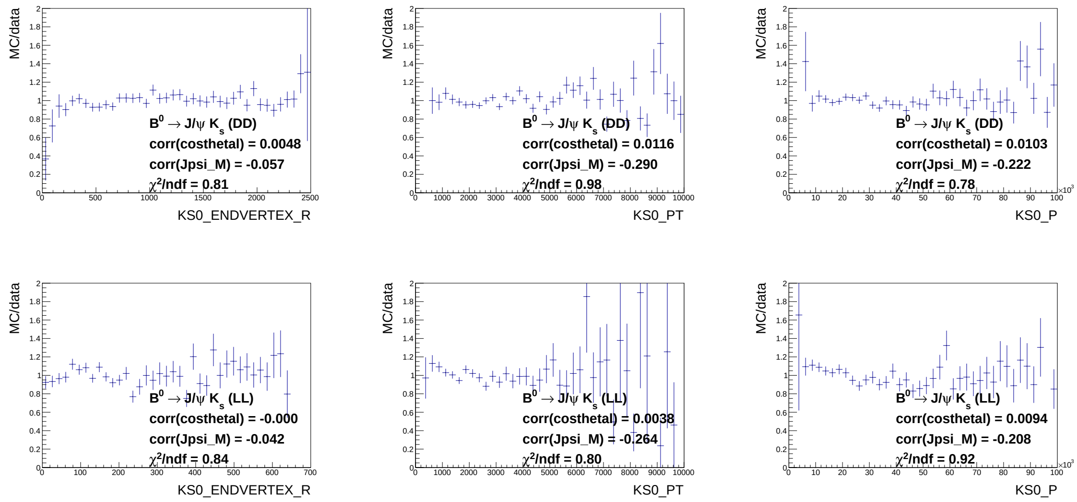


Figure 106: MC/data agreement for various K^0_S variables. For each variable the correlation to the dimuon mass in $B^0 \to K^0_S \mu^+\mu^-$ simulation is shown. The χ^2 of the data with a horizontal straight line at 1 is also shown.

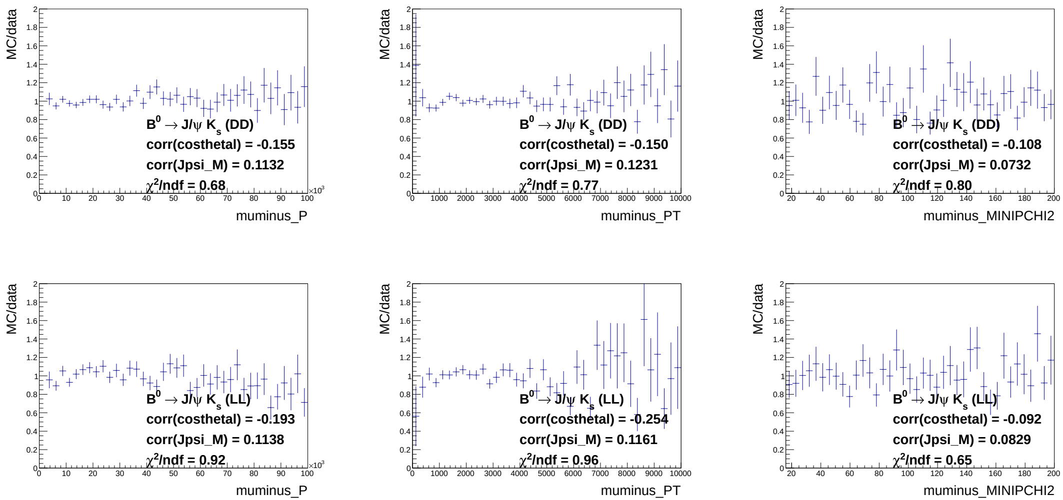


Figure 107: MC/data agreement for various muon variables. For each variable the correlation to the dimuon mass in $B^0 \to K^0_S\mu^+\mu^-$ simulation is shown. The χ^2 of the data with a horizontal straight line at 1 is also shown.

A.3 $B^+ \to (K^{*+} \to K^0_S \pi^+)\mu^+\mu^-$

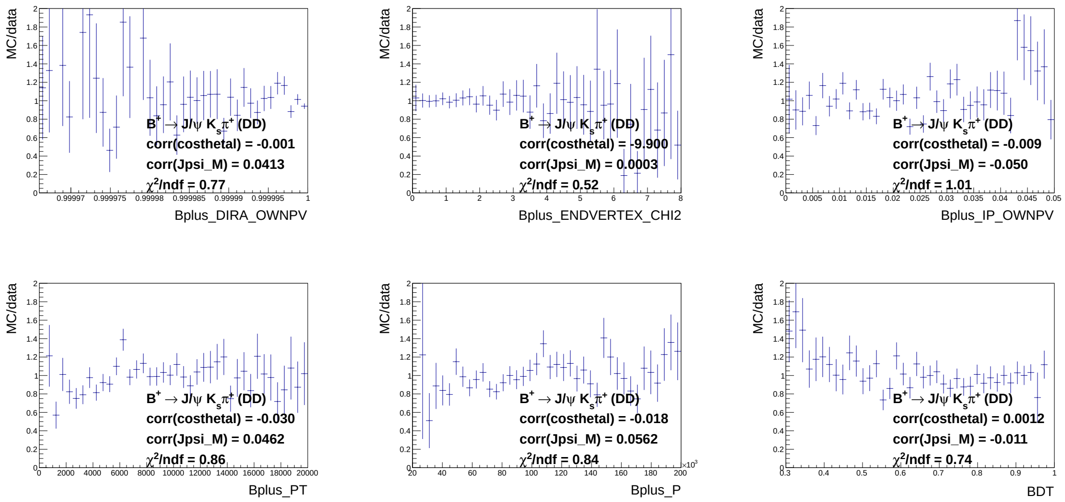


Figure 108: MC/data agreement for various B variables. For each variable the correlation to the dimuon mass in $B^0 \to K^0_S \mu^+\mu^-$ simulation is shown. The χ^2 of the data with a horizontal straight line at 1 is also shown.

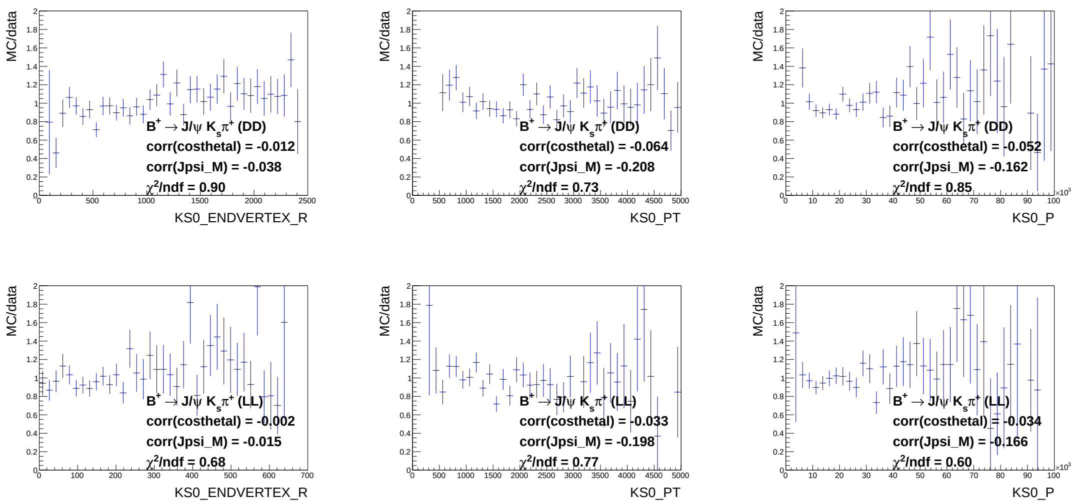


Figure 109: MC/data agreement for various K^0_S variables. For each variable the correlation to the dimuon mass in $B^0 \to K^0_S \mu^+\mu^-$ simulation is shown. The χ^2 of the data with a horizontal straight line at 1 is also shown.

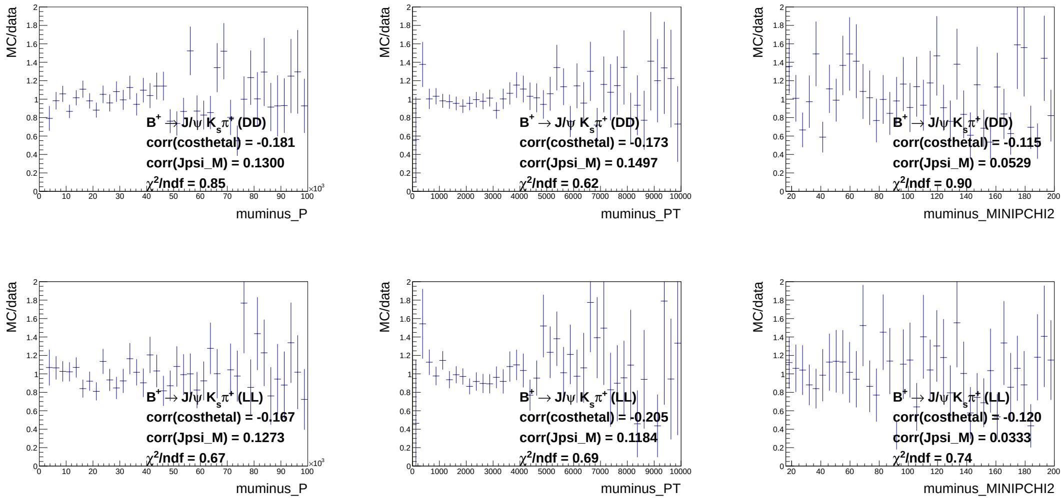


Figure 110: MC/data agreement for various muon variables. For each variable the correlation to the dimuon mass in $B^0 \to K^0_S\mu^+\mu^-$ simulation is shown. The χ^2 of the data with a horizontal straight line at 1 is also shown.

B Mass fits

B.1 $B^+ \to K^+\mu^+\mu^-$

Mass fits for each q^2 bin.

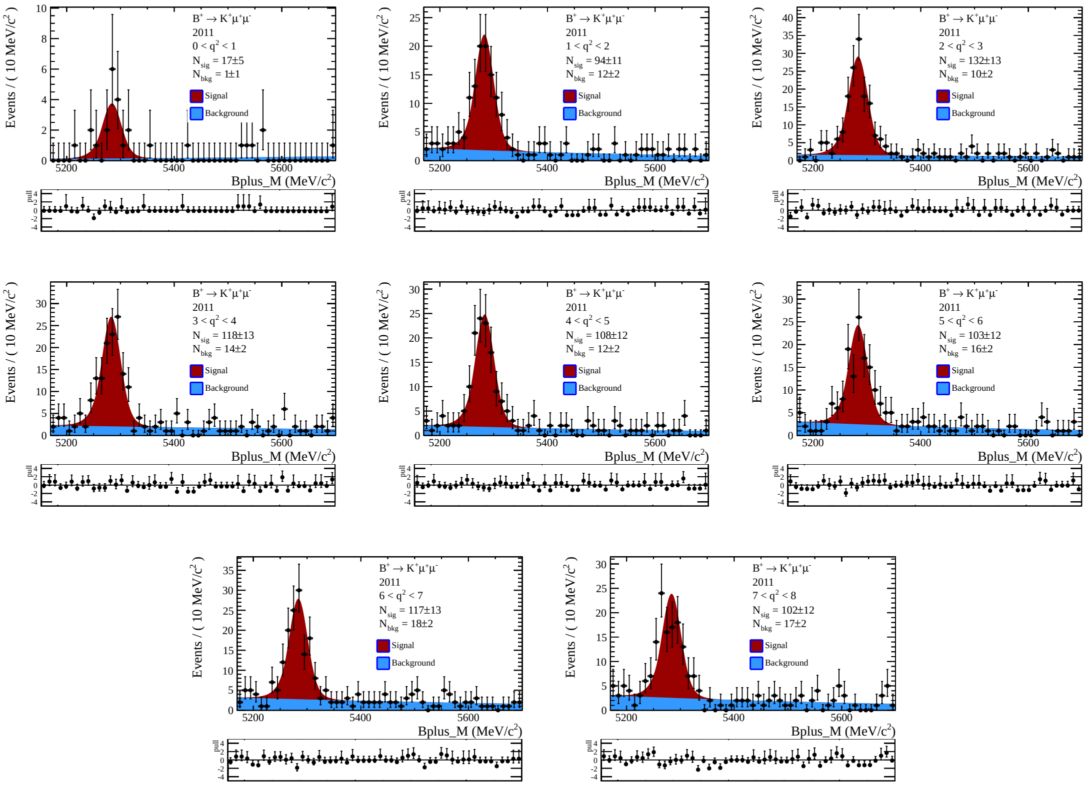


Figure 111: Mass fits to 2011 $B^+ \to K^+\mu^+\mu^-$ data in the q^2 bins below the J/ψ.

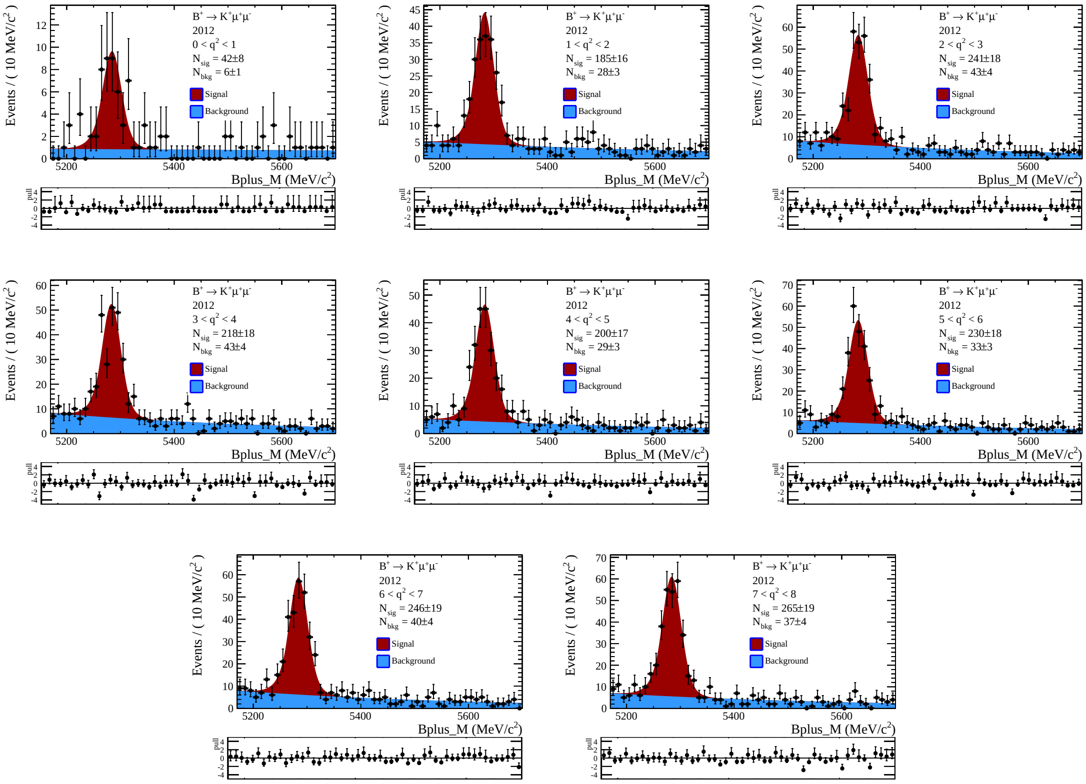


Figure 112: Mass fits to 2012 $B^+ \to K^+\mu^+\mu^-$ data in the q^2 bins below the J/ψ.

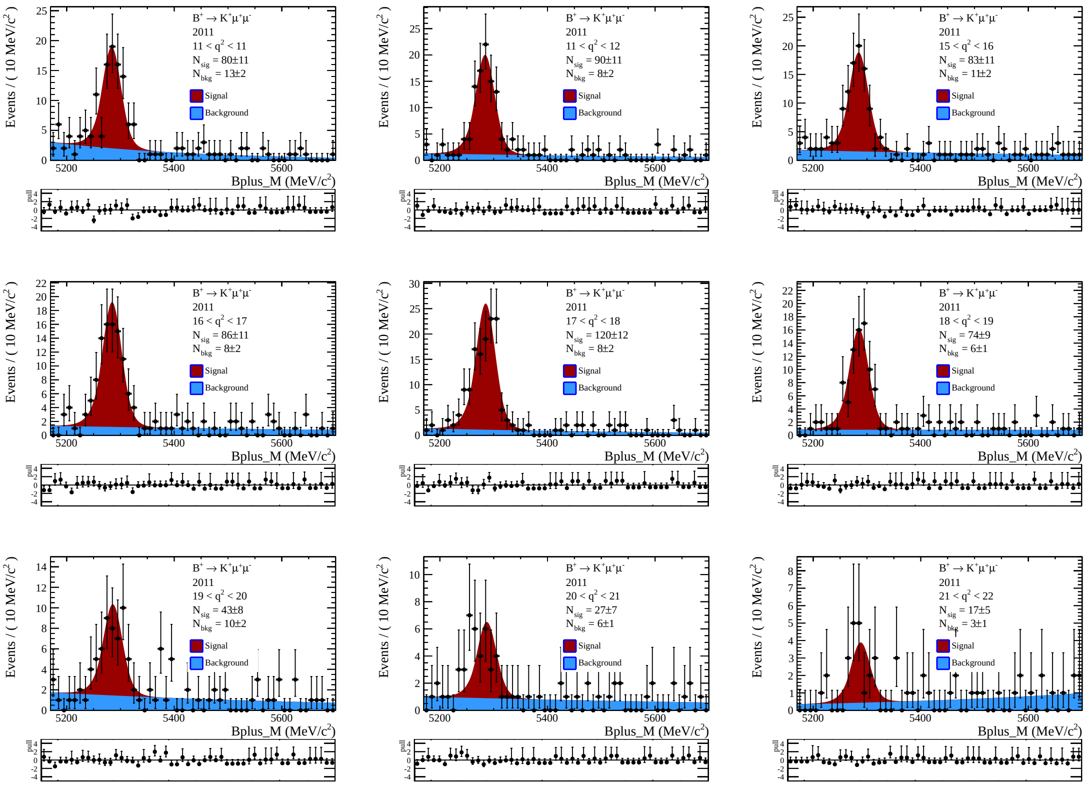


Figure 113: Mass fits to 2011 $B^+ \to K^+\mu^+\mu^-$ data in the q^2 bins above the J/ψ.

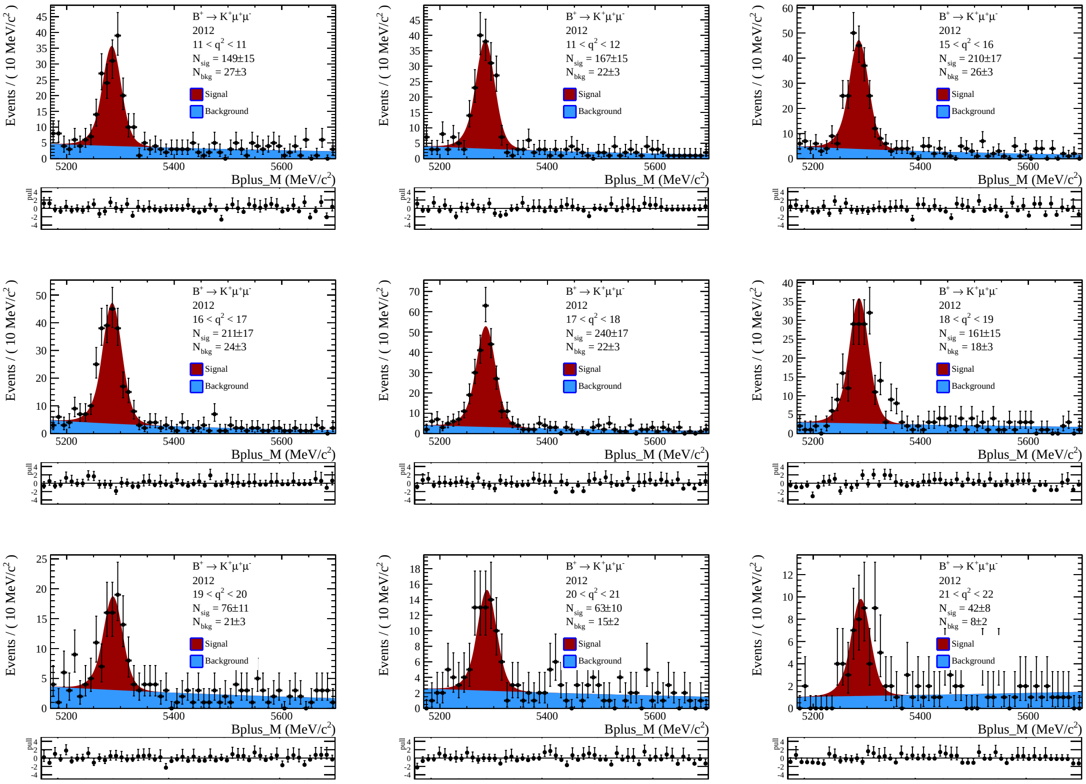


Figure 114: Mass fits to 2012 $B^+ \to K^+\mu^+\mu^-$ data in the q^2 bins above the J/ψ.

B.2 $B^0 \to K^0_S \mu^+ \mu^-$

B.3 $B^+ \to (K^{*+} \to K^0_S \pi^+) \mu^+ \mu^-$

B.4 $B^0 \to (K^{*0} \to K^+ \pi^-) \mu^+ \mu^-$

Mass fits for each q^2 bin.

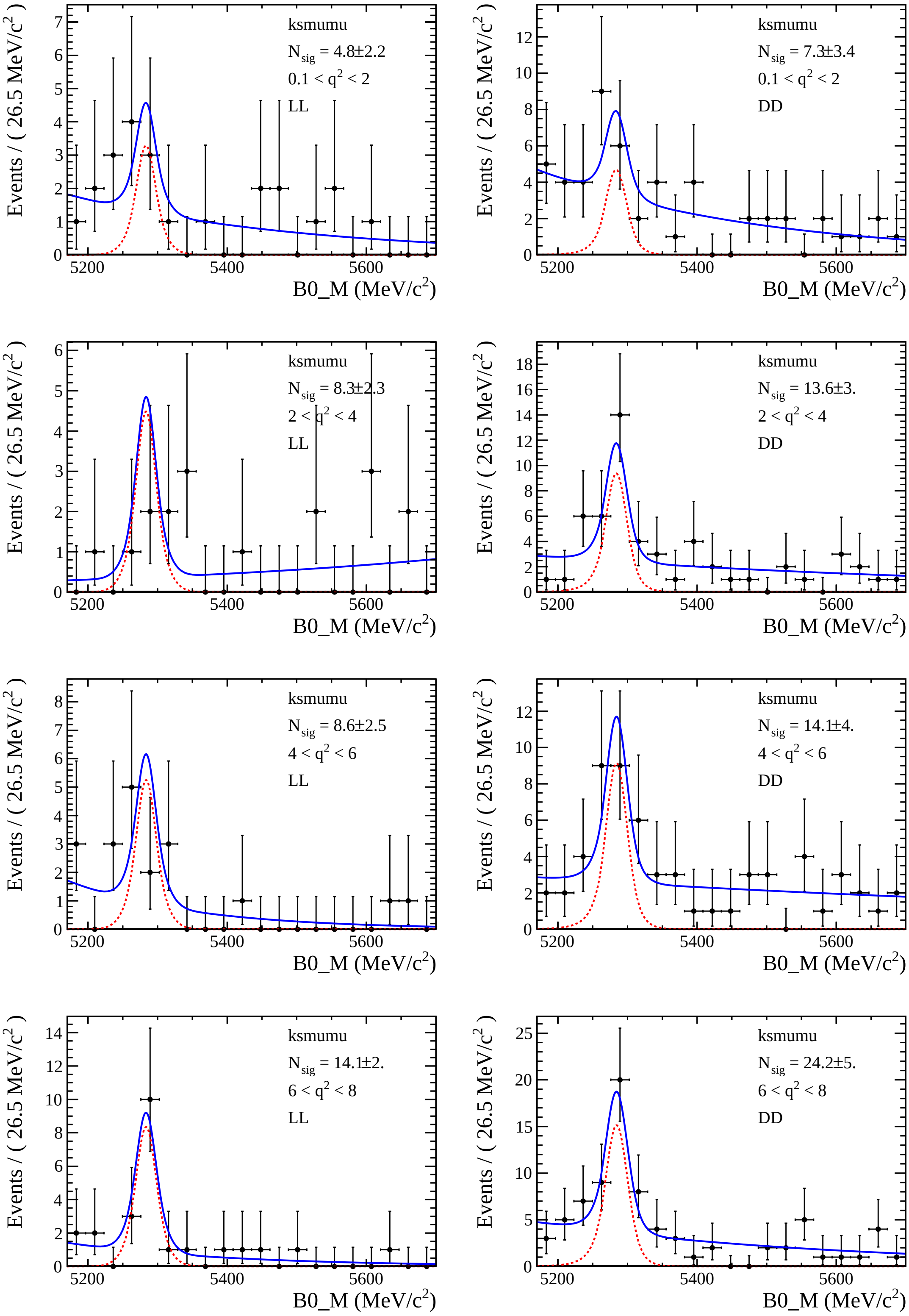


Figure 115: Mass fits to $B^0 \to K^0_S\mu^+\mu^-$ data (2011+2012) in the bins of q^2 below the J/ψ. Note that the LL and DD categories are fit simultaneously to ensure they have the same branching fraction.

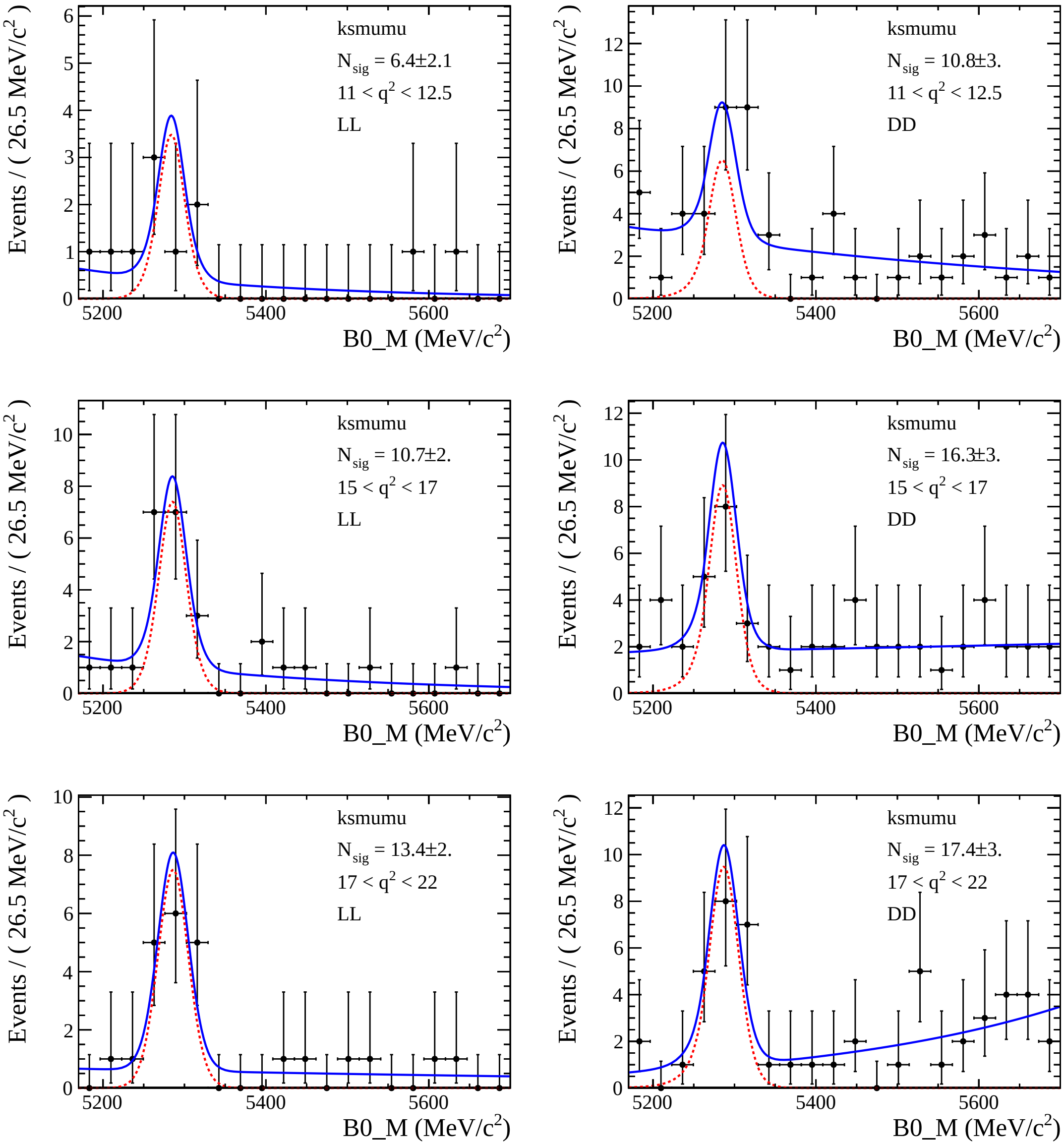


Figure 116: Mass fits to $B^0 \to K^0_S \mu^+\mu^-$ data (2011+2012) in the bins of q^2 above the J/ψ. Note that the LL and DD categories are fit simultaneously to ensure they have the same branching fraction.

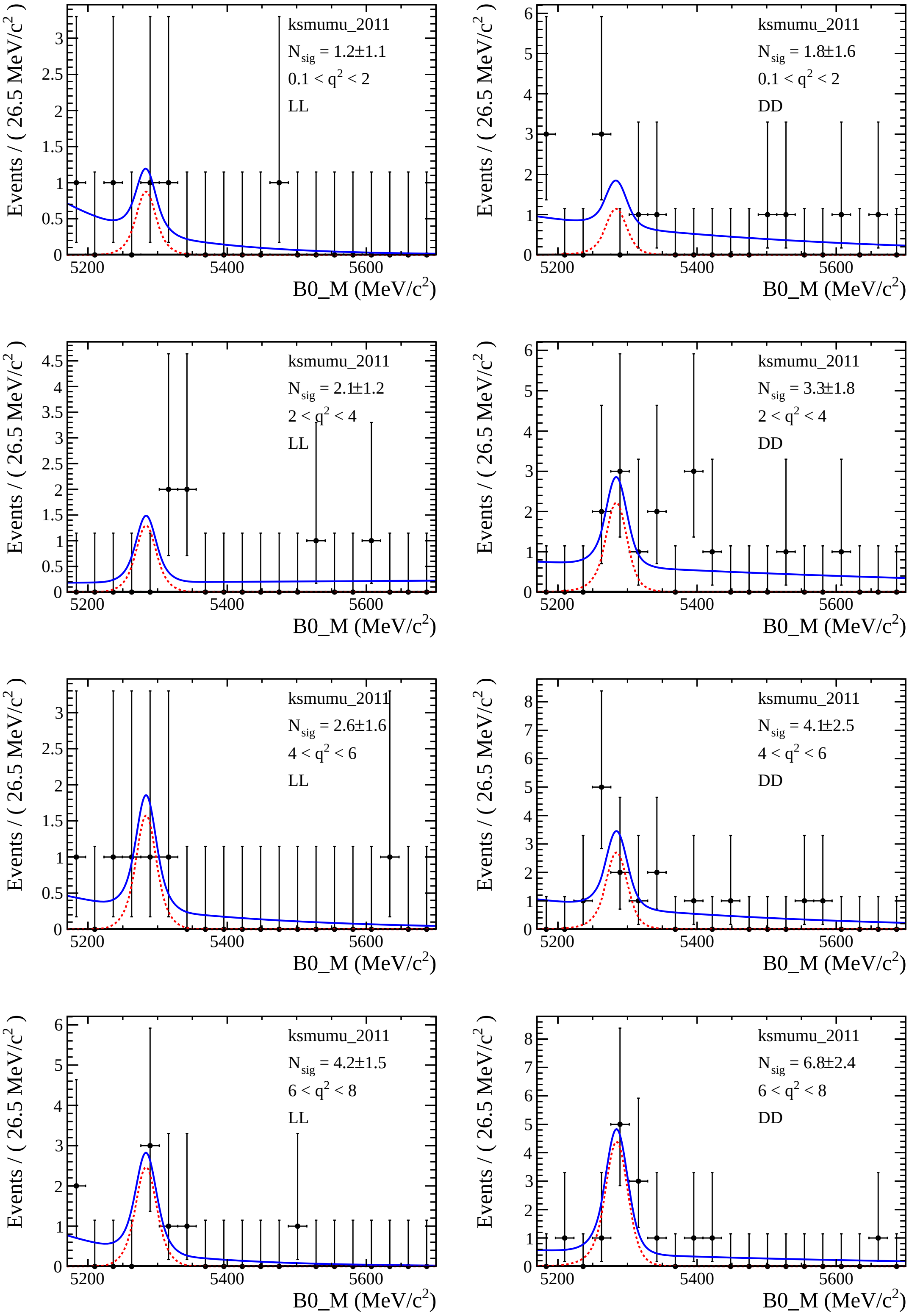


Figure 117: Mass fits to $B^0 \to K^0_S \mu^+ \mu^-$ 2011 data in the bins of q^2 below the J/ψ. Note that the LL and DD categories are fit simultaneously to ensure they have the same branching fraction.

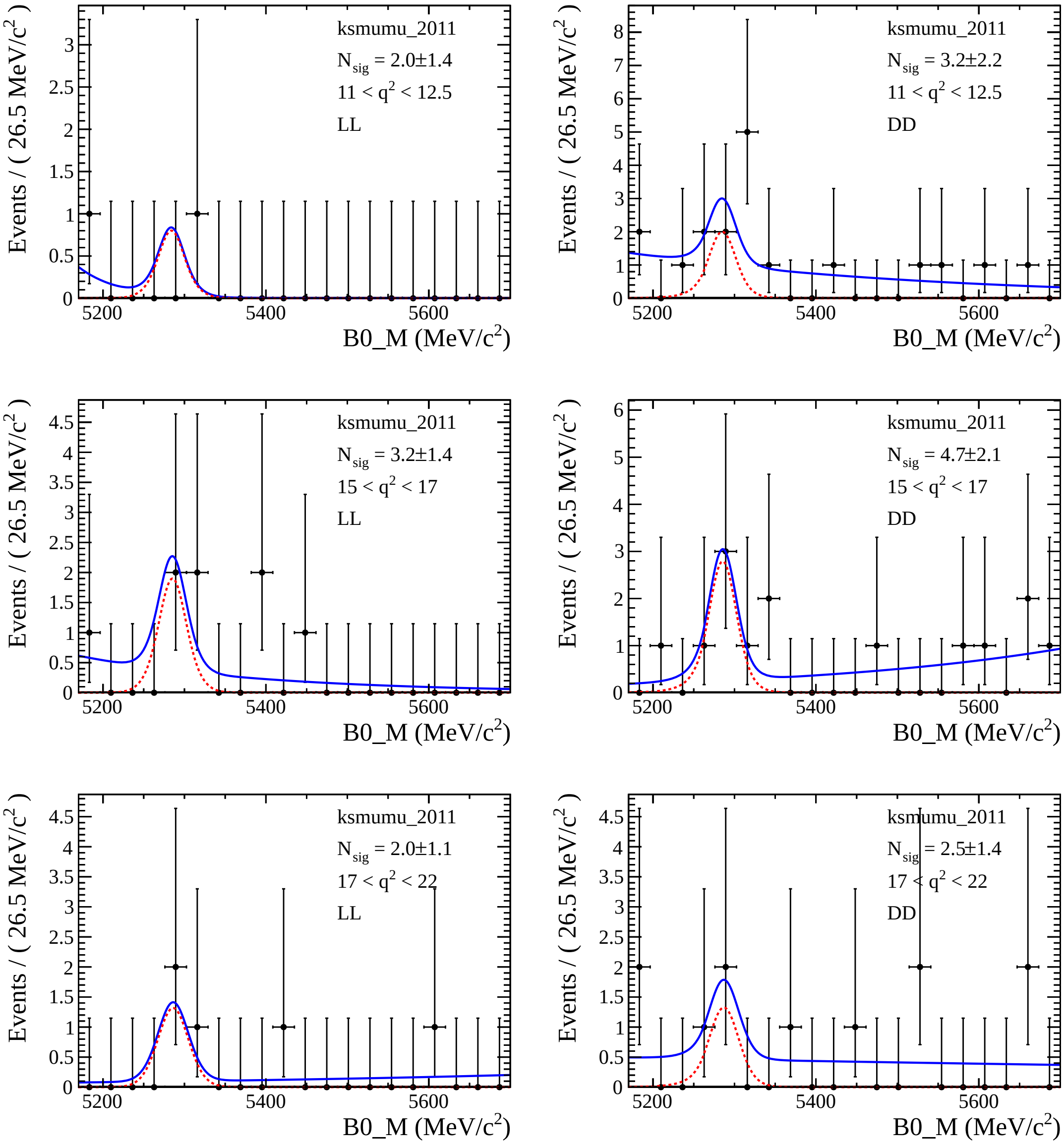


Figure 118: Mass fits to $B^0 \to K^0_S\mu^+\mu^-$ 2011 data in bins of q^2 above the J/ψ. Note that the LL and DD categories are fit simultaneously to ensure they have the same branching fraction.

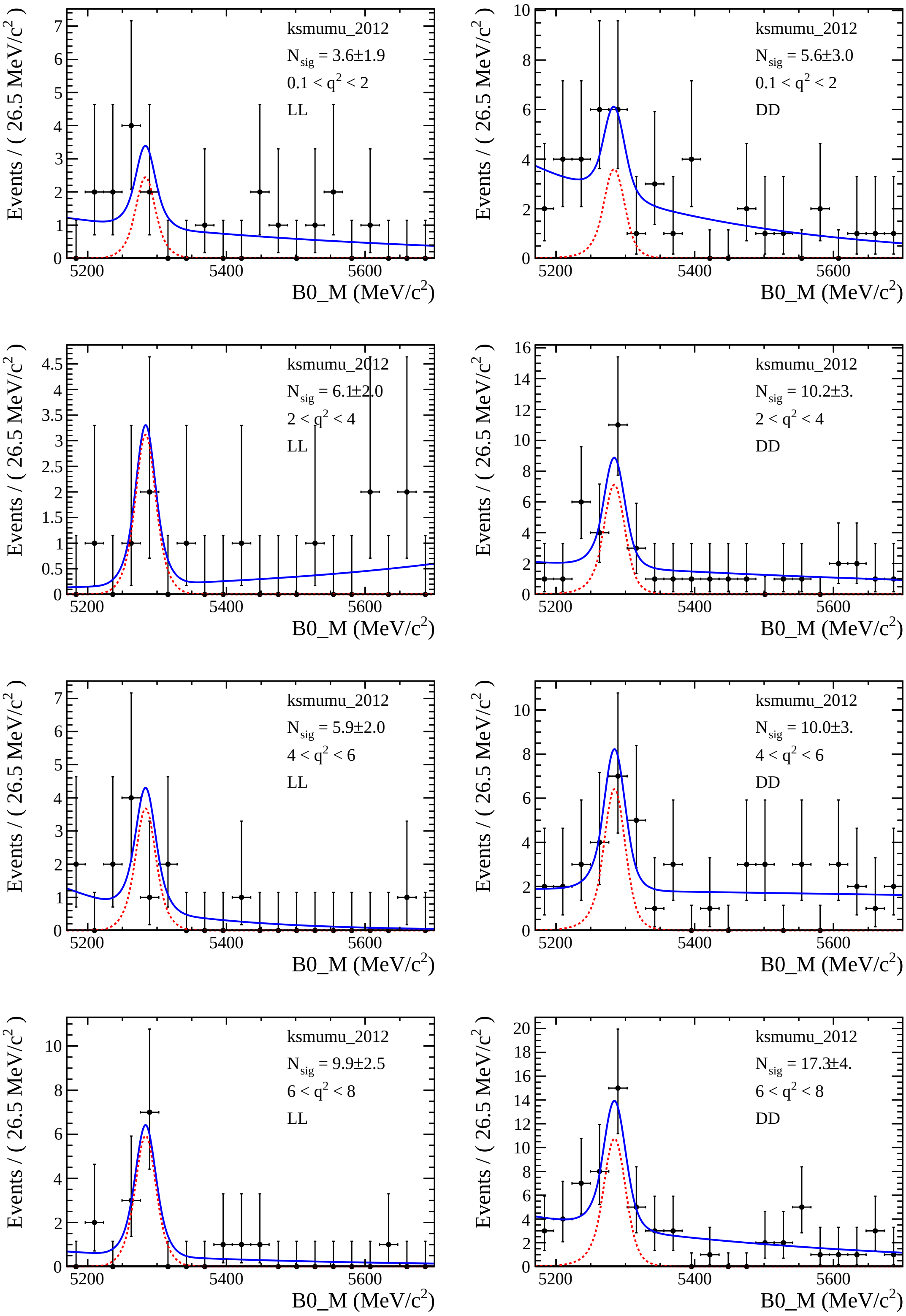


Figure 119: Mass fits to $B^0 \to K^0_{\rm S}\mu^+\mu^-$ 2012 data in bins of q^2 below the J/ψ. Note that the LL and DD categories are fit simultaneously to ensure they have the same branching fraction.

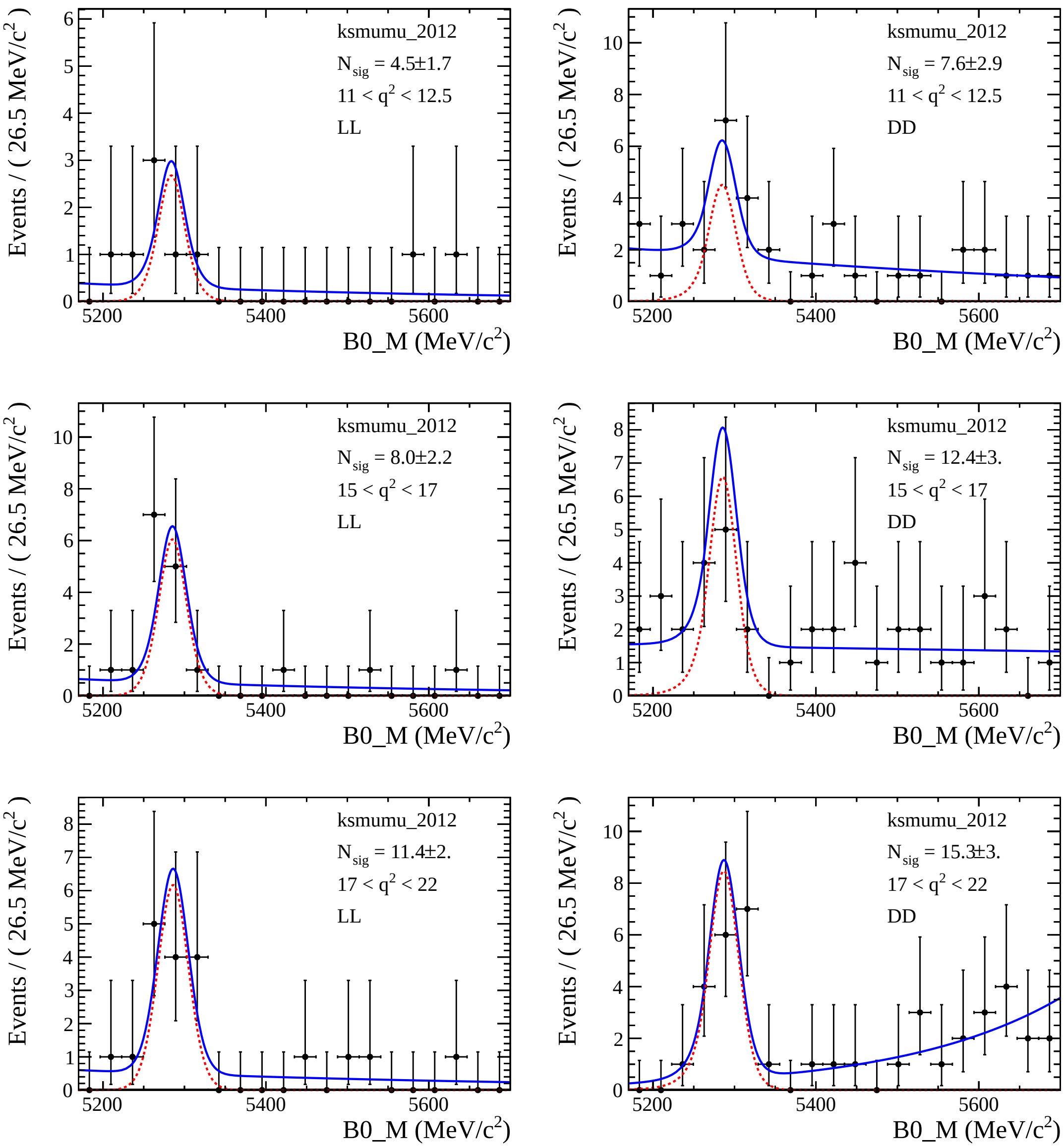


Figure 120: Mass fits to $B^0 \to K^0_\mathrm{S}\mu^+\mu^-$ 2012 data in bins of q^2 above the J/ψ. Note that the LL (left) and DD (right) categories are fit simultaneously to ensure they have the same branching fraction.

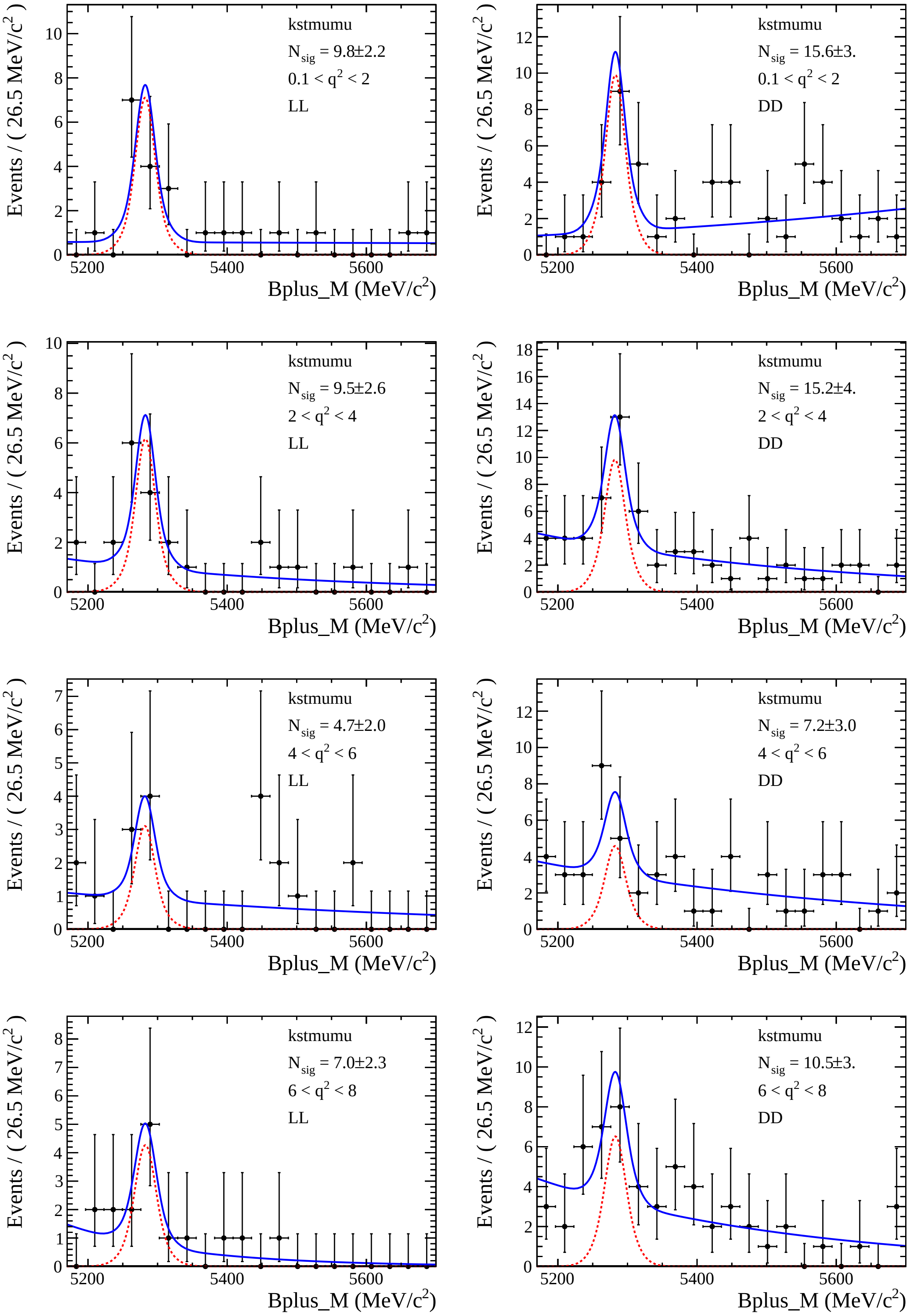


Figure 121: Mass fits to $B^+ \to (K^{*+} \to K^0_\mathrm{S}\pi^+)\mu^+\mu^-$ data (2011+2012) in the bins of q^2 below the J/ψ. Note that the LL and DD categories are fit simultaneously to ensure they have the same branching fraction.

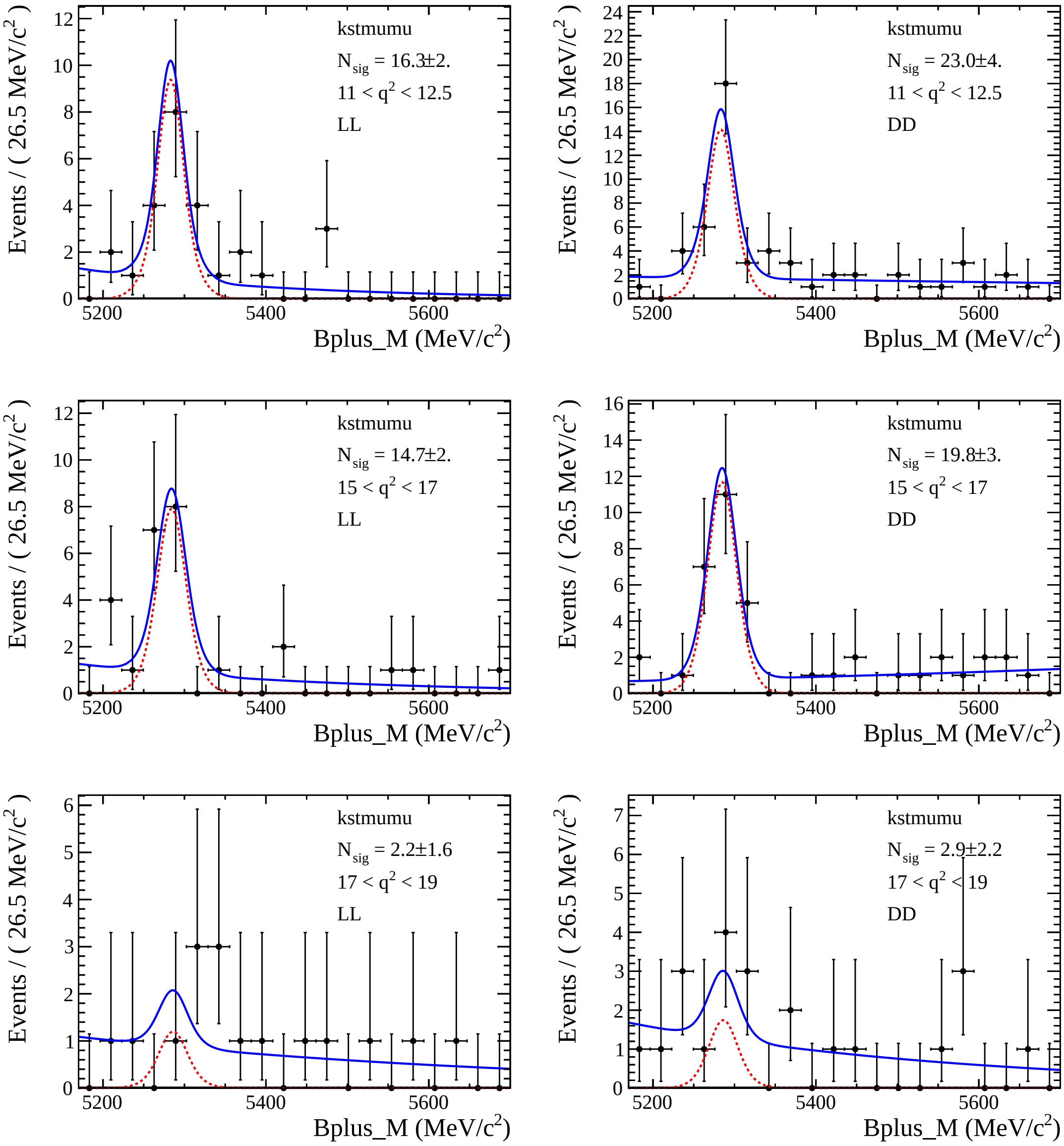


Figure 122: Mass fits to $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ data (2011+2012) in the bins of q^2 above the J/ψ. Note that the LL and DD categories are fit simultaneously to ensure they have the same branching fraction.

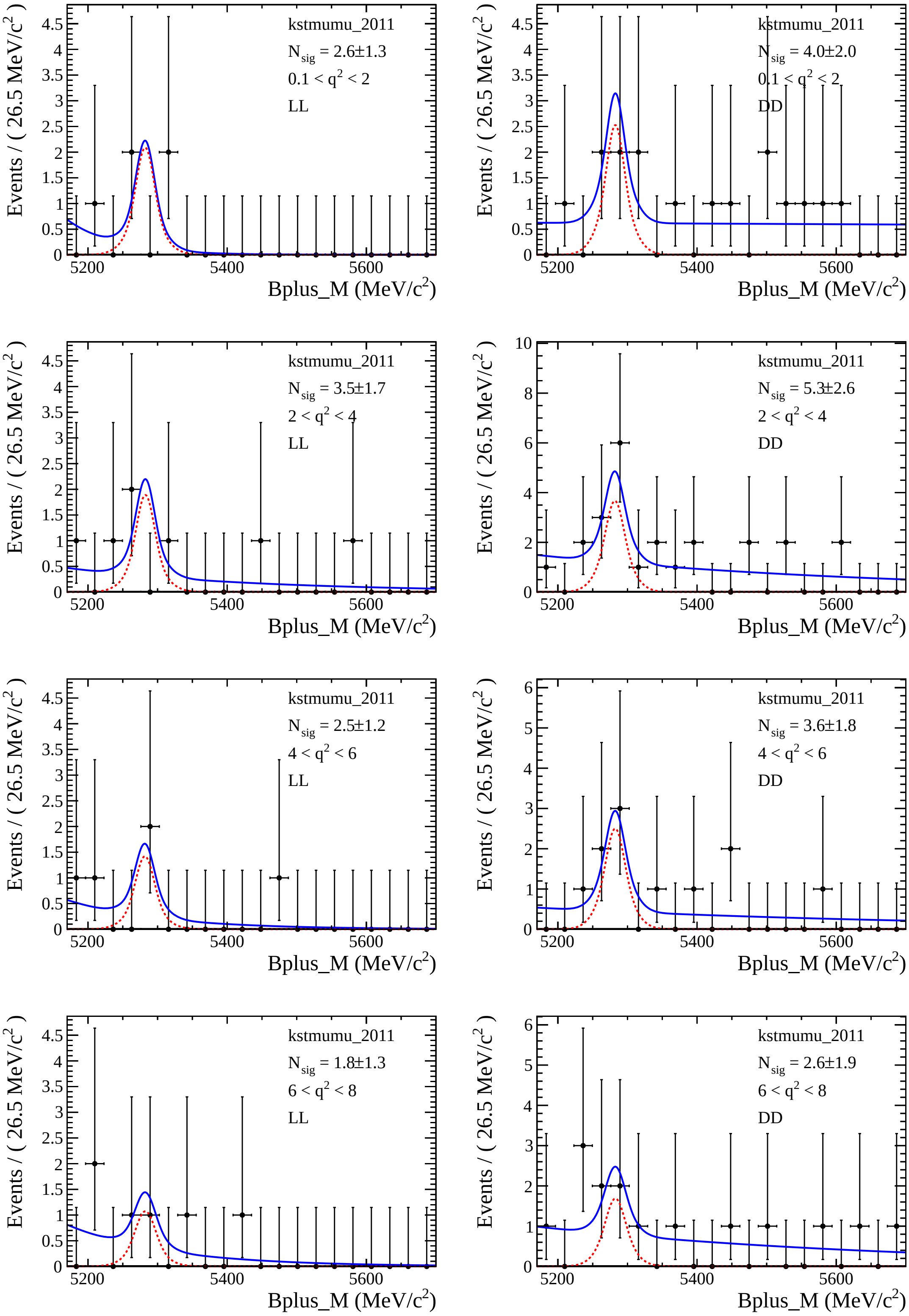


Figure 123: Mass fits to $B^+ \to (K^{*+} \to K^0_S \pi^+)\mu^+\mu^-$ 2011 data in the bins of q^2 below the J/ψ. Note that the LL and DD categories are fit simultaneously to ensure they have the same branching fraction.

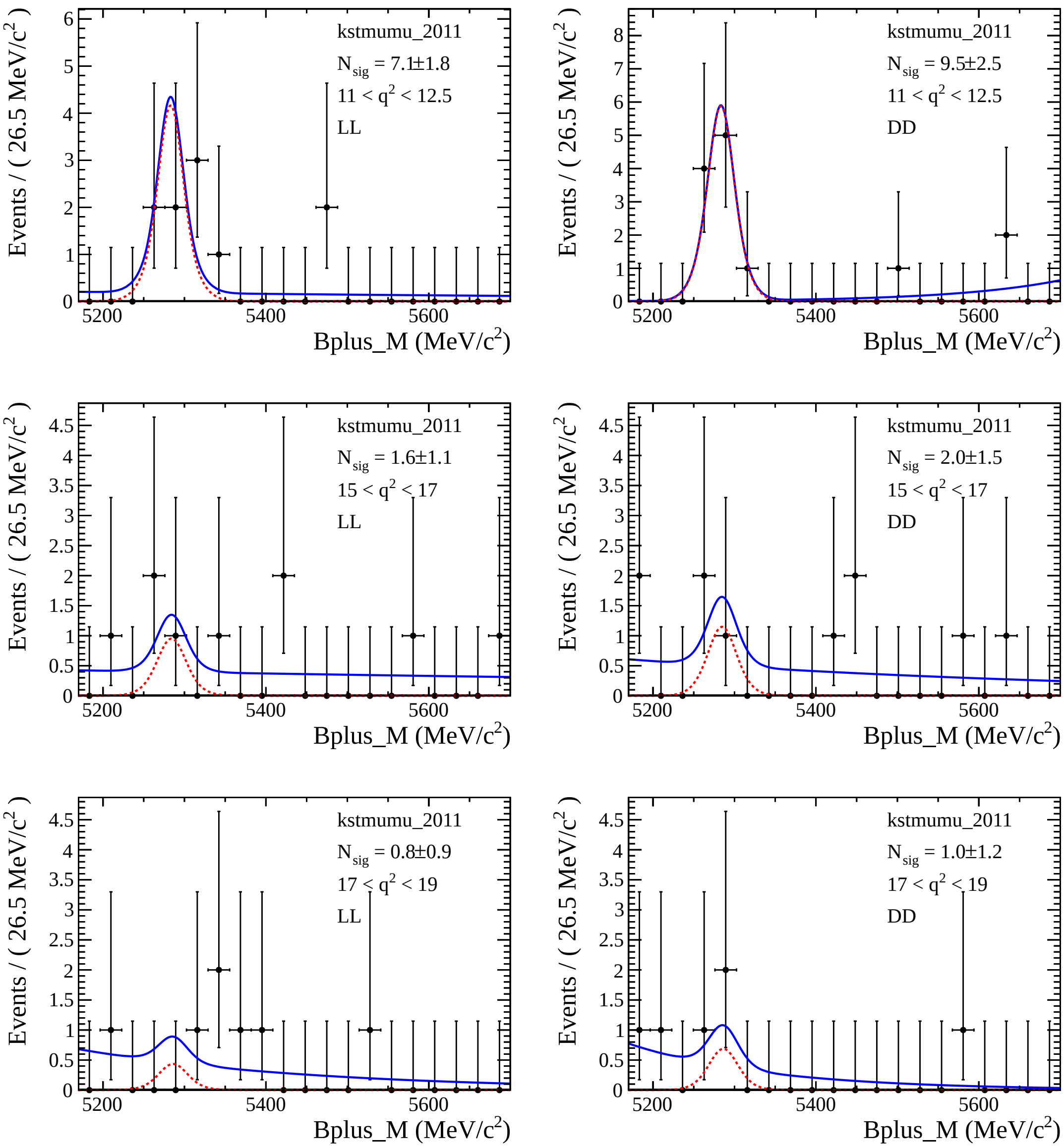


Figure 124: Mass fits to $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ 2011 data in bins of q^2 above the J/ψ. Note that the LL and DD categories are fit simultaneously to ensure they have the same branching fraction.

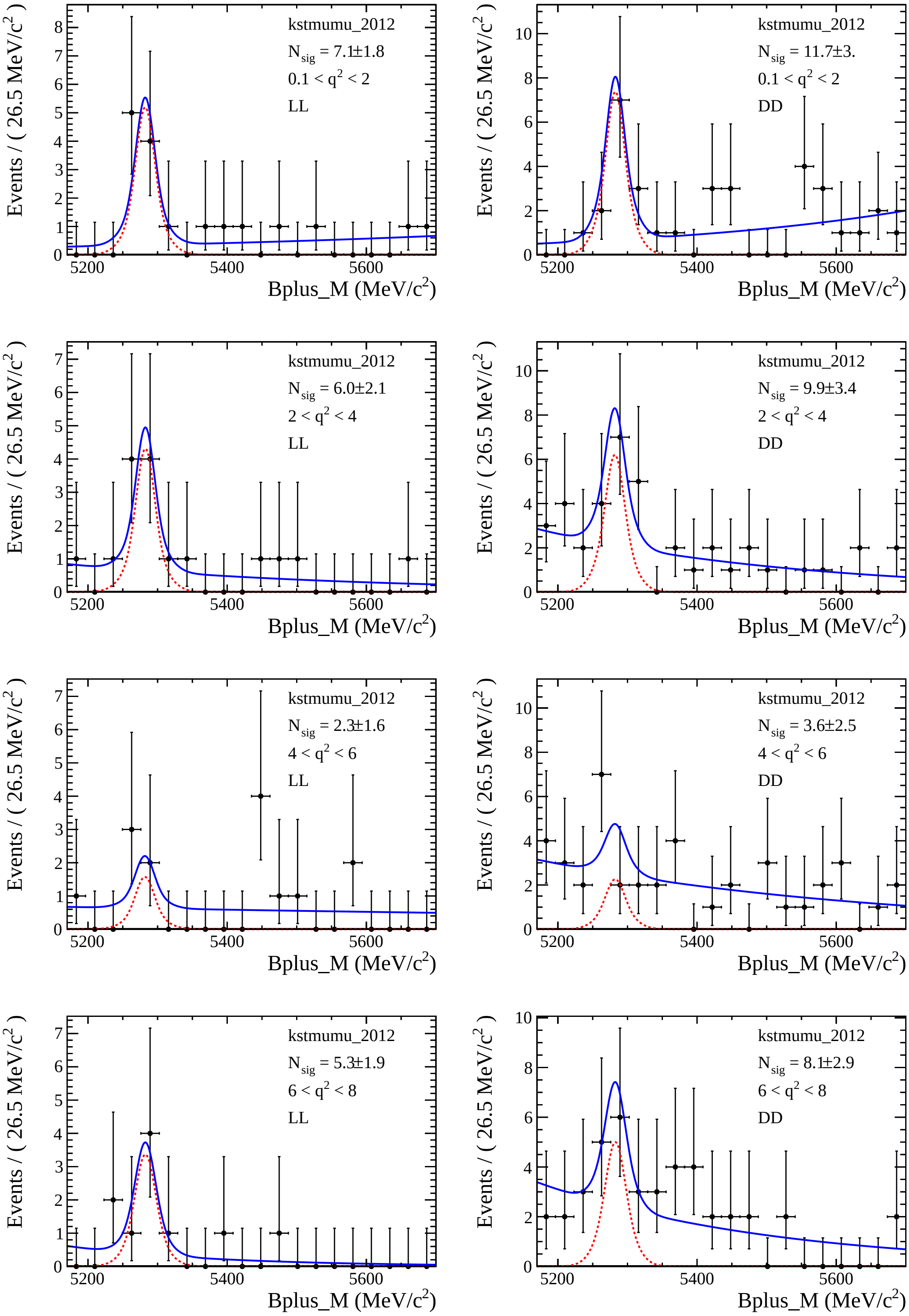


Figure 125: Mass fits to $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ 2012 data in bins of q^2 below the J/ψ. Note that the LL and DD categories are fit simultaneously to ensure they have the same branching fraction.

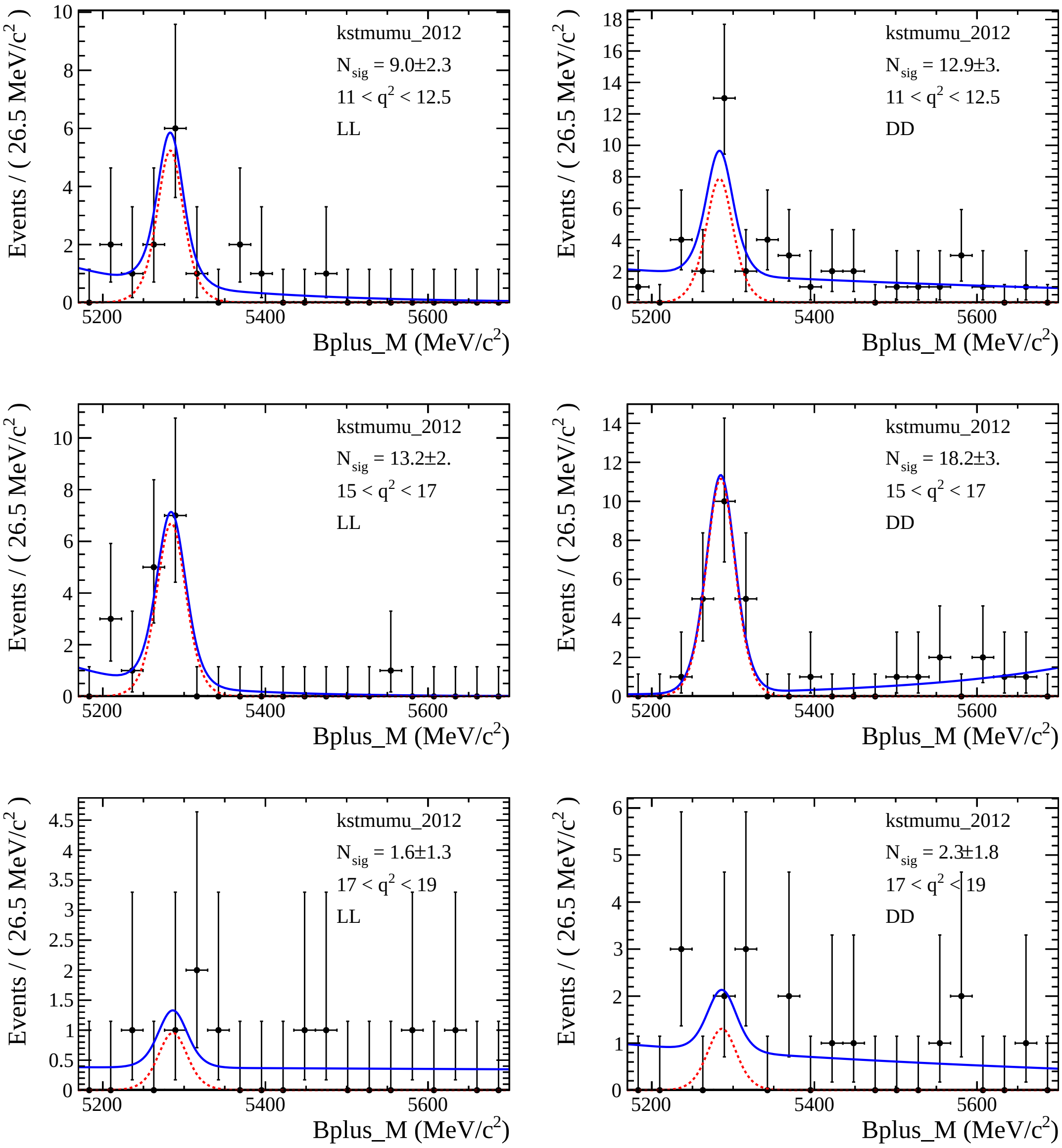


Figure 126: Mass fits to $B^+ \to (K^{*+} \to K^0_S\pi^+)\mu^+\mu^-$ 2012 data in bins of q^2 above the J/ψ. Note that the LL (left) and DD (right) categories are fit simultaneously to ensure they have the same branching fraction.

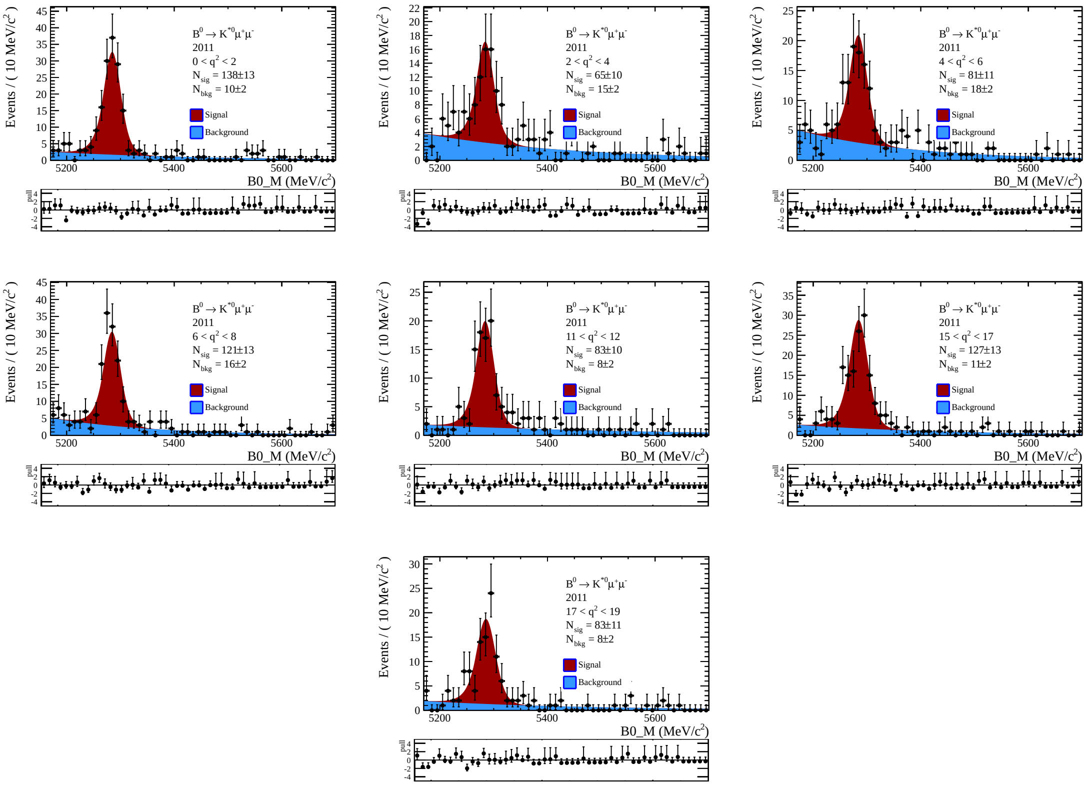


Figure 127: Mass fits to 2011 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ data in bins of q^2.

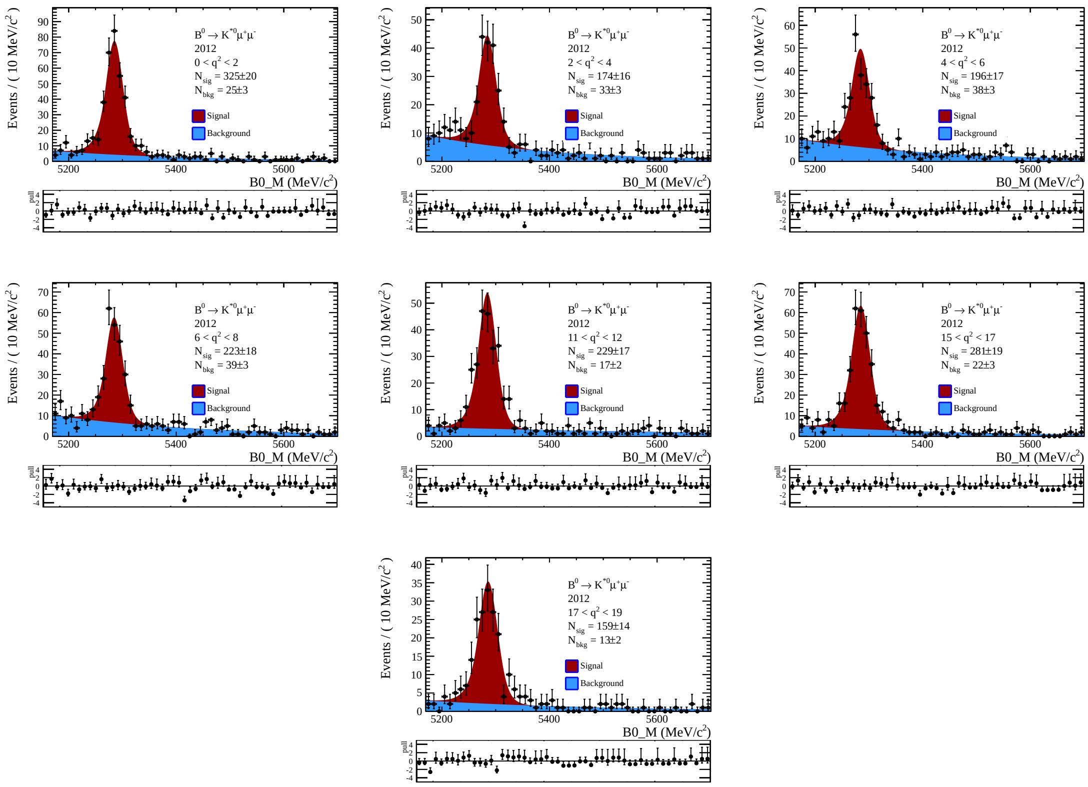


Figure 128: Mass fits to 2012 $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ data in bins of q^2.

C Trigger efficiency with level

In this section the trigger efficiency is shown for the three trigger levels. The HLT2 efficiency at low q^2 is different for the different running conditions. This due to the fact that in 2011 DD K^0_S did not participate in the Topological trigger and so the efficiency is lower in 2011. In early 2012, there was a bug which again rendered the K^0_S useless in the Topo. Finally, late in 2012 the K^0_S bug was fixed and so the trigger is more efficient in this period at low q^2. For the LL category, the situation is reversed, as there was a bug (lifetime cut at $< 10\,\text{ps}$) introduced into the topological trigger late in 2012 which lowers the efficiency for this period.

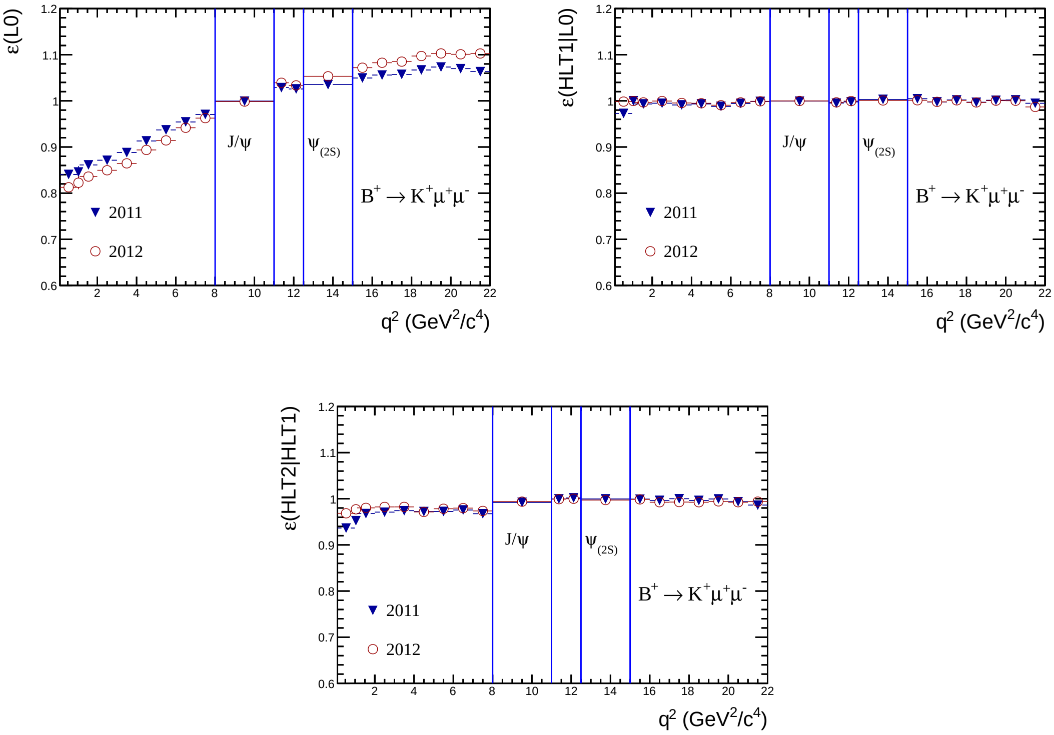


Figure 129: Relative efficiency between $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to J/\psi\, K^+$ for the three different trigger levels, L0, HLT1 and HLT2.

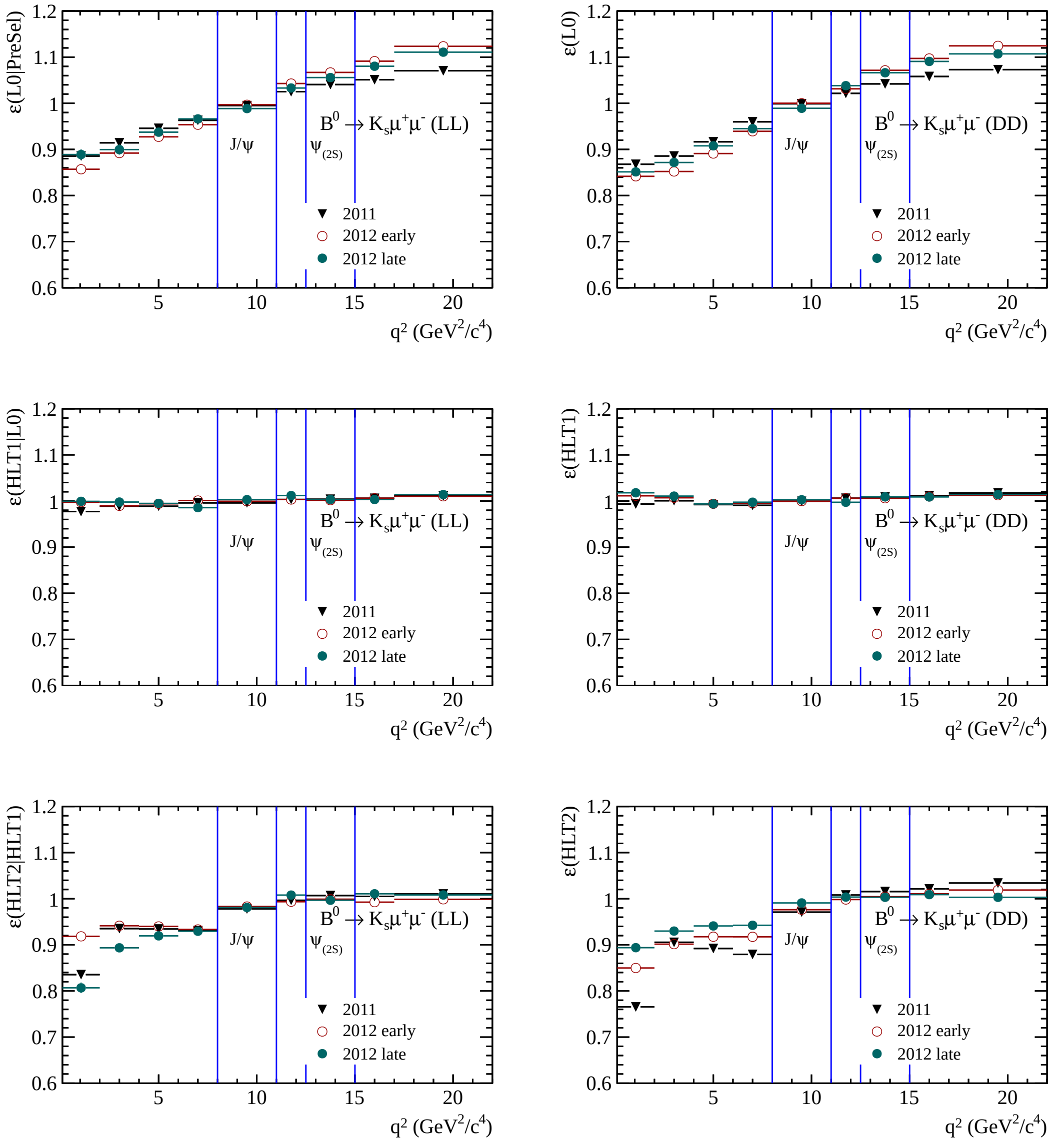


Figure 130: Relative efficiency between $B^0 \to K^0_S\mu^+\mu^-$ and $B^0 \to J/\psi K^0_S$ for the three different trigger levels, L0 on the top row, HLT1 on the middle row and HLT2 on the bottom row.

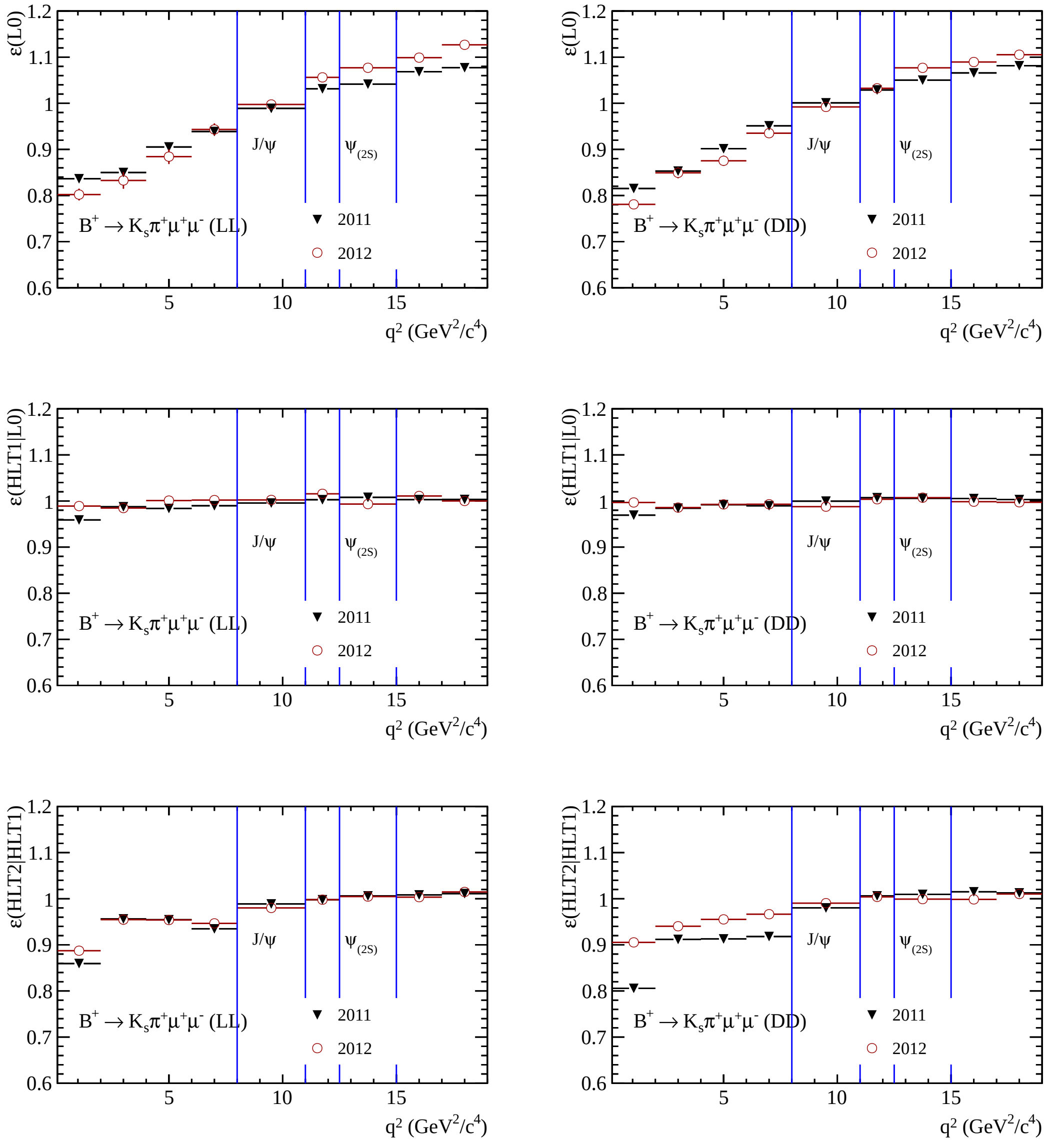


Figure 131: Relative efficiency between $B^+ \to (K^{*+} \to K^0_{\rm S}\pi^+)\mu^+\mu^-$ and $B^+ \to J/\psi(K^{*+} \to K^0_{\rm S}\pi^+)$ for the three different trigger levels, L0 on the top row, HLT1 on the middle row and HLT2 on the bottom row.

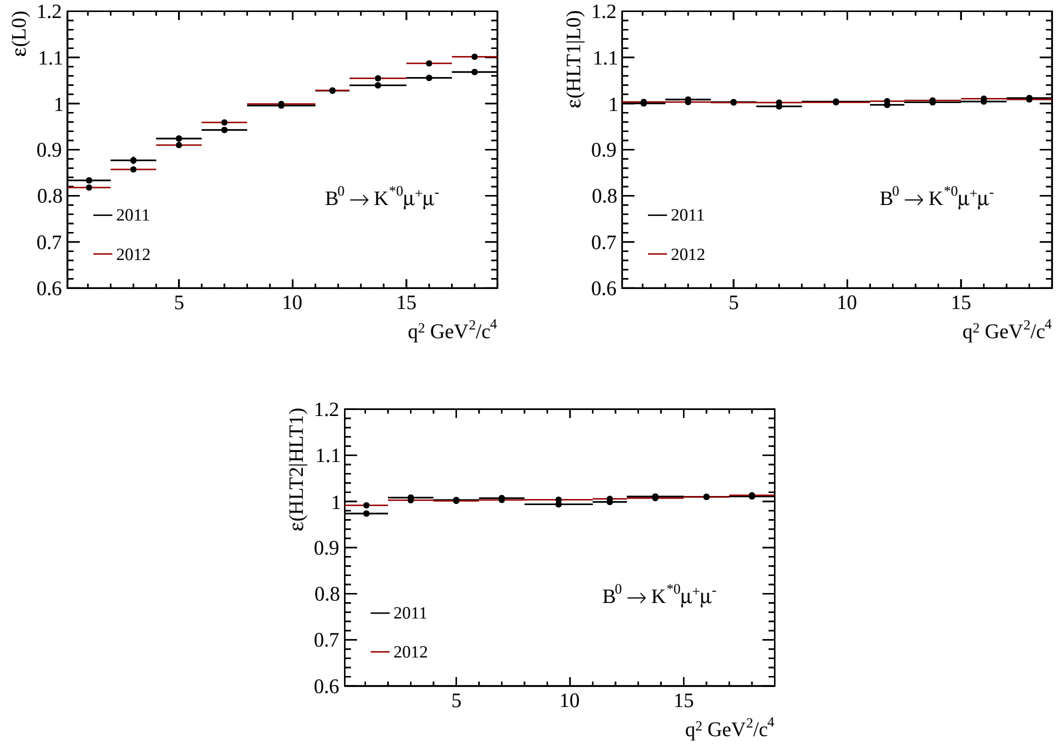


Figure 132: Relative efficiency between $B^0 \to (K^{*0} \to K^+\pi^-)\mu^+\mu^-$ and $B^0 \to J/\psi\,(K^{*0} \to K^+\pi^-)$ for the three different trigger levels, L0, HLT1 and HLT2.

D Data vs MC for K_S^0 kinematics

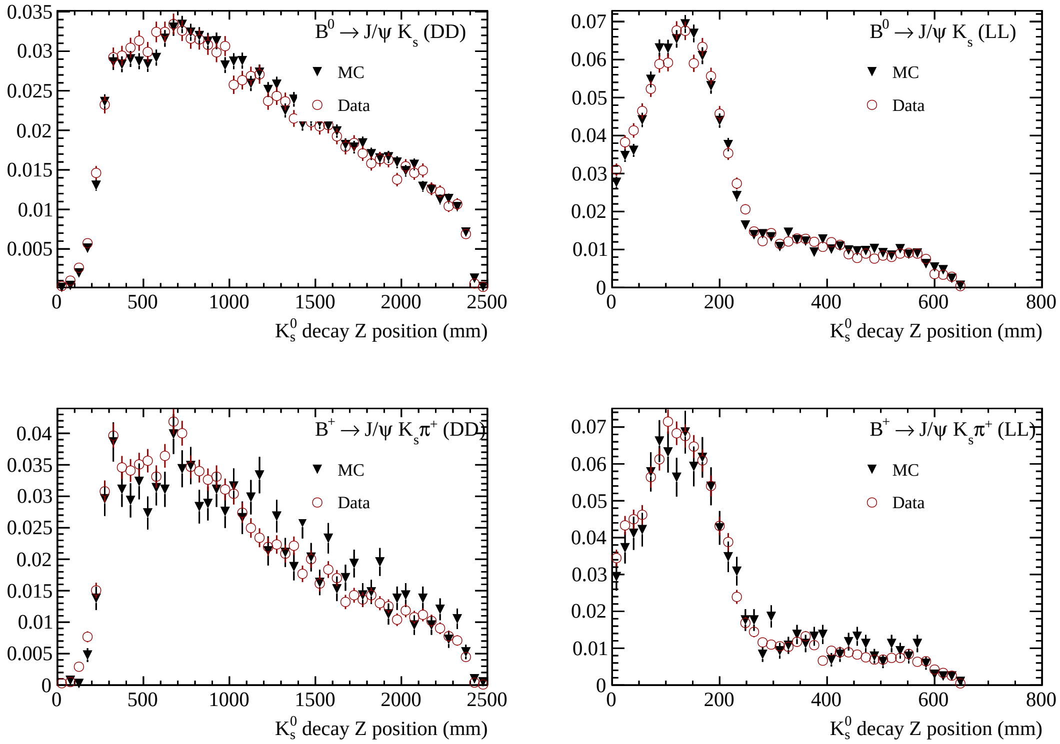


Figure 133: Distributions of the Z position of the K_S^0 decay vertex in data and simulation for $B^0 \to J/\psi\, K_S^0$ and $B^+ \to J/\psi\,(K^{*+} \to K_S^0 \pi^+)$ decays.

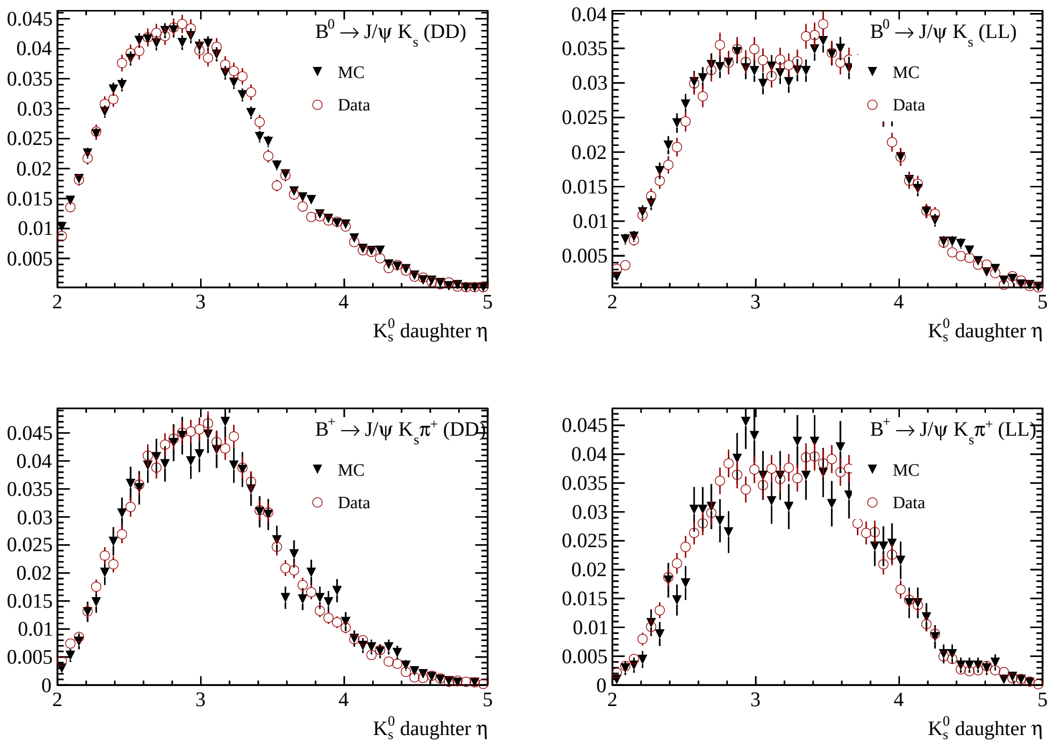


Figure 134: Distributions of the $K_{\rm S}^0$ daughter psuedo-rapidity in data and simulation for $B^0 \to J/\psi\, K_{\rm S}^0$ and $B^+ \to J/\psi\,(K^{*+} \to K_{\rm S}^0\pi^+)$ decays.

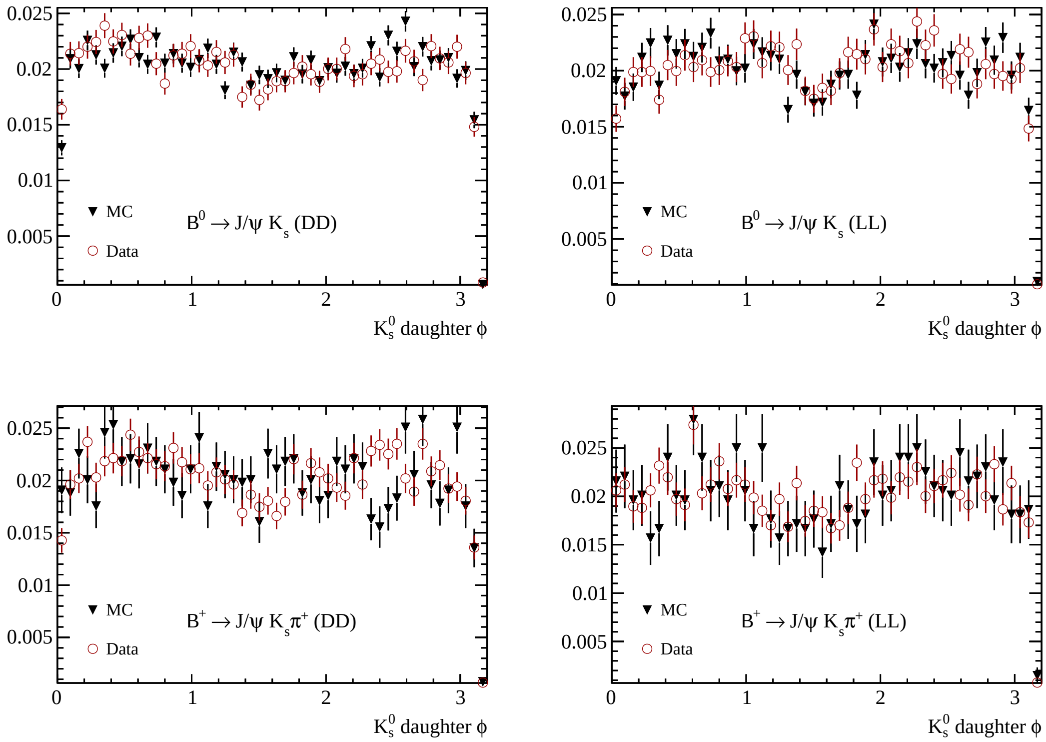


Figure 135: Distributions of the K^0_S daughter ϕ coordinate in data and simulation for $B^0 \rightarrow J/\psi\, K^0_S$ and $B^+ \rightarrow J/\psi\,(K^{*+} \rightarrow K^0_S\pi^+)$ decays.

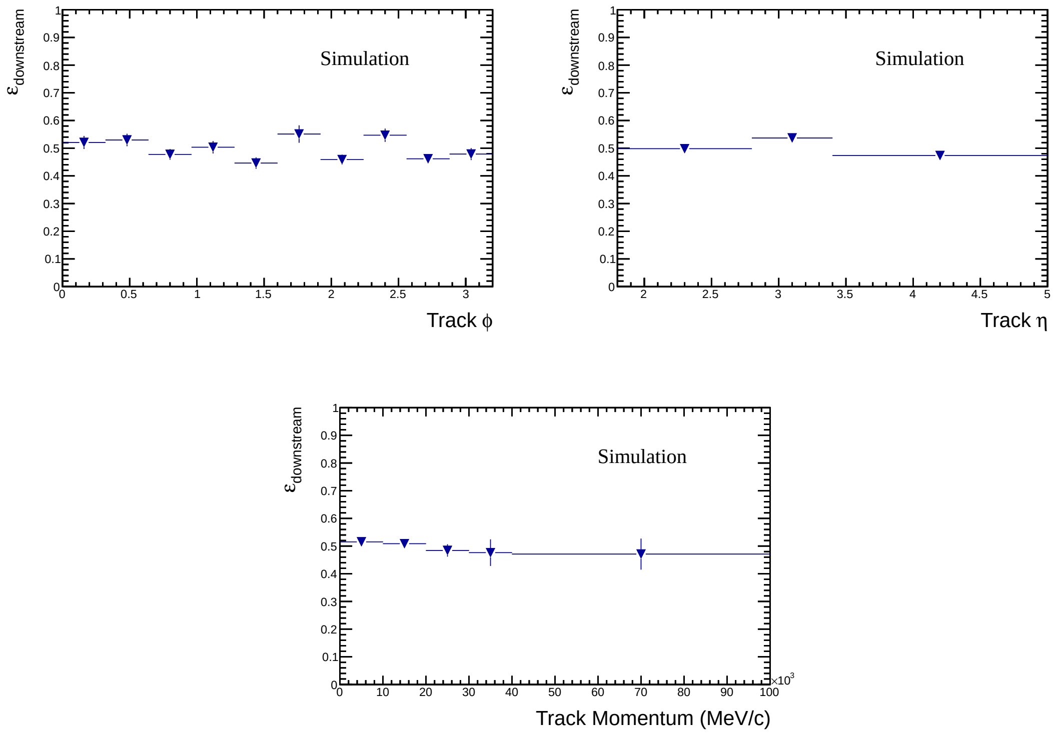


Figure 136: Downstream tracking efficiency as a function of ϕ, η and momentum.