



# Measurement of the CKM angle $\gamma$ using $B_s^0 \rightarrow D_s K \pi \pi$ decays

P. d'Argent<sup>1</sup>, E. Gersabeck<sup>2</sup>, M. Kecke<sup>1</sup>, M. Schiller<sup>3</sup>

<sup>1</sup>*Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany*

<sup>2</sup>*School of Physics and Astronomy, University of Manchester, Manchester, United Kingdom*

<sup>3</sup>*School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom*

## Abstract

We present the first measurement of the weak phase  $2\beta + \gamma$  obtained from a time-dependent (amplitude) analysis of  $B_s^0 \rightarrow D_s K \pi \pi$  decays using proton-proton collision data corresponding to an integrated luminosity of  $\text{xxx fb}^{-1}$  recorded by the LHCb detector.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Sensitivity studies</b>	<b>2</b>
2.1	PDF . . . . .	2
2.2	Estimation of coherence factor . . . . .	3
2.3	Results . . . . .	5
<b>3</b>	<b>Selection</b>	<b>8</b>
3.1	Cut-based selection . . . . .	8
3.2	Multivariate stage . . . . .	9
<b>4</b>	<b>Fits to invariant mass distributions of signal and normalization channel</b>	<b>14</b>
4.1	Signal models for $m(D_s\pi\pi\pi)$ and $m(D_sK\pi\pi)$ . . . . .	14
4.2	Background models for $m(D_s\pi\pi\pi)$ . . . . .	15
4.3	Background models for $m(D_sK\pi\pi)$ . . . . .	15
4.4	Fit to $B_s^0 \rightarrow D_s\pi\pi\pi$ candidates . . . . .	17
4.5	Fit to $B_s^0 \rightarrow D_sK\pi\pi$ candidates . . . . .	17
4.6	Extraction of signal weights . . . . .	17
<b>5</b>	<b>Flavour Tagging</b>	<b>19</b>
5.1	OS tagging calibration . . . . .	20
5.2	SS tagging calibration . . . . .	20
5.3	Tagging performance comparison between the signal and normalization channel . . . . .	20
5.4	Combination of OS and SS taggers . . . . .	21
<b>6</b>	<b>Decay-time acceptance</b>	<b>24</b>
6.1	Comparison of acceptance in subsamples . . . . .	25
6.2	Results . . . . .	27
<b>7</b>	<b>Decay-time Resolution</b>	<b>32</b>
7.1	Formalism . . . . .	32
7.2	Decay-time Error in Run I & Run II . . . . .	32
7.3	Fits to the decay time distributions . . . . .	33
7.4	Results . . . . .	34
<b>8</b>	<b>Time dependent fit</b>	<b>36</b>
8.1	sFit to $B_s^0 \rightarrow D_s\pi\pi\pi$ data . . . . .	36
8.2	sFit to $B_s^0 \rightarrow D_sK\pi\pi$ data . . . . .	36
8.3	Results . . . . .	36
<b>A</b>	<b>Detailed mass fits</b>	<b>37</b>
<b>B</b>	<b>Decay-time Resolution fits</b>	<b>41</b>
<b>C</b>	<b>MC corrections</b>	<b>42</b>

<b>D Data distributions</b>	<b>47</b>
D.1 Comparison of signal and calibration channel . . . . .	47
D.2 Comparison of Run-I and Run-II data . . . . .	49
D.3 Comparison of $D_s$ final states . . . . .	51
D.4 Comparison of trigger categories . . . . .	53
<b>References</b>	<b>55</b>

# 1 Introduction

- The weak phase  $\gamma$  is the least well known angle of the CKM unitary triangle. A key channel to measure  $\gamma$  is the time-dependent analysis of  $B_s^0 \rightarrow D_s K$  decays [1], [2].  
 The  $B_s^0 \rightarrow D_s K\pi\pi$  proceeds at tree level via the transitions shown in Fig. 1.1 a) and b).

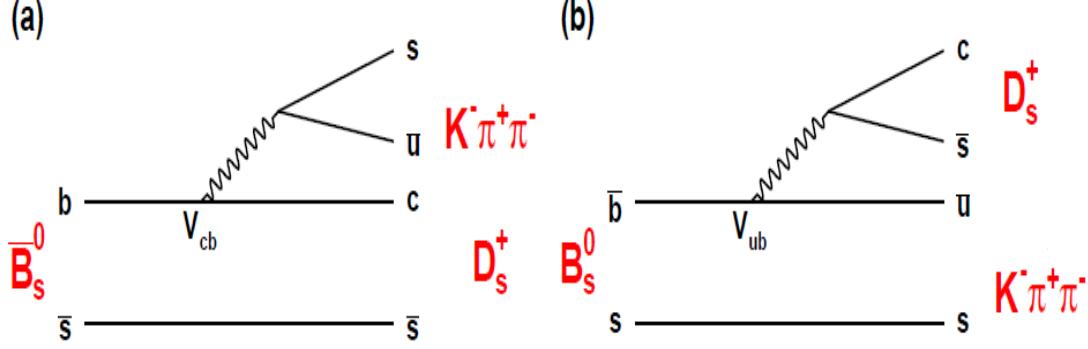


Figure 1.1: Feynman diagram of the  $B_s^0 \rightarrow D_s K\pi\pi$  decay, proceeding via a)  $b \rightarrow c$  transitions or b)  $b \rightarrow u$  transitions.

To measure the weak CKM phase  $\gamma \equiv \arg[-(V_{ud}V_{ub}^*)/(V_{cd}V_{cb}^*)]$ , a decay with interference between  $b \rightarrow c$  and  $b \rightarrow u$  transitions at tree level is needed [1]. As illustrated in Fig. 1.1, this is the case for the presented decay mode. A measurement of  $\gamma$  using  $B_s^0 \rightarrow D_s K\pi\pi$  decays, where the  $K\pi\pi$  subsystem is dominated by excited kaon states such as the  $K_1(1270)$  and  $K_1(1400)$  resonances, is performed. It is complementary to the above mentioned analysis of  $B_s^0 \rightarrow D_s K$ , making use of a fully charged final state, where every track is detected in the vertex locator. To account for the non-constant strong phase across the Dalitz plot, one can either develop a time-dependent amplitude model or select a suitable phase-space region and introduce a coherence factor as additional hadronic parameter to the fit.

This analysis is based on the first observation of the  $B_s^0 \rightarrow D_s K\pi\pi$  decay presented in [3] and [4], where its branching ratio is measured relative to  $B_s^0 \rightarrow D_s \pi\pi\pi$ . The result obtained by the previous analysis is  $0.052 \pm 0.005 \pm 0.003$ , where the uncertainties are statistical and systematical, respectively. In this note, we present a measurement of  $\gamma$ , making use of the full phase space by using a 6 dimensional time- and amplitude-dependent fit.

## 2 Sensitivity studies

### 2.1 PDF

First, I define the purely hadronic amplitudes for a given phasespace point  $x$ . The weak phase dependence is written latter explicitly in the pdf.

$$A(B_s^0 \rightarrow D_s^- K^+ \pi\pi) \equiv A(x) = \sum_i a_i A_i(x) \quad (2.1)$$

$$A(B_s^0 \rightarrow D_s^+ K^- \pi\pi) \equiv \bar{A}(\bar{x}) = \sum_i \bar{a}_i \bar{A}_i(\bar{x}) \quad (2.2)$$

$$A(\bar{B}_s^0 \rightarrow D_s^- K^+ \pi\pi) = \bar{A}(x) \text{ (Assuming no direct CPV)} \quad (2.3)$$

$$A(\bar{B}_s^0 \rightarrow D_s^+ K^- \pi\pi) = A(\bar{x}) \text{ (Assuming no direct CPV)} \quad (2.4)$$

The full time-dependent amplitude pdf is given by:

$$\begin{aligned} P(x, t, q_t, q_f) \propto & [(|A(x)|^2 + |\bar{A}(x)|^2) \cosh\left(\frac{\Delta\Gamma t}{2}\right) \\ & + q_t q_f (|A(x)|^2 - |\bar{A}(x)|^2) \cos(m_s t) \\ & - 2\text{Re}(A(x)^* \bar{A}(x) e^{-iq_f(\gamma-2\beta_s)}) \sinh\left(\frac{\Delta\Gamma t}{2}\right) \\ & - 2q_t q_f \text{Im}(A(x)^* \bar{A}(x) e^{-iq_f(\gamma-2\beta_s)}) \sin(m_s t)] e^{-\Gamma t} \end{aligned} \quad (2.5)$$

where  $q_t = +1$  ( $-1$ ) for a  $B_s^0$  ( $\bar{B}_s^0$ ) tag and  $q_f = +1$  ( $-1$ ) for  $D_s^- K^+ \pi\pi$  ( $D_s^+ K^- \pi\pi$ ) final states.

Integrating over the phasespace, we get

$$\begin{aligned} \int P(x, t, q_t, q_f) dx \propto & [\cosh\left(\frac{\Delta\Gamma t}{2}\right) \\ & + q_t q_f \left(\frac{1-r^2}{1+r^2}\right) \cos(m_s t) \\ & - 2 \left(\frac{\kappa r \cos(\delta - q_f(\gamma - 2\beta_s))}{1+r^2}\right) \sinh\left(\frac{\Delta\Gamma t}{2}\right) \\ & - 2q_t q_f \left(\frac{\kappa r \sin(\delta - q_f(\gamma - 2\beta_s))}{1+r^2}\right) \sin(m_s t)] e^{-\Gamma t} \\ = & [\cosh\left(\frac{\Delta\Gamma t}{2}\right) + q_t q_f C \cos(m_s t) - \kappa D_{q_f} \sinh\left(\frac{\Delta\Gamma t}{2}\right) - q_t \kappa S_{q_f} \sin(m_s t)] e^{-\Gamma t} \end{aligned} \quad (2.6)$$

where the  $C, D_{q_f}, S_{q_f}$  are defined exactly as for  $D_s K$ . The coherence factor is defined as :

$$\kappa e^{i\delta} \equiv \frac{\int A(x)^* \bar{A}(x) dx}{\sqrt{\int |A(x)|^2 dx} \sqrt{\int |\bar{A}(x)|^2 dx}} \quad (2.7)$$

$$r \equiv \frac{\sqrt{\int |\bar{A}(x)|^2 dx}}{\sqrt{\int |A(x)|^2 dx}} \quad (2.8)$$

31 and appears in front of the  $D_{qf}, S_{qf}$  terms. This means one additional fit parameter for  
 32 the lifetime fit. In the limit of only one contributing resonance  $\kappa \rightarrow 1$ .

33

## 34 2.2 Estimation of coherence factor

35 To estimate the coherence factor we could generate many toys with random  $a_i$  and  $\bar{a}_i$   
 36 values (see [https://twiki.cern.ch/twiki/pub/LHCbPhysics/Bu2DKstar/LHCb-ANA-2017-005\\_v1.pdf](https://twiki.cern.ch/twiki/pub/LHCbPhysics/Bu2DKstar/LHCb-ANA-2017-005_v1.pdf)) using the set of amplitudes show in our last talk. However with so many  
 37 interfering amplitudes, I would be surprised if you couldn't generate every possible value  
 38 for  $\kappa$ . In any case, this would give us a range where to expect possible values for  $\kappa$ . Worst  
 39 case would be  $0 \leq \kappa \leq 1$ .  
 40

41

42 Assumptions:

43 •  $A(x) = \sum_i a_i A_i(x)$   
 44     $\bar{A}(x) = \sum_i \bar{a}_i \bar{A}_i(x)$

45 • Use amplitudes from flavor-averaged, time-integrated fit

46 • Draw random  $a_i$  and  $\bar{a}_i$  values

47 • Constraints:

48     $\int(|a_i A_i(x)|^2 + |\bar{a}_i \bar{A}_i(x)|^2) dx/N = F_i^{eff}$   
 49     $r \approx 0.4$  (ration of CKM elements)

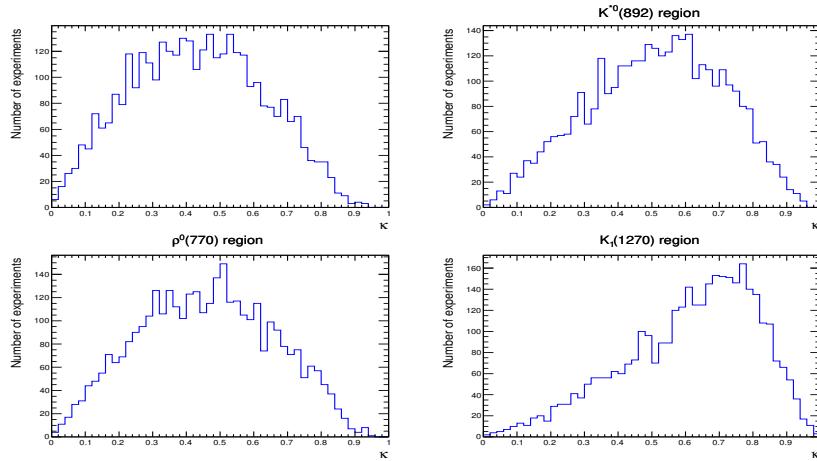


Figure 2.1

Table 2.1

Region	$<\kappa>$ (%)	Cut eff. (%)
Full	43	100
$K^*(892)$	51	43
$\rho^0(770)$	46	47
$K_1(1270)$	61	23

## 50 2.3 Results

51 Assumptions:

- 52 • Use amplitudes from flavor-averaged, time-integrated fit
- 53 •  $r = 0.4$  (ratio of CKM elements)
- 54 • PDG values for:  $\tau, \Delta m_s, \Delta\Gamma, \beta_s$
- 55 •  $\epsilon(x, t) = \text{const.}$ , perfect resolution
- 56 •  $\epsilon_{Tag} = 0.66, \langle \omega \rangle = 0.4$
- 57 •  $N_{signal} = 3000$  (Run1+15/16 data)

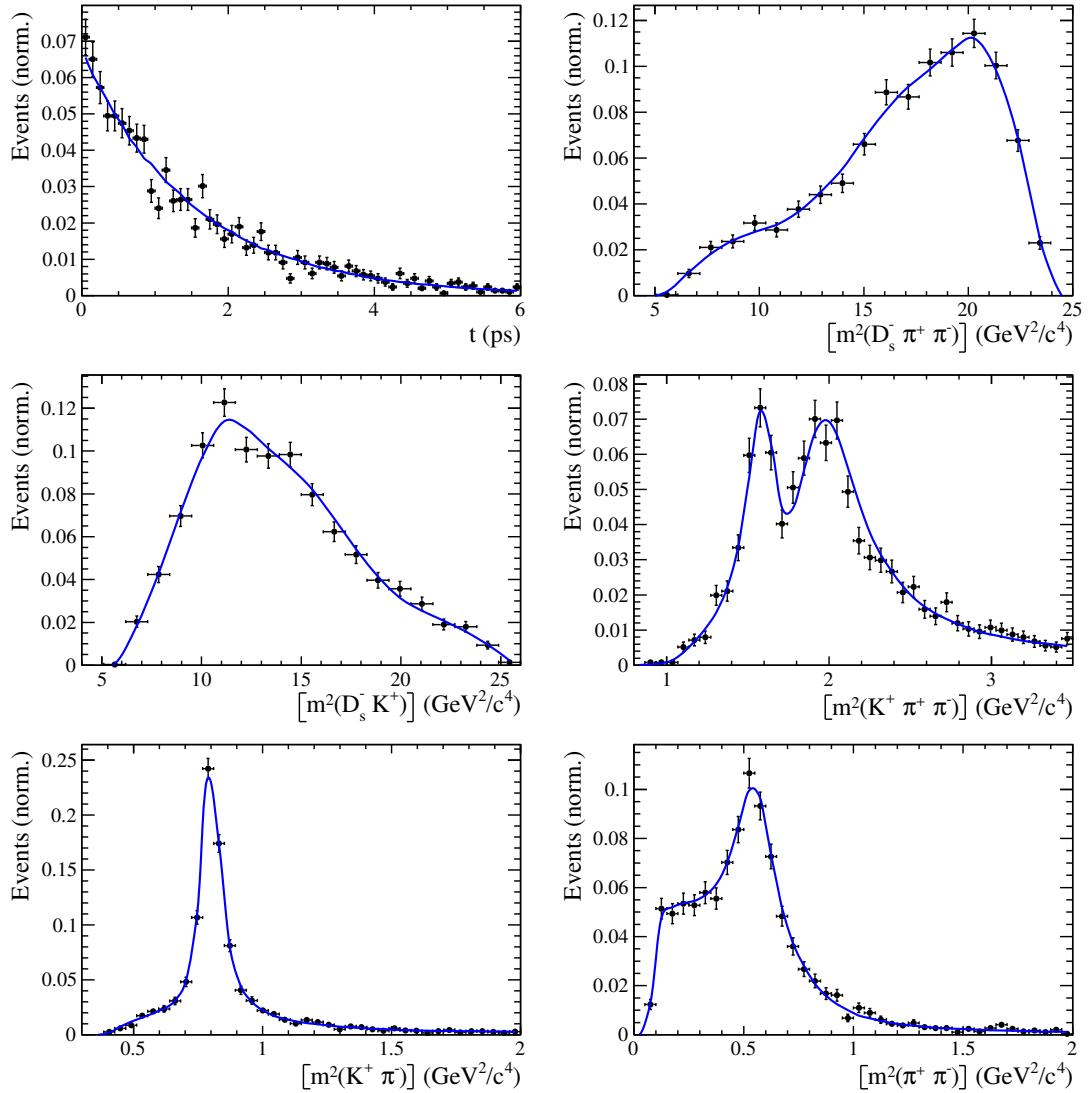
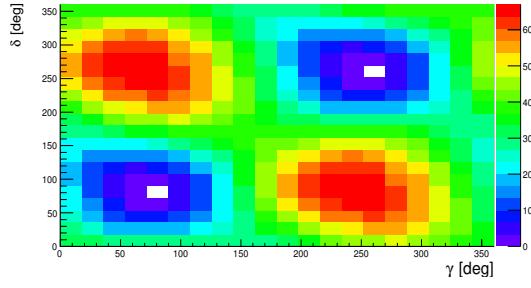


Figure 2.2: Example toy fit



Generated values:

$$\gamma = 70^\circ, \delta = 100^\circ$$

Fit result:

$$\begin{aligned} \gamma &= 74 \pm 15^\circ, \delta = 84 \pm 15^\circ \\ (\gamma &= 254 \pm 15^\circ, \delta = 264 \pm 15^\circ) \end{aligned}$$

Figure 2.3: Likelihood scan

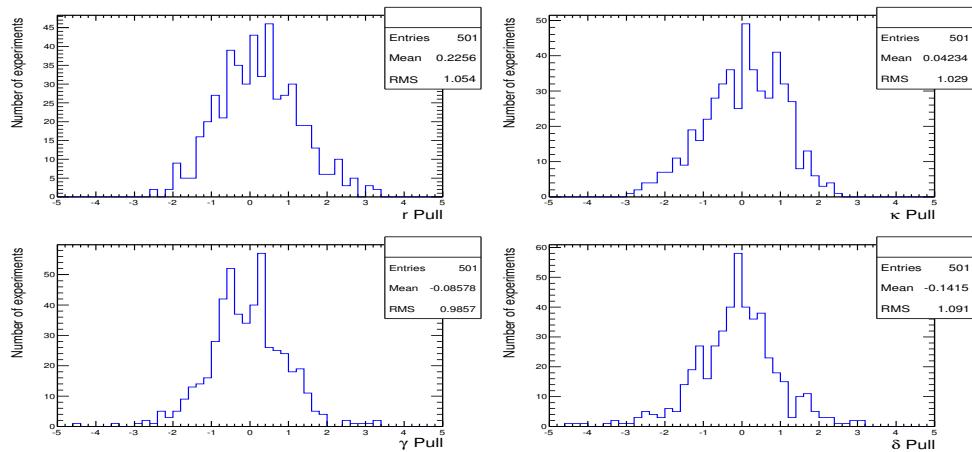


Figure 2.4: Pulls

Table 2.2

	Generated	Full PDF	Phasespace integrated
$r$	0.4	$0.38 \pm 0.06$	unstable
$\kappa$	<b>0.2</b>	$0.23 \pm 0.13$	0.2 (fixed)
$\delta$	100	$99 \pm 22$	unstable
$\gamma$	70	$70 \pm 17$	unstable

	Generated	Full PDF	Phasespace integrated
$r$	0.4	$0.44 \pm 0.07$	$0.43 \pm 0.11$
$\kappa$	<b>0.4</b>	$0.41 \pm 0.14$	0.4 (fixed)
$\delta$	100	$101 \pm 19$	$95 \pm 41$
$\gamma$	70	$69 \pm 16$	$66 \pm 40$

	Generated	Full PDF	Phasespace integrated
$r$	0.4	$0.41 \pm 0.08$	$0.39 \pm 0.11$
$\kappa$	<b>0.6</b>	$0.60 \pm 0.13$	0.6 (fixed)
$\delta$	100	$98 \pm 17$	$92 \pm 25$
$\gamma$	70	$68 \pm 17$	$65 \pm 28$

	Generated	Full PDF	Phasespace integrated
$r$	0.4	$0.42 \pm 0.09$	$0.39 \pm 0.09$
$\kappa$	<b>1.0</b>	$0.96 \pm 0.03$	1.0 (fixed)
$\delta$	100	$100 \pm 17$	$100 \pm 17$
$\gamma$	70	$66 \pm 17$	$67 \pm 17$

58 **3 Selection**

59 For the presented analysis, we reconstruct the  $B_s^0 \rightarrow D_s K\pi\pi$  decay through two different  
60 final states of the  $D_s$  meson,  $D_s \rightarrow KK\pi$  and  $D_s \rightarrow \pi\pi\pi$ . Of those two final states  
61  $D_s \rightarrow KK\pi$  is the most prominent one, while  $\mathcal{BR}(D_s \rightarrow \pi\pi\pi) \approx 0.2 \cdot \mathcal{BR}(D_s \rightarrow KK\pi)$   
62 holds for the other one.

63 A two-fold approach is used to isolate the  $B_s^0 \rightarrow D_s K\pi\pi$  candidates from data passing  
64 the stripping line. First, further one-dimensional cuts are applied to reduce the level of  
65 combinatorial background and to veto some specific physical background. This stage is  
66 specific to the respective final state in which the  $D_s$  meson is reconstructed, since different  
67 physical backgrounds, depending on the respective final state, have to be taken into  
68 account. After that, a multivariate classifier is trained which combines the information  
69 of several input variables, including their correlation, into one powerful discriminator  
70 between signal and combinatorial background. For this stage, all possible  $D_s$  final states  
71 are treated equally.

72 **3.1 Cut-based selection**

73 In order to minimize the contribution of combinatorial background to our samples, we  
74 apply the following cuts to the b hadron:

- 75     • DIRA > 0.99994
- 76     • min IP  $\chi^2 < 20$  to any PV,
- 77     • FD  $\chi^2 > 100$  to any PV,
- 78     • Vertex  $\chi^2/\text{nDoF} < 8$ ,
- 79     •  $(Z_{D_s} - Z_{B_s^0}) > 0$ , where  $Z_M$  is the z-component of the position  $\vec{x}$  of the decay vertex  
80       for the  $B_s^0/D_s$  meson.

81 Additionally, we veto various physical backgrounds, which have either the same final  
82 state as our signal decay, or can contribute via a single misidentification of  $K \rightarrow \pi$  or  
83  $K \rightarrow p$ . In the following, the vetoes are ordered by the reconstructed  $D_s$  final state they  
84 apply to:

85 1. All:

- 86     (a)  $B_s^0 \rightarrow D_s^+ D_s^-$  :  $|M(K\pi\pi) - m_{D_s}| > 20$  MeV/ $c^2$ .
- 87     (b)  $B_s^0 \rightarrow D_s^- K^+ K^- \pi^+$  : possible with single missID of  $K^- \rightarrow \pi^-$ , rejected by  
88       requiring  $\pi^-$  to fulfill  $\text{DLL}_{K\pi} < 5$ .

89 2.  $D_s \rightarrow KK\pi$

- 90     (a)  $B^0 \rightarrow D^+(\rightarrow K^+\pi^-\pi^+) K\pi\pi$  : possible with single missID of  $\pi^+ \rightarrow K^+$ , vetoed  
91       by changing particle hypothesis and recompute  $|M(K^+\pi^-\pi^+) - m_{D_p}| > 30$   
92       MeV/ $c^2$ , or the  $K^+$  has to fulfill  $\text{DLL}_{K\pi} > 10$ .

- 93 (b)  $\Lambda_b^0 \rightarrow \Lambda_c^+ (\rightarrow pK^-\pi^+)K\pi\pi$  : possible with single missID of  $p \rightarrow K^+$ , vetoed by  
 94 changing particle hypothesis and recompute  $M(pK^-\pi^+) - m_{\Lambda_c^+} > 30 \text{ MeV}/c^2$ ,  
 95 or the  $K^+$  has to fulfill  $(\text{DLL}_{K\pi} - \text{DLL}_{p\pi}) > 5$ .  
 96 (c)  $D^0 \rightarrow KK$  :  $D^0$  combined with a random  $\pi$  can fake a  $D_s \rightarrow KK\pi$  decay and  
 97 be a background to our signal, vetoed by requiring  $M(KK) < 1840 \text{ MeV}/c^2$ .

98 3.  $D_s \rightarrow \pi\pi\pi$

- 99 (a)  $D^0 \rightarrow \pi\pi$  : combined with a random  $\pi$  can fake a  $D_s \rightarrow \pi\pi\pi$  decay and be a  
 100 background to our signal, vetoed by requiring both possible combinations to  
 101 have  $M(\pi\pi) < 1700 \text{ MeV}/c^2$ .

102 The most prominent final state used in this analysis is  $B_s^0 \rightarrow D_s(\rightarrow KK\pi)K\pi\pi$ , where  
 103 the  $D_s$  decay can either proceed via the narrow  $\phi$  resonance, the broader  $K^{*0}$  resonance, or  
 104 non resonant. Depending on the decay process being resonant or not, we apply additional  
 105 PID requirements on this final state:

- 106 • resonant case:

- 107 –  $D_s^+ \rightarrow \phi\pi^+$ , with  $|M(K^+K^-) - m_\phi| < 20 \text{ MeV}/c^2$  : no additional requirements,  
 108 since  $\phi$  is narrow and almost pure  $K^+K^-$ .
- 109 –  $D_s^+ \rightarrow \bar{K}^{*0}K^+$ , with  $|M(K^-\pi^+) - m_{K^{*0}}| < 75 \text{ MeV}/c^2$  :  $\text{DLL}_{K\pi} > 0$  for kaons,  
 110 since this resonance is more than ten times broader than  $\phi$ .

- 111 • non resonant case:  $\text{DLL}_{K\pi} > 5$  for kaons, since the non resonant category has  
 112 significant charmless contributions.

113 For the  $D_s \rightarrow \pi\pi\pi$  final state, we apply global PID requirements:

- 114 •  $\text{DLL}_{K\pi} < 10$  for all pions.  
 115 •  $\text{DLL}_{p\pi} < 10$  for all pions.

116 3.2 Multivariate stage

117 We use TMVA [5] to train a multivariate discriminator, which is used to further improve  
 118 the signal to background ratio. The 17 variables used for the training are:

- 119 •  $\max(\text{ghostProb})$  over all tracks
- 120 •  $\text{cone}(p_T)$  asymmetry of every track, which is defined to be the difference between the  
 121  $p_T$  of the  $\pi/K$  and the sum of all other  $p_T$  in a cone of radius  $r = \sqrt{(\Delta\Phi)^2 + (\Delta\eta)^2}$   
 122  $< 1 \text{ rad}$  around the signal  $\pi/K$  track.
- 123 •  $\min(\text{IP}\chi^2)$  over the  $X_s$  daughters
- 124 •  $\max(\text{DOCA})$  over all pairs of  $X_s$  daughters
- 125 •  $\min(\text{IP}\chi^2)$  over the  $D_s$  daughters

- $D_s$  and  $B_s^0$  DIRA
- $D_s$  FD significance
- $\max(\cos(D_s h_i))$ , where  $\cos(D_s h_i)$  is the cosine of the angle between the  $D_s$  and another track i in the plane transverse to the beam
- $B_s^0$  IP $\chi^2$ , FD $\chi^2$  and Vertex  $\chi^2$

Various classifiers were investigated in order to select the best performing discriminator. Consequently, a boosted decision tree with gradient boost (BDTG) is chosen as nominal classifier. We use truth-matched MC as signal input. Simulated signal candidates are required to pass the same trigger, stripping and preselection requirements, that were used to select the data samples. For the background we use events from the high mass sideband ( $m_{B_s^0 \text{candidate}} > 5600 \text{ MeV}/c^2$ ) of our data samples. As shown in Fig. 3.1, this mass region is sufficiently far away from signal structures and is expected to be dominantly composed of combinatorial background. For completeness, the mass distribution of preselected  $D_s \rightarrow hh$  candidates (where  $h = \pi$  or  $h = K$ ) is also shown in Fig. 3.1.

140

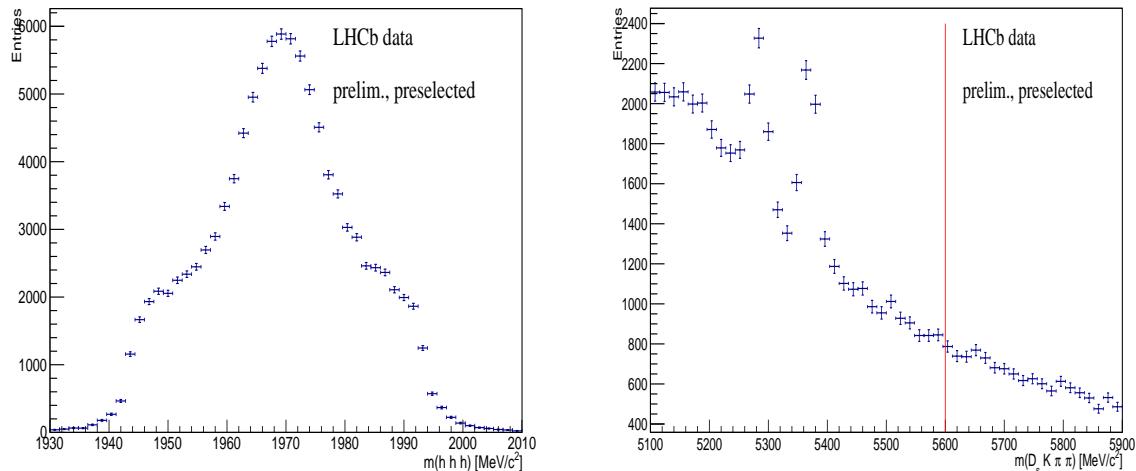


Figure 3.1: Invariant mass distribution of preselected (left)  $D_s \rightarrow hh$  and (right)  $B_s^0 \rightarrow D_s K\pi\pi$  candidates. For the  $B_s^0 \rightarrow D_s K\pi\pi$  candidates, the region right from the red coloured line with  $m_{B_s^0 \text{candidate}} > 5600 \text{ MeV}/c^2$  is used as background input for the boosted decision tree.

The distributions of the input variables for signal and background are shown in Fig. 3.2.

The relative importance of the input variables for the BDTG training is summarized in Table 3.1.

The BDTG output distribution for test and training samples is shown in Fig 3.3. No sign of overtraining is observed.

We determine the optimal cut value by maximizing the figure of merit  $S/\sqrt{S + B}$  where S is the signal yield and B the background yield in the signal region, defined to be within  $\pm 50 \text{ MeV}/c^2$  of the nominal  $B_s^0$  mass. To avoid a bias in the determination of the

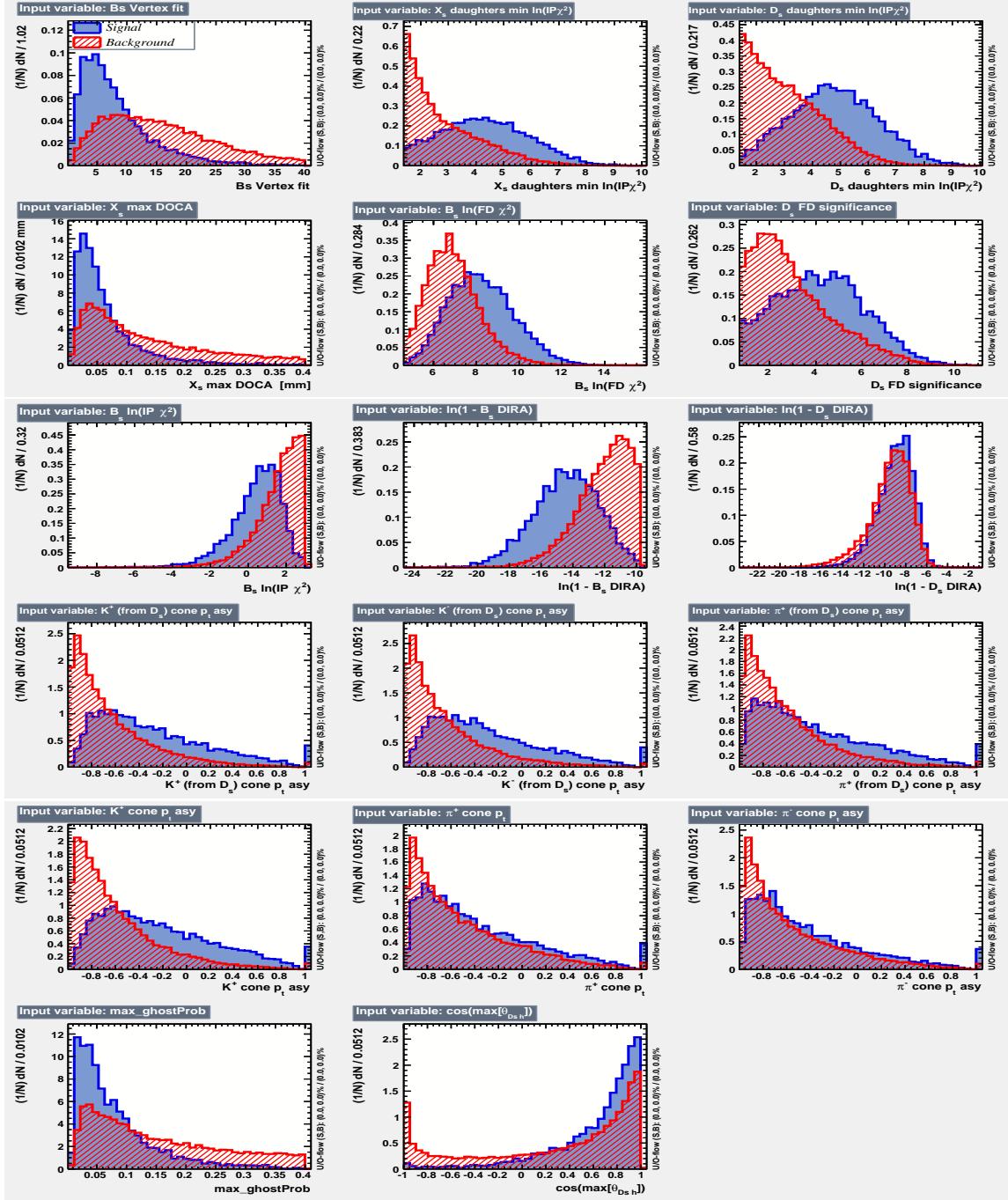


Figure 3.2: Distributions of the input variables used in the BDTG training. The background is shown as red hatched, while the signal is depicted solid blue.

branching fraction, we determine S and B using our normalization channel. All trigger, stripping and additional selection criteria described in this and the previous chapter are applied to the  $B_s^0 \rightarrow D_s \pi\pi\pi$  data samples. After that, we perform a simplified version of the fit to the invariant mass distribution of  $B_s^0 \rightarrow D_s \pi\pi\pi$  candidates described in Sec. ???. Here, a Gaussian function to model the signal and an exponential function to model combinatorial background is used. From this fit we estimate the number of signal events in our normalization channel. Multiplying that number with the PDG branching fraction

Variable	relative importance [%]
pi_minus_ptasy_1.00	7.32
log_Ds_FDCHI2_ORIVX	7.23
K_plus_ptasy_1.00	7.17
log_Ds_DIRA	6.96
Bs_ENDVERTEX_CHI2	6.82
max_ghostProb	6.76
pi_plus_ptasy_1.00	6.57
log_DsDaughters_min_IPCHI2	6.21
log_Bs_DIRA	6.15
K_plus_fromDs_ptasy_1.00	6.10
log_XsDaughters_min_IPCHI2	5.87
K_minus_fromDs_ptasy_1.00	5.62
cos(Ds h)	5.58
log_Bs_IPCHI2_OWNPV	5.08
log_Bs_FDCHI2_OWNPV	4.04
Xs_max_DOCA	3.98
pi_minus_fromDs_ptasy_1.00	2.59

Table 3.1: Summary of the relative importance of each variable in the training of the BDTG.

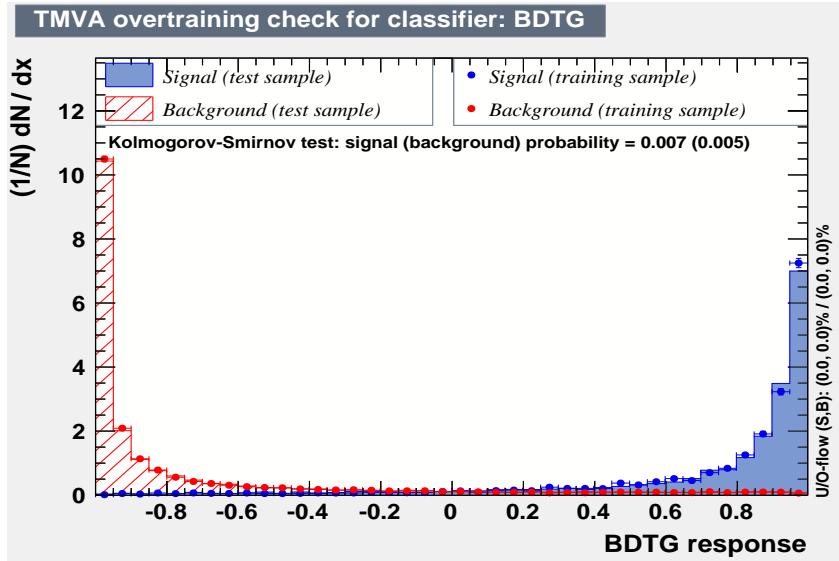


Figure 3.3: BDTG output classifier distribution for (blue) signal and (red) background. The response of an independent test sample (dots) is overlaid.

157 of  $\frac{\mathcal{B}(B_s^0 \rightarrow D_s K\pi\pi)}{\mathcal{B}(B_s^0 \rightarrow D_s \pi\pi\pi)}$  and the ratio of efficiencies discussed in Sec. ?? allows us to estimate the  
158 expected number of  $B_s^0 \rightarrow D_s K\pi\pi$  signal decays. The number of background events can  
159 then be computed as

$$N_{bkg} = N_{all} - N_{sig}|_{m_{B_s^0 \pm 50 \text{ MeV}/c^2}}. \quad (3.1)$$

160 The efficiency curves as a function of the cut value are shown in Fig. 3.4. The optimal  
 161 cut value is found to be  $\text{BDTG} > 0.7012$ . At this working point the signal efficiency is  
 162 estimated to be 72.47 %, while the background rejection in the signal region is 97.38 %.

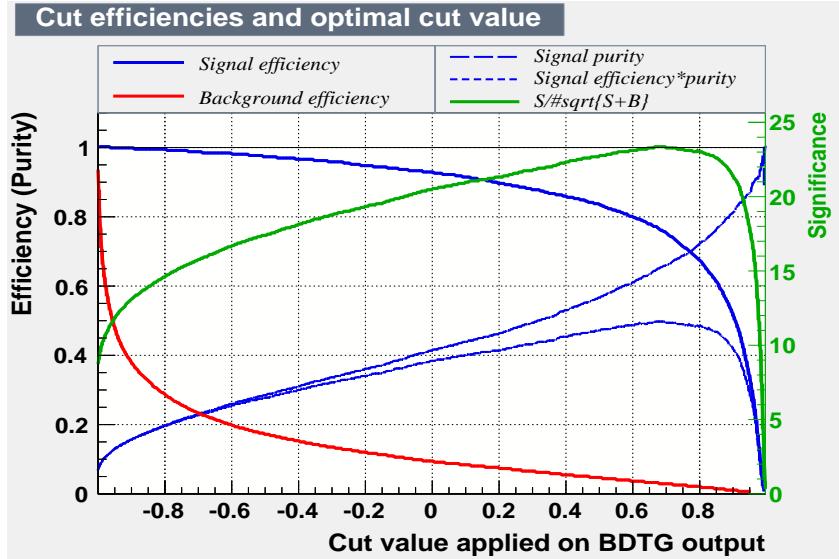


Figure 3.4: Efficiency and purity curves for (blue) signal, (red) background and the (green) FoM curve, as a function of the chosen cut value.

163 **4 Fits to invariant mass distributions of signal and**  
 164 **normalization channel**

165 In order to properly model the invariant mass distribution of  $B_s^0 \rightarrow D_s K \pi \pi$  and  $B_s^0 \rightarrow$   
 166  $D_s \pi \pi \pi$  candidates, the expected signal shape, as well as the expected shape for the  
 167 combinatorial and physical background has to be known. This model can then be used to  
 168 fit the distributions and obtain signal sWeights [6], which are employed to suppress the  
 169 residual background that is still left in the sample, for the time-dependent amplitude fit.

170 **4.1 Signal models for  $m(D_s \pi \pi \pi)$  and  $m(D_s K \pi \pi)$**

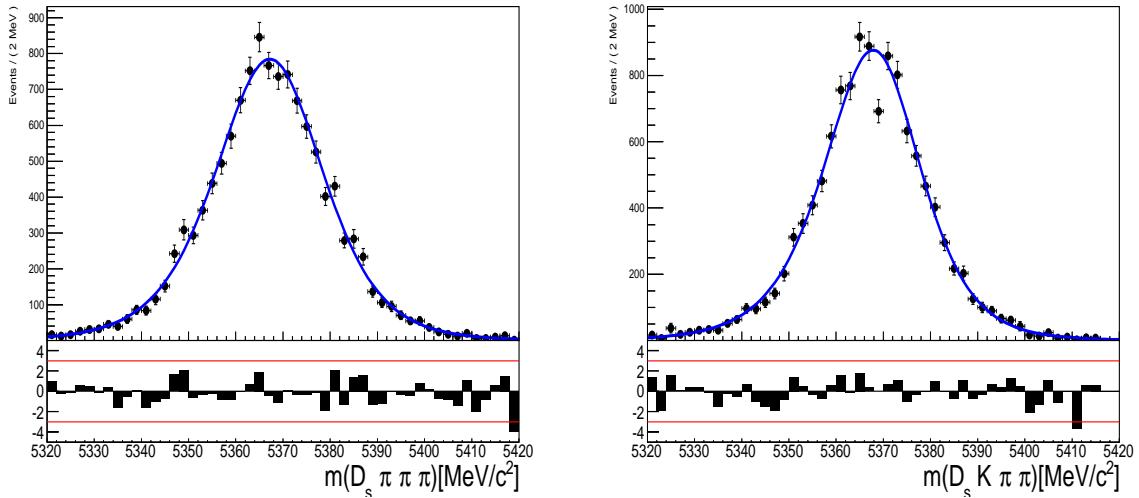


Figure 4.1: Invariant mass distributions of simulated (left)  $B_s^0 \rightarrow D_s \pi \pi \pi$  and (right)  $B_s^0 \rightarrow D_s K \pi \pi$  events. A fit of a RooJohnsonSU function to each distribution is overlaid.

171 The mass distribution of  $B_s^0 \rightarrow D_s K \pi \pi$  signals is modeled using a Johnson SU  
 172 function [7], which is a gaussian function with a Landau-like tail on one side,

$$J(m_{B_s^0}; \mu, \sigma, \gamma, \delta) = \frac{\delta}{\sigma 2\pi \sqrt{1 + (\frac{m_{B_s^0} - \mu}{\sigma})^2}} \exp\left(-\frac{1}{2}[\gamma + \delta \operatorname{Argsh}(\frac{m_{B_s^0} - \mu}{\sigma})^2]\right). \quad (4.1)$$

173 The sign of  $\gamma$  in Eq. 4.1 determines whether the tail is located at lower ( $\gamma > 0$ )  
 174 or higher ( $\gamma < 0$ ) invariant mass values than the mean  $\mu$  of the gaussian function and  
 175  $\delta$  describes the (a)symmetry of the fitted distribution. Higher values of  $\delta$  result in a  
 176 more symmetric, gaussian-like function. Another Johnson SU function function is used  
 177 to account for the contribution of the  $B^0 \rightarrow D_s K \pi \pi$  decay, which is also present in  
 178 the  $m(D_s K \pi \pi)$  spectrum. The width, as well as the tail parameters are fixed to values  
 179 obtained by a fit to the invariant mass distribution of simulated events shown in Fig 4.1.  
 180 A linear scaling factor for the mean  $\mu$  and width  $\sigma$  is floated in the fit to account for  
 181 possible differences between the simulation and real data.

182 The same approach is used to describe the invariant mass distribution of  $B_s^0 \rightarrow D_s \pi \pi \pi$

183 candidates. A Johnson SU function is used to model the signal, the parameters are  
184 determined by a fit to the invariant mass of simulated  $B_s^0 \rightarrow D_s\pi\pi\pi$  decays, shown in  
185 Fig 4.1. A scale factor for the width and the mean is floated to account for differences  
186 between data and MC.

## 187 4.2 Background models for $m(D_s\pi\pi\pi)$

188 Different background sources arise in the invariant mass spectrum of candidates in the  
189 normalization mode.

190 The following backgrounds have to be accounted for:

- 191 • Combinatorial background: This contribution arises from either a real  $D_s$ , which is paired with random tracks to form the  $B_s^0$  candidates, or via real  $X_d$ 's, which are combined with three tracks that fake a  $D_s$  candidate to form a fake  $B_s^0$ .
- 194 • Partially reconstructed  $B^0/B_s^0 \rightarrow D_s^*\pi\pi\pi$  decays, with  $D_s^* \rightarrow D_s\gamma$  or  $D_s^* \rightarrow D_s\pi^0$ , where the  $\gamma/\pi^0$  is not reconstructed in the decay chain.

196 In both cases of combinatorial background, the distribution in the invariant mass of  
197  $B_s^0$  candidates is expected to be smooth and decrease with higher masses. Therefore, one  
198 exponential function is used to model these contributions.

199 The shape of the  $B_s^0 \rightarrow D_s^*\pi\pi\pi$  contribution is expected to be peaking in the  $m(D_s\pi\pi\pi)$   
200 spectrum, with large tails due to the missing momentum, which is carried away by the  $\pi^0$   
201 or  $\gamma$ . The pion or photon from  $D_s^* \rightarrow D_s(\gamma/\pi^0)$  is excluded from the reconstruction. We  
202 model the shape of this contribution using the sum of three bifurcated Gaussian functions.  
203 The shape parameters, as well as the yield of this contribution, are directly determined  
204 on data from a fit to the  $m(D_s\pi\pi\pi)$  invariant mass distribution.

## 205 4.3 Background models for $m(D_sK\pi\pi)$

206 For the signal channel, the following background sources have to be considered:

- 207 • Combinatorial background: same contributions as discussed in Sec. 4.2.
- 208 • Partially reconstructed  $B_s^0 \rightarrow D_s^*K\pi\pi$  decays, with  $D_s^* \rightarrow D_s\gamma$  or  $D_s^* \rightarrow D_s\pi^0$ , where the  $\gamma/\pi^0$  is not reconstructed in the decay chain.
- 210 • Partially reconstructed  $B^0 \rightarrow D_s^*K\pi\pi$  decays, with  $D_s^* \rightarrow D_s\gamma$  or  $D_s^* \rightarrow D_s\pi^0$ , where the  $\gamma/\pi^0$  is not reconstructed in the decay chain.
- 212 • Misidentified  $B_s^0 \rightarrow D_s\pi\pi\pi$  decays, where one of the pions is wrongly identified as a kaon  $\pi \rightarrow K$ .
- 214 • Misidentified, partially reconstructed  $B_s^0 \rightarrow D_s^*\pi\pi\pi$  decays, where one of the pions is wrongly identified as a kaon  $\pi \rightarrow K$  and the  $\gamma/\pi^0$  from  $D_s^* \rightarrow D_s\gamma/\pi^0$  is not reconstructed.

217 The combinatorial background is expected to be non-peaking in the spectrum of the  
218 invariant mass of  $B_s^0 \rightarrow D_sK\pi\pi$  candidates. An exponential function is used to model  
219 this contribution.

220 The shape of the partially reconstructed background without misID is taken from our  
 221 normalization channel, where it can be directly fitted by the sum of three bifurcated  
 222 Gaussian functions as described above. In the signal mass fit, all shape parameters for  
 223 the  $B_s^0 \rightarrow D_s^* K\pi\pi$  background are fixed to the input values from our normalization fit.

224 For the contribution of the  $B^0 \rightarrow D_s^* K\pi\pi$  background, the same shape is used but  
 225 the means  $\mu_i$  of the bifurcated gaussians are shifted down by  $m_{B_s^0} - m_{B^0}$  [?]. The yields  
 226 of both contributions are directly determined in the nominal fit.

227 To determine the shape of misidentified  $B_s^0 \rightarrow D_s\pi\pi\pi$  candidates in the  $m(D_s K\pi\pi)$   
 228 spectrum, we take a truth-matched signal MC sample of our normalization channel. We  
 229 then use the PIDCalib package to determine the  $\pi \rightarrow K$  fake rate. For every candidate  
 230 in our MC sample, a (momentum)  $p$  and (pseudorapidity)  $\eta$ -dependent event weight is  
 231 computed and assigned. We flip the particle hypothesis from pion to kaon for the  $\pi$  with  
 232 the biggest miss-ID weight for each event and recompute the invariant  $B_s^0$  mass. This  
 233 distribution is then modeled using two Crystal Ball functions. The distribution and the  
 234 fit are shown in Fig. 4.2(left).

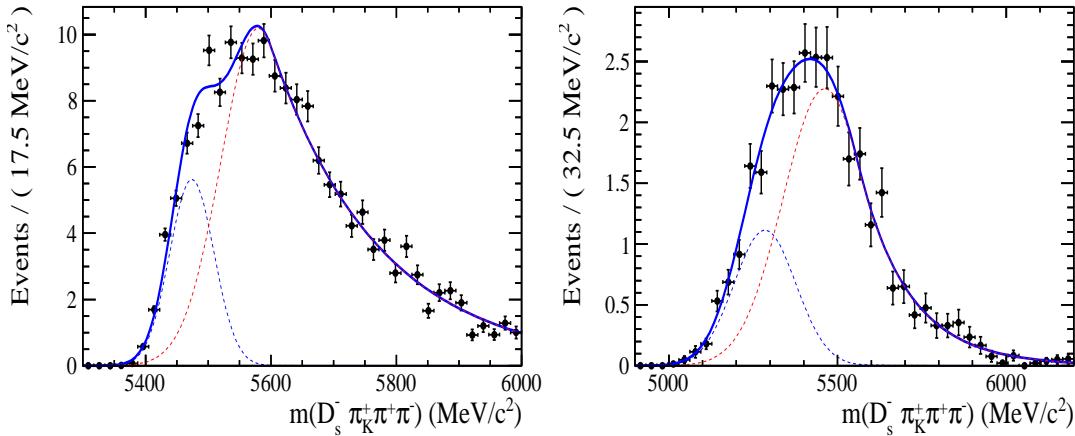


Figure 4.2: Invariant mass distribution of (left) simulated  $B_s^0 \rightarrow D_s\pi\pi\pi$  events, where one of the  $\pi$ 's is reconstructed as a  $K$  and the misID probability for each event is taken into account. The corresponding distribution for simulated  $B_s^0 \rightarrow D_s^*\pi\pi\pi$  events, where the  $\gamma/\pi^0$  from the  $D_s^*$  is excluded from reconstruction, is shown on the right. The solid, black curve on each plot corresponds to the fit consisting of two Crystal Ball functions.

235 The expected yield of misidentified  $B_s^0 \rightarrow D_s\pi\pi\pi$  candidates in the  $m(D_s K\pi\pi)$  spec-  
 236 trum is computed by multiplying the fake probability of  $\propto 3.2\%$ , which is derived from  
 237 PIDCalib, by the yield of  $B_s^0 \rightarrow D_s\pi\pi\pi$  signal candidates, determined in the nominal  
 238 mass fit of our normalization channel.

239 In the same way as mentioned above, we can determine the rate of misidentified, partially  
 240 reconstructed  $B_s^0 \rightarrow D_s^*\pi\pi\pi$  decays in our sample of  $B_s^0 \rightarrow D_s K\pi\pi$  decays using PIDCalib  
 241 and a MC sample of  $B_s^0 \rightarrow D_s^*\pi\pi\pi$  events. The invariant mass distribution we obtain  
 242 when we exclude the  $\gamma/\pi^0$ , flip the the particle hypothesis  $\pi \rightarrow K$  and apply the event  
 243 weights given by the fake rate, is shown in Fig. 4.2 (right). The fit of two Crystal Ball  
 244 functions to this distribution is overlaid. The yield of this contribution is determined  
 245 from the yield of  $B_s^0 \rightarrow D_s^*\pi\pi\pi$  candidates in the nominal mass fit of our normalization  
 246 channel, multiplied by the misID probability of  $\propto 3.6\%$ .

<sup>247</sup> **4.4 Fit to  $B_s^0 \rightarrow D_s\pi\pi\pi$  candidates**

<sup>248</sup> An unbinned maximum likelihood fit is performed simultaneously to the invariant mass  
<sup>249</sup> distribution of  $B_s^0 \rightarrow D_s\pi\pi\pi$  candidates. As discussed in Sec. 4.1, the fit is given  
<sup>250</sup> as a Johnson SU signal model for the  $B_s^0$  and  $B^0$  signal, the sum of three bifurcated  
<sup>251</sup> Gaussian functions to model the partially reconstructed  $B_s^0 \rightarrow D_s^*\pi\pi\pi$  background and  
<sup>252</sup> an Exponential function to account for combinatorial background. The invariant mass  
<sup>253</sup> distribution and the fit is shown in Fig. 4.3. All simultaneously performed fits to the  
<sup>254</sup>  $m(D_s\pi\pi\pi)$  distribution, ordered by the respective  $D_s$  final state, can be found in the  
<sup>255</sup> Appendix ???. The obtained yields are summarized in Table 4.1.

<sup>256</sup> **4.5 Fit to  $B_s^0 \rightarrow D_sK\pi\pi$  candidates**

<sup>257</sup> The shape of the invariant mass distribution of  $B_s^0 \rightarrow D_sK\pi\pi$  candidates is described by  
<sup>258</sup> Johnson SU functions for the  $B^0$  and  $B_s^0$  signal, two sums of three bifurcated Gaussians  
<sup>259</sup> for the  $B_s^0/B^0 \rightarrow D_s^*K\pi\pi$  partially reconstructed background contributions and two  
<sup>260</sup> sums of double Crystal Ball functions for the single misID  $B_s^0 \rightarrow D_s\pi\pi\pi$  and the partially  
<sup>261</sup> reconstructed, misidentified  $B_s^0 \rightarrow D_s^*\pi\pi\pi$  decays. A simultaneous unbinned maximum  
<sup>262</sup> likelihood fit is performed and the result is shown in Fig. 4.3. All simultaneously performed  
<sup>263</sup> fits to the  $m(D_sK\pi\pi)$  distribution, ordered by the respective  $D_s$  final state, can be found  
<sup>264</sup> in the Appendix ???. The obtained yields are summarized in Table 4.1.

<sup>265</sup> **4.6 Extraction of signal weights**

<sup>266</sup> The sPlot technique [6] is used to extract signal weights from the fits to the invariant  
<sup>267</sup> mass distributions of our signal and normalization channel. This statistical tool assigns  
<sup>268</sup> a weight to every event, according to its position in the respective mass distribution,  
<sup>269</sup> given the fitted signal and background models. The weights can then be used to suppress  
<sup>270</sup> the background components in every other observable distribution of interest. Figure 4.4  
<sup>271</sup> shows the distribution of weights across the invariant mass spectra of  $B_s^0 \rightarrow D_s\pi\pi\pi$  and  
<sup>272</sup>  $B_s^0 \rightarrow D_sK\pi\pi$  candidates.

fit component	yield 2011	yield 2012	yield 2015	yield 2016
$m(D_sK\pi\pi)$				
$B_s^0 \rightarrow D_sK\pi\pi$	$392 \pm 25$	$860 \pm 38$	$309 \pm 21$	$1984 \pm 55$
$B^0 \rightarrow D_sK\pi\pi$	$276 \pm 26$	$692 \pm 41$	$261 \pm 23$	$1385 \pm 58$
$B^0/B_s^0 \rightarrow D_s^*K\pi\pi$	$7 \pm 25$	$171 \pm 75$	$114 \pm 25$	$893 \pm 84$
$B_s^0 \rightarrow D_s^{(*)}\pi\pi\pi$	$63 \pm 0$	$158 \pm 0$	$53 \pm 0$	$314 \pm 0$
combinatorial	$1482 \pm 53$	$2884 \pm 100$	$605 \pm 43$	$4261 \pm 133$
$m(D_s\pi\pi\pi)$				
$B_s^0 \rightarrow D_s\pi\pi\pi$	$9183 \pm 105$	$22083 \pm 166$	$7574 \pm 95$	$43773 \pm 245$
$B^0 \rightarrow D_s\pi\pi\pi$	$289 \pm 58$	$716 \pm 95$	$229 \pm 54$	$968 \pm 147$
$B_s^0 \rightarrow D_s^*\pi\pi\pi$	$3640 \pm 130$	$9086 \pm 232$	$3047 \pm 110$	$17827 \pm 421$
combinatorial	$4991 \pm 154$	$11127 \pm 271$	$3728 \pm 126$	$24589 \pm 500$

Table 4.1: Summary of yields obtained from the fits to Run1 and Run2 data.

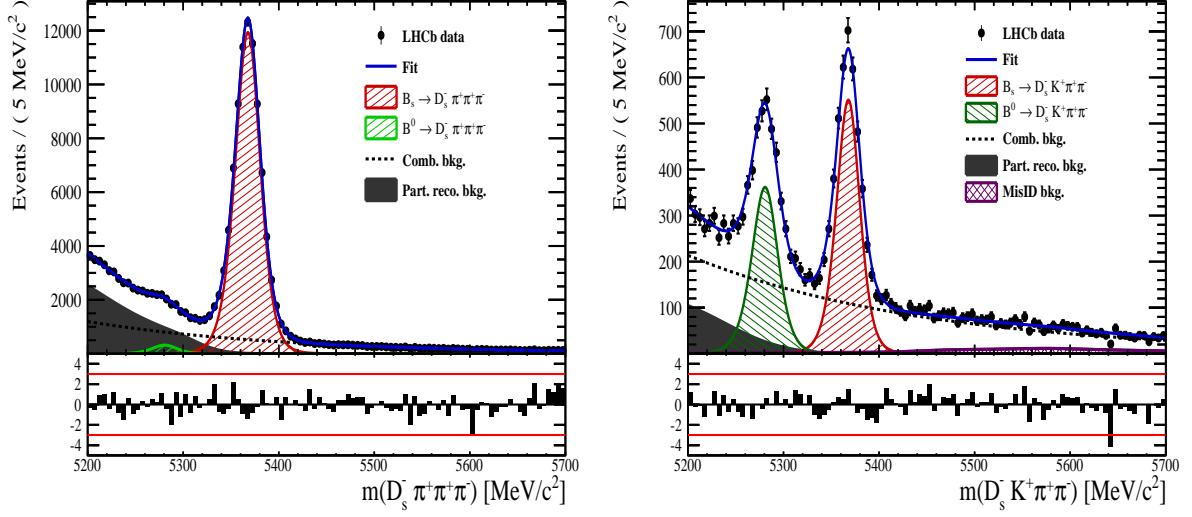


Figure 4.3: Invariant mass distribution of (left)  $B_s^0 \rightarrow D_s \pi \pi \pi$  and (right)  $B_s^0 \rightarrow D_s K \pi \pi$  candidates for Run1 and Run2 data. The respective fit described in the text is overlaid.

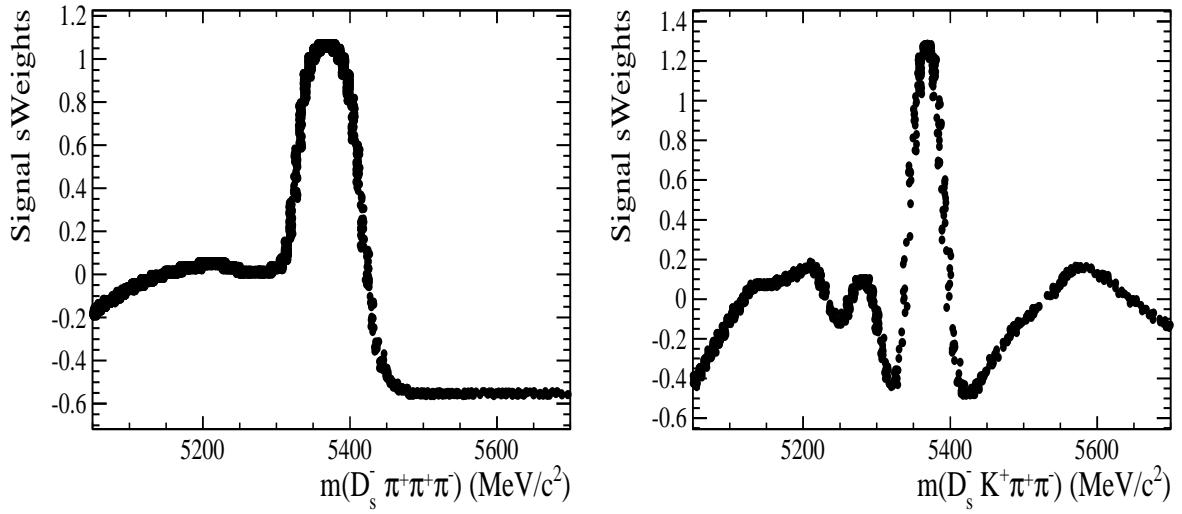


Figure 4.4: Distribution of sWeights across the invariant mass of (left)  $B_s^0 \rightarrow D_s \pi \pi \pi$  and (right)  $B_s^0 \rightarrow D_s K \pi \pi$  candidates for Run1 and Run2 data.

## 273 5 Flavour Tagging

274 To successfully perform a time- and amplitude-dependent measurement of  $\gamma$ , the identifi-  
 275 cation of the initial state flavour of the  $B_s^0$  meson is crucial. In the presented analysis,  
 276 a number of flavour tagging algorithms are used that either determine the flavour of  
 277 the non-signal b-hadron produced in the event (opposite site, OS), or they use particles  
 278 produced in the fragmentation of the signal candidate  $B_s^0/\bar{B}_s^0$  (same side, SS).  
 279 For the same side, the algorithm searching for the charge of an additional kaon that accom-  
 280 panies the fragmentation of the signal candidate is used (SS-nnetKaon). For the opposite  
 281 site, four different taggers are chosen: The Two algorithms that use the charge of an  
 282 electron or a muon from semileptonic B decays (OS-  $e,\mu$ ), the tagger that uses the charge  
 283 of a kaon from a  $b \rightarrow c \rightarrow s$  decay chain (OS-nnetKaon) and the algorithm that determines  
 284 the  $B_s^0/\bar{B}_s^0$  candidate flavour from the charge of a secondary vertex, reconstructed from  
 285 the OS b decay product (OS-VtxCharge). All four taggers are then combined into a signal  
 286 OS tagger.  
 287 Every single tagging algorithm is prone to misidentify the signal candidate at a certain  
 288 mistag rate  $\omega = (\text{wrongtags})/(\text{alltags})$ . This might be caused by particle misidentifica-  
 289 tion, flavour oscillation of the neutral opposite site B-meson or by tracks that are wrongly  
 290 picked up from the underlying event. For every signal  $B_s^0/\bar{B}_s^0$  candidate, each tagging  
 291 algorithm predicts a mistag probability  $\eta$ , which is calculated using a combination of  
 292 inputs such as the kinematics of the tagging particles. The inputs are then combined  
 293 to a predicted mistag using neural networks. These are trained on simulated samples  
 294 of  $B_s^0 \rightarrow D_s^- \pi^+$  (SS algorithm) and  $B^+ \rightarrow J/\psi K^+$  (OS algorithms) decays. For the  
 295 presented analysis, the measurable CP-violating coefficients are damped by the tagging  
 296 dilution  $D$ , that depends on the mistag rate:

$$297 \quad D = 1 - 2\omega. \quad (5.1)$$

298 This means that the statistical precision, with which these coefficients can be measured,  
 299 scales as the inverse square root of the effective tagging efficiency,

$$\epsilon_{eff} = \epsilon_{tag}(1 - 2\omega)^2, \quad (5.2)$$

300 where  $\epsilon_{tag}$  is the fraction of events that have a tagging decision. The flavour  
 301 tagging algorithms are optimised for highest  $\epsilon_{eff}$  on data, using the  $B_s^0 \rightarrow D_s^- \pi^+$  and  
 302  $B^+ \rightarrow J/\psi K^+$  samples.

303 Utilizing flavour-specific final states, the predicted mistag  $\eta$  of each tagger has to be  
 304 calibrated to match the observed mistag  $\omega$  on the data sample. For the calibration, a  
 linear model of the form

$$\omega(\eta) = p_0 + p_1 \cdot (\eta - \langle \eta \rangle), \quad (5.3)$$

305 where the values of  $p_0$  and  $p_1$  are determined using the  $B_s^0 \rightarrow D_s \pi \pi \pi$  normalization  
 306 mode and  $\langle \eta \rangle$  is the average estimated mistag probability  $\langle \eta \rangle = \sum_{i=1}^{N_{cand}} (\eta_i)/N_{cand}$ .  
 307 Following this model, a perfectly calibrated tagger would lead to  $\omega(\eta) = \eta$  and one would  
 308 expect  $p_1 = 1$  and  $p_0 = \langle \eta \rangle$ . Due to the different interaction cross-sections of oppositely  
 309 charged particles, the tagging calibration parameters depend on the initial state flavour of  
 310 the  $B_s^0$ . Therefore, the flavour asymmetry parameters  $\Delta p_0$ ,  $\Delta p_1$  and  $\Delta \epsilon_{tag}$  are introduced.  
 311 For this analysis, the calibrated mistag is treated as per-event variable, giving a larger

weight to events that are less likely to have an incorrect tag. This adds one additional observable to the time- and amplitude-dependent fit.  
 The tagging calibration is determined using a time-dependent fit to the full  $B_s^0 \rightarrow D_s\pi\pi\pi$  sample, where the mixing frequency  $\Delta m_s$  is fixed to the nominal PDG value [8]. The calibration procedure for the OS tagging algorithms (Sec.5.1) and the SS kaon tagger (Sec.5.2) is applied on the full Run I and 2015 and 2016 Run II  $B_s^0 \rightarrow D_s\pi\pi\pi$  data sample, which is selected following the steps described in Sec. 3. The similar selection ensures as close as possible agreement between the  $B_s^0 \rightarrow D_s\pi\pi\pi$  and  $B_s^0 \rightarrow D_sK\pi\pi$  samples in terms of the decay kinematics, which are crucial for the flavour tagging. Section 5.3 shows the compatibility of both samples. After applying the calibration, the response of the OS and SS taggers are combined, which is shown in Sec. 5.4.

## 5.1 OS tagging calibration

The responses of the OS electron, muon, neural net kaon and the secondary vertex charge taggers are combined for the mistag calibration. Figure ?? shows the distribution of the predicted OS mistag for signal candidates from  $B_s^0 \rightarrow D_s\pi\pi\pi$ . The extracted calibration parameters and tagging asymmetries are summarized in Table 5.1 and the measured tagging power for the OS combination is  $\epsilon_{eff,OS} = 4.81\%$ .

$p_0$	$p_1$	$<\eta>$	$\epsilon_{tag}$	$\Delta p_o$	$\Delta p_1$	$\epsilon_{eff} [\%]$
$0.025 \pm 0.005$	$0.944 \pm 0.048$	$0.347$	$0.517 \pm 0.002$	$0.028 \pm 0.005$	$0.037 \pm 0.045$	$4.81 \pm 0.04 \text{ (stat)} \pm 0.37 \text{ (cal)}$

Table 5.1: Calibration parameters and tagging asymmetries of the OS tagger extracted from  $B_s^0 \rightarrow D_s\pi\pi\pi$  decays.

## 5.2 SS tagging calibration

The SS neural net kaon tagger can be calibrated using the flavour-specific  $B_s^0 \rightarrow D_s\pi\pi\pi$  decay. It's development, performance and calibration is described in detail in [9]. Figure ?? shows the distribution of the predicted mistag of the neural net kaon tagger. The extracted calibration parameters and tagging asymmetries are summarized in Table 5.2 and the measured tagging power for this algorithm is  $\epsilon_{eff,SS} = 3.22\%$ .

$p_0$	$p_1$	$<\eta>$	$\epsilon_{tag}$	$\Delta p_o$	$\Delta p_1$	$\epsilon_{eff} [\%]$
$0.008 \pm 0.004$	$1.086 \pm 0.059$	$0.381$	$0.571 \pm 0.002$	$-0.017 \pm 0.004$	$0.135 \pm 0.058$	$3.22 \pm 0.03 \text{ (stat)} \pm 0.26 \text{ (cal)}$

Table 5.2: Calibration parameters and tagging asymmetries of the SS tagger extracted from  $B_s^0 \rightarrow D_s\pi\pi\pi$  decays.

## 5.3 Tagging performance comparison between the signal and normalization channel

To justify the usage of the tagging calibration, obtained using the  $B_s^0 \rightarrow D_s\pi\pi\pi$  sample, for our signal decay, the performance of the taggers in the two decay channels needs to be compatible. This is verified using both, simulated signal samples of both decays and

sweighted data, to compare the similarity of the mistag probabilities, tagging decisions and kinematic observables that are correlated with the tagging response, on simulation and data.

The distributions of the predicted mistag probability  $\eta$  for the OS combination and the SS kaon tagger are shown in Fig. ?? (simulation) and Fig. 5.1 (data).

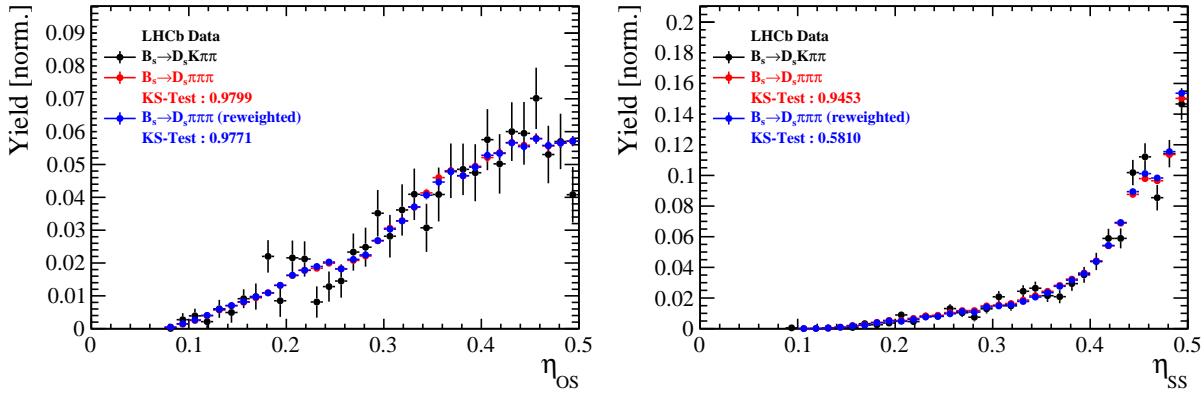


Figure 5.1: Distributions of the predicted mistag  $\eta$  for the OS combination (left) and the SS kaon tagger (right) for signal candidates in the  $B_s^0 \rightarrow D_s K\pi\pi$  (black) and  $B_s^0 \rightarrow D_s \pi\pi\pi$  (red) data samples.

Both, data and simulated samples, show good agreement between the signal and normalization channel. Compatibility is also seen in Fig. ??, which shows the comparison of the tagging decision distributions of the OS and SS tagger for sweighted data.

Fig. ?? shows the signal data distributions of the transverse  $B_s^0$  momentum  $p_T$ , the pseudorapidity  $\eta$  of the signal candidate and the number of reconstructed tracks per event. Sufficient agreement is observed.

To justify the portability of the flavour tagging calibration obtained from  $B_s^0 \rightarrow D_s \pi\pi\pi$  to the  $B_s^0 \rightarrow D_s K\pi\pi$  channel, besides the good agreement of the distributions shown above, the dependence of the measured mistag  $\omega$  on the predicted mistag  $\eta$  has to be compatible in both channel. This dependence is shown in Fig. 5.2 for simulated signal events of both channels, where good agreement is observed.

## 5.4 Combination of OS and SS taggers

In the time- and amplitude-dependent fit to  $B_s^0 \rightarrow D_s K\pi\pi$  data, the obtained tagging responses of the OS and SS tagger will be combined after the calibration described in the previous sections is applied. Events that acquire a mistag probability greater than 0.5 after the calibration will have their tagging decision flipped. For events where only one of the two taggers fired, the combination of the tagging decision is trivial. In those events where both taggers made a decision, we use the standard combination of taggers [10] provided by the flavour tagging group. In the nominal fit, the calibrated mistags  $\omega$  are combined event by event for the OS and SS tagger, thus adding one variable to observable to the fit procedure. This ensures that the uncertainties of the OS and SS tagging calibration parameters are propagated properly to the combined tagging response for each event. The tagging performance for the combined tagger in the categories SS tagged only, OS

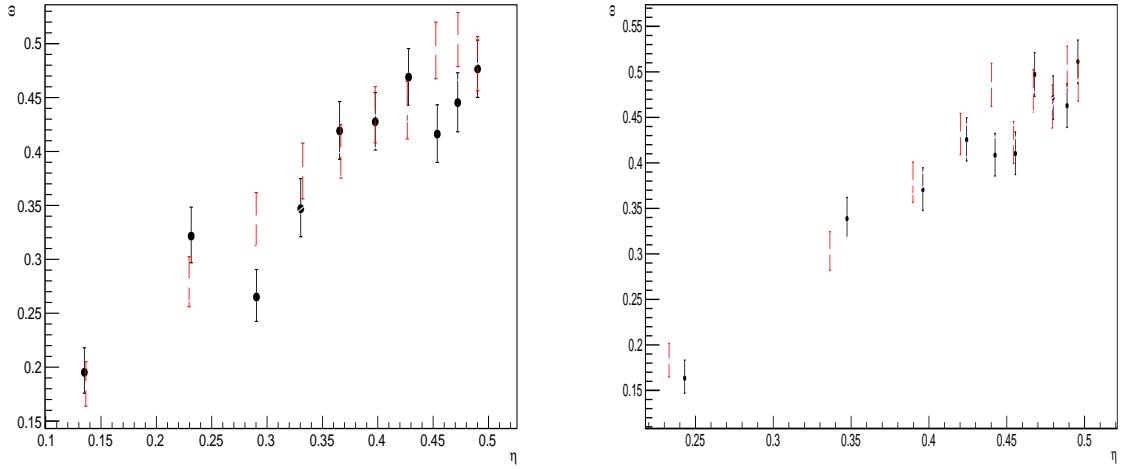


Figure 5.2: Dependence of the observed mistag  $\omega$  on the predicted mistag  $\eta$  for the (left) OS combination and the (right) SS kaon tagger, found in the simulated  $B_s^0 \rightarrow D_s K\pi\pi$  (black) and  $B_s^0 \rightarrow D_s \pi\pi\pi$  (red) signal samples.

368 tagged only and SS+OS tagged, are shown in Tab. 5.4 for the signal and normalization  
 369 channel. The distribution of the observed mistag  $\omega$  as a function of the combined mistag  
 370 probability  $\eta$  for  $B_s^0 \rightarrow D_s \pi\pi\pi$  decays is shown in Fig. 5.3.

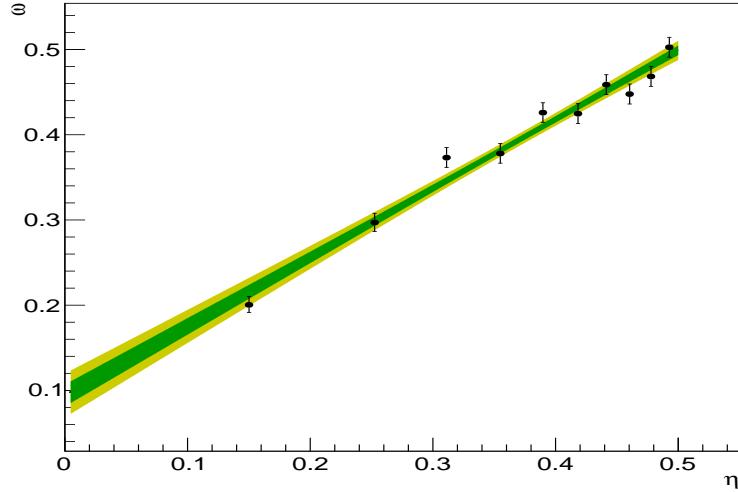


Figure 5.3: Distribution of the predicted combined mistag probability  $\eta$  versus the observed mistag  $\omega$  for  $B_s^0 \rightarrow D_s \pi\pi\pi$  signal candidates. The fit with a linear polynomial, used to determine  $p_0$  and  $p_1$  is overlaid.

$B_s^0 \rightarrow D_s\pi\pi\pi$		$\epsilon_{tag}$	$\epsilon_{eff}$
SS only		$(28.586 \pm 0.165)\%$	$(1.408 \pm 0.018(\text{stat}) \pm 0.082(\text{cal}))\%$
OS only		$(17.221 \pm 0.138)\%$	$(2.027 \pm 0.029(\text{stat}) \pm 0.100(\text{cal}))\%$
SS+OS		$(39.981 \pm 0.179)\%$	$(5.690 \pm 0.047(\text{stat}) \pm 0.196(\text{cal}))\%$
total			
$B_s^0 \rightarrow D_sK\pi\pi$		$\epsilon_{tag}$	$\epsilon_{eff}$
SS only		$(30.094 \pm 0.960)\%$	$(1.379 \pm 0.082(\text{stat}) \pm 0.085(\text{cal}))\%$
OS only		$(18.923 \pm 0.819)\%$	$(1.768 \pm 0.121(\text{stat}) \pm 0.099(\text{cal}))\%$
SS+OS		$(27.277 \pm 0.932)\%$	$(3.914 \pm 0.194(\text{stat}) \pm 0.220(\text{cal}))\%$
total			

Table 5.3: Flavour tagging performances for  $B_s^0 \rightarrow D_s\pi\pi\pi$  and  $B_s^0 \rightarrow D_sK\pi\pi$  events which are only OS tagged, only SS tagged or tagged by both.

## 371 6 Decay-time acceptance

372 The decay-time distribution of the  $B_s^0$  mesons is sculpted due to the geometry of the LHCb  
 373 detector and the applied selection cuts, which are described in Section 3. In particular,  
 374 any requirement on the flight distance (FD), the impact parameter (IP) or the direction  
 375 angle (DIRA) of the  $B_s^0$  mesons, as well as the direct cut on the lifetime, will lead to a  
 376 decay-time dependent efficiency  $a(t)$ . This efficiency will distort the theoretically expected,  
 377 time-dependent decay rate

$$\frac{\Gamma(t)^{observed}}{dt} = \frac{\Gamma(t)^{theory}}{dt} \cdot a(t), \quad (6.1)$$

378 and has to be modelled correctly, in order to describe the observed decay rate. We  
 379 use our control channel for this measurement, because for  $B_s^0 \rightarrow D_s K\pi\pi$  decays the  
 380 decay-time acceptance is correlated with the CP-observables which we aim to measure.  
 381 Therefore, floating the CP-observables and the acceptance shape at the same time is  
 382 not possible. Hence, a fit to the decay-time distribution of  $B_s^0 \rightarrow D_s \pi\pi\pi$  candidates is  
 383 performed and the obtained acceptance shape is corrected by the difference in shape found  
 384 for the  $B_s^0 \rightarrow D_s K\pi\pi$  and  $B_s^0 \rightarrow D_s \pi\pi\pi$  MC.

385 A PDF of the form

$$\mathcal{P}(t', \vec{\lambda}) = \left[ (e^{\Gamma_s t} \cdot \cosh(\frac{\Delta\Gamma_s t}{2}) \times \mathcal{R}(t - t')) \cdot \epsilon(t', \vec{\lambda}) \right], \quad (6.2)$$

386 is fit to the decay time distribution of  $B_s^0 \rightarrow D_s \pi\pi\pi$  candidates in data. Since the  
 387 fit is performed untagged, the PDF shown in Eq. 6.2 contains no terms proportional  
 388 to  $\Delta m_s$ . The values for  $\Gamma_s$  and  $\Delta\Gamma_s$  are fixed to the latest HFAG results [11]. The  
 389 decay-time acceptance  $\epsilon(t', \vec{\lambda})$  is modelled using the sum of cubic polynomials  $v_i(t)$ , so  
 390 called Splines [12]. The polynomials are parametrised by so-called knots which determine  
 391 their boundaries. Knots can be set across the fitted distribution to account for local  
 392 changes in the acceptance shape. Using more knots is equivalent to using more base  
 393 splines which are defined on a smaller sub-range. In total,  $n + 2$  base splines  $v_i(t)$  are  
 394 needed to describe an acceptance shape which is parametrised using  $n$  knots.

395 For fits shown in the following, the knots have been placed at  $t =$   
 396  $[0.5, 1.0, 1.5, 2.0, 3.0, 9.5] ps$ . To accommodate these 6 knot positions, 8 basic splines  
 397  $v_i$ ,  $i = [1, \dots, 8]$  are used. Since a rapid change of the decay time acceptance at low  
 398 decay times due to the turn-on effect generated by the lifetime and other selection cuts is  
 399 expected, more knots are placed in that regime. At higher decay times we expect linear  
 400 behavior, with a possible small effect due to the VELO reconstruction. Therefore fewer  
 401 knots are used. Furthermore,  $v_7$  is fixed to 1 in order to normalize the overall acceptance  
 402 function. To stabilise the last spline,  $v_8$  is fixed by a linear extrapolation from the two  
 403 previous splines:

$$v_N = v_{N-1} + \frac{v_{N-2} - v_{N-1}}{t_{N-2} - t_{N-1}} \cdot (t_N - t_{N-1}). \quad (6.3)$$

404 Here,  $N = 8$  and  $t_{N-1}$  corresponds to the knot position associated with  $v_{N-1}$ .

## 405 6.1 Comparison of acceptance in subsamples

406 It is possible that the decay-time dependent efficiency deviates in different subsamples of  
 407 our data. In particular, the acceptance could differentiate in subsamples with different  
 408 final state kinematics, such as the run I & run II sample, the various  $D_s$  final states and  
 409 the ways an event is triggered at the L0 stage. To investigate possible deviations, the  
 410 full selected  $B_s^0 \rightarrow D_s\pi\pi\pi$  sample is split into subsamples according to the categories  
 411 mentioned above (run,  $D_s$  state, L0 trigger). For each subsample, the fit procedure  
 412 described at the beginning of this chapter, using the pdf given by Eq. 6.2, is repeated  
 413 and the obtained values for the spline coefficients  $v_i$  are compared. Figure 6.1 shows the  
 414 comparison of the obtained spline coefficients for the different  $D_s$  final states.

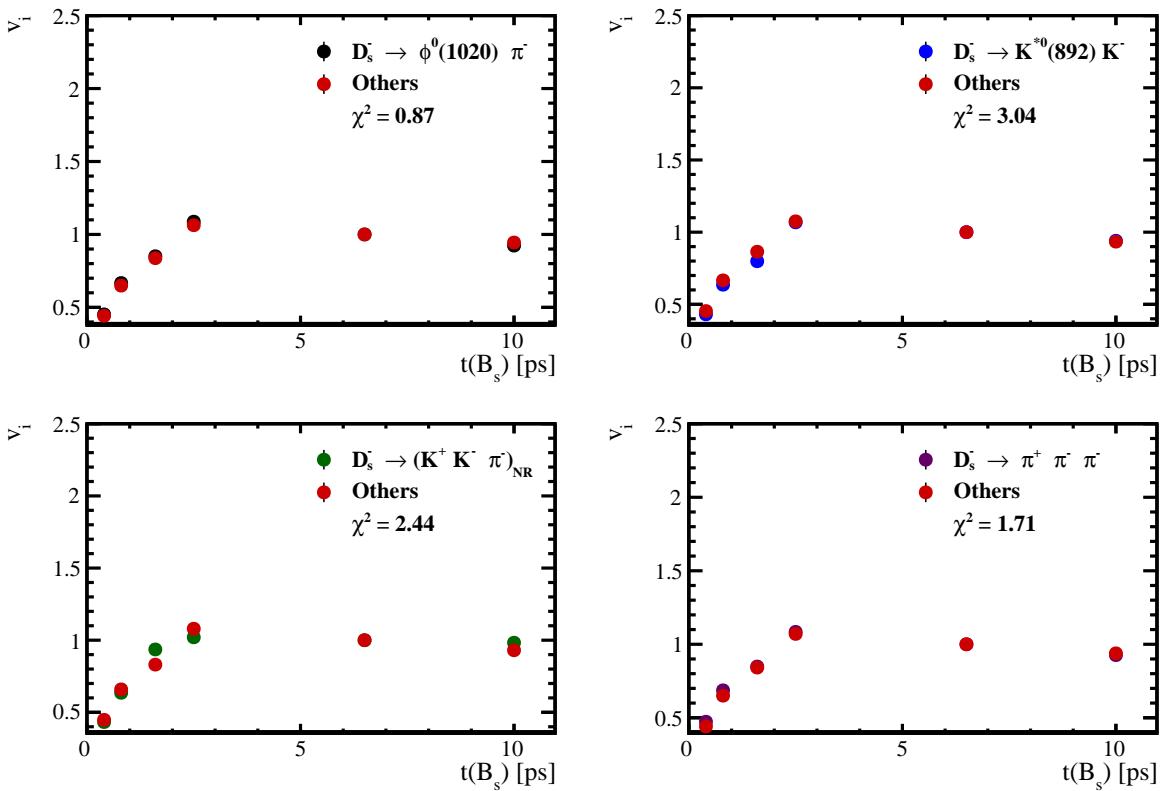


Figure 6.1: Comparison of the spline coefficients obtained from time-dependent fits to the  $B_s^0 \rightarrow D_s\pi\pi\pi$  subsamples of different  $D_s$  final states. The comparison of one particular  $D_s$  state against all other states is shown.

415 Investigating the obtained spline coefficients from different  $D_s$  final states, good  
 416 agreement is observed between all four channels and no need to distinguish between  
 417 different final states in the time-dependent amplitude fit is found.

418 The comparison between spline coefficients for the different runs and L0 trigger categories  
 419 is shown in Figure 6.2.

420 Significant deviations between spline coefficients obtained from the two different runs  
 421 and L0 trigger categories can be observed. The deviations are most pronounced in the  
 422 (0 – 5) ps region, where the majority of statistics is found. Therefore, the time-dependent  
 423 efficiency has to be treated separately for the runs and L0 categories. This is achieved by

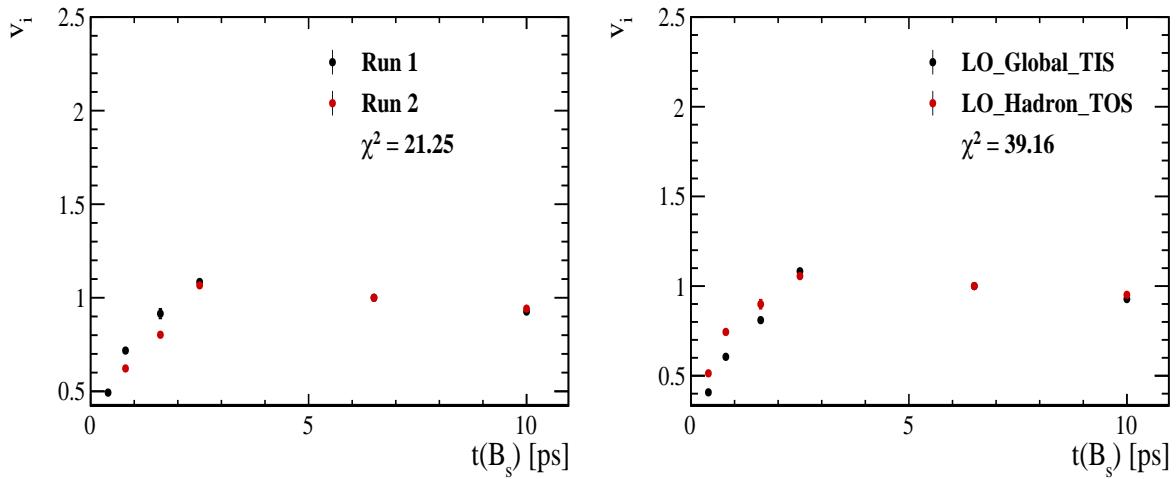


Figure 6.2: Comparison of the spline coefficients obtained from time-dependent fits to the  $B_s^0 \rightarrow D_s\pi\pi\pi$  subsamples of (left) the different runs and (right) L0 trigger categories.

<sup>424</sup> implementing a simultaneous fit, where the acceptance description is allowed to vary in  
<sup>425</sup> the subsamples.

## 6.2 Results

The nominal fit to  $B_s^0 \rightarrow D_s\pi\pi\pi$  data using this configuration is shown in Figure 6.3. Note that the normalization of the splines in the following figures is not in scale. The fit parameters obtained from the described fits to data and simulation are summarised in Table 6.5.

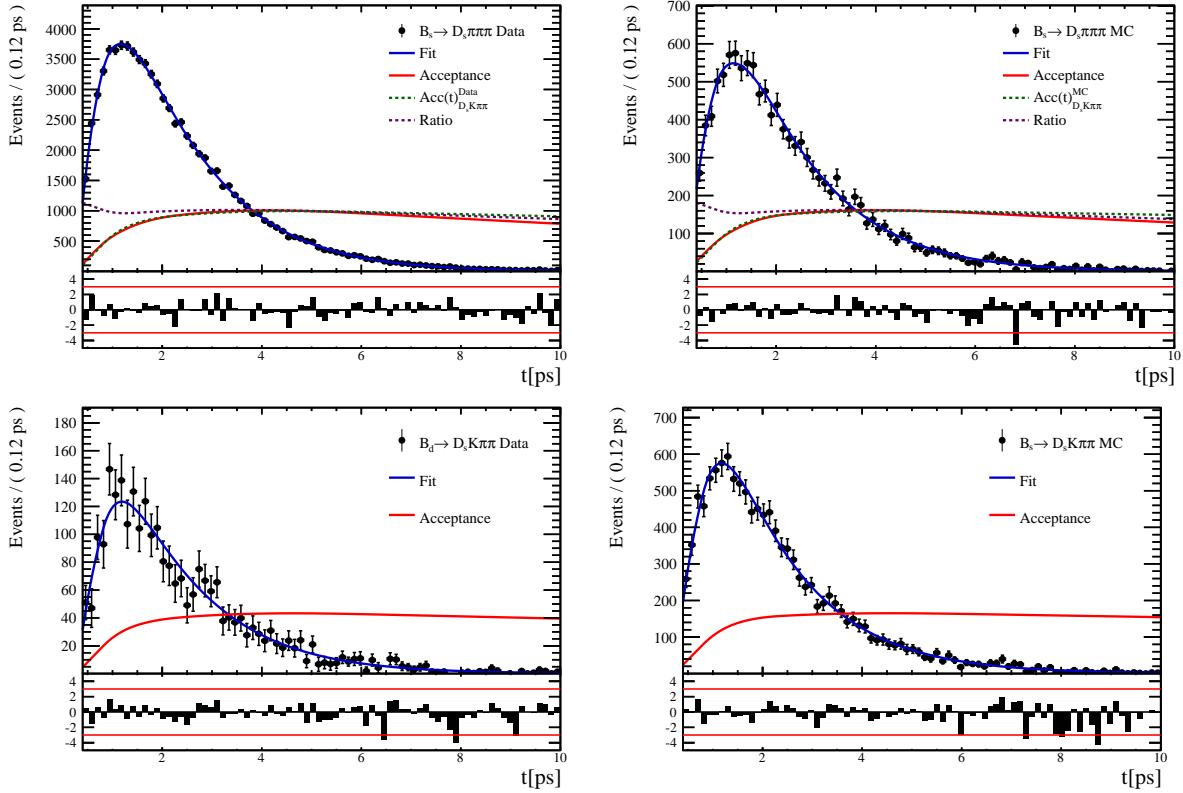


Figure 6.3: The red line shows the spline function describing the acceptance and the blue line depicts the total fit function.

Table 6.1: Summary of the obtained parameters from the acceptance fit

Knot position	Coefficient	$B_s^0 \rightarrow D_s K\pi\pi$ data	$B_s^0 \rightarrow D_s K\pi\pi$ MC	Ratio
0.4	$v_0$	$0.446 \pm 0.034$	$0.479 \pm 0.022$	$1.020 \pm 0.078$
0.8	$v_1$	$0.673 \pm 0.051$	$0.702 \pm 0.034$	$0.943 \pm 0.069$
1.6	$v_2$	$0.874 \pm 0.076$	$0.915 \pm 0.054$	$0.984 \pm 0.084$
2.5	$v_3$	$1.028 \pm 0.043$	$1.021 \pm 0.038$	$1.042 \pm 0.043$
6.5	$v_4$	$1.000 \pm 0.000$	$1.000 \pm 0.000$	$1.000 \pm 0.000$
10.0	$v_5$	$0.975 \pm 0.000$	$0.982 \pm 0.000$	$0.963 \pm 0.000$

Table 6.2: Summary of the obtained parameters from the acceptance fit

Knot position	Coefficient	$B_s^0 \rightarrow D_s K\pi\pi$ data	$B_s^0 \rightarrow D_s K\pi\pi$ MC	Ratio
0.4	$v_0$	$0.425 \pm 0.039$	$0.456 \pm 0.025$	$1.054 \pm 0.091$
0.8	$v_1$	$0.626 \pm 0.058$	$0.673 \pm 0.039$	$1.028 \pm 0.095$
1.6	$v_2$	$0.936 \pm 0.105$	$0.887 \pm 0.067$	$0.955 \pm 0.102$
2.5	$v_3$	$1.081 \pm 0.066$	$1.050 \pm 0.049$	$1.019 \pm 0.069$
6.5	$v_4$	$1.000 \pm 0.000$	$1.000 \pm 0.000$	$1.000 \pm 0.000$
10.0	$v_5$	$0.929 \pm 0.000$	$0.956 \pm 0.000$	$0.983 \pm 0.000$

Table 6.3: Summary of the obtained parameters from the acceptance fit

Knot position	Coefficient	$B_s^0 \rightarrow D_s K\pi\pi$ data	$B_s^0 \rightarrow D_s K\pi\pi$ MC	Ratio
0.4	$v_0$	$0.615 \pm 0.080$	$0.525 \pm 0.039$	$0.935 \pm 0.106$
0.8	$v_1$	$0.922 \pm 0.124$	$0.751 \pm 0.060$	$0.842 \pm 0.110$
1.6	$v_2$	$0.940 \pm 0.123$	$0.959 \pm 0.083$	$1.025 \pm 0.126$
2.5	$v_3$	$0.940 \pm 0.077$	$0.949 \pm 0.062$	$1.111 \pm 0.059$
6.5	$v_4$	$1.000 \pm 0.000$	$1.000 \pm 0.000$	$1.000 \pm 0.000$
10.0	$v_5$	$1.053 \pm 0.000$	$1.045 \pm 0.000$	$0.903 \pm 0.000$

Table 6.4: Summary of the obtained parameters from the acceptance fit

Knot position	Coefficient	$B_s^0 \rightarrow D_s K\pi\pi$ data	$B_s^0 \rightarrow D_s K\pi\pi$ MC	Ratio
0.4	$v_0$	$0.385 \pm 0.019$	$0.478 \pm 0.018$	$1.014 \pm 0.056$
0.8	$v_1$	$0.584 \pm 0.028$	$0.695 \pm 0.029$	$0.975 \pm 0.055$
1.6	$v_2$	$0.802 \pm 0.067$	$0.919 \pm 0.052$	$0.967 \pm 0.078$
2.5	$v_3$	$1.032 \pm 0.048$	$1.021 \pm 0.038$	$1.045 \pm 0.047$
6.5	$v_4$	$1.000 \pm 0.000$	$1.000 \pm 0.000$	$1.000 \pm 0.000$
10.0	$v_5$	$0.972 \pm 0.000$	$0.982 \pm 0.000$	$0.960 \pm 0.000$

Table 6.5: Summary of the obtained parameters from the acceptance fit

Knot position	Coefficient	$B_s^0 \rightarrow D_s K\pi\pi$ data	$B_s^0 \rightarrow D_s K\pi\pi$ MC	Ratio
0.4	$v_0$	$0.481 \pm 0.032$	$0.479 \pm 0.020$	$1.015 \pm 0.061$
0.8	$v_1$	$0.711 \pm 0.050$	$0.698 \pm 0.031$	$0.962 \pm 0.067$
1.6	$v_2$	$0.905 \pm 0.079$	$0.918 \pm 0.052$	$0.972 \pm 0.079$
2.5	$v_3$	$1.002 \pm 0.061$	$1.020 \pm 0.041$	$1.045 \pm 0.055$
6.5	$v_4$	$1.000 \pm 0.000$	$1.000 \pm 0.000$	$1.000 \pm 0.000$
10.0	$v_5$	$0.998 \pm 0.000$	$0.982 \pm 0.000$	$0.961 \pm 0.000$

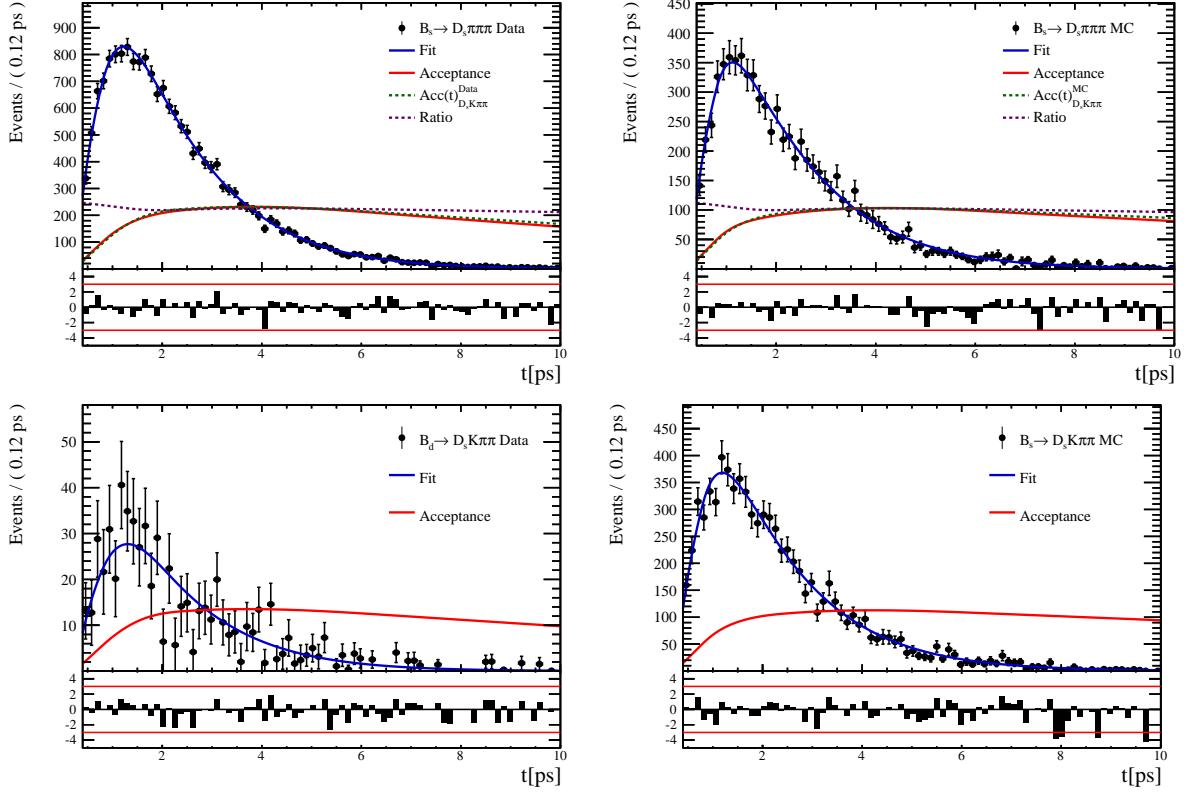


Figure 6.4

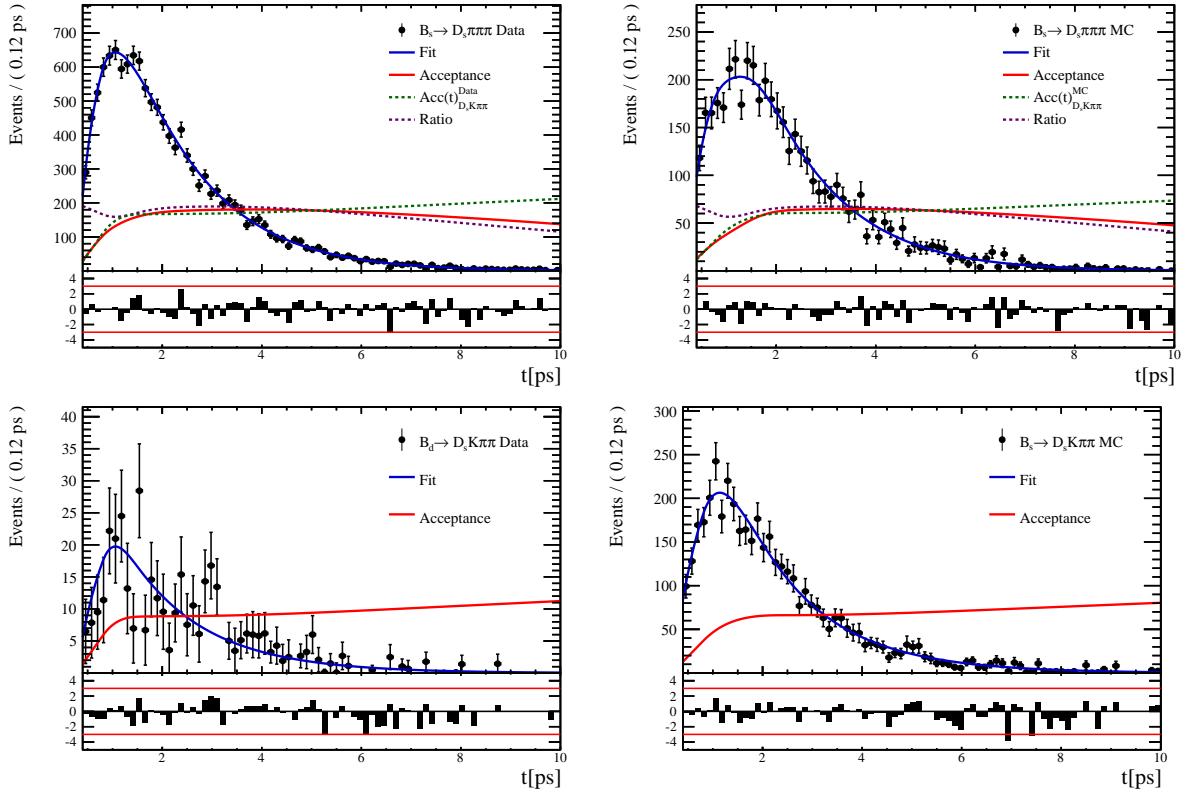


Figure 6.5:

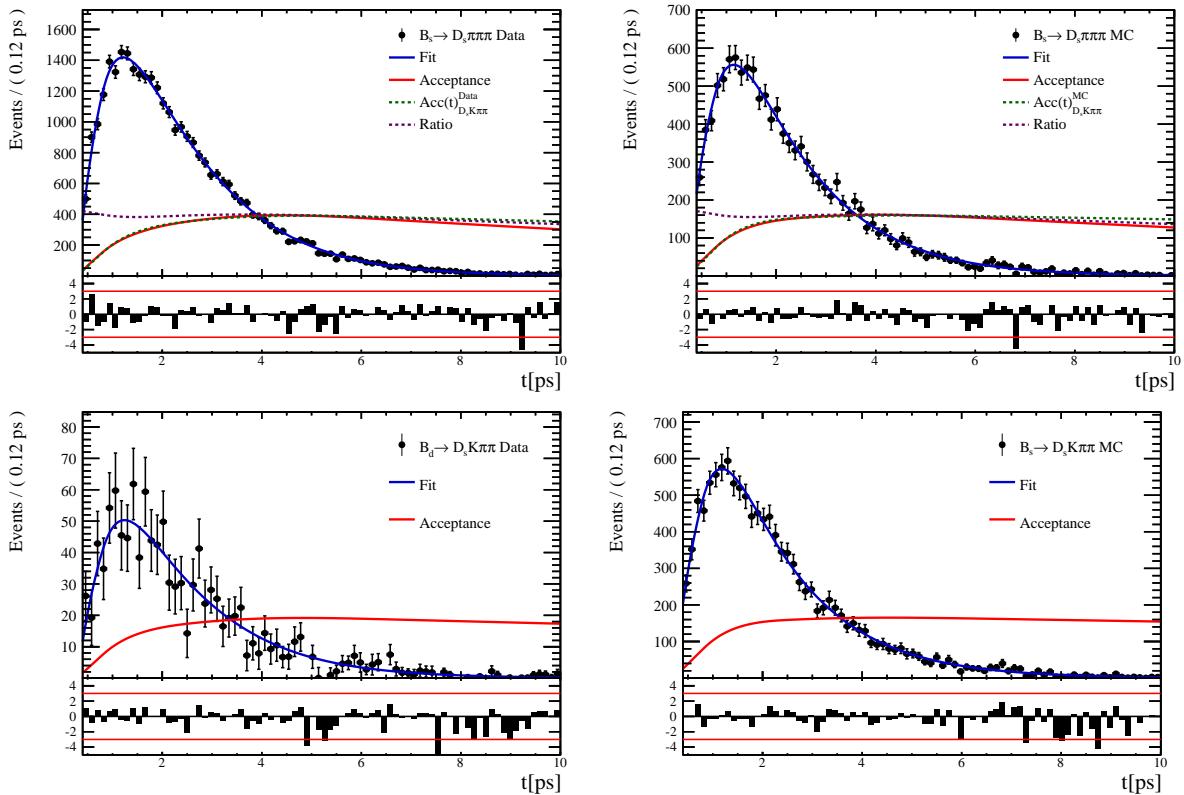


Figure 6.6

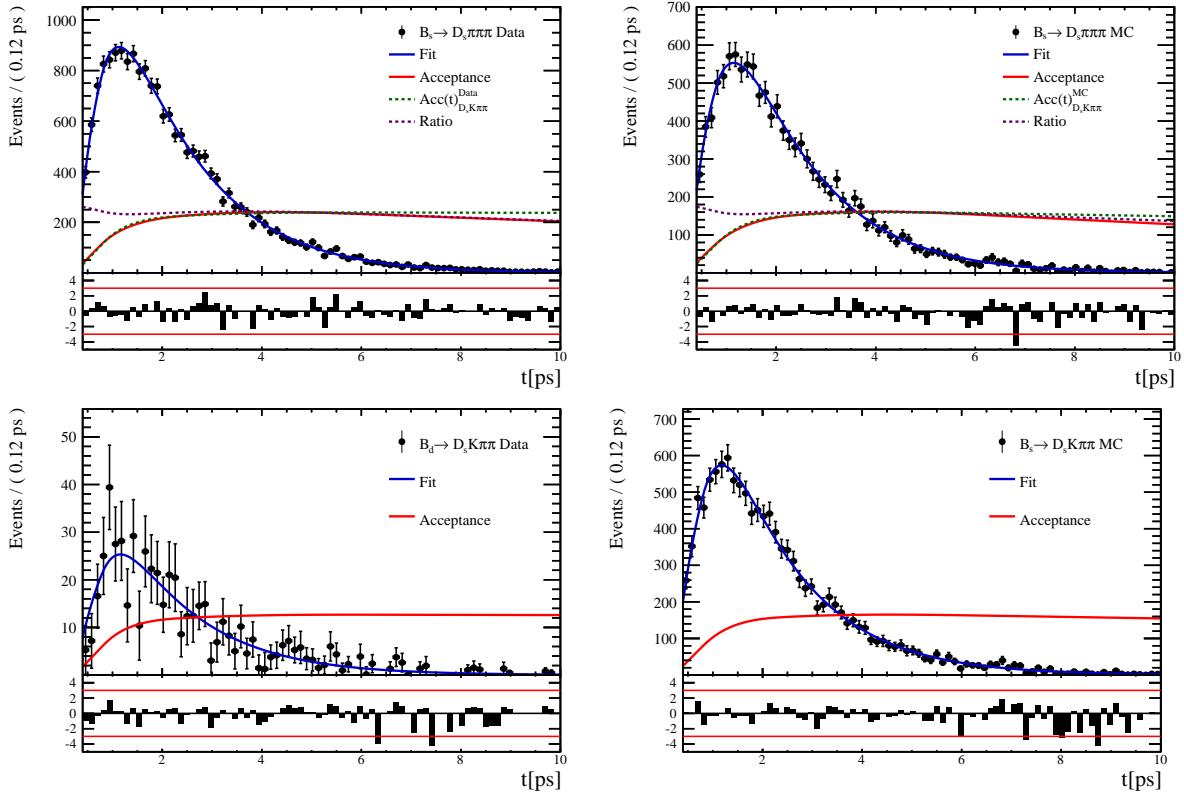


Figure 6.7:

## 431 7 Decay-time Resoution

432 The observed oscillation of B mesons is prone to dilution, if the detector resolution is  
433 of similar magnitude as the oscillation period. In the  $B_s^0$  system, considering that the  
434 measured oscillation frequency of the  $B_s^0$  [8] and the average LHCb detector resolution [13]  
435 are both  $\mathcal{O}(50 \text{ fs}^{-1})$ , this is the case. Therefore, it is crucial to correctly describe the  
436 decay time resolution in order to avoid a bias on the measurement of time dependent CP  
437 parameters.

438 In the presented analysis, we assume a gaussian resolution function with different  
439 widths for each event. This gives rise to a per-event decay time error  $\sigma_t$ , which is  
440 computed separately for every event along with the proper time  $t$ , by the decay time  
441 fitter. Furthermore, the per-event decay time error  $\sigma_t$  is usually underestimated by the  
442 decay time fitter, making it necessary to derive a scaling function, which matches the  
443 per-event error to the actually measured decay time resolution.

444 Due to the lack of a decay time unbiased sample of real  $B_s^0 \rightarrow D_s K \pi \pi$  decays, this  
445 analysis relies on simulation to describe the time resolution. The obtained results will  
446 be compared to those found in the closely related  $B_s^0 \rightarrow D_s K$  analysis and systematic  
447 uncertainties will be assigned conservatively. In the following, we investigate the Run1  
448 and Run2 MC samples to find the proper decay time resolution in bins of the per-event  
449 decay time errors and derive a scaling function from that.

### 451 7.1 Formalism

452 For simulated  $B_s^0 \rightarrow D_s K \pi \pi$  events, the information on the true  $B_s^0$  lifetime  $\tau_{true}$  assigned  
453 at production of the event, as well as the measured decay time  $\tau_{measured}$ , which is determined  
454 after the interaction with the LHCb detector, is stored. In this analysis, the difference  
455  $\Delta t = \tau_{true} - \tau_{measured}$  is obtained for each simulated  $B_s^0 \rightarrow D_s K \pi \pi$  candidate. The width  
456 of the distribution of  $\Delta t$  is a direct measure of the decay time resolution.

457 To analyse the relation between the per-event decay time error  $\sigma_t$  and the actual resolution,  
458 the simulated sample is split into 8 bins of  $\sigma_t$ . Each bin width is chosen using an adaptive  
459 binning scheme which ensures that approximately equal numbers of events are found  
460 in every bin. A fit is then performed to the distribution of  $\Delta t$  in each of the bins to  
461 determine the decay time resolution in the respective bin. After that, the correlation  
462 between the binned per-event decay time error and the measured decay time resolution is  
463 analyzed to determine the scaling function.

### 464 7.2 Decay-time Error in Run I & Run II

465 Due to the increase in center of mass energy from Run I to Run II, as well as (among  
466 others) new tuning in the pattern and vertex reconstruction, the distributions of the raw  
467 decay time error might not necessarily match each other between the two different runs.  
468 Significant deviations can be observed in the shape and mean of those two distributions  
469 for  $B_s^0 \rightarrow D_s K \pi \pi$  signal candidates shown in Figure 7.1.

470 It can be observed that the decay time error distribution for signal candidates from Run  
471 II is significantly broader and shifted to slightly higher values. Due to the discrepancies  
472 between the distributions of the decay time error  $\sigma_r$  for Run I and Run II data, the time

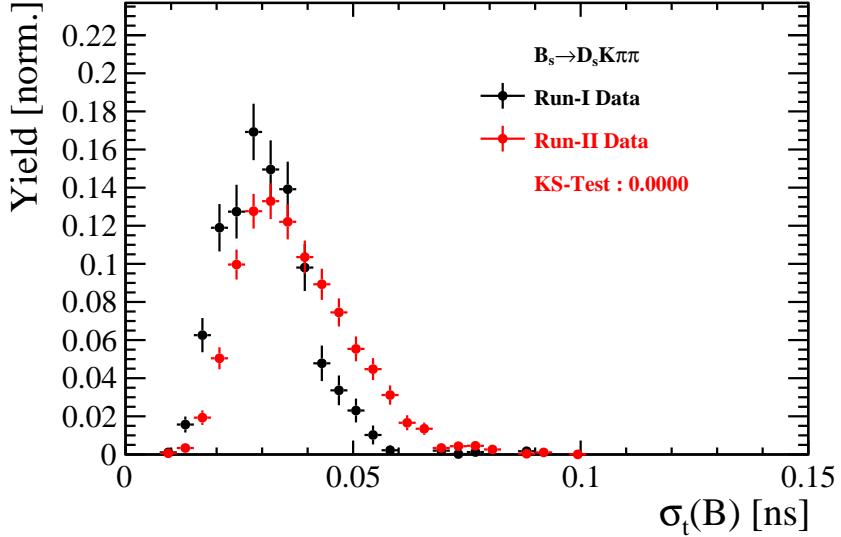


Figure 7.1: Distribution of the decay time error for  $B_s^0 \rightarrow D_s K\pi\pi$  signal candidates on data for (black) Run I and (red) Run II. The signal distributions are obtained using the sWeight technique.

<sup>473</sup> resolution studies have to be performed separately for both runs, which leads to two  
<sup>474</sup> different scaling functions to map  $\sigma_t \rightarrow \sigma_{eff}$ .

### <sup>475</sup> 7.3 Fits to the decay time distributions

<sup>476</sup> The sum of two Gaussian functions is used to fit the distributions of the decay time  
<sup>477</sup> difference  $\Delta t$  in each  $\sigma_t$  bin. One Gaussian function is relatively narrow and describes the  
<sup>478</sup> decay time of the majority of candidates in each bin, while the other, broader Gaussian  
<sup>479</sup> function describes candidates where the measured decay time differs considerably from  
<sup>480</sup>  $\tau_{true}$ . Those contributions are shifted to the tails of the distribution. From the two  
<sup>481</sup> Gaussian functions, the combined, effective width  $\sigma_{eff}$  is quoted as decay time resolution  
<sup>482</sup> in each bin. Figure 7.2 shows the double Gaussian fit to the distribution of the decay time  
<sup>483</sup> difference for events where  $20.7 \text{ ps} < \sigma_t < 24.3 \text{ ps}$ . All fits are shown in the Appendix ??.  
<sup>484</sup>

<sup>485</sup> For the combination of the two separate widths  $\sigma_1$  and  $\sigma_2$ , a method which takes the  
<sup>486</sup> damping effect of the finite time resolution on the CP observables into account, is used.  
<sup>487</sup> The effective damping of the CP amplitudes is described by the dilution  $\mathcal{D}$ , which can  
<sup>488</sup> take values between 1 and 0. In the case of infinite precision, there would be no damping  
<sup>489</sup> and therefore  $\mathcal{D} = 1$  would hold, while for a resolution that is much larger than the  $B_s^0$   
<sup>490</sup> oscillation frequency,  $\mathcal{D}$  would approach 0. For two Gaussians describing the resolution,  
<sup>491</sup> the dilution can be defined as [14]

$$\mathcal{D} = f_1 e^{-\sigma_1^2 \Delta m_s^2 / 2} + (1 - f_1) e^{-\sigma_2^2 \Delta m_s^2 / 2}, \quad (7.1)$$

<sup>492</sup> where  $f_1$  is the fraction of events described by the first Gaussian relative to the second  
<sup>493</sup> and  $\Delta m_s$  is the oscillation period of the  $B_s^0$  meson.

<sup>494</sup> The dilution is computed in every bin of the per-event decay time error and can be

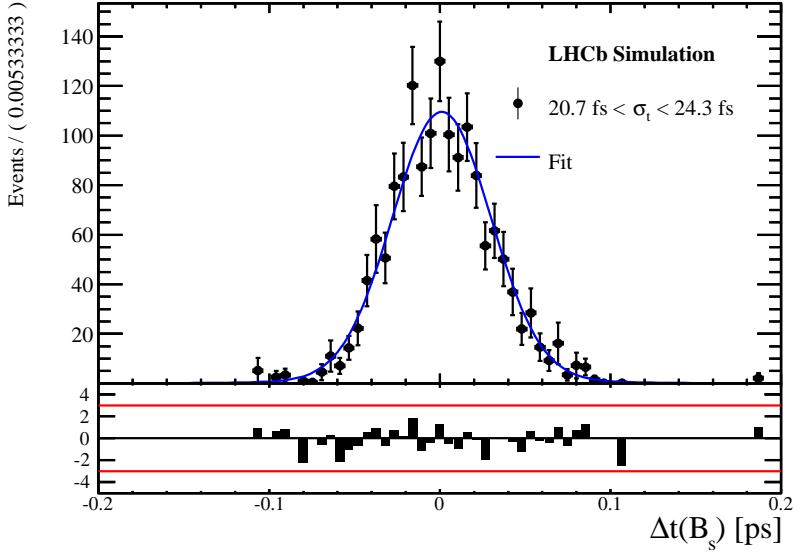


Figure 7.2: Difference of the true and measured decay time of  $B_s^0 \rightarrow D_s K \pi\pi$  candidates from MC in the bin  $20.7 \text{ ps} < \sigma_t < 24.3 \text{ ps}$ . A fit of the sum of two Gaussian functions is overlaid.

495 converted into the effective resolution

$$\sigma_{eff} = \sqrt{(-2/\Delta m_s^2) \ln \mathcal{D}}. \quad (7.2)$$

## 496 7.4 Results

497 The fitted values for the Gaussian widths  $\sigma_1$  and  $\sigma_2$ , the fraction of the first relative to  
 498 the second Gaussian function  $f_1$ , as well as the effective resolution  $\sigma_{eff}$ , found in each bin  
 499  $\sigma_t$ , are shown in Tab. 7.1. Figure 7.3 shows the obtained values for  $\sigma_{eff}$  as a function of  
 500 the per-event decay time error  $\sigma_t$ . A linear polynom of the form

$$\sigma(\sigma_t)_{mc} = s_0 + s_1 \cdot \sigma_t \quad (7.3)$$

501 is used to parametrise this distribution. The obtained values are

$$\sigma(\sigma_t)_{mc} = 0 + (1.257 \pm 0.017)\sigma_t, \quad (7.4)$$

502 where  $s_0$  is compatible with 0 in the fit and therefore is set to  $s_0 = \sigma(\sigma_t = 0) = 0$ . For  
 503 comparison, the linear scaling functions found for  $\sigma(\sigma_t)$  in the  $B_s^0 \rightarrow D_s K$  analysis [14] for  
 504 MC is also shown in Figure 7.3. Motivated by the similarity between the  $B_s^0 \rightarrow D_s K \pi\pi$   
 505 and  $B_s^0 \rightarrow D_s K$  decay, we assume a comparable scaling relation for data,

$$\frac{\sigma(t)_{D_s K \pi\pi, data}}{\sigma(t)_{D_s K \pi\pi, mc}} \approx \frac{\sigma(t)_{D_s K, data}}{\sigma(t)_{D_s K, mc}}. \quad (7.5)$$

506 This leads to a correction factor

$$\sigma(t)_{D_s K \pi\pi, data} \approx \frac{\sigma(t)_{D_s K, data}}{\sigma(t)_{D_s K, mc}} \cdot \sigma(t)_{D_s K \pi\pi, mc}, \quad (7.6)$$

507 where all elements of the right side of the equation are known.

508  
509 Taking the scaling function found in our simulation, as well as input from the  $B_s^0 \rightarrow$   
510  $D_s K$  analysis for  $\sigma(t)_{D_s K, mc/data}$ , we find

$$\sigma(t)_{D_s K \pi\pi, data} = 10.3 fs + 1.28 \cdot t$$

511 ,  
512 which is the scaling factor used for the per-event decay time error in the nominal time-  
513 and amplitude-dependent fit.

$\sigma_t$ Bin [fs]	$\sigma_1$ [fs]	$\sigma_2$ [fs]	$f_1$	D	$\sigma_{eff}$ [fs]
0to19	$22.57 \pm 0.96$	$45.57 \pm 4.061$	$0.827 \pm 0.057$	$0.89 \pm 0.067$	$27.46 \pm 8.82$
19to24	$24.64 \pm 1.03$	$46.65 \pm 3.109$	$0.768 \pm 0.061$	$0.86 \pm 0.070$	$30.64 \pm 8.48$
24to29	$30.96 \pm 0.90$	$58.76 \pm 5.684$	$0.884 \pm 0.045$	$0.83 \pm 0.05$	$34.66 \pm 5.28$
29to34	$35.28 \pm 1.54$	$57 \pm 6.698$	$0.839 \pm 0.098$	$0.79 \pm 0.10$	$39.09 \pm 10.47$
34to39	$37.05 \pm 2.36$	$61.98 \pm 5.769$	$0.707 \pm 0.12$	$0.73 \pm 0.12$	$44.76 \pm 11.78$
39to44	$68.38 \pm 8.33$	$42.15 \pm 3.583$	$0.331 \pm 0.18$	$0.66 \pm 0.16$	$50.98 \pm 15.11$
44to49	$199.9 \pm 100.1$	$53.72 \pm 1.419$	$0.020 \pm 0.014$	$0.62 \pm 0.02$	$54.89 \pm 1.60$
49to150	$68.75 \pm 165.3$	$68.92 \pm 4.603$	$0.001 \pm 0.97$	$0.47 \pm 0.65$	$68.92 \pm 63.42$

Table 7.1: Summary of the obtained parameters from the resolution fits described above.

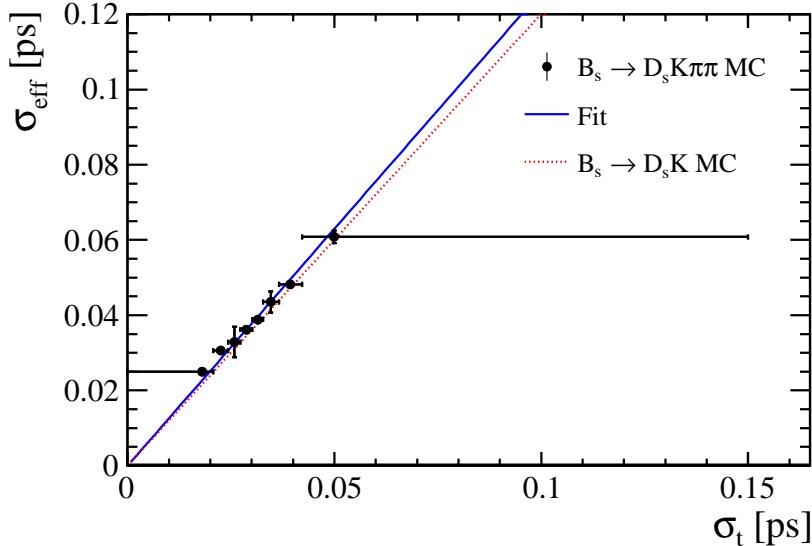


Figure 7.3: Decay-time resolution of  $B_s^0 \rightarrow D_s K \pi\pi$  candidates from MC. The scaling functions found in  $B_s^0 \rightarrow D_s K$  (dotted red line) MC is also shown for comparison. The fit described in the text is overlaid.

## 514 8 Time dependent fit

515 This section will cover the phasespace integrated, time dependent fit to  $B_s^0 \rightarrow D_s h\pi\pi$  data,  
 516 which is described by the PDF formulated in Eq. 2.6. For the phasespace integrated fit to  
 517  $B_s^0 \rightarrow D_s K\pi\pi$  data, the sensitivity to the CKM phase  $\gamma$  will depend on the magnitude of  
 518 the coherence factor  $\kappa$ , defined in Eq. 2.7, which is added as an additional fit parameter. In  
 519 order to avoid any pollution of the final data samples by background events, both samples  
 520 are weighted using the sWeights obtained by the fits to the invariant mass distributions,  
 521 described in Sec. 4. This fit approach is commonly known as *sFit*. As additional input to  
 522 the fit, the tagging information (Sec. 5), as well as the decay time acceptance (Sec. 6)  
 523 and resolution (Sec. 7) is used and fixed to the values obtained by the dedicated studies.  
 524 Taking all inputs into account, the final time dependent fit PDF is given by

$$525 \quad \mathcal{P}\mathcal{D}\mathcal{F}(t, \vec{\lambda}) = \left( \epsilon(t) \cdot \int P(x, t, q_t, q_f) dx \right) \times \mathcal{R}(t - t'), \quad (8.1)$$

525 where  $\int P(x, t, q_t, q_f) dx$  is the PDF given by Eq. 2.6,  $\epsilon(t)$  is the efficiency due to the  
 526 time acceptance effects and  $\mathcal{R}(t - t')$  is the Gaussian time resolution function.

### 527 8.1 sFit to $B_s^0 \rightarrow D_s \pi\pi\pi$ data

528 The phasespace-integrated, time-dependent fit is performed to the full sWeighted sample  
 529 of selected candidates from Run I and 2015+2016 Run II data, containing both possible  
 530 magnet polarities and  $D_s$  final states. In the fit, the values of  $\Gamma_s$  and  $\Delta\Gamma_s$  are fixed to the  
 531 latest PDG report. All tagging parameters are fixed to the central values found in the  
 532 tagging calibration, described in Sec. 5. Due to the fact that the  $B_s^0 \rightarrow D_s \pi\pi\pi$  decay is  
 533 flavour specific, the CP-coefficients can be fixed to  $C = 1$  and  $D_f = D_{\bar{f}} = S_f = S_{\bar{f}} = 0$ ,  
 534 reducing Eq. 2.6 to

$$535 \quad \int P(x, t, q_t, q_f) dx = [\cosh\left(\frac{\Delta\Gamma t}{2}\right) + q_t q_f \cos(\Delta m_s t)] e^{-\Gamma t}. \quad (8.2)$$

535 Note that in this case, the dependence on the coherence factor  $\kappa$  is dropped and the  
 536 same relation as found for  $B_s^0 \rightarrow D_s \pi$  decays is recovered. Therefore, the only free fit  
 537 parameter left is  $\Delta m_s$ . The data distribution with the overlaid fit is shown in Fig. xXx  
 538 and the obtained value for the mixing frequency is

$$539 \quad \Delta m_s = xx.xxx \pm 0.yyy. \quad (8.3)$$

### 539 8.2 sFit to $B_s^0 \rightarrow D_s K\pi\pi$ data

### 540 8.3 Results

## 541 A Detailed mass fits

542 In this section, all fits to the mass distribution of  $B_s^0 \rightarrow D_s\pi\pi\pi$  and  $B_s^0 \rightarrow D_sK\pi\pi$   
 543 candidates are shown. The fits are performed simultaneously for every year of datataking  
 544 (2011, 2012, 2015 and 2016) and the  $D_s$  decay ( $D_s \rightarrow KK\pi$  non-resonant,  $D_s \rightarrow \phi\pi$ ,  
 545  $D_s \rightarrow K^*K$ , or  $D_s \rightarrow \pi\pi\pi$ ) through which the final state is reached.

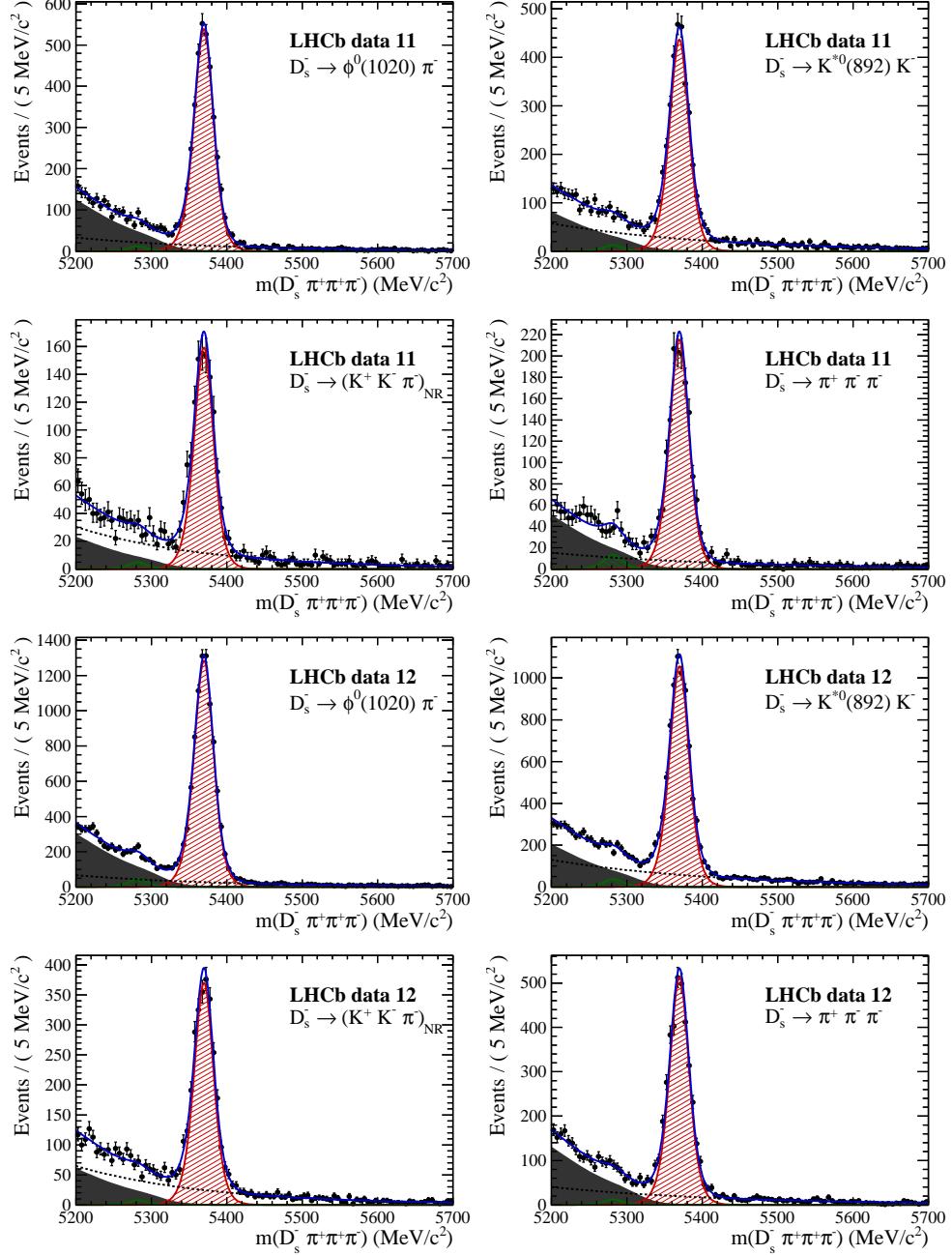


Figure 1.1: Invariant mass distributions of  $B_s^0 \rightarrow D_s\pi\pi\pi$  candidates, ordered by  $D_s$  final state, for Run1 data. The fit described in 4.4 is overlaid.

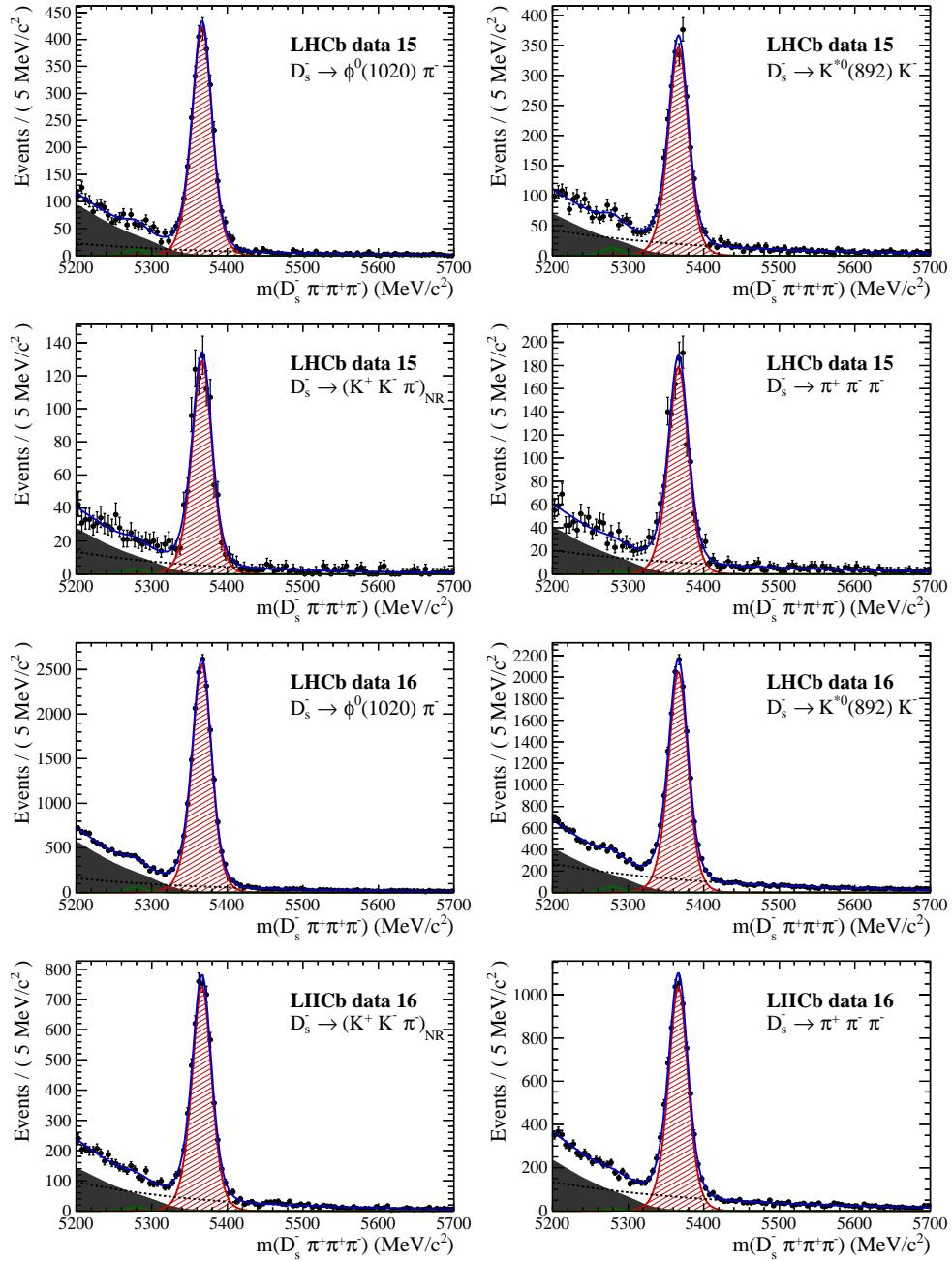


Figure 1.2: Invariant mass distributions of  $B_s^0 \rightarrow D_s \pi\pi\pi$  candidates, ordered by  $D_s$  final state, for Run2 data. The fit described in 4.4 is overlaid.

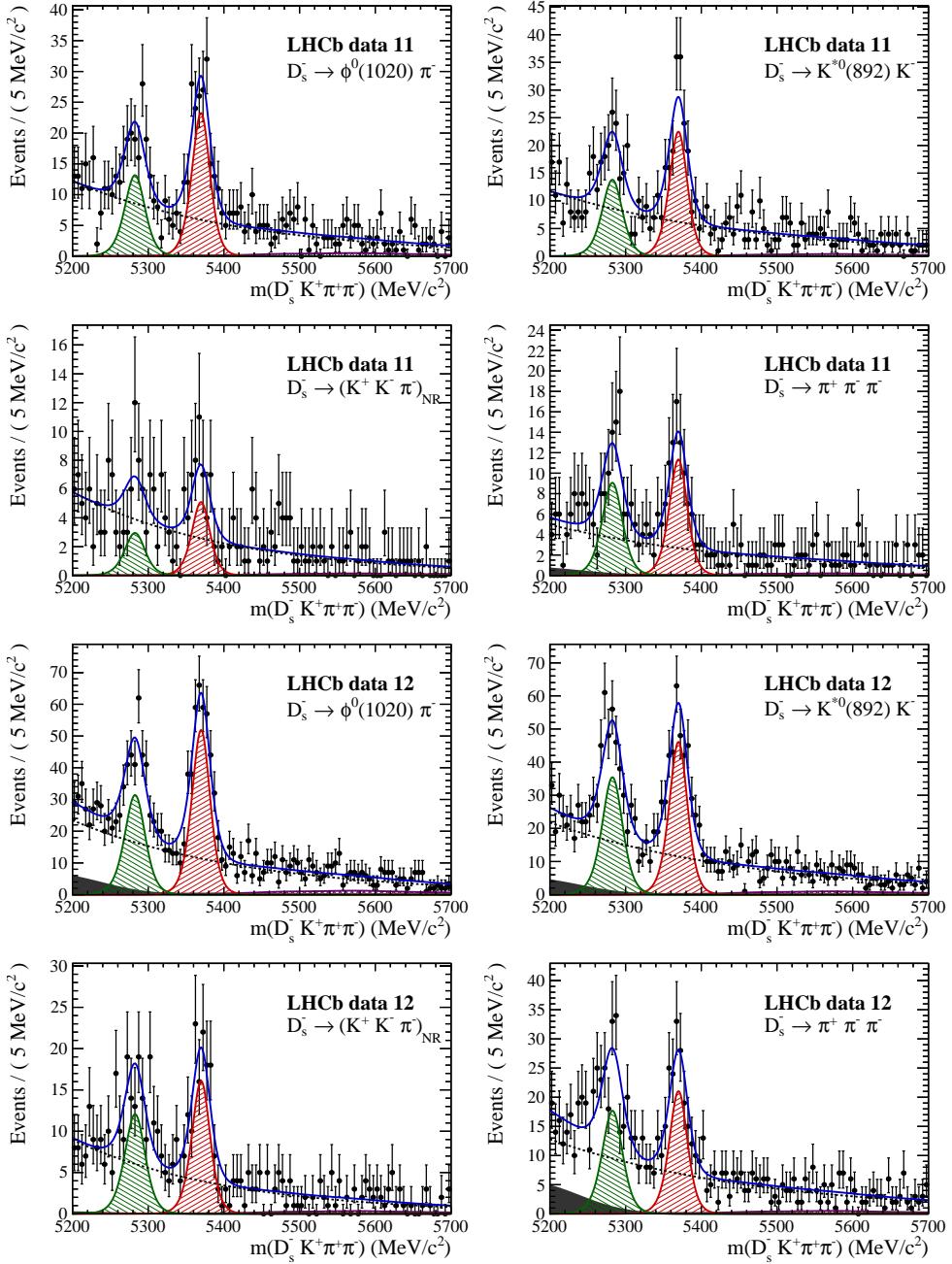


Figure 1.3: Invariant mass distributions of  $B_s^0 \rightarrow D_s K \pi \pi$  candidates, ordered by  $D_s$  final state, for Run1 data. The fit described in 4.5 is overlaid.

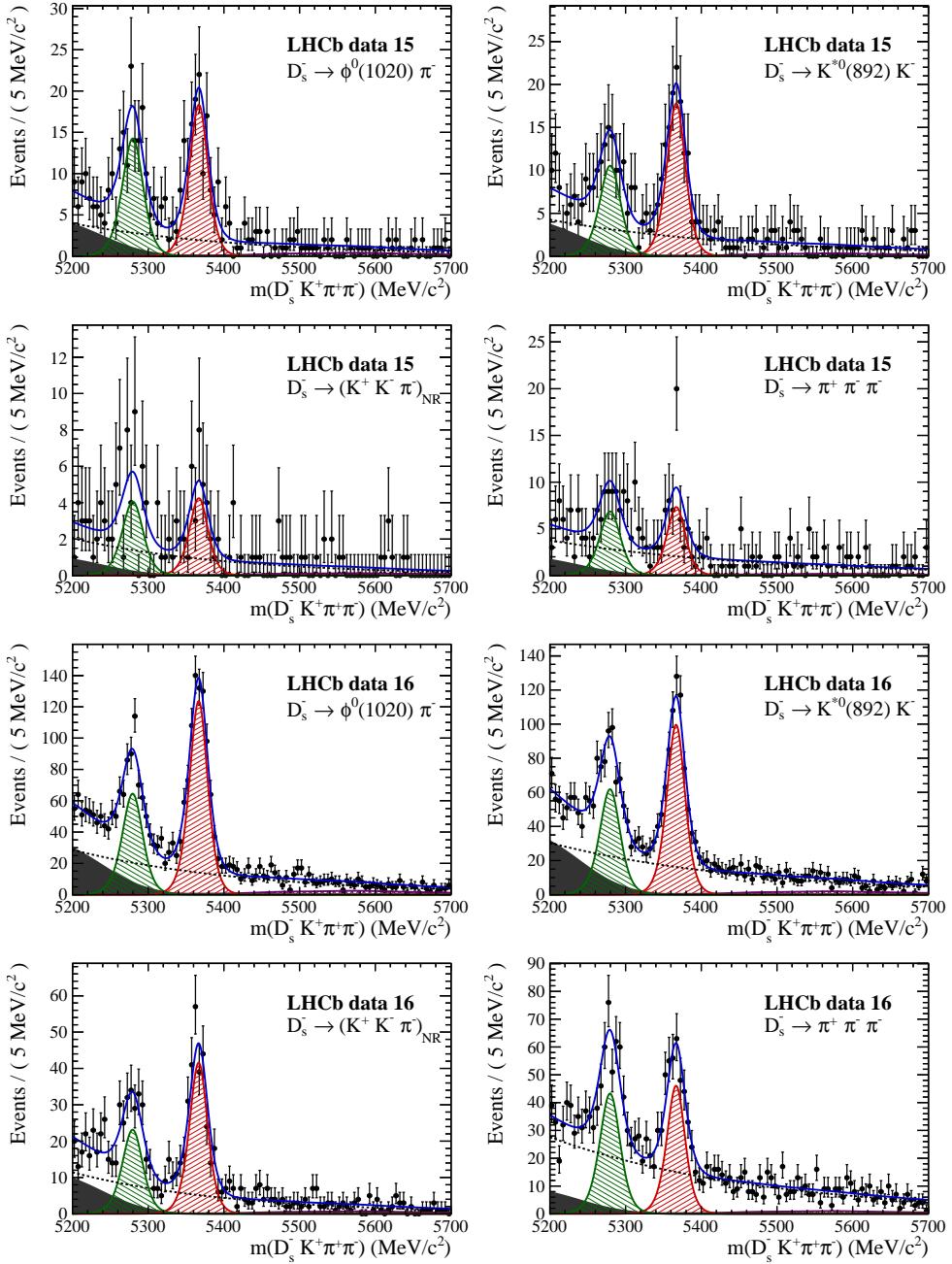


Figure 1.4: Invariant mass distributions of  $B_s^0 \rightarrow D_s K \pi \pi$  candidates, ordered by  $D_s$  final state, for Ru2 data. The fit described in 4.5 is overlaid.

546 **B Decay-time Resolution fits**

547 This section contains all fits to the distributions of the decay time difference  $\Delta t$  between  
 548 the true and the reconstructed decay time of the truth-matched  $B_s^0$  candidates on MC.  
 549 The fits are performed in bins of the decay time error  $\sigma_t$ , where an adaptive binning  
 550 scheme is used to ensure that approximately the same number of events are found in each  
 551 bin.

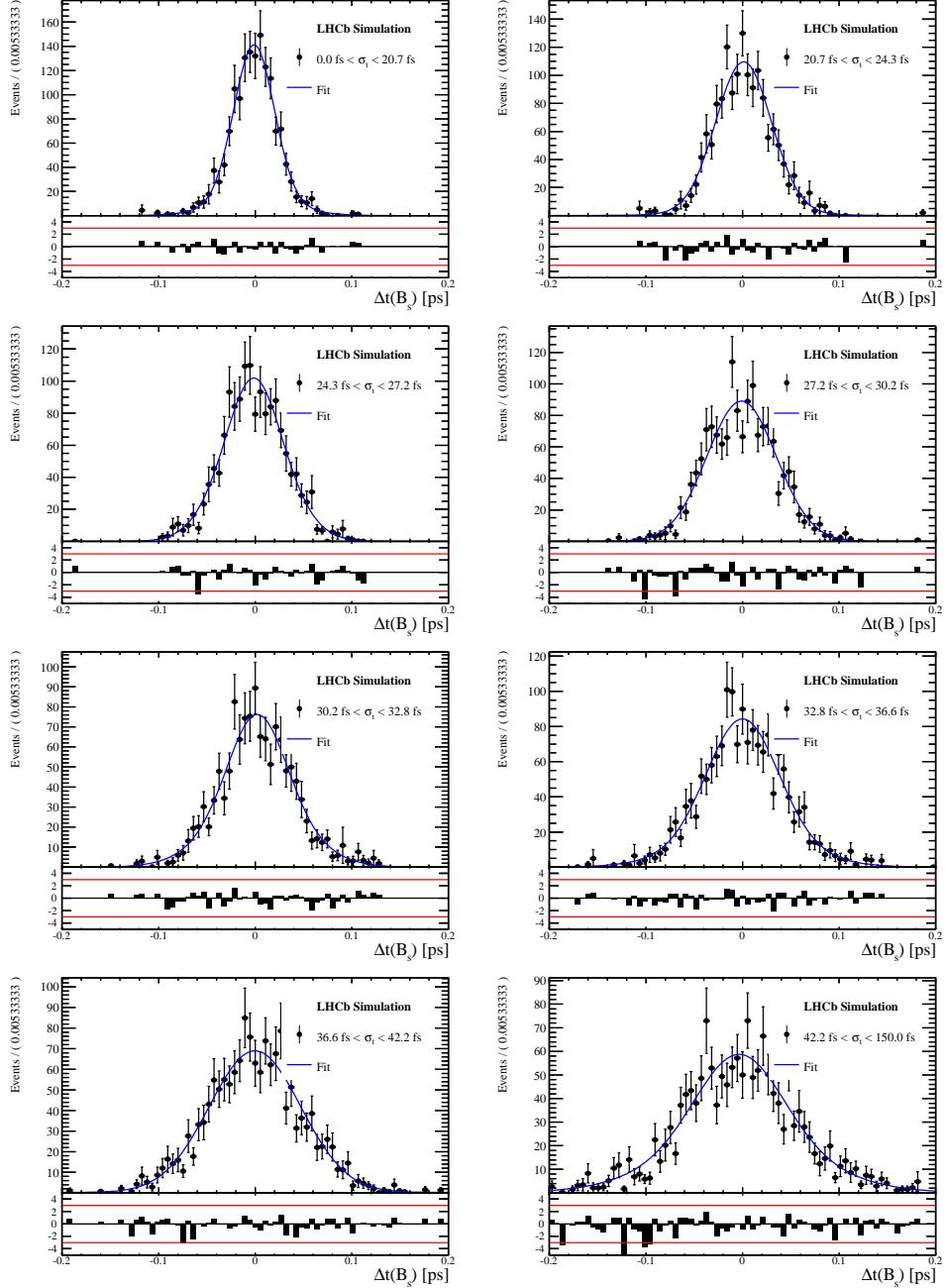


Figure 2.1: Difference of the true and measured decay time of  $B_s^0 \rightarrow D_s K \pi \pi$  candidates from MC in all bins. The fit described in 7 is overlaid.

552 C MC corrections

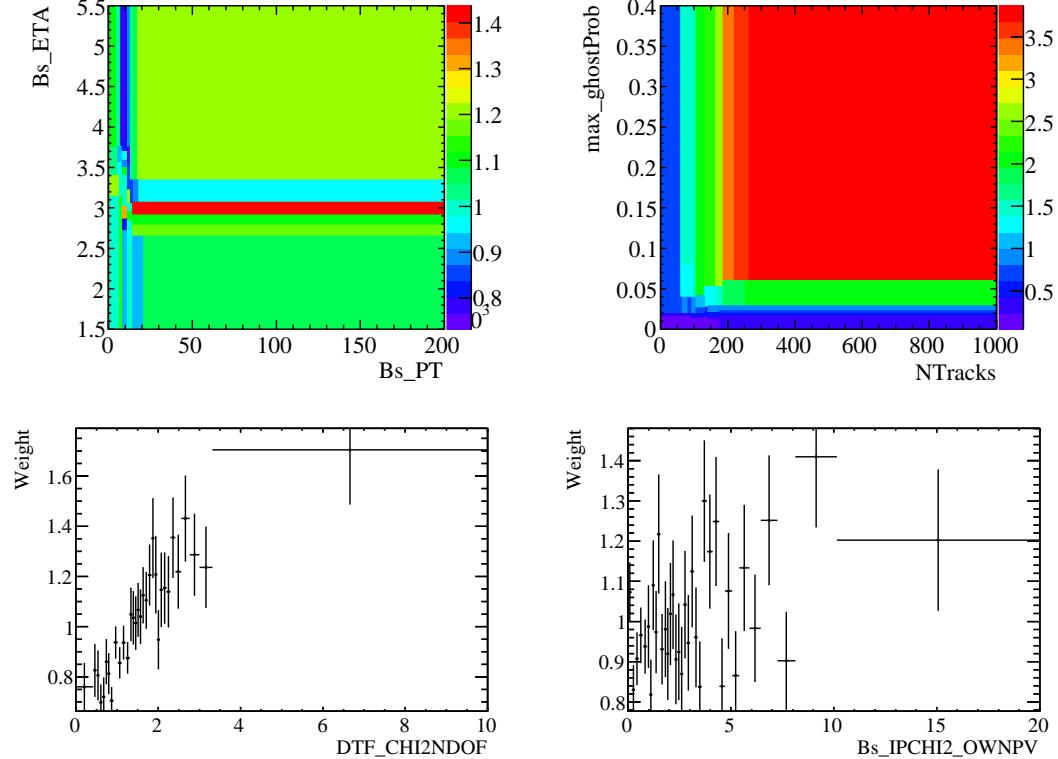


Figure 3.1: Weights applied to correct for Data/MC differences.

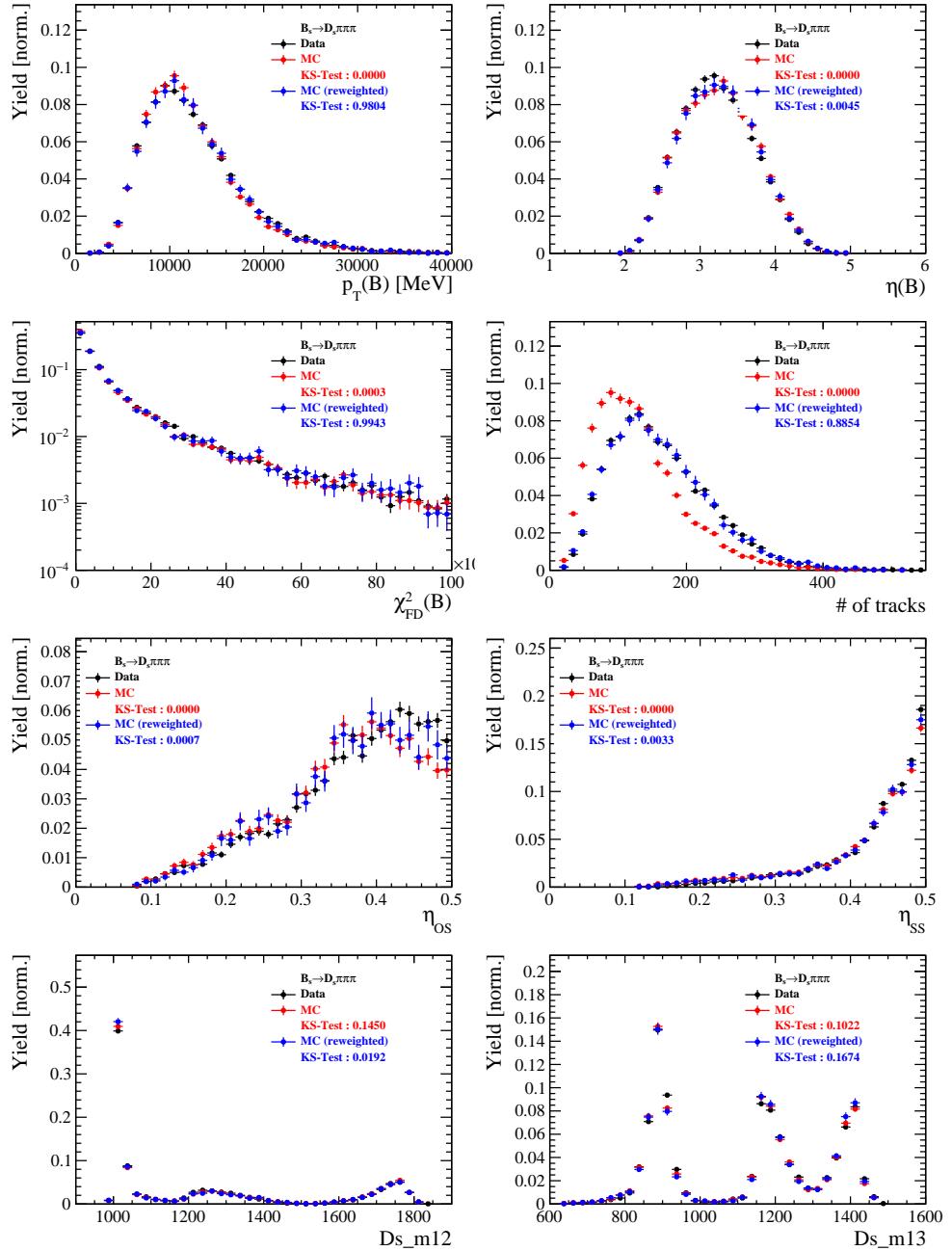


Figure 3.2: Comparison of selected variables for  $B_s \rightarrow D_s \pi\pi\pi$  decays.

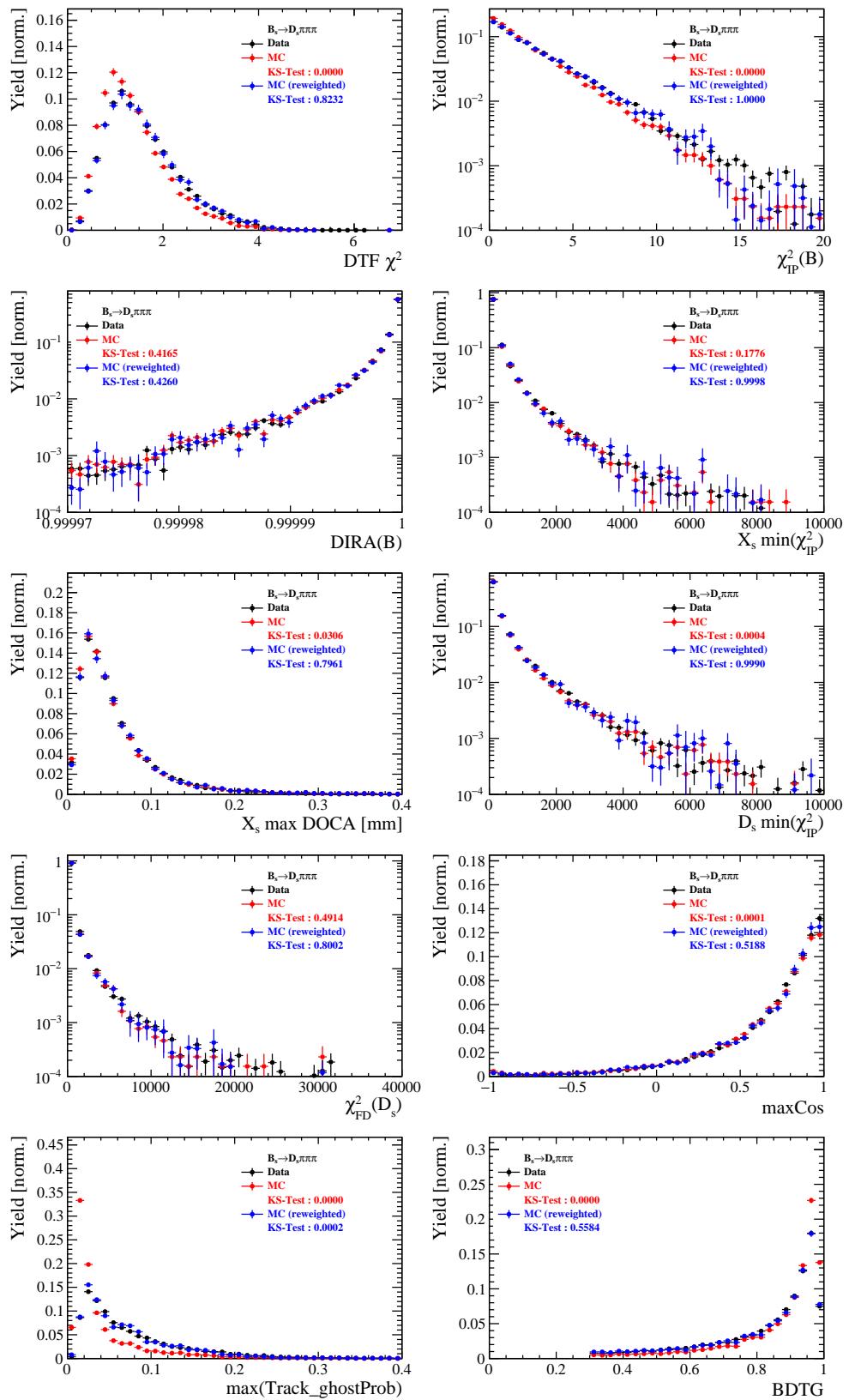


Figure 3.3: Comparison of BDTG input variables and classifier response for  $B_s \rightarrow D_s \pi\pi\pi$  decays.

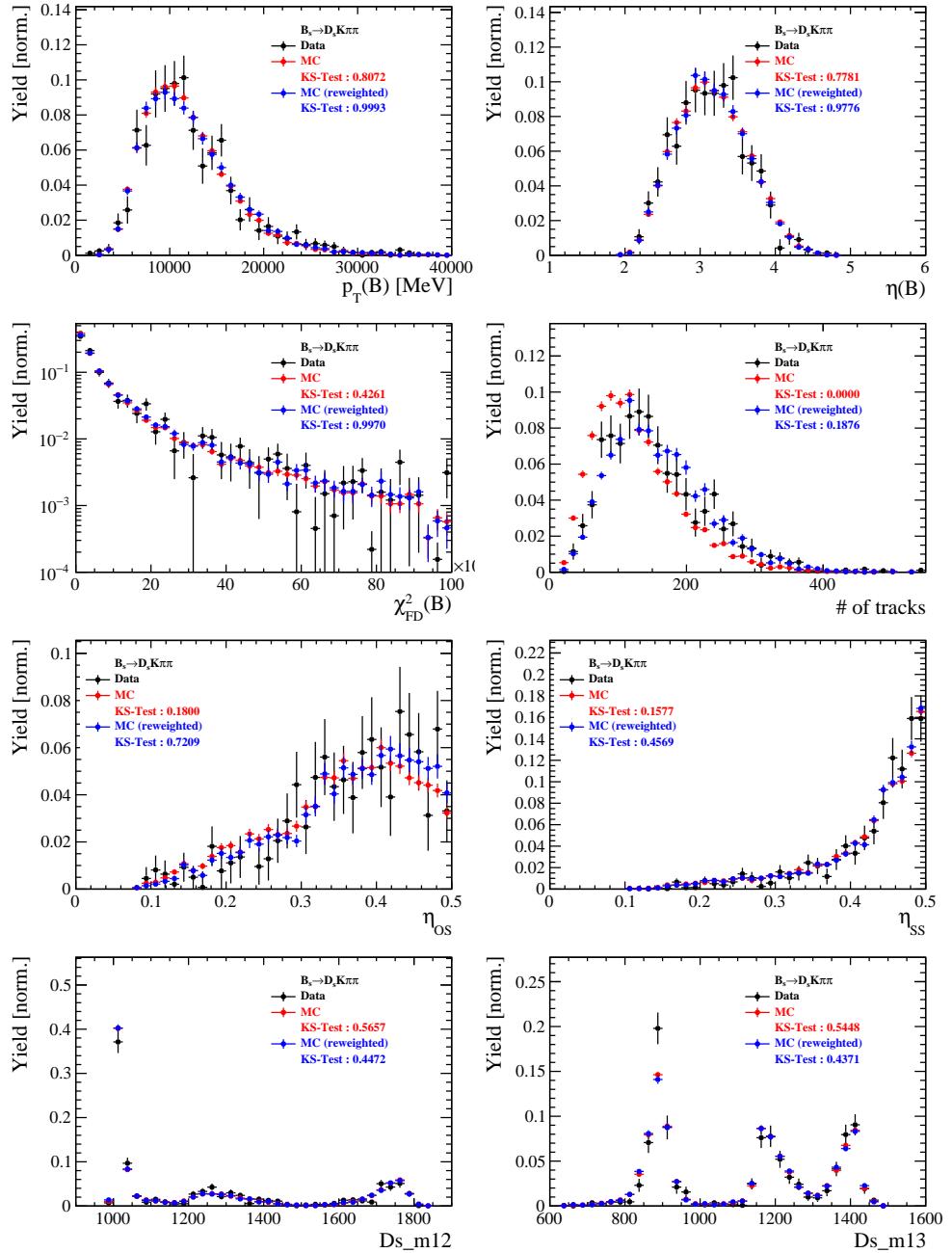


Figure 3.4: Comparison of selected variables for  $B_s \rightarrow D_s K\pi\pi$  decays.

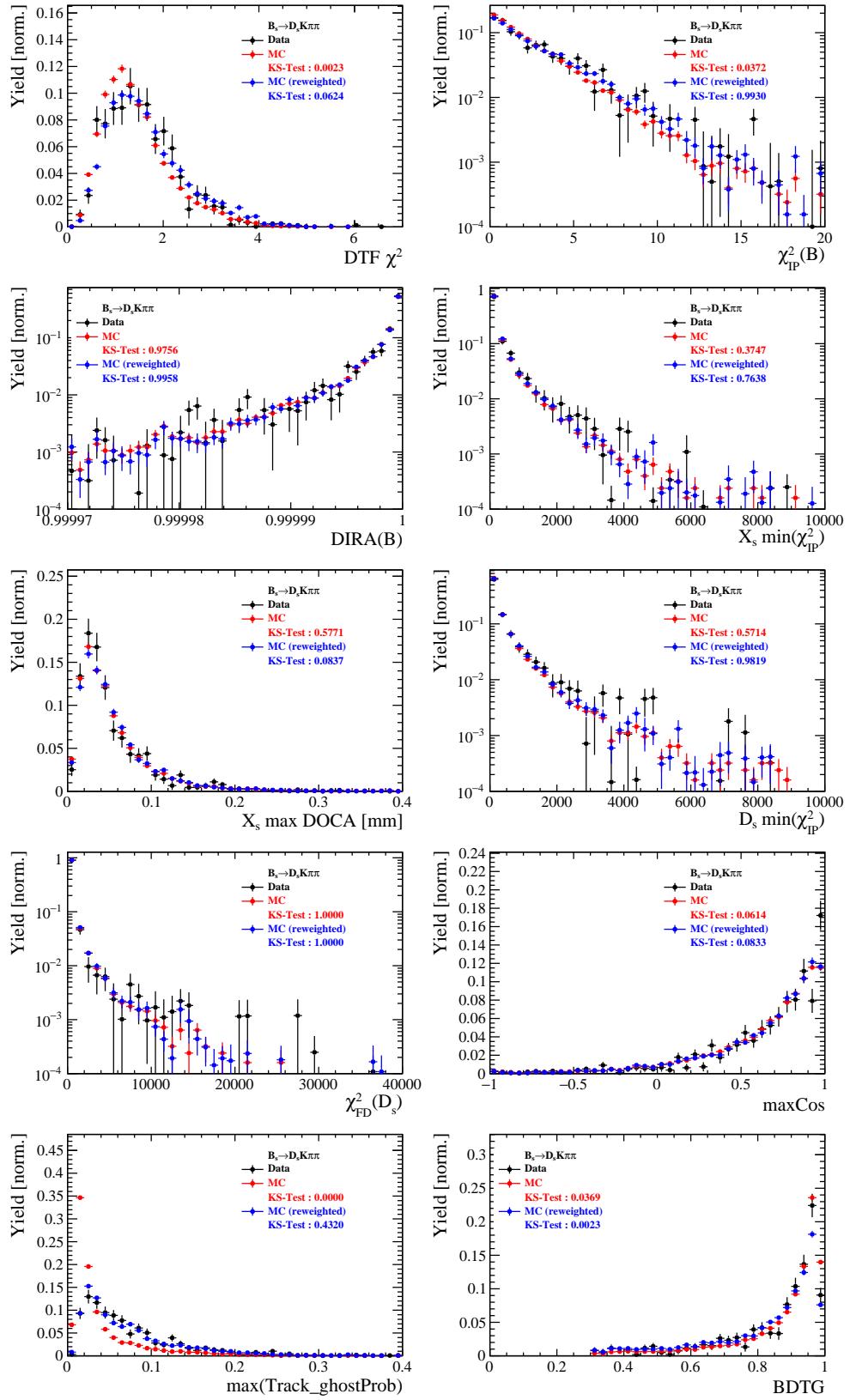


Figure 3.5: Comparison of BDTG input variables and classifier response for  $B_s \rightarrow D_s K\pi\pi$  decays.

553 **D Data distributions**

554 **D.1 Comparison of signal and calibration channel**

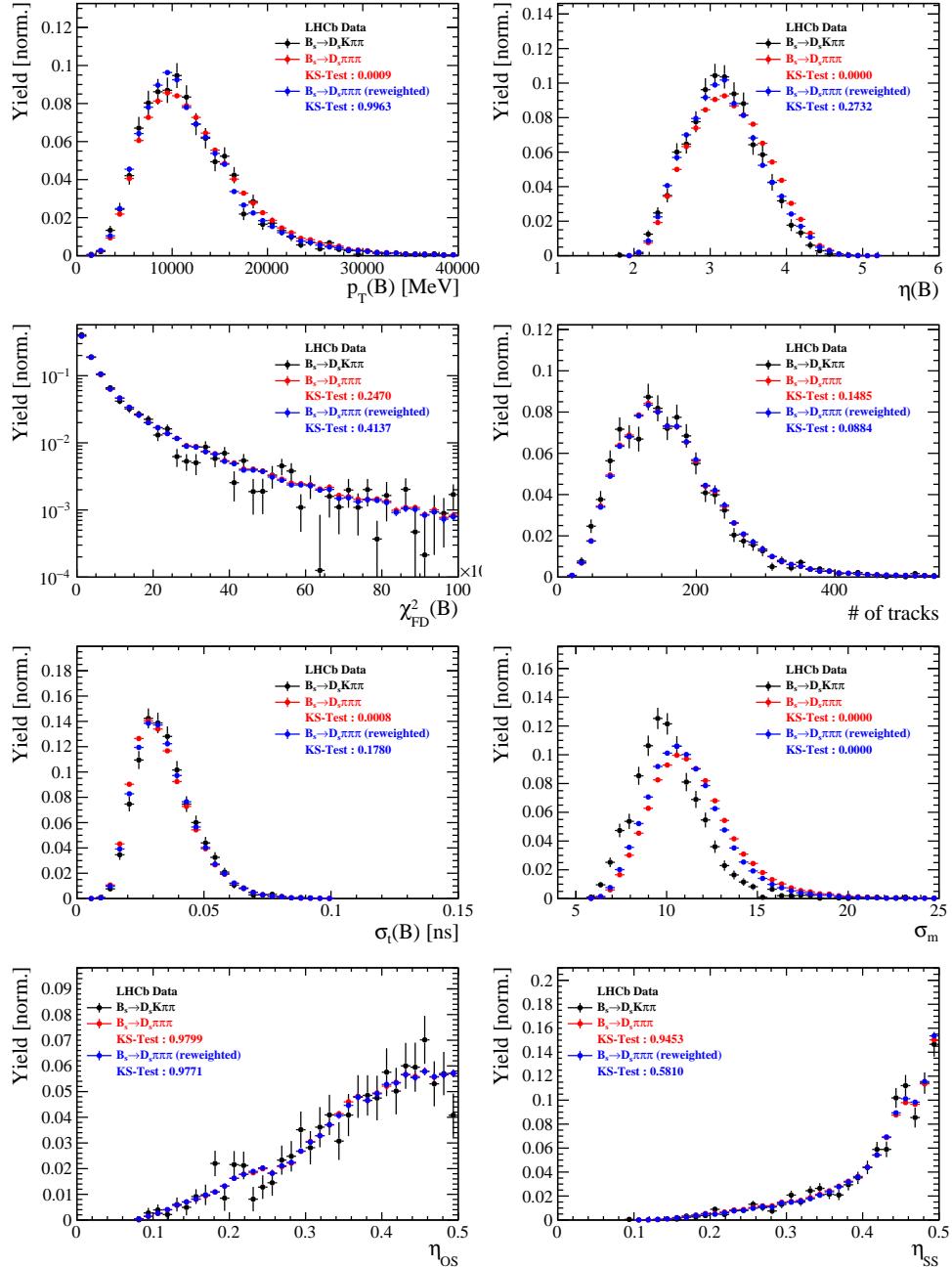


Figure 4.1: Comparison of selected variables.

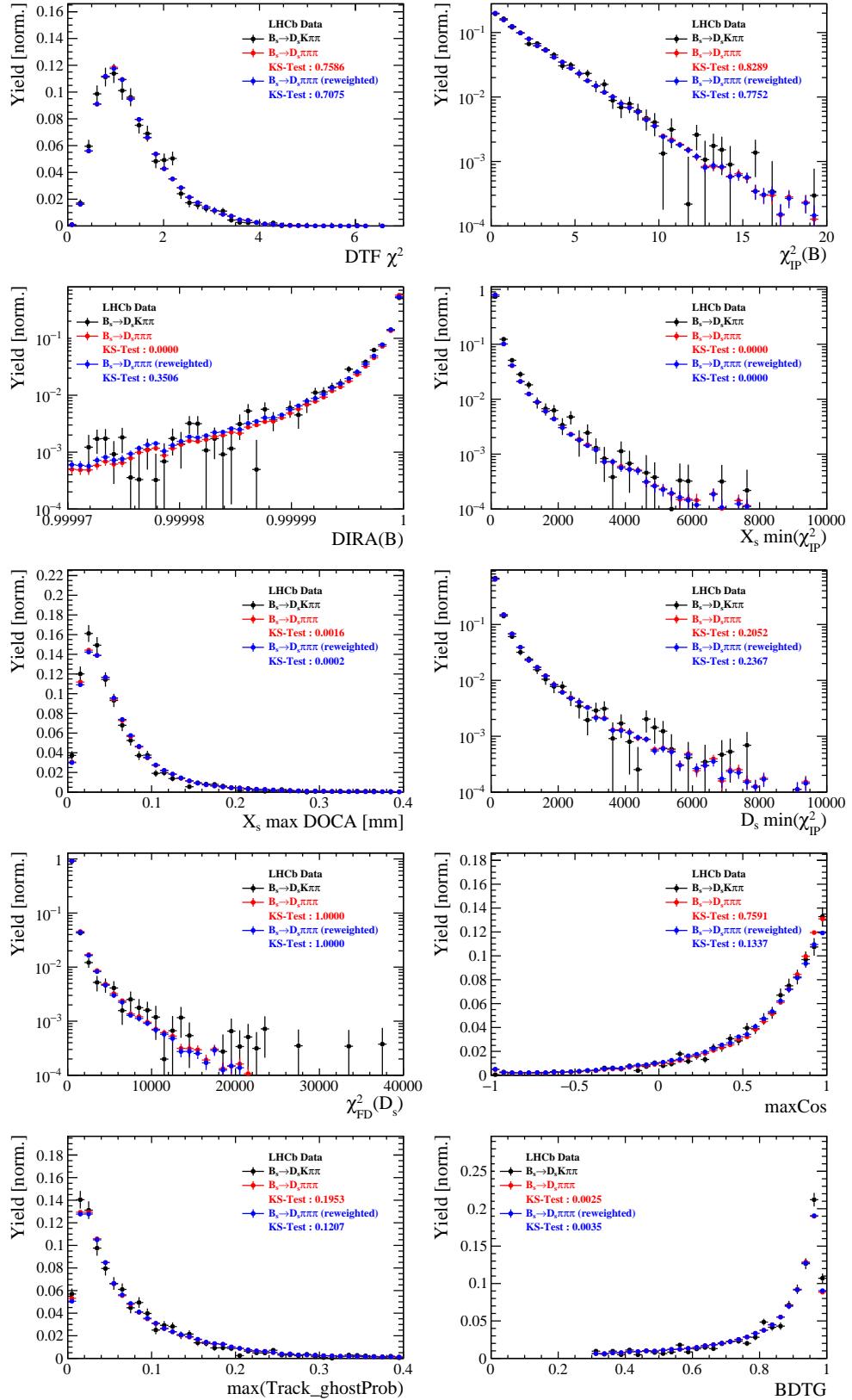


Figure 4.2: Comparison of BDTG input variables and classifier response.

555 D.2 Comparison of Run-I and Run-II data

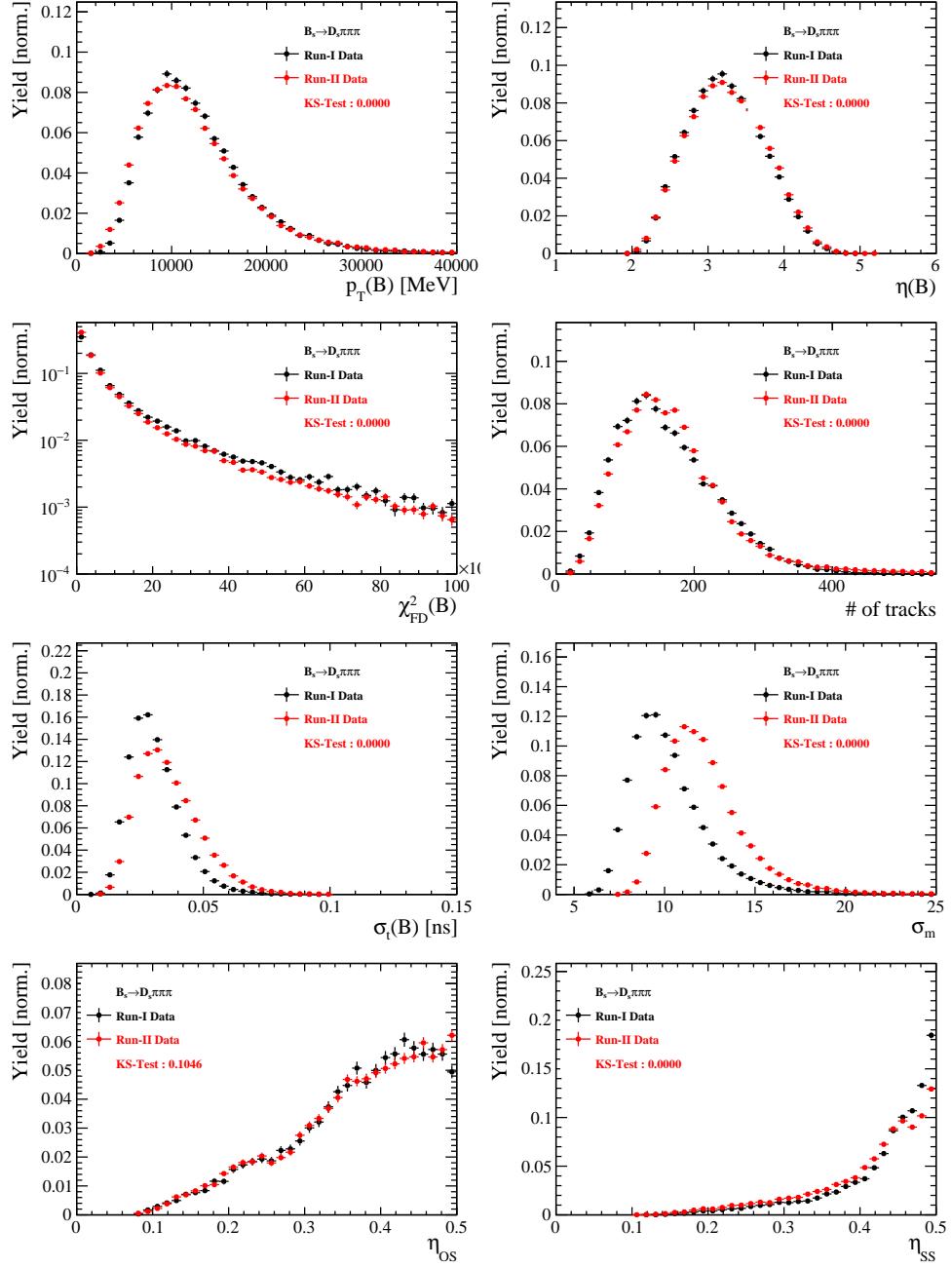


Figure 4.3: Comparison of selected variables.

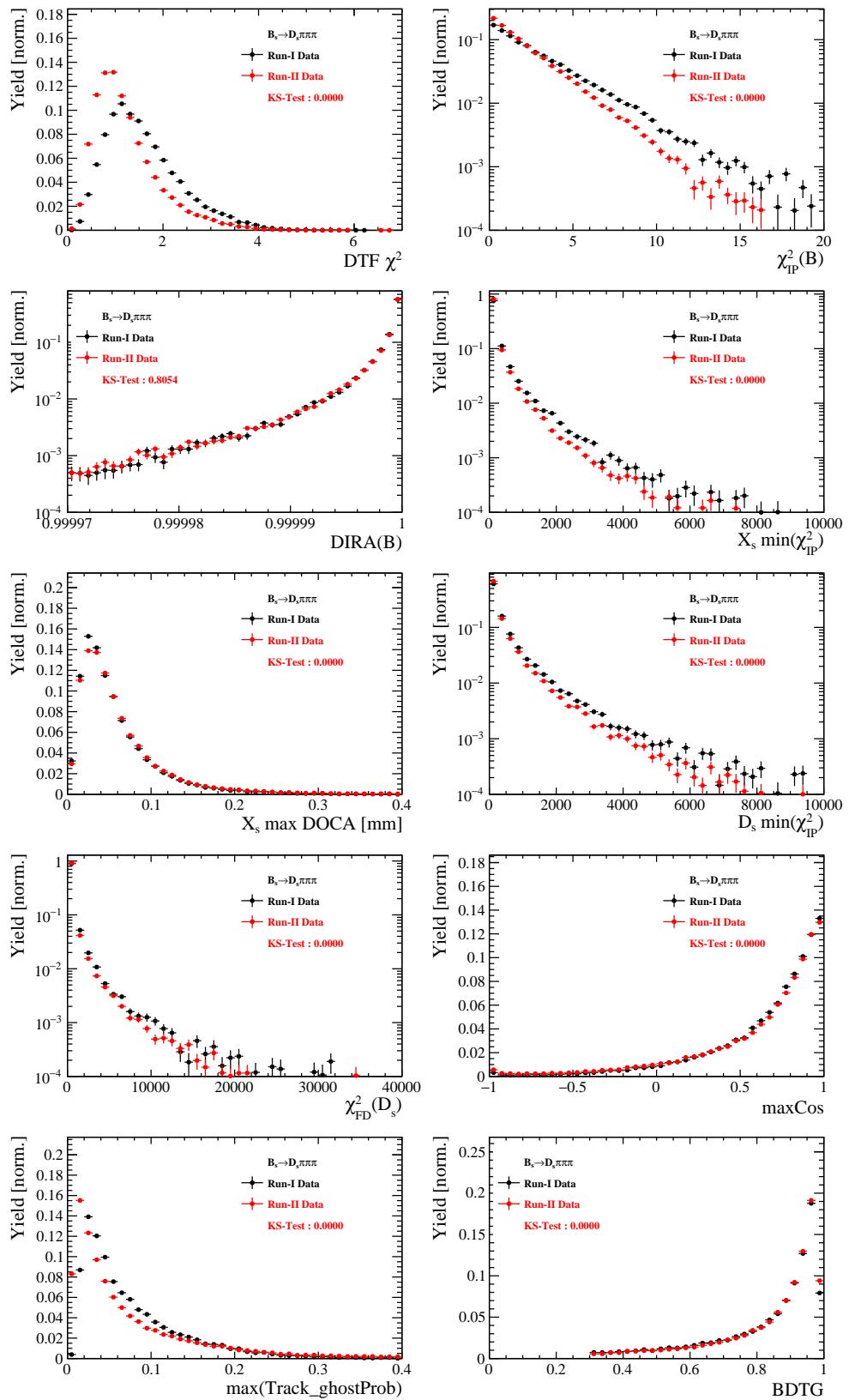


Figure 4.4: Comparison of BDTG input variables and classifier response.

556 D.3 Comparison of  $D_s$  final states

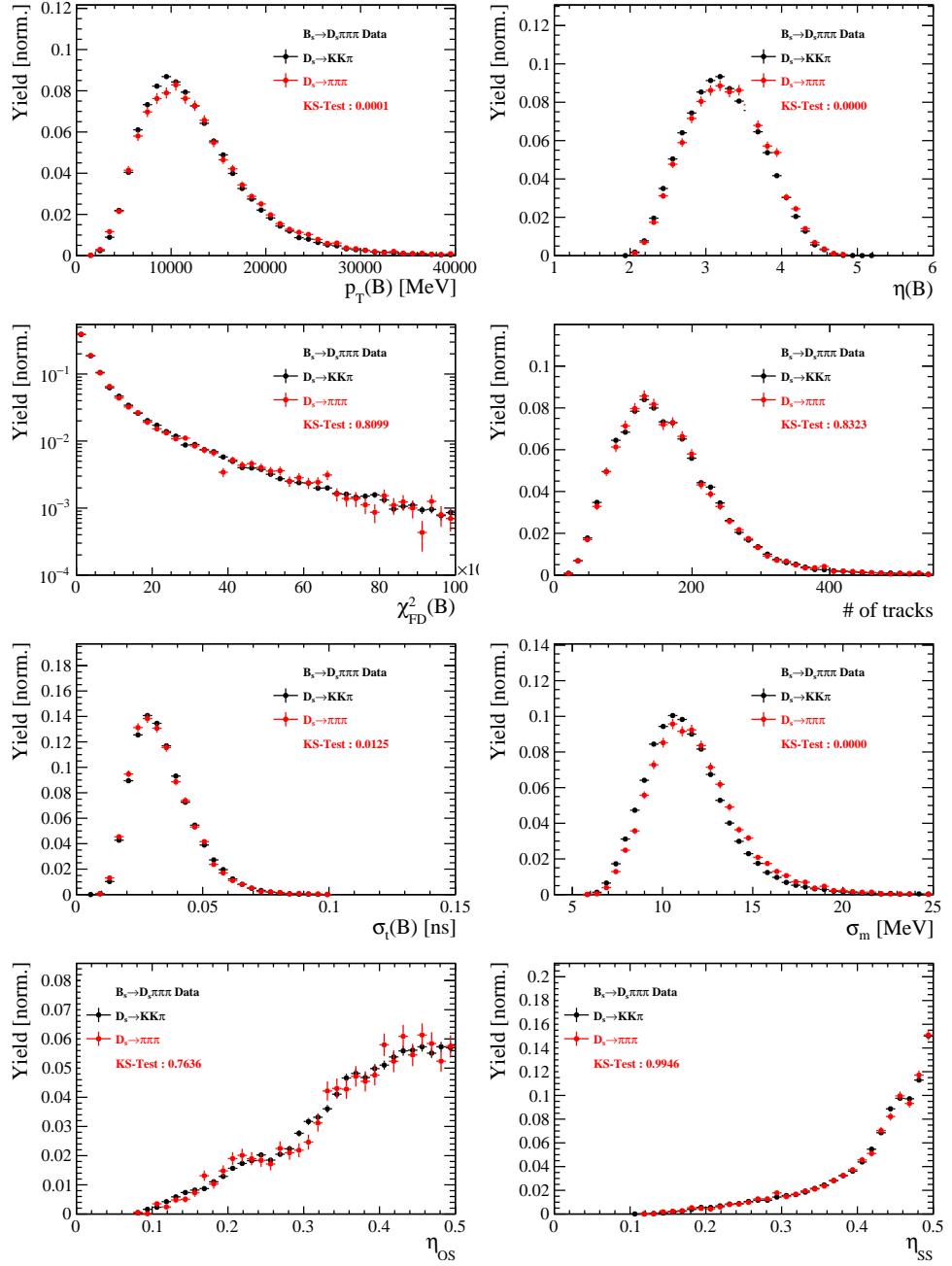


Figure 4.5: Comparison of selected variables.

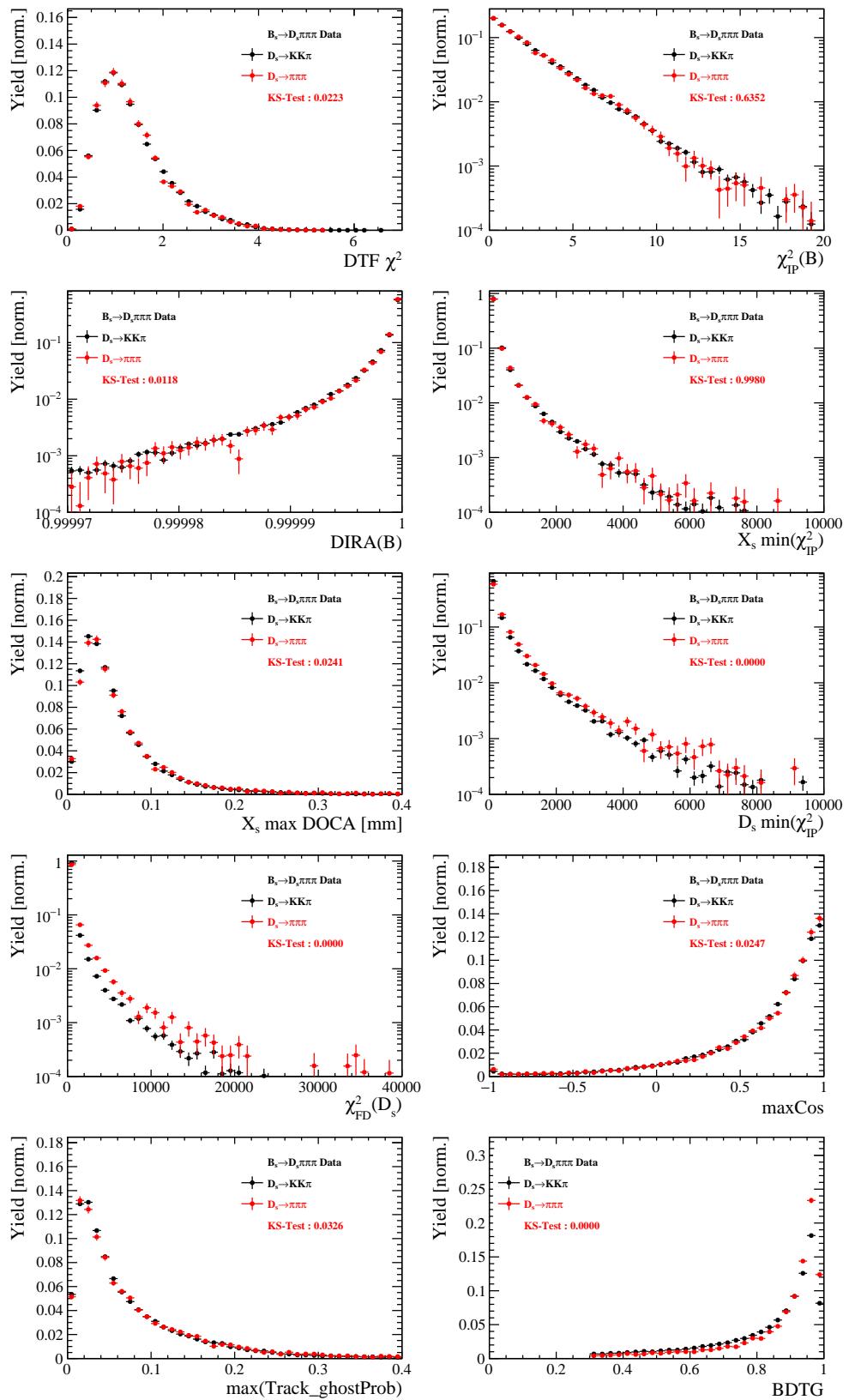


Figure 4.6: Comparison of BDTG input variables and classifier response.

557 D.4 Comparison of trigger categories

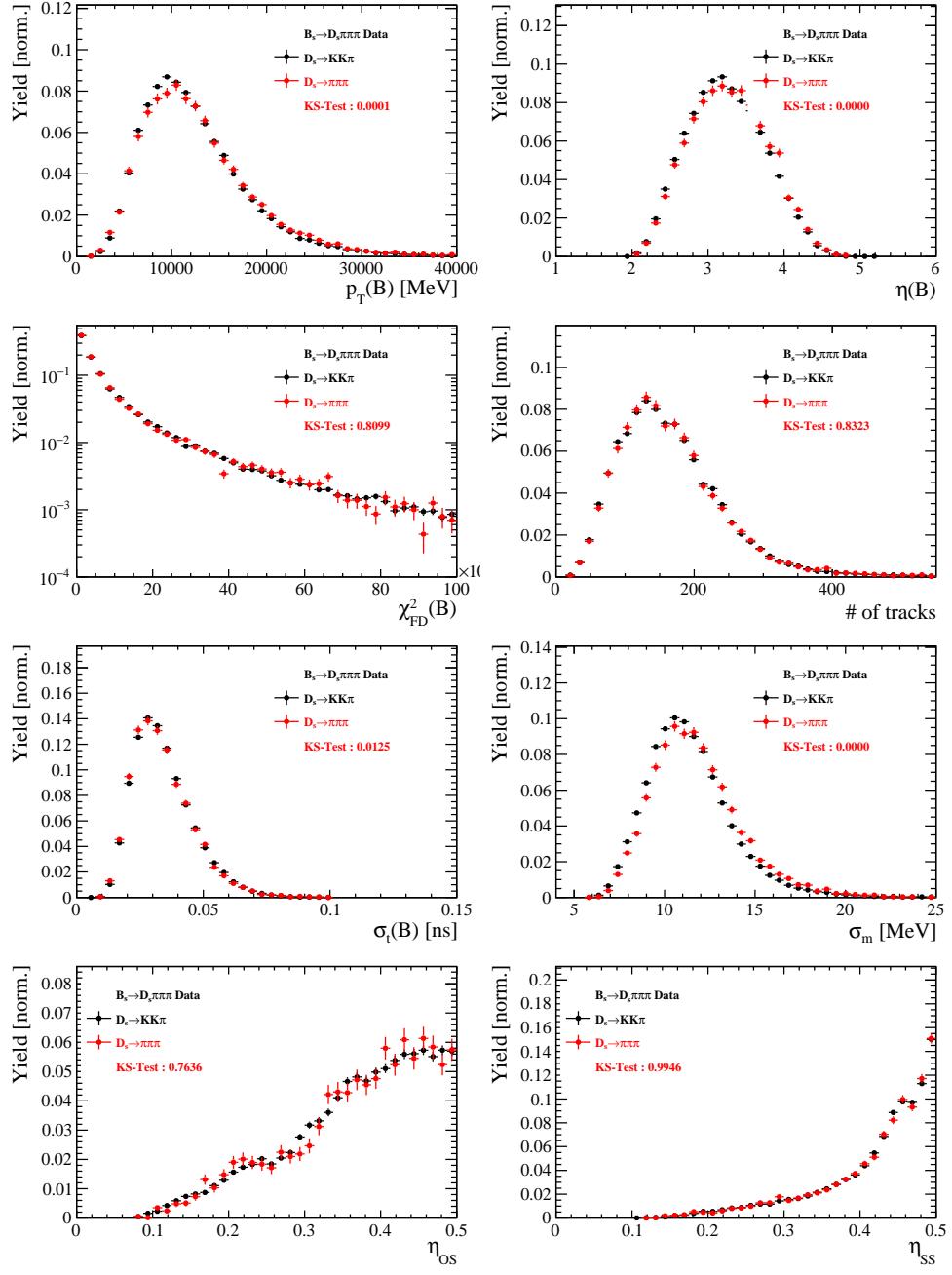


Figure 4.7: Comparison of selected variables.

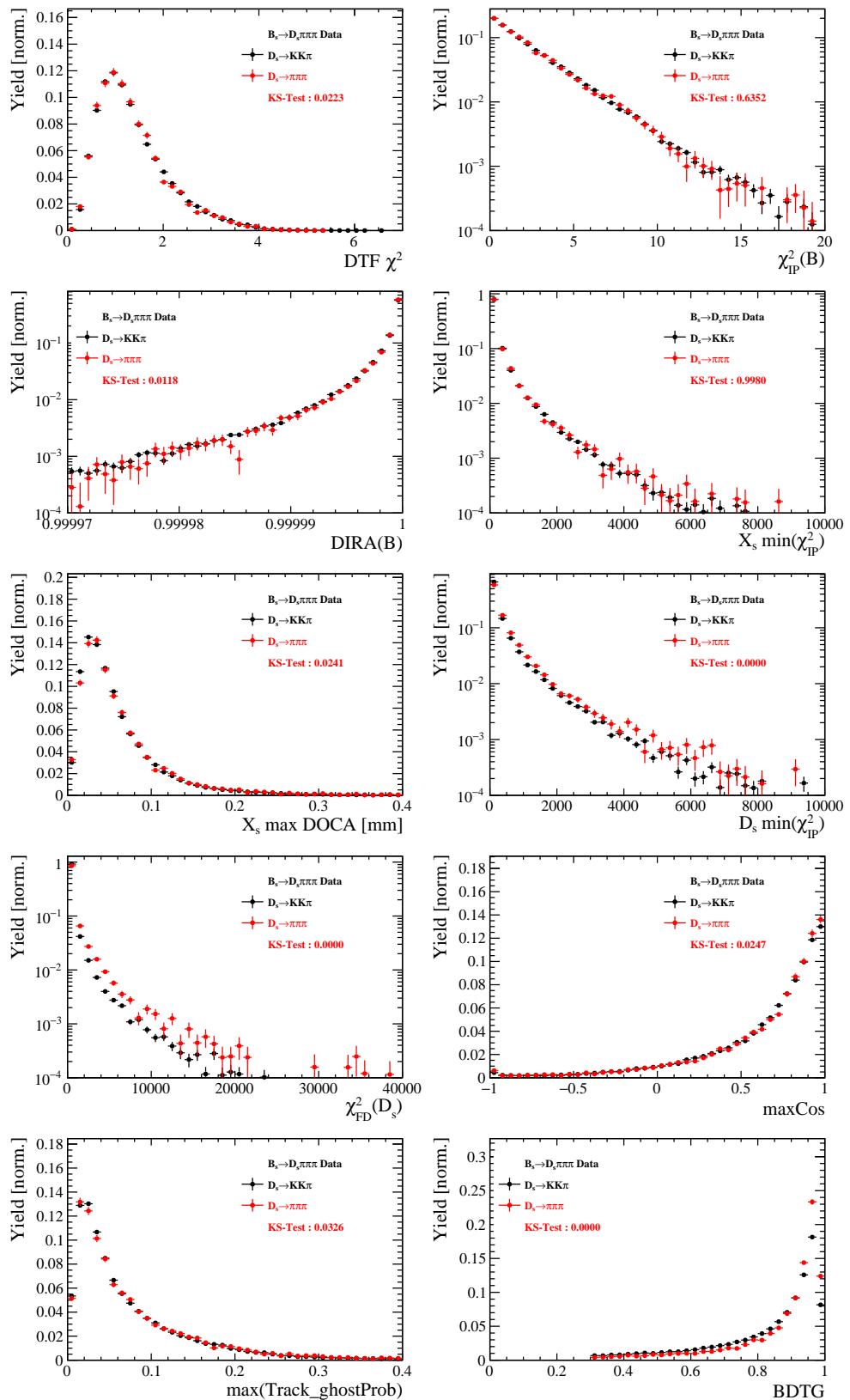


Figure 4.8: Comparison of BDTG input variables and classifier response.

558 **References**

- 559 [1] R. Fleischer, *New strategies to obtain insights into CP violation through  $B(s) \rightarrow D(s)$ ,  $D(s) \rightarrow K\pi$ ,  $D(s) \rightarrow D\pi$ , ... and  $B(d) \rightarrow D\pi$ ,  $D^* \rightarrow D\pi$ , ... decays*, Nucl.  
560 Phys. **B671** (2003) 459, [arXiv:hep-ph/0304027](https://arxiv.org/abs/hep-ph/0304027).
- 562 [2] K. De Bruyn *et al.*, *Exploring  $B_s \rightarrow D_s^{(*)\pm} K^\mp$  Decays in the Presence of a Sizable*  
563 *Width Difference  $\Delta\Gamma_s$* , Nucl. Phys. **B868** (2013) 351, [arXiv:1208.6463](https://arxiv.org/abs/1208.6463).
- 564 [3] S. Blusk, *First observations and measurements of the branching fractions for the*  
565 *decays  $\bar{B}_s^0 \rightarrow D_s^+ K^- \pi^+ \pi^-$  and  $\bar{B}^0 \rightarrow D_s^+ K^- \pi^+ \pi^-$* .
- 566 [4] LHCb, S. Blusk, *Measurement of the CP observables in  $\bar{B}_s^0 \rightarrow D_s^+ K^-$  and first obser-*  
567 *vation of  $\bar{B}_{(s)}^0 \rightarrow D_s^+ K^- \pi^+ \pi^-$  and  $\bar{B}_s^0 \rightarrow D_{s1}(2536)^+ \pi^-$* , 2012. [arXiv:1212.4180](https://arxiv.org/abs/1212.4180).
- 568 [5] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007)  
569 040, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
- 570 [6] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*,  
571 Nucl. Instrum. Meth. **A555** (2005) 356, [arXiv:physics/0402083](https://arxiv.org/abs/physics/0402083).
- 572 [7] N. L. Johnson, *Systems of frequency curves generated by methods of translation*,  
573 Biometrika **36** (1949), no. 1/2 149.
- 574 [8] Particle Data Group, K. A. Olive *et al.*, *Review of particle physics*, Chin. Phys. **C38**  
575 (2014) 090001, and 2015 update.
- 576 [9] LHCb, R. Aaij *et al.*, *A new algorithm for identifying the flavour of  $B_s^0$  mesons at*  
577 *LHCb*, JINST **11** (2016), no. 05 P05010, [arXiv:1602.07252](https://arxiv.org/abs/1602.07252).
- 578 [10] LHCb collaboration, R. Aaij *et al.*, *Opposite-side flavour tagging of  $B$  mesons at the*  
579 *LHCb experiment*, Eur. Phys. J. **C72** (2012) 2022, [arXiv:1202.4979](https://arxiv.org/abs/1202.4979).
- 580 [11] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and*  
581  *$\tau$ -lepton properties as of summer 2014*, [arXiv:1412.7515](https://arxiv.org/abs/1412.7515), updated results and plots  
582 available at <http://www.slac.stanford.edu/xorg/hfag/>.
- 583 [12] T. M. Karbach, G. Raven, and M. Schiller, *Decay time integrals in neutral meson*  
584 *mixing and their efficient evaluation*, [arXiv:1407.0748](https://arxiv.org/abs/1407.0748).
- 585 [13] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys.  
586 **A30** (2015) 1530022, [arXiv:1412.6352](https://arxiv.org/abs/1412.6352).
- 587 [14] LHCb, R. Aaij *et al.*, *Measurement of CP asymmetry in  $B_s^0 \rightarrow D_s^\mp K^\pm$  decays*,  
588 Submitted to: JHEP (2017) [arXiv:1712.07428](https://arxiv.org/abs/1712.07428).