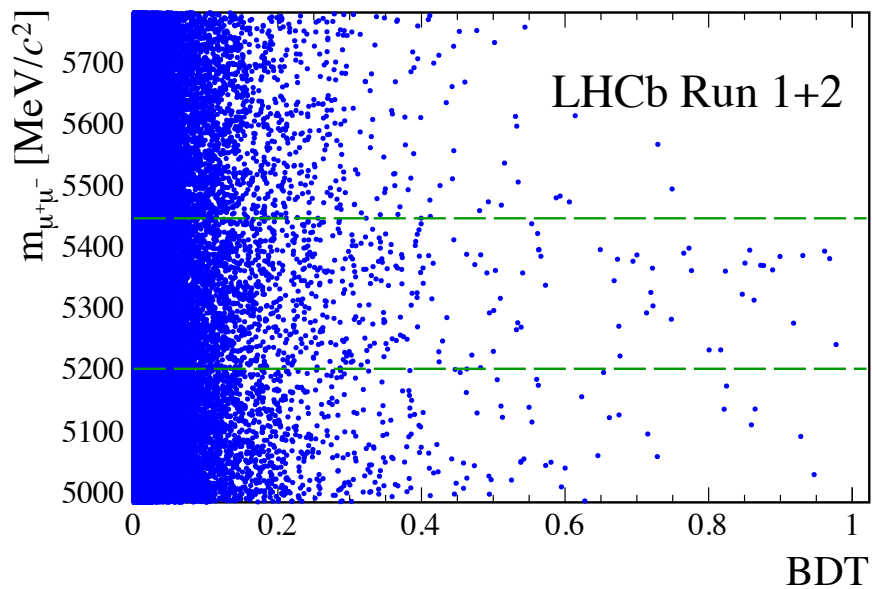# 1 Supplementary material

1.1 Selection

Figure 3: A scatter plot showing the mass and BDT distributions of events used for the analysis, for the full dataset used in Run 1 and Run 2. The BDT variable is shown on the x-axis and the invariant mass of the two muons is shown on the y-axis. The dashed green line indicates the signal mass region, for which $m_{\mu^+\mu^-} \in [5200, 5445]\,\mathrm{MeV}/c^2$.
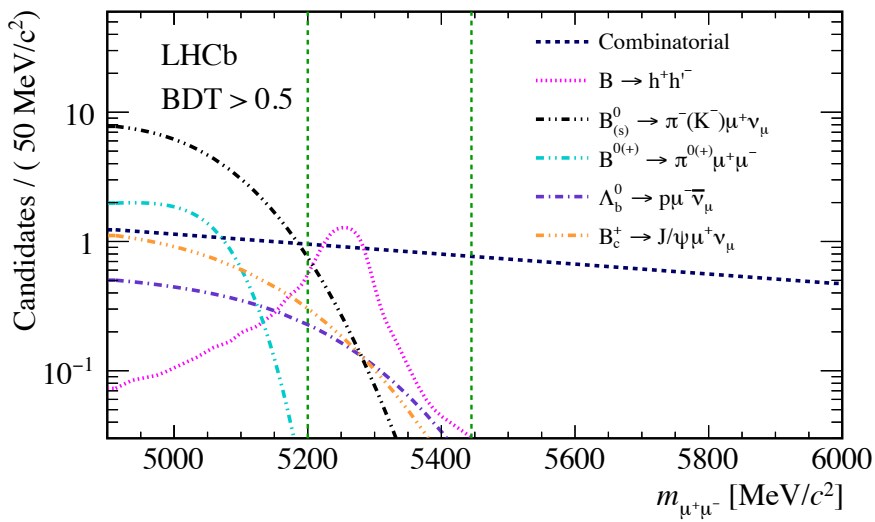
Figure 4: Mass distribution of the main background sources: combinatorial background (blue short dashed), $B^0_{(s)} \to h^+ h'^-$ (magenta dotted), $B^0 \to \pi^- \mu^+ \nu_\mu$ and $B^0_s \to K^- \mu^+ \nu_\mu$ (black dot-dashed), $B^{0(+)} \to \pi^{0(+)} \mu^+ \mu^-$ (cyan dot-dashed), $B^+_c \to J/\psi \mu^+ \nu_\mu$ (orange dot-dashed) and $\Lambda^0_b \to p \mu^- \bar{\nu}_\mu$ (violet dot-dashed). The dashed green line indicates the signal mass region, for which $m_{\mu^+\mu^-} \in [5200, 5445] \, \mathrm{MeV}/c^2$.
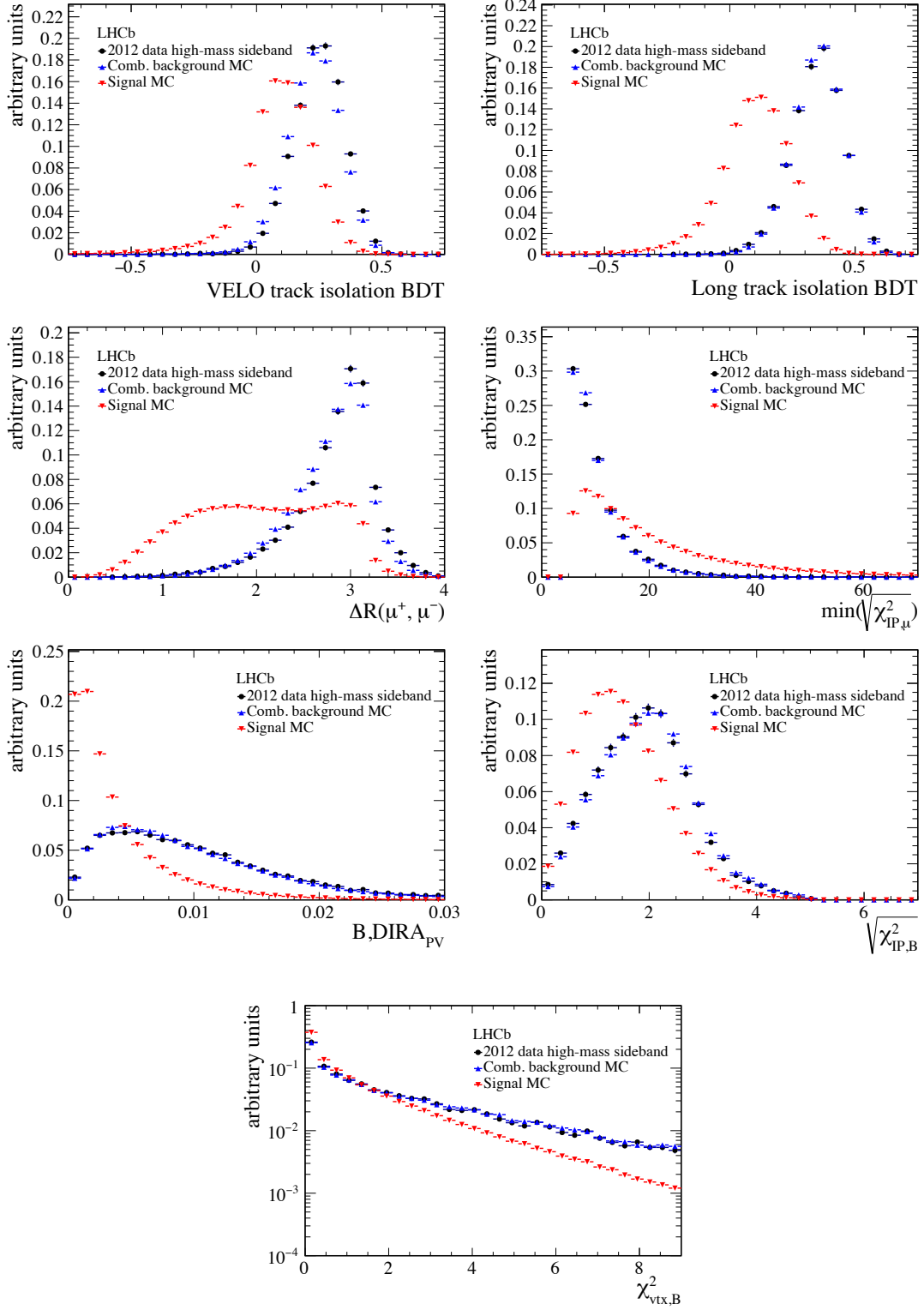
Figure 5: BDT input variable distributions in 2012 data high-mass sideband candidates (black dots), simulated combinatorial background events (blue triangles) and simulated $B_s^0 \to \mu^+\mu^-$ signal events (red triangles). The tracks used for Long track isolation BDT are defined as those that have been reconstructed both before and after the magnet.
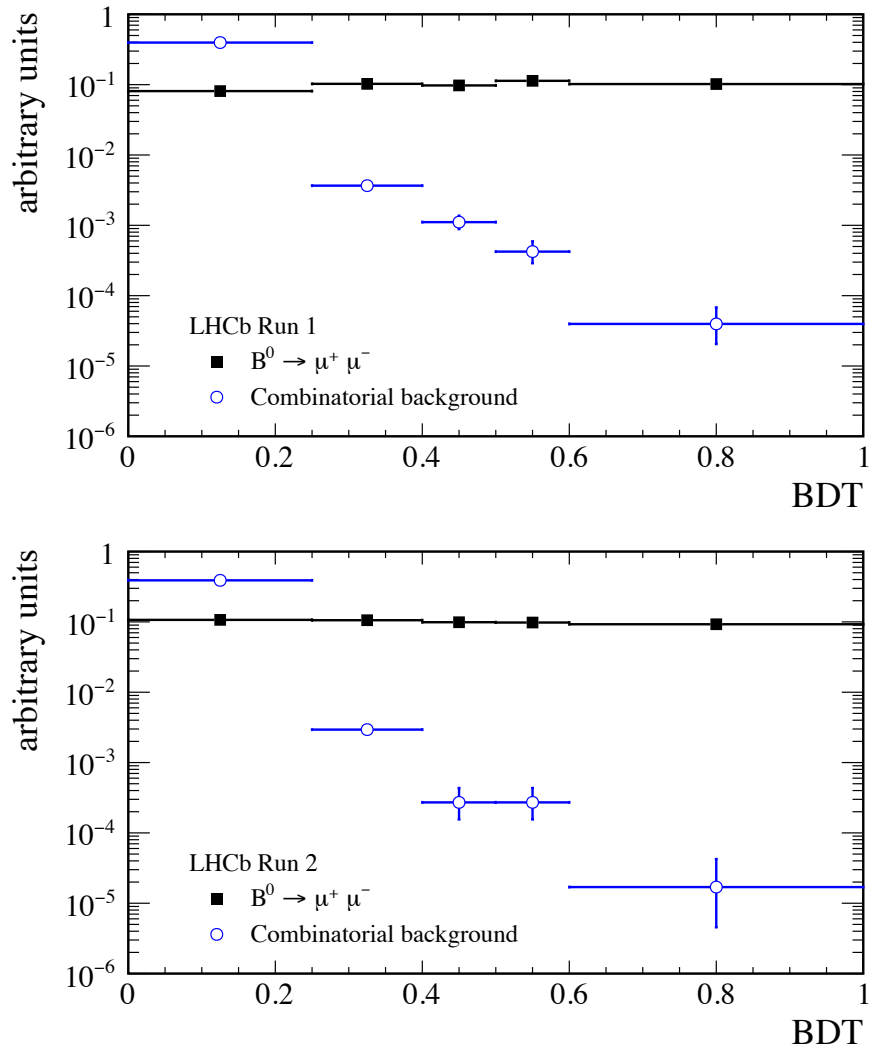
3

1.2 Calibration of the signal BDT distribution

Figure 6: Expected BDT distribution in (top) Run 1 and (bottom) Run 2 data for $B^0 \to \mu^+\mu^-$ signal calibrated with $B^0 \to K^+\pi^-$ control channel (black squares) and combinatorial background from high-mass sidebands (blue circles).
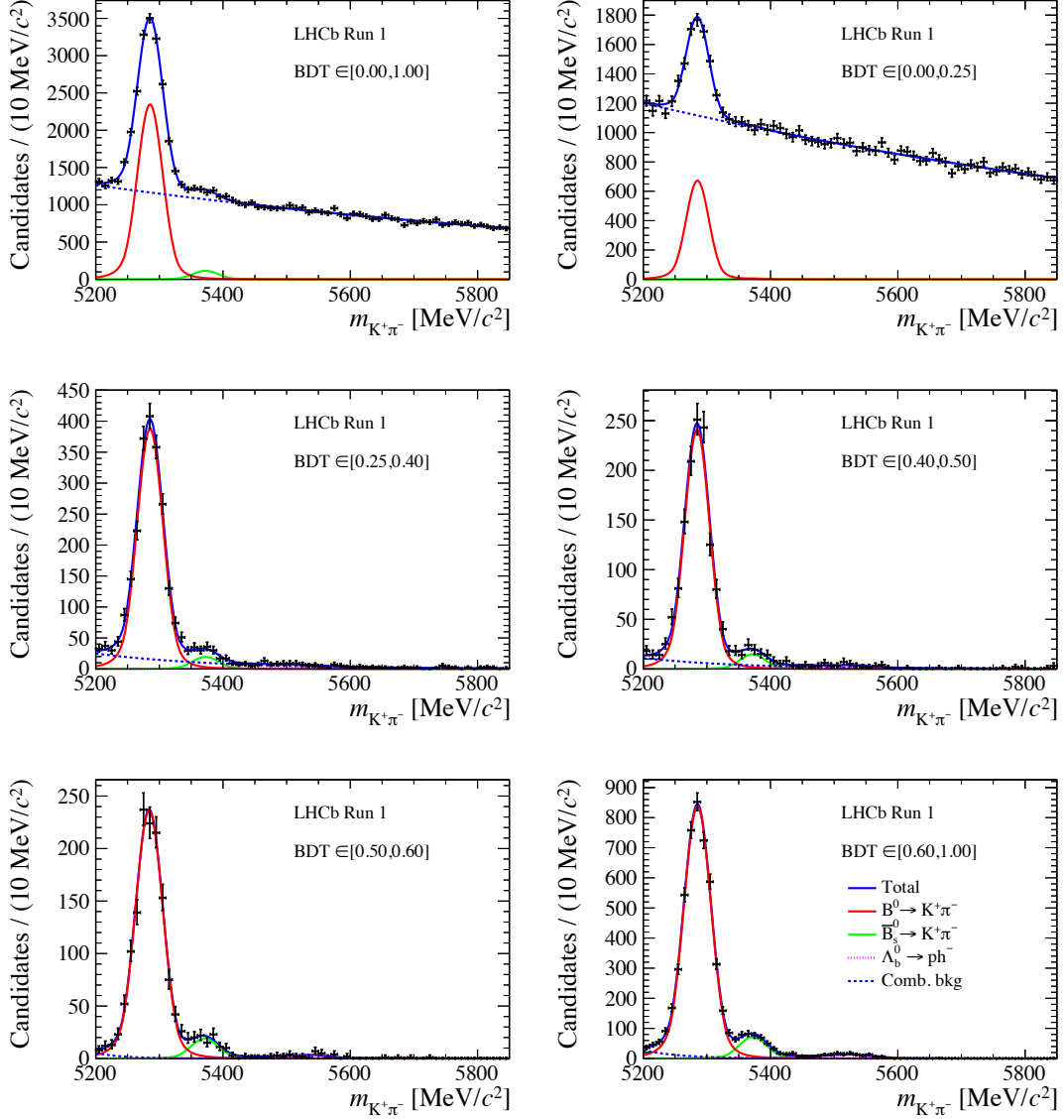
Figure 7: Mass distribution of the selected $B^0 \to K^+\pi^-$ candidates (black dots) in each BDT bin of Run 1 data, used to evaluate the BDT distribution for $B^0_{(s)} \to \mu^+\mu^-$. The result of the fit is overlaid (blue solid line) and the different components are detailed: $B^0 \to K^+\pi^-$ (red solid line), $\overline{B}^0_s \to K^+\pi^-$ (green solid line), $\Lambda^0_b \to ph^-$ (magenta dotted line) and combinatorial background (blue dashed line).
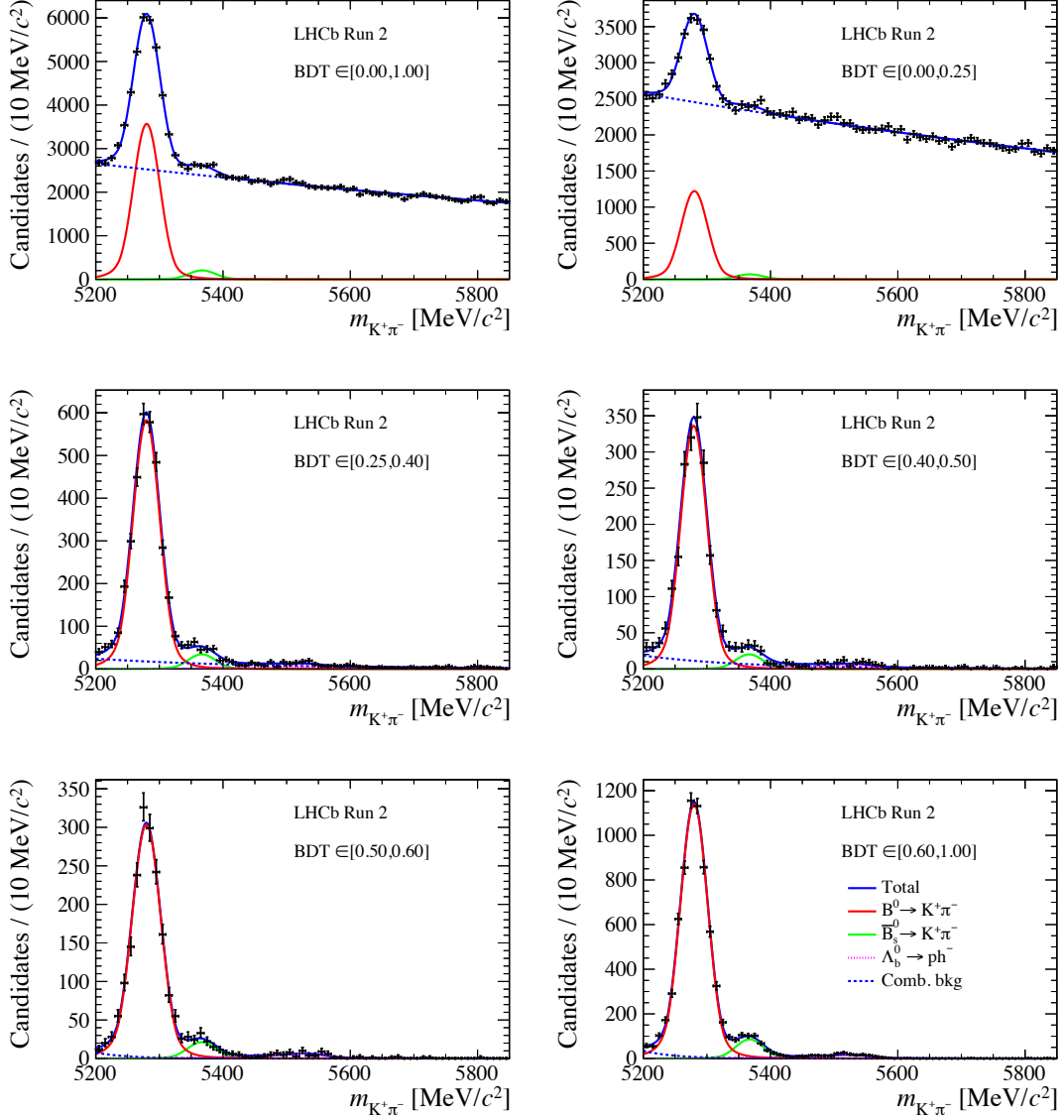
Figure 8: Mass distribution of the selected $B^0 \to K^+\pi^-$ candidates (black dots) in each BDT bin of Run 2 data, used to evaluate the BDT distribution for $B^0_{(s)} \to \mu^+\mu^-$. The result of the fit is overlaid (blue solid line) and the different components are detailed: $B^0 \to K^+\pi^-$ (red solid line), $\overline{B}^0_s \to K^+\pi^-$ (green solid line), $\Lambda^0_b \to ph^-$ (magenta dotted line) and combinatorial background (blue dashed line).
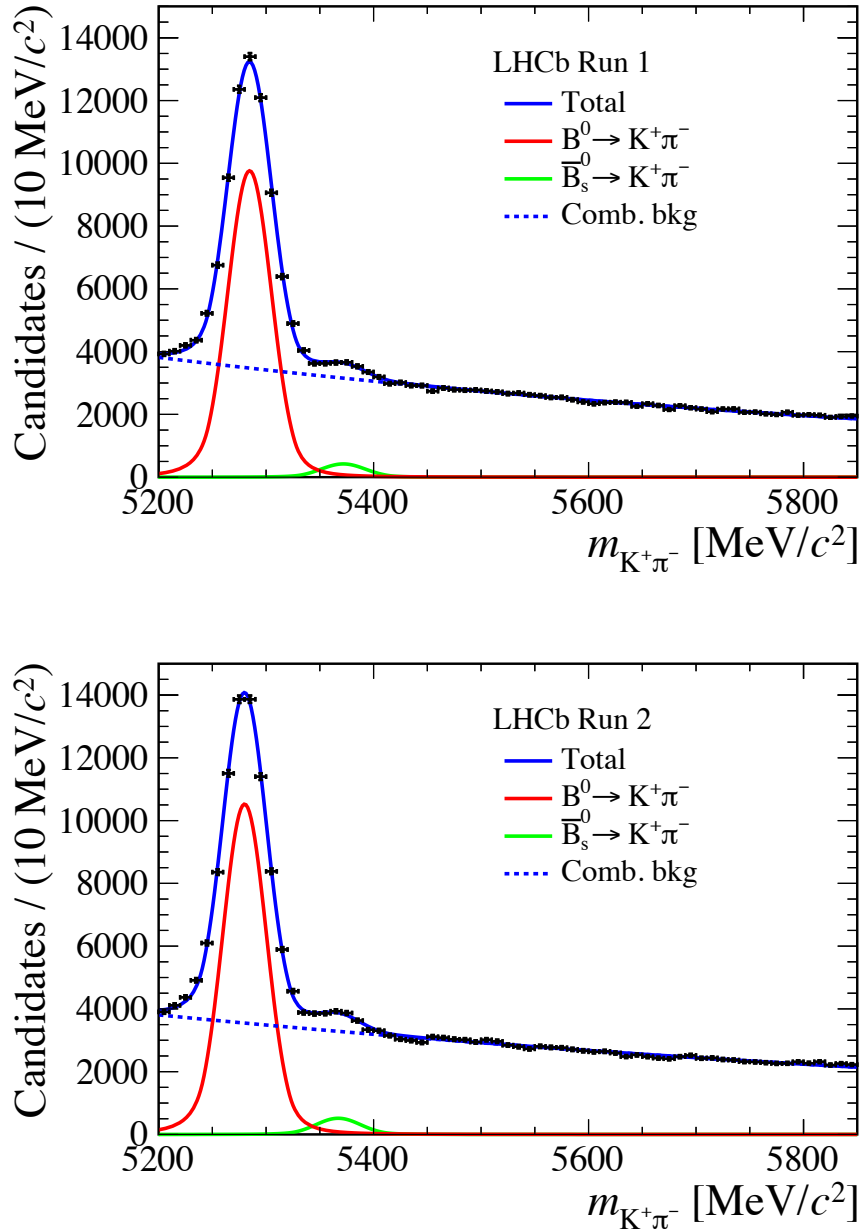
6

1.3   Calibration of the signal mass distribution

Figure 9: Mass distribution of the selected $B^0 \to K^+\pi^-$ candidates (black dots) in (top) Run 1 and (bottom) Run 2 data. The result of the fit to determine the peak value of the $B^0 \to \mu^+\mu^-$ mass distribution is overlaid (blue solid line) and the different components are detailed: $B^0 \to K^+\pi^-$ (red solid line), $\bar{B}_s^0 \to K^+\pi^-$ (green solid line) and combinatorial background (blue dashed line). Candidates shown in this plot do not have to pass any explicit trigger requirements.
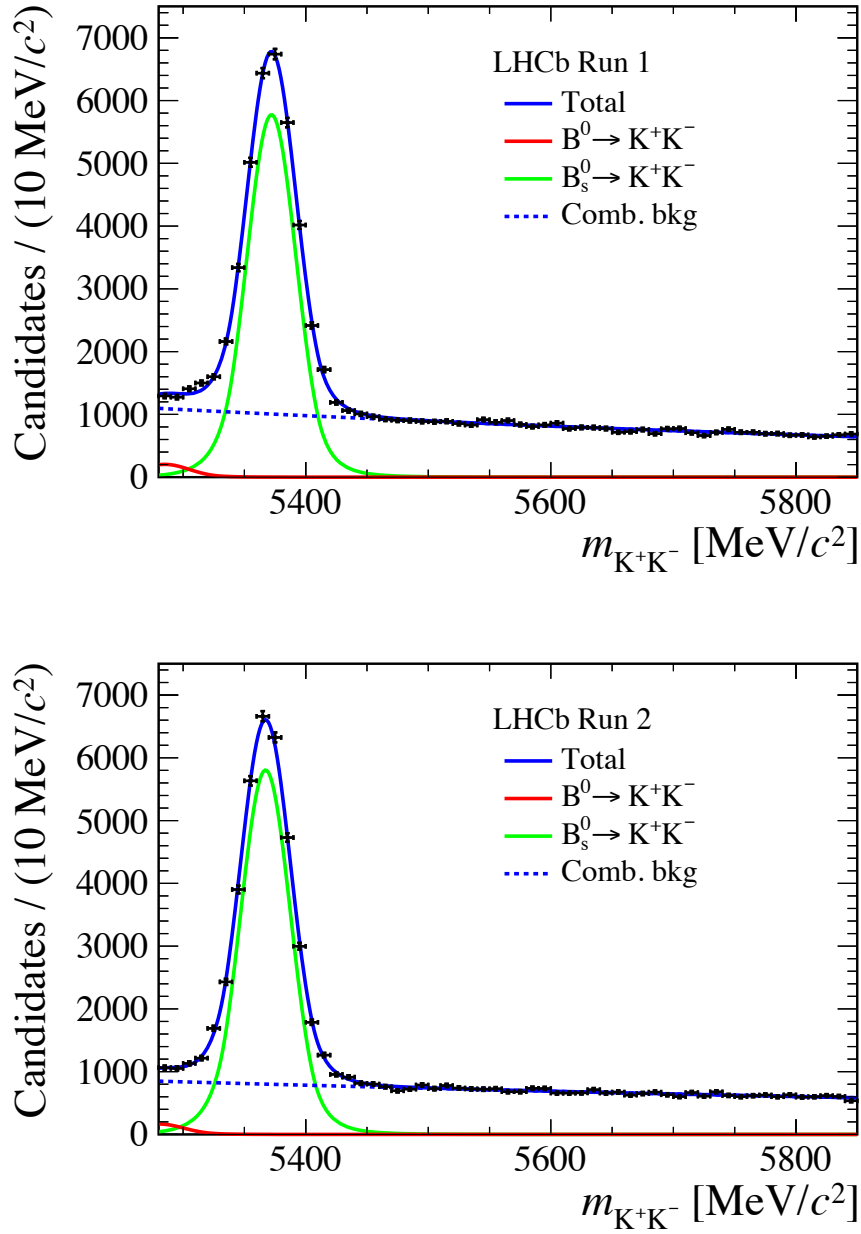
Figure 10: Mass distribution of the selected $B_s^0 \to K^+K^-$ candidates (black dots) in (top) Run 1 and (bottom) Run 2 data. The result of the fit to determine the peak value of the $B_s^0 \to \mu^+\mu^-$ mass distribution is overlaid (blue solid line) and the different components are detailed: $B^0 \to K^+K^-$ (red solid line), $B_s^0 \to K^+K^-$ (green solid line) and combinatorial background (blue dashed line). Candidates shown in this plot do not have to pass any explicit trigger requirements.
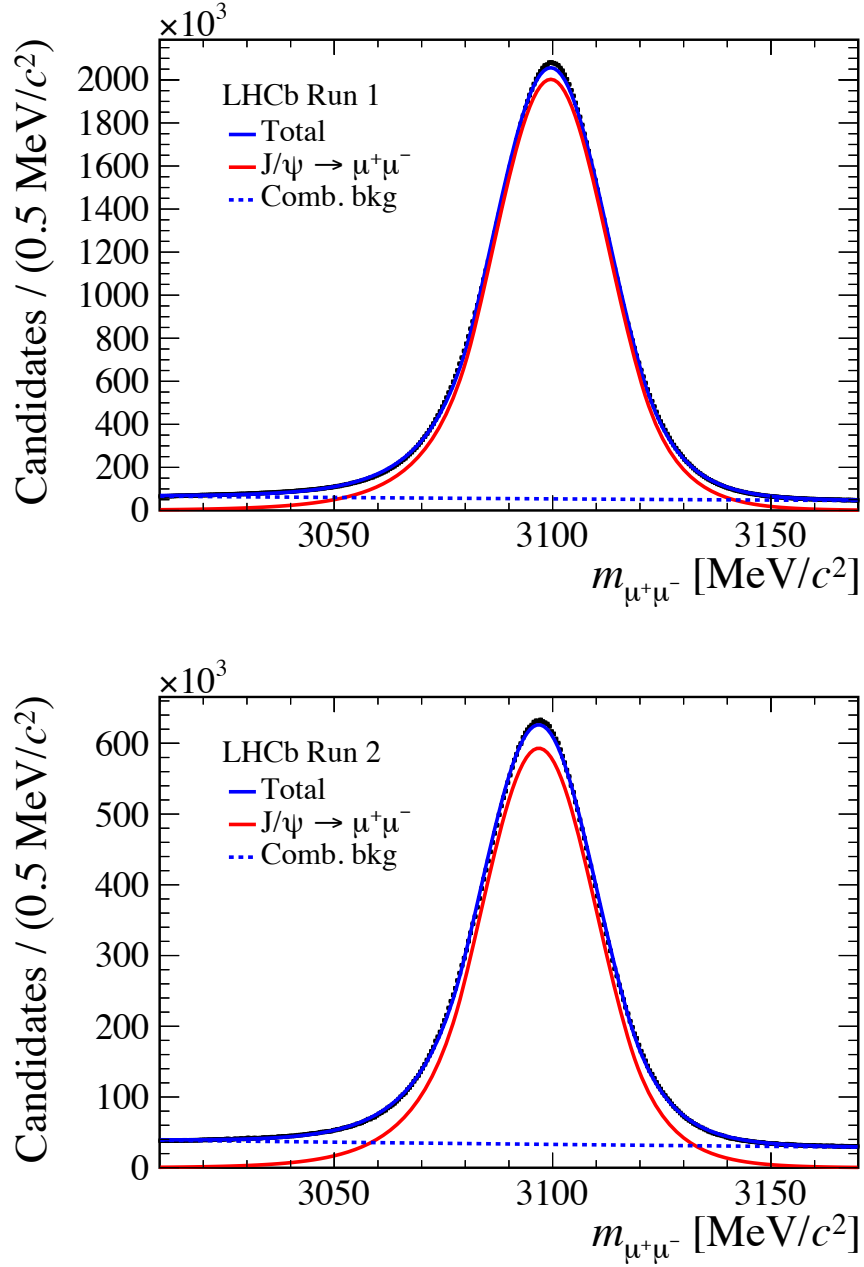
Figure 11: Mass distribution of $J/\psi \to \mu^+\mu^-$ candidates in (top) Run 1 and (bottom) Run 2 data. The widths of the distributions are used for the interpolation in Figure 14.
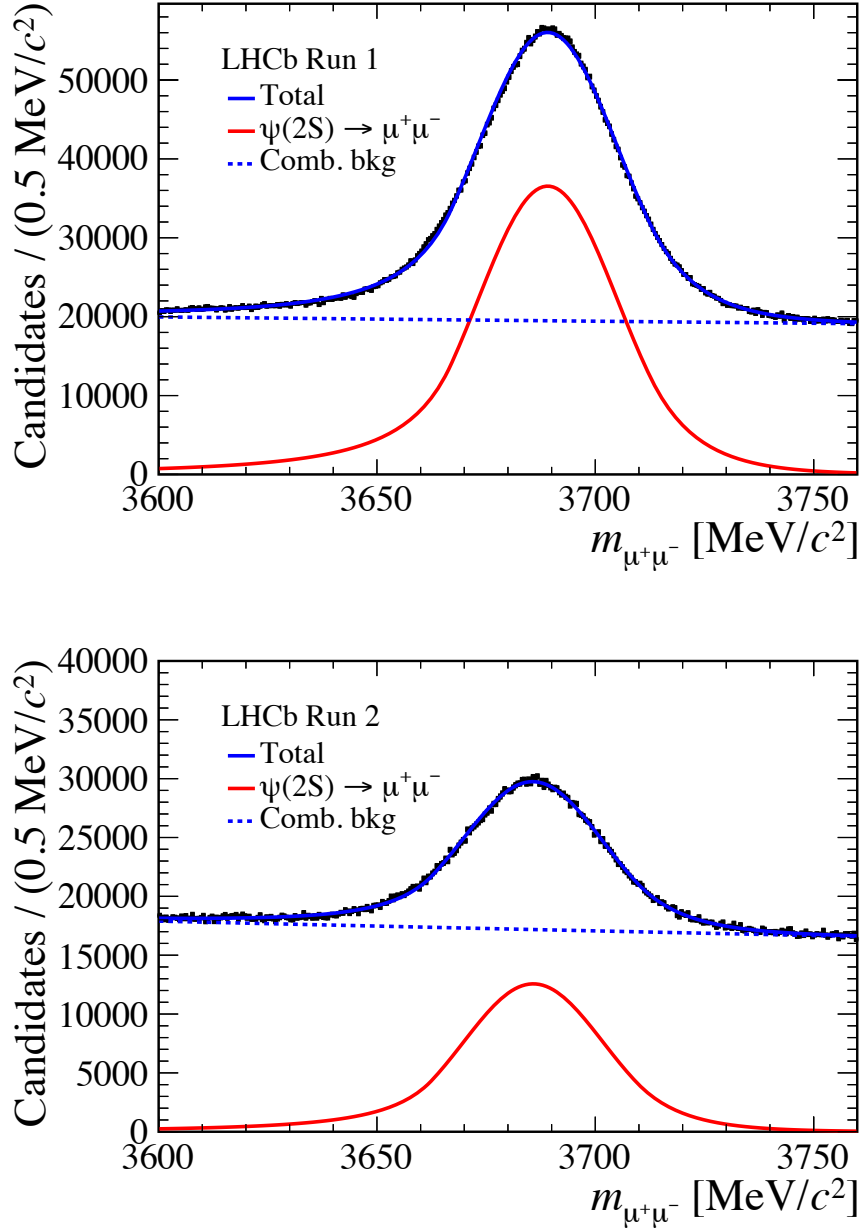
Figure 12: Mass distribution of $\psi(2S) \to \mu^+\mu^-$ candidates in (top) Run 1 and (bottom) Run 2 data. The widths of the distributions are used for the interpolation in Figure 14.
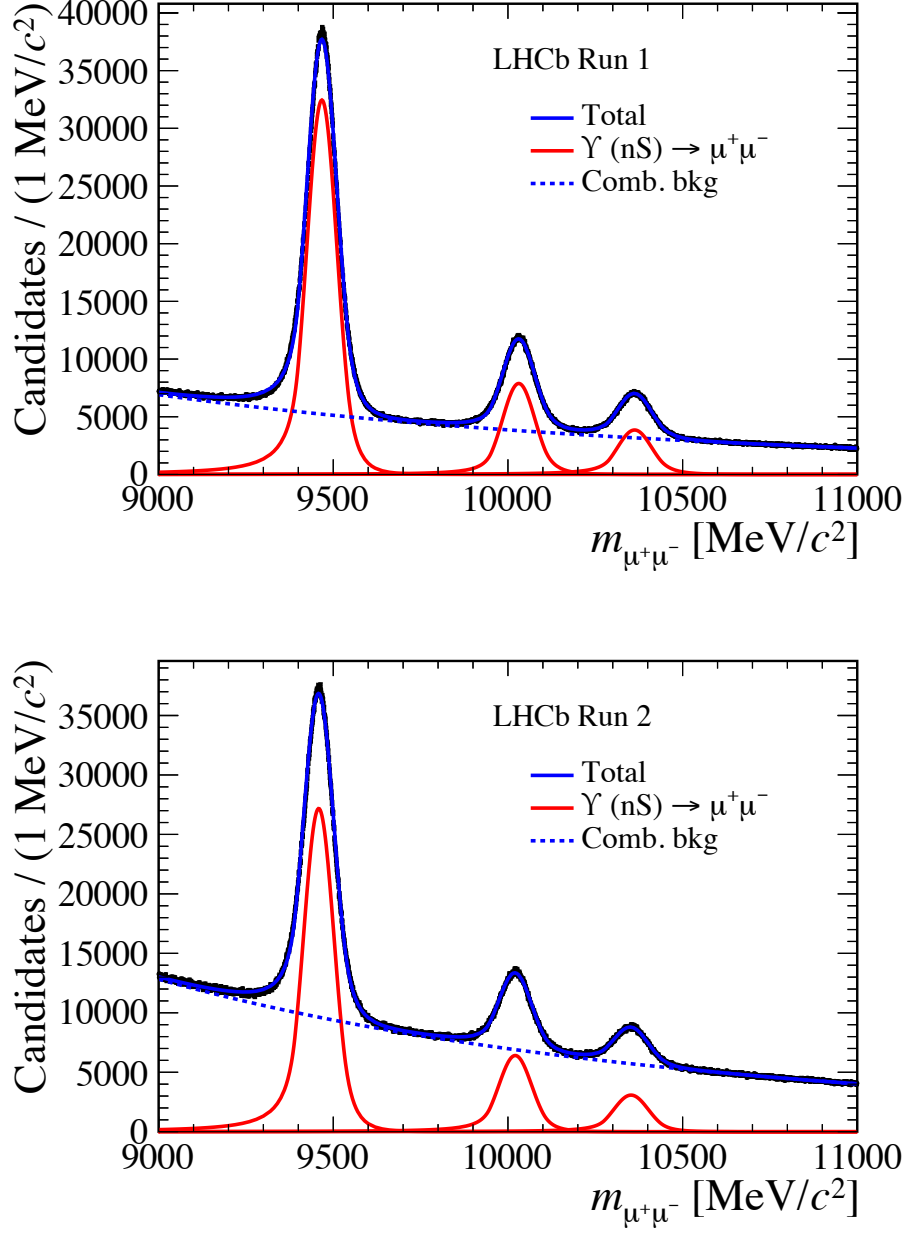
Figure 13: Mass distribution of $\Upsilon(1S, 2S, 3S) \to \mu^+\mu^-$ candidates in (top) Run 1 and (bottom) Run 2 data. The widths of the distributions are used for the interpolation in Figure 14.
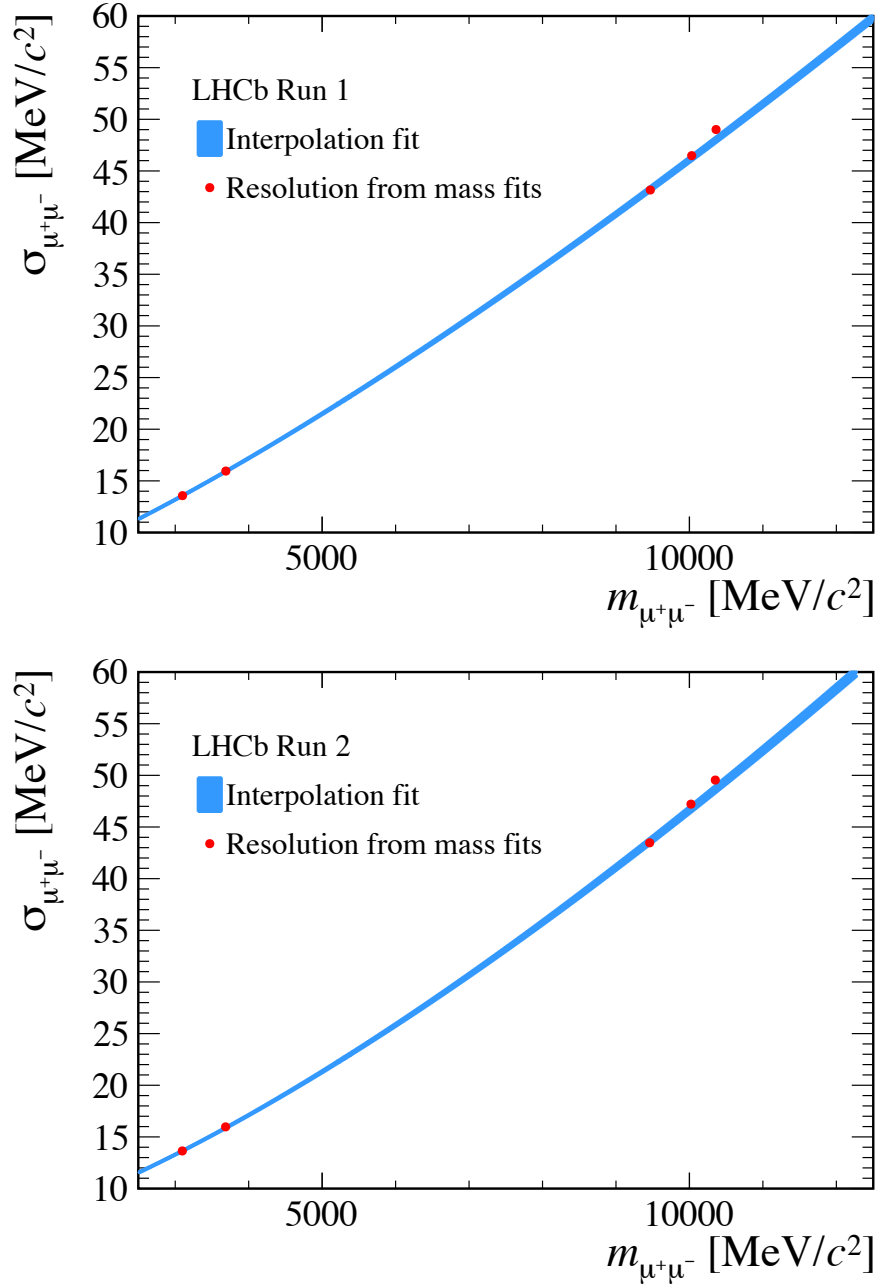
Figure 14: Fit of a power-law to the widths of charmonium and bottomonium resonances in (top) Run 1 and (bottom) Run 2 data.
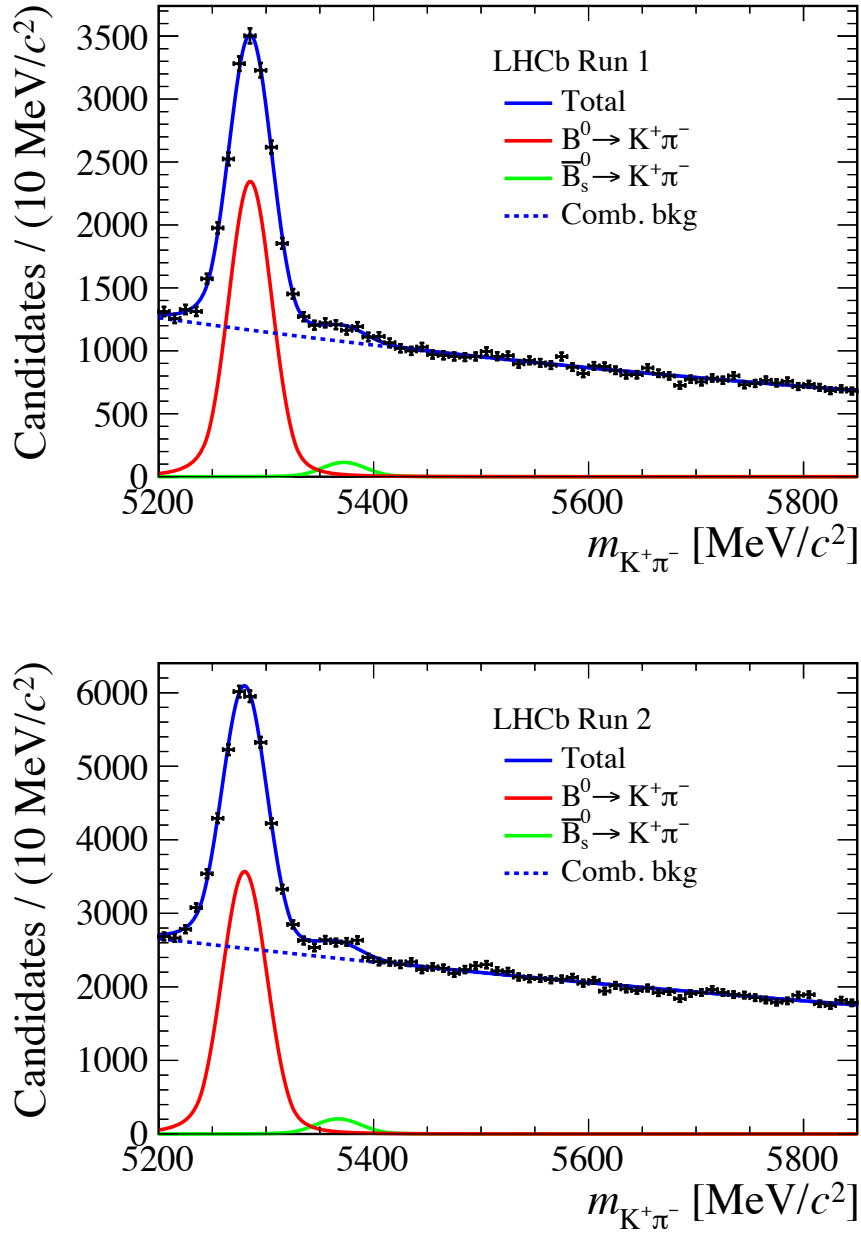
1.4 Normalisation

Figure 15: Mass distribution of the selected $B^0 \to K^+\pi^-$ candidates (black dots) in (top) Run 1 and (bottom) Run 2 data. The result of the fit to determine the normalisation yield is overlaid (blue solid line) and the different components are detailed: $B^0 \to K^+\pi^-$ (red solid line), $\overline{B}^0_s \to K^+\pi^-$ (green solid line) and combinatorial background (blue dashed line). Candidates shown in this plot are triggered almost independently of signal.
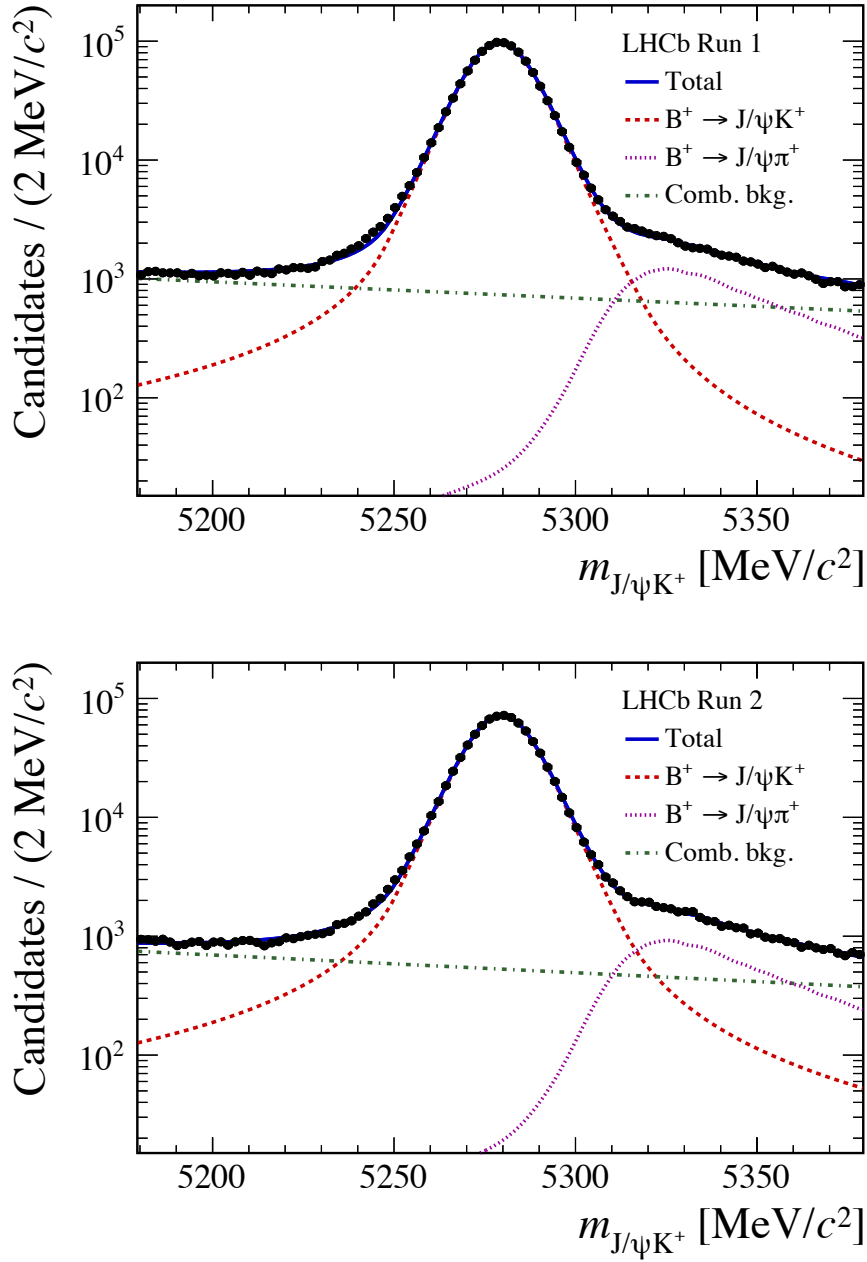
Figure 16: Mass distribution of the selected $B^+ \to J/\psi K^+$ candidates in (top) Run 1 and (bottom) Run 2 data. The result of the fit to determine the normalisation yield is overlaid and the different components are detailed: $B^+ \to J/\psi K^+$ (red dotted line), $B^+ \to J/\psi \pi^+$ (purple dashed line), and combinatorial background (green dashed-dotted line). Candidates shown in this plot are triggered almost independently of signal.
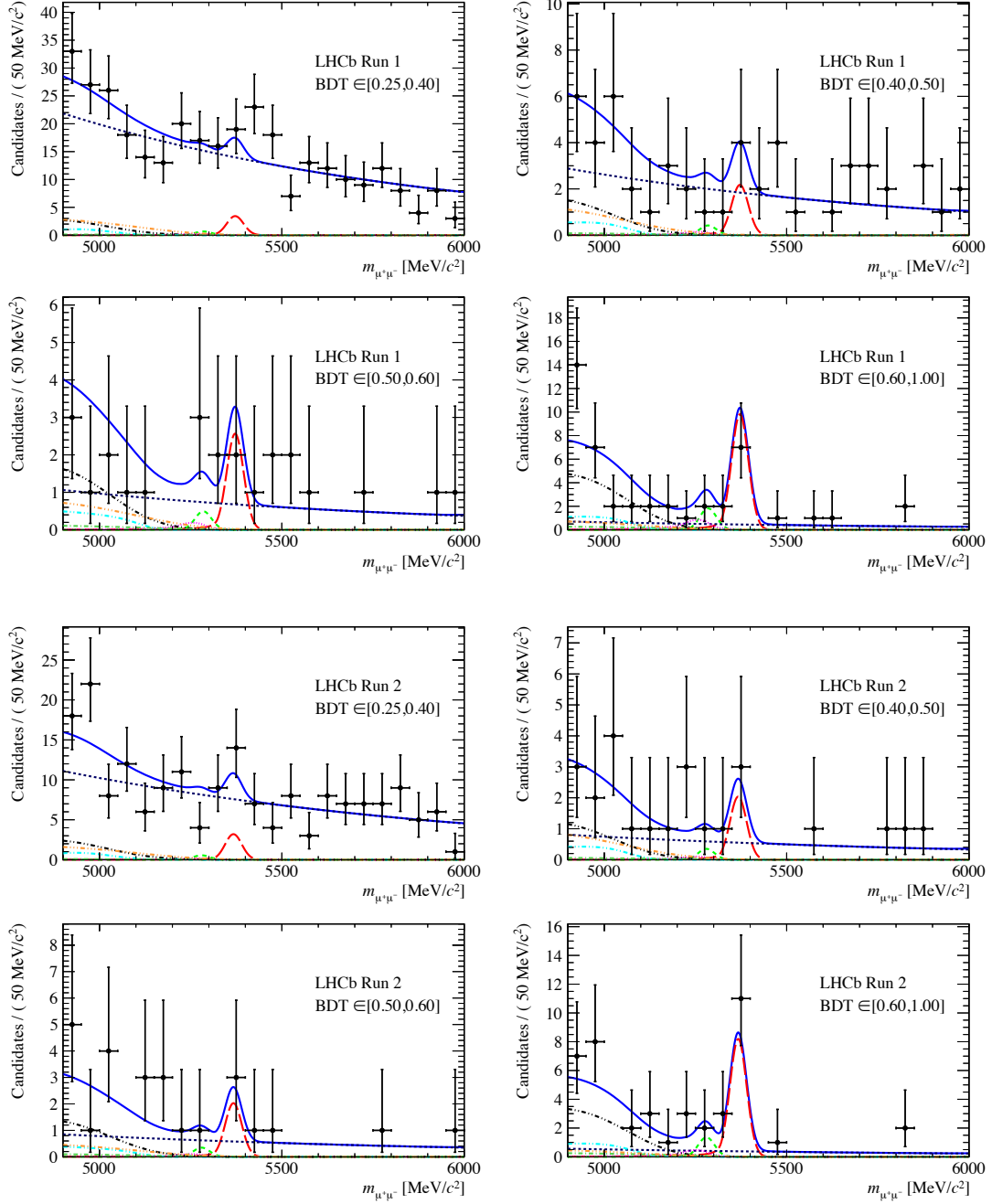
1.5 Branching fraction analysis results

Figure 17: Mass distributions of the selected $B^0_{(s)} \to \mu^+\mu^-$ candidates (black dots) in bins of BDT. The result of the fit is overlaid (blue solid line) and the different components are detailed: $B^0_s \to \mu^+\mu^-$ (red long dashed), $B^0 \to \mu^+\mu^-$ (green medium dashed), combinatorial background (blue short dashed), $B^0_{(s)} \to h^+ h'^-$ (magenta dotted), $B^0 \to \pi^-\mu^+\nu_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ (black dot-dashed), $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ (cyan dot-dashed), $B^+_c \to J/\psi\mu^+\nu_\mu$ (orange dot-dashed) and $\Lambda^0_b \to p\mu^-\bar{\nu}_\mu$ (violet dot-dashed).
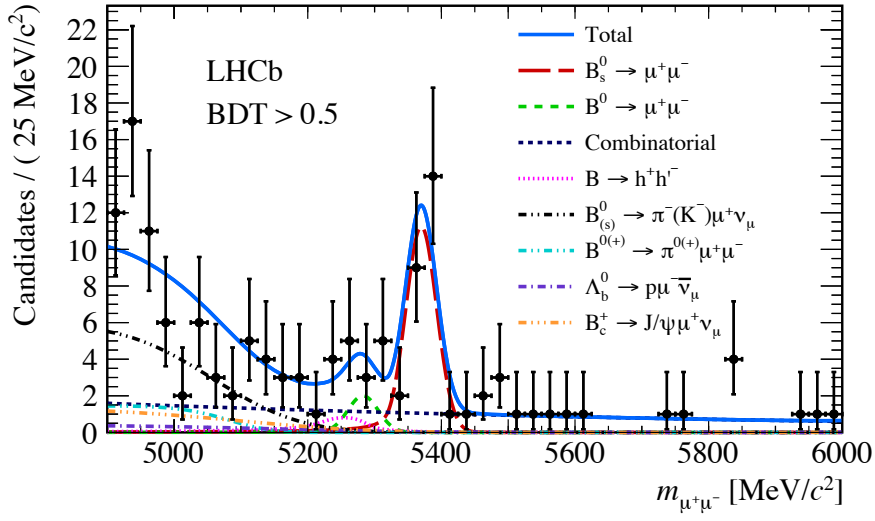
Figure 18: Mass distribution of the selected $B^0_{(s)} \to \mu^+\mu^-$ candidates (black dots) with BDT > 0.5 and twice the number of mass bins shown in Fig.1. The result of the fit is overlaid (blue solid line) and the different components detailed: $B^0_s \to \mu^+\mu^-$ (red long dashed), $B^0 \to \mu^+\mu^-$ (green medium dashed), combinatorial background (blue short dashed), $B^0_{(s)} \to h^+h'^-$ (magenta dotted), $B^0 \to \pi^-\mu^+\nu_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ (black dot-dashed), $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ (cyan dot-dashed), $B^+_c \to J/\psi\mu^+\nu_\mu$ (orange dot-dashed) and $\Lambda^0_b \to p\mu^-\bar{\nu}_\mu$ (violet dot-dashed).

━━━━━ Total

━ ━ ━ ⋅ $B_s^0 \to \mu^+\mu^-$

⋯⋯⋯⋯ $B^0 \to \mu^+\mu^-$

⋯⋯⋯⋯ Combinatorial

⋯⋯⋯⋯ $B \to h^+h'^-$

⋯⋯⋯⋯ $B_{(s)}^0 \to \pi^-(K^-)\mu^+\nu_\mu$

⋯⋯⋯⋯ $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$

⋯⋯⋯⋯ $\Lambda_b^0 \to p\mu^-\overline{\nu}_\mu$

⋯⋯⋯⋯ $B_c^+ \to J/\psi\mu^+\nu_\mu$
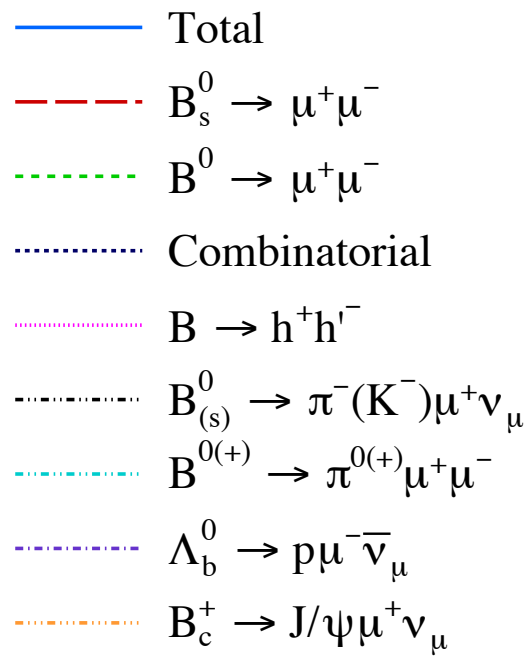
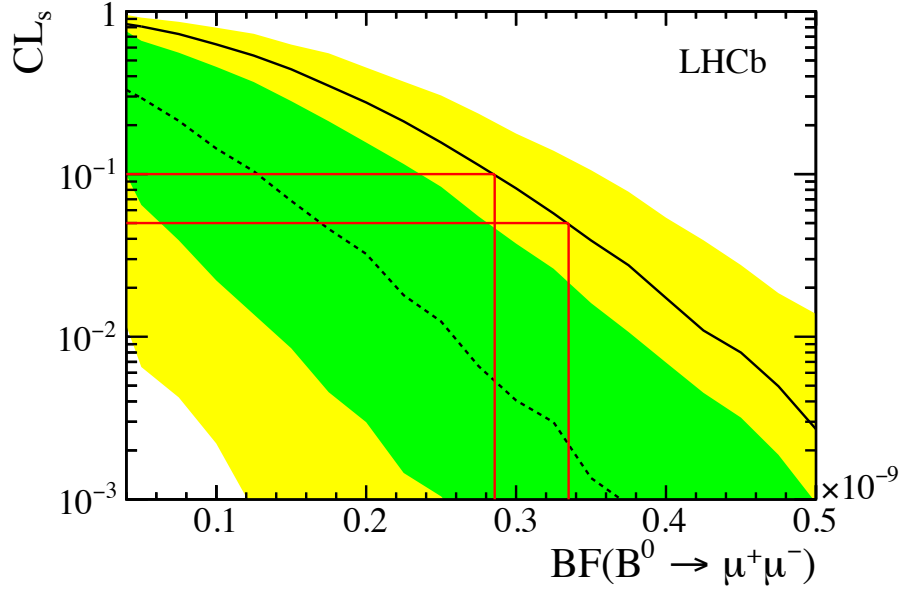Figure 19: Legend for the fit to the mass distribution of the selected $B_{(s)}^0 \to \mu^+\mu^-$ candidates.

Figure 20: Results from the CL$_s$ scan used to obtain the limit on $\mathcal{B}(B^0 \to \mu^+\mu^-)$. The background-only expectation is shown by the dotted line and the 1- and 2-σ bands are shown as green and yellow bands respectively. The observation is shown as the solid black line. The two red lines intersecting with the observation indicate the limits at 90% and 95% CL for the upper and lower line respectively.
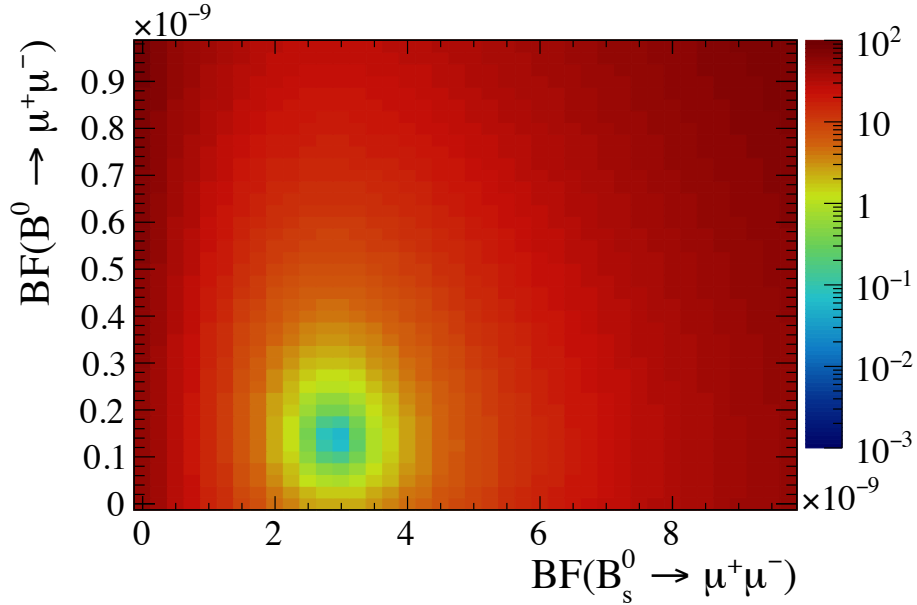

Figure 21: $-2\Delta \ln L$, where $\Delta \ln L$ is the difference between the log-likelihood values evaluated in the point under consideration in the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ versus $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ plane and in the best-fit point. This histogram is used to evaluate the contour levels in Fig. 22.

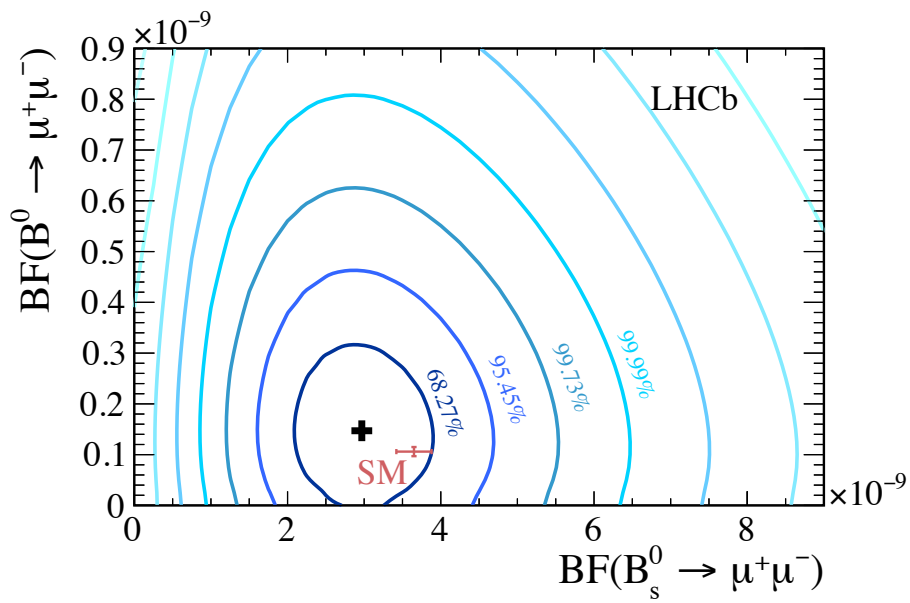Figure 22: A 2 dimensional representation of the branching fraction measurements for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$. The Standard Model value is shown as the red cross labeled SM. The central value from the branching fraction measurement is indicated with the black plus sign. The profile likelihood contours for 1,2,3,... σ are shown as blue contours.

1.6 Effective lifetime of $B^0 \to K^+\pi^-$ decays
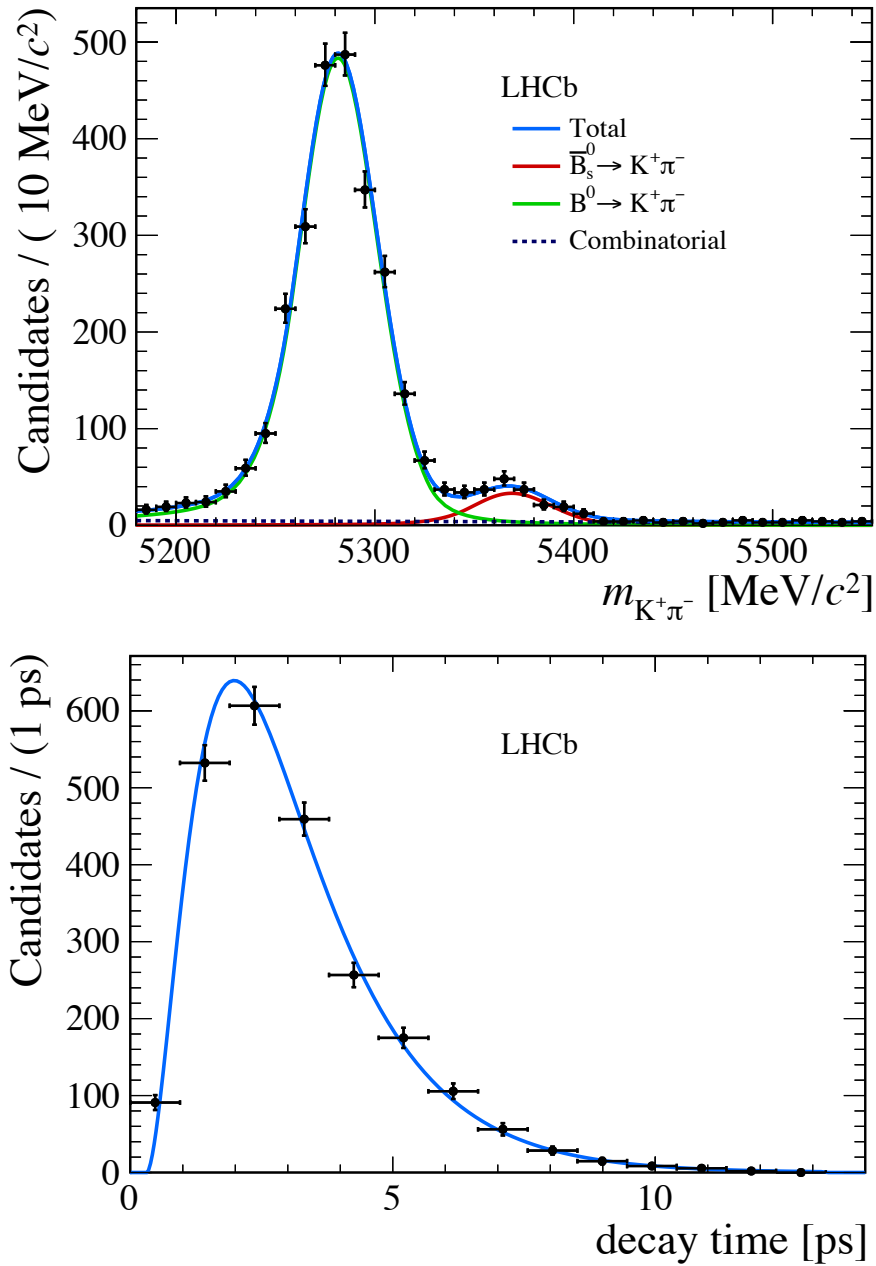

Figure 23: Extended unbinned maximum likelihood fit to the invariant mass distribution of $B^0 \to K^+\pi^-$ candidates (top), which is then used to compute *sWeights*, and unbinned maximum likelihood fit to the sWeighted decay time distribution of $B^0 \to K^+\pi^-$ candidates (bottom) used to determine the $B^0 \to K^+\pi^-$ lifetime. The invariant mass PDF contains a $B^0 \to K^+\pi^-$ component (green), $\overline{B}^0_s \to K^+\pi^-$ component (red) and a combinatorial background component (dark blue dot-dashed).

1.7 Acceptance function
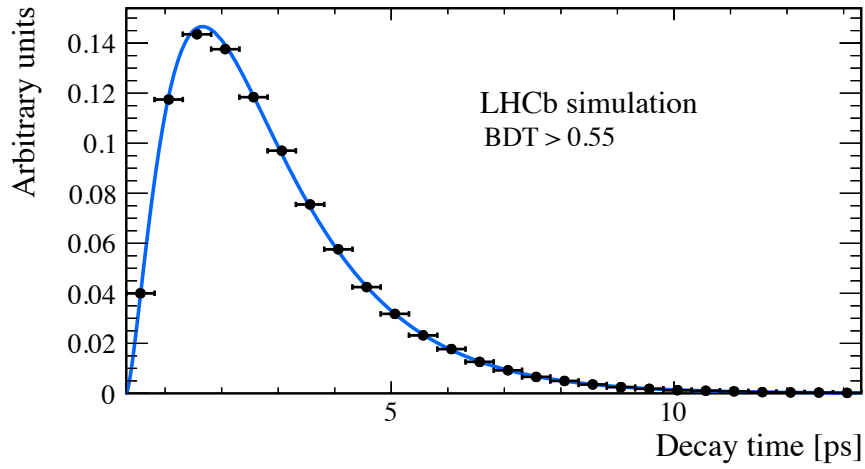

Figure 24: Unbinned maximum likelihood fit to the decay time distribution of simulated $B_s^0 \to \mu^+ \mu^-$ signal events. The fit is used to determine the acceptance function parameters.
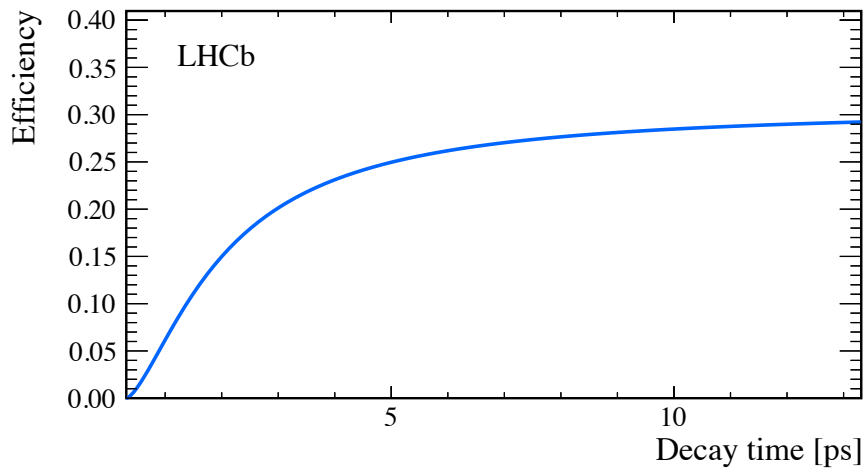

Figure 25: Acceptance function determined from the unbinned maximum likelihood fit to the decay time distribution of simulated $B_s^0 \to \mu^+ \mu^-$ signal events.