11 Supplementary material

This appendix contains supplementary material that appears only in the public CDS record and not in the journal paper.

11.1 Measurement of the branching fractions

Figure 18 shows the dimuon mass versus BDT distribution of the selected $B^0_{(s)} \to \mu^+\mu^-$ candidates in the Run 1 and Run 2 datasets. Figure 19 shows the mass distributions of the main background contributions for BDT > 0.5. Figures 20 and 21 show the mass distributions of the selected $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ candidates from which the corresponding normalisation yields are extracted. Figure 22 shows the mass distribution of the selected $B^0_{(s)} \to \mu^+\mu^-$ candidates with BDT > 0.5. The result of the fit is overlaid and the different components are shown. Figure 23 compares the background-subracted distribution of the B^0_s transverse momentum and pseudorapidity for $B^0_s \to \mu^+\mu^-$ candidates with BDT > 0.5 in data and simulation. Figure 24 reports the ratio of the efficiency-corrected yields of the two normalisation channels as measured in Run 1 and Run 2, and the PDG average. Figure 25 shows the mass distributions of the selected $B^+ \to J/\psi K^+$ candidates from which the corresponding normalisation yields are extracted, split by Run 1 and Run 2 datasets and with linear y-axis scale, as opposed to the logarithmic scale reported in Fig. 21.
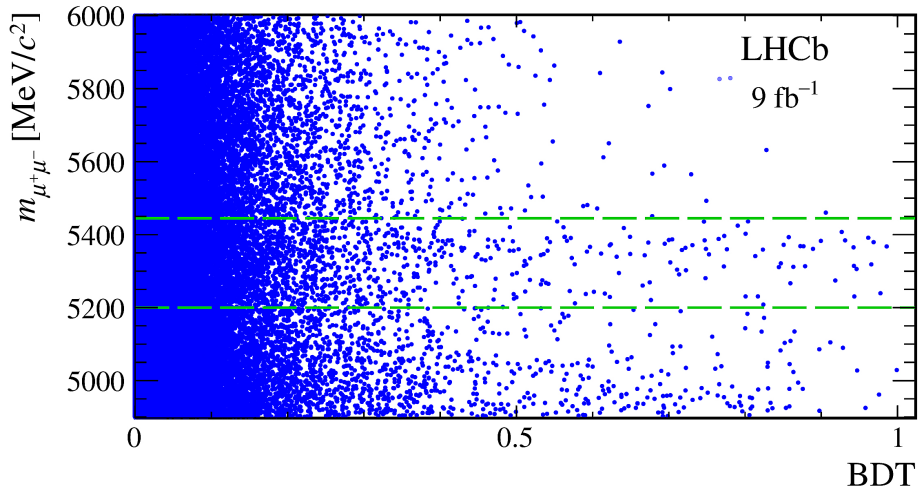

Figure 18: A scatter plot showing the dimuon mass versus BDT distribution of signal candidates, for the full data set used in Run 1 and Run 2. The dashed green line indicates the signal mass region, for which $m_{\mu^+\mu^-} \in [5200, 5445]\,\mathrm{MeV}/c^2$.
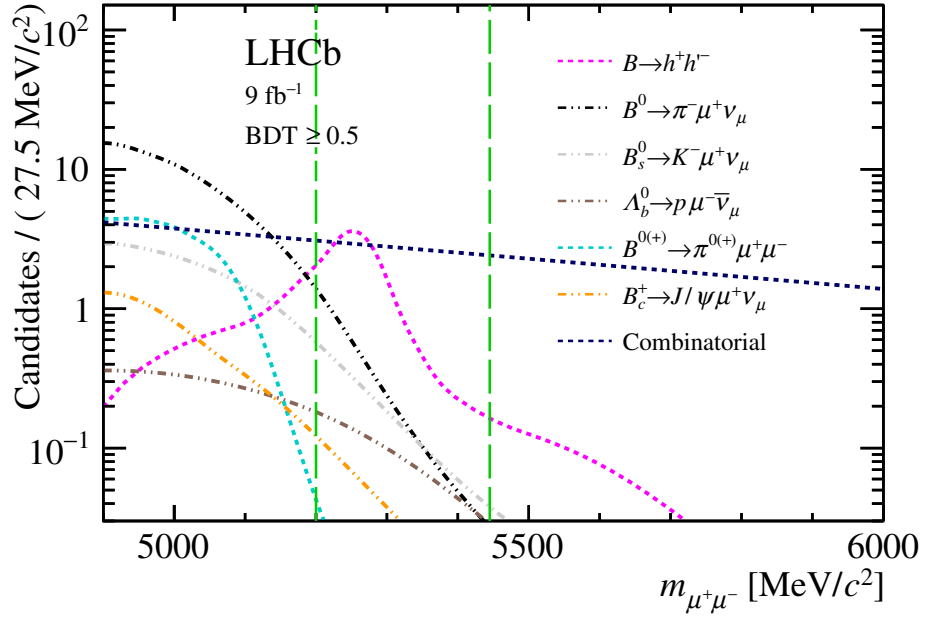
Figure 19: Mass distribution of the main background sources: combinatorial background (blue short dashed), $B^0_{(s)} \to h^+h'^-$ (magenta dotted), $B^0 \to \pi^-\mu^+\nu_\mu$ (black dot-dashed), $B^0_s \to K^-\mu^+\nu_\mu$ (grey dot-dashed), $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ (cyan dot-dashed), $B^+_c \to J/\psi\mu^+\nu_\mu$ (orange dot-dashed) and $\Lambda^0_b \to p\mu^-\overline{\nu}_\mu$ (brown dot-dashed). The dashed green line indicates the signal mass region, for which $m_{\mu^+\mu^-} \in [5200, 5445]\,\mathrm{MeV}/c^2$.
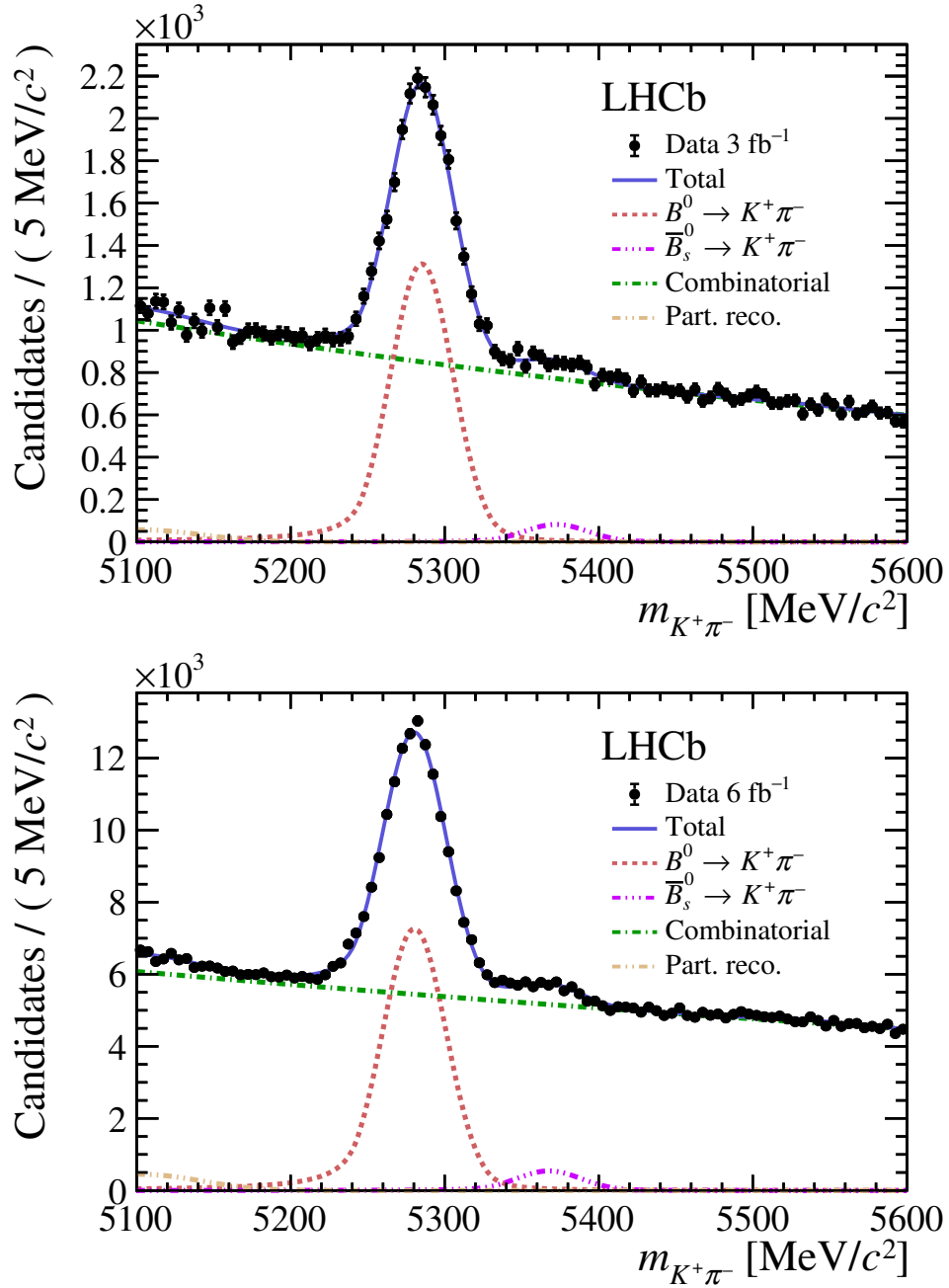
Figure 20: Mass distribution of the selected $B^0 \to K^+\pi^-$ candidates (black dots) in (top) Run 1 and (bottom) Run 2 data. The result of the fit to determine the normalisation yield is overlaid (blue solid line) and the different components are detailed: $B^0 \to K^+\pi^-$ (red solid line), $B_s^0 \to K^-\pi^+$ (green solid line) and combinatorial background (blue dashed line).
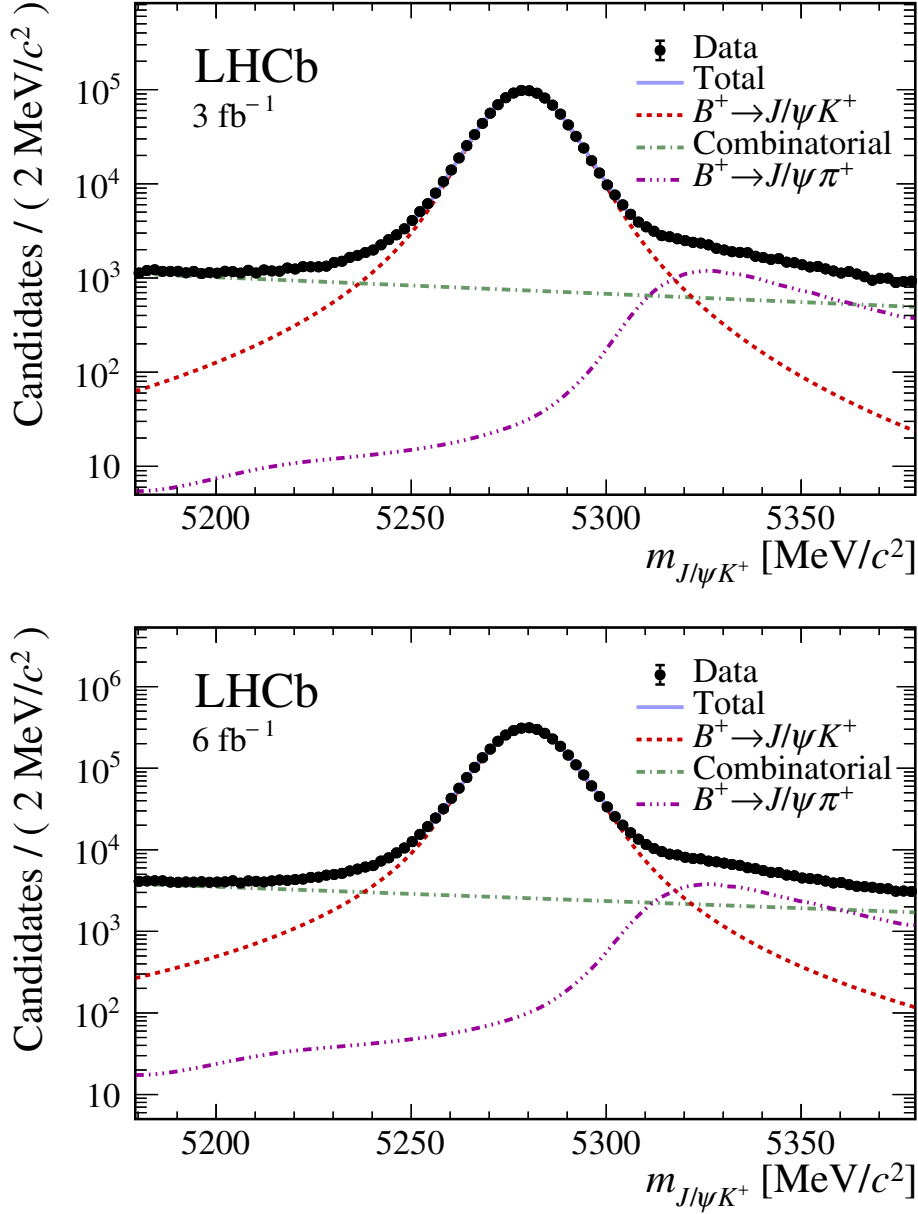
Figure 21: Mass distribution of the selected $B^+ \to J/\psi K^+$ candidates in (top) Run 1 and (bottom) Run 2 data. The result of the fit to determine the normalisation yield is overlaid and the different components are detailed: $B^+ \to J/\psi K^+$ (red dotted line), $B^+ \to J/\psi \pi^+$ (purple dashed line), and combinatorial background (green dashed-dotted line).

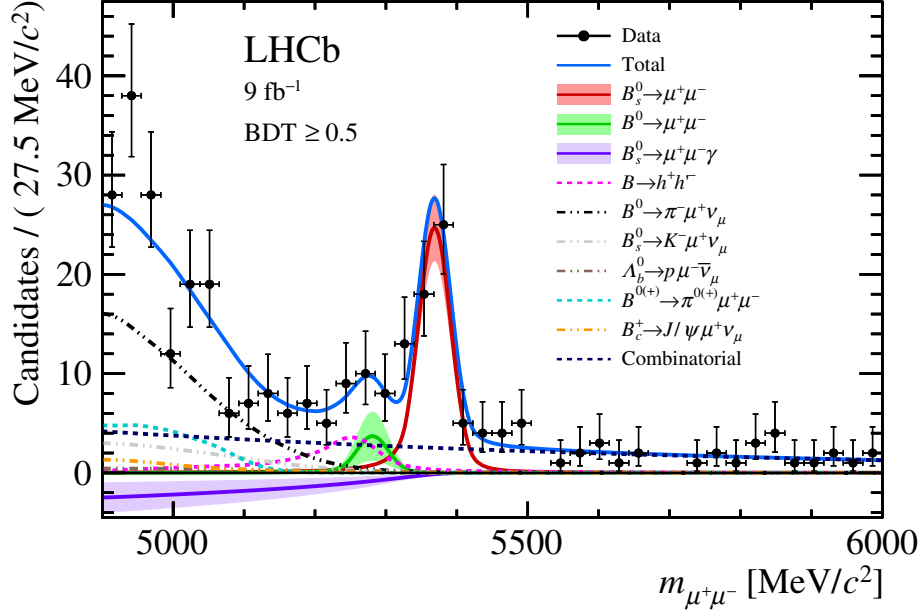Figure 22: Mass distribution of the selected $B^0_{(s)} \to \mu^+\mu^-$ candidates (black dots) with BDT > 0.5. The result of the fit is overlaid (blue solid line) and the different components detailed: $B^0_s \to \mu^+\mu^-$ (red solid line), $B^0 \to \mu^+\mu^-$ (green solid line), $B^0_s \to \mu^+\mu^-\gamma$ (violet solid line), combinatorial background (blue dashed line), $B^0_{(s)} \to h^+h'^-$ (magenta dashed line), $B^0 \to \pi^-\mu^+\nu_\mu$ (black dot-dashed line), $B^0_s \to K^-\mu^+\nu_\mu$ (grey dot-dashed line), $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ (cyan dashed line), $B^+_c \to J/\psi\mu^+\nu_\mu$ (orange dot-dashed line) and $\Lambda^0_b \to p\mu^-\overline{\nu}_\mu$ (brown dot-dashed line).
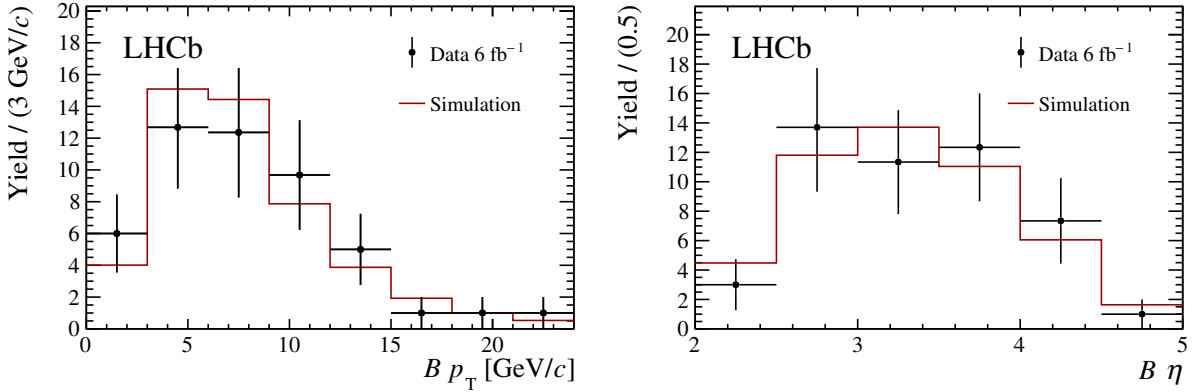

Figure 23: Distribution of (left) B^0_s transverse momentum and (right) B^0_s pseudorapidity, for $B^0_s \to \mu^+\mu^-$ candidates in data, compared to simulation. Only $B^0_s \to \mu^+\mu^-$ candidates with BDT > 0.5 are used. The distribution is background subtracted considering signal in $\pm 3\sigma$ from the B^0_s mass, where σ is the mass resolution, and background from the upper mass sideband. The simulation is normalised to the same integral as the data. The figures include Run 2 data only, as for Run 1 the statistics is small.
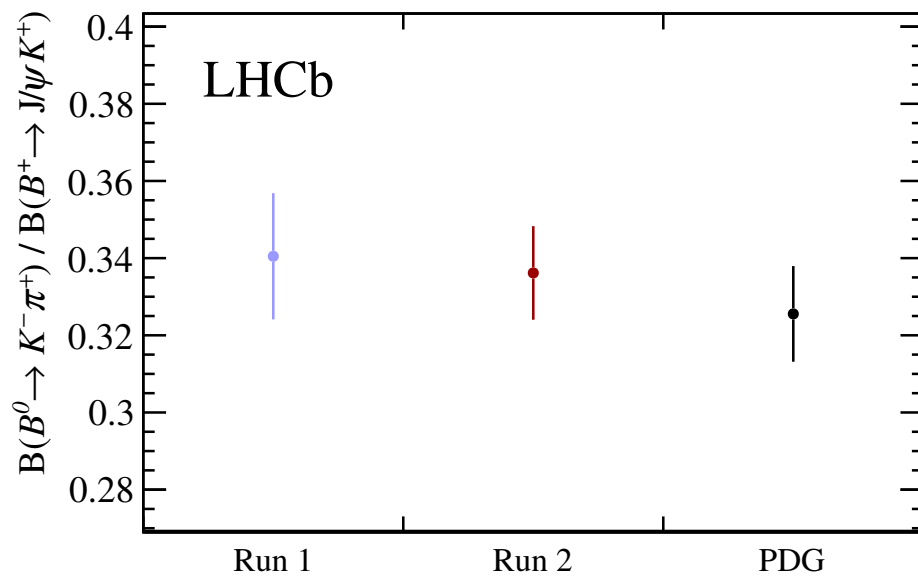
Figure 24: Ratio of the efficiency-corrected yields of the two normalisation channels for (violet) Run 1, (red) Run 2 and (black) the ratio from the PDG values.
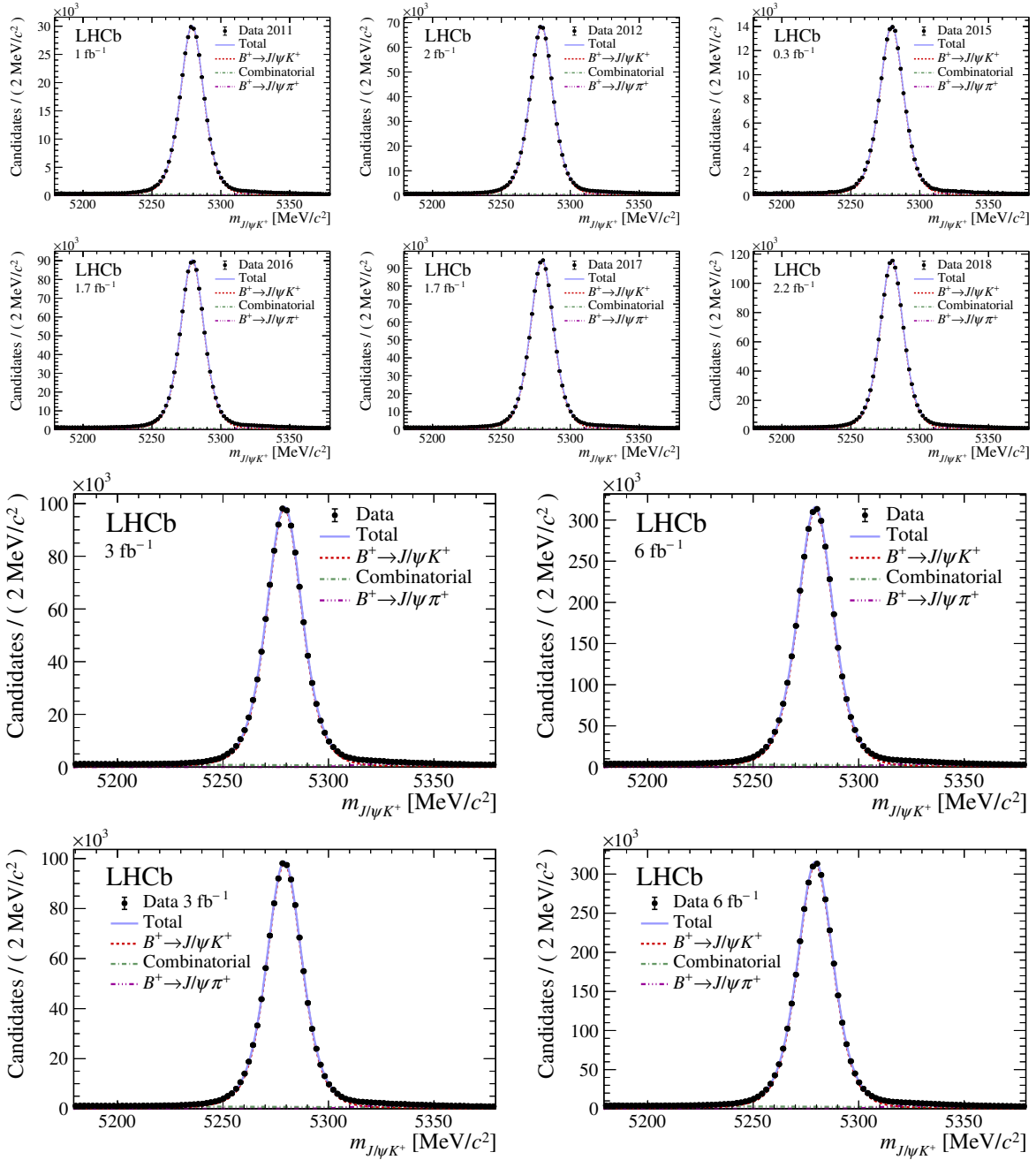
Figure 25: Mass distribution of $B^+ \to J/\psi K^+$ candidates in data for different data-taking years. Superimposed is a fit to the distribution: the blue line shows the total fit, the red dashed line is the $B^+ \to J/\psi K^+$ component, the green dash-dotted line is the combinatorial background, the purple dash-three-dotted line is the $B^+ \to J/\psi \pi^+$ misidentified background. These are the linear plots of the same as in the main text. For the linear plots there are the two different styles versions.

11.2 Measurement of the effective lifetime

Figure 26 shows the fit to the decay time distributions of simulated $B_s^0 \to \mu^+\mu^-$ events for the two BDT regions used in the measurement. Figure 27 shows the profile likelihood of the $B_s^0 \to \mu^+\mu^-$ effective lifetime measurement. Figure 28 displays the two-dimensional representation of the $B_s^0 \to \mu^+\mu^-$ branching fraction vs $A_{\Delta\Gamma_s}^{\mu\mu}$ in the physically allowed range.
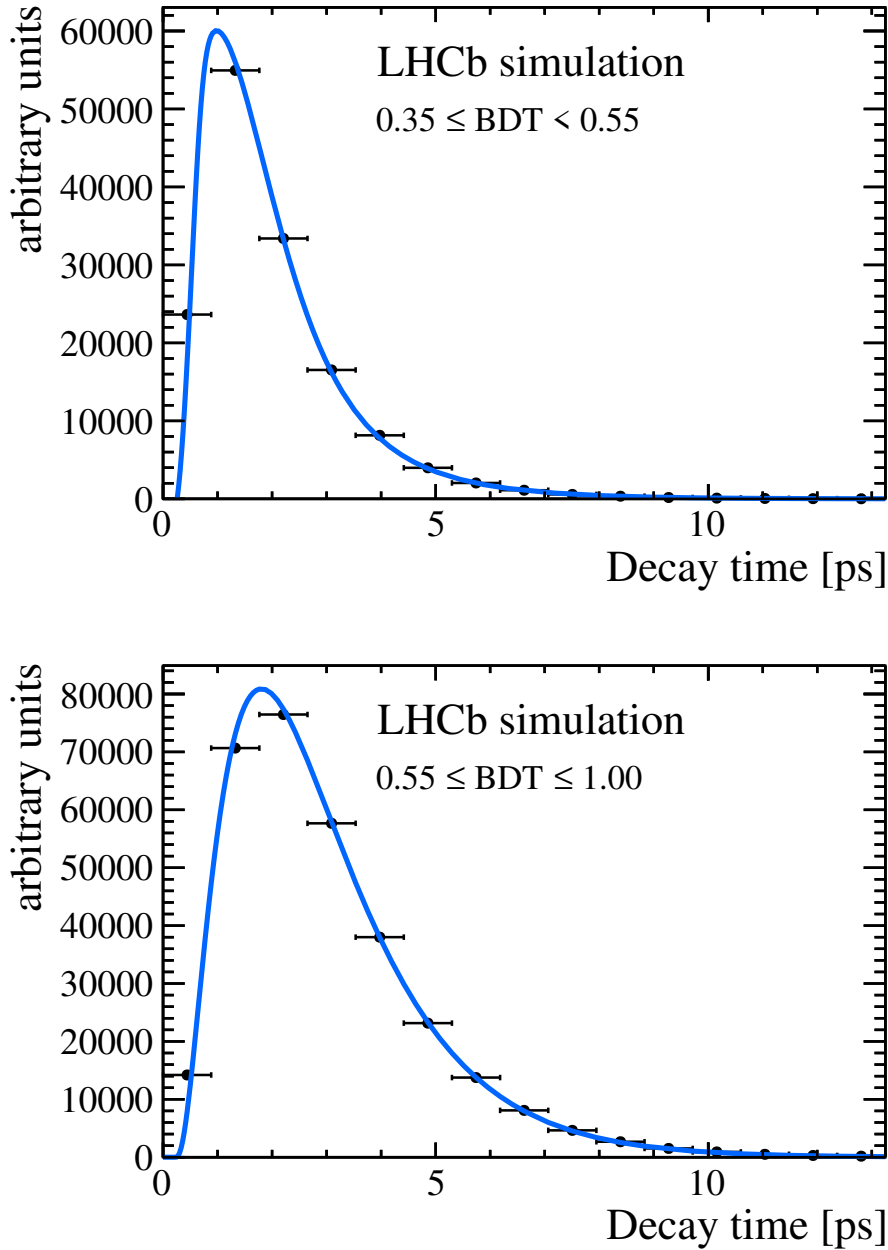

Figure 26: Unbinned maximum-likelihood fits to the decay time distribution of simulated $B_s^0 \to \mu^+\mu^-$ signal events in two BDT regions. The fit is used to determine the acceptance function parameters.
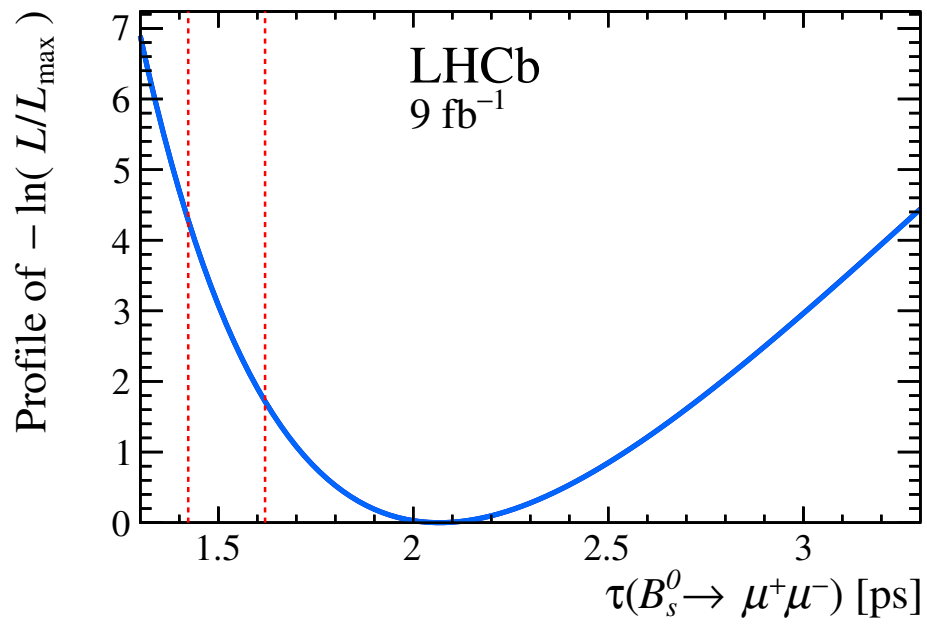
Figure 27: The profile likelihood for the effective lifetime measurement. The red dotted lines indicate the physically allowed range for the lifetime, with the left edge and right edges corresponding to the lifetimes of the light and heavy mass eigenstates, respectively.
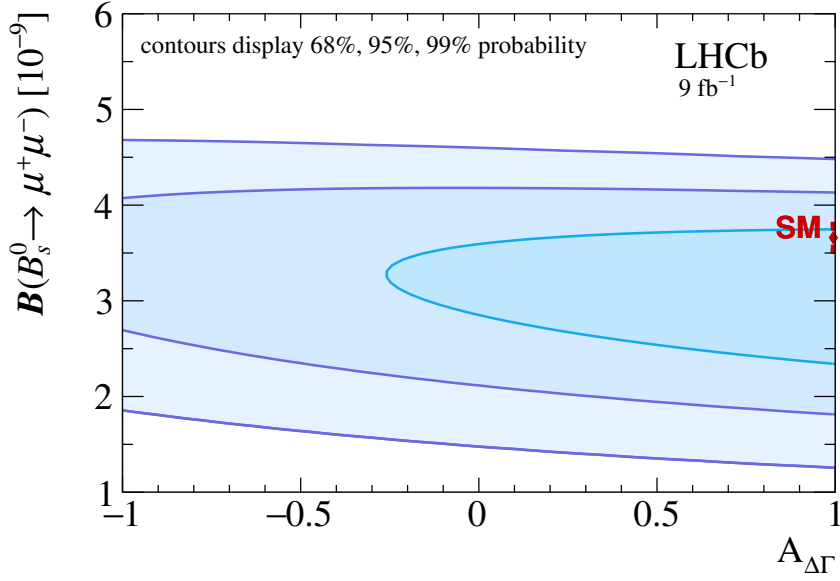
Figure 28: A two-dimensional representation of the combined measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime in the physically allowed range. The Standard Model value [7] is shown in red. The profile likelihood contours corresponding to 68%, 95%, 99% probability within this range are shown as blue contours. The profile likelihood is determined by approximating the likelihoods of each measurement with a variable-width Gaussian [103] (to take into account asymmetrical uncertainties) and by estimating the dependence of the branching fraction on $A^{\mu\mu}_{\Delta\Gamma_s}$ by fitting a linear function to the branching fraction results for the three $A^{\mu\mu}_{\Delta\Gamma_s}$ hypotheses, as described in Sec. 8. It is found that the branching fraction fit does not favour any $A^{\mu\mu}_{\Delta\Gamma_s}$ hypotheses and therefore does not provide additional sensitivity to the effective lifetime.