Statistical methods used on searches at LHCb, with special emphasis in the search for the very rare decay $B_s \to \mu^+\mu^-$

Jose A. Hernando, on behalf of the LHCb collaboration
Universidade de Santiago de Compostela, Spain

Abstract

The LHCb experiment searches for new physics in CP violation and rare decay processes of B and D mesons. We describe here the strategy followed to search for the rare decay $B_s^0 \to \mu^+\mu^-$. This is one of the key analyses of LHCb and serves as a model for other LHCb searches. Emphasis will be put on the statistical methods used.

1 Introduction

The LHCb experiment is one of the four experiments located at the Large Hadron Collider (LHC) at CERN. The experiment is designed to search for new physics (NP) beyond the Standard Model (SM) in charge-parity (CP) violation and rare decay processes of beauty (B) and charm (D) mesons. During 2010, the experiment collected 37 pb^{-1} of integrated luminosity. One of the first measurements published by the LHCb collaboration was the value of the $b\bar{b}$ cross-section $\sigma(pp \to b\bar{b}X) = (284 \pm 20 \pm 49)\,\mu\mathrm{b}$ [1]. This cross-section is high enough to produce thousands of B mesons per second at the nominal luminosity $\mathcal{L} = 2 - 5 \times 10^{32}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. With these large statistical samples, the physics reach of the LHCb experiment does not suffer significantly from the fact that the LHC runs at $\sqrt{s} = 7$ TeV.

The LHCb detector has performed beautifully during its first year of operation. The detector is a forward spectrometer with a vertex detector, a tracking system (before and after a warm dipole magnet), two RICH detectors, an electromagnetic and hadronic calorimeter and a muon system. A detailed description of the detector can be found in Ref. [2]. In order to identify specific B decays and separate them from the large background, the experiment has a flexible trigger system, good particle identification and excellent momentum and vertex resolution. Except for the vertex resolution, where the impact parameter of the tracks has been measured with an uncertainty $\sim 10\%$ greater than expected, the rest of the detector characteristics are within design specifications.

The LHCb physics program includes the search for very rare decays $B_{(s)}^0 \to \mu^+\mu^-$, $D \to \mu^+\mu^-$, lepton flavor violating decays $B \to \mu e$, etc. The $B_s^0 \to \mu^+\mu^-$ is one of the key searches of LHCb. The first results [3] obtained with $37\,\mathrm{pb}^{-1}$ integrated luminosity has been recently submitted to Phys. Lett. B. The $B_s^0 \to \mu^+\mu^-$ analysis has defined a strategy that is followed by other LHCb searches. For this reason, and given its mature state, this Paper is dedicated to the description of this analysis. Special emphasis is given to the statistical methods used. However it is a "classical" search as it uses well-known methods. The only exception is the use of a multi-variate method ($\Delta\chi^2$) that will be described here in detail.

2 The $B_s \to \mu^-\mu^+$ search

Within the SM, exclusive dimuon decays of B^0 and B_s^0 mesons occur only via loop diagrams and are helicity suppressed. The SM prediction is $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.2 \pm 0.2) \times 10^{-9}$ [4]. However, within NP models, especially those with an extended Higgs sector, the $\mathcal{B}$ can differ significantly. This is the case, for example, within the minimal supersymmetric SM (MSSM) [5]. The current limits have been set by the CDF and D0 collaborations [6]. The CDF collaboration has presented a preliminary result [7]

with the most stringent limit so far $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 4.3 \times 10^{-8}$ at 95 % C.L. with $3.7\,\mathrm{fb}^{-1}$ of data analyzed, but this is still an order of magnitude greater than the SM value.

The LHCb experiment is well suited for the search of this decay due to its excellent invariant mass resolution, vertex resolution, muon identification and trigger acceptance. The forward geometry of LHCb allow us to trigger on muons with low transverse momenta. For example, the first level trigger L0, which is a hardware trigger, accepts an event if there is a single muon with $p_T > 1.4\ \mathrm{GeV/c^2}$ or if there are two muons with $p_T > 0.48\,(0.56)\ \mathrm{GeV/c^2}$ for the muon with the lowest (highest) p_T. The signal trigger efficiency has been estimated with data to be $(90 \pm 4)\,\%$. $B_s^0 \to \mu^+\mu^-$ events are selected offline with soft criteria designed to remove the most obvious background events while keeping the signal efficiency as high as possible. The selection is based on the existence of two tracks identified as muons, that form a good secondary vertex separated from the primary one by a distance significance (distance/uncertainty) greater than 15. After the trigger and the selection, we expect 18 ± 2 signal events per fb^{-1} according to the SM prediction.

The search for the signal is done in a two-dimensional space. One coordinate of the space is the invariant mass. The second coordinate, which we refer to it as a Geometrical Likelihood (GL), is the output of a multivariate method that combines different discriminant variables taking into account their correlations. The invariant mass resolution of the signal has been measured in control channels and is $26.7 \pm 0.9\,\mathrm{MeV/c^2}$. We expect the overall mass distribution to be a Gaussian distribution for the signal plus an exponential distribution for the background. The background is dominated by random combinations of real muons coming from semileptonic decays of a $b\bar{b}$ pair ($b\bar{b} \to \mu^+\mu^- X$). The GL combines the following variables: the minimum impact parameter of the muons, the distance of closest approach between them, the impact parameter significance of the B candidate, the B proper time and an isolation variable that quantifies if any of the muons is attached to other secondary vertices besides the one from the B. The GL variable has a range between 0 and 1. For the background it peaks at 0 while there are almost no events left for GL> 0.5. The GL has been constructed in such a way that the signal events distribute uniformely between 0 and 1. The GL is defined using a sample of simulated $B_s \to \mu^+\mu^-$ and $b\bar{b} \to \mu^+\mu^- X$ events, but as we will comment later, its distribution has been validated with data. The region $\mathrm{GL} > 0.5$ and with an invariant mass within $60\,\mathrm{MeV/c^2}$ interval around the $B_{(s)}^0$ mass was blinded until the analysis was completely defined.

To construct the GL we use a multivariate method called $\Delta\chi^2$. This method is described in Ref. [8] and there is a first version in Ref. [9]. The method transforms a set of n initial variables $\{x_i\}$ into a set of n new variables $\{s_i\}$ which are distributed according to a Gaussian with zero mean and sigma unity. The transformed variables are mostly uncorrelated and the p.d.f. can be approximated by an n-dimensional Gaussian. Two transformations are defined to separate signal from background: one for the signal events and a second one for background. Given an event, the original $\{x_i\}$ variables are transformed into the $\{s_i\}$ variables using the transformation of the signal, and into the $\{b_i\}$ variables using the transformation of the background. Then we compute the quantities $\chi_s^2 = \sum_{i=1}^n s_i^2$ and $\chi_b^2 = \sum_{i=1}^n b_i^2$ that are related to the probability that the event is signal or background, and we use the difference between them as the final discriminanting variable $\Delta\chi^2 = \chi_s^2 - \chi_b^2$. For practical reasons, we transform the $\Delta\chi^2$ distribution for the signal events into a uniform distribution between 0 and 1.

The process of "Gaussianization" and de-correlation of the input variables is made in two steps. In the first step the initial variables are transformed into gaussian distributed variables. To do so, first the variables are transformed into a uniform distribution using the accumulative function of the distribution of the input variable; and later, they are transformed into a gaussian distribution using the inverse function of the accumulative function of a gaussian distribution. At this stage, the variables are gaussian but they are still correlated. To reduce the correlation, we compute the moments or the symmetry axis of the new variables. Then, we rotate them to the symmetry axis and we re-gaussianize the variables with the process described above. After this point, the variables are now gaussian distributed with unit variance and zero mean and they are mostly uncorrelated. They follow an n-dimensional Gaussian. We can relate

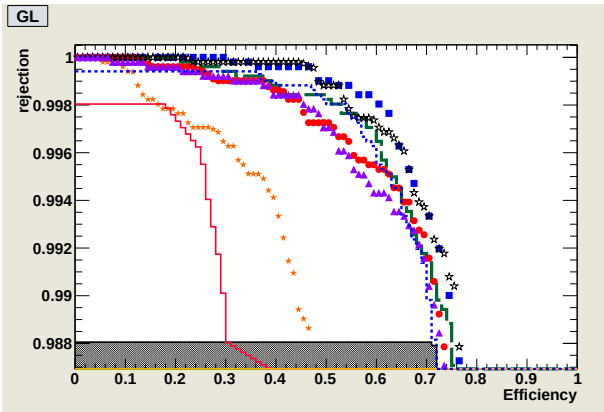

Fig. 1: Performance (rejection of background events *vs* efficiency on signal events) of the $\Delta\chi^2$ method (blue squares) compared with the default configuration of different TMVA methods: BDT (open stars), PDERS (sort dashed), Fisher Discriminant (violet trianges), Best performing NN (red circles), Support Vector Machine (green dashed line), RuleFit (red solid line), FDA (orange stars), kNN (black filled histogram).

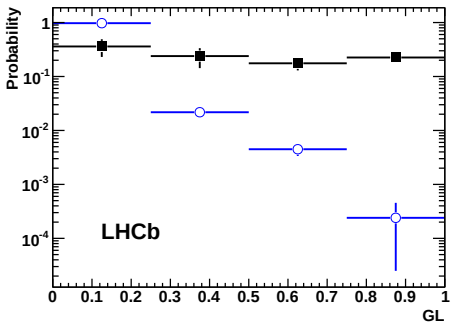

Fig. 2: GL p.d.f for signal (black) and background (blue) calibrated using data.

the probability of an event to belong to that sample (i.e signal) with the $\chi^2_s = \sum_{i=1}^{n} s_i^2$ quantity.

We have compared the performance of this method with some multi-variate methods implemented in the `ROOT TMVA` package [11]. Using as input variables the ones indicated above, the $\Delta\chi^2$ method performs as well as the Boosted Decision Trees (BDT), which of the MVA methods used performs the best. The default configurations of the methods provided by the `TMVA` package were used. The comparisons between all the methods were done under the same conditions. The sample was divided in two identical samples, one used for training the method and the second one to obtain the performance. Figure 1 shows the signal efficiency vs background rejection obtained with the different methods on the signal and background samples. For the background, we used $b\bar{b} \to \mu^+\mu^- X$ simulated events.

One of the strong points of the analysis is the fact that the mass and the GL pdfs have been calibrated with data. The resolution of the B_s mass is obtained from the interpolation between the J/Φ and Υ dimuon resonances. It has also been measured from the mass distribution of $B_s^0 \to K^+K^-$ events. The GL pdf of the signal has been calibrated using $B \to h^+h^{-'}$ events (where h, h' stand for kaon or pion). The selection of these events is identical to the signal ones (except for the muon identification requirement) but they are triggered differently. The effect of the trigger has been corrected using only $B \to h^+h^{-'}$ candidates where the events were triggered not using the candidates themselves. Therefore, they were not biased by the trigger. The GL pdf of the background has been calibrated using the $B_s \to \mu^+\mu^-$ events that are in the sidebands of the invariant mass (they are outside a $60\,\mathrm{MeV/c^2}$ window centered at the B_s mass but inside a larger window of $600\,\mathrm{MeV/c^2}$). The GL calibrated pdfs for

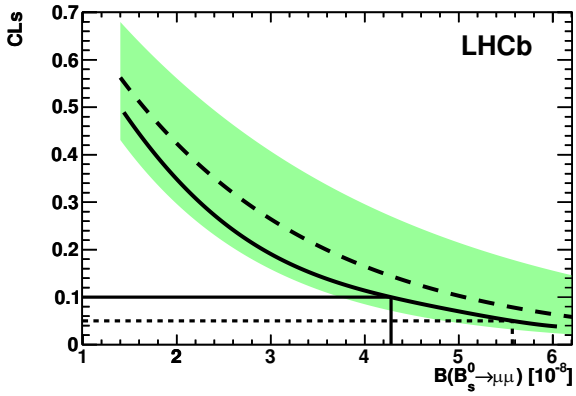

Fig. 3: CL_s vs $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ expected value if there is no signal events (dashed curve) and observed value (solid curve). The green shaded area contains the $\pm\sigma$ interval of possible results compatible with the expected value when only background is observed. The 90% (95 %) C.L. observed value is shown by the solid (dashed) line.

signal and data are shown in Fig. 2.

The region $\mathrm{GL} > 0.5$ and $60\,\mathrm{MeV/c^2}$ window around the B_s^0 mass has been divided in bins (2 for the GL, and 6 for the mass, equally spaced). In each bin the expected number of background and signal events are computed using the mass and GL calibrated pdfs. We used several control channels $B^+ \to J/\Psi K^+,\ B^0 \to K^+\pi^-,\ B_s \to J/\Psi\phi$ to normalize the signal expected events to a known $\mathcal{B}$ via the relation,

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \mathcal{B}_n \frac{\epsilon_n^{sel}}{\epsilon_s^{sel}} \frac{\epsilon_n^{trg/sel}}{\epsilon_s^{trg/sel}} \frac{f_n}{f_s} \frac{N_s}{N_n}, \tag{1}$$

where $\mathcal{B}_n$ is the branching ratio of any of the normalization channels noted above. The $\frac{\epsilon_n^{sel}}{\epsilon_s^{sel}}$ is the ratio of the selection efficiency (which includes also the acceptance of the detector and the reconstruction efficiency) between the normalization and the signal channel. The $\frac{\epsilon_n^{trig/sel}}{\epsilon_s^{trg/sel}}$ is the ratio of the trigger efficiency on the selected events between the normalization and the signal channels. f_n/f_s is the ratio of the fragmentation fractions, i.e. the ratio of the probabilities that a b quark produces a $B_s(B_n)$ meson. The ratio is 1 when normalizing to B_s, and $f_d/f_s = 3.71 \pm 0.47$ [10] when normalizing to a B^0. Finally N_s and N_n are the number of selected and triggered events for the signal and the control channel, respectively. The first ratio has been computed using MC simulations. The second one has been estimated using the data. Several cross-checks have been performed with data to verify the first ratio. Care has been taken in defining the selection of the normalization of the control channels so that it matches the signal selection as closely as possible in order to minimize the systematic errors.

To set a limit on the $\mathcal{B}$ we have used the CL_s method [12]. The CL_s method is well known in HEP; in particular, it was used in the Higgs searches performed at LEP. It uses as a test-statistic the ratio of the likelihood of the signal plus background hypothesis and the likelihood of the background-only hypothesis. The distribution of the test-statistic of the signal plus background (sb) and background-only (b) hypotheses are used to compute two p-values ($p_{sb} = \mathrm{CL}_{sb},\ p_b = 1 - \mathrm{CL}_b$). The CL_s quantity is computed as the ratio $\mathrm{CL}_s = \frac{\mathrm{CL}_{sb}}{\mathrm{CL}_b}$. This quantity has the advantage (over the pure p_{sb}-value) of not excluding a region where the experiment has no sensitivity to observe a signal. Figure 3 shows the CL_s vs $\mathcal{B}$ when the observation equals the expected number of background events (dashed curve). The shaded area contains the $\pm\sigma$ interval of possible results compatible with the expected value when only background events are observed. The solid curve corresponds to the LHCb observation with $37\,\mathrm{pb}^{-1}$. The horizontal solid (dashed) line corresponds to the 90% (95%) C.L. The LHCb limits are $\mathcal{B}(B_s \to \mu^+\mu^-) < 5.6 \times 10^{-8}$ at 95 % C.L. and $\mathcal{B}(B^0 \to \mu^+\mu^-) < 1.5 \times 10^{-8}$ at 95 % C.L. The systematics errors of the normalization

factor and of the pdfs of the mass and the GL are propagated into the calculation of the $\mathcal{B}$ C.L. using the technique described in Ref. [13]. Currently, the collaboration has started to discuss the possibility of using different methods other than the CL$_s$ to obtain the $\mathcal{B}$ limit.

Previous MC studies [14] have shown the potential of LHCb to observe the SM $B_s^0 \to \mu^+\mu^-$ decay. For the observation we use the p-value of the background-only hypothesis or 1-CL$_b$. Measuring the SM $\mathcal{B}(B_s \to \mu^+\mu^-)$ at 3 σ it will require collecting more than $2\,\mathrm{fb}^{-1}$ of data.

3 Conclusions

The LHCb detector has performed beautifully during the data taking period of the year 2010 and has collected a data-set corresponding to 37 pb^{-1} of integrated luminosity. The LHCb physics program includes the searches for rare or forbidden B and D meson decays. One of the most relevant LHCb analyses is the measurement of the $\mathcal{B}(B_{(s)}^0 \to \mu^+\mu^-)$. The LHCb collaboration has sent for publication in Physics Letter B the first results of this search.

The $B_s \to \mu^+\mu^-$ analysis serves as a model for other LHCb searches. It is based on the definition of a sensitive region, in this case it is a plane defined by the invariant mass and a second variable, the GL, that combines several discriminant variables into one using a multi-variate method (the $\Delta\chi^2$ method). The sensitive region of the plane was blinded until the analysis was completely defined. The main point of the analysis is the use of control channels to calibrate the signal and background pdfs and to normalize the number of observed events to a known $\mathcal{B}$ ratio using several normalization channels. To set a limit on the $\mathcal{B}$ the CL$_s$ method has been used. Discussions are ongoing to use other methods. The limits set by the LHCb collaboration with $37\,\mathrm{pb}^{-1}$ are $\mathcal{B}(B_s \to \mu^+\mu^-) < 5.6 \times 10^{-8}$ at 95 % C.L. and $\mathcal{B}(B^0 \to \mu^+\mu^-) < 1.5 \times 10^{-8}$ at 95 % C.L.

References

[1] R. Aaij *et al.*, [LHCb Collaboration], "Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7\,\mathrm{TeV}$ in the forward region", Phys. Lett. B **694** 209 (2010).

[2] A.A. Alves *et al.* [LHCb Collaboration] "The LHCb detector at LHC", JINST **3** (2008) S08005, and references therein.

[3] R. Aaik *et al.* [The LHCb Collaboration] "Search for the rate decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$", arXiv:1103.2465v1 [hep-ex], CERN-PH-EP-2011-029.

[4] A.J. Buras, G. Isidori and P. Paradisi, "EDMs vs CPV in $B_{s,d}$ mixing in two Higgs doublet models with MFV", arXiv:1007.5291; A. J. Buras, "Relations between $\Delta m_{s,d}$ and $B_{s,d} \to \mu^+\mu^-$ in Models with Minimal Flavour Violation", Phys. Lett. B **566**, 115 (2003)

[5] K.S. Baba, C.F. Kolda, "Higgs-mediated $B_q^0 \to \mu^+\mu^-$", Phys. Rev. Lett. **84**, 539 (2010).

[6] V. Abazov *et al.* [D0 Collaboration], "Search for rare decay $B_s^0 \to \mu^+\mu^-$, Phys. Lett. B **693**, 539 (2010). T. Aaltonen *et al.* [CDF Collaboration], "Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with $2\,\mathrm{fb}^{-1}$ pf $p\bar{p}$ Collisions", Phys. Rev. Lett. **100**, 101802 (2008).

[7] T. Aaltonen *et al.* [CDF Collaboration], "Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with $3.7\,\mathrm{fb}^{-1}$ pf $p\bar{p}$ collisions with CDF II", CDF Public Note 9892.

[8] D. Martinez Santos, J.A. Hernando and F. Teubert, "LHCb Potential to Measure/Exclude the Branching Ratio of the Decay $B_s^0 \to \mu^+\mu^-$", LHCb Note LHCb-2007-033.

[9] D. Karlen, "Using Projections and Correlations to Approximate Probability Distributions", Comp. Phys., **12** 69 (1998).

[10] D. Asner *et al.* [Heavy Flavour Averaging Group], "Averages of b-hadron, c-hadron, and τ-lepton properties", arXiv:1010.1589 [hep-ex] (online update at: http://www.slac.stanfrod.edu/xorg/hfag/osc/).

[11] P. Speckmayer *et al.* "The toolkit for Multivariate Data Analysis TMVA 4", J. Phys. Conf. Ser. **219** 033057 (2010).

[12] A.L. Read, "Presentation of search results: the CL_s technique", J. Phys. G **28**, 2693 (2002).

[13] T. Junk, "Confidence Level computation for combining searches with small statistics", Nucl. Instrum. Meth A **434** 435 (1999).

[14] D. Martinez Santos, 'Study of the very rare decay $B_s^0 \to \mu^+\mu^-$ in LHCb", CERN-THESIS-2010-068.