Selection of $B_s^0 \to \psi(2S) K_s^0$ events

Judith Gnade
judith.gnade@tu-dortmund.de

Hendrik Lauersdorf
hendrik.lauersdorf@tu-dortmund.de

John Wendel
john.wendel@tu-dortmund.de

July 25, 2022

TU Dortmund – Fakultät Physik

Contents

1 The LHCb detector

The LHCb is one of four large experiments of the LHC at CERN. Its detector, a forward spectrometer, is located at one of the four interaction points. There, mostly hadrons with a charm or bottom quark are studied. The structure of the detector is shown in figure 1 and its different components are briefly described in the following.

graphics/lhcb.png

Figure 1: Components of the LHCb detector. [1]

The Vertex Locator (VELO) tracks the trace of the particles after a collision. This way it is possible for the VELO to recontruct the position of the proton-proton collision: the primary vertex (PV).

The components TT, T1, T2 and T3 are four-layer silicon tracking detectors. They reconstruct the tracks of the particles after the collision. Furthermore T1-T3 contain tube modules filled with gas (Ar and CO_2). The crossing particles ionize the gas, which results in a signal.

Two magnets are used to measure the momenta of the particles. Due to the magnetic field the trace of the particles is curved and from the curvature the momenta of the particles can be determined. Because of existing systematic uncertainties up to 1%, the polarity of the magnetic field is reversed from time to time.

For hadrons, the ring-imaging Cherenkov detectors (RICH1 and RICH2) differentiate between pions, kaons and protons. They contain gas or aerogel and due to the Cherenkov effect, the velocity of the crossing particles can be determined. This is possible since the angle of the emitted Cherenkov photons is related to the velocity of the particles. Using the information about the particles' momenta and their velocities, the mass of the different particles (and therefore their flavour) can be computed.

A calorimeter system consisting of a scintillating pad (SPD), a preshower (PS) detector,

an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL) identifies the photons, electrons and hadrons. The particles deposit their energy in the different layers of the calorimeters. The photons are mainly absorbed in the ECAL, while hadrons are absorbed in the HCAL. Only muons and neutrinos pass the calorimeters whithout being absorbed. Therefore muon chambers (M1-M5) are added to the detector to register muons.
Since there are plenty events happening at a high rate, the LHCb uses a trigger system to differentiate between potentially interesting events and other events. This trigger system consists of a hardware component and a software component. First, the hardware component uses information gathered from the calorimeters and from the muon chambers to decide whether an event is of interest. After that, the software component reconstructs the chosen events in real time and makes a decision whether the events are interesting for further analysis. If they are not considered to be useful, they are neglected.

2 Theoretical foundations

2.1 The B_s^0 Decay

In this analysis signal candidates for $B_s^0 \rightarrow \psi(2S)K_S^0$ events, as shown in Figure 2, are extracted from a heavily polluted LHCb data sample. The final-state mesons of the B_s^0 decay (the $\psi(2S)$ and the K_S^0) are not detected directly. Instead, the $\psi(2S)$ is reconstructed from two oppositely charged muons and the K_S^0 is reconstructed from two oppositely charged pions.

graphics/decay.png

Figure 2: Leading order Feynman diagram of the decay $B_s^0 \rightarrow \psi(2S)K_S^0$.

A very similar decay is the $B^0 \rightarrow \psi(2S)K_S^0$ decay. Since it has the same final state as the B_s^0 decay and since it is also kinematically similar, it passes the same selection requirements as the B_s^0 decay. Not only these decays, but also a combinatorial background can be found in the data sample. This combinatorial background stems from events coincidently passing the selection. For example if by chance two unrelated muons have a point of intersection, the algorithm will reconstruct them as a $\psi(2S)$, which then can be reconstructed to a false B meson if combined with a K_S^0.
To differentiate the signals of the B_s^0 decay from the background a multivariate classifier is implemented.

3 Analysis Strategy

In this analysis three samples of data are used. One sample contains the actual data set of the LHCb experiment. Another sample contains simulated events of the $B_s^0 \to \psi(2S)K_S^0$ decay. The third sample contains simulated events of the similar decay $B^0 \to \psi(2S)K_S^0$.
At first, a background training sample is created from the upper sideband of the B_s^0 and a signal training sample is created from the simulated B_s^0 events. The B^0 events are used as control channel.
After this, a feature selection is done and a boosted decision tree is implemented as a multivariate classifier to differentiate between signal like events and combinatorial background. To train this classifier, the training samples are k-folded.
At last, the number of signal events in the data sample is determined and the significance of the result is computed.

4 Analysis

The reconstruction of the invariant mass of the final state particles in the signal channel is shown in figure 3 The distribution shows a clear B^0 peak. Since the branching ratio

plots/BM_Data.png

Figure 3: Invariant mass of final state particles in measured data of the signal channel.

of the B_s^0 decay is much lower, it is not seen in the distribution.

4.1 Define signal and background training samples

The shortest interval in which 99% of the expected B_s^0 decays are is chosen to be the signal window. This interval is not used until the last steps of this analysis and is determined by using the mass distribution of the signal channel MC shown in figure 4. This leads to a signal region in the area of $[5330, 5391]$ MeV. The upper side band

plots/BM_MC.png

Figure 4: Invariant mass distribution of simulated events in the signal channel.

(USB) is in the area > 5339 MeV and is used as a background sample since it mainly consists of combinatorial background. The signal training sample is obtained using the $B_s^0 \to \psi(2S)K_s^0$ simulations.
To get the best features for training the classifier on, it is neccessary to check which features are described well in simulations and which are good in separating background and signal. To evaluate which features are well simulated, data and simulation of every feature is compared for the control channel $B^0 \to \psi(2S)K_s^0$. Since this decay is kinematically similar to the signal channel it is possible to evaluate the features in this channel. To filter the B^0 decays in the data sample, sWeights are used. The invariant mass distribution of the data with and without applied sWeights is shown in figure 5. The weighted distribution shows a sharp peak at the B^0 mass, therefore it is assumed that these are the B^0 events. Since the simulation of the decays are not perfect, kinematic weights are applied on the MC before comparing them with the data. To quantify the similarity of two distributions, a metric is defined:

$$sup_n|F_n^1 - F_n^2|.$$

This metric gives the difference between two distributions F^1 and F^2 while n runs over all bins in the distributions. Well simulated features have a low value when comparing simulation and data. To get the features with the best separating power, the same metric is used when comparing data in the background sample with the simulated signal sample. This time only the well simulated features are compared. A higher value of the metric means a better separation power. To drop the features that are correlated plots/BM_Control_sweights.png

Figure 5: Invariant mass distribution of measured data with and without sWeights

with the reconstructed B mass, the Pearson correlation coefficient is used as a metric [3]. After the determination of the best simulated features with the best separation power without correlation to the B mass, unphysical features and those highly correlated with each other are removed as well.

4.2 Multivariate classification

The XGBoostClassifier [4] is used as a multivariate classifier. The best described and most discriminating features from the last section are used as features to train this classifier. Since there are much more events in the data than in the MC sample, data sample events are all weighted with a factor of 0.5. To train the classifier unbiased, the k-fold method is used: For this, the whole dataset is split into k subsets. One subset is used as a test set while the other $k-1$ subsets are used as training samples. This process is repeated k times so that every subset was tested on once. In this analysis a value of $k = 5$ was used. To cross check if the used BDTs are overtrained, the distribution of the BDT classification on the k-folded training and test sets is examined. One of these distributions is shown in figure 6, while the other four possess a similar distribution. It is seen that background events have classifications near the value of 0 while signal events are near a value of 1. Also the distributions of training and test samples are the same which leads to the conclusion that the BDT is most likely not overtrained. To evaluate the separating power of the BDT, a receiver operating characteristic (ROC curve) is used. This curve is plotting the false positive rate (fpr) against the true positive rate (tpr) of the BDT classification. An ideal classification would have constantly a tpr of 1 independent of the fpr. The area under the curve can be used as a metric to determine plots/TrainTest.png

Figure 6: Comparison of signal and background classifications in training and test subset in one of the k-fold samples.

the separating power of the BDT. Better separating power would result in a larger area under the ROC curve. The ROC curves of the 5 BDTs of the k-fold are shown in figure 7. The curves show an area of ≈ 0.95 showing that the BDTs are well capable of separating signal from background. To compute the ideal cut value of the BDT output, the Punzi figure of merit (FoM) is used as a metric [2]:

$$f = \frac{\epsilon_{\mathrm{BDT}}}{5/2 + \sqrt{N_{\mathrm{bkg}}}}$$

with ϵ_{BDT} as signal efficiency of the BDT and N_{bkg} as estimated background events in the signal region. To obtain the signal efficiency, the number of events before and after cutting on the BDT is compared. To obtain the background events in the signal region, the efficiency of the USB is being computed the same way. It is assumed that the BDT efficiency on background events is the same in the signal region as in the USB. Also it is assumed that the signal region consists mostly of combinatorial background, therefore the number of background events is estimated by neglecting possible signal events.

plots/ROC.png

Figure 7: ROC curves of all 5 BDTs. The blue dashed line shows the case of a total random classification.

plots/FoM.png

Figure 8: Punzi FoM for all k-fold BDTs.

Cutting on the BDT on various points results into the FoM shown in figure 8. The optimal BDT cutting value is the maximum of the FoM. Since 5 BDTs are evaluated at this point, the average of the FoM maxima is being used as the optimal cut value of the BDT classification. This results in an optimal value of 0.8725. The whole dataset is evaluated by all 5 BDTs and the classification of each event is the average of all BDTs. After cutting on the optimal value, this results in an invariant mass distribution shown in figure 9. The distribution shows the clear B^0 peak. To evaluate the small B_s^0 peak,

plots/BM_Data_afterBDT.png

Figure 9: Invariant mass distribution of the dataset after cutting on the BDT classification.

fits need to be made on the distribution.

4.3 Fitting mass distribution

To fit the mass distribution, a model needs to be created. This model consists of three components: a signal model for the signal channel, a control model for the control channel and a background model for combinatorial background.
The signal model consists of two normal distributions with the parameters μ, σ_1, σ_2 as well as the fraction of the first of the two components $frac_{sig}$. Fitting the model to the mass distribution of the signal MC gives the parameter values:

$$\begin{aligned} \mu &= (5367.11 \pm 0.02)\,\mathrm{MeV} \\ \sigma_1 &= 25.8 \pm 0.4 \\ \sigma_2 &= 4.97 \pm 0.02 \\ frac_1 &= 0.053 \pm 0.0014 \end{aligned}$$

For the control model there are two normal distributions used as well. This results in the parameters:

$$
\begin{aligned}
\mu_C &= (5279.92 \pm 0.02)\,\text{MeV} \\
\sigma_{1,C} &= 4.35 \pm 0.03 \\
\sigma_{2,C} &= 11.7 \pm 0.2 \\
frac_1 &= 0.8841 \pm 0.005
\end{aligned}
$$

The fit results of the signal and control MC are shown in figure 10a and 10b.

plots/Fit_MC.png

(a) Signal channel MC.

plots/Fit_Control.png

(b) Control channel MC.

Figure 10: Fits of signal channel and control channel MC.

As a background model, a simple exponential function is used to describe combinatorial background with the exponential parameter λ. Combining all three models using the computed parameters gives the fit shown in figure 11.
The parameters of this fit are:

$$
\begin{aligned}
\lambda &= 0.0020 \pm 0.00045 \\
frac_{control} &= 0.522 \pm 0.009 \\
frac_{signal} &= 0.011 \pm 0.003
\end{aligned}
$$

Since computing the significance would be too complicated for this analysis, a proxy is used to calculate it:

$$
m = \frac{N_{sig}}{\sqrt{N_{sig} + N_{bkg}}}
$$

with N_{sig} and N_{bkg} as the number of signal and background events in the signal region. These numbers are obtained by integrating the models with their parameters inside the plots/Fit.png

Figure 11: Plot of full model as well as its three submodels and data.

signal region. The expected number of events are

$$
\begin{aligned}
N_{sig} &= 45.51 \\
N_{bkg} &= 595.44\,.
\end{aligned}
$$

This results to a significance of $\approx 1.8\sigma$ and therefore not indicating any observation of the decay $B_s^0 \to \psi(2S)K_S^0$.

5 Discussion

The trained BDT showed no signs of overtraining while being very good in separating signal and background. The ROC curve showed a good relation of tpr against fpr. After the BDT classification, almost no background events seem to be inside the upper mass sideband. This indicates a well trained BDT for reducing combinatorial background.
Fitting the MC of the signal and control channel seemed to work very fine. The errors on these fit parameters are small, so this step was not problematic. The fit of the full model was good as well, indicating no errors were made during this step.
With a signifcance of 5σ the decay $B_s^0 \to \psi(2S)K_S^0$ was observed. Further steps in the event selection like training a second BDT to reduce partial reconstructed background events in the lower mass sideband or cutting on particle identification (PID) variables could result in even higher significances.

References

[1] The LHCb Collaboration et al. "The LHCb Detector at the LHC." In: *Journal of Instrumentation* 3.08 (2008).

[2] Giovanni Punzi. "Sensitivity of searches for new signals and its optimization." In: *Proceedings of PHYSTAT2003: Statistical Problems in Particle Physics, Astrophysics, and Cosmology* (Sept. 2003).

[3] Mehreen Saeed. *Calculating Pearson Correlation Coefficient in Python with Numpy.* URL: https://stackabuse.com/calculating-pearson-correlation-coefficient-in-python-with-numpy/.

[4] *XGBoost Documentation.* URL: https://xgboost.readthedocs.io/en/latest/.