Angular Analysis of $B^0 \to K^{*0}\mu^+\mu^-$ Decay in the Search for Beyond Standard Model Physics

Team 10

Imperial College London

March 2022

Outline

Abstract

- We obtain parameters for the angular distribution of the $B^0 \to K^{*0}\mu^+\mu^-$, for decay products up to $q^2 = 19$ GeV^2/c^4
- Manual cutting, neural networks and decision tree methods are explored to separate candidate data from background signal
- Negative loglikelihood minimisation performed to fit angular observables with uniangular distributions
- Obtained parameters show agreement to predicted Standard Model (SM) values almost entirely within less than one standard deviation

Theory: Introduction

- Study the decay:

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

- Dynamics described by three angles $(\theta_l, \theta_K, \phi)$ and the di-μ invariant mass squared q^2

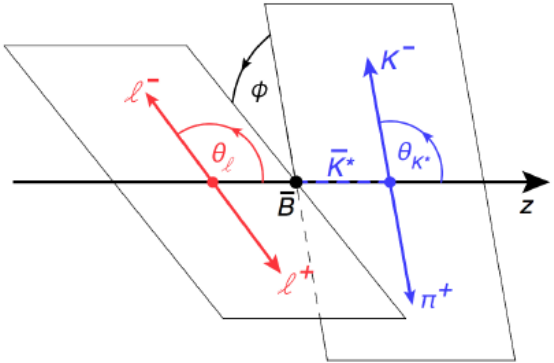


Figure 1: Geometry of Decay

Theory: Features of $B^0 \to K^{*0}\mu^+\mu^-$ decay

- B^0 lifetime means we expect it to decay at a secondary vertex
- K^{*0} decays by strong mechanism, so it is effectively instant with regards to detection
- Expect daughter particle tracks to meet at same point, have considerable impact parameters and transverse momentum
- Expect B^0 to have low impact parameter and transverse momentum

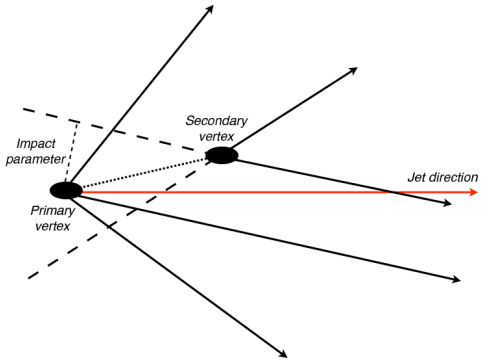


Figure 2: Diagram illustrating primary and secondary vertices [1]

Theory: Angular Distribution

- The full angular distribution follows (up to normalisation) [2]:

$$\begin{aligned}\frac{\mathrm{d}^4\Gamma}{\mathrm{d}q^2\mathrm{d}\cos\theta_l\mathrm{d}\cos\theta_K\mathrm{d}\phi} = \frac{9}{32\pi}\Bigg[& F_L\cos^2\theta_K + \frac{3}{4}(1-F_L)\sin^2\theta_K \\ & + \frac{1}{4}(1-F_L)\sin^2\theta_K\cos 2\theta_l \\ & - F_L\cos^2\theta_K\cos 2\theta_l + S_3\sin^2\theta_K\sin^2\theta_l\cos 2\phi \\ & + S_4\sin 2\theta_K\sin 2\theta_l\cos\phi + S_5\sin 2\theta_K\sin\theta_l\cos\phi \\ & + \frac{4}{3}A_{FB}\sin^2\theta_K\cos\theta_l + S_7\sin 2\theta_K\sin\theta_l\sin\phi \\ & + S_8\sin 2\theta_K\sin 2\theta_l\sin\phi + S_9\sin^2\theta_K\sin^2\theta_l\sin 2\phi\Bigg]\end{aligned}$$

Theory: Angular Distribution

- Integration renders uni-angular distributions [3, 4]:

$$\frac{\mathrm{d}^2\Gamma_p}{\mathrm{d}q^2\mathrm{d}\cos\theta_l} = \frac{3}{8}\left[\frac{3}{2} - \frac{1}{2}F_L + \frac{1}{2}\cos 2\theta_l(1-3F_L) + \frac{8}{3}A_{FB}\cos\theta_l\right] \tag{1}$$

$$\frac{\mathrm{d}^2\Gamma_p}{\mathrm{d}q^2\mathrm{d}\cos\theta_K} = \frac{3}{2}F_L\cos^2\theta_K + \frac{3}{4}(1-F_L)\sin^2\theta_K \tag{2}$$

$$\frac{\mathrm{d}^2\Gamma_p}{\mathrm{d}q^2\mathrm{d}\phi} = \frac{1}{2\pi}\left(S_3\cos 2\phi + S_9\sin 2\phi + 1\right) \tag{3}$$

- Other *optimised* observables can be obtained from these, such as:

$$P_1 = \frac{2S_3}{1-F_L}, \quad P_2 = \frac{2}{3}\frac{A_{FB}}{(1-F_L)}, \quad P_3 = \frac{-S_9}{1-F_L} \tag{4}$$

Candidate Selection: Overview

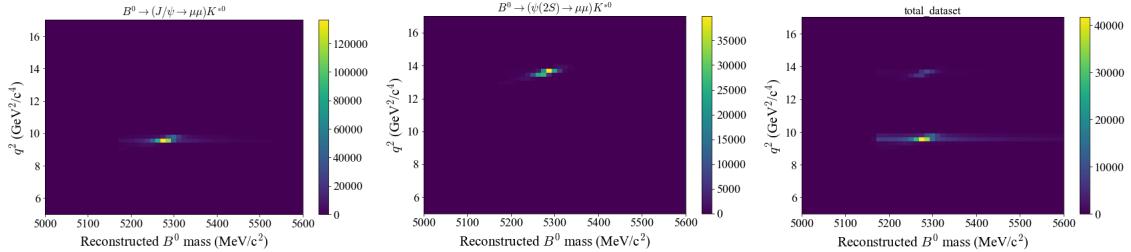
- Multiple approaches were explored for the selection of event candidates from the data
  - Manual cuts
  - Machine learning: Decision Tree
  - Machine learning: Neural Network
  - Combined filtering (Machine Learning and manual cuts)

Manual Cuts - Values

Cut	**Threshold Values**
q^2 Cuts	$0.1 < q^2 < 19$, $q^2 \notin [8.0, 11.0]$, $q^2 \notin [12.5, 15]$
Particle ID Checks	$XProbX > 0.5$ for all particles X
Invariant Mass Checks	$5170 < m(B^0) < 5700$, $790 < m(K^{*0}) < 1000$ $[MeV/c^2]$
Momentum Checks	$P_T(\pi), P_T(K) > 250$, $P_T(\mu^+), P_T(\mu^-) > 800$ $[MeV/c]$
IP and Vertex Checks	$\chi^2_{IP}(B^0) < 16$,$\chi^2_{FD}(B^0) > 64$, $\chi^2_{IP}(\mu, \mu, K, \pi) < 9$,$\chi^2_{EV}(B^0) < 8$, $\chi^2_{EV}(\mu, \mu) < 9$,$\chi^2_{FD}(K^{*0}) < 9$, $\chi^2_{FD}(K^{*0}) > 9$

Table 1: Table of manual cuts

Manual cuts: q^2



(a) Simulated $B^0 \to J/\psi K^{*0}$ (b) Simulated $B^0 \to \psi(2s)K^{*0}$ (c) Unfiltered Dataset

Figure 3: q^2 and $m(K\pi\mu\mu)$ mass of charm resonance simulations and the unfiltered dataset

Machine Learning (ML): Summary

- We can perform a supervised classification task for filtering signal
- Given training examples $(x^{(i)}, y^{(i)})$, where x are our features (the measured decay characteristics) and y is the target (the type of decay), learn a function $h(x)$ that can perform $h : \mathcal{X} \mapsto \mathcal{Y}$ by minimising an objective function.
- Advantages of a machine learning approach
  - Removes need for manual inspection of data
  - Can learn non-linear decisions (between different variables)

Machine Learning (ML): Decisions

1. Perform binary classification (signal vs other)
   - Fewer labels mean a model with the same complexity has higher predictive power
   - Final goal is to filter signal, multiclass classification is not needed
2. Perform manual cuts prior to ML (q^2 and PID) and trained on cut data to prevent domain shift
3. Remove certain features from discussion
   - Variables you are trying to measure e.g. $\cos\theta_k$, $\cos\theta_l$, q^2
   - Variables with no meaning e.g. year
4. Model evaluated with test-train split of 75%, 25% with different random seed
5. Train Model on all of the data
6. Optimise for precision (true signal/ predicted signal)

ML: Ensemble Learning- Random Forest

- Tree-based methods partition the feature space into a series of rectangles/ hyper-cubes
- Trees were trained via binary splitting using the Gini-Impurity metric
- Regularisation prevents overfitting
  - Hyperparameter tuning
  - Ensemble Learning: Bagging 100 trees to make a forest [5]

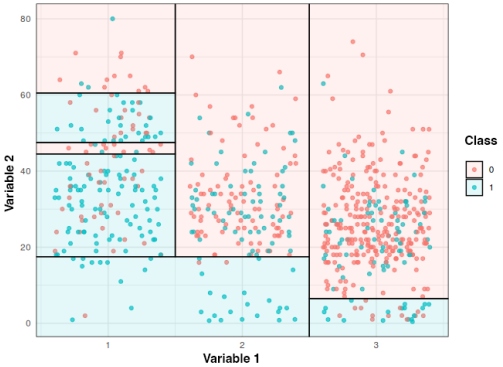


Figure 4: Diagram showing how a decision tree splits up the feature space in 1 dimension [6]

ML: Random Forest Performance

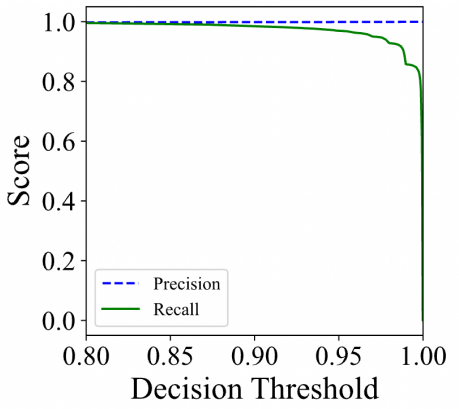


Figure 5: Plot illustrating how the decision threshold can be tuned (default 0.5) to vary the precision and recall

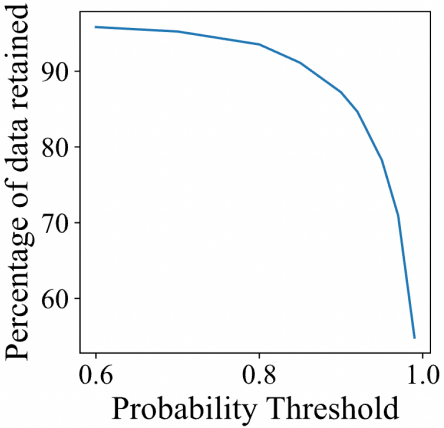


Figure 6: Plot showing how varying the random forest decision threshold changes the number of signal events retained

ML: Neural Network

- A layered network of neurons that apply and change weights of features based on a specified activation function [7]
- Tested for 8-64 neurons in each layer, and 1-3 middle layers [8]
- Compromised with 32 neurons in 2 middle layers, and one dropout layer for regularisation
- The resulting model has a 98.09% accuracy and a 6.81% loss.

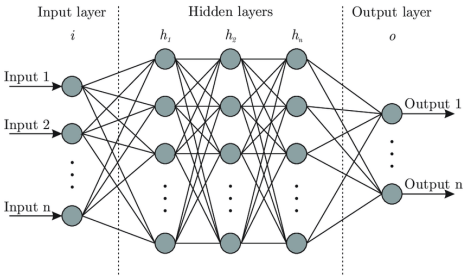
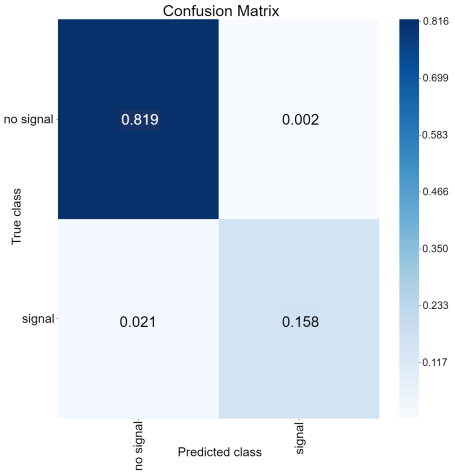


Figure 7: Structure of a neural network [9].

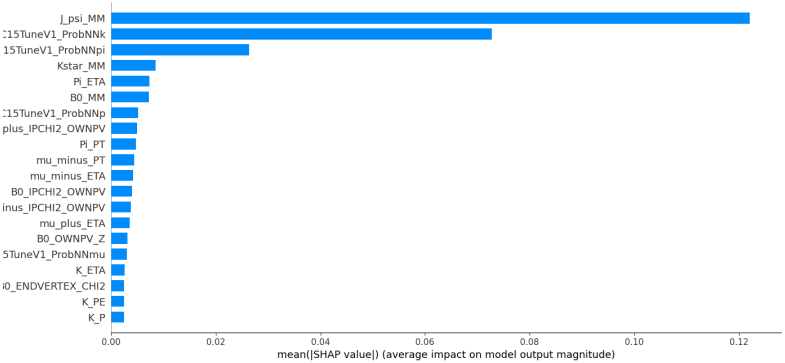
ML: Neural Network

- Confusion Matrix: Binary NN Classification



ML: Neural Network

- SHAP Plot: Binary NN Classification



Fitting: -Log-likelihood Minuit Minimization

PDF without background = angular distribution * obtained acceptance(TESE)

↓

PDF with background = normalised PDF without background + $\frac{f_{bkg}}{1-f_{bkg}}$ (WOE) * fitted background PDF (TESE),

↓

Log-likelihood Minuit Minimization

TESE = To Extract Systematic Error
WOE = With Obtained Error

Figure 8: Flowchart of Fitting Procedure

Fitting: Acceptance function

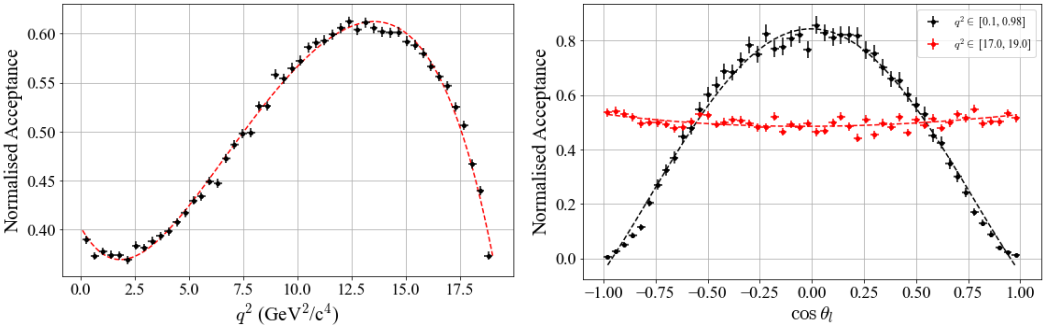
- Acceptance is parameterised in four dimensions as [2, 10]:

$$\varepsilon(\cos\theta_l, \cos\theta_K, \phi, q^2) = \sum_{ijmn} c_{ijmn} L_i(\cos\theta_l) L_j(\cos\theta_K) L_m(\phi) L_n(q^2) \tag{5}$$

  where L_p denotes Legendre polynomials, and all variables are rescaled to the range $[-1, 1]$
- L_p go up to 4th, 5th, 6th and 5th order in $\cos\theta_l$, $\cos\theta_K$, ϕ and q^2 respectiely
- Coefficients c_{ijmn} are determined using method of moments approach on (originally) flat dataset in all angular variables and q^2 [10]
- χ^2 analysis used to determine a suitable polynomial order

Fitting: Acceptance function

- Single-variable acceptance functions can be obtained through the integration of the 4D acceptance function
- For each bin, ε is integrated over other angular variables and the q^2 range of the bin, and then re-normalised



(a) Acceptance function in q^2

(b) Acceptance functions in $\cos\theta_l$ for bins 0 and 6

Figure 9: Acceptance function projections for NN dataset

Fitting: Mass Distribution Model

- The background distribution is modelled by an exponential decay:

$$N_b(m_{K\pi\mu\mu}) = Ae^{-\lambda m_{K\pi\mu\mu}} \tag{6}$$

  where A and λ are variables to be fitted for each q^2 bins
- The signal distribution is modelled by a Gaussian distribution

$$N_s(m_{K\pi\mu\mu}) = Be^{-\frac{1}{2}\left(\frac{m_{K\pi\mu\mu}-m_0}{\sigma}\right)^2} \tag{7}$$

  where B, m_0 and σ are determined by the fit.

Fitting: Signal Yield

- The mass model was used to calculate $B^0 \to K^{*0}\mu^+\mu^-$ signal yield
- Yields were obtained through integration of the mass model
- Signal candidate yields were:
  - 2048 ± 44 for Decision Tree
  - 2112 ± 46 for Neural Network

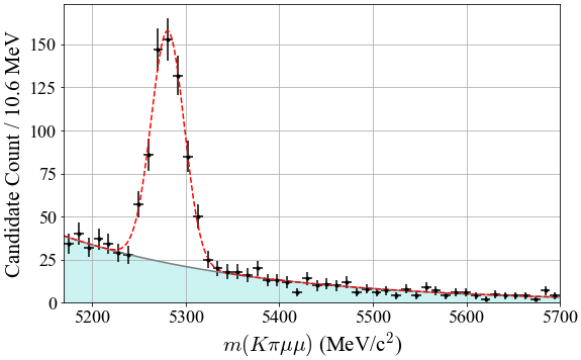


Figure 10: Mass spectrum for $q^2 \in [1.0, 6.0]$ GeV^2/c^4 for Decision Tree data

Fitting: Angular Background Model

- The background model is described by an second-order Chebyshev polynomial:

$$N_b(x) = \sum_{n=0}^{2} c_n T_n(x) \tag{8}$$

where x are angular variables $\cos\theta_k$, $\cos\theta_l$ and ϕ, c_n are the coefficients and T_n are the Chebyshev polynomials.

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1$$

- c_n were obtained by fitting the high mass range $m(K\pi\mu\mu) > 5355$ MeV/c^2

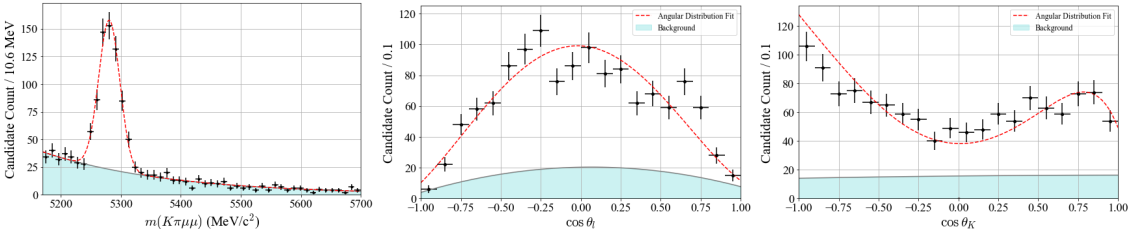
Fitting: Angular Distribution

- The angular distribution fitted, then, is of the form

$$\begin{aligned}\mathcal{P}_{\text{tot}} &= f_{\text{sig}}\mathcal{P}_{\text{sig}} + f_{\text{bkg}}\mathcal{P}_{\text{bkg}} \\ &= (1 - f_{\text{bkg}})\mathcal{P}_{\text{sig}} + f_{\text{bkg}}\mathcal{P}_{\text{bkg}}\end{aligned} \tag{9}$$

  where $\mathcal{P}_{\text{sig}}$ is the signal PDF, with acceptance, $\mathcal{P}_{\text{sig}}$ is the background model, and $f_{\text{sig}}, f_{\text{bkg}}$ are the fraction of signal and background respectively
- f_{bkg} is determined through the integration of the mass model

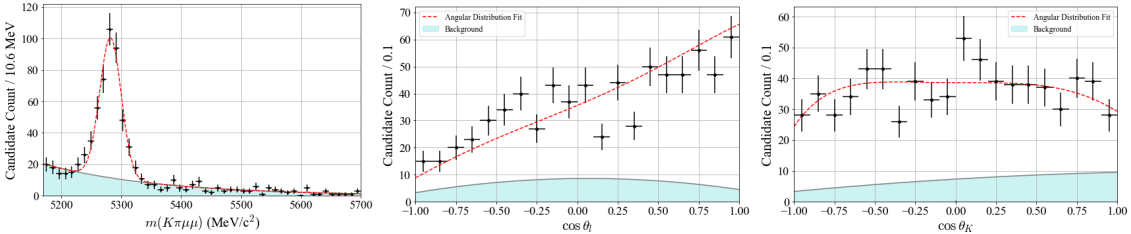
Fitting: Results



(a) $m(K\pi\mu\mu)$ mass distribution (b) $\cos\theta_l$ fit (c) $\cos\theta_K$ fit

Figure 11: Mass spectrum, $\cos\theta_l$ and $\cos\theta_K$ fits for $q^2 \in [1.0, 6.0]$ GeV^2/c^4 for Decision Tree data

Fitting: Results



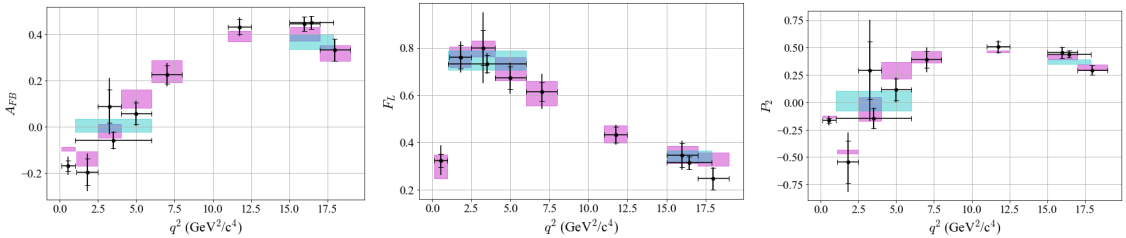
(a) $m(K\pi\mu\mu)$ mass distribution (b) $\cos\theta_l$ fit (c) $\cos\theta_K$ fit

Figure 12: Mass spectrum, $\cos\theta_l$ and $\cos\theta_K$ fits for $q^2 \in [15.0, 17.0]$ GeV^2/c^4 for Decision Tree data

Fitting: Error Analysis

- Statistical uncertainty on the acceptance is found by bootstrapping with the acceptance dataset
  - This error is found to be negligible compared to the uncertainty on the angular observables obtained from the negative loglikelihood minimisation
- Systematic error in the acceptance is found by carrying fits with higher order acceptances
  - Difference between the values obtained in higher and lower order is taken as the error
- Similarly, systematic error in the angular background model is found by performing fits that go up to fourth order
  - Difference between the values obtained in higher and lower order is taken as the error

Fitting: Results for Angular Observables



(a) A_{FB} from different q^2 bins (b) F_L from different q^2 bins (c) P_2 from different q^2 bins

Figure 13: Results for angular observables A_{FB}, F_L, P_2, average of all datasets and distributions fitted. Caps indicate extent of statistical error. Shaded area denotes SM predictions, different colours used to distinguish overlapping bins.

Fitting: Results for Angular Observables

Bin	q^2 range	A_{FB}	F_L	P_2
0	0.1 - 0.98	-0.17 $\pm$ 0.02 $\pm$ 0.01	0.32 $\pm$ 0.03 $\pm$ 0.03	-0.16 $\pm$ 0.02 $\pm$ 0.02
1	1.1 - 2.5	-0.20 $\pm$ 0.06 $\pm$ 0.02	0.76 $\pm$ 0.05 $\pm$ 0.02	-0.5 $\pm$ 0.1 $\pm$ 0.09
2	2.5 - 4.0	0.09 $\pm$ 0.07 $\pm$ 0.05	0.80 $\pm$ 0.07 $\pm$ 0.07	0.3 $\pm$ 0.2 $\pm$ 0.2
3	4.0 - 6.0	0.06 $\pm$ 0.04 $\pm$ 0.008	0.67 $\pm$ 0.05 $\pm$ 0.02	0.11 $\pm$ 0.09 $\pm$ 0.02
4	6.0 - 8.0	0.23 $\pm$ 0.04 $\pm$ 0.01	0.61 $\pm$ 0.04 $\pm$ 0.03	0.39 $\pm$ 0.08 $\pm$ 0.04
5	15.0 - 17.0	0.45 $\pm$ 0.03 $\pm$ 0.0008	0.34 $\pm$ 0.05 $\pm$ 0.01	0.45 $\pm$ 0.05 $\pm$ 0.006
6	17.0 - 19.0	0.33 $\pm$ 0.05 $\pm$ 0.0009	0.25 $\pm$ 0.05 $\pm$ 0.003	0.29 $\pm$ 0.05 $\pm$ 0.001
7	11.0 - 12.5	0.43 $\pm$ 0.03 $\pm$ 0.009	0.43 $\pm$ 0.03 $\pm$ 0.008	0.51 $\pm$ 0.05 $\pm$ 0.01
8	1.0 - 6.0	-0.06 $\pm$ 0.03 $\pm$ 0.003	0.73 $\pm$ 0.04 $\pm$ 0.008	-0.15 $\pm$ 0.09 $\pm$ 0.009
9	15.0 - 17.9	0.45 $\pm$ 0.03 $\pm$ 0.0005	0.31 $\pm$ 0.038 $\pm$ 0.002	0.44 $\pm$ 0.03 $\pm$ 0.0009

Table 2: Table of results for A_{FB}, F_L and P_2, averages of all fits. The first error is statistical and the second is systematic

Conclusion

- Several Candidate Selection methods were explored, resulting in hybrid ML and manual cuts approaches with neural networks and decision trees
- Negative loglikelihood minimisation fit was performed for uniangular distributions in each q^2 bin, taking into account angular acceptance and background model
- Agreement found with SM in measured observables, to within less than one standard deviation in almost all cases
- Hence we do not find conclusive evidence of beyond SM physics

References and works consulted I

[1] A. Coccaro, "Track Reconstruction and b-Jet Identification for the ATLAS Trigger System," CERN, Geneva, Tech. Rep., Dec 2011. [Online]. Available: https://cds.cern.ch/record/1402986

[2] R. Aaij *et al.*, "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb-1 of integrated luminosity," *Journal of High Energy Physics*, vol. 2016, no. 2, p. 104, Feb 2016. [Online]. Available: https://doi.org/10.1007/JHEP02(2016)104

[3] ——, "Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$," *Phys. Rev. Lett.*, vol. 108, p. 181806, May 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.108.181806

[4] J. Matias, F. Mescia, M. Ramon, and J. Virto, "Complete anatomy of $\bar{b}_d \to \bar{k}^{*0}(\to k\pi)\mu^+\mu^-$ and its angular distribution," *Journal of High Energy Physics*, vol. 2012, no. 4, p. 104, Apr 2012. [Online]. Available: https://doi.org/10.1007/JHEP04(2012)104

[5] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: https://doi.org/10.1023/a:1010933404324

[6] P. v. d. Laken and B. M. says:, "Visualizing decision tree partition and decision boundaries," Mar 2020. [Online]. Available: https://paulvanderlaken.com/2020/03/31/visualizing-decision-tree-partition-and-decision-boundaries/

[7] M. Williams, "Machine learning in particle physics," in *APS April Meeting Abstracts*, vol. 2018, 2018, pp. D05–003.

[8] D. Bourilkov, "Machine and deep learning applications in particle physics," *International Journal of Modern Physics A*, vol. 34, no. 35, p. 1930019, 2019.

[9] L. Shukla, "Designing your neural networks," *Towards Data Science, September*, 2019.

[10] F. J. Kress, "Angular Analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay at the LHCb experiment," Ph.D. dissertation, Imperial College London, 2020, presented 14 Aug 2020. [Online]. Available: https://cds.cern.ch/record/2732006

[11] T. Aaltonen *et al.*, "Measurements of the angular distributions in the decays $B \to K^{(*)}\mu^+\mu^-$ at cdf," *Phys. Rev. Lett.*, vol. 108, p. 081807, Feb 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.108.081807

[12] V. Khachatryan *et al.*, "Angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from pp collisions at $\sqrt{s} = 8$ Tev," *Physics Letters B*, vol. 753, pp. 424–448, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0370269315009685

References and works consulted II

[13] B. Aubert *et al.*, "Angular distributions in the decay $B \to K^* l^+ l^-$," *Phys. Rev. D*, vol. 79, p. 031102, Feb 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.79.031102

[14] R. Aaij *et al.*, "Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$," *Journal of High Energy Physics*, vol. 2013, no. 8, p. 131, Aug 2013. [Online]. Available: https://doi.org/10.1007/JHEP08(2013)131

[15] S. Chatrchyan *et al.*, "Angular analysis and branching fraction measurement of the decay $B^0 \to K^{*0}\mu^+\mu^-$," *Physics Letters B*, vol. 727, no. 1, pp. 77–100, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0370269313008101

[16] B. Aubert *et al.*, "Measurements of branching fractions, rate asymmetries, and angular distributions in the rare decays $B \to K\ell^+\ell^-$ and $B \to K^*\ell^+\ell^-$," *Phys. Rev. D*, vol. 73, p. 092001, May 2006. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.73.092001

[17] R. Aaij *et al.*, "Test of lepton universality with $B^0 \to K^* l^+ l^-$ decays," *Journal of High Energy Physics*, vol. 2017, no. 8, p. 55, Aug 2017. [Online]. Available: https://doi.org/10.1007/JHEP08(2017)055

[18] S. Coquereau, "Study of the $B^0 \to K^{*0}\mu^+\mu^-$ decay with the LHCb experiment: angular analysis and measurement of the ratio RK," Theses, Université Pierre et Marie Curie - Paris VI, Sep. 2015. [Online]. Available: https://tel.archives-ouvertes.fr/tel-01267248

[19] M. Williams, V. V. Gligorov, C. Thomas, H. Dijkstra, J. Nardulli, and P. Spradlin, "The HLT2 Topological Lines," CERN, Geneva, Tech. Rep., Jan 2011. [Online]. Available: https://cds.cern.ch/record/1323557

[20] LHCb Collaboration *et al.*, "Angular analysis of $B^0 \to D^{*-} D_s^{*+}$ with $D_s^{*+} \to D_s^+\gamma$," *JHEP*, vol. 2021, no. 6, Jun. 2021. [Online]. Available: https://doi.org/10.1007/JHEP06(2021)177

[21] M. Aaboud *et al.*, "Angular analysis of $B_d^0 \to K^{*0}\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ Tev with the ATLAS detector," *Journal of High Energy Physics*, vol. 2018, no. 10, p. 47, Oct 2018. [Online]. Available: https://doi.org/10.1007/JHEP10(2018)047

[22] M. Williams, "Machine learning in particle physics," Aug 2018. [Online]. Available: https://indico.fnal.gov/event/15893/contributions/34315/attachments/21384/26602/Fermilab_ML_Lecture1.pdf

References and works consulted III

[23] D. Bourilkov, "Machine and deep learning applications in particle physics," *International Journal of Modern Physics A*, vol. 34, no. 35, p. 1930019, 2019. [Online]. Available: https://doi.org/10.1142/S0217751X19300199

[24] S. Shramm, "Machine learning at CERN: ATLAS, LHCb, and more," *Proceedings of Science*, vol. 340, Aug 2019. [Online]. Available: https://doi.org/10.22323/1.340.0158

[25] M. De Cian, S. Farry, P. Seyfert, and S. Stahl, "Fast neural-net based fake track rejection in the LHCb reconstruction," CERN, Geneva, Tech. Rep., Mar 2017. [Online]. Available: https://cds.cern.ch/record/2255039

[26] HEP ML Community, "A Living Review of Machine Learning for Particle Physics." [Online]. Available: https://iml-wg.github.io/HEPML-LivingReview/

[27] L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman, "Weakly supervised classification in high energy physics," *Journal of High Energy Physics*, vol. 2017, no. 5, p. 145, May 2017. [Online]. Available: https://doi.org/10.1007/JHEP05(2017)145

[28] A. Galli, A. Petrolini, M. Riani, and E. Simonotto, "Use of neural network classifiers in high-energy physics: search for the Higgs boson," in *Applications of Artificial Neural Networks III*, S. K. Rogers, Ed., vol. 1709, International Society for Optics and Photonics. SPIE, 1992, pp. 990 – 1002. [Online]. Available: https://doi.org/10.1117/12.139975

[29] A. Rogozhnikov, "Hep-ml," Sep 2021. [Online]. Available: https://pypi.org/project/hep-ml/

[30] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, ser. Springer Texts in Statistics. Springer New York, 2013.