Centro de Astropartículas y Física de Altas Energías
Universidad Zaragoza

Departamento de Física Teórica
Universidad Zaragoza

Using Machine Learning techniques in phenomenological studies in flavour physics

Jorge Alda,
Universidad de Zaragoza/CAPA jalda@unizar.es

Based on **JA**, J. Guasch, S. Peñaranda
arXiv:2109.07405 [hep-ph]
Saturnalia, 21st December 2022

B anomalies

Experimental results for the decays of B mesons that don't conform[ed] with the SM predictions.

In particular, these anomalies show hints of violation of the Leptonic Flavour Universality (LFU).

B anomalies

Flavour-changing neutral currents $b \to s\ell^+\ell^-$, with $\ell = e, \mu$.

- Universality ratios $R_{K^{(*)}}$

$$R_{K^{(*)}} = \frac{\mathrm{BR}(B \to K^{(*)}\mu^+\mu^-)}{\mathrm{BR}(B \to K^{(*)}e^+e^-)}\,,$$

  In the SM, theoretically clean, and $R_{K^{(*)}} = 1$.
  LHCb measurements[1]:
  - $R_{K^+} = 0.846^{+0.042}_{-0.039}{}^{+0.013}_{-0.012}$,
  - $R_{K^{*0}} = 0.685^{+0.113}_{-0.069} \pm 0.047$. ($3.1\,\sigma$ tension)
- Angular observables for $B \to K^*\ell^+\ell^-$: P_4', P_5' ...
- Also $B_s \to \phi\mu^+\mu^-$ decays.

[1] R. Aaij *et al.* (LHCb) arXiv:2103.11769, arXiv:1705.05802

B anomalies

Flavour-changing charged currents $b \to c\ell\nu$, with $\ell = e/\mu, \tau$.

- Universality ratios $R_{D^{(*)}}$

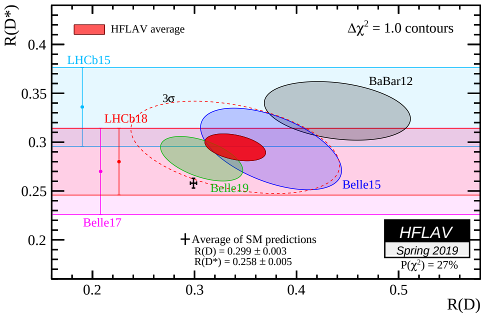
$$R_{D^{(*)}} = \frac{\mathrm{BR}(B \to D^{(*)}\tau\nu)}{\mathrm{BR}(B \to D^{(*)}\ell\nu)}\,,$$

$$R_D = 0.340 \pm 0.027 \pm 0.013\,,$$

$$R_{D^*} = 0.295 \pm 0.011 \pm 0.008\,.$$

Combined tension of $3.08\,\sigma$.[1]

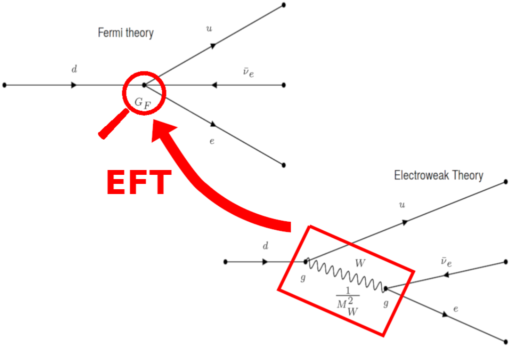
- Longitudinal polarization of D^*.
- Also $B_c \to J/\psi\ell\nu$ decays.



[1] Y. S. Ahmis *et al.* (HFLAV) arXiv:1909.12524

Effective Field Theories

We want to investigate whether these anomalies could be originated by some New Physics (NP) beyond the Standard Model. How can we examine [almost] all models at once?



The NP particles are much heavier than the energy scales of the B decays $\implies$ their effects are limited to alterations of the interaction vertices, and these alterations are suppressed by powers of p/Λ. We can study these in a systematical way.

Effective Field Theories

Effective Field Theories (EFTs) describe deviations from the SM in a model-independent way, using effective operators of dim > 4 and their corresponding Wilson coefficients.

At energies Λ above the electroweak scale, the dim-6 SMEFT (2499 operators)[1]

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \frac{1}{\Lambda^2} \sum C_i O_i \,.$$

We will focus on the following operators at $\Lambda = 1\,\mathrm{TeV}$:

$$O_{\ell q(1)}^{ijkl} = (\bar{\ell}_i \gamma_\mu \ell_j)(\bar{q}_k \gamma^\mu q_l), \qquad O_{\ell q(3)}^{ijkl} = (\bar{\ell}_i \gamma_\mu \tau^I \ell_j)(\bar{q}_k \gamma^\mu \tau^I q_l)\,.$$

[1] B. Grzadkowski, M. Iskrzynski, M. Misiak and J. Rosiek. arXiv:1008.4884

Effective Field Theories

Our setting: in the interaction basis, NP only affects the third generation,[1]

$$C_1 = C^{3333}_{\ell q(1)}\,, \qquad\qquad C_3 = C^{3333}_{\ell q(3)}\,.$$

We have to rotate to the mass basis, using rotation matrices λ^q and λ^ℓ,

$$C^{ijkl}_{\ell q(1)} = C_1 \lambda^{ij}_\ell \lambda^{kl}_q \qquad\qquad C^{ijkl}_{\ell q(3)} = C_3 \lambda^{ij}_\ell \lambda^{kl}_q\,.$$

The rotation matrices must be hermitian, idempotent and with $\mathrm{tr}\lambda = 1$. Parameterized as

$$\lambda = \frac{1}{1+|\alpha|^2+|\beta|^2}\begin{pmatrix} |\alpha|^2 & \alpha\beta^* & \alpha \\ \alpha^*\beta & |\beta|^2 & \beta \\ \alpha^* & \beta^* & 1 \end{pmatrix}\,.$$

We will fix $C_1 = C_3$ to avoid large deviations in the $B \to K^{(*)}\nu\bar{\nu}$ decays.

[1]F. Feruglio, P. Paradisi and A. Pattori. arXiv:1705.00929

Global fits

- The structure of the λ matrices creates Lepton Flavour Violating effects through λ_ℓ^{ij} $(i \neq j)$.
- RG evolution causes mixing between effective operators:
  - For example, $O_{\ell q(1)}$ mixes with $O_{\varphi\ell(1)} = (\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi)(\bar{\ell}\gamma^\mu \ell)$ that modifies the $Z\ell^+\ell^-$ couplings.

We need to consider the effects of the New Physics in a wide range of experimentally-measured observables: Global Fits.

Global fits

Tools used for the global fits (in `python3`):

- `wilson`[1]: RG evolution and mixing of the Wilson coefficients.
- `flavio`[2]: Calculation of observables in EFTs and database of experimental measurements.
- `smelli`[3]: Global likelihood function:

$$\Delta \log L = -\frac{1}{2}\Delta\chi^2 = -\frac{1}{2}\sum_{i,j}[\mathcal{O}_i^{\rm exp} - \mathcal{O}_i^{\rm th}(\vec{C})](\mathcal{C}^{\rm exp} + \mathcal{C}^{\rm th}(C))^{-1}_{ij}[\mathcal{O}_j^{\rm exp} - \mathcal{O}_j^{\rm th}(\vec{C})]\,.$$

$\mathcal{O}_i^{\rm exp}$ is the experimental measurements for the i-th observable, and $\mathcal{O}_i^{\rm th}$ its SMEMT prediction. $\mathcal{C}_{ij}^{\rm exp}$ and $\mathcal{C}^{\rm th}(C)$ are the covariance matrices (usually block-diagonal).

[1] J. Aebischer, J. Kumar and D. M. Straub. arXiv:1804.05033
[2] D. M. Straub. arXiv:1810.08132
[3] J. Aebischer, J. Kumar, P. Stangl and D. M. Straub. arXiv:1810.07698

Global fits

We perform global fits including $b \to s\ell^+\ell^-$ and $b \to c\ell\nu$ observables, and also electroweak precision tests, nuclear precision observables (superallowed β decays) and Leptonic Flavour Violating observables.

We consider two different scenarios:

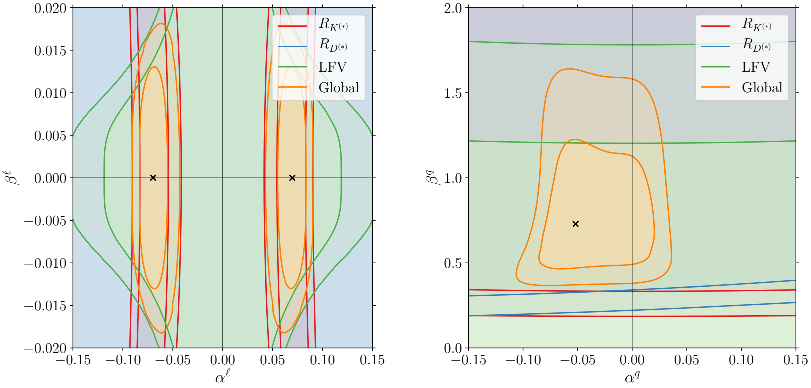
- **Scenario I:** Mixing to first generation α are negligible[1].
- **Scenario II:** Mixing to first and second generations[2].

	Scenario I	Scenario II
C	-0.13 ± 0.05	-0.13 ± 0.08
α^ℓ		$\pm(0.07^{+0.04}_{-0.07})$
β^ℓ	0 ± 0.025	0 ± 0.025
α^q		$-0.05^{+0.12}_{-0.07}$
β^q	$0.8^{+2.0}_{-0.5}$	$0.73^{+2.8}_{-0.6}$
$\Delta\chi^2_{\mathrm{SM}}$	40.32	57.06
SM Pull	$5.75\,\sigma$	$6.57\,\sigma$

[1] F. Feruglio, P. Paradisi and A. Pattori. arXiv:1705.00929
[2] **J.A.**, J. Guasch and S. Peñaranda. arXiv:2109.07404

Global fits: Results

Regions allowed at $1\ \sigma$:

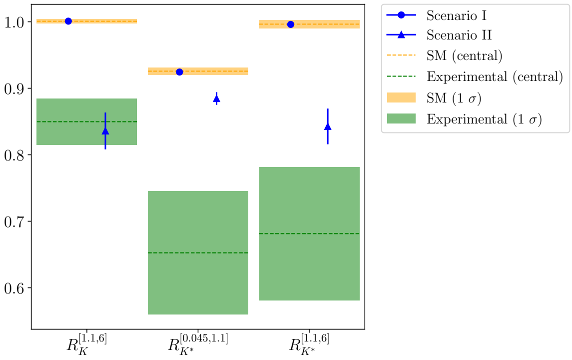


- α^ℓ is constrained by $R_{K^{(*)}}$.
- β^ℓ is constriained by LFV observables.

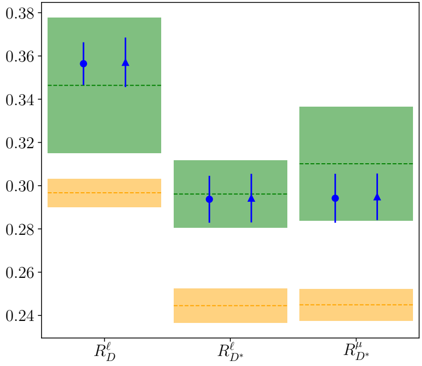
J.A., J. Guasch and S. Peñaranda. arXiv:2109.07404

Global fits: Results

$$R_{K^{(*)}} = \frac{\mathrm{BR}(B\to K^{(*)}\mu^+\mu^-)}{\mathrm{BR}(B\to K^{(*)}e^+e^-)}$$



$$R_{D^{(*)}} = \frac{\mathrm{BR}(B\to D^{(*)}\tau\nu)}{\mathrm{BR}(B\to D^{(*)}\ell\nu)}$$



$R_{K^{(*)}}$ requires mixing with the first generation (α^ℓ)

J.A., J. Guasch and S. Peñaranda. arXiv:2109.07404

Exploring the likelihood function

- We would like to explore the properties of the likelihood function around the best fit point.
- **Problem:** At each point $(C, \alpha^\ell, \beta^\ell, \alpha^q, \beta^q)$ requires the calculation of 471 observables, many of them need numerical integration.
- In a previous work[1], we used the Hessian approximation,

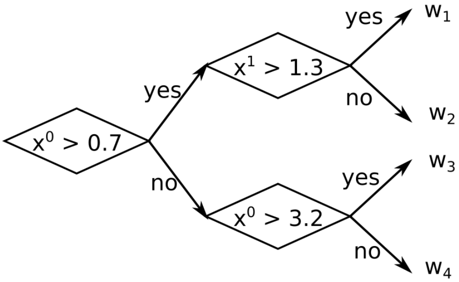
$$\Delta \log L(\vec{C}) \approx \Delta \log L_{\mathrm{bf}} + \vec{C}^T \mathbb{H} \vec{C}\,.$$

- Doesn't work in this case because of non-linear relation between fit parameters and Wilson coefficients (equi-probability surfaces are not ellipsoids).
- **Idea:** Teach a machine to approximate $\log L(\vec{C})$.

[1] **J.A.**, J. Guasch and S. Peñaranda. arXiv:2012.14799

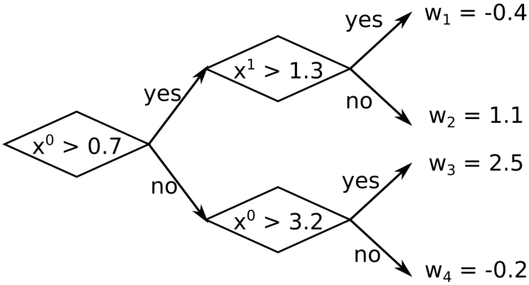
xgboost

Decision tree



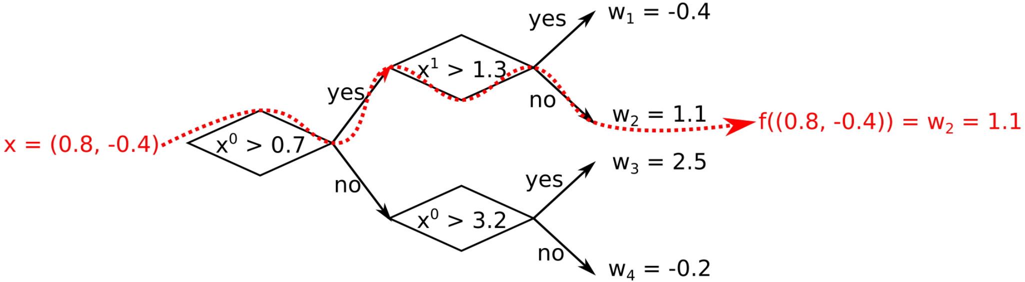
xgboost

Regression tree



xgboost

Regression tree



A regression tree k defines a function $f_k(x)$. We use an ensemble of trees F for the prediction:

$$\phi(x) = \sum_{k \in F} f_k(x) \,.$$

xgboost

Extreme Gradient Boosting (xgboost)[1].

We have a dataset $\{(x_i, y_i)\}$ with $x_i \in X(\sim \mathbb{R}^n)$ and $y_i \in \mathbb{R}$, and want to find $\phi(x)$ so $\phi(x_i) \approx y_i$. Defining the regularized objective

$$\mathcal{L}[\phi] = \sum_i \ell(\phi(x_i), y_i) + \sum_k \Omega(f_k)\,.$$

- $\ell(\tilde{y}_i, y_i)$ is the loss function (MAE: $|\phi(x_i) - y_i|$).
- $\Omega(f_k)$ penalizes the complexity of the trees.

The objective is optimized iteratively:

- The algorithm starts with one tree with a single leaf.
- At each step, one new tree with one more leaf is added to the ensemble (splitting).
- To prevent overfitting, the weights of the newly added leaf is multiplied by $\eta < 1$ (shrinkage).

[1] T. Chen and C. Guestrin. arXiv:1603.02754

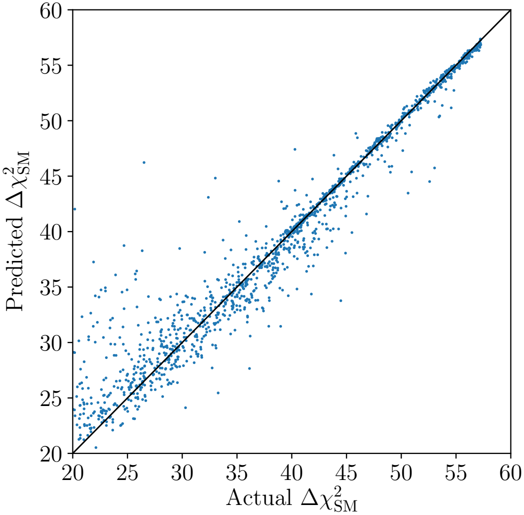
Machine Learning training

Data sample consisting of

- 5000 points re-used from likelihood plots.
- 5000 random points.
- Split in 75% training set, 25% validation set.

Results of the training:

- Pearson regression coefficient $r = 0.971$.
- Mean Absolute Error 0.655.



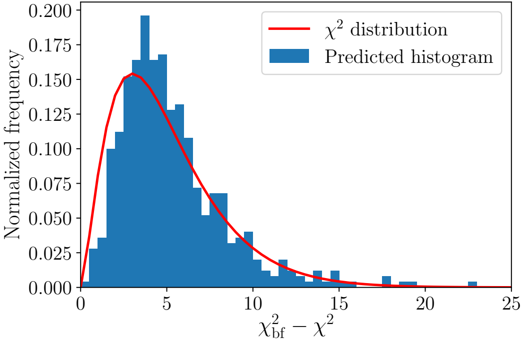
J.A., J. Guasch and S. Peñaranda. arXiv:2109.07404

Machine-Learning Montecarlo

We want to generate a sample of points distributed according to the χ^2 of the fit.
We generate random points, that are accepted if

$$\log \tilde{L}(\vec{C}) = \log L_{\text{bf}} + \log u\,,$$

with u a random number from the uniform distribution in $[0,1)$.



We use the trained model $\log \tilde{L}(\vec{C})$ to compute an approximation of the likelihood function.

J.A., J. Guasch and S. Peñaranda. arXiv:2109.07404

SHAP values

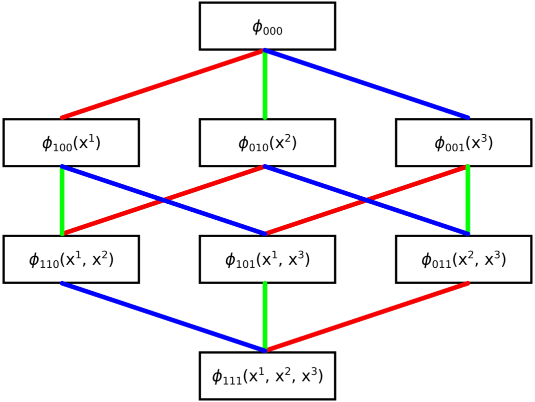
Measure of the importance of each parameter in the Machine Learning prediction for a single datapoint[1]. Based on game theory (Lloyd Shapley, Nobel Prize in Economics '12).

- **Local accuracy:** The sum of all SHAP values (plus a constant) is the ML prediction.
- **Missingness:** If some parameter is missing, its SHAP value is zero.
- **Consistency:** If the model is changed so one parameter has a larger impact, its SHAP value will increase.

[1] S. Lundberg, S. Lee. arXiv:1705.07874

SHAP values

We train 2^n simplified models



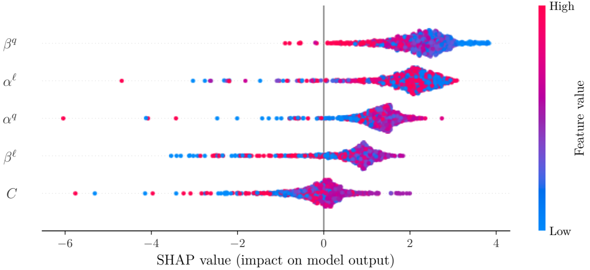
- Marginal contributions when adding x^1: $\phi_{100} - \phi_{000}$, $\phi_{110} - \phi_{010}$, $\phi_{101} - \phi_{001}$, $\phi_{111} - \phi_{011}$.
- SHAP value for x^1 is a weighted average of all its marginal contributions.
- ϕ_{000} is the constant value.

Efficient implementation (poly time) for tree-based Machine Learning models.[1]

[1] S. Lundberg, G. Erion, S. Lee. arXiv:1802.03888

SHAP values: results

The mixing β^q to the second quark generation and α^ℓ to the first lepton generation are in general the most important features.



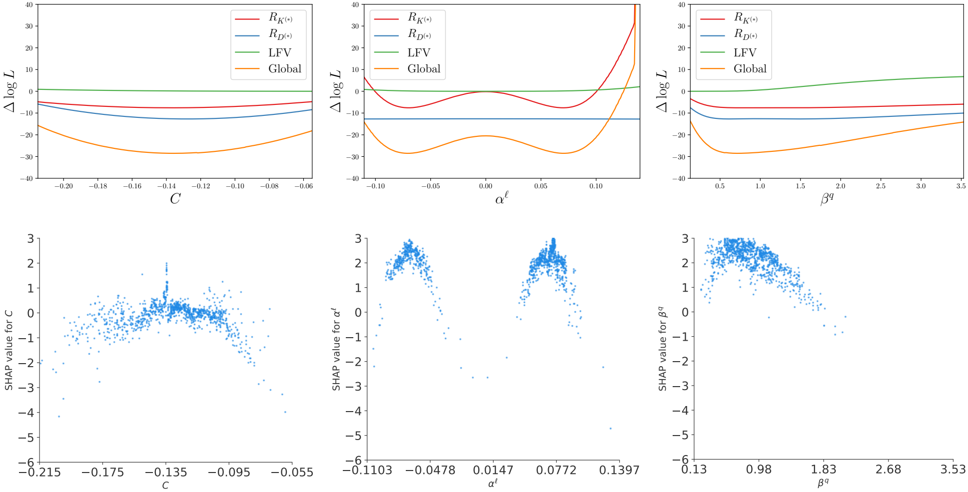
For the best fit point, the value of C is also important:

Base value	SHAP value for					Final prediction	Actual $\log L$
	C	α^ℓ	β^ℓ	α^q	β^q		
39.43	3.293	4.056	1.993	2.671	4.086	55.537	57.06

J.A., J. Guasch and S. Peñaranda. arXiv:2109.07404

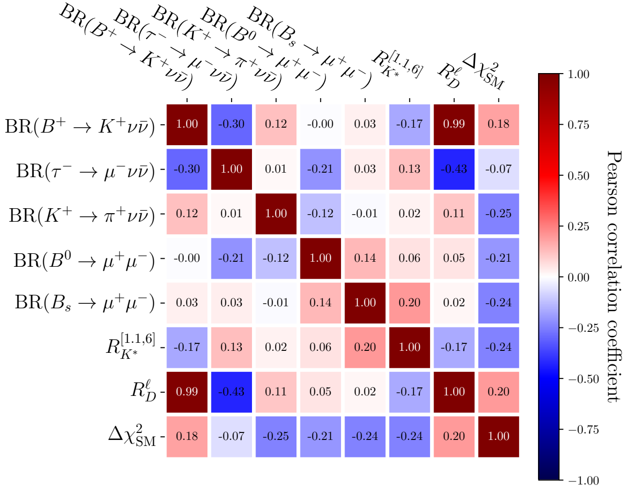
SHAP values: results

SHAP importances reproduce the dependence of the $-\log L$:



J.A., J. Guasch and S. Peñaranda. arXiv:2109.07404

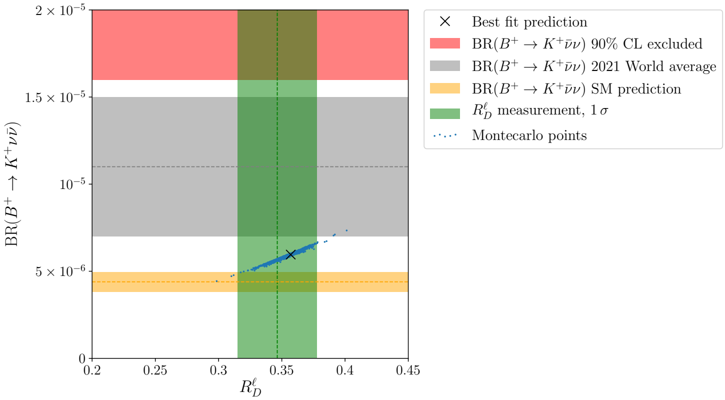
Correlations between observables



- Moderate correlation between R_K and $\mathrm{BR}(B_s \to \mu^+\mu^-)$ because $C_9^\mu \neq C_{10}^\mu$.
- Also moderate correlation between R_K and R_D.
- Perfect correlation between R_D and $\mathrm{BR}(B \to K^{(*)}\nu\bar{\nu})$.
- No observable displays large correlations to the global likelihood: global fits are needed.

J.A., J. Guasch and S. Peñaranda. arXiv:2109.07404

R_D and $\mathrm{BR}(B \to K^{(*)}\nu\bar{\nu})$



An excess in R_D implies an excess in $\mathrm{BR}(B \to K^{(*)}\nu\bar{\nu})$.
(Note that the 2021 World Average is not included in our fit).

J. Grygier *et al.* (Belle) arXiv:1702.03224; F. Dattola (Belle-II) arXiv:2105.05754; Y. S. Ahmis *et al.* (HFLAV) arXiv:1909.12524; **J.A.**, J. Guasch and S. Peñaranda. arXiv:2109.07404

Conclusions

- Using EFTs (SMEFT) to describe the $R_{K^{(*)}}$ and $R_{D^{(*)}}$ anomalies: $R_{D^{(*)}}$ generated at tree level and $R_{K^{(*)}}$ from interplay between tree and loop level.
- Global fits are needed to take in account possible side-effects.
- `xgboost` Machine Learning model is able to approximate the likelihood function.
- SHAP values capture the importance of each parameter in the fit.
- Interesting correlation between $R_{D^{(*)}}$ and $\mathrm{BR}(B \to K^{(*)} \nu\bar{\nu})$.
- Explanation in terms of vector leptoquark U_1 coupling to second and third generations of quarks and third ($R_{D^{(*)}}$) and first ($R_{K^{(*)}}$) generation of leptons.

Updated experimental values in late 2022

- Provisional values for $R_{D^{(*)}}$ from LHCb [https://indico.cern.ch/event/1187939]. Overall tension remains at $3.3\ \sigma$.
- Values for $R_{K^{(*)}}$ superseded by LHCb, arXiv:2212:09152-3 (submitted yesterday):

$$R_K^{\text{low}} = 0.994 \pm 0.086 \pm 0.028, \qquad R_{K^*}^{\text{low}} = 0.927 \pm 0.090 \pm 0.035,$$

$$R_K^{\text{cen}} = 0.949 \pm 0.041 \pm 0.022, \qquad R_{K^*}^{\text{cen}} = 1.027 \pm 0.070 \pm 0.027.$$

The $R_{K^{(*)}}$ anomaly has disappeared (only $0.2\ \sigma$)!

Back-up slides

Effective Field Theories

The B anomalies are defined at $\mu = m_b$, we need to integrate the heavy SM particles. We obtain the WET:

- $$\left.\begin{aligned} O_9^\ell &= (\bar{s}_L \gamma_\alpha b_L)(\bar{e}_\ell \gamma^\alpha e_\ell) \\ O_{10}^\ell &= (\bar{s}_L \gamma_\alpha b_L)(\bar{e}_\ell \gamma^\alpha \gamma_5 e_\ell) \end{aligned}\right\} \implies R_{K^{(*)}}.$$
- $$O_{VL}^\ell = (\bar{c}_L \gamma_\alpha b_L)(\bar{e}_{\ell\,L} \gamma^\alpha \nu_\ell) \implies R_{D^{(*)}}.$$
- $$O_\nu^\ell = (\bar{s}_L \gamma_\alpha b_L)(\bar{\nu}_\ell (1-\gamma_5)\nu_\ell) \implies \mathrm{BR}(B \to K^{(*)} \nu\bar{\nu}).$$

Effective Field Theories

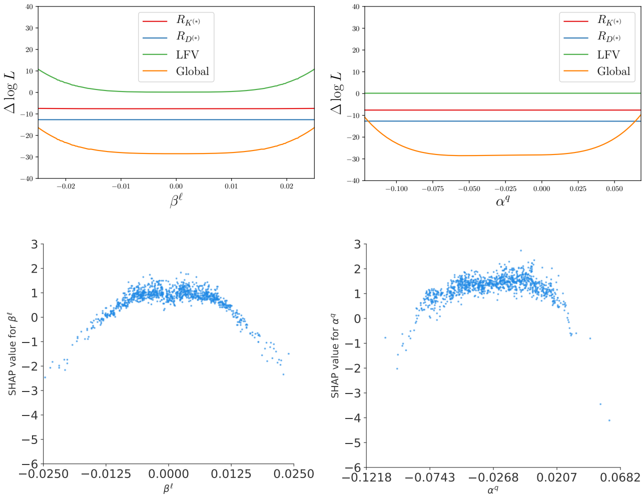
Translating between the two EFTs: 1-loop RG running $\Lambda \to \mu_{\mathrm{EW}}$ + matching:

- $C_9^\ell \approx \frac{\sqrt{2}}{V_{tb}V_{ts}^* G_F \Lambda^2} \left[\frac{2\pi^2}{e^2}(C_{\ell q(1)}^{\ell\ell 23} + C_{\ell q(3)}^{\ell\ell 23}) + \frac{\log(m_b/\Lambda)}{3}(C_{\ell q(1)}^{3323} + C_{\ell q(3)}^{3323})\right].$
- $C_{10}^\ell \approx \frac{-\sqrt{2}}{V_{tb}V_{ts}^* G_F \Lambda^2} \frac{2\pi^2}{e^2}(C_{\ell q(1)}^{\ell\ell 23} + C_{\ell q(3)}^{\ell\ell 23}).$
- $C_{VL}^\ell \approx \frac{-1}{\sqrt{2}G_F\Lambda^2}\left(C_{\ell q(3)}^{\ell\ell 33} + \frac{V_{cs}}{V_{cb}} C_{\ell q(3)}^{\ell\ell 23}\right).$
- $C_\nu^\ell \approx \frac{\sqrt{2}}{e^2 V_{tb}V_{ts}^* G_F \Lambda^2}\left[2\pi^2(C_{\ell q(1)}^{\ell\ell 23} - C_{\ell q(3)}^{\ell\ell 23}) + \frac{2g'^2}{3}\log(m_b/\Lambda) C_{\ell q(3)}^{\ell\ell 23}\right].$

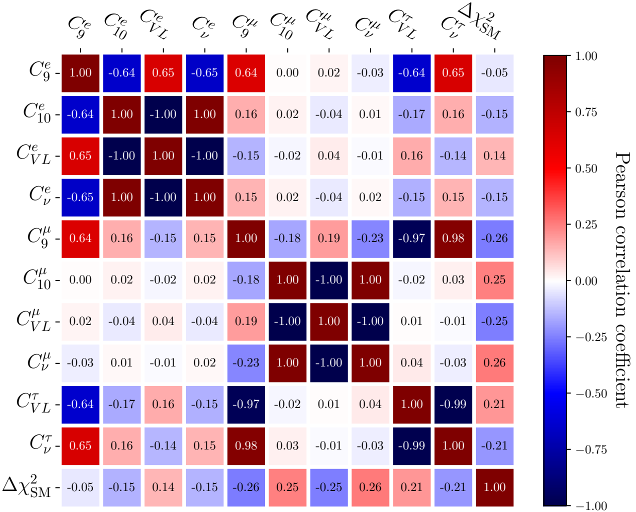
Loop-level corrections break the relation $C_9^\ell = -C_{10}^\ell$.

J.A., J. Guasch and S. Peñaranda. arXiv:2109.07404

Backup slides



Correlations between WCs



- Correlation in C_{10}^e, C_{VL}^e and C_ν^e, all proportional to $C\lambda_{23}^q\lambda_{11}^\ell$.
- Correlation in C_{10}^μ, C_{VL}^μ and C_ν^μ, all proportional to $C\lambda_{23}^q\lambda_{22}^\ell$.
- Correlation in C_{VL}^τ and C_ν^τ, all proportional to $C\lambda_{23}^q\lambda_{33}^\ell$.
- $C_9^\mu = C_9^{\mathrm{loop}}$ correlated to τ sector ($C\lambda_{23}^q \approx C\lambda_{23}^q\lambda_{33}^\ell$).
- $C_9^e = -C_{10}^e + C_9^{\mathrm{loop}}$ partially correlated to e and τ sectors.

Connection to Leptoquarks

Vector Leptoquark $U_1 \sim (\bar{\mathbf{3}}, \mathbf{1}, 2/3)$,

$$\mathcal{L}_{U_1} = x_L^{ij} \bar{q}_i \gamma_\mu U_1^\mu \ell_j + x_R^{ij} \bar{d}_{Ri} \gamma_\mu U_1^\mu e_{Rj} + \text{h.c.}$$

Matching to the SMEFT Wilson coefficients

$$C_{\ell q(1)}^{ijkl} = C_{\ell q(3)}^{ijkl} = \frac{-\Lambda^2}{2M_U^2} x_L^{li} x_L^{kj*} \,,$$

$$C_{ed}^{ijkl} = -\frac{1}{2} C_{qde}^{ijkl} = \frac{-\Lambda^2}{2M_U^2} x_R^{li} x_R^{kj*} \,,$$

In terms of the parameters of our fit,

$$|x_L^{ji}|^2 = -\frac{2}{M_U^2} C \lambda_{ii}^\ell \lambda_{jj}^q \,,$$

$$\text{Arg}(x_L^{ji}) = \text{Arg}(\lambda_{j3}^q) - \text{Arg}(\lambda_{i3}^\ell) + \theta \,.$$

Connection to leptoquarks

Couplings to vector leptoquark U_1 in Scenario II (mixing to first and second generation):

$$x_L = \begin{pmatrix} -2.27 \times 10^{-3} & -3.76 \times 10^{-10} & -0.0325 \\ 0.0319 & 5.29 \times 10^{-9} & 0.458 \\ 0.0437 & 7.25 \times 10^{-9} & 0.627 \end{pmatrix} \approx \begin{pmatrix} 0 & 0 & 0 \\ x_1 & 0 & x_3 \\ x_1 & 0 & x_3 \end{pmatrix} .$$

- Couplings x_L^{23} and x_L^{33}, previously proposed to describe the $R_{D^{(*)}}$ anomaly[1]. Compatible with experimental limits.
- Additionally, couplings x_L^{21} and x_L^{31} to describe the $R_{K^{(*)}}$ anomaly.

Other leptoquark models, and W' and Z' models, do not generate $C_{\ell q(1)} = C_{\ell q(3)}$.

[1] A. Bhaskar, D. Das, T. Mandal, S. Mitra and C. Neeraj. 2101.12069