Belle II Data Preservation Workshop, Rome, 7 October 2022

Theory perspective on Data Preservation: Experimental data in Flavio

Peter Stangl CERN

flavio: what can it do for me?

https://github.com/flav-io/flavio
D. Straub, arXiv:1810.08132

flavio: main features

1. Computing theory predictions

for a large number of observables (flavour physics, electroweak precision observables, Higgs physics, ...)

- **Standard Model** (SM) predictions
- Predictions in the presence of **new physics** (NP) (parameterized by Wilson coefficients)
- Theory **uncertainties** for SM and NP

flavio: main features

2. Database of experimental measurements

for all implemented observables that have been measured

- provided in terms of YAML file
- easy to update and extend

$\rightarrow$ more on this later!

flavio: main features

3. Likelihoods

Combining predictions with experimental data allows constructing likelihoods

- Likelihoods in parameters (e.g. CKM parameters) or Wilson coefficients
- Possibility to use Gaussian approximation for **fast likelihood** estimates
- Use external fitters to perform Bayesian or frequentist statistics with `flavio` likelihoods
- Basis for the **smelli** global **SME**FT **L**ike**LI**hood Python package

https://github.com/smelli/smelli
Aebischer, Kumar, PS, Straub, arXiv:1810.07698
PS, arXiv:2012.12211

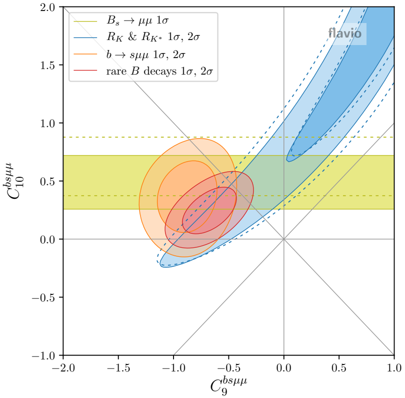
flavio: main features

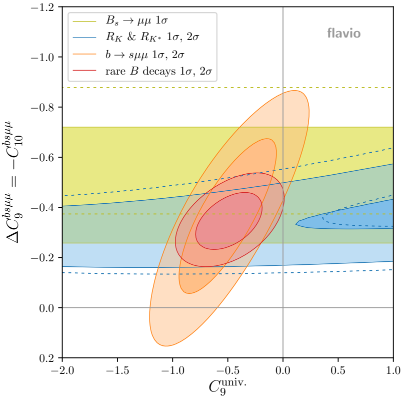
4. Plots

- Visualize experimental measurements & theory predictions
- Visualize your likelihoods

flavio: showcase

New physics in B-decays in Weak effective theory Wilson coefficients @ 4.8 GeV

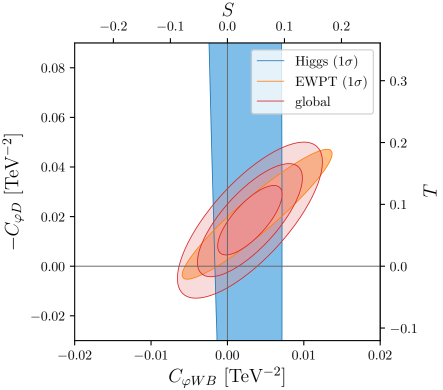




Altmannshofer, PS, arXiv:2103.13370

flavio: showcase

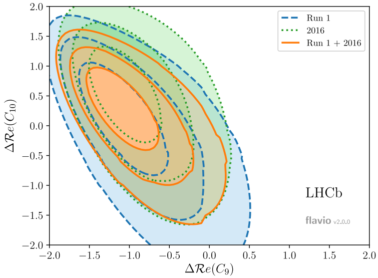
S-T fit using combined Higgs and electroweak likelihood in SMEFT

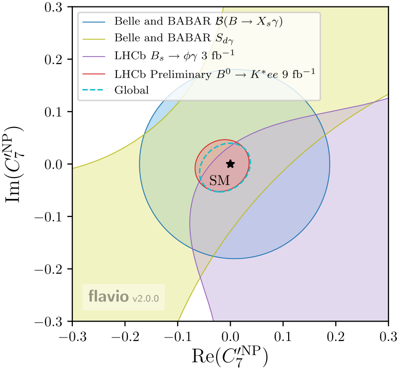


Falkowski, Straub, arXiv:1911.07866

flavio: showcase

Fits to new physics Wilson coefficients from recent LHCb analyses





LHCb-PAPER-2020-002
LHCb-TALK-2020-155

Measurements in flavio

Measurements in flavio

- A **measurement** in `flavio` corresponds to a **likelihood** associated with one (or more) **observable(s)**
- The likelihood is provided by experiments and included in `flavio`
- Predefined measurements contained in file `measurements.yaml`
- Simple example in `flavio` YAML format:

```
Belle Rmue(B->D*lnu) 2017:                    # name
  experiment: Belle                           # experiment
  inspire: Abdesselam:2017kjf                 # INSPIRE tex key
  values:
    Rmue(B->D*lnu): 0.96 ± 0.05 ± 0.01       # constraint
```

  - The constraint is interpreted as Gaussian likelihood
  - Different error contributions are summed in quadrature

Measurements in flavio

- Observables can have **asymmetric uncertainties**

```
K+->pinunu NA62 2019:                           # name
  experiment: NA62                              # experiment
  values:
    BR(K+->pinunu): 0.47 + 0.72 - 0.47 e-10     # constraint
```

  - The constraint is interpreted as combination of two half Gaussians

Measurements in flavio

- For measurement of **correlated observables**, the correlation can be specified

```
Belle RD* sl 2019:                                # name
  experiment: Belle                               # experiment
  inspire: Abdesselam:2019dgh                     # INSPIRE tex key
  values:
    Rtaul(B->Dlnu): 0.307 ± 0.037 ± 0.016        # constraint observable 1
    Rtaul(B->D*lnu): 0.283 ± 0.018 ± 0.014       # constraint observable 2
  correlation: -0.53                              # correlation
```

  - The constraints are interpreted as 2D Gaussian likelihood
  - Different error contributions are summed in quadrature

Measurements in flavio

- For more than two correlated observables, the **correlation matrix** can be specified (in quadratic or triangular form)

```
A_l SLD:                                          # name
  experiment: SLD                                 # experiment
  inspire: ALEPH:2005ab                           # INSPIRE tex key
  description: Table 3.6 of arXiv:hep-ex/0509008  # description
  values:
    A(Z->ee): 0.1516 ± 0.0021               # constraint observable 1
    A(Z->mumu): 0.142 ± 0.015               # constraint observable 2
    A(Z->tautau): 0.136 ± 0.015             # constraint observable 2
  correlation: [[1.0, 0.038, 0.033], [1.0, 0.007], [1.0]]#corr.matrix
```

  - The constraints are interpreted as 3D Gaussian likelihood

Measurements in flavio

- Simple **upper limits** on positive quantities can be specified

```
Belle B->K*emu 2018:                             # name
  experiment: Belle                              # experiment
  inspire: Sandilya:2018pop                      # INSPIRE tex key
  values:
    BR(B0->K*emu): < 1.2e-7 @ 90% CL             # constraint observable 1
    BR(B0->K*mue): < 1.6e-7 @ 90% CL             # constraint observable 2
```

  - The constraints are interpreted as half Gaussians with mode at 0 and cumulative probability below the limit being 0.90 (in this example)

Measurements in flavio

- More appropriate **upper limit** likelihood for low-statistics counting experiments if **number of events** is available

```
DELPHI Z LFV:                                   # name
  experiment: DELPHI                            # experiment
  inspire: Abreu:1996mj                         # INSPIRE tex key
  values:
    BR(Z->mutau):                               # constraint
      distribution: gamma_upper_limit           # type of distribution
      limit: 1.2e-5
      confidence_level: 0.95
      counts_total: 0
      counts_background: 0
```

  - The constraint is implemented as a Gamma distribution

Measurements in flavio

- If number of expected background events is uncertain, **background uncertainty** can be specified

```
Belle B->hnunu SL 2017:                              # name
  experiment: Belle                                  # experiment
  inspire: Grygier:2017tzo                           # INSPIRE tex key
  values:
    BR(B0->K*nunu):                                  # constraint
      distribution: general_gamma_upper_limit        # type of distribution
      limit: 1.8e-5
      confidence_level: 0.9
      counts_total: 13
      counts_signal: -2.0
      background_uncertainty: 1.8
```

  - The constraint is implemented as a Gamma distribution convoluted with a (folded) Gaussian for the background uncertainty

Measurements in flavio

- Arbitrary **non-Gaussian** likelihoods can be given in **numerical form**

```
LHCb RK 2021:
  experiment: LHCb
  url: http://moriond.in2p3.fr/2021/EW/slides/3_flavour_02_moise.pdf
  values:
    - name: <Rmue>(B+->Kll)       # constraint for binned observable
      q2min: 1.1                  # lower bin boundary
      q2max: 6.0                  # upper bin boundary
      value:
        distribution: numerical   # type of distribution
        x: [0.55, 0.56, 0.57, 0.58, 0.59, 0.60, 0.61, 0.62, ... ]
        y: [4.592421315031857e-16, 5.5901132147031404e-15, ... ]
```

  - `y` value of the likelihood with arbitrary normalisation
  - number of points should not be too small as likelihood is interpolated only linearly between them
  - likelihood is assumed to be zero outside of provided range, so range must be large enough, including possible tails

Measurements in flavio

- **Numerical likelihoods** can be specified in **arbitrary number of dimensions**, e.g.

```
LHCb Bs->mumu 2021:
  experiment: LHCb
  url: http://moriond.in2p3.fr/2021/EW/slides/3_flavour_01_archilli.
      pdf
  observables:
    - BR(Bs->mumu)
    - BR(B0->mumu)
  values:
    distribution: multivariate_numerical
    xi: [[8.161097450987314e-10, 9.172370129183739e-10, ... ],
         [-4.806317281473294e-13, 1.2345785573114215e-11, ... ]]
    y: [[1.3143971288076154e-07, 1.675574140528865e-07,  ... ],
        [3.878158386282714e-07, 4.923565057071156e-07, ... ],
        [1.2014436902795998e-06, 1.5185447893919208e-06, ... ],
        [3.5512584579519983e-06, 4.468647340339764e-06, ... ],
        ...
        [1.1380897209303134e-06, 1.1676827880277567e-06, ... ]]
```

  - N dimensional grid given by `xi` values (here 2D)
  - `y` provides value of the likelihood on the grid as N-dim array

Approximations used for likelihoods

Gaussian likelihoods

- Central values and uncertainties
  - have to be approximated as Gaussian
  - correlations have to be neglected
- Central values and uncertainties + correlations
  - multivariate Gaussian
  - often good approximation close to central values
  - deviations from Gaussian have to be neglected

Upper limits

- Upper limit with confidence level
  - can be approximated as Half-Gaussian
  - not necessarily a good approximation
- Upper limit with confidence level + event counts
  - can be modelled by Gamma distribution
  - more appropriate than Half-Gaussian for low-statistics counting experiment
- Upper limit with confidence level + event counts + background uncertainty
  - can be modelled by Gamma distribution convoluted with (folded) Gaussian

Generic non-Gaussian likelihoods

- Likelihoods in parameterized functional form
  - e.g. (convolutions of) parameterized probability distributions
  - can be stored in relatively compact way
- Numerical likelihoods on a grid
  - precision can be increased by finer grid
  - might require large amounts of data to be stored
- Samples from the likelihood distribution
  - precision can be increased by larger number of samples
  - might require large amounts of data to be stored
- New ideas?
  - Neural Networks trained to represent a likelihood
  - ...

Example: BR($B_{d,s} \to \mu\mu$)

Example: BR($B_{d,s} \to \mu\mu$)

LHCb, arXiv:1703.05747

- Most simple result: uncorrelated central values and uncertainties

Example: BR($B_{d,s} \to \mu\mu$)

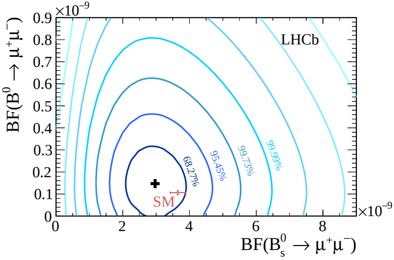
LHCb, arXiv:1703.05747

- Most simple result: uncorrelated central values and uncertainties
- More information: correlation between BR($B_d \to \mu\mu$) and BR($B_s \to \mu\mu$)

Example: BR($B_{d,s} \to \mu\mu$)

LHCb, arXiv:1703.05747

- Most simple result: uncorrelated central values and uncertainties
- More information: correlation between BR($B_d \to \mu\mu$) and BR($B_s \to \mu\mu$)
  - Non-Gaussian likelihood: contour plot

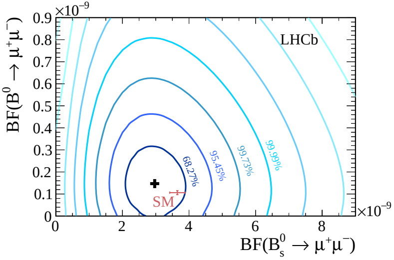


Example: BR($B_{d,s} \to \mu\mu$)

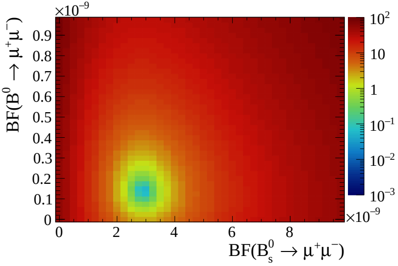
LHCb, arXiv:1703.05747

- Most simple result: uncorrelated central values and uncertainties
- More information: correlation between BR($B_d \to \mu\mu$) and BR($B_s \to \mu\mu$)

  - Non-Gaussian likelihood: contour plot



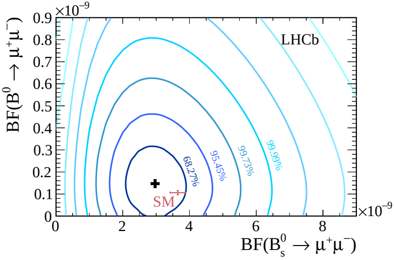
  - Non-Gaussian likelihood: heat map



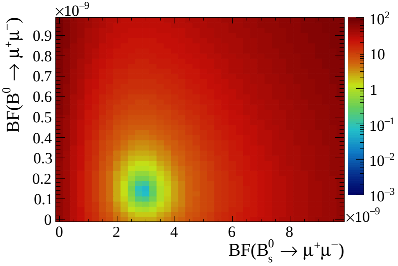
Example: BR($B_{d,s} \to \mu\mu$)

LHCb, arXiv:1703.05747

- Most simple result: uncorrelated central values and uncertainties
- More information: correlation between BR($B_d \to \mu\mu$) and BR($B_s \to \mu\mu$)
  - Non-Gaussian likelihood: contour plot



  - Non-Gaussian likelihood: heat map



- Even better: **numerical data** instead of only plots!

Wishlist

Wishlist

- Correlation matrices for measurements of correlated observables
- Event counts (and background uncertainties) for upper limits
- Likelihoods in parameterized functional form (e.g. parameterized Gamma distribution, convolutions of different parameterized distributions, etc.)
- Full non-Gaussian likelihoods
- Likelihoods in numerical form instead of only plots!
- Data files instead of only tables in PDF files!

Wishlist

- Correlation matrices for measurements of correlated observables
- Event counts (and background uncertainties) for upper limits
- Likelihoods in parameterized functional form (e.g. parameterized Gamma distribution, convolutions of different parameterized distributions, etc.)
- Full non-Gaussian likelihoods
- Likelihoods in numerical form instead of only plots!
- Data files instead of only tables in PDF files!

Thank you!

Backup slides