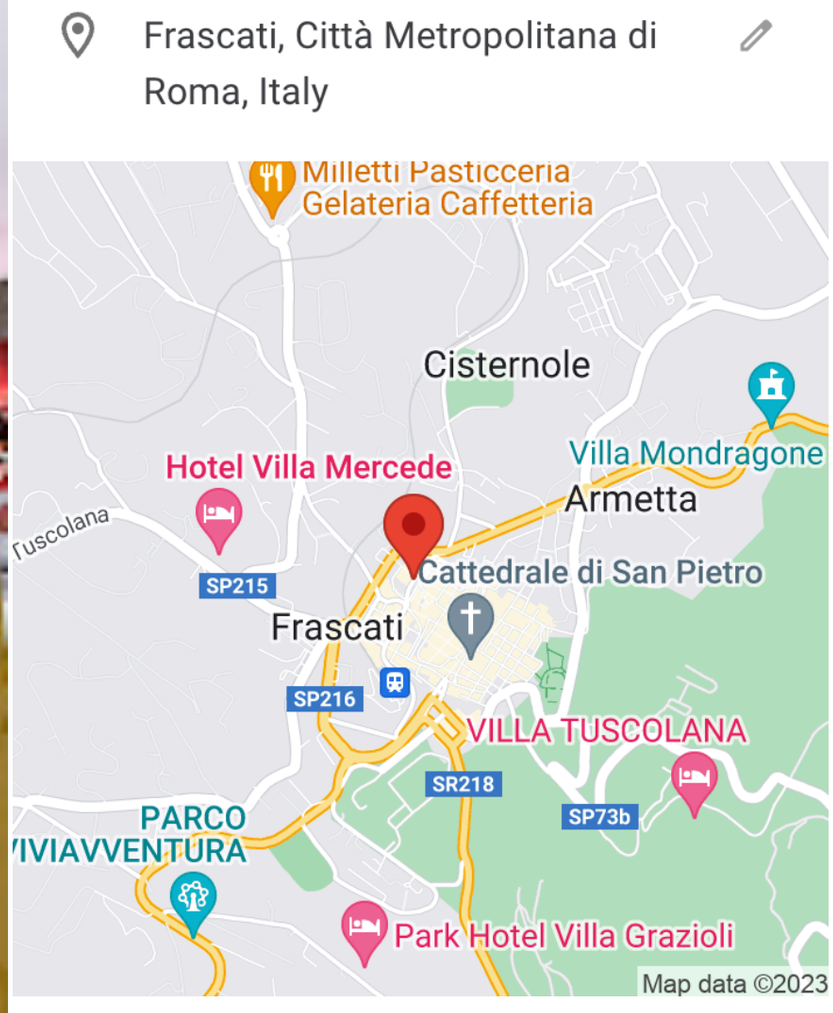
Fitting tools for semileptonics



Patrick Owen

13/04/23

Fitting in semileptonics

- A few features of fitting for semileptonics:
  - Don’t tend to parameterise functions analytically.
  - PDF shapes depend on decay model.
  - Multi-dimensional fits common to extract most info.
  - Large yields.
  - Low purity - strong dependence on background.
  - Simulation size leads to non-negligible systematics.

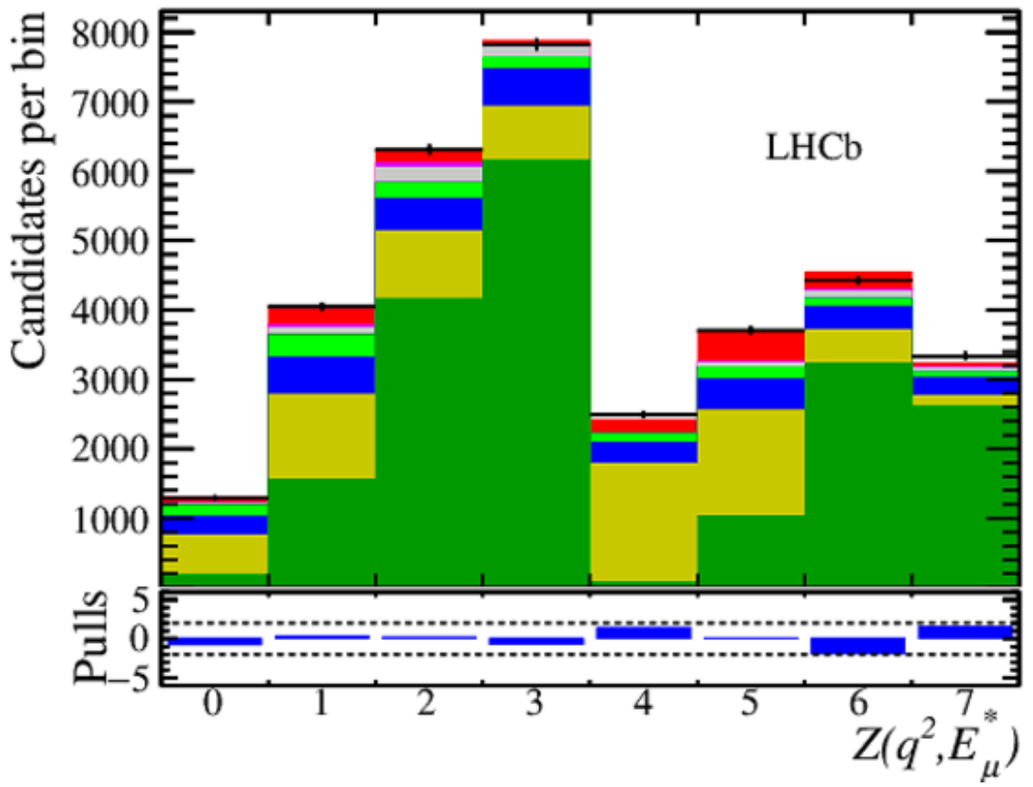
Fitting in semileptonics

- A few features of fitting for semileptonics:
  - Don’t tend to parameterise functions analytically. Template fitting
  - PDF shapes depend on decay model. Shape variations during minimisation
  - Multi-dimensional fits common to extract most info. 1D projections insufficient for fit quality
  - Large yields. Efficient data management needed
  - Low purity - strong dependence on background. Control samples and ad-hoc variations
  - Simulation size leads to non-negligible systematics. Barlow-Beeston or bootstrapping

RooFit implementation

- RooFit still most common tool.
  - Templates provided by RooHistPDF.
  - Limited to three explicit dimensions.
- Why histograms and not KDE?
  - Much faster and resolution wide c.f. bin width.

LHCb-PAPER-2017-035



- RooHistFactory builds upon RooFit.
  - Provides Barlow-Beeston-lite implementation.
  - Also allows ad-hoc systematic variations.
  - Nice tutorial from Phoebe (old but gold) [link]

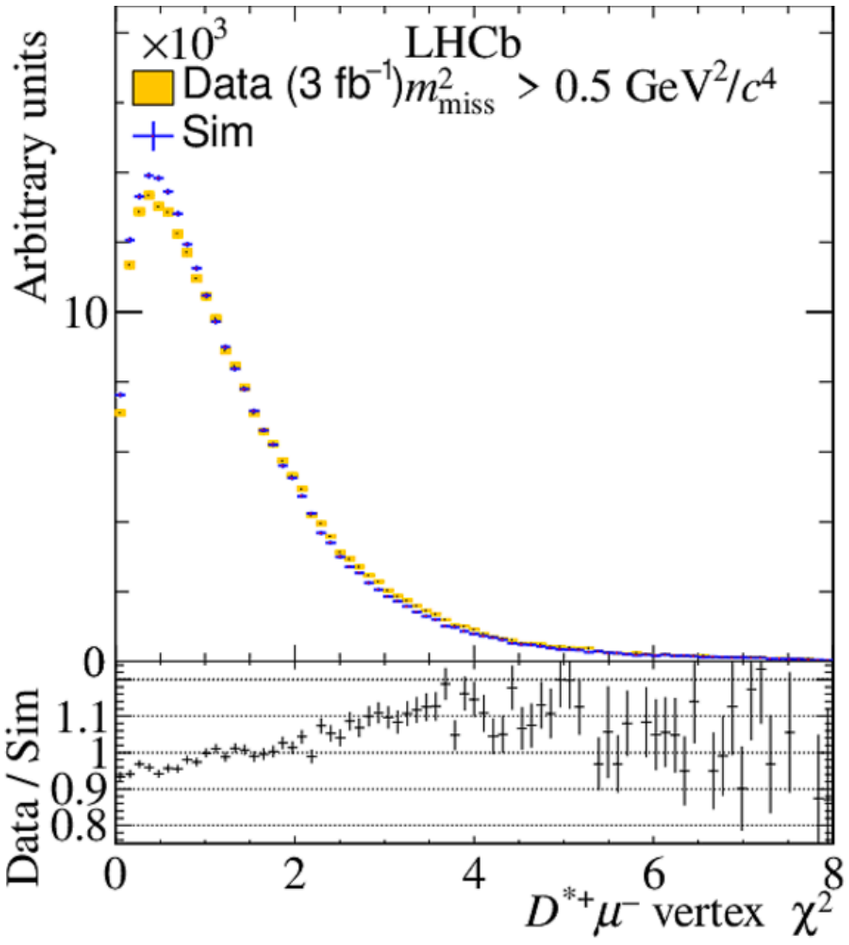
```
if(cfg->MCstat())
{
  DstTauNu.ActivateStatError();
  DplusTauNu.ActivateStatError();
  DplusMuNu.ActivateStatError();
  DstMuNu.ActivateStatError();
  Bd2DD.ActivateStatError();
  Bu2DD.ActivateStatError();
  Bu2Dstst.ActivateStatError();
}
```

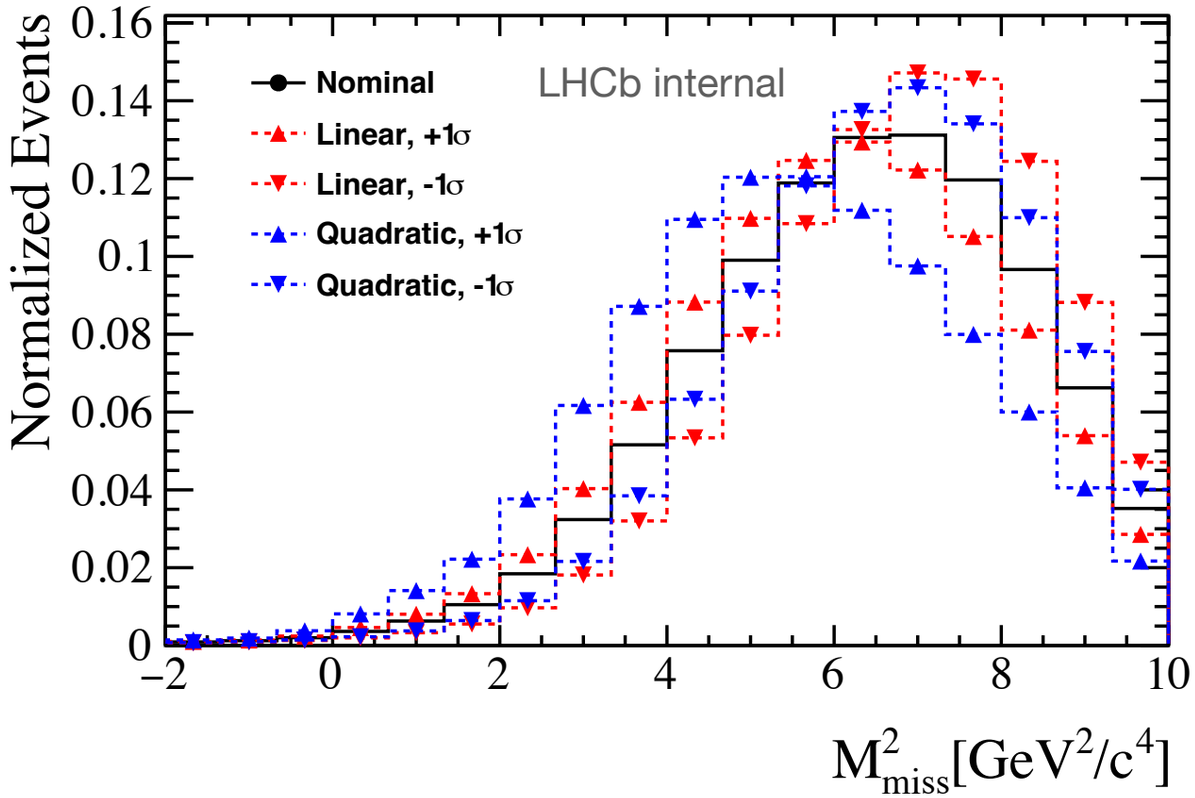
- If one only has one missing neutrino, more options open up.

Systematic shape variations

- Partial reconstruction imprints shape dependence.
- Decay model dealt with separately (see later), but what about detector effects such as data/MC agreement?
- HistFactory systematic variations allow for polynomial interpolation between 1σ variations.

LHCb-PAPER-2022-039





Decay model

- What about the decay model? Could also use HistFactory variations.
  - This ignores correlations in the uncertainties.
  - Fine for systematics but what if the shapes are parameters of interest? e.g. form factors or Wilson Coefficients.
- Brute force solution - event-by-event reweighing for every likelihood call.
  - Reweighting exact, no assumptions on correlations or interpolation.
  - Prohibitively slow, even for smaller size yields.
- One get out: If we measure form factors of muonic mode, uncertainties should be sub-leading and then one doesn’t need to vary them in the fit.

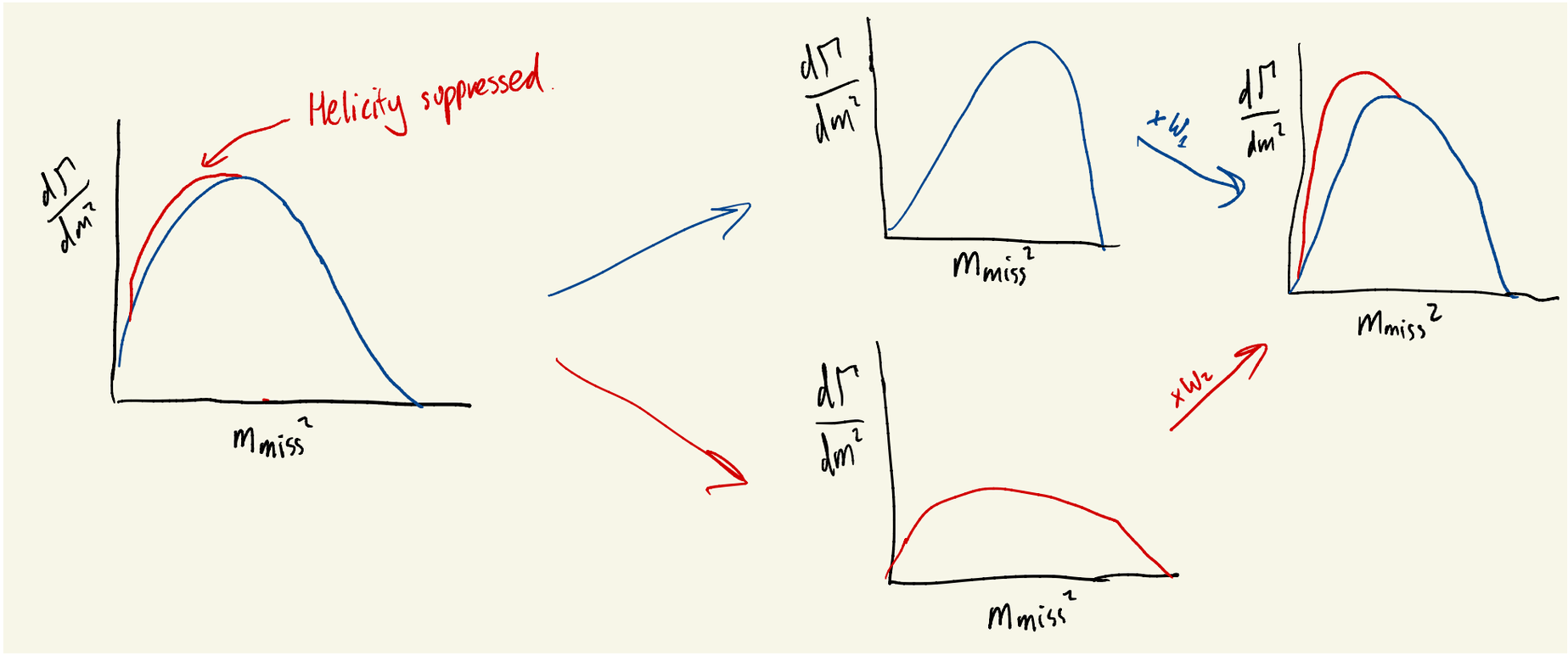
HAMMER

- Helicity Amplitude Module for Matrix Element Reweighting [Eur. Phys. J. C 80, 883 (2020)].

- Decompose amplitude into sub-parts, each with different linear dependence on parameters of interest.

$$\frac{d\Gamma^{\{s\}}}{d\mathcal{PS}} = \sum_{\alpha,i,\beta,j} c_\alpha c_\beta^\dagger \, F_i F_j^\dagger(\{q\}) \, \mathcal{A}_{\alpha i}^{\{s\}} \mathcal{A}_{\beta j}^{\dagger\{s\}}(\{q\})$$
$$= \sum_{\alpha,i,\beta,j} c_\alpha c_\beta^\dagger \, F_i F_j^\dagger(\{q\}) \, \mathcal{W}_{\alpha i \beta j} \, ,$$

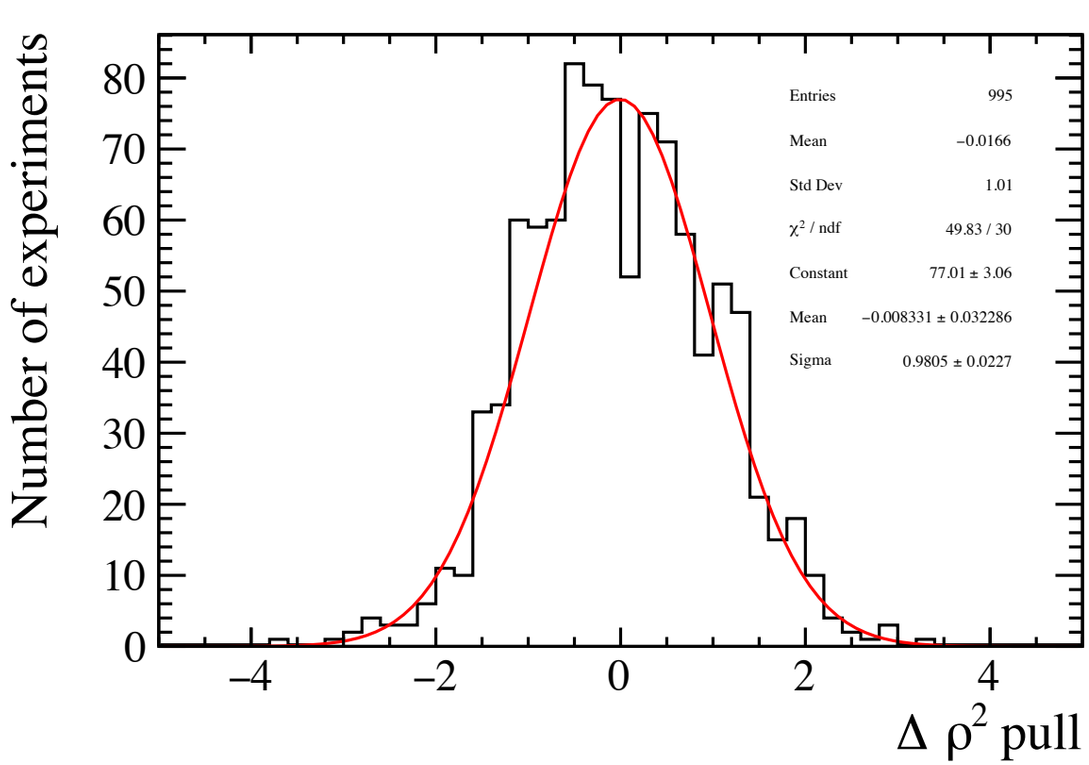
- Oversimplified example: scalar part of a histogram.



- Best place to find documentation is on the webpage: https://hammer.physics.lbl.gov/

HAMMER and RooFit

- HAMMER being used more and more at LHCb.
- Sometimes just for convenience of reweighing things - reliable place to get debugged models from.
- In some analyses, we are directly using the reweighing procedure in the fits.
  - Developed RooHammerModel [2007.12605] to interface with RooFit. Currently being used for the upcoming R(D^+) analysis.

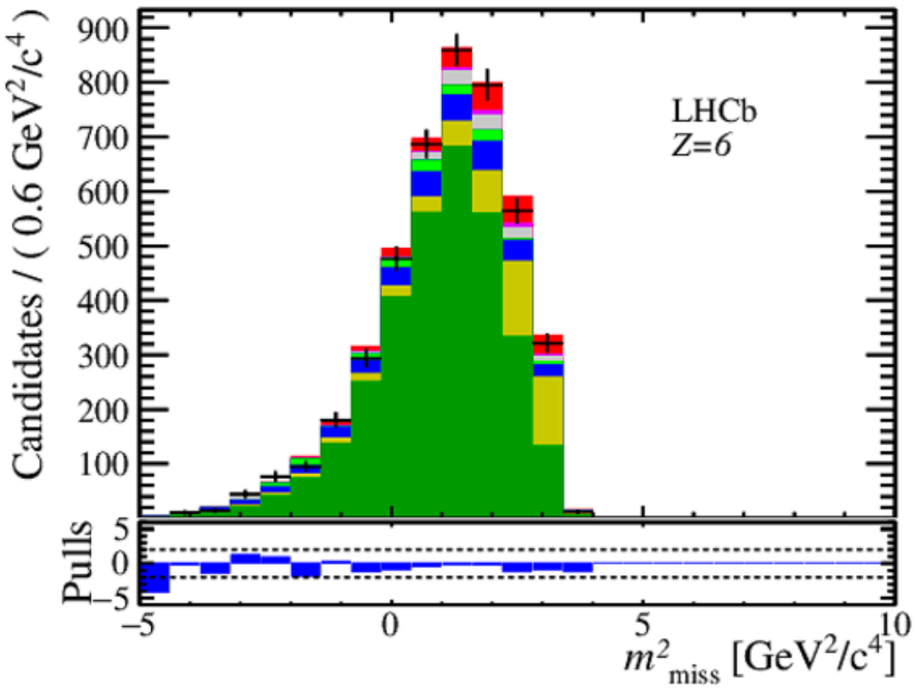


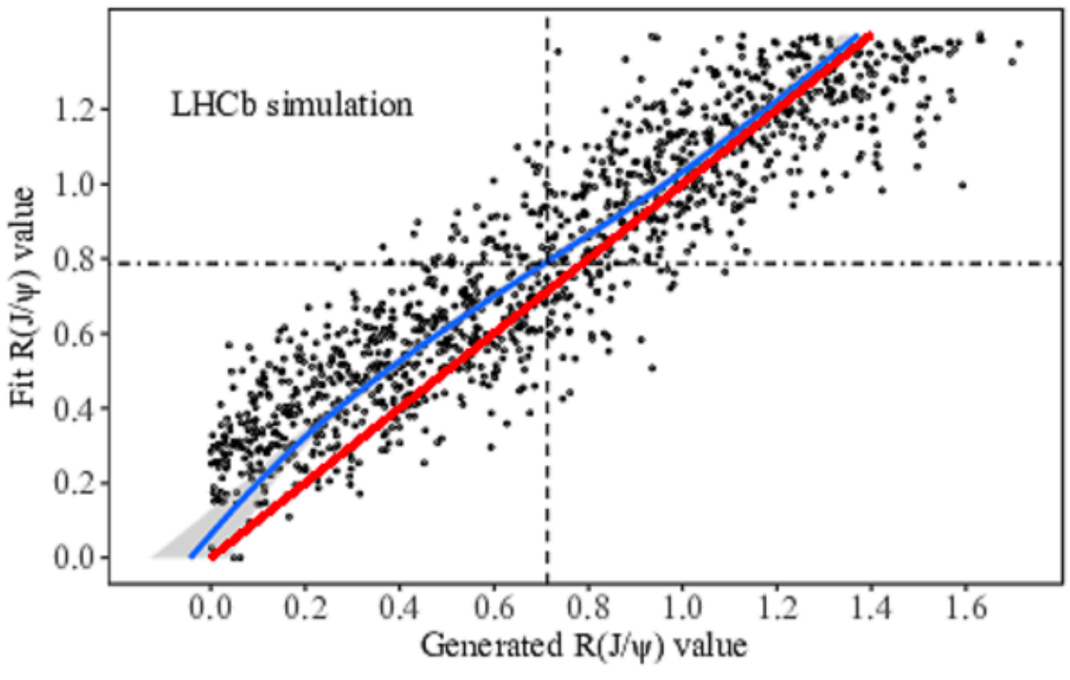
- Overall works well and gives good pulls in toys. Increases likelihood call time by factor ~100, worth digging into once this round is over.

Simulation uncertainties

- Despite our best efforts, simulation size still seems creep up as systematic uncertainty.
  - Lose a large fraction of effective size from the various re-weighting schemes.
- Barlow-Beeston-lite can undercover this uncertainty.
  - Necessary to check with boostraps.
  - Is Barlow-Beeston worth it? Fits faster and more stable without it.
  - As long as have enough simulation to explore disagreement, perhaps not needed.
- Can have issues with empty bins, particularly with low stats fits.

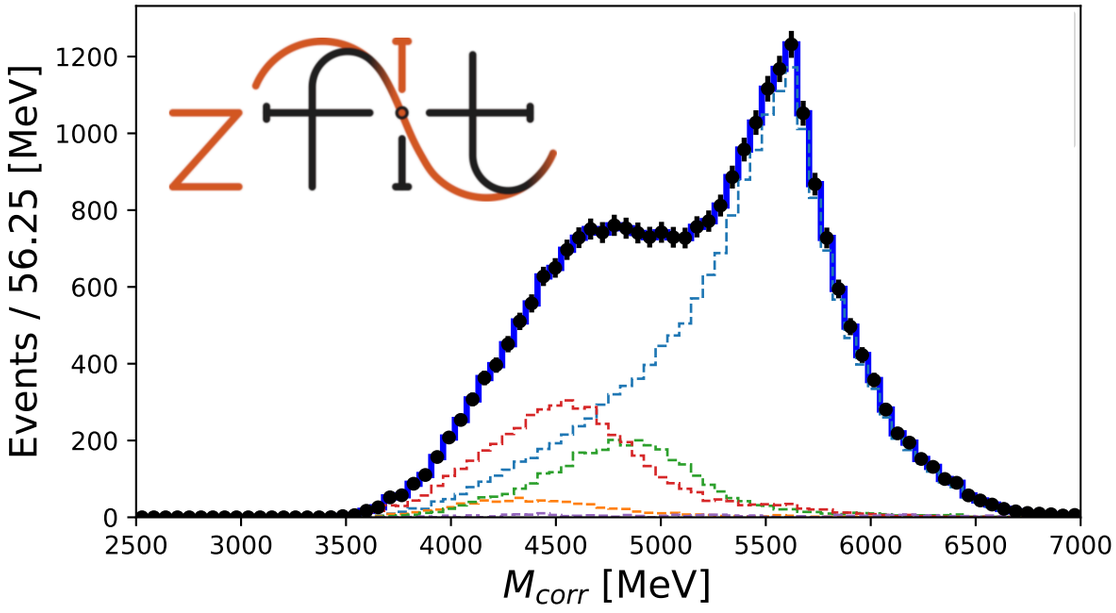
LHCb-PAPER-2017-035



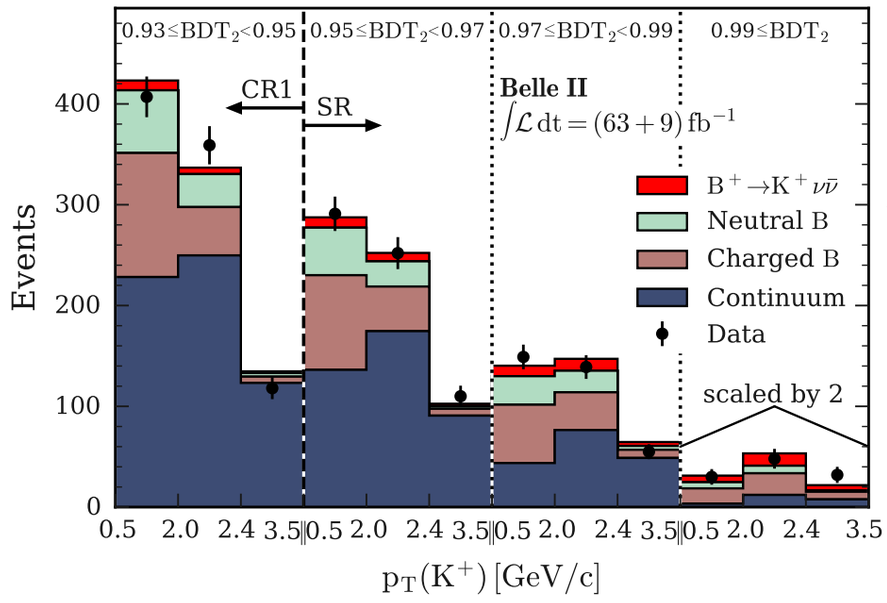


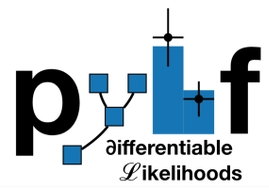
Beyond ROOT

- ROOT starting to compete with large ecosystem of analysis in python.
- Original plan for R(Λ_c) was to implement it in zFit [SoftwareX 11 (2020) 100508].
  - How to efficiently interface HAMMER here would be a big undertaking.
  - Pyhf also an option, as the python replacement for HistFactory [JOSS, 6(58), 2823]
- HistFactory works, do we stick with it, try to improve timing, or move to python.
  - Or both?



Phys. Rev. Lett. 127, 181802 (2021)





Summary Slide:

Title: Fitting Tools for Semileptonics - Key Points

1. Fitting in semileptonics:
   - Multi-dimensional fits
   - Large yields, low purity
   - Simulation size systematics
2. Template fitting with RooFit:
   - Efficient data management
   - 1D projections insufficient
   - Barlow-Beeston or bootstrapping
3. HAMMER & RooFit integration:
   - Increasing use at LHCb
   - RooHammerModel for interfacing
   - Slower likelihood calls
4. Addressing simulation uncertainties:
   - Barlow-Beeston-lite & bootstrap methods
   - Empty bins in low stats fits
5. Future directions:
   - HistFactory vs. Python alternatives
   - Efficient HAMMER integration in Python

P

can you suggest some questions for the speaker

1. How do you foresee improvements in simulation techniques impacting the fitting process for semileptonic decays in the near future?
2. Given the increasing popularity of Python in data analysis, do you think the community should prioritize transitioning from RooFit to Python-based alternatives like zFit or pyhf?
3. Are there any specific limitations of the Barlow-Beeston method that could be addressed to make it more reliable in handling simulation uncertainties?
4. How can researchers optimize the balance between computing resources and the accuracy of the fitting process in semileptonic decays?
5. In the context of HAMMER integration with RooFit, are there any plans to improve the likelihood call time or implement additional optimizations to reduce the computational burden?
6. Can you discuss any recent advances or ongoing research in semileptonic decay fitting that may help address the challenges outlined in your presentation?
7. How do you envision the role of machine learning and artificial intelligence in improving the fitting process for semileptonic decays in the coming years?