The UNIVERSITY of OKLAHOMA

MLHAD

Hadronization Tuning with Reweighting

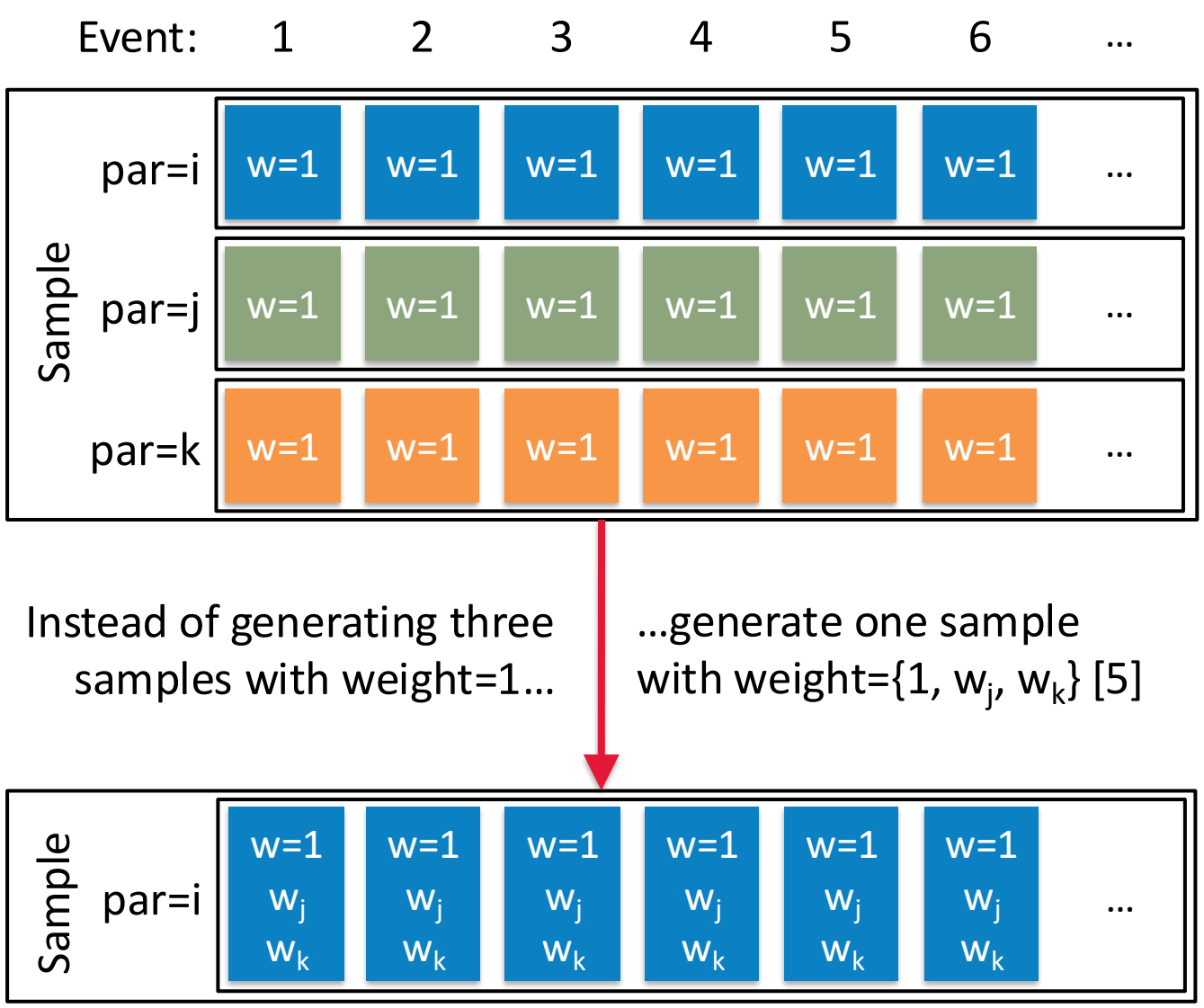
Michael K. Wilkinson

LPCC MCWG Tuning Forum

22 September 2026

Tuning with reweighting

- Tuning methods are about avoiding event generation wherever possible.
  - By hand (Monash tune [1]): Select relatively few parameter values, check agreement. Agreement hard to optimize.
  - Surrogates (Professor [2]): Generate at many parameter values, interpolate with fitted function (polynomial). Requires trusting model.
  - **Reweighting** ("the FRAM method" [3]): Reinterpret a single set of generated events **using the model itself**. Requires support.
- To tune with reweighting, **update event-by-event weights** at each function call to minimize the test-statistic (Migrad [4]) [3].



1. P. Skands and others. *Tuning PYTHIA 8.1: the Monash 2013 Tune*. Eur. Phys. J. C, 74(8):3024, 2014. arXiv:1404.5630.
2. A. Buckley and others. *Systematic event generator tuning for the LHC*. Eur. Phys. J. C, 65:331--357, 2010. arXiv:0907.2973.
3. B. Assi and others. Work in preparation.
4. F. James and M. Roos. *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*. Comput. Phys. Commun., 10:343--367, 1975.
5. M. K. Wilkinson. Simulating Hadronization with Machine Learning. EPJ Web Conf. 295 (2024) 09026. Contribution to CHEP 2023.

Hadronization flavor parameters

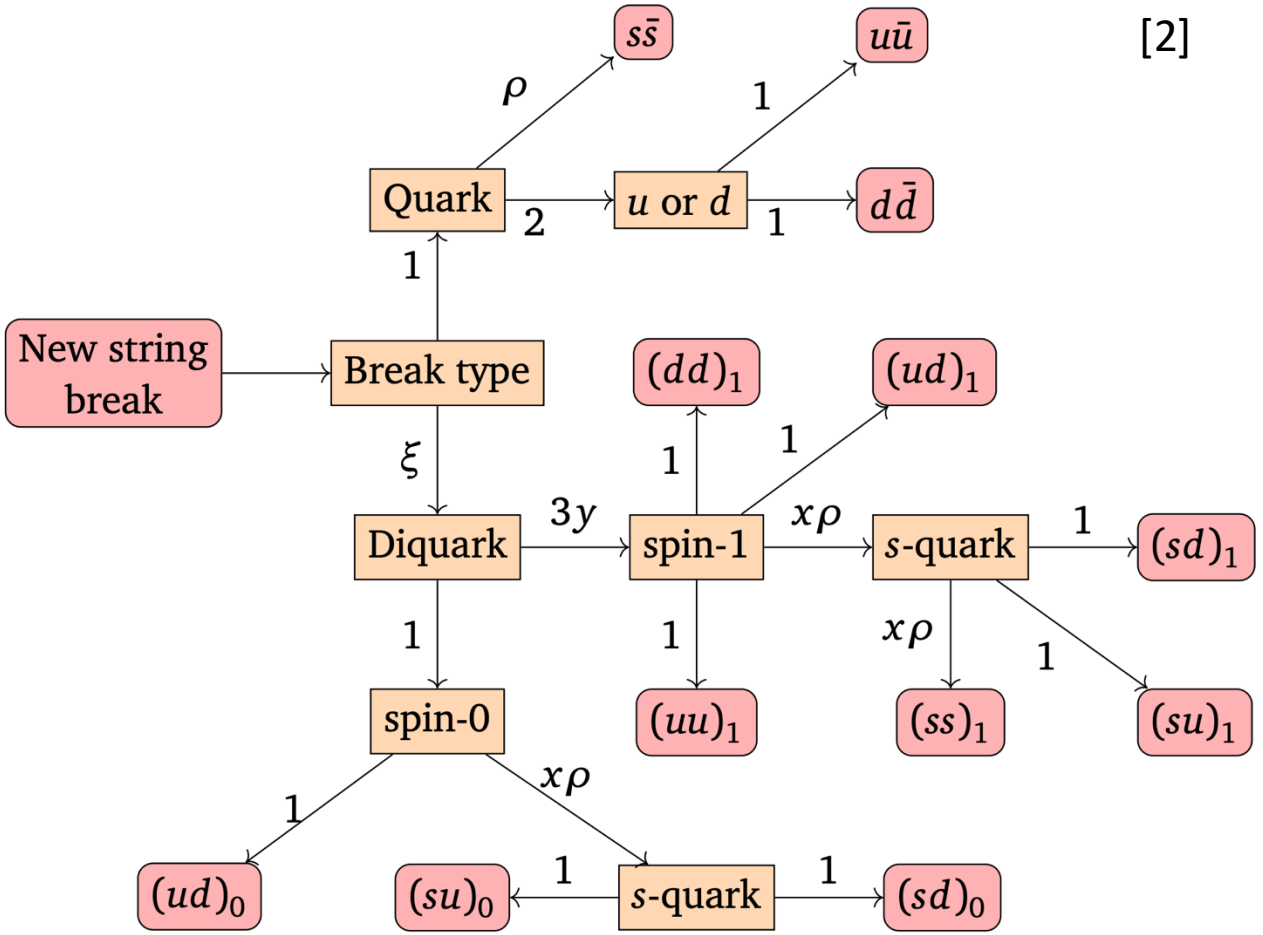
MLHAD

- The Lund string model [1] connects quark $q\bar{q}$ (or diquark $Q\bar{Q}$) pairs with a **string of color potential**.
- As the string ends move apart, the string breaks into pieces, a.k.a. hadrons.
- The relative rates of hadron production are determined by the (di)quark **flavors assigned to the string ends**,
- governed by **four parameters** that **suppress various rates** (valued 0–1) [2]:

  $\boldsymbol{\xi}$: $Q\bar{Q}$ vs. $q\bar{q}$ $\quad$ $\boldsymbol{\rho}$: $s\bar{s}$ vs. $u\bar{u}$ or $d\bar{d}$

  $\boldsymbol{x}$: $Q_s\bar{Q}_s$ vs. $Q\bar{Q}$ $\quad$ $\boldsymbol{y}$: $\{Q\bar{Q}\}_{\text{spin1}}$ vs. $\{Q\bar{Q}\}_{\text{spin0}}$

[2]



1. B. Andersson and others. *Parton Fragmentation and String Dynamics*. Phys. Rept., 97:31--145, 1983.

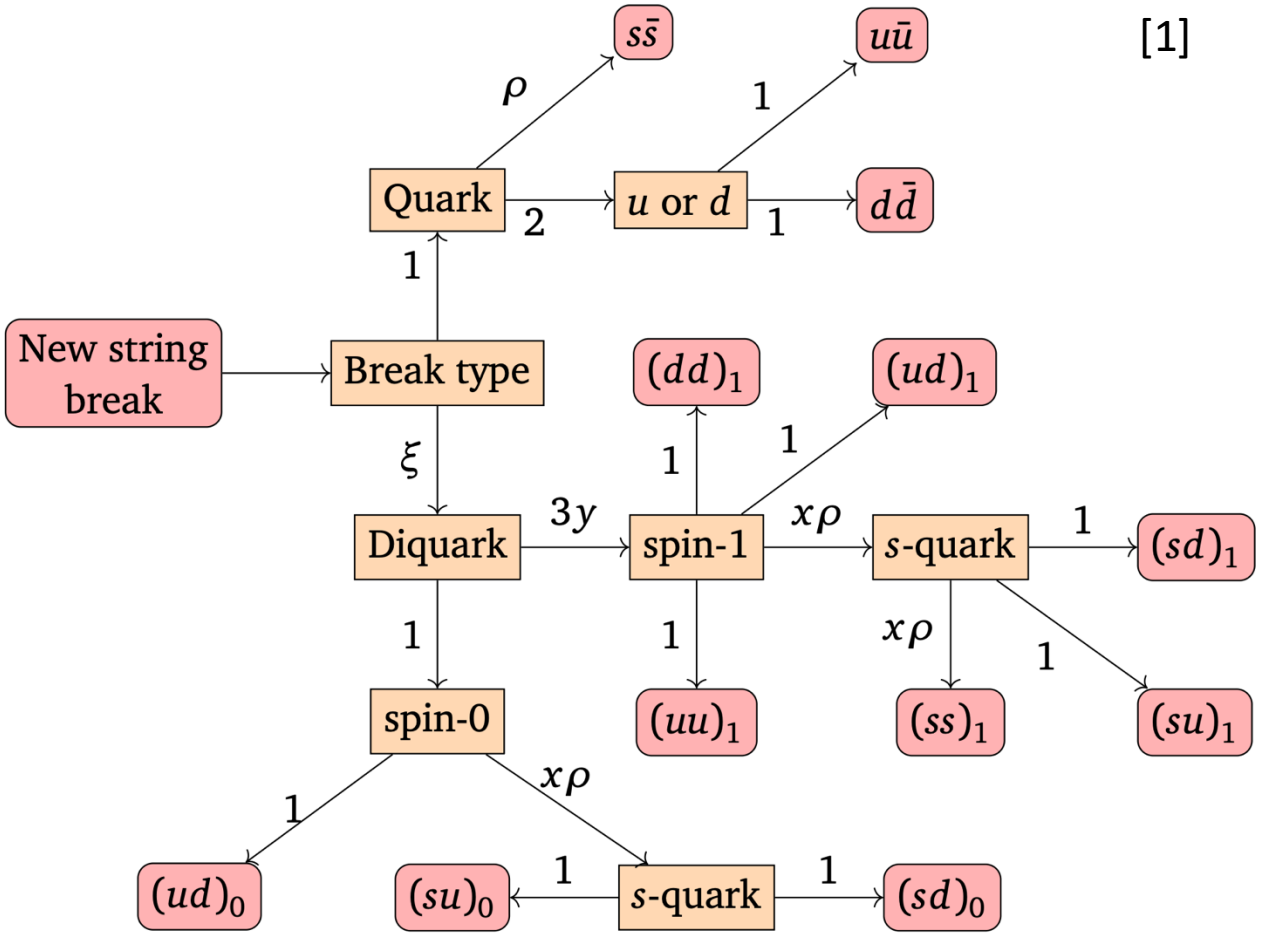
2. B. Assi and others. *Post-hoc reweighting of hadron production in the Lund string model*. SciPost Phys., 19(4):104, 2025. arXiv:2505.00142.

Post-hoc hadronization flavor

MLHAD

- **Four flavor parameters** (valued 0–1) [1]:
  ξ: $Q\bar{Q}$ vs. $q\bar{q}$ $\quad$ ρ: $s\bar{s}$ vs. $u\bar{u}$ or $d\bar{d}$
  x: $Q_s\bar{Q}_s$ vs. $Q\bar{Q}$ $\quad$ y: $\{Q\bar{Q}\}_{\mathrm{spin1}}$ vs. $\{Q\bar{Q}\}_{\mathrm{spin0}}$
- Post-hoc **reweighting implemented for flavor parameters** using a detailed analytic algorithm; about **1,500x faster** than event generation and only requires storing 20 extra integers per event [1].
- (In-situ reweighting implemented for kinematic [energy and momentum] parameters using modified accept/reject algorithm [2]. Post-hoc available as well, validation in progress [3].)

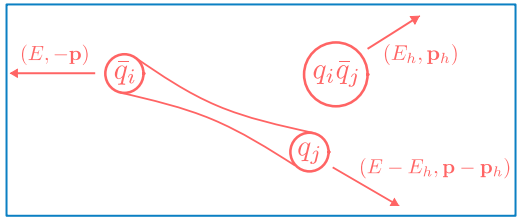
[1]

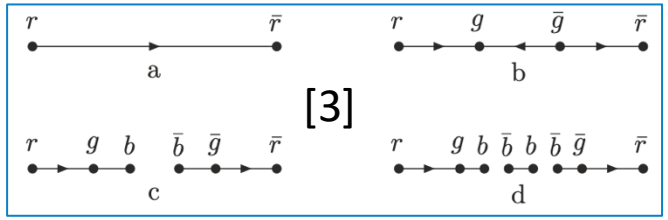


1. B. Assi and others. *Post-hoc reweighting of hadron production in the Lund string model*. SciPost Phys., 19(4):104, 2025. arXiv:2505.00142.
2. C. Bierlich and others. *Reweighting Monte Carlo predictions and automated fragmentation variations in Pythia 8*. SciPost Phys., 16(5):134, 2024. arXiv:2308.13459.
3. B. Assi and others. Work in preparation.

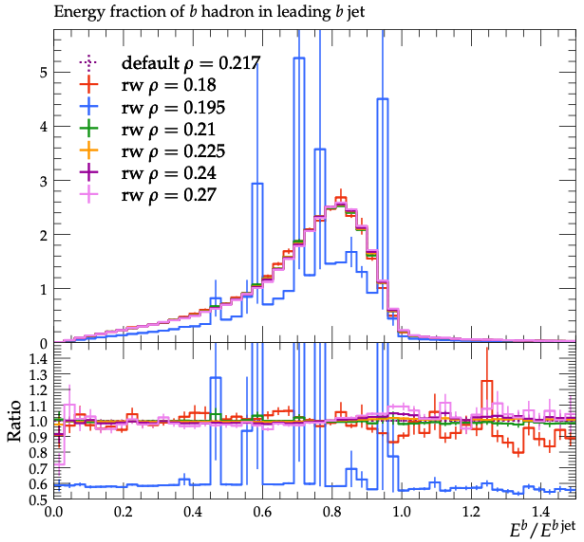
Aside: Popcorn reweighting

- The **popcorn model** [1] is an additional mechanism for baryon production besides string-end diquarks [2, 3].
- This **changes the probabilities** of different string breaks, breaking the reweighting calculations for ξ, ρ, x, y [4].
- ATLAS simulation, e.g., has popcorn enabled by default, making reweighting unusable without turning it off [5].
- Popcorn reweighting **in development** [6]

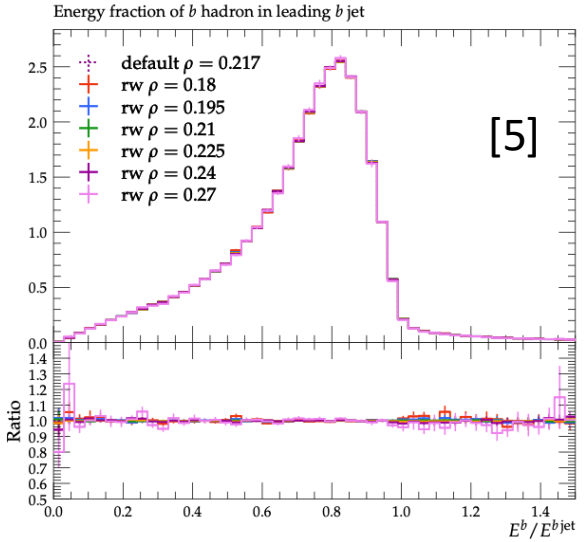




[3]



(a) Default popcorn rate

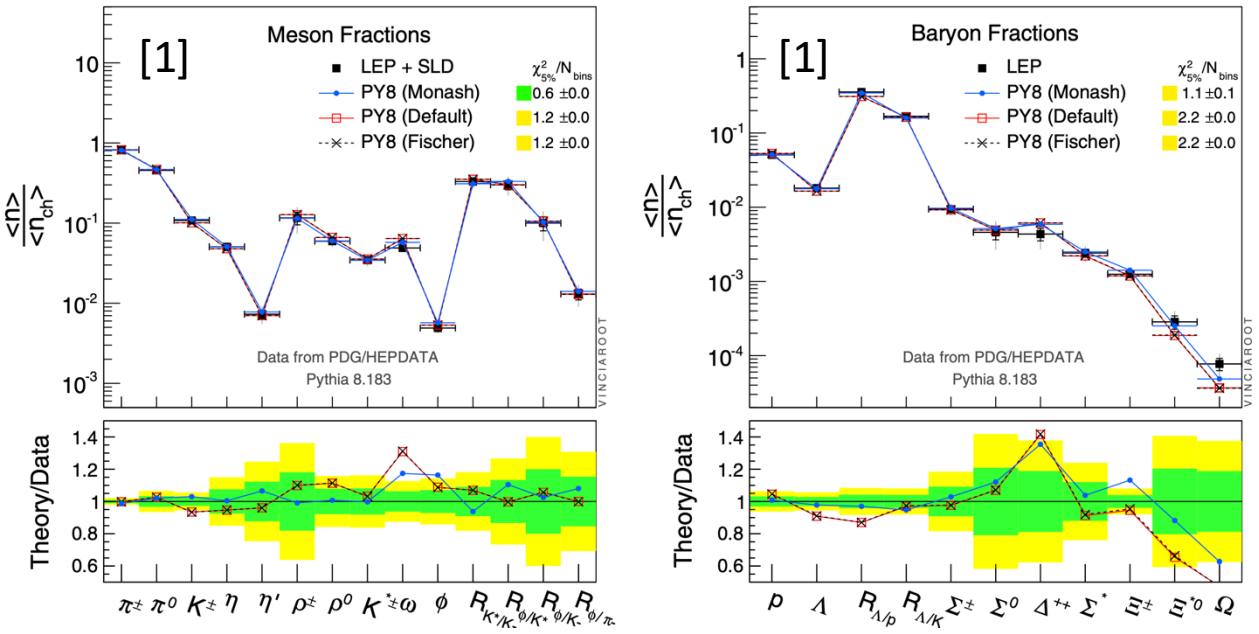


[5]

(b) Popcorn disabled

1. B. Andersson and others. *Baryon Production in Jet Fragmentation and Decay*. Phys. Scripta, 32:574, 1985.
2. C. Bierlich and others. *A comprehensive guide to the physics and usage of PYTHIA 8.3*. SciPost Phys. Codeb., 2022:8, 2022. arXiv:2203.11601.
3. C. Bierlich, G. Gustafson, L. Lönnblad, and A. Tarasov. *Effects of Overlapping Strings in pp Collisions*. JHEP, 03:148, 2015. arXiv:1412.6259.
4. B. Assi and others. *Post-hoc reweighting of hadron production in the Lund string model*. SciPost Phys., 19(4):104, 2025. arXiv:2505.00142.
5. A. Casali. *Improving hadron level modelling and uncertainty estimation with reweighting* . 2026. ATL-COM-PHYS-2026-163.
6. B. Assi and others. Work in preparation.

The Monash tune



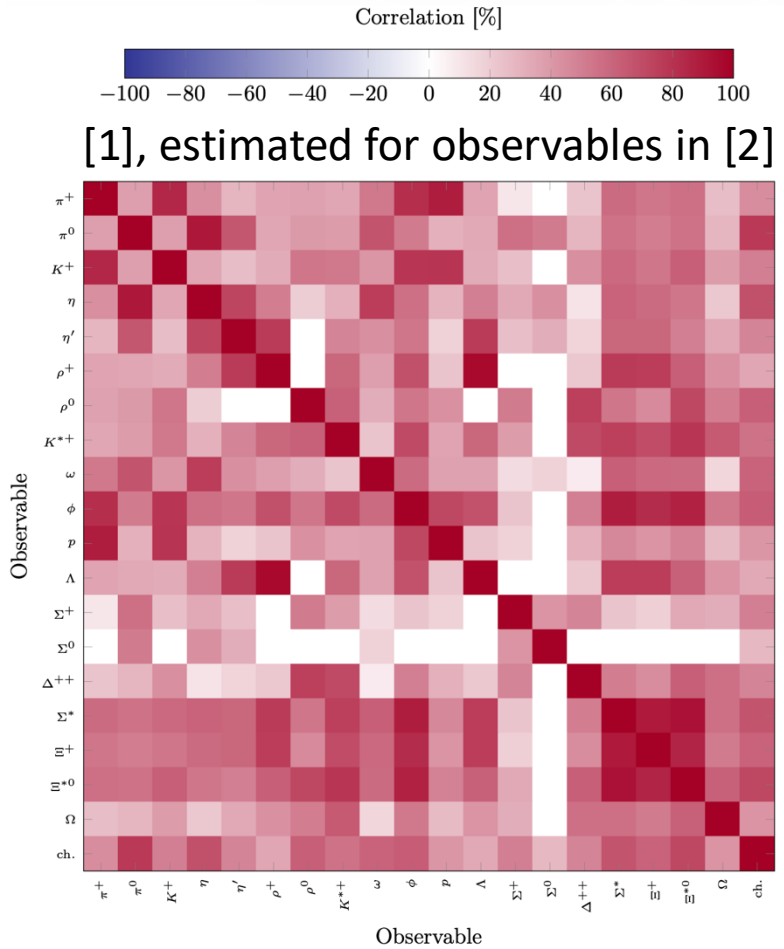
$$\xi = 0.085, \quad \rho = 0.217,$$
$$x = 0.915, \quad y = 0.0275$$

- The Monash Tune [1] (primarily) **hand tuned** to relative hadron production rates in (LEP) e^+e^- collisions at the Z pole, e.g., $\langle n_{\pi^\pm}\rangle/\langle n_{ch.}\rangle$, where the denominator is treated as exact.
- Characterized by (not optimized) $\langle \chi^2_{5\%} \rangle \equiv \left(\sum_{i=1}^{N_{\text{bins}}} \frac{(\text{MC}_i - \text{Data}_i)^2}{\sigma^2_{\text{Data},i} + (0.05\text{MC}_i)^2}\right)/N_{\text{bins}}$, **treating observables as independent**.
- The goal: improved tune
- Not the goal: **best fit, uncertainties**

1. P. Skands and others. *Tuning PYTHIA 8.1: the Monash 2013 Tune*. Eur. Phys. J. C, 74(8):3024, 2014. arXiv:1404.5630.

Aside: Unreported correlations

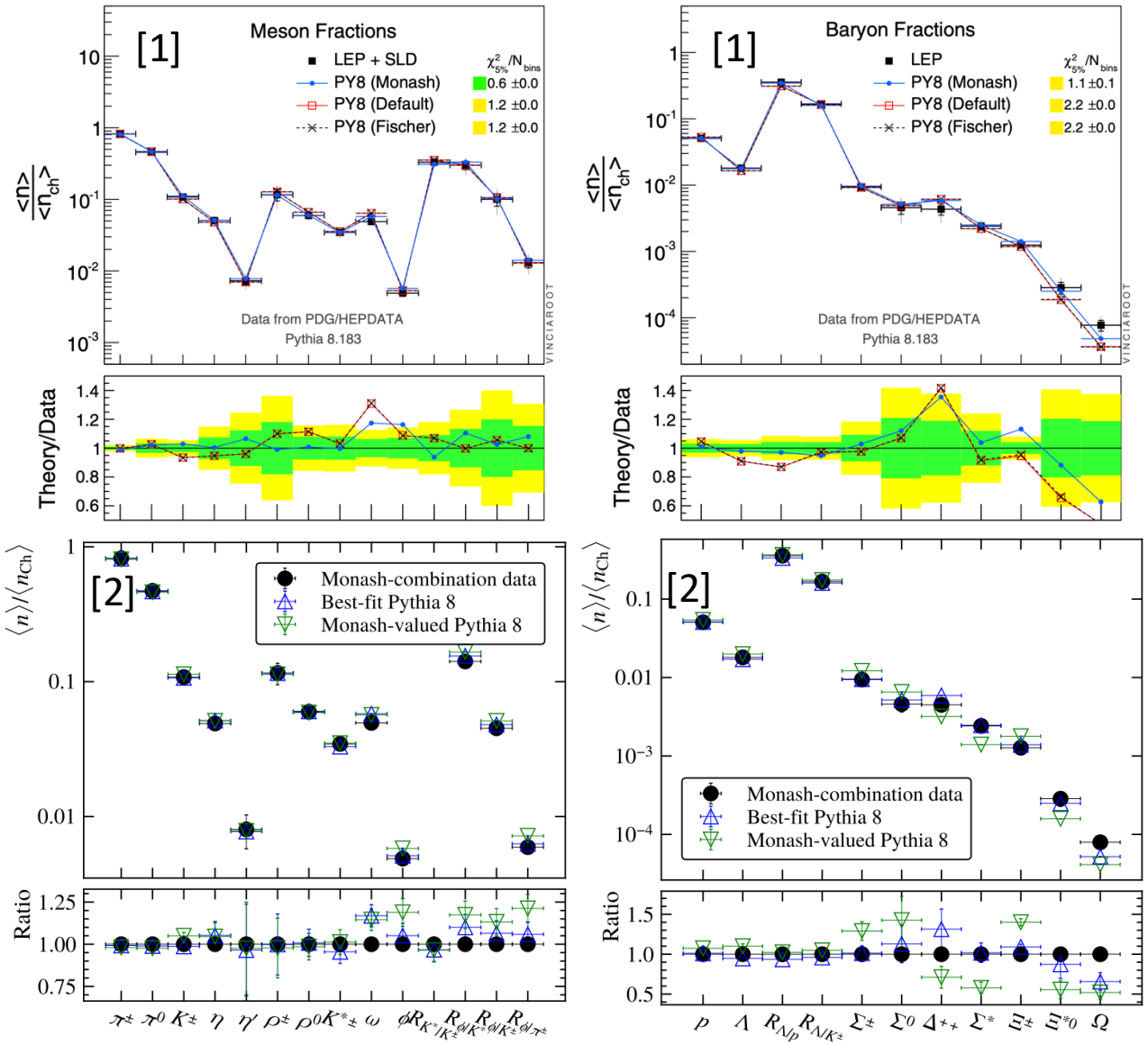
- The results used in the Monash tune do not report correlations. Conservatively [1]:
  - Assume 100% correlated systematic uncertainties within an experiment, independent across experiments.
  - Assume statistical uncertainties share the same correlation structure as raw Pythia.
- This level of correlations would significantly alter a tune [1].
  - Fitting to artificial data samples generated with this correlation matrix yielded **biased confidence intervals** when the test-statistic did not account for correlations [1].
  - Takeaway: In a future tune, **should not ignore correlations**. Choose data with a correlation structure you understand. More work needed to understand this, so
  - the following **neglects correlations** for now.



[1], estimated for observables in [2]

1. B. Assi and others. Work in preparation.
2. P. Skands, S. Carrazza, and J. Rojo. *Tuning PYTHIA 8.1: the Monash 2013 Tune*. Eur. Phys. J. C, 74(8):3024, 2014. arXiv:1404.5630.

FRAM fit to real data



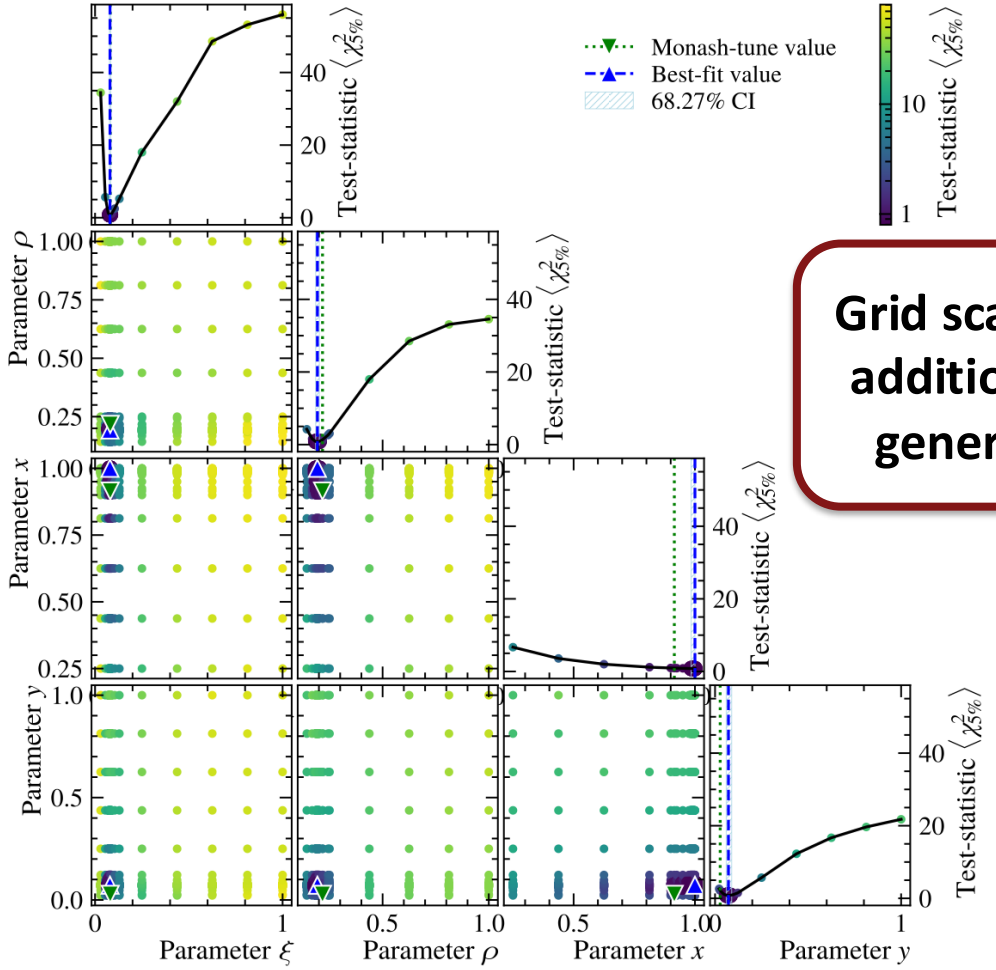
- We can fit 50,000 Pythia events using reweighting ("FRAM method") to get **optimized** $\langle \chi^2_{5\%} \rangle$ using Migrad [3].

	ξ	ρ	x	y	$\langle \chi^2_{5\%} \rangle$
Monash (popcorn) [1]	0.085	0.217	0.915	0.0275	0.82
Monash (no popcorn) [2]	0.085	0.217	0.915	0.0275	3.58
Us (no popcorn) [2]	0.080	0.193	1.000	0.0720	0.80

1. P. Skands and others. *Tuning PYTHIA 8.1: the Monash 2013 Tune*. Eur. Phys. J. C, 74(8):3024, 2014. arXiv:1404.5630.
2. B. Assi and others. Work in preparation.
3. F. James and M. Roos. *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*. Comput. Phys. Commun., 10:343--367, 1975.

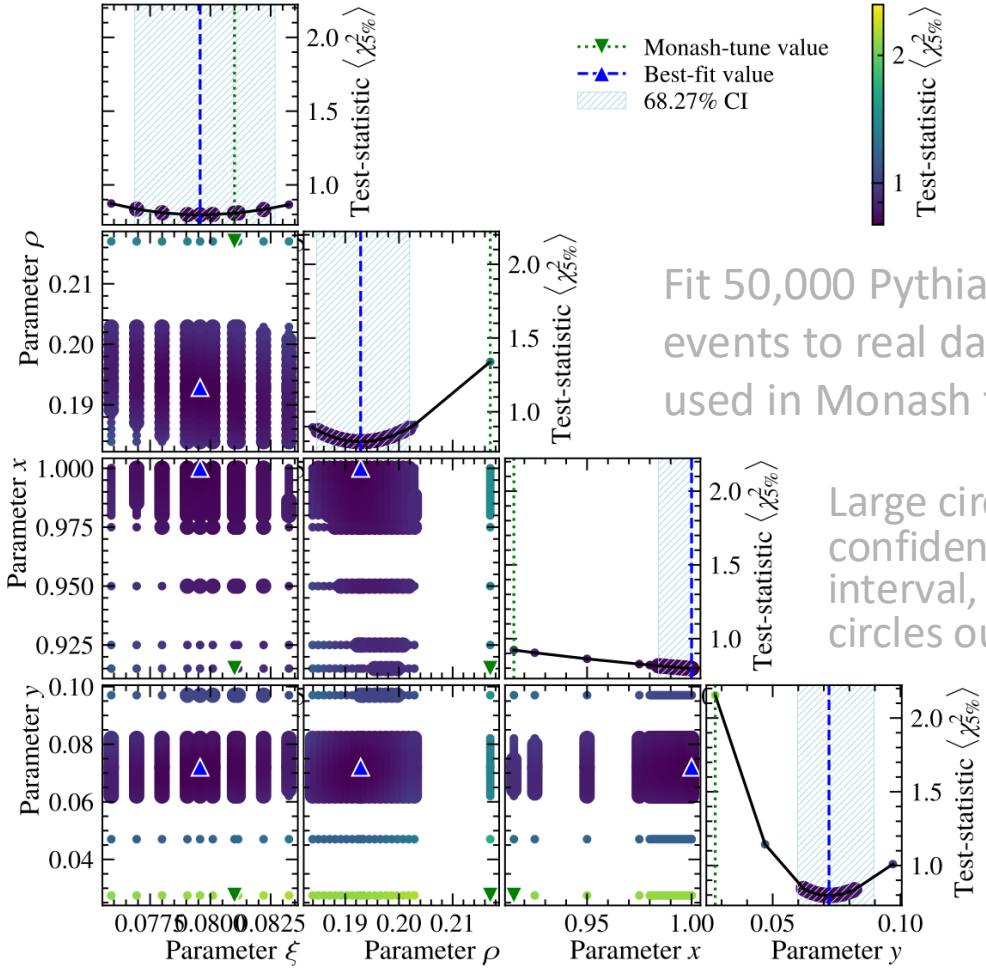
FRAM profile test-statistic map

Full parameter map



Grid scan without additional event generation [1]

Parameter map near the minimum



Fit 50,000 Pythia events to real data used in Monash tune

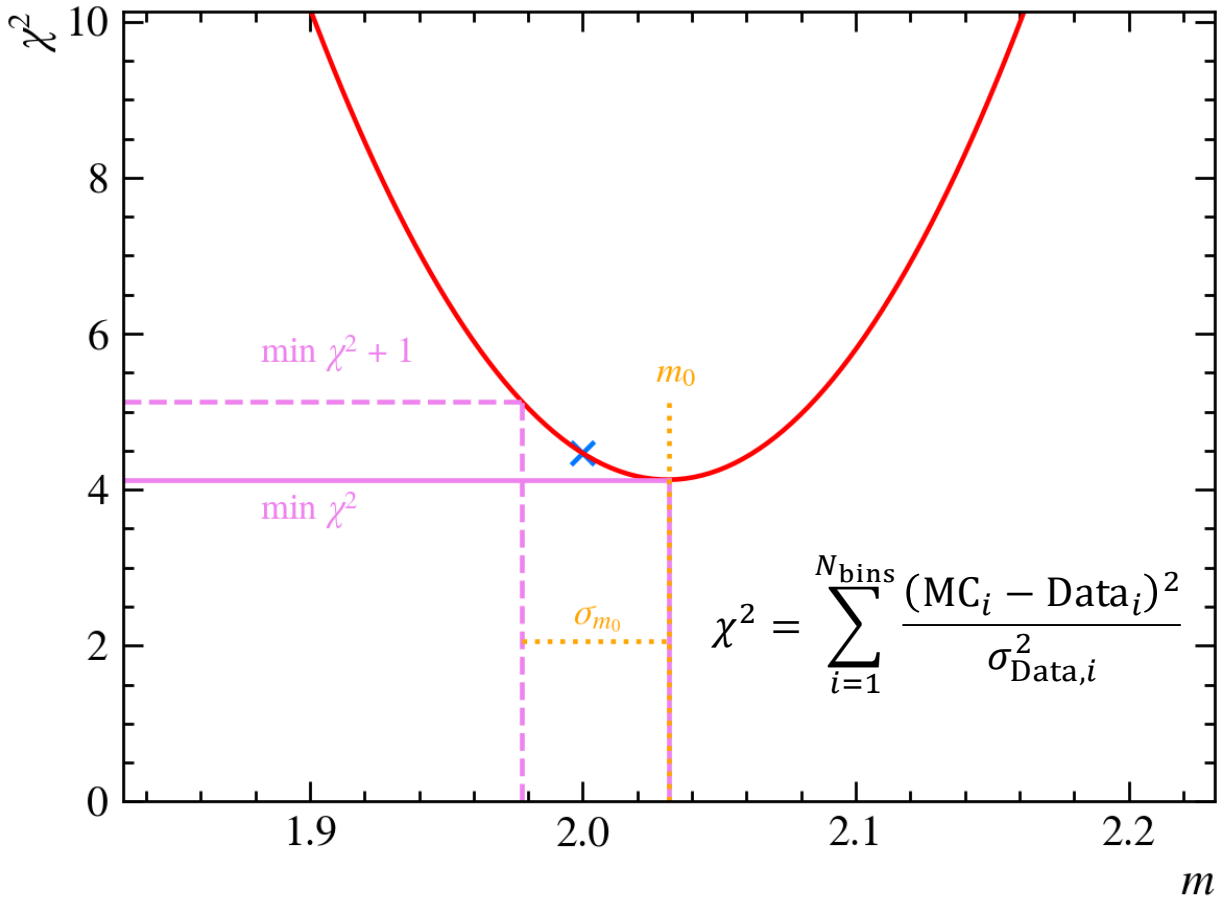
Large circles in confidence interval, small circles outside

1. B. Assi and others. Work in preparation.

The Hessian method (not used here)

- In a standard fit to one parameter, χ^2 parabolic and $\min \chi^2 + 1$ gives the uncertainty (e.g., [1])
- Only works for Gaussian-distributed likelihoods thanks to Wilks' theorem [2]
  - Many observables have Gaussian-distributed likelihoods near the minimum, and Poisson observables, such as particle counts, approximate Gaussian at sufficient statistics.
  - **Not guaranteed to hold** for $\langle \chi^2_{5\%} \rangle$ or for hadronization likelihoods in general.
- We want a more generic approach.

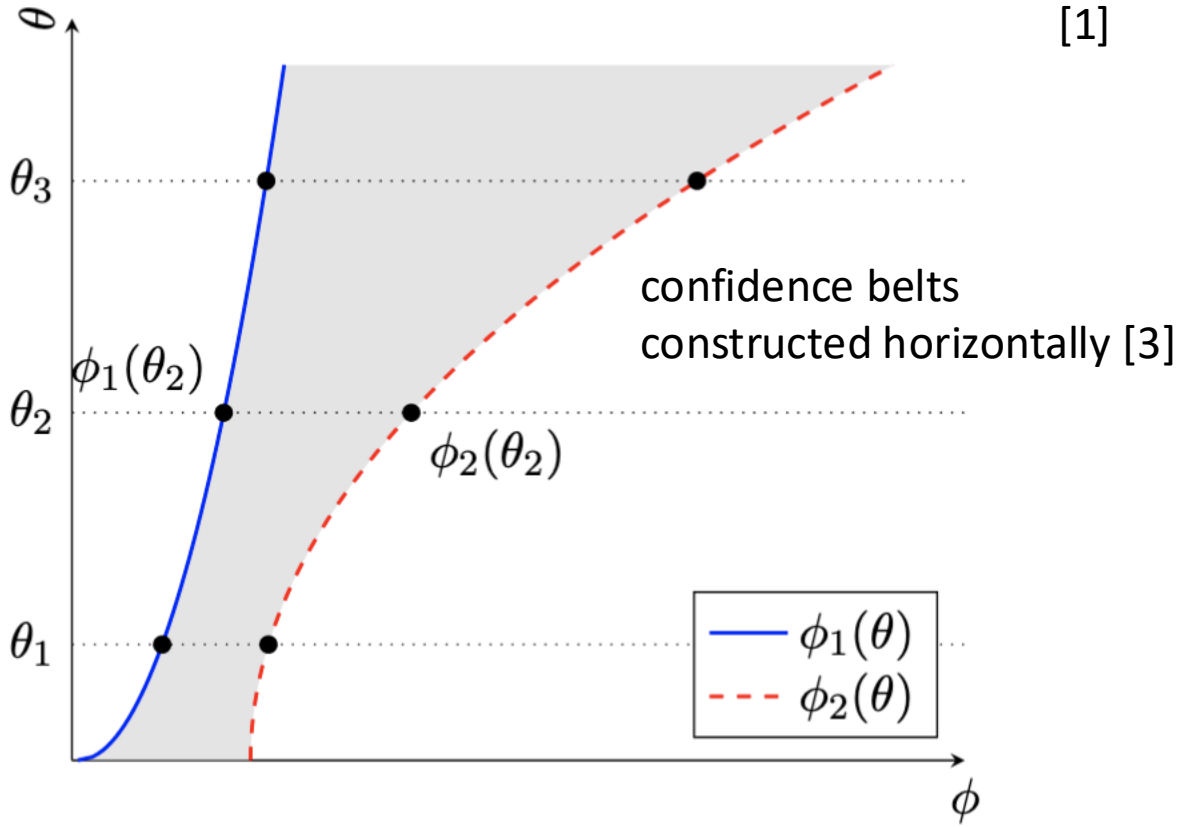
https://goi42.github.io/2024-06-24-UCQuarkNet/_episodes/iminuit



1. P. Risse, N. Derakshanian, T. Jezo, K. Kovarik, and A. Kusina. *Determination of proton PDF uncertainties with Markov Chain Monte Carlo*. 2025. arXiv:2510.16158.
2. S. S. Wilks. *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*. Annals Math. Statist., 9(1):60--62, 1938.

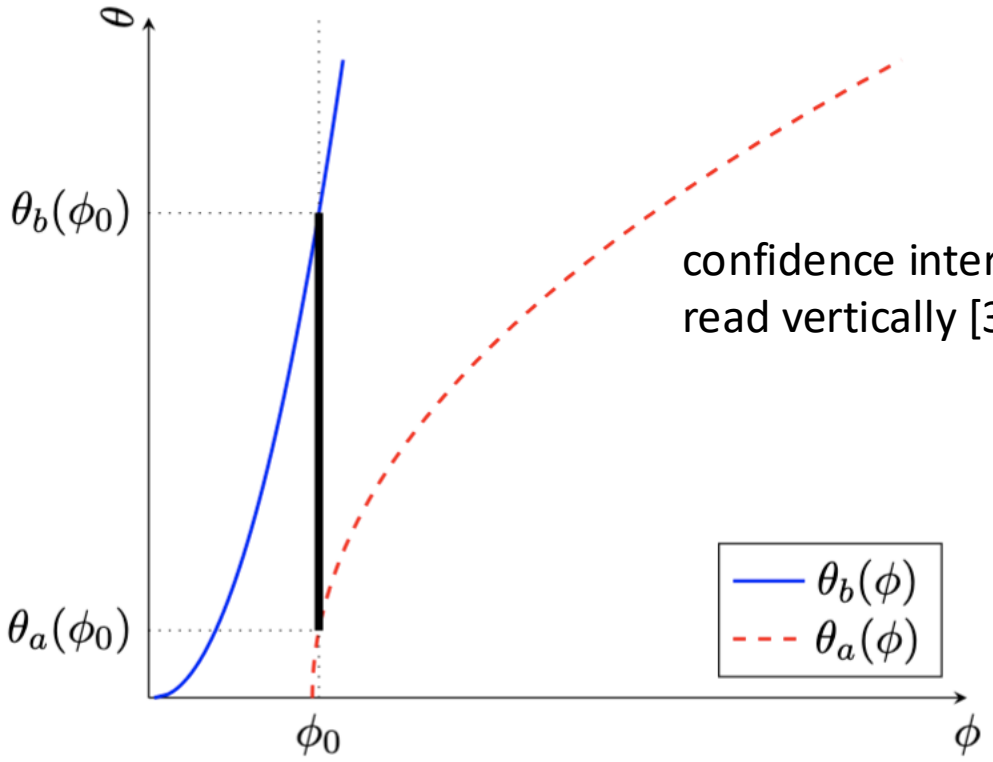
Frequentist uncertainties

[1]



confidence belts constructed horizontally [3]

[1]



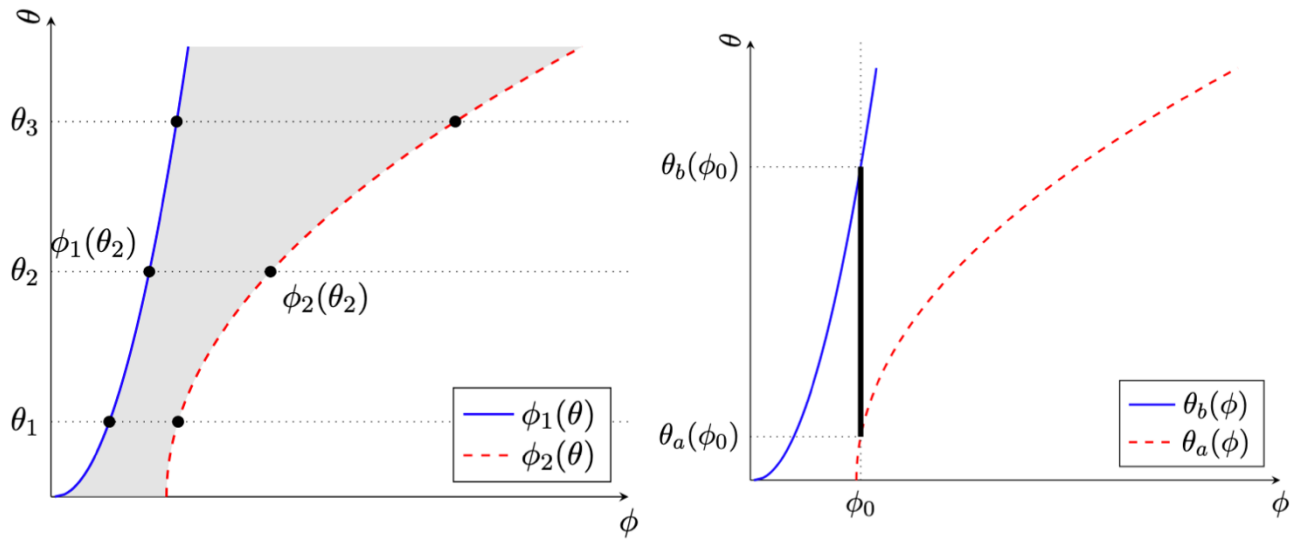
confidence interval read vertically [3]

If the same experiment were repeated the same way many times, e.g., 68% of them should contain the true parameter value θ within a confidence interval $[\theta_a, \theta_b]$ based on observation ϕ_0 [2, 3]. Expression of **uncertainty on the method**.

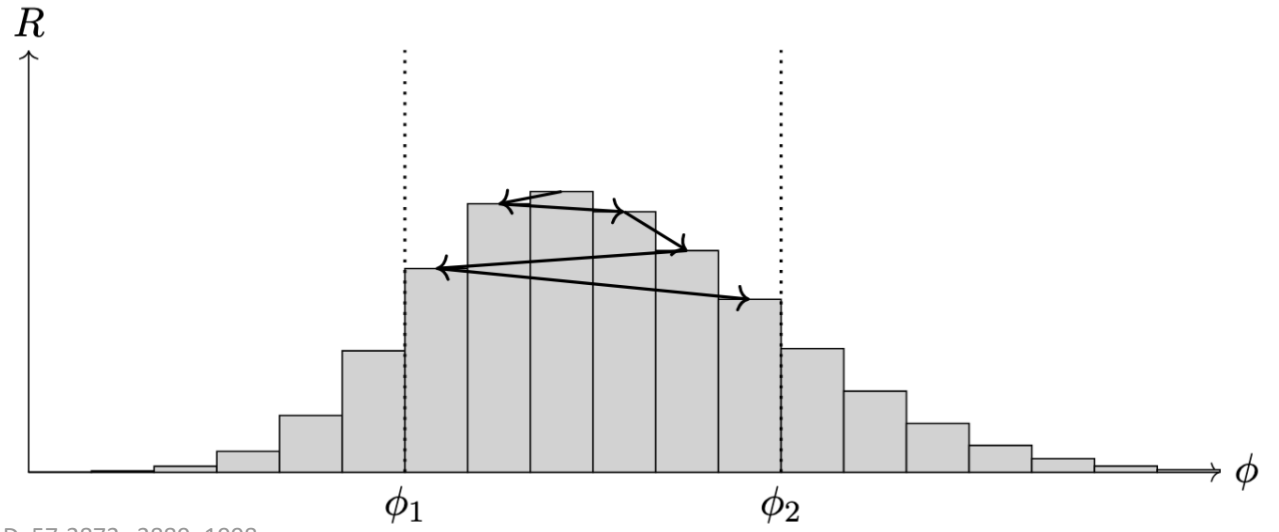
1. B. Assi and others. Work in preparation.
2. J. Neyman. *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*. Phil. Trans. Roy. Soc. Lond. A, 236(767):333--380, 1937.
3. F. E. James. *Statistical Methods in Experimental Physics*. World Scientific, Singapore, 2nd, 2006.

The Feldman-Cousins Unified Approach

- To draw the boundaries of the confidence belt,
  1. generate many "pseudo-experiments" at parameter value θ; for each,
  2. fit the model to find $\chi^2(\hat{\theta}|\phi)$ and
  3. calculate $-\ln R = \chi^2(\theta|\phi) - \chi^2(\hat{\theta}|\phi)$;
  4. add results to the confidence belt from highest to lowest R until 68% are included and determine the cutoff $-\ln R \equiv \Delta\chi^2$ [1].
- If the observed $\Delta\chi^2(\theta|\phi_0)$ is within the cutoff, θ is in the confidence interval for observation ϕ_0 [1].
- **"For true parameter value θ, how far off is the best-fit χ^2 likely to be?"**



[2]

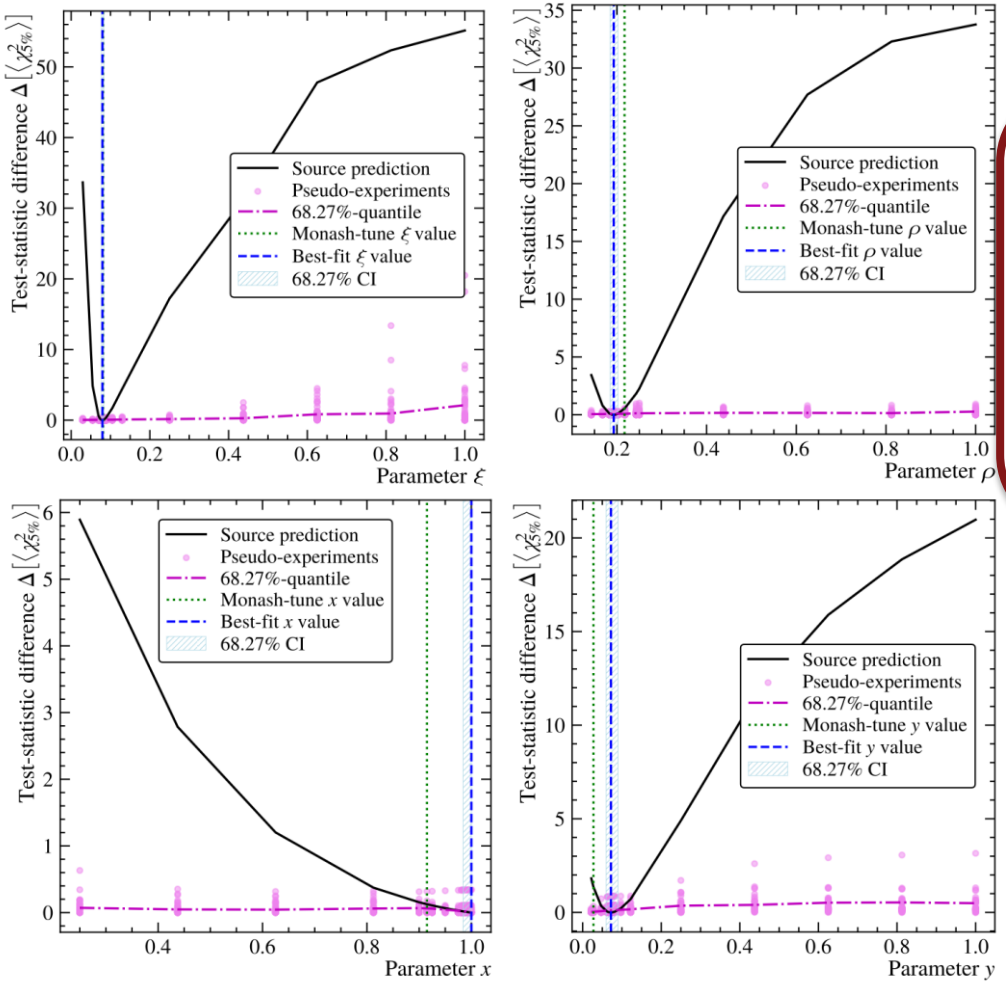


1. G. J. Feldman and R. D. Cousins. *A Unified approach to the classical statistical analysis of small signals*. Phys. Rev. D, 57:3873--3889, 1998.
2. B. Assi and others. Work in preparation.

FRAM confidence intervals

Full parameter scan

[1]



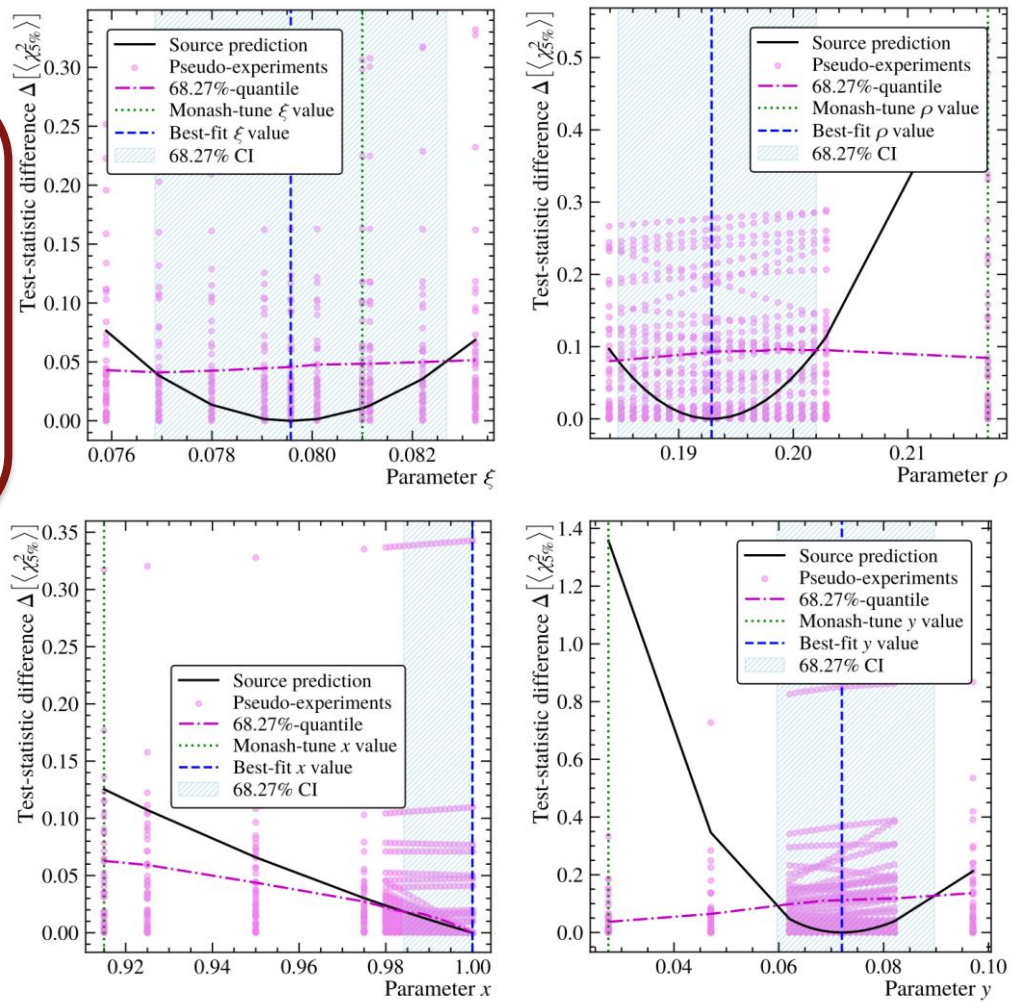
Hadronization tune with frequentist confidence intervals!

…albeit with the popcorn model disabled (for now)

Works even in non-Gaussian cases (e.g., x)!

Parameter scan near the minimum

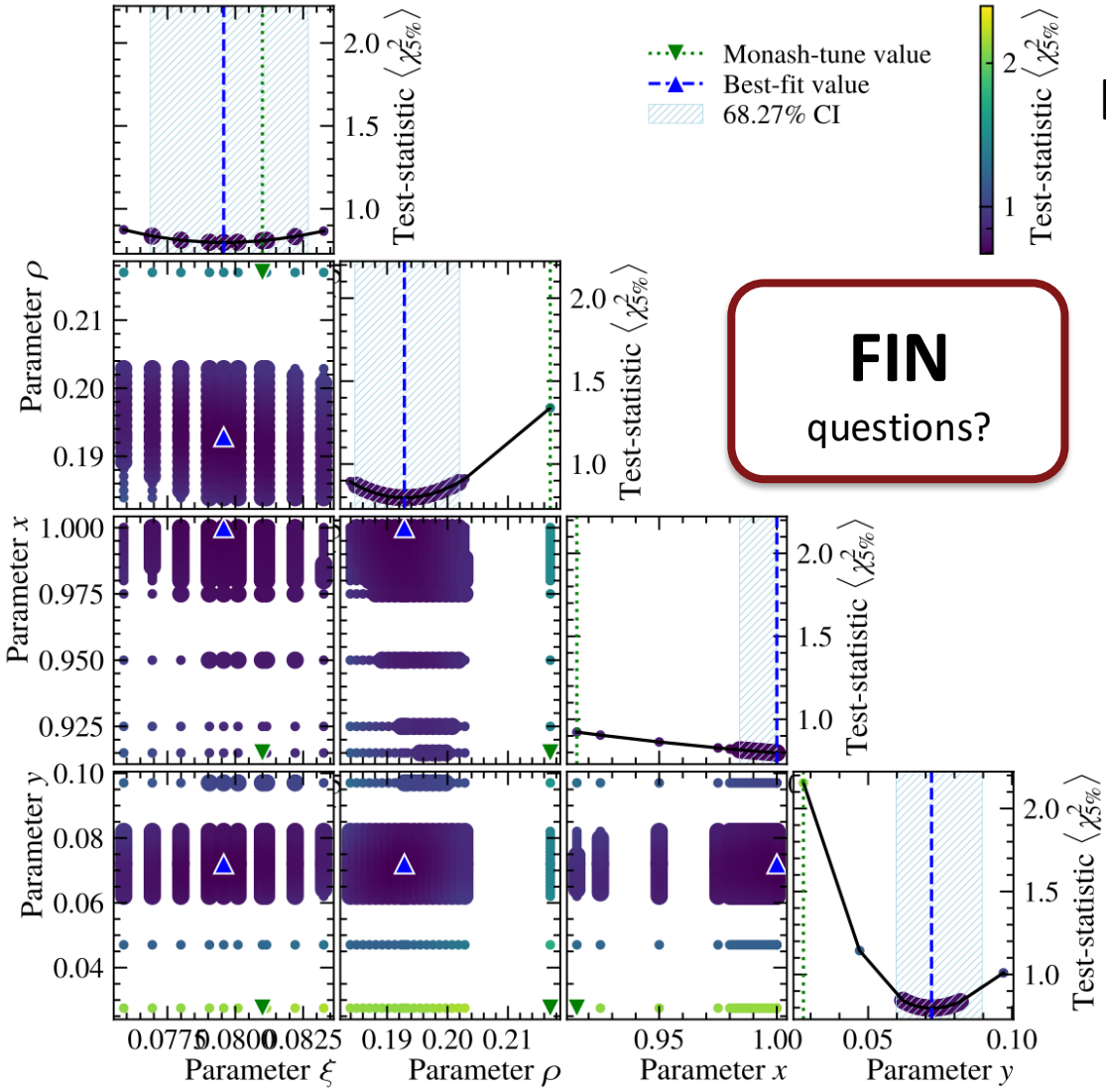
[1]



1. B. Assi and others. Work in preparation.

Conclusion

- The FRAM tuning method provides [1]
  - optimized parameter values
  - and **robust frequentist uncertainties**
  - **without a surrogate** or regeneration
  - for any reweightable model—**no Gaussian likelihood assumption required**.
- We have tuned, as a demonstration, [1]
  - four flavor-selection parameters in the Lund string model used in Pythia…
  - without popcorn (in development [2]).
- Correlations matter
  - for hadronization in a conservative limit [1]
  - and **should not be neglected**.



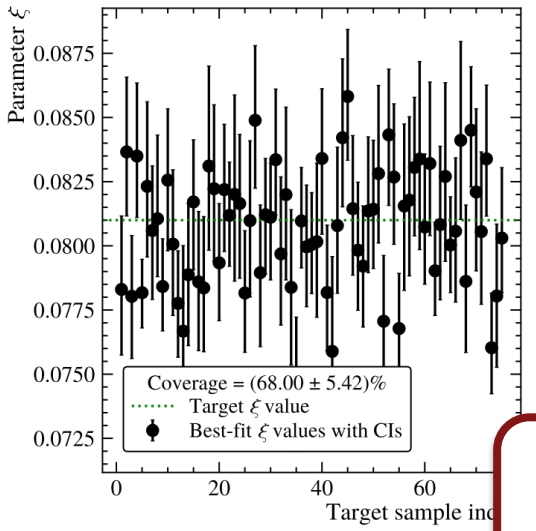
[1]

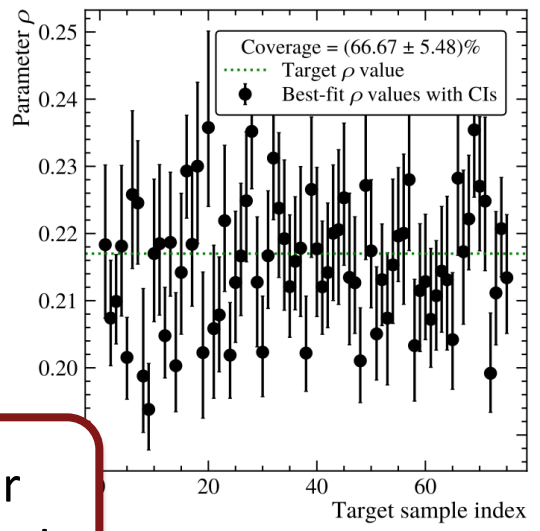
1. B. Assi and others. Work in preparation.
2. B. Assi and others. Work in preparation.

BACKUP

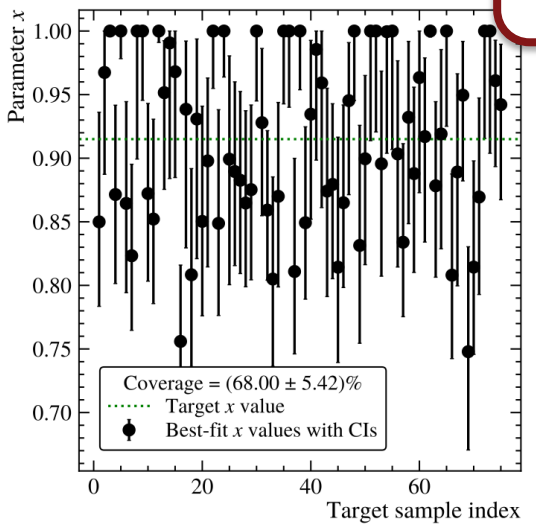
Closure test for hadronization

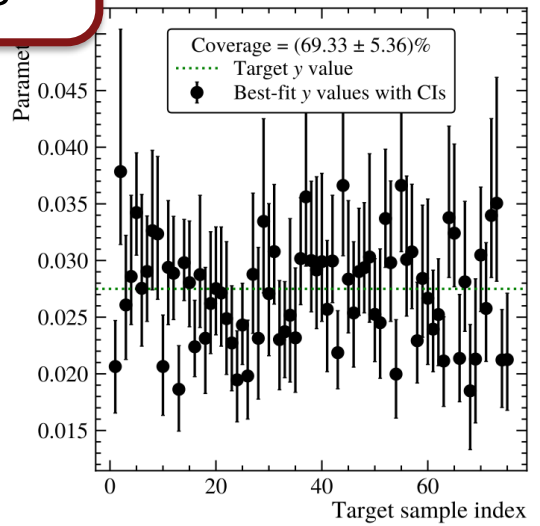
- The true parameter values ought to lie within uncertainties 68% of the time (“**proper coverage**”).
- Generate 75 **artificial data** samples
  - by Gaussian **smearing the mean particle counts** in a Pythia sample of 10M events using the Monash data uncertainties,
  - to simulate independent fitting targets,
- and 50x75 **pseudo-experiments**,
  - each comprising 20 independent Pythia samples, **one for each particle count**, of appropriate numbers of events,
  - to calculate the confidence intervals.





Proper coverage!

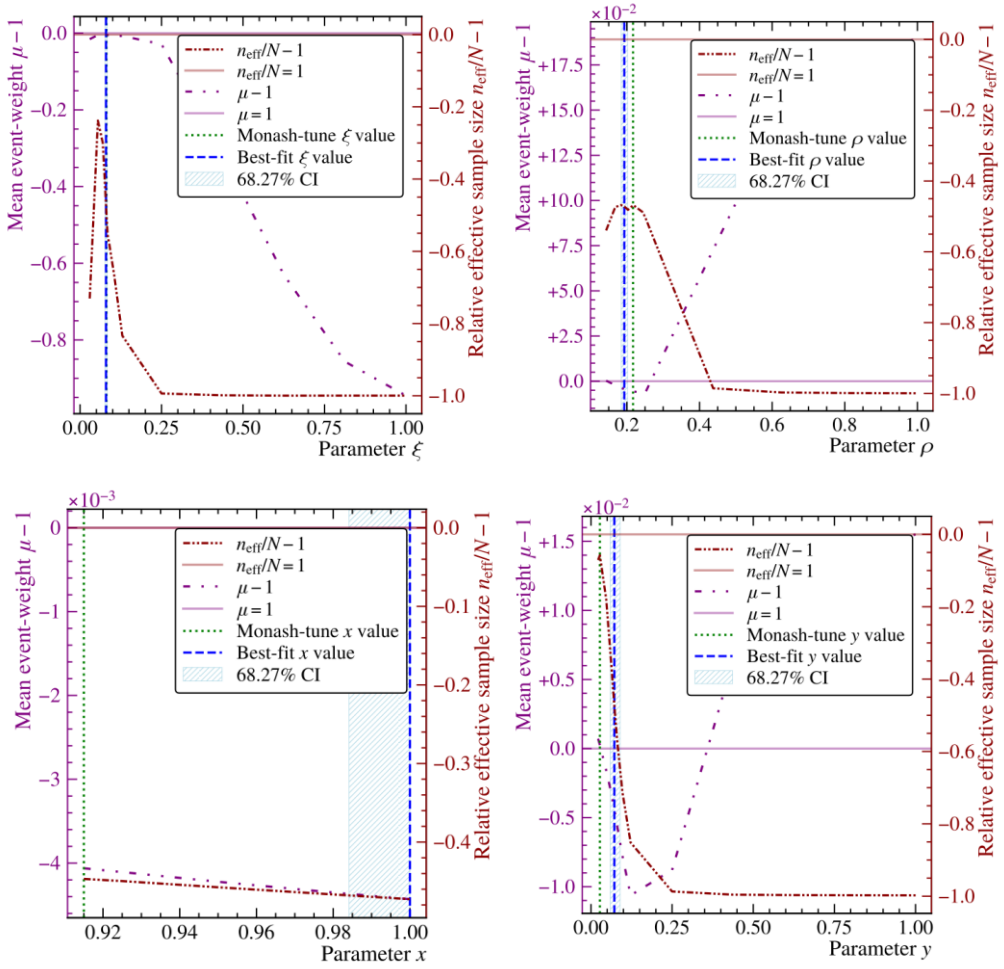




Reweighting support

Full parameter scan

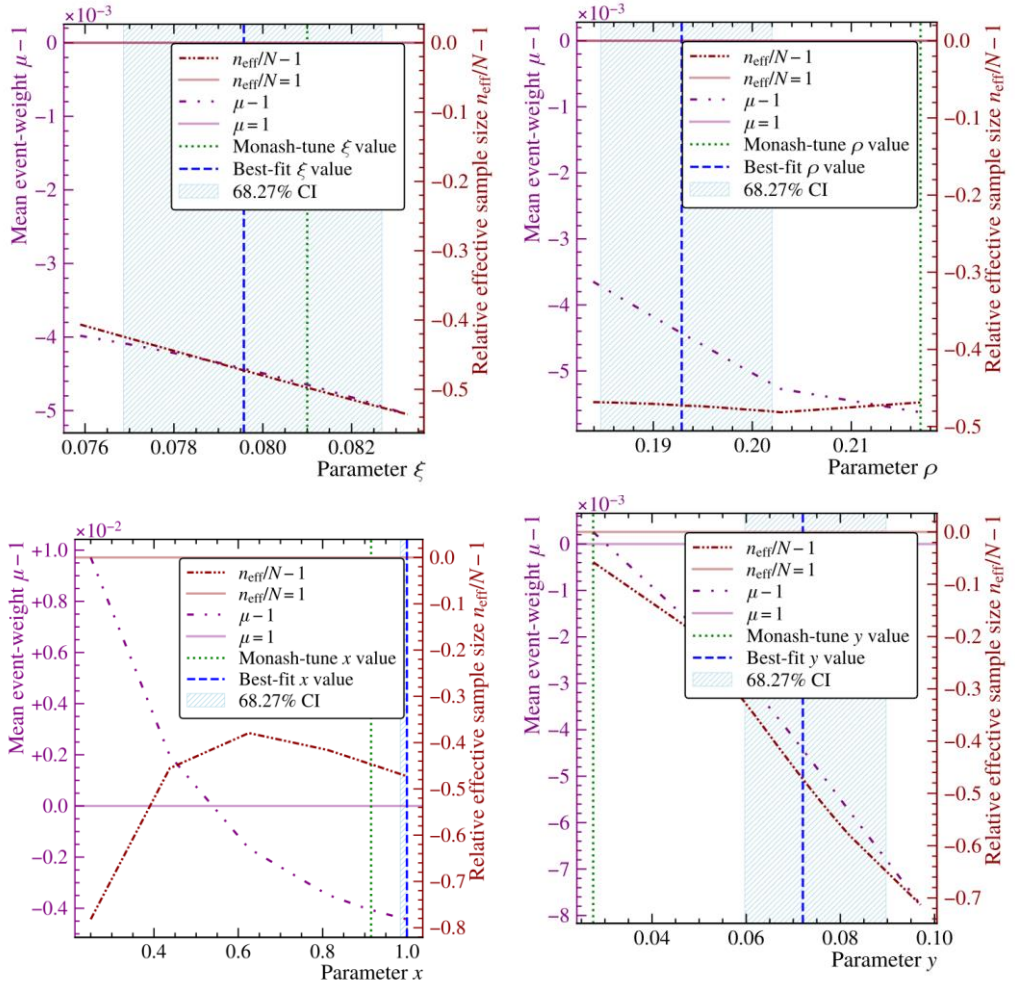
[1]



The mean weight and effective statistics are proxies for reweighting success [2].

Parameter scan near the minimum

[1]



1. B. Assi and others. Work in preparation.
2. C. Bierlich and others. *Reweighting Monte Carlo predictions and automated fragmentation variations in Pythia 8*. SciPost Phys., 16(5):134, 2024. arXiv:2308.13459.