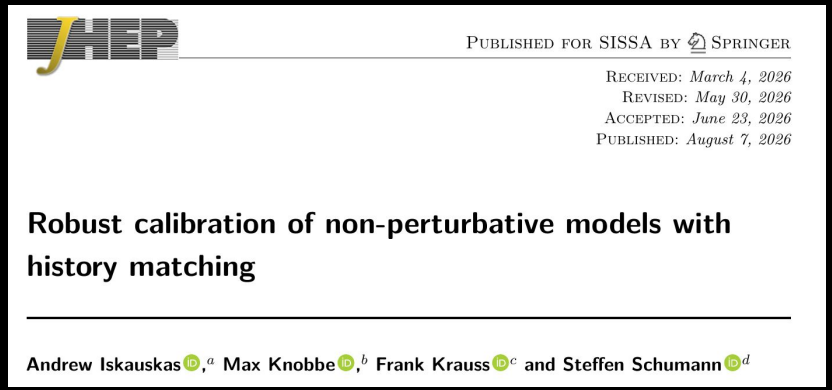
Model Calibration using History Matching

Bayesian uncertainty analysis

Steffen Schumann

Institut für Theoretische Physik

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737



JHEP

Published for SISSA by Springer

Received: March 4, 2026
Revised: May 30, 2026
Accepted: June 23, 2026
Published: August 7, 2026

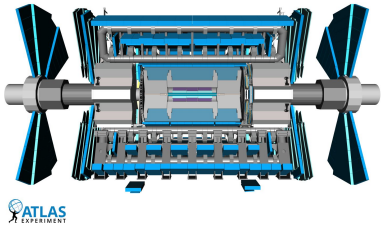
Robust calibration of non-perturbative models with history matching

Andrew Iskauskas [a], Max Knobbe [b], Frank Krauss [c] and Steffen Schumann [d]

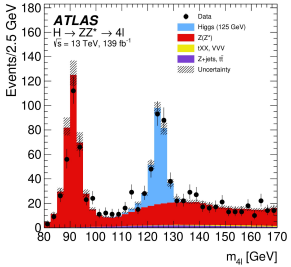
JHEP 08 (2026) 066

Computer models for complex physical systems

Contrast measured physical reality with (imperfect) model predictions

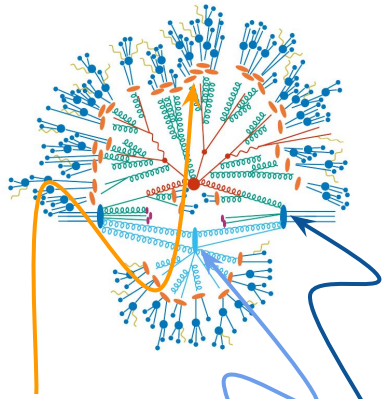


systematic errors
finite statistics



[ATLAS EPJC 80 (2020) 942]

truncation errors
model limitations



Hadronization
O(20) params

Intrinsic k_t
O(3) params

Underlying Event
O(5) params

Spotlight on non-perturbative models

- Specification of input parameters prior to simulator runs
  - High-dimensional input space [multimodal, correlations]
  - Costly to evaluate model predictions [simulator run time]
- Wide-range of potential model outputs and physical observations
  - Discrepancy of model and real-world process
  - Statistical and systematic measurement errors

Exploring high-dimensional spaces

Classic tuning strategy

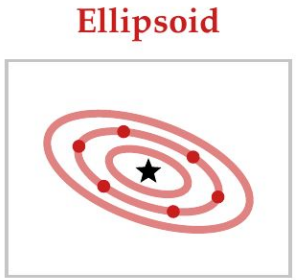
"find (single) best-fit parameter point and possibly uncertainty" based on *GoF measure*
guaranteed to find a solution
not necessarily unique, possibly unstable

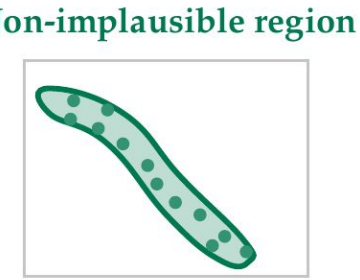
History Matching approach [Iskauskas et al. JSS 109 (2024) 1]

"identify parameter-space regions for which simulator output closely matches observations"
via complementarity to *implausible regions*
better captures multi-modality, correlations
requires samples from non-implausible region

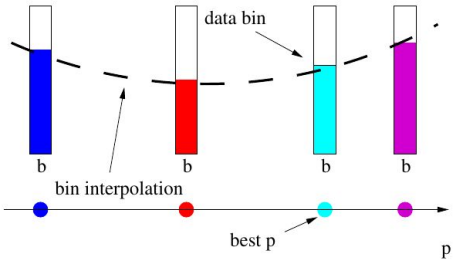
Commonalities: contrast input & observations

- select sensitive model outputs, i.e. observations
- use fast surrogate/emulator instead of simulator
- capture parameter dependence of model outputs





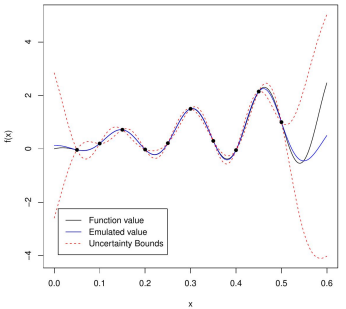
Professor/Apprentice approach
[Buckley et al. EPJC 65 (2010) 331]
[Krishnamoorthy et al. 2103.05748]



HM: Bayes Linear emulation

Construct emulator g(x) for each considered observation (aka observable bin i)

$$g_i(x) = \sum_{j=1}^{p} \beta_{ij} h_{ij}(x) + u_i(x)$$



- known basis functions h_{ij} and stochastic process u_i (zero mean)
- use Bayes Linear statistics as approximation to full Bayes analysis [Goldstein, Wooff '07 Wiley & Sons]
  - expectation and variance as primitives [no distributional statements needed]

$$E[g_i(x)] = \sum_{j=1}^{p} E[\beta_{ij}] h_{ij}(x)$$

$$Cov[g_i(x), g_i(x')] = Cov\left[\sum_{j=1}^{p} \beta_{ij} h_{ij}(x), \sum_{j=1}^{p} \beta_{ij} h_{ij}(x')\right] + \sigma_{u_i}^2 \exp\left(-\frac{(x-x')^2}{\theta_i^2}\right)$$

HM: implausibility measure

Exclude implausible parameter regions based on Bayes Linear emulators

- consider actual observation (data point) z_i with uncertainty $Var\left[e_i\right]$
- given emulator $g_i\left(x\right)$ with expectation $E\left[g_i\left(x\right)\right]$ and variance $Var\left[g_i\left(x\right)\right]$

$$I_i^2\left(x\right) = \frac{\left(E\left[g_i\left(x\right)\right] - z_i\right)^2}{Var\left[g_i\left(x\right)\right] + Var\left[e_i\right] + Var\left[\epsilon_i\left(x\right)\right]}$$

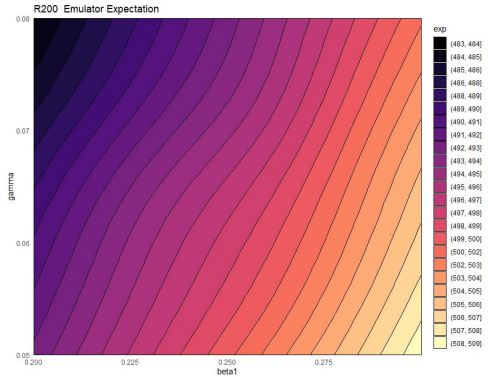
  - $Var\left[\epsilon_i\left(x\right)\right]$ accounts for additional modelling uncertainty [expert knowledge]

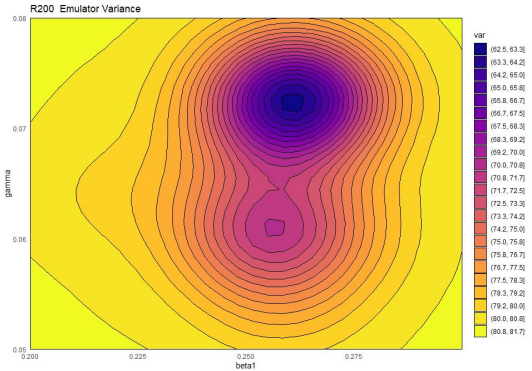
- discard input-parameter point if $I_M\left(x\right) = \max_{i=1,\ldots,N} I_i\left(x\right) > 3$

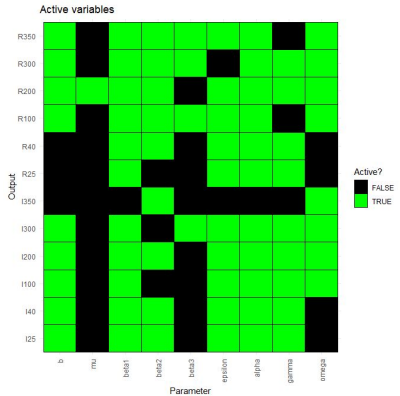
HM: emulator training & diagnostics

Generate collection of simulator predictions (50:50 training & validation)

- adjust expectation and variance of all emulators using training set
  - Bayes Linear update formulae for $E\left[g_i\left(x\right)\right]$ and $Var\left[g_i\left(x\right)\right]$ [Goldstein, Wooff '07 Wiley & Sons]
  - possibly identify reduced set of *active variables* [Vernon et al. Bayesian Analysis 5 (2010) 619]
  - yields updated beliefs about simulator output



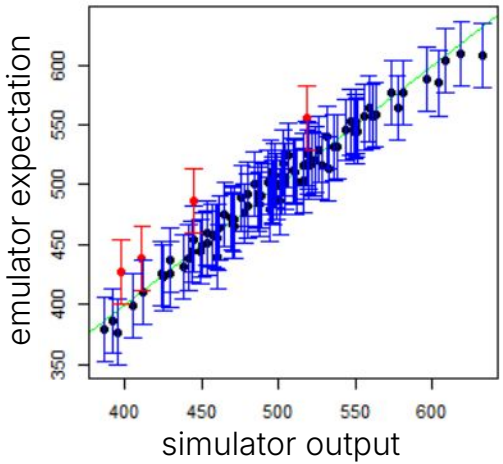


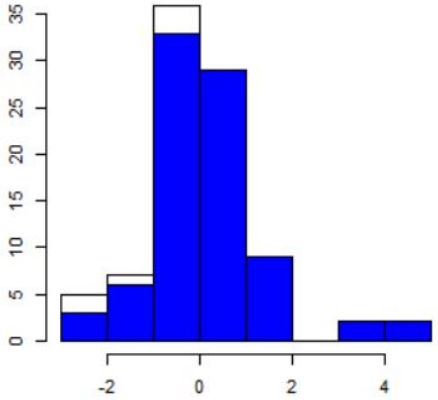


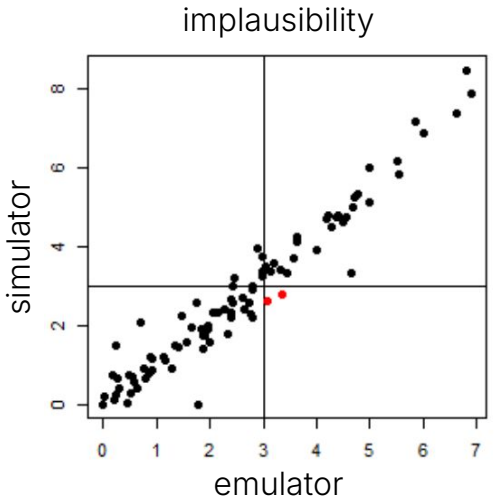
HM: emulator training & diagnostics

Generate collection of simulator predictions (50:50 training & validation)

- adjust expectation and variance of all emulators using training set
  - Bayes Linear update formulae for $E\left[g_i\left(x\right)\right]$ and $Var\left[g_i\left(x\right)\right]$ [Goldstein, Wooff '07 Wiley & Sons]
  - possibly identify reduced set of *active variables* [Vernon et al. Bayesian Analysis 5 (2010) 619]
  - yields updated beliefs about simulator output
- diagnostic analysis using validation runs







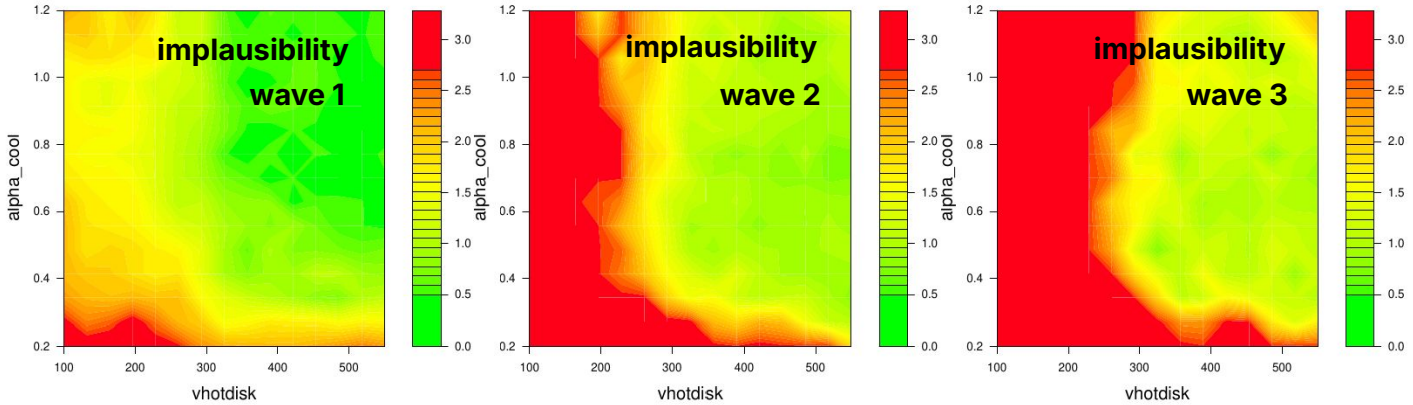
HM: consecutive space compression

subsequent waves of History Matching

wave 1: generate space-filling set of input-parameter points, train/validate emulators, *compress parameter-space volume* based on implausibility measure

wave k: generate new set of *points from remaining non-implausible volume*, train/validate *new* emulators, further compress volume based on implausibility measure

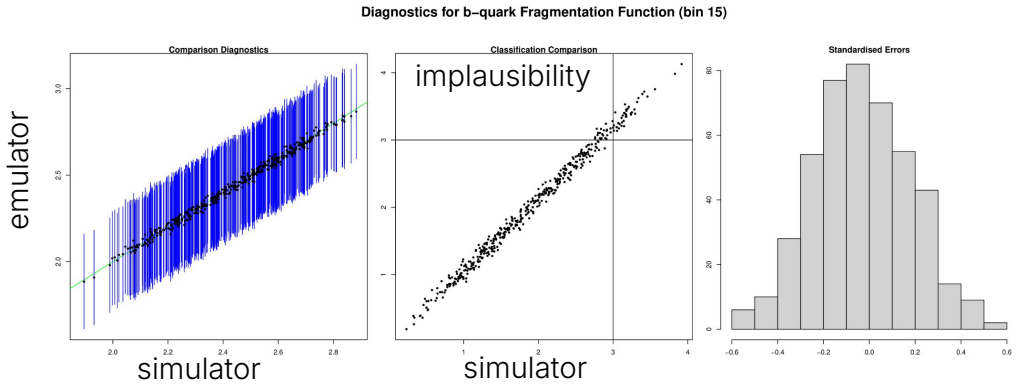
stopping criteria: relative compression factor less than 5%, emulator variance subdominant, ...



HM: hadronization-model calibration

Cluster- and Lund-string (via PYTHIA 8) fragmentation for SHERPA [Iskauskas et al. JHEP 08 (2026) 066]

- identical perturbative setup ee → qq+{0,1}j @ NLO + SHERPA dipole shower
  - facilitates assessment of hadronization-model dependence
- consider 19 parameters for AHADIC, 23 for PYTHIA Lund-string model
- constrained using large set of LEP1 data, e.g. event shapes, hadron multiplicities, …
  - total of 432 bins, determine reduced sets via Principal Variable Analysis
    [Cumming, Wooff Computational Statistics & Data Analysis 52 (2007) 550]
- $h_{ij}(x)$ polynomials up to 2nd order, O(800) simulator runs per wave (split 50:50)

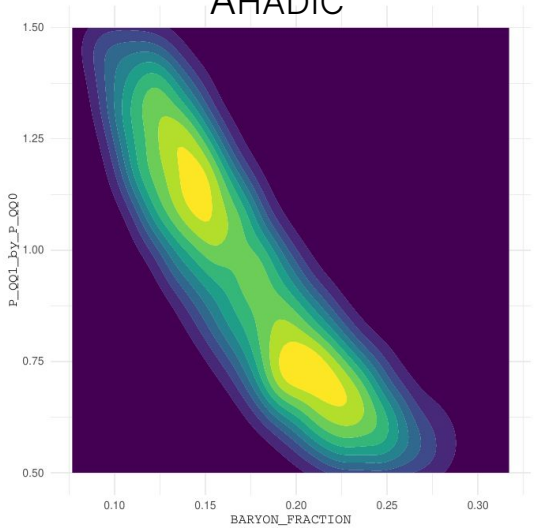


HM: hadronization-model calibration

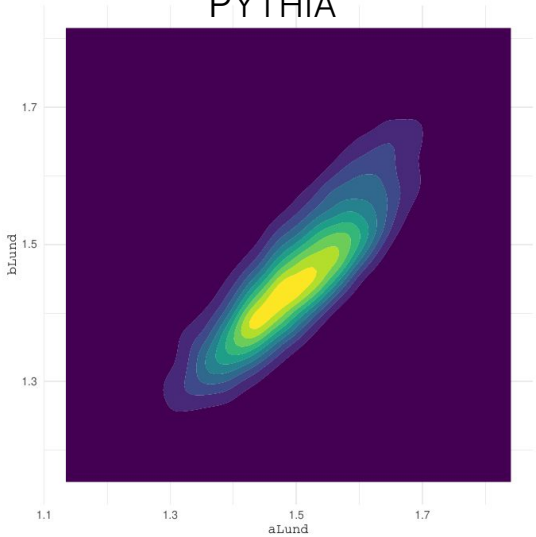
Cluster- and Lund-string (via PYTHIA 8) fragmentation for SHERPA [Iskauskas et al. JHEP 08 (2026) 066]

- after initial design, perform 3(5) waves of History Matching for AHADIC(PYTHIA)
- total compression factors: AHADIC ca. $5 * 10^{-4}$, PYTHIA ca. 10^{-12}
  - 2D projections of (final) non-implausible parameter space

AHADIC



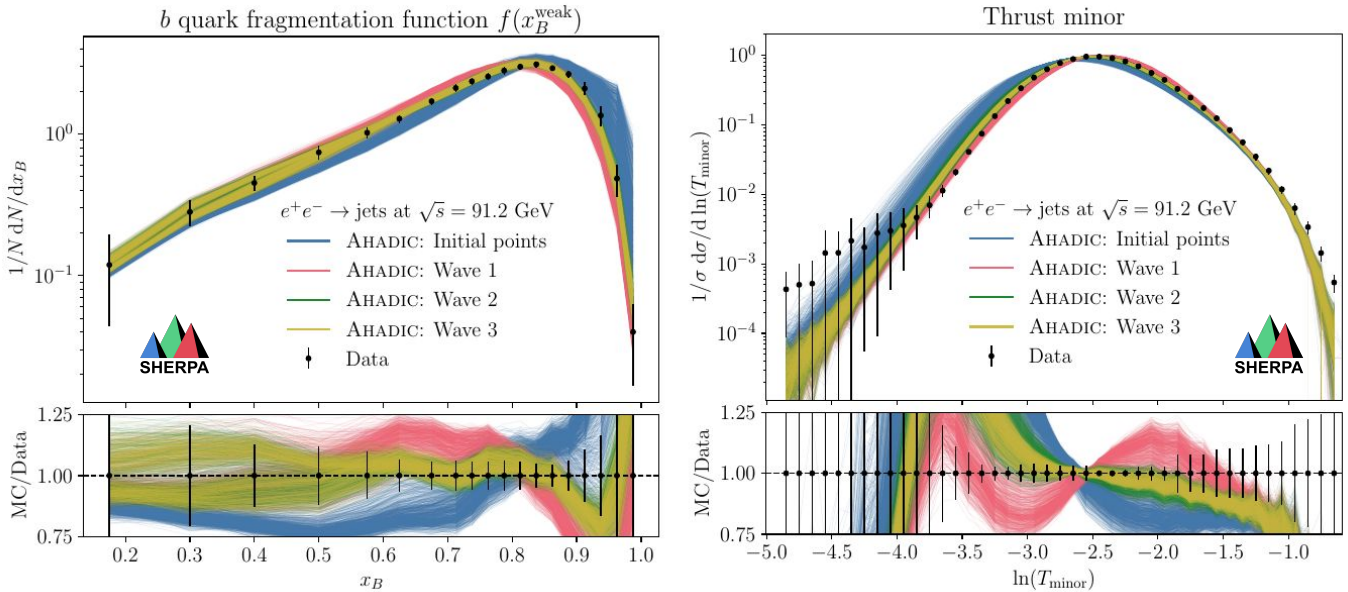
PYTHIA



HM: hadronization-model calibration

Cluster- and Lund-string (via PYTHIA 8) fragmentation for SHERPA [Iskauskas et al. JHEP 08 (2026) 066]

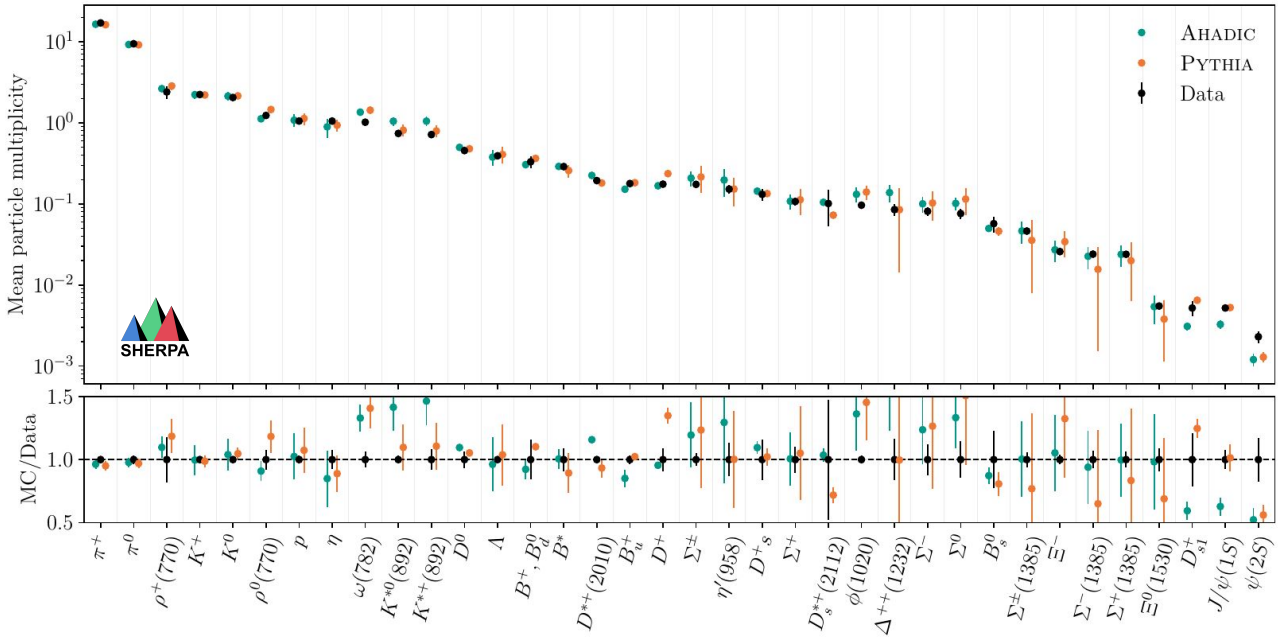
- wave-member envelopes for physical observables



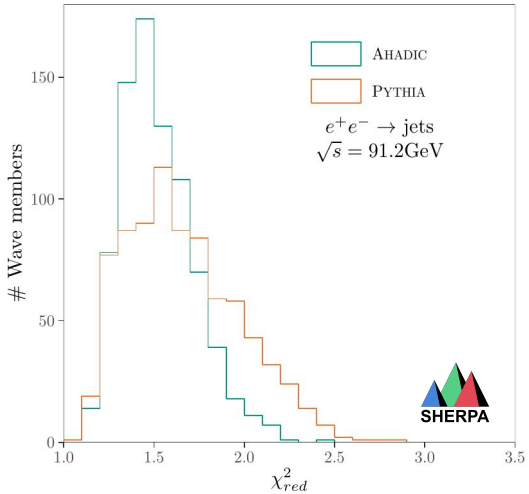
HM: hadronization-model calibration

Cluster- and Lund-string (via PYTHIA 8) fragmentation for SHERPA [Iskauskas et al. JHEP 08 (2026) 066]

- spread of final-wave members provides uncertainty estimate



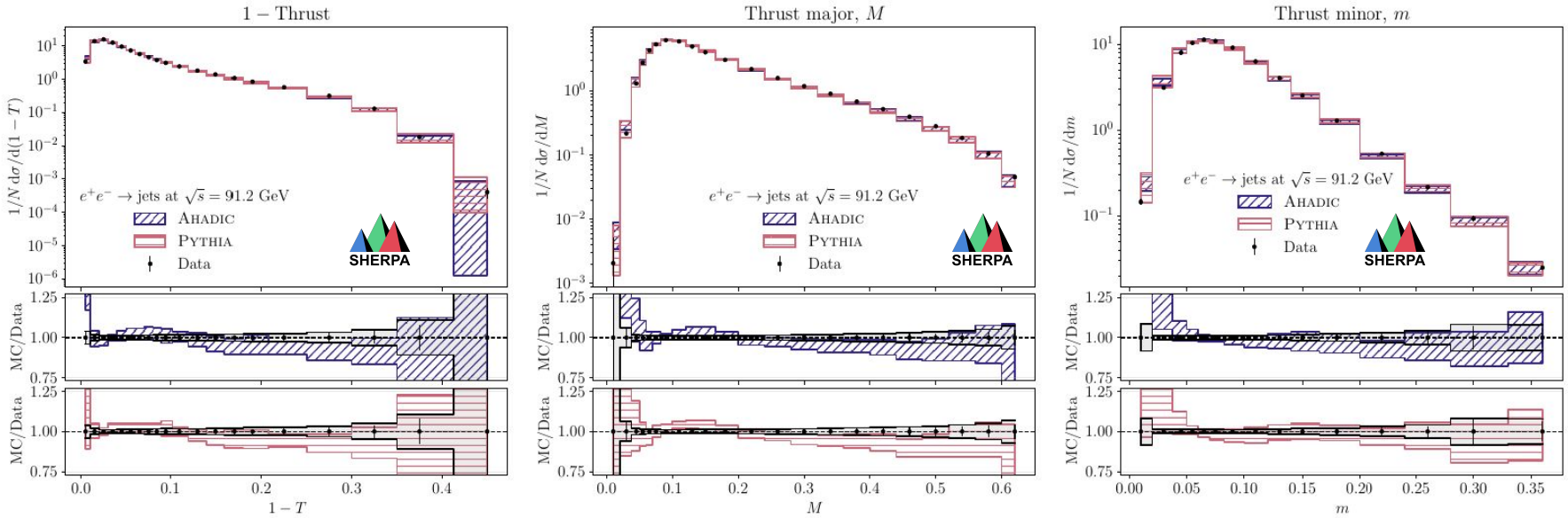
χ^2_{red} for full set of 432 outputs



HM: hadronization-model calibration

Cluster- and Lund-string (via PYTHIA 8) fragmentation for SHERPA [Iskauskas et al. JHEP 08 (2026) 066]

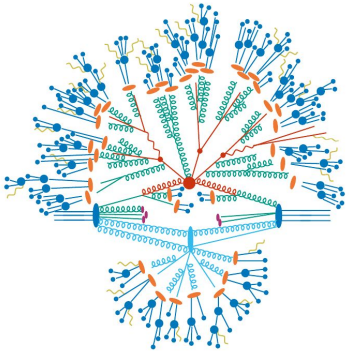
- spread of final-wave members provides uncertainty estimate



History Matching for Model Calibration

Analysis of non-perturbative uncertainties

- Careful calibration of model parameters (aka tuning)
- Quantification of model-inherent uncertainties
- Estimate dependence on choice of model



First application of HM for hadronization-model calibration

[Iskauskas et al. JHEP 08 (2026) 066]

- Identify *non-implausible regions* of parameter space
  - based on fast Bayes Linear emulators
- Compression of parameter-space volume (waves of History Matching)
  - account for observational errors and model discrepancy
- Spread of final-wave members provides uncertainty estimate
- Enables sensitivity studies in search for constraining observables

SHERPA

- Employed HMER package, written in R [Iskauskas et al. JSS 109 (2024) 1]
- Currently being adapted for workflows in HEP (use in conjunction with RIVET, YODA)
- Aim for wave-member generation and uncertainty evaluations via on-the-fly reweighting [Knobbe et al. JHEP 09 (2026) 087]

➔ see talk by Moritz