Multi-Differential and Unbinned Measurements of Hadronic Event Shapes in e^+e^- Collisions at $\sqrt{s} = 91$ GeV from ALEPH Open Data

Anthony Badea, Patrick Komiske, Eric Metodiev,
Ben Nachman, Yen-Jie Lee, Jesse Thaler, Austin Baty, Chris McGinn

ICHEP 2020 | Prague (Virtual)
July 30, 2020

Overview

Unfolding

3. Challenges with Traditional Unfolding
4. Omnifold

Data

5. ALEPH Archived Data

Physics

6. Observables: τ and $\log(\tau)$
7. Measurement

8. Summary & Comments
9. Thanks & Contact Info

Challenges with Traditional Unfolding

Previous methods are inherently binned

Binning fixed ahead of time, cannot be changed later

Performance of method sensitive to binning

Limited number of observables

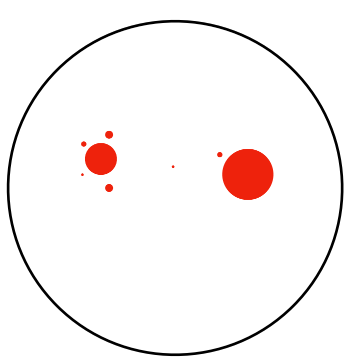
Binning induces curse of dimensionality

Response matrix depends on auxiliary features

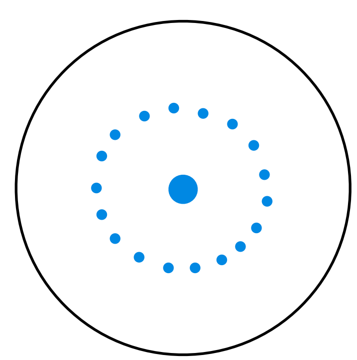
Detector-level quantity may not capture full detector effect

Ex. – Two jets acquiring the same mass in different ways

Jet 1
Two hard prongs



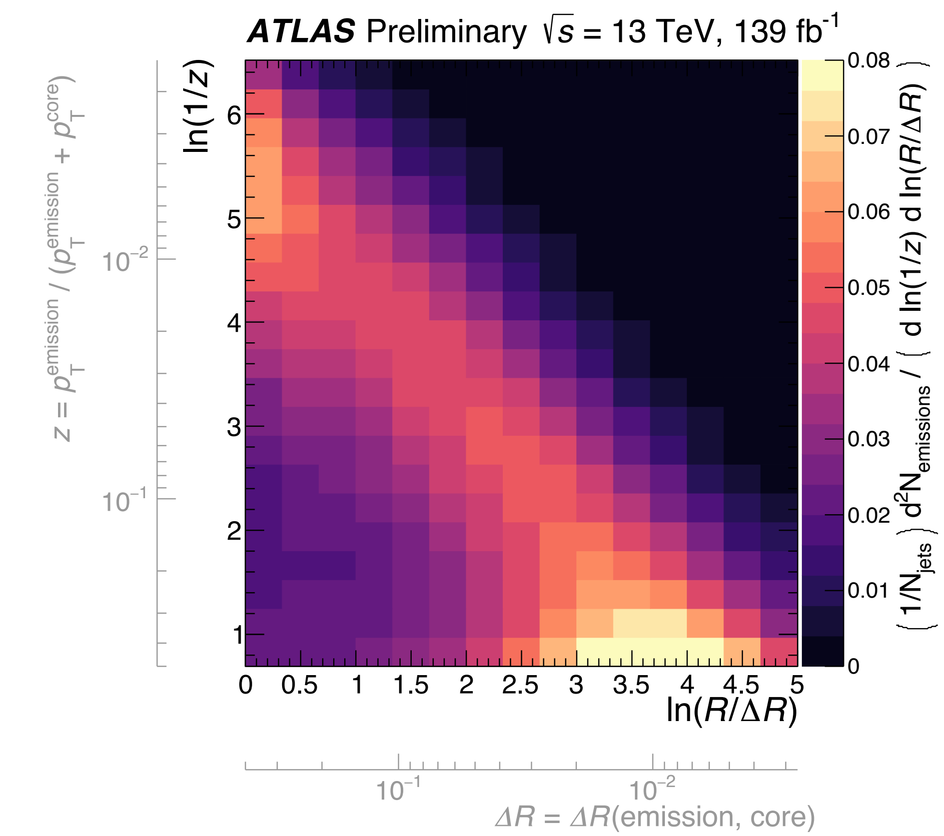
Jet 2
Hard core, diffuse spray



Example with IBU

ATLAS State-of-the-art Lund Plane Measurement

[ATLAS-CONF-2019-035]



21 x 15 bins in $\ln(1/z)$ x $\ln(R/\Delta R)$

- Must redo unfolding for other binnings e.g. finer/coarser, k_T (diagonal) binning, etc.

Limited to two observables

- 21^2 x 15^2 elements in response matrix R
- Going differential in n bins of p_T would multiply size of R by n^2

Slide borrowed from Patrick Komiske: click me

Omnifold

Multi-differential and unbinned machine learning based generalization of Iterative Bayesian Unfolding (IBU)

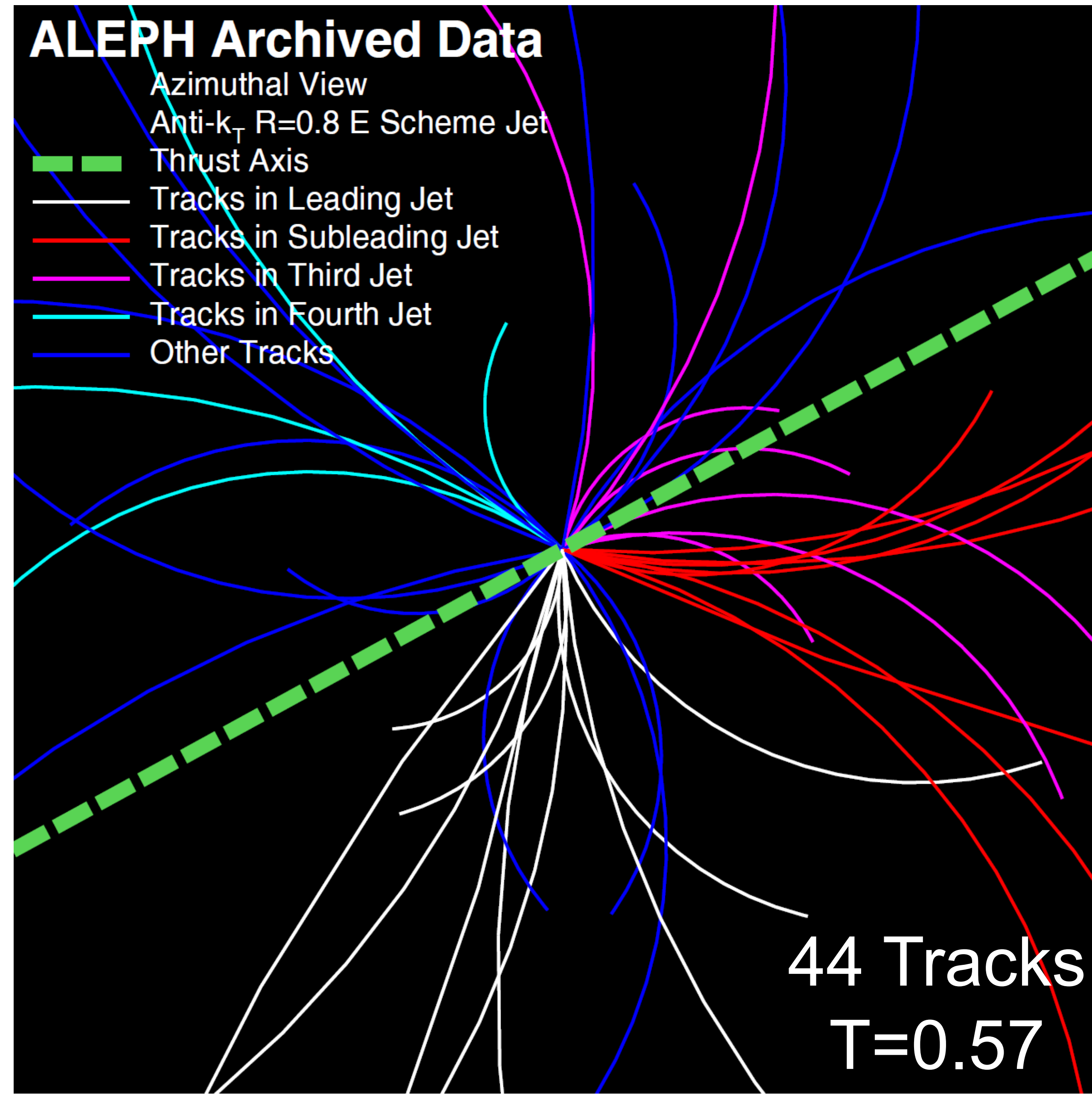
Key Insight:

1. Unfolding matrix accessible via the Likelihood ratios of data, reconstruction, and truth.
2. Likelihood ratio = optimal binary classifier by Neyman-Pearson Lemma
3. Use machine learning with softmax activation to get Likelihood ratios by training a network to classify events as data or reconstruction and reconstruction or truth

Andreassen, Komiske, Metodiev, Nachman, Thaler
Phys. Rev. Lett. 124, 182001
More details in backup

ALEPH Archived Data

- Collection at LEP between 1992 - 1995 with center-of-mass energy of 91 GeV
- Approx. 1.36 million e^+e^- collisions used for this study
- Collaboration event and track selection applied
- Unfolding utilizes archived Pythia 6.1 with matched truth and reconstructed events
- *Future*: compare with Pythia 8.230, Herwig 7.1.5, and Sherpa 2.2.6
- **More details in backup**



Many thanks to ALEPH Collaboration!!

Two-Particle Correlation: Badea et al. Phys. Rev. Lett. 123, 212002

Observables: τ and $\log(\tau)$

$$T \equiv \max_{\vec{n}} \left(\frac{\sum_{i \in E} |\vec{n} \cdot \vec{p}_i|}{\sum_{i \in E} |\vec{p}_i|} \right)$$

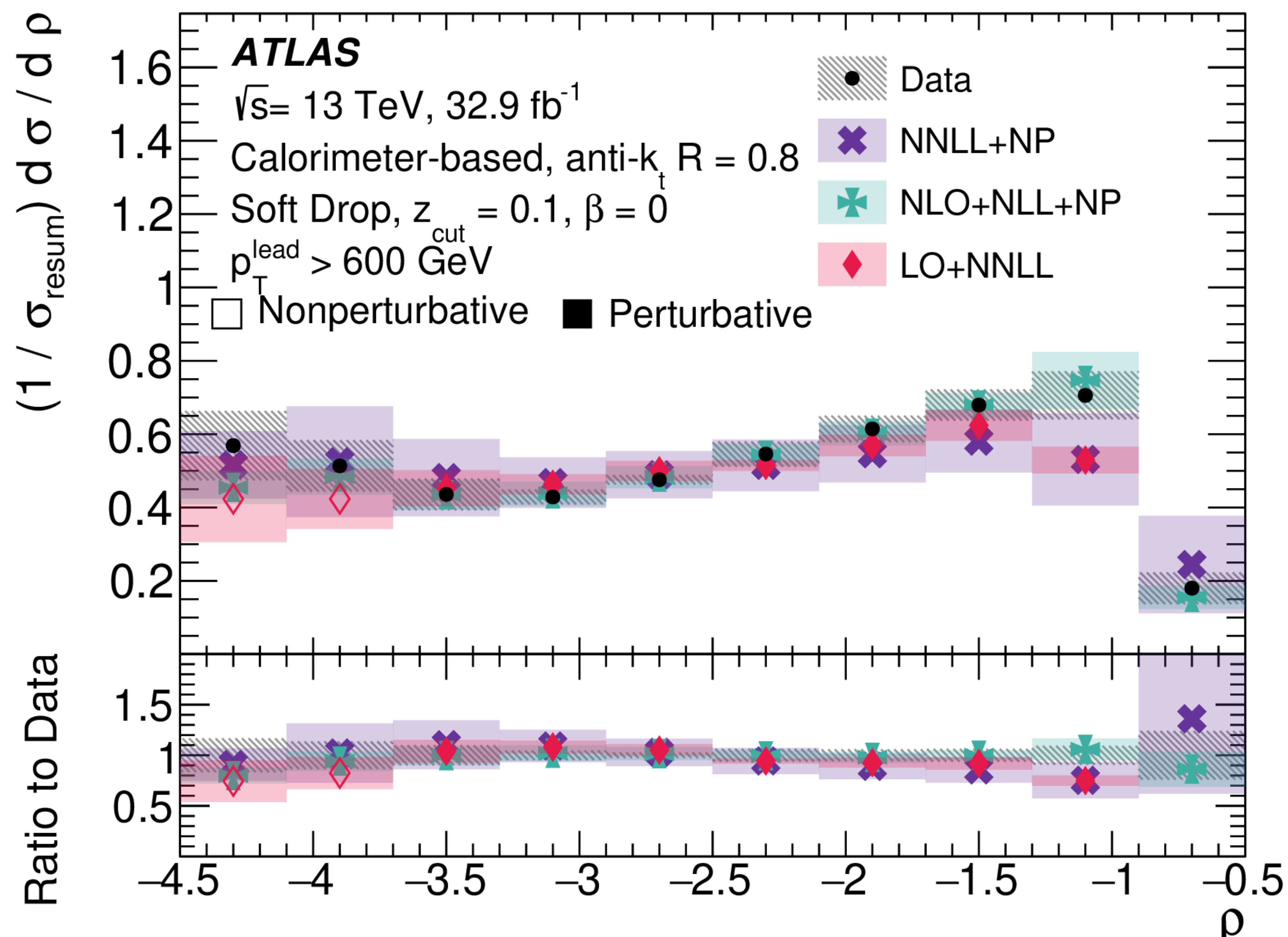
Why $\tau = 1$ - Thrust?

Discrepancy with theory predominantly in the highly perturbative regions = low T or high τ. Theoretical calculations often work with τ.

Why linear and log space?

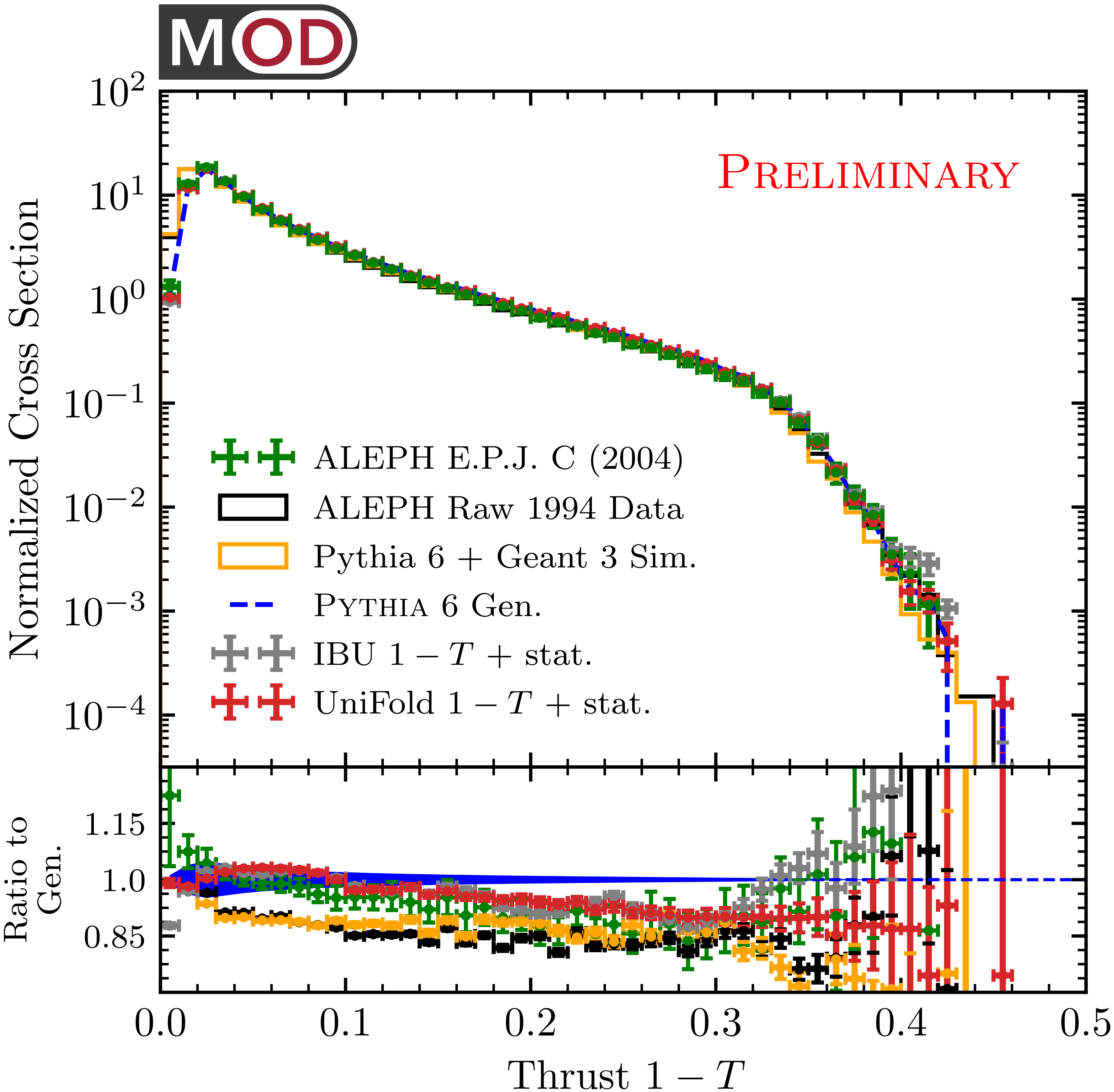
Linear and log space highlight different physics. Insights into perturbative vs. nonperturbative regimes.

Nonperturbative vs. Perturbative

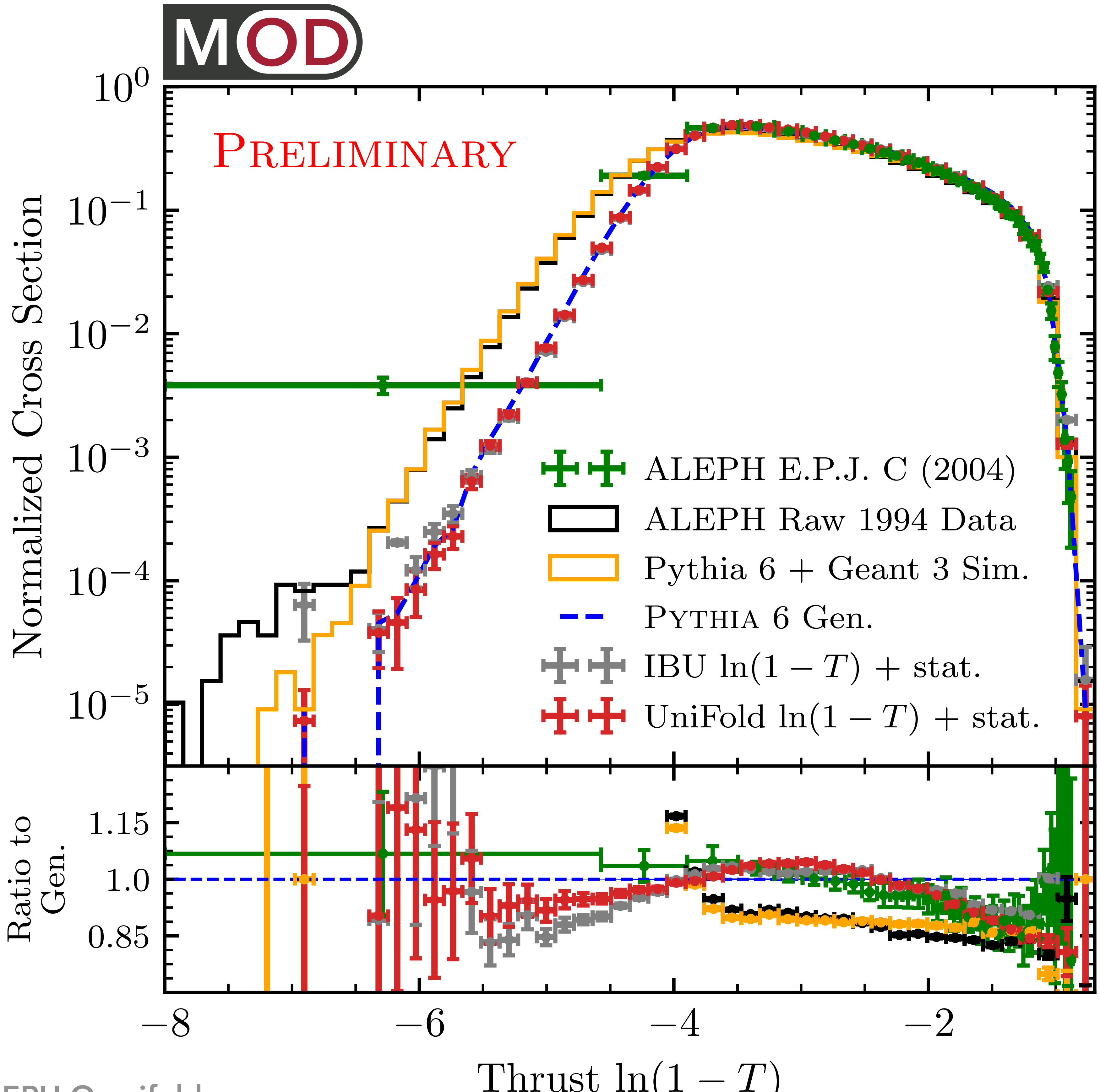


Phys. Rev. D **101**, 052007 FIG. 9b

Measurement: $\tau = 1 - T$



Measurement: $\ln(\tau) = \ln(1 - T)$



Summary & Comments

Significance of the Technique

Current and previous results suggest that Omnifold is a viable solution to challenges facing modern unfolding methods

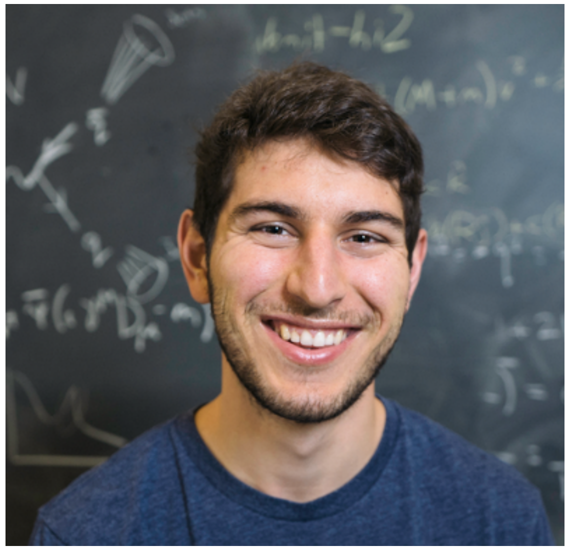
Utilizing Archived Data

Significant progress has been made in deploying Omnifold on real collider data via the ALEPH Archived Data

Ongoing Work

Handle uncertainties, compare results with different Monte Carlo, perform multi-differential unfolded measurements

Thank you from the analysis team



Anthony Badea

Patrick T. Komiske III

Eric Metodiev

Yen-Jie Lee

Ben Nachman

Jesse Thaler

Email us questions/comments: abadea@g.harvard.edu!

BACKUP

Omnifold

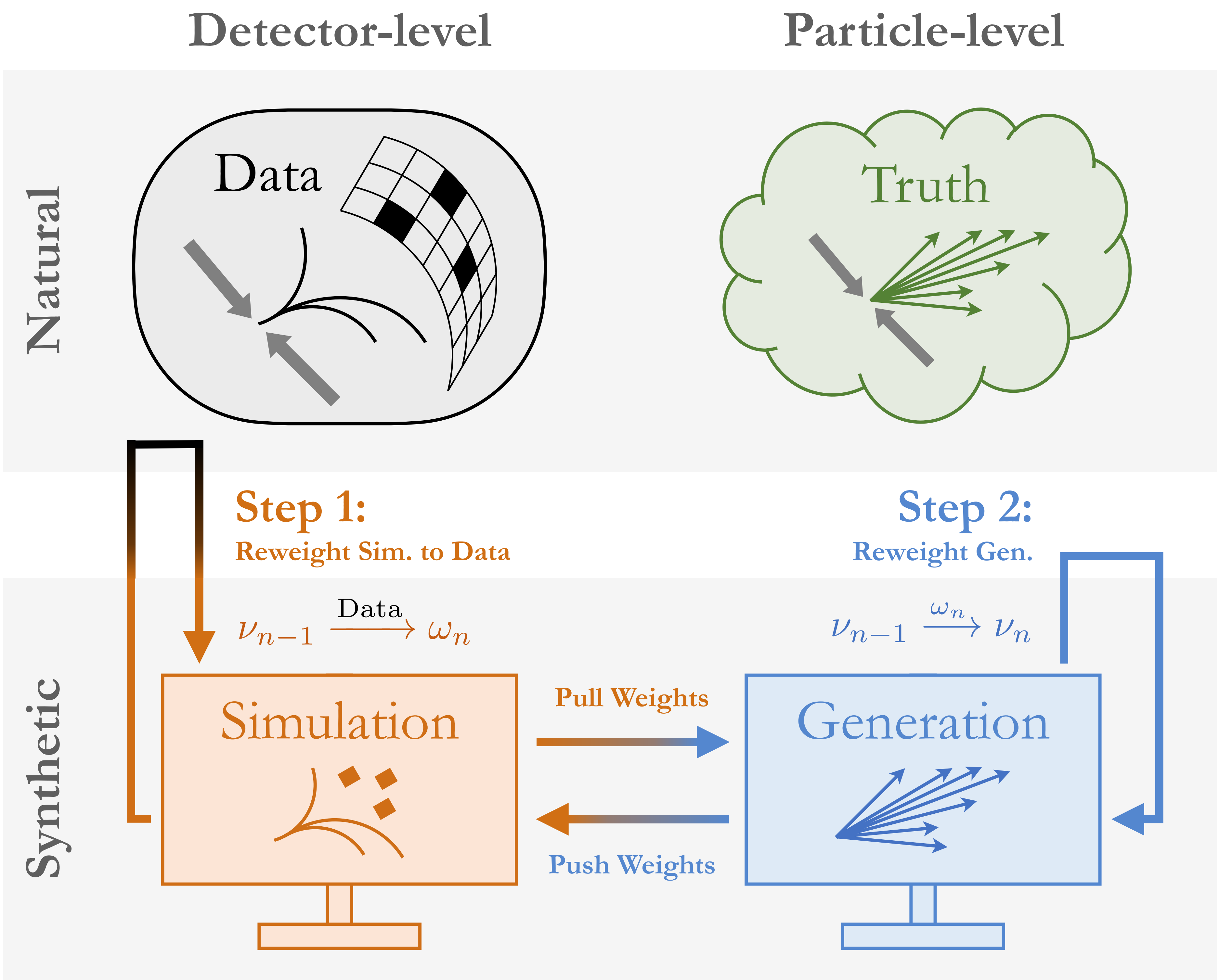
Challenges with Traditional Unfolding

- Goal of unfolding: correct the measured data for effects arising from the finite efficiency and resolution of the detector
- Current prominent methods: bin-by-bin correction factors, matrix inversion, template fits with regularization, iterative methods
- Above methods apply to binned distributions and **must be** repeated for each desired binning and observable
- **Dream**: all analyses use same unfolding done prior to choice of observable or binning. Unfold the *entire* phase space.

Omnifold Resources

- List of talks explaining the procedure in more depth: https://indico.cern.ch/event/906711/contributions/3851602/attachments/2043779/3424363/OmniFoldATLAS.pdf
- Jupyter notebook demo: https://mybinder.org/v2/gh/ericmetodiev/OmniFold/master?filepath=OmniFold%20Demo.ipynb

Omnifold

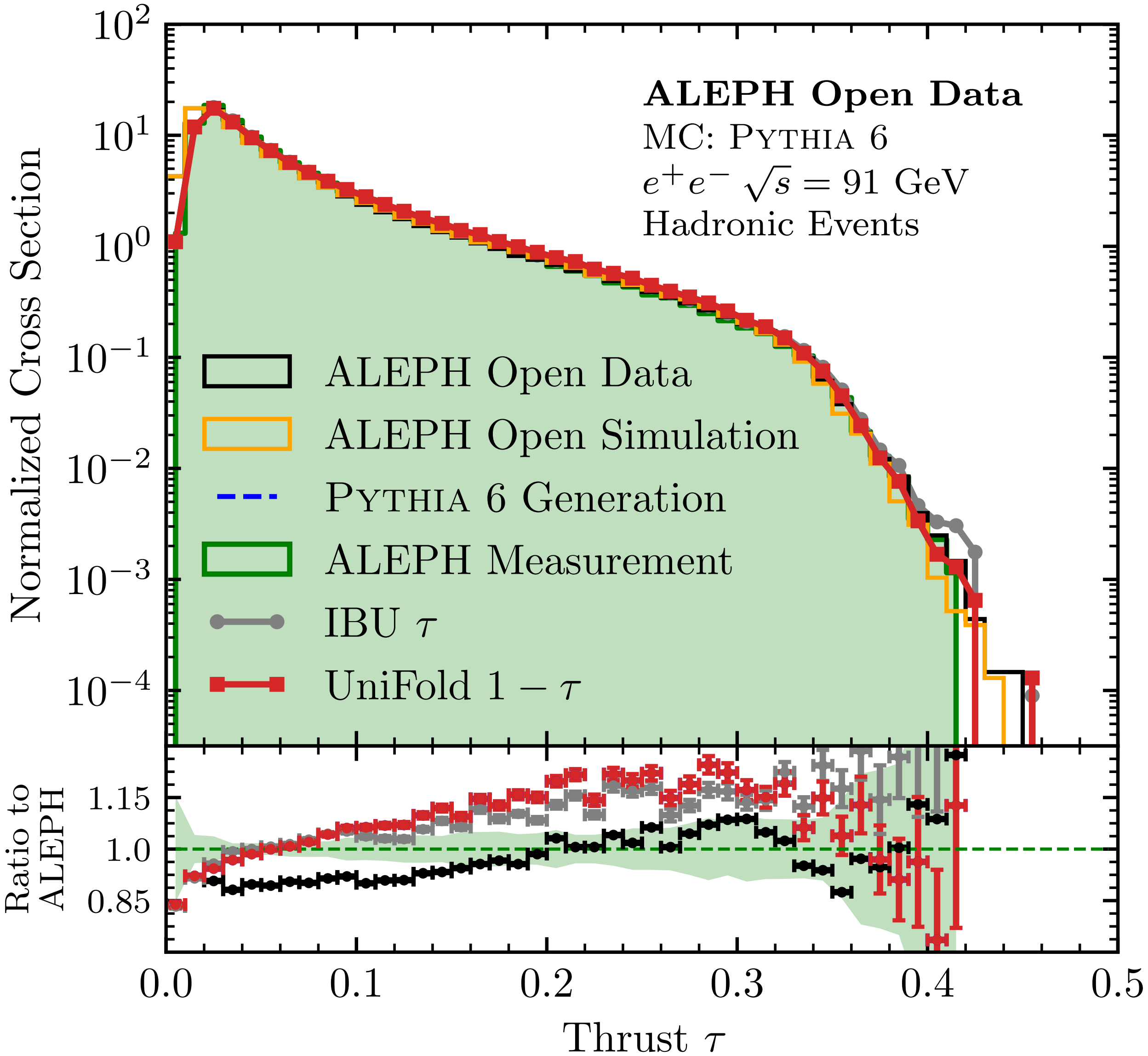


Andreassen et al. Phys. Rev. Lett. 124, 182001

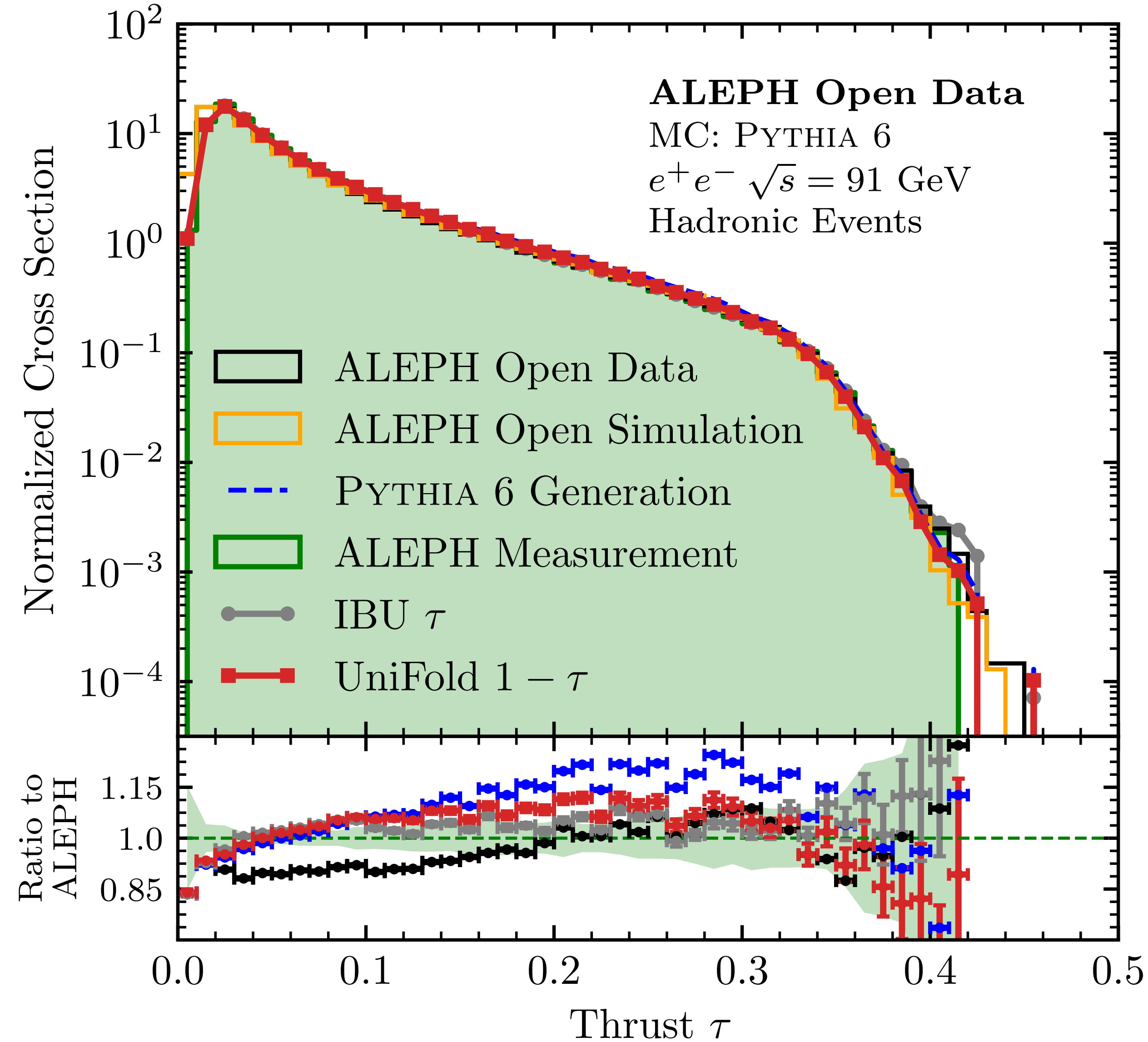
Outline of Uncertainty Calculation

- Statistical handled via bootstrapping
- Systematics
  - Vary event and track selections
- Method & Non-closure Uncertainties
  - Repeat unfolding with different Monte Carlo — Pythia 8, Herwig 7, Sherpa 2

Uncorrected



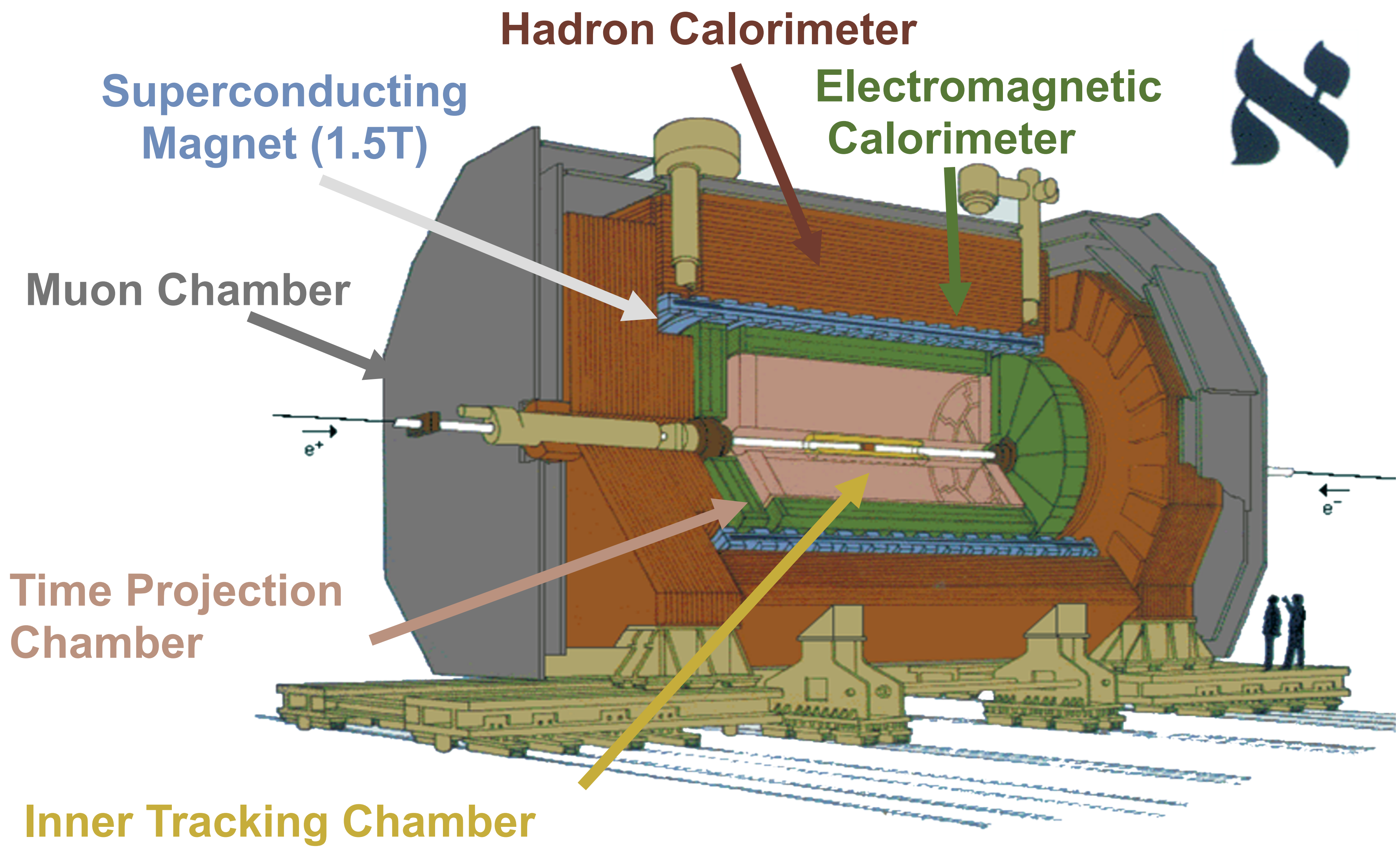
Corrected



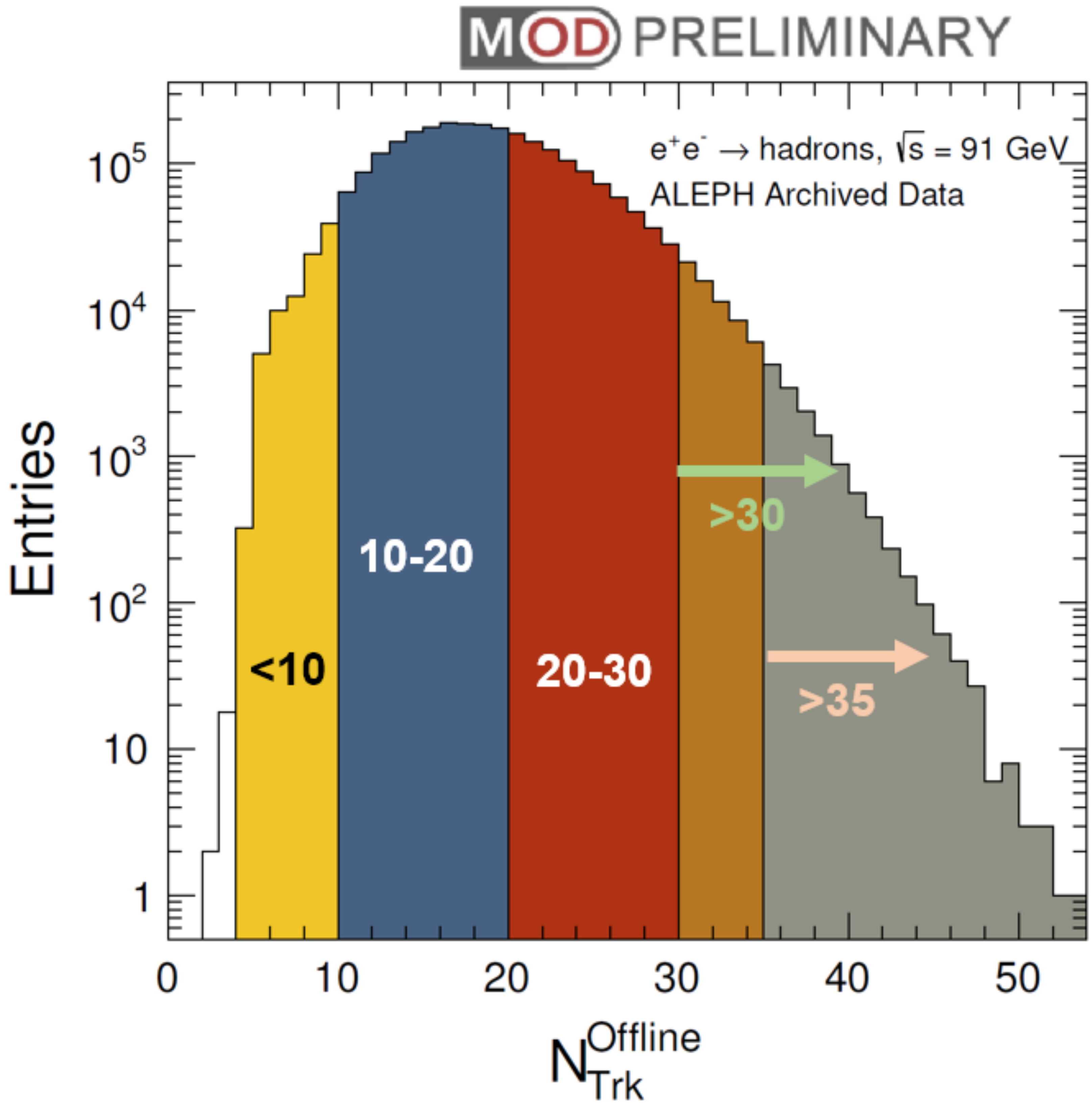
The truth Monte Carlo with an event-by-event matched reconstruction was used for the unfolding. This Monte Carlo, however, has hadronic event selections applied. There is another truth Monte Carlo without hadronic event selections that is event-by-event matched to the truth Monte Carlo with selections but not to reconstruction. A correction is applied to take us from truth with hadronic event selection to truth without hadronic event selection.

ALEPH Data

ALEPH Detector at LEP

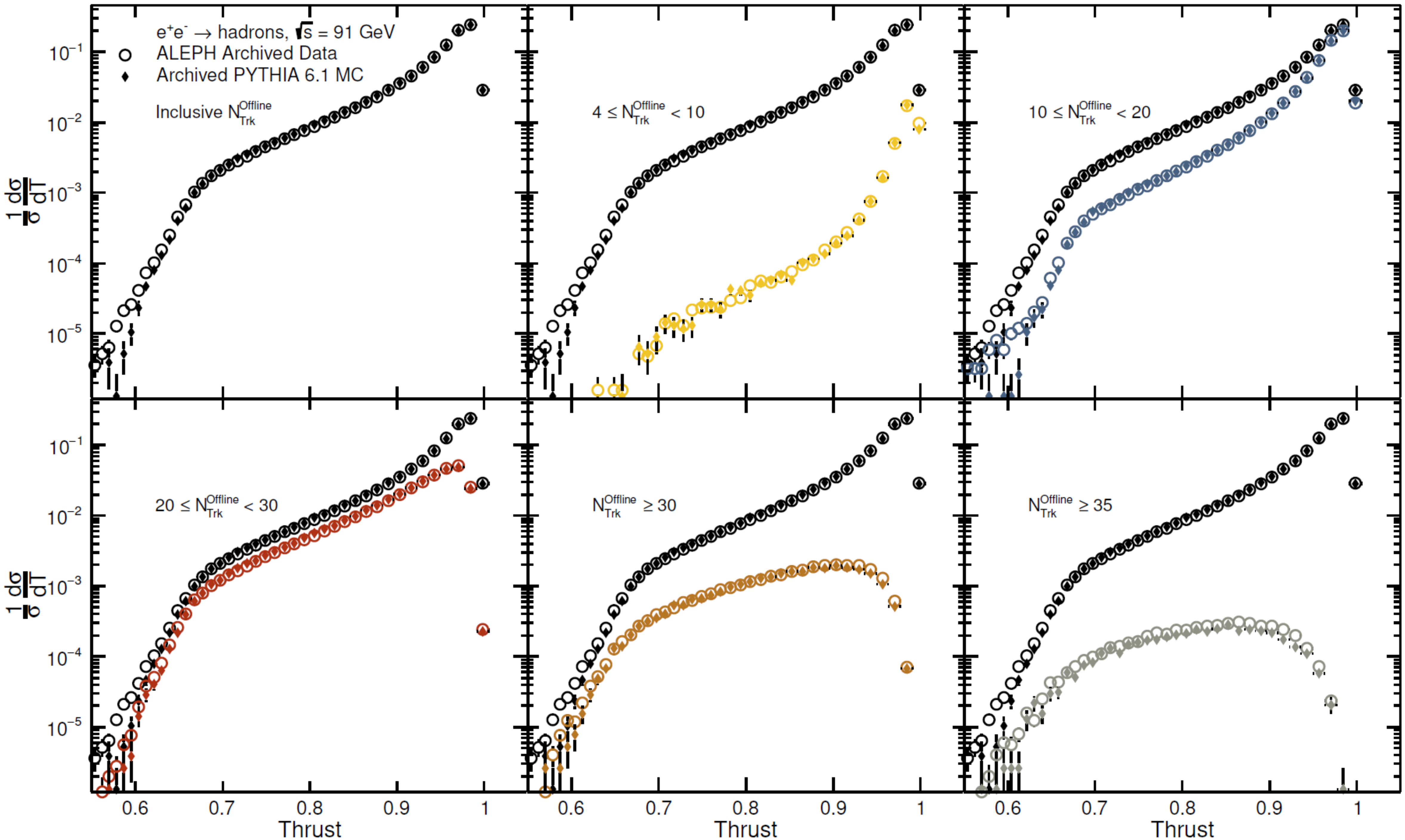


Multiplicity Distribution



Thrust Distribution split in Multiplicity Bins

MOD PRELIMINARY



Event Selection

LEP 1 Events satisfying selection have a 1 inside of the tBranch ‘passesAll’

passesAll defined by:

- passesNTupleAfterCut: Checks that the run and event number of the picked up data matches the recorded numbers. Used to verify that the data is recorded in the nTuple correctly. NOT A CUT.
- passesTotalChgEnergyMin: Total Energy of Charge Tracks >= 15 GeV
- passesNTrkMin: # of Charged Tracks >= 5
- passesNeuNch: # of Charged Tracks + Neutral >= 13
- passesSTheta: $|\cos\left(\theta_{Sphericity}\right)| \leq 0.82$
- passesMissP: Missing Momentum < 20 GeV (Only LEP2) → TURNED OFF
- passesISR: More detailed look at link below (Only LEP2) → TURNED OFF
- passesWW: More detailed look at link below (Only LEP2) → TURNED OFF

Following guidance of:
http://cds.cern.ch/record/690637/files/ep-2003-084.pdf, Section 2.2 page 2

% From Each Event Selection

	Raw nEvent	passesNTupleAfterCut	passesTotalChgEnergyMin	passesNTrkMin
LEP 1992	14024	1.0	1.000000	0.999715
LEP 1993	538601	1.0	0.999794	0.999870
LEP 1994	1365440	1.0	0.999801	0.999851
LEP 1995	595095	1.0	0.999807	0.999887

	passesSTheta	passesMissP	passesISR	passesWW	passesNeuNch	passesAll
LEP 1992	0.976683	0.971549	0.989375	0.989375	0.994438	0.945451
LEP 1993	0.977052	0.972796	0.989944	0.989775	0.994645	0.947002
LEP 1994	0.976877	0.972935	0.989827	0.989784	0.994567	0.947070
LEP 1995	0.977225	0.973085	0.989714	0.989531	0.994532	0.947402

Track Selection (High Purity)

Tracks satisfying selection have a 1 inside of the tBranch ‘highPurity’

highPurity defined by:

- passesPWFlag: 0,1,2 (only pick up charged tracks and leptons)
- Starting from 0 pwflag (via Marcello) - CHARGED_TRACK, CHARGED_LEPTONS1, CHARGED_LEPTONS2, V0, PHOTON, NEUTRAL_HADRON
- passesD0: $|d_0| \leq 2$
- passesZ0: $|z_0| \leq 10$
- passesNTPC: # hits in TPC ≥ 4
- passesAbsCosThCut: $|\cos\theta_{ch}| \leq 0.94$ (only includes charged tracks)
- passesPt: $p_T \geq 0.2$

  Calorimeter object cuts NOT implemented:
- $|\cos\theta_{cal.\ obj.}| \leq 0.98$
- $E_{cal.\ obj.} \geq 0.8$ GeV

% From Each Track Selection

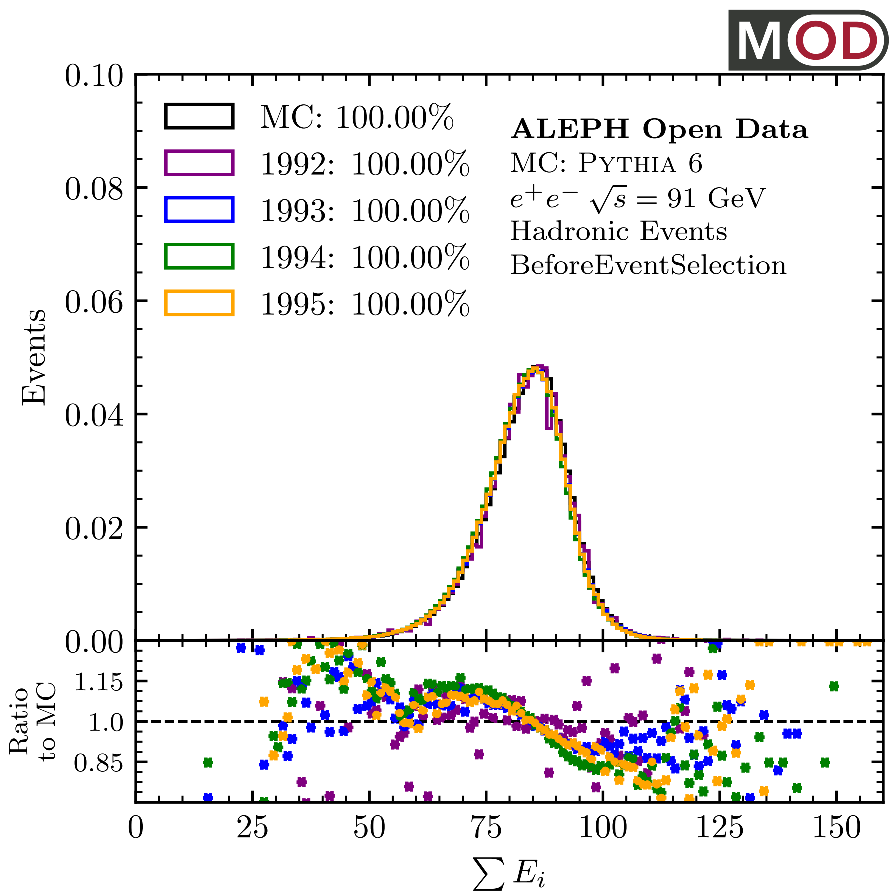
	pwflag	d0	z0	ntpc	charged theta	pt	highPurity
Cut	2.000000e+00	2.000000e+00	1.000000e+01	4.000000e+00	9.400000e-01	2.000000e-01	1.000000e+00
1992	4.128370e+05	4.128370e+05	4.128370e+05	4.128370e+05	2.615710e+05	4.128370e+05	4.128370e+05
% w/ Cut	6.335939e-01	6.336617e-01	6.336617e-01	6.336617e-01	9.999771e-01	9.831192e-01	9.830296e-01
1993	1.590741e+07	1.590741e+07	1.590741e+07	1.590741e+07	1.006942e+07	1.590741e+07	1.590741e+07
% w/ Cut	6.330015e-01	6.330727e-01	6.330727e-01	6.330727e-01	9.999800e-01	9.825680e-01	9.824973e-01
1994	4.005171e+07	4.005171e+07	4.005171e+07	4.005171e+07	2.550923e+07	4.005171e+07	4.005171e+07
% w/ Cut	6.369073e-01	6.369776e-01	6.369776e-01	6.369776e-01	9.999827e-01	9.825122e-01	9.824287e-01
1995	1.755621e+07	1.755621e+07	1.755621e+07	1.755621e+07	1.111647e+07	1.755621e+07	1.755621e+07
% w/ Cut	6.331930e-01	6.332619e-01	6.332619e-01	6.332619e-01	9.999831e-01	9.828150e-01	9.827427e-01

NOTES:

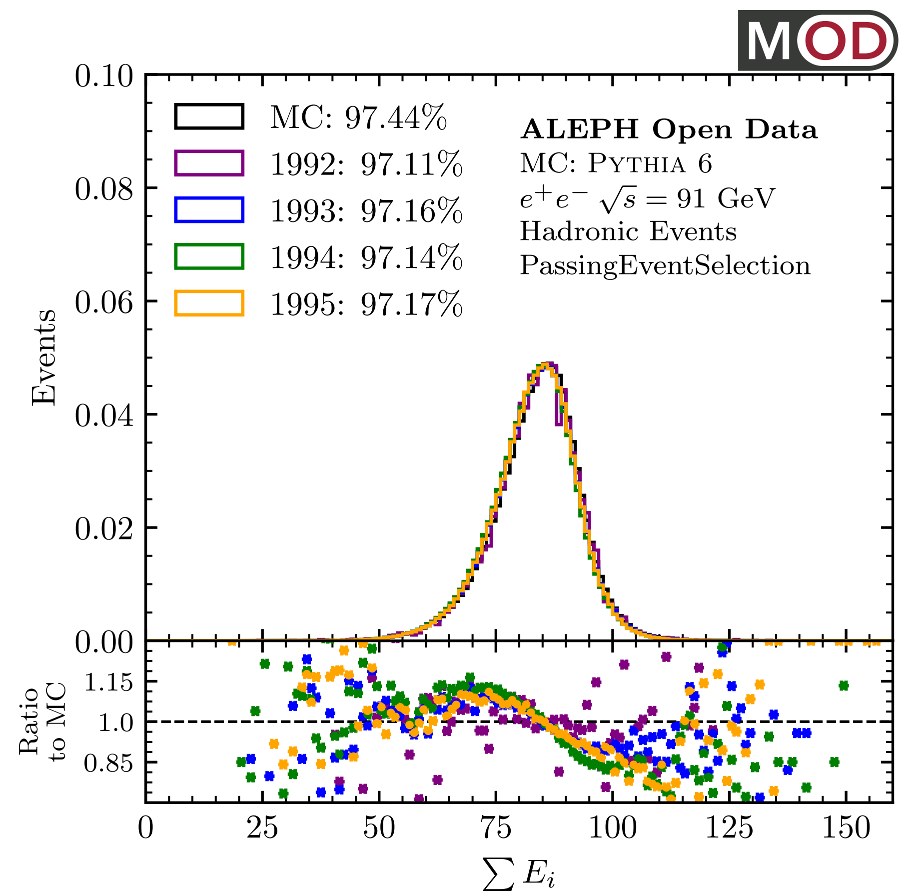
- d0, z0 = -999.0 if track was not selected by ALEPH. In MITHIG nTuples all tracks with d0, z0 != -999.0 pass the selection.
- nTPC = -127.0 if track was not selected by ALEPH. In MITHIG nTuples all tracks with nTPC != -127.0 pass the selection.

Total Event Energy

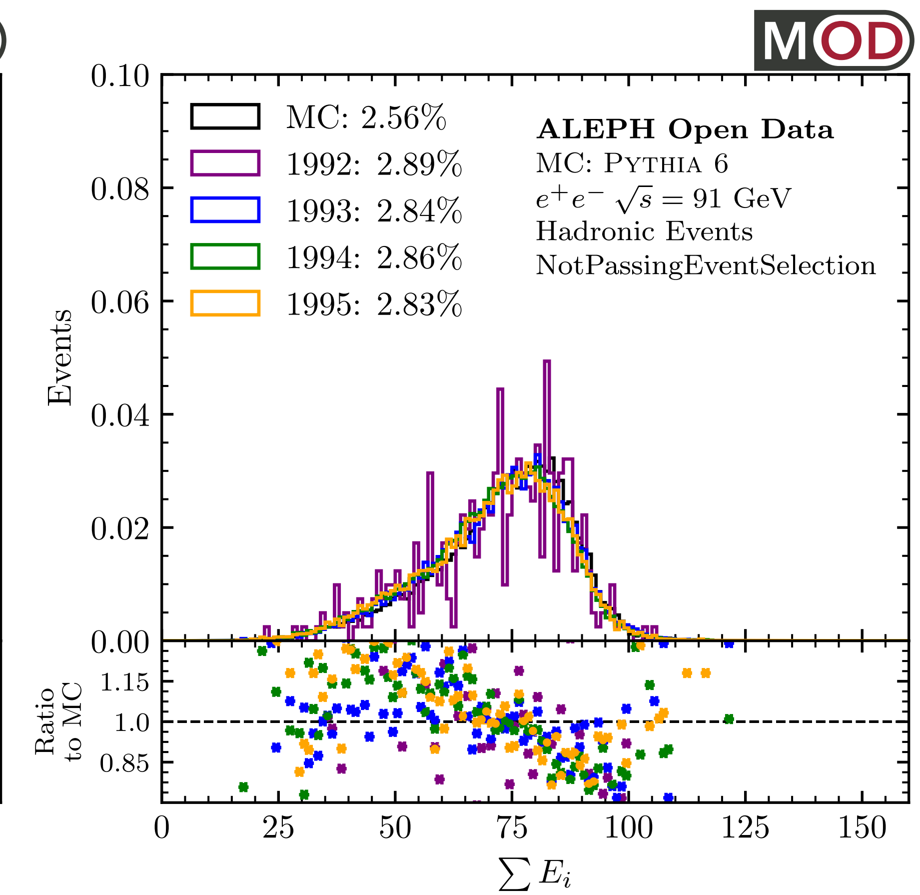
Before Event Selection



Passing Event Selection



Failing Event Selection



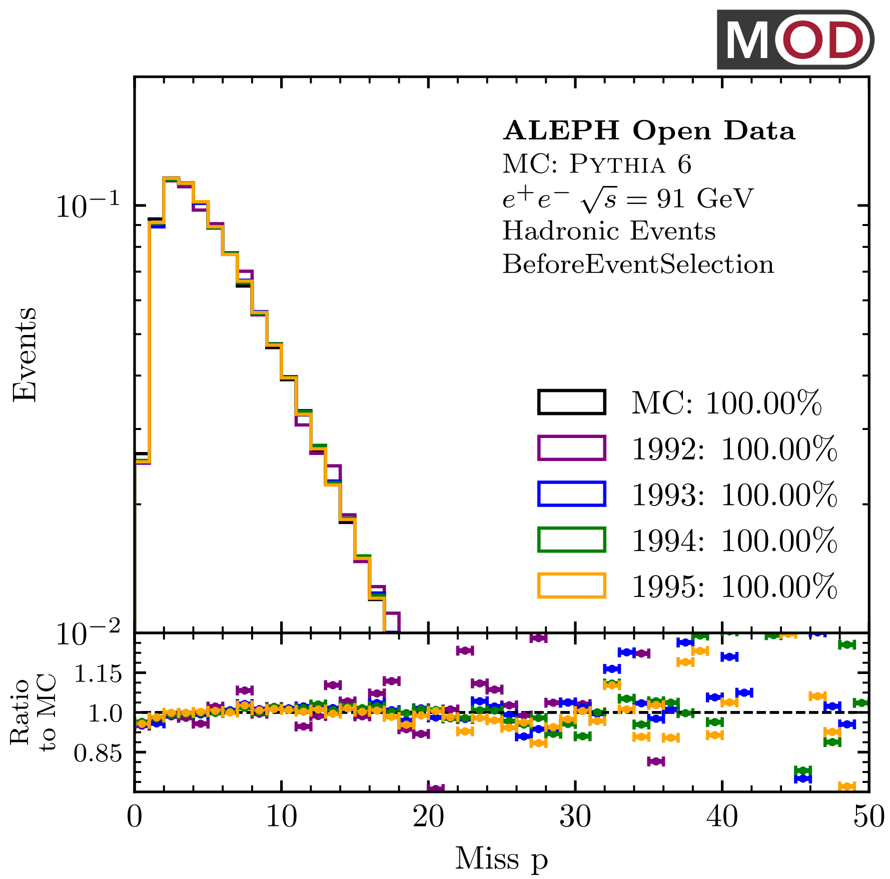
LEP1 1992 1993 1994 1995 : % Events

Pythia 6.1 (1994) MC

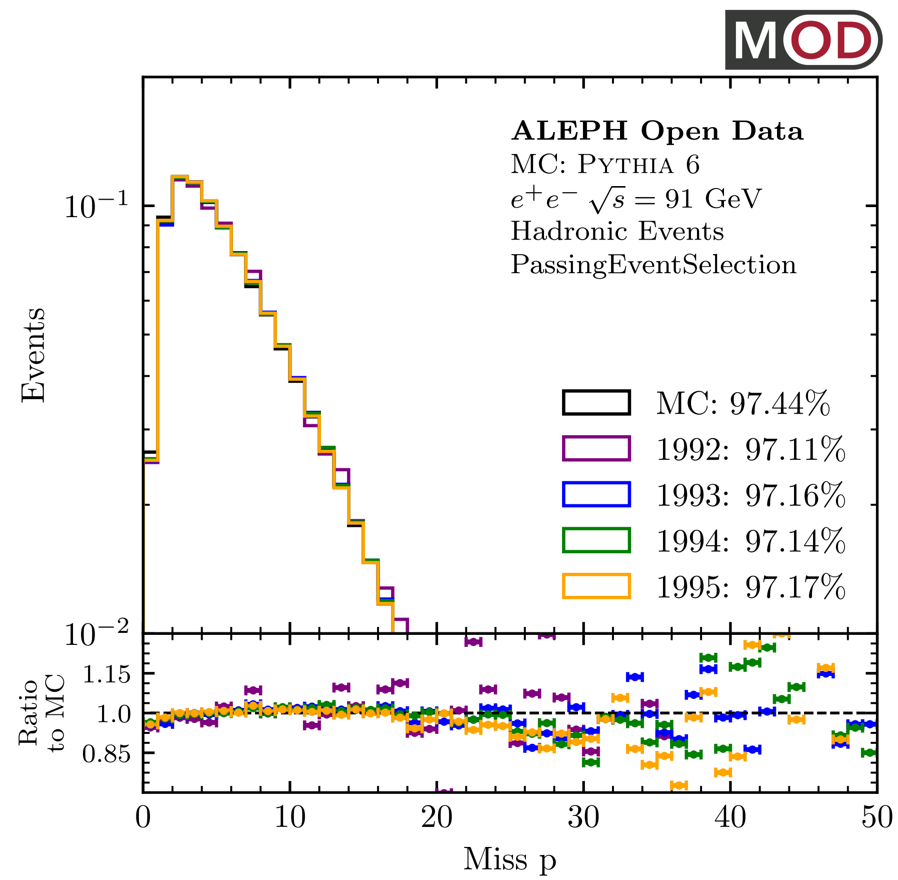
Ratio: 199X/MC

Missing Momentum

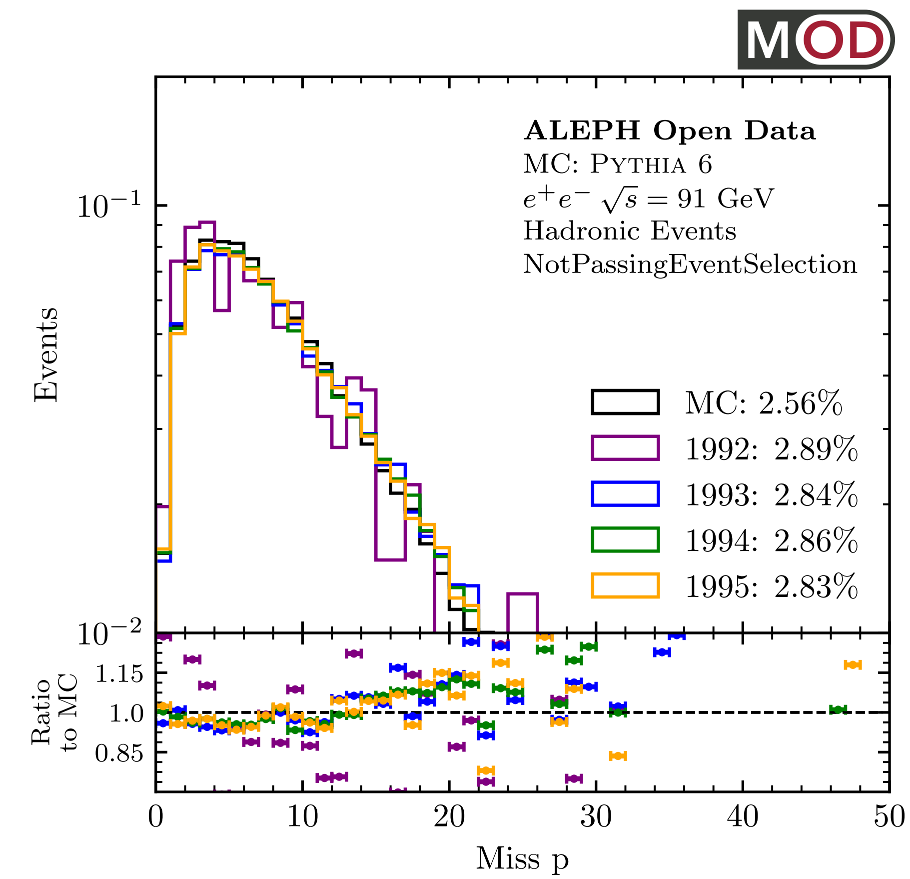
Before Event Selection



Passing Event Selection



Failing Event Selection



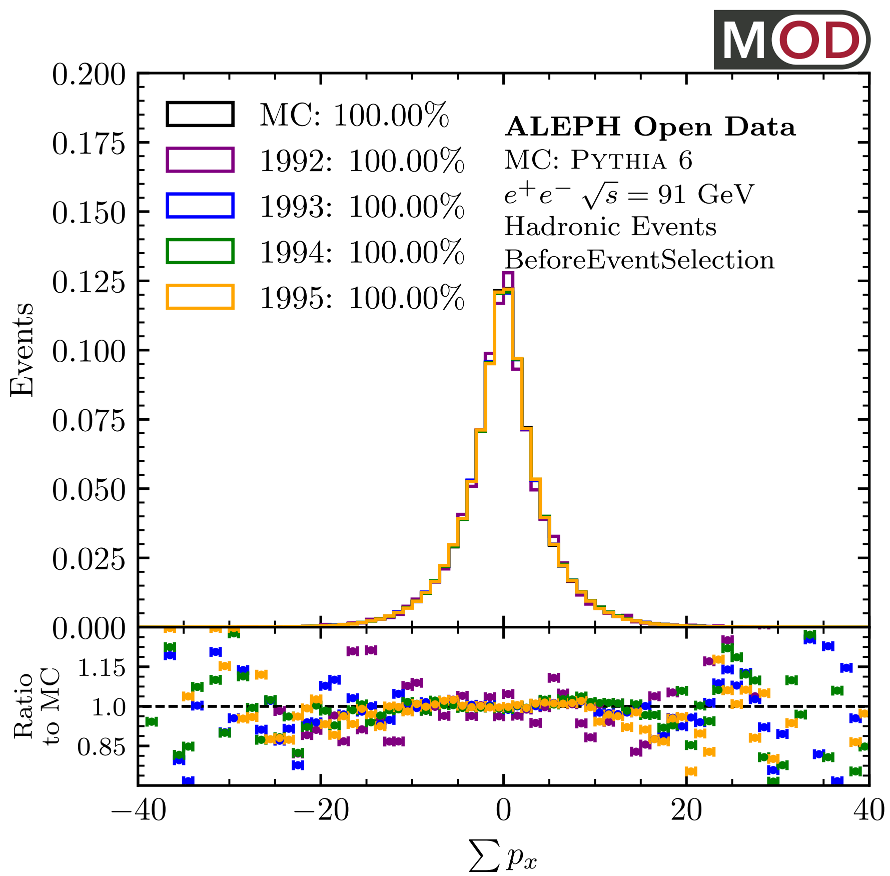
LEP1 1992 1993 1994 1995 : % Events

Pythia 6.1 (1994) MC

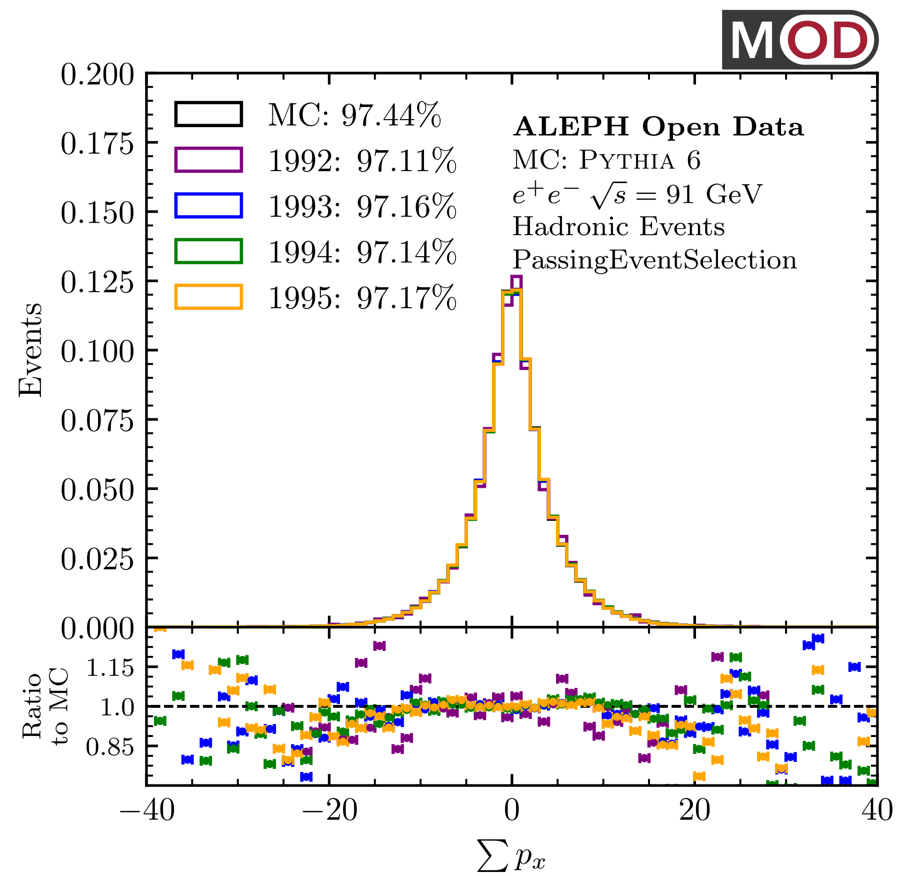
Ratio: 199X/MC

Total Event p_x

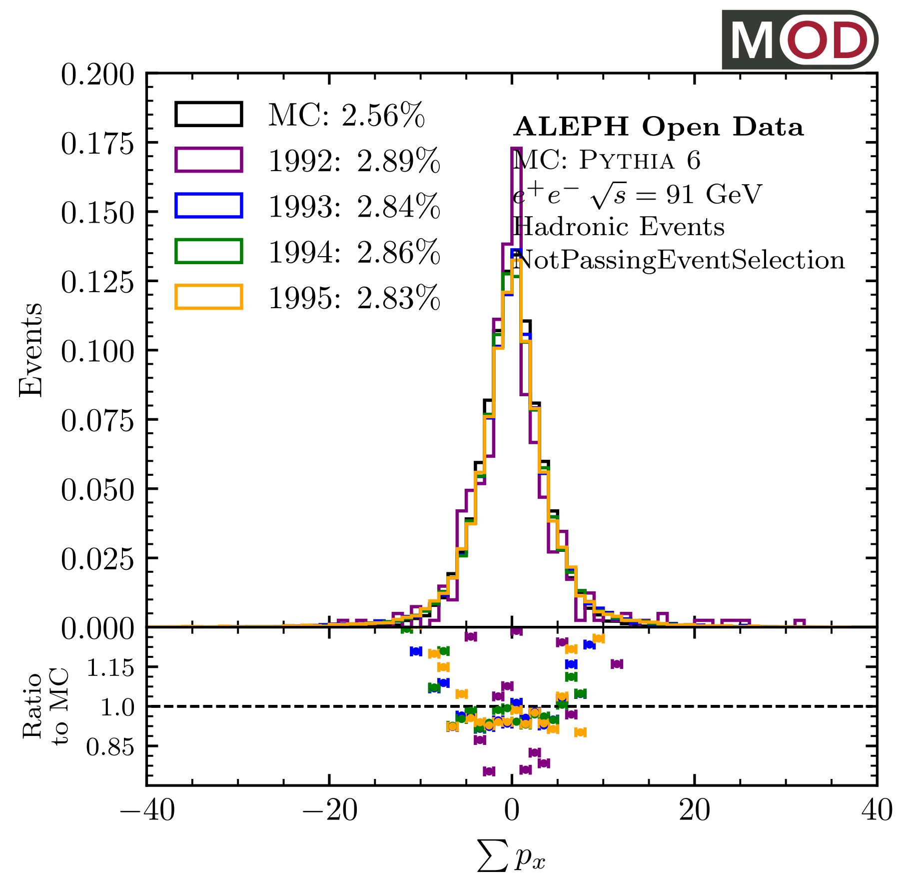
Before Event Selection



Passing Event Selection



Failing Event Selection



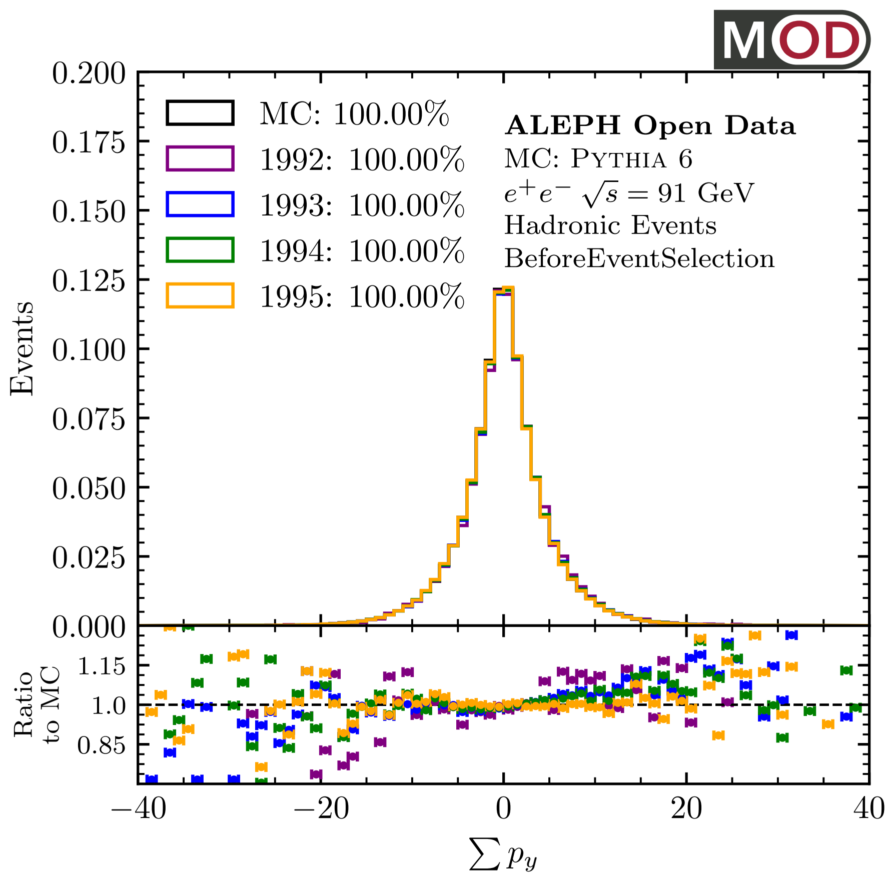
LEP1 1992 1993 1994 1995 : % Events

Pythia 6.1 (1994) MC

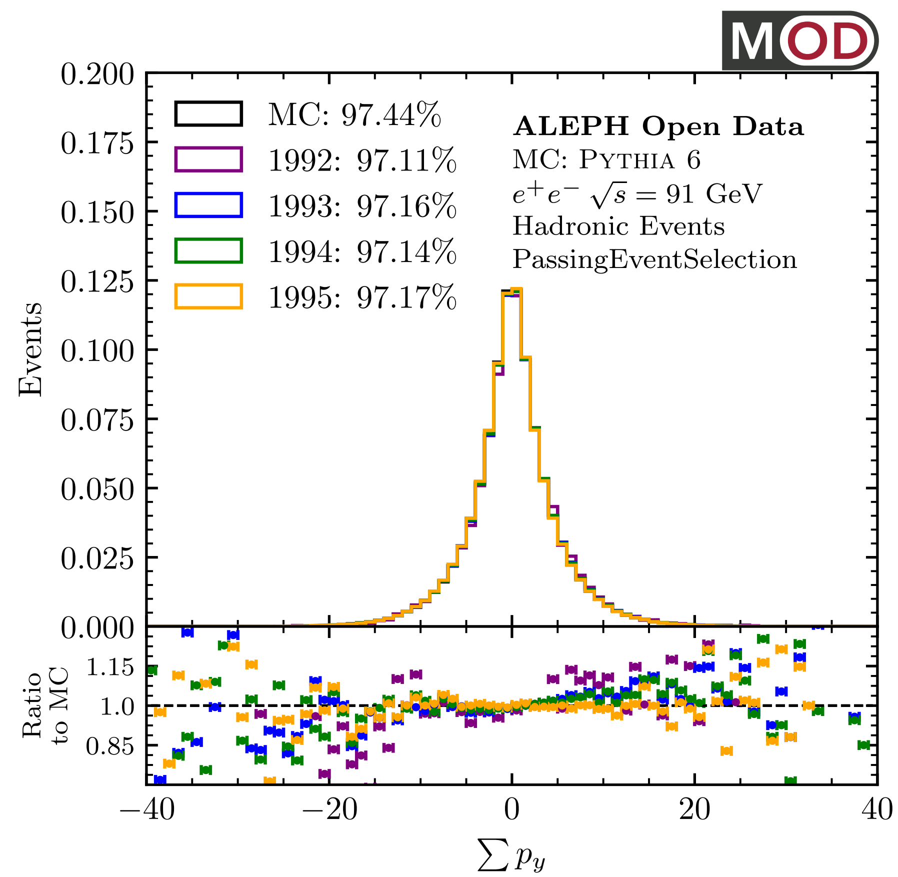
Ratio: 199X/MC

Total Event p_y

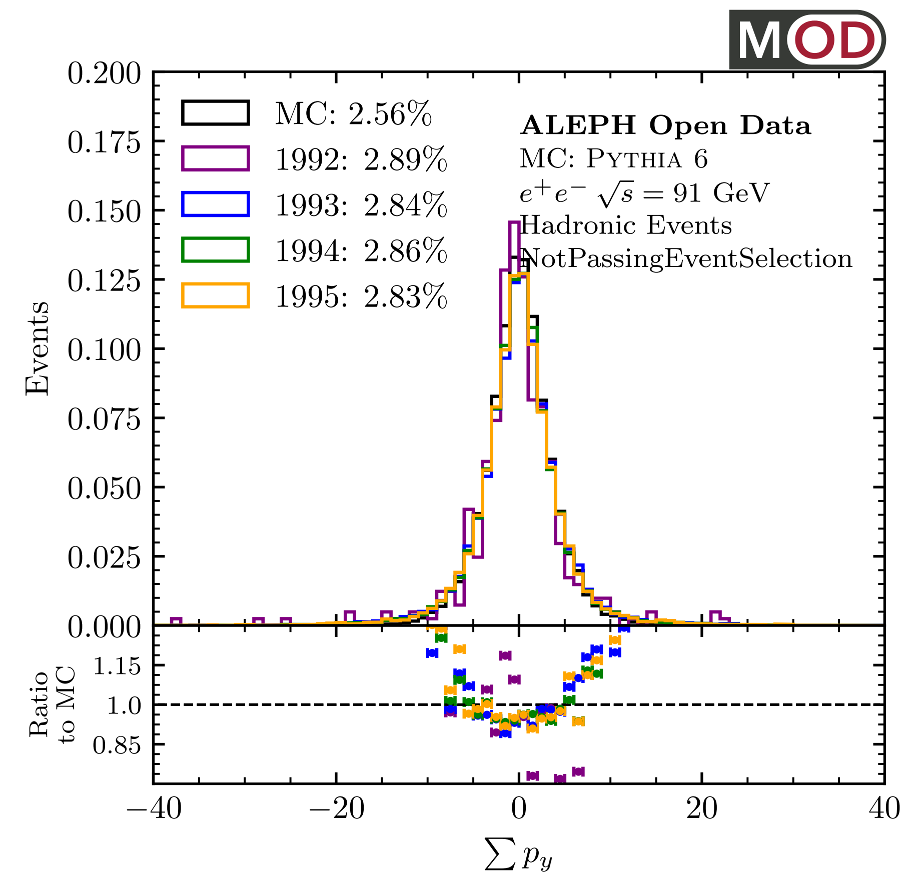
Before Event Selection



Passing Event Selection



Failing Event Selection



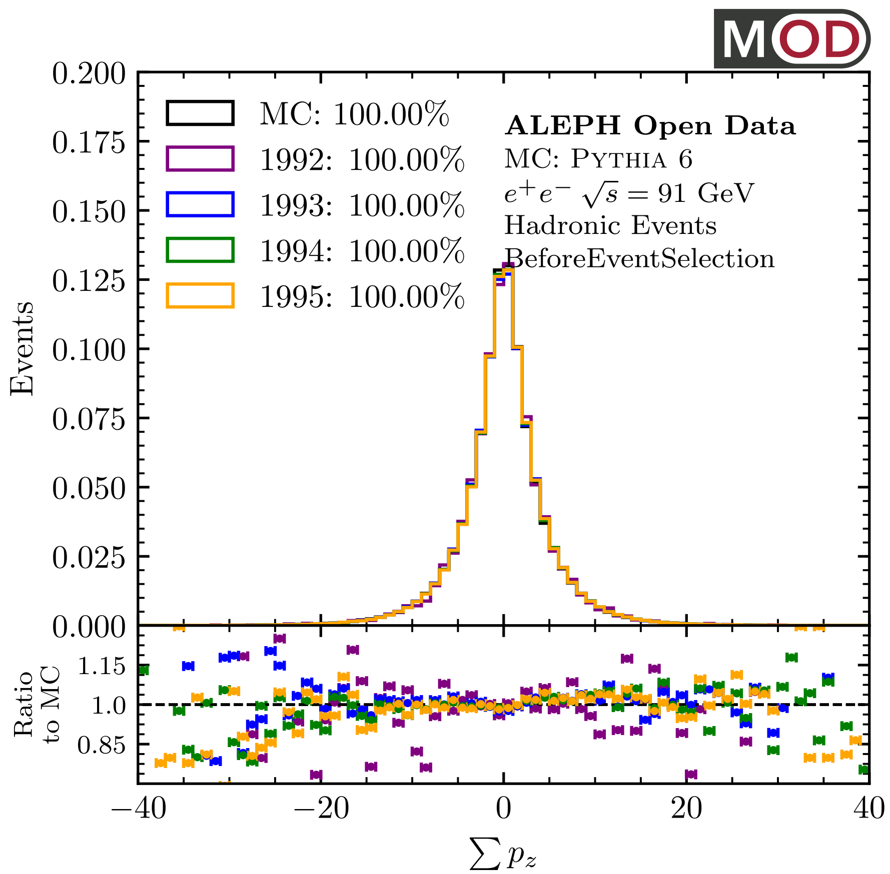
LEP1 1992 1993 1994 1995 : % Events

Pythia 6.1 (1994) MC

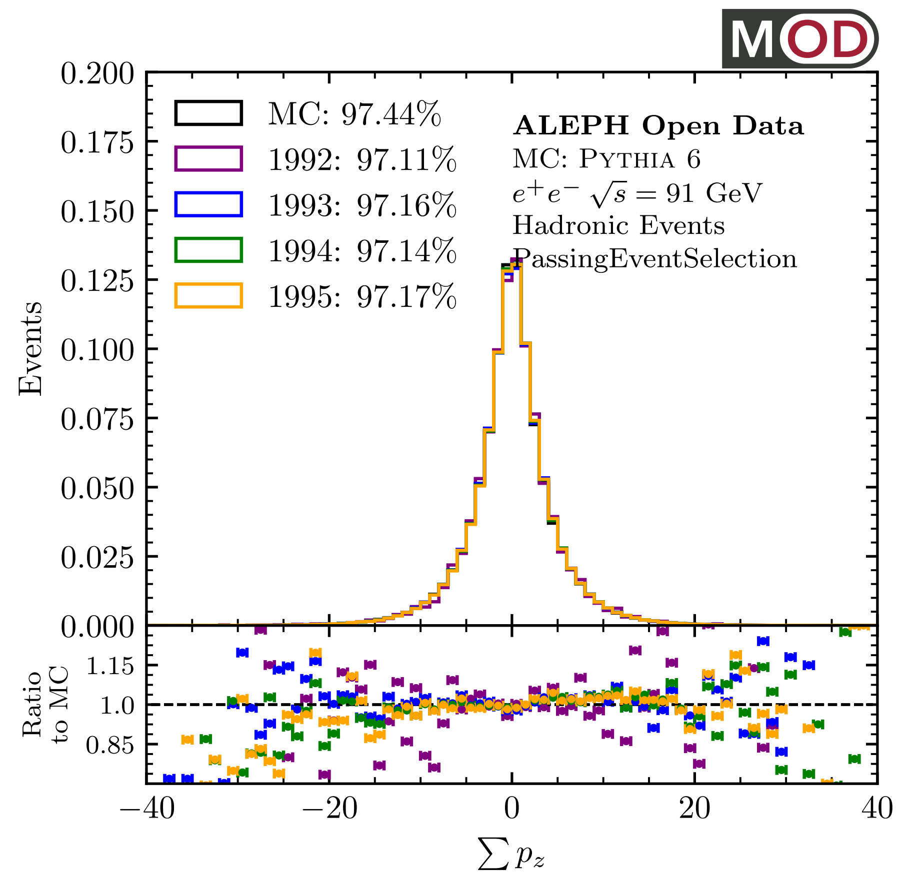
Ratio: 199X/MC

Total Event p_z

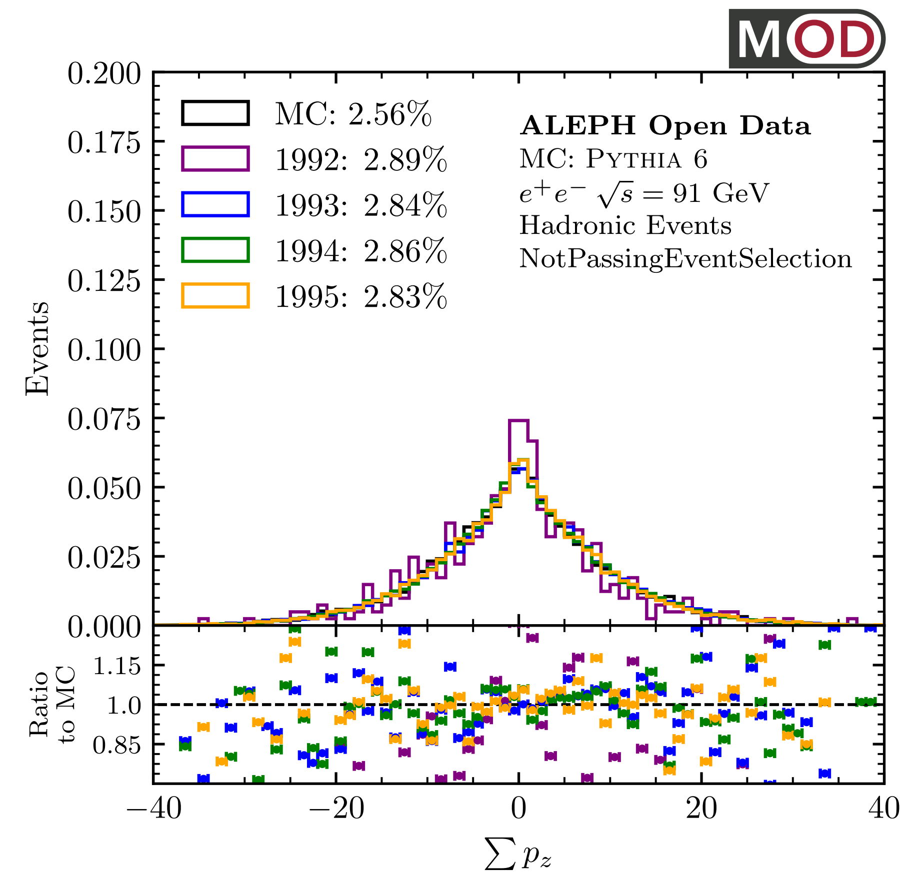
Before Event Selection



Passing Event Selection



Failing Event Selection

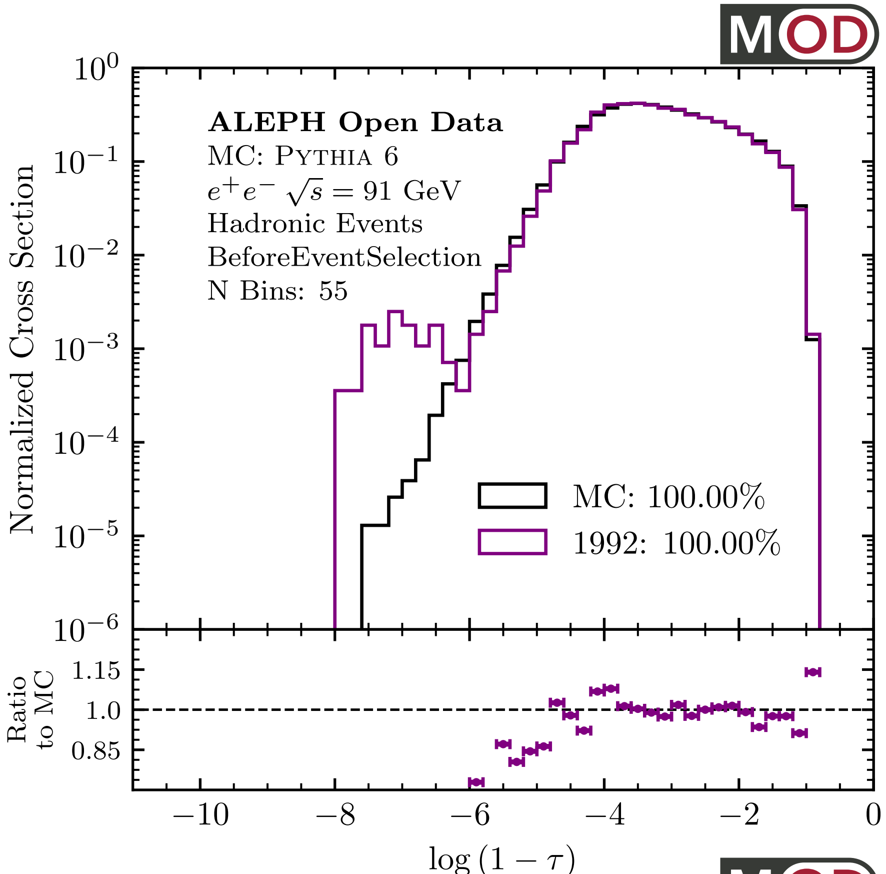


LEP1 1992 1993 1994 1995 : % Events

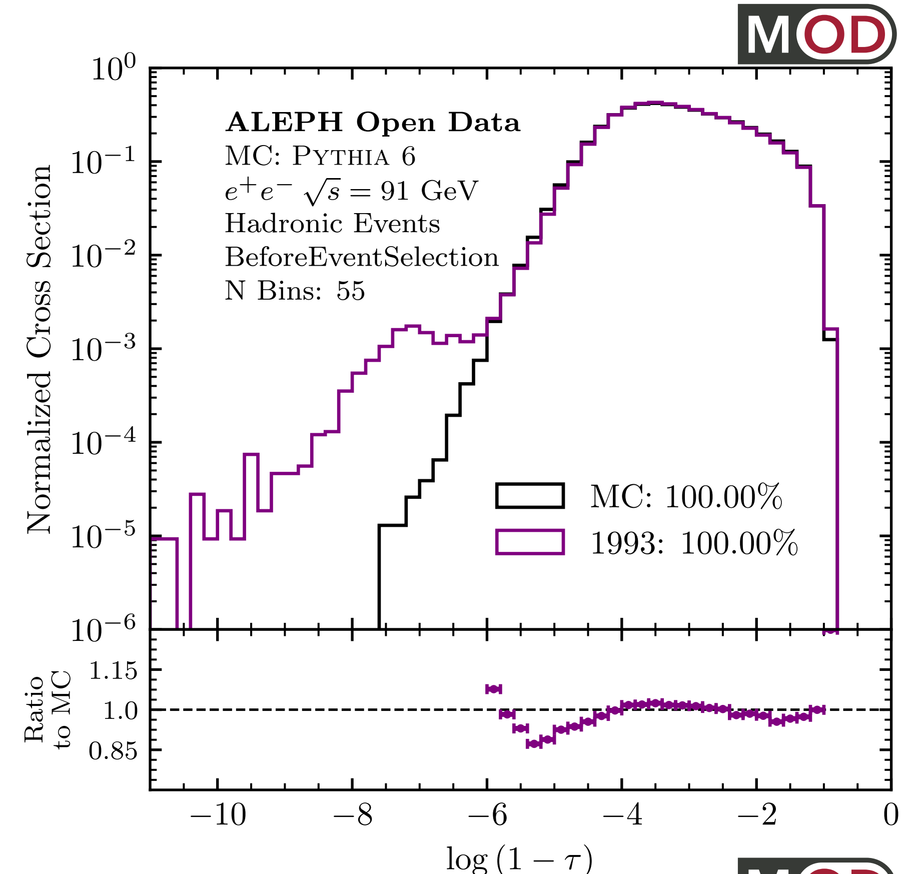
Pythia 6.1 (1994) MC

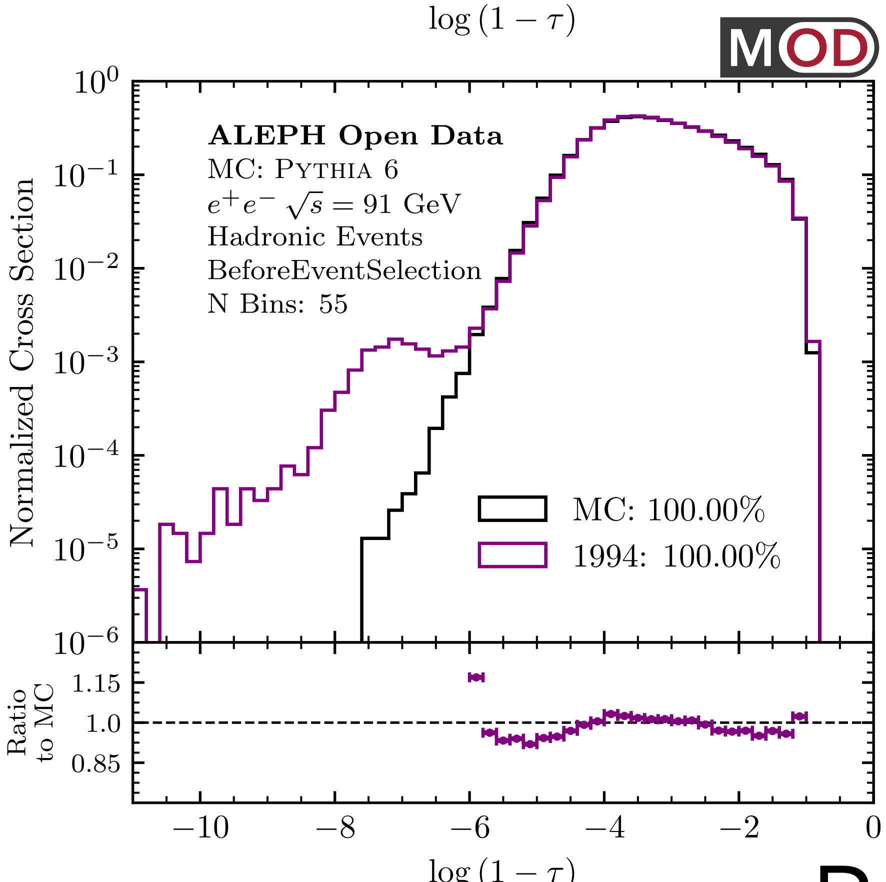
Ratio: 199X/MC

$\log(1-\tau)$, Detector Level, Before Event Selection



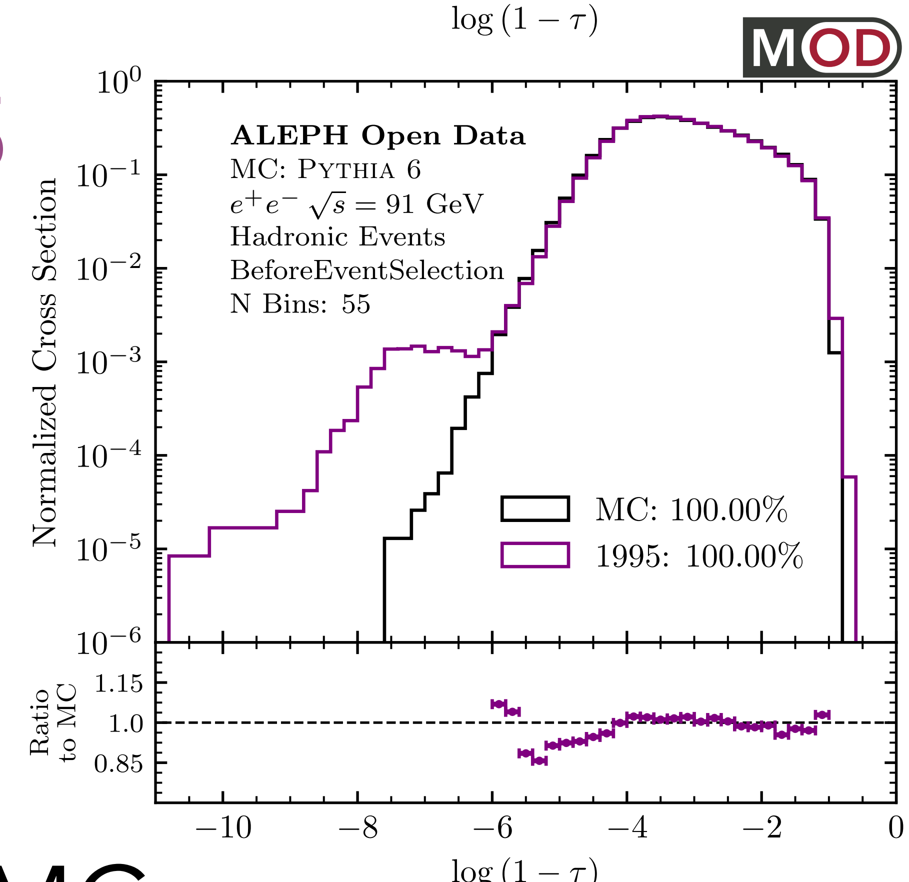
1992 1993





1994 1995

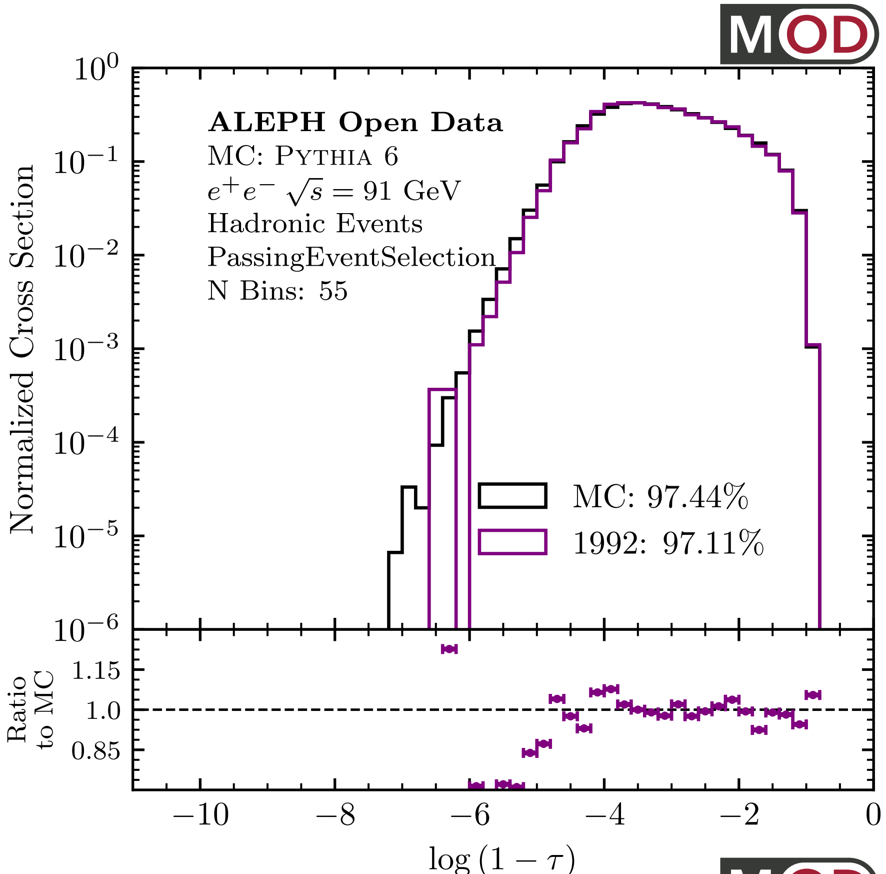
: % Events



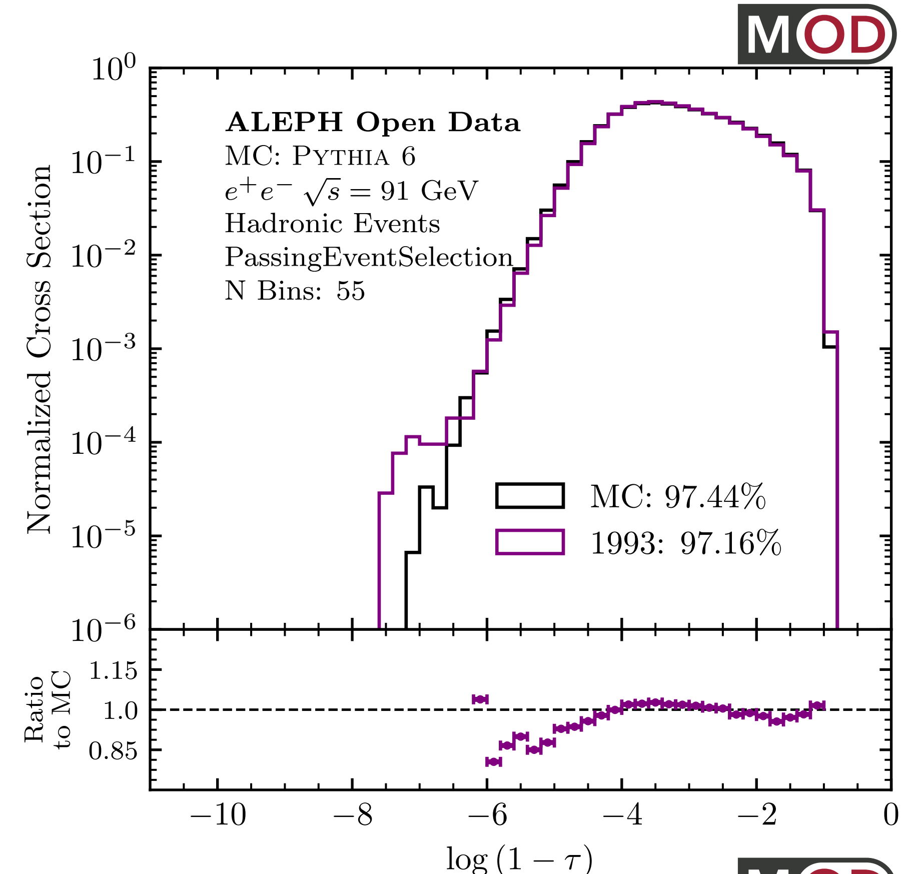
MC

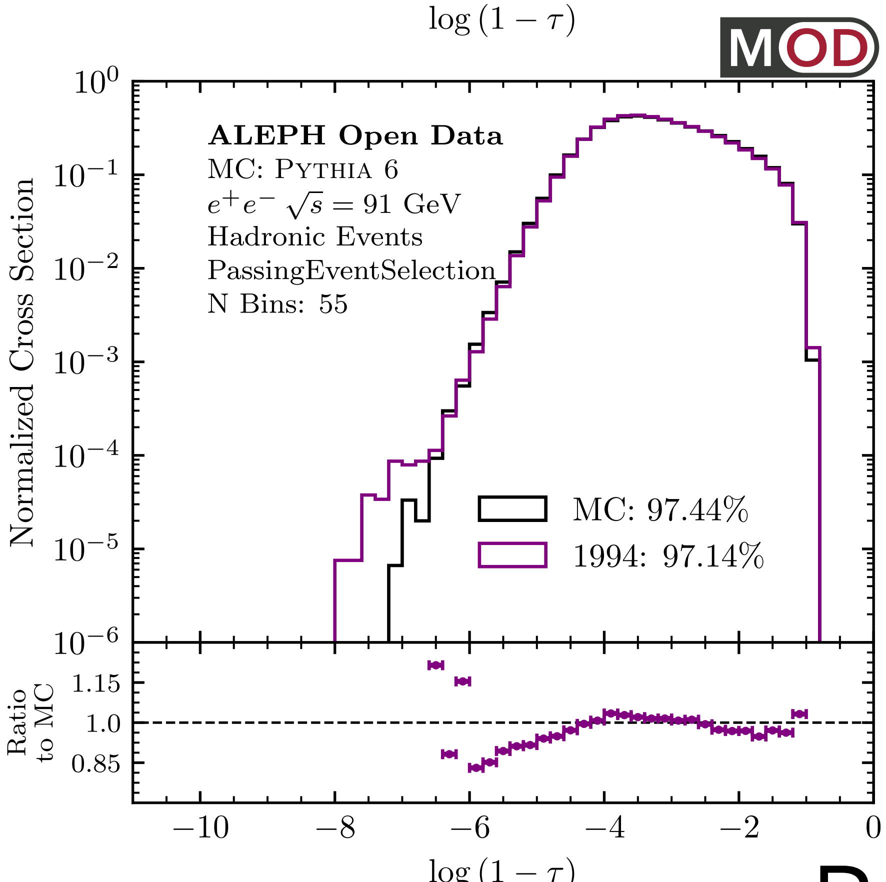
Ratio: 199X/MC

$\log(1-\tau)$, Detector Level, After Event Selection



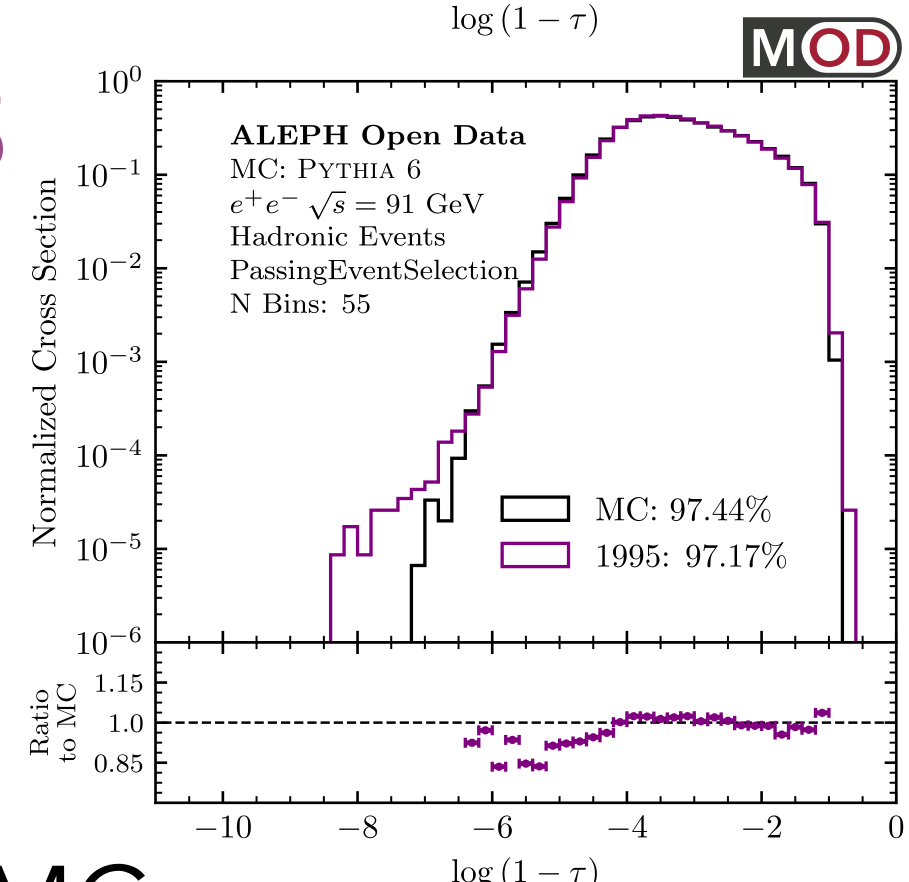
1992 1993





1994 1995

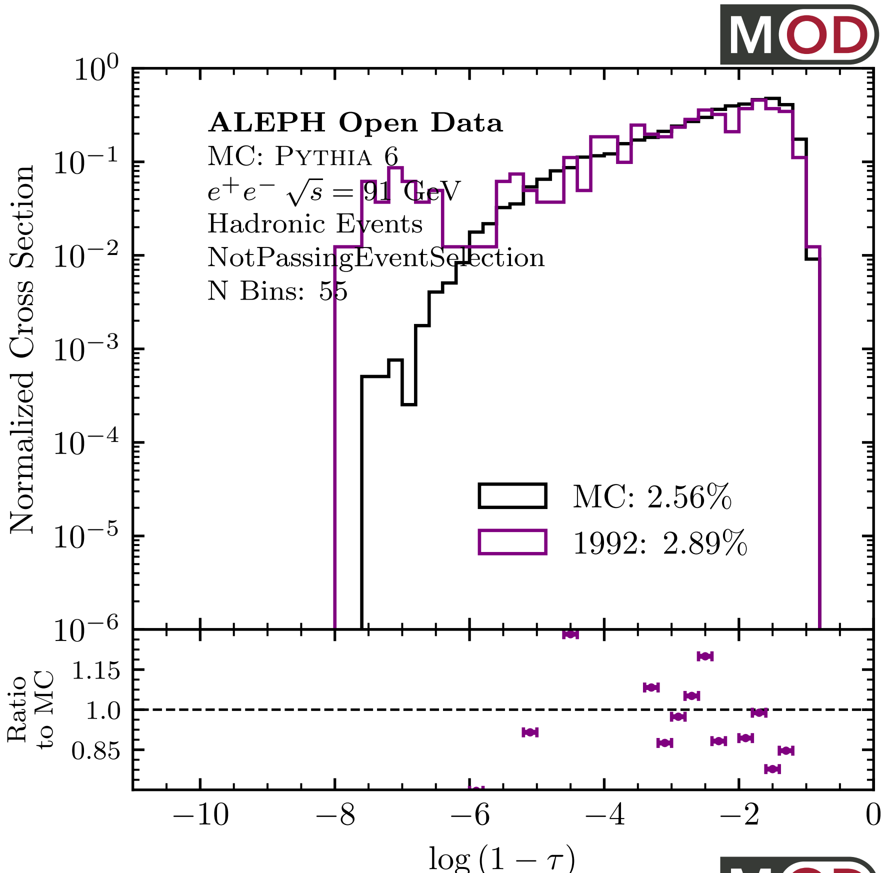
: % Events

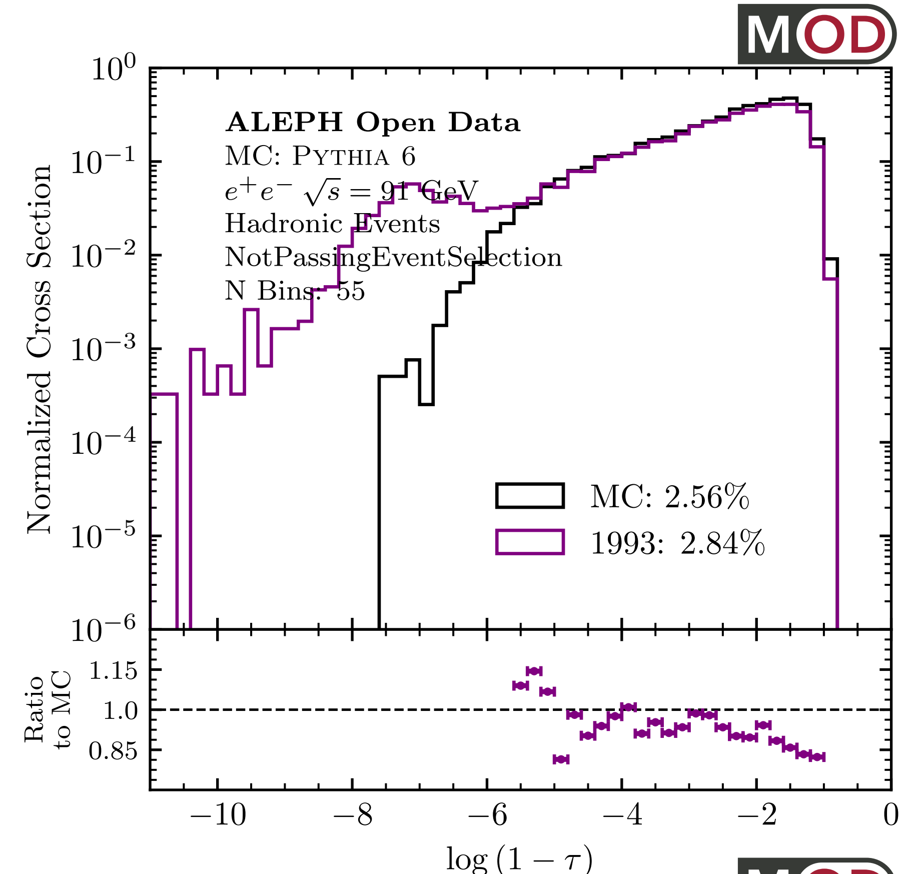


MC

Ratio: 199X/MC

$\log(1-\tau)$, Detector Level, Failing Event Selection

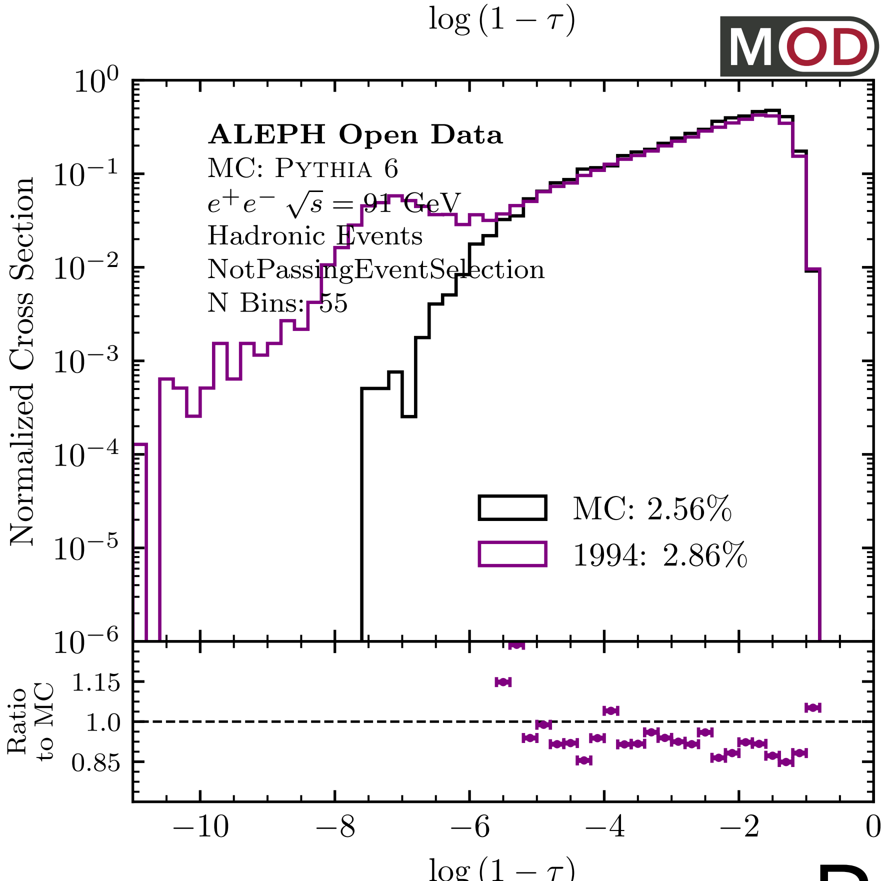


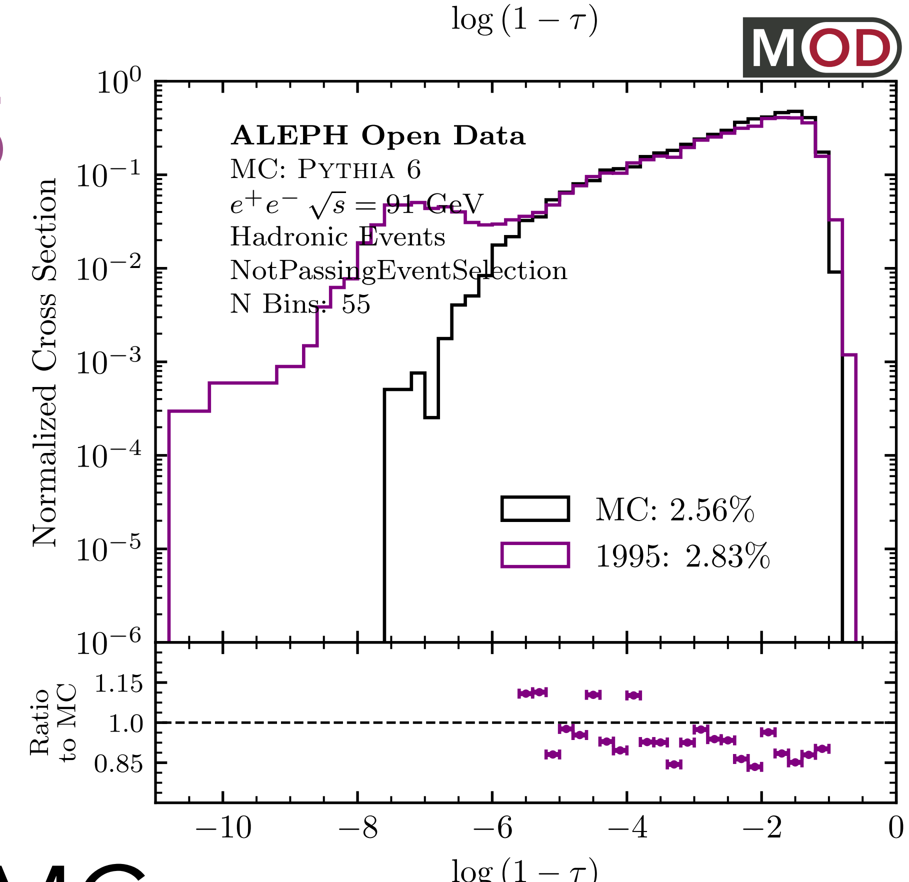


1992 1993

1994 1995

: % Events





MC

Ratio: 199X/MC