

Detector-Level/Reconstruction-Level Analysis

Extracting, Checking and Truth-matching of detector-level objects to validated B-hadron ancestry, producing a stable feature set for future exclusive B_c analysis

Saheed Oyeniran
Date: October 16, 2026

Physics target: $B_c^+ \rightarrow \tau^+ \nu_\tau$

A clean leptonic decay connected to $|V_{cb}|$

Channels under study (previous analysis [2305.02998](#))

- $B^+/B_c^+ \rightarrow \tau^+ \nu_\tau$: larger branching fraction;
- $B^+/B_c^+ \rightarrow \mu^+ \nu_\mu$: helicity suppressed but experimentally clean.

Theory motivation

- CKM extraction independent from exclusive–inclusive tension;
- Potential way to measure B_c hadronisation fraction (no current results);
- Cross check flavour anomalies ($b \rightarrow q \ell \nu$), [1611.06676](#). A new way to test LFU.

Clean probe of the Standard Model

$$\mathcal{B}(B_q^+ \rightarrow \tau^+ \nu_\tau)_{\text{SM}} = \tau_{B_q^+} \frac{G_F^2 |V_{qb}|^2 f_{B_q}^2 m_{B_q} m_\tau^2}{8\pi} \left(1 - \frac{m_\tau^2}{m_{B_q}^2}\right)^2, \quad q = u, c$$

→ Provides theoretically clean determination of $|V_{ub}|$ and $|V_{cb}|$

In the Standard Model, the purely leptonic decay rate is controlled by a small number of inputs:

- Fermi constant G_F
- CKM element $|V_{cb}|$
- decay constant f_{B_c} from lattice QCD
- B_c mass and lifetime
- τ mass and phase-space factor

Overarching Goal: Extraction of CKM Element Matrix $|V_{cb}|$

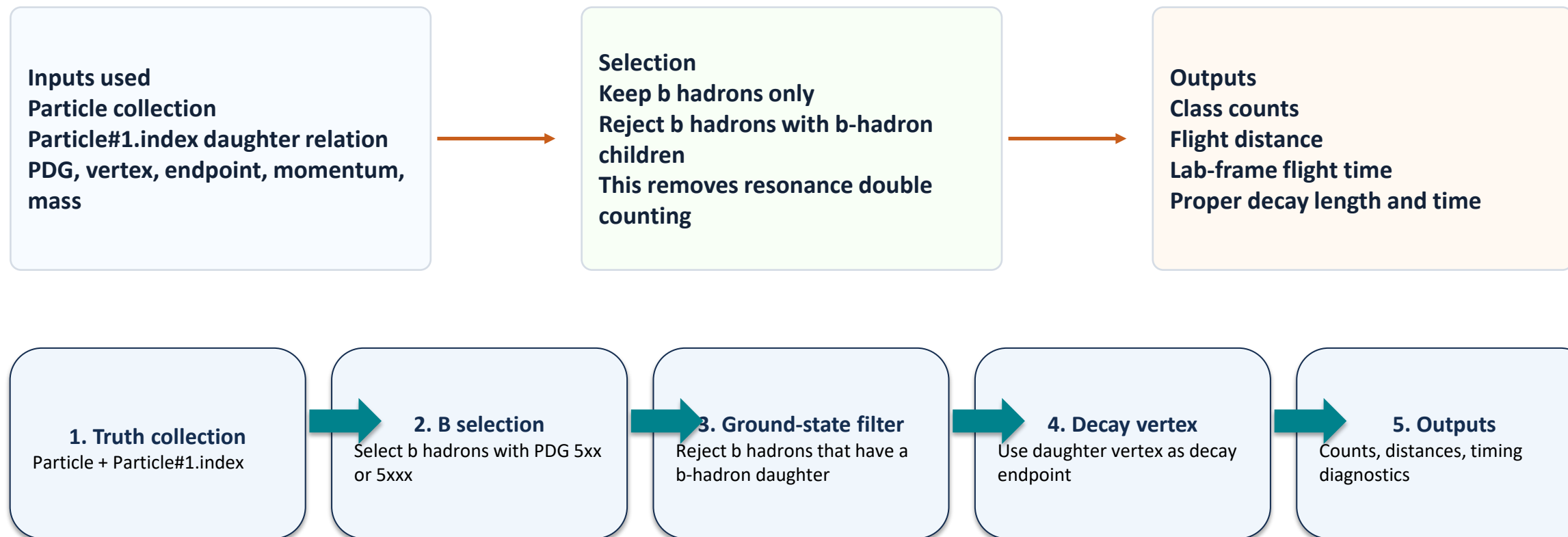
If the B_c branching ratio and production rate are known well enough, $B_c^+ \rightarrow \tau^+ \nu_\tau$ can become an independent handle on $|V_{cb}|$.

Overarching Goal: Estimation of Fragmentation/Hadronization fraction f_{B_c}

$f(b \rightarrow B_c)$ is the probability for a b quark to produce a B_c hadron in the final state.

What the Truth-level analysis did already

A truth-level, generator-particle analysis built around the EDM4HEP MCParticle / GenParticle content.



What question the RECO analysis answers

At generator level, true B-hadron identity is known and truth quantities (such as flight distance, flight time, proper time etc.) are computed.

The reconstruction-level study then asks what detector-level information is available and whether it can be connected back to the same truth classes.

Reconstructed-level objects such as “ReconstructedParticles, jets, MissingET, EFlowTrack, TrackState and MCRecoAssociations” are connected back to the validated truth B-hadron labels.

Question	Answer from current RECO analysis	Why it matters
Do the three stages read the same sample?	Yes: 100 ROOT files and 43,970,699 events in each stage.	Stage 1, 2, and 3 are comparable versions, not different event samples.
Are inclusive RECO observables stable?	mean n_Reco = 47.34, n_RecoCharged = 21.40, n_RecoNeutral = 25.94, n_Jet = 2.128.	Feature additions did not disturb the reconstructed-event audit.
Is Bc preserved at RECO level?	n_Bc = 5,387; n_Reco_fromBc = 61,313.	The rare Bc class is not lost in the detector-level feature construction.
Is this the final Bc fragmentation measurement?	No: this is the detector-level feature-building layer.	Exclusive candidates, efficiencies, backgrounds, and normalization modes remain future steps.

Current status: a validated truth-matched detector-level feature study, not yet an exclusive B_c decay reconstruction.

1. Reconstruction-level analysis flow

Analysis logic

Validates truth B-hadron collection: BHad_idx, BHad_pdg, BHad_class, n_BHad, n_Bc, per-class counts

Extracts inclusive ReconstructedParticle extraction: multiplicity, charge, energy, momentum, pT, η , ϕ , mass, track/cluster links

Event-level jet and missing-energy proxy quantities

MCRecoAssociations: match each reconstructed particle to an MC particle

Truth ancestry tracing: identify whether the matched MC particle descends from a selected B hadron

RECO B-hadron ancestry labels: Reco_BHadClass = 1, 2, ..., 8

Bc-focused variables: n_Reco_fromBc, sumE_Reco_fromBc, maxP_Reco_fromBc, BcEventOnly branches, BcReco_* arrays

Truth B-hadron labels
BHad_class
BHad_idx



Detector-level objects
ReconstructedParticles
Jets / MissingET / Tracks



Reco → MC matching
MCRecoAssociations



Ancestry tracing
Matched MC particle → selected B ancestor



Feature outputs
BcEventOnly
Jet/missing proxies

Role: build the bridge from truth validation to detector-level Bc selection variables.

2. Data sample and analysis flow in the script

The Stage-3 analysis uses the winter2023 IDEA production for p8_ee_Zbb_ecm91. The process list runs 100 chunks with fraction 0.1 and writes the output to BHadronReco_stage3_features.

```
PROCESS_LIST_FULL = {  
    "p8_ee_Zbb_ecm91": {"chunks": 100, "fraction": 0.1}  
}  
outputDirEos =  
"/eos/user/s/soyenira/OUTPUTS/Full_RECO/BHadronReco_stage3_features"
```

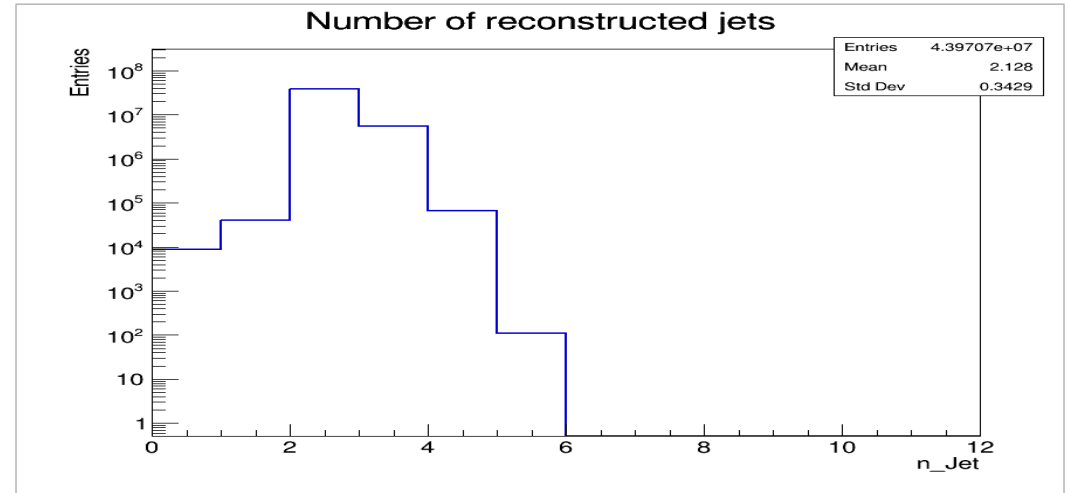
Physics interpretation

The sample is $Z \rightarrow b\bar{b}$ at $\sqrt{s}$ near the Z pole.

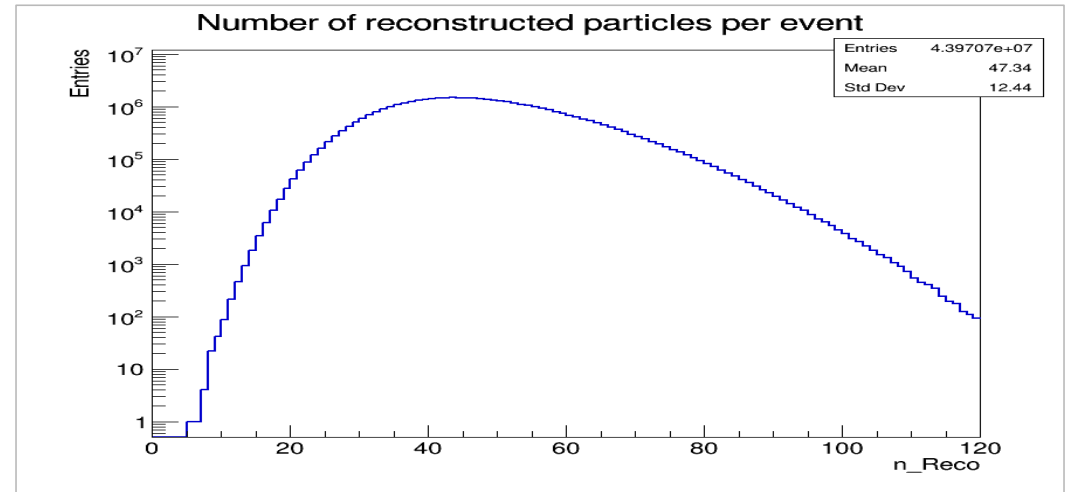
A typical event is expected to be two-jet-like because the b and $\bar{b}$ systems recoil against each other.

This is why $n_{\text{Jet}} \approx 2.13$ and two-leading-jet balance variables are meaningful sanity checks.

n_{Jet} peaks around two reconstructed jets



Stable reconstructed-particle multiplicity



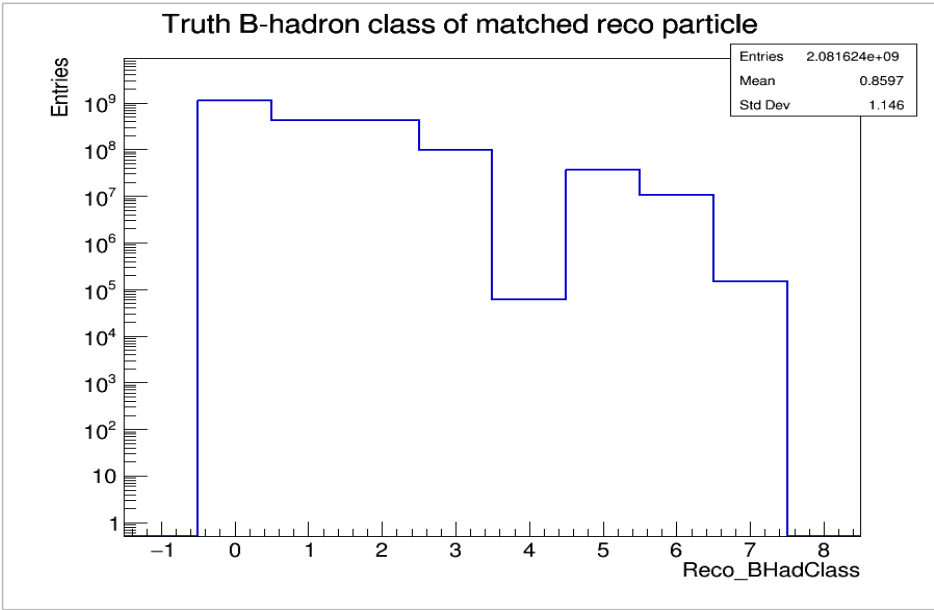
3. Preserving truth B-hadron labels inside the RECO ntuple

The reconstructed-level ntuple reuses the cleaned truth B-hadron definitions. This gives each detector-level object a trustworthy validation label during development.

```
.Alias("Particle0", "Particle#0.index")
.Alias("Particle1", "Particle#1.index")
.Define("BHad_idx", get_ground_bhadron_indices(...))
.Define("BHad_pdg", get_BHadron_pdg_from_indices(...))
.Define("BHad_class", get_BHadron_class_from_indices(...))
.Define("n_Bc", count_BHadron_class(BHad_class, 4))
```

Class	Physics label	Meaning
1	Bd / B0	neutral Bd-family B hadron
2	Bu / B+	charged Bu-family B hadron
3	Bs / Bs0	Bs-family B hadron
4	Bc / Bc+	key rare Bc class
5–7	Λ_b , Ξ_b , Ω_b	selected b-baryon classes
8	Other b baryon	catch-all b-baryon class

Truth B-hadron class of matched reco particle



Ground-state filtering prevents excited states or intermediate resonances from being double counted as weakly decaying B hadrons.

4. How ReconstructedParticle information is extracted

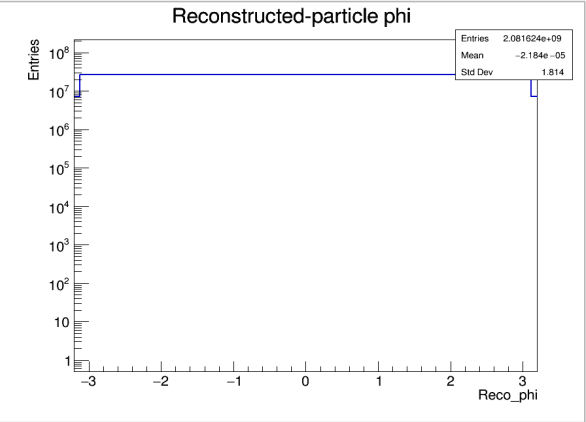
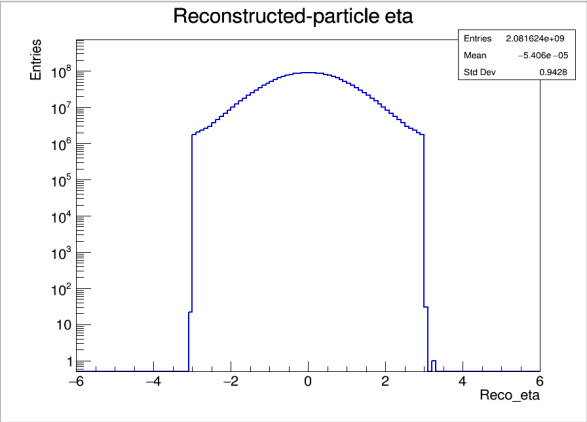
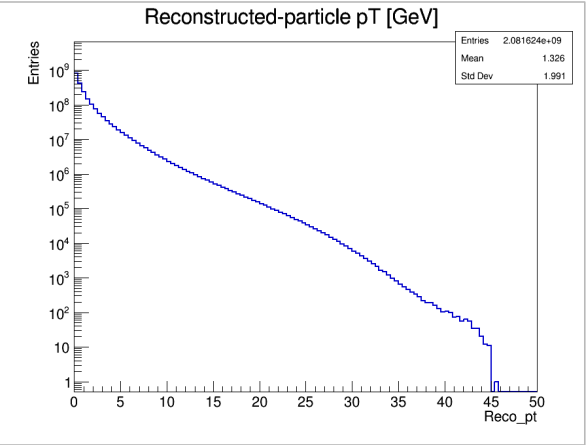
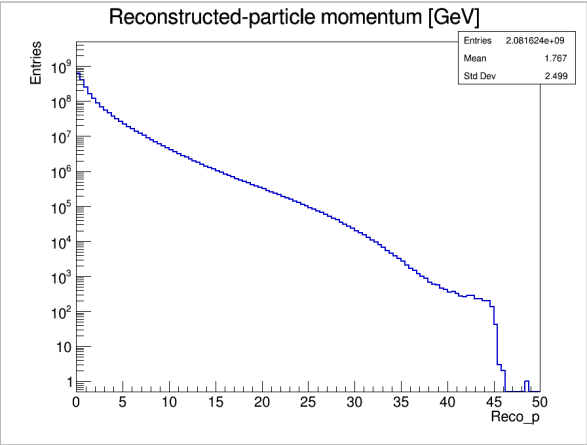
The ReconstructedParticle collection contains the detector-level particle flow objects.

The extracted Inclusive reconstructed quantities include: number of reconstructed particles, charged and neutral counts, energy, momentum components, total momentum, transverse momentum, pseudorapidity, azimuthal angle, charge, mass and track/cluster link counts.

These inclusive quantities are not intended to isolate B_c decays by themselves. They only show that the detector-level object collection is populated and physically shaped.

Quantity	Formula	Meaning
p	$\sqrt{p_x^2 + p_y^2 + p_z^2}$	overall momentum of a particle-flow object
p_T	$\sqrt{p_x^2 + p_y^2}$	momentum transverse to beam
η	$\frac{1}{2} \ln \frac{(p + p_z)}{(p - p_z)}$	angular coordinate relative to beam axis, central events cluster near $\eta = 0$
ϕ	$\tan^{-1} 2(p_y, p_x)$	azimuthal angle around beam axis, should be approximately uniform for an unbiased sample

```
.Define("n_Reco", get_n_reco(ReconstructedParticles))
.Define("n_RecoCharged", get_n_charged(ReconstructedParticles))
.Define("n_RecoNeutral", get_n_neutral(ReconstructedParticles))
.Define("Reco_e", reco_e(ReconstructedParticles))
.Define("Reco_p", reco_p(ReconstructedParticles))
.Define("Reco_pt", reco_pt(ReconstructedParticles))
.Define("Reco_eta", reco_eta(ReconstructedParticles))
.Define("Reco_phi", reco_phi(ReconstructedParticles))
```



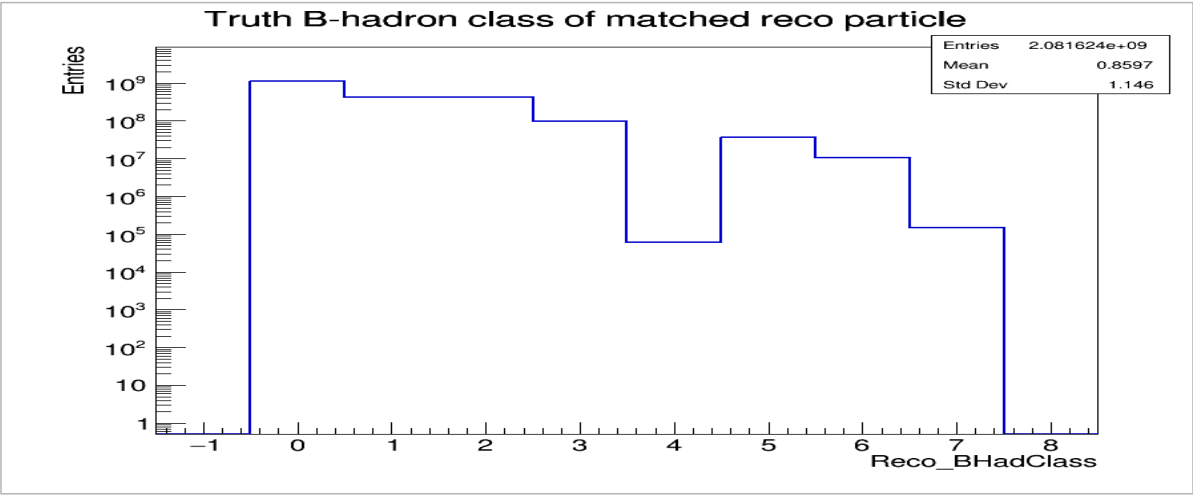
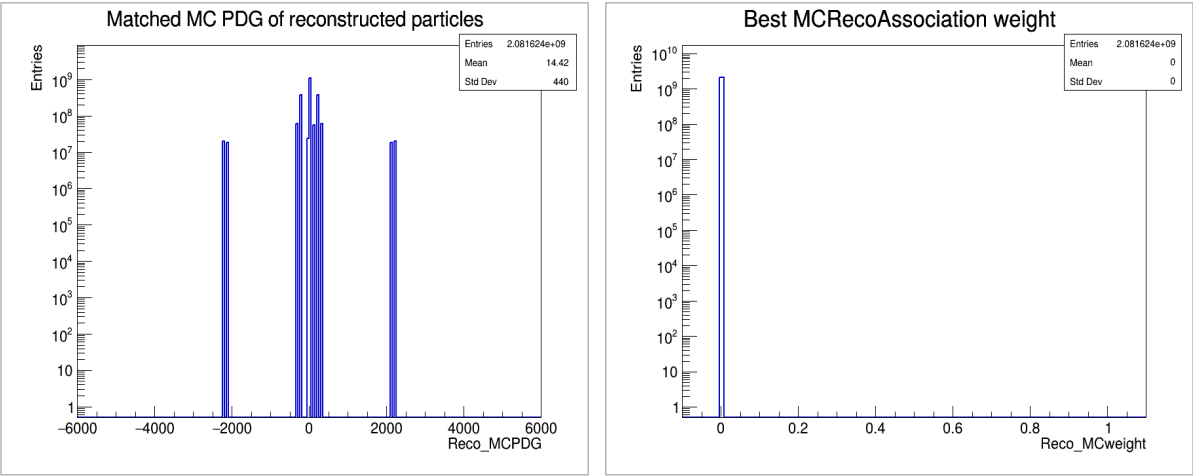
5. Reco-to-truth matching: why the default direction is kept

The crucial step is to match each reconstructed particle to an MC particle, then trace the matched MC particle upward to its selected B-hadron ancestor.

Reco_BHadClass tells which selected truth B-hadron ancestor a reconstructed particle comes from. It's a truth labelling tool for validation.

```
.Alias("MCRecoAssociations0", "MCRecoAssociations#0.index")
.Alias("MCRecoAssociations1", "MCRecoAssociations#1.index")
.Define("Reco_MCidx", get_reco_to_mc_best_index(...))
.Define("Reco_MCPDG", get_mc_pdg_from_indices(...))
.Define("Reco_BHadClass", get_reco_bhadron_class(...))
```

Branch	Matched entries	Interpretation
Reco_BHadClass	975,533,504	Default direction; use for physics-facing interpretation
Reco_BHadClass_alt	5,772,742	Alternate direction; debugging cross-check only
Reco_MCweight	mean = 0	Do not use as matching-quality variable here



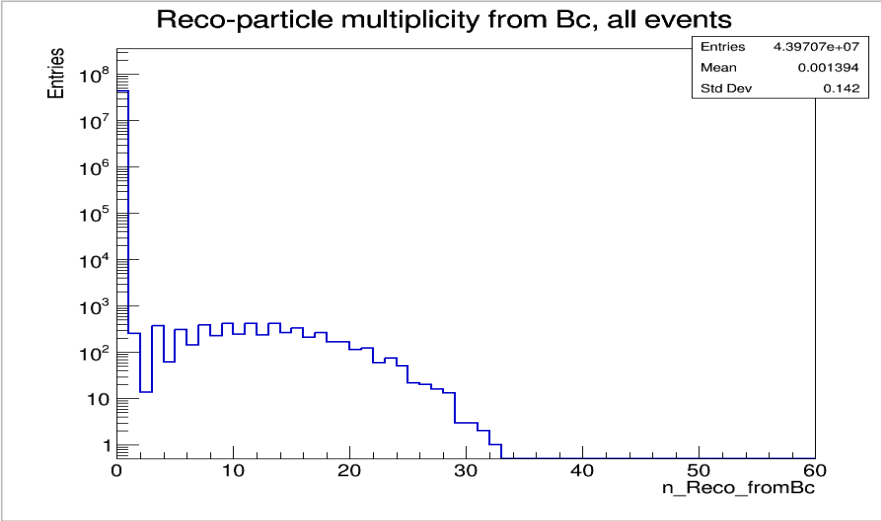
6. The core diagnostic numbers mean

The diagnostics are the numerical backbone: they verify output completeness, sample consistency, truth-label survival, and non-zero B_c content.

Diagnostic	Value	Meaning
ROOT files found	100	expected full output set
Tree entries	43,970,699	event-level sample size
n_Reco total	2.08162×10^9	total reconstructed particles
mean n_Reco	47.3412	average reco particles per event
mean n_RecoCharged	21.3994	charged PFO content
mean n_RecoNeutral	25.9418	neutral PFO content
mean n_Jet	2.12781	mostly two-jet $Z \rightarrow b\bar{b}$ topology
n_BHad total	8.82471×10^7	selected ground-state B-hadron count
n_Bc evt_hasBc	5,387 5,387	selected B_c preserved; one selected B_c per B_c event

Only 5,387 of 43,970,699 events have a selected B_c .
Therefore, all-event B_c histograms are expected to have a huge zero bin.

All-event B_c multiplicity: zero-dominated by rarity

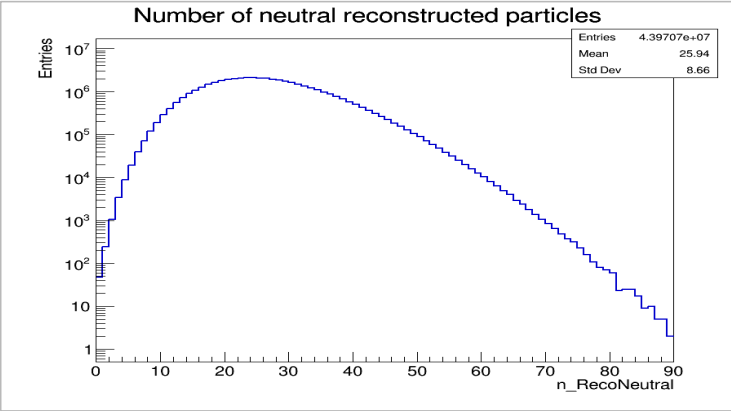
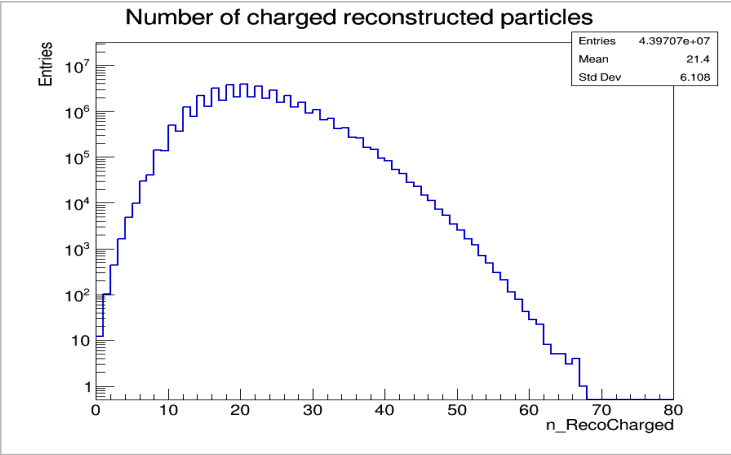
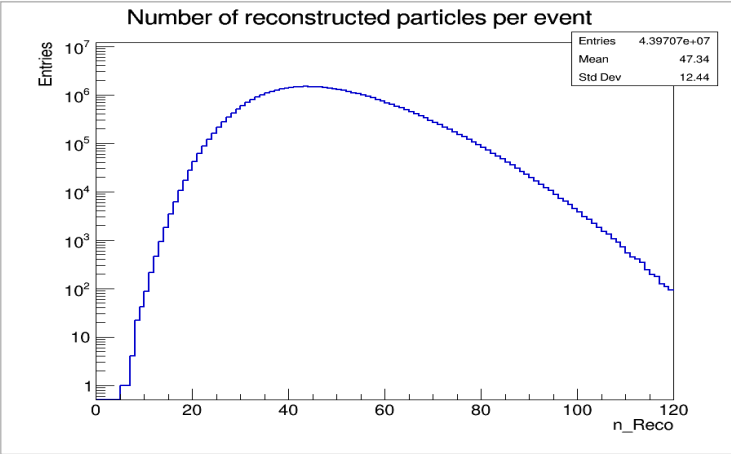


7. Physics interpretation of inclusive RECO plots

Inclusive plots are detector-level sanity checks, not Bc-signal selections.

They show that the reconstructed event content is physically shaped and stable, as expected.

Plot group	Observed behavior	Physics / analysis meaning
n_Reco n_RecoCharged n_RecoNeutral	Mean around 47.34 Mean around 21.40 Mean around 25.94 Broad multiplicity distributions	Hadronic $Z \rightarrow b\bar{b}$ events produce many PFOs; charged/neutral content is stable. The broad distribution reflects different hadronization and decay multiplicities.
Reco_e Reco_p Reco_pT	Steeply falling spectra	Most reconstructed hadrons are soft; few carry high energy/momentum.
Reco_eta, Reco_phi	centered near $\eta = 0$ ϕ nearly uniform	Central, symmetric sample with no obvious one-sided acceptance bias. Expected for an unbiased azimuthal detector/event sample
Reco_charge	Centered around neutral aggregate with charged state visible	Shows the mixture of charges and neutral reconstructed particles
Reco_nTracks	mostly 0 or 1	neutral PFOs have no track; charged PFOs usually have one.
Reco_MCPDG	Peaks at common final-state particle PDG	The matched MC content is dominated by ordinary final state particles from hadronization and decays
Reco_MCWeight	Concentrated at zero	Doesn't look informative and never used in the analysis
n_Jet	Mean around 2.128	Consistent with a two-jet $Z \rightarrow b\bar{b}$ topology, with extra jets or splitting in some events



11. Why the three stages exist and how to explain them

The stages are development versions, not a required sequential chain. Each reads the original RECO sample and writes its own output. Stage 3 is the current baseline for future work.

Stage	Role	Main result	Caveat / lesson
Stage 1	Baseline reconstructed-object audit	Confirmed ReconstructedParticle extraction, jets, MissingET, tracks, truth matching, and nonzero Bc ancestry.	Some Bc summary variables had placeholder/sentinel effects.
Stage 2	Cleaned control layer	Preserved stable means, made n_OtherBbaryon explicit, separated Bc controls, and fixed maxP_Reco_fromBc.	Mostly controls; not yet the final feature set.
Stage 3	Physics-feature layer	Added BcEventOnly variables, Bc-matched energy/momentum, charged/neutral Bc counts, jet ratios/asymmetry, and missing-energy proxies.	Still truth-matched; not an exclusive Bc reconstruction.

Stage 1 proved branch access; Stage 2 cleaned controls; Stage 3 produced the feature set for Bc-focused comparisons.

Stage 1: baseline reconstructed-object audit

What Stage 1 established

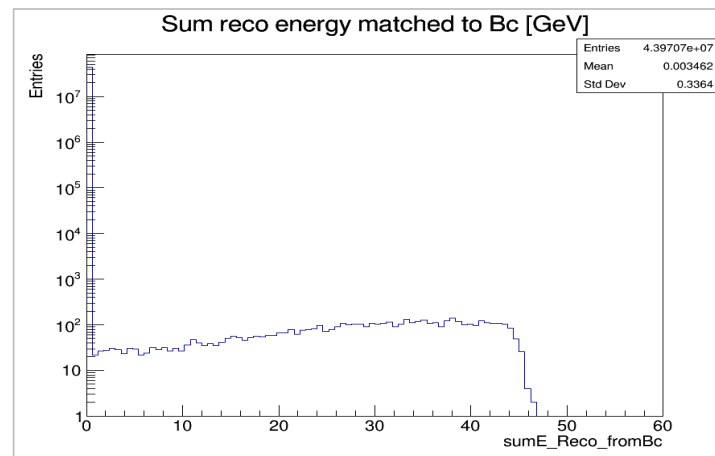
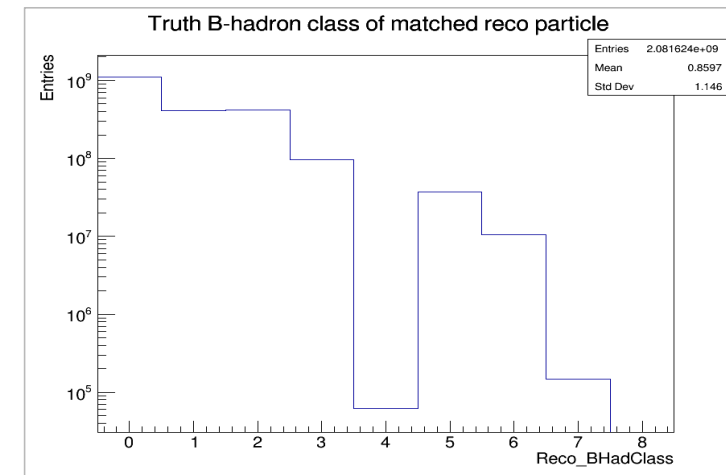
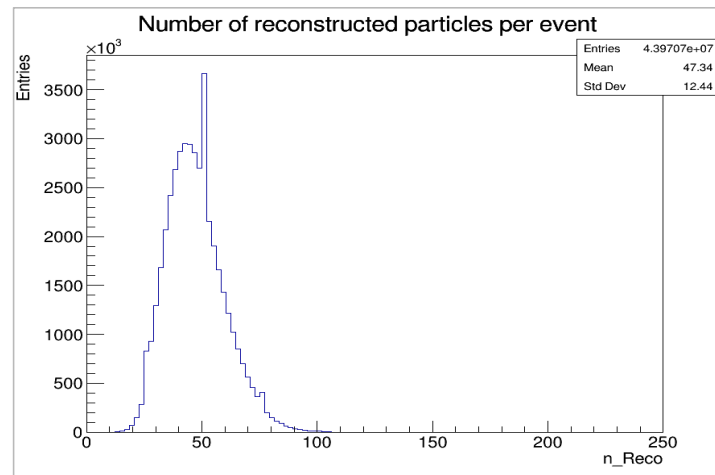
100 ROOT files and 43,970,699 events processed.

ReconstructedParticles, jets, MissingET, tracks, and truth matching were accessible.

Default MCReco association direction produced large matched population.

Bc ancestry was nonzero: $n_Reco_fromBc = 61,313$.

Known limitation: some Bc summary variables retained placeholder/sentinel effects.

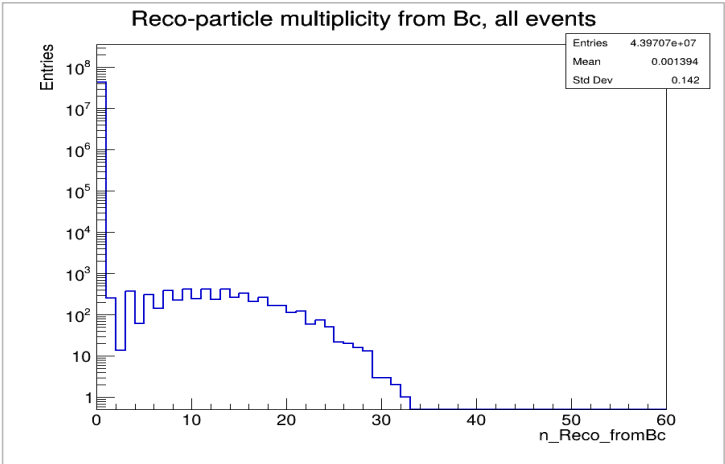
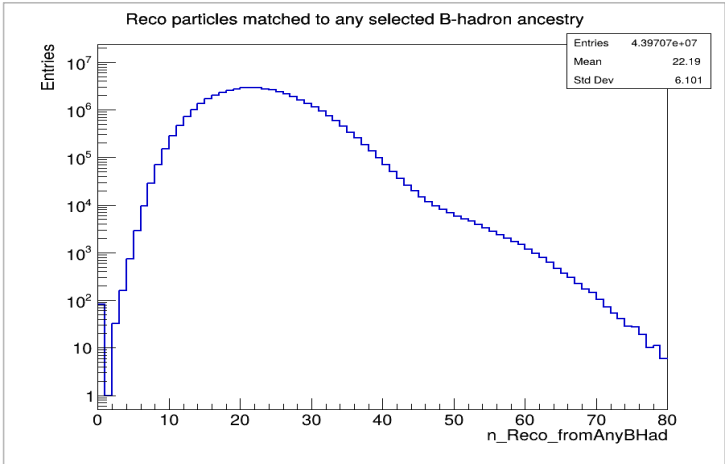


Stage 1 is best described as proof that the RECO object layer and B-hadron ancestry propagation work.

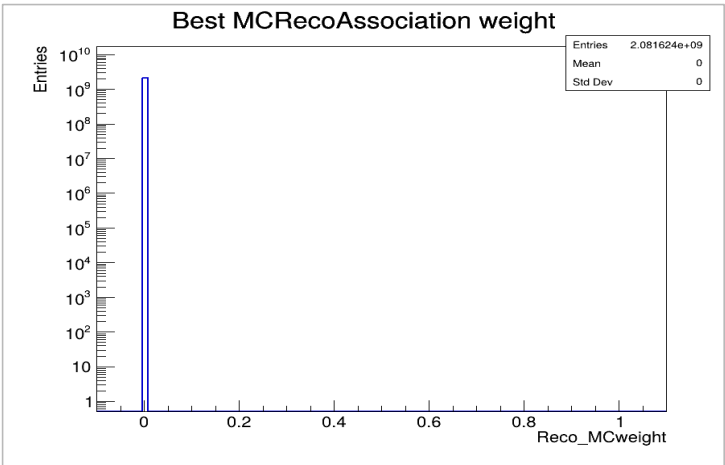
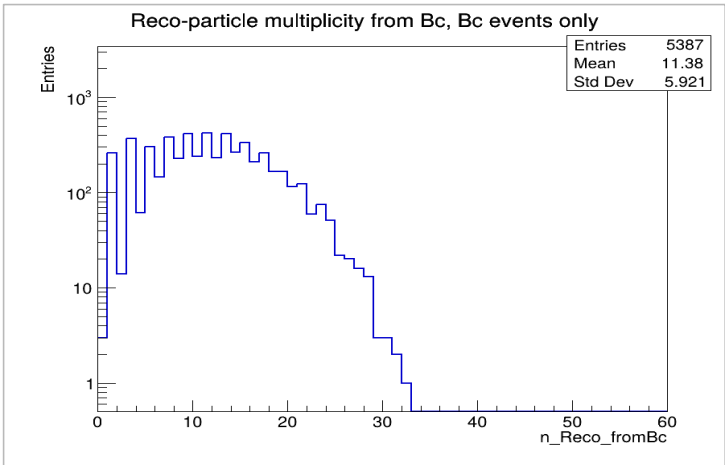
Stage 2: cleaned control layer and Bc-control distributions

Purpose of Stage 2

- Clean and harmonize baseline observables.
- Keep the same truth/RECO content and stable means.
- Make `n_OtherBbaryon` explicit as 0.
- Separate Bc-control plots from all-event dilution.



Quantity	Stage-2 result
<code>n_Reco_fromBc</code>	61,313
<code>n_Reco_fromBc / n_Bc</code>	11.38
<code>maxP_Reco_fromBc</code>	48,577.2
<code>n_OtherBbaryon</code>	0



Stage 3: physics-feature layer

Stage 3 preserves the validated common quantities and adds compact Bc-sensitive variables, BcEventOnly branches, and jet/missing-energy proxies.

Feature family	Key quantities	Use
Bc ancestry counts	n_Reco_fromBc, nCharged_Reco_fromBc, nNeutral_Reco_fromBc	Describe visible Bc detector footprint.
Bc energy/momentum	sumE_Reco_fromBc, sumP_Reco_fromBc, maxP_Reco_fromBc	Estimate visible energy and leading-particle scale.
BcEventOnly branches	n_Reco_fromBc_BcEventsOnly, sumE..., maxP..., mass proxy	Avoid dilution by zero-Bc events.
Jet proxies	Jet_Emax, Jet_Emin2, Jet_Eratio, Jet_Easym	Two-leading-jet balance controls.
Missing-energy proxies	Jet_Emissing2, Reco_missing_e_calc	Feature candidates for modes with neutrinos/tau decays.

Stage-3 diagnostic	Value
evt_hasBc	5,387
n_Reco_fromBc	61,313
reco particles / selected Bc	11.3817
sumE_Reco_fromBc / n_Bc	28.257 GeV
maxP_Reco_fromBc / n_Bc	9.017 GeV

Stage 3 would be used as the baseline reconstructed-level ntuple for the next analysis layer.

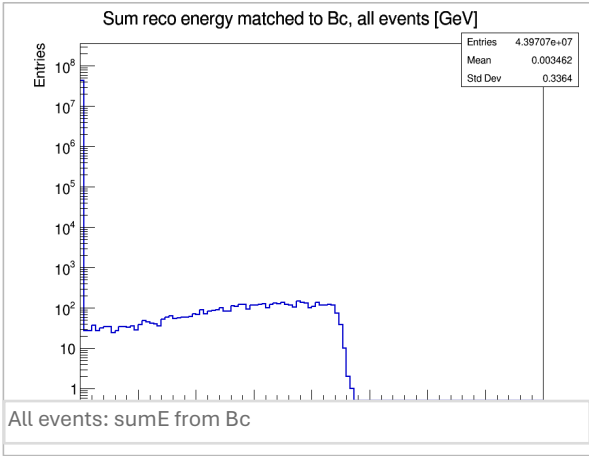
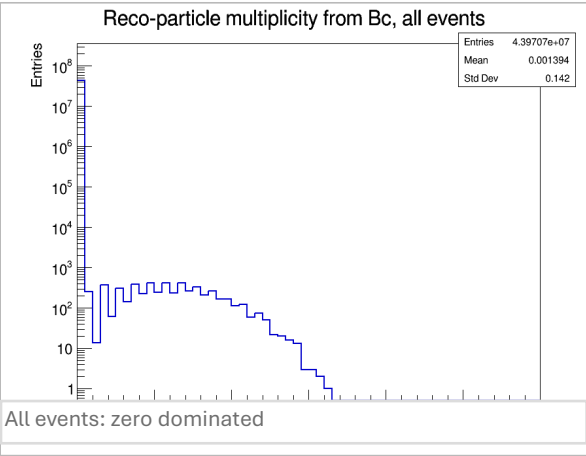
8. Physics interpretation of B_c -focused observables

These variables ask:

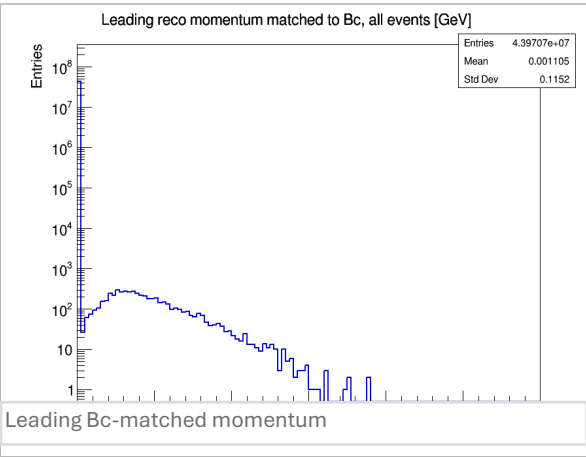
“among reconstructed particles in an event, which have MC ancestry tracing back to a selected B_c ?“

This is not yet exclusive B_c reconstruction; it is a truth-matched description of the detector-level B_c footprint.

They tell the approximate visible particles multiplicity and energy scale associated with Bc ancestry , which could become inputs for later stage selections.



Bc diagnostic	Value	Meaning
n_Reco_fromBc	61.313	total B_c -matched reco particles
n_Reco_fromBc/n_Bc	11.3817	Mean B_c -matched reco particle multiplicity per selected B_c event
nCharged_Reco_fromBc nNeutral_Reco_fromBc	30.178 31.135	Charged B_c footprint neutral B_c footprint
sumE_Reco_fromBc/n_Bc	28.257 GeV	mean visible B_c -matched energy
sumP_Reco_fromBc/n_Bc	28.159 GeV	mean B_c -matched total momentum scale
maxP_Reco_fromBc/n_Bc	9.017 GeV	mean leading B_c -matched momentum
mass proxy/n_Bc	4.960 GeV	Invariant mass-like proxy from all B_c -matched reco particles, not a reconstructed B_c mass peak



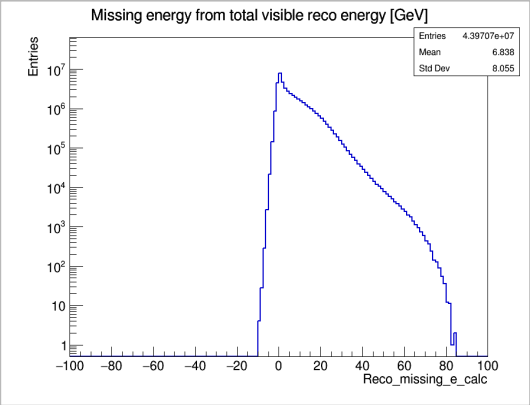
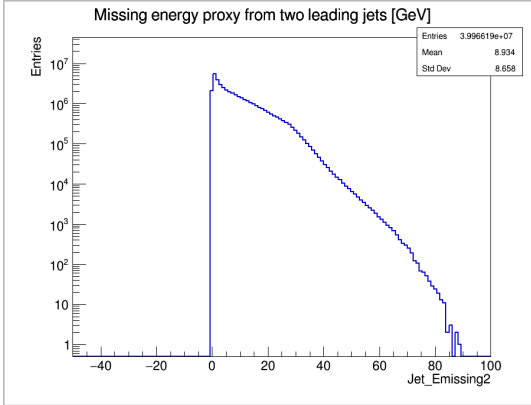
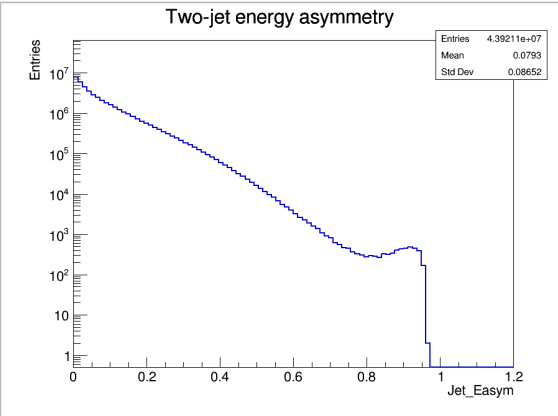
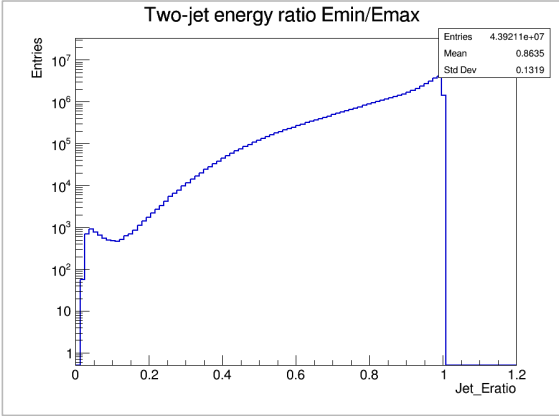
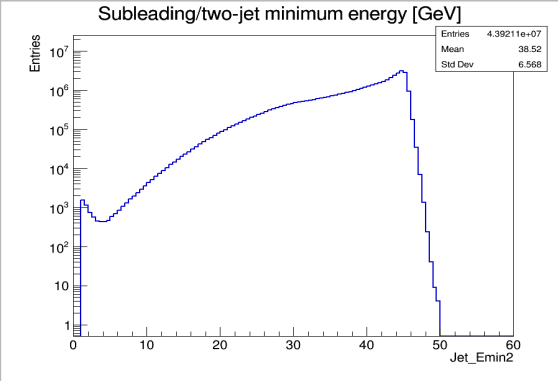
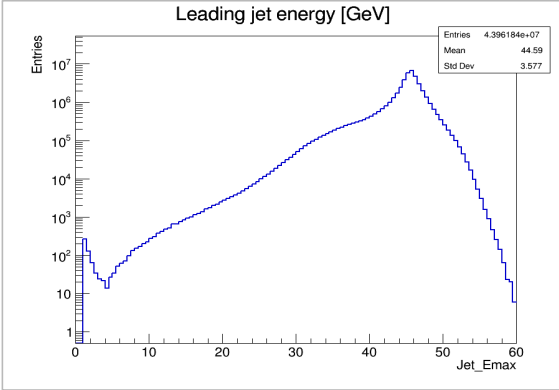
Use BcEventOnly quantities for Bc-specific interpretation because all-event distributions are dominated by zero-Bc events.

9. Physics interpretation of jet and missing-energy proxies

These variables are useful because B_c decays with neutrinos producing missing energy patterns.

They are useful controls and feature candidates, not yet full thrust/hemisphere observables.

Variable	Definition	Mean value	Interpretation
Jet_Emax	highest jet energy	44.5829 GeV	leading jet near Z-pole half-energy scale
Jet_Emin2	lower of two leading jets	38.4719 GeV	subleading side of energy balance
Jet_Eratio	E_{min}/E_{max}	0.861448	near 1 means balanced dijet topology
Jet_Easym	$(E_{max}-E_{min})/(E_{max}+E_{min})$	0.0780861	small value means balanced leading jets
Jet_Emissing2	$\sqrt{s}-(E_{max}+E_{min})$	8.0542 GeV	energy not captured by two leading jets
Reco_missing_e_scale	$\sqrt{s}$ -sum of visible reco energy	6.83776 GeV	Visible energy-based missing energy-proxy from all reconstructed particles



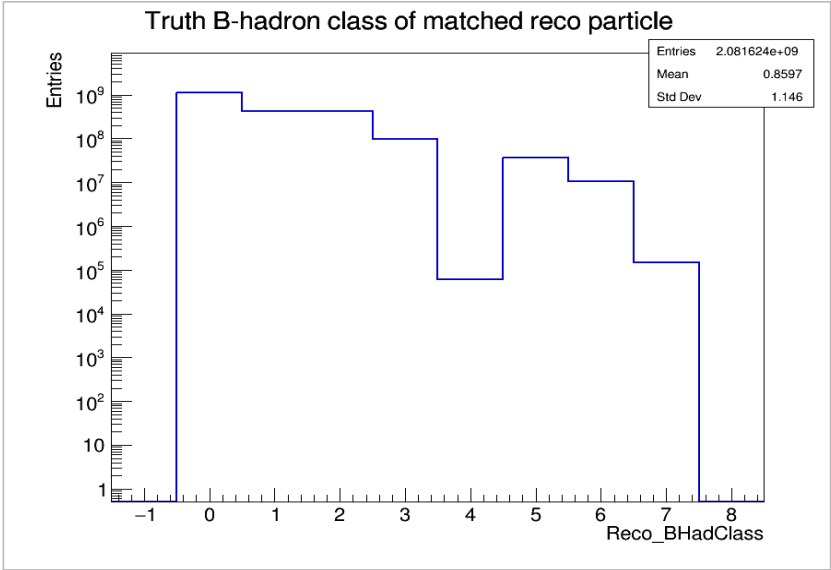
10. Reco-to-truth class content: what the class counts tell us

Class counts show how many reconstructed particles are assigned to each selected B-hadron ancestry.

The hierarchy mirrors the truth population: Bd/Bu dominate, Bs is smaller, Bc is rare but nonzero, and b-baryons appear at lower rates. This is as expected from the truth-matched RECO classes.

In the reconstruction stage, B_c is still rare but the ratio of B_c reco to B_c truth is close/comparable to other B-hadron classes. This shows that B_c events can be studied at detector level, even though they're statistically rare.

Class	Truth count	Reco matched	Reco/truth	Meaning
Bd/B0	37,845,248	412,884,880	10.91	dominant B-meson class
Bu/B+	37,862,515	417,820,288	11.04	dominant B-meson class
Bs	8,421,257	96,858,396	11.50	lower truth yield, similar reco scale
Bc	5,387	61,313	11.38	rare but preserved
Λ b	3,228,775	37,286,182	11.55	b-baryon class
Ξ b	872,176	10,475,131	12.01	smaller b-baryon class
Ω b	10,868	147,314	13.55	rare b-baryon; limited stats



Sanity check: B_c is rare in production, but once selected it has a reconstructed-particle footprint of similar order to other B-hadron classes.

12. How this complements the truth-level analysis

Truth-level analysis establishes the B-hadron identity and decay observables before detector effects. RECO-level analysis establishes what detector-level objects are available for those same truth classes.

Truth-level analysis gives	RECO-level analysis gives	Combined value
Ground-state B-hadron labels and counts	Reconstructed-particle multiplicities and kinematics	Connect true B species to visible detector signatures.
Validated flight distances and lifetimes	Track, jet, missing-energy, and PFO information	Move from lifetime validation toward detector-level selection variables.
Bc rarity and lifetime hierarchy	Bc-matched reco multiplicity and visible energy scale	Show that rare Bc events have a measurable detector-level footprint.
Cleaned truth endpoints and time conversion	Truth-matched reco ancestry labels	Validate future Bc training/selection studies before removing truth labels.

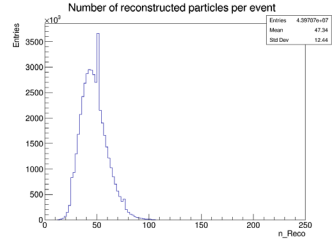
Truth-analysis validates the B-hadron object and lifetime observables; RECO validates that corresponding detector-level objects can be accessed, truth-labeled, and summarized into Bc-sensitive features.

15. Recommended next analysis steps

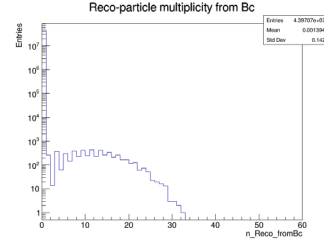
- 1 Use Stage 3 as baseline reconstructed-level ntuple. Re-run Stage 1/2 only for debugging or documenting development history.
- 2 Make per-class comparisons of Stage-3 feature variables: B_d , B_u , B_s , B_c , Λ_b , Ξ_b , Ω_b .
- 3 Add reconstructed vertex/displacement observables because truth-level flight observables are central to the project.
- 4 Implement richer hemisphere/thrust-like observables beyond the current jet proxies.
- 5 Choose explicit B_c decay modes and normalization channels for an LHCb-like fragmentation strategy.
- 6 Separate validation labels from data-like selections: use truth labels for training/validation, but define final candidate variables using reconstructed information.

Backup: Stage 1 output plot stack

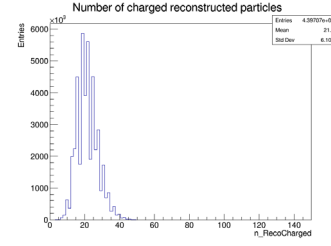
n_Reco



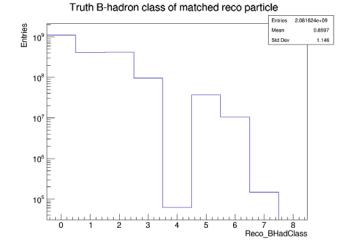
n_Reco_fromBc



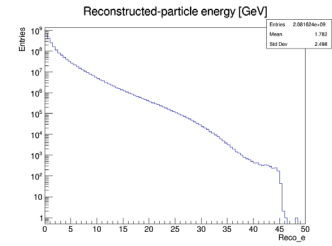
n_RecoCharged



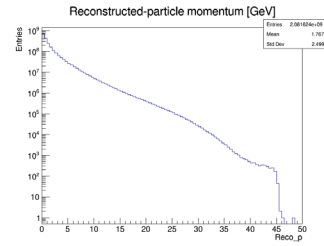
Reco_BHadClass



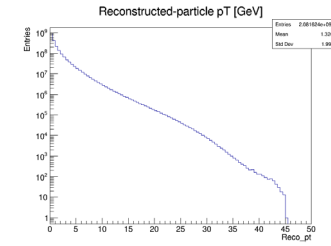
Reco_e



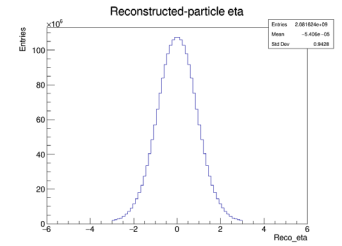
Reco_p



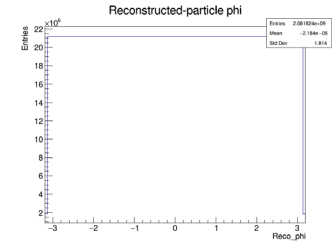
Reco_pt



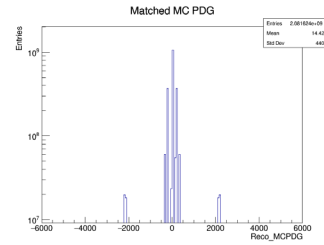
Reco_eta



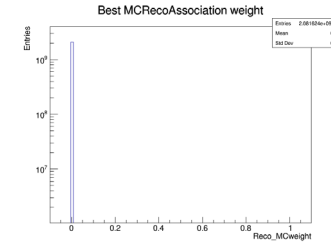
Reco_phi



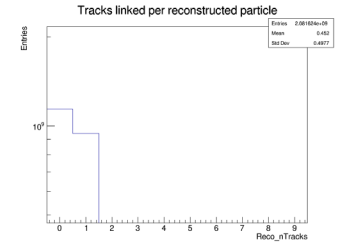
Reco_MCPDG



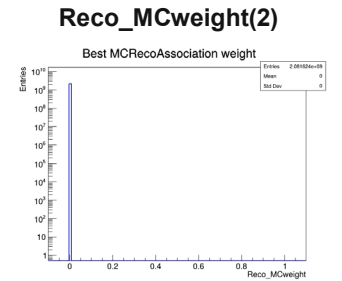
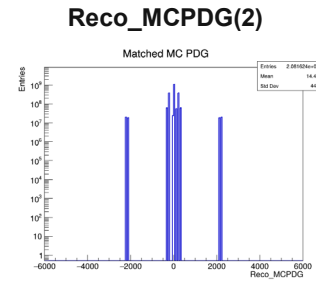
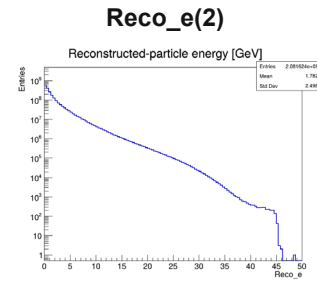
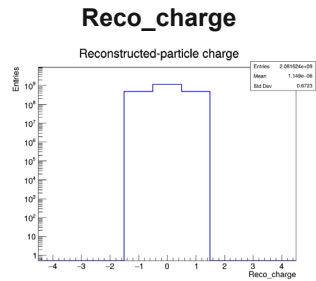
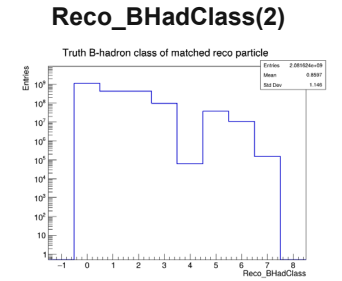
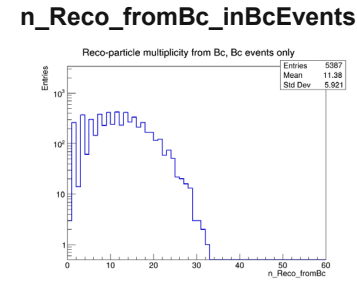
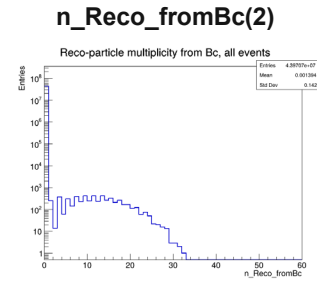
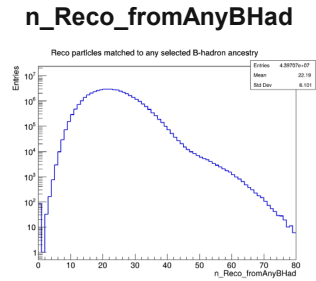
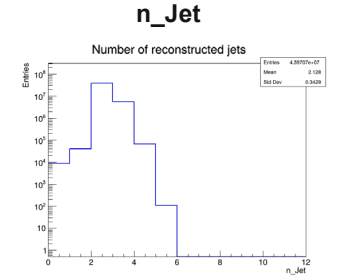
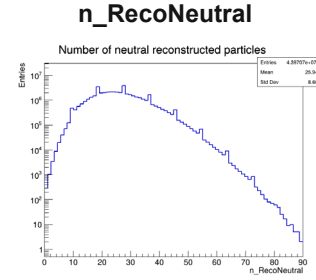
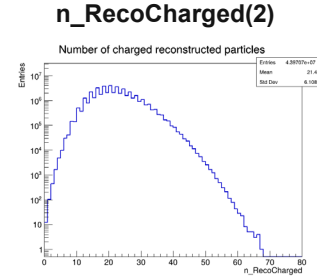
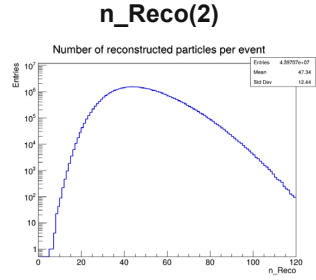
Reco_MCweight



Reco_nTracks

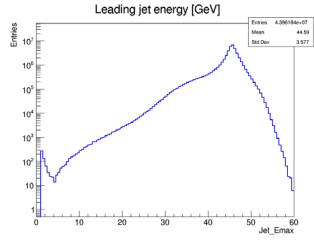


Backup: Stage 2 output plot stack

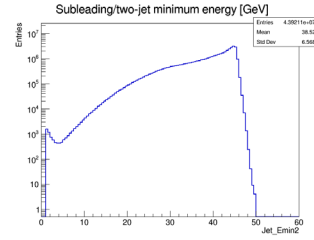


Backup: Stage 3 output plot stack

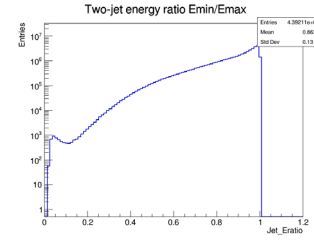
Jet_Emax



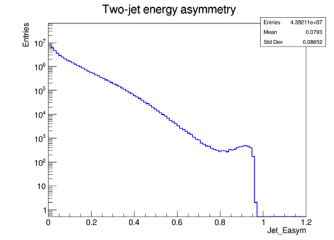
Jet_Emin2



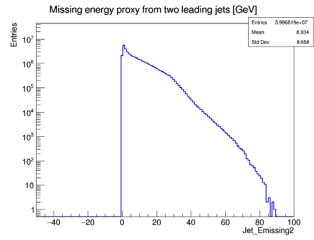
Jet_Eratio



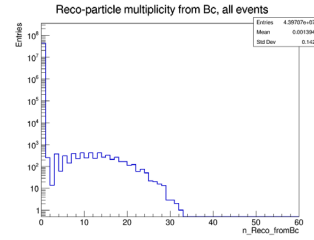
Jet_Easym



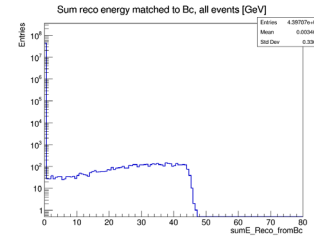
Jet_Emissing2



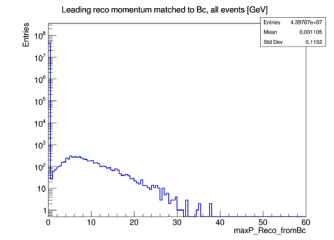
n_Reco_fromBc(3)



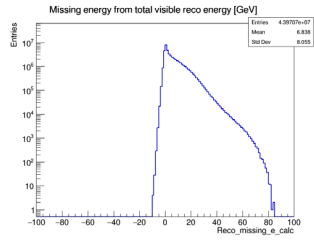
sumE_Reco_fromBc(2)



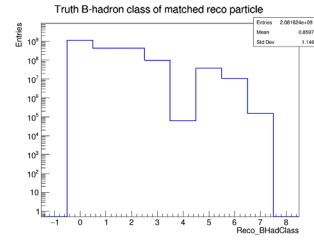
maxP_Reco_fromBc



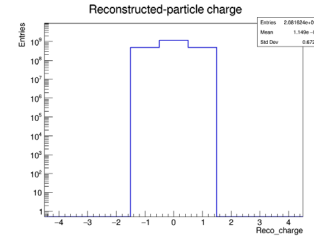
Reco_missing_e_calc



Reco_BHadClass(3)



Reco_charge(1)



Reco_pt(3)

