JetClass, five ways

Pythia · Sherpa · Herwig 7.3 · +DIRE · +VINCIA — one format, five shower / hadronisation models

NERSC Perlmutter (m2612) | Delphes + AK8 | exact JetClass ntuplization | 125M jets / edition

Generator robustness datasets for jet tagging · June 2026

Goal — two axes of generator variation

The public **JetClass** (100M jets, 10 boosted classes) was made with **MadGraph5 + Pythia8 + Delphes**. Tagger robustness to the **shower / hadronisation model** is a known concern, so we rebuild it in **byte-compatible format** along two axes:

Full-generator swaps
Sherpa 3 (own ME + CS-dipole shower + cluster hadr.) · **Herwig 7.3** (MG5 ME → angular-ordered shower + cluster hadr.)

Shower-only swaps inside Pythia
same MG5 ME + string hadronisation, change **only the shower**: **DIRE** (dipole) and **VINCIA** (antenna) — clean isolation of the shower

All six datasets (Pythia + 5 ours) share the **identical Delphes card, AK8 jets, and exact JetClass ntuplizer** (Huilin Qu's `makeNtuples`) — **125M jets** each, official layout (100k/file, train 100M + test 20M + val 5M), cross-comparable feature-by-feature.

Pipelines

ME + parton shower + hadronisation → **HepMC** → **Delphes** (AK8 EFlow, soft-drop, τ_N, track-IP smearing)
→ **makeNtuples** (gen-matched label) → **rebin** → 100k-jet files, train / test / val

Sherpa

NERSC native conda (OpenLoops + Sherpa 3 pkgs); ME+PS+hadr. all from Sherpa → HepMC3 → Delphes.

Herwig 7.3

el9 shifter container + CVMFS LCG_109a; MadGraph 3.6.4 → Herwig 7.3 → DelphesHepMC2.

DIRE / VINCIA

reuse the **same MG5 LHE** (container) → Pythia 8.3 shower (`model`=dire/vincia, conda) → HepMC2 → Delphes.

Reusing the Herwig MG5 LHE for DIRE/VINCIA means **only the shower differs** — and it sidesteps Sherpa's decay-forcing entirely (MG5+MadSpin make the decays).

The datasets

Edition	Status	Jets	Generation
Pythia (official)	ref	125 M	MG5 + Pythia8 simple shower
Sherpa (ours)	done	125 M	Sherpa 3 + OpenLoops (decay fixes, see next)
Herwig 7.3 (ours)	done	125 M	MG5 + Herwig 7.3 angular-ordered
Herwig (orig.)	ref	125 M	MG5 + older Herwig (JetClass authors)
Pythia+VINCIA (ours)	done	125 M	MG5 LHE + Pythia VINCIA antenna shower
Pythia+DIRE (ours)	done	125 M	MG5 LHE + Pythia DIRE dipole shower

Each: **10 classes × 12.5M jets**, batched SLURM arrays (transient HepMC on pscratch, cleaned per sub-batch). Pythia/Sherpa/Herwig integrity-**verified** (ALL

GOOD): file/entry counts, 41 branches, one-hot labels & particle types. All six editions now complete and verified.

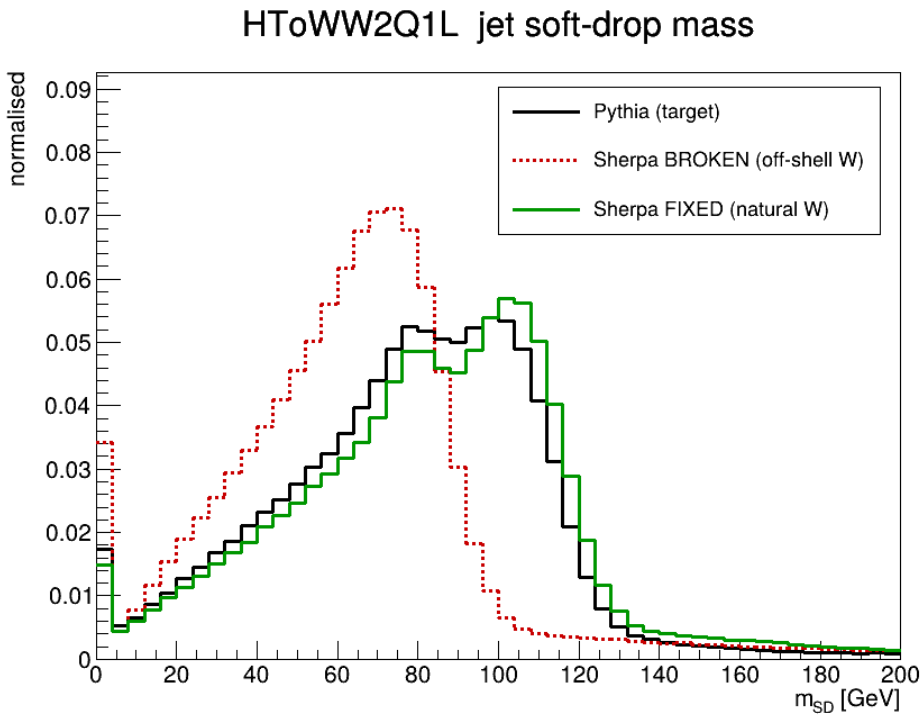
Sherpa caveat — forced decays that silently fail

Sherpa's `HARD_DECAYS` **silently drops a forced channel it cannot build** and falls back to the default decay — so **6/10 classes initially collapsed** (HToBB≡CC≡GG, HToWW4Q≡2Q1L, TTBar≡TTBarLep). Root causes + fixes:

- **h→bb / cc** — quark Yukawa = 0 by default → inject `PARTICLE\_DATA … Yukawa`.
- **h→gg** — loop-induced → `MODEL: HEFT` (like h→γγ).
- **h→WW*** — off-shell, 2-body invalid → force the **3-body** `25,∓24,q,q'` channels.
- **Leptonic W** — order must be (W, ν, ℓ); reversed → silently hadronic.
- **Verify the label, not that it ran:** gen-match every class.
- All fixes validated end-to-end (`gen\_match` 100% correct), regenerated, re-rebinned → **all 10 classes now distinct & verified.**

The semileptonic HToWW2Q1L needed one more fix → next slide.

HToWW2Q1L — the soft-drop mass you flagged



Forcing the exclusive decay pinned the hadronic quark pair to the **off-shell** W* (45 GeV) every event; the escaping ν means the jet mass tracks only the hadronic W, so m_{SD} peaked 25 GeV LOW (red, peak 73). Fix: force only h→WW* (BR-weighted) and let each W decay **naturally** → physical on/off-shell mix (green, peak 102, matching Pythia 98). Now generated, rebinned & swapped in.

Validation — Sherpa vs Pythia

Feature-distribution KS distance (averaged over 10 classes). **Close everywhere** — the small, expected signature of a different ME + shower + hadronisation:

Differ most (substructure):

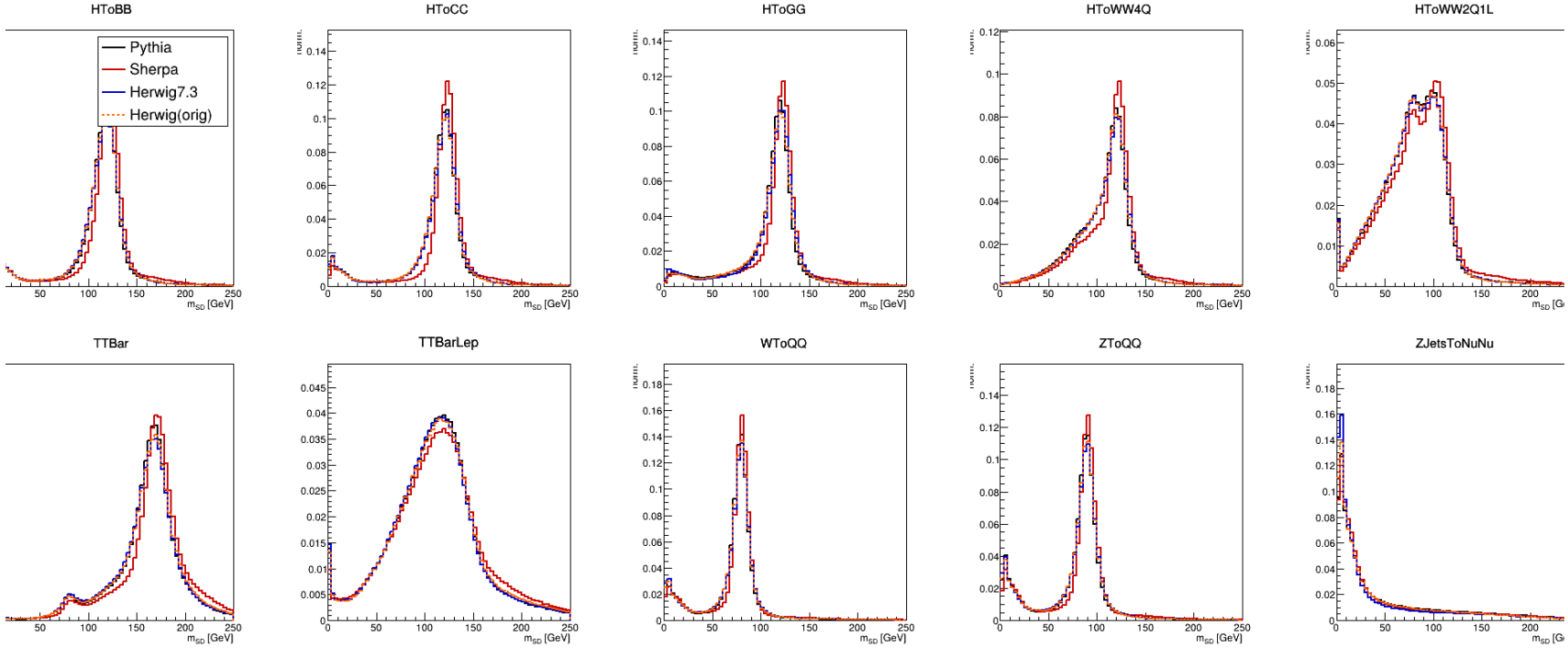
jet_sdmass	0.113
jet_tau2	0.085
jet_tau3	0.081
jet_nparticles	0.080
tau21	0.037

Agree closely (kinematics, detector):

jet_pt	0.016
jet_eta	0.003
part_deta	0.016
part_d0val	0.008
isElectron/Muon	0.004

Per-class mean KS **0.022–0.073**. Jet **kinematics, IP, and particle composition** match tightly; the residual — small (≤0.11) — lives in **substructure**, led by soft-drop mass.

Soft-drop mass, all 10 classes × 4 generators



Every class peaks at the **right resonance** (Higgs 123, top 169, W 81, Z 90; QCD falls off). Pythia, Herwig 7.3 and original Herwig sit almost on top of each other; **Sherpa** (red) grooms to a **slightly harder, narrower** peak across all classes — a coherent 0.1-KS shower signature, not a bug.

Why Sherpa grooms harder — it's the recoil, not the emissions

A natural guess: Sherpa's **dipole shower** emits less. The data says otherwise —

- Constituent **multiplicity is comparable** to Pythia (Herwig 7.3 is actually the lowest); Sherpa constituents are if anything slightly **softer**.
- The **ungroomed** jet mass is identical across generators; the difference appears **only after soft-drop** (Sherpa loses 2.5 GeV less to grooming).
- Groomed jet mass is the textbook probe of the **shower recoil scheme**. Sherpa's **CS-dipole local recoil** redistributes emission momentum differently from Pythia's pT-ordered and Herwig's angular-ordered showers → a harder groomed peak.
- Not a generic "dipole" effect: Pythia is **also** dipole-based yet clusters with Herwig. It's **Sherpa's specific CS-recoil + tune**.

This is exactly the question the DIRE / VINCIA editions are built to answer cleanly — swap the shower, hold everything else fixed.

Validation — Herwig 7.3 vs everything

Comparison	top-feature KS	Reading
Herwig73 **vs Pythia**	nparts 0.095, τ_4 0.081	Very close — both mature general-purpose showers; kinematics KS<0.02
Herwig73 **vs Sherpa**	nparts 0.129, τ_4 0.124	Slightly more — distinct ME + shower + hadronisation
Herwig73 **vs orig. Herwig**	nparts 0.107, τ_4 0.094	Reproduces the JetClass authors' (older) Herwig

Per-class mean KS: **vs Pythia 0.008–0.046**, **vs origHerwig 0.016–0.036**, **vs Sherpa 0.029–0.071** — all tight. **Herwig 7.3 ≈ Pythia** at jet level and faithfully reproduces the original JetClass Herwig; **Sherpa** sits a little further out but still

close. Residuals live in **multiplicity & N-subjettiness**, never in kinematics/IP (KS<0.02).

Shower isolation — the DIRE & VINCIA editions

Same MG5 matrix element, same string hadronisation, same Delphes — **change only the Pythia 8.3 shower**. This isolates the shower the way full-generator swaps cannot:

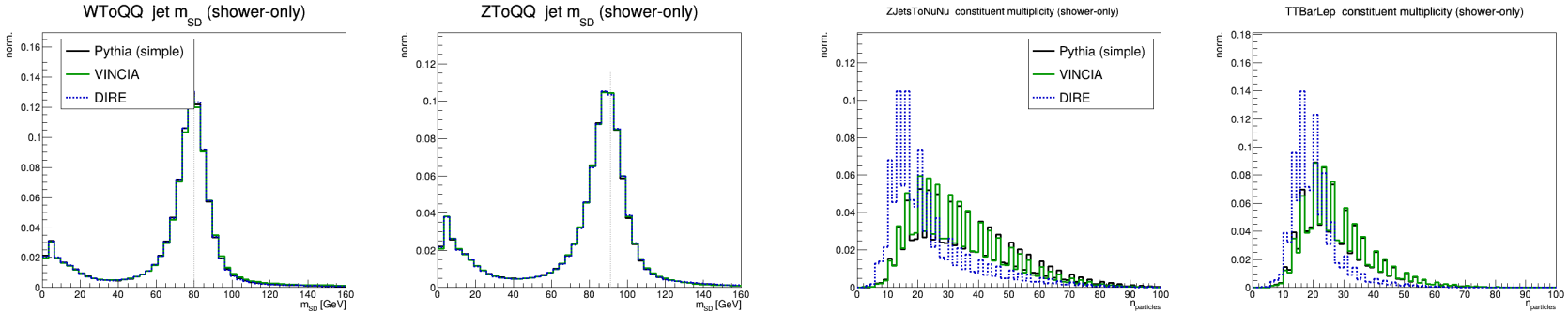
- **VINCIA** — antenna shower. **Indistinguishable from Pythia** at jet level: per-class KS **0.004–0.037**, every jet-level feature KS<0.024.
- **DIRE** — dipole shower (like Sherpa's CS). Matches Pythia on **quark** jets (Higgs/W/Z 0.01) but **softer on gluon-rich jets** — a genuine shower difference, **not** a bug.

DIRE vs Pythia — where it splits (KS):

class	nparts	m_{SD}
ZJetsToNuNu (QCD)	0.36	0.29
TTBarLep	0.32	0.03
TTBar	0.16	0.02
HToBB / WToQQ	<0.05	<0.01

Both editions reproduce the W/Z/top/Higgs mass peaks exactly (see next); DIRE's difference lives purely in **constituent multiplicity of gluon-radiation-rich jets**. Build gotcha: Delphes silently drops Pythia's HepMC3 → write **HepMC2**.

Shower isolation — agree on resonances, split on gluon jets



Left: W/Z soft-drop mass — Pythia (black), VINCIA (green), DIRE (blue, dashed) all peak on the boson mass. Right: constituent multiplicity for the gluon-rich classes — VINCIA sits on Pythia, while DIRE peaks lower with a shorter tail (fewer particles). Same MG5 events + string hadronisation throughout.

Where the editions agree — and where they split

Mean KS over 10 classes — the spread is **small and structured**:

feature	Sherpa –Pythia	Herwig –Pythia	Herwig –Sherpa
Kinematics / detector — universal			
jet_pt	0.016	0.016	0.016
jet_eta	0.003	0.005	0.004
part_deta	0.016	0.006	0.012
part_d0val	0.008	0.004	0.008
Jet substructure — small residual, no outlier			
jet_sdmass	0.113	0.027	0.112
tau21	0.037	0.030	0.055
jet_tau2	0.085	0.052	0.120
jet_nparticles	0.080	0.095	0.129

Two regimes, all close:

- **Kinematics** (pt, η, IP, particle ID) are fixed by the **matrix element + boost cut** → all editions essentially identical (KS<0.02).
- **Substructure** carries the only real spread, and even that is **modest** (≤0.13): **Pythia roughly central**, Herwig closest, Sherpa a touch out via its CS-dipole recoil.

No single generator is an outlier — the family clusters tightly, with substructure the lone axis of (small) disagreement. The shower-only DIRE / VINCIA editions extend this along a clean shower axis: VINCIA ≈ Pythia, DIRE splits only in gluon-jet multiplicity.

Engineering lessons (NERSC Perlmutter)

- **Native conda**, not home-grown micromamba (`module load conda`).
- **`$HOME` has no flock** → rattler/pip die with error 524; redirect caches to pscratch.
- **`/tmp` is RAM (tmpfs)** → multi-GB HepMC there OOM-kills; stage transient on **pscratch**.
- **ROOT 6.38 exits nonzero on success** → judge by the output file.
- **Herwig / MG need an el9 container** (cmssw/el9 + cvmfs) — Perlmutter nodes aren't el9.
- **Sherpa silently drops forced decay channels** → verify the gen-matched label, not just that it ran.
- **Forcing an exclusive decay biases kinematics:** HToWW2Q1L pinned the hadronic W off-shell → mass 25 GeV low. Force the boson, let sub-decays run naturally.
- **Delphes silently drops Pythia's HepMC3** (0 events) → write **HepMC2**.
- **DIRE is under-tuned for QCD jets** in 8.3 (CERN dropped it from newer LCG) — sound on resonances, softer on QCD. **Validate on a resonance before scaling.**

Summary

Delivered: byte-compatible JetClass editions —

Pythia Sherpa Herwig 7.3 +VINCIA +DIRE

each **125M jets**, official layout, side-by-side in `part_datasets/JetClass/`.

Validated: integrity ALL GOOD; Sherpa decay bugs found & fixed (incl. the HToWW2Q1L soft-drop mass); feature distributions cross-compared — physically sensible, kinematics consistent, **substructure carries the differences**.

Model spread (mean per-class KS):

VINCIA ↔ Pythia: **0.015** shower-only
Herwig73 ↔ Pythia: **0.023**
Herwig73 ↔ origHerwig: **0.025**
Sherpa ↔ Pythia: **0.034**
Herwig73 ↔ Sherpa: **0.043**
DIRE ↔ Pythia: **0.052** gluon jets

Delivered: a **6-way** shower / hadronisation matrix — full-generator (Sherpa, Herwig) and shower-only (VINCIA, DIRE) axes — for robustness

& domain-adaptation studies in jet tagging.

`jetclass-sherpa/` (Sherpa) · `jetclass_herwig73_cluster/` (Herwig) · `jetclass_pythia_showers/` (DIRE/VINCIA)