ThermoBench-Consist v1.0: A Tiny Benchmark for Thermodynamic Consistency of ML EoS/VLE Surrogates

Aryan Gupta
arygupta@umich.edu

Abstract

Machine-learned surrogates for equations of state (EoS) and vapor-liquid equilibrium (VLE) are increasingly used in design and CFD pipelines, but thermodynamic consistency (not only pointwise accuracy) determines solver stability. This benchmark presents ThermoBench-Consist v1.0, a CPU-only benchmark that exercises four lightweight checks: (C1) isotherm monotonicity $(\frac{\partial\rho}{\partial p})_T \geq 0$, (C2) positivity of isothermal compressibility, (C3) Clapeyron along saturation, and (C4) a CFD-relevant speed-of-sound sanity check a^2. The suite provides small seeded grids, guardrails near spinodal/critical regions, composite scores, and compact plots/reports plus a machine-readable JSON summary. On the shipped tiny grids CO_2, optional N_2, a reference CoolProp adapter passes Core+Plus checks, while a deliberately inconsistent toy surrogate fails as expected. All runs are CPU-only (CLI < 30 s; notebook < 60 s).

Code: https://github.com/guptaaryanr/ThermoBench-Consist.git
DOI: 10.5281/zenodo.17330440

1. Motivation

Surrogates for thermophysical properties offer speed and differentiability, yet solvers (flowsheeting, optimization, and especially compressible CFD) demand physically consistent state functions to remain well-posed. In a single phase, density must increase with pressure; compressibility should not be negative; VLE must satisfy Clapeyron; and the speed of sound should be plausible to avoid CFL violations and spurious acoustics [2,3,4]. ThermoBench-Consist is a diagnostics/benchmark, not an EoS. It complements reference libraries (e.g., CoolProp [1]) by providing fast, CI-friendly checks and reports that surface pathologies early.

2. Related Work

CoolProp provides reference-quality properties for pure fluids [1]. Classical texts summarize stability criteria, Clapeyron, and critical behavior [2,5,6]. In ML for physics, physics-informed or constrained objectives aim to embed governing laws [7]; the goal of this benchmark is orthogonal: a post-hoc consistency benchmark to evaluate any surrogate.

3. Methods

3.1 Checks

- C1 - Isotherm monotonicity: Mechanical stability in a single phase implies $(\frac{\partial \rho}{\partial p})_T \geq 0$. Centered finite differences along an isotherm are required to be non-negative within a small tolerance.
- C2 - Isothermal compressibility: With $\kappa_T = -\frac{1}{v}(\frac{\partial v}{\partial p})_T = \frac{1}{\rho}(\frac{\partial \rho}{\partial p})_T$ stability requires $\kappa_T \geq 0$. Slightly negative values within tolerance are treated as numerical noise.
- C3 - Clapeyron along saturation: $\frac{dP_{sat}}{dT} = \frac{\Delta h}{T \cdot \Delta v}$. The LHS is obtained from a reference saturation curve; the RHS uses phase-split properties from the surrogate (if exposed). The median relative error across sampled T is reported and passed if below a default tolerance.
- C4 - Speed of sound sanity (CFD sanity): $a^2 = (\frac{\partial p}{\partial \rho})_s$. a is compared between the surrogate (if provided) and a CoolProp baseline across a small set of T at a reference pressure p^*. The check passes if the median relative error is below a tolerance (default 0.2).

3.2 Adapter Interface

A model may expose: 'rho(T,p)' (required), 'h(T,p)' (optional), 'phase_split_at_T(T)' returning $(\rho_l, h_l, \rho_v, h_v)$ (optional), and 'speed_of_sound(T,p)' (optional). A capability descriptor declares which checks apply; unsupported checks are skipped (not penalized). Units are SI.

4. Sampling & Guardrails

This benchmark ships tiny seeded single-phase grids for CO_2 (e.g., $T \in [220,300]$ K) and optionally N_2. A near-spinodal flag is raised when derivative magnitudes approach zero, and an optional critical-band guard excludes a small neighborhood around (T_c, P_c) during sampling to avoid ambiguity, These guardrails inform interpretation but do not by themselves flip pass/fail.

5. Scoring and Reports

Per-check pass/fail aggregates to a composite 0-100 score (unsupported checks excluded). A Core badge (C1-C3) and a Plus badge (adds C4) are summarized. Each run emits a JSON with raw arrays and a compact Markdown/HTML report with small plots:

1. isotherm ρ-p with sign-aware shading,
2. Clapeyron LHS vs RHS (if phase-split is available), and/or
3. speed of sound vs T (reference vs surrogate).

6. Results Snapshot

Table 1: Tiny Pass/Fail Summary

Adapter	Fluid	C1	C2	C3	C4	Composite
CoolProp	CO_2	PASS	PASS	PASS	PASS	100.0
ToyInconsistent	CO_2	FAIL	PASS	FAIL	FAIL	33.3

Figure 1: Isotherm ρ-p With Sign-Aware Shading

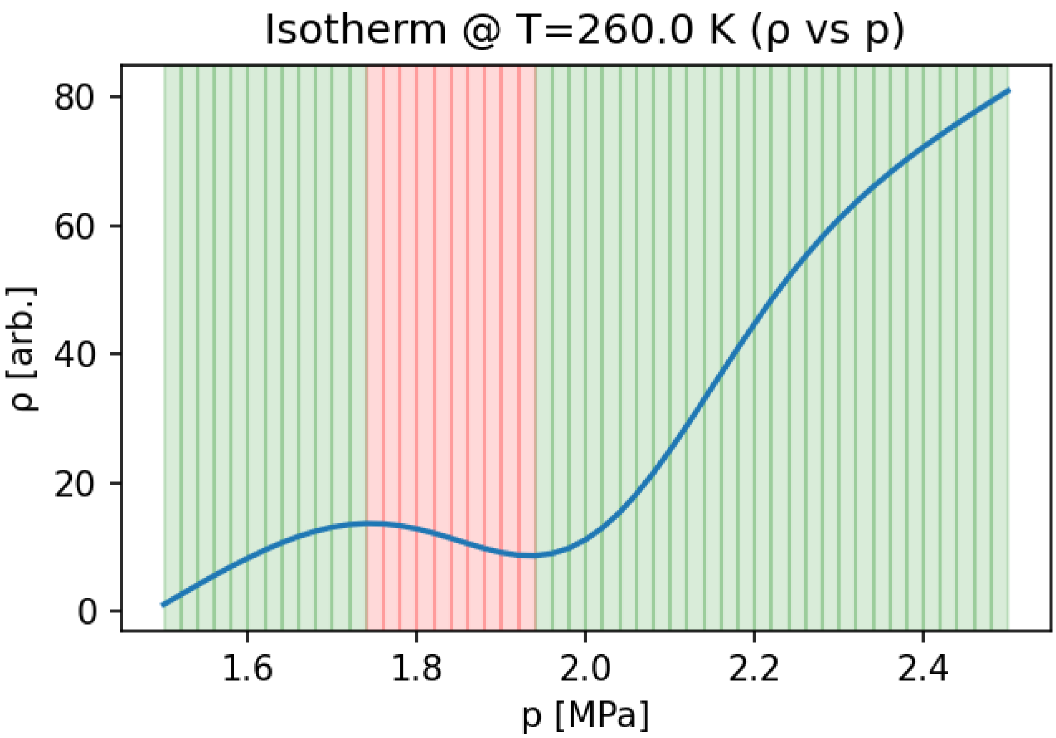


Figure 2: Clapeyron LHS vs RHS

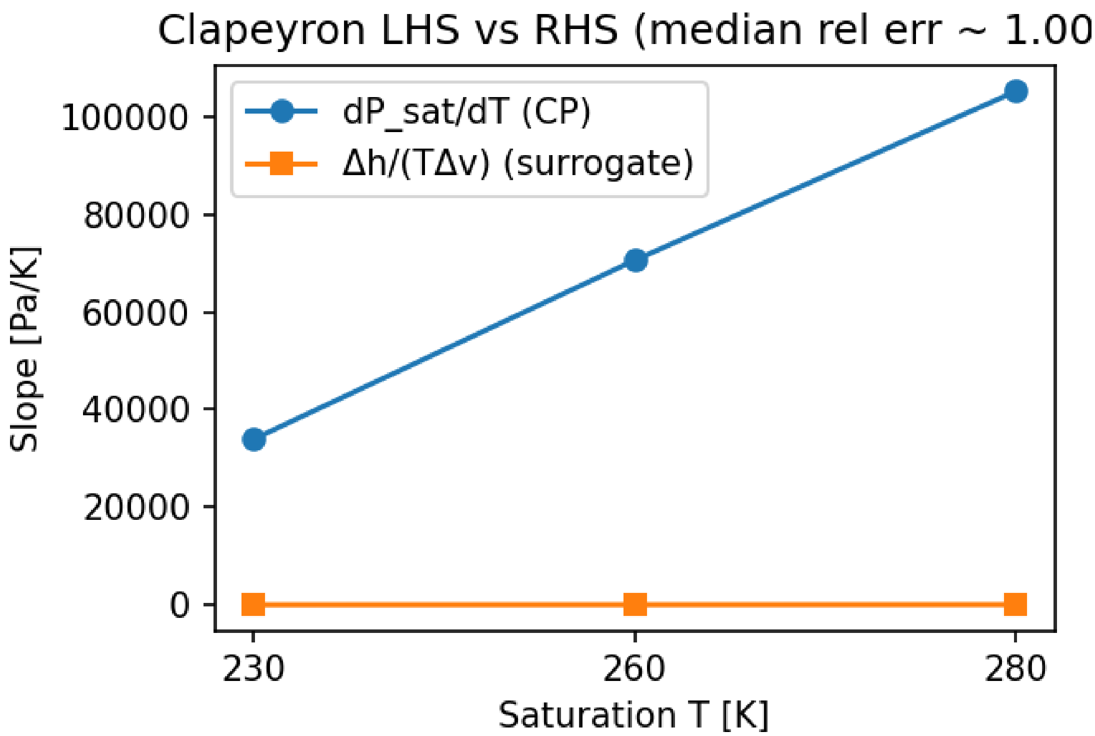
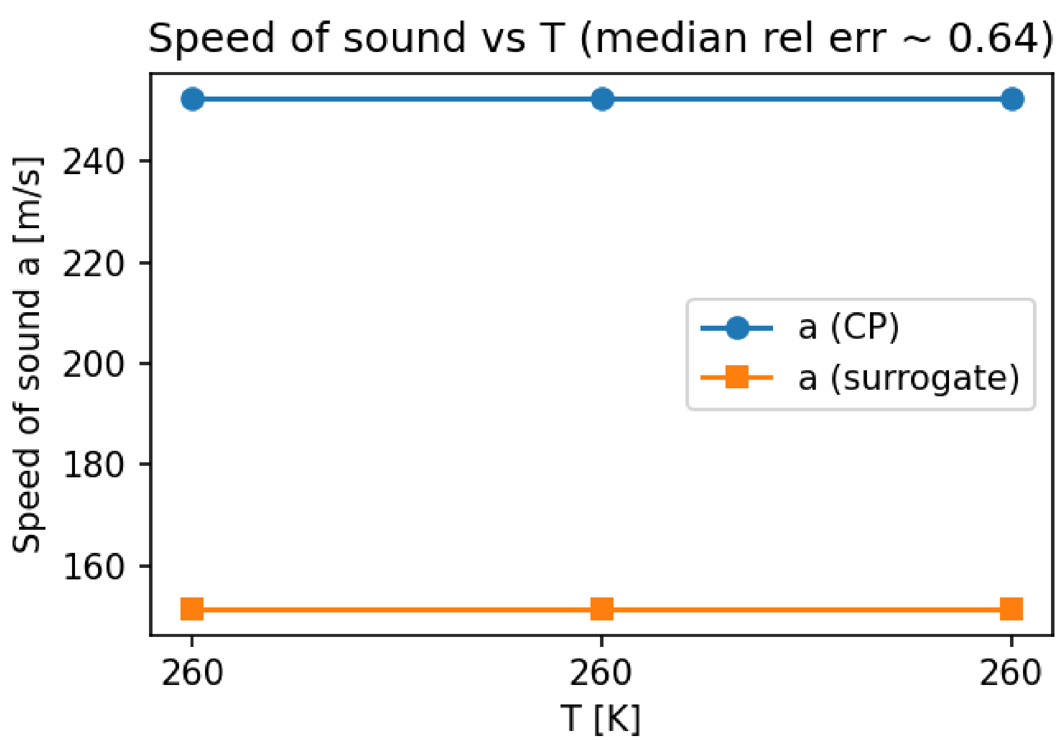


Figure 3: Speed of Sound vs T (Reference vs Surrogate)



7. Limitations

Current release targets pure fluids on small CPU-friendly grids; mixtures and composition derivatives are out-of-scope. Finite differences on tiny grids can misclassify edge cases; tolerances and near-critical guards mitigate but do not eliminate this risk. This benchmark uses CoolProp [1] as the baseline; multi-baseline comparisons are future work.

8. Availability & Reproducibility

All experiments are CPU-only and finish within the stated budgets (CLI < 30 s; notebook < 60 s). Each run writes a reproducible JSON plus plots used here.
GitHub: https://github.com/guptaaryanr/ThermoBench-Consist.git
DOI (code release): 10.5281/zenodo.17330440

References

[1] Bell, Ian H., et al. "Pure and Pseudo-Pure Fluid Thermophysical Property Evaluation and the Open-Source Thermophysical Property Library CoolProp." Industrial & Engineering Chemistry Research, vol. 53, no. 6, Jan. 2014, pp. 2498–508. Crossref, https://doi.org/10.1021/ie4033999.

[2] Callen, Herbert B., and H. L. Scott. "Thermodynamics and an Introduction to Thermostatistics, 2nd Ed." American Journal of Physics, vol. 66, no. 2, Feb. 1998, pp. 164–67. Crossref, https://doi.org/10.1119/1.19071.

[3] Moran, M. J., Shapiro, H. N., Boettner, D. D., & Bailey, M. B. (2018). Fundamentals of Engineering Thermodynamics (9th ed.). Wiley.

[4] Kundu, P. K., Cohen, I. M., & Dowling, D. R. (2015). Fluid Mechanics (6th ed.). Academic Press.

[5] Smith, J. M., Van Ness, H. C., & Abbott, M. M. (2017). Introduction to Chemical Engineering Thermodynamics (8th ed.). McGraw-Hill.

[6] Kontogeorgis, Georgios M., and Georgios K. Folas. Thermodynamic Models for Industrial Applications. Dec. 2009. Crossref, https://doi.org/10.1002/9780470747537.

[7] Raissi, M., et al. "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations." Journal of Computational Physics, vol. 378, Feb. 2019, pp. 686–707. Crossref, https://doi.org/10.1016/j.jcp.2018.10.045.