NASA-STD-7009B ASME V&V 40 DETERMINISTIC · NO LLM BUILT ON **UOFA** OPEN SOURCE

PhysMAP

A credibility guardrail for engineering surrogate models.

It tells you when a learned surrogate is being trusted **outside the physics it was validated for**, the failure a distance-to-training check can't see.

Vishnu Vettrivel · Doctoral Candidate, Systems Engineering, George Washington University

THE PROBLEM

A surrogate doesn't fail where you're watching. It fails on a physics variable it **never took as input**. Distance-to-training and uncertainty checks watch the input space, so when the model is pushed past the validity bound of a variable it doesn't even read, they stay quiet. That quiet zone is where deployed surrogates go wrong.

P1 · **KNOWS THE BOUNDS**

A curated closure corpus

53 published physics correlations, each with its validity bound, bound status, provenance, and geometry class. Hand-curated and literature-cited, so it can't be derived from your input statistics.

P2 · **SEES THE BLIND SPOT**

Observability-weighted detection

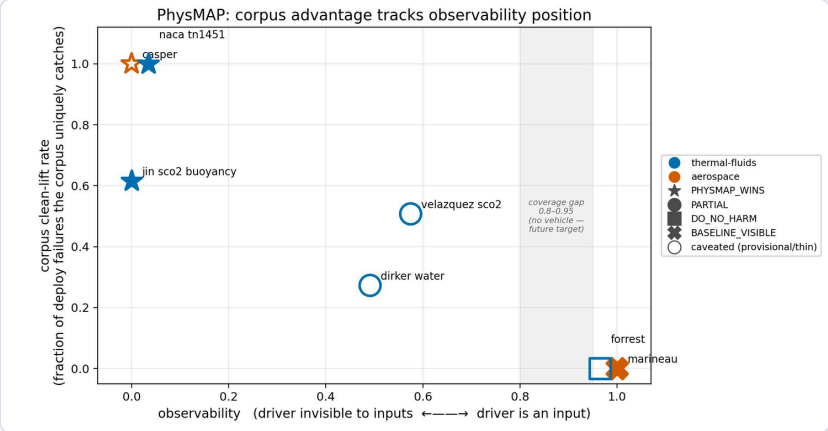
Each validity variable is classed as observable or unobservable to the surrogate's inputs. An unobservable breach is weighted higher, because that's the one your baseline structurally can't catch.

P3 · **RENDERS A VERDICT**

Deterministic, no LLM

Every operating point gets a verdict (TRUSTWORTHY, WARN, REJECT, UNCERTAIN) with a plain rationale and no threshold tuning. Auditable, and it pins in CI.

THE PROOF



Clean-lift rate is high where the failure driver is invisible to the surrogate's inputs and collapses to zero where the driver is an input. Both domains interleave along the axis. **7 vehicles, n=219.**

```
g = CredibilityGuardrail.fit(surrogate)
g.assess(op)
# op: Re=4.2e4 Pr=0.71 x/D=5
```

matched	gnielinski-1976
breach	x/D ≥ 10 (deploy 5)
observ.	x/D **UNOBSERVABLE**
baseline	QUIET
PhysMAP	**REJECT**

relies on gnielinski-1976 beyond its x/D ≥ 10 bound; x/D is not an input, so input-space checks can't see it.

WHY IT MIGHT FIT YOU

Works on any learned surrogate that maps inputs to a physical quantity. **Demonstrated on a 7-vehicle benchmark** across thermal-fluids, aerospace, and hemodynamics, on real measured data, where the corpus localizes the failures a steelman statistical baseline stays silent on. Verdicts are emitted as **machine-verifiable credibility evidence in the UofA vocabulary**, with a NASA-STD-7009B pack for aerospace. It sits beside your UQ and out-of-distribution checks; it catches what they can't see, it doesn't replace them.

BUILT ON UOFA

Verdicts as signed evidence

uofa.net
github.com/cloudronin/uofa

SEE IT RUN

Live walkthrough on request

Email physmap@uofa.net for a short session on the benchmark, baseline quiet versus corpus fires.

THE ASK

Bring one of your own

Send one surrogate you deploy and the operating point you use it at. I'll show you whether your current checks would have flagged it, and whether PhysMAP would.

PhysMAP builds on the Unit of Assurance (UofA).

Vishnu Vettrivel · physmap@uofa.net · uofa.net