

# Aero Blind Validation Challenge 01

## Reference-withholding protocol and implementation status

Alex Blythe / Aero engineering record

13 September 2026

### Abstract

This preparation release establishes a public evidence index and a testable framework for an externally referenced engineering challenge. NASA TM-4074 is selected as a candidate experimental source. The exact test series, permitted input package, sealed measurements and scientific acceptance thresholds have not been frozen. No production solve, frozen prediction, experimental reveal or comparison has occurred. Consequently the challenge is NOT EVALUATED. This paper documents the protocol, implemented safeguards, source-selection limitations and remaining work; it is not a validation-result paper or a preregistration.

## 1 Objective and existing evidence

The challenge asks whether a prediction made without runtime access to experimental answers agrees with an independent measurement series under criteria fixed before production solving. It complements, rather than replaces, numerical verification. Aero's retained Verification Bench Rev 1.0 demonstrates analytical pipe agreement, measured negative controls and readiness rejection. The channel study compares nine OpenFOAM solves against a parallel-plate reference. The thermal study is analytical screening. Seven structural configurations contain 21 retained CalculiX solves with mixed gates. None of these facts establishes universal validation.

## 2 Source selection

The selected candidate is C. L. Ladson's NASA TM-4074 (1988), NTRS 19880019495 [1]. NASA metadata identifies the Langley low-turbulence pressure tunnel and a campaign varying Mach number, Reynolds number and transition fixing. Its stated ranges are Mach 0.05–0.36 and chord Reynolds number 3–12 million. These ranges are not selected production conditions.

A complete matched  $C_L(\alpha)$  and  $C_D(\alpha)$  series is proposed. It is an operating-condition distribution rather than a spatial field. Availability of  $C_p(x/c)$  for a matched series is not asserted without source curation. The NASA wall-mounted hump was considered for spatial distributions but deferred because incoming-boundary-layer, blockage and configuration reconstruction add demands to the first challenge. Familiar local NACA examples and numerical-reference cases cannot substitute for an independent experimental series.

Public NASA access is not blanket clearance to republish every excerpt, comparison or derived dataset. A custodian must review the actual rights, source version, extraction accuracy and experimental methodology before accepting this candidate for production.

### 3 Blindness: allowed information and exposure

The input package may contain geometry, chord and trailing-edge convention, flow conditions, station locations, fluid properties, facility reconstruction information and non-answer-bearing uncertainty metadata. It must exclude experimental coefficient values, curves, author conclusions, published CFD comparisons and advice revealing successful modeling choices.

The reference package holds the measurements and any answer-bearing material. In the implemented CSV comparator, each measurement is identified by quantity and station, with a value and optional uncertainty. References with different formats require explicit reviewed extraction; they are not silently parsed or interpolated.

**Exposure already occurred in planning.** Web-search snippets returned experimental and published-comparison numbers to the implementation task. That task must not be reused as the blind predictor. Its exposure is documented in the source-selection record. A fresh predictor must receive only permitted inputs, without inherited conversation, RAG, browsing or benchmark memory files. Model pretraining may contain famous benchmarks; runtime withholding cannot prove total prior ignorance.

### 4 Roles and technical controls

1. The custodian curates source material and holds the encryption key, outside the predictor and repository.
2. A supervisor verifies input-only mounts, network restrictions, disabled retrieval and predictor identity before authorizing execution.
3. A fresh predictor and solver operate on the permitted package and frozen criteria.
4. The comparison operator receives the key only after the prediction and retained evidence have been anchored to source control.

The reference is encrypted with AES-256-GCM. Both plaintext and ciphertext SHA-256 commitments are recorded; the key is never printed or included in the package. Encryption protects bytes, not custodian honesty. An allowlisted schema does not prove that free-text fields are free of answer material.

The state machine records isolation attestations but does not itself launch or secure an Aero container, inspect a model's memory, or independently certify operator assertions. A restricted predictor/solver integration and reviewed supervisor evidence remain prerequisites to a scientific production run. This distinction is essential: unit-tested sequencing is not an accomplished blind experiment.

## 5 Chronology and freeze gates

| Scientific stage                | Current Challenge 01 status |
|---------------------------------|-----------------------------|
| Input frozen                    | Not created                 |
| Criteria frozen and committed   | Draft only                  |
| Production solve                | Not started                 |
| Prediction frozen and committed | Not created                 |
| Reference unsealed              | Not occurred                |
| Automated comparison            | Not evaluated               |

The implementation requires complete preregistration before freeze: source version and citation, package hashes, geometry, dimensional assumptions, flow conditions, solver and equations, turbulence/wall treatment, boundary/initial conditions, mesh levels and refinement, wall-resolution and convergence criteria, conservation, comparison metrics, thresholds, uncertainty, preprocessing restrictions, failure conditions, versions, date and repository commit.

Three mesh levels are required. The production template deliberately contains unresolved fields and a DRAFT status; the framework refuses to freeze it. Proposed OpenFOAM/SST choices are not commitments. Selecting thresholds now without a clean review of the experimental series and intended-use tolerances would be premature.

Anchoring fetches origin/main, verifies that the supplied full commit is reachable there, and compares every frozen file byte with the committed blob. A prediction cannot be unsealed merely by supplying a timestamp string. Git records are auditable but not a trusted timestamp service or immutable against administrators. Independent archival or signatures can strengthen custody.

Prediction freeze requires every registered quantity/station and hashed evidence for requirements, first principles, geometry, topology, mesh, mesh diagnostics, solver deck, runtime, log, residuals, conservation, wall diagnostics, grid study and extracted values. Unknown or failed numerical status is retained. An implementation defect or proposed correction must be a visible new version; it cannot overwrite the frozen original.

## 6 Comparison behavior

Exact registered station matching is enforced. Missing or duplicate points, non-finite values and omitted regions reject comparison. The implemented metrics for a quantity with  $N$  stations are

$$E_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - e_i)^2}, \quad E_{\text{max}} = \max_i |p_i - e_i|.$$

Thresholds must be positive, finite and declared separately for each quantity. Relative error at zero reference is undefined. Known uncertainty overlap is descriptive; missing uncertainty remains unknown and cannot widen an acceptance bound after unsealing.

## 7 Outcomes, reproduction and reporting

All declared quantities passing gives PASS; none passing gives FAIL; a mixture gives MIXED. Failed or unestablished numerical adequacy prevents an overall scientific PASS even if experimental errors happen to be small. Full signed/absolute errors and uncertainty availability remain in the result.

The result renderer accepts only an anchored comparison, produces complete-series prediction/reference overlays and signed-error plots, and generates a status-dependent LaTeX report. It does not supply fallback measurements or invent a missing run. Any later explanation or model improvement must be labeled post-unseal diagnosis or Challenge 01B, preserving the first result.

The preparation download contains source code, unit tests, the incomplete preregistration template, source selection and exposure record, reproduction instructions and this protocol paper. It contains no experimental paper, hidden measurement package, production result, solver deck or purported prediction. It is a protocol source release, not the final benchmark reproduction bundle.

The implemented tests use synthetic values explicitly labeled as such. They exercise encryption, early-unseal refusal, required anchoring, tampering, incomplete/non-finite criteria, full curve coverage, unknown uncertainty, all three comparison dispositions and a temporary Git-backed chronology. These tests support implementation confidence, not physical validation. The public manifest distinguishes this preparation package from scientific checkpoint hashes, which do not yet exist.

## 8 Remaining work and conclusion

A source custodian must select a complete matched experimental series, review rights and source version, curate geometry/flow/transition/facility information without answers, and escrow the measurements. A separate clean input-only review must fix numerical settings and meaningful acceptance tolerances. After public preregistration, a supervised restricted predictor/solver environment must retain the complete numerical trail, freeze its prediction publicly, then permit reference reveal and comparison.

No result is predetermined. The current status is **NOT EVALUATED**. This release makes the missing steps inspectable instead of presenting framework readiness as experimental success.

## 9 References

- [1] Charles L. Ladson, *Effects of independent variation of Mach and Reynolds numbers on the low-speed aerodynamic characteristics of the NACA 0012 airfoil section*, NASA TM-4074, 1988. Metadata reviewed 13 September 2026; production source-file version not yet frozen.  
<https://ntrs.nasa.gov/citations/19880019495>  
<https://ntrs.nasa.gov/api/citations/19880019495>