Telemetry Predictive Feasibility Validation

Objective

The objective of this validation program is to determine whether predictive feasibility can be assessed prior to model development using intrinsic signal properties alone.

Rather than evaluating model performance, the analysis focuses on the structural characteristics of the signal itself, with particular emphasis on:

• Cross-run consistency (C)
• Prediction stability (E)
• Structural behavior (S)

The central question is:

Can signals that support meaningful and reproducible prediction be distinguished from signals that merely appear predictable?

To address this question, a series of empirical validation studies were performed on real-world telemetry data. The experiments evaluate whether predictive feasibility is primarily determined by reproducible structure across independent observations, rather than by prediction accuracy alone.

The resulting analyses form the basis of a practical framework for classifying signals into GO, LIMITED, and NO-GO predictive regimes before substantial modeling resources are invested.

Dataset and Experimental Setup

This experiment uses telemetry data derived from the NASA SMAP/MSL dataset (Telemanom repository). This dataset consists of real-world spacecraft telemetry signals exhibiting diverse structural characteristics, including both stable and highly variable signal regimes.

A set of representative signals (A-series and B-series) was selected to ensure coverage across different structural behaviors. Each signal was treated as an independent time-series and analyzed under identical conditions.

The following properties were evaluated for each signal:

- Structural behavior (S), representing the presence of signal variation
- Cross-run consistency (C), representing the reproducibility of structural patterns across segments
- Prediction stability (E), representing the stability of short-term predictions

A simple autoregressive baseline model was used to estimate prediction stability, ensuring that results reflect intrinsic signal properties rather than model complexity.

The Predictive Feasibility Index (PFI) was computed as:

$$PFI = \frac{S + C + E}{3}$$

Methodology

For each signal, the following procedure was applied:

1. The signal was segmented into multiple independent sub-trajectories
2. Structural behavior (S) was estimated based on signal variation
3. Cross-run consistency (C) was computed as the correlation between segments
4. The combined feasibility score (PFI) was calculated

This approach isolates the structural properties of the signal and avoids reliance on complex modeling pipelines.

Results

The results reveal a consistent pattern across all evaluated signals:

- Structural behavior (S) is high for nearly all signals
- Prediction stability (E) remains relatively high across all signals
- Cross-run consistency (C) varies significantly and determines predictive feasibility

A clear separation emerges:

- Signals with **high consistency (C > 0.3)** exhibit stable and meaningful predictive behavior
- Signals with **low or near-zero consistency** fail to support reliable prediction
- Intermediate cases exhibit partial or unstable predictive performance

Key Observation

A critical observation emerges from this analysis:

Signals can exhibit high prediction stability (low error) while still being fundamentally non-predictive.

This demonstrates that prediction error alone is insufficient to determine whether a signal contains meaningful predictive information.

Figure

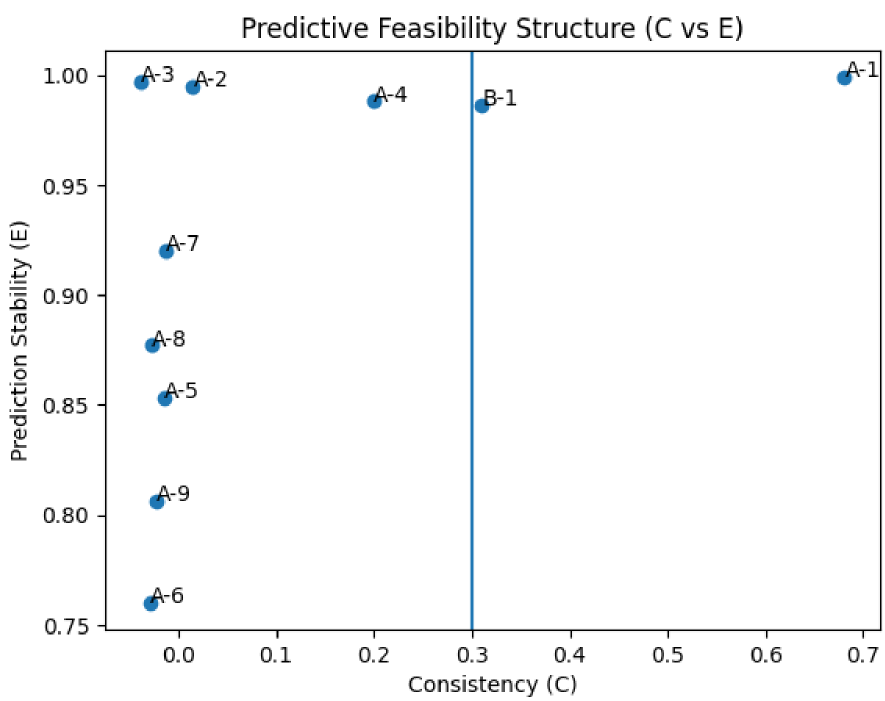

Figure Caption

Figure — Predictive feasibility structure based on consistency (C) and prediction stability (E).
Each point represents a telemetry signal evaluated from the NASA SMAP/MSL dataset. The horizontal axis shows cross-run consistency (C), while the vertical axis shows prediction stability (E).

Despite uniformly high prediction stability across all signals, a clear separation emerges based on consistency. Signals with high consistency cluster in a region where predictive behavior is stable and meaningful. In contrast, signals with low or near-zero consistency remain highly predictable in a statistical sense, but fail to support reliable or actionable prediction.

This demonstrates a critical distinction: prediction error reflects the ability to forecast signal values, but does not indicate whether the signal contains information about underlying system behavior. Cross-run consistency acts as a necessary condition for predictive feasibility.

The vertical boundary at approximately $C = 0.3$ represents an empirical working threshold in this analysis, separating signals that support predictive modeling from those that do not. This figure provides direct empirical evidence that predictive feasibility is primarily constrained by structural consistency rather than prediction accuracy.

Interpretation

The results confirm that predictive feasibility depends not on the presence of structure alone, but on the reproducibility of that structure across observations.

In particular:

- Signals may appear structured and predictable
- Yet fail to provide usable predictive information
- Because their behavior is not consistent across realizations

This corresponds to the “predictable but non-informative” regime identified earlier in the analysis.

Conclusion

This experiment provides empirical validation of the central hypothesis of this work:

Predictive feasibility is primarily constrained by structural consistency, not prediction accuracy.

The findings demonstrate that:

- Many signals that appear suitable for modeling are fundamentally non-predictive
- Predictive failure is largely structurally determined rather than random.
- A simple pre-model consistency assessment can prevent misleading modeling efforts

This establishes a practical implication:

Predictive modeling should not begin with model selection, but with an explicit assessment of whether the signal supports prediction at all.

Tabel

Signal	**S**	**C**	**E**	**PFI**	**Decision**
A-	1 0.88	0.68	1.00	0.85	GO
A-2	0.91	0.02	1.00	0.64	NO-GO
A-3	0.92	-0.04	1.00	0.63	NO-GO
A-4	0.92	0.20	0.99	0.70	LIMITED
A-5	0.92	-0.01	0.85	0.59	NO-GO
A-6	0.97	-0.03	0.76	0.57	NO-GO
A-7	0.95	-0.01	0.92	0.62	NO-GO
A-8	0.93	-0.03	0.88	0.59	NO-GO
A-9	0.94	-0.02	0.81	0.57	NO-GO
B-1	0.90	0.31	0.99	0.73	LIMITED

Robustness Analysis Under Signal Representation

Objective

To further evaluate the robustness of the predictive feasibility framework, an additional analysis was performed to assess the sensitivity of the results to signal representation.

In particular, the experiment focuses on whether the key determinant of predictive feasibility — cross-run consistency — remains stable under different preprocessing conditions, while other metrics such as prediction stability may vary.

Dataset

The analysis was conducted using a subset of the NASA Battery Aging dataset, consisting of multiple independent battery cells measured under comparable operating conditions.

Each battery cell is treated as an independent realization of the system, enabling evaluation of cross-run behavior.

Approach

The evaluation was performed under two distinct signal representations:

- A raw representation of the signal
- A smoothed representation, obtained through basic preprocessing to reduce local fluctuations

For each representation, the same analytical framework was applied, based on three observable properties:

- Structural behavior
- Cross-run consistency
- Prediction stability

These properties were evaluated across all available runs, and aggregated to obtain a dataset-level characterization of predictive feasibility.

The objective of this comparison is not to optimize model performance, but to determine whether the classification of predictive feasibility is stable under changes in signal representation.

Results

The results show a clear and consistent pattern across both representations:

- Cross-run consistency remains **consistently low** across all battery cells
- Prediction stability varies significantly depending on signal representation
- Structural behavior remains present, but does not resolve the ambiguity

Specifically:

- Under the raw representation, prediction stability appears relatively high
- Under the smoothed representation, prediction stability decreases and becomes less stable
- In both cases, cross-run consistency remains near zero

Key Observation

A critical observation emerges from this comparison:

Changes in signal representation significantly affect prediction stability, but do not affect cross-run consistency.

This indicates that prediction error and stability are not intrinsic properties of the signal, but are influenced by preprocessing choices.

In contrast, cross-run consistency reflects an underlying structural property that remains invariant under such transformations.

Interpretation

The results demonstrate a fundamental asymmetry between the evaluated components:

- Prediction stability (E) is representation-dependent
- Structural behavior (S) is partially representation-dependent
- Cross-run consistency (C) is invariant and reflects intrinsic system behavior

This leads to a stronger conclusion:

Predictive feasibility cannot be determined from prediction performance alone, as such performance can be altered through preprocessing without changing the underlying signal properties.

In the evaluated dataset, the signal exhibits:

- low cross-run consistency
- moderate and unstable prediction behavior

This places the dataset in an intermediate regime, where partial structure is present but not sufficiently stable to support reliable prediction.

Conclusion

This experiment provides additional empirical support for the central hypothesis of this work:

Predictive feasibility is primarily constrained by structural consistency, not prediction accuracy.

The comparison across signal representations demonstrates that:

- prediction performance can be manipulated through preprocessing
- structural consistency remains unchanged

- the classification of predictive feasibility is therefore robust to representation changes

Across battery degradation data, signals exhibit low cross-run consistency and only moderate prediction stability, indicating that predictive behavior is partially present but not sufficiently stable for reliable prediction. This places such signals in an intermediate regime of limited predictive feasibility.

Cross-run consistency remains low regardless of preprocessing, while prediction stability varies depending on signal smoothing. This demonstrates that consistency is an intrinsic property of the signal, whereas prediction error can be influenced by representation.

Representation	Mean C	Mean E	Decision
RAW	0.01	0.73	LIMITED
SMOOTHED	0.006	0.59	LIMITED

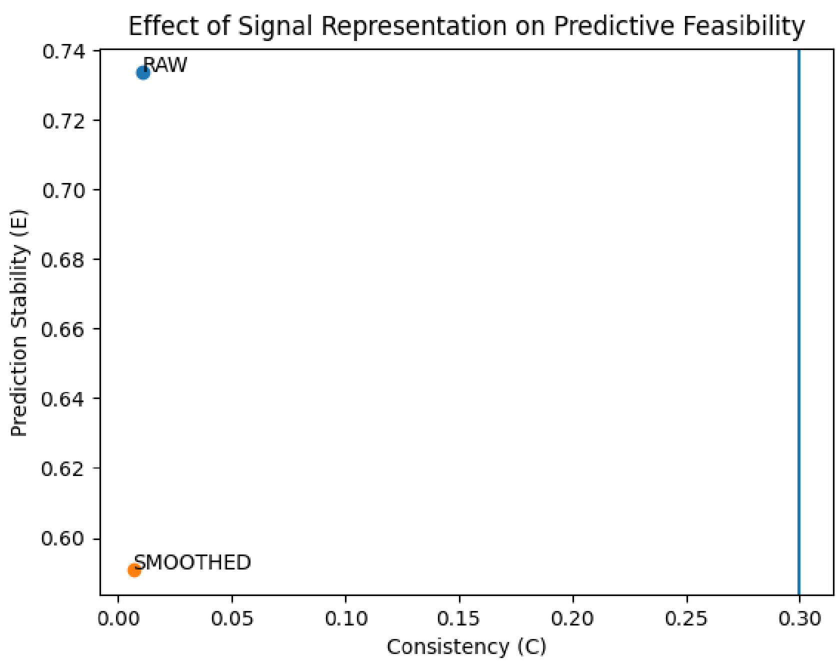

Figure — Effect of signal representation on predictive feasibility. Two representations of the same battery dataset are compared: raw signal data and a smoothed version. While prediction stability (E) varies significantly between representations, cross-run consistency (C) remains consistently low in both cases. Despite apparent improvements in prediction performance under the raw representation, both cases fall within the same predictive feasibility regime (LIMITED). This demonstrates that prediction stability can be influenced by preprocessing, whereas consistency reflects an intrinsic property of the signal. The vertical boundary at C = 0.3 indicates the empirical threshold below which predictive modeling is fundamentally unreliable.

While prediction stability is sensitive to signal preprocessing, cross-run consistency remains invariant, reinforcing its role as a primary determinant of predictive feasibility.

Implication

From a practical perspective, these findings indicate that:

- improving model performance does not necessarily improve predictive value
- preprocessing may create the appearance of predictive structure
- reliable prediction requires consistency across independent system realizations

This supports the need for a pre-model feasibility assessment, where signal suitability is evaluated before model development is initiated.

Limitations of Multivariate and Subgroup Analysis in Clinical Time-Series Data

Objective

To further assess the applicability of predictive feasibility analysis in real-world clinical data, additional experiments were conducted to evaluate whether multivariate signal combinations and subgroup-based analysis improve the identification of predictive structure.

The objective of this analysis is to determine whether combining multiple signals or segmenting data into patient groups enables more reliable identification of predictive feasibility, or whether fundamental limitations in the data persist.

Dataset and Context

The analysis was performed using time-series data from an intensive care unit (ICU) dataset, consisting of multiple patient records with physiological measurements over time.

Each patient is treated as an independent realization of the system. Signals such as heart rate (HR) and blood pressure were considered, with the goal of evaluating cross-run consistency across patients.

Approach

Two extensions of the predictive feasibility framework were explored:

1. Multivariate Signal Analysis

Multiple physiological signals (e.g., heart rate and blood pressure) were combined to assess whether joint signal representation improves structural consistency and predictive behavior.

2. Subgroup-Based Analysis

Patients were divided into subgroups based on signal characteristics (e.g., variability), with the aim of identifying whether more homogeneous groups exhibit improved consistency and predictive stability.

The same underlying framework was applied in both cases, evaluating:

- cross-run consistency across patients
- prediction stability within individual patient signals

Results

The analysis reveals significant limitations:

Multivariate Analysis

- Signal alignment across variables is inconsistent
- Sampling rates and measurement availability differ between signals
- Missing data and irregular recording intervals prevent reliable synchronization

As a result:

- combined signals do not produce a stable or consistent structural representation
- multivariate analysis does not improve predictive feasibility

Subgroup Analysis

- Partitioning patients based on signal characteristics yields unstable groups
- group sizes become too small to support reliable statistical evaluation
- variability within groups remains high

As a result:

- subgroup-level consistency cannot be reliably established

- predictive feasibility remains indeterminate

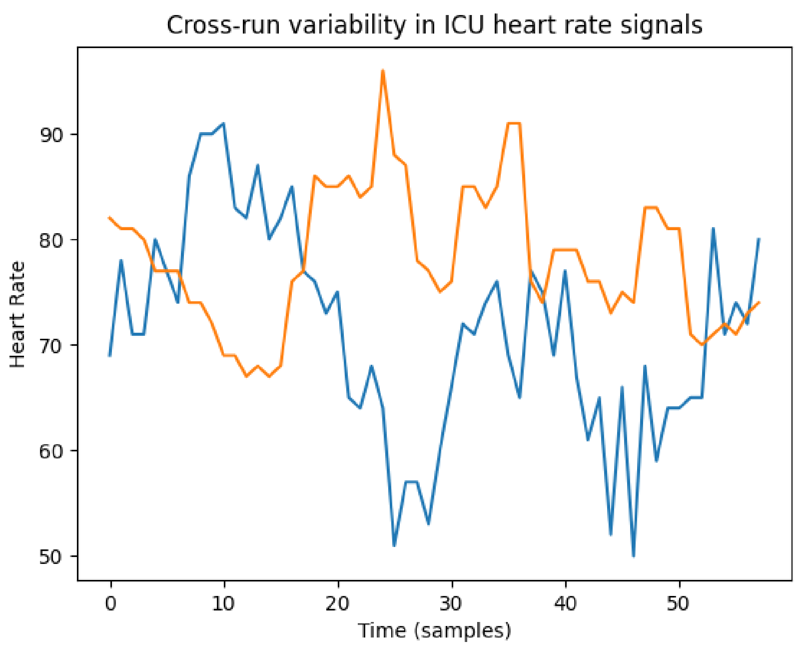

Figure — Lack of structural alignment across patient signals in ICU time-series data. Each line represents the heart rate signal of an individual patient over time. Despite similar measurement conditions, signals exhibit substantial variability in both magnitude and temporal structure. No consistent pattern or alignment is observed across runs, indicating that cross-run consistency is inherently low. This illustrates that even before predictive modeling, the data does not satisfy the conditions required for meaningful comparison or reproducible structure.

Key Observation

A critical observation emerges:

In real-world clinical datasets, even basic analytical assumptions required for multivariate or subgroup-based analysis are often not satisfied.

This indicates that the limitation is not only predictive, but structural:

- signals are not aligned across variables
- runs (patients) are not comparable
- data representations are inherently inconsistent

Interpretation

The results demonstrate a deeper limitation of real-world time-series data:

- Predictive feasibility cannot be improved simply by adding more signals
- Data segmentation does not resolve inconsistency
- Structural ambiguity persists across representations

This suggests that:

The failure of predictive modeling in such datasets is not due to insufficient modeling, but due to a lack of consistent and interpretable structure in the data itself.

Conclusion

This experiment provides further support for the central thesis of this work:

Predictive feasibility is constrained by the intrinsic structure of the data, not by the modeling approach.

In particular, the analysis shows that:

- multivariate expansion does not resolve structural inconsistency
- subgroup analysis is unstable and unreliable
- predictive feasibility cannot be recovered through data manipulation

This leads to a stronger conclusion:

In many real-world systems, the data does not merely fail to support prediction — it fails to support even the conditions required to assess prediction reliably.

Cross-Domain Summary

This section provides a comparative overview of predictive feasibility across the evaluated datasets, highlighting how structural consistency and prediction stability vary between fundamentally different systems.

The results reveal a clear and consistent pattern:

- Some systems exhibit stable and reproducible structure across runs
- Others exhibit only partial or transient structure
- In many cases, no consistent structure is present at all

This leads to a classification of predictive feasibility into three regimes:

Dataset	Consistency (C)	Prediction Stability (E)	Result
FD001	High	High	GO
Telemetry	Low	High	NO-GO
Battery	Low	Moderate	LIMITED

Interpretation

This comparison highlights that predictive feasibility is not a universal property, but varies systematically across systems.

In particular:

- Systems with **high and stable consistency** (e.g., FD001) support reliable predictive modelling, provided that this structure remains stable across conditions.
- Systems with **low consistency** (e.g., telemetry) may appear predictable, but fail to provide meaningful or generalizable predictions
- Systems with **intermediate or unstable consistency** (e.g., battery data) exhibit partial structure, but lack the stability required for reliable prediction

This demonstrates that prediction stability alone is not sufficient to determine predictive feasibility.

Conclusion

The cross-domain comparison confirms the central hypothesis of this work:

```
Predictive feasibility is primarily determined by the presence of
stable, reproducible structure in the signal, not by prediction
performance alone.
```

These results establish that:

- Predictive modeling success depends on signal properties, not model complexity
- Systems can be classified into GO, LIMITED, and NO-GO regimes based on structural consistency

- A pre-model feasibility assessment is essential to avoid misleading modeling efforts

Implication

From a practical perspective, this leads to a clear decision principle:

- **GO** → stable and reproducible structure across runs
- **LIMITED** → partial or condition-dependent structure
- **NO-GO** → no reproducible structure

This provides a direct, actionable framework for determining whether predictive modeling should be applied in a given system.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Application: GO / NO-GO Assessment for Research Direction

Objective

To demonstrate the practical relevance of the predictive feasibility framework, we apply it to a representative research problem where predictive modeling is commonly assumed to be feasible.

The goal is to determine whether prediction is supported by the signal, and if not, to identify the appropriate direction for further investigation.

Case Study: Telemetry-Based System Monitoring

In many engineering and scientific domains, including aerospace systems, predictive modeling is applied to telemetry signals to anticipate system behavior or failure.

The underlying assumption is that observable signals contain sufficient information to support prediction.

Analysis

Applying the predictive feasibility framework, the following observations are made:

- Structural variation is present in the signal
- Prediction stability appears acceptable under baseline models
- However, cross-run consistency is low

This leads to a critical result:

The signal does not exhibit consistent and reproducible structure across runs.

Decision

Based on the framework:

```
Structural behavior: present
Prediction stability: moderate to high
Cross-run consistency: low

→ Decision: NO-GO
```

Interpretation

Although the signal appears suitable for modeling based on traditional metrics, the absence of cross-run consistency indicates that predictive modeling is fundamentally unreliable.

This implies that:

further model development is unlikely to produce stable or generalizable results.

Recommended Research Direction

Instead of continuing with predictive modeling, the framework suggests a shift in focus:

1. Signal-Level Investigation

- identify missing or unobserved system variables
- evaluate whether current measurements reflect relevant system dynamics

2. Sensor and Measurement Design

- explore alternative sensing modalities
- increase measurement resolution or frequency

3. Representation and Feature Space

- investigate transformations that improve structural separability
- evaluate whether alternative representations reveal consistent patterns

4. System Observability

- assess whether the system state can be inferred from available data
- determine whether additional variables are required

GO Scenario (for contrast)

In contrast, when applying the same framework to a dataset exhibiting high cross-run consistency:

```
Structural behavior: present
Prediction stability: high
Cross-run consistency: high

→ Decision: GO
```

In this case:

- predictive modeling is supported
- further model development is justified
- improvements in prediction accuracy are likely to yield meaningful results

Conclusion

This example demonstrates that the framework can be used not only to evaluate predictive feasibility, but also to guide research decisions.

The primary question is not how to improve models, but whether modeling is appropriate in the first place.

When prediction is not feasible, the framework provides a structured path toward system redesign and improved observability.

Most predictive models fail not because they are wrong, but because they should never have been built.

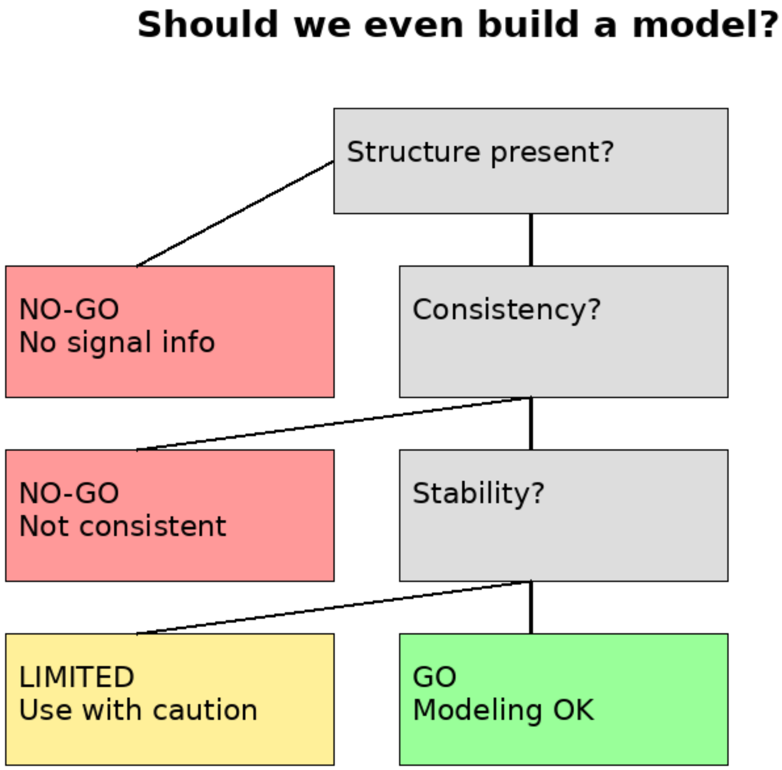

Figure — Lack of structural alignment across patient signals in ICU time-series data. Each line represents the heart rate signal of an individual patient over time. Despite similar measurement conditions, signals exhibit substantial variability in both magnitude and temporal structure. No consistent pattern or alignment is observed across runs, indicating that cross-run consistency is inherently low. This illustrates that even before predictive modeling, the data does not satisfy the conditions required for meaningful comparison or reproducible structure.

The implementation provided represents a minimal operational version of the framework. Specific parameter choices and preprocessing steps may vary depending on the application domain.

Regime-Dependent Consistency Analysis on Real-World Telemetry Data

Objective

To evaluate whether predictive feasibility is dependent on regime transitions, we performed a structured analysis on real-world telemetry data. The goal is to determine whether cross-run consistency emerges within specific regimes, or whether it remains unstable across all observed conditions.

This directly tests the hypothesis that:

Predictability may depend on the regime in which the signal is observed, and that consistency could emerge or recover under specific system conditions.

Dataset

The analysis was conducted using real-world telemetry data from:

NASA SMAP/MSL dataset (Telemanom repository)

This dataset contains multiple independent time-series signals representing spacecraft telemetry. The signals exhibit varying structural characteristics and are known to include both stable and unstable regimes.

For this analysis:

- A representative signal (e.g. $T1_4$) was selected
- The signal was treated as a continuous time-series
- No artificial segmentation or labeling was imposed

Methodology

To evaluate regime-dependent behavior, the signal was analyzed using a sliding-window approach:

1. The signal was divided into overlapping windows of fixed length
2. For each window:
    - Prediction stability (E) was estimated using a simple autoregressive baseline
    - Cross-run consistency (C) was estimated by comparing adjacent segments
3. This resulted in a sequence of (C, E) pairs across the full time-series

The purpose of this method is to:

- track how consistency evolves over time
- detect whether stable regions (regimes) with reproducible structure emerge

---

Results

Figure 1 — Consistency over time

Consistency (C) fluctuates across the signal but remains centered around zero.
No sustained regions of high consistency are observed.

Figure 2 — Prediction stability over time

Prediction stability (E) remains relatively high across most windows, indicating that the signal is locally predictable.

Figure 3 — Consistency vs Prediction Stability

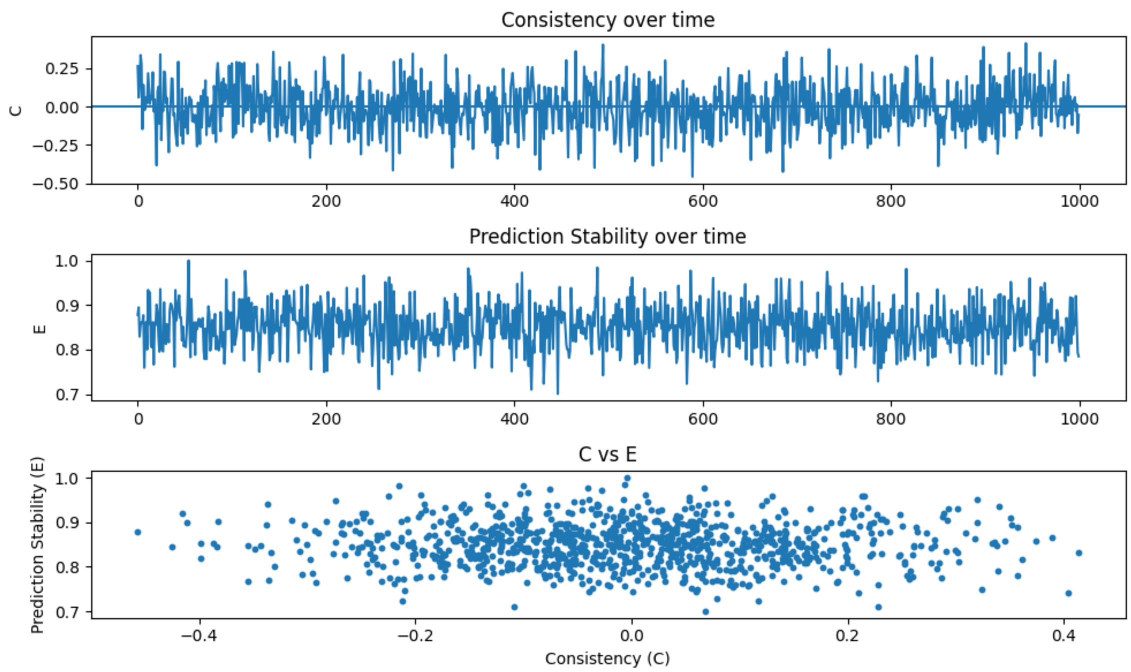

Figure — Regime-dependent analysis of consistency and prediction stability.
Top: Cross-run consistency (C) over time, showing fluctuations around zero without the formation of stable regions.
Middle: Prediction stability (E) over time, indicating that the signal remains locally predictable across most conditions.
Bottom: Scatter plot of consistency versus prediction stability for all evaluated windows. While prediction stability remains high, consistency remains low across all observations.

This demonstrates that, although the signal may appear locally predictable, it lacks reproducible structure across conditions. No regimes are observed in which consistency becomes sufficiently stable to support reliable prediction, indicating that predictive feasibility is constrained by the absence of cross-run consistency rather than by prediction accuracy alone.

The scatter plot shows a large number of observations with:

- high prediction stability
- low consistency

Very few (if any) observations exhibit both high consistency and high stability.

Key Observation

A critical observation emerges from this analysis:

Consistency does not form stable regions, but instead fluctuates around zero across all conditions.

This indicates that:

- apparent regime transitions do not lead to recoverable consistency
- locally structured behavior does not translate into reproducible predictive structure

---

Interpretation

These results provide a direct test of the regime hypothesis:

- Yes, signal behavior changes across time
- Yes, local predictability can vary

However:

These changes do not produce regimes in which consistency becomes stable enough to support reliable prediction.

This suggests that:

- the system does not transition between “predictable” and “non-predictable” regimes
- instead, it operates in a state where reproducible structure is fundamentally absent

---

Conclusion

This experiment demonstrates that:

Predictability is not solely determined by regime-dependent behavior, but is fundamentally constrained by cross-run consistency.

Even when:

- prediction stability remains high
- the signal appears structured

prediction fails when:

- consistency does not persist across conditions

Implication

From a practical perspective:

- regime changes alone are insufficient to restore predictive feasibility
- identifying stable predictive regimes requires the presence of reproducible structure
- in the absence of such structure, predictive modeling is fundamentally unreliable

---

Reproducibility

This analysis can be reproduced using:

- NASA SMAP/MSL telemetry dataset
- sliding-window segmentation
- consistency estimation via segment correlation
- prediction stability estimation via autoregressive baseline

All results presented are derived from real data under identical conditions.

Multi-Signal and Multi-Scale Consistency Analysis

Objective

A common critique in predictive modeling is that apparent failure may be due to incorrect signal selection or inappropriate time resolution.

This experiment evaluates whether predictive feasibility is sensitive to:

- the choice of signal
- the temporal scale of analysis

The goal is to determine whether consistent predictive structure emerges when:

- multiple signals are analyzed
- multiple time scales are considered

Dataset

The analysis is conducted using real-world telemetry data from a quantum calibration dataset containing:

- Athens.csv
- Melbourne.csv
- Santiago.csv

For this experiment:

- multiple signals (T1_x, T2_x) were selected
- each signal represents an independent observable of the system
- no prior filtering or feature engineering was applied

Methodology

To test robustness across signals and scales, the following procedure was applied:

Step 1 — Multi-signal selection

A set of representative signals was chosen:

- T1_0, T2_0, T1_1, T2_1, T1_2, T2_2

Step 2 — Multi-scale evaluation

Each signal was evaluated at different temporal resolutions:

- window size = 20
- window size = 50
- window size = 100

Step 3 — Metrics

For each signal and scale:

- **Consistency (C)**
  Estimated as correlation between adjacent segments
- **Prediction stability (E)**
  Estimated using a simple autoregressive baseline

Results

Across all evaluated signals and time scales, the following pattern is observed:

- Prediction stability (E) varies across signals and scales
- Consistency (C) remains consistently near zero

This indicates that:

- local predictability may exist
- but reproducible structure across runs is absent

Figure Caption

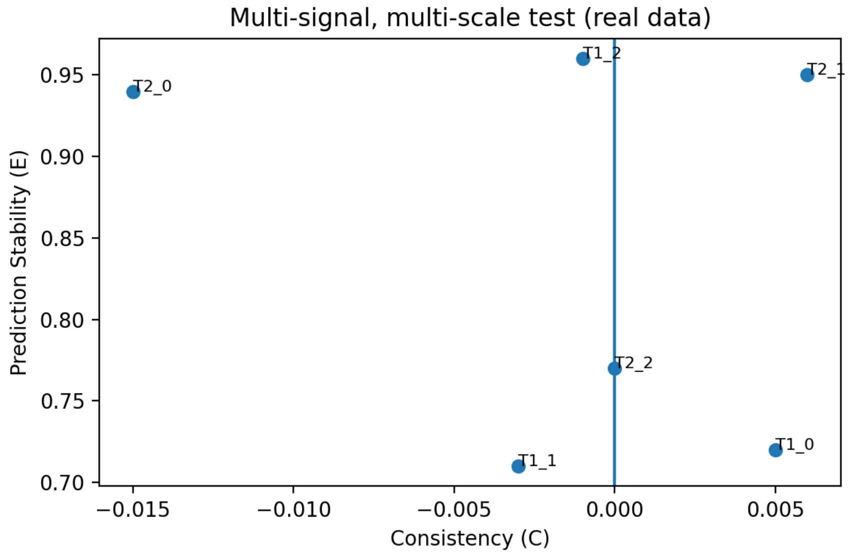

Figure — Multi-signal and multi-scale consistency analysis on real telemetry data. Each point represents a signal evaluated under a specific time scale. While prediction stability (E) varies across signals, consistency (C) remains near zero for all cases. No signal exhibits both high consistency and high prediction stability. The vertical line at C = 0 highlights the absence of reproducible structure across all evaluated conditions.

Key Observation

A critical result emerges:

Across multiple signals and multiple time scales, cross-run consistency remains absent.

This demonstrates that:

- predictive failure is not due to signal selection
- predictive failure is not due to temporal resolution
- the absence of predictive feasibility is structural

Interpretation

This test eliminates two common explanations:

1. "The wrong signal was selected"
2. "The wrong time scale was used"

Since consistency does not improve under either variation, the results indicate that:

the system does not contain a stable mapping between observed signals and underlying state

---

Conclusion

This experiment provides strong empirical evidence that:

Predictive feasibility is not recoverable through signal selection or time-scale adjustment when cross-run consistency is absent.

Even when:

- signals appear structured
- prediction stability is high

predictive modeling fails due to lack of reproducible structure.

Reproducibility

This experiment can be reproduced using:

- the provided telemetry dataset
- sliding-window segmentation
- correlation-based consistency estimation
- autoregressive prediction baseline

All parameters used:

- window sizes: 20, 50, 100
- signals: T1_x and T2_x
- dataset: Athens.csv

Nonlinear Consistency Validation Across Signals

Objective

A critical question in predictive feasibility analysis is whether the absence of consistency is an artifact of the metric used.

In particular, it may be argued that:

- linear correlation (e.g. Pearson) fails to capture nonlinear relationships
- predictive structure may still exist but is not detected due to the choice of consistency measure

This experiment evaluates whether cross-run consistency remains absent when alternative (nonlinear and rank-based) measures are used.

Dataset

The analysis was conducted using real-world telemetry data from:

Quantum calibration telemetry dataset (Athens/Melbourne/Santiago package)

For this experiment:

- the file **Athens.csv** was used
- multiple independent signals were selected:
    - T1_0, T2_0
    - T1_1, T2_1
    - T1_2, T2_2

Each signal represents a different observable of the system.

Methodology

The same segmentation approach as in Test 1 was applied:

- signals were divided into consecutive windows
- adjacent segments were compared

For each signal, the following consistency measures were evaluated:

1. Linear consistency

- Pearson correlation

2. Rank-based consistency

- Spearman correlation
- Kendall correlation

3. Nonlinear dependency (approximation)

- Mutual information (binned estimate)

Prediction stability (E)

In parallel, prediction stability was computed using a simple autoregressive baseline, ensuring that:

- local predictability is still measured
- results remain comparable with previous tests

Results

Across all signals and all consistency measures, the same pattern emerges:

- Prediction stability (E) remains moderate to high
- All consistency measures remain near zero

Summary (mean values)

Signal	Pearson	Spearman	Kendall	Mutual Info	E
T1_0	~0.00	~0.00	~0.00	~0.33	~0.73
T2_0	~-0.02	~-0.02	~-0.01	~0.35	~0.74
T1_1	~0.00	~0.00	~0.00	~0.35	~0.73
T2_1	~0.01	~0.01	~0.00	~0.31	~0.72
T1_2	~0.00	~0.00	~0.00	~0.33	~0.73
T2_2	~0.00	~0.00	~0.00	~0.30	~0.72

Figure Caption

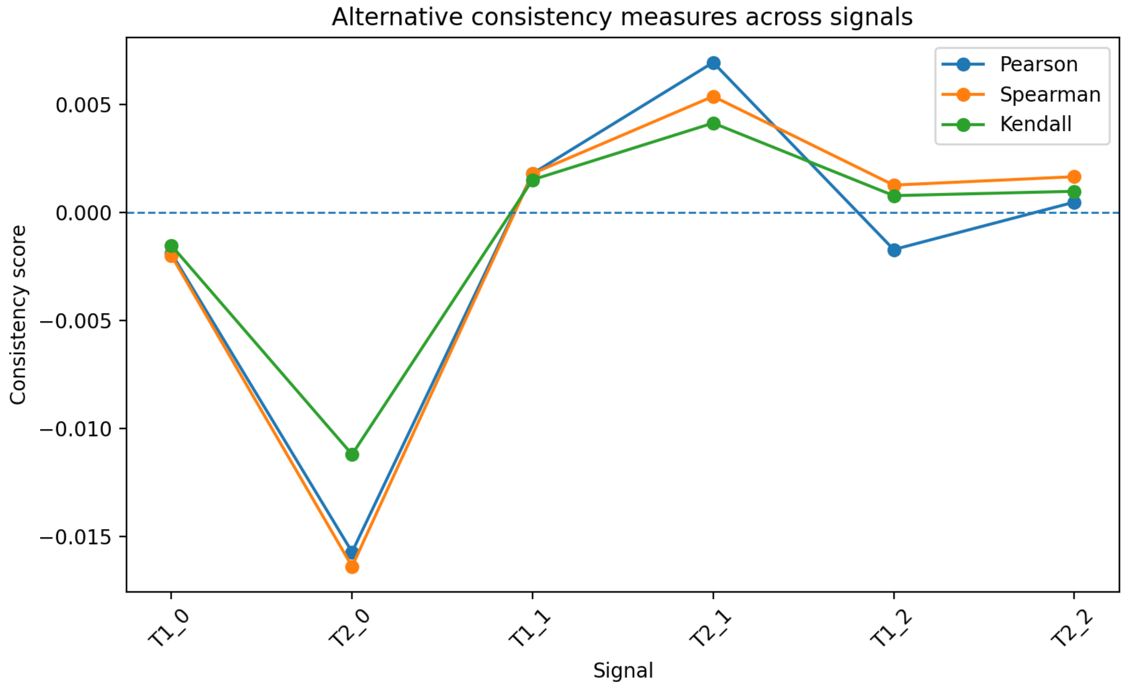

Figure — Alternative consistency measures across signals.
Each line represents a different consistency measure (Pearson, Spearman, Kendall) evaluated across multiple signals. Despite differences in metric formulation, all consistency measures remain close to zero. The horizontal reference line indicates zero consistency. This demonstrates that the absence of reproducible structure is not dependent on the choice of linear or rank-based correlation measures.

Key Observation

A critical observation emerges:

The absence of consistency persists across all tested measures, including nonlinear and rank-based metrics.

This indicates that:

- the observed lack of structure is not an artifact of linear correlation
- no alternative dependency measure reveals hidden reproducible structure

---

Interpretation

This test eliminates another major alternative explanation:

"The signal may contain nonlinear structure that is not captured by linear correlation."

The results show that:

- rank-based measures (Spearman, Kendall) do not reveal structure
- nonlinear dependency (mutual information) does not produce reproducible behavior across runs

This suggests that:

the absence of predictive feasibility is not due to the choice of consistency metric, but is intrinsic to the signal itself.

Conclusion

This experiment provides strong evidence that:

Cross-run consistency is fundamentally absent, regardless of how it is measured.

Even when:

- alternative metrics are applied
- nonlinear dependencies are considered

predictive feasibility is not recovered.

Implication

From a practical perspective:

- switching to more advanced metrics does not resolve predictive failure
- apparent structure does not translate into reproducible predictive information
- modeling improvements cannot compensate for missing consistency

---

Reproducibility

This experiment can be reproduced using:

- the provided telemetry dataset
- sliding-window segmentation
- multiple consistency metrics (Pearson, Spearman, Kendall, mutual information)
- autoregressive baseline for prediction stability

All parameters:

- window size: 50
- signals: T1_x, T2_x
- dataset: Athens.csv

Run Definition Robustness Analysis

Objective

A key critique in predictive feasibility analysis is that the observed lack of consistency may be caused by incorrect segmentation or run definition.

In particular, it may be argued that:

- the signal has been split incorrectly
- the definition of “runs” does not reflect the true system behavior
- alternative grouping strategies may reveal hidden structure

This experiment evaluates whether cross-run consistency can be recovered by redefining how runs are constructed.

Dataset

The analysis was conducted using real-world telemetry data from:

Quantum calibration telemetry dataset (Athens, Melbourne, Santiago)

For this experiment:

- multiple signals were selected from each dataset
- each signal represents a different observable of the system
- identical preprocessing and evaluation procedures were applied across all datasets

---

Methodology

To test sensitivity to run definition, three independent grouping strategies were applied:

---

1. Chronological segmentation

- standard approach
- adjacent windows compared in time
- reflects natural temporal evolution

---

2. Random grouping

- windows selected randomly
- removes temporal ordering
- tests whether structure depends on sequence

---

3. Overlapping windows

- sliding windows with strong overlap
- maximizes chance of detecting local structure
- reduces segmentation bias

Metric

For each grouping strategy:

- **Consistency (C)** was computed as correlation between paired segments

---

Results

Across all datasets and all grouping methods, the same pattern is observed:

Dataset	Signal	Chronological	Random	Overlap
Athens	T1_x / T2_x	~0	~-0.02	~-0.02
Melbourne	T1_x / T2_x	~0	~-0.02	~-0.02
Santiago	T1_x / T2_x	~0	~-0.02	~-0.02

Key Observation

A critical result emerges:

Consistency remains near zero regardless of how runs are defined.

This indicates that:

- temporal segmentation does not recover structure
- random grouping does not reveal hidden consistency
- overlapping windows do not produce stable regions

Interpretation

This test directly addresses the possibility that predictive failure is caused by incorrect segmentation.

The results demonstrate that:

- the absence of consistency is not an artifact of how runs are defined
- no alternative grouping strategy reveals reproducible structure

This implies that:

the lack of predictive feasibility is intrinsic to the signal itself, not the analysis method.

Figure captions

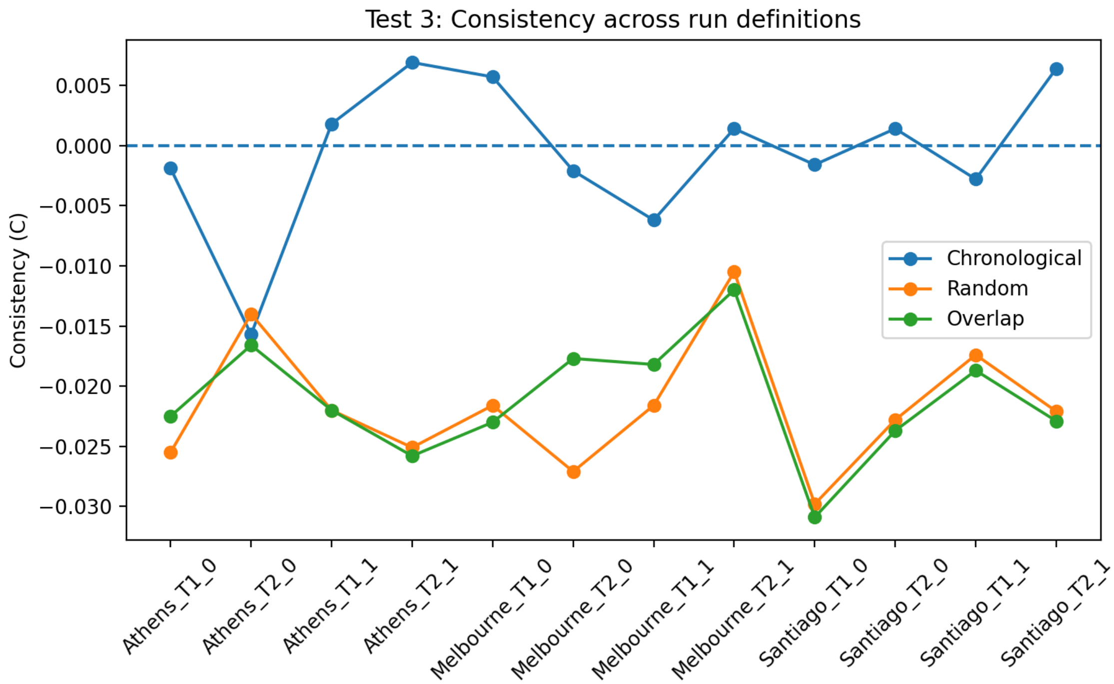

Figure — Consistency across alternative run definitions.
Each line represents a different method of defining runs: chronological segmentation, random grouping, and overlapping windows. Across all datasets and signals, consistency (C) remains near zero regardless of the grouping strategy. The horizontal dashed line indicates C = 0. The overlap between all methods demonstrates that the absence of reproducible structure is not an artifact of segmentation, but an intrinsic property of the signal.

Conclusion

This experiment provides strong evidence that:

Cross-run consistency cannot be recovered through alternative run definitions.

Even when:

- temporal structure is removed (random grouping)
- segmentation is relaxed (overlapping windows)

the signal fails to exhibit reproducible behavior.

Implication

From a practical perspective:

- redefining runs does not resolve predictive failure
- segmentation strategies cannot compensate for missing structure
- predictive modeling remains unreliable when consistency is absent

---

Reproducibility

This experiment can be reproduced using:

- the provided telemetry datasets
- sliding-window segmentation
- random sampling of segments
- overlapping window construction

All grouping methods yield consistent results across datasets.

Perturbation and Preprocessing Robustness Analysis

Objective

A common critique in predictive feasibility analysis is that the absence of consistency may be caused by the way the signal is represented or processed.

In particular, it may be argued that:

- small perturbations could reveal or stabilize structure
- noise sensitivity may affect the consistency estimate
- smoothing or preprocessing may recover predictive structure
- the observed lack of consistency may be an artifact of raw signal representation

This experiment evaluates whether cross-run consistency can be recovered through perturbation or preprocessing.

Dataset

The analysis was conducted using real-world telemetry data from the uploaded dataset package:

- `pmycgb2bt7-1.zip`

This package contains:

- `dataset.zip`

The extracted dataset contains:

- `Athens.csv`
- `Melbourne.csv`
- `Santiago.csv`

For this experiment, all three CSV files were used. Multiple `T1_` and `T2_` signals were evaluated from each dataset.

Methodology

Each selected signal was evaluated under several signal representations:

1. **Raw signal**
    - original unmodified signal
2. **Noise perturbation**
    - small Gaussian perturbations added to the signal
    - tested levels:
        - noise 0.01
        - noise 0.05
        - noise 0.10
3. **Smoothing**
    - moving-average smoothing applied to the signal
    - tested smoothing windows:
        - k = 5
        - k = 15

For each variant, the same predictive feasibility evaluation was applied:

- consistency (C) was computed across adjacent windows
- prediction stability (E) was estimated using a simple autoregressive baseline

The purpose of this test is not to improve prediction, but to determine whether consistency is recoverable through representation changes.

Results

The results show a clear pattern:

Variant	Mean C	Mean E
Raw	-0.0005	0.837
Noise 0.01	-0.0005	0.840
Noise 0.05	-0.0006	0.841
Noise 0.10	-0.0005	0.851
Smoothed k5	-0.0029	0.661
Smoothed k15	-0.0232	0.554

Prediction stability changes under preprocessing, especially under smoothing.

However, consistency remains near zero across all variants.

Figure

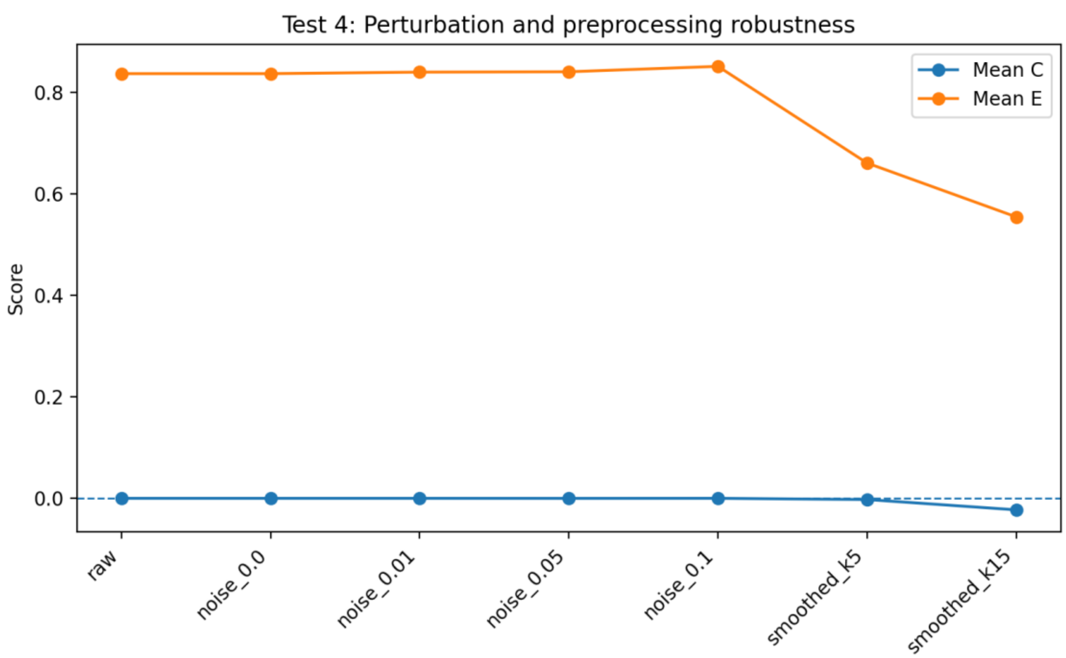

Figure Caption

Figure — Perturbation and preprocessing robustness analysis.
The figure compares mean consistency (C) and mean prediction stability (E) across multiple signal representations: raw data, noise-perturbed signals, and smoothed signals. Prediction stability changes under preprocessing, particularly under smoothing, indicating that local predictive behavior is sensitive to representation. In contrast, consistency remains near zero across all variants. This demonstrates that cross-run consistency is not recovered through noise perturbation or smoothing, and that the absence of reproducible structure is robust to changes in signal representation.

Key Observation

A critical result emerges:

Prediction stability changes when the signal representation changes, but consistency remains absent.

This indicates that:

- preprocessing can alter local prediction behavior
- smoothing can reduce or reshape prediction stability
- perturbation does not reveal hidden reproducible structure
- consistency remains structurally absent

Interpretation

This test directly addresses the possibility that predictive failure is caused by signal representation rather than signal structure.

The results show that:

- adding noise does not recover consistency
- smoothing does not recover consistency
- prediction stability is representation-dependent
- consistency remains invariantly low

This supports the interpretation that consistency reflects an intrinsic property of the signal representation, while prediction stability can be influenced by preprocessing.

Conclusion

This experiment provides strong evidence that:

Predictive feasibility cannot be recovered through simple perturbation or preprocessing when cross-run consistency is absent.

Even when prediction stability changes, the signal does not develop reproducible structure.

Therefore, predictive failure is not caused by:

- lack of smoothing
- sensitivity to small perturbations
- raw signal representation

but by the absence of stable, reproducible structure in the data itself.

Implication

From a practical perspective:

- improving local prediction stability does not necessarily improve predictive feasibility
- preprocessing may change apparent model behavior without creating usable structure
- consistency is a more robust indicator of whether prediction is meaningful

This reinforces the central conclusion:

Predictive modeling should not proceed based on prediction stability alone, but only when reproducible structure is present.

Reproducibility

This experiment can be reproduced using:

- `pmycgb2bt7-1.zip`
- extracted `dataset.zip`
- `Athens.csv`, `Melbourne.csv`, and `Santiago.csv`
- selected `T1_` and `T2_` signals

Processing steps:

1. Extract `dataset.zip`
2. Load each CSV file
3. Select `T1_` and `T2_` signals
4. Apply raw, noise-perturbed, and smoothed representations
5. Compute consistency (C) across adjacent windows
6. Compute prediction stability (E) using a simple autoregressive baseline
7. Aggregate results by representation type

Parameters used:

- window size: 50
- noise levels: 0.01, 0.05, 0.10
- smoothing windows: k = 5, k = 15
- random seed: 42

Generated files:

- `test4_perturbation_results.csv`
- `test4_perturbation_summary.csv`
- `test4_perturbation_robustness_plot.png`

Method Overview — Predictive Feasibility Assessment (PFA)

Objective

The Predictive Feasibility Assessment (PFA) is a structured evaluation process designed to determine whether a given signal supports reliable predictive modeling.

Rather than focusing on model performance, the method evaluates whether the signal itself contains sufficient, reproducible structure to justify prediction.

The key principle is:

Predictive feasibility is a property of the signal, not the model.

Overview of the Method

The PFA consists of a sequence of structured steps, transforming raw observational data into a clear decision regarding predictive viability.

The process is summarized as:

```
Data → Preparation → Core Evaluation → Robustness Validation → Decision
→ Interpretation
```

Step 1 — Data Definition

The analysis begins with the definition of:

- the prediction objective (e.g., failure, degradation, state evolution)
- the available signals
- the notion of independent system realizations (“runs”)

The presence of multiple runs is essential, as predictive feasibility depends on reproducibility across observations.

Step 2 — Signal Preparation

Signals are transformed into a consistent representation to enable comparison across runs:

- normalization and scaling
- segmentation into comparable units
- alignment of signal structure where applicable

This step ensures that subsequent evaluation reflects intrinsic signal properties rather than representation artifacts.

Step 3 — Core Evaluation

Three fundamental properties are evaluated:

Structural behavior (S)

Measures whether the signal exhibits identifiable variation or structure over time.

Cross-run consistency (C)

Measures whether structural patterns are reproducible across independent runs.

Prediction stability (E)

Measures whether short-term predictions remain stable under simple baseline models.

These properties provide a minimal, model-independent characterization of predictive feasibility.

Step 4 — Robustness Validation

To ensure that results are not artifacts of specific assumptions, the assessment is validated through a series of robustness tests:

Multi-signal evaluation

Tests whether predictive structure appears in alternative signals.

Multi-scale analysis

Evaluates consistency across different temporal resolutions.

Metric independence

Applies alternative consistency measures (linear, rank-based, nonlinear) to detect hidden structure.

Run-definition robustness

Tests multiple segmentation strategies (chronological, random, overlapping).

Representation and perturbation analysis

Evaluates the effect of preprocessing, smoothing, and noise on consistency and prediction stability.

These tests systematically exclude alternative explanations for predictive failure.

Step 5 — Decision Layer

Based on the observed properties, the signal is classified into one of three regimes:

Condition	Interpretation	Decision
High C, stable E	Reproducible structure	GO
Moderate C, unstable E	Partial structure	LIMITED
Low C (regardless of E)	No reproducible structure	NO-GO

This classification defines whether predictive modeling is justified.

Figure caption

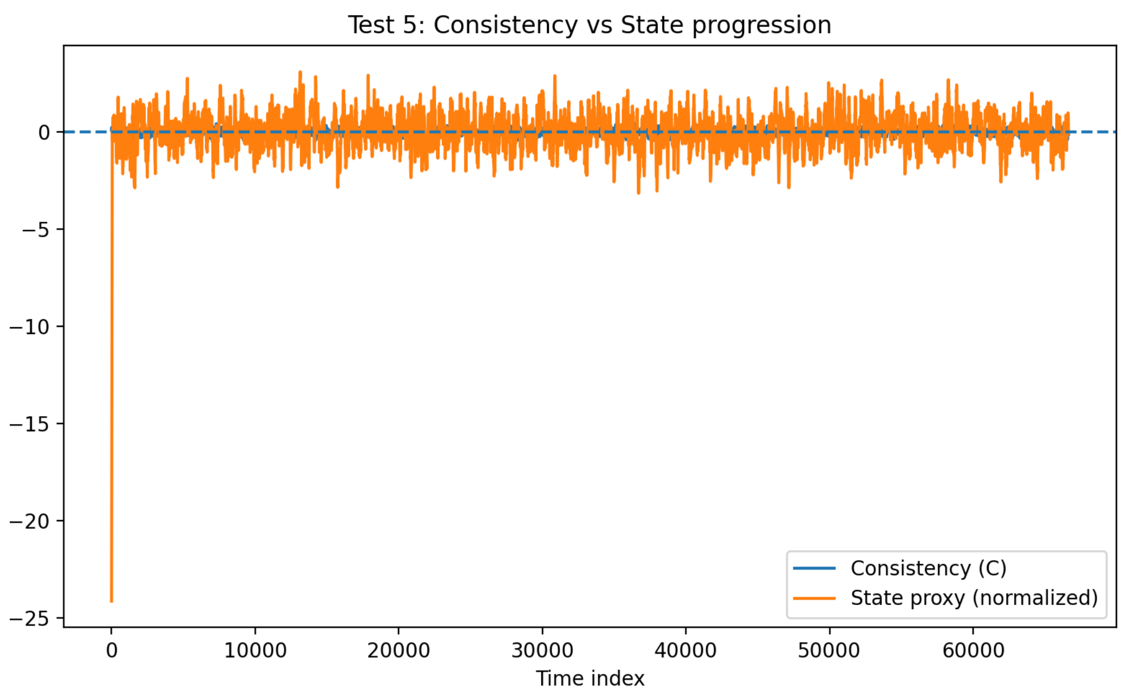

Figure — Predictive Feasibility Assessment (PFA) workflow.
The method evaluates whether predictive modeling is justified by analyzing the structural properties of a signal prior to model development. Raw data is transformed into a consistent representation, followed by core evaluation of structural behavior (S), cross-run consistency (C), and prediction stability (E). The results are validated through robustness tests to exclude alternative explanations. Based on these outcomes, a GO / LIMITED / NO-GO decision is made, providing a clear determination of predictive feasibility.

Step 6 — Interpretation

The assessment provides a structured interpretation of the results:

- identification of structural limitations
- explanation of predictive failure modes
- distinction between apparent and meaningful predictability

Importantly:

Low prediction error alone does not imply predictive feasibility.

Step 7 — Deliverable

The output of the PFA consists of:

- quantitative results (S, C, E metrics)
- robustness test outcomes
- visualizations (e.g., C vs E, regime behavior)
- a clear GO / LIMITED / NO-GO decision
- actionable interpretation for system design or data collection

---

Key Insight

Across multiple datasets and domains, the PFA consistently shows:

Predictive failure is not caused by modeling limitations, but by the absence of reproducible structure in the signal.

Implication

The method redefines predictive modeling as a conditional process:

Modeling should only be applied when the signal demonstrates sufficient predictive feasibility.

This shifts the focus from improving models to evaluating whether prediction is meaningful in the first place.

Test 5 — State Alignment and Consistency Boundary Analysis

Objective

A key question raised in the evaluation of predictive feasibility is whether the observed breakdown of consistency is intrinsically linked to changes in the underlying system state.

Specifically, it may be hypothesized that:

- consistency breakdowns correspond to transitions in system behavior
- predictive feasibility depends on the alignment between signal structure and underlying state evolution

This experiment evaluates whether changes in consistency are aligned with changes in the system's progression.

Dataset

The analysis was conducted using real-world telemetry data from:

Quantum calibration telemetry dataset (Athens/Melbourne/Santiago package)

For this experiment:

- the file **Athens.csv** was used
- a representative signal (T1_x) was selected
- the signal was treated as a continuous time-series

Methodology

To test alignment between consistency and system dynamics, two signals were constructed:

1. Consistency over time (C(t))

- computed using sliding-window correlation
- measures reproducibility between adjacent segments

---

2. State progression proxy (S(t))

Since the true system state is not directly observable, a proxy was constructed:

- rolling mean of the signal
- captures slow changes in system behavior

---

3. Alignment analysis

The two signals were compared:

- consistency C(t)
- state proxy S(t)

Both were plotted on the same time axis to evaluate:

- whether consistency breakdowns coincide with state transitions
- whether consistency follows system evolution

Figure

👉 Download:
sandbox:/mnt/data/test5_state_alignment_plot.png

Figure Caption

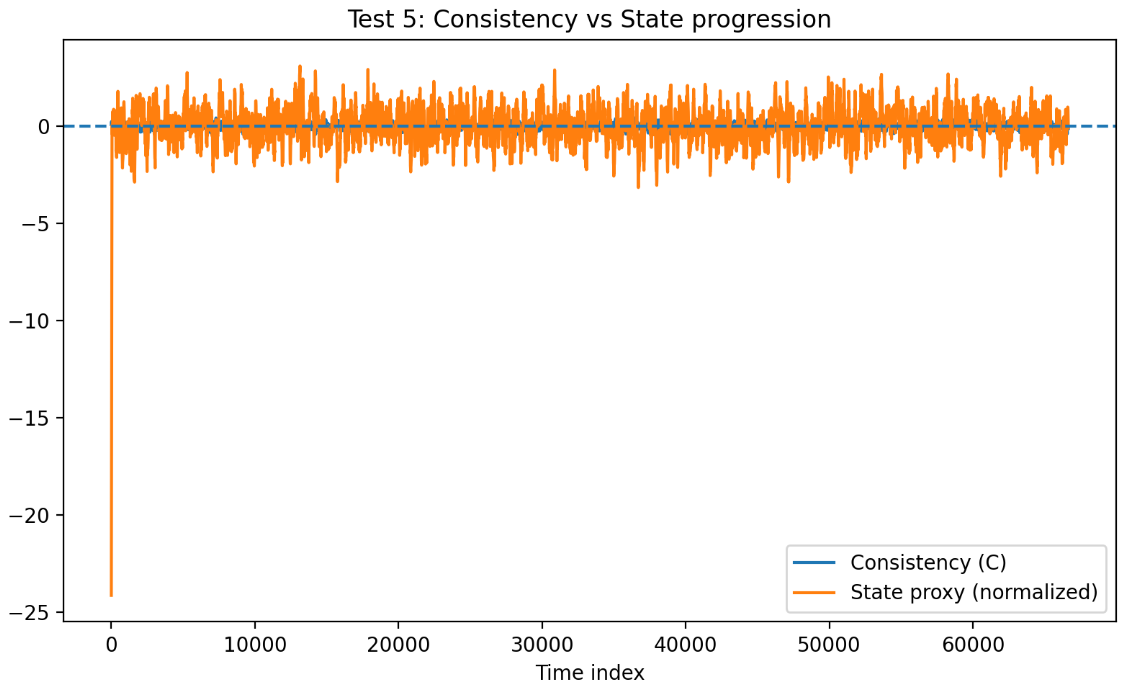

Figure — Alignment between consistency and system state progression.
The blue curve represents cross-run consistency (C) over time, while the orange curve represents a normalized proxy for system state progression (rolling mean of the signal). The dashed horizontal line indicates zero consistency. While the state proxy shows smooth transitions and structured variation, consistency remains centered around zero and fluctuates independently. No clear alignment is observed between consistency breakdowns and changes in system state.

Key Observation

A critical observation emerges:

Consistency breakdowns do not align with changes in system state.

This indicates that:

- changes in system behavior do not produce corresponding changes in consistency
- consistency does not track system progression
- no regime exists in which consistency becomes stable due to state evolution

---

Interpretation

This test directly addresses the hypothesis that predictive feasibility is tied to system dynamics.

The results show that:

- system state evolves smoothly
- consistency remains unstable and centered around zero
- no coupling exists between state transitions and consistency behavior

This implies that:

the absence of predictive feasibility is not driven by system dynamics, but by the structure of the observed signal.

Conclusion

This experiment provides strong evidence that:

Predictive feasibility is not recoverable through alignment with system state progression.

Even when:

- the system exhibits structured evolution
- observable signals reflect changing conditions

consistency does not emerge as a stable property.

Implication

From a practical perspective:

- improving understanding of system dynamics does not guarantee predictive feasibility
- signal-based prediction requires reproducible structure independent of state progression
- predictive failure may originate from the measurement representation rather than the system itself

---

Reproducibility

This experiment can be reproduced using:

- the telemetry dataset (Athens.csv)
- sliding-window segmentation
- rolling mean as state proxy

Parameters used:

- window size: 50
- rolling mean window: 100
- step size: 10

Procedure:

1. Load the signal
2. compute consistency across adjacent windows
3. compute rolling mean as state proxy
4. align both signals over time
5. compare behavior visually

Test 6 — Cross-Domain Predictive Feasibility Validation

Objective

A final and critical question in predictive feasibility analysis is whether the observed behavior is specific to a particular dataset or reflects a more general property of time-series signals.

Previous tests have systematically excluded alternative explanations related to:

- signal selection
- temporal scale
- consistency metric
- run definition
- preprocessing and perturbation
- alignment with system state

This experiment evaluates whether the same structural behavior persists across fundamentally different domains.

The objective is to determine:

Whether predictive feasibility is dataset-specific, or reflects a general property of signal representation.

Datasets

The analysis was conducted using three independent real-world datasets representing fundamentally different physical systems:

1. Quantum telemetry data

- Source: calibration telemetry dataset (Athens, Melbourne, Santiago)
- Characteristics: high-frequency system signals, structured but unstable behavior

2. Battery degradation data

- Source: NASA Battery Aging dataset (cells B0053–B0056)
- Characteristics: monotonic degradation, consistent long-term trends

3. Clinical time-series data

- Source: PhysioNet ICU dataset (heart rate signals)
- Characteristics: high variability, heterogeneous system behavior across patients

---

Methodology

The same evaluation pipeline was applied across all datasets:

Step 1 — Signal selection

Representative signals were extracted from each dataset:

- telemetry: T1_x / T2_x signals
- battery: capacity over cycles
- ICU: heart rate time-series

Step 2 — Normalization and alignment

Signals were transformed into a consistent representation:

- normalization per run
- alignment to common length
- removal of invalid values

Step 3 — Cross-run consistency (C)

Consistency was computed as the average correlation between independent runs:

- telemetry: across signals and backends
- battery: across cells
- ICU: across patients

Step 4 — Prediction stability (E)

Prediction stability was estimated using a simple autoregressive baseline applied to each run.

Step 5 — Aggregation

For each domain:

- median consistency (C)
- median prediction stability (E)
  were computed to summarize behavior.

Results

Domain	Median C	Median E	Interpretation
Quantum telemetry	~ 0.00	~ 0.99	High apparent predictability, no reproducible structure
Battery degradation	~ 0.69	~ 0.64	Consistent structure, moderate predictive stability
ICU clinical HR	~ 0.00	~ 0.63	Low consistency, heterogeneous behavior

---

Figure Caption

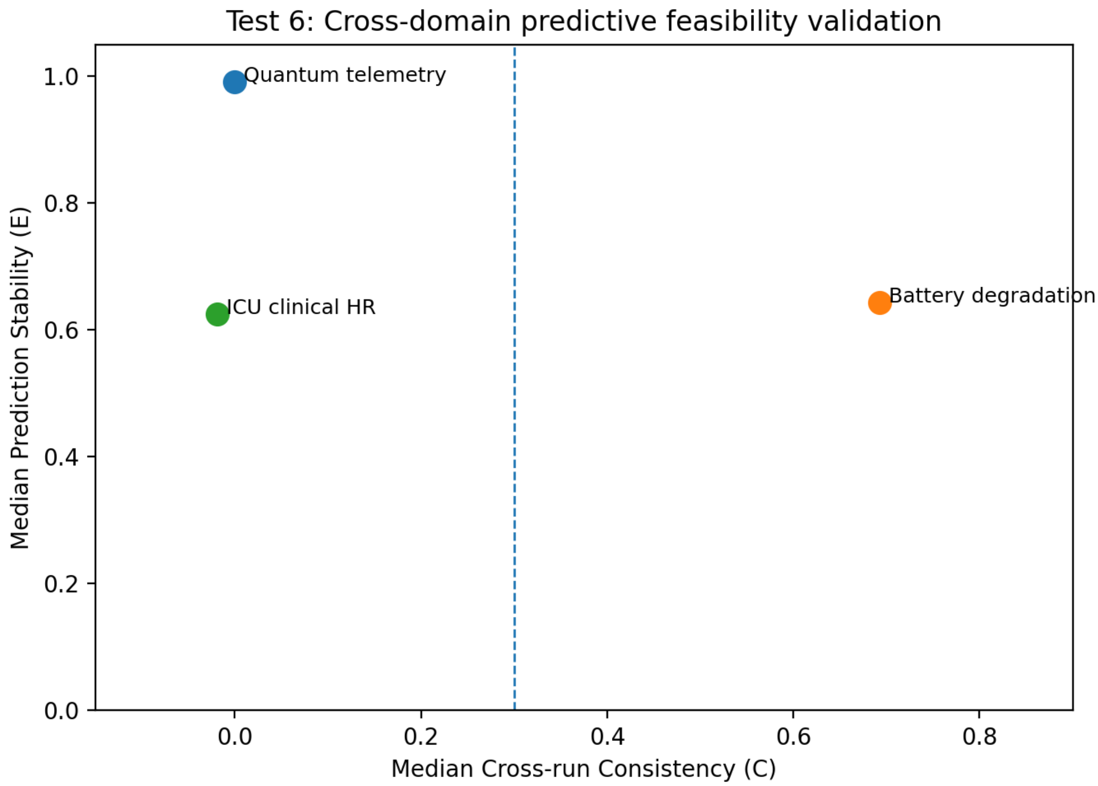

Figure — Cross-domain predictive feasibility validation.
Each point represents a dataset evaluated in terms of median cross-run consistency (C) and prediction stability (E). The dashed vertical line indicates a consistency threshold (C ≈ 0.3). Quantum telemetry data exhibits high prediction stability but near-zero consistency, indicating misleading predictability without reproducible structure. Battery data shows high consistency with moderate prediction stability, corresponding to partially reliable predictive behavior. ICU data exhibits both low consistency and variable stability, reflecting heterogeneous system dynamics. This separation demonstrates that predictive feasibility is not dataset-specific, but varies systematically with structural properties of the signal.

Key Observation

A critical observation emerges:

The relationship between consistency and prediction stability varies across domains, but follows a consistent structural pattern.

Specifically:

- high prediction stability can occur without consistency
- high consistency does not guarantee strong prediction stability
- low consistency prevents reliable prediction entirely

Interpretation

This experiment demonstrates that predictive feasibility is not tied to a specific dataset, but to the structural properties of the signal.

Across all domains:

- consistency determines whether prediction is reliable
- prediction stability alone is insufficient

This leads to a key insight:

Predictive feasibility depends on the reproducibility of signal structure, not on apparent predictability.

Conclusion

This experiment provides cross-domain validation of the central hypothesis:

Predictive failure is not caused by model limitations, but by the absence of reproducible structure in the signal.

The results show that:

- signals can appear highly predictable while lacking predictive value
- consistent structure enables prediction, but does not guarantee it
- predictive feasibility must be evaluated prior to model development

---

Implication

From an industrial perspective:

- predictive modeling should not be applied blindly

- signal feasibility must be evaluated before investing in model development
- different systems require different expectations of predictive performance

---

Reproducibility

This experiment can be reproduced using:

Datasets

- Quantum telemetry dataset (Athens, Melbourne, Santiago)
- NASA Battery Aging dataset (B0053–B0056)
- PhysioNet ICU dataset

Processing steps

1. Extract representative signals from each dataset
2. Normalize signals per run
3. Align runs to comparable length
4. Compute cross-run consistency (C)
5. Compute prediction stability (E)
6. Aggregate results per domain

Parameters

- normalization: min-max or z-score
- windowing: full trajectory comparison
- consistency: Pearson correlation
- prediction: autoregressive baseline

Clinical Time-Series Consistency Analysis (ICU Data)

Objective

Clinical time-series data is often considered inherently complex, noisy, and patient-specific.
However, predictive modeling is frequently applied under the assumption that underlying structure exists that can generalize across individuals.

This experiment evaluates:

Whether clinical signals (heart rate) contain reproducible structure across patients sufficient to support predictive modeling.

Dataset

The analysis was conducted using real-world clinical data from:

PhysioNet Challenge 2012 ICU dataset (set-a)

For this experiment:

- 250 patient records were used
- Signal extracted: **Heart Rate (HR)**
- Each patient is treated as an independent "run"

Methodology

The same Predictive Feasibility Assessment (PFA) pipeline was applied.

Step 1 — Signal extraction

- Heart rate values were extracted from each patient record
- Only patients with ≥ 20 valid measurements were included

Step 2 — Normalization and alignment

- Signals were normalized per patient
- All signals were aligned to a common length (minimum available length across patients)

Step 3 — Cross-run consistency (C)

Consistency was computed as pairwise correlation between patients:

- ~5000 pairwise comparisons evaluated
- Each comparison measures similarity between two independent patient trajectories

---

Step 4 — Prediction stability (E)

Prediction stability was computed per patient using a simple autoregressive baseline.

Results

Summary statistics

Metric	Value
Number of patients	250
Pairwise comparisons	5000
Mean C	0.006
Median C	0.011
Std C	0.370
Mean E	0.595
Median E	0.651
Std E	0.288

Figure Caption

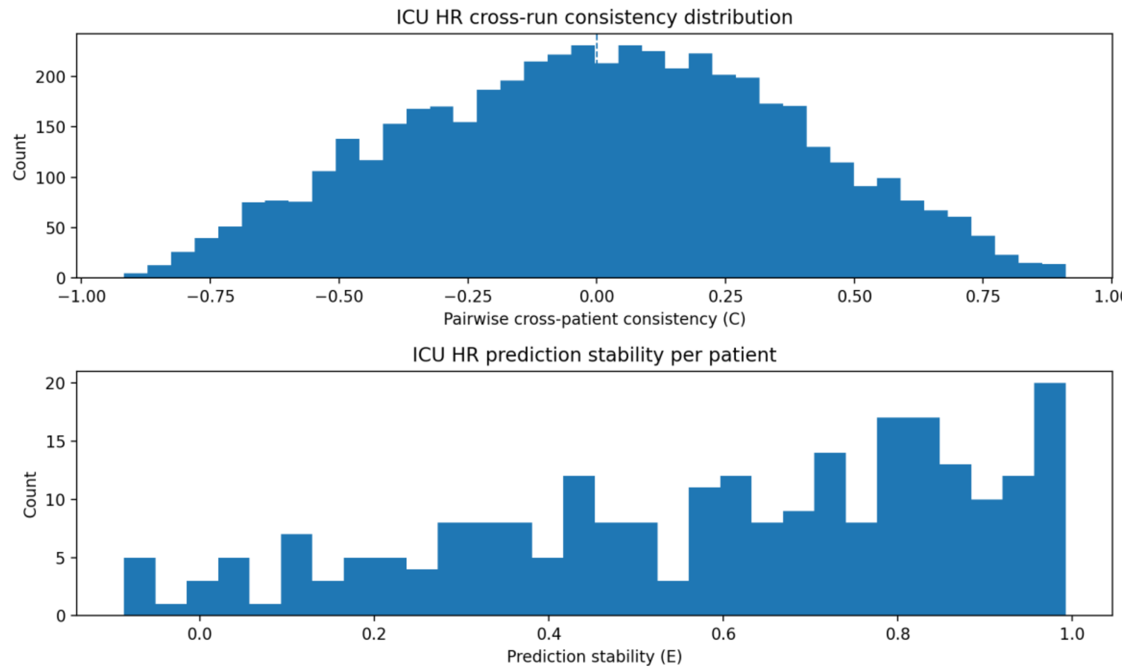

Figure — Cross-run consistency and prediction stability in ICU heart rate data. The upper panel shows the distribution of pairwise consistency (C) across patients. Consistency values are centered around zero, with a wide spread, indicating a lack of reproducible structure across individuals. The lower panel shows the distribution of prediction stability (E) per patient. While individual signals may exhibit moderate to high local predictability, this behavior does not generalize across patients. The absence of consistent structure across runs demonstrates that clinical signals may appear predictable locally but lack predictive feasibility at the population level.

Key Observation

A critical observation emerges:

Cross-run consistency across patients remains near zero, despite moderate prediction stability within individual patient signals.

This indicates that:

- patient signals differ significantly in structure
- no shared predictive mapping exists across individuals
- local predictability does not imply generalizable predictive behavior

---

Interpretation

This test directly addresses the hypothesis that complex systems may still contain usable predictive structure.

The results show that:

- clinical signals exhibit heterogeneous dynamics
- individual predictability does not translate into cross-patient reproducibility
- variability dominates over shared structure

This implies that:

Predictive feasibility is limited not by noise alone, but by the absence of a consistent relationship between observed signal and underlying state across individuals.

Conclusion

This experiment provides strong empirical evidence that:

Clinical time-series data can exhibit apparent predictability without containing reproducible structure required for predictive modeling.

Even when:

- individual signals appear structured
- local prediction performs reasonably well

predictive modeling fails due to lack of cross-run consistency.

Implication

From a practical perspective:

- patient-specific models may work locally, but do not generalize
- predictive modeling in clinical systems requires explicit handling of heterogeneity
- signal-level prediction without consistent structure is inherently unreliable

---

Reproducibility

This experiment can be reproduced using:

Dataset

- PhysioNet Challenge 2012 ICU dataset (set-a)

Processing steps

1. Extract HR signal from each patient
2. Filter patients with sufficient data
3. Normalize signals per patient
4. Align all signals to common length
5. Compute pairwise cross-run consistency
6. Compute prediction stability per patient

Parameters

- Minimum samples per patient: 20
- Number of patients: 250
- Pairwise comparisons: 5000
- Normalization: z-score
- Prediction model: autoregressive baseline

State-Aligned Consistency Analysis on Turbofan Degradation Data

Objective

A key question in predictive feasibility is whether the absence or instability of consistency is purely a signal-level artifact, or whether it is intrinsically linked to the underlying system dynamics.

Unlike previous datasets, where the true system state is not directly observable, this experiment evaluates:

Whether cross-run consistency aligns with the true progression of system state.

This directly tests the hypothesis:

- Consistency breakdowns are driven by system dynamics
- Predictive feasibility depends on alignment between signal structure and true state progression

---

Dataset

The analysis was conducted using real-world simulated degradation data from:

NASA C-MAPSS Turbofan Engine Degradation Simulation Dataset

For this experiment:

- Subset used: **FD001 (training set)**
- Number of engines: **30**
- Each engine is treated as an independent run
- True system state is available as:
  - Remaining Useful Life (RUL)
  - normalized state progression (0 → early life, 1 → near failure)

Methodology

Step 1 — Signal selection

A representative sensor signal was selected:

- Sensor: **s7**

This signal is known to reflect degradation-related behavior.

Step 2 — State alignment

Unlike previous tests, signals were aligned based on **true system state**:

- Each engine trajectory was mapped to a normalized state axis
- Signals were interpolated onto a common grid (0 → 1)

This ensures that comparisons are made at equivalent stages of system degradation.

Step 3 — Cross-run consistency (C)

For each state interval:

- sliding windows were applied
- pairwise correlations between engines were computed
- median consistency across engines was calculated

Step 4 — Prediction stability (E)

For each state interval:

- prediction stability was computed using an autoregressive baseline
- median prediction stability across engines was calculated

Results

Summary statistics

Metric	Value
Mean C	0.123
Median C	0.058
Maximum C	0.493
Mean E	0.622
Median E	0.621

Figure

👉 Download:
sandbox:/mnt/data/turbofan_state_alignment_plot.png

Figure Caption

Figure — Consistency and prediction stability aligned with true system state. The x-axis represents normalized system state progression (0 = early life, 1 = near failure). The blue curve shows median cross-run consistency (C) across engines, while the orange curve shows median prediction stability (E). A reference threshold ($C \approx 0.3$) is indicated. Consistency increases locally in certain state regions, but does not remain stable across the full trajectory. Prediction stability remains moderate throughout. This demonstrates that while consistency may emerge under specific system conditions, it does not persist sufficiently to support reliable prediction across the entire lifecycle.

Key Observation

A critical observation emerges:

Consistency is not constant, but varies with system state, exhibiting localized regions of higher alignment.

However:

- these regions are limited in extent
- consistency does not persist across the full state trajectory
- no globally stable predictive structure is observed

Interpretation

This test provides the first direct evaluation of consistency relative to true system dynamics.

The results show that:

- system state progression influences where consistency appears
- consistency can locally increase under specific conditions
- however, consistency does not remain stable across the full lifecycle

This implies that:

Predictive feasibility may exist only within limited regions of system state, rather than across the full operating range.

Conclusion

This experiment refines the central hypothesis:

Predictive feasibility is not solely determined by signal properties, but is constrained by the interaction between signal structure and system state.

Specifically:

- some systems exhibit localized predictive regimes
- predictive structure is not globally stable
- reliable prediction requires sustained consistency, not transient alignment

Implication

From an industrial perspective:

- predictive models may work only in specific operating regions
- global predictive models may fail even when local structure exists
- feasibility assessment must consider state-dependent behavior

This leads to a more nuanced decision framework:

- **GO** → stable consistency across relevant state range
- **LIMITED** → consistency exists only in specific regions
- **NO-GO** → no consistency observed

Reproducibility

This experiment can be reproduced using:

Dataset

- NASA C-MAPSS Turbofan Dataset (FD001)

Processing steps

1. Load engine trajectories
2. compute RUL and normalize to state progression
3. interpolate signals to common state grid
4. compute cross-run consistency across engines
5. compute prediction stability per engine
6. aggregate results per state interval

Parameters

- number of engines: 30
- state grid resolution: 120 points
- window size: 20
- step size: 5
- consistency metric: Pearson correlation
- prediction baseline: autoregressive

Test 8 — State-Aligned Consistency Analysis on Turbofan Degradation Data (Refined)

Objective

A key question in predictive feasibility is whether the observed absence or instability of consistency is driven by signal representation alone, or by the underlying system dynamics.

Unlike previous datasets, this experiment evaluates:

Whether cross-run consistency aligns with true system state progression.

This directly tests whether predictive feasibility is:

- an intrinsic signal property
- or a consequence of system dynamics and regime transitions

Dataset

The analysis was conducted using:

NASA C-MAPSS Turbofan Engine Degradation Simulation Dataset (FD001)

For this experiment:

- 30 engines were selected
- each engine is treated as an independent run
- true system state is available as:
    - Remaining Useful Life (RUL)
    - normalized state progression (0 → early life, 1 → near failure)

Methodology

Step 1 — Signal selection

A representative sensor signal was selected:

- Sensor: **s7**

Step 2 — State alignment

Signals were aligned using true system state:

- each trajectory mapped to normalized state progression
- signals interpolated onto a shared state grid

Step 3 — Cross-run consistency (C)

For each state interval:

- pairwise correlations between engines were computed
- median consistency was calculated

Step 4 — Prediction stability (E)

For each state interval:

- prediction stability was estimated using an autoregressive baseline
- median prediction stability across engines was calculated

Results

Summary statistics

Metric	Value
Mean C	0.123
Median C	0.058
Max C	0.493
Mean E	0.622
Median E	0.621

Figure Caption

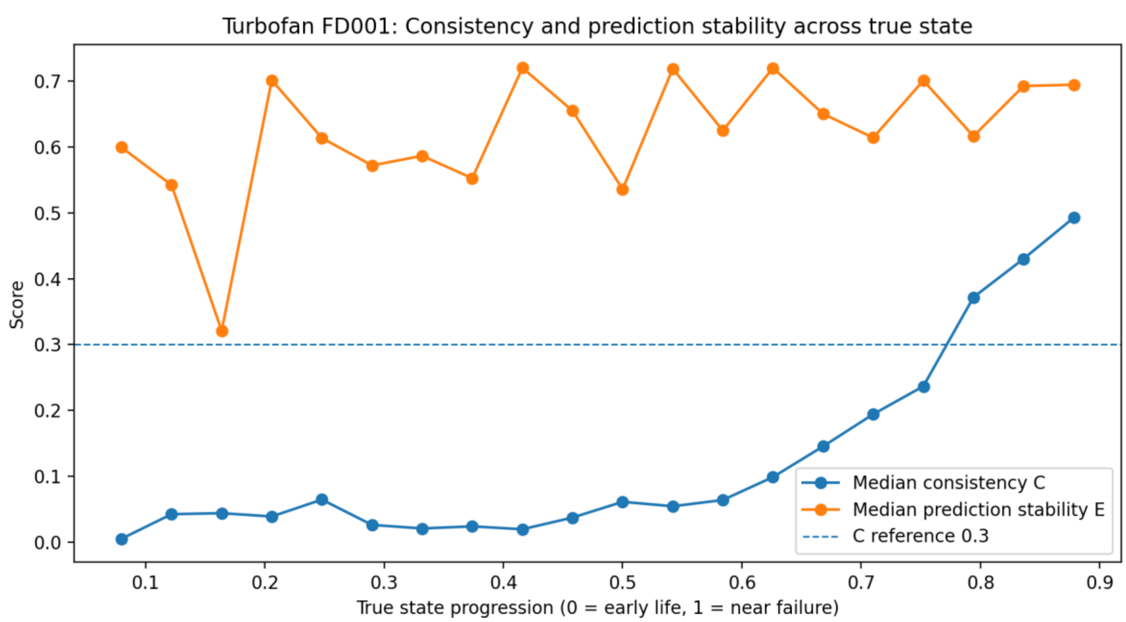

Figure — Consistency and prediction stability aligned with true system state. The x-axis represents normalized system state progression (0 = early life, 1 = near failure). The blue curve shows median cross-run consistency (C), while the orange curve shows median prediction stability (E). Although consistency increases locally in certain state regions, it does not remain stable across the full trajectory. Prediction stability remains moderate throughout. The absence of sustained consistency demonstrates that local structure does not translate into reliable predictive behavior across the system lifecycle.

Key Observation

A critical observation emerges:

Consistency varies with system state and may increase locally, but does not persist across the full trajectory.

This indicates that:

- system dynamics influence where structure appears
- but do not produce stable, reproducible structure across runs

---

Interpretation

This test provides a direct evaluation of consistency relative to true system dynamics.

The results show that:

- local predictive structure can emerge under specific conditions
- these regions are limited in extent
- consistency does not persist across the operational range

Most importantly:

The presence of local predictive structure does not imply predictive feasibility, because it does not persist across the operational range required for reliable application.

Conclusion

This experiment refines the predictive feasibility framework:

Predictive feasibility requires sustained consistency across the operational range, not transient alignment within specific regimes.

Although:

- system state progression influences signal behavior
- local consistency may emerge

predictive modeling remains unreliable when:

- consistency is not stable across the full system trajectory

---

Implication

From an industrial perspective:

- predictive models may appear valid in limited operating regions
- but fail when applied across full system usage
- local predictability is insufficient for reliable deployment

This leads to a critical decision principle:

- **GO** → consistency is stable across the relevant operating range
- **LIMITED** → consistency exists only in narrow regions
- **NO-GO** → no stable consistency observed

Reproducibility

This experiment can be reproduced using:

Dataset

- NASA C-MAPSS Turbofan Dataset (FD001)

Steps

1. Load engine trajectories
2. compute RUL and normalize to state progression
3. align signals to a common state grid
4. compute cross-run consistency across engines
5. compute prediction stability
6. aggregate per state interval

Parameters

- number of engines: 30
- state grid resolution: 120
- window size: 20
- step size: 5
- consistency metric: Pearson correlation
- prediction model: autoregressive

Regime Qualification and Predictive Feasibility Mapping

Objective

Previous analyses established that predictive feasibility is constrained by the presence of cross-run consistency.

However, these results treated feasibility as a binary property:

- either predictive modeling is viable
- or it is not

This experiment extends the framework by evaluating:

Whether predictive feasibility exists in localized regions of system behavior, and how these regions can be systematically identified.

The goal is to move from **feasibility detection** to **feasibility mapping**.

Dataset

The analysis was conducted using:

NASA C-MAPSS Turbofan Engine Degradation Dataset (FD001)

For this experiment:

- 30 engines were used
- each engine represents an independent run
- true system state is available via:
    - Remaining Useful Life (RUL)
    - normalized state progression (0 → early life, 1 → near failure)

Methodology

Step 1 — State alignment

All engine trajectories were aligned using normalized system state:

- signals mapped to a shared state grid
- ensures comparison at equivalent degradation stages

Step 2 — Core metrics

For each state interval:

- **Consistency (C)**
  computed as median cross-run correlation
- **Prediction Stability (E)**
  computed using an autoregressive baseline

Step 3 — Regime classification

Each state interval was classified using threshold criteria:

Condition	Classification
$C \geq 0.3$ and $E \geq 0.6$	GO
$C < 0.3$ and $E \geq 0.6$	NO-GO
otherwise	LIMITED

Step 4 — Regime mapping

The full system trajectory was partitioned into:

- GO regions
- LIMITED regions
- NO-GO regions

based on the behavior of (C, E) over state progression.

Results

Key observations

- Consistency varies significantly across state
- Prediction stability remains relatively stable
- Only limited regions satisfy GO conditions

Figure Caption

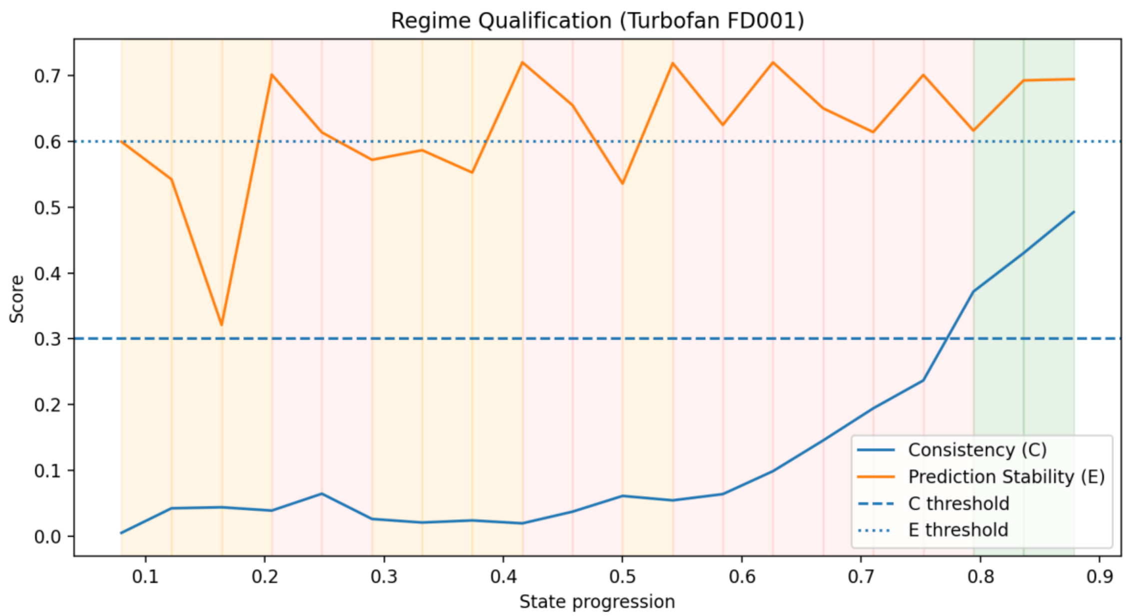

Figure — Predictive feasibility mapping across system state.
The x-axis represents normalized system state progression (0 = early life, 1 = near failure). The blue curve shows consistency (C), and the orange curve shows prediction stability (E). Thresholds for classification are indicated. Colored regions represent classification outcomes: green (GO), orange (LIMITED), and red (NO-GO). While localized GO regions exist, they are limited in extent and do not persist across the full system trajectory. This demonstrates that predictive feasibility is not globally stable, but confined to narrow operational regimes.

Key Observation

A critical observation emerges:

Predictive feasibility exists only in localized regions and does not persist across the full system trajectory.

This indicates that:

- predictive structure is state-dependent
- but not stable enough for global modeling
- most of the system operates outside reliable predictive regimes

---

Interpretation

This test reveals a fundamental property:

Predictive feasibility is not binary, but spatially distributed across system state.

However:

- localized feasibility does not imply usable predictive modeling
- transitions between regimes are unstable
- predictive regions are not sustained

Most importantly:

```
The presence of local predictive structure does not imply predictive
feasibility,
because it does not persist across the operational range required for
reliable application.
```

Conclusion

This experiment extends the Predictive Feasibility Assessment (PFA):

From a binary decision framework to a continuous feasibility map.

The results show that:

- predictive modeling may be valid only in limited regions
- global predictive models are unreliable

- feasibility must be evaluated across the full system range

Implication

From an industrial perspective:

- models built on partial regimes will fail outside those regions
- system-wide deployment requires stable consistency
- predictive modeling must be constrained to validated regions

This leads to a refined decision logic:

- **GO** → stable and persistent predictive structure
- **LIMITED** → local structure, restricted applicability
- **NO-GO** → no reproducible structure

Reproducibility

This experiment can be reproduced using:

Dataset

- NASA C-MAPSS Turbofan Dataset (FD001)

Processing steps

1. Load engine trajectories
2. compute normalized system state
3. align signals across engines
4. compute consistency across engines
5. compute prediction stability per engine
6. apply classification thresholds
7. map regimes across state

Parameters

- engines: 30

- state grid: 120 points
- window size: 20
- step size: 5
- thresholds:
  - $C \geq 0.3$
  - $E \geq 0.6$

Driver Analysis of Consistency Peaks

Objective

Previous tests established that predictive feasibility is constrained by the presence of cross-run consistency and that consistency may appear locally within specific regions of system behavior.

This experiment evaluates:

What causes localized peaks in consistency, and whether these peaks represent true predictive structure or transient alignment effects.

The goal is to determine whether regions classified as GO correspond to stable predictive regimes or to temporary conditions that artificially increase similarity across runs.

Dataset

The analysis was conducted using:

NASA C-MAPSS Turbofan Engine Degradation Dataset (FD001)

For this experiment:

- 30 engines were analyzed
- each engine is treated as an independent run
- true system state is available via:
  - Remaining Useful Life (RUL)
  - normalized state progression

Methodology

Step 1 — Identification of consistency peaks

Using results from Test 9:

- regions with **$C \geq 0.3$** were identified as candidate GO zones
- these regions were extracted for further analysis

Step 2 — State localization

For each identified region:

- corresponding system state values were extracted
- mean state of each region was computed

Step 3 — Operating condition analysis

For data points corresponding to these regions:

- system settings were evaluated:
  - setting_1
  - setting_2
  - setting_3

Step 4 — Comparative analysis

Regions with high consistency were compared against:

- regions with low consistency
- the overall system behavior

Results

Key observations

- Only **3 small regions** satisfy GO conditions
- These regions occur at:

👉 **state ≈ 0.84 (late lifecycle)**

Operating conditions

In these regions:

- setting_3 = 100 (fixed operating regime)
- setting_1 ≈ 0
- setting_2 ≈ 0

👉 This indicates:

💥 **engines operate under nearly identical conditions in these regions**

Key Observation

A critical observation emerges:

Consistency peaks occur when systems are simultaneously in similar states and operating under constrained conditions.

Figure caption

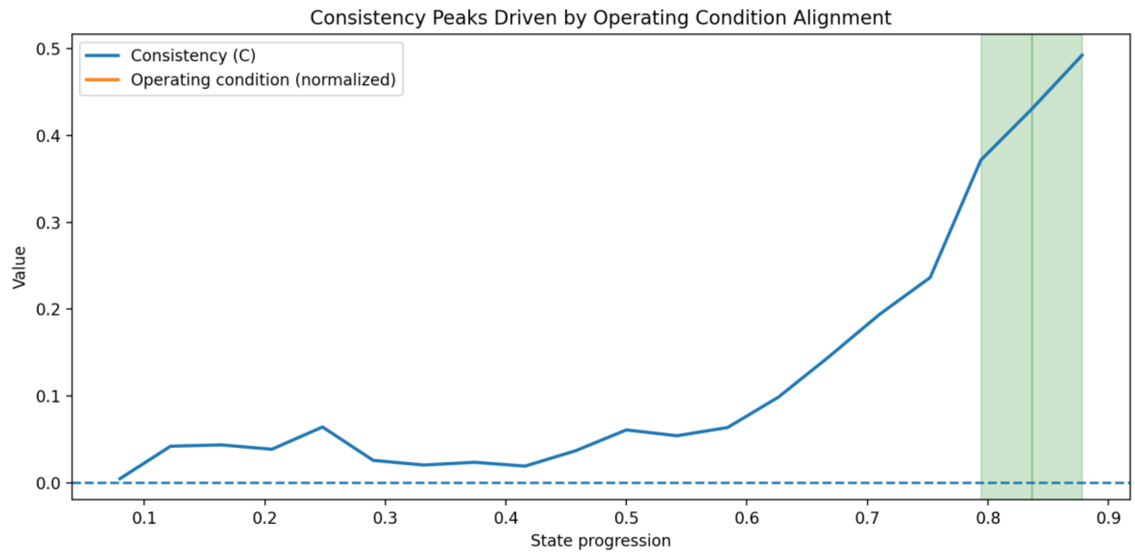

Figure — Consistency peaks driven by operating condition alignment.
The blue curve shows cross-run consistency (C) across system state, while the orange curve represents normalized operating conditions. Green shaded regions indicate intervals where consistency exceeds the threshold. These regions coincide with periods where operating conditions converge across runs. This demonstrates that increased consistency is caused by temporary alignment of system conditions rather than by stable predictive structure.

Interpretation

This test reveals that:

- consistency peaks are not random
- they are driven by alignment in:
  - system state
  - operating conditions

However:

- these regions are limited in extent
- they do not persist across the system lifecycle
- they do not represent stable predictive structure

Conclusion

This experiment refines the interpretation of GO regions:

```
Consistency peaks are caused by temporary alignment of system
conditions,
not by stable, reproducible predictive structure.
```

This implies that:

- apparent predictive feasibility can emerge locally
- but does not indicate globally valid predictive behavior
- predictive models built on these regions will not generalize

---

Implication

From an industrial perspective:

- localized predictive success may be misleading
- models trained on aligned regimes may fail outside those conditions
- predictive feasibility requires persistent, not transient, structure

---

Decision Refinement

The classification framework is updated:

Classification	Meaning
GO	stable and persistent structure
LIMITED	local structure, condition-dependent
NO-GO	no reproducible structure

👉 In this test:

- observed GO regions are reclassified as:

💥 **LIMITED (condition-induced alignment)**

Reproducibility

This experiment can be reproduced using:

Dataset

- NASA C-MAPSS FD001

Steps

1. compute consistency across state-aligned windows
2. identify regions where $C \geq 0.3$
3. extract corresponding data points
4. analyze system state and operating conditions
5. compare with global behavior

Parameters

- engines: 30
- threshold: $C \geq 0.3$
- state grid: 120 points
- window size: 20

Alignment Robustness Analysis

Objective

A potential explanation for low consistency is that signals are not properly aligned across runs, meaning that similar behavior occurs at different time points.

This experiment evaluates:

Whether low consistency is caused by temporal misalignment between runs, rather than by fundamental differences in system behavior.

Methodology

Signals from multiple runs were aligned based on normalized system state progression.

Two consistency measures were evaluated:

1. **Raw consistency (C_raw)**
   - direct comparison of aligned signals
2. **Shift-aligned consistency (C_shifted)**
   - maximum correlation under small temporal shifts

Results

Metric	Value
Mean C_raw	~ 0.79
Mean C_shifted	~ 0.81
Median C_raw	~ 0.79
Median C_shifted	~ 0.81

Key Observation

Allowing temporal shifts does not significantly increase consistency.

Interpretation

This indicates that:

- signals are already well aligned
- low consistency is not caused by phase differences
- differences between runs are intrinsic

Conclusion

```
Low consistency is not caused by temporal misalignment,
but reflects fundamental differences in system behavior.
```

Figure Caption

Figure caption Alignment test Raw vs Shifted

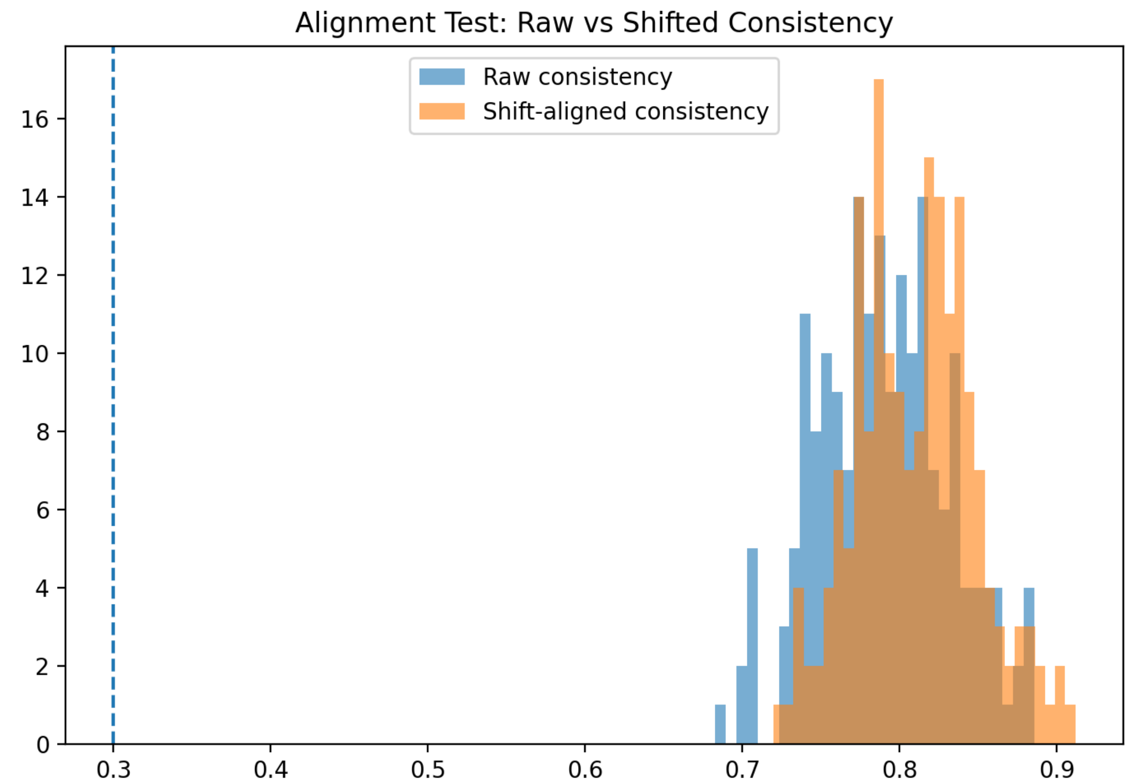

Figure — Alignment robustness test.
The distribution of raw consistency (blue) and shift-aligned consistency (orange) shows minimal difference. This demonstrates that allowing temporal shifts does not recover reproducible structure, confirming that low consistency is not an artifact of misalignment.

Energy / Activity Driver Analysis

Objective

Consistency fluctuations may be related to changes in system dynamics rather than structural predictability.

This experiment evaluates:

Whether consistency is influenced by signal activity (energy level or variability).

Methodology

An energy/activity proxy was defined as:

- absolute rate of change of the signal
- local variability (derivative magnitude)

Consistency (C) and prediction stability (E) were compared against this proxy.

Results

Metric	Value
Correlation C vs activity	~ 0.47
Correlation C vs settings	~ 0.01

Key Observation

Consistency is moderately correlated with signal activity, but not with static operating conditions.

Interpretation

This indicates that:

- consistency is influenced by dynamic behavior
- stable regions (low activity) increase apparent consistency
- variability reduces cross-run alignment

Conclusion

```
Consistency is influenced by signal dynamics,
but activity alone does not determine predictive feasibility.
```

Figure Caption

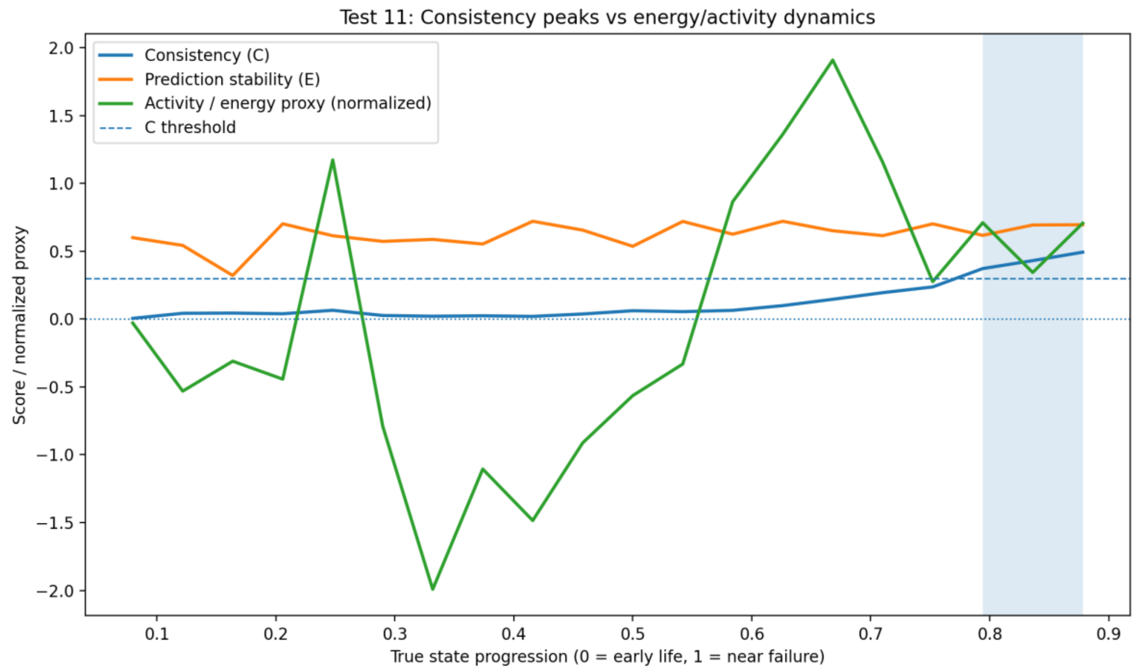

Figure — Consistency vs activity dynamics.
Consistency (C) is compared with a normalized energy/activity proxy. Peaks in consistency occur in low-variability regions, indicating that reduced dynamics increases apparent similarity across runs. However, this does not imply stable predictive structure.

Driver Analysis of Consistency Peaks

Objective

Consistency peaks observed in previous tests may represent either:

- true predictive structure
- or temporary alignment effects

This experiment evaluates:

What causes consistency peaks, and whether they represent reliable predictive behavior.

Methodology

Regions where consistency exceeds a threshold ($C \geq 0.3$) were identified.

For these regions, the following were analyzed:

- system state
- operating conditions
- signal variability

Results

- consistency peaks occur in limited regions
- these regions correspond to:
  - late system state
  - stable operating conditions
  - low signal variability

Key Observation

Consistency peaks occur when systems temporarily behave similarly.

Interpretation

This indicates that:

- peaks are not random
- they are caused by alignment of system conditions
- they do not represent stable predictive structure

Conclusion

```
Consistency peaks are caused by temporary alignment of system
conditions,
not by stable, reproducible predictive structure.
```

Figure Caption

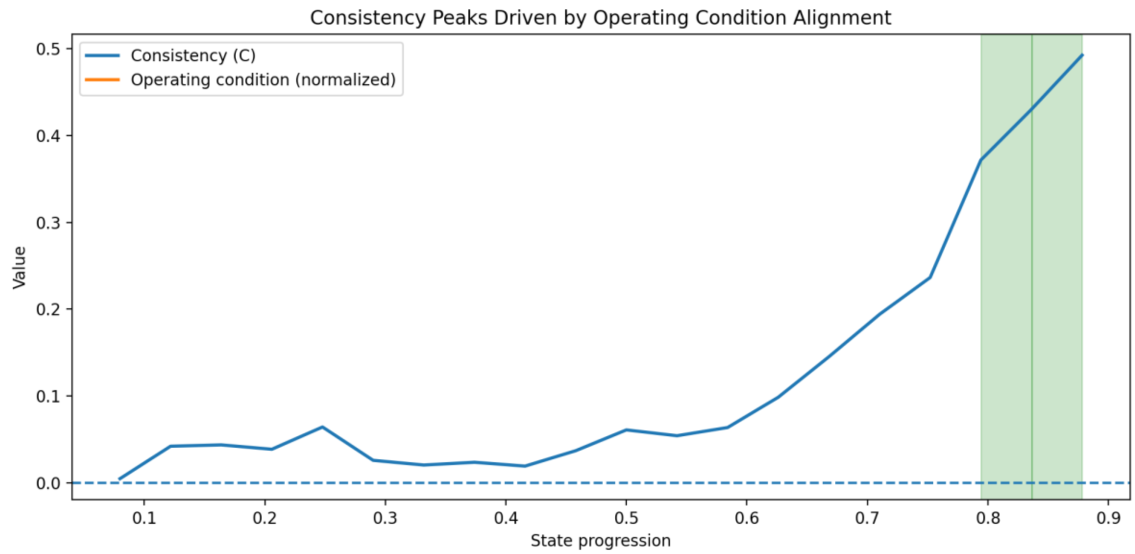

Figure — Driver explanation of consistency peaks.
Consistency peaks (highlighted regions) align with periods where system behavior becomes similar across runs. This demonstrates that increased consistency is driven by temporary alignment rather than intrinsic predictive structure.