Chapter Outline

Fault Diagnosis and Health Prognostics

Chapter 7 Health Management and Safety Assessment

7.1 Health Management Fundamentals

7.2 Unified Fault Modeling

7.3 Safety Reinforcement and Fault Data

7.4 Fault Hazards and Safety Assessment

7.5 Diagnosis and Prognostics

7.6 Safety Decisions and Improvement

7.7 Fault Injection Interfaces

7.8 Experiments and Validation

7.9 Exercises and Classroom Discussion

4 Units in This Section

7.5.1 Residuals and Baselines

Textbook: Secs. 7.1.2.1 and 7.1.5.4

7.5.2 Diagnostic Information Fusion

Textbook: Sec. 7.1.5.4

7.5.3 Evaluation across Operating Conditions

Textbook: Sec. 7.1.5.4

7.5.4 Health Prognostics

Textbook: Sec. 7.1.2.1 (battery degradation)

7.5.1 Residuals and Baselines: Health-Management Judgments

Detection, Diagnosis, and Prognostics Answer Different Questions

Prognostics and health management (PHM): sense current state, predict degradation, and guide mission and maintenance decisions.

Judgment	Inputs and evidence	Typical output
Anomaly detection	Compare observations with normal ranges or model predictions	Is the current window abnormal?
Fault diagnosis	Multichannel features, residuals, and known fault patterns	Component, fault class, and class scores
Health prognostics	Health-indicator history, degradation laws, and future use conditions	Future health trends and remaining useful life (RUL)

Health varies continuously; sufficient capability depends on the mission.

7.5.1 Residuals and Baselines: Operating Conditions and Anomalies

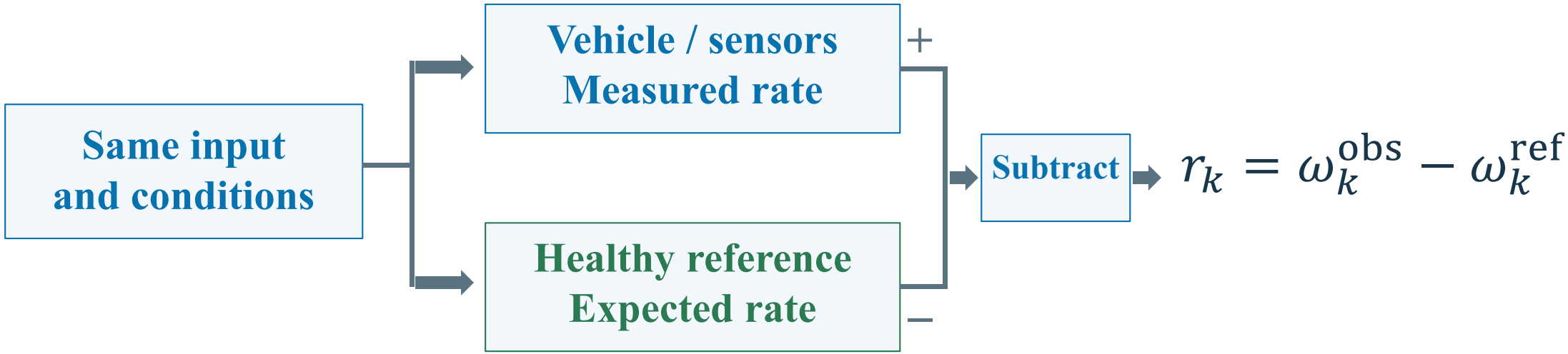
Normal Maneuvers Can Cause Large Sensor Fluctuations

Situation	Why observations change	Useful comparison
Acceleration / turns	Control inputs change; a healthy vehicle changes its response	Healthy-model response to the same commands
Sensor anomaly	A channel bias or noise change may break cross-channel consistency	Redundant observations and cross-channel relationships
Reduced motor effectiveness	Actuation falls; control output may rise to compensate	Actuator data, motion, and healthy-reference residuals

A condition-matched healthy reference makes anomaly evidence comparable.

7.5.1 Residuals and Baselines: Residual Sign and Units

Compare Measurements with a Matched Healthy Response



ω_k^{obs} is measured; ω_k^{ref} is the healthy reference. Both are body-frame roll rates here.

Align time, frames, and units. Residuals also contain model bias, noise, and timing error.

$$r_k = 0.70 - 0.50 = 0.20 \text{ rad/s}$$

7.5.1 Residuals and Baselines: Normal Residual Statistics

Use Normal Data to Calibrate Bias and Fluctuation Scale

Normal-training baseline: mean 0.02 rad/s; standard deviation 0.05 rad/s (teaching example).

$$z_k = \frac{r_k - \mu_0}{\sigma_0}$$

μ_0 is the normal mean; σ_0 is the normal fluctuation scale; z_k is dimensionless.
New samples use these training statistics; test data do not fit normalization parameters.

$$z_k = \frac{0.20 - 0.02}{0.05} = 3.6$$

Given $|z_k| > 3$: an abnormal sample; fault class needs further evidence.
Select thresholds on independent validation data: false alarms, misses, and alarm timing.

7.5.1 Residuals and Baselines: Updating a Persistence Counter

Persistence Checks Reject Transient Alarms

Teaching: two consecutive $|z_k| > 3$ samples confirm; interval 0.2 s; fault onset 0.75 s.

One spike is not enough; two consecutive crossings trigger an alarm.

t = 1.20 s

Same input, same time

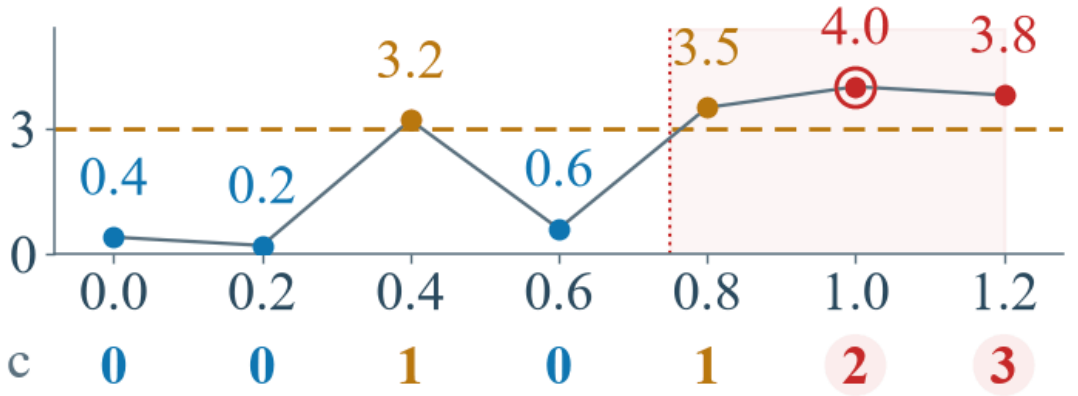
Healthy reference 0.500

Observed reading 0.710

r = 0.710 − 0.500
= 0.210 rad/s

t_k = 1.2 s

Standardized residual |z|



Run count c

3

$|z| > 3$

Two in a row

ALARM

Delay 0.25 s

Fault 0.75 s → Alarm 1.00 s; delay 0.25 s.

$$\Delta t_{\text{det}} = 1.00 - 0.75 = 0.25 \text{ s}$$

7.5.1 Residuals and Baselines: Enhancing the Reference Model

Correct Model Error while Retaining Fault Deviations

Pair real / virtual outputs
Compute normal error

→

Learn error behavior
Reconstruct the error

→

Model output + error
Corrected reference

$$y_{\text{ref}} = y_{\text{phys}} + \hat{e}_{\text{normal}}$$

y_{phys} is the physics-model output; $\hat{e}_{\text{normal}}$ reconstructs error under normal conditions. All share the reference-output units. This corresponds to $X_{\text{comp}} = X_{\text{virt}} + \Delta$ on the enhanced digital twin case page in Section 7.2.

Accept the normal reference; retain suspicious deviations in operation, then confirm faults. Learning residuals during unknown faults can absorb fault deviations as model error.

Adapted from enhanced digital twin Fig. 1(b); source: Section 7.2, enhanced digital twin case.

Chapter Outline

Fault Diagnosis and Health Prognostics

Chapter 7 Health Management and Safety Assessment

7.1 Health Management Fundamentals

7.2 Unified Fault Modeling

7.3 Safety Reinforcement and Fault Data

7.4 Fault Hazards and Safety Assessment

7.5 Diagnosis and Prognostics

7.6 Safety Decisions and Improvement

7.7 Fault Injection Interfaces

7.8 Experiments and Validation

7.9 Exercises and Classroom Discussion

4 Units in This Section

7.5.1 Residuals and Baselines

Textbook: Secs. 7.1.2.1 and 7.1.5.4

7.5.2 Diagnostic Information Fusion

Textbook: Sec. 7.1.5.4

7.5.3 Evaluation across Operating Conditions

Textbook: Sec. 7.1.5.4

7.5.4 Health Prognostics

Textbook: Sec. 7.1.2.1 (battery degradation)

7.5.2 Diagnostic Information Fusion: Method Families

Three Method Families Use Different Diagnostic Evidence

Method	Source of diagnostic evidence	Main dependencies and tradeoffs
Model-driven	Physics models, observers, analytical redundancy: expected versus measured response	Clear interpretation; model bias enters residuals
Data-driven	Principal component analysis, support vector machines, and neural nets learn normal/fault patterns	Needs class and condition coverage; unknown faults remain difficult
Fusion methods	Residual learning, Bayesian fusion, model-error correction, multiple observations	Check consistency of models, data, and interfaces

Diagnostic scope depends jointly on model validity and data coverage.

7.5.2 Diagnostic Information Fusion: Models and Data

Training Has Labels; Application Has Observations

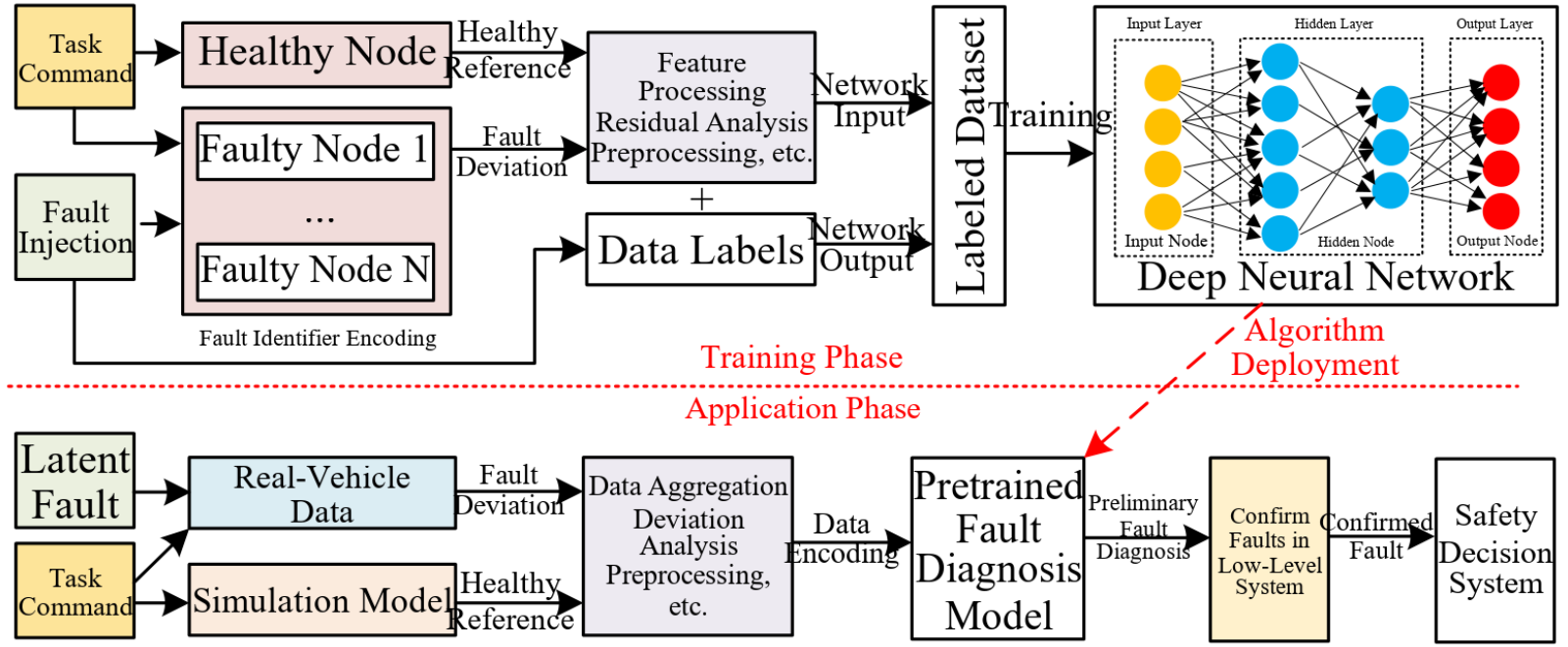


Fig. 7.10. Fault diagnosis training using models and data

Learn inputs → labels. In operation, use a healthy reference, then confirm faults at low level.

7.5.2 Diagnostic Information Fusion: Training Samples

Residuals Are Network Inputs; Injection Information Provides Labels

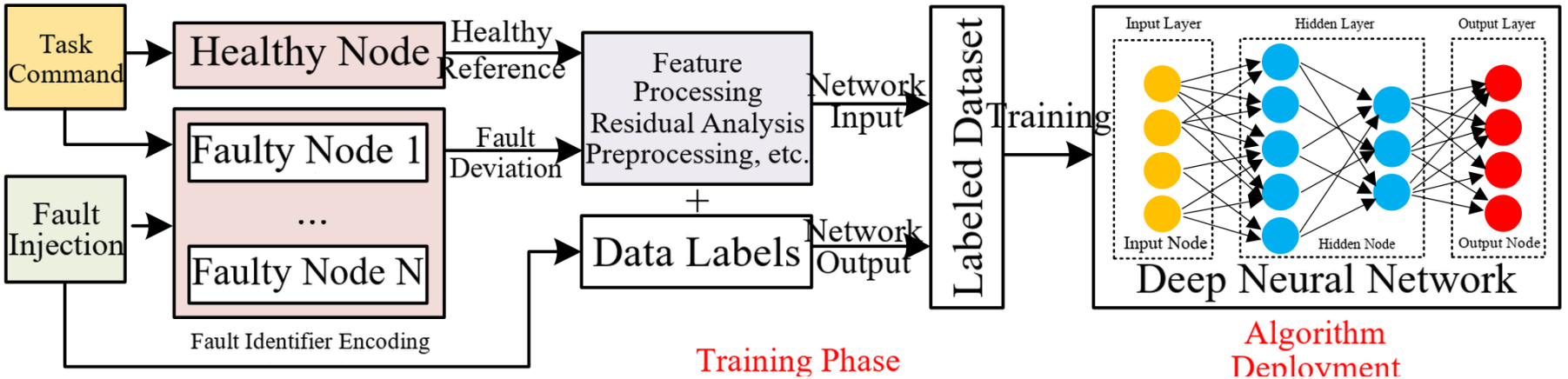


Fig. 7.10, training detail: separate feature and label paths

Nodes: software-in-the-loop (SITL; SIL in the textbook) or hardware-in-the-loop (HITL; HIL in the textbook), or real vehicles (Unit 7.3.1).

Healthy and faulty nodes run the same task in parallel; fault ID controls injection.

Features: aligned, normalized residual windows; injection class/time supply natural labels.

7.5.2 Diagnostic Information Fusion: Runtime Data and Outputs

Class Scores Still Need Fault Confirmation in Operation

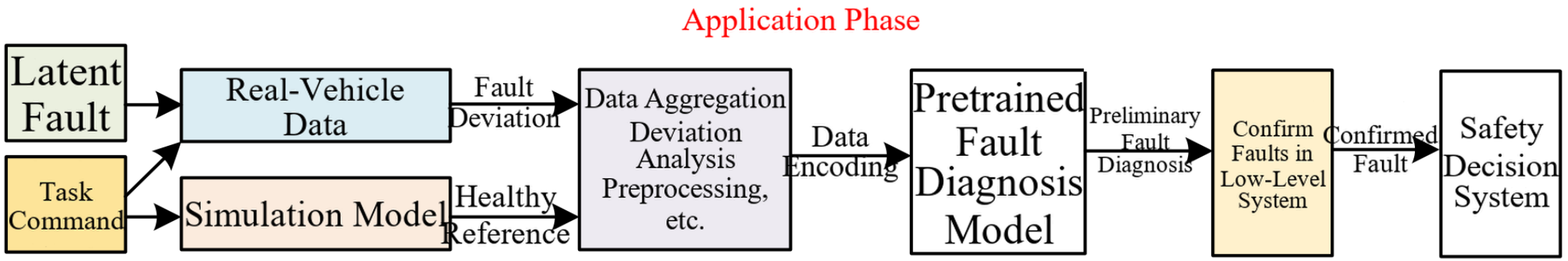


Fig. 7.10, application detail: current features enter the trained model

Interface: keep training channel order, signs, units, windows, and normalization statistics.

Class scores express relative preference; fault probabilities need independent calibration.

Before safety decisions, confirm faults using persistent, redundant, and task-state evidence.

7.5.2 Diagnostic Information Fusion: Paper Source and Symbols

Check the Paper and Entry Point Before Reading Its Symbols

Advanced graph-network pages end at class scores; main sequence → Unit 7.5.3.

Paper: Tu J, Nian X, Dai X. Digital Twin-Guided Spatiotemporal Graph Representation Learning for Reliable UAV Fault Diagnosis under Complex Flight Conditions. IEEE Internet of Things Journal, 2026.

DOI: https://doi.org/10.1109/JIOT.2026.3709310

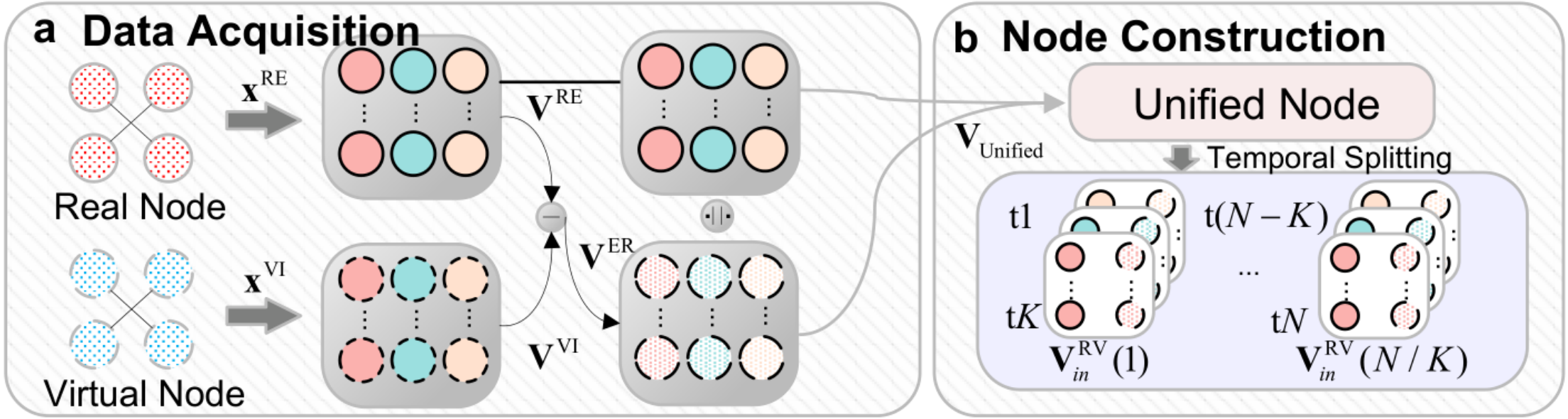
Folder: [installation directory]/RflySimAPIs/PaperRes/6.IoT-DT-Diagnosis/paper/

Real Node: unmanned aerial vehicle (UAV)
Virtual Node: healthy digital twin
V^{RE}: measured nodes
V^{VI}: virtual twin outputs

$\ominus$: subtract to form residual nodes V^{ER}
V_{Unified}: measured and residual nodes
Temporal Splitting: windows of K time steps
Node: feature unit, not a simulation instance

7.5.2 Diagnostic Information Fusion: Constructing Paper Nodes

Measured and Residual Nodes Retain Different Information

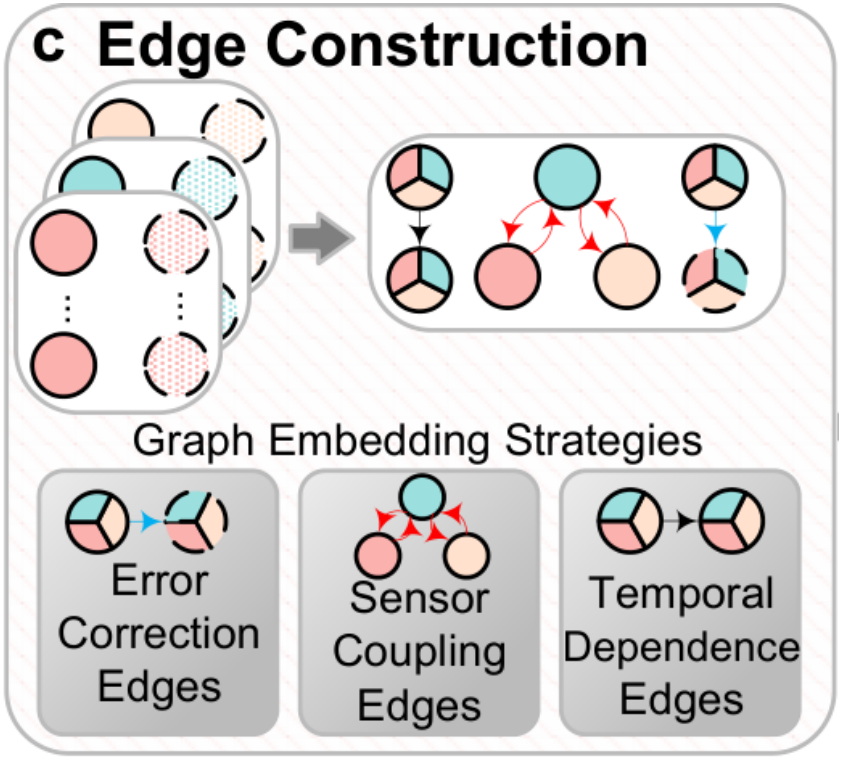


Same paper, Fig. 3(a,b): node construction

Measured nodes retain current responses; residual nodes show deviation from the healthy reference.

7.5.2 Diagnostic Information Fusion: Physical Relations and Graph Edges

Three Edge Types Carry Different Anomaly Evidence



Same paper, Fig. 3(c): relation edges

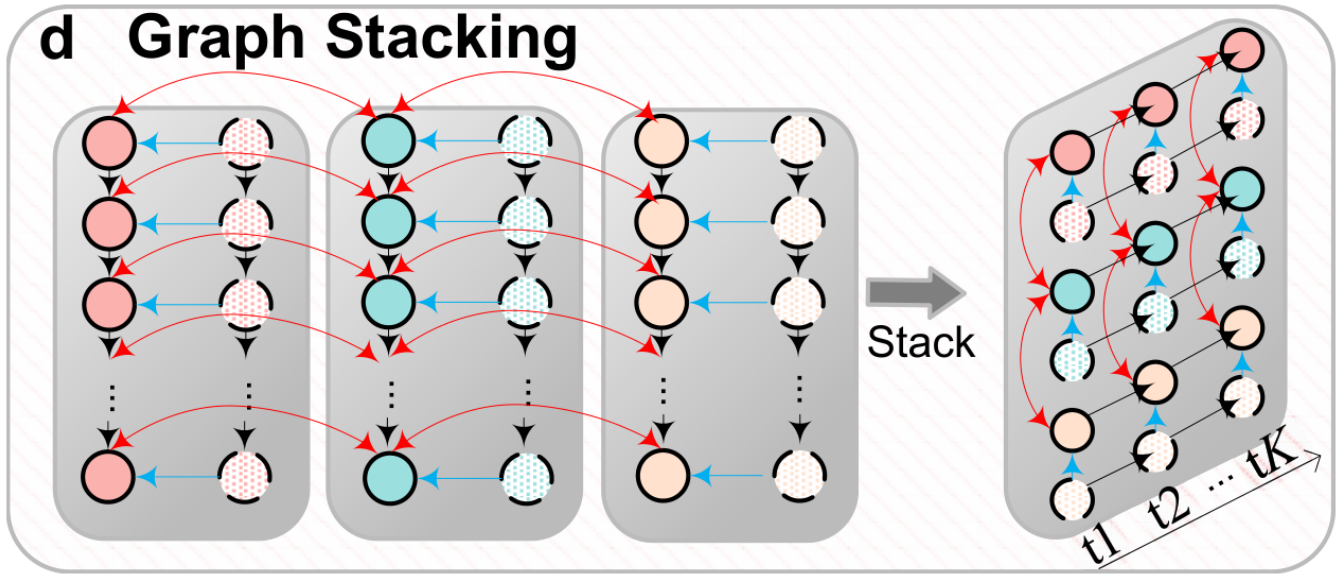
Error-correction edges connect measurements to corresponding residuals, providing healthy-reference consistency evidence.

Sensor-coupling edges connect channels at the same time to check consistency under shared motion.

Temporal-dependence edges connect like nodes over time, retaining continuous evidence of fault development.

7.5.2 Diagnostic Information Fusion: From Relation Graphs to Class Scores

Graph Sequences Retain Changes within the Observation Window



Same paper, Fig. 3(d): temporal graph sequence

Twin-Consistent Spatiotemporal Graph Diagnosis Network (TC-SGDN)

Attention: weight related neighbors.

Gated recurrent units: store and update sequence state.

Online windows end now; output current class scores.

Observations + residuals → **Graph + time sequence** → **Known-class scores**

Chapter Outline

Fault Diagnosis and Health Prognostics

Chapter 7 Health Management and Safety Assessment

7.1 Health Management Fundamentals

7.2 Unified Fault Modeling

7.3 Safety Reinforcement and Fault Data

7.4 Fault Hazards and Safety Assessment

7.5 Diagnosis and Prognostics

7.6 Safety Decisions and Improvement

7.7 Fault Injection Interfaces

7.8 Experiments and Validation

7.9 Exercises and Classroom Discussion

4 Units in This Section

7.5.1 Residuals and Baselines

Textbook: Secs. 7.1.2.1 and 7.1.5.4

7.5.2 Diagnostic Information Fusion

Textbook: Sec. 7.1.5.4

7.5.3 Evaluation across Operating Conditions

Textbook: Sec. 7.1.5.4

7.5.4 Health Prognostics

Textbook: Sec. 7.1.2.1 (battery degradation)

7.5.3 Evaluation across Operating Conditions: Confusion Matrix

Overall Accuracy Can Hide Missed Minority-Class Faults

Example: 200 windows (180 normal, 20 faulty). Fault is the positive class; threshold = 0.5.

Actual state	Predicted normal	Predicted fault
Normal: 180	172: true negatives (TN)	8: false positives (FP) False alarms
Faulty: 20	5: false negatives (FN) Misses	15: true positives (TP) Detected faults

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{187}{200} = 93.5\%$$

7.5.3 Evaluation across Operating Conditions: Diagnostic Metrics

Measure Misses among Faults and False Alarms among Normal Samples

Same teaching matrix: TP = 15, FN = 5, FP = 8, TN = 172.

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{15}{20} = 75\% \qquad \text{Miss} = \frac{FN}{TP + FN} = \frac{5}{20} = 25\%$$

$$\text{FPR} = \frac{FP}{FP + TN} = \frac{8}{180} \approx 4.4\% \qquad \text{Precision} = \frac{TP}{TP + FP} \approx 65.2\%$$

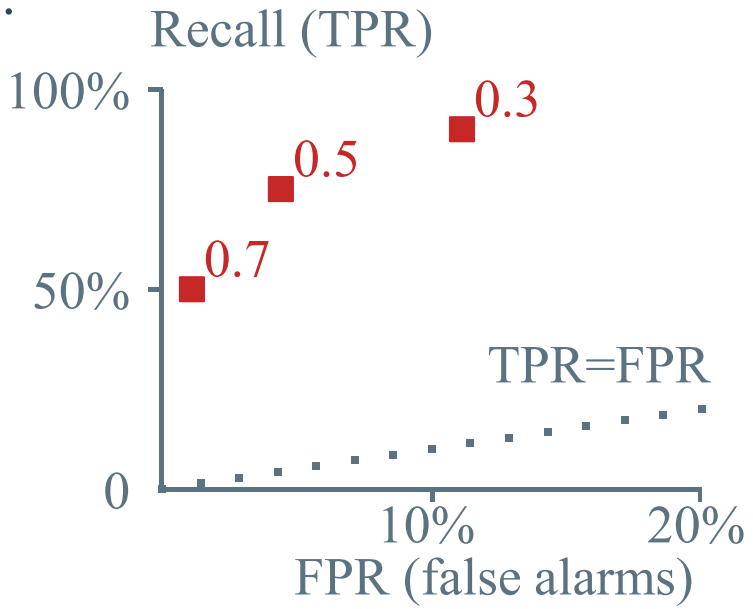
Recall: detections / faults. Miss rate: misses / faults.
False-positive rate (FPR): false alarms / normal samples. Precision: true faults / alarms.
Specify units: windows, complete flights, or alarms per hour.

7.5.3 Evaluation across Operating Conditions: Score Thresholds and ROC

The Threshold Trades False Alarms Against Missed Faults

Receiver operating characteristic (ROC): false-positive rate (FPR) on the horizontal axis. True-positive rate (TPR), or fault recall, is on the vertical axis.

Threshold	TP / FN	FP / normal	TPR; FPR
0.3	18 / 2	20 / 180	90%; 11.1%
0.5	15 / 5	8 / 180	75%; 4.4%
0.7	10 / 10	2 / 180	50%; 1.1%



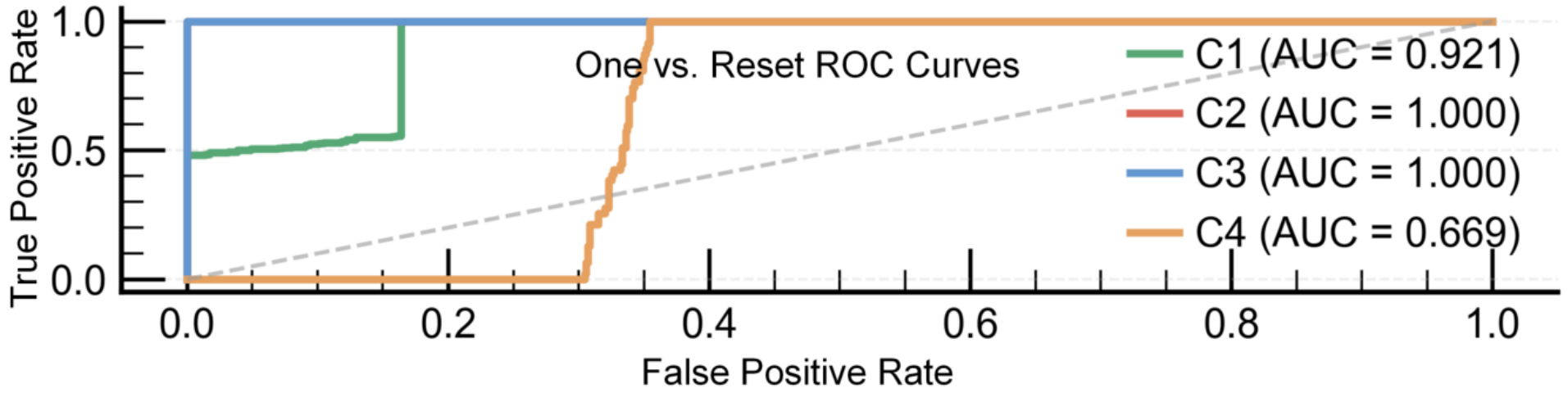
Teaching ROC operating points

Same 200 teaching scores: each threshold gives one ROC operating point.

7.5.3 Evaluation across Operating Conditions: Class ROC and AUC

Area under a Curve Measures Discrimination, Rather Than Accuracy

Area under the curve (AUC): one-vs-rest discrimination across thresholds.
C1 normal; C2 accelerometer noise; C3 single-motor signal jumps; C4 motor 3 effectiveness ×0.85.

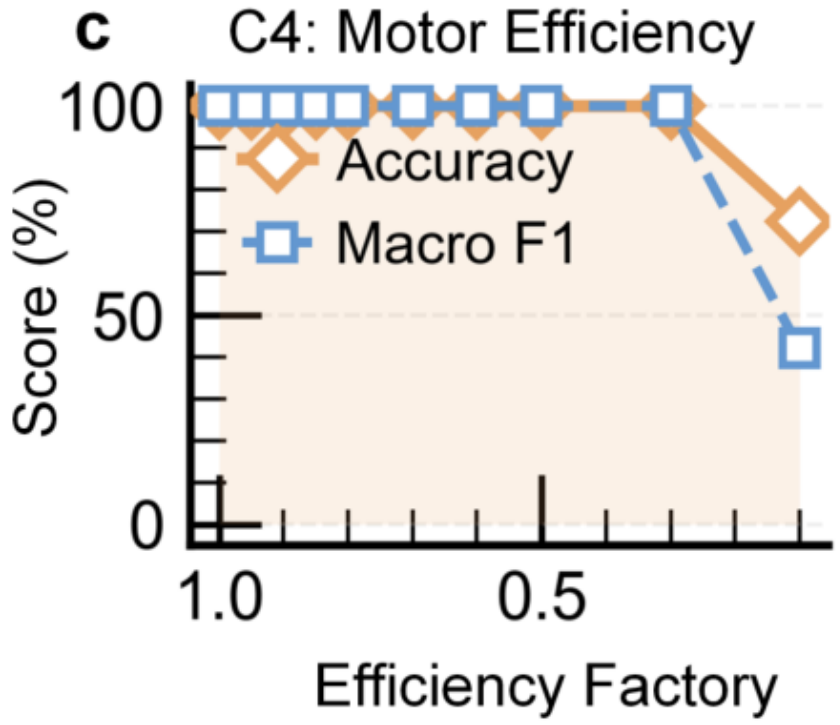


Same paper, Fig. 11(d): class ROC curves on the benchmark test set

Upper left is better: C2/C3 follow the boundary; C4 detects almost none at FPR < 0.3.

7.5.3 Evaluation across Operating Conditions: Fault-Intensity Scan

Accuracy and Macro F1 Drop at the Strongest Degradation Level



Same paper, Fig. 11(c): fault-intensity scan

Scale motor 3 effectiveness η on normal base samples.
This scan differs from the class-ROC benchmark (η = 0.85, AUC = 0.669).

Orange: accuracy. Blue: macro F1,
the equally weighted mean of class F1 scores.

η decreases to the right, so degradation grows.
Scores fall only at the strongest scaling end.

η scales simulated motor output.
Severity needs regression or a health model.

7.5.3 Evaluation across Operating Conditions: Conditions and Leakage

Adjacent Windows Overlap by 95% in a Paper-Code Entry Point

Group flights by condition → **Window within each group** → **Fit training scales** → **Independent test**

Window i: samples 1–80 → training

Window i+1: samples 5–84 → test

76 shared
Leakage

$$\text{overlap} = \frac{80 - 4}{80} = 95\%$$

Splitting rules: see the sliding-window leakage and train/validation/test pages in Unit 7.3.5.

Notebook setting: 80 samples, stride 4; adjacent windows share 76 samples.

Group whole flights by payload and maneuver; report each test condition separately.

Check splitting before windowing; consult the paper for its other validation splits.

7.5.3 Evaluation across Operating Conditions: Windows, Pipelines, and Alarms

Measure Detection Delay from When the Fault Takes Effect

Time / s	Known event	Alarm state
10.0	Fault takes effect	Timing origin
10.4	First threshold exceedance	Await persistent confirmation
10.6	Second consecutive exceedance	Alarm confirmed

$$\Delta t_{\mathrm{det}} = 10.6 - 10.0 = 0.6\ \mathrm{s}$$

Table VII: forward median (P50) 10.13 ms; acquisition-to-output pipeline 43–54 ms.
Inertial measurement unit (IMU): 50 Hz, 80 samples, 480 graph nodes; ground RTX3060.
Timing scopes differ: alarms await persistent evidence, timed from actual fault effect.

Chapter Outline

Fault Diagnosis and Health Prognostics

Chapter 7 Health Management and Safety Assessment

7.1 Health Management Fundamentals

7.2 Unified Fault Modeling

7.3 Safety Reinforcement and Fault Data

7.4 Fault Hazards and Safety Assessment

7.5 Diagnosis and Prognostics

7.6 Safety Decisions and Improvement

7.7 Fault Injection Interfaces

7.8 Experiments and Validation

7.9 Exercises and Classroom Discussion

4 Units in This Section

7.5.1 Residuals and Baselines

Textbook: Secs. 7.1.2.1 and 7.1.5.4

7.5.2 Diagnostic Information Fusion

Textbook: Sec. 7.1.5.4

7.5.3 Evaluation across Operating Conditions

Textbook: Sec. 7.1.5.4

7.5.4 Health Prognostics

Textbook: Sec. 7.1.2.1 (battery degradation)

7.5.4 Health Prognostics: Battery State and Lifetime

SOC and SOH Describe Now; RUL Describes the Future

State of charge (SOC)	Capacity state of health (SOH)	Remaining useful life (RUL)
Remaining charge / current usable full charge	Current full-charge capacity / rated capacity	Cycles or time remaining to a specified threshold
Conceptual example	Conceptual example	Conceptual example
$1.6/1.6 = 100\%$	$1.6/2.0 = 80\%$	Also needs a degradation law and future use conditions

1.6 Ah is a teaching assumption. SOC/SOH describe now; RUL describes the future.

7.5.4 Health Prognostics: Battery State and Lifetime

Why Does a Fully Charged Aged Battery Empty Sooner?

Both at SOC 100%, but full-charge capacities differ

New cell: rated 2 Ah

Full-charge capacity 2.00 Ah



Full

Remaining charge 2.00 Ah

Capacity-based SOH 100%

Aged cell: rated 2 Ah

Full-charge capacity 1.60 Ah



Full

Remaining charge 1.60 Ah

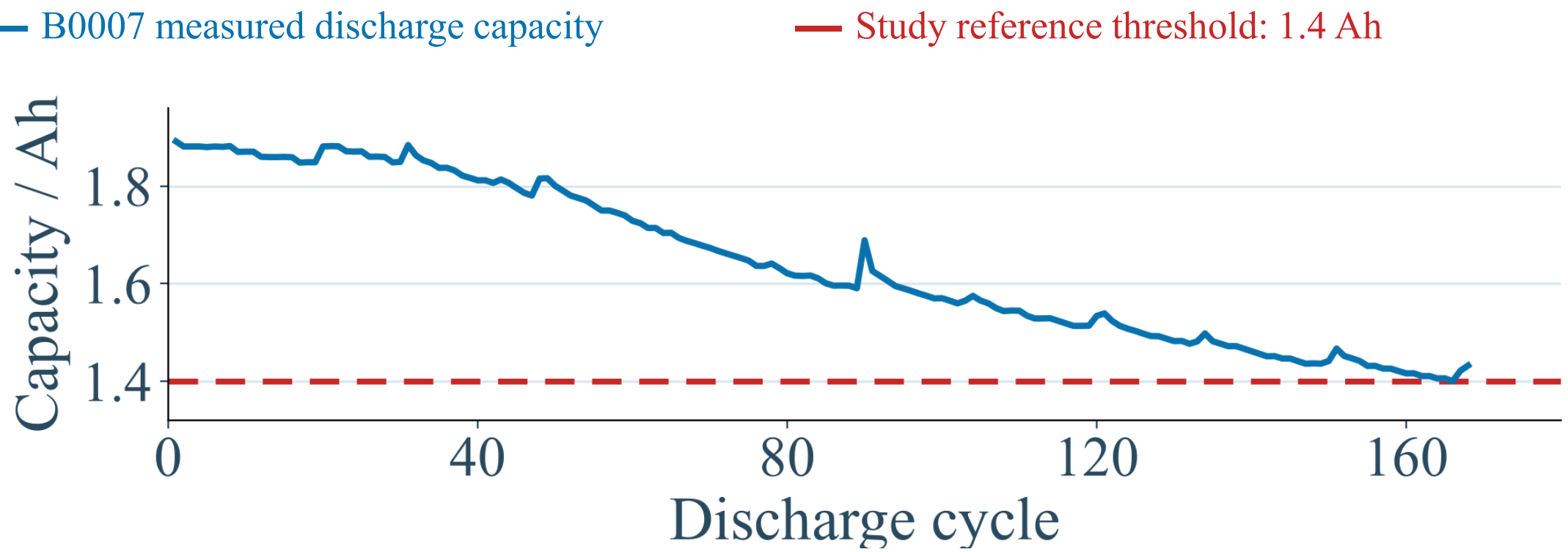
Capacity-based SOH 80%

Full SOC does not restore health; RUL needs degradation history

Capacity at the same current: charging restores SOC, but does not restore capacity already lost.
Capacity is the SOH indicator here; RUL also needs degradation history and future use conditions.

7.5.4 Health Prognostics: NASA Cell Records

Measured Capacity Degrades Overall but Can Rise Briefly



NASA Prognostics Center of Excellence (PCoE), item 5: B0007 discharge capacity; ambient 24 °C

7.5.4 Health Prognostics: One-Step Trend Extrapolation

Predict Threshold Crossing from History through Cycle 100

Predict at cycle 100; fit cycles 71−100. Units: Ah for capacity; cycle index is dimensionless.

Use only cycles 71–100 → **Least-squares line fit** → **Extrapolate to 1.4 Ah**

End of life (EOL): first cycle at capacity ≤1.4 Ah. Assume recent slope and use conditions persist.

$$\hat{Q}(n) = 1.8500 - 0.002731n$$

$$\hat{n}_{\mathrm{EOL}} = \frac{1.8500 - 1.4}{0.002731} \approx 164.8$$

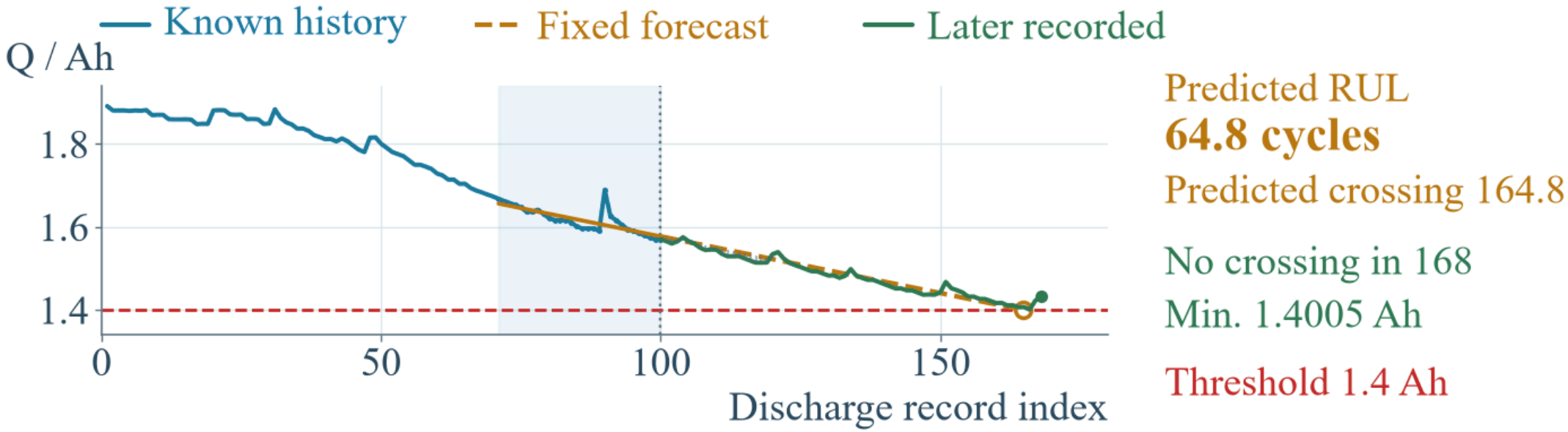
$$\widehat{\mathrm{RUL}} = \hat{n}_{\mathrm{EOL}} - 100 \approx 64.8 \text{ cycles}$$

7.5.4 Health Prognostics: Separating History from the Future

Freeze the Historical Prediction Before Revealing Later Observations

Frozen forecast, checked against later observations

NASA B0007

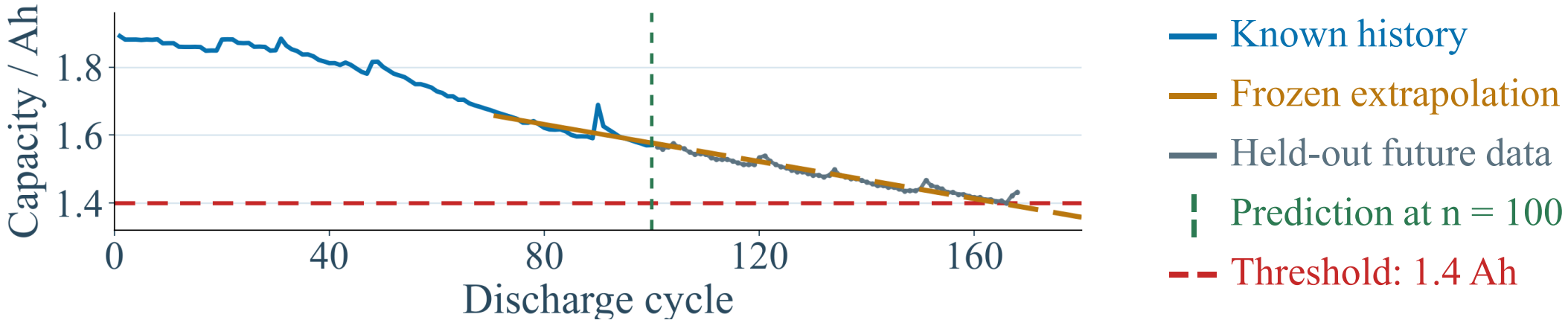


Next-20 MAE = 0.0099 Ah; true RUL remains unknown

NASA B0007 laboratory capacity: freeze at cycle 100; validate against later observations.
No crossing of 1.4 Ah is observed, so true remaining useful life is unknown.

7.5.4 Health Prognostics: Separating History from the Future

Test Predictions against Later Observations



B0007 history, frozen extrapolation, and held-out observations

Mean absolute error (MAE): mean |predicted $\hat{Q}(n)$ − measured $Q(n)$|.

$$\mathrm{MAE}_{101:120} = \frac{1}{20}\sum_{n=101}^{120} |\hat{Q}(n) - Q(n)| \approx 0.0099\ \mathrm{Ah}$$

7.5.4 Health Prognostics: Sources of Errors and Intervals

A Scenario Range Is Not a Confidence Interval

Fitted current capacity $\hat{Q}(100) = 1.5769$ Ah; threshold $Q_{\mathrm{lim}} = 1.4$ Ah. Endpoint rates d are teaching assumptions.

$$\widehat{\mathrm{RUL}} = \frac{\hat{Q}(100) - Q_{\mathrm{lim}}}{d}, \quad d > 0$$

Capacity loss / (Ah/cycle)	RUL / cycles	Meaning
0.0020	88.4	Slower-degradation scenario
0.002731	64.8	|Slope|, cycles 71–100
0.0040	44.2	Faster-degradation scenario

Intervals: error model or independent calibration, with coverage. Here: sensitivity only.

7.5.4 Health Prognostics: Validation and Right Censoring

Without a Threshold Crossing, End of Life Is Unobserved

Prediction cutoff	Fit last 30 points	RUL / cycles	Next-20 MAE / Ah
60	31–60	82.5	0.0259
80	51–80	38.5	0.0431
100	71–100	64.8	0.0099

Right censoring: 168 cycles, with no capacity ≤1.4 Ah; minimum ≈1.4005 Ah.
Observation stops before EOL is observed.
True EOL >168: the cycle-100 prediction of 164.8 is at least 3.2 cycles early.

7.5.4 Health Prognostics: NASA Data and Offline Tasks

Read the Data Structure Before Choosing a Prognostics Tool

Repository item 5 MAT files: cycle → select type = discharge → data.Capacity (Ah).
Offline task: truncate the supplied capacity history, fit a trend, then test the next 20 cycles.

NASA PCoE prognostics repository: aging-data structure and test description

https://data.nasa.gov/dataset/li-ion-battery-aging-datasets

ProgPy models and data-loading code

https://github.com/nasa/progpy

ProgPy nasa_battery: item 11 random-load cells RW1−RW28.
B0007: item 5, Battery Data Set; read its MAT structure.
End of discharge (EOD): one discharge ends. End of life (EOL): degradation threshold reached.

Exercises and Discussion

When Should Residuals Trigger Alarms? Which Inputs Leak Labels?

Q1: Normal residuals have $\mu_0 = 0.04$ rad/s and $\sigma_0 = 0.05$ rad/s. Alarm after two consecutive samples with $|z_k| > 3$.

At 0.1 s intervals from 0.0 to 0.5 s, residuals are 0.10, 0.18, 0.21, 0.12, 0.20, 0.22 rad/s. The fault takes effect at 0.35 s. Find each z_k, alarm time, and detection delay.

Q2: Train a diagnostic model using fault-injected simulation logs. Four candidate inputs are: ① residual window; ② logged fault ID; ③ injection time; ④ motor-output commands.

Which inputs cause label leakage, and why?

Reference Points

Subtract the Mean before Scaling; Use Labels as Targets

$$z_k = \frac{r_k - \mu_0}{\sigma_0} = \frac{r_k - 0.04}{0.05}$$

z_k: 1.2, 2.8, 3.4, 1.6, 3.2, 3.6

Counts: 0, 0, 1, 0, 1, 2; alarm confirmed at 0.5 s. Omitting μ_0 gives 2.0, 3.6, 4.2… and a false alarm at 0.2 s.

$$\Delta t_{\text{det}} = 0.5 - 0.35 = 0.15\ \text{s}$$

Q2: ②③ come from labels or the test setup and are unavailable in deployment: label leakage. Inputs ①④ are available in operation; still evaluate using complete-flight splits.

Natural labels: injection parameters become training targets, never inputs.

Exercises and Discussion

Should a Higher Overall Score Mean a Higher Threshold?

Teaching: 200 windows (180 normal, 20 faulty). Threshold 0.7: TP=10, FN=10, FP=2, TN=178.

Q1: At threshold 0.7, compute accuracy, precision, and F1, the harmonic mean of precision and recall.

Q2: Compare with threshold 0.5: accuracy 93.5%, recall 75%, and false-positive rate 4.4%. How would you choose?

Q3: Fault onset is 2.3 s; confirmation needs three consecutive samples with $|z_k| > 3$. At 0.2 s intervals from 2.0–3.2 s, z_k is 0.5, 3.3, 3.5, 2.8, 3.2, 3.6, 4.1. Find detection delay.

Reference Points

Higher Thresholds Reduce False Alarms but Miss More Faults

$$\text{Accuracy} = \frac{10 + 178}{200} = 94\%$$

$$\text{Recall} = \frac{10}{20} = 50\% \qquad \text{FPR} = \frac{2}{180} \approx 1.1\%$$

Precision: 10/12≈83.3% (65.2% at 0.5). F1=2TP/(2TP+FP+FN)=20/32=62.5%.

Fewer false alarms; misses rise 25%→50%. Choose by mission consequences and response costs.

Q3: reset at 2.6 s; confirm the third exceedance at 3.2 s. Fault onset: 2.3 s.

$$\Delta t_{\text{det}} = 3.2 - 2.3 = 0.9\ \text{s}$$

Exercises and Discussion

Does This Capacity Curve Validate the RUL Prediction?

Q1: At cycle 100, can a prediction still be history-only if cycles 101–120 are used for fitting or window selection?

Q2: Records end at cycle 168 without crossing 1.4 Ah. Can 168 be taken directly as end of life?

Q3: Capacity MAE over the next 20 cycles is about 0.0099 Ah. Does this establish a small RUL error?

Q4: The 44.2–88.4-cycle range comes from rates of 0.004–0.002. Why is it not a 95% confidence interval?

Reference Points

Capacity-Error Evidence Still Lacks an Observed End of Life

History isolation: the cycle-100 prediction uses only observations through 100. Future data validate the frozen model.

Right censoring: cycle 168 has not reached the threshold. The record endpoint is not end of life; true RUL is unobserved.

Evaluation scope: 0.0099 Ah is capacity MAE over the next 20 cycles. Censoring only bounds RUL prediction error.

Interval meaning: 44.2–88.4 cycles is sensitivity to hand-chosen rates. A 95% interval needs its generation method and coverage.

Section Summary

Support Diagnosis and Prognostics with Evidence

Residuals and baselines: match inputs, time, frames, and units. Calibrate normal bias while retaining fault deviations.

Fusion and evaluation: combine residuals, sensor relations, and time. Keep labels out of inputs; test independent conditions.

False alarms and delay: report denominators, thresholds, and the timing origin.
AUC differs from accuracy; forward-inference time differs from detection delay.

Health prognostics: distinguish SOC, SOH, and RUL. Use only history; state thresholds, interval sources, and censoring.