PINNeAPPle

STUDY REPORT · PUBLIC EXAMPLE

HX-201 digital twin a fouled heat exchanger

An accuracy contract signed before operation — and what happened when the fouling showed up.

0.216 °C
of RMSE against the held-out probe

95.5%
band coverage, calibrated outside training

2.6 h
of alarm lead time, with the forecast error



Document TW-201 · version 1.0 · 2026
Synthetic demonstration history: not a plant commissioning record
Reference: probe TT-105, held out of the filter
PINNeAPPle · LAIS & YAN LTDA · pinneapple-labs.github.io

A shell-and-tube exchanger loses efficiency to fouling. Operations does not want a forecast: it wants to know **how much time it has** before the downstream load stops closing.

The risk of a twin is not getting the forecast wrong — it is being heard without deserving it. So the study starts with a contract written before any data: the metrics, the ranges and the limits the model must meet in order to exist.

What was contracted

- Model `exchanger-ntu 1.0`, with a declared accuracy floor of 1 °C.
- Acceptance gates: RMSE ≤ 2.5% of span, bias ≤ 1%, coverage ≥ 90%, alarm lead ≥ 1 h.
- A redundant probe (TT-105) that never enters the filter: it is the judge.
- Validity range declared in the contract: T from 80 to 110 °C and $\dot{m}$ from 0.5 to 2.0 kg/s.

> **Order matters.** The contract is signed before the first filter step. A metric chosen after seeing the result is not a metric: it is an excuse.

Predicted fields and process limits

FIELD	RANGE	PROCESS LIMIT
T_ho (hot outlet)	35–75 °C	alarm above 60.0 °C
T_co (cold outlet)	45–75 °C	alarm below 57.8 °C
R_f (fouling)	0–0.0025 m²K/W	no engineering limit declared

Sensors

ROLE	SENSORS	USE
Inputs (operating point)	TT-101, TT-102, FT-101, FT-102	drive the model
Assimilated	TT-103, TT-104	enter the filter, noise 0.15 °C
Held out	TT-105	only compares; never enters the filter

Acceptance gates

GATE	CRITERION	WHY
RMSE	≤ 2.5% of the 40 °C span	≈ 1 °C, the model's own accuracy floor
Bias	≤ 1%	systematic error cannot hide in the mean
Band coverage	≥ 90%	the band has to contain what happens
Alarm lead	≥ 3,600 s	a warning that arrives late is not a warning

Validity range: T from 80 to 110 °C and $\dot{m}$ from 0.5 to 2.0 kg/s. Outside it the twin promises nothing — and the report says so instead of extrapolating quietly.

852 steps, 255,300 s (70.9 h) in the validation window. Metrics are computed only on the part that was not used to calibrate the band.

0.216 °C
RMSE against the held-out probe

0.168 °C
mean absolute error

0.432 °C
p95 of absolute error

−0.010 °C
bias

GATE	CRITERION	ACHIEVED	STATUS
RMSE	≤ 2.5% of span	0.54% of span	passes
Bias	≤ 1%	−0.026% of span	passes
Coverage	≥ 90%	95.5%	passes
Alarm lead	≥ 3,600 s	9,300 s	passes

Fouling estimated at the end of the window was **0.001652 m²K/W**, against 0.001637 in the scenario's true value — 0.9% apart. It is the variable operations cannot measure and that the twin starts estimating.

> **The judge is external.** Probe TT-105 stayed out of the filter for the whole commissioning. The metric shown here is the distance between the model and a measurement it never saw.

The uncertainty band is not a prediction interval: it is a state band, calibrated once on a separate window (the first 127,500 s) and then scored on the rest. The scale applied was ×1.336; without it, the raw ensemble band would cover 86.4% — below the 90% gate.

That is stated in the report, not hidden: a band that only works after being scaled is a scaled band, and the reader needs to know that to decide what to do with it.

The alarm

INDICATOR	VALUE	READING
Alarm lead	9,300 s (2.6 h)	time between the warning and the limit crossing
Crossing forecast error	3,900 s (1.1 h)	distance between the predicted and the measured instant
Where the lead comes from	linear extrapolation of the latent state	it grows with any change of regime

> **House rule.** Lead time without a stated error is a guess with a clock. The number operations uses is the pair: 2.6 h of warning, with 1.1 h of error on the crossing.

What is not proven

ITEM	WHAT THE REPORT SAYS
The band is a state band	not a prediction interval for a change of regime
Commissioned inside the declared range only	outside T 80–110 °C or $\dot{m}$ 0.5–2.0 kg/s, no claim
First 12 frames	assimilated, but excluded from the metrics
No lead time for R_f	no engineering limit was declared for fouling
Synthetic history	a demonstration; not a plant commissioning record
No re-fit	this is the twin as shipped, not a model retrained later

> **Why publish a twin's limits.** Because its value is being auditable: a model that declares where it does not hold can be used; one that does not, cannot.

95.5%	0.216 °C	2.6 h	0.001652
band coverage	RMSE against the probe	of alarm lead	m^2K/W estimated fouling

What operations does with it

Lead time is time to plan a cleaning, negotiate a shutdown window or shift load. The twin decides nothing: it warns, and states the error it is warning with.

What this study does not decide

The economic moment of the shutdown, the cleaning vendor and the cost of lost head. It delivers the fouling estimate and the warning window; the rest is an operations decision.

Provenance

WHAT	WHERE
Accuracy contract	`PINNeAPPle/apps/twin_ops/spec_hx201.yaml`
Commissioning evidence	`twin_evidence_hx201.json` · spec 6b86712f52ed2644
Preconditions	11 preflight checks, 11 passed
Public article	`pinneapple-labs.github.io/artigos/gemeo-hx201/`

> **This is an example report,** on a synthetic demonstration history. A twin for your plant follows the same structure: contract first, held-out probe as judge, and the same limits page.