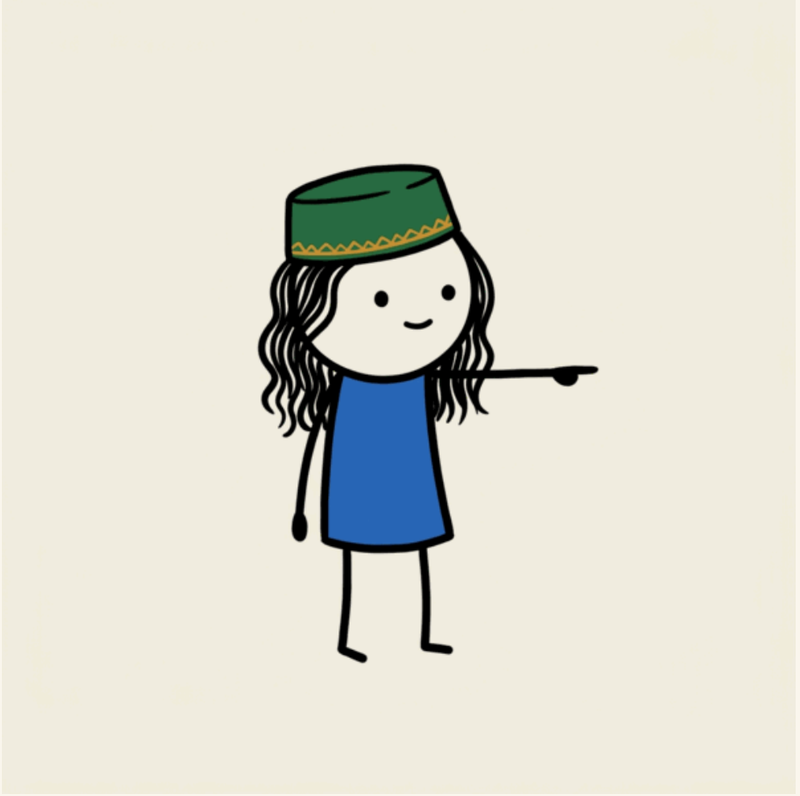
Would you keep this pump running?

Choose measurements and features that keep the intended fault visible.

One reading looks normal.

The recent history may tell another story.

Intro to ML · Two lectures + a practical lab



Time changes the question

> The order, spacing, and context of observations carry information.

A robot

IMU and joint currents

Is it moving normally?
What happens next?

A person

EEG and heart activity

Which state or event produced this epoch?

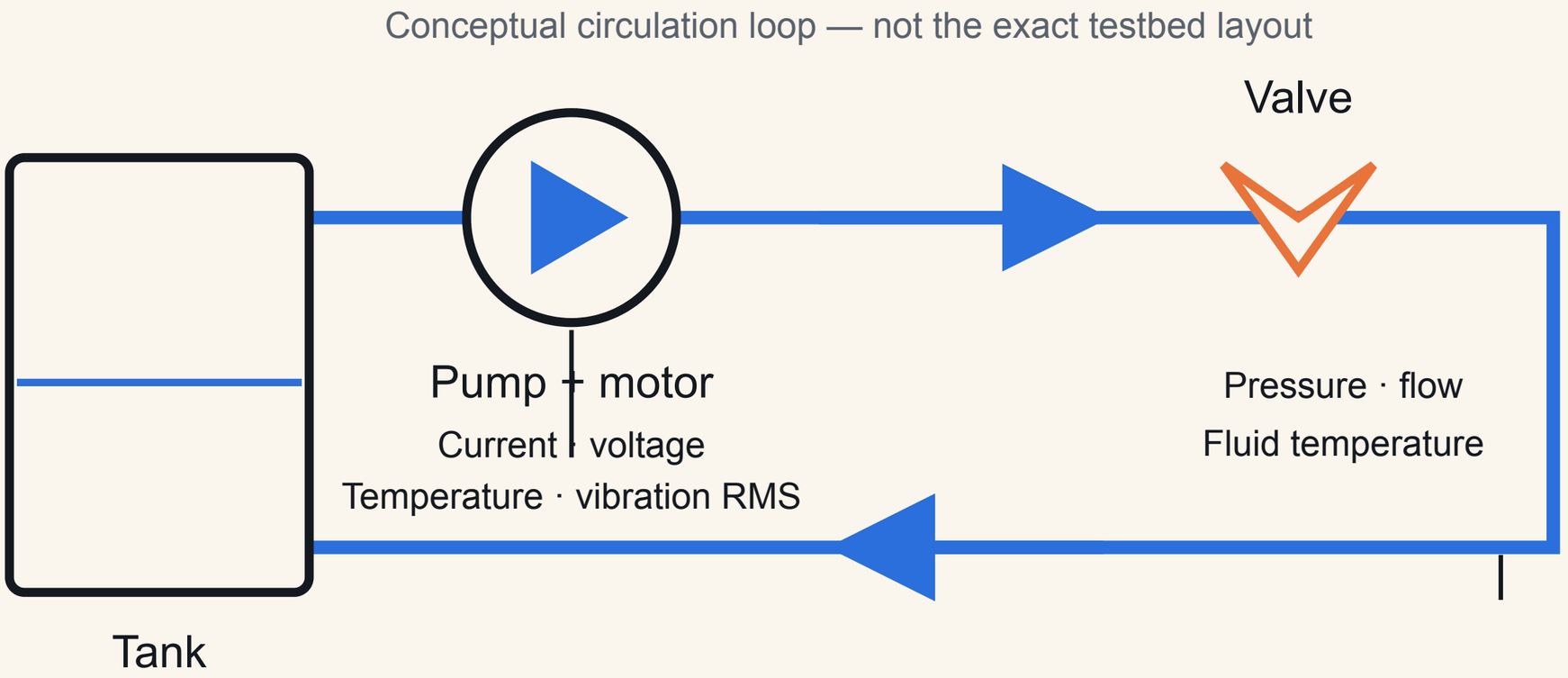
A service

Requests, latency, energy

Is this spike a fault, a daily pattern, or growth?

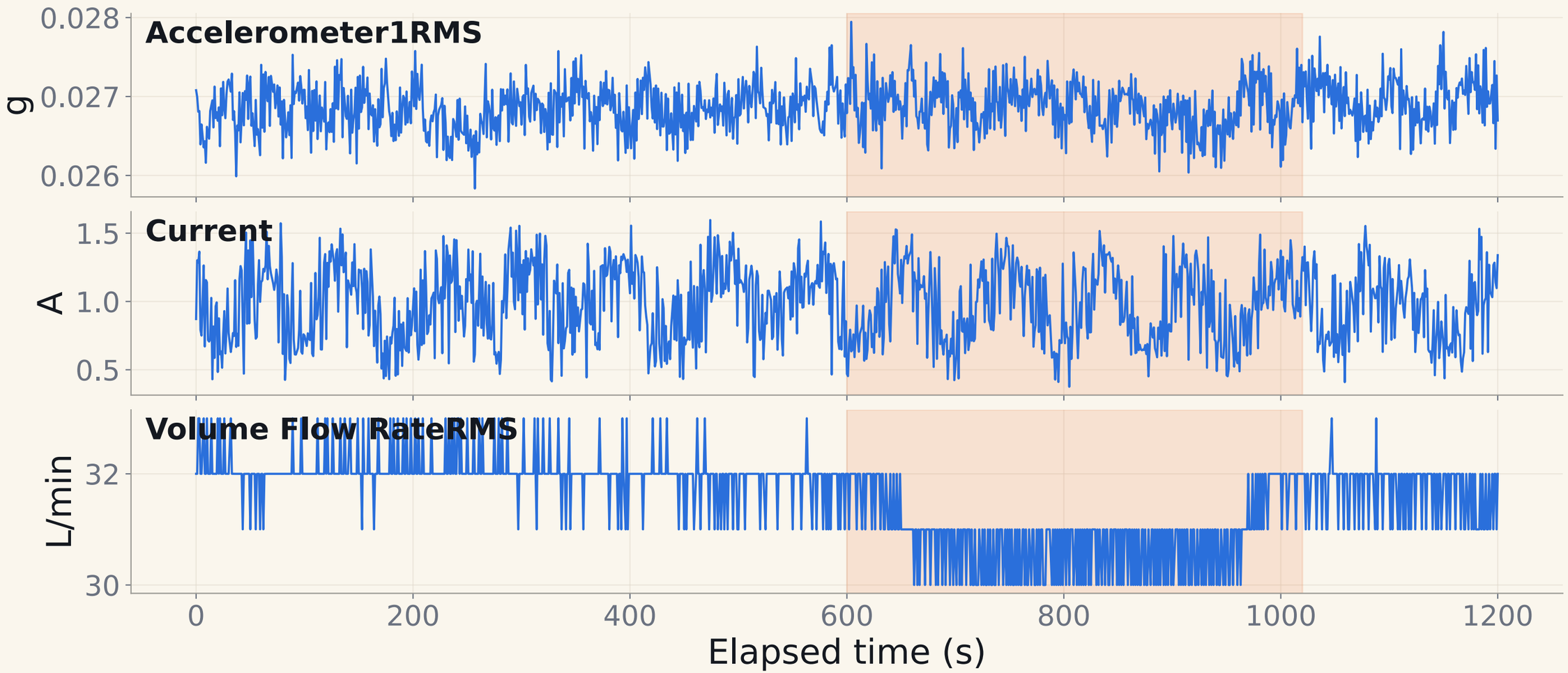
One physical system, eight channels

A water-circulation testbed links **motor, hydraulic, thermal, and vibration summaries**.



Keep the whole abnormal state visible

State detection: keep flagging the interval. Onset: mark when it begins.



Real SKAB valve1/1.csv · orange = provided anomalous-state label · no model predictions

An observation is more than a number

Timestamp
when?

→

Value + unit
what quantity?

→

Channel + quality
where? trustworthy?

$$X \in \mathbb{R}^{T \times C}$$

> A row is a synchronized observation, not an independent example.

What does one row mean?

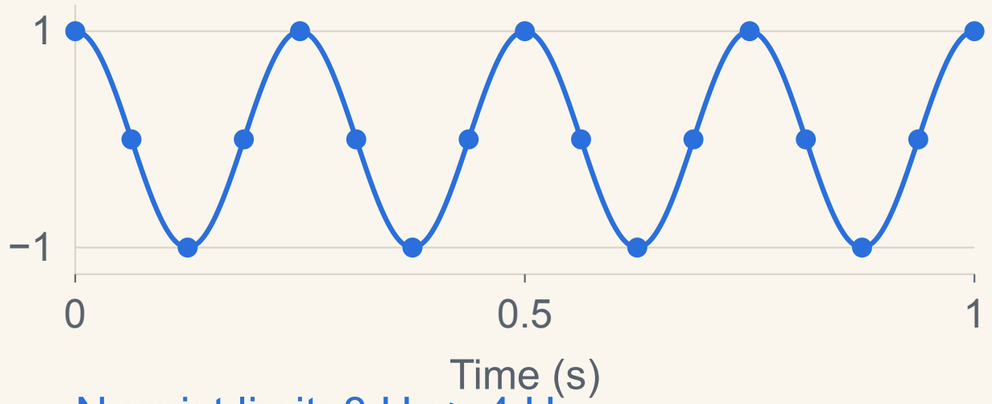
The clock sets units; the sampling rule sets which frequencies can be recovered.

Regular grid: $f_s = 1 / \Delta t$

Nyquist–Shannon: $f_s > 2\, f_{max}$

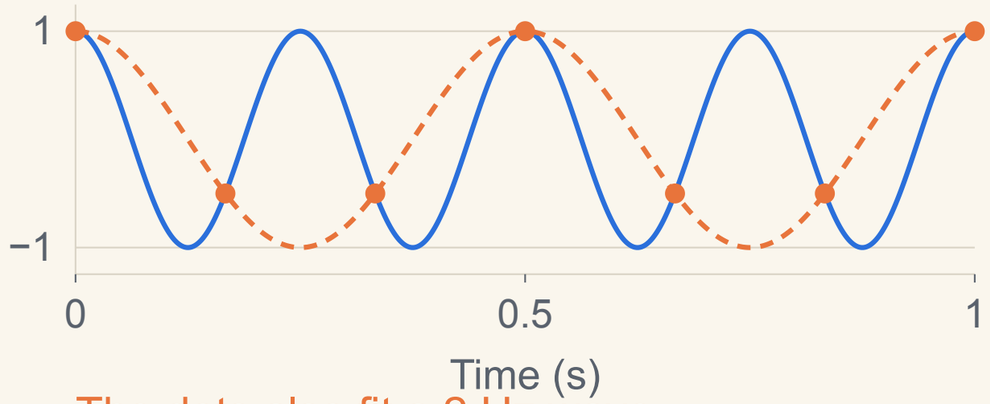
Band-limited signal + ideal uniform samples → unique reconstruction

4 Hz signal · sample at 16 Hz



Nyquist limit: 8 Hz > 4 Hz

4 Hz signal · sample at 6 Hz



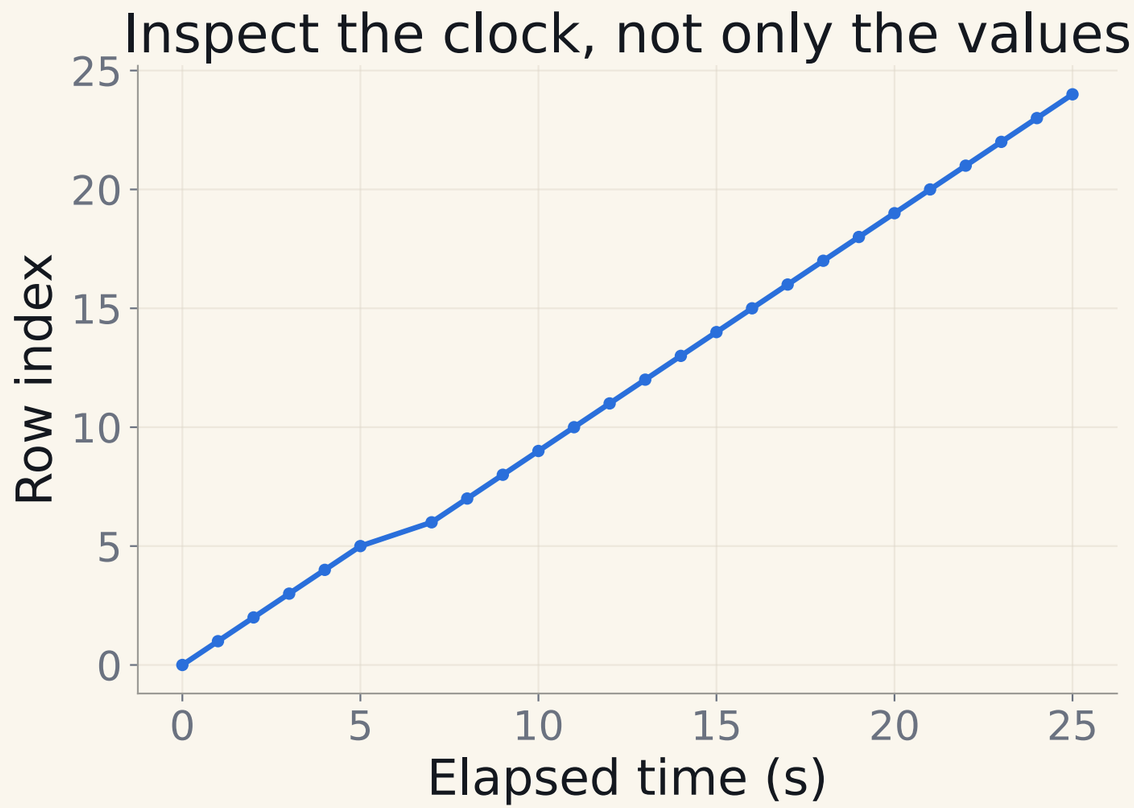
The dots also fit a 2 Hz wave

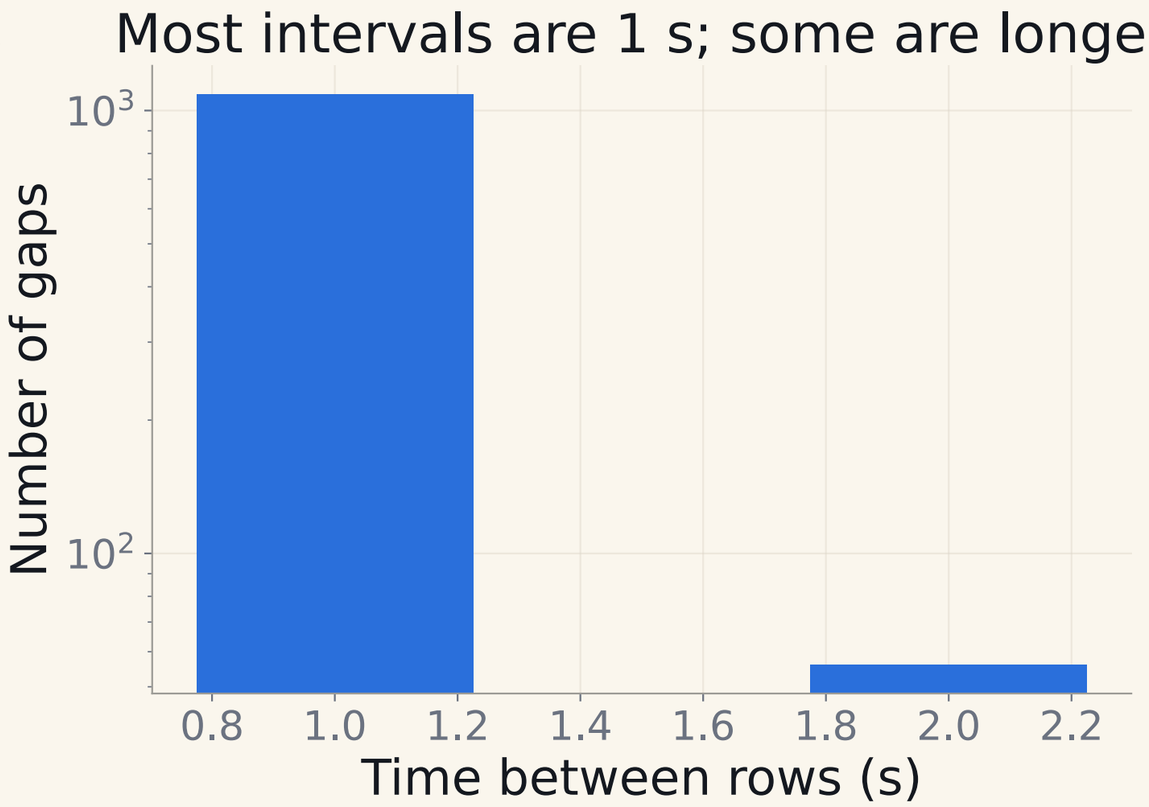
Before sampling: anti-alias filtering removes content above the usable band.

History looks backward; horizon looks forward. Both need the clock’s units.

One row is not always one second

Regularize deliberately. Keep an **observation mask**; fill only by a declared causal rule.

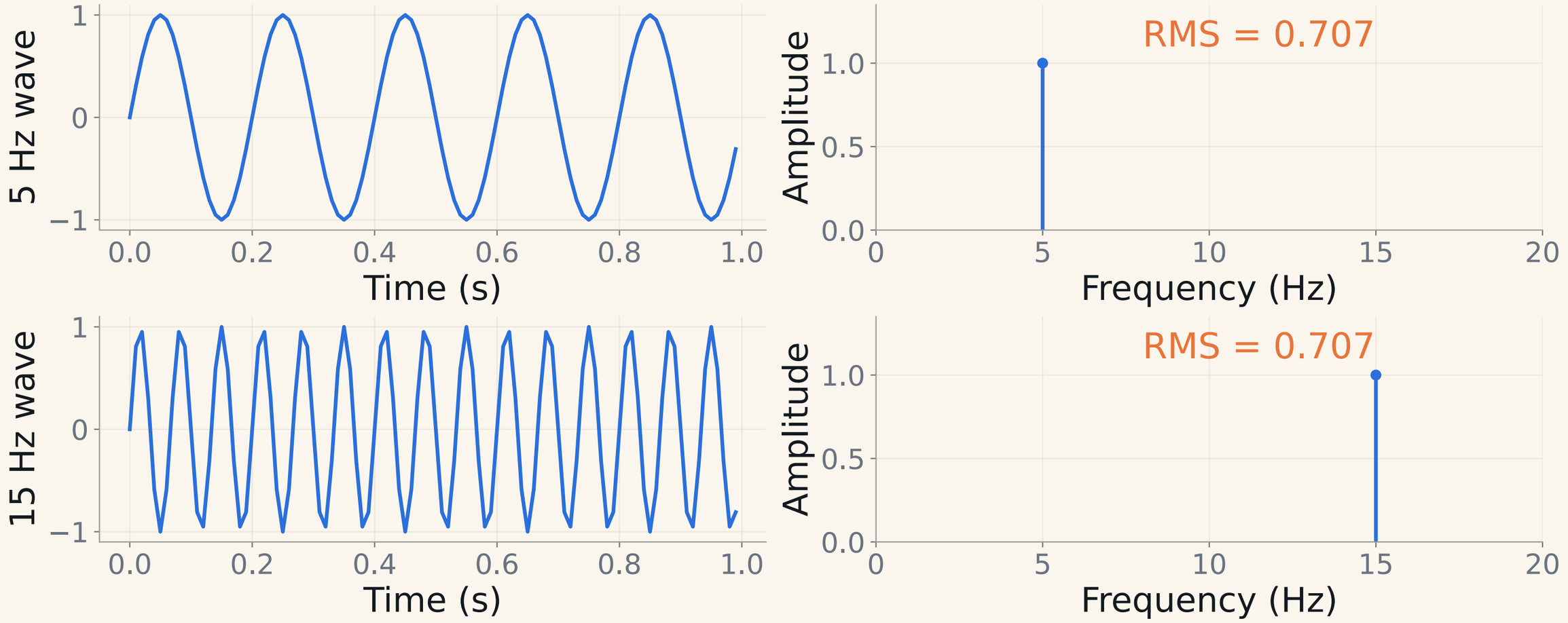




Real SKAB valve1/1.csv · inspect Δt before resampling

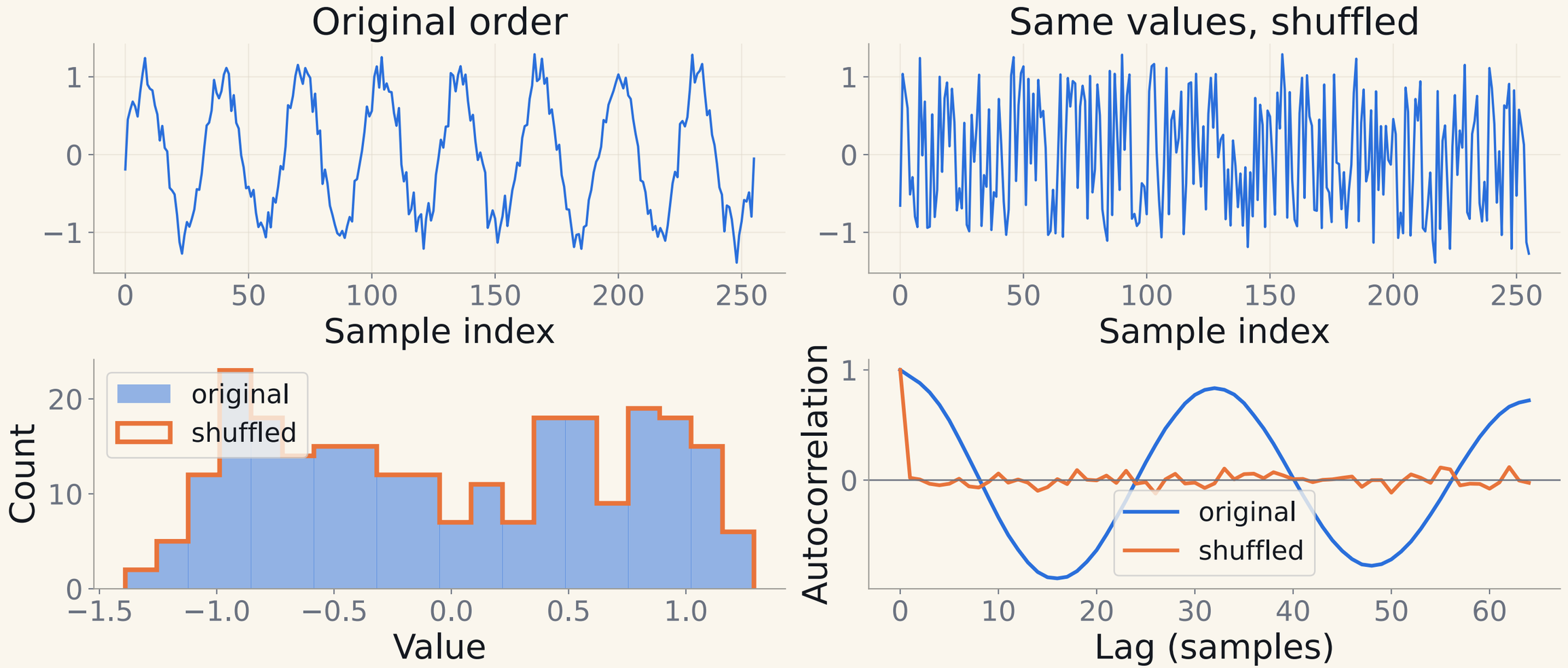
What did RMS erase?

Same energy summary. Different oscillation frequencies. **RMS cannot recover the original vibration spectrum.**



Constructed teaching signal · fixed seed 20261029

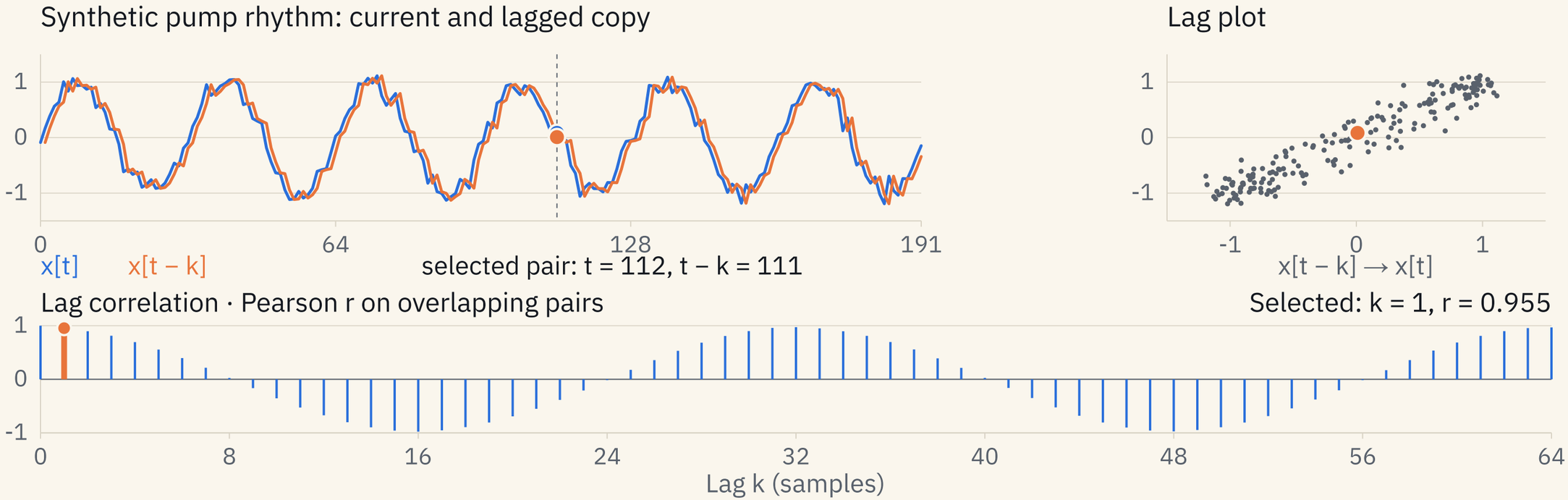
Same values. A different story.



Constructed teaching signal · fixed seed 20261029

How far does memory reach?

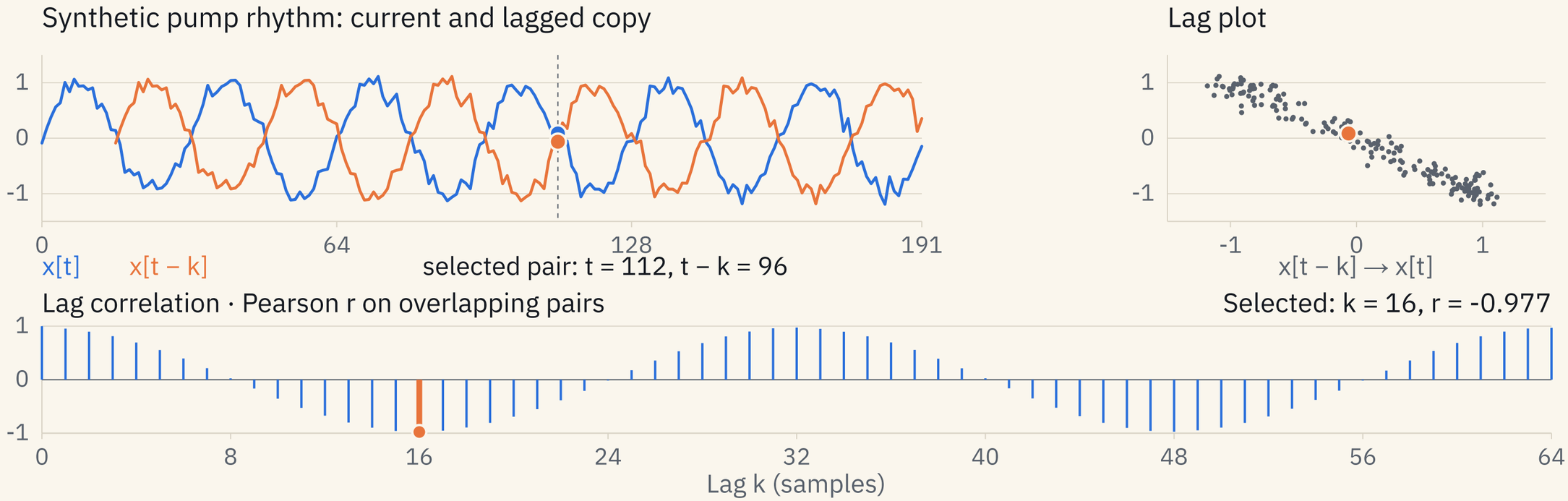
Compare the pump rhythm with a shifted copy. Which points form one pair?



Keyframe 1/3 · Shift by one sample — k = 1 · pair (111, 112) · Pearson r = 0.955 over 191 overlapping pairs

How far does memory reach?

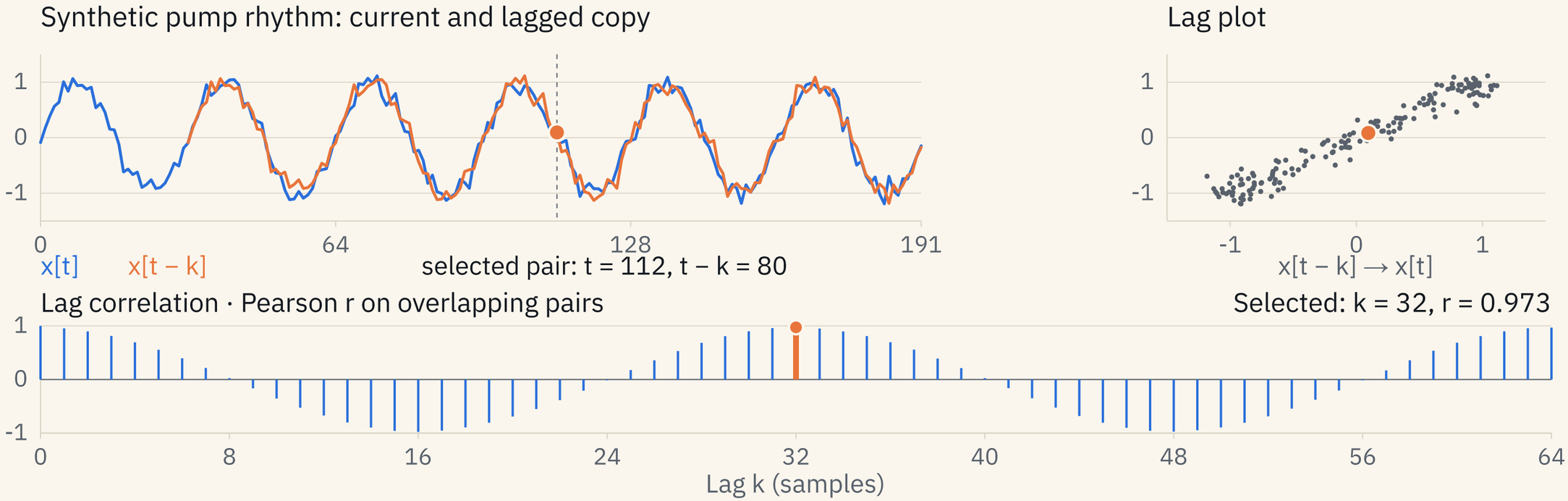
At lag 16, a high reading tends to align with a low reading.



Keyframe 2/3 · Half a cycle — k = 16 · pair (96, 112) · Pearson r = -0.977 over 176 overlapping pairs

How far does memory reach?

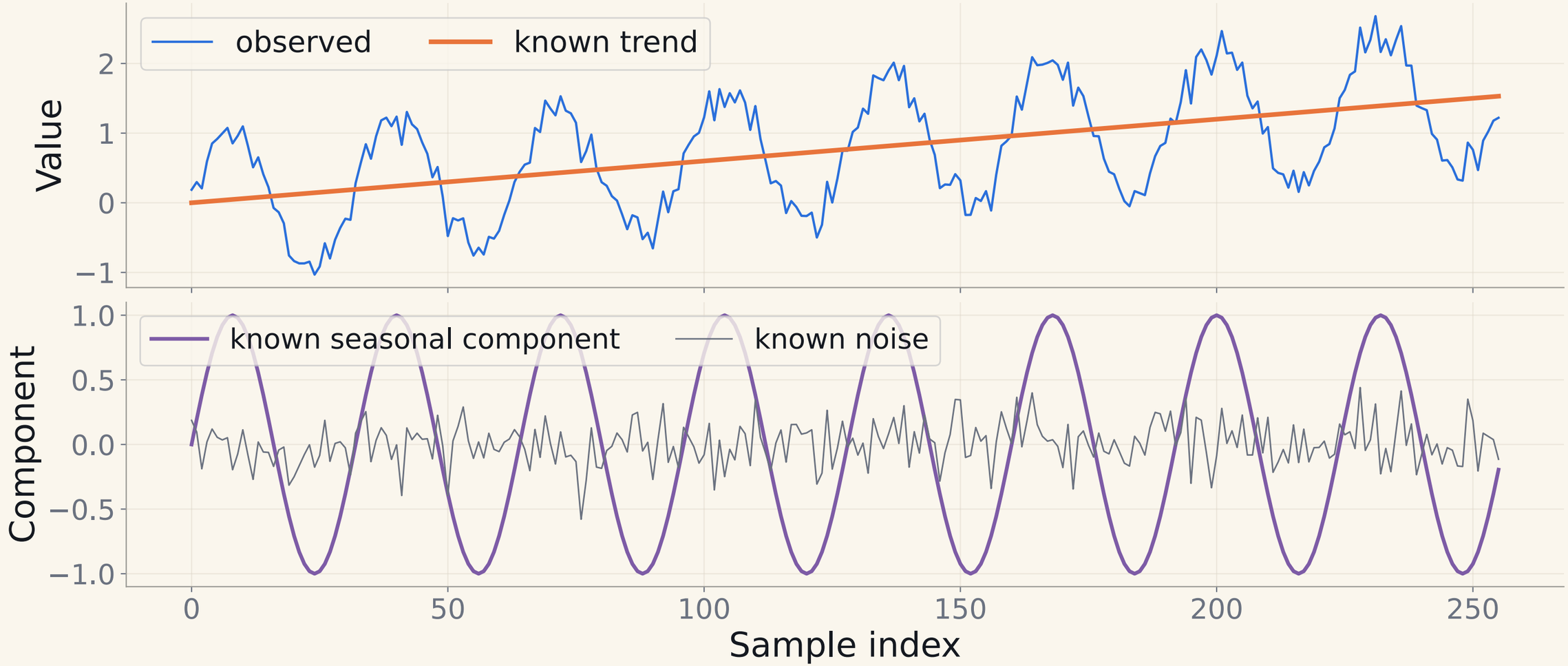
At lag 32, the rhythm aligns again. ACF describes dependence across lags.



Keyframe 3/3 · One whole cycle — k = 32 · pair (80, 112) · Pearson r = 0.973 over 160 overlapping pairs

Trend, seasonality, and residuals

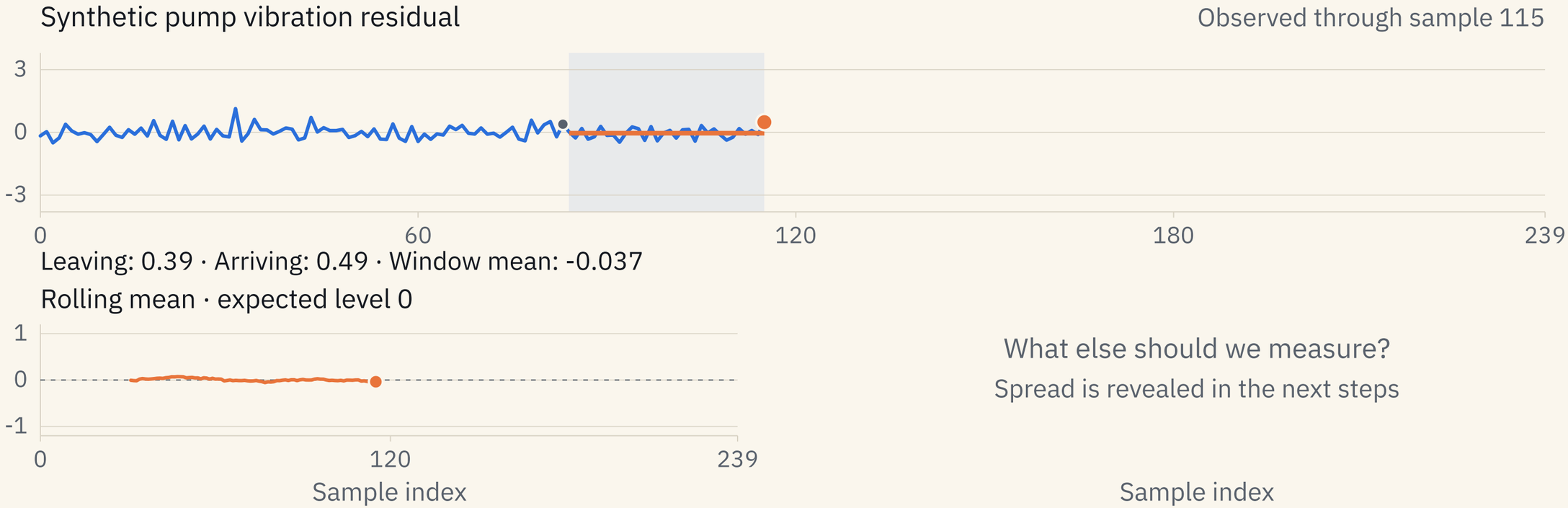
For this teaching signal: **x[t] = trend[t] + seasonal[t] + noise[t]**.



Constructed teaching signal · fixed seed 20261029

A flat mean can hide a change

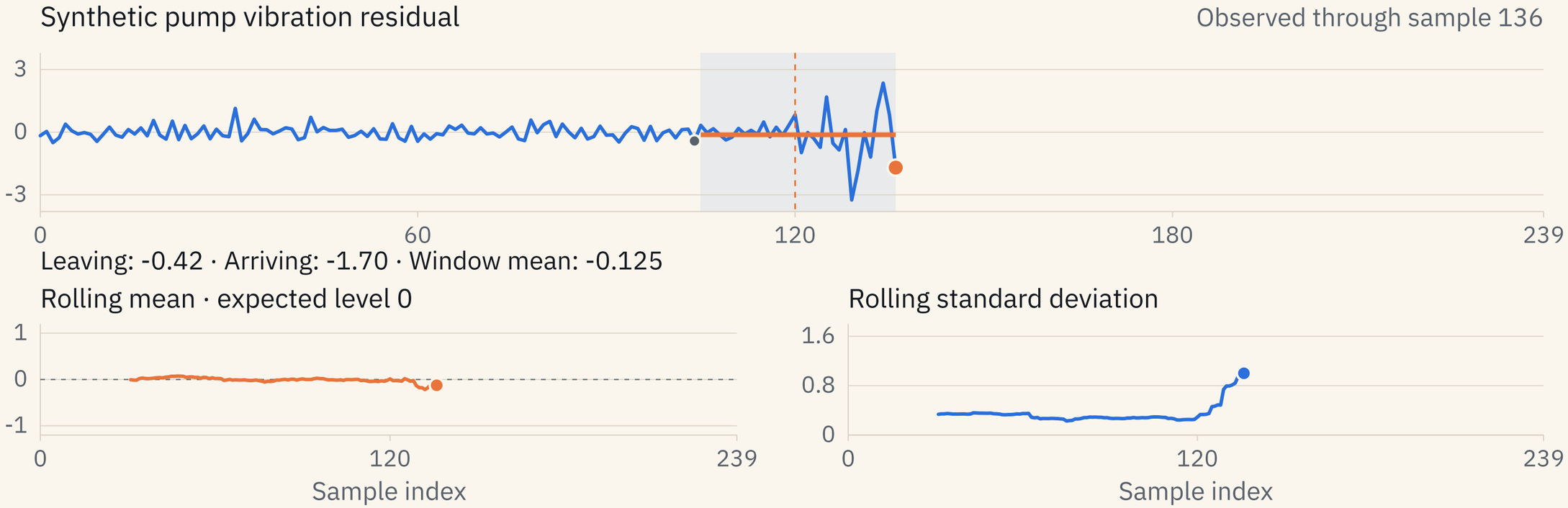
Predict the next curve: will a larger spread produce a sustained change in the mean?



Keyframe 1/3 · Before the change — Window 84–115 · mean -0.037

A flat mean can hide a change

Now reveal standard deviation. The window mixes the old and the new operating regimes.

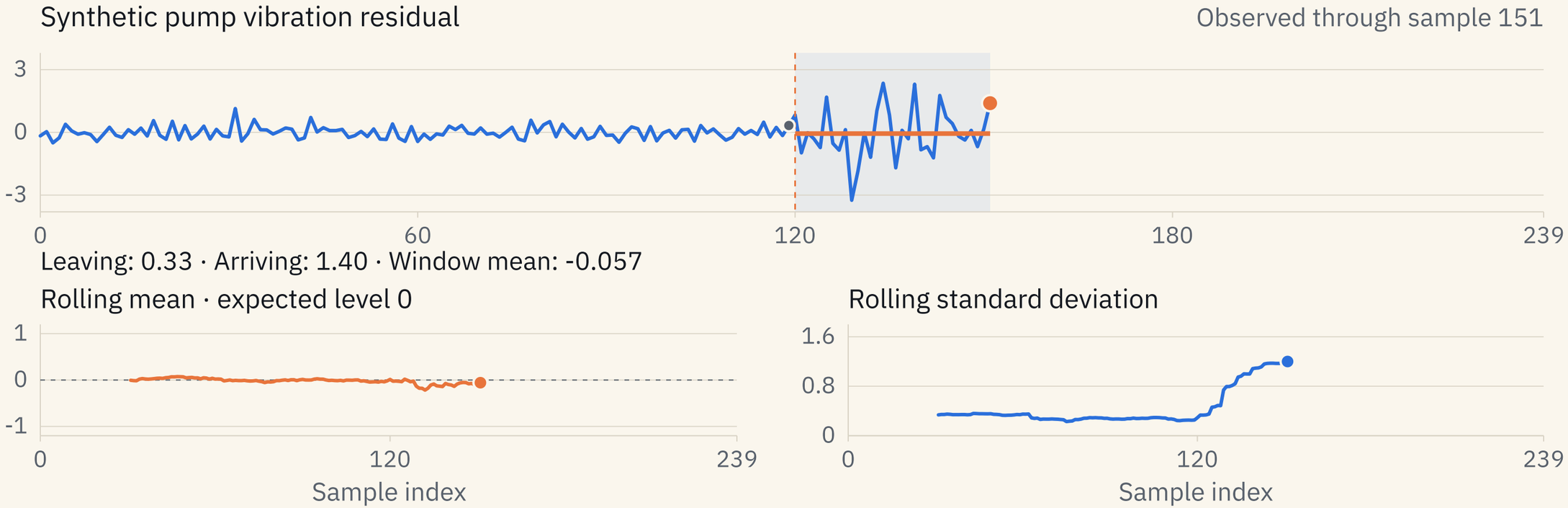


Keyframe 2/3 · Half a window later — Window 105–136 · mean -0.125 · standard deviation 1.000 · population divisor n

WORKED EXAMPLE

A flat mean can hide a change

Spread has increased; the expected level is still zero. These features answer different questions.



Keyframe 3/3 · A full window later — Window 120–151 · mean -0.057 · standard deviation 1.199 · population divisor n

Stationarity is an assumption to inspect

Focus: Treat stationarity as a modelling assumption that requires evidence.

Weak stationarity

Constant mean and finite variance

Autocovariance depends on **lag**, not absolute time

Useful diagnostics

Compare local means and spreads

Inspect ACF, regimes, and residuals

A flat plot or one statistical test cannot establish every assumption.

What else can change?

Property	Measure or plot	What it can reveal
Level / scale	Median, MAD, quantiles	Offsets, amplitude changes
Dependence	ACF, lag and channel plots	Memory, changed relationships
Observation quality	Gap lengths, masks, saturation	Sensor and logging failures
Local dynamics	Differences, local slope, spectrum	Transitions, oscillation changes

Normal depends on the operating regime

A current value may be normal during startup and unusual at steady speed.

A new level may be a fault, a commanded change, or a new operating mode.

> Labels and the monitoring goal give “unusual” its meaning.



One signal, three coordinate views

Explain what a representation preserves, loses and locates in time.

Time

What happened, and when?

Waveform, lags, local statistics

Frequency

Which rhythms are present?

Fourier coefficients, magnitude, power

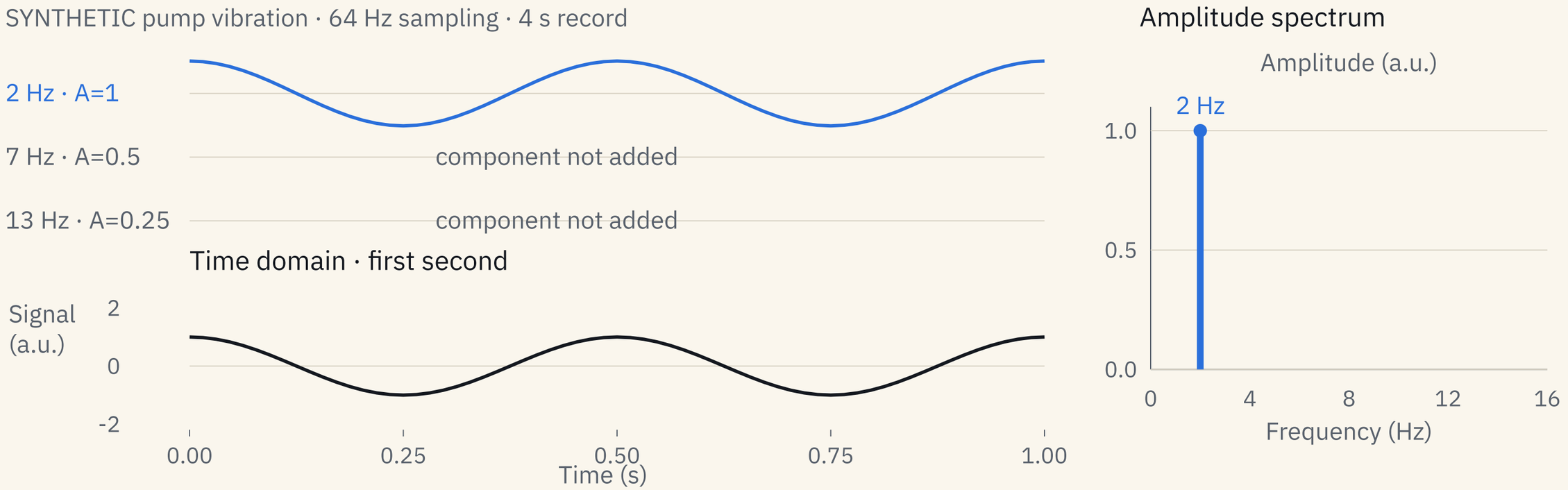
Time–frequency

When did each rhythm appear?

STFT spectrogram, wavelet scalogram

Build a signal from simple waves

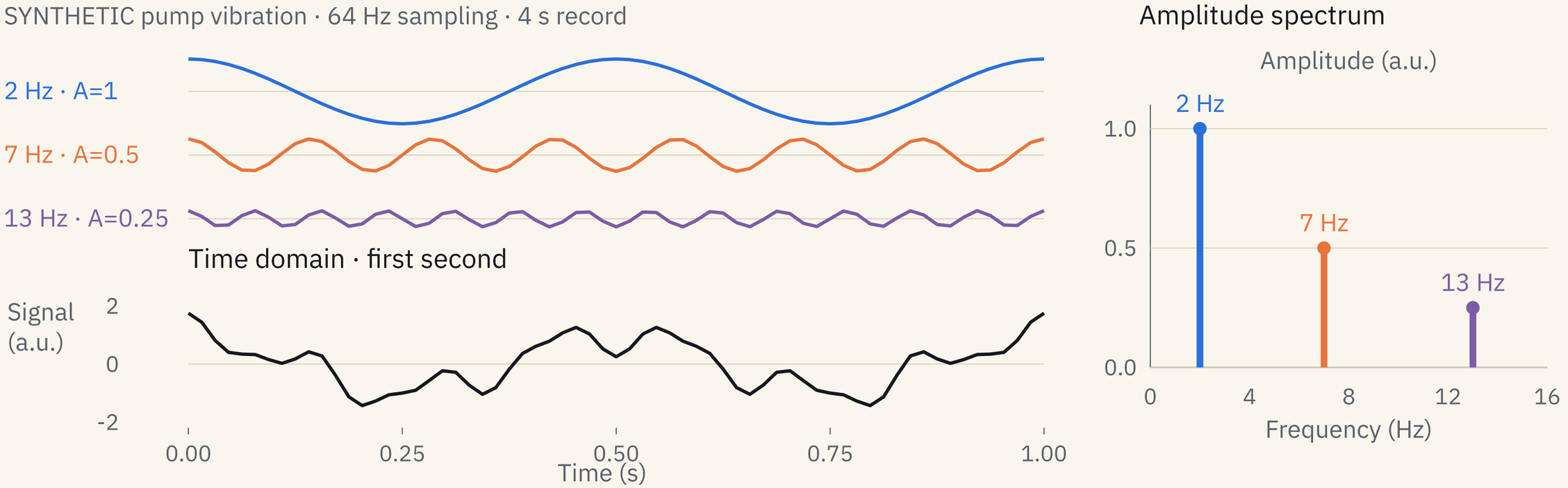
The rotor contributes a 2 Hz rhythm. What will adding a 7 Hz rhythm change?



Keyframe 1/3 · One rhythm — 1 component; 7 Hz phase = 0.00 rad. The spectrum uses the full 4 s record; the waveform shows a 1 s zoom.

Build a signal from simple waves

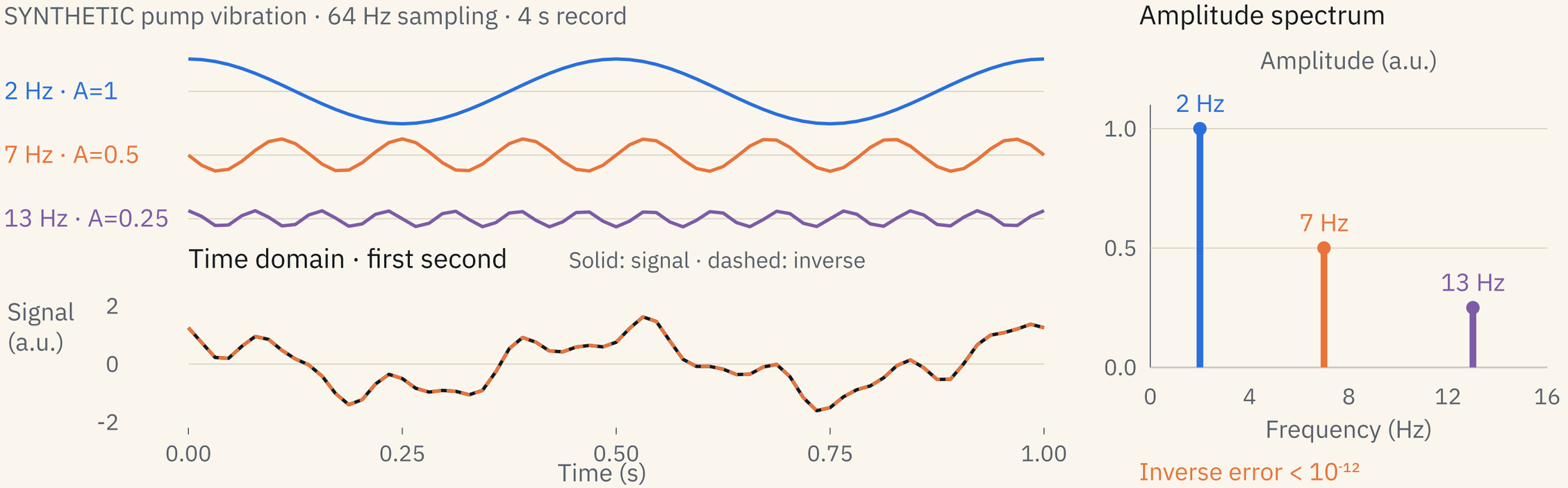
Three frequencies coexist in one sensor trace. Each component keeps its own amplitude.



Keyframe 2/3 · Add a third — 3 components; 7 Hz phase = 0.00 rad. The spectrum uses the full 4 s record; the waveform shows a 1 s zoom.

Build a signal from simple waves

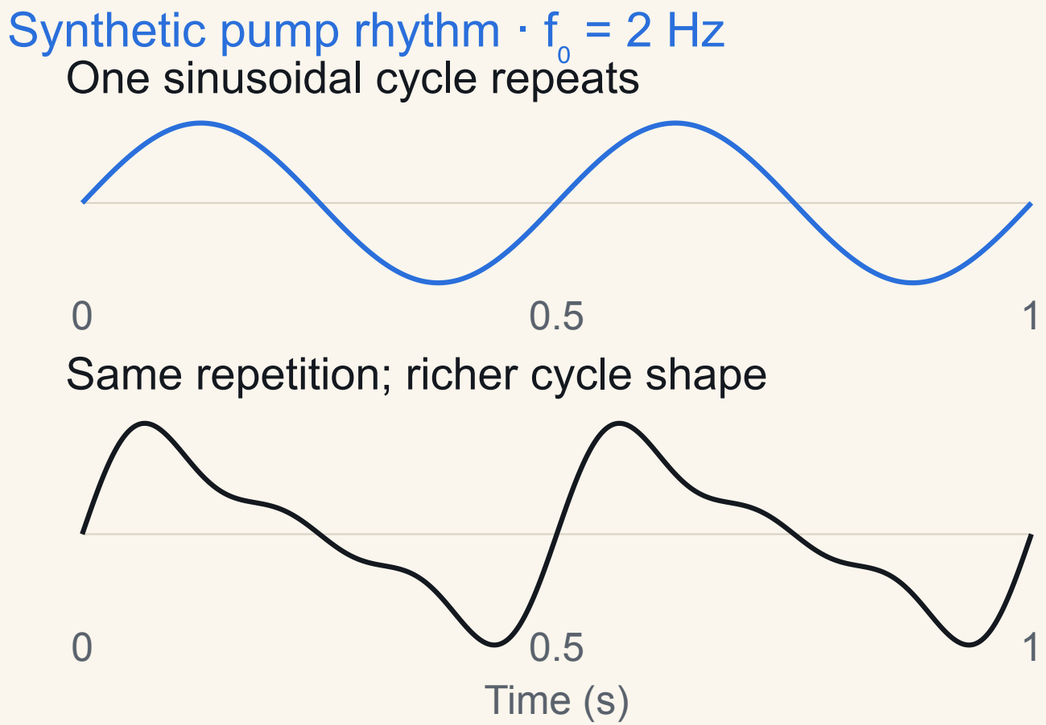
Keep complex coefficients, including phase: the inverse transform recovers every sample.

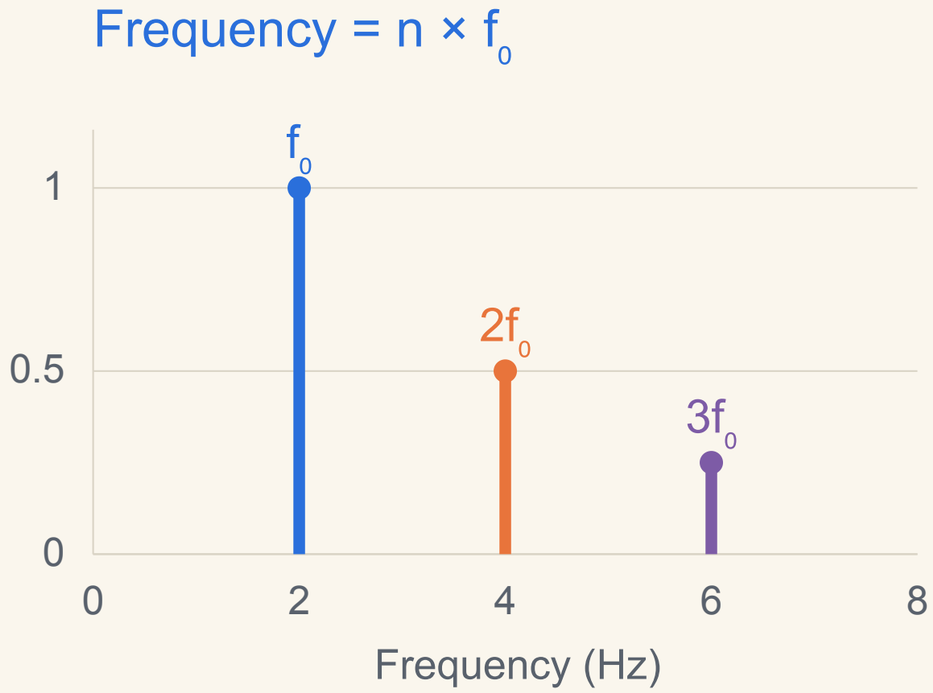


Keyframe 3/3 · Reconstruct — 3 components; 7 Hz phase = 1.57 rad. Inverse reconstruction error < 10^{-12}.

A repeated shape creates harmonics

Harmonics are integer multiples of a fundamental: $f_n = n\, f_0$.



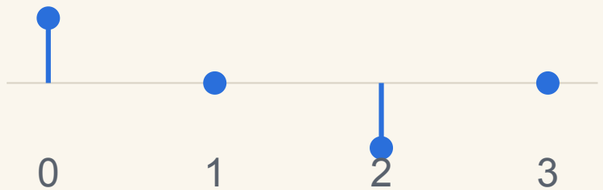


2, 7 and 13 Hz on the previous slide are not all harmonics of 2 Hz.

A Fourier coefficient measures agreement

Multiply samples by a candidate wave, then add: matching structure survives the sum.

Observed: x = (1, 0, −1, 0) · 4 samples at 4 Hz



Dot product: multiply, then add

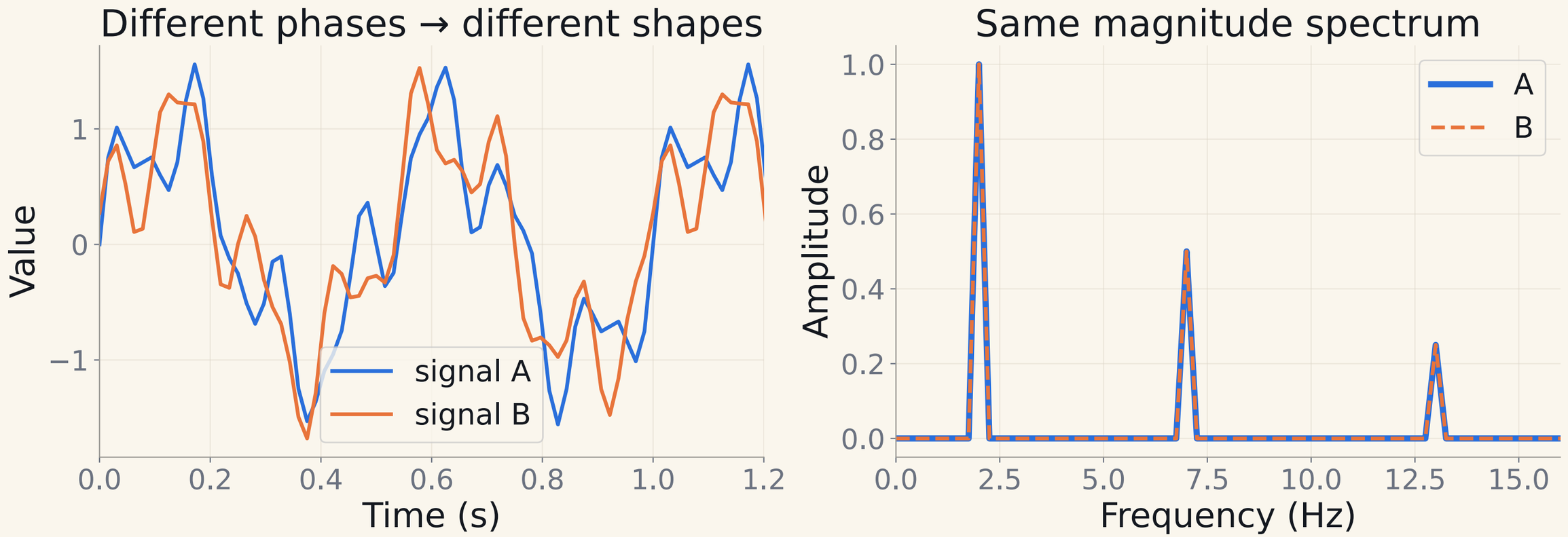
n = 0, 1, 2, 3

		Products x[n] × template[n]	Sum
cos · 1 Hz	1, 0, −1, 0	1 + 0 + 1 + 0 =	2
sin · 1 Hz	0, 1, 0, −1	0 + 0 + 0 + 0 =	0
cos · 2 Hz	1, −1, 1, −1	1 + 0 + (−1) + 0 =	0

Matched cosine: 2 / (N/2) = 1 amplitude. A sine–cosine pair also preserves phase.

Keep the phase

Full FFT coefficients ↔ original samples. Magnitude alone discards phase.



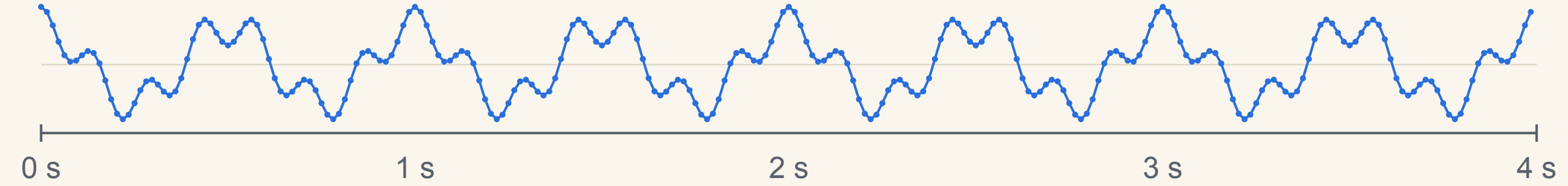
Constructed teaching signal · fixed seed 20261029

Put units on the frequency axis

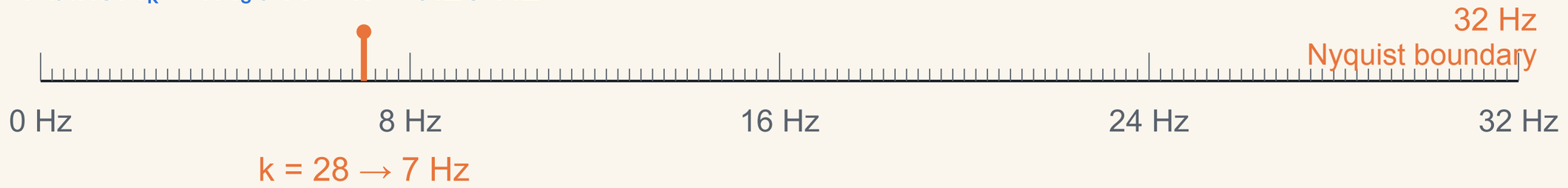
Bin spacing comes from observation duration; zero-padding only samples the spectrum more densely.

$f_s = 64$ Hz $N = 256$ samples $T = N / f_s = 4$ s

One observed record: 256 samples on [0, 4 s)



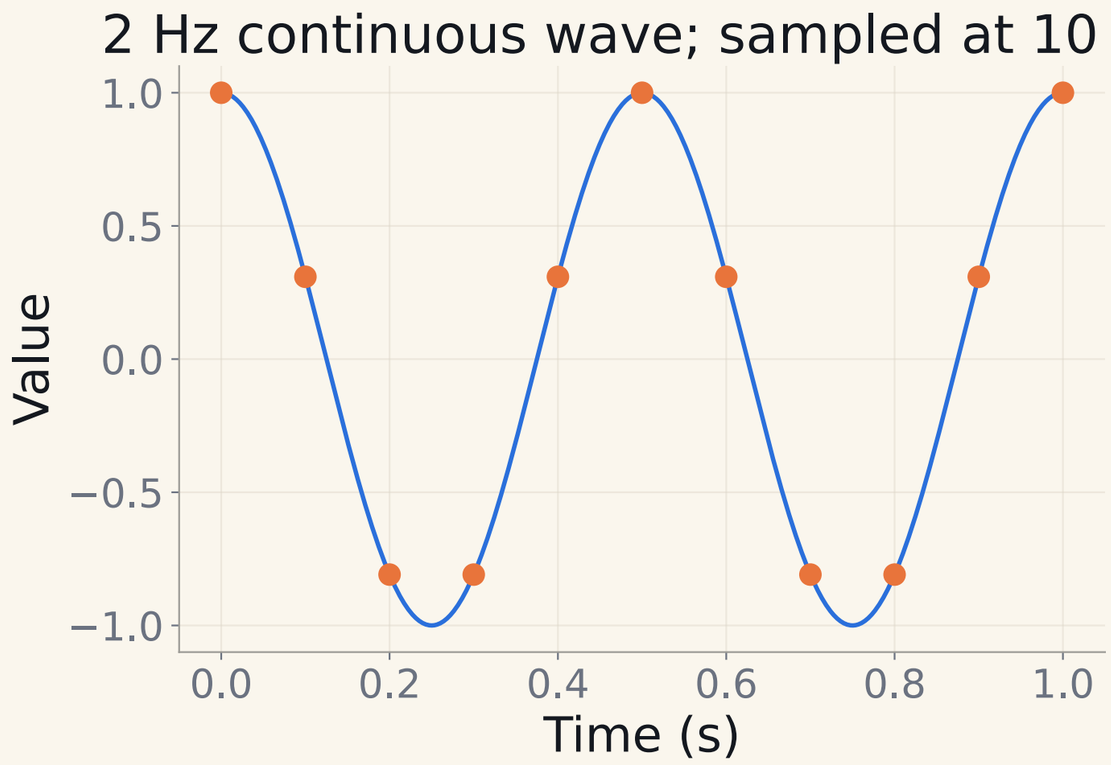
DFT bins: $f_k = k\, f_s / N = k \times 0.25$ Hz

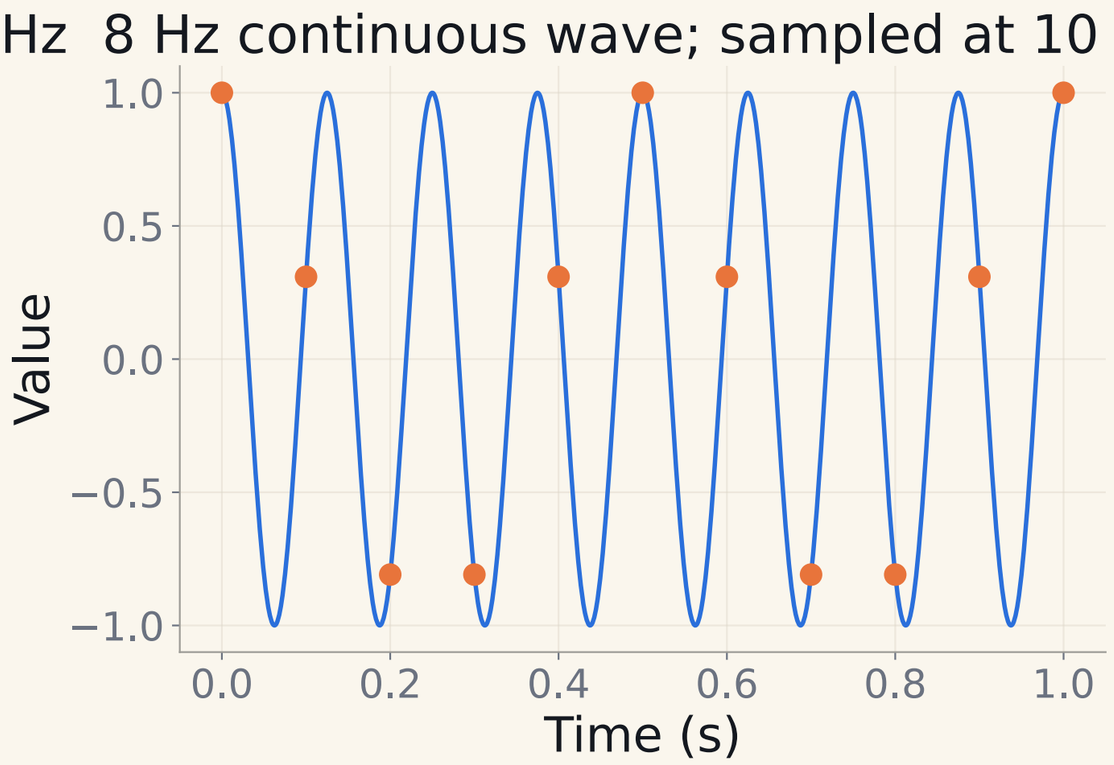


$\Delta f = 1/T = 0.25$ Hz. Zero-padding adds plotted bins, not a longer observation.

Can these samples tell two waves apart?

At **10 Hz sampling**, these sample values are identical. Sampling faster can distinguish them.

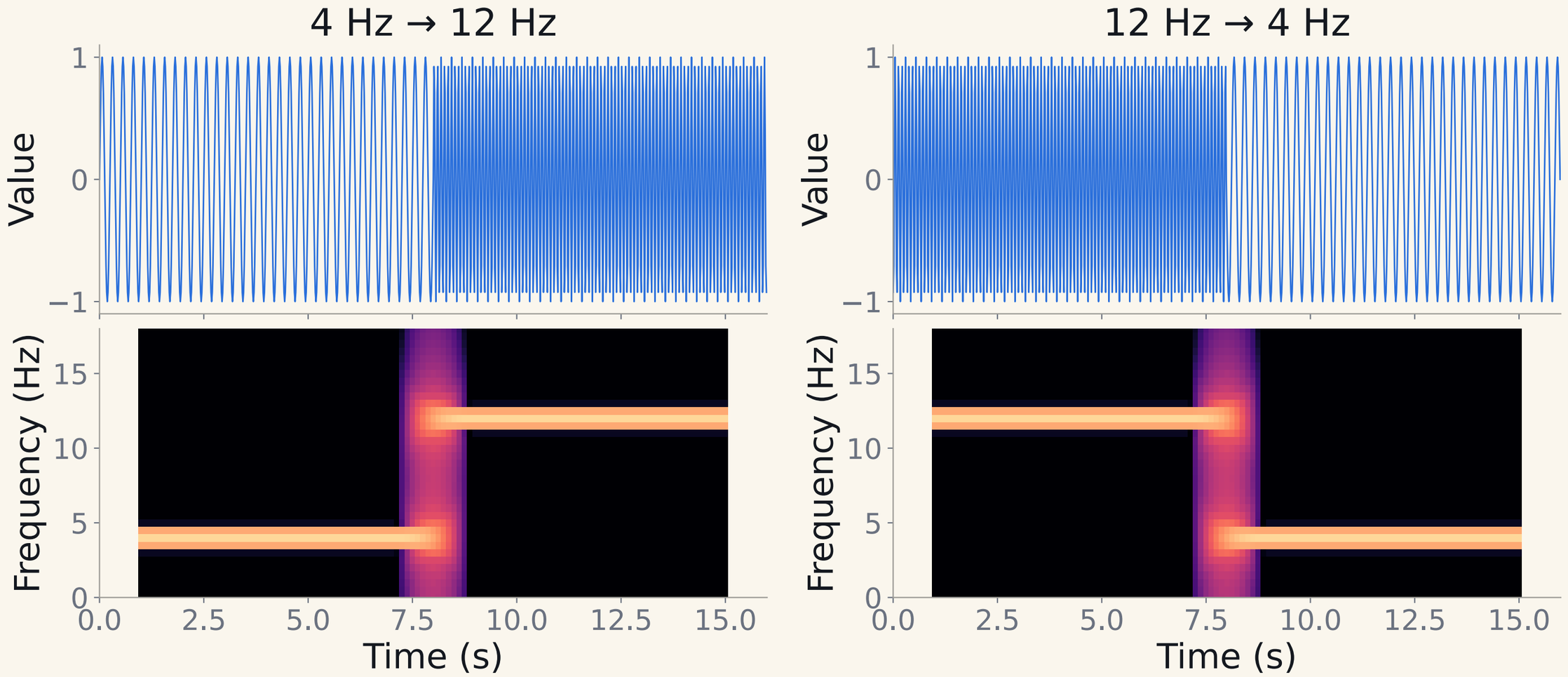




Constructed teaching signal · fixed seed 20261029

When did the frequency change?

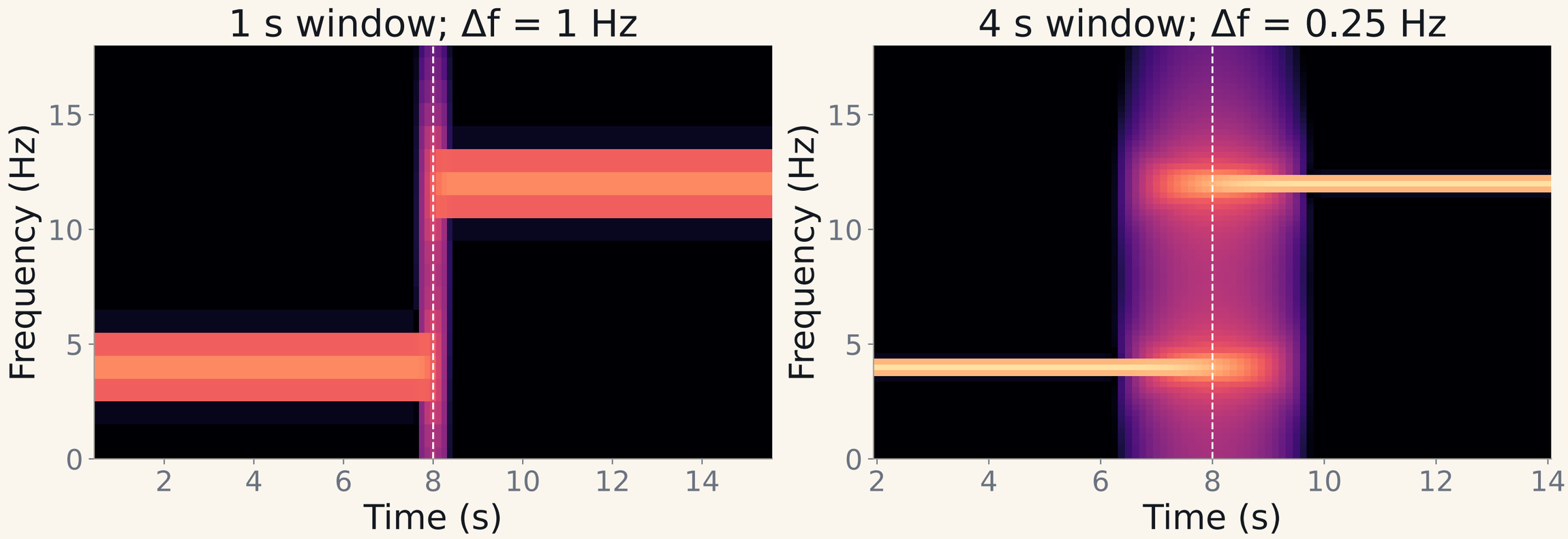
A time reversal preserves FFT magnitudes. A **spectrogram** exposes the order of events.



Constructed teaching signal · fixed seed 20261029

Choose a window. Choose a resolution.

Short windows localize transitions. Long windows separate close frequencies.



Constructed teaching signal · fixed seed 20261029

A wavelet stretches with the question

Scale changes duration and frequency; translation chooses where the probe looks.

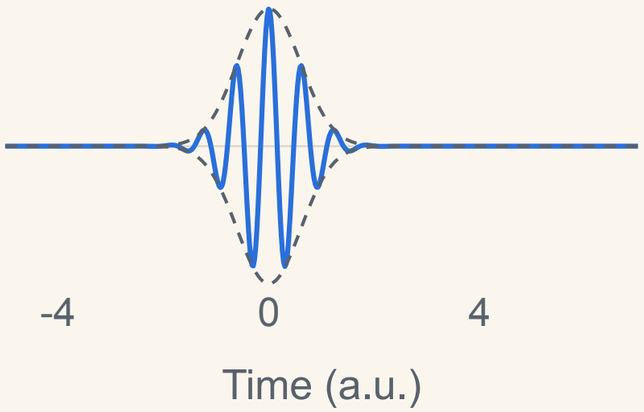
$$\psi_{a,b}(t) = a^{-1/2}\,\psi((t - b) / a),\ a > 0$$

a: scale (stretch) b: location (shift) $a^{-1/2}$: energy normalization

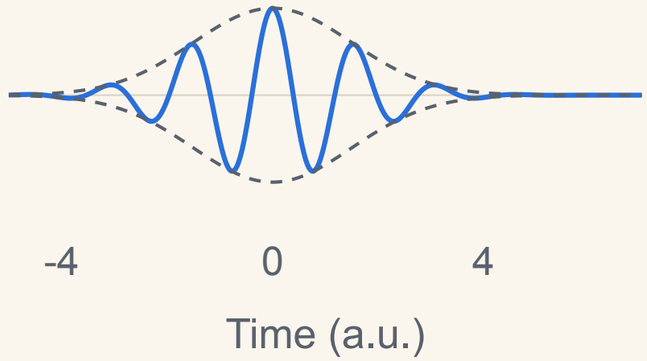
Example: $\psi(u) = [\cos(6u) - e^{-18}]\, e^{-u^2/2}$

real, zero-mean Morlet

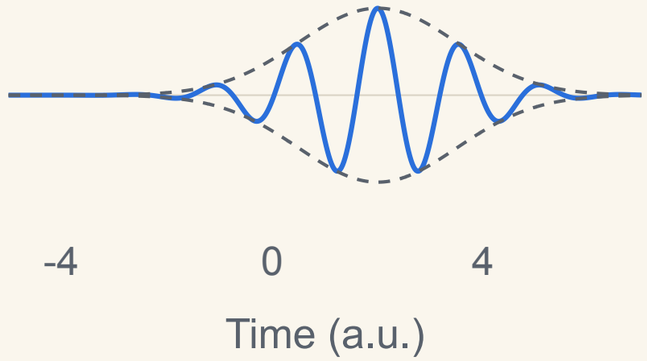
a = 0.6, b = 0



a = 1.5, b = 0



a = 1.5, b = 2

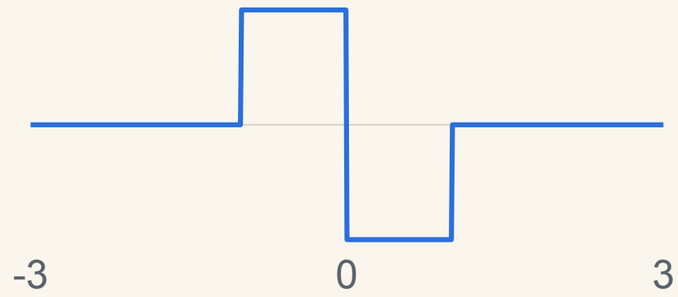


Choose or design the wavelet for the event

Different probes emphasize different shapes; their mathematical constraints still matter.

A wavelet is a choice of probe, not one universal curve.

Haar · scaled / shifted



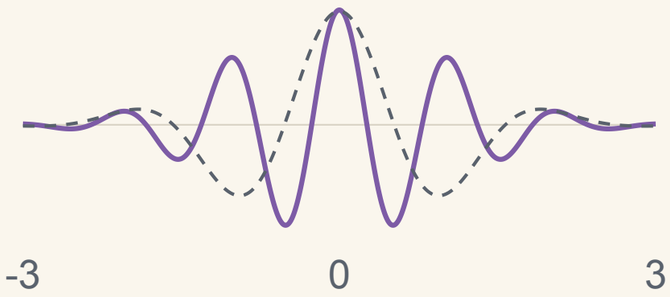
Abrupt local changes

Mexican hat



Peaks and curvature

Morlet



Solid: ω_0=6; dashed: ω_0=3

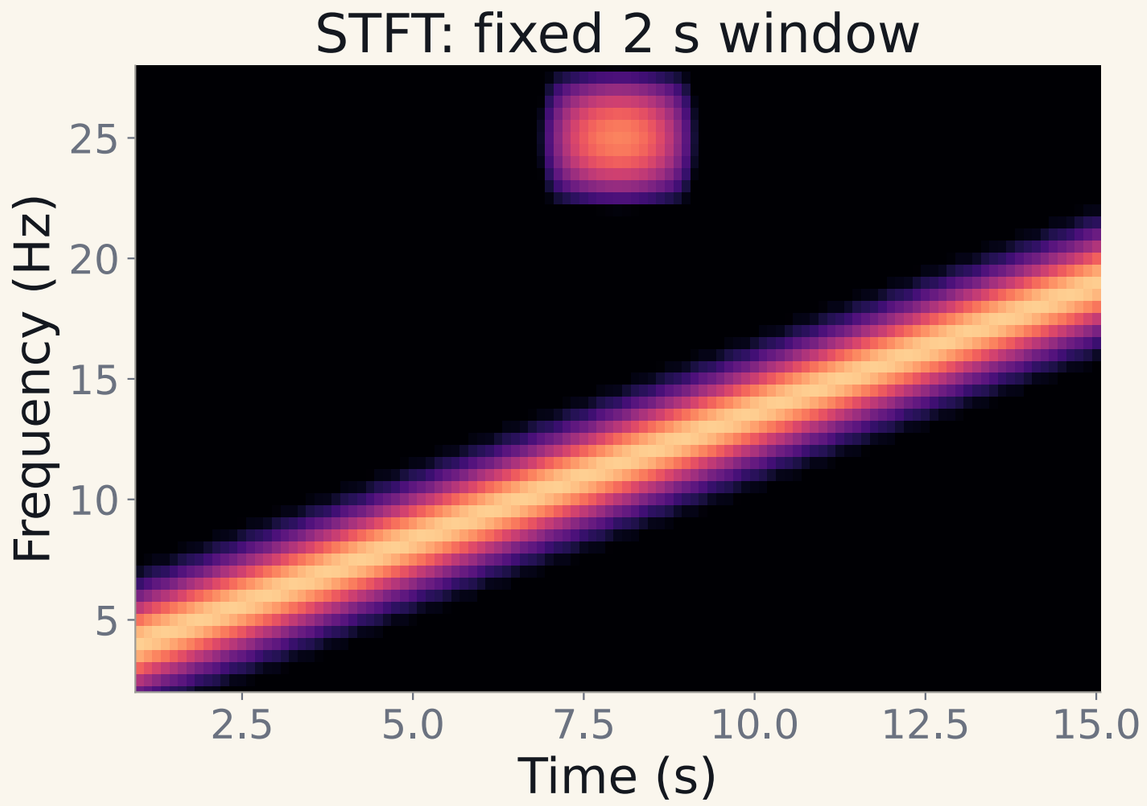
Design knobs: support · smoothness · symmetry · centre frequency / bandwidth

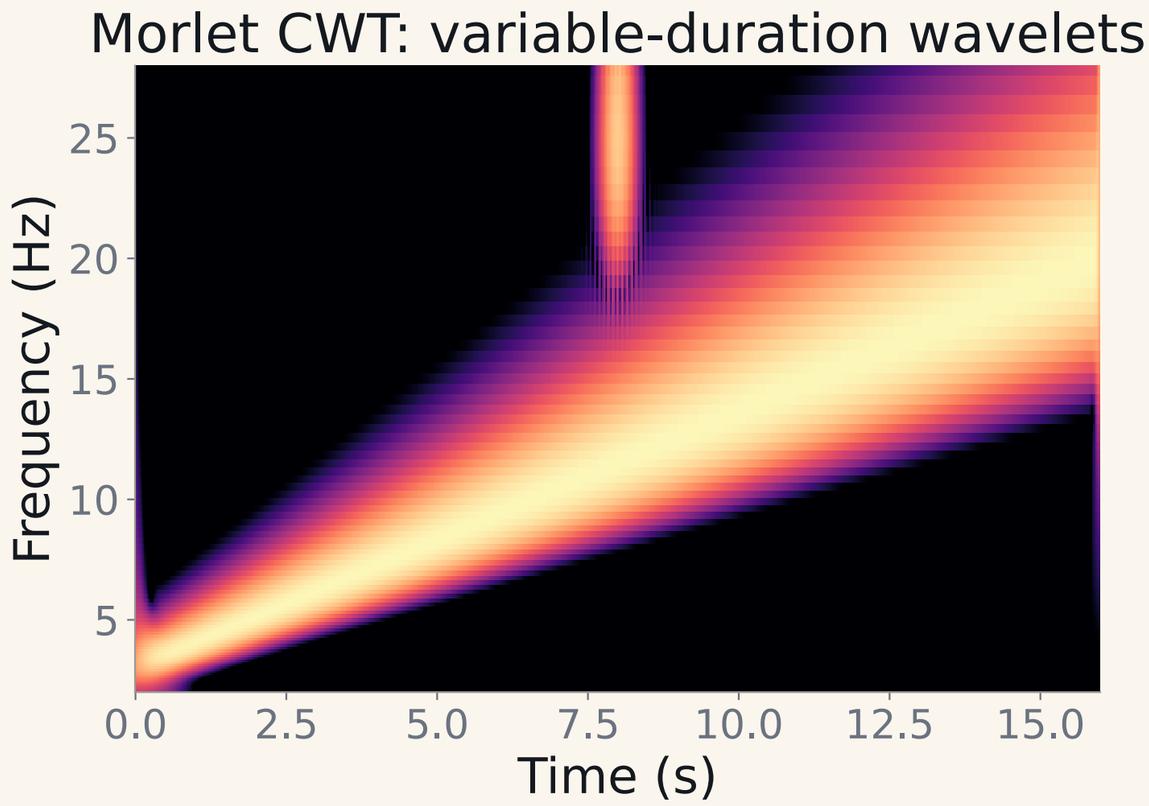
Validity: finite energy, zero mean, and a reconstruction condition (admissibility).

Custom wavelet filter banks are possible. A generic learned CNN filter is not automatically a wavelet.

STFT and CWT ask different local questions

STFT uses one duration. CWT uses multiple scales. **Choose the representation for the event.**



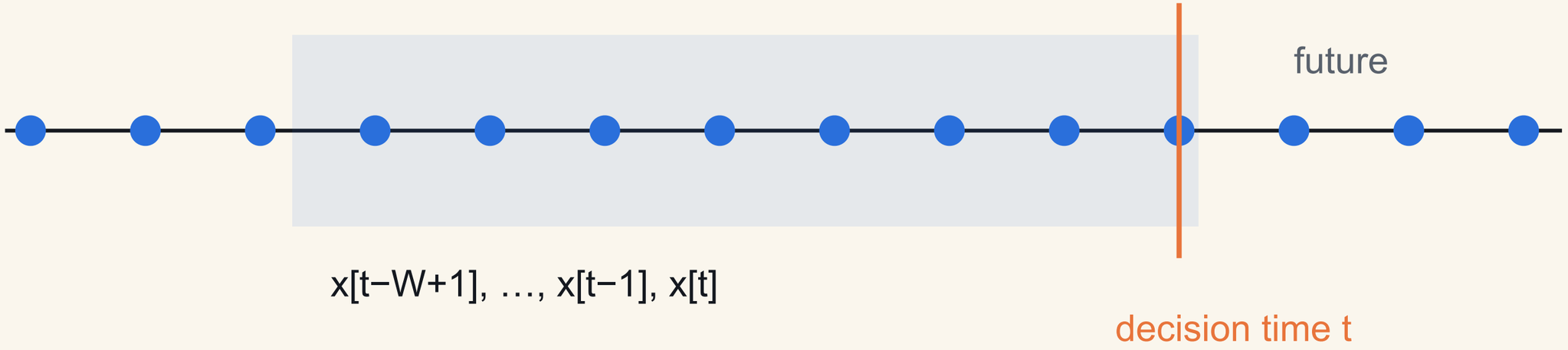


Constructed teaching signal · fixed seed 20261029

A window becomes one training example

For an anomaly decision after observing x[t], **every feature must use times ≤ t.**

At time t, use the latest W observations and predict a score for t



W time points × C channels → one feature vector

Lags preserve recent history

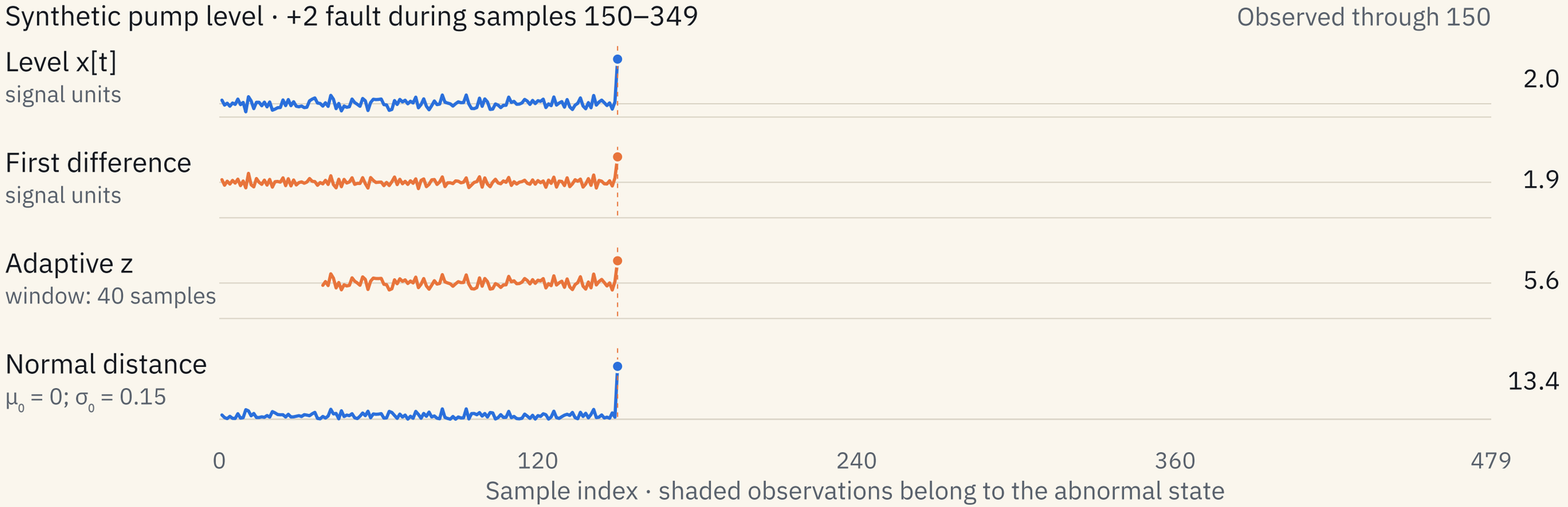
Decision time	x[t]	x[t−1]	x[t−2]	Optional target
2	4	2	1	x[3] = 3
3	3	4	2	x[4] = 5
4	5	3	4	not observed yet

> Lag features let a tabular model use order. The target determines the alignment.

WORKED EXAMPLE

Change is not the same as state

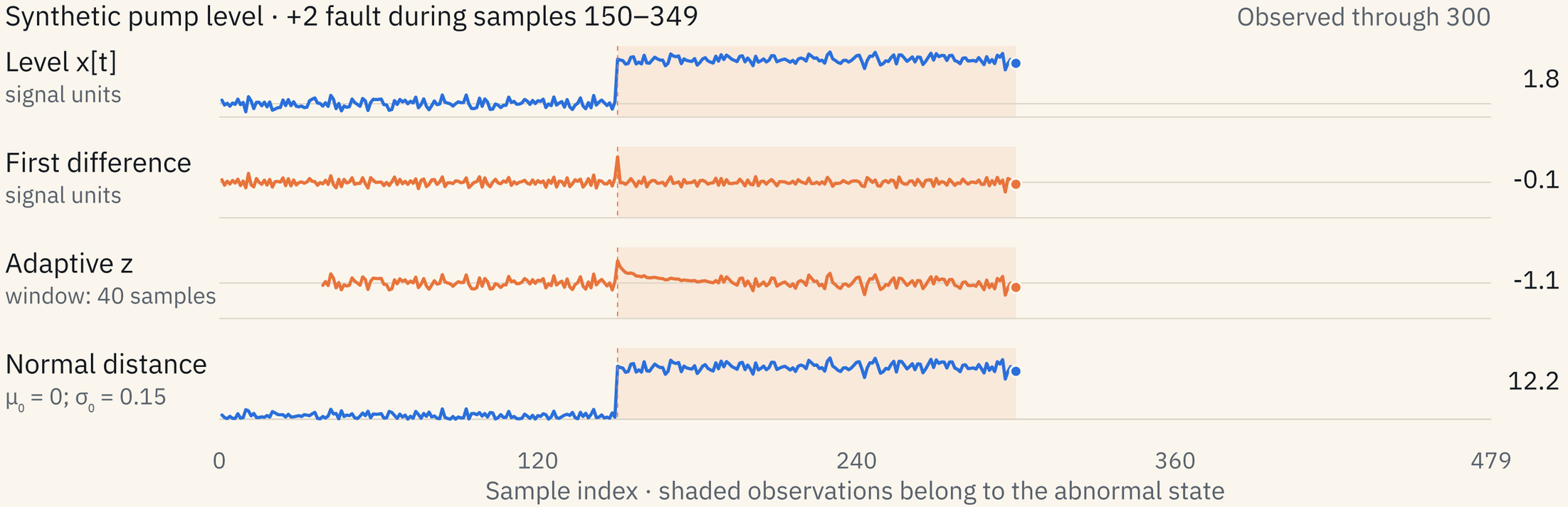
A +2 level shift arrives. Differences and adaptive z both react at the boundary.



Keyframe 1/3 · The fault begins — Sample 150 · faulty state · adaptive z 5.65 · fixed normal distance 13.42

Change is not the same as state

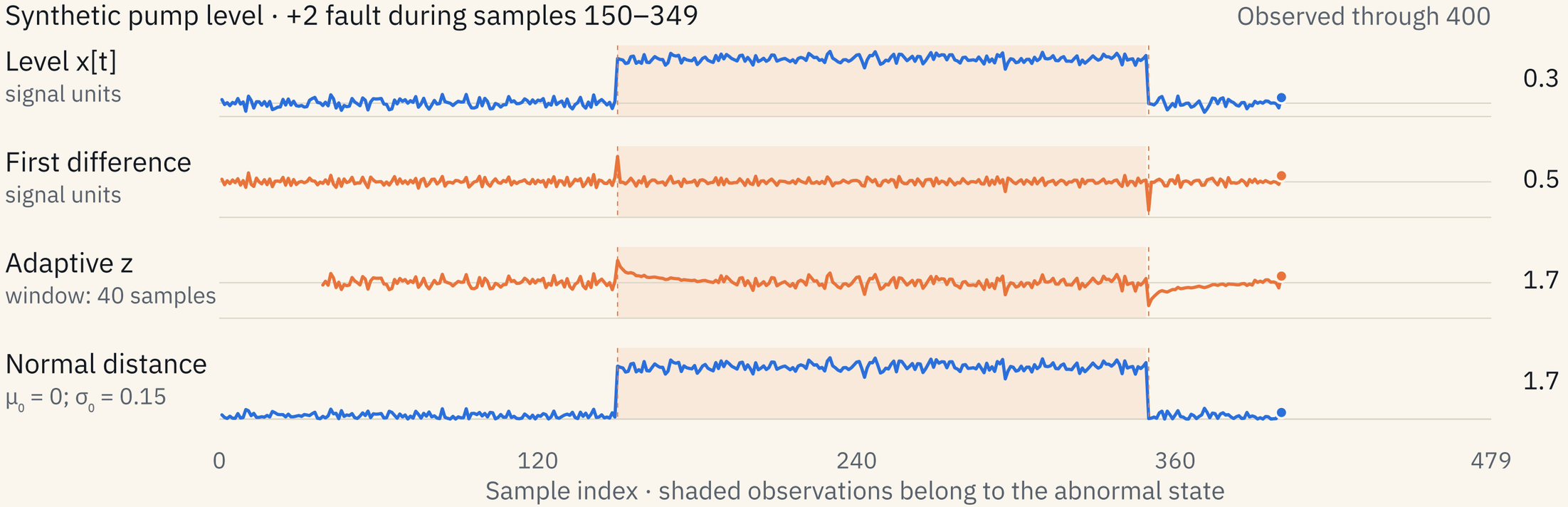
A fixed healthy reference retains the state; local change features can become ordinary.



Keyframe 2/3 · Inside the faulty state — Sample 300 · faulty state · adaptive z -1.11 · fixed normal distance 12.15

Change is not the same as state

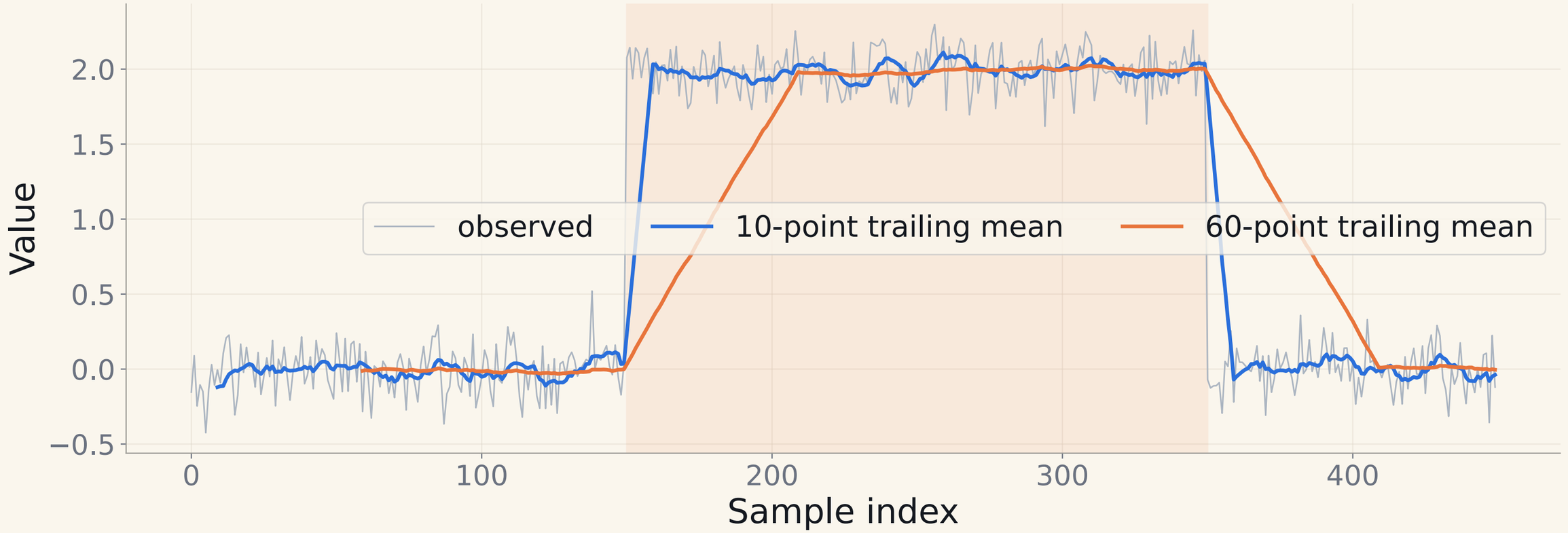
Explore time and window length. Match the representation to the entire-state label.



Keyframe 3/3 · After recovery — Sample 400 · healthy state · adaptive z 1.68 · fixed normal distance 1.68

Smoothing costs response time

Trailing mean / EWMA: more stable estimates, slower transitions, possible adaptation to a fault.



Constructed teaching signal · fixed seed 20261029

Summaries choose what survives

Family	Examples	Useful for	What it can hide
Level	Mean, median	Persistent offsets	Short transients
Spread	Std, MAD, IQR	Volatility, noisy states	Mean changes
Energy	RMS	Amplitude/energy summaries	Phase and waveform
Extremes	Min/max, quantiles	Spikes, tails, operating limits	Order inside the window
Dynamics	Difference, slope, ACF	Transitions and persistence	Absolute level

Spectral features summarize rhythm

> A spectrum of SKAB RMS describes modulation of RMS — not the missing raw vibration spectrum.

Fourier / STFT features

Peak frequency and band power

Spectral centroid or entropy

Pick the bands and window from the domain.

Wavelet features

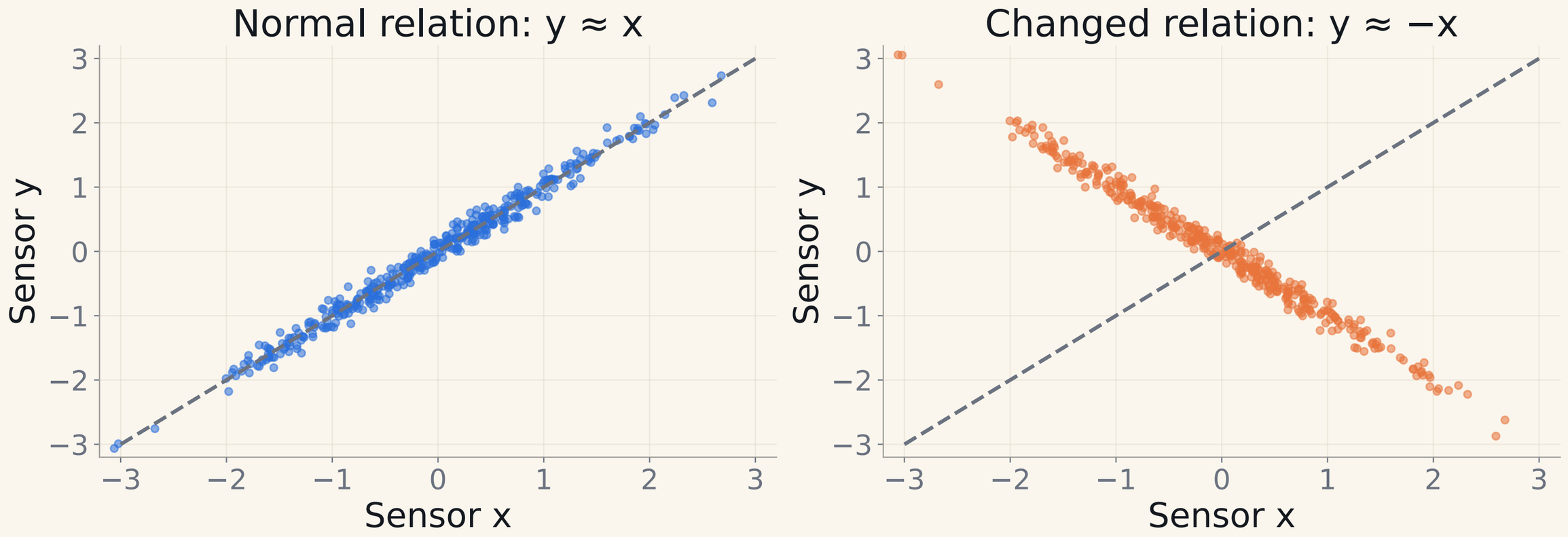
Energy across selected scales

Transient coefficient peaks

State the wavelet, scale mapping, and edge policy.

Relationships can fail first

Useful features: channel differences, rolling covariance, or **distance from a learned normal subspace**.



Constructed teaching signal · fixed seed 20261029

Fit transformations on the past

Build and evaluate a feature pipeline using only information available at deployment.

Check the clock
units · gaps · mask

→

Build causal features
raw + chosen history

→

Fit on train
scaler · detector

> Choose invariances deliberately: per-window centering can erase the very fault we want to detect.

Five tasks. Five different outputs.

Task	Output	Example
Classification	One label per window	Movement type from IMU / EEG epoch
Forecasting	Future values / distributions	Next-hour demand
Anomaly detection	Score, then abnormal-state decision	Pump monitoring
Segmentation	State label at each time	Walking → standing → sitting
Imputation	Values at missing positions	Lost telemetry blocks

Start with a task-shaped baseline

Task	First comparison	Next step
Classification	Statistics + logistic regression	ROCKET / EEGNet
Forecasting	Last value / seasonal naïve	Lag regression / foundation model
Anomaly detection	Robust normal-state distance	PCA / Isolation Forest / score ensemble
Segmentation	Thresholds / HMM	Temporal CNN; change-point methods
Imputation	Last value / interpolation	Kalman model / learned imputation

What was available at decision time?

> Causality depends on the decision time, not on the model’s name.

Forecast issued before x[t]

$$x[\leq t-1] \rightarrow \hat{x}[t]$$

Use the error only **after** x[t] arrives.

State decision after x[t]

$$x[\leq t] \rightarrow s[t]$$

A trailing statistic may include **the current observation**.

Overlapping windows are not independent

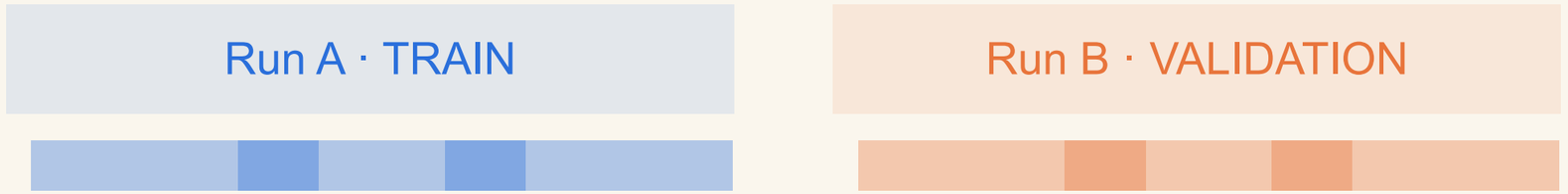
These length-eight windows share seven observations. A new-run evaluation holds out whole recordings.

Randomly splitting windows can send the same observations to both sides

Window A → train	x[0] +0.2	x[1] +0.5	x[2] +0.1	x[3] -0.3	x[4] -0.1	x[5] +0.4	x[6] +0.8	x[7] +0.5
Window B → validation	x[1] +0.5	x[2] +0.1	x[3] -0.3	x[4] -0.1	x[5] +0.4	x[6] +0.8	x[7] +0.5	x[8] +0.0

7 of 8 observations are shared; the validation window is nearly a duplicate.

Split whole recordings

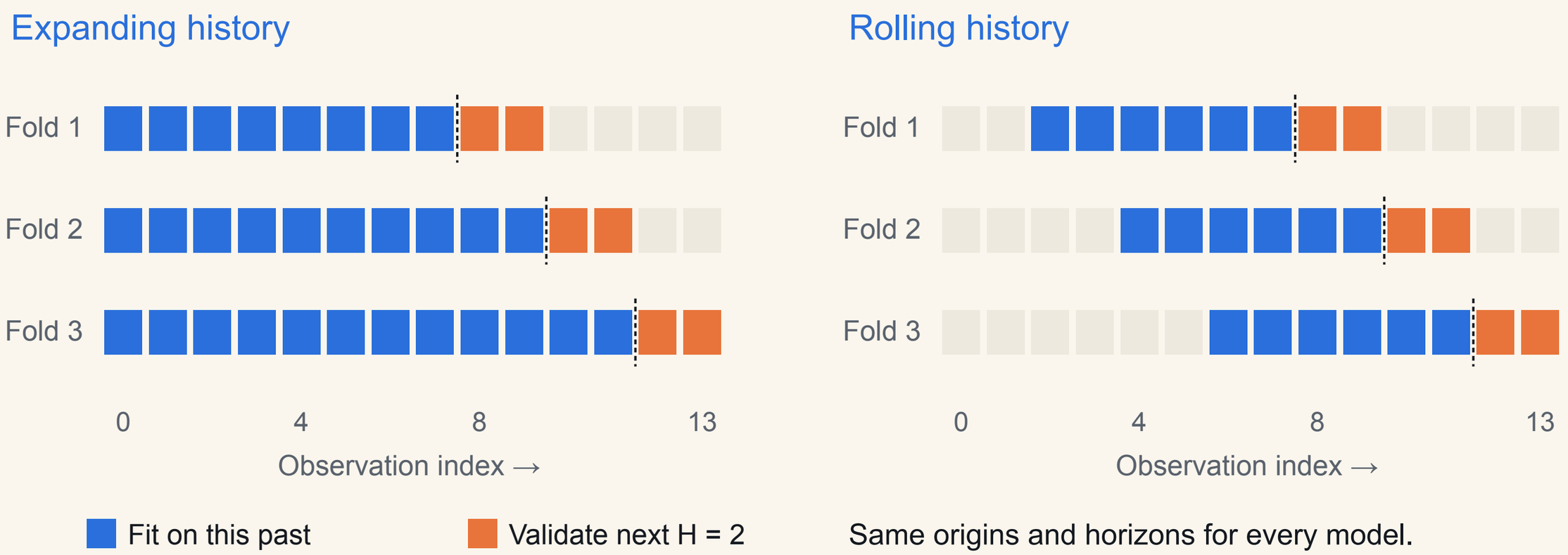


All windows from one run stay together. A new-run claim needs a new run.

Constructed evaluation examples · the SKAB protocol is stated explicitly

Roll the forecast origin forward

Expanding windows retain all earlier data; rolling windows keep a fixed recent history.



Refit preprocessing inside each fold. Report horizon-wise errors and variation across origins.

Constructed evaluation examples · the SKAB protocol is stated explicitly

Hold out the identity you expect to meet

Time splits test the future; group splits test new runs, people or devices. Some deployments require both.

Cross-validation over recording identity

A, B and C rotate through validation; D stays outside model selection.

	Pump run A	Pump run B	Pump run C	Final run D
Fold 1	train	train	validation	untouched
Fold 2	train	validation	train	untouched
Fold 3	validation	train	train	untouched

New person or device? Group by that identity. Future data? Also enforce time order.

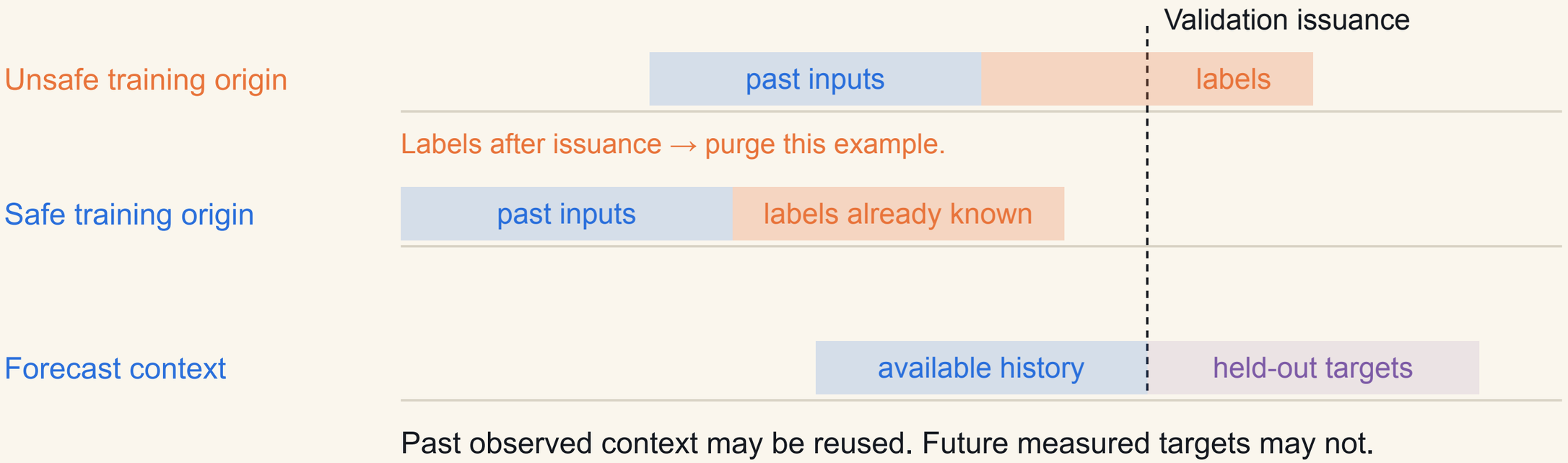
SKAB lab: runs are held out; calendar times overlap. This is a new-run experiment.

Constructed evaluation examples · the SKAB protocol is stated explicitly

Guard the information boundary

Choose the gap from what each training example needs to know; preserve legitimately available forecasting history.

Purge training examples whose required labels are not yet available



Past observed context may be reused. Future measured targets may not.

Gap / purge length follows target support, feature availability and the deployment question.
A gap is not automatically equal to the lookback window. Keep the final test out of all tuning.

Constructed evaluation examples · the SKAB protocol is stated explicitly

Choose the split for a new person and a later date

Deployment means a never-seen participant recorded after a fixed date. Which split enforces both identity and time?

1. Randomly split all windows; stratify by activity label.
2. Hold out people with GroupKFold; ignore dates.
3. Train other participants before the cutoff; validate held-out participants after it.

State separately which rule protects identity and which rule protects time.

Pause here. The next keyframe contains the explanation.

Choose the split for a new person and a later date

Deployment means a never-seen participant recorded after a fixed date. Which split enforces both identity and time?

1. Randomly split all windows; stratify by activity label.
2. Hold out people with GroupKFold; ignore dates.
3. **Train other participants before the cutoff; validate held-out participants after it. ✓**

State separately which rule protects identity and which rule protects time.

> **Reasoning**
>
> Keep every session of held-out participants out of training. Use only other participants' data from before the fixed cutoff; validate on held-out people later. Fit preprocessing inside training folds and reserve a separate final holdout.

Choose a feature for a new fault

Two pump sensors keep the same individual means and spreads, but their usual relationship breaks. Which feature targets this fault?

1. Threshold each sensor’s rolling mean and standard deviation.
2. Use only the first difference of each sensor separately.
3. Use a residual from a relationship learned on healthy sensor pairs.

Name the property that changed, then explain why an individual-channel threshold can miss it.

Pause here. The next keyframe contains the explanation.

Choose a feature for a new fault

Two pump sensors keep the same individual means and spreads, but their usual relationship breaks. Which feature targets this fault?

1. Threshold each sensor’s rolling mean and standard deviation.
2. Use only the first difference of each sensor separately.
3. **Use a residual from a relationship learned on healthy sensor pairs. ✓**

Name the property that changed, then explain why an individual-channel threshold can miss it.

> **Reasoning**
>
> The joint relationship changed. Fit a healthy relation using training data, then score deviations. A channel difference, regression residual or PCA residual can fit the task under its assumptions. Context still distinguishes faults from expected changes.

Sources, credits, and reproducible controls

Read and explore

Hyndman & Athanasopoulos: Forecasting: Principles and Practice

3Blue1Brown: Fourier transform intuition

PyWavelets: continuous wavelets, scale and frequency

Data and illustration

Katser & Kozitsin: SKAB

Serega artwork reused from Albert Nasybullin's Deep Learning for Search course.

Scientific control plots are computed; synthetic plots are labeled explicitly.

Clickable sources and teaching notes: press N. Figure provenance is packaged with this deck.