

How historian compression distorts machine learning

The series in a historian is not the series the sensor saw. Compression settings change what features and models learn.

In short

- A historian series is not the series the sensor saw, because a deadband or other lossy compression can change noise and peaks.
- A deadband can erase noise and produce an apparently flat stretch, which matters when the noise itself carries information.
- Forward-fill and linear interpolation are different assumptions, so the resampling method and its rationale should be recorded.

01 / What lossy compression is

Many historian products do not store a value unless it leaves a band around the previously stored value (a deadband). Some drop points that a straight line can represent. This reduces storage and makes trend screens look good, but it leaves a different series from the raw one.

02 / Which features are affected

The mean and last value are usually preserved. Standard deviation, derivatives, peaks and noise level change: a deadband erases noise and produces an apparently flat stretch. In predictive maintenance, where the vibration or current noise itself carries information, a compressed series may have removed it.

03 / Resampling to regular intervals

Models usually expect an evenly spaced series. Forward-fill and linear interpolation are different assumptions when turning compressed records into a regular grid. Record the method and its rationale, and keep a separate flag for “actually measured” versus “derived”.

04 / A practical checklist

Find out the compression setting for the tag you use. If possible, collect raw and compressed data side by side for a week and compare feature distributions. If the model will be trained on compressed data only, state that as a limitation. This guide is a suggested workflow; product-specific algorithm details should come from the vendor documentation.

Try it

Thin a series with a deadband and compute how the standard deviation and the number of peaks change.

References and online guide

[Open the online guide](#)

SciPy: signal processing

Proposed pilot steps do not claim field accuracy. Synthetic examples are not measurements from a production site.