UnQuantize: Overcoming Signal Quantization Effects in IoT Time-Series Databases

Matthew Torin Gerdes
Oracle Physical Sciences Research Center
Oracle Corporation
San Diego, CA
matthew.gerdes@oracle.com

Kenny Gross
Oracle Physical Sciences Research Center
Oracle Corporation
San Diego, CA
kenny.gross@oracle.com

Guang Chao Wang
Oracle Physical Sciences Research Center
Oracle Corporation
San Diego, CA
guang.wang@oracle.com

Abstract*—Low-resolution quantized time-series signals present a challenge to big-data Machine Learning (ML) prognostics in IoT industrial and transportation applications. The challenge for detecting anomalies in monitored sensor signals is compounded by the fact that many industries today use 8-bit sample-and-hold analog-to-digital (A/D) converters for almost all physical transducers throughout the system. This results in the signal values being severely quantized, which adversely affects the predictive power of prognostic algorithms and can elevate empirical false-alarm and missed-alarm probabilities. Quantized signals are dense and indecipherable to the human eye and ML algorithms are challenged to detect the onset of degradation in monitored assets due to the loss of information in the digitization process. This paper presents an autonomous ML framework that detects and classifies quantized signals before instantiates two separate techniques (depending on the levels of quantization) to efficiently unquantize digitized signals, returning high-resolution signals possessing the same accuracy as signals sampled with higher bit A/D chips. This new "UnQuantize" framework works in line with streaming sensor signals, upstream from the core ML anomaly detection algorithm, yielding substantially higher anomaly-detection sensitivity, with much lower false-alarm and missed-alarm probabilities (FAPs/MAPs).

Keywords*—Internet of Things, Quantized Signals, ML Prognostics, A/D Converter.

I. Introduction

Quantized signals are prevalent in many Internet-of-Things (IoT) industries, including Utilities, Oil&Gas, Transportation, Manufacturing, and even in business-critical enterprise computing data centers. Quantized signals originate from inexpensive analog-to-digital (A/D) chips with low-bit resolution, which are used to convert the analog transducer signals into digitized time series. Machine learning (ML) algorithms perform poorly with quantized signals, prohibiting its wide applications in many dense-sensor IoT industries. As the price of high-bit resolution analog to digital (A/D) converters has decreased, their adoption throughout sensor dense industries would seem to be inevitable; however, transition to 12-bit and 16-bit A/D chips is slow in new assets. Furthermore, even as high resolution A/D chips become more widespread in future assets, older legacy assets will still need to be serviced and data from this hardware will need to be analyzed for the remainder of their lifetimes.

For large scale IoT prognostic applications, it is not possible to have humans look at all the signals to decide between quantized vs unquantized, and to count how many levels of quantization there are for the quantized signals. As examples: a modern oil refinery these days has 1M sensors recording time series signals 24x7x365. In commercial aviation, for example one Airbus airplane has 75,000 sensors. Similarly, due to the proliferation of sensors inside enterprise IT servers and storage platforms in data center, a medium size enterprise or cloud data center today has 1M sensors, very many of which are quantized, and at various levels of quantization. What is needed is a technique that can examine a large universe of sensor time series signals, automatically identify signals that are quantized, determine the number of levels of quantization, and "unquantize" the signals identified as quantized. It is also essential to unquantize the signals in a manner that is optimized per the number of levels of quantization, including for the "worst case" lowest resolution that yields only 2 quantization levels (and we've seen 2-level quantization in many collections of real telemetry measurements). This paper explores a novel analytical framework that automates the discovery of quantized signals in "big data" databases that may contain thousands of sensors signals, identifies the exact quantization levels for all of those signals, and unquantizes the signals with a novel algorithm that is optimized to different quantization levels for the various individual signals.

The remainder of this paper is organized as follows. Section II introduces the implementation of our UnQuantize framework step by step. Section III.A and B illustrate the impact of quantized signals and demonstrate the performance of our unquantize methodology, and Section III.C presents how our solutions address the challenges in ML prognostic caused by quantized signals. Section IV provides the conclusions.

II. Methodology

This paper focuses on two aspects of unquatization. The first aspect is the accuracy of unquantized signals, when compared to the known high-resolution signal, relative to their quantized counterparts. The second aspect is the prognostic performance gains that unquantization imparts to the performance of ML pattern recognition and automated anomaly discovery. When quantized signals are included in training data sets there is a much higher likelihood for false-alarms and missed-alarms. In most industries where ML prognostic surveillance is valuable, false alarms are very costly, resulting in taking revenue-generating assets out of service unnecessarily. Moreover, missed alarms can be catastrophic. We demonstrate in this paper that the introduction of an autonomous framework for unquantization of signals requires no hardware modifications and, when combined with Oracle's advanced ML pattern recognition, is helping to substantially increase component reliability margins and system availability goals while reducing (through improved root cause analysis) costly sources of "no trouble found" events that have become a significant warranty-cost issue for asset manufactures.

A. Overview of UnQuantize Methodology

The telemetry time series signals are unquantized on a signal by signal basis. Figures 1 and 2 exemplify quantized sensor readings and demonstrate the drastic difference between the recorded signal and the genuine signal characteristics. The first step of unquantizing is to identify the number of quantization levels in each signal. If the computed number of quantization levels is smaller than 20 or smaller than 5% of number of observations, the signal is deemed ‘quantized’. Signals possessing a number of quantization levels greater than four are unquantized by Fourier decomposition and reconstruction. When there are between two and four quantization levels (e.g. -1 and 1, or -2, -1, 1 and 2), the signal is unquantized by computing the bin-switching frequency between higher and lower quantization levels in a sliding window. The scaled bin frequency to match the quantization levels is served as the unquantized signal. Figure 3 shows the flow chart of the unquantizing process.

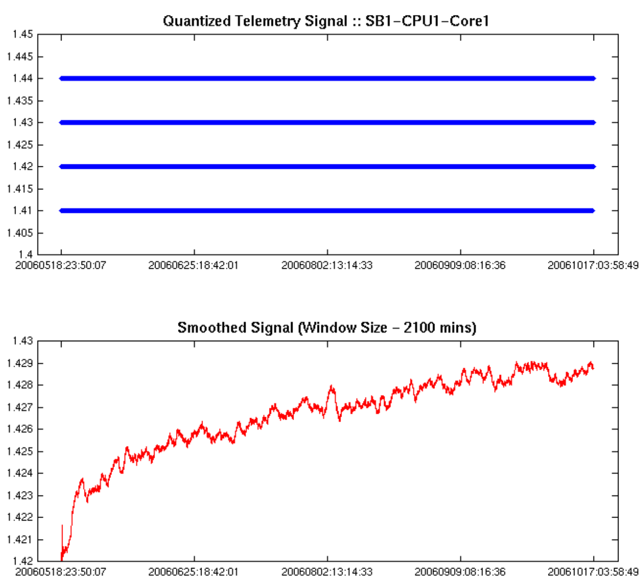


Figure 1 Example of quantized telemetry signal reported from 8-bit A/D chips used in the servers and the unquantized telemetry signal produced by our technique.

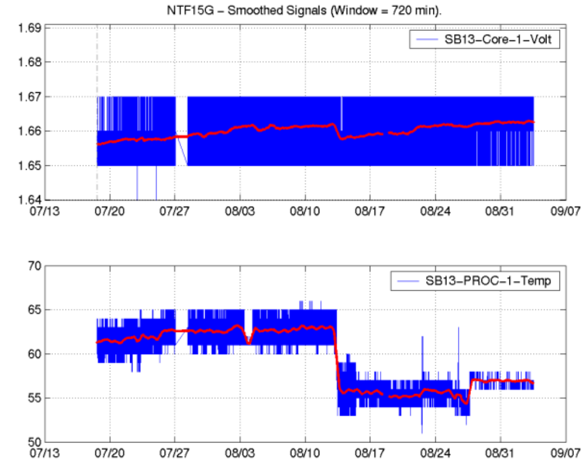


Figure 2: The raw voltage (upper) and temperature (lower) signals reported from 8-bit A/D chips used in many electronic systems. The red signal shows the actual values of the monitored variables.

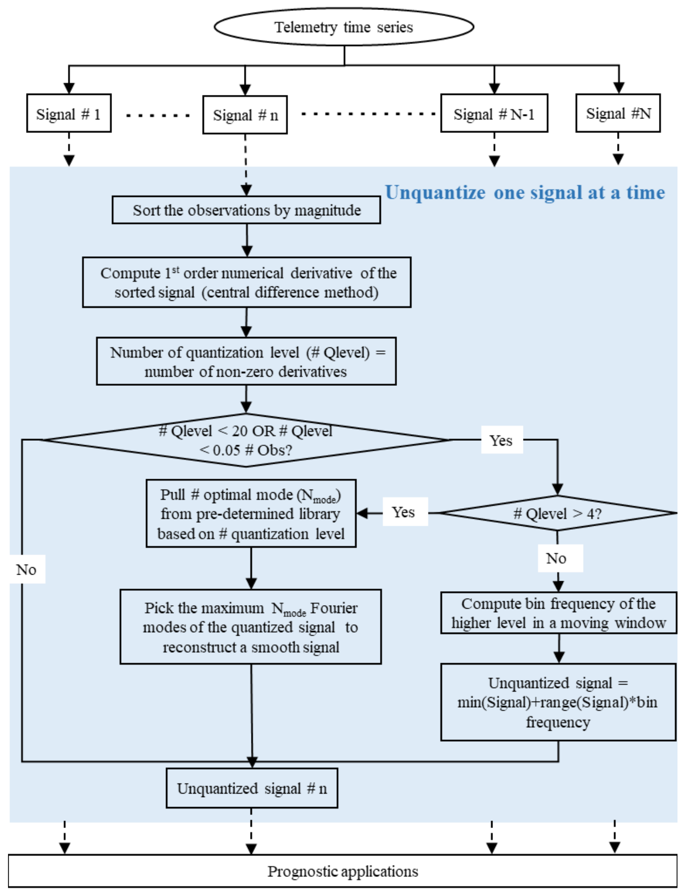


Figure 3: Flowchart of framework for autonomous Unquantizations analysis that can be used upstream of any ML algorithms.

B. Testing Signals

The time-series signals used in the case study have been synthesized with a high-fidelity signal synthesis algorithm from real time series signatures across a variety of IoT industrial use cases. These signals are synthesized, not simulated, which match real IoT sensor signals in all statistical characteristics important to ML prognostics, including serial correlation content, cross correlation between/among signals, and stochastic content (variance, skewness, kurtosis), as real IoT sensor signals. For the large scale database of synthesized signals used in this investigation, OracleLabs' Telemetry Parameter Synthesis System (TPSS) has been employed [1-3].

Once the signals are synthesized they are quantized by the mid-rise and mid-tread uniform quantizing method (Eqn. 1). To emulate the low-resolution A/D chips more realistically Δ is calculated using the minimal and maximum values from the original signal that is noiseless. The quantized signals are then passed through the unQuantize framework.

$$f = \begin{cases} \Delta \cdot \text{floor}\left(\frac{x}{\Delta} + \frac{1}{2}\right), & \text{even Quantization Level} \\ \Delta \cdot \left(\text{floor}\left(\frac{x}{\Delta}\right) + \frac{1}{2}\right), & \text{odd Quantization Level} \end{cases} \quad (1)$$

where $\Delta = \frac{\max(x) - \min(x)}{QL - 1}$.

C. Determine Quantization Level (QL)

To determine the quantization level the signal is sorted in ascending order and then, and a numerical central difference scheme is applied to find the 1st order derivative of the sorted signals. The sum total of the nonzero values in the derivative of the sorted signal determines the QL of the signal. If QL is 20 or less or the QL is less than 5% of the number observations the signal determined to be quantized. Figure 4 illustrates the process of determining the quantization lever of a signal.

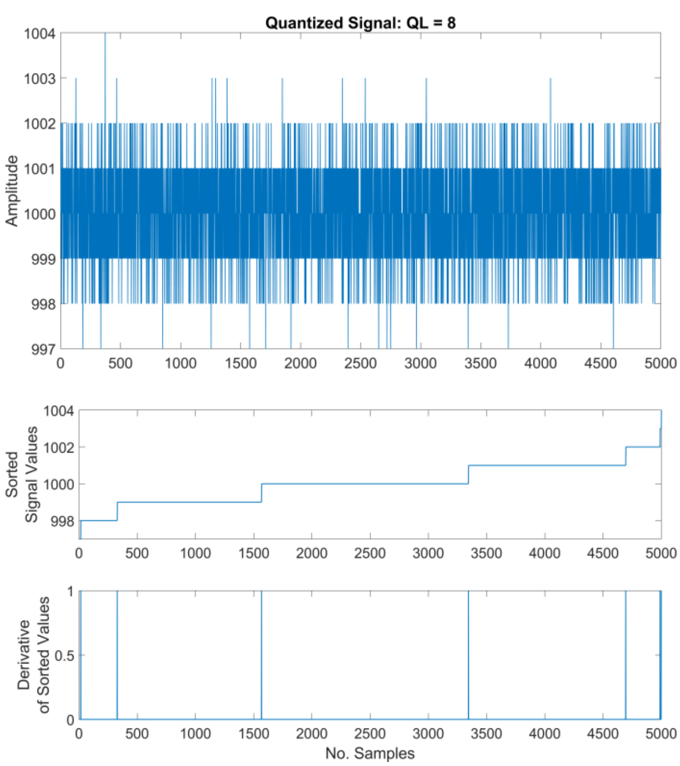


Figure 4: Illustration of finding the number of quantization levels in a quantized signal. The upper plot showcases a typical quantized signal. The middle plot presents the sorted observations in ascending order and the lower plot shows the 1st order derivative of the sorted values. The number of quantization levels is found to be 8 for this example.

D. Fourier Decomposition (for QL>4)

If QL is greater than four, Fourier Decomposition is used to unquantize the signals. The quantized signal is converted into the frequency domain using a Fourier transform (FFT). In the frequency domain the most prominent harmonic modes are extracted to generate a composite frequency signal. The number of Fourier modes (N largest modes) used is precomputed and stored in a mode library, where the number of modes is a function of the number of quantization levels. Then the new composite signal is converted back to the time domain through and inverse Fourier transform (iFFT) [6].

E. Bin Switching Frequency (for QL≤4)

If QL is two, the signal is processed with the bin switching frequency algorithm. The algorithm passes a sliding window over the quantized signal, determines the frequency of the highest level in the window, and normalizes that frequency value by the length of the window. The normalized value becomes one data point in the unquantized signal. Because the output of the bin switching frequency algorithm is normalized, the range is between 0 and 1. This requires a rescaling of the signal.

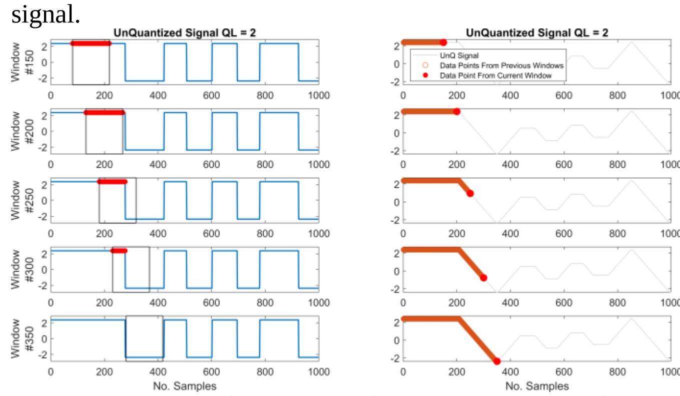


Figure 5: Illustration of bin-switching frequency method. The left column contains a portion of the quantized signal (in blue) as the sliding window (in black) moves forwards in time. The data points in red determine the corresponding red point in the right column. The right column illustrates the construction of the unquantized signal (in grey). The point in red indicates the unquantized sample determined by the current window (e.g. window 300 is equivalent to the 300th unquantized sample). The points in orange are the points determined by the previous windows. The windows are increasing by increments of 50 from top to bottom.

If QL is three or four, additional upstream and post processing are required to use the bin switching frequency algorithm. The upstream addition consists of splitting the quantized signal into multiple quantized signals where QL is two. For example if the signal is simple, and has levels that are equivalent to -1, 0, and 1, QL would be three and can be separated in two signals where the signals would have QL equivalent to two: -1 and 0, 0 and 1. The new split signals are now processed with the bin switching in the same manner as the when QL is two. The post processing addition sums the scaled signals and then subtracts off the mean of quantized signal.

III. EVALUATION AND DISCUSSIONS

A. Negative Impact of Quantized Signals

When low-resolution A/D converters record physical phenomena, the resulting signal is an abstraction. Many meaningful physical characteristics, such as periodicity, noise ratio, number of modes are lost due to this abstraction, or quantization. Figure 6 illustrates the loss of information that can occur from quantization. When comparing the sinusoidal signal to its quantized version it is very apparent that many of the identifying patterns that characterize the time-series are indecipherable, such as the number of modes and the underlying frequencies.

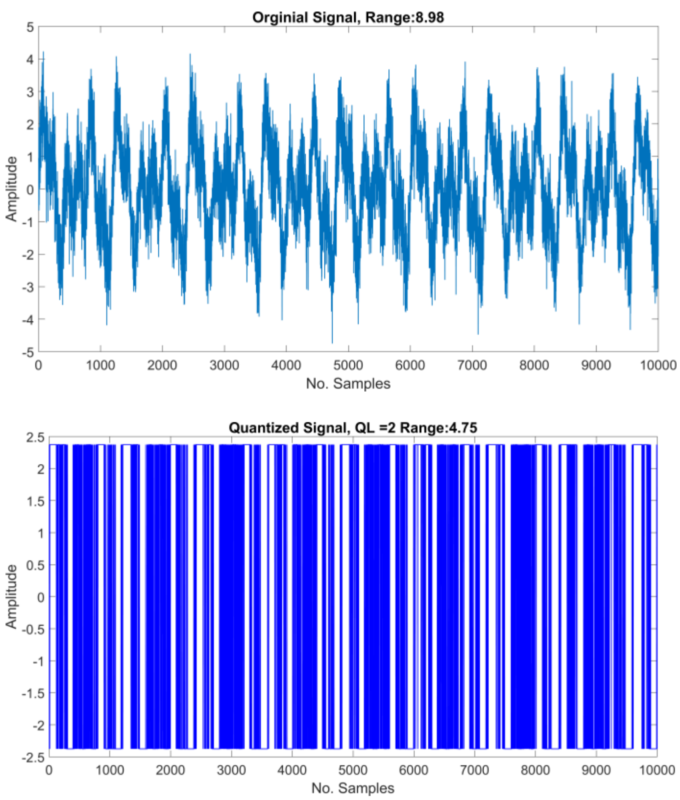


Figure 6: Comparison between the intact telemetry signal (top) and the quantized version with QL = 2 (bottom). While the original signal was found to be a composition of three sin waves, the quantized signal only exhibits a periodic movement oscillating between two points.

Unfortunately, the signal dynamics that quantization obscures are of importance for any ML algorithms to build accurate and meaningful models.

B. Comparions of Signal Reproductions between Quantized and Unquantized Signals

Oracle innovators have developed an algorithm that analytically unquantizes signals in effect taking low-resolution input signals and turning them into high-accuracy output signals. These unquantized signals extract the dynamics of the ground truth much more closely and accurately than their quantized counterparts making them much more conducive to prognostic ML modeling. Figure 7 quantitatively assesses and compares the deviations of the quantized and unquantized signal from the original signal. The continuous signal in Figure 7(a) is quantized into three different quantization levels: 2, 3, and 4. In Figure 7(b) and (c), the quantized and unquantized signals overlaid on top of the corresponding original signals (left column) yield the deviations that are evaluated by RMSE metric (right column) respectively. As evidenced by the smaller RMSE values and more consistent residuals, the unquantized signals are much closer reproductions of the original signals.

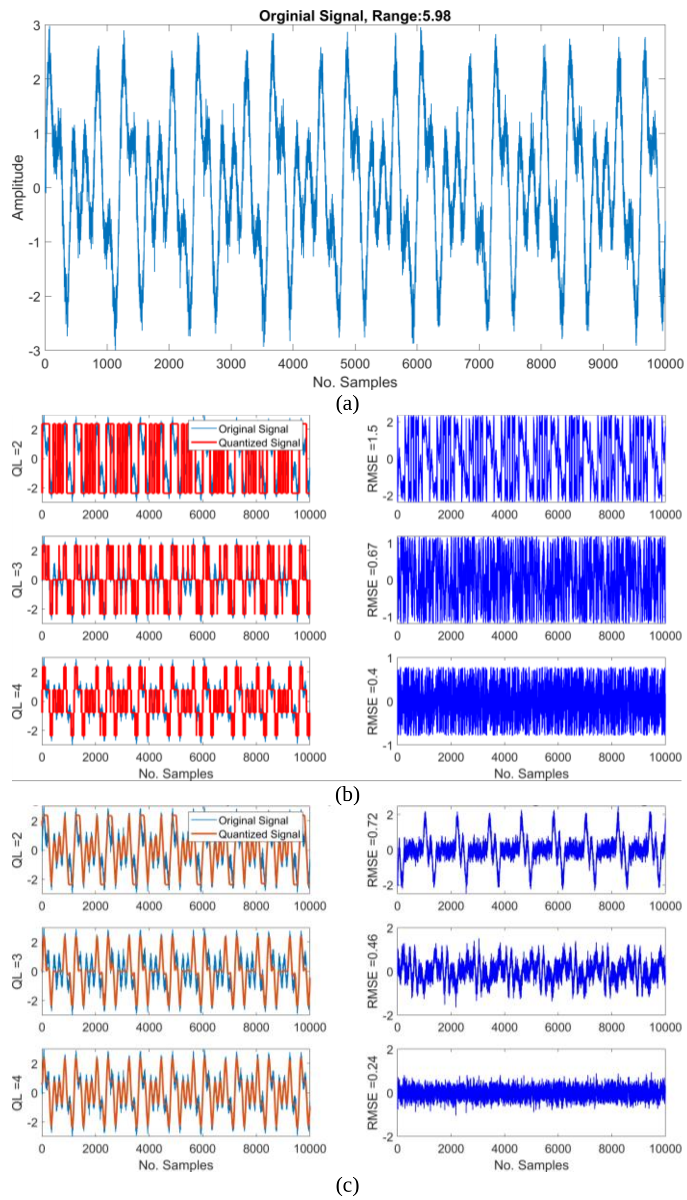


Figure 7: (a) The original testing signal is quantized into three different levels, resembling the typical outcome of the low-bit A/D chips. (b) The quantized signals (red) in three different levels are compared with the original signal (blue) in the left column, and their respective deviations characterized by RMSE are presented on the right. (c) Same as (b) except the quantized signals have been unquantized according to their quantization levels.

C. *Performance Gains with Unquantized Signals in ML Prognostics*

The case study presented herein demonstrates the UnQuantize framework upstream of the Oracle's preferred ML prognostic solution, which is the Multivariate State Estimation Technique (MSET, refer to [4-7] for more details). MSET provides high sensitivity for proactive warnings of incipient anomalies, and ultra-low false-alarm and missed-alarm probabilities. The increased prognostic accuracy afforded by the unquantization of signals allows for better anomaly detection performance, which is evaluated and demonstrated in this section.

Figures 8-10 demonstrate the unquantization of signal yields better false alarm probability (FAP) in an anomaly detection example. A 50 sec long continuous signal was equally divided into two parts: the first part (Figure 8) was used for building up a MSET model, which was then used to examine the second part as the surveillance data that had been quantized (QL = 4) deliberately (Figure 9a). While zero false alarm is expected since the surveillance data has the same characteristic as in the training data, the fact that quantization causes loss of meaningful physical characteristics leads to significant deviations between the two data, and subsequently yields false alarms (Figure 9b). However, if we had the quantized signal undergo the unquantization process, the prior false alarms (red dots in Figure 9b) were eliminated, as demonstrated in Figure 10.

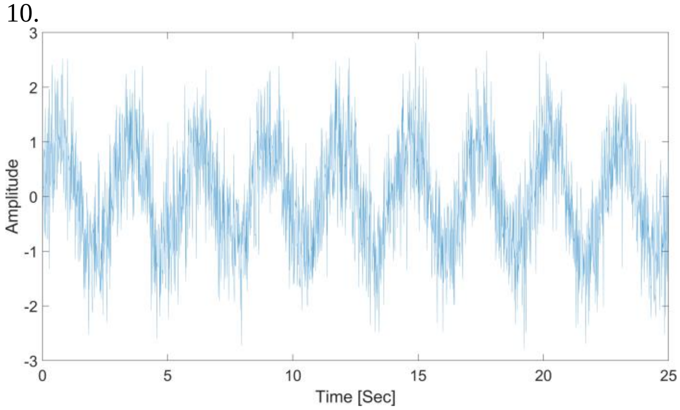
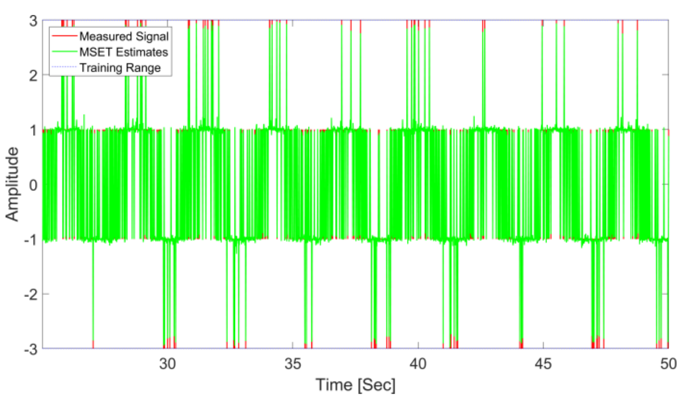
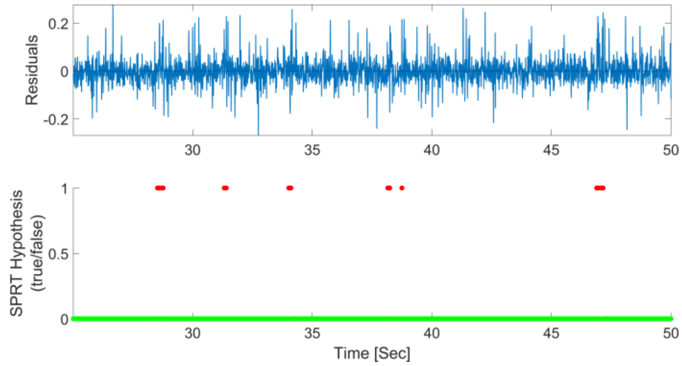


Figure 8: The first half of a 50 sec long signal sampled at 100Hz is used as the training data to create an MSET model.

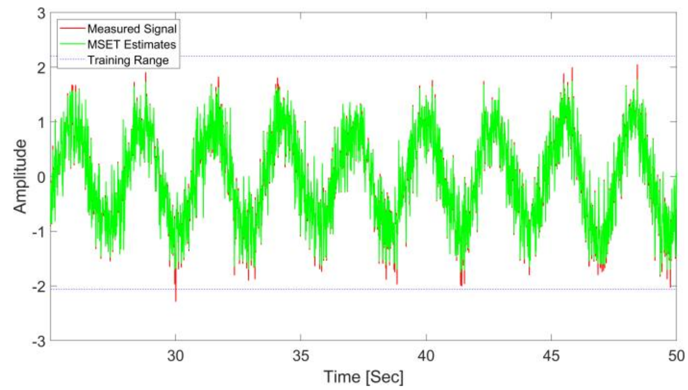


(a)

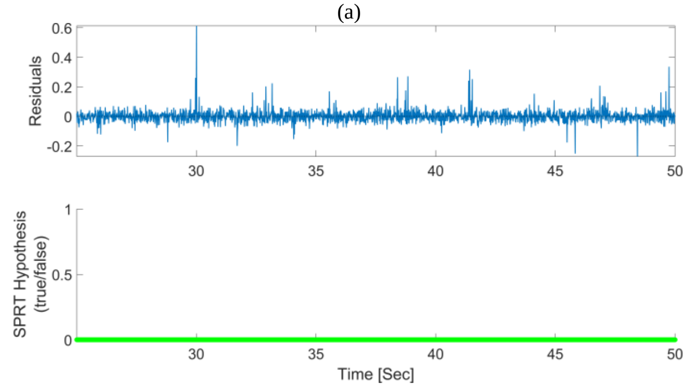


(b)

Figure 9: The second half of the same continuous signal is quantized at QL = 4 before sending to the prior built MSET model as the surveillance data for anomaly detection (a). The residuals between the surveillance data (green) and the corresponding MSET estimates (red) and the subsequent anomaly detection results are illustrated in the top and bottom subplots in (b) respectively.



(a)



(b)

Figure 10: Same as Figure 9 except the prior quantized surveillance data is unquantized first before compared with the corresponding MSET estimates (a). The resulting residuals do not trigger any false alarms (b).

Figures 11-14 illustrate the prognostic performance gains of the unquantization technique with respect to lower missed alarm probability (MAP) through another anomaly detection example. Figure 11 presents the testing signal with degradations starting at #3750. Similar to the previous example, the first half of the signal was used to train a MSET model, which was then used to find out the degradations in the surveillance data (i.e. the second half). The deviations between the surveillance data and the corresponding MSET estimates started to become significant at #4100, where the prognostic alarms were triggered (Figure 12).

On the other hand, the surveillance data was quantized (QL = 3) and the same anomaly detection process was repeated. Figure 13 illustrates the quantized version of the degraded signal caused much fewer prognostic alarms, indicating significant missed alarms which can be costly in the safe critical industries. To proceed, we applied the unquantization technique to the quantized surveillance data and repeated the anomaly detection again. As illustrated in Figure 14, while the resulting alarms did not begin as early as in the original example in Figure 12, they appeared to begin much more earlier and revealed the severity of the degradation more accurately when compared with the scenario in Figure 13, because the unquantization process was able to retrieve the meaningful physical characteristics to a great extent from the quantized signal.

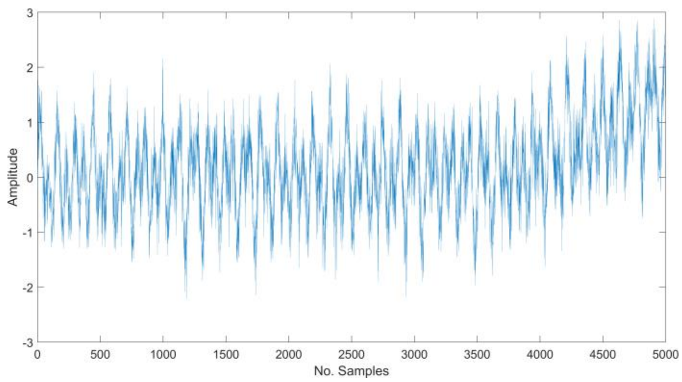


Figure 11: A time-series testing signal with a ramp inserted between #3750 and #5000 resembling degradations.

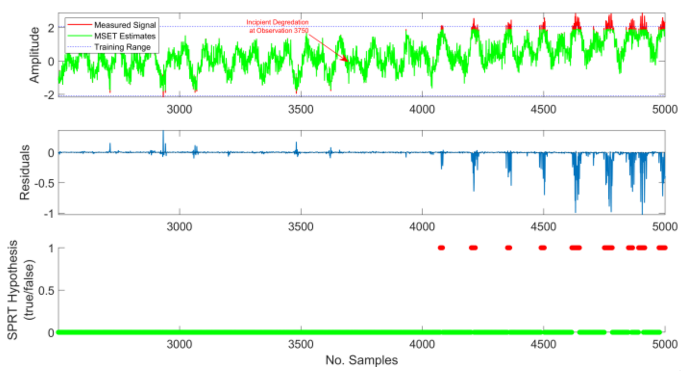


Figure 12: A MSET based anomaly detection is executed over the second half of the signal as the surveillance data where the degradations locate and the first half of the signal as the training data. The top plot compares the surveillance data (green) with the corresponding MSET estimates (red). The resulting residuals and the trigged alarms are presented in the middle and bottom plots respectively.

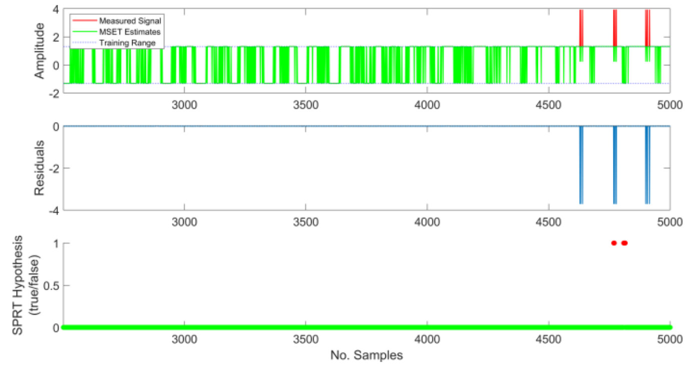


Figure 13: Same as Figure 12 except the surveillance data has been quantized at QL = 3 before the same anomaly detection is executed again. Substantial missed alarms are observed in reference to Figure 12.

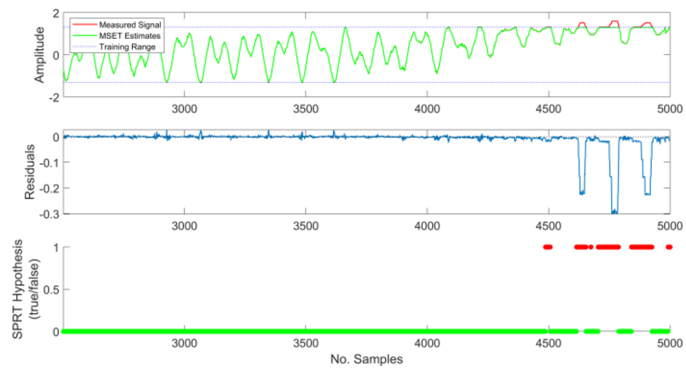


Figure 14: Same as Figure 13 expect the quantized surveillance data has been unquantized before the same anomaly detection is executed once again. Most of the missed alarms in Figure 13 are discovered comparted to Figure 12.

IV. CONCLUSIONS

Addressing the negative impacts of quantized signals benefits the IoT sectors of Utilities, Oil&Gas, Manufacturing, Transportation, and other sensor dense industries when it comes to ML prognostics. In this paper we propose a novel technique that is able to convert the quantized signals that are typically unanalyzable to smooth signals that matches the original sensor output as closely as possible. With this technique, signals that previously required human surveillance or were monitored with simple thresholds can now be analyzed automatically with greater precision. Another major benefit is that this technique can improve all types of ML algorithms, such as Neural Nets or Support Vector Machine. Any ML algorithm intended for time-series analysis will attain higher prognostic accuracy for discovering subtle anomalies in critical assets and processes, and with much lower false-alarm and missed-alarm probabilities. While unquantization already presents major strides in time-series analysis, more research on this topic is currently being done at OracleLab.

REFERENCES

[1] Wang, G. C. and Gross., K. C. (Dec., 2018). “Telemetry Parameter Synthesis System for Enhanced Tuning and Validation of Machine Learning Algorithmics,” IEEE 2018 Intn'l Symposium on Internet of Things & Internet of Everything (CSCI-ISOT), Las Vegas, NV.

[2] Dhanekula., R. C. and Gross, K. C. (Jul., 2012). "High Fidelity Telemetry Signal Synthesis for Improved Electronic Prognostics," IEEE World Congress in Computer Science, Computer Engineering, and Applied Computing (WorldComp2012), Las Vegas, NV.

[3] Gross, K. C., & Schuster, E. (2005). "Spectral Decomposition and Reconstruction of Telemetry Signals from Enterprise Computing Systems," In CDES (pp. 240-246).

[4] Singer, R. M., Gross, K. C., Herzog, J. P., King, R. W., & Wegerich, S. (1997). "Model-based nuclear power plant monitoring and fault detection: Theoretical foundations," Proc. 9th Intnl. Conf. On Intelligent Systems Applications to Power Systems, pp. 60-65, Seoul, Korea (July 6-10, 1997).

[5] Gross, K. C., Singer, R. M., Wegerich, S. W., Herzog, J. P., VanAlstine, R., & Bockhorst, F. (1997). "Application of a model-based fault detection system to nuclear plant signals," Proc. 9th Intnl. Conf. On Intelligent Systems Applications to Power Systems, pp. 66-70, Seoul, Korea (July 6-10, 1997).

[6] F. Zhang, S. Boring, J. W. Hines, J. Coble, and K. C. Gross, Combination of Unquantization Technique and Empirical Modelling for Industrial Applications," 2017 American Nuclear Society Intn'l Conf., Washington D.C. (Nov 2017).

[7] K. C. Gross and G. C. Wang, "AI Decision Support Prognostics for IoT Asset Health Monitoring, Failure Prediction, Time to Failure," IEEE 2019 Intn'l Symposium on Artificial Intelligence (CSCI-ISAI), Las Vegas, NV (Dec 5-8, 2019).