Application of Generative Autoencoders in the Detection of Anomalies in Hypercompressors

Zoroastro Fernandes Filho[1*], Alex Álisson Bandeira Santos[1]
[1]SENAI CIMATEC University Center; Salvador Bahia Brazil

Hypercompressors are essential assets that compress high-flow-rates of ethylene to pressures between 100~350 MPa in the LDPE industry. They are sources of essential risks and costs. This work proposes an unsupervised and univariate monitoring method for detecting anomalies in hypercompressors through data collected from an online monitoring system in an actual installation. A variational autoencoder learns the process of generating shapelets associated with vibrational patterns. A combination of matrix profile algorithms automatically selects the training data set. A *β*-VAE composed of MLP layers is trained and applied on the input space so that a voting operation and a box-cox transformation on the absolute residual errors between the inputs and outputs lead to the upper outlier detection threshold, obtained by the Tukey fence method. The model detected suspicious vibration patterns classified a priori as potential anomalies.
Keywords: Hypercompressors. Matrixprofile. *β*-VAE. Anomalies.

Introduction

The polymerization of low-density polyethylene (LDPE) is processed on an industrial scale under temperatures of 200 to 300 °C and pressures of 100 to 350 MPa, achieved through the compression of ethylene by special reciprocating compressors (hypercompressors, occasionally also called secondary compressors). Modern LDPE mega plants are equipped by hypercompressors with capacities of up to 400 kNm³ and 33,000 kW [1]. Due to the flammability and high pressures involved, potential risks are inherent in this process, and thus, these machines are sources of essential risks and costs for the LDPE industry. The development of anomaly detection methods based on monitoring data is essential.

According to Park and colleagues [2], a large amount of data, and appropriate sampling rates, becomes mandatory for an acceptable mathematical model of a hypercompressor. There are some examples of conventional reciprocating compressors in the literature, but it is rare for hypercompressors [3,4]. The principal component analysis (PCA) is widely used among the techniques available for creating models. Considering LDPE plants, according to Park and colleagues [2], some researchers have developed fault detection models based on PCA to predict dangerous thermal decomposition reactions due to sudden compression or excess peroxide injection that occur in autoclave reactors of LDPE plants [5,6]. However, the theoretical assumption of the classical PCA transformation, assuming linearity and projecting the data into a low-dimensional latent space, may need to fit better with the nonlinear nature of vibrations typically monitored in reciprocating compressors in general. We propose the application of a variational autoencoder (VAE) to overcome this limitation.

According to Sivalingam and colleagues [7], while a classical autoencoder (AE [8]) learns to make predictions from some observations, the variational autoencoder (VAE [9]) learns to simulate the data generation process. An essential effect of VAE is the possibility of revealing a potential understanding of the inherent causal relationships of the input data entangled in the latent space, providing a better generalization of the data and not only considering time-variant or invariant data. In this way, VAE allow its application in monitoring

Received on 19 March 2023; revised 24 May 2023.
Address for correspondence: Zoroastro Fernandes Filho. Avenida Orlando Gomes, 1845, Piatã. Salvador, Bahia, Brazil. Zipcode: 41650-010. E-mail: zorofernandes@hotmail.com.

stationary or non-stationary processes, typically displayed in monitoring hypercompressors.

The data used in this work consists of 36 vibration variables (for each hypercompressors cylinder) and the motor's electrical current. A set of operating data considered "non-suspicious" was automatically selected by a method composed of an algorithms combination based on matrixprofile (MP) [10] and MP snippet [11]. Then, a β-VAE [13] was trained, and the latent space obtained by the encoder and the residual error about the encoder inputs led to the threshold of outliers, discriminating "suspicious" vibratory patterns from "normal" ones through a box-cox transformation combined with the "Tukey fence" method [14]. The extracted suspicious patterns were classified as potential anomalous cylinder vibration shapelets [15]. Data sampling (~ six continuous months) was performed every minute, starting from an online vibration monitoring system in an actual installation.

Materials and Methods

Process Description

Figure 1 describes the LDPE production process (autoclave reactor) in a simplified way. Before bagging the pellets, LDPE is produced in three stages: Compression, Reaction, and Extrusion. First, the ethylene is compressed sequentially by the two compressors (primary and secondary), reaching the reaction pressure. In the reactors, in addition to ethylene, a reaction initiator is also injected, up to a maximum of 350 MPa, and thus the polymerization reaction is initiated. Ethylene gas is the primary raw material. The compression section requires most of the energy of the production process, and the hyper compressor is the leading equipment in the chain described here.

Sensor Location and Compression Cycle

An accelerometer measures the impacts occurring in the intermediate body, those resulting from the inertia of the parts, and the loads arising from the effort to compress the gas ([Figure 2 (side view of a hyper compressor cylinder)].

With each rotation of the shaft (360°), starting from the Top Dead Center (TDC), the compression cycle is divided into four stages: 1) 0° to 25° – re-expansion, followed by the opening of the suction valve, 2) 25° to 180° – suction, followed by closing the suction valve, 3) 18° to 290° – compression, followed by opening the discharge valve and 4) 290° to 360° – expansion, followed by closing the discharge valve.

Figure 1. LDPE: Gas flowchart.

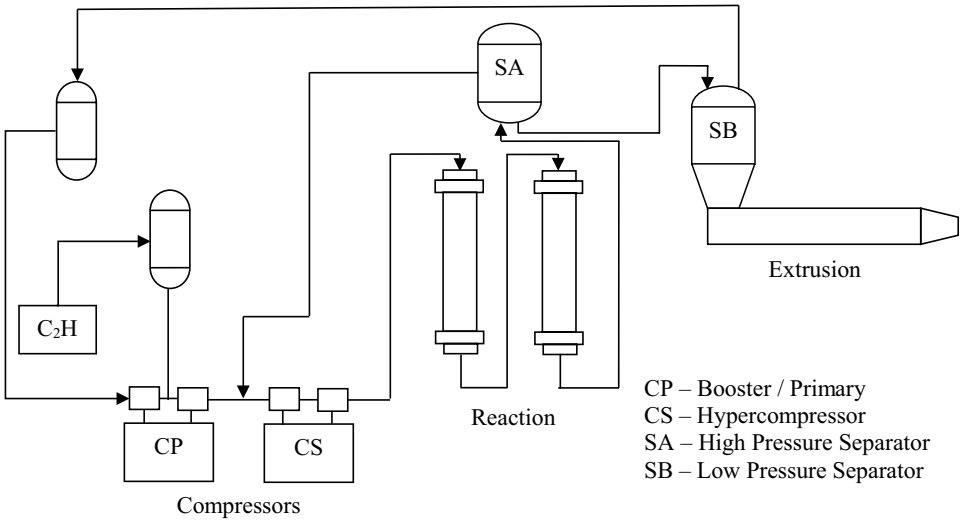
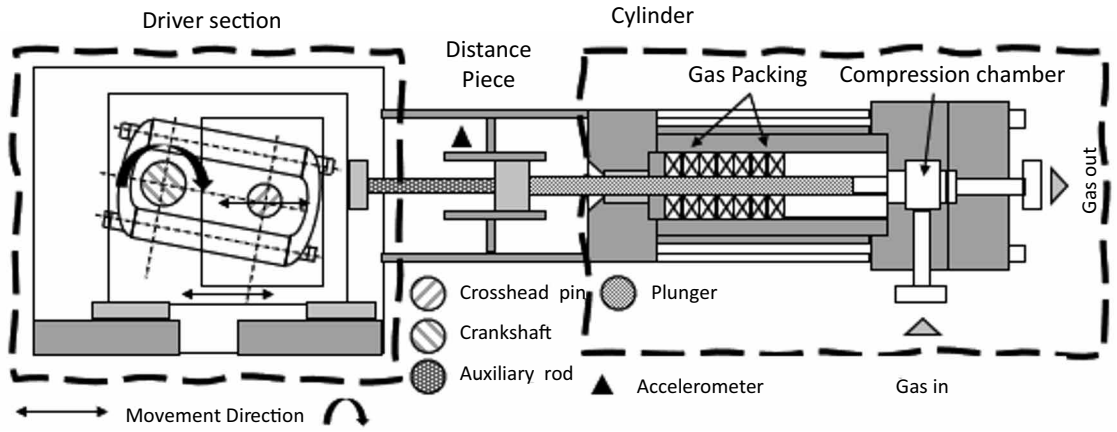


Figure 2. Hypercompressor – Accelerometer location.



Each shaft rotation triggers a phase sensor, which is not shown in the Figure.

Dynamic Patterns (shapelets)

Vibration (acceleration) data is available in the Root Mean Square (RMS) metric and collected at sampling rates of the order of 20 kHz. The RMS values are measured and calculated in segments, following time intervals relative to each 1/36th crankshaft rotation every minute. Each measured segment is stored in a database. Thus, typical vibration patterns at each shaft turn, composed of 36 segments (shapelets), emerge as objects of investigation (Figure 3). The numbering of the segments in the figure corresponds to the angular movement of the crankshaft shaft, which corresponds to 10-degree angles.

Each segment is monitored as a univariate time series in the collected data. The shaft rotation speed is approximately 200 RPM. The compressor has eight cylinders and two stages, four in each, labeled 1-1A, 1-1B, 1-2A, 1-2B for the 1st stage, and 2-1A, 2-1B, 2-2A and 2-2B for the 2nd stage (Figure 4).

Monitored Variables and Database

In this work, the data from the monitored segments were rearranged sequentially and displayed into a univariate time series (Figure 5). The 36 time series for each cylinder were converted into a single univariate time series, referred to as RMS_Avg. The conversion provides information about the evolution of vibration patterns over time. Consequently, the final length of RMS_Avg is a multiple of 36. Data from all cylinders were used.

Table 1 shows the variables (in univariate form) monitored for 6 months.

Figure 3. Typical vibration with each turn of the shaft.

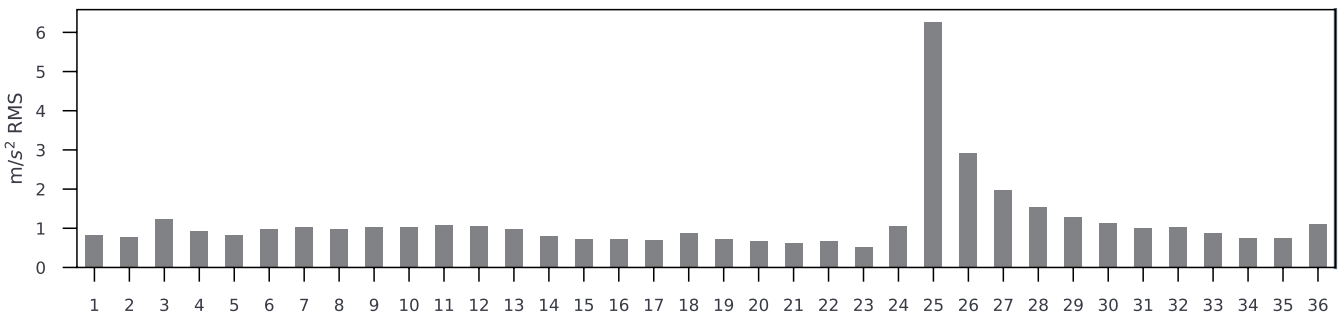
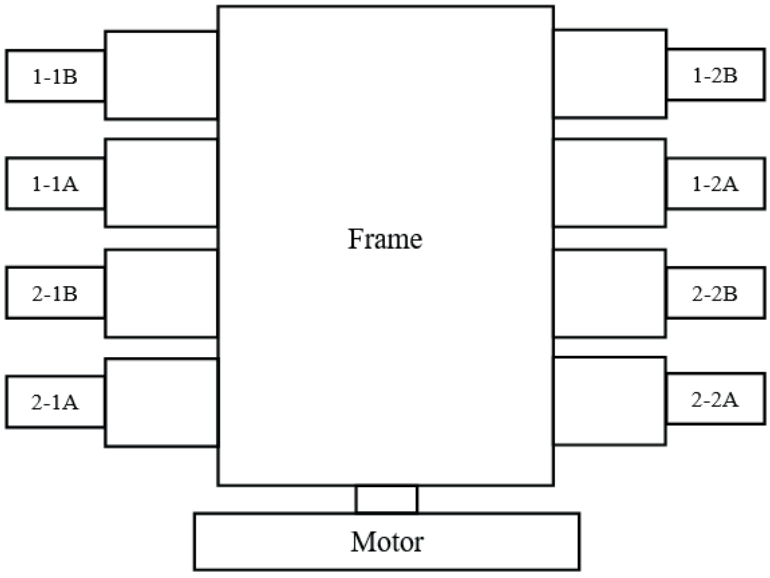


Figure 4. Physical position of the cylinders.



Training and Testing Data

In extensive collections of reciprocating compressor monitoring data, the challenge arises of automatically labeling the most typical and/or representative data in various operating contexts considered "normal". We proposed the use of an algorithm based on matrixprofile (MP) [10] and MP snippet to overcome this barrier, [11]. In Imani and colleagues [11], the authors demonstrated the usefulness of the MP snippet algorithm to, at a high level, seek to answer the question "*Show me some representative/typical data...*". According to the authors, despite being trivial in many domains, surprisingly, the difficulty persists in extensive data collections, given the complexity of answering what "representative/typical data" actually means. In this work, the balance in the choice of hyperparameters of the proposed algorithm is expected to maximize the probability

Figure 5. Univariate time series (RMS_Avg).

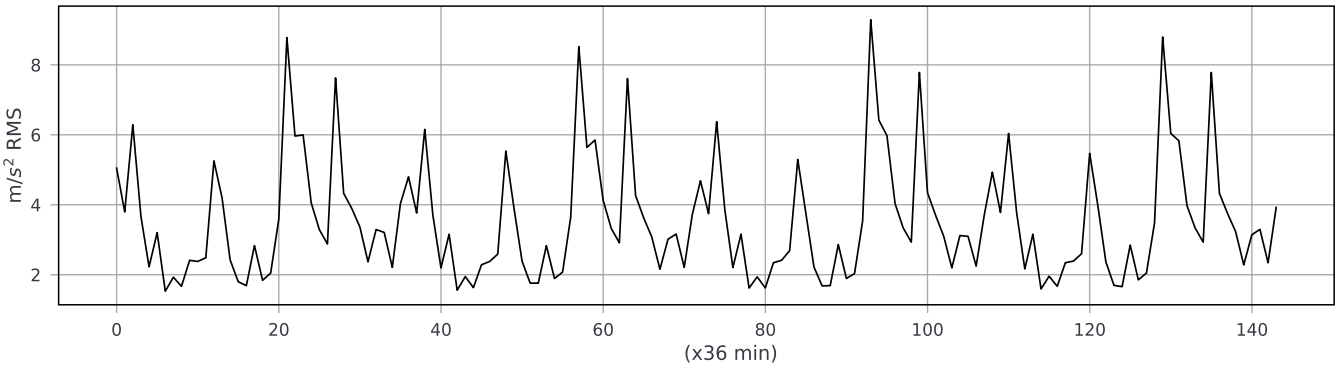


Table 1. Monitored variables.

Series	Description	Variables	Instances
RMS_Avg	Acceleration	8	32054868
Current	Electric current	1	890413

of selecting the most representative data possible for training instances.

Steps for Creating the Model

Figure 6 shows the model development steps, from cleaning the data to obtaining outliers.

Data cleaning (step 1) excludes sensor errors and removes instances with missing data and periods with vibration values above pre-defined alarms. Step 2 requires some heuristics, complementing data cleaning, excluding stop and start intervals. To this end, we selected operational periods when the motor operates above a minimum power value (electric current > 300 A), consistent with full LDPE production operational regimes.

An unusual shapelet (discord [10]) tends to have a high distance profile value (PD [10]). In the RMS_Avg series, discords are expected to be extremely rare due to the cyclic vibration patterns typical in reciprocating compressors. A "normal" shapelet (motif [10]), more common, arising in typical operational situations, tends to lower PD values. The two classes (motifs and discords) are fundamental objects in the search motors contained in MP algorithms. The idea is to use such mechanisms to select training instances.

For each cylinder, we divided the RMS_Avg series into k contiguous subseries, Figure 5, step 3, so all subseries contained only groups of 36 instances equidistant one minute apart.

In step 4 of Figure 5, the PD's, (series T') are calculated by applying the MP algorithm on each of the k separate subseries, $T = \{T_1, T_2 \dots T_k\}$, $T_i = \left\{a_1^i, a_2^i, \dots a_{p_i}^i\right\}$, $(i = 1,2,\dots k)$. The MP algorithm was applied sequentially on sections of sequential windows of length w and moving windows of length $m = 36$ on each T_i, obtaining the set of series $T_i' = \{T_1', T_2' \dots T_k'\}$, $T_i' = \left\{a'^{\,i}_{\,1}, a'^{\,i}_{\,2} \dots a'^{\,i}_{\,p}\right\}$, $(i = 1,2,\dots k)$. w and m are hyperparameters.

For step 5, the idea is to take advantage of the periodic nature of vibrational patterns and the consequent expectation of regularity (albeit relative) of these patterns in different sections of each T_i'. We applied the MP snippet algorithm on T' to reduce the probability of selecting "suspicious" shapelets as "normal". The algorithm searches each consecutive section of size w of each T_i', adopting moving windows of size m ($m = 36$), extracting its most representative shapelets. This occurrs because very sudden changes in PD values (suspicious transients) tend not to represent the fixed window considered and, therefore, not detected by the MP snippet algorithm as a training instance. The integer value Qt indicates the desired number of shapelets retrieved from each T'. The indices of the snippets retrieved in T' indicate the location of the shapelets retrieved in T. Qt is a hyperparameter.

By setting the parameter $m = 36$, we maintain the expectation that the balance between the choice of Qt and w guarantees the adequate selection of training instances. At the end of the process, $k * Qt$ indices of the T shapelets will be selected for training for each cylinder. Chart 1 shows the pseudocode described in the previous paragraphs. Data from the eight cylinders were used for training the VAE.

As an example, taking m fixed and equal to the length of the shapelets $m = 36$, the hyperparameters in Table 2 are expected to automatically select two "normal" shapelets every 30 minutes.

Intuitively, a higher value of w implies a lower probability of selecting more representative data because a smaller number of representative shapelets will be recovered for the same section of size w. The converse is accurate; however, in this case, a more significant risk is expected in selecting suspicious shapelets as more representative data because there is a greater probability of persistent suspicious values occurring, with a duration of the order of the size of the fixed window w.

A higher value of Qt implies a more significant amount of data for training; however, due to the inherent repeatability of shapelet patterns, it is expected that from a particular value of Qt onwards, an increase in the number of training instances does not necessarily imply greater capability to learn patterns..

In Figure 6, steps 6 to 10 describe the process from calculating the VAE to the threshold of outliers obtained by the Tukey fence method.

Figure 6. Steps for creating the model.

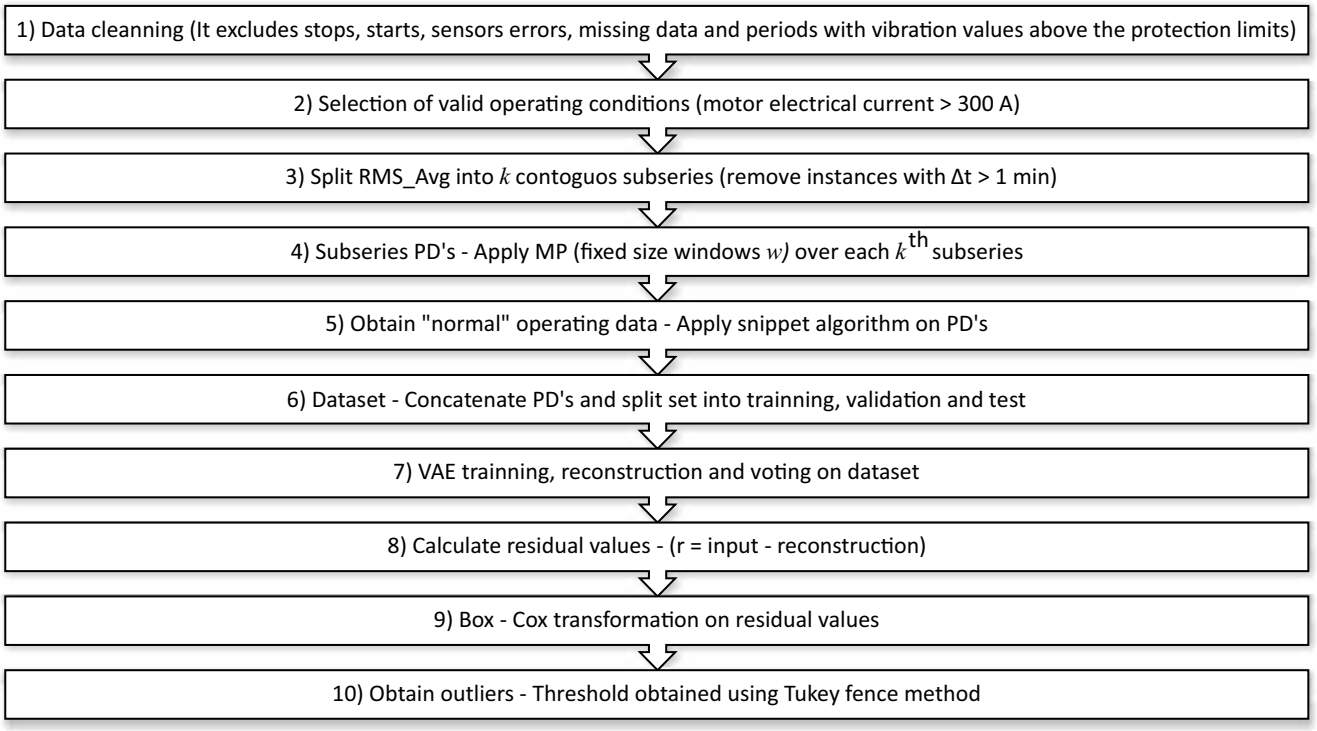


Table 2. MP hyperparameters and MP snippet.

MP				**MP snippet**					
m	36	*w*	30 min	*Qt*	2	*m*	36	*w*	30 min

<u>β-VAE and Vibration Monitoring</u>

Compressor vibration patterns are diverse, but this diversity arises from a relatively small and coherent set of mechanical properties, physical rules, construction, and operating principles. VAEs can learn representations in the latent structure of data and extract concepts based on the learned patterns.

Variations of VAEs are available in the literature. In β-VAE [13], the cost function $l(x, \hat{x})$ includes the real hyperparameter β, with $l(x, \hat{x}) = l_{rec} + \beta l_{KL}$. While the reconstruction term l_{rec} leads to the separation of points in the latent space, the second term βl_{KL} (Kulback-Leibler divergence multiplied by β) does the opposite, avoiding the learning of non-existent representations in the original data space [7]. According to the authors, β "encourages/ discourages" the disentanglement of latent variables. Table 3 shows the architecture used in β-VAE and the hyperparameters.

<u>Inference</u>

For each cylinder, we concatenate the T_i subseries and rearrange them into sequential batches of 36 units, one-time unit apart concerning the previous batch, along the entire length (size) of the concatenated subseries, obtaining a new series of sequential batches L^{obs}, of size (p_i-36, 36).

Then, the trained β-VAE reconstructed each batch contained in L^{obs}, obtaining the analogous batch series L^{rec}.

Chart 1. Pseudocode for automatic selection of VAE training instances.

	Input: $= T\ \{T_1 T_2,\ \ldots T_k\} T_i\ ,\ a\left\{ = a_1^i, a_2^i, \ldots\ a_{p_i}^i \right\}, i = 1, 2 \ldots k$
	Output: $ind\_train\_set = \{ind_1, ind_2 \ldots\ ind_n\}$, *n=1...k*Qt*
1	$ind\_train\_set = \{\}$ *// empty index set*
2	$T' = \{\}$ *// empty set of subseries*
3	***for (i=1, k, 1)*** *Do:// Scroll k (sub)subseries* T_i *calculating the PD's*
4	***for (j=1, w, w)*** *Do:// Compute the MP algorithm sequentially over* T_i *with m=36 and fixed window=*w *obtained* T_i'
5	$T_i' \leftarrow MP(ts = T_i[\,j{:}j + w + m - 1], window\_size = m)$ *// Algorithm MP according Law SM [12]*
6	$T' = T' + \{T_i'\}$ *// concatenate getting* T'
7	***for (i=1, k, 1)*** *Do:// Scroll k subseries* T_i', $T' = \{T_1', T_2', \ldots T_k'\}$
8	***for (j=1, w, w)*** *Do: // Compute the algorithm MP snippet sequentially over* T_i' *with m=36, quantity of snippets = Qt and fixed window=*w
9	$train\_set_i \leftarrow\ MP\_discover\_snippet(ts = T_i'[j{:}j + w], window = m, num\_snippets = Qt)$ *// set of selected instances of the j-th section of* T_i'. *Algorithm* $MP\_discover\_snippet$ *Law SM [12]*
10	$ind\_train\_set_i \leftarrow train\_set_i.indices$ *// retrieve the indexes of selected instances on j-th section of* T_i'
11	***for (i=1...k)*** *Do: // Concatenate the indexes retrieved from the selected instances*
12	$ind\_train\_set = ind\_train\_set + \{ind\_train\_set_i\}$

Table 3. β-VAE architecture.

Scaler	MinMax
Architecture	MLP
Optimizer	Adam
Training + validation data	(36495, 36)
Test data	(1921, 36)
Input + encoder layers	(Dense + BatchNormalization) (36, 27, 18, 18, 9)
Loss function	LeakyRelu
Latent space	1
Decoder + output layers	(Dense + BatchNormalization) (9, 18, 18, 27, 36)
Regularizer	1E-11
Epochs	80
Batch size	1
β	0.0001

A new series containing the residual values $r = abs(L^{obs} - L^{rec})$ was computed. The lst indices of the best residuals ($lst = argmin(r)$ define, by voting, the choice of the best inference over L^{rec}.

Chart 2 presents the pseudocode. As additional information, the VAE training and test sets were transformed according to steps 6 and 7, aiming to more comprehensive learning of the shapelet generation process.

Results and Discussion

Table 4 shows the MAE and MSE metrics reconstructions obtained in our study.

Figure 7a shows an example of an anomaly detected in cylinder 2-2B. The filled area (in red) highlights the difference between the observed and reconstructed values. Figure 7b shows the histogram and the detection threshold calculated on the transformed residuals (box-cox transformation) over T. The decision threshold is shown in the vertical red line ("Tukey fences", k=3) - values higher than the threshold are considered "suspicious". The anomaly detection rate reached 0.0142% across the entire dataset.

Figure 7c shows four shapelets reconstructed from equidistant values of the latent space [-3.748, -0.416, 2.915, 6.247] within the quantile interval [q0.0001, q0.9999]. Blue vertical lines separate the four patterns shown. This example illustrates how it is possible to understand the diversity of vibration patterns that arise from operational scenarios.

The histogram in Figure 7d shows the distribution of the latent space in 4 equally distributed compartments, also within the quantile range mentioned in item 7c.

Conclusion

This work aimed to establish a model for detecting potential vibration anomalies in hypercompressor cylinders.

The model was developed using operational data collected every minute for six months on a hypercompressor in an actual installation. The model considered 288 vibration variables, which were transformed into 8 univariate time series, each corresponding to a distinct cylinder. In parallel, another univariate time series, from the same time interval, representing the motor's electrical current used to finalize data cleaning, selecting periods in operational regime of interest.

Operation data labeled "normal" was extracted from the operation data using a combination of the MP and MP snippet algorithms.

A box-cox transformation was applied to the residual absolute errors between the values reconstructed by β-VAE and its inputs, and an outlier detection threshold was obtained by applying the Tukey fence method.

The anomaly detection rate reached 0.0142% on the selected dataset.

The research demonstrated the possibility of establishing a simplified model to detect potential vibration anomalies in a hypercompressor, allowing specialist technicians to focus their investigations on these instances and to correlate detected deviations with incipient operational and maintenance problems, improving safety and operational costs in LDPE plants.

For future studies, we proposed to carry out a qualitative assessment of the physical significance of the detected anomalous shapelets.

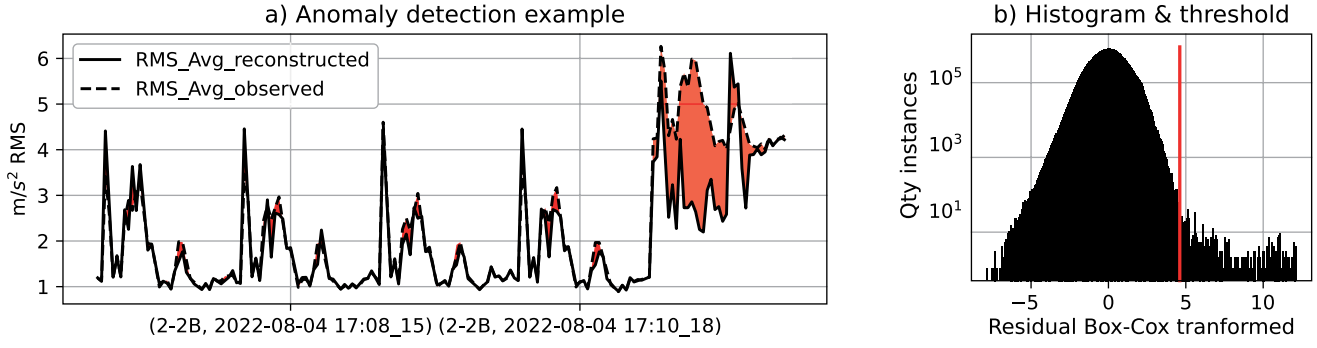
Table 4. Reconstruction metrics

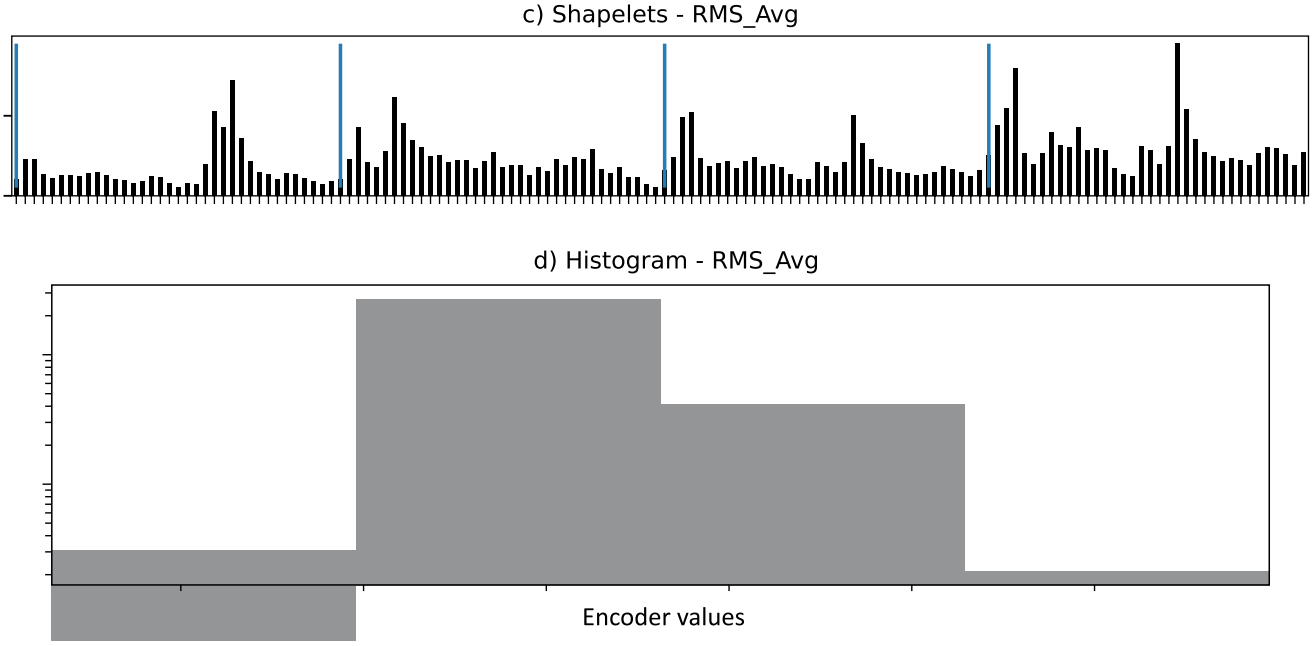
	MAE	MSE
RMS_Avg - training + validation	0.1055	0.0805
RMS_Avg – test	0.0841	0.0316

Chart 2. Pseudocode for VAE inference.

	Input: $T = \{T_1, T_2 \dots T_k\},\ T_i = \left\{a_1^i, a_2^i, \dots\ a_{p_i}^i\right\}, i = 1, 2 \dots k$ *// Observed values* $(T = RMS\_Avg_{obs})$ *//* p_i *is the lenght of series* T_i
	Output: $T^{rec} = \{T_1^{rec}, T_2^{rec} \dots T_k^{rec}\},\ T_i = \left\{a\_r_1^i, a\_r_2^i, \dots\ a\_r_{p_i}^i\right\}, i = 1, 2 \dots k$ *// Reconstructed observed values* $(T^{rec} = RMS\_Avg_{rec})$
1	*size* $\leftarrow (\sum_1^k p_i)$ *// Sum of the length of all subseries* T_i
2	$L^{obs} \leftarrow \{\}$ *// Empty set*
4	$T \leftarrow concat(\{T_1, T_2 \dots T_k\})$ *// Concatenate the subseries* T_i *into a single series of length=size*
5	$T \leftarrow scaler.apply(T)$ *// Compute scaler used in VAE trainning over T, for reconstruction inference*
6	***for (i=0, size-36, 1)*** *Do: // Cycle through the observed values and rearrange them in batches of 36 elements*
7	$L^{obs} \leftarrow L^{obs} + T[i: i + 36]$ *// Add to the set* L^{obs}*, sequential batches of 36 elements contained in T*
8	$L^{rec} \leftarrow encoder(decoder(L^{obs}))$ *// Inference on the batches of 36 elements contained in L*
9	$L^{rec} \leftarrow reshape(L^{rec})$ *// Reshape of (size-36, 36) to ((size-36)*36,1)*
10	$L^{rec} \leftarrow scaler.inverse.apply(L^{rec})$*// Scaler inversion of the step 5*
11	$L^{rec} \leftarrow reshape(L^{rec})$ *// Reversion of reshape of the step 9*
12	***for (i = 36, size, 1)*** *Do: // Go through the entire timeseries and rearrange in batches taking diagonals to i of* L^{rec} *e* L^{obs}
13	$diag^{rec} \leftarrow L^{rec}.diagonal(i)$ *// Reconstructed values - Each new batch with 36 (distinct) inferences of the same instance*
14	$diag^{obs} \leftarrow L^{obs}.diagonal(i)$ *// Observed values - Each new batch with 36 (same) inferences of the same instance*
15	$r_i \leftarrow abs(L^{obs} - L^{rec})$ *// Compute the residuals* r_i
16	$lst \leftarrow argmin(r_i)$ *// Find the indexes of the best values by vote (lower residuals)*
17	$T^{rec} \leftarrow L^{rec}[lst]$ *// Assign the best values inferred from the lst index list*

Figure 7 . a) Example of anomaly detection, b) Histogram, outliers, and detection threshold, c) Examples learned by VAE, d) Relative frequency of shapelets occurrence.





As a second option, compare the nature of the detected anomalies by varying the size of the latent space, the family and architecture of the autoencoder, and the hyperparameters of the automatic selection of the training dataset.

References

1. Compression B. Compressors-for-LDPE. https: //www.burckhardtcompression.com/, 2018. Disponivel em: https://www.burckhardtcompression.com/wp-content/uploads/2021/03/Compressors-for-LDPE_210305.pdf. Accessed on January 31, 2023
2. Park BE et al. Anomaly detection in a hyper-compressor in low-density polyethylene manufacturing processes using WPCA-based principal component control limit. Korean Journal of Chemical Engineering 2020;37:11-18.
3. Xiao S et al. Dynamic behavior analysis of reciprocating compressor with subsidence fault considering flexible piston rod. Journal of Mechanical Science and Technology 2018;32:4103-4124.
4. Xiao S et al. Dynamic analysis for a reciprocating compressor system with clearance fault. Applied Sciences 2021;11(23):11295.
5. Sivalingam G, Soni NJ, Vakil SM. Detection of decomposition for high pressure ethylene/vinyl acetate copolymerization in autoclave reactor using principal component analysis on heat balance model. The Canadian Journal of Chemical Engineering 2015;93(6):1063-1075.
6. Kumar V et al. Multivariate statistical monitoring of a high-pressure polymerization process. Polymer Reaction Engineering 2003;11(4):,1017-1052.
7. Canziani A. Generative Models - Variational Autoencoders, 2020. Disponivel em: <https://atcold.github.io/pytorch-Deep-Learning/pt/week08/08-3/>. Accessed on January 25, 2023.
8. Hinton GE, Salakhutdinoc RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504-507.
9. Kingma DP, WwllingM. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013.
10. Yeh CCM et al. Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In: 2016 IEEE 16th International Conference on Data Mining (ICDM). IEEE 2016:1317-1322.
11. Imani S et al. Matrix profile XIII: Time series snippets: A new primitive for time series data mining. In: 2018 IEEE International Conference on Big Knowledge (ICBK). IEEE 2018:382-389.
12. Law SM. STUMPY: A powerful and scalable python library for time series data mining. Journal of Open Source Software 2019;4(39):1504.
13. Higgins I et al. beta-vae: Learning basic visual concepts with a constrained variational framework. In: International conference on learning representations, 2016.
14. Tukey JW. Exploratory data analysis. Addison-Wesely, 1977.
15. Ye L, Keogh e. Time series shapelets: a new primitive for data mining. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2009:947-956.