Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-6 (2022) 366–371

Health Indicator for Batch Processes Based on SP-LASSO ⋆

Dima EL JAMAL * **Bouchra ANANOU** *
Guillaume GRATON ** **Mustapha OULADSINE** *
Jacques PINATON ***

* *Aix Marseille Univ, Univ de Toulon CNRS, LIS (UMR 7020), 13397 Marseille, France (e-mail: dima.el-jamal@lis-lab.fr)*
** *Ecole Centrale Marseille, Technopôle de Château-Gombert, 13451 Marseille, France*
*** *STMicroelectronics Rousset, 13106 Rousset, France*

Abstract:
For decades, manufacturers have been collecting and storing high amounts of data with the aim of better controlling and managing their processes. With the vast amount of information and hidden knowledge in all of these data, the challenge for these manufacturers to monitor their equipment units, is the extraction of an appropriate health indicator from these data that illustrates the actual state of their equipment units. In this paper, we are interested in extracting the health indicator of semiconductor equipment where manufacturing is performed by batch. For that, a novel automatic approach named Significant Points combined to the Least Absolute Shrinkage and Selection Operator (SP-LASSO) is proposed. This approach is mainly based on LASSO regression model. Its accuracy is illustrated by numerical application on simulated data.

Keywords: Health indicator extraction, feature selection, LASSO, batch processes, semiconductor manufacturing

1. INTRODUCTION

Over the last few years, with the increasing worldwide competition, industries have had to constantly innovate in order to enhance their performance, productivity and minimize the downtime. Monitoring the state of health of their equipment units is important to achieve these goals as they can always go through a process of degradation which affects the stability of fabrication and the yield's rate. The equipment degradation level can be qualified by an index which is called Health Indicator (HI). As the modern production equipment units are increasingly equipped with sensors and due to the improvement of the technologies of signal processing and data storage, large amount of data is displayed to be exploited, processed and analyzed to extract HIs.

In the case of the semiconductor manufacturing, batch processes data collected from the equipment units are usually referred to Fault Detection and Classification (FDC) data. These data are collected continuously during wafer processing, and contain comprehensive information about process conditions and equipment state. The batch processes data are handled in three dimensional matrix (wafer × sensor × observation) as explained in Rostami et al. (2018), Du (2019) and Korabi et al. (2021). Their structure is different from the two dimensional data sets .

In the literature, dealing with the irregular three dimensional data (3D) is not sufficiently explored, and most of the works applied on batch processes proceed firstly by converting the data set into a two dimensional matrix (2D). Among the different ways of converting the data from 3D to 2D, the sensor-based structure used in Nguyen et al. (2015b): the batch processes data are restructured into a 2D matrix as (wafer × mean (sensor)), where the sensor signal is resumed to its mean for each wafer. Another way is proposed by Rostami et al. (2017) which consists in transforming the data into 2D matrix by Wavelet Packet Decomposition. Each sensor signal is decomposed into energy values that are considered as features. Hence, the obtained matrix is represented as (wafer × energies). In Nguyen et al. (2015b), the data are batch-wise unfolded: each row of the obtained 2D matrix corresponds to a wafer and each column corresponds to the point (sensor, observation). In this paper, the batch process data are transformed into 2D matrix and the batch-wise unfolding is adopted. Contrary to the average computing restructuring, the changes along the sensor profiles are taken into account. Besides, the variables signification is kept as all original data are used without transformation in opposition to the method in Rostami et al. (2017).

However, the problem is that the batch-wise restructuring leads to the curse of dimensionality when the data set is quite large. It becomes difficult to build and interpret a HI extraction approach that takes into consideration all the features. In addition, some features can contain redundant information, constant values or even attributes which do not reflect the equipment state, thus the reduction of features dimension while keeping the relevant information is a solution to this problem.

⋆ This work is supported by MADEin4

10.1016/j.ifacol.2022.07.156

Two strategies for dimension reduction are resumed in Korabi et al. (2021): feature projection and feature selection. The feature projection consists in building a new set of features based on projecting the initial set in a low dimensional space that captures most of the information. The main issue related to this strategy is the lack of interpretability in the context, for instance the root cause analysis. However, feature selection reduces the initial set of features to a smaller subset without transforming the data and while eliminating the irrelevant features. This strategy is opted in this work.

Feature selection methods can be divided into three types (summarized in Zebari et al. (2020), Malekloo et al. (2021) and Lee et al. (2019)): filter, wrapper and embedded. In the filter type methods, the features are ranked according to a specific statistical criteria. Features having values above a predefined threshold for this criteria are selected. However, this type of methods presents some drawbacks: the interactions between the features are not considered. In addition, it is a non automatic threshold setting-up method as it requires expert intervention to set the threshold. The wrapper type methods generate different subsets of features, each one is then used to train the learning algorithm. Then the best subset is selected by evaluating the performance of this algorithm (e.g. Babajanian Bisheh et al. (2019)). Whereas in the embedded type methods, the selection is based on learning the feature importance as part of the learning algorithm. This type is adopted in this work as it is less complex in computing than the wrapper type methods and it has higher accuracy than the filter type methods.

Among the embedded type methods, LASSO, a l_1-penalized regression method is widely used in the literature. It has proved its effectiveness in different applications on real industrial data like semiconductor, ships, logistics, etc. Several works proposed LASSO-based approaches for different purposes as: (1) Feature selection (Kim et al. (2018)), (2) Prediction (Wang et al. (2018), Melhem et al. (2015)) and (3) Process monitoring (Zou and Qiu (2009), Dong et al. (2019)). However, the applications of LASSO on batch processes data for health monitoring was not performed to the best of our knowledge.

The novelty of this paper is the use of LASSO for HI extraction for batch processes through the proposed approach named as Significant Points combined to the Least Absolute Shrinkage and Selection Operator (SP-LASSO).

The effectiveness of SP-LASSO is demonstrated through the comparison to an existing approach performed in the same context: the Significant Point combined to Principal Component Analysis (SP-PCA) approach developed in Nguyen et al. (2015b). It performs feature selection and HI extraction using respectively a filter-type method and the PCA.

The remainder of this paper is structured as follows: section 2 presents a detailed description of the LASSO-based regression model. The proposed approach is illustrated in section 3. Section 4 presents the approach application on simulated data and a comparison with the SP-PCA approach. Finally, conclusion and perspectives are given in section 5.

2. MODEL DESCRIPTION

2.1 Overview

LASSO (Least Absolute Shrinkage and Selection Operator) was first formulated by Tibshirani (1996). It is a powerful method that performs both regularization and feature selection. The regularization can be performed by constraining the l_1 norm of the coefficients directly, or by adding the l_1 norm of coefficients to the objective function as a penalty. Thus, it encourages sparsity in the component coefficients forcing the coefficients of non-relevant features to be shrunk to zero.

2.2 LASSO-based regression model

Generally, standard linear regression models are formulated as:

$$Y = \beta X + \varepsilon \tag{1}$$

where Y is the response vector and X is the matrix of the explanatory variables. β is the vector of regression coefficients and ε is the residual vector.

To estimate the model parameters, LASSO minimizes the sum of the squared errors with an upper bound s on the sum of the absolute values of model parameters as:

$$\hat{\beta} = \arg\min_{\beta}\{\|Y - X\beta\|_2^2\} \tag{2}$$

$$\text{subject to } \|\beta\|_1 \leq s \tag{3}$$

where $\|.\|_1$ and $\|.\|_2$ are respectively the norm l_1 and l_2.

This problem is equivalent to:

$$\hat{\beta} = \arg\min_{\beta}\{\|Y - X\beta\|_2^2 + \lambda \|\beta\|_1\}, \lambda \geq 0 \tag{4}$$

where λ is the tuning parameter that controls the strength of the penalty.
The relation between λ and s is a reverse relation: when λ is zero, s is infinity and vice versa (Fonti and Belitser (2017)).

3. METHODOLOGY

The proposed approach for HI extraction is depicted in Fig. 1. It consists of three steps: (1) data unfolding, (2) dimension reduction and (3) HI estimation. These steps are detailed in the following.

3.1 Data Unfolding

Batch processes data are stored in a 3D matrix $\mathbf{X_{3D}}$ of dimension $I \times J \times K$, where I is the number of monitored wafers, J is the number of sensors, and K is the number of observations of each sensor for each wafer.

The 3D matrix is batch-wise unfolded as illustrated in Fig. 2. The obtained matrix $\mathbf{X_{2D}}$ is of dimension $I \times P$ where $P = J \times K$. Each element of this matrix is noted as x_i^p for $i = 1, \ldots, I$ and $p = 1, \ldots, P$.

In this restructuring, each row contains all the measurements within a batch, and each column depicts the evolution of the point (j, k) chronologically, where $j = 1, \ldots, J$ and $k = 1, \ldots, K$.

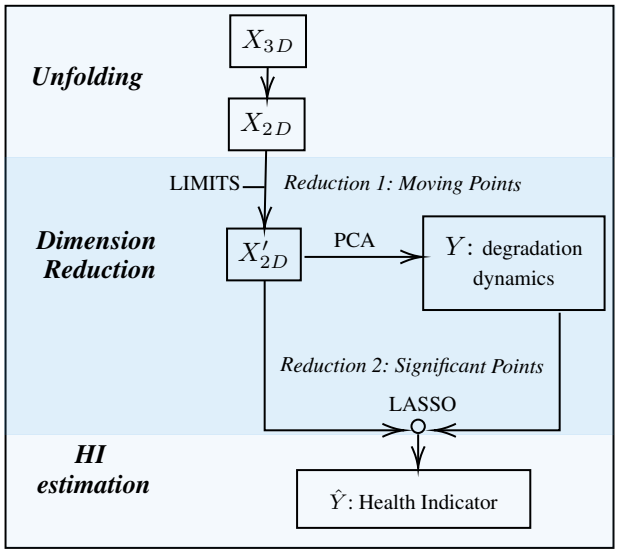

Fig. 1. Schema of the proposed SP-LASSO approach

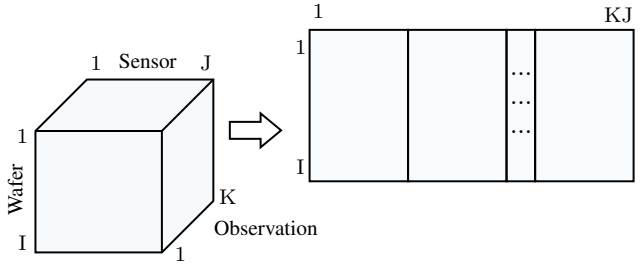

Fig. 2. Batch processes data are illustrated and unfolded

3.2 Dimension reduction

As the number of points increased after unfolding, two reductions are performed. The first one aims to select the points which present variations over time, called in this paper the moving points. These points, with non zero-variability, are identified for the use of Principal Component Analysis (PCA) in the next step. The second reduction intends to select the significant points among the moving ones to be used in estimating the HI.

First Reduction: selection of Moving Points Based on the assumption that the machine deterioration is gradual over time, the first n wafers are assumed to be reliable products and are considered to be respecting the good quality norms. Thus for each point $p \in \{1, \dots P\}$, the limits of its normal functioning (UL stands for the Upper Limit and LL for the Lower Limit) are defined based on the data collected from these first n wafers as:

$$\begin{cases} UL^p = max(x_i^p, i = 1, ..., n) \\ LL^p = min(x_i^p, i = 1, ..., n) \end{cases} \quad (5)$$

The moving points are those for which the measurements of a degraded wafer exceed their limits. Based on the assumption, the last product (the I^{th} wafer) can be used to select these points satisfying the condition:

$$x_I^p > UL^p \text{ or } x_I^p < LL^p \quad (6)$$

The set of moving points $mp = \{p_1, p_2, \dots, p_M\}$ of length M, generates the reduced matrix $\mathbf{X'_{2D}}$:

$$X'_{2D} = \begin{pmatrix} x_1^{p_1} & x_1^{p_2} & \dots & x_1^{p_M} \\ x_2^{p_1} & x_2^{p_2} & \dots & x_2^{p_M} \\ \vdots & \vdots & & \vdots \\ x_I^{p_1} & x_I^{p_2} & \dots & x_I^{p_M} \end{pmatrix} \quad (7)$$

Second Reduction: selection of Significant Points The reduced matrix $\mathbf{X'_{2D}}$ is then mean-centered and unit-deviation scaled. It is decomposed by PCA as:

$$X'_{2D} = T \times P^\top \quad (8)$$

where T and $P^\top$ are respectively the score and the transpose of the loading matrix.

It was proved in Nguyen (2016) that the first principal component (the first column of T) depicted from the decomposition of $\mathbf{X'_{2D}}$ by PCA represents the equipment degradation dynamics. Due to the assumption that the machine deterioration is gradual overtime, its features which are progressively increasing or decreasing should be depicted by the principal components. The principal features of machine over time are: gradual drifts of degradation, abrupt drifts, noises and disturbances; among them, only gradual drifts of degradation and noises always occur on all the products, thus, they are depicted by the first principal component noted as Y.

Since the equipment unit degrades gradually overtime, the FDC data collected during its use might reflect its health state. Thus, the LASSO-based regression model is fitted to the data to select the points (j, k) that reflect the equipment degradation based on modeling the relationship between $\mathbf{X'_{2D}}$ and the equipment degradation dynamics Y.

The model coefficients β are then estimated. The points with non-zero coefficient constitute the significant points.

3.3 Health indicator estimation

After fitting the model to the data and estimating the coefficients $\hat{\beta}$, only the significant points corresponding to non-zero $\hat{\beta}$ intervene in estimating the HI by the constructed model, as:

$$\hat{Y} = \hat{\beta} X'_{2D} \quad (9)$$

4. APPLICATION

The proposed approach is evaluated by numerical experiments on data simulated on MATLAB. The data set corresponds to an equipment unit of the semiconductor manufacturing process. It consists of measurements collected from 6 sensors and 351 observations for a recipe that was performed on 2000 wafers. To demonstrate the effectiveness of our approach, the comparison with the SP-PCA approach of Nguyen et al. (2015b) is performed. This comparison concerns the feature selection and the HI extraction methods which are different for the two approaches.

4.1 Experimental Setup

To generate $\mathbf{X_{3D}}$, a simulated signal illustrating the degradation of the equipment unit is embedded into the data set of the 2000 wafers ($I = 2000$) as explained in the following. An illustration of the measurements of the 6

sensors ($J = 6$) along the production time ($K = 351$ observations) of a wafer are displayed in Fig. 3.

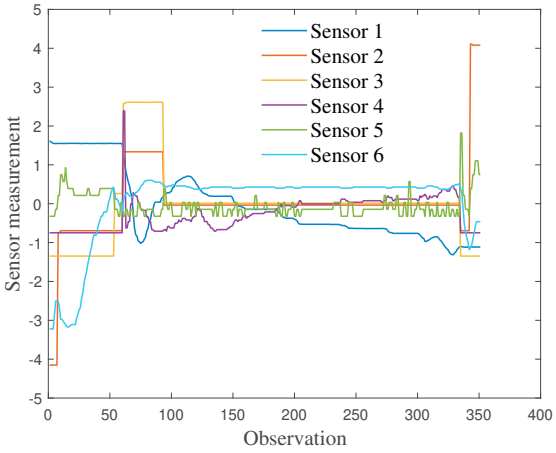

Fig. 3. Sensors measurements of a wafer

The simulation of the degradation signal is based on the assumption of Nguyen et al. (2015a) and Saarela et al. (2014) that the common form of the HIs in real case applications is a progressive trend embedded in noise. Thus, the HI can be considered as the synthesis of three elements: a real degradation state as a monotonic profile, disturbances as step or waves forms and noise. The simulated degradation signal is displayed in Fig. 4.

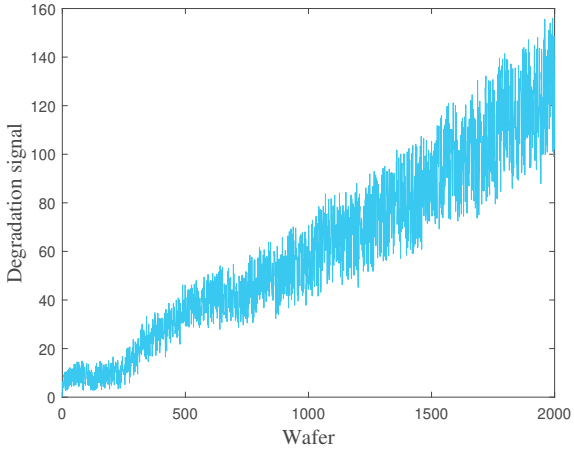

Fig. 4. Degradation signal

The degradation signal is distributed as illustrated in Fig. 5 on some points (j, k) called degraded points among the total P points where $P = J \times K = 2106$. It is remarkable that the degraded points of sensor 1 belong to the observations between 60 and 140. Whereas, the sensor 5 for example does not carry the degradation.

The generated $\mathbf{X_{3D}}$ is then batch-wise unfolded. For the first dimension reduction, the measurements of the first 200 wafers ($n = 200$) are used to define the limits of the normal functioning for each point as in (5). After this step, 595 moving points are identified.

Next, to select the significant points, the LASSO-based regression model is built depending on the λ parameter.

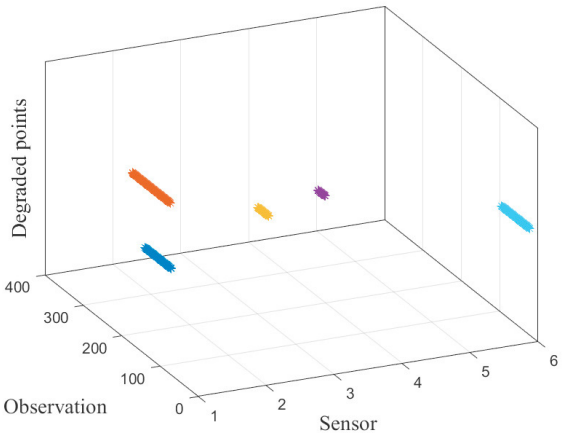

Fig. 5. Positioning of degraded points

The 10-fold cross-validated Mean Squared Error (MSE) are computed for different values of λ as in Fig. 6. For each λ, the original data set is randomly partitioned into 10 equally sized subsets and each one is used once as a hold-out set for testing the model, and the remaining subsets are used for training. The MSE is calculated as the average of the errors on all the hold-out subsets. The λ that minimises the MSE is chosen.

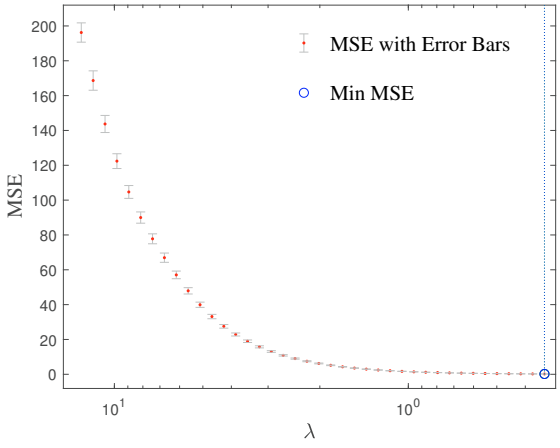

Fig. 6. Cross-validated MSE of LASSO fit

4.2 Experimental Study

Compared methods A comparison between the proposed approach and the SP-PCA approach of Nguyen et al. (2015b) is performed. SP-PCA similarly to SP-LASSO aims to extract a HI from batch processes data. The difference between the two approaches resides in the feature selection and the HI extraction steps. In the SP-PCA, the correlation belonging to the filter type methods, is adopted for the feature selection. It is calculated between the moving points and the degradation dynamics Y extracted from the data as defined previously. A threshold representing the 50^{th} percentile of the set of this correlation allows the selection of the significant points. For the HI extraction, these identified points are rearranged in a new matrix which is decomposed by PCA. The first principle component is considered as the HI.

Results In the following, the results of the comparison between the two approaches are presented graphically and evaluated by numerical metrics.

Fig. 7 illustrates the results of the feature selection methods of the two approaches. Among the P points, only the degraded ones are displayed on the x-axis of this plot. The identified significant points by each of the two selection methods are also presented.

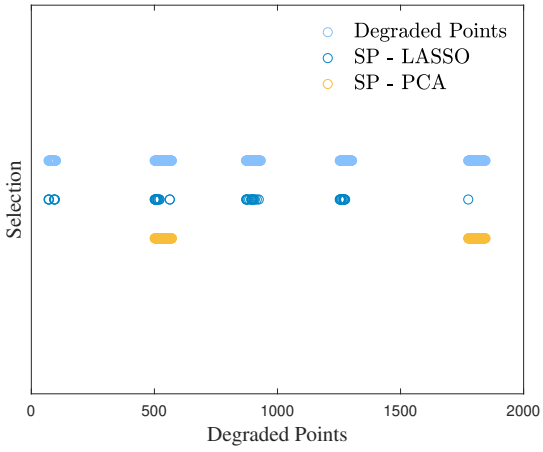

Fig. 7. Significant points selected in SP-LASSO and SP-PCA

First it can be noticed that there are no "False Negative" selections by the two methods. The identified points are subset of the degraded feature set. It is also remarkable that the number of identified points by the correlation is greater than that of LASSO: 142 vs 48 selections. The large difference in the number of selected points is explained by the fact that LASSO introduces sparcity in the model, thus the points that do not add any information to the HI estimation are excluded. Instead, the correlation method used in SP-PCA selects all the points which are correlated with the degradation dynamics up to a predefined threshold.

The significant points identified by each method are used to generate the health indicators. These indicators are normalized and displayed in Fig. 8. One can remark that the HI estimated by SP-LASSO is more accurate as it presents almost the same tendency as the degradation signal which is not the case for the HI extracted by SP-PCA.

The different approaches are evaluated numerically by two metrics detailed in Lei et al. (2018): the Root Mean Squared Error (RMSE) and the consistency metric.

The RMSE is calculated as follows :

$$RMSE(Y_1, Y_2) = \sqrt{\frac{\sum_{i=1}^{I}(Y_{1,i} - Y_{2,i})^2}{I}} \qquad (10)$$

where $Y_{1,i}$ and $Y_{2,i}$ are the values of the two HIs Y_1 and Y_2 for wafer i such as $i = 1, \ldots, I$.

The consistency metric describes the correlation among multiple HIs. Since the different HIs of a single unit contain the information of the same degradation process, they are

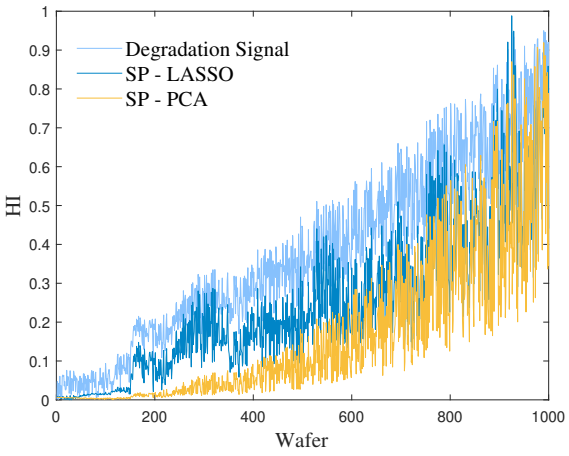

Fig. 8. HIs extracted by SP-LASSO and SP-PCA

supposed to present some kinds of correlation with each other (Liu et al. (2016)):

$$Con(Y_1, Y_2) = \frac{\sum_{i=1}^{I}(Y_{1,i} - \bar{Y}_1)(Y_{2,i} - \bar{Y}_2)}{\sqrt{\sum_{i=1}^{I}(Y_{1,i} - \bar{Y}_1)^2 \sum_{i=1}^{I}(Y_{2,i} - \bar{Y}_2)^2}} \qquad (11)$$

where $\bar{Y}_1$ and $\bar{Y}_2$ are the means of Y_1 and Y_2. The consistency value belongs to the range of $[0, 1]$ where a larger value means a higher similarity between two HIs.

Table 1. Metrics results

Metric	SP-LASSO	SP-PCA
RMSE	0.14	0.25
Consistency	0.96	0.87

The results illustrated in Table 1 indicate that the proposed approach shows improvements over the SP-PCA approach in term of RMSE and consistency: SP-LASSO has lower error and higher consistency.

Discussions The advantage of the proposed approach SP-LASSO compared to SP-PCA is that fewer features are selected to estimate the HI while conserving the information. Thus the SP-LASSO leads to reduce the storage space in databases. In addition, it makes easier the cause-effect analyses since the reduced set contains the features that have an impact on the target output.

Another asset of this approach is the adequacy of LASSO to handle the dimensionality problems in the industrial context where the number of features can exceed the number of manufactured wafers. Indeed, this problem comes from the fact that the number of sensors can easily reach a hundred and the measurements are often sampled at very high frequencies, which exponentially increase the volume of data especially after unfolding.

In addition to these advantages, the proposed approach is simpler in implementation compared to SP-PCA for two reasons: 1) it does not require experts intervension as the λ value is set using cross-validation contrary to SP-PCA method where the percentile is chosen subjectively by an expert and 2) the same model 'LASSO-based model' realises both feature selection and the HI estimation.

5. CONCLUSION

This paper proposes a new approach, the SP-LASSO, for health indicator extraction for batch processes. It deals with the problem of high dimensional data. The SP-LASSO approach is mainly based on the LASSO regression model for performing feature selection and health indicator estimation. The evaluation of the proposed approach carried out on simulated data shows its effectiveness compared to an existing approach.

In the future, we aim to test this approach on real data provided by ST-Microelectronics. We also intend to adapt this method for online applications where the operating conditions, the type of the product or its measurement during the normal functioning can change. Hence, updating the reduced matrices and the model parameters constitute the upcoming work for online HI estimation.

ACKNOWLEDGEMENTS

This paper is conducted in the framework of the project MADEin4, which has received funding from the ECSEL JU (Electronic Components and Systems for European Leadership Joint Undertaking) under grant agreement No 826589. The JU receives support from the European Union's Horizon 2020 research and innovation program and France, Germany, Austria, Italy, Sweden, Netherlands, Belgium, Hungary, Romania and Israel.

REFERENCES

Babajanian Bisheh, H., Ghodrati Amiri, G., Nekooei, M., and Darvishan, E. (2019). Damage detection of a cable-stayed bridge using feature extraction and selection methods. *Structure and Infrastructure Engineering*, 15(9), 1165–1177.

Dong, J., Sun, R., Peng, K., Shi, Z., and Ma, L. (2019). Quality monitoring and root cause diagnosis for industrial processes based on Lasso-SAE-CCA. *IEEE Access*, 7, 90230–90242.

Du, X. (2019). Fault detection using bispectral features and one-class classifiers. *Journal of Process Control*, 83, 1–10.

Fonti, V. and Belitser, E. (2017). Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*, 30, 1–25.

Kim, K.J., Kim, K.J., Jun, C.H., Chong, I.G., and Song, G.Y. (2018). Variable selection under missing values and unlabeled data in semiconductor processes. *IEEE Transactions on Semiconductor Manufacturing*, 32(1), 121–128.

Korabi, T.E., Borodin, V., Michel, J., and Roussy, A. (2021). A hybrid feature selection approach for virtual metrology: Application to CMP process. In *2021 32nd Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 1–5. IEEE.

Lee, D.H., Yang, J.K., Lee, C.H., and Kim, K.J. (2019). A data-driven approach to selection of critical process steps in the semiconductor manufacturing process considering missing and imbalanced data. *Journal of Manufacturing Systems*, 52, 146–156.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., and Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical systems and signal processing*, 104, 799–834.

Liu, Z., Zuo, M.J., and Qin, Y. (2016). Remaining useful life prediction of rolling element bearings based on health state assessment. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 230(2), 314–330.

Malekloo, A., Ozer, E., AlHamaydeh, M., and Girolami, M. (2021). Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Structural Health Monitoring*.

Melhem, M., Ananou, B., Djeziri, M., Ouladsine, M., and Pinaton, J. (2015). Product's quality prediction with respect to equipments data. In *Proceedings of the International Conference on Data Science (ICDATA)*, 17. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Nguyen, T.B.L. (2016). *Approche statistique pour le pronostic de défaillance: application à l'industrie du semi-conducteur*. Ph.D. thesis, Aix-Marseille.

Nguyen, T.B.L., Djeziri, M., Ananou, B., Ouladsine, M., and Pinaton, J. (2015a). Remaining useful life estimation for noisy degradation trends. *IFAC-PapersOnLine*, 48(21), 85–90.

Nguyen, T.B.L., Djeziri, M.A., Ananou, B., Ouladsine, M., and Pinaton, J. (2015b). Health index extraction methods for batch processes in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 28(3), 306–317.

Rostami, H., Blue, J., and Yugma, C. (2018). Automatic equipment fault fingerprint extraction for the fault diagnostic on the batch process data. *Applied Soft Computing*, 68, 972–989.

Rostami, H., Blue, J., Yugma, C., and Pinaton, J. (2017). Equipment health modeling for deterioration prognosis and fault signatures diagnosis. In *2017 28th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 357–362. IEEE.

Saarela, O., Hulsund, J.E., Taipale, A., and Hegle, M. (2014). Remaining useful life estimation for air filters at a nuclear power plant. In *PHM Society European Conference*, volume 2.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

Wang, S., Ji, B., Zhao, J., Liu, W., and Xu, T. (2018). Predicting ship fuel consumption based on LASSO regression. *Transportation Research Part D: Transport and Environment*, 65, 817–824.

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., and Saeed, J. (2020). A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 1(2), 56–70.

Zou, C. and Qiu, P. (2009). Multivariate statistical process control using LASSO. *Journal of the American Statistical Association*, 104(488), 1586–1596.