2022 European Control Conference (ECC)
July 12-15, 2022. London, United Kingdom

LASSO-based Health Indicator Extraction Method for Semiconductor Manufacturing Processes*

Dima EL JAMAL[1], Bouchra ANANOU[1], Guillaume Graton[2], Moustapha OULADSINE[1] and Jacques PINATON[3]

***Abstract*— Over the last few years, with the increasing worldwide competition, semiconductor 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 avoid machine failures and to plan maintenance actions. For that, a novel approach for health indicator extraction named Significant Points combined to the Least Absolute Shrinkage and Selection Operator (SP-LASSO) is proposed in this paper. It deals with the problem of high dimensional data and the specificity of the health indicator in real industrial cases. The proposed method performs feature selection and health indicator extraction and it is mainly based on LASSO. A numerical application on simulated data illustrates the accuracy of this approach.**

***Index Terms*— Health indicator extraction, feature selection, LASSO, semiconductor manufacturing**

I. INTRODUCTION

A semiconductor manufacturing line is made of hundreds of sequential batch processing stages. Each of these stages consists of many steps carried out by expensive equipment units [1] [2]. The failure of these latter can deteriorate the wafer quality and cause a catastrophic loss in the yield of manufacturing. Hence, it is important to monitor these equipment units in order to assure a stable fabrication and a high yield rate. To achieve this goal, the challenge is to extract from the data an appropriate Health Indicator (HI) that illustrates the actual state of the equipment units.

The difficulty of the HI extraction lies firstly in the large volume of data. As the equipment units are monitored with high number of sensors whose measurements are often sampled at very high frequencies [3], large databases are displayed to be exploited, processed and analyzed in order to select the significant features to be used to extract the HI.

The second difficulty resides in the composition of the HI in real industrial cases. Based on [4] and [5], the HIs have a common form: a progressive trend embedded in noise. They can be considered as the synthesis of three elements which are: a real degradation state as a monotonic profile, disturbances as step and waves forms and noise. Hence, the HI extraction methods should illustrate these elements and the form of the HI.

*This work is supported by MADEin4

[1]D. EL JAMAL, B. ANANOU, M. OULADSINE are Aix Marseille Univ, Universite de Toulon, CNRS, LIS (UMR 7020), Avenue Escadrille Normandie-Niemen, F-13397 Marseille Cedex 20, France dima.el-jamal@lis-lab.fr

[2] G. GRATON is with Ecole Centrale Marseille, Technopôle de Château-Gombert, 38 rue Frédéric Joliot-Curie, F-13451 Marseille, France

[3] J. PINATON is STMicroelectronics Rousset, 190 avenue Celestin Coq, ZI - Rousset, F-13106 Rousset, France

To overcome these problems, the Significant Points combined to the Least Absolute Shrinkage and Selection Operator (SP-LASSO) is proposed. It is an automated approach that performs both feature selection and HI extraction based mainly on LASSO.

LASSO is a l_1-penalized regression method [6]. It is used in this work since it is able to perform feature selection and HI prediction simultaneously. In addition, three types of feature selection methods exist in the literature which are: filter, wrapper and embedded [7] [8]. LASSO is part of the embedded type methods. The advantages of this type compared to the filter and the wrapper type methods are respectively higher accuracy and less computing complexity. Finally, LASSO provides good prediction accuracy and helps increasing the model interpretability by eliminating irrelevant features making it appropriate to handle high-dimensional data.

The effectiveness of SP-LASSO is demonstrated through the comparison to an existing approach performed in the same context which is the Significant Points combined to Principal Component Analysis (SP-PCA) developed in [9]. It performs feature selection and HI extraction using respectively a filter-type method and the PCA. The comparison is performed on two different scenarios. The remainder of this paper is structured as follows: the proposed approach is illustrated in section II. A detailed description of the LASSO-based regression model is also given. Section III presents the approach application on simulated data and the comparison with the SP-PCA approach [9] on two different scenarios. Finally, conclusion and perspectives are given in section IV.

II. METHODOLOGY

Batch processes data [9],[10] are stored in a three dimensional (3D) matrix 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 proposed approach for HI extraction is depicted in Fig. 1. It consists of three steps. At first, X_{3D} is unfolded into a two-dimensional matrix noted as X_{2D}. Second, the new structured matrix undergoes dimension reduction to obtain X'_{2D} and to extract the degradation dynamics. Finally, the LASSO-based regression model is applied to estimate the HI. In the following, the LASSO-based model is firstly described then the different parts constituting the approach are detailed.

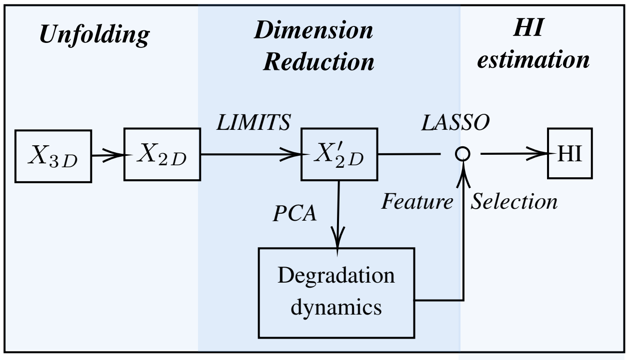


Fig. 1: Schema of the proposed approach

A. LASSO

1) Overview: LASSO (Least Absolute Shrinkage and Selection Operator) was first formulated by Robert Tibshirani in 1996 [11]. It is a powerful method that performs 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 the coefficients to the objective function as a penalty. LASSO encourages sparsity by forcing the coefficients of non-relevant features to be shrunk to zero.

LASSO is widely used in the literature and has proved its effectiveness in different applications on real industrial data. Several works proposed LASSO-based approaches for different purposes as: feature selection [12], prediction [13], [14], [15] and process monitoring [16], [17]. In this paper, it is adopted for health monitoring of batch processes.

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 anupper 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 [6].

In order to find the optimal λ, the LASSO-based regression models are constructed for various values of λ and the chosen value is the one having the minimal cross-validated mean squared error.

B. Proposed Approach

The approach three parts are described in the following.

1) Data Unfolding: The 3D matrix X_{3D} is batch-wise unfolded as illustrated in Fig. 2. The obtained matrix X_{2D} is of dimension $I \times P$ where $P = K \times J$. Each element of this matrix is noted as x_i^p for $i = 1, \ldots, I$ and $p = 1, \ldots, P$. The rows of X_{2D} contain the measurements within a batch and the columns depict the chronological evolution of the features corresponding to the points (j, k) of sensors and observations, where $j = 1, \ldots, J$ and $k = 1, \ldots, K$.

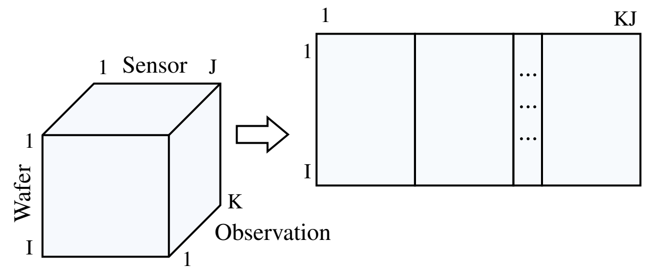


Fig. 2: Batch process data is illustrated and unfolded

2) Dimension reduction: As the number of points increased after unfolding, two reductions are performed. The first one aims to select the moving points which present variations over time. These points, with non zero-variability, are identified for the use of 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.

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 good quality norms. Thus, the data collected from these wafers are used to define the limits of the normal functioning for each point $p \in \{1, \ldots P\}$ as:

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

where UL^p and LL^p stand respectively for the Upper and the Lower Limits of the point p.

In the first reduction, the moving points are those for which the measurements of a degraded wafer exceed the limits. If an arbitrary wafer N ($n < N < I$) verified as bad wafer by a measured test is available, it is used in order to select the moving points satisfying the condition:

$$x_N^p > UL^p \text{ or } x_N^p < LL^p \tag{6}$$

Otherwise, the last product (the I^{th} wafer) is used instead to select the moving points verifying:

$$x_I^p > UL^p \text{ or } x_I^p < LL^p \tag{7}$$

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

$$X'_{2D} = \begin{pmatrix} x_1^{p_1} & x_1^{p_2} & \dots & x_1^{p_M} \\ x_2^{p_1} & x_2^{p_2} & \dots & x_2^{p_M} \\ \vdots & \vdots & & \vdots \\ x_I^{p_1} & x_I^{p_2} & \dots & x_I^{p_M} \end{pmatrix} \tag{8}$$

For the second reduction, the reduced matrix X'_{2D} is mean-centered and unit-deviation scaled. It is decomposed by PCA as:

$$X'_{2D} = T \times P^{\top} \tag{9}$$

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

It was proved in [18] that the first principal component (the first column of T) depicted from the decomposition of X'_{2D} by PCA illustrates the equipment degradation dynamics. Based on 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 noise always occur on all the products, thus they are depicted by the first principal component noted as Y.

The LASSO-based regression model is fitted to the data in order to select the significant points that reflect the equipment degradation based on modeling the relationship between X'_{2D} and the equipment degradation dynamics Y. The model coefficients β are then estimated. The points with coefficient equal to zero are excluded from the model whereas, the points with non-zero coefficient are only selected and are called the significant points.

C. Health indicator estimation

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

$$HI : \hat{Y} = \hat{\beta} X'_{2D} \tag{10}$$

III. APPLICATION

The proposed approach is tested 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.

The proposed approach is evaluated on two different scenarios and compared to the SP-PCA approach detailed in [9]. The first scenario aims to illustrate the benefits provided by our approach compared to [9]. The second one intends to study the effectiveness and the adaptability of the proposed approach to real industrial applications.

A. Experimental setup

To generate the data set, the signal which represents the degradation of the equipment due to the manufacturing of the wafers is first built. This signal is simulated as a synthesis of three elements [9]: (1) a monotonic profile representing the degradation state of the equipment, (2) noise and (3) disturbances as step and waves forms as illustrated in Fig. 3. It is remarkable that the generated signal is not strictly monotonous as it illustrates real cases applications where engineers interventions can occur in order to perform small repairs.

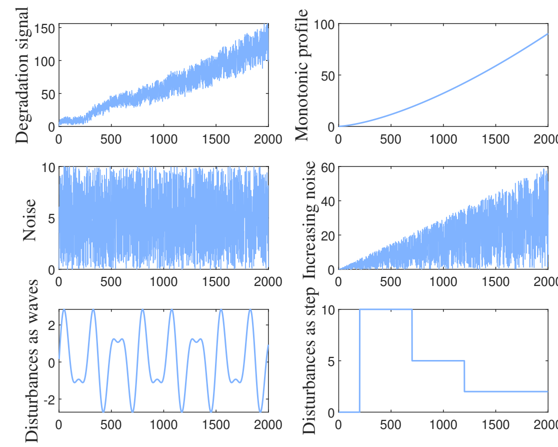


Fig. 3: Degradation signal and the signal and its components

The signal is then embedded into the data set of the 2000 wafers ($I = 2000$) in two different ways as explained further in the two scenarios. The aim of the SP-LASSO approach is to extract this signal supposed unknown in real applications. An illustration of the measurements of the 6 sensors ($J = 6$) along the production time ($K = 351$ observations) for non degraded wafers are displayed in Fig. 4.

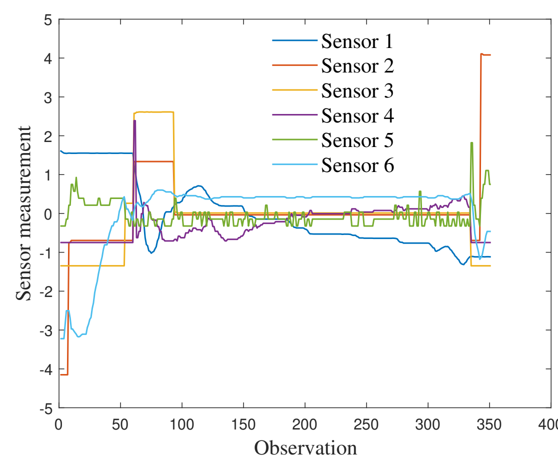


Fig. 4: Example of sensors measurements of a wafer

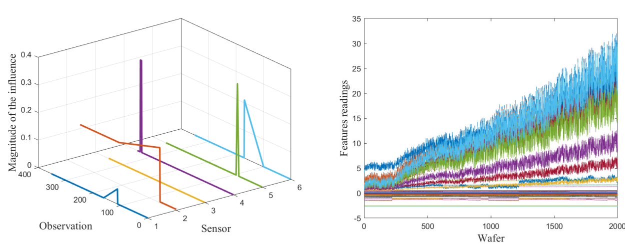
B. Experimental study

In the following, the comparison between the proposed approach and the SP-PCA is illustrated on two different scenarios. This comparison concerns the feature selection and the HI extraction methods which are different for the two approaches. In the SP-PCA approach, the correlation, belonging to a filter-type method, was adopted for the feature selection. It was calculated between the moving points and the degradation dynamics (Y) extracted from the data as in II-B.2. A threshold defined as the 50^{th} percentile of the set of this correlation allows the selection of the significant points. Next for the HI extraction, these identified points were rearranged in a new matrix X'_{2D} which was decomposed by PCA. The first principle component was considered as the HI.

1) Scenario 1:

a) Data generation: The generation of the three dimensional matrix X_{3D} is performed as in [9]. The degradation signal is embedded into the data set according to the amplitude of the degradation defined for each sensor and observation as illustrated in Fig. 5a. As it might be seen, not all the sensors carry the degradation (e.g. sensor 3). It is distributed on some features called degraded points among the total P points where $P = J \times K = 2106$.

The generated X_{3D} is then batch-wise unfolded. The readings of the $P = 2106$ points of the obtained X_{2D} matrix are plotted over the 2000 wafers as in Fig. 5b. We can see that the readings of some points are unchanged over the production. They reflect the points with zero amplitude of degradation in Fig.5a. Unlike the others points whose readings present exactly the same trend as the degradation signal.



(a) Relative amplitude of degradation carried by sensors and observations (b) Reading of the P points after unfolding

Fig. 5: Data generation in Scenario 1

In this application, n is equal to 200 and λ is chosen using 10-fold cross-validation.

b) Results and discussion: Fig. 6 illustrates the results of the features selection methods of the two approaches.

Among the total 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. First, there is no "False Negative" selections by the two methods as the identified points are subsets of the degraded feature set. It is also remarkable that the number of identified points by the correlation is greater than that of LASSO: 223 vs 114 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 adopted in SP-PCA selects all the points that are correlated with the degradation dynamics up to a defined threshold.

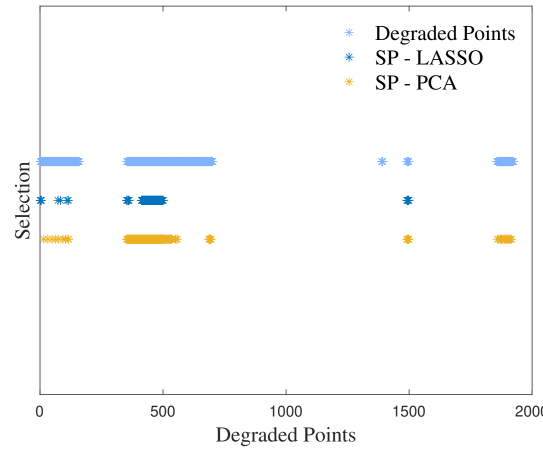


Fig. 6: Significant points selected in SP-LASSO and SP-PCA

The identified significant points are then used to generate the health indicators.

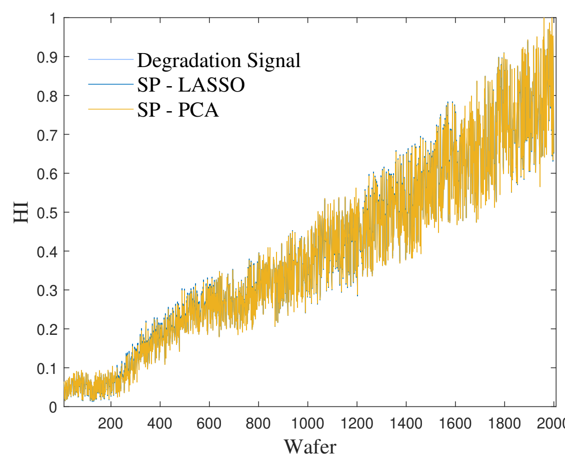


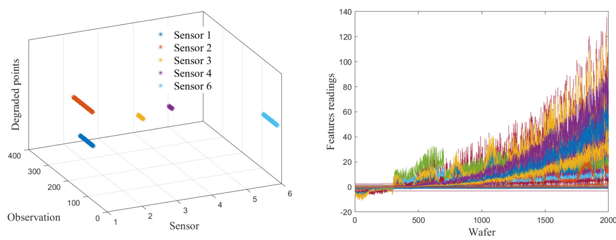
Fig. 7: HIs extracted by SP-LASSO and SP-PCA

In Fig. 7 the normalized HI estimated by the SP-LASSO is presented as well as the HI extracted by the SP-PCA approach. These two indicators illustrate well the equipment deterioration as they are both superposed to the degradation signal.

Therefore the advantage of the proposed approach compared to the 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 only the features that have an impact on the target output.

2) Scenario 2:

a) Data generation: The degradation signal is embedded into the data in such a way to present a real case application. For that, contrary to the first scenario, the degradation is not carried similarly by all the degraded points. This is based on the fact that the sensors do not provide the same information: their measurements depend on their type and objective. Thus, it is assumed that each sensor carries in its measurements one of the HI elements as illustrated in Fig. 8a. The noise is embedded into the readings of sensor 1, the increasing noise (variability) in sensor 2, the disturbance as waves forms (trigonometric forms) in sensor 3, the disturbance as step forms in sensor 5 and the monotonic profile illustrating the degradation in sensor 6.



(a) Positioning of degraded points (b) Reading of the P points after unfolding

Fig. 8: Data generation in Scenario 2

After unfolding X_{3D}, the readings of each point are displayed in Fig. 8b. It is notable that the trend of some degraded points is not similar to that of the degradation signal as in the first scenario. It is explained by the fact that each sensor illustrates a specific information.

b) Results and discussion: The result of the feature selection methods are illustrated in Fig. 9.

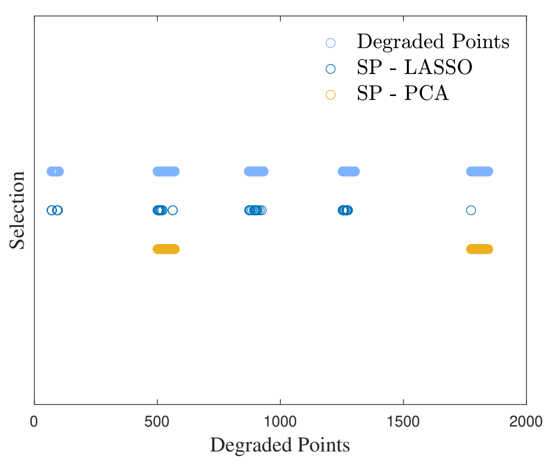


Fig. 9: Significant points selected in SP-LASSO and SP-PCA

As in the first scenario, the proposed approach SP-LASSO selects fewer points: 48 vs 142. This is explained by the fact that LASSO tends to select few features from a group of correlated features (carrying the same HI element) contrary to the correlation where all the features verifying a given characteristic are selected up to a defined threshold. In addition, LASSO performs heterogeneous selections. Its identified significant points cover all signal elements: from each sensor, few degraded points are selected. Unlike the correlation where only the points having an increasing tendency are identified, the points belonging to the sensors 2 and 4 are selected where the increasing variability and the monotonic profile are measured.

The two extracted HI by SP-LASSO and SP-PCA are shown in Fig. 10. Even though the two approaches did not reproduce exactly the same HI in this scenario, the HI estimated by SP-LASSO present almost the same tendency as the degraded signal which is not the case for the HI extracted by SP-PCA.

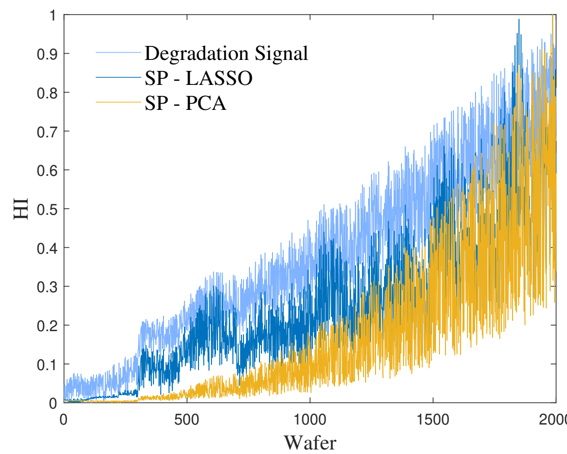


Fig. 10: HIs extracted by SP-LASSO and SP-PCA

The different approaches are also evaluated numerically by the Root Mean Squared Error (RMSE) and the consistency metric computed between the extracted HI and the degradation signal.

- The RMSE is calculated as follows:

$$RMSE(Y_1, Y_2) = \sqrt{\frac{\sum_{i=1}^{I}(Y_{1,i} - Y_{2,i})^2}{I}} \tag{11}$$

  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:

$$Con(Y_1, Y_2) = \frac{\sum_{i=1}^{I}(Y_{1,i} - \bar{Y}_1)(Y_{2,i} - \bar{Y}_2)}{\sqrt{\sum_{i=1}^{I}(Y_{1,i} - \bar{Y}_1)^2 \sum_{i=1}^{I}(Y_{2,i} - \bar{Y}_2)^2}} \tag{12}$$

  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.

The proposed approach presents smaller error: 0.152 vs 0.2469 and higher consistency: 0.96 vs 0.87. Thus, the proposed approach is more suitable for real applications where the degradation information is not distributed uniformly in data. The SP-LASSO performs more accurate estimation of HI compared to SP-PCA. It performs heterogeneous selection of degraded features that illustrate the equipment health state.

C. Discussions

It was proved in the two above scenarios that the proposed approach outperforms the SP-PCA in both feature selection and HI extraction. Another asset of the SP-LASSO is the adequacy of LASSO to handle 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 simple in implementation and it does not require expert intervension as the λ value is set using cross-validation contrary to SP-PCA method where the percentile is chosen subjectively by expert.

IV. CONCLUSIONS

This paper proposes the SP-LASSO approach for health indicator extraction for semiconductor manufacturing processes. It is based on the LASSO regression model for feature selection and health indicator estimation. The evaluation of the proposed approach, carried out on two scenarios, shows its effectiveness compared to an existing approach.

In the future, we aim 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.

ACKNOWLEDGMENT

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

[1] S.-P. Lee, A.-K. Chao, F. Tsung, D. S. H. Wong, S.-T. Tseng, and S.-S. Jang, "Monitoring batch processes with multiple on–off steps in semiconductor manufacturing," *Journal of Quality Technology*, vol. 43, no. 2, pp. 142–157, 2011.

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

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

[4] T.-B.-L. Nguyen, M. Djeziri, B. Ananou, M. Ouladsine, and J. Pinaton, "Remaining useful life estimation for noisy degradation trends," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 85–90, 2015.

[5] O. Saarela, J. E. Hulsund, A. Taipale, and M. Hegle, "Remaining useful life estimation for air filters at a nuclear power plant," in *PHM Society European Conference*, vol. 2, no. 1, 2014.

[6] V. Fonti and E. Belitser, "Feature selection using lasso," *VU Amsterdam Research Paper in Business Analytics*, vol. 30, pp. 1–25, 2017.

[7] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 56–70, 2020.

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

[9] T.-B.-L. Nguyen, M. A. Djeziri, B. Ananou, M. Ouladsine, and J. Pinaton, "Health index extraction methods for batch processes in semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 28, no. 3, pp. 306–317, 2015.

[10] H. Rostami, J. Blue, and C. Yugma, "Automatic equipment fault fingerprint extraction for the fault diagnostic on the batch process data," *Applied Soft Computing*, vol. 68, pp. 972–989, 2018.

[11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[12] K.-J. Kim, K.-J. Kim, C.-H. Jun, I.-G. Chong, and G.-Y. Song, "Variable selection under missing values and unlabeled data in semiconductor processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 32, no. 1, pp. 121–128, 2018.

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

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

[15] X. Chen, G. Jin, S. Qiu, M. Lu, and D. Yu, "Direct remaining useful life estimation based on random forest regression," in *2020 Global Reliability and Prognostics and Health Management (PHM-Shanghai)*. IEEE, 2020, pp. 1–7.

[16] C. Zou and P. Qiu, "Multivariate statistical process control using LASSO," *Journal of the American Statistical Association*, vol. 104, no. 488, pp. 1586–1596, 2009.

[17] J. Dong, R. Sun, K. Peng, Z. Shi, and L. Ma, "Quality monitoring and root cause diagnosis for industrial processes based on Lasso-SAE-CCA," *IEEE Access*, vol. 7, pp. 90 230–90 242, 2019.

[18] T. B. L. Nguyen, "Approche statistique pour le pronostic de défaillance: application à l'industrie du semi-conducteur," Ph.D. dissertation, Aix-Marseille, 2016.