A Modeling Tool for Equipment Health Estimation Using a System Identification Approach

Alexander Dementjev, Ilkay Wunderlich, Klaus Kabitzsch
Department of Computer Science, Technical University of Dresden
01062 Dresden, Germany
Email: *firstname.lastname*@tu-dresden.de

Germar Schneider
Infineon Technologies Dresden GmbH
01099 Dresden, Germany
Email: germar.schneider@infineon.com

Abstract —**In high-technology manufacturing industries like pharmaceuticals, semiconductors or photovoltaics, best and stable yields are very important. This can only be reached if the appropriate equipment is still operating in the specified working ranges according to the current recipes or process steps. Because of the high complexity of the corresponding equipment and processes, monitoring solutions like fault detection and classification or predictive maintenance are already established. The combination of such techniques with advanced process control leads to a mechanism for avoiding product scraps and, finally, to the maximization of the production efficiency and business competitiveness. The equipment health factor is an important index of the status of a processing equipment and at the same time a key enabler for advanced monitoring and control strategies. Three families of estimation techniques will be briefly explained in this paper: physics-inspired, statistical-based and data-driven. Then, the focus will be set on the system identification-based approach as the most promising and emerging equipment health estimation strategy. Its backgrounds and advantages will be explained and illustrated. At the end, a universal tool for investigation and modeling of the equipment health factor will be described and discussed.**

Keywords* —*Equipment Health; Data-Driven Modeling; System Identification; Predictive Maintenance; Time Series Analysis.

I. Introduction

The Equipment Health Factor (EHF) (also known as equipment health index [1]) is a quantitative factor of the status of a processing equipment or a tool, which can be estimated from observable equipment parameters, e.g., on the basis of the history data of the process.

For a sustainable application, the EHF calculation should also be supported by the use of available metrology data (including virtual metrology data) and maintenance information. So, the appropriate data integration of different information sources like process control system, metrology or maintenance is essential for the overall production success. In the literature, there are other similar equipment health definitions under terms like “equipment condition”, “tool health” or “machine health condition”.

In this article, we focus on the following EHF application fields:

- **Predictive Maintenance** [2] is based on condition monitoring [3]. With Predictive Maintenance (PdM), maintenance actions are performed only if needed and offer cost savings compared to time-based Preventive Maintenance (PvM) [4]. Another advantage towards to PvM is to avoid equipment breakdown by performing proactive maintenance. EHF can be used as a main indicator to predict equipment status.
- **Dynamic Sampling** avoids expensive measuring operations on lots of products (e.g. wafers) without increasing the “material at risk” in production [5]. In case of dynamic sampling, EHF can supply the sampling system with information about current equipment condition for avoiding the metrology actions [6].
- **Equipment Prioritization and Production Scheduling** by giving each equipment a certain priority according to its current EHF. If critical tasks with a high quality demand occurs, the equipment with the highest EHF is recommended to be used [7]. This technique motivates to include the EHF into global production scheduling [1].

In the following, an overview of the already existing methods for equipment health estimation will be given (Section II). Then, a system identification-based approach will be introduced and its advantages underlined in Section III. An appropriate software tool for investigation of the identification-based EHF estimation will be presented in Section IV. Achieved results will be discussed in Section V and, in the end, a short conclusion with an outlook will be given.

II. Methods for Equipment Health Estimation

In this section several families of EHF techniques, which can be classified as follows, are discussed:

- **Physics inspired:** These methodologies are considered to be the most complex to transfer between different equipment types, as they entirely rely on what is known (and can be modeled) of the equipment’s behavior [8]. This family of techniques usually consists of very ad-hoc approaches to tackle a specific use case and is therefore the most expensive [9].
- **Statistical or forecast-based:** In this case, relevant time series coming from the target equipment are predicted by exploiting their statistical properties. Their probabilistic future outcome is compared with a preset failure threshold [10]. Notable examples of problems easily solved with this kind of technique include helium flow prediction for edge ring consumption by plasma etching in semiconductor production [11] and health monitoring of electronics under shock loads in packaging and manufacturing [12]. While such techniques can be extremely

 ISBN: 978-1-61208-681-1

powerful in targeting specific problems, it must be said that they lack generality and most of the time cannot be directly transferred between equipment classes and problems.

- **Data-driven:** in contrast to the aforementioned techniques, this class of methods allows quick transfer of methodology between different equipment and processes [13]. The appropriate EHF models will be estimated from the input data using supervised machine learning algorithms like linear or nonlinear regression [14], neural networks [15] or system identification algorithms with gradient-based optimization. Also, classification methods (support vector machines, decision trees, etc.) can be used for equipment health assessment.

  The required training, validation and test data sets could be obtained in a semi-automatic way (for instance by exploiting associated metrology information) as well as in a manual manner, by manually selecting a subset of important parameters to be controlled based on the actual knowledge of the process engineer [16]. It should be noted that such methodologies still require some expert knowledge in order to provide proper results and minimize the false positive rate.

Depending on the use case, each of the approaches may prove itself useful, but no tool is capable for solving every issue in the whole "EHF galaxy" [17]. For this reason, since the beginning of the planning stage, much emphasis has been put on the modularity of the system in order to be able to accommodate a wide variety of mathematical techniques with minimal effort.

III. EHF Estimation Using a System Identification Approach

Most popular statistical-based and data-driven EHF estimation methods presume that changes in the process data can be adequately described with standard statistical functions like variance, mean value, standard deviation or linear regression models. But this limits their applicability to narrow process windows. System identification methods do not rely on that steady-state assumption and characterize processes based on identified physical properties, especially on typical responses to well-defined step changes.

Many equipment signal charts represent in fact responses after control actions like "Heater On", "RF Power On" or "N_2 Primary Valve On" with characteristic trends in according to sensor readings [18]. These trends can be compared with each other or with trends from a reference model to detect process anomalies or, indirectly, decreasing of the equipment health.

By using of a system identification approach, an automatic identification of the reference models should be performed for each existing process context (e.g. different recipes) to describe the characteristic trends in signals not only in steady-state mode. Then, based on an identified reference model, the reference time series should be calculated and compared to the new measurements from the same production context. In the last step, well-known metrics like Mean Squared Error (MSE) can be used to characterize the current equipment health. Finally, the calculated MSE values will be applied to an already existing Fault Detection and Classification (FDC) or EHF software to track the estimated EHF parts.

The proposed common approach consists of 5 steps, as shown Figure 1. The first three steps are performed offline during EHF system deployment:

1) Selection of relevant process variables,
2) Preliminary estimation of the model parameters for the reference time series and
3) Parameter estimation for different model structures.

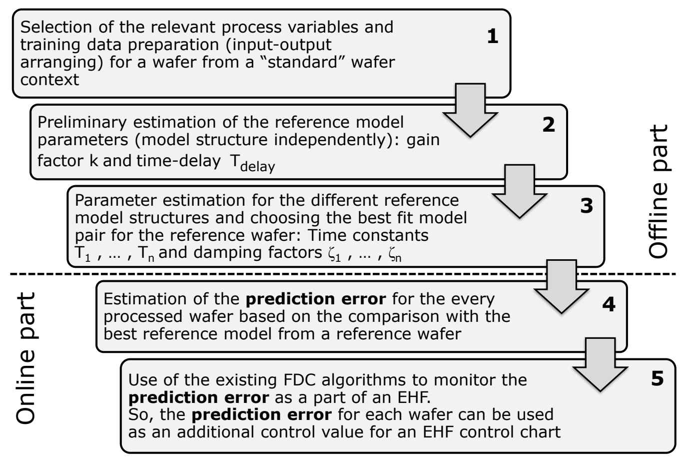

Figure 1. Steps of the proposed approach at a wafer manufacturing example divided into offline and online part.

The system identification based EHF estimation occurs in the two remaining steps which will be executed during an online operation of the EHF modeling tool:

4) EHF online estimation on the basis of model identification on process data and
5) Use of the existing FDC algorithms to monitor and control the EHF (like control charts etc.).

The fourth step can alternatively be performed by identification of the model parameters for each measured time series. In this case the EHF metrics are the differences in the model parameters form different modeled time series according to the reference model.

Figure 2 shows the most popular structures for reference models.

In Figure 3, an example of a reference time series generated by a dynamic second order reference model is given. Figure 4 shows the identified time shift in time series caused by equipment malfunction, respectively.

For the second order model, the estimated model parameters are time delay T_{delay}, gain factor k and two time constants T_1 and T_2. Such models can adequately describe relatively slow processes like heating. Note that for successful and meaningful identification, the measurement frequency must

Copyright (c) IARIA, 2018. ISBN: 978-1-61208-681-1

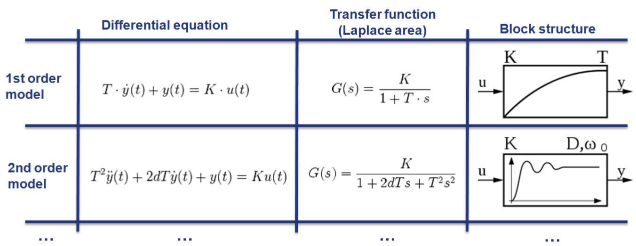

Figure 2. Most popular reference model structures.

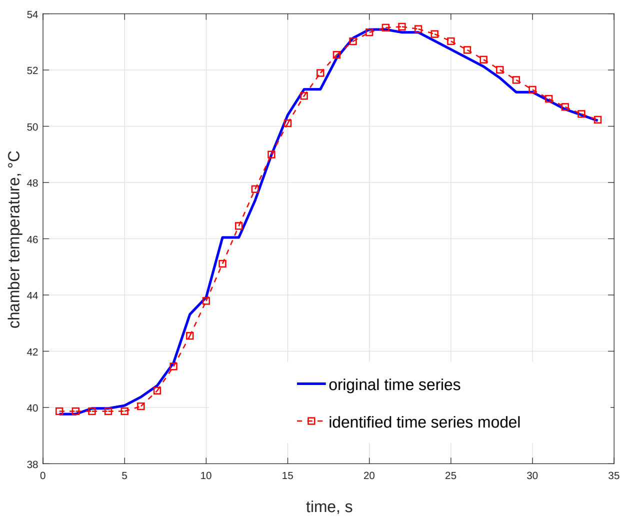

Figure 3. An example of a reference time series.

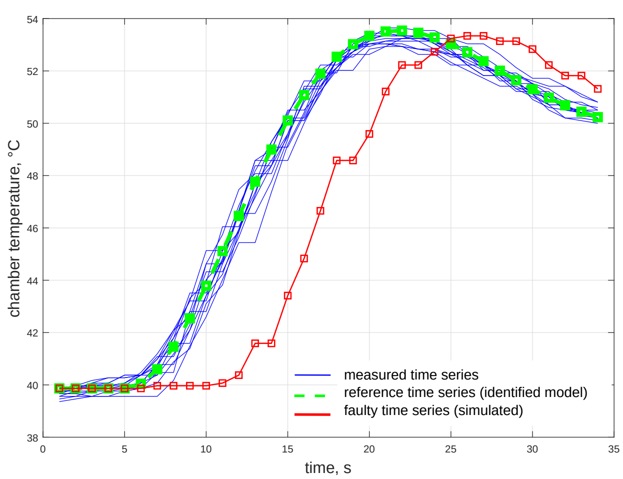

Figure 4. An example for the identified time shift in time series which can be caused by equipment malfunction.

meet the sampling theorem.

Among the common approach (Figure 1), the model identification occurs only on process data from the reference product (e.g., wafer etc.), then generate the reference time series and calculate the MSE on the residuals for the next processed products. In this case the monitored EHF is the MSE.

IV. A Tool for Investigation of the System Identification-Based Equipment Health Estimation

The following section introduces a tool for EHF estimation based on system identification. The tool is programmed in MATLAB using the MATLAB GUIDE environment for the Graphical User Interface (GUI).

A. Structure of the Tool

The EHF estimation tool consists of five GUI panels. The functionality of the first two panels is shortly described in the following:

- **Visualization**: In the first panel the time series can be loaded and displayed. (Ideally the corresponding time series are collected in a data set beforehand)
- **Data Preprocessing**: In case of inhom6ogeneities, e.g., dead time delays, of certain time series can be processed by changing start or end time.

B. Model Identification

Figure 5 shows the third panel after the model identification process is completed. At the top of the figure the settings area can be seen. The following preferences can be changed:

- Selection of the appropriate **model type** (e.g., second order model)
- Choice of the **reference time series** corresponding to the specified course of the process.
- Pick the (virtual) **input signal** of your system (either positive/negative unit step or positive/negative impulse).

The model identification is performed by the MATLAB function for Prediction Error Estimate for Linear and Nonlinear Model (PEM). After identifying the reference model, which is denoted with $m_{ref}(k)$, the estimated model parameters and the fitting quality (in percent) of the model (Figure 5 right) are displayed. In the diagram a comparison of the model and the reference time series is plotted.

C. Reference Based Approach

After finishing the model identification step, one can continue with the reference based approach, which is the implementation of the EHF algorithm introduced Section III. In order to suppress noise components of the signals the model of the reference time series is used to calculate the MSE.

$$MSE(i) = \sum_{k=1}^{K} (x_i(k) - m_{ref}(k))^2 \qquad (1)$$

with i corresponding to the i^{th} time series of the data set and K the length of the each time series $x(k) = [x_1, x_2, ...x_K]$.

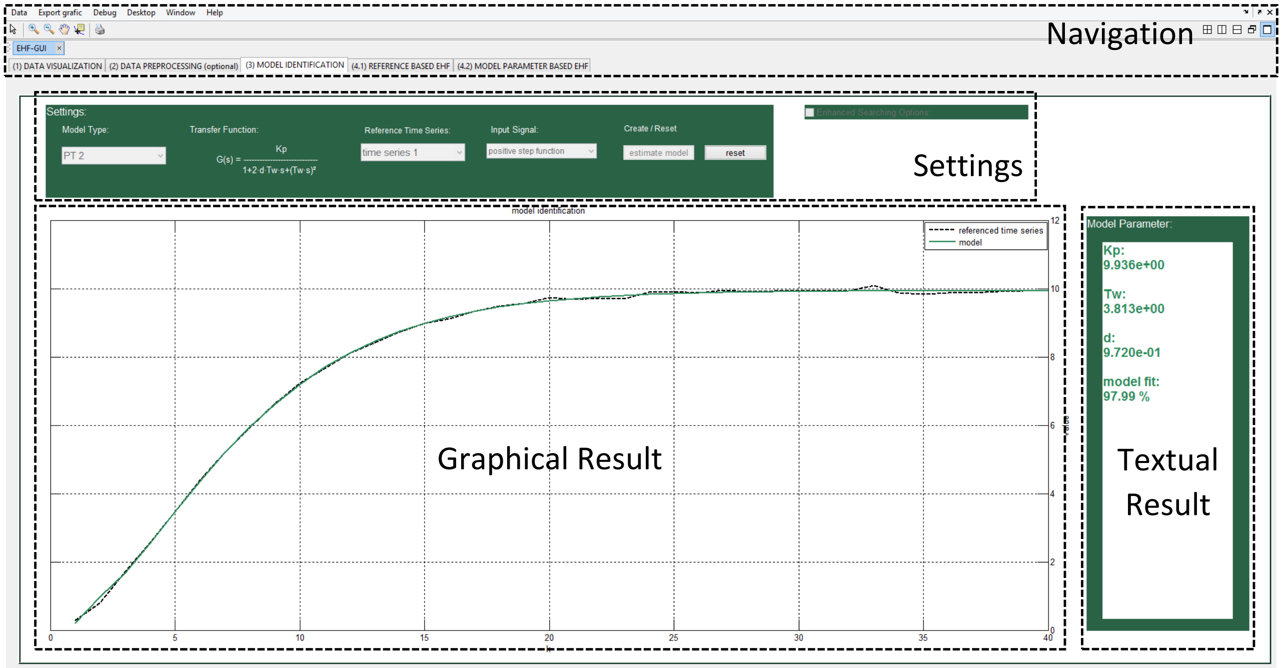

Figure 5. Panel 3 - Model Identification: At the top of the GUI the menu, tool bar and the panel navigation are located. In the middle, the settings for the model identification can be changed. The diagram and the text box at the bottom show the results of the model identification.

Additionally it is possible to set an alert limit in multiples of the standard deviation of the MSE σ_{MSE} (Figure 8, dashed line).

D. Model Parameter Based Approach

The model parameter based approach was introduced in Section III. In order to obtain a normalized EHF between 0% and 100% the identified parameter of the reference model are defined as 100%.

The parameters of the remaining time series are estimated with the same settings which are chosen for the reference model in the third panel. The calculation of the EHF will be introduced in the following steps:

1) Estimation of the reference model parameters $P_{j,ref}$ with $j = 1, \ldots, J$ different parameters (excluding the fitting percentage)
2) Estimation of the parameters of all time series $P_j(i)$ (i corresponding to the i^{th} time series)
3) Calculation of the absolute deviation between each parameter and the corresponding reference parameter:

$$\Delta_j(i) = \frac{|P_j(i) - P_{j,ref}|}{P_{j,ref}} \tag{2}$$

Note that the absolute deviation at the index $i = ref$ (reference time series) is always zero and therefore does not need to be calculated:

$$\Delta_j(i = ref) = \frac{|P_j(i = ref) - P_{j,ref}|}{P_{j,ref}} \equiv 0, \ \forall j$$

4) Summation of the absolute deviations:

$$\Delta_{total}(i) = \frac{1}{J}\sum_{j=1}^{J} \Delta_j(i) \tag{3}$$

5) Mapping of the total deviation to the EHF:

$$EHF(i) = 100\% \cdot (1 - \Delta_{total}(i)) \tag{4}$$

(Negative values will be mapped to 0%.)

E. Validation of the Identification Procedure

In order to validate the estimation process and to further investigate on noise tolerance, data sets are generated with MATLAB SIMULINK and analyzed by the tool.

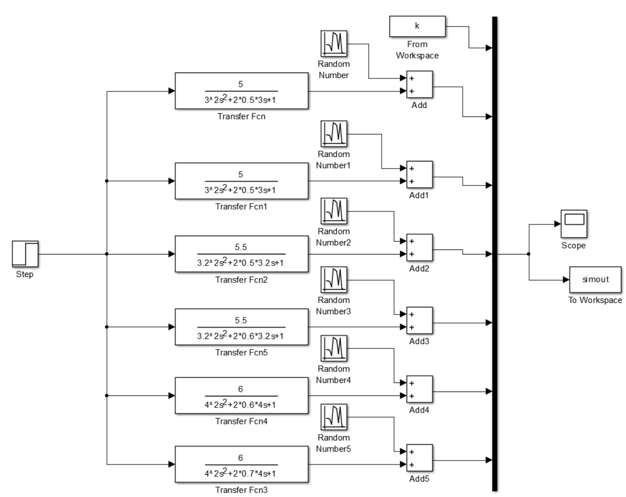

Figure 6. SIMULINK structure for test data generation.

 ISBN: 978-1-61208-681-1

One example data set, which is generated by the structure shown in Figure 6, consists of 6 time series created by a second order system with varying parameters and a unit step as input-signal. These time series are appended with an additive white Gaussian noise component.

The estimated model parameters of the time series are displayed in Table I. Deviations from the generated second order parameters are denoted in brackets. With this comparison it is shown, that the identification procedure still works for time series with additive white Gaussian noise components.

TABLE I. ESTIMATED MODEL PARAMETERS AND THEIR DEVIATION FROM THE GENERATED SECOND ORDER PARAMETERS AT A SIGNAL TO NOISE RATION OF $SNR \approx 27dB$.

time series	1	2	3
K_p	5 (+0)	5 (+0)	5.5 (+0)
T_w	2.9 (−0.1)	3.1 (+0.1)	3.5 (+0.3)
d	0.48 (−0.02)	0.48 (−0.02)	0.51 (+0.01)
fitting:	74.4%	74.8%	76.8%

time series	4	5	6
K_p	5.5 (+0)	6 (+0)	6 (+0)
T_w	3.1 (−0.1)	3.8 (−0.1)	3.7 (−0.3)
d	0.52 (−0.08)	0.61 (+0.01)	0.69 (−0.01)
fitting:	78.4%	81.6%	83.1%

F. Application Example

The following generated data set consists of 15 generated time series, which can be seen in Figure 7.

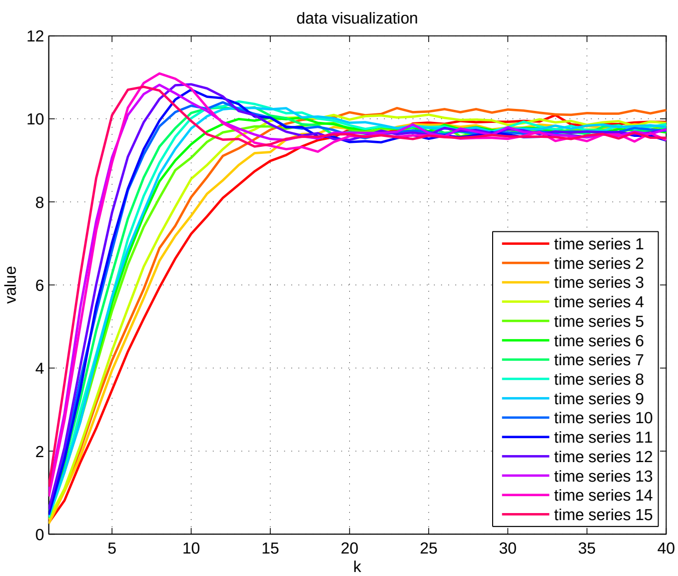

Figure 7. Visualisation of the data set, consisting of 15 time series.

The first time series is selected as the reference time series and is modeled with a second order model type. In Figure 8 the result of the reference based EHF estimation is shown. One can see that the time series 13, 14 and 15 exceed the deviation of $2.5{\cdot}\sigma_{MSE}$ (highlighted in red). Figure 9 illustrates the model parameter based approach, which is calculated with the given in Section IV-D. The reference time series, which is highlighted with a green bar, corresponds to an EHF of 100%.

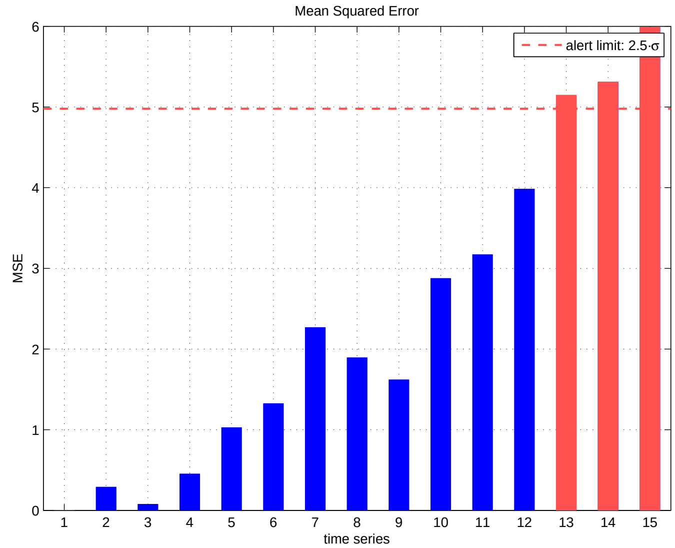

Figure 8. Reference based EHF estimation.

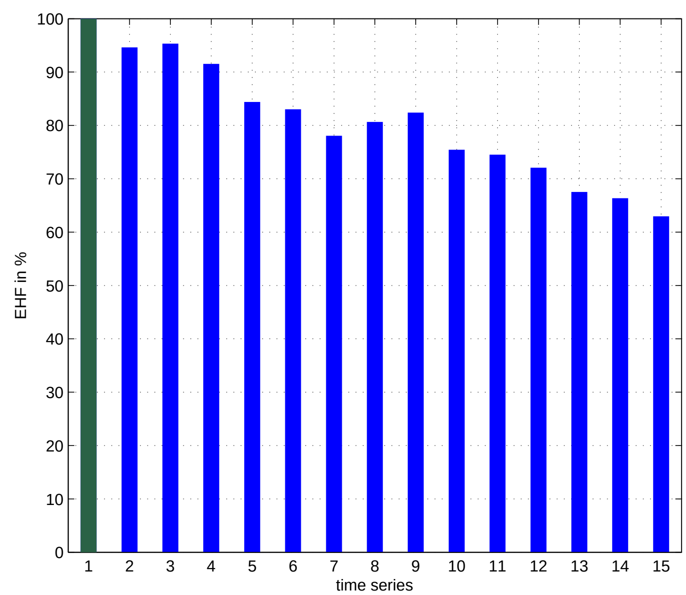

Figure 9. Model parameter based EHF estimation.

It is observable, that the critical time series from the reference based EHF estimation (Figure 8) are corresponding to an EHF below 70%.

V. RESULTS AND DISCUSSION

This section compares the approaches reference based and model parameter based with respect to calculation time and error prediction.

A. Calculation Time

Compared to the the PEM-function the other arithmetic operations such as MSE or absolute deviations are negligible.

 ISBN: 978-1-61208-681-1

With this negligence the two algorithms can be evaluated with respect to the calculation time as follows:

- **Reference based approach:**
  This approach needs one execution of the PEM algorithm for determining the reference model
- **Model parameter based approach:**
  The PEM algorithm has to be executed for each time series in the data set.

⇒ The reference based approach has lower calculation time, which can be critical if there are limited calculation capacities or real-time demands.

B. Error Prediction

- **Reference based approach:**
  With the MSE, high deviations between reference model and time series can be detected. However, the cause of the high deviations cannot be analyzed any further.
- **Model parameter based approach:**
  It is possible to detect which individual parameter is causing the deviation from the reference parameters and as a result the low EHF. This information can be helpful for more precise maintenance work.

⇒ The model parameter based approach provides more information for predicting the equipment health decrease.

VI. Conclusion and Future Work

A tool for the system identification-based equipment health modeling was developed. The EHF modeling tool was tested both on the real equipment data sets and on generated ones. The highlights of the tool are:

- Support for EHF system deployment in offline mode (i.e. using history data),
- Parameterization of the EHF models occurs automatically after the model type selection,
- In-depth analysis of the process background is not necessary (but still welcome!) and
- Processes in “out of steady-state” mode can be considered.

As a future work the application of the developed approach on other processes will be tested, including the preciseness analysis of the EHF estimation. Also the use cases from a process engineer’s point of view should be investigated and possibly adapted.

Acknowledgment

The investigation of the described system identification-based EHF estimation method was a part of the EU cooperative project “Enhanced Power Pilot Line” (EPPL) which was co-funded by grants from Austria, Germany, The Netherlands, France, Italy, Portugal and the ENIAC Joint Undertaking.

References

[1] Y.-T. Kao, S. Dauzère-Pérès, J. Blue, and S.-C. Chang, “Impact of integrating equipment health in production scheduling for semiconductor fabrication,” *Computers & Industrial Engineering*, vol. 120, pp. 450 – 459, 2018.

[2] R. K. Mobley, *An introduction to predictive maintenance*. Butterworth-Heinemann, 2002.

[3] S. Alaswad and Y. Xiang, “A review on condition-based maintenance optimization models for stochastically deteriorating system,” *Reliability Engineering & System Safety*, vol. 157, pp. 54 – 63, 2017.

[4] R. Ahmad and S. Kamaruddin, “An overview of time-based and condition-based maintenance in industrial application,” *Computers & Industrial Engineering*, vol. 63, no. 1, pp. 135 – 149, 2012.

[5] J. N.-M. et al., “A literature review on sampling techniques in semiconductor manufacturing,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 26, no. 2, pp. 188–195, May 2013.

[6] C. Yugma, J. Blue, S. Dauzère-Pérès, and A. Obeid, “Integration of scheduling and advanced process control in semiconductor manufacturing: review and outlook,” *Journal of Scheduling*, vol. 18, no. 2, pp. 195–205, Apr 2015.

[7] C. Krauel and L. Weishäupl, “Multivariate approach for equipment health monitoring,” *IFAC-PapersOnLine*, vol. 49, no. 12, pp. 716 – 720, 2016, 8th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2016.

[8] H. M. Hashemian and W. C. Bean, “State-of-the-art predictive maintenance techniques*,” *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 10, pp. 3480–3492, Oct 2011.

[9] C. S. Kulkarni, J. R. Celaya, K. Goebel, and G. Biswas, “Physics based electrolytic capacitor degradation models for prognostic studies under thermal overstress,” *in Proc. of the First European Conference of the Prognostics and Health; Dresden; Germany*, 2012.

[10] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, “Remaining useful life estimation a review on the statistical data driven approaches,” *European Journal of Operational Research*, vol. 213, no. 1, pp. 1 – 14, 2011.

[11] A. Schirru, S. Pampuri, and G. D. Nicolao, “Particle filtering of hidden gamma processes for robust predictive maintenance in semiconductor manufacturing,” in *2010 IEEE International Conference on Automation Science and Engineering*, Aug 2010, pp. 51–56.

[12] P. Lall, P. Choudhary, S. Gupte, and J. Hofmeister, “Statistical pattern recognition and built-in reliability test for feature extraction and health monitoring of electronics under shock loads,” *IEEE Transactions on Components and Packaging Technologies*, vol. 32, no. 3, pp. 600–616, Sept 2009.

[13] P. O’Donovan, K. Leahy, K. Bruton, and D. T. J. O’Sullivan, “An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities,” *Journal of Big Data*, vol. 2, no. 1, p. 25, Nov 2015.

[14] T. Le, M. Luo, J. Zhou, and H. L. Chan, “Predictive maintenance decision using statistical linear regression and kernel methods,” in *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, Sept 2014, pp. 1–6.

[15] S. j. Wu, N. Gebraeel, M. A. Lawley, and Y. Yih, “A neural network integrated decision support system for condition-based optimal predictive maintenance policy,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 37, no. 2, pp. 226–236, March 2007.

[16] S. S. Biswas, A. K. Srivastava, and D. Whitehead, “A real-time data-driven algorithm for health diagnosis and prognosis of a circuit breaker trip assembly,” *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3822–3831, June 2015.

[17] D. An, N. H. Kim, and J.-H. Choi, “Practical options for selecting data-driven or physics-based prognostics algorithms with reviews,” *Reliability Engineering & System Safety*, vol. 133, pp. 223 – 236, 2015.

[18] A. Dementjev, A. Gellrich, A. Schirru, and A. Kästner, “System identification based equipment health estimation,” in *in Proc. of the 15th APCM Conference, Freising, Germany*, 2015.