

Production trend identification and forecast for shop-floor business intelligence

Dr. Zsolt János Viharos^{1,3}, Jenő Csanaki², Dr. János Nacsa¹, Márton Edelényi², Csaba Péntek²,
Krisztián Balázs Kis¹, Ádám Fodor¹, János Csempez¹

¹ *Fraunhofer Project Center for Production Management and Informatics, Institute for Computer Science and Control, Hungarian Academy of Sciences, H-1111, Budapest, Kende u. 13-17. Hungary, {viharos.zsolt, nacsa.janos, kis.krisztian, fodor.adam, csempez.janos}@sztaki.mta.hu*

² *Opel Szentgotthárd Ltd., H-9970, Szentgotthárd, Füzesi út 15., Hungary, {jeno.csanaki, marton.edelenyi, csaba.pentek}@gm.com*

³ *Kecskemét College, H-6000, Kecskemét, Izsáki út 10., Hungary, viharos.zsolt@gamf.kefo.hu*

Abstract – The paper introduces a methodology to define production trend classes and also the results to serve with trend prognosis in a given manufacturing situation. The prognosis is valid for one, selected production measure (e.g. a quality dimension of one product, like diameters, angles, surface roughness, pressure, basis position, etc.) but the applied model takes into account the past values of many other, related production data collected typically on the shop-floor, too. Consequently, it is useful in batch or (customized) mass production environments. The proposed solution is applicable to realize production control inside the tolerance limits to proactively avoid the production process going outside from the given upper and lower tolerance limits. The solution was developed and validated on real data collected on the shop-floor; the paper also summarizes the validated application results of the proposed methodology.

I. INTRODUCTION

The paper introduces a methodology to define production trend classes and also the results to serve with a trend prognosis in a given manufacturing situation.

The overview on production trend forecast methods in the next paragraph concludes that the identification and forecast of production trends are key issues on the shop-floor of manufacturing plants; moreover, many artificial intelligence techniques are applied in this field [1][2]. Trend types can be formulated to define a classification model for the prognosis; Control Chart Pattern (CCP) is the mostly used keyword for these classes [3][4], however the definition of the existence or absence of a trend situation is not specified in the literature, so the current paper proposes a novel methodology for that issue described in paragraph IV.

The paragraph III. gives an overview of the usually applied production databases and emphasizes the typical difficulty of their connections and integration; moreover a

novel method is illustrated how to establish a connection of databases having no common keys but varying time handling capabilities.

The integration of quality, manufacturing execution, alarm handling and machine log information together with the definition of trend situations and the availability of developed machine learning techniques allowed building up production trend identification learning models: their modelling capabilities are reported in the next paragraph.

Conclusions on all the results, acknowledgement and referred literature close the paper.

II. PRODUCTION TREND FORECAST METHODS

Production forecasting has received significant attention in the last decades [5][6][7]. There are many approaches for trend identification and forecasting in time domain to ensure with the required level of production quantity and quality. This is one of the important assignments of the Industry 4.0, direction, too [8], that is mentioned in USA as Industrial Internet or in more general as Cyber-Physical Production Systems (CPPSs) [9][10].

El-Midany et al. used ANNs to recognize a set of subclasses of multivariate abnormal patterns [11] in machining of a crank case as one of the main components of compressor. They used a simulated and a real world data set as well; furthermore they can identify the responsible variable(s) on the occurrence of the abnormal pattern. Ranaee and Ebrahimzadeh used a hybrid intelligent method [3] to recognize whether a process runs in its planned mode or it has unnatural patterns. This method includes three modules: a feature extraction module, a multi-class SVM-based classifier module (MCSVM) and an optimization module using genetic algorithm. They tested the algorithm on synthetically generated control charts. Control Chart Patterns (CCPs) with different levels of noise were analyzed by Lavangananda and Khamchai [4]. They implemented

and compared three different classifiers: Decision Tree, ANN, and the Self-adjusting Association Rules Generator (SARG) for process CCPs that were generated by predefined equations of GARH (Generalized Autoregressive Conditional Heteroskedasticity) Model for \bar{X} chart. Pelegrina et al. used different Blind Source Separation (BSS) methods in the task of unmixing concurrent control charts to achieve high classification rates. [12] Gutierrez and Pham presented a new scheme to generate training patterns for ML algorithms: Support Vector Machine (SVM) and Probabilistic Neural Network (PNN) [13]. Yang et al. proposed a hybrid approach that integrates extreme-point symmetric mode decomposition (ESMD) with extreme learning machine (ELM) to identify typical concurrent CCPs [14]. Motorcu and Güllü constructed X-R control charts for each production line on the data obtained from shop-floor to provide high quality production by eliminating key problems: undesirable tolerance limits, poor surface finish or circularity of spheroidal cast iron parts during machining [15].

Huybrechts et al. applied standardization, trend modeling, and an autoregressive moving average (ARMA) model to determine short-term correlation between subsequent measurements. The out-of-control observations can be determined precisely with the Dijkstra model and cumulative sum chart of the corrected residuals between the measured and predicted values. Milk yield data from two Automatic Milking System (AMS) farms and one farm with a conventional milking system were used for the case study [16].

Köksal et al. in 2011 reviewed the quality management related applications of various data mining techniques in manufacturing industry published between 1997-2007 [1]. They grouped the quality related assignments into four groups: product/process quality description, predicting quality, classification of quality, and parameter optimization. They proved the increasing importance of such research and application techniques and their relevance in industry. Their analysis on the literature also indicated that data mining applications were mostly encountered in the metal, computer and electronic products manufacturing industries, and relatively less observed in plastics, glass, paper, food processing and chemical manufacturing industries. The importance of integrating production and quality data was highlighted in their paper, too. Applications involving classification of quality were not as many as those in the predicting quality category.

Viharos and Monostori presented an approach, already in 1997 [17] for optimization of process chains by artificial neural networks and genetic algorithms using quality control charts. It was shown that the control of “internal” parameters (temporal parameters along the production chain) is a necessity, by this way, early decisions can be made whether to continue the production

of a given part or not. Also continuous optimization of the production system is possible using the proposed solution.

Concerning the applied techniques, the most prevalent approaches are based on statistical methods, such as autoregression, moving average and their combinations: autoregressive integrated moving average model (ARIMA) [18] with use of linear regression analysis, quasi-linear autoregressive model [19] or Markov chain models (MCM) [20]. These methods based on historical production or time series data for modelling and prediction.

Another approaches has appeared with the evolution of artificial intelligence, such us modelling with artificial neural networks (ANN), support vector machines (SVM) or nearest neighbor approaches based on pattern sequence similarity [21]. There are several curve-fitting methods in this field for small sample data, such as genetic algorithm [22]. By using artificial neural networks combined with statistical methods to compensate drawbacks of the separate approaches in trend forecasting lead to better classification and approximation results.

III. MANUFACTURING ASSIGNMENT & INFORMATION SOURCES

Many different IT systems are running on shop-floor, all of them have their main functionality they are supporting. In a business intelligence approach these data can be linked together to result more knowledge about the details of the production system. Without the appropriate linkage a significant part of this knowledge is hidden from the operators and the production engineers. Vogel-Heuser et al. reported that the appropriate integration of IT systems in production environment is still a challenge for the industry, since “on the software side, a typical problem is the consistency between interfaces of components both on the syntactical as well as semantic levels” [23], so, the continuous maintenance of overlapping information is a key question. The Fig. 1. shows the systems related to the published analysis, as a typical example of shop-floor IT environment.

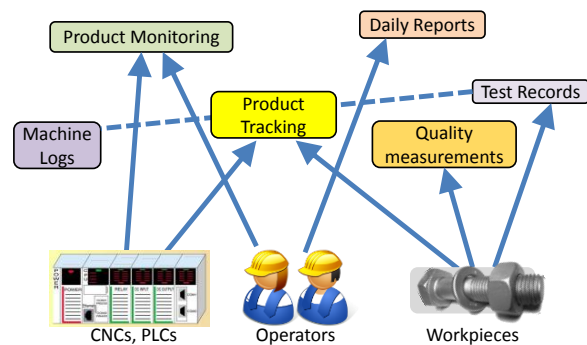


Fig. 1. Data sources on the shop-floor.

Physically, information is inherited from different sensors and controllers (PLCs and CNCs), measuring machines or sometimes the operators; the figure represents the logical sources that are relevant in a business intelligence point of view: the machines, the operators and the workpieces. One of the main scopes of the paper is to collect and link all the information about individual workpieces, and to detect possible trends and relations in the production based on the measurement series of the workpieces.

The quality control systems check the geometry and other features of the workpieces. Measuring equipments process the products independently and the correlations among the measure sets of different machines are typically unknown. It is often simple to link the given measured data to the workpiece, e.g. using a serial code, the time of the measure. One of the main difficulties is the frequency of the measures, because in the most cases only every 'nth' are measured, according to a certain sampling frequency.

The Manufacturing Execution System (MES, Product Tracking in Fig. 1.) logs all the operations and the production time of the individual workpieces. This is the main information source to link the manufacturing machines/operations to the individual workpieces. On the basic level the production monitoring system provides measurement information with the related timestamps about the status of the manufacturing machines and the possible reasons when they are not in production. Other reports are mainly created by the human operators (e.g. daily working reports, special test results). During ETL process, an important feature of a business intelligence system is to correct inconsistency among manually recorded data.

The machine logs are typically not stored in the factory central IT systems and the data they have are deleted after a certain time period. On the other hand many machines have internal measuring systems, moreover, they store internally the measured data and/or the e.g. machining parameter corrections performed on the basis of these measurement. To solve this synchronization/mapping problem an algorithm was developed to find the best fitting delta time between the two independent time series provided by the two independent IT systems (machine logs and MES).

The appropriately connected data set can be used as an integrated data source of the comprehensive shop-floor business intelligence system.

IV. PRODUCTION TREND DEFINITION

The preliminary results of the shop-floor business intelligence applications allowed analyzing/reporting special manufacturing related problems using connected data sets. The overview on production trend forecasting proved the importance of identifying and estimating the production trends already before the process is going

outside the tolerance limits, e.g. to ensure with the requested process capability and stability indexes. The review concluded and also the experiences are mirroring that it is far not an easy task to define and formulate production trends. A decision is needed about a time sequence (time series) of some measurement points whether they form a trend or not. Some aspects of this decision is e.g. what trend length is expected/required, what is the minimum and maximum value, what is the minimum frequency of the measurement points, in what direction the trend is going, what form it has (e.g. linear, exponential), where it starts, etc. Even so if a production process that can be described with equations, or by any other form of knowledge representation method, the answers for these questions are not easy, moreover it is also dependent on the engineering aspects of the given manufacturing assignment. In a majority of industrial cases:

- the plant engineers do not have fix definition what time sequence they consider as trend,
- usually, there exists a hypotheses that there are trends in the processes,
- many and various ideas arise how to use trend forecasts if they are recognized..

Consequently, the identification and forecast of "any" trends have significant manufacturing potentials and benefits. The next illustration gives a novel methodology to automatically determine "*What is a production trend?*".

In the next analysis, a critical production feature (e.g. quality measure) was selected that is manufactured in three operation steps in three different machines in the linear production line. The behavior of the most critical second operation was investigated. As the third operation changes very little in the final value of the selected feature, it was assumed that the measurements at the end of line (EOL) show significant correlations of the circumstances of the second operation. This critical processing is done in three alternative machines and the results of all these three were investigated.

The prepared, linked database allowed working with the EOL values ordered by the machining sequence according to the time of the second operation; consequently, in general, with this approach it is possible to analyze an operation through the measured data collected at another operation. This is an important advantage of the presented approach, e.g. analyses based on varying, available data frequencies and measurement time points are possible. Four different time-window periods were selected for the examination but results of one, having the highest industrial relevance is presented in the paper.

The Fig. 2. shows one example where the average and the deviation on a certain time-window of the individual measurements are also visualized.

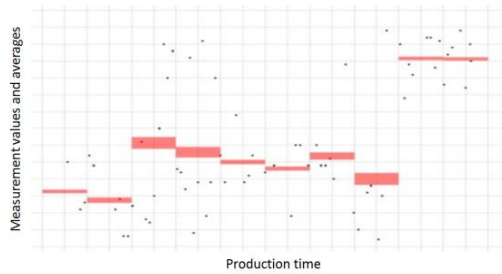


Fig. 2. Measurement points, averages (lines' vertical positions) and deviation values (lines' thickness) in the investigated data set.

The figure shows also that sudden average jumps can be recognized in the given production process. Based on these visual investigations a systematic algorithm was defined to explore trends in the given data sets. Average values and also linear regression and average jumps were sought after. The average jumps were defined at a given time point of one workpiece machining as difference between the average of the previous values and the average of the following values for a certain time window. Average deviation on these time windows was also calculated.

It was clear that the trend exploration is not possible during the time periods when the production is running with moderate speed, consequently, only those periods were accepted (presented and used) when the production

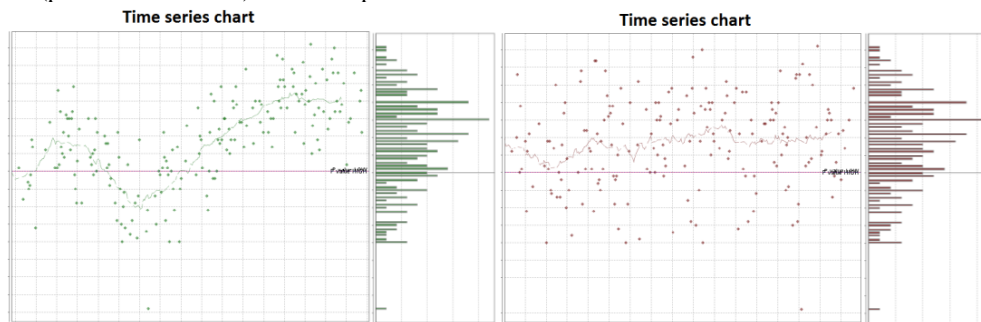


Fig. 4. The same measured production/workpiece values (real case) with their identical distributions, having trends in the left and without any trend in the right (horizontal axis: time, vertical axis: production quality measure values).

This implication (Fig. 4.) leads to the approach taking the measured real production values that are pairs of values & time points from the shop-floor but the mixing of their time points ("fully") randomly results such a measurement series that have the same distribution than the original production sequences but for sure all time domain trends are eliminated. The preparation of such a "trendless" dataset and its representation in the same method as in Fig. 4., together with the original production dataset having trends formulates clearly those non-overlapping zones where production trends are given (Fig. 5.). The data points of the original production trend are on the left side of the figure, representing much lower distribution than the mixed, trendless data points on the

(data) reached a certain, minimum frequency level.

The Fig. 3. below shows all the average jump levels up or down (vertical axis) and their calculated deviations (horizontal axis) in the analyzed (long) production horizon.

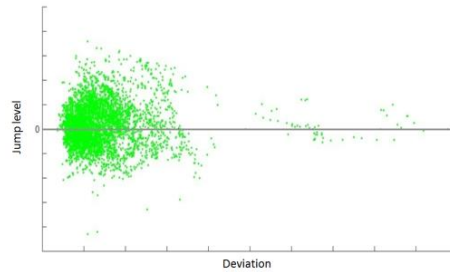


Fig. 3. Jump level (vertical axis) and deviation (horizontal axis) values in the investigated data set

This representation (jumps in the averages versus distribution in the data) has motivated the definition of the trend cases. Trend means a special sequence of data in time and in the value domain together, otherwise no trend is given. Fig. 4. shows typical production trends on the left side, while the same production data are shown on the right side, but with random mixing the time points of the same measured data (so no trend is given).

right side, while the trend jumps are similar or slightly larger in the original production measurements. This experience indicates the existence of real trends since:

- the trend "sizes" (jump levels, vertically) are slightly larger, moreover
- the distribution of the related measurement values are significantly smaller (horizontally) than in case of the mixed dataset and
- the two areas are mainly separated with only slightly overlapping zones.

A range could be specified between the two datasets, defining the border between trend and not trend situations (Fig. 6.).

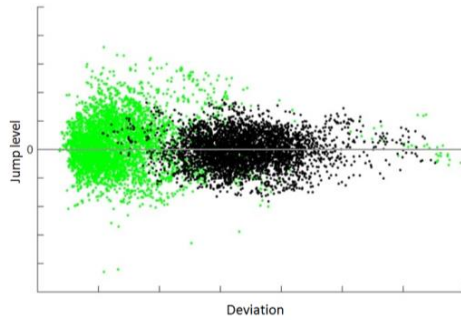


Fig. 5. Data sets with trends (left side) and without trends (right side).

The near to ellipsoidal characteristics of the trendless point zone implicated fitting an ellipsoid around these values. Principal component analysis with e.g. 95% confidence interval can be applied to determine an exact border between the two datasets (Fig. 6.). An interesting experience arose that the largest principal vector was almost parallel to the horizontal axis and almost crossing at the zero vertical value, but not exactly.

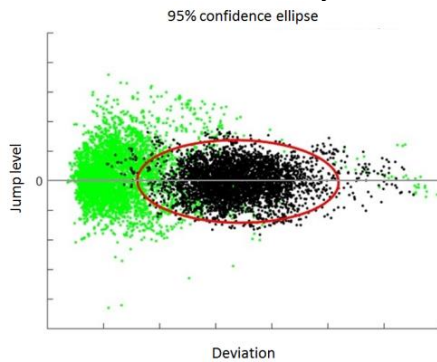


Fig. 6. The exact border between trend and trendless production situations.

Having an exact border between the trend and not trend cases allows defining a classification problem for trend identification (Fig. 7.). When the distribution values of the production points are horizontally above the middle point of the ellipsoid they are considered as no trend situations. Consequently, three classes can be formulated:

- *Trend up:* Production measures of the analyzed workpieces in the considered interval form a trend up class through the points in Fig. 7. above the zero vertical axis and horizontally less than the ellipsoid center when they are outside the identified ellipsoid.
- *Trend down:* these situations can be defined similarly to the trend up cases but the values are below the zero horizontal axis.
- *No trend:* points inside the ellipsoid and also that have distribution higher than the center of the identified ellipsoid.

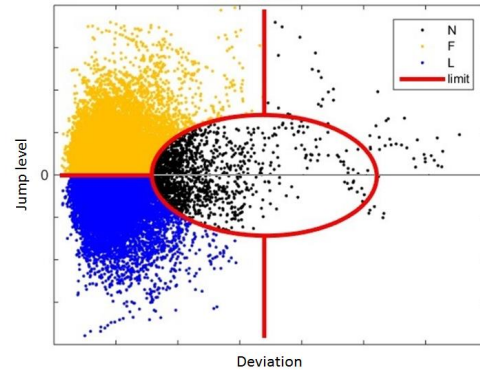


Fig. 7. Identified fields and borders of trend classes (F: trend goes upwards, L: trend goes downwards, N: no trend is given)

Each point in Fig. 7. represents a trend class and can be ordered to one individual workpiece, consequently, for each workpiece trend class can be ordered representing the trend in its's future. Consequently, it is a classification assignment where trend classes define the outputs while past values of many other, related measurements collected typically on the shop-floor form the available input information.

Naturally, the relationship between past production measurements and the trend classes is unknown, consequently, it is impossible to form a closed form, or any equation based model for this task. However given measurements at each workpiece (even with sampling) lead to the possibility to use any learning model based on available shop-floor data. The probably non-linear and clearly multidimensional relationship among input and output data indicated the application of artificial neural network model for the formulated classification assignment [24], where the model is able to forecast the trend in form of trend class specification at each manufactured and measured workpiece.

V. PRODUCTION TREND FORECAST RESULTS

The previously defined trend class specification can be applied dynamically to the production data resulting in a large labelled dataset which can be used for model building and testing.

The precision of the forecast model can be measured by the amount of data points correctly labelled by the model which can be expressed as a percentage of the whole dataset (Fig. 8.).

Fig. 8. shows the classification results of the applied forecast model by measuring the percentage of data points falling into the nine possible recognition categories. As there are three real and three prognosed classes (F: trend up, L: trend down, N: no trend) the number of different recognition cases is nine. On the diagram the blue, middle columns show the amount of

correct recognition and remaining, red columns show the incorrect recognitions. One can see that the ratio of correct trend classification is around 60-70%.

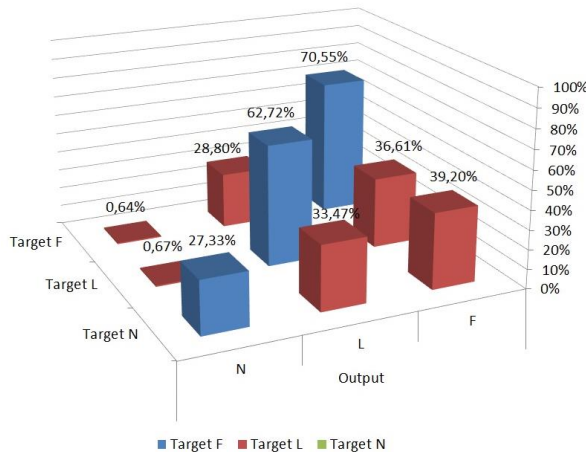


Fig. 8. Classification results: Target represents the real measurements while Output means the model classification decision.

As the classes are connected to the individual elements of a time series (manufactured workpieces) the recognition categories can be arranged in continuous trend sections (Fig. 9).

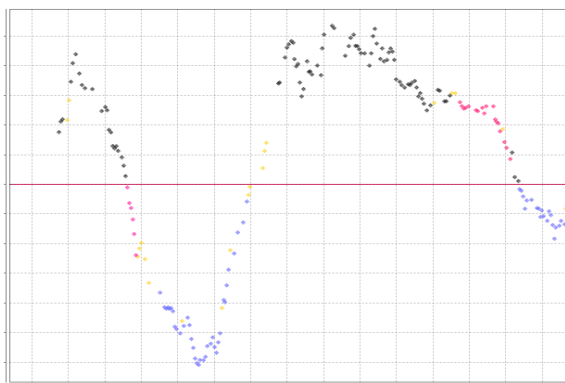


Fig. 9. Trend decisions at each workpiece (horizontal axis: production time, vertical axis: quality measure values, one-by-one of the individual produced products)

Fig. 9. shows the trend sections on a small part of the e.g. whole year time series. Black (on the upper side) denotes the correctly diagnosed F (trend up) trends; blue (on the lower side) denotes the correctly diagnosed L (trend down) trends and red (mainly in the middle) denotes any form of misclassification. The y axis is the measure from which the classes were directly derived. It can be seen that the incorrectly diagnosed data points are concentrated in the middle area.

Consequently, it is proven that the production trends can be recognized with fair accuracy with the proposed

approach.

VI. CONCLUSIONS

The paper presented a methodology to define production trend classes and also the results to serve with trend prognosis in a given manufacturing situation. The solution is useful in batch or (customized) mass production environments because it continuously collects and analyses data from the shop floor sensors and is applicable to realize production control inside the tolerance limits to proactively avoid the production process going outside from the given upper and lower tolerance limits.

The review of the production trend forecast methods concluded and also the experiences are mirroring that it is far not an easy task to define and formulate production trend classes. The developed solution for solving this issue collects shop-floor data and based on the concrete manufacturing values it is able to define three trend classes: i) trend goes up, ii) trend goes down and iii) no trend is given. It is an automatic process, consequently, for this decision no prescriptions are needed from the plant engineers or shop-floor operators. However, the resulted classes can be modified by them when any hypothesis or experiences are given in this field.

The described method for production trend prognosis considers past values from any related (e.g. physically previous) operations e.g. measurements, alarms, etc. Using historical data the applied artificial neural network model determines the prognosis at each produced workpiece. This model is built up also on shop-floor data of the process analyzed, consequently, it is valid for that concrete process under the given, prompt situations.

The solution was developed and validated on real data collected on the shop-floor; the paper summarizes the validated application results of the proposed methodology, too.

VII. ACKNOWLEDGEMENT

The authors acknowledge the support of grants of the Highly industrialised region on the west part of Hungary with limited R&D capacity: Research and development programs related to strengthening the strategic future oriented industries manufacturing technologies and products of regional competences carried out in comprehensive collaboration, VKSZ_12-1-2013-0038. This research has been supported also by the Hungarian Scientific Research Fund (OTKA), Grant No. 113038.

REFERENCES

- [1] G. Köksal, I. Batmaz, M.C. Testik, "A review of data mining applications for quality improvement in manufacturing industry", *Expert Systems with Applications*, Elsevier, Vol 38., 2011, pp. 13448–13467.
- [2] P. Bilski, "Analysis of the classifier fusion efficiency

- in the diagnostics of the accelerometer”, *Measurement*, Vol. 67., Elsevier, 2015, pp. 116-125.
- [3] V. Ranaee, A. Ebrahimzadeh, "Control chart pattern recognition using a novel hybrid intelligent method", *Applied Soft Computing*, vol.11, 2011, pp.2676–2686.
- [4] K. Lavangnananda, S. Khamchai, "Capability of Control Chart Patterns Classifiers on Various Noise Levels", *Procedia Computer Science*, vol.69, 2015, pp.26–35.
- [5] R. Schmitt, L. Monostori, H. Glöckner, Zs.J. Viharos, "Design and assessment of quality control loops for stable business processes", *CIRP Annals – Manufacturing Technology*, Vol. 61, No. 1, 2012, pp. 439-444.
- [6] D. Galar, A. Thaduria, M. Catelani, L. Ciani, "Context awareness for maintenance decision making: A diagnosis and prognosis approach", *Measurement*, Vol. 67., Elsevier, 2015, pp. 137-150.
- [7] L. Cristaldi, M. Faifer, M. Lazzaroni, M. M. A. F. Khalil, M. Catelani, L. Ciani, "Diagnostic architecture: A procedure based on the analysis of the failure causes applied to photovoltaic plants", *Measurement*, Vol. 67., Elsevier, 2015, pp. 99-107.
- [8] E. Geisberger, M. V. Cengarle, P. Keil, J. Niehaus, C. Thiel, H.-J. Thönnißen-Fries, „Cyber-Physical Systems Driving force for innovation in mobility, health, energy and production”, *acatech – National Academy of Science and Engineering*, Germany, position paper, December, 2011, p. 48.
- [9] H. Gill, "Cyber-Physical Systems", *Fourth Annual Carnegie Mellon Conference on the Electricity Industry*, March 10-11, 2008, p. 28.
- [10] R. Rajkumar, I. Lee, L. Sha, J. Stankovic, "Cyber-Physical Systems: The Next Computing Revolution", *Design Automation Conference (DAC), 2010 47th ACM/IEEE*, 13-18 June 2010., pp. 731 – 736.
- [11] T.T. El-Midany, M.A. El-Baz, M.S.Abd-Elwahed, "A proposed framework for control chart pattern recognition in multivariate process using artificial neural networks", *Expert Systems with Applications*, vol.37, 2010, pp.1035–1042.
- [12] G. D. Pelegrina, L. T. Duarte, C. Jutten, "Blind source separation and feature extraction in concurrent control charts pattern recognition: Novel analyses and a comparison of different methods", *Computers & Industrial Engineering*, vol.92, 2015, pp. 105-114.
- [13] H. De la T. Gutierrez, D.T. Pham, "Estimation and generation of training patterns for control chart pattern recognition", *Computers & Industrial Engineering*, Vol. 95, 2016, pp. 72-82.
- [14] W.-A. Yang, W. Zhou, W. Liao, Y. Guo, "Identification and quantification of concurrent control chart patterns using extreme-point symmetric mode decomposition and extremelearning machines", *Neurocomputing*, Vol. 147, 2015, pp. 260-270.
- [15] A.R. Motorcu, A. Güllü, "Statistical process control in machining, a case study for machine tool capability and process capability", *Materials and Design*, vol.27, 2006, pp.364–372.
- [16] T. Huybrechts, K. Mertens, J. De Baerdemaeker, B. De Ketelaere, W. Saeys, "Early warnings from automatic milk yield monitoring with online synergistic control", *American Dairy Science Association*, Vol.97, 2014, pp. 3371-3381.
- [17] Zs.J. Viharos, L. Monostori, "Optimization of process chains by artificial neural networks and genetic algorithms using quality control charts", *Proceedings of Danube - Adria Association for Automation and Metrology*, Dubrovnik, 1997. pp. 353-354.
- [18] J. Xie, Y. Wang, X. Zheng, Q. Yang, T. Wang, Y. Zou, J. Xing, Y. Dong, "Modeling and forecasting *Acinetobacter baumannii* resistance to set appropriate use of cefoperazone-sulbactam: Results from trend analysis of antimicrobial consumption and development of resistance in a tertiary care hospital", *American Journal of Infection Control*, vol.43, 2015, pp.861-864.
- [19] M. Gan, Y. Cheng, K. Liu, G. Zhang, "Seasonal and trend time series forecasting based on a quasi-linear autoregressive model", *Applied Soft Computing*, vol.24, 2014, pp.13-18.
- [20] C.R. Clarkson, J.D. Williams-Kovacs, F. Qanbari, H. Behmanesh, M.H. Sureshjani: "History-matching and forecasting tight/shale gas condensate wells using combined analytical, semi-analytical, and empirical methods", *Journal of Natural Gas Science and Engineering*, vol.26, 2015, pp.1620-1647.
- [21] I. Koprinska, M. Rana, A.T. Lora, F. Martínez-Álvarez, "Combining pattern sequence similarity with neural networks for forecasting electricity demand time series", *The 2013 International Joint Conference on Neural Networks (IJCNN)*, 2013, pp.1-8.
- [22] V.K. Semenychev, E.I. Kurkin, E.V. Semenychev, "Modelling and forecasting the trends of life cycle curves in the production of non-renewable resources", *Energy*, Vol.75., 2014, pp. 244-251.
- [23] B.Vogel-Heusera, A. Fay, I. Schaefer, M. Tichy, "Evolution of software in automated production system: Challenges and research directions", *The Journal of Systems and Software*, Vol. 110., 2015., pp. 54-84.,
- [24] Zs.J. Viharos, "AI based, automatic production system decomposition", *5th International Conference on Digital Enterprise Technology, DET2008*, October 22-24, 2008, Nantes, France, paper nr. 94.