Advanced Machine Learning in Recursive Data-based Modelling

Esko K. Juuso

Control Engineering, Environmental and Chemical Engineering, Faculty of Technology, P.O. BOX 4300, 90014 University of Oulu, Finland; esko.juuso@oulu.fi

SNE 32(2), 2022, 113-120, DOI: 10.11128/sne.32.tn.10608
Received: 2021-02-10 (selected EUROSIM 2019 Postconf. Publication; Revised: 2022-05-30; Accepted: 2022-05-31
SNE - Simulation Notes Europe, ARGESIM Publisher Vienna, ISSN Print 2305-9974, Online 2306-0271, www.sne-journal.org

Abstract. Recursive data-based modelling is needed for making decision online in varying operating conditions. Recursive algorithms are useful in adapting the parameters within selected memory horizons. Abrupt changes can be handled when the situation change is approved to be drastic. The nonlinear scaling based on generalized norms includes additional alternatives: the norm orders adapt to the gradually changing operating conditions. The drastic shape changes of the scaling functions require full analyses of the orders. The orders can also be stored for different situations and re-used later. Fuzzy inequalities are useful in finding out if the feasible ranges of the most recent period are different from the current active ranges or similar with some of previous feasible ranges. Machine learning is integrated in the system in three levels: (1) finding the appropriate time windows, (2) interactions of feasible levels, and (3) finding decision support when some of feasible ranges need to change. These decisions are supported by expert knowledge. Other model parameters can be included in the analysis. The solution has been tested with measurement data from several application cases. The recursive approach is beneficial in the control and maintenance in varying operating conditions.

Introduction

Models understood as relationships between variables are used for predicting of properties or behaviours of the system. Variable interactions and nonlinearities are important in extending the operation areas of control and fault diagnosis, where the complexity is alleviated by introducing software sensors [1]. Recursive data-based modelling is needed in varying operating conditions. In industry, where very large datasets have been common already long time, the problem has been tackled by *data analytics* and *intelligent systems*, where linear regression and parametric models are used in the recursive tuning of the interaction equations [2]. Adaptation of the parametric systems is essential in varying operating conditions. Various nonlinear multivariable systems combine statistical and intelligent methodologies with sensor fusion based on data pre-processing, signal processing and feature extraction [3]. Dynamic models are based on additional parametric model structures.

Artificial intelligence (AI) mimic human cognitive functions, including e reasoning, knowledge, planning, learning, natural language processing, perception and the ability to move and manipulate objects. *Machine learning (ML)* is a subset of AI: the iterative seeking of solutions is done by using new architectures, techniques and algorithms in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. *artificial neural networks* have been widely used in these studies as a behavioural model to map a systems input to its output regardless of the nature of the system.

Fuzzy logic extends the approximate reasoning, and the connection of fuzzy rule-based systems and expert systems is clear. Fuzzy set systems separate meanings and interactions which is an important key in the adaptation in varying operating conditions. *Linguistic equation (LE)* approach originates from fuzzy set systems: rule sets are replaced with linear equations, and meanings of the variables with nonlinear scaling functions [4]. Constraints handling [5] and data-based analysis [6], facilitate the recursive updates of the systems [7]. The nonlinear scaling revises the meanings and linear models represent the interactions [1].

Table 1: Adaptation level and learning.

Adaptation	Smart adaptive systems	Machine learning
Changing environment	Recursive data analytics for parametric solutions	Statistical techniques, network tuning.
Similar setting, not explicitly ported	Computational intelligence, expertise and data analytics, including variable selection	Data mining utilizing previous experience
New or unknown application	Computational intelligence, expertise and data analytics, including variable selection.	Deep learning with big data ideas.

Statistical analysis with steady state models form the basic elements of different intelligent models. Decomposed systems can be based on weighting the local models with *linear parameter varying (LPV) models* [8]. External dynamic models provide the dynamic behaviour for the LE models developed for a defined sampling interval.

Smart adaptive systems (SAS) are aimed for developing successful applications in different fields by using three levels of adaptation [9]: (1) adaptation to a changing environment, (2) adaptation to a similar setting without explicitly being ported to it, and (3) adaptation to a new or unknown application (Table 1). The recursive analysis is important in all these levels. The smart adaptive data analysis and the data processing form a five-layer advanced deep learning (ADL) platform which supports levels of smart adaptive systems and development of cyber-physical systems (CPS). [2]

Machine learning (ML) is focusing on algorithmic data-based analyses which provide promises for automatic modelling. Performance is good in small scale systems and interesting interactions are found in big data analysis where the previous knowledge is limited. Additional levels bring more challenges for explanations. How far we can go in complex systems? How can we guide the automatic analysis? Different measurements have specific processing requirements [2]. Therefore, the recursive adaptation has application specific requirements.

This article focuses on the possibilities of applying machine learning in data-based modelling. Data analysis (Section 1) is the key part of the modelling. The resulting parametric systems are recursively updated (Section 2) and the solutions are analysed in four different applications (Section 3). The overall system is analysed in Section 4 and conclusions about the applicability of the advanced machine learning in these applications are drawn in Section 5.

1 Data analysis

The main part of the data analysis is variable specific extended with parametric models.

1.1 Variable specific analysis

Machine learning is suitable for the variable specific analysis: features, parameters of the nonlinear scaling functions and intelligent indicators are obtained by using a set of algorithms. Several measurements and sets of features can be analysed in parallel.

Features The arithmetic means and medians are suitable for recursive tuning, but the resulting scaling functions are narrow and sensitive to outliers. More flexible solutions can be obtained with generalized norms defined by

$$||^{\tau}M_j^p||_p = (M_j^p)^{1/p} = [\frac{1}{N}\sum_{i=1}^{N}(x_j)_i^p]^{1/p}, \qquad (1)$$

where the order of the moment $p \in R$ is non-zero, and N is the number of data values obtained in each sample time τ. The norm (1) calculated for variables x_j, $j = 1, \ldots, n$, have the same dimensions as the corresponding variables [10]. These norms can be extended to variables including negative values [7]. The norm values are monotonously increasing with the norm order.

The norm values are updated by including new equal sized sub-blocks in calculations since the computation of the norms can be done from the norms obtained for the equal sized sub-blocks, i.e. the norm for several samples can be obtained as the norm of the norms of the individual samples:

$$||^{K_s\tau}M_j^p||_p\{\frac{1}{K_s}\sum_{i=1}^{K_s}[(^{\tau}M_j^p)_i^{1/p}]^p\}^{1/p}=[\frac{1}{K_s}\sum_{i=1}^{K_s}[(^{\tau}M_j^p)_i]^{1/p}, \tag{2}$$

where K_s is the number of samples $\{x_j\}_{i=1}^N$. [10] In automation and data collection systems, the sub-blocks are normally used for arithmetic mean ($p=1$).

Nonlinear scaling Meanings of the feature or measurement values are represented by using monotonously increasing functions $x_j = f(X_j)$ where x_j is the variable and X_j the corresponding scaled variable. The function $f()$ consist of two second order polynomials, one for the negative values of X_j and one for the positive values, respectively. The corresponding inverse functions $x_j = f^{-1}(X_j)$ based on square root functions are used for scaling to the range [-2, 2]. Both functions, $f()$ and $f^{-1}()$, are monotonously increasing. [4, 5]

The parameters of these scaling functions are extracted from measurements by using generalized norms and moments. The support area is defined by the minimum and maximum values of the variable, i.e. the support area is $[\min(x_j), \max(x_j)]$ for each variable $j, j = 1,\ldots,m$. The central tendency value, c_j, divides the support area into two parts, and the core area is defined by the central tendency values of the lower and the upper part, $(c_l)_j$ and $(c_h)_j$, correspondingly. This means that the core area of the variable j defined by $[(c_l)_j,(c_h)_j]$ is within the support area. [6] Monotonicity constraints and special requirements are used if needed [5].

The scaling is defined by five parameters which allow highly asymmetric functions, i.e. different shapes for upper and lower parts of the functions. The range [-2, 2] provides a good basis for natural language representations [11]

Intelligent indicators All features and measurements processed with the nonlinear scaling can be used as intelligent indicators. Several indicators can combined and additional indicators, including trends and fluctuations, can be constructed with temporal analysis [12]. *Trend indices*,

$$I_j^T(k) = \frac{1}{n_S+1}\sum_{i=k-n_S}^{k}X_j(i) - \frac{1}{n_L+1}\sum_{i=k-n_L}^{k}X_j(i), \tag{3}$$

are based on the means obtained for a short and a long time period, defined by delays n_S and n_L, respectively. Time periods are variable specific. The index value represents the strength of both the decrease and increase of the variable x_j. The derivative of the index $I_j^T(k)$, denoted as $\Delta I_j^T(k)$, is used for analyzing triangular episodic representations. Severity of the situation evaluated by a *deviation index*

$$I_j^D(k) = \frac{1}{3}(X_j(k) + I_j^T(k) + \Delta I_j^T(k)). \tag{4}$$

All the indicators are in the range [-2, 2], which facilitates the natural language representation also them [11].

Fluctuation indicators are based on the moving range of variable values obtained as a difference of two moving generalized norms:

$$\Delta x_j^F(k) = ||^{K_s\tau}M_j^{p_h}||_{p_h} - ||^{K_s\tau}M_j^{p_l}||_{p_l}, \tag{5}$$

where the orders $p_h \in \Re$ and $p_l \in \Re$ are large positive and negative, respectively. The moments are calculated from the latest K_s+1 values, and an average of several latest values of $\Delta x_j^F(k)$ is used as an indicator of fluctuations. [13]

1.2 Parametric models

The performance of the scaling functions can be analysed by using combinations of several indicators, e.g. the deviation index (4) combines three indicators. The analysis is expanded with models, which have only linear interactions between the indicators. All the scaled variables and linear combinations of them are in the same range [-2, 2] where integer numbers correspond labels, e.g. {very low, low, normal, high, very high}. [11]

2 Recursive analysis

Recursive data analysis facilitates the adaptation of the scaling functions to changing operating conditions, also the orders of the norms are re-analyzed if needed. The existing scaling functions provide a basis for assessing the quality of new data: outliers should be excluded, but the suspicious values may mean that the operating conditions are changing. The scaling functions are extended for analysing outliers and suspicious values to select data for the adaptive scaling. The borders represent the data distribution for different shape factors α_j^+ (Figure 1).

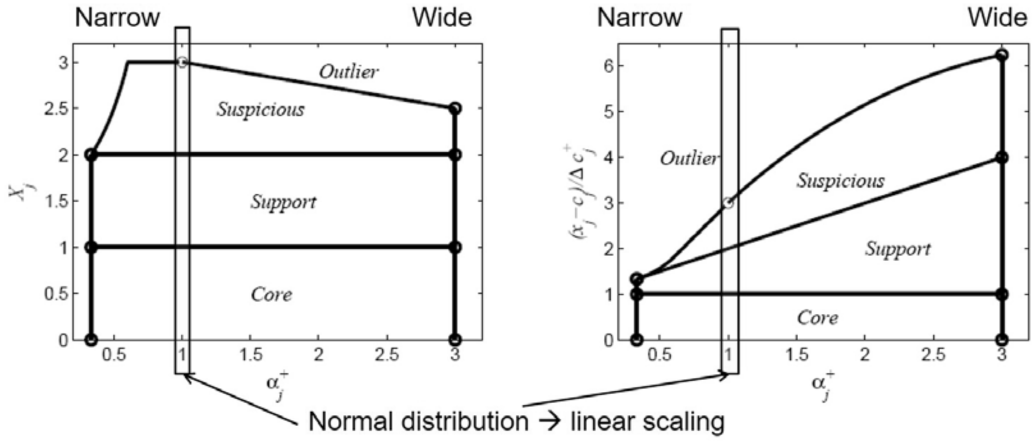


Figure 1: Limits for the core, support, suspicious and outlier areas as a function of shape factor $\alpha_j^+ \in [\frac{3}{3}, 3]$ [14]

The parameters of the nonlinear scaling functions can be recursively updated with (2) by including new equal sized sub-blocks in calculations. The number of samples K_s can be increasing or fixed with some forgetting, and weighting of the individual samples can be used in the analysis. If the definitions should cover all the operating areas, also suspicious values are included as extensions of the support area. In each adaptation step, the acceptable ranges of the shape factors α_j^- and α_j^+ are checked and corrected if needed. The analysis has two levels: the parameters of the scaling functions and the corresponding orders of the norms. [15]

Outliers Clear outliers need to be excluded in both the first analysis and the subsequent recursive steps. In linear scaling, the z-score values outside the range [-3, 3] are often considered as an indication of an outlier (Figure 1). The scaled values are in the range [-2, 2], and this is also the range for the monotonous increase if the minimum and maximum points are obtained from the derivatives of the scaling functions.

Recursive adaptation The parameter of the scaling functions can be recursively updated by using the norms (1) with five defined orders, $(p_{min})_j$, $(p_l)_j$, $(p_0)_j$, $(p_h)_j$ and $(p_{max})_j$, which correspond the corner points of the scaling function. A highly negative and highly positive orders are used instead of *min* and *max*, respectively. Calculations are done in two windows: short and long. If the corner points are not drastically different, the new blocks are included in the calculation of the parameters.

Drastic changes are needed for the corner points if the smooth adaptation does not provide suitable parameters for the new data distribution, i.e. the distribution is changing considerably with new measurements. The orders of the corresponding norms need to be re-analysed. The new situation may require a totally new set of parameters.

Change detectors and decision making Intelligent trend analysis may provide an early indication of the coming changes [12]. Generalised statistical process control (GSPC) notifies if the limits are exceeded more often [16].

3 Applications

The solution has been tested with measurement data from application cases.

3.1 Solar thermal collectors

Solar power plants should collect any available thermal energy in a usable form at the desired temperature range. Irradiation varies considerably between days and on cloudy periods, the variations are very fast and strong variations (Figure 2). The efficient collection requires a fast start-up and reliable operation in the varying cloudy conditions without unnecessary shutdowns and start-ups.

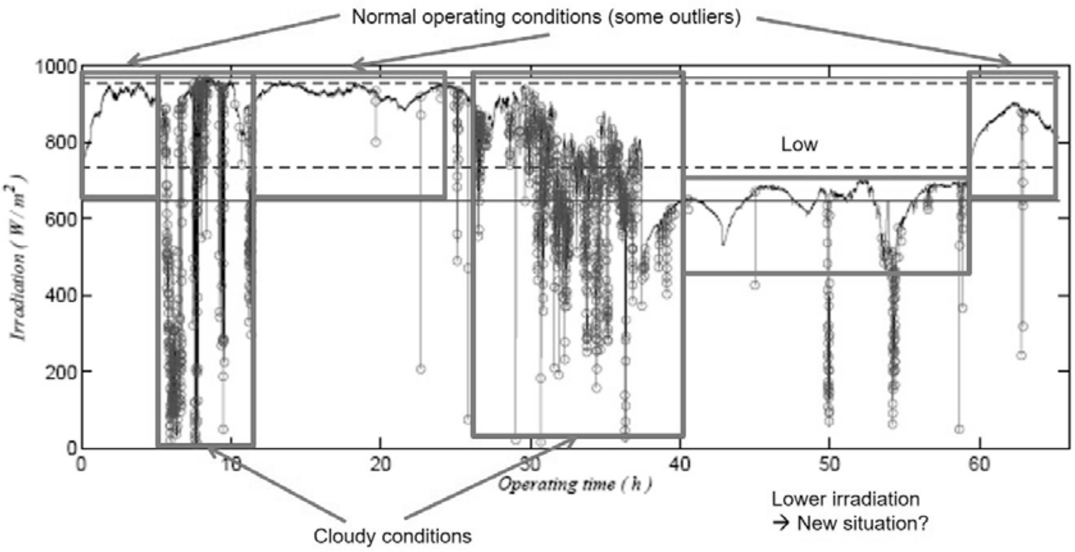


Figure 2: Varying irradiation on a solar collector field: measurements are from 65 hours in 12 days.

The irradiation measurements shown in Figure 2 include typical operating periods. In the first normal period, the feasible area expands until the solar noon is achieved and scaling functions are only slightly modified in the afternoon. The second normal period after cloudy conditions continues from the scaling functions of the first period. The fast dropping single values are considered as outliers which do not affect on the recursive updates of the parameters. This period is followed by a short cloudy situation.

Cloudy periods are detected with the fluctuation indicator (5). The norms are calculated, but they are not used for the recursive updating. In the long cloudy period, the irradiation is fluctuating and the level is going down. The controller limits the acceptable range of control actions by changing the working point in these situations.

After recovering from the cloudy conditions, the situation is compared with the active set of scaling functions which were updated during the second normal period. A new situation is detected: the irradiation is lower than during the previous normal operating conditions. The new set of scaling functions are activated and gradually refined. The parameters of the previous functions are stored for future use. Short cloudy periods disturb operation in this period as well.

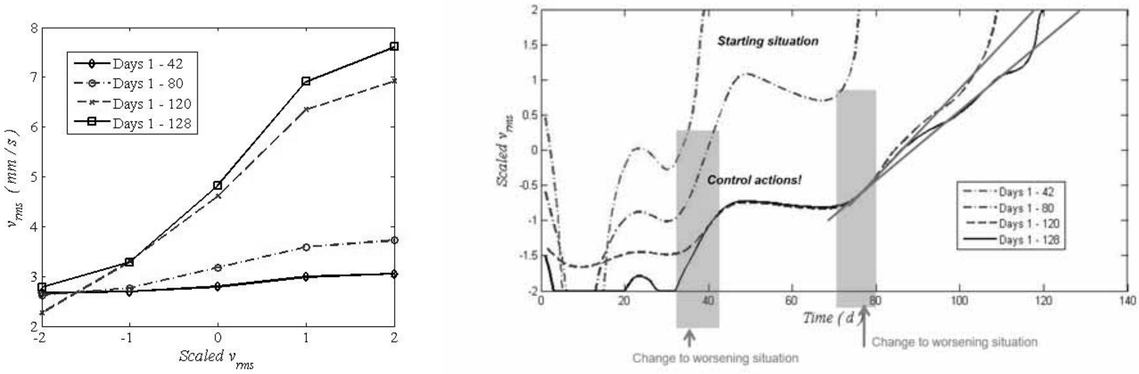
The parameters of the first normal period were recovered for the last day of the measurement period.

Trend indices (3) and deviation indices (4) are used for the early detection of changes for adaptive control [17]. For example, the decrease of the irradiation during the long cloudy period.

This research analysed the irradiation measurements. The same methodologies can be used e.g. for the energy demand, temperature difference over the collector field, and properties of the field devices and environment. The machine learning will be used for this extended problem.

3.2 Prognostics

In the prognostics, the range of the scaling functions need to be expanded when new phenomena activate. This is quite typical when wearing progresses. Recursive data analysis has been demonstrated by using root mean square (rms) velocity, v_{rms}, measurements collected from a paper machine: resin problems of a press roll in the felt washer. The scaling functions were recursively updated three times (Figure 3(a)) after expanding the data set by using spline interpolation to get sufficiently long data sets for the recursive predictions (Figure 3(b)) [18]. The system clearly shows the point when decisions are needed. Domain expertise is needed to assess the situation. There are three alternatives: (1) update the scaling functions, (2) change control actions or (3) start maintenance.



(a) Updating of the parameters.

(b) Forecasting of the failure.

Figure 3: Recursively scaled values of v_{rms} for a press roll of the washer, modified from [18]

.

The decision is based on the assessment of condition, remaining useful time and alternative schedules of maintenance. The control actions to reduce the speed of wearing are good options. In this case, the operation time was extended and the maintenance was postponed to a better time period. The analysis was extended to uncertainty processing and natural language in [19].

3.3 Wastewater treatment

Biological water treatment depends strongly on the inlet water quality. Load and nutrient should be balanced since both an exceptionally high load and excess nutrients cause problems. The operating conditions are modified by oxygen, temperature and flow. Much slower changes in the biological state drastically influence the purification result and subsequent process phases. Scaled values are used together with intelligent trend indices. [20]

The recursive updates of the scaling functions are important for modelling in different situations. Also, the interaction coefficients can depend on operating conditions. Therefore, the early indications of changes provided by the temporal analysis are beneficial in getting warnings and avoiding alarming situations.

3.4 Fatigue

Fatigue is caused by repeated loading and unloading. The mechanism proceeds through cracks formed when the load exceeded certain thresholds. Structures fracture suddenly when a crack reaches a critical size. Stress-cycle (S-N) curves, also known as Wöhler curves, are represented by a linguistic equation

$$I_S(k) = log_{10}(N_C(k)), \tag{6}$$

where the stress index $I_S(k)$ is obtained the stress based on the torque measurements [21]. The scaling of the logarithmic values of the number of cycles, $N_C(k)$, is linear. As the LE model is nonlinear, the LE based S-N curve covers a wide operating range. The continuous model (6) extends the principle of the Palmgren-Miner linear damage hypothesis. In each sample time, τ, the cycles $N_C(k)$ obtained from $I_S(k)$ by (6), and the resulting contribution $\tau/N_C(k)$ is summarised to the previous contributions in the risk analysis. Since the stress is not constant for the whole cycle, the sample time is taken as a fraction of the cycle time. The previous history can be updated whenever the scaling functions are changed. [22]

4 Discussions

Compact linear models enhanced with the nonlinear scaling are used in the selected application cases. Recursive data-based modelling is needed for making decision online in varying operating conditions (Table 2): disturbances, activation of new phenomena, different operating conditions and material properties are taken into account. Recursive algorithms are useful in adapting the parameters within selected memory horizons. Abrupt changes can be handled when the situation change is approved to be drastic.

Table 2: Recursive adaptation in applications.

Application	Phenomena	Recursive analysis
Solar thermal collectors	Daily and seasonal variations Cloudy conditions	Temporal analysis Fluctuations
Prognostics	New phenomena activation	Smoothly extending scaling range
Wastewater treatment	Several operating conditions	Early detection with temporal analysis
Fatigue	Stress scaling	Tuning of risk analysis

Temporal analysis provides early indications in slow processes. Risk analysis is needed to set appropriate labels for the calculation results.

Machine learning is integrated in the system in three levels: (1) finding the appropriate time windows, (2) interactions of the feasible levels, and (3) finding decision support when some feasible ranges need to change. Selecting the time windows, sample times, frequencies and the weights of different indicators are supported by expert knowledge.

In applications, linear interactions can be widely used together with the nonlinear scaling. This does not need to be taken as a limitation, since the approach can be understood as a data pre-processing for any type of nonlinear models, including fuzzy and neural models. Dynamic models are based dynamic structures. All the variables, features and indicators are represented with natural language.

5 Conclusions

The machine learning can focus on the variable specific analysis: algorithms for extracting features, tuning parameters of the nonlinear scaling functions and developing intelligent indicators can be integrated in the machine learning approach. Several measurements and sets of features can be analysed in parallel. The decision-making required in the recursive analysis and the adaptation solutions can be performed within the machine learning. The algorithmic solutions can be improved by using domain expertise and feedback information through other methodologies of computational intelligence.

References

[1] Juuso EK. Intelligent Methods in Modelling and Simulation of Complex Systems. *Simulation Notes Europe SNE*. 2014;24(1):1–10.

[2] Juuso E. Smart Adaptive Big Data Analysis with Advanced Deep Learning. *Open Engineering*. 2018; 8(1):403–416.

[3] Fortuna L, Graziani S, Rizzo A, Xibilia MG. *Soft Sensors for Monitoring and Control of Industrial Processes*. Advances in Indusrial Control. New York: Springer. 2007. 270 pp.

[4] Juuso EK. Integration of Intelligent Systems in Development of Smart Adaptive Systems. *International Journal of Approximate Reasoning*. 2004; 35(3):307–337.

[5] Juuso EK. Tuning of Large-Scale Linguistic Equation (LE) Models with Genetic Algorithms. In: *Revised selected papers of the International Conference on Adaptive and Natural Computing Algorithms - ICANNGA 2009, Kuopio, Finland, Lecture Notes in Computer Science*, edited by Kolehmainen M, vol. LNCS 5495, pp. 161–170. Heidelberg: Springer-Verlag. 2009;.

[6] Juuso E, Lahdelma S. Intelligent scaling of features in fault diagnosis. In: *7th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2010 - MFPT 2010, 22-24 June 2010, Stratford-upon-Avon, UK*, vol. 2. 2010; pp. 1358–1372.

[7] Juuso EK. Recursive Tuning of Intelligent Controllers of Solar Collector Fields in Changing Operating Conditions. In: *Proceedings of the 18th World Congress The International Federation of Automatic Control, Milano (Italy) August 28 - September 2, 2011*, edited by Bittani S, Cenedese A, Zampieri S, pp. 12282–12288. IFAC. 2011;.

[8] Ljung L. Perspectives on system identification. In: *Plenary papers, milestone reports & selected survey*

papers, 17th IFAC World Congress, Seoul, Korea, July 6-11, 2008, edited by Chung MJ, Misra P, pp. 47–59. IFAC. 2008;http://www.ifac-papersonline.net/.

[9] Anguita D. Smart Adaptive Systems - State of the Art and Future Directions for Research. In: *Proceedings of Eunite 2001 - European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems, July 13-14, 2001, Tenerife, Spain*, pp. 1–4. Aachen: Wissenschaftsverlag Mainz. 2001;.

[10] Lahdelma S, Juuso E. Signal processing and feature extraction by using real order derivatives and generalised norms. Part 1: Methodology. *International Journal of Condition Monitoring*. 2011;1(2):46–53.

[11] Juuso EK. Informative process monitoring with a natural language interface. In: *2016 UKSim-AMSS 18th International Conference on Modelling and Simulation, 6-8 April, 2016, Cambridge, UK*, pp. 105–110. IEEE Computer Society. 2016;.

[12] Juuso EK. Intelligent Trend Indices in Detecting Changes of Operating Conditions. In: *2011 UKSim 13th International Conference on Modelling and Simulation*, pp. 162–167. IEEE Computer Society. 2011;.

[13] Juuso EK. Model-based Adaptation of Intelligent Controllers of Solar Collector Fields. In: *Proceedings of 7th Vienna Symposium on Mathematical Modelling, February 14-17, 2012, Vienna, Austria, Part 1*, edited by Troch I, Breitenecker F, vol. 7, pp. 979–984. IFAC. 2012;.

[14] Juuso E. Integration of intelligent systems in development of smart adaptive systems: linguistic equation approach. Ph.D. thesis, University of Oulu. 2013. 258 pp., http://urn.fi/urn:isbn:9789526202891.

[15] Juuso EK. Recursive Dynamic Modelling in Changing Operating Conditions. In: *Proceedings of the 56th Conference on Simulation and Modelling (SIMS 56), October, 7-9, 2015, Linköping University, Sweden*. Linköping, Sweden: Linköping University Electronic Press. 2015; pp. 169–174.

[16] Juuso EK. Generalised statistical process control GSPC in stress monitoring. *IFAC-PapersOnline*. 2015; 48(17):207–212.

[17] Juuso E. Intelligent trend analysis for a solar thermal energy collector field. vol. 136. 2018; .

[18] Juuso EK. Recursive Data Analysis and Modelling in Prognostics. In: *12th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2015 - MFPT 2015, 9-11 June 2015, Oxford, UK*, pp. 560–567. BINDT. 2015; ISBN: 978-1-5108-0712-9.
URL `www.proceedings.com`

[19] Juuso EK. Intelligent Performance Analysis with a Natural Language Interface. *Management Systems in Production Engineering*. 2017;25(3):168–175.

[20] Juuso EK. Data-based development of dynamic models for biological wastewater treatment in pulp and paper industry. In: *SIMS 2010 Proceedings, The 51st Conference on Modelling and Simulation, Oulu, 14-15 October, 2010*. 2010;9 pp.

[21] Juuso EK, Ruusunen M. Stress Indices in Fatigue Prediction. In: *International Conference on Maintenance, Condition Monitoring and Diagnostics, and Maintenance Performance Measurement and Management, MCMD 2015 & MPMM 2015, Oulu, Finland. 30 September - 1 October 2015*. Oulu, Finland: POHTO. 2015;8 pp., ISBN 978-951-98113-7-6.

[22] Juuso EK. Advanced prognostics based on intelligent data analysis. In: *WCCM 2017 - 1st World Congress on Condition Monitoring 2017*, vol. 2. Red Hook, NY: Curran Associates. 2017; pp. 782–794.