

Recursive dynamic modelling in changing operating conditions

Esko K. Juuso

Control Engineering Group, Faculty of Technology, University of Oulu, Finland
esko.juuso@oulu.fi

Abstract

Changing operating condition may require updates for the dynamic models. Recursive updates are needed especially when sufficient information about the new situations is not available. In machine diagnostics and prognostics, the analysis starts from good conditions and new phenomena activated with time may considerably change the model. In biological wastewater treatment processes, the condition of the biomass drastically changes during the operation of the treatment process. Measurements based on the image analysis of the biomass condition is recently included. Recursive modelling is clearly needed in these situations. The usual approach is to modify the model equations. However, the interactions do not necessarily change if the meanings of the variables are modified. This paper keeps the model equations constant and modifies the non-linear scaling of the variables by extending the data-driven scaling to a recursive approach. The recursive methodology is tested in two applications: machine diagnostics and wastewater treatment.

Keywords: intelligent modelling, recursive statistical analysis, adaptive modelling, prognostics, transitions

1 Introduction

Changing operating conditions need to be taken into account in prognostics since new phenomena activate gradually with time. In the condition-based maintenance (CBM), the most obvious and widely used form of prognostics is to predict how much time is left before a failure occurs. The time left before observing a failure is usually called remaining useful life (RUL) (Jardine et al., 2006). It is even better if the maintenance actions are taken before the failure occurs, especially when a fault or a failure is catastrophic. Dynamic models include hazard functions and stochastic processes for the process states which activate with time. A time-dependent proportional hazard model (PHM) has a hazard function of the form

$$h(t) = h_0(t) \exp\left(\sum_{j=1}^n \eta_j x_j(t)\right) \quad (1)$$

where $h_0(t)$ is a baseline hazard function, $x_j(t)$, $j = 1, \dots, m$, are covariates which are functions of time and

$\eta_j(t)$, $j = 1, \dots, m$ are coefficients. The baseline hazard function $h_0(t)$ can be in non-parametric or parametric form, e.g. a Weibull hazard function, which is the hazard function of the Weibull distribution. The covariates $x_j(t)$, $j = 1, \dots, m$, can be any condition variables such as health indicators and features in condition monitoring. (Jardine et al., 2006)

In biological wastewater treatment in pulp and paper industry, a lot of process measurements are available, but measurement sets do not include sufficient information on the special features of the influent nor on the microbial composition of the sludge (Juuso, 2009a). Activated Sludge Models provide a basis for phenomenological modelling and can be linked to process expertise (Henze et al., 1987; Lindblom, 2003). Their use has been limited by complexity of the models. Hybrid models with a cascade approach are needed in biological wastewater treatment to cover different operating conditions. The condition of the biomass need to modelled as well (Juuso, 2009a).

Many variables are normally measured in a plant, but some of them are strongly cross-correlated. Data-based analysis is needed for variable selection (Teppola et al., 1997; Mujunen et al., 1998; Oliveira-Esquerre et al., 2002) and for detection of operating conditions has used in (Heikkinen et al., 2008b) to analyse specialised submodels. As the sludge settling properties have remarkable effects on the treatment results, the modelling of the diluted sludge volume index (DSVI) is important (Heikkinen et al., 2008b). Heikkinen et al. (2008a) used models for predicting the chemical oxygen demand (COD) of the effluent in an industrial activated sludge plant. Image analysis has been studied in (Tomperi et al., 2015) to find out dependencies to the process variables of the plant and to the quality of the treated wastewater.

The main idea of time series modelling is to fit the waveform data to a parametric time series model and extract features based on this parametric model, where the output at time t can depend on many signal values y from previous time instants. The amount of signals should be chosen according to the appropriate time delays (Ljung, 1999). Fuzzy and neural models are based on the same structures. The most common structure for the input-output models is the NARX /Nonlinear AutoRegressive

with eXogenous input model, where the input and output values are chosen according to appropriate system orders, as in the ARX model. The regressor vector consists of a finite number of past inputs and outputs (Babuška and Verbruggen, 2003). Another possibility is to use recurrent networks, e.g. Elman networks are two-layer feedforward networks, with the addition of a feedback connection from the output of the hidden layer to its input (Elman, 1990).

The basic form of the linguistic equation (LE) model is a static mapping in the same way as fuzzy set systems and neural networks, and therefore dynamic models will include several inputs and outputs originating from a single variable (Juuso, 2004). External dynamic models provide the dynamic behaviour, and LE models are developed for a defined sampling interval in the same way as in various identification approaches discussed in (Ljung, 1999). Nonlinear scaling reduces the number of input and output signals needed for the modelling of nonlinear systems. For the default LE model, all the degrees of the polynomials become very low:

$$Y(t) + a_1 Y(t-1) = b_1 U(t - n_k) + e(t) \quad (2)$$

for the scaled variables Y and U .

The nonlinear scaling maps from the operation area of the (sub)system, defined with feasible ranges, to the linguistic values represented inside a real-valued interval $[-2, 2]$. The feasible range is defined by a membership function, and membership functions for finer partitions can be generated from membership definitions (Juuso et al., 1993). The basic scaling approach presented in (Juuso, 2004) has been improved later: a new constraint handling was introduced in (Juuso, 2009b), and a new skewness based methodology was presented for signal processing in (Juuso and Lahdelma, 2010).

Recursive modelling can be done by including new model components as in (1) or by modifying the coefficients as is done in identification. The highly complex model structures, which are based on data fitting, can be difficult to understand. In LE models, the structure is kept compact, see (2), to keep process insight clear.

This paper combines compact model structures with the data-driven analysis of variable meanings and discusses results and requirements in two applications: machine diagnostics and wastewater treatment.

2 Nonlinear scaling

Scaling functions are 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 $[-2, 2]$, denoted linguistification. In LE models, the results are scaled to the real values by using the function $f()$.

The parameters of the functions are extracted from measurements by using generalised 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 $j, j = 1, \dots, m$. 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.

2.1 Generalised norms

The corner points can be extracted from measurements with generalised norms defined by

$$\|\tau M_j^p\|_p = (\tau M_j^p)^{1/p} = \left[\frac{1}{N} \sum_{i=1}^N (x_j)_i^p \right]^{1/p}, \quad (3)$$

where $p \neq 0$, is calculated from N values of a sample, τ is the sample time. With a real-valued order $p \in \mathfrak{R}$ this norm can be used as a central tendency value if $\|\tau M_j^p\|_p \in \mathfrak{R}$, i.e. $x_j > 0$ when $p < 0$, and $x_j \geq 0$ when $p > 0$. For variables with only negative values, the norm is the opposite of the norm obtained for the absolute values. If a variable has both positive and negative values, each norm is an average of two norms where the data sets are made positive and negative by subtracting a value $x_L < \min((x_j))$ and a value $x_H > \max(x_j)$, respectively (Juuso, 2011). These norms generalise the earlier analysis based on arithmetic means or medians (Juuso, 2004).

The computation of the norms can be divided into the computation of equal sized sub-blocks, i.e. the norm for several samples can be obtained as the norm of the norms of the individual samples:

$$\|^{K_s} \tau M_j^p\|_p = \left\{ \frac{1}{K_s} \sum_{i=1}^{K_s} [(\tau M_j^p)_i]^{1/p} \right\}^{1/p} = \left[\frac{1}{K_s} \sum_{i=1}^{K_s} [(\tau M_j^p)_i]^{1/p} \right]^{1/p}, \quad (4)$$

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

2.2 Parameters

The generalised skewness γ_3^p is used when choosing appropriate methods for defining the central tendency. The central tendency value and the core area are based on the normalised moment

$$\gamma_k^p = \frac{1}{N \sigma_j^k} \sum_{i=1}^N [(x_j)_i - \|\tau M_j^p\|_p]^k \quad (5)$$

where k is a positive integer. The standard deviation σ_j , which is calculated about the origin, is used to obtain a dimensionless feature. (Juuso and Lahdelma, 2010) The

central tendency value is chosen by the point where the skewness changes from positive to negative, i.e. $\gamma_3^p = 0$. Then the data set is divided into two parts: a lower part and an upper part. The same analysis is done for these two data sets. The estimates for the corner points, $(c_l)_j$ and $(c_h)_j$, are the points where $\gamma_3^p = 0$ for the lower and upper data sets, respectively.

3 Recursive modelling

Recursive data analysis facilitates the adaptation of the functions to changing operating conditions, also the orders of the norms are re-analysed if needed. The existing scaling functions provide a basis for assessing the quality of the new data: outliers should be excluded, but the suspicious values may mean that the operating conditions are changing. In this research, the scaling functions are extended for analysing outliers and suspicious values to select data for the adaptive scaling. Different operating areas can be analysed from previous measurements with various clustering methods and the statistical process control (SPC) provides additional tools for detecting changes, anomalies and novelties in connection with the operation.

3.1 Data selection

Process data often contains outliers, which must be removed before generating the feasible area, because the procedure described above is sensitive to them. This is the idea in medians and trimmed means, which are used for the data samples containing outliers. A good estimate for the support area can be obtained with the generalised norms (3) with large negative and large positive orders since these features are less sensitive to the outliers than the minimum and maximum values. Discarding values at the high and low end can be used together with the generalised norms if there are obvious outliers. Trimming does not need to be the same for the low and high values.

Clear outliers need to be excluded in both the first analysis and the subsequent adaptation steps. In linear scaling, the z-score values outside the range $[-3, 3]$ are often considered as an indication of an outlier. This is a feasible solution for the normal distribution, but it is not optimal for steeper and wider distributions. The effect of the shape factor is presented in Figure 1.

For LE models, the scaled values X_j are in the range $[-2, 2]$ (Figure 2), and this is also the range for the monotonous increase if $\alpha_j^- = \alpha_j^+ = \frac{1}{3}$. The minimum and maximum points depend on the shape factors: the maximum point in the linguistic range

$$(X_j)_{max} = -\frac{b_j^+}{2a_j^+} = \frac{3 - \alpha_j^+}{2(\alpha_j^+ - 1)}, \quad (6)$$

which goes to infinity when $\alpha_j^+ \rightarrow 1$, and the upper polynomial does not have any maximum point when $\alpha_j^+ > 1$.

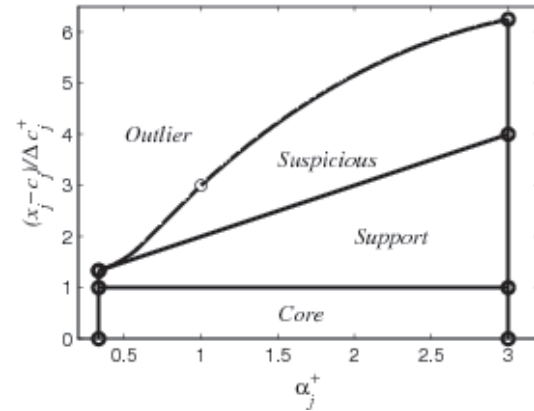


Figure 1. Relative score values for the core, support, suspicious and outlier areas as a function of the shape factor $\alpha_j^+ = 3$.

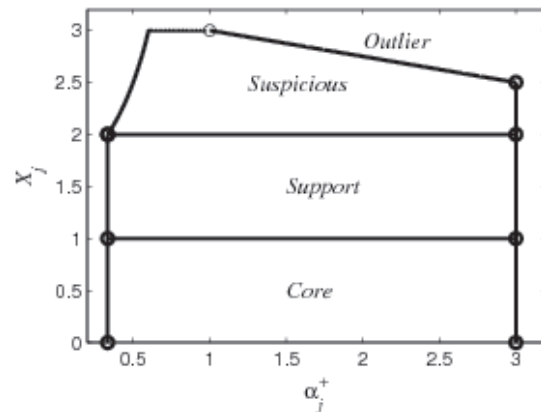


Figure 2. Scaled score values for the core, support, suspicious and outlier areas as a function of the shape factor $\alpha_j^+ = 3$.

3.2 Adaptation

The parameters of the nonlinear scaling functions can be recursively updated with (4) 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 orders $(p_l)_j$, $(p_0)_j$ and $(p_h)_j$ of the corresponding norms are re-analysed if the distribution is changing considerably with new measurements.

Since the search of these points is performed by using the order of the moment, the resulting orders $(p_l)_j$, $(p_0)_j$ and $(p_h)_j$ are good estimates when additional data sets are used. The norm values are recursively updated with (4), and a new search for the orders is done only if the values change considerably (Juuso, 2011).

Data values are classified with fuzzy logic to three categories: (1) extended area, (2) new operating area, or

(3) outliers. Decision is based on the difference of the feasibility areas of the recent data and the current definitions. Outliers were discussed in the previous section. High overlap supports the extension alternative. Lower and lower overlap proposes a new operating area. The approach goes smoothly and requires several new data points to change the parameters. The window of the short recent time period depends on the process. The decisions should be based on expert knowledge, especially when really drastic changes are detected.

3.3 LE Models

The same model, e.g. (2), is used all the time, even with same coefficients. The recursive adaptation is principally done by updating for the parameters of the scaling functions. The coefficients are updated in the second level, and naturally, it also possible to include more equations if needed. The solution is more compact than repeated regression analysis to get the coefficients of the models. Naturally, the scaling approach can be enhanced with regression if needed.

4 Applications

The approach is aimed for processes where the online modelling is needed to adapt in new situations without sufficient information needed in predefined adaptation.

4.1 Prognostics

Remaining useful life (RUL) is a conditional random variable, which is estimated by using the current age and the past condition profile up to the current moment. The failure is assumed as a case where the fault defined by the condition variables reaches a predetermined level. The need for detailed prognostics could be detected by monitoring indices with SPC. The indication could be given when the index exceeds the warning level. (Lahdelma and Juuso, 2011)

Trend monitoring of vibration features is a more useful maintenance tool than a one-time survey of absolute magnitudes (Kaufman, 1975). Models have been presented for the increase of vibrations with the machine operation time, e.g. according to (Sankar and Xistris, 1972; Collacott, 1977) the increase is first linear and turns into exponential when 75 % of the machine life time is exceeded. However, the life time is not known as it varies from machine to machine. Degradation curves can be based on time series analysis where time series models are developed to trend the condition variables for failure prediction. A dynamic model can also be used. Heng et al. (2009) used a feedforward neural network.

Trends from two paper machines are analysed in (Lahdelma and Juuso, 2011). In the first case, the resonance of the press section resulted in a fast increase of the

root mean square velocity, v_{rms} . The machine speed was reduced 4 %, and a breakdown and an additional stoppage were avoided. The machine operated with reduced speed for two weeks, and the same or lower vibration level was kept for one week. In the second case, resin problems of a press roll in the felt washer were seen as a typical trend represented by

$$\ln \frac{v_{rms}(t)}{(v_{rms})_0} = b_0 + b_1 t + b_2 t^2, \quad (7)$$

where b_0 , b_1 and b_2 are constants. The velocity v_{rms} is growing with an accelerating speed (Figure 3). The models for the periods one and two are linear in a semilogarithmic scale, but the model for the whole data set requires the quadratic term.

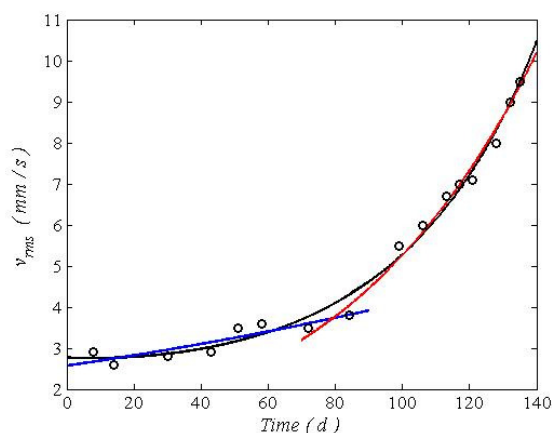


Figure 3. Exponential models v_{rms} (Lahdelma and Juuso, 2011).

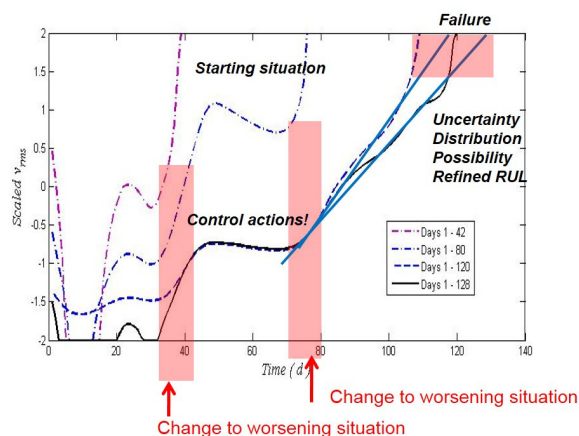


Figure 4. Recursive adaptation in prognostics.

In prognostics, the scaling functions are expanding when the new phenomena activate (Figure 4). The first models do not predict a failure: they are only used to detect the need for the update. Later, the LE models, which are in Figure 4 shown as lines, predict well the failure. In the beginning, the time window includes only good operation. Since the window is continuously increased, it contains the whole data only in the end.

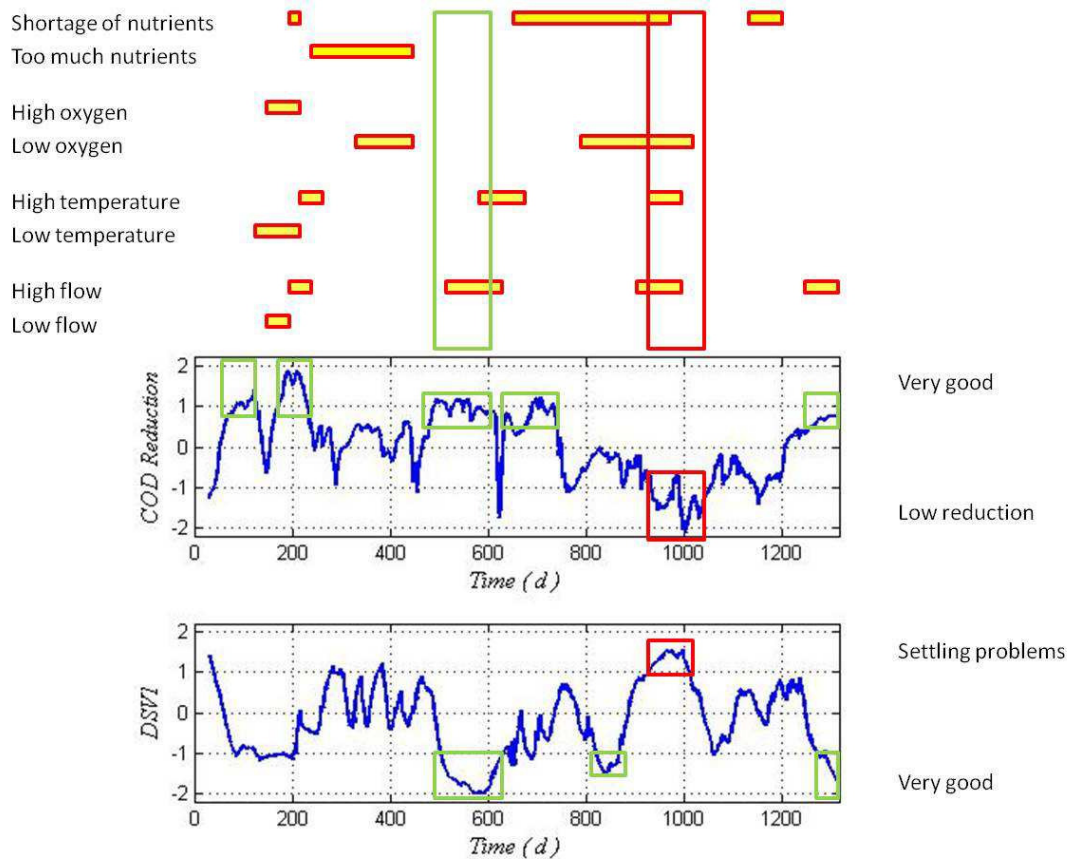


Figure 5. Treatment results of an activated sludge treatment plant: reduction of chemical oxygen demand (COD) and diluted sludge volume index (DSVI), see (Juuso, 2010).

In prognostics, the data points do not always cover the whole area of operation, e.g. only the close neighbourhood of the normal operation point may be covered, or we would like to extend the model of upper part later to the lower part. In fault diagnosis, only one part may be in use. In these cases, expert knowledge is used in extending the feasible range or selecting the methodologies.

4.2 Wastewater treatment

Biological water treatment depends strongly on the operating conditions. Load and nutrient levels should be balanced since both an exceptionally high load and excess nutrients cause problems. The operating conditions are modified by oxygen, temperature and flow. The worst case, which includes both low reduction and settling problems, arise when there multiple warnings and alarms (Figure 5). Correspondingly, good reduction and very good settling is achieved when there were very few warnings. As it takes some time to lose good conditions and recover from problematic conditions, the intelligent indices are useful for supervisory control.

In wastewater treatment, all the alternative decisions are feasible. Drastic changes take place, but they start with smooth changes (Figure 5). Strong transitions are especially seen in settling: different operating conditions are

seen in the parameters of the scaling functions of the measurements: temperature, oxygen, flow and nutrient feed. A moving window is suited for these applications to recursively update the parameters of the scaling functions. The coefficients and variables of the model are the same in all operating conditions. The new measurements reported in (Tomperi et al., 2015) are included in the analysis.

5 Conclusions

Changing operating condition can be handled by updating the scaling functions without specific information about the new situations. The recursive analysis operates well and the nature of the process can be taken into account. In machine diagnostics and prognostics, the scaling functions are expanded when new phenomena activate. In wastewater treatment, drastic changes require completely new parameters. The model equations are not changed.

Acknowledgment

This work was done within the research program “Measurement, Monitoring and Environmental Efficiency Assessment (MMEA)” funded by the TEKES (the Finnish Funding Agency for Technology and Innovation).

References

- R. Babuška and H. Verbruggen. Neuro-fuzzy methods for non-linear system identification. *Annual Reviews in Control*, 27(1):73–85, 2003.
- R. A. Collacott. *Mechanical Fault Diagnosis and Condition Monitoring*. Chapman and Hall, London, 1977.
- J. L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.
- M. Heikkinen, T. Heikkinen, and Y. Hiltunen. Modelling of activated sludge treatment process in a pulp mill using neural networks. In *The 6th International Conference on Computing, Communications and Control Technologies: CCCT 2008, Orlando, Florida, USA, June 29th - July 2nd 2008.*, page 6 pp. 2008a.
- M. Heikkinen, T. Latvala, E. Juuso, and Y. Hiltunen. SOM based modelling for an activated sludge treatment process. In *Tenth International Conference on Computer Modelling and Simulation, EUROSIM/UKSim, Cambridge, UK, April 13, 2008.*, pages 224–229. The Institute of Electrical and Electronics Engineers IEEE, 2008b. doi: 10.1109/UKSIM.2008.78.
- A. Heng, A. C. C. Tan, J. Mathew, N. Montgomery, D. Banjevic., and A. K. S. Jardine. ‘intelligent condition-based prediction of machinery reliability. *Mechanical Systems and Signal Processing*, 23(5):1600–1614, 2009.
- M. Henze, C. P. L. Grady Jr., W. Gujer, G. V. R. Marais, and T. Matsuo. Activated sludge model no. 1., IAWQ scientific and technical report no. 1. London, UK, 1987.
- A. K. S. Jardine, D. Lin, and D. Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.
- E. Juuso and S. Lahdelma. 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*, volume 2, pages 1358–1372, 2010. URL www.scopus.com.
- E. K. Juuso. Integration of intelligent systems in development of smart adaptive systems. *International Journal of Approximate Reasoning*, 35(3):307–337, 2004. doi: 10.1016/j.ijar.2003.08.008.
- E. K. Juuso. Hybrid models in dynamic simulation of a biological water treatment process. In J. Kunovský, P. Hanáček, F. Zboril, Al-Dabass, and A. Abraham, editors, *Proceedings First International Conference on Computational Intelligence, Modelling and Simulation, 7- 9 September 2009, Brno, Czech Republik*, pages 30–35. IEEE Computer Society, 2009a. doi: 10.1109/CSSim.2009.52.
- E. K. Juuso. Tuning of large-scale linguistic equation (LE) models with genetic algorithms. In M. Kolehmainen, editor, *Revised selected papers of the International Conference on Adaptive and Natural Computing Algorithms - ICANNGA 2009, Kuopio, Finland, Lecture Notes in Computer Science*, volume LNCS 5495, pages 161–170. Springer-Verlag, Heidelberg, 2009b. doi: 10.1007/978-3-642-04921-7_17.
- E. K. Juuso. 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.
- E. K. Juuso. Recursive tuning of intelligent controllers of solar collector fields in changing operating conditions. In S. Bittani, A. Cenedese, and S. Zampieri, editors, *Proceedings of the 18th World Congress The International Federation of Automatic Control, Milano (Italy) August 28 - September 2, 2011*, pages 12282–12288. IFAC, 2011. doi: 10.3182/20110828-6-IT-1002.03621.
- E. K. Juuso, J.C. Bennavil, and M.G. Singh. Hybrid knowledge-based system for managerial decision making in uncertainty environment. In N. Piera Carreté and M. G. Singh, editors, *Qualitative Reasoning and Decision Technologies, Proceedings of the IMACS International Workshop on Qualitative Reasoning and Decision Technologies -QUARDET’93, Barcelona, June 16 - 18, 1993*, pages 234–243, Barcelona, 1993. CIMNE.
- A. B. Kaufman. ‘measure machine vibration – it can help you anticipate and prevent failures. *INSTRUMENTS & CONTROL SYSTEMS*, 48:50–62, 1975.
- S. Lahdelma and E. Juuso. Trend analysis in condition monitoring of process equipments. In *Proceedings of the 8th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2011 - MFPT 2011, 20-22 June 2011, Cardiff, UK*, volume 2, pages 904–913. Curran Associates, NY, USA, 2011. ISBN 978-1-61839-014-1.
- E. Lindblom. *Dynamic modelling of nutrient deficient wastewater treatment process. M.Sc. Thesis*. Lund University, Lund, Sweden, 2003. TEIE-5175.
- L. Ljung. *System Identification - Theory for the User*. Prentice Hall, Upper Saddle River, N.J., 2nd edition, 1999.
- S.-P. Mujunen, P. Minkkinen, P. Teppola, and R.-S. Wirkkala. Modeling of activated sludge plants treatment efficiency with PLSR: a process analytical case study. *Chemometrics and Intelligent Laboratory Systems*, 41(1):83–94, 1998.
- K. P. Oliveira-Esquerre, M. Mori, and R. E. Bruns. Simulation of an industrial wastewater treatment plant using artificial neural networks and principal component analysis. *Brazilian Journal of Chemical Engineering*, 19(4):365–370, 2002.
- T. S. Sankar and G. D. Xistris. Measure machine vibrations - It can help you anticipate and prevent failures. *Journal of Engineering for Industry*, 94:133–137, 1972.
- P. Teppola, S.-P. Mujunen, and P. Minkkinen. Partial least squares modeling of an activated sludge plant: A case study. *Chemometrics and Intelligent Laboratory Systems*, 38(2):197–208, 1997.
- J. Tomperi, E. Koivuranta, A. Kuokkanen, E. K. Juuso, and K. Leiviskä. Real-time optical monitoring of the wastewater treatment process. *Environmental Technology*, pages 1–8, 2015. doi:10.1080/09593330.2015.1069898.