Intelligent Methods in Modelling and Simulation of Complex Systems

Esko K. Juuso*

Control Engineering Laboratory Department of Process and Environmental Engineering, P.O.Box 4300, 90014 University of Oulu, Finland; *Esko.Juuso@oulu.fi

Simulation Notes Europe SNE 24(1), 2014, 1 - 10
DOI: 10.11128/sne.24.on.102221
Received: Jan.10, 2014 (Selected SIMS 2013 Postconf. Publ.);
Accepted: February 15, 2014;

Abstract. Data mining with a multitude of methodologies is a good basis for the integration of intelligent systems. Small, specialised systems have a large number of feasible solutions, but developing truly adaptive, and still understandable, systems for highly complex systems require domain expertise and more compact approaches at the basic level. This paper focuses on the integration of methodologies in the smart adaptive applications. Statistical methods and artificial neural networks form a good basis for the data-driven analysis of interactions and fuzzy logic introduces solutions for knowledge-based understanding the system behaviour and the meaning of variable levels.

Efficient normalisation, scaling and decomposition approaches are the key methodologies in developing large-scale applications. Linguistic equation (LE) approach originating from fuzzy logic is an efficient technique for these problems.

The nonlinear scaling methodology based on advanced statistical analysis is the corner stone in representing the variable meanings in a compact way to introduce intelligent indices for control and diagnostics. The new constraint handling together with generalised norms and moments facilitates recursive parameter estimation approaches for the adaptive scaling.

Well-known linear methodologies are used for the steady state, dynamic and case-based modelling in connection with the cascade and interactive structures in building complex large scale applications. To achieve insight and robustness the parameters are defined separately for the scaling and the interactions.

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 (Figure 1).

Adaptive systems can be developed for nonlinear multivariable systems by various statistical and intelligent methodologies, which are in sensor fusion combined with data pre-processing, signal processing and feature extraction [14]. Fault diagnosis is based on symptoms generated by comparing process models and measurements [15], signal analysis [38], limit checking of measurements [14] and human observations [19]. All these are used in intelligent control and detection of operating conditions, which introduce reasoning and decision making to the smart adaptive systems, whose hybrid nature is seen in literature where these topics are combined from different perspectives.

The linguistic equation (LE) approach originates from fuzzy set systems [35]: rule sets are replaced with equations, and meanings of the variables are handled with scaling functions which have close connections to membership functions [27]. The nonlinear scaling technique is needed in constructing nonlinear models with linear equations [28]. New development methodologies [32,23], improve possibilities to update the scaling functions recursively [33,24].

This paper classifies methodologies and focuses on combining advanced statistical analysis and soft computing in developing LE based applications for complex systems.

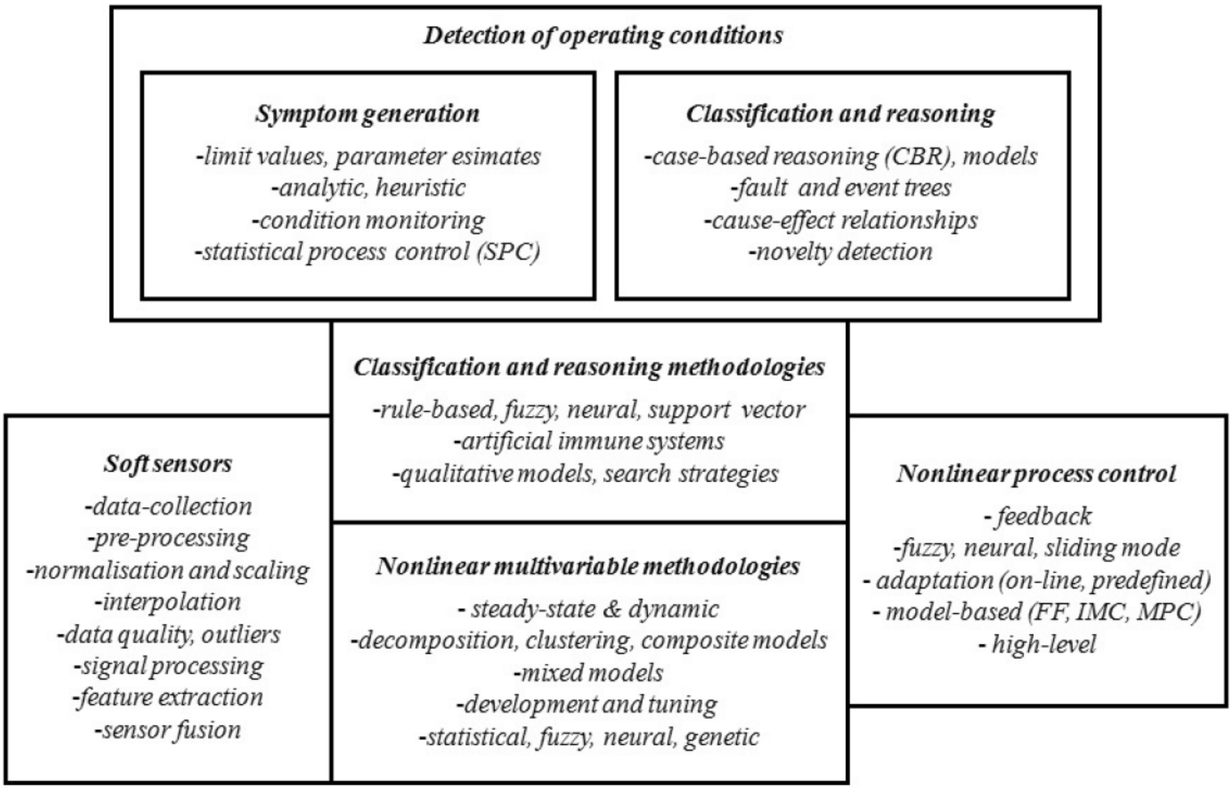

Figure 1: Methodologies for modelling of complex system.

1 Steady-State Modelling

The steady-state simulation models can be relatively detailed nonlinear *multiple input, multiple output* (*MIMO*) models $\boldsymbol{y} = F(\boldsymbol{x})$, where the output vector $\boldsymbol{y} = (y_1, y_2, \dots, y_n)$ is calculated by a nonlinear function F from the input vector $x = (x_1, x_2, \dots, x_m)$. More generally, the relationship could also be a table or a graph. Fuzzy set systems, artificial neural networks and neurofuzzy methods provide additional methodologies for the function $F(x)$.

Statistical modelling in its basic form uses linear regression for solving coefficients for linear functions. In the *response surface methodology* (*RSM*), the relationships are represented with *multiple input, single output* (*MISO*) models, which contain linear, quadratic and interactive terms [5]. Application areas of the linear modelling can also be extended by arbitrary nonlinear models, e.g. semi-physical models, developed by using appropriate calculated variables as inputs, see [39]. Principal component analysis (PCA) compresses the data by reducing the number of dimensions: each principal component is a linear combination of the original variables, usually the first few principal components are used. Various extensions of PCA are referred in [21]. Partial least squares regression (PLS) uses potentially collinear variables [17].

Fuzzy logic emerged from approximate reasoning, and the connection of fuzzy rule-based systems and expert systems is clear, e.g. the vocabulary of AI is kept in fuzzy logic [13]. *Fuzzy set theory* first presented by Zadeh (1965) form a conceptual framework for linguistically represented knowledge. *Extension principle* is the basic generalisation of the arithmetic operations if the inductive mapping is a monotonously increasing function of the input. The interval arithmetic presented by Moore (1966) is used together with the extension principle on several membership α-cuts of the fuzzy number x_j for evaluating fuzzy expressions [6-8]. The fuzzy sets can be modified by intensifying or weakening modifiers [11]. *Type-2 fuzzy* models introduced by Zadeh in 1975 take into account uncertainty about the membership function [42]. Most systems based on interval type-2 fuzzy sets are reduced to an interval-valued type-1 fuzzy set.

Linguistic fuzzy models [12], where both the antecedent and consequent are fuzzy propositions, suit very well to qualitative descriptions of the process as they can be interpreted by using natural language, heuristics and common sense knowledge. *Takagi-Sugeno (TS) fuzzy models* [48], where each consequent $y_i, i = 1, \ldots, n$ is a crisp function of the antecedent variables $\boldsymbol{x}$, can be interpreted in terms of local models. For linear functions, the standard weighted mean inference must be extended with a smoothing technique [2]. *Singleton models*, where the consequents are crisp values, can be regarded as special cases of both the linguistic fuzzy models and the TS fuzzy models. *Fuzzy relational models* [44] allow one particular antecedent proposition to be associated with several different consequent propositions.

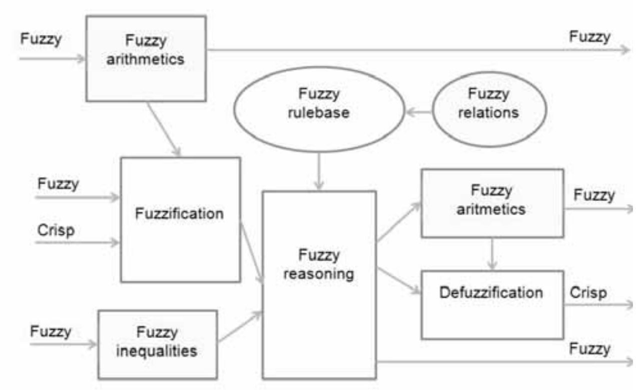

Figure 2: Combined fuzzy modelling.

Several fuzzy modelling approaches are combined in Figure 2: fuzzy arithmetics is suitable both for processing fuzzy inputs for the rule-based fuzzy set system and the fuzzy outputs; fuzzy inequalities produce new facts like $A \leq B$ and $A = B$ for fuzzy inputs A and B; fuzzy relations can be represented as sets of alternative rules, where each rule has a degree of membership.

Artificial neural networks (*ANN*) are used as behavioural input-output models consisting of neurons. Network architectures differ from each other in their way of forming the net input, use of activation functions and number of layers. *Linear networks* correspond to the models with linear terms in RSM models. The most popular neural network architecture is the *multilayer perceptron* (*MLP*) with a very close connection to the *backpropagation learning* [45].

Neurofuzzy systems use fuzzy neurons to combine the weight factors and the inputs (Figure 3). The activation function is handled with the extension principle from the fuzzy input, which is obtained by the fuzzy arithmetics [16]. Also cascade architectures of fuzzy set systems and neural networks are often called neurofuzzy systems. Neural computation is used for tuning fuzzy set systems which can be represented by neural networks, see [20].

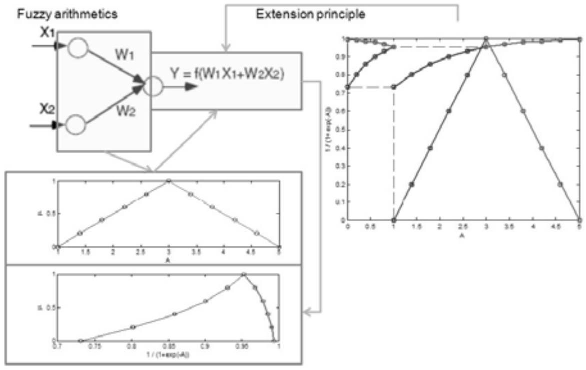

Figure 3: A fuzzy neuron.

A *function expansion* presented in [40] provides flexible way to present several types of black box models by using basis functions, which are generated from one and the same function characterised by the scale (dilation) and location (translation) parameters. The expansion can contain, for example, radial basis functions, one-hidden-layer sigmoidal neural networks, neurofuzzy models, wavenets, least square support vector machines, see [39].

Approximate reasoning based on T-norms and S-norms, also called T-conorms, is an essential part of combining antecedents and rules in fuzzy logic [12]. T-norms and S-norms can be used in neurofuzzy systems if the inputs are normalised to the range [0,1] [16].

2 Decomposition Methodologies

A modelling problem can be divided into smaller parts by developing separate models for independent subprocesses (Figure 4). Cluster analysis can be used in the data-driven approach. Composite local models can be used, and fuzzy set systems provide feasible techniques for handling the resulting partially overlapping models. The system may also include models based on the first principles.

A process plant consists of several processing units interconnected with process streams. *Decomposition* can be continued within process units. In an electric furnace presented in [22], a cylindrically symmetrical one-electrode model was based on two-dimensional areas defined by overlapping rectangular grids where the amount of detail can be increased in selected parts [25]. In addition to spatial or logical blocks decomposed modelling can be based on different frequency ranges.

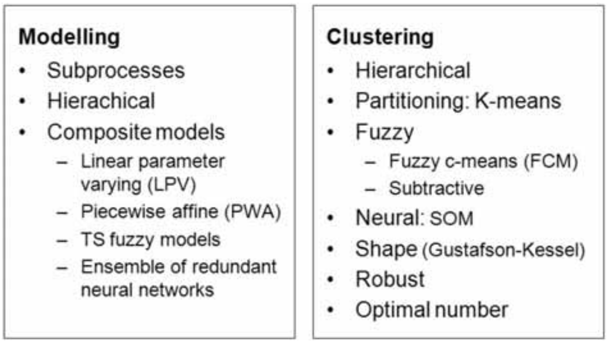

Figure 4: Decomposition for modelling.

Hundreds of clustering algorithms have been developed for the data-driven analysis by researchers from a number of different scientific disciplines. *Hierarchical clustering* groups data by creating a cluster tree, where clusters at one level are joined as clusters at the next higher level. *Partitioning-based clustering algorithms*, e.g. K-means, minimise a given clustering criterion by iteratively relocating data points between clusters until a (locally) optimal partition is attained.[1].

Numerous *fuzzy clustering* algorithms have been proposed and applied to a variety of real-world problems [4]. *Fuzzy c-means (FCM)* clustering is a partitioning-based method: each data point belongs to a cluster to some degree membership. *Subtractive clustering* [10] is an algorithm for estimating the number of clusters and the cluster centres according to the parameters of the algorithm. *Neural clustering* use competitive networks based on competitive layers, e.g. self-organising maps (SOM) [36] have several alternatives for calculating the distance in the competitive layer. The response of a radial basis functions (RBF) neuron is obtained from an exponential function [9].

The clustering algorithms have limitations in shape, cluster centres and generalisation of the results. The algorithm with the standard Euclidean norm imposes a spherical shape on the clusters, regardless of the actual data distribution [2]. Gustafson and Kessel (1979) extended the standard by employing an adaptive distance norm to detect clusters of different geometrical shapes. *Robust clustering*, which is based on a spatial median, is aimed for problems where classical clustering methods are too sensitive to erroneous and missing values [1]. Optimal *number of clusters* is selected iteratively by using some quality criteria, see [49].

Composite local model approach constructs a global model from local models, which usually are linear approximations of the nonlinear system in different neighbourhoods. If the partitioning is based on a measured regime variable, the partitioning can be used in weighting the local models. *Linear parameter varying (LPV) models*, where the matrices of the state-space model depend on an exogeneous variable measured during the operation, are close related to local linear models [40]. *Piecewise affine (PWA) systems* are based on local linear models, more specifically in a polyhedral partition [47]. The models can be state-space models or parametric models. The model switches between different modes as the state variable varies over the partition [40].

Fuzzy models can be considered as a class of local modelling approaches, which solve a complex modelling problem by decomposing into number of simpler understandable subproblems [2,3]. The smoothing problem around the submodel borders of TS fuzzy models needs special techniques, e.g. smoothing maximum [2], or by making the area overlap very strong. *Multiple neural network systems* improve generalisation through task decomposition and an ensemble of redundant networks [46].

A *mixed approach* using both the rigorous first principles and the black box modelling in an integrated environment is an interesting alternative for complex systems, see [41]. [40] classifies the models as a palette of grey shades from white to black into six categories: first principles, identified parameters, semi-physical models, composite models, block oriented models, and black box models. In semi-physical models, linear modelling used together with nonlinear transformations which are based on process insight.

3 Adaptive Nonlinear Scaling Membership

Membership definitions provide nonlinear mappings 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 [34]. The basic scaling approach presented in [28] has been improved later: a new constraint handling was introduced in [32], and a new skewness based methodology was presented for signal processing in [23].

3.1 Working point and feasible ranges

The concept of feasible range is defined as a trapezoidal membership function. In the fuzzy set theory [51], support and core areas are defined by variable, x_i, specific subsets,

$$\operatorname{supp}(F_j) = \left\{x_j \in U_j \,|\, \mu_{F_j}(x_j) > 0\right\}, \tag{1}$$

$$\operatorname{core}(F_j) = \left\{x_j \in U_j \,|\, \mu_{F_j}(x_j) = 1\right\}, \tag{2}$$

where U_j is an universal set including F_j; $\mu_{F_j}(x_j)$ is the membership value of x_j in F_j. The main area of operation is the core area, and the whole variable range is the support area.

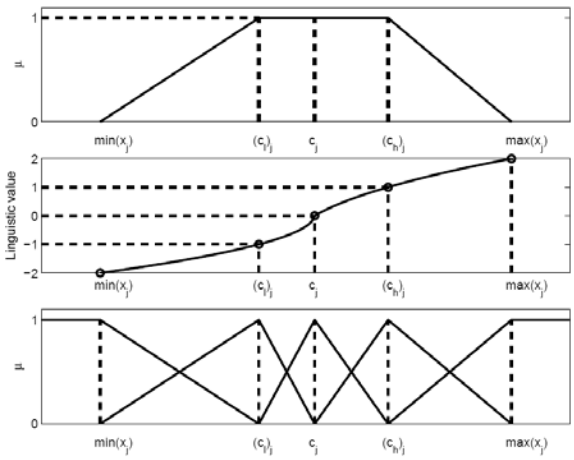

Figure 5: Nonlinear scaling [28]

For applications, a trapezoidal function providing linear transitions between 0 and 1 is sufficient (Figure 5). The corner parameters can be defined on the basis of expert knowledge or extracted from data.

The slope can be different on upper and lower part depending on the linearity or nonlinearity of the system. The complement of a fuzzy set is defined as a subset [51]

$$\bar{F}_j = \left\{x_j \in U_j \,|\, \mu_{\bar{F}_j}(x_j) = 1 - \mu_F(x_j)\right\}, \tag{3}$$

where $\mu_{\bar{F}_j}(x_j)$ is the membership value of x_j in $\bar{F}_j$. The membership function of the complement corresponds to the highest and lowest membership functions (Figure 5).

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.

The corner points can be extracted from existing rule-based fuzzy systems or defined manually. Feasible ranges should be consistent with membership definitions, and therefore they are defined together in the data-driven approach. Earlier the analysis of the corner points and the centre point has been based on the arithmetic means or medians of the corresponding data sets [28]. The norm defined by

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

where $p \neq 0$, is calculated from N values of a sample, τ is the sample time. With a real-valued order $p \in \Re$ this norm can be used as a central tendency value if $\left\| {}^{\tau}M_j^p \right\|_p \in \Re$, i.e. $x_j > 0$ when $p < 0$, and $x_j \geq 0$ when $p > 0$. The norm (4) is calculated about the origin, and it combines two trends: a strong increase caused by the power p and a decrease with the power $1/p$. All the norms have same dimensions as x_j. The norm (4) is a Hölder mean, also known as the power mean. The generalised norm for absolute values $|x_j|$ was introduced for signal analysis in [37].

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 obtained from data sets made positive and negative. [33]

The operating area of each variable is defined by a feasible range represented with a trapezoidal membership function whose corner points are $\min(x_j), (c_l)_j, (c_h)_j$ and $\max(x_j)$. Warnings and alarms can be generated directly from the degrees of membership of the complement (3).

3.2 Membership definitions

A membership definition is defined as a (nonlinear) mapping of variable values x_j inside its range to $X_j \in [-2,2]$, denoted as *linguistic range*. It more or less describes the distribution of variable values over its range. The range $[-2,2]$ includes the normal operation in the range $[-1,1]$ and the areas with warnings and alarms. The values X_j are called *linguistic values* since the scaling idea originates from the fuzzy set systems: values $-2, -1, 0, 1$ and 2 associated to the linguistic labels are defined with membership functions (Figure 5). The number of membership functions is not limited to five: the values between these integers correspond to finer partitions of the fuzzy set system. The early applications of the linguistic equations used only integer values [27].

In present systems, membership definitions are used in a continuous form consisting of two second order polynomials: one for negative values, $X_j \in [-2, 0)$, and one for positive values, $X_j \in [0, 2]$. The polynomials

$$\begin{aligned} f_j^-(X_j) &= a_j^- X_j^2 + b_j^- X_j + c_j, \quad X_j \in [-2,0), \\ f_j^+(X_j) &= a_j^+ X_j^2 + b_j^+ X_j + c_j, \quad X_j \in [0,2], \end{aligned} \tag{5}$$

should be monotonous, increasing functions in order to result in realisable systems. The upper and lower parts should overlap at the linguistic value 0. [28].

The functions are monotonous and increasing if the ratios

$$\begin{aligned} \alpha_j^- &= \frac{(c_l)_j - \min(x_j)}{c_j - (c_l)_j}, \\ \alpha_j^+ &= \frac{\max(x_j) - (c_h)_j}{(c_h)_j - c_j}, \end{aligned} \tag{6}$$

are both in the range $\left[\frac{1}{3}, 3\right]$ see [25]. If needed, the ratios are corrected by modifying the core $[(c_l)_j, (c_h)_j]$ and/or the support $[\min(x_j), \max(x_j)]$.

Errors are checked independently for f^- and f^+: each error can always be corrected either by moving the corner of the core or the support. In some cases, good results can also be obtained by moving c_j. If these constraints allow a non-empty range, the maximum of the lower limits and the minimum of the upper limit are chosen to define the limits for continuous definitions.

The coefficients of the polynomials can be represented by

$$\begin{aligned} a_j^- &= \frac{1}{2}(1 - \alpha_j^-)\Delta c_j^-, \\ b_j^- &= \frac{1}{2}(3 - \alpha_j^-)\Delta c_j^-, \\ a_j^+ &= \frac{1}{2}(\alpha_j^+ - 1)\Delta c_j^+, \\ b_j^+ &= \frac{1}{2}(3 - \alpha_j^+)\Delta c_j^+, \end{aligned} \tag{7}$$

where $\Delta c_j^- = c_j - (c_l)_j$ and $\Delta c_j^+ = (c_h)_j - c_j$. Membership definitions may contain linear parts if some coefficients α_j^- or α_j^+ equals to one.

For each variable, the membership definitions are configured with five parameters, which can be presented with three consistent sets. The working point (centre point) c_j belongs to all these sets, where the other parameters are:

- the corner points (Figure 5) are good for visualisation;
- the parameters $\{\alpha_j^-, \Delta c_j^-, \alpha_j^+, \Delta c_j^+\}$suit for tuning;
- the coefficients $\{a_j^-, b_j^-, a_j^+, \Delta b_j^+\}$ are used in the calculations.

3.3 Adaptation of nonlinear scaling

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 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 with clustering, and statistical process control provide additional tools for detecting changes, anomalies and novelties.

The parameters of the nonlinear scaling functions can be recursively updated by including new equal sized subblocks into calculations. The number of samples 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.

4 Intelligent Systems

Nonlinear models can be constructed by using scaled values in linear modelling based on data and expertise [27,28]. Compact model structures are beneficial in building and tuning dynamic and case-based models for complex systems. The recursive analysis provides new tools for both the adaptation of the scaling functions and the model interactions to changing operating conditions. Linear interactions are used in steady-state models and extended to dynamic systems by parametric structures used in identification. Decomposition of the modelling area is used for case-based systems which can include both steady state and dynamic models. The nonlinear scaling is performed twice: first scaling from real values to the interval $[-2,2]$ before applying linguistic equations, and then scaling from the interval $[-2,2]$ to real values after applying linguistic equations. Variable selection is needed in large-scale systems.

4.1 LE models

The nonlinear scaling with membership definitions transforms the nonlinear model $\boldsymbol{y} = F(\boldsymbol{x})$ to a linear problem. The basic element of a linguistic equation (LE) model is a compact equation

$$\sum_{j=1}^{m} A_{ij} X_j(t - n_j) + B_i = 0 \qquad (8)$$

where X_j is a linguistic value for the variable $j, j = 1 \ldots m$. Each variable j has its own time delay n_j compared to the variable with latest time label. Linguistic values in the range $[-2, 2]$ are obtained from the actual data values by membership definitions. The directions of the interaction are represented by interaction coefficients $A_{ij} \in \Re$.

In the original system [35], the linguistic labels $\{very\ low, low, normal, high,\ \ very\ high\}$ were replaced by numbers $\{-2, -1, 0, 1, 2\}$. The approach was generalized for finer fuzzy partitions in [34]. The bias term $B_i \in \Re$ was first introduced as an additional component in fuzzy LE models [26], and later extended for fault diagnosis systems [28].

The coefficients A_{ij} and B_i in (8) have a relative meaning, i.e. the equation can be multiplied or divided by any nonzero real number without changing the model. A LE model with several equations can be represented as a matrix equation

$$A\boldsymbol{X} + \boldsymbol{B} = 0, \qquad (9)$$

where the interaction matrix A contains all coefficients $A_{ij}, i = 1, \ldots, n, j = 1, \ldots, m,$ and the bias vector $\boldsymbol{B}$ all bias terms $B_i, i = 1, \ldots, \boldsymbol{n}$. The time delays of individual variables are equation specific. As linear equations, each model can be used in any direction, i.e. the output variable can be chosen freely.

4.2 Hybrid LE systems

Statistical analysis is an essential part of the development and tuning of the LE systems: data-driven development of the scaling functions, which is based on advanced generalised norms and moments, is suitable for different statistical distributions. The linguistic equation approach originates from the fuzzy set systems which keeps the connections of the methodologies strong. Compact LE models provide a good basis for multimodel systems, where local LE models are combined with fuzzy logic, to handle transitions between models, some special situations and uncertainty with fuzzy set systems. Fuzzy reasoning is an important part of the LE based fault diagnosis and the decision making in the recursive adaptation.

The LE based development of fuzzy systems on any partition can be done if a sufficient number of variables are known or variated by selecting membership locations. Fuzzy set systems, which represent gradual changes by interpolating with membership functions, can be handled by membership definitions and linguistic equations, i.e. the system does not necessarily need any uncertainty or fuzziness. Fuzzy set systems have been moved to higher levels in applications, when first modelling and control, and later also the detection of the operating conditions was realised with the LE approach.

Fuzzy numbers can be handled in LE models by the *extension principle* [29-31. LE models are extended to fuzzy inputs with this approach if the membership definitions, i.e. functions f_j^- and f_j^+ defined by (5) and the corresponding inverse functions, are replaced by the corresponding extensions of these functions. The square root functions are used in the *linguistification* part. The extension principle is needed for fuzzy inputs. The result of the fuzzy extension is a nonlinear membership function for the output even if the membership function of the input is linear. The number of membership levels should increase with the growing fuzziness of the input.

5 Conclusions

Data mining need to be combined with domain expertise to develop practical systems. The LE approach provides a feasible integration framework for practical intelligent applications. The process insight is maintained since all the modules can be assessed by expert knowledge and the membership definitions relate measurements to appropriate operating areas. The nonlinear scaling methodology based on statistical analysis enhanced with domain expertise is the corner stone of the approach, which represents the variable meanings in a compact way to introduce intelligent indices for control and diagnostics.

Different statistical and intelligent methods are used together with the LE approach. The weighting of submodels also is based on the scaled values and fuzzy logic. The cascade and interactive model structures are used in building more complex large scale applications.

Acknowledgement

This contribution is a post-conference publication from SIMS 2013 Conference (54th SIMS Conference, Bergen University College, Norway, October 16-18, 2013). The contribution is a modified publication from the paper published in the Proceedings of SIMS 2013, to be found at http://www.scansims.org/sims2013/SIMS2013.pdf.

References

[1] Äyrämö S, Kärkkäinen T. (2006). *Introduction to partitioning-based clustering methods with a robust example.* Reports of the Department of Mathematical Information Technology Series C. University of Jyväskylä, Software and Computational Engineering; 2006; No. C. 1/2006.

[2] Babuška R. *Fuzzy Modeling and Identification*. Boston: Kluwer Academic Publisher; 1998.

[3] Babuška R, Setnes M, Kaymak U, Verbruggen HB. (1997). Fuzzy Modelling: a Universal and Transparent Tool. In: Yliniemi L, Juuso E, editors. TOOLMET'97. Proceedings of Tool Environments and Development Methods for Intelligent Systems; 1997 April 17-18, Oulu. P.1-27

[4] Bezdek JC. *Pattern Recognition with Fuzzy Objective Function*. New York: Plenum Press. 1981.

[5] Box GEP, Wilson KB. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society*. 1951; Series B, 13(1): 1–45.

[6] Buckley JJ, Feuring T. (2000). Universal approximators for fuzzy functions. *Fuzzy Sets and Systems*. 2000; 113: 411–415.

[7] Buckley JJ, Hayashi Y. (1999). Can neural nets be universal approximators for fuzzy functions? *Fuzzy Sets and Systems*. 1999; 101: 323–330.

[8] Buckley JJ, Qu Y. On using α-cuts to evaluate fuzzy equations. *Fuzzy Sets and System.* 1990 December; 38(3): 309–312. doi: 10.1016/0165-0114(90)90204-J

[9] Chen S, Cowan C, Grant PM. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks.* 1991; 2(2): 302–309. doi: 10.1109/72.80341

[10] Chiu S. Fuzzy Model Identification Based on Clustering Estimation. *Journal of Intelligent & Fuzzy Systems.* 1994; 2(3): 267–278

[11] De Cock M, Kerre EE. Fuzzy modifiers based on fuzzy relations. *Information Sciences*, 2004; 160(1- 4): 173–199. doi: 10.1016/j.ins.2003.09.002

[12] Driankov D, Hellendoorn H, Reinfrank M. *An Introduction to Fuzzy Control*. Berlin: Springer; 1993.

[13] Dubois D, Prade H, Ughetto L. Fuzzy logic, control engineering and artificial intelligence. In: Verbruggen HB, Zimmermann HJ, Babuska R, editors. Fuzzy Algorithms for Control, International Series in Intelligent Technologies; 1999; Kluwer, Boston. P.17–57

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

[15] Frank PM, Garcia EA, Köppen-Seliger B. Modelling for fault detection and isolation versus modelling for control. Mathematics and Computers in Simulation. 2000; 53: 259–271

[16] Fuller R. Introduction to Neuro-Fuzzy Systems. Advances in Soft Computing. Springer. 2000. 289 pp.

[17] Gerlach RW, Kowalski BR, Wold HOA. Partial least squares modelling with latent variables. Anal. Chim. Acta; 1979; 112(4): 417–421

[18] Gustafson DE, Kessel WC. (1979). Fuzzy clustering with a fuzzy covariance matrix. *In Proceedings of IEEE CDC*, 1979; San Diego. 761–766

[19] Isermann R. Supervision, fault-detection and fault-diagnosis methods - an introduction. Control Engineering Practice. 1997; 5(5): 639–652

[20] Jang JSR. ANFIS: Adaptive-Network-based Fuzzy Inference Systems. *IEEE Transactions on Systems, Man, and Cybernetics.* 1993; 23(3): 665–685

[21] Jolliffe IT. Principal Component Analysis. 2nd ed. New York: Springer. 2002. 487 pp.

[22] Juuso E. A computer analysis of temperature distribution and energy consumption in a submerged arc furnace used in the production of high-carbon ferrochromium. Acta Universitatis Ouluensis, Series C Technica No. 17, Technica Processionum. 1980; No. 3. 42 pp.

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

[24] Juuso E, Lahdelma S. Intelligent trend indices and recursive modelling in prognostics. In: Proceedings The 8th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies; 2011 June 20-22; Cardiff, UK. 1: 440– 450. BINDT

[25] Juuso EK. Multilayer simulation of heat flow in a submerged arc furnace used in the production of ferroalloys. In: Amouroux M, Jai AE, editors. Control of Distributed Parameter Systems 1989. Selected Papers from the 5th IFAC Symposium, Perpignan, France. 1989 June 26-29; IFAC Symposia Series; 1990; Pergamon, Oxford, UK. 3: 459–464

[26] Juuso EK. Computational intelligence in distributed interactive synthetic environments. In: Bruzzone AG, Kerckhoffs EJH, editors. Simulation in Industry. Proceedings of the 8th European Simulation Symposium, Simulation in Industry, ESS'96; 1996 October 2-5; Genoa, Italy; 157–162

[27] Juuso EK. Fuzzy control in process industry: The linguistic equation approach. In: Verbruggen HB, Zimmermann HJ, Babuska R, editors. *Fuzzy Algorithms for Control, International Series in Intelligent Technologies;* 1999; Kluwer, Boston; 243–300. Doi: 10.1007/978-94-011-4405-6 10

[28] Juuso EK. Integration of intelligent systems in development of smart adaptive systems. *International Journal of Approximate Reasoning;* 2004; 35: 307–337. doi: 10.1016/j.ijar.2003.08.008

[29] Juuso EK. Forecasting batch cooking results with intelligent dynamic simulation. In: Zupančič B, Karba R, Blažič S, editors. Proceedings of the 6th EUROSIM Congress on Modelling and Simulation; 2007 September 9-13; Ljubljana, Slovenia. 2: 8 pp.

[30] Juuso EK. Intelligent modelling of a fluidised bed granulator used in production of pharmaceuticals. In: Bunus P, Fritzson D, Führer C, editors. Conference Proceedings of SIMS 2007 - The 48th Scandinavian Conference on Simulation and Modeling; 2007 October 30-31; Göteborg (Särö); 101–108. Linköping University Electronic Press, Linköping, Sweden.

[31] Juuso EK. Intelligent dynamic simulation of a fed-batch enzyme fermentation process. In: 10th International Conference on Computer Modelling and Simulation, EUROSIM/UKSi; 2008 April 13; Cambridge, UK. 301–306. The Institute of Electrical and Electronics Engineers IEEE. Doi: 10.1109/UKSIM.2008.133

[32] Juuso EK. Tuning of large-scale linguistic equation (LE) models with genetic algorithms. In: Kolehmainen M, 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: 161–170. Springer-Verlag, Heidelberg. Doi: 10.1007/978- 3-642-04921-7_17

[33] Juuso EK. Recursive tuning of intelligent controllers of solar collector fields in changing operating conditions. In: Bittani S, Cenedese A, Zampieri S, editors. Proceedings of the 18th World Congress The International Federation of Automatic Control; 2011 August 28 - September 2; Milano; P.12282–12288.

[34] Juuso EK, Bennavail J, Singh M. Hybrid knowledge-based system for managerial decision making in uncertainty environment. In: Carrete NP, Singh MG, editors. Qualitative Reasoning and Decision Technologies. QUARDET'93. Proceedings of the IMACS International Workshop on Qualitative Reasoning and Decision Technologies. 1993 June 16 – 18; Barcelnoa; P. 234–243

[35] Juuso EK, Leiviskä K. Adaptive expert systems for metallurgical processes. In: Jämsä-Jounela SL, Niemi AJ, editors. Expert Systems in Mineral and Metal Processing. Proceedings of the IFAC Workshop; 1991 August 26-28; Espoo, Finland; IFAC Workshop Series; 1992; Pergamon, Oxford, UK; Number 2: 119–124

[36] Kohonen T. *Self-Organizing Maps*. Berlin: Springer. 1995.

[37] Lahdelma S, Juuso E. Signal processing in vibration analysis. In: Proceedings of the Fifth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies; 2008 July 15-18; Edinburgh, UK; 867–878. BINDT.

[38] Lahdelma S, Juuso E. Signal processing and feature extraction by using real order derivatives and generalised norms. Part 1: Methodology. The International Journal of Condition Monitoring; 2011; 1(2), 46–53. doi: 10.1784/204764211798303805.

[39] Ljung L. System Identification - Theory for the User. 2nd edition. Upper Saddle River, N.J.: Prentice Hall. 1999.

[40] Ljung L. Perspectives on system identification. In: Chung MJ, Misra P, editors. Plenary papers, milestone reports & selected survey papers, 17th IFAC World Congress; 2008 July 6-11; Seoul, Korea; P.47–59

[41] Macias-Hernandez J. Opportunities of smart adaptive systems in the oil refining industry. In: Proceedings of Eunite 2004 - European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems; 2004 June 10-12; Aachen, Germany. Wissenschaftsverlag Mainz, Aachen; P.1–12.

[42] Mendel JM. Advances in type-2 fuzzy sets and systems. *Information Sciences*. 2007; 177(1): 84–110

[43] Moore RE. Interval Analysis. Englewood Cliffs, NJ: Prentice Hall. 1966.

[44] Pedrycz W. An identification algorithm in fuzzy relational systems. *Fuzzy Sets and Systems*. 1984; 13(2): 153– 167.

[45] Rummelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rummelhart DE, McClelland J, editors. Parallel Data Processing; 1986; M.I.T. Press, Cambridge, MA.; P.318–362

[46] Shields MW, Casey MC. A theoretical framework for multiple neural network systems. *Neurocomputing*, 2008; 71: 1462–1476

[47] Sontag E. Nonlinear regulation: The piecewise linear approach. *IEEE Transactions Automatic Control*. 1981; 26(2): 346–358

[48] Takagi T, Sugeno M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*. 1985; 15(1): 116–132

[49] Windham MP. Cluster validity for fuzzy clustering algorithms. *Fuzzy Sets and Systems*. 1981; 5(2): 177–185

[50] Zadeh LA. Fuzzy sets. ***Information and Control.*** 1965; 8(June): 338–353

[51] Zimmermann HJ. *Fuzzy set theory and its applications*. Kluwer Academic Publishers. 1992.