Esko Juuso

INTEGRATION OF INTELLIGENT SYSTEMS IN DEVELOPMENT OF SMART ADAPTIVE SYSTEMS

LINGUISTIC EQUATION APPROACH
ESKO JUUSO

INTEGRATION OF INTELLIGENT SYSTEMS IN DEVELOPMENT OF SMART ADAPTIVE SYSTEMS
Linguistic equation approach

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in Arina-sali (Auditorium TA105), Linnanmaa, on 29 November 2013, at 12 noon

UNIVERSITY OF OULU, OULU 2013
Abstract

Smart adaptive systems provide advanced tools for monitoring, control, diagnostics and management of nonlinear multivariate processes. 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 highly complex systems require domain expertise and more compact approaches at the basic level. Linguistic equation (LE) approach originating from fuzzy logic is an efficient technique for these problems. This research is focused on the smart adaptive applications, where different intelligent modules are used in a smart way.

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. The LE based intelligent analysers are useful in the multilevel LE control and diagnostics: the LE control is enhanced with the intelligent analysers, adaptive and model-based modules and high level control. The operating area is extended with the predefined adaptation and specific events activate appropriate control actions. The condition, stress and trend indices are used for the detection of operating conditions. The same overall structure is extended to the scheduling and managerial decision support. The linguistic representation becomes increasingly important when the human interaction is essential.

The new scaling approach is used in control and diagnostic applications and discussed in connection with previous multivariate modelling cases. The LE based intelligent analysers are the key modules of the system integration, which produces hybrid systems: fuzzy systems move gradually to higher levels, neural networks and evolutionary computing are used for tuning. The overall system is reinforced with advanced statistical analysis, signal processing, feature extraction, classification and mechanistic modelling.

Keywords: control, decision support, diagnostics, intelligent methods, linguistic equations, modelling, smart adaptive systems, statistical analysis
Juuso, Esko, Älykkäiden menetelmien yhdistäminen viisaiden mukautuvien järjestelmien kehittämisessä. Lingvististen yhtälöiden menetelmä
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Teknillinen tiedekunta, Prosessi- ja ympäristötekniikan osasto
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä
Viisaat mukautuvat järjestelmät sisältävät kehittyneitä työkaluja epälineaaristen monimuuttujaisen prosessien valvontaan, säätöön, diag nostistiikkaan ja johtamiseen. Laajaan menetelmäpohjaan perustuva tiedonrik astus on pohjana älykkäiden järjestelmien yhdistämisele. Pienille erikoistuneille järjestelmille on monia toteutettavissa olevia ratkaisuja, mutta erittäin monimutkaiset järjestelmät vaativat alan asiantuntemusta ja kompak teja lähestymistapoja perustasolla. Sumean logiikkaan pohjautuva lingvististen yhtälöiden (linguistic equation, LE) menetelmä on tehokas ratkaisu näissä ongelmissa. Tämä tutkimus kohdistuu viisaihin mukautuviin sovelluksiin, jossa useita älykkäitä moduleuja käytetään yhdessä viisaalla tavalla.


Asiasanat: diagnostiikka, lingvistiset yhtälöt, päätöksenteontuki, säätö, tilastollinen analyysi, viisaat mukautuvat järjestelmät, älykkäät menetelmät
To my family
Acknowledgements

My deepest gratitude goes to Professor Kauko Leiviskä for his support and encouragement as my supervisor and the reviewers, Professor Sirkka-Liisa Jäämsä-Jounela and Doctor Olli Ventä, for their valuable comments and suggestions.

I thank the people at the Control Engineering Laboratory for the friendly and encouraging working atmosphere there while I was doing research and finishing the manuscript. My special thanks go to the researchers with whom I have worked in numerous research projects. I also offer my sincere thanks to Professor Sulo Lahdelma and his research group at the Mechatronics and Machine Diagnostics Laboratory. This cooperation has introduced valuable new topics to my research.

The control applications at Plataforma Solar de Almería in Spain have had an essential role in this research work. Especially, I’d like to thank Doctor Luis Yebra and his research group for facilitating the testing of recent developments in intelligent analysers and model-based control. The early control applications in the pulp and paper industry were important in forming the overall framework. I express a special thank you to Doctor Mika Järvensivu in the lime kiln control and Marjatta Piironen ja Iiris Joensuu in water applications for their cooperation and the chance to develop effective realizations in industry. The long-term research of industrial wastewater treatment with Ilkka Laakso and Jani Tomperi was instrumental to developing the trend analysis. The fault diagnosis research with Doctor Timo Ahola and Doctor Sébastien Gebus provided a basis for the higher levels of the overall framework.

The Measurement, Monitoring and Environmental Assessment (MMEA) research programme has provided an innovative integration environment for completing this thesis. I also thank my colleagues in European networks, ERUDIT, EUNITE and NiSIS, and international scientific societies, Eurosim, IFAC, ISCM and SIMS, for interesting discussions of the research field.

Many parts of this research were partly funded by the Finnish Funding Agency for Technology and Innovation and by companies involved in the projects. European research programmes have provided funding during the research. The latest control tests were supported by the "Solar Facilities for the European Research Area (SFERA)" as a part of the 7th Framework Programme of the EU (SFERA Grant Agreement 228296).
Automation Foundation has helped me to keep the international contacts active. All sponsoring, financial support, and project partners are gratefully acknowledged.

My warmest appreciation goes to Seija for her motivation and trust, to Ilkka and Anna-Maija and their families for enriching everyday life. Ilkka and Anna-Maija have also been important in helping with the practicalities. Special thanks go to Anna-Maija for helping me to improve the language of the thesis. I am grateful to my grandson Aleksi for introducing efficient breaks to research work.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial intelligence</td>
</tr>
<tr>
<td>ANFIS</td>
<td>Adaptive Network based Fuzzy Inference System</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>ARMAX</td>
<td>AutoRegressive Moving Average with eXogeneous inputs</td>
</tr>
<tr>
<td>ARX</td>
<td>AutoRegressive with eXogeneous inputs</td>
</tr>
<tr>
<td>ASA</td>
<td>ASymmetrical Action</td>
</tr>
<tr>
<td>BSS</td>
<td>Blind Source Separation</td>
</tr>
<tr>
<td>CBM</td>
<td>Condition-Based Maintenance</td>
</tr>
<tr>
<td>CBR</td>
<td>Case-Based Reasoning</td>
</tr>
<tr>
<td>CDG</td>
<td>Causal Directed Graphs</td>
</tr>
<tr>
<td>CLA</td>
<td>Cooking Liquor Analyser</td>
</tr>
<tr>
<td>CH</td>
<td>Constraints Handling</td>
</tr>
<tr>
<td>CM</td>
<td>Condition Monitoring</td>
</tr>
<tr>
<td>COD</td>
<td>Chemical Oxygen Demand</td>
</tr>
<tr>
<td>DSS</td>
<td>Decision Support System</td>
</tr>
<tr>
<td>DSVI</td>
<td>Deluted Sludge Volume Index</td>
</tr>
<tr>
<td>ETA</td>
<td>Event Tree Analysis</td>
</tr>
<tr>
<td>FB</td>
<td>FeedBack</td>
</tr>
<tr>
<td>FCM</td>
<td>Fuzzy C-Means</td>
</tr>
<tr>
<td>FF</td>
<td>FeedForward</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>FFT(^{-1})</td>
<td>Inverse Fourier Transform</td>
</tr>
<tr>
<td>FGS</td>
<td>Fuzzy Gain Scheduling</td>
</tr>
<tr>
<td>FLC</td>
<td>Fuzzy Logic Controller</td>
</tr>
<tr>
<td>FLS</td>
<td>Fuzzy Least Squares</td>
</tr>
<tr>
<td>FMEA</td>
<td>Failure Mode and Effect Analysis</td>
</tr>
<tr>
<td>FRM</td>
<td>Fuzzy-ROSA (Rule Orientated Statistical Analysis)</td>
</tr>
<tr>
<td>FTA</td>
<td>Fault Tree Analysis</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic algorithm</td>
</tr>
<tr>
<td>G2</td>
<td>G2 software</td>
</tr>
<tr>
<td>GPC</td>
<td>Generalized Predictive Control</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>ICA</td>
<td>Independent Component Analysis</td>
</tr>
<tr>
<td>IMC</td>
<td>Internal Model Control</td>
</tr>
<tr>
<td>JOT</td>
<td>Just On Time</td>
</tr>
<tr>
<td>K-means</td>
<td>Clustering methodology</td>
</tr>
<tr>
<td>KPI</td>
<td>Key Performance Indicator</td>
</tr>
<tr>
<td>LARS</td>
<td>Least Angle Regression</td>
</tr>
<tr>
<td>LE</td>
<td>Linguistic Equation</td>
</tr>
<tr>
<td>LECont</td>
<td>Linguistic Equation Controller</td>
</tr>
<tr>
<td>LEGS</td>
<td>Linguistic Equation Gain Scheduling</td>
</tr>
<tr>
<td>LNFS</td>
<td>Linguistic NeuroFuzzy System</td>
</tr>
<tr>
<td>LCC</td>
<td>Life Cycle Cost</td>
</tr>
<tr>
<td>LPC</td>
<td>Linguistic Principal Component</td>
</tr>
<tr>
<td>LPCA</td>
<td>Linguistic Principal Component Analysis</td>
</tr>
<tr>
<td>LPV</td>
<td>Linear parameter varying</td>
</tr>
<tr>
<td>LRBF</td>
<td>Linguistic Radial Basis Function</td>
</tr>
<tr>
<td>LSOM</td>
<td>Linguistic Self-Organising Map</td>
</tr>
<tr>
<td>LTS</td>
<td>Linguistic Takagi-Sugeno model</td>
</tr>
<tr>
<td>LVQ</td>
<td>Learning Vector Quantisation</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple Input, Multiple Output</td>
</tr>
<tr>
<td>MISO</td>
<td>Multiple Input, Single Output</td>
</tr>
<tr>
<td>MIT</td>
<td>Measurement Index</td>
</tr>
<tr>
<td>MLP</td>
<td>MultiLayer Perceptron</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean Absolute Error</td>
</tr>
<tr>
<td>MDPA</td>
<td>Model-based Diagnostical Process Analysis</td>
</tr>
<tr>
<td>MPC</td>
<td>Model Predictive Control</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>NARX</td>
<td>Nonlinear AutoRegressive with eXogeneous inputs</td>
</tr>
<tr>
<td>ODE</td>
<td>Ordinary Differential Equation</td>
</tr>
<tr>
<td>OE</td>
<td>Output-Error</td>
</tr>
<tr>
<td>OEE</td>
<td>Overall Equipment Effectiveness</td>
</tr>
<tr>
<td>PBA</td>
<td>Predictive Braking Action</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PCB</td>
<td>Printed Circuit Board</td>
</tr>
<tr>
<td>PCI</td>
<td>Process Capability Index</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>PDE</strong></td>
<td>Partial Differential Equation</td>
</tr>
<tr>
<td><strong>PFT</strong></td>
<td>Possibilistic Fault Tree</td>
</tr>
<tr>
<td><strong>PHM</strong></td>
<td>Proportional Hazard Model</td>
</tr>
<tr>
<td><strong>PID, PI, PD</strong></td>
<td>Controller type: P=Proportional, I=Integral and D=Derivative</td>
</tr>
<tr>
<td><strong>PLC</strong></td>
<td>Programmable Logic Controller</td>
</tr>
<tr>
<td><strong>PLS</strong></td>
<td>Partial Least Squares regression</td>
</tr>
<tr>
<td><strong>PSA</strong></td>
<td>Plataforma Solar de Almeria</td>
</tr>
<tr>
<td><strong>PWA</strong></td>
<td>PieceWise Affine</td>
</tr>
<tr>
<td><strong>RAMS</strong></td>
<td>Reliability, Availability, Maintainability and Safety</td>
</tr>
<tr>
<td><strong>RBF</strong></td>
<td>Radial Basis Function</td>
</tr>
<tr>
<td><strong>RBM</strong></td>
<td>Risk-Based Maintenance</td>
</tr>
<tr>
<td><strong>RCM</strong></td>
<td>Reliability-Centered Maintenance</td>
</tr>
<tr>
<td><strong>RSM</strong></td>
<td>Response Surface Methodology</td>
</tr>
<tr>
<td><strong>RUL</strong></td>
<td>Remaining Useful Life</td>
</tr>
<tr>
<td><strong>SAS</strong></td>
<td>Smart Adaptive Systems</td>
</tr>
<tr>
<td><strong>SC</strong></td>
<td>Stabilising Controller</td>
</tr>
<tr>
<td><strong>SCEMM</strong></td>
<td>Scandinavian Center for Maintenance Management</td>
</tr>
<tr>
<td><strong>SISO</strong></td>
<td>Single Input, Single Output</td>
</tr>
<tr>
<td><strong>SMART</strong></td>
<td>Specific, Measurable, Attainable, Realistic and Timely</td>
</tr>
<tr>
<td><strong>SMC</strong></td>
<td>Sliding Mode Control</td>
</tr>
<tr>
<td><strong>S-N curve</strong></td>
<td>Stress-Cycle curve</td>
</tr>
<tr>
<td><strong>SOC</strong></td>
<td>Self-Organizing Controller</td>
</tr>
<tr>
<td><strong>SOM</strong></td>
<td>Self-Organising Map</td>
</tr>
<tr>
<td><strong>SPC</strong></td>
<td>Statistical Process Control</td>
</tr>
<tr>
<td><strong>SSE</strong></td>
<td>Sum Squared Error</td>
</tr>
<tr>
<td><strong>SVM</strong></td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td><strong>TS fuzzy</strong></td>
<td>Takagi-Sugeno fuzzy models</td>
</tr>
<tr>
<td><strong>TPM</strong></td>
<td>Total Productive Maintenance</td>
</tr>
<tr>
<td><strong>TQM</strong></td>
<td>Total Quality Management</td>
</tr>
<tr>
<td><strong>VDI</strong></td>
<td>Verein Deutscher Ingenieure</td>
</tr>
</tbody>
</table>

**A** Interaction matrix defined by coefficients $A_{ij}$ of variables $j, j = 1, \ldots, m$ in equations $i, j = 1, \ldots, n; j = \text{out}$ used for the output variable

$A_S, B_S, C_S, D_S$ Coefficient matrices of a state-space model
Model coefficients in ARX models
\[ a_j, b_j, c_j \]
Coefficients of the polynomial \( f_j = a_j X_j^2 + b_j X_j + c_j, \ j = 1, \ldots, m \) or \( j = \text{out} \)
\[ a^+_j, b^+_j, c^+_j \]
Coefficients of the polynomial \( f^+_j = a^+_j X_j^2 + b^+_j X_j + c^+_j, \ j = 1, \ldots, m \) or \( j = \text{out} \)
\[ a_{rms} = ||\mathbf{M}_2^2||_2 \]
rms acceleration
\[ \mathbf{\bar{B}} \]
Bias vector defined by the bias terms \( B_i \) of the equations \( i, i = 1, \ldots, n \)
\[ b_{type} \]
Bias terms, types: feedforward (FF), neural network (NN), working point (wp), control power (cp)
\[ bc_i(k) \]
Predictive braking coefficient
\[ C(k), C(k-1) \]
Fatigue contributions of the stress
\[ c^f_i(k) \]
Weighted sum of the scaling actions
\[ c_j = \frac{1}{2}(c^-_j + c^+_j) \]
Operating point
\[ (c_i)_j, (c_k)_j \]
Limits of the core area
\[ c^-_j, c^+_j \]
Centre value from left and right, respectively
\[ (c_A)_j, (c_B)_j \]
Asymmetry and braking constants, \( (c_A)_j \in [0, 1], (c_B)_j \in [0, 1] \)
\[ c_{in}, c_{out}, c_{diff} \]
Coefficients of the additional change of control
\[ cp_i(k) \]
Control power of the manipulating variable \( i \)
\[ cr_i(k) \]
Cumulative rate of control actions of the manipulating variable \( i \)
\[ D^+_j, D^-_j \]
Derivatives of the scaling functions
\[ e(i) \]
Noise term
\[ e_j(k), e_j(k-1) \]
Error of the controlled variables \( j \)
\[ \hat{e}_j(k), \hat{e}_j(k) \]
Scaled error of the controlled variables \( j \)
\[ e'_j(k), e'_j(k-1) \]
New and previous initial error of the controlled variable \( j \)
\[ \hat{e}'_j(k), \hat{e}'_j(k-1) \]
New and previous scaled initial error of the controlled variable \( j \)
\[ e^-_j, e^+_j \]
Deviation thresholds of the initial error of the controlled variable \( j \)
\[ F \]
Oil flow rate \( (ls^{-1}) \)
Feasible ranges in recursive adaptation

\( F(\vec{x}), F(t, \vec{x}), F(x_j) \)

Functions

\((F_D, \mu_D)\) Discrete membership functions for \( x_j \) and \( X_j \)

\( F_i() \) Activation function of the neuron \( i \)

\( F_j, \bar{F}_j \) Fuzzy set and its complement in \( U_j \)

\( F_k^{(a)} \) Features obtained from a real order derivative

\( f_l(), f() \) Basis functions, \( l = 1, \ldots, m_f \)

\( f_j^{-1}, f_j, f_{out} \) Scaling functions of the variable \( j \): \( x_j \rightarrow [-2,2] \rightarrow \mathbb{R} \)

\((f_{type})^{-1}_j\) Scaling functions of the variable \( j \), type: \( e, \Delta e, \delta e \) and \( \Delta u \)

\( f_j^{-1}, f_j \) Scaling functions of the variable \( x_j \) from \([-2,0)\) and \([0,2]\) to \( \mathbb{R} \)

\((f_a)^{-1}\) Scaling functions of the relative norm \(|\tau M_a^0|\)

\( \vec{I}_{C}(k), \bar{I}_C(k), (I_C(k))_L \)

Condition indices obtained from the case models, \( L = 0, 1, \ldots, n_{cat} \)

\( \vec{I}_C^{(a)}(k) \) Condition index obtained from a real order derivative

\( I_j^{D}(k) \) Deviation index of variable \( j \)

\( I_{eff} \) Effective irradiation (Wm\(^{-2}\))

\( \vec{I}_H(k), \bar{I}_H(k) \) Health indices

\( \vec{I}_M(k), \bar{I}_M(k) \) Measurement indices

\( \vec{I}_Q(k), \bar{I}_Q(k) \) Quality indices obtained from the case models

\( \vec{I}_S(k), \bar{I}_S(k) \) Stress indices obtained from the case models

\( \vec{I}_S^{(a)}(k) \) Stress index obtained from a real order derivative

\( \vec{I}_S(k), \bar{I}_S(k), (I_S(k))_L \)

Stress indices obtained from the case models, \( L = 0, 1, \ldots, n_{cat} \)

\( \vec{I}_j^T(k) \) Trend index of variable \( j \)

\( I_X, I_Y \) Identity matrices

\( K_P, K_D, K_I \) Controller coefficient matrices defined by coefficients \( K_P(i, j), K_D(i, j) \) and \( K_I(i, j) \) (Proportional, Derivative, Integral), rows \( i, j = 1, \ldots, n \) and columns \( j, j = 1, \ldots, m \)

\( K_S \) Number of samples

\( k \) Time step, integer

\( L_e, L_{\Delta e}, L_{\Delta e} \) Linguistic labels (error, derivative of error, sum of errors)

\( L_{C}, L_{\Delta C} \) Linguistic labels (control, change of control)
\( L_{x_i} \) Linguistic label of a variable \( x_j \)
\( L_m \subset \mathbb{R}^m, L_n \subset \mathbb{R}^n \)
- Sets limited to \([-2, 2]\)
\( m \) Number of variables
\( \tau M_p^\alpha \) Generalised central absolute moment about \( \hat{x}^{(\alpha)} \), normalised by standard deviation \( \sigma \)
\[ ||\tau M_p^\alpha||_p = (\tau M_p^\alpha)^{1/p} \]
- Generalised norm of \( x_j, x_j > 0 \) when \( p < 0 \), and \( x_j \geq 0 \) when \( p > 0 \)
\( \tau M_p^\alpha = \tau M_p^\alpha \) Generalised moment of \( x^{(\alpha)} \)
\[ ||\tau M_p^\alpha||_p = (\tau M_p^\alpha)^{1/p}, ||x_i^p||_{p_i} \]
- Generalised norm of \( x^{(\alpha)} \)
\( m_{type} \) Number of variables, \( type \): controlled variables (C), feedforward (FF), working point variables (W), variables in asymmetry action (A), variables effecting to the control power (P), variables with a trend index (T), variables with a fluctuation index (F), variables in condition indices (CM), variables in stress indices (S).
\( N \) Number of values in the sample
\( N_C(k) \) Number of cycles in fatigue analysis
\( N_{predicted} \) Predicted category
\( N_s \) Number of signal values in a second
\( n \) Number of outputs
\( n_{type} \) Number of variables, \( type \): inputs (i), manipulating variables (M), feedback (FB), manipulating variables used in feedforward control (FF), categories (cat), basis functions (f)
\( n_{window} \) Length of window: cumulative sum (R), long window (L), short window (S)
\( n_c(L) \) Number of cases in category \( L, L = 1, \ldots, n_{cat} \)
\( n_D = n_F + n_C \)
- Number of discrete points: \( n_F \) from the calculation level and \( n_C \) from the most possible area
\( n_e(l) \) Number of equations in case \( l, l = 1, \ldots, n_c(L) \)
\( n_k, n_j \) Number of delay from input to output, variable \( x_j \)
\( p \in \mathbb{R}, q \in \mathbb{R} \) Order of norm or moment
\( \vec{p} = (p_1, p_2, \ldots, p_m) \)
- Normalised values of \( x_j \) (min-max or z-score)
Orders of the norm corresponding to \((c_l)_j\), \(c_j\) and \((c_h)_j\), respectively

\(p_l, p_h\) Lower and higher orders of the norms used in the fluctuation analysis

\(R^2\) Coefficient of determination

\(s_i\) Constant for manual adjustment

\(sc_j(k)\) Scaling coefficient

\(T_{amb}\) Ambient temperature (°C)

\(T_{diff} = T_{out} - T_{in}\) Temperature difference (°C)

\(T_{in}\) Inlet temperature (°C)

\(T_{out}\) Outlet temperature (°C)

\(T_{ref}\) Setpoint of the outlet temperature (°C)

\(T_{ref}^{wp}\) Setpoint of the outlet temperature (°C) calculated from the working point \(wp_{min}\)

\(t\) Time

\(U_j\) Universal set

\(U_i(k)\) Model input in the linguistic range

\(u_i(k), u_i(k - 1)\) Manipulating variable

\(u_i^{type}(k)\) Model input and control outputs, see bias correction (BC), feedback (FB), feedforward (FF)

\(u_i^{ll}, u_i^{hi}\) Low and high limits of the control output

\(u_i^{lo}, u_i^{ho}\) Dynamic low and high limits of the control output

\(u_{ij}(k)\) Change of control \(i\) based on variable \(j\)

\(\tilde{u}_{ij}(k)\) Scaled change of control \(i\) based on variable \(j\)

\(v_{rms} = \sqrt{x_{rms}} = ||\mathbf{M}^2_{x}||_2\) root-mean-square (rms) velocity

\(w_{ij}\) Weight factors, types: control (C), neural network (NN), working point (wp), control power (cp)

\(w_{il}\) Weight factor of the function \(F_i(\vec{x}), l = 1, \ldots, m_f\)

\(w^{wp}_i, w^{cp}_i, w^{cr}_i\) Weight factors of the scaling actions: working point (wp), control power (cp), cumulative rate (cr)

\(w^{T1}_j, w^{T2}_j\) Weight factors of the trend index
\(w_{\text{cat}}(L)\) Weight factor of the category \(L, l = 1, \ldots, n_{\text{cat}}\)
\(w_p(i, l)\) Weight factor of equation \(i, i = 1, \ldots, n_r(l)\), in case \(l, l = 1, \ldots, n_r(L)\)
\(w_{p\text{min}}(k)\) Working point of the manipulating variable \(i\)
\(X_j^+, X_j^-\) Negative and positive scaled values of \(x_j\)
\((X_j)_{\text{min}}, (X_j)_{\text{max}}\) Scaled outlier limits of \(x_j\)

\((X_j)_{2L-1}, (X_j)_{2L}\) Limits of \(\alpha - \text{cuts}\)
\((x_j)^{\text{hl}}, (x_j)^{\text{hl}}\) Limits of the support area of \(x_j\)
\((x_j)_{\text{min}}, (x_j)_{\text{max}}\) Outlier limits of \(x_j\)
\(\vec{X}, X_j = \tilde{x}_j\) Scaled values of \(x\)

\(\vec{y} = (y_1, y_2, \ldots, y_n), y_i\) Model output
\(y_{j}(k), y_{j}(k-1)\) The most recent moving average values of the measurement
\(y_p^p(k), y_p^p(k-1)\) Set point, target, reference value

\(\alpha \in \mathbb{R}\) Order of derivation
\(\alpha^{-}, \alpha^{+}\) Shape parameters of the scaling functions, \(\alpha^{-} \in [\frac{1}{4}, 3], \alpha^{+} \in [\frac{1}{4}, 3]\)
\(\tilde{\beta}_i\) Dilation parameters of the basis function \(f\)
\(\gamma_k = \frac{\gamma_k^m}{\sigma_j}\) Normalised moment \((k \text{ order of moment, } k = 3 \text{ skewness, } k = 4 \text{ kurtosis})\) about \(c_j\)
\( \gamma_p^k \) Generalised moment \((k\ \text{order of moment, } p\ \text{order of norm, } k = 3)\) skewness, \(k = 4\) kurtosis) about \(|\| M_p^j ||_p\)

\( \bar{h} \) Location parameters of the basis function \( f \)

\( \Delta c_j, \Delta c_j^-, \Delta c_j^+ \) Deviations (symmetric, negative, positive)

\( \Delta e_j(k) \) Derivative of the error of the controlled variable \( j \)

\( \Delta e_j(k) \) Scaled derivative of error of the controlled variable \( j \)

\( \Delta e_j(k) \) Moving average of the derivative of the error

\( \Delta I_{eff} \) Derivative of effective irradiation

\( \Delta F^F \) Fluctuation index of effective irradiation

\( \Delta F^F \) Fluctuation index of oil flow

\( \Delta I^T(k) \) Derivative of the trend index of variable \( j \)

\( \Delta T^F_{diff} \) Fluctuation index of temperature difference

\( \Delta T^H_{in}(k) \) Fast change of inlet temperature (\( ^\circ\)C)

\( \Delta T^H_{in}(k) \) Scaled \( \Delta T^H_{in}(k) \)

\( \Delta T^H_{out} \) Change of outlet temperature (\( ^\circ\)C), type: range \( R \), overshoot \( H \)

\( \Delta T^H_{out} \) Scaled \( \Delta T^H_{out} \)

\( \Delta u^T_{i}, \Delta u^{hl}_l \) Low and high limits of the change of control output

\( \Delta u^{type}_{i}(k) \) Change of control, \( type: \) constraint handling \( (CH) \), stabilising control \( (SC) \)

\( \Delta u_{ij}(k), \Delta u_{i}(k) \) Change of control \( i \) based on variable \( j \)

\( \Delta u_{ij}(k) \) Scaled change of control \( i \) based on variable \( j \)

\( \Delta w_{p_i}(k) \) Change of working point of the manipulating variable \( i \)

\( \Delta x_j^F(k) \) Fluctuation indicator

\( \delta e_j(k) \) Sum of the errors of the controlled variable \( j \) in a window of \( n_R \) sample times

\( \delta u_i(k) \) Cumulative sum of control actions \( i \)

\( \delta H \) Health index corresponding to \( I_c(k) = -2 \)

\( \delta u_i(k) \) Scaled cumulative sum of control actions \( i \)

\( \varepsilon_i \) Residual of the equation \( i \), denoted as fuzziness

\( \varepsilon^-_1, \varepsilon^+_1 \) Lower and upper thresholds of the trend index of variable \( j \)

\( \varepsilon^-_2, \varepsilon^+_2 \) Lower and upper thresholds of the derivative of the trend index of variable \( j \)
κ  Shape parameter of a Weibull distribution

\( \lambda_i^-, \lambda_i^+ \)  Ratios for the dynamic low and high limits of the control output

\( \lambda_p \)  Expectation value of a Poisson distribution

\( \lambda_W \)  Scale parameter of a Weibull distribution

\( \mu_c(l), \mu_c(i, L) \)  Degree of membership of case \( l \) and \( i, i = 1, \ldots, n_c(L) \) in category \( L \), \( L = 1, \ldots, n_{cat} \)

\( \mu_{cat}(L) \)  Degree of membership of category \( L, L = 1, \ldots, n_{cat} \)

\( \mu_e(i, l) \)  Degree of membership of equation \( i, i = 1, \ldots, n_e(l) \), in case \( l, l = 1, \ldots, n_e(L) \)

\( \mu_F(), \mu_{\overline{F}}() \)  Membership function of fuzzy set \( F \) and its compliment \( \overline{F} \)

\( \mu_{\bar{F}}() \)  Membership functions of the complement sets \( \bar{F}_j \)

\( \mu_{F^i}() \)  Membership functions of fuzzy sets \( F^i_j \)

\( \mu_{L_l} \)  Degree of membership of the point \( \bar{x}_l \) in the cluster \( L \)

\( \sigma_j = \sqrt{\frac{M^2_j}{\tau}} \)  Standard deviation of \( x_j = 2\text{nd absolute moment about } \overline{x}_j \)

\( \sigma_{\alpha} = \sqrt{\frac{M^2_{\alpha}}{\tau}} \)  Standard deviation of \( x^{(\alpha)} = 2\text{nd absolute moment about } \overline{x}^{(\alpha)} \)

\( \tau \)  Sample time
Contents

Abstract
Tiivistelmä
Acknowledgements 9
Abbreviations 11
Contents 21
1 Introduction 25
  1.1 Background ................................................. 25
  1.2 Research problem and asserted hypothesis .................. 29
  1.3 Contribution .................................................. 32
  1.4 Main results ................................................... 34
  1.5 Contents ...................................................... 35
2 Adaptive systems 37
  2.1 Nonlinear multivariate methodologies ....................... 38
    2.1.1 Steady-state systems .................................. 38
    2.1.2 Dynamic systems ....................................... 40
    2.1.3 Decomposition methodologies ........................ 41
    2.1.4 System development and tuning ........................ 43
  2.2 Soft sensors .................................................. 45
    2.2.1 Data collection and pre-processing .................... 45
    2.2.2 Signal processing and feature extraction ............. 47
    2.2.3 Sensor fusion ............................................ 50
  2.3 Nonlinear process control ................................ 50
    2.3.1 Feedback control ....................................... 51
    2.3.2 Adaptive control ....................................... 53
    2.3.3 Model-based control .................................... 54
    2.3.4 High-level control ...................................... 55
  2.4 Detection of operating conditions .......................... 56
    2.4.1 Symptom generation ..................................... 57
    2.4.2 Classification and reasoning ........................... 58
  2.5 Discussion ................................................... 61
    2.5.1 Nonlinear multivariate modelling ....................... 61
2.5.2 Nonlinear multivariate control and diagnostics ........................................ 67
2.5.3 Smart adaptive systems ................................................................. 71

3 Adaptive nonlinear scaling ................................................................. 75
  3.1 Nonlinear scaling ............................................................................... 75
    3.1.1 Working point and feasible ranges .............................................. 75
    3.1.2 Membership definitions and functions ......................................... 79
    3.1.3 Statistical distributions ............................................................. 86
  3.2 Adaptation of nonlinear scaling ........................................................ 88
    3.2.1 Data selection ............................................................................... 89
    3.2.2 Recursive adaptation .................................................................. 91
    3.2.3 Decomposition ............................................................................ 92
    3.2.4 Statistical process control ............................................................. 92

4 Linguistic equation modelling ............................................................... 95
  4.1 LE models ....................................................................................... 95
    4.1.1 Interactions .................................................................................. 95
    4.1.2 Steady-state LE models ............................................................... 96
    4.1.3 Dynamic LE models ................................................................. 97
    4.1.4 Case-based LE models ................................................................ 101
    4.1.5 Large-scale LE models .............................................................. 102
  4.2 Data-based modelling ....................................................................... 103
  4.3 Adaptation of LE models .................................................................. 106

5 Linguistic equation control ................................................................. 111
  5.1 Feedback LE controllers ............................................................... 111
  5.2 Adaptation ...................................................................................... 114
    5.2.1 Adaptive scaling ...................................................................... 114
    5.2.2 Predictive braking action ......................................................... 116
    5.2.3 Asymmetrical action ................................................................. 119
    5.2.4 Constraint handling ................................................................. 119
  5.3 Model-based LE control .................................................................. 120
  5.4 High-level LE control ...................................................................... 122
  5.5 Intelligent indices ............................................................................ 122
  5.6 Adaptation of LE control systems .................................................. 127

6 Linguistic equations in hybrid systems .............................................. 133
  6.1 Linguistic equations and statistical analysis .................................... 133
1 Introduction

Process automation is essential for industrial processes, where systems developed for different subprocesses need to be efficiently integrated. Sustainability requirements have introduced new applications, where intelligent methodologies are increasingly important. Similar solutions are also useful in production planning and scheduling. Managerial decision making based partly on uncertain information completes the overall environment for the development of smart adaptive systems. Linguistic equations, which provide feasible solutions for many applications in this framework, form the methodological platform of this thesis.

1.1 Background

Process automation has a long history in pulp and paper industry: centralised control rooms and standardised signal systems have been used since the 1950s, and in the 1970s process computers were introduced to control different processes. Digital automation systems provided tools for combining distributed control functions with centralised process monitoring. System integration on the mill and corporate level together with the integration with office automation, personal data processing and general networks became possible in the 1990s. Introduction of intelligent systems to this kind of application environment has resulted various integrated tools. (Leiviskä 2001) Intelligent applications have been developed for fibre line, chemical recovery, bleaching, paper mill, and water treatment (Juuso 2004c). The handling of nonlinearities is important in these applications with nonlinear reaction kinetics, long time delays, variable raw material characteristics and many interconnected variables. The control actions must be designed to meet multiple and sometimes conflicting objectives, changing operating conditions and uncertainty of measurements. The operation may also be upset by disturbances, such as changes in the composition and/or properties of the raw materials. The process variables are maintained within predefined constraints to assure safe operation.

Efficient integration of subprocesses is important in the pulp and paper industry because of advanced recycle flows: the chemical recovery of the chemical pulping process includes an additional recovery loop for the lime, and internal water circulation is essential as washing is done in many stages of the process. The water is tried to keep in the circulation with a capable treatment although the amount and quality of water
fluctuates greatly depending on the process conditions. All disturbance of the purification result will later cause new disturbances in the pulping or paper making processes. (Juuso 2004c) If the dosing control of chemicals in water treatment processes is far from ideal, occasional inefficient plant operations cause unnecessary costs and decreasing water quality. However, the development of control strategies has been increased due to tightened environmental requirements and decreasing water consumption in pulp and paper mills. (Joensuu et al. 2004) In pulp mills, the efficient use of chemicals, water and energy requires the adaptation of the nonlinear multivariable processes into an ever changing operating environment (Juuso 2004c).

Sustainable energy plants provide a viable, cost effective alternative to power production if the fluctuations of the primary energy source are efficiently handled. For solar thermal power plants, the energy source, the sunlight, has seasonal and daily cyclic variations, and the irradiance also depends on atmospheric conditions. The efficient operation of such plants requires fast control actions, which are adapted to the operating conditions. Various control methodologies have been tested at a solar thermal power plant in the south of Spain. (Juuso 1999a) In waste handling, actions are required from operators to optimise the burn-out of the waste in a given position on the grate. Reliable sensors are needed for automatic detection of the burn-out position. The measurements do not always give the correct information, which leads to wrong control actions, and eventually to switch-off of the automatic control strategy. In many things, the control strategy relies on inputs from the operator, who has lots of additional information from the manual detection of the process. (Oestergaard 2004) In addition biofuels generated from biomass can be used in production processes, e.g., there are good results from a lime kiln of a pulp mill (Järvensivu et al. 2001).

In production planning and scheduling, more real-time information should be used for optimization. Especially, short term scheduling requires adaptation to the changing operating conditions. In midterm scheduling, an aggregate optimization problem can be handled by an optimization-based inference engine. This optimization based procedure is combined with rule-based approaches (Jagdev 1989), but an inappropriate sequence of scheduling rules may drop out some reasonable alternatives (Juuso & Jagdev 2007). In water power stations, even narrow cavitation-free power ranges can be utilised in load optimisation if the cavitation risk is detected real-time. (Juuso & Lahdelma 2009).

Managerial decision making in the business world differs considerably from theoretical models, which comprise of both qualitative and quantitative procedures (Singh & Bennavil 1991). Significant interaction between the strategic and operational
decision making is required to ensure the effectiveness of the acquisition, the use of resources and the performance of the operations. Strategic decision making is based on relatively uncertain information and predictions over a long term horizon. Operational decisions are made more frequently over short time horizon. (Juuso et al. 1993) The management of engineering assets has traditionally been dealt with reliability, utilization and availability. Companies started to recognise maintenance as a specific and autonomous function around 70 years ago, and preventive maintenance became very important in the 1940s (Farinha 2010). Condition-based maintenance (CBM) attempts to avoid unnecessary maintenance tasks by taking maintenance actions only when there is evidence of abnormal behaviour (Jardine et al. 2006). The concept of total productive maintenance (TPM) meant a move from corrective maintenance to "deterioration prevention" and "maintenance reduction", which has closely connected to the concepts of total quality management (TQM) in operations and just on time (JOT) in material flows (SCEMM 1998). In all these approaches, the whole process chain and all the employees are important (Tennant 2001). Plant top performance requires integration of three central areas: TPM, TQM and JOT (SCEMM 1998). The reliability of operation, high quality, safety and environmental issues are all important parts of competitiveness.

Changes in operation, process equipment and products should be detected at an early stage to minimise harmful effects. Condition monitoring (CM) provides a reliable and economical way of carrying out CBM. Considerable benefits can be achieved by recognizing which machines are most important in terms of production and devising an appropriate health monitoring strategy for them. An increasing number of measurement points and more demanding problems require automatic fault detection. Condition monitoring professionals have more time to identify root causes when a system takes care of some routine tasks (Lahdelma et al. 1999). Signal processing and feature extraction provide new possibilities for expanding the CM activities. The efficient integration of automation and condition monitoring introduce more real-time information to maintenance and asset management processes (Juuso & Lahdelma 2011b). Inspection is used in manufacturing, but false alarms should be kept at minimum without introducing escape faults (Wei et al. 1999). Functional testing is needed for automating the production of highly customized fault-free products (Gebus & Juuso 2002, Gebus et al. 2009).

Adaptive systems are needed in nonlinear multivariable systems for modelling, control and diagnostics. Multilayer simulation combined with expert systems provides a
method for comparing the production alternatives (Juuso 1991) of process designs, in which the parameters of the detailed models depend on the process conditions (Juuso & Leiviskä 1992). Although the process control, which is based on linearised models, has a limited range of applicability in nonlinear systems, sophisticated nonlinear control methods have not been widely used. Both fuzzy control and conventional stabilizing control should be used together in complex applications (Juuso 1993). Classical supervision and monitoring based on the limit value checking of some measurable variables can be improved by using the analytical redundancy between the process variables (Isermann 1993b): earlier fault detection is possible by combining symptoms based on measurements and symptoms obtained by human operators or by stating the life history, maintenance status and fault statistics. The boolean-probabilistic AI approaches of fault-decision-making by fault trees processing are inflexible and lead to inefficient computational techniques (Ulmer 1993). Soft sensors extend the operation area when the complexity of the control systems increases (Juuso 2005b).

**Uncertainty** is an unavoidable part of many applications, which include decision making. For process design, the range of alternatives is too wide for using detailed calculations already at the initial stage. In business organizations, strategic decision making is often based on relatively unverifiable, often inconsistent, broad scope, aggregated, qualitative, and sometimes rather old information which is obtained from external sources (Juuso et al. 1993). Control applications also have some unmeasurable variables which affect the process operation (Juuso 1999a). In fault diagnosis, many symptoms are based on subjective operator observations or from expert knowledge, which is often inexact and incomplete (Juuso 1994a). Moreover, unpredictable changes may occur in resources and requirements affecting production scheduling (Juuso & Jagdev 2007).

*Smart adaptive systems (SAS)* can be used to develop successful applications in different fields. Three levels of adaptation have been identified (Anguita 2001):

1. Adaptation to a changing environment;
2. Adaptation to a similar setting without explicitly being ported to it;
3. Adaptation to a new or unknown application.

On the first level, a short-term memory is needed for incremental or on-line learning, a long-term memory for recognising context drifting. Successful past solutions and the idea of reasoning by analogy are used on the second level. The most challenging requirement is to adapt to new applications. In real applications, the constraint of starting
from zero knowledge is modified to building new knowledge or, at least, improving existing knowledge.

Different application areas have similar requirements: *changes in operating condition need to be detected, and decision making, including control, must adapt to the changes.* There are increasing challenges: faster action should be taken to respond to the changes, which should be detected earlier. Using a higher automation level with a multitude of suitable methodologies makes this possible. However, *data mining* needs to be combined with *domain expertise* to develop better systems for measurement handling, diagnostics and adaptation procedures utilising, e.g., rule-based systems and learning from experience. *Complex smart adaptive applications need an integration framework to combine different specific intelligent systems.*

1.2 Research problem and asserted hypothesis

Development of *smart adaptive systems* for nonlinear, complex, multivariable and highly interactive industrial processes is a challenging task. In overall production processes, control systems handle several subprocesses, some of which contain important quality variables that can only be estimated from other measured variables. The systems need to take into account many and long time varying delays, process feedbacks at several levels, closed control loops, factors cannot be measured and interactions between physical and chemical factors. Significant interactions between process variables cause interactions between the controllers. Also physical limitations of the actuators must be taken into account. The time delays are highly dependent on operating conditions and they can dramatically limit performance or even destabilise the closed loop system. Uncertainty is an unavoidable part of the process control in real world applications since there are always some unknown factors affecting the process conditions. Successful applications require the integration of data-based methods and expertise, especially if fast reactions to changing operating conditions are needed.

The industrial focus is on the *smart use of intelligence*, i.e., the use of the practical and interactive small scale systems in process and production control to handle complexity. Combining the functions and features of smart adaptive systems provides a basis for adaptation both to a changing environment and to a similar setting. Classical on-line adaptation is not sufficient when there are strong and fast changes. Fuzzy self-organising approaches extend adaptation possibilities, but they are too slow to react, if the operating
conditions change continuously. Dynamic intelligent simulation is a fast and reliable method for the tuning of the adaptation mechanisms. (Juuso 2004b)

Individual subsystems can be constructed with various statistical and intelligent methods, there is a number of efficient simulation and programming tools available. However, a huge number of interactions should be handled in smart adaptive systems. Complexity is the main problem in developing integrated applications. Therefore, the basic modelling approach should be compact and easy to tune. The *linguistic equation (LE)* approach introduced in 1991 produces very compact nonlinear models from expertise and data (Juuso 2004e). The results of the research have led to various applications (Fig. 1). In LE models, the meaning and the interactions are separated as in fuzzy set systems, and the interactions are handled with various linear methodologies.

This thesis concentrates on the LE approach and smart adaptive systems (SAS). The research problem is twofold: *How to implement SAS in modelling and control using LE, and how to use LE in developing hybrid SAS.* The overall research problem is divided into the following subproblems:

– Nonlinear scaling in modelling, control and diagnostics,
– Model structures and statistics in regression analysis,
– Tuning of multilevel, interacting LE systems, and
– Links to other intelligent methodologies.

The applications also involve variable selection, data pre-processing, signal processing, feature extraction, clustering, and approximate reasoning. However, *this thesis focuses on adaptive nonlinear scaling, which is the core of the LE approach.* System structures, modeling and tuning are important, but application specific analysis is needed to compare their applicability. This thesis examines, if *the scaling methodology can be used to represent variable meanings in a compact way and to introduce intelligent indices for control and diagnostics and whether these indices are useful in building smart adaptive applications or not?*
Fig. 1. Evolution of the LE approach.
1.3 Contribution

The author of this thesis created the basic idea of the linguistic equation (LE) approach already in 1991 (Juuso & Leiviskä 1992). Since then he has developed the LE methodology further through active participation and contribution in various research projects on intelligent methods. The research is divided into five phases. The first three phases (1991-2004) form the structural basis for the integration of intelligent systems:

- 1991-1995: The LE approach was originally used for simulation and modelling. The first phase was closely connected with fuzzy logic: fuzzy rulebases were represented by an integer equation, and membership functions were generated with nonlinear scaling functions, denoted as membership definitions. Fuzzy models and controllers were tuned with the LE approach. The main applications were in decision support and production scheduling. Fault diagnosis was studied in connection with the fuzzy logic controllers (FLCs).

- 1996-1999: In the second phase, the data-driven LE modelling was introduced. Applications expanded to process control and first systems were developed for fault diagnosis. The LE approach was mainly used in fuzzy set systems. Genetic algorithms and neural networks were used in tuning and the idea of linguistic neural networks was introduced. The first direct LE-based controller was implemented in 1996 for a solar thermal power plant. The structure of the basic LE controller was reused in the kiln control with some modifications. Intelligent analysers, e.g. the fuel quality analyser, became important parts of control solutions. Case-based reasoning (CBR) with LE-based models was introduced to analyse web breaks in paper machines. Applications extended to inspection and functional testing in electronics manufacturing.

- 2000-2004: The third phase started with extensive use of dynamic LE simulation leading to controller tuning and forecasting. Fuzzy set systems were used for smooth transitions between operating areas and for modelling special cases. Distributed parameter models were developed for the solar power plant. Applications were mainly based on the LE approach. The main areas in LE control were solar energy collection, lime kiln, and, as a new area, water treatment. A new water quality indicator improved LE control in water treatment. New application areas in forecasting were granulation, batch cooking, and fed-batch fermentation. LE-based CBR was extended to continuous brewing and multisensor fault detection for condition monitoring.
The research of these three phases is discussed in (Juuso 1999a, 2004e). The integration aspects are presented in a roadmap (Juuso & Leiviskä 2004) and methodologies are summarized in white papers (Juuso 2004g,f). The LE approach was introduced as a methodology which can be connected with other intelligent methodologies in various smart applications.

The first application already included the idea of monotonously increasing nonlinear scaling functions. This thesis concentrates on the new mathematical approaches developed after 2004 to improve the development and adaptation of the scaling functions and new intelligent indices based on them. The research is classified in two phases:

– 2005-2008: In the fourth phase, the LE approach was combined with fuzzy calculus to handle uncertainties of the models, and new approaches for feature extraction and fault diagnosis were introduced in model-based condition monitoring. LE-based condition and cavitation indices were taken into use. Several grouping methodologies were integrated into the LE modelling to select variables. Detection of operating conditions became important in new more complex modelling applications, e.g. biological water treatment. Generalised moments and norms provided new tools for feature extraction.

– 2009-2012: The fifth phase started with new constraint handling methods and a compact tuning approach and continued with a novel methodology to extract membership definitions from data. This methodology is suitable for recursive analysis. The present LE approach also provides a basis for generating type-2 fuzzy systems to be used for detecting the need for the recursive analysis. These new methodologies are essential in hybrid systems. A LE-based trend index was developed and extended to detecting trend episodes, triggering recursive tuning and modelling for prognostics. New intelligent analysers were developed for LE controllers and LE-based condition monitoring. The first application area, decision support, got new tools when the overall equipment effectiveness (OEE) was analysed with the LE methodology.

The phases IV and V return to the main idea of linguistic equations: the meaning of variable values and trends, quality control and detection of the changes in operation and equipment condition trigger actions in smart adaptive systems. The integrating aspects introduced during the phases I - III are discussed to present the ideas of building smart adaptive applications for complex systems.

The author has supervised many projects, and his main contribution has been in developing new methodologies to be used in applications. Each application has required
a lot of work in process studies, experiments, data acquisition, data pre-processing, signal processing, interactive modelling, and comparing with other methodologies. In most projects, researchers have done this application specific work, and the resulting experience and feedback have been the drivers for new features in the LE approach. The detailed analysis of the results is not included as it also requires a detailed process analysis, which is partly done by the researchers. Detailed discussion would also require a deeper state of the art analysis in each application area. The LE approach forms an integrating basis for smart adaptive systems: each application covers only a small fraction of the properties of the overall system.

The aim of this thesis is to present the methodology of the adaptive nonlinear scaling and the integration principles in different application types.

1.4 Main results

Smart adaptive systems (SAS) are handled as hybrid systems developed by integrating compact linguistic equation (LE) modules. The basic idea of the LE approach was introduced 22 years ago, and since then it has been developed in close connection to various applications in modelling, control and diagnostics. This thesis concentrates on the new adaptive scaling approach, intelligent indices and the parametrisation of the systems developed during the research phases IV and V. The new methodologies of the LE-based development of smart adaptive applications provide following results:

– The nonlinear scaling methodology based on statistical analysis enhanced with domain expertise is the corner stone of the LE approach. The new data-driven methodology, which is suitable for wide variety of nonlinear distributions, extends the scaling to recursive adaptation.

– The compact parametric approach extends the well-known and widely used linear methodologies to nonlinear systems in steady-state, dynamic and case-based modelling. This thesis concentrates on summarising the overall model structures, which are used together with nonlinear scaling.

– The multilevel LE control approach, which combines feedforward and feedback LE control with intelligent analysers, is enhanced with adaptive and model-based modules and high level control, where different control modules are activated smoothly when needed.
– The condition, stress and trend indices are used directly to detect changes of operating conditions and indirectly to introduce facts with uncertainty for fault diagnosis and performance monitoring.
– In decision making, the nonlinear scaling provides new tools for performance analysis, scheduling and allocation. The meanings have natural language interpretations to use domain expertise efficiently.
– The LE systems are efficiently parameterised for tuning even in large scale systems. The parameters are defined and adapted separately for the nonlinear scaling and the systems, which provides insight and robustness.

This thesis focuses on the key elements of the LE systems and application specific differences in the requirements of smart adaptive systems.

1.5 Contents

The advanced data analysis, modelling, control and detection of operating conditions form the basis of this thesis. Chapter 2 provides a short overview of adaptive system alternatives and their applications to software sensors, nonlinear multivariable control and fault diagnosis. As this huge area cannot be adequately covered, this section concentrates on the main features needed in assessing the approach presented in this thesis. Emphasis is on intelligent methods and their possibilities and usefulness in smart adaptive systems. Also the advances of the LE approach during the first three phases are discussed in this context.

Meanings of the variable levels and the interactions of the variables are important in the LE approach. The LE methodology is presented in Chapters 3, 4, 5 and 6 with special emphasis on the new advances after 2004. The adaptive nonlinear scaling explained in Chapter 3 is the key methodology. The LE modelling and the tuning possibilities of the LE systems are the main topics in Chapter 4. Contributions of the LE approach to process control with emphasis on intelligent analysers and model-based adaptation are presented in Chapter 5. The new approaches of trend analysis, intelligent stress and condition indices are discussed as a part of the process control. Chapter 6 extends the LE approach to hybrid systems with a focus on LE, and how it is integrated with other methodologies.

The application examples of the LE approach are summarised in Chapter 7 by concentrating on the LE methodologies in multivariate modelling and simulation, nonlinear multivariate control, detection of operating conditions, and decision making.
Chapter 8 integrates methodologies and application types and compares the LE approach to the methodologies presented in Chapter 2. Chapter 8 includes the main ideas of the LE-based smart adaptive systems and applications. Data collection, pre-processing, signal processing, variable selection and feature extraction are essential in all applications. This thesis concentrates on the LE methodology, and therefore these topics as well as learning and optimisation approaches are not discussed in detail. Chapter 9 presents the conclusions of the thesis with special emphasis on the recent developments of the LE methodology and its applications.
2 Adaptive systems

Nonlinear multivariable systems can be developed with various statistical and intelligent methodologies, which can focus on modelling, clustering and classification (Fig. 2). Smart adaptive systems also need filtering, soft sensors, model-based control and extracting symptoms for fault diagnostics. Sensor fusion combines models with data pre-processing, signal processing and feature extraction (Fortuna et al. 2007). Fault diagnosis is based on symptoms generated by comparing process models and measurements (Frank et al. 2000), signal analysis (Lahdelma & Juuso 2011a), limit checking of measurements (Fortuna et al. 2007) and human observations (Isermann 1997). All these are used in the intelligent control and detection of operating conditions, which introduce reasoning and decision making to the smart adaptive systems. The hybrid nature is seen in literature where these topics are usually combined from methodological perspectives, e.g. fuzzy and neural methodologies are used in various types of applications (Fig. 2).

This chapter classifies known statistical and intelligent methodologies to find feasible combinations for smart adaptive applications. The first three development phases of the LE approach form the background of this thesis (Section 2.5).

Fig 2. Methodologies for developing smart adaptive applications.
2.1 Nonlinear multivariate methodologies

Data-driven methodologies have been introduced to steady-state modelling and extended to dynamic systems by special structures. In the control area, these techniques known under the term system identification (Ljung 2008) include model development and tuning. Steady-state models can be applied in dynamic simulation by calculating the derivative and then using numerical integration methods. Statistical, fuzzy, neural and neurofuzzy methods have been applied first separately and later in combined approaches where cascaded models are based on decomposition. The increased complexity means additional challenges for development and tuning: models should show good agreement with data, but too complex structures may cause overfitting or switching problems between models.

2.1.1 Steady-state systems

The steady-state simulation models can be relatively detailed nonlinear multiple input, multiple output (MIMO) models $\vec{y} = F(\vec{x})$, where the output vector $\vec{y} = (y_1, y_2, \ldots, y_n)$ is calculated by a nonlinear function $F$ from the input vector $\vec{x} = (x_1, x_2, \ldots, x_m)$. 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 $F(\vec{x})$.

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 (Box & Wilson 1951). 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 (Ljung 1999). 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 (Jolliffe 2002). Partial least squares (PLS) regression uses potentially collinear variables (Gerlach et al. 1979). Nonparametric models for $y_i$ at each $\vec{x}$ can be constructed from data as a weighted average of the neighbouring values of $y_i$. The smaller the neighbourhoods, the more complex or flexible model. (Wasserman 2007)
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 (Dubois et al. 1999). 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 \(\alpha\)-cuts of the fuzzy number \(x_i\) for evaluating fuzzy expressions (Buckley & Qu 1990, Buckley & Hayashi 1999, Buckley & Feuring 2000). The fuzzy sets can be modified by intensifying or weakening modifiers (De Cock & Kerre 2004). Type-2 fuzzy models introduced by Zadeh in 1975 take into account uncertainty about the membership function (Mendel 2007). Most systems based on interval type-2 fuzzy sets are reduced to an interval-valued type-1 fuzzy set.

Linguistic fuzzy models (Driankov et al. 1993), 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. The input-output mapping is realized by the fuzzy inference mechanism equipped with conversion interfaces, fuzzification and defuzzification. Fuzzy set systems can also handle contradictory data (Krone & Kiendl 1994, Krone & Schwane 1996). Takagi-Sugeno (TS) fuzzy models (Takagi & Sugeno 1985), where each consequent \(y_i, i = 1, \ldots, n\), is a crisp function of the antecedent variables \(\vec{x}\), can be interpreted in terms of local models. The consequent is a parameterized function, whose structure remains constant. For linear functions, the standard weighted mean inference must be extended with a smoothing technique (Babaška 1998). 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 (Pedrycz 1984) allow one particular antecedent proposition to be associated with several different consequent propositions, i.e. all the antecedents are tied to all the consequents with different weights.

Artificial neural networks (ANN) are used as behavioural input-output models consisting of neurons. The response of each neuron is obtained by an activation function, whose input is calculated from a weighted sum of the normalised variables and a bias term. Network architectures differ from each other in their way of forming the net input, the use of activation functions and the number of layers. Activation functions are binary, linear or nonlinear. Rosenblatt (1961) created many variations of the perceptron for
pattern classification. Linear networks correspond to the models with linear terms in RSM models. The most popular ANN architecture is the multilayer perceptron (MLP) with a very close connection to the backpropagation learning (Rumelhart et al. 1986).

Neurofuzzy systems use fuzzy neurons to combine the weight factors and the inputs. The activation function is handled with the extension principle from the fuzzy input, which is obtained by the fuzzy arithmetics. Different combinations with fuzzy and crisp weight factors and elements can be used in these models (Fullér 2000). Also the 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 (Jang 1993).

A function expansion presented in (Ljung 2008) provides a 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 (SVMs), see (Ljung 1999).

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 (Driankov et al. 1993). T-norms and S-norms can be used in neurofuzzy systems if the inputs are normalised to the range $[0, 1]$ (Fullér 2000).

Dependencies between variables can also be described with Bayesian (or belief) networks, where a probability model is fitted to a set of data. The result is summarised by a probability distribution of the parameters. (Gelman et al. 2004).

2.1.2 Dynamic systems

The dynamic models are represented by first order ordinary differential equations (ODE), which are solved by integration over time from the initial conditions. The derivative is a nonlinear function, and some algebraic equations are needed for handling material properties and experimental correlations. Dynamic phenomena are usually tackled with numerical integration methods. Stability, accuracy, calculation time and memory together with special requirements as stiff systems, discontinuities and event handling, are important topics in modelling and simulation. Step size control based on error estimates is efficient if there are both rapidly changing transients and almost steady-state conditions.
For **parametric models**, the output is computed as a linear combination of past inputs and past outputs. The number of delayed inputs and outputs is usually referred as the model order(s). Linearised theoretical models are the basis of **transfer functions** and **linear state-space models**. Some unknown of uncertain parameters could be estimated with nonlinear identification but in practice the problem is very demanding (Ljung 2008). Linearised models are typically developed around some physical equilibrium. (Ljung 1999)

**Dynamic fuzzy models** can be constructed on the basis of state-space models, input-output models or semi-mechanistic models (Babuška et al. 1997). In the state-space models, fuzzy antecedent propositions are combined with a deterministic mathematical presentation of the consequent. The **NARX /Nonlinear AutoRegressive with eXogenous input model** use a finite number of past inputs and outputs (Babuška & Verbruggen 2003). **Dynamic ANN models** are based on similar structures as the dynamic fuzzy models, e.g. a static feedforward network and an external feedback connection when the delays are taken into account in the input vector (Babuška & 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 weight factors can also depend on time.

### 2.1.3 Decomposition methodologies

A modelling problem can be divided into smaller parts by developing separate models for independent subprocesses which are interconnected with process streams. **Decomposition** can be continued within process units, e.g. 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 (Juuso 1980, 1990). In addition to spatial or logical blocks decomposed modelling can be based on different frequency ranges.

Hundreds of **clustering algorithms** have been developed for 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. (Äyrämö & Kärkkäinen 2006). Numerous **fuzzy clustering** algorithms have been proposed and applied to a variety of real-world problems (Bezdek 1981). **Fuzzy c-means (FCM)**
clustering is a partitioning-based method: each data point belongs to a cluster to some degree membership. *Subtractive clustering* (Chiu 1994) is an algorithm for estimating the number of clusters and the cluster centers according to the parameters of the algorithm. *Neural clustering* use competitive networks based on competitive layers, e.g. self-organising maps (SOM) have several alternatives for calculating the distance in the competitive layer (Kohonen 1995). The response of a radial basis functions (RBF) neuron is obtained from an exponential function (Chen *et al.* 1991).

The algorithm with the standard Euclidean norm imposes a spherical shape on the clusters, regardless of the actual data distribution (Babuška 1998). Gustafson & Kessel (1979) extended the standard algorithm 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 (Äyrämö & Kärkkäinen 2006). An optimal *number of clusters* is selected iteratively by using some quality criteria, see (Windham 1981).

*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 (Ljung 2008). *Piecewise affine (PWA) systems* are based on local linear models, more specifically in a polyhedral partition (Sontag 1981). The models can be state-space models or parametric models. The model switches between different modes as the state variable varies over the partition (Ljung 2008).

*Fuzzy models* can be considered as a class of local modelling approaches, which solve a complex modelling problem by decomposing into a number of simpler understandable subproblems (Babuška *et al.* 1997, Babuška 1998). The smoothing problem around the submodel borders of TS fuzzy models needs special techniques, e.g. smoothing maximum (Babuška 1998), or by making the area overlap very strong. *Multiple neural network systems* improve generalisation through task decomposition and an ensemble of redundant networks (Shields & Casey 2008).

A *mixed approach* using both the rigorous first principles and the black box modelling in an integrated environment is a interesting alternative for complex systems, see (Macías-Hernández 2004). Ljung (2008) 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.

### 2.1.4 System development and tuning

Model development, also referred as estimation or learning, is based on the minimisation of the difference between the measurements and the model. System identification is aimed for developing dynamic models from observed input-output data, see (Ljung 2008). The graphical inspection of dispersion plots is a starting point of variable selection. Rallo et al. (2002) use self-organizing maps to project the subsets of input variables along with the output variables onto network output space. More quantitative criteria are reviewed in (Warne et al. 2004). Knowledge-based methods can be used in decreasing the number of variables before variable grouping with data analysis (Ahola et al. 2007).

Performance of the model, also denoted as the model fit, can be measured as a residual, the difference between the observed and predicted data. Performance measures include, among others, the least mean square error (MSE), the mean absolute error (MAE) and the sum squared error (SSE). The patterns of the residual may provide hints for a better model structure, e.g. polynomial and exponential functions. Validation is needed in all modelling approaches to gain confidence on the model. The average fit to validation is worse than the fit to estimation data, and more deterioration is expected when the complexity of the model increases. There are several analytical criteria to quantify this, see (Ljung 2008). The complexity penalty terms, also known as regularisation terms, depend on the number of data, the number of parameters, the norm of the parameter vector, the number of rules, etc. For example, MSE with regularisation improves generalisation by adding a term that consists of the mean of the sum of squares of the coefficients.

The regression is usually based on assumption that the distribution is normal, but special methods have been developed for other distributions, e.g. binomial, gamma, inverse Gaussian, lognormal or Poisson. Robust regression methods have been developed for making the models less sensitive to the effects of outliers by gradually reducing the weight of the points that are far from the predicted values. Model selection algorithms are important in the RSM methodology, e.g. least angle regression (LARS), are used.
for finding an efficient prediction with less parameters (Efron et al. 2004). Expert knowledge and robust fitting techniques can improve RSM models (Morgenthaler & Schumacher 1999).

In neural computing, learning rules are defined as procedures for modifying the weights and biases of a network: various optimisation methods, e.g. conjugate gradient, quasi-Newton and Levenberg-Marquardt, have been used in these networks to speed up learning. Since overfitting is the main problem, the generalisation is tried to improve for example with ensemble learning with early stop (Lampinen 1997) and regularisation techniques (Foresee & Hagan 1997).

Fuzzy linear regression has two types of approaches: linear programming minimises the total vagueness (Tanaka et al. 1982), and fuzzy least squares (FLS) calculates the distance of the fuzzy numbers (Diamond 1988). Bargiela et al. (2007) proposed an iterative algorithm for multiple regression with fuzzy variables. In multiobjective identification of TS fuzzy models, a huge number of parameters need to be tuned and updated when the process conditions change (Johansen & Babuška 2003). A fuzzy set system can be seen as a layered structure (network), similar to radial basis networks (Babuška & Verbruggen 2003). This representation provides various techniques for tuning, e.g. the ANFIS method (Adaptive-Network-based Fuzzy Inference Systems) introduced by Jang (1993) is suitable for the tuning of membership functions for TS fuzzy models.

Genetic algorithms (GAs) produce satisfactory solutions in large search space with noisy data (Szczerbicka 1994) and assist other methods of computational intelligence in structure optimisation (Juuso 1996). The GAs suit well for tuning the gradual rule-based systems to approximate nonlinear mappings between input and output signals, e.g. a strict fuzzy partition and triangular membership functions provide a basis for an efficient coding (Lotvonen et al. 1997). Complexity penalty terms are needed to handle changes in the number of parameters.

Recursive methods are used to estimate models to make some decision on-line, as in adaptive control, adaptive filtering, or adaptive prediction. Recursive identification, adaptive parameter estimation, sequential estimation, and on-line algorithms are used for these algorithms, which update the parameters within selected memory horizons. Also abrupt changes can be handled. (Ljung 1999). Incremental training can be applied to both static and dynamic ANNs: adaptive linear networks are typical examples of this. Recourse learning algorithm for TS fuzzy models has been presented in (Angelov & Filev 2004).
2.2 Soft sensors

Soft sensors compensate the lack of appropriate on-line measurements and extend the use of direct measurements with more informative measurements. The soft sensors may reduce the need for measuring devices, but they can also work in parallel with hardware sensors. Typical steps of the data-driven development are the selection of data, outlier detection, data filtering, model structure selection, model estimation and model validation (Fortuna et al. 2007). Signal processing and feature extraction are needed to provide useful material for the nonlinear multivariate methods in practical sensor fusion.

2.2.1 Data collection and pre-processing

Pieces of the informative and reliable datasets are selected in such a way that the data may contain measurement sets from several experiment periods. The multiple-experiment sets and selected data periods must be handled appropriately, especially in dynamic modelling. Feedback effects, narrow operating areas and unknown disturbances cause problems in modelling. Designed experiments are needed if the data material is not sufficient for modelling (Hinkelmann & Kempthorne 2008). In industrial applications, the primary goal is to extract the maximum amount of unbiased information regarding the factors affecting a production process from as few (costly) observations as possible.

Normalisation or scaling of the data is needed since measurements with considerably different magnitudes cause problems in modelling. Widely used min-max normalisation matches the values between the minimum and maximum to the range $[0, 1]$. The operating point $c_j$ is fixed in Z-score,

$$p_j = \frac{x_j - c_j}{\Delta c_j},$$

which is calculated about the arithmetic mean, $c_j = \bar{x}_j$, by using the standard deviation of the variable $\Delta c_j = \sigma_j$, transforms the values to a distribution with mean of 0 and standard deviation 1. The arithmetic mean and standard deviation are optimal if the data sample comes from a normal distribution. The sample arithmetic mean also is sensitive to data entry errors, e.g. outliers.

Data distributions should be taken into account in estimating the centre point $c_j$ and developing the scaling functions. The geometric mean and harmonic mean are useful when the sample is distributed lognormal or heavily skewed. The median and trimmed mean are two measures that are resistant (robust) to outliers. The trimmed mean ignores
a small percentage of the highest and lowest values of a sample when determining the centre of the sample. Scaling with the median and median absolute deviation, i.e. $c_j = \text{median}(x_j)$ and $\Delta c_j = \text{median}(|x_j - \text{median}(x_j)|)$, provides a solution, which is insensitive to outliers and the points in the extreme tails of the distribution. Decimal scaling, where the values are scaled by $10^{\log_{10}\max(x_j)}$, suits for cases where the ranges of the variables vary by a logarithmic factor. Minimum and maximum values are very sensitive to outliers.

The outliers, which are unusually large disturbances caused for example by temporary sensor or transmitter failures, should be removed from the data. This can be done step by step by examining more thoroughly the data corresponding to the unusually large residual values. An observation is often considered as an outlier if the absolute value $|p_j|$ obtained by (1) is greater than 3. The Joliffe method has been introduced to detect observations that do not confirm with the correlation structure of the data (Fortuna et al. 2007, Warne et al. 2004). A survey of outlier detection methods is reported in (Englund & Verikas 2005). As statistical inspection of process data tend to remove peaks which can carry precious information about system dynamics, all available information, including expert knowledge and input-output relationships, should be used (Fortuna et al. 2007).

**Nonlinear** activation functions, log-sigmoid and hyperbolic tangent, are used to generate the neuron outputs from the sum of the weighted inputs and the bias. These functions have been modified to improve the normalisation of the matching scores in multimodal biometric systems (Jain et al. 2005, Snelick et al. 2005). The log-sigmoid function, \((1 + \exp(-2p_j))^{-1}\), can be used for nonlinear scaling from the z-score values $p_j$ to the range $[0, 1]$. The double sigmoid function extends this with different linear characteristics in the intervals $[c_j - \Delta c_j, c_j]$ and $[c_j, c_j + \Delta c_j]$. The operating point $c_j$ and the edges $\Delta c_j^-$ and $\Delta c_j^+$ are tuned. The sigmoid function is related to the hyperbolic tangent $\tanh(\frac{1}{2}p_j)$, which scales to the range $[-1, 1]$. The functions introduced by Snelick et al. (2005) are based on the scaling of the min-max normalised values with two functions, which are quadratic, logistic and combined linear and quadratic: the inflection point is in the range $[0, 1]$.

The clustering algorithms discussed in Section 2.1.3 can also be used for compressing large datasets: the cluster centres will replace the corresponding datapoints, and the compressed data is used for modelling. Interpolation is needed if measurements are not frequent enough or if the sampling period is not constant, e.g. various laboratory measurements are based on samples taken infrequently compared to the on-line measure-
ments. In practice, some measurements are missing because of failures in sensors or in data acquisition. These values are either reported as missing or recognised as erroneous values. Missing data can be replaced by using imputation with constants, e.g. the feature or class mean (Enders 2010). Outliers are handled in the same way but with extra care as their difference from the acceptable values can be fairly small. For large data sets, missing values are simply left out, since the imputation may bias the data. Multiple solutions based on clustering or model-based correction form a basis for iteration.

Quality control systems are developed

- to make quality control more effective and closer to real time,
- to identify calibration, measurement and communication errors as close to the observation source as possible,
- to focus on automatic quality control algorithms development,
- to develop a comprehensive flagging system to indicate data quality level,
- to make it easier for data users to identify suspicious and erroneous data, and to highlight corrected values.

Numerous methods are used real-time and non real-time for the spatial and temporal checks of meteorological data (Vejen et al. 2002).

2.2.2 Signal processing and feature extraction

Digital signal processing is needed in software sensor applications to extract informative parts from the measurements (Diniz et al. 2010). Trend removal is used if the variation around the trend is important for the modelling (Thornhill et al. 2004). On the other hand, the data can be smoothed with moving averages or medians. Process signals usually contain many distinct components from various sources (Stephanopoulos & Han 1996).

The filtering on specific frequency bands, e.g. low-pass, high-pass and band-pass, are based on models. Numerous methods have been developed for noise reduction by using linear and polynomial models. State estimation is aimed for predicting the next state from the current, and updating or correcting the prediction when using noisy measurements. The recursively operating Kalman filter, which is based on linear state-space models, assumes that the noise terms are uncorrelated, Gaussian and white with zero mean. Advanced versions developed to handle nonlinear systems have emulated with ANN, e.g. RBFs are used in (Härter & Velho 2008). Wiener filter
seeks a linear time-invariant filter whose output would come as close to the original signal as possible (Wiener 1964). *Adaptive linear filters* can predict the next value of a stationary random process by adapting weights for a set of delayed previous values. The performance is optimised by comparing with the desired response obtained from a reference input (Widrow & Winter 1988). Barreto & Souza (2006) studied adaptive filtering with SOM and RBF.

*Subspace methods* are used for finding the optimal subsets of a larger space. *Independent component analysis (ICA)* has been developed for isolating the effects of a single phenomenon from measured values that include mixtures of several underlying phenomena (Lee *et al.* 2004). Subset selection techniques are in literature understood in wide meaning, e.g. feature selection techniques include modelling, optimisation and classification in (Saeyes *et al.* 2007). Various methodologies are reported as *blind source separation (BSS)* methods to emphasise their use in separating signals to find useful signals (Deville *et al.* 2010).

*Principal component analysis (PCA)* is used to find groups of variables which move together as they are measuring the same driving forces (Jolliffe 2002). The static analysis has been extended to dynamic systems by including past measurements (Ku *et al.* 1995, Li & Qin 2001). Moving PCA is another variant of PCA developed by Kano *et al.* (2001) for detecting changes in operating conditions. A special multi-way approach has been developed for analysing variations from the normal trajectories in batch processes (Nomikos & MacGregor 1994).

*Frequency domain analysis* in its basic form uses the ordinary amplitude spectrum. Advanced analysis methods select specific frequency ranges or further modify the spectra: popular methods are the envelope and PeakVue spectrum, cepstrum and high-order spectra to search for nonlinear interactions (Lahdelma & Juuso 2011a,b). *Fourier analysis* decomposes continuous periodic functions into its constituent by producing an infinite sum of sine and cosine waves. The *fast Fourier transform (FFT)* transforms a function of time into a function of frequency (Newland 1993).

*Derivation* used in vibration analysis means the construction of new signals from measurement signals. The measurement parameters traditionally used in condition monitoring are displacement, velocity and acceleration. The first time and the second derivative of acceleration are very suitable for the condition monitoring of slowly rotating bearings (Lahdelma 1995). Use of real order derivatives allows stepless differentiation (Lahdelma & Juuso 2007). The *FFT* and the inverse Fourier transform (*FFT*\(^{-1}\)) are used in the calculation of the time domain real order derivatives on the basis of a rigorous
mathematical theory (Samko et al. 1993). The mathematical background of fractional
derivatives and feature extraction methods are discussed in (Lahdelma & Juuso 2011a,b).

*Wavelet methods* and *multiresolution* approaches have opened new possibilities for
on-line processing of signal information (Thuillard 2004): the original signal or function
is represented by coefficients in a linear combination of the wavelet functions (Graps 1995).

*Statistical features* calculated in moving windows, i.e. moving variances, standard
deviations or value ranges are highly informative in the time domain analysis. Also
moving skewness and kurtosis can be obtained. Selecting an appropriate window
is important. For statistical analysis, see (Cramér 1971, Rinne 2008, Spiegel et al. 2000). The root-mean-square and peak values, which are the most commonly used in
the vibration analysis, are special cases of the generalised norms (Lahdelma & Juuso 2011a,b). Features can also be obtained from specific frequency ranges, e.g. spectral
kurtosis (Antoni & Randall 2006).

*Temporal reasoning* is a very valuable tool for diagnosing and controlling slow
processes. Manual process supervision relies heavily on the visual monitoring of the
shapes of changes in process variables, especially their trends. Although humans are
very good at visually detecting such patterns, for control system software it is a difficult
problem. Trend indices are obtained by comparing samples with different length.
Cheung & Stephanopoulos (1990) developed a formal framework for the extraction and
representation of process trends with triangular episodic representations. The elements
are defined by the signs of the first and second derivative, respectively. These elements,
which are also known as triangular episodic representations (Cheung & Stephanopoulos
1990), have their origin in qualitative reasoning and simulation (Forbus 1984, Kuipers
1985). This methodology has been applied in fermentation processes (Kivikunnas et al.
1996b, Stephanopoulos et al. 1997). Typical reasoning systems have three components:
a language to represent the trends, a technique to identify the trends, and a mapping from
trends to operational conditions (Dash et al. 2003). More methods for trend analysis are
reviewed in (Kivikunnas 1999, Juuso 2011a).

*Resampling* can be used for estimating the uncertainty by using different subsets
of available data (jackknifing) or drawing randomly with replacement from a set of
data points (bootstrapping) (Garatti & Bitmead 2010). Cross-validation is a statistical
method for validating a predictive model by leaving out a single observation or a subset
at a time.
2.2.3 **Sensor fusion**

In multi-sensor systems, variable selection problem originates the high dimensionality in the data used caused by a high number of sensors or many features extracted, or both. *Gray-box modelling* approaches combine mechanistic models, multivariate statistics and intelligent methods. Any knowledge, regarding the variable selection, the system order, the operating range, time delays, the degree of nonlinearity, sampling times, etc., represents a valuable source of information that should be taken into account (Fortuna et al. 2007).

*Variable grouping and selection* is needed in data-driven modelling (Ahola et al. 2007): expert knowledge gives a clear basis for selecting variables in small systems, but variable selection becomes important when the number of variables increases, especially when normal process data is used. This rapid increase of combinations is known as the combinatorial explosion (Pyle 1999) and it can easily defeat even powerful computers. PCA and PLS regression are a widely used in variable selection, see e.g. (Westad et al. 2003, Cadima et al. 2004, Zarzo & Ferrer 2004). Abrahamsson et al. (2003) studied four different methods - genetic algorithm, iterative PLS, uninformative variable elimination by PLS and interactive variable selection for PLS - and compared them in PLS regression to a calibration made with manually selected wavelengths. A large number of different methods have been used in process monitoring (Venkatasubramanian et al. 2003b).

The sensor fusion is highly applications an methodology specific, e.g. a comparative study of soft sensors derived using multiway PLS and an extended Kalman filter for a fed-batch fermentation process is presented in (Zhang et al. 2005), fuzzy models are used in (Liu 2005) to realize a piecewise linear time-varying model for a polyethylene process, and a model-free trainable fuzzy conversion estimator for a set-up of lactic acid fermentation is presented in (Kivikunnas et al. 1996a).

2.3 **Nonlinear process control**

Normal *feedback (FB)* and *feedforward (FF)* controllers can be extended to changing operating conditions with adaptation, model based approaches and high level knowledge based systems. Intelligent methods provide a good basis for nonlinear multivariable control systems, e.g. a large number of highly successful *fuzzy logic control (FLC)* applications are implemented in process industry. The system structures are similar to the fuzzy modelling, i.e. a FLC is interpreted as a real-time expert system combining
qualitative, fuzzy variables which are linked to the real world by membership functions. Numerous advanced control algorithms have been tested in solar energy systems (Camacho et al. 2012).

2.3.1 Feedback control

Feedback controllers use error \( e_j(k) \), derivative of the error \( \Delta e_j(k) \), and sum of errors \( \delta e_j(k) \) calculated over a predetermined time interval by

\[
\begin{align*}
e_j(k) &= y_j(k) - y_j^{sp}(k), \\
\Delta e_j(k) &= e_j(k) - e_j(k-1), \\
\delta e_j(k) &= \sum_{r=1}^{k-1} e_j(r),
\end{align*}
\]

where \( e_j(k-1) \) is the previous error of the controlled variable \( j, j = 1, \ldots, m \), \( y_j(k) \) the most recent moving average values of the measurement, and \( y_j^{sp}(k) \) the target.

Conventional PID controllers use proportional (P), integral (I) and derivative (D) terms: the coefficients, \( K_P(i, j) \), \( K_D(i, j) \) and \( K_I(i, j) \) are used for the error, the derivative of the error and the integral of the error, respectively. In the PI controllers, the new control, \( u_i(k) \), is obtained by adding the change of control,

\[
\Delta u_i(k) = K_P(i, j) \Delta e_j(k) + K_I(i, j) e_j(k)
\]

to the previous control, \( u_i(k-1) \).

Fuzzy logic controllers (FLCs) can be classified in the same way as the conventional controllers. The controllers operate in a considerably different way. PID-type fuzzy controllers are represented by rules where the antecedents are handled with linguistic labels: \( L_e \) for the error variable, \( e_j(k) \), \( L_{\Delta e} \) for the the change of control, \( \Delta u_i(k) \), \( L_{\delta e} \) for the sum of error, \( \delta e_j(k) \), \( L_{\Delta u} \) for the change of control, \( \Delta u_i(k) \), and \( L_u \) for the control, \( u_j(k) \). A TS fuzzy controller is a collection of linear controllers, which are adapted to different operating conditions, and an interpolation system between these controllers. In singleton type fuzzy controllers, the consequent is represented as a crisp value. The control action, \( \Delta u_i(k) \) or \( u_i(k) \), is obtained as a weighted average.

FLCs are flexible: a rule can contain several controlled variables and a specialised set of rules is built for each manipulating variable, but the antecedent part of the controller can also be the same for several variables, or a single manipulating variable can be in several sets of rules. Traditionally, the rule base is constructed from expert knowledge by trial-and-error. Many controllers are implemented as two dimensional fuzzy logic
decision tables where control actions and input conditions are expressed in terms of membership functions. The rule base is often essentially linear, and the nonlinearity can be introduced by adjusting membership functions or by modifying rules.

Some experiments with automatic control based on deterministic decision tables were carried out with real plants in the 1970s (Oestergaard 1996). The first industrial FLCs were realised in 1978 for a cement kiln (Holmblad & Oestergaard 1982) and in 1979 for a lime kiln (Oestergaard 1996). Industrial kilns have been active control application areas (Oestergaard 1993, Juuso et al. 1996, Järvensivu et al. 2001). Most early applications were based on the acquisition of operator experience, and process knowledge was the basis of several control strategies (Juuso 1999a). The FLC implemented for a cement kiln consisted of several rule blocks corresponding different control objectives (Hantikainen 1992). The knowledge-based FLC implemented for a lime kiln was originally developed in cooperation with the personnel (Haataja & Ruotsalainen 1994) and later extended to a model-based approach (Juuso et al. 1996). A combination of five strategies included in the FLC of the lime kiln was well balanced in normal operation but different tuning was needed in other operating points.

The pulp and paper industry has been active in implementing FLCs as there are nonlinear multivariable processes with heterogeneous raw materials, complicated variable interactions and long delays, see (Lampela et al. 1996, Myllyneva et al. 1997, Juuso 1999a). In steel industry, the FLC developed for the heating control of a coke oven battery improved the adaptation to changing coal blends (Palmu 1998). In an effluent treatment plant, a FLC controls the sludge amounts and the nutrient dosages (Juuso 1999a). For a solar power plant, a PI-type FLC was implemented in 1994 on the basis of the experience of the previous experiments (Rubio et al. 1995) and later replaced by an automatic genetic design technique and used together with a FF controller (Gordillo et al. 1997).

The fuzzy expert control is easy to comprehend and adapt because of its inherent pragmatic and human modelling approach. However, increasingly advanced and complicated control strategies lead problems in system maintenance (Oestergaard 1996). The trial-and-error approach is very time-consuming and forms a bottleneck in controller design, but the number of rules included can be reduced by handling the nonlinearities properly by membership functions. (Juuso 1999a)

Neural control with a large number of control structures is discussed in (Liu 2002). Applications in the lime kiln processes are reported for the identification and control Ribeiro & Correia (1993), for the FF control in conjunction with high-level
FB controllers (Järvensivu et al. 1994) and for the quality prediction of the burnt lime (Järvensivu & Seaworth 1998, Ribeiro 1998). A rule-based kiln control system, where ANNs are used to represent the rule set, has been reported in (Bo et al. 1997).

Sliding mode control (SMC) (Utkin 1977) is designed to converge to a sliding surface based on the plant dynamics, and model based switching controllers use a finite set of fixed controllers. A combination of sliding mode control and fuzzy logic is presented in (Iglesias et al. 2002). The conventional sliding surface is modified by using a set of fuzzy rules, which are similar to that of a FLC. This confers to the controller robustness and flexibility to handle the highly nonlinear behaviour found in a neutralization reactor.

2.3.2 Adaptive control

Adaptive controllers generally contain two extra components, a process monitor and an adaptation mechanism, compared with the standard controller (Driankov et al. 1993). The process monitor detects changes in the process characteristics either by the performance measure or by the parameter estimator. The adaptation mechanism, which updates the controller parameters, can be on-line or predefined. In normal operation, the efficient reuse of controllers developed for different operating conditions is a good practice as the adaptation requires always time.

On-line adaptation includes self-tuning, autotuning and self-organisation. For on-line adaptation, changes in process characteristics can be detected through the on-line identification, or by the assessment of the control response (performance analysis). The choice of performance measures depends on the target response: alternative measures include overshoot, rise time, settling time, delay ratio, frequency of oscillations, gain and phase margins and various error signals (Driankov et al. 1993). The identification block typically contains some kind of recursive estimation algorithm which aims at determining the best model of the process at the current instant.

Plant results with adaptive schemes, e.g. a self-tuning PI controller (Camacho et al. 1992) and a prescheduled adaptive control (Meaburn & Hughes 1994), have clearly demonstrated the importance of adaptive tuning in the solar plant application. Intelligent methods provide additional techniques for on-line adaptation, e.g. Jantzen & Poulsen (2003) studied the update mechanism of a fuzzy self-organising controller (SOC), and a self-tuning fuzzy control of a rotary dryer is presented in (Pirrello et al. 2002). In a meta-rule approach, parameters of a low-level controller are changed by a meta-rule supervisory system whose decisions are based on the performance of the

### 2.3.3 Model-based control

Model-based control is widely applied to industrial applications (Camacho & Bordons 1995). Phenomenological models provide a useful insight into the interactions and time delays of the process, see e.g. (Castro et al. 2001), but the control of an industrial kiln requires adequately accurate models which are not easy achieve (Barreto 1997). Developing kiln control systems based on empirical models, as described e.g. in (Uronen & Aurasmaa 1979, Bailey & Willison 1986) was the second step. The first commercial supervisory-level system for the lime kiln already appeared at the end of the 1970s (Elsilä et al. 1979).

Feedforward control (FF) can be based on models, e.g. most of the controllers tested in the solar collector field use model-based feedforward control based directly on the steady-state energy balance relationships can be based on the measurements of solar radiation and inlet temperature (Camacho et al. 1992). Then even PID controllers operate (Valenzuela & Balsa 1998) and FLCs are improved considerably (Rubio et al. 1995, Gordillo et al. 1997).

Fuzzy logic controllers can use normal state variables, \((x_1, \ldots, x_n)\), instead of error, change of error and sum of error. Then the controllers are presented by rules relating the corresponding labels \((L_{x_1}, \ldots, L_{x_n})\). This is a MISO-type controller with several input variables and one output variable. An example of a FF controller based on an inverted fuzzy model is presented in (Juuso 1999a). An alternative for this controller could be based on Takagi-Sugeno type rules, where the antecedent is a function \(u_i(k) = f(x_1, \ldots, x_n)\). Various methodologies have been developed for inverting fuzzy process models, in most cases an approximate inverse is used. Babuška (1998) introduced an exact fuzzy inverse scheme for fuzzy singleton models.

Internal model control (IMC) compares the process output with the predicted output and uses the inverse model to remove the difference. In principle any types of models can be used, including fuzzy models (Babuška 1998), models based on partial differential
equations (Farkas & Vajk 2002), and nonlinear models based on local linear models (Fink et al. 2002). The classical IMC can operate efficiently in varying time delay conditions (Farkas & Vajk 2002). The IMC approach is a good solution if the model is not too complicated. The scheme can also contain on-line adaptation, e.g. a fuzzy model can be adapted and the consequent parameters are transferred to the inverse model (Babuška 1998). Feedback controllers are needed to cope with modelling errors and disturbances.

In model predictive control (MPC), models are used for predicting the process output over a prediction horizon. Control actions are calculated over a control horizon in such a way that the predicted output is as close as possible to a desired reference, and the first control action in sequence is applied in each step. A considerable amount of identification work is needed to obtain sufficiently accurate models for a wide operating area, see e.g. (Morari & Lee 1999). Intelligent methods can be used at the modelling level, in optimisation and in the specification of the control objectives (Babuška 1998). Stages in the development of algorithms and methods of incorporating fuzzy models into controllers are described in (Postlethwaite 2002). Fuzzy internal models have been used in the MPC approach (Årzén et al. 1999). A MPC using TS fuzzy models is discussed in (Roubos et al. 1999).

Several applications for rotary kilns have been reported in (Smith & Aggarwal 1997, Valiquette et al. 1999, Carter & Rozek 2000). For solar plant applications, a generalized predictive control (GPC) based on a gain scheduling algorithm was handling different operating conditions and sudden perturbations caused by clouds (Camacho & Berenguel 1994). An adaptive GPC approach provided a fast response with high overshoots and some oscillations (Camacho et al. 1994). A nonlinear neural MPC presented in (Arahal et al. 1997) worked fairly well since the ANN models introduced a feedforward effect.

2.3.4 High-level control

Several control objectives should be fulfilled as much as possible in industrial processes, see early rule-based expert systems for kiln control discussed in (Dekkiche 1991, Hall 1993, Hagemoen 1993). High-level control is a natural area for FLCs, which extends the ideas of rule-based systems (Oestergaard 1996, Juuso 1999a): the first FLCs were inspired by manual operation experiences. Oestergaard (1996) presents for cement kilns a control strategy, which contains state index values for the process conditions, an arrangement of the groups of control objectives, smooth priority management, control
objectives for individual goals. The rule blocks are normally formulated as fuzzy rules, but other techniques, like PID, neural networks, mathematical models etc. can be used as well. The program selects the control actions on the basis of process constraints and indices.

Switching control strategies are based on selecting a controller from a finite set of fixed controllers. In (Mosca et al. 1989), a supervisor controller selects the controller by choosing the model with the best fit. This has shown robust characteristics with respect to unmodelled dynamics in the solar plant application (Coito et al. 1997). Each controller has been tuned in order to match a region in the plant operating conditions (Rato et al. 1997). An important difference from the gain scheduling scheme is that each controller can have a completely different structure. Each individual controller can have specific advanced features, i.e. switching control can be used for higher level decisions. A combination of a switching algorithm and MPC is presented in (Lemos et al. 2002).

Event based control, also known as aperiodic or asynchronous control, uses sampling which is event-triggered rather than time-triggered (Åström & Bernhardsson 1999, Årzén 1999, Sandee 2006). The nature of the significant event can vary, e.g. in statistical process control, a new control action is only calculated when a statistically significant deviation has occurred. It is close to a way a human behaves as a controller, and suits for distributed control systems. There are two structures: arrival of new sensor data triggers control, or control is updated only when it required because of the control performance.

2.4 Detection of operating conditions

Advanced supervision and fault diagnosis improve the reliability, safety and economy of technical systems. A fault may lead to a malfunction or failure of the system (Isermann 1997). Knowledge-based procedures are based on analysing analytical and heuristic symptoms with different classification methods. Complexity of analytical, data- and knowledge-based modelling approaches in control and fault detection are compared in (Frank et al. 2000). Detection of operating conditions, novelties and anomalies is a wider task since the process states are not necessarily faults. Symptoms are developed from deviations in measurements, soft sensors or models and used in classification based on nonlinear multivariate methodologies (Fig. 2). Specialised methodologies are based for example on causalities, tree analysis and remembering previous situations.
2.4.1 Symptom generation

Classical supervision and monitoring based on limit value checking of some measurable variables can be improved by using analytical redundancy between process variables (Isermann 1993b): methods are simple and reliable but require relatively large changes in feature values. Signal analysis and feature extraction provide additional possibilities for symptom generation: correlation functions, frequency spectra, autoregressive moving average or characteristic values, e.g. variances, amplitudes or frequencies. Principal components help to understand the driving forces of the system by showing which variables are moving together (Jolliffe 2002).

Symptom generation based on parameter estimation is useful since process faults are to a large extent directly related to process parameter changes (Isermann & Freyermuth 1991). Analytic symptoms are generated by comparing residuals or parameters with the normal behaviour, both fixed and fuzzy thresholds can be used (Isermann 1997, Frank et al. 2000). Heuristic symptoms based on human observations and inspection are related to special noises, colours, smells, vibration, wear, tear, etc. Maintenance history, repairs, former faults, life time and load measures, as well as fault probabilities, constitute additional heuristic information, or part of it can be used for generating analytic symptoms, if fault statistics exists. As uncertainty is an unavoidable part of the fault diagnosis, heuristic symptoms can be represented as linguistic variables or as vague numbers. (Isermann 1997)

In control systems, the purpose of the diagnostics is to detect changes in the controller performance or in the process conditions. Fuzzy controllers include diagnostic features on the coarse fuzzy partition level (Juuso & Leiviskä 1994). If some variables become very high or very low, the working point has changed beyond the area where the fuzzy control system can take care of the problem. Very high and very low values are considered as symptoms.

Condition monitoring of machines is based on measurable changes, e.g. in vibration, temperature, power consumption and number of particles in lubricant. Vibration measurements are widely used for detecting mechanical problems and many electrical failures. The variations in trends and rates of change are basic features when estimating the severity of the faults. Advanced signal processing methods are more sensitive and expose failures earlier: Lahdelma has used higher derivatives for monitoring slowly rotating machinery (Lahdelma et al. 1999). Sensitivity of a feature is at maximum on a feature specific order of the differentiation (Lahdelma 1997, Lahdelma & Kotila 2003).
Integration of maintenance and operation focus on five key words: production, quality, just on time, lean and quick response (SCEMM 1998). Maintenance is defined as risk based (Arunraj & Maiti 2007), preventive (Garg & Deshmukh 2006) or condition-based (Jardine et al. 2006, Hägg 2010). There are several performance measures: harmonised indicators for asset management (Olsson & Svantesson 2009); several performance indicators for process units (Al-Shammasi & Al-Shakhoyry 2010); statistical analysis for reliability-centered maintenance (RCM) to find critical process units (Carpignano et al. 2010). Also, environment is increasingly important (Parida & Kumar 2009). Many aspects of these measures are incorporated in the OEE measure (SCEMM 1998). However, this view could be too narrow, as all these measures are related to the outcomes of the process and maintenance improvements. Performance should be understood in connection with a life cycle (Hägg 2010).

Statistical process control (SPC) is used in monitoring a process and comparing process capability indices (PCIs) related to Six Sigma to identify and remove the causes of defects (errors) and minimizing variability in manufacturing and business processes. Total quality management (TQM) has strongly effected on Six Sigma approach. Performance is evaluated by PCIs, which assume that process output is approximately normally distributed. The Taguchi capability index calculates an off-center process mean. (Tennant 2001, Wu et al. 2009).

2.4.2 Classification and reasoning

Decision making between different operating conditions is based on different types of reasoning. All symptoms or features are presented as facts in a unified form with confidence numbers (Isermann & Freyermuth 1991) or with membership functions (Ulieru 1993). Symptom vectors are mapped into fault vectors by classification or clustering if fault-symptom causalities are not available.

Model-based approach, which uses analytic models, has been a long time an essential part of the fault detection and diagnosis (Isermann 1984). Diagnostic reasoning strategies are based on known fault-symptom causal networks. Faults in processes, actuators and sensors are detected by using process models which represent dependencies between measurable signals. A straightforward method is to run a fixed model parallel to the process and detect an output error. Differences in model parameters or output error, as well as state estimators or observers are used to model-based fault detection.
Data-driven modelling with combinations of various methodologies is widely used in classification and reasoning. ANNs provide various methods for classification starting from perceptrons, which form linear decision boundaries between the classification regions. Frank et al. (2000) replace analytical dynamic models MLP networks. In probabilistic neural networks, the responses of the RBF neurons are fed to a competitive layer to detect classes. A competitive layer is used in SOM for detecting classes. In learning vector quantization (LVQ), a linear layer detects the classification classes by using the subclass output of the competitive layer. ANN models have been used in condition monitoring (Samanta & Al-Balushi 2003, Samanta 2004). SVMs create a line or hyperplane between two sets of data for classification and tries to orient the boundary such that the distance between the boundary and the nearest data point in each class is maximal (Samanta 2004, Samanta et al. 2003). Classification can be based on artificial immune systems (Liang et al. 2006, Strackeljan & Leiviskä 2008), and a generative topographic map, which is a probabilistic counterpart of SOM (Liao et al. 2002). GAs are suitable for tuning the classifiers, for example ANN (Samanta & Al-Balushi 2003) and SVM models (Samanta et al. 2003).

Rule-based systems use IF-THEN rules representing domain expertise. Inference methods are data-driven forward chaining and goal-driven backward chaining. Conflict resolution is applied to select one rule out of the active ones. Fuzzy logic provides feasible solutions for resolving the conflicts and handling the uncertainty in fault diagnosis (Isermann 1993c). The whole structure of the possibilistic fault tree (PFT) can be considered as a texture of chains of abductive links interconnected via fuzzy operators (Uliéru 1993): each cause node reflects a complex fuzzy rule with all antecedents and branches required to enter the respective node, i.e. the complex multiple inference to be divided into a number of simple inferences.

Qualitative models and search strategies are reviewed in (Venkatasubramanian et al. 2003a): causal directed graphs (CDG) represent physical cause-effect relations between variables; fault tree analysis (FTA) and event tree analysis (ETA) provide the pathways which interconnect the basic cause events and conditions to a foreseeable, critical event. Bayesian networks represent probability relations among random variables as a graphical model to be used in probability inference (Pourret et al. 2008). Petri nets are used as a general purpose graphical tool for describing relations between conditions and events (Rochdi et al. 1999). The boolean-probabilistic AI approaches of fault-decision-making by fault trees processing are inflexible and lead to inefficient computational techniques (Ulieru 1993). The fault trees and processing of statistical data neglect the
causal-deterministic structure. Usually, these trees can be used only as an aggregation support (Isermann 1993a). Probabilistic reasoning based on Bayesian networks can include conditional probabilities for causalities, but often the computational effort is reduced by assuming statistically independent symptoms. (Isermann 1997)

**Case-based reasoning (CBR)** is a problem solving paradigm for finding out the solution to a new problem by remembering a previous similar situation and by reusing information and knowledge of that situation (Aamodt & Plaza 1994). In structural point of view, CBR is a cyclical method that stresses the reuse of solutions to similar problems, where solutions are maintained in carefully indexed memory, known as case base. Vehí & Mujica (2003) combined CBR, SOM and wavelet transforms for damage identification. The system resembles the human decision making process and does not need a complete set of information to make a decision.

**Fault-symptom causalities** can be used in forward and backward chaining, for example with Boolean algebra for binary facts and approximate reasoning for probabilistic or possibilistic facts. A fault decision indicates the type, size and location of the most possible fault, as well as its time of detection. (Isermann 1997) In prognostics, models are used to estimate the remaining useful life (RUL). A time-dependent proportional hazard model (PHM) a sum of several exponential terms which are activated through the fault progress. Stochastic processes based on a finite number of states can be defined by hidden Markov models (HMM) (Jardine et al. 2006).

**Quality control** of measurements includes several levels. For example, a Nordic research project classifies four levels for meteorological observations: (1) quality control at a station site, (2) real-time on-line quality control, (3) non-real-time quality control, and (4) human quality control. The methods include limit, step, consistency, spatial and homogeneity checks for the temperature, pressure, precipitation, wind and humidity. (Vejen et al. 2002)

In *novelty detection*, one wants to detect significant changes from the normal or near-normal behaviour, whereas no examples of the current abnormal behaviour leave unobserved. In this sense, it is an one-class approach, where a model (prototypes, probability density, domains in feature space) describes the areas in feature space where normal behaviour is expected. The problem resembles the outlier detection problem. Methods and applications for this problem have been reported in (Tax 2001, Tarassenko et al. 1990, Ypma & Duin 1997, Munoz & Muruzabal 1998). Novelty detection has an extremely challenging task to detect events, which are not present in the training data. A classifier detects if an input is a part of the data used in training, or if it is unknown.
Novelty detection is useful in signal processing, data mining, pattern recognition etc. Several models perform well in these applications, e.g. statistical modelling (Markou & Singh 2003a) and neural networks (Markou & Singh 2003b) are used.

2.5 Discussion

Nonlinear multivariate methodologies are widely applicable in developing smart adaptive systems for modelling, control and diagnostics. Intelligent methods are the key to successful large scale multivariable systems since they combine data and expertise efficiently (Juuso 2004g,f). Methodologies summarised in Figure 2 are discussed in two parts: soft sensors and nonlinear multivariate methodologies form the basis of modelling and measurement handling for the control and diagnostics needed in smart adaptive systems. The analysis also includes contributions of the linguistic equation approach during the phases I-III.

2.5.1 Nonlinear multivariate modelling

Integration of simulators requires consistent modelling approaches which combine modelling traditions on different application areas: process design community relies on phenomenological models, automation design is usually based on data-driven methodologies, and computational intelligence has been applied in some special cases (Fig. 3). For complex industrial processes, emphasis should be on data-driven techniques and computational intelligence since phenomenological models need additional models for selecting appropriate parameter combinations. The process insight is important both in selecting dynamic structures and in evaluating the applicability of the models. Batch processes have additional requirements for forecasting in a quite long horizon. Cascade structures and decomposition are needed for large-scale linear systems, and trade-off between accuracy and complexity becomes increasingly important in nonlinear multivariable systems.

Statistical modelling with data-driven methods uses general function approximators ("black-box" structures) to capture the nonlinear phenomena. In the RSM models, the number of parameters is fairly low, if only few input variables are used, but increases when more interaction terms are introduced (Fig. 4). Identification procedures for parameter estimation for dynamic models are quite straightforward and easy if appropriate data is available. As the structure and parameters of these models do not
necessarily have any physical or chemical significance, these models can be adapted to
different operating conditions only by re-running the identification steps. Since the
operating areas of these models are rather limited, it is natural that recursive methods are
widely used in parameter estimation. Composite local models provide extensions to the
changing operating conditions but limitations of linear LPV and PWA systems should be
kept in mind.

Computational intelligence provides flexible presentations for nonlinear functions.
The maintenance problems of rule-based programming become easier by using fuzzy
systems, which provide clear explanations of the operations through membership
functions and rules. Qualitative knowledge can be incorporated in linguistic fuzzy
models, or in fuzzy relational models if there are several alternative rules. Locally valid
linear models can be collected by TS fuzzy models. Data-based tuning is needed in
computational intelligence since the number of parameters required for the membership
functions becomes very high with an increasing number of variables. More and more
rules are needed when the number of variables increases (Fig. 4). In neural models, the
number of parameters, i.e. weight factors, depends strongly on the number of neurons in
the hidden layer (Fig. 4). Intelligent dynamic models are widely based on the model
structures of parametric identification approaches. To avoid overfitting, the high number
of parameters is reduced with advanced modelling methods: LARS is used for choosing
appropriate coefficients in RMS models, complete rule sets are not usually needed in
fuzzy models, and regularisation methods reduce the number of the active connections

Fig 3. Methodologies and application types of modelling and simulation, modified from

Computational intelligence provides flexible presentations for nonlinear functions.
Fig 4. Number of parameters for different types of multiple input single output (MISO) models: quadratic models and MLP networks with 2\ldots5 neurons in the hidden layer (left), and MLP networks with 5 neurons in the hidden layer and fuzzy set systems with one rule base and a cascade of rules, all variables with 5 membership functions (right).

in ANN models. Evolutionary computing extends the optimisation of parameters and model structures to wide search spaces.

*Cascade modelling* divides the problem into sequential parts to further alleviate the problem of parameters (Fig. 5). The number of parameters is further reduced with principal components, e.g. Model A and/or Model B in Figure 5(a) could produce principal components for Model C. The cascade modelling principle and linear models are essential in various fuzzy and neural methodologies as well. TS models use fuzzy reasoning for weighting local linear models. Radial basis networks are linear combinations of the outputs of the RBF, e.g. in Figure 5(a) Model A and Model B could be radial basis functions and Model C the linear model. The LVQ combines a competitive layer with a linear model. The output of a model can be used as an input of several models (Fig. 5(b)), and the models may also contain interactions or recycle flows (Fig. 5(c)). Feedback structures are needed in dynamic simulation, e.g. feedback connections in Elman networks can be generalised for interactive models (Fig. 5(c)). Neurofuzzy systems can be constructed as sequential combinations of neural and fuzzy parts. Variable grouping is important in cascade model structures.

*Operating conditions* can have drastic effects on the system. Heuristic knowledge and know-how can be introduced to the fuzzy set systems by a trial and error approach.
Data-based approaches usually rely on an automatic generation of rules from predefined simple sets of membership functions. Both knowledge and data need to be used together in developing practical applications. The classification depicted in Figure 6 is mainly based on the ideas of numerical computation with fuzzy rules and membership functions. The increasing complexity of the application requires combined approaches classified to the upper right corner. The knowledge-based approach uses nonlinear sets of membership functions and a fairly simple set of rules. The actual functions can be constructed from linear or nonlinear parts. The linguistic fuzzy models are well suited
to human input based on domain expertise, and the operation of the system can be improved by tuning of the membership functions. First systems operate fairly well but improving the system is time consuming as numerous parameters affect the resulting system. The key issues are suitable modelling areas and understandable rules. Nonlinear behaviour should be included as much as possible in the distribution of the membership functions. The number of rules is tried to keep in minimum. Fuzzy set systems can be used to implement smooth transitions between submodels developed for process phases, special situations or different variables.

**Multimodel** approaches are aimed for changing operating conditions, e.g. clustering can be used for finding suitable areas for modelling. Grid partitioning, which divides each linguistic variable into at least two linguistic terms, is not satisfactory for large-scale systems as too many modelling areas result overfitting. The ANFIS tuning increases the overlap of the TS fuzzy models and destroys the meanings of the individual linear models, e.g. the role of some submodels may transform into a part of a smoothing algorithm. Fitting results are very good with strongly overlapping models, but the process insight is lost. Sharper borders require nonlinear consequent models, i.e. smoothing should be a part of each individual local model. Hybrid modelling approaches try to combine the advantages of the physical and data–driven modelling techniques, e.g. parameters for mechanistic models are approximated by black–box techniques. Since the identification is on a practical level only for linear systems, a lot of work with local linear models is needed. Physical insight is needed, and many nonlinear models can be nonlinearly transformed using linear models, e.g. Model A and Model B before Model C in Figure 5(a)).

**Large-scale models** of multivariable nonlinear systems can be classified to the upper right corner of Figure 6. Membership functions and approximate reasoning are used to combine partially overlapping submodels. Crisp levels and rules are useful for special cases with abrupt changes. Simple fuzzy models classified to the lower right corner of Figure 6 contain only the decomposition level. Parametric steady-state and dynamic models are suitable for the submodels which are understood as consequents. Submodels can be based on phenomenological, statistical, fuzzy and neural models. Variable step numerical integration methods are needed since the models have wide operating areas. Also stiff systems, discontinuities and event handling are important issues in simulation. Fuzzy, neural and neurofuzzy models can be developed equally well for MISO or MIMO systems, also coupled MISO models can be used. Cascade models clarify model structure in these complex systems.
Soft sensors, or intelligent analysers, which use on-line analysers, image analysers and simulation models combined with process measurements, are essential parts of the system. Designed experiments are used to get sufficient variety in operating conditions. Redundancy of measurements is useful in the development of robust solutions. Indirect measurements should be based on real-time data, and therefore, data pre-processing is essential in the modelling phase: data selection, outlier detection, handling of missing values, scaling and removing noise are needed in most applications. (Juuso 2004f) Normalisation is needed in modelling, and symmetric nonlinear scaling can be done for example with log-sigmoid or hyperbolic tangent. Also two-quadrics and logistic functions presented in (Snelick et al. 2005) can be used. The logistic function is symmetrical, but the double sigmoid function based on two log-sigmoid and the two-quadrics function provide asymmetrical nonlinear solutions. All these nonlinear scaling functions are monotonous and increasing, and change from convex to concave at the operating point. The total number of parameters is two for the double sigmoid, three for the two-quadrics and four for the logistic function. Sigmoid functions can be represented as basis functions used the function expansion (Ljung 1999).

Signal processing together with feature extraction and model-based sensor fusion may improve performance considerably. Several signals can be constructed from one signal by independent component analysis or by derivation. Features are often developed with statistical methods, also labels obtained by fuzzification or responses of radial basis functions and neurons of competitive layers can be considered as features. Trends are informative especially in slow processes. High dimensional data can be combined in sensor fusion to produce understandable signals and features. Finding the correct time delays, or number of delay in system identification, is essential for obtaining good modelling results.

The LE approach combines the ideas of the variable meanings originating from fuzzy logic with the linear regression and modelling. The development started from a representation tool of deterministic models and fuzzy set systems (Juuso & Leiviskä 1992) and continued to data-driven modelling and tuning of fuzzy set systems (Juuso 2004e). Combining expertise and data was a key idea in early applications. Dynamic LE models introduced in (Juuso 1999b, 2003a) used the same parametric structures as other intelligent modelling methods. Dynamic models were essential in soft sensors (Juuso 1999b, Ainali et al. 2002). Also a distributed parameter model was developed (Juuso 2004d). Operating conditions, cascade structures and case-based models opened application for forecasting (Juuso 2003b, Mäki et al. 2004, Saarela et al. 2003a).
Integration with other methods of computational intelligence has been important all the time: fuzzy systems provided smooth transitions between the forecasting models; neural networks (Juuso 1999a) and genetic algorithms with penalty terms were used in tuning in (Lotvonen et al. 1997); special situations were modelled with fuzzy set systems in (Juuso et al. 2000). The LE models provide new possibilities for adaptation in modelling and simulation (Juuso 2004e,g, Juuso & Leiviskä 2004). Benefits of the compact LE models become increasingly important when the complexity of the system increases.

During the research phases I - III, the steady-state and dynamic LE models had complex structures, including various cascade and interactive structures presented in Figure 5. The parameters were extracted from data by using arithmetic means and medians and iterated to achieve monotonously increasing scaling functions. The LE approach already included solutions to the equation selection in case-based models, the tuning of the parameters with neural networks and genetic algorithms. The constraints were taken into account with penalty functions. The arithmetic mean is suitable for recursive tuning, but the resulting scaling functions are narrow and sensitive to outliers. The medians provide better solutions but the recursive analysis is more complicated. Since existing model structures and tuning methodologies are used in the LE modelling, the research focuses on the adaptive nonlinear scaling and model adaptation.

### 2.5.2 Nonlinear multivariate control and diagnostics

Various control approaches exist for coping with nonlinearities in changing environment:

- **Nonlinear** control extends the operating area of the control systems, especially intelligent methods provide new tools for this.
- **Adaptation** is primarily devoted to new operating conditions but advances in modelling have improved possibilities for predefined adaptation.
- **Model-based** control has already been used in the very beginning of automatic process control. Model-based predictive control has recently become an increasingly popular research topic, including various modelling approaches.
- **Human** operators can successfully control processes which are very difficult for classical automatic control. Switching and event based control provide channels for introducing these high level actions to the control systems.
- **Multivariable** control should take into account a large number of variables, which introduced a need for software sensors to handle the complexity of the systems.
During the years, these approaches have been used mostly separated, but practical industrial applications require combining these approaches in a hybrid control system. The resulting huge variety of features means that the tuning of the control system must be done with modelling and simulation. Both knowledge and data have been used to make control systems adaptive. The overall structure combines ideas from classical and advanced approaches (Fig. 7), where high-level control provides a platform for combining control and diagnostics.

![Fig 7. Modules of adaptive control, modified from Juuso (2004b).](image)

*Expert control* with rule-based systems may lead to serious testing and maintenance problems in large-scale applications. Representing conventional expert systems with *fuzzy expert systems* needs much less rules since nonlinearities are taken into account in membership functions. Complicated expert control strategies can be included in the practical control by using FLCs, which are suitable for modelling operator’s actions and for taking incomplete, even contradictory, information into account. Technically, a huge number of variables can be combined but at the same time the size of the rule base increases considerably, especially if a complete set of rules is needed (Fig. 4). Also acquiring the required knowledge can be a tedious and time-consuming task. Fuzzy controllers should be kept transparent to interpretation and analysis on the basis of understanding the process behaviour since this insight is highly important.

The results of various applications have shown that the FLCs increase the overall controllability of the processes by using process knowledge together with continuous measurements, different analysers and laboratory analysis. A stable plant operation is achieved by early reactions to disturbances and the smooth control with considerably
smaller steps than operators, and thus the operators can concentrate on higher level tasks or on drastic changes in operating conditions. Intelligent analysers improve performance by providing more informative measurements for the controller. The combined system is similar with the cascade structure shown in (Fig. 5(a)).

On-line adaptation is based on changes detected either in controller performance or in process operation. Adaptation requires time, and classical adaptive schemes do not cope easily with strong and fast changes unless the adaptation rate is made very high. Nonlinear control surfaces can be constructed with FLCs, but there is a need for adaptation to changing operating conditions. Scaling factors of the membership functions can be tuned for example by using heuristic rules or by constructing an analytical relation between the scaling factors and the dynamic response of the closed loop system. On-line modelling with recursive identification, or estimation, methods can be used for obtaining parameters for controllers. The structure and parameters of empirical models do not necessarily have any physical significance, and therefore, these models cannot be directly adapted to different operating conditions. Since the identification is on a sufficient level only for linear systems, a lot of work with local linear models is needed.

Predefined adaptation is based on a priori knowledge about the plant dynamic behaviour. This can be done with gain scheduling or with a multiple model adaptive scheme, which selects a controller from a finite set of predefined fixed controllers. This scheme can be understood as a high level control approach, which is based on the detection of operating conditions. The gradual differences of the operating conditions can be handled with the scaling (Fig. 7).

Inverse models are used in feedforward control to get the process into a good operating conditions. Feedback control is used to compensate the incompleteness of the models. IMC uses the response difference between the model and the process. MPC, which uses simplified models in a prediction horizon to optimise the control actions, suits well for multivariable systems. Also fuzzy and neural models have been used. Combined controllers and predefined adaptation, e.g. with gain scheduling, extend the operating area of the MPC approach. Invertible compact models would be useful in the model-based control.

High-level control acts as a human operator in making decisions, it also supervises adaptation procedures (Fig. 7). This is important since multivariable control has a serious problem in combining the controllers in such a way that individual controllers do not disturb each other. If the weighting of several control strategies is based on
operating conditions, several control loops can operate consistently in changing operating conditions. *Multiobjective controllers*, aimed for drastic changes, can be interpreted as switching controllers: a gradual switching strategy can make the operation smooth. Many FLC applications show that various control strategies can be balanced fairly successfully if the proposed control actions are not too contradictory. For frequently conflicting objectives, priorities should depend on operating conditions.

*Detection of operating conditions* can be based on variables or features, whose low or high values are considered as symptoms in fault diagnosis or in condition monitoring. Non-parametric models, fuzzy rules or self-organising maps can be developed for these purposes by clustering the data. RBF and LVQ networks combine clusters and linear layers. Generalised regression networks have a slightly different linear layer. Fault trees provide a causal-deterministic structure for this analysis. Fuzzy logic is primarily meant for multiple-valued logic and approximate reasoning and fault trees can be represented with more flexible fuzzy rules. An understandable set of structured rules and nonlinear membership functions are essential for practical fuzzy systems. Parametric models, which are developed for different operating conditions, can be used for detection by comparing the performance of the alternative models. In on-line modelling, the model parameters are used for detecting. These model-based approaches can produce indirect measurements or intelligent analysers, including trend analysers.

*Case-based reasoning* uses previous experience stored in the case base, and actually *novelty detection* is a part of the cycle. Models are increasingly used in CBR applications. Software sensors and diagnostic indicators apply for fairly similar purposes in multivariate control. The high level control, detection of operating conditions and fault diagnosis interact strongly with *operator actions*. The continuous operation is guaranteed if there is a real co-operation between the control system and the operator, i.e. the operator can make both manual control actions and small adjustments without switching off the controller, and the controller can continue operation after these interventions. These actions will improve the overall performance if the operator is familiar and agrees with the controller actions. In fault diagnosis, also fault-symptom causalities need to be analysed.

Advanced control systems can be built for integrating various control strategies and different types of models, and modelling and simulation is an essential part of the control design (Juuso 2004f). Invertible models, model-based scaling, several control strategies, multiobjective switching and event-based control are important in nonlinear multivariable control. Cascade structures are needed when combining modelling, control
and diagnostics. The intelligent analysers and model-based detection of operating conditions should be included.

The early applications of the LE approach were in decision support, production scheduling and controller tuning. The first LE controller was implemented in 1996 for a solar power plant (Juuso et al. 1997b, 1998b). Later new applications were developed for a lime kiln (Järvensivu et al. 2001) and a water treatment plant (Joensuu et al. 2004). New advanced modules were introduced to the controller of the solar plant (Juuso & Valenzuela 2003). LE-based fault diagnosis, which started in connection with fuzzy controllers (Juuso 1994a), was later based on case detection in functional testing (Komulainen et al. 1997, Gebus & Juuso 2002) and X-ray inspection (Wei et al. 1999). The solutions expanded to CBR-type applications for paper machines (Ahola et al. 2004) and multisensor fault detection for condition monitoring (Juuso et al. 2004).

During the research phases II-III, the LE control already included PI-type controllers with adaptive scaling and additional tools for handling large changes and removing offset, inverted models for feedforward control, advanced actions for fast changes and combinations of control strategies. Soft sensors simplified the control applications considerably (Juuso 2004g), but the fault diagnosis and detection of operating conditions were based on highly complex and detailed models, which used measurements from processes, testing and inspection as widely used statistical features. The research on control and diagnostics focuses on the soft sensors to make the control and the detection of operating conditions easier to understand and adapt to the process requirements.

### 2.5.3 Smart adaptive systems

Adaptation to changes and similar situations can be done with different on-line and predefined adaptation mechanisms and models. Detection of operating conditions and decomposition of the solution are important in the CBR-type learning, which provides tools for gradual extensions. Adaptation to new applications discussed in (Anguita 2001) is related to data mining, also called knowledge discovery, used for finding relevant information from large databases.

All three levels of SAS can be done on-line, but practical applications need to combine expertise and data (Fig. 8). Expert systems are feasible channels for introducing expertise, and the main benefit of fuzzy logic is that it provides flexible tools for using expertise. Knowledge extraction and case-based reasoning support these solutions. Neural computing started from the data-driven modelling, but both fuzzy set systems
and neural networks are coming closer: data-driven techniques have use in fuzzy set systems, and utilisation of expertise is also an important topic in neural computing. Various neurofuzzy systems are examples of these synergy effects. Top-down and bottom-up approaches are also combined in Bayesian networks and hyperplane methods, e.g. SVM. Signal processing and feature extraction form the basis for the data-driven development. Data mining can be based on multiresolution search and extensive systems can be tuned with evolutionary computing.

Fig 8. Methodologies of smart adaptive systems, modified from Juuso & Leiviskä (2004).

Methodologies have been modified, extended and combined in various ways, also in connection with the feature extraction. This analysis focuses on the main building blocks of the SAS alternatives (Fig. 2). The active research, which is going on in all these methodological areas, would need an extensive presentation. The scenario of the SAS environment integrates the research, system design and implementation in focus areas, including adaptive decision making, integration of human expertise, performance analysis, condition monitoring, and learning, by using model-based approaches, agent-based methodologies, steady-state and dynamic simulation. Information intensive systems based on heterogeneous and inconsistent data are combined with expertise in hybrid multilevel architectures.

The research and development of SAS focuses on three areas (Juuso & Leiviskä 2004):

- Adaptive decision making expands the ideas of adaptive control and intelligent analysers to new application areas, where the human expertise is a natural part. Multiobjective decision making takes into account even contradictory strategies.
Performance analysis and condition monitoring concentrate on detecting changes in operating conditions, including novelty and anomaly detection.

Learning integrates both knowledge based and data-based methodologies ranging from intelligent agents to learning new cases.

Extensions to the environmental monitoring and feasible use of different forms of energy increase importance of all these focus groups.

In the integration of intelligent systems, the normalisation, scaling approaches and membership functions are the key methodologies. The scaling functions used in the LE approach already provided compact nonlinear solutions for modelling and control during the first three research phases. A better understanding of the meanings of the variables, features and symptoms is essential in applications. Adaptation of the parameters of the functions to different distributions and changes with time are focused in the SAS framework, which also includes possibilities to adapt to new operating conditions. The LE approach can be considered as a generalised hyperplane method (Fig. 8) since models are modified with the scaling functions. Connections to other methodologies are important in large-scale complex interconnected distributed systems and in handling uncertainties in smart adaptive applications (Fig. 2). Comparison of the LE approach with all these approaches is beyond the topic of this dissertation. The discussion above includes only the main characteristics of the approaches: more methodologies are available and new solutions are introduced.

The research on the LE approach in developing smart adaptive applications focuses on the adaptation of the nonlinear scaling and the efficient parametrisation of the overall system. Linear methodologies are suitable for modelling by using the scaled values: the residuals are interpreted as fuzziness. The modelling methodologies extend the soft sensors, which are increasingly important in the control and diagnostics applications. Signal processing and feature extraction are important in practical applications and close links are maintained with advanced statistical methods, which are essential in this analysis, and fuzzy set systems, which are used in uncertainty handling.
3 Adaptive nonlinear scaling

Linguistic equation (LE) approach originates from fuzzy set systems: rule sets are replaced with equations (Juuso & Leiviskä 1992), and meanings of the variables are handled with scaling functions which have close connections to membership functions (Juuso 1999a). The nonlinear scaling technique is needed in constructing nonlinear models with linear equations (Juuso 2004e). New development methodologies (Juuso 2009e, Juuso & Lahdelma 2010) improve possibilities to update the scaling functions recursively (Juuso 2011b, Juuso & Lahdelma 2011a).

3.1 Nonlinear scaling

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 (Juuso et al. 1993). The basic scaling approach presented in (Juuso 2004e) has been improved later: a new constraint handling was introduced in (Juuso 2009e), and a new skewness based methodology was presented for signal processing in (Juuso & Lahdelma 2010).

3.1.1 Working point and feasible ranges

The concept of feasible range is defined as a trapezoidal membership function. In the fuzzy set theory (Zimmermann 1992), support and core areas are defined by variable, $x_j$, specific subsets,

$$supp(F_j) = \{ x_j \in U_j \mid \mu_{F_j}(x_j) > 0 \}, \quad (4)$$

$$core(F_j) = \{ x_j \in U_j \mid \mu_{F_j}(x_j) = 1 \}, \quad (5)$$

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. For applications, a trapezoidal function providing linear transitions between 0 and 1 is sufficient (Fig. 9). 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 (Zimmermann 1992)

\[ \bar{F}_j = \{ x_j \in U_j \mid \mu_{\bar{F}_j}(x_j) = 1 - \mu_{F_j}(x_j) \}, \quad (6) \]

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 (Fig. 9).

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 (Juuso 2004e). The norm defined by

\[ ||\tau^p_M j||_p = (\tau^p_M j)^{1/p} = \left( \frac{1}{N} \sum_{i=1}^{N} (x_j^p)^{1/p} \right), \]  

(7)

where \( p \neq 0 \), is calculated from \( N \) values of a sample, \( \tau \) is the sample time. With a real-valued order \( p \in \mathbb{R} \) this norm can be used as a central tendency value if \( ||\tau^p_M j||_p \in \mathbb{R} \), i.e. \( x_j > 0 \) when \( p < 0 \), and \( x_j \geq 0 \) when \( p > 0 \). The norm (7) 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 (7) is a Hölder mean, also known as the power mean. The generalised norm for absolute values \(|x_j|\) was introduced for signal analysis in (Lahdelma & Juuso 2008a).

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:

\[ ||\tau^p_M j||_p = \frac{1}{2} \left\{ \left( \frac{1}{N} \sum_{i=1}^{N} [(x_j^p)_i - x_L]^p \right)^{1/p} + x_L \right\} - \frac{1}{2} \left\{ \left( \frac{1}{N} \sum_{i=1}^{N} [(x_j^p)_i - x_H]^p \right)^{1/p} + x_H \right\}, \]  

(8)

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 2011b)

The norm values increase with increasing order, i.e.

\[ (\tau^p_M j)^{1/p} \leq (\tau^q_M j)^{1/q}, \]  

(9)

if \( p < q \). The increase is monotonous if all the signals are not equal. The arithmetic mean, the harmonic mean and the root mean square (rms) are special cases where the order \( p \) is 1, -1 and 2, respectively. The norm (7) represents the norms from the minimum to the maximum corresponding the orders \(-\infty \leq p < \infty\), i.e. the definition includes the \( l_p \) norms defined for \( 1 \leq p < \infty \). The geometric mean is obtain from (7) when the order \( p \to 0 \).

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:

\[ ||K_s\tau^p_M j||_p = \left( \frac{1}{K_s} \sum_{i=1}^{K_s} [(\tau^p_M j)^{1/p}]^{1/p} \right)^{1/p} = \left( \frac{1}{K_s} \sum_{i=1}^{K_s} [(\tau^p_M j_i)^{1/p}] \right)^{1/p}, \]  

(10)

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 (\( p = 1 \).
Distributions of the data can be analysed with dimensionless features obtained by normalising the moments $M^k_j$, for example by standard deviation $\sigma_j$:

$$\gamma_k = \frac{\tau M^k_j}{\sigma_j} = \frac{1}{N \sigma_j} \sum_{i=1}^{N} [(x_j)_i - c_j]^k,$$

(11)

where the moment $M^k_j$ is obtained about some central value, usually arithmetic mean. Variance $\sigma^2_j$ is the second moment $M^2_j$. The feature $\gamma_3$ is called the coefficient of skewness, or briefly skewness, and the feature $\gamma_4$ as the coefficient of kurtosis. The skewness is a measure of asymmetry: $\gamma_3 = 0$ for a symmetric distribution. If $\gamma_3 > 0$, the skewness is called positive skewness and the distribution has a long tail to the right, and vice versa if $\gamma_3 < 0$. The kurtosis is a measure of the concentration of the distribution near its mean. The generalised moment for absolute values $|x_j|$ was introduced for signal analysis in (Lahdelma & Juuso 2008b).

The normalised moments (11) are generalised by using the generalised norm (7) as the central value:

$$\gamma^p_k = \frac{1}{N \sigma_j} \sum_{i=1}^{N} [|x_j)_i - \| \tau M^p_j \|_p]^k$$

(12)

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

The generalised skewness $\gamma^p_3$ is used when choosing appropriate methods for defining the central tendency. Juuso & Lahdelma (2010) introduced a new approach based on (12) for estimating the central tendency value and the core area. The central tendency value is chosen by the point where the skewness changes from positive to negative, i.e. $\gamma^p_3 = 0$. 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 of the corner points, $(c_l)_j$ and $(c_u)_j$, are the points where $\gamma^p_3 = 0$ for the lower and upper data sets, respectively. 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 $(p_u)_j$ are good estimates when additional data sets are used. The norm values can be recursively updated with (10), and a new search for the orders is done only if the values change considerably (Juuso 2011b).

In practical applications, 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.
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 (7) 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.

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 (6).

### 3.1.2 Membership definitions and functions

A membership definition is defined as a (nonlinear) mapping of variable values inside its range to a range \([-2, 2]\), denoted as *linguistic range*. It more or less describes the distribution of variable values over its range. Membership definitions are presented by a function

\[
x_j = f_j(X_j) \forall \min(x_j) \leq x_j \leq \max(x_j), \ X_j \in [-2, 2],
\]

where \( x_j \) is the value of variable \( j \) and \( X_j \) is the corresponding value in the range \([-2, 2]\), which 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 can be associated to the linguistic labels, e.g.

\[
\{ \text{very low}, \text{low}, \text{normal}, \text{high}, \text{very high} \}
\]

are defined with membership functions (Fig. 9). 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 (Juuso 1999a).

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, $X \in [0, 2]$. So

$$
\begin{align*}
  x_j &= f_j^- (X_j), X_j \in [-2, 0), \\
  x_j &= f_j^+ (X_j), X_j \in [0, 2].
\end{align*}
$$

Two facts must be pointed out concerning $f_j^- (X_j)$ and $f_j^+ (X_j)$. They should be monotonous, increasing functions in order to result in realisable systems, i.e.

$$
\begin{align*}
  f_j^- ((X_j)_1) > f_j^- ((X_j)_2) \forall (X_j)_1 > (X_j)_2, \\
  f_j^+ ((X_j)_1) > f_j^+ ((X_j)_2) \forall (X_j)_1 > (X_j)_2.
\end{align*}
$$

The lower part function is defined by values corresponding linguistic levels -2, -1 and 0, and the upper part function by values corresponding linguistic levels 0, 1 and 2. The upper and lower parts should overlap at the linguistic value 0. (Juuso 2004e)

Five points,

$$\{\min(x_j), -2), ((c_j)_j, -1), (c_j, 0), ((c_h)_j, 1), (\max(x_j), 2}\},$$

define the coefficients of the second order polynomials,

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

The centre point, $c_j$, defines the operating point. Four linear equations are needed for solving the other coefficients:

$$
\begin{align*}
  4a_j^- - 2b_j^- + c_j &= \min(x_j), \\
  a_j^- - b_j^- + c_j &= (c_j)_j, \\
  a_j^+ + b_j^+ + c_j &= (c_h)_j, \\
  4a_j^+ + 2b_j^+ + c_j &= \max(x_j).
\end{align*}
$$

In order to keep the functions monotonous and increasing, the derivatives of functions $f_j^-$ and $f_j^+$ should always be positive (Fig. 10). As a second order polynomial has either a minimum or a maximum point, this requirement is fulfilled only if these points are outside the ranges $(-2, 0)$ and $(0, 2)$ for functions $f_j^-$ and $f_j^+$, respectively. The derivatives,

$$
\begin{align*}
  D_j^- &= 2a_j^- X_j + b_j^- , \quad X_j \in [-2, 0), \\
  D_j^+ &= 2a_j^+ X_j + b_j^+ , \quad X_j \in [0, 2],
\end{align*}
$$

are corrected to positive in the areas $(-2, 0)$ and $(0, 2)$, respectively, by changing the coefficients of the polynomials (Juuso 2004e). The membership definitions are continuous functions but derivatives can have discontinuities in the centre point.
The functions are monotonous and increasing if the ratios,
\[
\alpha_j^- = \frac{(c_l)_j - \min(x_j)}{(c_l)_j - (c_l)_j},
\alpha_j^+ = \frac{\max(x_j) - (c_h)_j}{(c_h)_j - c_j}, \quad (21)
\]
are both in the range \([\frac{1}{3}, 3]\), see (Juuso 2009e). 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 \(f_j^-\) and \(f_j^+\): each error can always be corrected either by moving the corner of the core or the support (Table 1). In some cases, good results can also be obtained by moving \(c_j\) in the range defined by
\[
(c_l)_j + \frac{1}{4}((c_l)_j - \min(x_j)) \leq c_j \leq (c_l)_j + 3((c_l)_j - \min(x_j)),
(c_h)_j - 3(\max(x_j) - c_j) \leq c_j \leq (c_h)_j - \frac{1}{4}(\max(x_j) - (c_h)_j). \quad (22)
\]
If these constraints allow a non-empty range, the maximum of the lower limits and the minimum of the upper limits are chosen to define the limits for continuous definitions (Fig. 11).
The coefficients of the polynomials can be represented by
\[
\begin{align*}
    a_j^- &= \frac{1}{2}(1 - \alpha_j^-) \Delta c_j^- , \\
    b_j^- &= \frac{1}{2}(3 - \alpha_j^-) \Delta c_j^- , \\
    a_j^+ &= \frac{1}{2}(\alpha_j^+ - 1) \Delta c_j^+ , \\
    b_j^+ &= \frac{1}{2}(3 - \alpha_j^+) \Delta c_j^+ ,
\end{align*}
\]
(23)
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 \(\alpha_j^-\) or \(\alpha_j^+\) equals to one (Fig. 10).

**Table 1.** Parameter corrections by changing the core and/or the support.

<table>
<thead>
<tr>
<th>Error</th>
<th>Core</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f_j^-)</td>
<td>(\alpha_j^- &gt; 3)</td>
<td>((c_l)_j = c_j - \frac{1}{2}(c_j - \min(x_j))) (\min(x_j) = c_j - 4(c_j - (c_l)_j))</td>
</tr>
<tr>
<td>(f_j^-)</td>
<td>(\alpha_j^- &lt; \frac{1}{2})</td>
<td>((c_l)_j = c_j - \frac{1}{2}(c_j - \min(x_j))) (\min(x_j) = c_j - \frac{4}{3}(c_j - (c_l)_j))</td>
</tr>
<tr>
<td>(f_j^+)</td>
<td>(\alpha_j^+ &gt; 3)</td>
<td>((c_h)_j = c_j + \frac{1}{4}(\max(x_j) - c_j)) (\max(x_j) = c_j + 4((c_h)_j - c_j))</td>
</tr>
<tr>
<td>(f_j^-)</td>
<td>(\alpha_j^+ &lt; \frac{1}{2})</td>
<td>((c_h)_j = c_j + \frac{1}{2}(\max(x_j) - c_j)) (\max(x_j) = c_j + \frac{4}{3}((c_h)_j - c_j))</td>
</tr>
</tbody>
</table>

The centre point is not known if the feasible range is defined manually. It can be calculated by defuzzifying the feasible range with the centre of gravity:
\[
c_j = \frac{1}{4} ((c_l)_j + (c_h)_j + \min(x_j) + \max(x_j)).
\]
(24)
For strongly asymmetrical feasible ranges, this value may be outside the core (Juuso 2004e). The requirement (24) can be fulfilled by modifying the corner points.

Additional constraints can be taken into account, e.g. a good solution can be to use a locally linear function in the neighbourhood of the centre point. Then a continuous derivative is chosen at the centre point: \( b_j^- = b_j^+ \), which can be represented by

\[
6 c_j - 4 (c_l)_j - 4 (c_h)_j + \min(x_j) + \max(x_j) = 0.
\]

This can be achieved by modifying the centre point or the corner points of the feasible range. There can be several acceptable modifications, for which the ratios (21) remain in the range \( [\frac{1}{3}, 3] \).

Monotonously increasing membership definitions can be constructed by adjusting the centre point \( c_j \), the core \( [(c_l)_j, (c_h)_j] \) and the support \( [\min(x_j), \max(x_j)] \). An easier way for manual approach was introduced in (Juuso 2009e): first define the centre point \( c_j \), then the core by choosing the ratios (21) from the range \( [\frac{1}{3}, 3] \), and finally calculate the support \( [\min(x_j), \max(x_j)] \). The norms (7) and (8) are used together with the generalised skewness (12) in the data-driven approach to define the centre and corner points. The ratios (21), which are checked in all data-driven cases, are also guiding the manual construction of the membership definitions. Additional constraints like (24) and (25) are used if they are feasible.

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 \( \{\min(x_j), (c_l)_j, (c_h)_j, \max(x_j)\} \) are good for visualisation;
- the parameters \( \{\alpha^-_j, \Delta c^-_j, \alpha^+_j, \Delta c^+_j\} \) suit for tuning;
- the coefficients \( \{a^-_j, b^-_j, a^+_j, b^+_j\} \) are used in the calculations.

The upper and lower parts of the scaling functions can be convex or concave independently. Also simplified functions can be used: a linear membership definition needs only two parameters: \( c_j \) and \( b_j = b_j^- = b_j^+ \) or \( \Delta c_j = \Delta c^-_j = \Delta c^+_j \), since \( \alpha^-_j = \alpha^+_j = 1 \) and \( a^+_j = a^-_j = 0 \); an asymmetrical linear definition has \( \Delta c^+_j \neq \Delta c^-_j \) and \( b^+_j \neq b^-_j \). Local linear functions defined by (25) are used if appropriate.

Feasible ranges can be used in high level fuzzy set systems. More detailed membership functions can be generated from membership definitions on a chosen partition. A strong partition

\[
\forall x_j \in U_j, \sum_{i=1}^{n} \mu_{F_i}(x_j) = 1,
\]

83
where $\mu_{F_i}$ are membership functions of fuzzy sets $F_i, n = 1, \ldots, n$, suits well for automatic generation. Figure 9 shows the default locations corresponding to values -2, -1, 0, 1 and 2 in the linguistic range. The overlap between adjacent linguistic terms expresses the smooth transition from one term to the other. Different shapes of membership definitions result different sets of default membership functions: the locations depend on the core, the support and the centre point (Fig. 12). The whole range of the ratios is covered in these examples: the ratios are 3 for 'narrow core' and 'wide support', 1/3 for 'wide core' and 'narrow support', and 1 for 'asymmetric linear functions'. The ratios $\alpha^-_j$ and $\alpha^+_j$ do not need to be equal. The number of membership functions can be variable specific (Juuso 2004e).

\begin{align}
\text{Narrow core} &= [-1.20, 0.08, 0.50, 0.70, 1.30] \\
\text{Wide core} &= [-1.20, -0.77, 0.50, 1.10, 1.30] \\
\text{Asymmetric linear functions} &= [-1.20, -0.35, 0.50, 0.90, 1.30] \\
\text{Wide support} &= [-2.90, -0.35, 0.50, 0.90, 2.10] \\
\text{Narrow support} &= [-0.63, -0.35, 0.50, 0.90, 1.03] \\
\text{Defuzzified feasible range} &= [-1.20, -0.35, 0.28, 0.90, 1.30] \\
\text{Continuous derivative} &= [-1.20, -0.35, 0.35, 0.90, 1.30]
\end{align}

Discretised membership functions defined in the form of two related vectors,

\begin{align*}
F_D &= [(x_j)_1, (x_j)_2, \ldots, (x_j)_{n_D}]^T, \\
\mu_D &= [\mu((x_j)_1), \mu((x_j)_2), \ldots, \mu((x_j)_{n_D})]^T,
\end{align*}

(27)

provide a channel for using deterministic models in uncertain environment (Juuso 1989). Each value, $(x_j)_l, l = 1, 2, \ldots, n_D$, has its own degree of membership, $\mu((x_j)_K), K = 1, 2, \ldots, n_D$, generated on the basis of the membership function of variable $j$. In addition
to the deterministic level, where the degree of membership equals one, discretized levels,

\[ \{\mu(x_j)\}_{n_F} = \left\{ \frac{K}{n_F + 1} \right\}_{K = 1,2, \ldots, n_F}, \quad (28) \]

are taken into account in the fuzzy calculation level \( n_F \). The total number of discrete points \( n_D = 2n_F + n_C \) where \( n_C \) is the number of the points selected from the most possible area. The membership function can be for example triangular, trapezoidal or rectangular (Juuso 1989). Triangular functions are used for modelling and control, and trapezoidal functions for fault diagnosis. A piecewise linear membership function is defined by the vectors (27) when sorted in ascending order of the variable values \((x_j)_K\).

Nonlinear membership functions are developed by generating linear membership functions within the linguistic range, since the nonlinear scaling transforms the linear parts into a nonlinear form in the original real range. The scaling functions are not necessarily available for knowledge-based information, but this approach can be used if the labels are ordered, e.g. from very low to very high. The membership functions defined in the linguistic range are useful in analysing interactions.

Discretised membership functions can be obtained by choosing the locations and their degree of membership in the linguistic range. The discrete values in real range are defined from the locations by the membership definitions. Nonlinear effects are even better captured by selecting the points \((x_j)_K\), \( K = 1,2, \ldots, n_D \), by using scaled values \((X_j)_K\), and the corresponding degrees of membership \( \mu((X_j)_K) \). The fuzzy set has \( n_F + 1 \) \( \alpha \)-cuts, which are asymmetric to the centre point \( c_J \). For the scaled values the corresponding \( \alpha \)-cuts \([((X_j)_{2L-1}, (X_j)_{2L}]\), where \( L = 1,2, \ldots, n_F + 1 \), are symmetric to \( c_J \) in the linguistic range in all levels \( n_F + 1 \). This approach can be also gradually refining, each finer partition introduces additional discrete points between the points in the previous partition. Taking the new points from the centre of the previous points in the linguistic range, \( n_F = 1,3, \ldots \), suits well for gradually refining optimisation used in (Juuso et al. 1993).

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 by (10) as the norm for the norms of individual samples. The membership functions defined in the linguistic range will be updated through changing membership definitions, when the points (17), the parameters \( \{\alpha^{-}_j, \Delta c^{-}_j, \alpha^{+}_j, \Delta c^{+}_j\} \), and the coefficients (23) change.

In a type-2 fuzzy feasible range, all the corner points are defined by an interval, and the centre point is within the core. The constraint handling (Juuso 2009e) and the skewness based selection of corner points (Juuso & Lahdelma 2010) provide a basis for
generating type-2 fuzzy membership functions. Feasible ranges and all the membership functions are generated by using fuzzy scaling functions, i.e. the parameters are fuzzy numbers. To ensure consistency with monotonous and increasing scaling functions, fuzziness is introduced to the centre point $c_j$ and the parameters $\{\alpha_j^-, \Delta c_j^-, \alpha_j^+, \Delta c_j^+\}$. The ratios (21) are fuzzy but limited to the range $[\frac{1}{3}, 3]$.

The alternative ways explained above are used to correct the parameters if these constraints are violated. Since the centre point $c_j$ can also be a fuzzy number, the functions $f_j^-$ and $f_j^+$ do not need to have the same constant term. It is sufficient to have both $c_j^-$ and $c_j^+$ in the core. If corrections are needed, changing the core or the support provide good solutions: $c_j^-$ and $c_j^+$ are used in Table 1 for $f_j^-$ and $f_j^+$, respectively.

Resampling procedures (page 49) can be used in data-based analysis: each sample results its own set of parameters as singletons. The fuzzy numbers for the parameters are obtained from these singletons.

### 3.1.3 Statistical distributions

In data-based analysis, the nonlinear scaling functions are based on data samples. The parameters obtained by statistical analysis depend strongly on the statistical distribution. The functions provide extensions to the normalisation and scaling methods discussed in 2.2.1. The z-score (1) is a symmetric special case, where $c_j = |^1\| M_j^1 |$ and $\Delta c_j = \sigma_j = |^2\| M_j^2 |$, i.e. generalised norms (7) with orders $p = 1$ and $p = 2$, respectively. Other special cases, geometric mean ($p = 0$) and harmonic mean ($p = -1$), are used in defining the centre of the sample for lognormal or heavily skewed data. Trimmed or truncated means, medians and median absolute deviations are generally recommended for the cases with outliers. The generalised norms can also be trimmed by discarding values at the high and low end. For heavily skewed data, the discarding limits are defined by the norms with high positive and negative orders, respectively.

In the skewness based approach presented above, all the parameters are analysed from the data. As expected, the $c_j$ is close to the arithmetic mean ($p = 1$) when the sample is taken from a normal distribution. Normalisation with the z-score is the first phase since the core is symmetrical, i.e. $\Delta c_j^+ = \Delta c_j^- = \frac{1}{2} |^2\| M_j^2 |$. The resulting shape factors are equal, $\alpha_j^- = \alpha_j^+ = 3$, and the support is $[c_j - 2\sigma_j, c_j + 2\sigma_j]$. The size of the random sample effects on the analysis: the centre point is correctly obtained from a small sample ($N = 10000$), and also the core is fairly accurate. The limits of the support area and the shape factors require larger samples, e.g. 10000 points provides fairly good
estimate, but 50000 points are required for highly accurate estimates. Only a slight adjustment of the core or preferably the support is needed for these samples.

The scaling functions become asymmetrical about the centre $c_j$ in random samples of Poisson and Weibull distributions (Fig. 13). For the Poisson distribution, the order $(p_0)_j$ is almost constant, $1.68 \pm 0.03$ when the expectation number $\lambda_p \geq 2$, and $(p_0)_j = 1.73$ when $\lambda_p = 1$ (Fig. 13(a)). For the Weibull distribution, the order $(p_0)_j$ decreases smoothly from 2.8 to $-1.85$ when the shape parameter increases from one to ten (Fig. 13(b)). The order range $[(c_l)_j, (c_h)_j]$ increases for both: from $[1, 4.34]$ to $[-1.15, 6.05]$ for Poisson and from $[2.2, 3.2]$ to $[-4.75, 6.15]$ for Weibull distributions whose scale parameter $\lambda_W = 3$.

Poisson distributions have the same shape factor $\alpha_j^+$ as the normal distributions, but the shape factor $\alpha_j^-$ increases from 0.6 to almost 3 when the expectation number $\lambda_p$ increases from one to ten (Fig. 13(c)). The core also is asymmetrical: $\Delta c_j^+ > \Delta c_j^-$. The difference is high, when $\lambda_p$ is small, and becomes negligible, when $\lambda_p > 10$ (Fig. 13(c)). Weibull distributions are very asymmetrical when the shape parameter $\kappa$ is small: $\Delta c_j^+ \gg \Delta c_j^-$, $\alpha_j^+ \approx \frac{1}{4}$ and $\alpha_j^- = 3$, when $\kappa = 0.5$ (Fig. 13(d)). This exponential distribution becomes more symmetrical when $\kappa$ increases, but becomes again asymmetrical for higher $\kappa$ values (Fig. 13(d)). The Poisson distributions have only integer values, which causes irregular changes in orders $(c_l)_j$ and $(c_h)_j$ obtained from random samples.

For all these distributions, the core area becomes wider than in the previous approaches where the mean or the median were used. Higher sensitivity around the centre point was already detected in (Juuso & Lahdelma 2010). High positive and negative orders are used in selecting the limits for the core area if small deviations are not important. Asymmetrical scaling functions can be obtained by analysing the upper and the lower part separately.

The scaling functions consisting of two second order polynomials operate well for versatile distributions, and various sigmoid functions can be interpreted as special cases. The centre points, which define the operating point of the model, can be defined manually. For the error, the derivative of error, the sum of error, the original error and the change of control, the centre point is zero. Also the core and support areas can be defined manually for any membership definition. Monotonous increase needs to be checked for the manually defined functions.

The shape factors define the type of the feasible range (Fig. 12): narrow and wide cores correspond to high and low shape factors, respectively. The factors also allow an
asymmetric core, i.e. the core can be narrow on one side of the centre point \( c_j \) and wide on the other side. The support can depend strongly on the number of points as seen in the comparisons of different statistical distributions. The shape factors \( \alpha_j^- \) and \( \alpha_j^+ \) can be chosen by using expert knowledge and physical limitations. The factors can be set to three if the data set is fairly limited and there is no specific additional knowledge. Linear scaling functions, i.e. \( \alpha_j^- = \alpha_j^+ = 1 \) are used if the material is very limited.

### 3.2 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 the statistical process control provides additional tools for detecting changes, anomalies and novelties.

### 3.2.1 Data selection

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 (Fig. 14). For LE models, the scaled values are in the range $[-2, 2]$, and this is also the range for the monotonous increase if $\alpha_j = \alpha_j^- = \frac{1}{3}$. The minimum and maximum points are obtained from the derivatives (20): the maximum point in the linguistic range

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

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

The outlier border $(X_j)_{\text{max}}$ is limited to 3 in the range $\alpha_j^+ \in \left[\frac{1}{3}, 1\right]$. For the normal distribution, $p_j = 1 \approx \sigma_j^2$. Since $\alpha_j^+ = 3$, the common outlier limit $3\sigma_j$ corresponds to $X_j = \sqrt{6} \approx 2.4495$. A limit between the points $(1, 3)$ and $(3, 2.5)$ is approximated by $(X_j)_{\text{max}} = 3 - \frac{1}{4}(\alpha_j^+ - 1)$, which results a limit of $6.25 = 3.125\sigma_j$ when $\alpha_j^+ = 3$. The scaled limits shown in Figure 14(a) are used in calculating the relative limits (Fig. 14(b)) with (18). The corresponding lower limits, $(X_j)_{\text{min}}$, are obtained by using $a_j^-$, $b_j^-$ and $\alpha_j^-$. The support area includes the core area, and the area between the support and outlier area is considered as suspicious area, where the degree of membership of outlier increases towards the outlier area.

For Poisson distributions, the upper outlier limit $(x_j)_{\text{max}}$ obtained by using $\alpha_j^+ = 3$ (Fig. 14(b)) is increasing considerably with increasing expectation value, (Fig. 15(a)). The limit $(x_j)_{\text{max}}$ has very high values for the Weibull distributions, which are close to the exponential distribution (Fig. 15(b)). The actual lower outlier limit is zero for these types of distributions. The calculated value is equal to the $x_{ll}$ when $\alpha_j^- = \frac{1}{4}$ and is negative in a wide area. The calculated limit becomes active when the shape parameter is greater than eight.

Even steeper distributions can be represented by logarithmic functions as a first step (Fig. 16), e.g. the shape parameters of Weibull distributions can be less than 0.465, which is the limit in Figure 15(b). In this case $c_j \approx 0.2 \ll 21.4$, and the core is in much

89
lower level [0, 2.71] compared to [5.4, 68.7]. The upper border of the support area is slightly higher: $(c_h)_j = 377 > 317$, and the outlier limit is very high $9 \cdot 10^4$. (Fig. 16(a)). In these examples, the shape factors are almost constant: $\alpha^+_j = 3$ and $\alpha^-_j = 1.85 \pm 0.20$.

For logarithmic function $\log_{10}$, $c_j \approx 0.19$, and the core is on similar level $[0, 2.57]$. The upper border of the support area is lower: $(c_h)_j = 145$, and the outlier limit is very high $1.5 \cdot 10^4$ (Fig. 16(b)). The shape factor $\alpha^+_j \in [1.5, 1.8]$; $\alpha^-_j = 3$, when $\lambda_W \leq 0.6$, and starts to decrease towards two. For both logarithmic functions, the lower limits go very close zero: the lower core border become less than $e^{-5.5} = 10^{-2.3} = 0.005$ when $\lambda_W \approx 0.46$, the centre point $c_j \approx 0.005$ when $\lambda_W \approx 0.18$. The lower outlier limit is very close to zero: the maximum is 0.012 for the cases shown in Figure 16. The norms are calculated by (8), since the logarithmic values include both negative and positive values.

Since the generalised norms cover the whole range from minimum to maximum, the norms with a high negative and a high positive order are promising for choosing the lower and upper limits, respectively. The orders are chosen interactively to get appropriate levels for different applications, e.g. orders -30 and 30 have worked for this purpose. In this method, the orders need to be adjusted to get proper shape factors since often the limits $x^{ll}_j$ and $x^{hl}_j$ are quite uncertain. Assumptions on the shape can be used together with the centre point and the limits of the core to define the support and the outlier limits. The suspicious range is used for defining a degree of membership for the reliability of the data value, e.g. a linear decrease from $\mu = 1$ on the support border to $\mu = 0$ on the outlier limit.
3.2.2 Recursive adaptation

The parameters of the nonlinear scaling functions can be recursively updated with (10) 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 \( \alpha_j^- \) and \( \alpha_j^+ \) 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.
3.2.3 **Decomposition**

Domain expertise is essential in choosing variables, which are handled with overall membership definitions, and variables, which require condition specific functions. Decomposition is needed for *case specific* definitions. In the recursive analysis, suspicious values are collected for the further analysis to find new operating conditions. Different data samples may result very different parameters for the scaling functions, i.e. the feasible range becomes highly fuzzy type-2 fuzzy number. This means that the samples come from different operating conditions, and corresponding membership definitions are either case specific or the variables should be used as working point variables.

Various clustering methods can be used for introducing several operation areas. Fuzzy clustering and radial basis networks provide slightly overlapping clusters, where the cluster centres (Babuška 1998) and weight vectors (Chen *et al.* 1991) suit for estimating the centre points $\tilde{c}_l$ of the clusters $l = 1, \ldots, n_{clus}$. The overlapping areas and the shape of the cluster provide information about the feasible ranges of cases defined by the clusters. An appropriate number of the clusters $n_{clus}$ is obtained interactively by iterating the parameters of the clustering algorithm. Hierarchical and partitioning based clustering can be used for finding the cluster centers. Robust clustering based on the spatial median provides good estimates. Different shapes require modifications to the clustering algorithms, see (Gustafson & Kessel 1979). Nonlinear scaling expands the application area of the symmetrical clustering methods since the resulting clusters can be interpreted as the clusters of different shapes in the original data sets.

3.2.4 **Statistical process control**

*Statistical process control (SPC)* provides algorithms for detecting deviations from the defined operating areas of individual variables. The SixSigma approach is based on normal distributions, and therefore, it can be understood as a symmetric special case, where the shape ratios $\alpha_j^- = \alpha_j^+ = 3$. The upper and lower control limits, $[c_j + 3\sigma_j$ and $c_j - 3\sigma_j]$, correspond to the scaled values $X_j = \sqrt{6}$ and $X_j = -\sqrt{6}$, respectively. The upper and lower warning limits are $c_j + 2\sigma_j$ and $c_j - 2\sigma_j$, i.e. the scaled values $X_j = 2$ and $X_j = -2$.

The nonlinear scaling with the second order polynomials to $[-2, 2]$ provides a basis for variable specific control charts, which take into account the distributions. First the
center line \(c_j\) is defined with an appropriate generalised norm. Then the shape ratios \(\alpha_j^+\) and \(\alpha_j^-\) are used to define the upper and lower control limits by using the limits \((X_j)_{\text{max}}\) and \((X_j)_{\text{min}}\), and the upper and lower warning limits by using the limits of the support area. The limits can be highly asymmetrical, when \(\Delta c_j^+ \neq \Delta c_j^-\) and/or \(\alpha_j^+ \neq \alpha_j^-\), e.g. in Poisson distributions \(\alpha_j^+ = 3\) and \(\alpha_j^- = 1\) with strongly asymmetrical core areas. The analysis with the scaled values \(X_j\) is beneficial in short term SPC, e.g. specific scaling functions and limits can be developed for the operating conditions seen in Figure 14.
4 Linguistic equation modelling

Nonlinear models can be constructed by using scaled values in linear modelling based on data and expertise (Juuso 1999a, 2004e). 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.

4.1 LE models

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.

4.1.1 Interactions

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

\[
\sum_{j=1}^{m} A_{ij} X_j(t - n_j) + B_i = 0,
\]

where \( X_j \) is a linguistic value for the variable \( j, j = 1...m \). 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 (Section 3.1.2). The directions of the interaction are represented by interaction coefficients \( A_{ij} \in \mathbb{R} \). In the original system (Juuso & Leiviskä 1992), 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 (Juuso et al. 1993). The bias term \( B_i \in \mathbb{R} \) was first introduced as an additional component in fuzzy LE models (Juuso 1996), and later extended for fault diagnosis systems (Juuso 2004e).
The coefficients $A_{ij}$ and $B_i$ in (30) have a relative meaning, i.e. the equation can be multiplied or divided by any non-zero real number without changing the model. A LE model with several equations can be represented as a matrix equation

$$AX + B = 0,$$

(31)

where the interaction matrix $A$ contains all coefficients $A_{ij}, i = 1, \ldots, n, j = 1, \ldots, m$, and the bias vector $B$ all bias terms $B_i, i = 1, \ldots, n$. 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.

### 4.1.2 Steady-state LE models

In the case of polynomial membership definitions, the linguistic level of the input variable $j$ is calculated according to equation

$$X_j = \begin{cases} 
2 & \text{with } x_j \geq \max x_j \\
\frac{-b_j^+ \sqrt{b_j^+ + 4a_j^+ (c_j - x_j)}}{2a_j^+} & \text{with } c_j \leq x_j \leq \max (x_j) \\
\frac{-b_j^- \sqrt{b_j^- + 4a_j^- (c_j - x_j)}}{2a_j^-} & \text{with } \min (x_j) \leq x_j \leq c_j \\
-2 & \text{with } x_j \leq \min (x_j).
\end{cases}$$

(32)

where $a_j^+, b_j^+, a_j^-$, and $b_j^-$ are coefficients of the polynomials (18), $c_j$ is real value corresponding to the linguistic value 0 and $x_j$ is the real value. Limits $\min (x_j)$ and $\max (x_j)$ are the minimum and maximum values of the real data corresponding to values -2 and 2 in the linguistic range. The membership definitions contain linear parts, if at least one of the coefficients $a_j^+$ and $a_j^-$ is zero. These parts are handled with linear functions instead of (32).

This scaling is done for all input variables with the linguistification block (Fig. 17(b)). Coefficients of the polynomials are selected from the parameter structures by the variable index $j$ and the type of the definition which is 1 for the membership definition of the variable value. Type 2 is needed for variable time delays if they are used in the model (Section 4.1.3). Other types of membership definitions, types 3\ldots8, are for feedback LE controllers summarised in Section 5.6.

The linguistic value of the output variable is obtained by

$$X_{out} = -\frac{1}{A_{out}} \left( \sum_{j=1, j \neq out}^m A_{ij}X_j(t - n_j) + B_i \right).$$

(33)
In the equation block the weighted sum of the linguistic values of the inputs is multiplied by $-1/A_{i\text{out}}$ where $A_{i\text{out}}$ is the interaction coefficient of the output variable $\text{out}$ (Fig. 17(c)). A linguistic equation (30) can be used to any direction, i.e. one variable can be solved from each equation if other variables with non-zero coefficients are known. Advanced matrix based methods can be used for matrix equations (31).

After the linguistic level of the model output, $X_{\text{out}}$, is calculated with linguistic equation model, it is converted to the real value of the output, $x_{\text{out}}$, using the following equation:

$$x_{\text{out}} = \begin{cases} a_{\text{out}}^- X_{\text{out}}^2 + b_{\text{out}}^- X_{\text{out}} + c_{\text{out}}^- & \text{with } X_{\text{out}} < 0 \\ a_{\text{out}}^+ X_{\text{out}}^2 + b_{\text{out}}^+ X_{\text{out}} + c_{\text{out}}^+ & \text{with } X_{\text{out}} \geq 0 \end{cases}$$

(34)

where $a_{\text{out}}^-$, $b_{\text{out}}^-$, $a_{\text{out}}^+$ and $b_{\text{out}}^+$ are coefficients of the polynomials (18), and $c_{\text{out}}$ is the real value corresponding to the linguistic value 0. This scaling is done with the delinguistification block (Fig. 17(d)). Coefficients of the polynomials are taken from the parameter structures with a Matlab function in the same way as for the linguistification.

The model is represented by

$$x_{\text{out}} = f_{\text{out}}( - \frac{1}{A_{\text{out}}^-} \sum_{j=1, j\neq \text{out}}^m A_{ij} f_j^{-1}(x_j(t - n_j)) + B_i ) ,$$

(35)

where the functions $f_j$ and $f_{\text{out}}$ are membership definitions. In the general case, the the weight factors

$$w_{ij} = - \frac{A_{ij}}{A_{i\text{out}}^-} ,$$

(36)

and the bias term

$$B_i = - \frac{B_i}{A_{i\text{out}}^-} ,$$

(37)

mean $n_i + 1$ parameters. The total number of additional parameters is $4 + 4n_i$ for $n_i$ input variables since the scaling functions of each variable require three additional parameters to two parameters needed for normalisation. The bias term $B_i = 0$ if the operating point is correctly defined. The coefficients $A_{ij}$ and $A_{i\text{out}}$ can be set to one or to a chosen value by modifying the scaling functions.

4.1.3 Dynamic LE models

Dynamic LE models provide additional methodologies for the nonlinear function and the algebraic equations. The basic form of the LE model is a static mapping in the same way as fuzzy set systems and neural networks, and therefore dynamic models will
Fig 17. Linguistic equation model in Simulink: parameters for the linguistification and delinguistification are selected by the variable index and the type of the definition (MATLAB Function). The types of membership definitions are following: 1 variable value, 2 time delay, 3 error, 4 change of error, 5 sum of errors, 6 original error, 7 change of control, 8 sum of control actions, and 9 fluctuation index.
include several inputs and outputs originating from a single variable. 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 presented 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 in the parametric models resulting

\[ Y_i(k) + a_1 Y_i(k - 1) = b_1 U_i(k - n_k) + e(t). \]  

This model is a special case of (30) with three values, \( Y_i(k), Y_i(k - 1) \) and \( U_i(k - n_k) \), in the linguistic range, the interaction matrix \( A = [1 \ a_1 - b_1] \), the bias term \( B_i = 0 \) and \( e(t) \) the noise term. The current value of the simulated variable \( x_j \) and an appropriate value of the control variable \( u_i \) as inputs and the new value of the simulated variable as an output.

Linear state-space models can be used in LE models by combining the coefficient matrices \( A_S, B_S, C_S \) and \( D_S \) to the interaction matrix

\[
A = \begin{pmatrix}
A_S & B_S & -I_x & 0 \\
C_S & D_S & 0 & -I_y \\
\end{pmatrix}
\]  

(39)
to handle the state variables, \( \vec{X}(t) \), and the inputs, \( \vec{U}(t) \), in the linguistic range. The identity matrices \( I_x \) and \( I_y \) introduce variables \( \vec{X}(k) \) and \( \vec{Y}(k - 1) \) to the model. The bias vector \( \vec{B} = 0 \). Equations can be used sequentially to solve one variable per equation. Through nonlinear scaling LE models can combine several linear local models (Section 2.1.3). In addition, gradual changes can be taken into account in membership definitions and interaction coefficients.

These models are used for calculating the change \( \Delta y \) for a sampling interval (Fig. 18(a)): the linguistification and delinguistification blocks are the same as in Figure 17. Although the bias vector \( \vec{B} = 0 \), the equation requires \( n_i + 2 \) parameters since two values of the output \( y \) are included. As the membership definition of the variable \( y \) does not depend on time, \( 4n_i + 5 \) parameters are needed in addition to the normalisation. An alternative approach is to make membership definitions also for the change \( \Delta y \) and obtain the derivative directly from the corresponding LE model. The number of parameters is \( 4n_i + 8 \) when nonlinear scaling functions are used. The output, the derivative of the variable \( y \), is integrated with numerical integration methods (Section 2.1.2). Step size control is often needed to adapt the simulation to changing operating conditions since the LE models are developed for wide operating areas. 

99
An appropriate handling of delays extends the operating area of the model considerably. For small systems, time delays can be taken into account by moving the values of input variables correspondingly, e.g. $n_k$ sampling intervals in (38). Initial estimates of the delays can be developed by correlation analysis, but similarities detected by the correlation analysis can be accidental in some cases. Scaled values are used since the basic correlation analysis is a linear approach. Local time delays can be analysed with different parametric models used in system identification: the number of delay $n_k$ in (38) is the effective time delay. The resulting delays should be assessed against process knowledge, especially for normal on-line process data (Juuso 1999a).

Effective delays depend on the working conditions, e.g. the production rate in many industrial processes. In the block shown in Figure 18(b), the delay of the variable $Var1$ depends on the variable $Var2$: the linguistic level of the variable $Var2$ is multiplied by 1 or -1 to get the linguistic level of the delay for the variable $Var1$, coefficient 1 means that the delay increases when the variable $Var2$ increases. The real value of the delay is obtained by the delinguistification block (Fig. 17(d)) with one difference, the type of the
definition is 2 (delay). This approach requires from two to five additional parameters per modelled time delay.

4.1.4 Case-based LE models

Feasible ranges of the variables \(x_j, j = 1, \ldots, m\), define areas for modelling in the LE approach (Fig. 9). A set of feasible ranges of the input and output variables defines a case, but the area can also be divided into subareas. Overlapping does not need to be based on the strong partition defined by (26). An effective approach to the data-based modelling of complex nonlinear systems is to partition the data set into subsets and approximate each subset by a special LE model. Decomposition can be based on expertise or data-based clustering methods applied on data, scaled data or principal components obtained from data or scaled data. Often the modelling areas are based on variables or facts which are not known on-line.

The shape of clusters is important in the multimodel approach. Most clustering methods lead to more or less symmetric clusters for normalised or unnormalised data. Linguistic fuzzy clustering or linguistic neural networks based on nonlinear scaling detect clusters of different geometrical shapes and thus considerably reduce the number of submodels needed. This also reduces the overfitting risk. Many clustering methodologies can be used in all these forms. The clustering variables are not necessarily used as inputs in the case models. Each case may even have a completely different set of variables. LE models, where the variables are the same for the cases and all the case models, belong to a special model class. This kind of model can be represented as an extension of TS fuzzy models, where the local linear models are LE models. Transitions between the submodels become smooth if the nonlinear LE submodels are overlapping.

Also these linguistic Takagi-Sugeno (LTS) fuzzy models can be tuned with the ANFIS method (Juuso 2009a). Steeply changing nonlinear models are handled with crisp or very narrow transition ranges, see Figure 10.

Case models are normal LE models (31): each equation has from two to five variables, whose membership definitions can be partly general and partly case specific. Cases may also include dynamic models. In small dynamic models, a single equation includes all the interactions, i.e. also variables affecting to the working point of the model are included in the model. For larger models, the equation system is a set of equations where each equation describes an interaction between two to five variables. The development work starts with the generation of membership definitions, which are
then used in the generation of interaction alternatives. Case models and fuzzy reasoning are used for the detection of operating conditions, and hybrid multimodel systems can be constructed by introducing fuzzy decisions to select submodels, see Section 6.2.2.

4.1.5 Large-scale LE models

Each equation has usually from three to five variables to keep the process insight. Interactions with more than five variables require special attention. Principal components compress the data by reducing the number of dimensions. This analysis can also be done for the scaled data: this approach is called linguistic principal component analysis (LPCA). The resulting features can be considered as indirect measurements, which can be used in symptom generation for fault detection.

For large systems, the number of possible variable combinations becomes very large (Fig. 19), e.g. the case models of the web break indicator system includes 24 variables, which means that there are 2024 alternative three variable interactions. In a paper machine application, 72 variables were used, which means 59,640 three variable groups, 1,028,790 four variable groups and over 13,991,544 five variable groups. Most of these alternatives are useless, and therefore, several methodologies are used for selecting alternative groups for processing. (Ahola et al. 2007, Juuso et al. 2008)

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig19.png}
\caption{Variable combinations (Juuso et al. 2008).}
\end{figure}

Variable selection is started with manual methods, and the final variable selection is based on generated alternative interactions assessed with domain expertise. Some variable combinations should be avoided, e.g. calculated variables should not be used
together with the variables which are used to calculate them. Also a group containing a controlled variable and its set point is not usually appropriate. These problems are avoided by defining the inappropriate groups as non-groups. A non-group should not be a part of any acceptable variable group.

4.2 Data-based modelling

Data-based modelling with linguistic equations is based on membership definitions, which developed and adapted variable by variable (Section 3.1.2). The data-based modelling has the following stages (Fig. 20): (1) generation of linguistic relations, (2) development of equation alternatives, (3) selection of equations. Adaptation modifies membership definitions, linguistic equations or both to improve fitting to the training data (Section 4.3). Linguistic relations are generated from the data by the nonlinear scaling of all the variables to the linguistic range [-2, 2] with the membership definitions. Each relation contains all the variables. Fuzzy rulebases are handled in a similar way: linguistic labels are replaced by real numbers and the mapping of these numbers to the linguistic range produces membership definitions. The locations of membership functions are important tuning parameters, see Section 6.2. For small rule bases the generation of data is needed to get enough material for fitting.

Equation alternatives are generated from the linguistic relations with linear regression for selected variable combinations. The interaction coefficients $A_{ij}$ and the bias term $B_i$ of (30) are obtained as least squares solutions of the alternative variables. The output is known for forecasting models but variables with low interaction coefficient are difficult to forecast. All alternative output variables are compared for case detection and the output variable giving the best fit is chosen if output variables are not specified. The structures of dynamic models are taken into account in the interaction matrix, see (38) and (39). The regression analysis is usually based on assumption that the distribution is normal. Methods based on other distributions are compared to find out if they produce a better fit. The effects of outliers can be reduced by using robust regression, but the development of membership definitions should already remove this problem.

Time delays, denoted as $n_j$, must be taken into account in the analysis of the interactions. Variable time delay, which is important in many applications, can be handled with membership definitions, see Figure 18(b). All the outlier and suspicious points are excluded after handling all effects of time delays.
Selected equations from alternatives are based either on the overall fit or on the prediction performance, several output alternatives are compared if appropriate. The performance measure of each equation \( i, i = 1, \ldots, n \), is evaluated in the linguistic range by the residual

\[
\epsilon_i = \sum_{j=1}^{m} A_{ij}X_j(t - n_j) + B_i. \tag{40}
\]

Each vector \((A_{i1} A_{i2} \ldots A_{im} B_i)\) is normalised to get the error measures comparable between equation alternatives. The equations are selected in such a sequence that each new equation will bring at least one new variable to the model, i.e. only one of the equation alternatives of each group can be included.

Case models are normal LE models, which are developed case by case from data as explained above. The model can be single equation or a set of equations, where each equation has from two to five variables. The membership definitions can be partly general and partly case specific. The residual \( \epsilon_i \) has a linguistic meaning, denoted as fuzziness of the equation, corresponding to the fuzziness used in fuzzy regression, see (Tanaka et al. 1982, Diamond 1988, Bargiela et al. 2007). For each equation \( i \), the distribution of the fuzziness in the training material is the basis for a feasible range of
the new variable $\varepsilon_i$. Detection of the operating conditions with LE models is based on this interpretation: the width of the distribution provides information about uncertainties, and the centre of the distribution moves from zero if there are systematic differences. Additional selection rules are needed for the case-based systems aimed for the detection of the operating conditions: each model set should be accurate in its own case but at the same time its fit to other cases should be much worse (Juuso 2004e).

**Variable grouping** is important in large-scale models. In small systems, even the directions of interaction are usually quite clear: only the absolute values of the coefficients need to be defined. For more complex systems, a set of alternative variable groups is developed first, and all the selected combinations are taken into account in generating the alternatives of the equations by using all acceptable three variable as a basis (Fig. 20). Four and five variable groups are based on three variable groups with one or two important variables, respectively.

The automatic variable grouping, which has several phases and alternative approaches for gradually reducing the number of variable groups, require appropriate preprocessing, feature extraction and correct time delays between the variables. Correlation analysis is used for selecting interesting groups from the acceptable groups. For forecasting models, input variables should have high correlation to the output variable but low correlation between each other. For case detection, causality is not always as evident: there is not necessarily any definite output variable, i.e. groups where several variables have high correlation between each other are acceptable. The nonlinear scaling with the membership definitions improves the correlation analysis for curvilinear relationships since the correlation analysis is a linear methodology. PCA is applied either to the original or to the scaled data to find large groups of variables which move together. These groups are divided into groups with three, four and five variables. Several clustering methods are used for dividing the data sets into different operating areas. Variable selection and modelling are done interactively and the final selection and grouping results from the selection of the linguistic equations. (Juuso et al. 2008)

**Nonlinear model structures** can be used but the goal is to capture the nonlinear effects with the membership definitions. The validity of the linearity assumption can be tested by applying nonlinear regression analysis, e.g. higher order polynomials or exponential functions, or alternative dynamic structures defined by parametric models and state-space models. For example interaction and quadratic terms of RSM model or nonlinear activation functions and complex structures in function expansions presented in (Ljung 2008) can be compared. TS fuzzy models and ANFIS tuning method can be
used for evaluating the effects of multimodel approaches (Juuso 2009a), see Section 6.2.1.

Data pre-processing is an essential part of modelling: appropriate filtering or interpolation needs to be included to find for each variable values, which are suitable for modelling. Calculation of moving averages, medians or value ranges already has a time delay which depends on the calculation window and the applied methodology. Imputation of the outliers and missing values can be used for assessing the test results of the dynamic models.

4.3 Adaptation of LE models

The structural restrictions of the LE models are beneficial in tuning and adaptation, which consists of two independent parts: scaling and interactions. Since only five parameters are needed to define the meaning for each variable, the LE systems can be adapted to various operating conditions with manual, neural or genetic methods.

The whole model can be constructed manually: the interaction coefficients and bias terms are then based on domain expertise and thus can be chosen freely, and the scaling functions can be manually defined for all the variables. Also the manually defined scaling functions should fulfill the constraints of the shape factors $\alpha_j^-$ and $\alpha_j^+$. In large-scale systems, domain expertise is important in the variable grouping (Fig. 20).

The neural tuning algorithm reduces the error between model and training data (Juuso 1999a). Membership definitions are tuned for one variable of the equation at a time with a linear neural network using the set of equations defined by (19). Recursive implementations of the regression or the linear network make on-line adaptation possible. The input of the linear network consists of a column vector

$$\vec{P} = [(X_j^+)^2 X_j^- (X_j^-)^2 X_j^-]^T, \quad (41)$$

where $X_j^+ = \max(0, X_j)$, $X_j^- = \min(0, X_j)$ and $X_j$ is the linguistic value of the variable $j$. The values $X_j$ are obtained from the linguistic equation which is used for tuning, and the target consists of the measured real values of $x_j$. The centre points $c_j$ are obtained by generalised norms or medians (Section 3.1.1), and the linear network is used to obtain the coefficients $a_j^-, b_j^-, a_j^+$ and $b_j^+$. The input is a $4 \times N$ matrix where $N$ is the number of measurement points. The values $X_j^+$ and $X_j^-$ have non-zero values when the real values $x_j > c_j$ and $x_j < c_j$, respectively. Requirements of monotonous, increasing functions are taken into account by analysing the derivatives of $f_j^+$ and $f_j^-$ and the
corner points (17) obtained by (19). The centre point $c_j$ must fill the inequalities (22). A locally linear model results at the centre points if $c_j$ is obtained from (25).

For dynamic models, all the old and new values of the variable $j$ are included, the linguistic values in the input and the real values in the output of the network. For the matrix equation (31), only one variable from each equation can be selected for tuning, i.e. membership definitions for other variables are either fixed or tuned with other equations.

Genetic tuning can handle the whole system simultaneously by comparing the effects of the parameters defined by the scaling functions, interactions, time delays, and the weights of the models. The fitness values of the models are obtained through simulation. To keep the solutions comparable, the combined coefficient vector $(A_{i1} A_{i2} \ldots A_{im} B_i)$ is normalised. Genetic tuning is controlled by the population size, the number of bits, and the probabilities of crossover and mutation.

The main part of the LE modelling is to tune the parameters of the scaling functions. The genetic tuning method can handle several variables at a time by varying the parameters defined by (17). The centre points $c_j$ are the real values corresponding to the normal (or the origin); the other parameters of the membership definitions are defined by the differences, i.e. the tuned parameters for variable $j$ are

$$(c_j, c_j - (c_l)_j, (c_l)_j - \min(x_j), (c_h)_j - c_j, \max(x_j) - (c_h)_j).$$

(42)

In this approach, which is similar to the methodology used in (Lotvonen et al. 1997), penalty functions are needed to ensure that the requirements of the functions are filled.

The constraint handling provides a basis for the genetic tuning approach introduced in (Juuso 2009e). No penalties are needed if coefficients $\alpha_j^-$ and $\alpha_j^+$ are restricted to the range $\left[\frac{1}{3}, 3\right]$, and the tuning is done for

$$(c_j, \Delta c_j^-, \alpha_j^-, \Delta c_j^+, \alpha_j^+).$$

(43)

For asymmetric linear functions, only three parameters are tuned since $\alpha_j^- = \alpha_j^+ = 1$. For linear definitions, the number of parameters reduces to two since $\Delta c_j^- = \Delta c_j^+$. In dynamic models, the definitions of variables are time invariant.

Interaction coefficients $A_{ij}$ and bias terms $B_i$ can be included in the genetic tuning in the same way as the parameters of the membership definitions (Juuso 2009e). Variable time delays (Fig. 18(b)) are handled by introducing additional bit sequences for the membership definitions of the time delays. The genetic tuning can be extended into multimodel systems by introducing corresponding parameters. For example, working
point models contain scaling functions, interactions, and effective time delays. The cascade models (Fig. 5) introduce additional parameters.

In current applications, real numbers are represented with binary coding, i.e. the population of the solution is represented with chromosomes described as bits. Different parts of the bit sequences define different parameters. The coded real values are then used in simulation to obtain the fitness of the model. The initial population is constructed by using limited value ranges: the ranges of the parameters used in the scaling functions are shown in Table 2, where $\bar{x}_j$ is the mean of $x_j$. All the interaction coefficients and bias terms have the same default range $[-1, 1]$. Outliers need to be removed when defining the minimum and maximum values. Expert knowledge can be used to modify the ranges. It is also possible to concentrate the initial population close to the corner points obtained from data analysis (Fig. 9).

Table 2. Ranges for the parameters of the scaling functions (Juuso 2009e, published by permission of Springer).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Lower limit</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_j$</td>
<td>$\frac{1}{4}(3\bar{x}_j + \min(x_j))$</td>
<td>$\frac{1}{4}(3\bar{x}_j + \max(x_j))$</td>
</tr>
<tr>
<td>$\Delta c^-_j$</td>
<td>$\frac{1}{4}(\bar{x}_j - \min(x_j))$</td>
<td>$\frac{1}{4}(\bar{x}_j - \max(x_j))$</td>
</tr>
<tr>
<td>$\Delta c^+_j$</td>
<td>$\frac{1}{4}(\max(x_j) - \bar{x}_j)$</td>
<td>$\frac{1}{4}(\max(x_j) - \bar{x}_j)$</td>
</tr>
</tbody>
</table>

Performance measures and complexity penalties are in principle the same as in the normal regression analysis. The objective function consists of a weighted average of several measures and penalty terms (Juuso 2009e). Since the residuals have a linguistic meaning, all the measures can be scaled to the linguistic range $[-2, 2]$ : the mean absolute error (MAE) already is in the range $[0, 2]$, and the mean square error (MSE) can be scaled to that but the scaled values are the same as the MAE values. Similarly, the correlation coefficients and the coefficient of determination, $R^2$, provide the same information when scaled to the range $[-2, 2]$ but the parameters of the scaling function can be chosen more flexibly. Maximum error and relative error are scaled to the range $[0, 2]$. The best linear fit of the calculated values to the scaled values provides two additional measures: the slope and the base. Penalty levels can be defined manually for the complexity of the model, e.g. the number of inputs, equations and submodels. A good solution has high correlation and $R^2$ values, small values for errors, base values and penalties, and a slope value close to one. Linear scaling without clipping is used
first to also handle very large residuals. The nonlinear scaling is taken into account after reducing the population.

The data-based framework shown in Figure 20 provides tools for developing smart adaptive systems: recursive tuning of scaling functions and equations is useful for the adaptation to changing environment and similar settings. Adaptation to new/unknown application can be done technically but the domain expertise is essential e.g. in variable grouping, data pre-processing and case definitions. Cascade and case-based solutions take into account new combinations of applications.
5 Linguistic equation control

Fuzzy controllers can be converted to linguistic equation form by replacing the symmetric parts of the rules with linguistic equations where linguistic levels for error, error derivative and change of control are represented by linguistic values. The first direct LE controller was implemented in 1996 for a solar power plant (Juuso et al. 1997b, 1998b), and later the multilevel LE controller was installed in an industrial lime kiln (Järvensivu et al. 2001). A genetic tuning method (Juuso 2006) has been improved by using the constraint handling method (Juuso 2009e) and the advanced nonlinear scaling (Juuso & Lahdelma 2010). New intelligent indices, which provide information about changing conditions, are increasingly important in the LE control (Juuso 2011b, 2012b,c).

5.1 Feedback LE controllers

Feedback linguistic equation (LE) controllers use error $e_j(k)$, derivative of the error $\Delta e_j(k)$, and sum of errors $\delta e_j(k)$ calculated by (2). These real values are mapped to the linguistic range by nonlinear scaling with variable specific membership definitions:

$$\tilde{e}_j(k) = \begin{cases} 
2 & \text{with } e_j(k) \geq \max(e_j(k)) \\
(f_{e_j})_{-1}(e_j(k)) & \text{with } \min(e_j(k)) \leq e_j(k) \leq \max(e_j(k)) \\
-2 & \text{with } e_j(k) \leq \min(e_j(k)) 
\end{cases}, \quad (44)$$

where the function $(f_{e_j})_{-1}$ is a nonlinear function which maps the real values of $e_j(k)$ to the range $[-2, 2]$;

$$\tilde{\Delta e}_j(k) = \begin{cases} 
2 & \text{with } \Delta e_j(k) \geq \max(\Delta e_j(k)) \\
(f_{\Delta e_j})_{-1}(\Delta e_j(k)) & \text{with } \min(\Delta e_j(k)) \leq e_j(k) \leq \max(\Delta e_j(k)) \\
-2 & \text{with } e_j(k) \leq \min(\Delta e_j(k)) 
\end{cases}, \quad (45)$$

where the function $(f_{\Delta e_j})_{-1}$ is a nonlinear function which maps the real values of $\Delta e_j(k)$ to the range $[-2, 2]$;

$$\tilde{\delta e}_j(k) = \begin{cases} 
2 & \text{with } \delta e_j(k) \geq \max(\delta e_j(k)) \\
(f_{\delta e_j})_{-1}(\delta e_j(k)) & \text{with } \min(\delta e_j(k)) \leq e_j(k) \leq \max(\delta e_j(k)) \\
-2 & \text{with } e_j(k) \leq \min(\delta e_j(k)) 
\end{cases}, \quad (46)$$

111
where the function \((f_{\delta e})^{-1}\) is a nonlinear function which maps the real values of \(\delta e_j(k)\) to the range \([-2, 2]\). All these membership definitions, whose centre points are zeros, are defined by the core and the ratios (21).

A PI-type LE controller is represented by

\[
\widehat{\Delta u}_{ij}(k) = e_j(k) + \Delta e_j(k),
\]

which is a special case of the matrix equation \(AX = 0\) with the interaction matrix \(A = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}\), and variables \(X = \begin{bmatrix} \hat{e}_j & \Delta e_j & \Delta u_{ij} \end{bmatrix}^T\). The linguistic value of the change of control, denoted as \(\Delta u_{ij}\), is obtained by (47) and limited to the range \([-2, 2]\) and the scaled to the real values \(\Delta u_{ij}(k)\) by the membership definitions of the change of error denoted as \((f_{\Delta u})_i\).

![Relations](a) Relations. ![Control surface](b) Control surface.

**Fig 21.** The rule base of a fuzzy PI controller represented by integer numbers and the control surface of the corresponding LE controller (Juuso 1999a, published by permission of Springer).

The PI-type fuzzy controller presented in (Juuso 1999a) can be replaced by a PI-type LE controller (Fig. 21). Each linguistic relation in Fig. 21(a) corresponds to a rule: linguistic labels \(NB, NS, ZO, PS, PB\) are replaced by integer numbers \(-2, -1, 0, 1\) and \(2\). All these rules can be obtained from (47) if \(-2\) and \(2\) are the minimum and maximum values, respectively. Control law (47) is also applicable on any fuzzy partition, and the procedure always produces a rule set which is complete, consistent, and continuous. The output \(i\) of a single input single output (SISO) controller is calculated by adding the
effect of the controlled variable $j$ to the manipulated variable $i$:

$$u_i(k) = u_i(k - 1) + \Delta u_{ij}(k).$$  \hspace{1cm} (48)

Different **PID-type fuzzy controllers** can be represented by linguistic equations (Tab. 3): the interaction matrix is defined by the coefficients $K_P(i, j)$, $K_D(i, j)$ and $K_I(i, j)$, which can be used in the tuning of the controllers. The directions of the control action are application specific. The inputs are handled with a vector $\vec{X} = [\tilde{e}_j \, \Delta \tilde{e}_j \, \delta \tilde{e}_j(k) \, \tilde{u}_{ij}]^T$. The linguistic value of the control, which is obtained for PD and PID controllers, is scaled to the real values $u_{ij}(k)$ by the membership definitions of the control denoted as $f_i$. The strengths for the effects of error, derivative of the error, and sum of errors can be tuned by membership definitions $(f_e)_j$, $(f_{\Delta e})_j$, and $(f_{\delta e})_j$, respectively.

**Table 3. PID-type LE controllers.**

<table>
<thead>
<tr>
<th>Type</th>
<th>$\tilde{e}_j(k)$</th>
<th>$\Delta \tilde{e}_j(k)$</th>
<th>$\delta \tilde{e}_j(k)$</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI</td>
<td>$K_i(i, j)$</td>
<td>$K_P(i, j)$</td>
<td></td>
<td>$\Delta u_{ij}(k)$</td>
</tr>
<tr>
<td>PD</td>
<td>$K_P(i, j)$</td>
<td>$K_D(i, j)$</td>
<td>$K_I(i, j)$</td>
<td>$u_{ij}(k)$</td>
</tr>
<tr>
<td>PID</td>
<td>$K_P(i, j)$</td>
<td>$K_D(i, j)$</td>
<td>$K_I(i, j)$</td>
<td>$u_{ij}(k)$</td>
</tr>
</tbody>
</table>

Several controlled variables can have effect on a single manipulating variable. The SISO controllers presented in Table 3 can be extended to MISO controllers by weighting the SISO controllers with weight factors $w^C_{ij}$

$$\Delta u_i(k) = \sum_{j=1}^{m} w^C_{ij} \Delta u_{ij}(k).$$ \hspace{1cm} (49)

for PI-type controllers, and

$$u_i(k) = \sum_{j=1}^{m} w^C_{ij} u_{ij}(k).$$ \hspace{1cm} (50)

for PD- and PID-type controllers.

Several manipulating and controlled variables can be taken into account with MIMO controllers defined by matrix equations. Errors, changes of error and sums of error are handled with vectors $\tilde{e}(k) \in L_m \subset \mathbb{R}^m$, $\Delta \tilde{e}(k) \in L_m \subset \mathbb{R}^m$, and $\delta \tilde{e}(k) \in L_m \subset \mathbb{R}^m$, respectively. Coefficients $K_P$, $K_D$ and $K_I$ are $n \times m$ matrices. The result, either the change of control vector or the control are column vectors $\Delta \vec{u}^{\tilde{e}}(k) \in L_n \subset \mathbb{R}^n$ and $\vec{u}^{\tilde{e}}(k) \in L_n \subset \mathbb{R}^n$. 

Each element of these vectors defined in the sets $L_m$ and $L_n$ is limited in the range $[-2, 2]$. The weights of the different SISO controllers are included in the coefficient matrices $K_P$, $K_D$ and $K_I$.

Basic PID-type LE controller handles the normal operation with symmetrical membership definitions $(f_e)_j$, $(f_{\Delta e})_j$, and $(f_{\delta e})_j$. The structure of the basic LE controller is similar to the structure of a steady-state model (Fig. 17). A PID-type controller needs three linguistics blocks, and the type of the definition is different for each of these blocs: denoted as 3 for $(f_e)_j$, 4 for $(f_{\Delta e})_j$, and 5 for $(f_{\delta e})_j$.

The PI-type LE controller (47) was introduced for a single manipulated variable in the solar thermal power plant (Juuso et al. 1998b). Several manipulated variables are important in the lime kiln control (Järvensivu et al. 2001).

Feedback LE controllers are not restricted to these types: the controller inputs can be also values of variables, changes of variables, etc. These variables were used for example in the LE controller which in tuning replaced the fuzzy logic controller of the flue gas fan in a lime kiln (Juuso et al. 1996).

## 5.2 Adaptation

Adaptation to changing operating conditions is necessary in industrial processes Juuso (1999a). As the basic level LE controller is implemented in a compact manner, additional higher-level structures can be introduced for adaptation purposes. The operation of the LE controller is modified by means of adaptive scaling, which is used to adjust the control surface in accordance with the changing operating conditions of the process. It, therefore, extends the accomplished working area of the basic LE controller. Gain scheduling type predefined adaptation (Section 2.3.2) is extended to LE models. Dynamic modelling and simulation is needed for comparing alternatives in controller design (Section 4.1.3). Facts for detecting operating conditions can be integrated and tested with case-based modelling (Section 4.1.4).

### 5.2.1 Adaptive scaling

The adaptive scaling of the manipulating variable $i$ can be based on the working point

$$wp_i(k) = \sum_{j=1}^{m} w_{ij}w^{wp}x_j(k) + b^{wp}_i,$$  

(51)

114
where \( \tilde{x}_j(k) \) and \( w_{ij}^{wp} \) are the linguistic values of the variable \( j, j = 1, \ldots, m \), and the corresponding weight factors. The bias term \( b_{i}^{wp} \) is specific to the manipulating variable. The working point \( wp_i(k) \in [-2, 2] \) indicates a deviation from the normal operation.

The linguistic value of the control power can be determined by the same principle as the working point of the process. In this case, \( x_j \) stands for the variable related to the gain of the manipulated variable.

\[
cp_i(k) = \sum_{j=1}^{m} w_{ij}^{cp} \tilde{x}_j(k) + b_{i}^{cp}, \tag{52}
\]

where \( \tilde{x}_j(k) \) and \( w_{ij}^{cp} \) are the linguistic values of the variable \( j, j = 1, \ldots, m \), and the corresponding weight factors, respectively. The bias term \( b_{i}^{cp} \) is specific to the manipulating variable. For instance variations in the heat energy content of the fuel can be handled with this technique. The control power \( cp_i(k) \in [-2, 2] \).

For slow processes with long time delays it is important to reduce the cumulative rate of control actions over the preset time period. The cumulative sum of control actions is computed as

\[
\delta u_i(k) = u_i(k - 1) - u(k - n_R), \tag{53}
\]

where \( n_R \) is the length of the window. The scaled value \( \tilde{\delta u}_i(k) \) computed by

\[
\tilde{\delta u}_i(k) = \begin{cases} 
2 & \text{with } \delta u_i(k) \geq \max(\delta u_i(k)) \\
(f_{\delta u_i}^{-1}(\delta u_i(k))) & \text{with } \min(\delta u_i(k)) \leq u_i(k) \leq \max(\delta u_i(k)) \\
-2 & \text{with } \delta u_i(k) \leq \min(\delta u_i(k)) 
\end{cases} \tag{54}
\]

defines the cumulative rate of control actions:

\[
\cr_i(k) = \begin{cases} 
\max\{0, \tilde{\delta u}_i(k)\} & \text{with } \Delta u_i(k) > 0 \\
\cr_i(k - 1) & \text{with } \Delta u_i(k) = 0 \\
\min\{0, -\tilde{\delta u}_i(k)\} & \text{with } \Delta u_i(k) < 0 
\end{cases} \tag{55}
\]

The rate \( \cr_i(k) \) receive only values between 0 and 2. According to (55), this scaling effect is active only following appropriate directions. The membership definition \( (f_{\delta u_i}) \) is specific to the manipulating variable \( i \).

Operation condition controller changes the control surface of the basic LE controller by modifying membership definitions for the change in the control variable \( \Delta u_i(k) \). The correction factor is obtained in the linguistic range as a weighted sum of these scaling actions:

\[
c_l^i(k) = w_{ij}^{wp} wp_i(k) - w_{ij}^{cp} \cp_i(k) - w_{ij}^{cr} \cr_i(k), \tag{56}
\]
where the weight factors $w_{w,\text{p}}^i$, $w_{c,\text{p}}^i$ and $w_{c,\text{r}}^i$ are all positive real values in the range $[0, 1]$, i.e. any of these adaptation methods can also be switched off on the high-level control (Section 5.4). Directions of the different scaling effects are defined by the signs of these weight factors. The scaling coefficient $sc_i(k)$ is obtained by scaling $sc_i(k) = 1 + c^i_F(k)$.

The adaptive scaling is tuned for each manipulated variable which are controlled with PI-type LE controllers. The output of a MISO-type LE controller is calculated as the weighted average of the corrections determined independently on the basis of the controlled variables $j$, $j = 1, \ldots, m$ as follows:

$$\Delta u_i(k) = sc_i(k) \sum_{j=1}^{m} w_{C,ij} \Delta u_{ij}(k), \quad (57)$$

where $\Delta u_i(k)$ is the change of control for the manipulated variable $i$: $\Delta u_{ij}(k)$ and $w_{C,ij}$ are, respectively, the change and weight related to the controlled variables $j$, $j = 1, \ldots, m$.

Membership definitions are needed for all variables used in models (51) and (52), the cumulative sum of control actions, $\delta u_i(k)$, and the scaling coefficient, $sc_i$. Weight factors $w_{w,\text{p}}^i$ and $w_{c,\text{p}}^i$ belong to the models. Weight factors $w_{w,\text{p}}^i$, $w_{c,\text{p}}^i$ and $w_{c,\text{r}}^i$, as well as $w_{C,ij}$ in (57), are used for balancing different aspects and control strategies. Bias terms $b_{w,\text{p}}^i$ and $b_{c,\text{p}}^i$ are aimed for fine tuning during the operation.

The working point adaptation (51) was introduced for a single manipulated variable in the solar thermal power plant (Juuso et al. 1998b) and used in water treatment (Joensuu et al. 2004). The control power, the cumulative rate and several manipulated variables are important in the lime kiln control (Järvensivu et al. 2001).

5.2.2 Predictive braking action

The predictive braking action (PBA) is used to ensure smooth recovery after severe disturbances or smooth change to a considerably different set point. The PBA reduces the risk of oscillation and large overshoot after a sizeable deviation in the controlled variable(s), both of which are common complications especially in processes with long dead times.

For each controlled variable $j$, the PBA changes membership definition for the derivative of the error $\Delta e_j$ by using a braking rate coefficient $bc_j$ which is based on the initial error $e^I_j$ and the current change of error $\Delta e_j$.
The initial error $e^I_j$ is determined as the error at the turning point where the derivative of the error changes from positive to negative, or vice versa. The initial error is defined by the following principle:

if $e^I_j(k-1) = 0$ and $e_j(k) \leq e^-_j$ then

$$e^I_j(k) = \begin{cases} 
0 & \text{if } \Delta e_j < 0, \\
 e^I_j(k-1) & \text{if } \Delta e_j \geq 0, 
\end{cases}$$

(58)

if $e^I_j(k-1) = 0$ and $e_j(k) \geq e^+_j$ then

$$e^I_j(k) = \begin{cases} 
0 & \text{if } \Delta e_j > 0, \\
 e^I_j(k-1) & \text{if } \Delta e_j \leq 0, 
\end{cases}$$

(59)

if $e^I_j(k-1) \neq 0$ then

$$e^I_j(k) = \begin{cases} 
0 & \text{if } e^I_j(k-1) \Delta e_j > 0, \\
 e^I_j(k-1) & \text{if } e^I_j(k-1) \Delta e_j \leq 0, 
\end{cases}$$

(60)

where $e^I_j(k)$ and $e^I_j(k-1)$ are the new initial error and the previous initial error, respectively. $e^-_j$ and $e^+_j$ are the preset low and high boundaries for the initial error (note: the braking action is activated only when a relatively large deviation occurs, i.e. the error is above $e^+_j$ or below $e^-_j$). The moving averages of the derivative of the error $\Delta e_j(k)$ are used in (58) . . . (60) to filter out the influence of fluctuations in the measurement. If the target value of the controlled variable is changed, then the initial error is correspondingly updated by the following equation:

$$e^I_j(k) = e^I_j(k-1) + (y^{fp}(k-1) - y^{sp}(k)).$$

(61)

The procedure (58) . . . (61) has been developed for slow processes. For faster processes, averaging can be done over the sampling period. The original error $e^I_j(k)$ can be obtained from the low and high values of the controlled variable:

if $y^{fp}(k) > x_j^-(k)$ then $e^I_j(k) = y^{fp}(k) - x_j^-(k)$

else if $y^{fp}(k) > x_j^+(k)$ then $e^I_j(k) = y^{fp}(k) - x_j^+(k)$

else $e^I_j(k) = 0.$

(62)

The low and high values, $x_j^-(k)$ and $x_j^+(k)$, are updated by

if $x_j(k) < x_j(k-1) + \varepsilon$ then $x_j^-(k) = x_j(k),$

if $x_j(k) > x_j(k-1) - \varepsilon$ then $x_j^+(k) = x_j(k).$

(63)
where \( \varepsilon \) is a small real number.

The scaled value \( \tilde{e}_j(k) \) is computed by

\[
\tilde{e}_j(k) = \begin{cases} 
2 & \text{with } e_j(k) \geq \max(e_j(k)) \\
(f_x)_j^{-1}(e_j(k)) & \text{with } \min(e_j(k)) \leq e_j(k) \leq \max(e_j(k)) \\
-2 & \text{with } e_j(k) \leq \min(e_j(k)) 
\end{cases},
\]

where the membership definition \((f_x)\) is a nonlinear function which maps the real values of \( e_j(k) \) to the linguistic range \([-2, 2]\).

The PBA can be made stronger when the controlled variable goes closer to the set point:

\[
bc_j(k) = \left( (c_B) \right)_j \tilde{e}_j(k) \left( 1 - \frac{e_j(k)}{e_j(k)} \right),
\]

where \( bc_j(k) \) is the braking rate coefficient, \( bc_j(k) \in [0, 2] \), related to the controlled variable \( j \). Coefficient \((c_B)_j \) is a manually adjustable braking constant in \([0, 1]\) used for fine-tuning the force of braking (Juuso et al. 1998b). Close to the set point this force is reduced:

\[
\text{if } |\tilde{e}_j(k)| < 1 \text{ then } (c_B)_j = \frac{1}{2} (c_B)_j (1 + |\tilde{e}_j(k)|).
\]

Another solution is to start with a strong braking and reduce the effect when the controlled variable goes closer the set point (Järvensivu et al. 2001):

\[
bc_j(k) = \left( (c_B) \right)_j \tilde{e}_j(k) \frac{e_j(k)}{\tilde{e}_j(k)},
\]

In this case, braking reaches a maximum immediately after the turning point, and then decreases as the error declines and the controlled variable approaches the target value, see the term \( e_j(k)/\tilde{e}_j(k) \) in (67). Membership definitions and threshold values of the initial error \( e_j(k) \) are related to the membership definitions of the corresponding error \( e_j(k) \).

In practice, \( bc_j(k) \) is used to emphasise the influence of the derivative of the error by means of the following equation:

\[
K_P(i,j) = (1 + bc_j(k)) K_P(i,j)
\]

where \( K_P(i,j) \) is the coefficient used in Table 3. The control surface is modified only when PBA action has been activated. The braking coefficient \((c_B)_j \in [0, 1]\) is used manually on the high-level control (Section 5.4).

The predictive braking action was introduced in the solar thermal power plant (Juuso et al. 1998b). The braking coefficient (65) was modified to a different form (67) in the
lime kiln control (Järvensivu et al. 2001). This action is not used in water treatment control (Joensuu et al. 2004).

### 5.2.3 Asymmetrical action

Asymmetrical action (ASA) is aimed for fine-control in cases where a precise set point tracking is required. The action is activated only close to the set point if there are no strong fluctuations of the controlled variable evaluated by \( e_j^- \) and \( e_j^+ \).

\[
K_P(i, j) = \begin{cases} 
(1 + (c_A)_j \Delta wp(k))K_P(i, j) & \text{with } (f_e)_j^{-1}(e_j(k))\Delta wp_i(k) > 0 \\
(1 + (c_A)_j \Delta wp(k))^{-1}K_P(i, j) & \text{with } (f_e)_j^{-1}(e_j(k))\Delta wp_i(k) < 0 
\end{cases}, \quad (69)
\]

The asymmetry coefficient \((c_A)_j\) is aimed for manual use. This action was already in the first LE controllers of the solar thermal power plant (Juuso et al. 1998b) and later updated drastically: the calculation is now based on the changes of the corrected irradiation (Juuso 2012c). The ASA is not needed in the lime kiln (Järvensivu et al. 2001) nor in water treatment (Joensuu et al. 2004).

### 5.2.4 Constraint handling

Each manipulated variable \(u_i\) should be on a variable specific acceptable range \([u_{il}^i, u_{ih}^i]\) where \(u_{il}^i\) and \(u_{ih}^i\) are the predetermined high and low limits, respectively. These limits, which are determined by certain physical constraints, are set in order to protect personnel and safeguard equipment.

The range can further be reduced on the basis of process knowledge, e.g. the acceptable range for the manipulated variable is calculated by the system as a pipe around the most recent output of the FFM module \(u_{FF}^i(k)\):

\[
u_i(k) = \begin{cases} 
u_{hr}^i & \text{if } u_i(k) \geq u_{hr}^i \\
u_i(k) & \text{if } u_{lr}^i < u_i(k) < u_{hr}^i \\
u_{lr}^i & \text{if } u_i(k) \leq u_{lr}^i \end{cases}, \quad (71)
\]
where the dynamic limits,

\[ u_{hr}^i(k) = \begin{cases} 
  u_{hl}^i & \text{if } \lambda_i^+ u_{FF}^i(k) \geq u_{hl}^i, \\
  (1 + \lambda_i^+) u_{FF}^i(k) & \text{if } \lambda_i^+ u_{FF}^i(k) < u_{hl}^i,
\end{cases} \]

(72)

and

\[ u_{lr}^i(k) = \begin{cases} 
  (1 - \lambda_i^-) u_{FF}^i(k) & \text{if } \lambda_i^- u_{FF}^i(k - 1) > u_{ll}^i, \\
  u_{ll}^i & \text{if } \lambda_i^- u_{FF}^i(k) \leq u_{ll}^i,
\end{cases} \]

(73)

are between the high and low boundaries \( u_{hl}^i \) and \( u_{ll}^i \). The dynamic boundaries are changed in accordance with the actual loading state of the process. Constants \( \lambda_i^+ \) and \( \lambda_i^- \) are used to calculate the width of the range for acceptable control actions. (Järvensivu et al. 2001)

The constraint handling (CH) may also produce additional changes of control \( \Delta u_{FF}^i(k) \) if large deviations or fast changes are detected either in the process input or output. These smart actions are presented in the application part (Section 7.3) as they are highly process specific.

### 5.3 Model-based LE control

*Feedforward control* can be performed with inverted LE models, since nonlinear scaling with membership definitions and LE models can be used to in any direction (Juuso 1999a). The feedforward controllers operate in the same way as steady-state LE models. Each equation can have one output: \( Y_i(k), i = 1, \ldots, n \). The converted real outputs \( y_i(k) \) are limited to the range \([u_{ll}^i, u_{hl}^i]\) defined by membership definitions. The control system may contain additional constraints and parameters for fine tuning purposes (Järvensivu et al. 2001):

\[ u_{FF}^i(k) = \begin{cases} 
  u_{hl}^i & \text{if } y_i(k) + b_{FF}^i \geq u_{hl}^i, \\
  y_i(k) + b_{FF}^i & \text{if } u_{ll}^i < y_i(k) + b_{FF}^i < u_{hl}^i, \\
  u_{ll}^i & \text{if } y_i(k) + b_{FF}^i \leq u_{ll}^i,
\end{cases} \]

(74)

where \( u_{FF}^i(k) \) is the output of the feedforward controller linked to the manipulated variable \( i \). The output is limited by the limits \( u_{ll}^i \) and \( u_{hl}^i \) to be sure that the output is in an appropriate range even after fine tuning with the bias term \( b_{FF}^i \).

Feedforward controllers can also be based on heuristic LE systems and manually constructed membership definitions. These controllers can be built also from fuzzy logic controllers (FLCs) as in Figure 22 where the linguistic relations (Fig. 22(a)) are
developed from the rule set presented in (Juuso 1999a). The corresponding linguistic equation is developed from these relations provides a control surface (Fig. 22(b)) which is much smoother than the control surface of the original FLC. The development of LE systems from fuzzy set systems is discussed in Chapter 6.

The working point model (51) can be used in cascade control to limit setpoints $y_{sp}^j(k)$ for a chosen working point $w_{p}(k)$ (Juuso & Valenzuela 2003). The limiting working point is calculated by using intelligent indices, which detect changes in operating conditions (Juuso 2012b). In the current control system, the set point is the minimum of the alternative setpoint values: manual setpoint and setpoint obtained from the limited working point $w_{p_{min}}$, which is calculated and limited by the corresponding manually defined value (Juuso 2012c).

LE models are feasible for introducing nonlinearity to the IPC and MPC controllers: invertible LE models can replace linear models in IPC approach; also interactions of several variables can be included in the LE models, which is beneficial in MPC approach.
5.4 High-level LE control

The output of each LE controller, \( \Delta u_i(k) \), is already limited by the membership definitions \((f_{\Delta u_i})\), but further checking of the acceptable ranges is important in all applications with long process delays. Therefore, the outputs of the controller, \( \Delta u_i(k), i = 1, \ldots, n \), are applied in the system for control purposes only after checking them with respect to the acceptable ranges as follows:

\[
\Delta u_{\text{SC}}^i(k) = \begin{cases} 
\Delta u_i^{hl} & \text{if } s_i \Delta u_i(k) \geq \Delta u_i^{hl}, \\
 s_i \Delta u_i(k) & \text{if } \Delta u_i^{ll} < s_i \Delta u_i(k) < \Delta u_i^{hl}, \\
 \Delta u_i^{ll} & \text{if } s_i \Delta u_i(k) \leq \Delta u_i^{ll}.
\end{cases}
\] (75)

Each output of the stabilising controller (SC) module, denoted by \( \Delta u_{\text{SC}}^i(k) \), can be adjusted manually with a constant \( s_i \in [0.5, 1.5] \), which can be used for fine-tuning purposes. \( \Delta u_i^{ll} \) and \( \Delta u_i^{hl} \) are the preset low and high limits for the FB correction per scan interval.

The feedback control action combines the effects of the adjusted stabilising controller and the constraint handling calculated on the basis of the latest FB corrections of the SC module \( \Delta u_{\text{SC}}^i(k) \) and the stepwise correction of the CH module \( \Delta u_i^{\text{CH}}(k) \):

\[
u_i^{FB}(k) = u_i^{FB}(k - 1) + \Delta u_{\text{SC}}^i(k) + \Delta u_i^{\text{CH}}(k).
\] (76)

The feedback effect is added to the most recent output of the feedforward module, \( u_i^{FF}(k) \), i.e. \( u_i^{FB}(k) \) is corrections to the output of the FF module.

\[
u_i(k) = u_i^{FF}(k) + u_i^{FB}(k) + u_i^{RC}(k),
\] (77)

where \( u_i^{RC}(k) \) is a bias correction, which can be adjusted manually by the operators when the controller is turned on.

5.5 Intelligent indices

Any scaled variable \( X_j, j = 1, \ldots, m \), or residual \( \varepsilon_i, i = 1, \ldots, n \), defined by (40) can be used as an intelligent index. The working point (51) and the control power (52) are examples of these indices. Since the scaled variables can be handled as zero centered symmetrical distributions, also an absolute value of the residual \( |\varepsilon_i| \) is a suitable index in detecting fluctuations from the operating point.
Trend analysis provides useful indirect measurements for high-level control. For any variable $j$, a trend index $I^T_j (k)$ is calculated from the scaled values $X_j$ by a linguistic equation

$$I^T_j (k) = \frac{1}{n_s + 1} \sum_{i=k-n_s}^{k} X_j(i) - \frac{1}{n_L + 1} \sum_{i=k-n_L}^{k} X_j(i),$$

which is based on the means obtained for a short and a long time period, defined by delays $n_s$ and $n_L$, respectively. The index value is in the linguistic range $[-2, 2]$ representing the strength of both decrease and increase of the variable $x_j$. (Juuso et al. 2009)

The derivative of the index $I^T_j (k)$, denoted as $\Delta I^T_j (k)$, is used for analysing triangular episodic representations (p. 49). An increase is detected if the trend index exceed a threshold $I^T_j (k) > \epsilon^+_1$. Correspondingly, $I^T_j (k) < -\epsilon^-_1$ for a decrease (Fig. 23). These trends are linear if the derivative is close to zero: $-\epsilon^-_2 < \Delta I^T_j (k) < -\epsilon^+_2$. Concave upward monotonic increase (D) and concave downward monotonic decrease (B) are dangerous situations. Concave downward monotonic increase (A) and concave upward monotonic decrease (C) mean that an unfavourable trend is stopping. (Juuso 2011a)

Fig 23. Triangular episodic representations defined by the index $I^T_j (k)$ and the derivative $\Delta I^T_j (k)$ (Juuso 2011a, published by permission of IEEE).

The trend detection logic is similar to a typical PI-type LE controller (47). The system could be implemented as a fuzzy set system by generating the membership functions. In the present system, fuzzy logic is used only for explaining the progress of the analysis.
Severity of the situation can be evaluated by a deviation index

\[ I^D_j(k) = \frac{1}{3} \left( I^T_j(k) + I^T_j(k) + \Delta I^T_j(k) \right). \]  

(79)

This index has its highest absolute values, when the difference to the set point is very large and is getting still larger with a fast increasing speed. (Juuso et al. 2009) This can be understood as an additional dimension in Figure 23.

The trend analysis is tuned to applications by selecting the time periods \( n_L \) and \( n_S \). Further fine-tuning can be done by adjusting the weight factors \( w^{T1} \) and \( w^{T2} \) used for the indices \( I^T_j(k) \) and \( \Delta I^T_j(k) \). The thresholds \( \varepsilon^+ = \varepsilon^- = \varepsilon^2 = 0.5 \) (Juuso & Laakso 2011).

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

\[ \Delta x^F_j(k) = ||^{K_M^{M\alpha}}|p_h| - ||^{K_M^{M\alpha}}|p_l|, \]  

(80)

where the orders \( p_h \in \mathbb{R} \) and \( p_l \in \mathbb{R} \) are large positive and negative, respectively. The moments are calculated from the latest \( K_a + 1 \) values, and an average of several latest values of \( \Delta x^F_j(k) \) is used as an indicator of fluctuations. (Juuso 2012b)

Condition and stress indices can be calculated from continuously measured signals. The analysis is based on consecutive equally sized samples. Duration of each sample is the sample time \( \tau \) in seconds. The number of signal values \( N = \tau N_s \) where \( N_s \) is the number of signal values which are taken in a second. Norms can also be calculated by using sub-blocks, see (10). The sample time connects the moment to control applications. This idea was introduced in (Juuso & Lahdelma 2006) as a cavitation index. The combined approach can be based on generalised moments \( ^aM^p \) and norms,

\[ ||^{aM^p}\|_p = (^{aM^p})^{1/p} = \left[ \frac{1}{N} \sum_{i=1}^{N} |(x^{(a)}_j)^{(p)}|^1/p \right], \]  

(81)

which is the norm (7) calculated from the derivative \( x^{(a)} \) for a sample time \( \tau \). The order of derivation \( (a) \) and the order of the moment \( (p) \) are real numbers. In vibration analysis, rapid changes in acceleration become emphasised upon the derivation of the signal \( x^{(2)} \). Higher order derivatives, especially \( x^{(4)} \), work very well in the whole range from slowly to very fast rotating rolling bearings. Real order derivatives \( x^{(a)} \) provide additional possibilities. Generalised moments \( ^aM^p \) introduced in (Lahdelma & Juuso 2008b) were normalised with the standard deviation \( \sigma_a \). The norms (81) introduced in (Lahdelma & Juuso 2008a) have the same dimensions as the signal to be analysed.
The analysis can be further improved by taking into account nonlinear effects by monotonously increasing scaling functions, which were already used in the first applications in 2006. The new scaling approach presented in Section 3.1.1 was introduced in (Juuso & Lahdelma 2010). Condition indices can be based on several scaled features, which are all scaled to the range \([-2, 2]\):

$$I_C^{(α)}(k) = \sum_{j=1}^{n} w_j f_j^{-1}\left(\max(||\mathbf{M}_p^α||)\right) + \sum_{j=n+1}^{n+m} w_j f_j^{-1}(F_j^{(α)}),$$  \hspace{1cm} (82)

where \(w_j\) is the coefficient and \(f_j^{-1}\) the scaling function of the feature \(j\). Features include maximum norms \(\max(||\mathbf{M}_p^α||)\) and other features \(F_j^{(α)}\), e.g. the bins of the histograms. Maximum is obtained from several samples, see (10). Features can have specific frequency ranges. Good conditions correspond to value 2 and not allowable conditions to value -2. Index (82) is obtained from the features of the signal \(x^{(α)}\), but an index can also combine the features of different physical signals, e.g. stress caused by deviations from good operating conditions can obtained by the residual (40).

Stress indices \(I_S^{(α)}(k)\) are calculated in the same way as \(I_C^{(α)}(k)\): low stress corresponds to value -2 and not allowable stress to value 2. As operating conditions with high stress should be avoided, stress indices can be used in control: the cavitation index has been tested for power control (Juuso & Lahdelma 2009). The control system has a feedforward controller, which allocates the load to the turbines by means of cavitation indices, and a feedback controller, which is based on the LE approach. Correspondingly, condition indices \(I_C^{(α)}(k)\) can be used in adapting the control to the condition of process equipment.

Cases \(l, l = 0, \ldots, n_c\), in case-based LE models may include several equations. Then a degree of membership \(\mu_c(l)\) can be calculated for each case \(l\) by

$$\mu_c(l) = \frac{\sum_{i=1}^{n_e(l)} w_e(i,l)\mu_e(i,l)}{\sum_{i=1}^{n_e(l)} w_e(i,l)},$$  \hspace{1cm} (83)

where degree of membership of equations \(i, i = 1, \ldots, n_e(l)\), denoted as \(\mu_e(i,l)\), is based on the corresponding residuals \(\varepsilon_e(i,l)\), which are calculated by (40) and compared with the distribution of error, denoted as the fuzziness of the equation. The distribution of error is represented as a trapezoidal membership function developed on the basis of the train case. Each equation \(i\) has a case specific weight \(w_e(i,l)\), which is based on the performance of the equation in detecting the case \(l\).

Similar condition, stress or quality level can also be achieved in several cases where the number of cases is category specific: \(n_c(L), L = 1, 2, \ldots, n_{cat}\). The case with the
maximum degree of membership is chosen in each category:

\[ \mu_{\text{cat}}(L) = \max\{ \mu_{\text{cat}}(1), \mu_{\text{cat}}(2), \ldots, \mu_{\text{cat}}(n_c(L)) \}, \]  

(84)

and finally, the predicted category is the category, which has the case with the highest degree of membership. Maximum is used since only one of the cases \(1, 2, \ldots, n_c(L)\) is needed to activate a category \(L\). (Juuso & Ahola 2008)

The performance analysis can be based on normal (good) conditions by obtaining the condition index \((I_c(k))_0\) from the degree of membership defined by (84):

\[ (I_c(k))_0 = 4 \mu_{\text{cat}}(0) - 2. \]  

(85)

More detailed analysis is not needed for very good condition, \(\mu_{\text{cat}}(0)\) close to 1, if the categories do not have too much overlapping, i.e. the degree of membership is low for all fault cases. If \((I_c(k))_0 < 2 - \delta_c\), also the fault cases \(l, l = 1, \ldots, n_c\), are taken into account:

\[ I_c(k) = w_{\text{cat}}(0)(I_c(k))_0 - \sum_{L=1}^{n_c} w_{\text{cat}}(L)(4 \mu_{\text{cat}}(L) - 2), \]  

(86)

The weight factors \(w_{\text{cat}}(L)\) are based on the severity of the category in an application specific way. The indices (85) and (86) can also be understood as quality indices \(I_Q(k)\) if the cases are related to quality.

Deviations from the good operating conditions indicate stress, which is represented with the stress index

\[ (I_S(k))_L = w_{\text{cat}}(L)(4 \mu_{\text{cat}}(L) - 2), \]  

(87)

i.e. each category represents a component of the stress vector \(\vec{I}_S\). The severity of the component is given by the weight \(w_{\text{cat}}(L)\). In quality control, the corresponding terms \((I_Q(k))_L\) are understood as quality reducing factors.

Health indices \(I_H(k)\) can be calculated from the condition index:

\[ I_H(k) = 1 - \frac{2 - I_c(k)}{4}(1 - \delta_H), \]  

(88)

where \(\delta_H\) is the value of \(I_H(k)\) index when the condition index \(I_c(k) = -2\). The measurement index \(I_M(k)\) is an inverse of the index \(I_H(k)\). If the parameter \(\delta_H = 0.2\), the highest values of the index \(I_M\) are 5. These indices can combine several measurements, e.g. vibration, pressure, temperature, electric current, rotation speed. Originally, the \(I_M\) and \(I_H\) indices, known as \(MIT\) and \(SOL\), were obtained from normalised features where the good conditions are the reference cases: \(I_H(k) = I_M(k) = 1\) (Lahdelma &
Intelligent indices extend this to nonlinear analysis. Indices are used in diagnostics, see Section 6.2.3. Changes of condition and stress can be indicated by trend and deviation indices.

5.6 Adaptation of LE control systems

Multiple input multiple output (MIMO) controllers are represented as a single matrix controller which contains both feedforward and feedback actions. For large-scale systems, the matrices are sparse. The LE controller can combine various control strategies: all the modules shown in Figure 7 can be realised with the LE methodology. The implementation is very compact since no special switching programs or rules are needed. A LE controller has the same structure as the steady-state LE models (Fig. 17(a)). Control blocks are built with nonlinear scaling, limits, control laws and models, and the blocks are selected into use in the high-level control. Additional manual control actions can be introduced in the high-level control.

The controllers are configured stagewise: first the membership definitions (Tab. 4) with the default values of the coefficients listed in Table 5, and then fine-tuning of the coefficients; intelligent indices and advanced procedures are introduced gradually; different actions are activated with the high-level control (Tab. 6).

Nonlinear scaling is the key of the LE control: membership definitions are essential parts of the controllers, and the high and low limits and thresholds are related to them (Tab. 4). All the definitions are variable specific, referred by the variable index, and do not depend on the control module, i.e. the scaling functions are defined only once for each variable and used in adaptive scaling, intelligent indices or feedforward controllers when needed. Feedforward control is used for a part of the manipulating variables: \( n_{FF} \) out of \( n_M \).

Additional scaling functions defined in Figure 17(d) are needed for feedback control, predictive braking and fluctuation indices. For each controlled variable membership definitions are needed for \( e_j, \Delta e_j \) and \( \delta e_j \), depending on the type of the feedback controller, i.e. the types of membership definitions (Fig. 17(b)) are chosen from 3, 4 and 5. For the predictive braking the original error (type 6) is used. The moving fluctuation indices are variable specific (type 9). The controller output is either \( u_i \) or \( \Delta u_i \) corresponding to type 1 or 7, respectively. The change of control is defined in the same way for all the controllers with the same manipulating variable, i.e. \( \Delta u_{ij} \).
Table 4. Membership definitions and limits of a MIMO LE controller, numbers of the variables are denoted in the following way: $n_M$ manipulating variables, $m_C$ controlled variables, $n_{FF}$ manipulating variables used in feedforward control, $m_{FF}$ variables used in feedforward control, $n_{FB}$ manipulating variables used in feedback control, $m_W$ working point variables, $m_P$ variables effecting to the control power, $m_T$ variables with a trend index, $m_F$ variables with a fluctuation index, $m_A$ variables in asymmetry action, $m_{CM}$ variables in condition indices, and $m_S$ variables in stress indices.

<table>
<thead>
<tr>
<th>Action</th>
<th>Definitions and parameters</th>
<th>Maximum number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedforward</td>
<td>Membership definitions for $e_j$, $e_j$, $\delta e_j$, $\Delta u_j$ and $u_i$</td>
<td>$5 \cdot (n_{FF} + m_{FF})$</td>
</tr>
<tr>
<td>Feedback</td>
<td>Membership definitions</td>
<td>$9 \cdot n_{FB} + 12 \cdot m_C$</td>
</tr>
<tr>
<td>Adaptive scaling</td>
<td>Membership definitions for $sc_i$</td>
<td>$5 \cdot m_W$</td>
</tr>
<tr>
<td>Working point</td>
<td>Membership definitions</td>
<td>$5 \cdot m_P$</td>
</tr>
<tr>
<td>Control power</td>
<td>Membership definitions for $(f_{sw})_i$</td>
<td>$4 \cdot n_M$</td>
</tr>
<tr>
<td>Cumulative rate</td>
<td>Membership definitions for $\epsilon_i$</td>
<td>$4 \cdot m_C$</td>
</tr>
<tr>
<td>Predictive braking</td>
<td>Deviation thresholds $\epsilon_{i-1}$ and $\epsilon_{i+1}$</td>
<td>$2 \cdot m_C$</td>
</tr>
<tr>
<td>Asymmetrical action</td>
<td>Linear membership definitions</td>
<td>$m_A$</td>
</tr>
<tr>
<td>Trend index</td>
<td>Membership definitions</td>
<td>$5 \cdot m_T$</td>
</tr>
<tr>
<td>Fluctuation index</td>
<td>Linear membership definitions</td>
<td>$m_F$</td>
</tr>
<tr>
<td>Condition index</td>
<td>Membership definitions</td>
<td>$5 \cdot m_{CM}$</td>
</tr>
<tr>
<td>Stress index</td>
<td>Membership definitions</td>
<td>$5 \cdot m_S$</td>
</tr>
</tbody>
</table>

and $\Delta u_i \forall j = 1, \ldots, m_C$. The sum of control actions (type 8) is used in assessing the cumulative rate $\delta u_i$ (Tab. 4).

Control laws are realised as equation blocks (Fig. 17(c)). Feedforward controllers and adaptive scaling with the working point and the control power have model structures like (35) defined by weight factors (36) and bias terms (37). Development of feedforward controllers is based on modelling if inverted models are used. Feedback LE controllers have similar parameters as PID controllers. All these parameters and the window size of the cumulative rate are controller specific (Tab. 5). The basic feedback controller is usually a PI-type LE controller. Trend, condition and stress indices can be calculated for any variable. Short and long time windows are important parameters for the trend index. The condition and stress indices may contain weight factors (Tab. 5). The deviation index (79) does not need any additional parameters. Correspondingly, health indices are calculated from condition indices. Trend and deviation indices, fluctuation indicators,
Table 5. Coefficients of a MIMO LE controller, notations are shown in Table 4.

<table>
<thead>
<tr>
<th>Action</th>
<th>Definitions and parameters</th>
<th>Maximum number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedforward</td>
<td>Weight factors $w_{ij}^{FF}$ for $m_{FF}$ variables</td>
<td>$m_{FF}$</td>
</tr>
<tr>
<td></td>
<td>Bias term $b_{ij}^{FF}$</td>
<td>$n_{FF}$</td>
</tr>
<tr>
<td>Feedback</td>
<td>Coefficients $K_p(i, j)$, $K_i(i, j)$ and $K_0(i, j)$</td>
<td>$3 \cdot n_{FB} \cdot m_{C}$</td>
</tr>
<tr>
<td>Adaptive scaling</td>
<td>Working point</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weight factors $w_{ij}^{wp}$ for $m_{w}$ variables</td>
<td>$m_{w} \cdot n_{FB}$</td>
</tr>
<tr>
<td></td>
<td>Bias term $b_{ij}^{wp}$</td>
<td>$n_{FB}$</td>
</tr>
<tr>
<td></td>
<td>Control power</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weight factors $w_{ij}^{cp}$ for $m_{p}$ variables</td>
<td>$m_{p} \cdot n_{FB}$</td>
</tr>
<tr>
<td></td>
<td>Bias term $b_{ij}^{cp}$</td>
<td>$n_{FB}$</td>
</tr>
<tr>
<td></td>
<td>Cumulative rate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Window sizes $n_{R}$</td>
<td>$n_{FB}$</td>
</tr>
<tr>
<td>Intelligent indices</td>
<td>Trend index</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Window sizes $n_L$ and $n_S$</td>
<td>$2 \cdot m_{T}$</td>
</tr>
<tr>
<td></td>
<td>Weight factors $w_{T1}$ and $w_{T2}$</td>
<td>$2 \cdot m_{T}$</td>
</tr>
<tr>
<td></td>
<td>Condition index</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weight factors $w_j$</td>
<td>$m_{CM}$</td>
</tr>
<tr>
<td></td>
<td>Stress index</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weight factors $w_j$</td>
<td>$m_{S}$</td>
</tr>
</tbody>
</table>

stress, condition, quality and health indices can be represented as weighted sums of scaled features. Membership definitions are needed for these features (Tab. 4).

The maximum number of parameters listed in Tables 4 and 5 is never realised since all the manipulated variables do not have all the features described above and all the scaling functions do not need to be nonlinear. Defaults can be used many parameters. For example, the multilevel LE controller of the solar collector field can have 34 parameters, but the basic nonlinear controller requires only 14 parameters (Juuso 2006).

Modules of the MIMO LE controller are selected in high-level control to come up with application requirements. The basic controller with adaptative scaling is a feasible solution for many applications. Feedforward control is included if the process changes are not detected in time in the feedback control. During the operation, a full control to the existing control strategies can be taken by the weight factors and coefficients shown in Table 6. Different adaptive scaling actions can be weighted for each manipulating variable. The working point level $wp_{min}$ can be used for defining effective set points with cascade controllers based on (51). The working point limit is temporarily increased when strong fluctuations are detected. Even higher working points can be used. The PBA action and constrain handling are used if fast corrections for large changes are needed. The ASA action is used only for fine-control to the set point. The actions, whose strengths are defined by $(c_B)_j \in [0, 1]$ and $(c_A)_j \in [0, 1]$, respectively, are not
used if the operating conditions are strongly fluctuating. This can be done automatically by using fluctuation indices: all \((c_B)_j\) and \((c_A)_j\) are set to zero for the problematic time period.

**Table 6. Coefficients of a high-level LE controller, notations are shown in Table 4.**

<table>
<thead>
<tr>
<th>Action</th>
<th>Definitions and parameters</th>
<th>Maximum number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive scaling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>working point</td>
<td>Weight factor (w_{iP}^w)</td>
<td>(n_{FB})</td>
</tr>
<tr>
<td>control power</td>
<td>Weight factor (w_{iP}^T)</td>
<td>(n_{FB})</td>
</tr>
<tr>
<td>cumulative rate</td>
<td>Weight factor (w_{iP}^r)</td>
<td>(n_{FB})</td>
</tr>
<tr>
<td>Minimum working point</td>
<td>(w_{Pmin})</td>
<td>1</td>
</tr>
<tr>
<td>Predictive braking</td>
<td>Coefficient ((c_B)_j)</td>
<td>(m_C)</td>
</tr>
<tr>
<td>Asymmetry</td>
<td>Coefficient ((c_A)_j)</td>
<td>(m_C)</td>
</tr>
<tr>
<td>Feedback</td>
<td>weight (w_{ij}^F)</td>
<td>(n_{FB} \cdot m_C)</td>
</tr>
<tr>
<td>Constraint handling</td>
<td>Scale coefficients (\lambda_i^-) and (\lambda_i^+)</td>
<td>(2 \cdot n_{FB})</td>
</tr>
<tr>
<td></td>
<td>High limit (u_{iP}^H)</td>
<td>(n_M)</td>
</tr>
<tr>
<td></td>
<td>Low limit (u_{iP}^L)</td>
<td>(n_M)</td>
</tr>
<tr>
<td>High-level control</td>
<td>High limit (\Delta u_{iP}^H)</td>
<td>(n_{FB})</td>
</tr>
<tr>
<td></td>
<td>Low limit (\Delta u_{iP}^L)</td>
<td>(n_{FB})</td>
</tr>
<tr>
<td></td>
<td>Weight factor (s_i)</td>
<td>(n_{FB})</td>
</tr>
<tr>
<td></td>
<td>Bias correction (u_{BC}^i)</td>
<td>(n_M)</td>
</tr>
</tbody>
</table>

The effects of different controlled variables \(j\) can be weighted by using (49) and (50) for each manipulating variable \(i\). The weights can be changed in high-level control. The high and low limits used in constraint handling and high-level control are related to the membership definitions of the variables and the change of control, respectively. Adaptive scaling should keep the control action in a safe range. Scale coefficients \(\alpha_i\) and \(\beta_i\) and limits \(\{u_{iP}^L, u_{iP}^H, \Delta u_{iP}^L, \Delta u_{iP}^H\}\) can be set in the high-level control to safeguard the operation. Possibilities to scale the feedback actions with \(s_i \in [0.5, 1.5]\) and add manual control actions \(u_{BC}^i\) are important for anomalous situations, which are not taken into account in control design.

**Smart adaptive control** combines control modules which are activated smoothly by using different intelligent indices: the coefficients of the high-level control (Tab. 6) limit the effects of the special actions. Models are included in the feedforward control, adaptive scaling and model-based control. All these models are adapted in the same
way as any other LE model (Fig. 20). The control related membership definitions listed in Figure 17 and the controller coefficients (Tab. 5) can be tuned with simulation experiments. The operating areas and the control and model surfaces depend on both the scaling functions and the coefficients.

*Genetic tuning* of LE controllers operates in the same way as for LE models, which are fully parameterised. The penalties used in (Juuso 2006) are not needed when the approach introduced in (Juuso 2009e) is used: the parameters of the scaling functions have a limited search are (Tab. 2), which fill the constraints of the shape factors \( \alpha_j^- \) and \( \alpha_j^+ \). Since the parameters of all membership definitions (Tab. 4) and the coefficients of the controllers and intelligent analysers (Tab. 5) are handled in the same way, the process models, intelligent analysers and controllers can be tuned together with simulation experiments. The wide operating area of the LE control system must be taken into account in the adaptation.

Other controllers may contain LE modules. The *on-line adaptation* of the models and controllers is based on linear methodologies enhanced with the recursive adaptation of the scaling functions. However, the complex applications use predefined adaptation and advanced model-based control. *Switching* is done smoothly with an *event-based* procedure, which combines modules developed for specific situations. Some parts of the adaptation schemes of the LE controllers are identical to the gain scheduling. A LE controller can represent a controller with a monotonously increasing or decreasing control surface, including

- a FLC with a symmetric rule base,
- a sliding mode controller, where several sliding modes can be combined without loosing the monotonous change,
- a neural controller,
- a LE-based SOC,
- LE-based IMC and MPC are reserved for more complex variable interactions.

In addition, inverted model FF controllers can be introduced the existing model can be transformed to a single LE model, since all such LE models are invertible.
6 Linguistic equations in hybrid systems

Integration of various intelligent techniques and statistical analysis is needed for development and tuning of smart adaptive systems for practical applications. The linguistic equation approach provides an efficient modelling environment and many tools for building intelligent systems for modelling and control. Other intelligent methods are used in the development and tuning but they can also extend the application area of the LE methodology through a multimodel approach. This chapter presents linguistic equations in connection with other intelligent methods to evaluate their possible contributions to the smart adaptive systems. Also the links to the signal processing, feature extraction, advanced statistical analysis and classification are discussed. In the applications, hybrid systems are needed especially in fault diagnosis and performance monitoring.

The knowledge-based LE systems, the principles of multimodel systems and case-based diagnostics were developed during the phases I-III (Juuso 2004e). The later extensions, including recursive tuning, fuzzy LE calculus, intelligent indices, the constraint handling and the presentation of knowledge-based information, have changed the hybrid LE systems drastically (Juuso 2007a,b, 2008, 2009a,e, Juuso et al. 2009, Juuso 2011a, Juuso & Lahdelma 2011b, Juuso 2012a). The current approach integrates all these into the recursive tuning of parameters.

6.1 Linguistic equations and statistical analysis

Statistical analysis is an essential part of the development and tuning of the LE systems: the data-driven development of the scaling functions, which is based on advanced generalised norms and moments, is suitable for different statistical distributions. Both the scaling functions and the linear equation systems can be recursively adapted. Well-known performance measures are calculated by using the scaled values. Several LE models, which normally include from two to five inputs, can be combined by using the principles of composite local, linear parameter varying and piecewise affine modelling. The compact parametric approach is beneficial for all these model types.

Hierarchical, K-means or robust clustering of the working point variables is suitable for finding specific modelling areas. An optimal number of clusters is important for
successful modelling. Different geometrical shapes of the clusters are introduced through the nonlinear scaling. Principal components are used for variable selection.

The linear models can be extended with more complicated structures, e.g. RSM models in steady-state and parametric models in dynamic models, but robust methods are needed to avoid overfitting. Modelling methodologies are embedded in the design of experiments. The nonlinear scaling defined by a parametric approach also provides a good basis for this extension.

6.2 Linguistic equations and fuzzy logic

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.

6.2.1 Linguistic equations and fuzzy set systems

The membership definitions, which extend the idea of membership functions, form a close connection between fuzzy set systems and LE models (Fig. 24). Triangular membership functions generated from the membership definitions have consistent meanings for different variables (Fig. 12). In the early applications, the scaling approach was used for generating membership functions for fuzzy modelling and control. Fuzzy modelling approaches provide channels for combining different sources of information. The FuzzEqu Toolbox (Juuso 2000b) includes routines for building a single LE system from large fuzzy systems including various rule blocks implemented in FuzzyCon or Matlab FuzzyLogic Toolbox.

Knowledge-based LE modelling and control can be developed from existing fuzzy set systems with the data-driven approach (Fig. 20) after replacing the linguistic labels with the linguistic values in the range $[-2, 2]$. As these linguistic values correspond to the centre points of the membership functions they are denoted as membership locations. Usually equidistant locations are used as a starting point, e.g. negative large, negative small, zero, positive small, and positive large with numbers -2, -1, 0, 1 and 2 (Fig. 9). The membership definitions are generated from adjusted locations and the centre points
of the corresponding membership functions (Fig. 24). This approach has been used for representing fuzzy controllers, e.g. a feedback controller with five labels and a feedforward controller with seven labels (Juuso 1999a). The resulting LE controllers (Figs. 22 and 21) have much smoother operation. Additional points are generated for very small systems, which have too few relations for the parameter based approach. Fuzzy relational models encode several associations between the linguistic terms of the input and output domains. The different strengths of these associations provide a good basis for the development of fuzzy LE models. (Juuso 2004e)

![Diagram](image)

**Fig 24. Fuzzy set systems and linguistic equations (Juuso 2004e, published by permission of Elsevier).**

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 (Fig. 24). For diagnostical purposes, the coarse partition with five terms is quite adequate (Figure 9). To use the same methodology in control applications, a finer fuzzy partition is employed (Juuso & Leiviskä 1994). If a non-complete set is satisfactory, a part of the rules can be deleted already before tuning. The best similarity with the LE controller is achieved if the positions of all the fuzzy sets are coordinated with the real valued linguistic equations.

Fuzzy rules can be generated in two ways: the simultaneous approach produces a compact rule set, and the sequential approach a set which is easier to interpret and prune. Singleton models represent the LE model quite accurately if the linear surfaces are in appropriate areas. TS fuzzy models have the same requirement, but the locations of the membership functions are different. Linguistic fuzzy models are developed from
singleton models. Fuzzy relational models developed from LE systems will have fairly few nonzero elements since nonlinearities are included in the membership functions. The labels of the output can be pruned by clustering the membership locations to obtain an appropriate fuzzy partition. Hierarchical sets of membership functions developed from the membership definitions introduced a gradually refining decision making methodology presented in (Juuso et al. 1993).

Principles of the knowledge-based LE systems and the LE-based development of fuzzy systems were already introduced during the phases I-III (Juuso 2004e). In the approach introduced in (Juuso 2009e), the monotonous and increasing scaling functions are achieved by requiring that the shape factors $\alpha_j^-$ and $\alpha_j^+$ are in the range $[\frac{1}{3}, 3]$. Knowledge-based information obtained from natural language is translated to the same value range $[-2, 2]$ with the indices and indicators calculated from numerical values. The fuzzy labels are extracted from linguistic information given by humans. Therefore, the resulting fuzzy numbers may contain vagueness and uncertainty. To take this into account, the membership functions are made sharper or wider with modifiers 'extremely', 'very', 'more or less' and 'roughly'. The powering modifier should reflect the reliability of the information, or of the source of information. Several linguistic terms can be combined or further modified with 'and', 'or' and 'not' (Juuso 2012a).

Correspondingly, numeric values can be presented by linguistic labels generated by membership definitions (Juuso & Lahdelma 2011b).

### 6.2.2 Fuzzy set systems in LE models

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. The explanation of the operation suits well fuzzy set systems, also when the calculations are done with LE systems. Labels of all the variables are related to the range $[-2, 2]$, and the partition can be chosen freely. The linguistic meanings of the labels can be generated from the scaled values in a straightforward way. Domain expertise can be included by using these membership functions.
**Multimodel approach**

A multimodel approach based on fuzzy LE models has been developed for combining specialised submodels (Juuso 1999b). The approach is aimed for systems that cannot be sufficiently described with a single set of membership definitions because of very strong nonlinearities. Additional properties can be achieved since the equations and delays can be different in different submodels. In the multimodel approach, the working area defined by a separate working point model. The submodels are developed by the case-based modelling approach (Section 4.1.4). The multimodel system has several submodels and a fuzzy decision system for selecting a good model for each situation by using several working point variables as inputs. If several inputs are combined into a single working point index, the fuzzy set systems is reduced to a fuzzification block (Fig. 25).

![Multimodel LE system with a fuzzy decision module](image)

**Fig 25. A multimodel LE system with a fuzzy decision module (Juuso 2008a).**

Fuzzy models for special situations are useful additions to compact LE models, which suit well for more or less smooth input–output dependencies. The Fuzzy–ROSA method (FRM) serves for a data–based rule generation to model a given input–output dependency and is efficient for modeling complicated local nonlinear structures. These properties are combined in a hybrid data–based modeling concept which is applied to dynamic simulation of a solar power plant (Fig. 26). The performance of the simulator is considerably enhanced with this concept, and the hybrid simulator can be used in the control design. (Juuso et al. 2000)

The submodels are developed for different operating conditions with fuzziness $\varepsilon_i$ defined by (40) as a fitness measure. A hierarchical LE clustering introduced for this purpose is based on a gradually refining set of LE models. The first model is considered
normal and it is valid for the part of the data set where the fuzziness is low. The areas of high positive and high negative fuzziness are then considered to form two new cases. Additional models are developed if there is training material available.

The LTS models introduced in (Juuso 2009a) are multimodels with one limitation: the fuzzy partition is defined with the same variables as the models. As LE models are nonlinear, also these local models are nonlinear: nonlinear scaling is used for both inputs and outputs. LTS models are initialised by clustering the scaled values. Special smoothing algorithms, which are commonly used in TS models, are not needed in LTS models. Genetic tuning presented in Section 4.3 is preferable since the process insight is in many cases destroyed by the ANFIS tuning.

![Cascaded modelling (left) and resulting model (right) (Juuso et al. 2000).](image)

**Fuzzy calculus in LE models**

Fuzzy numbers can be handled in LE models by the extension principle (Juuso 2007a,b, 2008). LE models are extended to fuzzy inputs with this approach if the membership definitions, i.e. functions $f_j^-$ and $f_j^+$ defined by (18) and the corresponding inverse functions, are replaced by the corresponding extensions of these functions. Square root functions are used in the linguistification part (Fig. 18(a)). The extension principle is needed for fuzzy inputs (Tab. 7). 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 $\alpha$ levels should increase with the growing fuzziness of the input.

The argument of the function $f_{out}$ in (35) is obtained by fuzzy arithmetic. Only the addition and subtraction are needed if the interaction coefficients are crisp. Fuzzy coefficients are obtained if LE models are generated for several samples selected by
resampling, either singletons or fuzzy numbers (page 85) can be used. Results of the fuzzy LE models with fuzzy inputs can be constructed by using fuzzy multiplication and division as well. The fuzzy extension of the classical interval analysis (Moore 1966) suits very well to these calculations. Results of the fuzzy interval analysis always have maximal uncertainty as it takes the worst case. Negative associations between the input variables reduce the uncertainty considerably. In the calculations, this can be taken into account by using specific membership functions for the upper and lower parts of the value range.

Finally, the delinguistication block uses second order polynomials defined by (18). The extension principle is needed if at least one input or equation is fuzzy (Tab. 7). The fuzzy output can be defuzzified to obtain a crisp output. The uncertainty presented by type-2 fuzzy numbers is an important part of the inputs and the models, not just additional tuning parameters of the system. Since the type reduction (page 39) is not used, the output is still fuzzy in these cases.

<table>
<thead>
<tr>
<th>Model</th>
<th>Input</th>
<th>Linguistification</th>
<th>Equations</th>
<th>Delinguistification</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crisp</td>
<td>Crisp</td>
<td>Crisp</td>
<td>Crisp</td>
<td>Crisp</td>
<td>Crisp</td>
</tr>
<tr>
<td>Fuzzy</td>
<td>Fuzzy extension</td>
<td>Fuzzy + and -</td>
<td>Fuzzy extension</td>
<td>Fuzzy</td>
<td></td>
</tr>
<tr>
<td>Crisp</td>
<td>Crisp</td>
<td>Crisp</td>
<td>Fuzzy + and -</td>
<td>Fuzzy extension</td>
<td>Fuzzy</td>
</tr>
<tr>
<td>Fuzzy</td>
<td>Fuzzy extension</td>
<td>Fuzzy +, -, x and /</td>
<td>Fuzzy extension</td>
<td>Fuzzy</td>
<td></td>
</tr>
</tbody>
</table>

**Fuzzy logic in recursive parameter estimation**

The decomposition alternative is taken into consideration in the recursive parameter estimation of the membership definitions. The centre point, the corner points and the outlier limits are calculated for the recent data points define a membership function by using the date-driven approach presented in Chapter 3 and variable specific time windows. The resulting membership function $F_r$ is compared with the current feasible range $F_0$ to obtain the degrees of membership for three inequalities: $F_r < F_0$, $F_r = F_0$ and $F_r > F_0$. The calculation is done as a similarity analysis with the upper and lower complements of $F$: $F_r = F^-$, $F_r = F$ and $F_r = F^+$. A high degree of membership for
either \( F_i < F \) or \( F_i > F \) mean that parameters for a new case are started to extract. A sufficient length is required for introducing a new case: short periods are considered as outliers or anomalies. The time windows and the limits for the degrees of membership are based on expertise.

6.2.3 **Fuzzy logic in LE-based diagnostics**

Case-based analysis was the main approach in the LE-based diagnostics during the phases I-III (Juuso 2004e). New intelligent indices presented in Section 5.5 have moved the diagnostics to another level.

**Case-based analysis**

The main application area of the case-based LE models is the detection of operating conditions. The fuzziness \( \varepsilon_i \) defined by (40) is the fitness measure of the LE model. The range of the fuzziness can be different for different equations: narrow range means a sensitive fact and a wide range an unsensitive one. These levels are related to the uncertainty of the fact: the range defines the degree of the membership for the fact.

For the detection of operating conditions, a sensitive fact should have a higher weight than the less sensitive ones. Therefore, the weights of the equations are defined on the basis of the ranges of the fuzziness. The mapping from the fuzziness to the weight is chosen case by case. The interaction coefficients and bias terms are normalised for the comparison of equations. The fuzziness \( \varepsilon_i \) can be assessed as a feasible range (Fig. 9) of the fact related to the equation: the core corresponds to the normal area related to the fact, and the other support area produces warnings and alarms.

The positive and negative values of fuzziness have different meaning as they can be considered as an additional variable with positive and negative value, correspondingly. In the model (30), this indirect measurement can be represented as three levels \(-1, 0, \text{ and } 1\), which have specific degrees of membership \( \mu_{-1}, \mu_0, \text{ and } \mu_1 \), respectively (Juuso 1999a). The fuzziness increases for example if all the variables are not included to the model. The fact that experts do not always agree with interactions can also be taken into account by using several interaction matrices with different coefficient values. On the other hand, the directions of interactions can depend on the working area in nonlinear systems. In these cases, different interaction matrices have different degrees
of membership. Different combined effects can be taken into account as well (Juuso 1999a).

Also fuzzy rule-based systems can be developed by dividing the feasible range into membership functions, e.g. large negative, small negative, zero, small positive and large positive deviation for the fact corresponding to the membership functions shown in Figure 9.

**Intelligent indices**

All the scaled variables and intelligent indices presented in Section 5.5 can be used in fuzzy diagnostics by choosing appropriate membership functions, see Figure 9. There are different levels of complexity: only a single index is needed in some diagnostics, a linear combination of two, three or more indices solve some problems, or the indices are antecedents of a fuzzy rule base. Anomaly detection can be based on intelligent indices, since an unknown fault causes a clear decrease of the condition index, \( (I_C(k))_0 \ll 2 \), without increasing any fault index \( (I_C(k))_L, L = 1, \ldots, n_{ca} \). Linear membership functions based on these indices can be used in fuzzy fault and anomaly detection.

In the LE-based trend analysis, the features of dynamic behaviour are handled by scaling with membership definitions; fuzzy reasoning methods are transformed into linguistic equations. This approach, which is integrated to the LECont control concept implemented in G2 (Juuso 2004e), is the basis of the intelligent trend analysis introduced in (Juuso et al. 2009, Juuso 2011a). Intelligent indices (78) and (79) are developed for single measurements. A degree of membership can be obtained for the trend episodes by introducing fuzzy limits to the areas in Figure 23. The indices based on case-based analysis include fuzzy reasoning, see (83) and (86).

### 6.3 Other methodologies in LE systems

Hybrid systems combine phenomenological modelling, data-driven modelling and computational intelligence (Fig. 3): the structures are taken from parametric models or neural networks, and the systems are tuned e.g. with genetic algorithms. Integration with other methodologies are also important in signal processing and classification.
Signal processing and feature extraction

Signal processing is a wide research application area, which has high potential for the nonlinear scaling and LE systems. Noise reduction with filtering is widely used, but the fluctuations may also include useful information. The LE systems can for example use following methodologies:

- The membership function \( F_r \) obtained for the recent data in recursive parameter estimation can be used in filtering: the method based on defuzzification (Murtovaara et al. 1996, Juuso 2005b) is extended by using generalised norms and moments. The nonlinear scaling extends the applicability of linear systems in modelling and control. This can be beneficial in many filtering methods, which use embedded linear models, e.g. Kalman and Wiener filters and also adaptive linear filters.
- Subspace methods and blind source can precede the LE-based analysis to find suitable signal subsets, which can then be analysed separately. Interactive compact LE systems reduce the need of the LPCA or the advanced versions of PCA.
- Specific frequency ranges can be emphasised with real-order derivation is already used in condition monitoring application to produce more informative signals in time domain (Lahdelma & Juuso 2011a,b).
- Wavelet analysis, multiresolution approaches and frequency domain analysis can be a source for informative features. Spectral features, e.g. spectral kurtosis, are obtained for specific frequency ranges.

Semi-mechanistic models

In semi-mechanistic models, the approximation rules of the model parameters are combined into linguistic equations. The approximation rules can be based on qualitative knowledge. Uncertainty can be taken into account in deterministic systems by using discretised membership functions for each input variable \( x_j \). Each result has a degree of membership which is calculated by T-norms, e.g. minimum. Several levels defined by (28) are used. The output can be defuzzified as a weighted average, and uncertainty of the result is presented as a fuzzy number. A gradually refining set of discretized values was used in (Juuso et al. 1993). Nonlinear effects are better captured if the points are selected from scaled input values (Section 3.1.2). In dynamic models, the mechanistic models take care of the basic level dynamic simulation. This approach extends the
operating area of the mechanistic models. The dynamic solution of the overall model, i.e. integration, is done in the mechanistic model.

**Neural networks**

Adaptation of the scaling functions with linear neural networks, which was already used in early LE applications (Juuso 1999a), suits for recursive tuning (Juuso 2004e). The nonlinear scaling with the membership definitions provides an extension to neural networks, i.e. the scaling replaces the normalisation. Also linguistic principal components (LPC) can be used as inputs. **Linguistic neural models** are extensions of LE models: the linear equations are replaced by networks and weight factors. The response of each neuron is obtained by an activation function $F_i$:

$$a_i = F_i\left(\sum_{j=1}^{n} w_{ij}^{NN} f_j^{-1}(x_j) + b_i^{NN}\right),$$

(89)

where $w_{ij}^{NN}$ is the weight factor between the neurons $j$ and $i$, $x_j$ is an element $j$ of the input vector, and $f_j^{-1}$ the corresponding membership definition. A linear linguistic network is identical to the basic LE model. Also perceptron networks are extended to detecting borders between nonlinear areas by the membership definitions.

Nonlinear multivariable models can be developed by using SOM as a clustering method, and generating LE models from these neurons (Juuso & Juuso 1999). The resulting LE model, which can be used in any direction, represents the SOM network very accurately. The LE model can be considered as a new type of neural network, linguistic SOM network (LSOM), where each neuron weight also has a linguistic meaning. This network can be adapted to changing conditions, e.g. different regions and industry, by adjusting membership definitions. This approach provides tools for domain experts to adapt the networks to different operating conditions without reconstructing the network.

Linguistic radial basis functions (LRBFs) provide asymmetrical neurons, or clusters, in the real range since the inputs are scaled. Fuzziness $\varepsilon_i$ defined by (40) can be considered as a LRBF. The use of the linear layer in LVQ is also a reasonable way if the competitive layer uses linguistic values. Fuzzy inputs can be handled in (89) by the extension principle for the functions $f_j^{-1}$, fuzzy arithmetic and the extension principle for the function $F_i$. **Linguistic neurofuzzy systems (LNFS)** provide additional, more complex, structures for LE models.
**Classification**

In the LE systems, classification is based on intelligent indices, which are calculated from the different features or residuals $\varepsilon_i$ defined by (40). Since each LPC is defined by a LE model, the corresponding residual can also be used for classification. The multidimensional applications need various other methodologies: rule-based systems are enhanced with fuzzy logic based on approximate reasoning with $T$-norms and $S$-norms; artificial neural networks include several alternatives; support vector machines focus on the boundaries between cases; directed graphs analyse causalities; fault and event trees combine various symptoms; Bayesian networks use conditional probabilities; hidden Markov models are based on stochastic processes. CBR combines different methodologies with previous experience. Outliers, novelties, quality control, anomalies, faults and weighting of the submodels are special classification tasks. The complex methodologies are raised to the higher levels, which combine the compact low level LE solutions.

**Genetic algorithms**

Adaptation of LE models and controllers with genetic algorithms (Juuso 2004e) has already been used in early applications since this approach can handle a large number of parameters in a wide search space. The new constraint handling (Juuso 2009e) integrates the adaptation of the scaling functions and the models. Fuzzy set systems, neural networks and mechanistic models introduce additional parameters to the overall system. The models, soft sensors, controllers and diagnostic tools are tuned with the same methodologies.
7 Application examples of linguistic equations

The linguistic equation (LE) approach is aimed for nonlinear multivariate modelling by using nonlinear scaling (Chapter 3) and linear interactions (Chapter 4). The approach can be used in many ways in control (Chapter 5), and for fault diagnosis it provides informative indices and facts (Chapter 6). Adaptive nonlinear scaling is a new solution to development smart adaptive systems (Section 7.1). Earlier multivariable modelling applications are discussed in Section 7.2 to explain the multitude of modelling structures: the adaptive scaling approach is now coming to this type of applications. The main principles of the LE-based control are extended with several intelligent analysers, which are used in recent control tests (Section 7.3). Scheduling and decision making also have a long history, which is now extended to performance analysis (Section 7.5). Detection of operating conditions, which is essential in fault diagnosis and performance monitoring, is currently based on intelligent indices: the earlier case-based solutions are moving to higher levels (Section 7.4).

7.1 Adaptive nonlinear scaling

The new adaptation methods have been tested in the Acurex Solar Collectors Field of the Plataforma Solar de Almeria (PSA) located in the desert of Tabernas (Almeria), in the south of Spain. The Acurex field (Fig. 27) supply thermal energy (1 MW) in form of hot oil to an electricity generation system or a Multi–Effect Desalination Plant (Zarza 1995). The energy is collected with the oil, which is circulated by pumping first in the field (start-up) and then through the storage tank when the outlet temperature is high enough. The energy is taken into use from the storage tank or in the tests, the oil is cooled as shown in Figure 27. The operating modes are changed by opening and closing valves.

The analysis of outliers and anomalies is based on the previous knowledge or data. Measurements of over 65 hours from 12 days are here analysed by using the scaling functions generated from the first 5 hours 25 minutes. Most irradiation measurements of the first 28.5 hours are accepted in the range and step analysis (Fig. 28(a)). In this analysis, the cloudy periods from 5.4 to 11.4 hours would be outliers, but the measurements are correct. After the normal operation, these periods could be considered as anomalies. Since the cloudy periods can be used for energy collections, these periods novel operating conditions, which need to be detected to adapt the operation, see Section
7.3.2. Very short disturbances are taken as outliers since they do not have any effect on operation.

![Diagram of Acurex solar collector field](image)

Fig 27. Layout of the Acurex solar collector field including three operating modes: start-up, energy collection, and cooling.

The irradiation is on much lower level after the cloudy period (28.5 - 37.2 hours). If the limits are based on the first day measurements, the whole period (37.2 - 59.4 hours) is at least in suspicious area. However, most measurements are acceptable, and the scaling function and the limits need to be updated with the data of the new operating point. This is an example of changes which trigger the recursive tuning of the scaling functions. There are shorter cloudy periods and few outliers in this time period. During the last day, the field is getting towards the high irradiation conditions, where only disturbances are a single outlier and a short cloudy period (Fig. 28(a)).

In the recursive adaptation, the membership function $F_r$ calculated for a working day becomes smaller than the previously defined feasible range $F$ after Day 8. Since the definitions of the irradiation should cover all the operating areas, a new case is not formed. The membership definitions are updated by including the new measurements. During the following three days, the values are in the new feasible range and the membership definitions continue to move to lower values. Again during the last day, the irradiation is within the feasible range. The cloudy periods and outliers are not used in this analysis. Several operating days are needed to extract widely applicable membership definitions.
The inlet temperature increases strongly during each operating day (Fig. 28(b)). Drastic changes, which result from load disturbances, are seen during the first and the last day. The analysis, which is based on the first 4.25 hours, needs to be updated with the period from 10.2 to 37.5 hours. This also widens the upper limits. The highest temperatures are reached on the last day. The inlet temperature drops in some start-ups, but values are within the support area. On Day 8, the operation is started from the ambient temperature. The inlet temperatures are kept fairly low during the following three days to adapt the operation to lower irradiation.

Fig 28. Outliers and anomalies in solar collector data: measurements with upper and lower warning (−−) and alarm (-) limits.
The variable specific analysis is extended to consistency checking when the outlet temperatures of the collector loops are compared. There are some differences after drastic changes, but a real anomaly means that a loop drops out from the operation, i.e. the outlet temperature of one loop is not rising in the same way as in the other loops.

Three operating areas, high, low and cloudy, are detected in the irradiation measurements shown in Figure 28(a). The irradiation is used together with the inlet temperature in the working point control of the solar collector field. The working point varies in a wide range during each day: the start-up is very different from the rest of the day. Generalised statistical process control with new limits reveals the changes of operating conditions, e.g. changes shown in Figure 28.

The new scaling approach provides a systematic data-driven approach for defining membership functions. In this analysis, the feasible range is represented as a type-2 fuzzy number if the corner points are defined by intervals. Since the location $c_j$ is allowed to be a fuzzy number, all the membership functions generated from the membership definitions can be type-2 fuzzy numbers. In the solar collector case, the parameters of the scaling function of the irradiation are clearly different for the high, low and cloudy periods (Fig. 28(a)). The combined feasible ranges and membership functions are type-2 fuzzy numbers, but they can be combined into normal fuzzy numbers.

7.2 Multivariate modelling and simulation

Nonlinear scaling forms the basis for the LE modelling, i.e. linear models can be used for the scaled variables. An important benefit of the linear approach is that the models can be inverted, technically to any direction. The compact basic solution makes extensions to dynamic and case-based systems possible. Complex models for steady-state and dynamic systems (Tab. 8) can be built with the cascade and interactive structures (Fig. 5). The sets of equations can also be used in any sequence. Extensions to multimodels and uncertainty handling require fuzzy logic. The structures of the LE models developed during the periods I-III are useful for for smart adaptive applications. New models are developed for fatigue, condition and wastewater treatment. Uncertainty is taken into account in batch cooking, granulation, fermentation and wastewater treatment.
7.2.1 Steady-state LE modelling

Linear regression and linear networks are used model development. In principle the equations could also be nonlinear, but only linear equations have been needed in the applications so far. Other model structures, including RSM models with interactive and quadratic terms, ANN models with hidden layers, TS fuzzy models and fuzzy relational models, are used with the scaled data to test if the linear LE model sufficiently captures the behaviour of the system.

Table 8. Application examples of LE models.

<table>
<thead>
<tr>
<th>Case</th>
<th>Model type</th>
<th>Application</th>
<th>Special</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electric furnace</td>
<td>steady-state</td>
<td>process design</td>
<td>interactions</td>
</tr>
<tr>
<td>Lime kiln</td>
<td>steady-state</td>
<td>feedforward control</td>
<td></td>
</tr>
<tr>
<td>Solar collector field</td>
<td>steady-state</td>
<td>fuel quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dynamic</td>
<td>controller tuning</td>
<td></td>
</tr>
<tr>
<td>Solar collector field</td>
<td>dynamic (ODE)</td>
<td>adaptation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dynamic (PDE)</td>
<td>controller tuning</td>
<td>working point</td>
</tr>
<tr>
<td>Continuous cooking</td>
<td>steady-state</td>
<td>forecasting</td>
<td>Kappa</td>
</tr>
<tr>
<td>Fatigue</td>
<td>steady-state</td>
<td>stress contribution</td>
<td>S-N curves</td>
</tr>
<tr>
<td></td>
<td>dynamic</td>
<td>forecasting</td>
<td>risk</td>
</tr>
<tr>
<td>Gas furnace</td>
<td>dynamic</td>
<td>modelling</td>
<td>tuning</td>
</tr>
<tr>
<td>Water treatment</td>
<td>steady-state</td>
<td>feedforward control</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dynamic</td>
<td>water quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>controller tuning</td>
<td></td>
</tr>
<tr>
<td>Batch cooking</td>
<td>dynamic</td>
<td>on-line forecasting</td>
<td>interactions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>samples</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>uncertainty</td>
</tr>
<tr>
<td>Granulation</td>
<td>dynamic</td>
<td>forecasting</td>
<td>interactions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>samples</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>uncertainty</td>
</tr>
<tr>
<td>Fed-batch fermentation</td>
<td>dynamic</td>
<td>on-line forecasting</td>
<td>phases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>interactions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>uncertainty</td>
</tr>
<tr>
<td>Wastewater treatment</td>
<td>steady-state</td>
<td>forecasting</td>
<td>interactions</td>
</tr>
<tr>
<td></td>
<td>dynamic</td>
<td>operating</td>
<td>trends</td>
</tr>
<tr>
<td></td>
<td></td>
<td>conditions</td>
<td>uncertainty</td>
</tr>
<tr>
<td>Condition monitoring</td>
<td>dynamic</td>
<td>prognostics</td>
<td>condition</td>
</tr>
</tbody>
</table>

Nonlinear scaling also extends the applicability of some data pre-processing methods: the linear versions can be used for correlation analysis, principal component analysis and
clustering. Using effective time delays already provides a basis for simple forecasting models (Tab. 8). Outliers are removed in model development after taking into account the time delays between the variables, and only those points, where all the corresponding values are acceptable, are used in modelling, i.e. imputation is not used.

Electric furnace

The first LE system was developed for designing submerged arc furnaces used in the production of ferroalloys. The interaction matrix $A$ was based on five well known relations, including four three variable relations and one five variable relation. The vector $\vec{X}$ includes the scaled values of ten variables: useful power, voltage, current and resistance are used together with the calculation of the resistance, which depends on the average conductivity and the immersion depth of the electrode; the average conductivity depends on four variables: the conductivities of the slag and the charge, the immersion depth and the electrode current. The conductivity of the slag depends on the basicity and the temperature. (Juuso & Leiviskä 1992)

Lime kiln

A steady-state LE model was developed in an early control application from the process measurements of a lime kiln. Different operating conditions are represented by a steady-state LE model $A \cdot \vec{X} = 0$, where

$$A = \begin{bmatrix} -1 & 3 \\ -3 \\ \end{bmatrix},$$

and $\vec{X}$ includes the scaled values of variables: the lime feed, the fuel feed and the hot end temperature. This LE model provided better results than a fuzzy TS model since the weaknesses of the data were clearly seen and corrected. (Juuso et al. 1997a)

Solar thermal power plant

A steady-state LE model is used as a working point model in the modelling and control of a solar power plant. The model is based on experiments carried out in the solar collector field shown in Figure 27. The calculations are based on the energy balance $\text{energy stored} = \text{irradiance} - \text{energy transferred} - \text{heat loss}$. During the start-up the volumetric heat capacity increases very fast in the start-up stage but later remains almost
constant because the normal operating temperature range is fairly narrow. This nonlinear effect is handled with the working point LE model developed for the temperature difference, $T_{\text{diff}} = T_{\text{out}} - T_{\text{in}}$, ambient temperature, $T_{\text{amb}}$ and effective irradiation, $I_{\text{eff}}$. The model has a quite high correlation to the real process data, since the working point variables already define the overall normal behaviour of the solar collector field. The differences have a clear relation to operating conditions, e.g. oscillatory behaviour is a problem when the temperature difference too high for the effective irradiation level. Variable time delay (Fig. 18(b)) is taken into account in steady-state modelling.

The working model is an essential part of the LE control in the adaptation of parameters (Juuso et al. 1997b, 1998b), in limiting the setpoints (Juuso & Valenzuela 2003) and currently also in the model-based control (Juuso 2012b, c).

**Continuous cooking**

For continuous cooking, a LE model has been developed for predicting the Kappa number, which is widely used quality variable. The model is based on the concentrations of alkali, total dissolved solids and lignin measured on-line with CLA 2000 Cooking Liquor Analyser developed by ABB. As the digester process is far from a linear and simple input-output system, the analysis must be nonlinear. Different approaches have been used for the mathematical modelling of the cooking result (Leiviskä et al. 2001): fuzzy logic, PLS, ANN and LE. In modelling the Kappa number in a continuous digester, all these methods seem to learn the process behaviour in a similar manner, but the LE models are the best in process environment since they can be adapted to various operating conditions in an understandable way. ANN and PLS models are sensitive to changes in process conditions, and fuzzy models need a large number of membership functions and rules that are too time-consuming to adapt.

**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

$$I_S(k) = \log_{10}(N_C(k)), \quad (91)$$

151
where the stress index $I_S(k)$ can be a scaled value of stress or a scaled value of a generalised norm obtained from vibration signals. 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 system may also contain several specific equations corresponding to different operating points, e.g. low, normal and high stress. (Juuso & Lahdelma 2012) The dynamic LE model is used for fatigue prediction (Juuso & Ruusunen 2013).

### 7.2.2 Dynamic LE modelling

Dynamic models are needed in control design, and in batch processes they are used for on-line forecasting (Tab. 8). Differences between measured and predicted are used in intelligent analysers, e.g. indicators for fuel quality and water quality (Tab. 8). The dynamic LE models, which are parametric models in the ARX form (38), are developed for steps equal to the sample time. The number of delay $n_k$ of the input is important in these models. Simulation is based on numerical integration methods, known as solvers. Since the stiff solver ode15s with variable step size is used, the effective time delay (Fig. 18(b)) is used as a real value.

The models are developed for the scaled values of the case specific input and output variables listed in Table 9. In most cases, the models are based on the structure shown in Figure 18(a). Specific LE models are combined with fuzzy models to provide a smooth and fairly accurate overall behaviour of the LE model with fuzzy systems generated for special situations. Cascaded models A, B and C are used to handle more complicated interactions, where each subsystem is a dynamic LE model, see Figure 5. These models are discussed in Section 7.2.3.

**Gas furnace**

The dynamic LE modelling approach was first tested in a gas furnace system, where air and methane are combined to form a mixture of gases containing carbon dioxide ($CO_2$). Air fed into the gas furnace is kept constant, while the methane ($CH_4$) feed rate is varied in any desired manner and the resulting $CO_2$ concentration is measured in the exhaust gases at the outlet of the furnace. A time series for predicting the concentration is provided in (Box & Jenkins 1970). The data set with 292 points has been used for
dynamic modelling, interactive testing and tuning with neural networks and genetic algorithms.

The LE models have been developed with the first 115 data points, and the rest of the data set has been used for testing. Considerable differences can be found only in the end part of the data set, where also the fuzziness of the equation becomes higher suggesting that there are unknown additional effects. The simulation results are very accurate even when comparing each calculated value with the interpolated measurement values. The genetic tuning of the membership definition for the $CO_2$ concentration slightly improve the fit to the training data but the improvements to the validation and testing data were negligible.

**Solar thermal power plant**

The dynamic of the solar collector process is characterised by the following aspects:

- Time varying transport delay depends on the manipulated variable (oil flow rate, $F$).
- The dynamics, in particular high frequency peaks in the frequency response of the plant, is difficult to model.
- The plant has a nonlinear behaviour, and therefore linearised models depend on the operating point.
- The solar irradiation, $I_{eff}$, acts as a fast disturbance with respect to the dominant time constant of the process.

The dynamic LE models are based on test campaigns, which cannot be planned in detail because of changing weather conditions. Usually, test campaigns include step changes and load disturbances. As the process must be controlled all the time, modelling is based on process data from controlled process. Weather conditions take care of irradiation disturbances. The start-up is very different from the normal operation, which is divided into three operating conditions. The model structure and variables are the same for all of them (Tab. 9), but the interaction coefficients are case specific. (Juuso 2003a, 2004d, 2009b)

**Water treatment**

In the pulp and paper industry a huge amount of water flows through different processes. The purification can be done with chemical water treatment, which includes complex
Table 9. Examples of dynamic LE models: one delayed value of each input, the current and new value of the output, and a bias term.

<table>
<thead>
<tr>
<th>Case</th>
<th>Input variables</th>
<th>Output variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas furnace</td>
<td>Methane</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>Solar collector field</td>
<td>Oil Flow</td>
<td>Temperature difference</td>
</tr>
<tr>
<td></td>
<td>Effective irradiation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inlet temperature</td>
<td></td>
</tr>
<tr>
<td>Water treatment</td>
<td>Chemical dosages</td>
<td>Outlet turbidity</td>
</tr>
<tr>
<td></td>
<td>Suspended solids</td>
<td></td>
</tr>
<tr>
<td>Lime kiln</td>
<td>Fuel feed</td>
<td>Hot end temperature</td>
</tr>
<tr>
<td></td>
<td>Production level</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fuel feed trend</td>
<td></td>
</tr>
<tr>
<td>Batch cooking</td>
<td>Alkali</td>
<td>Change of alkali</td>
</tr>
<tr>
<td>Models A, B and C</td>
<td>Lignin</td>
<td>Change of lignin</td>
</tr>
<tr>
<td></td>
<td>Dissolved solids</td>
<td>Change of dissolved solids</td>
</tr>
<tr>
<td></td>
<td>H-factor</td>
<td></td>
</tr>
<tr>
<td>Fluidised bed granulation</td>
<td>Change of air humidity</td>
<td>Temperature difference</td>
</tr>
<tr>
<td>Models A, B and C</td>
<td>Air flow</td>
<td>Humidity difference</td>
</tr>
<tr>
<td></td>
<td>Granule size</td>
<td></td>
</tr>
<tr>
<td>Fed-batch fermentation</td>
<td>Glucose feed rate</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>Models A, B and C</td>
<td>Mixing power</td>
<td>Oxygen transfer rate</td>
</tr>
<tr>
<td></td>
<td>Aeration rate</td>
<td>Dissolved oxygen</td>
</tr>
<tr>
<td></td>
<td>Cooling water flow</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pressure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Volumetric oxygen</td>
<td></td>
</tr>
<tr>
<td></td>
<td>transfer coefficient</td>
<td></td>
</tr>
<tr>
<td>Wastewater treatment</td>
<td>Chemical oxygen demand</td>
<td>COD reduction</td>
</tr>
<tr>
<td>Models A, B and C</td>
<td>Inflow</td>
<td>Sludge settling</td>
</tr>
<tr>
<td></td>
<td>Suspended solids</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oxygen</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nutrients</td>
<td></td>
</tr>
<tr>
<td>Fatigue</td>
<td>Stress</td>
<td>Number of cycles</td>
</tr>
</tbody>
</table>

nonlinear phenomena such as coagulation and flocculation processes. The modelling of these complicated processes is mainly data-based or empirical due to lack of comprehensive physical models. Experimental design has been used in (Ainali et al. 2002). In the flotation unit, process water is treated with a polymer which reacts with extractives and forms pitch sludge. The dynamic LE model is similar to the model shown in Figure 18(a): the outlet turbidity is calculated on the chemical dosages, previous calculated turbidity and properties of incoming water, which are represented by the
concentration of suspended solids. The dynamic simulator contains a dynamic LE model for the flotation basin, controllers for two chemicals and a soft sensor for the detection of incoming water quality. The basic dynamic flotation model is the core of the quality indicator. (Ainali et al. 2002, Joensuu et al. 2005)

Fatigue

The continuous model (91) extends the principle of the Palmgren-Miner linear damage hypothesis (Palmgren 1924, Miner 1945). In each sample time, $\tau$, the cycles $N_C(k)$ obtained from $I_S(k)$ by (91), and the resulting contribution $\tau/N_C(k)$ is summarised to the previous contributions

$$C(k) = C(k-1) + \frac{\tau}{N_C(k)},$$

which can also be used for predictions based on use scenarios. 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. (Juuso & Lahdelma 2012)

7.2.3 Decomposition in LE models

The multimodel LE system with a fuzzy decision module shown in Figure 25 was first used for a lime kiln (Juuso 1999b) and then for a solar thermal power plant Juuso (2003a). Special case models based on fuzzy set system were introduced in the solar application (Juuso et al. 2000). Cascade and interactive structures were taken into use in forecasting models: an interactive model (Fig. 5(c)) was used for batch cooking (Juuso 2003b), and a cascade model (Fig. 5(a)) for a fluidised bed granulator (Mäki et al. 2004). Both the multimodel system (Fig. 25) and the cascade structure (Fig. 5(b)) were combined in fed-batch fermentation (Saarela et al. 2003a). The approach is generic for batch processes (Juuso 2001, 2004a). These applications, which were introduced during the phases I-III, are here discussed as examples of complex structures (Fig. 9).

Lime kiln

A dynamic LE model was developed for the same lime kiln as the steady-state model (Juuso 1999b). The multimodel simulator was based on six operating areas defined by
the production level (high, normal and low) and the trend of the fuel feed (increasing, decreasing). The structure of each submodel is the same (Tab. 9): the differences between the submodels are introduced by the membership definitions. Smooth changes between six submodels are controlled by a fuzzy decision module as in Figure 25. Since the working point is not calculated, two inputs are fed directly to the fuzzy module. Time delays were smoothly changing fuzzy numbers as each submodel had its own time delay vector. Some considerable differences revealed fairly short disturbances in fuel quality. This result lead to a new fuel quality indicator which had an essential part in the development of a successful control system, see Section 5.5. The simulator was used for controller design Juuso (1999a): the first versions of the multilevel controller presented in (Järvensivu et al. 2001) were tuned with this simulator.

\textit{Solar thermal power plant}

The dynamic of the process depends on the general field operating conditions. The model structure in Figure 25 is modified to have four submodels corresponding to start-up, low, normal and high operation. Each submodel has the same variables (Tab. 9), but the interaction coefficients are model specific. According to the tests, the fuzzy LE system with four operating areas is clearly the best overall model Juuso (2003a): the simulator moves smoothly from the start-up mode via the low mode to the normal mode and later visits shortly in the high and low mode before returning to the low mode in the afternoon. Even oscillatory conditions are handled correctly.

The dynamic LE simulator predicts well the average behaviour but requires improvements for predicting the maximum temperature since the process changes considerably during the first hour. For radiation disturbances, the LE simulator operates quite well: the temperature is on an appropriate range all the time and the timing of the changes is very good. For handling special situations, additional fuzzy models have been developed on the basis of the Fuzzy–ROSA method. For the period after radiation disturbances, the combined model (Fig. 26) improves the result considerably and is feasible for controller tuning but more special cases need to be analysed to expand the operating area of the dynamic simulator (Juuso et al. 2000).

The membership definition of the outlet temperature does not depend on time and the bias term is zero. Model coefficients and the scaling functions for $T_{\text{diff}}$, $I_{\text{eff}}$ and $F$ are all model specific. The total number of parameters is 60 for variables, 12 model coefficients, and 12 for time delays.
**Batch cooking**

For batch cooking, the overall LE model structure is similar to Figure 5(c): the interactive dynamic models are constructed for alkali (Model A), lignin (Model B) and dissolved solids (Model C). The model is a set of three equations, in the present these equations are used for the calculation of changes for alkali, lignin and dissolved solids. The dynamic models in (Juuso 2001) were based on the structure shown in Figure 18(a). The analysis of the on-line operation changed the submodels completely although the high level model shown in Figure 5(c) remained unchanged. The models calculate directly the change of predicted variable (Juuso 2003b). All three submodels are interacting strongly (Tab. 9).

**Fluidised bed granulation**

Test campaigns based on experiment design were done in a bench-scale fluidised bed granulator (Mäki et al. 2004). On-line measurements were collected with the sampling time of one second, and particle size distributions were analysed for on an average eight samples per batch. The overall model consists of three models as presented in Figure 5(a), the models are: temperature (Model A), humidity (Model B), and granule size (Model C). The input variables in these models are temperature difference between granules and air in the granulator, humidity difference between incoming and outgoing air and the amount of incoming air. Output variables were the temporary value of the granule temperature, the new value of humidity difference and the new estimated value of the granule size, correspondingly, see Table 9.

**Fed-batch fermentation**

Linguistic equations, neural networks and fuzzy modelling with several variants have been compared by using the process data obtained from an industrial fed-batch fermenter (Saarela et al. 2003b). Steady-state modelling is easy: linguistic equations, linear neural networks, feedforward neural networks and TS fuzzy models created by subtractive clustering appeared to be the best. However, the dynamic simulation is demanding for most of the methodologies.

A multimodel approach is shown in Figure 25, where the submodels 1, 2 and 3 correspond to the three growth phases of the fermentation process: (1) lag phase, (2)
exponential growth phase, and (3) steady-state phase. The fuzzy decision system calculates the weights for the submodel phases of the process. The overall model consists of three prediction models (Fig. 5(b)): carbon dioxide concentration (Model A), oxygen transfer rate (Model B), and dissolved oxygen concentration (Model C). Inputs to the models include measurements from the process, such as mixing power, aeration rate, pressure, and substance concentrations (Tab. 9). Each model contains three submodels and a decision system as shown in Figure 25. The overall dynamic model contains an additional model for calculating the volumetric mass transfer coefficient.

**Wastewater treatment**

In the biological wastewater treatment, the model consists of three interactive models (Fig. 5(b)): the treatment model (Model C) is combined with Model A, which calculates the load from the inflow, chemical oxygen demand (COD) and suspended solids, and sludge settling (Model B). The load, nutrients, oxygen, and temperature are used in Model B, which is the heart of the system: the load and the nutrients should be balanced, i.e. the difference between the scaled load and nutrient levels should be close to zero. A too high nutrient level compared to the load causes poor settling seen as an increase of the diluted sludge volume index (DSVI), which continues as an oscillating behaviour. On the other hand, a too low nutrient level causes problems in settling. The normal levels are the best also for the temperature and the oxygen: too high and low temperatures affect the biomass; too low oxygen levels are harmful and too high levels mean excess energy consumption. (Juuso 2009c)

The multimodel system shown in Figure 25 should be based on the biomass population, which is here assessed by DSVI. Smooth transitions between the submodels can be handled by using the degrees of membership obtained from the difference between the load and the nutrients. The LE models provide continuous smooth transition between different operating conditions. Also oxygen and temperature levels are used in the model as amplifying features. The time perspective of all these variables is taken into account by using moving averages.

**Condition monitoring**

For prognostics, the scaling functions are updated in recursive modelling, which is triggered by a fast increase of the indices. Some estimates of the limits are needed at the
beginning, but later recursive tuning can be used for fine tuning the scaling functions. (Juuso & Lahdelma 2011a) The shape factors $\alpha_j^-$ and $\alpha_j^+$ are adjusted since the higher or lower limits are changing during the recursive analysis. Specific S-N curves of the different operation modes can be combined with fuzzy logic or simply with abrupt changes (Juuso & Lahdelma 2012).

7.2.4 Distributed parameter LE models

In distributed parameter models, the solar collector field is divided into modules, where the dynamic LE models are applied in a distributed way Juuso (2004d). Element locations for partial differential equations (PDEs) are defined by the flow rate. In cloudy conditions, the heating effect can be strongly uneven. These effects are simulated by introducing irradiation disturbances. Uneven distributions of the oil flow are important if the oil flow changes are rapid since some loops may be unable to follow.

The distributed parameter model is aimed for control design. It extends the operability of the simulator to evaluating the controller performance for drastic changes, e.g. start-up and large load disturbances, and local disturbances and malfunctioning. This extension is important for controlling the maximum temperature of the collector field as the previous models were capable only for the simulation of the average temperature. Since there is only one irradiation measurement, which is outside the field, these models are used in sensitivity analysis.

7.2.5 Uncertainty in LE models

The fuzzy LE approach combines the extension principle and fuzzy arithmetic with LE models (Section 6.2.2). The principle was introduced in the forecasting of batch cooking in a pulp mill (Juuso 2007a), and later adapted to the dynamic modelling of a fluidised bed granulator used in production of pharmaceuticals (Juuso 2007b) and in the fed-batch fermentation (Juuso 2008). All these models use fuzzy inputs and crisp models (Tab. 7). Fuzzy inputs can be estimated with the same methods, which are used for extracting the corner points in the recursive adaptation. This extends the idea, which was used in a fuzzy edge detection algorithm developed for improving the performance of the image processing system in a recovery boiler (Murtovaara et al. 1996, Juuso 2005b).
7.3 Nonlinear multivariate process control

Linear control algorithms enhanced with nonlinear scaling and special actions provide a widely applicable integrated control approach (Tab. 10). The LE controllers were first implemented for a solar collectors field at a solar thermal power station at PSA (Juuso et al. 1997b, 1998b). Intelligent analysers and adaptive scaling procedures have later been included to reduce the risks of overheating and oscillations. The model-based controller is also used to limit the acceptable range of the setpoints (Juuso & Valenzuela 2003), and recently an indicator of variable irradiation conditions is used for adjusting the working point limit (Juuso 2012b). A new extension of the model-based control includes the calculated limit to the constraints of the control actions in (Juuso 2012c).

The multilevel LE controller has been applied in the control of the burning end of the lime kiln, where the changes of operating conditions are handled with several feedforward controllers (Järvensivu et al. 2001). An intelligent analyser and adaptive scaling are important in combined feedback and feedforward control developed for water treatment (Joensuu et al. 2004, Piironen et al. 2008).

Applications are in this section analysed by using the classification shown in Figure 7. In practice several modules are needed in each application (Tab. 10): the FB control is fairly similar in all cases, but the importance of the FF control and the intelligent analysers is case specific. The overall performance and control design are discussed here. More detailed results related to the gradual improvements of the controllers can be found in application specific publications.

7.3.1 Feedback LE control

The feedback LE controllers, whose control laws are similar to those of the PID controllers, are through nonlinear scaling closely related to the FLCs, i.e. each LE controller produces a singleton value, $\Delta u_{ij}(k)$, and the balanced change of control is obtained by the weighted sum method. The weights are defined manually for fine-tuning purposes. The basic LE controller could be represented as a fuzzy controller, but the additional levels described above would make the fuzzy controller too complicated for practical applications. As the implementation is very compact, the basic LE controller can be extended with adaptive and model-based features.
Table 10. Applications of LE controllers.

<table>
<thead>
<tr>
<th>Case</th>
<th>Control</th>
<th>Variables</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar power plant</td>
<td>Feedback</td>
<td>Oil flow</td>
<td>Adaptive scaling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Outlet temperature</td>
<td>Braking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Working point</td>
<td>Asymmetry</td>
</tr>
<tr>
<td></td>
<td>Model-based</td>
<td>Working point</td>
<td>Fluctuations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Anomalies</td>
</tr>
<tr>
<td>Lime kiln</td>
<td>Feedforward</td>
<td>Draught fan speed</td>
<td>Inverse model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kiln rotational speed</td>
<td>Inverse model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sawdust feed rate</td>
<td>Inverse model</td>
</tr>
<tr>
<td></td>
<td>Feedback</td>
<td>Fuel oil flow rate</td>
<td>Inverse model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fuel feed</td>
<td>Adaptive scaling</td>
</tr>
<tr>
<td>Water treatment</td>
<td>Feedforward</td>
<td>Chemical 1</td>
<td>Braking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Quality analyser</td>
</tr>
<tr>
<td></td>
<td>Feedback</td>
<td>Chemical 2</td>
<td>Adaptive scaling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Quality analyser</td>
</tr>
</tbody>
</table>

**Solar thermal power plant**

Solar power plants should be designed to collect all the available thermal energy in a usable form within a desired temperature range. In cloudy conditions, the collector field is maintained in a standby mode ready for full-scale operation when the intensity of the sunlight rises again. Control is achieved by means of varying the flow of oil pumped through the pipes during the plant operation. For the solar collector field, the goal is to reach the nominal operating temperature 180 - 295 °C and keep it in changing operating conditions.

The feedback controller is a PI-type LE controller (Tab. 3) with one manipulating variable, oil flow $F$, and one controlled variable, the maximum of the outlet temperatures of the loops (Fig. 27), or shortly denoted as the outlet temperature $T_{out}$. The original controller shown in Figure 21(b) was defined by the coefficients $K_P(i, j) = K_I(i, j) = 1$ (Juuso et al. 1997b, 1998b), and extended to real-valued coefficients in (Juuso 2006). The basic LE controller is defined for the normal working point $wp_i = 0$. 

161
Lime kiln

Lime kilns are large machines with very slow rotation speeds and high temperatures in the hot end. Depending on production conditions, the kiln must run at different production capacities and rotation speeds. Speed is controlled together with fuel feed and draft fan speed in order to obtain good operating conditions (Järvensivu et al. 2001). The flue gas fan was controlled with an FLC, which was tuned with linguistic equations (Juuso et al. 1996): each equation corresponds to a group of rules describing the effect of one or two input variables on the manipulated variable, change in the fan speed. The FB controller is represented by \( A \cdot X = 0 \), where

\[
A = \begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix},
\]

and \( X \) includes the scaled values of the variables: (1) feed end temperature, (2) change in feed end temperature, (3) hot end temperature, (4) change in hot end temperature, (5) excess oxygen, (6) change in excess oxygen, (7) temperature after the cyclone and (8) change in the fan speed. This MISO controller resembles the PI-type LE controller defined by (49).

The fuzzy FB controller mitigated the process and made control work easier but it was limited to a narrow operating area. Later this controller has been replaced by a combination of feedforward and feedback LE controllers (Tab. 10). The FB control, which is required in order to maintain the hot-end temperature within the most favourable range for the lime quality, are used for the fuels: sawdust and oil. The controllers are PI-type LE controllers (Tab. 3) based on two controlled variables: the hot end temperature and the cold end temperature with weights 0.7 and 0.3, respectively. The error is calculated as a difference of two moving averages: 5 minutes and 30 minutes for the hot end and 4 minutes and 30 minutes for the cold end. (Järvensivu et al. 2001)

Water treatment

The key to efficient chemical dosage control in water treatment processes is the addition of proper amount of chemicals to the process. If too much chemical is added, treatment targets will be reached but costs will increase and extra sludge waste is produced. Too
small a dosage will lead to poor treatment results and problems in subsequent processes such as filtration, flotation and sedimentation. Two chemicals are used in the flotation unit: a coagulant neutralises the charge on the surface of particles to enable particles to coalesce and form small flocs; a flocculant forms bridges between flocs and increases the floc size. (Joensuu et al. 2005)

The feedback controller is a PI-type LE controller with one manipulating variable, the flocculant dosage, and one controlled variable, the outlet turbidity (Tab. 3). (Joensuu et al. 2004, Piironen et al. 2008). The appropriate dose depends on the quality of the raw water and the purification target. As the reaction is fast and the delay of the flotation basin is short, the response time is short enough for the FB control.

### 7.3.2 Intelligent analysers

Intelligent analysers are used for detecting changes in operating conditions to activate adaptation and model-based control and to provide indirect measurements for high-level control. The working point $wp_i(k)$ defined by (51) is an important intelligent index for adaptive control in all three applications. They can also be based on the model-based performance analysis of the control actions, e.g. chemical dosage and fuel feed. New intelligent indices are highly important in the control solutions of the solar power plant (Juuso 2012b,c).

**Solar thermal power plant**

In the solar application, the working point variables are the effective solar irradiation, $I_{eff}$, and the temperature difference, $T_{diff} = T_{out} - T_{in}$. The bias term is used for fine tuning, normally $b_{wp}^p = 0$ (Juuso et al. 1997b). Tuning with additional data would still be needed for using the ambient temperature, $T_{amb}$, which was included in Juuso (2000a). The idea of changing the meaning of $\Delta u_{ij}$ came in (Juuso et al. 1997b), and the PBA and ASA actions were introduced in (Juuso et al. 1998b). These actions are represented as corrections based on intelligent analysers and adaptation mechanisms, see Figures 29 and 30.

The predictive braking coefficient $bc_j(k)$ defined by (65) is activated for large initial errors, see Section 5.2.2. PBA is activated in the start-up: first during the circulation in the field and again when the oil flow is taken to the storage system (Fig. 29(a)). Model based adaptation reduces the latter impact (Fig. 29(b)). Large corrections are also
introduced during cloudy periods (Fig. 29(a)), and small corrections are needed for large setpoint changes and load disturbances (Fig. 29(b)). The overall level is restricted with the variable specific coefficient \( (c_B)_{j} \in [0, 1] \). The irradiation disturbances, which are seen as fluctuations of the working point (Fig. 31), activate PBA. The peaks seen after the start-up period in Figure 29(a) correspond to the fluctuations seen in Figure 31(a).

The change of the working point, \( \Delta wp_i(k) \), indicates asymmetry (Section 5.2.3). The asymmetrical action is activated for short periods, and the level depends on the variable specific coefficient \( (c_A)_{j} \in [0, 1] \). The correction is negative before solar noon and positive after that (Fig. 30(b)), but irradiation disturbances may cause differences to that (Fig. 30(a)). Also in Figure 30(b) the positive values before solar noon are caused by very short irradiation disturbances (Fig. 31(d)). Longer ASA periods (Fig. 30(b)) are seen only during smooth operation on a clear day (Fig. 31(d)). In the current controller, the change \( \Delta wp_i(k) \) is evaluated from the derivative of the irradiation \( \Delta I_{eff} \). The actions PBA and ASA are not activated at the same time.
Fig 31. Working point and state indicators for the irradiation, temperature difference and oil flow (Juuso 2012b, published by permission of IFAC).
The cloudy conditions are detected by calculating the difference of the high and the low values of the corrected irradiation, see (80). These values are calculated as moving norm values from the latest minute, which includes five measurement values, \( \tau = 15 \text{s} \). Then an average of the 25 latest values is used as the indicator, i.e. latest five minutes will have an effect on the indicator value. The norms (7) with high positive and high negative order are more robust than the maximum and minimum. In this case, the orders are 30 and -30, respectively. (Juuso 2012b) The difference, denoted by \( \Delta I_{\text{eff}} \), indicates variable irradiation well (Figs. 31(a), 31(b) and 31(c)). Severity of cloudy periods is detected efficiently: difficult periods shown in Fig. 31(b) result much higher indicator values than weaker cloud periods shown in Fig. 31(a) and very short periods of clouds shown in Fig. 31(c). Just a small sign of clouds is seen in Fig. 31(d). (Juuso 2012b)

The same orders and time periods are used for the norms in cases of the temperature difference and the oil flow. A fast temperature increase is detected with \( \Delta T_{\text{diff}}^F \) in the start-up phase in all cases shown in Fig. 31. Also irradiation disturbances introduce a fast temperature increase (Figs. 31(a) and 31(b)). If this increase is not taken into account in control, also the oil flow starts to oscillate as was seen in (Juuso & Valenzuela 2003). State indicators \( \Delta T_{\text{diff}}^F \) and \( \Delta F_{\text{out}}^F \) react to this as seen in Figures 31(a) and 31(b).

Predictive actions introduced in (Juuso & Valenzuela 2003) for avoiding oscillatory conditions and too high temperatures can be based on intelligent indices which detect anomalies:

- fast change of the inlet temperature obtained by
  \[
  \Delta T_{\text{in}}^H(k) = T_{\text{in}}(k) - \frac{1}{n_L + 1} \sum_{i=k-n_L}^{k} T_{\text{in}}(i),
  \]
  \(94\)
- too fast outlet temperature increase by the value range
  \[
  \Delta T_{\text{out}}^R(k) = \max_{i=k-n_L...k} \{ T_{\text{out}}(i) \} - \max_{i=k-n_L...k} \{ T_{\text{out}}(i) \},
  \]
  \(95\)
  if \( T_{\text{out}} \) has increased during the period,
- too high temperature difference by an overshoot
  \[
  \Delta T_{\text{out}}^H(k) = \max \{ 0, T_{\text{out}}(k) - T_{\text{out}}^H \}.
  \]
  \(96\)

The window for the recent values is defined by delay \( n_L \). A robust estimate for \( T_{\text{out}}^R \) can also be obtained by using (80).

Large differences are detected in the start-up with \( \Delta T_{\text{in}}^H \) and \( \Delta T_{\text{out}}^R \) (Fig. 32). Load disturbances are seen in \( \Delta T_{\text{in}}^H \), and \( \Delta T_{\text{out}}^H \) reacts to overshoot and oscillations, which
become considerable in variating cloudy conditions and after load disturbances. On a clear day, these indices are negligible if there are no load disturbances. Adjusting the working point limit with the new fluctuation indices $\Delta I_{f eff}$, $\Delta T_{diff}$ and $\Delta F^F$ prevent oscillatory conditions and too high temperatures, which has reduced the need of the indices (94), (95) and (96). The PBA is also used less frequently since the smooth operation reduces the number of large changes. The ASA, which requires smooth operation, is started more frequently. (Juuso 2012c)

![Graphs showing temperature changes](image1)

(a) Short cloudy periods.  
(b) A clear day with load disturbances.

**Fig 32. Intelligent indices for detecting fast temperature increase and high temperature differences.**

Trend indices (78) and corresponding episodes are useful in warning about the changes of irradiation, inlet temperature, outlet temperature, temperature difference and oil flow. The indices can be obtained without additional scaling. Trends can also be detected for fluctuation, asymmetry and anomaly indices of any of these variables. The scaling functions provide limits for the statistical process control.

**Lime kiln**

In the lime kiln control, the working point variables are the production rate and the draught fan speed. The PBA is used in an modified form to cope with the long
measurement delays. The coefficient $bc_j(k)$ is defined by (67). The fuel quality indicator is handled with a control power $cp_j(k)$ defined by (52). This is an important index for the saw dust, but not needed for the fuel oil. The cumulative rate of control actions $cr_j(k)$ defined by (55) has an important effect, since the effective time delay of the process is long. Indicators have been developed for detecting large deviations or fast changes in the temperature or excess oxygen. (Järvensivu et al. 2001)

### Water treatment

In the water treatment control, the working point variables are the water quality and the setpoint of turbidity (Piironen et al. 2008). The water quality indicator introduced in (Ainali et al. 2002) is essential in avoiding on the other hand oscillations and on the other hand too slow operation, since the quality and amount of incoming water can fluctuate greatly. The basic dynamic flotation model discussed in Section 7.2.2 is the core of the quality indicator: the predicted outlet turbidity is compared with the on-line measurement to calculate the impurity level.

#### 7.3.3 Adaptive LE control

Adaptive LE control is predefined in applications, which need fast adaptation in a wide operating area (Juuso et al. 1998b). The predictive braking action is used for large errors. The asymmetrical action, additional smart control actions and cascade control were developed for the solar power plant (Juuso & Valenzuela 2003). The new intelligent indices improve the working point control by preventing the oscillating situations (Juuso 2012b,c).

### Solar thermal power plant

The operation of the feedback LE controller is modified by the working point $wp_i(k)$, the predictive braking coefficient $bc_j(k)$ and the change of the working point, $\Delta wp_i(k)$. The adaptive scaling, which makes the control surface steeper or flatter, was already in (Juuso et al. 1997b). The predictive braking and asymmetry actions were introduced in (Juuso et al. 1998b). The coefficients $K_P(i, j)$ and $K_I(i, j)$ of the PI-type LE controller are modified by the coefficient $bc_j(k)$ and the change of the working point, $\Delta wp_i(k)$, see (68), (69) and (70).
In the beginning of the start-up, the oil temperature decreases in the field, i.e. $T_{\text{diff}} < 0$, leading to negative power and efficiency values for a while (Juuso 2011b). Working point $wp_i(k) = 2$ during the circulation in the field. The temperature rises first to the level of the inlet temperature, i.e. $T_{\text{diff}} = 0$, and then to sufficiently high temperature to be taken to the storage system. Start-up is very fast in good weather conditions: the temperature $180 \, ^\circ\text{C}$ is achieved in all active loops, i.e. effective energy collection starts, even in less than 30 minutes (Fig. 33). The irradiation increase is utilised efficiently, and $T_{\text{diff}}$ goes to normal level, i.e. $wp_i(k)$ also goes to zero if the irradiation is normal (Fig. 31(a)).

![Graph](a) Start-up time.  
![Graph](b) Temperature variation.

**Fig 33. Start-up period during a two week test period.**

The change of control is in the beginning multiplied by 1.5 which corresponds to $wp_i(k) = 2$, and the coefficient goes to 1.0 when $wp_i(k) \to 0$. Smaller changes are used if $wp_i(k) < 0$. Naturally, the length of the start-up period depends on the inlet temperature: a quadratic model fits to the cases shown in Figure 33(a). Deviations from this model are explained by the differences in irradiation, which was between 303 and 706 W/m², the average was 495 W/m². An afternoon start is fast since the irradiation is high, but the start-up can be fast even in cloudy conditions if the inlet temperature is high. The maximum temperature variation between the loops during the start-up periods is high in cloudy conditions (Fig. 33(b)). However, the lowest variation was achieved when the temperature was rising very slowly. In this lightly cloudy case, also the inlet temperature was rising slowly. For clear days, high variation is related to the high values of the average working point.

The scaling coefficient $sc_i(k)$, which is based on the working point values $wp_i(k)$, has fast changes in cloudy periods (Fig. 34(a)). Since the braking action is activated
(a) Short cloudy periods.  
(b) Heavy cloudy periods.  
(c) Setpoint tracking.  
(d) A clear day with load disturbances.

Fig 34. Correction factors on four test days.

(Fig. 29(a)), the combined oil flow changes use the whole range (Fig. 35(a)). The asymmetrical action activates shortly in the end of start-up and the setpoint changes (Fig. 30(a)). On a clear day, the predictive braking is active in the start-up and again when the load disturbances take place (Fig. 29(b)). Asymmetrical action activates several times (Fig. 30(b)). Large oil flow changes are introduced in the start-up, load disturbance and in the beginning of setpoint changes (Fig. 35(b)).

The additional intelligent features (94), (95) and (96), which detect anomalies, introduce an additional change of control:

\[
\Delta u^C_j(k) = c_{in} \Delta T^H_{in}(k) + c_{out} \Delta T^R_{out}(k) + c_{diff} \Delta T^H_{out}(k),
\]  

(97)

where the coefficients \(c_{in}\), \(c_{out}\) and \(c_{diff}\) are chosen from the range \([0, 1]\). The first two actions are predictive, and the third one is corrective. If \(T_{diff}\) is too high, also the setpoint is corrected correspondingly to avoid low working point \(wp_i(k) \ll 0\). Additional control actions are introduced in start-up and after cloudy periods (Fig. 35(a)). On a clear day, these additions are very small (Fig. 35(b)).
Fig 35. Oil flow changes on two test days.

**Lime kiln**

The predefined adaptation method was applied in the lime kiln control (Järvensivu *et al.* 2001): the scaling of the change of control depends on the working point \( wp_i(k) \), the control power \( cp_j(k) \), and the cumulative rate of the control actions \( cr_j(k) \), see (56). The weights for these terms are defined in the high level control. The coefficient \( K_P(i,j) \) of the PI-type LE controller is modified by the coefficient \( bc_j(k) \). The production rate has the highest effect: the control actions are at their largest when the production rate is high, and vice versa.

In the case of large deviations or fast changes in the temperature or excess oxygen, CH module is activated to make appropriate stepwise corrections to the fuel supply. Both the sawdust and fuel oil have their own CH modules. Adaptive limits (72) and (73) are used in CH module. Stepwise corrections \( \Delta u^{CH}_i(k) \) are used to decrease the fuel supply, often after severe disturbances, in order to prevent extensive temperature excursions and to protect the refractory linings from excessive heat. Corrections that increase the fuel supply are aimed to prevent a decline in the temperature, e.g. due to rapid deterioration in sawdust quality, down to the level at which substantial deterioration
occurs in the lime quality. These corrections can be even 5-10 times larger than the changes introduced by the feedback control. Reasonably large changes are used to bring the process immediately back to within the safe operating range. After each correction, the module is blocked for a certain time in order to give the process time to become stable after stepwise control actions. The CH module is also used to prevent the LE controller from further increasing the fuel supply if the excess oxygen content drops to too low a level. (Järvensivu et al. 2001)

**Water treatment**

In the water treatment case, the adaptation is needed to avoid oscillations in low polymer doses when the process conditions change strongly and fast. Respectively, if the polymer dose is high, the change of control has to be bigger to avoid the slow control. The dosage curve of the polymer is nonlinear, i.e. with low polymer doses changes have more effect on outgoing turbidity than with high doses. The predefined adaptation method uses an adaptation coefficient obtained from the working point obtained from the water quality and the setpoint of turbidity (Piironen et al. 2008). Pre-tuning facilitates fast operation in changing process conditions: the controller does not need time for finding correct parameters, since the changes are detected by the water quality indicator.

### 7.3.4 Model-based LE control

The working point models, which are used predefined adaptation, are the basis for limiting the setpoints in the solar plant (Juuso & Valenzuela 2003, Juuso 2012b,c). Feedforward control is an essential control module in the lime kiln (Järvensivu et al. 2001) and the water treatment (Joensuu et al. 2004). MPC has been used in controller tuning in the solar thermal power plant (Juuso 2006). Model-based approaches can be used for optimising the energy collection (Juuso 2012c).

**Solar thermal power plant**

The manual adaptation of the working point limit \( w_{\text{min}} \) improves the operation by limiting the setpoint of the outlet temperature (Juuso & Valenzuela 2003, Juuso 2005a): the setpoint \( T_{\text{ref}}^{\text{wp}} \) has abrupt changes (Fig. 36). The limited setpoint improved operation considerably in the start-up stage and when the inlet temperature changes. During the
tests on a clear day (Fig. 36), the \( w_{p_{\text{min}}} = 0 \) was used in the morning, and before 13:00 hours \( w_{p_{\text{min}}} \) was raised for a while to two and then level one was kept for the rest of the afternoon. The start-up was very smooth when the setpoint was defined by \( w_{p_{\text{min}}} = 0 \). After the load disturbance, there was hardly any overshoot and the oscillations were damped fast although the temperature change was drastic. (Juuso 2005a)

![Graphs showing temperature, oil flow, and irradiation changes](image)

Fig 36. Test results of the Linguistic Equation Controller: temperatures, oil flow and irradiation (Juuso 2012b, published by permission of IFAC).

Indicators are needed to make the working point changes automatic for cloudy conditions and load disturbances, which introduce a fast increase and oscillations of temperature and oil flow (Juuso 2012b,c). Manual actions for changing the setpoint, or the limit \( w_{p_{\text{min}}} \), are easily late and kept active too long time. The new indicators increase the limit \( w_{p_{\text{min}}} \) in the beginning of the start-up, and a short period of additional limits is enough on a clear day (Figs. 31(c) and 31(d)). In these cases the limit \( T_{\text{diff}} = 0 \) operates well, but in cloudy conditions the correction continues a longer time (Figs. 31(a) and 31(b)). Setpoints, which are under the limit defined by the working point limit, can be reached smoothly on a clear day (Figs. 31(c) and 31(d)). There is some
improvement potential for short cloudy periods (Fig. 31(c)). The whole working point range is not necessarily available. The irradiation level was too high for the setpoint after 14 o’clock in the case shown in Figure 31(d): the oil flow reached the maximum but the temperature was still increasing. On the other hand, the limits $wp_{min} < 0$ can be acceptable if the oscillations are handled sufficiently well (Fig. 31(c)).

The fluctuation indicators improve the operation through the working point adaptation on clear days (Fig. 37), in cloudy periods (Fig. 38) and for load disturbances (Fig. 39) (Juuso 2012c). The operating conditions can change during a day, e.g. two very different periods can be seen in Figure 38(b): the start was very bright and the irradiation was rising smoothly, but everything was changed just before the solar noon, and the heavy cloudy period continued the whole afternoon.

On a clear day with high irradiation, the setpoint tracking was very fast (Fig. 37(a)): step changes from 15-25 degrees were achieved in 20-30 minutes with minimal oscillations. The temperature followed the increase and decrease steps in spite of the irradiation changes. The working point limit activated the setpoint correction when the temperature difference exceeded the limit corresponding to the irradiation level. On a fairly clear day with a lower and slightly varying irradiation, the setpoint correction was activated more often throughout the day (Fig. 37(b)). The temperature followed the setpoint well with small offsets. Working point corrections and limiting the fast change were negligible in both cases. The operation is smooth during sunny spells (Fig. 38) and between load disturbances (Fig. 39).

In the start-up, the temperature follows the setpoint defined by the working point limit: the operation is very smooth despite the increasing irradiation and inlet temperature (Fig. 38(b)). The operation can be started from low inlet temperatures with the minimum flow and stopped on any level (Figs. 39(a) and 39(b)). Also the small temperature increase, which was caused when a new loop was taken into use, was efficiently corrected (Fig. 38(b)).

Three cloudy periods are seen in Figure 38(a): a long period in the morning, a short light one close the solar noon and a short, but heavy, in the afternoon. The temporary setpoint correction operated well in these situations. In the first case, the temperature went down with 20 degrees but rose back during the short sunny spells, and finally, after the irradiation disturbances, high temperatures were achieved almost without oscillations with the gradually changing setpoint defined by the working point limit although the inlet temperature was simultaneously rising. The oil flow was changed smoothly also during the other two cloudy periods. The working point corrections were
Fig 37. Test results of the LE controller in a solar power plant: clear days (Juuso 2012c).

now very strong, but limiting the fast changes was hardly needed. Strong braking was used in the beginning and in the recovery from the first cloudy period. There were problems with some loops during that day. The heavy clouds mean going back to the minimum flow, but also lower setpoints. In the case shown in Figure 38(b), the field was ready for normal operation and short sunny spells raised the temperature, but also the oil flow. The controller was ready to prevent a high overshoot, if the sky clears up. The field was in temperatures 160 - 210 °C for more than two hours although the loops were not tracking the sun all the time. The working point corrections were during this period very strong, but limiting the fast changes was hardly needed.

Fig 38. Test results of the LE controller in a solar power plant: cloudy periods (Juuso 2012c).
The load disturbances seen as variable inlet oil conditions were handled well by using the limit $w_{\text{min}} = 1$ in the case shown in Figure 36. Oscillations of the oil flow are further reduced by using the fluctuation indicators, which propose slightly smaller changes for the setpoint (Fig. 31(d)). During the day shown in Figure 39(a) there was an unintentional drop of 16.9 degrees in the inlet temperature. The disturbance lasted 20 minutes. The controller reacted by introducing a setpoint decrease of 19.8 degrees. The normal operation was retained in 50 minutes with only an overshoot of two degrees, but with some oscillations. The setpoint correction was too early and too large. The disturbance was repeated on the sixth day (Fig. 39(a)): maximum 13.5 degrees and 15 minutes. Now the setpoint was changed when the inlet temperature reached the minimum. The working point limit was changed to allow a higher setpoint in the recovery. The effect was very small in practice: the temperature drop was 7.5 degrees, the overshoot 2.5 degrees and the recovery time 30 minutes.

(a) A fairly clear day with a load disturbance.  
(b) A fairly clear day: asymmetrical action.

Fig 39. Test results of the LE controller in a solar power plant: load disturbances (Juuso 2012c).

The asymmetrical action was not used on first two days (Fig. 37)) resulting a small offset: the outlet temperature exceeded the setpoint with 0.5-1 degrees, when the irradiation was increasing, and remained about 1.0 degrees lower when the irradiation decreased. The offset was on later days removed in the morning with the new asymmetrical action (Figs. 38(b) and 39(a)). The action was better tuned for the afternoon as well on the sixth day (Fig. 39(b)). Then the setpoints were achieved in the range $\pm 0.5$ degrees with hardly any offset, which was a considerable improvement to the first days. Around the solar noon, the setpoint was achieved very accurately.
Energy collection depends on the oil flow, the temperature difference and the properties of the oil. High temperature differences are achieved by using low oil flow, and high flow leads to low temperature differences. The highest energy collection in a time unit is achieved by selecting the optimal temperature difference (Fig. 40(a)). The working point is defined by the optimal \( T_{\text{diff}} \) and the irradiation, which is the highest close to the solar noon. The power surface (Fig. 40(b)) is highly nonlinear because of the nonlinear properties of the oil. The acceptable range of the working point is limited: oscillation risks and the high viscosity of the oil during the start-up must be taken into account. In the latest tests, the inlet temperatures are high already in the start-up, since the oil flow was not first circulated in the field. High irradiation periods would lead to too high outlet temperatures, if the oil flow is too low, but this is avoided by keeping the working point under two. The maximum collected power is achieved when the oil flow is close to 6 l/s. Another maximum area close to the upper limit of the oil flow is achieved around the solar noon on a clear day.

![Graphs showing calculated temperature difference vs. oil flow and power vs. oil flow on a fairly clear day.](image)

(a) Calculated temperature difference vs. oil flow on a fairly clear day.  
(b) Power vs. oil flow on a fairly clear day.

**Fig 40. Optimisation of the LE control in a solar plant (Juuso 2012c).**

The irradiation is at PSA measured far from the collector field. Therefore, the MPC methodology has been restricted to the tuning of the braking action to get good trajectories for selected simulation cases (Juuso 2006). Practical improvements are fairly limited, since the periods of active PBA are very short after the start-up (Fig. 29). The strongest corrections are done during cloudy periods, where the new state indicators will lower the setpoint (Fig. 31).
Lime kiln

In the lime kiln control, feedforward LE controllers move the operating conditions smoothly to keep good process operation. The steady-state model (90) was inverted into a FF controller, which calculates the appropriate fuel feed to come up with the lime feed and the burning end temperature requirements (Fig. 22(b)), see (Juuso 1999a). The FF controller was updated for both fuels in (Järvensivu et al. 2001). The saw dust feed rate controller moves the hot end temperature to an appropriate level by using the 60 minutes moving average of the production rate (t/h) updated every 20 minutes and the 2 hour moving average of the torque of the kiln drive (%), which correlates satisfactorily with the total amount of the solids in the kiln. The fuel oil flow rate (kg/s) operates in a similar way: only the scaling functions of the outputs are different.

The controller discussed in (Juuso 1999a) was based on the FF fuel controller and a FB draught fan controller. The control system was in (Järvensivu et al. 2001) updated by using the draught fan controller as a FB controller and introducing more FF controllers. The rotational speed control moderates the filling rate of the kiln by the inputs are the 4 hour moving average of the production rate (t/h) and the torque of the kiln drive (%), which provides a reasonably accurate prediction of the loading state of the process. The feedback was introduced to the fuel feed control described above: the FF fuel controller also restricts the acceptable range for the fuel feed $u_i(k)$: the dynamic range $[u_l^F_i(k), u_h^F_i(k)]$ is defined by (72) and (73) for both the fuels from the most recent $u_i^F(k)$ by using relative thresholds (Section 5.2.4).

Water treatment

The slowly affecting chemical, coagulant, is controlled with a feedforward LE controller, which modifies the typical dosage with effects of the water quality (Section 7.3.2) and the difference of the setpoint from the typical setpoint. The feedforward controller takes into account the process changes in advance (Piironen et al. 2008). The faster effecting chemical, flocculant, is controlled by the adaptive feedback LE controller.

7.3.5 High-level LE control

The weighting of different control strategies is based on the operating conditions, e.g. five controllers in (93) can have different weights. The high-level LE control is based on
activating, weighting and closing different actions of the control system (Tab. 6). The set of active actions depends on the applications.

**Solar thermal power plant**

In the solar thermal power plant, smooth setpoint changes reduce the need of the predictive braking and the additional control actions (97). A faster temperature increase is accepted in the start-up phase to avoid a high flow of the oil whose viscosity is still high.

The high-level control uses the coefficients shown in Table 6 with several restrictions. Only one controller with one controlled variable \( j \) and one manipulating variable \( i \) was used: \( w_{ij} = 1 \). Adaptive scaling is based only on the working point, i.e. \( w_{ij}^{wp} \in [0, 1] \), \( w_{ij}^{cr} = w_{ij}^{ct} = 0 \), and also the predictive braking and asymmetry actions can be chosen from \([0, 1]\) with limitation \( (c_B)_j = (c_A)_j = 0 \) if \( \Delta T_{eff}^i \) is high. The additional special case indicators can be weighted in the same way, see (97). The working point \( wp_{min} \) and the set point \( T_{ref} \) can be changed any time in the high-level control. In the adaptive set point procedure, the set point is not allowed to be higher than the setpoint \( T_{ref}^{wp} \) calculated from the working point \( wp_{min} \). The acceptable range of the oil flow \( [u^{ll}, u^{hl}] = [2, 10] \) (l/s). The change of the flow \( \Delta u_j(k) \) and additional change of flow \( \Delta u_i(k) \) are limited by the scaling functions. Scale coefficients \( \lambda^{−}_{ij} \) and \( \lambda^{+}_{ij} \), the weight factor \( s_i \) and the bias correction \( u_{BC} \) are not used.

Two limiting cases can be seen in Figure 36: (1) oil flow is limited by the viscosity during the start-up, and (2) the maximum flow defines the lower limit for the setpoint. The fluctuation index \( \Delta F_F \) can be used as an indication to turn the flow to the storage tank (Fig. 31). Increasing \( F \) toward maximum flow without simultaneous increases of \( T_{out} \) can be detected by comparing the intelligent trend indices of \( F \) and \( T_{out} \).

**Lime kiln**

The feedforward controllers discussed in Section 7.3.4 are needed all the time. In the feedback control, there are two fuels \( i \), saw dust and fuel oil, and two controlled variables \( j \), hot-end-temperature and cold-end-temperature. The feedback controller (57) has two weight factors \( w_{ij} \) for both fuels. The weight factors and the predictive braking constants \( (c_B)_j \in [0, 1] \), are chosen independently for both fuels in the high level control; the asymmetrical action is not used. The working point and the cumulative rate are used for
both fuels, \( w_i^{np} \in [0, 1] \), \( w_i^{f} \in [0, 1] \). The control power is activated, \( w_i^{sp} \in [0, 1] \), only for the saw dust, \( w_i^{sp} = 0 \) for the fuel oil. Finally, the change of control can be adjusted manually with a factor \( s_i \in [0.5, 1.5] \) and then limited to the range \([u_i^l(k), u_i^h(k)]\). The tube around the feedforward control \( u_i^f \) is defined by the coefficients \( \lambda_i^- \) and \( \lambda_i^+ \), which can be adjusted in the high level control.

The control actions contain both feedforward and feedback terms, and the lime kiln controller also allows the operator to make a bias correction \( u_i^{bc}(k) \) to (77), which provides a possibility for manual control if both the feedback and feedforward control are closed.

Water treatment

In principle, activating coefficients could be chosen in the high-level control for adaptive scaling and quality indicator. Also feedforward control could be excluded if the coagulant control is done manually.

7.3.6 Performance and control design

LE controllers have been tested in three real processes: solar power plant (Juuso 2005a, 2009d, 2012b,c), lime kiln (Järvensivu et al. 2001) and water treatment (Piironen et al. 2008). Dynamic LE models have been used for control design in these three processes: solar power plant (Juuso 2003a), lime kiln (Juuso 1999b) and water treatment (Ainali et al. 2002, Joensuu et al. 2005). The models operate very well in different operating conditions: the outlet temperature is predicted with the collector field model, the hot end temperature with the lime kiln model, and the outlet turbidity with the flotation model, respectively. For the control design, a realistic dynamic behaviour, including oscillations, is necessary and achieved in all these models. Benefits of predefined operation increase with the complexity of the control system.

Solar thermal power plant

The solar collector field is controlled by the flow of oil pumped through the pipes during the plant operation (Fig. 27). The first LE controller based on (47) provided good results close to the solar noon, and after a fast start-up the oscillations were considerably reduced by the adaptive scaling (Juuso et al. 1997b, 1998b). Further improvements were
achieved with the PBA action, and the ASA action removed the offset after set point changes (Juuso et al. 1998b). Additional smart actions and cascade control used in the start-up to facilitate a fast temperature increase without an oscillatory behaviour, and a similar reduction of the set point is activated for load disturbances (Juuso & Valenzuela 2003).

The test results is shown in Figure 36 correspond to a very careful operation on a sunny day: there are hardly any oscillations. The irradiation increase is utilised efficiently: the field goes quite smoothly to the operating temperatures, and handles the increase of $T_{in}$ and the load disturbance well. Short irradiation disturbances do not have any effect on the outlet temperature. Juuso (2005a) When the field was started from the ambient temperature, all the active loops reached in 54 minutes the operating temperature with the maximum temperature difference of 35.2 °C between the loops. The start-up was an average one, since also the average irradiation, 499 W/m², and the average inlet temperature, 110 °C, were normal for the two-week test period, see Figure 33.

According to the test results, the field operates well also in less favourable conditions if the multilevel LE controller is used. Start-up is efficient even in cloudy conditions (Fig. 33), and recovery from the load disturbances (Fig. 36) are clear benefits resulting from the intelligent analysers, predefined adaptation and the cascade control. The model-based cascade control is essential in cloudy condition since a fast irradiation increases may cause severe overshoot and oscillation. Fairly good operating conditions can be kept in spite of drastic changes in irradiation. The multilevel LE controller is necessary in these cases: all the special features described above are used. The plant can operate closer to the temperature limit if the set point is based on appropriate working point. (Juuso 2005a, 2009d) The new state indicators introduced in (Juuso 2012b) operate well in the automatic adjustment of the limit $w_{p_{min}}$ (Juuso 2012c).

The multilevel LE controller extends considerably the operating time of the plant. In good weather conditions, the thermal power around 1 MW, and the efficiency is around 0.5 for most of the time. The start-up period has lower efficiencies, and higher efficiencies are achieved when the field is at the high operation limit. (Juuso & Valenzuela 2003, Juuso 2009d) On a cloudy day, the operation time is shorter, but quite high temperatures are achieved. In cloudy conditions, irradiation is fluctuating and there are periods of very low irradiation. Naturally, the efficiency decreases, but even short sunny spells are utilised quite well. The thermal power almost reaches 1 MW, and the efficiency is almost 0.5 for long periods, even for low irradiation. The temperature
difference is usually kept rather low for safety reasons. Although the efficiencies are almost on the same level as in good weather conditions, the total collected thermal energy reduces, since there is less solar energy available. The collected energy can be approximated from the effective solar energy by a quadratic model. In many cases, this energy can be collected only with the aid of the advanced LE controller tuned with the full multilevel LE simulator.

In the latest tests, the temperature difference was limited under 70-75 degrees to take the condition of the field into account. This reduced the collected solar energy between 2.2 and 3.2 MWh corresponding the efficiency from 0.24 to 0.33. A part of the time only seven or eight loops were used, which means that the average efficiency of a loop was from 0.31 to 0.34. For the day with heavy clouds in the afternoon (Fig. 38(b)), the collected energy was 1.4 MWh corresponding the efficiency of 0.28. The daily maximum power was from 0.66 to 0.9 MW. The lower number of active loops and the lower temperatures reduced the collected daily power with about 1 MWh and efficiency with 0.10 - 0.15.

Trial and error type controller tuning does not work since the operating conditions cannot be reproduced or planned in detail because of changing weather conditions. As the process must be controlled all the time, modelling is based on process data from the controlled process. The simulation system, which consists of the working point model and the dynamic models, has been tuned by using data from several test campaigns. This model has then used for tuning the LE control system in different operating conditions. The distributed parameter models and the fuzzy models have been used for testing the controller. The fuzzy LE model is aimed for choosing proper working points.

The LE controller is mostly defined by the membership definitions. Full membership definitions are developed for two working point variables, $I_{eff}$ and $T_{diff}$. Zero centered scaling functions are developed for the error, change of error and initial error of the controlled variable, the temperature $T_{out}$. The manipulating variable, the oil flow, needs a zero centered scaling funtions for the change of control. The change of irradiation, which is used in the asymmetrical action, and the working point corrections based the fluctuations of $I_{eff}$, $T_{out}$ and $F$ are all handled with linear definitions. The threshold values of PBA are defined from the definitions of the controlled variable. The adaptive scaling is based on the correction factor, which is delinguistified from the working point: $c_F(k) = f(wp_i(k))$. A zero centered scaling function is needed for $c_F$. The fast increase and too high values of the temperatures are handled with two linear and one asymmetric
linear functions. Totally 38 parameters are used for the membership definitions and limits of the LE controller in the solar application (Tab. 11).

In this case, the coefficients listed in Table 5 are in default values: feedforward control, control power and cumulative rate are not used; default values are used for $K_p(i, j) = K_i(i, j) = 1$; derivative control is not used; condition and stress indices are not used; trend analysis is not integrated in the control. The adaptive scaling is based on the working point, only $w_{wp} = 1$ is needed in Table 6. There is only one manipulating and one controlled variable, and $w_{ij} = 1$ for them. The minimum working point and the setpoint are given by the operator. Coefficients $(c_A)_j$ and $(c_A)_j$ are used to activate, scale and close the corresponding actions.

The trend analysis of the errors, changes and the working point variables is based on the membership functions: only additional parameters are the window sizes and weight factors. For other variables, like the oil flow, also the membership definitions are needed. Further tuning can be done for the coefficients of the PI-type LE controller, and working point variables, and three for the additional change of control $\Delta u^{CH}(k)$.

**Lime kiln**

The LE control system combines both FF and FB controllers strengthened with intelligent analysers and predefined adaptation models. All the properties are combined in the same controller structure, where special modes operate simultaneously, each one activated and closed smoothly (Järvensivu et al. 2001). With this efficiently operating controller, it is finally possible to optimise the kiln operation and reduce safety margins (Juuso 2004c). Temperature excursions in the hot end became very uncommon, which reduces the thermal stress and probably increases the refractory life. (Järvensivu et al. 2001) The production capacity was increased, the energy efficiency was improved and heat energy consumption decreased. The major quantifiable benefits from the ecological point of view were a decrease in the mean of the total reduced sulfur emissions, and a reduction of short-term emission peaks. (Järvensivu 2003)

The LE control system includes four FF controllers (Tab. 10), which based on three variables, have seven full membership definitions requiring 35 parameters. The FB controller (93) requires 36 parameters, since there are four full and four zero centered membership definitions. Later the FB controllers were realised for two controlled variables, hot-end and cold-end temperatures, and two manipulating variables, saw dust feed rate and fuel oil flow. Both controllers, which are PI-type LE controllers, require 20
parameters, and additional four parameters for the PBA action. Adaptive scaling is based on the production rate, the draught fan speed and the cumulative rate of control actions \( \delta u_i \), which requires four additional parameters. The control power is not used for the fuel oil. The upper and lower limits are defined by the corresponding limits of the support areas. Most coefficients shown in Table 11 are in the default value: the FF controllers have totally eight active coefficients; the coefficients of the FB controllers are defined by the directions of the interactions. The dynamic window parameters \( \lambda_i^- \) and \( \lambda_i^+ \) are needed for the outputs of the FF controllers. The correction factor \( c_F \) requires four parameters.

**Water treatment**

LE controllers have been successfully implemented at a mill (Piironen et al. 2008). The adaptive FB controller of the faster effecting chemical reacts efficiently to the change of the water quality and to the halving of incoming flow: the setpoint is kept, and there is no offset. The FF controller of the slowly affecting chemical is needed for fast changes.
Dynamic LE models have been used for developing, testing and tuning the controllers in changing process conditions without disturbing the process. Simulation made the implementation faster, and no re-tuning of control parameters was needed. (Joensuu et al. 2005)

The water quality index, which is used in the FF controller and in the adaptation of the FB controller, needs full membership definitions for four variables, i.e. 20 parameters, see Section 7.2.2. Totally 20 parameters are needed for these variables. The FB controller is based on three zero centered membership definitions (error, change of error, change of control), requiring 12 parameters. The turbidity is used in the quality index, adaptation and FF control: the membership definitions are the same for measurements and setpoints in all time steps, and the parameters are taken into account in the quality indicator. The flow rate of incoming water, which is used in the FF controller, needs five parameters. The correction factor $c_F$ requires four parameters.

The total number of parameters of the membership definitions is 41 (Tab. 11). The coefficients are fine-tuned for the water quality index and the FF and FB control laws.

7.4 Detection of operating conditions

The first applications of LE-based diagnostics were linked with fuzzy set systems (Juuso 1994a): symptoms were generated from a fuzzy controller and a set of linguistic equations by using the complement of the feasible range (6). The overall model-based diagnostical process analysis (MDPA) was introduced in (Juuso 1997) and extended to a CBR-type approach in a paper machine application (Juuso et al. 1998a). Parameters of the scaling functions were analysed from data (Tab. 12). Fuzzy set systems and linguistic equations were applied to functional testing in electronics manufacturing (Komulainen et al. 1997). Membership functions were based on domain expertise, and the interactions for the corresponding LE system were analysed from fuzzy rules (Tab. 12). Automatic membership function generation from human inspector actions was used in solder joint inspection to find proper thresholds (Wei et al. 1999).

Case-based LE models were used for condition monitoring in (Juuso et al. 2004). The same LE methodology provided good results in detecting the fluctuations of flavour ingredients in brewing, in predicting web break sensitivity in paper machines and in condition monitoring of machines (Juuso & Leiviskä 2005).

The condition monitoring applications are important in the current methodological development during the research phases IV and V. Dimensionless indices obtained
Table 12. LE applications in detecting operating conditions.

<table>
<thead>
<tr>
<th>Application</th>
<th>Scaling</th>
<th>Interactions</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Process</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brewing</td>
<td>pilot data</td>
<td>data -&gt; LE</td>
<td>normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>special cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>off-line</td>
</tr>
<tr>
<td>Paper machine</td>
<td>process data</td>
<td>data -&gt; LE</td>
<td>break categories</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>web break sensitivity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>on-line &amp; off-line</td>
</tr>
<tr>
<td>Wastewater treatment</td>
<td>process data</td>
<td>data -&gt; LE</td>
<td>performance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>interpolation</td>
</tr>
<tr>
<td><strong>Testing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional testing</td>
<td>expertise</td>
<td>fuzzy rules</td>
<td>normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-&gt; LE</td>
<td>several faults</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>on-line</td>
</tr>
<tr>
<td>Inspection</td>
<td>inspection data</td>
<td>fuzzy rules</td>
<td>normal / faulty</td>
</tr>
<tr>
<td></td>
<td>on-line</td>
<td></td>
<td>on-line</td>
</tr>
<tr>
<td><strong>Condition monitoring</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centrifuge</td>
<td>CM data norms</td>
<td>data -&gt; LE</td>
<td>several faults</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>condition index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>very fast rotation</td>
</tr>
<tr>
<td>Lime kiln</td>
<td>CM data norms</td>
<td>data -&gt; LE</td>
<td>several faults</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>condition index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>slow rotation</td>
</tr>
<tr>
<td>Paper machine</td>
<td>CM data features</td>
<td>data -&gt; LE</td>
<td>condition index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>trend index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>severity</td>
</tr>
<tr>
<td>Test rig</td>
<td>CM data features</td>
<td>data -&gt; LE</td>
<td>normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 faults</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>multisensor</td>
</tr>
<tr>
<td>Water turbine</td>
<td>CM data norms</td>
<td>data -&gt; LE</td>
<td>cavitation index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>stress</td>
</tr>
</tbody>
</table>

by comparing each feature value with the corresponding value in normal operation provide useful information on different faults, and even more sensitive solutions can be obtained by selecting suitable features (Lahdelma & Juuso 2007). Generalised moments and norms include many well-known statistical features as special cases and provide compact new features capable of detecting faulty situations (Lahdelma & Juuso 2008a, 2011a,b). Vibration measurements provide useful data for this research.
Intelligent indices presented in Section 5.5 are used for detecting operating conditions. Condition and stress indices were introduced in condition monitoring. The first application was done for cavitation studies (Juuso & Lahdelma 2006), where the original indices were first improved by using generalised moments (Lahdelma & Juuso 2008b) and later with generalised norms (Juuso & Lahdelma 2008). In lime kiln applications, the solution was first based on histograms (Lahdelma & Juuso 2006) and later improved by generalised norms (Juuso & Lahdelma 2008). The new scaling approach (Juuso & Lahdelma 2010) improved results considerably and facilitated intelligent trend analysis (Juuso & Lahdelma 2011a).

7.4.1 Stress indices

Stress indices $I_S^{(α)}$ are calculated a weighted sum of scaled features, see (82). Very low stress corresponds to value -2 and not allowable high stress to value 2. High-stress operating conditions should be avoided in order to keep the process in good condition. Feature selection and nonlinear scaling are essential steps in developing stress indices from measurement signals (Juuso & Lahdelma 2006, Juuso et al. 2007, Lahdelma & Juuso 2008b, Juuso & Lahdelma 2010). Large deviations from the normal operating area can also cause stress, e.g. examples in the solar power plant: the pump if a high flow is required already in low temperature, or the oil tubes if a high temperature difference is required (Juuso 2012c).

Water turbine

A Kaplan water turbine, which has sleeve bearings and four blades, has been used in the cavitation research. The cavitation index developed in (Juuso & Lahdelma 2006) is based on the nonlinear scaling of two features: peak height and the fraction of the peaks exceeding the normal range $[-3σ_α, 3σ_α]$ obtained from the signal $x^{(α)}$, $α = 1, 3$ and 4. The velocity $x^{(1)}$ was replaced by the acceleration $x^{(2)}$ in (Juuso et al. 2007). The generalised central absolute moment about $\bar{x}^{(α)}$ was introduced in (Lahdelma & Juuso 2008b):

$$
τ_{σ^p} M_{0}^{(α)} = \frac{1}{N(σ_α)^p} \sum_{i=1}^{N} |x_i^{(α)} - \bar{x}^{(α)}|^p,
$$

where the real number $α$ is the order of derivation, the real number $p$ is the order of the moment, $τ$ is the sample time (s), and $\bar{x}^{(α)}$ and $σ_α$ the mean and the standard deviation,
respectively, calculated for the signal $x^{(a)}$. The moment $\mathcal{M}_2^{(a)}$ = 1, and the moment $\mathcal{M}_4^{(a)}$ correspond to the kurtosis of the signal $x^{(a)}$. Short sample times were found to be good in (Lahdelma & Juuso 2008b). The generalised norm $\|\mathcal{M}_p^{\alpha}\|$ defined by (81) was compared with a knowledge-based cavitation index in (Lahdelma & Juuso 2008a). Short sample times and order $p$ between 2 and 4 provide the best candidates for a single feature to operate throughout the whole power range. The order $p$ of the norm was chosen to be 2.75 and sample time $\tau = 3$ s which provides a good balance between low and high power ranges, since sample time has a strong effect in the low power range.

![Fig 41. The cavitation index $I^{(4)}_S$ for a Kaplan water turbine (Juuso & Lahdelma 2010, published by permission of BINDT).](image)

The cavitation index (Fig. 41) is obtained by

$$I^{(a)}_S = f^{-1}_\alpha(\text{relative max}(\|\mathcal{M}_p^{\alpha}\|)),$$

which is based on the relative max(\$\|\mathcal{M}_p^{\alpha}\|\$), which is obtained by comparing max(\$\|\mathcal{M}_p^{\alpha}\|\$) at each power with max(\$\|\mathcal{M}_p^{\alpha}\|\$) the maximum norm at 15 MW. The index $I^{(1)}_S$ based on the velocity signal has the lowest $R^2$ value, and also the classification result is the worst of these three. The coefficient of determination $R^2$ was 0.897 for $I^{(4)}_S$. Index $I^{(3)}_S$ has the highest $R^2$ value, but the index $I^{(4)}_S$ has the best classification results. Good operating conditions, clear signs of cavitation and very strong cavitation were detected. Variation with time can be handled as uncertainty by presenting the indices as time-varying fuzzy numbers. The classification limits can also be considered fuzzy. The reasoning system will produce the degrees of membership for different cases.
The new nonlinear scaling approach was used in (Juuso & Lahdelma 2010) for the relative \( \max (|1^3M_1^{2.75}|) \). The resulting corner points satisfy the conditions of the ratios (21). The derivative of the scaling function is chosen to be continuous: only \( c_j \) is changed by (25) in the corner points (Fig. 42). The new scaling function improves the sensitivity of the cavitation index for short periods of cavitation (Fig. 41).

### 7.4.2 Condition indices

A condition index (82) can be based on several features: high values mean low condition indices. Generalised norms are used in two levels (Fig. 43): feature values are obtained from vibration signals, and the resulting features are scaled to the range of \([-2, 2]\). The methodology has been tested in a very slowly rotating lime kiln (Juuso & Lahdelma 2007) and in a very fast rotating centrifuge (Lahdelma et al. 2006). Different faults and damages can be detected with specific combinations of the scaled features. The features, which are obtained by using the orders \( \alpha \) and \( p \) for derivation and generalised norms for analysing the signals in a specific frequency range, replace the earlier used histogram based features. (Juuso & Lahdelma 2008, 2010)
Lime kiln

Condition indices (82) have been used for the supporting rolls of a lime kiln (Lahdelma & Juuso 2006, Juuso & Lahdelma 2007). The condition index is defined by the standard deviation $\sigma_\alpha = \text{max}(||^{15}M_\alpha^\| \|)$ and five bins of the histograms $F_\alpha^k$, $k = 2, \ldots, 6$: 

$(k = 2) |x^{(\alpha)}| < 2\sigma_\alpha$, $(k = 3) 2\sigma_\alpha \leq |x^{(\alpha)}| < 3\sigma_\alpha$, $(k = 4) 3\sigma_\alpha \leq |x^{(\alpha)}| < 4\sigma_\alpha$, $(k = 5) 4\sigma_\alpha \leq |x^{(\alpha)}| < 5\sigma_\alpha$, and $(k = 6) |x^{(\alpha)}| \geq 5\sigma_\alpha$. The standard deviation and $F_\alpha^k$, $k = 5$ and 6 are related to faulty situations, and large values for the fractions $F_\alpha^k$, $k = 2$ and 3 and 4 are obtained in normal conditions. The weight matrix $[-2 +1 +1 +1 -1 -1]$ is based on expertise. The same model structure is used for $x^{(3)}$ and $x^{(4)}$. The condition index obtained from the signal $x^{(4)}$ is already suitable for practical applications. The index obtained from signal $x^{(3)}$ requires further tuning. All the supporting rolls can be analysed using the same system. The velocity signal hardly shows any difference between a serious surface problem and an excellent condition. (Juuso & Lahdelma 2007)

The norms $\text{max}(||^{15}M_\alpha^\| \|)$ and $\text{max}(||^{15}M_4^{\alpha \| \|})$ obtained from $x^{(4)}$ are highly sensitive to faulty situations (Lahdelma & Juuso 2008a). The new estimation method was used for the norms, and the resulting scaling functions, which are locally linear at the centre point (Fig. 43), provide good results (Fig. 44). Surface damage and alignment problems are clearly visible, and an early indication of the friction increase is also perceivable. The data set covers the following cases: (1) surface problems, (2) good conditions after grinding, (3) misalignment, (4) stronger misalignment, (5) very
The results are consistent with the vibration severity criteria, see Table 17. The overall condition index $I_C$ can be defined as an average of the scaled features, see Figure 44. All the very good cases are close to the lower corner $[-2, -2]; I_C^{(4)} < -1$. The index is below zero for the good cases and cases with small fluctuations. These cases are clearly usable. For still acceptable cases, the index is below one. All the faults and clear friction cases are in the area where i.e. the severity level is 'not acceptable'. The upper corner $[2, 2]$ is related to the case where the temperature was reduced by cooling. The very high signal levels are caused by noise, but the hidden faults can be...
detected by stopping the cooling for a moment. The abnormal feature levels mean that some additional analysis needs to be carried out. Special cases, such as the effects of the rotation counter and jingles, can be detected by comparing the indices obtained from the maximum values to those obtained from the mean values. All the condition levels can be defined with fuzzy membership functions. Fuzzy rules or specialised condition indices could be developed for different types of fault. The corner points can be used as warning and alarm limits. (Juuso & Lahdelma 2010)

Very fast rotating rolling bearings

Detecting bearing faults and unbalance in very fast rotating rolling bearings was first based on standard deviations calculated for the signal $x^{(4)}$ in three frequency ranges (Lahdelma et al. 2006). The faults are correctly detected (Fig. 45). The same measurements were used in (Lahdelma & Juuso 2009) to test the norm (81): each calculated value of the norm was divided by the average of the values of the same norm in good conditions. The sample time, $\tau = 3.9 \text{ ms}$, corresponded to two rotations when the rotation frequency was 525 Hz. In this case, very high frequency ranges were also used for the vibration measurements. The condition index (82) is obtained as a weighted sum of the scaled norms of the signal $x^{(4)}$ calculated for three frequency ranges $k = 1, 2,$ and 3: $10 - 1000 \text{ Hz}$, $10 - 10000 \text{ Hz}$, and $10 - 50000 \text{ Hz}$.

![Classification results for fast rotating bearings, the rotation frequency was 525 Hz, modified from Lahdelma & Juuso (2007).](data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAABAgAAAQcAMAAACUfGoAABAgAEl2cnR hf8AAADHklEQVR42m38zZ/zA8/wAAACUfGoAAAAABJRU5ErkJggg==)

The results shown in Figure 45 are also promising for the identification of the fault. The sequential algorithm introduced in (Lahdelma et al. 2006) can be generalised to form:

192
– Calculating the condition index.
– The condition is normal if \( I_C^{(\alpha)} < 1.5 \).
– There is an outer race fault in the bearings if \( I_C^{(\alpha)} < 0 \).
– The condition is unbalance if the standard deviation for the low frequency range is very high.
– Otherwise the condition is inner race fault in the bearings.

The minimum of the index \( I_C^{(\alpha)} \) is 2 which is achieved when all the features are at the lowest level and all the weight factors \( w_k = -1/3 \). The weight factors are transformed to obtain the general form (82). The faults were detected correctly in (Lahdelma et al. 2006, Juuso & Lahdelma 2008) by using the standard deviation \( \sigma_4 \). The applicability of different norms has been studied in (Lahdelma & Juuso 2009). Unbalance can be clearly detected with norms based on all the signals \( x^{(\alpha)} \) in the frequency range 10-1000 Hz. The norms \( \| T M_p^4 \| \) based on the signal \( x^{(4)} \) provide the best results in all the frequency ranges. The high range has the best indication results: all the sensitivity values are high, and the values for different faults are at specific levels. In the low frequency range, the order \( p \) has a strong effect on the indication of the inner and outer race faults, especially when \( p \) increases from 1 to 3. It is interesting that the signal \( x^{(4)} \) provides better sensitivities for the outer race faults in the range 10-1000 Hz than in the range 10-10000 Hz, if the order \( p \) is high enough. This means that the number of required signal points can be reduced to 1/10 when the signal \( x^{(4)} \) is used.

Good results can be obtained e.g. by using the rms value of velocity and the peak value of signal \( x^{(4)} \). The absolute mean is sufficient in many cases. (Lahdelma & Juuso 2009) For the early detection of the faults, the overlapping levels make the identification more difficult than in Figure 45.

### 7.4.3 Case detection

Case detection is needed if there are several possible faults, which can also occur simultaneously (Fig. 44). The performance index can be obtained by using the deviation from the normal model (Juuso & Kronlöf 2005) or from the sum of test measurements (Gebus et al. 2009). Complex systems, e.g. paper machines, can have similar risk levels in very different cases (Juuso & Ahola 2008). Complexity of the models can be reduced with intelligent indices (Juuso 2010b) and feature selection (Juuso et al. 2010).
Continuous brewing

A model-based LE system for the monitoring and diagnostics of fermentation and flavour formation in an immobilized yeast fermentation process has been presented in (Juuso & Kronlöf 2005). Some fluctuations of flavour ingredients were detected for only short time periods as the process was stable. The time span of these fluctuations was usually too short for the development of specialised models. A more detailed model with seven equations for all the prereactor measurements indicates clear differences in the end of the test period. All the equations have a high fuzziness and the degree of membership for the model goes to very low values. By creating grounds for the prediction of quality factors the models increase options to control the product quality in different cases. The modular model library is expandable.

Functional testing

A fuzzy rule-based approach and linguistic equations were applied to functional testing in electronics manufacturing (Komulainen et al. 1997). In the LE approach, defects are identified by linear interactions among linguistic variables (Gebus & Juuso 2002). The number of variables in each equation is limited to avoid overfitting caused by redundancy. For each variable, functional tests are designed in such a way that an area is considered as defective if the measured variable crosses a certain control limit. The centre points, which correspond to the normal system conditions, were in (Gebus 2006) defined by the measured average of each variable during the steady-state of the system. The functions are highly asymmetric in respect to the origin, e.g. measurements shifted towards the lower control limit and a 'low voltage' can have a much bigger effect on the linguistic variable than a 'high voltage' and will therefore be detected much faster.

In functional testing of a printed circuit board (PCB), the LE model (30) is in steady-state form, i.e. \( n_j = 0 \). The bias term \( B_i = 0 \) and the coefficients \( A_{ij} \) are limited to the values 0 and 1. The models are specific to the areas on the PCB, i.e. \( A_{ij} = 1 \) only if \( X_j \) is relevant to the area \( i \) on a PCB. Areas are defined according to their functionality. The area is OK if every variable \( X_j \approx 0 \). All the scaled variables should be within the control limits: \( |X_j| \leq 1 \). A deviation \( \sigma \) is calculated by

\[
|X_1| + \cdots + |X_m| < 1 + \eta \frac{m-1}{m} \sum_{j=1}^{m} \sigma_j,
\]

(100)
where \( \sigma_j \) is the standard deviation of the variable \( X_j \), and \( \eta \) is a coefficient that depends mainly on the amount of available data and the spread of the variables. A good case would be to have a lot of data for many variables with a similar standard deviation. In all other cases, coefficient \( \eta \) can be used to adjust the system if it generates, for example, false alarms. (Gebus 2006, Gebus et al. 2009)

**Solder joint inspection**

The on-line membership function generation system was an essential part of the Fuzzy Expert System for X-ray Inspection presented in (Wei et al. 1999). The training data included the solder joint measurement values and human inspectors verifying results. Different membership functions were obtained for different package types, joint types and defect types. The system can successfully learn the knowledge from human inspectors by analysing on-line data and then use the knowledge to support decision making based on fuzzy rules. This expert system has been tested in real world mass production environment.

**Test rig**

A multisensor approach has been used for fault diagnosis in a test rig that consists of an electric motor and a transmission between two axes with SKF single row ball bearings 6002. Two sensors measured axial vibration and five accelerometers radial vibration in the vertical direction. LE models were developed in (Juuso et al. 2004) for the normal operation and nine fault cases including two levels of rotor unbalance, a bent shaft, three levels of misalignment and three bearing faults. Classification is based on the degrees of membership developed for each case from the fuzziness defined by (40). The classification results of the experimental cases were very good and logical.

Each sensor specific equation is based on the rotation speed, five features obtained from vibration measurements and a bias term. A large number of equations and features are not necessarily needed. The number of variables and sensors required to diagnose each fault can be reduced considerably by selecting the most sensitive features by comparing the health indices of individual features. The health indices \( (I_{HI})_L \) combine several feature specific indices:

\[
(I_{HI})_L = \sum_{i=1}^{n_r} \frac{\left( \| x^r(\alpha_i) \|^{p_r} \right)_0}{\| x^r(\alpha_i) \|^{p_i}},
\]  

(101)
were calculated by dividing the reference value \( ||x^{(n)}||_{p_1} \) by the feature value \( ||x^{(n)}||_{p_2} \).

Good Rolling Inner Outer
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
SOL( x p(3) )
SOL( \text{s} \text{rms} )
SOL( \gamma_{2,4} )
element
SOL
Bearing fault
Sensor 7
race race

(a) Bearing faults.

(b) Misalignment.

(c) Unbalance.

Fig 46. Health indices of a test rig, modified from Juuso et al. (2010).

Sensitivities of several features compared with health indices (Fig. 46) are used in feature selection: three sensors are needed from seven, and totally seven features are used to calculate eight indices (Tab. 13). The rms velocity, \( v_{\text{rms}} = ||M_1^2||_2 \), is calculated for Sensors 1 and 2 with specific frequency range. The rms acceleration, \( a_{\text{rms}} = ||M_2^2||_\infty \), and the peak value of jerk, \( x_p(3) = ||M_2^\infty||_\infty \), are obtained for Sensors 2 and 7. Also the kurtosis of the acceleration, \( \gamma_{2,4} \), is used for Sensor 7.

Faults are analysed in two groups: (1) bearing faults and misalignment need four features, and (2) unbalance and bent shaft five features. Two features are used in both groups. The first stage is to detect with the indices \( (I_{H})_1 \), \( (I_{H})_5 \) and \( (I_{H})_6 \) if everything is in ‘good condition’. If both \( (I_{H})_1 \) and \( (I_{H})_5 \) are less than one, there is a fault which belongs to the group of misalignment and bearing faults. The fault type is identified by comparing indices \( (I_{H})_2 \), \( (I_{H})_3 \) and \( (I_{H})_4 \) (Fig. 46(a)). The strength of misalignment can be estimated by \( (I_{H})_5 \) (Fig. 46(b)). If \( (I_{H})_7 \) is less than one, the fault is either unbalance or a bent shaft. More information is obtained by \( (I_{H})_4 \) and \( (I_{H})_8 \). The strength of unbalance can be estimated by \( (I_{H})_8 \) (Fig. 46(c)). The classification results are in line with the previous studies with the same data set. (Juuso et al. 2010)
Table 13. Feature selection.

<table>
<thead>
<tr>
<th>(l_1)</th>
<th>Sensor 1</th>
<th>Sensor 2</th>
<th>Sensor 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>((v_{rms})_2)</td>
<td>((v_{rms})_1)</td>
<td>(a_{rms})</td>
</tr>
<tr>
<td>1</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>•</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>•</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>•</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>8</td>
<td>•</td>
<td>•</td>
<td></td>
</tr>
</tbody>
</table>

Wastewater treatment

Wastewater treatment within Finnish pulp and paper industry is most commonly done in an activated sludge plant, which is a complex biological process, where several physical, chemical, and microbiological mechanisms simultaneously affect purification results. Limits of the emissions are defined by authorities. A lot of process measurements are available, but measurements do not include sufficient information on the special features of the influent nor on the microbial composition of the sludge. The populations of microorganisms are highly important, e.g. sludge bulking cause especially poor treatment efficiency results when biosludge escapes from secondary clarification (Fig. 47).

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 indices (78) and (79) to detect these changes (Juuso et al. 2009). The nonlinear scaling approach presented above were used in (Juuso 2010b) for the variables obtained from process measurements and laboratory analysis. Interpolation was needed for some variables. The balance between the load and the nutrients was evaluated by the difference of the corresponding scaled values. The worst case with low reduction and settling problems arise, when there multiple warnings and alarms (Fig. 47). Correspondingly, good reduction and very good settling was achieved when there were very few warnings. As it
Fig 47. Treatment result of an activated sludge treatment plant: reduction of chemical oxygen demand (COD) and diluted sludge volume index (DSVI), see (Juuso 2010a).

takes some time to lose good conditions and recover from problematic conditions, the intelligent indices are useful for process control.

**Web break sensitivity**

The web break sensitivity stands for the possibility of the paper web to break, predicting the amount of breaks during one day. The main area of interest in indicator development is the paper making process before the actual paper machine. This includes also the short circulation and the wet end of the paper machine. In this area, the paper making process is typically nonlinear with many and long delays, numerous process feedbacks at several levels, several closed control loops, interactions between physical and chemical factors, and various unmeasurable factors. Delays and interactions change in time and with process conditions. (Ahola et al. 2004)
The web break indicator shown in Figure 48 uses CBR and actual measurements from the paper machine. The LE modelling was done for variable groups including up to five variables. In order to detect different operating conditions each model set should be accurate in its own case but at the same time its fit to other cases should much worse. This methodology is suitable for detecting the most important variables and variable groups for separate cases. In the web break indicator, the degree of membership for each case is obtained by combining the membership degrees and weight factors of individual equations by (83), and the final selection of the active cases and the corresponding web break category is based on fuzzy reasoning (Ahola & Leiviskä 2005, Ahola 2005). Maximum is used to select one case at a time, since the cases of a certain category are dissimilar (Fig. 48). There are many ways to run the process on average break sensitivity, and much less ways to run it better or worse. The testing of the indicator has been done simultaneously on-line and off-line. New versions have first been tested with
the simulator. If new version improved the performance, the latest version was installed for the on-line testing. (Juuso & Ahola 2008)

7.4.4 Trend analysis

Intelligent trend indices were introduced for wastewater treatment (Juuso et al. 2009) and extended to the analysis of trend episodes in (Juuso 2011a). However, trends of the fuel feed were used in the LE models of the lime kiln (Juuso 1999b), and the intelligent indicators of the fast changes (94), (95) and (96) are used in the solar power plant (Juuso & Valenzuela 2003, Juuso 2012c). In condition monitoring, trends were analysed from rms velocity, \( v_{rms} = x_{rms} \), measurements of two paper machines (Juuso & Lahdelma 2011a): the resonance of the press section and resin problems of a press roll in the felt washer are discussed in (Lahdelma 1977). In both cases, the \( v_{rms} \) measurements are scaled with the nonlinear scaling method described above, and the trend index (78) is calculated using appropriate short and long time periods. (Juuso & Lahdelma 2011a)

The thresholds \( \epsilon^{+} = \epsilon^{-} = \epsilon^{+2} = \epsilon^{-2} = 0.5 \) are kept by using case specific weight factors \( w_{T1}^{j} \) and \( w_{T2}^{j} \). Trend are followed with trend episodes and deviation indexed (Figs. 49 and 50).

Paper machines

Periodic vibration measurements from a press roll in the felt washer provided useful data for detecting the gradual increase of the resin problem at an early stage (Juuso & Lahdelma 2011a). Operating conditions had a major effect on the resonance of the press section. In the first case, the machine speed was reduced 4%, and a breakdown and an additional stoppage were avoided. The machine was operated with reduced speed for two weeks (Fig. 49(a)). The same or lower vibration level was kept for one week. Measurements are quite infrequent in this case. The trend index is zero at the beginning, the increase corresponding to \( I_{T}^{j}(k) \approx 1 \), and finally the index reaches the value 1.5 (Fig. 49(b)). The short and long periods are two and four measurements, respectively. The weight \( w_{T1}^{j} = w_{T2}^{j} = 1.5 \). The episode analysis starts from normal and moves to a concave upward increase, then returns to normal and again starts to speed up at the end (Fig. 49(c)). The roll was changed when the concave upward area was reached. The change is very strong. The deviation index \( I_{D}^{j}(k) \) is around or below zero a long time at
the beginning (Fig. 49(d)). Higher values after Day 26 give a warning, and the highest values at the end of the period can be regarded as alarms.

In the second case, the resin problems of a press roll in the felt washer were seen as a typical trend, and different order derivatives were used in severity assessment (Juuso & Lahdelma 2011a). The scaled values increase from -2 to 2: at the beginning the increase is slow but becomes faster at the end of the period (Fig. 50(a)). The trend index $I^T_j(k)$ is zero at the beginning, the slight increase corresponds to $I^T_j(k) \approx 1$, and finally the index reaches the value 1.5 (Fig. 50(b)). The short and long periods are three and nine measurements, respectively. The weight $w_{T1} = w_{T2} = 1.5$. The episode analysis starts from normal, gives the first warnings about the increase of $I^T_j(k)$, then a slight linear increase, which first turns into a concave upward increase and then returns to a quite strong linear increase that again starts to speed up at the end (Fig. 50(c)). The roll was changed just before the concave upward area would have been reached. The index $I^D_j(k)$ is around or below zero a long time at the beginning (Fig. 50(d)). Higher values after
Day 84 give a warning, and the highest values at the end of the period can be considered as alarms.

**Wastewater treatment**

Scaled values of the treatment results (Fig. 47) and other variables are intelligent indices, which can be used in trend analysis (Fig. 51). Changes in operating conditions are detected at an early stage (Juuso & Laakso 2011). The operating conditions are classified by oxygen, temperature and flow. As the flow comes from the production process, it cannot be made suitable for the treatment process. The process stoppages can be seen as fast changes of trend episodes. The normal levels are the best both for the temperature and the oxygen: too high and low temperatures affect the biomass; too low oxygen levels are harmful and too high levels mean excess energy consumption. The period of low temperature is during the long stoppage of the process, and high temperatures are related to high flow in summer time.
Load and nutrient should be balanced since both an exceptionally high load and excess nutrients cause problems. The data set of three years and eight months starting from January 2005 includes examples of these problems. For most time during the first year, the feed of nutrients was on a high level, and then it was decreased to a low level during the second year (Juuso 2010a). During the first autumn, there was a high peak of nutrient feed when the load was fairly low. The nutrient feed returned to normal by the
spring. However, later the load was strongly increased, which have caused an opposite situation. Averages obtained from longer time periods have been used in calculating the load nutrients balance. The changes of the balance are clearly seen in the deviation index (Fig. 51(a)): the low levels detected when $250 < t < 450$ were corrected before $t = 600$, but after that the increase of the load was not balanced when $630 < t < 680$. Too low nutrition levels were kept until $t = 950$. Better results were obtained only when the load went down.

Treatment results are here analysed by comparing the COD reduction and DSVI values: the deviation index should be high for the COD reduction and low for the DSVI. The load and the nutrients should be balanced, i.e. the difference between the load and nutrient levels should be close to zero (Fig. 51(b)). A too high nutrient level compared to the load causes poor settling seen as an increase of the DSVI, which continues as an oscillating behaviour when $250 < t < 480$ (Fig. 51(f)). The treatment result is on the normal level, but the deviation index already indicates reduced performance (Fig. 51(d)). The treatment result improves when the nutrient feed corresponds better the load level. Even high flow around $t = 600$ is acceptable if it does not cause a considerable increase in the temperature. Very good operation achieved when $480 < t < 620$ and again when $650 < t < 750$. These periods end when the nutrient balance becomes worse (Fig. 51(b)).

7.5 Scheduling and decision making

The LE-based decision support was started in the early applications (Juuso & Leiviskä 1992, Juuso et al. 1993). Domain-specific knowledge can be coupled with simulation models in the evaluation of alternatives is based on fuzzy and linguistic simulation. In combined simulation and expert systems, vagueness is gradually reduced. The reliability of the results is clearly seen and it is easier to concentrate on the most important areas, even with partial differential equations. More details about the uncertainty processing in detailed simulation models are presented in (Juuso 1994b). The analysis is based on the development of membership functions and definitions (Juuso et al. 1993, Frantti & Juuso 1996). The models can be deterministic (Juuso et al. 1993) and rulebases symmetric (Juuso & Jagdev 1999, 2007). Links to the knowledge-based information are essential (Juuso & Lahdelma 2011b, Juuso 2012a). The condition and stress indices of process equipment are useful in scheduling and decision making (Juuso & Lahdelma 2009).
Decision support for tactical pricing

A hybrid knowledge-based system was developed for managerial decision making emphasising tactical pricing decisions (Juuso et al. 1993). The aim was to combine qualitative and quantitative methods for pricing support in industries. The systems are applicable to price problems with several products and competitors. Constraints have an essential effect on the aggregated sets of relations in management applications. The linguistic values of the decision variables are presented by membership functions, which are based on the differential constraints represented by fuzzy numbers. Scenarios for experts are developed with these functions. The simulated response from the regression analysis produces data for the estimation of output variable’s membership functions. The decision making is based on optimisation. The profits or cost are related to the input and output variables of the linguistic model by the objective function. The gradually refining hierarchical sets of membership functions provide an efficient optimisation method for these applications. The refinement is based on the membership definitions as presented in Section 3.1.2.

Demand forecasting

An adaptive, hierarchical fuzzy logic advisory tool was developed for anticipating the demand of transmission products for component purchasing in electronics manufacturing (Frantti & Juuso 1996). To control a very complex and highly nonlinear materials purchasing system, the inference is performed by hierarchical reasoning where linguistic relations in each subsystem are expressed by equations. The membership functions are generated by an on-line system for demand and product order.

Short term scheduling

In the short term scheduling system, linguistic equations are used for handling preferences and priorities in selecting production tasks from several feasible alternatives (Juuso & Jagdev 1999, 2007). The main idea is to select clusters and items in such a way that the risk of delay in the end of the scheduling period is minimised. The scheduler consists of three phases: (1) allocation, (2) selecting clusters to the active queue and (3) selecting items for processing. Factory topology, resource requirements and product demand may change within the scheduling period, and the scheduler returns to the
previous phases if significant changes take place. Allocation is performed for clusters in each subset by the longest processing time first rule. For detailed scheduling, the linguistic equation approach provides a method for a smooth adaptation of scheduling rules to the changing operation conditions and an efficient method in representing the preferences and priorities from process requirements and available resources.

With linguistic equations this set of rules can be represented by a single equation: the overall importance = importance of the variable + usefulness of the filter, where the original linguistic labels correspond to the real values -2, -1, 0, 1 and 2. According to test runs with combined fluctuating problems, the intelligent scheduling approach is a robust solution for short term scheduling (Juuso & Jagdev 2007).

Performance indices

Performance measures can be handled with nonlinear scaling to produce information in natural language, e.g. the OEE improvements shown in Table 14 can be based on following classification: [0.25, 0.75) slight, [0.75, 1.25) good, [1.25, 1.75) very good, and ≥ 1.75 excellent improvement (Juuso & Lahdelma 2011b, 2013). The feasible range defined by 48.0%, 62.6%, 76.4%, 86.6% and 95.0% was analysed from the OEE values presented in (Willmott 2010).

Table 14. Examples of OEE improvements, modified from Juuso & Lahdelma (2011b).

<table>
<thead>
<tr>
<th>Process</th>
<th>From</th>
<th>To</th>
<th>Improvement</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steel Plant</td>
<td>-0.18</td>
<td>1.38</td>
<td>1.56</td>
<td>Good</td>
<td>Very good</td>
</tr>
<tr>
<td>White Goods</td>
<td>0.24</td>
<td>1.15</td>
<td>0.91</td>
<td>Good</td>
<td>Very good</td>
</tr>
<tr>
<td>Automotive</td>
<td>-2.00</td>
<td>-0.11</td>
<td>1.89</td>
<td>Excellent</td>
<td>Good</td>
</tr>
<tr>
<td>Flour Mill</td>
<td>0.93</td>
<td>1.74</td>
<td>0.81</td>
<td>Good</td>
<td>Very good</td>
</tr>
<tr>
<td>Chemical Plant</td>
<td>0.52</td>
<td>2.00</td>
<td>1.48</td>
<td>Very good</td>
<td>Excellent</td>
</tr>
<tr>
<td>Filling Line 1</td>
<td>-1.53</td>
<td>0.83</td>
<td>2.36</td>
<td>Excellent</td>
<td>Poor</td>
</tr>
<tr>
<td>Filling Line 2</td>
<td>-0.62</td>
<td>0.33</td>
<td>0.95</td>
<td>Good</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Packing Line 1</td>
<td>-0.76</td>
<td>1.04</td>
<td>1.80</td>
<td>Excellent</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Packing Line 2</td>
<td>-1.87</td>
<td>-0.11</td>
<td>1.76</td>
<td>Excellent</td>
<td>Poor</td>
</tr>
</tbody>
</table>

The explanations of numeric values are case specific, e.g. the ranges of the OEE values are taken into account when comparing the performance within a certain industrial sector. Translation of the knowledge-based information obtained from natural language to numeric indices is needed for decision making and optimisation (Juuso 2012a).
Modifiers and membership locations (Fig. 24) are important in this analysis, which returns to developing LE systems from fuzzy set systems.

Power control

Stress indices depend on the condition of the turbine, which can be used in power control (Juuso & Lahdelma 2009). In a simulation study, power control minimises the cavitation risk by dividing the load between three turbines, whose conditions are normal, bad and very good. Knowledge-based cavitation indices are used for load allocation by minimising the overall cavitation index. The solution is easy if all the turbines operate in power ranges, which are free of cavitation. For higher loads, the system increases production in the turbines in such a way that short-term cavitation is minimised. The optimisation procedure can also use the low power ranges of the turbines efficiently. However, the low power ranges in some turbines are selected only when all the turbines cannot operate in the cavitation-free area. The load allocation operates as a feedforward control. Feedback control is needed since the cavitation depends on the flow conditions in the turbine.
8 Discussion

Linguistic equation applications are based on the nonlinear scaling of measurements and features. Novel methodologies developed for data analysis and intelligent analysers introduce new solutions for control and diagnostics. The principles of integrating the LE approach with other methods of computational intelligence are discussed to define hybrid systems, which are useful building blocks of smart adaptive systems.

8.1 LE approach in applications

The linguistic equation (LE) approach is here presented as an integrating methodology in developing smart adaptive systems for complex applications. The key methodology is nonlinear scaling, which builds on the idea of membership function to represent the meaning of variables (Fig. 52). Recursive adaptation is important for developing smart adaptive systems. Adaptive nonlinear scaling and modelling produces intelligent indices, which are used in monitoring and control, detection of operating conditions and decision support systems (DSS). These ideas have been demonstrated within various applications (Fig. 1).

Fig 52. LE based intelligent analysers in application, modified from (Juuso & Leiviskä 2010).
8.1.1 Adaptive nonlinear scaling

Feature extraction from process measurements provides material for the development of intelligent systems (Fig. 52). In the process control, the measurements are available as average values obtained for the application specific sample times. The condition monitoring has usually been based on the root mean square (rms) and peak values of the displacement, velocity and acceleration measurements. The generalised norms (7) and moments (11) create additional possibilities for feature extraction. In the condition monitoring, the norms are calculated for the absolute dynamic part of the signal to accept all the real valued orders. The signal processing and feature extraction by using real order derivatives and generalised norms is discussed in (Lahdelma & Juuso 2011a,b). Advanced signal processing is too broad a topic to discuss here in detail. Interpolation is needed to process the measurements obtained by analysers, e.g. CLA 2000, and especially for the laboratory analyses are less frequent than the sample time. The interpolation is an independent topic, which is not discussed here in detail.

The nonlinear scaling functions are defined by the location, dilation and shape: the location \( c_j \) and the dilation \( \Delta c_j \) are understood in the same way as in the function expansion (102), and the shape factors \( \alpha_j^- \) and \( \alpha_j^+ \) introduce a family of nonlinear functions, which provide an understandable parametric way to handle nonlinear systems. The dilations and shape factors asymmetrical around the operating point: \( \Delta c_j^- \neq \Delta c_j^+ \) and \( \alpha_j^- \neq \alpha_j^+ \). The scaling functions based on two second order polynomials cover versatile definitions, see Figures 10 and 12.

The scaling functions can be understood as basis functions used the function expansion (Ljung 2008):

\[
y_i = \sum_{l=1}^{m_f} w_{il} F_l(\vec{x}) = \sum_{l=1}^{m_f} w_{il} f(\vec{\beta}_l \cdot (\vec{x} - \vec{\gamma}_l))
\]  
(102)

with some basis functions \( F_l(\vec{x}) \), \( l = 1, \ldots, m_f \), is represented by the LE model (35) where the coefficient \( w_{il} \) is defined by the ratio (36); the bias term \( B_i = 0 \). The basis functions \( F_l \) are variable specific functions \( f_{-j}^{-1}(p_j) \), which are defined by the location parameters \( \gamma_j = c_j \) and the dilation parameters \( \beta_j \). The scaling is allowed to be asymmetrical in respect to the location \( c_j \): the functions \( f(p_j) \) are defined by the shape factors \( \alpha_j^- \) and \( \alpha_j^+ \), and the score values \( p_j \) are calculated by using the location \( c_j \) and the dilations \( \beta_j^- = \frac{1}{\Delta c_j^-} \) and \( \beta_j^+ = \frac{1}{\Delta c_j^+} \). The shapes are limited by \( \alpha_j^- \in [\frac{1}{3}, 3] \) and \( \alpha_j^+ \in [\frac{1}{3}, 3] \) to keep the second order polynomials (18) monotonously increasing in
versatile forms shown in Figure 10. The z-score is generalised to asymmetrical form: \( \Delta c_j \neq \Delta c_j^+ \) before applying shape factors. For the linear parts, i.e. \( \alpha_j^- = 1 \) or \( \alpha_j^+ = 1 \), scaling is same as normalisation: \( X_j = p_j \). The scaling function is \( X_j = \text{sgn}(p_j) \sqrt{|p_j|} \) for the upper limit 3. For the lower limit \( \frac{1}{3} \), \( X_j = \text{sgn}(p_j)(2 - \sqrt{4 - 3|p_j|}) \) with a limitation that \( |p_j| \leq \frac{4}{7} \).

The parameters can be analysed from data by generalised norms, and the approach allows the important value ranges to be taken into account. Mean and median have been used in most of the applications discussed in Section 7.2. Improved sensitivity around normal operation is achieved with the method which uses skewness, and high sensitivity for the low and high values corresponds to large negative and positive orders, respectively.

Nonlinear monotonously increasing activation functions are widely used in neural networks: the log-sigmoid function and tan-sigmoid function scales to \([0, 1]\) and \([-1, 1]\), respectively. In the log-sigmoid and tan-sigmoid functions, the centre point can be moved by the bias term. To compare to the membership definitions, the sigmoid function is represented by using the z-score \( p_j \) and scaled to the range \([-2, 2]\). For \( p_j = 1 \) the function value is required to be 1, then the scaled values are obtained by

\[
X_j = \frac{e+1}{e-1} \tanh\left( \frac{p_j}{2} \right) = \frac{e+1}{e-1} \frac{e^{p_j} - 1}{e^{p_j} + 1}.
\]

(103)

The same scaling function is obtained from both the hyperbolic tangent and the log-sigmoid function. Delinguistification is done by

\[
p_j = \ln(X_j + \frac{e+1}{e-1}) - \ln\left( \frac{e+1}{e-1} - X_j \right).
\]

(104)

The functions (103) and (104) are close to the quadratic scaling functions, whose \( \alpha^- = \alpha^+ = \ln\left( \frac{2e-1}{e} \right) - 1 \approx 2.2346 \), which is in the range \([\frac{1}{3}, 3]\). The maximum difference between this function and (106) is 0.125 and 0.316 in the ranges \([-1, 1]\) and \([-2, 2]\), respectively. The double sigmoid function extends the log-sigmoid function to the case \( \Delta c_j^- \neq \Delta c_j^+ \).

The scaling with the hyperbolic sine function results

\[
X_j = \frac{2e}{e^2 - 1} \sinh(p_j) = \frac{e}{e^2 - 1} (e^{p_j} - e^{-p_j}).
\]

(105)

when the function is required to be 1 for \( p_j = 1 \). Delinguistification is done by

\[
p_j = \frac{1}{1 + \sqrt{2}} \text{arsinh}(X_j) = \frac{1}{1 + \sqrt{2}} \ln\left( \frac{2e}{e^2 - 1} X_j + \sqrt{X_j^2 + 1} \right).
\]

(106)
The functions (105) and (106) are close to the quadratic scaling functions, whose $\alpha^- = \alpha^+ \approx 0.66$, which is in the range $[\frac{1}{3}, 3]$. The maximum difference between this function and (106) is neglicable, 0.023, in the range $[-2, 2]$.

Table 15. Sigmoid functions.

<table>
<thead>
<tr>
<th>$X_j$</th>
<th>$p_j$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_j = f^{-1}(p_j)$</td>
<td>$p_j = f(X_j)$</td>
<td>0.2649</td>
</tr>
<tr>
<td>$X_j = \frac{p_j}{\sqrt{2}-p_j^2}$, $\forall</td>
<td>p_j</td>
<td>\leq 2\sqrt{5}$</td>
</tr>
<tr>
<td>$X_j = \frac{p_j}{2-p_j}$, $\forall</td>
<td>p_j</td>
<td>\leq \frac{4}{5}$</td>
</tr>
</tbody>
</table>

The linguistic interpretation requires that the core corresponds to the range $[-1, 1]$, which also confines the sigmoid functions shown in Table 15 close to the limiting quadratic membership definition $\alpha = \frac{1}{3}$. Some limitations are needed in the linguistification stage. The scaling functions introduced in (Snelick et al. 2005) provide more special cases defined by the min-max normalisation to $[0, 1]$ and two nonlinear functions. The centre $c_j$, scaled to the min-max range defines the inflection point of the functions. The norm (7) in the min-max range can be used instead of (8) for variables, which have both negative and positive values. Logarithmic functions are used as a first step if very steep scaling is needed. For example, pH is used in the skewness based approach instead of molar concentrations.

The scaling functions can be combined with additional nonlinear transformations to include process insight. All the functions must be monotonous increasing. For example, logarithmic scaling suits for the number of cycles in the fatigue model (91). Nonlinear scaling improves the feasibility of analytical models, e.g. cylindrical and spherical symmetry can be taken into account with the mixed approach.

Outliers and suspicious value ranges are obtained from the scaling functions (Fig. 14). The parameters can be recursively updated since the norms can be calculated by using equal sized sub-blocks, see (10). Type-2 fuzzy numbers provide information about the uncertainty of the functions. In addition, the fuzziness of the centre point $c_j$ has an important effect on the shape of the scaling functions. Suspicious values and high uncertainty of the parameters trigger adaptation procedure for the orders of the norms. Outliers are removed, but the imputation is used with care.

The scaling functions are useful in quality control systems:

- The quality control becomes more effective and closer to real time;
– The suspicious area is important in identifying calibration, measurement and communication errors;
– The automatic quality control algorithms are based on scaled values;
– The flagging systems can be based on the scaled quality levels;
– The scaled values make it easier for data users to identify suspicious and erroneous data, and to highlight corrected values.

Outliers and suspicious values may also mean that the operating conditions are changed, see Figure 28.

The scaled variables are intelligent indices, which can be used, even without modelling, as intelligent analysers in control and diagnostics. Good results have been obtained with stress and condition indices in condition monitoring.

Nonlinear scaling is integrated with various data analysis methods. It can be preceded by different signal processing and feature extraction methods, i.e. signals are divided or combined. The analysis of special features is not limited to the time domain. Frequency domain analysis expands the analysis, e.g. problem specific frequency ranges are important in vibration analysis. Also the parameters of wavelets can be handled with the membership definitions. The variable and feature selection is usually based on linear methods, which is wellsuited for nonlinear scaling, see Section 4.1.5. Generalised norms and nonlinear scaling functions also generalise SPC algorithms.

The main benefit of using the membership definitions is that the interpretation of the scaled values is straightforward and the same for all the variables. The domain expertise or on various kinds of data are used in producing scaled values to be used in linear models, which can be understood and explained in natural language. The generalised norms have improved the feasibility of the recursive data-based analysis.

### 8.1.2 Modelling and simulation

A steady-state LE model is an essential part of the system generated for diagnosis, soft sensors and trend analysis. They are also used in control (Tab. 8): inverted LE models are suitable for feedforward control, and working point models are used directly in the adaptation of the feedback control. Cascade structures make complex models more understandable. Dynamic modelling is needed for the prediction of trends and detecting of fluctuations (Fig. 52) and for controller tuning. In case-based models, interaction coefficients are specific to operating areas.
Steady-state models

A LE model is linear only if all membership definitions are linear. A part of the quadratic models presented by RSM models can be constructed as special cases of LE models. Interaction terms are not used in LE models, but quadratic terms are embedded if the membership definitions of the output variable have second order terms. For this special case, the number of parameters is \( n_i + 4 \). Also interaction terms are embedded in the LE models, since a sum of two scaled variables represents a product of two original variables. Correspondingly, a ratio is represented by subtracting the scaled variables. For a high number of variables, LE models need much less parameters than RMS models in the general case shown in Figure 4.

LE models were originally developed for representing fuzzy models. The model \( (35) \) corresponds to a one neuron model with an activation function represented by the membership definition of the output variable, if the weight factor \( (36) \) and all the elements \( p_j \) are obtained by linear membership definitions.

The number of parameters in LE models is comparable to that in quadratic, full RSM and simple neural models (Tab. 16). Normalisation of all variables is included in all the models. The LE model with linear scaling functions is equal to a linear model. The number of the model parameters depends on the number of variables, since the additional complexity is handled by asymmetry and nonlinearity, which require additional parameters. For the cascade and interactive models, the number of model parameters increases, but the total number of parameters increases only slightly since the number of the scaling parameters is the same. In addition, the parameters of LE models have constraints as discussed in Section 3.1.2, and the whole LE system can be assessed with expert knowledge. The RSM and MLP models, where complexity is handled with model parameters, have a high risk of overfitting since the number of parameters becomes very high, especially in the cascade and interactive models. In LE models, the overfitting risk needs to be taken into account if all the parameters are tuned simultaneously, e.g. with neural networks or genetic algorithms.

Steady-state LE models have been developed for different purpose. The first application, the electric furnace, was a representation of an analytic model for process design. Performances of several modelling methodologies have been compared in the continuous cooking case, where the model was used for predicting a quality variable. The lime kiln model \((90)\) was developed for FF controller. The working point model of the solar power plant is used for the adaptation of the FB control. For fatigue detection,
<table>
<thead>
<tr>
<th>Model</th>
<th>Single model</th>
<th>Cascade model</th>
<th>Single model</th>
<th>Cascade model</th>
<th>Single model</th>
<th>Cascade model</th>
</tr>
</thead>
<tbody>
<tr>
<td>LE 1</td>
<td>14 + 7 + 14 + 7 = 42</td>
<td>14 + 7 + 14 + 11 = 46</td>
<td>12 + 6 + 12 + 6 = 36</td>
<td>12 + 6 + 12 + 11 = 41</td>
<td>8 + 4 + 8 + 4 = 24</td>
<td>8 + 4 + 8 + 12 = 32</td>
</tr>
<tr>
<td>LE 2</td>
<td>14 + 6 + 12 + 7 = 39</td>
<td>14 + 6 + 12 + 11 = 43</td>
<td>12 + 5 + 10 + 6 = 33</td>
<td>12 + 5 + 10 + 11 = 40</td>
<td>8 + 3 + 6 + 4 = 21</td>
<td>8 + 3 + 6 + 12 = 29</td>
</tr>
<tr>
<td>LE 3</td>
<td>14 + 7 + 2 + 7 = 30</td>
<td>14 + 7 + 2 + 11 = 34</td>
<td>12 + 6 + 2 + 6 = 26</td>
<td>12 + 6 + 2 + 11 = 31</td>
<td>8 + 4 + 2 + 4 = 18</td>
<td>8 + 4 + 2 + 12 = 26</td>
</tr>
<tr>
<td>LE 4</td>
<td>14 + 1 + 2 + 7 = 24</td>
<td>14 + 1 + 2 + 11 = 28</td>
<td>12 + 1 + 2 + 6 = 21</td>
<td>12 + 1 + 2 + 11 = 26</td>
<td>8 + 0 + 0 + 11 = 19</td>
<td>8 + 0 + 0 + 33 = 41</td>
</tr>
<tr>
<td>MLP 3</td>
<td>14 + 0 + 0 + 25 = 39</td>
<td>14 + 0 + 0 + 45 = 59</td>
<td>12 + 0 + 0 + 22 = 34</td>
<td>12 + 0 + 0 + 45 = 57</td>
<td>8 + 0 + 0 + 16 = 24</td>
<td>8 + 0 + 0 + 48 = 56</td>
</tr>
</tbody>
</table>

The bias term is tuned to zero. MLP 2 has two neurons and MLP 3 three neurons in the hidden layer.
the LE model (91) represents a Wöhler curve by using at least five points from the original curve.

**Dynamic models**

Dynamic LE models only require very compact linear structures (Fig. 18(a)). In all applications, the model is based on an autoregressive with exogeneous inputs (ARX) model. Comparisons of (38) with different parametric models, e.g. a more complex ARX, an autoregressive moving average with exogeneous inputs (ARMAX), Box-Jenkins and Output-Error (OE), have shown that the performance improvement with additional values is negligible. Actually, numerical integration methods divide the time step into several parts when variable step integration and higher order solvers are used. In all these models, the new output is calculated first and the difference from the previous value is used together with the step size to obtain the derivative. Alternatively, the derivative can be calculated directly as it is done in the batch cooking models (Tab. 9). Transfer functions and linear state space models can be used in introducing analytic modelling to LE models.

The effective time delay (Fig. 18(b)) is important in the solar power plant models, since the operating conditions are highly dependent on the oil flow. The time delay is taken into account in the gas furnace case as a constant value.

Dynamic simulation is understood as a technique, which summarises the effects of the time steps, e.g. in the fatigue detection where the load contributions $N_C(k)$ of each sample time $\tau$ obtained from (91).

**Multimodel simulators**

Cascade models, e.g. alternatives shown in Figure 5, can be built with the blocks shown in Figure 17: intermediate variables can be handled as linguistic values, i.e. linguistification is needed only for the original inputs and delinguistification only for the final outputs. The number of parameters increases as the number of equations increases (Tab. 16). The cascade models I and II constructed with quadratic RSM models have less parameters than the corresponding LE 1 models. The interactive model (Fig. 5(c)) contains two equations more than the single LE model, which requires eight additional parameters (Tab. 16). A corresponding quadratic model requires six parameters more, and ANN models require even more parameters. The number of parameters needs to be
reduced in both the quadratic and neural models, see Section 2.1.4. This benefit is even more important with cascade and interactive models. However, the cascade structures can also be combined into a single set of equations.

Decomposition and clustering are used for finding the operating areas. Parameters \(\Delta c_j, \alpha_j, \Delta c_j^+\) and \(\alpha_j^+\) of different operating areas can be used in developing higher level models to be used in adapting the basic level models and the LE controllers (Section 5.2.1). Decomposition and multimodel simulators (Fig. 25) are needed to come up with different operating conditions. Similar models with case specific parameter values are used for the lime kiln and the solar power plant. The overall model moves smoothly between the submodels: the fuzzy decisions are based on the working point model and the fuzzy rules for the solar plant and the lime kiln, respectively.

Interactions of the dynamic models are realised in a modular solution based on process insight: the temperature, humidity and granular size in the fluidised bed granulator case (Fig. 5(a)), and alkali, lignin and dissolved solids in the batch cooking case, which also includes feedbacks (Fig. 5(c)). In the fed-batch fermentation case, the multimodel approach (Fig. 25) is used within interactive models (Fig. 5(b)). In the fluidised bed granulator case, the interactive models are embedded in the different phase models of the batch process.

The LE approach is closely linked to the fuzzy set systems (Fig. 24). The methodologies are combined in multimodel systems (Fig. 25) and in the fuzzy calculus (Tab. 7). In the solar application, an additional combination is used: special situations modelled with a fuzzy set system are combined with the dynamic LE model (Fig. 26). Complex structures can also be taken from neural networks, e.g. (89).

The multimodel approach is needed to realise simulators for nonlinear multivariable systems in wide operating areas. The number of parameters increases with an increasing number of variables (Fig. 4). Therefore, the number of variables should exceed five, but even these systems have a huge number of alternative variable groups (Fig. 19). Performance was compared with other methodologies in continuous cooking and fed-batch fermentation cases.

The models are based on normal process operation in most of the applications, including the lime kiln, the batch cooking and the fed-batch fermentation. Experimental design was used for the fluidised bed granulation and partly in the water treatment, where the unavoidable changes of the operating conditions broke the design matrix. In the solar application, the controller testing and model development depended on the
weather conditions, which meant that the experimental plans always had to be modified and the individual tests cannot be redone.

Applications are in forecasting and control design (Fig. 3). In on-line forecasting the simulator, which is started on chosen time intervals, takes into account the history preceding the starting point. As prediction also needs assumptions on the control actions, this is very close to MPC. The same structures and scaling methods as the LE models are used in the LE controllers with some additional types of membership definitions, see Figure 17.

The main benefit of LE models is the compact structure, which can be extended to highly complex systems by using reliable linear modelling methodologies. The process insight is retained by using the scaled values, which can be understood in linguistic terms. The compact structure of the LE models provides a basis for intelligent distributed parameter models.

8.1.3 Monitoring and control

Nonlinear multivariable LE control is based on parameterised feedback and feedforward controllers: almost all the modules presented in Figure 7 have been used real-time. The combination of the intelligent analysers and control shown in Figure 53 is the main approach. Nonlinear LE controllers have a wide operating area, which is further extended with adaptation. Although the LE approach allows on-line identification and classical adaptation mechanisms, the applications are based on predefined adaptation. On-line adaptation has too many risks in these processes characterised by strong and fast disturbances. In addition to the feedforward control, models are used in predefined adaptation and cascade control.

Feedback control is based on a PI-type LE controller (Tab. 3), which is designed for normal operation. Multiple control strategies with several manipulating and controlled variables can be used. A SISO controller is used for the solar thermal power plant and water treatment. A MISO controller is realised as a weighted sum, e.g. seven controllers defined by (93) were combined with (49) in lime kiln control. In MIMO controllers, each manipulating variable has its own weighted sum.

Intelligent analysers are developed to provide more informative indirect measurements, which can directly be used in adaptive, model-based and high level control. The working point is important in all three applications, the predictive braking is used for handling large errors in solar and lime kiln applications, asymmetrical control surfaces
react to changing working points in solar application. The working point is defined by application specific variables: the irradiation $I_{eff}$ and the difference $T_{diff} = T_{out} - T_{in}$ in the solar plant, the production rate and the draught fan speed in the lime kiln, and the water quality and the setpoint of the turbidity in the water treatment. Several analysers are developed for detecting changes in the inputs, including water quality in the water treatment and inlet temperature in the solar collectors, in the process environment, including irradiation in the solar collectors, in control power, including fuel quality in the lime kiln, and in the cumulative changes of control in the lime kiln. State indicators based on the fluctuations of the irradiation, temperature difference and oil flow react to cloudiness and oscillations in the solar application. The intelligent trend analysis would be useful in asymmetrical action. The intelligent indicators of the fast temperature changes have a link to the intelligent trend analysis. This monitoring is done for the inlet and outlet temperatures and their difference in the solar application.

*Adaptation* is based on the scaling of the change of control by using the working point $wp_i(k)$, the control power $cp_j(k)$ and the cumulative change of control $cr_j(k)$. Intelligent analysers are the main indicators for the need of adaptation. The working point (51) is essential in adaptation: the control power (52) and the cumulative sum of the control actions (53) can be understood as additional parts of the working point model. The predictive braking action for large changes was introduced for the solar plant and later extended to the lime kiln. The solutions are different: one for fast reactions and the
other for smooth corrections. The asymmetrical action is used for removing the offset in the solar application. Adaptive scaling changes the control surface of the PI-type LE controller in a way which is related to the gain scheduling approaches: the controller could be called a LE-based gain scheduling (LEGS) controller.

*Model-based* control is used either by limiting the allowed ranges of the setpoints, or by using feedforward control based on inverted models. The set points are limited in the solar application during the start-up and other drastic changes to avoid oscillations and temperatures that are too high. Since fast reactions are needed in the solar application, the online fluctuation indices are used. The lime kiln should be controlled smoothly, due to the long time constants in the measurements and the process operation. Therefore, feedforward control is important in the supervisory control of the lime kiln process: the operating conditions are changed by controlling the rotational speed and the draft fan speed. The feedforward control of the fuels moves the fuel feeds on an appropriate level and defines an allowed tube for the feedback control outputs. In the water treatment application, feedforward control is used for the slowly affecting chemical, and in addition the dynamic turbidity model is a part of the water quality indicator.

*High level control* focuses on switching, activating, weighting and closing controllers and their special features, e.g. adjusting the strengths of PBA and ASA. The fine tuning of the adaptation can be done by changing the weights for the working point, control power and cumulative rate. The manual actions can be automated stage by stage, e.g. the limit $w_{p_{\text{min}}}$, which was earlier defined manually, can now be set in the model-based control. However, a possibility of *manual control actions* is important industrial use, since unknown situations may occur. In the lime kiln control, the bias correction $u_{BC}^{i}(k)$ has reduced the need to interrupt the automatic control.

*Event based control* is a part of model-based and high level control. The events are detected with intelligent analysers and they can introduce additional automatic control actions, for example in the solar power plant:

- Large initial error activates the braking action.
- Cloudy conditions detected as strong fluctuations of the irradiation introduce changes to the working point.
- Oscillations detected as strong fluctuations of the outlet temperature and oil flow introduce changes to the working point.
– Additional control actions are introduced if the special features detect fast changes or overshoot for temperatures (inlet, outlet, difference) in the solar plant and for temperatures and oxygen in the lime kiln.
– Steady operating conditions and a small error activate the asymmetrical actions.

The strength of the braking and asymmetrical actions, the weigh factors of fast change indicators, the working point limit and the setpoint are defined in the high-level control. A high setpoint means that the control is based on a chosen working point level and fluctuation corrections. A fixed setpoint can be taken into use on any time if it does not exceed the chosen limits. The controller can operate unattended, but the operator can take full command at any time. The scaling functions are process specific.

The LE controllers can be combined with other controllers. Fuzzy controllers, which were earlier developed and tuned with the LE approach, are now in the high-level control for the weighting of different control strategies. Neural LE controllers can be introduced by replacing normalisation with the nonlinear scaling. Advanced model-based controllers like MPC and IMC have not been used in the real-time LE control. In solar thermal power plants, the controller should react quickly in changing operating conditions. In principle, MPC could handle the working point adaptation since it can combine the interactions of several variables. However, in this case, the irradiation, which is import in MPC, is in this case measured far from the collector field. On sunny days MPC could be an option, but the operating condition change very quickly in varying cloudy conditions. The IMC can cope with varying time delay, but the quick changes in cloudy conditions are challenging. For these reasons, the MPC and IMC approaches were restricted to controller tuning: MPC can be used in the tuning of the braking action to get good trajectories for selected simulation cases (Juuso 2006). However, in practice the PBA is activated for very short time periods. The feedforward control was also left out, since the changes of the operating conditions can be handled well with the LE controller, which combines intelligent analysers and adaptation mechanisms discussed above. The operation is further improved by adjusting the setpoint with the working point and the working point corrections. The very fast predefined actions remove the need for on-line identification, or for classical adaptation mechanisms based on performance analysis.

The main benefit of the LE controllers is the flexible modular combination of several control strategies: the same controller can operate in the whole working area. The efficient nonlinear scaling is the key in extending the well-known linear control solutions.
to nonlinear multivariable applications. All these properties are implemented into a very compact control program: the solar plant controller as Matlab and C-code, the lime kiln controller in the G2 software (Järvensivu et al. 2001) and the water treatment controller in a programmable logic controller (PLC) (Piironen et al. 2008).

8.1.4 Detection of operating conditions

Intelligent stress and condition indices obtained by using the nonlinear scaling and the LE models provide good tools for detecting operating conditions and faults (Fig. 52). The intelligent trend analysis extends these solutions to temporal reasoning and prognostics. Steady-state and dynamic models are embedded in multimodel simulators, where specific submodels are developed for different situations, or cases. In large scale systems, the LE models are combined with CBR, see Figure 48. Data quality control and anomaly detection are important parts of the data selection. Compact intelligent analysers can be used in the intelligent control, risk analysis and detection of sensor failures (Fig. 53). The scaling functions are also useful when cavitation is taken into account in power control. As the corner points shown in Figure 42 can be used as warning and alarm limits.

The LE-based condition indices are consistent with the vibration severity criteria, which originate from VDI 2056 (VDI 1964, Collacott 1977). Several severity criteria are compared in Table 17. Rathbone (1939) introduced the first widely known recommendations for the acceptable vibration levels as a function of frequency. Rathbone presented these classes (Table 17) for displacement. Blake (1964) connected the severity levels to the rotation frequencies. Boyce (1978) and Goldman (1984) presented the severity levels for displacement, velocity and acceleration. Similar classification principles hold for the higher derivatives as well. The severity criteria based on VDI 2056 are used in Figure 41.

In condition monitoring, the faults are also identified, if it is possible to get suitable features for that, e.g. by performing tests for different faults with the test rig. In paper machines, the detected information of the identified cases is used for predicting the break sensitivity of the paper web from process data. As the analysis is based the same methodology in all these applications, the monitoring of the machines can be combined with the process data. The smooth operation and high quality of products is the main goal of all these applications, and this can be achieved by combining these indicators
Table 17. Scaled features and vibration severity criteria.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1.5, 2]</td>
<td>Very rough</td>
<td>Not</td>
<td>Dangerous</td>
<td>Extremely rough</td>
</tr>
<tr>
<td>[1, 1.5)</td>
<td>Rough</td>
<td>Still acceptable</td>
<td>Failure is near</td>
<td>Very rough</td>
</tr>
<tr>
<td>[0.5, 1)</td>
<td>Rough</td>
<td>Faulty</td>
<td>Faulty</td>
<td>Rough</td>
</tr>
<tr>
<td>[0, 0.5)</td>
<td>Slightly rough</td>
<td>Usable</td>
<td>Minor faults</td>
<td>Fair</td>
</tr>
<tr>
<td>[-0.5, 0)</td>
<td>Fair</td>
<td>Usable</td>
<td>Minor faults</td>
<td>Fair</td>
</tr>
<tr>
<td>[-1, -0.5)</td>
<td>Good</td>
<td>Good</td>
<td>No faults</td>
<td>Good</td>
</tr>
<tr>
<td>[-1.5, -1)</td>
<td>Good</td>
<td>Good</td>
<td>No faults</td>
<td>Good</td>
</tr>
<tr>
<td>[-2, -1.5)</td>
<td>Very smooth</td>
<td>Good</td>
<td>Good</td>
<td>Smooth</td>
</tr>
</tbody>
</table>

with process control in the same way as it has been done for smaller indicators used in the lime kiln control and in water treatment.

Case-based models are generated for large-scale systems, and usually much less than one percent of the original alternatives are needed in the case models. The model selection is important since the models should be sensitive to the cases, i.e. models which are good in many highly different cases are not useful for detecting these cases. Considerable overlap is acceptable for nearby cases. The case sensitive selection requires a useful feature for making the decision on the sequence of cases. The final set of selected groups can be built in a modular way: some groups may include also four or five variables, and different subgroups can be developed for different subsets of variables. Redundant measurements are included in the same groups only for the fault diagnosis of measurement devices. All these tasks can be performed automatically. However, it is more important that all intermediate results can be modified on the basis of expertise.

The main benefits of the LE diagnostics are the intelligent stress and condition indices, which can be used in the same way as the normal process measurements.

8.1.5 Scheduling and decision making

Intelligent analysers clarify decision making and even make it automatic in the high level control, where the trade-offs between control strategies can be done with fuzzy logic. Gradually refining levels based on the linguistic meanings provide an efficient optimisation approach for hierarchical decision making. Short term scheduling adapts to changing operating conditions if the rules are represented by preferences and priorities,
which can be understood as soft sensors. The stress index can be used in the power allocation of water turbines to minimise cavitation levels.

Decision support can use the same principles as the control when the information is represented as quantitative indicators. The scale \([-2, 2]\) can also be used to interpret information presented in natural language. The OEE-value, which is a good tool for monitoring the performance improvement, is explained in natural language in Table 14. Harmonised indicators provide valuable information about the performance: improvements towards more cost effective maintenance with less unscheduled actions can clearly be seen. SMART criteria, defined by specific, measurable, attainable, realistic and timely, bring key performance indicators (KPIs) to a closer connection to the process. The criticality analysis is provided by the RCM and RAMS (Reliability, Availability, Maintainability and Safety) analysis. All these can be considered as feedback information for management.

More process related information is needed to find suitable actions to improve safety, reliability and performance. Condition, stress and quality indicators are useful even in real time risk-based maintenance (RBM). Intelligent trend analysis brings these ideas to detecting slow changes. Statistical process control (SPC), which is used for monitoring variability, can also be extended to monitor harmonised indicators, KPIs, RAMS data and the OEE. In addition, SPC can be used to compare the results of improvements when developing best practices. Also the trend indices are suitable for SPC.

The linguistic meanings are emphasised in decision support and advisory systems, since the humans in the loop are important in new situations. Expertise is needed for finding the appropriate subsystems for development and tuning. Interpretation into natural language is beneficial in these applications.

8.2 Linguistic equations in hybrid systems

Linguistic equations provide techniques for combining expertise and data to make the overall system development and tuning easier. The adaptive nonlinear scaling expands the idea of membership functions. The LE systems extend the possibilities of fuzzy set systems: they can be extracted from fuzzy set systems and tuned with neural networks and genetic algorithms to adjust the membership definitions, model coefficients and variable time delays. The LE approach provides a channel for transforming existing fuzzy systems into different forms of fuzzy systems. The linguistic neural networks, which combine neural computing and nonlinear scaling, extend the application areas.
of the old neural networks, e.g. linguistic perceptron networks and linear networks operate also in nonlinear applications. Methodologies used in the development of expert systems are useful in extracting knowledge from domain experts. The case-based LE modelling also produces the feasible ranges for the submodels. The feasible ranges and the corresponding weights are essential in hybrid systems. Rules are generated only if they are required by the programming system used in final implementation.

The LE approach combines various intelligent modelling techniques in a unified framework: a close connection to fuzzy set systems was already important in the early applications; data-driven modelling properties have brought the approach close to ANN techniques. Various methodologies of computational intelligence have their strong areas but also weaknesses (Juuso 1996). Expert systems and fuzzy set systems are knowledge-oriented, and neural networks data-oriented. Small fuzzy systems are easy to build but tuning them is difficult in complicated systems. Fuzzy clustering and rule generation methods extend fuzzy methods towards data-based techniques. Domain expertise is used in selecting the structures for neural networks. Neuro-fuzzy methods are used in developing fuzzy systems with neural networks. Genetic algorithms are suitable for the optimisation of heterogeneous systems, including also fuzzy set systems and neural networks.

Data-driven fuzzy modelling can be based on various methodologies, e.g. fuzzy clustering, self-organizing maps, neurofuzzy methods and linguistic equations. Clusters can be represented by the feasible ranges (Fig. 9). Different approaches are combined in the tuning phase. Connections to fuzzy set systems are used in the knowledge-based part, and the data-driven parts use neural computation for comparing the applicability of the linear approach. Fuzzy set systems introduce uncertainty handling, join together local LE models, and provide tools for the modelling of special cases. The membership definitions have the same number of parameters as the set of membership functions shown in Figure 9. Fuzzification and defuzzification are needed if the system includes fuzzy rules (Fig. 54).

Several sets of linguistic equations can be combined and shown as a matrix presentation (31), where each row is generated separately. Some of the equations may be generated automatically from data, some from expertise presented by fuzzy set systems, and some are defined directly from prior knowledge. Large rulebases can be packed into very compact matrix equations. Each linguistic equation represents a multivariable interaction, and only the variables with nonzero coefficients belong to the interaction. This set of equations is the main part of the decision making logic shown in Figure 54.
Fuzziness and probabilities are handled with higher levels of the system. The models can be built in a flexible way for different levels in multilevel systems (Fig. 25) and specialised models (Fig. 26).

Detection of operating conditions is based on the facts produced by the case-based LE modelling (Section 4.1.4). Reasoning on these facts is based on fuzzy logic which brings additional parameters to the tuning of the hybrid LE systems. The fuzziness (40) of the facts provides a good basis for the uncertainty processing. For decision making, linguistic equations provide new aggregated facts to be used in approximate reasoning on the basis of the degree of memberships. This reasoning could contain both feedforward and feedback chaining, but only data-driven forward chaining is needed since all the variables in the set of linguistic equations can be handled in the same way.

The core of the decision making logic in Figure 54 is a set of linguistic equations: linguistic input values come from other systems or as an input from experts; linguistification and delinguistification blocks (Fig. 17) are needed for other types of input. The fuzzy calculus (Tab. 7) is embedded in the decision making logic. Deterministic inputs are handled either with membership definitions in the normal way or with fuzzification and linguistification if the system contains fuzzy rules. Probabilistic inputs are either used random numbers or converted into fuzzy numbers. The discretised fuzzy numbers shown in Figure 13 are handled as deterministic values, which have a degree of membership. The system shown in Figure 54 also handles the type-2 fuzzy numbers and probabilistic inputs. The output is kept in fuzzy form when the type-2 fuzzy sets are not reduced to type-1 sets.
The new parameterised scaling functions with clearly defined search areas provide very efficient coding for genetic tuning. Penalty functions used in the previous approach are not needed. The new skewness based method for defining the working point and the feasible area reduces the need for tuning in many cases. The relations are tuned by adjusting the meanings of the linguistic values on the basis of the working point conditions.

In intelligent control design, hybrid techniques combining different modelling methods in a smooth and consistent way are essential for the successful comparison of alternative control methods. Switching between different submodels in multiple model approaches should be as smooth as possible. For slow processes, the predictive model–based techniques are necessary at least on the tuning phase. The adaptation to various nonlinear multivariable phenomena requires a highly robust technique for the modelling and simulation. Dynamic LE simulators are continuously used in the development of multilayer LE controllers.

8.3 Smart adaptive systems

The smart adaptive systems in this research are based on the smart use of intelligent subsystems: the smartness of the overall systems depends on the integration of these subsystems. The levels of adaptation are in (Anguita 2001) analysed as data mining with fuzzy systems, artificial neural networks, machine learning, evolutionary computation and case-based reasoning. In the LE approach, the importance of the complex structures is reduced by the scaling functions, which also provide limits for the core, support, suspicious and outlier areas (Fig. 14). The different types of membership definitions (Fig. 17) are informative for combining data and expertise since the shape factors $\alpha_j^-$ and $\alpha_j^+$ cover a wide range of data distributions. The same scaling approach is also used in modelling, control and diagnostics. The focus is on combining data and expertise in the development of applications, where adaptive and intelligent methods are important.

Smart adaptive applications are built as integrated sets of subsystems (Fig. 55) which combine expertise and data with several methodologies (Fig. 8), especially fuzzy logic and linguistic equations. The interpretation of the scaled values is important when there are the humans in the loop. The research and development focuses on adaptive decision making, performance analysis and learning by integrating different methodologies (Fig. 8).
Adaptation levels

The adaptation to a changing environment is handled in two parts: (1) recursive tuning of the parametric scaling functions, and (2) on-line identification of the model interaction parameters. The combined cascade and interactive structures are used in the multimodel approach, where the tuning and identification of the parameters are specific to the operating conditions and process phases. Novelty and anomaly detection is aimed at finding situations where the adaptation of the parameters is needed. The feasible ranges represented as type-2 fuzzy numbers provide useful information about the fuzziness. Values, which are within the feasible range, are accepted in recursive adaptation. Suspicious values are included first in the recent data set and special acceptance is needed before expanding the definitions. The values, which were first notified as outliers, are used if they are rendered acceptable by the new definitions.

The parameters of the scaling functions may change drastically or vary strongly when the operating conditions change, e.g. a cloudy period followed by a lower irradiation conditions can be seen in Figure 28(a): the membership definitions were updated with the data of the lower irradiation conditions, but the cloudy conditions are used to adapt the working point for the model-based control (Fig. 31). The same definitions are used in the trend analysis to detect changes in operating conditions (Figs. 49 and 50). The recursive modelling is essential in the prognostics. In the applications discussed above, the model structures are fixed, i.e. only the parameters are updated, when the operating conditions change. The changes of the parameters are managed with LE-based working point models or fuzzy set systems.

The adaptation to a similar setting can use the same procedure of extracting the membership definitions and the interaction parameters. New operating conditions are analysed on the basis of similar settings, where the LE model structure should not change. The number of model parameters is fairly small (Tab. 16). Even the default parameters operate well in many applications, e.g. in the solar application the coefficients of the working point model are one, i.e. the adaptation is completely based on the membership definitions. Complex model structures can be realised and used in various specific control actions (Tab. 11): most of the parameters are related to understanding of the meaning of the variable values. The same membership definitions are used in trend analysis to detect changes in the operating conditions. All the parameters can be tuned for example with genetic algorithms in a well-defined way without using additional constrained methodologies.
The adaptation to a new or unknown application is a challenging task in complex systems. The future potential is in the integration of functions and features in hybrid environment (Fig. 55), where the methods are selected from an expanding pool of methodologies. The connection alternatives are changing much faster than the functions and features. The same compact system structures have been used in many applications together with the procedure of extracting the membership definitions. Technically, the development could be done with CBR after selecting the data for different cases as it was done in the web break sensitivity analysis (Fig. 48). The recursive parameter estimation is the main extension to this approach.

Domain expertise and human contributions should not be overlooked: expertise is important in practical applications, especially in variable selection and grouping. Similar functions and features are useful in intelligent analysers, controllers and the detection of operating conditions in new or unknown applications. The difference between similar settings and new applications is gradual: a new similar application may require only the parameter estimation for a totally new application can be based on similar knowledge-based structures.
Applications

Integrated modelling approaches produce smart adaptive models for forecasting and control design: the smoothly operating LE models are combined with fuzzy set systems in many ways to enhance the properties of the multivariate models. Fuzzy set systems or working point models facilitate smooth transitions between LE-based submodels, which are developed for different process phases and operating conditions. LTS models are special cases, where the modelling areas and the local models use the same variables. The nonlinear local models reduce the hedges and valleys and the risks of chattering. The system may contain phenomenological models whose parameter handling is based on discrete membership functions (Fig. 13). Dynamic LE models have been combined with fuzzy set systems to handle special cases together with the smoothly operating LE model (Fig. 26). Uncertainty of inputs and models is handled with fuzzy calculus based on the extension principle and interval analysis (Tab. 7). Type-2 fuzzy numbers can be used for the centre point and the feasible area, i.e. the parameters are fuzzy. The combined models are located in the upper right corner of Figure 6. As the basic LE models are flexible in representing nonlinear behaviour, the linguistic neural networks (89) have not been used so far. The LSOM is used for generating linguistic interpretation for SOM neurons. The recursive parameter estimation is the key to the adaptation in these applications.

Intelligent analysers and controllers form a flexible modular combination of several control strategies: the same controller operates in the whole working area by activating special actions in a proper way. The decision making unit is the control block (Fig. 55), but putting all the used measurements in it would result in systems that are too complicated. A better alternative is to make a generic and configurable control block, whose inputs are (calculated) variables that relate to system properties which really should be controlled. This is done with software sensor type of intelligent analysers (Fig. 53), which are especially important in connection with continuous on-line analysers. Intelligent analysers are also needed for the detection of operating conditions, partly combined with trend analysers. Stress and condition indices are soft sensors, which are based on continuous condition monitoring, signal processing and feature extraction. The LE models are suitable for representing the operation limitations of actuators, e.g. in the solar plant the oil flow through the pump is limited by the viscosity in the start-up. Moving close to the process brings new challenges for the implementation of
the intelligent systems. The intelligent analyser and control system shown in Figure 53 has the first two feedback loops in the overall system (Fig. 56).

The periodic condition monitoring produces additional stress and condition indices to be used together with the continuous indices and other measurements in diagnostics and prognostics (Fig. 56), where a huge number of different approaches are used. Failure mode and effect analysis (FMEA) is aimed at identifying potential failure modes based on past experience. Causal directed graphs (CDG) represent physical cause-effect relations between variables. Computational intelligence, cluster analysis, modelling and CBR form the basis for the analysis. The time and frequency domains are important in signal analysis. Stress and condition indices, measurement and health indices ($I_M$ and $I_H$) are compared with severity criteria (Tab. 17). Cumulative times on different stress levels are useful in prognostics and risk analysis to estimate remaining useful life (RUL) or number of operating cycles. Diagnostics and prognostics should have effects on the adaptation of the control strategy and the optimisation and coordination of control.

Decision support systems (DSS) can use the same structure as the intelligent analysers and control (Fig. 53). The information is more uncertain in tactical pricing and demand forecasting. The short term scheduling is closer to the control: the importance and usefulness can be considered as intelligent analysers of the work load. Condition and stress indices produce useful information for the allocation of power to several turbines. Short-term scheduling adapts the control strategy (Fig. 56) in these cases. The linguistic interpretation of the OEE values shown in Table 14 is an example of using the scaling approach in the management applications. Other performance measures like harmonised indicators, KPIs and PCIs can be explained in a similar way. The scaling approach and the limits of the suspicious and outlier areas provide feasible extensions to the SPC, especially to the short-term SPC. Performance monitoring should be based more on the real-time information aggregated from the subprocesses and process units.

Condition-based maintenance (CBM) is cost effective way of improving the timing of the maintenance actions. The evidence of abnormal behaviour is obtained by diagnostics and prognostics based intelligent indices and models. This provides time to plan the required maintenance actions in connection with the control strategy and process requirements. The condition monitoring data should be used together with the process data to get reliable information about the operation. RCM and RBM are natural platforms for using the scaling approach in maintenance. Utilisation of the RAMS data and estimating the life cycle cost (LCC) are integrated into the overall system (Fig. 56).
Process and automation design is intensified by gradually refining simulators. For a new application or process, the first part of process development is a feasibility study with fairly light and flexible models. Automation design has traditionally done this with separate simplified simulators but the present software and hardware performance provide a good basis for the combined use of process and control models. Simulators and intelligent analysers store the knowledge accumulated during the design process. Training simulators, which support process operation, are useful in building experience on difficult operating conditions if they are really exceptional. Smart adaptive design starts with prototype simulators and forms links between the simulators as they provide feasible solutions for collecting the experience during the design process. Data-driven modelling techniques will update the models during the operation as well. Safety analysis and optimisation with simulators may also produce new ideas for process changes.

Efficient integration of subprocesses and application areas in a changing environment is facilitated by a unified methodology, in which individual solutions are adapted to the
amount and uncertainty of the data and expertise. Signal processing, interpolation and feature extraction are combined in an application specific way (Fig. 52). Since similar compact structures and scaling methodologies are used in different subsystems (Fig. 55), the LE-based smart adaptive system combines various aspects of the overall system (Fig. 56). The inputs and outputs of the systems can be deterministic, linguistic, fuzzy or probabilistic (Fig. 54). The compact LE solutions facilitate the use of methodologies shown in Figure 8 in the higher level systems, where the analysis becomes easier by representing all the variables, features and indices in the same scale $[-2, 2]$. 
9 Conclusions and future work

Different application areas have similar requirements: changes in operating condition need to be detected earlier, and decision making, including control, must adapt to changes faster. Data mining needs to be combined with domain expertise to develop better systems. Earlier research demonstrates that 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. This thesis concentrates on the new adaptive scaling approach, intelligent indices and parametrisation of the systems for advanced tuning.

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. Fluctuations and trends are analysed with the adaptive scaling functions, which also provide limits for suspicious and outlier values. The recursive parameter estimation extends the definitions or reveals changes in operating conditions. In principle this provides a basis for the development of smart adaptive systems in three levels: changing environment, similar settings and new or unknown applications. Different statistical and intelligent methods are used together with the LE approach. However, data mining is not sufficient since the domain expertise should not be overlooked in practical applications. Therefore, the research is focused on the smart adaptive applications, where different intelligent modules are used in a smart way.

The compact parametric approach for the steady-state, dynamic and case-based modelling from data and expertise is based on linear methodologies: piecewise affine models are extended to nonlinear applications by using nonlinear scaling to avoid the chattering behaviour and hedges and valleys in the model borders. The weighting of submodels is also based on the scaled values and fuzzy logic. The cascade and interactive model structures are used in building more complex large scale applications. The application specific performance analysis and comparison to other methodologies are excluded in this thesis. The model structures have already been successfully used in previous LE applications: the scaling methodologies were updated during the phase V.
The LE-based intelligent analysers are useful in the multilevel LE control and diagnostics: the feedforward and feedback LE control is enhanced with the intelligent analysers, adaptive and model-based modules and high level control. The operating area is extended with the predefined adaptation and specific events activate appropriate control actions, like braking and asymmetrical corrections. Overshoot and oscillations are minimised by adapting the setpoints to the chosen working point limits, which are modified with intelligent analysers. The multilevel controller can follow the adapted working point limit unattended, but at any time, different operating modes can be closed, started and weighted by the operator. Feedforward controllers are important in slow processes.

The condition, stress and trend indices introduced during the phases IV and V are used in the same way as normal process measurements. Combined indices also introduce facts with uncertainty for fault diagnosis and performance monitoring. The indices are consistent with the vibration severity criteria. Generalised norms and moments provide efficient tools for feature extraction and scaling: the same methodology is applicable to various processes, whose rotation speed ranges from slow to very fast. The number of sensors and features can be reduced drastically in multisensor applications. In more complex models, fuzzy logic and case-based solutions are moved to higher aggregation levels by using these informative features and indices.

Scheduling and managerial decision support can be represented by the same overall structure as the control and diagnostics. The linguistic representation becomes increasingly important in these applications where the human interaction is essential. Diagnostics, prognostics and performance monitoring are taken into account in condition-based maintenance. The unified scaling approach provides flexible extensions to SPC, TPM, TQM and JOT.

The efficient parametrisation of the LE systems is beneficial in tuning and adaptation, especially in large scale systems. To achieve insight and robustness the parameters are defined separately for the nonlinear scaling and the interactions. The performance measures of different variables are represented with the scaled values interpreted as fuzziness. The system integration leads to a hybrid system: fuzzy set systems move gradually to higher levels, neural networks and evolutionary computing are used for tuning, and the whole system is reinforced with advanced statistical analysis, signal processing, feature extraction, classification and mechanistic modelling and simulation.

The recursive data analysis and modelling approach presented above will be applied in many ways in the future development of smart adaptive applications: the mean and
median previously used in the simulation and modelling are replaced with the generalised moments and norms; the recursive modelling is essential in new prognostics applications. The development methodology also allows on-line modelling and adaptation, which can be compared with the current approaches in control application. Simplified LE models are tested in MPC and IMC environment for applications which have strong interactions between controlled variables. The focus on the smart adaptive applications is maintained in practical cases although the compact LE approach provides a feasible solution for the most challenging type of data-driven smart adaptive systems: the adaptation to a new application.
References


Järvensivu M (2003) Intelligent control of the lime kiln process with respect to environmental requirements. Helsinki University of Technology Laboratory of Process Control and Automation, no. 8.


248


249


VDI (1964) VDI 2056 Beurteilungstäbe für mechanische Schwingungen von Maschinen - VDI-Richtlinien.


460. Koelba, Adrian (2013) Theory of rational decision-making and its applications to adaptive transmission
461. Lauri, Janne (2013) Doppler optical coherence tomography in determination of suspension viscosity
462. Kukkola, Jarmo (2013) Gas sensors based on nanostructured tungsten oxides
463. Reiman, Arto (2013) Holistic work system design and management:—a participatory development approach to delivery truck drivers’ work outside the cab
464. Tammela, Simo (2013) Enhancing migration and reproduction of salmonid fishes: method development and research using physical and numerical modelling
465. Yadav, Animesh (2013) Space-time constellation and precoder design under channel estimation errors
468. Remes, Jukka (2013) Method evaluations in spatial exploratory analyses of resting-state functional magnetic resonance imaging data
469. Orvisjärvi, Kati (2013) Industry and traffic related particles and their role in human health
470. Czajkowski, Jakub (2013) Optical coherence tomography as a characterization method in printed electronics
471. Haapalainen, Mikko (2013) Dielectrophoretic mobility of a spherical particle in 2D hyperbolic quadrupole electrode geometry
Esko Juuso

INTEGRATION OF INTELLIGENT SYSTEMS IN DEVELOPMENT OF SMART ADAPTIVE SYSTEMS

LINGUISTIC EQUATION APPROACH