

Application of Elastic Intuitions to Process Engineering

Kalle Halmevaara*

*Helsinki University of Technology
Control Engineering laboratory
PO Box 5500, 02 015 TKK, Espoo, Finland
kalle.halmevaara@tkk.fi

Heikki Hyötyniemi*

*Helsinki University of Technology
Control Engineering laboratory
PO Box 5500, 02 015 TKK, Espoo, Finland
heikki.hyotyniemi@tkk.fi

Abstract

In this paper the intuitions of the elastic systems and neocybernetics are applied to large scale industrial systems. The process performance enhancement can be seen as technical evolution of the system. The objectives of technical evolution are somewhat different from those of natural neocybernetic systems. The performance of the system can be characterized by means of quality measures and the parameters of the system can be tuned along the axes of freedoms in the parameter space towards values that result in better process performance. Results from a simulated case study on a continuous pulp digester model are presented and discussed.

1 Introduction

Traditionally, managing complexity in large-scale industrial systems has been tackled with reductionist engineering approaches. One concentrates on the analysis of individual signal realizations – however, it is only different kinds of *quality measures* emerging from statistical considerations that are the most important from the point of view of mastering the process and production. The low-level structures are typically well-known – ideal mixers, etc. – but how are the properties of the high-level structures – networks of such mixers – related to them?

It turns out that thinking of the complex plants as *elastic systems* helps to find the conceptual and practical tools to reach the level of *emergent behaviors* of such systems (Hyötyniemi, 2006). One never knows all relationships among variables in a complex system – but in the elasticity framework this does not matter. It only has to be assumed that the underlying interactions and feedbacks manage to keep the system in balance in varying conditions – and this assumedly is a realistic assumption: Only realistic, successfully operating processes are studied.

Whereas the interactions and feedbacks determine the *constraints*, it is the remaining *freedoms* that are more interesting. The number of degrees of freedom is typically much lower than the total number of system parameters is, so that when concentrating on the emergent functionalities, or the manifestations of the freedoms, low-dimensional models can be found. When the system is “pressed”, the

system yields along the freedom axes – if the disturbances are small, one can assume that there is a linear relationship between the *tensions* and the corresponding *deformations*.

The industrial plant is very much like a biological cell (Haavisto, 2006): It also tries to maintain its *homeostasis*, and it tries to keep up its “metabolism”, or production, regardless of the external disturbances. There are differences, too: “Evolution” in such a system is implemented not by nature but by an “intelligent designer”, the human. In a natural system there is no predetermined direction of evolution – adaptation simply takes place in the direction of maximum cross-correlation among signals, in the sense of *principal components* (see Hyötyniemi, 2006); however, in the industrial case explicit goals exist. There is now more structure, as the signals have different roles: There are the input variables, the *qualifiers* that can be affected, and then there are separately the output variables, or *qualities*. The process of adaptation thus becomes more sophisticated: When the outputs are to be maximized, there exist plenty of alternative regression structures (like *canonical correlations*) that can be exploited to modify the directions of the freedom axes to make them better reflect the design goals.

After the degrees of freedom are found, the qualifiers (model or control parameters, set points, etc.) can be gradually modified to “push” the qualities in the desired directions. This way, a “holistic” way of iterative optimization of the complex system can be implemented.

In this paper, a holistic Iterative Regression Tuning (IRT) method applying neocybernetic ideas for

tuning multiparameter systems is presented. Originally the IRT method and results from earlier case studies were presented in (Halmevaara and Hyötyniemi, 2004; Halmevaara and Hyötyniemi, 2005). First, the IRT method and some practical instruments are introduced. After that, an application on continuous pulp digester control system is presented. Finally, the results and the potential of the IRT method in other applications are discussed.

2 IRT method

Any dynamic (or static) system with multiple input and output variables, $x = (x_1 \ x_2 \ \dots)^T$ and $y = (y_1 \ y_2 \ \dots)^T$, respectively, can be described simplistically as,

$$y(t) = G(x(t); \theta) + e(t), \quad (1)$$

where θ are the system parameters and e stands for the stochastic variations in y . The function G determines the internal structure of the system. For example, G can stand for a large process model including its control system in which case θ correspond to control parameters, setpoint values and other numerical process parameters. The performance of the system can be evaluated by examining the values of the output variables with respect to the applied inputs. The performance can be evaluated in general by means of quality measures q , see Figure 1. E.g., robustness and accuracy are important concepts in describing the performance of a control system. Such concepts can be measured, e.g., with quality measures like variance and setpoint tracking ability of controlled variables.

The values of the m quality measures, $q = (q_1 \ q_2 \ \dots \ q_m)^T$, corresponding to certain values of n parameters, $\theta = (\theta_1 \ \theta_2 \ \dots \ \theta_n)^T$, are calculated from a set of input and output signals, x and y , using some user defined smooth continuous functions. The signals can be recorded from the existing system or they can be produced using a simulation model of the system in consideration. Since the system outputs y are affected by random variations the quality measures q are also random variables. However, it turns out that there exists a statistical dependency between θ and q , which can be modelled if enough data is available. Starting from some initial values of θ and varying their values randomly in a local domain, a set of data, (Θ, Q) , is obtained. Notice that the lower case symbols (e.g., θ and q) are used here for the single samples of (multidimensional) data vectors and the upper case symbols (e.g., Θ and Q) are used for the data matrices such that $Q = (q(1) \ q(2) \ \dots \ q(k))^T$ and $\Theta = (\theta(1) \ \theta(2) \ \dots \ \theta(k))^T$, where k is the number of data points. The data acquisition can be

done with Markov Chain Monte Carlo simulation, or to minimize k , with a pseudo-random sampling of the parameter space (experiment design).

If the joint distribution of Θ and Q forms a single multivariate Gaussian distribution, the dependency between θ and q is linear in the maximum likelihood sense (see, e.g., Pindyck and Rubinfeld, 1991),

$$\Delta q = F^T \cdot \Delta \theta + \varepsilon, \quad (2)$$

or using the matrix notation,

$$\Delta Q = \Delta \Theta \cdot F + E. \quad (3)$$

In equations (2) and (3), Δ corresponds to deviation from the current performance \bar{q} , i.e., $\Delta q = q - \bar{q}$, and ε and E are the residual errors of the linear model. Notice that in the quality measure calculation the time axis is abstracted away and therefore even the dynamic systems are now examined using static models. Now, θ and q can be interpreted as variables on the slower emergent scale, the dynamics of the original variables x and y being considerably faster.

The performance of a large process and its control system requires several quality measures to be characterized thoroughly. Also the number of the parameters rises along the size of the system. Thus, the size of the linear equation system (2) also grows making the determination of the mapping matrix F nontrivial. If a sufficient amount of data from the system can be obtained the local linear model can be estimated using statistical multivariate methods like Partial Least Squares (PLS) or Canonical Correlation Analysis (CCA) based regression techniques (Hyötyniemi, 2001). Instead of concentrating on the correlations among θ (Principal Component Analysis) the aim is now to capture cross-correlations between θ and q .

The tuning task of the parameters can be formulated as a quadratic optimization problem,

$$J = q^T W q, \quad (4)$$

where J is the cost function to be minimized and W is a diagonal weighting matrix. The problem cannot be solved analytically and nonlinear (numerical) gradient based optimization methods, such as conjugate gradient method (see, e.g., Gill et al., 1981), have to be applied. The cost function can be approximated locally as

$$\begin{aligned} \hat{J} &= \hat{q}^T W \hat{q} \\ &= (\Delta \theta^T F + \bar{q}) W (F^T \Delta \theta + \bar{q}), \end{aligned} \quad (5)$$

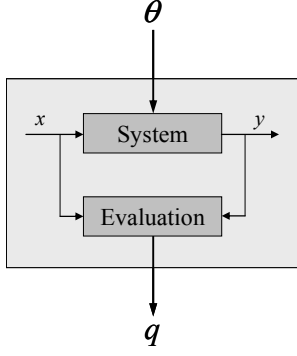


Figure 1: The statistical dependency between parameters, θ , and quality measures, q , describes the relevant properties of the underlying dynamic MIMO system, having inputs u and outputs y (Hyötyniemi, 2002).

which is differentiable with respect to θ . Using the following update principle, the current values of the parameters, $\bar{\theta}$, can be tuned iteratively,

$$\bar{\theta}(K+1) = \bar{\theta}(K) + \gamma d(K), \quad (6)$$

where K is the update step index, γ is the length of the update step and d is the search direction, which is determined by the chosen optimization procedure. E.g., in the basic gradient descent method d is calculated by differentiating (5) with respect to θ , and in the moment method exponential smoothing is applied to combine the successive gradients to obtain a more robust search direction.

The IRT method requires reasonable initial values for θ since local optimization methods are used. When concentrating on control parameter tuning there exists a great number methods to assist in single input single output (SISO) controller tuning (see, e.g., Harris et al., 1999; Thornhill and Hägglund, 1997; Hägglund, 1999). In simulation model tuning the domain area expertise is used to obtain reasonable starting values for the model parameters. The IRT procedure requires also some steering, e.g., with respect to range of the local variation of θ , appropriate selection of γ , and number of data points suitable for obtaining the local model F . To obtain somewhat automated version of the same procedure the following enhancements are proposed.

2.1 Gaussianity testing

Based on domain area expertise the level of the local variation of θ is specified. To enhance the signal-to-noise ratio it is desirable to increase the local variation, but on the contrary, the local linearity between

θ and q is lost if too large area of the parameter space is studied in one iteration step. Therefore, some means for Gaussianity testing on q are required.

When selecting an appropriate test one has to consider the power of the test, i.e., what is the probability that the incorrect null hypothesis on Gaussianity is rejected. Second, the distribution of the test statistic or at least some critical points of it should be known. And third, the hypothesis testing can be done using either known or unknown distribution as an alternative hypothesis. If the alternative distribution (the form of non-Gaussianity) is unknown, an omnibus test has to be chosen (like in this case).

First it is assumed that it is sufficient to consider the marginal distributions of the multinormal distribution which give the necessary (but not sufficient) conditions for the multivariate normality. This simplification can be argued with the fact that the multivariate normality tests are computationally fairly intensive, and their sensitivity to different types of departures from the multinormality vary significantly, i.e., it has been recommended to use more than one test especially when the characteristics of the non-normality are unknown a priori (Thode, 2002). Further, since many of the multivariate tests are extensions of the corresponding univariate tests, they require the calculation of the order statistics, which cannot be done explicitly in the multivariate case.

Three most powerful univariate omnibus tests are Wilk-Shapiro, Anderson-Darling and the joint kurtosis-skewness tests (Thode, 2002). The Wilk-Shapiro test is the most attractive since the distribution of the test statistic is known. Also, if the improvements suggested by Royston (Royston, 1982) are applied, the Wilk-Shapiro test is valid for sample sizes $7 \leq k \leq 2000$. The test compares the observations to the theoretical order statistics of the Gaussian distribution.

The Gaussianity testing also serves as an outlier detection method, since the tests are typically sensitive to strongly differing samples. It can also point out the improper quality measure definitions.

2.2 Reliability of the local linear model

No matter how well the linearity assumption is satisfied, the tuning procedure always returns some search direction at every iteration step. Since the local linearity assumption may have been violated severely, the reliability of the following update step in that direction should be somehow taken into account. One way to accomplish this is to study the confidence intervals of the estimated parameters F of the linear model which gives at the same time confidence intervals for the gradient vector. This

can be done, e.g., using Bayesian analysis of the regression model coefficients (Gelman et al., 2004). The posterior joint distribution of the model parameters F and the residual variance σ_ε^2 ,

$$p(F, \sigma_\varepsilon^2 | q, \theta) = p(F | \sigma_\varepsilon^2, q, \theta) p(\sigma_\varepsilon^2 | q, \theta), \quad (7)$$

can be utilized to obtain confidence intervals for the estimated linear model. In (7) the first term on the right is the conditional posterior distribution of F , given σ_ε^2 , and the second term is the marginal posterior distribution of σ_ε^2 . The confidence intervals can assist to define a suitable length for the update step, γ . If it is uncertain that the update step in the search direction results in an improvement, the value of γ should be reduced. Further, wide confidence intervals suggest that more data should be applied in the parameter estimation to obtain more reliable prediction.

2.3 Iterative PLS regression

In a data-based iterative tuning of θ (6) the local linear model (2) is used in the approximation of the local gradient to determine d . Usually, if n and m are reasonably small, it is sufficient to solve F using standard multilinear regression (MLR). In many cases, however, MLR fails due to collinearities and noise in the multidimensional data. These numerical problems can be avoided using more advanced multivariate regression methods.

The PLS regression model can be defined using an eigenproblem-oriented framework (Hyötyniemi, 2001). The resulting PLS latent basis coincides with the traditional algorithmic definition of PLS only with respect to the most significant basis vector. However, the method provides one with means for iterative PLS regression.

The starting point in finding a good latent variable basis is to maximize the correlation between the x and y . Assuming that the input data X is projected onto a latent basis φ , and the output data Y is projected onto basis ϕ , the objective is to find such bases that the correlations between the projected variables are maximized. To obtain a unique solution the lengths of the basis vectors have to be restricted somehow. This results in a constrained optimization problem,

$$\begin{cases} f(\varphi_i, \phi_i) = \frac{1}{k} \cdot \varphi_i^T X^T \cdot Y \phi_i \\ g_1(\varphi_i) = 1 - \varphi_i^T \varphi_i \\ g_2(\phi_i) = 1 - \phi_i^T \phi_i \end{cases} \quad (8)$$

where the subscript i refers to a single latent basis vector. Using the method of Lagrangian multipliers to solve (8) results in a pair of eigenvalue problems,

$$\begin{cases} \frac{1}{k^2} X^T Y Y^T X \varphi_i = \lambda_i \varphi_i \\ \frac{1}{k^2} Y^T X X^T Y \phi_i = \lambda_i \phi_i \end{cases} \quad (9)$$

where λ_i is the eigenvalue corresponding to the eigenvectors φ_i and ϕ_i . It is sufficient to solve one of the equations in (9). The eigenvalue decomposition of the cross covariance matrix of the projected data gives the complete set of latent basis vectors from which the N most significant can be chosen. E.g., if the first equation in (9) is utilized, the calculation of the covariance matrix can be expressed in an iterative form using exponential smoothing,

$$R(K+1) = \mu R(K) + (1-\mu) X^T Y Y^T X, \quad (10)$$

where $\mu \in (0, 1)$ is the forgetting factor and R is the cross covariance matrix. Using the iteratively updated R in the determination of the PLS latent basis, one gets rid of the zigzag effects plaguing the standard gradient descent algorithm. This can be seen as an alternative optimization method to moment method for example. Now the whole covariance structure of the data is preserved from one iteration step to another instead of only the gradient vector.

3 Pulp digester case study

A model of a Finnish pulp mill was applied in the test case. The process model was constructed with the APROS software that is a professional simulation environment for modeling combustion and nuclear power plants, and pulp and paper mills (see, e.g., Juslin and Paljakka, 1997). APROS provides large libraries of process and automation components, which can be combined into rigorous models of industrial plants. APROS has been used successfully, e.g., in various training simulator and process analysis projects. In the following, the model of the pulp mill is introduced, after which the tuning targets of the case study are formulated.

3.1 Model and problem description

A general overview of the model is presented in Figure 2. First, the woodchips and the impregnation liquor are mixed, after which the mixture is fed into the impregnation vessel. The flow continues to the top of the digester, where the mixture is heated with

steam to the cooking temperature. The chips were modelled to consist of several cellulose, carbohydrate and lignin components, and the liquor was assumed to contain sodium hydroxide and sodium sulphide in addition to the organic compounds dissolved from the woodchips. Chemical reactions during the cooking phase were modelled according to (Gustafson et al. 1983). Several circulation streams of black liquor to the preceding parts of the process were modelled. The subsequent washing and bleaching operations were excluded from the model.

In the case study, the production of pulp in a steady state operation was considered, i.e., changes of neither production rate nor quality targets were simulated. Initially, several control loops were behaving poorly. The level controllers in the impregnation and the digester vessels were oscillating heavily due to their inappropriate tuning. Consequently, also the washing coefficient control failed to meet its targets. The two model-based controllers were behaving even more detrimentally. The other one predicts the kappa number of the produced pulp based on the digester top temperature measurement. The five hour time delay in the cooking process makes the prediction essential for the process control. The other model calculates the setpoint value for the H factor based on the measured temperature profile of the digester, the amount of applied alkali and the kappa number of the produced pulp. Both models were giving strongly biased predictions, causing naturally serious problems. Due to the improper controller tuning the process was producing continuously out of specification pulp.

3.2 Optimization targets

The optimization cost function was formed with six quality measures. The six considered variables were:

- y_1 : kappa number in the digester blow,
- y_2 : washing coefficient,
- y_3 : liquor level in the digester,
- y_4 : chip level in the digester,
- y_5 : chip level in the impregnation vessel, and
- y_6 : H factor prediction.

The control error of the variables $y_1 - y_5$ and the prediction error of y_6 was tried to minimize. The tuning involved the parameters of seven PI controllers and two models that were applied in prediction and setpoint calculation, i.e., the total number of parameters was $n = 20$. The PI controllers were responsible for regulation of the chip and liquor levels, washing coefficient, production rate, H factor, and the digester steam chamber temperature.

Here are presented the results of $K = 30$ iteration steps. The fairly great number of global steps is due to the applied gradient method. Also, since the initial performance was purposefully set to extremely poor, long tuning time was understandable. In the case study $k = 50$ data points on the average were simulated on each iteration step and the length of a simulation run was $T = 8$ h. Simulations were run about 25 times faster than real time.

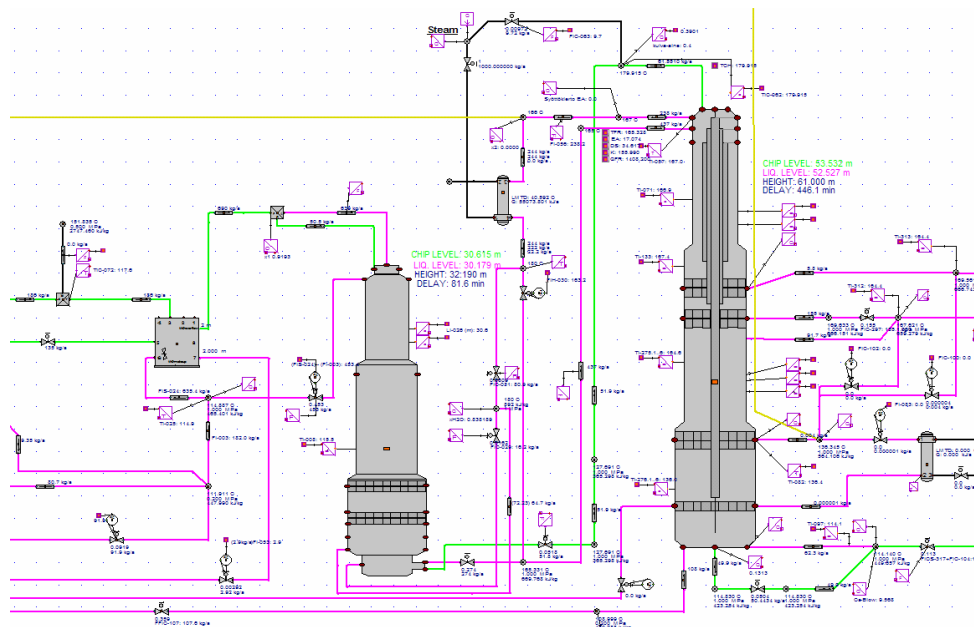


Figure 2: The model of the pulp mill: The chip feed screw conveyor (left), the impregnation vessel (in the middle), and the pulp digester (right).

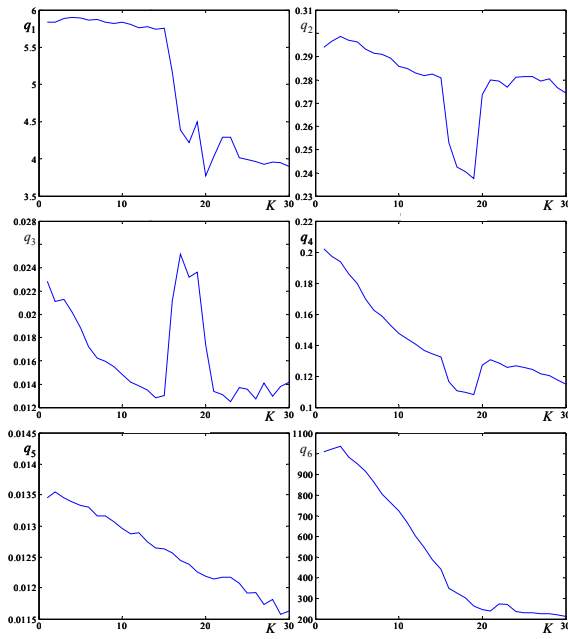


Figure 3: The quality measure values, q_i , $i = 1 \dots 6$, in iterative optimization steps, $K = 1 \dots 30$.

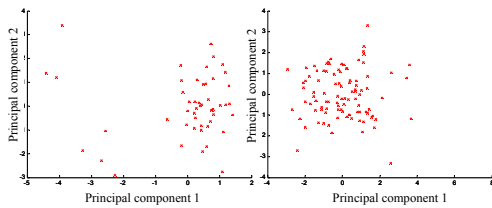


Figure 4: Distributions of q values projected to the plane spanned by the two major principal components, $K = 18$ (left) and $K = 30$ (right).

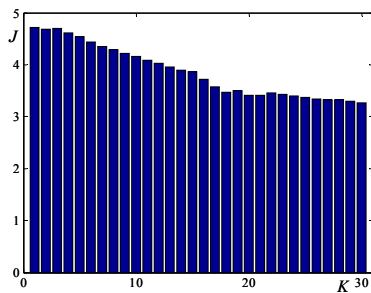


Figure 5: The values of the cost function, J , in the iterative optimization, $K = 30$.

The IRT method succeeded in improving the process performance regarding the six quality measures (see Figure 3). E.g., the kappa number deviation from the setpoint, q_1 , diminished from about 6 to 4 units, and the absolute value of the H factor

prediction error, q_6 , was reduced from about 1000 to 200 units. For softwood pulp the kappa number is typically tried to keep within $\pm 2-3$ units range from the target. In that sense the control is not yet satisfactory but the tuning succeeded to improve its performance notably. The values of q include some stochastic variation and therefore the trends are not monotonically descending. Also occasional “outliers” can be perceived in the values of the quality measures in the global iteration steps 15-20. The convenient linear model was justified with the assumption of Gaussian data. The distributions of the quality measure data in global steps $K = 18$ and $K = 30$ are shown in Figure 4. Obviously the distribution in the left figure does not fulfil the Gaussianity assumption, which explains the outliers in Figure 3.

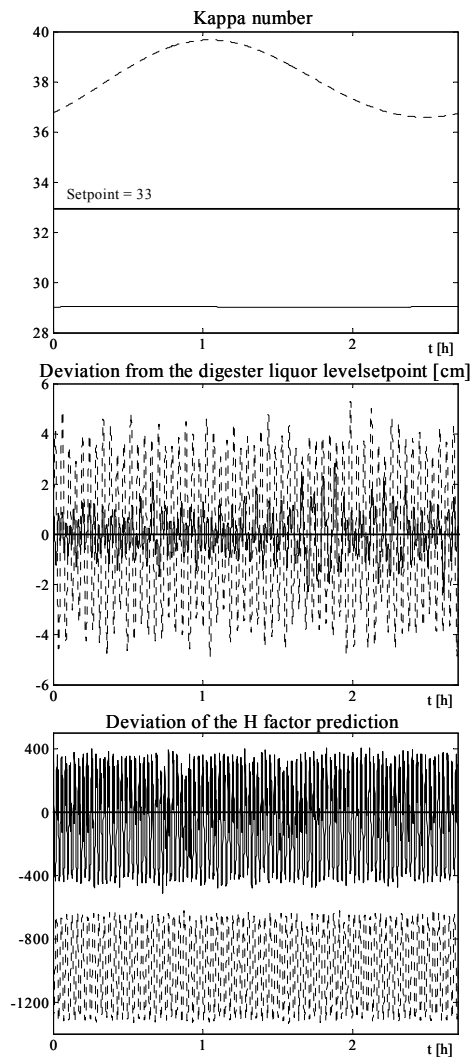


Figure 6: Simulation results with initial and tuned parameters, $K = 1$ (dotted) and $K = 30$ (solid).

The success of the optimization can be followed also from the values of J (see Figure 5). It can be seen how the conflicting targets finally started to slow down the tuning procedure. If the objectives are not met at the end, one has to reconsider the weighting of q_i . It is the only way to continue the optimization among Pareto optimal solutions. Alternatively, reformulation of the quality measures may yield in better results. Examples of the improved performance are presented in Figure 6. The fluctuation of the kappa number has stopped, variance of the digester liquor level has diminished and the H factor prediction has improved tremendously.

4 Conclusions

The IRT method offers a holistic approach to manage large complex systems, like control systems of industrial processes. Dimensional complexity in systems becomes manageable by using statistical multivariate methods. The IRT method opens up new possibilities for control engineering: Different control structures can be tuned to their maximum performance and compared with each others before implementation.

The stochastic nature of the optimization problem advocates the use of methods that diminish the effects of noise. In that sense IRT as a local optimization method using linear regression to approximate the cost function locally seems to be a more tempting approach than, e.g., genetic algorithms.

The development work on IRT method continues. One object of concern is the minimization of the cost function evaluations during the tuning procedure, since one function evaluation of (6) requires reasonably long time series data from the system. This increases the time and computational power needed for the tuning procedure. One way to minimize the amount of data is to involve a pseudo-random parameter space sampling in the data acquisition instead of completely random variation.

Future research will also include studies on simulation model tuning. The objective is to tune the model parameters to give the best output prediction possible with the chosen model structure. An example of the model parameter tuning is given by Halmevaara and Hyötyniemi (2006).

The goals of evolution in technical systems differ from those of natural neocybernetic systems. The energy transfer maximization does not result in performance that satisfies the ever-tightening standards of profitability in process industry. Instead, maximization of the environment with the intended system functionalities gives practically applicable working practices for control engineering. The tuning of environmental variables towards better system performance along the axes of freedoms (de-

termined by the PLS latent basis) in the parameter space can be seen as a directed technical evolution of the system.

References

- Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin (2004). *Bayesian Data Analysis*, 2nd Edition, Chapman & Hall / CRC Press, Boca Raton, FL, USA.
- Gill, P.E., W. Murray and M.H. Wright (1981). *Practical Optimization*, Academic Press, London, Great Britain.
- Gustafson, R.R., C.A. Sleicher, W.T. McKean and B.A. Finlayson (1983). Theoretical Model of the Kraft Pulp Process. *Industrial & Engineering Chemistry Process Design and Development*, 22, 87-96.
- Haavisto, O. and H. Hyötyniemi (2006). Neocybernetic Modeling of a Biological Cell. (To be published in) *Proceedings of the 9th Scandinavian Conference on Artificial Intelligence SCAI 2006, Espoo, Finland*.
- Hägglund, T. (1999). Automatic detection of sluggish control loops. *Control Engineering Practice*, 7, 1505-1511.
- Halmevaara, K. and H. Hyötyniemi (2004). Iterative Simulation Based Multivariate Control Parameter Tuning Method. *Proceedings of the 5th EUROSIM Congress on Modeling and Simulation*, Paris, France.
- Halmevaara, K. and H. Hyötyniemi (2005). Performance Optimization of Large Control Systems – Case Study on a Continuous Pulp Digester. *Proceedings of the 15th IFAC World Congress*, Prague, Czech Republic.
- Halmevaara, K. and H. Hyötyniemi (2006). Data-based parameter optimization of dynamic simulation models. (To be published in) *Proceedings of the 47th Scandinavian Simulation Conference SIMS'06, Helsinki, Finland*.
- Harris, T.J., C.T. Seppala and L.D. Desborough (1999). A review of performance monitoring and assessment techniques for univariate and multivariate control systems. *Journal of Process Control*, 9, 1-17.
- Hyötyniemi, H. (2001). *Multivariate Regression – Techniques and Tools*, Helsinki University of Technology, Control Engineering Laboratory Report, Helsinki, Finland.

- Hyötyniemi, H. (2002). On Emergent Models and Optimization of Parameters. *Proceedings of 42nd Scandinavian Simulation Conference SIMS'02*, 45-50. Oulu, Finland.
- Hyötyniemi, H. (2006). *Neocybernetics in Biological Systems*. (To be published in) Helsinki University of Technology, Control Engineering Laboratory Report, Helsinki, Finland.
- Juslin, K. and M. Paljakka (1997). APROS - A Multifunctional Modelling Environment, *Proceedings of the 2nd CSNI Specialist Meeting on Simulators and Plant Analysers*, Espoo, Finland.
- Pindyck, R.S. and D.L. Rubinfeld (1991). *Econometric Models and Economic Forecasts*, 3rd edition, McGraw-Hill, New York, USA.
- Royston, J.P. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics*, 31, pp. 115-124.
- Thode, H.C. (2002). *Testing for Normality*, Marcel Dekker, New York, USA.
- Thornhill, N.F. and T. Hägglund (1997). Detection and diagnosis of oscillation in control loops. *Control Engineering Practice*, 5, 1343-1354.