DATA-BASED DEVELOPMENT OF HYBRID MODELS FOR BIOLOGICAL WASTEWATER TREATMENT IN PULP AND PAPER INDUSTRY

Esko K. Juuso

Control Engineering Laboratory
Department of Process and Environmental Engineering
90014 University of Oulu, Finland

Esko.Juuso@oulu.fi(Esko K. Juuso)

Abstract

Modelling and simulation of biological wastewater treatment in pulp and paper industry requires hybrid models since the operating conditions can fluctuate drastically. A lot of process measurements are available, but measurement sets do not include sufficient information on special features of the influent nor on microbial composition of the sludge. Populations of microorganisms are highly important. A compact dynamic simulation is realized with linguistic equation (LE) models. The models consist of two parts: interactions are handled with linear equations, and nonlinearities are taken into account by membership definitions. Process insight is maintained, while data-driven tuning relates the measurements to the operating areas. Genetic algorithms are well suited for LE models. A new approach based on generalised norms and skewness has been developed for analyzing scaling functions from data sets, which include various operating conditions. Sensitivity to small changes from the optimal conditions is increased considerably. The resulting model has a cascade structure with specialized LE models.

Keywords: dynamic models, nonlinear systems, pulp and paper industry, wastewater, linguistic equations.

Presenting Author's Biography

Esko K. Juuso has M.Sc. (Eng.) in Technical Physics from University of Oulu. He is currently a university teacher in Control Engineering at University of Oulu, Oulu, Finland. He is active in Finnish Simulation Forum (FinSim), Scandinavian Simulation Society (SIMS) and EUROSIM, currently he is secretary of EUROSIM Board, and chairman of FinSim and SIMS. His main research fields are intelligent systems and simulation in industrial applications, including control and fault diagnosis. In 1991 he introduced the linguistic equation (LE) methodology, which is currently used in many applications of modelling, control and fault diagnosis.

1 Introduction

In the pulp and paper industry, a huge amount of water flows through different processes. For environmental and economical reasons, the plant recycles the water as much as possible. Before recycling the water is purified to a certain degree. The chemical treatment is one of the purification methods. Chemical water treatment includes complex nonlinear phenomena such as coagulation and flocculation processes. The dosing control of chemicals is very demanding, and chemicals are quite often dosed on the basis of the flow rate which does not always guarantee the adequate purification efficiency. Modelling of these complicated processes is mainly data-based or empirical due to a lack of comprehensive physical models. Intelligent methods such as linguistic equations and neural networks can be applied for the modelling of nonlinear interactions between input and output variables. [1, 2, 3]

Waste water 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 measurement sets do not include sufficient information on special features of the influent nor on microbial composition of the sludge. Populations of microorganisms are highly important, e.g. sludge bulking cause especially poor treatment efficiency results when biosludge escapes from secondary clarification.

Process simulators are effective for developing, testing and tuning the controllers. Different control methods can be tested safely in changing process conditions without disturbing the process. Furthermore, the chemical dosage can be optimised and the quality of water can be analysed in the simulator. However, a reliable process model is essential for process simulations. For activated sludge plants, modelling is even more demanding since the condition of the biomass need to modelled as well. Mechanistic models have been two decades in active use. The first Activated Sludge Models (ASM) was presented in 1987 [4]. However, the use has been limited by complexity of the models. Lindblom [5] reduced the ASM1 to an activated sludge plant in pulp and paper industry.

AS models are constructed through a step-wise procedure: model purpose definition, model selection, data collection, data reconciliation, calibration of the model parameters and model unfalsification. The model purpose, defined at the beginning of the procedure, influences the model selection, the data collection and the model calibration. In the model calibration a process engineering approach, i.e. based on understanding of the process and the model structure, is needed. [6]

Calibration of the models is challenging because of a large number of variables and parameters. Black-box, stochastic grey-box and hybrid models are useful in waste water applications for prediction of the influent load, for estimation of biomass activities and effluent quality parameters. These modelling methodologies thus complement the process knowledge included in white-box models with predictions based on data in areas where the white-box model assumptions are not valid or where white-box models do not provide accurate predictions. [7]

Many variables are normally measured in a plant, but some of them are strongly cross-correlated. Data-based analysis have been used for variable selection [8, 9, 10]. Clustering data for detection of operating conditions has used in [11] and [12] to get basis for specialised submodels. As the sludge settling properties have remarkable effects on the treatment results, modelling of the diluted sludge volume index (DSVI) is important [11]. In [12] models were used for predicting the chemical oxygen demand (COD) of the effluent in an activated sludge plant treating pulp and paper mill wastewater.

Complex data-based models can be tuned with *genetic algorithms (GAs)*: penalty functions can be used to reduce complexity, i.e. the number of neurons, layers, rules, active coefficients etc. The main challenge is the efficient coding of the alternatives, because formulating the constraints for the parameters is difficult. Good solutions can be found for simple fuzzy models since the labels have a clear sequence [13]. This idea can be extended into the nonlinear scaling approach used in *linguistic equation (LE)* systems [14, 15]. Fitness can be evaluated efficiently with compact, parameterised LE models. In earlier applications, monotonous scaling functions were forced by penalty functions [16]. This facilitates the detection of constraint violations, but in large scale systems the approach is rather ineffective. Improved coding presented in [17] does not require penalty funtions.

This paper presents a new methodology for data-based development of scaling functions for hybrid LE models of biological water treatment.

2 Modelling approach

Data-driven and mechanistic modelling need to be combined with cascade structures and uncertainty handling to obtain simulators for practical applications in biological wastewater treatment.

2.1 Data-driven intelligent modelling

Linguistic equation (LE) models consist of two parts: *interactions* are handled with linear equations, and nonlinearities are taken into account by *membership definitions* [14]. In the LE 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. The linguistic level of the input variable x_j is calculated the inverse functions of the polynomials [15].

Steady state LE models are represented by

$$x_{out} = f_{out}\left(-\frac{\sum_{j=1,j\neq out}^{m} A_{ij}\, f_j^{-1}(x_j) + B_i}{A_{i\ out}}\right) \quad (1)$$

where the functions f_j and f_{out} are membership definitions of input variables x_j and output x_{out}, respectively.

Rather simple input-output LE models, where the old value of the simulated variable and the current value of the control variable as inputs and the new value of the simulated variable as an output, can be used since nonlinearities are taken into account by membership definitions. For the default LE model, all the degrees of the polynomials in parametric models become very low, i.e. all the parametric models become the same

$$y(t) + a_1 y(t-1) = b_1 u(t-n_k) + e(t). \quad (2)$$

This model is a special case with three variables, $y(t), y(t-1)$ and $u(t-n_k)$, the interaction matrix $A = [1\ \ a_1\ \ -b_1]$ and the bias term $B = 0$.

The output, the derivative of the variable y, is integrated with numerical integration methods:

$$y = \int_0^T F(t, y, u)dt + y_0, \quad (3)$$

where T is the time period for integration, and y_0 *the initial condition*. Usually, several values from the integration step or the previous steps are used in evaluating the new value. Step size control adapts the simulation to changing operating conditions.

In water treatment, dynamic LE approach has been used in modelling of flotation units, where the process water is treated with chemicals which react with extractives forming pitch sludge. The dynamic LE model is similar to the model (2): the outlet turbidity, $x_{turb}(t + T_s)$, is here calculated on the properties of incoming water, chemical dosages and previous calculated turbidity, $x_{turb}(t)$. The model is developed for steps equal to the sampling time, T_s. Normal process data and some test campaigns based on experimental design [1].

2.2 Cascade and interactive modelling

Cascade modelling divides the problem into sequential parts to further alleviate the problem of parameters, and the number of parameters is further reduced with principal components. Cascade modelling principle and linear models are essential in various fuzzy and neural methodologies as well, e.g. Takagi-Sugeno (TS) fuzzy models and radial basis networks. [18] The output of a model can be used as a input of several models, and the models may also contain interactions or recycle flows (Fig. 1). Feedback structures are needed in dynamic simulation. Neurofuzzy systems can be constructed as sequential combinations of neural and fuzzy parts, i.e. fuzzy set system provides inputs for a neural network, or neural preprocessing is used for inputs of a fuzzy set system. Variable grouping is important in cascade model structures.

The submodels are developed by the case-based modelling approach. The multimodel system has several submodels and a fuzzy decision system for selecting a good model for each situation (Fig. 2). Linguistic Takagi-Sugeno type fuzzy models (LTS) belong to this class but with one limitation: the fuzzy partition is defined with same variables as the models. As *LE* models are nonlinear, also these local models are nonlinear, see [19].

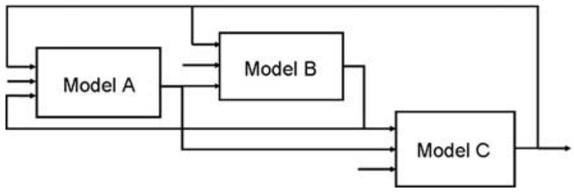

Fig. 1 Interactive models.

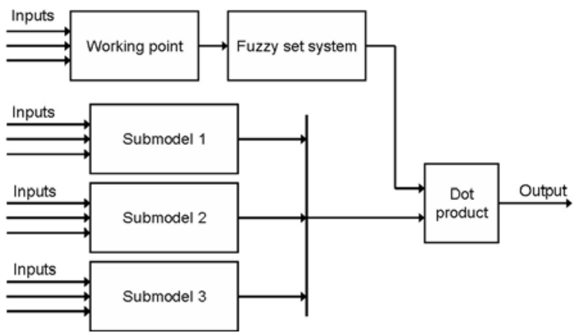

Fig. 2 A multimodel system with a fuzzy decision module.

2.3 Uncertainty

Universal approximators for fuzzy functions can be constructed as extension principle extensions of continuous real-valued functions which continuously map fuzzy numbers into fuzzy numbers [20, 21]. The dynamic LE models with fuzzy inputs were introduced in [22], and later adapted to dynamic modelling in several applications [23]. LE models are extended to fuzzy inputs with this approach if the membership definitions f_j and the corresponding inverse functions, are replaced by corresponding extension principle extensions of these functions. Square root functions are used in the linguistification part.

The argument of the function f_{out} in (1) is obtained by fuzzy arithmetic. Fuzzy LE models with fuzzy inputs can be constructed by using multiplication and division as well. Fuzzy extension of the classical interval analysis [24] suits very well to these calculations.The delinguistication block uses also second order polynomials. Fuzzy extension results a nonlinear membership function for the output even if the membership function of the input is linear. Results of the fuzzy interval analysis have always maximal uncertainty as it takes the worst case. Negative associations between the input variables reduce the uncertainty considerably [23].

3 Data analysis

3.1 Statistical features

The mathematical expectation, expected value, or briefly the expectation, of a random variable is a very important concept in probability and statistics. For a discrete random variable having possible values , the expectation of is defined as

$$E(X) = \sum_{i=1}^{N} x_i P(X = x_i) = \sum_{i=1}^{N} x_i f(x_i), \quad (4)$$

where the probability of each value is defined by means of a probability function P or a density function $f(x_i)$. We assume that all the values are equally probable, i.e. $P(x = x_i) = f(x_i) = \frac{1}{N}$ lx$_i$, $i = 1, \ldots, N$. More details are presented in [25].

The moments can be defined about some central value, e.g. the moments about the mean defined by

$$M^k = E[(X - E(X))^k], \quad (5)$$

where k is a positive integer. Other central values are medians and modes, for example. The variance $Var(X) = \sigma_X^2$ can be represented by

$$\sigma_X^2 = M^2 = E[(X - E(X))^2] = E(X^2) - [E(X)]^2. \quad (6)$$

The positive square root of the variance, σ_X, or briefly σ, which is called the standard deviation, is used more often since it has the same dimension as the variable.

Dimensionless features can be obtained by normalising the moments (5), for example by standard deviation σ_X:

$$\gamma_k = \frac{E[(X - E(X))^k]}{\sigma_X^k}. \quad (7)$$

The feature γ_3 is called the coefficient of skewness, or briefly skewness, and the feature γ_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. For a Gaussian signal $\gamma_4 = 3$. Flatter, also described as long-tailed or heavy-tailed, distributions have $\gamma_4 < 3$ and for distributions with high peakedness $\gamma_4 > 3$. For a sinusoidal signal $\gamma_4 = 1.5$. An alternative definition of kurtosis reduces the ratio by 3 to give the value zero for the normal distribution.

3.2 Generalised norms

A norm defined by

$$||^\tau M_\alpha^p|| = (^\tau M_\alpha^p)^{1/p} = (\frac{1}{N} \sum_{i=1}^{N} |x_i^{(\alpha)}|^p)^{1/p}, \quad (8)$$

where $\alpha \in R$ is the order of derivation, the order of the moment $p \in R$ is non-zero, τ is the sample time. This is l_p norm of $x^{(\alpha)}$ was introduced in [26]. We can write it in an alternative way

$$||^\tau M_\alpha^p|| \equiv ||x^{(\alpha)}||_p, \quad (9)$$

It has same dimensions as $x^{(\alpha)}$. The l_p norms are defined in such a way that $1 \leq p < \infty$. In this study, the order p is allowed to be less than one. The norm (8) is a Hölder mean, also known as the power mean. The norm values increase with increasing order, i.e. for the l_p and l_q norms holds

$$(^\tau M_\alpha^p)^{1/p} \leq (^\tau M_\alpha^q)^{1/q}, \quad (10)$$

if $p < q$. The increase is monotonous if all the signals are not equal.

The absolute mean,

$$||x^{(\alpha)}||_1 = (\frac{1}{N} \sum_{i=1}^{N} |x_i^{(\alpha)}|, \quad (11)$$

the rms value,

$$||x^{(\alpha)}||_2 = (\frac{1}{N} \sum_{i=1}^{N} |x_i^{(\alpha)}|^2)^{1/2}, \quad (12)$$

and the absolute harmonic mean,

$$||x^{(\alpha)}||_{-1} = \frac{N}{\sum_{i=1}^{N} \frac{1}{|x_i^{(\alpha)}|}}, \quad (13)$$

are special cases where the order is 1, 2 and -1, respectively. When the order $p \to 0$, we obtain from (8) the absolute geometric mean, i.e.

$$\lim_{p \to 0} ||x^{(\alpha)}||_p = (\prod_{i=1}^{N} |x_i^{(\alpha)}|)^{1/N}, \quad (14)$$

The norm (9) represents the norms from the minimum to the maximum, which correspond the orders $p = -\infty$ and $p = \infty$, respectively. When $p < 0$, all the signal values should be non-zero, i.e. $x \neq 0$. Therefore, the norms with are reasonable only if all the variable values are positive. [26, 27]

3.3 Generalised moments

The normalised moments (6) are here generalised by replacing with the norm (7) as the central value :

$$\gamma_k = \frac{E[(X^{(\alpha)} - ||^\tau M_\alpha^p||_p)^k]}{\sigma_X^k}. \quad (15)$$

where σ_X is calculated about the origin, and k is a positive integer. [28]

4 Nonlinear scaling in modelling

Membership definitions provide nonlinear mappings from the operation area of the (sub)system to the linguistic values represented inside a real-valued interval $[-2, 2]$, denoted as the *linguistic range*, see [15].

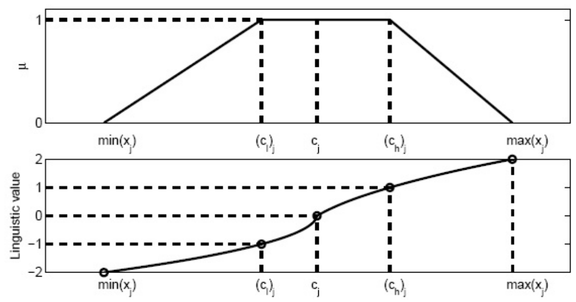

Fig. 3 Feasible range and membership definitions [15].

The concept of a feasible range is defined as a trapezoidal membership function (Fig. 3), which is based on the support and core areas defined in the fuzzy set theory [29]. The support area is defined by the minimum and maximum values of the variable, $\min(x_j)$ and $\max(x_j)$, respectively. The value range of x_j is divided into two parts by the central tendency value c_j, and the core area, $[(c_l)_j, (c_h)_j]$, is limited by the central tendency values of the lower and upper part.

4.1 Scaling functions

In current systems, the membership definitions consist of two second order polynomials: one for negative values, $X \in [-2, 0)$, and one for positive values, $X \in [0, 2]$:

$$\begin{array}{ll} f_j^- = a_j^- X_j^2 + b_j^- X_j + c_j, & X_j \in [-2, 0), \\ f_j^+ = a_j^+ X_j^2 + b_j^+ X_j + c_j, & X_j \in [0, 2]. \end{array} \quad (16)$$

The values X_j are called *linguistic values* because the scaling idea is based on the membership functions of fuzzy set systems. The coefficients of the polynomials are defined by Table 1.

Tab. 1 Corner points of the feasible are.

x_j	X_j
$\min(x_j)$	-2
$(c_l)_j$	-1
c_j	0
$(c_h)_j$	1
$\max(x_j)$	2

4.2 Constraints

As the membership definitions are used in a continuous form, the functions $f_j^-(X_j)$ and $f_j^+(X_j)$ should be monotonous, increasing functions in order to produce realisable systems. In order to keep the functions monotonous and increasing, the derivatives of the functions f_j^- and f_j^+ should always be positive (Fig. 4).

The inequalities for the core and the support are satisfied with

$$\begin{array}{lll} (c_l)_j - \min(x_j) & = & \alpha_j^- \, (c_j - (c_l)_j), \\ \max(x_j) - (c_h)_j & = & \alpha_j^+ \, ((c_h)_j - c_j) \end{array} \quad (17)$$

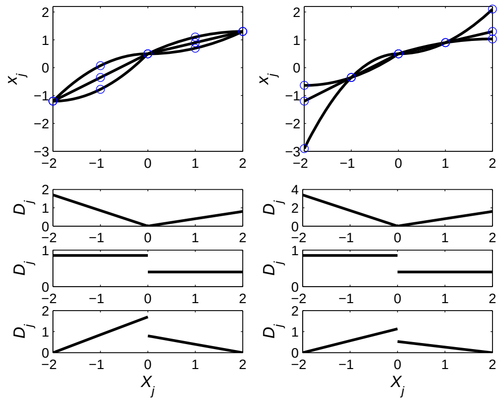

Fig. 4 Feasible shapes of membership definitions f_j and corresponding derivatives D_j: coefficients adjusted with core (left) and support (right). Derivatives are presented in three groups: (1) decreasing and increasing, (2) asymmetric linear, and (3) increasing and decreasing.

if the coefficients α_j^- and α_j^+ are both in the range $\frac{1}{3} \ldots 3$. Corrections are done by changing the borders of the core area, the borders of the support area or the centre point. Additional constraints for derivatives can also be taken into account. The coefficients of the polynomials can be represented by

$$\begin{array}{lll} a_j^- & = & \frac{1}{2}(1 - \alpha_j^-) \, \Delta c_j^-, \\ b_j^- & = & \frac{1}{2}(3 - \alpha_j^-) \, \Delta c_j^-, \\ a_j^+ & = & \frac{1}{2}(\alpha_j^+ - 1) \, \Delta c_j^+, \\ b_j^+ & = & \frac{1}{2}(3 - \alpha_j^+) \, \Delta c_j^+, \end{array} \quad (18)$$

where $\Delta c_j^- = c_j - (c_l)_j$ and $\Delta c_j^+ = (c_h)_j - c_j$. Membership definitions may contain linear parts if some coefficients α_j^- or α_j^+ equals to one (Fig. 4).

The best way to tune the system is to first define the working point and the core $[(c_l)_j, \ (c_h)_j]$, then the ratios α_j^- and α_j^+ from the range $\frac{1}{3} \ldots 3$, and finally to calculate the support $[\min(x_j), \ \max(x_j)]$. The membership definitions of each variable are configured with five parameters, including the centre point c_j and three consistent sets: the corner points $\{\min(x_j), (c_l)_j, (c_h)_j, \max(x_j)\}$ are good for visualisation; the parameters $\{\alpha_j^-, \Delta c_j^-, \alpha_j^+, \Delta c_j^+\}$ are suitable for tuning; and the coefficients $\{a_j^-, b_j^-, a_j^+, b_j^+\}$ are used in the calculations. The upper and the lower parts of the scaling functions can be convex or concave, independent of each other. Simplified functions can also be used: a linear membership definition only requires 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 $b_j^+ \neq b_j^-$.

Additional constraints can be taken into account for derivatives, e.g. locally linear function results if con-

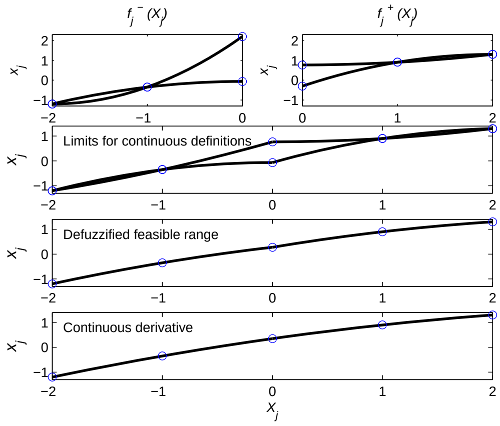

Fig. 5 Membership definitions in the core: coefficients adjusted with the centre point c_j.

tinuous derivative is chosen in the centre point:

$$c_j = \frac{1}{6}\left(4\,(c_l)_j + 4\,(c_h)_j - \min(x_j) - \max(x_j)\right). \qquad (19)$$

The continuity requirement limits the ranges of the ratios α_j^- and α_j^+ if the functions are adjusted by moving the centre point (Fig. 5). Inequalities (17) should be satisfied also after changing c_j, which may require iteration.

4.3 Corner points with generalised moments and norms

The analysis of the corner points has earlier based on mean or median values. The value range of x_j is divided into two parts by the central tendency value c_j and the core area, $[(c_l)_j, (c_h)_j]$, is limited by the central tendency values of the lower and upper part. There are problems when the value range is very wide or the distribution is very concentrated. A new approach based on (15) is introduced here for estimating the central tendency value and the core area, $\alpha = 0$. The central tendency value is chosen by the point where the skewness changes from negative to positive, i.e. $\gamma_3 = 0$. 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 $(c_h)_j$, are the points where the direction of the skewness changes. The iteration is performed with generalised norms. Then the ratios α_j^- and α_j^+ are restricted to the range $[\frac{1}{3}, 3]$ moving the corner points $(c_l)_j$ and $(c_h)_j$ or the upper and lower limits $\min(x_j)$ and/or $\max(x_j)$. The linearity requirement (19) is taken into account, if possible.

The nonlinear scaling methodology based on generalised norms and skewness provides good results for the automatic generation of scaling functions [28]. Sensitivity to small faults and anomalies was increased considerably. The approach was tested with normal, Poisson and Weibull distributions and with two applications of condition monitoring.

4.4 Genetic tuning

Structural restrictions of the LE models are beneficial for tuning and adaptation. Manual tuning consists two independent parts: scaling and interactions. Both can be defined manually but the constraints must be taken into account when defining the scaling functions. Neural tuning reduces the error between the model and the training data [14]. Genetic algorithms 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 [17]. Scaling functions of the variables and time delays are tuned with the parameter sets

$$\{c_j,\ \alpha_j^-,\ \Delta c_j^-,\ \alpha_j^+,\ \Delta c_j^+\}. \qquad (20)$$

This coding is essential for handling hybrid models. The fitness values of the models are obtained through simulation, and the new approach reduces the simulation runs considerably. Genetic tuning is controlled by population size, number of bits, and probabilities of crossover and mutation.

5 Activated sludge plant

Biological water treatment depends strongly on changes in inlet water quality. Changes in biological state influence on the purification result and subsequent process phases. The objective of the project is to develop a model based optimisation and control concept for detecting process conditions and comparing control actions to improve operation of an activated sludge plant (Fig. 6). On-line measurements and laboratory analysis are combined to build indirect measurements and intelligent dynamic models. Intelligent methods are needed in the models of the aeration basin. Uncertainty handling is an essential part of the models. The concept is tested in connection to industrial purification processes.

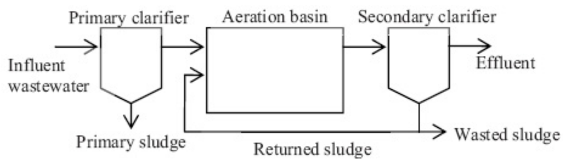

Fig. 6 Activated sludge plant.

5.1 Measurements

Influent quality depends on suspended solids (SS), chemical oxygen demand (COD), biological oxygen demand (BOD) and concentrations of nitrogen and phosphorus. In pulp and paper industry, additional nitrogen and/or phosphorus dosing is needed to keep the biomass in good condition. Changes in biomass population may cause sludge bulking which is seen as deterioration of sludge settling properties, described with sludge volume index (SVI) or diluted sludge volume index (DSVI). For example, if there is lack of oxygen or

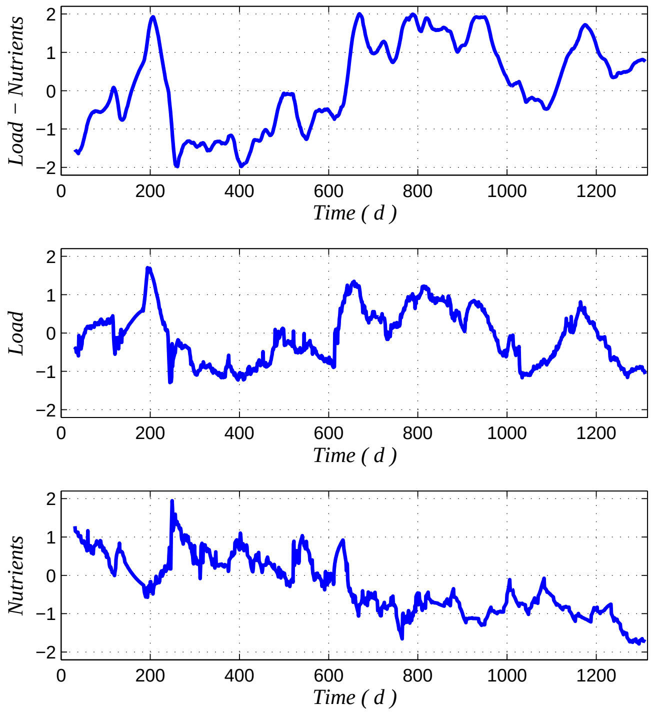

Fig. 7 Levels of load and nutrients.

nutrients compared to biomass population, filamentous sludge leads into poor settling properties.

Changes in activated sludge process are slow, especially recovering from the bulking state to normal operation takes time. There significant seasonal effects, e.g. temperature is typically some degrees lower in winter time. On the other hand, cooling problems may cause temperature rise in summer time. In addition pH, dissolved oxygen profile have obvious effects to the biomass population. Considerable changes of influent quality can be seen in conductivity.

The control variables such as sludge age, COD/nutrient rate, sludge loading, and recycle ratio can be derived from the measurements. The treatment efficiency is assessed by reduction of total nitrogen, total phosphorus, and total COD. Effective time delays should be taken into account, and an additional challenge is that these time delays are varying. Naturally, the delays depend on the flow rates, but also the changes of kinetics have their effects.

5.2 Load and nutrients

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. The feed of nutrients has decreased most of the time (Fig. 7). There was a period of high nutrient feed when the load was fairly low. However, later the load has been increased, which have caused an opposite situation. Both the load and the nutrients have been scaled to the range $[-2, 2]$ with the new method of defining the corner points (Section 4.3). Sensitivity to small deviations from the normal is clearly increased. Averages obtained from longer time periods have been used in calculating the balance *Load - Nutrients*.

5.3 Operating conditions

The operating conditions are modified by oxygen, temperature and flow (Fig. 8). As the flow comes from the production process, it cannot be made suitable for the treatment process. The temperature should be kept in optimal range. However, in summertime heat exchanger and additional cooler water have not been sufficient to do that, when the flow has been very high. Correspondingly, in winter time the temperature of the basin has decreased when the flow has been on a lower level. Effective time delays also depend on the flow, which is taken into account in all calculations. The oxygen is tried to keep on a low level in order to reduce energy costs.

5.4 Treatment results

Treatment results are here analysed by comparing the COD reduction and DSVI values (Fig. 9). The load and the nutrients should be balanced, i.e. the difference of the load and nutrient level should be close to zero

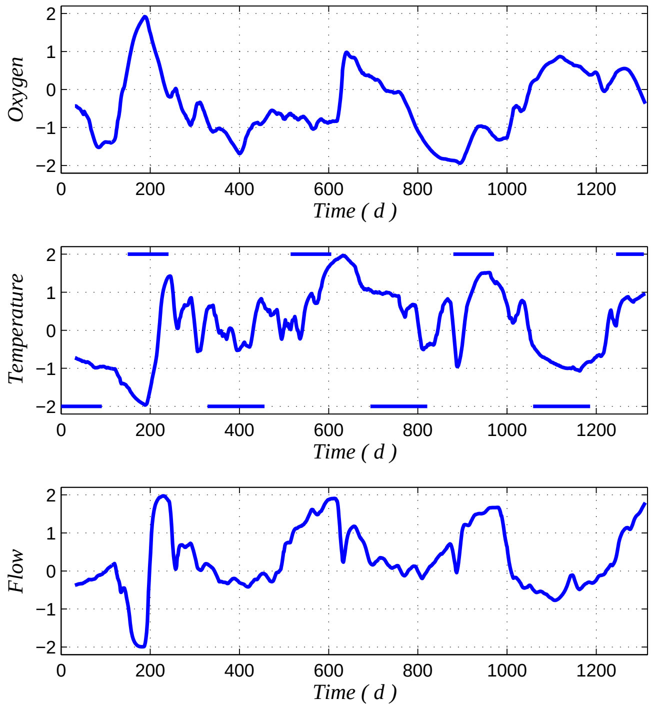

Fig. 8 Levels of working point variables: load, nutrients, oxygen and temperature.

(Fig.7). Too high nutrient level compared to the load causes poor settling seen as an increase of the DSVI, which continues as an oscillating behaviour. On the other hand too low nutrient level causes problems in settling. The normal levels are the best also for the temperature and the oxygen (Fig. 8): too high and low temperatures affect to the biomass; too low oxygen levels are harmful and too high levels mean excess energy consumption.

Problems of simple data-driven modelling approaches clearly understood by comparing these results. The operation of the treatment process depends strongly on the condition of the biomass, whose changes are fairly slow as can be seen from the DSVi values (Fig.9).

5.5 Intelligent modelling

Modelling methods used first in chemical water treatment [1, 2, 3] have been extended to biological wastewater treatment. The nonlinear scaling approach presented in [15] is the basis of these models. Measurements were extracted from databases of the pulp mill of Stora Enso Fine Paper in Oulu. Hybrid models combine mechanistic models [7], data-driven models [30, 31], and intelligent analysers [18]. The models are aimed for process control [32]. The LE models developed for the intelligent analysers are the basis of the hybrid approach as well. Mechanistic part is mainly devoted to analysing the effects of the fluctuations in the influent wastewater. The models of the clarifiers originate from material balances.

The model consists of three interactive models (Fig. 1): Model A calculates the load from the inflow, COD and suspended solids. Model B is the heart of the system. The load, nutrients, oxygen and temperature are used in the model of the sludge settling. The multimodel system shown in Figure 2 should be based on the biomass population, which is here assessed by DSVI. Smooth transitions between the submodels can be handled with fuzzy reasoning by using the degrees of membership obtained from the difference *Load - Nutrients*. In the present model, this part is done with LE models, which provide continuous smooth transition between different operating conditions and make the coding easier. 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. As the COD and BOD reduction depend strongly on the biomass.

The individual dynamic models are developed by using similar structures as in flotation models. Effective time delays are taken into account in these models. Uncertainty handling needs to be included, since the measurement material is rather sparse, especially for on features of the influent and microbial composition.

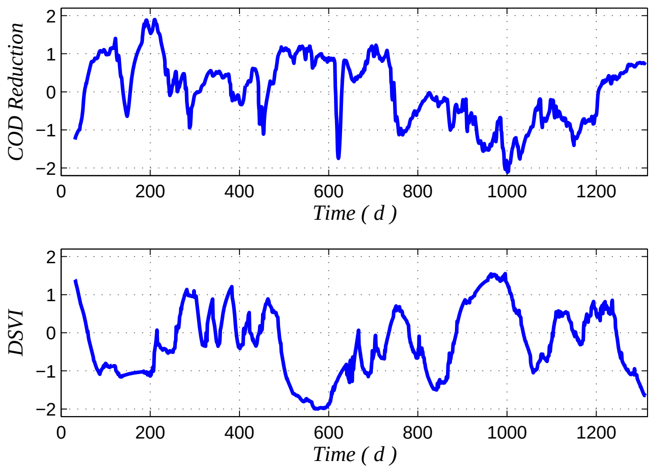

Fig. 9 COD reduction and diluted sludge volume index (DSVI).

6 Conclusions

The LE approach provides compact models in which all the parameters have clear constraints, and the whole system can be assessed with expert knowledge. This advantage becomes increasingly important in cascade and interactive models. The nonlinear scaling methodology based on generalised norms and skewness provides good results for the LE modelling. Sensitivity to small changes from the optimal conditions is increased considerably.

Acknowledgements

This research was supported by the Finnish Funding Agency for Technology and Innovation (Tekes), Kemira Oyj, Stora Enso Oyj and UPM-Kymmene Oyj.

7 References

[1] I. Ainali, M. Piironen, and E. Juuso. Intelligent water quality indicator for chemical water treatment unit. In *Proceedings of SIMS 2002 - the 43rd Scandinavian Conference on Simulation and Modelling, September 26-27, 2002, Oulu, Finland*, pages 247–252. Oulu, 2002.

[2] I. Joensuu, M. Piironen, and E. Juuso. Adaptive feedback controller for dosage of water treatment chemicals. In K. Leiviskä, editor, *Proceedings of AFNC'04 - the 2nd IFAC Workshop on Advanced Fuzzy/Neural, September 16-17, 2004, Oulu, Finland*, pages 127–131. Finnish Automation Society, Helsinki, 2004.

[3] I. Joensuu, M. Piironen, and E. Juuso. Dynamic simulator for dosing of water treatment chemicals. In L. Puigjaner and A. Espuna, editors, *Proceedings of European Symposium on Computer Aided Process Engineering-15 (Escape-15), Barcelona, Spain, May 29 - June 1, 2005, Computer-aided chemical engineering, 20A*, pages 301–306. Elsevier, Amsterdam, 2005.

[4] M. Henze, C. P. L. Grady Jr., W. Gujer, G. V. R. Marais, and T. Matsuo. Activated sludge model no. 1., IAWQ scientific and technical report no. 1. London, UK, 1987.

[5] E. Lindblom. *Dynamic modelling of nutrient deficient wastewater treatment process. M.Sc. Thesis*. Lund University, Lund, Sweden, 2003. TEIE-5175.

[6] K. V. Gernaey, M. C. M. van Loosdrecht, M. Henze, M. Lind, and S. B. Jørgensen. Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environmental Modelling & Software*, 19:763–783, 2004.

[7] J. La Cour Jansen, J. Keskitalo, and E. Juuso. Oxygen uptake rate measurements as a tool for deterministic modelling of pulp and paper wastewater treatment plants. In I. Troch and F. Breitenecker, editors, *6th Vienna Conference on Mathematical Modelling - MATHMOD 2009, February 11-13, 2009, Argesim Report no. 35*, pages 1008–1013. Argesim, 2009.

[8] P. Teppola, S.-P. Mujunen, and P. Minkkinen. Partial least squares modeling of an activated sludge plant: A case study. *Chemometrics and Intelligent Laboratory Systems*, 38(2):197–208, 1997.

[9] S.-P. Mujunen, P. Minkkinen, P. Teppola, and R.-S. Wirkkala. Modeling of activated sludge plants treatment efficiency with PLSR: a process analytical case study. *Chemometrics and Intelligent Laboratory Systems*, 41(1):83–94, 1998.

[10] K. P. Oliveira-Esquerre, M. Mori, and R. E. Bruns. Simulation of an industrial wastewater treatment plant using artificial neural networks and principal component analysis. *Brazilian Journal of Chemical Engineering*, 19(4):365–370, 2002.

[11] M. Heikkinen, T. Latvala, E. Juuso, and Y. Hiltunen. Som based modelling for an acti-

vated sludge treatment process. In *Tenth International Conference on Computer Modelling and Simulation, EUROSIM/UKSim, Cambridge, UK, April 13, 2008.*, pages 224–229. The Institute of Electrical and Electronics Engineers IEEE, 2008.

[12] M. Heikkinen, T. Heikkinen, and Y. Hiltunen. Modelling of activated sludge treatment process in a pulp mill using neural networks. In *The 6th International Conference on Computing, Communications and Control Technologies: CCCT 2008, Orlando, Florida, USA, June 29th - July 2nd 2008.*, page 6 pp. 2008.

[13] S. Lotvonen, S. Kivikunnas, and E. Juuso. Tuning of a fuzzy system with genetic algorithms and linguistic equations. In H.-J. Zimmermann, editor, *Proceedings of the 5th European Congress on Intelligent Techniques and Soft Computing - EUFIT'97, Aachen, September 8 - 11, 1997*, volume 2, pages 1289–1293, Aachen, 1997. Mainz.

[14] E. K. Juuso. Fuzzy control in process industry: The linguistic equation approach. In H. B. Verbruggen, H.-J. Zimmermann, and R. Babuska, editors, *Fuzzy Algorithms for Control, International Series in Intelligent Technologies*, pages 243–300. Kluwer, Boston, 1999.

[15] E. K. Juuso. Integration of intelligent systems in development of smart adaptive systems. *International Journal of Approximate Reasoning*, 35:307–337, 2004.

[16] E. K. Juuso. Modelling and simulation in development and tuning of intelligent controllers. In I. Troch and F. Breitenecker, editors, *Proceedings of 5th Vienna Symposium on Mathematical Modelling, February 8-10, 2006, Vienna, Austria, Argesim Report 30*, pages 6–1 – 10. Argesin Verlag, Vienna, 2006. ISBN 3-901608-30-3.

[17] E. K. Juuso. Tuning of large-scale linguistic equation (LE) models with genetic algorithms. In M. Kolehmainen, editor, *Proceedings of the International Conference on Adaptive and Natural Computing Algorithms - ICANNGA 2009, Kuopio, Finland, Lecture Notes in Computer Science*, volume LNCS 5495, pages 161–170. Springer-Verlag, Heidelberg, 2009.

[18] E. Juuso, , T. Latvala, and I. Laakso. Intelligent analysers and dynamic simulation in a biological water treatment process. In I. Troch and F. Breitenecker, editors, *MATHMOD 2009, February 11-13, 2009, Argesim Report no. 35*, pages 999–1007. Argesim, 2009.

[19] E. K. Juuso. Development of multiple linguistic equation models with takagi-sugeno type fuzzy models. In *Proceedings of the IFSA World Congress, Lisboa, Portugal*. 2009.

[20] J. J. Buckley and T. Feuring. Universal approximators for fuzzy functions. *Fuzzy Sets and Systems*, 113:411–415, 2000.

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

[22] E. K. Juuso. Forecasting batch cooking results with intelligent dynamic simulation. In B. Zupančíč, R. Karba, and S. Blažíč, editors, *Proceedings of the 6th EUROSIM Congress on Modelling and Simulation, Ljubljana, Slovenia, 9-13 Sept., 2007*, volume 2, page 8 pp. University of Ljubljana, Ljubljana, Slovenia, 2007.

[23] E. K. Juuso. Hybrid models in dynamic simulation of a biological water treatment process. In J. Kunovský, P. Hanácek, F. Zboril, Al-Dabass, and A. Abraham, editors, *Proceedings First International Conference on Computational Intelligence, Modelling and Simulation, 7- 9 September 2009, Brno, Czech Republik*, pages 30–35. IEEE Computer Society, 2009.

[24] R. E. Moore. *Interval Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1966.

[25] S. Lahdelma and E. Juuso. Generalised moments and lp norms in vibration analysis. In *The Sixth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 23-25 June, 2009, Dublin, Ireland*, pages 607–619. BINDT, 2009.

[26] S. Lahdelma and E. Juuso. signal processing in vibration analysis. In *The Sixth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 15-18 July, 2008, Dublin, Ireland*, pages 867–878. BINDT, 2008.

[27] S. Lahdelma and E. Juuso. Generalised lp norms in vibration analysis of process equipments. In *The 7th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies - CM 2010 - MFPT 2010, 22-24 June, 2010, Dublin, Ireland*, page 13 pp. BINDT, 2010.

[28] E. Juuso and S. Lahdelma. Intelligent scaling of features in fault diagnosis. In *Proceedings The CM 2010 - MFPT 2010, Dublin, Ireland*, page 15 pp. BINDT, 2010.

[29] H.-J. Zimmermann. *Fuzzy set theory and its applications*. Kluwer Academic Publishers, 1992.

[30] J. Keskitalo, A. Sorsa, T. Heikkinen, and E. Juuso. Predicting cod concentration of activated sludge plant effluent using neural networks and genetic algorithms. In I. Troch and F. Breitenecker, editors, *6th Vienna Conference on Mathematical Modelling - MATHMOD 2009, Argesim Report no. 35*, pages 1014–1020. Argesim, 2009.

[31] M. Heikkinen, H. Poutiainen, M. Liukkonen, T. Heikkinen, and Y. Hiltunen. Som-based subtraction analysis to process data of an activated sludge treatment plant. In I. Troch and F. Breitenecker, editors, *MATHMOD 2009, Argesim Report no. 35*, pages 1021–1026. Argesim, 2009.

[32] S. Halttunen. Model based on-line system in activated sludge process treating pulp and paper wastewaters. In I. Troch and F. Breitenecker, editors, *MATHMOD 2009, Argesim Report no. 35*, pages 1027–1032. Argesim, 2009.