Dynamic modelling of activated sludge process IV. Multivariable statistical model

[a]J. FARKAŠ*, [b]M. KRÁLIK, [a]J. DERCO, and [a]P. FARKAŠOVÁ

[a]*Department of Environmental Chemistry and Technology, Faculty of Chemical Technology, Slovak Technical University, CS-812 37 Bratislava*

[b]*Department of Organic Technology, Faculty of Chemical Technology, Slovak Technical University, CS-812 37 Bratislava*

Received 6 July 1990

Industrial wastewater treatment process was described using the statistical multi-input, multi-output, and multi-internal variable model. Dynamics of the process was described in a discrete form. Confidence of parameters was evaluated using the Student's criterion. The adequacy of the model was determined according to the Fisher's criterion.

Results of the identification of a real plant for the Carrousel activated sludge process are shown in this paper. It was found that the model coefficients have some physical meaning (flow type corresponds to the ideal mixing, hydraulic retention time).

The results show that the statistical modelling of dynamic changes in activated sludge process can be used for the process control when, for different reasons, another type of model, *i.e.* the structural one, cannot be used.

Activated sludge process of wastewater treatment consists of physical, chemical, and biochemical reactions around the reactor. Overall process description is difficult and complicated [1].

The problem of description lies mainly in biodegradability of organic compounds depending on conditions in the reactor. *Vančo* [2] describes three groups of biodegradability-limiting factors:

— physicochemical (temperature, solubility of organic compounds in the substrate, pH, concentration of dissolved oxygen in the process),

— biological (adaptation of the culture of microorganisms to substrate and sludge age, substrate toxicity, influence of other substrates),

— chemical (molecular size of the compounds in the substrate, their chain length and substituents).

Structural models consider all these factors to be constant in the whole

* *Present address*: Slovak Commission for Environment, 812 35 Bratislava.

system or in individual subsystems. It is assumed that nitrification, denitrification, and carbonization occur in "abstract" separate reactors. When structural models or modelling of Carrousel activated sludge process are used the following problems may appear:

— carbonization, nitrification, and denitrification run simultaneously and depend on the influent concentration and composition [3],

— industrial wastewaters are mixture of compounds with different degradability [4],

for industrial wastewaters treatment systems with high sludge age are often used. In such systems only a small part of biomass [5] is viable-active.

When the above-mentioned problems are solved in structural models, long computational time is necessary and models are quite difficult. Input variables requirements are higher, too. For practical usage a lot of simplifications are made, which lead to less accurate, but sufficient models used for given purposes. One of the simplest models is based on neglecting physicochemical properties and considering only the input—output dependence. This approach leads to the so-called statistical model. The authors of reports [6—9] recommend to use statistical models for the purpose of wastewater treatment plant control.

In our previous work [9] we discussed the statistical model with one input and one output. One of our conclusions was that there are very strong bindings among variables (COD, NO_2^- NO_3^- NH_4^+ concentration, *etc.*) and one output—one input model is not able to describe them. Therefore we have developed a multivariable statistical model, in which internal bindings of variables are considered.

Model description

The considered system is illustrated in Fig. 1. It is the multi-input and multi-output system (u_i inputs, y_k outputs, x_j internal variables). Internal variables are divided into two groups: *1.* without direct relationship (binding) to output x_j and *2.* with the same value as that of output variables x_{y_k}.

From the viewpoint of the dimension of the system (number of internal variables) this division can lead to its enlargement and to complication of the system description. This complication may appear in models, which take into account the relationship among all inputs, internal variables, and outputs. When the model is described as "black-box", this complication does not appear. In turn, all outputs could be included into model description as internal variables, which leads to the improvement of description. Physicochemical interpretation of this process is easy in systems with lumped parameters, especially in completely mixed systems. All internal variables are also output quantities in

such a system. Real biochemical systems are always systems with distributed parameters, *i.e.* the internal variable value is different in individual points of the system. In high recycled systems ($R > 100$) this difference is not significant. For example, in the Carrousel activated sludge process, the concentration of all components, except oxygen, is approximately the same around the bioreactor.

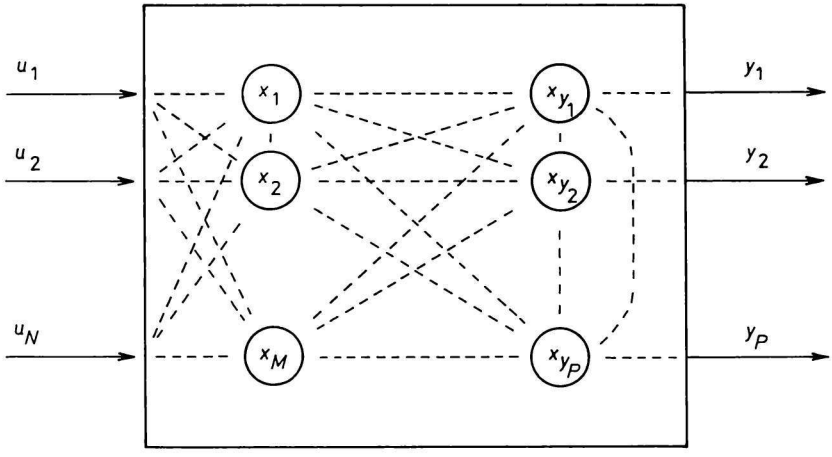

Fig. 1. Modelled system.

According to this, a discrete statistical dynamic model of activated sludge process may be expressed in the form of eqn (*1*).

$$y_{rt} = a_{r0} + \sum_{k=1}^{P} \sum_{l=1}^{P_{rk}} a_{rkl} y_{k,\,t-l\Delta t} + \sum_{i=1}^{N} \sum_{l=0}^{n_{ri}} b_{ril} u_{i,\,t-l\Delta t} \qquad (1)$$

In eqn (*1*) the influence of some inputs can be negligible and after that the relation may be simplified.

Model parameters evaluation

The least-square criterion was used for evaluation of statistical model coefficients

$$\omega_r = \sum_{h=1}^{N_c} (y_{rh}(\text{calc}) - y_{rh}(\text{exp}))^2 \qquad (2)$$

Substituting y_{rh}(calc) in eqn (*1*) and deriving *vs.* coefficient one can obtain system of linear algebraic equations

$$\mathbf{D}^{\mathrm{T}}\mathbf{D}\boldsymbol{C} = \mathbf{D}^{\mathrm{T}}\boldsymbol{Y} \qquad (3)$$

Table 1

Intervals of the variables used

Variables	Inputs		Outputs	
	min	max	min	max
$c(\mathrm{COD})/(\mathrm{mg\,dm^{-3}})$	1000	5000	380	1100
$c(\mathrm{NH_4^+})/(\mathrm{mg\,dm^{-3}})$	35	385	4.4	111
$\mathrm{MLVSS}/(\mathrm{g\,dm^{-3}})$	2.0	4.8		

Table 2

Parameters of model No. 1

Variable	Parameter (eqn (*1*))	Full model	Reduced model
	a_{10}	325.5	328.0
Conc. COD	a_{111}	0.2610	0.3282
	a_{112}	0.1600	
	a_{113}	0.0883	
	a_{114}	0.3415	0.3472
	a_{115}	−0.2584	−0.2782
	a_{116}	−0.1345	
	a_{117}	0.0054	
	a_{118}	−0.0039	
Conc. COD	b_{110}	0.0491	0.0468
	b_{111}	−0.0306	−0.0324
	b_{112}	−0.0114	
	b_{113}	−0.0063	
	b_{114}	0.0040	
	b_{115}	−0.0134	
Conc. $\mathrm{NH_4^+}$	b_{120}	0.5774	0.1975
	b_{121}	−0.5508	
	b_{122}	0.1880	
	b_{123}	0.4617	
	b_{124}	−0.0313	
	b_{125}	−0.2331	
MLVSS	b_{130}	10.97	
	b_{131}	−24.93	
	b_{132}	33.02	
	b_{133}	65.39	28.40
	b_{134}	89.20	70.12
	b_{135}	34.66	
	b_{136}	−92.58	−77.99
	b_{137}	−24.74	

Elements of $\boldsymbol{C}$ vector were computed using Cholesky's method, with the next iterative improvement [10]. The confidence of parameters was evaluated using the correlation matrix. Diagonal elements of this matrix are equal to coefficient dispersion (estimation error). Nondiagonal elements represent correlation among parameters. Student's criterion t_f for the confidence of all parameters was evaluated using diagonal elements of correlation matrix. Coefficients, for which

$$t_f > t_a(n) \tag{4}$$

were considered as significant, the other ones were nonsignificant. After the elimination of nonconfident coefficients the new computation has been made for more accurate estimation of the coefficient value. The adequacy of the model was determined according to the Fisher's criterion

$$F_r = \frac{s^2_{\mathrm{ad},\,r}}{s^2_{\mathrm{Y},\,r}} < F_{\mathrm{T}} \tag{5}$$

Since the critical value of the Fisher's criterion is always greater than 1, the relation

$$s^2_{\mathrm{ad},\,r} \leq s^2_{\mathrm{Y},\,r} \tag{6}$$

was used as sufficient adequacy condition [11]. The method described herein is realized in the program STAMOD written in Fortran 77 [12] based on [13, 14].

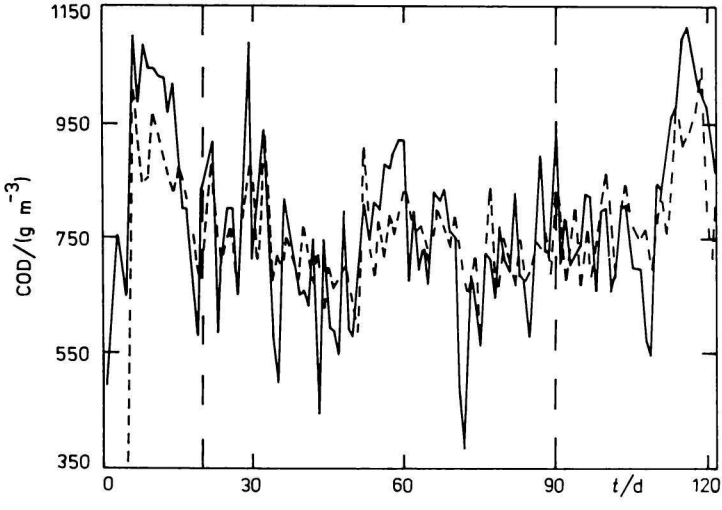

Fig. 2. Illustration of measured and computed values — model No. 1. —— Experimental, ———— reduced model.

Real plant data processing

Described identification method was verified by the data obtained in a real plant for the Carrousel activated sludge process. Composition and biodegradability of wastewaters (Chemko Strážske) is described in [2]. COD is formed by cyclohexanol, cyclohexanone, formaldehyde, pentaerythritol and hexamethylenetetramine. Wastewater inflow rate was about $280\,m^3h^{-1}$ and concentrations are given in Table 1.

Models which give the relationship among output COD and NH_4^+ concentration and inputs were searched for. Data from three months (92 d) were used — one value for one day for model verification.

Table 3

Parameters of model No. 2

Variable	Parameter (eqn (*1*))	Full model	Reduced model
	a_{20}	155.0	180.6
Conc. COD	a_{211}	0.6479	0.6862
	a_{212}	0.2792	0.2026
	a_{213}	−0.0581	
	a_{214}	0.3338	0.2809
	a_{215}	−0.4056	−0.3361
	a_{216}	−0.0716	
	a_{217}	0.1151	
	a_{218}	−0.1273	
Conc. COD	b_{210}	0.0565	0.0473
	b_{211}	−0.0483	−0.0407
	b_{212}	−0.0102	
	b_{213}	0.0048	
	b_{214}	0.0122	
	b_{215}	−0.0001	
Conc. NH_4^+	b_{220}	0.5505	−0.0056
	b_{221}	−0.7167	
	b_{222}	0.2372	
	b_{223}	0.5758	
	b_{224}	−0.2450	
	b_{225}	−0.5502	−0.4522
MLVSS	b_{230}	11.20	
	b_{231}	−12.94	
	b_{232}	31.23	
	b_{233}	59.83	
	b_{234}	59.47	
	b_{235}	−73.61	56.96
	b_{236}	−95.22	−68.59
	b_{237}	31.72	

Model No. 1 represents influence of COD, NH_4^+, and biomass concentrations on the output COD concentration. Nonfiltrate measured values were used for computation. Computed model parameters are in Table 2. Measured and computed profiles of values are shown in Fig. 2. Two vertical lines deliminate which measured values were used for the evaluation of parameters. Model No. 2 is analogous to model No. 1, but for computation filtrate values were used. Results are shown in Table 3 and Fig. 3.

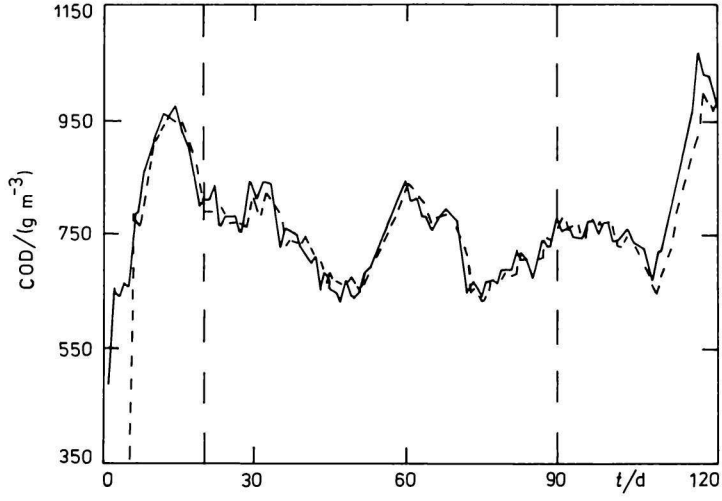

Fig. 3. Illustration of measured and computed values — model No. 2. Denotation as in Fig. 2.

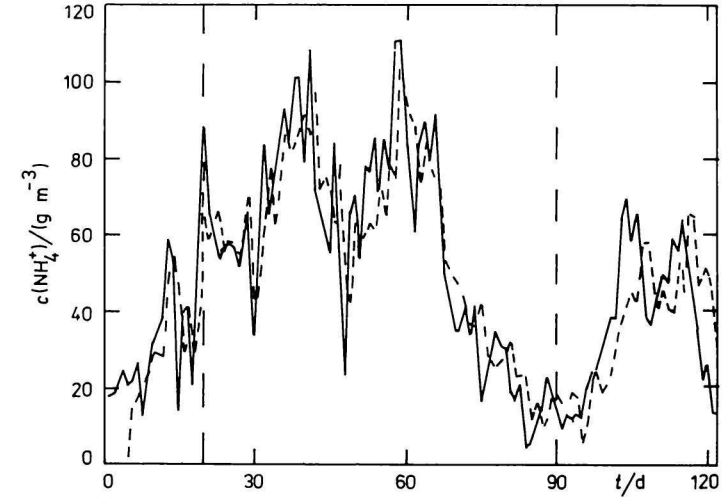

Fig. 4. Illustration of measured and computed values — model No. 3. Denotation as in Fig. 2.

Model No. 3 investigates influence of COD, NH_4^+, and biomass concentrations on the output NH_4^+ concentration. Measured nonfiltrate values were used. Computed parameters are in Table 4. Experimental and computed values are illustrated in Fig. 4. Model No. 4 is analogous to model No. 3, but for computation filtrate values were used. Results are presented in Table 5 and Fig. 5.

Figs. 4 and 5 provide a very good description of the process. Comparison of the models on the basis of residual sum of squares and adequacy criterion is given in Table 6. When the model with COD concentration output was used, the measured data had to be filtered. Comparison of full and filtrated model is given

Table 4

Parameters of model No. 3

Variable	Parameter (eqn (*1*))	Full model	Reduced model
	a_{30}	−21.84	
Conc. NH_4^+	a_{311}	0.5015	0.6038
	a_{312}	0.1672	
	a_{313}	0.2990	0.2920
	a_{314}	−0.1624	
	a_{315}	0.2631	0.0594
	a_{316}	−0.1309	
	a_{317}	0.0730	
	a_{318}	−0.1578	
Conc. COD	b_{310}	0.0023	
	b_{311}	−0.0034	−0.0031
	b_{312}	0.0004	
	b_{313}	0.0031	0.0007
	b_{314}	0.0052	0.0029
	b_{315}	−0.0001	
Conc. NH_4^+	b_{320}	0.0172	
	b_{321}	0.0359	
	b_{322}	0.0586	
	b_{323}	−0.1008	−0.0544
	b_{324}	0.0174	
	b_{325}	−0.0422	
MLVSS	b_{330}	−9.97	−6.06
	b_{331}	6.04	
	b_{332}	11.99	13.57
	b_{333}	−8.12	−5.57
	b_{334}	−2.42	
	b_{335}	6.06	
	b_{336}	1.49	
	b_{337}	−1.62	

in Table 6. The model with NH_4^+ concentration output was adequate enough also without data filtration. From this it follows that the COD concentration values are probably measured with less accuracy than the NH_4^+ concentration values, *i.e.* these values are the reason for the largest disturbances in the system.

When filtrate data are used for computation, the significance values are changed in such a way that they confirm the hypothesis that the Carrousel process should be described as one completely mixed bioreactor (for COD and NH_4^+ concentrations). The system has a "weak" memory. The most important parameter is the one from the first time period. This rule breaks down only

Table 5

Parameters of model No. 4

Variable	Parameter (eqn (*1*))	Full model	Reduced model
	a_{40}	−10.91	−3.78
Conc. NH_4^+	a_{411}	0.9337	1.0917
	a_{412}	0.1337	
	a_{413}	0.1409	
	a_{414}	−0.3306	−0.1092
	a_{415}	0.2063	
	a_{416}	−0.1603	
	a_{417}	0.0391	
	a_{418}	−0.0442	
Conc. COD	b_{410}	0.0014	
	b_{411}	−0.0034	
	b_{412}	0.0016	
	b_{413}	−0.0042	−0.0018
	b_{414}	0.0043	0.0009
	b_{415}	−0.0031	
Conc. NH_4^+	b_{420}	0.0356	
	b_{421}	0.0129	
	b_{422}	0.0376	
	b_{423}	−0.1154	−0.0498
	b_{424}	0.0298	
	b_{425}	0.0134	
MLVSS	b_{430}	−9.98	−6.39
	b_{431}	6.45	
	b_{432}	13.16	16.55
	b_{433}	−14.83	−9.66
	b_{434}	−1.85	
	b_{435}	7.70	
	b_{436}	1.01	
	b_{437}	−0.77	

in case of biomass concentration, which may be explained as a consequence of 4—6 d hydraulic retention time. Time of computation is short (approximately 100 s).

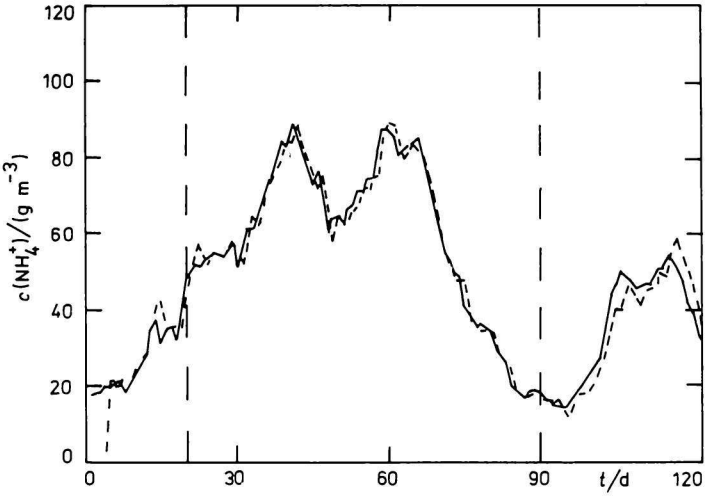

Fig. 5. Illustration of measured and computed values — model No. 4. Denotation as in Fig. 2.

Table 6

Model comparison

Model number	Full model		Reduced model	
r	ω_r	F_r	ω_r	F_r
1	0.6522×10^6	0.9715	0.7703×10^6	0.7841
2	0.3534×10^5	0.2386	0.4699×10^5	0.2205
3	0.1061×10^5	0.3309	0.1284×10^5	0.2736
4	0.6331×10^3	0.0353	0.7825×10^3	0.0294

Conclusion

From the measured values obtained at a real wastewater treatment plant it follows that there exists statistically significant dependence between inlet and outlet of wastewater. A simple computer program for fast statistical model parameters evaluation with consequent wastewater treatment plant working prediction and analysis of some dependences between input and output variab-

les was developed in our laboratory. Validity and usefulness of this model were confirmed using industrial wastewater treatment plant data. We have found that the model coefficients have also some physical meaning in the qualitative sense. For example, ideal mixing in system was illustrated by the significance of the first coefficient and the sludge age also determines position of significant coefficients.

Statistical modelling of dynamic changes in activated sludge process may be used, when sets of data from the treatment plant work are collected, and when, for different reasons, another model, *i.e.* the structural one, cannot be used.

Developed model provides easy and fast results, and therefore it may be used for the purpose of wastewater treatment plant control, especially in nontypical cases.

Symbols

a_{r0}	absolute coefficient of the r-th model
a_{rkl}	l-th coefficient of the k-th output in the r-th model
b_{ril}	l-th coefficient of the i-th input in the r-th model
$\boldsymbol{C}$	vector of coefficients
COD	chemical oxygen demand
$\mathbf{D}$	input matrix
$\mathbf{D}^{\mathrm{T}}$	transponed input matrix
F_r	Fisher's r-th model criterion
F_{T}	tabulated Fisher's criterion
M	number of internal variables
MLVSS	mixed liquid volatile suspended solids
N	number of inputs
N_{c}	number of measurements for evaluation, *i.e.* number of matrix $\mathbf{D}$ rows
n_{ri}	number of coefficients of the i-th input in the r-th model
P	number of outputs = number of models
P_{rk}	number of coefficients of the k-th output in the r-th model
R	recycle ratio
$s^2_{\mathrm{ad},\,r}$	dispersion of the model adequacy
$s^2_{\mathrm{Y},\,r}$	dispersion of the output parameters
t_f	calculated Student's criterion for all parameters
$t_a(n)$	tabulated Student's criterion
u_i	i-th input
$u_{i,\,t\,-\,l\Delta t}$	i-th input value in time $t - l\Delta t$
x_j	internal variable without direct binding to output
x_{y_k}	internal variable with direct binding to output
$\boldsymbol{Y}$	output vector

y_k	k-th output
y_{rh}(exp)	h-th measured output value used for the enumeration of ω_r
y_{rh}(calc)	h-th calculated value
y_{rt}	observed output value of the r-th model in time t
$y_{k,\,t-l\Delta t}$	output value in time $t - l\Delta t$
ω_r	squared deviation sum of the r-th model

References

1. Tchobanoglous, G., *Wastewater Engineering (Treatment, Disposal, Reuse)*. McGraw-Hill, New York, 1979.
2. Vančo, D., unpublished results.
3. Argaman, Y. and Miller, E., *J. Water Pollut. Contr. Fed. 51,* 749 (1979).
4. Deshpande, S. D., Chakrabarti, T., and Subrahmanyan, P. V. R., *Environ. Sci. Technol. 21,* 1003 (1987).
5. Green, M. and Shelef, G., *Water Res. 15,* 953 (1981).
6. Berthouex, P. M. *et. al., Water Res. 10,* 689 (1976).
7. Debelak, K. A. and Sims, C. A., *Water Res. 15,* 1173 (1981).
8. Hansen, J. L., Fiok, A. E., and Hovious, J. C., *J. Water Pollut. Contr. Fed. 52,* 1966 (1980).
9. Králik, M., Derco, J., Farkaš, J., and Gubová, P., *Chem. Papers 45,* 313 (1991).
10. Olehla, M., Věchet, V., and Olehla, J., *Řešení úloh matematické statistiky ve Fortranu.* (Solution of Problems of Mathematical Statistics in Fortran.) Nakladateľstvo dopravy a spojov (Publishing House of Transport and Communications), Prague, 1982.
11. Kovaleva, N. G. and Kovalev, B. G., *Biokhimicheskaya ochistka stochnykh vod predpriyatii khimicheskoi promyshlennosti.* (Biochemical Wastewater Treatment in Plants of Chemical Industry.) Khimiya, Moscow, 1987.
12. Farkaš, J., *Diploma Thesis.* Faculty of Chemical Technology, Slovak Technical University, Bratislava, 1988.
13. Kafarov, V V., *Metody kibernetiki v khimii i khimicheskoi tekhnologii.* (Methods of Cybernetics in Chemistry and Chemical Technology.) Khimiya, Moscow, 1976.
14. Rektorys, K. *et. al., Přehled užité matematiky.* (Survey of Applied Mathematics.) Nakladatelství technické literatury (Publishers of Technical Literature), Prague, 1981.

Translated by P Farkašová