Efficient Operation of Urban Wastewater Treatment Plants
(EOLI)

WP2 – Model Selection and Parameter Identification
EM1 Associated to Inhibitory Toxicants Removal Processes

EOLI MODEL 1 candidates BC3, C3 and D3
LabScale Reactor 1, UNAM (Partner 6)

CONTENTS

By Manuel J. BETANCUR, Engineering Institute, UNAM, Jun 200, Updated Oct 2004 for notation.

Contract Number ICA4-CT-2002-10012

INTRODUCTION

This work reports the identification of EOLI Model one (EM1) using a new set of experiments (ID3, ID4, VAL2) and a new procedure (C3). This C3 proposal adds two refinements to the model. It includes the sensor filtering effect on the oxygen signal and includes also a modification to the Specific Biomass Growth Rate curve for very low substrate concentrations. A procedure D3 without these additions is also performed to those same experiments for comparison purposes. The user then may now choose, for example, between the simpler D3 type or, if more stringent fitting needs arise, the C3 type instead.

To test the hypothesis that different biomasses may have significantly different kinetic models, a mixture of all ID experiments available was made, and a unified EM1-BC3 was generated. It is a mixture of procedures B and C and is the best single representation for all experiments in set (ID1, ID2, ID3, ID4, VAL1 and VAL2).

In total 3 new EM1 candidates BC3, C3 and D3 are now added to the repertory of models. For completeness, all annexes and data tables from former reports are also included and, when appropriated, updated with the new procedures results.

EXPERIMENT SELECTION

After careful examination UNAM's WP1 team recommended a validation VAL2 experiment for set (ID3, ID4). These experiments belong to the "Starvation" mode. However all comply with standard specifications. The biomass used in all of them is the same, but different to that used in (ID1, ID2, VAL1) for former procedures A and B.

Annex 1 shows the full names and classification of all experiments.

MODEL STRUCTURE

Equation 1 represents the basic structure of the SBR mathematical model, for both the fill-reaction and full-reaction phases in all model candidates. Equation 2 complements it by representing the effect of D.O. Sensor on oxygen signal for candidate "C". Such effect is approximated as a second order filter. Definitions and units are found in Annex 4.

$$\dot{X} = \mu X - X\frac{q_{in}}{V}$$

$$\dot{S}_C = -k_1 \mu X + (S_{C,in} - S_C)\frac{q_{in}}{V} \qquad \text{Eq.(1)}$$

$$\dot{S}_O = -k_2 \mu X - bX + K_L a * (S_{Os} - S_O) + (S_{O,in} - S_O)\frac{q_{in}}{V}$$

$$\dot{V} = q_{in}$$

$$\dot{S}_{Oa} = (S_O - S_{Oa}) / \tau_a$$

$$\dot{S}_{Osen} = (S_{Osen} - S_{Oa}) / \tau_b \qquad \text{Eq.(2)}$$

Equation 3 represents the *Haldane* graph for the Specific Biomass Growth Rate as function of *Substrate* and *Dissolved Oxygen*. Equation 4 represents the Oxygen limitation dependence and Equation 5 represents the proposed modification for the inhibitory substrate dependence (Betancur et al, 2002; Moreno et al 2003). Figure 1 shows a graph for Equation 5 displaying the threshold piece-wise discontinuity proposed.

$$\mu = \mu_S \mu_{DO} \qquad \text{Eq.(3)}$$

$$\mu_{DO} = \frac{O}{O + K_o} \qquad \text{Eq.(4)}$$

$$\mu = \begin{cases} \dfrac{\mu_0 S_C}{K_S + S_C + S_C^2 / K_I} & for \quad S_C \geq S_t \\ \mu_t S_C / S_t & for \quad S_C \prec S_t \end{cases} \qquad \text{Eq.(5)}$$

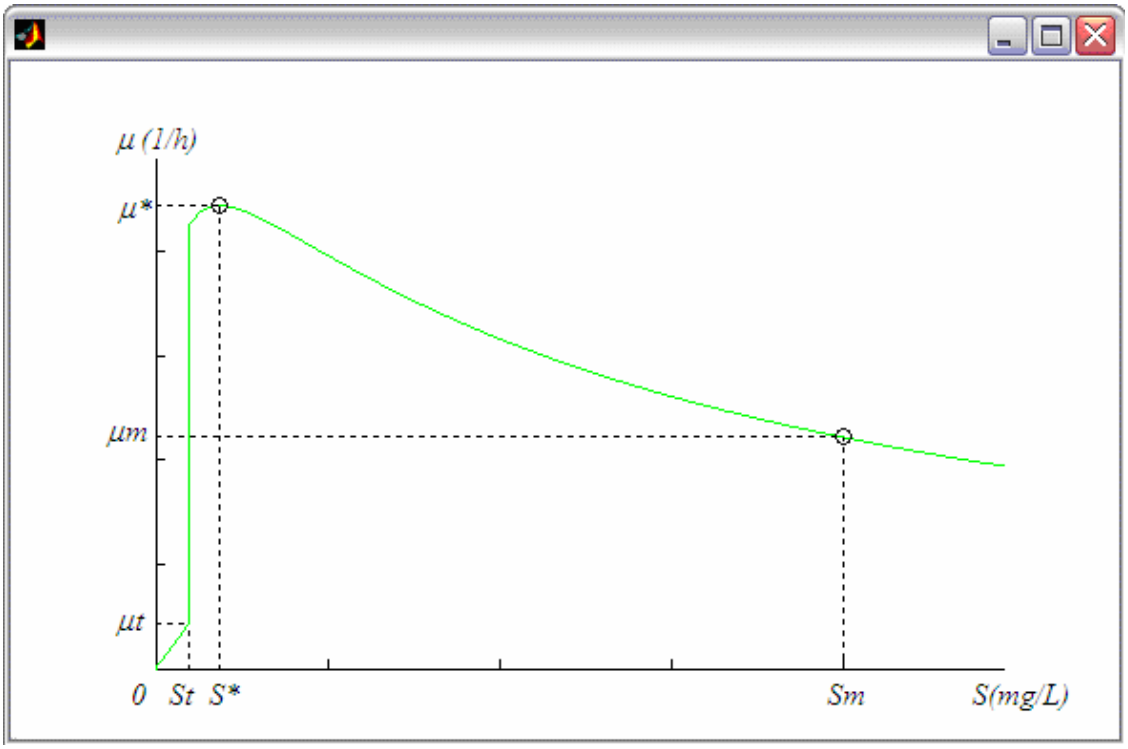

Figure 1. Graph of Eq. 5 representing μ_s for procedure C proposal.
By definition $\mu_m=\mu^*/2$ and $\mu_t=\mu^*/10$
See annex 5 for equivalences with K_I, K_S and μ_o
For procedures A, B and D make $S_t=0$ to obtain the Haldane standard graph.

IDENTIFICATION PROCEDURES

Each EM1 candidate is made by combining two procedures, one for the Yield coefficients (tagged with numbers) and other mainly for the kinetic ones (tagged with letters). This work is concerned with procedures BC3, C3 and D3. However, for completeness, all procedures options are described in Table 1 (for Yield parameters) and Table 2 (for Kinetic parameters).

Table 1. Yield coefficients Estimation/Identifying procedures.

Version	Procedure (Yield coefficients)
1	K_1=3,7 mg4CP/mgVSS is estimated from data coming from a series of batches where total substrate consumption and total biomass generation was measured. See Annex 3.1. K_2 is calculated using the identified relation k_2 / k_1 using the separation technique.
2	K_1 and k_2 are calculated from stechiometric relation (Gaudy et al, 1988) and the identified relation k_2 / k_1. See Annex 3.2.
3	K_1=3,7 mg4CP/mgVSS is estimated from data coming from a series of batches where total substrate consumption and total biomass generation was measured. See Annex 3.1. k_2 is identified using the same program as for kinetic parameters (no separation).

Table 2. Kinetic coefficients identifying procedures.

Version	Procedure (Kinetic coefficients)
A	FTC experiments ID1 and ID2 are used. See Annex 1 for experiment classification. Influent's Oxygen assumed negligible Experiment VAL1 is used for validation
B	FTC experiments ID1 and ID2 are used. See Annex 1 for experiment classification. Influent's Oxygen assumed unknown, but constant for each reaction. Experiment VAL1 is used for validation
BC	Is a combination of B and C i.e. it includes ID1, ID2, ID3, ID4, VAL1 and VAL2
C	FTC experiments ID3 and ID4 are used. See annex 1 for experiment classification. A mathematical model for representing the D.O. sensor is added. A piece-wise discontinuity to μ(S) is added. Experiment VAL2 is used for validation
D	Similar to procedure B, but applied to different experiments. FTC experiments ID3 and ID4 are used. See annex 1 for experiment classification. Experiment VAL2 is used for validation

Initial conditions and estimable parameters for procedure C and D are treated in a similar way as for procedures A and B. See Annex 7. The added filter, that emulates the sensor effect on the measured Oxygen in procedure C, is initialized to S_{Oo} value. D.O. sensor's model identification is described in Annex 8.

A kinetic parameter *"St"* is added to modify the Haldane curve by creating a piecewise discontinuity at low substrate concentrations. This parameter should be identified also.

Fitting criteria and software description is similar to the one used for procedures A and B. See Annex 9 for explanation.

RESULTS AND VALIDATION

Table 3 shows fitting and validation results for EM1-C3 and EM-D3 candidates. For completeness and for comparison all former results are also listed. It can be seen that C3 fits the best and validates the best (i.e. displays the lowest Mean Quadratic Error (MQE)). Table 4 shows the improvement in fitting for EM1-C3 compared to EM1-D3. It is about 18% for the global evaluation, i.e. fitting infinite time horizon (ITH) simulations, using the same single EM1 model under test, for all ID and VAL experiments involved. MQE improvement may go as high as 35% for individual experiments. Table 5 shows all results for all identified parameter values, for all EM1 candidates, and their respective confidence intervals.

Table 3. Validation fitting results for EM1 model candidates using *Fit_SO* criteria

EM1 candidate	Experiment tag	MQE kid (own_parameters)	MQE test (using_EM1)	Mean EM1 test MQE	
				by group	Global
A1	ID1	758	961	804	707
	ID2	812	647		
	VAL1		514	514	
A2	ID1	742	929	788	701
	ID2	794	647		
	VAL1		527	527	
A3	ID1	444	446	392	430
	ID2	328	337		
	VAL1		506	506	
B1	ID1	369	515	390	343
	ID2	373	264		
	VAL1		251	251	
B2	ID1	355	516	383	342
	ID2	353	250		
	VAL1		259	259	
B3	ID1	220	231	186	244
	ID2	127	141		
	VAL1		360	360	
(Unifying) BC3	ID1	220	1339	1512	1525
	ID2	127	1158		
	ID3	12	1824		
	ID4	27	1727		
	VAL1		1866	1550	
	VAL2		1233		
C3	ID3	12	42	51	81
	ID4	27	60		
	VAL2		141	141	
D3	ID3	17	49	58	99
	ID4	41	66		
	VAL2		182	182	

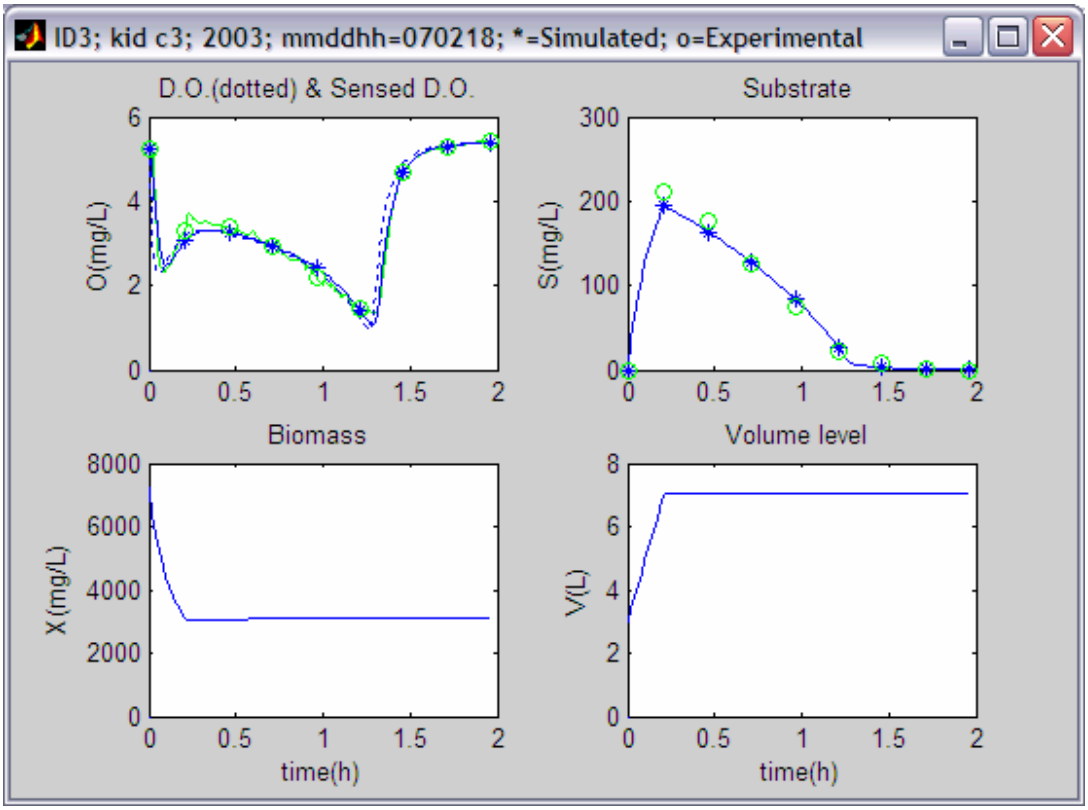

Figure 2. ITH simulation result compared to experimental data for C3 identification of ID3experiment.

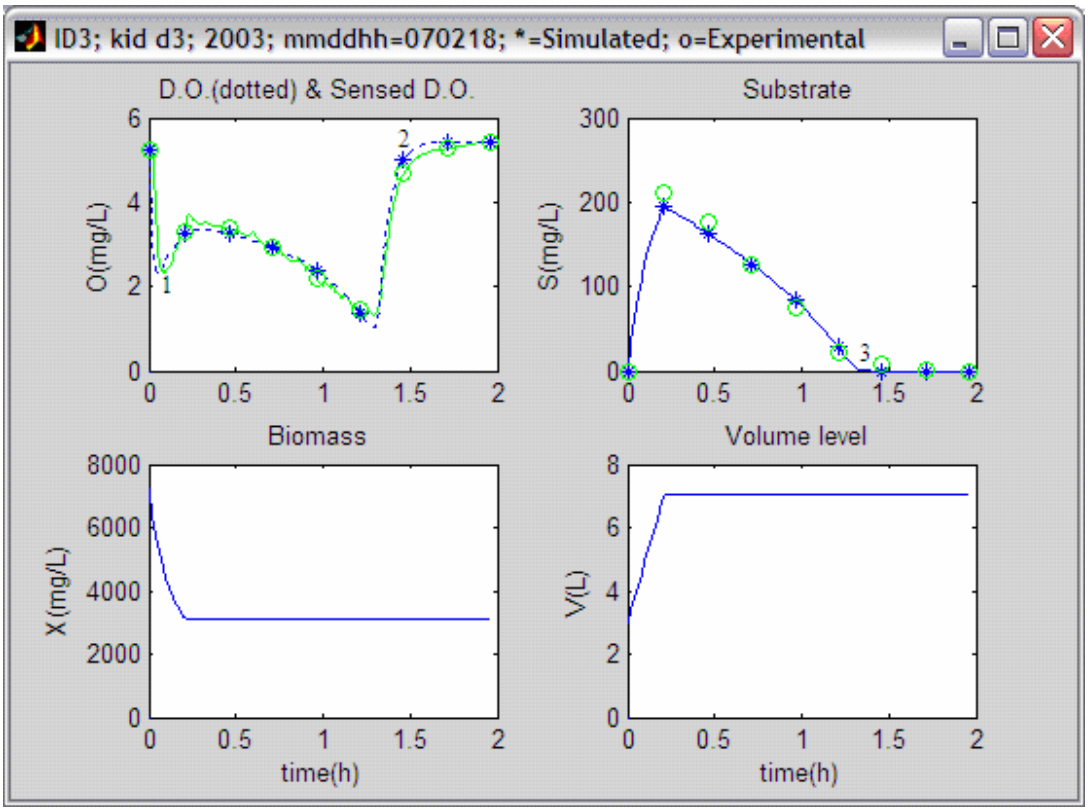

Figure 3. ITH simulation result compared to experimental data for D3 identification of ID3 experiment.

Table 4. Fitting Improvement for EM1-C3 compared to EM1-D3

Fit to	"C3" over "D3" MQE Improvement
kid ID3	31%
kid ID4	35%
test VAL2	22%
Global test	18%

Figure 2 shows the result for ID3 identification using procedure C3, and Figure 3 shows the same experiment with procedure D3. Validation tests for resulting candidates EM1-C3 and EM1-D3 are shown in Figures 4 and 5.

For Figure 2 the fitting graph is the continuous blue line (D.O. sensor simulation), while for Figure 3 the fitting graph is the dotted blue one (without D.O. sensor simulation). The green line is the experimental on-line D.O., and the marked points indicate the manual off-line samplings. There are 3 key-points to compare between Figures 2 and 3. Those will explain the improving mechanism of the EM1-C3 added features and their possible practical implications with respect to EM1-D3:

1. The initial drop of the D.O. right after t_o and into first D.O. valley. See number "1" inside subfigure D.O. of Figure 3. This happens because the system is first quickly gaining activity and then crossing the optimal operational point as S increases, so there is a big D.O. initial demand. EM1-D3 (Fig. 3) represents this without considering the sensor. Any system (software sensors, controllers, supervisory or fault detection systems) using such model will have a response delay error if it uses the real sensed D.O. instead of the "real D.O.". On the contrary, EM1-C3 does represent this filtering (Fig. 2). Hence, if EM1-C3 is used, better practical results are to be expected.

2. Entrance to the last D.O. plateau when reaction is about to finish. See number "2" inside subfigure D.O. of Figure 3. Real D.O. suggest that there is some activity going on at this point as saturation is usually not achieved as quick as EM1-D3 would predict. Instead, EM1-C3 (continuous blue line) in Fig. 2 represents this phenomena by means of the threshold S_t in $\mu(S)$

3. Substrate depletion point. See number "3" inside subfigure Substrate of Figure 3. This is directly related to point 2. Real kinetics display slowly disappearing remaining substrate after the main reaction. By considering it less error in predicting the reaction's finishing time is expected, if finalization criterion happens to be below S_t.

Links to all related documents, graphs, datasheets etc. are provided in Annex 10.

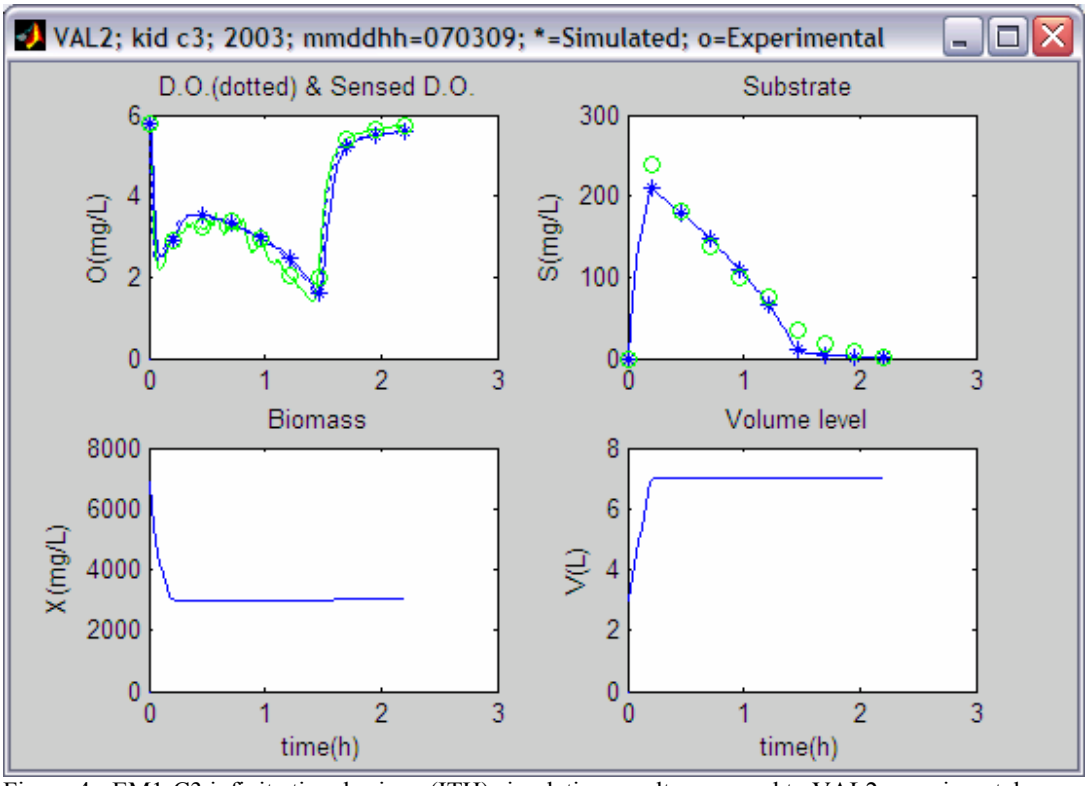

Figure 4. EM1-C3 infinite time horizon (ITH) simulation result compared to VAL2 experimental data for validation.

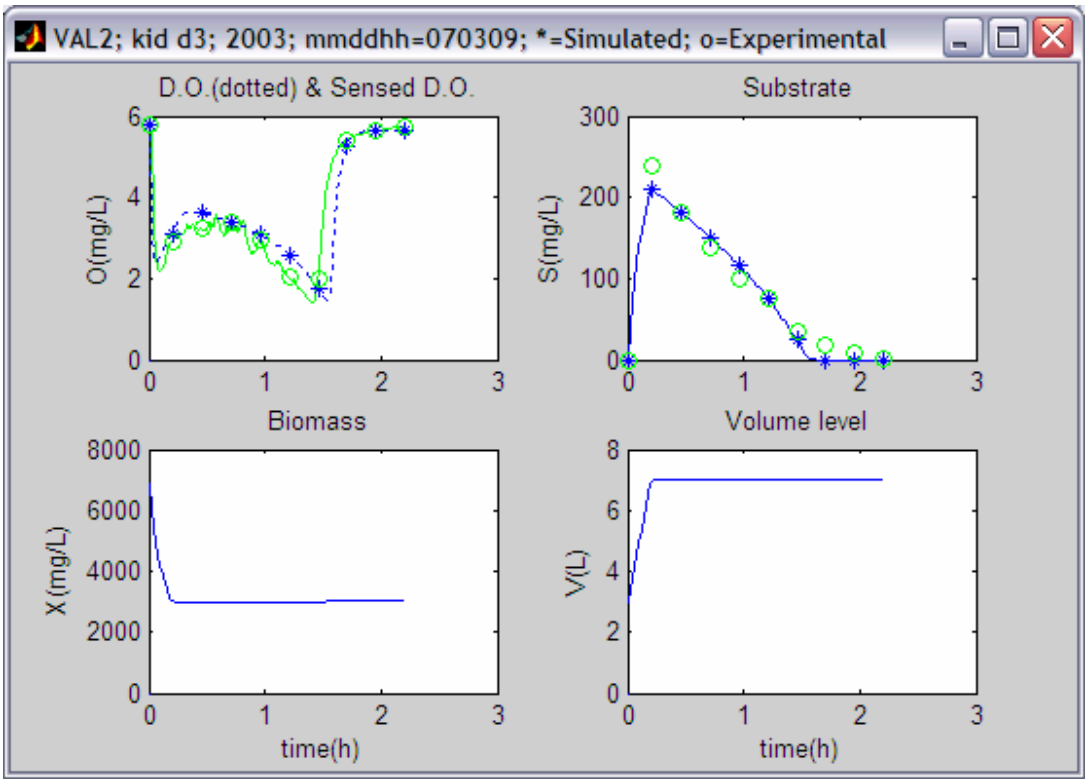

Figure 5. EM1-D3 ITH simulation result compared to VAL2 experimental data for validation.

Table 5. Identification results. Parameter values for EM1 candidates and their confidence intervals

EM1 candidate results		b	$K_L a$	k_2/k_1	k_1	k_2	μ_o=mu0[1]	K_I	K_S	Ko	μ^*=mue[2]	S*=Se	Sm	St
EM1-A1	Confidence Interval 95%	0.0004	0.1	0.0714		0.2643	0.1075	5.4292	41.2296	0.0631	0.0009	1.9653	1.5915	
	Confidence/Mean ratio	7%	1%	16%		16%	77%	84%	106%	49%	4%	14%	3%	
	Mean Value	0.0059	16.8	0.4585	3.7	1.6963	0.13915	6.43745	38.8812	0.1283	0.0219	13.821	58.442	0
EM1-A2	Confidence Interval 95%	0.0004	0.1	0.0714	0.1275	0.1722	0.05233	0.91687	13.426	0.0051	0.0048	0.1581	0.3867	
	Confidence/Mean ratio	7%	1%	16%	8%	24%	12%	23%	25%	5%	9%	1%	1%	
	Mean Value	0.0059	16.8	0.4585	1.593	0.7327	0.4275	3.9491	52.8808	0.1093	0.0514	14.339	57.746	0
EM1-A3	Confidence Interval 95%	0.0004	0.1			0.0578	0.00255	0.00755	0.01499	0.0038	0.0002	0.017	0.0714	
	Confidence/Mean ratio	7%	1%			6%	1%	0%	0%	4%	1%	0%	0%	
	Mean Value	0.0059	16.8		3.7	1.0363	0.1916	3.75305	60.0076	0.1021	0.0213	15.007	60.03	0
EM1-B1	Confidence Interval 95%	0.0004	0.1	0.0714		0.2643	0.01784	1.01075	8.46744	0.0919	0.0008	0.7886	1.875	
	Confidence/Mean ratio	7%	1%	16%		16%	15%	17%	28%	107%	4%	6%	3%	
	Mean Value	0.0059	16.8	0.4585	3.7	1.6963	0.1197	6.1161	30.1745	0.0862	0.022	13.497	56.901	0
EM1-B2	Confidence Interval 95%	0.0004	0.1	0.0714	0.1275	0.1722	0.04087	0.00706	0.0391	0.0208	0.0045	0.0192	0.0792	
	Confidence/Mean ratio	7%	1%	16%	8%	24%	9%	0%	0%	24%	9%	0%	0%	
	Mean Value	0.0059	16.8	0.4585	1.593	0.7327	0.46225	3.598	59.8005	0.0872	0.0505	14.668	58.6	0
EM1-B3	Confidence Interval 95%	0.0004	0.1			0.177	0.09565	2.27003	31.4181	0.1065	0.0005	1.2448	2.2331	
	Confidence/Mean ratio	7%	1%			14%	54%	55%	73%	159%	2%	10%	4%	
	Mean Value	0.0059	16.8		3.7	1.2718	0.1777	4.1116	43.1102	0.0671	0.0231	12.581	51.351	0
EM1-BC3	Confidence Interval 95%	0.0015	0.24			0.2766	0.08851	44.2531	26.349	0.0583	0.0027	1.0368	48.186	3.446
	Confidence/Mean ratio	32%	1%			26%	82%	103%	116%	79%	11%	8%	51%	114%
	Mean Value	0.0046	16.6		3.7	1.0641	0.10753	42.8947	22.772	0.0733	0.0254	13.287	93.776	3.012
EM1-C3	Confidence Interval 95%	0.0002				0.3106	0.00421	13.7396	0.76478	0.0934	0.0014	1.0373	9.9086	1.237
	Confidence/Mean ratio	6%				36%	11%	17%	31%	117%	5%	7%	7%	21%
	Mean Value	0.0033	16.4		3.7	0.8565	0.03735	81.6779	2.4339	0.0796	0.0277	13.992	136.2	6.025
EM1-D3	Confidence Interval 95%	0.0002				0.3	0.00118	5.99573	0.29703	0.0465	0.0017	0.6857	3.4015	
	Confidence/Mean ratio	6%				35%	3%	6%	17%	31%	6%	5%	2%	
	Mean Value	0.0033	16.4		3.7	0.8684	0.0343	95.3118	1.76295	0.1505	0.027	12.94	145.92	0

[1] K-parameters (*mu0, K_I, K_S*) computations are included only for reference. Their mean value is not used for EM1 candidates. Reasons are given in Annex 5.
[2] S-parameters (*mue, Se, Sm, St*) are used for each EM1 candidate final computation. To convert them to K-parameters use *StoK.m* or equations from Annex 5.

BIOMASSES COMPARISON

Figure 6 shows four different biomass kinetic models in graphical form. The parameters to generate these graphs come from Table 5. It contains EM1-B3 (red dotted line) and EM1-C3 (blue dashed line) models and also a combination of both, the EM1-BC3 model. There are 2 options for calculating such combined model. $BC3_{(S)}$ (black continuous line) uses the S-parameters to obtain a mean value for each one. $BC3_{(K)}$ does the same but uses K-parameters instead. Comparing both results it is evident that only $BC3_{(S)}$ represents a "mean combined model", so that one is the natural choice to build EM1-BC3.

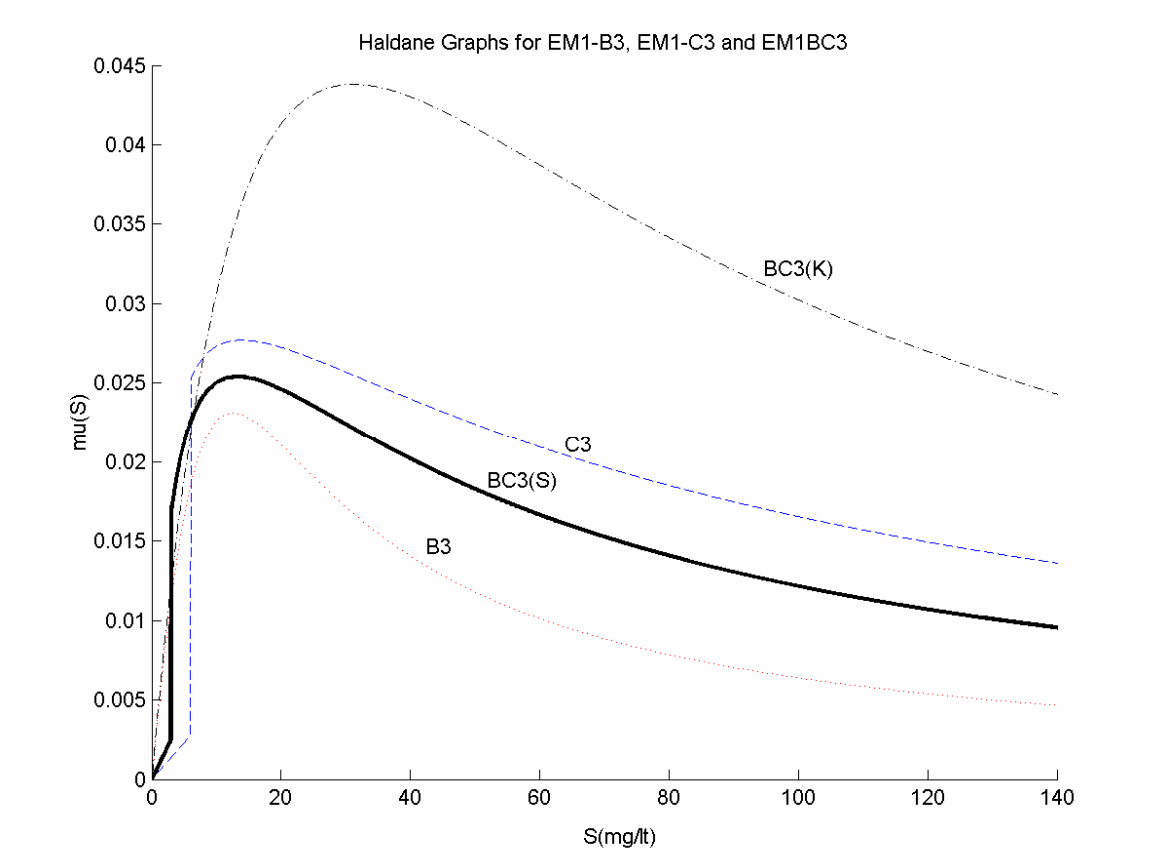

Figure 6. The red(dotted) and blue(dashed) lines are two identification results (EM1-B3 and EM1-C3). The Black-continuous one ($BC3_{(S)}$) is their mean graph obtained averaging each S-parameter (mue, Se, Sm, Sk). The Black-dotted-dashed one ($BC3_{(K)}$) is the mean graph obtained averaging each traditional K-parameter (mu0, Ki, Ks).

These two biomasses, lets call them by year and month as 0307 and 0312 respectively, are different. Trying to represent them by using a combined model will produce the results given in Figures 7 and 8 for VAL1 and VAL2 respectively. Table 3 shows their fitting results. They are the worst of all candidates. Hence it is fair to say that 0307 and 0312 biomasses require each one a different model, as initially hypothesized.

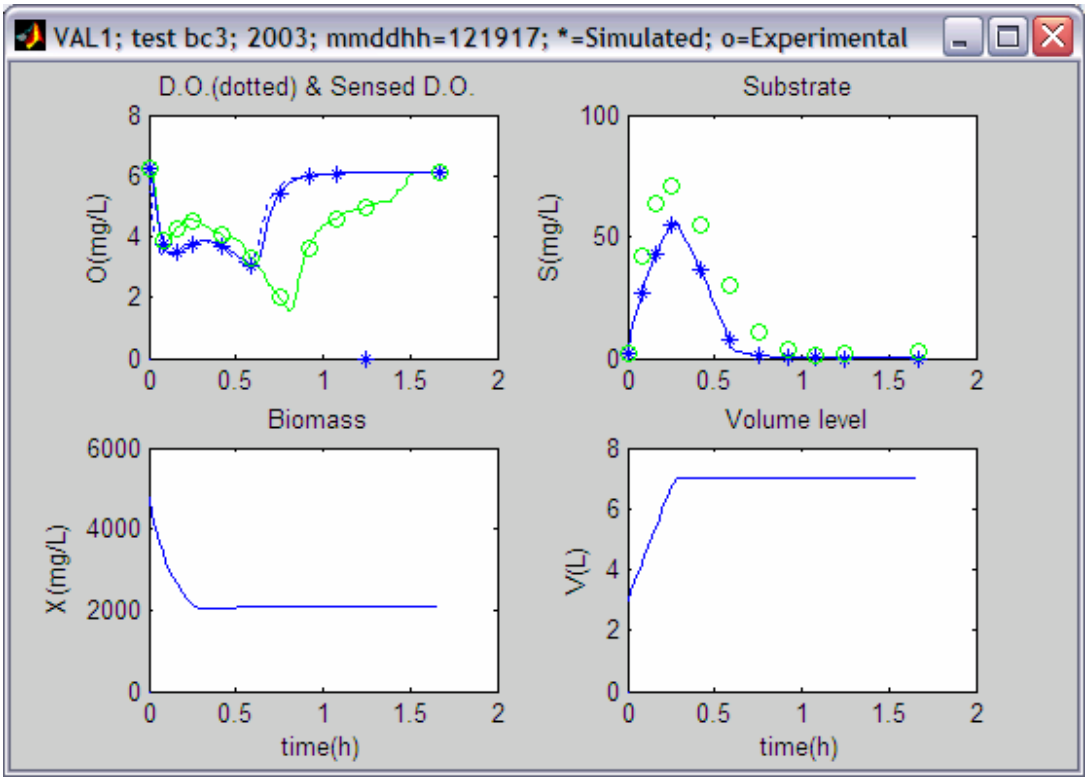

Figure 7. Unified EM1-BC3 model applied to VAL1 (biomass 0312)

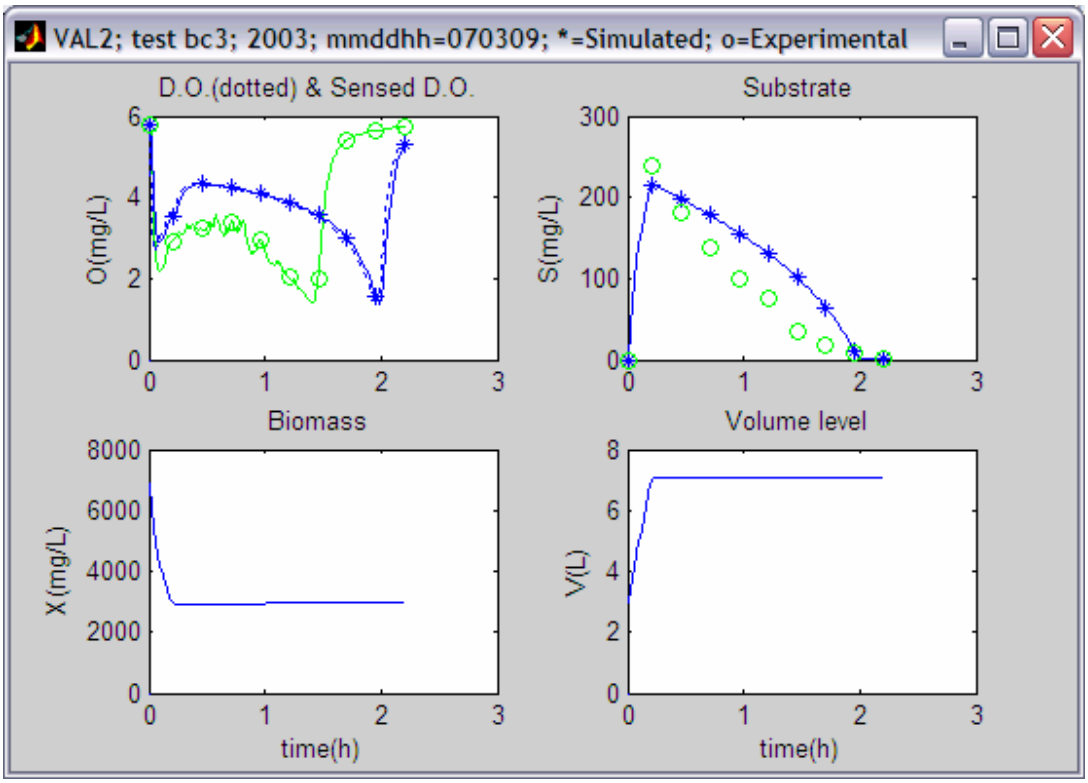

Figure 8. Unified EM1-BC3 model applied to VAL2 (biomass 0307)

Some interesting results come from this exercise however. First, if a unified model is required anyway, even accepting low accuracy, this is the one. Second, by looking at the confidence intervals for the parameters one could get a preliminary idea of the variability range of the biomass parameters. More study is necessary on the subject, i.e. more biomass batches modelized. But meanwhile any system using EM1 should be robustly designed to cope with the fact that EM1 may vary in time, i.e. when new biomass is put to the SBR.

Temperature is a factor that might affect biomass performance. In order to be sure this variable is not responsible for the differences in EM1-B3 and EM1-C3 the mean temperature was analyzed from online logs. They are shown in Table 6. Global temperature difference is less than 0.5 ºC

Table 6. Mean temperature during t=(0, 2)h

Biomass	Experiment	Mean Tmp (ºC)	Global Mean
0312	(ID1) 2003-12-19_09	19.73	20.23
	(ID2) 2003-12-19_13	19.96	
	(VAL1) 2003-12-19_17	20.99	
0307	(ID3) 2003-07-02_18	20.90	20.71
	(ID4) 2003-07-04_14	20.90	
	(VAL2) 2003-07-03_09	20.34	

CONCLUSIONS

A refinement EM1-C3 of EM1 former options was proposed. Model fitting evaluation improved 18% for experimental set (ID3, ID4 and VAL2). Fitting criterion is the same as before, the sum of a weighed combination of D.O. quadratic errors and Substrate quadratic errors.

EM1-C3 refinements include D.O. sensor modeling and also the addition of a threshold discontinuity to Haldane inhibitory model. Modeling the D.O. sensor allows EM1-C3 to better represent the D.O. signal when sudden and big changes occur in it. In Fixed Time Control (FTC) mode this is the case when substrate feeding starts, thus quickly taking the system to its optimum operational point and then over passing it. On the other hand, inclusion of a threshold, in the form of a piecewise discontinuity, to former Haldane model allows EM1-C3 to represent end of reactions more accurately.

Two biomass batches have been used in experiments for EM1 identification. Biomass 0312 was used for experiments set (ID1, ID2 and VAL2), while biomass 0307 was used for experiments set (ID3, ID4 and VAL2). Models for each set are different because biomasses behave differently even if they come from the same source. Nevertheless a unifying EM1-BC3 model was found. Results suggest that a unified model is not viable and that all designs based on EM1 should robustly cope with model variations when biomasses changes occur.

REFERENCES

Betancur, M., Moreno, J., Buitrón, G. (2002) "Refined Model of an aerobic bioreactor" (in Spanish). *Proceedings of (IFAC) Congreso Latinoamericano de Control Automático*, CLCA 2002, Guadalajara, Jal., Mexico, 4-6 December 2002. CD, Paper: 120_Betancur.pdf

Fibrianto, Hadyan and Dochain, Denis (2004) "WP2 – Model Selection and Parameter dentification, EM2 Associated to C/N Removal Processes". *Efficient Operation of Urban Wastewater Treatment Plants (EOLI) project report.* Nov 2002 – May 2004, EM2_Cnmodel.doc, *(CESAME) – Université Catholique de Louvain, Belgium*

Gaudy Jr. AF, Rosich AF, Garniewsky S, Moran NR, Ekambaram A. (1988), "Methodology for utilizing respirometric data to asses biodegradation kinetics", *42nd Purdue University Waste conference proceedings*, chap 59, 573-584.

Moreno, J., Betancur, M., Buitrón, G. (2003). "Refinement of an aerobic bioreactor model". *Proceedings of the 4th IMACS Symposium on Mathematical Modelling - 4th MATHMOD*, Vienna, Austria, 5-7 February 2003. pp. 1599-1606. (CD, Paper: 482-Text-Jaime-Moreno.pdf), ISBN 3-901608-24-9

Annex 1. Experimental files names and classification

Table 1 shows the selection of files for identification (ID) and validation (VAL) of standard model EM1. Table 2 relates files to procedures for identifying a specific EM1 candidate.

Table 1. Tags and Names of experimental files

Experiment ID (tag)	ShortName (mmddhh)	Experiment data file (Full name)
ID1	121909	unam1_2003-12-19_09 Std FTC 100mgL 1.xls
ID2	121913	unam1_ 2003-12-19_13 Std FTC 100mgL 2.xls
ID3	070218	unam1_2003-07-02_18h49_FT_starvation200mgL-before12h-dataoffline
ID4	070414	unam1_2003-07-04_14h00_FT_starvation200mgL-before24h-dataoffline
VAL1	121917	unam1_2003-12-19_17 Std FTC 100mgL 3.xls
VAL2	070309	unam1_2003-07-03_09h44_FT_starvation200mgL-after12h-dataoffline

Table 2. Experiments for each procedure combination

EM1 candidate	Experiment tag
A1	ID1 ID2
	VAL1
A2	ID1 ID2
	VAL1
A3	ID1 ID2
	VAL1
B1	ID1 ID2
	VAL1
B2	ID1 ID2
	VAL1
B3	ID1 ID2
	VAL1
BC3	ID1 ID2 ID3 ID4
	VAL1 VAL2
C3	ID3 ID4
	VAL2
D3	ID3 ID4
	VAL2

Only FTC kinetics are used for identification as no inflow is present after filling and this facilitates separation of yield coefficients for procedures 1 and 2.

Some kinetics in EDTOC mode may be later used for validation along with FTC ones. Acclimatization kinetics are not used for identification of standard model. Some before-starvation kinetics are used as they are done with the reactor in full standard conditions.

A file-directory-code is assigned as "ammddhh" ("a" is optional for year 2003 but mandatory thereafter) for storage and retrieval simplicity. Each directory will contain all the data of experiments, excel and Matlab files used for identifying or validating it. In the tables below such code is at the beginning of each line.

Standard FTC

3121909 unam1_2003-12-19_09 Std FTC 100mgL 1.xls (ID1)
3121913 unam1_ 2003-12-19_13 Std FTC 100mgL 2.xls (ID2)
3121917 unam1_2003-12-19_17 Std FTC 100mgL 3.xls (VAL1)

50 ACCLIMATIZATION

3040717 unam1_2003-04-07_17h24_FT_acclimatization50mgL-cycle1-dataoffline.xls (refused: no online)
3040915 unam1_2003-04-09_15h40_FT_acclimatization50mgL-cycle2-dataoffline.xls (refused: no online)
3041013 unam1_2003-04-10_13h28_FT_acclimatization50mgL-cycle4-dataoffline.xls (refused: no online)
3041108 unam1_2003-04-11_08h58_FT_acclimatization50mgL-cycle7-dataoffline.xls (refused: no online)
3041410 unam1_2003-04-14_10h55_FT_acclimatization50mgL-cycle8-dataoffline.xls (refused: no online)
3041511 unam1_2003-04-15_11h00_FT_acclimatization50mgL-cycle10-dataoffline.xls (refused: no online)

100 ACCLIMATIZATION

3051717 unam1_2003-05-17_17h45_FT_acclimatization100mgL-cycle1-dataoffline.xls (refused: no online) (14?)
3051817 unam1_2003-05-18_17h30_FT_acclimatization100mgL-cycle3-dataoffline.xls (refused: no online)
3051916 unam1_2003-05-19_16h44_FT_acclimatization100mgL-cycle4-dataoffline.xls
3052010 unam1_2003-05-20_10h21_FT_acclimatization100mgL-cycle6-dataoffline.xls
3052018 unam1_2003-05-20_18h16_FT_acclimatization100mgL-cycle7-dataoffline.xls
3052109 unam1_2003-05-21_09h54_FT_acclimatization100mgL-cycle10-dataoffline.xls

200 ACCLIMATIZATION

3060913 unam1_2003-06-09_13h30_FT_acclimatization200mgL-cycle1-dataoffline.xls
3060917 unam1_2003-06-09_17h32_FT_acclimatization200mgL-cycle2-dataoffline.xls
3061009 unam1_2003-06-10_09h41_FT_acclimatization200mgL-cycle6-dataoffline.xls
3061113 unam1_2003-06-11_13h16_FT_acclimatization200mgL-cycle13-dataoffline.xls (rejected: O trouble)

STARVATION

3042411 unam1_2003-04-24_11h49_FT_starvation50mgL-before8h-dataoffline.xls (rejected: OS mismatch)
3042421 unam1_2003-04-24_21h57_FT_starvation50mgL-after8h-dataoffline.xls

3042819 unam1_2003-04-28_19h41_FT_starvation50mgL-before12h-dataoffline.xls (rejected: No F)
3042910 unam1_2003-04-29_10h24_FT_starvation50mgL-after12h-dataoffline.xls

3050120 unam1_2003-05-01_20h14_FT_starvation50mgL-before24h-dataoffline.xls (rejected: OS mismatch)
3050308 unam1_2003-05-03_08h47_FT_starvation50mgL-after24h-dataoffline.xls

3052612 unam1_2003-05-26_12h59_FT_starvation100mgL-before12h-dataoffline.xls (rejected: anoxic, no F)
3052702 unam1_2003-05-27_02h32_FT_starvation100mgL-after12h-dataoffline.xls

3052710 unam1_2003-05-27_10h37_FT_starvation100mgL-before24h-dataoffline.xls (rejected: anoxic, no F)
3052812 unam1_2003-05-28_12h26_FT_starvation100mgL-after24h-dataoffline.xls

3053020 unam1_2003-05-30_20h55_FT_starvation100mgL-before36h-dataonline.xls xls (rejected: Anoxic)
3060115 unam1_2003-06-01_15h11_FT_starvation100mgL-after36h-dataoffline.xls

3070218 unam1_2003-07-02_18h49_FT_starvation200mgL-before12h-dataoffline.xls (ID3)
3070309 unam1_2003-07-03_09h44_FT_starvation200mgL-after12h-dataoffline.xls (VAL2)

3070414 unam1_2003-07-05_14h00_FT_starvation200mgL-before24h-dataoffline.xls (ID4)
3070518 unam1_2003-07-05_18h05_FT_starvation200mgL-after24h-dataoffline.xls

3070710 unam1_2003-07-07_10h54_FT_starvation200mgL-before36h-dataoffline.xls (rejected: SO mismatch)
3070902 unam1_2003-07-09_09h05_FT_starvation200mgL-after36h-dataoffline.xls (No starvation registered)

Standard ED-TOC

3071116 unam1_2003-07-11_16h12_ED-TOC_standardkinetic-dataoffline.xls
3071515 unam1_2003-07-15_15h28_ED-TOC_replicate1ofstandardkinetic-dataoffline.xls
3071518 unam1_2003-07-15_18h28_ED-TOC_replicate2ofstandardkinetic-dataoffline.xls

VARIATION 25 4CP

3071202 unam1_2003-07-12_02h13_ED-TOC_kineticwithvariationofC-X(+25%in4CP_in)-dataoffline.xls

VARIATION 50 4CP

3071119 unam1_2003-07-11_19h22_ED-TOC_kineticwithvariationofC-X(+50%in4CP_in)-dataoffline.xls
3071123 unam1_2003-07-11_23h30_ED-TOC_KineticwithvariationofC-X(-50%in4CP_in)-dataoffline.xls

VARIATION 75 4CP

3071409 unam1_2003-07-14_09h52_ED-TOC_kineticwithvariationofC-X(+75%in4CP_in)-dataoffline.xls
3071415 unam1_2003-07-14_15h42_ED-TOC_replicate1ofkineticwithvariationofC-X(+75%in4CP_in)-dataoffline.xls

VARIATION +/-50% Air

3071521 unam1_2003-07-15_21h27_ED-TOC_kineticwithvariationof+50%intheAirFR_in-dataoffline.xls
3071605 unam1_2003-07-16_05h20_ED-TOC_kineticwithvariationof-50%intheAirFR_in-dataoffline.xls
3071600 unam1_2003-07-16_00h11_ED-TOC_replicate1ofthekineticwithvariationof+50%intheAirFR_in-dataoffline.xls

Annex 2. Precision of measurements

Table below is an extract taken from EOLI report "Methods WP1.xls"

Parameter	Analytical method	Methodology	Observations
COD	Reactor digestion micro-method. Hach method 420 COD Low Range Colorimetric determination.	10 mL of mixed liquor were filtered with a 0.45 mm membrane (cellulose nitrate or Durapore) and collected in glass vials. They were preserved adding phosphoric acid (pH <2) and refrigerated at 4 ºC before the analysis. To determine the COD, 2 ml of the filtrate were placed in a vial according to the concentration range of the sample (0 to 150 or 0 to 1500 mg/L).	Precision: ± 10 %. Detection limit: 2 mg/L (vials of 0 to 150 mg/L), 5 mg/L (vials of 0 to 1500mg/L).
4-CP	4-aminoantipyrine method (colorimetric method). Modified technique 5530C from APHA (1992).	10 mL of mixed liquor were filtered with a glass microfibre filter GFA (Whatman). Samples were preserved adding phosphoric acid (pH <2) and refrigerated at 4 ºC before the analysis. Generally, 3 mL of the sample were place in a 50 mL flask (final concentration must be lower than 7 mg/L, if not, a more important dilution is required) and were added 40 mL of distilled water, 1.5 mL of buffer solution, 0.5 mL of aminoantipyrine solution and 0.5 mL of K3Fe(CN)6 and finally, complete to 50 mL with water. The phenolic compounds have a reaction with the 4-aminoantipyrine at a pH of 7.9 +/- 0.1 in presence of alkaline medium to form an aminoantipyrine colored complex with color. This is read at 500 nm.	Precision ± 5% (0 to 200 mg/L). Detection limit 0.01 mg/L
DOC	Total Organic Carbon Analyzer	The sample was filtered with a 0.45 mm membrane (cellulose nitrate or Durapore) and collected in glass vials. They were preserved adding phosphoric acid (pH <2) and refrigerated at 4 ºC before the analysis. The analysis were carried out with a Carbon Organic Total Analyzer (Shimadzu 5050) coupled to an autosampler (ASI-5000A). Acidified samples were bubbling for 3 minutes with high purity air to eliminate the inorganic carbon. 150 mL of sample were injected to the devise and the NPOC was determined. In our case this corresponds to the dissolved organic carbon.	Precision: ± 2 mg/L. Detection limit 1mg/L.
Metabolite	Spectrophotometric method	This is an indirect measure of the quantity of The sample was filtered with a glass microfibre filter GFA (Whatman), and measured at 380 nm in a HACH spectrophotometer with a 1cm length cell.	The reference product was not avaliabable to quantify the amount of metabolite
VSS	Volatile Solid Suspended S*tandard Methods* (APHA, 1992).	10 mL of sample were filtered in a glass microfibre filter GFA (Whatman), dry at 103ºC for one hour, weighted and incinerated during 15 min at 550ºC. The volatilized material was evaluated.	Precision: ± 10%.
TSS	Total Solid Suspended. S*tandard Methods* (APHA, 1992).	10 mL of sample were filtered in a glass microfibre filter GFA (Whatman), dry at 103ºC for one hour.	Precision: ± 10%.

Annex 3. Estimation of yield coefficients

1. Extract of 1st year report (covering period from 1 November 2003 to 31 October 2003)

"A unified methodology for identifying the yield coefficients and kinetic coefficients of the model is been proposed by the project leader. Nevertheless, for the same reason explained in section 3.7.2.2.1, it is not feasible to gather trustable kinetics for the biomass. Even then, some manipulation of the equation (1) is possible to theoretically avoid this problem. But some other practical direct methods are available. That is why we want to calculate the k_1 substrate/biomass yield coefficient using various methods and then decide the best approach for this case.
One direct method for calculating it is to measure the total amount of Substrate fed to the reactor in a sequence of consecutive reactions, and then measure the biomass increase for the whole sequence. The number of reaction cycles for the sequence is chosen in such a way that the Biomass can be measured within some acceptable error.
Such experiment was registered between cycles 37 and 48 for the UNAM 1 datasets.

The results are:
k_1=3,7 mg4CP/mgSSV
k_1=5,0 mgCOD/mgSSV

The initial biomass was VSS=2240 mg/L and it increased by 340 mgSSV/L at the end of the sequence for a total of 2580mg/L in the 12 cycles (combined). This means an increase of 15,17% during the whole sequence, and hence less than 1,2% for each reaction cycle."

- Another approach is to use the stechiometric relation:

k_2[mgCOD/mgSSV]=k_1[mgCOD/mgSSV] - 1.42

but [mgCOD/mg4CP]= 5 / 3.7 = 1.35135 ; then

k_2[mgCOD/mgSSV]=1.35135k_1[mg4CP/mgSSV] - 1.42

And then, dividing by k_1 to obtain the estimable (k_2 / k_1)

1.35135 - (k_2 / k_1)[mgCOD/mg4CP] = 1.42/ k_1[mg4CP/mgSSV]

k_1[mg4CP/mgSSV]=1.42/ (1.35135 - (k_2 / k_1)[mgCOD/mg4CP])

Annex 4. Definitions and Units

The variables and parameters given are valid during fill-reaction and full-reaction phases, while the liquor inside the reactor's tank is being homogenized and aerated.

Independent variable and relatives

t	h	Time in hours
t_0	h	Initial time, when fill-reaction phase starts
t_1	h	Time when fill-reaction phase ends and full-reaction starts
t_f	h	Final time, when full-reaction phase ends

EM1 state variables ξ:

S_O=O	mgD.O./L	Dissolved Oxygen (D.O.)
S_C=S	mg4CP/L	Substrate concentration (toxicant)
V	L	Volume level
X	mgVSS/L	Active Biomass concentration as VSS
S_{Oa}=Oa	mgD.O./L	Intermediate state variable inside D.O. sensor (optional)
S_{Osen}=Osen	mgD.O./L	Sensed D.O. (optional)
$\xi_0 = \xi 0$		State variable initial value, at $t = t_0$
ξ_f		State variable final value, at $t = t_f$

Parameters, external variables and functions:

μ=μ(S,O)	h^{-1}	Specific biomass growth rate
k_1=1/$Y_{x/s}$=1/Y	mg4CP/mgVSS	Yield coefficient (Substrate to Biomass)
k_2=1/$Y_{x/o}$	mgD.O./mgVSS	Yield coefficient (Oxygen to Biomass)
b	h^{-1}mgD.O./ mgVSS	Specific Endogenous respiration rate
k_La(V)	h^{-1}	Oxygen mass transportation coefficient as $k_l a(V) = K_l a * V_{max}/V$
K_La	h^{-1}	Value for $k_l a$ when Volume is maximum as $K_l a = k_l a(V_{max})$
K_o=	mgD.O./L	Monod constant for 50% oxygen limitation effect in μ
$S_{O,in}$=O_{in}	mgD.O./L	D.O. in influent flow
V_{max}	L	Final volume level as V_{max}=7L
$S_{C,in}$=S_{in}	mg4CP/L	Substrate concentration in influent flow.
S_{Os}=O_s=O_{sat}	mgD.O./L	Saturation oxygen for liquor. Vary with temperature.
X_f=VSS	mgVSS/L	X at end of the reaction phase
τ_a=Ta	h	Time constant of first first order filter stage for D.O. sensor
τ_b=Tb	h	Time constant of second first order filter stage for D.O. sensor

Kinetic parameters: (K-parameters)

μ_0= mu0	h^{-1}	Maximum Non-inhibited Specific Biomass Growth Rate
K_I=Ki	mg4CP/L	Inhibition constant
K_S=Ks	mg4CP/L	Saturation (Michaelis-Menten) constant

Kinetic parameters: (S-parameters)

μ*= mue	h^{-1}	Maximum value of Specific Biomass Growth Rate
μ_t= 0.1 μ*	h^{-1}	μ* fraction at discontinuity threshold (optional)
S*= Se	mg4CP/L	Substrate concentration for maximum Biomass growth rate
Sm	mg4CP/L	Substrate concentration for medium Biomass growth rate
St	mg4CP/L	Threshold value (optional)

Annex 5. Kinetic Parameterization

The kinetic parameters that model the Specific Biomass Growth rate, for inhibitory Haldane type substrates, are traditionally K_i, K_s, μ_o (K-parameters) as in Eq.1 of this annex:

$$\mu = \frac{\mu_0 S}{K_S + S + S^2 / K_I} \qquad \text{Eq. (1)}$$

An equivalent simple re-parameterization is suggested (S-parameters) in Eq. 2. It offers various advantages. It gives a different and direct graphical/physical meaning to the parameters in relation to key-points in the specific biomass growth rate graph. Explanation of the meaning of each S-parameter is given in Table 1.

Table 1. S-parameters

S-Parameter names	Definition	Concept
mue=μ*	μ* = max(μ (S))	Maximum of the biomass growth rate. (S>0)
Se=S*	μ(S*) = μ*	Substrate at which the maximum μ* happens
Sm	μ(Sm) = μ*/2	Substrate for medium decay (for Sm>S*)

$$S^* = \sqrt{K_I K_S}$$
$$\mu^* = \frac{\mu_o}{1 + 2K_S / S^*} \qquad \text{Eq. (2)}$$
$$S_m = \frac{K_I + 4S^* + \sqrt{K_I^2 + 8K_I S^* + 12S^{*2}}}{2}$$

Formulas for parameter conversion are given in StoK.m and KtoS.m Matlab files.

One advantage is that by changing *Se* or *Sm* the real gain does not get affected, as is the case when changing K_I or K_S. In fact, the peak remains exactly at the same point if *Sm* is changed, allowing to fix the behavior of the growth rate near the optimal operational point (by using *Se* and *mue*) and then explaining the inhibitory behavior by moving, basically, only *Sm*.

The same concept applies if *Sm* is fixed and *Se* moves. In this case the inhibitory behavior gets almost fixed, and by moving *Se* it is possible to adjust the position of the optimal operational point without worrying too much for misadjusting the right part of the graph.

One other key advantage of S-parameters is that of getting the mean values of different experiments. As an example, results from procedure A1 are taken and the respective Haldane graphs plotted in figure 1 of this annex.

In this example it is clear that the mean of K-parameters is not related to the mean of the Haldane graphs. Using the S-parameters instead reduces the error and gives logical results.

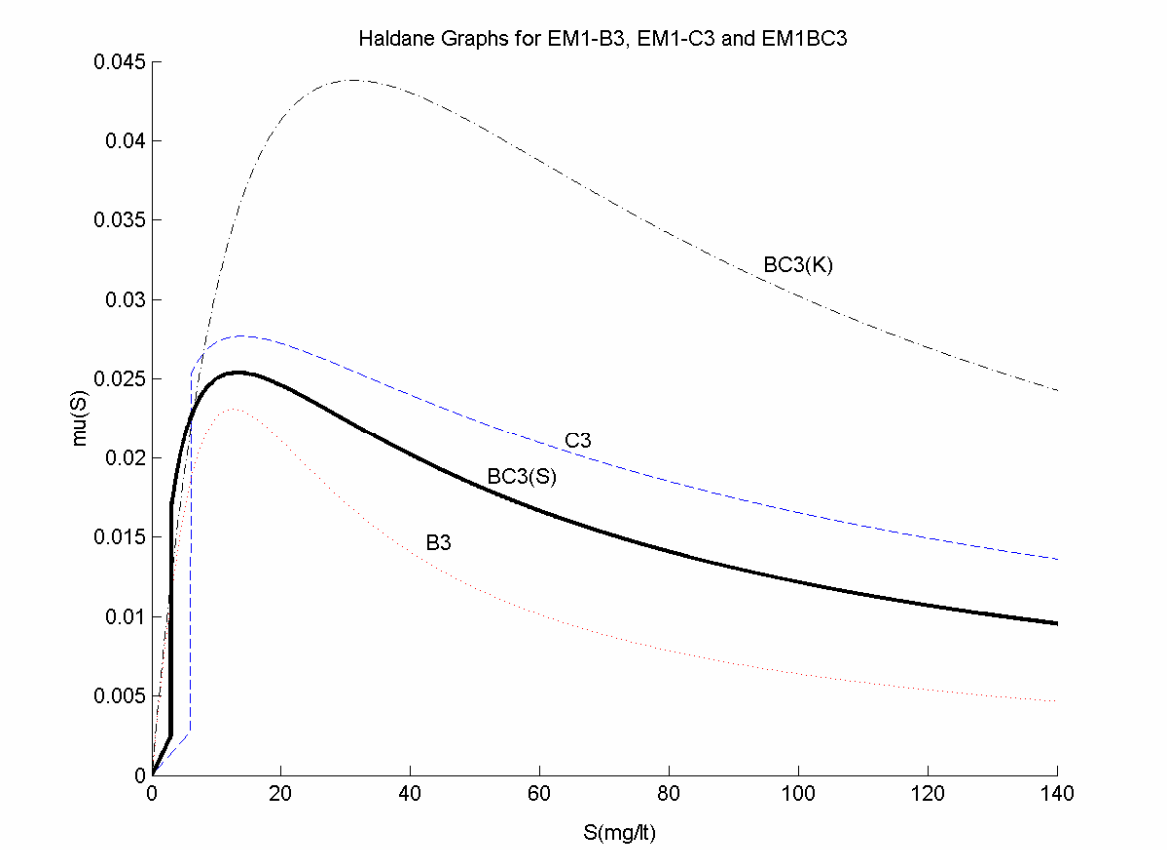

Figure 1. The red(dotted) and blue(dashed) lines are two identification results (EM1-B3 and EM1-C3). The Black-continuous one ($BC3_{(S)}$) is their mean graph obtained averaging each S-parameter (mue, Se, Sm, Sk). The Black-dotted-dashed one ($BC3_{(K)}$) is the mean graph obtained averaging each traditional K-parameter (mu0, Ki, Ks).

Annex 6. Modeling Assumptions

The assumptions are taken, and modified, from a previous work presented in the midterm report, available in file WP2 (model) Mid-Term P6 (unam)1.4.doc

Original Assumption 13 was changed. A Monod term to explain oxygen limitation is added to the model of the Specific Biomass Growth Rate, and the corresponding *Ko* is added to the list of parameters to identify.

Original Assumption 2 was also changed. Confirmation from UNAM's WP1 team about this assumption revealed that influent may contain oxygen at appreciable levels.

New Assumptions for procedure C and D are:

CD1. Saturation DO remains constant throughout the reaction, i.e. the temperature inside the reactor is constant, and so is pressure and other factors that might alter its value.

CD2. DO level in the influent is important, but unknown.

CD3. Aeration and biomass distribution are isotropic inside the reactor

CD4. Transport time is neglectable, i.e. the system can be modeled using ordinary differential equations instead of partial ones.

CD5. Biomass growth in one standard reaction is small, below 2%, i.e that the biomass concentration changes mainly only because of volume changes.

CD6. No biomass is added to the SBR, except when acclimating. EM1 does not model that.

CD7. No biomass is lost during decantation, except when purging. EM1 does not model that.

CD8. Biomass will not die appreciably during reaction time. EM1 does not model that.

CD9. *Kla* will vary with volume but not with substrate concentration.

CD10. For control purposes only the fill-reaction and full-reaction phases are of direct importance, hence no modelization or special control is necessary for other phases as sedimentation, decantation, etc… This means that standard timed open loop control may be used for the other phases.

CD11. μ is of the Haldane type or close to it. This means it grows monotonically from substrate concentration zero to some optimal substrate concentration S^*, where it takes the maximum value μ^*, and beyond that it decreases monotonically.

CD12. *$\mu(S)$ parameters do* not change appreciably in one standard reaction.

CD13. D.O. concentration will remain above some given minimum concentration during a standard reaction. Failing to do so is interpreted as a sign of malfunction. Even then modeling for oxygen depletion limitation should be provided.

CD14. Before starting to fill, some pre-aeration will be allowed in order to fulfill the assumption above.

Original Assumptions were:

1. Saturation DO remains constant throughout the reaction, i.e. the temperature inside the reactor is constant, and so is pressure and other factors that might alter its value.
2. DO level in the influent is neglectable (not assumed in procedure B)
3. Aeration and biomass distribution are isotropic inside the reactor
4. Transport time is neglectable, i.e. the system can be modeled using ordinary differential equations instead of partial ones.
5. Biomass growth in one standard reaction is small, below 2%, i.e that the biomass concentration changes mainly only because of volume changes.
6. No biomass is added to the SBR, except when acclimating. EM1 does not model that.
7. No biomass is lost during decantation, except when purging. EM1 does not model that.
8. Biomass will not die appreciably during reaction time. EM1 does not model that.
9. *Kla* will vary with volume but not with substrate concentration.
10. For control purposes only the filling and reaction phase is of direct importance, hence no modelization or special control is necessary for other phases as sedimentation, decantation, etc… This means that standard timed open loop control is used for the other phases.
11. μ is of the Haldane type or close to it. This means it grows monotonically from substrate concentration zero to some optimal substrate concentration S^*, where it takes the maximum value μ^*, and beyond that it decreases monotonically.
12. μ does not change, as a function of S, appreciably in one standard reaction.
13. D.O. concentration will remain above some given minimum concentration during a standard reaction. Failing to do so is interpreted as a sign of malfunction. This means no modeling for oxygen depletion is necessary for control purposes.
14. Before starting to fill, some pre-aeration will be allowed in order to fulfill the assumption above.

Annex 7. Initial Conditions and Estimable Parameters

Initial conditions for simulations come from experimental data for the state variables, and the initial parameter values were given by simulations using a procedure similar as that reported in the midterm report WP2 (model) Mid-Term P6 (unam)1.4.doc

The initial conditions used for the kinetic identification in C procedure were (EM1_Matlab_library\sIni_02.m) :

μ^*	=	*mue*	$=0.095/k_1$
S^*	=	*Se*	=12
Sm	=	150	
St	=	8	
Ko	=	0.1283	
Oin	=	0	

"D" procedure uses these same conditions, except that *St* is set to zero.

As before, the initial time t0=t_0=0 is assumed as the instant at which the influent's pump is turned on and the filling begins. This is true for simulations tests and for identification of kinetic parameters. However, for identifying the yield coefficient relation (K_2/K_1), it is necessary to avoid the fill-react phase. That is one reason why no ED-TOC kinetics may be used for such purpose.

Description of initial values for the variables follows:

1. Substrate: Initial (at t=t_0) substrate concentration inside the reactor is usually neglected, i.e. considered zero, unless a direct measure is available.
   $S_{C,o}$= S0=0 (mg/L), (or_use_measured_value_if_available)

2. Volume: Initial volume is always assumed to be 3L. No level sensor is available. Calibration of the draw-pump suction configuration is assumed to be such that draw always take away volume in excess of 3L. Final volume is estimated to 7L by integrating the commanded input flow.
   Vo = V0=3 (L), Vmax=Vfin=7 (L)

3. Oxygen: An automated log-file for capturing the oxygen readings is available. This data was preferred over manually taken data. The initial value for Dissolved oxygen concentration is taken from there.
   $S_{O,o}$= O0=first_datum_from_automatic_log_file (mg/L)

4. Biomass: VSS is measured at the end of reaction. Biomass is assumed to be equal to VSS at this time. Estimation for initial biomass is made by using the yield coefficient and the known quantity of substrate degraded, using the following equations:

TS=(Vf-V0)*Sin	Total substrate mass treated
Bp=TS/K_1	Total biomass produced during the reaction
Bf=Vss*Vf	Total biomass at the end of reaction
X0=B0/V0	Initial Biomass concentration
X0=(Bf-Bp)/V0	

   $S_{C,o}$= X0=(Vss*Vf-(Vf-V0)*Sin/K_1)/V0

From table in annex 2 it can be seen that the precision of the measurement of VSS, and hence of the estimation of X0, is 10%. No provisions are taken to compensate this fact as for the Yield part identification its effect is not seen, and for the kinetics part identification most of the error's effect is masked by an equivalent but opposite error in estimating mu0 (or mue). This thinking shows that the tolerance in the confidence interval for mu0 (mue) may be affected by this error, but the simulation as a whole is not.

5. Inflow Substrate concentration: Again, annex 2 shows that the precision for measuring the 4CP concentration is 5%. This is important as an error in S_{in} may affect greatly the simulation performance for EM1. A possible action to minimize the associated error that propagates from using a badly measured S_{in} is to include it in the identification procedure for the kinetic part. Lower and upper bounds may be given to restrict the S_{in} estimation to the given precision. However, initial tests conducted to include S_{in} in the list of parameters to identify did show poor performance, so the idea was discarded. That is why the value used is the one measured.
   $S_{C,in} = S_{in}$=as_measured (mg4CP/L). *Tolerance=5%*

6. Saturation Dissolved Oxygen: Changes with temperature and Biomass and Substrate concentration. As all of them change over time, it was preferred to calculate it (to choose the best constant-value) for each reaction. An estimation is made by using the equilibrium point of model in equation 1 after finishing the reaction (Dilution=0, X=kte, S=0, O_l=$O_{max(V=Vmax;S=0)}$).
   $\boldsymbol{S_{Os} = O_s = O_{max} + b*VSS/K_la}$

7. Influent's Dissolved Oxygen: Unlike in procedure A, procedures B,C and D assume that there might be dissolved oxygen in the influent flow. However no direct measures were taken for all experiments. One solution to avoid uncertainty is not to use Oxygen data while filling. The other is to include it as a parameter to identify. This last option was included in Procedure B, C and D. Future experiments should report this value explicitly.
   $\boldsymbol{S_{O,in} = O_{in} = 0}$ *initially, but become part of the parameters to identify*

Annex 8. D.O. Sensor Model

An experiment was conducted for identifying the sensor's model. Two beakers filled with water were provided. One of them had zero D.O. by means of an added chemical. The other was saturated by means of an aerator. The sensor (COS4, Endress+Hauser) was let in the saturated beakers and then quickly moved to the other, before stabilizing, to generate step responses, up and down. An additional beaker with saturated clean water was provided for quickly washing the probe in order not to contaminate the clean water beaker with the chemical added to the other one. The sampling time was 0.005h (18 sec). Data is contained in file "ModeloSensorOxigeno 2002-07-11"

The model proposed is represented in Eq. 1 of this annex. A comparison between simulated and sensed D.O is shown in figure 1 of this annex.

$$\dot{O}_a = (O_{real} - O_a) / \tau_a$$
$$\dot{O}_{sensor} = (O_{sensor} - O_a) / \tau_b \qquad \text{Eq. (1)}$$

With Ta=0.01h and Tb=0.02h

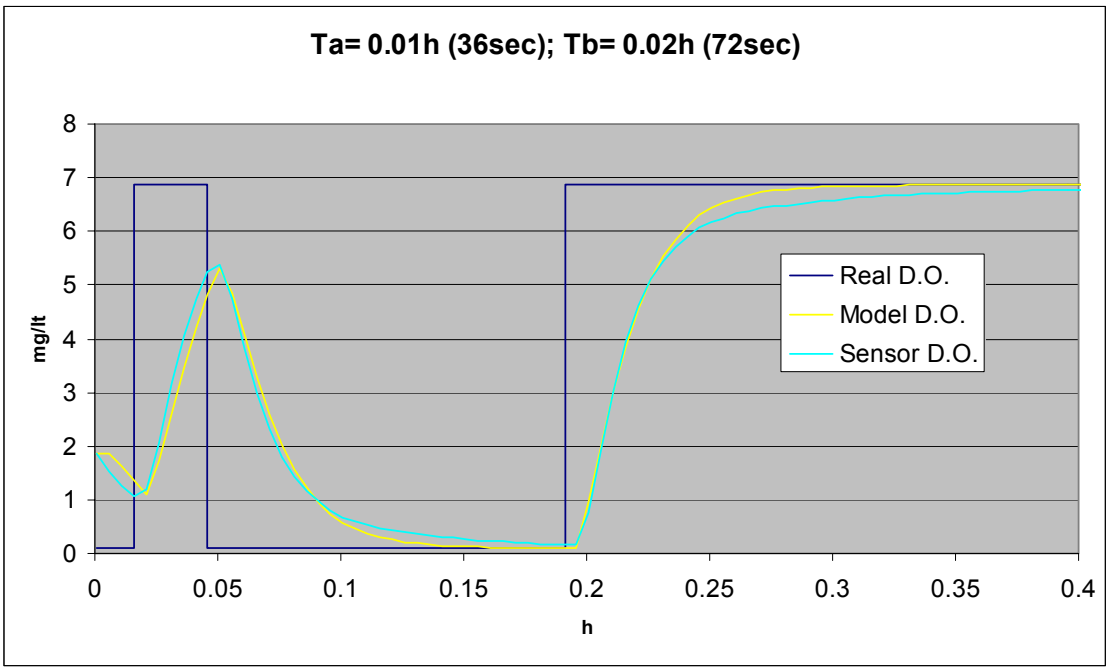

Figure 1. (Annex 8). Experiment for identifying D.O. sensor. The blue straight line is the real oxygen to which the probe was exposed (by quickly changing it from one beaker to another with opposite D.O. conditions). The blue one is the sensed signal. The yellow one is the model.

Annex 9. Software, Fitting criteria and Statistical guidelines

The parameters to identify in procedure C are μ^*, S^*, Sm, St, K_2, K_0 and O_{in}. For doing this, a series of Matlab m-files were written. Both Levenberg-Marquardt and Gauss-Newton algorithms were tested. No significant difference was found in the former tests performed so Gauss-Newton was used, for that being the default one.

The versions of the software used are MATLAB Version 6.5.0.180913a (R13) and Optimization Toolbox Version 2.2 (R13).

To test the capability of the designed software a TEST example was constructed. Using a known model, a virtual experiment was generated in an independent simulator. The data gathered was treated as data coming from a real reactor. A fitting criterion including substrate and oxygen error was designed. Initial trials to identify this model revealed the poor performance of the program when the given initial conditions were not close to the real ones. Then the fitting criterion was changed and again the algorithm behaved well for some cases but no so well for others. Finally a combination of criteria was used to iteratively apply the identification method. The software now is such that the interactivity may be preprogrammed or manually oriented.

To test a dataset of a given experiment, an infinite time horizon (ITH) simulation is run, and data from simulation is compared to experimental data at real sampling points using a given fitting criterion. ITH means that, other than initial conditions, no experimental data is directly incorporated to EM1 simulation at any other time.

Fitting criteria: 3 different criteria were defined, all of the type "MQE=Sum(F_k^2)/n". One of them (1) forgives errors in Oxygen, the other (2) forgives errors in Substrate, and the last one (3) includes all of them and weights the oxygen over the substrate data:

(1). Fit_S: $F_k=FS_k=S_k$simulated-$S_k$measured (k=1,…,n; n=number of samples)
(2). Fit_O: $F_k=FO_k=O_k$simulated-$O_k$measured
(3). Fit_SO: $F_k=W_o*FO_k$; $F_{k+n}=W_s*FS_k$

For reference, if Substrate concentration is assumed to be in the range [0 , 200]mg/L and Oxygen concentration in the range [0 , 10]mg/L, an error of *error%* (of maximum range) in ALL samples will yield a MQE for Fit_SO criterion (3) as calculated in file MQE_interpretation.xls and shown in Table 1.

The criterion (3) is the natural one. It combines the first two, establishing a compromise between Substrate and Oxygen fitting errors. This is considered to be of importance as the optimal control law and the supervision system will both depend on Oxygen measurements to software-sense what happens with the unmeasured Substrate.

Criterion (3) was the one first tested, but convergence was not appropriated for all cases. Even when OPTIMSET (algorithm set-up in Matlab) was set to more stringent settings (to allow more iterations, diminish tolerances, change methods…), convergence was still poor for some cases. That is why the two different additional criteria were defined. The "lsqnonlin" routine may be run with any of these criteria, and one single run will be called "a step". The idea then is to combine steps, combining criteria, one different criterion in each step. This pushes the fit-search in different directions at each identification step. A combination of steps rendered better global

convergence for most initial conditions and for most cases. Note that the global final criterion for a sequence of steps is always defined as to minimize Fit_SO, regardless of the amount of steps and independently of the criterion used in each particular step.

The possibility to manually intervene in the fitting process was incorporated to the software. This means that at any step the user may (optionally) modify some of the parameters, and then continue with the next step. In the end it is the user who decides which one of the multiple tools to use and how many steps to apply.

To avoid the use of suspicious data, a flag was associated to each datum. If the flag value is "1" then it is used in the fitting procedure and its graph consists of an asterisk at the moment of the sample. If the flag is set to zero, it is not used and its graph will appear as an asterisk at "zero" height (but again at the time the sample was taken). In the "***kid***" (***k***inetic ***id***entification) files this is made by the use of vectors *Ofit* and *Sfit*, wich elements relate to those in *texp*, *Oexp* and *Sexp* (experimental data)

Table 1 (Annex 9) Mean Quadratic Error (MQE) as function of error for ideal case of all errors being equal

% error	MQE
2	7
3	15
4	27
5	42
6	60
7	82
8	107
9	135
10	167
11	202
12	240
14	327
16	427
18	540
20	667
22	807
24	960
25	1042
30	1500

Statistical treatment of parameter values was made in the same manner for all WP2. Extracts of annex 2 from EM2_CNmodel.doc (Fibrianto, H and Dochain D., 2004) follow:

"The confidence interval is a range on either side of a sample mean. For example, if you order a product through the mail, you can determine, with a particular level of confidence, the earliest and latest the product will arrive.

To compute the confidence interval of a set of data, the following information have to be provided:

1. Alpha (α)

Alpha is the significance level used to compute the confidence level. The confidence level equals to 100 × (1–α) %, or in other words, an alpha of 0.05 indicates a 95 percent confidence level.

2. Standard deviation

The standard deviation, denoted σ, is a measure of how widely values are dispersed from the average value (the mean). It is given by the following expression:

$$\sigma = \sqrt{\frac{n\sum x^2 - \left(\sum x\right)^2}{n(n-1)}}$$

where *n* is the sample size, and *x* is the sample values.

3. Sample size

The sample size is the number of data that will be treated.

If we assume α equals 0.05, we need to calculate the area under the standard normal curve that equals (1 - α), or 95 percent. This value is ± 1.96. The confidence interval, CI, is therefore:

$$\text{CI} = \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

Hence for a random population of a mean value $\bar{x}$, its confidence interval with 95 percent of confidence level includes:

$$\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

with $\bar{x} = \frac{1}{n}\sum_n x_i$ "

Annex 10. Links to related documents (reports, graphs, data, code…)

Link to numerical results for parameter values and confidence intervals, and validation tests.

Additional files\EM1_UNAM1_Confidence_Intervals.xls

Links to former reports:

Midterm preliminary/exploratory report

Additional files\WP2 (model) Mid-Term P6 (unam)1.4.doc

Former EM1-A and EM1-B report

Additional files\EM1_AB (WP2_UNAM)1.5.doc

Links to Kinetical Identification graphs, data (kid.doc) an code:

Kids for EM1-C

(ID3) 070218\kid070218c3.doc, (ID3) 070218\kid070218c3.m
(ID4) 070414\kid070414c3.doc, (ID4) 070414\kid070414c3.m

Kids for EM1-D

(ID3) 070218\kid070218d3.doc, (ID3) 070218\kid070218d3.m
(ID4) 070414\kid070414d3.doc, (ID4) 070414\kid070414d3.m

Links to EM1 Tests graphs, data (kid.doc) and code:

Tests for EM1-C

(ID3) 070218\test070218c3.doc, (ID3) 070218\test070218c3.m
(ID4) 070414\test070414c3.doc, (ID4) 070414\test070414c3.m
(VAL2) 070309\test070309c3.doc, (VAL2) 070309\test070309c3.m

Tests for EM1-D

(ID3) 070218\test070218d3.doc, (ID3) 070218\test070218d3.m
(ID4) 070414\test070414d3.doc, (ID4) 070414\test070414d3.doc
(VAL2) 070309\test070309d3.doc, (VAL2) 070309\test070309d3.m

Tests for EM1-BC

(ID1) 121909\test121909bc3.m
(ID2) 121913\test121913bc3.m
(ID3) 070218\test070218bc3.m
(ID4) 070414\test070414bc3.m
(VAL1) 121917\test121917bc3.m
(VAL2) 070309\test070309bc3.m

Link to data for D.O. sensor modeling

Additional files\ModeloSensorOxigeno 2002-07-11.xls

Links to experiments data (offline and online)

(ID1) 121909\unam 2003-12-19_09 Std FTC 100mgL 1.xls
(ID2) 121913\unam 2003-12-19_13 Std FTC 100mgL 2.xls
(ID3) 070218\unam1_2003-07-02_18h49_FT_starvation200…line.xls
(ID4) 070414\unam1_2003-07-04_14h00_FT_starvation200…line.xls
(VAL1) 121917\unam 2003-12-19_17 Std FTC100mgL 3.xls
(VAL2) 070309\unam1_2003-07-03_09h44_FT_starvation200…line.xls