EM1 IDENTIFICATION PROCEDURES A,B

(Reactor 1 : LabScale)

EOLI, WP2 (Model), UNAM (Partner 6)

This document contains only version for models EM1Ax and EM1Bx.
See "EM1_CD (WP2_UNAM)" for models EM1Cx and EM1Dx

CONTENTS

Manuel J. BETANCUR

31 May 2004
Updated for nomenclature 29 October 2004

Contract Number ICA4-CT-2002-10012

1. INTRODUCTION

The objective of the identification process is to render a mathematical EM1 (Eoli model 1) suitable for studying the classical FTC (Fixed Time Control) and the ED-TOC (Event Driven Time Optimal Control) strategies applied to an SBR treating toxicants. Some previous work was presented in the first annual report and also in the midterm report. In the document WP2 (model) Mid-Term P6 (unam)1.4.doc there is an explanation of the process and the initial assumptions for modeling, as agreed in Mexico meeting. Two of those assumptions are revised now (see annex 6).

In particular, it was decided that FTC experiments are to be used for the identification, so the model should explain the experimental results in FTC mode. A more demanding objective would be to explain the behavior of the plant NEAR the optimal operational zone, as when the ED-TOC strategy is applied. Nevertheless, only FTC kinetics are suitable for Yield parameters identification using the proposed method, the software available and the given low sampling frequency.

The identification procedure proposed here may take several forms depending on the combinations of methods for identifying, separately, yield and kinetic parameters. Initially six of such combinations are to be studied, each one producing an EM1 candidate. "Kinetic" options will be tagged by letters, while "Yield" options will be tagged by numbers, as follows:

A relation among Yield coefficients (k_2 / k_1) can be found by separating them from dynamical ones. Two versions appear for calculating k_2 by using such relation:

1) k_1 is estimated from experimental measures. See estimation procedure in annex 3.
2) A stechiometric relation is used to compute k_1 and, simultaneously, k_2

The two versions conform to the one method agreed in Mexico and follow the instructions of the work-package leader. Another option, not included in and not conforming to the original plan, is also tested:

3) k_1 is estimated independently as in option1 (annex 3) but, to identify k_2, the same program as the one used for identifying the kinetic parameters is used.

Once yield coefficients are available, kinetic ones are identified using MatLab "lsqnonlin" function. This means there are 3 "Yield" versions for each "Kinetic" procedure, each version with different yield coefficients (different because different methods were used for obtaining them).

The proposed options of procedures for kinetic estimation are:

A) Two (2) FTC Standard Experiments are chosen for identification and one for validation. $S_{O,in}$ is assumed zero. Parameters to identify are: μ_0 , K_I , K_S,, Ko

This version conforms to the original assumptions. However, preliminary tests did show a mismatch of the oxygen graph during the filling time. As there is no experimental report for the $S_{O,in}$ value, another option was added to test the possibility that some oxygen is coming in from the inflow:

B) Similar to procedure A, but $S_{O,in}$ is supposed unknown. Parameters to identify are: μ_0, K_I, K_S, Ko, $S_{O,in}$

Other than the 6 possible combinations, each one rendering a possible model, optional future procedures to explore non FTC standard kinetics may be included in future works:

X) ED-TOC kinetics may be added also for (kinetic only) identification/validation.

2. EXPERIMENT SELECTION

Before the first trial for finding yield parameters using FTC data, an assessment was made about whether or not the consistency of data was ok for all experiments listed in annex 1. If there were missing data the experiment was refused for validation or identification purposes. If anoxic conditions, or any non standard conditions are present, or if substrate or inflow data does not match substrate reasonably, such experiments were also rejected.

All 3 of the FTC standard experiments were accepted. One of these was chosen for validation, so the other two are used for identification. Two other experiments from the starvation series seemed also plausible to use for identification/validation. However, after discussing the preliminary Yield coefficients identification results with UNAM's WP1 team, an agreement was made not to use such experiments as they correspond to a different biomass. In particular it is (stressed) biomass in STARVATION cycles and it was changed before performing the new FTC standard experiments. As biomass is taken from a municipal WWTP and subjected to an acclimation process to the specific 4CP toxicant, it is reasonable to expect differences in model parameters from one biomass to the next. That is exactly what results form Yield Identification parameters suggested.

Separation of the yield coefficients relation (k_2 / k_1) allowed identifying it separately using a linear regression. That was made for the two chosen standard experiments and also for the two starvation ones. After the individual experiments identification of (k_2 / k_1) were made, an Excel file was used to present results and to calculate the confidence intervals, using STDEV Excel function. Such function for calculating the confidence interval was agreed, in LLN meeting, with the work-package leader.

Once a given (k_2 / k_1) is accepted, identification of the other parameters is made. For this, a Matlab environment was used. The results are presented in the same Excel file as above.

The experiments selected from Annex 1 are:

IDENTIFICATION PROCEDURES A,B:
3121909 unam1_2003-12-19_09 Std FTC 100mgL 1.xls (ID1)
3121913 unam1_ 2003-12-19_13 Std FTC 100mgL 2.xls (ID2)

ADDITIONAL EXPERIMENTS FOR IDENTIFICATION OF YIELD COEFFICIENTS:
3070218 unam1_2003-07-02_18h49_FT_starvation200mgL-before12h-dataoffline.xls (ID3)
3070414 unam1_2003-07-05_14h00_FT_starvation200mgL-before24h-dataoffline.xls (ID4)

VALIDATION :
3121917 unam1_2003-12-19_17 Std FTC 100mgL 3.xls (VAL1)

3. INITIAL CONDITIONS AND ESTIMABLE PARAMETERS (FROM DIRECT MEASUREMENTS)

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 filling phase. That is one reason why no ED-TOC kinetics may be used for such purpose.

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)

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)

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)**

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)*$S_{C,in}$/k_1)/Vo
From table in annex 2 it can be seen that the precision of the measurement of VSS, and hence of the estimation of Xo, 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 μ_0 (or μ^*). This thinking shows that the tolerance in the confidence interval for μ_0 (μ^* respectively) may be affected by this error, but the simulation as a whole is not.

Inflow Substrate concentration: Again, annex 2 shows that the precision for measuring the 4CP concentration is 5%. This is important as an error in $S_{C,in}$ may affect greatly the simulation performance for EM1. A possible action to minimize the associated error that propagates from using a badly measured $S_{C,in}$ is to include it in the identification procedure for the kinetic part. Lower and upper bounds may be given to restrict the $S_{C,in}$ estimation to the given precision. However, initial tests conducted to include $S_{C,in}$ 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%*

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_C=0, S_{Of}=$S_{Omax(V=Vmax;S=0)}$).
$S_{Os} = O_s = S_{Omax} + b*VSS/K_La$

Influent's Dissolved Oxygen: It is assumed that there is no 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. Future experiments should report this value explicitly.
$S_{O,in}$= 0 for procedure A. *May become part of the parameters to identify (Procedure B)*

4. YIELD PARAMETERS IDENTIFICATION

The model in Eq. 1 is used. See annex 4 for definitions and units. See annex 6 for modeling assumptions.

$$\begin{aligned} \dot{X} &= \mu X - X\frac{q_{in}}{V} \\ \dot{S}_C &= -k_1 \mu X + (S_{C,in} - S_C)\frac{q_{in}}{V} \\ \dot{S}_O &= -k_2 \mu X - bX + K_L a * (S_{Os} - S_O) + (S_{O,in} - S_O)\frac{q_{in}}{V} \\ \dot{V} &= q_{in} \end{aligned} \qquad \text{Eq.(1)}$$

For FTC batches identification is made for $t > t_1$ where t_1 is assumed to be the end of filling time, and $t_a > t_1$ is the next sampling time, i.e. the initial time for the usable data for identification. Hence, for yield coefficients separation, it can be assumed that $q_{in}=0$ in Eq. 1 for $t>t_1$

As no kinetics is available for biomass (its net growth in one reaction is comparable to the measuring error, so it is useless to measure its kinetic), biomass equation (first line of Eq. 1) is useless for identification purposes in one single experiment. That is why only Substrate and Oxygen equations are used (second and third lines in Eq. 1).

After replacing μX from second line into the third one and integrating:

$$S_O - S_{O\,a} = \frac{k_2}{k_1}(S_C - S_{Ca}) - \int_{t_a}^{t} (bX - K_L a * (S_{Os} - S_O))dt \qquad \text{Eq.(2)}$$

But it is possible to simplify the procedure if some S_{Omax} is defined in such a way that it is the maximum D.O. when endogenous respiration is active and no substrate is present. It is possible to estimate S_{Omax} from the end of each experimental kinetic. In this way S_{Os} temperature dependence is eliminated and, at the same time, bX disappears from Eq. 2. And, because temperature varies slowly and X kinetic is really slow (less than 2% growth in each standard batch), it is logical to assume that, approximately, $bX=K_La(S_{Os} - S_{Omax})$. Replacing in Eq. 2:

$$S_O - S_{Oa} - K_L a \int_{t_a}^{t} (S_{O\max} - S_O)dt = \frac{k_2}{k_1}(S_C - S_{Ca}) \qquad \text{Eq.(3)}$$

Eq. 3 is used for linear regression. The integral is made assuming triangular reconstruction of Dissolved Oxygen graph. Note that for such procedure suspicious and non relevant data points were taken out.

The files used for the computations are (See notation equivalences in annex 4):
(ID1) 3121909 Std FTC 100mgL 1\unam 2003-12-19_09 Std FTC 100mgL 1.xls
(ID2) 3121913 Std FTC 100mgL 2\unam 2003-12-19_13 Std FTC 100mgL 2.xls
(ID3) 3070218 stv200bef12\unam1_2003-07-02_18h49_FT_starvation200mgL-before12h-dataoffline.xls
(ID4) 3070414 stv200bef24\unam1_2003-07-04_14h00_FT_starvation200mgL-before24h-dataoffline.xls

Results are summarized in page "Yield" of file EM1_UNAM1_Confidence_Interval.xls

5. KINETIC PARAMETERS IDENTIFICATION

An approximation to the kinetics for μ in Eq. 1 is given by Eq. 4. There, two terms describe the relation of the growth rate with respect to substrate concentration (assumed Haldane for inhibitory behavior when substrate is toxicant) and also with respect to oxygen limitation (assumed Monod to explain the loss of activity in the case of acute absence of oxygen, while having negligible effect otherwise).

$$\mu = \mu_0 \left(\frac{S_C}{K_S + S_C + \frac{S_C^2}{K_I}} \right) \left(\frac{S_O}{S_O + K_o} \right) \qquad \text{Eq.(4)}$$

The parameters to identify are μ_o, Ki_i, K_s, K_o from Eq. 4 and optionally $S_{O,in}$ and k_2 from Eq. 1. 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 results. When one of them stalled, so did the other. When one converged, also did the other.

The versions of the software used are:

- MATLAB Version 6.5.0.180913a (R13)
- 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.

Fitting criteria: 3 different criteria were defined, all of the type "Criteria=Sum(F_k^2)". 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: $F_k=W_o*FO_k$; $F_{k+n}=W_s*FS_k$

The last criterion 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 control law will depend on Oxygen measurements to software-sense what happens with the unmeasured Substrate. It 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.

But convergence was still not always good enough, as for some initial conditions the algorithm converged to local minima, thus not finding the "real" (TEST) model. Then complementary fitting m-files were added, using a re-parameterization of the kinetic model, as this had shown in the past to be useful for visual analysis. Re-parameterization and its advantages are explained in annex 5.

The result is that combining the 6 types of identifying algorithm (3 for the classical K-parameters and 3 for the re-parameterized S-parameters), better global convergence was achieved in most cases. Even then, 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)

Initial condition for the simulations come from the experimental data for the state variables, and the initial parameter values were given by "manual" simulations done in BIOREV, 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 were (EM1_Matlab_library\sIni_01.m) :

μ^* = *mue* =0.08/k_1
S^* = *Se* =15
Sm = *Sm* =60

RESULTS

Numerical parameter results for each experiment, their mean values, and also their confidence intervals, are given in file EM1_UNAM1_Confidence_Interval.xls. One page for each possible model candidate A1, A2, A3, B1, B2, B3. Table 1 summarizes the yield coefficients identifying procedure versions and Table 2 summarizes the kinetic coefficients identifying procedure versions.

Table 1. Yield coefficients identifying procedure versions.

Version	Procedure (Yield coefficients)
1	k_1=3,7 mg4CP/mgSSV is estimated from data coming from a series of batches where total substrate consumption and total biomass generation was measured 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 and the identified relation (k_2/k_1)
3	k_1=3,7 as in 1) k_2 is identified using the same program as for kinetic parameters.

Table 2. Kinetic coefficients identifying procedure versions.

Version	Procedure (Kinetic coefficients)
A	2 FTC experiments are used, input Oxygen assumed negligible
B	2 FTC experiments are used, input Oxygen assumed unknown, but constant (identify)
C	2 FTC experiments and 2 Starvation experiments are used (Optional, Pending)

Table 3. "Mean Quadratic Error" (MQE) for combined Identification Procedures

P.Kinetic \ P.Yield	1	2	3
A	785	768	386
B	371	354	173

Table 3 compares the fitting error for different combination of versions. Results show that the best fit is that of procedure B3, but this identification method is not the one originally proposed. Each procedure has its advantages, so the user may chose the model depending on the necessity and the accepted tradeoffs. Remember that procedures B1, B2 and B3 assume there is oxygen in the influent, which is not normally the case, but they fit better. Procedures A3 and B3 do not follow the procedure agreed in Mexico, but again, they fit better than A1, A2 and B1, B2

respectively. Models using the k_1 found experimentally (A1, A3, B1, B3) will explain the long term biomass growth, the others will not (remember that the growth in one standard reaction is small).

It is important to clarify the meaning of the given Mean Quadratic Error (MQE). 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 as calculated in file MQE_interpretation.xls and shown in Table 4.

Table 4. MQE as function of error for ideal case of all errors being equal

error%	MQE
5	42
10	167
11	202
12	240
14	327
16	427
18	540
20	667
21	735
22	807
24	960
25	1042

Of course, in practice, not all samples do have the same error. So it must be kept in mind that errors contribute in a quadratic (not linear) way to the total error.

Results and graphs for procedure A1 can be seen in files:
(ID1) 3121909 Std FTC 100mgL 1\Fit 121909A1.doc
(ID2) 3121913 Std FTC 100mgL 2\Fit 121913A1.doc
Results and graphs for procedure A2 can be seen in files:
(ID1) 3121909 Std FTC 100mgL 1\Fit 121909A2.doc
(ID2) 3121913 Std FTC 100mgL 2\Fit 121913A2.doc
Results and graphs for procedure A3 can be seen in files:
(ID1) 3121909 Std FTC 100mgL 1\Fit 121909A3.doc
(ID2) 3121913 Std FTC 100mgL 2\Fit 121913A3.doc

Results and graphs for procedure B1 can be seen in files:
(ID1) 3121909 Std FTC 100mgL 1\Fit 121909B1.doc
(ID2) 3121913 Std FTC 100mgL 2\Fit 121913B1.doc
Results and graphs for procedure B2 can be seen in files:
(ID1) 3121909 Std FTC 100mgL 1\Fit 121909B2.doc
(ID2) 3121913 Std FTC 100mgL 2\Fit 121913B2.doc
Results and graphs for procedure B3 can be seen in files:
(ID1) 3121909 Std FTC 100mgL 1\Fit 121909B3.doc
(ID2) 3121913 Std FTC 100mgL 2\Fit 121913B3.doc

6. VALIDATION

The mean value of each parameter was found for each procedure, thus giving 6 possible EM1 (Eoli Model 1). Then each EM1 version was applied in simulation to all experiments, that is, to the validation experiment (VAL1) and also to the identification experiments (ID1 and ID2). Results are shown in table 5.

Table 5. Validation results for all 6 EM1 model options

EM1 (procedure)	Experiment	MQE	Mean MQO	
			by group	Global
A1	ID1	961	804	707
	ID2	647		
	VAL1	514	514	
A2	ID1	929	788	701
	ID2	647		
	VAL1	527	527	
A3	ID1	446	391.5	430
	ID2	337		
	VAL1	506	506	
B1	ID1	515	389.5	343
	ID2	264		
	VAL1	251	251	
B2	ID1	516	383	342
	ID2	250		
	VAL1	259	259	
B3	ID1	231	186	244
	ID2	141		
	VAL1	360	360	

Validation criterion must be chosen according to the end use of the model. Independently of that choice, it is clear that global error is less for the “B” type procedures, and also for the “3” type procedures.

If UNAM experimental team states that no Oxygen was present in the inflow, then all “B” models should be rejected. But after reviewing the method for preparing the synthetic wastewater WP1 accepted that most of the times it was Oxygen saturated, even if initially we did think it was not.

If adhesion to the separation of identification of kinetic and yield coefficients is maintained, all “3” type procedures should be also rejected.

All models exhibited a good behavior for the validating experiment, in the sense that the MQE of the validation rests of the same order, or even lower, as for identifications.

If limits for accepting a model as valid are chosen considering information in Table 4 as an indicative, then the validation results positive for all models for a chosen limit of MQE<1042 (25%), but models A1 and A2 are rejected if the limit is chosen at MQE<540 (18%)

Graphs and data resulting from validation tests can be found in the following files:

(ID1) 3121909 Std FTC 100mgL 1\Test 121909A1.doc
(ID2) 3121913 Std FTC 100mgL 2\Test 121913A1.doc
(VAL1) 3121917 Std FTC 100mgL 3\Test 121917A1.doc

(ID1) 3121909 Std FTC 100mgL 1\Test 121909A2.doc
(ID2) 3121913 Std FTC 100mgL 2\Test 121913A2.doc
(VAL1) 3121917 Std FTC 100mgL 3\Test 121917A2.doc

(ID1) 3121909 Std FTC 100mgL 1\Test 121909A3.doc
(ID2) 3121913 Std FTC 100mgL 2\Test 121913A3.doc
(VAL1) 3121917 Std FTC 100mgL 3\Test 121917A3.doc

(ID1) 3121909 Std FTC 100mgL 1\Test 121909B1.doc
(ID2) 3121913 Std FTC 100mgL 2\Test 121913B1.doc
(VAL1) 3121917 Std FTC 100mgL 3\Test 121917B1.doc

(ID1) 3121909 Std FTC 100mgL 1\Fit 121909B2.doc
(ID2) 3121913 Std FTC 100mgL 2\Test 121913B2.doc
(VAL1) 3121917 Std FTC 100mgL 3\Test 121917B2.doc

(ID1) 3121909 Std FTC 100mgL 1\Test 121909B3.doc
(ID2) 3121913 Std FTC 100mgL 2\Test 121913B3.doc
(VAL1) 3121917 Std FTC 100mgL 3/Test 121917B3.doc

Additional verifications to test the hypothesis that different biomasses do have different yield/kinetical properties, even if they come from the same but uncontrolled source, were made. The 6 model candidates were applied to one other experiment (ID3) from the starvation cycle. This experiment was made before the starvation, so there is no direct effect of starvation in the kinetic. Nevertheless, the biomass used for ID3 is not the same than the one used for ID1, ID2 and VAL1. Remember that the biomass is a consortium, taken from a municipal WWTP and acclimated to the specific toxicant. Results are shown in Table 6. The results in Table 6 indicate that there is no fit at all of any of the EM1 options for such experiment. However, the model found in previous WP2 (model) Mid-Term P6 (unam)1.4.doc work does fit ID3 (and ID4).

Table 6. Test of EM1 options with experiment using different biomass

EM1 (procedure)	Experiment	MQE
A1	ID3	1933
A2	ID3	1986
A3	ID3	2283
B1	ID3	2032
B2	ID3	2109
B3	ID3	2365

Data and Graphical results may be seen in files:

(ID3) 3070218 stv200bef12\Test 070218A1.doc
(ID3) 3070218 stv200bef12\Test 070218A2.doc
(ID3) 3070218 stv200bef12\Test 070218A3.doc
(ID3) 3070218 stv200bef12\Test 070218B1.doc
(ID3) 3070218 stv200bef12\Test 070218B2.doc
(ID3) 3070218 stv200bef12\Test 070218B3.doc

7. CONCLUSIONS

1. Six different identification procedure combinations were applied to "Standard FTC" experimental data, producing six options for EM1 (Eoli Model 1), namely A1, A2, A3, B1, B2 and B3. Models of the "A" type assume there is no oxygen in the inflow. "B" models assume there is. Models of types "1" and "2" identify the yield parameters independently of kinetic ones. Type "3" identifies k_1 independently, but k_2 is identified simultaneously with the kinetic ones.

2. Substrate kinetics, and associated finishing reaction time, were satisfactorily modeled in FTC mode by all 6 model candidates.

3. Better mean results to merge various $\mu_n(S)$ into one single $\mu(S)$ to model a set of experiments (for n=2 experiments) were obtained using S-parameters (μ^*, S^*, Sm) instead of using K-parameters (μ_o, K_I, K_S)

4. Confidence intervals are good for the kinetic parameters if represented as S-parameters (worst case 7% for A1; best case 0.4% in A3) while large if represented as K-parameters (worst case 89% for A1; best case 0.5% in A3)

5. "A" type procedures were not able to reproduce with the same closeness, and simultaneously, the D.O. behavior during fill-reaction phase and full-reaction phase. To solve this undesirable feature, "B" procedures assuming there is D.O. in the influent were proposed. "B" type procedures exhibited better fit than "A" type procedures. Experimental data in new WP1 experiments later confirmed the validity of such approach.

6. "1 and 2" type procedures conform to the principle of separating the identification of Yield coefficients from Kinetic coefficients. A different "3" type procedure was proposed that estimates k_1 from complementary experimental data, and then identifies k_2 along with the rest of the Kinetic coefficients. "3" type procedures exhibited better fit than "1 and 2" type procedures for the same identification conditions.

7. Both yield and kinetic coefficients are biomass dependant. As biomass is an unknown consortium, which might even evolve in time, and is taken from an uncontrolled environment (municipal WWTP) it seems not possible to use a single set of parameters to represent the SBR at all times. The parameters produced in this work are then valid only for the biomass in the conditions exhibited at the time of the experiments used.

8. All model assumptions in Annex 6 seem to hold.

9. Future work may address the variability of parameters for different biomasses.

10. If considered necessary, more work may be addressed also to explain the apparent mismatch of the model's D.O. between the fill-reaction phase and the full-reaction phases.

11. Future work may include the use of ED-TOC kinetics for identification/validation, as long as they are produced with the same biomass as the FTC to which they are going to relate.

Annex 1. Experimental files names and classification

The tables below show the selection of files for identification and (ID) validation (VAL) for identification of standard model EM1A.

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

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

The file-directory-code is assigned as "ammddhh" 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

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

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)

Annex 2. Precision of measurements (extract of UNAM table "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 coefficient $k_1 = 1/Y_{x/s}$

- 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:

$\mu=\mu(S,O)$	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 $k_la(V)= K_la*V_{max}/V$
K_La	h^{-1}	Value for k_la when Volume is maximum as $K_la = k_la(V_{max})$
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:

$$\mu = \frac{\mu_0 S}{K_s + S + K_i S^2} \qquad \text{Eq. (1)}$$

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 EM1_Matlab_library\StoK.m and EM1_Matlab_library\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.

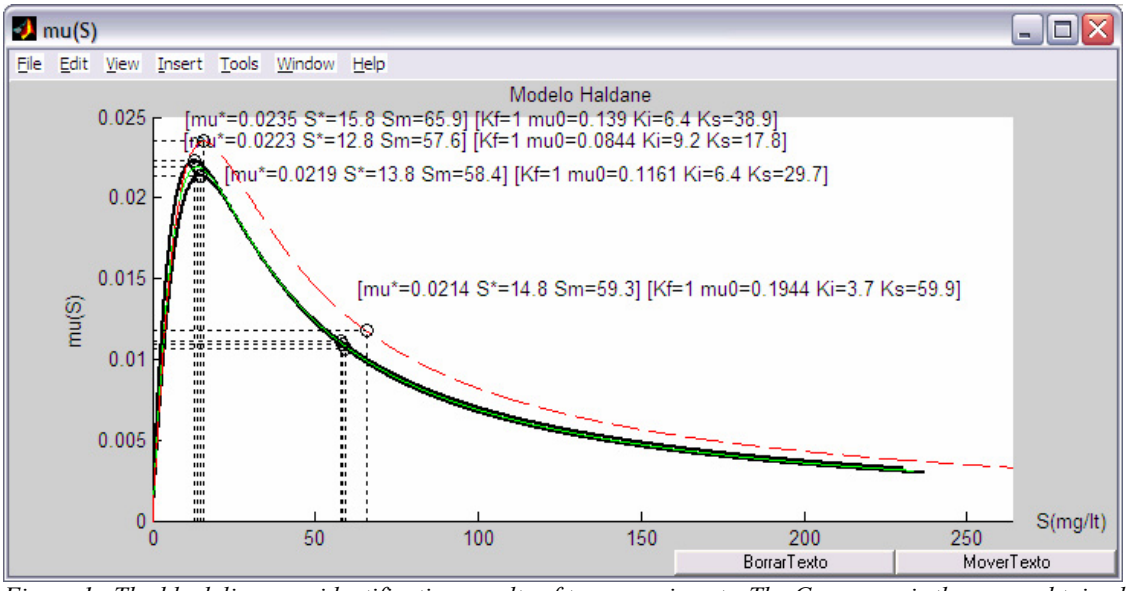

Figure 1. The black lines are identification results of two experiments. The Green one is the mean obtained by means of doing the mean of each S-parameter (mue, Se, Sm). The red-dashed one is the mean obtained by doing the mean of each traditional K-parameter (mu0, K_I, K_S).

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

Assumption 13 is changed by suggestion of project leader. No objections to such a change were expressed by the other experts involved. 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.

Assumption 2 is put to test as the shape of the oxygen curve while filling suggest there is more oxygen in the reactor than that explained by the model with its actual structure. Oxygen in the influent or different K_La while filling may explain such behavior. Confirmation from the experimental team about this assumption is asked. Meanwhile, procedure B assumes there is oxygen in the influent, while procedure A assumes there is not. See Tables 1 and 2 (main document) for an explanation of procedures.

Assumptions

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.