Identification of analytical redundancy relations design for sensor fault diagnosis of a wastewater treatment station

Yvon Tharrault, Gilles Mourot, José Ragot, David Fiorelli, Serge Gillé

IDENTIFICATION OF ANALYTICAL REDUNDANCY RELATIONS DESIGN FOR SENSOR FAULT DIAGNOSIS OF A WASTEWATER TREATMENT STATION

Yvon Tharrault * **Gilles Mourot** * **José Ragot** *
David Fiorelli ** **Serge Gillé** **

* *Centre de Recherche en Automatique de Nancy*
UMR 7039 Nancy-Université CNRS
2, avenue de la Forêt de Haye, 5416
Vandoeuvre-lès-Nancy, France

** *Centre de Recherche Public Henri Tudor*
LTI, Laboratoire des Technologies Industrielles, 70, route de Luxembourg, L - 4009 Esch-Alzette

Abstract: This article deals with identification design for sensor fault diagnosis of a wastewater treatment station. First, we established a physical model of this station by simplifying and linearizing the model ASM1 (Henze *et al.*, 1987). Then, we study the measurement values in order to find other redundancy equations. Multiple linear regression allowed the determination of a static black box type model, coherent with the physical model. The addition of temporal lags between the signals makes possible to improve the performances of this model. We show finally the use of this model for detection-localization of faults measurements.

Keywords: Diagnostic, ASM1, Data validation, Identification.

1. INTRODUCTION

In the case of an activated sludge wastewater treatment station, the complexity of the biological mechanisms (non-linear kinetics, parameters varying in time, ...) imposes the development and the use of advanced techniques of automatic control to develop a strategy of real time fault diagnosis. In our work, we use the model-based diagnosis methods based on analytical redundancy, that correpond to execute a test of coherence between measurement and the estimate provided by a model.

In a first part, we present the wastewater treatment station of Bleesbrück. To highlight the existing relations between the variables characterizing the activated sludge process of the biology, the second part of the talk is devoted to the application of the Activated Sludge Model 1 (ASM1) (Henze *et al.*, 1987) to this station. Initially, we seek a linear model which depends on the available measurements; it was then necessary to reduce the complexity and to linearize the ASM1 model. Next we search the relations between the variables of the station by multiple linear regression. This analysis allows us, at the same time, to check in the data the relations of the second part and to underscore the redundancy relations used to determine a black box type model. We show then that this approach is useful for fault localization. In the last part, we conclude on the current results of the study.

NOMENCLATURE ASM1

S_I	Concentration of soluble inert organic matter
S_S	Concentration of readily biodegradable substrate
S_{O_2}	Concentration of dissolved oxygen
$S_{O_2}^{sat}$	Dissolved oxygen saturation concentration
S_{NO}	Concentration of nitrate and nitrite nitrogen
S_{NH}	Concentration of ammonia nitrogen
S_{ND}	Concentration of soluble biodegradable organic nitrogen
S_{ALK}	Molar concentration of alkalinity
X_I	Concentration of particulate inert organic matter
X_S	Concentration of slowly biodegradable substrate
X_H	Concentration of active heterotrophic biomass
X_A	Concentration of active autotrophic biomass
X_{ND}	Concentration of particulate biodegradable organic nitrogen
X_P	Concentration of particulate products from biomass decay
b_A	Decay rate coefficient for autotrophic organisms
b_H	Decay rate coefficient for heterotrophic organisms
f_{XI}	Fraction of biomass generating the particulate products
i_{NBM}	Mass of nitrogen in the biomass
i_{NXI}	Mass of nitrogen in the inert particulate organic matter
k_a	Ammonification rate
k_h	Maximum specific hydrolysis rate
k_La	Coefficient of oxygen rate
$K_{(.)}$	Half-saturation coefficient:
K_{NH}	of ammonia for autotrophs
K_{NO}	of nitrate for denitrifying heterotrophs
$K_{O_2,A}$	of oxygen for autotrophs
$K_{O_2,H}$	of oxygen for heterotrophs
K_S	for heterotrophic organisms
K_X	for hydrolysis of slowly biodegradable substrate
Y_A	Yield coefficient for autotrophic organisms
Y_H	Yield coefficient for heterotrophic organisms
μ_A	Maximum specific growth rate for autotrophic organisms
μ_H	Maximum specific growth rate for heterotrophic organisms
$\eta_{NO_{3g}}$	Correction factor for anoxic growth of heterotrophs
$\eta_{NO_{3h}}$	Correction factor for anoxic hydrolysis

2. BLEESBRÜCK STATION

The station has a maximum capacity of treatment of 100 000 equivalent inhabitants and it treats wastewater of urban and industrial origins. It is located near the Sûre, river which collects purified water.

2.1 Description of operation

In the purification procedure of water, we can distinguish three phases, the pretreatment, the primary treatment and the secondary treatment. The pretreatment operation consists in the elimination of all the components which could block or cause an abrasion of the conduits and pumps or disturb the treatments of decontamination. The primary treatment is composed of a rectangular aeration tank called the first biology which is an activated sludge biology and two settling tanks. The sludges in the settling tanks are recirculated in the biology. The goal of this biology is to eliminate mainly organic pollution. Phosphorus is eliminated chemically by the addition of a product. The secondary treatment is composed of a second biology split into two aeration tanks (Bio 2.1 and Bio 2.2) and two settling tanks. The aim of this treatment is to eliminate ammonium by nitrification. An overflow, limits the input flow of the second biology.

N°	Position	Name	Measure type
1	Before biology 1	Q_{in}	Flow
2		T	Temperature
3		σ	Conductivity
4		A	UV absorption
5		PH	PH
6	In biology 1	$MES1$	Suspended matter
7		$O1$	Dissolved oxygen
8	After biology 1	$NH1$	Ammonium
9		$NO1$	Nitrate
10		$OP1$	Orthophosphate
11	Before biology 2	H	Height of wastewater before the overflow
12	In biology 2	$O21$	Dissolved oxygen
13		$MES21$	Suspended matter
14		$O22$	Dissolved oxygen
15		$MES22$	Suspended matter
16	After biology 2	$NH2$	Ammonium
17		$NO2$	Nitrate
18		$OP2$	Orthophosphate

Table 1. Available sensors

2.2 Available sensors

For this application, we seek to validate the data of the sensors detailed in the table 1.

3. MODEL OF KNOWLEDGE

With an aim of determining the relations between the variables and their structures, we model a system made up of a biology (activated sludge process) and of a settling tank.
In the literature, there are various types of models which describe the mechanisms of biological degradation in wastewater treatment stations, as the ASM1, the AMS2, ASM2d, and the ASM3 (Gernaey *et al.*, 2004). In our case we will use ASM1, which describes the biological mechanisms of degradation of the organic matter, nitrification and denitrification.

The objective is to simplify model ASM1 in order to as much as possible remove the unmeasured variables. For that we will use assumptions generally allowed in the literature and validated by the practice.

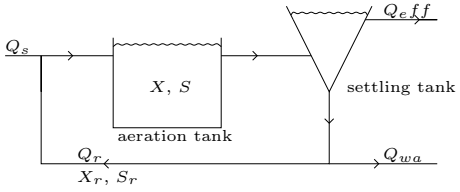

Fig. 1. Aeration tank and settling tank

The figure 1 presents the biology and the settling tank. We call Q_s the flow entering the station, Q_r the flow of recirculation of sludges, Q_{wa} the flow of extraction of sludges in excess and Q_{eff} the

flow on the outlet side of the settling tank. Next, we consider, because of absence of measurement in the settling tank, that this one is perfect, *i.e.* no sludge leaves by the overflow the settling tank. This assumption leads to the two following equations:

$$S = S^r \tag{1}$$

$$(Q_s + Q_r)X = (Q_r + Q_{wa})X^r \tag{2}$$

where X and S are the respective concentrations of particle and substrate in the aeration tank, and X^r and S^r the concentrations of particle and substrate in the circuit of sludge recirculation.

3.1 ASM1 Modeling

In this study two traditional simplifications are applied to the original model (Julien, 1997) (the alkalinity S_{ALK} is omitted and the particulate compound interns X_I and X_P are not distinguished). Then ASM1 model includes 11 variables of states. It is a general model, which can be used to model the first or the second biology.
The figure 1 represents the biological engine. The equations of mass balances are written:

$$\dot{\xi}_\alpha = D_s\xi_\alpha^{in} + D_r\xi_\alpha^r - (D_s + D_r)\,\xi_\alpha + \mathcal{R}_\alpha \tag{3}$$

where $D_s = \dfrac{Q_s}{V}$, $D_r = \dfrac{Q_r}{V}$, V is the volume of the aeration tank and $\mathcal{R}_\alpha$ correspond to the degradation of the compound α and ξ_α^{in}, ξ_α^R, ξ_α indicate the concentration of the compound α in the influential, the recycling and the aeration tank, *i.e.*:

$$\xi = [S_I, S_S, S_{NO}, S_{NH}, S_{ND}, S_{O_2}, X_I, X_S, X_H, X_A, X_{ND}]^T \tag{4}$$

The various values of $\mathcal{R}_\alpha$ are explicited in table 2 and table 3. The parameters Y_A, Y_H, f_{XI}, i_{NBM}, i_{NXI}, $\eta_{NO_{3g}}$, $\eta_{NO_{3h}}$, k_a, k_h as well as the terms $K_{..}$ are constant. The values of these constants are given for a temperature of 10°C and of 20°C (Henze *et al.*, 1987). The other parameters vary according to time. The $k_La\left(S_{O_2}^{sat} - S_{O_2}\right)$ term corresponds to the supplying of air in biology, *i.e.* with the operation of the aerators of biology. The terms in ρ_j with j=1...8, correspond to the kinetics of the degradation model and are explicit in table 3.

α	$\mathcal{R}_\alpha$
1	0
2	$-\dfrac{1}{Y_H}(\rho_1 + \rho_2) + \rho_7$
3	$-\dfrac{1 - Y_H}{2.86Y_H}\rho_2 + \dfrac{1}{Y_A}\rho_3$
4	$-i_{NBM}(\rho_1 + \rho_2) - \left(i_{NBM} + \dfrac{1}{Y_A}\right)\rho_3 + \rho_6$
5	$-\rho_6 + \rho_8$
6	$k_La\left(S_{O_2}^{sat} - S_{O_2}\right) - \dfrac{1 - Y_H}{Y_H}\rho_1 - \dfrac{4.57 - Y_A}{Y_A}\rho_3$
7	$f_{XI}(\rho_4 + \rho_5)$
8	$(1 - f_{XI})(\rho_4 + \rho_5) - \rho_7$
9	$\rho_1 + \rho_2 - \rho_4$
10	$\rho_3 - \rho_5$
11	$(i_{NBM} - f_{XI}i_{NXI})(\rho_4 + \rho_5) - \rho_8$

Table 2. Stoichiometric parameters of the degradation model

To be able to apply model ASM1 to our situation, it should be reduced for taking into account only the measured variables.

3.2 Reduction of model ASM1

The simplifying assumptions used are inspired by (Chachuat, 2001) and (Gomez-Quintero, 2002):

- *Absence of anoxic growth of heterotrophs.* Indeed, the wastewater treatment station does not have a period of significant anoxia, therefore the term ρ_2 is removed from the various values of degradation $\mathcal{R}_\alpha$ (table 2).
- *Independence of S_I.* The variations of the inert soluble organic matter concentration (S_I) do not have any influence on the others composed and conversely, we then omit this compound in the model, by removing the state variable of the equation (3) with $\alpha = 1$.
- *Simplification of model dynamic.* Theory of the singular disturbances makes possible to consider that X_I, X_A and X_H have slow dynamic. Thus these variables are considered constant over a few hours. Then the corresponding states variables are removed from the equation (3) corresponding to $\alpha = 7, 9$ and 10.
- *Simplification of the organic compounds.* The measurement of the chemical oxygen demand (COD), does not make possible to distinguish the soluble part (S_S) and the particulate part (X_S) (Smets *et al.*, 2003). A single organic compound then is considered: $X_{COD} = S_S + X_S$
- *Simplification of the nitrogenized compounds*, Model ASM1 distinguishes four nitrogenized fractions S_{NO}, S_{NH}, S_{ND} and X_{ND}. The process of ammonification of soluble organic nitrogen (kinetic ρ_6) is assumed constant. The concentration S_{ND} being constant, the modeling of X_{ND} is then useless, because this concentration is only used in the estimate of S_{ND}. With this assumption, the ammonium internal formation is also condidered constant.

In the station, the flow of recirculation is higher than the flow of extraction of sludges, one supposes that $\frac{D_{wa}}{D_r + D_{wa}} << 1$. Thus the X_S term is neglected from the equation $X_{COD} = S_S + X_S$. Next one defines the soluble fraction of the organic compounds by $f_{SS} = \frac{S_S}{X_{COD}}$, X_{COD} can be written:

j	Process	Kinetic, ρ_j
1	Aerobic growth of heterotrophs	$\mu_H \frac{S_S}{S_S+K_S}\frac{S_{O_2}}{S_{O_2}+K_{O_2,H}}\frac{S_{NH}}{S_{NH}+K_{NH,H}}X_H$
2	Anoxic growth of heterotrophs	$\mu_H \frac{S_S}{S_S+K_S}\frac{K_{O_2,H}}{S_{O_2}+K_{O_2,H}}\frac{S_{NH}}{S_{NH}+K_{NH,H}}\frac{S_{NO}}{S_{NO}+K_{NO}}\eta_{NO_{3g}}X_H$
3	Aerobic growth of autotrophs	$\mu_A \frac{S_{O_2}}{S_{O_2}+K_{O_2,A}}\frac{S_{NH}}{S_{NH}+K_{NH,A}}X_A$
4	'Decay' of heterotrophs	$b_H X_H$
5	'Decay' of autotrophs	$b_A X_A$
6	Ammonification of soluble organic nitrogen	$k_a X_H S_{ND}$
7	'Hydrolysis' of entrapped organics	$k_h \frac{X_S}{K_X+X_S/X_H}\left[\frac{S_{O_2}}{S_{O_2}+K_{O_2,H}}+\eta_{NO_{3h}}\frac{K_{O_2,H}}{S_{O_2}+K_{O_2,H}}\frac{S_{NO}}{S_{NO}+K_{NO}}\right]$
8	'Hydrolysis' of entrapped organic nitrogen	$k_h \frac{X_{ND}}{K_X+X_S/X_H}\left[\frac{S_{O_2}}{S_{O_2}+K_{O_2,H}}+\eta_{NO_{3h}}\frac{K_{O_2,H}}{S_{O_2}+K_{O_2,H}}\frac{S_{NO}}{S_{NO}+K_{NO}}\right]$

Table 3. Kinetic of the degradation model

$$\dot{X}_{COD} = D_s\left(X^{in}_{COD} - f_{SS}X_{COD}\right) - \frac{1}{Y_H}\rho_1 + (1-f_{XI})(\rho_4+\rho_5) \quad (5)$$

It is also supposed that the concentration, at the input of the station, of dissolved oxygen is negligible in front of the concentration in the tank ($S^{in}_{O_2}$=0). Then, starting from the equations (3, 5), the following reduced model is as follow:

$$\begin{cases} \dot{X}_{COD} = D_s\left(X^{in}_{COD} - \frac{K_S}{K_{COD}}X_{COD}\right) - \frac{1}{Y_H}\theta_1 \\ \qquad \frac{X_{COD}}{K_{COD}+X_{COD}}\frac{S_{O_2}}{K_{O_2,H}+S_{O_2}}\frac{S_{NH}}{S_{NH}+K_{NH,H}} + \theta_2 \\ \dot{S}_{NO} = D_s\left(S^{in}_{NO} - S_{NO}\right) + \frac{\theta_3}{Y_A}\frac{S_{NH}}{K_{NH,A}+S_{NH}}\frac{S_{O_2}}{K_{O_2,A}+S_{O_2}} \\ \dot{S}_{NH} = D_s\left(S^{in}_{NH} - S_{NH}\right) - i_{NBM}\theta_1\frac{X_{COD}}{K_{COD}+X_{COD}} \\ \qquad \frac{S_{O_2}}{K_{O_2,H}+S_{O_2}}\frac{S_{NH}}{S_{NH}+K_{NH,H}} - \left(i_{NBM}+\frac{1}{Y_A}\right)\theta_3 \\ \qquad \frac{S_{NH}}{K_{NH,A}+S_{NH}}\frac{S_{O_2}}{K_{O_2,A}+S_{O_2}} + \theta_4 \\ \dot{S}_{O_2} = -D_S S_{O_2} - \frac{1-Y_H}{Y_H}\theta_1\frac{X_{COD}}{K_{COD}+X_{COD}} \\ \qquad \frac{S_{O_2}}{K_{O_2,H}+S_{O_2}}\frac{S_{NH}}{S_{NH}+K_{NH,H}} - \frac{4,57-Y_A}{Y_A}\theta_3 \\ \qquad \frac{S_{NH}}{K_{NH,A}+S_{NH}}\frac{S_{O_2}}{K_{O_2,A}+S_{O_2}} + k_La\left(S^{sat}_{O_2} - S_{O_2}\right) \end{cases} \quad (6)$$

with constants:

$$\begin{cases} D_S = \frac{Q_S}{V} \\ \theta_1 = \mu_H X_H \\ \theta_2 = (1-f_{XI})(b_H X_H + b_A X_A) \\ \theta_3 = \mu_A X_A \\ \theta_4 = k_a X_H S_{ND} \end{cases} \quad (7)$$

Hence the model obtained depends only on nine variables: D_S, X^{in}_{COD}, X_{COD}, S_{NO}, S^{in}_{NO}, S^{in}_{NH}, S_{NH}, S_{O_2} and of the aerators working. For simplicity, one tries now to find linear relations.

3.3 Linearization of the reduced model

We linearize the model (6) with the following considerations:

- In our case, biology 1 does not have an anoxia period, then it is supposed that the dissolved oxygen concentration is important, i.e. $S_{O_2} >> K_{O_2,H}$ and thus $\frac{S_{O_2}}{S_{O_2}+K_{O_2,H}} \approx 1$; one uses the same assumption with $K_{O_2,A}$, thus $\frac{S_{O_2}}{S_{O_2}+K_{O_2,A}} \approx 1$.
- To linearize the equations (6) compared to the state variables (X_{COD}, S_{NH} et S_{NO}), one simplifies:
  - $\frac{X_{COD}}{K_{COD}+X_{COD}} \rightarrow a_{COD}X_{COD} + b_{COD}$ where a_{COD} and b_{COD} are constants.
  - $\frac{S_{NH}}{K_{NH}+S_{NH}} \rightarrow a_{NH}S_{NH} + b_{NH}$ where a_{NH} and b_{NH} are constants.
  - $\frac{S_{NO}}{K_{NO}+S_{NO}} \rightarrow a_{NO}S_{NO} + b_{NO}$ where a_{NO} and b_{NO} are constants.
- Moreover, the products of state variables are also replaced by terms of type $S_xS_y = \bar{S}_yS_x + \bar{S}_xS_y$ with $\bar{S}_x$ and $\bar{S}_y$ constant average values of the variables S_x and S_y.

Then one finds a linear system with variable parameters, however the dilution ratio relating to the input flow ($D_S(t)$) and the operation of the aerators ($k_La(t)\left(S^{sat}_{O_2} - S_{O_2}(t)\right)$) depend on time.

Discretization of the linear model from a numerical point of view gives with a sampling step of Δ:

$$\begin{cases} X_{COD}(k+1) = \Delta\left(D_S(k)X^{in}_{COD}(k) \right. \\ \qquad \left. + A_{COD}(k)X_{COD}(k) + B_{NH}S_{NH}(k) + k_1\right) \\ S_{NO}(k+1) = \Delta\left(D_S(k)S^{in}_{NO}(k) + C_{NO}(k)S_{NO}(k) \right. \\ \qquad \left. + D_{NH}S_{NH}(k) + k_2\right) \\ S_{NH}(k+1) = \Delta\left(D_S(k)(k)S^{in}_{NH}(k) \right. \\ \qquad \left. + E_{NH}(k)S_{NH}(k) + F_{COD}X_{COD}(k) + k_3\right) \\ S_{O_2}(k+1) = \Delta\left(G_{O_2}(k)S_{O_2}(k) + H_{NH}S_{NH}(k) \right. \\ \qquad \left. + I_{COD}X_{COD}(k) + k_La(k)S^{sat}_{O_2} + k_4\right) \end{cases} \quad (8)$$

where parameters $A_{COD}(k)$, B_{NH}, $C_{NO}(k)$, D_{NH}, $E_{NH}(k)$, F_{COD}, $G_{O_2}(k)$, H_{NH}, I_{COD}, k_1, k_2, k_3 and k_4 are identified according to the linear model, for example $G_{O_2}(k) = \frac{1}{\Delta} - D_S(k) - k_La(k)$.

Explained var.	Corr.	Explanatory var.
1	0.925	11 55
2	0.773	8 3 7 4 55
3	0.658	2 8 4 55
4	0.523	2 8 3 55
5	×	
6	0.360	11 1 55
7	0.798	2 8 9 16 14 55
8	0.882	2 10 7 16 55
9	0.825	7 8 16 10 55
10	0.763	8 16 9 2 55
11	0.925	1 55
12	0.973	14 16 55
13	0.976	15 55
14	0.973	12 7 55
15	0.976	13 58
16	0.852	18 8 9 7 12 10 55
17	0.258	1 55
18	0.668	16 55

Table 4. Explanatory variables in the static case

Explained var.	Corr.	Explanatory var.
1	0.929	11 19 55
2	0.926	20 8 26 3 21 55
3	0.803	21 2 20 8 4 55
4	0.590	2 20 22 8 3 55
5	0.028	25 55
6	0.955	24 19 29 10 55
7	0.780	25 20 26 2 55
8	0.987	26 44 20 2 55
9	0.958	27 25 7 26 8 55
10	0.944	28 2 8 55
11	0.934	1 19 29 55
12	0.982	14 30 32 55
13	0.992	15 31 33 55
14	0.987	12 32 30 55
15	0.991	13 33 31 55
16	0.996	34 52 44 26 55
17	0.986	35 53 55
18	0.997	36 54 52 34 16 55

Table 5. Explanatory variables in the dynamic case

To extract the relationship between the variables of biologies and obtaining redundancy relations connecting the variables of the two biology sets, we will now used multiple linear regression.

4. BLACK BOX MODEL

To build a black box model by multiple linear regression, we choose a period where the water temperature is varying a bit thus allowing to free itself from this disturbance. The period include variations of polluant because of the week-end (close of industries).

4.1 Static model

Initially, explained variables (table 4) are determined by multiple linear regression by adding the most explanatory variables in first (variable 55 corresponds to a constant). If one calls $y_i(t)$ the explained variable, $u_j(t)$ the explanatory variables and θ_j the parameters associated with the explanatory variables, then one finds a linear model of the form:

$$y_i(t) = \sum_{j=1}^{p} \theta_j \, u_j(t) \tag{9}$$

where p represent the number of explanatory variable.
The estimates $\hat{\theta}_j$ of the parameters and $\hat{y}_i(t)$ of the measurement are determined by minimization of the quadratic criterion corresponding to the sum of the squares of the differences between measurement and estimate.

The table 4 specifies the coefficient of correlation of each estimated variable $\hat{y}_i(t)$ towards its measurement $y_i(t)$. Let us note that these calculations were carried out on a data set not used for the identification.

One notices that in the case of the measurement of the pH (5), it is not possible to build a model, because of weak correlations with the other variables. Whereas the flow (1) is explained only from the height measure of the overflow (11) and vice-versa, which translates a strong correlation between these two measurements.

4.2 Dynamic model

By putting back certain parametres, one can improve the quality of modeling. One considers, moreover, the same delay of 120 points (signal number from 19 to 36) and 240 points (signal number from 37 to 54), *i.e.* a lag of 1h and 2h, which corresponds roughly to the residence times of the water used in the first and the second biology. The table 5 presents the same information as in the table 4 but in the dynamic case. The constructed models are, when the redundancy are near direct (like parameters 12 and 14 or 13 and 15), good. The other model must be improved.

The determined model for NO1 is then:

$$\begin{aligned} NO1(k) = {} & 0.38\, NO1(k-120) \\ & -0.03\, O1(k-120) \\ & -0.01\, O1(k) \\ & +0.02\, NH1(k-120) \\ & +0.75\, NH1(k) + 0.23 \end{aligned} \tag{10}$$

The figure 2 represents the estimate of the concentration of NO1 with the static model and the difference between its estimate and its measurement, this difference is named residual. The figure 3 shows the estimate and the residual of this same variable but with the dynamic model. These two estimates are carried out on validation data, one notices an better estimation in the dynamic case. For the residual determined with the dynamic model, it is necessary to adapt the test of fault detection according to the nature and the amplitude of the defect.

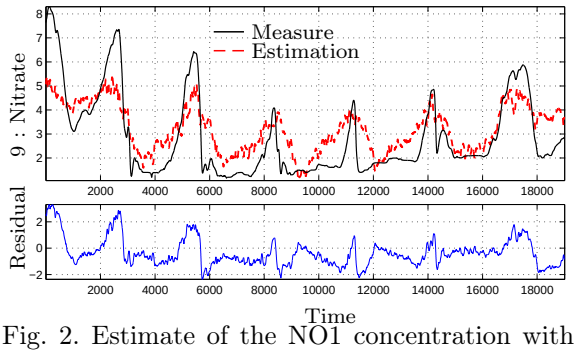

Fig. 2. Estimate of the NO1 concentration with the static model

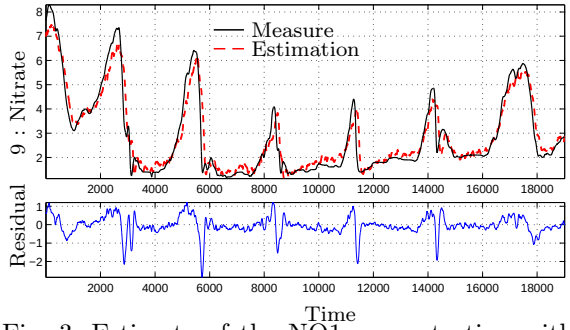

Fig. 3. Estimate of the NO1 concentration with the dynamic model

4.3 Use of the models for the diagnosis

The construction of the incidence matrix of the various variables for the dynamic models makes possible to highlight the ability of locating the sensors faults. Indeed, this matrix shows the relationship between two classes of objects. If we consider two classes, X for the residual and Y for the parameters , the matrix has one row for each element of X and one column for each element of Y. The entry in row x and column y is 1 if x and y are related (called incident in this context) and 0 if they are not.

For example, if a fault appears on the temperature measurement (2), then the tests of fault detection of residuals 3, 4, 7, 8 and 10 must be positive, one can thus locate the defect. This table also makes it possible to highlight impossibility of dissociating a defect on the sensor MES21 (13) and on the sensor MES22 (15), because in both cases the tests of fault detection of residuals 13 and 15 must be positive, it is then necessary to have additional information to locate the defect. One is in the same case to locate a defect appearing on the sensor O21 (12) and that P22 (14).

5. CONCLUSION

To build models for fault diagnosis, the first stage is the comprehension of the system in order to release the relationship between the measured variables. In optics to highlight linear redundancy relations between the variables, we used ASM1 model. But not having measurements of all the variables of this model, it should initially have been reduced by assuming realistic and commonly accepted assumptions. To determie the simplest possible relations, we linearized it. We thus determined the structure of the relations between the variables.

After having found the structure of the relations between the variables describing the operation of biologies, we were then interested in the determination of the redundancy relations between the variables of the station, by multiple linear regression. This regression allowed, by adding the most explanatory variables successively, to build black box models. It is however necessary to take into account, in modeling, the need for locating a fault when it is detected. The last part of this article approaches this problem.

ACKNOWLEDGEMENTS

The authors gratefully thank the financial support from the Ministry of Culture, Higher Education and Research of the Luxembourg government.

REFERENCES

Chachuat, B. (2001). Méthodologie d'optimisation dynamique et de commande optimale des petites stations d'épuration à boues activées. PhD thesis. Institut National Polytechnique de Lorraine.

Gernaey, K. V., M. C. M. van Loosdrecht, M. Henze, M. Lind and S. B. Jorgensen (2004). Activated sludge wastewater treatment plant modelling and simulation : state of the art. *Environmental Modelling & Software* **19**(9), 763–783.

Gomez-Quintero, C. (2002). Modélisation et estimation robuste pour un procédé boues activées en alternance de phases. PhD thesis. Université Toulouse III.

Henze, M., C. P. L. Grady Jr., W. Gujer, G. v. R. Marais and T. Matsuo (1987). Activated Sludge Model No. 1. In: *IAWPRC Scientific and Technical Reports No. 1.*

Julien, S. (1997). Modélisation et estimation pour le contrôle d'un procédé boues activées éliminant l'azote des eaux résiduaires urbaines. PhD thesis. Institut National Polytechnique de Toulouse.

Smets, I. Y., J. V. Haegebaert, R. Carrette and J. F. Van Impe (2003). Linearization of the activated sludge model asm1 for fast and reliable predictions. *Water Research* **37**(8), 1831–1851.