The estimation of the parameters of a fed-batch Bakers' yeast fermentation process

Mols, I.M.

Award date:
1990

Eindhoven University of Technology
Department of Electrical Engineering
Measurement and Control section

THE ESTIMATION OF THE PARAMETERS OF A FED-BATCH BAKERS' YEAST FERMENTATION PROCESS

I.M. Mols

M.Sc. Thesis
carried out from january to august 1990
Commissioned by: prof. dr. ir. P. Eykhoff
Under supervision of: ir. M. Keulers.
dr. ir. A.J.W. van den Boom.
Date: August 1990

Summary.

At Unilever Research Laboratory research is done to develop new controllers for the fed-batch production of Bakers' Yeast. In order to test the controllers on a model to model basis, a simulation model has been built using mass-balance equations. The model, however, contains several unknown parameters that have to be estimated. This is done using the identification scheme proposed by Backx and Damen. Some changes in the scheme had to be made to deal with the non-linear and time varying behavior of the process.

A number of tests has been done to obtain information to design a set of experiments that can be used for the parameter estimation. The input signals for this set are PRBNS-signals added to the open loop setpoints of the process. To deal with the time varying behavior the amplitude of the PRBNS-signal applied to one input increases during the experiment.

As the process dynamics were faster than expected, special arrangements had to be made to sample the process at a sufficiently high rate. Inputs of other experimental setups are used to speed up the sample frequency. The accuracy of the combined set is not high due to synchronization problems. Now these problems have been solved. The final experiments, done to obtain the data sets for minimizing the output error, should be done again. This will improve the accuracy of the data set.

To test the minimization routine a model to model simulation without noise has been done. The estimation of the parameters has been successful. However the estimation of the real parameters has not been successful because of problems due to the routine solving the differential equations of the model. A way has to be found to handle this problem. Until now only a local minimum has been found. The simulation results are good; consequently it is expected that the design of the experiments is good and minimizing the output error is possible using data sets obtained with these experiments.

Contents.

1. Introduction.

At Unilever Research laboratory (URL) research is done to optimize the production of Bakers' Yeast. One goal is to obtain a good simulation model, so testing of new controllers can be done on computers rather than on an experimental configuration. This will save time and money. This master degree thesis deals with the estimation of parameters of this simulation model.

Earlier a physical simulation model has been built. This model describes states such as glucose-concentration, biomass-concentration and culture volume, through mass-balance equations. This model is described in chapter 2. Sixteen parameters are not known and need to be estimated.

These parameters are estimated using the least squares output error method. This means that on the experimental configuration proper signals are applied to the process-inputs to obtain rich output signals. This data set is used to minimize the error between the process outputs and the outputs of the simulation-model in the least squares sense.

To obtain proper input signals the method proposed by Backx and Damen is used. Because this method is designed for linear time-invariant processes and this fermentation-process is highly non-linear and strongly varying in time, the method is modified in order to be used here. The method as it is used, is described in chapter 3.

In chapter 4 to 7 each step along the way to obtain the proper input signals, as well as the results of these steps, are given. The results of these steps are that final experiments can be designed in order to obtain a dataset that can be used to estimate the parameters (chapter 8).

Chapter 9 deals with the estimation of the unknown parameters. First the model to model simulations are discussed and then the estimation of the real parameters. Problems in the estimation routine are described. Recommendations to solve the problems and conclusions of the work done are given in the last chapter, chapter 10.

2. Fed-batch bakers' yeast fermentation.

A simulation model with some unknown parameters has been built at Unilever Research Laboratory (URL) [Keulers, 1988]. It will be used to test new controllers on computers instead of on a laboratory scale, real process. The model is built on the physical knowledge of the process, using mass-balance equations of the most important states. In this chapter the real fermentation process is described and the model is given.

2.1 The fermentation process.

Bakers' Yeast is cultivated in a fermentation vessel. A schematic diagram is given in figure 2.1. In the broth (the grey area in the vessel) the cultivation of yeast takes place. The broth contains yeast, water, dissolved oxygen, substrate and other necessary elements like vitamins. To keep the broth homogeneous, it is mixed by a stirrer. The stirrer dissolves the inlet air. The temperature (T) and pH are controlled primary, The setpoints are 30°C and 5 respectively.

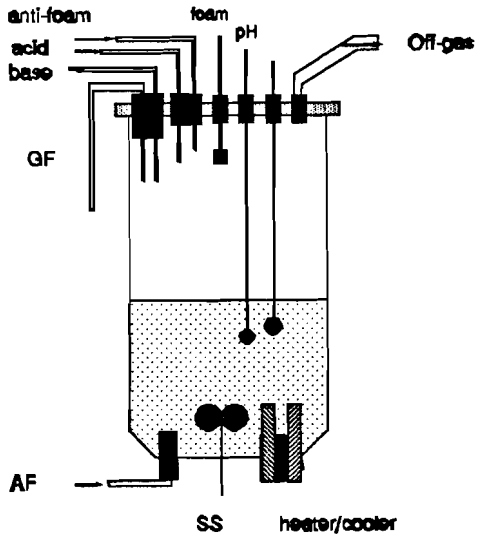

Figure 2.1. A schematic diagram of a fermentation vessel.

The process discussed in this thesis is operated in the fed-batch mode. This means that the substrate is added in an exponential way and no product is removed during the fermentation.

One of the inputs is the substrate flow. The main substrate content is glucose, thus this input is referred to as the glucose flow (GF). The other inputs are the airflow (AF), and the stirrer speed (SS).

The outputs are the dissolved oxygen tension (P) and the off-gas analyzed by the mass-spectrometer. The results of the mass-spectrometer are converted to the ethanol concentration in the broth (E), the oxygen uptake rate (OUR) and the carbon dioxide production rate (CPR). This is also given in figure 2.2.

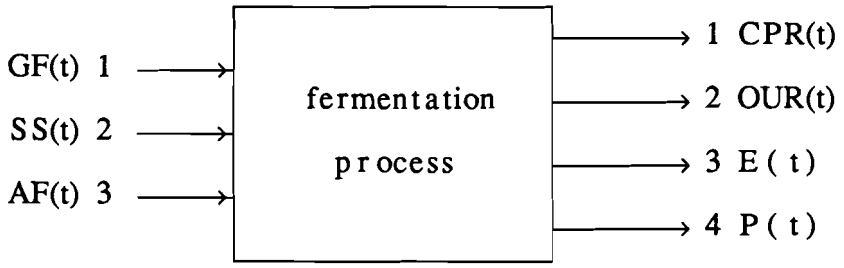

Figure 2.2. In- and outputs of the fermentation process.

Inputs:

GF, glucose flow	liter per hour
SS, stirrer speed	revolutions per second
AF, airflow	liter per minute

Outputs:

P, concentration of dissolved oxygen	mole per liter
E, ethanol concentration	mole per liter
OUR, oxygen uptake rate	mole per hour
CPR, carbon dioxide production rate	mole per hour

2.2 Process model.

The model of the fermentation of yeast is based on the hypothesis that the growth conditions can be divided in 3 pathways [Sonnleitner, 1986].

pathway 1. Under optimal conditions yeast performs an oxidative reaction with

the carbon source:

$$CH_2O + a.O_2 + b.nx\ [NH_3] \rightarrow b.CH_{hx}O_{ox}N_{nx} + c.CO_2 + d.H_2O$$

Pathway 2. If the substrate consumption rate is in excess to the oxidative capacity of the reaction, the reductive reaction occurs:

$$CH_2O + g.nx\ [NH_3] \rightarrow g.CH_{hx}O_{ox}N_{nx} + h.CO_2 + i.H_2O + j.C_2H_6O$$

Pathway 3. Ethanol is metabolized according to the following aerobic reaction:

$$CH_3O_{0.5} + k.O_2 + l.nx\ [NH_3] \rightarrow l.CH_{hx}O_{ox}N_{nx} + m.CO_2 + n.H_2O$$

The $CH_3O_{0.5}$ and C_2H_6O stand for ethanol, CH_2O stands for glucose and NH_3 stands for ammonia, and $CH_{hx}O_{ox}N_{nx}$ stand for biomass composition. Note that all formulas have been normalized to the carbon source.

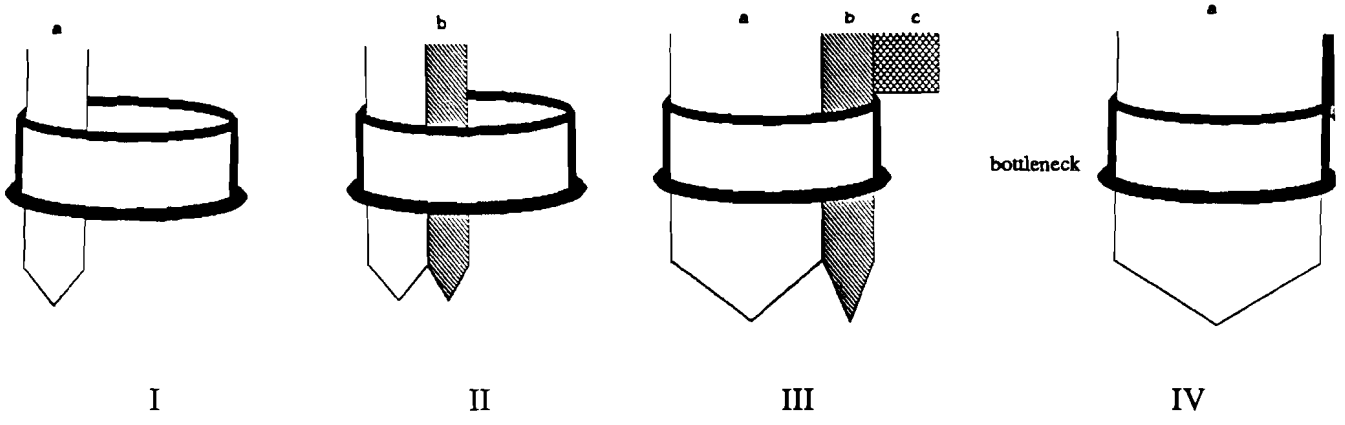

Figure 2.3. Choice of pathways according the oxygen bottleneck.

I) Oxidative reaction of glucose (pathway 1). Glucose flux fits into the bottleneck.

II) Oxidative reaction of glucose and ethanol (pathway 1 and 3). Fluxes fit into the bottleneck.

III) Oxidative reaction of glucose and ethanol (pathway 1 and 3). Fluxes do not fit into the bottleneck.

IV) Partly oxidative reaction of glucose and partly reductive (pathway 1 and 2). Flux does not fit into the bottleneck.

(a) = glucose metabolized oxidatively.

(b) = ethanol metabolized.

(c) = ethanol unchanged.

(d) = glucose metabolized reductivey.

Normally the biomass metabolizes according to pathway 1, or a combination of pathway 1 and one of the other pathways. The pathway chosen to metabolize depends on the available oxygen and glucose. The concentration of dissolved oxygen forms a bottleneck, that gives a restriction on the maximal amount of glucose that can metabolize oxidatively (pathway 1). If the glucose exceeds the maximal amount the exceeding part is metabolized according pathway 2. If ethanol is available and the concentration of glucose is less then the maximal amount that can be metabolized oxidatively, a part of the ethanol can be consumed according pathway 3. This is explained in figure 2.3.

The model built describes the most important states, based on the mass-balance equations. It is valid if the fermentation is in fed-batch mode, i.e. glucose is added to the broth, and the broth is perfectly mixed. The describing states are:

Glucose concentration, G(t)	mole/l
Biomass concentration, X(t)	mole/l
Ethanol concentration, E(t)	mole/l
Concentration of dissolved oxygen P(t)	mole/l
Culture volume, V(t)	l
Enzyme activity function for the uptake of oxygen, Q_{omax}	1/h
Enzyme activity function for the uptake of glucose, Q_{gmax}	1/h
Enzyme activity function for the uptake of ethanol, Q_{emax}	1/h

The differential equations are:

$$\frac{dG(t)}{dt} = -\sigma(t).\frac{G(t)}{K_g + G(t)}X(t) - \frac{GF(t)}{V(t)}.\{ G(t) - G_f\} \qquad (2.1a)$$

$$\frac{dX(t)}{dt} = \left\{\mu_1(t).\frac{G(t)}{K_g + G(t)} + \mu_2(t).\frac{P(t)}{K_o + P(t)} + \mu_3(t).\frac{E(t)}{K_e + E(t)}.\frac{K_i}{K_i + G(t)} - Ms/b\right\}.X(t) - \frac{GF(t)}{V(t)}.X(t) \quad (2.1b)$$

$$\frac{dE(t)}{dt} = \left\{\pi_1(t).\frac{G(t)}{K_g + G(t)} + \pi_2(t).\frac{P(t)}{K_o + P(t)} + \pi_3(t).\frac{E(t)}{K_e + E(t)}.\frac{K_i}{K_i + G(t)}\right\}.X(t) - \frac{GF(t)}{V(t)}.E(t) \quad (2.1c)$$

$$\frac{dP(t)}{dt} = OUR + OTR - \frac{GF(t)}{V(t)}.P(t) \quad (2.1d)$$

with

$$OUR = o(t).\frac{P(t)}{K_o + P(t)} \quad (2.1e)$$

$$OTR = C_1.SS(t)^{C2}AF(t)^{C3}\{P_s - P(t)\} \quad (2.1f)$$

$$\frac{dV(t)}{dt} = GF(t) \quad (2.1g)$$

$$\frac{dQ_{gmax}(t)}{dt} = \left[Q_{gm} \times \frac{G(t)}{K_n + G(t)} - Q_{gmax}(t)\right] / \tau_g \quad (2.1h)$$

$$\frac{dQ_{omax}(t)}{dt} = \left[Q_{om} \times \frac{P(t)}{K_o + P(t)} \times \frac{2G(t) + E(t)}{K_m + 2G(t) + E(t)} - Q_{omax}(t)\right] / \tau_o \quad (2.1i)$$

$$\frac{dQ_{emax}(t)}{dt} = \left[Q_{em} \times \frac{E(t)}{K_e + E(t)} \times \frac{K_i}{K_i + G(t)} \times \frac{P(t)}{K_p + P(t)} - Q_{emax}(t)\right] / \tau_e \quad (2.1j)$$

$$CPR = \left\{\gamma_1(t).\frac{G(t)}{K_g + G(t)} + \gamma_2(t) \times \frac{P(t)}{K_o + P(t)} + \gamma_3(t) \times \frac{E(t)}{K_e + E(t)} \times \frac{K_i}{K_i + G(t)} + a \times Ms\right\} \quad (2.1k)$$

where

			unit
μ_i	=	growth rate of the biomass.	1/h
σ	=	consumption rate of glucose.	1/h
π_i	=	production or consumption rate of ethanol.	1/h
o	=	consumption rate of oxygen.	mole/h
γ_i	=	production rate of carbon dioxide.	mole/h

$^{*}K_g$	=	saturation parameter for glucose uptake.	mole/l
$^{*}K_o$	=	saturation parameter for oxygen uptake.	mole/l
$^{*}K_e$	=	saturation parameter for growth on ethanol.	mole/l
$^{*}K_i$	=	inhibition parameter: free glucose inhibits ethanol uptake.	mole/l
$^{*}K_n$	=	glucose saturation parameter for the induction of the production of glucose consumption capacity.	mole/l
$^{*}K_m$	=	substrate saturation parameter for the induction of the production of oxidation consumption capacity.	mole/l
$^{*}Q_{gm}$	=	maximum specific glucose consumption rate.	1/h
$^{*}Q_{om}$	=	maximum specific oxygen uptake rate.	1/h
$^{*}Q_{em}$	=	maximum specific ethanol consumption or production rate.	1/h
Q_{gmax}	=	specific glucose consumption rate.	1/h
Q_{omax}	=	specific oxygen uptake rate.	1/h
Q_{emax}	=	specific ethanol consumption or production rate.	1/h
$^{*}\tau_g$	=	time constant for glucose uptake.	h
$^{*}\tau_o$	=	time constant for oxygen uptake.	h
$^{*}\tau_e$	=	time constant for ethanol uptake.	h
$^{*}C_1$	=	constant.	s.min/h
$^{*}C_2$	=	constant.	
$^{*}C_3$	=	constant.	
Ps	=	saturated dissolved oxygen concentration.	mole/l
Gf	=	glucose concentration of the glucose flow.	mole/l
*Ms	=	maintenance term of biomass.	1/h

Parameters marked with * are unknown, and need to be estimated.

This model is implemented in Fortran and given in appendix E.
It can be divided in 4 parts:

a). Calculate the combination of pathways.
First the oxygen bottleneck is calculated. Then the maximal glucose and ethanol flow that fit in the bottleneck are calculated.

b). Calculate the influences of the AF and the SS on the OTR.
Formula (2.1f) is used.

c). Calculate the new states and the influence of GF on the states.
Formulas (2.1a) - (2.1j) are used.

d). Calculate the outputs.

Formula (2.1c), (2.1e) and (2.1k) are used.

The differential equations are solved using a Runga-Kutta-Merson algorithm, that is implemented in the NAG-library.

3 Survey of the method used for parameter estimation.

Output error least squares criteria are used to estimate the unknown parameters in the model of Bakers' Yeast fermentation. The result is well-suited for the simulation on the basis of process inputs only. Good estimation results make high demands upon the input signals. For example one should use certain enriched signals that give on all outputs the same, high signal to noise ratios. The method proposed by Backx and Damen [Backx and Damen, 1989] is used to obtain the proper input signals. This method consists roughly of 6 phases that are discussed in the next paragraphs.

3.1 Talks with the operator and study of sensors and actuators.

To get a base for the experiment design, one needs to have an indication of the process characteristics, like the smallest time constants, range of inputs and outputs and delay times. In most cases the operators can give this indication. They also can give information on how the process is operated normally.

In this case results of each phase are also discussed with the operator and scientists of URL in order to know if the results of the experiments are to be expected and if they can be improved.

To overcome non-linearity most of the inputs are controlled with primary SISO controllers. If so, the controllers have to be tested to verify the accuracy and the linearity of the controller-actuator combination. If not, the actuator is tested separately.

It is also possible that the actuators have transfer functions that can not be neglected (non-linear or large time constants or delays). If so the transfer functions have to be included into the process model; sometimes it is even necessary to estimate the parameters of the transfer functions too.

The sensors have to be linear or linearizable over the range they are used. One also needs to have an idea of the delay times. If necessary these effects can be included in the model.

Note that this information is obtained by study of databooks and by measurements.

3.2 Free run experiments.

Free run experiments are experiments where the noise signals at the output can be separated from the other signals. In this way the noise can be examined. In order to get a good estimation of the parameters, one needs to know the variance of the noise. The noise is supposed to be white gaussian with zero mean and additive at the output; if not, the model is to be extended to include the noise characteristics. To test the whiteness of the noise, the bandwidth of the noise is calculated.

The outputs contain signals due to the input actions, the growth of the biomass, trends and noise. A trend is a slow drift or variation in the output. Sometimes the cause of the trend is known but it is impossible to prevent the trend.

In this experiment only the noise signals are wanted. To obtain the noise characteristics all the deterministic signals, have to be removed. If the process inputs are set at constant level the outputs do not show signals due to the inputs. However the process is normally operated in such a way that the biomass grows exponentially. Noise characteristics can change during the experiments if the noise is not adaptive and the process is non-linear. Therefore the process has to be operated in the normal way so changes in noise characteristics can be detected. The inputs are varying slowly so the reactions of the process at the outputs can be seen as a slowly changing trend. The effects of the growth of the biomass are also varying slow and can be seen as a trend too. Trends can be removed by filtering off-line using a causal low-pass filter of appropriate cut-off frequency consecutively both in forward and reverse time. This filter procedure will give no phase shift. The remaining output signals contain noise only.

An exponential growth of the biomass is achieved as the glucose flow (GF) increases exponentially too (see figure 3.1). The increase of the flow is calculated from the growth setpoint that is given by the operator, and the

mass of the yeast at the beginning of the experiment. The inputs airflow (AF) and stirrer speed (SS) are kept constant on a sufficiently high level. These process conditions will be referred to as open loop setpoints. The open loop setpoints are the normal operating points if there is no feed-back control.

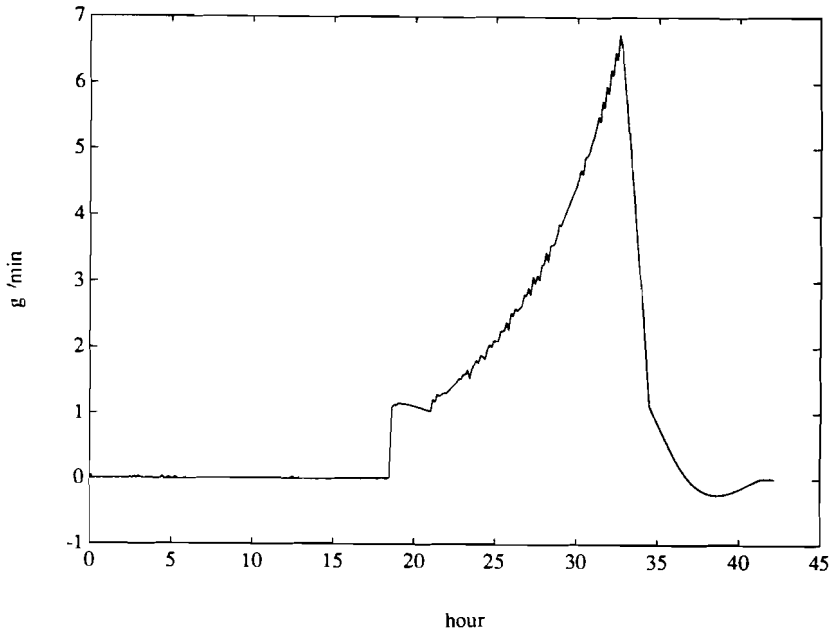

Figure 3.1. Glucose flow setpoints of an open loop experiment.

After detrending the output signals of the process the remaining noise signal can be processed:

- The standard deviation (σ) of the signal is estimated using the formula:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - y_{mean})^2 \qquad (3.1)$$

Where N stands for the length of the used detrended dataset, y_i is one sample of the set and y_{mean} is the mean value of the detrended dataset. To get a meaningful result, one should take care that the characteristics of the noise (mean value and standard deviation) are constant. This is done by visual inspection. One also can calculate σ^2 using the the mean amplitude of the peaks of the noise signal. The mean amplitude equals the variance if the noise is white. Of course both methods should give the same results.

- The average power of the noise (P_n) is calculated using:

$$P_n = \frac{1}{N} \sum_{i=1}^{N} y_i^2 \tag{3.2}$$

If the noise is white, gaussian and zero mean, the power equals the variance of the noise for a sufficient long dataset.

- Using the fast fourier transform the power spectrum of the noise can be calculated. This is done to know the energy in the disturbance and how the energy is distributed over the different frequencies.

- To detect changes in the noise characteristics visual inspection of the data set is used. The noise should behave in the same way during the whole process.

3.3 Staircase experiments.

In order to obtain an indication of the steady state gains and largest time constants staircase signals (see figure 3.2) are applied to the process inputs separately. In linear processes this is done by applying one step at the inputs separately. However process states react non-linear according Monod kinetics and the gain depends on the level of the states. Therefore several steps at different levels have to be made. The magnitude of each step is quite arbitrary but it should have at least a certain minimum signal to noise ratio, thus the responses of the outputs can be used to estimate the gains and time constants by visual inspection. The duration of one step is long enough to be sure that the responses of all outputs are damped. The levels of the staircase should cover the whole input range of interest.

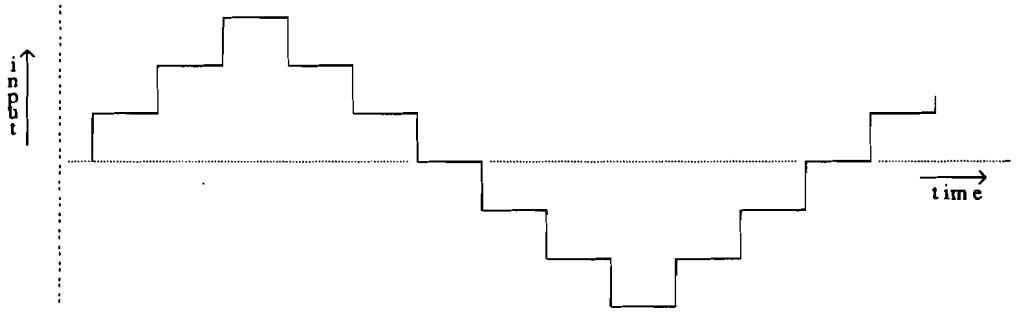

Figure 3.2. Staircase input signal.

If the model that is to be expected is linear and time invariant, then these experiments are used to check the linearity of the process. In this case the

process and the model are non-linear. The experiment is used to investigate the amount of non-linear behavior of the process.

The steady state gain is the difference of the final output level after one step and the level at the beginning of this step, divided by the difference of the input levels. The largest, dominant time constant is 1/3 the time needed to reach 95% of the final output level.

3.4 First PRBNS-experiments.

For this experiment mutually independent pseudo random binary noise sequence (PRBNS) signals are applied to the inputs simultaneously. These signals are symmetric around the normal operating points. The clock frequency (f_{PRBNS}) equals the sample frequency (f_s), this guarantees a white frequency spectrum on the used bandwidth of 1/2 f_s as is necessary for identification. The duration of the experiment needed to allow a reliable estimation of the process values, depends on the signal to noise ratio. But a rule of thumb is that they have to cover at least 5 to 10 times the longest time constants. The time constants are estimated in the staircase experiments.

In the case parameters are estimated for linear processes the amplitudes of the signals are chosen such that the signal to noise ratios (S/N- ratios) of all the outputs are equal. This is done by setting the S/N- ratio of one output equal to:

$$\left(\frac{S}{N}\right)_n = \frac{a_1^2 \times T}{\sigma_n^2} G_{n,1}^2 + \frac{a_2^2 \times T}{\sigma_n^2} G_{n,2}^2 + \ldots + \frac{a_p^2 \times T}{\sigma_n^2} G_{n,p}^2 \quad (3.3)$$

and

$$\left(\frac{S}{N}\right)_1 = \left(\frac{S}{N}\right)_2 = \ldots = \left(\frac{S}{N}\right)_n = \ldots = \left(\frac{S}{N}\right)_q$$

where P stands for the number of the inputs, Q for the number of outputs, $G_{i,j}$ stands for the gain from the input j to the output i obtained in the staircase experiments, n is one of the outputs and a_k is the amplitude of the k^{th} input PRBNS-signal. T is the minimal time the PRBNS-signal is constant (1/f_{PRBNS}). The estimate of the variance of the noise σ^2 was obtained in the free run

experiments. Obtaining the amplitudes for the PRBNS- signals that are applied to the inputs is done by solving the set of equations where all S/N- ratios are equal and a_k is maximized for all k, although the process has physical limitations that limit the amplitude a_k. In this case the equations can not be solved because there are 5 equations and 4 variables (a_1, a_2, a_3, S/N), a best possible fit can be used instead.

However eq. (3.3) can not be used as the process is not linear. The summation of the signal to noise ratio calculated for each input separately, is not necessarily the signal to noise ratio at the output. Reactions of the output due to two different input signals, can be summed in linear cases. This gives the reaction of the output if both signals are applied simultaneously. This process is not linear, therefore it is not allowed.

If the process is excited in such way that all states of the process are excited, the input signals are 'rich' enough. The inputs have to be sufficiently 'rich' to be able to estimate the unknown parameters of the process. In linear cases the signal to noise ratio is set equal for all outputs to give an equal variance in the estimated parameters. As the signal to noise ratio due to the inputs can not be calculated using eq.(3.3), the idea is simplified and it is tried to keep the amplitude of an output due to the (PRBNS-) signal of one input equal to the amplitude of that output due to the signals of the other inputs. Because the output reacts with the same amplitude on each input, it is suspected that all states affected by that input are excited enough. To obtain the amplitudes of the PRBNS-signals the following equations are solved and A_{ni} are maximized:

$$\left.\begin{aligned} A_{n1} &= a_1 \times G_{n,1} \\ A_{n2} &= a_2 \times G_{n,2} \\ A_{n3} &= a_3 \times G_{n,3} \\ A_{n1} &= A_{n2} = A_{n3} \end{aligned}\right\} \quad \text{for all n} \qquad (3.4)$$

A_{ni} stands for the amplitude of the n^{th} output due to the i^{th} input.
If the set can not be solved for all outputs, the amplitudes have to be chosen so the equations are fulfilled in the best way. The noise variance is not used

to obtain equal signal to noise ratios at all outputs. To get equal variance in the estimates of the parameters the noise variance is used as a weight function in the estimation algorithm.

The bandwidth can be calculated from this dataset. After removal of outliers the fourier transform is calculated. The bandwidth is estimated by visual inspection of the spectrum. The largest bandwidth of all outputs is then taken as the process bandwidth.

The cross-correlations of the inputs and outputs are calculated to obtain the delay times. The position of the significant peaks gives the delays in the outputs. These delays have to be included in the model.

3.5 Final PRBNS-experiments.

The final PRBNS-experiment is the most critical of all. This dataset is used for the minimizing procedure. Thus the accuracy of the parameter estimation can be decreased if the experiment is not designed correctly.

The amplitude of the PRBNS-input signals are calculated in the same way as given in § 3.4 eq. 3.4. The clock frequency of the PRBNS-signal equals the Nyquist frequency, i.e. twice the bandwidth of the process. The sample frequency is higher than the Nyquist frequency because this improves the signal processing. Thus $f_s = \alpha \times f_{nyquist} = \alpha \times f_{PRBNS}$, where α is an integer called oversampling ratio. The duration of the experiment is as long as possible. This guarantees the smallest variance in the parameter estimates. In an industrial environment the loss of production will limit the duration, but the fermentation experiments are done in a laboratory so this will not give a limitation of the duration of the process. As the changes of the setpoints are done by hand, the maximal time the process can be operated is one work day (8 hours).

The output signals are processed. Normally this would involve the removal of outliers and detrending and a compensation of the delay times. In this case, however, the signals can not be detrended because information involving the growth of the biomass will be lost too. The experiments are done in a laboratory where the conditions are kept constant and the sensors are

calibrated before every experiment, thus the trends are expected to be very small and can be neglected.

3.6 Minimizing the output error.

The method for estimation of the parameters of the model, proposed by Backx and Damen [Backx and Damen, 1989], uses a special strategy to minimize the computer time and obtain a low order, linear model. This can not be used for this model because it is non-linear and the model set is already given. Therefore a direct method for estimation is used.

The minimization criterion is defined thus:

$$\hat{\theta} = \min_{\theta} \left\{ \sum_{j=1}^{M} \sum_{i=1}^{N} \left(y_{j,i}^{t}(\theta_t) - y_{j,i}^{m}(\theta) \right)^2 / \sigma^2 \right\} \qquad (3.5)$$

where $\hat{\theta}$ is the solution of the minimization, θ_t stands for the real process parameters, θ is the set of unknown parameters that are to be estimated, M stands for the number of outputs, N stands for the length of the data set, $y_{j,i}^{t}$ stands for the i^{th} sample of the measured output j and $y_{j,i}^{m}$ stands for the i^{th} sample of the simulated model output j. σ_j stands for the noise variance of the j^{th} output as estimated in the free run experiments.

The solution of the problem will be unchanged if the error function is multiplied by a positive constant. The performance of the minimization method will be significantly increased. Since convergence tolerances and other criteria are necessarily based on an implicit definition of 'small' or 'large', problems with unbalanced scaling may cause difficulties for the algorithm. As there is an indication of the expected values of the parameters $\hat{\theta}$ the parameters are scaled by dividing the parameters by the expected value. All estimated parameters are then normalized and the expected value is near one.

The output error is minimized using the Marquardt method [Marquardt, 1963], that combines the fast descent in the error of a Gauss-Newton method and the robustness of a steepest descent method. Because it is very hard to give the gradient of the model analytically, it is calculated numerically [Gill and Murrey, 1978].

4 Talks with the operators and study of actuators and sensors.

The talks with the operators and the study of the actuators and sensors are done to design the experiments needed for the estimation of the process parameters. It is necessary to have some knowledge of the process and how it is controlled. A summary of the work done is given in this chapter, the discussion of all results will be given in appendix F.

4.1 Talks with the operators.

The operators give an indication of the normal operating points of the fermentation process. The time constants of the process are supposed to be large, near 20 minutes. Only the dissolved oxygen tension (DOT) will react much faster, near 6 minutes.

4.2 Study of the actuators.

The process contains 3 actuators: the stirrer motor, the airflow pump/valve and the glucose flow/pump. The inputs are governed by the Applikon unit. This unit contains SISO controllers for the inputs.

- Stirrer motor.
No control loop is implemented in the Applikon unit. The error between the setpoint and the real speed is large (up till 10%). The setpoint has to be corrected before estimation of the parameters of the process.

- Airflow pump/valve.
This input is controlled well by the Applikon unit. The error is small (1% of the setpoint) as well as the delay times and the time constants (less then 5 seconds). Non of these actuator characteristics have to be included in the model.

- Glucose flow/pump.
The glucose flow is controlled by the Vax. This makes an exponential growth of the flow possible. The controller does not function well. It is too slow to track the fast changes that are necessary for the PRBNS-experiments. The signal to noise ratio of the measured flow is very high and can be used

either. Therefore the pump voltage is used to change the flow. The setpoints of the pump voltage is scaled to the measured flow. These scaled setpoints are used for the estimation of the parameters instead of the measured flow.

4.3 Study of the sensors.

Although there are 4 outputs there are only 2 sensors. These are the DOT-probe and the mass-spectrometer. The last one measures the OUR, CPR and the ethanol concentration in the broth.

- DOT-probe.
The probe functions well, it reacts fast on the changes of the oxygen concentration that is dissolved in the broth. As the DOT changes fast, it has to be sampled every 5 seconds.

- The mass-spectrometer.
The mass-spectrometer measures the oxygen, ethanol and carbon-dioxide concentration of the off-gas. From these values the OUR, ethanol concentration in the broth and CPR are calculated. The delay time of the mass-spectrometer is expected to be near 40 seconds. The sample frequency depends on the number of users of the spectrometer. However the minimum sampling period is about 30 seconds.

- Off line measurements.
It is possible to measure the biomass concentration of the broth off line. However the noise of the method is high and the accuracy low, therefore it is not used as an output.

5 Free run experiments.

The free run experiments are used to determine the noise characteristics, i.e. bandwidth, variance and power spectrum of the noise signals. It is assumed that the noise is white, gaussian and additive at the output. This whiteness is tested by calculating the power spectra of the noise. If it is not white the noise characteristics have to be included in the model. If the noise is added to the output its characteristics will not change during the process. This is checked visually. If the noise is white and additive, it is supposed to be gaussian too.

5.1 Experiment design.

As discussed in § 3.2, open loop setpoints are used for the inputs during the free run experiments. Air flow is set at 4 liter per minute and stirrer speed at 750 RPM. The glucose flow increased exponentially as calculated by the controller. The sample rate of CDAS is set at 4 minutes, but the mass-spectrometer used approximately 6 minutes for a measurement cycle. At that moment that was the highest possible sample frequency.

Experiment numbers and initial settings are given in appendix A.

Parts of the output data (especially DOT, OUR and CPR) showed strong low frequent behavior due to the actions of the glucose flow controller (see appendix C, figure C.1). As this behavior is due to large input variations, the results after signal processing can not be seen as noise only. Therefore these parts are not used for determination of the noise characteristics. It was necessary to set the airflow at 5 liter per minute after approximately 4 hours. The reaction on this change of setpoint has to be completely damped before the data could be used.

Only 50 samples of the ethanol output were available after these 3 experiments, divided in short data sets. To minimize the effects of the glucose pump, the pump was set at a fixed setpoint. The setpoints will not track the open loop setpoints, but if the noise is additive at the output then the noise characteristics will not change during the experiment and data at any setpoint can be used for determining the variance and power of the noise.

At the ethanol output the noise is clearly additive at the output, there is no difference in the noise characteristics at the end of the experiment and at the begin (there are already estimations of the noise characteristics available). Visual inspections of the other outputs give no reason to believe the noise is not additive at the output (although it is very hard to check this, due the control actions of the glucose flow). The data obtained in these experiments are used to determine the noise characteristics of the other outputs.

5.2 The results.

Data analysis was done using the interactive program package MATLAB. The data is detrended using IPCOS[1]. The proper cut-off frequency of the detrend filter is determined using the function OFF-TRE.M[2]. Removal of outliers was not necessary as there were no spikes in the data. The remaining data sets are given in appendix C, figure C.2. The power spectra are calculated using the functions SPECTRUM.M and SPECPLOT.M (appendix C, figure C.3). The spectra are white over the measured frequency range (sample frequency is 6 minutes thus the maximal frequency of the data is 5×10^{-3} Hz).

The noise is supposed to be gaussian, therefore the power of the noise equals the variance. The variance is estimated both visually as calculated using STD.M (calculates the standard deviation using eq.(3.1)). The mean amplitude of the noise peaks are estimated visually. This amplitude equals the noise variance. The results were identical for both methods and are given in table 5.1. The estimates are done using only a part of the set (samples 250 - 350) as in this region the noise characteristics seem constant.

[1] IPCOS is a tool developed by DATEX industries to estimate parameters of linear time- invariant processes based on the scheme as proposed by Backx and Damen [Backx and Damen, 1989]. It works within the MATLAB environment. In this tool an easy to use data processing part is included.

[2] OFF-TRE.M is a function developed for IPCOS, but until now it is not integrated yet.

Table 5.1 The standard deviation (σ) and variation (σ^2) of the noise of the process.

output measured units	CPR $\frac{mmole}{hour}$	OUR $\frac{mmole}{hour}$	DOT %	Ethanol mmole
σ	1.36	3.48	0.11	6.9
σ^2	1.85	12.11	0.012	47.61

The results will be used in the procedure that minimizes the output error as described in § 3.7.

6 Staircase experiments.

The staircase experiments are used to estimate the steady state gains and the largest time constants, and not at least to give an idea of the non-linearities of the process. The gain is the difference of the final level of the output and the level at the beginning of one step, divided by the difference of the input levels. The time constant is 1/3 the time needed to reach 95% of the final output level after one step is applied. Non-linearities are detected by comparing the results of each step. The magnitude of the difference between the results is an indication for the non-linearities.

6.1 Experiment setup.

The experiment starts as the ethanol, that is produced in the batch phase, is consumed. The stirrer speed (SS) is set at 750 RPM and the airflow (AF) at 4.0 l/min. The glucose flow (GF) is controlled in the normal way for an open loop experiment. For the experiments where staircase signals are applied to the stirrer speed, a step of 10% is made every half an hour. After half an hour the outputs do not react on the input step anymore. During the airflow experiment also a step of 10% is made every half an hour. During the glucose flow experiment the step is made by entering a new value of the biomass concentration (X_o) in the workstation, the controller calculates a new setpoint for the glucose flow using eq. (4.1). As the setpoint of the glucose flow is linear with the biomass concentration, a change of 10% of the biomass will give a change of 10% of the flow. After one step it showed that 10% change was too small and for the other steps a 20% change is used. The inputs for the different experiments are given in figure 6.1.

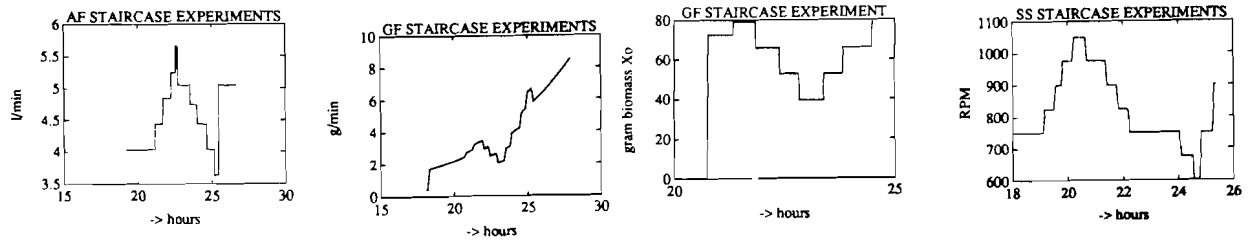

Figure 6.1. The inputs used for the staircase experiments.

6.2 The results.

The gains and time constants are very hard to estimate due to the fact that the outputs show trends caused by the growth of the biomass. The results are a rough indication of the real values. It was tried to correct the trend by dividing the outputs by the estimate of the biomass at every sample moment. This gives bad results because the estimation of the biomass is not accurate. Therefore the estimation is done without correction of the trend. Results of the 3 staircase experiments are combined in table 6.1 and table 6.2 and given in appendix C, figure C.4.

6.2.1 The gains of the process.

The gain of the last step (step 7) applied to the glucose flow is very high. At that moment all the ethanol produced earlier was consumed. The outputs did not react on the step alone but also on the change of pathway, used by the metabolism.

Table 6.1 The difference between the final output level and the level at a start of a step as measured during the 3 staircase experiments.

	GF level to	te	SS level to	te	AF level to	te	ΔL RO GF mmole/h	ΔL RO SS mmole/h	ΔL RO AF mmole/h	ΔL RCO2 GF mmole/h	ΔL RCO2 SS mmole/h	ΔL RCO2 AF mmole/h	ΔL DOT GF %	ΔL DOT SS %	ΔL DOT AF %
1	72.6	79.2	750	825	4.0	4.4	20	*	79	30	*	93	- 4.3	12.21	17
2	79.2	66.0	825	900	4.4	4.8	-179	*	46	-175	*	50	17.1	4.31	12
3	66.0	52.8	900	975	4.8	5.2	- 61	*	41	-168	*	47	18.5	3.32	*
4	52.8	66.0	975	1050	5.2	4.7	- 98	*	60	- 80	*	54	12.8	3.50	*
5	66.0	79.2	1050	975	4.7	4.4	- 73	*	- 58	- 78	*	- 56	7.7	- 7.95	*
6			975	900	4.4	4.0	70	*	- 48	85	*	- 57	-16.3	- 7.30	*
7			900	850	4.0	3.6	384	*	- 28	248	*	- 34	-40.0	-14.10	-8.9
8			850	700	3.6	5.0		- 89	- 75		80	- 84		-16.50	-4.3
9			700	675				- 81	- 42		*	- 37		- 0.70	-0.8
10			675	600				184	295		102	295		- 0.77	0.7
11			600	750				170			*			1.39	

Level 'to' stands for the input level at the beginning of a step and 'te' for the level at the end if all inputs are damped.
ΔL stands for the difference between the output level at 'te' and 'to'.

* indicates that no gain could be estimated. This is not possible by high values of the outputs because saturation occurs at these levels. As these levels will not be used for the estimation of the parameters of the process, it is not necessary to obtain these values in another way. For the ethanol output no change in level could be measured. It is known that ethanol does not react strongly on changes of the inputs. The data of the ethanol output is not measured in the broth but calculated from the off-gas. The change over of the dissolved ethanol to ethanol in the broth acts as a filter at the output. It is unknown how large this effect is. If the effects are large, it might be necessary to model these effects as it is not modeled yet. If validation of the model show large differences between simulation and measured output, it may be considered to model the filter.

Most outputs do not show a large non-linearity, except the DOT, which reacts on a step of the stirrer speed in 3 different ways:
a. If the DOT is high, a step applied to the input has a very small gain. It could not be measured. This is caused by the saturation of the DOT.
b. If the DOT is between 5 to 25%, the DOT will react strongly. This region is (most frequently) used for controlling the process.
c. If the DOT is less than 5%, the DOT reacts hardly at all. This is caused by the oxygen limitation. There is so much glucose available that a step applied at the input, give no significant change anymore. The small changes it gives can not be detected anymore because the accuracy of the DOT-probe is not good enough at this low level. These effects can be seen in appendix C, figure C.4.

It is expected that the gains depend on the biomass concentration. The gains of the stirrer speed and air flow did not show this expected change. This is probably due to the relatively short time the experiments are done (8 - 9 hours). In the table the gain of the glucose flow does not change either, but amplitude of a step is a percentage of the setpoint. The setpoint will increase exponentially as is normal for open loop setpoints. Therefore the amplitude of the glucose step will increase too and the gain will decrease as the biomass grows.

The gains are given in table 6.2. Note that the gains are normalized at the inputs. Therefore the gain of the stirrerspeed is small because it is divided by 75, and the gain of airflow is high because it is divided by 0.4.

Table 6.2 The estimated gains of the staircase experiments.

	OUR mmole/h	CPR mmole/h	DOT %
GF % X_o	5.8	5.3	0.73
SS RPM	1.8	1.3	0.5 / 1.6
AF l/min	-134	-142	16.5

The DOT has 2 values if a step at stirrerspeed is applied. The first one (0.5% /RPM) is valid for a high DOT level (more than 30%), the second one (1.6% /RPM) is valid for a normal level (5 and 25%). The third one, valid for low level of DOT (smaller than 5%), could not be estimated because it was too small.

6.2.2 The estimation of the largest time constants.

The sample time for the DOT is 5 seconds. For the other outputs, measured by the mass-spectrometer, the sample time is between the 3 and 6 minutes. The outputs react faster than expected, OUR and CPR reached 95% of the final level within 1 or 2 sample periods. The largest time constant is smaller than $\frac{2}{3}$ of the sample time (near 4 minutes). As the sample frequency is too low the results are only a rough indication of the real time constants. Only the DOT, as it was sampled fast, could be estimated reliably.

There are 3 phases to be recognized in the stirrerspeed experiment although they differ from the ones discussed in the paragraph of the gains.

a. The DOT reacts quickly, the time constant (τ) is estimated to be 0.006 h ($\cong$ 20 seconds), if the step is positive and no oxygen limitation occur. But then this is a pure physical process.
b. The DOT reacts slowly, τ is 0.04 h (2.4 min), if the step is negative and no oxygen limitations occurs. Now the decrease of the DOT is caused by the biochemical process of the biomass.
c. If oxygen limitation occurs, the changes in the DOT are that small that no time constant can be estimated.

The difference between the physical and biochemical process can be explained as follows: as the stirrerspeed increases more oxygen dissolves in the broth, this reaction is quite fast (see § 4.2.4). After a dead time the yeast culture reacts on the increased DOT. It will metabolize more substrate, if available. Now the increase of DOT will be slower. For large changes of the stirrerspeed even overshoot can occur before a balance between the dissolving oxygen and oxygen uptake arise. As the step in the staircase experiments are quite large, a small overshoot occurs and the 95% level is reached fast.

It is also expected that time constants change as the biomass grows. This effect can not be found in the results of these experiments. This is probably caused by the rough estimation due to the large trends and the low sample frequency.

6.3 Conclusion.

The gains of the process have been estimated. The results are reliable as they are the steady state gains.

The time constants are not estimated reliably, but it is found that they are much smaller than expected. They are essential in order to determine the duration of the PRBNS-experiments. The experiments can cover one work day. As no production loss is made, this is done. The time constants are in minutes, so experiments that cover a day are certainly long enough.

Large non-linearities and time varying reactions are found as expected. Effect of the non-linearity in the time constants are not found as the sample time was too large with the exception of the DOT that was sampled fast.

The most important results are obtained in these experiments. There is no need to do the experiments again, although the sample frequency was too low.

7. First PRBNS-experiments.

The first PRBNS-experiments are done to obtain the bandwidth of the process. As the clock frequency of the PRBNS-signals equals the sample frequency, all frequencies of interest are covered. The power spectra of the output signal are used to estimate the bandwidth. It is necessary to estimate the bandwidth in order to determine the frequency of the final PRBNS-signals and the sample rate. The clock frequency of the PRBNS-signals equals the Nyquist frequency and the sample frequency is at least the Nyquist frequency, but it is preferred to sample faster.

7.1 Experiment design.

The PRBNS-sequence is made using a 10 bit shift register. This gives a set of $(2^{10}-1)$ bits that are Uncorrelated. This set is divided in 3 subsets of 340 samples each. These sets are used to generate sequences that are applied to the inputs simultaneously. This ensures that the input signals are mutually uncorrelated.

In linear cases the PRBNS-sequence is applied symmetrically around the open loop setpoints, i.e. if the next number in the sequence is a '1' a positive step with amplitude a_i is made from the open loop setpoint, but if the number is a '0' a negative step is made with the same amplitude. Thus the input signal switches from the open loop setpoint plus a_i to the setpoint minus a_i and visa versa, according to the PRBNS-sequence. This is done for the airflow (AF) and stirrer speed (SS). However if the glucose flow (GF) is higher than the open loop setpoint, ethanol production might occur. In most experiments this is not wanted and therefore the PRBNS-sequence is not symmetric around this setpoint. On a '1' the glucose flow is set to the open loop setpoint and on a '0' it is set $2 \times a_i$ less than the open loop setpoint. Note that a_i stands for the amplitude of the i^{th} input as calculated using eq (3.4).

In these experiments the glucose flow (GF) is not controlled by the primary controller. Instead the voltage of the pump is changed to change the glucose flow. The non-linearity of the pump is neglected to simplify the calculation of a new setpoint. This is necessary because changing the setpoints is done manually. The glucose flow is supposed to change 20% as the pump setting is

changed 20%. The pump setting PS^* is calculated from the calculated glucose flow setpoint (GF^*) as used in an open loop experiment:

$$PS^*_{(\%)} = \begin{cases} 10 \times GF^*_{(g/min)} & \text{as PRBNS} = 1 \\ 8 \times GF^*_{(g/min)} & \text{as PRBNS} = 0 \end{cases} \tag{7.1}$$

This pump setting gives not exactly the glucose flow calculated for the open loop setpoints, but the flow increases exponentially and that is most important.

However for small values of the glucose flow setpoint the change is that small that the reaction of the outputs can not be measured. Therefore PS^* is set at 0% if the next value in the PRBNS-sequence is '0'. This is done until PS^* reaches 10% (see figure 7.1). This is also done to improve the estimation of the glucose parameters. They can be estimated best if the glucose concentration is low.

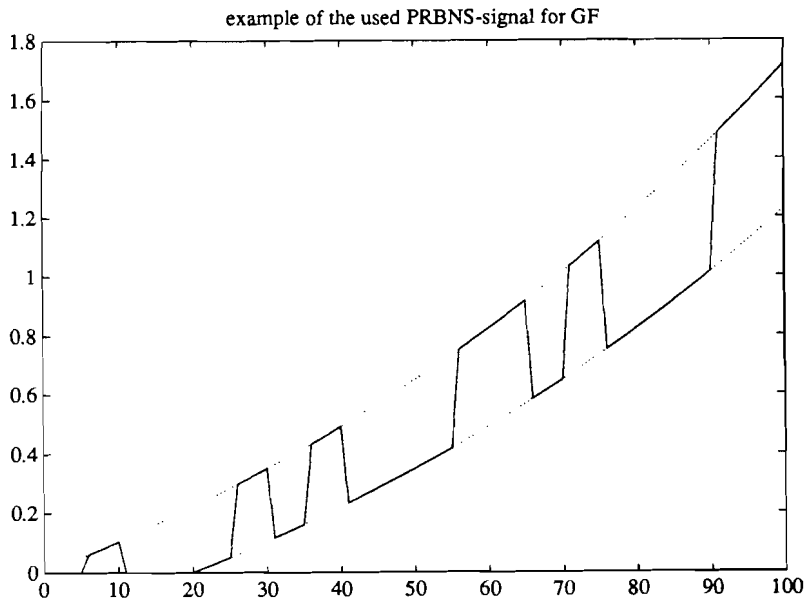

Figure 7.1. Glucose flow setpoints for the PRBNS-experiments.

The experiments are designed for a change in the DOT of maximal 15%. A larger change in the output could give oxygen limitation and ethanol production. It also could take the process away from the open loop setpoints. As we want to keep near the normal operation points this is not wanted. The in- and outputs are numbered as given in figure 2.2.

This would mean that the amplitude of the output due to the inputs (A_{3i} i=1, 2, 3) should equal 7.5% in order to avoid 15% change in DOT. As the ethanol does not react on the inputs this output is not used to obtain the amplitudes (a_i) of the PRBNS-signals. Using eq. (3.4) and table 6.2 gives:

$$\left.\begin{aligned} 7.5 &= a_1 \times G_{3,1} \\ 7.5 &= a_2 \times G_{3,2} \\ 7.5 &= a_3 \times G_{3,3} \end{aligned}\right\} \qquad \begin{aligned} a_1 &= 10\ \% \\ a_2 &= 47\ \text{RPM} \\ a_3 &= 0.45\ \ 1/\text{min} \end{aligned}$$

The changes in CPR and OUR are then:

OUR:

$$\begin{aligned} a_1 \times G_{1,1} &= 58 \\ a_2 \times G_{1,2} &= 85 \\ a_3 \times G_{1,3} &= 60 \end{aligned}$$

CPR:

$$\begin{aligned} a_1 \times G_{2,1} &= 53 \\ a_2 \times G_{2,2} &= 61 \\ a_3 \times G_{2.3} &= 64 \end{aligned}$$

The influence of the glucose flow is small on the OUR and CPR, therefore the stirrer speed and air flow are decreased to equal the influences. The amplitudes (a_i) are chosen as follows:

$$\begin{aligned} a_1 &= 10\ \% \\ a_2 &= 37.5\ \text{RPM} \\ a_3 &= 0.3\ \ 1/\text{min} \end{aligned} \qquad (7.2)$$

Note that the amplitude of the PRBNS-signal applied to the glucose flow is in percentage of the open loop setpoints. Therefore the amplitude will increase during the experiment. The setpoints of the stirrer speed and air flow are increased if the DOT becomes that low that oxygen limitation could occur.

As indicated in § 3.4 the clock frequency of the PRBNS-signals equals the sample frequency. As the highest possible sample frequency for the mass-spectrometer is 6 minutes, the clock frequency is set at 6 minutes. The new setpoints are set by hand if new outputs are available from the mass-spectrometer, as can be seen on the personal computer that controls it.

7.2 Results.

The outputs of the experiment is given in appendix C, figure C.5. The power

spectra are calculated for the 4 outputs. They are given in appendix C, figure C.6. The spectrums are calculated using the function SPECTRUM.M over a set of 130 samples for CPR, OUR and E. For the spectrum of the DOT, that is sampled at 5 seconds instead of 6 minutes, 2470 samples are used. The bandwidth is chosen where the amplitude of the spectrum is 1000 times smaller than the amplitude at the lowest frequency (note, not the D.C. amplitude).

The sample frequency of 6 minutes is to low for the CPR and OUR. Folding effects caused by alliasing can be seen in the power spectrums of those outputs. The power spectra do not reach a noise level but instead they are raising at higher frequency. The bandwidth of the DOT is near 0.01 Hz (time constant is 1.6 minutes). The Nyquist frequency is twice the bandwidth and is 0.02 Hz. This implies that the minimum sample frequency is also 0.02 Hz (sample rate is 50 seconds).

To simplify the minimization the sample frequency of all outputs are chosen equal. Axelsson [Axelsson, 1989] uses first order processes to model OUR and CPR, this means that the largest time constant is also the smallest time constant. The bandwidth of these outputs will be the reciprocal to the time constant. As the time constant is estimated to be 4 minutes the estimated bandwidth is 0.004 Hz. The process bandwidth is the largest bandwidth of all outputs and is 0.01 Hz. Therefore a sample frequency of 50 seconds is to be used for the final PRBNS-experiments.

8 Final PRBNS-experiments.

The final PRBNS-experiments are done to gather the data sets for the estimation of the unknown parameters. Delay times and bandwidths are estimated. Delay times are estimated using step responses and bandwidth are estimated using power spectra. This could not be done using the data of the first PRBNS-experiments because the sample frequency was too low.

8.1 Experiment design.

The amplitudes of the PRBNS-signals are chosen as in the first PRBNS-experiments. The minimum sample rate is 50 seconds. The clock frequency of the PRBNS-signals should equal the sample frequency. However as the signals are entered by hand, this clock frequency can not be reached. The maximal frequency that could be reached is 0.008 Hz. The minimum time the PRBNS-signal remains constant is 2 minutes. At this frequency the spectrum of the PRBNS-signal is not white (see figure 8.1).

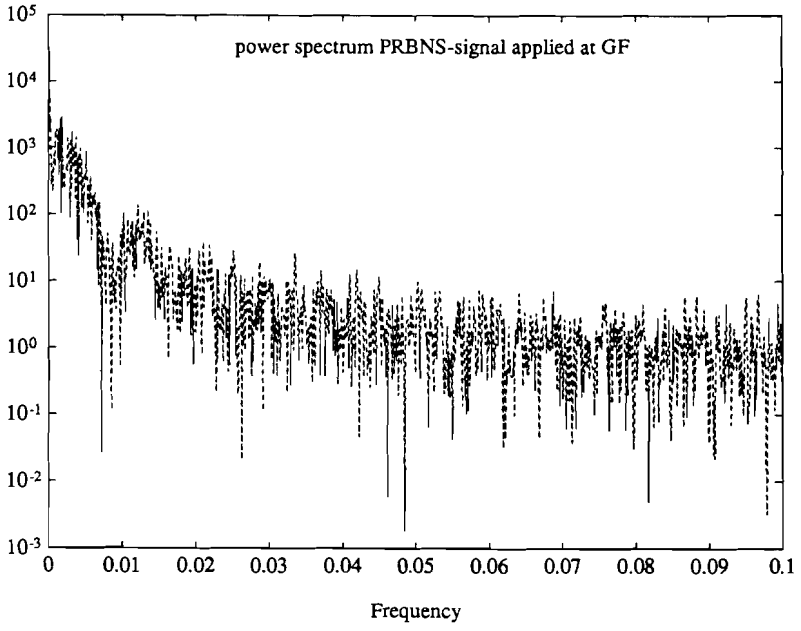

Figure 8.1. Power spectrum of the PRBNS-signal applied to the glucose flow.

The bandwidth of an input (GF) is almost identical to the bandwidth of the DOT (appendix C, figure C.6). This means that the DOT is not excited at all frequencies and therefore the estimation of the DOT in the model could be bad.

However it is thought this will not happen. To understand this one has to remember that the fast reaction of the DOT only occurs if the stirrer speed is changed. In § 6.2 this was called the pure physical process of dissolving the oxygen in the broth. The reaction of the biomass on the changing DOT is much slower. This was called the bio-chemical process. Only this last reaction is modeled. If the process is excited on all frequencies of the bio-chemical process the model can give a good prediction of the process. It is expected that the bio-chemical process can be approximated using a first order model [Axelsson, 1990]. In that case the largest time constant gives a rough indication of the bandwidth. The process is excited at these frequencies, so an estimation of the parameters should be possible.

At Unilever work is done to generate PRBNS-signals automatically. Then it will be possible to increase the clock frequency of the PRBNS-signals, so the process can be excited at all frequencies of interest. If new experiments are done it can be checked whether the bio-chemical process has a small bandwidth indeed.

To reach the minimum sample rate of 50 seconds the following measures are taken:
- Technical improvements are made to speed up the mass-spectrometer. The measurement cycle for all experimental sets is decreased from 6 to 3 minutes.

- Experiments are done when no or only a few other experiments are performed. The mass-spectrometer only analyzes the gas of the experiments that are active. If only one experiment is running (ours) the time needed for one measurement cycle is decreased to 50 seconds.

- The off-gas of the fermentor, that is analyzed by the mass-spectrometer, is also led to other, unused, inputs (figure 8.2). The gas is analyzed more times every measurement cycle of the mass-spectrometer. The time for 1 cycle increases (35 seconds for every extra input) but more samples are taken every cycle. This speeds up the measurements because every measurement cycle starts with a calibration. The more inputs are used the more samples are taken before a calibration takes place.

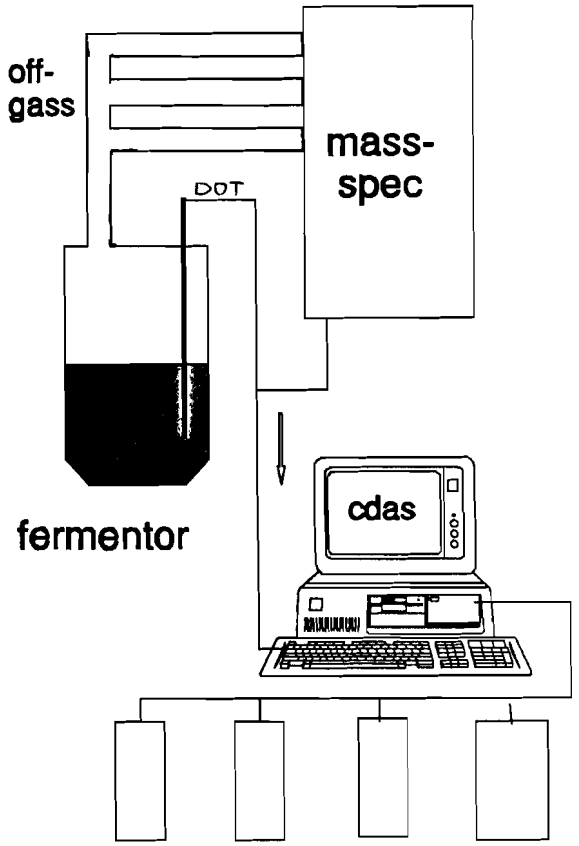

Figure 8.2. The setup if more inputs of the mass-spectrometer are used to sample the off-gas.

This is even necessary if only one experiment is done because CDAS logs the data of the mass-spectrometer every minute or a multiple of a minute. If every new sample has to be logged, the measurement cycle of the mass-spectrometer has to cover 1 minute or more. Otherwise new measurements can be done before the old ones are logged. To sample more than once in 50 seconds more samples have to be taken every measurement cycle.

It is tried to chose those inputs of the mass-spectrometer that the time between 2 samples is equal. This is not always possible because inputs that can be used have to be free. If this is not possible other inputs can be used as long as the time between 2 samples is less than 50 seconds. This will give no problem in the minimization routine because the Runge-Kutta method of solving the differential equations needs a start and stop time and these times are known. However analysis in frequency domain could give doubtful results. These analysis are done only with data sets whereby the time between 2 samples is equal.

The measurements of each input of the mass-spectrometer is stored in a file. As more inputs are used to sample the experiment, the data is divided over

more files. These files have to be combined into one file. This is done by taking one sample of each file and place them in the new file in the same order as they were measured. As the logging of CDAS and the measurements of the mass-spectrometer are not synchronized, data processing was necessary. As there is no information on the exact moment the samples are taken there is a possibility the data sets are not combined correctly. The equipment will be improved in order to log this moment too. If that is done better data sets can be obtained.

Four experiments were planned, but because we ran out of time only two were done. The original experiments were set up as follows:
Three experiments were designed, one for each growth way. The parameters in the model that do not used to simulate the process in that growth way are not estimated. This reduction of parameters simplifies the estimation procedure.
Finally the 3 sets could be joined in one large set to estimate all parameters. This can be done by placing them one after the other in the file. It will give discontinuities but the routine used for minimizing the output error can handle some discontinuities. One experiment was designed to change between the growth ways as much as possible. This would give good estimations for the parameters involved in the change over (enzyme functions).

The experiments done were set up as follows:
One experiment was done as was planned. The experiment started as the ethanol produced in the batch-phase, was consumed. The maximal glucose flow equaled the open loop setpoints and therefore no ethanol production occurred. The growth way was purely an oxidative metabolism of the glucose.
The other experiment started before the ethanol was consumed. Thus the experiment started with the oxidative reduction of ethanol. As the ethanol was consumed the process changed to the first growth way. At the end of the experiment the glucose flow was increased until ethanol production occurred. In this way all 3 growth ways occurred and all parameters are to be to estimate.

As discussed above several files containing the results of the measurements of the mass-spectrometer have to be combined into one file. The way this is done is described in appendix B, The results are given in appendix C, figure C.7.

8.2 Results.

Power spectra are calculated using 545 samples for CPR, OUR and E. DOT is not calculated as this is done in the previous experiment. The spectra are calculated using the function SPECTRUM.M. The results are given in appendix C, figure C.8. The bandwidth of the CPR and OUR are 1/4 minutes, as was expected. The smallest time constant is indeed the same as the largest one. All results are combined in table 8.1.

Table 8.1 Time constants, bandwidth and delay times of the process. All times are in minutes. BW stands for bandwidth, TC stands for the largest time constant and TΔ stands for the delay time.

	E	DOT	OUR	CPR
BW	1/16	1/1.6	1/4	1/4
TC.	*	2.1	4	4
TΔ	*	2	3	3

Normally the delay times are estimated using correlation techniques but these results are only meaningful if the process is linear. If a cross correlation between an input and output is calculated the reaction of the output due to an other input is averaged. This is possible if the inputs are mutually independent and the process is linear. However if the gains change during the process the reactions on other inputs are not averaged. Therefore the reactions of other inputs will be found in the results of a cross correlation reactions. It is not possible to use these techniques in this case.

Step responses have to be used to estimate the delay times. Visual inspection gives the time the output starts to react on the step applied at the input, this is the delay time. The final PRBNS-experiments are the only ones that have a sample frequency that is high enough to estimate the delay times. As steps on the inputs are applied simultaneously it is impossible to do this. Some mistakes are made during the experiments as wrong setpoints are entered. The reaction of the outputs on these very high or low setpoints can be recognized. This gives an indication of the delay times (table 8.1). Not all outputs react on the inputs and these delay times can not be estimated, for example the ethanol concentration does not react on all inputs, therefore no

good estimations can be made. During the estimation of the parameters, estimation has to be made using different delay times. The simulation that gives the best results will have the proper delay times.

8.3 Conclusions.

Two data sets are obtained that can be used for the estimation of the parameters. It is possible that the accuracy of the sets is not very high due to the synchronization problems. It can not be guaranteed that the combination of the sets is done correctly. It is possible that the samples are not placed in the proper order because the moment samples are taken is not known exactly. As synchronization is improved (The time the mass-spectrometer finishes the analysis of the gas is logged too) and PRBNS-signals can generated automatically, it is better to do these experiments again. A higher accuracy is possible this way.

The data sets obtained by these experiments will be used to test the minimizing routine and to check if it is necessary to improve this experiment design (for example if no minimum could be found).

9 Parameter estimation.

Parameters of the process are estimated using the output error criterion. The routine is tested first on a model to model base. Afterwards it is tried to estimate the parameters of the real process. This is tried using the dataset obtained in the final PRBNS-experiments.

9.1 Model to model simulations.

The output error is minimized as given in § 3.6 eq.(3.5). The minimize routine used is MINIQD[3]. It minimizes the error in the least squares sense using an algorithm that is selected by the user. The algorithm selected for this problem is the Marquardt method with numerically calculated derivatives. This method combines the fast descent in error of a Gauss-Newton method and the robustness of the steepest descent method. It calculates the minimum of the function:

$$\hat{\theta} = \min_{\theta} \left\{ \sum_{i=1}^{N} (y_i^t (\theta_t) - y_i^m (\theta))^2 / \sigma_i^2 \right\} \tag{9.1}$$

where $\hat{\theta}$ is the solution of the minimization, θ_t stands for the real process parameters, θ is the set of unknown parameters that are to be estimated, N stands for the length of the data set, y_i^t stands for the i^{th} sample of the measured output and y_i^m stands for the i^{th} sample of the simulated model output. σ_i stands for weight factor defined by the user.

As there are 4 outputs the following function is minimized (identical to eq.3.5):

$$\hat{\theta} = \min_{\theta} \left\{ \sum_{i=1}^{N} (f_{1,i}^2(\theta))/\sigma_1^2 + \sum_{i=1}^{N} (f_{2,i}^2(\theta))/\sigma_2^2 + \sum_{i=1}^{N} (f_{3,i}^2(\theta))/\sigma_3^2 + \sum_{i=1}^{N} (f_{4,i}^2(\theta))/\sigma_4^2 \right\}$$

and

$$f_{j,i}(\theta) = \left(y_{j,i}^t(\theta_t) - y_{j,i}^m(\theta) \right) \tag{9.2}$$

$f_{j,i}$ stands for the i^{th} residual of the j^{th} output, $y_{j,i}^t$ stands for the i^{th} sample of the measured output j and $y_{j,i}^m$ stands for the i^{th} sample of the simulated model output j. σ_j stands for the noise variance of the j^{th} output

[3] MINIQD is a library routine of the RC Eindhoven University of technology to minimize least squares problems. The user can select different minimizing algorithms with or without analytical derivatives.

as estimated in the free run experiments. Note that the noise variance is used as a weight function. This way 3 discontinuities in the dataset are introduced, because the error is calculated for each output, one after the other. However the algorithm is robust enough to find a minimum anyway.

To calculate the model output the model described in § 2.2 is used. The model is also used to calculate a dataset to test the routine. The inputs of the test set are the same as the inputs that are used for the first PRBNS-experiments. However the sample frequency was too low in these experiments. The sample frequency used for these simulations is 11 seconds. This means that samples actually measured every 6 minutes are used every 11 seconds. The relative and absolute accuracy, that are input parameters for the MINIQD routine, are both 1.10^{-5}. Starting values for the parameters are the expected values.

The data set is given in appendix C, figure C.9. No noise was added. In this set no ethanol production occurred. It is expected that this would make it impossible to estimate the parameters involved with ethanol production or consumption (K_e, K_i, Q_e and τ_e). This expectation was confirmed by the result of the minimization. No minimum could be found. However if these 4 parameters were not estimated but kept at the expected value, the minimum could be found with a variance of 1.10^{-7}. The report created by the routine is given in appendix D. Stop criterion 915 means no minimum could be found (see [manual TUE-RC-68438, 1989]). Note that the parameters are divided by the expected values because it improves the minimization (see § 3.6).

The model to model estimation is done using different starting parameters. The start value of the normalized parameters are in one simulation 1.2 and in an other 0.8. The minimum could always be found. To make ethanol production possible the glucose flow was increased a factor 30. Now ethanol was produced as can be seen in appendix C, figure C.10. The minimum is calculated using all 16 parameters. Again 3 simulations are done using the same starting values as before.

Here also the minimum could be found with the accuracy of 1.10^{-7}. Therefore it is expected that the experiment design, using PRBNS-signals that are applied at the inputs, is sufficiently good for minimizing the output error of the

real process.

9.2 Estimation of the parameters of the real process.

The final PRBNS-experiment designed for keeping the process in the first growth way is used for the estimation of 12 parameters. The parameters involved in
ethanol production or consumption can not be estimated. This was also a result of the simulations. The data is processed as described in appendix D. The delay times of the process outputs were estimated during the final PRBNS-experiments. However this could not be done for the ethanol output. The estimated delays are neutralized by shifting the output data. The dataset is then 3 samples shorter. As the delay of the ethanol output is not known, its shifted the same number of points (3) as the other outputs that are measured by the mass-spectrometer.

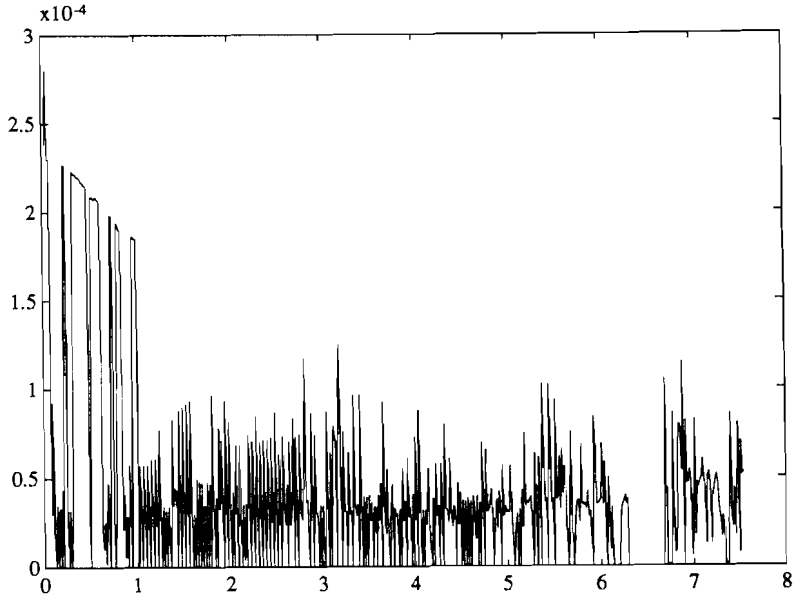

Figure 9.1. A wrong solution of the differential equation, using D02BAF.

During the minimization problems occurred using the routine that solves the differential equations. This routine of the NAG-library (D02BAF) solves equation according the Runge-Kutta Merson method, that uses a variable number of steps to find the solution with the required accuracy. The sample time of the dataset is 50 seconds. The time is too long to solve the differential equations reliably (see figure 9.1 for an example of the DOT output).

Therefore the step is divided in 10 small steps of 5 seconds each. The routine is called 10 times to solve the equations. This way the solution is reliable. To ensure this, the equations are also calculated using the MATLAB routine ODE45.M that uses the Runge-Kutta method too. The results of both simulation is given in figure 9.2 and are identical. The small changes are due to the higher accuracy of ODE45.M. However ODE45.M is not used because the routine has to be used not only for the simulation of the process but also for the minimizing routine. This routine is written in Fortran. Therefore the Fortran routine of the NAG-library (D02BAF) is preferred.

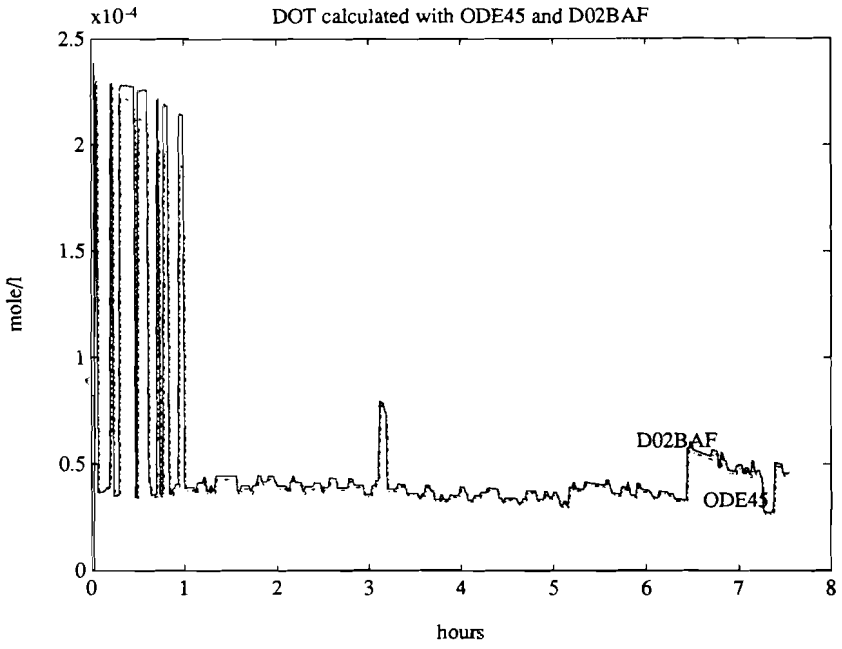

Figure 9.2. The solutions of the differential equation after improvements calculated by D02BAF (_._.) and ODE45.M (—).

If the normalized starting values of the parameters are 1, a minimum can be found with a variance of 1.10^{-4}. The estimated parameters are still 1. However the simulation with these parameters gives bad results as can be seen in appendix C, figure C.11. The minimum is a local minimum. Note that the dynamic behavior of the DOT output is simulated correctly. The error is due to the wrong estimation of the width of the bottleneck. Therefore DOT is too low and ethanol production occurs in the simulation. If the estimation is started with other parameters, the routine stops giving a floating point

overflow. This means the problem is singular.

If the simulation is run with the last set of parameters MINIQD used before a floating point overflow occurred, a overflow is given too. The overflow occurs in solving the differential equations and not during the search for a minimum. The differential equations are probably almost singular. The NAG-library manual suggests to find a analytical solution of the set. This however is very hard to do and therefore it can not be used.

The problem does not occur always. Therefore if MINIQD tries only a small number of parameters sets it should be possible to find a minimum before a set of parameters is tried that gives a floating point overflow. This means that the user should try to find a set of parameters that is close to a minimum. It is hoped that the routine will find the minimum because it will not need many steps. The chance on an overflow is small.

Another solution could be to use a routine of the NAG-library that can be controlled better. There is a routine (D02EAF) that checks for a possible floating point overflow. Just before an overflow occurs the routine can be stopped. Depending on how far it was from the end, the error in the solution can be very large. But maybe a minimum of the output error can be found anyhow.

It is possible other solutions for the problem can be found. The best solution must be tried to minimize the output error. This has to be done before the parameters can be estimated.

10 Conclusions and recommendations.

A number of experiments are done to obtain a dataset that is 'rich' enough to estimate the unknown parameters of the process. It was found that the time constants of the process were much smaller than was expected. The DOT has time constants of 1.66 minutes and the other outputs near 4 minutes. This is about 6 times less than expected.

Sample times of the mass-spectrometer and database are not synchronized nor is the moment recorded the mass-spectrometer finishes the analysis of the gas. As these moments are not known estimations are made of the moments the data is measured. The accuracy of these estimations is not very high. It could be improved by synchronizing the sampling times, or if that is very difficult by logging the moment the mass-spectrometer finishes the measurements.

As the time constants are smaller than expected, the sample frequency needed to be much higher. Especially the mass-spectrometer could not reach that high frequency. More inputs of the mass-spectrometer had to be used. This enlarged the synchronization problems. The more inputs of the mass-spectrometer are used the more difficult it is to synchronize all files. This decreases the accuracy even more.

The stirrer speed is not controlled. Measurements show an error between the setpoint and the measured speed. A feed-back controller or at least a measurement device should be implemented to minimize errors. A feed-back control minimizes the error between the setpoint and the real speed. If only the measurement device is implemented the real input is measured and that minimizes the error due to the uncertainties in the input.

The measurement of the glucose flow has a low signal to noise ratio due to the quantization noise. This could be improved if the Applikon has a digital input for the signals of the balance. This will give less noise than converting the signal from digital to analog and then back from analog to digital as is done now.

If the problems discussed above, are solved it is suggested to repeat the final PRBNS-experiments to increase the reliability of the estimation.

Model to model simulations give good results for estimation of the parameters. Therefore it is believed the input signals are rich enough to estimate the unknown parameters. This means the experiment design of the final PRBNS-experiments is good.

Until now no results of the estimation are available because problems occurred solving the differential equations of the model as they are almost singular. No solution for this problem is found yet. One has to find a better way for solving the equations.

References.

Axelsson, J.P. (1989) *Modelling and control of fermentation processes.* Ph. D. Thesis, Lund University of Technology.

Backx A.C.P.M., and A.A.H. Damen (1989) *Identification of industrial MIMO processes for fixed controllers. Part 1 General theory and practice.* Journal A, (1) pag. 3-12.

Gill, P.E. and W. Murray (1978) *Algorithms for the solution of the nonlinear least-squares problem.* SIAM J. Numer. Anal., (5) pag. 977-993

Keulers, M. (1988) *Modelling and self-tuning control application for fed-batch Bakers' Yeast fermentation.* Unilever report.

Keulers, M. (1988) *More data and figures for the master degree thesis 'Modelling and self-tuning control application for fed-batch Bakers' Yeast fermentation'.* Unilever report.

Marquardt, D.W. (1963) *An algorithm for least-squares estimation of nonlinear parameters.* J. Soc. Indust. Appl. Math., (11), pag. 431-441.

Mols, I.M. (1990) *Appendices for the master degree thesis 'The estimation of the parameters of the fed batch Bakers' Yeast process'.* Unilever report.

Nelder, J.A. and R. Mead. *A simplex method for function optimization.* Computer J., (7), pag 308-313.

Scales, L.E. (1985) *Introduction to non-linear optimization.* (MacMillan computer science series). London, MacMillan.

Sonnleitner, B. and O. Käppeli (1986) *Growth of saccharomyces cerevisiae is controlled by its limited respiratory capacity: formulation and*

verification of a hypothesis. Biotechnology and bioengineering, (28), pag 927-937.

Sweere, A. (1988) *Response of Bakers' Yeast to transient environmental conditions relevant to large-scale fermentation processes*. Ph. d. thesis, Delft University of Technology.

Van den Boom, A.J.W. and A.A.H. Damen. (1987) *Stochastiche systeem theorie* Internal lecture notes, Eindhoven University of technology.

List of symbols.

URL = Unilever Research Labatory.

h = Hours.
l = liter.
min = minute.

θ = Parameter set.
$\hat{\theta}$ = Estimated parameters.
θ_t = Real process parameters.
y_i^t = i^{th} measured output sample.
y_i^m = i^{th} sample of the simulated output.
σ_i = Weight factor for the i^{th} output.

AF = Air flow.
GF = Glucose flow.
SS = Stirrer speed.
E = Ethanol concentration in the broth.
CPR = Carbon dioxide production rate.
OTR = Oxygen transfer rate.
OUR = Oxygen uptake rate.
P = Concentration of dissolved oxygen.
DOT = Dissolved oxygen concentration in procents.
G = Glucose concentration.
X = Biomass concentration.
X_o = Biomass concentration at the beginning of the experiment.
V = Culture volume.

μ_i = Growth rate of the biomass.
σ = Consumption rate of glucose.
π_i = Production or consumption rate of ethanol.
o = Consumption rate of oxygen.
γ_i = Production rate of carbon dioxide.
K_g = Saturation parameter for glucose uptake.
K_o = Saturation parameter for oxygen uptake.

K_e = Saturation parameter for growth on ethanol.

K_i = Inhibition parameter: free glucose inhibits ethanol uptake.

K_n = Glucose saturation parameter for the induction . of the production of glucose consumption capacity.

K_m = Substrate saturation parameter for the induction of the production of oxidation consumption capacity.

Q_{gm} = Maximum specific glucose consumption rate.

Q_{om} = Maximum specific oxygen uptake rate.

Q_{em} = Maximum specific ethanol consumption or production rate.

Q_{gmax} = Specific glucose consumption rate.

Q_{omax} = Specific oxygen uptake rate.

Q_{emax} = Specific ethanol consumption or production rate.

τ_g = Time constant for glucose uptake.

τ_o = Time constant for oxygen uptake.

τ_e = Time constant for ethanol uptake.

to = Time at the beginning of a step in a staircase experiment.

te = Time at the end of a step in a staircase experiment.

ΔL = Difference between output level at to and te.

C_1 = Constant.

C_2 = Constant.

C_3 = Constant.

Ps = Saturated dissolved oxygen concentration.

Gf = Glucose concentration of the glucose flow.

Ms = Maintenance term of biomass.

T = Temperature.

PRBNS= Pseudo Random Noise Sequence.

a_i = Amplitude of PRBNS-signal of the i^{th} input.

PS^* = Pump setting.

G_{ij} = Gain of input j to output i.

BW = Bandwidthe.

TC = Largest time constant.

$T\Delta$ = Delaytime.

Eindhoven University of Technology
Department of Electrical Engineering
Measurement and Control section

APPENDICES FOR THE MASTER DEGREE THESIS 'THE ESTIMATION OF THE PARAMETERS OF A FED-BATCH BAKERS' YEAST FERMENTATION PROCESS'

I.M. Mols

Date: August 1990

This report contains the appendices not supported by the master degree thesis of I.M. Mols.
The contents of the report is:

Appendix A. Experiment numbers and conditions.
Appendix B. Data processing.
Appendix C. Data sets of the experiments.
Appendix D. Report of the results of the minimizing routine MINIQUAD.
Appendix E. Fortran Routine ESTIM.FOR that is used for simulation.
Appendix F. Results of the study of sensors and actuators.

Appendix A. Experment numbers and conditions.

In this appendix all experiments are given. In the first column the name of the data file is given. In the name E stands for experiment, then the date of the experiment is given and then the number the file is registered by URL. In the second column a short explanation of the purpose of the experiment is given. There were a lot of problems with the equipment, especially at the beginning (until May). Air valves blocked, glucose pumps did not function etc.. Therefore some data could not be used. The disturbance was too large.

In the second part a description of the starting values of the process are given in the batch report that is generated during the experiments at URL.

Experiment name	Description of experiment	Remarks
E_27_2_900419	Free Run	not used, large disturbance
E_6_3_900427	GF step	not used, instead stair case
E_21_3_900480	Free Run	Not used, large disturbance
E_28_3_900489	Free Run	chapter 5
E_3_4_900490	AF staircase	chapter 6, also used in 5
E_10_4_900499	GF staircase	not used, large disturbance
E_19_4_900509	GF staircase	chapter 6
E_2_5_900532	SS staircase	chapter 6, also used in 5
E_7_5_900551	SS staircase	not used, large disturbance
E_10_5_900555	SS staircase	not all data recovered
E_6_7_900635	First PRBNS	chapter 7
E_6_7_900636	First PRBNS	not used, low sample rate
E_17_7_900655	Final PRBNS	not used, DOT not sampled
E_17_7_900656	Same Final PRBNS	second input mass-spec.
E_17_7_900657	Same Final PRBNS	third input mass-spec.
E_20_7_900660	Final PRBNS	chapter 8, E. production
E_20_7_900662	Same Final PRBNS	second input, disconnected
E_20_7_900663	Same Final PRBNS	third input
E_20_7_900665	Same Final PRBNS	second input, 1 hour later
E_20_7_900667	Same Final PRBNS	first input, starts after 6h
E_27_7_900672	Final PRBNS	chapter 8, 9, no E prod.
E_27_7_900675	Same Final PRBNS	second input

Batch report of the experiments. For all experiments these conditions are maintained constant.

Experiment date : 06-07-1990
Experiment number : 90....

>>>>> BATCH REPORT <<<<<<

BATCH DEFINITION

Micro-organisme : *SU32*
Product / Aim : *PRBNS for estimation of parameters.*

Total work volume (L) : *8.000*
Class of fermentation : *GRAS*

Responsible person : *Marc/Ivo*
Project number/name : *170090*
CDAS recipe : *FBMARC*

INITIAL SETTINGS

Start weight of biomass	: 66	g
Growth rate	: 0.15	1/H
Estimated yield	: 0.45	YSX
Substrate concentration	: 0.44	gS/gF

MEDIUM COMPOSITION FOR FED-BATCH Phase

Medium : *FEED0001*
Amount : *4.000 Kg*

Ingredient	Quantity	Units	Check
Glucose. 1aq	1760	g	.....
(NH4)2SO4	40	g	.....
KH2PO4	16	g	.....
K2HPO4	16	g	.....
CaCl2 **(ster. seperate)**	3.2	g	.....
MgSO4. 7aq	24	g	.....
Tracement Sol. (EGLI)	64	ml	.....
Vitamin cocktail (EGLI)	16	ml	.....

Note : **Steralization 20 minutes at 120 C. (not the tracement sol.)**
Vitamins filter sterilize.
Medium set at Ph = 3.0 before sterilization.

PARAMETER SETTINGS FOR FED-BATCH Phase

Temperature	: 30	(C)
Ph	: 5.0	
Stirrer speed	: 750	(RPM)
Air Flow	: 4	(L/Min)
DO2	: 0.0	(% of saturation)
Run time	: 24.0	(H)

CONTROLLER STATUS (1 = MANUAL, 2 = AUTO, 3 = CASCADE)

DO2	: 1
Anti foam	: 2
Ph	: 2
Temperature	: 2

Appendix B. Data processing.

In order to get a sufficiently high sample frequency more inputs of the mass-spectrometer are used. For each input a file is generated containing the data. These files have to be processed in order to get one file containing the final data set. The way this is done is described here.

- At 27 July 1990 a experiment is done, using 2 inputs of the mass-spectrometer. The following data files are generated:

900672.EX4.	In this file the operating actions are given.
900672.EX6.	The temperature, pH, glucose flow setpoints for open loop experiments and the doses of ammonia is logged every 16 minutes.
900672.EX7.	Signal of the balance, stirrerspeed and airflow are logged every minute.
900672.EX8.	OUR, CPR, RQ, E, DOT, CO_2, O_2 are logged every minute. This file contains the analysis of the gas, measured at input 12 of the mass-spectrometer.
900675.EX8.	OUR, CPR, RQ, E, DOT, CO_2, O_2 are logged every minute. This file contains the analysis of input 2.
900672.FSC.	This file contains the DOT sampled every 5 seconds.

- Using the Vax editor all unnecessary information (date, heading etc.) is removed of the files with extension EX6, EX7, and EX8. These files are flat ASCII files now. The experiment starts as all log files start. The data in log files that start before the last one are removed until the start time. These files are ready to be read by MATLAB.

- The setpoints of the glucose flow is obtained by extracting time and setpoints from the file 'operator actions' (900672.EX4) using the program INPUT that is written in C for this purpose. The glucose flow is calculated as indicated in § 4.2.3. The measured glucose flow is calculated from the balance signal that is logged by CDAS. This is done using the function WEEG.M[1], which differentiates the signal by dividing the difference of two proceeding samples of the balance by the difference of the times the two samples are taken. The

[1] Procedures mentioned in this section are written especially for these experiments.

resulting file is given in figure B.1. The pump setpoints are fitted to the measured glucose flow using eq. (4.2) (also included in figure B.1).

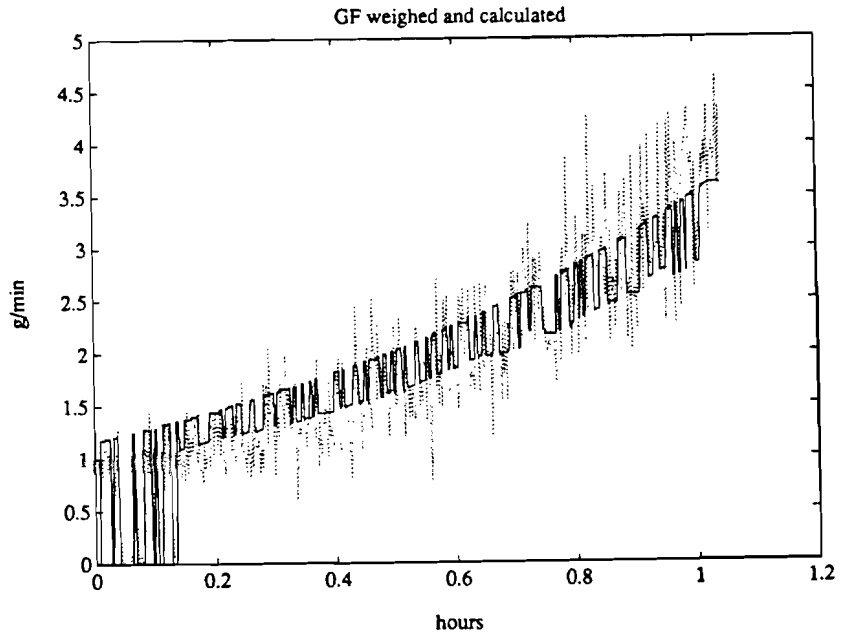

Figure B.1. The weighed glucose flow and the calculated using the setpoints of the pump.

- The inputs are sampled every minute. To get the same sample frequency as used by the outputs, more samples are added. This is done by copying the previous data sample using the procedure INPUT_UP.M.

- CDAS does not log the time the mass-spectrometer finishes the analysis of the off-gas. Therefore the time generated by CDAS can not be used for the data sets. A new time axis has to be made that represents the time that samples are taken by the mass-spectrometer. During the experiments the time of one measurement cycle of the mass-spectrometer is recorded (by hand) and is 1:40 minutes. The axis starts at 0, an other point in the time axis is the indices of that point multiplied with the cycling time(see table B.1). Time axis for the other set is made by adding a delay to the time axis. This delay (50 seconds) is noted during the experiments too. Note that the inputs are chosen in a way that the sample time is 50 seconds.

- The cycling time of the mass-spectrometer is more than a minute in order to log all the measurements. The log time of CDAS is 1 minute. Therefore CDAS sometimes logs the result of the mass-spectrometer twice. These double

loggings have to be removed. The last 2 steps are combined in one procedure TIME_REDUCTION.M. The data belonging to a moment of the time axis is the first by CDAS recorded data after this moment. For example the new time axis is made using a sample time of 1.66 minutes (1:40 minutes), the data of the mass-spectrometer is logged every minute. The moments the same data is logged is given in the same column of table B.1.Double samples are samples 0 and 1, 2 and 3, 5 and 6, 7 and 8, 10 and 11. The last points are removed.

Table B.1 The first moments the mass-spectrometer ends an analysis and the moments they are logged.

Time axis (minutes)	0	1.67	3.33	5.00	6.66	8.33	10.00
Logged time (minutes)	0 1	2 3	4	5 6	7 8	9	10 11

- As the DOT in the files 900672.EX8 is sampled every minute and that is not at the moment the mass-spectrometer samples, it can not be used. The DOT in file 900675.EX8 is not sampled at all. Therefore the DOT is sampled fast (every 5 seconds in file 900672.FSC). As the recorded time of the DOT is correct, those samples can be selected that are measured at the same moment the mass-spectrometer finishes the analyses.

- The air flow is measured. These measurements are used to calculate the CPR and OUR. As they are not logged in an file 900675.EX7 ,the CPR and OUR in file 900675.EX8 have to be calculated off line using eq.(4.3).

- The gas analyses done at other inputs of the mass-spectrometer have a large influence on the ethanol measurement of our inputs. The inputs are not washed with a gas of a high temperature but with gas of a lower temperature to speed up the mass-spectrometer. If file 900672.EX8 and 900765.EX8 are combined a large noise signal is introduced due to the difference in the level (see figure B.2). Therefore these signals have to be normalized to reduce the noise.

Input 1 is used for analyses of another experiment, input 3 to 11 are not used. The analyses of input 12 is supposed to be correct because the previous input (input 2) is used for our experiment too. As those measurements are almost identical the influence can not be large. If the ethanol concentration

of our experiment is relatively low, the measured concentration by input 2 is higher than the one measured by input 12. But if the concentration is relatively high the result of input 2 is smaller. At about 30 PPM the results of both inputs are equal.

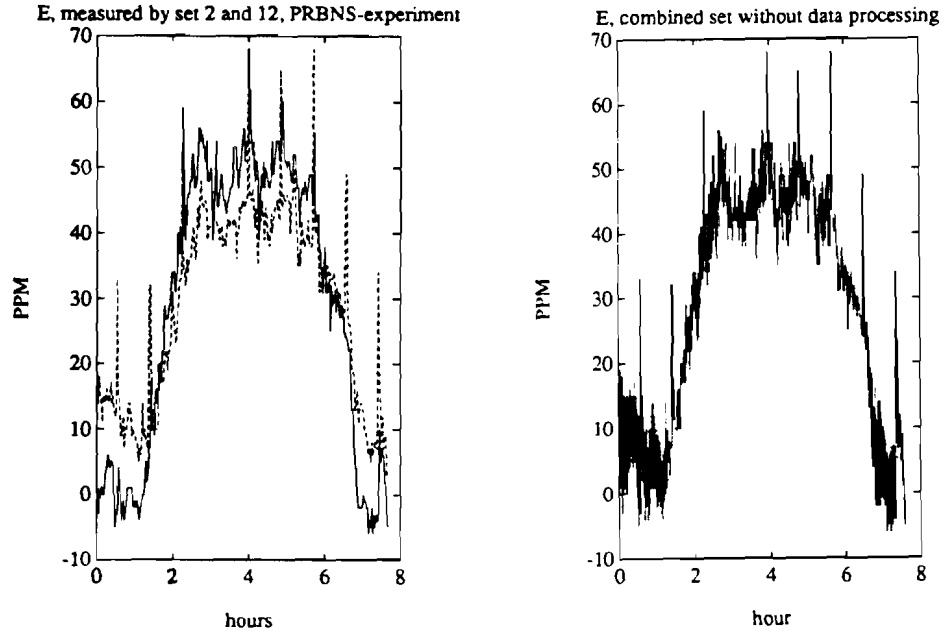

Figure B.2. The ethanol measured using two inputs. A) both sets with large difference in level. b) Combined set without signal processing.

The correction procedure is as follows: Of both data sets 30 PPM is subtracted. Then data set of input 2 is normalized by dividing it by the other dataset and multiplying the it by the coefficient, so it fits the results of input 12, afterwards 30 PPM is added again. This way the noise is reduced significantly as can be seen in appendix C, figure C.7.

- All time axes and data sets are copied in one file each. The exact way is not important as the time axis is sorted using the function SORT.M. This function also gives the indexes used to sort the time axis. The data sets are sorted using these indexes.

- The data set is peak shaved. This is necessary because the calibration of the mass-spectrometer (approximately every 2 hours) has a large influence in the data samples just after the calibration (see figure B.2). These spikes are removed using the peak shaving function in IPCOS.

Appendix C. Data sets of the experiments.

In this appendix the outputs of the experiments done are given. The appendix contains:

C.1 Output of the noise experiments.
C.2 The set used to determine the noise characteristics.
C.3 The power spectra of the noise set.

C.4 Outputs of the staircase experiments.
- A) Staircase input applied at the glucose flow (GF).
- B) Staircase input applied at the stirrer speed (SS).
- C) Staircase input applied at the air flow (AF).

C.5 Outputs of the first PRBNS-experiments.
C.6 Power spectra of the first PRBNS-experiments.

C.7 Outputs of the final PRBNS-experiments.
C.8 Power spectra of the final PRBNS-experiments.

C.9 Dataset use to simulate the model to model estimation, without ethanol.
C.10 Dataset use to simulate the model to model estimation, with ethanol.

C.11 Simulation with the estimated parameters. Local minimum.

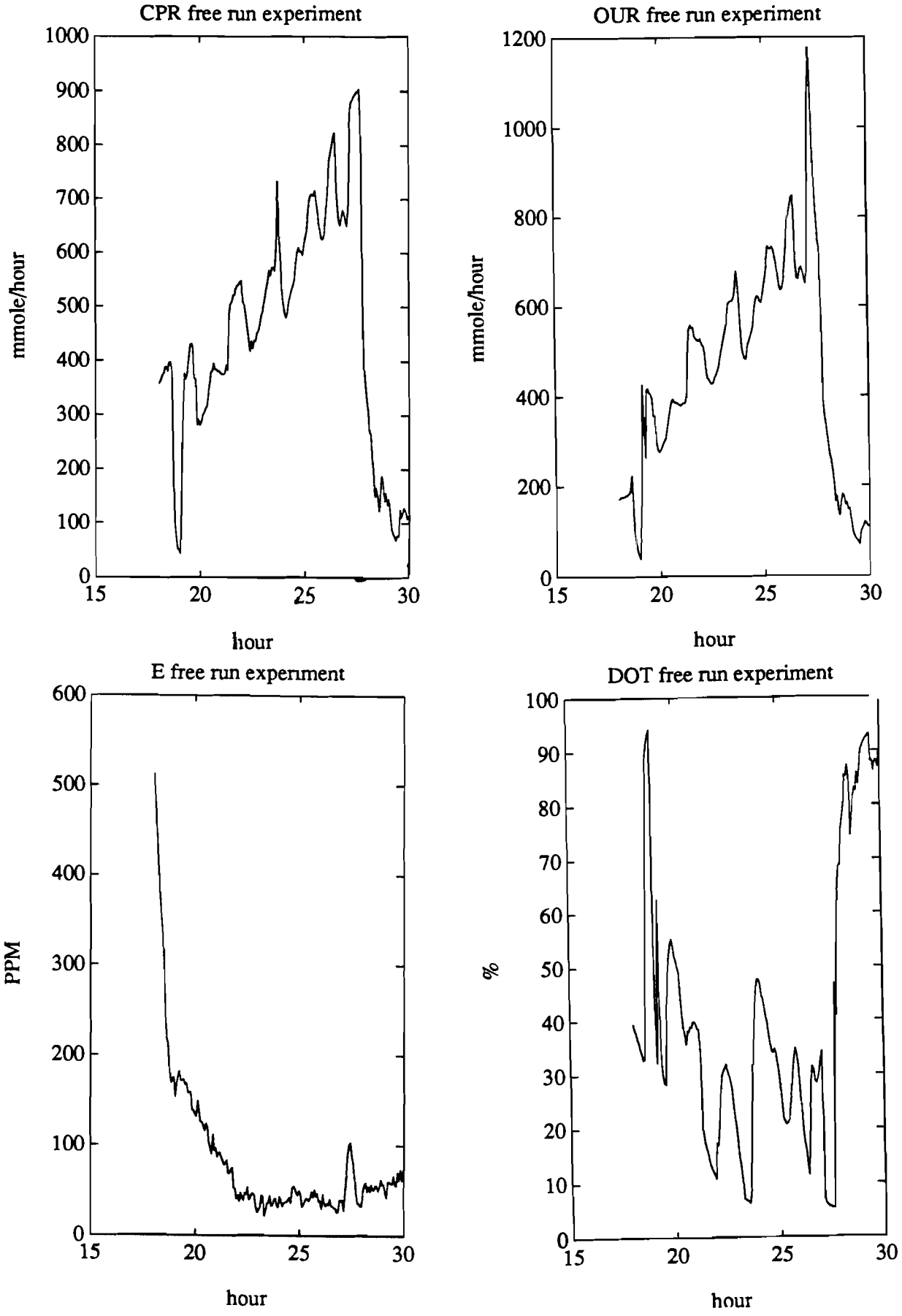

Figure C.1. Outputs obtained in the noise experiments. Large influence of the GF can be seen especially in the DOT, CPR and OUR.

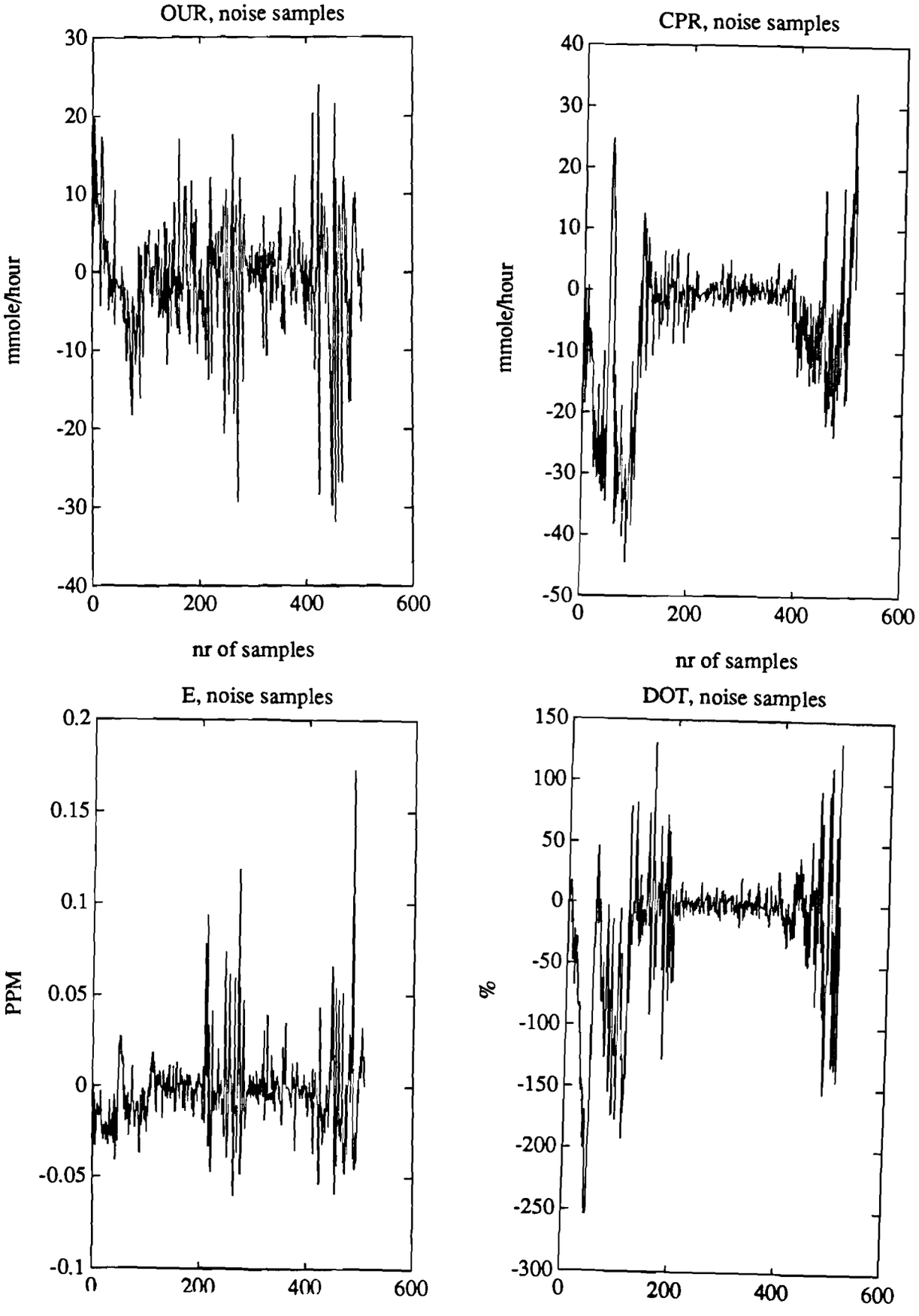

Figure C.2. The combined dataset for the estimation of the noise characteristics. The data is obtained in the staircase experiments where the input were applied at the SS and AF. At the end of the experiments all inputs were kept at a constant level.

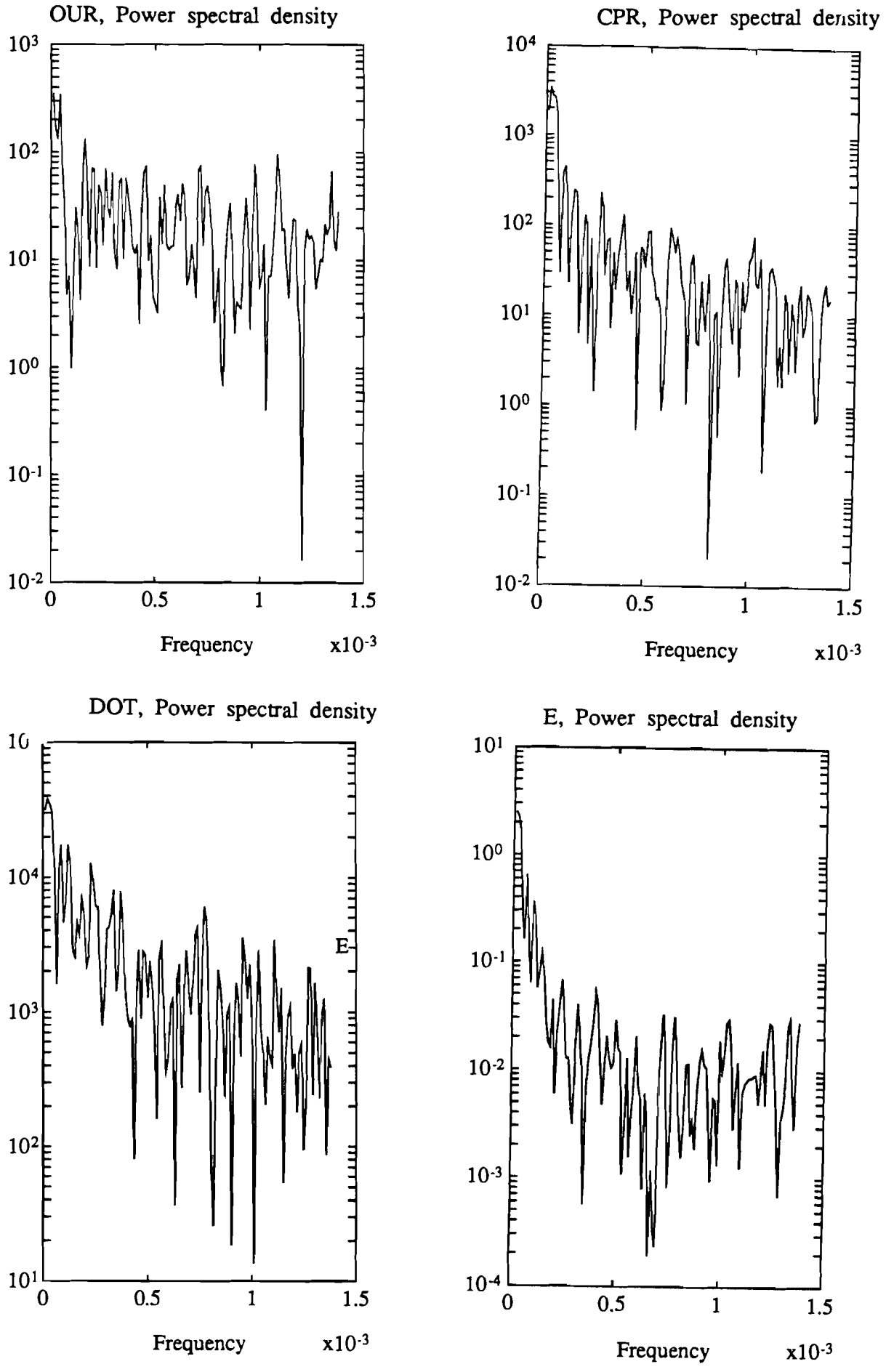

Figure C.3. The power spectra of the noise set as it is given in figure C.2.

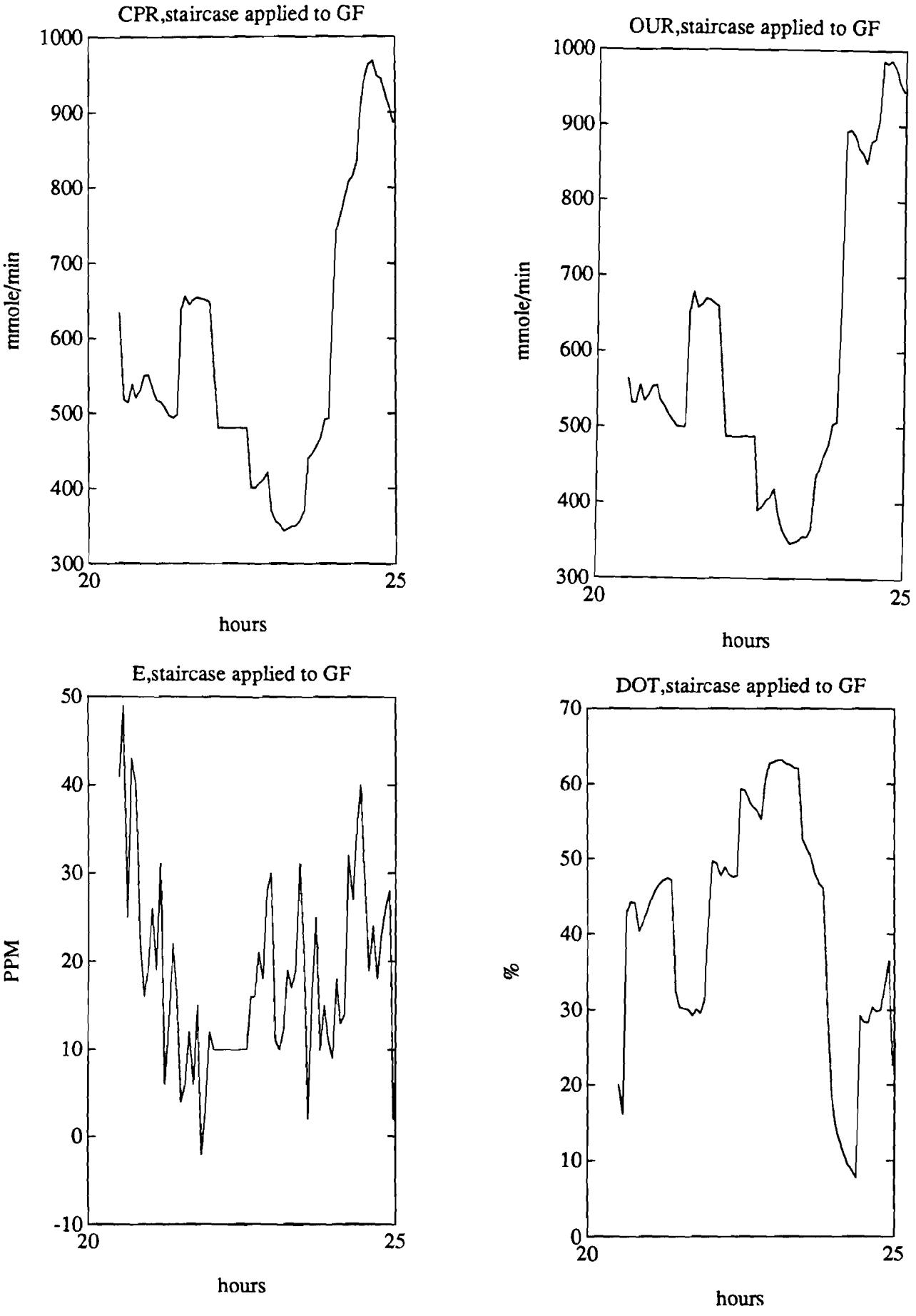

Figure C.4. A) Outputs of the staircase experiment, signal applied at the GF.

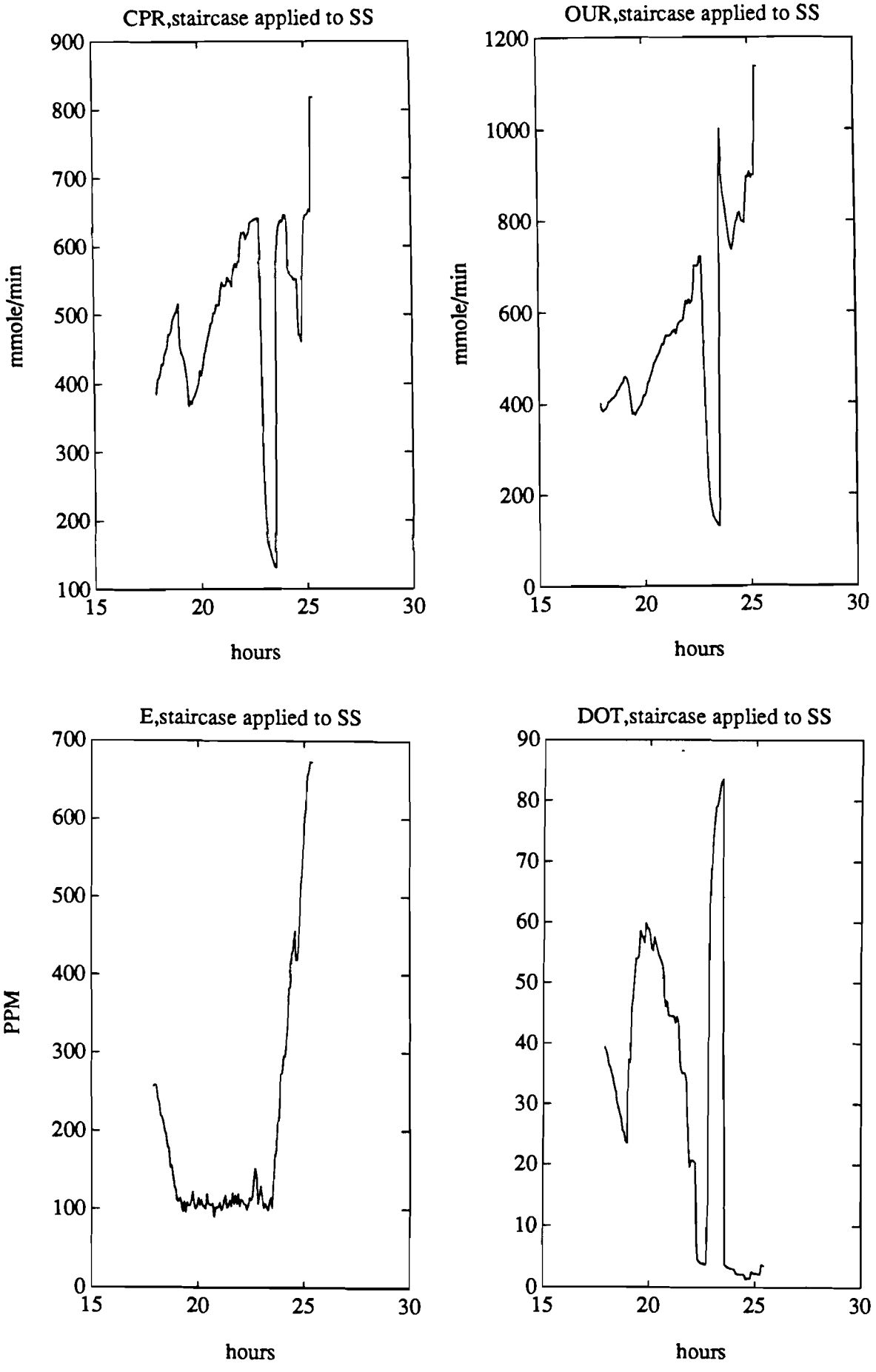

Figure C.4. B) Outputs of the staircase experiment, signal applied at the SS

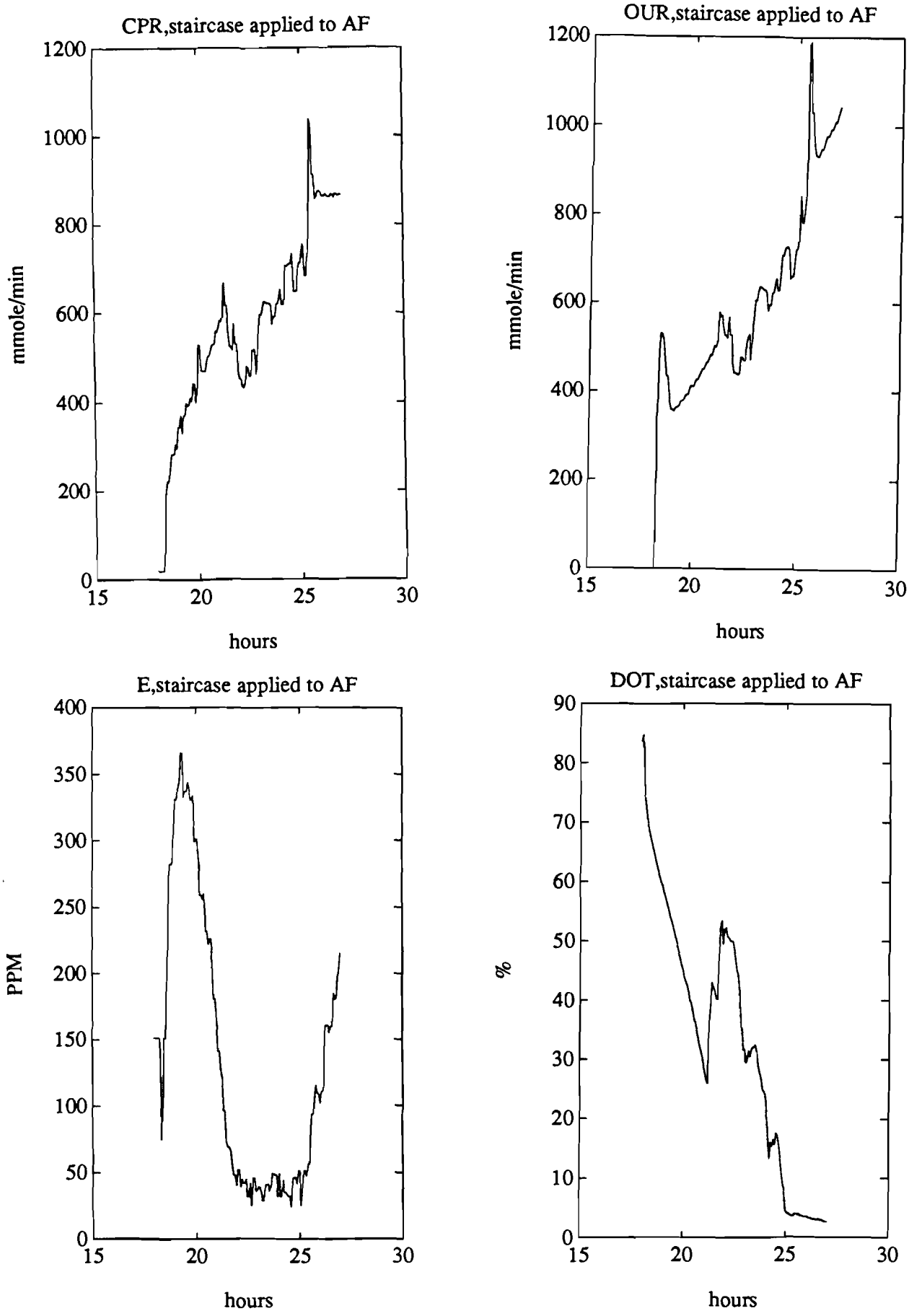

Figure C.4. C) Outputs of the staircase experiment, signal applied at the AF.

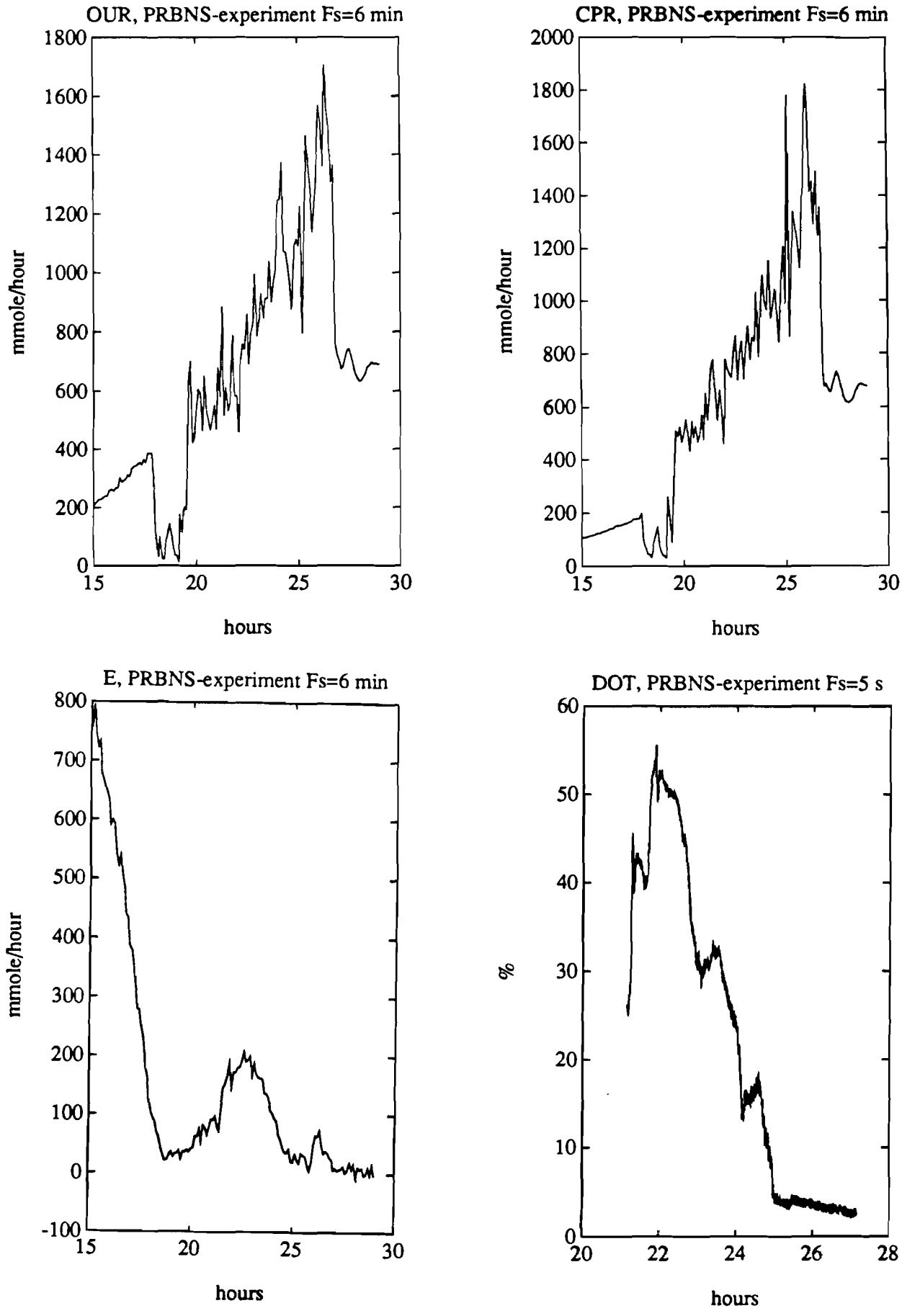

Figure C.5. Outputs of the first PRBNS-experiment.

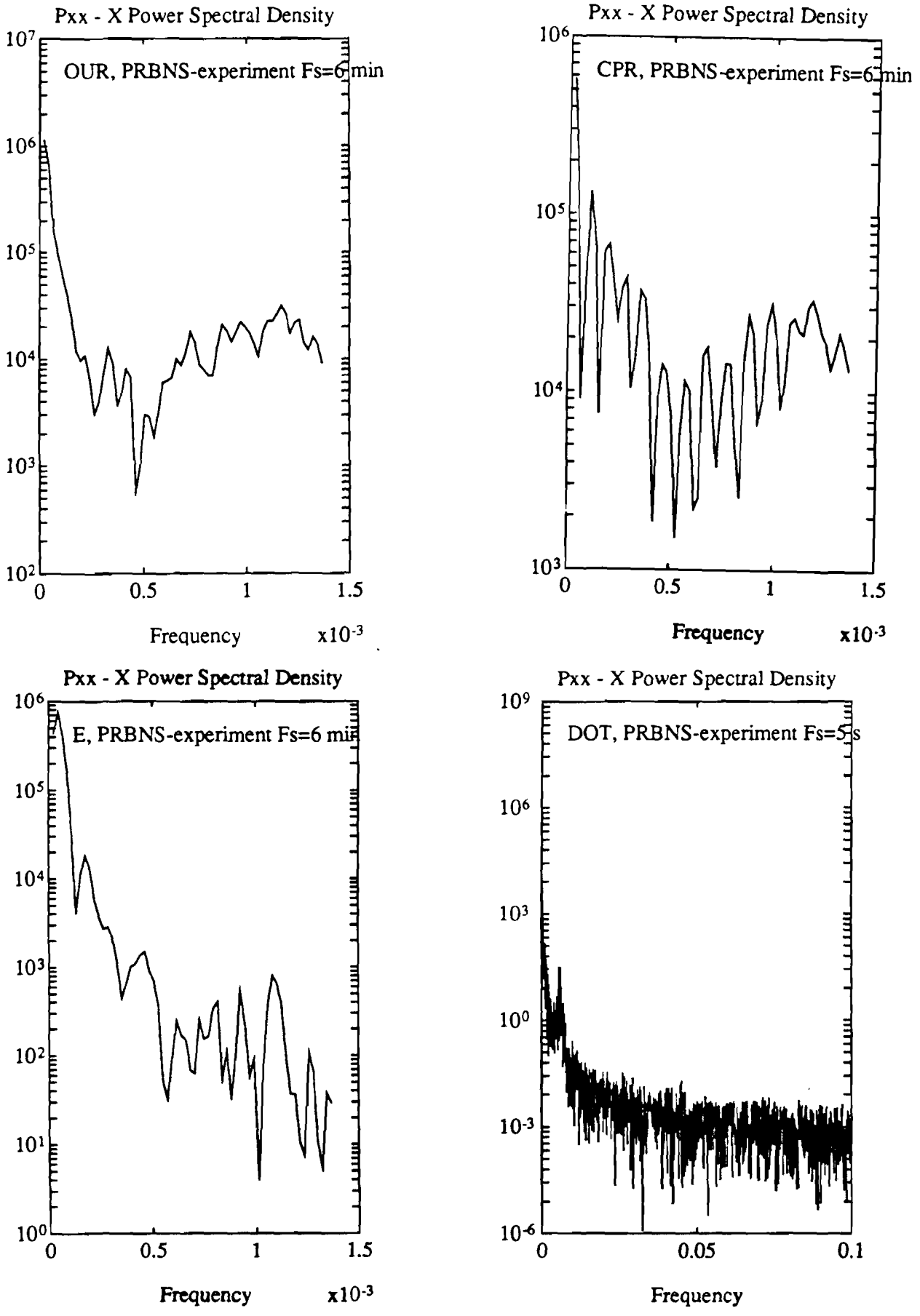

Figure C.6. Power spectra of the outputs of the first PRBNS-experiments.

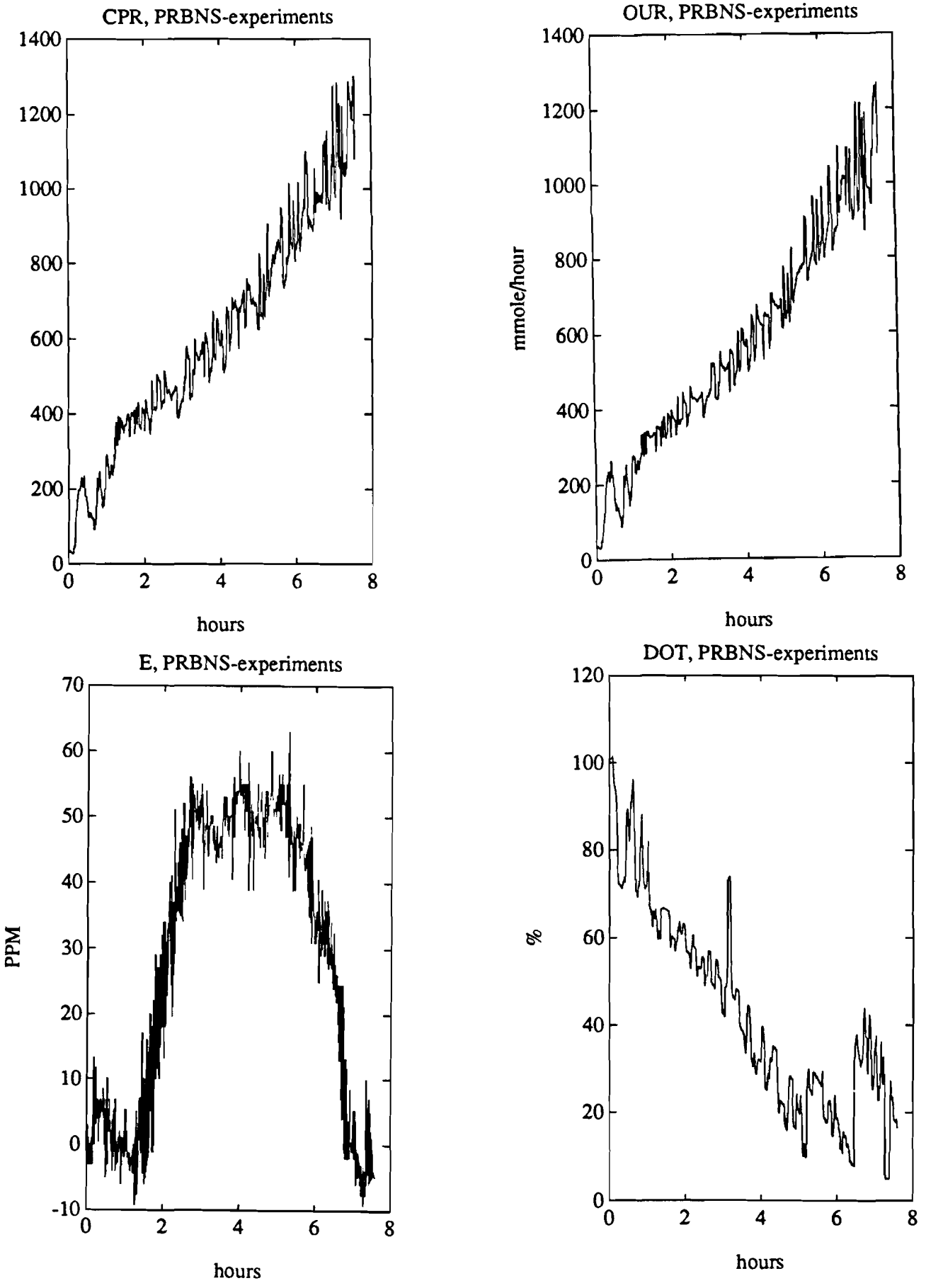

Figure C.7. Outputs of the final PRBNS-experiment. This set is also used for the parameter estimation.

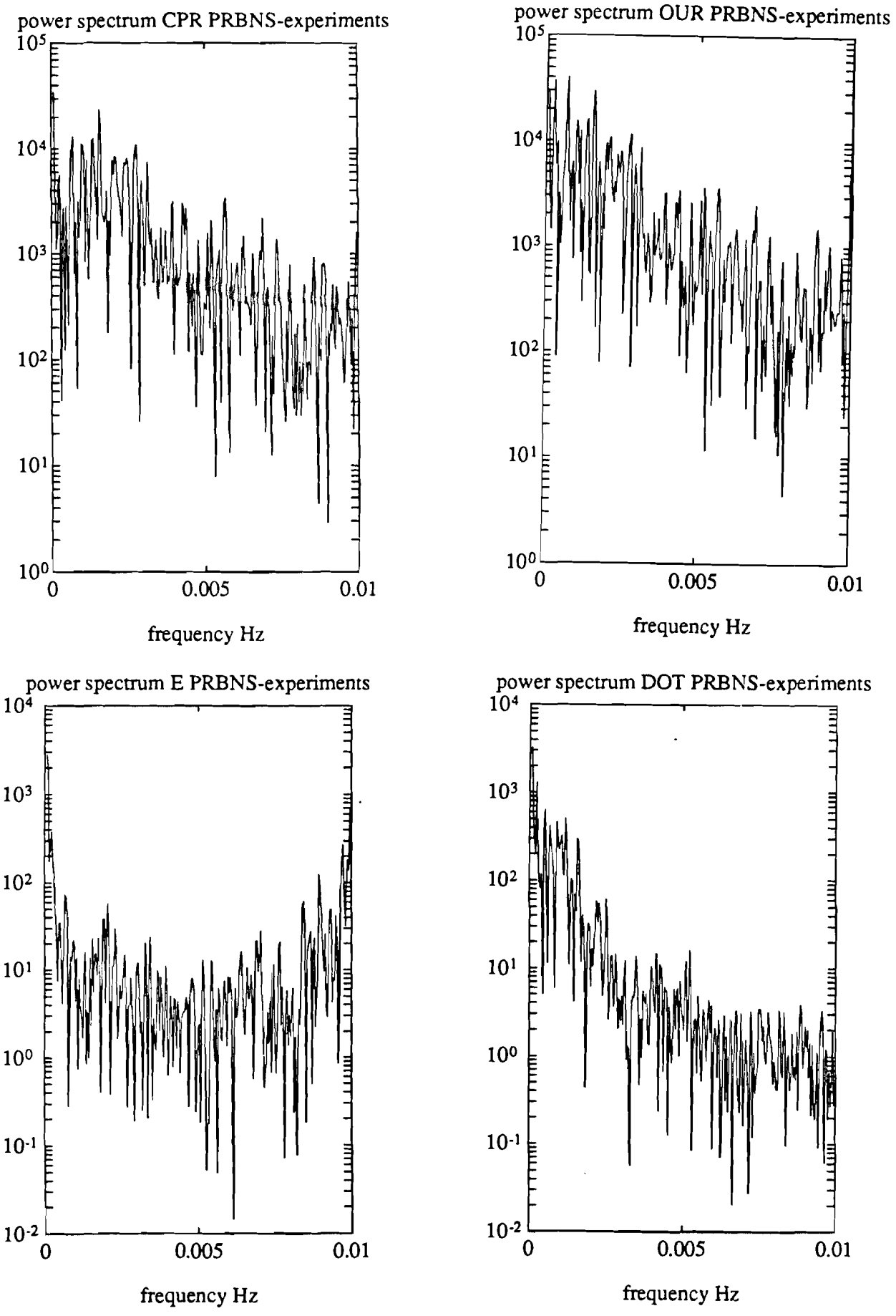

Figure C.8. Power spectra of the outputs of the final PRBNS-experiments.

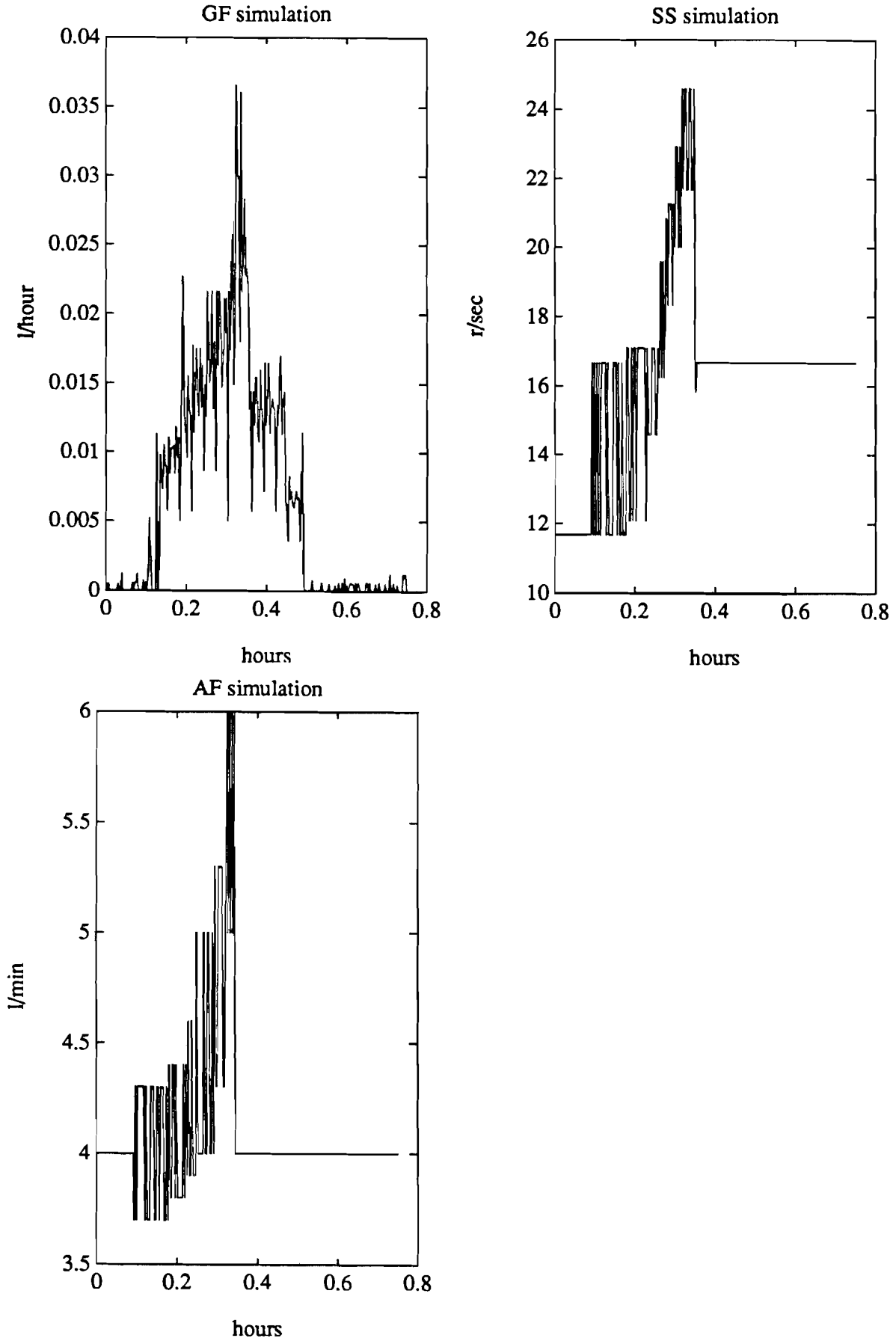

Figure C.9. A) Inputs of the dataset, used for model to model estimation of the parameters, without the production of ethanol

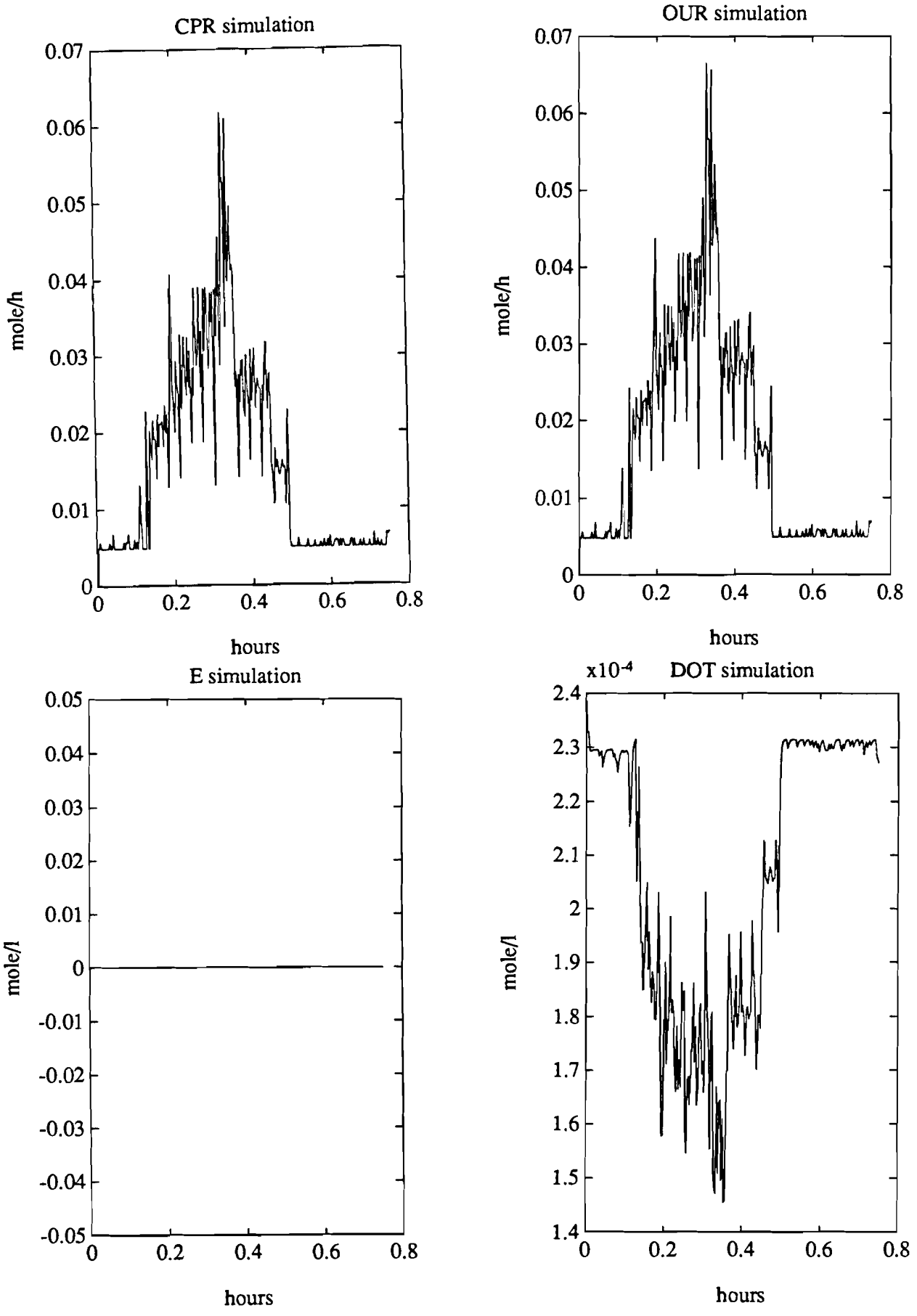

Figure C.9. B) Outputs of the dataset, used for model to model estimation of the parameters, without the production of ethanol

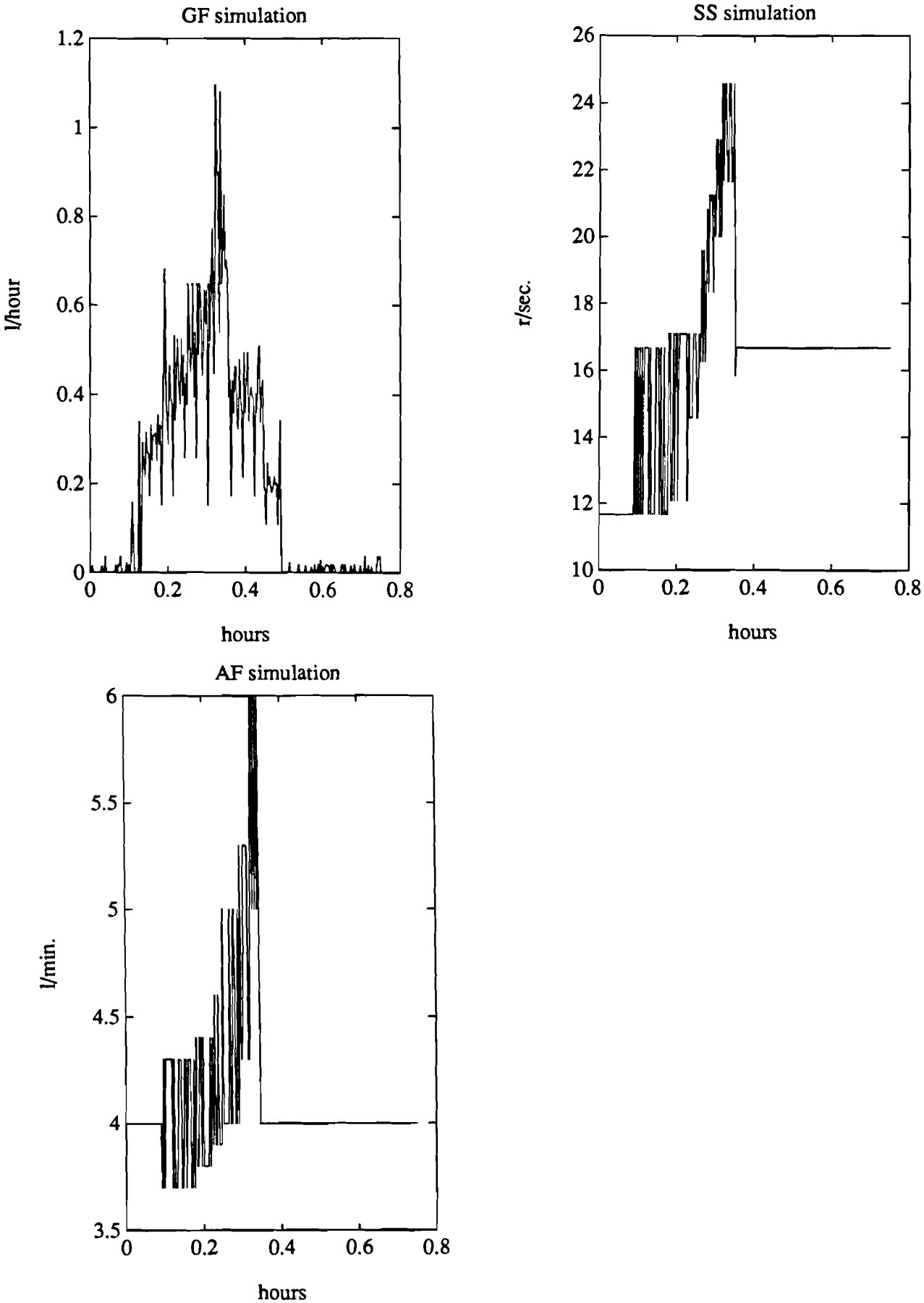

Figure C.10. A) Inputs of the dataset, used for model to model estimation of the parameters, with ethanol production.

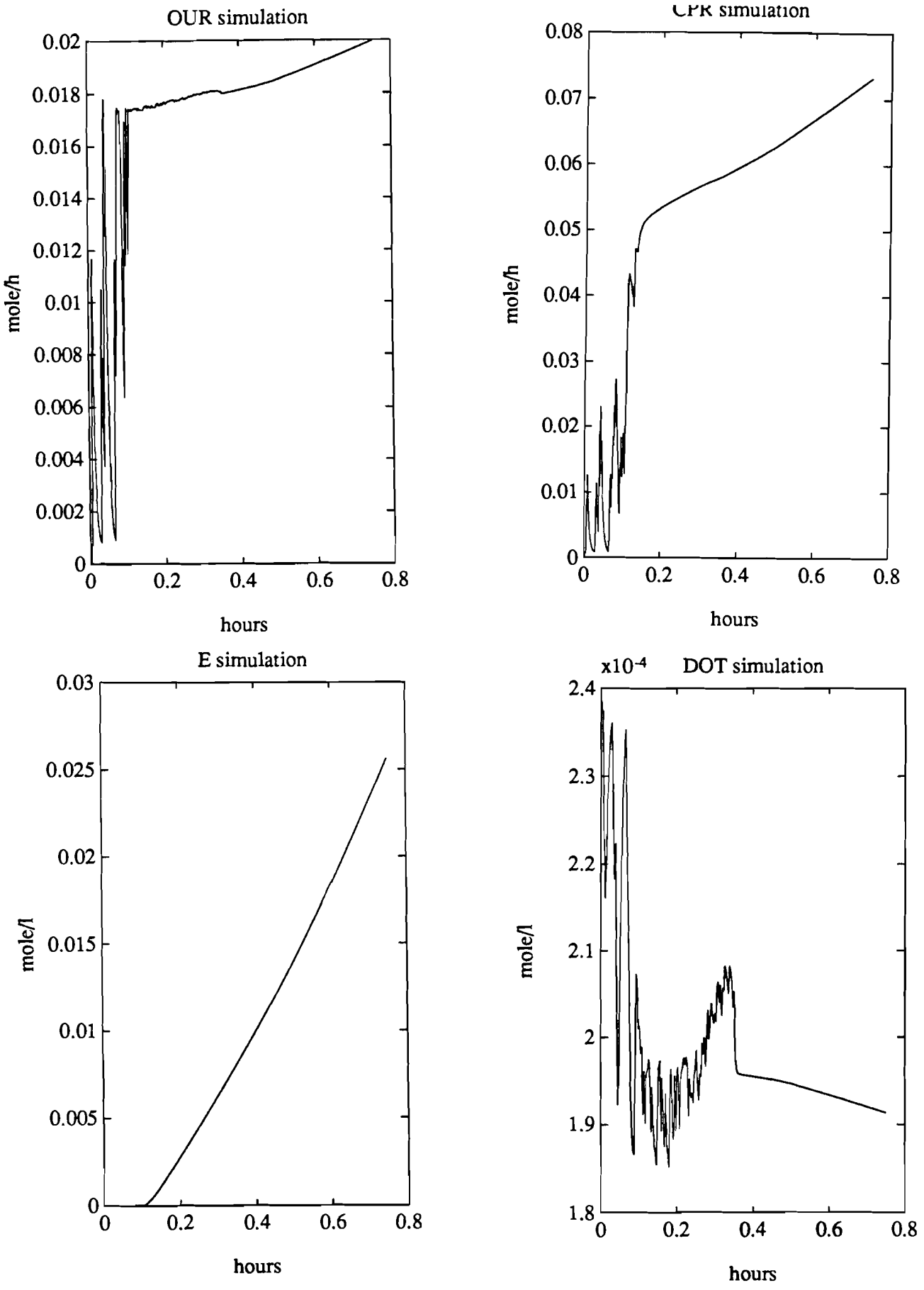

Figure C.10. B) Outputs of the dataset, used for model to model estimation of the parameters, with ethanol production.

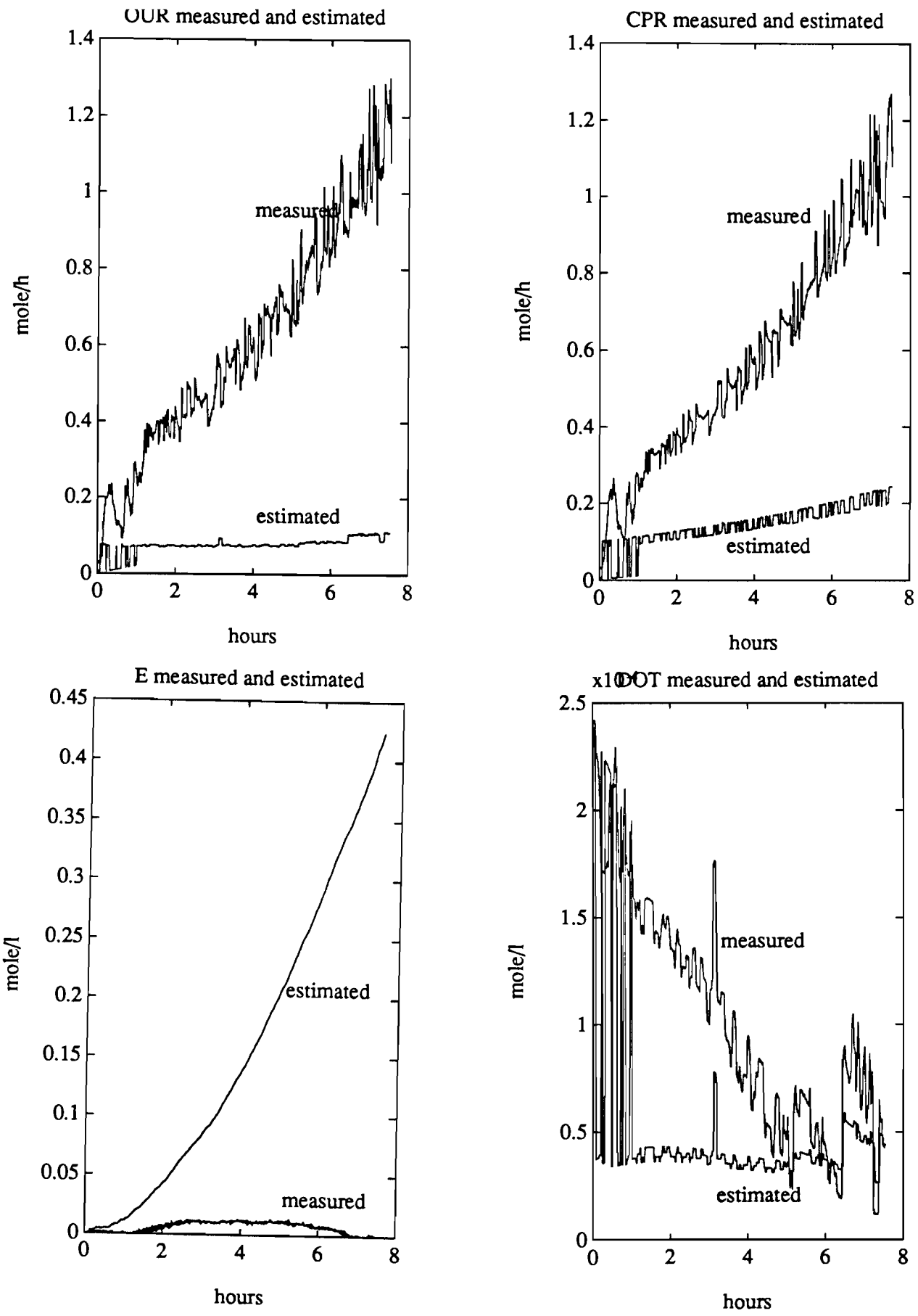

Figure C.11. Results of the parameter estimation of the real set

Appendix D. Report of the results of the minimizing routine MINIQUAD.

The report of an estimation is given in this appendix as an example of the output.

```
*************************************************************************
***************** M I N I Q U A D:  INPUT  REPORT *********************
*************************************************************************


METHOD =                          MARQUARDT

DERIVATES =                       NUMERICAL

LINEMIN METHOD =                  LINEAR SEARCH

NUMBER OF VARIABLES =               16

NUMBER OF RESIDUALS =              329

MAXIMUM NUMBER OF ITERATIONS =     999

OUTPUT PARAMETER =                   2

EPS FOR ABSOLUTE ACCURACY =       .100E-04

EPS FOR RELATIVE ACCURACY =       .100E-04

STOPCR =  12 FROM  1=NORMDELTAX  < EPSABS+EPSREL*NORM(XK)
                   2=DELTAFX     < EPS+EPSREL*ABSFX(K)
                   3=NORMGRAD    < EPSABS+EPSREL*NORMGRAD(X0)

*** VARIABLES ***
 X     STARTING
 1   0.120000000E+01
 2   0.120000000E+01
 3   0.120000000E+01
 4   0.120000000E+01
 5   0.120000000E+01
 6   0.120000000E+01
 7   0.120000000E+01
 8   0.120000000E+01
 9   0.120000000E+01
10   0.120000000E+01
11   0.120000000E+01
12   0.120000000E+01
13   0.120000000E+01
14   0.120000000E+01
15   0.120000000E+01
16   0.120000000E+01

FX = 0.132166733E+12
```

```
*******************************************************************************
********************* M I N I Q U A D:  EXTENDED **********************
*******************************************************************************

*******************************************************************************

ITERATION NR    =  0                NORMDELTAX = 0.000000000E+00
FUNCTION VALUE  = 0.132166733E+12   DELTAFX    = 0.000000000E+00
NR OF FIE EVAL. = 17                NORMGRAD   = 0.128602713E+13

 I   X[I]              GRADIENT[I]
 1  0.120000000E+01  -0.318253142E+04
 2  0.120000000E+01  -0.872221755E+00
 3  0.120000000E+01   0.391822953E+01
 4  0.120000000E+01  -0.755160393E+00
 5  0.120000000E+01   0.000000000E+00
 6  0.120000000E+01  -0.436009347E+03
 7  0.120000000E+01  -0.604903189E+04
 8  0.120000000E+01  -0.743087735E+03
 9  0.120000000E+01   0.237319149E+05
10  0.120000000E+01   0.387101781E+01
11  0.120000000E+01   0.191132081E+04
12  0.120000000E+01  -0.390593200E+05
13  0.120000000E+01  -0.369172005E+01
14  0.120000000E+01   0.220279270E+12
15  0.120000000E+01   0.123979137E+13
16  0.120000000E+01   0.261266502E+12

OUTPUT MARQUARDT-SEARCH
STEP    LAMBDA FUNCTION VALUE   GAMMA
 0  0.10E-01  0.132166733E+12    0.00
 1  0.10E-01  0.390874004E+11   90.00

*******************************************************************************

ITERATION NR    =  1                NORMDELTAX = 0.276646460E+11
FUNCTION VALUE  = 0.390874004E+11   DELTAFX    = 0.930793328E+11
NR OF FIE EVAL. = 34                NORMGRAD   = 0.349749829E+13

 I   X[I]              GRADIENT[I]
 1  0.612746330E+07  -0.569220563E+06
 2  0.270421515E+11   0.000000000E+00
 3  0.177625775E+09   0.000000000E+00
 4 -0.564559268E+10   0.000000000E+00
 5  0.120000000E+01   0.000000000E+00
 6  0.883020085E+05  -0.253743124E+08
 7  0.845076586E+07   0.563206894E+06
 8 -0.210855379E+07  -0.187725678E+07
 9 -0.602723658E+06   0.000000000E+00
10  0.987437520E+09   0.000000000E+00
11 -0.134215694E+06  -0.458783378E+08
12  0.337163909E+06   0.000000000E+00
13 -0.108409684E+10   0.000000000E+00
14  0.962753537E+00   0.142636999E+12
15  0.115836721E+01  -0.189226517E+13
16  0.999781495E+00  -0.293793831E+13

OUTPUT MARQUARDT-SEARCH
STEP    LAMBDA FUNCTION VALUE   GAMMA
 0  0.10E-02  0.390874004E+11    0.00
 1  0.10E-02  0.390253418E+11   90.00

*******************************************************************************
```

```
ITERATION NR    =  2                NORMDELTAX = 0.133906197E+04
FUNCTION VALUE  = 0.390253418E+11   DELTAFX    = 0.620586445E+08
NR OF FIE EVAL. = 51                NORMGRAD   = 0.118252029E+13

 I   X[I]              GRADIENT[I]
 1  0.612742869E+07  -0.251606491E+06
 2  0.270421515E+11   0.000000000E+00
 3  0.177625775E+09   0.000000000E+00
 4 -0.564559268E+10   0.000000000E+00
 5  0.120000000E+01   0.000000000E+00
 6  0.882943563E+05   0.216485448E+08
 7  0.844943202E+07   0.123539978E+06
 8 -0.210844581E+07   0.116165260E+07
 9 -0.602723658E+06   0.000000000E+00
10  0.987437520E+09   0.000000000E+00
11 -0.134183669E+06   0.352301659E+08
12  0.337163909E+06   0.000000000E+00
13 -0.108409684E+10   0.000000000E+00
14  0.962591470E+00   0.408902404E+12
15  0.115832677E+01   0.107834088E+13
16  0.100002205E+01  -0.261407710E+12

OUTPUT MARQUARDT-SEARCH
STEP    LAMBDA FUNCTION VALUE   GAMMA
 0  0.10E-03  0.390253418E+11    0.00
 1  0.10E-03  0.389452461E+11   90.00

*************************************************************************

ITERATION NR    =  3                NORMDELTAX = 0.917074116E+03
FUNCTION VALUE  = 0.389452461E+11   DELTAFX    = 0.800956303E+08
NR OF FIE EVAL. = 68                NORMGRAD   = 0.260590088E+13

 I   X[I]              GRADIENT[I]
 1  0.612832715E+07  -0.998480876E+06
 2  0.270421515E+11   0.000000000E+00
 3  0.177625775E+09   0.000000000E+00
 4 -0.564559268E+10   0.000000000E+00
 5  0.120000000E+01   0.000000000E+00
 6  0.883102568E+05  -0.655402055E+08
 7  0.844951913E+07  -0.263617565E+06
 8 -0.210860318E+07  -0.202113972E+07
 9 -0.602723658E+06   0.000000000E+00
10  0.987437520E+09   0.000000000E+00
11 -0.134218047E+06  -0.289201369E+08
12  0.337163909E+06   0.000000000E+00
13 -0.108409684E+10   0.000000000E+00
14  0.962551705E+00  -0.698495783E+12
15  0.115796137E+01  -0.155218527E+12
16  0.100016749E+01  -0.250573946E+13

OUTPUT MARQUARDT-SEARCH
STEP    LAMBDA FUNCTION VALUE   GAMMA
 0  0.10E-04  0.389452461E+11    0.00
 1  0.10E-04  0.388307826E+11   90.00

*************************************************************************

ITERATION NR    =  4                NORMDELTAX = 0.612005309E+03
FUNCTION VALUE  = 0.388307826E+11   DELTAFX    = 0.114463530E+09
NR OF FIE EVAL. = 85                NORMGRAD   = 0.299371642E+12

 I   X[I]              GRADIENT[I]
 1  0.612887133E+07  -0.586268251E+06
 2  0.270421515E+11   0.000000000E+00
```

```
 3  0.177625775E+09   0.000000000E+00
 4 -0.564559268E+10   0.000000000E+00
 5  0.120000000E+01   0.000000000E+00
 6  0.883217970E+05  -0.253624010E+08
 7  0.844926129E+07  -0.712060911E+05
 8 -0.210849455E+07  -0.681335632E+05
 9 -0.602723658E+06   0.000000000E+00
10  0.987437520E+09   0.000000000E+00
11 -0.134216064E+06  -0.172368737E+08
12  0.337163909E+06   0.000000000E+00
13 -0.108409684E+10   0.000000000E+00
14  0.962441789E+00   0.144850355E+11
15  0.115774233E+01  -0.487730563E+10
16  0.100024975E+01  -0.298981228E+12

OUTPUT MARQUARDT-SEARCH
STEP    LAMBDA FUNCTION VALUE   GAMMA
 0  0.10E-05  0.388307826E+11    0.00
 1  0.10E-05  0.387860184E+11   90.00

*************************************************************************

ITERATION NR    =  5                NORMDELTAX = 0.112646042E+04
FUNCTION VALUE  = 0.387860184E+11   DELTAFX    = 0.447642379E+08
NR OF FIE EVAL. =102                NORMGRAD   = 0.218962159E+13

 I   X[I]              GRADIENT[I]
 1  0.612986398E+07  -0.653344161E+06
 2  0.270421515E+11   0.000000000E+00
 3  0.177625775E+09   0.000000000E+00
 4 -0.564559268E+10   0.000000000E+00
 5  0.120000000E+01   0.000000000E+00
 6  0.883240740E+05  -0.234170575E+08
 7  0.844873340E+07  -0.386765700E+06
 8 -0.210856357E+07   0.118237503E+07
 9 -0.602723658E+06   0.000000000E+00
10  0.987437520E+09   0.000000000E+00
11 -0.134205730E+06  -0.271824791E+08
12  0.337163909E+06   0.000000000E+00
13 -0.108409684E+10   0.000000000E+00
14  0.962341788E+00  -0.546124699E+12
15  0.115766145E+01  -0.413325257E+12
16  0.100021179E+01  -0.207974824E+13

OUTPUT MARQUARDT-SEARCH
STEP    LAMBDA FUNCTION VALUE   GAMMA
 0  0.10E-06  0.387860184E+11    0.00
 1  0.10E-06  0.386833975E+11   90.00

*************************************************************************

ITERATION NR    =  6                NORMDELTAX = 0.981113153E+03
FUNCTION VALUE  = 0.386833975E+11   DELTAFX    = 0.102620900E+09
NR OF FIE EVAL. =119                NORMGRAD   = 0.439934692E+12

 I   X[I]              GRADIENT[I]
 1  0.613064922E+07  -0.267930756E+06
 2  0.270421515E+11   0.000000000E+00
 3  0.177625775E+09   0.000000000E+00
 4 -0.564559268E+10   0.000000000E+00
 5  0.120000000E+01   0.000000000E+00
 6  0.883135416E+05   0.250498973E+07
 7  0.844917281E+07   0.172798678E+06
 8 -0.210895273E+07  -0.868962622E+06
 9 -0.602723658E+06   0.000000000E+00
10  0.987437520E+09   0.000000000E+00
```

```
11 -0.134169059E+06   0.216298984E+07
12  0.337163909E+06   0.000000000E+00
13 -0.108409684E+10   0.000000000E+00
14  0.962140573E+00  -0.225789948E+12
15  0.115740590E+01  -0.120605809E+12
16  0.100043810E+01  -0.357792777E+12

OUTPUT MARQUARDT-SEARCH
STEP    LAMBDA FUNCTION VALUE   GAMMA
 0  0.10E-07  0.386833975E+11    0.00
 1  0.10E-07  0.386401797E+11   90.00

*************************************************************************

ITERATION NR    =  7                NORMDELTAX = 0.106308056E+04
FUNCTION VALUE  = 0.386401797E+11   DELTAFX    = 0.432177228E+08
NR OF FIE EVAL. =136                NORMGRAD   = 0.221596466E+13

 I   X[I]              GRADIENT[I]
 1  0.613110408E+07  -0.155428151E+06
 2  0.270421515E+11   0.000000000E+00
 3  0.177625775E+09   0.000000000E+00
 4 -0.564559268E+10   0.000000000E+00
 5  0.120000000E+01   0.000000000E+00
 6  0.883120408E+05   0.286109740E+08
 7  0.844824265E+07   0.423579979E+06
 8 -0.210871184E+07   0.785373213E+06
 9 -0.602723658E+06   0.000000000E+00
10  0.987437520E+09   0.000000000E+00
11 -0.134171826E+06   0.334262307E+08
12  0.337163909E+06   0.000000000E+00
13 -0.108409684E+10   0.000000000E+00
14  0.962096421E+00  -0.132947188E+12
15  0.115736916E+01   0.150950536E+13
16  0.100045880E+01   0.161685435E+13

OUTPUT MARQUARDT-SEARCH
STEP    LAMBDA FUNCTION VALUE   GAMMA
 0  0.10E-08  0.386401797E+11    0.00
 1  0.10E-08  0.385559953E+11   90.00

*************************************************************************

ITERATION NR    =  8                NORMDELTAX = 0.122827569E+04
FUNCTION VALUE  = 0.385559953E+11   DELTAFX    = 0.841844086E+08
NR OF FIE EVAL. =153                NORMGRAD   = 0.108074355E+13

 I   X[I]              GRADIENT[I]
 1  0.613208450E+07   0.141269694E+06
 2  0.270421515E+11   0.000000000E+00
 3  0.177625775E+09   0.000000000E+00
 4 -0.564559268E+10   0.000000000E+00
 5  0.120000000E+01   0.000000000E+00
 6  0.883179748E+05   0.691243299E+07
 7  0.844750356E+07  -0.457622215E+06
 8 -0.210868807E+07   0.161656750E+07
 9 -0.602723658E+06   0.000000000E+00
10  0.987437520E+09   0.000000000E+00
11 -0.134196126E+06  -0.244662549E+07
12  0.337163909E+06   0.000000000E+00
13 -0.108409684E+10   0.000000000E+00
14  0.962235761E+00  -0.105717991E+13
15  0.115710019E+01  -0.170328310E+12
16  0.100032646E+01  -0.146169520E+12

OUTPUT MARQUARDT-SEARCH



*************************************************************************
********************* M I N I Q U A D:  ENDREPORT  **********************
*************************************************************************

THE PROCES IS TERMINATED BECAUSE STOPCRIT = 915
```

Appendix E. Fortran routine ESTIM.FOR that is used to simulate the model.

```
      SUBROUTINE ESTIM(T,Z,F)
C
C     ORDINARY DIFFERENTIAL EQUATION FOR D02BAF
C
      INTEGER           I, NN, MM, PP, NNR, MDEC
      PARAMETER         (MDEC=6001)
      COMMON            /NUMBER/NN, MM, PP, NNR
      DOUBLE PRECISION  Z(8), F(8), T, LTHETA(16)
     *                  ,A,B,C,G,H,J,K,L,M,KS,KO,KE,KI,KM,KN
     *                  ,MS,TAUS,TAUO,TAUE,QSM,QOM,QEM,GF,PF,C1,C2,C3
     *                  ,QSMAX,QOMAX,QEMAX,Z1,Z2,Z3,Z4,Z5,Z6,QO,QS,QE
     *                  ,KLA,OTC,RO1,RO2,RO3,U0,U1,U2,OMAX,RO,RCO
      COMMON            /EST/LTHETA, U0,U1,U2, RO,RCO
C
C     THE PARAMETERS THAT ARE NOT ESTIMATED.
C
      PARAMETER         (A=2.1675,B=3.65,C=2.35,G=0.36, H=1.8920,J=1.8740
     *                  ,K=1.6140,L=1.32,M=0.68,GF=2.933,PF=0.0,OMAX=23.84E-5)
C
C     THE PARAMETERS THAT ARE TO BE ESTIMATED.
C     IF NN=12 THEN IS PATHWAY 1 USED ELSE ARE THE OTHER PATHWAYS USED.
C     ALL PARAMETER ARE NORMED (AT 1)
C
C     PATHWAY 1 WITH OUT MS
C
      IF (NN .EQ. 3) THEN
         KS  =LTHETA(1) *1E-4
         KO  =LTHETA(2) *3E-7
         KE  =          2.2E-3
         KI  =          2E-4
         KM  =LTHETA(3) *1.7E-4
         KN  =          6.43E-5
         MS  =          0.002
         TAUS=          2.5
         TAUO=          1.6
         TAUE=          2.8
         QSM =          0.5
         QOM =          0.2
         QEM =          0.13
         C1  =          123.
         C2  =          0.7
         C3  =          0.25
      ELSE IF (NN .EQ. 11) THEN
         KS  =LTHETA(1) *1E-4
         KO  =LTHETA(2) *3E-7
         KE  =          2.2E-3
         KI  =          2E-4
         KM  =LTHETA(3) *1.7E-4
         KN  =LTHETA(4) *6.43E-5
         MS  =          0.002
         TAUS=LTHETA(5) *2.5
         TAUO=LTHETA(6) *1.6
         TAUE=          2.8
         QSM =LTHETA(7) *0.5
         QOM =LTHETA(8) *0.2
         QEM =          0.13
         C1  =LTHETA(9) *123.
         C2  =LTHETA(10)*0.7
         C3  =LTHETA(11)*0.25

C
C     PATHWAY 1
C
      ELSE IF (NN .EQ. 12) THEN
         KS  =LTHETA(1) *1E-4
```

```
         KO  =LTHETA(2) *3E-7
         KE  =          2.2E-3
         KI  =          2E-4
         KM  =LTHETA(3) *1.7E-4
         KN  =LTHETA(4) *6.43E-5
         MS  =LTHETA(5) *0.002
         TAUS=LTHETA(6) *2.5
         TAUO=LTHETA(7) *1.6
         TAUE=          2.8
         QSM =LTHETA(8) *0.5
         QOM =LTHETA(9) *0.2
         QEM =          0.13
         C1  =LTHETA(10)*123.
         C2  =LTHETA(11)*0.7
         C3  =LTHETA(12)*0.25
C
C     PATHWAY 1, 2, 3
C
      ELSE IF (NN .EQ. 16) THEN
         KS  =LTHETA(1) *1E-4
         KO  =LTHETA(2) *3E-7
         KE  =LTHETA(3) *2.2E-3
         KI  =LTHETA(4) *2E-4
         KM  =LTHETA(5) *1.7E-4
         KN  =LTHETA(6) *6.43E-5
         MS  =LTHETA(7) *0.002
         TAUS=LTHETA(8) *2.5
         TAUO=LTHETA(9) *1.6
         TAUE=LTHETA(10)*2.8
         QSM =LTHETA(11)*0.5
         QOM =LTHETA(12)*0.2
         QEM =LTHETA(13)*0.13
         C1  =LTHETA(14)*123.
         C2  =LTHETA(15)*0.7
         C3  =LTHETA(16)*0.25
      ELSE
         PRINT *,'***********************************************'
         PRINT *,'*                                             *'
         PRINT *,'*          WRONG NUMBER OF THE                *'
         PRINT *,'*     ESTIMATED PARAMETERS, SHOULD BE         *'
         PRINT *,'*       N = 16 OR N = 12 OR N = 11            *'
         PRINT *,'*                                             *'
         PRINT *,'***********************************************'
         STOP
      END IF
C
C     LIMITATION OF OUTPUTS
C
      DO 20 I=1,8
         IF (Z(I) .LT. 1E-10) THEN
             Z(I)=0
         END IF
   20 CONTINUE
C
C     CALCULATE MONOD KINETICS
C
      QSMAX=Z(6)
      QOMAX=Z(7)
      QEMAX=Z(8)
      Z1=Z(1)/(KS+Z(1))
      Z2=Z(3)/(KE+Z(3))
      Z3=Z(4)/(KO+Z(4))
      Z4= KI /(KI+Z(1))
      Z5=Z(1)/(KN+Z(1))
      Z6=(2*Z(1)+Z(3))/(KM+2*Z(1)+Z(3))
C
```

```
C     CALCULATE PATHWAY
C
      QS=QSMAX*Z1
      QO=QOMAX*Z3

      IF (QS .GE. QO/A) THEN
         RO1=QO/A
         RO2=QS-RO1
         RO3=0
      ELSE
         RO1=QS
         RO2=0
         RO3=QEMAX*Z2*Z4
         IF (RO3 .GE. ((QO-A*QS)/K)) THEN
            RO3=(QO-A*QS)/K
         END IF
      END IF
C
C     CALCULATE KLA AND OTC
C
      KLA=C1*(U1**C2)*(U2**C3)
      OTC=(OMAX-Z(4))
C
C     THE DIF. EQ. ARE:
C
      F(1)=(-RO1-RO2)*Z(2)-(U0*(Z(1)-GF))/Z(5)
      F(2)=(B*RO1+G*RO2+L*RO3-MS/B)*Z(2)-(U0*Z(2))/Z(5)
      F(3)=(J*RO2-RO3)*Z(2)-(U0*Z(3))/Z(5)
      F(4)=(-A*RO1-K*RO3-A*MS)*Z(2)+KLA*OTC-(U0*Z(4)-PF)/Z(5)
      F(5)=U0
      F(6)=(QSM*Z5-Z(6))/TAUS
      F(7)=(QOM*Z3*Z6-Z(7))/TAUO
      F(8)=(QEM*Z2*Z3*Z4-Z(8))/TAUE
C
C     F(1)=dZ(1)/dT   :SUBSTRATE
C     F(2)=dZ(2)/dT   :BIOMASS
C     F(3)=dZ(3)/dT   :ETHANOL
C     F(4)=dZ(4)/dT   :DOT
C     F(5)=dZ(5)/dT   :VOLUME
C     F(6)=dZ(6)/dT   :QSMAX  -
C     F(7)=dZ(7)/dT   :QOMAX   |  ENZYM
C     F(8)=dZ(8)/dT   :QEMAX  -
C
C     RO2 AND RCO2 ARE CALCULATED FOR RESIDU, THIS IS NOT DONE FOR D02BAF
C
      RO =(A*RO1+K*RO3+B*MS)*Z(2)
      RCO=(C*RO1+H*RO2+M*RO3+B*MS)*Z(2)
C
      RETURN
      END
```

Appendix F. The results of the study of the sensors and actuators.

In this appendix the results of the talks with the operators and the study of the actuators and sensors are discussed. These results are necessary to design the experiments needed for the estimation of the process parameters.

F.1 Talks with the operators.

The operators give information on the normal operating points of the process. The setpoint of the glucose flow is calculated and the flow is controlled to track the normal operating points. It will increase exponentially depending on the growth factor that is given and the concentration of the biomass at the beginning of the experiment. The stirrer speed is normally between 700 and 900 RPM, the airflow is set between 3 and 4 liter per minute.

The time constants of most outputs are expected to be very large, near 20 minutes. Dissolving the oxygen is a fast process, therefore the DOT will react much faster. The time constant of DOT is supposed to be near 5 minutes. Therefore a sampling time of 6 minutes should be enough. Only the dissolved oxygen has to be sampled faster.

F.2 Study of the actuators and sensors.

The process has 3 actuators: the stirrer motor, the airflow pump/valve and the glucose flow/pump. The sensors are the dissolved oxygen probe, the mass-spectrometer and the off-line measurements of the biomass. These actuators and sensors are tested. The result of these tests is discussed in the next paragraphs.

All inputs are governed by an Applikon unit (see figure F.1). The unit contains SISO controllers for the inputs. The pH and the temperature are controlled by the Applikon too. The Applikon also measures and discretizes the continuous signal from the dissolved oxygen probe. A Vax system controls the Applikon. All outputs but the concentration of dissolved oxygen are measured by the mass-spectrometer. The in- and output data is stored in a database.

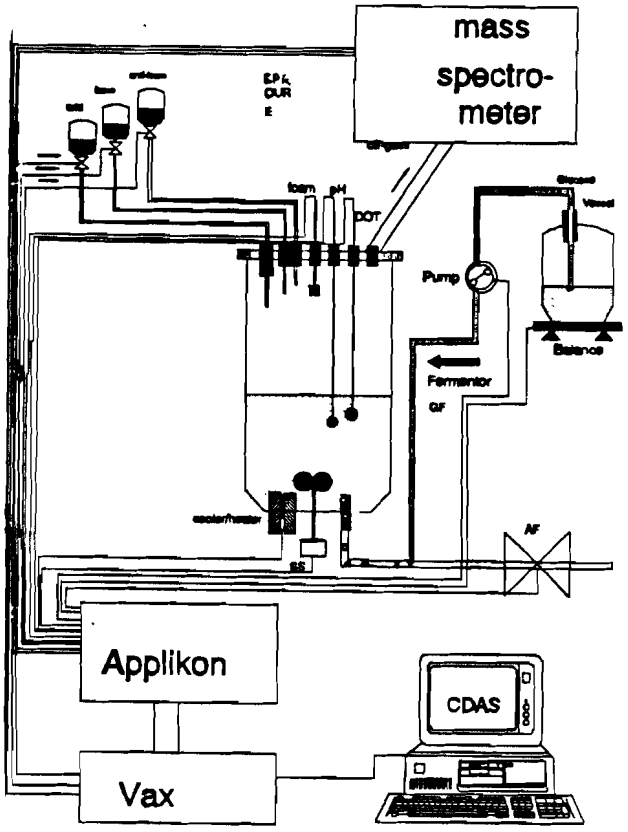

Figure F.1. The process with actuators and sensors.

The sample rate of the Applikon is 5 seconds. CDAS logs the outputs of the Applikon (setpoints and measured values) and the mass-spectrometer every minute or a multiple of a minute. If the mass-spectrometer has not calculated new values yet, CDAS copies the old values. If CDAS logs the data faster than the mass-spectrometer samples the data file contains a lot of duplicate samples. Note that the sample frequencies of CDAS and Applikon as well as the mass-spectrometer are independent and not synchronized. The moment, that the mass-spectrometer finishes the calculation of the new data, is not stored by CDAS. This means that samples are stored twice but then with a new time as if the sample is new.

F.2.1 Stirrer motor.

There is no control loop implemented in the Applikon to control the output of the stirrer motor. The linearity of the motor is tested. The error of the real RPM and the set point is quite large, about 2 till 10% (see figure F.2). This is tested without load, but the motor is assumed to be powerful enough to rotate with the same speed whatever the process conditions are. The setpoints of the stirrer speed are stored by CDAS. The input stirrer speed (SS) has to be corrected, using the values measured in the calibration, before estimation.

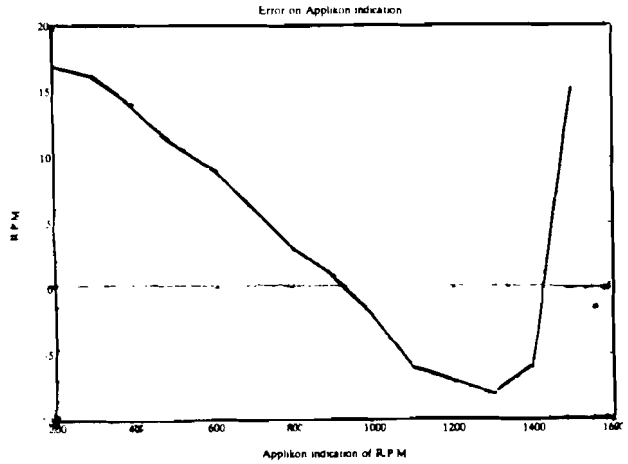

Figure F.2. Error in the stirrer speed between the setpoint and measured value.

The delay and the time constants are too small to measure. They are assumed to be less than 5 seconds which is the minimum sample rate.

F.2.2 Airflow pump/valve.

The air flow is controlled by the Applikon unit. The mean error is less then 1% of the setpoint. The standard deviation is smaller than 0.006 l/min. The delay time and time constants are smaller then 5 seconds and could not be measured.

The controller functions well, none of the actuator characteristics have to be included in the model.

F.2.3 Glucose flow/pump.

The glucose flow is not controlled by the Applikon unit but the setpoint of the pump is calculated by a control-program on the Vax. The exponential increase of the flow depends on the concentration of the biomass at the beginning of the experiment (X_o) and the growth-factor (μ).

$$GF^*(t) = K_1 \times X_o \exp(\mu \times t) \qquad (F.1)$$

where GF^* stands for the setpoint of the glucose flow and K_1 is given by $K_1 = \mu \setminus YSX$ where YSX stands for the estimated yealt (see batch report). t stands for the time passed during the experiment.

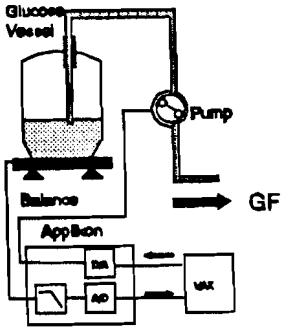

Figure F.3. The setup to measure the glucose flow.

The actual flow is measured indirectly. The glucose is stored in a vessel that stands on a balance. The glucose is pumped out of the vessel (see figure F.3). The decrease of the weight is differentiated in time to obtain the glucose flow. The analog output of the balance is filtered by a low pass filter with unknown characteristics and then sampled by the Applikon. The range is from 0 to 100 gram. The signal of the balance switches from 0 to 100 gram and vice versa if it is getting out of range (see figure F.4), so the signal is modulus 100. This signal as well as the calculated flow are stored by CDAS.

The flow is controlled by a recursive least squares controller. Slow changing setpoints are tracked good, but the controller needs 5 to 10 minutes to settle at a new setpoint if a step to the setpoint is applied. This is too slow for a PRBNS-signal and the controller can not be used. It is possible that the poor settling of the controller is due to the filtering of the balance signal.

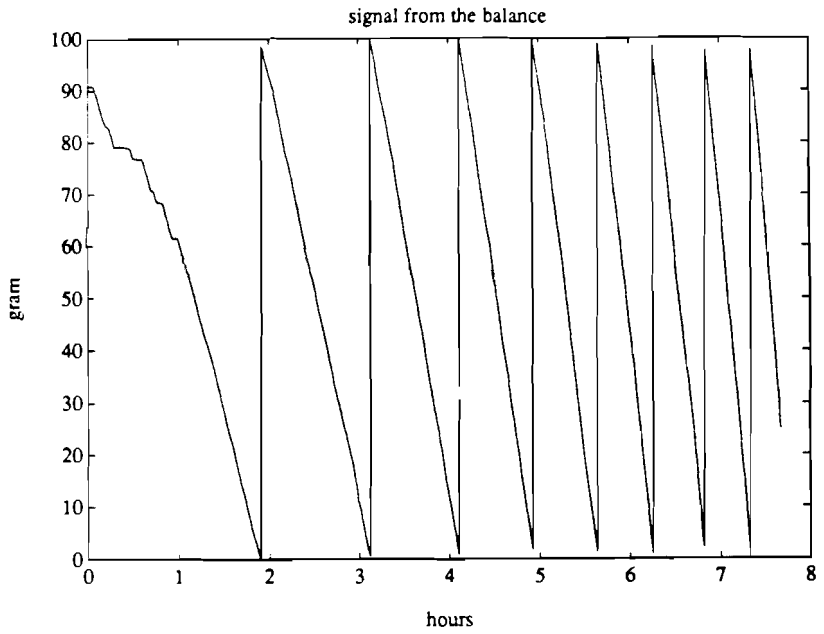

Figure F.4. The signal of the balance as it is measured by the Applikon.

The signal to noise ratio of the calculated flow is poor (figure F.5). This is caused by the measurement of the balance. The balance has a digital output. The Applikon unit, however, has analog inputs only. Therefore a digital/analog converter is mounted at the output of the balance. In the Applikon this signal is converted again to a digital value. Noise is introduced this way. If the sample frequency is high, the difference between 2 samples is small. This means the relative error is large. However if the sample frequency is low, the difference between 2 samples is large. The same absolute error will give a small relative error i.e. a better signal to noise ratio after differentiating the data of the balance.

To adjust the flow of the glucose pump one can also change the voltage of the pump. Of course the setpoint of the voltage (SP_{pump}) does not equal the glucose flow. If the pump is linear, the actual glucose flow (GF) is the setpoint multiplied by a constant. The constant can easily be calculated thus: the mean values of the setpoints and the measured glucose flow (GF_{meas}) are calculated. The constant is the quotient of the mean value of the glucose flow and the setpoints.

$$GF = \frac{mean(GF_{meas})}{mean(SP_{pump})} \times SP_{pump} \qquad (F.2)$$

As the pump is not linear visual inspection of the results is necessary to see if the calculated flow (GF) fits the measured flow (GF_{meas}). The advantage of this method is that it is very simple and wear of the tubes, changing pump settings or changing the pump has no influence on the results as this is calculated for each experiment. An example of this method is given in figure F.5.

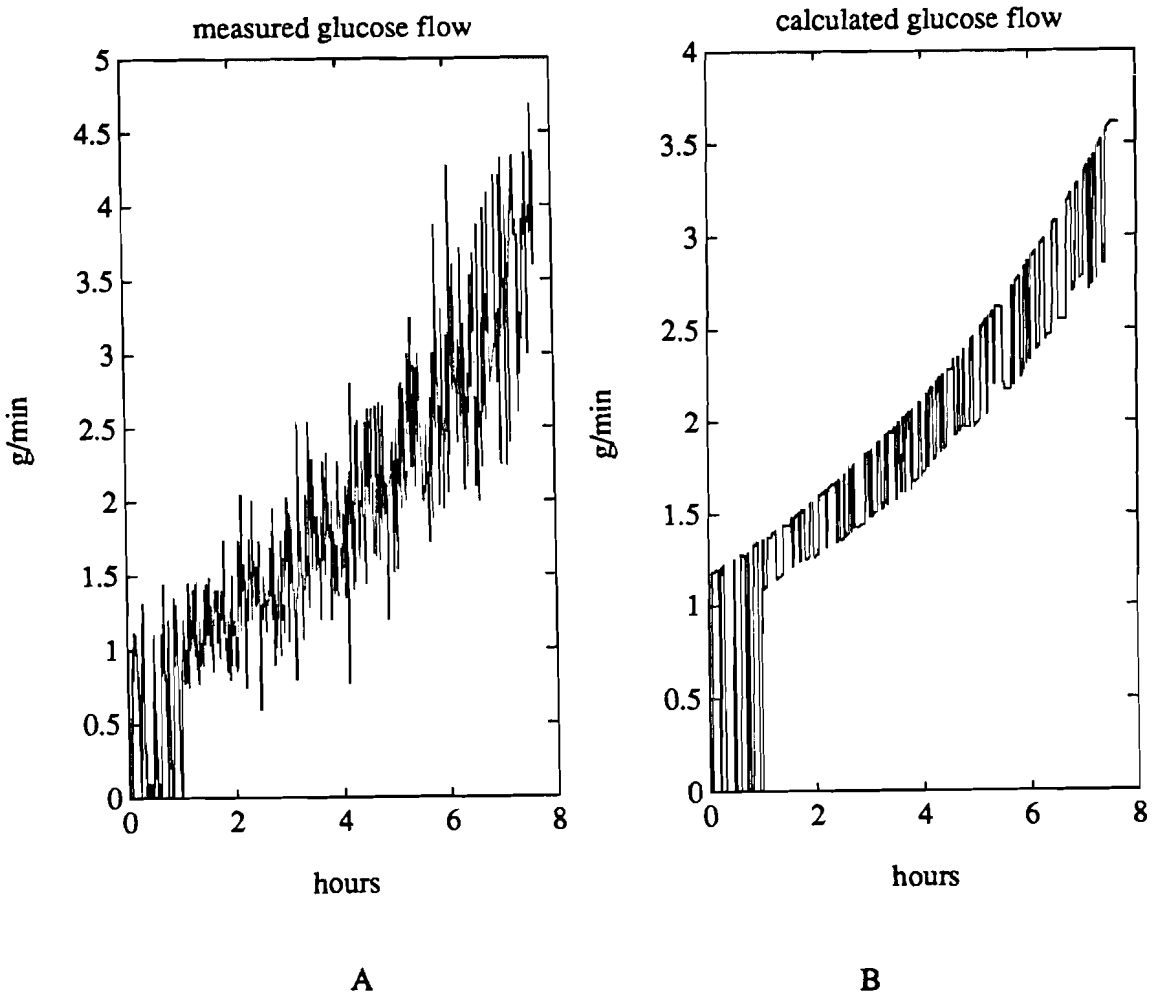

Figure F.5. The glucose flow, A) measured and B) calculated using the setpoints.

F.2.4 The DOT-probe.

The DOT-probe measures the dissolved oxygen tension of the broth. This is the concentration of the dissolved oxygen (P) as a percentage of the maximal concentration.

If a step is made from 100% to 0% the largest time constant of the DOT-probe can be measured, because the oxygen can be removed very fast using chemicals. The time constant is very small, within 5 seconds. The step from 0 to 100% took about 110 seconds settling time. This however is not the time needed for the probe to settle at the new level, but the time needed to dissolve the oxygen in the water. The noise of the probe is small. Its variance is 0.04% (figure F.6). With special adjustments the DOT can be sampled every 5 seconds.

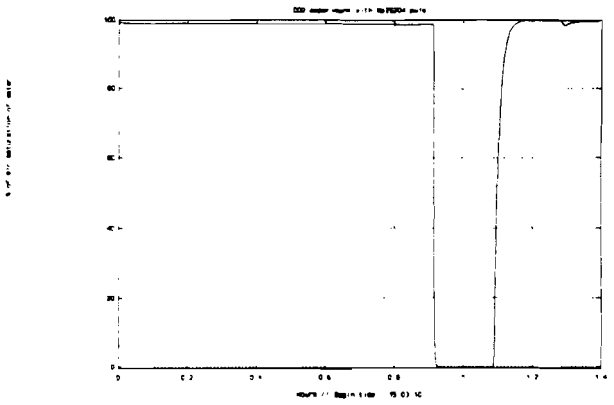

Figure F.6. Test of the DOT-probe.

F.2.5 The mass-spectrometer.

The mass-spectrometer measures O_2, CO_2 and ethanol concentration of the outlet gas. These values are used to calculate the ethanol concentration of the broth, the CPR and the OUR, using:

$$CPR(t)=AF(t)\times(CO_{2,off}(t) - CO_{2,in}(t))\times K_1$$
$$OUR(t)=AF(t)\times(O_{2,in}(t) - O_{2,off}(t))\times K_2 \qquad (F.3)$$

where $K_1 = 26.8517$

$$K_2 = 25.9699$$

These parameters are estimated from the data of the file E_27_7_900672. This is done using the dividing operator of MATLAB:

$$K_1 = CPR \;/\; ((CO_{2,off}(t) - CO_{2,in}) \times AF(t))$$

$$K_2 = OUR \;/\; ((O_{2,in} - O_{2,off}(t)) \times AF(t))$$

where $CO_{2,off}(t)$ stands for the concentration of the carbon dioxide in the off-let gas, $O_{2,off}(t)$ stands for the concentration of oxygen in the off-let gas, $CO_{2,in}$ stands for the CO_2 concentration in the air (=0.03%) and $O_{2,in}$ for the O_2 concentration in the air(=21%).

The mass-spectrometer needs at least an outlet airflow of 0.5 l/min. But normally it is about 5 l/min. The delay of the flow through the tube is about 8 seconds (5 l/min), the mass-spectrometer needs 20 to 30 seconds for calculation of the gas contents. Time constants nor noise spectra are known. The expected delay time is at least 28 to 38 seconds.

It is very hard to test the mass-spectrometer with calibrated gases, this is not done.

As the mass-spectrometer is used by more than one user, its sample time depends on the number of users. The minimum sampling time is near 30 seconds.

F.2.6 Off line measurements.

Off line it is possible to measure the concentration of the biomass by measuring the optical density of a sample of the broth. The noise of the method is very high and the off-set is not known. Therefore the measurement is not used to estimate the parameters of the process.

It is also possible to dry a sample of the broth and weigh the biomass. The accuracy of this method is high but the sample needed is too large to neglect. This makes it impossible to take more than a few samples. The method can not be used either.