CHAPTER 6

VALIDATION AND EVALUATION OF THE METHODOLOGY WITH REAL PROCESS DATA

6.1 INTRODUCTION

The proposed methodology was implemented on real closed-loop process data to try and identify a part of a reactor in the MIBK process. The process description of the MIBK plant can be found in Chapter 2.

In Section 6.2 the implementation and validation procedure is described in terms of each of the identification steps. This procedure is very similar to the procedure followed in the simulation, see Section 5.3. In Section 6.3 the expected validation results are discussed and the final results obtained are then given and also discussed. In Section 6.4 the implemented methodology is thoroughly evaluated and further recommendations, with regards to successful implementation of this methodology, are made.

Finally, it is concluded in Section 6.5 that, although it is possible to identify satisfactory models from closed-loop data for controller design, structured tests that ensure good SNR and PE reference signals should still be performed.

6.2 IMPLEMENTATION OF THE METHODOLOGY

The proposed methodology, as described in Section 4.8, was again implemented in MATLAB. This implementation is briefly discussed in terms of the five SID steps, as well as the methodology validation step. The similarities to Section 5.3 are mentioned and the differences are emphasized.

6.2.1 Experiment Design

Signals to be Measured: The measured CVs, $y_i(t)$, and MVs, $u_i(t)$, of a part of reactor A for October and November 1998 were obtained. These signals are described in Section 2.5.1 and the signal tags and labels are given in Table 6.1.

Table 6.1: Signal tags and labels.

Variable No.	Tag	Label
1	05TIC280A.MV	y_1
2	05TIC232A.MV	y_2
3	05TIC279A.MV	y_3
4	05FIC223A.MV	y_4
5	05TIC280A.SV	u_1
6	05TIC232A.SV	u_2
7	05TIC279A.SV	u_3
8	05FIC223A.SV	u_4

Sampling Time: The plant data were sampled every 30s, which is twice as fast as the controller execution time of 1min. Therefore, the plant input and output data sets were inter-sampled by a factor of two. In MATLAB the data sets were also resampled, making use of the *resample* function, to obtain a sampling time of 1min. The plant was identified from these data sets, as well as from the inter-sampled data sets. Thus, the influence of the inter-sampling could also be seen.

Excitation Signals: No structured closed-loop tests were allowed on the plant and the reference (set-point) values were also not logged and could thus not be retrieved. Therefore, it is not known wether these signals were PE.

6.2.2 Data Collection

Collection: The desired data sets were retrieved from a database on which the relevant data sets are stored.

Preprocessing: No high frequency disturbances or bursts and outliers in the data record were observed. There were, however, some missing data sets and at some stages the data records were non-continuous. A routine was written to search for the sections in the data records where the data sets were not sampled every 30s. These sections were not considered for estimation or validation.

Again, the trends in the data sets were removed, making use of the MATLAB function *dtrend*.

The continuous data sets were visually inspected to select the ranges where the MVs were the most excited. Models were estimated from different sections in the data. For estimation the range that gave the best results is 9 November 1998 13:00:00 - 10 November

1998 17:12:00. The models were validated with different data ranges. The validation results for the data range from 10 November 1998 17:13:30 - 11 November 1998 15:16:00 are shown in Section 6.3.

Time Delay: From visual inspection of the data, as well as from the known open-loop identified models, it was concluded that all the time delays were for practical purposes approximately zero. Just to make sure of these values, second-order models with different delays were compared by making use of the estimation data set and the *compare* function. The best fit selected the delays. From these tests, very small time delays (relative to the sampling time) of 30s were selected.

6.2.3 Model Structure Selection

Type of Structure: Again, the multivariable ARX type model structure was used.

Order Selection: The old step-response models obtained from the open-loop tests were examined. Since these step response models are nonparametric with 180 parameters each, they did not supply much information regarding a suitable model structure. The state-space models obtained from these step response models are of a very high order and therefore did not give a good indication of the appropriate model order. Therefore, the multivariable ARX models were fitted to the estimation data set for different model structure orders, starting at a low order. For each of these models, the sum of squared prediction errors were computed with the *compare* function, as they were applied to the estimation data set. The percentage of fit did not improve significantly from the second-order model onwards (up to an eight-order model was tested). Therefore, a second-order model was chosen.

Since only single-output data sets are handled by the *arxstruc* and *selstruc* routines, these functions could not be used in the structure selection.

6.2.4 Model Estimation

Again, similar to the simulation in Chapter 5, the *idarx* command, which uses the LSE PEM estimation method, was used to fit the chosen models to the estimation data. The model was also transformed from discrete-time to continuous-time.

6.2.5 Model Validation

Simulation and Prediction: As in the simulation, the pure simulated and 6-step ahead predicted outputs, as well as the percentage of fit for the identified models were computed with the *compare* command. The prediction horizon of the real controller was not known, therefore, an arbitrary horizon of 6 steps was assumed.

Residual Analysis: Again, the *resid* command was used to calculate and display the auto-correlation function of the residuals, as well as the cross-correlation between the residuals and the plant inputs.

Model Reduction: The *minreal* function can be used to cancel pole-zero pairs in a model transfer function and thus to reduce the model order if necessary. This function was used to try and cancel pole-zero pairs in the identified model. There were no pole-zero pairs to cancel and the model order is thus not too high.

6.2.6 Methodology Validation

Preparation of the Open-Loop Identified Models: The AspenTech DMCplus model file of the open-loop identified MIBK plant was imported into an EXCEL spreadsheet. The model is in a step response format, with each of the SISO transfer functions made up of 180 step response coefficients. These data sets were sorted and the step response coefficients of the relevant SISO transfer functions were imported into MATLAB. In MATLAB a function was written to construct state-space models from the step response coefficients, making use of singular value decomposition (SVD). This function is described in Addendum B. SISO models with orders between 19 and 59 were obtained. The state-space transfer functions could and were then reduced to fourth-order models. The SISO step responses of this model are shown in Fig. 6.1. This figure shows that the resulting open-loop identified model of the selected part of reactor A consists of ten non-zero SISO transfer functions and that the transfer functions from u_1 to y_2, y_3 and y_4, from u_2 to y_4, from u_3 to y_4 and from u_4 to y_2 are zero.

The closed-loop identified model was compared with both the full order and the reduced order models. Since the results are very similar, the validation results obtained for the fourth-order model are shown in this chapter.

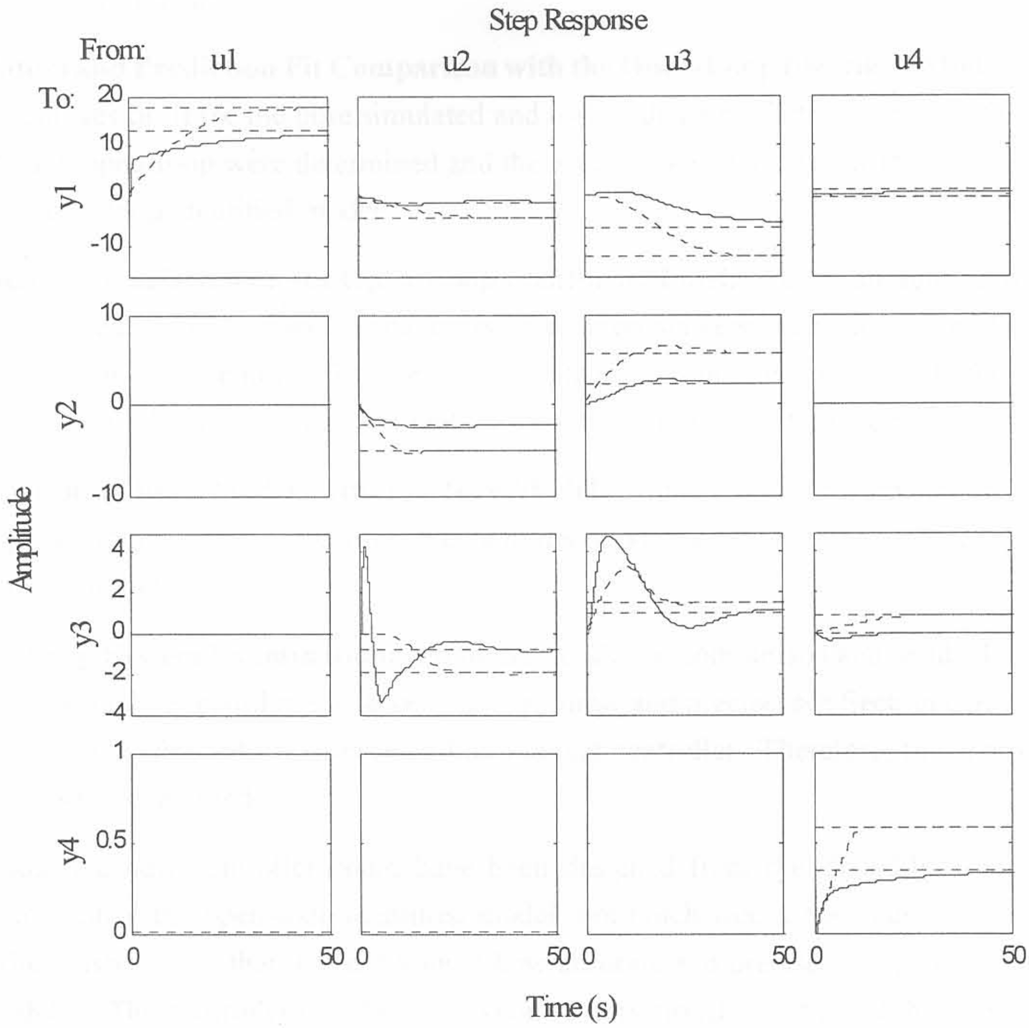

Figure 6.1: The SISO step responses of the open-loop identified (dotted lines) and closed-loop identified (solid lines) models. The horizontal dotted lines are the steady sate values of these responses.

Visual Time and Frequency Domain Comparison with the Open-Loop Identified Model: Again, the *step* command was used to plot the step responses and the *impulse* command was used to plot the impulse responses for each of the SISO transfer functions of the models. The *bode* command was also used to plot both the amplitude and phase for each of the SISO transfer functions of the models. The plots for the open-loop and closed-loop identified models are visually compared.

Simulation and Prediction Fit Comparison with the Open-Loop Identified Model: Again, the percentages of fit for the pure simulated and 6-step ahead predicted output of the model identified in open-loop were determined and these values are compared with those obtained for the closed-loop identified model.

Residual Comparison with the Open-Loop Identified Model: Again, the auto-correlation function of the residuals, as well as the cross-correlation between the residuals and the plant inputs for the model identified from open-loop data were computed and these functions are visually compared with the ones obtained for the closed-loop identified model.

Comparison of Both Models with the True Model: Since this is not a simulation and the "true" plant model is thus not known, the identified models could not be compared with the "true" plant model.

Closed-Loop System Examination: From the simulation comparison and residual analysis the open-loop model could not be accepted as accurate and precise, see Section 6.3.3. There was also no available information regarding the real controller. Therefore, the closed-loop system was not examined.

Although a new controller could have been designed from the closed-loop identified model to control the open-loop identified model, not much would have been gained from this. The reason being that it is not known how accurate and precise the open-loop identified model is. The controller may be able to control this model, but it is not the "true" plant. Also, it should be determined if the real controller, redesigned from the closed-loop identified model for the original design parameters, is able to control the plant. This cannot be determined, since the controller parameters are not known and the real controller can thus not be modelled.

6.3 Validation Results

In this section the expected validation results are discussed, the obtained results are shown and finally, the obtained results are discussed and compared with the expected results.

6.3.1 Expected Results

In Chapter 5 it is concluded that the proposed closed-loop SID methodology and open-loop SID method deliver comparable identification results only when structured tests, which ensure PE reference signals and good SNRs, are performed. Therefore, it was expected that an unsatisfactory model, which is not a good description of the plant, would be identified, because structured tests were not performed and the reference inputs were thus probably not PE.

It is also concluded in Chapter 5 that when the reference signals are not PE, or result in bad SNRs for the plant input signals, the methodology should be reconsidered, since, in general, methods that ensure identifiability, e.g. inter-sampling and nonlinear controllers, still do not guarantee satisfactory models with small variances. Therefore, even though the measured data sets were informative, since the data signals were inter-sampled, a large variance in the identified model was still expected.

It was also expected that even if satisfactory structured tests were performed, with the noise model of the ARX structure not an accurate description of the true noise model, the identified model would contain a bias. A possible bad fit in the low frequency regions was also expected, since the ARX model structure penalises the high-frequency misfit behaviour more than low-frequency misfit behaviour.

Even if a good model were identified from closed-loop, a possible difference between the models identified from open-loop and closed-loop data was still expected, since the frequency weighting for these two types of models are different and the plant may have exhibited closed-loop dynamics different form the open-loop.

6.3.2 Obtained Results

The validation results for the second-order model, estimated from the closed-loop data, measured from 9 November 1998 13:00:00 to 10 November 1998 17:12:00, are given in this section. This model is the most accurate identified model. Models were also estimated from other ranges of the measured data sets. Although there are similar characteristics in these models, the models vary considerably for the step responses, bode plots, percentage of fit, etcetera.

When only the estimation data are used in the validation all the results are good. However, the validation results for other ranges in the measured sets of data are less attractive, as

Table 6.2: Percentage of fit between measured and predicted outputs for chosen validation range.

Signal	# Steps ahead Prediction	Open-loop identified model	Closed-loop identified model (inter-sampled)
y_1	∞	55.27%	32.22%
	6	-	91.28%
y_2	∞	-4.306%	58.41%
	6	-	93.21%
y_3	∞	49.13%	34.09%
	6	-	10.55%
y_4	∞	-39.76%	-71.21%
	6	-	74.28%

shown. The validation results for the data range 10 November 1998 17:13:30 - 11 November 1998 15:16:00 are given in this section. The model was also validated with some other data ranges from October and November. For each of these sets different results were obtained.

For the resampled data, a model was also identified. This model, obtained from data that were not inter-sampled, caused MATLAB to warn that *the matrix is close to singular.* Although this model and the model obtained from the inter-sampled data have similar characteristics, it is worse in accuracy for the step responses, bode plots, percentage of fit, etcetera.

6.3.2.1 Simulation and Prediction Analysis and Comparison with the Open-Loop Identified Model

The percentage of fit between the measured and pure simulated output signals, as well as the 6-step ahead predicted output signals, for the open-loop and closed-loop identified models are compared in Table 6.2. Since the open-loop identified model is in state-space form, the 6-step ahead predicted outputs were not computed. In Figs. 6.2, 6.3 and 6.4 the pure simulated, predicted and measured outputs are also shown. These results show that the open-loop and closed-loop identified model have very low percentages of fit. With the 6-step ahead prediction, this fit improves significantly.

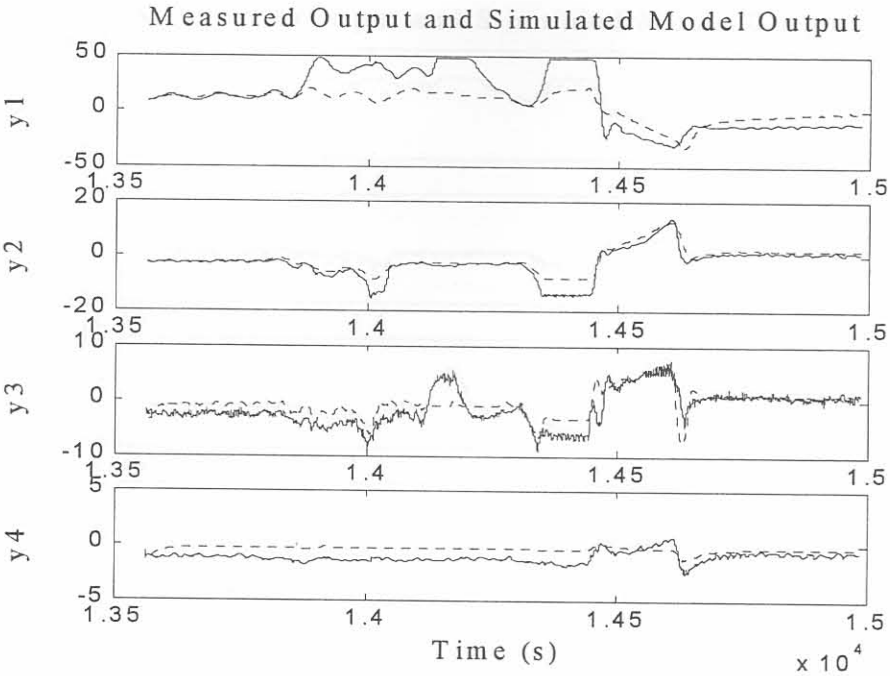

Figure 6.2: A comparison of the measured outputs (solid lines) and the simulated outputs of the closed-loop identified model (dotted lines).

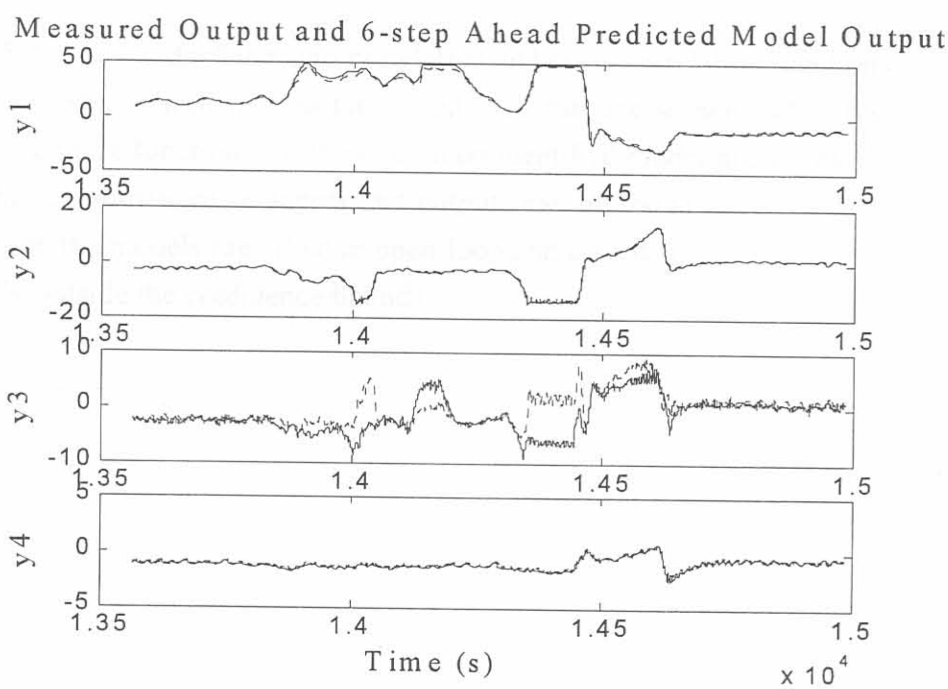

Figure 6.3: A comparison of the measured outputs (solid lines) and the 6-step ahead predicted outputs of the closed-loop identified model (dotted lines).

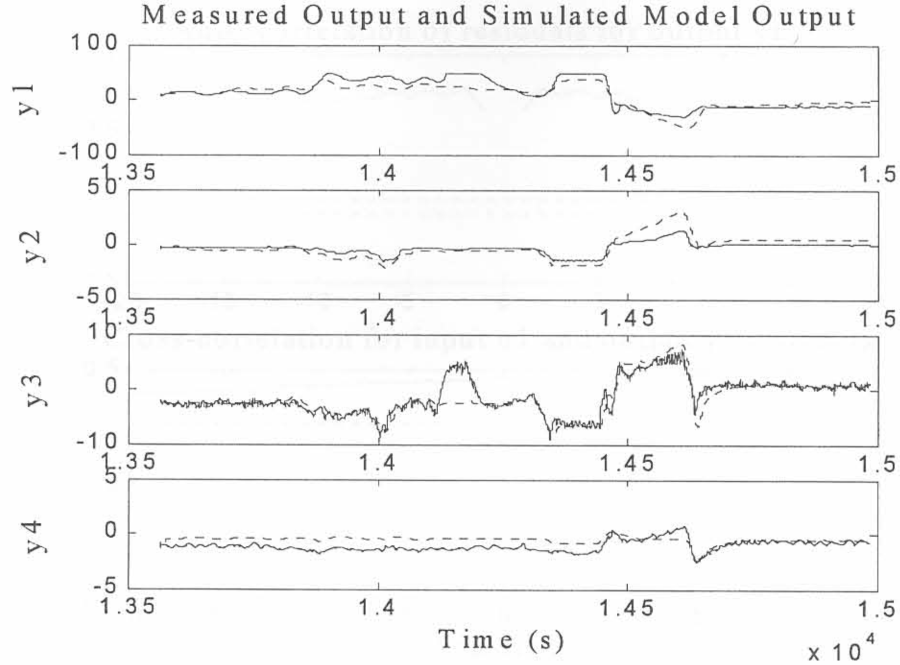

Figure 6.4: A comparison of the measured outputs (solid lines) and the simulated outputs of the open-loop identified model (dotted lines).

6.3.2.2 Residual Analysis and Comparison with the Open-Loop Identified Model

In Figs. 6.5, 6.6, 6.7 and 6.8 the auto-correlation and cross-correlation functions of the residuals for the model identified from the closed-loop data are shown and in Figs. 6.9, 6.10, 6.11 and 6.12 these functions for the open-loop identified model are shown. The functions for the other combinations of inputs and outputs can be found in Addendum C. In these plots, for both the models identified in open-loop and closed-loop, most of the functions go significantly outside the confidence bounds.

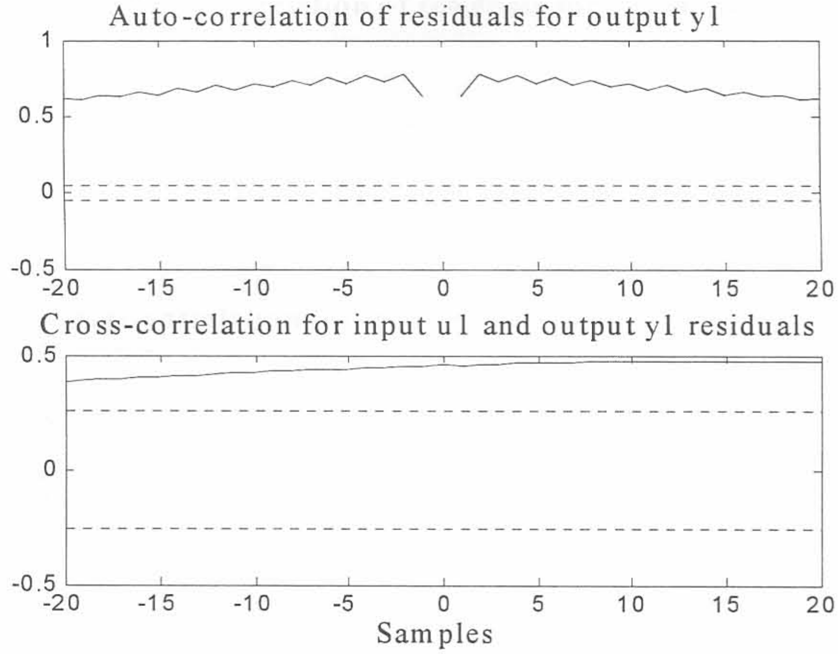

Figure 6.5: The cross-correlation and auto-correlation of the residuals for y_1 and u_1 of the closed-loop identified model.

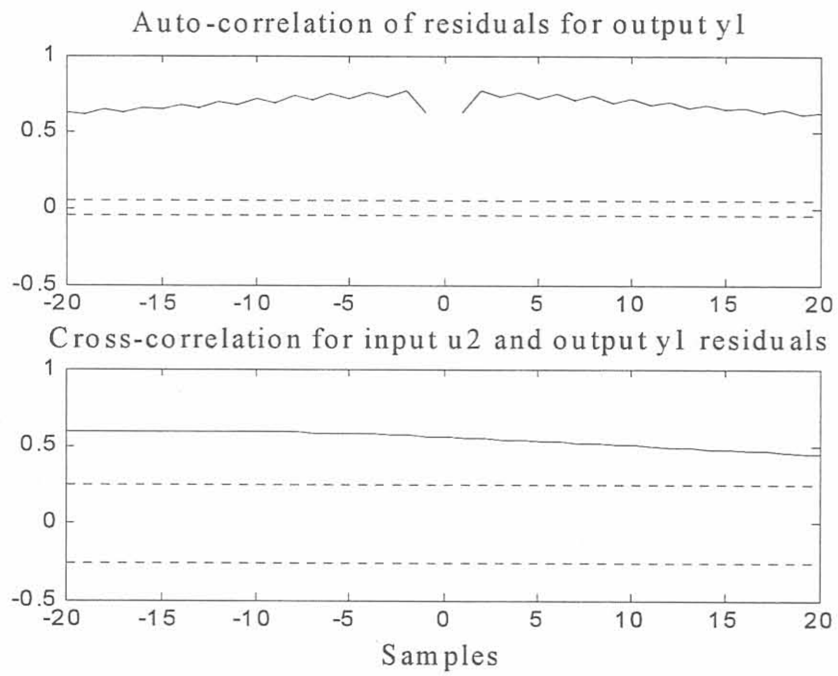

Figure 6.6: The cross-correlation and auto-correlation of the residuals for y_1 and u_2 of the closed-loop identified model.

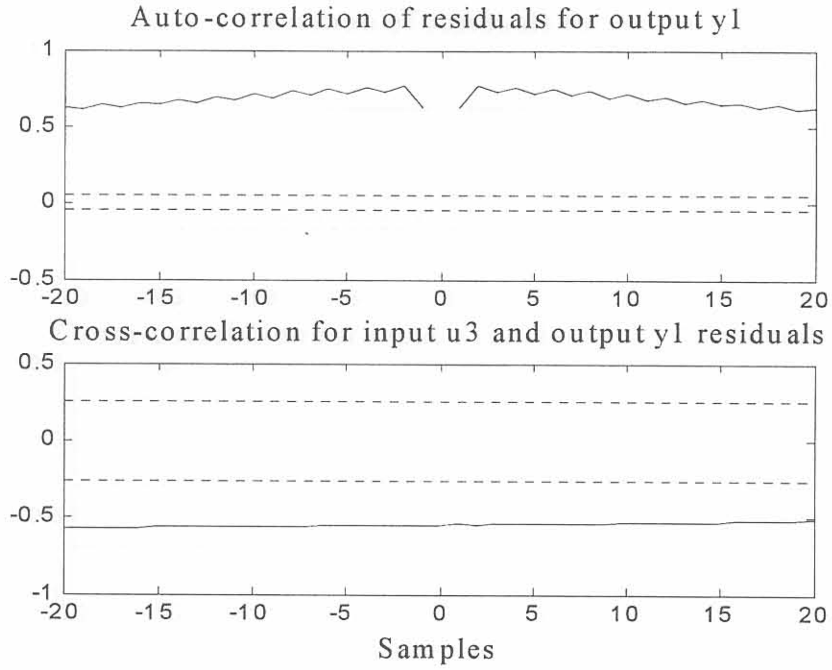

Figure 6.7: The cross-correlation and auto-correlation of the residuals for y_1 and u_3 of the closed-loop identified model.

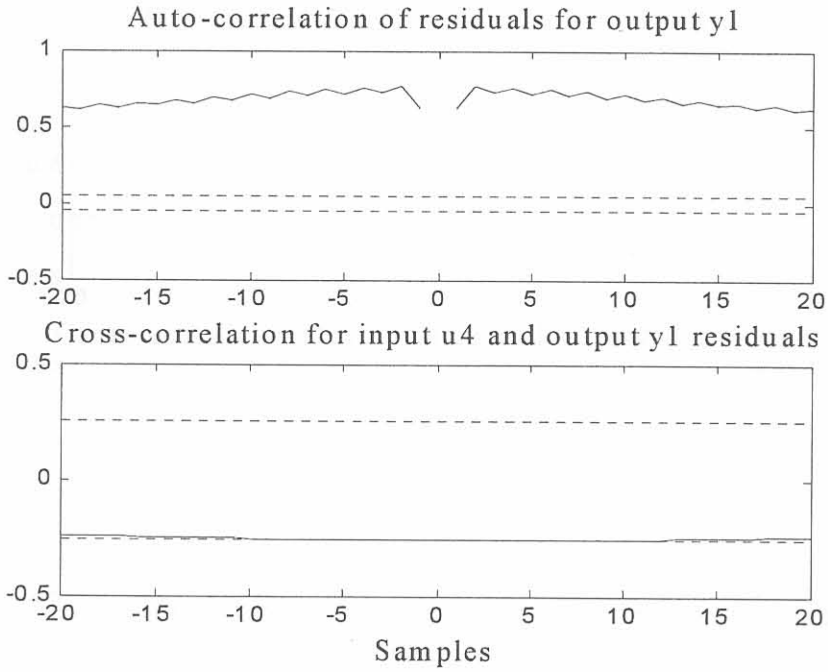

Figure 6.8: The cross-correlation and auto-correlation of the residuals for y_1 and u_4 of the closed-loop identified model.

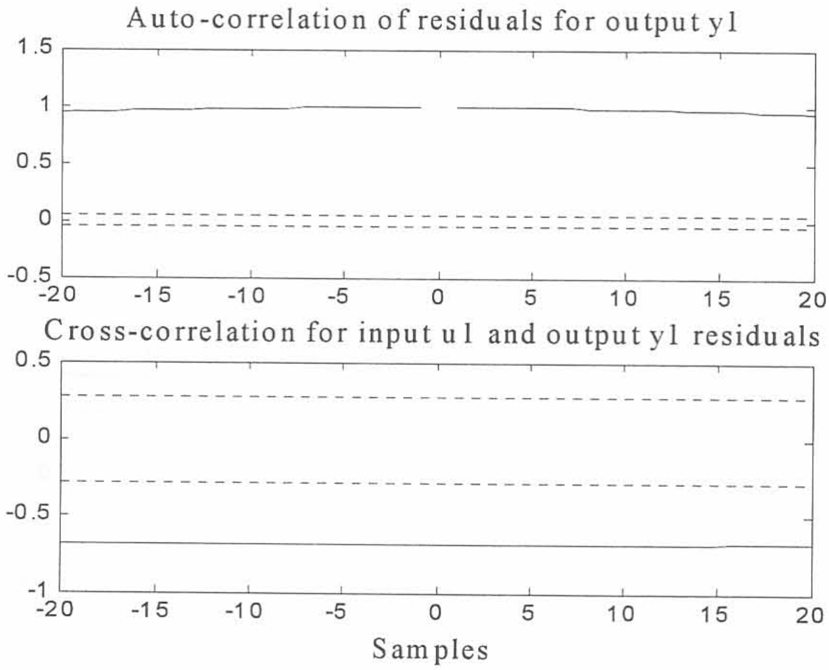

Figure 6.9: The cross-correlation and auto-correlation of the residuals for y_1 and u_1 of the open-loop identified model.

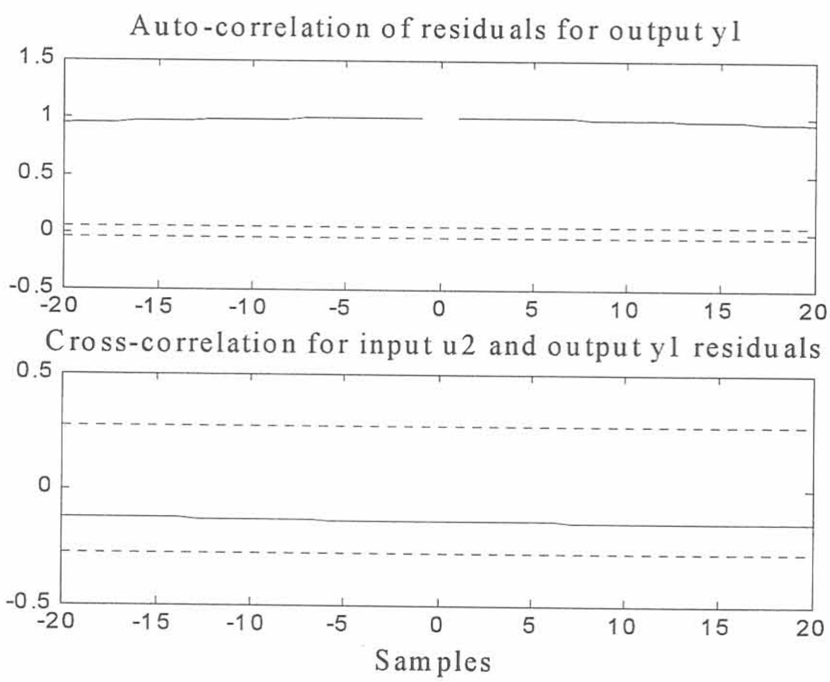

Figure 6.10: The cross-correlation and auto-correlation of the residuals for y_1 and u_2 of the open-loop identified model.

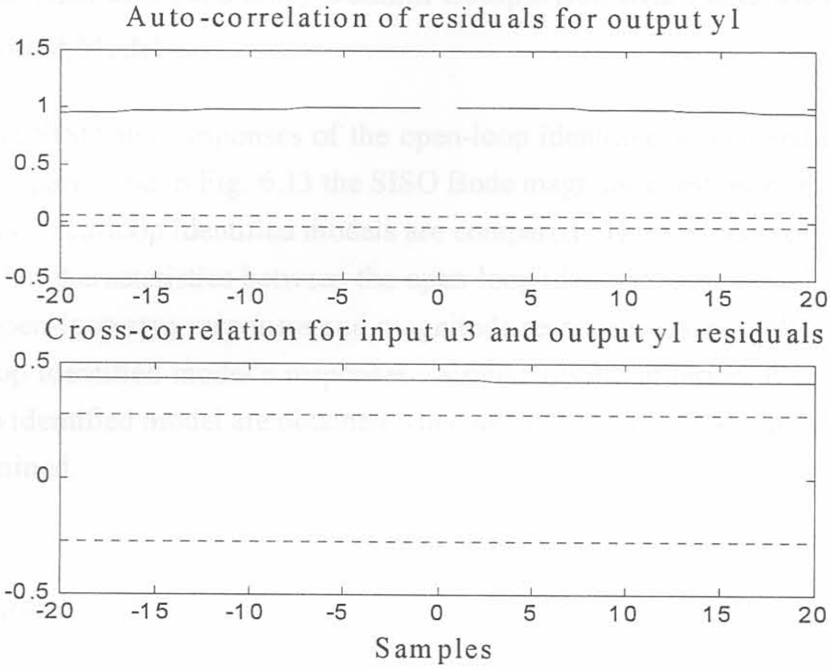

Figure 6.11: The cross-correlation and auto-correlation of the residuals for y_1 and u_3 of the open-loop identified model.

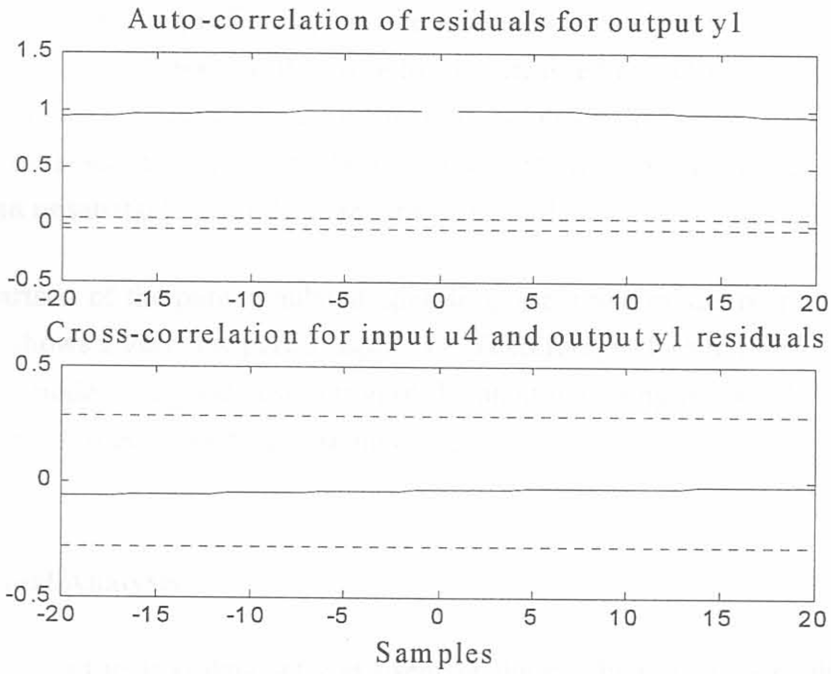

Figure 6.12: The cross-correlation and auto-correlation of the residuals for y_1 and u_4 of the open-loop identified model.

6.3.2.3 Visual Time and Frequency Domain Comparison with the Open-Loop Identified Model

In Fig. 6.1 the SISO step responses of the open-loop identified and closed-loop identified models are compared and in Fig. 6.13 the SISO Bode magnitude responses of the open-loop identified and closed-loop identified models are compared. These plots show that, although there are similar characteristics between the open-loop identified and closed-loop identified models, the open-loop step responses and magnitude responses are not followed closely by the closed-loop identified model's responses. Similar results in terms of comparison with the open-loop identified model are obtained when the impulse responses and the Bode phase plots are examined.

6.3.3 Discussion

6.3.3.1 Simulation and Prediction

When the closed-loop data from which the model was estimated, are reproduced (pure simulation and 6-steps ahead prediction), all the results are quite good, i.e. high percentage fits are obtained. This shows that the closed-loop identified model is able to reproduce the data from which it is estimated and the estimation process is thus satisfactory. However, the model do not satisfactorily reproduce the validation data (low percentage of fit). Therefore, as expected, an unsatisfactory model is identified from the measured closed-loop data.

The comparison of the pure simulated open-loop identified model outputs and the measured outputs shows a very low percentage of fit. Therefore, it does not look as if this open-loop identified model is a good description of the plant operating in closed-loop, during the time in which the closed-loop data were measured.

6.3.3.2 Residual Analysis

Again, when the estimation data set was used for the residual analysis of the closed-loop identified model, all the results are good, i.e. the functions are inside the confidence bounds. This shows that the closed-loop identified model is able to reproduce the data from which it was estimated. However, for the validation data, the functions go significantly outside the

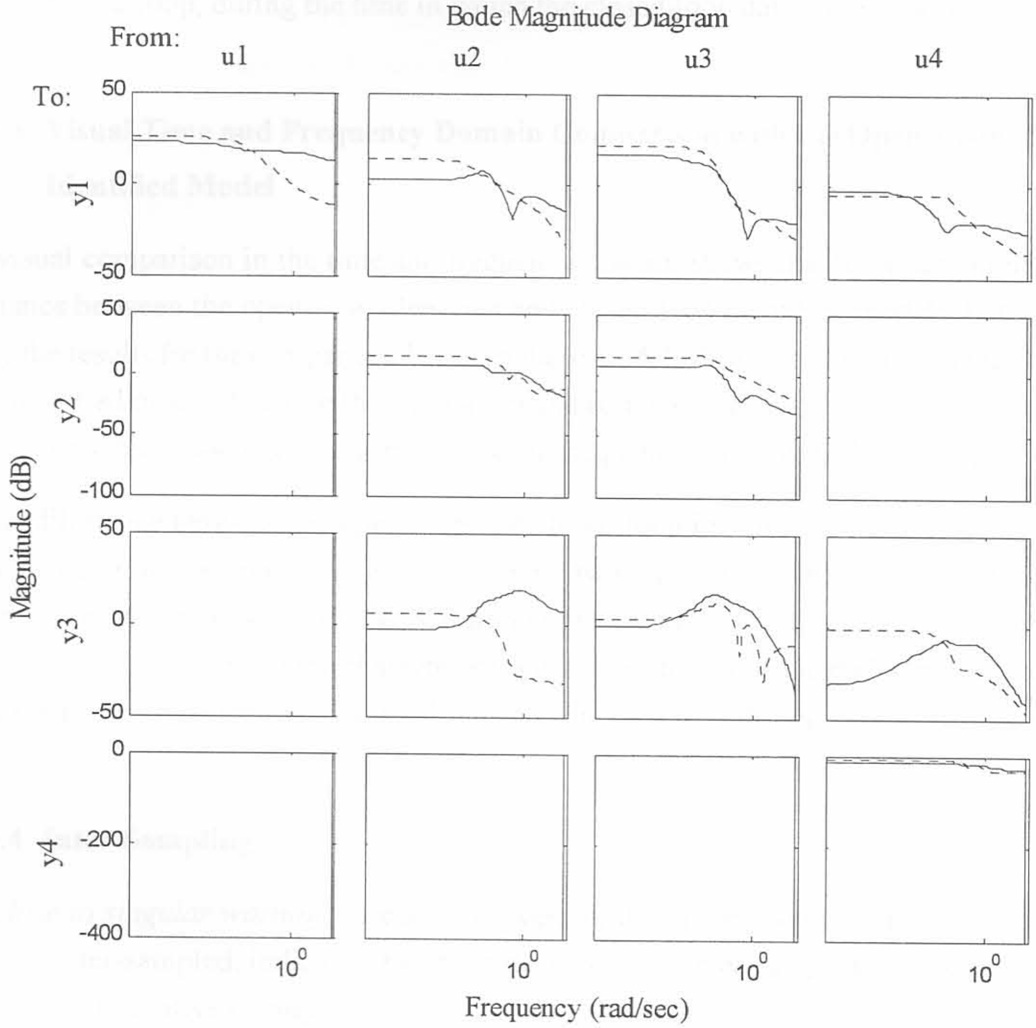

Figure 6.13: The SISO Bode magnitude plots of the open-loop identified (dotted lines) and closed-loop identified (solid lines) models.

confidence bounds. Again, this result shows that an unsatisfactory model is identified from the measured closed-loop data.

The residual analysis of the open-loop identified model also shows that the auto-correlation and cross-correlation functions go outside the confidence bounds. Thus, these results also indicate that the open-loop identified model is not a good description of the plant operating in closed-loop, during the time in which the closed-loop data were measured.

6.3.3.3 Visual Time and Frequency Domain Comparison with the Open-Loop Identified Model

The visual comparison in the time and frequency domain shows that there are similar characteristics between the open-loop identified and closed-loop identified models, but unfortunately the results for the comparison between these models do not say much about the closed-loop identified model, because the simulation and residual analysis show that the open-loop identified model is not a good representation of the plant operating in closed-loop.

The difference between the open-loop and closed-loop identified models can be because of the fact that the frequency weighting for these two types of models are very different and that the plant may have exhibited very different dynamics in closed-loop than in open-loop. Another reason may be that the characteristics of the plant changed in the short time between when the plant was identified in open-loop and when the closed-loop data were measured.

6.3.3.4 Inter-Sampling

The *close to singular warning*, which was given for the model identified from the data that were not inter-sampled, indicates that if the data were not inter-sampled, the data would not have been informative enough.

Therefore, inter-sampling did make the data informative enough. However, the fact that the models, estimated from different ranges in the measured data sets, vary considerably, shows that an imprecise model resulted, as expected.

6.3.3.5 Synopsis

Not much can be learned from the comparison between the open-loop and closed-loop identified models, since the open-loop model is not able to satisfactory reproduce the measured closed-loop data.

From the simulation and prediction, as well as residual analysis, it is concluded that an unsatisfactory model for controller design was identified from data measured under normal closed-loop control. Although the data were informative enough, the absence of structured tests that ensure PE reference signals and good SNRs resulted in an undesirable large variance in the closed-loop identified model. This result emphasizes the requirement for structured tests, which ensure PE reference signals and good SNRs.

6.4 Evaluation of the Proposed Methodology

6.4.1 Reasons for Success/Failure of the Implemented Methodology

From the experiment results it cannot be concluded that the proposed closed-loop SID methodology failed or succeeded, only that a satisfactory model was not identified when the methodology was implemented on the measured process data.

For the simulation data, it is shown that the methodology does work, but that satisfactory results depend on the excitation properties of the reference signals, as well as the SNR.

Since structured tests were not performed on the plant, the excitation properties of the reference signals and the SNRs could not be controlled. Thus, the most probable reason for the unsatisfactory identification of the model in the experiment is that, although the data were informative enough, since it was inter-sampled, the reference signals were probably not PE and the SNRs not good, which resulted in an undesirable large variance in the closed-loop identified model.

6.4.2 Recommendation for Future Implementation of the Methodology

From the results obtained it can be concluded that data, measured when the plant is under normal closed-loop control, are not necessarily informative enough and the SNRs are not necessarily good. Thus, structured test should be conducted. Even with structured tests the proposed identification technique is less intrusive and can reduces re-identification time considerably. These structured tests must ensure that the SNRs are good. In order to ensure informative data, it is easy to use these structured tests to ensure that the reference signals are PE of a sufficiently high order. Care should be taken to design appropriate test signals for the relevant plant.

Typical knowledge available from closed-loop experiments is:

1. measurements of $y(t)$ and $u(t)$,
2. knowledge about excitation properties of $r_a(t)$ and $r_b(t)$,
3. measurements of $r_a(t)$ and $r_b(t)$, and
4. knowledge of $C(q)$.

At least knowledge on (1) and (2) should be obtained in order to implement the proposed methodology satisfactorily. The knowledge of the excitation properties of the reference signals (2) will help in choosing informative data sets with good SNR for identification. The extra knowledge of either (3) or (4) will allow for consistent SID of $G_0(q)$, irrespective of the noise model $H_o(q)$ [31], since then the indirect or joint input-output approach can also be used.

It is also advisable that when future implementation of the proposed methodology is planned, a record should be kept of all previously identified models, i.e. record of time delays, orders, etcetera, as well as all changes to the plant. This extra information will simplify the implementation of the methodology considerably, since one will have an idea of what the structure looks like and how the process changed.

Furthermore, preprocessing of the data is also very important, since the data are not likely to be in shape for immediate use in identification algorithms.

6.5 CONCLUSION

The results obtained in this experiment validate the proposed methodology's requirement for structured tests that ensure PE reference signals and good SNRs. From this experiment it can be concluded that, an unsatisfactory model will most probably be identified from data measured under normal closed-loop control.

According to the simulation results in Chapter 5, the proposed closed-loop system identification methodology will give reliable results for MIMO plants controlled by MPC controllers, for the type of system disturbances and constraints used in the simulation, as long as structured tests that ensure PE reference signals and good SNRs are used.

Thus, a preliminary conclusion from this experimental results is that structured tests are needed to ensure that the proposed closed-loop SID methodology delivers reliable results. However, in future, structured test, with PE reference signals, on a real process are still needed, to validate the simulation results.