A NARMAX model representation for adaptive control based on local models

TOR ARNE JOHANSEN† and BJARNE A. FOSS†

Here we address the problem of representing NARMAX (nonlinear ARMAX) models with application to adaptive control. We propose a nonlinear model representation where a number of simple local models are combined. The local models are valid in specific operation regimes of the process. Explicitly defined model validity functions make it possible to combine the local models by interpolation. During online identification, only the local models corresponding to the current process operation regime are updated. It is therefore not necessary to relearn the model each time there is a change in the operation regime of the process. The concept is illustrated by a simulation example of a nonlinear pH-neutralization process.

1. Motivation

A process must often work under different operation conditions. A change in operating conditions may be caused by process nonlinearities, time-varying parameters, significant disturbances, batch operation, setpoint changes, startup and shutdown, component failure or process maintenance. There are several approaches to control a process under different operation conditions:

- The use of a simple model of the process, and the design of a robust controller that works well under significant model uncertainty.
- A controller based on a nonlinear model of the process that is valid for a large set of operation conditions.
- An adaptive controller that will try to compensate for varying process characteristics using online identification.
- Gain scheduling (Åström and Wittenmark, 1989) can be applied to choose a pre-designed controller depending on current operation conditions.
- Some kind of 'model scheduling' (Skeppstedt, 1988) (Skeppstedt *et al.* 1992) (Hilhorst *et al.* 1991) and (Sørheim, 1990) to choose the best model corresponding to the current operation regime.

However, all these have a number of drawbacks. A robust controller will have to sacrifice bandwidth to achieve robustness. A nonlinear state-space model that is valid for different operation conditions may be difficult to expensive to find. Gain scheduling usually does not contain a feedback loop to decide if the chosen controller performs well. Adaptive or self-tuning controllers are often based on a linearized input/output-model, for instance an ARMAX model. When the process changes operation regime, it
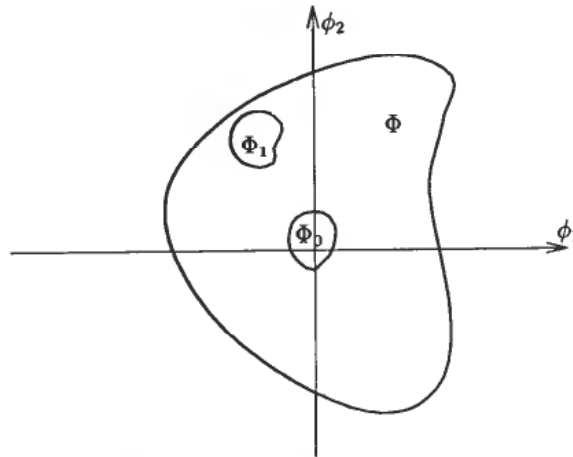
T. A. Johansen and B. A. Foss

Figure 1.   A typical situation. The set of possible operation points is Φ, but the process typically operates in the operation regimes Φ_0 and Φ_1.
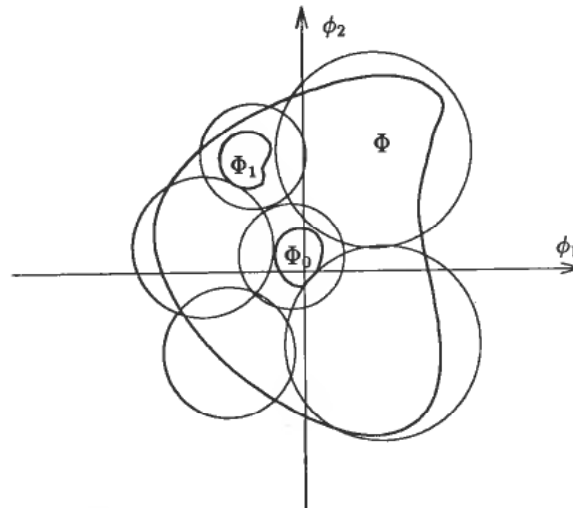

Figure 2.   Several local models are used to cover all possible operation points in Φ.

is necessary to discard old information in order to adapt to the process characteristics in the new operation regime, since the linear ARMAX model will be a linearization of the system about the current operation point. Since the parameter estimator should be slower than the process dynamics in order to be robust, the controller can exhibit poor performance during rapid transition between operation regimes.

Our approach (Johansen and Foss, 1992a) to this problem is illustrated by Fig. 1, which shows a typical situation. Let Φ be the set of all operation points. A vector $\phi \in \Phi$ is a possible operation point for the process. The process will typically operate in $\Phi_0 \subset \Phi$ (typically a small subset) or $\Phi_1 \subset \Phi$. If we have a set of simple linear local models describing the process well in different parts of Φ, as shown in Fig. 2, we can form a complete model by introducing smooth interpolation between the local models. This requires explicitly defined model validity functions for each local model. These functions are defined over the set of operation points, and indicate the validity of the local models as a function of the operation point, which is assumed to be known.

During on-line identification, only the local models which have significant weight will be updated. This implies that only the local models which are relevant for the current operation regime will have their parameters updated, the other local models will remain unchanged. Hence, information about other operation regimes will not be forgotten, and this information can be recalled instead of being relearned when the process re-enters an operation regime. The goal of a controlled process is often to keep certain process states near their set-points. Hence, a very small part of the input-space is normally excited and it is important not to forget information from operation regions that are rarely entered.

2. Model representation

For linear systems, a popular family of models is the ARMAX models

$$y(t) = \theta^T \psi(t-1) + e(t) \tag{1}$$

where $y(t)$ is the system output at time t, $e(t)$ is the equation error at time t, θ is a parameter vector and $\psi(t-1) = [y(t-1) \ldots y(t-n)u(t-1) \ldots u(t-n)e(t-1) \ldots e(t-n)]^T$ is the information vector at time $t-1$. The ARMAX model family are input/output models. This means that the model structure and parameters have no direct interpretation in terms of physical phenomena or physical parameters. Input/output-models may be preferred instead of state-space models when our process knowledge is limited and it is a resource demanding task to develop first principles models. This will often be the case for complex nonlinear systems. In that case the NARMAX (nonlinear ARMAX) (Leontaritis and Billings, 1985), (Chen and Billings, 1989) model representation can represent a large class of dynamical systems

$$y(t) = f(\psi(t-1)) + e(t) \tag{2}$$

where f is a nonlinear function with range and domain given by the output- and information-vectors. The problem is now how to represent this function. If we assume that our process knowledge is limited, a generic nonlinear structure for f is required. Among the alternatives are polynomials (Chen and Billings, 1989) and neural networks (Chen and Billings, 1990), (Chen *et al.* 1990).

A generic representation based on local models is proposed (Johansen and Foss, 1992a). By definition, a local model is valid only within a limited operation regime. With the local model representation we try to reduce the problem of building a global model to the problem of building a set of local models. In order to sew these local models together to a global model, we use interpolation.

Assume that we have given a set of local models $\hat{f}_i$ such that

$$\hat{y}(t) = \hat{f}_i(\psi(t-1)) \tag{3}$$

To each such local model we associate a model validity function $\rho_i : \Phi \to [0, 1]$ which by definition is close to 1 for those operation points ϕ where the model $\hat{f}_i$ is a good model and close to 0 elsewhere. Having knowledge of these is a strong assumption, but experience shows that the choice of ρ_i is not critical. A typical choice for ρ_i is the Gaussian function shown in Fig. 3. We now define a set of normalized interpolation functions $w_i : \Phi \to [0, 1]$

$$w_i(\phi) = \frac{\rho_i(\phi)}{\sum\limits_{j=0}^{N-1} \rho_j(\phi)} \tag{4}$$
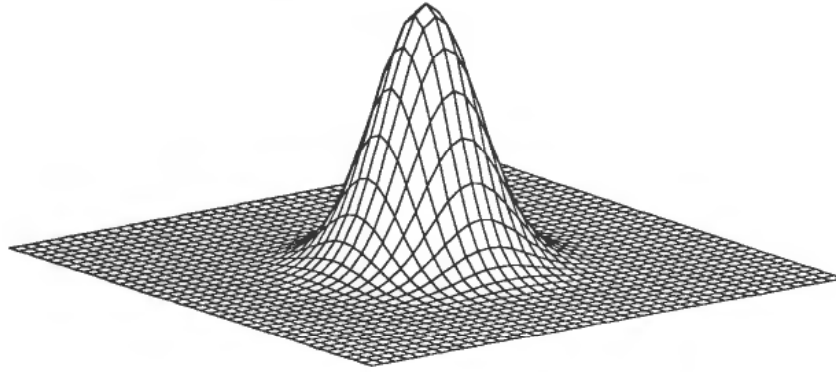
T. A. Johansen and B. A. Foss

Figure 3.   A typical model validity function ρ_i defined on the set of possible operation points Φ.

By this definition we know that for any $\phi \in \Phi$ will $\Sigma_{i=0}^{N-1} w_i(\phi) = 1$. If we assume that the real system is given by (2), then the following identity will hold

$$y(t) = f(\psi(t-1)) + e(t) \equiv \sum_{i=0}^{N-1} f(\psi(t-1))w_i(\phi(t-1)) + e(t) \qquad (5)$$

Since w_i is a normalization of ρ_i, we know that $w_i(\phi)$ is close to 0 for operation points ϕ where $\hat{f}_i$ is not assumed to be a good model, expect for those $\phi \in \Phi$ where none of the local models are good. Hence we can substitute $\hat{f}_i$ for f on the right-hand side of (5) without losing too much accuracy in the model. This is rigorously analysed in (Johansen and Foss, 1992c). Given a sufficient set of local models and sensible choices for the operation point vector and model validity functions, the model will approximate the system with any given accuracy. This leads to the proposed representation

$$\hat{y}(t) = \sum_{i=0}^{N-1} \hat{f}_i(\psi(t-1))w_i(\phi(t-1)) \qquad (6)$$

The operation point $\phi(t-1)$ will typically be directly related to $\psi(t-1)$, since this vector contains the inputs, outputs and in some cases the measured disturbances of the process. In most cases there will exist a mapping $\psi(t) \mapsto \phi(t)$, such that the operation point at any time can be calculated as a function of the information vector. This mapping will typically be a projection onto some subspace of the information space. In some cases it is not possible to find such a direct mapping. This is the case if some estimator is used to find quantities describing the operation point. Typically process knowledge about different operation regimes or the validity of the available local models will be used to define the validity functions ρ_i.

The local models can have different structures. A 0th order Taylor-series expansion of f about ψ_i gives a local model

$$\hat{f}_i(\psi(t-1)) = f(\psi_i) = \theta_i^0 \qquad (7)$$

where θ_i^0 is a parameter. In this case we typically have $\phi = \psi$. The model can be interpreted as a radial basis-function neural network if ρ_i is a radial function (Moody and Darken, 1989), (Stokbro *et al.* 1990). This is a very simple local model, since the value of f at a point ψ is extrapolated to nearby points. By interpolation between such simple models, this simple local representation is sensible.

A 1st order Taylor-series expansion of f around ψ_i gives a local linear model

$$\hat{f}_i(\psi(t-1)) = f(\psi_i) + \nabla f(\psi_i)(\psi(t-1) - \psi_i) = \theta_i^0 + \theta_i^T(\psi(t-1) - \psi_i) \qquad (8)$$

where θ_i^0 and θ_i are parameters. The local model will in this case be an ARMAX model. This is related to the approaches by Stokbro *et al.* (1990) and Jones *et al.* (1991).

Higher order expansions or nonlinear local models like neural networks may be applied. Local state-space models based on first principles can be used. Physical knowledge can thus be integrated with input/output type models using this representation (Johansen and Foss, 1992b).

3. Learning

An important property of (6) is that if all the local models f_i are linear in the parameters, the nonlinear model composed by interpolating the local models will also be linear in the parameters. Hence, in the simple case when we have uncorrelated noise, the model can be written in the linear regression form

$$\hat{y}(t) = \theta^T \varphi(t-1) \qquad (9)$$

where θ contains all the parameters, and $\varphi(t-1)$ is a nonlinear regression vector. In this case a least squares estimator may be used. When the information vector $\psi(t-1)$ contains delayed error terms, a prediction error estimator must be used if unbiased estimates are desired. Standard system identification algorithms may therefore be applied (Söderström and Stoica, 1988). The performance index used for learning can be based on the total prediction error

$$J = \frac{1}{2} \sum_{t=1}^{T} \left(y(t) - \sum_{i=0}^{N-1} \hat{f}_i(\psi(t-1)) w_i(\phi(t-1)) \right)^2 \qquad (10)$$

or the local prediction errors

$$J_i = \frac{1}{2} \sum_{t=1}^{T} (y(t) - \hat{f}_i(\psi(t-1)))^2 \qquad (11)$$

Using the local prediction errors, we may choose to keep some of the local model parameters fixed, and choose different algorithms for each local model. The disadvantage of the local indices (11) is that the interpolation-functions are not included in the index, implying that the model validity functions that have not been well chosen will not be compensated for during learning. The advantage of the index (10) is that because the shape of the interpolation-functions is incorporated in the model, this criteria may give a more accurate prediction. However, a local linear model will not be a linearization of f in (2), since the optimal parameter values will be influenced by the model validity functions.

Only local models that have a weight above some limit should be updated, if forgetting is to be reduced. Another possibility is to only update the local model that has most weight. The amount of information a model is capable of storing is limited. Using a standard ARMAX model and an RLS- or RPE-algorithm (Söderström and Stoica, 1988) with a constant forgetting factor, information will be forgotten at a constant rate. If the information vector do not contain much information, e.g. due to lack of persistence of excitation, the model will eventually contain very little information about the process. Using a model composed of several local models, and updating only the local models with large weight will reduce this problem considerably.

Only the information content of the model being updated will be changed. Hence, there will be no forgetting about the operation regimes the process is not currently operating in. Since relearning because of changing operation regimes is generally not needed using this model representation, learning is only needed to compensate for slowly varying process parameters and disturbances. If the disturbances vary slowly compared to the process dynamics, the learning rate may be chosen to be small.

The model (6) can be viewed mathematically as a series expansion of the function f. Each term in the series expansion will be orthogonal to all other terms except those which have overlapping model validity functions. Informally, if the model validity functions do not overlap too much, a local model (corresponding to a term) will be 'almost orthogonal' to most other local models. From this one may believe that it is possible to include a new local model without disturbing the other local models, except in the neighbourhood of the new local model. This means that the structure of the model representation (6) can be changed without losing significant information. It also implies that the model accuracy will improve by adding new 'almost orthogonal' terms (Johansen and Foss, 1992b).

4. Simulation example

We have performed a simulation study on a pH-control system, in order to illustrate the concept. Consider the stirred mixing tank in Fig. 4. The mass balance for the tank is

$$V\frac{d}{dt}\,\delta x(t) = -(q_A(t)+q_B(t))\delta x(t)+c_A(t)q_A(t)-c_B(t)q_B(t) \qquad (12)$$

and the pH is given by the titration curves in Fig. 5. The symbols are defined in Table 1.

An ideal level controller keeps the tank volume constant. This gives a holdup-time $T=V/(q_A+q_B)$ in the tank in the interval $100\,\mathrm{s} \leqslant T \leqslant 400\,\mathrm{s}$. Due to stirring there is a dead-time τ, and the pH-electrode and the valve have first order dynamics.
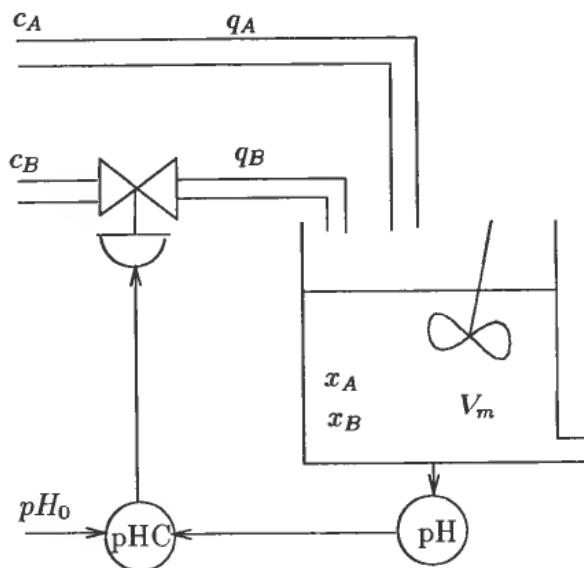

Figure 4.   Stirred tank with pH-control. A level controller (not shown) keeps the level constant.
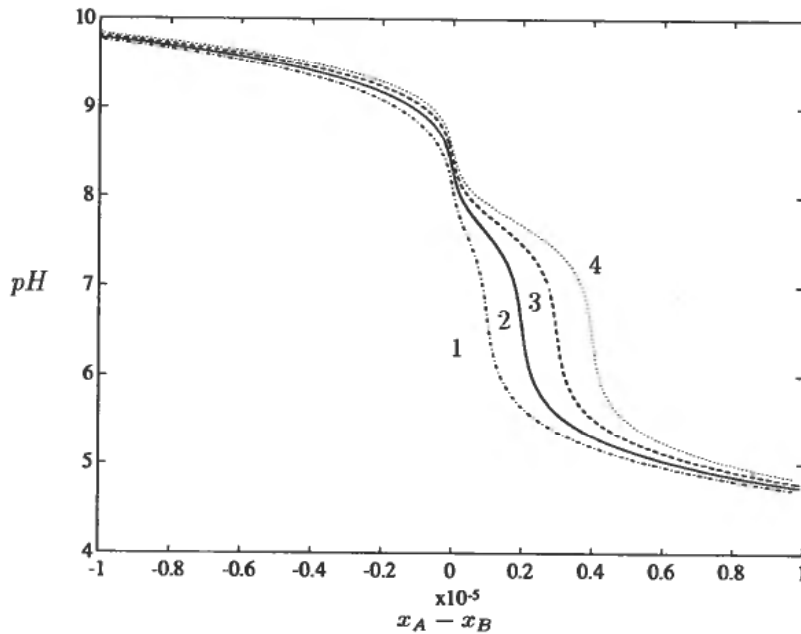
Figure 5. Titration-curves for acid-base mixture. The different curves are for different buffer contents.

Symbol	Value	Unit	
c_A	0·000025	mol/l	Acid concentration in input, fixed
c_B	0·0001	mol/l	Base concentration in input fixed
q_A	0·025 − 0·1	l/min	Acid flow, disturbance
q_B	≈ 0·01	l/min	Base flow, control input
x_A		mol/l	Acid concentration in tank
x_B		mol/l	Base concentration in tank
δx	$x_A - x_B$	mol/l	
pH			pH in tank, measurement
pH_0	7·2		Setpoint
V	10	l	Tank volume
τ	5	s	Dead-time due to stirring
T_{pH}	5	s	Time-constant in pH-sensor
T_q	2	s	Time-constant in valve

Table 1. Definition of symbols.

Due to varying buffer contents the titration curve will vary continuously between the ones shown in Fig. 5. This variation is assumed to be slow compared to the dynamics of the process and control system. The process gain is proportional to the slope of the titration curve. We see that the slope of the titration curve varies both with pH and also with buffer contents. This is a difficult nonlinear control problem due to several factors (Waller and Gustafsson, 1983). The strong nonlinearity of the process gain requires a compensating nonlinearity in the controller. The time-variation of the gain due to variations in buffer contents calls for adaptation. The kinetics of the neutralization process is very fast, and together with dynamics in the pH-electrode and high gain near pH = 7, this makes the pH measurement signal inaccurate. The dead-time due to stirring gives some limitation on the attainable bandwidth of the closed loop.

The simulations cover all these factors, but the disturbances and time-variations on the process may be smaller in the simulations than in many industrial plants, hence the simulations should be regarded as semi-realistic with the only purpose of illustrating the modeling concept.

We define the operation point to be equal to the measured pH-value at any time, i.e. $\phi(t) = \text{pH}(t)$. In a more advanced control system the operation point could also include information about the flows into the tank and buffer contents (titration curve). Since such information is not directly available, we only use measured pH, which should make the model able to capture variations in process gain due to different pH-values. Since the buffer contents is slowly varying, a slow estimator is used to identify the parameters of each of the local models on-line. 5 local models are applied. They are centred at the following pH-values, $\text{pH}_1 = 6.2$, $\text{pH}_2 = 6.7$, $\text{pH}_3 = 7.2$, $\text{pH}_4 = 7.7$ and $\text{pH}_5 = 8.2$ in order to take care of disturbances taking the pH away from the setpoint at 7.2. Each of the local models is an ARX model

$$\widehat{\Delta \text{pH}}(t) = a_i \Delta \text{pH}(t-1) + b_i \Delta q_B(t-1) = \theta^T(\psi(t-1) - \psi_i) \tag{13}$$

where $\theta = [a_i \, b_i]^T$, $\psi(t-1) = [\text{pH}(t-1) \, q_B(t-1)]^T$ and $\psi_i = [\text{pH}_i \, q_B^0]^T$ where $q_B^0 = 0.01 \, \text{l/s}$ is nominal flow. The two parameters (for each of the 5 local models) are estimated on-line. A slow LMS-estimator (Åström and Wittenmark, 1989) is used to illustrate that adaptation is not needed to adapt to changes in pH, but only to buffer contents. We update only local models that have weight above some threshold, and we use the local prediction errors (11)

$$\hat{b}_i(t) = \begin{cases} \hat{b}_i(t-1) + \eta \Delta q_B(t-1) \epsilon_i(t) & \text{if } \rho_i(\phi(t)) \geqslant 0.4 \\ \hat{b}_i(t-1) & \text{otherwise} \end{cases} \tag{14}$$

where the local model prediction errors are

$$\epsilon_i(t) = y(t) - \hat{f}_i(\psi(t-1)) \tag{15}$$

Using the global prediction error gives qualitatively equal results in this case, and may in general be preferred. The prediction horizon for the model is chosen as 15 s because of the dead-time and dynamics in sensor and actuators. The sampling interval for the controller is 1 s. As illustrated in Fig. 6 the local model validity functions are chosen as Gaussians centred at each of the 5 operation points.

The controller is a discrete PI-controller, where the integral time is kept fixed at $T_i = 60 \, \text{s}$ and the controller gain is proportional to the inverse of the estimated process gain. The process gain estimate is

$$\hat{b}(t) = \sum_{i=1}^{N} \hat{b}_i(t) w_i(\phi(t)) \tag{16}$$

For comparison purposes we use a PI-controller with fixed gain $K_p = 0.005$, and a gain scheduling controller with 5 gains as shown in Table 2. The gain is scheduled as a function of pH.

Figure 7 shows the simulation results for titration curve 4, see Fig. 5. We see that the adaptive-PI controller based on local models will compensate faster for the disturbance in q_A than both the fixed-gain PI-controller and the gain-scheduled PI-controller. Comparing Figs 7, 9 and 5 we see that the adaptive PI-controller increases the controller gain when the process gain is low and vice versa. The gain-scheduler does the
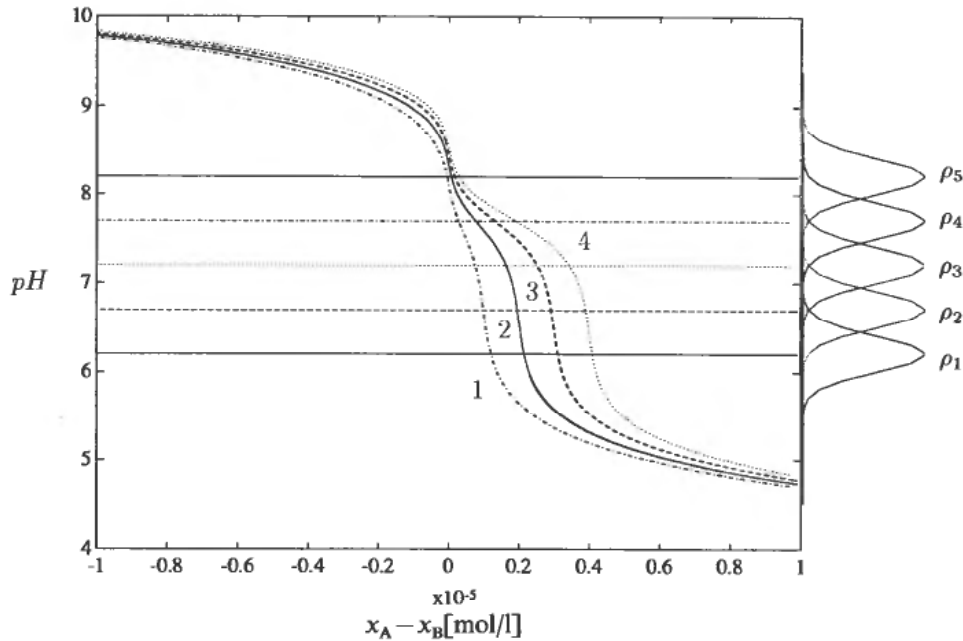
Figure 6.   Five local models are used. The associated model validity functions ρ_i are shown at right.

Measured pH	K_p (PI-controller gain)
pH > 7·95	0·01
7·95 ⩾ pH > 7·45	0·008
7·45 ⩾ pH > 6·95	0·005
6·95 ⩾ pH > 6·45	0·003
pH ⩽ 6·45	0·004

Table 2.   Gain scheduling table.

same, but ends up with slightly worse compensation. From Fig. 8 we see that improved control does not require significantly larger control inputs, and from figure 10 we see that the parameter estimates $\hat{b}_i$ is updated slowly, so the adaptation does not have significant influence on the controller gain.

When the titration curve slowly approaches the curve 1 in Fig. 5 the controller adapts and gives results as illustrated in Fig. 11. The difference between the controllers is now smaller, as would be expected, since the process gain is now larger for all pH-values. The controller gain is not increased as much as earlier, cf. Fig. 13, and the control signals do not differ much for the different controllers. The parameter estimates have grown since the process gain has increased throughout the operation regimes, see Fig. 14.

5.   Discussions and conclusions

The simulations have shown that using a local models representation, we have been able to build a model that is able to describe the nonlinear behaviour of the process and that this model is applied to a simple adaptive PI-controller where only the gain is
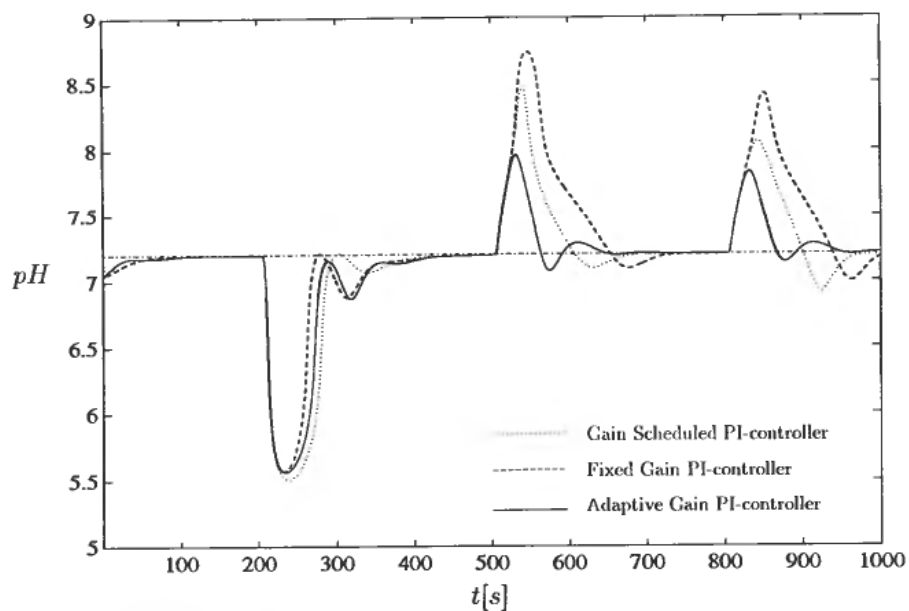
T. A. Johansen and B. A. Foss

Figure 7. The figure illustrates the pH for 3 controllers when the titration curve is curve 4.
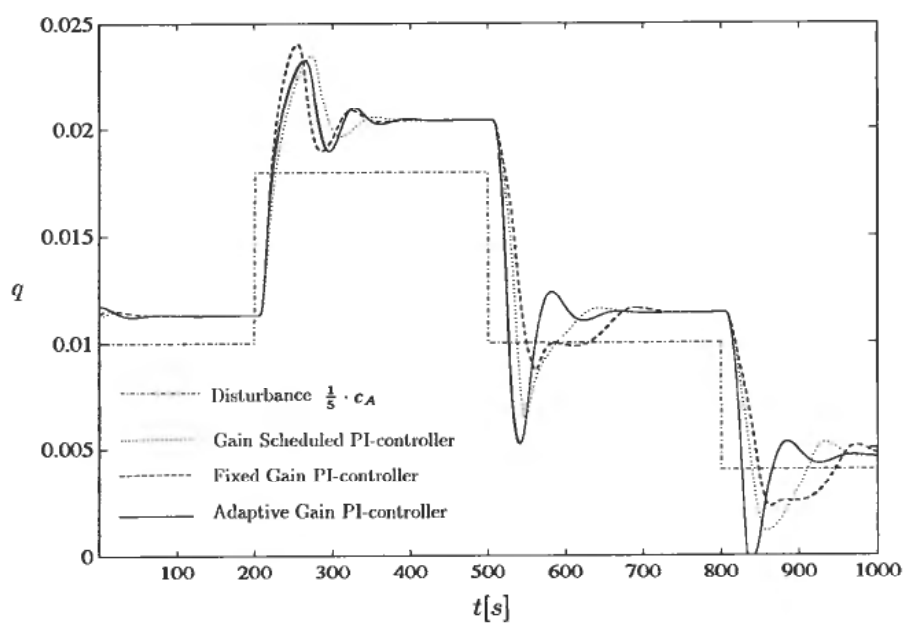

Figure 8. The figure illustrates the flow for 3 controllers when the titration curve is curve 4.
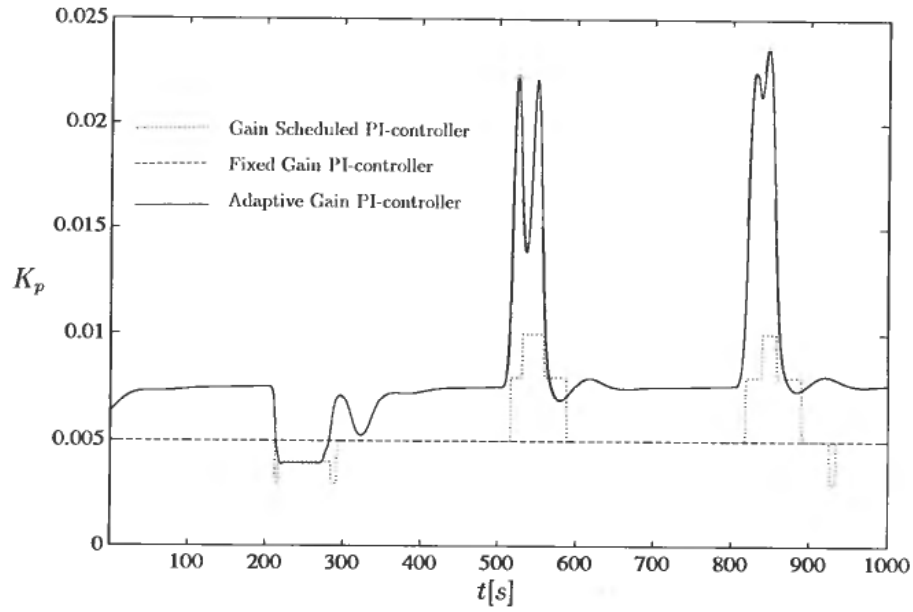
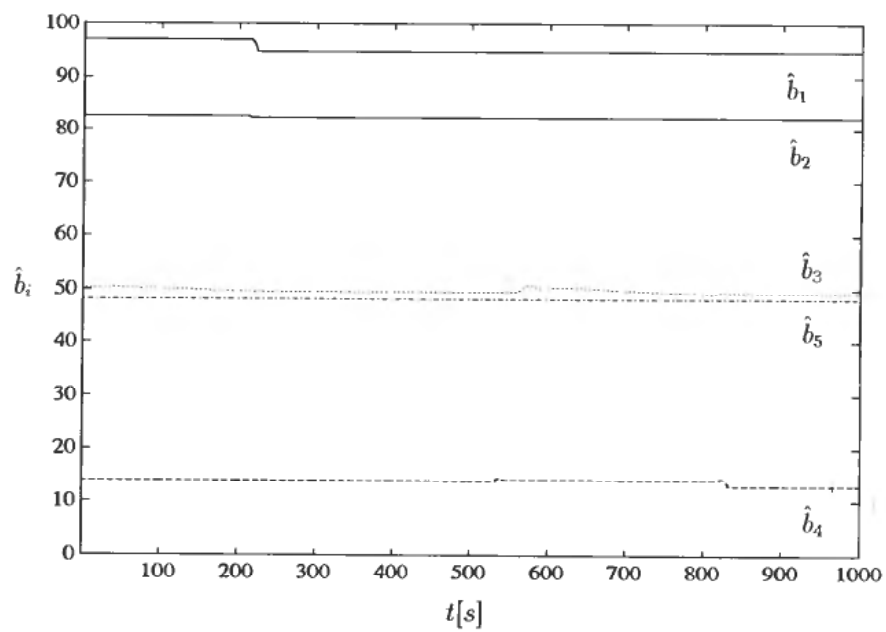Figure 9. The figure illustrates 3 controller gains when the titration curve is curve 4.

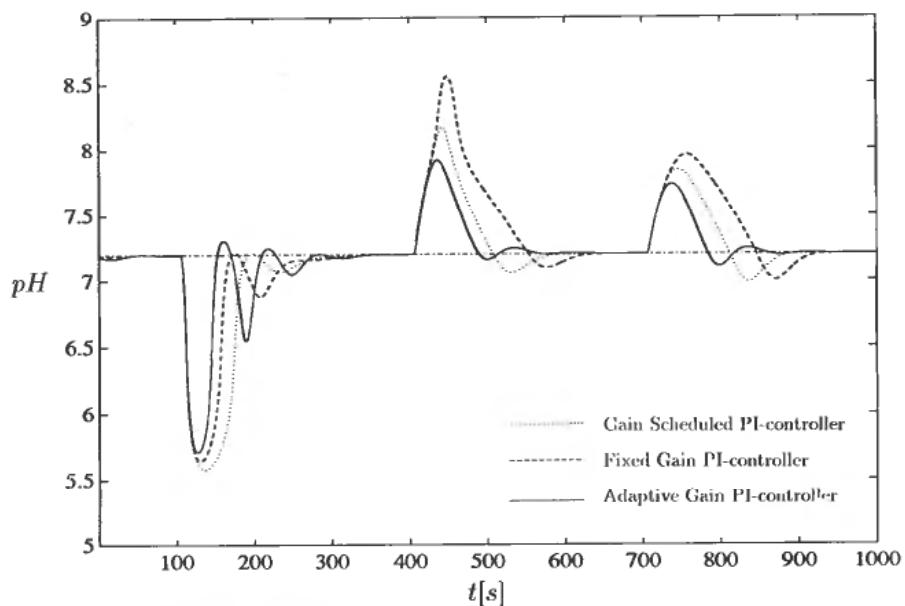Figure 10. The figure illustrates parameter estimates when the titration curve is curve 4.

Figure 11.   The figure illustrates the pH for 3 controllers when the titration curve is curve 2.
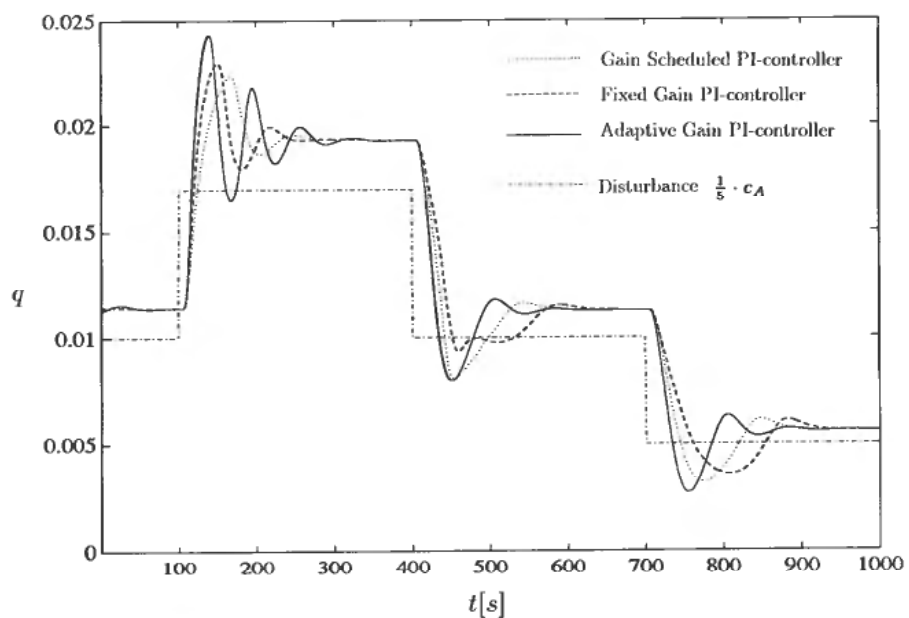

Figure 12.   The figure illustrates the flow for 3 controllers when the titration curve is curve 2.
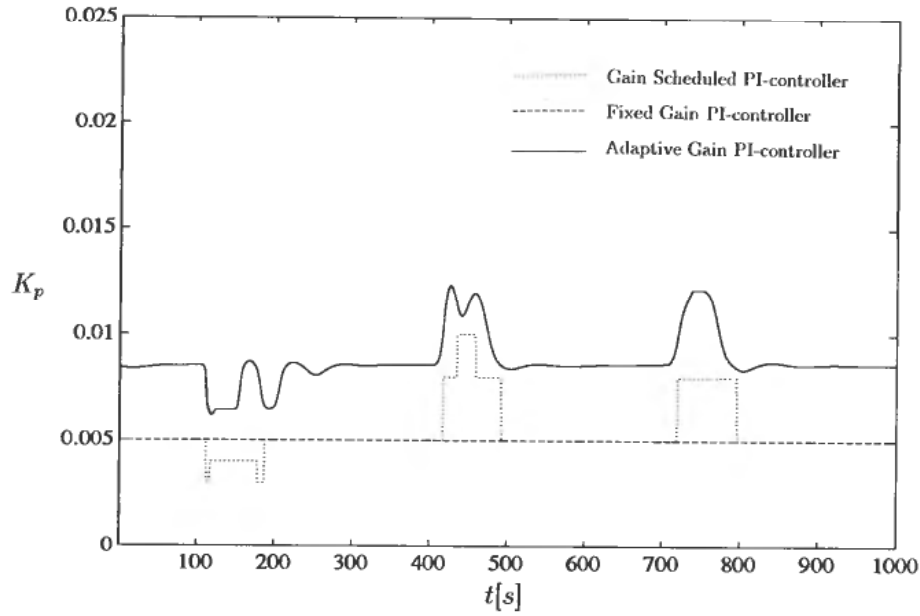
Figure 13. The figure illustrates 3 controller gains when the titration curve is curve 2.
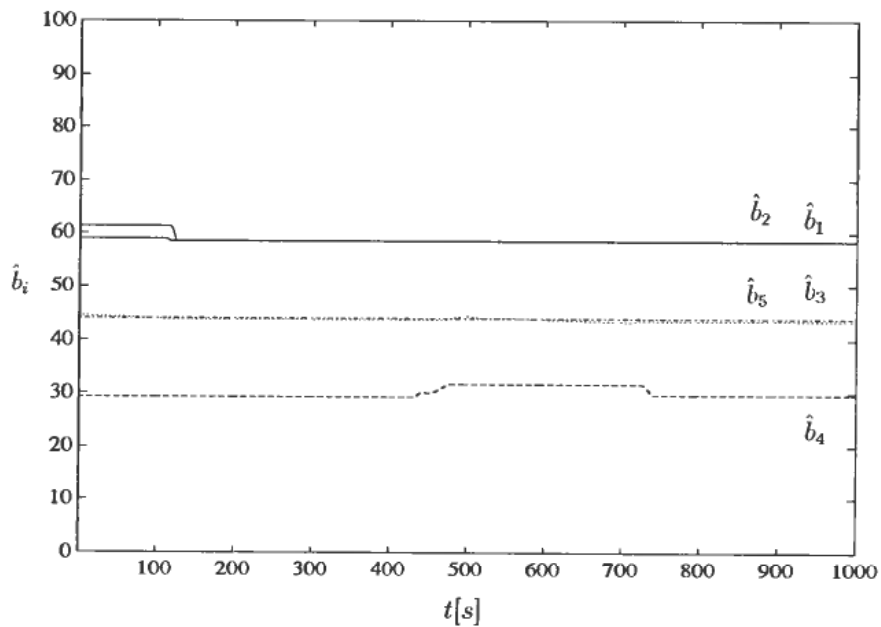

Figure 14. The figure illustrates parameter estimates when the titration curve is curve 2.

adapted on the basis of the nonlinear model. The controller performs well compared to a fixed-gain PI-controller and a gain-scheduled PI-controller, indicating that the model is sensible.

The main idea is to store information about several operation regimes in the model, making relearning unnecessary. Hence, adaption is only needed to take care of slowly varying parameters or disturbances. A global NARMAX model is built using simple local models like ARMAX models. It is also possible to use the proposed model representation in a non-adaptive context to build models.

A possible disadvantage of the proposed model representation is that the number of parameters may be larger than necessary. From the Parsimony principle we know that it is important to try to keep the number of local models at a minimum.

The method is presented here for SISO-models, but can be easily extended to MIMO-models in a straightforward manner (Johansen and Foss, 1992c). There are several open questions, particularly regarding the robustness with respect to the choice of model validity functions. These validity functions must generally be chosen heuristically, and it is desirable that the model is not sensitive to the choice of shape and width of these functions.

REFERENCES

ÅSTRÖM, K. J. and WITTENMARK, B. (1989). *Adaptive Control* (Addison Wesley).

CHEN, S. and BILLINGS, S. A. (1989). Representation of non-linear systems: the NARMAX model. *Int. J. Control*, **49**, 1013–1032.

CHEN, S., BILLINGS, S. A. and GRANT, P. M. (1990). Non-linear system identification using neural networks. *Int. J. Control*, **51**, 1191–1214.

CHEN, S., BILLINGS, S. A., COWAN, C. F. N. and GRANT, P. M. (1990). Practical identification of NARMAX models using radial basis functions. *Int. Journal of Control*, **52**, 1327–1350.

HILHORST, R. A., VAN AMERONGEN, J. and LÖHNBERG, P. (1991). Intelligent adaptive control of mode-switch processes. In *Proc. IFAC International Symposium on Intelligent Tuning and Adaptive Control*, Singapore, January 1991.

JOHANSEN, T. A. and FOSS, B. A. (1992a). Nonlinear local model representation for adaptive systems. *Int. Conf. on Intelligent Control and Instrumentation*, February 1992, Vol. 2, 677–682.

JOHANSEN, T. A. and FOSS, B. A. (1992b). Representing and learning unmodeled dynamics with neural network memories. To be presented at *American Control Conference*, Chicago, June 1992.

JOHANSEN, T. A. and FOSS, B. A. (1992c). Constructing NARMAX models using ARMAX models. Submitted to *Int. J. Control*.

JONES, R. D. *et al.* (1991). Nonlinear adaptive networks: A little theory, a few applications. Technical Report 91-273, Los Alamos National Lab., New Mexico.

LEONTARITIS, I. J. and BILLINGS, S. A. (1985). Input–output parametric models for non-linear systems. *Int. Journal of Control*, **41**, 303–344.

MOODY, J. and DARKEN, C. J. (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation*, **1**, 281–294.

SKEPPSTEDT, A. (1988). Construction of Composite Models from Large Data-sets. PhD thesis, University of Linköping.

SKEPPSTEDT, A., LJUNG, L. and MILLNERT, M. (1992). Construction of composite models from observed data. *Int. J. Control*, **55**, 141–152.

SÖDERSTRÖM, T. and STOICA, P. (1988) *System Identification* (Prentice Hall).

SØRHEIM, E. (1990). A combined network architecture using ART2 and back propagation for adaptive estimation of dynamical processes. *Modeling, Identification and Control*, **11**, 191–199.

STOKBRO, K., HERTZ, J. A. and UMBERGER, D. K. (1990). Exploiting neurons with localized receptive fields to learn chaos. Preprint 28, Niels Bohr Institute and NORDITA, Copenhagen.

WALLER, K. V. and GUSTAFSSON, T. K. (1983) Fundamental properties of continuous pH control. *ISA Transactions*, **22**, 25–34.