Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Hybrid Operating Regime Selection Algorithm in Local Modeling ⋆

Katsuji Uosaki * **Toshiharu Hatanaka** **

* *Department of Information and Management Sciences, Fukui University of Technology, Gakuen, Fukui 910-8505, Japan (e-mail: uosaki@fukui-ut.ac.jp)*
** *Department of Information and Physical Sciences, Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565-0871, Japan (e-mail: hatanaka@ist.osaka-u.ac.jp)*

Abstract: Recently, local modeling has been received much attention to identify the complex systems. In local modeling, global system model is obtained by combining a number of local models, each of which has simpler structure and has a range of validity less than the full range of operation. Since the local models are identified for corresponding local operating regimes, the performance of the global model is highly affected by the choice of the local operating regimes. This paper addresses automatic selection algorithms of suitable local regimes in local modeling. Based on three criteria, Kullback Discrimination Information (KDI), Akaike Information Criterion (AIC), and Mean Square Error (MSE), new hybrid regime selection algorithms are developed by combining with regime integration and partition processes. Numerical simulation studies illustrate the applicability of the proposed selection algorithms.

1. INTRODUCTION

Recent technological development makes engineering systems much more complex, and practical approaches to deal with such systems easily are requested. Since these complex systems are usually composed by a huge number of components, which are strongly related with each other and have wide range of operation, it is difficult to construct a global model applicable to the full range of operation. Hence, an idea of local modeling (Johansen and Foss [1995], Johansen and Foss [1997], Murray-Smith and Johansen [1997]), which closely relates to piecewise affine systems, switching systems, and hybrid systems (Roll [2003], Pepke *et.al.* [2004], Juloski *et al.* [2006], Padletti *et al.* [2007]), has attracted much attention in modeling of such complex systems.
Local modeling is a modeling framework that is based on combining a number of local models, each of which has simpler structure and has a range of validity less than the full range of operation. The range in which each local model is valid is called the operating regime. It is necessary to select a number of operating regimes, which completely cover the full range of operating range of the system based on suitably chosen variables to characterize the operating conditions of the system. Then, for each local operating, an adequate local model is found and then corresponding local model validity function is specified, which indicates the validity of the local model for each local operating regime at the specified operating condition. A global model is constructed by combining local models with an interpolation technique based on the local model validity function.
Since the local models and local model validity functions are closely related to the selection of local operating regimes, quality of the global model is highly dependent on the selection of local operating regimes. This paper is concerned with regime selection in local modeling, and propose new hybrid automatic regime selection algorithms, based on the observed input and output data with three criteria, Kullback Discrimination Information (KDI), Akaike Information Criterion (AIC) and Mean Square Error (MSE), for integration and partition of regimes to build up suitable local regimes. They are modifications of previous algorithms (Uosaki *et al* [2002]) and efficiently find the better regime selections compared to the previous ones and prevent necessary increase of local regimes, which occurs in LOLIMOT (local linear model tree) (Nelles [1997], Nelles [2001]), by introducing parsimony principles.
Remainder of this paper is organized as follows. After introduction of local modeling approach in section 2, the local regime selection procedure is proposed with brief review of three criteria of KDI, AIC and MSE in section 3. Section 4 presents results of some numerical examples. Finally, a conclusion is presented in Section 5.

2. LOCAL MODELING

Real systems usually have complex and nonlinear structures. For such systems, any models have a limited range of operating conditions and hence do not provide sufficient accuracy or performance over the full range of operations $\mathcal{R}$. Hence, a number of local models, each of which has simpler structure but serves well in a region less than the full range of operating region, are developed and then a global model is constructed by combining the local models with an interpolation technique. In order to develop local models, first, the system's full range of operation is decomposed into a number of operating regimes where simple local models can be applied. In this approach, suitable choice of operating regimes is a key issue for building up a good global model.
Consider nonlinear dynamical systems expressed by the following nonlinear autoregressive models with exogenous input (NARX model):

⋆ This work is partially supported by the Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture (C)(2)14550447, and FUT Research Promotion Fund.

$$y(t) = f(y(t-1), \cdots, y(t-n_y), \\ u(t-1), \cdots, u(t-n_u)) + e(t), \\ = f(\boldsymbol{\phi}(t-1)) + e(t) \quad (1)$$
$$\boldsymbol{\phi}(t-1) = (y(t-1), \cdots, y(t-n_y), \\ u(t-1), \cdots, u(t-n_u))^T$$

Here $y(t)$, $u(t)$, and $e(t)$ are output, input, and noise, respectively, and $\boldsymbol{\phi}(t-1)$ is called the information vector. Here the orders n_y, n_u are assumed to be known.

The total operating regime $\mathcal{R}$ is decomposed into a set of disjoint operating regimes $\{\mathcal{R}_i\}$ such that

$$\mathcal{R} = \cup_{i-1}^{n_r} \mathcal{R}_i, \\ \mathcal{R}_i \cap \mathcal{R}_j = 0 \quad \text{(empty)} \qquad i \neq j \quad (2)$$

For each operating regime $\mathcal{R}_i$, a local model

$$y(t) = \hat{f}_i(\boldsymbol{\phi}(t-1)) + e(t) \qquad i = 1, \ldots, n_r \quad (3)$$

is available, and the different local model is sufficiently valid under different operating conditions. Thus, there may be several local models M_i which are valid under some operating conditions, while no local models are valid under other conditions. The relative validity function $\tilde{\rho}_i(\boldsymbol{\phi}) \in [0, 1]$ indicates the validity of each local model at the operating condition $\boldsymbol{\phi}$. The local model M_i is accurate for the operating condition $\boldsymbol{\phi}$ when $\tilde{\rho}_i(\boldsymbol{\phi})$ is close to one, while local model M_j is in accurate if $\rho_i(\boldsymbol{\phi})$ is close to zero. And then, these local models with the relative validity functions are combined to fit for the full range of operating region as follows (Fig.1):

$$y(t) = \hat{f}(\boldsymbol{\phi}(t-1)) + e(t), \\ \hat{f}(\boldsymbol{\phi}) = \sum_{i=1}^{n_r} \hat{f}_i(\boldsymbol{\phi}) w_i(\boldsymbol{\phi}), \quad (4) \\ w_i(\boldsymbol{\phi}) = \frac{\tilde{\rho}_i(d)}{\sum_{j=1}^{n_r} \tilde{\rho}_j(d)}$$

where $\boldsymbol{\phi}$ and d indicate the current operating condition and the distance between the current operating condition $\boldsymbol{\phi}$ and the operating condition $\boldsymbol{c}_i$ that fits most for specified local model M_i, respectively, i.e.,

$$d = \|\boldsymbol{\phi} - \boldsymbol{c}_i\|, \qquad i = 1, \ldots, n_r \\ \boldsymbol{c}_i = \arg\max_{\boldsymbol{\phi}} \tilde{\rho}_i(\boldsymbol{\phi}) \quad (5)$$

For some cases, Gaussian functions are employed as the validity functions.

$$\tilde{\rho}_i(\boldsymbol{\phi}) = \exp(-d^2(\boldsymbol{\phi}, \boldsymbol{c}_i, \sigma_i)/2), \\ d(\boldsymbol{\phi}, \boldsymbol{c}_i, \sigma_i) = \sqrt{(\boldsymbol{\phi} - \boldsymbol{c}_i)^T \sigma_i^{-2} (\boldsymbol{\phi} - \boldsymbol{c}_i)}, \quad (6)$$

where σ_i is introduced to weight the influence of possible different covariances. Since the global model (4) are weighted combination of local models that are, of course, depend on the choice of local regimes $\mathcal{R}_i$, the regime selection will highly affect on the modeling performance. Thus, suitable choice of local regimes should be considered.

3. LOCAL REGIME SELECTION PROCEDURE

Here, automatic regime selection algorithms for suitable local modeling are proposed. The algorithms are based on three criteria, Kullback Discrimination Information (KDI), Akaike Information Criterion (AIC) and Mean Square Error (MSE), for integration and partition of regimes to build up suitable local regimes.

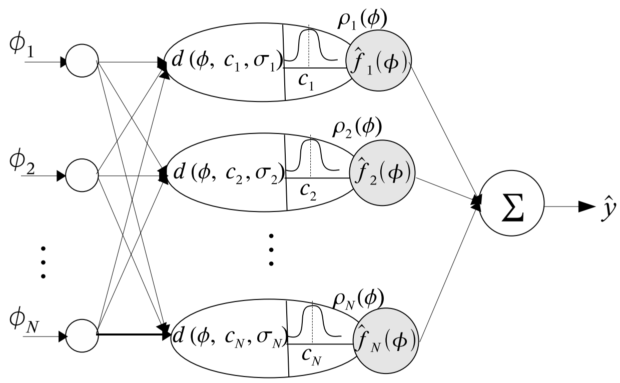


Fig. 1. Weighted combination of local models

3.1 Kullback Discrimination Information (KDI)

Kullback Discrimination Information (KDI) is well-known information criterion for model discrimination. It is a measure for discriminating in favor of the model M_1 over the model M_2, and is defined by

$$I_t[1:2;\ y^t] = \int p_1(y^t|u^{t-1}) \log \frac{p_1(y^t|u^{t-1})}{p_2(y^t|u^{t-1})} dy^t \quad (7)$$

where $p_j(y^t|u^{t-1})$ is the probability density function of $y^t = (y(t), y(t-1), \ldots, y(1))^T$ given $u^{t-1} = (u(t-1), u(t-2), \ldots u(1)^T$ under the model M_j $(j = 1, 2)$, respectively. Since KDI is non-negative and equals to zero if and only the models are identical, it can be employed as the index of distance between models M_1 and M_2.

Consider the following two stable autoregressive models with exogenous input (ARX) models.

$$M_1: \ A_1(q)y(t) = B_1(q)u(t) + e^{(1)}(t) \\ M_2: \ A_2(q)y(t) = B_2(q)u(t) + e^{(2)}(t) \quad (8)$$

where $A_j(q), B_j(q)$ are

$$A_j(q) = 1 + \sum_{t=1}^{n_j} a_t^{(j)} q^{-t}, \\ B_j(q) = \sum_{t=1}^{m_j} b_t^{(j)} q^{-t}, \qquad (j = 1, 2) \quad (9)$$

$e(t)^{(j)}$ is independently normally distributed with mean zero and variance σ_j^2, and q^{-1} is delay operator. Since the noise distribution $p(e(t))$ is assumed to be normal as above, the conditional probability distribution $p(y(t)|u(t-1))$ can be computed easily and KDI for these models is constructed as follows (Hatanaka and Uosaki [1999]).

$$I_t[1:2;\ y^t] = -\frac{1}{2}\Bigg(t + \log\frac{|\Sigma^{(1)}|}{|\Sigma^{(2)}|} \\ -\left(\mu^{(1)t} - \mu^{(2)t}\right)\left(\Sigma^{(2)}\right)^{-1}\left(\mu^{(1)t} - \mu^{(2)t}\right) \\ -\text{trace}\Big(\left(\Sigma^{(2)}\right)^{-1}\Sigma^{(1)}\Big)\Bigg) \quad (10)$$

where

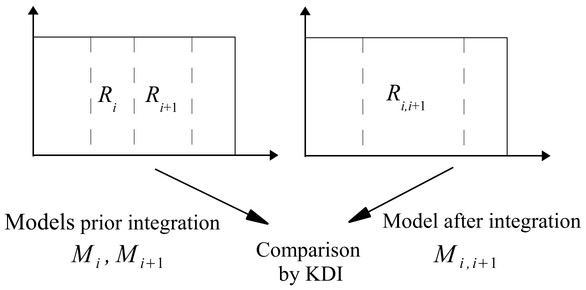


Fig. 2. Models before and after integration

$$\mu^{(j)t} = E_j[y^t|u^{t-1}] = \int_{-\infty}^{\infty} y(t)p_j(y^t|u^{t-1})dy(t)$$
$$\Sigma^{(j)} = E_j[(y^t - \mu^{(j)t})(y^t - \mu^{(j)t})^T]$$

with conditional probability density function $p_j\ (y(t)\ |u^{t-1})$ of $y(t)$ given $u^{t-1} = (u(t-1), \ldots, u(1))\ (j = 1, 2)$. For two local regimes, $\mathcal{R}_i$ and $\mathcal{R}_{i+1}\ \ (i, i+1 \in [1, \cdots, n_r])$, which are adjacent each other, it is examined whether it is better to integrate these two local regimes $\mathcal{R}_i,\ \mathcal{R}_{i+1}$ into single regime $\mathcal{R}_{i,i+1} = \mathcal{R}_i \cup \mathcal{R}_{i+1}$, or not. The KDI for discriminating in favor the local models M_i & M_{i+1} for the local regimes prior to integration over the model $M_{i,i+1}$ for the regime $\mathcal{R}_{i,i+1} = \mathcal{R}_i \cup \mathcal{R}_{i+1}$ after regime integration is calculated, where

$$\begin{array}{ll} M_i \ \&\ M_{i+1}: & \\ A_i(q)y(t) = B_i(q)u(t) + e(t), & \boldsymbol{\phi} \in \mathcal{R}_i \\ A_{i+1}(q)y(t) = B_{i+1}(q)u(t) + e(t), & \boldsymbol{\phi} \in \mathcal{R}_{i+1} \end{array} \quad (11)$$

and

$$M_{i,i+1}:\ A_{i,i+1}(q)y(t) = B_{i,i+1}(q)u(t) + e(t)$$
$$\boldsymbol{\phi} \in \mathcal{R}_{i,i+1} \quad (12)$$

If the KDI is small, the distance between models M_i & M_{i+1} and $M_{i,i+1}$ is small. It indicates the possibility to integrate the adjacent regimes $\mathcal{R}_i$ and $\mathcal{R}_{i+1}$ into single regime $\mathcal{R}_{i,i+1}$.

3.2 Akaike Information Criterion (AIC)

When a global model is built up by using larger number of local models, the fitting error becomes smaller. But, in some cases, the phenomenon of ‘over-fit” is occurs; additional (unnecessary) increase of local regimes adjust themselves to particular features of the particular realization of noise realization, and the models obtained do not work for different possible operating conditions. Hence idea of ‘parsimony principle’ is introduced. It says that among the models which explain the data well, the model with the smallest number of independent parameters should be chosen. This indicates that the number of local models, or the number of local regimes should not be increased so much. One of the ideas to realize this parsimony principle is introduction of a penalty for model complexity. Akaike Information Criterion is an example. It is defined by

$$\begin{aligned}\text{AIC} = & -2\log(\text{maximum likelihood}) \\ & +2(\text{number of parameters})\end{aligned} \quad (13)$$

For ARX models, AIC is given by

$$\text{AIC} = N \log V + 2(n + m) \quad (14)$$

where N is number of data, n and m are number of parameters $\boldsymbol{\theta} = [a_i, b_j]$ in ARX models, respectively, and V is the Mean Square Error (MSE) of the identified ARX models,

$$V(\boldsymbol{\theta}) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2(t, \hat{\boldsymbol{\theta}})$$
$$\varepsilon(t, \hat{\boldsymbol{\theta}}) = y(t) - \hat{y}(t, \hat{\boldsymbol{\theta}}) \quad (15)$$

where $\hat{\boldsymbol{\theta}}$ is the minimum mean square error estimate of the ARX parameters $\boldsymbol{\theta} = (a_1, \cdots, a_{n_j}, b_1, \cdots, b_{n_u})^T$ and $\hat{y}(t, \hat{\boldsymbol{\theta}})$ is the prediction of $y(t)$ based on the estimates $\hat{\boldsymbol{\theta}}$. Hence, the best choice of orders is

$$(\hat{n}, \hat{m}) = \arg\min_{n,m} \text{AIC} \quad (16)$$

In local modeling, the best choice of number of local regimes is given by the number of local regimes minimizing the value of AIC, since the number of parameters is proportional to the number of local regimes.

3.3 Regime Selection

Both regime integration and regime partition process are applied to find good local regimes. The criteria MSE and AIC are used for regime integration and regime partition processes, and AIC is used for stopping the whole selection process (integration and partition).

Regime integration process When the KDI for discriminating in favor the local models M_i & M_{i+1} for the local regimes prior to integration in $\mathcal{R}_i$ & $\mathcal{R}_{i+1}$ over the model $M_{i,i+1}$ for the regime $\mathcal{R}_{i,i+1} = \mathcal{R}_i \cup \mathcal{R}_{i+1}$ after regime integration is small, it is likely that the distance between models M_i & M_{i+1} and $M_{i,i+1}$ is small and integration of the adjacent regimes $\mathcal{R}_i$ and $\mathcal{R}_{i+1}$ into single regime $\mathcal{R}_{i,i+1}$ is possible. Hence, the regimes $\mathcal{R}_j$ and $\mathcal{R}_{j+1}$, which give the minimum of KDI among the KDI's for all the combination of adjacent local regime $\mathcal{R}_i$ and $\mathcal{R}_{i+1}$ and their integration $\mathcal{R}_{i,i+1}$, will be integrated into single regime $\mathcal{R}_{i,i+1}$. Then local model should be reconstructed for the regime $\mathcal{R}_{j,j+1}$.

Regime partition process For each local model, the observations and their estimates based on the local model are compared. If the discrepancy measured by MSE is large, the fitness is insufficient. It may come from that the regime is too large to fit the local model. Hence the local regime with the worst fitness or the largest MSE, will be divided into two equi-partitioned local regimes and the local models for the partitioned regimes are re-constructed.

Two algorithms, which conduct regime integration process only and regime partition process only, respectively, are developed in (Uosaki *et al.* 2002) and their usefulness is shown in local modeling. Here, new two hybrid algorithms, Series Algorithm and Parallel Algorithm are proposed by combining these regime integration and regime partition processes.

(a) *Series Algorithm*
- **Step 1 :** Identify a model assuming the system is composed by single regime.
- **Step 2 :** Calculation of AIC.
- **Step 3 :** Execution of regime partition process based on MSE as above and re-calculation of AIC.
- **Step 4 :** Comparison of AIC's of Step 2 and Step 3. If AIC of Step 2 is smaller, stop the regime partition with the model prior to partition process and go to Step 5.

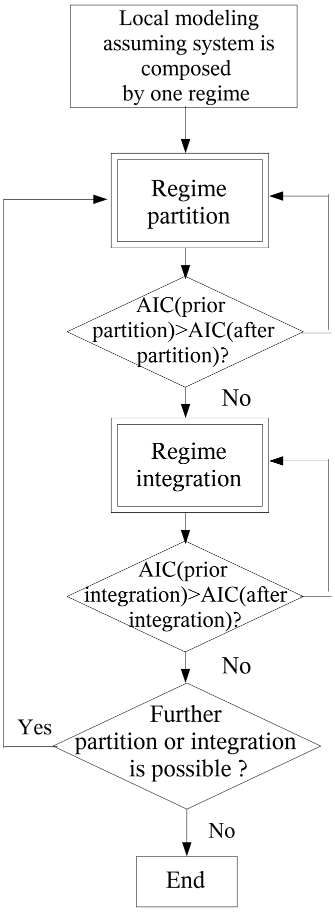


Fig. 3. Series Algorithm

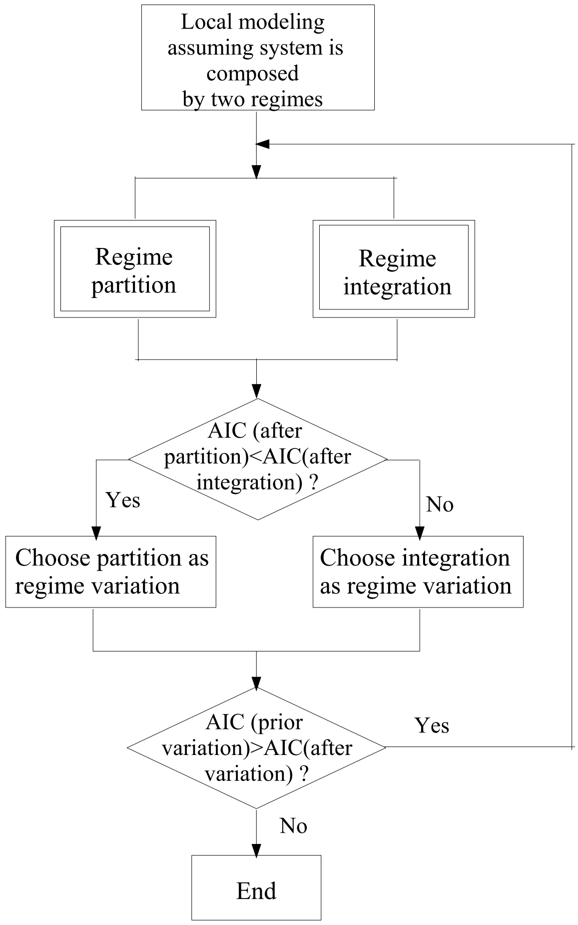


Fig. 4. Parallel Algorithms

Otherwise, renew the model with after partition, and go back to Step 2.

Step 5 : Calculation of AIC.

Step 6 : Execution of regime integration process based on KDI as above and re-calculation of AIC.

Step 7 : Comparison of AIC's of Step 5 and Step 6. If AIC of Step 5 is smaller, stop the regime integration with the model prior to integration and go to Step 8. Otherwise, renew the model with after partition, and go back to Step 5.

Step 8 : Repeat Steps 2 through 7 while AIC is decreasing.

(b) *Parallel Algorithm*

Step 1 : Identify a model assuming the system is composed by two regimes with equivalent volume.

Step 2 : Calculation of AIC.

Step 3 : Execution of regime partition process based on MSE and regime integration process based on KDI as above and re-calculation of AIC's for each process.

Step 4 : Comparison of AIC's after regime partition and regime integration and choose the process with smaller AIC.

Step 5 : Repeat Steps 2 through 4 while the smaller AIC in Step 4 is decreasing.

4. NUMERICAL SIMULATION STUDIES

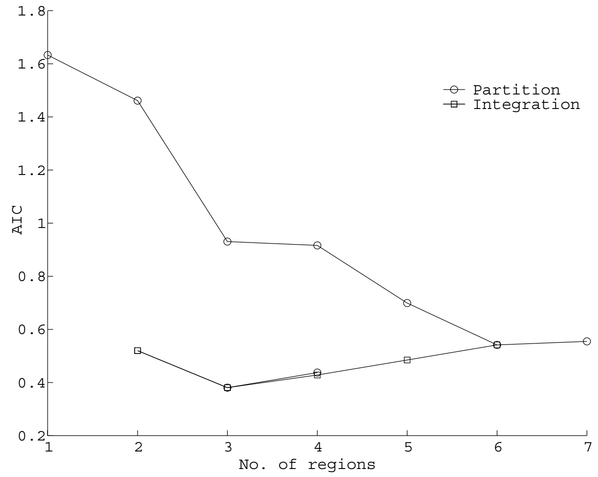
Numerical simulation studies have been carried out to examine the applicability of the proposed regime selection algorithms. Consider the following nonlinear time series model.

$$x(t+1) = \begin{cases} 0.5x(t) + 1.0u(t) + 0.6u(t-1) \\ \quad u(t) \geq 0.7 \\ -1.2x(t) - 0.9u(t) + 1.8u(t-1) \\ \quad 0.2 \leq u(t) < 0.7 \\ -0.5x(t) + 1.8u(t) + 1.3u(t-1) \\ \quad u(t) < 0.2 \end{cases}$$
$$x(0) = 0,$$
$$y(t) = x(t) + e(t)$$
$$u(t) : \text{random number distributed uniformly in [0,1]} \tag{17}$$

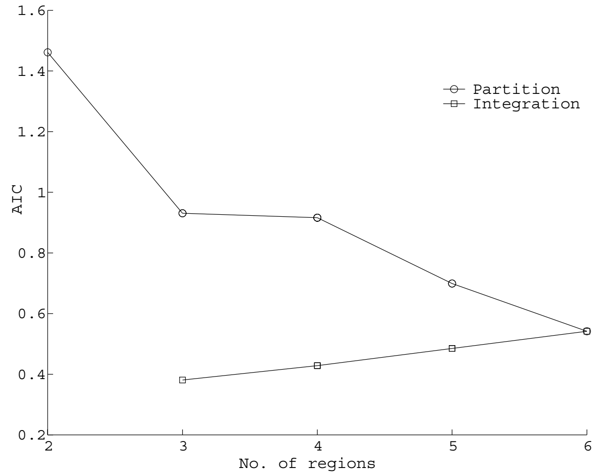
where observation noises $e(t)$ are white Gaussian with mean 0 and variance 0.1. Number of observations is $N = 200$, and they are divided into two parts; the first half is used for the regime selection and the latter half for validation. The following ARX model is employed here as a local model.

$$\hat{y}_i(t+1) = a_1^{(i)}\hat{y}(t) + b_1^{(i)}u(t) + b_2^{(i)}u(t-1), \quad i = 1, \cdots, n_r$$

The following final identification result, which is same by both Series and Parallel Algorithms, is obtained after 8 steps of regime partition and integration processes:



(a) Series Algorithm



(b) Parallel Algorithm

Fig. 5. Variation of AIC

$$\hat{y}(t+1) = \begin{cases} 0.4968\hat{y}(t) + 0.9672u(t) \\ \quad +0.5717u(t-1) \quad u(t) \geq 0.6815 \\ -1.2123\hat{y}(t) - 0.9114u(t) \\ \quad +1.8238u(t-1) \\ \quad\quad 0.2478 \leq u(t) < 0.6815 \\ -0.6289\hat{y}(t) - 0.5738u(t) \\ \quad +1.6242u(t-1) \quad u(t) < 0.2478 \end{cases}$$

Here, corresponding variations of AIC are shown in Fig.5(a) for Series Algorithm, and Fig.5(b) for Parallel Algorithm, respectively. By comparison of the observations and the estimates by the proposed algorithm as shown in Fig.6, the good performance of the identified model can be found. It should be noted that, regime selection process, which employs Regime Partition Algorithm (Uosaki *et al.* 2002) with partition process only, stops at the stage with 6 sub-regions corresponding to AIC=0.5416 and never finds the optimal result corresponding to AIC=0.3808. This and other examples not shown here (Manabe [2001]) indicate the applicability of the proposed algorithms.

5. CONCLUSIONS

This paper has considered automatic selection algorithm of suitable operating regime in local modeling. Based on three criteria, Kullback Discrimination Information (KDI), Akaike Information Criterion (AIC), and Mean Square Error (MSE), new two regime selection algorithms, Series Algorithm and Parallel Algorithm, are proposed by combining regime integration and partition processes. Though their better performance in local modeling has been shown here, the algorithms do not assure the best solution analytically, and theoretical analysis should be pursued.

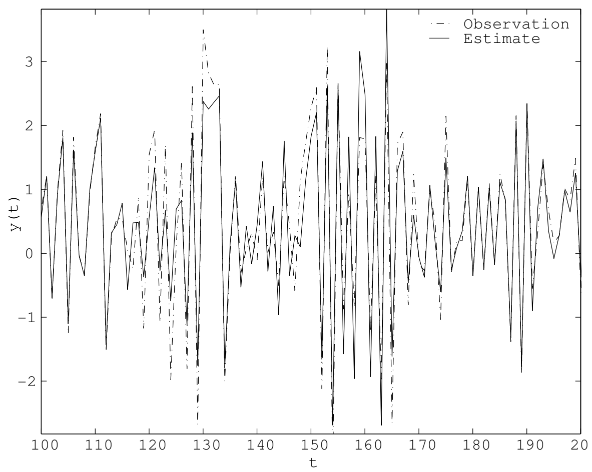


Fig. 6. Modeling result (Number of regimes is 3)

REFERENCES

T. Hatanaka, and K. Uosaki. Optimal auxiliary input design for fault diagnosis. *Proc. 14th IFAC World Congress*, vol. H, pages 91–96, 1999.

T. A. Johansen, and B. A. Foss. Local modeling as a tool for semi-empirical and semi-mechanistic process. In A. B. Bulsari, editor, *Neural Networks for Chemical Engineers*, Elsevier, Amsterdam, pages 297-334, 1995.

T. A. Johansen, and B. A. Foss. Operating regime based process modeling and identification. *Computers and Chemical Engineering* 21: pages 159–176, 1997.

A. L. Juloski, S. Paoletti, and J. Roll, editors. *Recent Techniques for the Identification of Piecewise Affine and Hybrid Systems*, Birkhäuser, Boston, 2006.

N. Manabe. *Regime Selection in Local Modeling*. MS. Thesis, Department of Information and Knowledge Engineering, Tottori University 2001. (*in Japanese*)

R. Murray-Smith, and T. A. Johansen, editors. *Multiple Model Approaches to Modeling and Control*. Taylor & Francis, 1997.

O. Nelles. Orthogonal basis functions for nonlinear system identification with local linear model trees (LOLOMOT), *Proc. IFAC Symposium on System Identification*, pages 667–672, 1997.

O. Nelles. *Nonlinear System Identification*. Springer-Verlag, 2001.

S. Padletti, A. L. Juloski, G. Ferrari-Trecate, and R. Vidal. Identification of hybrid systems: A tutorial. *European J. of Control*, 13: 242-260, 2007.

K. M. Pekpe, G. Mourot, K. Gasso, and J. Ragot. Identification of switching systems using change detection technique in the subspace framework. *Proc. 43rd IEEE Conference on Decision and Control*, pages 3720–3725, 2004.

J. Roll. *Local and Piecewise Affine Approaches to System Identification*. PhD. dissertation, Department of Electrical Engineering, Linköping University 2003.

K. Uosaki, T. Hatanaka, and N. Manabe. Automatic operating regime selection in local modeling. *Proc. 15th IFAC World Congress*, T-We-M21, CD-ROM, 2002.