15th Triennial World Congress, Barcelona, Spain

AUTOMATIC OPERATING REGIME SELECTION IN LOCAL MODELING

Katsuji Uosaki [*,1] **Toshiharu Hatanaka** [*] **Norio Manabe** [*,2]

[*] *Department of Information and Physical Sciences*
Graduate School of Information Science and Technology
Osaka University
Suita, Osaka 565–0871, Japan

Abstract: Recently, local modeling has attracted 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), we develop regime integration and partition algorithms. Numerical simulation studies illustrate the applicability of the proposed selection algorithms.

Keywords: System identification, Modeling, Local structure, Complex systems, Information analysis

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 (Eds.), 1997) has been received much attention in modeling of such complex systems. It 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. We 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, we find an adequate local model and find a local model validity function, 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

[1] Partially supported by the Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (10650432, 10045043).

[2] Now with Developing Center for Mobile Computing and Communications, Toshiba Corp. Hino 191-8555, Japan

dependent on the selection of local operating regimes. This paper is concerned with regime selection in local modeling, and propose 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 Squared Error (MSE), for integration and partition of regimes to build up suitable local regimes. Though similar partitioning idea, LOLIMOT (local linear model tree), was proposed independently by Nelles (Nelles, 1997),(Nelles, 2001), it is basically an incremental algorithm and may provide unnecessary complex models. The proposed partitioning algorithm in this paper prevents unnecessary increase of local regimes by introducing the parsimony principle.

Remainder of this paper is organized as follows. After introduction of local modeling approach in section 2, we describe 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, we develop 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, and then construct a global model by combining the local models with an interpolation technique. In order to develop local models, we, first, decompose the system's full range of operation into a number of operating regimes where a simple local model 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):

$$\begin{aligned} y(t) &= f(y(t-1), \cdots, y(t-n_y), \\ &\qquad u(t-1), \cdots, u(t-n_u)) + e(t) \\ &= f(\boldsymbol{\phi}(t-1)) + e(t) \qquad (1) \\ \boldsymbol{\phi}(t-1) &= (y(t-1), \cdots, y(t-n_y), \\ &\qquad u(t-1), \cdots, u(t-n_u))^T \end{aligned}$$

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

We decompose the total operating regime $\mathcal{R}$ into a set of disjoint operating regimes $\{\mathcal{R}_i\}$ such that

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

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 functions $\tilde{\rho}_i(\boldsymbol{\phi}) \in [0, 1]$ indicate 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, we combine these local models with the relative validity functions to fit for the full range of operating region as follows (Fig.1):

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

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.,

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

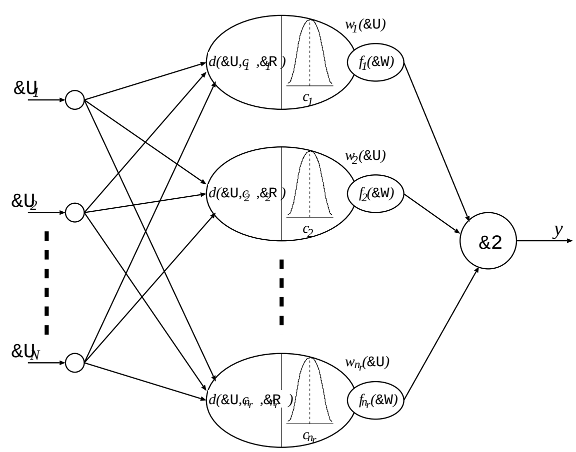

Fig. 1. Weighted combination of local models

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

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

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

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

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

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

$\mathrm{e}(t)^{(j)}$ is independently normally distributed with mean zero and variance σ_j^2, and q^{-1} is delay operator. Assuming that the noise distribution p(e(t)) is normal as above, we can easily compute the conditional probability distribution $p(y(t)|u(t-1))$ and obtain KDI for these models as follows (Hatanaka and Uosaki, 1999).

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

where

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

with conditional probability density function $p_j\,(y(t)\,|u^{t-1})$ of $y(t)$ given $u^{t-1} = (u(t-1), \ldots, u(1))$ $(j = 1, 2)$. We consider two local regimes, $\mathcal{R}_i$ and $\mathcal{R}_{i+1}$ $(i, i+1 \in [1, \cdots, n_r])$ which are adjacent each other, and examine 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. We calculate the KDI for discriminating in favor the local models M_i & M_{i+1} for the local regimes prior to integration, where

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

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, where

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

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 $\mathcal{R}_{i,i+1}$.

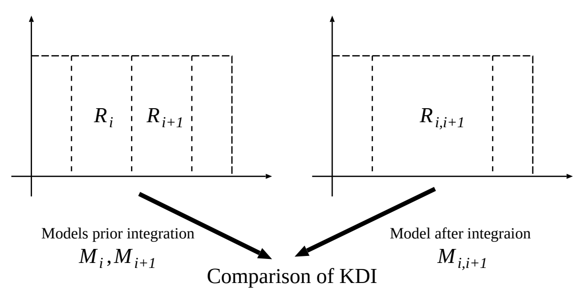

Fig. 2. Models before and after integration

3.2 *Akaike Information Criterion (AIC)*

When we increase the number of local regimes and build up a global model 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

$$\mathrm{AIC} = -2\log(\text{maximum likelihood}) \\ +2(\text{number of parameters}) \tag{13}$$

For ARX models, AIC is given by

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

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,

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

where $\hat{\boldsymbol{\theta}}$ is the 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

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

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*

We apply both regime integration and regime partition process 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).

(i) 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 re-constructed for the regime $\mathcal{R}_{j,j+1}$.

(ii) 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 we will divide the local regime with the worst fitness, which gives the largest MSE, into two equi-partitioned local regimes and the local models for the partitioned regimes are re-constructed.

Two algorithms are considered, which conducts regime integration process only and regime partition process only, respectively.

(a) *Regime Integration Algorithm* (Regime integration process only)
  - **Step 1:** Identification of local models for $m \times n$-divided local regimes.
  - **Step 2:** Calculation of AIC.
  - **Step 3:** Execution of regime integration process as above.
  - **Step 4:** Re-calculation of AIC after regime integration process.
  - **Step 5:** Comparison of AIC's of Step 2 and Step 4. If AIC of Step 2 is smaller, stop the regime selection with the model prior to integration process. Otherwise, renew the model with after integration process, and go back to Step 2.

(b) *Regime Partition Algorithm* (Regime partition process only)
  - **Step 1:** Identification of a local model for the whole operating regime.
  - **Step 2:** Calculation of AIC.
  - **Step 3:** Execution of regime partition process as above.
  - **Step 4:** Re-calculation of AIC after regime partition process.
  - **Step 5:** Comparison of AIC's of Step 2 and Step 4. If AIC of Step 2 is smaller, stop the regime selection with the model after partition process. Otherwise, renew the model with after partition process, and go back to Step 2.

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.

$$x(t+1) = \begin{cases} 0.5x(t) + 1.0u(t) + 0.6u(t-1) \\ \qquad\qquad\qquad u(t) \geq 0.5 \\ -0.5x(t) + 1.8u(t) + 0.3u(t-1) \\ \qquad\qquad\qquad u(t) < 0.5 \end{cases}$$

$x(0) = 0$

$y(t) = x(t) + e(t)$

$u(t)$: random number distributed uniformly in $[0, 1]$

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 operating points $\phi(t)$ are assumed to be a 2-dimensional variable $(u(t), u(t-1))$. In the *Regime Integration Algorithm*, the whole operating regime are first partitioned with $n_r = 16$ (4×4) rectangular regimes by $\phi(t)$. While, in the *Regime Partition Algorithm*, the regime is equi-partitioned into two regime with smaller MSE (Fig.3).

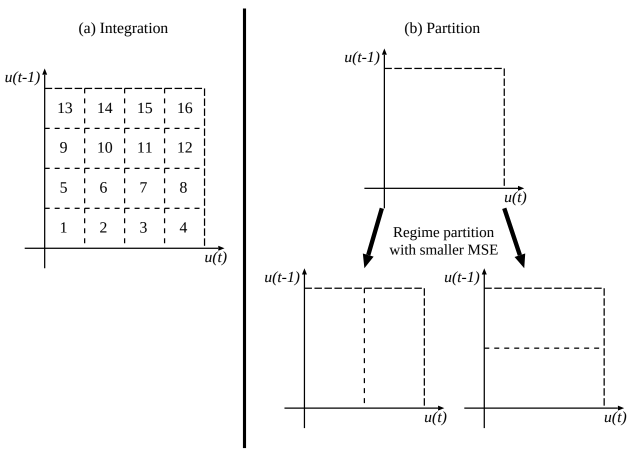

Fig. 3. Regime setting in two dimension

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 regime selection processes by using the integration and partition algorithms are shown in Figs.4 and 5, respectively. The same final result is obtained by both algorithms, and is given by

$$\hat{y}(t+1) = \begin{cases} 0.4947\hat{y}(t) + 1.0078u(t) + 0.5896u(t-1) \\ \qquad\qquad\qquad u(t) \geq 0.5 \\ -0.4750\hat{y}(t) + 1.8473u(t) + 0.2641u(t-1) \\ \qquad\qquad\qquad u(t) < 0.5 \end{cases}$$

By comparison of the observations and the estimates by the proposed algorithm as shown in Fig.6, we can find the good performance of the identified model. This and other examples not shown here (Manabe, 2001) indicate the applicability of the proposed algorithms.

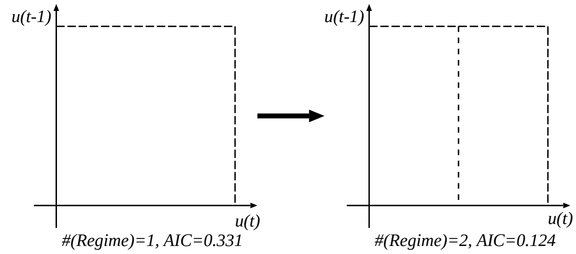

Fig. 5. Regime variation by dividing algorithm

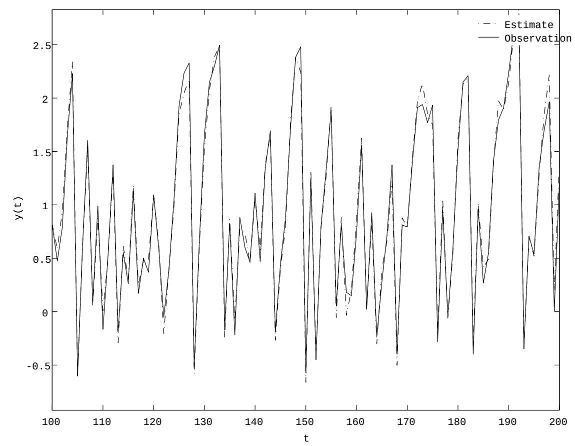

Fig. 6. Modeling result (Number of regimes is 2)

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), we have developed regime integration and partition algorithms and showed their possible applicability in local modeling. By combining these two algorithms, better regime selection will be expected.

6. REFERENCES

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

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

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

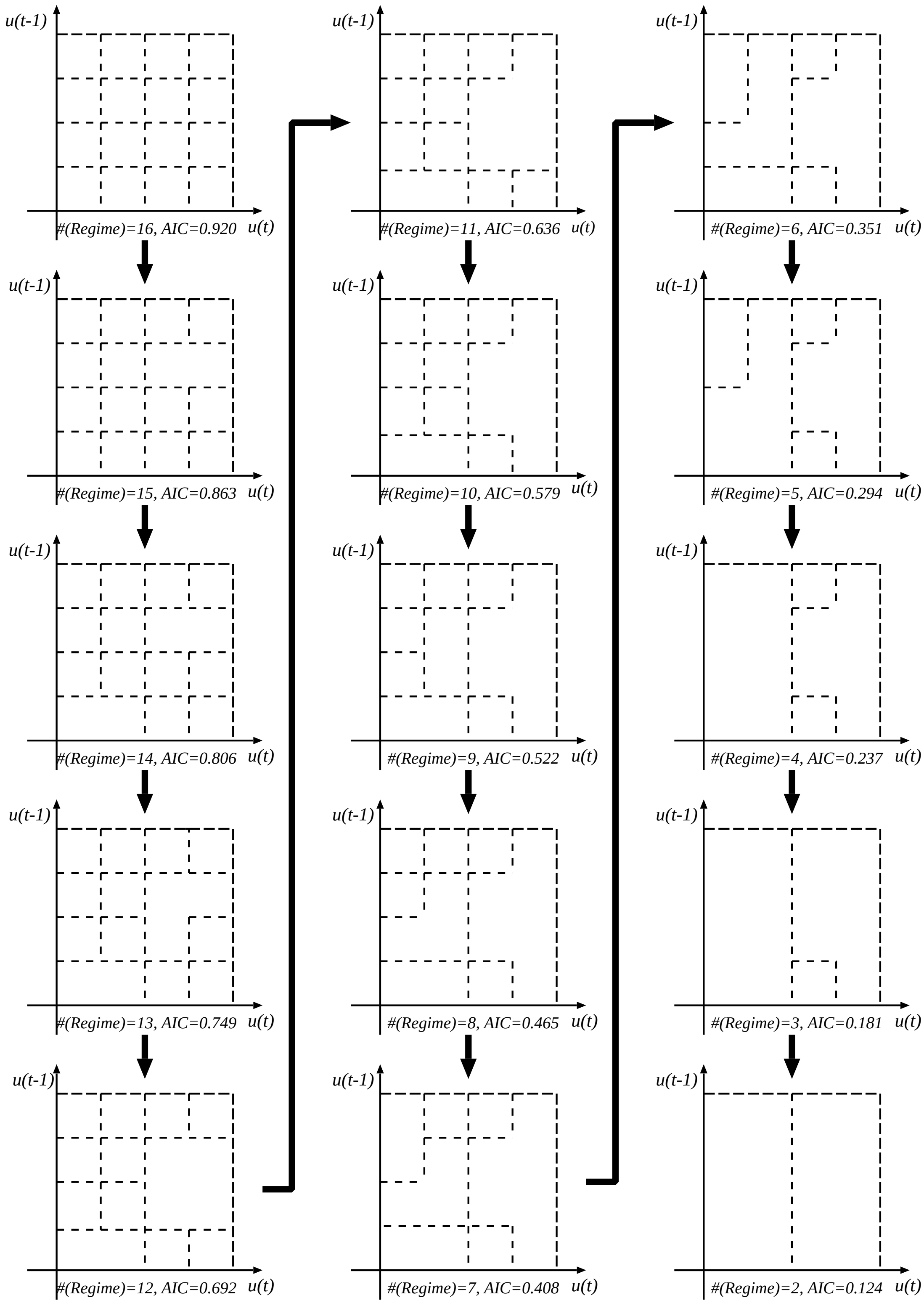

Fig. 4. Regime variation by integration algorithm

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

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

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

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