A REAL-TIME MULTIPLE-MODEL BASED CONTROL AND IDENTIFICATION
OF A NONLINEAR PROCESS

A. Aminzadeh  A.A. Safavi  A. Khayatian

Department of Electrical Engineering, School of Engineering
Shiraz University, Shiraz, Iran
safavi@shirazu.ac.ir

Keywords: Real Time Control, Multiple Model, Operating Regime, Validity Function, ARMAX model, Data Acquisition.

Abstract

This paper presents implementations of Real-Time Control and Identification based on Multiple-Modeling approach for an experimental thermal process. The thermal process is a nonlinear plant; therefore, based on variations of the input and disturbance, four local operating regimes are defined. Then linear local ARMAX models are identified and integrated into a NARMAX model by combining local models via proper validity and interpolation functions. Results of a single model and multiple model approach show superior performance of multiple modeling technique which is also more flexible. Moreover, the Real-Time control of the plant with four locally designed controllers is addressed. The platform used for the Real-Time implementation is Matlab/Simulink/Real-Time-Workshop with Visual C++ and Watcom compilers using a DAQ interface. The Real-Time application of the global controller based on multiple model approach proves excellent performance of this design in comparison to a single PID controller.

1 Introduction

Usually it is extremely difficult to identify a model that accurately matches a nonlinear plant in all operating regimes [1],[3],[5],[6],[7],[9],[12],[14],[16],[25],[26]. Even if such a model can be identified, controller design may be also difficult. Therefore it is quite attractive to use an alternative approach wherein local “Multiple Models” are identified at the different regions of operation and controller design is carried out based on these models. This allows to invoke simple linear models to represent a nonlinear system and then design systematic controllers. When we find different local models, each local model has a “Relative Validity” in its operating regime [12]. There are many advantages to use multiple models; such as their flexibility in selecting the modeling methods (e.g. transfer function and state space), different presentations (e.g. continuous time and discrete time) [12],[14] and other cases such as noise and disturbance reduction. Moreover, this method can be used to apply online control signals to the systems based on this modeling approach with high speed and accuracy [9],[25],[26].

Multiple Modeling control has been a research tool in various applications. One conventional method that guarantees the stability of a global system and includes some local subsystems or controllers is addressed by Fernandez-Anaya and Escandon-Alcazar [7] and is called “Simultaneous Stabilization”. Necessary and sufficient conditions for control stability are presented and the method is used for three SISO plants to m SISO plants. Johansen and Foss [14] show an empirical modeling of a heat transfer process using local models and interpolation. Narendra, Balakrishnan and Ciliz [25] introduced a general methodology for the improvement of performance of dynamical system operating in rapidly varying environments. Both linear and nonlinear plants are considered and an indirect approach based on multiple models is used for control. Also the article has shown a general architecture for indirect adaptive control using neural networks (NN) and an architecture of the control system for robotic manipulators with N models and controllers. Gregorcic and Lightbody [9] compared pole-placement self-tuning control with the multiple model approach for the control of highly nonlinear process. A nonlinear “Continuous Stirred Tank Reactor” (CSTR) process is used to highlight some of the difficulties associated with self-tuning control. Doya et al [6] introduced a modular reinforcement learning architecture for non-linear, non-stationary control tasks which is called “Multiple Model-based Reinforcement Learning” (MMRL). There are two other methods used for complex system with several operation behaviors: The first method is introduced by Aarhus [1] called “Partial Least Squares regression” (PLS) models that usually is used for nonlinear empirical modeling. The second one is addressed by Angelis [3] named “Polytopic Linear Models” (PLM) which is used for modeling, control and identification. One of the latest application of the multiple modeling approach in radar and communication is introduced by Bar-Shalom and Dale Blair [5] which is called “Interacting Multiple Model” (IMM) estimator and provide superior tracking performance compared to maneuver detection schemes. Each of the above methods has some advantages and special complexity. Besides, none of them, except [25], has applied the method for Real-Time applications. In this paper, we try not only to develop the multiple model based method in a simple way, but also its Real-Time implementation is also considered. The system used in this paper is an experimental “Heating Plant” with an “air tube” which contains a “heating element” as input, “temperature sensor” as output and an “air damper” as a disturbance [12],[14],[26]. It is desirable to control the output temperature of this system. The main reason for using the multiple modeling approach is the special nonlinear behavior of this system [10],[11].
Developing “Data Acquisition” (DAQ) software for real-time control can be a difficult task, often resulting in a software that is inflexible, hard to maintain and difficult to modify, especially if the specifications of the hardware involved change [24]. So another goal of this paper is applying the “Real-Time-Control” signals to the real process. This needs advanced methods of sending and receiving data that match the special software and hardware equipments such as (DAQ)[20],[24]. The “Thermal-Process” is connected to a computer with MATLAB/SIMULINK environment using required interfaces [2],[22],[23].

2 Problem Definition

2.1 The Thermal Process and Its DAQ Interface

Consider the experimental heating plant schematically depicted in Fig 1. The practical process consists of a tube, an air damper, a heating element and a temperature measuring device.

![Figure 1: Schematic of the Thermal-Process](image)

Air enters the tube and is warmed up by the heating element. The temperature of the air is measured by the temperature sensor and is feedback to the controllers to make a proper signal. The variables are:

- Input voltage $u(t)$ which is applied to the heater and changes by the fire angle of a BT-137 Triac.
- Fan driver $v(t)$ that is considered as a disturbance and changes by the potentiometer which controls the fan driver containing two BD-140 and 2N-3055 Transistors.
- Output temperature $y(t)$ which is measured by an LM-35 Transistor and amplified by an OP-07 Op-Amp. The measured output sensitivity is 1V/20ºC.

For implementation of identification and control algorithms, the thermal process is connected to a computer via a PCL-818HG DAQ-card of Advantech Company which is a high-gain, high-performance multifunction data acquisition card for IBM PC/XT/AT or compatible computers. It offers the five most desired measurement and control functions: 12-bit A/D conversion, D/A conversion, digital input, digital output and timer/counter [2]. The platform to implement the control and identification procedures is developed within Matlab/Simulink/Real-Time-Workshop. This can be done by defaulting Visual C++, Watcom and Java compilers to make a link with ISA slot [22],[23]. This algorithm needs to generate the codes such as “Target Module”, “Dynamic Link Library (DLL) Files”, “Intermediate Object Files”, “Batch Files”, “Data Type Transition C Files” and “Model Header Files” which is called “Build a Model”.

2.2 The Multiple Modeling Approach Based on the Process Operating Regimes

Any model will have a limited range of validity. The model restrictions may be due to the assumptions made for a mechanistic model, or by the experimental conditions under which the data was logged for an empirical model. To emphasize this, a model that has a range of validity less than the desired range of validity will be called a local model, as opposed to a global model that is valid in the full range of operation. We will be concerned with a modeling framework that is based on combining a number of local models, where it is of particular importance to describe the region in which each local model is valid. We call such a region an operating regime [17].

![Figure 2: The set of two-dimensional operating points is decomposed into four regimes. The vector $z(t) = (z_1(t); z_2(t))$ is the operating point](image)

The framework can be conceptually illustrated as in Fig. 2. The system full range of operation is completely covered by a number of possibly overlapping operating regimes. In each operating regime the system is modeled by a local model, and the local models can be combined into a global model using an interpolation technique. One motivation behind this framework is that global modeling is complicated because one will need to describe the interactions between a large number of phenomena that appear globally. Local modeling, on the other hand, may be considerably simpler because locally there may be a smaller number of phenomena that are relevant, and their interactions are simpler [12].

For some applications, one may need a model that only describes the input/output behavior of the system (i.e. the system is considered a black box). The ARMAX model representation is a well known linear input/output model representation, while the NARMAX (Nonlinear ARMAX) model representation is an extension that represents the model as a nonlinear mapping of past inputs, outputs and noise terms to future outputs [13]. The NARMAX model representation:
\[ y(t) = f(y(t-1), \ldots, y(t-n_{y}), u(t-1), \ldots, u(t-n_{u}), \\
\quad e(t-1), \ldots, e(t-n_{e}) + e(t) \]  
\[ y(t) = f(y(t-1)) + e(t) \]  
(2.1)

is able to represent the observable and controllable modes of a large class of discrete-time non-linear systems. Here \( y(t) \in Y \subset \mathbb{R}^n \) is the output vector, \( u(t) \in U \subset \mathbb{R}^m \) is the input vector, and \( e(t) \in E \subset \mathbb{R}^r \) is the noise vector. We assume \( n_{y} \), \( n_{u} \) and \( n_{e} \) are known. The problem is to construct the nonlinear function \( f : \Psi \to \mathbb{R}^n \). So we introduce the \((m(n_{y} + n_{u}) + r_{n})\)-dimensional information vector:

\[ \psi(t-1) = [y(t-1) \quad \cdots \quad y(t-n_{y}) \quad u(t-1) \quad \cdots \quad u(t-n_{u}) \\
\quad e(t-1) \quad \cdots \quad e(t-n_{e})]^{T} \]  
(2.2)

belonging to the set \( \Psi = Y^n \times U^n \times E^n \). This enables us to write (2.1) in the form:

\[ y(t) = f(\psi(t-1)) + e(t) \]  
(2.3)

Provided that necessary smoothness conditions on \( f \) are satisfied, a general way of approximating \( f \) is by series expansions. A first order Taylor-series expansion about an equilibrium point yields an ARMAX model. Second- and third-order Taylor-expansions are also possible, while higher-order Taylor-expansions are not very useful in practice because the number of parameters in the model increases rapidly with the expansion order, and because of the poor extrapolation and interpolation capabilities of higher-order polynomials. Splines offer another possible solution to this problem, where the idea is to patch together low-order polynomials. A representation that is closely related to splines in spirit, but still very different for multi-dimensional modeling problems, is based on patching together local models [12]. For the optimal combination of local models suppose \( N \) local models (indexed by \( i \in \{1,2,\ldots, N\} \))

\[ y(t) = \hat{f}_{i}(\psi(t-1)) + e(t) \]  
(2.4)

are available, and the different local models are accurate under different operating conditions. Hence, under some operating conditions there may be several local models that are accurate, while no local model may be accurate under other conditions. Suppose the relative validity (or relevance) of each local model is indicated by the weighting functions \( \tilde{\rho}_{1}, \tilde{\rho}_{2}, \ldots, \tilde{\rho}_{N} : \Psi \to [0,1] \).

If at a given \( \psi \in \Psi \) the local model indexed with \( i \) is accurate, then \( \tilde{\rho}_{i}(\psi) \) will be close to one, while \( \tilde{\rho}_{i}(\psi) \) is close to zero for all \( \psi \in \Psi \) where local model \( i \) is inaccurate. We essentially seek a global model:

\[ y(t) = \hat{f}(\psi(t-1)) + e(t) \]  
(2.5)

based on a combination of the local models (2.4). From the definition of \( \tilde{\rho}_{i} \), it is natural to require that, \( \hat{f}(\psi) \) should be close to \( \hat{f}_{i}(\psi) \) at those \( \psi \in \Psi \) where \( \tilde{\rho}_{i}(\psi) \) is large. The subset of \( \Psi \) where \( \tilde{\rho}_{i}(\psi) \) is large [12] is denoted \( \Psi_{i} \). This suggests that \( \hat{f} \) should minimize a criterion given by:

\[ M(\hat{f}) = \sum_{i=1}^{N} \int \left| \hat{f}(\psi) - \hat{f}_{i}(\psi) \right|^2 \tilde{\rho}_{i}(\psi) d\psi \]  
(2.6)

where \( \| \cdot \| \) is the Euclidean norm.

It can be shown by a theorem [3],[12] that if the function \( \hat{f}_{1}, \hat{f}_{2}, \ldots, \hat{f}_{N} \) belong to \( C^{m}(\psi) \), the set of all continuous m-dimensional functions defined on \( \Psi \) and also \( \sum_{i=1}^{N} \rho_{i}(\psi) > 0 \), for all \( \psi \in \Psi \), then the function \( \hat{f} \) defined by:

\[ \hat{f}(\psi) = \sum_{i=1}^{N} f_{i}(\psi) \tilde{\rho}_{i}(\psi) \]  
(2.7)

which

\[ \tilde{\rho}_{i}(\psi) = \frac{\hat{f}_{i}(\psi)}{\sum_{i=1}^{N} \hat{f}_{i}(\psi)} \]  
(2.8)

minimizes \( M \) on \( C^{m}(\psi) \) [3],[12],[16].

3 Modeling and Identification of the Thermal Plant

3.1 Semi-Mechanistic Model Identification

Semi-physical modeling is an application of system identification where physical insight is used to come up with suitable nonlinear transformations of the raw measurements, so as a good model structure is achieved. Semi-physical modeling is less ambitious than physical modeling in that no complete physical structure is sought, just suitable inputs and outputs that can be subjected to more or less standard model structures such as linear regressions are considered [14],[18]. This method can be applied for the thermal process by identifying the parameters and solving the following energy balance equation [10],[11]:

\[ \rho V c_p \frac{dT}{dt} = \rho c_p q v (t - \tau_{c}) (T_{o} - T(t)) + U (T_{i} - T(t)) \]

\[ + \, G v ^{2} (t - \tau_{c}) + Q \]  
(3.1)

where the parameters are defined in Table 1. This model will be discretized and the parameter identification equations are solved to reach at a semi-mechanistic NARMAX model. This requires cumbersome calculations such as RLS methods [8] to find the unknown parameters of the system especially time delays of input and disturbance (\( \tau_{c} \) and \( \tau_{a} \)) thus is not considered in this paper.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variable</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>Mass density of air</td>
<td>( g/l )</td>
</tr>
<tr>
<td>( V )</td>
<td>Volume of tube</td>
<td>( l )</td>
</tr>
<tr>
<td>( c_{p} )</td>
<td>Specific heat capacity of air</td>
<td>( J/gK )</td>
</tr>
<tr>
<td>( q )</td>
<td>Volumetric air flow rate</td>
<td>( l/s )</td>
</tr>
<tr>
<td>( G )</td>
<td>Conductance in resistor</td>
<td>( l^2 )</td>
</tr>
<tr>
<td>( U )</td>
<td>Effective heat transfer coefficient</td>
<td>( J/Ks )</td>
</tr>
<tr>
<td>( Q )</td>
<td>Heat added through fan</td>
<td>( W )</td>
</tr>
<tr>
<td>( u )</td>
<td>Voltage resistor (input)</td>
<td>( V )</td>
</tr>
<tr>
<td>( v )</td>
<td>Speed of fan (disturbance)</td>
<td>( l/s )</td>
</tr>
<tr>
<td>( T_{i} )</td>
<td>Air temperature in tube (output)</td>
<td>( ^\circ C )</td>
</tr>
<tr>
<td>( T_{o} )</td>
<td>Air temperature in environment</td>
<td>( ^\circ C )</td>
</tr>
<tr>
<td>( T_{a} )</td>
<td>Temperature in equipment</td>
<td>( ^\circ C )</td>
</tr>
<tr>
<td>( \tau_{c} )</td>
<td>Time delay of disturbance to output</td>
<td>( s )</td>
</tr>
<tr>
<td>( \tau_{a} )</td>
<td>Time delay of input to output</td>
<td>( s )</td>
</tr>
</tbody>
</table>

Table 1: Parameters used in the thermal process mechanistic model
3.2 An “ARMAX” Model Identification

For the purpose of identification the data sequences shown in Figure 3 are used. The sampling interval due to the long time constant of the system and its slow dynamics behavior is chosen $\Delta t=0.1$ sec and the sequence contains about 10000 samples. The input $u(t) \in [0,2]$ volts, is an exciting signal with normal random distribution which covers the full range of input operation. The speed of the fan varies with its driver voltage $v(t) \in [3,5]$ volts, and acts as the disturbance. This variable has deviation over the full range of its operation in a pseudo-random manner. Application of these random input and disturbance to the experimental plant result an output temperature $y(t)$ shown in Fig.3.

![Figure 3: Data sequences used for identification](image)

Using these data sequences collected at the room temperature and acts as offset, the following ARMAX model for the overall range of variations of input and disturbance will be obtained [19]. This model is estimates the temperature of the heating range of variations of input and disturbance will be obtained $\Delta y$. Using these data sequences collected at the room temperature and acts as offset, the following ARMAX model for the overall range of variations of input and disturbance will be obtained $\Delta y$. This model is estimates the temperature of the heating range of variations of input and disturbance will be obtained $\Delta y$.

\[
Y(z) = H(z)U(z) + G(z)V(z)
\]

where the output is given by:

\[
y(t) = H(z)u(t) + G(z)v(t)
\]  

![Figure 4: Simulation of an ARMAX modeling](image)

3.3 The Multiple Model Based Identification

The single ARMAX model developed before is not sufficiently accurate for the whole operating region of the non-linear process considered here. Thus, one may look for accurate models valid at some smaller operating regions of the process. This means, first the different operating regimes of the thermal process should be identified. For this purpose, we applied a wide range of step changes to both input and disturbance and collected the resulting output samples. This is done by applying constant input with deviations in disturbance and fixed disturbance by changing the input. This provides rich information of the system operating regimes. It is obvious that the steady state response of the plant depends on both $u(t)$ and $v(t)$.

There are several methods of searching algorithms for optimal decomposition of the overall plant into different operating regimes [13],[15]. Most of these algorithms are heuristic and depend on exhausting search methods to find the best operating regimes. According to the non-linear steady-state response gain characteristic of the system, we chose to combine four local ARMAX model structures into an NARMAX model structure. The input and disturbance deviations are thus decomposed into the following separate regimes. In advanced applications, intersections between regimes can be considered:

- Regime #1: $u(t) \in [0,1]$, $v(t) \in [3,4]$
- Regime #2: $u(t) \in [0,1]$, $v(t) \in [4,5]$
- Regime #3: $u(t) \in [1,2]$, $v(t) \in [3,4]$
- Regime #4: $u(t) \in [1,2]$, $v(t) \in [4,5]$

Four separate data sequences with the necessary deviations in each operating regime are generated to identify and construct local ARMAX models. Based on these data sets, the following local models are obtained:

\[
H_{11}(z) = -0.000458z^{-2} + 0.000526z^{-2} - 1.9838z^{-1} + 0.9839
\]

\[
H_{12}(z) = -0.00072z^{-2} + 0.000804z^{-2} - 1.9765z^{-1} + 0.9766
\]

\[
H_{21}(z) = -0.00113z^{-2} + 0.00118z^{-2} - 1.9765z^{-1} + 0.9766
\]

\[
H_{22}(z) = -0.00396z^{-2} + 0.007835z^{-2} - 0.7273z^{-2} - 0.2697
\]

where the output of each regime is obtained from (3.2). The errors have been calculated based on the “Normalized Root Mean Square Error” (NRMSE) [4].

\[
\varepsilon = \sqrt{\frac{\sum (\hat{y} - y)^2}{\sum y^2}}
\]

where $\hat{y}$ is the output vector of model and $y$ is the actual output of the thermal process. The errors between model and actual plant in each operating regime are shown in Table 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>NRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Model #1</td>
<td>0.0801</td>
</tr>
<tr>
<td>Local Model #2</td>
<td>0.0973</td>
</tr>
<tr>
<td>Local Model #3</td>
<td>0.1006</td>
</tr>
<tr>
<td>Local Model #4</td>
<td>0.0253</td>
</tr>
</tbody>
</table>

\[\text{Table 2: Errors of the identified local models}\]
4 Validity and Interpolation Functions

To combine the four local ARMAX model structures into an NARMAX model in a smooth manner as mentioned in local multiple model section, we need to define a validity function which shows the relative validation of each local model. The validity functions are considered two dimensional Gaussian functions which are illustrated in Fig.5:

\[
\rho_i(u,v) = e^{-\frac{1}{2}\left(\frac{(u-u_i)^2}{\sigma_u^2} + \frac{(v-v_i)^2}{\sigma_v^2}\right)}
\]

(4.1)

According to (2.7) interpolation functions can be defined as:

\[
w_j(u,v) = \frac{\rho_j(u,v)}{\sum_{i=1}^{4} \rho_j(u,v)}
\]

(4.2)

The output of the model can be found by combining the outputs of local models with interpolating functions which varies by the value of input and disturbance. This can be done by an offline algorithm shown in the block diagram of Fig. 6.

The output of the model can be found by combining the outputs of local models with interpolating functions which varies by the value of input and disturbance. This can be done by an offline algorithm shown in the block diagram of Fig. 6.

The NRMSE of the single ARMAX model and the NARMAX multiple model approach are compared in Table 3b. It is evident that multiple model performs much better than a single ARMAX model acting globally.

<table>
<thead>
<tr>
<th>Optimal Variances</th>
<th>Input Variance</th>
<th>\sigma_u=0.43</th>
<th>Disturbance Variance</th>
<th>\sigma_v=0.40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Single ARMAX</td>
<td>0.1453</td>
<td>Multiple Model</td>
<td>0.0317</td>
</tr>
</tbody>
</table>

Table 3a: Optimal variances

Table 3b: Errors of identified models

5 Real-Time Control and the Implementation Results

In this section the procedure for designing the Real-Time control system is presented. The global controller for the thermal process consists of four local digital PID filters [27] with the structure given below:

\[
D(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2}}{1 + b_1 z^{-1} + b_2 z^{-2}}
\]

(5.1)

where \(a_i\) and \(b_i\) are defined as:

\[a_i = K_p + K_i \frac{K_T}{T} + K_0 \frac{2K_p}{T} \quad a_1 = -K_p + K_i \frac{K_T}{T} \quad a_2 = K_p \quad b_1 = -1 \quad b_2 = 0\]

and \(K_p, K_i\) and \(K_0\) are the proportion, integration and differentiation coefficients. There are several methods to tune the PID controllers by auto-calibration [28]. The criteria in most methods depends on overshoot and settling time or gain margin and phase margin. One of the practical methods to apply is a dynamic system simulation for MATLAB which is called nonlinear control design (NCD) [21]. The NCD blockset uses time domain constraint bounds to represent lower and upper bounds on response signals. Constraint bounds can be changed to meet the best performance. Based on this technique, for each of the local models a digital controller is obtained as:

\[D_1(z) = \frac{195.5 - 385.5 z^{-1} + 190 z^{-2}}{1 - z^{-1}} \quad D_2(z) = \frac{205 - 405 z^{-1} + 200 z^{-2}}{1 - z^{-1}}\]

\[D_3(z) = \frac{186 - 366 z^{-1} + 180 z^{-2}}{1 - z^{-1}} \quad D_4(z) = \frac{215 - 425 z^{-1} + 210 z^{-2}}{1 - z^{-1}}\]
These controllers are combined via the validity and interpolation functions to obtain the global controller for the thermal plant. It is noted that the validity and interpolation functions were found in the previous section. The block diagram of this implementation is shown in Fig.9. The saturation function at the output port limits the level of the control signal which applies to the DAQ card to avoid the unbounded signals.

![Block Diagram of Real-time control system](image)

Figure 9: Block Diagram of Real-time control system

To illustrate the performance of the multiple model controller three random setpoint changes at 15, 120 and 160 sec. together with disturbance changes at 35, 140 and 270 sec. with 20 sec. duration were applied. The data sequences are shown in Fig.10. The results of an experiment with a single PID controller and its comparison with the Multiple-Model based control are shown in Fig.11. As it is clear from this experiment, the multiple models show better performance (such as lower overshoot) versus the single PID controller.

![Applied setpoints and disturbances](image)

Figure 10: Applied setpoints and disturbances

![Closed-loop responses](image)

Figure 11: Closed-loop responses

6 Conclusions

In this paper Multiple-Model based control and identification of a nonlinear thermal process are presented. After defining several operating regimes for the operation of the nonlinear process, local models and local controllers are developed. Then these models and controllers were put together to find a global model and a global controller. This method will simplify modeling and control of complex systems. Besides, a useful environment were set up for the Real-Time implementation on the experimental nonlinear thermal process.

References