

Nonlinear System Identification in Frequent and Infrequent Operating Points for Nonlinear Model Predictive Control

Alireza Fatehi^{*1}, Behnaneh Sadeghpour², Batool Labibi³

^{1,2,3} *Advanced Process Automation & Control (APAC) research group,
Industrial Control Center of Excellence, Faculty of Electrical & Computer Engineering,
K.N. Toosi University of Technology, Seyyed-Khandan, P.O.Box: 16315-1355, Tehran, Iran
e-mail: fatehi@kntu.ac.ir¹, sadeghpour@yahoo.com², labibi@eetd.kntu.ac.ir³*

crossref <http://dx.doi.org/10.5755/j01.itc.42.1.997>

Abstract. This paper studies identification of a process in both frequent and infrequent operating points to design a nonlinear model predictive controller. Although, many of industrial processes normally work around an operating point, however they should seldom work in some infrequent points as well. In this case, due to low ratio of data points, identification of the processes based on all data set results in poor identification of the infrequent operating points. To resolve this problem, in this paper, at the first step, a data clustering strategy is used to group the data in different operating points. Since the ratio of infrequent to frequent data points is extremely low, the strategy used is the fuzzy Gath-Geva clustering methodology. Then, at the second step, a new approach has been proposed to compromise performance of identification of the nonlinear model for frequent and infrequent operating points. It is shown that if the ratio of data associated with frequent operating point to data of infrequent operating point is appropriately selected, the performance of the model remains satisfactory in the frequent operating point while the performance in the infrequent operating point is significantly improved as well. The proposed method gives an interval for appropriate ratio of data set in the highly nonlinear pH neutralization process.

Keywords: Nonlinear System Identification; Nonlinear Model Predictive Control (NMPC); Fuzzy Clustering; Multilayer Perceptron (MLP) Neural Networks.

1. Introduction

Model predictive control is widely used in industrial processes because of simplicity, understandability, handling constraints and dealing with time delays. The method is based on prediction of future outputs to generate a future control action to minimize a performance index of error and control signals [15], [3]. Since this strategy strongly depends on the underlying model, its performance is violated if the quality of the predictor model is low. Due to severe nonlinearity of most of industrial processes and efficacy of MPC in control of nonlinear systems, nonlinear MPC (NMPC) has attracted more attention in the last two decades [14]. This kind of control strategy uses nonlinear models to predict future output. In [20] & [16] Wiener models, in [2] & [17] Volterra series models and in [11] & [5] Hammerstein model are used as the nonlinear models in NMPC. Multiple models strategy [4] & [21] is another suitable choice for a nonlinear plant where combination of some local linear models is used to model the

nonlinear plant. Moreover, neural networks have been extensively used as nonlinear models in model predictive control of nonlinear processes [19], [13], [1], [11] & [23]. In all of these works, nonlinear model is the fundamental part of the controller which quality of control heavily depends on.

To identify a process, the most important step is to design excitation signals according to important frequency range of the nonlinear system at different operating points [18]. However, in many real applications, it is not always possible to apply the designed excitation signals to the real system. This necessitates using usual data of the process which are logged into historical data servers. These data are obtained from different operating points with different data ratio and different weights of importance. Industrial processes usually work in several limited main or frequent operating areas. However, changes in quality of input materials, disturbances or so on necessitate employing a process in infrequent and unusual operating points. If the data set used to

* Corresponding author

identify the nonlinear model of the plant does not involve enough data of these operating points, the plant will not be accurately identified. Consequently, the prediction results in large error which may cause large control error or even instability. On the other hand, using the same number of data points for both frequent and infrequent operating points decreases quality of the identified model for frequent operating points and subsequently increases overall cost of using the model.

In this paper, we propose a method to design a nonlinear MPC so that it has satisfactory performance in frequent operating point while the quality of control at infrequent operating point is not less than an acceptable threshold. The proposed method recommends using a portion of data in the infrequent area much larger than its real portion in the actual data set but much less than the portion of data of frequent operating area. This strategy improves identification quality in the infrequent operating area while it keeps the relative identification quality in frequent operating area as well.

However, there are two main problems associated with this strategy. The first one is how to specify various operating points of the process as frequent and infrequent data points. For this purpose, data clustering is employed to separate data in different operating points. Since the ratio of data in frequent to infrequent operating points is enormously high, fuzzy Gath-Geva clustering [7] has been used. The second problem is how to select the ratio of identification data to accurately identify a nonlinear model for the system. In this direction, we specified an approximate value for the ratio of training data by providing an especial experiment.

The pH neutralization process is selected in order to show the efficacy of the proposed algorithm. Due to severe nonlinearity of this process, it is known as one of the most difficult processes to be controlled [22]. Nonlinearity of pH neutralization is more on its steady state gain which may change by a factor of 30 for a low nonlinear process up to 1000 for high nonlinear one. A pH neutralization process can be modeled well by Wiener [16] or Hammerstein [5] model. Indeed, linear model with adaptive dynamic [6] or general nonlinear dynamic model has also been used [8] in some researches. In this paper, we use a neural network ARX (NARX) model for the process. By implementing the proposed algorithm in identification of a nonlinear model for NMPC of a pH process, error

$$J(M_1, M_2, M_u) = \sum_{k=M_1}^{M_2} \delta(k) [\hat{y}(t+k) - r(t+k)]^2 + \sum_{k=0}^{M_u-1} \lambda(k) [\Delta u(t+k)]^2 \quad (1)$$

where $\hat{y}(t+k)$ is k steps ahead prediction of the output, $r(t+k)$ is k steps ahead reference signal, $u(t+k)$ is k steps ahead control signal and Δ is the backward difference operator. $M_1 > 1$ and M_2 present interval of output prediction and M_u is the control horizon. Normally, M_1 is equal to the delay of the

of frequent operating points remains low while at the same time error of infrequent operating points is in an acceptable range. Indeed, the overall error is near to minimum.

The rest of the paper is organized as follows: In the next section a review of the model predictive control strategy is given. In Section 3, the clustering method has been explained to specify frequent and infrequent data points. Section 4 presents the main result on nonlinear identification in both frequent and infrequent operating points, in which the proper ratio in identification data set is obtained for a pH process. In section 5, the proposed method is used in a neural network model predictive control system and finally section 6 gives conclusions.

2. Review of model predictive control

Figure 1 shows the general block diagram of model predictive control strategy. Contrary to the classical control which only uses past and present output or states of the plant, the model predictive control considers prediction of future signals as well. All the MPC algorithms have common components which are predictive model, cost function and calculation of future control to minimize the cost function.

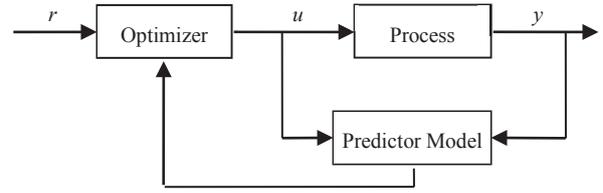


Figure 1. The general structure of MPC

Predictive model is a fundamental part of an MPC. At the first step, a mathematical model of the process is identified based on past inputs ($u(t+i); i \leq -1$) and past outputs ($y(t+i); i \leq 0$). This model, together with future inputs ($u(t+k); 0 \leq k \leq M_u$) is employed to predict future outputs ($\hat{y}(t+k|t), 1 < k < M_2$). Since the future inputs are unknown, practically the model gives future outputs as a function of future inputs. At the next step, an optimization problem is solved using a cost function of predicted error and future control effort. A general expression for such cost function can be given as follows:

process [15]. Coefficients $\delta(k)$ and $\lambda(k)$ affect performance of the controller and usually are considered to be constant.

Industrial processes are subject to some constraints in input signal, output signal, their differentiation through time, acceptable overshoot and so on. These

constraints can be dealt in MPC with minimization of Eq. (1) subject to defined constraints. Solution of this optimization problem is the future input to apply as control signal to the process. In order to obtain $u(t+k)$, it is necessary to minimize the cost function J . To do this, predicted output $\hat{y}(t+k)$ is substituted in the cost function based on the prediction model as a function of previous and present inputs and previous outputs which are known and future control signal which is unknown [11].

Despite of the fact that most of industrial processes are nonlinear in nature, many applications of MPC so far have been applied to linear models [3]. However, there exist processes whose nonlinearities are severe enough not to be negligible. Accordingly, nonlinear model predictive control, NMPC which uses a nonlinear model in MPC strategy is very helpful and justifiable [14]. In nonlinear modeling, it is possible to use neural models, fuzzy models, a combination of both or other nonlinear models such as Wiener and Hammerstein models. In this paper, multilayer perceptron (MLP) neural network is selected as the process model.

To identify a nonlinear model, it is necessary to excite the system at different operating points. Acquired data are saved in a database to be used in identification of the plant. The structure of model (number of neurons), algorithm of parameter adjustment (training of neural networks), quality of produced data and preprocessing of data (noise elimination, delay estimation and so on) affect quality of the model [3].

Since in neural network training, mean of square of errors is used as the criterion of parameter adjustment, volume of produced data in any area heavily affects quality of identification in that area. If the number of data points in an operating point is much less than the number of data points in another operating point, the effect of modeling error for the first data set is less than that for the second data set in the cost function. This reduces the efficacy of the cost function on designing a good model for the first operating point, despite the error is considerable in this area. In fact, training algorithms have more tendencies to reduce small errors in regions with a large amount of data than to reduce large errors in regions with small amount of data.

3. Data separation by fuzzy Gath-Geva clustering

Identification data may be produced either actively using some special experiment or passively using the ordinary data of the process during its operation. In the first case, one may excite the process in different operating points and select necessary data in each one.

$$d^2(x_i, v_j) = \frac{1}{p_j} \sqrt{\det(A_j)} \exp\left(\frac{1}{2}(x_i - v_j)A_j^{-1}(x_i - v_j)\right) \quad (3)$$

However, this may not possible in a plant during its normal operation. In this case, the only available data for identification is the ordinary data logged during the operation. These data include both frequent and infrequent data points. Therefore, it is necessary to separate data of different operating points and use specific ratio of them for identification, as will be presented in Section 4. For this purpose clustering is employed to separate frequent and infrequent data points.

There are various clustering methods [12]. C-means clustering is a basic method in which data are clustered in some hyper-sphere clusters. Gustafson-Kessel method was introduced in [9] in which the clustering distribution variation can be different in each direction which means each cluster is a hyper-ellipsoid. However, the total clustering error proportionally depends on the number of data points in that cluster in both of the above methods. Therefore, the total clustering error for a small size but scattered and high distributed cluster may be much less than the total clustering error for a condense but large size cluster. As a result, the clustering algorithm may divide the large cluster in two parts and includes small groups of data in each of these two large clusters to minimize the overall clustering error. Gath and Geva [7] presented another clustering in which the number of data points in each cluster is inversely affects the clustering error. Therefore, a small cluster may have the same effect in the overall clustering error as a large one with more or less the same distribution of data. In this section, fuzzy Gath-Geva clustering algorithm is used to cluster the data in frequent and infrequent operating points.

3.1. Gath-Geva clustering

Clustering is a nonsupervisory classification technique in which the data set is divided into c clusters based on the similarity of the members of each other and their difference from those of other clusters. Consider data set X with data points $x_i = (x_{i1}, x_{i2}, \dots, x_{il})$, $i = 1, 2, \dots, M$, where M is the number of data points in X and l is the dimension of X . The problem is to find c cluster centers $v_j = (v_{j1}, v_{j2}, \dots, v_{jl})$, $j = 1, 2, \dots, c$, and membership of x_i to the cluster C_j such that an overall cost function of distance of data from their cluster center is minimized. The distance function in fuzzy Gath-Geva clustering is defined as [7]:

$$J = \sum_{x_i \in X} \sum_{v_j \in V} f^m(x_i, v_j) d^2(x_i, v_j) \quad (2)$$

where $d^2(x_i, v_j)$ is the distance of x_i from center of v_j :

$f(x_i, v_j)$ is the degree of membership of x_i to cluster v_j which is computed during the clustering, m is used to change the fuzziness of the clusters and A_j is the covariance matrix of cluster C_j . P_j is cardinality of cluster C_j which is defined as the proportion of sum of the membership values of all data points to C_j to sum of all membership values of all data points to all of the clusters.

3.2. Clustering of pH neutralization process

pH neutralization process is selected to illustrate clustering frequent and infrequent operating points. pH process, as demonstrated in Figure 2, has three reaction streams: HNO₃ as an acid, NaOH as a base and water. It has two output variables: liquid level and pH of the output stream. Water flow rate is deployed to control the liquid level and base flow rate is the control signal on pH control loop where acid flow rate is disturbance. A detailed model of this process is presented in [10] which is used in this study.

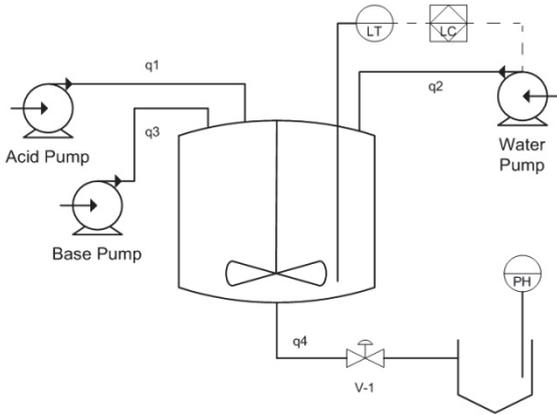


Figure 2. pH Neutralization process

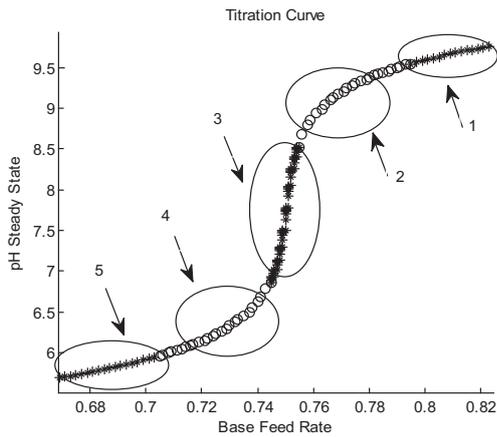


Figure 3. The titration curve of pH process

This process is a severe nonlinear process. Figure 3 shows a typical titration curve of a pH neutralization process. As it is shown in this figure, gain of the process could be categorized into 3 different areas; low gain in areas 1 & 5, medium gain in areas 2 & 4 and high gain in area 3.

If we design controller of this process in any area other than area 3, changing the operating point from that area to area 3 may result in oscillation or even instability. Suppose the pH process usually operates in area 2 but operates at a very small interval of time in area 3 due to some disturbances. Accordingly, we consider area 2 as the main or frequent area and area 3 as the rare or infrequent area. The question is how to separate data points of area 2 and 3 from each other.

Since dynamic behavior of processes is concerned in this paper, each data point consists of the regression vector used later in nonlinear identification of the process. This vector, as will be stated in sub-section 4.1, is $(y(t), y(t - 1), y(t - 2), y(t - 3), u(t - 1), u(t - 2))$, where $u(t)$ is the inlet base feed rate and $y(t)$ is the outlet pH value. Suppose N_1 is the number of data points in the infrequent area and N_2 is the number of data points in the frequent area and $N = N_2/N_1$. In other words, N is the ratio of period of time the process operates in the frequent area with respect to that of the infrequent area.

We apply an amplitude modulation pseudo random binary signal (APRBS) [18] with length $N_1 = 2000$ step. The amplitude of the inlet flow is selected so that the operating point remains in area 3. Another APRBS is also applied with different length of N_2 and with the amplitude so that it operates in area 2. The regressor data are clustered using the Gath-Geva clustering method. Figure 4 illustrates the clustering result for $N = 100$ where the number of clusters is $c = 5$. Figure 4a shows the first element (the pH value) of each data point. The horizontal axis is the element number in each cluster. The clustering is near optimal after 10 training steps. Figure 4b illustrates the center of the clusters for these 10 steps of training. Although the number of infrequent data points is too small Gath-Geva clustering could accurately cluster them. The frequent data points are divided into 4 different clusters. To merge these clusters, we define a minimum threshold for the distance between the centers of the clusters.

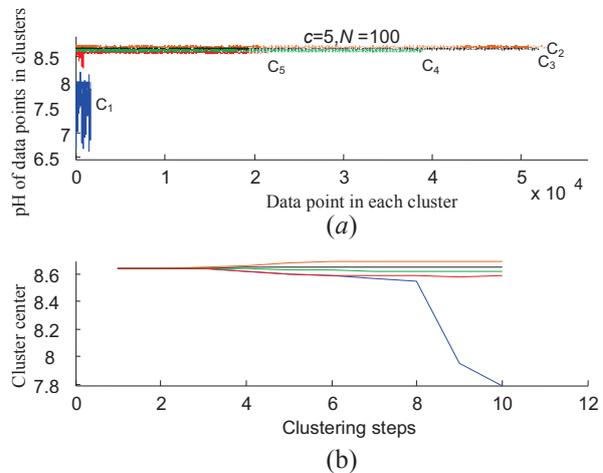


Figure 4. Clustering of pH neutralization process into 5 clusters

4. Nonlinear identification for infrequent operating points

The main question in simultaneous identification of frequent and infrequent operating areas is how to select the ratio of the data in these areas to identify them such that the following objectives are satisfied:

1. Identify the main operating area with a high quality such that the model error in this area is low.
2. Identify the process in the infrequent area such that the model error in this area is not high.
3. The overall identification error of the model over all the operating range is optimally low.

To satisfy these objectives, we have to find the ratio of data to be employed for identification. In subsection 4.2, through several experiments, we will find the appropriate ratio. To this end, first we present some practical concerns in modeling of pH process in subsection 4.1.

In identification of pH process similar to other systems, several practical issues should be considered which are explained in details in [21]. One of the important factors in a good identification is quality of the identification data. Accordingly, the identification data need to be informative, i.e. excite different frequencies and operating points of the process dynamics. To this end, we use a binary signal with admissible amplitude to excite all modes of the system. Also, we add noise to that signal to increase the quality of identification and to have a better excitation. We consider the amplitude of noise roughly 1/15 of the signal amplitude. To collect data, open loop identification is employed which is acceptable due to stability of the pH neutralization process. Since the time constant of the process is 350 to 400 seconds, we select the sample time as 30 seconds. The input vector of the neural network consists of $y(t-1), y(t-2), y(t-3), u(t-1), u(t-2)$ and its output is $y(t)$. The number of neurons is chosen based on an error threshold. Consequently, the neural network is composed of a hidden layer with 5 and, in some of experiments, with 7 neurons.

4.1. Practical concerns in pH process identification

This section explains how the amount of required data in different areas is selected to satisfy the mentioned objective of simultaneous identification of frequent and infrequent operating points. In order to find the best ratios, we provide the following experiment. Suppose N is the ratio of period of time the system operates in the frequent areas with respect to that of the infrequent areas. After clustering, part of data in frequent cluster and part of that from the infrequent cluster is selected to identify the plant. If we select n_1 and n_2 as the numbers of training data sets in infrequent area and frequent area, respectively, then $n = n_2/n_1$ is the ratio of training data sets. In our experiment, for different values of N we identify the

process with different value of n . Also, another data set is used as test data set, which its ratio is N_2 to N_1 as well, to measure the infrequent data error, frequent data error and the total error in each case. Table 1 compares MSE_1 , MSE_2 and MSE_T which are mean square of identification errors of infrequent data, frequent data and total data sets, respectively, for different values of n and N .

As depicted in the table, if for example the length of the infrequent operating area of the process is one tenth of that of the main operating area and the ratio of the training sets for the main and the infrequent data sets is selected as 2 to 1, then the total error of the performance of the test data in the infrequent operating point is equal to $0,86 \times 10^{-6}$. Figures 5, 6, and 7 show MSE for infrequent operating point, main operating point and total operating points for different values of n and N , respectively.

In Figure 5, we observe that by increasing n , the error associated with the infrequent operating point increases. In fact, the ability of the neural model to predict the output of the infrequent operating point decreases due to decrease of proportion of data in this operating point to the main operating point in the total data. However, in all of the error curves in Figure 5 up to $n \leq 25$, the MSE of data in the infrequent operating point has a negligible value and performance of the model is acceptable.

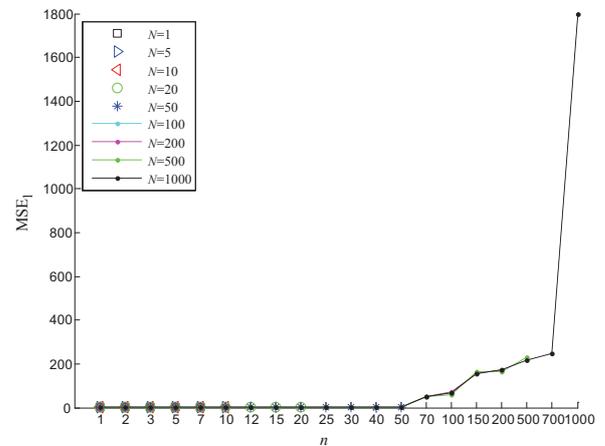


Figure 5. Identification error of the infrequent operating point for different values of N and n

As it can be observed from Figure 6, the error decreases for the main operating point by increasing the proportion of data of the main operating point to the infrequent operating point for all different values of N . This is because the neural network model better identified the main operating point. Hence, the error associated with this area decreases. However, the rate of this decrease for lower ratios is rapid and for larger ratios is very slow. In all the error curves for $n \geq 12$ this rate is very slow and approximately equal to zero.

By studying Figure 7, we observe that MSE value for data sets associated with the combination of infrequent and main operating points initially

Table 1. The identification MSE for total data set, infrequent data set and frequent (main) data set (all values are in 10^{-6})

N	n	1	2	3	5	7	10	12	15	20	25	30	40	50	70	100	150	200	500	700	1000	
1	MSE _T	0.4706	0.4758	0.4749	0.4645	0.469																
	MSE ₁	0.8198	0.8553	0.8602	0.843	0.8529																
	MSE ₂	0.1215	0.0962	0.0896	0.086	0.0852																
5	MSE _T	0.2278	0.221	0.2252	0.2125	0.2144	0.1486															
	MSE ₁	0.8187	0.8541	0.8993	0.8377	0.8518	0.8634															
	MSE ₂	0.1099	0.0947	0.0906	0.0877	0.0872	0.0773															
10	MSE _T	0.188	0.155	0.1485	0.1469	0.1477	0.1486															
	MSE ₁	1.07	0.8565	0.8219	0.8381	0.8525	0.8636															
	MSE ₂	0.1	0.085	0.0813	0.078	0.0774	0.0773															
20	MSE _T	0.1403	0.1278	0.1246	0.1206	0.1208	0.1211	0.1216	0.1217	0.1224												
	MSE ₁	0.8189	0.8524	0.8707	0.8346	0.8537	0.8616	0.8734	0.8739	0.8874												
	MSE ₂	0.1065	0.0917	0.0873	0.085	0.0843	0.0842	0.0841	0.0842	0.0843												
50	MSE _T	0.1348	0.1148	0.1106	0.1069	0.1064	0.1064	0.1065	0.1066	0.1065	0.1067	0.135	0.139	0.141								
	MSE ₁	0.8378	0.8534	0.8618	0.8481	0.8569	0.8633	0.8724	0.8721	0.868	0.8724	2.316	2.523	2.602								
	MSE ₂	0.1208	0.1001	0.0956	0.0921	0.0914	0.0913	0.0912	0.0913	0.0913	0.0914	0.092	0.092	0.092								
100	MSE _T	0.117	0.099	0.0965	0.0934	0.0937	0.0929	0.093	0.093	0.0932	0.0931	0.0928	0.0925	0.0923	0.6	0.73						
	MSE ₁	0.829	0.8453	0.858	0.8429	0.8992	0.8635	0.8733	0.8676	0.8898	0.8721	0.8529	0.83	0.8094	51.84	65.77						
	MSE ₂	0.1099	0.0916	0.0889	0.0859	0.0857	0.0852	0.0852	0.0853	0.0853	0.0853	0.0853	0.0852	0.0851	0.0851	0.0856						
200	MSE _T	0.1112	0.0958	0.092	0.0888	0.0882	0.0881	0.0881	0.0882	0.0882	0.0883	0.0881	0.0881	0.098	0.34	0.43	0.9	1				
	MSE ₁	0.8198	0.8553	0.8602	0.843	0.8563	0.8635	0.8742	0.8734	0.8682	0.8711	0.8458	0.8317	2.748	50.69	70.33	156	174.7				
	MSE ₂	0.1077	0.092	0.0881	0.085	0.0844	0.0843	0.0842	0.0843	0.0843	0.0844	0.0843	0.0843	0.084	0.084	0.084	0.084	0.088				
500	MSE _T	0.1128	0.0978	0.0944	0.0906	0.0899	0.0897	0.0897	0.0898	0.0898	0.0898	0.0897	0.093	0.093	0.19	0.2	0.4	0.4	0.5			
	MSE ₁	0.8189	0.8524	0.903	0.8346	0.8537	0.8616	0.8734	0.8739	0.8874	0.862	0.8531	2.521	2.637	51.39	57.17	162.4	166	231.5			
	MSE ₂	0.1114	0.0963	0.0928	0.0891	0.0883	0.0882	0.0881	0.0882	0.0882	0.0883	0.0882	0.089	0.088	0.088	0.088	0.088	0.088	0.088			
1000	MSE _T	0.1393	0.1249	0.1216	0.1188	0.1177	0.1174	0.1174	0.1175	0.1175	0.1174	0.1174	0.1173	0.119	0.17	0.18	0.3	0.3	0.3	0.4	1.9299	
	MSE ₁	0.8187	0.849	0.8737	0.876	0.8545	0.863	0.8737	0.8718	0.8884	0.8613	0.8527	0.8301	2.606	50.49	66.5	154.9	173.9	215.4	245.5	1800	
	MSE ₂	0.1387	0.1241	0.1209	0.118	0.1169	0.1167	0.1167	0.1167	0.1167	0.1167	0.1167	0.1166	0.117	0.116	0.116	0.116	0.116	0.116	0.116	0.116	0.116

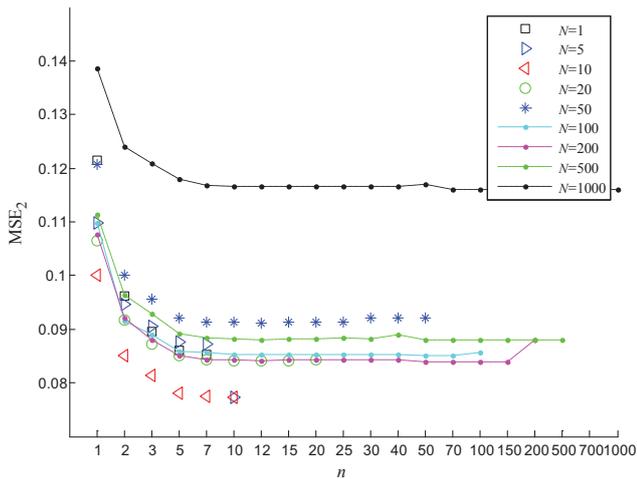


Figure 6. Identification error of the main operating point for different values of N and n

decreases with slow rates and then increases with rapid rates. Hence, to minimize the error cost function, the appropriate ratio is in an interval around the optimal point.

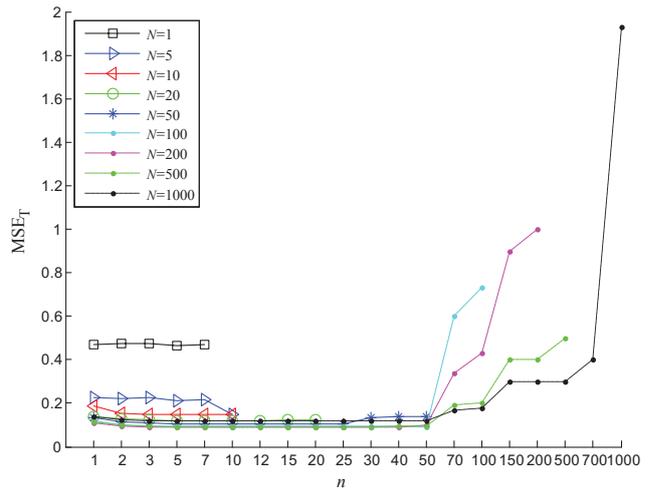


Figure 7. Identification error of the total operating points for different values of N and n

According to Figures 5 to 7, we can conclude that for the main operating point, the ratios greater than 12 and for the infrequent operating point the ratios less than 25 are acceptable. Therefore, the common

interval of [12,25] is acceptable. Nevertheless, for $N < 12$ we can set $n = N$.

Thus, we find an appropriate value for the ratio of the identification data of the main operating point to that of the infrequent operating point for the pH process such that the modeling objectives are satisfied. In fact, through this selection, despite reduction of amount of selected data from the main operating point, MSE does not increase significantly. On the other hand, the model is sufficiently trained in the infrequent operating area to improve not only the model in this area but also the overall performance of the system.

As a consequence, the overall nonlinear identification of a process in frequent and infrequent operating points can be summarized in the following steps:

- Step 1.** Gather the actual data of the process during its operation.
- Step 2.** Cluster the data using hierarchical Gath-Geva fuzzy clustering.
 - Sub step 2.1.** If standard deviation of any of the clusters is high, divide it into smaller clusters.
 - Sub step 2.2.** If centers of some clusters are close to each other, merge these clusters.
- Step 3.** Select part of data of each cluster for identification according to the following ratio rule:

Ratio rule: If actual ratio of data between clusters is less than 25, use the same ratio of data for identification; otherwise use some ratio between 12 to 25.
- Step 4.** Identify the nonlinear model using the selected data.

5. Nonlinear neural model predictive control

In the previous section, we found an optimal interval for ratio of the main data to the infrequent data when a nonlinear model is identified for both operating points. In this section, we use this ratio to train a neural network to identify a nonlinear model of the process to be used in an NMPC structure.

The actual ratio of data is supposed to be $N = 1000$. In the identification process we selected different training ratios as $n = 1, 15$ and 1000 . Then, each of the identified models for any of these ratios is used in the model predictive control structure. The closed loop tracking results for these controllers are shown in Figures 8 to 10. In these figures, the plant is in infrequent operating area during the first 2000 sample times. After that it works in the main operating area for 2,000,000 sample times, however just portion of it is shown in these figures. MSE values of reference input tracking for neural model predictive control using different models are given in Table 2.

Table 2. MSE of tracking in NMPC for different ratios of training data sets (all values are in 10^{-6}).

n	1	15	1000
MSE _T	3.48	0.90	0.98
MSE ₁	31.66	22.68	135.20
MSE ₂	3.46	0.88	0.84

Figure 8 illustrates the case $n = N = 1000$, i.e. the training data ratio is the same as the real data ratio. As it is shown, the output response in the infrequent operating area is oscillating. The reason is that, only one thousand of the training data are from infrequent operating point. This means that this area has not been accurately identified. So the model is not suitable for this area and the NMPC cannot control the process in this operating point. When the same number of data points are selected for both main and infrequent operating areas, i.e. $n = 1$, the results given in Figure 9 are obtained. The response is stable in all the operating points. However, comparison of the results of main operating points in Figures 8 to 10 shows that the quality of the controller degrades. This is also depicted in Table 2, where the MSE of the main operating area with $n = 1$ is 4 times more than the MSE for the same area when $n = 1000$.

The results given in Table 2 illustrate that the controller performance is the best when $n = 15$. For the main operating area, increasing n from 1 to 15 results in a significant improvement in the MSE. However, more increase of this ratio to $n = 1000$ does not have a large effect on the performance of this area. On the other hand, increasing n from 15 to 1000 significantly degrades the performance of the infrequent operating area. MSE_T values recommend that ratio 15 is also the best ratio among the others for the total test data set. Summarizing the above results, we conclude if the identification (or training) data ratio is selected as $n = 15$, all three objectives of Section 2 would be satisfied. It is important to emphasize that there is no large difference between the results when n is any value between 12 and 25. This means that the results are robust and are not largely sensitive to the ratio values.

6. Conclusions

Industrial processes generally operate in a specific operating point. However, it is a well known practice that the processes may operate in some infrequent operating point for a short period of time due to any kind of disturbances like changes in input materials quality, faults in other interacted control loops and so on. For these processes, in addition to preserving performance quality in the frequent operating area, closed loop oscillation or instability must be avoided in the infrequent operating area. However, this is a difficult task for the control system of a plant with severe nonlinearity. Many of control strategies, like

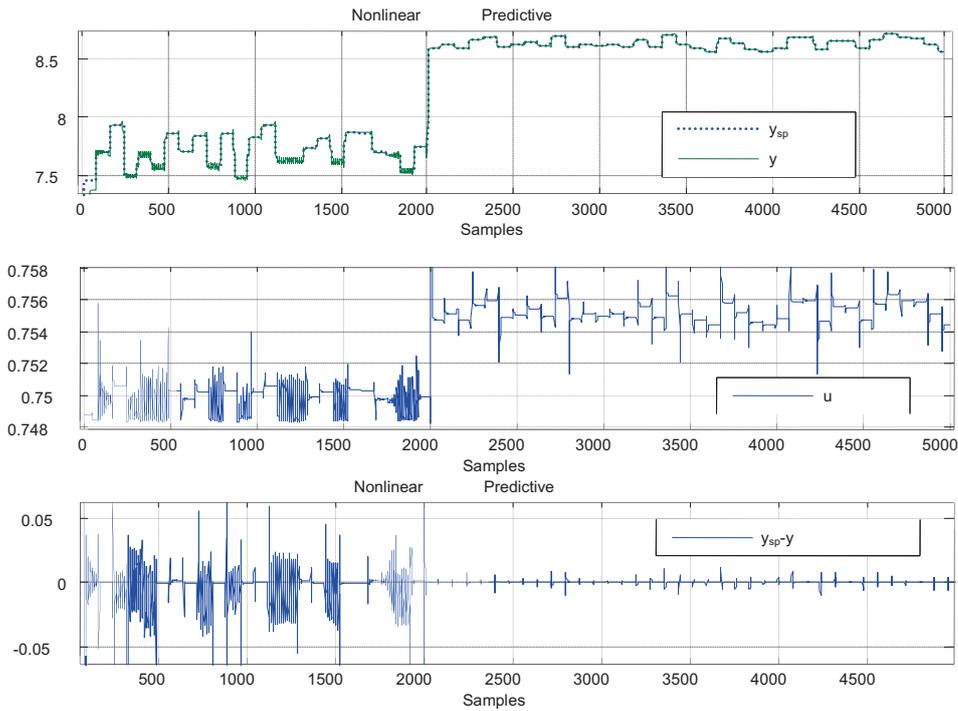


Figure 8. Nonlinear model predictive control with training ratio of $n = 1000$

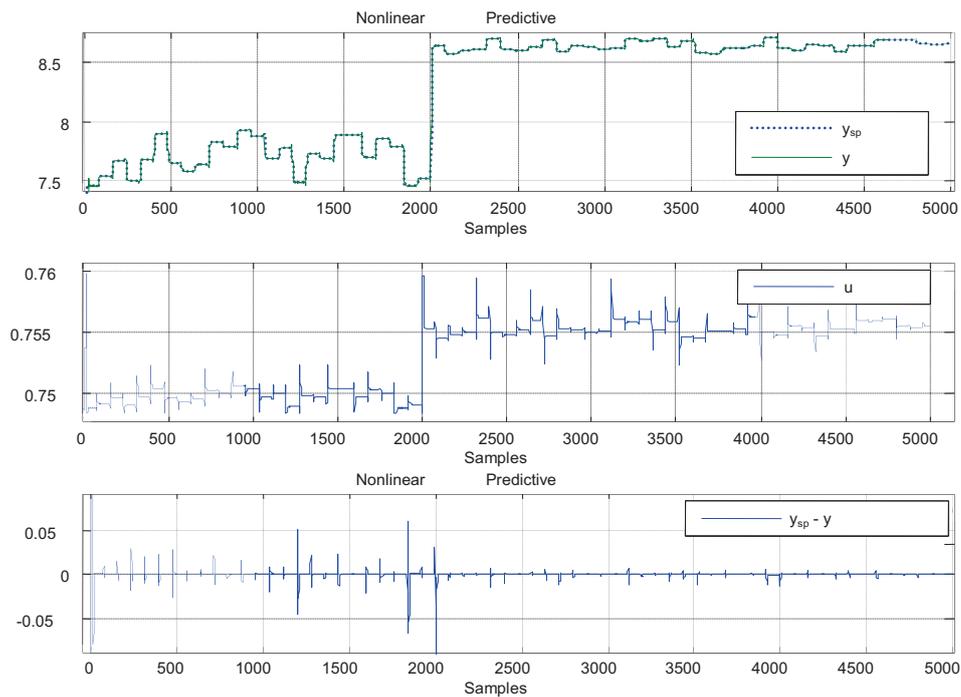


Figure 9. Nonlinear model predictive control with training ratio of $n = 1$

nonlinear model predictive control, heavily depend on the identified model. Therefore, deviation of the model from the actual behavior of the plant degrades performance of such controllers. Using low number of data point in infrequent operating space causes significant inaccuracy in that space which may end to instability. On the other hand, using the same number of data points in both frequent and infrequent operating points, causes low accuracy in most of the

time which reduces the overall efficiency. To resolve this problem, in this paper we proposed to choose an appropriate data ratio of the main operating area to that of the infrequent operating area in the identification.

The usual assumption in system identification is that we should design the excitation signal according to important frequency range and, in nonlinear identification also according to different operating

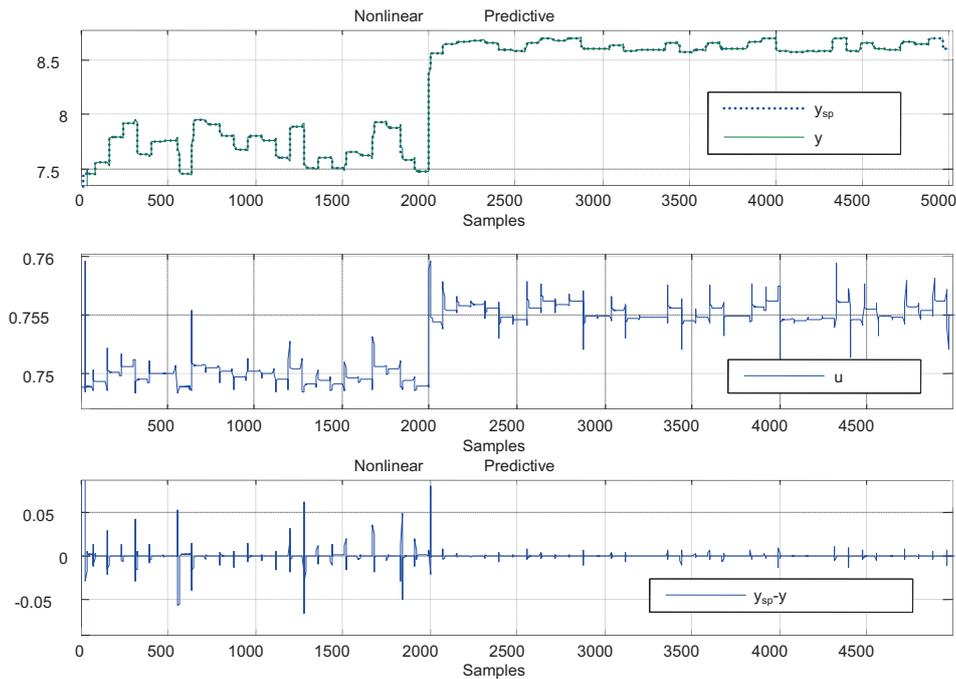


Figure 10. Nonlinear model predictive control with training ratio of $n = 15$

points [18]. However, to identify part of a process in a working factory, we are not allowed to apply our designed excitation signal due to some technical safety and conservatism. Therefore, we have to use the usual data of the process which is logged into historical data servers. These data come from different operating points with different data ratio and different importance. So, to apply the above suggested data ratio, it is necessary to distinguish various operating points. We use clustering for this purpose. This is not new, but since it is supposed that the amount of data points in various data sets are substantially different, usual method of clustering cannot properly cluster the data sets. We suggest using Gath-Geva clustering method. In addition, to improve its performance a hierarchical algorithm is presented for it.

We showed that for the pH neutralization process, which is a highly nonlinear process, by selecting data ratios between 12 and 25, in addition to high accuracy of the identified model in the main operating area, modeling error in the infrequent operating area does not exceed a specific threshold. Besides, error associated with the total data is minimized as well when this data ratio is selected. Since the ratio interval is somehow wide, the algorithm is robust to the selection of the data ratio.

Although nonlinear model predictive control of a pH neutralization process is studied in this paper as a case study, the idea of selection of specific ratio for identification of a nonlinear process can be used in other types of model based nonlinear control and processes. Also, using a neural network for model

structure is not crucial. Using any other kind of black box nonlinear models has the same problem. So the same algorithm can be applied to them for nonlinear system identification in both frequent and infrequent operating points.

References

- [1] **B. M. Akesson, H. T. Toivonen.** A neural network model predictive controller In: *Journal of Process Control*, 2006, Vol. 16, pp. 937-946.
- [2] **T. L. Aliyev, A. T. Stamps, E. P. Gatzke.** Improved crude oil processing using second-order Volterra models and nonlinear model predictive control. In: *American Control Conference, Seattle, USA*, Jun. 11-13, 2008, No. 13, pp. 2215-2220.
- [3] **E. F. Camacho, C. Bordons.** Model predictive control, Springer-Verlag, Berlin, 1999.
- [4] **B. A. Foss, T. A. Johansen, A. V. Sørensen.** Nonlinear predictive control using local models - applied to a batch fermentation process. In: *Control Engineering Practice*, 1995, Vol. 3, pp. 389-396.
- [5] **K. P. Fruzzetti, A. Palazoglu, K. A. McDonald.** Nonlinear model predictive control using Hammerstein models. In: *Journal of Process Control*, 1997, Vol. 7, No. 1, pp. 31-41.
- [6] **V. Galvanauskas.** Adaptive pH control system for fed-batch biochemical processes. In: *Information Technology and Control*, 2009, Vol. 38, No. 3, pp. 225-231.
- [7] **I. Gath, A. B. Geva.** Unsupervised optimal fuzzy clustering. In: *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1989, Vol. 11, pp. 773-781.

- [8] **A. Grancharova, J. Kocijan, T.A. Johansen.** Explicit output-feedback nonlinear predictive control based on black-box models. In: *Engineering Applications of Artificial Intelligence*, 2011, Vol. 24, No. 2, pp. 388-397.
- [9] **D. E. Gustafson, W. C. Kessel.** Fuzzy clustering with a covariance matrix. In: *Proc. IEEE Conference on Decision & Control, San Diego, USA*, 1979, pp. 761-766.
- [10] **M. A. Henson, D. E. Seborg.** Adaptive nonlinear control of a pH neutralization process. In: *IEEE Trans. Control Systems Technology*, 1994, Vol. 2, pp. 169-182.
- [11] **H. B. Huo, X. J. Zhu, W. Q. Hu, H. Y. Tu, J. Li, J. Yang.** Nonlinear model predictive control of SOFC based on Hammerstein model. In: *Journal of Power Sources*, 2008, Vol. 185, pp. 338-344.
- [12] **A. K. Jain, M. N. Murty, P. J. Flynn.** Data Clustering: A Review. In: *ACM Computing Surveys*, 1999, Vol. 31, pp. 264-323.
- [13] **A. Jazayeri, A. Fatehi, H. Sadjadian, A. Khaki-Sedigh.** Disturbance rejection in neural network model predictive control. In: *17th IFAC World Congress, Seoul, Korea, July 6-11, 2008*.
- [14] **B. Kouvaritakis, M. Cannon.** Nonlinear predictive control: theory and practice. In: *Institution of Engineering and Technology (IET)*, 2001.
- [15] **J. M. Maciejowski.** Predictive control with constraints, Prentice-Hall, N.J., 2002.
- [16] **S. Mahmoodi, J. Poshtan, M. R. Jahed-Motlagh, A. Montazeri.** Nonlinear model predictive control of a pH neutralization process based on Wiener-Laguerre model. In: *Chemical Engineering Journal*, 2009, Vol. 146, pp. 328-337.
- [17] **B. Maner, F.J. Doyle, B.A. Ogunnaike, R.K. Pearson.** Nonlinear model predictive control of a multivariable polymerization reactor using second-order Volterra models", *Automatica*, 1996, Vol. 32, pp. 1285-1301.
- [18] **O. Nelles.** Nonlinear system identification, Springer-Verlag, Berlin, 2001.
- [19] **M. Norgaard, O. Ravn, N.K. Poulsen, L.K. Hansen.** Neural networks for modeling and control of dynamic systems, Springer-Verlag, Berlin, 2003.
- [20] **R. K. AlSeyab, Y. Cao.** Nonlinear model predictive control for the ALSTOM gasifier. In: *Journal of Process Control*, 2006, Vol. 16, pp. 795-808.
- [21] **E. Peymani, A. Fatehi, P. Bashivan, A. Khaki Sedigh.** An Experimental Comparison of Adaptive Controllers on a pH Neutralization Pilot Plant. In: *IEEE Conference & Exhibition on Control, Communication and Automation (INDICON), Kanpur, India*, 11-13 Dec. 2008.
- [22] **V. S. R. Rao, R. Pickhardt, M. Chidambaram, D. Swati.** Nonlinear Control of pH System for Change Over Titration Curve. In: *Chemical and Biochemical Engineering Quarterly*, 2005, Vol. 19, No. 4, pp. 341-349.
- [23] **B. Vatankhah, A. Fatehi, A. Khaki-Sedigh.** Neural Network Model Based Predictive Control for Multivariable nonlinear systems. In: *International Multi Conferences on Systems and Control*, Yokohama, Japan, Sep. 7-10, 2010.

Received December 2011.