

# Melt Index Prediction Based on Two Compensation by Compound Basis Function Neural Network and Hidden Markov Model

Hongmei Chen

College of of Mechatronics and Vehicle Engineering

Weifang University

Weifang261061, Shandong P.R., China

chmwfu@163.com

Xinggao Liu

State Key Laboratory of Industrial Control Technology, Dept. of Control Science and Engineering

Zhejiang University

Hangzhou 310027, Zhejiang P.R., China

liuxg@iipc.zju.edu.cn

**Abstract**—To estimate the Melt Index(MI) value accurately and quickly in the quality control process of polypropylene (PP), the paper proposed a forecast model of MI (PHKT-GRBFNN-CRBFNN-HMM) integrating the technologies of data mining, model constructing and two error compensation based on the data. Firstly applied the integrating data processing algorithm including PCA, Holt Exponential Smoothing, Kernel Density Estimation and Time-Variable scale Weighting to mine the data deeply to extract the useful information of the data; Then constructed the MI NARMA prediction model based on Gaussian Radial Basis Function Neural Network and Compound RBFNN on the basis of the data mining; Due to Markov property of the error sequence, used Hidden Markov Model to predict the error as the second compensation for the MI prediction values. The proposed model has been checked based on a real plant history data and the MRE(%), RMSE, STD and TIC of the generalization database is respectively 1.40, 0.045, 0.0457 and 0.0088. The results indicate that the proposed model has better comprehensive characteristics and is worth popularization and application in the PP industry process.

**Keywords:** *integrating data processing; compound basis function Neural Network; HMM; exponential smoothing; melt index prediction Of polypropylene*

## I. INTRODUCTION

Propylene polymerization (PP) process is a typical nonlinear process which is influenced by a variety of factors. Different grades of PP product determine different properties and uses. Melt Index (MI) is one of the most important parameters which judges the grade of product and determines the control strategy of process parameters. At present, most chemical enterprises still use traditional method of manual sampling and off-line analyzing data to get the analytical value of MI between which and actual production process there is a big lag(2-4h)[1]. It can't guide industrial production real-time with the lagged MI. So it is important to measure MI in time and accurately. By digging the inherent laws of PP MI analytical value and historical data of its process variables and establishing the model between them to predict the MI value has become a popular way. How to improve the accuracy and generalization performance of the model is the task of our study.

Many scholars have researched the statistical modeling by RBF neural network(NN). But most people focus on network architecture improvements[2-6] or algorithms optimization[7-12],etc. This paper mainly introduces the basis function improvement of RBFNN for error compensation and its modelling, besides, the methods of data processing before modeling and the second error compensation by HMM are also the exploratory topics in the text. The proposed model of applying the above technologies achieves better accuracy and generalization performance in the MI prediction by test.

## II. MODELING PROCESS DESCRIPTION

### A. Brief introduction to Propylene polymerization process

The PP process with continuous stirred tank reactor is shown in Fig .1 in brief[13]. The PP MI is affected by many factors which are called process variables referring to propylene polymerization mechanism and the subjacent diagram. Among there are nine variables which are commonly used to estimate the MI value[14]. They are primary and auxiliary catalyst feeding flow rate, three streams of propylene feeding flow rate, temperature, pressure, liquid level and the volume concentration of hydrogen in reactor. Measure these process variables and build a precise model between them and the analytical value of MI from the laboratory, the MI value can be accurately predicted on-line by the model with the measurements of process variables. This helps to effectively control these process variables to meet the production requirement and greatly shorten the production cycle.

In this article, a total of 85 groups of time series data are extracted from the same petrochemical enterprise to be used to model, including nine process variables' measurements obtained from the DCS database of actual PP production and the MI analyses got from the laboratory. Among them 70 groups of data ahead belong to the same batch consisting of train set (50 groups ahead) and test set(20 groups behind) chronological, the remaining 15 groups of data attribute to another batch used as generalization set.

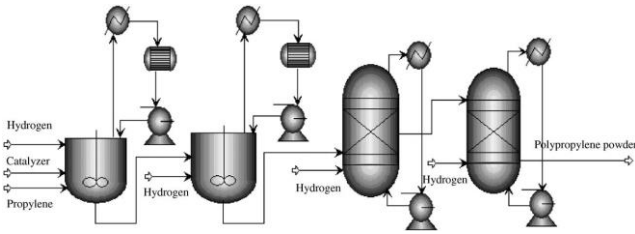


Figure 1. flow chart of polypropylene production

### B. Integrating data processing before modeling

The measurement values of PP process variables from the real plant are multidimensional, correlative each other and contain a amount of information. Thus it is vital to deal with data with some methods before modeling. Before this, the measurements of the nine process variables are firstly standardized to be zero-mean and one-standard-deviation.

#### 1) Principal Component Analysis(PCA)

PCA is one of the methods of de-correlation and dimensionality reduction[15-16], and is applied widely in many fields. In the text, if  $X$  denotes the standardized process variables, uses PCA to simplify the  $X$  firstly(see Eq. 1).

$$X = TP^T \approx t_1 p_1^T + t_2 p_2^T + \dots + t_{m_0} p_{m_0}^T. \quad (1)$$

Where,  $T$  is score matrix,  $T = [t_1, t_2, \dots, t_{m_0}]$ ;  $P$  is load matrix,  $P = [p_1, p_2, \dots, p_{m_0}]$ ; the superscript  $T$  is the transpose symbol; the subscript  $m_0$  is the number of principal components, which generally is selected the minimal number of the principal components that their accumulative contribution rate is greater than 85 percent. Here,  $m_0$  takes 5 by testing.

#### 2) Holt-Winters Exponential Smoothing(HES)

Drawing each principal component data sequence of the PP process variables in figure, it is easy to find these data sequences with slow trends and periodicities. For improvement the accuracy and generalization ability of the MI prediction model, HES[17-18] is led into these principal components data sequences: Firstly, the horizontal items  $S$ , trend items  $B$  and periodic items  $C$  are respectively updated by following the principles of using the weighted average of past values to predict the future value, and the weights are assigned exponentially-decreasing according to the distance to the current time from near to far(see Eq. 2); Then these principal components data sequence are smoothed with the trend items  $B$  and periodic items  $C$  be got rid of the data sequence. To prevent the data series over-smoothing, the trend items  $B$  and periodic items  $C$  should be removed in certain proportion from these principal components data sequence (see Eq. 3).

$$\begin{aligned} S_t &= \alpha X_t / C_{t-l} + (1-\alpha) (S_{t-1} + B_{t-1}) \\ B_t &= \beta (S_t - S_{t-1}) + (1-\beta) B_{t-1} \\ C_t &= \gamma X_t / S_t + (1-\gamma) C_{t-l} \end{aligned} \quad (2)$$

Where:  $X$  denotes the principal components data sequence of the PP process variables after PCA;  $S$  is a horizontal items;  $B$  is a trend items;  $C$  signifies the periodic items;  $\alpha, \beta, \gamma$  are respectively smooth coefficients of the horizontal items, trend items and periodic items which values are between 0 and 1; the subscript  $l$  is the period of periodic items in  $X$ ; the subscript  $t$  represents the  $t$  moment.

$$X_t = X_t - m_1 B_t - m_2 C_t. \quad (3)$$

Where:  $m_1, m_2$  are respectively the removal proportion factors of trend items and periodic items.

The smooth coefficients( $\alpha, \beta, \gamma$ ) affect the smoothing effect, which selections have no universal method at present. In the paper they are determined by enumeration method under the following guidelines: the more marked the data sequence feature including horizontality, trend and periodicity is, the greater the corresponding coefficient is; conversely, the smaller. Moreover, the initial values of the three feature items( $S_0, B_0, C_0$ )also have certain effect on the smooth effect. Here they are selected by experience:  $S_0$  is taken for the mean of  $X$ ,  $B_0$  is taken for the average value of the differences of the corresponding sample values between the first two period of  $X$  and  $C_0$  is taken as 0.6. Besides, the other factors in HES are respectively chosen by test according to the principle of minimum error. After HES the smoothness of  $X$  data sequences is improved obviously than before.

#### 3) Kernel Density Estimation(KDE)

In order to obtain more reasonable distribution properties of the  $X$  data sequence after HES, Kernel Density Estimation method[19-20] is led into the  $X$  data sequence to re-estimate their distributional rule only taking advantage of the data themselves. Here the work is accomplished by 'ksdensity' of Matlab function and uses Gaussian kernel. Finished this step the  $X$  data sequence spread more logically and reliably to benefit to improve the model generalization.

#### 4) Time-Variable scale Weighting(TVW)

The time series of the PP process variables have both time scale characteristic and variable dimension feature. Usually the data column at different moment plays different role in the modeling of multi-input single-output system, and the role of each dimension variable is also different. So the data had better be given the weights respectively in time scale and variable dimension. Take the following function(see Eq. 4) as a weight function and the  $X$  data sequence after KDE is weighted based on the Euclidean distance to the space origin in time scale.

$$\psi(d_i) = \exp(- (2^a d_i - b) / T_0). \quad (4)$$

Where,  $d_i$  is the Euclidean distance of the  $X$  data column to the space origin at  $t$  moment;  $a, b$  is respectively the scale factor and the shifting factor of  $d_i$ , which are introduced to enhance the weighting effect and their values are both positive;  $T_0$  is the width factor of the weight function. These three factors are determined by the  $X$  data sequence in accordance with the rule of minimal error.

In order to reasonably weight in the dimension of the  $X$  data, each dimension variable makes a Gray Correlation Analysis (GCA)[21-22] with target vector  $y$  on the train dataset (see Eq. 5). The obtained gray correlation coefficient reflects the relevant degree of each dimension variable to the goal vector value at every time, which is normalized as Eq. 6 to get the relative gray correlation coefficients used for the weight of the value of each dimension variable of  $X$  at each time.

$$L_i(n) = \frac{\min_n \min_i |x_i(n) - y(n)| + \rho \max_n \max_i |x_i(n) - y(n)|}{|x_i(n) - y(n)| + \rho \max_n \max_i |x_i(n) - y(n)|}. \quad (5)$$

Where,  $\rho$  is the resolution factor( $\rho \in (0,1)$ ), which is selected as 0.092 here;  $i$  is the  $i$ th dimension;  $n$  is the  $n$ th sample on the timeline.

$$Q_i(n) = \sum_{i=1}^k L_i(n). \quad (6)$$

Different variables of the X data should be given different weights according to their contribution to the prediction precision, and the weights are ascertained in the following manner: establishing the linear regression model of the aim variable y and each dimension variable of the X data on the train database abiding by the minimal error criterion, then calculating the reciprocal of each regression error, finally normalizing each error reciprocal as the weight of each dimension variable.

It is worth mentioning that the above obtained weights of the X data on time scale and variable dimension need to be adjusted to fluctuate around one within a tiny scope if necessary.

### C. model architecture

Real industrial PP process is complicated, full of strong nonlinearity, large delay and unknown interference[23]. Therefore, adopts the nonlinear model to predict the MI value in this paper. Given a multidimensional input variable  $x(t) \in R^v \times 1$ , the approximation of output variable  $y(t)$  is gained by the following Nonlinear Autoregressive (NAR) model (see Eq. 7) [24-27]:

$$y = f_1(x, y) + f_2(x, y)x. \quad (7)$$

Where,  $f_1(\bullet)$  and  $f_2(\bullet)$  are two different nonlinear continuous functions involving input x and output y.

Owing to NAR model only having low-order terms, the approximation of output  $y(t)$  by NAR model is bound to lower precision. If an error term is added to the Eq. 1 for compensation, Nonlinear Autoregressive Moving Average (NARMA) model[24-27] comes into being (see Eq. 8).

$$y = f_1(x, y) + f_2(x, y)x + \varepsilon(x, y). \quad (8)$$

Where,  $\varepsilon(\bullet)$  is the high-order error term involving input x and output y.

Because NARMA model has higher precision than NAR model, NARMA model is selected as the prediction model of the PP MI in this paper. To improve the performance of the model, the strategy of model predictive control is introduced to the modeling process. At the current time k, taking  $U[k]$  ( $U[k] = [y[k], y[k-1], x[k-1]]^T$ ) as input, the MI output of the  $[k+d]$ th time will be gotten by Eq. 9.

$$\hat{y}[k+d] = f_1(y[k], y[k-1], x[k-1]) + f_2(y[k], y[k-1], x[k-1])x[k] + \varepsilon(y[k], y[k-1], x[k-1]) \quad (9)$$

Where,  $d_r$  represents the relative order of system which is an index of characterizing time-delay and its value takes 2 here by verification.

### D. RBF Neural Network

#### 1) The prediction model by Gaussian RBFNN

RBFNN is a multilayer feed-forward NN, of which three-layer network has been widely used in many engineering fields. The three layers are generally input layer, hidden layer and output layer. Here employs three-layers-RBFNN to approximate the nonlinear functions in the MI NARMA model because of their better approximation characteristic, faster convergence rate, higher precision level, etc[28]. Corresponding to the three functions  $f_1$ ,  $f_2$  and  $\varepsilon$  in the model, three different RBFNNs with the common input  $U[k]$  are built successively to respectively approach them(see Eq. 10).

$$\begin{cases} \hat{f}_1(U[k]) = \sum_{j=1}^{N1} \omega_{1j} \varphi_{1j}(\|U[k] - c_{1j}\|) \\ \hat{f}_2(U[k]) = \sum_{j=1}^{N2} \omega_{2j} \varphi_{2j}(\|U[k] - c_{2j}\|) \\ \hat{\varepsilon}(U[k]) = \sum_{j=1}^{N3} \omega_{3j} \varphi_{3j}(\|U[k] - c_{3j}\|) \end{cases} \quad (10)$$

Where,  $\varphi(\cdot)$  is the selected radial basis function;  $w_j$  is the weight of the  $j$ th hidden-layer-node output;  $c_j$  is the  $j$ th center vector of the hidden layer;  $\|\cdot\|$  represents the euclidean distance between the input  $U[k]$  and the center  $c_j$ ;  $N$  represents the node number of the hidden layer.

The radial basis function  $\varphi(\cdot)$  may adopt many forms, which influences the network's approximation capability. Here the most frequently used Gaussian radial basic functions ( $\varphi_g(r)$ ) is employed as the mapping function from *input* to *output* (see Eq. 11).

$$\varphi_1(r) = \varphi_2(r) = \varphi_g(r) = \exp\left(-\frac{r^2}{2\delta^2}\right). \quad (11)$$

Where,  $\delta$  represents the breadth of the Gaussian radial basis function.

Owing to the sample size is not larger in the paper, this text picks fewer hidden node number for RBFNNs of predicting the MI to prevent over-fitting and the networks parameters are initially determined on the basis of unsupervised method in the training. To acquire the desired predicting performance, the centers, width and the weights of the two approximation RBFNNs are carried out collaborative optimization with gradient descent algorithm

and adaptive-changing learning rate in the model training.

#### 2) The error compensation by non-gaussian RBFNN

Because PP process is highly nonlinear and non-gaussian, the RBFNNs for predicting the MI take Gaussian basis function, and the RBFNN for error compensation adopts the non-gauss

basis functions( $\varphi_3(\cdot)$ ), e.g. Wavelet function, Thin-Plate-Spline function, Descending-Cosine function (see Eq. 12 to 14), etc[28]. The complementary NN system consisting of Gauss basis function and non-gauss basis function is useful to improve the forecasting precision and generalization performance of the model. Those functions mentioned in Eq. 12 to 14 are freely assembled or separately combined with quadratic polynomial function to form the convergent compound basis functions, like Gaussian-Wavelet, Gaussian-polynomial function (see Eq. 15 and 16), and so on. The error compensation NN based on the formed compound basis functions play a better compensatory effect on the MI prediction through verification. Moreover, the mixed basis function error networks having different types of basis functions at different hidden nodes can yet achieve good error compensation characteristics by test. In short, the basis

function( $\varphi_3(\cdot)$ ) of RBFNN for approximating the error item( $\varepsilon(\cdot)$ ) in the MI prediction may choose the arbitrary function forms in the Eq. 12 to 16, and the error networks based on these non-gauss basis functions would acquire better compensation effect on the MI prediction as long as their relevant parameters adjusted to appropriate values based on the principle of minimum MSE.

a) *Wavelet function*

$$\varphi_w(U) = \exp\left(-\frac{\left\|2^{\frac{a}{2}}(2^a U - b) - c_j\right\|^2}{2\delta^2}\right)$$

$$\text{or } \varphi_w(r) = \exp\left(-\frac{[2^{\frac{a}{2}}(2^a r - b)]^2}{2\delta^2}\right). \quad (12)$$

Where,  $a$  is the stretch factor;  $b$  is the shift factor.

b) *Thin-Plate-Spline function*

$$\varphi_s(r) = r^2 \log(cr^2). \quad (13)$$

Where,  $c$  is the coefficient of the thin-plate-spline function.

c) *Descending-Cosine function*

$$\varphi_{dc}(r) = \exp(-c_1 r) \cdot \cos(c_2 r + \alpha). \quad (14)$$

Where,  $c_1$  and  $c_2$  are both the coefficients of the descending-cosine function;  $\alpha$  is the initial phase angle of the cosine function.

d) *Gaussian-Wavelet Compound function*

$$\psi(\theta) = 2^{\frac{a_w}{2}} (2^{a_w} \theta - b_w) = 2^{\frac{a_w}{2}} (2^{a_w} \varphi_g(r) - b_w). \quad (15)$$

Where,  $a_w$  is the stretch factor;  $b_w$  is the shift factor.

e) *Gaussian-Polynomial Compound function*

$$\psi(\theta) = c_0 + c_1 \theta + c_2 \theta^2 = c_0 + c_1 \varphi_g(r) + c_2 \varphi_g^2(r). \quad (16)$$

Where,  $c_0, c_1$  and  $c_2$  are the polynomial coefficients.

During the network training of error compensation, uses the regularized cost function to simultaneously ensure small training error and simple network structure.

#### E. The second error compensation by HMM

HMM is embedded two random processes, of which one is implicit state transition sequence corresponding to a discrete Markov process, another is the observable sequence[29-30]. The parameters of a simplified HMM are commonly expressed as  $\lambda = \{A, B, \pi\}$ , among which  $\pi$  represents hidden state initial probability matrix and  $A$  is hidden state transition probability matrix,  $B$  is observed state radiation probability matrix. HMM can precisely infer the observation sequence output tendency by double analyzing of hidden state operating mechanism and observed state probability, so it is able to predict time series and there are also successful examples in engineering field[31-32].

The error is generally thought to have Markov property owing to having three cases of: positive, negative and approximate zero, so the hidden states number of HMM is naturally taken for three. Here uses HMM with three hidden states to estimate the future values of the deviation between the analytic MI values and the MI prediction values after the first compensation based on compound RBFNN as the second compensation of the MI prediction. Through this step the accuracy of the MI prediction values is improved again by testing.

The predicting modeling progress via HMM describes as follows[29-32]:

Step 1: Define the initial parameters of the HMM model:  $\lambda = \{A, B, \pi\}$ . (1) Suppose hidden state set:  $S = \{S_1, \dots, S_Q\}$ ,  $Q$  is hidden state number:  $Q=3$  here. (2) Set the initial value of implicit state transition probability distribution at random:  $A =$

$\{a_{ij}\}$ ,  $a_{ij} = P\{qt+1=s_j \mid qt=si\}$ ,  $1 \leq i, j \leq Q$ , and  $a_{ij}$  must meet the

$$\sum_{j=1}^Q a_{ij} = 1$$

following conditions:  $a_{ij} \geq 0$  and (3) If  $V$  is the sample space of observed variables, set the initial value of output observed variables probability distribution:  $B = \{b_i(v), 1 \leq i \leq Q, v \in V\}$ ,  $b_i(v) = P\{O_t = v \mid qt=si\}$ . Among,  $O_t$  is observation random variable at  $t$  time. This paper considers the stochastic deviation sequence after compound RBFNN error compensation as observation sequence written as  $O = (e_1, \dots, e_{T1})$ . Among,  $T1$  represents the observation sequence length. (4) Set randomly the initial state probability distribution:

$$\pi = \{\pi_i, 1 \leq i \leq Q\}, \quad \sum_{i=1}^Q \pi_i = 1$$

Step 2: Derive the  $t$ th time forward probability variables:

$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = S_i \mid \lambda)$ . (1) Initialize the parameter:

$\alpha_1(i) = \pi_i b_i(o_1)$ ,  $1 \leq i \leq Q$ . (2) Deduce  $\alpha_{t+1}(j)$ :

$\alpha_{t+1}(j) = \sum_{i=1}^Q \alpha_t(i) a_{ij} b_j(o_{t+1})$ ,  $1 \leq j \leq Q$ ,  $1 \leq t \leq T1-1$ . (3) Calculate the output probability of the observation sequence with  $\lambda$ :

$$P(O \mid \lambda) = \sum_{i=1}^Q \alpha_{T1}(i)$$

Step 3: Derive the  $t$ th time backward probability variables:

$\beta_t(i) = P(o_t, o_{t+1}, o_{t+2}, \dots, o_{T1} \mid q_t = S_i, \lambda)$ . (1) Initialize the parameter:  $\beta_{T1}(i) = 1$ ,  $1 \leq i \leq Q$ . (2) Deduce  $\beta_t(i)$  at the  $t$ th time

from back to front:  $\beta_t(i) = \sum_{j=1}^Q \alpha_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$ ,  $1 \leq j \leq Q$ ,  $t = T1-1, T1-2, \dots, 1$ .

Step 4: Optimize the HMM model parameters using Baum-Welch algorithm. (1) Derive the probability of  $S_i$  state at  $t$  time:

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^Q \alpha_t(j) \beta_t(j)}, \quad 1 \leq i \leq Q.$$

(2) Derive the probability of  $S_i$  state at  $t$  time and  $S_j$  state at  $(t+1)$  time:

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(O \mid \lambda)} = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{m=1}^Q \sum_{n=1}^Q \alpha_t(m) a_{mn} b_n(o_{t+1}) \beta_{t+1}(n)}$$

(3) HMM parameters can be re-estimated using the following formula: (a) The state probability distribution is re-

estimated:  $\bar{\pi}_i = \gamma_1(i)$ ,  $1 \leq i \leq Q$ . (b) The state transition

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T1-1} \xi_t(i, j)}{\sum_{t=1}^{T1-1} \gamma_t(i)}$$

probability distribution is re-estimated:

$1 \leq i, j \leq Q$ . (c) The weight of the observed value in  $S_i$  state at  $t$

$$\bar{v}_t(i) = \frac{\sum_{m=1}^M \zeta_t(i, m)}{\sum_{i=1}^Q \sum_{m=1}^M \zeta_t(i, m)}$$

time is re-estimated:  $1 \leq i \leq Q$ . Where,

$\varsigma_t(i, m)$  is the probabilities of the system which is in  $S_i$  state at  $t$  time and corresponds to the  $m$ th density estimate in the Gaussian kernel density estimations.

$$\varsigma_t(i, m) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^Q \alpha_t(i)\beta_t(i)} \frac{f_m(v_t(i) \cdot o_t)}{\sum_{m=1}^M f_m(v_t(i) \cdot o_t)}, \quad 1 \leq i \leq Q, 1 \leq m \leq M.$$

Among,  $f_m(\cdot)$  is the  $m$ th density estimation of the Gaussian kernel density estimates for the observed values in  $S_i$  state at  $t$  time. (d) The observation sequence is re-estimated:  $\bar{O}_t = \{v_t(i) \cdot O_t\}$ ,  $1 \leq t \leq T_1$ ,  $1 \leq i \leq Q$ . (e) The model parameters are updated:  $\bar{\lambda} = \{\bar{A}, \bar{B}, \bar{\pi}\} = \{\bar{a}_{ij}, \bar{O}_t, \bar{\pi}_i\}$ .

Step5: Go back to step 2 to re-execute the steps and exit the loop while the Step2.3 value meets certain requirements or the iterations number attains maximum, then remember the corresponding model parameters, denoted as  $\lambda_{opt} = \{a_{ij_{opt}}, O_{t_{opt}}, \pi_{i_{opt}}\}$ . Assign  $\lambda_{opt}$  to  $\lambda$ :  $\lambda = \lambda_{opt}$ ,  $a_{ij} = a_{ij_{opt}}$ ,  $O_t = O_{t_{opt}}$ ,  $\pi_i = \pi_{i_{opt}}$ .

Step6: Determine the path of hidden states corresponding to the maximum output probability of the observation sequence on the train database based on Viterbi algorithm. (1) define two functions  $\delta_t(i)$  and  $\psi_t(j)$ :  $\delta_t(i)$  represents the maximum output probability of the observation sequence in  $i$  state at  $t$  time:  $\delta_t(i) = \max_{q_1, \dots, q_{t-1}} P(q_1, \dots, q_t = i, o_1, \dots, o_t | \lambda)$ ;  $\psi_t(j)$  is a function of storing the optimal state sequence which is corresponding to the maximum output probability of the observation sequence at  $t$  time. (2) Initialize the variable:

$$\begin{cases} \delta_1(i) = \pi_i b_i(o_1), & 1 \leq i \leq Q \\ \psi_1(i) = 0, & 1 \leq i \leq Q \end{cases}; \quad (3) \text{ Derive the variable:}$$

$$\begin{cases} \delta_t(j) = \max_{1 \leq i \leq Q} [\delta_{t-1}(i) a_{ij}] b_j(o_t), & 2 \leq t \leq T_1, 1 \leq j \leq Q \\ \psi_t(j) = \arg \max_{1 \leq i \leq Q} [\delta_{t-1}(i) a_{ij}], & 2 \leq t \leq T_1, 1 \leq j \leq Q \end{cases};$$

$$(4) \text{ Terminate the derivation: } \begin{cases} P^* = \max_{1 \leq i \leq Q} [\delta_{T_1}(i)] \\ q_{T_1}^* = \arg \max_{1 \leq i \leq Q} [\delta_{T_1}(i)] \end{cases}; \quad (5)$$

Derive inversely to get the optimal hidden state sequence of the deviations after compound RBFNN error compensation on the train database and store them in  $\psi_t(j)$  function:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), t = T_1 - 1, \dots, 1.$$

Step7: According to the best hidden state sequence of the deviations on the train database, predicts the deviations

sequences on test database and generalization database by HMM.

### III. MELT INDEX PREDICTION

#### A. Error indicators

In order to measure the performance of the above proposed PP MI prediction models, this paper adopts four conventional indicators to quantize the differences between the MI predicted values and the MI analytic values from the laboratory: Mean Relative Error (MRE), Root of Mean Square Error (RMSE), error Standard Deviation (STD) and Theil's Inequality Coefficient (TIC). Among, the first two indicators are used to characterize the accuracy of the model, STD measures the stability of the model, and TIC measures the consistency of the MI predicted value and the MI analytic values. Define the parameters as follows:

$$\text{MRE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (i=1, 2, \dots, N). \quad (17)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (i=1, 2, \dots, N). \quad (18)$$

$$\text{STD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (e_i - \bar{e})^2} \quad (i=1, 2, \dots, N). \quad (19)$$

$$\text{TIC} = \frac{\sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^N y_i^2}} \quad (i=1, 2, \dots, N). \quad (20)$$

Among,  $y_i$  denotes the laboratory analysis value of MI;  $\hat{y}_i$  represents the MI prediction value.

#### B. Results and discussions

On the basis of different data processing methods (HES, KDE and TVW) and error compensation by HMM and RBF NN, several various MI prediction models are established in this paper. To compare the effect of the different data processing method and the integrating data processing means on the MI forecasting, the results of the prediction models are listed in TABLE I in the form of the above four error indicators, and the TABLE II offers the performance comparisons of the models with various basis function compensation network in the premise of HES\_KDE\_TVW data processing and the HMM error compensation.

The results of the TABLE I show that: (1) the prediction values of the models using data processing methods are more accurately than the predicting value of the model without data processing; (2) the prediction precisions of the models using multiple data processing technologies are higher than the models using single data processing technology; (3) the prediction values of the models that their error compensation networks have compound basis function have better generalization ability than the predicting value of the models with Gaussian basis function compensation networks; (4) the model with error compensation by HMM improves the accuracy of the prediction result further; (5) the RBFNN compensatory model with a

complex basis function is better than the model with a single function as the basis function on the matching effect and a variety of data processing technologies, so the prediction value of the model (PHKT\_GRBFNN\_CRBFNN\_HMM) with compound basis function NN and HMM as error compensation on the basis of HES\_KDE\_TVW data processing has both higher prediction precision and generalized ability.

From the TABLE II it is seen that:(1) the compensatory models with non-gauss basis function have better prediction performance of precision and generalization;(2) in all the models of the compensatory networks with non-gauss basis functions, the model with thin-plate-spline basis function(PHKTRTSRH) has best prediction accuracy and better generalized property, the model with wavelet basis function(PHKTRWRH) has higher prediction precision and lower generalization, and the models with compound basis functions including Gaussian-Polynomial function(PHKTRCR-GPH) and Gaussian-Wavelet function (PHKTRCR-GW H) have best generalization ability and slightly lower prediction precision, the model with descending-cosine basis function(PHKTRDRH) has the same predicting performance as the models with compound basis functions.

For simple and clear, the models in the graphs express by simplified form as follows: PRR denotes PCA\_GRBF\_GRBF, PHKTRR denotes PCA\_HKT\_GRBF\_GRBF, PHKTRCR represents PCA\_HKT\_GRBF\_CRBF, and PHKTRCRH represents PCA\_HKT\_GRBF\_CRBF\_HMM. The predictive results of the above four models are displayed on the Matlab graphs (see Fig .2 and Fig .3). The predicted results of the models which basis functions of the compensatory networks are non-gauss functions are shown in Fig .4 and Fig .5. It is easily to see that the PP MI prediction values got from the models in the graphs can more accurately approximate the MI analysis values from the laboratory in the overall trend whether on the test database or on the generalization database.

TABLE I . performance comparison of MI prediction models

Models	Test database				Generalization database			
	MRE (%)	RMSE	STD	TIC	MRE (%)	RMSE	STD	TIC
PRCR	5.14	0.1540	0.1465	0.0648	0.93	0.0319	0.0313	0.0125
PRCRH	5.14	0.1535	0.1454	0.0645	0.94	0.0322	0.0313	0.0126
PKRCR	4.56	0.1454	0.1451	0.0611	2.54	0.0705	0.0574	0.0276
PRR	4.75	0.1449	0.1472	0.0304	1.51	0.0442	0.0314	0.0087
PKRCRH	4.56	0.1442	0.1434	0.0606	2.56	0.0712	0.0574	0.0278
PRRH	4.75	0.1442	0.1464	0.0606	1.49	0.0437	0.0314	0.0171
PKRR	4.47	0.1426	0.1463	0.0600	1.66	0.0558	0.0577	0.0218
PTRCR	4.46	0.1414	0.1390	0.0595	7.10	0.1928	0.0684	0.0754
PKRRH	4.47	0.1412	0.1448	0.0594	1.68	0.0558	0.0577	0.0218
PTRCRH	4.46	0.1410	0.1385	0.0593	7.09	0.1925	0.0686	0.0753
PTRR	4.44	0.1379	0.1390	0.0580	7.67	0.2072	0.0684	0.0810
PTRRH	4.44	0.1376	0.1386	0.0579	7.65	0.2070	0.0685	0.0809
PHRCR	4.47	0.1328	0.1262	0.0559	1.00	0.0341	0.0337	0.0134

PHRCRH	4.47	0.1323	0.1251	0.0556	1.02	0.0344	0.0337	0.0135
PHRR	4.11	0.1244	0.1270	0.0523	1.54	0.0458	0.0339	0.0179
PHRRH	4.11	0.1235	0.1260	0.0519	1.51	0.0453	0.0339	0.0177
PHKTRR	3.72	0.1123	0.1111	0.0238	2.64	0.0794	0.0461	0.0157
PHKTRRH	3.71	0.1109	0.1101	0.0466	2.61	0.0787	0.0461	0.0308
PHKTRCR	3.48	0.1067	0.1084	0.0224	1.41	0.0452	0.0456	0.0089
PHKTRCRH	3.47	0.1057	0.1071	0.0222	1.40	0.0450	0.0457	0.0088

TABLE II . performance comparison of various basis function models

Models	Test database				Generalization database			
	MRE (%)	RMSE	STD	TIC	MRE (%)	RMSE	STD	TIC
PHKTRCR <sub>GP</sub> H	3.47	0.1057	0.1071	0.0222	1.40	0.0450	0.0457	0.0088
PHKTRCR <sub>w</sub> H	3.47	0.1057	0.1071	0.0222	1.40	0.0450	0.0457	0.0088
PHKTRDRH	3.47	0.1057	0.1071	0.0222	1.40	0.0450	0.0457	0.0088
PHKTRWRH	3.50	0.1051	0.1073	0.0222	1.76	0.0571	0.0459	0.0112
PHKTRTSRH	3.46	0.1046	0.1072	0.0220	1.66	0.0535	0.0459	0.0105

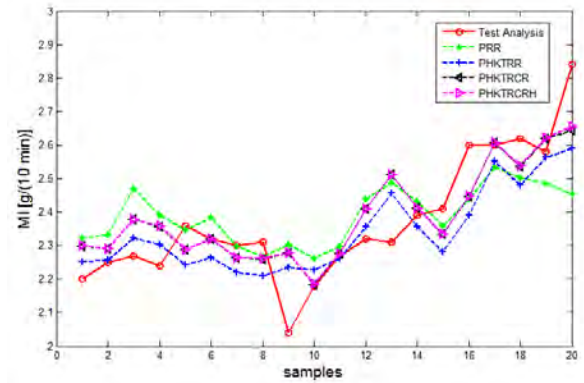


Figure 2. Performances of models on test database

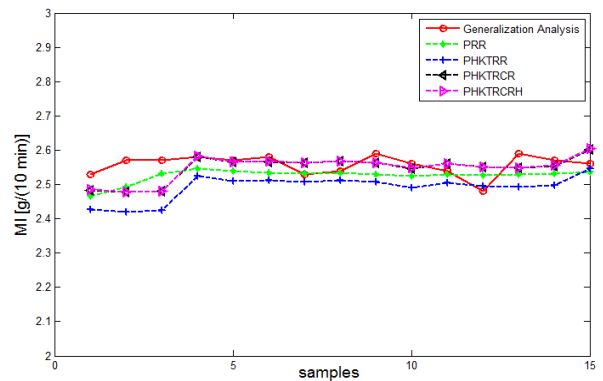


Figure 3. Performances of models on generalization database

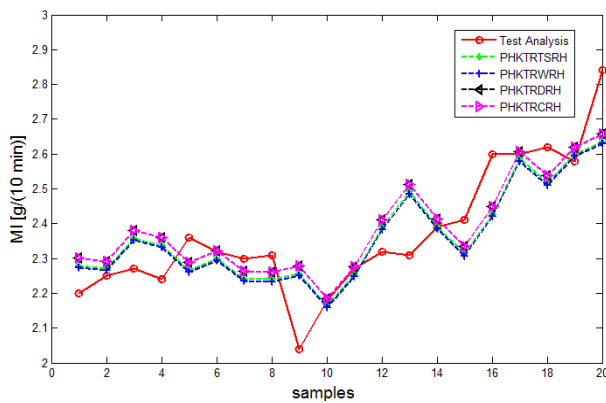


Figure 4. Performances of various basis function models on test database

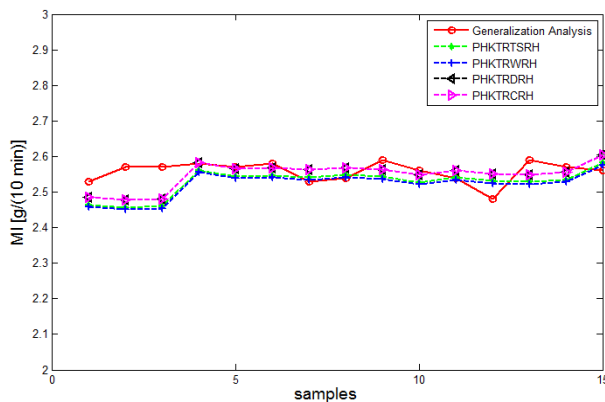


Figure 5. Performances of various basis function models on generalization database

#### IV. CONCLUSION

This article applies technologies of data processing including PCA, Holt-Winters exponential smoothing, kernel density estimation and data weighting on the time scale and variable dimension, model constructing of RBF neural network and two kinds of error compensations to build the forecast models of PP MI. Compared with other PP MI forecasting models, the models proposed in this paper have the following advantages: (1) HES method help to get rid of the trend term and periodic item from the input data to restore the authentic data sequence; Gaussian KDE method is helpful to correctly estimate the data sequence distribution; TVW method is good for highlighting the predominant variables and superior data in the database. Through data processing of the integrating technology, the data sequence for modeling, i.e. the data sequence of PP process variables, becomes closer to the real data distribution to make the prediction results of the models more believable; (2) the integrating data-processing algorithm run the shorter time and the result is more stable contrasted with the usual intelligence swarm algorithm; (3) the error sequence which is predicted by the compound basis function RBFNN or the other non-gauss RBFNN has more general nonlinear characteristics, so it is more accordance with the real error of the PP MI in the industry process; (4) the error sequence of MI that is inferred by HMM would be more reasonable as the second compensation for the MI prediction values because it is obtained

based on the probability distributions of the MI real error; (5) lastly it has been proven that the proposed models of integrating multifold data-processing technologies and two error compensations show better overall performance in precision, generalization, reliability, efficiency, etc.

#### ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China(Grant U1162130), Shandong province high school teachers of home visiting scholar project funding, and their supports are thereby acknowledged.

#### REFERENCES

- [1]Bafna S S,Beall A M.A design of experiments study on the factors affecting variability in the melt index measurement [J]. Journal of Applied Polymer Science,Vol.65, No.2,1997, pp.277-288.
- [2]Felipe Miguel Aparicio Acosta.Radial basis function and related models: An overview[J]. Signal Processing,Vol.45,1995,pp.37-58.
- [3]Liu Fang,Mao Zhizhong,Li Lei.Process industrial time series prediction method based on improved RBF network[J]. Journal of Northeastern University(Natural Science),Vol.31,No.12,2010,pp.1693-1696.
- [4] Ying Zhang,Meng Xin Li,Ran Ran Wang.Utilization of Improved RBF Network in Modeling of Wave Energy Absorption System[J]. Applied Mechanics and Materials,Vol.483,No.12,2013,pp.578-582.
- [5]WEI Yan-ling.Application Research on An Improved RBF Neural Network in Pests Forecast[J]. Science Technology and Engineering, Vol.13,No.1,2013,pp.136-139.
- [6] LI Quanshan,ZHANG Yishan,CAO Liulun,et al. Improved dual model structure of RBF neural network and its application[J]. CIESC Journal,Vol.62,No.8,2011,pp.2345-2349.
- [7] TAN Qian,JIANG Yi,LIN Fan. Load forecasting of Swift cloud storage based on AHP-RBF [J]. Computer Engineering and Applications, Vol.50,No.8,2014, pp.35-39.
- [8] Chengye Zhao1Xinggao Liu1Feng Ding2. Melt Index Prediction Based on Adaptive Particle Swarm Optimization Algorithm-Optimized Radial Basis Function Neural Networks[J]. Chem. Eng. Technol,Vol.33,No.11,2010,pp. 1909-1916.
- [9] Jiubao Li,XinggaoLiu.Melt index prediction by RBF neural network optimized with an MPSO-SA hybrid algorithm[J]. Neurocomputing, Vol.74,2011, pp.735-740.
- [10]Lou W,Liu XG.Melt index prediction of polypropylene by neural networks model based on PCA-GA-RBF[J]. J Petrochem Univ.,Vol.20,2007, pp.82-85.
- [11]ZHANG Zhimeng ,LI Jiubao ,LIU Xinggao. Melt index prediction of polypropylene based on a new ant colony optimization[J]. CIESC Journal, Vol.62,No.8,2011,pp.2270-2274.
- [12]W Kong, J Yang. Prediction of polypropylene melt index based on RBF networks[J]. J. Chem. Ind. Eng. (China),Vol.54,No.8,2003,pp.1160-1163.
- [13] Jian Shi, Xinggao Liu, Youxian Sun.Melt index prediction by neural networks based on independent component analysis and multi-scale analysis [J]. Neurocomputing,Vol.70,2006,pp.280-287.
- [14] Xinggao Liu and Chengye Zhao. Melt Index Prediction Based on Fuzzy Neural Networks and PSO Algorithm with Online Correction Strategy[J]. AIChE Journal,Vol.00,No.0,2011,pp.1-9.
- [15] Chu-dong TONG,Xue-feng YAN,Yu-xin MA.Statistical process monitoring based on improved principal component analysis and its application to chemical processes[J]. Journal of Zhejiang University-Science A(Applied Physics & Engineering),Vol.14,No.7,2013,pp.520-534.
- [16]SUN S G. Application multivariable statistical analysis [M] Beijing: Science Press, 2011.
- [17] Holt C C. Forecasting seasonals and trends by exponentially weighted moving averages[J]. International Journal of Forecasting,Vol.20,2004,pp.5-10.

- [18] Billah B, King M L, Snyder R D, et al. Exponential smoothing model selection for forecasting [J]. International Journal of Forecasting, Vol.22, 2006, pp.239-247.
- [19] ZHANG Baoqiang, CHEN Guoping, GUO Qintao. Static Frame Model Validation with Small Samples Solution Using Improved Kernel Density Estimation and Confidence Level Method[J]. Chinese Journal of Aeronautics, Vol.25, No.6, 2012, pp.879-886.
- [20] Xie zhonghua. MATLAB statistical analysis and applications: 40 Case Studies[M]. Beijing: Beijing University of Aeronautics and Astronautics Press, 2010.
- [21] M Luo, B T Kuhnell. Fault detection using grey relational grade analysis[J]. The Journal of Grey System, Vol.5, No.4, 1993, pp.319-328.
- [22] Su Bo, Liu Lu, Yang Fangting. Neural network model based on gray relational analysis[J]. Systems Engineering Theory and Practice, Vol.9, No.9, 2008, pp.98-101.
- [23] Funahashi K. On the approximate realization of continuous mapping by neural Networks[J]. Vol.2, 1989, pp.183-192.
- [24] Chen S, Billings S A. Representation of non-linear systems: the NARMAX model[J]. International Journal of Control, Vol.49, 1989, pp.1013-1032.
- [25] PANG S W, YU K P, ZOU J X. Nonlinear time-varying system identification based on varying NARMA model[J]. Engineering Mechanics, Vol.23, No.12, 2006, pp.25-30.
- [26] JIN Chao, WU Bo and HU Youmin. Wavelet Neural Network Based on NARMA-L2 Model for Prediction of Thermal Characteristics in a Feed System[J]. Chinese Journal of Mechanical Engineering, Vol.24, No.1, 2011, pp. 33-41.
- [27] DING Feng. System identification new theory [M] Beijing: Science Press, 2013, pp.69-72.
- [28] Zhong Luo, Rao Wenbi, Zou chengming. Artificial Neural Network and its integration application technology[M]. Beijing: Science Press, 2007.
- [29] Lei HE, Chang-fu ZONG, Chang WANG. Driving intention recognition and behaviour prediction based on a double-layer hidden Markov model[J]. Journal of Zhejiang University-Science C(Computers and Electronics), Vol.13, No.3, 2012, pp.208-217.
- [30] YUAN Li-chi. Improved hidden Markov model for speech recognition and POS tagging[J]. Journal of Central South University, Vol.19, No.2, 2012, pp.511-516.
- [31] González A M, Roque A M, González J G. Modeling and forecasting electricity prices with input/output hidden Markov models[J]. IEEE Transactions on Power Systems, Vol.20, No.1, 2005, pp.13-24.
- [32] Cui X D, Gong Y F. A study of variable-parameter Gaussian mixture hidden Markov modeling for noisy speech recognition[J]. IEEE Transactions on Audio, Speech and Language Processing, Vol.15, No.4, 2007, pp.1366-1376.



Hongmei Chen\*(1977-), female, lecturer, Master, her research topics include: multivariate statistical modeling and optimization, measurement and control technology; Tel:13696363899, Email:chmwfu@163.com, Address: No. 5147, Dongfeng East Street, Weifang City, Shandong P.R, China, post code:261061.  
\*: Corresponding author.



Xing gaoLiu( 1968-, male, professor, Doctor, Doctoral tutor, his main research direction include: multivariate statistical modeling, model identification, optimization and control of complex industrial process, and dynamic optimization. Tel:0571-87988336(F), Email: liuxg@iipc.zju.edu.cn.