2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Wiener Model Identification of Blast Furnace Ironmaking Process Based on Laguerre Filter and Linear Programming Support Vector Regression

Xia Xu, Changchun Hua, Yinggan Tang and Xinping Guan

Abstract— **As a highly complex multi-input and multi-output system, blast furnace plays an important role in industrial development. Although much research has been done in the past few decades, there still exist many problems, such as the modeling and control problems. In view of these reasons, this paper is concerned with developing a Wiener model to predict the silicon content of blast furnace. Unlike traditional Wiener model, this paper avoids the optimization of high number of model parameters. The Wiener model here is composed of a basis filter filter expansion named Laguerre filter and a linear programming support vector regression (LP-SVR). They are used to represent the linear dynamic component and the nonlinear static element. Take the advantages that Laguerre filter can approximate linear systems with a lower model and order and LP-SVR can achieve a sparse solution, the proposed Wiener model not only improves the prediction accuracy but also reduces the computation complexity. Simulation results show that this Wiener model is suitable for the prediction of blast furnace silicon content.**

I. Introduction

IRON and steel making, a typical high energy-consuming, high emitters and high pollution industry, is a pillar industry for our national economy. Among all the processes, blast furnace is the major source of sulfur dioxide emissions and energy consumption. It is a complex industrial reactor used to produce hot metal from iron ore for subsequent processing into steel. When a blast furnace runs, most chemical reactions happen among different phases including gas phase, liquid phase and solid phase, along with high temperature, high pressure [1]. These facts bring the reaction process large time delay and spatiotemporal characteristics, such as it will take about 6∼8 hours for a cycle of ironmaking [2]. Due to the complicated process and hostile environment, current study focuses on multifields, such as modeling, control method, performance optimizing and metallurgical analysis. The main purpose of all the efforts is to obtain a stable and efficient operation of blast furnace [3]. However, the control of blast furnace is still implemented by experienced foremen, who determine status of the reaction process through comprehensive analysis of numerous measured data. Therefore, modeling of blast furnace is still a crucial problem.

For past decades, a great deal of models including mechanism-based white-box models and data-based black-box models have been developed and most of them are used to predict the hot metal silicon content. Due to the white-box models fail to capture the dynamic disturbances in ironmaking process [4], data-based models have attracted more attention and obtained more achievements which aim at capturing the intricate interaction between the measured variables. The main data-driven models include neural networks [5-7], fuzzy logic [8], support vector regression [9-10], state space [11-12], partial least squares recursive analysis [13] and evolutionary networks [14-15]. They have been successfully applied to predict the silicon content. For data-driven models, the modeling process only needs a large set of observations but no need for the prior information about the process.

Wiener model can approximate any complex nonlinear system with short-term memory and it has also been applied to some practical systems. A sub-space based Wiener model has been applied to identify the hot metal model in [16], where the linear dynamic subsystem has a state space representation and the inverse of the nonlinear part is described as a linear combination of polynomial basis functions. Most Wiener models are based on the standard least square methods. The model output is a nonlinear function of numerous parameters after both the linear dynamic subsystem and nonlinear static subsystem are identified. Then the final task is to solve a nonlinear least square problem which may has local optima [17]. In order to avoid the high number of model parameters, in this paper, an orthonormal filter expansion—Laguerre filter is used as the linear part and a regularization term—linear programming support vector regression is used as the nonlinear part.

As a linear black-box modeling method, orthonormal filter expansions have achieved much attention in the past decades. In [18], the advantages of using orthonormal filter network for system identification are summarized. One is that a good approximation can be obtained with no much restriction. Another is that the modeling process does not need any explicit knowledge about system time constant and time delay. Later, this orthonormal filter method is combined with nonlinear map and forms the Wiener-Laguerre model which

Xia Xu, Changchun Hua, Yinggan Tang and Xinping Guan are with the Department of Institute of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei, China (email:xuxiajnz@gmail.com).

This paper is partially supported by the Science Fund for Hundred Excellent Innovation Talents Support Program of Hebei Province,Hebei Province Applied Basis Research Project (13961806D), and the National Natural Science Foun- dation of China (61290322, 61322303).

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

is used to model the nonlinear dynamic system.

Support vector machine (SVM) is a popular machine learning method. Its main idea is to embed the inputs into a feature space through a high dimension mapping, and then find an optimal decision hyperplane among the high dimension embedded data points[19]. The essence of SVM is a structure risk minimization principle. It is based on Vapnik-Chervonenkis (VC) theory. Initially, SVM was used to solve pattern recognition problems. In order to find a decision rule with good generalization capability, the so-called support vectors (SVs) consisted by a small subset of the training data, are selected to support the optimal hyperplane [20]. Later, support vector regression (SVR) was exploited to cope with the regression estimation and function approximation problems. Although the traditional SVR is very effective for function estimation, there still exist shortcomings, such as the high computational complexity and the solution is not sparse enough. Therefore, the idea of linear programming support vector regression (LP-SVR) is born and used to capitalize on the advantages of the mode sparsity. It is developed to change the support vector regression from a quadratic programming to a linear programming problem which uses the l_1 norm of the coefficient vector [20]. Specifically, the nonlinear regression problem is treated as a linear one in the kernel space [21].

In this work, a Wiener model is presented to model the blast furnace ironmaking process, viz. the silicon content. For the proposed model, the linear dynamic subsystem is described by Laguerre filter and the nonlinear static part is represented by linear programming support vector regression.

This paper is organized as follows. Brief review of related methods for the modeling process is in Section 2. Section 3 gives the experimental validation in nonlinear blast furnace system which conclude the experimental data and screening and the predictive results. Finally, Section 4 concludes this paper.

II. Problem Formulation

A. Wiener model

Nonlinear system can be approximated by Wiener model which consists of a linear dynamic element followed by a nonlinear static element. In this paper, Wiener model is used to identify a multi-input singler-output discrete-time nonlinear dynamical system as follows

$$\mathbf{X}(k+1) = f(\mathbf{X}(k), \mathbf{u}(k)) \tag{1}$$

$$\hat{y}(k) = h(\mathbf{X}(k)) \tag{2}$$

where $\mathbf{u}$ is the scalar input signal which is selected in advance, $\mathbf{X}$ is the state vector which is both the linear dynamic element output and the static nonlinear component input. $\hat{y}$ is the model output. $f(\cdot,\cdot)$ and $h(\cdot)$ are the linear and nonlinear mappings. The structure of Wiener model is shown in Fig. 1.

In this paper, the linear dynamic component is parameterized by Laguerre filter while the nonlinear static part is represented by linear programming support vector regression.

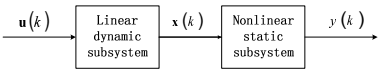


Fig. 1: Structure of Wiener model

Our goal is to find an appropriate Wiener model which can explain the input-output data $\{\mathbf{u}(k), y(k)\}$ as well as possible, where y is the measured output signal.

B. Laguerre filters

The idea of representing the linear dynamic block using orthonormal filter networks such as Laguerre and Kautz filter has received increasing attention in the past decades. A discrete Laguerre filter which does not need any explicit information about system time constant and time delay can expand the output of the linear dynamics as an expression of finite input bases [22]. Consider a SISO linear system modeled by a Laguerre filter as follows [18]

$$\hat{y}(z) = \left(\sum_{i=1}^{n} c_i L_i(z)\right) u(z) \tag{3}$$

where

$$L_i(z) = \frac{\sqrt{(1-a^2)T}}{(z-a)} \left(\frac{1-az}{z-a}\right)^{i-1} \tag{4}$$

Here $L_i(z)$ indicates the ith order Laguerre filter, n is number of Laguerre filter used for model development, c_i is the Laguerre coefficient, $a(-1 < a < 1)$ is the Laguerre filter time-scaling factor, T is the sampling interval, $\hat{y}$ is the model output, and u represents manipulated input.

The state vector is defined as

$$X(k) = [x_1(k), x_2(k), \ldots, x_n(k)]^T \tag{5}$$

where $x_i(k)$ represents the output from ith order Laguerre filter at the kth sampling instant, a discrete state space realization of the Laguerre filter network can obtained as

$$X(k+1) = \varphi(a)X(k) + \gamma(a)u(k) \tag{6}$$

where $u(k)$ is the system input, $\gamma(a)$ is an n dimensional vector defined as

$$\gamma(a) = [\sqrt{(1-a^2)T}, \ldots, (-a)^{n-1}\sqrt{(1-a^2)T}]^T \tag{7}$$

and $\varphi(a)$ is an $n \times n$ lower triangular matrix defined as

$$\varphi(a) = \begin{bmatrix} a & 0 & 0 & \cdots & 0 \\ (1-a^2) & a & 0 & \cdots & 0 \\ -a(1-a^2) & (1-a^2) & a & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (-1)^n a^{n-2}(1-a^2) & (-1)^{n-1}a^{n-3}(1-a^2) & (-1)^{n-2}a^{n-4}(1-a^2) & \cdots & a \end{bmatrix} \tag{8}$$

For the SISO linear model (3), the model output can be expressed as the weighted sum of states

$$\hat{y}(k) = \mathbf{c}^T X(k) \tag{9}$$

where $\mathbf{c} = [c_1, c_2, ..., c_n]^T$.

Consider to develop a Wiener-Laguerre SISO nonlinear model, a nonlinear state-output map can be constructed so that the model output is represented as

$$\hat{y}(k) = \Psi(X(k)) \tag{10}$$

where $\Psi(\cdot)$: $R^n \longrightarrow R$ represents the nonlinear mapping.

Then let's expand the SISO model to MISO condition which has m inputs. For simplicity, we assume each input has the same number of filters $n_i(n)$ and the same Laguerre coefficient $a_i(a)$. The input vector is defined as

$$U(k) = [u_1(k), u_2(k), \ldots, u_m(k)]^T \tag{11}$$

and state vector is defined as

$$\mathbf{X}(k) = [X_1(k), X_2(k), \ldots, X_m(k)]^T \tag{12}$$

$$X_i(k) = [x_{i1}(k), x_{i2}(k), \ldots, x_{in}(k)]^T, i = 1, 2, \ldots, m \tag{13}$$

where $x_{ij}(j = 1, 2, \ldots, n)$ is the state vector correlating the jth order Laguerre filter with ith input.

The corresponding state dynamics can be defined as

$$\mathbf{X}(k+1) = \Phi(a)\mathbf{X}(k) + \Gamma(a)U(k) \tag{14}$$

and

$$\Phi(a) = block\ diag[\varphi(a), \varphi(a), \ldots, \varphi(a)]_{N\times N} \tag{15}$$

$$\Gamma(a) = block\ diag[\gamma(a), \gamma(a), \ldots, \gamma(a)_{N\times m}] \tag{16}$$

where

$$N = m \times n \tag{17}$$

The output of the MISO Wiener-Laguerre model can be expressed as

$$\hat{y}(k) = \Psi[\mathbf{X}(k)] \tag{18}$$

where $\Psi(\cdot)$: $R^N \longrightarrow R$ represents the nonlinear mapping.

C. Linear programming support vector regression

SVM is a kernel-based method and frequently used for nonlinear classfication. It is performed by using a real-valued function f: $X \subseteq R^n \rightarrow R^m (m > n)$ in the following way: the input $x = (x_1, \ldots, x_n)$ is assigned to the positive class if $f(x) \geq 0$, and otherwise to the negative class. The basic idea is to construct a so-called optimal separating hyperplane in a high-dimensional (even infinite-dimensional) feature space by maximizing the margin between the nearest training data points of the two classes. It is based on Vapnik's ε-insensitive loss function and structural risk minimization that the SV solution derived can be sparse. The goal of SVR is to bound the mean approximation error of a finite data set [8]. The select of kernel function and the homologous kernel parameters is very important for the accuracy of results.

For a given training set $D = \{(x_i, y_i), i = 1, \ldots, l\}$, where x_i is the ith input vector, y_i is the corresponding response variable, l is the total number of exemplars. In this paper, x_i is the ith state of the Laguerre filters. We consider the case can be written as a nonlinear mapping as follows

$$f(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b \tag{19}$$

where $f(\mathbf{x})$ maps a data point $\mathbf{x}_i$ into a higher dimensional space. In $\varepsilon-$SVR, the goal is to find a function $f(\mathbf{x})$ that has at most ε deviation from the actually obtained target y_i. In the traditional support vector method, the problem of the $\varepsilon-$SVR is formulated as

$$\begin{aligned} &\min_{\mathbf{w},b} \ \tfrac{1}{2} \parallel \mathbf{w} \parallel^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) \\ &s.t. \ \begin{cases} y_i - \left(\mathbf{w}^T \cdot \varphi(\mathbf{x}_i) + b\right) \leqslant \varepsilon + \xi_i \\ \mathbf{w}^T \cdot \varphi(\mathbf{x}_i) + b - y_i \leqslant \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geqslant 0, \ i = 1, \ldots, l \end{cases} \end{aligned} \tag{20}$$

where the parameter $C > 0$ decides the trade-off between the large margin and a small error penalty, $\varepsilon > 0$ controls the admissible uncertainty on the data points, ξ_i and ξ_i^* are slack variables, ξ_i denotes the training error above ε, whereas ξ_i^* denotes the training error below ε. This is a classic quadratic optimization problem with inequality constraints. By defining the $\varepsilon-$insensitive loss function

$$H_\varepsilon (y_i - f(\mathbf{x}_i)) = \max \ \{0, |y_i - f(\mathbf{x}_i)| - \varepsilon\} \tag{21}$$

the optimization problem (20) is equivalent to the following regularization problem [20]

$$\min_f \left\{ \lambda \parallel \mathbf{w} \parallel^2 + \sum_i^l H_\varepsilon (y_i - f(\mathbf{x}_i)) \right\} \tag{22}$$

where $f(\mathbf{x})$ is in the form of (19) and $\lambda \parallel \mathbf{w} \parallel^2$ is the regularization term. According to the well-known Representer Theorem, the solution to the optimal function (22) can be written as the linear combination of the kernel functions centering the training examples

$$f(\mathbf{x}) = \sum_{i=1}^{l} \beta_i k(\mathbf{x}, \mathbf{x}_i) \tag{23}$$

where $k(\mathbf{x}, \mathbf{x}_i)$ is the kernel function which can be chosen from the following functions

- Gaussian radial basis function (GRBF) kernel:

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(\frac{- \parallel \mathbf{x} - \mathbf{x}' \parallel^2}{2\sigma^2}\right) \tag{24}$$

- Polynomial kernel:

$$k(\mathbf{x}, \mathbf{x}') = (1 + \langle \mathbf{x}, \mathbf{x}' \rangle)^q \tag{25}$$

- Sigmoid kernel:

$$k(\mathbf{x}, \mathbf{x}') = \tanh(\alpha \langle \mathbf{x}, \mathbf{x}' \rangle + \gamma) \tag{26}$$

where $\sigma, q, \alpha, \gamma$ are the adjustable parameters of the above kernel functions. In this paper, the GRBF kernel is the choice.

Linear programming support vector regression changes the l_2 norm in (22) to l_1 norm as

$$\min_f \left\{ \lambda \parallel \boldsymbol{\beta} \parallel_1 + \sum_i^l H_\varepsilon \left(y_i - f(\mathbf{x}_i) \right) \right\} \tag{27}$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_l]^T$, $f(\mathbf{x})$ is in the form of (23). This regularization programming is equivalent to the following optimization problem

$$\min_{\boldsymbol{\beta}} \ \tfrac{1}{2} \parallel \boldsymbol{\beta} \parallel_1 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*)$$
$$s.t. \begin{cases} y_i - \sum\limits_{j=1}^{l} \beta_j k(\mathbf{x}_j, \mathbf{x}_i) \leqslant \varepsilon + \xi_i \\ \sum\limits_{j=1}^{l} \beta_j k(\mathbf{x}_j, \mathbf{x}_i) - y_i \leqslant \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geqslant 0, \ i = 1, \ldots, l \end{cases} \tag{28}$$

From [21], it can be followed that $\xi_i \xi_i^* = 0$ in SV regression and it is sufficient to just introduce slack variable ξ_i in the constrained optimization problem (28). Thus, the optimization problem can be written as

$$\min_{\boldsymbol{\beta}} \ \tfrac{1}{2} \parallel \boldsymbol{\beta} \parallel_1 + 2C \sum_{i=1}^{l} \xi_i$$
$$s.t. \begin{cases} y_i - \sum\limits_{j=1}^{l} \beta_j k(\mathbf{x}_j, \mathbf{x}_i) \leqslant \varepsilon + \xi_i \\ \sum\limits_{j=1}^{l} \beta_j k(\mathbf{x}_j, \mathbf{x}_i) - y_i \leqslant \varepsilon + \xi_i \\ \xi_i \geqslant 0, \ i = 1, \ldots, l \end{cases} \tag{29}$$

In order to transform the above optimization problem into a linear programming problem, the variable β_i and $|\beta_i|$ can be decomposed to the following form

$$\beta_i = \alpha_i^+ - \alpha_i^-, \ |\beta_i| = \alpha_i^+ + \alpha_i^- \tag{30}$$

where $\alpha_i^+, \alpha_i^- \geqslant 0$. It is worth noting that the decompositions of (30) are unique that for a given β_i there is only one pair $\left(\alpha_i^+, \alpha_i^-\right)$ which fulfills both equations. $\alpha_i^+ \cdot \alpha_i^- = 0$ is guaranteed implicitly to guarantee both variables cannot be larger than zero at the same time. In this way, the l_1 norm of $\boldsymbol{\beta}$ can be written as

$$\parallel \boldsymbol{\beta} \parallel_1 = \left(\underbrace{1, \ldots, 1}_{l}, \underbrace{1, \ldots, 1}_{l} \right) \begin{pmatrix} \boldsymbol{\alpha}^+ \\ \boldsymbol{\alpha}^- \end{pmatrix} \tag{31}$$

where $\boldsymbol{\alpha}^+ = \left(\alpha_1^+, \alpha_2^+, \ldots, \alpha_l^+\right)^T$ and $\boldsymbol{\alpha}^- = \left(\alpha_1^-, \alpha_2^-, \ldots, \alpha_l^-\right)^T$. Then plug (30) into (29), we obtain

$$\min_{\boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-} \ \tfrac{1}{2} \sum_{i=1}^{l} \left(\alpha_i^+ + \alpha_i^- \right) + 2C \sum_{i=1}^{l} \xi_i$$
$$s.t. \begin{cases} y_i - \sum\limits_{j=1}^{l} \left(\alpha_j^+ - \alpha_j^- \right) k(\mathbf{x}_j, \mathbf{x}_i) \leqslant \varepsilon + \xi_i \\ \sum\limits_{j=1}^{l} \left(\alpha_j^+ - \alpha_j^- \right) k(\mathbf{x}_j, \mathbf{x}_i) - y_i \leqslant \varepsilon + \xi_i \\ \xi_i \geqslant 0, \ \alpha_i^+, \alpha_i^- \geqslant 0, \ i = 1, \ldots, l \end{cases} \tag{32}$$

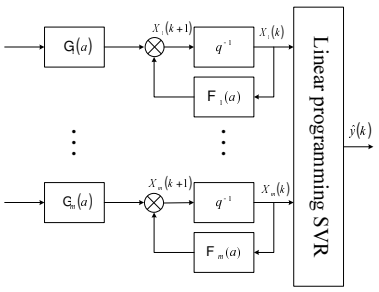


Fig. 2: Wiener model with Laguerre filter and LP-SVR

For the sake of simplicity, the above optimization problem can be written as the following linear programming problem

$$\begin{aligned} &\min \ \mathbf{c}^T \mathbf{x} \\ &s.t. \ A\mathbf{x} \leqslant \mathbf{b} \end{aligned} \tag{33}$$

where

$$\mathbf{c} = \left[\underbrace{1, \ldots, 1}_{l}, \underbrace{1, \ldots, 1}_{l}, \underbrace{2C, \ldots, 2C}_{l} \right]^T,$$

$$\mathbf{x} = \left[\underbrace{\alpha_1^+, \ldots, \alpha_l^+}_{l}, \underbrace{\alpha_1^-, \ldots, \alpha_l^-}_{l}, \underbrace{\xi_1, \ldots, \xi_l}_{l} \right]^T,$$

$$A = \begin{bmatrix} K & -K & -I \\ -K & K & -I \end{bmatrix} \text{ with } K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j),$$

$$\mathbf{b} = \left[\underbrace{\varepsilon + y_1, \ldots, \varepsilon + y_l}_{l}, \underbrace{\varepsilon - y_1, \ldots, \varepsilon - y_l}_{l} \right]^T,$$

I is $l \times l$ identity matrix.

After (33) is solved, we can get the predictive result as

$$f(\mathbf{x}) = \sum_{\mathbf{x}_i \in SV} \left(\alpha_i^+ - \alpha_i^- \right) k(\mathbf{x}, \mathbf{x}_i) \tag{34}$$

Here, SV represents the small set of support vectors.

The state space illustration of the whole MISO Wiener model is below in Fig. 2

III. Experimental validation in nonlinear blast furnace system

In this section, the proposed Wiener model is applied to data collected from No.2 blast furnace of Liuzhou Steel in China. The volume of the blast furnace is 2000 m^3. There are, in total, 500 consecutive data points available from the selected blast furnace with the preceding 300 data points to be the training set and the rest 200 data points to be the test set.

A. Experimental data and preprocessing

There are 23 candidate variables from which to select model inputs. As mentioned in Section 2, not all monitored process variables have an influence on the fluctuation of silicon content. Additionally, too many inputs will increase the complexity of the model while too little inputs will reduce the model precision. So it is necessary to select the more

important variables as inputs from all these listed variables. Here grey correlation method is used for the selection [23].

The grey correlation degree is a quantitative value of the correlation between the factors. Positive correlation between main-array and sub-array indicates that the sub-factor will enhance the main-factor. However, negative correlation indicates that the sub-factor will weaken the main-factor. The higher the value of grey correlation degree is, the more relevant the main-factor and the sub-factor are. To get the grey relational degree and the grey relational order, the grey correlation analysis method can be summarized as follows

1) Get the reference $X_0 = (x_0(1), x_0(2), \ldots, x_0(n))$ and comparison sequences $X_i = (x_i(0), x_i(1), \ldots, x_i(n)), i = 1, 2, \ldots m$.
2) Calculate the relational degree. To determine the relational degree between the reference and comparison sequences, a discrete function of the relational degree coefficient (the grey relational coefficient) is represented by

$$\varsigma_{0i}(k) = \frac{\Delta_{\min} + \rho\Delta_{\max}}{\Delta_{0i}(k) + \rho\Delta_{\max}}, k = 1, 2, \ldots, n \qquad (35)$$

where

$$\Delta_{0i}(k) = \mid x_i(k), x_0(k) \mid$$

$$\Delta_{\max} = \max_i \max_k \{\Delta_{0i}(k)\}$$

$$\Delta_{\min} = \min_i \min_k \{\Delta_{0i}(k)\}$$

and $\rho \in [0, 1]$ is the coefficient to distinguish the degree of proximity of X_0 and X_i such that $\varsigma_{0i} \in [0, 1]$. This value can be adjusted based on the actual system requirements. In this paper, we consider $\rho = 0.5$. After the grey relational coefficients have been obtained, the mean of the coefficients usually adopted as the grey relational degree. Then

$$\gamma(x_0, x_i) = \frac{1}{n}\sum_{k=1}^{n} \varsigma_{0i}(k) \qquad (36)$$

is called the grey correlation degree of the ith comparison sequence X_i to the reference sequence X_0.

3) Order the grey correlation degree. From the ordered $\gamma(x_0, x_i)$, we shall pick the sequence with the greatest relational degree.

Table I lists the grey correlation degree of candidate variables. The variables with grey correlation degree larger than 0.8800 are selected as the inputs of the model.

TABLE I: The grey correlation degree of candidate variables.

Variable	Grey correlation degree
The latest silicon content	**0.9008**
Cold wind flowrate	**0.8850**
Feed wind ratio	**0.8849**
Hot blast pressure	**0.8835**
Furnace top pressure	**0.8823**
Pressure difference	**0.8847**
Top pressure blast volume ratio	**0.8817**
Gas permeability	**0.8839**
Drag coefficient	**0.8810**
Hot blast temperature	0.8762
Oxygen enrichment flowrate	0.7966
Oxygen enrichment percentage	0.7964
Pulverized coal injection	0.8702
Blast humidity	**0.8860**
Standard wind speed	**0.8850**
Actual wind speed	0.8788
Blast momentum	0.8718
Bosh gas volume	**0.8847**
Bosh gas index	**0.8847**
Top temperature (northeast)	**0.8825**
Top temperature (southwest)	**0.8830**
Top temperature (northwest)	**0.8849**
Top temperature (southeast)	**0.8847**

B. Parameter settings

There are three parameters to be determined in advance while using RBF kernel, viz. ε, C, and σ^2. We examine the forecasting performance of the method with $\varepsilon = 0.01$. C and σ^2 are determined by Grid-search method. The theory of this search method is to divide the feasible region of each parameter into some small subregions. Computer can calculate the parameters combination sequentially and provide the corresponding error. Parameters will be updated if the error value is reduced. Finally, the best parameters will be given. The final search results are C = 500, σ^2 = 2.

The laguerre filter and order have an effect on the model accuracy. Through analysis, we select the laguerre filter pole $a = 0.2$ and filter orders $M = 4$.

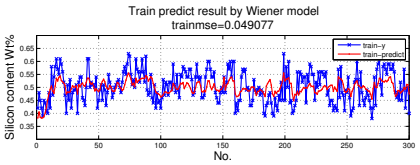
C. Predictive results and evaluation

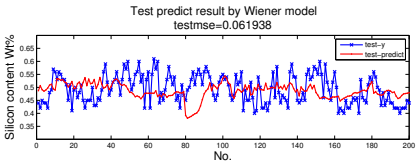
To better exhibit the performance of the proposed method, RMSE (Root Mean Square Error) is defined as a derivation measurement between the target and the predictive values. All experiments are carried out using MATLAB R2009a with 2GB memory and the operating system of Windows XP.

$$RMSE = \sqrt{\frac{1}{l}\sum_{i=1}^{l}(\hat{y}_i - y_i)^2} \qquad (37)$$

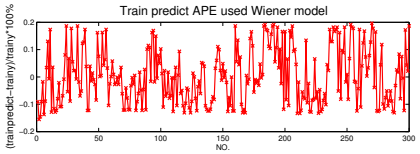
The RMSEs of the train set and test set are trainRMSE = 0.049077, testRMSE = 0.061938.

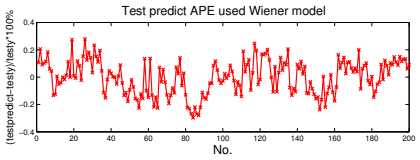
Fig. 3 shows the predictive results and APEs (Absolute





(a)





(b)

Fig. 3: Predictive results and the APEs.

Percentage Error) compared with the measured output.

$$APE_i = \frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \tag{38}$$

It can be seen that the proposed Wiener model can successfully model the silicon content of blast furnace. The Laguerre filter part and LP-SVR part approximate the linear and nonlinear part without knowing the inner information of industrial process but still predict the next silicon content with small error. In addition, linear programming support vector regression enables us to obtain a much sparser approximation model which makes the model process simpler. In conclusion, the proposed Wiener model is suitable to the blast furnace system and it can be spread to other complex industrial systems.

IV. Conclusion

This paper has developed a Wiener model to identify the silicon content of blast furnace ironmaking process, where Laguerre filter and linear programming support vector regression represent the linear and nonlinear component. Experimental results show that this model structure is suitable for hot metal prediction. The proposed Wiener model structure reduces the parameters to be optimized, only several parameters need to be selected. This modeling process greatly reduce the modeling complexity and achieve a good result. Due to the complexity of blast furnace smelting process, we should deepen the analysis of blast furnace and improve the prediction technique so that the method could be more suitable to the smelting process.

Future work research work will focus on the deep analysis of blast furnace ironmaking process and the exploration of modeling approach.

References

[1] C.H. Gao, L. Jian, X.Y. Liu, J.M. Chen and Y.X. Sun, "Data-driven modeling based on volterra series for multidimensional blast furnace system," *IEEE Transactions on Neural Network,* vol. 22, no. 12, pp. 2272-2283, Dec. 2011.

[2] L. Jian, C.H. Gao, "Bianry coding SVMs for the multiclass problem of the blast furnace system," *IEEE Transactions on Industrial Electronics,* vol. 60, no. 9, pp. 3846-3856, Sep. 2013.

[3] C.H. Gao, J. Ling and S.H. Luo, "Modeling of the thermal state change of blast furnace heart with support vector machines," *IEEE Transactions on Industrial Electronics,* vol. 59, no. 2, pp. 1134-1145, Feb. 2012.

[4] C.H. Gao, Q.H. Ge and L. Jian, "Rules extraction from fuzzy-based blast furnace SVMs multiclassifier for decision-making," *IEEE Transactions on Fuzzy Systems*, avilible online, June. 2013.

[5] V.R. Radhakrishnan and A.R. Mohamed, "Neural networks for the identification and control of blast furnace hot metal quality," *Journal of Process Control*, vol. 10, pp. 509-524, 2000.

[6] J. Chen, "A predictive system for blast furnaces by integrating a neural network with qualitative analysis," *Engineering Applications of Artificial Intelligence*, vol. 14, pp. 77-85, 2001.

[7] H. Saxén, and F. Pettersson, "Nonlinear prediction of the hot metal silicon content in the blast furnace," *ISIJ International*, vol. 47, pp. 1732-1737, 2007.

[8] R.D. Martín, F. Obeso, J. Mochón, R. Barea and J. Jiménez, "Hot metal temperature prediction in blast furnace using advanced model based on fuzzy logic tools," *Ironmaking & steelmaking*, vol. 34, pp. 241-247, 2007.

[9] X.L. Tang, L. Zhang and C.J. Jian, "Prediction of silicon content in hot metal using support vector regression based on chaos particle swarm optimization," *Expert Systems with Applications*, vol. 36, pp. 11835-11857, 2009.

[10] L. Jian C.H. Gao and Z.Q. Xia, "A sliding-window smooth support vector regression model for nonlinear blast furnace system," *Steel Research International*, vol. 82, pp. 169-179, 2011.

[11] M.S. Phadke and S.M. Wu, "Identification of multiinput-multioutput transfer function and noise model of a blast furnace from closed-loop data," *IEEE Transactions on Automatic Control,* vol. AC-19, no. 6, pp. 944-951, Dec. 1974.

[12] H. Saxén, and R. Östermark, "State realization with exogenous variables—A test on blast furnace data," *European Journal of Operational Research*, vol. 89, pp. 34-52, 1996.

[13] T. Bhattacharya, "Prediction of silicon content in blast furance hot metal using partial least squares (PLS)," *ISIJ International*, vol. 45, pp. 1943-1945, 2005.

[14] F. Pettersson, N. Chakraborti and H. Saxén, "A genetic algorithms based multi-objective neural net applied to noisy blast furnace data," *Applied Soft Computing*, vol. 7, pp. 387-397, 2007.

[15] F. Pettersson, H. Saxén and K. Deb, "Genetic algorithm-based multi-criteria optimization of ironmaking in the blast furnace," *Materials and Manufacturing Processes*, vol. 24, pp. 343-349, 2009.

[16] J.S. Zeng, X.G. Liu, C.H. Gao, S.H. Luo and L. Jian, "Wiener model identification of blast furnace ironmaking process," *ISIJ International*, vol. 48, pp. 1734-1738, 2008.

[17] S. Tötterman and H.T. Toivonen, "Support vector method for identification of Wiener models," *Journal of Process Control*, vol. 19, pp. 1174-1181, 2009.

[18] P. Saha, S.H. Krishnan, V.S.R. Rao and S.C. Patwardhan, "Modeling and predictive control of MIMO nonlinear systems using Wiener-Laguerre models," *Chemical Engineering Communications*, vol. 191, pp. 1083-1119, 2004.

[19] H. Saxén, C.H. Gao and Z.W. Gao, "Data-driven time discrete models for dynamic prediction of the hot metal silicon content in the blast furnace," *IEEE Transactions on Industrial Informatics,* vol. 9, no. 4, pp. 2213 - 2225, Nov. 2012.

[20] Y.P. Zhao and J.G. Sun, "Multikernel semiparametric linear programming support vector regression," *Expert Systems with Applications*, vol. 38, pp. 1611-1618, 2011.

[21] Z. Lu and J. Sun, "Non-mercer hybrid kernel for linear programming support vector regression in nonlinear systems identification," *Applied Soft Computing*, vol. 9, pp. 94-99, 2009.

[22] S.H. Hwang, C.Y. Hsieh, H.T. Chen and Y.C. Huang, "Use of discrete Laguerre expansions for noniterative identification of nonlinear Wiener models," *Industrial & Engineering Chemistry Research*, vol. 50, pp. 1427-1438, 2011.

[23] J. Zhao, W. Wang, Y. Liu, W. Pedrycz, "A Two-Stage Online Prediction Method for a Blast Furnace Gas System and Its Application," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 3, pp. 507-520, 2011.