Data-Based Multiscale Modeling for Blast Furnace System

Yanxu Chu and Chuanhou Gao
Dept. of Mathematics, Zhejiang University, Hangzhou 310027, China

DOI 10.1002/aic.14426
Published online March 11, 2014 in Wiley Online Library (wileyonlinelibrary.com)

The operation of a blast furnace system involves strong multiscale features to which enough attention should be paid, when describing its complex dynamics. For this purpose, a data-based multiscale modeling algorithm that can extract and pick out some subscale variables most relevant to the output from the original input variables set is presented. These selected subscale variables acting as inputs together with the output are introduced into a general linear or nonlinear model framework to form the corresponding multiscale model. Through model validation, the constructed multiscale models are found to exhibit large advantage compared with those traditional models based on the averaging idea over a fixed scale, especially in the cases of nonlinear models, in which high agreement between the predicted values and the real ones is observed. These results indicate that the proposed multiscale modeling algorithm on the one hand, can provide a kind of thought to develop a data-based multiscale model from the viewpoint of methodology, and conversely, can serve as a potential blast furnace modeling tool from the viewpoint of engineering applications. © 2014 American Institute of Chemical Engineers *AIChE J*, 60: 2197–2210, 2014
Keywords: multiscale modeling, blast furnace system, data based, prediction

Introduction

Blast furnace ironmaking process is a highly complex nonlinear process with the main purpose of producing pig iron, often called hot metal, for primary steelmaking. A blast furnace operates like a countercurrent reactor, from the top of which the solid raw materials including ore and coke are charged layer by layer, while preheated air together with pulverized coal or oil is blown into at the bottom. The complex chemical reactions and multiphase transport phenomena, accompanied by high temperature, high pressure, distributed parameters, and strong noise, take place during the whole ironmaking process. It may be said that the blast furnace is one of the most complex industrial reactors. This together with the substantial connection of iron and steel industry to the national economy and the living standard makes the blast furnace attracting wide interest for the purposes of academical research and engineering applications. This is further emphasized with the advent of the global warming and strongly rising prices for the blast furnace raw materials.

A great deal of effort has been made in the past decades to construct the blast furnace mathematical models for energy saving and consumption reduction, including first principles based,[1–3] data-driven models,[4–15] and even rule-based expert system.[16] These models have contributed to improving the blast furnace performance remarkably; however, they still lack of enough reliability to be used in practice. First principles-based blast furnace models cannot handle discontinuous phenomena, such as unstable slip of furnace charge,[17] external noise disturbance originated by control actions[18]; data-driven models are essentially black-box ones, and their transparency and comprehensibility are extremely lacking; rule-based expert system often works more effective when it is combined with a blast furnace mathematical model, so it still needs the reliable blast furnace models. What is especially important is that most of the above models are based on the idea of average over a fixed scale, and the contributions of different sublevel scales to the system outputs have not been distinguished. In fact, the blast furnace ironmaking process has obvious multiscale features, such as atomic/molecular chemical reactions, device/boundary layer transport phenomena, process systems engineering, and so forth occurring at different locations. Meantime, there are great differences in the average residence time of solid, liquid, and gaseous phase, like 5–10 h for the iron ore and liquid phase, 10–20 s for the gaseous phase, and 1–4 or more weeks for the coke in the hearth.[19] Our earlier work[20,21] also detected the multiscale nature and scales coupling phenomena of the blast furnace system through analyzing the operating data. Theoretically, every subscale structure has its own control mechanism and makes a certain contribution to the system behaviors. It is of high importance to capture these different scales information, when constructing the blast furnace mathematical model. For capturing the complex blast furance dynamics, it is, thus, better to construct a multiscale model to reflect the multiscale features exhibited during the blast furnace ironmaking process.

Multiscale methodology, formed in the late 1980s,[22] has yielded extensive applications in the control discipline, such as multivariate statistical process control based on wavelet transforms (or wavelet packets) and principal component analysis to extracts those features that represent abnormal operation,[23–25] multiscale model based on the dynamic Monto Carlo

Correspondence concerning this article should be addressed to C. Gao at gaochou@zju.edu.cn.

method to predict polymer structure distributions, to capture the evolution of both macroscopic and molecular phenomena during the sol-gel silica film formation process,[26] chemical industrial process modeling[27,28] and so forth. However, little attention is paid to the multiscale modeling for the blast furnace system. Given the multiscale features exhibited in the blast furnace ironmaking process, this work is devoted to designing a data-based multiscale algorithm for the blast furnace modeling that can fully consider the multiscale features and the differences of different sublevel scales contributing to the system outputs. The rest of this article is organized as follows. In the next section, problem formulation is presented. This is followed by the descriptions of the related analytical methods for multiscale decomposition and variables selection. Then, the blast furnace process including the silicon transfer mechanism, data collection, and analysis on the interdependence of variables is introduced. Next, the multiscale modeling algorithm and blast furnace experimental validation are given. Finally, conclusions are summarized.

Problem Formulation

Consider a multiple-input and single-output system, formulated by

$$y(t)=f(\mathbf{x}(t)) \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{R}^M$ is the input vector at time t and $y(t) \in \mathbb{R}$ is the corresponding output. In general, the next effort will be made toward finding an appropriate project $f(\cdot)$ to approximate the input–output relationship for modeling, control, and optimization purposes. However, it should be mentioned that the inputs involved in Eq. 1 are usually collected at a fixed scale, and the contributions of other sublevel scales to the system output are not considered. Therefore, strictly speaking, Eq. 1 is unsuited to be used to describe a system with multiscale features. For this reason, it is better to take into account the contributions of different sublevel scales when modeling a system with multiscale features. Attention is turned to this task below:

Assume the inputs in Eq. 1 to be $\mathbf{x}(t)=(x_1(t), x_2(t), \cdots, x_M(t))$ and that for every input $x_i (i=1, \cdots, M)$ there are N_i subscale signals, denoted as $c_{i1}, c_{i2}, \cdots, c_{iN_i}$. Theoretically, every sublevel scale will have a contribution to the system output, so the multiscale manner of Eq. 1 may be written as

$$y(t)=g(c_{11}(t), \cdots, c_{1N_1}(t), \cdots, c_{M1}(t), \cdots, c_{MN_M}(t)) \tag{2}$$

Compared with Eqs. 1 and 2 includes more inputs and much finer scale information. These new added inputs may provide richer information to characterize the underlying dynamics, and it is, thus, taken for granted that this equation can be more accurate and more suitable for describing a system with multiple scales. Despite the above facts, it also poses a problem of the great increase in the dimension of the inputs from Eq. 1 to 2. In the multiscale manner, the dimension of the inputs is

$$L=N_1+N_2+\cdots+N_M \tag{3}$$

which will make Eq. 2 a rather high-dimensional multiscale model. For this kind of high-dimensional problems, collinearity of variables and model overfitting are often encountered.[29] The former means that certain variables involved in model can be perfectly represented by a linear combination of some other variables, and inclusion of these variables may result in excessively redundant information and further yield a model with poor performance. The latter, in turn, means that the model describes random noise instead of the underlying input–output relationship, which will make the model have strong learning ability but poor generalization performance. To reduce the effect of the above two phenomena, it is necessary to perform inputs selection before constructing the underlying multiscale model. Among all the inputs appeared in Eq. 2, assume that the following inputs are picked out $c_{1S_{11}}, \cdots, c_{1S_{1v_1}}, c_{2S_{21}}, \cdots, c_{2S_{2v_2}}, \cdots, c_{MS_{M1}}, \cdots, c_{MS_{Mv_M}}$, where $\{S_{i1}, \cdots, S_{ip_i}\} \subset \{1, \cdots, N_i\}, i=1, \cdots, M$, then the multiscale model will change to be

$$y(t)=h(c_{1S_{11}}(t), \cdots, c_{1S_{1v_1}}(t), \cdots, c_{MS_{M1}}(t), \cdots, c_{MS_{Mv_M}}(t)) \tag{4}$$

At this moment, the dimension of model inputs reduces to be $v_1+v_2+\cdots+v_M < L$. This black-box model will act as the main basis for the subsequent research.

Analytical Methods

In this section, the related analytical methods used for designing the mulitiscale modeling algorithm are described, including the multiscale decomposition method from a time series, variables relativity measuring approach, and variables selection technique.

Hilbert–Huang transform

Hilbert–Huang transform (HHT) is a kind of powerful multiscale resolution tool from a time series,[30] in which two procedures are included: one is empirical mode decomposition (EMD), the other is Hilbert transform. The main idea of EMD is to continuously sift out the intrinsic mode functions (IMFs) embedded with a time series. Here, the IMF means any function with the same number of extrema and zero crossings, and with its envelopes being symmetric with respect to zero. As an example of the time series $x(t)$, the EMD sifting process can be written as

$$x(t)=\sum_{j=1}^{n} c_j(t)+r_n(t) \tag{5}$$

where $c_j(t)$ is the jth IMF, $r_n(t)$ the residual term, and n is the number of IMFs, that is, the number of the time scales. A detailed procedure to perform EMD on a time series may be found in the work of Huang et al.[30] Unlike the traditional multiscale decomposition method, such as wavelet transform which is a linear analytical tool, the EMD method has strong local and adaptive properties, and can, thus, handle a time series with high nonlinearity and nonstationarity.[30] As nonlinearity and nonstationarity are the essence of a multiscale structure,[28] the EMD method will be used to perform multiscale decomposition on the time series measured from the blast furnace system in the subsequent investigation.

After multiscale decomposition, HHT performs the Hilbert transform on every IMF and further yields the corresponding analytic signal $z_j(t)$ following

$$z_j(t)=c_j(t)+i\tilde{c}_j(t)=a_j(t)e^{i\varphi_j(t)} \tag{6}$$

where $\tilde{c}_j(t)$ is the Hilbert transform of $c_j(t)$, i is the imaginary unit, $a_j(t)$ is the instantaneous amplitude, and $\varphi_j(t)$ is the phase angle, satisfying

$$\begin{cases} a_j(t)=\sqrt{c_j^2(t)+\tilde{c}_j^2(t)} \\ \tan\varphi_j(t)=\tilde{c}_j(t)/c_j(t) \end{cases} \tag{7}$$

The phase angle further defines the instantaneous frequency $\omega_j(t)$ as

$$\omega_j(t)=d\varphi_j(t)/dt \tag{8}$$

the average value of which is representative of a physical time scale hidden in the time series $x(t)$.[30] Therefore, every decomposed signal is representative of a subscale signal produced by the corresponding subscale system contained in the studied system.

Mutual information

Mutual information is a measure that quantifies the mutual dependence of a coupled system (X, Y), including linear and nonlinear dependence. According to information theory,[31] it is commonly defined to be the difference between the self entropies of all variables and their joint entropy. Let $\mathbf{x}(t)$ and $y(t)$ stand for the measured values of variables X and Y, respectively, then their mutual information may be expressed as

$$I(X,Y)=H(X)+H(Y)-H(X,Y) \tag{9}$$

where I (X, Y) is the mutual information, and $H(X)$, $H(Y)$, and $H(X, Y)$ are the self and joint entropies, respectively, defined as

$$H(X)=-\int \mu_X(\mathbf{x})\log\mu_X(\mathbf{x})d\mathbf{x} \tag{10}$$

$$H(X,Y)=-\int\int \mu_{X,Y}(\mathbf{x},y)\log\mu_{X,Y}(\mathbf{x},y)d\mathbf{x}dy \tag{11}$$

Here, $\mu_X(\mathbf{x})$ and $\mu_{X,Y}(\mathbf{x},y)$ are the probability density function of the isolated variable X and the coupled variables (X,Y), also called marginal and joint probability density, respectively. They follow

$$\mu_X(\mathbf{x})=\int \mu_{X,Y}(\mathbf{x},y)dy \tag{12}$$

Substituting Eqs. 10, 11, and 12 into Eq. 9 will give

$$I(X,Y)=\int\int \mu_{X,Y}(\mathbf{x},y)\log\frac{\mu_{X,Y}(\mathbf{x},y)}{\mu_X(\mathbf{x})\mu_Y(y)}d\mathbf{x}dy \tag{13}$$

There are some available methods to estimate the mutual information, including histogram,[32] bin counting,[33] k-nearest neighbor (k-NN) statistics,[34] and so forth, among which the first two methods are generally restricted to handle one or two-dimensional case, whereas the k-NN statistics can address high-dimensional problems. As the blast furnace modeling refers to high dimension, the k-NN statistics will be used to estimate the probability density.

The k-NN is a type of instance-based learning algorithm. For a set of input–output measurement pairs in the discrete form $\{(\mathbf{x}(i),y(i))\}_{i=1}^N$, the main idea of k-NN is to compute the distance between every reference input–output pair $\{(\mathbf{x}(i),y(i))\}_{i=1}^N$ and its k-NNs point, then within this distance count the input/output points to every reference input/output point $\{\mathbf{x}(i)\}/\{y(i)\}_{i=1}^N$.[34] Here, for simplicity, the direct formula for estimating the mutual information is used

$$I(X,Y)\approx\psi(k)+\frac{1}{N}\sum_{i=1}^{N}[\psi(n_{\mathbf{x}}^{(i)}+1)+\psi(n_y^{(i)}+1)]+\psi(N) \tag{14}$$

where $\psi(\cdot)$ is the digamma function, defined as the logarithmic derivative of the gamma function. Note that k is set in advance and the distance between the input–output point pair $(\mathbf{x}(i),y(i))$, and its k-NNs point pair $(\mathbf{x}(k_i),y(k_i))$ is defined as[34]

$$||(\mathbf{x}(i),y(i))-(\mathbf{x}(k_i),y(k_i))||_\infty=\max(||\mathbf{x}(i)-\mathbf{x}(k_i)||,||y(i)-y(k_i)||) \tag{15}$$

where $||\cdot||$ is the Euclidean norm. $n_{\mathbf{x}}^{(i)}$ appearing in Eq. 14 can be obtained by counting the points $\mathbf{x}(j)(j=1,\cdots,N)$ satisfying

$$||\mathbf{x}(j)-\mathbf{x}(i)||<||(\mathbf{x}(i),y(i))-(\mathbf{x}(k_i),y(k_i))||_\infty \tag{16}$$

and $n_y^{(i)}$ is obtained in the same manner. More details may be found in [34].

Mutual information-based input variables selection

Variables selection plays an important role for reducing redundant information and decreasing model complexity. There are many methods reported for selecting appropriate variables from the original variables, such as stepwise selection, cross-correlation analysis, best subset variable selection and so forth. Here, the main concern is focused on the mutual information-based input variables selection (MIBIVS) method due to its high validity for nonlinear modeling.[29] It gives the optimal inputs subset when the mutual information between this inputs subset in all inputs subsets and output attains the maximum. Assume the original variables set to be $\{\mathbf{x}(i)=(x_1(i),x_2(i),\cdots,x_M(i))\}_{i=1}^N$, then the following procedure serves to screen the optimal inputs subset from this set.[29]

1. Initialization: set $i=1$; num $=K$
2. Select the first input variable x_{r1} satisfying

$$x_{r_1}=\arg\ \max_{x_j}\{I(x_j,y)\}\ 1\le j\le M \tag{17}$$

and denote $\mathcal{A}=\{x_{r_1}\}$

3. Select the next variable $x_{r_{i+1}}$

$$x_{r_{i+1}}=\arg\ \max_{x_j}\{I(x_j,y)\}\quad 1\le j\le M\qquad x_j\notin\mathcal{A} \tag{18}$$

and let $\mathcal{A}=\mathcal{A}\cup\{x_{r_{i+1}}\}$

4. If $i<$ num, $i=i+1$, go to Step **3**;
5. Initialization: set $s=1, x'_{r_1}=x_{r_1}, \mathcal{B}=\{x'_{r_1}\}$;
6. Select the next variable $x'_{r_{s+1}}$ satisfying

$$x'_{r_{s+1}}=\arg\ \max_{x_j}\{I(\{x_j,\mathcal{B}\},y)\}\quad 1\le j\le M\qquad x_j\notin\mathcal{B} \tag{19}$$

and let $\mathcal{B}=\mathcal{B}\cup\{x'_{r_{s+1}}\}$

7. Backward Selection: select the variable x_{rd} needed to be removed from $\mathcal{B}$

$$x_{rd}=\arg\ \max_{x'_{r_k}}\{I(\mathcal{B}/x'_{r_k},y)\},\quad 1\le k\le s+1 \tag{20}$$

where $\mathcal{B}/x'_{r_k}$ means to remove x'_{r_k} from $\mathcal{B}$

If $I(\mathcal{B}/x_{rd},y)>I(\mathcal{B},y)$,let $\mathcal{B}=\mathcal{B}/x_{rd}$

8. If $x'_{r_{s+1}}\in\mathcal{B}$, $s=s+1$, go to Step **6**;
9. Construct the set $\mathcal{L}=\mathcal{A}\cup\mathcal{B}$, and compute the mutual information between every subset of $\mathcal{L}$ and y. The optimal

Table 1. List of the Blast Furnace Variables and the Corresponding Statistics

Symbol	Variable Name	Unit	Mean	St. D
y	Silicon content	wt %	0.63	0.16
x_1	Last silicon content	wt %	0.63	0.16
x_2	Pulverized coal injection rate	Ton/h	23.77	9.04
x_3	Blast volume	nm^3/min[a]	5010.10	292.38
x_4	Gas permeability	nm^3/min kPa	3206.03	160.95
x_5	Blast temperature	°C	1194.26	27.37
x_6	Feed discharged speed	ton/h	225.30	26.57
x_7	Pressure difference	kPa	183.30	19.84
x_8	O_2 enrichment of the blast	vol %	9.86	2.65

[a]where nm^3 represents the normal cubic meters.

inputs subset is a subset of $\mathcal{L}$ that has the maximum mutual information with the output.

Blast Furnace Process

Silicon transfer mechanism

The blast furnace control often means to stringently control the hot metal temperature, silicon, and sulfur contents within acceptable upper and lower bounds, among which the hot metal silicon content plays an important role on evaluating hot metal quality, indicating in-furnace thermal state and reflecting energy consumption. Generally, an increasing silicon implies that the furnace heat level increases, which corresponds to an increasing fuel use and excessive production of heat, whereas a decreasing silicon typically means a cooling of the furnace, which may result in a chilled hearth in the worst situation.[6] Therefore, the main attention is, here, paid to the prediction of the hot metal silicon content for the future control. It is naturally selected as the output of the prediction model.

The silicon inside the blast furnace process is originally in the form of silica, mostly SiO_2, exist in the burdens. As the ironmaking proceeds, SiO_2 is reduced by CO to generate volatile $SiO_{(g)}$, and then $SiO_{(g)}$ continuous to react with carbon saturated iron droplets producing Si, which results in silicon transfer to the iron.[35] The related chemical reactions include

$$(SiO_2)+CO=SiO_{(g)}+CO_2 \tag{21}$$

$$SiO_{(g)}+C=Si+CO \tag{22}$$

$$CO_2+C=2CO \tag{23}$$

Combining these three reactions yields the whole silicon transfer reaction as

$$(SiO_2)+2C=Si+2CO \tag{24}$$

Ozturk and Fruehan[35] pointed out that the silicon content in the hot metal is mainly determined by reaction (22) but not the direct reaction between SiO_2 and carbon saturated iron droplets. The former has very fast rate of silicon transfer while the latter's reaction rate is quite slow.

Data collection and interdependence of variables

Based on the silicon transfer mechanism,[35] the most relevant variables to silicon are those related to the charged solid raw materials and the gaseous materials, so they are expected to be relevant inputs. Intuitively, the more these variables are measured, the richer information we can get for building the silicon prediction model. However, the hostile measurement environment and expensive cost limit these relevant variables measured as many as possible. For the investigated blast furnace, there are totally seven variables related to the solid and gaseous materials that can be collected. They are:

- Pulverized coal injection rate, defined as the weight of the injected pulverized coal per hour;
- Blast volume;
- Blast temperature;
- Feed discharged speed, defined as the speed of material discharging from hopper;
- Pressure difference, defined as the difference between blast pressure and the top gas pressure;
- Gas permeability, defined as the ratio of blast volume to pressure difference;
- Oxygen enrichment of the blast, defined as the ratio of the enriched oxygen volume to blast volume.

A list of the above variables is presented in Table 1, and the corresponding time series of these variables and of the hot metal silicon content are exhibited in Figure 1. The reason of including the last hot metal silicon content as an input is that it has close relation to the current one, and can also capture higher-order dynamics.[8] Of course, inclusion of this autoregressive term will strengthen the inertia of the prediction model so that the model cannot capture the large change of output in time.[9] Table 1 also reports the mean and standard deviation (St. D) of every candidate input and output variables. Clearly, there is great difference in the magnitude of these variables.

All the experimental data are collected from a medium-sized blast furnace with the inner volume of about 2500 m^3, and the size of every time series is 1000 and the sampling interval is about 1.5 h. As a real industrial process, the reliability of input and output data is highly important from the viewpoint of modeling, especially when data-driven model is adopted. For such a traditional industrial process, data collections are generally made with strict demands and the measurements of some blast furnace variables have even formed as the national standards. As an example, the silicon content in hot metal is measured according to some standards, often called ladle-wise analysis in the metallurgical field. To obtain the silicon value of a tapping, two samples are usually taken from the ladle: the first sample is collected when the hot metal flows across one third of the ladle's volume, while the second one is taken when two-thirds of the ladle's volume is full of the hot metal. The arithmetical mean of the silicon measurements of these two samples acts as the final silicon value for this tapping. Therefore, it is credible enough to use the measured data points for the subsequent modeling.

As realistic industrial measurements, the blast furnace data are inevitably contaminated by noise, such as process noise and measurement noise. Directly feeding the noisy data as training set into a data-driven model often makes it have

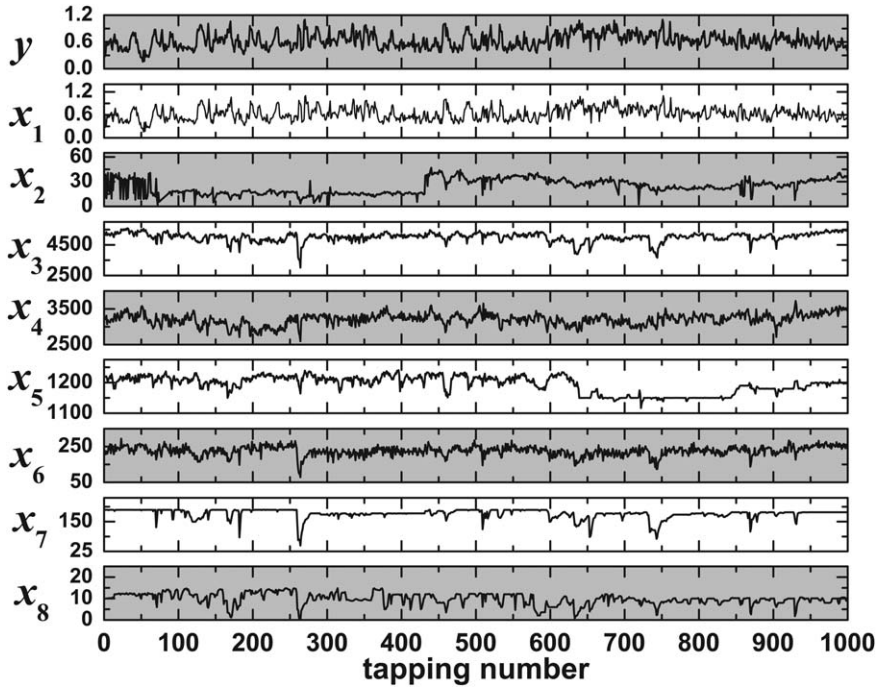

Figure 1. Time series of the original input and output variables collected from the studied blast furnace.

poor generalization performance, such as overfitting and even instability. To circumvent this problem, it needs to perform noise reduction of the blast furnace data. However, it is always a dilemma to reduce industrial process noise due to its unknown statistic properties.[36] The effective component may be filtered from a variable signal if an improper noise reduction method is used. To avoid this point, the measured blast furnace datasets are directly used without preprocessing, but the noise problem will be revisited in the subsequent modeling process.

The above analysis indicates that the listed eight variables in Table 1, $x_1, \cdots, x_8$, will act as the candidate inputs for modeling. As for a model the inputs often play an important role on accuracy and complexity, they are expected to be uncorrelated with each other to avoid redundant information. For this reason, the interdependence of these candidate input variables needs to be analyzed first. As a general measure, mutual information can measure linear and nonlinear relationship, and is, thus, used to quantify the dependence of any two candidate inputs. Given the fact that there are large differences in the magnitudes of the candidate inputs, they are normalized according to

$$\bar{x}_j = \frac{x_j - \text{mean}\,(x_j)}{\max\,(x_j) - \min\,(x_j)} j = 1, \cdots, 8 \tag{25}$$

Table 2 reports the results of the mutual information between any two inputs. Clearly, there is strong dependence between x_7 and x_8, x_2 and x_7, and x_2 and x_5. Hence, it is not a good modeling strategy to include a pair of variables (x_7,x_8), or (x_2,x_7), or (x_2,x_5) as inputs. Table 2 also exhibits the dependence of every input and the output. As expected, the last silicon content x_1 is closely related to the current one, that is, the output y, due to high mutual information. The variable x_8 seems not related to y because of very small mutual information, while the remaining variables have moderate correlation with the output owing to the mutual information of the same magnitude. Principally, the variable exhibiting high correlation with the output has priority to be selected as an input. The results on the interdependence of the blast furnace variables will act as reference for the subsequent multiscale modeling.

To evaluate the performance of the multiscale modeling algorithm more valid, the depicted data points in Figure 1 will be segmented into two groups, the first group containing the preceding 900 points as the training set and the second group containing the remaining 100 points as the testing set.

Multiscale Modeling and Experimental Validation

Multiscale modeling algorithm design

Based on the mentioned methods in Analytical Methods section, it is possible to design a multiscale modeling algorithm that takes into account the contribution of different sublevel scales of a system to the output. Let $\{(\mathbf{x}(i), y(i))\}_{i=1}^{N}$ be the input–output pairs $\mathbf{x}(i) \in \mathbb{R}^M$ and $y(i) \in \mathbb{R}$, then the

Table 2. Interdependence of the Candidate Inputs and the Output through Mutual Information

	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y
x_1	0.12	0.15	0.09	0.13	0.11	0.13	0.07	0.39
x_2		0.26	0.11	0.49	0.13	0.51	0.36	0.11
x_3			0.48	0.15	0.43	0.41	0.25	0.16
x_4				0.07	0.13	0.13	0.11	0.10
x_5					0.06	0.43	0.30	0.10
x_6						0.34	0.19	0.14
x_7							0.60	0.12
x_8								0.06

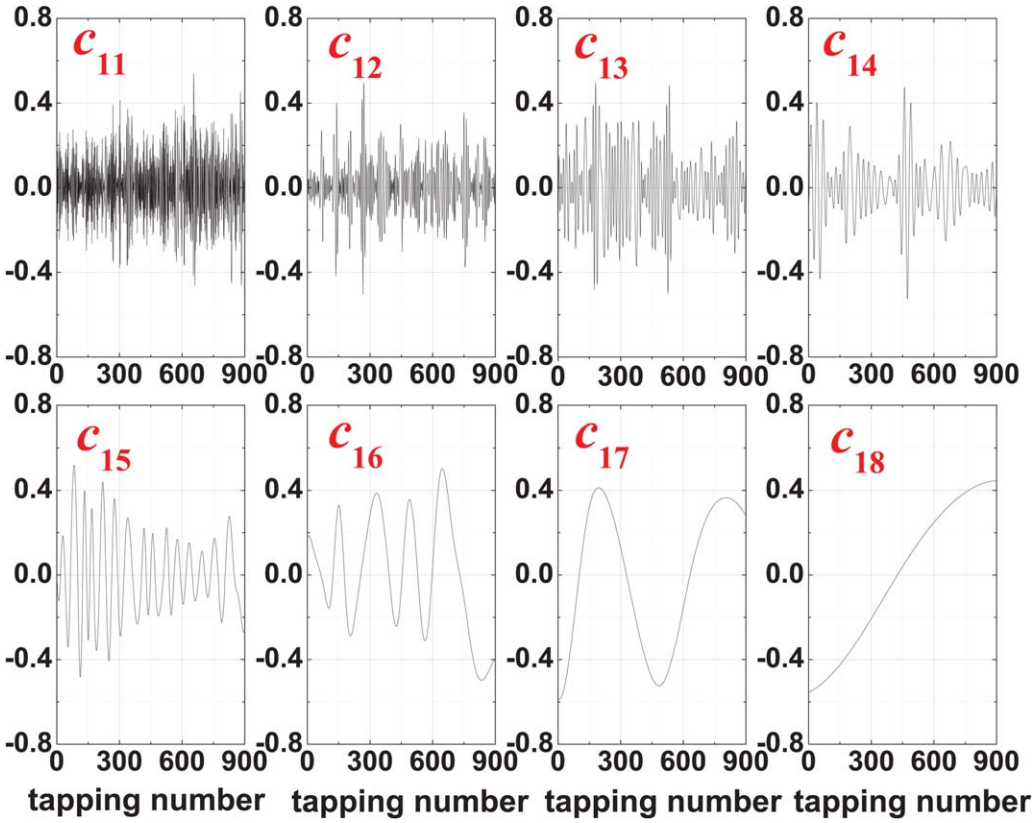

Figure 2. EMD on the time series of the last hot metal silicon content.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

detailed procedure for the multiscale modeling can be described as follows.

Algorithm: Multiscale Modeling Algorithm

Input: M-dimensional variables $\{\mathbf{x}(i) \in \mathbb{R}^M\}_{i=1}^N$
Output: $\{y(i) \in \mathbb{R}\}_{i=1}^N$

1: **for** each series $x_j, j=1,\cdots,M$ **do**
2: Perform the EMD process on the series $\{x_j(i)\}_{i=1}^N$ to obtain the possible subscale signals $\{c_{j1}(i), c_{j2}(i),\ldots,c_{jN_j}(i)\}_{i=1}^N$.
3: **end for**
4: Denote all subscale variables to be

$$U=\{c_{11},\cdots,c_{1N_1},\cdots,c_{M1},\cdots,c_{MN_M}\}.$$

5: **for** every variable in U **do**
6: Perform the variables normalization so that all variables in U are in the same magnitude.
7: **end for**
8: Denote the normalized subscale variables set to be

$$\bar{U}=\{\bar{c}_{11},\cdots,\bar{c}_{1N_1},\cdots,\bar{c}_{M1},\cdots,\bar{c}_{MN_M}\}.$$

9: **for** every variable in $\bar{U}$ **do**
10: Compute the mutual information between this variable and the output y using the k-NN estimator.
11: **end for**
12: Acquire the subset that has the largest mutual information with the output among all normalized subscale variables.
13: Based on the normalized subscale variable acquired in the last step, find the optimal inputs set using the MIBIVS algorithm, denoted as U_{opt}.
14: Feed U_{opt} and the output y into a general linear or nonlinear model framework, and get a corresponding multiscale model.

It should be noted that the above algorithm is not presented in direct reference to tangible blast furnace process variables and control objectives, although it is intended for direct implementation to real-world industrial problems. Instead, some abstract variable symbols, like $\mathbf{x}$ and y, are used to express input and output. The main reason is that the algorithm is expected to be a general modeling method. It does not rely on which the inputs and outputs are, or which blast furnace needs to be modeled, or even on whether the investigated system is a blast furnace. The algorithm can be applied to any real-world industrial problem as long as multiscale feature is exhibited. When it is applied, the input and output of the investigated system are directly introduced into the algorithm framework to yield a multiscale model. Of course, as viewed from the completeness of constructing a data-driven model, this algorithm does not include the necessary procedures for selecting measured variables and preprocessing data. Actually, these two issues are addressed directly referring to the blast furnace variables in Data collection and interdependence of variables section. Therefore, from the viewpoint of the whole multiscale modeling process, there are clear parallels between mathematical formulation of the proposed algorithm and the blast furnace implementation, and meantime, the generality of the algorithm is not lost.

Experimental validation

The hot metal silicon prediction problem of the studied blast furnace will act as a benchmark to test the performance

Table 3. Number of Sublevel Scales of Every Candidate Input Variable Time Series

Variable Symbol	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Numbers of subscales	8	9	9	8	9	9	9	8

of the proposed multiscale modeling algorithm. For this purpose, multiscale EMD of the original eight input variables in the training set is first performed. Note that multiscale decomposition is performed on all candidate input variables without selection according to their interdependence. The main reason is that inputs selection (Step 13) is included in the algorithm after all measured variables are decomposed into subscale ones. The unrelated variables to the output can be still excluded then. As an example, Figure 2 exhibits the decomposed results of the last silicon content. It is clear that there are eight subseries including seven IMFs and the residual. Denote them by $c_{11}, c_{12}, \cdots, c_{18}$ with the first subscript to indicate the variable number and the second subscript to express the subseries number. Also, it should be observed that the decomposed signal tends to be regular as the subseries number increases despite large oscillation emerging in the low IMFs part. Since every subseries is associated with a time scale,[30,37] there are eight physical time scales hidden in the time series of the last silicon content with the low IMFs corresponding to small time scales (high frequency) while the high subseries related to large time scales.[21] Similar to the last silicon content, the sublevel signals of other candidate input variables can be also obtained through the EMD process. Shown in Table 3 are the numbers of the sublevel scales decomposed from every candidate input variable. As can be seen from Table 3, the numbers of the subscales contained in every variable are not the same. Generally, the number of IMFs extracted from a time series depends on the inherent characteristics of this time series, the experimental conditions and the stopping criterion of EMD, among which the stopping criterion is a little subjective because of the subjective selection of a threshold in this criterion,[30] denoted by ε here. Huang et al.[30] suggested to use $\varepsilon \in [0.2, 0.3]$ for a general application. To weaken the subjective effect of ε, a search of ε in[0.2, 0.3] with step 0.01 every time is carried out to observe the change of the numbers of IMFs decomposed from those candidate input variables. The searching results indicate that the current studied cases are not sensitive to ε during this search process. Therefore, ε is finally set to be a routine value 0.2 to perform EMD, and the reported results in Table 3 are corresponding to this ε. From those eight inputs time series, the multiscale decomposition can produce 69 sublevel signals $c_{11}, c_{12}, \cdots, c_{18}, \cdots, c_{81}, \cdots, c_{88}$

Furthermore, the instantaneous frequency hidden in every subscale signal according to Eqs. 6, 7, and 8 was computed, and the respective average one is then obtained as shown in Table 4. Based on these results, a few qualitative conclusions can be drawn to help understand the blast furnace operation, for example, we can know, what dynamical level the future analysis need be made on, and how many sublevel scales a blast furnace process contains approximately.[38] Furthermore, from these results, we can estimate how complex some sublevel structures may be, and what kind of methods may be appropriate to describe them.[38] For example, those low frequency subscales, corresponding to larger temporal scale structures, change steadily and regularly (see Figure 2), so general less sophisticated averaging methods may be enough to deal with them[38]; while for the high frequency subscales, more sophisticated chaos, or random method may be needed.[21] Especially, for the first IMF of every measured variable, it is thought to mainly contain noise,[21,38] which is also embodied in the strong fluctuating behavior emerging in the first subgraph of Figure 2. The noise problem mentioned in Mutual information section can be tackled flexibly through excluding the first IMF of every measured variable as input.

For the decomposed subscale signals as there is very great difference in the magnitudes of those eight original input variables, it naturally leads to the difference among their magnitudes. In general, the subscale variable with a large magnitude will have a stronger effect on the model parameters than the one with a small magnitude. As a result, these 69 subscale variables cannot be directly used for modeling, and the variables normalization is needed. Utilizing Eq. 25, these subscale variables can be normalized to be $\bar{U} = \{\bar{c}_{11}, \bar{c}_{12}, \cdots, \bar{c}_{18}, \cdots, \bar{c}_{81}, \cdots, \bar{c}_{88}\}$, which will form new candidate input set for the subsequent multiscale modeling.

Next, the optimal input variables must be selected. For this, the mutual information between every element in $\bar{U}$ and the output, hot metal silicon content, should be estimated first. For this 69-dimensional problem, the k-NN method is utilized to estimate the mutual information, which, using Eq. 14. As stated in Mutual information section, k must be set in advance for computing the mutual information. Too large k will result in the estimator having a small variance and a large bias, and vice versa. Here, motivated by Rossi et al.,[29] k is set to a midrange value of six, and the mutual information between every normalized subscale variable and the output y is shown in Figure 3. Apparently, $\bar{c}_{69}$ is the most related to the y, so $\bar{c}_{69}$ is the first element for producing $\mathcal{A}$ and $\mathcal{B}$ in the MIBIVS algorithm.[29] Figure 3 also renders a ranking normalized subscale variables from high mutual information to low one with the output, and those having high mutual information with the output will constitute $\mathcal{A}$. Based on $\bar{c}_{69}$, iteratively using Eq. 19 can yield other elements in $\mathcal{B}$. The detailed seeking process may be observed from Figure 4. A look at this figure suggests that the mutual

Table 4. Mean Instantaneous Frequency of Every Subscale Signal (h^{-1})

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
x_1	0.738	0.459	0.237	0.108	0.046	0.006	0.017	0.011	
x_2	0.554	0.484	0.251	0.133	0.067	0.041	0.010	0.005	0.011
x_3	0.706	0.373	0.192	0.092	0.061	0.011	0.001	0.011	0.001
x_4	0.787	0.441	0.212	0.105	0.061	0.012	0.013	0.011	
x_5	0.684	0.585	0.288	0.145	0.085	0.028	0.014	0.011	0.003
x_6	0.821	0.512	0.290	0.131	0.088	0.019	0.022	0.003	0.001
x_7	0.505	0.324	0.237	0.115	0.072	0.028	0.021	0.004	0.003
x_8	0.541	0.330	0.155	0.098	0.061	0.031	0.009	0.002	

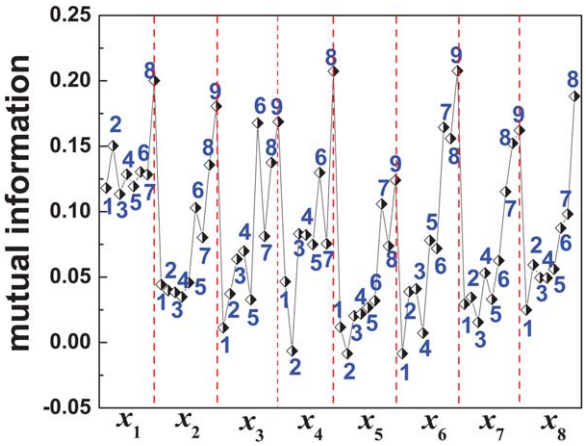

Figure 3. Mutual information between every normalized subscale variable and the output.

The Arabic numerals indicate the subscale numbers of the corresponding variable. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

information will attain the maximum after $\bar{c}_{79}$ is included in $\mathcal{B}$, at this moment $\mathcal{B}=\{\bar{c}_{69},\bar{c}_{14},\bar{c}_{12},\bar{c}_{11},\bar{c}_{13},\bar{c}_{38},\bar{c}_{16},\bar{c}_{79}\}$. To avoid obtaining a local optimal $\mathcal{B}$, a backward procedure is performed on $\mathcal{B}$ to remove some redundant elements, but no elements need to be removed. As mentioned in Mutual information based input variables selection section, it is difficult to stop seeking the elements in $\mathcal{A}$. A common method is to limit the number of the elements in $\mathcal{L}=\mathcal{A}\cup\mathcal{B}$ beforehand, denoted as m, then to select those variables having high mutual information with the output but different from the elements $\mathcal{B}$ to form $\mathcal{A}$.[29] The setting of m is usually the tradeoff between the algorithm precision and efficiency. As there are eight normalized subscale variables in $\mathcal{B}, m$ is set as 12 here. Even so, there will be 2^{12} subset of $\mathcal{L}$ to be searched for the optimal inputs set. With these constraints, $\mathcal{A}$ can be obtained to be $\{\bar{c}_{48},\bar{c}_{18},\bar{c}_{88},\bar{c}_{29}\}$, hence $\mathcal{L}=\{\bar{c}_{69},\bar{c}_{14},\bar{c}_{12},\bar{c}_{11},\bar{c}_{13},\bar{c}_{38},\bar{c}_{16},\bar{c}_{79},\bar{c}_{48},\bar{c}_{18},\bar{c}_{88},\bar{c}_{29}\}$. Further performing the procedure of the MIBIVS algorithm will produce the optimal inputs set to be $U_{\text{opt}}=\{\bar{c}_{11},\bar{c}_{12},\bar{c}_{13},\bar{c}_{14},\bar{c}_{16},\bar{c}_{29},\bar{c}_{69}\}$, which has the mutual

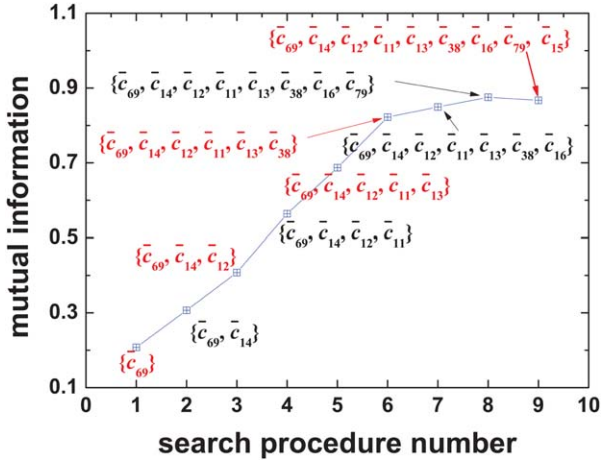

Figure 4. Search process for the elements in $\mathcal{B}$ by the MIBIVS algorithm.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

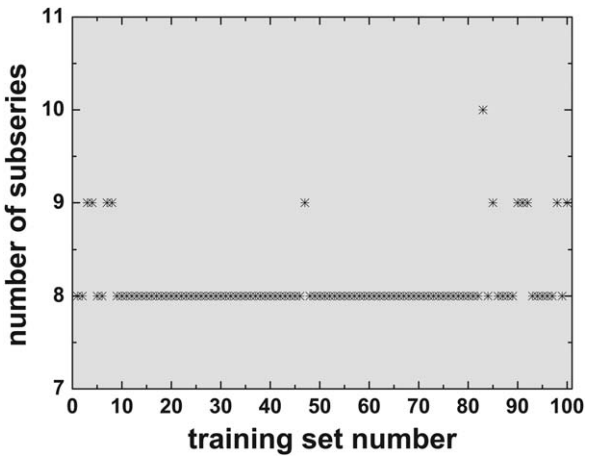

Figure 5. Numbers of the subseries resolved from the last silicon content at different training set as the sliding window moves.

information to the output 0.8755. Clearly, the optimal input set contains information only related to three original variables, x_1: the last silicon content, x_2: the pulverized coal injection rate and x_6: the feed discharged speed. Parts of their sublevel scales will mainly contribute to the evolution of the hot metal silicon content, as far as the studied blast furnace is concerned. A further look at $U_{\text{opt}}=\{\bar{c}_{11},\bar{c}_{12},\bar{c}_{13},\bar{c}_{14},\bar{c}_{16},\bar{c}_{29},\bar{c}_{69}\}$ suggests that most of subscale variables come from the last silicon content, and only one subscale variable from the pulverized coal injection rate and one from the feed discharged speed. This confirms again the fact that the last silicon content will affect the current one much. Linking these subscale variables with their frequencies (0.738, 0.459, 0.237, 0.108, 0.006, 0.011, 0.001) reported in Table 4 reveals that these variables almost cover all representative subscales information contained in the studied blast furnace system. c_{11}, c_{12}, c_{13}, and c_{14} have high frequencies, and thus, capture the oscillating behavior of the silicon evolution, whereas c_{16}, c_{29}, and c_{69} capture the slow level fluctuations or gradual level changes of the silicon evolution due to having low frequencies. Moreover, their frequencies look quite diffusive between the minimum 0.001 (c_{69}) and the maximum 0.821 (c_{61}) in Table 4. So the optimal input set includes enough subscales information but avoiding redundancy for constructing a multiscale model of the studied blast furnace system. Good performance is expected for the constructed model with such input set. From the viewpoint of noise, only the first IMF of the last silicon content is included in U_{opt}, so most of noise contained in the blast furnace measured variables has no effect on the blast furnace multiscale model. In addition, combining the optimal input set with Table 2 suggests that the most relevant variable to the output, x_1, is expected to be included as an input while the most irrelevant variable x_8 is excluded. Furthermore, by studying the mutual information presented in Table 2, one may conclude that little correlation can be observed between the selected three input variables x_1, x_2, and x_6.

The optimal inputs set U_{opt} and the output y will be fed into a general linear and nonlinear model framework to verify the validity of the multiscale modeling algorithm. Here, two kinds of linear models including multiple linear regression (MLR) and multiple stepwise linear regression (MSLR) and two kinds of nonlinear models including radial basis

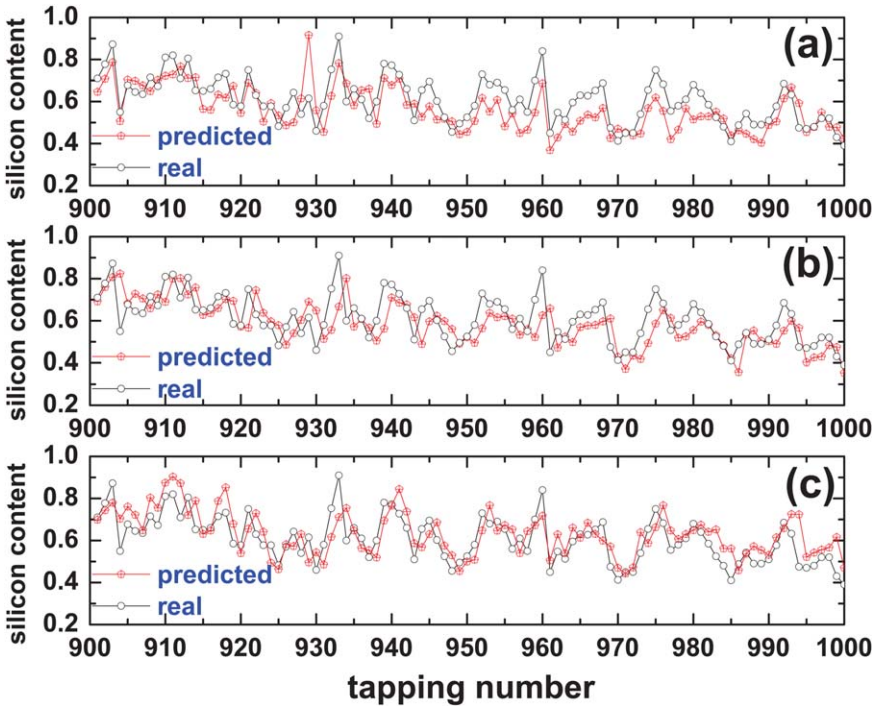

Figure 6. Silicon prediction results in the RBFN framework: (a) RBFN; (b) RBFN$_{vs}$; and (c) multiscale-RBFN.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

function network (RBFN)[39] and the least-square-support vector machines (LS-SVM)[40] are taken as the reference models framework, among which the RBFN is defined as a linear combination of some Gaussian kernels and the LS-SVM is with Gaussian kernel. These models with the input–output pairs $\{(U_{\text{opt}}, y)\}$ are identified with adding a prefixion "Muliscale-" in the context. The training set part of U_{opt} and y are used to estimate the model parameters for these four models while their testing sets are used to validate the models performance. To avoid the slow learning resulting from the increase of the size of the training sets, a sliding window technique is adopted through which the size of the training sets keeps unchanged but the training sets are continuously updated by including the latest measured input–output pair

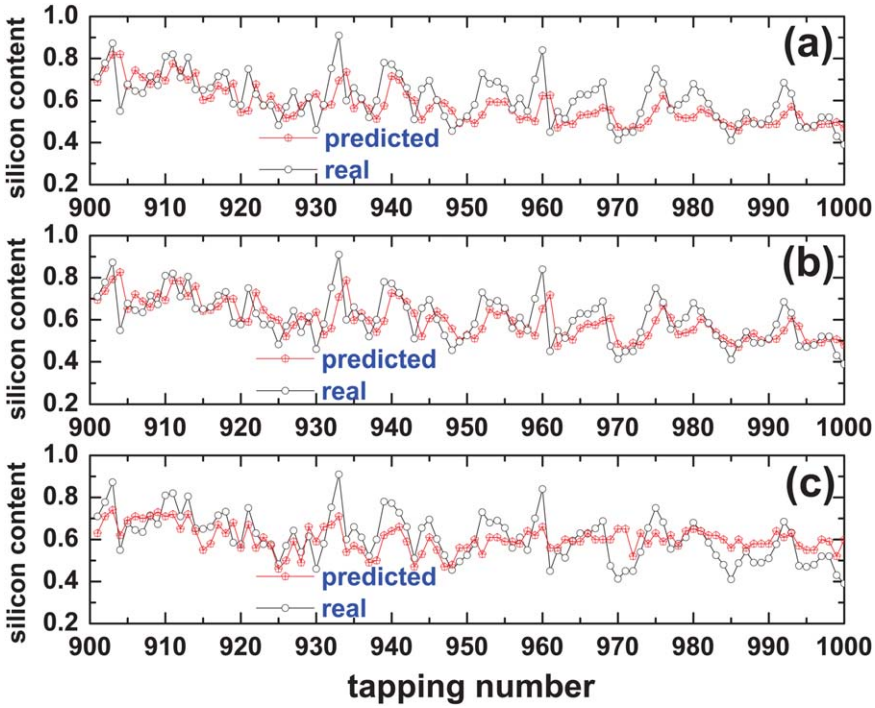

Figure 7. Silicon prediction results in the LS-SVM framework: (a) LS-SVM; (b) LS-SVM$_{vs}$; and (c) multiscale-LS-SVM.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 5. Trained Parameters in the RBFN and LS-SVM Frameworks

Model Framework	Model Name	Parameters[a]
RBFN framework	RBFN	$\tau=0.50$
	$RBFN_{vs}$	$\tau=0.45$
	Multiscale-RBFN	$\tau=0.52$
LS-SVM framework	LS-SVM	$(\upsilon,\sigma^2)=(0.05,15)$
	$LS\text{-}SVM_{vs}$	$(\upsilon,\sigma^2)=(0.10,20)$
	Multiscale-LS-SVM	$(\upsilon,\sigma^2)=(0.46,50)$

[a]τ is the tolerance chosen by trial and error, an important parameter in balancing the accuracy and the complexity of the final RBFN; υ is the regularization parameter and σ^2 is the Gaussian kernel parameter.

and removing the first input–output pair. It should be pointed out that during every new prediction task on the testing set, the EMD process will be carried out on the updated training sets according to the determined numbers of subscales listed in Table 3. To observe if the numbers of subscales decomposed from EMD will change as the training sets are updated, the last silicon content is taken as an example for this purpose. Figure 5 presents the corresponding results, most of which are consistent with what has been obtained from the initial training set. Therefore, it is reasonable to perform the EMD process on the continuously updated training sets according to those determined numbers of subscales listed in Table 3. Additionally, the procedure for selecting input variables is also left out, using the subscale variables with the same subscripts as those in U_{opt} produced by the initial training sets directly act as the candidate inputs. These candidate inputs are normalized renewedly according to Eq. 25, and are then introduced into the above four models with the same models parameters as those learned from the initial training sets to predict the next hot metal silicon content. The lower panels of Figures 6 and 7 illustrate the silicon prediction results on the testing set by the multiscale-RBFN and multiscale-LS-SVM models, respectively. Here, the main concern is made on the nonlinear models, because this kind of models are thought to be more suitable for describing the complex dynamics of the blast furnace system.

To demonstrate the good performance of the multiscale models more fully, the original input variables (after normalization) are directly fed into those two nonlinear reference models to make silicon prediction as comparisons, including two cases, one case without input variables selection and the other case with input variables selection through the mutual information technique, identified by adding the subscript "vs." For the latter case, the computation of the mutual information indicates that the combination $\{x_1,x_2,x_3,x_5\}$ has the largest mutual information to the output among all the subsets of $\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}$, and the largest mutual information is 0.5118, so the variables set $\{x_1,x_2,x_3,x_5\}$ acts as the inputs for the $RBFN_{vs}$ and $LS\text{-}SVM_{vs}$ models. For the models without input variables selection, the mutual information between the inputs set $\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}$ and the output is 0.4565. The prediction results in these two cases are also given in Figures 6 and 7. It should be noted that during every prediction as the sliding window moves, the models are time-invariant. The trained parameters learned from the initial training set serve for all predictions in the testing set. Table 5 gives the trained parameters in the nonlinear model frameworks. From Figures 6 and 7, the following information is clear: (1) in the case of the RBFN model framework, the RBFN model (Figure 6a) cannot capture the evolution of the silicon series well, and there are a lot of disagreement between the real values and the predicted ones, but the delay phenomena, that is, the predicted values lagging the real ones, are not serious except around the tapping numbers 993. The variables selection seems not to bring great improvement on the $RBFN_{vs}$ model precision (Figure 6b). Moreover, the delay phenomena become serious, such as around the tapping number 940, 953, 976, 993, and so forth. For the multiscale-RBFN model (Figure 6c), the predictions obviously become good, only a little disagreement appears. However, there are still serious delay phenomena taking place, such as around the tapping numbers 934, 941, 953, 976, 993, and so forth. (2) in the case of the LS-SVM model framework, the occasions are very similar to those of the RBFN framework. Namely, neither the LS-SVM model nor the $LS\text{-}SVM_{vs}$, one can work satisfactorily for the silicon prediction of the studied blast furnace. In a similar manner, the multiscale-LS-SVM model (Figure 7c) can basically follow the silicon change, but this multiscale pattern is slightly delayed and is strongly damped at the end of the interval.

Discussion

To quantitatively evaluate the models performance, the agreement between the real values and the predicted ones is measured in the light of the following two criteria, that is, the hitting rate

$$P=\frac{1}{N_t}\sum_{i=1}^{N_t}H(i)\times 100\% \tag{26}$$

where N_t is the size of the testing set and

$$H(i)=\begin{cases}1, \text{if } |\hat{y}(i)-y(i)|<0.1\\ 0, \text{if } |\hat{y}(i)-y(i)|\geq 0.1\end{cases} \tag{27}$$

and the root mean square error (RMSE)

$$\text{RMSE}=\sqrt{\frac{1}{N_t}\sum_{t=1}^{N_t}(\hat{y}(i)-y(i))^2} \tag{28}$$

The hitting rate is the most concerning criterion in the metallurgical field as it can provide important indication of controlling blast furnace operations, whereas the RMSE is a common criterion for evaluating models precision. Based on these two criteria, the performance of every kind of model can be measured, and the results are reported in Table 6. As Figures 6 and 7 might suggest, those two kinds of nonlinear models in multiscale manner can outperform the corresponding models at a fixed scale completely, an obvious increase in the hitting rate, a slight decrease in the RMSE, and less variables

Table 6. Performance Comparisons of the Multiscale Models and the Corresponding Models at a Fixed Scale

Model Category	Model Name	Hitting Rate (%)	RMSE
Linear case	MLR	67	0.106
	Multiscale-MLR	71	0.120
	MSLR	68	0.104
	Multiscale-MSLR	71	0.120
Nonlinear case	RBFN	72	0.094
	$RBFN_{vs}$	74	0.093
	Multiscale-RBFN	86	0.084
	LS-SVM	76	0.096
	$LS\text{-}SVM_{vs}$	77	0.092
	Multiscale-LS-SVM	82	0.089

(only three) needed to be measured from the blast furnace system. Great advantage is reflected by considering the multiscale features for the silicon prediction models, which in turn renders that the blast furnace system contain multiscale characteristics. Moreover, as far as the multiscale RBFN model is concerned, the high hitting rate 86% and low RMSE 0.084 makes it a candidate for the practical silicon prediction task. The reliable prediction results can provide important guide for blast furnace operator to take control actions. By comparing the predicted silicon to the desired one, the operator is able to estimate the direction that the next silicon content will deviate so that a reasonable control action can be decided to make the silicon content increase or decrease. At this point, the output information of the multiscale-RBFN model is helpful in enhancing furnace operation. Further comparing the multiscale-RBFN model with our early chaotic model[11] which has the hitting rate and the RMSE to be 79.04% and 0.0854, respectively, indicates that the former still has rather large competitive power. More importantly, the current multiscale-RBFN model is trivial from the viewpoint of the RBFN structure. The model is in the original RBFN form, and is also time-invariant as the training sets are updated, so there is much room for further improvement.

In addition, the results reported in Table 6 also indicate that in the same nonlinear model framework, the precision improvement from a fixed scale model to a multiscale model mainly originate from the multiscale decomposition of every original variable but not the input variables selection as there is only slight improvement from the RBFN (LS-SVM) model to the RBFN_{vs} ($\text{LS-SVM}_{\text{vs}}$) model. The inherent reason leading to the above great improvement may be the increase of the mutual information between the inputs and the output, 0.8755 in the multiscale models, which is greater than 0.5118 in the fixed scale models with input variables selection and 0.4565 in the ones without input variables selection. At the same time, the mutual information results can be also used to account for that there is only slight improvement from the RBFN (LS-SVM) model to the RBFN_{vs} ($\text{LS-SVM}_{\text{vs}}$) as the mutual information increase is slight. Although the multiscale models outperform the fixed scale models completely in the nonlinear case, it is not always true in the linear case. The RMSE criterion in Table 6 states that the multiscale-MLR and multiscale-MSLR models have poorer performance than the respective fixed scale models despite a little increase in the hitting rate. Moreover, the 71% hitting rate implies that the studied multiscale linear models are hardly competent for the silicon prediction. The possible reason may be that in the linear model framework, the relationship between every input and the output is assumed to be linear, but the EMD process and the MIBIVS method are both nonlinear. Their implementation may destroy the supposed linear relationship, and thus cannot produce positive effect on the succeeding multiscale models. To verify this point, Figure 8 measures the linear relationship between every subseries resolved from the original eight blast furnace input variables using the Pearson's correlation coefficient. The results indicate that the first seven subscale variables most linearly dependent on the output are $\{\bar{c}_{13}, \bar{c}_{12}, \bar{c}_{14}, \bar{c}_{11}, \bar{c}_{78}, \bar{c}_{69}, \bar{c}_{43}\}$, which are not the complete same with what has been obtained by the MIBIVS method, $\{\bar{c}_{11}, \bar{c}_{12}, \bar{c}_{13}, \bar{c}_{14}, \bar{c}_{16}, \bar{c}_{29}, \bar{c}_{69}\}$. Namely, the latter is not the optimal input set for a linear model framework. However, the linear cases cannot put off the development and further applications of the proposed multiscale modeling algorithm

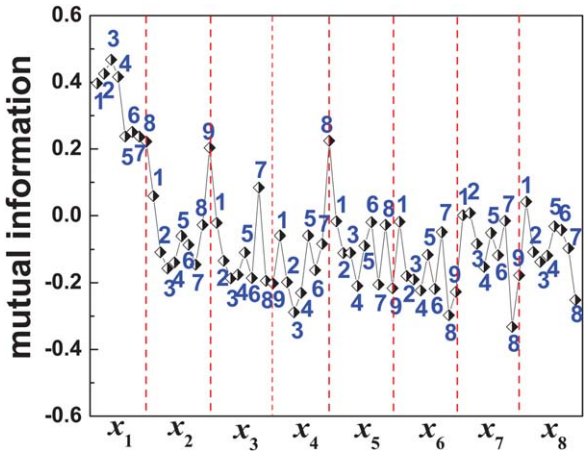

Figure 8. Pearson's correlation coefficients between every normalized subscale variable and the output.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

as nonlinearity is an inherent characteristic of a multiscale system and nonlinear models are expected to have greater possibility to capture the complex blast furnace dynamics.

Of course, it has to be mentioned that it will take more time to do a prediction using the mutliscale models than the corresponding fixed scale models due to performing EMD and variables selection. In this study, the multiscale-RBFN and multiscale-LS-SVM models will spend about 5 min doing a prediction after determining the optimal inputs set during the training stage, which is much lower than the silicon sampling interval, 1.5 h, so the constructed multiscale models are of practical use for predicting the hot metal silicon content. Additionally, there are some delay phenomena in the constructed multiscale models, the main reason of which may be the relative more subscales resolved from the last silicon content in the optimal inputs set, including five subscales $\{\bar{c}_{11}, \bar{c}_{12}, \bar{c}_{13}, \bar{c}_{14}, \bar{c}_{16}\}$. These autoregressive terms are thought as the source of high inertial in the predictive models.[9] Weakening the delay phenomenon may further give room for improvement of the multiscale models, which will be the focus of our future investigations.

Conclusions and Points of Future Research

In this work, a multiscale modeling algorithm that can fully consider the multiscale features of the blast furnace system is developed to address the prediction of the silicon content of hot metal in the blast furnace. Through this algorithm, the original blast furnace variables can be decomposed into sublevel scale variables from which the most relevant subscale variables to the output are singled out, and then fed into a general linear or nonlinear model framework to construct the corresponding multiscale model. During the modeling process, the contribution of different subscales of the blast furnace system to the output is distinguished effectively. Four kinds of reference model frameworks including two linear models (MLR, MSLR) and two nonlinear models (RBFN, LS-SVM) are used to perform the blast furnace experimental validation. The results indicate that the multiscale-RBFN and multiscale-LS-SVM can outperform completely the respective models at a fixed scale no matter whether the variables selection is made

or not. Moreover, the inherent reason for the above results is also given by comparing the mutual information between the respective model inputs and the output. As far as the constructed multiscale-RBFN model is concerned, the high hitting rate, low RMSE, and less input variables needed to be measured from the blast furnace system make it full of potential for the silicon prediction. More importantly, there is much room for further improvement of the blast furnace multiscale models, including improvement of the reference models, considering the time-variant manner, weakening the inertia resulting from the selection of autoregressive terms as inputs and so forth.

Also, it should be pointed out that the proposed multiscale algorithm is not limited to being used for the studied blast furnace. It is inherently a general method, and easily transferable to other metallurgical plants instrumented well or poorly. As far as the silicon prediction is concerned, the input variables can be any ones related to the charged solid raw materials and the gaseous materials but not necessary to be the same with those listed in Table 1. In actual practice, as long as the collected inputs and output of the blast furnace system are introduced into the algorithm framework, a multiscale model is available. To demonstrate this, data from another poorly instrumented blast furnace with a volume of 750 m^3 and sampling interval of 2 h was used for validation. As expected, the proposed multiscale algorithm also exhibits apparent advantage like that observed from the current blast furnace. The details are presented in the Appendix. More broadly, the multiscale algorithm can be extended to other chemical processes with multiscale features. In like manner, the multiscale model can be developed by feeding the inputs and outputs of the studied systems into the algorithm framework.

Of course, as a kind of black-box modeling methods, the data-based multiscale modeling approach can provide important guide for the blast furnace section chief to take control action, but it does not provide the detailed information on how to control the silicon content. In this respect, the proposed method can be said just for the silicon prediction as a soft sensor. However, the prediction results might be utilized to enhance furnace operation with the help of metallurgical know-how and blast furnace expert knowledge. For example, a prediction of the next silicon content is said to be high, which means the consumption of the pulverized coal injection rate is high according to the metallurgical know-how, so it requires to reduce the pulverized coal injection rate such that the next silicon content will not be too high. The reduction amount may be estimated based on the blast furnace expert knowledge. At this point, the multiscale model can work more effectively if combing some metallurgical know-how. Actually, in the process of selecting the blast furnace measured variables as inputs, the metallurgical know-how has been invoked. Another means to promote the current model serving for the furnace operation is to enhance its transparency by rules extraction. The extracted rules (e.g., if $\text{input}_1, \cdots, \text{input}_n$, then output) can account for the output results with detailed and definite inputs information, which may further serve for control purpose by linking the output results with controlled variables. Direct indication on how to adjust inputs can be provided. The corresponding work is under way. Notwithstanding some limitation, it is expected that the proposed multiscale modeling algorithm can serve as a useful framework for constructing a data-based multiscale model, and also as an elicitation to develop a blast furnace mathematical model of practical use.

Acknowledgments

The authors acknowledge the financial support by the National Natural Science Foundation of China under Grant No. 11271326, the Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20130101110040, and the Fundamental Research Funds for the Central Universities.

Literature Cited

1. Nogami H, Chu MS, Yagi J. Multidimensional transient mathematical simulator of blast furnace process based on multi-fluid and kinetic theories. *Comput Chem Eng*. 2005;29:2438–2448.
2. Chu MS, Yang XF, Shen FM, Yagi J, Nogami H. Numerical simulation of innovative operation of blast furnace based on multi-fluid model. *J. Iron Steel Res Int*. 2006;13:8–15.
3. Jindal A, Pujari S, Sandilya P, Ganguly S. A reduced order thermochemical model for blast furnace for real time simulation. *Comput Chem Eng*. 2007;31:1484–1495.
4. Radhakrishnan VR, Mohamed AR. Neural networks for the identification and control of blast furnace hot metal quality. *J Process Control*. 2000;10:509–524.
5. Chen J. A predictive system for blast furnaces by integrating a neural network with qualitative analysis. *Eng Appl Artif Intell*. 2001;14:77–85.
6. Waller M, Saxén H. Time-varying event-internal trends in predictive modeling methods with applications to ladlewise analyses of hot metal silicon content. *Ind Eng Chem Res*. 2003;42:85–90.
7. Bhattacharya T. Prediction of silicon content in blast furnace hot metal using partial least squares (PLS). *ISIJ Int*. 2005;45:1943–1945.
8. Pettersson F, Chakraborti N, Saxén H. A genetic algorithms based multi-objective neural net applied to noisy blast furnace data. *Appl Soft Comput*. 2007;7:387–397.
9. Saxén H, Pettersson F. Nonlinear prediction of the hot metal silicon content in the blast furnace. *ISIJ Int*. 2007;47:1732–1737.
10. Jian J, Gao CH, Li L, Zeng JS. Application of least squares support vector machines to predict the silicon content in blast furnace hot metal. *ISIJ Int*. 2008;48:1659–1661.
11. Gao CH, Chen JM, Zeng JS, Liu XY, Sun YX. A chaos-based iterated multistep predictor for blast furnace inromaking process. *AIChE J*. 2009;55:947–962.
12. Jian L, Gao CH, Xia ZQ. A sliding-window smooth support vector regression model for nonlinear blast furnace system. *Steel Res Int*. 2011;82:169–179.
13. Zhao J, Wang W, Liu Y, Pedrycz W. A two-stage online prediction method for a blast furnace gas system and its application. *IEEE Trans Control Syst Technol*. 2011;19:507–520.
14. Nurkkala A, Pettersson F, Saxén H. Nonlinear modeling method applied to prediction of hot metal silicon in the ironmaking blast furnace. *Ind Eng Chem Res*. 2011;50:9236–9248.
15. Gao CH, Jian L, Luo SH. Modeling of the thermal state change of blast furnace hearth with support vector machines. *IEEE Trans Ind Electron*. 2012;59:1134–1145.
16. Warren P, Harvey S. Development and implementation of a generic blast-furnace expert system. *Trans Inst Min Metall Section C-Mineral Process Extractive Metall*. 2001;110:C43–C49.
17. Johansson A, Medvedev A. Detection of incipient clogging in pulverized coal injection lines. *IEEE Trans Ind Appl*. 2000;36:877–883.
18. Miyano T, Kimoto S, Shibuta H, Nakashima K, Ikenaga Y, Aihara K. Time series analysis and prediction on complex dynamical behavior observed in a blast furnace. *Physica D*. 2000;135:305–330.
19. Saxén H, Östermark R. State realization with exogenous variables-a test on blast furnace data. *Eur J Oper Res*. 1996;89:34–52.
20. Chu YX, Gao CH, Liu XG. Multiscale dynamic analysis of blast furnace system based on intensive signal processing. *Chaos*. 2010;20:033102.
21. Gao CH, Zeng JS, Zhou ZM. Identification of multiscale nature and multiple dynamics of the blast furnace system from operating data. *AIChE J*. 2011;57:3448–3458.
22. Benveniste A, Nikoukhah R, Willsky AS. Multiscale system theory. *IEEE Trans Circuits Syst I-Fundam Theor Appl*. 1994;41:2–15.
23. Bakshi BR. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE J*. 1998;44:1596–1610.
24. Reis MS, Saraiva PM. Multiscale statistical process control with multiresolution data. *AIChE J*. 2006;52:2107–2119.

25. Reis MS, Saraiva PM, Bakshi BR. Multiscale statistical process control using wavelet packets. *AIChE J*. 2008;54:2366–2378.
26. Li X, Rankin SE. Multiscale dynamic Monte Carlo/continuum model of drying and nonideal polycondensation in sol-gel silica films. *AIChE J*. 2010;56:2946–2956.
27. Balaji B, Du J, White CM, Ydstie BE. Multi-scale modeling and control of fluidized beds for the production of solar grade silicon. *Powder Technol*. 2010;199:23–31.
28. Li JH, Kwauk M. Exploring complex system in chemical engineering-the multi-scale methodology. *Chem Eng Sci*. 2003;58:521–535.
29. Rossi F, Lendasse A, Francois D, Wertz V, Verleysen M. Mutual information for the selection of relevant variables in spectrometric nonlinear modeling. *Chemom Intell Lab Syst*. 2006;80:215–226.
30. Huang NE, Shen Z, Long SR, Wu MD, Shih HH, Zhen Q, Yen NC, Tung CC, Liu HH. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc London A*. 1998;454:903–995.
31. Shannon CE, Weaver W. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.
32. Scott DW. *Multivariable Density Estimation: Theory, Practice, and Visualization*. New York: Wiley, 1992.
33. Nichols JM, Nichols JD. Attractor reconstruction for non-linear systems: a methodological note. *Math Biosci*. 2001;171:21–32.
34. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev E*. 2004;69:066138.
35. Ozturk B, Fruehan RJ. Kinetics of the reaction of $SiO_{(g)}$ with carbon saturated iron. *Metall Trans B*. 1985;16:121–127.
36. Gao CH, Qian JX. Evidence of chaotic behavior of nosis from industrial process. *IEEE Trans Signal Process*. 2007;55:2877–2884.
37. Huang NE, Wu ML, Qu WD, Steven RL, Shen SP. Applications of Hilbert-Huang transform to non-stationary financial time series analysis. *Appl Stochastic Models Business Ind*. 2003;19:245–268.
38. Briongos JV, Aragon JM, Palancar MC. Phase space structure and multi-resolution analysis of gas-solid fluidized bed hydrodynamics: Part I-The EMD approach. *Chem Eng Sci*. 2006;61:6963–6980.
39. Chen S, Cowan CFN, Grant PM. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans Neural Networks*. 1991;2:302–309.
40. Suykens JAK, Lukas L, Vandewalle J. Sparse least squares support vector machine classifiers. In; Proceedings of the European Symposium on Artificial Neural Networks ESANN'2000, Bruges, Belgium, 2000:37–42.

Appendix

In this Appendix, the proposed multiscale modeling algorithm is applied to another blast furnace system with the volume of 750 m^3 and the sampling interval of 2 h. To make a difference from the previous blast furnace, this small one is denoted as SBF in the following. Table A1 presents the seven measured variables as candidate

Table A1. Measured Variables and Their Statistics of the SBF

Symbol	Variable Name	Unit	Mean	St. D
y	Silicon content	wt %	0.45	0.10
x_1	Last silicon content	wt %	0.45	0.10
x_2	Pulverized coal injection rate	ton/h	12.75	2.16
x_3	Blast volume	nm^3/min*	1705.08	292.38
x_4	Gas permeability	nm^3/min kPa	15.84	1.53
x_5	Blast temperature	°C	1006.36	55.11
x_6	Feed discharged speed	ton/h	86.36	11.98
x'_7	Iron difference	ton	9.75	58.51

*where nm^3 represents the normal cubic meters.

Table A2. Trained Parameters in the RBFN and LSSVM Frameworks for the SBF

Model Framework	Model Name	Parameters*
RBFN framework	RBFN	$\tau = 0.47$
	$RBFN_{vs}$	$\tau = 0.45$
	Multiscale-RBFN	$\tau = 0.53$
LS-SVM framework	LS-SVM	$(\upsilon, \sigma^2) = (0.05, 20)$
	$LS\text{-}SVM_{vs}$	$(\upsilon, \sigma^2) = (0.10, 10)$
	Multiscale-LS-SVM	$(\upsilon, \sigma^2) = (0.10, 25)$

*τ is the tolerance chosen by trial and error, an important parameter in balancing the accuracy and the complexity of the final RBFN; υ is the regularization parameter and σ^2 is the Gaussian kernel parameter.

Table A3. Performance Comparisons of the Multiscale Models and the Corresponding Models at a Fixed Scale for the SBF

Model Category	Model Name	Hitting Rate (%)	RMSE
Linear case	MLR	77	0.108
	Multiscale-MLR	77	0.115
	MSLR	78	0.107
	Multiscale-MSLR	77	0.115
Nonlinear case	RBFN	79	0.104
	$RBFN_{vs}$	81	0.103
	Multiscale-RBFN	88	0.073
	LS-SVM	79	0.105
	$LS\text{-}SVM_{vs}$	79	0.105
	Multiscale-LS-SVM	83	0.087

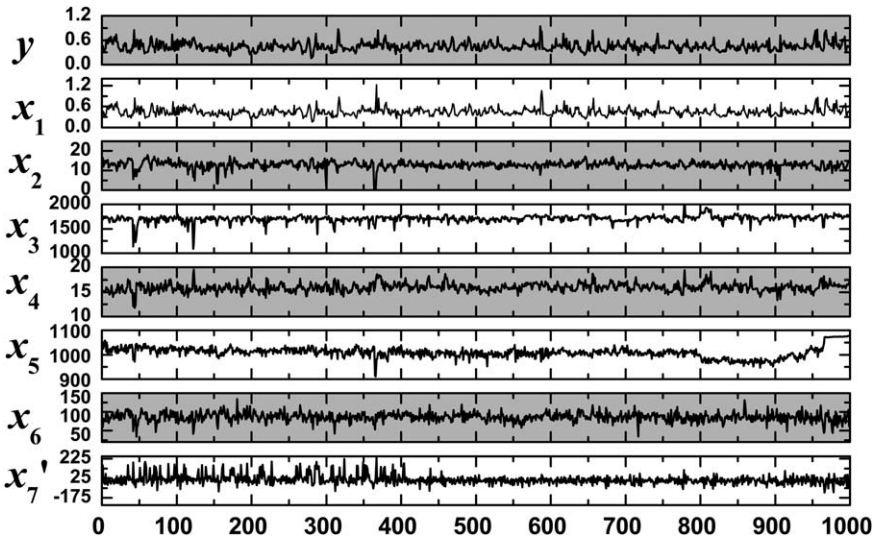

Figure A1. Time series of the original input and output variables collected from the SBF.

inputs and their statistics of the SBF. The first six variables $x_1, \cdots, x_6$ share the same inputs as those of the previous blast furnace while the last variable "Iron difference" differs and is defined by the difference between the theoretical tapping weight and the actual tapping weight. Figure A1 exhibits the evolution of these variables time series. As what have been done on the previous blast furnace, the same experiments are made on the SBF to yield the optimal input set to be $U_{\text{opt}} = \{\bar{c}_{12}, \bar{c}_{13}, \bar{c}_{14}, \bar{c}_{25}, \bar{c}_{46}, \bar{c}_{69}\}$, and the trained parameters to be shown in Table A2. The inputs for the RBFN_{vs} model and $\text{LS-SVM}_{\text{vs}}$ model are $\{x_1, x_2, x_4, x_6\}$. The performance comparisons of the multiscale models and the corresponding models at a fixed scale for this blast furnace are reported in Table A3. Clearly, the multiscale modeling strategy also exhibits advantage in the Hitting Rate and RMSE.

Manuscript received Apr. 22, 2014, and revision received Feb. 9, 2014.