Information Sciences 193 (2012) 104–114

Contents lists available at SciVerse ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Data-driven based model for flow prediction of steam system in steel industry

Ying Liu [a], Quanli Liu [a,*], Wei Wang [a], Jun Zhao [a], Henry Leung [b]

[a] Faculty of Electronics Information and Electrical Engineering, Dalian University of Technology, China
[b] Department of Electrical and Computer Engineering, University of Calgary, Canada

A R T I C L E   I N F O

Article history:
Received 10 March 2011
Received in revised form 23 October 2011
Accepted 31 December 2011
Available online 8 January 2012

Keywords:
Steam system
Data-driven
Time series prediction
Bayesian ESN

A B S T R A C T

The steam system is one of the main energy systems in steel industry, and its operational scheduling plays a crucial role for energy utility and resources saving. For a reasonable resources operation, the accurate prediction of steam flow is required. Considering the large amount of production data in energy system, a data-driven based model is proposed to perform a time series prediction for steam flow, in which a Bayesian echo state network (ESN) is established. This method combines Bayesian theory with ESN to obtain optimal output weight via maximizing the posterior probability density of the weights to avoid over-fitting in the training process of sample data. To pursue optimized hyper-parameters in the proposed Bayesian ESN, the evidence framework based on sample data is further adopted in this work. Experimental results using the real production data from Shanghai Baosteel show the validity and practicality of the proposed data-driven based model in providing scientific decision guidance for the steam system.

Published by Elsevier Inc.

1. Introduction

Steel industry has a relatively high demand on energy consumption. The energy cost is usually about 20–40% of the total product cost. Saving energy has become a main problem for the steel industry. As an indirect product in the steel making process, it is reported that steam consumption gets up to 10% of the total energy cost. Not only is it used in the general manufacturing process, but it is also transformed into electricity power when internal power supply of the steel plant is insufficient [16]. Therefore, a reasonable utility of steam resource is a crucial measure to improve the competitiveness of steel enterprises.

A fundamental challenge in steam resource utility is the scientific steam flow prediction. Once the variation tendency of steam flow can be accurately predicted, an effective scheduling can then be implemented. For example, when the amount of steam generated is more than the consumed amount at time, boilers or power generators can be used to reduce the steam supply. In such way, the steam diffusion can be largely avoided. On the other hand, when the steam flow is insufficient, boilers and power generators can produce steam in time so as to complement the demand for production. Thereby, online prediction for steam generation and consumption flow is an effective way for energy utility and beneficial to economy profit of steel plant. At present, the studies on metallurgic steam system automation mainly focus on control system design or equipment management of steam recovery system in the literatures such as coking dry quenching technique [3], recovery of sintering waste heat and gas [23] and recycle of converter afterheat steam [17]. However, these mentioned technologies emphasize on the perspectives of equipment or technological development, and do not consider the benefits of operational

* Corresponding author.
E-mail address: liuql@dlut.edu.cn (Q. Liu).

0020-0255/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.ins.2011.12.031

decision on energy scheduling. As a matter of fact, a sound steam flow estimation can effectively improve energy utility and reduce the waste of useful steam resources. In practice, there are a large number of steam suppliers and users in a steel plant and the pipeline networks are very complicated, therefore it is difficult to establish a physical model for steam flow prediction. In addition, we notice that the steel enterprise has accumulated various types of energy data related to their daily production. Not only do these data indicate the functionalities of real-time monitoring and control for industrial process, but they also provide us with the useful knowledge that reflects the practical system characteristics and operational measures. In such perspective, data-driven based method will be a realizable approach to study the steam flow prediction problem.

As a data-driven modeling method, neural network approach is widely used. It is reported that multilayered neural networks have theoretically been a good approximator of any nonlinear mapping with accuracy to any desired degree [21]. Recently, such modeling method is applied to energy prediction [4,8]. In [18], neural networks are used to forecast hydropower plant reservoir water inflow, and the prediction results provide valuable information for system operators. A neural network based energy consumption prediction is also presented in [6], and the method is applied to some industries in Greece. In [1], neural network is combined with genetic algorithm to predict power consumption and compare with time series based model. In addition, a combination of regression analysis, decision tree and neural network based method is proposed in [7], where the forecasting of power system load is studied. The above mentioned methods demonstrate that neural networks are effective as data-driven based modeling tool to predict energy resource flow variation. To realize a short term prediction for energy resource, one would like to fulfill the memory function for system dynamics. Currently, for predictive modeling problems, recurrent neural network is most widely used. Echo state network (ESN), which is a recurrent network, exhibits a sound modeling performance on time series based prediction [10]. The prediction accuracy of ESN is has been shown to be superior to the other network models for some classical prediction problems [11,12]. Although the approach is appealing, there are still some inherent limitations in the original formulation. In [14], leaky-integrator ESN is proposed. Units of this type have individual state dynamics, which can be exploited in various ways to accommodate the network to the temporal characteristics of a learning task. In [9], Georg suggests two enhancements of ESN. First, the idea of filters in neurons is extended to arbitrary infinite impulse response (IIR) filter neurons. This enables such networks to learn multiple attractors and signals at different timescales. Second, a delay and sum readout is introduced, which adds trainable delays in the synaptic connections of output neurons and therefore vastly improves the memory capacity of echo state networks. ESN has been applied to byproduct gas flow prediction in steel industry [24]. Nevertheless, since the output weight of ESN is usually calculated by linear regression or least square algorithm, an ill-conditioned solution might occur, which may result in a large performance degradation. To address this issue, a technique adding noises to the process of dynamic reservoir training is proposed in [13]. Yet, it does not completely overcome the drawback of ill-conditioned solution, and the precision of the established model is still found to be degraded. Ref. [22] proposes a decouple ESN model to resolve the multiple superimposed oscillator (MSO) and sea clutter problem. But, designing a series of reservoirs is needed in that method; meanwhile, the parameters determination of reservoirs is a complex and time-consuming process. The ridge regression technique is proposed in [19] to optimize the output weight of ESN. Nevertheless, the prediction is better than the generic ESN only when the ill-conditioned situation appears, and the ridge parameters are rather hard to be selected. Ref. [20] proposes a regularization method based on support vector machine. However, it is hard to determine the parameters of regularization coefficient, and the corresponding cross-validation procedure is rather time-consuming [5].

Considering the practice of steam system in steel industry where a great number of data are accumulated, a data-driven based steam flow prediction method is proposed in this study, in which a Bayesian ESN is established. Different from the least square estimation, the Bayesian based method takes the probability density over the weight space into account and obtains the optimal weight matrix via maximizing the posterior distribution of weights so as to avoid the over-fitting phenomenon. Compared to the regularizational methods, the computational load can be greatly reduced. For demonstrate the effectiveness of the proposed method, real steam flow data in Baosteel are used in this study. Results demonstrate that the proposed method is effective and can be applied to online steam flow prediction.

The paper is organized as follows. In Section 2, the practical steam system structure and related problem formulation are given. The drawbacks of ESN are analyzed and a Bayesian ESN is proposed in Section 3. Using the real data from Shanghai Baosteel Co. Ltd., we verify the effects of the proposed model for steam flow tendency prediction and the results are reported in Section 4. Section 5 presents conclusions of this study.

2. Problem formulation

Steam is one of the important resources in steel plant, which provides energy in the production process and has a close relationship with other energy media in steel industry. For instance, steam can be generated by burning gas, and the produced steam may be further utilized by internal electricity power plant. Fig. 1 presents a brief structure of steam system of Shanghai Baosteel Co. Ltd., where steam suppliers consist of coke dry quenching (CDQ) generator unit, low pressure boilers, steel-making boilers, sintering boilers, etc.; and steam users consist of steel making plant, hot rolling plant, cold rolling plant, blast furnace, etc.

Due to a large quantity of steam suppliers and users, and the complex transportation network structure, it is usually a tough task for operational workers to precisely estimate the steam variation tendency. For example, once there are high productivities simultaneously in steel making process and low pressure boilers, the steam transported into pipeline

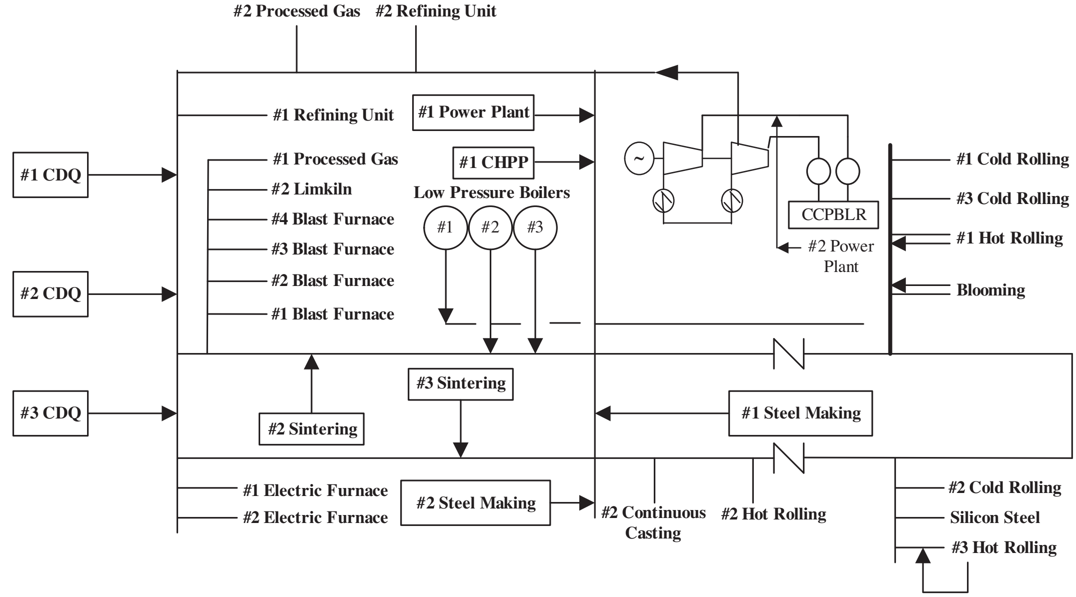


Fig. 1. A brief chart of steam system structure in Shanghai Baosteel.

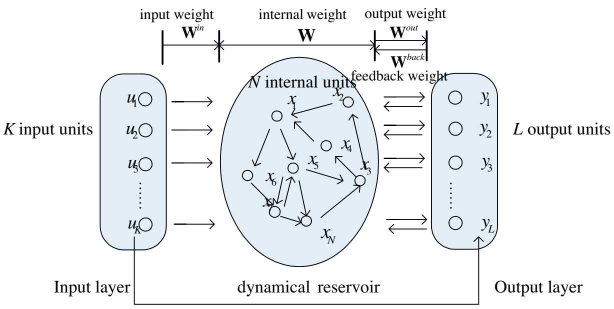


Fig. 2. Diagram of an echo state network.

networks will bring out a high pipeline pressure. At that time, if the scheduling operation by workers is out of time, then the instantaneous redundant steam will have to be diffused into the environment for protecting the energy system security, which creates the resources waste. But if the flow variation of steam supply and consumption can be aware of in advance, then boilers or power generators can be adjusted in time to use the redundant steam. Thus, the operators have to concentrate on the steam flow tendency in the future and perform resource scheduling for energy balance. However, the estimation for steam flow based on personal experience is still widely adopted in current steel plant. They estimate the steam flow depending on their personal experiences and rough production plan or manufacturing capacity of equipments, which results in rather poor estimation accuracy.

3. Bayesian ESN for steam flow prediction

Since the steam system is highly associated with other energy resources and its pipeline networks structure are rather complicated, it is difficult to establish a physical model for steam flow prediction. With the development of computer and database technology, many steel enterprises have accumulated a large number of real-time data and historical data during the production process. Not only do these data indicate the practical production real-time information, but they also reflect the dynamic characteristics of the system. According to the idea of data-based modeling, we establish a data-driven model based on time series analysis for generation and consumption amount prediction in steam system. Echo state network, a novel of artificial recurrent neural network, has been found very effective in predicting nonlinear time series [11].

3.1. Generic ESN

As illustrated in Fig. 2, the traditional ESN consists of an input layer, dynamical reservoir (DR) and an output layer. The basic idea of ESN is to use a large DR (the order ranging from tens to thousands) as a supplier of interesting dynamics from which the desired output is combined.

The recursive relation governing the ESN is described as follows:

$$\begin{aligned} \mathbf{x}(k) &= f(\mathbf{W}^{in}\mathbf{u}(k) + \mathbf{W}\mathbf{x}(k-1) + \mathbf{W}^{back}\mathbf{y}(k-1)) \\ \mathbf{y}(k) &= f^{out}(\mathbf{W}^{out}(\mathbf{u}(k), \mathbf{x}(k))) \end{aligned} \tag{1}$$

where the number of input units equals to K; the number of neurons in the DR is N; and the size of the output layer is equal to L. At the kth step, the input is $\mathbf{u}(k) = [u_1(k), u_2(k), \ldots, u_K(k)]^T$, the states of DR is $\mathbf{x}(k) = [x_1(k), x_2(k), \ldots, x_N(k)]^T$, and the output is $\mathbf{y}(k) = [y_1(k), y_2(k), \ldots, y_L(k)]^T$. $\mathbf{W}^{in}$ is the input weight connection matrix describing the relationships between the elements located in the input and the DR. $\mathbf{W}$ is the weight matrix of the neurons in DR. Note that in order to provide sufficient memorization capabilities, $\mathbf{W}$ is a sparse matrix whose connectivity level is about 1–5% and the spectral radius is less than 1. $\mathbf{W}^{back}$ is the feedback weight matrix from output layer to the DR. $\mathbf{W}^{out}$ denotes the output weight. f is the internal unit's activation function (typically sigmoid function), f^{out} is the output unit's activation function. Different from other recurrent neural networks, input weight matrix $\mathbf{W}^{in}$, internal weight matrix $\mathbf{W}^{out}$ and feedback weight matrix $\mathbf{W}^{back}$ of ESN are determined before learning process, and once initialed, the value of them are not changed in the learning and testing process. As for the output weight matrix $\mathbf{W}^{out}$, it is not need to initialized it before ESN learning process, and it will be trained by the samples data set.

The key of modeling an ESN lies on the calculation of the output weight matrix. The linear regression is usually employed such that over-fitting phenomenon occurs sometimes during the training process. Based on real data collected through many experiments, it is observed that the reservoir sometimes has an eigenvalue spread in the order of 1E12 or even higher. And, that is typically accompanied by very large learnt output weights (order of 1E8 easily reaches), which often leads to a poor prediction result. The large eigenvalue spread can increase the difficulty of implementing online algorithm [13], while ESN with large output weight led to a poor generalization. One can analyze that the traditional ESN determines the output weights matrix $\mathbf{W}^{out}$ based on generic linear regression, which means to solve the linear system about the input and reservoir states matrix $\mathbf{M}$ and the teacher collecting matrix $\mathbf{T}$.

$$\mathbf{T} = \mathbf{M}\mathbf{W}^{out} \tag{2}$$

where $\mathbf{M} = [\mathbf{x}(T_0)^T, \mathbf{x}(T_0+1)^T, \ldots, \mathbf{x}(T)^T]^T$, $\mathbf{T} = [\mathbf{y}(T_0), \mathbf{y}(T_0+1), \ldots, \mathbf{y}(T)]^T$, and T_0 is the initial time of the network. The mean least square error of $\widehat{\mathbf{W}}^{out}$ is given by

$$MSE(\widehat{\mathbf{W}}^{out}) = \sigma^2 \sum_{i=1}^{t} \frac{1}{\lambda_i} \tag{3}$$

where λ_i denotes the eigenvalue of $\mathbf{M}^T\mathbf{M}$, r is the rank of matrix $\mathbf{M}^T\mathbf{M}$ and σ^2 is the variance of the teacher output. When noise or outliner exist in the processed time series, $\mathbf{M}$ may become ill-conditioned. $\mathbf{M}^T\mathbf{M}$ will become singular meaning that at least one characteristic root is approaching to zero. $MSE(\widehat{\mathbf{W}}^{out})$ will then be amplified, and results in large learnt output weights.

3.2. Bayesian echo state network

Due to the complexity and steam flow fluctuations of the production process as well as the occurrence of detection error, the steam data collected by supervisory control and data acquisition (SCADA) system are usually noisy. If the generic ESN is used directly for modeling, the prediction performance can be degraded, and over-fitting might occur. In particular, if the number of training samples is less than the size of dynamical reservoirs, the possibility of over-fitting will also be largely increased.

In this paper, we propose a Bayesian ESN that considers a probability density function over the weight space, and the optimal output weight matrix is calculated by maximizing the posterior probability density function of the weight. The update equations of Bayesian ESN are given as follows (we omit back projections from output units to internal units or connections between output units here).

$$\begin{aligned} \mathbf{x}(k) &= f(\mathbf{W}^{in}\mathbf{u}(k) + \mathbf{W}\mathbf{x}(k-1)), \\ t(k) &= y(k) + \varepsilon(k) = f^{out}(\mathbf{W}^{out}(\mathbf{u}(k), \mathbf{x}(k))) + \varepsilon(k) \end{aligned} \tag{4}$$

where ε is an independent stochastic variable satisfying Gaussian distribution with zero mean value and variance σ_t^2 $(p(\varepsilon) \sim \mathcal{N}(\varepsilon|0, \sigma_t^2))$. Suppose that $\mathbf{z}(k) = (\mathbf{u}(k), \mathbf{x}(k))$ and the output of the proposed Bayesian ESN is $y(\mathbf{z}; \mathbf{W}^{out})$, then t satisfies a Gaussian distribution with mean value $y(\mathbf{z}; \mathbf{W}^{out})$ and variance $\sigma_t^2\,(p(t|\mathbf{W}) \sim \mathcal{N}(t|y(\mathbf{z}; \mathbf{W}^{out}), \sigma_t^2))$.

In the training stage, the samples denoted as $\{(\mathbf{u}_i, t_i)\ |i = 1, \ldots, T\}$ are trained by (4) taking $T_0 (T_0 < T)$ as the initial time, collecting input $\mathbf{u}$, reservoir state vector $\mathbf{x}$, and the teacher output on $T_0 - T$, then we get the "(input and reservoir states)-output" pairs $\{(\mathbf{z}_i, t_i)|i = 1, \ldots, n\}$. Let $D = \{t^n, \mathbf{z}^n\}$ and $\beta = 1/\sigma_t^2$, where $\mathbf{z}^n = \mathbf{z}_1, \ldots, \mathbf{z}_n$, $t^n = t_1, \ldots, t_n$. Then, the Bayesian method can be used to estimate the output weight matrix of the network, in which the probability density function over the weight space is mainly considered.

In general, there is no priori information on the output weight values. A fairly broad distribution can be chosen as a priori [2]. Here, we express this distribution in term of exponentials, that is,

$$p(\mathbf{W}^{out}) = \frac{1}{z_W(\alpha)} \exp(-\alpha E_W) \tag{5}$$

where α controls the distribution of other parameters (weights), known as the hyper-parameter, and $z_W(\alpha)$ is a normalization factor. For simplicity, a Gaussian distribution prior is chosen for the output weight, i.e.,

$$E_W = \frac{1}{2}\|\mathbf{W}^{out}\|^2 = \frac{1}{2}\sum_{i=1}^{L}(\mathbf{W}_i^{out})^2 \tag{6}$$

For a Gaussian prior, the normalization factor $z_W(\alpha)$ can be expressed by

$$z_W(\alpha) = \int e^{(-\alpha E_W)} d\mathbf{W}^{out} = \left(\frac{2\pi}{\alpha}\right)^{L/2} \tag{7}$$

where L is the total number of Bayesian ESN output weights.

As for the likelihood function $p(D|\mathbf{W}^{out})$, it can also be written as, i.e.,

$$p(D|\mathbf{W}^{out}) = \frac{1}{z_D(\beta)} \exp(-\beta E_D) \tag{8}$$

where $z_D(\beta)$ is a normalization factor, β is another hyper-parameter, and E_D is an error function. Assuming that the target t is generated from a smooth function with additive zero-mean Gaussian noise, the probability of t for a given $\mathbf{z}$ (input and reservoir states) can be written as, i.e.,

$$p(t|\mathbf{z}, \mathbf{W}^{out}) \propto \exp\left[-\frac{\beta}{2}(y(\mathbf{z};\mathbf{W}^{out}) - t)^2\right] \tag{9}$$

the hyper-parameter β controls the noise variance of the target output. Accordingly, the likelihood function (8) becomes

$$p(D|\mathbf{W}^{out}) = \prod_{i=1}^{n} p(t_i|\mathbf{z}_i, \mathbf{W}^{out}) = \frac{1}{z_D(\beta)} \exp\left(-\frac{\beta}{2}\sum_{i=1}^{n}\{(y(\mathbf{z}_i;\mathbf{W}^{out}) - t_i\}^2\right) \tag{10}$$

and the normalization factor $z_D(\beta)$ is given by [2]

$$z_D(\beta) = \int e^{(-\alpha E_D)} dD = \left(\frac{2\pi}{\beta}\right)^{n/2} \tag{11}$$

Based on the derived prior probability density $p(\mathbf{W}^{out})$ and likelihood $p(D|\mathbf{W}^{out})$, the posterior probability density by Bayesian theorem can be expressed by $p(\mathbf{W}^{out}|D) = \frac{p(D|\mathbf{W}^{out})p(\mathbf{W}^{out})}{p(D)}$.

Combining (5) with (9), we obtain the posterior distribution by making use of the fact that the likelihood does not depend on α and the prior does not depend on β. That is,

$$p(\mathbf{W}^{out}|D) = \frac{1}{Z_M} \exp(-\beta E_D - \alpha E_W) = \frac{1}{Z_M} \exp[-M(\mathbf{W}^{out})] \tag{12}$$

where,

$$\begin{aligned} M(\mathbf{W}^{out}) &= \beta E_D + \alpha E_W, \\ Z_M &= \int \exp(-\beta E_D - \alpha E_W) d\mathbf{W}^{out} \end{aligned} \tag{13}$$

The optimal output weight $\mathbf{W}_{MP}^{out}$ of the Bayesian ESN corresponds to the maximum of the posterior probability density function, i.e., maximizing (12). Since the normalizing factor Z_M is independent of the parameters, this is equivalent to minimizing $M(\mathbf{W}^{out})$.

To solve for the posterior distribution of the output weights, we apply Gaussian approximation to the output weight. This approximation is obtained through the second-order Taylor expansion of $M(\mathbf{W}^{out})$ around its minimum value or most probable set of weights $\mathbf{W}_{MP}^{out}$ [2]. Thus,

$$M(\mathbf{W}^{out}) = M(\mathbf{W}_{MP}^{out}) + \frac{1}{2}(\mathbf{W}^{out} - \mathbf{W}_{MP}^{out})^T A(\mathbf{W}^{out} - \mathbf{W}_{MP}^{out}) \tag{14}$$

where A is the Hessian matrix of $M(\mathbf{W}^{out})$. The posterior probability density of the Bayesian ESN output weight is then given by

$$p(\mathbf{W}^{out}|D) = \frac{1}{Z_M^*} \exp\left(-M(\mathbf{W}_{MP}^{out}) - \frac{1}{2}(\Delta\mathbf{W}^{out})^T A \Delta\mathbf{W}^{out}\right) \tag{15}$$

where $\Delta\mathbf{W}^{out} = \mathbf{W}^{out} - \mathbf{W}_{MP}^{out}, Z_M^*$ is a normalization.

3.3. Hyper-parameters selection

There are two hyper-parameters in Bayesian ESN. α controls the prior distribution of output weight, and β controls the distribution of likelihood function. In general, little prior information are available for α and β selection. The rules of probability theory included in the Bayesian framework propose a response through the concept of marginalization. The posterior distribution of Bayesian ESN output weight is given by

$$p(\mathbf{W}^{out}|D) = \iint p(\mathbf{W}^{out}, \alpha, \beta|D)d\alpha d\beta = \iint p(\mathbf{W}^{out}|\alpha, \beta, D)p(\alpha, \beta|D)d\alpha d\beta \tag{16}$$

Since the posterior probability distribution $p(\alpha,\beta|D)$ for the hyper-parameters is sharply peaked around their most probable values α_{MP} and β_{MP} [2]. Then, the formula (16) can be written as

$$p(\mathbf{W}^{out}|D) \approx p(\mathbf{W}^{out}|\alpha_{MP}, \beta_{MP}, D) \iint p(\alpha, \beta|D)d\alpha d\beta = p(\mathbf{W}^{out}|\alpha_{MP}, \beta_{MP}, D) \tag{17}$$

For the optimal values α_{MP} and β_{MP}, the evidence framework proposed by Mackay [15] allows us to infer the posterior

$$p(\alpha, \beta|D) = \frac{p(D|\alpha, \beta)p(\alpha, \beta)}{p(D)} \tag{18}$$

where $p(\alpha,\beta)$ is the prior of the hyper-parameter. Since the normalization factor $p(D)$ is independent of α and β, maximizing the posterior $p(\alpha,\beta|D)$ is equivalent to maximize the likelihood $p(D|\alpha,\beta)$, called the evidence term for α and β. Considering that the likelihood function is independent of β and the prior is independent of α, the evidence term writes as

$$p(D|\alpha, \beta) = \int p(D|\mathbf{W}^{out}, \alpha, \beta)p(\mathbf{W}^{out}|\alpha, \beta)d\mathbf{W}^{out} = \int p(D|\mathbf{W}^{out}, \beta)p(\mathbf{W}^{out}|\alpha)d\mathbf{W}^{out} \tag{19}$$

Combining (7) with (11), the above formula can be written as

$$p(D|\alpha, \beta) = \frac{1}{Z_D(\beta)} \frac{1}{Z_W(\alpha)} \int \exp(-M(\mathbf{W}^{out}))d\mathbf{W}^{out} = \frac{Z_M(\alpha, \beta)}{Z_D(\beta)Z_W(\alpha)} \tag{20}$$

Considering (10) with (12), and substituting the posterior distribution of output weight with Gaussian approximation, then the log of the evidence is then given as (21) where $Z_M(\alpha,\beta)$ is calculated by (13).

$$\ln p(D|\alpha, \beta) = -\alpha E_W^{MP} - \beta E_D^{MP} - \frac{1}{2}\ln(\det A) + \frac{L}{2}\ln(\alpha) + \frac{n}{2}\ln(\beta) - \frac{n}{2}\ln(2\pi) \tag{21}$$

The optimal values of the hyper-parameters can be obtained by differentiating the derived log evidence, see [12] for more calculation details. Hence, the optimal values are

$$\alpha^{new} = \gamma/2E_W, \beta^{new} = (n - \gamma)/2E_D \tag{22}$$

where γ is given by $\gamma = \sum_{i=1}^{W} \frac{\lambda_i}{\lambda_i+\alpha}$, and $\{\lambda_i|i = 1, \ldots, L\}$ are the eigenvalues of the Hessian matrix of the un-regularized error E_D, i.e. $H = \beta\nabla\nabla E_D$.

3.4. Prediction steps of steam flow tendency

The predicting steps for the steam flow tendency by using the proposed Bayesian ESN are summarized as follows.

Step 1: Obtain the time series data of steam flow via the SCADA system, which is long enough to be trained, denoted as $\{(\mathbf{u}_i, t_i)|i = 1, \ldots, T\}$. Initialize the Bayesian ESN structure and set the weights $\mathbf{W}^{in}$ and $\mathbf{W}$.
Step 2: Train the network using $\{(\mathbf{u}_i,t_i)|i = 1, \ldots, T\}$ and update the states by (4).
Step 3: Set the initial time $T_0(T_0 < T)$, collect the input vector $\mathbf{u}$, reservoir state vector $\mathbf{x}$ and the teacher output $\mathbf{t}$ on $T_0 - T$ to create "(input and reservoir states)-output" pairs $\{(\mathbf{z}_i,t_i)|i = 1, \ldots, n\}$.
Step 4: Initialize values for the hyper-parameters α and β, and the output weights of Bayesian ESN.
Step 5: At current i, based on current hyper-parameters α and β, and the output weights $\mathbf{W}^{out}$, the estimate of $M(\mathbf{W}^{out})$ is computed by (14).
Step 6: Optimize the weight matrix via minimizing $M(\mathbf{W}^{out})$ by using methods such as Gauss–Newton method or conjugate gradient descent.

Step 7: Calculate E_W and E_D by using the current output weights $\mathbf{W}^{out}$, and update the hyper-parameters α and β using (22).
Step 8: Check whether the accuracy requirement is satisfied. If so, the optimal output weights of Bayesian ESN $\mathbf{W}^{out}_{MP}$ are obtained, and go to step 9; otherwise, go back to step 5.
Step 9: When the process of network training is completed, all of the weight matrixes are not changed during the predicting process, and the prediction is performed by using formula (4).

4. Simulation and industrial application

To verify the effectiveness of the proposed Bayesian ESN method and quantify its performance, we employ the real-world data of low-pressure steam system, obtained from Shanghai Baosteel energy center in December 2010. The sampling frequency is 1 min, and the simulation is performed in MATLAB running on an IBM Workstation.

We first take the consumption amount of low-pressure steam in #1 cold rolling plant. The first 1400 points of a continuous 1440-min data are used as training samples, and the remainder 40 points are as testing samples. Considering the steam flow variation of production and consumption in one hour can basically reflect overall dynamic characteristics of the corresponding units, we empirically set the embedded dimension as 60. The activation function for the internal neurons is tanh(), and the output activation function is linear. The input weight $\mathbf{W}^{in}$ and reservoir weight $\mathbf{W}$ are randomly generated, the initial value of the hyper-parameter is empirically set as $\alpha = 5$, $\beta = 2$, and other parameter settings of the reservoir for #1 cold rolling plant time series are shown in Table 1. The prediction result compared to the real flow data is shown in Fig. 3, which indicates that the predicted values well trace the practical data. For presenting the superiority of the proposed method, we compare the prediction performance to the generic ESN modeling, in which the parameters of network are similar to the proposed method. The prediction results using the same training samples data by the generic ESN and the Bayesian one are illustrated as Figs. 4 and 5. The sub-figure (a) illustrates the results of the training samples, while (b) for the testing ones. It is apparent from the two figures that although the generic ESN gives relatively good fitting results for the training samples, the predicted results in sub-figure (b) are worse than those by the proposed Bayesian ESN. That is, the generalization of the proposed ESN is superior to the traditional version. To quantify the prediction accuracy, the independent experiments are performed 100 times respectively by ESN and the Bayesian version. Normalized Root Mean Square Error (NRMSE) and Mean Absolute Percentage Error (MAPE) are employed to evaluate the average experiment results shown in Table 2,

$$\mathrm{NRMSE} = \sqrt{\frac{1}{T\|y_d\|^2}\sum_{t=1}^{T}(y_*(t) - y_d(t))^2}, \quad \mathrm{MAPE} = \frac{100}{T}\sum_{t=1}^{T}\frac{|y(t) - y_d(t)|}{y_d(t)},$$

Table 1
The parameter settings of the reservoir for #1 cold rolling plant time series.

Items	Values
DR dimension	300
Sparse interconnectivity of DR	2%
Spectral radius of **W**	0.85

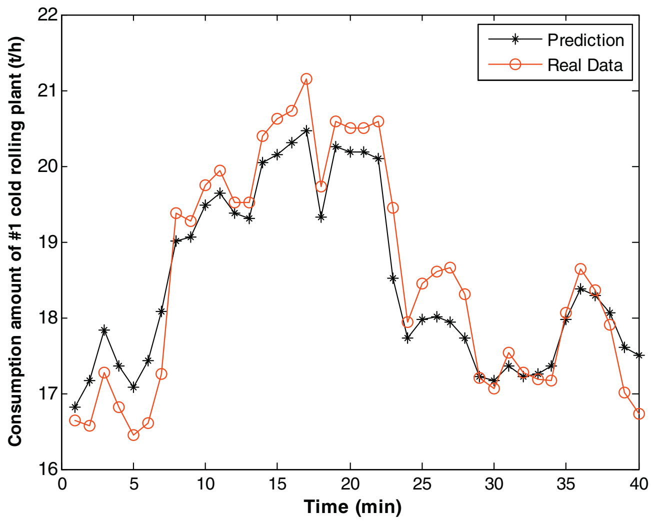
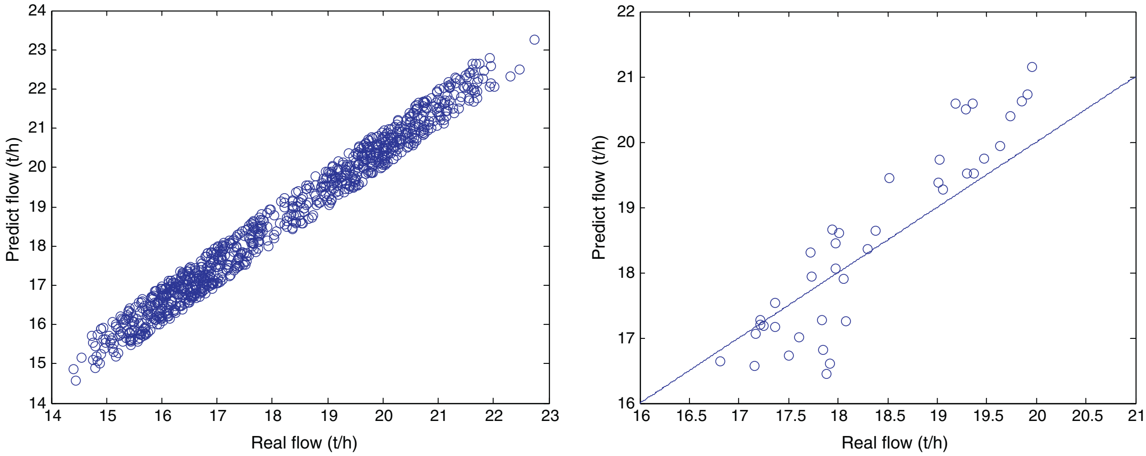
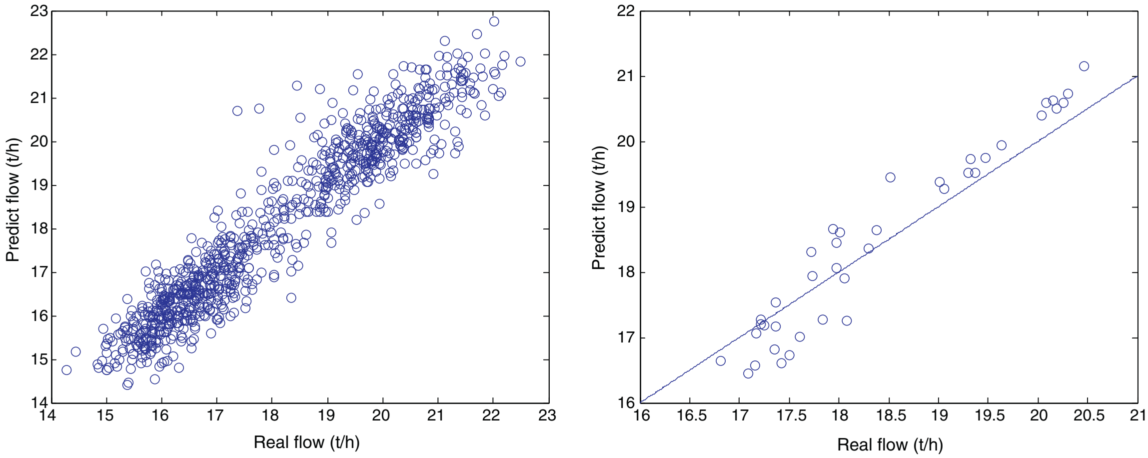


Fig. 3. The prediction results of steam consumption flow for #1 cold rolling plant.



(a) for training sample (b) for testing data

Fig. 4. The comparison of real flow and prediction results by generic ESN model.



(a) for training sample (b) for testing data

Fig. 5. The comparison of real flow and prediction results by the proposed Bayesian ESN model.

Table 2
The prediction error comparison by the two modeling methods.

Method	Training sample		Testing data	
	NRMSE	MAPE (%)	NRMSE	MAPE (%)
Generic ESN	0.2984	2.8725	0.4917	3.0242
Bayesian ESN	0.3025	2.9225	0.3178	2.0964

where T is the number of predicted points, $y_*(t)$ is the predicted value, and $y_d(t)$ is the real value. From the listed table, the prediction error of the proposed modeling is obviously lower than that of generic ESN in terms of testing data.

To further analyze the robustness of the DR dimensionality of Bayesian ESN, we set various DR dimensionalities to train the network, and evaluate the prediction error with NRMSE and MAPE for the steam consumption amount of #1 cold rolling plant. We also carry out the comparative experiments with the generic ESN, and complete them for 100 times. The statistical results are listed in Table 3, where we find that the prediction accuracy can be improved along with the increase of DR dimension when using the generic ESN; meanwhile, the computational cost is also grown. As for the proposed network, the prediction accuracy maintains a relatively stable level even for large DR dimension. In other words, Bayesian ESN can provide with a higher robustness compared to the generic ESN. With respect to the over-fitting phenomenon, the

Table 3
The prediction comparison when setting different DR dimension in the two network structure.

DR dimension	Method	NRMSE	MAPE (%)
100	ESN	0.4853	5.5970
	Bayesian ESN	0.3120	3.8253
200	ESN	0.4803	5.0368
	Bayesian ESN	0.3132	3.9112
300	ESN	0.4602	4.8925
	Bayesian ESN	0.3106	3.7796
400	ESN	0.4039	4.5367
	Bayesian ESN	0.3015	3.6732
500	ESN	0.3910	4.4535
	Bayesian ESN	0.3101	3.7092

Table 4
The statistics of over-fitting phenomenon amount by the two modeling method.

DR dimension	Over-fitting amounts happened in the 100-times experiments	
	Bayesian ESN	ESN
100	0	1
200	0	4
300	0	5
400	0	7
500	0	11

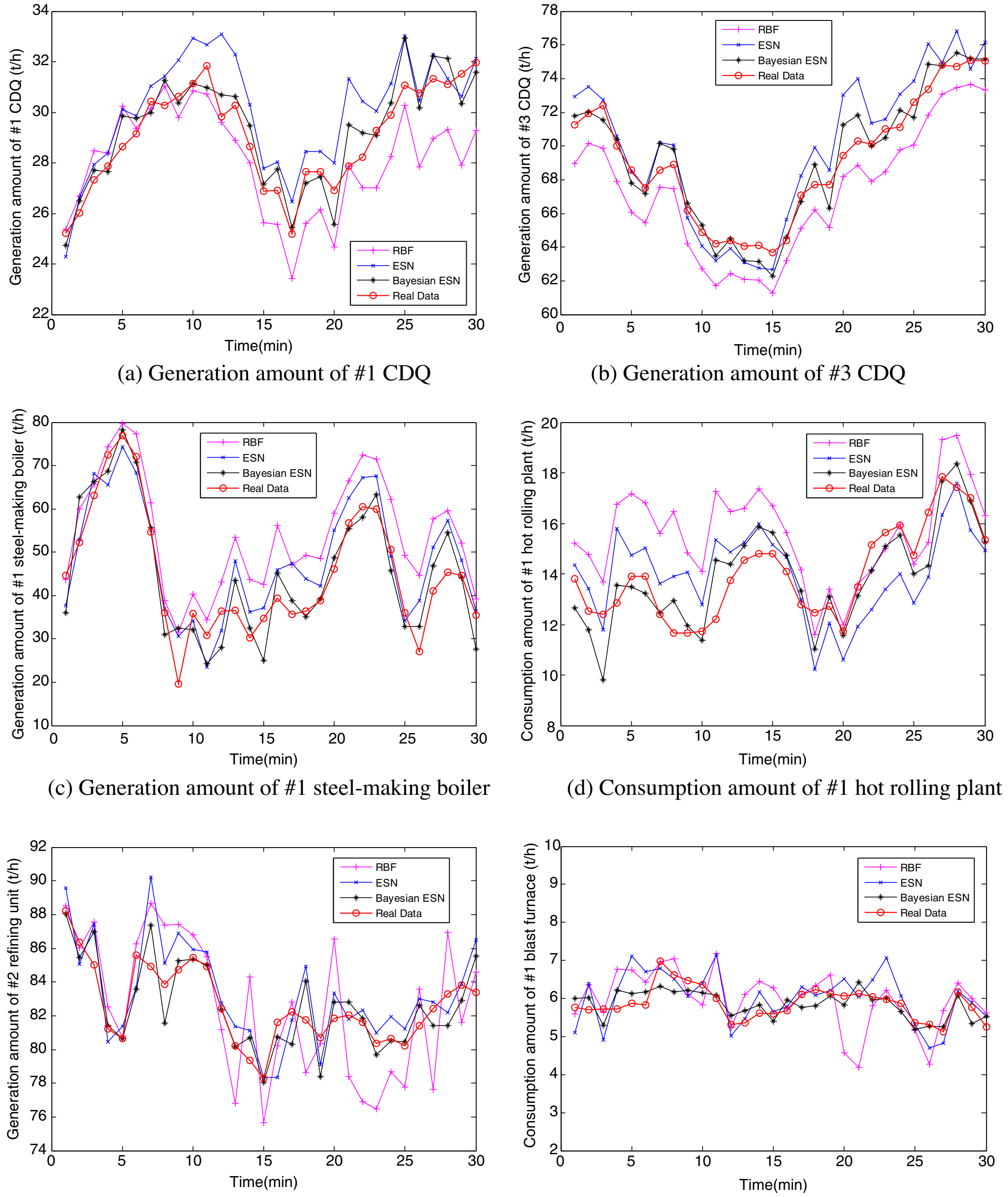
Table 5
The parameter settings of the Bayesian ESN.

Prediction objectives	DR dimension	Spectral radius of **W**	Sparse interconnectivity of DR (%)
#1 CDQ steam turbine	200	0.75	2
#3 CDQ steam turbine	200	0.75	1.5
#1 Steel-making boiler	300	0.85	1.5
#1 Hot rolling plant	200	0.85	1.5
#2 Refining uni	200	0.80	2
#1 Blast furnace	500	0.90	2

comparative experiment presents the statistics for the number of ill-conditioned solution, as shown in Table 4. It is clear that the proposed ESN does not result in the ill-conditioned solution, while the probability of over-fitting increases along with the increasing of the DR dimension when adopting the traditional ESN structure.

In order to understand the adaptivity of the proposed Bayesian ESN, we employ different steam resource suppliers and users in the energy system to realize the flow prediction process including the generation amounts of #1 CDQ steam turbine, #3 CDQ steam turbine, #1 steel-making boiler, and the consumption amount of #1 hot rolling plant, #2 refining unit and #1 blast furnace. The parameter settings of the Bayesian ESN are show in Table 5. The experiments adopting the generic ESN, RBF network and the proposed Bayesian ESN are performed and the results are reported in Fig. 6. The NRMSE and MAPE are used to evaluate the prediction quality, and the performances of those methods are shown in Table 6. Apparently, the proposed Bayesian ESN produces the best prediction performance and the prediction accuracy of 30-min steam generation. The consumption flow in steam system is higher than 90%, and it satisfies the industrial requirements.

From the experimental analysis, the proposed Bayesian ESN is found to have improved fitting and generalization performance. Although the fixed weights of DR of ESN can store the information received from the teacher signal, which avoids the drawback of the local minimum via calculating the required regression output weights between DR state and output, the traditional ESN suffers from the over-fitting problem for striking the network weights using linear regression or least squares methods. In this study, Bayesian rule is incorporated into the echo state networks. We focus on the probability distribution of the weights when solving the output weights of the network in order to overcome the over-fitting and improve the generalization ability.



(a) Generation amount of #1 CDQ (b) Generation amount of #3 CDQ

(c) Generation amount of #1 steel-making boiler (d) Consumption amount of #1 hot rolling plant

(e) Consumption amount of #2 refining unit (f) Consumption amount of #1 blast furnace

Fig. 6. The comparison of prediction tendencies using the three modeling methods.

The proposed prediction approach for steam system has been successfully applied to the energy center of Shanghai Baosteel Co. Ltd. The application software based on this method is developed using Visual C# language, the prediction model runs on an IBM P_SERIES server. The background database using Oracle 10 g acquires the steam flow data of each energy user via a real-time database called iHistorian on the SCADA system. This application system has been successfully run for six months so far, which plays a well guiding role for the steam resources prediction and real-time scheduling for energy.

5. Conclusions

Taking the steam resources system in steel industry as the study background, this paper proposes a data-driven based prediction model that combines the Bayesian rule with the structure of echo state network in order to avoid the over-fitting and improve the prediction accuracy. Through the practical data from Baosteel, the proposed data-driven based model is ver-

Table 6
The quantified prediction accuracy by the three methods.

Prediction objectives	Method	NRMSE	MAPE (%)
#1 CDQ steam turbine	Bayesian ESN	0.3097	2.3593
	RBF	0.5954	4.4811
	ESN	0.5124	4.1222
#3 CDQ steam turbine	Bayesian ESN	0.2435	1.0610
	RBF	0.5362	2.7641
	ESN	0.4225	1.8022
#1 Steel-making boiler	Bayesian ESN	0.4066	12.8680
	RBF	0.7373	23.7612
	ESN	0.4787	14.8743
#1 Hot rolling plant	Bayesian ESN	0.5631	5.7072
	RBF	1.2787	14.0300
	ESN	0.9039	10.0550
#2 Refining unit	Bayesian ESN	0.5687	1.2052
	RBF	1.2076	2.8358
	ESN	0.7836	1.6862
#1 Blast furnace	Bayesian ESN	0.6977	4.4032
	RBF	1.6215	8.9335
	ESN	1.2342	6.7913

ified to be viable to realize the short term flow prediction for the steam system in steel industry, which can provide with a scientific guidance to the energy scheduling.

Acknowledgments

This work is supported by the National Natural Sciences Foundation of China (Nos. 61034003 and 61104157). The cooperation of energy center of Shanghai Baosteel Co. Ltd., China, in this work is greatly appreciated. And, the authors would like to thank the Associate Editor and the anonymous reviewers for their valuable comments and constructive suggestions.

References

[1] A. Azadeh, S.F. Ghaderi, S. Tarverdian, M. Saberi, Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption, Applied Mathematics and Computation 186 (2007) 1731–1741.
[2] M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.
[3] G. Bisio, G. Rubatto, Energy saving and some environment improvements in coke-oven plants, Energy 25 (1998) 247–265.
[4] K.C. Choon, Passive learning and input-to-state stability of switched Hopfield neural networks with time-delay, Information Sciences 180 (2010) 4582–4594.
[5] L.O. David, D. Dursun, Advanced Data Mining Techniques, Springer, 2007.
[6] L. Ekonomou, Greek long-term energy consumption prediction using artificial neural network, Energy 35 (2010) 512–517.
[7] K.F. Geoffrey, K.W. Kelvin, Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks, Energy 32 (2007) 1761–1768.
[8] C. Godfrey, Design of hybrid differential evolution and group method of data handling networks for modeling and prediction, Information Sciences 178 (2008) 3616–3634.
[9] G. Holzmann, H. Hauser, Echo state networks with filter neurons and a delay and readout, Neural Networks 23 (2010) 244–256.
[10] H. Jaeger, Tutorial on training recurrent neural networks, covering BPTT, RTRL, EKF and echo state network approach, GMD Report 159, German National Research Center for Information Technology, 2002.
[11] H. Jaeger, Adaptive nonlinear system identification with echo state networks, Advances in Neural Information Processing Systems 15 (2003) 593–600.
[12] H. Jaeger, H. Haas, Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication, Science 304 (2004) 78–80.
[13] H. Jaeger, Reservoir riddles: suggestions for echo state network research, in: Proceedings of International Joint Conference on Neural Networks, Canada, IEEE, 2005.
[14] H. Jaeger, M. Lukosevicius, D. Popovici, U. Siewert, Optimization and applications of echo state networks with leaky integrator neurons, Neural Networks 20 (2007) 335–352.
[15] J.C. Mackey, A practical Bayesian framework for back-propagation networks, Computation and Neural Systems 4 (1992) 448–472.
[16] M.S. Mohsen, B.A. Akash, Energy analysis of the steel making industry, International Journal of Energy Research 22 (1998) 1049–1054.
[17] C.J. Rick, Strategies for use of superheated steam during stainless steel refining in converters, in: Association for Iron and Steel Technology Conference Proceedings, vol. 1, 2010, pp. 1103–1114.
[18] S. Rodrigo, C.O. Mustafa, C.P. Jose, et al., Water inflow forecasting using the echo state network: a Brazilian case study, in: Proceedings of International Joint Conference on Neural Networks, USA, 2007.
[19] Z.W. Shi, M. Han, Ridge regression learning in ESN for chaotic time series prediction, Control and Decision 22 (2007) 258–267.
[20] Z.W. Shi, M. Han, Support vector echo-state machine for chaotic time-series prediction, IEEE Transactions on Neural Networks 18 (2007) 359–372.
[21] H.T. Siegelmann, E.D. Sontag, Turing computability with neural nets, Applied Mathematics Letters 4 (1991) 77–80.
[22] Y.B. Xue, L. Yang, S. Haykin, Decoupled echo state networks with lateral inhibition, Neural Networks 20 (2007) 365–376.
[23] L.H. Zhang, J.J. Cai, The research on steam utilization way of iron and steel enterprises, Industrial Heating 36 (2007) 1–3.
[24] 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 19 (3) (2011) 507–520.