Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Echo state network based prediction intervals estimation for blast furnace gas pipeline pressure in steel industry

Chunyang Sheng*, Jun Zhao*, Wei Wang*, Quanli Liu*

**School of Control Science and Engineering, Dalian University of Technology, Dalian, China, (e-mail: zhaoj@dlut.edu.cn)}*

Abstract: The pipeline pressure of blast furnace gas (BFG) system in steel industry provides effective information for the energy scheduling operations. However, due to the complexity of the byproduct gas pipeline network and the large fluctuations of the gas flow, it is rather difficult to establish an accurate prediction model for the pipeline pressure. Additionally, the quantitative reliability of the prediction accuracy is the key concerns of energy scheduling workers since there are always a variety of uncertainties in industrial process. In this study, an echo state network (ESN) modelling with output feedback is proposed to predict the BFG pipeline pressure. Given the gas flows data and the pressures sampled from the sensors are generally accompanied with noise, a Bayesian framework for the prediction intervals (PIs) is designed, which can quantify the input and output noises. To verify the effectiveness of the proposed method, a number of prediction experiments coming from industrial data are conducted here. And, the experimental results indicate that the proposed approach has a satisfactory performance on PIs for the pipeline pressure.

Keywords: Echo state network; gas pipeline pressure; prediction interval; Bayesian framework.

1. INTRODUCTION

The pipeline pressure of BFG is one of the most important evaluation indexes for real time energy scheduling operations in steel industry. And, the accurate prediction for the pipeline pressure can provide effective information for the energy scheduling with low risks. However, the pipeline pressure is usually unstable due to the variation of the byproduct gas flow. Currently, a few of researches on the prediction of gas pipeline pressure were carried out based on point-oriented mode, which were however lack of the reliability analysis and any indications of the accuracy, see e.g. Zhu and Leis (2012),Taware and Brown (1999). Nevertheless, since the industrial data is sparse and the data collection processes are always accompanied with uncertainties, the energy scheduling workers paid more attentions on not only the predicted results but also the reliability analysis of prediction, which is rather difficult to be quantified by point-oriented prediction. A prediction interval comprised of the upper and lower bounds with a confidence probability can provide general point-oriented predictive values with the supplementary indication of their accuracy (Khosravi et al., 2011). Furthermore, the prediction interval can illustrate the range that the targets may appear with a confidence level.

As for the existing PIs construction methods, a large number of them were built on the numerical characteristics analyses of the predicted distribution, such as the delta (Chryssolouris et al., 1996), the MVE (mean-variance estimation) (Nix and Weigend, 1994), the bootstrap (Heskes, 1997) and the Bayesian techniques, see e.g. Sheng et al. (2013), Lauret et al. (2008), Zhang and Luh (2005). It was feasible for the delta and the MVE to construct PIs based on the assumption of the distribution of noise, where the noise variance was viewed as a constant for the delta method, and the MVE assumed that the noise was normally distributed around the true mean of the targets. The unreasonable assumptions made the two methods generate low-quality PIs, see e.g. Chryssolouris et al. (1996), Nix and Weigend (1994). Bootstrap method must be combined with a network ensemble for PIs, so its performance depends on the accuracy of many neural networks. It was frequently that one of these networks was rather biased, and then an inaccurate estimation had to be created, see e.g. Khosravi et al. (2011), Heskes (1997). Besides, the Bayesian method for PIs construction was based on a strong mathematical foundation, which typically had a better generalization than some other networks (Sheng et al., 2013). However, Bayesian learning technique is also computationally intensive in modeling stage due to the calculation of the Hessian matrix (Lauret et al., 2008).

Echo state network (ESN), a recurrent neural network (RNN), had been proposed for prediction, in which its advantage lies in that the input weights, the internal weights and the feedback weights are fixed before training, and only the output weights are required to be determined (Jaeger, 2004). Combining with the Bayesian learning technique, ESN can reduce the computational cost for the Hessian matrix because the output equation of ESN is linear. The existing Bayesian ESNs, see e.g. Sheng et al. (2013), Liu et al. (2012), Han and Mu (2010), mainly used for time series prediction may not be suitable for multiple factor relationship regression, because the input uncertainties levels associated with each factor are different. Furthermore, the output feedback uncertainty was not considered in the above studies, which is however necessary for practical problems.

In this study, an ESN model with output feedback is proposed for an industrial prediction. Since the industrial data generally contain noises, the output feedback uncertainty and the input uncertainties were considered for PIs construction by using the Bayesian framework, where the PIs contributed by the output feedback uncertainty are derived first, and then the total PIs are established based on the uncertainties of the input and the output feedback. To verify the quality of the proposed method, a pipeline pressure prediction of the BFG system is studied here, and the prediction experiments are conducted based on the industrial data. And, the results indicate that the proposed approach has a satisfactory performance on PIs for the pipeline pressure.

2. PROBLEM STATEMENT

BFG system is the one of the most important energy systems in steel industry, which consists of the blast furnaces, the gas tanks, the transportation pipelines and a series of gas users. Taking a certain steel plant as an example, the structure of the BFG system can be illustrated as Figure 1, in which four blast furnaces viewed as the generation units provide the gas into the pipeline network, and the consumption users primarily include coke oven, hot rolling, cold rolling, chemical product recovery, boilers, power plant, etc.. Besides, a part of gas is stored in the gas tanks which are regarded as a storage device.

In practice, the scheduling of BFG system can be carried out by the prediction of a number of variables, such as the gas flow, the gas tank level or the pipeline pressure of some locations. In Figure 1, the locations marked by ovals are the most concern points, seeing the outlet pressure of the gas tanks and those of the blast furnaces. Since the pipeline pressures are affected by the previous status and the flows of the other locations, the pressure estimation of these key locations has to consider the influence factor uncertainties (input) and the uncertainties of previous status of the location (output feedback).

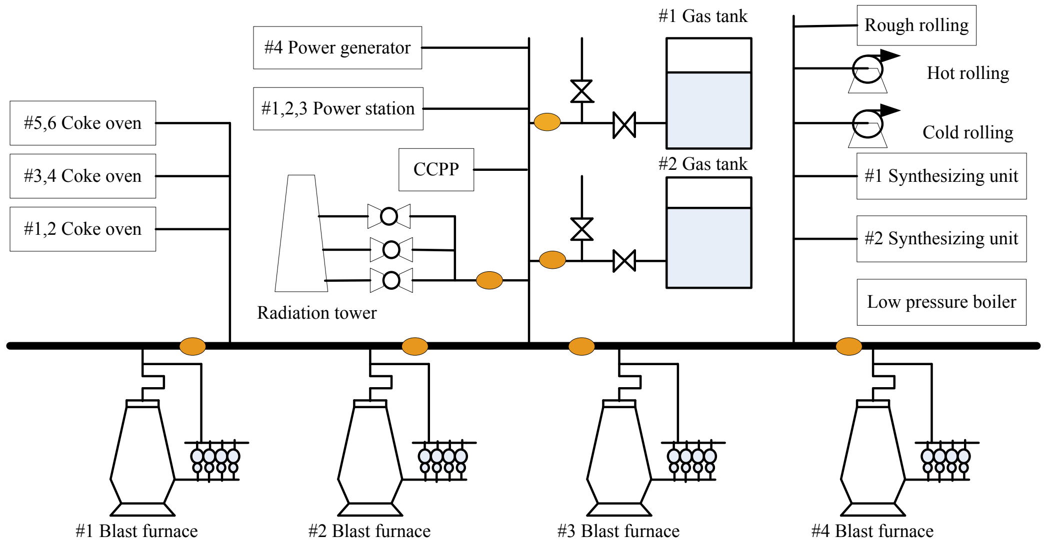


Fig.1 Pipeline structure of the BFG system in steel industry

3. BAYESIAN ECHO STATE NETWORK REGRESSION WITH UNCERTAINTIES

3.1 Prediction model based on echo state network

A typical ESN contains the input layer, the reservoir and the output layer (Jaeger, 2002a). And, an ESN with output feedback can be formulated as

$$\mathbf{x}_k = f(\mathbf{W}^{in}\mathbf{u}_k + \mathbf{W}\mathbf{x}_{k-1} + \mathbf{W}^{back} y_{k-1}) \quad (1)$$

$$t_k = y_k + n_k = f^{out}(\mathbf{W}^{out}(\mathbf{u}_k, \mathbf{x}_{k-1}, y_{k-1})) + n_k \quad (2)$$

where $\mathbf{u}_k$ is the exogenous input with dimensionality m, $\mathbf{x}_k$ is the internal states with dimensionality N, and y_k is the output. $\mathbf{W}^{in} = (W_{i,j}^{in}) \in \mathbb{R}^{N\times m}$ denotes the input weights, $\mathbf{W} = (W_{i,j}) \in \mathbb{R}^{N\times N}$ denotes the internal weights of the neurons in the reservoir. To provide proper memorization capabilities, $\mathbf{W}$ should be sparse whose connectivity level ranges from 1%~5% and its spectral radius should be less than 1. $\mathbf{W}^{back} = (W_{i,j}^{back}) \in \mathbb{R}^{N\times 1}$ denotes the output feedback weights, and $\mathbf{W}^{out} = (W_{i,j}^{out}) \in \mathbb{R}^{1\times(m+N+1)}$ denotes the output weights. f and f^{out} are the activation functions of internal neurons and output neurons, respectively. n_k is independent white Gaussian noise sequences reflecting the output uncertainty.

Here, t_k is the noisy output and y_k is the output of the ESN. Given the output feedback is the prior output of the ESN, the feedback uncertainty should be considered in this study.

As for the input uncertainties, one can assume a random vector $\mathbf{z}_k$ as the noisy input, i.e.,

$$\mathbf{z}_k = f(\mathbf{u}_k, \boldsymbol{\varepsilon}_k) \quad (3)$$

where $\mathbf{u}_k$ is the hidden input, and $\boldsymbol{\varepsilon}_k$ is a random noise vector, independent of $\mathbf{u}_k$. Because the gas flow in the locations of the pipeline network are accompanied with different level of noise, the covariance of $\boldsymbol{\varepsilon}_k$ can be denoted as $\boldsymbol{\Sigma} = diag\left(\sigma_1^2, \sigma_2^2, \cdots, \sigma_m^2\right)$.

3.2 Regression with output feedback uncertainty

Considering the modeling without output feedback (Wright, 1999), the predicted distribution $p(t^* \mid \mathbf{u}^*, D)$ can be similarly approximated by a Gaussian distribution in terms of the marginal distribution

$$p(t^* \mid \mathbf{u}^*, D) = \int p(t^* \mid \mathbf{u}^*, t^-, \mathbf{W}^{out}) p(\mathbf{W}^{out} \mid D) d\mathbf{W}^{out} \quad (4)$$

where t^- denotes the noisy output feedback.

Assuming that y^- is the output feedback without noise, it is possible to linearize $f^{out}(\mathbf{u}^*, \mathbf{x}^*, y^-; \mathbf{W}^{out})$ around t^-, then one can neglect the second-order terms, i.e.,

$$f^{out}(\mathbf{u}^*, \mathbf{x}^*, y^-; \mathbf{W}^{out}) = f^{out}(\mathbf{u}^*, \mathbf{x}^*, t^-; \mathbf{W}^{out}) + \mathbf{h}_{t^-}^T \delta t^- \quad (5)$$

where $\delta t^- = y^- - t^-$, $\mathbf{h}_{t^-}^T$ is the partial derivative of $f^{out}(\mathbf{u}^*, \mathbf{x}^*, y^-; \mathbf{W}^{out})$ with respect to y^-, i.e.,

$$\begin{aligned} \mathbf{h}_{t^-}^T &= \nabla_{y^-} f^{out}(\mathbf{u}^*, \mathbf{x}^*, y^-; \mathbf{W}^{out})|_{y^- = t^-} \\ &= \mathbf{W}_{y^-}^{out} + \mathbf{W}_{\mathbf{x}}^{out} \cdot \mathbf{W}^{back} \left(\left[\cosh(\mathbf{x}^*)\right]^T \left[\cosh(\mathbf{x}^*)\right] \right)^{-1} \end{aligned} \quad (6)$$

where $\mathbf{W}_{y^-}^{out}$ is a block matrix of the output weights corresponding to y^-, and $\mathbf{W}_{\mathbf{x}}^{out}$ is a block matrix of the output weights corresponding to the internal states. As such, using the Bayesian rule, the output distribution is

$$\begin{aligned} p(t^* \mid \mathbf{u}^*, t^-, \mathbf{W}^{out}) &= \int p(t^* \mid \mathbf{u}^*, y^-, \mathbf{W}^{out}) p(y^- \mid t^-) dy^- \\ &= \int \exp\left(\frac{\beta}{2} \left\{ \mathbf{W}^{out}\left(\mathbf{u}^*, \mathbf{x}^*, y^-\right) - t^* \right\}^2 \right. \\ &\quad \left. - \frac{1}{2\sigma_t^2} (\delta t^-)^T (\delta t^-) \right) dy^- \end{aligned} \quad (7)$$

where $\beta = 1/\sigma_t^2$. Then, the output distribution can be further approximated by a Gaussian distribution

$$p(t^* \mid \mathbf{u}^*, t^-, \mathbf{W}^{out}) = \frac{1}{Z_{t^-}} \exp\left(-\frac{\beta'}{2} \left\{\Delta t^*\right\}^2 \right) \quad (8)$$

where $\Delta t^* = t^* - f^{out}(\mathbf{u}^*, \mathbf{x}^*, t^-; \mathbf{W}^{out})$, Z_{t^-} is the normalizing constant and $1/\beta' = 1/\beta + \sigma_t^2 = \left(1 + \mathbf{h}_{t^-}^T \mathbf{h}_{t^-}\right)/\beta$.

As for the $p(\mathbf{W}^{out} \mid D)$ term in (4), one can expand $D = \left\{t_i, \mathbf{u}_i\right\}\big|_{i=1,2,\cdots,n.}$ and marginalize over the output feedback y_i^-, then

$$p(\mathbf{W}^{out} \mid D) = \int p(\mathbf{W}^{out}, y_i^- \mid \mathbf{u}_i, t_i, t_i^-) dy_i^- \quad (9)$$

Using the Bayesian rule and the conditional independence of t_i on t_i^- given y_i^-, (9) can be re-written as

$$\begin{aligned} p(\mathbf{W}^{out} \mid D) &= \frac{p(\mathbf{W}^{out})}{p(t_i \mid t_i^-)} \\ &\quad \cdot \int p(t_i \mid \mathbf{u}_i, y_i^-, \mathbf{W}^{out}) p(y_i^- \mid t_i^-) dy_i^- \end{aligned} \quad (10)$$

where $p(\mathbf{W}^{out}) \propto \exp(-\alpha/2 \|\mathbf{W}^{out}\|^2)$, α is a hyper-parameter. Based on (10), the posterior distribution of the output weights reads as

$$\begin{aligned} p(\mathbf{W}^{out} \mid D) &\propto \prod_{i=1}^{N} \exp\left(-\frac{\beta'}{2} \left\{\Delta t_i\right\}^2 - \frac{\alpha}{2} \|\mathbf{W}^{out}\|^2 \right) \\ &= \exp\left(-\frac{\beta'}{2} \sum_{i=1}^{n} \left\{\Delta t_i\right\}^2 - \frac{\alpha}{2} \|\mathbf{W}^{out}\|^2 \right) \end{aligned} \quad (11)$$

where $\Delta t_i = t_i - f^{out}\left(\mathbf{u}_i, \mathbf{x}_i, t_i^-; \mathbf{W}^{out}\right)$. Taking (8) and (10) into (4), then

$$\begin{aligned} p(t^* \mid \mathbf{u}^*, D) &= \frac{1}{Z_{t\mathbf{u}}} \int \exp\left(-\frac{\beta'}{2} \left\{\Delta t^*\right\}^2 \right) \\ &\quad \cdot \prod_{i=1}^{n} \exp\left(-\frac{\beta'}{2} \left\{\Delta t_i\right\}^2 - \frac{\alpha}{2} \|\mathbf{W}^{out}\|^2 \right) d\mathbf{W}^{out} \end{aligned} \quad (12)$$

where $Z_{t\mathbf{u}}$ is the normalizing constant, and

$$M(\mathbf{W}^{out}) = \frac{\beta'}{2} \sum_{i=1}^{n} \left\{\Delta t_i\right\}^2 + \frac{\alpha}{2} \|\mathbf{W}^{out}\|^2 \quad (13)$$

For the prediction distribution, (13) can be linearly approximated by the Taylor expanding with respect to $\mathbf{W}_{MP}^{out}$. That is,

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

where $\Delta \mathbf{W}^{out} = \mathbf{W}^{out} - \mathbf{W}_{MP}^{out}$ and $\mathbf{A}$ is the Hessian matrix of $M(\mathbf{W}^{out})$. Taking (14) into (12), a Gaussian approximation of the predictive distribution can be obtained.

$$p(t^* \mid \mathbf{u}^*, D) = \frac{1}{(2\pi\sigma_{t^*}^2)^{1/2}} \exp\left(\frac{\left\{t^* - f^{out}\left(\mathbf{u}^*, \mathbf{x}^*, t^-; \mathbf{W}^{out}\right)\right\}^2}{2\sigma_{t^*}^2} \right) \quad (15)$$

where $\sigma_{t^*}^2 = 1/\beta' + \mathbf{g}^T\mathbf{A}^{-1}\mathbf{g} = 1/\beta(1+\mathbf{h}_{t^-}^T\mathbf{h}_{t^-}) + \mathbf{g}^T\mathbf{A}^{-1}\mathbf{g}$. As such, the variance of the predictive distribution consists of three components. 1) the variance in the distribution over the observed targets, 2) an estimation of the output feedback uncertainty, and 3) the uncertainty induced by the weights.

3.3 Regression with input uncertainties

Given that the input of the ESN is noisy, the predictive distribution can be formulated as

$$p(t^* \mid \mathbf{z}^*, D') = \int p(t^* \mid \mathbf{z}^*, t^-, \mathbf{W}^{out}) p(\mathbf{W}^{out} \mid D') d\mathbf{W}^{out} \quad (16)$$

Given the noisy input $\mathbf{z}^*$ and the feedback t^-, the output distribution can be written as

$$\begin{aligned} p(t^* \mid \mathbf{z}^*, t^-, \mathbf{W}^{out}) &= \int p(t^* \mid \mathbf{z}^*, t^-, \mathbf{W}^{out}) p(\mathbf{z}^* \mid \mathbf{u}^*) d\mathbf{u}^* \\ &= \int \exp\left\{ \frac{\beta'}{2}(\Delta t^*)^2 - \frac{1}{2}(\Delta \mathbf{z}^*)^T \boldsymbol{\Sigma} (\Delta \mathbf{z}^*) \right\} d\mathbf{z}^* \end{aligned} \quad (17)$$

where $\mathbf{u}^*$ is the hidden input without noise, $\Delta\mathbf{z}^* = \mathbf{u}^* - \mathbf{z}^*$. To simplify (17), $f^{out}(\mathbf{u}^*; \mathbf{W}^{out})$ can be linearly approximated by Taylor expansion about $\mathbf{z}^*$, i.e.,

$$f^{out}(\mathbf{u}^*; \mathbf{W}^{out}) = f^{out}(\mathbf{z}^*; \mathbf{W}^{out}) + \mathbf{h}_{\mathbf{z}}^T \delta\mathbf{z} \quad (18)$$

where $\delta\mathbf{z} = \mathbf{u}^* - \mathbf{z}^*$ and $\mathbf{h}_{\mathbf{z}}^T = \nabla_{\mathbf{u}^*} f^{out}(\mathbf{u}^*; \mathbf{W}^{out})|_{\mathbf{u}^*=\mathbf{z}^*}$. Furthermore, (17) can be approximated by a Gaussian distribution (Bishop, 1995), i.e.,

$$p(t^* \mid \mathbf{z}^*, t^-, \mathbf{W}^{out}) = \frac{1}{Z_S} \exp\left(-\frac{\beta''}{2} \{\Delta t^*\}^2 \right) \quad (19)$$

where $1/\beta'' = 1/\beta(1+\mathbf{h}_{t^-}^T\mathbf{h}_{t^-}) + \mathbf{h}_{\mathbf{z}}^T\boldsymbol{\Sigma}\mathbf{h}_{\mathbf{z}}$ and $\Delta t^* = t^* - \mathbf{W}^{out} \cdot [\mathbf{z}^*, \mathbf{x}^*, t^-]$.

One can consider the $p(\mathbf{W}^{out} \mid D')$ term in (16). Expanding $D' = \{t_i, \mathbf{z}_i\}|_{i=1,2,\cdots,n.}$ and marginalizing over $\mathbf{u}_i$, we have

$$p(\mathbf{W}^{out} \mid D') = \int p(\mathbf{W}^{out}, \mathbf{u}_i \mid t_i, \mathbf{z}_i) d\mathbf{u}_i \quad (20)$$

Using the Bayesian rule and the conditional independence of t_i on $\mathbf{z}_i$ given $\mathbf{u}_i$, (20) can be re-written as

$$p(\mathbf{W}^{out} \mid D') = \frac{p(\mathbf{W}^{out})}{p(t_i \mid \mathbf{z}_i)} \int p(t_i \mid \mathbf{u}_i, \mathbf{W}^{out}) p(\mathbf{u}_i \mid \mathbf{z}_i) d\mathbf{u}_i \quad (21)$$

Taking (19) and (21) into (16), a Gaussian approximation of the predictive distribution can be obtained

$$p(t^* \mid \mathbf{z}^*, D) = \frac{1}{(2\pi\sigma_r^2)^{1/2}} \exp\left(\frac{\{t^* - f^{out}(\mathbf{z}^*, \mathbf{x}^*; \mathbf{W}^{out})\}^2}{2\sigma_r^2} \right) \quad (22)$$

where $\sigma_r^2 = 1/\beta'' + \mathbf{g}^T\mathbf{A}^{-1}\mathbf{g}$.

Then, according to (19), the variance σ_r^2 of the predictive distribution reads as

$$\sigma_r^2 = \frac{1}{\beta} + \frac{1}{\beta}\mathbf{h}_{t^-}^T\mathbf{h}_{t^-} + \mathbf{h}_{\mathbf{z}}^T\boldsymbol{\Sigma}\mathbf{h}_{\mathbf{z}} + \mathbf{g}^T\mathbf{A}^{-1}\mathbf{g} \quad (23)$$

where $\mathbf{A}$ is the Hessian matrix of $S(\mathbf{W}^{out})$, which is the summation of the likelihood and the priori over the weights.

$$S(\mathbf{W}^{out}) = \frac{\beta''}{2}\sum_{i=1}^{n}\left\{t_i - f^{out}(\mathbf{z}_i, \mathbf{x}_i; \mathbf{W}^{out})\right\}^2 + \frac{\alpha}{2}\left\|\mathbf{W}^{out}\right\|^2 \quad (24)$$

It is noticeable from the above formula that compared to the previous variance $\sigma_{t^*}^2 = 1/\beta + 1/\beta\mathbf{h}_{t^-}^T\mathbf{h}_{t^-} + \mathbf{g}^T\mathbf{A}^{-1}\mathbf{g}$, the estimate of uncertainty contains an additional term $\mathbf{h}_{\mathbf{z}}^T\boldsymbol{\Sigma}\mathbf{h}_{\mathbf{z}}$ that reflects the contribution of the prediction distribution from the variance of the input noise process. Given $\mathbf{z}^*$, the total variance of the predictive distribution is known, a $(1-\alpha)\%$ PI can be constructed

$$y^* \pm z^{1-\frac{\alpha}{2}}\left(\frac{1}{\beta} + \frac{1}{\beta}\mathbf{h}_{t^-}^T\mathbf{h}_{t^-} + \mathbf{h}_{\mathbf{z}}^T\boldsymbol{\Sigma}\mathbf{h}_{\mathbf{z}} + \mathbf{g}^T\mathbf{A}^{-1}\mathbf{g} \right)^{1/2} \quad (25)$$

where $z^{1-(\alpha/2)}$ is the $1-(\alpha/2)$ quantile of a normal distribution function with zero mean and unit variance.

4. EXPERIMENTS AND ANALYSIS

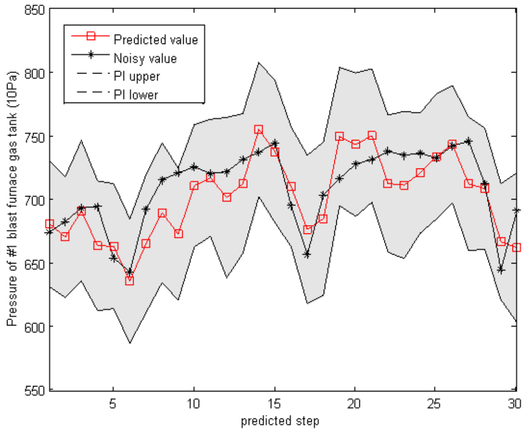
To verify the effectiveness of the proposed ESN model with output feedback, the gas flow and the pressure data coming from the energy data center of a certain plant of China are employed for the experiments. We randomly select the outlet pressure of #1, 2 gas tank as the validation point.

The related parameters of the proposed method are listed in Table 1. According to Figure 1, the outlet pressure of #1 gas tank is relevant to the outlet flows of #2 tank, the generation of #2,3 blast furnace, the flow of diffusion tower, the consumptions of #1,2,3 power plants, the consumption of #4 power generator and the consumption of CCPP. Similarly, the outlet pressure of #2 tank is relevant to the outlet flow of #1 gas tank, the generation of #2,3 blast furnace, the flow of diffusion tower, the consumption of #1,2,3 power plant, the consumption of #4 power generator and the consumption of CCPP. The dimensionality of the reservoir and other parameters of the ESN are set according to (Jaeger, 2002b). The initial value of hyper-parameters α and β'' is set as 5 and 50, respectively. The optimal process of hyper-parameters can be found in literature (Lauret, Fock, &Randrianarivony, 2008). For outlet pressure prediction of #1 gas tank, the final values of hyper-parameters α and β'' are optimized as 14.4587 and 103.8986, respectively. And, for those of #2 gas tank, the values of α and β'' are 19.5674 and 112.6395, respectively.

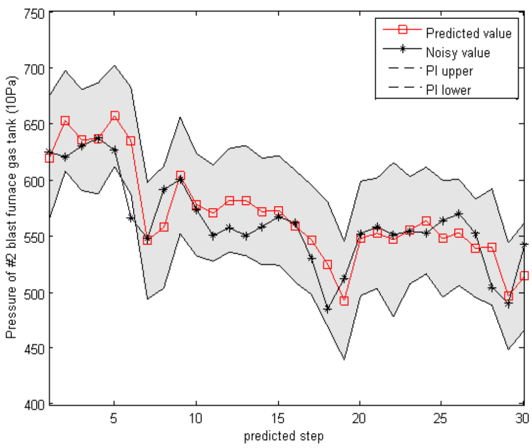
Table 1 The parameters of the proposed model

Parameters	value
Reservoir dimensionality	200
Sparse degree of the internal weights	0.02
Spectral radius of the internal weights	0.8
The number of training samples	1000
Initial value of α	5.0
Initial value of $\beta^{"}$	50.0

Fig.2 shows the results of the proposed method, the predicted pressures of #1 and #2 gas tanks, respectively. It can be seen that the proposed method exhibits a good performance with a prediction length 30min. And, the constructed PIs can provide some indicated information about the reliability of the prediction accuracy and present the possible interval that the targets located.



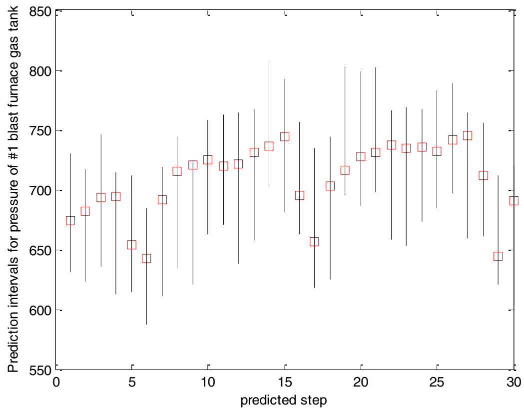
(a)



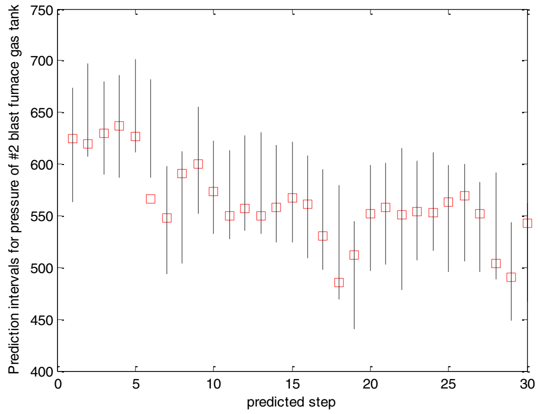
(b)

Fig. 2 The predicted pressure for #1 and #2 gas tank

Fig.3 shows a clearer illustration of PIs based on the proposed method with the confidence level 95%, in which the red square are used to represent the targeted values and the vertical lines denote the PIs. It is evident that the target values can be basically covered by the constructed PIs. The results in Fig.3 indicate that the proposed model is effective for the PIs construction of the outlet pressures of the gas tanks.



(a)



(b)

Fig.3 PIs for the pressure of #1 and #2 gas tank

Furthermore, to quantitatively evaluate the performance, statistical results are also reported in Table 2. In order to guarantee the indication of the statistics, the PIs construction is repeated by 50 times. Here, the root mean square error (RMSE) is used to measure the prediction quality,

$$RMSE = \sqrt{\sum_{i=1}^{n}(Y_i - F_i)^2 / n} \tag{26}$$

where n is the number of predicted data points, Y_i is the observation, and F_i is the predicted value. Although the RMSE of the pressure prediction surpasses 200 (Pa), the accuracy can completely meet the demands of the energy scheduling with consideration of the pressure order of 7000~8000 (Pa). Besides, the PI coverage probability (PICP) and the mean PI width (MPIW) are further adopted here.

$$PICP = \sum_{i=1}^{n_{test}} c_i / n_{test} \tag{27}$$

$$MPIW = \sum_{i=1}^{n_{test}} (U_i - L_i) / n_{test} \tag{28}$$

where c_i equals to 1 when the target is placed in the interval range; otherwise, c_i equals to 0. U_i is the upper bound, and

L_i is the lower bound; and , CWC that gives concerns to both the PICP and MPIW is a more comprehensive index, see

$$CWC = MPIW(1+\gamma(PICP)\exp(-\eta(PICP-\mu))) \quad (29)$$

where $\gamma(PICP)$ is given by

$$\gamma = \begin{cases} 0, & PICP \geq \mu \\ 1, & PICP < \mu. \end{cases}$$

From Table 2, it is apparently that the PI coverage probability of the pressures is superior to 95%. In this case, when the variable μ in (29) is set to 95%, the value of CWC is equal to that of MPIW. That is to say, the PI coverage probability based on an acceptable interval width can meet the practical requirements. The last row in Table 2 shows the computational cost (CC) of the proposed method, which indicates that the computing efficiency is acceptable compared to the prediction horizen. To sum up, the proposed approach obtains a satisfactory performance on PIs for the pipeline pressure.

Table 2 Statistical analysis of the PIs for pipeline pressure

Index	Pipeline pressure of #1 gas tank	Pipeline pressure of #2 gas tank
RMSE (10Pa)	22.5253	20.7134
PICP	1.0	0.9667
MPIW	104.4642	98.8186
CWC	104.4642	98.8186
CC (s)	3.65	2.47

5. CONCLUSIONS

The real-time prediction of BFG pipeline pressure in steel industry is a very significant issue on the energy scheduling task. An echo state network with output feedback is established in this study, where the reliability and the possible predicted ranges are quantitatively presented by using a Bayesian framework. Since the industrial data is always with noise, the input noise and the feedback noise are considered in the proposed model for PIs contribution. To verify the effectiveness of the proposed method, a series of experiments with the data coming from the practical database is conducted here, and the results indicate that the proposed approach has a satisfactory performance on the PIs construction of the pipeline pressure.

ACKNOWLEDGEMENTS

National Natural Sciences Fundation of China (61104157,61034003,61273037), National High-Tech R&D Program (2013AA040703), and the Fundamental Research Funds of China for the Central Universities (No. DUT12ZD214).

REFERENCES

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press.

Chryssolouris, G., Lee, M., &Ramsey, A. (1996). Confidence interval prediction for neural network model. *IEEE Transactions on neural networks*, vol. 7, pp. 229-232.

Han, M., & Mu, D. Y. (2010). Multi-reservoir echo state network with sparse Bayesian learning, *Adv. Neural Networks*, pp. 450–456.

Heskes, T. (1997) Practical confidence and prediction intervals. in *NeuralInformation Processing Systems*, vol. 9, T. P. M. Mozer and M. Jordan,Eds. Cambridge, MA: MIT Press, pp. 176–182.

Jaeger, H. (2002a). Tutorial on training recurrent neural networks, covering BPTT, RTRL, EKF and echo state network approach. *German National Research Center for Information Technology*, GMD Rep. 159.

Jaeger, H. (2002b). Short term memory in echo state networks. *German National Research Center for Information Technology*, GMD Report 152.

Jaeger, H., &Haass, H. (2004). Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless telecommunication. *Science,* vol. 304, pp. 78–80.

Khosravi, A., Nahavandi, S., Creighton, D., &Atiya, A. F. (2011). Comprehensive review of neural network based prediction intervals and new advances. *IEEE Transactions on neural networks*, vol. 22, pp. 1341-1356.

Lauret, P., Fock, E., &Randrianarivony, R. N. (2008). Bayesian neural network approach to short time load forecasting. *Energy Conversion and Management,* vol. 49, pp. 1156-1166.

Liu, Y., Liu, Q., Wang, W., Zhao, J., &Leung, H. (2012). Data-driven based model for flow prediction of steam system, Information Sciences, vol. 193, pp.104-114.

Nix, D. A., & Weigend, A. S. (1994). Estimating the mean and variance of thetarget probability distribution. in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 1, pp. 55–60.

Sheng, C., Zhao, J., Wang, W., & Leung, H. (2013). Prediction intervals for a noisy nonlinear time series based on a bootstrapping reservoir computing network ensemble. *IEEE Transactions on neural networks and learning systems*, vol. 24, pp. 1036-1048.

Taware, A., &Brown, R. H. (1999). Dynamic linear finite element model for pressure prediction in a gas pipeline. *Proceedings of the 38th IEEE conference on decision and control*, vol. 4, pp. 3248-3252.

Wright, W. A. (1999). Bayesian approach to neural-network modeling with input uncertainty. *IEEE Trans. Neural Network,* vol. 10, pp. 1261–1270.

Zhang, L. &Luh, P. B. (2005). Neural network based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method. *IEEE Transactions on power system*, vol. 20, pp. 59-66.

Zhu, X., &Leis, B. N. (2012). Evaluation of burst pressure prediction models for line pipes. *International Journal of Pressure Vessels and Piping,* vol. 89, pp. 85-97.