Int. J. Appl. Math. Comput. Sci., 2021, Vol. 31, No. 2, 321–335
DOI: 10.34768/amcs-2021-0022

DATA–DRIVEN ONLINE MODELLING FOR A UGI GASIFICATION PROCESS USING MODIFIED LAZY LEARNING WITH A RELEVANCE VECTOR MACHINE

SHIDA LIU [a], HONGHAI JI [a,*], ZHONGSHENG HOU [b], JIASHUO ZUO [c], LINGLING FAN [d]

[a] School of Electrical and Control Engineering
North China University of Technology
5 Jinyuanzhuang Road, Shijingshan District, Beijing, 100093, China
e-mail: lsdshiwo@hotmail.com, jhh@ncut.edu.cn

[b] School of Automation
Qingdao University
308 Ningxia Road, Laoshan District, Qingdao, 266071, China
e-mail: zhshhou@bjtu.edu.cn

[c] Robot Control Laboratory, School of Electronic Engineering
Yanshan University
No. 438, West Section of Hebei Street, Baitaling Street, Haigang District, Qingdao, 066004, China
e-mail: Chinazuojiashuo@stumail.ysu.edu.cn

[d] School of Automation
Beijing Information Science and Technology University
Qinghe Xiaoying Dong 12 Road, Haidian District, Beijing, 100092, China
e-mail: linglingfan@bistu.edu.cn

A modified lazy learning algorithm combined with a relevance vector machine (MLL-RVM) is presented to address a data-driven modelling problem for a gasification process inside a united gas improvement (UGI) gasifier. During the UGI gasification process, the measured online temperature of the produced crude gas is a crucial aspect. However, the gasification process complexities, especially severe changes in the temperature versus infrequent manipulation of the gasifier and the unknown noise in collected data, pose difficulties in dynamics process descriptions via conventional first principles. In the MLL-RVM, a novel weighted neighbour selection method is adopted based on the proposed dynamic cost functions. Moreover, the RVM is utilized in the implementation and design of the proposed online local modelling owing to its short test time and sparseness. Furthermore, the leave-one-out cross-validation technique is used for local model validation, by which the modelling performance is further improved. The MLL-RVM is applied to a series of real data collected from a pragmatic UGI gasifier, and its effectiveness is verified.

Keywords: data-driven modelling, UGI gasification process, relevance vector machine, modified lazy learning.

1. Introduction

Today, there exist various types of gasifiers applied in the coal gasification industry. Based on their working conditions, gasifiers can be classified into many types, including the Texaco gasifier (Yu *et al*., 2013), the Shell gasifier (Sun *et al*., 2011), the ALSTOM gasifier (Huang *et al*., 2013), the Lurgi gasifier (Papole *et al*., 2012), the UGI gasifier (Ren *et al*., 2004), the Winkler gasifier (Mondal *et al*., 2011) and others (Wei and Liu, 2014; 2013). In China, most coal gasification reactors are UGI gasifiers, which are widely applied in the synthetic ammonia industry.

The temperature of the gasification zone inside a UGI

*Corresponding author

gasifier is important during the gasification process but has not been measured straightforwardly to date. As an alternative, another two indices, the downdraft crude gas temperature (D-temperature) and the updraft crude gas temperature (U-temperature), are usually considered in practice. If both the U-temperature and D-temperature remain steady within permissible ranges, the gasification process is considered to operate in good conditions.

For practical industrial applications, it is expected that the U/D-temperature can be properly modelled such that the engineers can make prognoses as well as assess and monitor the gasification process. Moreover, by predicting the temperature and its tendency, we can devise reasonable manual/automatic control for the gasification process to achieve the objective of stable and high-efficiency production. Therefore, the research in this manuscript has its unique and important merit for the UGI gasification industry. Our objective is to create a mathematical model that describes the dynamics of U-temperature and D-temperature inside a UGI gasifier.

Some modelling methods based on physical or chemical principles have been proposed for different applications of different gasifiers. For instance, in fluidized-bed gasifiers, methods based on two-phase models are applied for different purposes, such as assessing the gas composition (Raman *et al*., 1981), optimizing the pressure and gas velocity (Fiaschi and Michelini, 2001), predicting the bed temperature, the heating value and production (Sadaka *et al*., 2002), and predicting the temperature and concentration profiles of gases (Gordillo and Belghit, 2011). Moreover, for a general class of fixed-bed biomass gasifiers, methods based on equilibrium models are applied in a wide scope, including in exergy analysis (Srinivas *et al*., 2009), assessment of the syngas composition variability (Simone *et al*., 2013), adjustment of gasifier working parameters (Altafini *et al*., 2003), and prediction of the composition and calorific value of syngas (Ruggiero and Manfrida, 1999). Furthermore, other unnamed first principle modelling methods are also used for gasification, such as calculation of the syngas composition (Petersen and Werther, 2005), generation of temperature profiles for the reduction zone of a downdraft biomass gasifier (Babu and Sheth, 2006), calculation of the temperature and axial profiles of concentration (Babu and Sheth, 2006; Corella and Sanz, 2005), etc.

However, the above-mentioned methods strongly depend on first principles, whose derivation is based on some assumptions that are rarely satisfied in practice. Furthermore, the first principles for complicated processes often result in sophisticated models that are difficult to use. In addition, large amounts of data are generated daily during industrial production processes since these available data contain all of the valuable information about the system. For these reasons, the data-driven modelling approach can be considered to study gasification processes.

In recent years, many global offline data-driven modelling methods aimed at different applications have been proposed in the gasification field, including artificial neural networks (Puig-Arnavat *et al*., 2013; Nougués *et al*., 2000; Simani *et al*., 2018), fuzzy methods (Shabbir *et al*., 2012; Zanoli *et al*., 2012), support vector machines (SVMs) (Han *et al*., 2008), multivariate regression (MVR) (Kalita *et al*., 2013; Chavan *et al*., 2012), reinforcement learning (RL) methods (Zhao *et al*., 2019), etc. However, in most cases, when applied to a system, it is difficult for a global model to cover the entire scope of the operation. In addition, the models established by applying these methods are laborious to renew online.

Recently, a memory-based technique for local learning, named lazy learning (LL), has developed rapidly (Aha, 1997). The modelling procedure of LL uses a query-based method to select the best model configuration at each query by assessing and comparing different alternatives. To date, LL has been successfully applied in modelling (Bontempi *et al*., 1999), time series prediction (Taieb *et al*., 2010), data-driven control (Hou and Wang, 2013; Hou and Xu, 2009), etc.

Although various innovative methods have been developed for different gasifiers (e.g., fluidized-bed gasifiers and downdraft biomass gasifiers) with different applications (e.g., estimation of the composition and adjustment of the gasifier's working parameters), few studies are proposed for modelling the gasification process, especially for the UGI gasifier. The contribution of this paper is to establish an online model within the UGI gasifier crude gas temperature with a modified lazy learning algorithm and relevance vector machine (MLL-RVM), in which two enhancement methods are introduced based on the classical lazy learning modelling algorithm.

One of the enhancements aims at the neighbour selection problem of the new query. The gasification process is known as a particular multiple-input, multiple-output (MIMO) nonlinear dynamic system, and some of its characteristics increase the difficulty of modelling, such as small and infrequent changes in some inputs, serious coupled outputs and unknown irregular noise existing in collected data. For these reasons, the classical neighbour selection method merely based on the distance of the Euclidean norm is not suitable here. Relying on these features, a new local modelling neighbour selection method is proposed. It uses a new dynamic cost function to select the neighbours of the query object.

The second improvement is for local modelling problems in the modelling process. The relevance vector machine (RVM) is considered. The RVM is a new machine learning implementation that is valid

for modelling and prediction of small-sample nonlinear systems (Tipping, 2001). To date, the RVM has been adopted in some pragmatic fields, including hyperspectral image classification (Demir and Erturk, 2007), channel equalization applications (Chen *et al*., 2001), automatic detection of clustered microcalcifications (Wei *et al*., 2005), etc. Based on the sparse Bayesian learning method, compared with the SVM and support LSSVM, the RVM has better performance and fewer kernel functions and is not limited by the Mercer condition (Tipping and Faul, 2003). Moreover, RVM sparsity has short test time characteristics and can be substantially constructed of nonlinear partial models.

The MLL-RVM is applied to a series of data collected from an actual UGI gasifier at the RuiXing Chemical Group, Shandong Province, China, whose effectiveness is verified via comparison with the conventional LL methods.

This is organized as follows. In Section 2, the gasification process is briefly introduced and the complexity of the problem is formulated. In Section 3, the MLL-RVM for modelling the gasification process is presented. The experimental results are shown in Section 4. A conclusion is drawn in Section 5.

2. UGI gasification process

2.1. Principle and work flow of the UGI gasification process. During the gasification process, some complex gas-solid reactions take place inside the UGI gasifier. Air, steam and oxygen can be supplied to the reactions as gasifying agents. Furthermore, a variety of gaseous products are generated, including CO_2, H_2, CO, H_2O and other gaseous hydrocarbons as well as small quantities of char, ash and several condensable compounds (tars and oils).

The schematic of a single gasification system is illustrated in Fig. 1. As a type of batch process, the gasification process occurring in a UGI gasifier is intermittent, periodic and repeated; that is, during each gasification cycle (batch) lasting for 150 seconds, gases (reactants) and feedstock are supplied to the gasifier in a fixed sequence for reaction by switching dozens of valves installed on different pipes that are connected to the gasifier. Basically, the UGI gasification process consists of several intermittent stages as follows.

Stage 1 (Aeration): At this stage, it will take approximately five seconds for a selected quantity coal to be fed from the top of the gasifier. At the same time, the air is blown into the gasifier through the aeration pipe (P2, Fig. 1) for approximately 28–40 seconds. An oxidation reaction then occurs to accumulate enough heat that will be used in the next step of the gasification cycle

Stage 2 (Steam upflow): Then, the depressurized steam is blown up into the gasifier through the pipe P3 and lasts approximately 10 seconds. During this period, the crude gas produced is discharged from the updraft rough gas pipe (P5, Fig. 1)

Stage 3 (Steam downflow): After the second stage, the vapor is blown downward through the steam downflow pipe P1 to the gasifier in a short time. Compared with Stage 2, the difference lies in the FAC that the produced crude gas at this stage is gathered and transmitted through the pipe P6.

When released from the dust collector, the crude gas must pass through the scrubber tower. Finally, the produced syngas is obtained and saved into the gas cabinet.

As an extra product, the slags gathered in the ash zone (f, Fig. 1) are continuously generated and squeezed into the left and right ash lock hopper (components B and C, Fig. 1) with the help of a rotating grate (component A, Fig. 1) at the bottom of the gasifier.

In each gasification cycle, Stages 1–3 will be performed in order.

2.2. Formulations of the U- and D-temperatures. As shown in Fig. 1, during the gasification process, the interior of a UGI gasifier is often divided into six zones. They are, from top to bottom, the (a) carbon-free zone, (b) drying zone, (c) pyrolysis zone, (d) reduction zone, (e) oxidation zone, and (f) ash zone. In practice, the gasification zone, which is an integration of the reduction zone and oxidation zone, is important during the gasification process since most reactions are conducted within it. If the gasification zone temperature is steadily maintained at an approximately ideal value within a permissible range, then the production efficiency, including the quantity and quality of the crude gas, will reach a satisfactory level required for practical production.

In fact, the gasification zone temperature cannot be measured in practice. As a substitution, process engineers in the field usually observe two other indices related to the gasification zone temperature: the updraft crude gas temperature (U-temperature) and the downdraft crude gas temperature (D-temperature), measured by Sensors 1 and 2 in Fig. 1, respectively. If both the D-temperature and U-temperature remain stable in the allowable ranges, it can be assumed that the gasification process is carried out under a fine status.

Many first-principle modelling methods are applicable on the basis of numerous strict assumptions. In practical UGI gasification processes, however, most of these assumptions cannot be satisfied. Furthermore, there are abundant U/D-temperature data that are collected daily from industrial fields, which can reflect the status and tendency of the gasification zone temperature. Therefore, it is necessary and possible to develop a

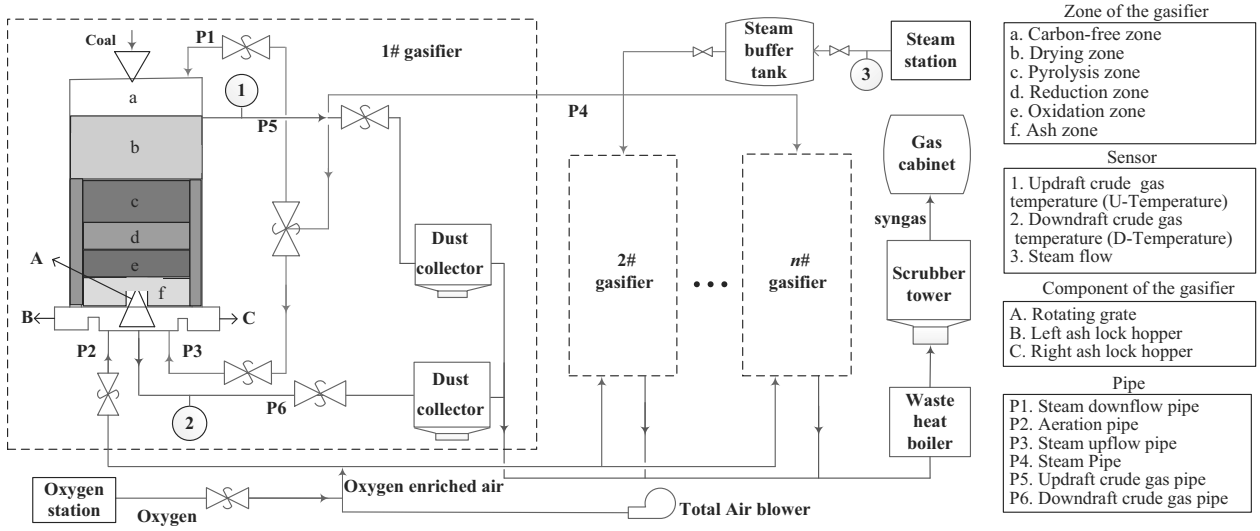

Fig. 1. Schematic of a single gasification system.

dynamic model for the U/D-temperature in a data-driven manner.

In the process of UGI gasification, there are many factors that affect the U/D-temperature and the corresponding data collection and storage of the distributed control system (DCS) including the steam upwelling time, steam downwelling time, rotating grate speed, steam flow, aeration time, etc. Only the inflation time and steam upflow are considered here because the time of each gasification cycle is a constant (150 seconds). Moreover, by regulating the speed of rotation of the grate, the gasifier can indirectly control the thickness of the gray zone, thus affecting the gasification zone.

From the above analysis, a general MIMO nonlinear discrete-time system describing the dynamical process of U-temperature $y^1(t)$ and D-temperature $y^2(t)$ is formulated as

$$\begin{aligned}\boldsymbol{y}(k+1) =& f(\boldsymbol{u}(k), \cdots, \boldsymbol{u}(k-n_u), \boldsymbol{y}(k),\\ & \cdots, \boldsymbol{y}(k-n_y), \boldsymbol{d}(k)),\end{aligned} \tag{1}$$

where

$$\boldsymbol{u}(k) = [u^1(k), u^2(k), u^3(k), u^4(k)]^T, \tag{2}$$

$$\boldsymbol{y}(k) = [y^1(k), y^2(k)]^T \tag{3}$$

indicate the control input and process output, respectively, k is the sampling index, and $f(\cdot)$ is an unknown nonlinear function. In particular, $u^1(t)$, $u^2(t)$, $u^3(t)$ and $u^4(t)$ represent the speed of the rotating grate, duration of aeration, duration of the steam upflow, and the steam flow, respectively.

Moreover, $\boldsymbol{d}(k)$ represents the noise and any disturbances that have a weak impact on the U/D-temperature, including observable factors such as the air temperature, air pressure, U-pressure and D-pressure. Furthermore, $\boldsymbol{d}(k)$ also represents other factors that cannot be measured in a quantitative manner, such as the shape of the slag inside the rotating grate. Because all data collected under formal working conditions are bounded, this gasification process is considered bounded-input, bounded-output (BIBO) stable.

3. DATA-driven modelling for the UGI gasification process

It can be seen in Section 2 that the U/D-temperature of the UGI gasification process can be expressed as (1)–(3). In this section, we will set up this relation by using the suggested MLL-RVM.

The idea of the classical lazy learning (CLL) method for modelling was systematically introduced by Aha (1997) and Aha *et al*. (1991), as well as Bontempi *et al*. (1999). The MLL-RVM is a modified version based on CLL. The enhancements are presented in two aspects: a novel local neighbour selection method and an efficient local modelling algorithm (RVM). The MLL-RVM will be introduced in Sections 3.1–3.3.

3.1. Local neighbor selection based on the dynamic cost function.

3.1.1. Introduction of the dynamic cost function.

For the input data $u(t)$, a relevant information vector $\boldsymbol{\varphi}(t)$

and weight vector $\mathbf{w}(t,i)$ are defined as

$$\boldsymbol{\varphi}(t) = [\boldsymbol{u}(t)^T, \ldots, \boldsymbol{u}(t-n_u)^T, \boldsymbol{y}(t)^T, \ldots, \boldsymbol{y}(t-n_y)^T]^T, \quad (4)$$

$$\boldsymbol{w}(t,i) = [\boldsymbol{w}_u^0(t,i), \ldots, \boldsymbol{w}_u^{n_u}(t,i), \boldsymbol{w}_y^0(t,i), \ldots, \boldsymbol{w}_y^{n_y}(t,i)]^T, \quad (5)$$

where

$$\begin{aligned}\boldsymbol{w}_u^j(t,i) &= [c_u^j \cdot \exp(-\beta_1 \cdot \left\|u^1(t-j) - u^1(i-j)\right\|_p), \ldots, \\ &\quad c_u^j \exp(-\beta_1 \cdot \|u^m(t-j) - u^m(i-j)\|_p)], \\ &\quad j = 0, \ldots, n_u. \end{aligned} \quad (6)$$

$$\begin{aligned}\boldsymbol{w}_y^l(t,i) &= [c_y^l \exp(-\beta_2 \cdot \left\|y^1(t-l) - y^1(i-l)\right\|_p), \ldots, \\ &\quad c_y^l \cdot \exp(-\beta_2 \cdot \|y^n(t-l) - y^n(i-l)\|_p)], \\ &\quad l = 0, \ldots, n_y, \end{aligned} \quad (7)$$

where $\|\cdot\|_p$ denotes the Minkowski L_p norm, and $\beta_1, \beta_2, c_u^j, c_y^l$, $j = 0, \ldots, n_u$, $l = 0, \ldots n_y$ are weight parameters. With this information vector, the original training set can be rebuilt as a set consisting of data pairs $\{\boldsymbol{\varphi}(t), \boldsymbol{y}(t+1)\}$. Remarkably, the $\|\cdot\|_p$ norm is a description of the distance of the neighbours, and the distance can also be replaced with other kinds of norms.

The distance similarity between the new query $\boldsymbol{\varphi}(t)$ and the i-th information vector $\boldsymbol{\varphi}(i)$ in the original training set is defined as

$$J_1(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(i)) = \left\|\boldsymbol{w}(t,i)^T(\boldsymbol{\varphi}(t) - \boldsymbol{\varphi}(i))\right\|_p. \quad (8)$$

Recent research has proven that similarity combining the distance and angle is more effective than using the distance similarity alone to improve the prediction accuracy of local models (Cheng and Chiu, 2004b). The angle similarity is defined as follows:

$$\cos(\theta_{ti}) = \frac{\langle \Delta\boldsymbol{\varphi}(t), \Delta\boldsymbol{\varphi}(i) \rangle}{\|\Delta\boldsymbol{\varphi}(t)\|_p \cdot \|\Delta\boldsymbol{\varphi}(i)\|_p}, \quad (9)$$

where $\Delta\boldsymbol{\varphi}(i) = \boldsymbol{\varphi}(i) - \boldsymbol{\varphi}(i-1)$. Then the dynamic cost function describing the comprehensive similarity between $\boldsymbol{\varphi}(t)$ and $\boldsymbol{\varphi}(i)$ in the data set is defined as

$$\begin{aligned}&J(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(i)) \\ &= \lambda \cdot \exp(-J_1(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(i))) + (1-\lambda) \cdot \cos(\theta_{ti}), \end{aligned} \quad (10)$$

where $J_1(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(i))$ and $\cos(\theta_{ti})$ are the distance similarity and the angle similarity between $\boldsymbol{\varphi}(i)$ and $\boldsymbol{\varphi}(i)$, respectively. In this way, λ is the weight parameter and only the distance similarity (or angle similarity) is adopted when $\lambda = 1$ (or $\lambda = 0$). It is worth mentioning that, if $\cos(\theta_{ti})$ is negative, then the corresponding $\boldsymbol{\varphi}(i)$ will be ignored since it has an opposite direction from the query.

Furthermore, the weight of each item in $J(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(i))$ is adjusted by $w_u^j(t,i)$, $j = 0, \ldots, n_u$, and $w_y^l(t,i)$, $l = 0, \ldots, n_y$, or especially by the relevant weight parameter. Taking $w_u^j(t,i)$ as an example, c_u^j, $j = 0, \ldots, n_u$ in $w_u^j(t,i)$ are used to approximately estimate the size of $w_u^j(t,i)$. When c_u^j is fixed, β_1 in $w_u^j(t,i)$ is used to accurately regulate the weights of all $\|u^s(t-j) - u^s(i-j)\|$, $s = 1, 2, \ldots m$ in $w_u^j(t,i)$. The effect of c_y^l and β_2 in $w_y^l(t,i)$ is similar to that of c_u^j and β_1.

Remark 1. Compared with the classical distance function based on the simple Euclidean norm (Bontempi *et al.*, 1999), this paper summarizes the novelty of the dynamic cost function from two perspectives. One is to consider the angular similarity, and the other is to introduce a special weight vector $\boldsymbol{w}(t,i)$ into the distance similarity $J(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(i))$ as described in (5)–(7). Note that $\boldsymbol{w}(t,i)$ may not be the focus of most existing research, but in this paper $\boldsymbol{w}(t,i)$ plays a leading role in the choice of neighbors.

The dynamic cost function (10) can be used to calculate the similarity between the new query $\boldsymbol{\varphi}(t)$ and $\boldsymbol{\varphi}(i)$ and each $\boldsymbol{\varphi}(i)$ in the training set. With regard to a single-input, single-output (SISO) or multiple-input, single-output (MISO) system, k data pairs $\{\boldsymbol{\varphi}(t_m), y(t_m+1)\}_{m=1,\cdots,k}$ are selected in descending order starting from the one with the largest similarity, and the local model can be built. With regard to a MIMO system, we can use different weight vectors $\boldsymbol{w}^r(t,i)$ to define different dynamic cost functions for each output, indicated as $J^r(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(i))$, $r = 1, \ldots, n$. Consequently, the relevant data pairs $\{\boldsymbol{\varphi}^r(t_m), y^r(t_m+1)\}_{m=1,\ldots,k}$ are chosen and employed for the local modelling for each output y^r, $r = 1, \ldots, n$.

3.1.2. Application of local neighbour selection in a gasification process. The MIMO gasification system depicted in Section 2 has many features that make the modelling process difficult. Aiming at these difficulties, based on the proposed dynamic cost function, a corresponding efficient neighbour selection strategy is adopted.

One feature is that very few changes in some input occur, and they are small. Generally, the duration of aeration u^2 and of steam upflow u^3 may not be adjusted more than once in a few hours, and the value must be an integer, often changing one or two values within the allowable range. For example, as shown in Fig. 2, u^2 is adjusted from 32 to 33 at approximately 22:42, and

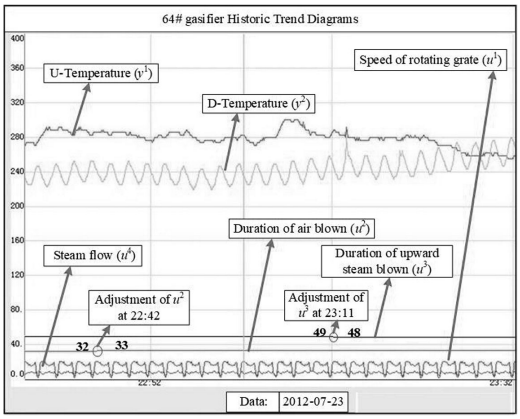

Fig. 2. Inputs and outputs displayed in the host computer of a DCS.

u^3 is adjusted from 49 to 48 at approximately 23:11. The corresponding outputs of the U-temperature and D-temperature, however, change sensitively during this period.

In the MLL-RVM, the approach to solve this problem is to determine the appropriate $\boldsymbol{w}(t,i)$ based on the correlation between $\boldsymbol{u}(t)^T,\ldots,\boldsymbol{u}(t-n_u)^T$ and $\boldsymbol{y}(t)^T,\ldots,\boldsymbol{y}(t-n_y)^T$ in $\boldsymbol{\varphi}(t)$ when selecting local neighbours from the training set. In particular, since u^2 and u^3 have infrequent changes, β_1 and c_u^j in $\boldsymbol{w}_u^j(t,i)$ should be set specifically smaller than β_2 and c_y^l when we select neighbours from the training set according to (10).

Remarkably, although u^2 and u^3 are adjusted infrequently and slightly as described above, they should not be ignored when selecting local neighbours since tuning u^2 and u^3 can affect the long-term tendency of the U/D-temperature. For instance, in Fig. 2, after adjusting the u^3 at approximately 23:11, U-temperature y^1 and D-temperature y^2 have downward and upward trends in the next 20 minutes, respectively.

Another characteristic is that there exists severe coupling during the gasification process between the D-temperature and U-temperature, which cannot be accurately obtained. In this case, the MLL-RVM uses the method of decomposing the MIMO system into two MISO systems: the U-temperature and D-temperature. When modelling the MISO system for the U-temperature (D-temperature), the message of the U-Temperature (D-temperature) is still taken into consideration. By doing this, although the severe coupling between outputs is unknown, the MIMO model for the UGI gasification process can still be guaranteed by the MLL-RVM with the help of two dynamic cost functions $J^r(\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(i))$ with different weight vectors $\boldsymbol{w}^r(t,i), r=1,2$.

3.1.3. Selection criterion of weight coefficients. Generally, the selection of weight coefficients c_u^j, β_1 (in $\boldsymbol{w}_u^j(t,i)$, $j=0,\ldots,n_u$) and c_y^l, β_2 (in $w_y^l(t,i)$, $l=0,\ldots n_y$) closely depends on the practical problem we studied. Therefore, it is difficult to give a series of strict mathematical methods to select the perfect weight coefficients for any given problem. For these reasons, we give a heuristic weight coefficient selection method only for the gasification process modelling problem investigated in this manuscript as follows. For other problems, the selection method can be ensured similarly.

First, the sum of all c_u^j and c_y^l, $j=0,\ldots,n_u$, $l=0,\ldots,n_y$ is set as constant 1. Furthermore, $c_u^j\in(0,0.4]$, $c_y^l\in(0,0.3]$. In practice, we usually first set each c_u^j as the same nonzero value and each c_y^l as a different one. After that, if the performance cannot achieve a satisfactory level, we will adjust the c_u^j and c_y^l slightly (plus or minus 0.1 for them gradually). As a result, the weight coefficients can be easily tuned in practice.

3.2. Relevance vector machine for local modelling. After the local neighbours for each output have been selected, the corresponding local models should be created next. Taking U-temperature y^1 as a example, an selected local training set consisting of k neighbours around $\boldsymbol{\varphi}(t)$ is described as

$$\{(\boldsymbol{\varphi}^1(t_1),y^1(t_1+1)),(\boldsymbol{\varphi}^1(t_2),y^1(t_2+1)),\ldots,(\boldsymbol{\varphi}^1(t_k),y^1(t_k+1))\}. \quad (11)$$

After the local training set has been selected, the next step is to use these neighbours to build a local model. In most previous works, the local weighted regression (Atkeson *et al.*, 1997) (LWR) based on the least-squares method (LSM) is commonly used for local modelling. However, given the small and infrequent changes of the duration of aeration u^2 and of steam upflow u^3, some matrices existing in th LSM become singular such that the corresponding regression parameter of the local linear model cannot be calculated.

As a substitution, this paper uses a small-sample RVM implemented under the probabilistic Bayesian learning framework to construct a basic nonlinear local model (Tipping, 2001). The RVM is more suitable for online modelling because of its sparse and shorter test time. Compared with the standard SVM and LS-SVM, the RVM based on the sparse Bayesian learning method can achieve better performance with a significantly reduced number of kernel functions and is not limited by the *Mercer conditions*. The process of creating a local RVM model is briefly described as follows.

The local RVM model can be described as learning a mapping $f:\boldsymbol{\varphi}(t)\rightarrow y^1(t+1)$ using the local 'training set.' Each $y^1(t+1)$ can be expressed as the sum of its

estimate $\hat{y}^1(t+1)$ and noise as

$$\begin{aligned} y^1(t+1) &= f(\omega^1(t);\boldsymbol{\varphi}(t)) + \varepsilon^1(t) \\ &= \sum_{m=0}^{k} (\omega_m^1(t)\cdot \boldsymbol{\phi}_m(\boldsymbol{\varphi}(t))) + \varepsilon^1(t), \end{aligned} \tag{12}$$

where $\omega^1(t)$ is a parameter vector, $\varepsilon^1(t)$ is an independent zero-mean Gaussian process and $\boldsymbol{\phi}_m(\boldsymbol{\varphi}(t))$ is the basis function denoted as $\boldsymbol{\phi}_m(\boldsymbol{\varphi}(t)) = K(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}^1(t_m))$. $K(\cdot)$ is the Gaussian kernel function adopted in most nonlinear modelling cases. The output is a linearly weighted sum of k nonlinear basis functions. Here, our goal is to assess the optimum $\omega^1(t)$. In this way, the output can be well generalized to new data since many elements of $\omega^1(t)$ are zero.

According to the assumption of the independence of $\boldsymbol{\varphi}(t)$, the likelihood of the local training set is denoted as

$$\begin{aligned} &p(\boldsymbol{y}^1(t)|\omega^1(t),\sigma^1(t)^2) \\ &\quad = (2\pi)^{-k/2}(\sigma^1(t))^{-k} \\ &\qquad \times \exp\Big(-\frac{\left\|\boldsymbol{y}^1 - \boldsymbol{\psi}^1(t)\omega^1(t)\right\|^2}{2(\sigma^1(t))^2}\Big), \end{aligned} \tag{13}$$

where $\boldsymbol{y}^1 = [y^1(t_1+1), \cdots, y^1(t_k+1)]^T$, ψ is a $k\times(k+1)$ 'design' matrix with

$$\boldsymbol{\psi}^1(t) = [\boldsymbol{\phi}(\boldsymbol{\varphi}^1(t_0)), \ldots, \boldsymbol{\phi}^1(\boldsymbol{\varphi}(t_k))]^T, \tag{14}$$

$$\boldsymbol{\phi}(\boldsymbol{\varphi}^1(t_0)) = 1, \tag{15}$$

$$\begin{aligned} \boldsymbol{\phi}(\boldsymbol{\varphi}^1(t_m)) = [&K(\boldsymbol{\varphi}^1(t_m), \boldsymbol{\varphi}^1(t_1)), \ldots, \\ &K(\boldsymbol{\varphi}^1(t_m), \boldsymbol{\varphi}^1(t_m))]^T, \quad m \neq 0. \end{aligned} \tag{16}$$

In the RVM, a Bayesian probabilistic framework is introduced to learn the general model of (12). First, the *a priori* probability distribution for $\omega^1(t)$ is given as

$$\begin{aligned} &p(\omega^1(t)|\alpha^1(t)) \\ &\quad = (2\pi)^{-k/2}\prod_{m=0}^{k}(\alpha_m^1(t))^{1/2} \\ &\qquad \times \exp\Big(-\frac{1}{2}\alpha_m^1(t)(\omega_m^1(t))^2\Big), \end{aligned} \tag{17}$$

where $\alpha^1(t)$= $[\alpha_0^1(t), \ldots, \alpha_k^1(t)]^T$ is a hyperparameter vector, and each $\alpha_m^1(t)$, $m = 0, \ldots, k$ is associated with $\omega_m^1(t)$ to moderate the strength of the prior. It is this form of prior that is ultimately responsible for the sparse properties of the model.

From the Bayesian rule, the posterior probability distribution of $\omega^1(t)$ is calculated as

$$\begin{aligned} &p(\boldsymbol{\omega}^1(t)|\boldsymbol{y}^1, \boldsymbol{\alpha}^1(t), (\sigma^1(t))^2) \\ &\quad = \frac{p(\boldsymbol{y}^1|\boldsymbol{\omega}^1(t),\sigma^1(t)^2)\cdot p(\boldsymbol{\omega}^1(t)|\boldsymbol{\alpha}^1(t))}{p(\boldsymbol{y}^1|\boldsymbol{\alpha}^1(t),\sigma^1(t)^2)} \\ &\quad = N(\boldsymbol{\omega}^1(t)|\boldsymbol{\mu}^1(t), \Sigma^1(t)), \end{aligned} \tag{18}$$

where the posterior mean and covariance are

$$\boldsymbol{\mu}^1(t) = (\sigma^1(t))^{-2}\Sigma^1(t)(\boldsymbol{\psi}^1(t))^T\boldsymbol{y}^1, \tag{19}$$

$$\Sigma^1(t) = (\sigma^1(t))^{-2}(\boldsymbol{\psi}^1(t))^T\boldsymbol{\psi}^1(t) + \boldsymbol{B}^1(t), \tag{20}$$

with $\mathbf{B}^1(t)$ denoting $\mathrm{diag}(\alpha_0^1(t), \ldots, \alpha_k^1(t))$.

The optimal $\boldsymbol{\alpha}_{\text{opt}}^1(t)$ can be found via the *type-II maximum likelihood* procedure. First, the marginal likelihood function is

$$\begin{aligned} &L(\boldsymbol{\alpha}^1(t)) \\ &\quad = \log(p(\boldsymbol{y}^1|\boldsymbol{\alpha}^1(t), (\sigma^1(t))^2)) \\ &\quad = \log\int_{-\infty}^{+\infty} p(\boldsymbol{y}^1|\boldsymbol{\omega}^1(t), \sigma^1(t))^2) \\ &\qquad \times p(\omega^1(t)|\alpha^1(t))\,\mathrm{d}(\omega^1(t)) \\ &\quad = -\frac{1}{2}[k\log 2\pi + \log|\boldsymbol{C}| + (\boldsymbol{y}^1)^T\boldsymbol{C}^{-1}y^1], \end{aligned} \tag{21}$$

with

$$\boldsymbol{C} = (\sigma^1(t))^2)\boldsymbol{I} + \boldsymbol{\psi}^1(t)(\boldsymbol{B}^1(t))^{-1}(\boldsymbol{\psi}^1(t))^T. \tag{22}$$

Then, the *fast marginal likelihood maximization method* (Tipping and Faul, 2003) is applied for maximization of the marginal likelihood (21), and $\boldsymbol{\alpha}_{\text{opt}}^1(t)$ is finally obtained via numeric iteration. The optimal $\boldsymbol{\omega}_{\text{opt}}^1(t) = \boldsymbol{\mu}_{\text{opt}}^1(t)$ and $\sigma_{\text{opt}}^1(t)$ is then obtained by (19) and (20).

It is worth mentioning that the final optimal values of many hyperparameters are infinite (Tipping, 2001). According to (19), this result leads to some posterior parameters that are infinitely peaked at zero for many weights $\omega_m^1(t)$. Correspondingly, $\omega_{opt}^1(t)$ comprises a small number of nonzero elements. As a consequence, the RVM model (12) has been obtained. For a new query $\boldsymbol{\varphi}(t)$, the estimated output is

$$\hat{y}^1(t+1) = (\boldsymbol{\mu}_{\text{opt}}^1(t))^T\cdot\boldsymbol{\phi}(\boldsymbol{\varphi}(t)). \tag{23}$$

Likewise, for the D-temperature, the corresponding local model can be calculated as

$$\hat{y}^2(t+1) = (\boldsymbol{\mu}_{\text{opt}}^2(t))^T\cdot\boldsymbol{\phi}(\boldsymbol{\varphi}(t)). \tag{24}$$

3.3. Local model validation. Based on the dynamic cost function, k ($k_{\min} \le k \le k_{\max}$) related samples can be chosen for building the RVM model for each output. Generally, two parameters $k_{\min}$ and $k_{\max}$ are chosen. For each $\boldsymbol{u}(t)$ and the corresponding output (taking y^1, for example), there are ($k_{\max} - k_{\min}$+1) local models created. How to choose an optimal one as the final local model from these candidates will be discussed in this section.

To date, the *leave-one-out* (LOO) cross-validation method has been the most common way to perform model validation since it is efficient and easy to implement in

practice (Garcia *et al*., 2010). The main steps for selecting the optimal local model for $y^1(t)$ are described as follows.

Step V1: Initialize $k = k_{\min}$. Given a $\boldsymbol{u}(t)$, $\boldsymbol{\varphi}(t)$ is denoted, and then the local RVM model $f_k^1(\boldsymbol{\varphi}(t))$ can be created by using the local training set (11).

Step V2: Remove $(\boldsymbol{\varphi}^1(t_1), y^1(t_1+1))$ from the local training set and create a submodel using the remaining $(k-1)$ points. After that, $\hat{y}^1(t_1+1)$, and the estimate of $y^1(t_1+1)$ can be calculated. Then the error $e_k^1(t_1)$ is calculated as

$$e_k^1(t_1) = \hat{y}^1(t_1+1) - y^1(t_1+1). \tag{25}$$

Step V3: Likewise, each data pair of the local training set is removed once, and then the generalization error vector $\boldsymbol{E}_k^1(t)$ of $f_k^1(\boldsymbol{\varphi}(t))$ is calculated as

$$\begin{aligned} E_k^1(t) = {} & \frac{1}{\sum_{m=1}^{k} a_m} \\ & \times [a_1 e_k^1(t_1), a_2 e_k^2(t_2), \ldots, a_k e_k^k(t_k)]^T, \end{aligned} \tag{26}$$

where a_m is the weight coefficient of the corresponding element of $\boldsymbol{E}_k^1(t)$, here $a_m = 1/k$, $m = 1, \ldots, k$.

Step V4: Define $CV_k^1(t)$ as the LOO generalization error, which is calculated as

$$CV_k^1(t) = \boldsymbol{E}_k^1(t)^T \cdot \boldsymbol{E}_k^1(t). \tag{27}$$

Step V5: Set ε as a sufficiently small positive constant:

If $CV_k^1(t) \le \varepsilon$, then k is the optimal size for local modelling, denoted as k_{opt} and go to Step V7.

If $CV_k^1(t) > \varepsilon$ and $k < k_{\max}$, add the $(k+1)$-th nearest point to the local training set, let $k = k+1$ and go to Step V6;

If $CV_k^1(t) > \varepsilon$ and $k = k_{\max}$, this indicates that the validation vector

$$CV^1(t) = [CV_{k_{\min}}^1(t), CV_{k_{\min}+1}^1(t), \ldots, CV_{k_{\max}}^1(t)]^T$$

has been obtained. The minimal element of $CV^1(t)$ as $CV_{\text{opt}}^1(t)$ and go to Step V7;

Step V6: Let $k = k+1$, add the $(k+1)$-th neighbour $(\boldsymbol{\varphi}^1(t_{k+1}), y^1(t_{k+1}+1))$ to the local training set, form $f_{k+1}^1(\boldsymbol{\varphi}(t))$, and calculate $CV_{k+1}^1(t)$ and $\hat{y}_{k+1}^1(t+1)$ similarly to Steps V1–V4.

Step V7: Finally, the optimal local model is obtained, denoted as $CV_{\text{opt}}^1(t)$. Correspondingly, $f_{\text{opt}}^1(\boldsymbol{\varphi}(t))$ is selected as the optimal local model for estimating $y^1(t_1+1)$, by which $\hat{y}^1(t+1)$ is calculated as

$$\hat{y}^1(t+1) = \hat{y}_{\text{opt}}^1(t+1) = f_{\text{opt}}^1(\boldsymbol{\varphi}(t)). \tag{28}$$

3.4. Layout of the MLL-RVM for the UGI gasification process. The detailed modelling steps for the UGI gasification process by using the MLL-RVM are as follows:

Step L1: Let $k \in [k_{\min}, k_{\max}]$ and initialize $k = k_{\min}$. For the UGI gasification process, $m = 4$ and $n = 2$. Given $\boldsymbol{u}(t)$, the information vector $\boldsymbol{\varphi}(t)$ is first calculated.

Step L2: First, for y^1, k neighbours around $\boldsymbol{\varphi}(t)$ are selected from the original training set according to the dynamic cost function $J^1(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(i))$ mentioned in Section 3.1.1. Then the local training set such as (11) can be formed.

Step L3: Create the candidate local models $f_k^1(\boldsymbol{\varphi}(t))$ and select the optimal size k_{opt} for local modelling by using LOO cross-validation mentioned in Subsection 3.3.

Step L4: A local RVM model $f_{\text{opt}}^1(\boldsymbol{\varphi}(t))$ is determined using the local training set and $\hat{y}_{\text{opt}}^1(t+1)$ can be calculated.

Step L5: For the estimation of $y^2(t+1)$, the corresponding dynamic cost function is defined as $J^2(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(i))$, and then a local training set for $y^2(t+1)$ is selected as

$$\{(\boldsymbol{\varphi}^2(t_1), y^2(t_1+1)), \ldots, (\boldsymbol{\varphi}^2(t_k), y^2(t_k+1))\}. \tag{29}$$

Similarly to Steps L1–L4, keeping $\boldsymbol{\varphi}(t)$ unchanged and calculating $\hat{y}^2(t+1)$ as

$$\hat{y}^2(t+1) = \hat{y}_{\text{opt}}^2(t+1) = f_{\text{opt}}^2(\boldsymbol{\varphi}(t)), \tag{30}$$

the final output vector is calculated as

$$\begin{aligned} \hat{y}(t+1) &= [f_{\text{opt}}^1(\boldsymbol{\varphi}(t)), f_{\text{opt}}^2(\boldsymbol{\varphi}(t))]^T \\ &= [\hat{y}^1(t+1), \hat{y}^2(t+1)]^T. \end{aligned} \tag{31}$$

Step L6: Store $(\boldsymbol{u}(t), \boldsymbol{y}(t))$ and wait for $\boldsymbol{u}(t+1)$; if $\boldsymbol{u}(t+1)$ appears, $\boldsymbol{\varphi}(t+1)$ is inferred and $\hat{\boldsymbol{y}}(t+2)$ is calculated similarly to Steps L1–L5.

Remark 2. Compared with other online modelling methods, such as recurrent neural networks and RBF neural networks, the proposed MLL-RVM is a local modelling method and creates a local model at each operating point. To achieve an optimal local model, model structures and model parameters at each operating point are different. Moreover, compared with the BP neural network, the greatest advantage of the proposed MLL-RVM is the online modelling property. At each time instant, the local model can be established with the most recent controller parameters. However, in the BP neural network, the model should be established in advance with an offline database.

Remark 3. Moreover, in practice, if the computing power of the device is limited, the local model validation procedure can be simplified. Specifically, the values of $k_{\max}$, $k_{\min}$, and $k_{\max} - k_{\min}$ can be set as relatively small according to practical situations. Moreover, the size of the original training set is also flexibly determined by the capacity and memory of the facilities. When the memory is full, many rarely used data would be removed from the original training set since an excessively large database will cause a prohibitively long computing time. The proposed MLL-RVM method can also be applied to other industrial processes in which the precise mathematical description of the system cannot be established, such as a water level control system and other complex systems.

Remark 4. Nelles and Isermann (1996) propose a local linear model tree (LOLIMOT) method based on the idea to approximate a nonlinear function with piece-wise linear models. Compared with LOLIMOT, the advantages of the proposed method are concluded as follows. Firstly, in the LOLIMOT method, a global linear model needs to be first established as a hyper-rectangle using all data points. Then, the hyper-rectangle is cut from every input dimension and the output is finally estimated by summing parts of local linear models with normalized Gaussian weighted functions. However, in the gasification process, two control inputs (the duration of aeration, and of the steam upflow) must be set as integers and change infrequently, and the changes have significant influence is the gasification process. Consequently, the data set of the UGI gasification process cannot satisfy the continuous excitation condition, which makes the global linear model reasonably difficult. In MLL-RVM, the local training set used to build a local model is first obtained by selecting a subset of the whole data set, which has a high similarity with the current quire. Consequently, the global linear model is not needed. Moreover, the LOLIMOT method in the work of Nelles and Isermann (1996) mainly focuses on the single-input and single-output nonlinear system. However, the UGI gasification process is a typical multi-input and multi-output system, and the coupling between the outputs is difficult to handle using the LOLIMOT method. In the MLL-RVM the coupling between the outputs is also considered to select the local training set by dynamic cost functions, such that the modelling performance is satisfactory.

Remark 5. The support vector machine is also a common modelling method. Compared with the conventional SVM, the advantages of the MLL-RVM can be concluded as follows. First, the SVM method is essentially a small-sample modelling one. When the training set is large, the amount of computations using an SVM will be greatly increased, whereas the proposed MLL-RVM creates a local model at each operating point using l local neighbours. Thus, the amount of computations of the MLL-RVM method is reduced. Furthermore, compared with existing global modelling methods, the advantage of the MLL-RVM is also its small computation burden, since the local model is created at each operating point using only l local neighbours selected from the training set. Moreover, the MLL-RVM is an online modelling method, and different RVM local models can be established online with a small amount of computations using the corresponding local neighbours such that the modelling accuracy can be ensured.

4. Experimental results

To confirm the usefulness and applicability of the MLL-RVM, this section uses data from the Ruixing Chemical Group Company in the Shandong Province, China, for experiments. The actual data used are from the database of the Plant 64 # gasifier collected during the period of 1–20 April 2012. As discussed in Section 2.2, although the collection consists of many variables that are stored in a database, only six variables are ultimately used for modelling, including the steam surge time, aeration time, rotary grate speed, U/D-temperature, and steam flow. The original data set contains 3000 consecutive records, with the first 2400 data records used as the training set and the remaining 600 as the test data set. The data measured by various sensors in the field implementation are converted by ADC and stored in a database every five seconds. To investigate the data conveniently for a long-term prediction, the sampling time is extended to five minutes.

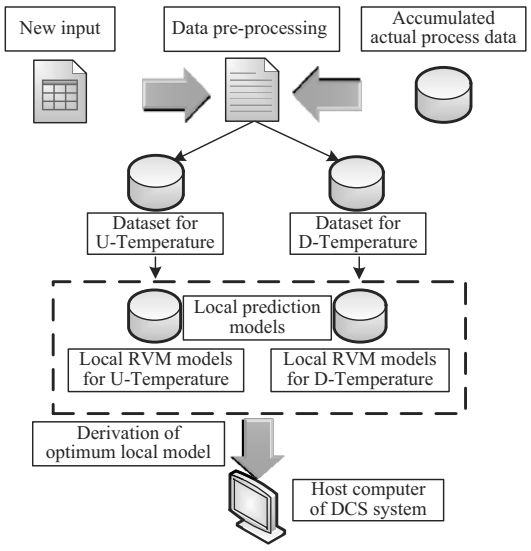

Fig. 3. Screen of a general-purpose module for online modelling of the U/D-temperature with the MLL-RVM.

Table 1. Simulation parameters of the MLL-RVM.

Public parameters of $\boldsymbol{J}^1$ and $\boldsymbol{J}^2$			β_1	β_2	P	λ
			0.1	0.1	2	0.8
	n_u	n_y	c_u^1	c_u^1	c_y^1	c_y^2
$\boldsymbol{J}^1$	2	2	0.2	0.2	0.3	0.3
$\boldsymbol{J}^2$	2	2	0.1	0.1	0.4	0.4

Table 2. Different weight coefficient settings of the MLL-RVM.

Coefficient		c_u^1	c_u^1	c_y^1	c_y^2
Scenario 1	$\boldsymbol{J}^1$	0.2	0.2	0.3	0.3
	$\boldsymbol{J}^2$	0.1	0.1	0.4	0.4
Scenario 2	$\boldsymbol{J}^1$	0.4	0.4	0.1	0.1
	$\boldsymbol{J}^2$	0.4	0.4	0.1	0.1

4.1. Design of the U/D-temperature predicting module. To perform the U/D-temperature prediction, the proposed method requires a large volume of historical data that have been preprocessed in advance. Consequently, all of the procedures of the MLL-RVM are implemented by utilizing a commercially available, general-purpose data processing module, as presented in Fig. 3.

As is shown in Fig. 3, the icon of 'Read database' has a function to read the actual process data from the database. The icon of 'Data preprocess' denotes preprocessing the collected data to check and delete anomalous data. The icons of 'Dataset for U/D-Temperature' have the corresponding dynamic cost functions described in Section 3.1.1, which are used to extract data for modelling the U-temperature and D-temperature, respectively. The icons of 'Local RVM models for U/D-Temperature' have functions that are equivalent to the modules for generating and validating local RVM models described in Sections 3.2 and 3.3. When clicking the icon of 'New input data', the current and history query can be seen in the corresponding screen. The icon of 'Host computer' is used to display the modelling results of U/D-temperature. By using this module, the MLL-RVM can be taken into account for the practical system and easily implemented by operators who are not specialists in computer science or programming.

4.2. Analysis of weighting coefficients. As discussed in Section 3.1.2, during the modelling process using the MLL-RVM, the weight coefficients, including $J^{r_2}(\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(i))$, $r_2 = 1, 2$ and $\beta_1, \beta_2, c_u^j, c_y^l$, are fairly important for local neighbour selection. Thus, the influence of the weight coefficients for the MLL-RVM will be analyzed in this section. The criterium is the mean absolute error

$$\text{MAE} = \frac{1}{N_{ts}} \sum_{k=1}^{N_{ts}} |\hat{y}(k) - y(k)|, \tag{32}$$

where N_{ts} denotes the size of the test set and $\hat{y}(k)$ denotes the model output at the k-th time. The parameters used in MLL-RVM are listed in Table 1. To facilitate the comparison, two types of weight setting scenarios are adopted in the implementation of the MLL-RVM, as shown in Table 2. Here β_1 and β_2 are set to be the same value for both scenarios.

Figure 4 shows the U-temperature distribution of the MLL-RVM for 600 consecutive test sets and different weighting coefficient settings. It can be seen that the performance of Scenario 1 is better than that of Scenario 2 since the MAE of Scenario 1 is 5.6738°C whereas that of Scenario 2 is 6.3324°C.

However, although a rational weight setting scenario can achieve a comparatively well-pleasing performance, the U-temperature error between the measured value and the predicted output has scarcely been further decreased. The reason is that the current data set cannot fully collect useful information that may come from other adjustable factors for U-temperature, such as the type of coal and the height of the free carbon layer, etc.

Figure 5 shows the D-temperature distribution of the MLL-RVM for 600 consecutive test sets and different weighting coefficient settings. Similarly to the U-temperature, compared with that in Scheme 2 (4.6672°C), the calculation error of the D-temperature in Scheme 1 (3.8491°C) has smaller MAE. In addition, from the comparison between the results shown in Figs. 4 and 5, the predicted results of the D-temperature for both scenarios (3.8491°C for Scenario 1 and 4.6672°C for Scenario 2) are superior to those of the U-temperature (5.6738°C for Scenario 1 and 6.3324°C for Scenario 2). The reason is that the measurement point of the D-temperature (Sensor 2, Fig. 1) is closer to the gasification zone (marked 'a', Fig. 1) than that of the U-temperature, and the data collected in this way contain more D-temperature information than U-temperature information. Thus, a higher accuracy of the D-temperature is guaranteed.

Summarizing Figs. 4 and 5, the results show that, regardless of the U-temperature or D-temperature, the precision of Scheme 1 is higher than that of Scheme 2. In fact, in the gasification process, the aeration u^2 and steam upwelling u^3 undergo minor changes and are not frequent. From (4), the relevance of $\boldsymbol{u}(t)^T, \ldots, \boldsymbol{u}(t - n_u)^T$ to $\boldsymbol{\varphi}(t)$ in (4) would play a subordinate role compared with that of $\boldsymbol{y}(t)^T, \ldots, \boldsymbol{y}(t - n_y)^T$. Just as in Table 2, in Scenario 1, $c_u^i, i = 1, 2$ of $\boldsymbol{J}^1$ and $\boldsymbol{J}^2$ are set as 0.2 and 0.1, respectively, whereas $c_y^i, i = 1, 2$ for both dynamic cost functions is 0.4. Under such a factor setting, the performance of Scenario 1 at the U/D-temperature can be

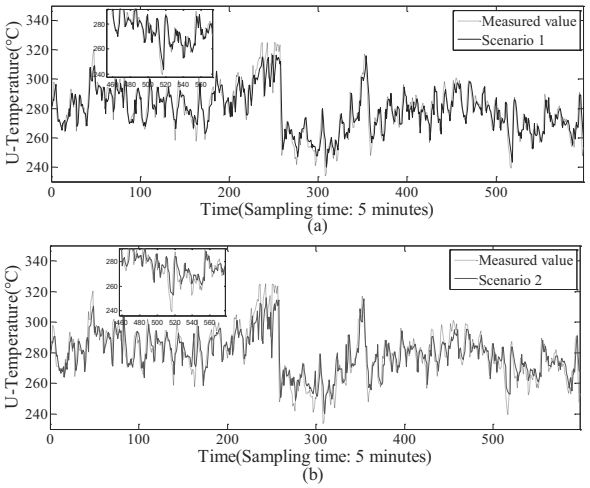

Fig. 4. U-temperature with different weight setting scenarios: Scenario 1, MAE = 5.6738°C (a), Scenario 2, MAE = 6.3324°C (b).

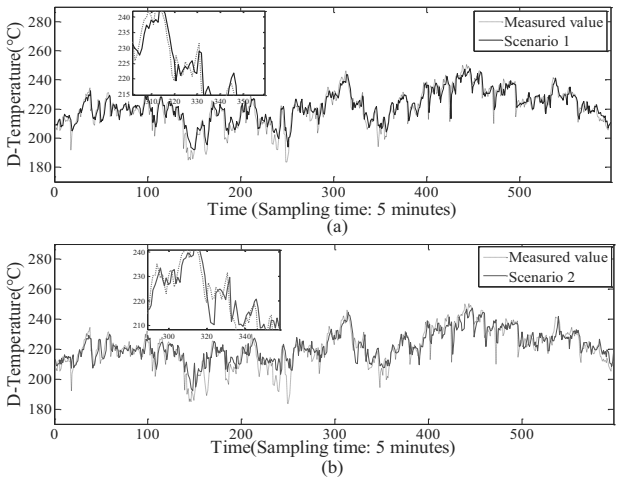

Fig. 5. D-temperature with different weight setting scenarios: Scenario 1, MAE = 3.8491°C (a). Scenario 2, MAE = 4.6672°C (b).

guaranteed to be better than that of Scenario 2.

Remark 6. From Figs. 4 and 5, it can also be seen that U-temperature is higher than the D-temperature. The reason is that, during the gasification process, the temperature of the gasification zone is highest, and the location of the U-temperature is closer to the gasification zone.

4.3. Local model validation via LOO cross-validation. As discussed in Section 3.3, for each $\boldsymbol{\varphi}(t)$, there exist $(k_{\max} - k_{\min} + 1)$ local models created with a different number of neighbors $k \in [k_{\min}, k_{\max}]$. The one with the satisfactory LOO generalization error can be selected as the final local model. The effectiveness and necessity of local model selection and validation are analyzed as follows.

4.3.1. Effectiveness of local model validation. Figure 6 illustrates the results of the optimal number of neighbours $k^1_{\text{opt}}(t)$ for the U-temperature and the corresponding generalization error $CV^1_{\text{opt}}(t)$. Figures 6(a) and (b) present the profile of the optimal number of neighbors $k^1_{\text{opt}}(t)$ changing from four to ten at each query $\boldsymbol{\varphi}(t)$ in the test set. Figure 6(c) illustrates the corresponding generalization error of $k^1_{\text{opt}}(t)$. From Figs. 6(a) and (b), it can be seen that, for the same $\boldsymbol{\varphi}(t)$, there are different numbers of neighbors for different weight setting scenarios. Furthermore, in Fig. 6(c), it is obvious that the generalization error $CV^1_{\text{opt}}(t)$ of Scenario 1 is superior to that of Scenario 2 since the mean value of $CV^1_{k_{\text{opt}}}(t)$ is 34.8127°C, decreasing by approximately 12°C compared with that of Scenario 1 (48.513°C). Therefore, the accuracy of the MLL-RVM with weight setting Scenario 1 is more satisfactory than that of Scenario 2.

Likewise, similar results for the D-temperature are shown in Fig. 7. Obviously, although there exist different $k^2_{\text{opt}}(t)$ with different weight setting scenarios similar to Fig. 7, the corresponding $CV^2_{\text{opt}}(t)$ for both scenarios are superior to those of $CV^1_{\text{opt}}(t)$. Therefore, a higher accuracy for the D-temperature can be achieved than for the U-temperature. It is worth mentioning that the generalization error is only an index to estimate the generalization ability of the local model, rather than the real model output error. From Figs. 6 and 7, it can be concluded that the MLL-RVM with weight setting Scenario 1 should have a higher accuracy than that of Scenario 2. Furthermore, the same conclusion reached in Section 4.4. Consequently, the effectiveness of local model validation by using *LOO* cross-validation is verified.

4.3.2. Necessity of local model selection. In some special cases, the number of neighbors of local modelling for each $\boldsymbol{\varphi}(t)$ is fixed. In this manuscript, however, given the existing noise in the original data and the characteristics of the MIMO gasification process, the data become irregular. To achieve a satisfactory modelling accuracy, different numbers of neighborus ($k^1_{\text{opt}}(t)$ and $k^2_{\text{opt}}(t)$ for the U/D-temperature, respectively) are selected for the local models at each $\boldsymbol{\varphi}(t)$.

Table 3 illustrates different range settings for $k^i_{\text{opt}}(t)$, $i = 1, 2$, and the corresponding *LOO* generalization error $CV^i_{opt}(t)$, $i = 1, 2$ for the U/D-temperature. Obviously, in Scheme 1, by automatically selecting the optimal $k^i_{\text{opt}}(t)$, $i = 1, 2$, more satisfactory local models for

Table 3. Different range settings for $k^i_{\text{opt}}(t)$, $i = 1, 2$ and the corresponding LOO cross-validation error $CV^i_{\text{opt}}(t)$, $i = 1, 2$, for the U/D-temperature.

Range of $k^i_{\text{opt}}(t)$, $i = 1, 2$		U-temperature Mean of $CV^1_{\text{opt}}(t)$	D-temperature Mean of $CV^2_{\text{opt}}(t)$
Scheme 1	[4,10]	34.8127°C	20.1817°C
Scheme 2	7	54.2228°C	33.4350°C

the U/D-temperature, respectively, can be determined by LOO cross-validation at each $\boldsymbol{\varphi}(t)$. In Scheme 2, however, since $k^i_{\text{opt}}(t)$, $i = 1, 2$, are fixed as 7, only one local model at query $\boldsymbol{\varphi}(t)$ is created for each output and the model must be used without consideration of the LOO generalization error. Consequently, the mean values of $CV^1_{\text{opt}}(t)$ and $CV^2_{\text{opt}}(t)$ for the U/D-temperature using Scheme 1 are superior to those of Scheme 2. Therefore, the necessity of local model selection is verified.

4.4. Comparison with the classical local modelling algorithm. In the MLL-RVM, the RVM is utilized for local modelling. To verify its effectiveness, the linearly weighted average (LWA) local modelling method is introduced here for comparison. Based on the LWA, the local modelling method is designed as (taking U-temperature $y^1(t+1)$, for example)

$$\begin{aligned} \hat{y}^1(t+1) &= \sum_{m=1}^{k} w(t_m) y^1(t_m), \\ \sum_{m=1}^{k} w(t_m) &= 1, \end{aligned} \tag{33}$$

where $\hat{y}^1(t+1)$ is the model output at time t, $y^1(t_m)$ and $w(t_m)$ are the measured output and the corresponding weight of the m-th neighbour in the 'local training set', respectively. k is the size of neighbours and $w(t_m) = 1/k$.

Table 4 illustrates the prediction errors of the U/D-temperature by using different local modelling algorithms. The definition of MaxAE and MinAE are defined as $\max_{i \in N_{ts}} |\hat{y}(i) - y(i)|$ and $\min_{i \in N_{ts}} |\hat{y}(i) - y(i)|$. Obviously, the performance of the MLL-RVM is superior to that of MLL-LWA for the U-temperature, and the MAE of the MLL-RVM is 5.67°C, improving approximately by 7% compared with that of the MLL-RVM (6.00°). For the D-Temperature, the MAE of MLL-RVM is 3.84°C, also improving approximately 7% compared with that of the MLL-LWA (4.16°C). Furthermore, MaxAE and MinAE, key indexes to verify the maximum and minimum error of each sampling point throughout the whole test set, respectively, are also improved.

Table 4. Prediction errors of the U/D-temperature with different local modelling algorithms.

Criterion	U-temperature		D-temperature	
	MLL-LWA	MLL-LWA	MLL-RVM	MLL-RVM
MAE	6.00°C	**5.67**°C	4.16°C	**3.84**°C
MaxAE	44.86°C	**41.78**°C	25.97°C	**24.42**°C
MinAE	2.02°C	**0.04**°C	0.02°C	**0.01**°C

5. Conclusions

In this work, focusing on the online modelling of the gasification process in a UGI gasifier in the fixed-bed intermittent gasification industry, a combination of an improved lazy learning method and a correlation vector machine (MLL-RVM) was proposed to study the problem. For the gasification process expressed as a class of automated nonlinear MIMO systems in terms of the U/D-temperature, there are some factors that make it very difficult to build a model from first principles. Based on the above, two improved technologies based on classical lazy learning are developed in the MLL-RVM, including a novel neighbour selection method based on the proposed dynamic cost function and the introduction of the RVM for establishing the local model. Furthermore, the *leave-one-out* cross-validation technique is used for local model validation and selection, by which the modelling performance is further improved. The MLL-RVM is successfully applied to a set of data collected from a practical UGI gasifier in a large chemical group of China. By comparison with the standard local weighted average (LWA) algorithm, the effectiveness of the MLL-RVM is verified.

In the future, we will design a controller to control the U-temperature and D-temperature based on the established data model in the simulation platform. For the established model of this paper, data driven control approach can be used. Specifically, the data driven control methods design the controller only using the input and output data of the controlled plant, and do not employ any model parameters of the control plant. As a typical data driven method, the model free adaptive control (MFAC) method can be used for the temperature control of the UGI gasification process, and the corresponding detailed works are described by Liu *et al*. (2020), where the controlled model is an offline and global model. Unlike Liu *et al*. (2020), the model in this manuscript is a local and online model, and it has more adaptivity during the controller design process. Consequently, modelling and control for the UGI gasification process can be realized. Furthermore, based on the control performance, one can make a prediction for the gasification process and realize improvement of crude gas production.

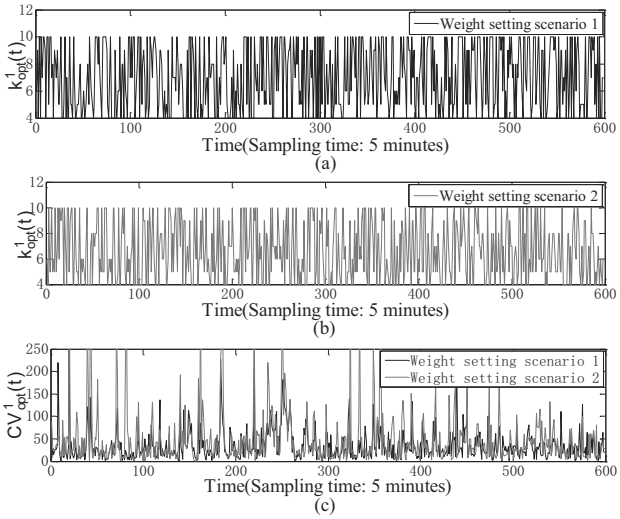

Fig. 6. U-temperature: profile of $k^1_{\text{opt}}(t)$ and $CV^1_{\text{opt}}(t)$ at sampling point t by using two weight setting scenarios of the MLL-RVM: Scenario 1 with $\bar{k}^1_{\text{opt}}(t) = 7.0903°\text{C}$ (a), Scenario 2 with $\bar{k}^1_{\text{opt}}(t) = 6.6756°\text{C}$. $CV^1_{\text{opt}}(t)$ (b), $\overline{CV}^1_{\text{opt}}(t) = 34.8127$ for Scenario 1, $\overline{CV}^1_{\text{opt}}(t) = 48.5130$ for Scenario 2 (c).

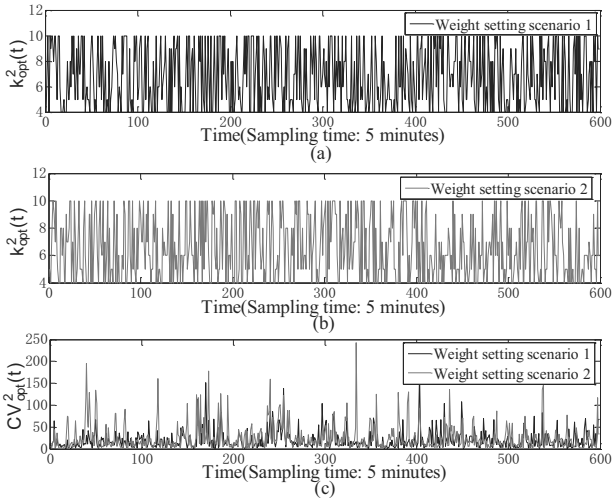

Fig. 7. D-temperature: profile of $k^2_{\text{opt}}(t)$ and $CV^2_{\text{opt}}(t)$ at sampling point t by using two weight setting scenarios of MLL-RVM: Scenario 1 with $\bar{k}^2_{\text{opt}}(t) = 6.6812°\text{C}$ (a), Scenario 2 with $\bar{k}^2_{\text{opt}}(t) = 6.5987°\text{C}$ (b), $CV^2_{\text{opt}}(t)$, $\overline{CV}^2_{\text{opt}}(t) = 20.1847$ for Scenario 1, $\overline{CV}^2_{\text{opt}}(t) = 26.3144$ for Scenario 2 (c).

Acknowledgment

This work is supported by the Beijing Municipal Natural Science Foundation under the grants 4204098 and 4212035, the National Natural Science Foundation (NNSF) of China under the grants 61903004, 61703019, 61833001 and 61803036, the North China University of Technology Scientific Research Foundation, the North China University of Technology YuYou Talent Training Program, and the Scientific Research Common Program of Beijing Municipal Commission of Education (KM201911232015).

References

Aha, D.W. (1997). *Lazy Learning*, Kluwer Academic Publisher, Dordrecht.

Aha, D.W., Kibler, D. and Albert, M.K. (1991). Instance-based learning algorithms, *Machine Learning* **6**(1): 37–66.

Altafini, C.R., Wander, P.R. and Barreto, R.M. (2003). Prediction of the working parameters of a wood waste gasifier through an equilibrium model, *Energy Conversion and Management* **44**(17): 2763–2777.

Atkeson, C.G., Moore, A.W. and Schaal, S. (1997). Locally weighted learning, *in* D.A. Kidwell (Ed.), *Lazy Learning*, Springer, Dordrecht, pp. 11–73.

Babu, B. and Sheth, P.N. (2006). Modeling and simulation of reduction zone of downdraft biomass gasifier: Effect of char reactivity factor, *Energy Conversion and Management* **47**(15–16): 2602–2611.

Bontempi, G., Birattari, M. and Bersini, H. (1999). Lazy learning for local modelling and control design, *International Journal of Control* **72**(7–8): 643–658.

Chavan, P., Sharma, T., Mall, B., Rajurkar, B., Tambe, S., Sharma, B. and Kulkarni, B. (2012). Development of data-driven models for fluidized-bed coal gasification process, *Fuel* **93**: 44–51.

Chen, S., Gunn, S.R. and Harris, C.J. (2001). The relevance vector machine technique for channel equalization application, *IEEE Transactions on Neural Networks* **12**(6): 1529–1532.

Cheng, C. and Chiu, M.-S. (2004a). A new data-based methodology for nonlinear process modeling, *Chemical Engineering Science* **59**(13): 2801–2810.

Cheng, C. and Chiu, M.-S. (2004b). A new data-based methodology for nonlinear process modeling, *Chemical Engineering Science* **59**(13): 2801–2810.

Corella, J. and Sanz, A. (2005). Modeling circulating fluidized bed biomass gasifiers. A pseudo-rigorous model for stationary state, *Fuel Processing Technology* **86**(9): 1021–1053.

Demir, B. and Erturk, S. (2007). Hyperspectral image classification using relevance vector machines, *IEEE Geoscience and Remote Sensing Letters* **4**(4): 586–590.

Fiaschi, D. and Michelini, M. (2001). A two-phase one-dimensional biomass gasification kinetics model, *Biomass and Bioenergy* **21**(2): 121–132.

Garcia, E.K., Feldman, S., Gupta, M.R. and Srivastava, S. (2010). Completely lazy learning, *IEEE Transactions on Knowledge and Data Engineering* **22**(9): 1274–1285.

Gordillo, E. and Belghit, A. (2011). A two phase model of high temperature steam-only gasification of biomass char in bubbling fluidized bed reactors using nuclear heat, *International Journal of Hydrogen Energy* **36**(1): 374–381.

Han, P., Li, D.-Z. and Wang, Z. (2008). A study on the biomass gasification process model based on least squares SVM, *Energy Conservation Technology* **1**(147): 3–7.

Hou, Z.-S. and Wang, Z. (2013). From model-based control to data-driven control: Survey, classification and perspective, *Information Sciences* **235**: 3–35.

Hou, Z.-S. and Xu, J.-X. (2009). On data-driven control theory: The state of the art and perspective, *Acta Automatica Sinica* **35**(6): 650–667.

Huang, C.-E., Li, D. and Xue, Y. (2013). Active disturbance rejection control for the ALSTOM gasifier benchmark problem, *Control Engineering Practice* **21**(4): 556–564.

Kalita, P., Clifford, M., Jiamjiroch, K., Kalita, K., Mahanta, P. and Saha, U. (2013). Characterization and analysis of thermal response of rice husk for gasification applications, *Journal of Renewable and Sustainable Energy* **5**(1): 013119.

Liu, S., Sun, J., Ji, H. and Hou, Z. (2020). Model free adaptive control for the temperature adjustment of UGI coal gasification processes in synthetic ammonia industry, *IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS), Guangxi, China*, pp. 1–6.

Mondal, P., Dang, G. and Garg, M. (2011). Syngas production through gasification and cleanup for downstream applications: Recent developments, *Fuel Processing Technology* **92**(8): 1395–1410.

Nelles, O. and Isermann, R. (1996). Basis function networks for interpolation of local linear models, *Proceedings of the 35th IEEE Conference on Decision and Control, Kobe, Japan*, pp. 470–475.

Nougués, J., Pan, Y., Velo, E. and Puigjaner, L. (2000). Identification of a pilot scale fluidised-bed coal gasification unit by using neural networks, *Applied Thermal Engineering* **20**(15–16): 1561–1575.

Papole, G., Focke, W.W. and Manyala, N. (2012). Characterization of medium-temperature Sasol–Lurgi gasifier coal tar pitch, *Fuel* **98**: 243–248.

Petersen, I. and Werther, J. (2005). Experimental investigation and modeling of gasification of sewage sludge in the circulating fluidized bed, *Chemical Engineering and Processing: Process Intensification* **44**(7): 717–736.

Puig-Arnavat, M., Hernández, J.A., Bruno, J.C. and Coronas, A. (2013). Artificial neural network models for biomass gasification in fluidized bed gasifiers, *Biomass and Bioenergy* **49**: 279–289.

Raman, P., Walawender, W.P., Fan, L. and Chang, C.-C. (1981). Mathematical model for the fluid-bed gasification of biomass materials: Application to feedlot manure, *Industrial & Engineering Chemistry Process Design and Development* **20**(4): 686–692.

Ren, Y.-Q., Xu, S.-S. and Gao, S.-W. (2004). Development status and tendency of coal gasification technology with dry coal feed in china, *Electric Power* **37**(6): 49–52.

Ruggiero, M. and Manfrida, G. (1999). An equilibrium model for biomass gasification processes, *Renewable Energy* **16**(1–4): 1106–1109.

Sadaka, S.S., Ghaly, A. and Sabbah, M. (2002). Two phase biomass, air-steam gasification model for fluidized bed reactors. Part I: Model development, *Biomass and Bioenergy* **22**(6): 439–462.

Shabbir, Z., Tay, D.H. and Ng, D.K. (2012). A hybrid optimisation model for the synthesis of sustainable gasification-based integrated biorefinery, *Chemical Engineering Research and Design* **90**(10): 1568–1581.

Simani, S., Farsoni, S. and Castaldi, P. (2018). Data-driven techniques for the fault diagnosis of a wind turbine benchmark, *International Journal of Applied Mathematics and Computer Science* **28**(2): 247–268, DOI: 10.2478/amcs-2018-0018.

Simone, M., Barontini, F., Nicolella, C. and Tognotti, L. (2013). Assessment of syngas composition variability in a pilot-scale downdraft biomass gasifier by an extended equilibrium model, *Bioresource Technology* **140**: 43–52.

Srinivas, T., Gupta, A. and Reddy, B. (2009). Thermodynamic equilibrium model and exergy analysis of a biomass gasifier, *Journal of Energy Resources Technology* **131**(3): 98–107.

Sun, B., Liu, Y., Chen, X., Zhou, Q. and Su, M. (2011). Dynamic modeling and simulation of shell gasifier in IGCC, *Fuel Processing Technology* **92**(8): 1418–1425.

Taieb, S.B., Sorjamaa, A. and Bontempi, G. (2010). Multiple-output modeling for multi-step-ahead time series forecasting, *Neurocomputing* **73**(10–12): 1950–1957.

Tipping, M.E. (2001). Sparse Bayesian learning and the relevance vector machine, *Journal of Machine Learning Research* **1**(Jun): 211–244.

Tipping, M.E. and Faul, A.C. (2003). Fast marginal likelihood maximisation for sparse Bayesian models, *Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics, Key West, USA*, pp. 1–14.

Wei, L., Yang, Y., Nishikawa, R.M., Wernick, M.N. and Edwards, A. (2005). Relevance vector machine for automatic detection of clustered microcalcifications, *IEEE Transactions on Medical Imaging* **24**(10): 1278–1285.

Wei, Q. and Liu, D. (2013). Adaptive dynamic programming for optimal tracking control of unknown nonlinear systems with application to coal gasification, *IEEE Transactions on Automation Science and Engineering* **11**(4): 1020–1036.

Wei, Q. and Liu, D. (2014). Data-driven neuro-optimal temperature control of water–gas shift reaction using stable iterative adaptive dynamic programming, *IEEE Transactions on Industrial Electronics* **61**(11): 6399–6408.

Yu, G.W., Wang, Y.M. and Xu, Y.Y. (2013). Modeling analysis of Shell, Texaco gasification technology's effects on water gas shift for Fischer–Tropsch process, *Advanced Materials Research* **608–609**: 1446–1453.

Zanoli, S., Astolfi, G. and Barboni, L. (2012). Application of a new dataset selection procedure for the prediction of the syngas composition of a gasification plant, *IFAC Proceedings Volumes* **45**(15): 868–873.

Zhao, D., Liu, J., Wu, R., Cheng, D. and Tang, X. (2019). An active exploration method for data efficient reinforcement learning, *International Journal of Applied Mathematics and Computer Science* **29**(2): 351–362, DOI: 10.2478/amcs-2019-0026.

Shida Liu received his BSc degree from Inner Mongolia University, Hohhot, China, in 2011, and his PhD degree from Beijing Jiaotong University, China, in 2017. He was a post-doctoral researcher with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, from 2017 to 2019. Now, he is a lecturer in the School of Electrical and Control Engineering, North China University of Technology, Beijing. His current research interests include intelligent traffic control, data-driven control, and artificial intelligence.

Honghai Ji received his PhD degree at the Advanced Control Systems Laboratory of Beijing Jiaotong University in 2017. He is currently a lecturer at the School of Automation, North China University of Technology University. His research interests are in the field of adaptive control, iterative learning control, Kalman consensus filtering, urban road traffic control and automatic train operation control.

Zhongsheng Hou (IEEE Fellow) received his BSc and MSc degrees in applied mathematics from the Jilin University of Technology, Changchun, China, in 1983 and 1988, respectively, and his PhD degree in control theory and applications from Northeastern University, Shenyang, China, in 1994. From 1997 to 2018, he was with Beijing Jiaotong University, China. He is currently the chair professor with the School of Automation, Qingdao University, China. His research interests are in data-driven control, model-free adaptive control, learning control, and intelligent transportation systems.

Jiashuo Zuo, from Hebei Province, has graduated from Yanshan University with an MSc degree. At present, his research direction is automation technology and application, unmanned control and intelligent control.

Lingling Fan received her PhD degree from Beijing Jiaotong University, China, in 2017. She is currently an associate professor at the School of Automation, Beijing Information Science and Technology University. Her research interests cover multi-agent systems, pursuit-evasion and data-driven control.

Received: 24 April 2020
Revised: 26 October 2020
Re-revised: 1 February 2021
Accepted: 9 February 2021