Applied Energy 197 (2017) 405–415

Contents lists available at ScienceDirect

Applied Energy

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

# A Monte Carlo and PSO based virtual sample generation method for enhancing the energy prediction and energy optimization on small data problem: An empirical study of petrochemical industries

Hong-Fei Gong, Zhong-Sheng Chen, Qun-Xiong Zhu *, Yan-Lin He *

*College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China*
*Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing 100029, China*

## HIGHLIGHTS

- A novel Monte Carlo and PSO based virtual sample generation method is proposed.
- Effective virtual samples are generated for solving the small data problem.
- Petrochemical industry empirical studies are carried out for performance validation.
- Simulation results show the proposed method can improve energy prediction accuracy.
- Guidance can be given to the production departments for improving energy efficiency.

## ARTICLE INFO

*Article history:*
Received 4 January 2017
Received in revised form 31 March 2017
Accepted 1 April 2017
Available online 9 April 2017

*Keywords:*
Energy optimization
Energy prediction
Small data
Virtual sample generation
Petrochemical industries

## ABSTRACT

Due to the imbalanced and uncompleted characteristics of complex petrochemical small datasets, it is a challenge to build an accurate prediction and optimization model of energy consumption of petrochemical systems. Therefore, this paper proposes a novel virtual sample generation (VSG) approach based on the Monte Carlo (MC) and Particle Swarm Optimization (PSO) algorithms to improve the accuracy of the energy efficiency analysis on small data set problems. The proposed approach utilizes the MC and PSO algorithms to generate appropriate virtual samples based on the underlying information extracted from the small datasets. An accurate prediction model is presented using the extreme machine learning (ELM) in view of the synthetic data. The performance of the proposed model is validated via an application using a purified Terephthalic acid (PTA) solvent system and an ethylene production system. The experiment results demonstrate that the accuracy of the prediction model can be improved, and guidance for the production department to improve the energy efficiency, energy savings and emission reduction is provided under the small data circumstance.

## 1. Introduction

Energy is a crucial factor in the effective development and sustainable prosperity of a society, economy, and environment for both developed and developing countries. Energy consumption has been rising constantly throughout the world over the past few decades and has caused an energy crisis since 1970s. Due to the booming economy in China, energy demands have skyrocketed. The increase energy demand can be an enormous threat to the finite energy supplement and available resources such as fossil fuels. The challenge of energy-savings and emission-reduction remains formidable. Hence, the improvement of energy efficiency is of great significance for the high energy-intense petrochemical industries. An accurate and robust forecasting model of energy consumption and production is indispensable for decreasing the probability of errors in energy project early stage programing.

Ethylene is one of the world's largest chemical products, and ethylene industry is the core of the petrochemical industry. Ethylene products account for more than 75% of petrochemical products and play a vital role in the national economy. Ethylene production is regarded as one of the major indexes to assess the development of a country's petrochemical industry. The energy consumption of crude oil, water, electricity, fuel and steam in

* Corresponding authors at: College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China.

*E-mail addresses:* zhuqx@mail.buct.edu.cn (Q.-X. Zhu), heyl@mail.buct.edu.cn (Y.-L. He).

http://dx.doi.org/10.1016/j.apenergy.2017.04.007
0306-2619/© 2017 Elsevier Ltd. All rights reserved.

ethylene production is considered a comprehensive indicator of a factory's technical capability. Reports show that the ethylene production from China Petrochemical Corporation was 10,420 kt/a, and the average fuel and power consumption (converted to the standard oil) was 571.39 kg/t in the year of 2014 [1]. China National Petroleum Corporation's production capacity and the average fuel and power consumption of the ethylene was 4976kt/a and 616.7 kg/t in 2014, respectively [2]. The domestic energy efficiency is remarkably lower than the advanced level, which creates huge opportunity for reduction of the energy consumption in the ethylene industries. Additionally, PTA is one of the most important bulk organic raw materials, widely used in chemical fiber, light industry, electronics, construction and other aspects of the national economy [3]. More than 90% of the world's PTA is utilized for the production of polyethylene terephthalate (PET). The downstream extension of PTA is mainly the production of polyester fiber. In the domestic market, 75% of the PTA accounts for the production of polyester fiber, 20% for bottle grade polyester and the remaining 5% for film grade polyester [4]. The accurate prediction and optimization of PTA process (PTAP) plays an essential role in energy-savings and emission reduction of the petrochemical industry.

Numerous academic research studies concerning the techniques for accurate predictions of energy consumption and production have been carried out during the last several decades. The petrochemical production process always includes numerous complex physical and chemical reactions. The physical principle (or first principle) requires a high level of knowledge about the entire process, which brings about difficulty in modeling. Data-driven modeling methods mainly focus on process data rather than the complicated process mechanism. A wide range of artificial intelligence (AI) methods such as neural networks and so on have been widely utilized to build prediction models. However, these approaches have specific requirements for data characteristics and sample size. The performance of the data-driven models can be easily influenced by differing sample sizes and the quantities of input and output variables. Sufficient sample size and uniform sample distribution are two key factors in determining the accuracy and robustness of the forecasting model [5]. Moreover, adequate training samples provide an important guarantee for improving the generalization ability of data-driven models. Considering the high acquisition cost or incapability in obtaining competent samples, insufficiently small datasets may have a bad influence on the prediction accuracy and robustness, as they are usually imbalanced and thus deteriorate the gap interval between training data. Therefore, the population characteristics of small datasets cannot be completely reflected. The number of observations of the small samples modeling is less than 30 in general [5], which occurs in various practical applications, such as energy detector (ED) of spectrum sensing and radar application, image recognition, flexible manufacturing system (FMS) scheduling [6–8], and so on.

There are many academic techniques to address the small dataset problem such as grey modeling, feature extraction, virtual sample generation (VSG), and so on. VSG is an effective technique that can improve the accuracy of prediction models established using small samples. It was originally introduced by Poggio and Vetter [7], who generated virtual examples by incorporating prior knowledge to improve the recognition performance. Following their study, Cho et al. [9] used a population of networks in order to suggest a scheme to generate virtual samples. A novel technique of extending attributes was proposed by Li et al. [8] with the mega-trend-diffusion (MTD) function, in which the experimental results demonstrated that good classification performance of small datasets was achieved. To avoid the disadvantages of a single distribution, Zhu et al. [5] designed a novel method of multi-distribution MTD in order to generate virtual samples based on the uniform and triangular distribution. In terms of energy efficiency analysis, Song et al. utilized a Super-SBM model to measure and calculate the energy efficiency of BRICS and modify the values based on DEA derived from small sample data using bootstrap [10]. Garshasbi et al. employed a hybrid Genetic Algorithm (GA) and Monte Carlo (MC) simulation approach to predict the energy consumption and renewable energy generation in a cluster of Net Zero Energy Buildings (NZEBs) over any time period [11]. A small-sample hybrid model for forecasting energy-related $CO_2$ emissions in developing countries was developed by Meng et al. [12] to obtain more precise prediction results.

In order to better analyze and predict the energy efficiency of petrochemical industries, a prediction method based on the MC and PSO algorithms and ELM model is proposed. The intent is to improve the prediction modeling accuracy of energy consumption against small size samples. The proposed method is able to inspect the production level of the petrochemical industry and enhance the energy efficiency. Standard ELM algorithm is utilized to construct the prediction model for small size samples and casts insights on how to find a number of feasible virtual samples. This paper explores the underlying information between the data gaps and thus improves the accuracy of prediction for an observed small dataset via the MC and PSO algorithms.

The remaining parts of this paper are organized as follows. Section 2 briefly reviews and introduces the ELM model, the MC and PSO algorithms. In Section 3, the proposed method in detail is described. Two real world cases concerning the energy prediction of petrochemical industries are investigated in Section 4. Finally, Section 5 draws the conclusions.

## 2. Preliminaries

In this section, the ELM algorithm is briefly reviewed and introduced followed by the Monte Carlo (MC) and particle swarm optimization (PSO) algorithms.

### 2.1. Extreme learning machine

As effective instruments for function approximation, artificial neural networks (ANNs) are able to fit any complicated nonlinear forecasting model without the physical mechanisms of systems. ANNs possess the strengths of easy realization and enjoys a wide range of practical applications to building prediction models for chemical industries [13–15]. However, the weaknesses of the gradient descent based algorithm like BPNN lie in its slow speed of convergence and its inclination to fall into the local optimum. In order to avoid the above-mentioned flaws, the extreme learning machine (ELM) algorithm was initially suggested. ELM algorithm allows for the extreme fast learning speed for single-hidden layer feedforward neural networks (SLFNs). Different from some other feed-forward neural networks, the ELM algorithm randomly selects the input weights and analytically determines the output weights of SLFNs via solving the generalized Moore-Penrose inverse of the hidden layer outputs [16]. Afterwards, a functional link least square (FLLS) structure of extreme learning machine was explored to predict key process variables and improve the prediction performance [17]. He et al. [18] developed an improved functional link neural network integrating with partial least square to deal with the nonlinear data from industrial processes. To tackle the challenges of accurately predicting the difficult-to-measure variables, soft sensor based on a novel robust bagging nonlinear model integrating improved extreme learning machine with partial least square (RB-PLSIELM) was developed [19].

The standard ELM algorithm consists of an input layer, a hidden layer and an output layer. Unlike the gradient descent-based single-hidden layer feedforward neural networks (SLFNs), ELM does not carry out the weight adjustments and optimizations based on the trial-and-error process step by step. In order to reach the minimum training error and obtain the best generalization performance, the input weights between the input layer and the hidden layer can be assigned randomly, and the output weights between the hidden layer and the output layer are analytically determined using the minimum norm least-squares solution, which makes ELM simple to understand and fast to implement. The learning strategy of ELM is totally different from that of the error back propagation algorithm. The difficulty in determining some initial parameters of error back propagation algorithm can be avoided in the learning algorithm of ELM. The structure of ELM is shown in Fig. 1.

Consider a training dataset with $M$ samples $S=(x_i,y_i)i=1,2,\ldots,M$, in which $\boldsymbol{x_i}=[x_{i1},x_{i2},\ldots x_{il}]^{\mathrm{T}}\in \mathrm{R}^l$ is the input variable vector, $\boldsymbol{y}_i=[y_{i1},y_{i2},\ldots,y_{im}]^{\mathrm{T}}\in \mathrm{R}^m$ is the output vector. The output of ELM with $N_h$ hidden nodes is given by

$$y_i=\sum_{j=1}^{N_h}\beta_j f(\omega_j\cdot x_i+b_j)i=1,2,...,M \tag{1}$$

where $\boldsymbol{\omega}_j=[\omega_{j1},\omega_{j2},\ldots,\omega_{jl}]^T$ is the input weight vector that connects the $j$th hidden layer node the inputs nodes, $\omega_j\cdot x_i$ is the inner product, $f$ represents the activation function, and $b_j$ denotes the bias of each hidden layer node. Eq. (1) can be described in the matrix format as follows:

$$\boldsymbol{Y}=\boldsymbol{H}\beta=\begin{bmatrix} f(\omega_1\cdot x_1+b_1) & \cdots & f(\omega_{N_h}\cdot x_1+b_{N_h}) \\ \cdots & f(\omega_j\cdot x_i+b_j) & \cdots \\ f(\omega_1\cdot x_M+b_1) & \cdots & f(\omega_{N_h}\cdot x_M+b_{N_h}) \end{bmatrix}\times\begin{bmatrix}\beta_1\\ \cdots \\ \beta_{N_h}\end{bmatrix} \tag{2}$$

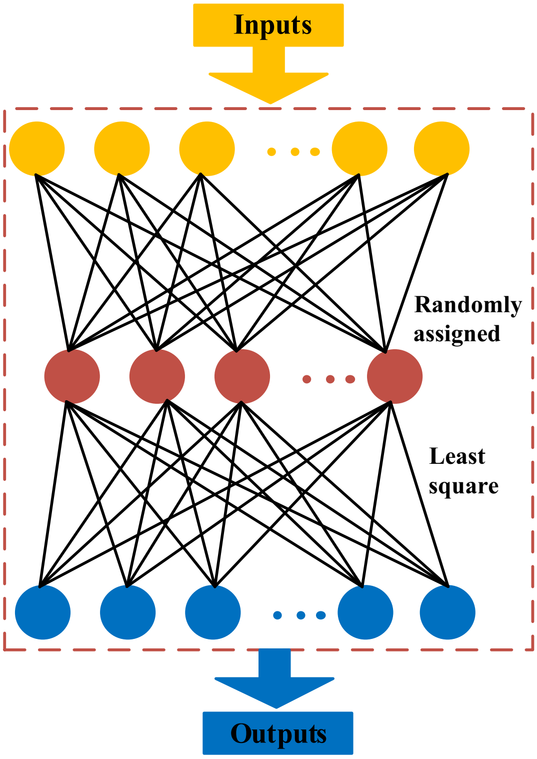

**Fig. 1.** The structure of traditional ELM.

where $\beta=[\beta_1,\beta_2,\ldots,\beta_{N_h}]^{\mathrm{T}}$ represents the vector that contacts the hidden layer nodes and the output layer node, and $H$ is the hidden layer output matrix. In this paper, the sigmoid function is defined as

$$f(x)=\frac{1}{1+e^{-x}} \tag{3}$$

Given that the input weights $\omega$ and the biases of the hidden layer neuron $b$ are already fixed, the weights $\beta$ can be analytically determined by the optimal systematic least-square solution $\hat{\beta}$, shown as follows:

$$\hat{\beta}=H^{+}Y \tag{4}$$

where $H^{+}$ is the Moore-Penrose generalized inverse of $\boldsymbol{H}$, which can be calculated by the full rank decomposition.

The whole procedure of the ELM training algorithm can be summed up as follows:

(1) Randomly generate the input weights $\boldsymbol{\omega}_j$ and the biases of the hidden layer nodes $\boldsymbol{b}_j$;
(2) Compute the hidden layer output matrix $\mathbf{H}$;
(3) Calculate the output weights $\boldsymbol{\beta}$ via the optimal least-square solution $\hat{\beta}$.

### 2.2. Monte Carlo approach

The Monte Carlo method's sampling technique utilizes the ability simulation and experimental data to provide more reliable latent distributional samples. This study intends to demonstrate the forecasting process and the application of the numerical non-parametric Monte Carlo method to obtaining samples when limited data are available.

Sufficient sample size plays a vital role in non-linear modeling. The stability and accuracy of the forecasting model cannot be assured if sufficient samples are not available. A random number generation (RNG) is required to produce samples in accordance with the underlying data distribution of the targeted outcomes, which can be used for calculating the desired amounts of data and attaining systematic data augmentation when the data is limited.

As the Monte Carlo method enjoys the practicality and effectiveness in dealing with data augmentation, it has been widely applied to various fields. Wilson et al. [20] applied the Monte Carlo Bootstrap Method in statistical analysis of radioecological data and obtained appropriate predictors for the population parameters. Pereira et al. [21] presented a methodology where the Monte Carlo Method (MCM) was used to estimate the behavior of economic parameters which may help decision, considering the risk in project sustainability. Fargesab et al. [22] proposed an alternative approach using the Monte Carlo Methods (MCM) to evaluate yearly collected energy in the context of optimizing concentrated solar power plants (CSP).

A widely used sampling technique in Monte Carlo method is Latin Hypercube Sampling (LHS), whose superiority over random sampling has been illustrated effectively already. The first theoretical analysis of the performance of Latin hypercube and basic random sampling found that LHS is more efficient than random sampling when the desired output is monotonic in all inputs [23]. The implementation of Latin hypercube sampling intends to achieve a better space-filling character given the information gaps caused by limited raw samples.

### 2.3. Particle swarm optimization algorithm

As one of the population-based evolutionary algorithms to solve the optimization problems, particle swarm optimization (PSO) was originally suggested by Kennedy [24] and was proven to achieve better performance in solving the highly complex optimization problems compared with conventional optimization approaches. The core algorithm simulates the social behavior of swarms like fish schooling and bird flocking to achieve precise objectives in multi-dimensional space. The advantages of PSO lie in fewer parameters to be adjusted, easy understanding and implementation, and highly computational efficiency. Being low cost, fast, and robust, the PSO algorithm enjoys a wide range of applications. PSO can rapidly converge towards an optimum and is simple and flexible to compute and implement. Piltan et al. [25] estimated the electricity demand function in the Iranian metal industry, where the unknown parameters of the energy consumption forecasting models were estimated using PSO and RCGA. Hu et al. [26] presented a data-driven method for estimating the capacity of Li-ion battery based on the charge voltage and current curves, and optimized the feature weights using PSO to minimize the CV error. Chang [27] assessed wind energy in Taiwan according to Weibull function using PSO to find the Weibull parameters.

In PSO, a population consists of individuals (called particles) that are updated by numerous iterations. A particle represents a potential solution and moves in the parameter space shaped by constrains in accordance with a specific optimization problem. Each particle is characterized by a position vector and a velocity vector, and makes use of the best previous experience (pBest) on its own and the best experience (gBest) of all the other members in search of a direction as well as to change its positions. pBest and gBest are two essential factors. During the optimization process, the initial population is randomly generated in the search space. The positions and velocities of the particles of the swarm are generated randomly as a start point. Then the objective function value of each particle is evaluated each iteration to find pBest and gBest. If the fitness value is better than the best fitness value (pBest) before, then the current value will be set as the new pBest. After the particle with the best fitness value is chosen as gBest, the particle position and velocity will be calculated and updated. The procedure is repeated until the stopping criterion is met.

The particle velocity and position update equations are described as follows:

$$v_{id} = \omega_{id} v_{id} + c_1 * r_1 * (pbest_{id} - x_{id}) + c_2 * r_2 * (gbest_{id} - x_{id}) \quad (5)$$

$$x_{id} = x_{id} + v_{id} \quad (6)$$

where $v_{id}$ is the velocity of particle I; $x_{id}$ denotes the position of the *i*-th particle. $c_1$ and $c_2$ are two learning factors, usually set as $c_1 = c_2 = 2$ [28]. $r_1$ and $r_2$ are random numbers in the range of [0,1], and $\omega_{id}$ is the inertia weight.

## 3. The proposed method

The purpose of this study is to improve the prediction accuracy of the forecasting model by adding virtual samples to a specific small size dataset. The following section introduces the proposed method in detail, starting from selecting data to construct the extreme learning machine (ELM) model, producing the initial values for virtual samples, and then generating virtual samples to build the modified forecasting model.

### 3.1. Constructing the extreme learning machine model

In the small dataset learning problem, the number of samples is usually considered as no more than 30. Therefore, ahead of the first step of the proposed approach, no more than 30 samples are randomly selected from the corresponding datasets to constitute the training dataset and the remaining ones become the testing dataset. Due to the stability and robustness, ELM algorithm is chosen as the modeling tool. The model is constructed in four steps, including the overall ELM modeling, production of the initial values of artificial samples through Monte Carlo method, the PSO virtual sample generation, and at last the modification of ELM model by adding virtual training samples.

The mean absolute percentage error (MAPE) is suggested as the tolerant error. The computing equation is shown as follows:

$$MAPE = \frac{1}{n_s} \sum_{n=1}^{n_s} \left| \frac{y - \hat{y}}{y} \right| \times 100\% \quad (7)$$

Determination of the attribute ranges $[X_l, X_u]$ is set using the actual value of the training dataset. The acceptable value of each attribute is defined using $X_l$ as the lower bound and $X_u$ as the upper bound. The ELM model and attribute ranges are combined as the subjective constraints.

### 3.2. Use the Monte Carlo method to produce initial values

The Monte Carlo (MC) method basically intends to implement a sufficiently large sequence of simulations p, each with different values for the n attributes to quantify the underlying distributions of the targeted outcomes. With the Latin Hypercube Sampling technique, MC method achieves selecting p samples out of each of the n attribute distribution [23]. In general, the p n-arrays are first drawn from uniform distributions in the n-dimensional hypercube $[0, 1]^n$ and then transformed to a series of p samples in accordance with the distributions of the n interdependent attributes. This guarantees that the sampling points are evenly distributed and that no two design points coincide when projected onto a lower number of dimensions.

Consider the $l$ attributes of the training dataset as $x_i = [x_{i1}, x_{i2}, \ldots x_{il}]^T \in R^l, i = 1, 2, \ldots, M$. Subsequently, the underlying distribution can be obtained by the sample frequency distribution function *F*. With the MC method, the sampling distribution can be attained and the new samples $x_1* = [x_{11}, x_{12}, \ldots x_{1l}]^T \in R^l$, $x_2* = [x_{21}, x_{22}, \ldots x_{2l}]^T \in R^l$; …; $x_n* = [x_{n1}, x_{n2}, \ldots x_{nl}]^T \in R^l$ that follow the same distribution; the new data are available to fill in the missing value ranges, where *n* is the number of sampling times. In order to examine the sampling distribution and outliers' presence, a widely used testing method called as chi-square test is utilized to check the convergence. Through the computed value, it is possible to increase the scattering of the original data with a number of new samples because this value is calculated between the maximum and minimum values. These new samples are able to fill the information gaps of the training dataset. The procedure is realized with the @risk6.0 software.

### 3.3. Virtual sample generation with PSO

The Monte Carlo method lacks the ability of completely reflecting the overall characteristics of the population. In other words, the samples generated by the MC method do not necessarily fit for small sample characteristics. An unknown number of unreasonable samples beyond the population can negatively influence on the prediction ability of the model. In addition, another drawback of the previous methods is that the attributes are usually coped with independency. This study takes the integrated effects of attribute into consideration. Through the PSO optimization process, a new combination of inputs can be found to minimize the relative percentage error between the predicted value and the observed value and guarantee better virtual samples generation. Hence, the MC method is used to address this data scarcity together with PSO and the fitness functions are formulated using MAPE. On the basis of the abovementioned modeling method and newly generated samples, the procedure of the proposed VSG is introduced in the following subsection.

The relationship among the original small samples, the virtual samples and population is shown in Fig. 2, from which it can be seen that the small sample set (green circle) consisting of only few raw samples (green triangles) is a subset of the population (blue circle). Plenty of virtual samples fill the information gap between the population and the original dataset (green circle). It is noticeable that the gaps among raw samples in the small dataset are reduced by the virtual samples. Hence, the virtual sample generation (VSG) procedure gives devotion and dedication to the improvement of accuracy.

The essence of the proposed MC-PSO method is that it attempts to find a novel association of attributes to minimize the smallest relative percent error between the predicted values and the observed values through the optimization process and ensure the performance of the whole procedure. Compared with the previous optimization technique, the new samples obtained by MC simulations are set as the starting points in the searching space. The whole optimization process will not be terminated until one of the following criterions is satisfied: the maximum number of iterations M or the desired fitness value is met. A virtual sample is generated around the original samples in the search area shaped by the upper and lower bounds of attributes after each iteration. The above process is repeated until the desired number of virtual samples is reached. The optimization problem is described as follows:

Minimize $f(x)$

$$\text{Subject to}: \quad x_{li} \leqslant x_i \leqslant x_{ui}; \quad i = 1, 2, \ldots, \mathrm{n}_1 \tag{8}$$

$$G_k(x) \leqslant 0; \quad k = 1, 2, \ldots, \mathrm{n}_2 \tag{9}$$

$$-0.1 \leqslant \frac{y - \hat{y}}{y} \leqslant 0.1 \tag{10}$$

The synthetic dataset is made up of the original small dataset and the newly generated virtual dataset. Then the final ELM prediction model is learned by training on the synthetic dataset and validated on the original testing dataset.

### 3.4. The operation steps of the proposed method

According to the abovementioned descriptions, the proposed method consists of the following four steps:

(1) Constructing a prediction model via ELM which is learnt using the training dataset and tested using the testing dataset. The optimal number of hidden nodes $N_{\text{best}}$ is determined by the trial-and-error method.

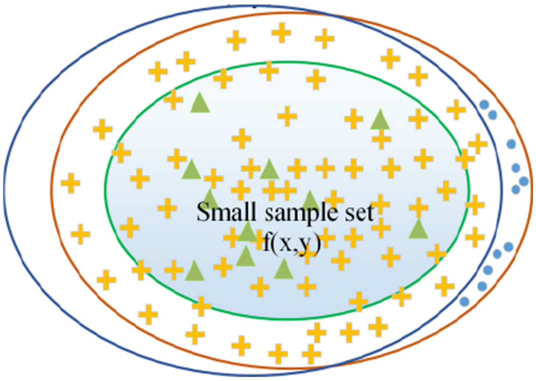

**Fig. 2.** Relationship among population, small-sample-sets and virtual datasets.

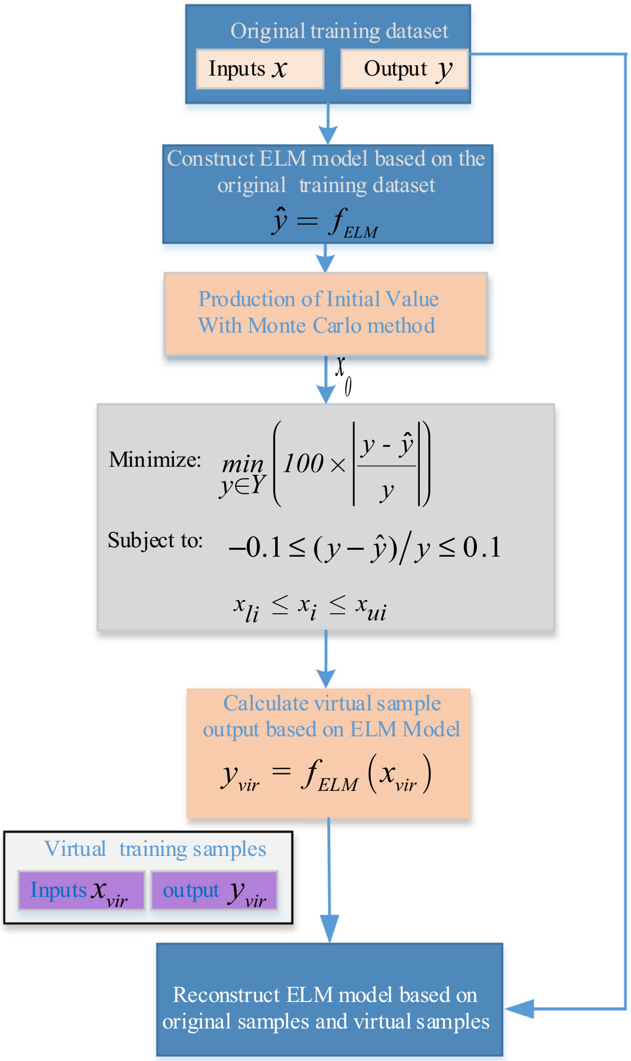
Original training dataset
Inputs $x$
Output $y$
Construct ELM model based on the original training dataset
$\hat{y} = f_{ELM}$
Production of Initial Value With Monte Carlo method
$x_0$
Minimize: $\min_{y \in Y} \left(100 \times \left|\frac{y - \hat{y}}{y}\right|\right)$
Subject to: $-0.1 \le (y - \hat{y})/y \le 0.1$
$x_{li} \le x_i \le x_{ui}$
Calculate virtual sample output based on ELM Model
$y_{vir} = f_{ELM}(x_{vir})$
Virtual training samples
Inputs $x_{vir}$
output $y_{vir}$
Reconstruct ELM model based on original samples and virtual samples

**Fig. 3.** Flowchart of the proposed method.

(2) Through the Monte Carlo method, the initial values of each input are obtained by numerous simulations and set as starting points for the search space in the optimization process described in Eqs. (8–10).

(3) Generate virtual samples according to the aforementioned procedures of virtual sample generation, where the Monte Carlo method based initial values is set as the starting point for PSO.

(4) Form a combined sample set $S_{\text{new}}$ by integrating the original training samples $S_{train}$ with the generated virtual samples $S_{\text{vir}}$. A modified forecasting ELM model can be obtained using the combined sample set $S_{\text{new}}$ and validated using the same testing set $S_{\text{test}}$.

Based on the above analyses, the flowchart of the proposed method is shown in Fig. 3.

## 4. Case study: production capacity prediction and energy saving analysis of the petrochemical industry

This section analyzes the acetic acid consumption of the purified Terephthalic acid process (PTAP) and the ethylene production process using the proposed method followed by the prediction and interpretation of the productivity and energy efficiency status of ethylene production to verify the effectiveness and application meaning of this approach.

### 4.1. Predictive modeling and energy saving of PTA solvent system

The purified Terephthalic acid (PTA) is one of the most significant bulk stocks in the petrochemical industry and enjoys a wide range of applications such as producing polyester fiber and covering beverage. The PTA process consists of three major parts: the reflux tank, the reboiler and the solvent dehydration tower. A vital index to evaluate the production process and the technique level of PTA is the acetic acid consumption, which plays a vital role in the optimization of PTA production process and economic efficiency improvement [18]. It is of great significance to establish a stable and accurate model for predicting the consumption of acetic acid. A schematic flowchart of the PTA process is shown in Fig. 4.

According to the preceding studies that introduced the operational characteristic of the solvent dehydration tower [19], 17 attributes that effectively influence the acetic acid consumption in PTAP are selected as the input variables. The 17 selected input variables are described in Table 1. The content of the acetic acid is the emphasis of this analysis; yet due to the absence of online detector at the top of the tower, the solvent dehydrating tower conductivity is considered as the output variable.

This work makes random selections of 50 samples out of the entire 300 PTA samples, 30 of which are stochastically selected as the training samples and the remaining becomes the testing dataset. The optimal number of hidden layer neurons is determined by using the trial-and-error method. Various experiments are carried out with various numbers of virtual samples ranging from 10 to 60 with increments of 10. For each experiment, 10 independent runs are implemented. During simulations, the root mean square error (RMSE) is recorded.

$$RMSE = \sqrt{\frac{1}{N_{te}}\sum_{i=1}^{N_{te}}(y_{te,i} - \hat{y}_{te,i})^2} \qquad (11)$$

where $N_{te}$ is the number of samples in the testing dataset; $y_{te,i}$ and $\hat{y}_{te,i}$ is the observed value and the predicted value of the $i$th element of the training output $y_{te}$, respectively. RMSE is related to the accuracy of the model. The smaller the value of RMSE is, the higher accuracy the model achieves.

**Table 1**
The 17 selected input variables in PTAP.

| Nos. | Variable description | Nos. | Variable description |
|---|---|---|---|
| 1 | Feed composition (acetic acid content) | 2 | Feed quantity |
| 3 | Water reflux | 4 | NBA main reflux |
| 5 | NBA side reflux | 6 | Steam flow |
| 7 | Produced quantity of top tower | 8 | Feed temperature |
| 9 | Reflux temperature | 10 | Temperature of top tower |
| 11 | Temperature point above the 35th tray | 12 | Temperature point between the 35th tray and 40th tray |
| 13 | Temperature point between the 44th tray and the 50th tray | 14 | Tray temperature near the up sensitive plate |
| 15 | Tray temperature near the low sensitive plate | 16 | Temperature point between the 53rd tray and the 58th tray |
| 17 | Reflux tank level | | |

Table 2 displays the RMSE values before and after adding different number virtual samples. As the results shown in Fig. 5, the ELM prediction model with virtual samples achieves better performance compared with that without virtual samples in the term of RMSE. Also, the more virtual samples are generated, the higher accuracy and better performance the prediction model obtains.

To clearly show the testing performance of the proposed method, the modeling relative errors of PTA testing samples with and without virtual samples are depicted in Fig. 6. From Fig. 6, it is clear that the modeling relative errors of PTA testing samples with virtual samples are closer to the zero line than those of PTA testing samples without virtual samples. It can be concluded that the proposed method can improve the accuracy of the prediction model.

The comparison among the results of the observed outputs, the ELM predicted outputs with and without virtual samples is shown in Fig. 7.

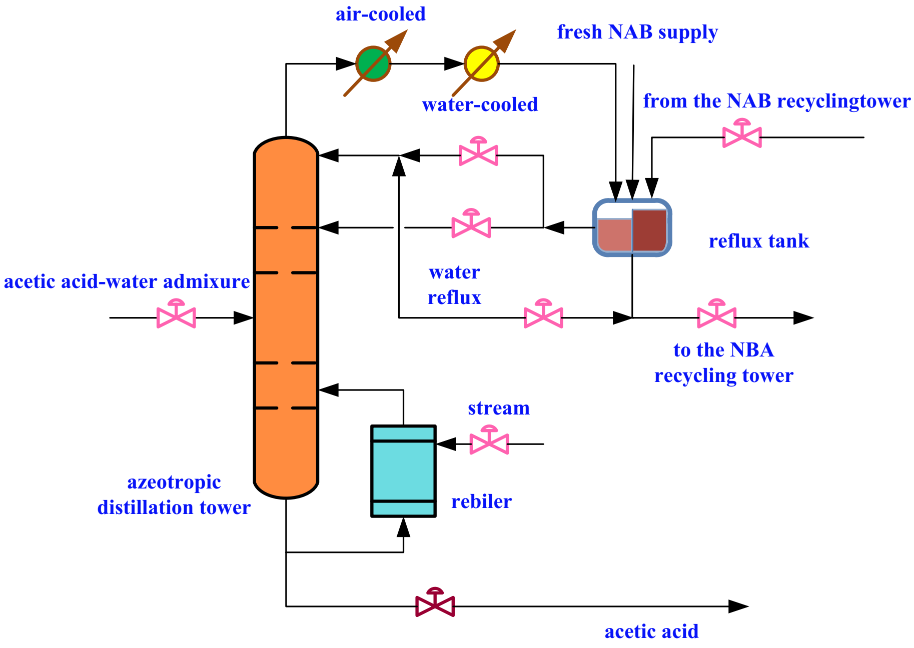

**Fig. 4.** Schematic flowchart of a PTA plant.

**Table 2**
Evaluation of the models for MC-PSO VSG.

| Number of samples | RMSE |
|---|---|
| 30 | 0.4058 |
| 30 + 10 | 0.3538 |
| 30 + 20 | 0.3467 |
| 30 + 30 | 0.3258 |
| 30 + 40 | 0.3159 |
| 30 + 50 | 0.2926 |
| 30 + 60 | 0.2895 |

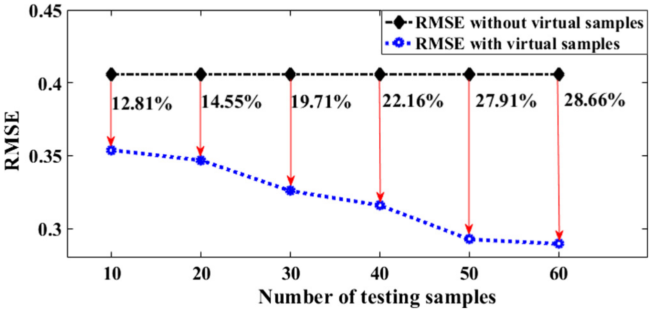

**Fig. 5.** Accuracy results with different numbers of virtual samples.

According to Fig. 7, the predicted content of the acetic acid of the 11th sample is smaller than the actual counterpart. The prediction outcome gained by the proposed method is 48.86, and the actual value is 49.41. This difference indicates that the PTAP device at this point is not operated on the condition of full load and the efficiency of energy consumption is relatively low. In contrast, the 10th sample's predicted acetic acid actual content (47.79) is larger than the actual value (48.39) and it indicates that the device at this point is operated in a more proper production status and the energy efficiency improves. The PTA production department is able to derive guidance from the samples like the 10th one as reference and tune production parameters in the coming production to increase energy efficiency for the petrochemical industry.

For instance, input variables No. 8 to No. 16 are all temperature parameters. As shown in Fig. 8, the input variables No. 12, No. 13 and No. 16 should be increased by 0.05982, 0.09758 and 0.06411 respectively to make more effective production and enhance energy optimization. And the input variables No. 8, No. 9, No. 10, No. 11, No. 14, and No. 15 should be decreased by 0.1183, 0.04108, 0.4418, 0.03658, 0.04711 and 0.03813. From the above analysis to input adjustments, the plant can reduce the consumption of acetic acid 1.62.

### 4.2. Predictive modeling and efficiency analysis for ethylene industrial production

The ethylene industry is one of the most essential systems of the petrochemical industry. Its production tends to be a major indicator of a country's industrialization level and a large slice of the petrochemical industry is taken up by its energy consumption [29]. The ethylene production process consists of two parts: cracking and separation. The cracking reactions in the tube are provided by an enormous amount of fuel and via recovering the waste heat; a large amount of steam is produced by a transfer line exchanger. The steam should be injected when the hydrocarbon is fed into the cracking furnace. As for the separation operation, the following three parts: rapid cooling, compression, and separation are executed. The main energy consumption is the power of compressor, the heat supply in separation such as by steam, and the cooling of compressor and cold box [30]. A typical flow diagram of an ethylene is shown in Fig. 9.

For the sake of building the prediction model and investigating the ethylene production efficiency, it is indispensable to confirm the inputs and outputs variables. The ethylene production efficiency is concerned with the following three aspects: (1) raw stock; (2) fuel and consumption of power; (3) products [31]. According to the energy battery limit of ethylene production, crude oil (naphtha, light diesel fuel, raffinate, hydrogenation tail oil, light hydrogenation tail oil, $C_{345}$ and others), fuel (light weight oil, heavy weight oil, fuel gas), steam (super high pressure steam, high pressure steam, medium pressure steam, low pressure steam), water (circulating water, industrial water, boiler water and other water) and electricity are selected as the input variables in ethylene production. The main output products are ethylene, propylene and $C_4$. Ethylene trade standard DB 37/751-2007 and GB/T 2589-2008 [32,33] are referred to perform the division of ethylene production plant battery limit, for various ethylene energy efficiency analysis battery limits and calculation approaches are utilized by different enterprises.

Due to the diverse physical units of input variables such as fuel, steam, water, and electricity, it is essential to convert them into GJ unanimously on the basis of the Table 3.0.2 and Table 3.0.3 in the "Petrochemical Design Energy Consumption Calculation Method" (SH/T3110-2001) [34]. As for output variables, ethylene, propylene and C4 are all measured in tons.

Production data of every month from 2010 to 2013 in 6 ethylene main production plants throughout China with 7 types of ethylene production techniques are regarded as analysis targets [29–31]. Each plant's production data from 2010 to 2012 constitutes the training dataset and the data of 2013 is set as the testing samples. The training dataset is modeled by the MC-PSO VSG integrated ELM to compare the prediction performance to that without

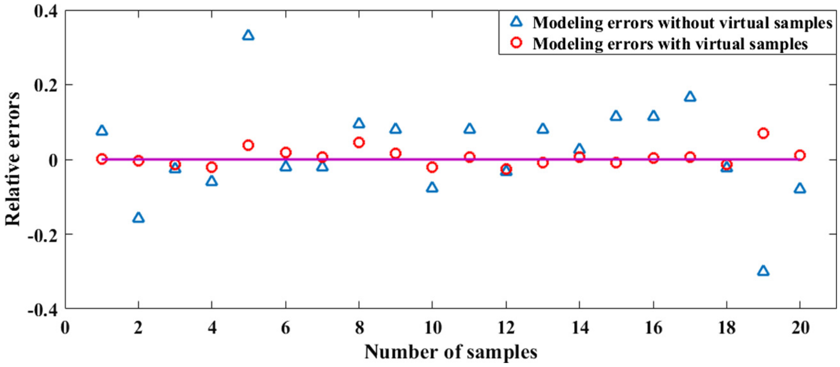

**Fig. 6.** Modeling errors of PTA testing samples with and without virtual samples.

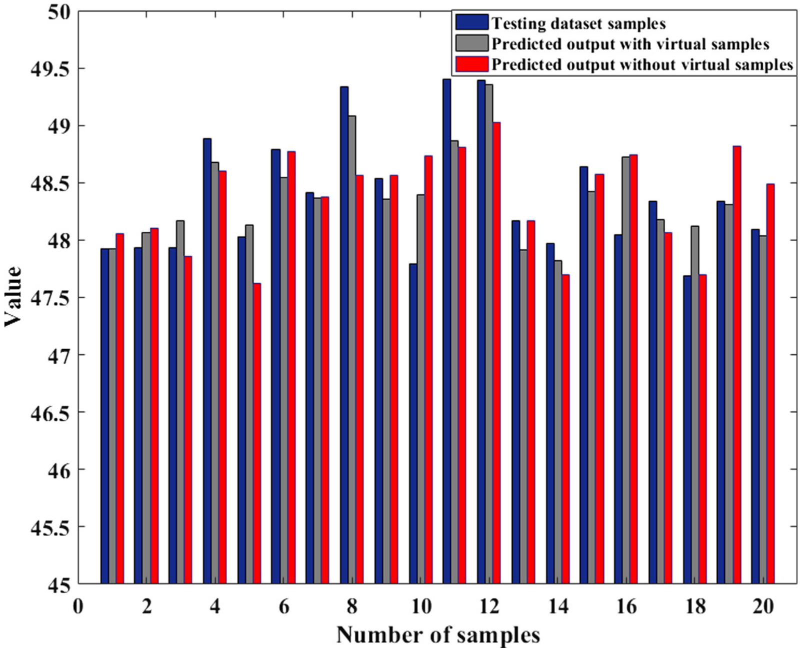

**Fig. 7.** Comparison of testing results of three types of outputs for PTA.

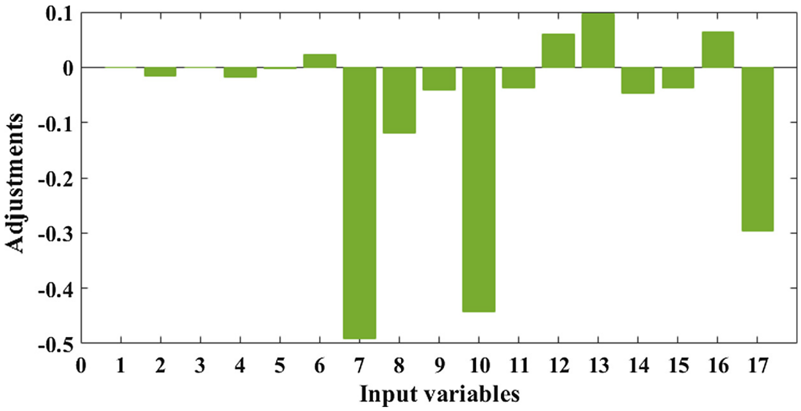

**Fig. 8.** Adjustments of PTA input variables of the 11th sample.

any virtual sample. The optimal number of hidden neurons is determined by using the trial-and-error method. Table 3 displays the RMSE values before and after adding different number virtual samples.

Fig. 10 shows the improvement of prediction accuracy compared to that without any virtual sample. According to what has been analyzed above, the prediction model based on ELM algorithm with MC-PSO VSG is able to achieve better performance of generalization with the complicated ethylene industrial production data and also increase the accuracy for ethylene industrial capacity forecasting.

In order to clearly show the testing performance of the proposed method, the modeling relative errors of the ethylene testing samples with and without virtual samples are depicted in Fig. 11. From Fig. 11, it is clear that the modeling relative errors of ethylene testing samples with virtual samples are closer to the zero line than those of ethylene testing samples without virtual samples. It can be concluded that the proposed method can improve the accuracy of the prediction model.

Fig. 12 displays the testing results of actual output, outputs with and without any virtual sample, respectively. Fig. 12 also shows that the total production of Plant 1 and Plant 2 is different from each other in different months. Whether the feed of ethylene and the production technology should be adjusted to improve the energy efficiency of ethylene production depends on the smaller RMSE. For example, Plant 1 in Tianjin City used 53,972 ton crude, 18.36GJ fuel, 3.17GJ stream, 1.94GJ water, and 1.87GJ electricity to produce ethylene, propylene and C4 in December 2013 (the 16th sample point). The actual value of total output is 32,056 tons, and the predicted value is 31,060 tons. On the basis of the aforementioned data, the production condition is good in December and is supposed to be maintained in the future. In the meantime, Plant 2 in Shandong Province used 181654.67 ton crude, 20.03 GJ fuel, 1.02 GJ stream, 2.14 GJ water, and 0.55 GJ electricity to produce ethylene, propylene and C4 (the 8th sample). The actual value of total output is 107,044 tons, and the predicted value is 102,438 tons. Therefore, the data indicates that Plant 2 has low energy efficiency in December and should adjust the input to improve the efficiency of production in the future. The adjustments of production parameters of Plant 2 are shown in Figs. 13 and 14.

Figs. 13 and 14 indicate that Plant 2 can enhance the energy consumption efficiency and realize energy conservation and emis-

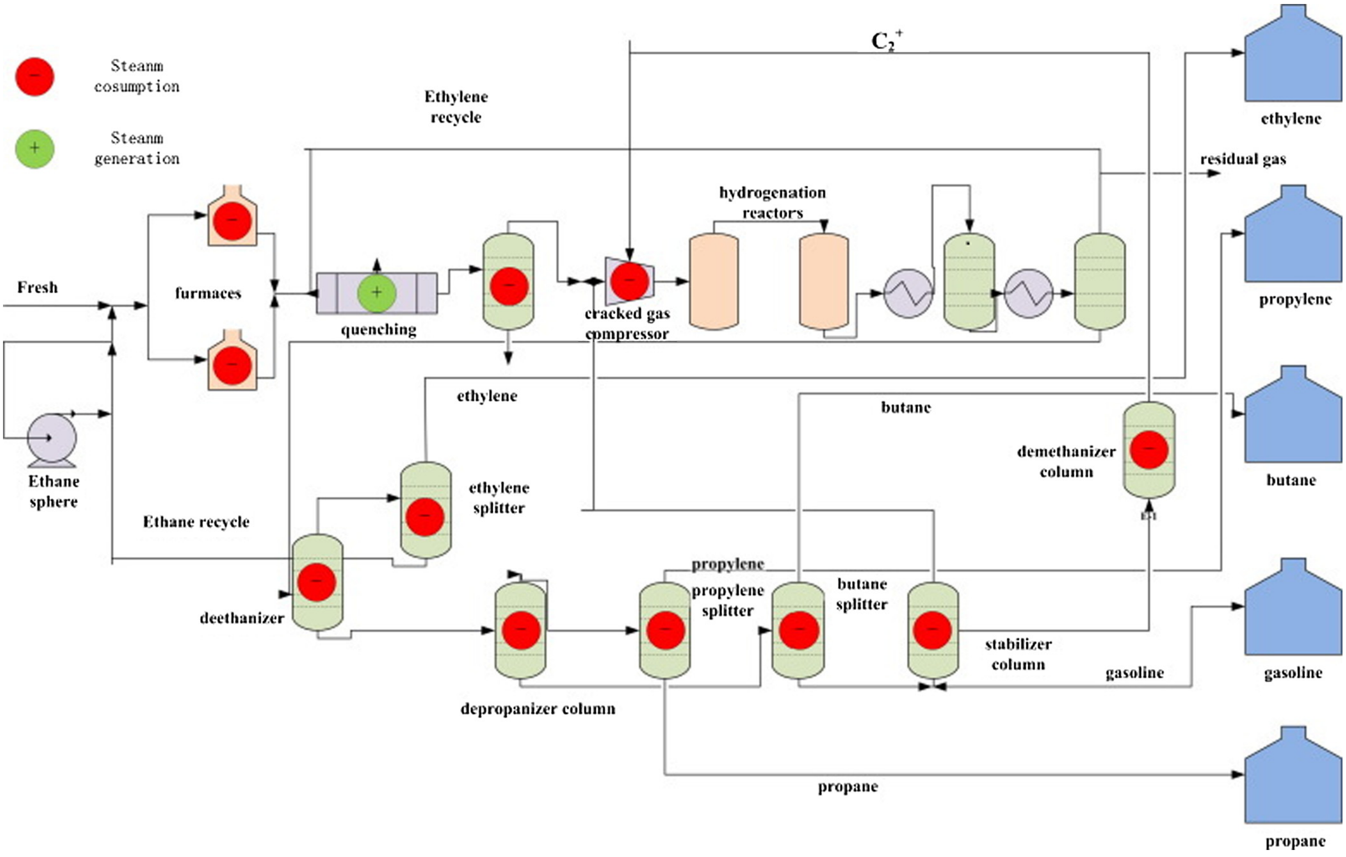

**Fig. 9.** A typical flow diagram of an ethylene.

**Table 3**
Evaluation of the models for MC-PSO VSG.

| Number of samples | RMSE |
|---|---|
| 30 | 0.03867 |
| 30 + 10 | 0.03476 |
| 30 + 20 | 0.03205 |
| 30 + 30 | 0.03051 |
| 30 + 40 | 0.02732 |
| 30 + 50 | 0.02489 |
| 30 + 60 | 0.02204 |

sion reduction by increasing the input of steam and electricity by 2.15 GJ and 1.32 GJ, respectively. At the same time, the input of crude oil, fuel, and water should be reduced by 127,682.67 tons, 1.67 GJ, and 0.2 GJ, respectively. In addition, for each ton of production, Plant 1 consumed 1.6837-ton crude and Plant 2 used up 1.6970-ton crude oil. In view of carbon emission factor for various types of fuels [35], Plant 2 will emit 0.0227-ton more $CO_2$ emission compared to Plant 1. The analysis can give useful guidance for the corresponding production departments to adjust corresponding parameters and if the production scale is 800,000 tons, then the carbon emission can be reduced by 18,154.5 tons.

### *4.3. Case study summary*

In the two abovementioned case studies, a prediction model with higher accuracy can be built using the synthetic dataset containing the original small size samples and the effective virtual samples generated using the proposed MC-PSO VSG method. According to the difference between the actual values and the

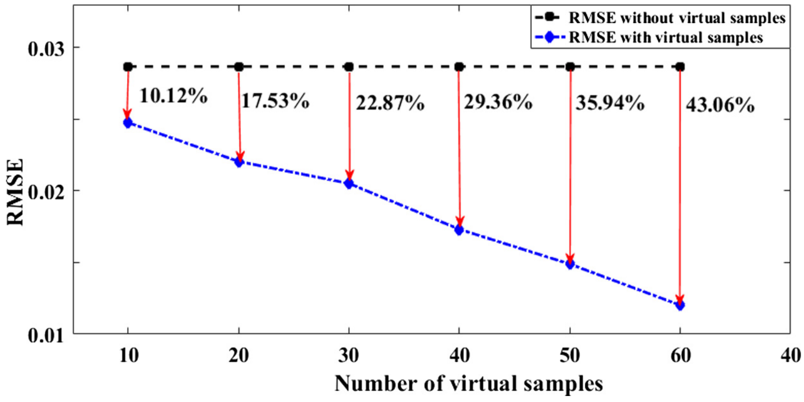

**Fig. 10.** The improvement of accuracy with various numbers of virtual samples.

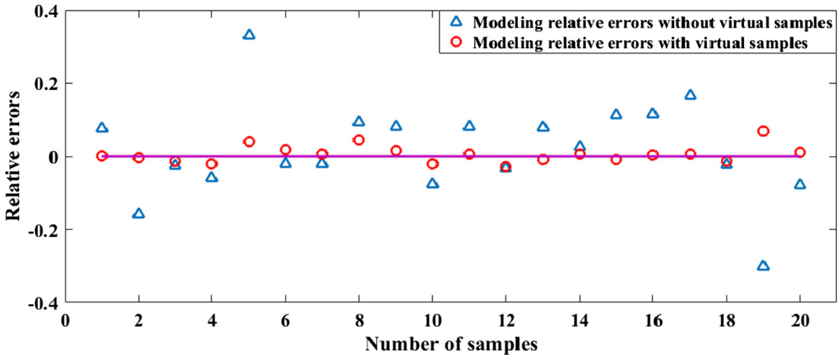

**Fig. 11.** Modeling errors of ethylene testing samples with and without virtual samples.

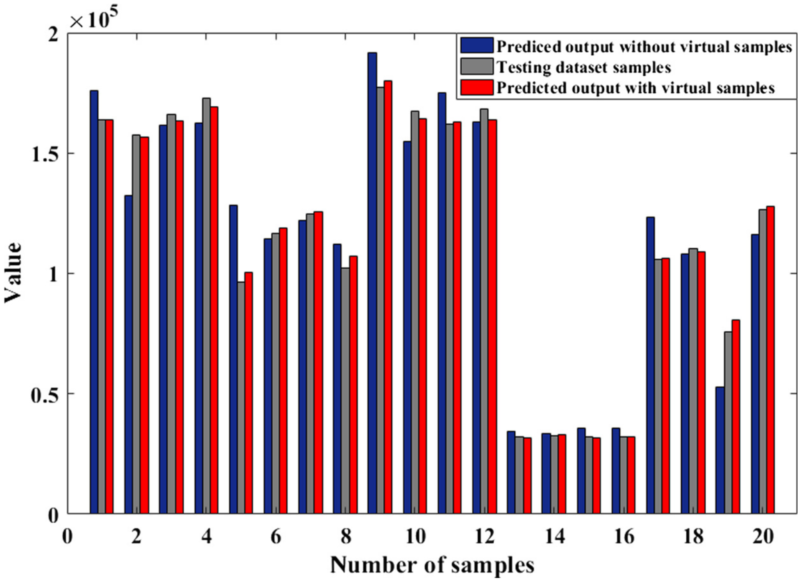

**Fig. 12.** Comparison of the actual production output and the predicted outputs.

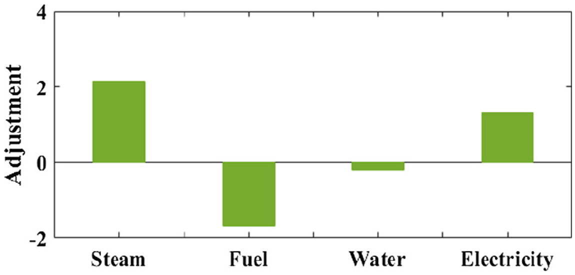

**Fig. 13.** The adjustment of production parameters (water, steams, fuels, electricity) of ethylene Plant 2.

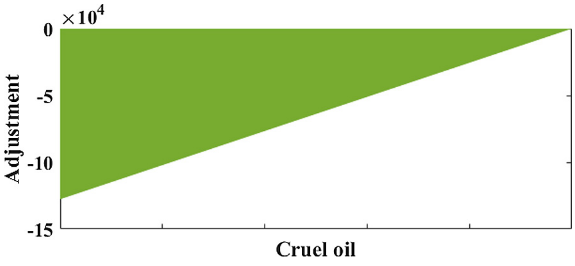

**Fig. 14.** The adjustment of production parameters (crude oil) of ethylene Plant 2.

predicted output values of the prediction model, the production condition and the energy efficiency of the production plant in different months are analyzed at each case study, and the corresponding production departments can accordingly adjust the input parameters of the production plant to improve the production condition and the energy efficiency referring to the production plant with good production condition and high energy efficiency.

## 5. Conclusions

A novel approach based on the Monte Carlo method and PSO virtual sample generation is presented in this study to improve the learning accuracy of small data set problems in the analysis of energy efficiency of the petrochemical industry. The proposed method is applied to enhancing the prediction modeling analysis of the energy efficiency of the petroleum production process. Two real-world cases are selected in this paper to verify the effectiveness of the forecasting model by comparing them with and without adding various numbers of virtual samples. The experimental results show that the prediction model with adding virtual samples performed better than that without virtual samples in terms of the forecasting ability and accuracy. The proposed method makes accurate predictions for PTA process to reduce the consumption of acetic acid. And then the proposed method could demonstrate the orientation of energy conservation for ethylene production enterprises in order to enhance production improvement. In conclusion, the proposed method could be considered an effective and robust prediction tool for energy efficiency analysis on the small data problem.

## Acknowledgements

This research is supported by the National Natural Science Foundation of China under Grant Nos. 61533003 and 61473026. Many appreciations are given to Prof. Victoria Mantzopoulos from University of Detroit Mercy for her careful proofreading and editing.

## References

[1] Ma GF, Xu YH, Guo X. Ethylene business review of china petrochemical in 2014. Ethyl Indust 2015;27(1):1–5.

[2] Zhang LJ, Hu J. Ethylene business review of china petroleum in 2014. Ethyl Indust 2015;27(1):6–10.

[3] Li YL, Sun XT. Development status and trend of China PTA industry. China Petrol Chem Indus Anal 2013;8:46–9.

[4] Market research report. Purified terephthalic acid (PTA) markets in China. Asia market information and development 2016;1–195.

[5] Zhu B, Chen ZS, Yu LA. A novel mega-trend-diffusion for small sample. Chin J Chem Eng (China) 2016;67(3):820–6.

[6] Rugini L, Banelli P, Leus G. Small sample size performance of the energy detector. IEEE Commun Lett 2013;17(9):1814–7.

[7] Poggio T, Vetter T. Recognition and structure from one 2D model view: observations on prototypes, object classes and symmetries. Massachusetts Inst of Tech 1992;1347:1–25.

[8] Li DC, Wu CS, Tsai T, Lina YS. Using mega-trend-diffusion and artificial samples in small data set learning for early flexible manufacturing system scheduling knowledge. Comput Oper Res 2007;34:966–82.

[9] Cho Z, Jang M, Chang SJ. Virtual sample generation using a population of networks. Neural Process Lett 1997;5(2):21–7.

[10] Song ML, Zhang LL, Liu W, Fisher R. Bootstrap-DEA analysis of BRICS' energy efficiency based on small sample data. Appl Energy 2013;112:1049–55.

[11] Garshasbi S, Kuinitski J, Mohammadi Y. A hybrid genetic algorithm and Monte Carlo simulation approach to predict hourly energy consumption and generation by a cluster of net zero energy buildings. Appl Energy 2016;179:626–37.

[12] Meng M, Niu DX, Shang W. A small-sample hybrid model for forecasting energy-related $CO_2$ emissions. Energy 2014;64:673–7.

[13] Zhang EL, Hou L, Shen C, Shi YL, Zhang YX. Sound quality prediction of vehicle interior noise and mathematical modeling using a back propagation neural network (BPNN) based on particle swarm optimization (PSO). Meas Sci Technol 2015;27(1):015801.

[14] Li MQ, Jin T, Chen FZ. Improving multiclass pattern recognition with a co-evolutionary RBFNN. Pattern Recogn Lett 2008;29(4):392–406.

[15] Zhu QX, Jia YW, Peng D, Xu Y. Study and application of fault prediction methods with improved reservoir neural networks. Chin J Chem Eng 2014;22 (7):812–9.

[16] Huang GB, Zhu QY, Siew CK. Extreme learning machine: a new learning scheme of feedforward neural networks. Proc Int Joint Conf Neural Netw 2004;2(2):985–90.

[17] He YL, Zhu QX. A novel robust regression model based on functional link least square (FLLS) and its application to modeling complex chemical processes. Chem Eng Sci 2016;153:117–28.

[18] He YL, Geng ZQ, Zhu QX. Hybrid robust model based on an improved functional link neural network integrating with partial least square (IFLNN-PLS) and its application to predicting key process variables. ISA Trans 2016;61:155–66.

[19] He YL, Xu Y, Geng ZQ, Zhu QX. Soft sensor development for the key variable of complex chemical processes using a novel robust bagging nonlinear model integrating improved extreme learning machine with partial least square. Chemom Intell Lab Syst 2016;151:78–88.

[20] Silva ANC, Amaral RS, Santos JA, Vieira JW Menezes RSC. Statistical analysis of discrepant radio ecological data using Monte Carlo bootstrap method. J Radioanal Nucl Ch 2015;306:571–7.

[21] Pereira EJDS, Pinho JT, Galhardo MAB, et al. Methodology of risk analysis by Monte Carlo Method applied to power generation with renewable energy. Renew Energ 2014;69(3):347–55.

[22] Farges O, Bézian JJ, Bru H, et al. Life-time integration using Monte Carlo Methods when optimizing the design of concentrated solar power plants. Sol Energy 2015;113:57–62.

[23] Janssen H. Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence. Reliab Eng Syst Safe 2013;109:123–32.

[24] Eberhart R, Kennedy J. A new optimizer using particle swarm theory. Proc Int Symp on Micro Mac Human Sci 1995:39–43.

[25] Piltan M, Shiri H, Ghaderi SF. Energy demand forecasting in Iranian metal industry using linear and nonlinear models based on evolutionary algorithms. Energ Convers Manage 2012;58(58):1–9.

[26] Hu C, Jain G, Zhang P, Schmidt C, Gomadam P, Gorka T. Data-driven method based on particle swarm optimization and k-nearest neighbor regression for estimating capacity of lithium-ion battery. Appl Energy 2014;129:49–55.

[27] Tian PC. Wind energy assessment incorporating particle swarm optimization method. Energ Convers Manage 2011;52(3):1630–7.

[28] Chan CL, Chen CL. A cautious PSO with conditional random. Expert Syst Appl 2015;42:4120–5.

[29] Han YM, Geng ZQ, Gu XB, Wang Z. Performance analysis of China ethylene plants by measuring malmquist production efficiency based on an improved data envelopment analysis cross-model. Ind Eng Chem Res 2015;54:272–84.

[30] Han YM, Geng ZQ, Zhu QX, Lin XY. Energy consumption hierarchical analysis based on interpretative structural model for ethylene production. Chin J Chem Eng 2015;23:2029–36.

[31] Geng ZQ, Han YM, Gu XB, Zhu QX. Energy efficiency estimation based on data fusion strategy: case study of ethylene product industry. Ind Eng Chem Res 2012;51:8526–34.

[32] China standards: the general computing guide of special energy consumption (GB/T2589-2008), 2008.

[33] China standards: the limitation of energy consumption for ethylene product (DB37/751-2007), 2008.

[34] Calculation method for energy consumption in petrochemical engineering design (SH/T3110-2001). 2002.

[35] Zhao M, Zhang WG, Yu LZ. Carbon emissions from energy consumption in Shanghai city. Res Environ Sci 2009;22(8):984–9.