Advanced Mathematical Models & Applications
Vol.6, No.3, 2021, pp.322-334

EFFECTIVE DATA PARALLEL OPTIMIZATION FOR QUANTIFYING THE MATHEMATICAL MODEL OF SUBSYSTEMS IN MULTIVARIABLE SYSTEMS

Shokhan Ali Omar[1], Rizhin Nuree Othman[2], Nawroz Fadhil Ahmed[3], Hawkar Saeed Ezat[4], Zana Azeez Kakarash[3]*

[1]Directorate of Information Technology and Statistics, University of Sulaimany, Sulaymaniyah, KRG, Iraq
[2]College of Computer Science and Engineering, Lebanese French University, Erbil, KRG, Iraq
[3]Department of Information Technology, Kurdistan Technical Institute, Sulaymaniyah, KRG, Iraq
[4]Kurdistan Institution for Strategic Studies and Scientific Research, Sulaimani, KRG, Iraq

Abstract. Aiming at the problem that intelligent algorithm can not accurately quantify the mathematical model of each subsystem in the process of multi-variable system identification combined with historical large data, an effective solution of data parallel optimization calculation is proposed. The method combines mechanism modelling, experiment modelling and intelligent modelling. The model structure and initial range of parameters are determined by step experiments of simulation model. The historical data of field operation are mined. The model is corrected by intelligent optimization algorithm, and the transfer function model of the system is obtained. The algorithm quantizes both the particle swarm coding and the original evolutionary search strategy. The experimental results show that the improved algorithm outperforms PSO and QPSO in search ability. Finally, the parameters are estimated by DQPSO algorithm based on the historical data of field operation. The design solution is applied to the identification of transfer function of load control system in thermal power plant. The obtained model lays a foundation for the design and optimization of the controller.

Keywords: ultra-super critical, optimization, intelligent algorithm, PSO, QPSO.
AMS Subject Classification: 68M68, 68T68, 68R06, 68U68.
Corresponding author: Zana Kakarash, Department of Information Technology, Kurdistan Technical Institute, Sulaymaniyah, KRG, Iraq, Tel.: +9647701024959, e-mail: *zana.azeez@kti.edu.krd*
Received: 24 September 2021; Revised: 27 October 2021; Accepted: 16 November 2021; Published: 27 December 2021.

1 Introduction

Establishing the mathematical model of the control system is one of the preconditions for system characteristic analysis and controller design and performance optimization (Chunhwan et al., 2016; Martins et al., 2017). Identification theory has been developed for decades, but in the study of multi-variable identification problems, the single variable open-loop step test method is basically used (Ke et al., 2018). Its essence is to decompose the multi-variable system into a single variable system for model identification, that is, to disturb one of the input variables and keep other inputs unchanged until the system runs steadily again, and then through single variable (Xia et al., 2018; Prasad et al., 2018; Gabarrón et al., 2018). The system model under this input is obtained by quantitative identification method. This method must disturb the normal production process frequently for a long time, seriously affecting the safety of production. It is generally not allowed, and it is difficult to ensure that other input variables remain unchanged (Wu et al., 2019). Literature used state space model to study the identification of multi-variable systems (Deng & Wang, 2017). Literature proposed to construct data product matrix directly from input and output data (Wu et al., 2019). The equivalent relationship between observable normality of system state equation I and II and normality of input and output difference equation was established. By judging the singularity of matrix, the parameters of each subsystem were determined. Literature put forward the F-test discrimination method of multi-variable CARMA model order, and extended it to the F-test discrimination of sub-order and time delay, forming a relatively complete multi-variable CARMA model identification method (Boussé, 2017). In recent years, subspace model identification methods have been widely used. Literature used subspace identification methods to obtain system controller model, process model and disturbance model respectively (Raissi et al., 2017). Literature used the progressive method in identification (Yang et al., 2017). First, the higher-order model was identified, then the model was reduced. Finally, the model was applied to MPC (Nakata, 2017). These identification methods are based on least squares method, and need to add disturbances frequently in the production process, which will affect the safety of production (Downey et al., 2017). Generally, they are not allowed. In addition to these traditional identification methods, intelligent identification methods such as particle swarm optimization, genetic algorithm and ant colony algorithm have also been applied in the field of identification. In the literature, the transfer function model of some circulating fluids bed systems is studied by mining historical data of field operation and using intelligent algorithm (Tsuchiya et al., 2017). However, most of the research objects are single-input single-output systems, lacking in-depth study of multiple-input multiple-output systems (Batselier et al., 2017). Based on the idea of collecting, a simplified non-linear model of coordinated control system for ultra-super-critical once-through boiler units composed of differential equations and algebraic equations is established, and the unknown parameters of the model are identified according to the experimental data (Tang et al., 2017; Lai et al., 2019; Dam et al., 2017; Ramos & Mercère, 2018; Pourbahrami et al., 2020) and (Ahmed et al., 2021). The model is validated by field operation data, but the interaction between the various devices is not fully considered in the process of modelling (Lai et al., 2019). Based on TS fuzzy model, the thermal process is modelled in literature, but the problem of initial value of multi-variable system model is not fundamentally solved (Dam et al., 2017). In view of the iteration nature of intelligent optimization algorithm, choosing an appropriate initial estimation range plays an important role in the optimization results and directly affects the quality of identification model (Ramos & Mercère, 2018; Pourbahrami et al., 2020). QPSO algorithm evolved from PSO is a newly developed optimization method. There are two main forms. One is to code the position of particles by the probability amplitude of quantum bits and update the particles according to the moving mode of quantum revolving gate. It not only extends the traversal ability of particles to the solution space, but also makes the search more precise, and the performance of the algorithm has been improved to a certain extent. The other is to quantize the particle evolutionary search strategy. Particle evolution is realized by median optimal position, particle optimal position and population optimal position. Velocity vectors are eliminated, and the evolution equation is simpler in form, fewer parameters and easier to control. The double quantum particle swarm optimization (QPSO) algorithm proposed in this paper combines the two methods to improve both the number of qualified convergence and the precision of optimization. The problems existing in the process of multi-variable system identification with the combination of intelligent algorithm and historical large data are analyzed through several groups of experiments. A solution of parallel optimization operation for multi-group data is proposed. The influence degree of each input variable on output is quantified. The scheme is applied to the modelling experiment of multi-variable coordinated control system in thermal power plant. The identification results show that the scheme is effective.

2 Double quantum particle swarm optimization

The double quantum particle swarm optimization (QPSO) firstly codes the particle population using the probability magnitude of quantum bits, expands the traversal ability of the particle to the solution space, and then quantizes the evolution process of the population, making the evolution equation simpler in form, fewer parameters and easier to control. The calculation steps are as follows:

2.1 Solution space transformation

In general, the continuous optimization problem can be expressed as min $f(x_1, x_2, ..., x_n)$, $x_i \in [a_i, b_i]$, n is the number of optimization variables,$[a_i, b_i]$ is the definition domain of independent variables, and $f(x)$ is the objective function. Using the probability amplitude of quantum bits as the current position coding of particles:

$$P_i = \left[\left(\left| \begin{matrix} \cos(\theta_{i1}) \\ \sin(\theta_{i1}) \end{matrix} \right| \left| \begin{matrix} \cos(\theta_{i2}) \\ \sin(\theta_{i2}) \end{matrix} \right| ... \left| \begin{matrix} \cos(\theta_{in}) \\ \sin(\theta_{in}) \end{matrix} \right| \right) \right] \tag{1}$$

where θ_{ij} is a random number between (0,1). $i = 1, 2, \cdots M, j = 1, 2, \cdots n$, n is the spatial dimension (the number of optimization variables), m is the size of the population. The probability amplitudes of quantum states $\mid 0 > and \mid 1 >$ corresponding to each particle in the population are as follows:

$$\left\{ \begin{array}{l} P_{ic} = [\cos(\theta_{i1}), \cos(\theta_{i2}) \cdots \cos(\theta_{in})] \\ P_{is} = [\sin(\theta_{i1}), \sin(\theta_{i2}) \cdots \sin(\theta_{in})], \end{array} \right. \tag{2}$$

where P_{ic} the cosine is position of the particle and P_{is} is the sinusoidal position of the particle. The range of P_i is $[-1, 1]$. The transformation from unit space n to solution space of optimization problem is as follows:

$$\left\{ \begin{array}{l} x_{ic}^j = \left[b_i \left(1 + \alpha_i^j \right) + a_i \left(1 + \alpha_i^j \right) \right] /2 \\ x_{ic}^j = \left[b_i \left(1 + \beta_i^j \right) + a_i \left(1 + \beta_i^j \right) \right] /2, \end{array} \right. \tag{3}$$

where each particle corresponds to two solutions of the optimization problem. The probabilistic amplitude of quantum state α_i^j corresponds to $\boldsymbol{X}_{ic}^j$, and the probabilistic amplitude of quantum state β_i^j corresponds to x_{ic}^j.

2.2 Quantization of Particle Renewal Process

In the space-time construction of quantum, the state of the probability amplitude of the quantum bit corresponding to the particle can be expressed by the wave function $\varphi(\vec{\theta}, t)$. According to the superposition state and probability expression characteristics in quantum theory, the probability density of the particle at a certain point in three-dimensional space satisfies:

$$\int_{-\infty}^{+\infty} |\varphi|^2 d\theta dy dz = \int_{-\infty}^{+\infty} Q d\theta dy dz = 1. \tag{4}$$

The evolution of particles is represented by Schrodinger equation:

$$\left\{ \begin{array}{l} i + \frac{\partial}{\partial \theta} \varphi(\vec{\theta}, t) = \hat{H} \varphi(\vec{\theta}, t) \\ \hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + V(\vec{\theta}) \end{array} \right. \tag{5}$$

where $\hbar$ is a Planck constant; $\hat{H}$ is Hamilton function.

By solving the Schrodinger equation $\mathrm{d}^2\varphi/dy^2 + 2m/h^2[E + \gamma\delta(y)]\varphi = 0$, the probability density of particles at some point in space can be obtained

$$\begin{cases} \varphi(y) = e^{-\beta|y|} \\ Q(y) = \frac{1}{L}e^{-2|y|/L} \\ L = 1/\beta = \hbar^2/m\gamma. \end{cases} \tag{6}$$

According to Monte Carlo method, we can get:

$$s = \frac{1}{L}u = \frac{1}{L}e^{-2|y|/L}, \tag{7}$$

where s a random number is evenly distributed between $(1, 1/L)$ and u is a random number between (0, 1). It can be obtained:

$$\begin{cases} u = e^{-2|y|/L} \\ y = \pm\frac{L}{2}\ln(1/u). \end{cases} \tag{8}$$

Then the position equation of the probability amplitude of quantum bit is obtained:

$$\theta(t) =| \theta + y. \tag{9}$$

L value is determined by $L(t-1) = 2\beta\,|\theta best - \theta(t)|$.

Finally, the evolution equation of particles is obtained as follows

$$\begin{cases} \theta_{\text{best}} = \sum_{i=1}^m \frac{\theta_i}{m} = \sum_{i=1}^m \frac{\theta_{i1}}{m}, \sum_{i=1}^m \frac{\theta_{i2}}{m} \cdots \sum_{i=1}^m \frac{\theta_{id}}{m} \\ \theta_j(j) = (\phi_{1j}\theta_{ij} + \phi_{2j}\theta_{gj}) \,/\, (\phi_{1j} + \phi_{2j}) \\ \theta(t+1) = \theta \pm \beta\,|\theta_{best} - \theta(t)|\ln(1/u), \end{cases} \tag{10}$$

where θ_j is the individual extreme value of the probability amplitude of the j-th individuality; θ_{best} is the median optimal position of the probability amplitude of all individuality; m is the population size; ϕ_1, ϕ_2 are the random number between $(0, 1)$; and β is the contraction factor. The experimental results show that the algorithm can basically achieve a better optimization effect when linear reduction from 1.0 to 0.5.

2.3 Variation of Particles

In order to avoid premature convergence of population, mutation operator is introduced into evolutionary algorithm. The mutation operation is realized by quantum non-gate: each quantum is assigned a random value, if the value is less than the given mutation probability, then the [n/2] qubits in the particle are randomly selected, and the quantum non-gate is converted into two probability amplitudes. The optimal position and rotation vector of its memory remain unchanged.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos(\theta_{ij}) \\ \sin(\theta_{ij}) \end{bmatrix} = \begin{bmatrix} \cos(\theta_{ij} + \pi/2) \\ \sin(\theta_{ij} + \pi/2) \end{bmatrix}. \tag{11}$$

3 System Identification of Multi variable Intelligent Mining in Big Data Framework

3.1 The Big Data Framework for Intelligent Mining and System Identification

We study and compare the related technologies of big data and data mining, determine the core engine of big data processing, and implement parallel data mining algorithm based on programming model. However, it is not enough to do this. It is inefficient for users to use cluster directly for data processing and data mining. It requires developers to learn and operate a lot of knowledge related to cluster environment which is not related to actual business needs, which reduces the efficiency of development and improves the use of the engine. This chapter is to design and implement a large transparent data mining platform at the bottom to provide services in the form of platform as a service. Let developers concentrate on data mining business, without paying attention to the underlying cluster configuration, management and other tedious work. The design goal of this system includes two aspects: first, to realize the transparency of the underlying data processing engine cluster to the user layer; second, to realize friendly and easy-to-use user inter-face. The significance of bottom transparency is to liberate users from cluster operations and configurations that are not related to data mining business and to focus on business logic. This is a very tedious and repetitive work, which seriously reduces the development efficiency. In addition, if you want to use data processing, you have to learn the relevant operations, according to its programming model to achieve business logic.

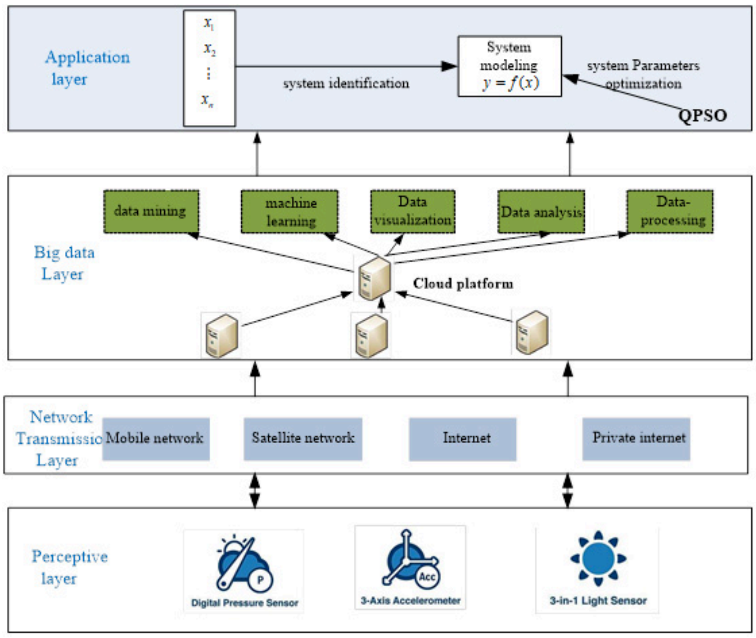

Figure 1: The Big Data Framework for Intelligent Mining and System Identification

Because it is a newly emerging data processing engine, the relevant information is relatively small, and few people are familiar with its operation mode, many data mining practitioners do not have relevant experience. Good tools are not used by people who can use them, nor can they play their role. Another problem is that cluster environments are usually operating system-based, which requires developers to be familiar with environment programming and operation, which is often unrealistic. If the data source access, environment configuration and other operations can be encapsulated and provided in the way of interface, and the related operations can be provided in the way of development toolkit (to achieve cross-platform remote calls, so that users can not feel the existence of the underlying engine, using it as local data, we can break through the limitations of the system environment, programming model and reduce the quotation. Using of engine door planting to improve work efficiency.

As is shown in the Fig.1, the Big Data Framework is composed of perception layer, network transmission layer, Big Data layer and application layer. In the perceptual interaction layer, the meteorological data and sensors data are collected to achieve all-weather real-time acquisition of scenic data. It is the foundation of the big data framework.

The network transmission layer transmits data through special equipment such as base station and public network, and realizes the upload of perceptual data and the delivery of system and control management instructions.

The application layer analyzes and processes the multi-source heterogeneous data, establishes the perceptual linkage relationship between the multiple inputs and outputs. The double quantum particle swarm optimization (QPSO) are used to update the system parameters.

3.2 The Model identification mechanism and model correction with QPSO algorithm

The modelling method proposed in this paper is as follows: Firstly, a simulation model of ultra-super critical units is established by mechanism modelling method. Step experiments are carried out on the mechanism model, and the structure of the transfer function model and the initial values of each parameter are preliminary determined. Then, the trans-fer function model of MIMO system is corrected by excavating the massive historical data of field operation. Finally, the transfer function model of MIMO system is obtained.

Even if the static characteristics of the unit simulated by the mechanism model are very close to the actual system, the dynamic trend can be better approximated to the field. Because the inevitable simplification process in the process of modelling makes the error between the model output and the actual output in the field unavoidable, the identification results based on the mechanism model can't accurately describe the actual unit characteristics.

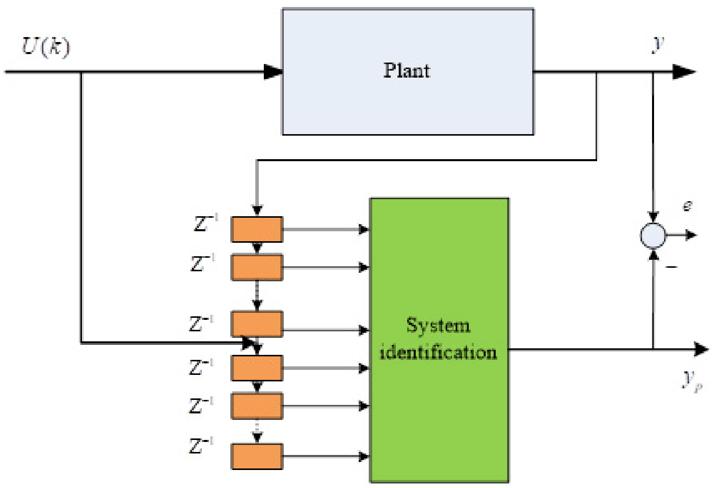

Figure 2: The System Identification using Big Data framework

For a SISO nonlinear control system, $u(k)$ is the control signal of the system, y and y_p are the output of the actual system and the output of the identification system respectively, where k is a time series, e is the error between the output of the actual system y and the output of the identification system. For an identification system based on empirical data sets, it is usually a departure. The output of the scattered system $k + 1$ time can be expressed as n previous outputs and M previous input functions (NARMAX model). The model of nonlinear identification system based on big data is as shown in Fig. 2.

With the wide application of monitoring and management information system, each system has a large amount of historical data of unit operation. After mining these data, it can be used for identification. Using the selected data to correct and optimize the identification results of the preliminary simulation model, the accurate unit model can be obtained.

4 Results and discussion

4.1 Numerical optimization experiment and analysis of classical functions

The validity of the algorithm is validated. The D-QPSO algorithm is tested by several benchmark functions and compared with the optimization results of QPSO and PSO. The selected test functions are shown as follows:

$$\max f_1(x,y) = 0.5 - \frac{\left[\sin\sqrt{x^2+y^2}\right]^2 - 0.5}{\left[1+0.001^*\left(x^2+y^2\right)\right]^2} \tag{12}$$

The function in (12) is Shaffer's function has infinite local maximum points, of which only one $(0,0)$ is the global maximum and the maximum value is 1. And the variable range is between the -100 and 100 .

$$\begin{aligned} \max f_1(x,y) = & -x\sin\sqrt{y+1-x}\cos\sqrt{y+1-x} \\ & +(y+1)\cos\sqrt{y+1-x}\sin\sqrt{y+1-x}. \end{aligned} \tag{13}$$

The function in (13) is multi-peak function with infinite maximum points and global maximum 511.7319, and variable range is $x, y \in [-512, 512]$.

$$\max f_3(x,y) = \frac{1}{4000}\sum_{i=1}^{n}\left(x_i-1000\right)^2 - \prod_{i=1}^{n}\cos\left(\frac{x_i-100}{\sqrt{i}}\right)+1. \tag{14}$$

The function in (14) is Griewank function which is a multidimensional function, a multidimensional function, and the global minimum value is 0 , and variable range is $x, y \in [100, 600]$.

The parameters of the algorithm are as follows: the particle swarm size is 50, the maximum number of iterations is 500, the optimization repetition is 50, and the final results are averaged. In D-QPSO, the shrinkage factor decreases linearly from 1.0 to 0.5; in QPSO (Lai et al. 2017), the inertia weight is 0.5, the self-factor value is 2.0, the global factor value is 2.0, and the variation probability is 0.05; in PSO (Dam et al. 2017), the inertia weight is 0.5, the self-factor value is 2.0, and the global factor value is 2.0. When the optimization result satisfies the condition, it is judged to be qualified and the iteration terminates.

The experimental results are shown in Table 1. It can be seen that D-QPSO algorithm has improved the number of qualified convergence times and the accuracy of optimization, especially in high-dimensional optimization problems.

Table 1: Simulation results

function	algorithm	best	worst	mean	theoretical	$\Delta^{1)}$	Qualification times
f_1	D-QPSO	1	0.9903	0.9966	1	< 0.01	50
	QPSO	1	0.9856	0.9929			44
	PSO	0.999	0.9847	0.9899			23
f_2	D-QPSO	511.7251	510.9182	511.5933	511.7319	< 0.23	48
	QPSO	511.7088	510.5853	511.3939			46
	PSO	511.708	499.4963	511.292			39
f_3	D-QPSO	21.3395	283.0107	106.663	0	< 150	42
	QPSO	134.424	377.1028	231.103			9
	PSO	206.929	1.2799*103	819.83			0

$\Delta^{1)} = |f^* - f|$, f^* is optimal value, f is theoretical value

4.2 Numerical optimization experiment and analysis of classical functions

The coordination system of ultra-super critical direct current units in thermal power plants is a typical multi-variable system. The input of the system is feed water flow, total fuel flow and turbine valve opening, and the output is turbine power, main steam pressure and intermediate point temperature. In this section, the mathematical model between unit power and three inputs is identified.

By analyzing the mechanism of load control system, the form of reference model set can be determined as shown in Formula (22). The fitness function is the mean square deviation function, as shown in equation (21).

$$G(s) = \left[k_2 + \frac{k_1(1-as)}{(T_1s+1)\,(T_2s+1)\cdots(T_ns+1)}\right] e^{-ds}. \tag{15}$$

In ultra-super critical once-through units, when each input is disturbed separately, the influence on unit power is as follows: when the feed water flow increases, the steam flow at the outlet of super heater increases, and the steam turbine power rises first, but because the total fuel quantity (total energy) of the boiler remains unchanged, the steam parameters decrease correspondingly, and the steam turbine power falls back to the original level, and finally it is slightly lower than the original level. When the total energy of the boiler is reduced, the steam pressure and temperature will be reduced to a certain level, and then the power of the steam turbine will be reduced, and finally stabilized at the level corresponding to the amount of fuel. When the turbine regulating valve is opened, the steam intake increases and the power of the turbine increases rapidly, but the total energy of the boiler remains unchanged. After a period of time, the power of the turbine restores to the original level. The deviation of the boiler from the optimum operating state will have a certain impact on the boiler efficiency and turbine efficiency, so the unit power will be slightly lower than the previous level.

Experiments were carried out on the established mechanism model. The experimental conditions were stable operation of the unit at full load, operation of five coal mills and two steam feed pumps, 100% opening of the main valve, middle main valve and middle regulating valve, load adjustment through high regulating valve, main control of the unit boiler, main control of the turbine and automatic release of feed water.

Firstly, the feed water disturbance experiment is carried out, then the coal feed disturbance experiment and the high-profile gate disturbance experiment is carried out. Because the feed water disturbance experiment can ensure that the feed coal and the high-profile gate do not move at all, and the feed water disturbance experiment and the high-profile gate disturbance experiment will inevitably cause the change of steam-water system pressure, which will cause the fluctuation of the feed water flow. For the correctness of the model, this part of the feed water disturbance cannot be ignored. Therefore, in the identification, firstly, the feed water disturbance identification is carried out, and then the coal feed disturbance and high-profile gate disturbance identification are carried out based on the feed water disturbance identification. The step increase of feed water flow is about 10%, and the response curve is shown in Fig. 3. The increase of feed water flow will lead to the increase of steam flow at the outlet of super heater, and the power of steam turbine will rise first. Be-cause the total fuel flow (total energy) of boiler re-mains unchanged, the temperature of super heated steam will decrease, the quality of steam will also decrease accordingly, and the power of steam turbine will fall back to the original level. With the increase of the speed of feed water pump, the feed water flow and the pressure of each node will increase, and the main steam pressure will increase. The outlet temperature (midpoint temperature) of the steam-water separator decreases to a certain level due to the decrease of the fuel-water ratio.

On the basis of the above experimental data, the parameters $k_1, k_2, T_1, T_2, T_3, a, d$ in formula are optimized by the double quantum particle swarm optimization algorithm to minimize the fitness function. The parameters from $k_1 \in [-10, 10]$ $\quad k_2 \in [-10, 10]$, T_1, T_2, $\mathrm{T}_3 \in [0, 1000]$, $\mathrm{a} \in$

$[0, 100]$, population size 50, mutation probability 0.05, maximum iteration number 500, shrinkage factor linearly reduced from number 500, shrinkage factor linearly reduced from 1.0 to 0. The identification results are shown in Fig. 4 and the corresponding parameters are shown in Table 2.

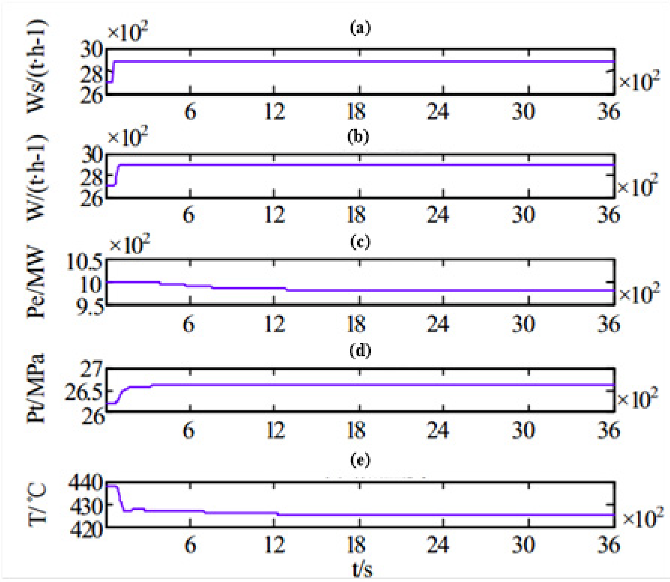

Figure 3: Feed water flow response curve under step signal (a) Setting Value of Outlet Flow of Feed Pump (b) Feed water flow rate (c) Actual power (d) Main steam pressure (e) Intermediate point temperature

Table 2: Simulation results of system identification parameter

	k1	k2	a	T1	T2	T3	d	Err
G11	-0.1	0	171.3	742.3	34.8	0	18	0.131
G12	0.0022	0	-141.3	179.3	24.1	0	6	0.002
G13	-0.066	0	-760.9	914.4	9.7	0	12	0.122
G21	3.47	0	0	182.2	73.6	0	0	0.232
G22	0.11	0	0	19.1	19.1	144.2	10	0.006
G23	0.7	0.35	-621.5	345.3	364.1	0	20	0.105
G31	-5.14	4.2	281.8	265.3	86.7	0	6	0.206
G32	-0.51	-0.07	0	180.2	0	0	0	0.007
G33	-2.15	0.19	0	134.7	94.19	23.4	21	0.059

Step wise increase of comprehensive valve position opening of steam turbine regulating valve is about 1% (about 2% of operation of high regulating valve). Its response curve is shown in Fig. 4. Because of the opening of the regulating valve, the steam intake of the steam turbine will increase rapidly, but the fuel flow and feed water flow will remain unchanged eventually. After a period of time, the power of the steam turbine will return to the original level. The main steam pressure decreases with the opening of the regulating door, and maintains at a certain level after balance. Because of the fluctuation of the main steam pressure, the feed water will fluctuate to a certain extent, resulting in the fluctuation of the mid-point temperature, and finally stabilize at a slightly low level. After the test, the deviation of the boiler from the optimal operating state will have a certain impact on the boiler efficiency and turbine efficiency, so the power and intermediate temperature will be slightly lower than the previous level. The experimental results show that our model can closely track the actual values and achieve good results.

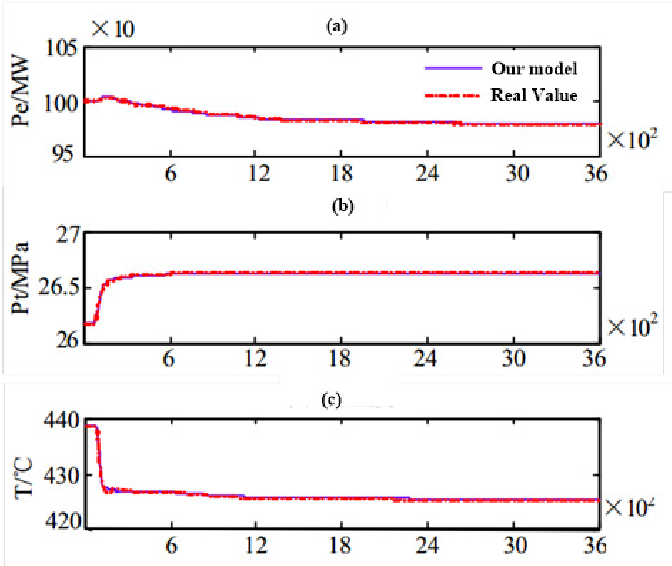

Figure 4: 4 Disturbance Identification Results of Feedwater Flow under step signal (a) Actual power (b) Main steam pressure (c) Intermediate point temperature

4.3 Model Identification and Verification of Field Data

Taking a 1000MW ultra-super critical unit in China as the research object, the sampling period is 5 seconds and the sampling time is 50 minutes for the data of the working condition near 800MW. Fig. 5 (a) is the unit operation data from 13:22 to 14:12 on Dec. 10, 2018, and Fig. 5 (b) is the unit operation data from 19:15 to 20:05 on Dec. 26, 2018. Among them: W is the feed water flow, t/h; C is the total fuel flow, t/h; M is the high-profile valve position, The original data are processed by outlier elimination, five-point cubic smoothing filter and zero-mean method (Ke et al., 2018). The identification results are shown in Fig. 5 (a) and (b). The optimal values of model parameters are obtained as shown in table 3. Among them, Gw, Gc and Gm are respectively the transfer function models of feed water flow, total fuel flow and high-profile valve position to unit load. The mean square deviation of the identification results is 0.788.

In order to verify that the model can represent the thermal characteristics of the load control system at this working point, the historical data of 20:20-21:10 on January 30, 2019 are selected for model verification. The unit operation data are shown in figure5. The mean square deviation was 0.434. From the results, the identification model can reflect the dynamic characteristics of the system.

Table 3: Verification results of system identification parameter

parameters	Gw	Gc	Gm
k_1	0.049	2.812	6.325
k_2	0.108	0.545	3.73
n	1	3	3
T_1	139.581	78.733	539.026
T_2	0	264.928	818.267
T_3	0	125.519	849.617
a	245.465	3.794	496.231
d	5	0	9

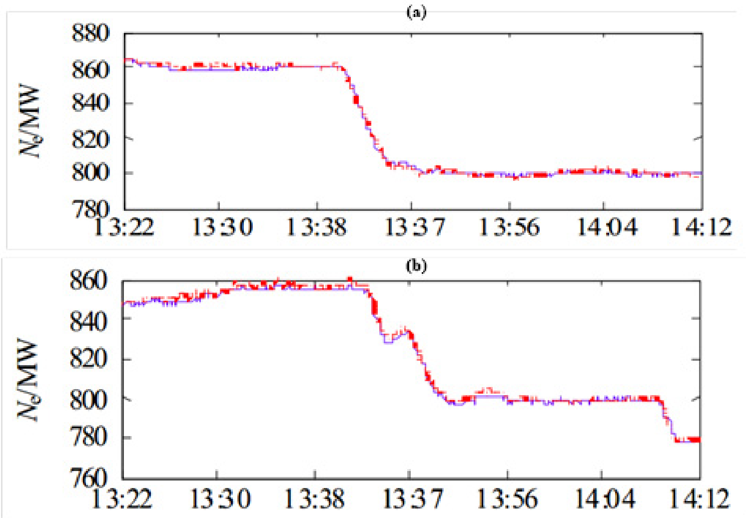

Figure 5: Result of identifying process under different time: Actual data(solid line) and model data (dashed line)

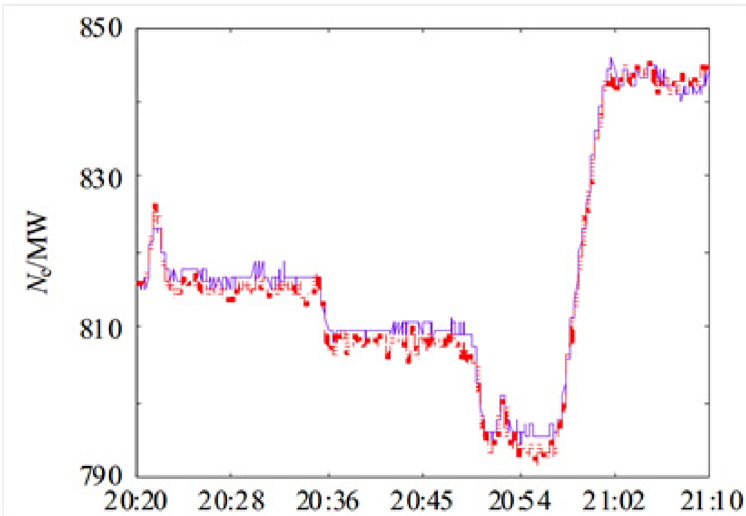

Figure 6: Result of verification process under different time: Actual data (solid line) and model data (dashed line)

In the selection principle of massive historical data, besides the premise of corresponding working conditions, the following principles should be followed as far as possible: (1) The starting and ending conditions should be as smooth as possible; (2) the data in the process of change should have a sufficiently large signal-to-noise ratio; (3) the sampling period should be appropriate. By using the double quantum particle swarm optimization algorithm, the parameters identified in the above section are doubled as the base points, and then the transfer function model of the unit's accurate coordinated control system is obtained by re-optimizing the parameters. The model obtained in the previous section can basically reflect the numerical simulation relationship between input data and out-put data, but it does not prove that the model can fully represent the thermal characteristics of the working point. Therefore, it needs to be validated by data independent of identification data. The validation data is the operation data of another time period near the identifying operating point. The input and output variables are the same as before. The output of the transfer function model and the actual output of the unit are shown in Fig 6. The mean square errors are 0.2591, 0.0949 and 0.3604, respectively. From the verification results, it can be seen that the output values of the three output models are basically consistent with the operating values of field units. The model can be used for the design and optimization of coordinated control system control.

5 Conclusion

Aiming at the problem that intelligent algorithm cannot accurately quantify the mathematical model of each subsystem in the process of multi-variable system identification combined with historical large data, an effective solution of data parallel optimization calculation is proposed. The method combines mechanism modelling, experiment modelling and intelligent modelling. The model structure and initial range of parameters are determined by step experiments of simulation model. The historical data of field operation are mined.

In this paper, mechanism modelling, experimental modelling and intelligent modelling are combined, and a modelling method suitable for multi variable systems is proposed. The development and application of the mechanism model avoids the step response of the production site, determines the structure and initial parameters of the MIMO model, corrects the transfer function model obtained from the preliminary test by using the double quantum particle swarm optimization algorithm and selected field operation history data, and finally obtains the accurate transfer function model. In this paper, the idea of modelling is successfully applied to the coordination system modelling of ultra-super critical units, and a multi-input multi-output transfer function model for the optimization of the controller of the coordinated control system is obtained. The transfer function model of any working point of other thermal systems can be established by using this method, which is also the next step of the author's work.

6 Acknowledgement

The authors acknowledge Collaborative Innovation Center for Integration of Terrestrial and Marine Economies (Grant: 2019YB15), the National Natural Science Foundation of China (Grant: 61762006, 71862003).

References

Ahmed, N.F., Omar, S.A., Othman, R.N., & Kakarash, Z.A. (2021). SQL Processor based on Intelligence Technique: Fuzzy Petri Net Database Applications. *Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12*(14), 1364-1371.

Boussé, M., Debals, O., & Lathauwer, L. D. Tensor-based large-scale blind system identification using segmentation. *IEEE Transactions on Signal Processing, 65*(21), 5770-5784.

Batselier, K., Chen, Z., & Wong, N. (2017). A tensor network kalman filter with an application in recursive mimo volterra system identification. *Automatica*, 84, 17-25.

Chunhwan, K., Kicheol, S., Youngbin, J., Chankyu, L., Doogyung, M., & Song, S.Y. (2016). Establishment of discrimination system using multivariate analysis of ft-ir spectroscopy data from different species of artichoke (Cynara cardunculus var. scolymus l.). *Korean Journal of Horticultural Science & Technology*, 35 Suppl 7(7 Supplement), 89-91.

Dam, M., Brøns, M., Juul Rasmussen, J., Naulin, V., & Hesthaven, J.S. (2017). Sparse identification of a predator-prey system from simulation data of a convection model. *Physics of Plasmas, 24*(2), 022310.

Deng, W., Wang, G. (2017). A novel water quality data analysis framework based on time-series data mining. *Journal of Environmental Management*, 196, 365-375.

Downey, B., Schmitt, J., Beller, J., Russell, B., Quach, A., Hermann, E., ... & Breit, J. (2017). A system identification approach for developing model predictive controllers of antibody quality attributes in cell culture processes. *Biotechnology progress, 33*(6), 1647-1661.

Gabarrón, M., Faz, A., & Acosta, J.A. (2018). Use of multivariable and redundancy analysis to assess the behavior of metals and arsenic in urban soil and road dust affected by metallic mining as a base for risk assessment. *Ultramicroscopy*, 181, 150.

Ke, Q., Wu, S., Wang, M., & Zou, Y. (2018). Evaluation of developer efficiency based on improved DEA model. *Wireless Personal Communications, 102*(4), 3843-3849.

Lai, C.F., Chien, W.C., Yang, L.T., & Qiang, W. (2019). LSTM and edge computing for big data feature recognition of industrial electrical equipment. *IEEE Transactions on Industrial Informatics, 15*(4), 2469-2477.

Martins, A.R., Talhavini, M., Vieira, M.L., Zacca, J.J., & Braga, J.W.B. (2017). Discrimination of whisky brands and counterfeit identification by $UV - V$ is spectroscopy and multivariate data analysis. *Food Chemistry*, 229, 142-151.

Nakata, K., Orihara, R., Mizuoka, Y., & Takagi, K. (2017). A comprehensive big-data-based monitoring system for yield enhancement in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing, 30*(4), 339-344.

Pourbahrami, S., Balafar, M.A., Khanli, L.M., & Kakarash, Z.A. (2020). A survey of neighborhood construction algorithms for clustering and classifying data points. *Computer Science Review,* 38, 100315.

Prasad, R., Deo, R.C., Yan, L., & Maraseni, T. (2018). Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors. *Soil & Tillage Research,* 181, 63-81.

Raissi, M., Karniadakis, G.E. (2017). Hidden physics models: machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357, 125-141.

Ramos, J.A., Mercère, G. (2018). A stochastic subspace system identification algorithm for state-space systems in the general 2-D Roesser model form. *International Journal of Control, 91*(12), 2743-2771.

Tang, W., Wu, J., & Shi, Z. (2017). Identification of reduced-order model for an aeroelastic system from flutter test data. *Chinese Journal of Aeronautics, 30*(1), 337-347.

Tsuchiya, T., Fujii, M., Matsuda, N., Kunida, K., Uda, S., & Kubota, H., et al. (2017). System identification of signaling dependent gene expression with different time-scale data. *Plos Computational Biology*, *13*(12), e1005913.

Wu, S., Liu, J., & Liu, L. (2019). Modeling method of internet public information data mining based on probabilistic topic model. *The Journal of Supercomputing, 75*(9), 5882-5897.

Wu, S., Zhang, Q., Chen, W., Liu, J., & Liiu, L. (2019). Research on trend prediction of internet user intention understanding and public intelligence mining based on fractional differential method. Chaos, *Solitons & Fractals,* 128, 331-338.

Xia, D., Lu, X., Li, H., Wang, W., Li, Y., & Zhang, Z. (2018). A MapReduce-based parallel frequent pattern growth algorithm for spatiotemporal association analysis of mobile trajectory big data. *Complexity*, 2018.

Yang, D., Dong, Z., Lim, L.H.I., & Liu, L. (2017). Analyzing big time series data in solar engineering using features and pca. *Solar Energy,* 153, 317-328.