Multi-objective optimization for oil-gas production process based on compensation model of comprehensive energy consumption using improved evolutionary algorithm

Tan Liu¹,², Qinyun Yuan¹,², Lina Wang³, Yonggang Wang¹,² and Nannan Zhang¹,²

Abstract
This paper establishes an error compensation multi-objective optimization model of oil-gas production process for optimizing these production indices, including overall oil production, overall water production and comprehensive energy consumption per ton of oil. In order to reduce the error between the model output and the actual value of comprehensive energy consumption per ton of oil, combining the mechanism model with least squares support vector machine (LS-SVM) error model optimized by Bayesian optimization algorithm (BOA), a hybrid model is established to predict the comprehensive energy consumption, in which the mechanism model is used to describe the overall characteristics of oil-gas production process, and LS-SVM error model is established to compensate the mechanism model error. Then, in order to improve the performance of Pareto non-dominated solutions, an improved non-dominated sorting genetic algorithm-II with multi-strategy improvement (IMS-NSGA-II) is proposed to solve the error compensation multi-objective optimization model. Finally, the effectiveness and superiority of the proposed optimization method are verified by the experiment results on some standard test problems and the optimization problem for the oil-gas production process in a block of an oil production operation area.

¹College of Information & Electrical Engineering, Shenyang Agricultural University, Shenyang, China
²Liaoning Engineering Research Control Information Technology Agricultural, Shenyang, China
³College of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou, China

Corresponding author:
Qinyun Yuan, Shenyang Agricultural University, Shenyang, China.
Email: yqy8_29@126.com
Keywords
Oil-gas production process, LS-SVM error model, Bayesian optimization algorithm, error compensation, NSGA-II with multi-strategy improvement (IMS-NSGA-II)

Introduction
Although technological progress has increased the diversity of energy alternatives, fossil energy is still the main energy in the modern society. Oil-gas is a kind of finite and scarce resource, and modern society has a strong dependence on it. In fact, the world’s demand for petroleum continues to grow, but with the continuous exploitation of petroleum, the world’s petroleum reserves continue to decline. In addition, with the continuous exploitation of oil, the industry of oil-gas production generally faces with some challenges such as the relative decline of formation energy, the increase of comprehensive water content, the gradual decline of surface gathering system load, the low efficiency and the gradual increase of the cost per ton of oil. These problems have caused the waste of resources and environmental pollution in the oil industry, and seriously affected the sustainable development of the oil industry. Therefore, in order to improve production, save energy, and make the production process safer and more efficient, advanced modeling, control and optimization technology have become extremely important and indispensible in the oil-gas production process (Al-Fehdly et al., 2019). Due to the complexity and uncertainty of oil-gas production system (Foss, 2012; Kosmidis et al., 2005), the modeling and optimization of oil-gas production process is a formidable challenge for comprehensive optimization of the process (Codas et al., 2012; Lei et al., 2019). Since the energy consumption is usually an important objective of optimization of oil-gas production process, and the energy consumption of pumping system and oil-gas gathering system accounts for a very large proportion of energy consumption in oil-gas production process, many scholars have proposed the optimization models and energy-saving methods of relevant systems.

In the optimization of sucker-rod pumping system, Zheng and Deng (2007) build an optimization model with the minimum energy consumption as objective function. Comprehensively considering the constraints of oil well productivity coordination, oil well production allocation, pumping unit bearing capacity, sucker rod string strength and motor power utilization, and taking the maximum system efficiency as the objective function, Dong et al. (2008a) propose the corresponding optimization simulation method of suction parameters for sucker-rod pumping. In order to overcome some shortcomings of single well research, Dong et al. (2010) also put forward the overall optimization method of an oil production block with the minimum overall input power as the objective function and the overall oil production as the constraint. The actual application shows the superiority of this optimization method. Considering the influence of plunger motion law, valve ball motion law, oil well inflow performance, etc, Wang et al. (2018) establish the dynamic parameter simulation model of pumping wells and study the methods to improve the system efficiency. For gas injection wells, Jun et al. (2017) use Monte Carlo simulation method to select the optimal gas recovery scheme, then genetic algorithm is used to find the optimal development scheme. In order to overcome the difficulties brought by the low oil prices and provide meaningful and valuable information for operators, Rui et al. (2018)
develop an integrated evaluation model by analyzing a large number of actual historical
data (from the subsurface reservoir to infield flow line).

For the optimization of oil-gas gathering and transferring system, some optimization
methods have been reported (Chebouba et al., 2009; Cui et al., 2008; Dong et al., 2017;
He et al., 2007; Liu and Chen, 1999; Tan et al., 2016; Wu, 2011; Wu et al., 2014), these
methods are mainly to pursue the minimum production operation cost and energy con-
sumption (Cheng et al., 2020). For example, in order to minimize the capital expenditure
such as pipeline cost, power consumption cost and heat energy cost, Liu and Chen (1999)
establish a mathematical optimization model of oilfield surface pipeline system by using
fuzzy optimization method. He et al. (2007) establish the optimization model of gathering
and transportation for electrical heating about ramiform net, in which the lowest investment
and operating cost are taken as the objective functions. The experiment results show that the
method can reduce the total investment and heating cost of gathering and transferring
system. According to the characteristics of energy consumption increasing year by year in
the gathering and transmission system, taking the tree-like water-mixed gathering and trans-
portation network as the research object, and taking the water-mixed quantity, water-mixed
temperature and water-mixed pressure of the network as the parameter variables, the genetic
algorithm is used to optimize the parameters (Cui et al., 2008) to make the energy con-
sumption of the network operation significantly lower than before optimization. For heavy
oil pipeline heating transportation, Wu (2011) establishes the corresponding energy con-
sumption model, which significantly reduces the energy consumption. Balancing the max-
imum operation benefits and the maximum transmission capacity, Wu et al. (2014) built an
optimization model of natural gas pipelines by using the weight sum method, and according
to the nonlinear characteristics of the optimization model, they propose an improved par-
ticle swarm optimization (PSO) algorithm to solve the optimization model. In view of the
complex composition and behavior of gathering and transportation system, based on a large
number of data accumulated in the process of oil and gas gathering and transportation, Tan
et al. (2016) establish time series prediction model of production energy consumption in
gathering and transportation system, and study the correlation between energy consumption
factors. Wang et al. (2019) analyze the gelation behavior of waxy crude oil emulsion, and
establish its dynamic model to provide a certain reference for waxy crude oil production and
gathering and transportation optimization. For large-scale oil-gas gathering system, Liu
et al. (1999) establish a high-dimensional mixed integer nonlinear optimization mathemat-
ical model including pipeline design parameters, and propose a modified MPSO algorithm
with global search ability for optimization solution. Aiming at the increase of water content
in the old areas, the existing gathering and transportation modes cannot meet the require-
ments of energy saving in oilfield production, Zhao et al. (2019) derive the temperature drop
and pressure drop models of oil, gas and water three-phase flow, and determine the oper-
ating parameters of single well gathering, branch connection gathering and annular gath-
ering for four wells.

The above optimization methods for oil-gas production process have achieved remark-
able results in practice, but they often only consider a single objective and lack of the
comprehensive consideration of oil-gas production process. In the actual production pro-
cess, oil recovery enterprises may have different production requirements during different
periods, i.e., at the initial stage of production, the main goal is to maximize the oil produc-
tion within a specified range of comprehensive energy consumption. At the middle and later
stage, the main goal is to minimize the comprehensive energy consumption within a specified
range of oil production. Additionally, the requirements of the overall oil production during different periods are not the same. So, it is necessary to consider some production indices simultaneously to satisfy the different requirements of the overall oil production, the energy consumption and so on. Ideally, these indices are expected to be the best at the same time. However, due to the conflict among these indicators, it is impossible for all the indices to reach the optimal values simultaneously. For example, when an oil well has a certain amount of oil production, it needs more input power of pumping system for continuing to increase oil production, which will consume more energy consumption of oil-gas gathering and transportation system. Ultimately, the more oil output tends to consume more energy. Additionally, when the output changes in a certain range, the increase of output will also lead to the increase of water production. Therefore, the improper allocation of these indices may lead to poor product quality, difficult to ensure the continuity and stability of production, and ultimately affect the economic benefits of oil recovery enterprises. In addition, due to the complexity of oil-gas production process, there may be a large error between the prediction value and the actual value of objective function. Therefore, it is necessary to establish a more accurate objective function model in the comprehensive optimization of oil-gas production process.

To address the above problems, this paper establishes an error compensation multi-objective optimization model for the integrated oil-gas production process. In order to improve the prediction accuracy of objective function model, a LS-SVM error compensation method optimized by Bayesian optimization algorithm (BOA) is put forward to compensate the mechanism model error. Then, an improved non-dominated sorting genetic algorithm-II with multi-strategy improvement (IMS-NSGA-II) is proposed to solve the multi-objective optimization model. Finally, the oil-gas production process of one block in an oil recovery operating area is taken as the research background. The simulation results show the effectiveness and superiority of the established optimization model and algorithm.

### Oil-gas production process

In this paper, rod pumping wells are taken as production wells, and the flowchart of the oil-gas production process is shown in Figure 1.

![Flowchart of integrated oil-gas production system.](image-url)
From Figure 1, it can be seen that a typical integrated oil-gas production system consists of reservoirs, production wells, manifolds, surface flow lines and surface facilities (Jahanshahi et al., 2020). First, the mixture of oil, gas and water is transported from the reservoir to each oil well, then to the oil well platform (manifold), and then to the surface facilities through the surface flow lines for oil, gas and water separation, processing and recycling. In the entire production system, the oil wells are the important component of production, and also the main production equipment to transport oil, gas and water multi-phase flow from reservoir to the surface. In actual production, the oil production, water production and comprehensive energy consumption are usually important production indices, in which the comprehensive energy consumption mainly includes the power consumption in sucker-rod pumping system, the power consumption and the natural gas consumption in gathering system.

Multi-objective optimization model of oil-gas production process based on error compensation of comprehensive energy consumption

This section is organized as follows: The first part is the objectives of the optimization model for the integrated oil-gas production process: including overall oil production, overall water production, comprehensive energy consumption per ton of oil. Then, the second part describes the multi-objective optimization model. Finally, the error compensation of comprehensive energy consumption per ton of oil is given in the third part: including hybrid model structure based on error compensation, LS-SVM error model based on Bayesian optimization algorithm.

Objectives of the optimization model for the integrated oil-gas production process

Overall oil production. Overall oil production is the sum of the oil production of each well in the block platform, which is expressed as follows:

\[
Q_0 = \sum_{j=1}^{N} Q_j (1 - w_{0j})
\]  

(1)

where \(Q_0\) is the overall oil production (m\(^3\)/d), \(Q_j\) is the actual flow rate (m\(^3\)/d) of the \(j\)-th oil well, \(w_{0j}\) is the water content (%), and \(N\) is the number of rod-pumped wells.

For each oil well, the wellbore model consists of two parts, including the multiphase flow from the reservoir into the wellbore and the flow rate in the pipe, they are represented by the inflow performance relation (IPR) curve of oil wells and outflow characteristic curve of pump respectively. When the static pressure is lower than the saturation pressure, the inflow performance relationship of the reservoir is expressed by Vogel equation (2) (McCoy and Podio, 2001).

\[
q^{IN} = q_{max} \left\{ 1 - 0.2 \left( \frac{p_{wf}}{p_r} \right) - 0.8 \left( \frac{p_{wf}}{p_r} \right)^2 \right\}
\]  

(2)
where $q_{\text{max}}$ is the maximum liquid production ($p_{\text{wf}} = 0$) of oil well (m$^3$/d), $p_{\text{wf}}$ is the bottom hole flow pressure, $p_r$ is the reservoir pressure (Pa), $p_{\text{sat}}$ is defined as the saturation pressure (Pa). When $p_r > p_{\text{sat}}$, equation (2) can be expressed as follows (Patton and Goland, 1980):

$$q^{IN} = \begin{cases} \frac{p_{\text{sat}}}{p_r - p_{\text{sat}}} (p_{\text{wf}} - p_{\text{sat}}) & q_{\text{wf}} \geq p_{\text{sat}} \\ p_{\text{sat}} + (q_{\text{max}} - p_{\text{sat}}) \left\{ 1 - 0.2 \left( \frac{p_{\text{wf}}}{p_{\text{sat}}} \right) - 0.8 \left( \frac{p_{\text{wf}}}{p_{\text{sat}}} \right)^2 \right\} & q_{\text{wf}} < p_{\text{sat}} \end{cases}$$ (3)

The inflow performance of oil well reflects the oil supply capacity of the reservoir to the oil well, which depends on reservoir pressure, geological properties of well surroundings, and the wellbore itself, etc.

Due to the oil flow resistance between reservoirs and well bores, the pressure of oil production well bore changes slowly with time with time, and the following equation can be obtained (Horne, 1998).

$$p_{\text{wf}} - p_r = \frac{141.2q^{IN}B\mu}{k_c h} \left( \frac{1}{2} \ln \frac{0.00264kt}{\phi \mu c_p r_B^2} + 0.80907 \right)$$ (4)

where $B$, $\mu$, $k_c$, $h$, $c_t$, $r_B$ and $\phi$ are experimentally parameters related to the reservoir and the oil well.

On the other hand, for rod pumping wells, the actual flow rate of the $j$-th well is calculated as follows:

$$Q_j = 1440\pi D_j^2 S_j n_j x_j/4$$

$$x_j = \eta_S \eta_F \eta_L \eta_V$$ (5)

In equation (5), $D_j$, $S_j$, $n_j$, $x_j$ represent the pump diameter (m), the stroke (m), the stroke times (min$^{-1}$) and the pump efficiency (%) of the $j$-th rod-pumped well, respectively. $\eta_S$, $\eta_F$, $\eta_L$, $\eta_V$ represent the effective stroke coefficient of oil well, fullness coefficient, leakage coefficient, and volume coefficient of gas dissolution crude oil under submerging pressure, which can be calculated as follows (Dong et al., 2008a; Dong et al., 2008b):

$$\begin{align*}
\eta_S &= \frac{S_{\text{pump}}}{S} \\
\eta_F &= \frac{1}{R} \left\{ 1 - \frac{KR}{1 + R(p_s/p_d)^{1/k}} \left[ 1 - (p_s/p_d)^{1/k} \right] \right\} \\
\eta_L &= \frac{A_p S \eta_S \eta_F \eta_V - \Delta Q}{A_p S \eta_S \eta_F \eta_V} \\
\eta_V &= \frac{1}{[(1 - w_0)B_{op} + w_0 B_{wp}]} 
\end{align*}$$ (6)
In the above formula, $S_{pmup}$ is the effective stroke length of plunger (m), $K$ is the clearance coefficient, $R$ denotes the gas liquid ratio of pump suction ($\text{m}^3/\text{m}^3$), $p_s$ and $p_d$ represent submerging pressure (Pa) and discharge pressure (Pa) shown in equation (7). $A_p$ is the cross section area of plunger pump ($\text{m}^2$), $k$ is the index of natural gas polytropic process, $w_0$ is the water content (%), $\Delta Q$ is the liquid clearance leakage between the plunger and pump barrel ($\text{m}^3$). $B_{ops}$, $B_{wps}$ are the oil volume coefficient and the water volume coefficient under the pump suction, which have a certain relationship between $p_d$ and reservoir temperature $t_L$.

\[
\begin{align*}
    p_d &= p_0 + [w_0 \rho_w + (1 - w_0) \rho_0]gL \\
    p_s &= p_c + [w_0 \rho_w + (1 - w_0) \rho_0]g(L - H_d)
\end{align*}
\]

(7)

where $\rho_w$ is the water density ($\text{t}/\text{m}^3$), $\rho_0$ is the crude oil density ($\text{t}/\text{m}^3$), $p_c$ is the casing pressure (MPa), $p_o$ is the oil pressure (MPa) and $H_d$ is the working fluid level.

When the oil well is producing under stable working conditions, the flow rate from the reservoir into the wellbore is equal to the actual production flow rate of the oil well, which is to meet the oil well production coordination conditions as follows.

\[q_{fj}^{IN} = Q_j\]  

(8)

In addition, the relation between the bottom hole flow pressure and the submerging pressure is given as follows:

\[p_{jwf} = p_{sj} + (L_Z - L_j) \rho_j g\]  

(9)

where $L_Z$ is the central reservoir depth (m), $L_j$ is the pump depth (m) of the $j$-th rod-pumped well, $\rho_j$ is the density of oil water mixture ($\text{t}/\text{m}^3$). So far, equation (9) can be substituted into equation (8) to determine the nonlinear relationship between flow pressure, submerging pressure, reservoir characteristics, crude oil physical properties and suction parameters. When other relevant production parameters are determined and measured, the flowing pressure and the submerging pressure can be determined according to different strokes and stroke times. Therefore, the actual oil flow rate $Q_j$ can be obtained, and finally the overall oil production can be calculated as shown in equation (1).

**Overall water production.** Overall water production of the block is expressed as follows:

\[Q_w = \sum_{j=1}^{N} Q_{jw0j}\]  

(10)

where $Q_w$ is the overall water production ($\text{m}^3/\text{d}$).

**Comprehensive energy consumption per ton of oil.** The comprehensive energy consumption per ton of oil is the energy consumption for producing one ton of crude oil, described as follows:

\[y = [Q_2 Q_{DW} + (W_1 + W_2)R]/R_1\]  

(11)
where \( y \) is the comprehensive energy consumption per ton of oil (kg/t), \( Q_{DW} \), \( R \) and \( R_1 \) are the low calorific value of fuel (kJ/m³), the electrical energy conversion coefficient (kJ/kW.h) and the equal value to the heat of per kilogram standard coal (kJ/kg), respectively. \( W_1 \) represents the power consumption in the sucker-rod pumping system (kW.h/t), \( W_2 \) and \( Q_g \) represent the power consumption (kW.h/t) and the natural gas consumption (m³/t) in the gathering system, respectively. \( W_1, W_2 \) and \( Q_g \) can be calculated by the following equations.

\[
\begin{align*}
W_1 &= \sum_{i=1}^{N} W_{1i}/Q_0 \\
W_2 &= Q_p t/(\eta_p R Q_0) \\
Q_g &= Q_T t/(\eta_T Q_{DW} Q_0)
\end{align*}
\]

where \( W_{1j} = P_j t \) is the electricity consumption of the \( j \)-th oil well (kW.h), \( P_j \) is the input power of suction system (kW), \( t \) is the operation time of the corresponding oil well in one day (h), \( P_j \) is a function of suction parameters when the management level and the technical equipment are invariant (Dong et al., 2010). \( Q_p \) and \( Q_T \) are the total pressure loss (kJ/h) and heat loss (kJ/h) of gathering pipelines, respectively. And they can be calculated by the related equations (Dong et al., 2017; Liu et al., 1999). When reservoir characteristics, crude oil physical property, technical equipment and management level are certain, \( Q_T \) and \( Q_p \) can be considered as the function of \( Q_j \), wellhead heating temperature \( T_{wj} \) (°C) and pump head \( H_j \) (m), \( \eta_T \) and \( \eta_p \) are the equipment efficiency.

**Multi-objective optimization model**

According to references (Dong et al., 2008a; Dong et al., 2008b; Dong et al., 2017; He et al., 2007; Liu et al., 1999; Zheng and Deng, 2007), the decision variables of the optimization model are selected as follows:

\[
X = \{ S, n, T_w, H \}
\]

where \( S = [S_1, S_2, \ldots, S_N] = [n_1, n_2, \ldots, n_N] \), \( T_w = [T_{w1}, T_{w2}, \ldots, T_{wN}] \) and \( = [H_1, H_2, \ldots, H_N] \) represent the stroke, the stroke times, the wellhead heating temperature and the pump head vector of the rod-pumped wells, respectively. \( N \) is the number of rod-pumped wells. So, a multi-objective optimization model of oil-gas production process is established, which is composed of the decision variables in equation (13) and the objective functions described by equations (1), (10), (11), under the corresponding constraints, it is formulated as follows:

\[
\begin{align*}
\{ \max Q_0, \min Q_w, \min y \} \\
\text{s.t.} \\
S_{\min} \leq S_j \leq S_{\max} \\
n_{\min} \leq n_j \leq n_{\max} \\
T_{w\min} \leq T_{wj} \leq T_{w\max} \\
H_{\min} \leq H_j \leq H_{\max} \\
T_j \geq T_{oil}, \ T_o \leq T_{omax} \\
p_{qi} \leq [p_o], \ p_z \geq [p_t]
\end{align*}
\]

\[ (14) \]
where $S_{j_{\text{min}}}$ and $S_{j_{\text{max}}}$ are the minimum and maximum values of the $j$-th oil well stroke. $n_{j_{\text{min}}}$ and $n_{j_{\text{max}}}$ are the minimum and maximum values of the stroke times. $T_{w_{\text{min}}}$ and $T_{w_{\text{max}}}$ are the minimum and maximum values of heating temperature, respectively. $H_{\text{min}}$ and $H_{\text{max}}$ are the minimum and maximum limits of pump head vector, respectively. $T_j$ is the temperature of the oil entering the oil transfer station (production platform) from the oil gathering pipeline and $T_{oil}$ is the maximum allowable value of $T_j$. $T_o$ ($^\circ\text{C}$) is the temperature of oil flowing from station and $T_{o_{\text{max}}}$ ($^\circ\text{C}$) is the maximum allowable value of $T_o$. $p_{qi}$ (Pa) is the wellhead back pressure and $p_{zi}$ (Pa) is the pressure of oil getting into station and $p_{zi_{\text{max}}}$ (Pa) is the minimum allowable value of $p_z$.

**Error compensation of comprehensive energy consumption per ton of oil**

*Hybrid model structure based on error compensation.* According to the optimization model in equation (14), the expression of objective function $y$ is complex. Because the mechanism model of the whole oil-gas production process is established under some assumptions and simplifications, there may be a large error between the output of the mechanism model and the actual value. In order to reduce the error, we propose a hybrid model based on the mechanism model and the LS-SVM error model, its structure is shown in Figure 2.

In this hybrid model, the mechanism model is used to describe the overall characteristics of oil-gas production process, and then a LS-SVM error model based on Bayesian optimization algorithm is established to compensate the model error which can’t be described by the mechanism model, the output of hybrid model is as follows:

$$y_c = y + \hat{e}$$

where $y_c$ is the prediction output of hybrid model, $y$ is the output of the mechanism model for the comprehensive energy consumption per ton of oil, and $\hat{e}$ is the prediction error between the output of the mechanism model and the actual value.

Through the analysis of oil-gas production process, the input vector $X$ is composed of the stroke and the stroke times of each oil well, oil heating temperature and pump head constitute, and the comprehensive energy consumption per ton of oil is taken as the output of the model.

**LS-SVM error model based on Bayesian optimization algorithm.** For the convenience of description, we select $\{x_i, e_i\}_{i=1}^L$ as the sample set, where $x_i = X$, $e_i = y_{\text{Actual}} - y$, $L$ is the number of samples. The optimization problem of LS-SVM error model is described as follows (Chen et al., 1999):

$$\min_{w, b, \xi} J(w, \xi) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^L \xi_i^2 $$

s.t. $e_i = \langle w, \varphi(x_i) \rangle + b + \xi_i$ $i = 1, 2, \cdots, L$

where $w$ is the weight vector, $\gamma$ is the regularization parameter (also known as the penalty factor), $\xi_i$ is the error variables, $b$ is the bias term. To solve this optimization problem, Lagrange function is constructed as follows:

$$L(w, b; \xi; a) = J(w, b, \xi) - \sum_{i=1}^L a_i \{ \langle w, \varphi(x_i) \rangle + b + \xi_i - e_i \}$$

(17)
where $\alpha_i$ are Lagrange multipliers. The solution of equation (17) can be determined by partial differentiation to $w$, $b$, $\zeta_i$ and $\alpha_i$:

\[
\begin{align*}
\frac{\partial L}{\partial w} &= 0 \rightarrow w = \sum_{i=1}^{L} \alpha_i \phi(x_i) \\
\frac{\partial L}{\partial b} &= 0 \rightarrow \sum_{i=1}^{L} \alpha_i = 0 \\
\frac{\partial L}{\partial \zeta_i} &= 0 \rightarrow \alpha_i = \gamma \zeta_i, \quad i = 1, 2, \ldots, L \\
\frac{\partial L}{\partial \alpha_i} &= 0 \rightarrow \langle w, \phi(x_i) \rangle + b + \zeta_i - e_i = 0, \quad i = 1, 2, \ldots, L
\end{align*}
\]

After removing the variables $w$ and $\zeta_i$, the equation is obtained as follows:

\[
\begin{bmatrix}
0 \\
Z \\
\Omega + 1/\gamma
\end{bmatrix} \begin{bmatrix}
b \\
\alpha \\
\zeta
\end{bmatrix} = \begin{bmatrix}
0 \\
e
\end{bmatrix}
\]

(19)

where $e = [e_1, e_2, \ldots, e_L]^T$, $Z = [1, 1, \ldots, 1]$, $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_L]^T$ and $\Omega = K(x_i, x_j)$. Because there are various forms of kernel function $K(x, x_i)$, the most widely used Gaussian function is selected in this paper.

\[
K(x, x_i) = \exp\left(-\frac{||x - x_i||^2}{2\sigma^2}\right)
\]

(20)

where $\sigma$ is kernel parameter, it denotes the width of the RBF which controls the regression ability.
So far, the final expression of error estimation model of comprehensive energy consumption error is obtained as follows:

\[
\hat{e} = \sum_{i=1}^{L} a_i K(x, x_i) + b
\]  

(21)

During the training process of LS-SVM error model, the selection of penalty factor \( c \) and kernel function parameter \( r \) are important to improve the accuracy and generalization ability of the model. In order to reduce the influence of uncertainty in the actual production process on the model, and further improve the accuracy and stability of the model, a Bayesian optimization algorithm (BOA) (Shahriari et al., 2016) is used to obtain the optimal values of \( \gamma \) and \( \sigma \) by maximizing the posterior probability of the hyperparameters. Bayesian inference is constructed with three levels of inference (Xie et al., 2019):

1. First level of inference: Let \( H \) as the model space, \( D \) as the sample training set, for a given hyperparameter \( \lambda = 1/\gamma \), if the training samples are independent and identically distributed, and the prior distribution \( p(w|\lambda, H) \) of the parameter \( w \) obeys the normal distribution, then the posterior probability of \( w \) can be obtained as follows:

\[
p(w|D, \lambda, H) \propto \exp\left( -\frac{1}{2} w^T w - \sum_{i=1}^{L} e_i^2 \right) \]  

(22)

where \( e_i \) is the regression error of each training sample, the optimal value \( w_{best} \) is obtained by maximizing equation (22).

2. Second level of inference: Let \( E_1 = \frac{1}{2} w^T w \), \( E_2 = \sum_{i=1}^{L} e_i^2 \), the posterior probability of hyperparameter \( \lambda \) is shown as follows:

\[
p(\lambda|D, H) \propto \frac{(2\pi\lambda)^{k/2}}{(\text{det}A)^{1/2}} \exp\left( -\lambda E_{best}^1 - E_{best}^2 \right) \]  

(23)

where \( A = \lambda \nabla^2 E_1 + \nabla^2 E_2 \), \( E_{best}^1 \) and \( E_{best}^2 \) are the values of \( E_1 \) and \( E_2 \) at \( w_{best} \), respectively. \( \nabla \) is the vector differential operator, and then the optimal value \( \lambda_{best} \) is obtained by maximizing \( \ln p(\lambda|D, H) \).

3. Third level of inference: Assuming that the prior probability of all sample data \( p(H) \) is a flat distribution, and then the posterior probability \( p(H|D) \) of parameter \( \sigma \) can be obtained as follows

\[
p(H|D) \propto \frac{p(D|\lambda_{best}, H)}{\sigma} \]  

(24)
Then, the optimal kernel parameter $\sigma_{\text{best}}$ can be calculated by maximizing $\ln p(H|D)$.

A multi-objective optimization model based on error compensation of comprehensive energy consumption is constructed by replacing $y$ in equation (14) with the output of hybrid model $y_c$ in equation (15). Other objective function models can also be compensated by the same method, and then a multi-objective optimization model based on error compensation is built, that is error compensation multi-objective optimization model.

**Optimization solution of error compensation multi-objective optimization model using IMS-NSGA-II algorithm**

**NSGA-II algorithm with multi-strategy improvement (IMS-NSGA-II)**

Through the above analysis, the multi-objective optimization model of oil-gas production process is a typical complex multi-objective optimization problem with multiple variables and multiple constraints. At present, the main methods to solve this kind of multi-objective optimization problems are multi-objective evolutionary algorithms (MOEAs), including dominant relation-based MOEAs, index-based MOEAs and decomposition-based MOEAs. The basic idea of dominant relation-based MOEAs is to find all non-dominant individuals from the current population by using Pareto-based fitness allocation strategy, such as non-dominated sorting genetic algorithm-II (NSGA-II) (Deb et al., 2002), Strength Pareto Evolutionary Algorithm2 (SPEA2) (Zitzler et al., 2001) and Pareto Envelope-based Selection Algorithm (PESA-II) (Corne et al., 2001). NSGA-II is one of the most popular evolutionary algorithms, it is an improved version of “NSGA”, its main feature is the fast non-dominated sorting for ranking solutions at its selection stage, and an effective elite strategy is adopted. Compared with previous evolutionary algorithms, NSGA-II has lower computational complexity and better convergence (Noori-Darvish and Tavakkoli-Moghaddam, 2012). NSGA-II has been successfully applied to many practical complex projects (Maity et al., 2018; Wang et al., 2017; Yang et al., 2018). In recent years, many scholars have introduced the differential evolution algorithm (Rainer and Price, 1997) into the field of multi-objective problems, forming a multi-objective differential evolution (MODE). Some scholars have also integrated DE algorithm into NSGA-II, using crossover and mutation operators instead of simulated binary crossover operators to improve the convergence performance. However, many traditional methods do not consider how to maintain the diversity of Pareto optimal solutions, and are prone to converge prematurely, when the control parameters are not set properly, the algorithm is easy to fall into local optimum in the late evolution period (Guo et al., 2012). In addition, the traditional DE algorithm cannot use the gradient information to further explore the local optimal solution after obtaining the optimal solution of a certain generation of population. Therefore, through the above analysis, based on the NSGA algorithm, this paper introduces the chaos idea to avoid falling into local optimum, and uses the gradient information near the optimal solution to further improve the convergence ability of the algorithm, and proposes an improved NSGA-II algorithm with multi-strategy improvement (IMS-NSGA-II). The specific ideas are as follows: In IMS-NSGA-II, a chaotic sequence is first introduced to initialize the population to maintain the diversity of the initial population. Then, the gradient operator is introduced into the crossover operator and mutation operator to improve the local search speed of NSGA-II algorithm. Finally, during the evolutionary process, when an improved population is selected, the cyclic crowding distance is adopted to
control the distribution of Pareto non-dominated solutions, so as to further improve the diversity and uniformity of the Pareto non-dominated solutions. Through these improved operations, IMS-NSGA-II algorithm not only guarantees the uniformity of initial population and the diversity of Pareto non-dominated solutions, but also improves the convergence of the algorithm to a great extent.

**Population initialization based on chaotic sequence.** Population initialization plays an important role in the accuracy and diversity of NSGA-II algorithm. In the standard NSGA-II algorithm, the uniformly distributed random number generator is usually used for crossover and mutation operations, but there are still some shortcomings in maintaining horizontal diversity and obtaining the Pareto-front with high uniformity (Guo et al., 2010). Chaotic motion has the characteristics of regularity, randomness and ergodicity. This process is not convergent, but bounded, and it is extremely sensitive to external parameters and initial values. Therefore, the population initialization based on chaotic sequence can traverse the search space within a certain range according to its own law, which significantly improves the accuracy and diversity of the solution obtained by the algorithm (Chen et al., 2014; Lu et al., 2013). Therefore, chaotic initialization is first introduced to enrich the diversity of the initial population, thus laying the foundation for global optimization.

In this paper, the chaotic sequence generated by Skew Tent mapping is used for population initialization (Hasler and Maistrenko, 1997; Long et al., 2019), and its mathematical model is expressed as follows:

\[
\begin{align*}
  x_{k+1} &= \frac{x_k}{\varphi}, \quad 0 \leq x_k \leq \varphi \\
  x_{k+1} &= \frac{(1-x_k)}{(1-\varphi)}, \quad \varphi < r_k \leq 1
\end{align*}
\]  

(25)

When \( \varphi \in (0,1) \) and \( x_k \in [0,1] \), equation (25) is in chaotic state. In equation (25), if \( x_k \) is the \( j \)-dimensional variable, the generated chaotic variables are mapped to the decision variables, and the \( j \)-th decision variable \( x_{ji} \) of the \( i \)-th initialized individual \( X_i^0 = (x_{i1}^0, \cdots, x_{ji}^0, \cdots, x_{in}^0) \) is obtained as follows:

\[
\begin{align*}
  x_{ji} &= x_{j_{\text{min}}} + x_{j_{k}}(x_{j_{\text{max}}} - x_{j_{\text{min}}}) \\
  j &= 1, 2, \cdots, m, \quad i = 1, 2, \cdots, N_p
\end{align*}
\]  

(26)

where \( x_{j_{\text{max}}} \) and \( x_{j_{\text{min}}} \) are the upper and lower limits of the \( j \)-th decision variable, respectively. \( m \) is the dimension of decision variable vector, \( N_p \) is the population size.

**The gradient-based operator.** As a random search algorithm, the convergence speed of evolutionary algorithm cannot be ignored. Since the negative gradient direction is the fastest direction to reduce the value of the objective function, in order to improve the optimization speed, we introduce the gradient operator into NSGA-II algorithm, and combine it with the crossover and mutation operator to form the comprehensive operator (Yu et al., 2011), and then use the comprehensive operator to generate a new generation of population, expressed as follows:

\[
\omega_1^* \text{operator}_{c}^* \oplus \omega_2^* \text{operator}_{m}^* \oplus \omega_3^* \text{operator}_{g}^*
\]  

(27)
where $\omega_i^* \geq 0$, $i = 1, 2, 3$, equation (27) means that the new solutions are generated by the crossover operator, the mutation operator and the gradient-based operator with the probability $\omega_1^*$, $\omega_2^*$ and $\omega_3^*$ respectively. For gradient operator, in practical engineering problems, it is often impossible to get the analytical expressions directly, so it is necessary to estimate the gradient of function. The standard approach of gradient estimation is the finite difference (FD) method, i.e., $f(x) = [f_1(x), f_2(x), \ldots, f_m(x)]^T, x \in \mathbb{R}^n$ represents a multi-objective function, $x = (x_1, x_2, \ldots, x_n)$ is a $n$-dimensional decision vector, the one-sided FD method used for the $j$-th objective gradient estimation requires $(n + 1)$ function evaluations, as follows:

$$g_{j}(x) = \frac{f_j(x + ce_i) - f_j(x)}{c}$$ (28)

where $e_i$ represents the unit vector in the $i$-th direction, and $c$ is the step size.

FD method is usually expensive in function evaluation. In contrast, the calculation time of Simultaneous Perturbation (SP) method is less (Zhou, 2017), so the SP method is used in the paper to estimate the gradient of the $j$-th objective function, and the $i$-th component of gradient is expressed as follows:

$$g_{ji}(x) = \frac{f_j(x + c\Delta) - f_j(x)}{c\Delta_i}$$ (29)

where $\Delta$ is a $n$-dimensional vector of random perturbations, and $\Delta_i$ is the $i$-th component of $\Delta$. For each component of $\Delta$, a simple choice is used to have a Bernoulli distribution $\pm 1$ and the probability of each $\pm 1$ outcome is 0.5. The step size $c$ at the $k$th iteration (denoted by $c_k$) is given as follows: $c_k = c_0 / (k + 1)^\gamma$, the values of $c_0$ and $\gamma$ are set to 0.001 and 1/6, respectively. Then the search direction can be obtained by using the gradient operator in equation (30).

$$\begin{align*}
\bar{e} &= d/\|d\| \\
d &= -\sum_{j=1}^{m} \lambda_j \frac{g_j(x)}{\|g_j(x)\|}
\end{align*}$$ (30)

where the search direction $\bar{e}$ is the normalized vector of $d$, and the convex cone combination of negative gradient direction of each objective function, which makes each selected point moves to the Pareto front along a common descent direction of some objectives, $\lambda_j (j = 1, 2, \cdots, m) \in (0,1)$. Obviously, the direction $d$ in equation (30) has a higher chance to minimize all the objectives (or at least one objective) than the variation direction in the whole space by the pure random genetic operator, which can also reduce the invalid trial times of crossover and mutation, and improve the local search ability.

After the $\bar{e}$ is determined, a new solution $x_{k+1}$ can be created by the gradient-based operator.

$$x_{k+1} = x_k + t_k \bar{e}$$ (31)

where $t_k$ is the step size of the $k$-th generation.
Circular crowded sorting strategy. In order to further improve the diversity of Pareto non-dominated solutions and overcome the shortcomings of traditional crowded sorting, and to make the algorithm can jump out of local optimum and prevent premature convergence when solving complex multi-objective optimization problems. In this paper, a circulating crowded ranking algorithm is adopted (Luo et al., 2010).

The specific steps are as follows:

- Step 1: Calculate the crowding distance of each individual (solution), and delete the individual with the smallest crowding distance in the population.
- Step 2: Among the remaining individuals, recalculate the crowding distance, and continue to delete the individuals with the smallest crowding distance in the population.
- Step 3: Determine the number of remaining solutions \( N' \). If \( N' < N_p \), return to Step 2, otherwise, skip to Step 4.
- Step 4: End the selection.

Optimization solution process of error compensation multi-objective optimization model using IMS-NSGA-II algorithm

The design variables in the multi-objective optimization model of oil-gas production process include discrete variables and continuous variables. Here, we use binary encoding for discrete variables \( S \) and real encoding for continuous variables \( n, T_w \) and \( H \). The flow chart of proposed IMS-NSGA-II algorithm for solving the multi-objective optimization problem of oil-gas production process is shown in Figure 3.

The specific steps of optimization solution of error compensation multi-objective optimization model using IMS-NSGA-II algorithm are as follows:

- Step 1: The measured data in actual production and the data generated by mechanism model of the comprehensive energy consumption per ton of oil are collected to establish the hybrid model of comprehensive energy consumption per ton of oil.
- Step 2: Hybrid model of comprehensive energy consumption per ton of oil is introduced into the multi-objective optimization model of oil-gas production process, and then the multi-objective optimization model based on error compensation is built, that is the error compensation multi-objective optimization model.
- Step 3: IMS-NSGA-II algorithm is put forward for solving the error compensation multi-objective optimization model established in Step 2.
- Step 4: The values of manipulated variables can be adjusted according to the optimization results obtained in Step 3.

Simulation results and analysis

Firstly, the production process with six oil wells in an oil recovery operation area is taken as background to verify the effectiveness of the established optimization model and the proposed IMS-NSGA-II algorithm. The basic data of oil well pumping equipment is shown in Table 1, where pumping unit model CYJ10-3-53HB and motor model Y250M-6 are used for oil wells numbers 11–22, 11–17, 11–25 and 11–20, and pumping
unit model CYJ14-5-73HB and motor model YQ280M2-8 are used for wells numbers
11–23 and 11–27.

Besides the basic data of the oil well, some production data, such as reservoir character-
istic parameters, wellhead parameters and physical characteristic parameters of crude oil
and natural gas, are also needed for model verification (Dong et al., 2010). According to the
actual situation, some production parameters are fitted or selected as follows:

\[ B = 1.21 \text{ m}^3/\text{m}^3, \quad \mu = 10 \text{ mPa.s}, \quad k_c = 20 \text{ md}, \quad h = 23 \text{ m}, \quad c_i = 8.32 \times 10^{-6} \text{ psi}, \quad r_R = 0.5 \text{ m}, \quad \phi = 0.21. \]

oil reservoir pressure \( p_r = 15.3 \text{ Mpa}, \) reservoir temperature \( T_L = 50.23 \text{°C}, \) saturation pressure of crude oil \( p_{sat} = 10.5 \text{ Mpa}, \) water density \( \rho_w = 1 \text{ t/m}^3, \) crude oil density \( \rho_0 = 0.863 \text{ t/m}^3, \) crude oil viscosity \( \mu_0 = 2.1 \text{ mPa.s}, \) water viscosity \( \mu_w = 1.005 \text{ mPa.s}, \) natural gas relative density \( \delta_{gs} = 0.75, \) natural
gas polytropic process index \( k = 1.1. \)

For the maximum liquid production of oil well \( q_{max}, \) under the condition of known reservoir pressure and saturation pressure, if a certain liquid
production of oil well \( q_0 \) and its corresponding flow pressure \( p_{wf} \) are measured, they are
substituted into equation (2) or equation (3) to obtain the corresponding maximum liquid
production of the oil well.

Figure 3. Flow chart of optimization solution using IMS-NSGA-II algorithm.

Table 1. Basic data of rod pumping equipment.

| Oil well number | Central reservoir depth (m) | Casing diameter (mm) | Oil tube diameter (mm) | Pump diameter (mm) | Sucker rod string | Pump depth (m) |
|-----------------|-----------------------------|----------------------|------------------------|-------------------|------------------|----------------|
| 11-22           | 2167.3                      | 127                  | 89                     | 44                | \( 22 \times 436.4 + 19 \times 1363.0 \) | 1899.4         |
| 11-17           | 2198.3                      | 139.7                | 73                     | 56                | \( 22 \times 703.6 + 19 \times 1247.16 \) | 1950.76        |
| 11-25           | 2236.8                      | 124                  | 62                     | 38                | \( 22 \times 795.9 + 19 \times 1354.19 \) | 2150.09        |
| 11-23           | 2245.2                      | 127                  | 89                     | 44                | \( 22 \times 710.1 + 19 \times 1390.62 \) | 2100.72        |
| 11-27           | 2183.39                     | 127                  | 89                     | 44                | \( 22 \times 725.6 + 19 \times 1257.63 \) | 1983.23        |
| 11-20           | 2175.5                      | 124                  | 62                     | 38                | \( 22 \times 746.8 + 19 \times 1253.32 \) | 2000.12        |
In oil well production, some production data such as oil pressure $p_o$, casing pressure $p_c$ and the water content, are not constant and can be measured every day. Table 2 shows the corresponding production data of each oil well in the block on a certain day. As the water content will not change greatly in a short term and only fluctuate in a small range, the water content can be regarded as a constant in the simulation experiment.

According to the basic parameters of pumping units, oil pipelines and the actual production requirements, the parameters in constraints are set as follows: The stroke constraints of rod pumped well are $s_{j_{\text{max}}} = 6$ m and $s_{j_{\text{min}}} = 1$ m, respectively. The stroke times constraints of rod pumped well are $n_{j_{\text{max}}} = 12$ min$^{-1}$ and $n_{j_{\text{min}}} = 2$ min$^{-1}$, respectively. The allowable temperature of crude oil getting into station $T_{\text{oil}} = 35$ °C and the allowable temperature of crude oil out of station $T_{\text{omax}} = 80$ °C, the pressure of oil gathering pipeline getting into the oil transfer station and the processing station is 0.2 MPa.

Parameters of IMS-NSGA-II are set as follows: the population size $N_p = 100$, the maximum evolutionary generation $G_{\text{max}} = 150$, distribution indexes for real encoding crossover operator and mutation operator of $mu = 20$ and $mum = 20$, crossover probability $\omega^*_1 = 0.8$, mutation probability $\omega^*_2 = 1/n$ ( $n$ is the dimension of decision vector), the probability of gradient-based operator $\omega^*_3 = 1$, $\varphi = 0.5$, the step size $t_k = 0.1(gen-k+1)/gen$, $gen = 100$.

Firstly, 200 groups of data are collected from the actual production process, and the data are generated by the mechanism model of comprehensive energy consumption. Among them, 150 groups are taken as training samples and 50 groups are taken as testing samples. During training LS-SVM error model, the Bayesian optimization is applied to obtain the optimal values of ( $\gamma$, $\sigma$ ), the values of $\gamma$ and $\sigma$ are determined to be 31.26 and 25.37, respectively. Then, the trained LS-SVM error model is used to compensate the error of mechanism model, and the hybrid model is compared with the mechanism model, the comparison results are shown in Figure 4. The performance comparison of the two models is shown in Table 3.

From Table 3, it can be seen that the average absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) of the hybrid model are smaller than those of the mechanism model.

IMS-NSGA-II algorithm is used to solve the multi-objective optimization models before and after error compensation. In order to further validate the reliability of optimization results, the manipulated variables are adjusted according to the optimization results, and the corresponding actual values of objectives are obtained. Part of Pareto non-dominated solutions are listed in Table 4, where $y$ represents the comprehensive energy consumption per ton of oil.

| Oil well number | casing pressure $p_c$ (MPa) | oil pressure $p_o$ (MPa) | Water content (%) |
|-----------------|-----------------------------|--------------------------|-------------------|
| 11-22           | 2.3                         | 0.3                      | 69                |
| 11-17           | 2.5                         | 0.25                     | 76.4              |
| 11-25           | 5.9                         | 0.4                      | 76                |
| 11-23           | 8.3                         | 0.3                      | 82                |
| 11-27           | 2.9                         | 0.3                      | 73                |
| 11-20           | 3.8                         | 0.2                      | 74.5              |
From the comparison results in Table 4, it can be seen that the corresponding actual values obtained by solving error compensation multi-objective optimization model are better than those obtained by solving multi-objective optimization model before error compensation, such as solution 1, the actual value (40.7745, 135.1560, 74.3573) is better than the value (40.1320, 135.9883, 77.6981). It shows that the method not only improves the overall oil production rate, but also reduces the comprehensive energy consumption and the overall water production to a certain extent. Moreover, due to the higher prediction accuracy of the error compensation multi-objective optimization model, the optimization results obtained by the multi-objective optimization model based on the error compensation are closer to the actual values than that obtained by the uncompensated model, which can reduce the impact of uncertainty and greatly improves the reliability of optimization results.

Then, in order to verify the superiority of the proposed algorithm, IMS-NSGA-II algorithm is tested on the standard test problems including ZDT1 and ZDT2 (Zitzler et al., 2000), and compared with other algorithms, such as the original NSGA-II algorithm, SPEA2 and hybrid multi-objective intelligent algorithm (C-NSGA-II) (Yan et al., 2018) for multi-objective optimization.

The parameters of these algorithms are set as follows: For these algorithms, the population size is 100, the dimension of decision variables of test problems ZDT1 and ZDT2 is $n = 30$. For IMS-NSGA-II algorithm, the parameters are set as above, and the other operations are consistent with those used in the relevant literatures (Deb et al., 2002; Yan et al., 2018; Zitzler et al., 2001).

In this paper, the hypervolume indicator $I_H$ (Zitzler et al., 2000) and spatial evaluation method (Deb et al., 2000) are applied to evaluate the performance of the algorithm. Hypervolume indicator $I_H$ is defined as follows: $I_H(A)$ of a solution set $A \subseteq X$ can be

---

### Table 3. Performance comparison of the two models.

| Model            | MAE     | RMSE    | MAPE   |
|------------------|---------|---------|--------|
| mechanism model  | 1.2561  | 2.0615  | 1.9202 |
| hybrid model     | 0.5108  | 0.8147  | 0.7926 |

---

Figure 4. Comparison of prediction results of comprehensive energy consumption per ton of oil.

(a) (b)
defined as the hypervolume of the space, which is dominated by the set $A$ and bounded by a reference point $r = (r_1, \ldots, r_m) \in \mathbb{R}^m$, expressed as follows:

$$IH(A) = k \left( \bigcup_{a \in A} [f_1(a), r_1] \times \ldots \times [f_m(a), r_m] \right)$$

(32)

where $\lambda(S)$ is the Lebesgue measure of a set $S$ and $[f_1(a), r_1] \times \ldots \times [f_m(a), r_m]$ is the $m$-dimensional hypercuboid composed of all points weakly dominated by the point $a$ rather than by the reference point. When the value of $IH(A)$ is higher, the convergence is better.

These algorithms are run on each test problem for 10 times, respectively, and each run selects a different initial population. The maximum number of function evaluations of all algorithms is 10000. The optimization results of a random run on ZDT1 and ZDT2 are shown in Figures 5 and 6, respectively. The corresponding hypervolume indicator $IH$ and the diversity metric $\Delta$ of these algorithms are listed in Tables 5 and 6, respectively. In order to show the significant difference between these algorithms, we also give the optimization results in Figures 5 and 6, they are obtained by these algorithms that the number of function evaluations is 5000.

From Figures 5 and 6 and Tables 5 and 6, it can be seen that the performance of IMS-NSGA-II on the problems: ZDT1 and ZDT2 is better than NSGA-II, SPEA2 and slightly better than C-NSGA-II in terms of the hypervolume indicator $IH$ and the diversity metric $\Delta$. The results also indicate that the Pareto non-dominated solutions obtained by IMS-NSGA-II have better convergence and diversity than other algorithms when these solutions are

| Pareto non-dominated solutions | Multi-objective optimization | $Q_0$ (m$^3$/d) | $Q_w$(m$^3$/d) | $y$(kg/t) |
|--------------------------------|-------------------------------|-----------------|----------------|-------------|
| 1 Before the error compensation | 38.5341                      | 134.2653        | 74.6722        |
| Actual value                    | 40.1320                      | 135.9883        | 77.6981        |
| After the error compensation    | 40.1914                      | 134.3697        | 73.5061        |
| Actual value                    | 40.7745                      | 135.1560        | 74.3573        |
| 2 Before the error compensation | 35.3816                      | 115.6281        | 67.2615        |
| Actual value                    | 33.9282                      | 113.6112        | 70.2185        |
| After the error compensation    | 33.7540                      | 111.5943        | 67.6088        |
| Actual value                    | 34.3298                      | 112.0596        | 68.3626        |
| 3 Before the error compensation | 28.6029                      | 90.7451         | 66.7500        |
| Actual value                    | 26.2670                      | 88.7538         | 64.0897        |
| After the error compensation    | 27.2575                      | 89.5439         | 62.9282        |
| Actual value                    | 26.7489                      | 88.8115         | 63.3762        |
| 4 Before the error compensation | 19.4133                      | 66.0509         | 58.9031        |
| Actual value                    | 17.5297                      | 68.1072         | 56.7124        |
| After the error compensation    | 18.1671                      | 65.8618         | 57.3927        |
| Actual value                    | 18.9598                      | 66.4186         | 56.6079        |
| 5 Before the error compensation | 11.7958                      | 43.3517         | 53.4215        |
| Actual value                    | 10.1160                      | 46.0211         | 56.4906        |
| After the error compensation    | 11.2061                      | 45.7571         | 53.7122        |
| Actual value                    | 10.4965                      | 45.1350         | 54.3072        |
distributed along the real Pareto-optimal set. In addition, when solving these problems, it not only doesn’t significantly increase the running time of these gradient-based operators, but also reduces the average running time used by IMS-NSGA-II.

Finally, both NSGA-II algorithm and IMS-NSGA-II algorithm are used to solve the error compensation multi-objective optimization model. The two algorithms have been run

| Algorithm     | Average $I_H$ | Average $\Delta$ | Average running time (min) |
|---------------|---------------|-------------------|----------------------------|
| NSGA-II       | 0.80215       | 0.42064           | 0.04819                    |
| SPEA2         | 0.80329       | 0.79345           | 0.34566                    |
| C-NSGA-II     | 0.81650       | 0.23498           | 0.03843                    |
| IMS-NSGA-II   | 0.82703       | 0.21523           | 0.02975                    |
10 times on the multi-objective optimization problem that the number of evolutionary generations is 100. The optimization results of a random run are shown in Figure 7. The corresponding hypervolume indicator $I_H$ and $\Delta$ of the two algorithms are listed in Tables 7 and 8, respectively.

It can be seen from Figure 7 that the Pareto non-dominated solution set obtained by IMS-NSGA-II algorithm are superior to that obtained by the traditional NSGA-II algorithm in terms of convergence and diversity. So, it is considered that IMS-NSGA-II algorithm can provide more superior solutions for decision-making departments in oil production enterprise, and the decision makers can choose more reasonable oil production programs to satisfy the different requirements of actual production.

From Table 7, it can be seen that each $I_H$ obtained by IMS-NSGA-II algorithm is larger than that obtained by NSGA-II algorithm. So, it further proves that the Pareto non-dominated solutions obtained by the IMS-NSGA-II algorithm is closer to the real Pareto front and has better convergence. In Table 8, it can be seen that $\Delta$ of the Pareto

### Table 6. The convergence metric, the diversity metric and average running time on ZDT2.

| Algorithm     | Average $I_H$ | Average $\Delta$ | Average running time (min) |
|---------------|---------------|------------------|----------------------------|
| NSGA-II       | 0.50256       | 0.49253          | 0.04896                    |
| SPEA2         | 0.50327       | 0.78326          | 0.36425                    |
| C-NSGA-II     | 0.52212       | 0.30126          | 0.04237                    |
| IMS-NSGA-II   | 0.53018       | 0.27250          | 0.03049                    |

![Figure 7. Distribution chart of Pareto set obtained by NSGA-II and IMS-NSGA-II algorithm.](image)
non-dominated solution set obtained by IMS-NSGA-II algorithm is smaller, which demonstrates that IMS-NSGA-II algorithm has a better diversity and uniformity. In addition, it can be seen from Tables 7 and 8 that the maximum value, the minimum value and the average value of hypervolume indicator $I_H$ obtained by IMS-NSGA-II algorithm for 10 runs are 0.5137, 0.4896 and 0.4990, respectively. The maximum value, the minimum value and the average value of $D$ obtained by IMS-NSGA-II algorithm are 19.9750, 17.2326 and 18.3082, respectively. And for each run, the values of $I_H$ and $D$ obtained by IMS-NSGA-II algorithm are better than that obtained by NSGA-II algorithm. Thus, all the results indicate that the improved algorithm is better than NSGA-II algorithm and has good robustness.

## Conclusions

In this paper, we propose an error compensation multi-objective optimization model for the integrated oil-gas production process. Then an improved NSGA-II (IMS-NSGA-II)
algorithm is presented to solve the optimization model. When IMS-NSGA-II algorithm is test on the problems: ZD1 and ZD2, the experimental results indicate the effectiveness and superiority of the proposed IMS-NSGA-II algorithm. Besides, the application results on oil-gas multi-objective optimization demonstrate that the proposed optimization method can achieve better production indices. It not only satisfies the requirements of the actual oil-gas production process, but also reduce the impact of uncertainty and improves the reliability of the optimization results. Thus, the proposed optimization method is a more effective tool for optimization and control of oil-gas production process. These simulation results also imply that the proposed optimization method has a good application protest in dealing with the optimization problems of other similar complex production processes.

Acknowledgments
The author(s) would like to thank the editor and the anonymous reviewers for their helpful comments and high-quality suggestions, we also thank for the support from the Liaohe Oilfield of China, providing us research and experimental conditions.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by National Natural Science Foundation of China (32001415, 31901400, 61673281, 61903264), China Postdoctoral Science Foundation (2018M631820), Liaoning Province Doctoral Research Initiation Fund Guidance Project (2019-BS-207), Liaoning province natural fund guidance plan project (2019-ZD-0720).

ORCID iD
Tan Liu https://orcid.org/0000-0003-1680-0355

References
Al-Fehdly H, ElMaraghy W and Wilkinson S (2019) Carbon footprint estimation for oil production: Iraq case study for the utilization of waste gas in generating electricity. *Procedia CIRP* 80: 389–392.
Chebouba A, Yalaoui F, Smati A, et al. (2009) Optimization of natural gas pipeline transportation using ant colony optimization. *Computers & Operations Research* 36(6): 1916–1923.
Chen Y, Wen C, Gong Z, et al. (1999) Realization of robot motion based on a learning method. *IEEE Transactions on Automation Control* 44: 271–376.
Chen Z, Yuan X, Ji B, et al. (2014) Design of a fractional order PID controller for hydraulic turbine regulating system using chaotic non-dominated sorting genetic algorithm II. *Energy Conversion and Management* 84: 390–404.
Cheng QL, Li ZD, Song DM, et al. (2020) Study on weak link of energy utilization in oil transfer station system: Insights from energy level analysis method. *Energy Reports* 6: 1097–1105.
Codas A, Campos S, Camponogara E, et al. (2012) Integrated production optimization of oil fields with pressure and routing constraints: The Urucu field. *Computers & Chemical Engineering* 46: 178–189.
Corne DW, Jerram NR, Knowles JD, et al. (2001) PESA-II: region-based selection in evolutionary multiobjective optimization. In: Spector L, Goodman ED, Wu A, et al. (eds) *Proceedings of the
genetic and evolutionary computation conference (GECCO’2001). San Francisco, CA: Morgan Kaufmann Publishers, pp.283–290.

Cui ZJ, Wang Q, Zhang XY, et al. (2008) Optimization of the operation efficiency of oil-gas gathering pipeline network based on genetic algorithm. Journal of Xi’an Shiyou University (Natural Science Edition) 23(6): 59–62.

Deb K, Agrawal S, Pratap A, et al. (2000) A fast-elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In: 6th International conference on parallel problem solving from nature, Paris, France. 18-20 September 2000, pp.849–858. Germany: Springer Verlag.

Deb K, Pratap A, Agarwal S, et al. (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2): 182–197.

Dong SM, Li HQ, Yan XM, et al. (2008a) Simulating maximum of system efficiency of rod pumping wells. Journal of System Simulation 20: 3533–3537.

Dong SM, Wang HB, Sun XR, et al. (2017) Oil gathering parameters optimization method based on the lowest energy consumption of machine mining and oil gathering system. Journal of China University of Petroleum 41(5): 123–129.

Dong SM, Wang SJ, Lu DF, et al. (2008b) Simulation models for optimization design of suction parameters for rod pumping system in directional wells. Acta Petrolei Sinica 29: 120–123.

Dong SM, Zhang XS, Wu CJ, et al. (2010) Simulation and optimization method for the integral energy-saving suction parameters of rod pumping wells in an oil field. Acta Petrolei Sinica 31: 475–479.

Foss B (2012) Process control in conventional oil and gas fields – Challenges and opportunities. Control Engineering Practice 20(10): 1058–1064.

Guo D, Wang JP, Huang J, et al. (2010) Chaotic-NSGA-II: An effective algorithm to solve multi-objective optimization problems. In: IEEE international conference on intelligent computing and integrated systems (ICISS), Guilin, China, 22-24 October 2010, pp.20–23. USA: IEEE.

Guo J, Gui WH, and Yang CH (2012) An improved hybrid differential evolution algorithm used for multi-objective optimization of aluminum electrolysis. Journal of Central South University (Science and Technology) 43(1): 185–188.

Hasler M and Maistrenko YL (1997) An introduction to the synchronization of chaotic systems: Coupled skew tent maps. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications 44(10): 856–866.

He YF, Zhu RH and Xu YC (2007) Gathering and transferring parameter optimization of 401 transfer station of Xingliu joint station. Journal of Daqing Petroleum Institute 31: 34–37.

Horne RN (1998) Modern Well Test Analysis. 2nd ed. Palo Also: Petroway Inc.

Jahanshahi E, Krishnamoorthy D, Es Codas A, et al. (2020) Plantwide control of an oil production network. Computers and Chemical Engineering 136: 1–12.

Jun JY, Kang JY, Jeong D and Lee H (2017) An efficient approach for optimizing full field development plan using Monte-Carlo simulation coupled with genetic algorithm and new variable setting method for well placement applied to gas condensate field in Vietnam. Energy Exploration & Exploitation 35(1): 75–102.

Kosmidis VD, Perkins JD and Pistikopoulos EN (2005) A strategy for simulation and optimization of gas and oil production. Computers & Chemical Engineering 29(7): 1523–1541.

Lei Q, Weng DW, Luo JH, et al. (2019) Achievements and future work of oil and gas production engineering of CNPC. Petroleum Exploration and Development 46(1): 145–152.

Liu Y and Chen GR (1999) Optimal parameters design of oilfield surface pipeline systems using fuzzy models. Information Sciences 120(1–4): 13–21.

Liu Y, Zhao HJ and Wang P (1999) An optimizing design and analysis of oil gas gathering of ring type mixing hot water. Acta Petrolei Sinica 20: 77–81.

Long W, Cai SH, Jiao JJ, et al. (2019) An improved grey wolf optimization algorithm. Acta Electronica Sinica 47(1): 169–175.
Lu H, Niu R, Liu J, et al. (2013) A chaotic non-dominated sorting genetic algorithm for the multi-objective automatic test task scheduling problem. *Applied Soft Computing* 13(5): 2790–2802.

Luo CY, Chen MY and Zhang CY (2010) Improved NSGA-II algorithm with circular crowded sorting. *Control and Decision* 25(2): 227–231.

McCoy JN and Podio AL (2001) Rotaflex efficiency and balancing. In: *Proceedings of SPE production and operations symposium*, Oklahoma City, Oklahoma, USA, 24-27 March 2001, pp.453–461. USA: Society of Petroleum Engineers.

Maity S, Roy A and Maiti M (2018) A rough multi-objective genetic algorithm for uncertain constrained multi-objective solid travelling salesman problem. *Granular Computing* 46: 1–18.

Noori-Darvish S and Tavakkoli-Moghaddam R (2012) Minimizing the total tardiness and make span in an open shop scheduling problem with sequence-dependent setup times. *Journal of Industrial Engineering International* 8(1): 1–13.

Patton D and Golland M (1980) Generalized IPR curves for predicting well behavior. *Petroleum Engineering Institute* 52(7): 74–82.

Rainer S and Price K (1997) Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* 11(4): 341–359.

Rui Z, Wang X, Zhang Z, et al. (2018) A realistic and integrated model for evaluating oil sands development with steam assisted gravity drainage technology in Canada. *Applied Energy* 213: 76–91.

Shahriari B, Swersky K, Wang ZY, et al. (2016) Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE* 104(1): 149–175.

Tan CD, Xiang Y, Zhao XM, et al. (2016) Energy consumption prediction and application in oil and gas gathering and transferring system production based on large data. *Acta Petrolei Sinica* 37(S2): 158–164.

Wang H, Fu Y, Huang M, et al. (2017) A NSGA-II based memetic algorithm for multiobjective parallel flowshop scheduling problem. *Computers & Industrial Engineering* 113: 185–194.

Wang HB, Dong SM, Gan QM, et al. (2018) Method for improving system efficiency by dynamic parameters simulation model of low-production pumping wells and the ways to improve system efficiency. *Acta Petrolei Sinica* 39(11): 1299–1307.

Wang ZH, Bai Y, Zhang HQ, et al. (2019) Investigation on gelation nucleation kinetics of waxy crude oil emulsions by their thermal behavior. *Journal of Petroleum Science and Engineering* 181: 106230.

Wu X, Li CJ, Jia WL, et al. (2014) Optimal operation of trunk natural gas pipelines via an inertia-adaptive particle swarm optimization algorithm. *Journal of Natural Gas Science and Engineering* 21: 10–18.

Wu YG (2011) A study on optimal operation and energy saving of super heavy oil pipeline. In: *International conference on computer distributed control and intelligent environmental monitoring (CDCIEM)*, Changsha, China, 19-20 February 2011, pp.400–403. USA: IEEE.

Xie W, Wang YC, Liu XQ, et al. (2019) Nonlinear joint PP-PS AVO inversion based on improved Bayesian inference and LSSVM. *Applied Geophysics* 16(1): 64–76.

Yan Y, Zhang CH, Li K, et al. (2018) The optimal operation strategy for hybrid combined cooling, heating and power microgrid with compressed air energy storage. *Proceedings of the CSEE* 40: 6925–6935.

Yang S, Cao D and Lo K and (2018) Optimizing the impact of economic restructuring on Shanghai’s carbon emissions using STIRPAT and NSGA-II. *Sustainable Cities and Society* 40: 44–53.

Yu G, Chai T and Luo X (2011) Multi-objective production planning optimization using hybrid evolutionary algorithms for mineral processing. *IEEE Transactions on Evolutionary Computation* 15(4): 487–513.

Zhao Y, Yao LR, Liu LJ, et al. (2019) Optimization of energy saving gathering and transportation mode in a block of Tahe oilfield. *Case Studies in Thermal Engineering* 13: 100378–100376.

Zheng HJ and Deng JB (2007) Research and application on designing method of sucker-rod pumping system with the least energy consumption. *Acta Petrolei Sinica* 28: 129–132.
Zhou K, Hou J, Fu HF, et al. (2017) Estimation of relative permeability curves using an improved Levenberg-Marquardt method with simultaneous perturbation Jacobian approximation. *Journal of Hydrology* 544: 604–612.

Zitzler E, Laumanns M, Thiele L, et al. (2001) SPEA2: Improving the strength Pareto evolutionary algorithm. TIK-report 103: 1–22.

Zitzler E, Thiele L and Deb K (2000) Comparison of multi-objective evolutionary algorithms: Empirical results. *Evolutionary Computation* 8(2): 173–195.