processes

MDPI

Article

Energy Consumption Prediction and Optimization of the Electrical Submersible Pump Well System Based on the DA-RNN Algorithm

Xianfu Sui [1,2], Guoqing Han [3], Xin Lu [3,*], Zhisheng Xing [3] and Xingyuan Liang [3]

[1] State Key Laboratory of Offshore Oil and Gas Exploitation, Beijing 100029, China; suixf@cnooc.com.cn
[2] CNOOC Research Institute, Beijing 100028, China
[3] School of Petroleum Engineering, China University of Petroleum-Beijing, Beijing 102249, China; hanguoqing@163.com (G.H.); xingzhisheng0214@163.com (Z.X.); xyliang@cup.edu.cn (X.L.)
* Correspondence: luxincup@outlook.com; Tel.: +86-151-2965-8215

Abstract: The electrical submersible pump (ESP) well system is widely used in the oil industry due to its advantages of high displacement and lift capability. However, it is associated with significant energy consumption. In order to conserve electrical energy and enhance the efficiency of petroleum companies, a deep learning-based energy consumption calculation method is proposed and utilized to optimize the most energy-efficient operating regime. The energy consumption of the ESP well system is precisely determined through the application of the Pearson correlation coefficient analysis method, which is utilized to examine the relationship between production parameters and energy usage. This process aids in identifying the input parameters of the model. Following this, an energy consumption prediction model is developed using the dual-stage attention-based recurrent neural network (DA-RNN) algorithm. To evaluate the accuracy of the DA-RNN model, a comparison of its errors is carried out in comparison to three other deep learning algorithms: Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Transform. Lastly, an orthogonal experiment is executed using the chosen model to pinpoint the most energy-efficient operating regime. Analysis of 325 ESP wells in the Bohai PL oil field indicated that ten parameters, including choke diameter, casing pressure, pump inlet pressure, pump outlet pressure, motor temperature, frequency, oil production, gas production, water production, and GOR significantly impact the energy consumption of the ESP well system. Consequently, these parameters were selected as input variables for the deep learning model. Due to the attention mechanisms employed in the encoding and decoding stages, the DA-RNN algorithm achieved the best performance during model evaluation and was chosen for constructing the energy consumption prediction model. Furthermore, the DA-RNN algorithm demonstrates better model generalization capabilities compared to the other three algorithms. Based on the energy consumption prediction model, the operating regime of the ESP system was optimized to save up to 12% of the maximum energy. The energy consumption of the ESP well system is affected by numerous parameters, and it is difficult to comprehensively evaluate and predict quantitatively. Thus, this work proposes a data-driven model based on the DA-RNN algorithm, which has a dual-stage attention mechanism to rapidly and accurately predict the energy consumption of the ESP well system. Optimization of production parameters using this model can effectively reduce energy consumption.

Keywords: electrical submersible pump; energy efficiency; power consumption optimization; deep learning; DA-RNN algorithm

check for updates

Academic Editor: Qingbang Meng

Received: 29 November 2024
Revised: 26 December 2024
Accepted: 1 January 2025
Published: 6 January 2025

Citation: Sui, X.; Han, G.; Lu, X.; Xing, Z.; Liang, X. Energy Consumption Prediction and Optimization of the Electrical Submersible Pump Well System Based on the DA-RNN Algorithm. *Processes* **2025**, *13*, 128. https://doi.org/10.3390/pr13010128

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Artificial lift is essential for oil fields around the world. Employing various artificial lift methods can extend the lifespan of wells and enhance oil recovery. One of the most widely used artificial lift methods is ESP. Approximately 15% to 20% of the world's one million wells utilize ESP due to its capability to achieve high production rates in deep wells [1]. ESP wells contribute to around 60% of the annual global crude oil production, with ESP representing 43% of the global yearly artificial lift costs [2]. A study has demonstrated that ESP systems utilized in offshore fields consume between 3 and 18 kWh of electricity per barrel of crude oil produced. Furthermore, the study revealed that ESP systems contribute to 67.3% of the total electricity consumption in offshore fields [3]. Currently, most oil fields have advanced to the late stages of development [4]. In these fields, oil production is marked by either a decreasing rate of oil extraction, a rise in the water cut of well production, and an increase in the factors that pose challenges to oil production and transportation processes [5]. Consequently, the primary objective for oil field managers is to optimize the expenses associated with maintaining their continuous operation [6,7].

Accurate calculation of the energy consumption of the ESP system is crucial for reducing operating costs and improving oil field management. The energy consumption of an ESP system is affected by several factors, including the physical and operating parameters of the pump, the downhole environmental conditions, and the physical characteristics of the fluid. Primarily, well depth significantly impacts energy consumption as deeper wells require more energy to overcome gravitational and frictional forces during fluid lifting. Additionally, pump efficiency directly affects energy transformation and utilization, where more efficient pumps can markedly reduce energy consumption. Furthermore, the density and viscosity of the fluid are critical determinants in the required lifting energy, with higher values indicating greater resistance and, thus, higher power requirements. Lastly, the quality of pump installation and maintenance condition also significantly impact energy use; improper installation and lack of maintenance lead to decreased efficiency and increased energy expenditure. To accurately calculate ESP energy consumption, the factors influencing ESP energy consumption must be considered comprehensively.

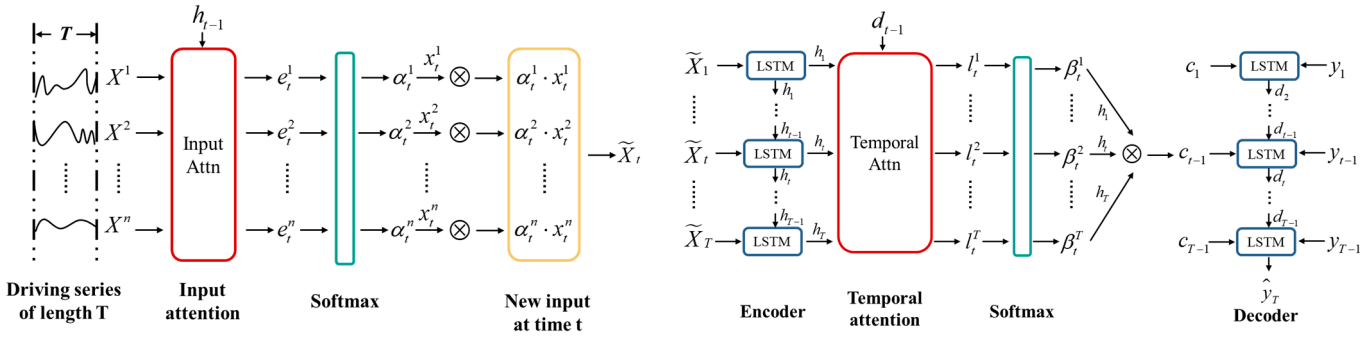
Liang [8] proposed a pump efficiency calculation method to calculate the energy loss of ground equipment, cables, motors, and pumps in an ESP system, but this study should have considered the effect of gas on pump efficiency. The mathematical model proposed by Petrochenkov [9] is more comprehensive, taking into account not only the effect of the gas but also the well pressure curve, degradation during pumping, fluid properties, demulsifier supply, and equipment operating parameters. Lyakhomskii [10] improved the calculation of energy consumption by considering the impact of harmonics on energy usage based on previous research. Grassian [3] not only meticulously calculated the energy consumption of ESP but also conducted a detailed analysis of the energy return on investment (EROI) for ESP. Although mathematical calculations of ESP energy consumption are available, ESP is a complex and integrated system, and it is still difficult for these mathematical methods to take all the influencing factors into account. For instance, it is hard to measure the effects of gas, and it is also difficult to model the impact of cable, motor, and pump performance degradation on energy consumption. The most challenging part is considering human factors, installation quality, and operating conditions. Therefore, new methods are required to conduct ESP energy consumption prediction research. The data-driven method can more effectively address the challenge of multifactor and nonlinear forecasting optimization. Recent studies have explored the use of machine learning and deep learning techniques to predict and optimize ESP energy consumption. For example, a hybrid model combining LSTM and CNNs was proposed to predict ESP operational conditions and diagnose faults, showcasing the ability of deep learning to

handle complex time series data [11]. Similarly, machine learning techniques were applied to estimate production rates in ESP-operated wells, indirectly demonstrating their potential in analyzing energy consumption trends [12]. Additionally, hybrid frameworks combining mechanistic models with data-driven approaches have also been explored to predict ESP performance, further highlighting the increasing reliance on data-driven methodologies [13]. Despite the progress made by these methods, several limitations remain. Many models struggle to effectively integrate and prioritize multiple input factors, particularly in systems where noise and irregularities are present in the data. In addition, some approaches lack the flexibility to capture both long-term dependencies and localized dynamics in time series data, which are critical for accurate energy consumption prediction in ESP systems. These challenges highlight the need for a more advanced and robust methodology that can overcome these constraints.

To address these issues, this study employs DA-RNN to enhance the accuracy and efficiency of ESP energy consumption prediction. By leveraging its dual-attention mechanism, the DA-RNN algorithm effectively identifies and prioritizes the most relevant input parameters through feature attention, while temporal attention captures long-term dependencies and dynamic trends in time series data. This approach enables the model to extract meaningful patterns from extensive and complex production data, bypassing the need to explicitly model the intricate nonlinear interactions among multiple parameters. Consequently, DA-RNN allows the energy consumption of ESP systems to be calculated swiftly and accurately, providing a robust basis for optimizing operating regimes and reducing operational costs in oil field applications.

2. DA-RNN Algorithm

The DA-RNN is a type of network used for time series prediction tasks. It combines an attention mechanism with a recurrent neural network (RNN) to improve the model's ability to identify and utilize important information in the input data [14]. The network consists of an encoder and a decoder, with an attention mechanism integrated into each part. The attention mechanism in the encoder allows it to determine the importance weights of relevant driver sequences when predicting the current target sequence. The temporal attention mechanism in the decoder enables automatic selection of the temporal importance of the encoder's hidden state. The use of attention mechanisms in the DA-RNN network offers several significant benefits. Firstly, the attention mechanism improves the model's interpretability by assigning importance weights to input features, allowing the network to focus on the most relevant parts of the data when making predictions. This reduces the influence of irrelevant or less important features, improving the efficiency and accuracy of the model. Secondly, attention mechanisms enhance the ability of the network to capture long-term dependencies in time series data by dynamically weighting past information based on its relevance to the current prediction. This is particularly important for time series with complex patterns or irregular dependencies. Studies such as Vaswani et al. [15]. have highlighted the ability of attention mechanisms to improve model performance in sequence modeling tasks. Similarly, Soydaner et al. [16]. have emphasized the role of attention mechanisms in enabling deep learning models to handle complex data relationships effectively. These attention mechanisms allow the DA-RNN network to selectively choose relevant input features, filter out noise, and accurately capture the long-term dependence of the time series, ultimately predicting the future trend of the sequence. In the context of the paper, the network is used to accurately predict the energy consumption of the ESP system. The structure of the DA-RNN network is depicted in Figure 1.



(**a**) (**b**)

Figure 1. Architecture of the DA-RNN. (**a**) The input attention mechanism. (**b**) The temporal attention mechanism.

2.1. Encoder with Input Attention Mechanism

An encoder is essentially an RNN. In a time series prediction problem, the encoder can be represented by the following mapping relationship:

$$h_t = f_1(h_{t-1}, X_t) \tag{1}$$

where the expression $h_t \in R^m$ refers to the hidden state vector of the encoder at time t, and m represents the size of the hidden state vector. The nonlinear activation function, denoted as f_1, can be a LSTM or a GRU. X_t denotes the value of the ESP energy consumption features at time t. Within a given sequence of ESP energy consumption features $X = (X_1, X_2, \ldots, X_T)$, $X_t \in R^n$, n indicates the number of features involved in the particular prediction of ESP energy consumption. The nonlinear activation function utilized in this study is the LSTM cell, and the operation process of the LSTM cell can be expressed using the following equations:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$\widetilde{s}_t = \tanh(W_s \cdot [h_{t-1}, x_t] + b_s) \tag{4}$$

$$s_t = f_t \cdot s_{t-1} + i_t \cdot \widetilde{s}_t \tag{5}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{7}$$

where σ is the sigmoid activation function, and the matrix $[h_{t-1}, x_t] \in R^{m+n}$ is created by concatenating the hidden state vector from the previous time step with the input feature vector from the current time step. Additionally, W_f, W_i, W_s, and $W_o \in R^{m \times (m+n)}$ denote the weight matrix, while b_f, b_i, b_s, and $b_o \in R^m$ represent the bias vector. It is important to note that the weight matrix and the bias vector are parameters that need to be learned. Using LSTM as an activation layer effectively addresses the issue of gradient vanishing and allows for capturing long-term dependencies within time series data. In the encoder's LSTM cell, an input attention mechanism is implemented as a multilayer perceptron, and the computational process can be expressed using the following equation:

$$e_t^k = V_e^\top \tanh(W_e[h_{t-1}, s_{t-1}] + U_e X^k) \tag{8}$$

$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{i=1}^{n} \exp(e_t^i)} \tag{9}$$

In Equation (8), the parameters to be learned are $V_e \in R^T$, $W_e \in R^{T \times 2m}$, and $U_e \in R^{T \times T}$. The attention weight α_t^k measures the importance of the k-th feature at time t. As demonstrated in Equation (9), the softmax function is employed to ensure that the sum of the attention weights equals 1. The input attention mechanism serves as a feedforward neural network that can be included in the training process alongside the rest of the RNN. By utilizing the attention weights, features highly correlated with ESP energy consumption can be adaptively determined and represented by the following equation:

$$\widetilde{X}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \ldots, \alpha_t^n x_t^n)^T \tag{10}$$

The features processed by the input attention mechanism are used to update the hidden state of the LSTM cell at time t. The update process can be expressed by the following equation:

$$h_t = f_1(h_{t-1}, \widetilde{X}_t) \tag{11}$$

The LSTM cell denoted by f_1 can be computed using Equations (2)–(7), and the use of $\widetilde{X}_t$ instead of x_t allows the encoder to selectively focus on the feature sequences highly correlated with the predicted target.

2.2. Decoder with Temporal Attention Mechanism

Within the decoder, an LSTM cell is employed to decode the input information from the encoder, ultimately yielding the predicted output upon completion of the decoding process. Nonetheless, the accuracy of the encoding–decoding network's predictions diminishes swiftly as the length of the input feature sequence increases [17]. Consequently, it becomes imperative to integrate a temporal attention mechanism into the decoder, enabling it to dynamically ascertain the relevance of the encoder's hidden state at every time step. This temporal attention mechanism can be mathematically expressed as follows:

$$l_t^i = V_d^\top \tanh(W_d[d_{t-1}, s'_{t-1}] + U_d h_i) \tag{12}$$

$$\beta_t^i = \frac{\exp(l_t^i)}{\sum_{j=1}^{T} \exp(l_t^i)} \tag{13}$$

where $d_{t-1} \in R^p$ is the hidden state vector of the decoder at the time $t-1$, and $s'_{t-1} \in R^P$ is the hidden state vector of the LSTM cell at the time $t-1$. $[d_{t-1}, s'_{t-1}]$ is a matrix created by concatenating the hidden state vector of the decoder at the time $t-1$ with the hidden state vector of the LSTM cell at the time $t-1$. $V_d \in R^m$, $W_d \in R^{m \times 2p}$, and $U_d \in R^{m \times m}$ are the parameters that need to be learned. β_t^i is the temporal attention weight, which is used to measure the importance of the ith encoder. After obtaining the attention weights, the context vector c_t at the current time is further computed. c_t is the weighted average of the hidden state of each encoder, which can be expressed by the following equation:

$$c_t = \sum_{i=1}^{T} \beta_t^i h_i \tag{14}$$

After obtaining the context vector c_t, it is combined with the target sequence.

$$\widetilde{y}_{t-1} = \widetilde{w}^\top [y_{t-1}, c_{t-1}] + \widetilde{b} \tag{15}$$

The matrix is created by concatenating the decoder input and context vectors. $\widetilde{w}^T \in R^{m+1}$ and $\widetilde{b} \in R$ are the parameters to be learned. The $\widetilde{y}_{t-1}$ in the equation is used to compute the hidden state of the decoder at time t, as shown in the following equation:

$$d_t = f_2(d_{t-1}, \widetilde{y}_{t-1}) \tag{16}$$

The nonlinear activation function f_2 is still an LSTM cell, and the process of updating d_t using the nonlinear activation function can be referenced to the update process of h_t. Finally, the following linear function is utilized to obtain the final target output:

$$\begin{aligned} \hat{y}_T &= F(y_1, \ldots, y_{T-1}, x_1, \ldots, x_T) \\ &= v_y^\top (W_y[d_T, c_T] + b_w) + b_v \end{aligned} \tag{17}$$

where $[d_T, c_T]$ denotes the concatenation matrix of the decoder's hidden state vector and context vector. The parameter $W_y \in R^{p \times (p+m)}$ and $b_w \in R^p$ map the concatenated matrix to the hidden state of the decoder. The final target output is obtained by a linear mapping of parameter $v_y \in R^p$ and parameter $b_v \in R$.

3. Models and Methods

The main process of the data-driven energy consumption prediction and operating regime optimization of the ESP system is shown in Figure 2.

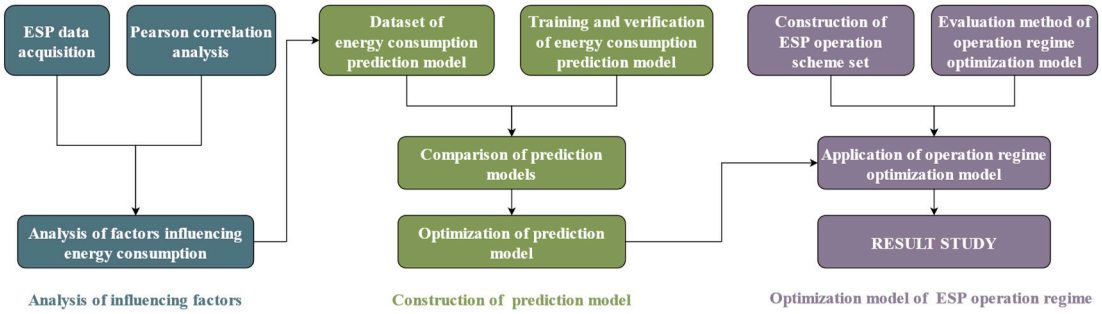


Figure 2. Data-driven research process of ESP energy consumption prediction and optimization.

Step 1. Constructing the dataset. The raw data are first pre-processed, then the main controlling factors of energy consumption are identified, and the parameters that have a small impact on energy consumption are eliminated. Finally, the dataset is established.

Step 2. Split the dataset into training, validation, and test sets, which will be utilized for model training, hyperparameter tuning, and model performance evaluation, respectively.

Step 3. Train the deep learning model using the training set and tune the hyperparameters on the validation set to obtain the selected ESP energy consumption prediction model.

Step 4. Analyze historical data of the ESP system to establish an operating regime adjustment scheme set.

Step 5. Based on the ESP operating regime set constructed in Step 4, a sensitivity analysis using the selected prediction model is performed to optimize the ESP operating regime.

3.1. Analysis of ESP Energy Consumption Influencing Factors

The Pearson correlation coefficient method was utilized to assess the correlation between various variables and the energy consumption of ESPs. Variables demonstrating significant correlation were included, while those showing low correlation were excluded. The Pearson correlation coefficient, which ranges from -1 to 1, indicates a stronger correlation with a higher absolute value. This coefficient is calculated using the Equation (18).

The application of Pearson correlation for feature selection has been well-documented in the literature. For instance, Shahzad et al. [18] used Pearson correlation to identify key input variables for predicting the power consumption of an air conditioner, demonstrating its ability to reduce input dimensionality while maintaining model performance. Similarly, Rita et al. [19] applied the Pearson correlation coefficient in a study to rank influential features in energy forecasting, highlighting its effectiveness in selecting critical inputs that improve model interpretability. These examples reinforce the applicability of the Pearson correlation method in energy consumption prediction and support its use in this study to identify the most relevant variables for ESP energy consumption modeling.

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \quad (18)$$

where r represents the Pearson correlation coefficient, X_i and Y_i are the observed values of the two variables, and $\overline{X}$ and $\overline{Y}$ are the mean values of the two variables, respectively.

This study utilizes data from 325 ESP wells in the Bohai PL oil field to investigate the factors influencing energy consumption. The dataset encompasses wellhead sensor parameters, downhole sensor parameters, electrical parameters, and fluid parameters. Since the data were collected from different types and models of ESPs, the current model does not specifically account for pump type. The fundamental statistical details of each parameter are detailed in Table 1. Prior to analysis, the data from ESP wells underwent pre-processing to eliminate outliers beyond the normal range, fill in any missing values, and standardize the dataset for further analysis.

Table 1. Basic overview of 325 ESP wells in the Bohai PL oil field.

No.	Variable Symbol	Influencing Factors	Unit	Minimum	Mean	Maximum
1	V1	Choke diameter	1/64″	6.10	38.20	59.00
2	V2	Tubing pressure	psi	194.35	289.7	433.66
3	V3	Casing pressure	psi	78.32	247.96	362.59
4	V4	Pump inlet pressure	psi	131.98	474.91	1667.93
5	V5	Pump outlet pressure	psi	1314.04	1790.82	2158.16
6	V6	Pump inlet temperature	temp	51	62.66	75.1
7	V7	Motor temperature	temp	60.7	88.98	134.4
8	V8	Motor current frequency	Hz	30	46.14	61.5
9	V9	Oil production	bbl	50.19	410.38	1246.51
10	V10	Gas production	Mscf	2.01	95.81	396.84
11	V11	Water production	bbl	1.01	1261.62	5335.83
12	V12	Gas–Oil Ratio (GOR)	Mscf/bbl	0.01	0.24	0.84

The correlation between 12 variables and ESP energy consumption was calculated using the Pearson coefficient method. The results of the calculations are shown in Figure 3 in the form of a thermodynamic diagram.

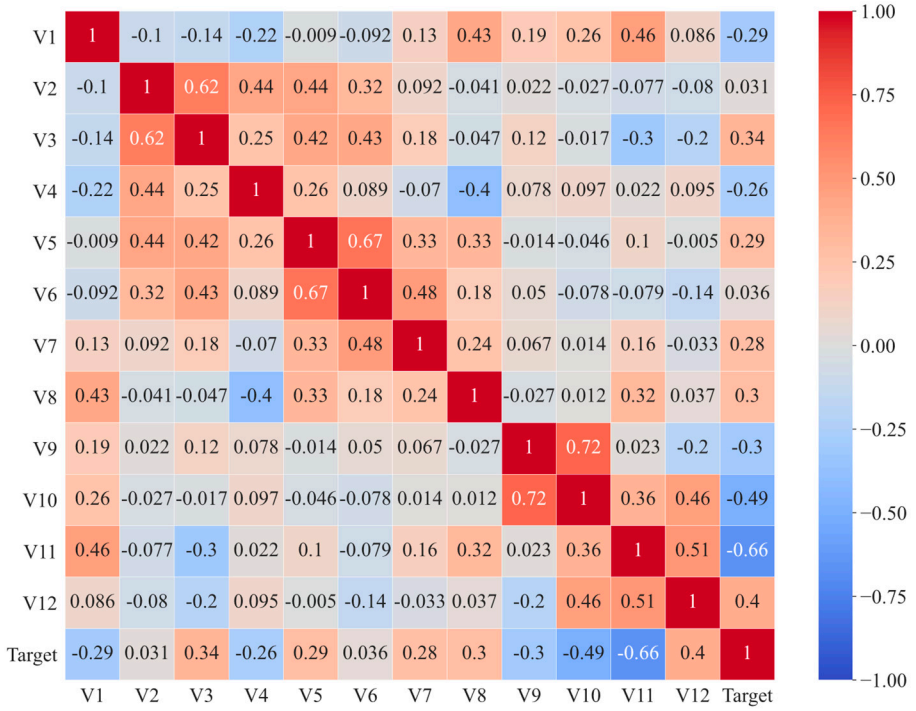


Figure 3. Thermodynamic diagram of Pearson correlation coefficient.

Figure 3 illustrates the correlations among the 13 variables. Table 1 displays V1 to V12, while Target represents the ESP energy consumption, serving as the prediction target for the energy consumption prediction model. The energy consumption of different ESP types varies significantly, mainly due to differences in their lifting capacities. So, it is essential to normalize the energy consumption across various types of ESPs to a standardized scale. This study defines ESP energy consumption as the electrical energy consumed to lift a unit volume of fluid, which can be calculated using the following equation:

$$EC_{\text{scale}} = \frac{DEC}{DLP} \tag{19}$$

EC_{scale} represents the normalized ESP energy consumption, measured in kWh/bbl. The daily energy consumption (DEC) indicates the total energy consumed by the ESP in one day, measured in kWh/d. The daily liquid production (DLP) represents the total liquid production by the ESP in one day, measured in BPD. This normalization enables a more accurate comparison and prediction of energy requirements, accounting for the inherent differences in performance and capacity among the different ESP models.

As observed from Figure 3, all variables except for the tubing pressure and pump inlet temperature exhibit a certain degree of correlation with the target variable. Variables with a Pearson correlation coefficient lower than 0.1 are considered to have no significant correlation with the target variable, which is why tubing pressure and pump inlet temperature were excluded from further analysis. The target variable shows a negative correlation with choke diameter, pump inlet pressure, oil production, gas production, and water production. This suggests that lower values of these parameters are associated with higher energy consumption of the ESP. A smaller choke diameter limits liquid flow, potentially increasing the pump's load and, consequently, energy consumption. Lower pump inlet pressure forces the pump to overcome a larger pressure differential to lift the fluid, also leading to increased energy consumption. Typically, higher oil, gas, and water production results in greater energy consumption of ESPs. However, this study found the opposite trend. This is attributed to the fact that the pumping capacity of the ESPs in the covered fields is generally

greater than the supply capacity of the formation, causing the ESPs not to operate at their optimal efficiency point. Therefore, higher fluid production allows for better matching of the pumping capacity of the ESP and the supply capacity of the formation, leading to lower energy consumption.

On the other hand, the target variable exhibits a positive correlation with casing pressure, pump outlet pressure, motor temperature, frequency, and GOR. This indicates that higher values of these parameters correspond to higher energy consumption of the ESP. Excessive casing pressure can inhibit the flow of reservoir fluids into the wellbore, reducing the fluid intake by the pump, which in turn lowers pump efficiency and increases energy consumption. Higher pump outlet pressure means the pump must overcome a greater pressure differential, thereby increasing energy consumption. Elevated motor temperature results in increased resistance in the motor windings, which reduces the motor's energy conversion efficiency and thus increases energy consumption. Additionally, higher motor temperature degrades the performance of lubricants, exacerbates mechanical wear, and increases friction, requiring more power to overcome these additional frictional forces. Higher frequency generally indicates higher flow rates and energy consumption. A high GOR signifies a substantial gas content in the liquid, which affects pump efficiency and increases energy consumption. Tubing pressure and pump inlet temperature did not show a significant correlation with the target variables. For tubing pressure, a higher pump inlet pressure makes it easier to achieve a higher tubing pressure, which means that the pump only has to overcome a small differential pressure. Therefore, higher tubing pressure results in lower energy consumption. Conversely, with a lower pump inlet pressure, high tubing pressure can also be achieved by increasing the frequency of ESP operation, thereby increasing the pump outlet pressure. This implies that higher tubing pressure can result in higher energy consumption. Consequently, the correlation between tubing pressure and energy consumption depends on the specific production situation, which explains why tubing pressure does not show a significant correlation with energy consumption. For pump inlet temperature, higher inlet temperatures typically reduce the viscosity of the liquid, thereby decreasing the pump load and energy consumption. Conversely, excessively high temperatures can also negatively impact the material performance of the pump and the efficiency of the motor, leading to increased energy consumption. Therefore, pump inlet temperature also does not show a significant correlation with the target variable.

In summary, the tubing pressure and pump inlet temperature are excluded, and the choke diameter, casing pressure, pump inlet pressure, pump outlet pressure, motor temperature, frequency, oil production, gas production, water production, and GOR are retained as the input features of the energy consumption prediction model.

3.2. Energy Consumption Prediction Model and Performance Evaluation

To establish an optimal energy consumption prediction model for ESP, this study first employs four deep learning algorithms—GRU, LSTM, Transformer, and DA-RNN—to develop an energy consumption prediction model. The prediction performance of these four deep learning algorithms is compared. By analyzing these models, the one with the highest prediction accuracy is selected to further predict ESP energy consumption and optimize the operating regime of ESP.

3.2.1. Model Comparison

LSTM is a recurrent neural network designed to handle long-term dependencies using input, output, and forget gates. These gates manage the cell state, enabling LSTMs to effectively learn from long sequences, which is ideal for speech recognition and language modeling tasks. GRU is a recurrent neural network that addresses the vanishing gradi-

ent problem using update and reset gates. These gates control the flow of information, making GRUs efficient for time series prediction and sequence modeling with fewer parameters than the LSTM. The Transformer model uses self-attention mechanisms to process sequences in parallel, capturing dependencies between all elements without relying on sequential data processing. This architecture improves training efficiency and performance, making the Transformer suitable for NLP tasks and time series analysis. The prediction model of ESP energy consumption was constructed using the above three deep learning algorithms and the DA-RNN algorithm; 60% of the dataset is allocated to a training set, 20% to a validation set, and 20% to a testing set. These are used for model training, hyperparameter optimization, and model performance evaluation, respectively.

This study utilizes three indicators to assess the performance of the four deep learning models. The three indicators are the magnitude of relative error (MRE), the mean absolute error (MAE), and the root mean square error (RMSE). The specific calculations are shown in Table 2.

Table 2. Energy consumption prediction model evaluation metrics.

Evaluation Metrics	Calculation Formula
MRE	$MRE = \sum_{i=1}^{n} \left\| \frac{y_t - y_{pre}}{y_t} \right\| \times \frac{1}{n}$
MAE	$MAE = \frac{\sum_{i=1}^{n} \left\| y_t - y_{pre} \right\|}{n}$
RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_t - y_{pre})^2}$

The performance of the deep learning model is evaluated based on the above three metrics, and the best-performing energy prediction model is selected. The results of the model performance comparison are shown in Figure 4. It can be seen that the DA-RNN model has the best prediction results for the training and test sets, and the MRE, MAE, and RMSE of the training and test sets are smaller. Therefore, DA-RNN is selected as the final energy consumption prediction model in this study.

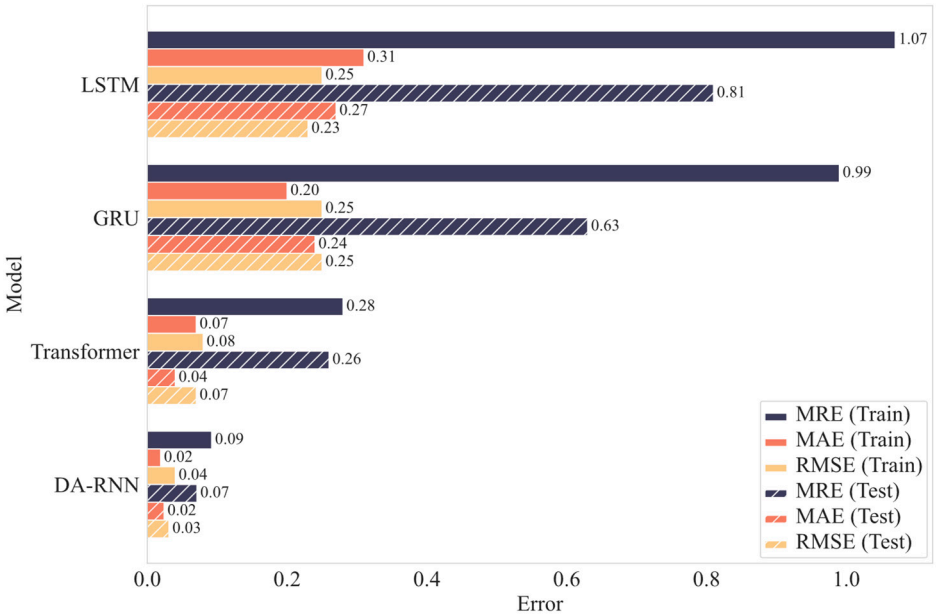


Figure 4. Performance comparison of four energy consumption prediction models.

3.2.2. Hyperparameter Optimization of the DA-RNN Model

In order to obtain the best-performing DA-RNN model, the hyperparameters were optimized using a grid search method with k-fold cross-validation to ensure that the information in the dataset is fully utilized. Hyperparameter optimization plays a critical role in enhancing the performance of neural networks by ensuring that the model's architecture and configuration are well-suited for the given task. For instance, in the study by Huang et al. [20]., the authors emphasized the importance of hyperparameter tuning in achieving superior prediction accuracy, particularly for periodic time series data. Similarly, Yu and Zhu [21] highlighted that hyperparameter optimization is essential for improving model accuracy and efficiency. They discuss how methods such as grid search, random search, and Bayesian optimization help refine neural network performance and reduce overfitting. By systematically searching for the optimal combination of hyperparameters, grid search combined with k-fold cross-validation ensures a more robust and reliable model, making it particularly valuable for tasks like energy consumption prediction. The k-fold cross-validation requires dividing the dataset into k parts, utilizing k-1 parts for training, and retaining one part for testing. The test accuracy is obtained for each test. In this study, k was set to 10. The hyperparameters of the energy consumption prediction model based on the DA-RNN include the length of the data window, learning rate, number of neurons in the encoder, and number of neurons in the decoder. The value range of the data window length is 7–28, and the adjustment step is 7. The value range of the learning rate is 0.001–0.1, the adjustment step is 0.001, the value range of the number of encoder and decoder neurons is 64–512, and the adjustment step is 64. Some of the hyperparameter optimization results are shown in Table 3.

Table 3. Performance comparison of different parameters of the DA-RNN model.

Data Window Length	Learning Rate	Number of Neurons in the Encoder	Number of Neurons in the Decoder	MRE	MAE	RMSE
27	0.001	128	128	0.0508	0.0032	0.0041
27	0.001	256	256	0.0554	0.0049	0.0051
14	0.001	128	128	0.0687	0.0062	0.0058
14	0.003	256	128	0.0761	0.0078	0.0118
7	0.002	128	128	0.0873	0.0081	0.0129
7	0.003	128	128	0.0913	0.0108	0.0131

Table 3 illustrates that the DA-RNN network demonstrates the lowest error when the window length is set to 27, the learning rate is 0.001, and both the number of encoder and decoder neurons are 128. The optimization results indicate that a longer data window can extract more detailed feature information from a sequence, and a lower learning rate can enhance the model's convergence and generalization performance, but an excessive number of neurons may lead to overfitting and result in lower prediction accuracy on the test set.

After fine-tuning the hyperparameters, the DA-RNN achieves high accuracy on both the training and test sets. The model's predictions on the dataset are illustrated in Figure 5, where the horizontal axis represents real energy consumption, and the vertical axis shows predicted energy consumption. The clustering of sample points from the training and test sets around the $y = x$ diagonal in the figure indicates that the model's predictions closely align with the actual observations.

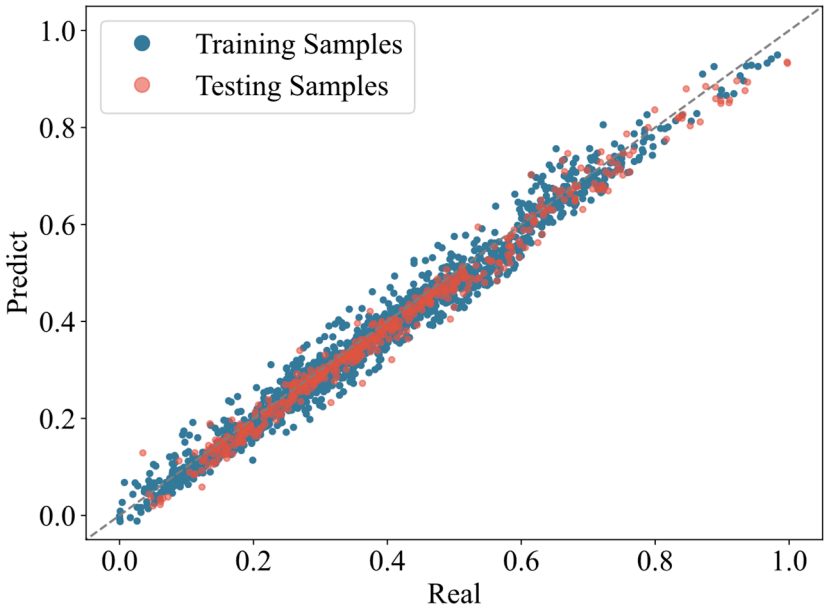


Figure 5. DA-RNN model predictions for the entire sample.

3.3. ESP Operation Regime Optimization Model

The optimization workflow is shown in Figure 6.

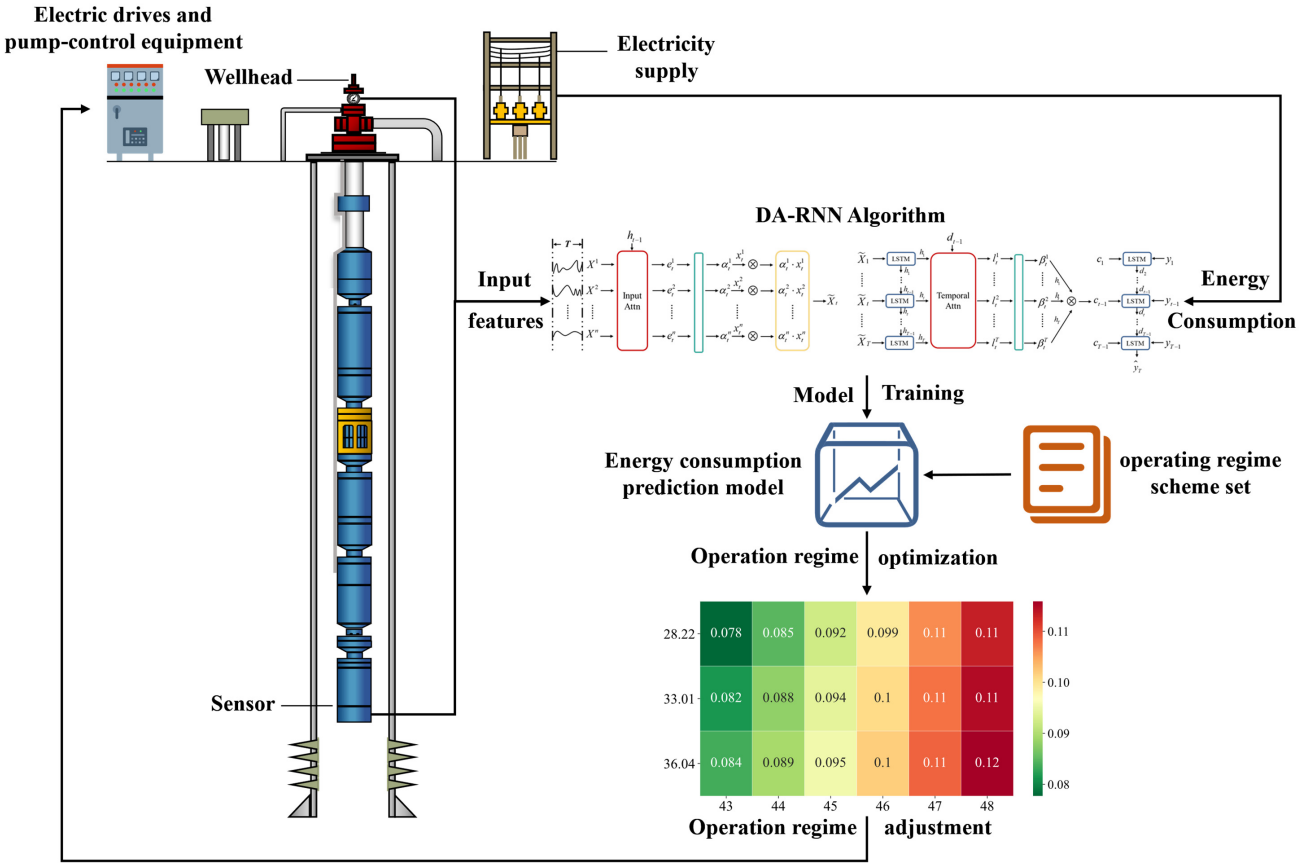


Figure 6. Optimization workflow for ESP operating regime.

There are several core aspects of optimizing the operation regime of ESPs using the energy consumption prediction model. First, accurate estimation of the energy consumption of the ESP by the prediction model is a prerequisite for optimizing the operation

mechanism. Second, the input features of the prediction model contain adjustable ESP operating parameters. Finally, it is crucial to establish a reasonable evaluation method to optimize the operating regime of the ESP, and the optimization objective function is shown in the following equation:

$$\begin{cases} \min \mathrm{EC}_{scale}(V_i),\ i \in (1,3,4,5,7,8,9,10,11,12) \\ M_1 \le V_1 \le M_2 \\ M_3 \le V_8 \le M_4 \end{cases} \tag{20}$$

In Equation (20), V_i is the 10 model input features identified above. $M_{i,i=1,3}$ is the minimum value of the corresponding parameter that can satisfy the on-site production, and $M_{i,i=2,4}$ is the maximum value of the corresponding parameter that can guarantee the safe operation of the production system.

In order to establish a set of ESP operating regime schemes, the first step is to calculate the outflow performance curves at various ESP operating frequencies and choke diameters and then determine their intersection with the IPR curves to find the operating points. Following this, the wellbore pressure distributions, temperature distributions, and fluid flow rates at the corresponding operating points should be calculated. Finally, the motor efficiencies at the corresponding operating points should be determined, followed by the calculation of the motor temperatures based on the motor efficiencies. This series of steps leads to the construction of the ESP operating regime set. The ESP operation regime set is fed into the energy consumption prediction model to find the optimal operation regime.

4. Result and Discussion

The DA-RNN-based ESP energy consumption prediction model accurately calculates the electrical energy consumption of ESPs. This model enables real-time online energy consumption calculation and lays the foundation for optimizing the ESP operation regime. The effect of energy consumption prediction for well A in the Bohai PL field is shown in Figure 7.

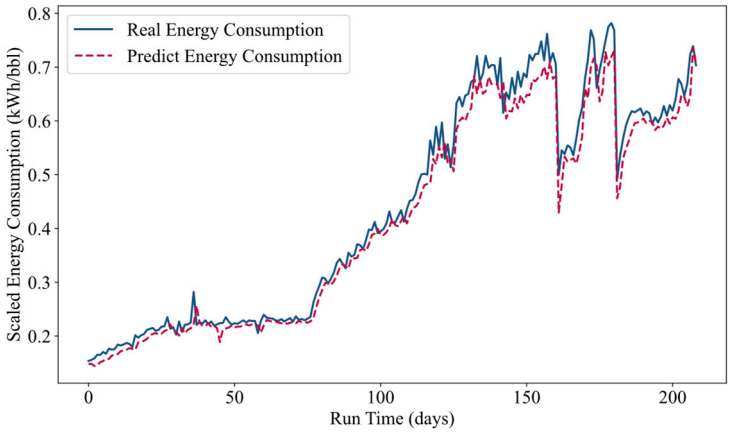


Figure 7. Predicted effect of energy consumption for well A.

Figure 7 depicts the actual energy consumption (blue solid line) compared to the predicted energy consumption (red dashed line) over the duration of the runtime. Overall, the predicted curve closely follows the trend of actual energy consumption. While there are minimal differences between the predicted and actual energy consumption for most time periods, some deviations are noticeable in regions with significant fluctuations, particularly around runtimes close to 150 and 200. Despite these occasional discrepancies, the prediction model effectively captures the overall trend of actual energy consumption with high reliability.

On the basis of the prediction model, the operating regime of the ESP was optimized. Using the optimization workflow mentioned above, the most energy-efficient combination of ESP frequency and choke diameter can be selected. In practical operations, the adjustment of frequency is achieved through variable frequency drives (VFDs), which provide precise control over the ESP motor speed. The results of the sensitivity analysis for different operating frequencies and choke diameters are shown in Figure 8.

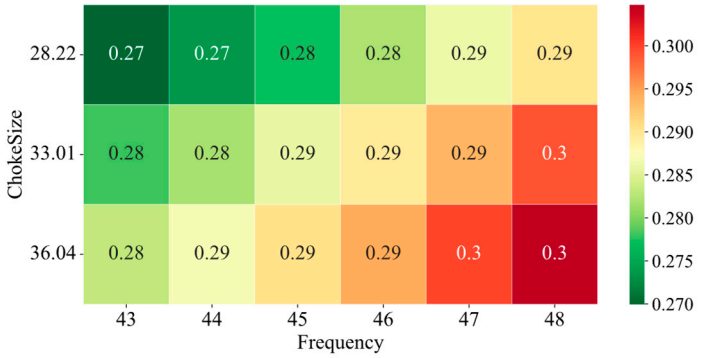


Figure 8. Sensitivity analysis of frequency and choke diameter.

The figure shows the energy consumption thermogram at different frequencies and choke diameters. The vertical axis represents the choke diameter, the horizontal axis represents the frequency, and the colors represent the energy consumption values, which are marked in each cell. The color changes gradually from green to red, representing lower to higher energy consumption values. Choosing the right combination of frequency and choke diameter to allow the ESP to operate in the high-efficiency zone can effectively reduce power consumption. This result aligns with established ESP operational principles, as a higher frequency typically increases motor speed and energy consumption due to greater mechanical and electrical losses, while larger choke diameters reduce system backpressure but may cause the pump to operate away from its optimal efficiency point. The thermogram highlights how careful selection of frequency and choke diameter can mitigate these effects and maintain the pump in a high-efficiency operating range, thereby optimizing energy usage. The optimization effect of well B is shown in Figure 9.

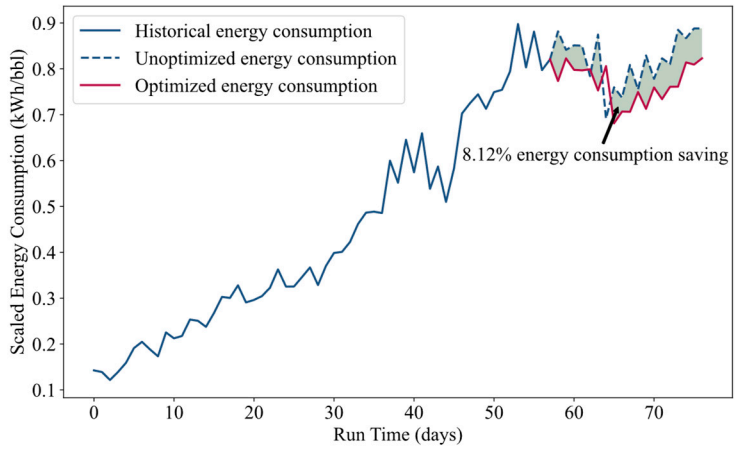


Figure 9. Optimization effect of well B.

Following 55 days of operation, the operating regime for well B was optimized. Well B represents the well with the best optimization effect among the 42 wells analyzed during the study. In Figure 9, the solid blue line represents the historical energy consumption, while the red solid line represents the measured energy consumption after the optimization. The

blue dashed line shows the energy consumption predicted by the model while maintaining the original operating regime. The green shading in the figure represents the percentage reduction in energy consumption. After the optimization, the energy consumption of well B decreased by 8.12%, which was the maximum energy savings observed among all wells. Across the 42 wells in the Bohai PL oil field, the implementation of optimized operating regimes over a one-month period resulted in energy savings ranging from a minimum of 3.37% to a maximum of 8.12%, with an average reduction of 5.19%. This demonstrates that the optimization method is not only effective but also statistically reliable across multiple wells. It is worth noting that traditional frequency control systems primarily aim to adapt to reservoir changes and regulate production. In contrast, the proposed method not only achieves these objectives but also systematically minimizes energy consumption. This dual benefit emphasizes the advantage of integrating choke size adjustments into the frequency control process, providing a more comprehensive solution for energy efficiency in ESP operations.

In terms of implementation costs, the proposed method leverages data that are already collected by standard ESP monitoring systems, such as pressure, temperature, frequency, and flow rates. As a result, no significant investment in additional sensors or hardware is required. The main cost lies in the computational resources needed to train and deploy the DA-RNN algorithm, which are relatively minimal compared to the potential energy savings it delivers. Additionally, the algorithm can be implemented using existing digital infrastructure and integrated into current ESP monitoring and control systems with minimal modifications, making it a cost-effective solution for improving energy efficiency.

While the current study focuses on optimizing choke size and frequency—the two parameters that can be manually adjusted in real time—the proposed method can be adapted to meet different optimization objectives. For example, the algorithm can identify the most energy-efficient operating regime for maintaining current production levels or find the optimal conditions to increase production with the least possible additional energy consumption. This flexibility underscores the versatility of the proposed method and its potential applicability to a wide range of operational goals in ESP systems.

5. Conclusions

The data-driven approach allows accurate prediction of ESP energy consumption. The DA-RNN algorithm uses attention mechanisms in both the encoding and decoding phases. This method achieves the highest accuracy in energy consumption prediction. The DA-RNN has the lowest error among all models, with an MRE of 0.09 for training and 0.07 for testing, an MAE of 0.02 for training and 0.04 for testing, and an RMSE of 0.04 for training and 0.03 for testing. Compared to the Transformer, the DA-RNN reduces testing MRE by approximately 73% (0.07 vs. 0.26) and testing RMSE by over 57% (0.03 vs. 0.07). Similar improvements are observed against the LSTM and GRU models, where DA-RNN achieves significant reductions in error metrics. The energy consumption prediction model is built using the DA-RNN algorithm. It is then used to optimize the ESP operation regime. This workflow helps find the best combination of ESP operating frequency and choke diameter. It reduces ESP energy consumption effectively. However, one limitation of the current method is that it relies on historical data for model training and prediction. If the operating conditions deviate significantly from the data used to train the model, the prediction accuracy may decrease. To address this limitation, future work could incorporate online learning techniques or adaptive models to handle changes in operating conditions more effectively.

Author Contributions: Methodology, X.S. and G.H.; Data curation, Z.X.; Writing—original draft, X.L. (Xin Lu); Writing—review & editing, X.L. (Xingyuan Liang). All authors have read and agreed to the published version of the manuscript.

Funding: This work was financial supported by the Science Foundation of China University of Petroleum, Beijing (No. 2462023YJRC019), and the National Natural Science Foundation of China (No. 52204059).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Xianfu Sui was employed by the company CNOOC Research Institute. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The company had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Pham, S.T.; Vo, P.S.; Nguyen, D.N. Effective Electrical Submersible Pump Management Using Machine Learning. *Open J. Civ. Eng.* **2021**, *11*, 70–80. [CrossRef]
2. Yang, J.; Wang, S.; Zheng, C.; Feng, G.; Du, G.; Tan, C.; Ma, D. Fault Diagnosis Method and Application of ESP Well Based on SPC Rules and Real-Time Data Fusion. *Math. Probl. Eng.* **2022**, *2022*, 72–82. [CrossRef]
3. Grassian, D.; Olsen, D. Detailed Energy Accounting of Electrical Submersible Pumping Systems. *Energies* **2020**, *13*, 302. [CrossRef]
4. Galkin, S.V.; Krivoshchekov, S.N.; Kozyrev, N.D.; Kochnev, A.A.; Mengaliev, A.G. Accounting of geomechanical layer properties in multi-layer oil field development. *J. Min. Inst.* **2020**, *244*, 408–417. [CrossRef]
5. Nurgalieva, K.S.; Saychenko, L.A.; Riazi, M. Improving the Efficiency of Oil and Gas Wells Complicated by the Formation of Asphalt–Resin–Paraffin Deposits. *Energies* **2021**, *14*, 6673. [CrossRef]
6. Bakker, S.J.; Kleiven, A.; Fleten, S.-E.; Tomasgard, A. Mature offshore oil field development: Solving a real options problem using stochastic dual dynamic integer programming. *Comput. Oper. Res.* **2021**, *136*, 105480. [CrossRef]
7. Wan, X.; Xie, G.; Ding, Y. Exploration of geology-engineering integration in hard-to-recover reserves in the Shengli Oilfield. *China Pet. Explor.* **2020**, *25*, 43. [CrossRef]
8. Liang, X.; Fleming, E. Electrical Submersible Pump Systems: Evaluating Their Power Consumption. *IEEE Ind. Appl. Mag.* **2013**, *19*, 46–55. [CrossRef]
9. Petrochenkov, A.; Ilyushin, P.; Mishurinskikh, S.; Kozlov, A. Development of a Method for Improving the Energy Efficiency of Oil Production with an Electrical Submersible Pump. *Inventions* **2023**, *8*, 29. [CrossRef]
10. Lyakhomskii, A.; Petrochenkov, A.; Romodin, A.; Perfil'eva, E.; Mishurinskikh, S.; Kokorev, A.; Kokorev, A.; Zuev, S. Assessment of the Harmonics Influence on the Power Consumption of an Electric Submersible Pump Installation. *Energies* **2022**, *15*, 2409. [CrossRef]
11. Liu, D.; Feng, G.; Feng, G.; Xie, L. Hybrid Long Short-Term Memory and Convolutional Neural Network Architecture for Electric Submersible Pump Condition Prediction and Diagnosis. *SPE J.* **2024**, *29*, 2130–2147. [CrossRef]
12. Agwu, O.E.; Alkouh, A.; Alatefi, S.; Azim, R.A.; Ferhadi, R. Utilization of machine learning for the estimation of production rates in wells operated by electrical submersible pumps. *J. Pet. Explor. Prod. Technol.* **2024**, *14*, 1205–1233. [CrossRef]
13. Du, C.; Gong, F.; Zhou, Y.; Lu, Y.; Wang, H.; Gao, J. A mechanistic-based data-driven modelling framework for predicting production of electric submersible pump wells in offshore oilfield. *Geoenergy Sci. Eng.* **2024**, *246*, 213603. [CrossRef]
14. Luo, Y.; Chen, Z.; Yoshioka, T. Dual-path RNN: Efficient long sequence modeling for time-domain single-channel speech separation. *arXiv* **2020**, arXiv:1910.06379. [CrossRef]
15. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need.(Nips), 2017. *arXiv* **2017**, arXiv:1706.03762.
16. Soydaner, D. Attention mechanism in neural networks: Where it comes and where it goes. *Neural Comput. Appl.* **2022**, *34*, 13371–13385. [CrossRef]
17. Cho, K.; van Merrienboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv* **2014**, arXiv:1409.1259.
18. Yousaf, S.; Bradshaw, C.R.; Kamalapurkar, R.; San, O. A gray-box model for unitary air conditioners developed with symbolic regression. *Int. J. Refrig.* **2024**, *168*, 696–707. [CrossRef]
19. Banik, R.; Das, P.; Ray, S.; Biswas, A. Prediction of electrical energy consumption based on machine learning technique. *Electr. Eng.* **2021**, *103*, 909–920. [CrossRef]

20. Huang, B.; Zheng, H.; Guo, X.; Yang, Y.; Liu, X. A Novel Model Based on DA-RNN Network and Skip Gated Recurrent Neural Network for Periodic Time Series Forecasting. *Sustainability* **2022**, *14*, 326. [CrossRef]
21. Yu, T.; Zhu, H. Hyper-Parameter Optimization: A Review of Algorithms and Applications. *arXiv* **2020**, arXiv:2003.05689. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.