



Prediction of Oil Pipeline Process Operating Parameters Based on Mechanism and Data Mining

Lixin Wei

School of Petroleum Engineering,
Northeast Petroleum University,
Daqing 163318, China
e-mail: weilixin73@163.com

Lan Wang¹

School of Petroleum Engineering,
Northeast Petroleum University,
Daqing 163318, China
e-mail: wanglan2516@163.com

Qiang Zhou

School of Petroleum Engineering,
Northeast Petroleum University,
Daqing 163318, China
e-mail: qzhou807@163.com

Yuhang Gao

School of Petroleum Engineering,
Northeast Petroleum University,
Daqing 163318, China
e-mail: 1721943555@qq.com

Precisely forecasting the operational characteristics of oil pipelines is essential for developing rational design, production, and operation strategies, as well as reducing energy consumption and saving energy. Due to significant disparities in the computation outcomes of conventional mechanism models and the inadequate performance of machine learning models when handling limited sample data, their conclusions likewise lack tangible significance. In this study, a novel physics-guided neural network (PGNN) model, which integrates mechanisms with machine learning models, is introduced. The proposed model incorporates essential physical intermediate factors that impact the temperature and pressure of oil pipelines as artificial neurons within the loss function. Additionally, an adaptive moment estimate approach is employed to optimize the parameters of the model. Through a comparative analysis of various models' predictive capabilities on an oil pipeline, it was shown that PGNN has the highest level of accuracy in forecasting pipeline temperature and pressure. Furthermore, PGNN demonstrates the ability to generate satisfactory prediction outcomes even with a limited sample size. Simultaneously, the predictive outcomes of PGNN exhibit a stronger correlation with variables that have a direct impact on temperature and pressure. [DOI: 10.1115/1.4065951]

Keywords: oil pipeline, parameter prediction, PGNN, mechanism, energy systems analysis, petroleum transport/pipelines/multiphase flow

1 Introduction

As the main mode of oil transportation, pipeline transportation is the core of the oil transportation industry [1]. In the transportation process of wax-containing crude oil, due to its unique physical properties, it shows poor fluidity in low-temperature environments, and the transportation process generally requires a heating transportation process [2]. This transportation requirement makes the waxed crude oil transportation pipeline need more energy consumption compared with the conventional pipeline, which mainly contains two aspects on the one hand, the pumping unit for the pipeline pressurization consumption of electric energy; on the other hand, is the heating furnace for the heating of the crude oil consumption of thermal energy [3]. Therefore, in order to respond to corporate energy conservation and emission reduction policy requirements and ensure the safe and economical operation of oil pipelines, accurately grasping the changes in pipeline temperature and pressure is the basis for pipeline simulation and optimization operations [4].

In crude oil pipeline transportation systems, for effective management and operation optimization, the traditional parameter prediction mainly includes two major categories: process calculation

method and statistical method [5]. The process calculation method establishes relevant mechanism models based on kinetic and thermodynamic analyses [6], calculates relevant parameters according to the pipeline's process, and optimizes the pipeline's operation plan [7]. Commonly used pipeline parameter simulation software using physical models include OLGA [8], SPS [9], and PIPESIM [10]. Due to the calculation process involving station equipment and pipeline, many parameters cannot be accurately obtained, affecting the accuracy of the predicted parameters. The statistical method is based on the pipeline historical production data on the parameters of the regression prediction do not need to take into account the complexity of the actual process [11]; the construction process is relatively simple, commonly used statistical analysis methods, including linear regression (LR) [12], autoregressive integral moving average model [13], gray prediction model method [14], and so on. The prediction accuracy increases with the increase of data, so the amount of data required is large, and the prediction accuracy is limited [15].

To overcome the limitations of traditional parameter prediction methods, machine learning methods have better advantages in the field of process prediction. For example, Yarveicy et al. [16,17] used different machine learning methods to predict the minimum initial temperature required for natural gas after throttling and the solubility of CO₂ in an aqueous Piperazine (PZ) solution, respectively, and both of them achieved good predictions, which are important for practical production. In addition, they applied Extra Trees combined with the least squares support vector machine

¹Corresponding author.

Contributed by the Petroleum Division of ASME for publication in the JOURNAL OF ENERGY RESOURCES TECHNOLOGY. Manuscript received March 12, 2024; final manuscript received June 30, 2024; published online July 26, 2024. Assoc. Editor: Roberto Carapellucci.

(LSSVM) algorithm to model the phase equilibrium of gas hydrate and found that Extra Trees is better than LSSVM in prediction accuracy, which is of significant application value for the prediction of gas hydrate phenomenon [18]. In the prediction of pipeline operating parameters, Sun et al. [19] developed a split systems approach-based model to predict pipeline reliability, which effectively helps one to make optimal pipeline maintenance decisions. Su et al. [20] integrated deep learning techniques with controllability theory to forecast the operational parameters of natural gas pipelines, achieving an average prediction accuracy of 0.99. Peng et al. [21] proposed a combination of support vector regression, principal component analysis, and chaotic particle swarm optimization model, which is effective in predicting the corrosion rate of multiphase flow pipelines. Temperature and pressure are some of the most important parameters in pipeline production and operation; machine learning methods are of great significance in realizing their accurate prediction, which helps to ensure the safe operation of pipeline systems. Shadloo et al. [22] utilized a multilayer perceptron neural network model to forecast the pressure drop in horizontal pipelines, yielding superior accuracy when compared to conventional pressure drop computation techniques. Zhang et al. [23] assessed the predictive performance of three neural networks (backpropagation neural network (BPNN), radial basis function neural network (RBFNN), and general regression neural network) for pipeline temperature and pressure and identified the most effective one. The utilization of the RBFNN neural network for forecasting leads to a reduction in the overall cost of energy by up to 10.75%. Li et al. [24] developed a three-dimensional computational model for oil pipelines in cold regions to predict the oil temperature along the pipeline, which effectively reduces the risk of pipeline freezing and plugging. Wei et al. [25] conducted a comparison between the BPNN model and the backpropagation (BP) neural network model improved by the particle swarm method to analyze pipeline pressure drop. They ultimately implemented the PSO-BP model and effectively attained an accurate prediction of pressure drop.

Although the previous process of predicting oil pipeline parameters has achieved good prediction accuracy, most models only focus on the relationship between input and output, resulting in a lack of physical meaning in the calculation process and biased predictions for some specific operating conditions [22–26]. To address this limitation, more and more studies have focused on combining mechanistic models with machine learning to achieve accurate prediction [27]. Zheng et al. [28] proposed a theory-guided short-term and long-term memory model for pipeline shutdown pressure prediction, which not only exhibits good prediction performance but also follows the physical principles and engineering theories. Zhang et al. [29,30] make full use of the advantages of both mechanistic models and machine learning to predict liquid retention and pressure drop in gas–liquid two-phase flow of pipelines and energy consumption of crude oil pipelines, respectively, and effectively improve the prediction accuracy of the model. Yuan et al. [31] combined a mechanistic model based on the Austin–Palfrey equation with machine learning to predict the length of the pipeline mixing section, which not only improved the accuracy of prediction but also increased the ability to recognize outliers. However, there are still relatively few studies combining mechanistic modeling and machine learning in the prediction of oil pipeline temperature and pressure.

Therefore, in order to fully integrate the advantages of mechanistic modeling and machine learning methods, a physically guided neural network (PGNN) model is developed in this paper to predict the temperature drop and pressure drop of oil pipelines. The model adds the physical intermediate variables affecting the temperature and pressure of the oil pipeline as neurons to the loss function and optimizes the model parameters using an adaptive moment estimation algorithm, which has a good prediction performance. It not only makes up for the shortcomings of the mechanistic model but also provides solid theoretical support for the prediction. The paper is organized as follows: Sec. 2 introduces the adopted

PGNN model and its background. The data sources, as well as the processing, are described in Sec. 3. Section 4 assesses the accuracy value and usefulness of the model predictions and, finally, explains the results of the model predictions and summarizes the research work in the paper.

2 Research Methods

2.1 Modeling of Temperature and Pressure Loss Mechanisms in Crude Oil Pipelines. For hot oil transportation pipelines in alpine regions, without considering the impact of frictional heat generation on the temperature drop of the pipeline, the mechanism model of the temperature drop can generally be calculated using the Sukhov formula, which is expressed as follows:

$$t_L = t_0 + (t_R - t_0)e^{-aL} \quad (1)$$

$$a = \frac{K\pi D}{Gc_h} \quad (2)$$

In the formula, G is the mass flowrate of oil, kg/s; c_h is the specific heat capacity of the oil, J/(kg °C); D is the pipe outer diameter, m; L is the length of pipe, m; K is the total heat transfer coefficient of the pipe, W/(m²·°C); t_R is the starting temperature of the pipeline, °C; t_L is the temperature at distance L from the starting point, °C; t_0 is the soil temperature around the pipe, °C; and a is a parameter.

The total heat transfer coefficient K value is a crucial element in the formula above, impacting the temperature drop. Testing is challenging because of the intricate environment of underground pipelines and the heat transmission process, which comprises the three primary stages of hot oil pipeline–soil. This study utilizes real production data and uses the formula inverse method to determine the K value in the oil pipeline temperature drop model.

Pressure is a crucial factor in ensuring the safe functioning of oil pipelines. The formula for pressure energy loss is as follows:

$$H = h_f + h_c + (Z_m - Z_c) \quad (3)$$

In the formula, H is the pressure drop in the pipeline, m; h_f is the friction loss along the pipeline, m; h_c is the total local friction loss, m; Z_m is the elevation of the endpoint, m; and Z_c is the elevation of the initial point, m.

The friction resistance h_f along the path can be determined using Darcy's formula, which is expressed as follows:

$$h_f = \lambda \frac{L}{D} \frac{v^2}{2g} \quad (4)$$

In the formula, λ is the hydraulic friction coefficient; D is the pipe diameter, m; L is the length of the pipe, m; V is the flowrate, m/s; and g is the acceleration due to gravity, m/s².

The friction loss occurring within the local section of the long-distance pipeline, as mentioned in this paper, often accounts for only 1–2% of the overall friction loss. Additionally, the pipeline is situated on a level terrain, so the decrease in elevation can be disregarded. Therefore, the friction loss along the way can be used as the actual value of the pipeline pressure drop, and the hydraulic friction coefficient also adopts the inverse algorithm as the training data for predicting the oil pipeline pressure drop model.

2.2 Mechanism-Based Prediction Model Construction. Artificial neural networks exhibit self-organization, self-learning, associative memory, and rapid problem-solving capabilities. This model mimics the functioning of the human brain and is extensively applied in research across different fields [32,33]. Neurons, serving as the fundamental processing units of artificial neural networks, achieve the learning and representation of intricate models through connections, learning mechanisms, and activation

functions [17]. The mathematical model is represented as follows:

$$y = b(w_1x_1 + w_2x_2 + \dots + w_nx_n + e) \quad (5)$$

In the formula, b is the activation function; w_i is the weight of the i th input, $i \in [1, n]$; x_i is the i th input, $i \in [1, n]$; e is the response sensitivity of neural units to signals; and y is the output of the neural unit.

The traditional artificial neural network model consists of three main parts: the input layer, the hidden layer, and the output layer [34]. Figure 1 displays the structure.

Traditional neural networks learn patterns from data without prior knowledge of physical principles, which can lead to conflicts between the findings of the data-driven model and the physical mechanism [35]. Outliers in the input data might lead to erroneous prediction findings and hinder generalization performance with fresh data. This research forecasts the temperature and pressure of oil pipelines by incorporating the mechanism into a neural network to ensure that the prediction results align with changes in physical consistency. The total heat transfer coefficient and hydraulic friction coefficient were back-calculated using historical production data, temperature drop, and pressure drop mechanism model to provide data for training the neural network model. Choose the mass flowrate, ambient temperature, and pipeline starting point temperature as the input layer of the neural network. Utilize the hidden layer to capture the non-linear relationship between the input layer and the λ and K values. The results are sent to the P-T layer, where the appropriate pipelines and oil parameters are entered. The parameters determine the pressure and temperature at the pipe's outlet. The neural network model structure (PGNN) based on the physical modeling method of pipeline temperature decline and pressure drop is illustrated in Fig. 2.

2.3 Loss Function Optimization. The loss function is an important component of the neural network model and generally

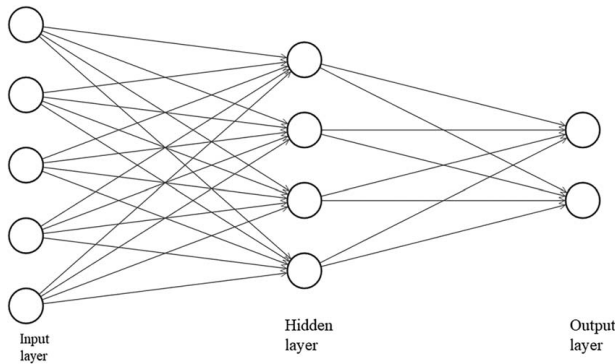


Fig. 1 Three-layer artificial neural network structure

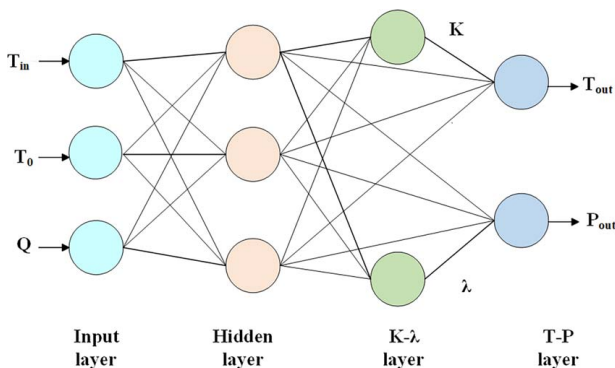


Fig. 2 Neural network structure diagram combining mechanism

responds to the difference between the true value and the predicted output. The objective of a prediction model is to reduce the error induced by the loss function while undergoing training [36]. Traditional neural networks' loss function prioritizes the output value and overlooks the influence of intermediate factors on prediction accuracy. This research analyzes the hydrothermal properties of oil pipelines. It includes the total heat transfer coefficient and hydraulic friction coefficient in the loss function as physical intermediate variables to predict temperature drop and pressure drop. The loss function has been optimized as follows: Eq. (6) represents the loss function for the temperature drop prediction model, whereas Eq. (7) represents the loss function for the pressure drop prediction model.

$$\text{LOSS}_{\text{MSE},T} = \partial_k \text{MSE}(K, \hat{K}) + \partial_T \text{MSE}(T_{\text{out}}, \hat{T}_{\text{out}}) \quad (6)$$

$$\text{LOSS}_{\text{MSE},P} = \partial_\lambda \text{MSE}(\lambda, \hat{\lambda}) + \partial_P \text{MSE}(P_{\text{out}}, \hat{P}_{\text{out}}) \quad (7)$$

In the formula, MSE is the mean square error; \hat{x} is an estimate of x ; ∂_k , ∂_T , ∂_λ , and ∂_P are the weights of the error term.

2.4 Parameter Optimization Using the Adaptive Moment Estimation Algorithm.

Due to the introduction of four error term weight values in the aforementioned loss function, the adaptive moment estimation algorithm (Adam) method is employed to improve the PGNN model by adjusting parameters. Adam is an optimization algorithm that was introduced by Kingma and Ba [37]. The algorithm combines the concept of moment estimation by calculating the first-order and second-order moments of the gradient to design adaptive learning rates for different parameters. This effectively balances convergence speed and stability and is primarily used for updating neural network parameters [38].

The Adam approach incorporates momentum and considers past gradient information by computing the exponential moving average of the gradient. The gradient's first-order moment estimation updating rule is as follows:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (8)$$

In the formula, m_t is the momentum at time-step t ; g_t is the current time-step gradient; and β_1 is the momentum attenuation coefficient, usually 0.9.

To ensure different learning rates in different weight directions, the update rule for adjusting the learning rate using the second-order moment estimate of the calculated gradient is as follows:

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (9)$$

In the formula, v_t is the exponential moving average of the squared gradient at time-step t ; β_2 is the attenuation coefficient of the squared gradient, usually 0.999.

Based on the above momentum term and adaptive learning rate, the update expression of the weight is as follows:

$$\theta_{t+1} = \theta_t - \frac{\alpha \cdot m_t}{\sqrt{v_t} + \epsilon} \quad (10)$$

In the formula, θ_t is the weight of time-step t ; α is the learning rate, usually 0.001; ϵ is a small constant added for numerical stability, usually 10^{-8} .

The error term weights (∂_k , ∂_T , ∂_λ , ∂_P) in the loss function of the PGNN model are chosen as the decision variable to minimize the loss function error. The flowchart of the optimization type using Adam's algorithm is shown in Fig. 3.

3 Data Preprocessing

3.1 Data Source. The research data for this paper come from a long-distance thermal oil pipeline. The data are gathered via authentic on-site operation reports, which mostly consist of operating data such as pipeline pressure, temperature, transportation volume, and

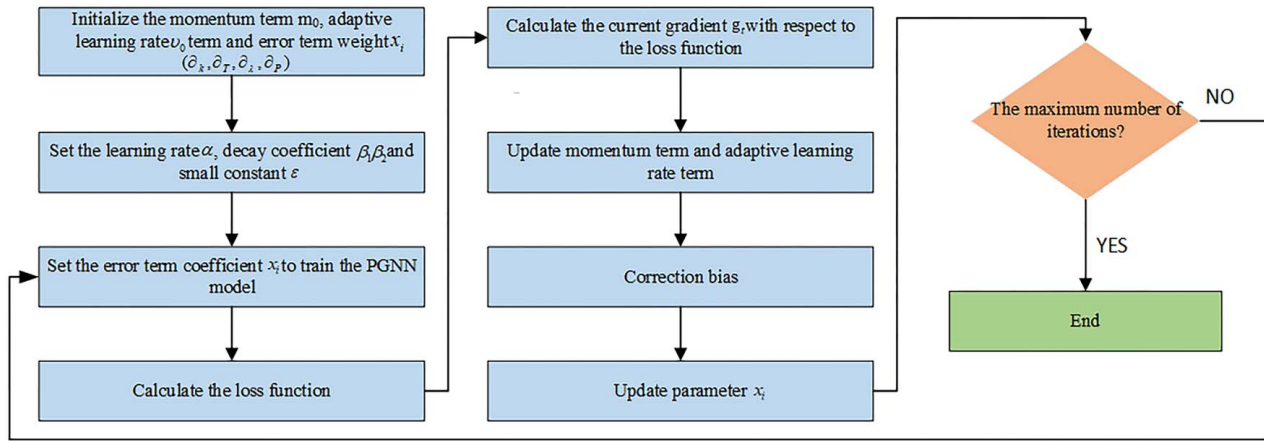


Fig. 3 Adam optimizes the neural network flowchart

ambient temperature. Table 1 displays the specifications of the pipeline and the characteristics of the oil. The actual operating data of the pipeline are shown in Table 2.

3.2 Noisy Data Processing. When collecting data, due to the many factors involved, data that do not conform to the actual operating conditions may be generated, thereby affecting the training effect of the model. Therefore, in order to improve data quality, wavelet denoising technology is used to smooth and denoise the data collected by oil pipelines.

The wavelet denoising approach efficiently addresses noise from various frequencies by decomposing and reconstructing the frequency domain, thus enhancing data correctness [39]. The threshold function in wavelet denoising is influenced by the noise level of the original data and the signal length n . Due to the limited data obtained on site within the safe production range, a fixed threshold principle is used because the fluctuation range is tiny. The threshold T expression is as follows:

$$T = \sqrt{2 \log(n)} \quad (11)$$

Typical threshold processing techniques consist of soft thresholding and strong thresholding. Hard thresholding highlights the distinctiveness of data by directly setting wavelet coefficients that are less than the threshold to zero. Contrary to this, soft thresholding decreases wavelet coefficients that are smaller than the threshold by

Table 1 Pipeline and oil parameters

Number	Parameter (Unit)	Value
1	Pipeline length (m)	31,360
2	Pipeline outer diameter (mm)	630
3	Pipeline wall thickness (mm)	8
4	Oil density (20 °C) (kg/m ³)	871.3
5	Oil freezing point (°C)	35
6	Oil viscosity (mPa·s)	25.5–931
7	Oil specific heat capacity (J/kg·K)	1745–1962

Table 2 Actual pipeline operating parameters

Parameter (Unit)	Minimum	Maximum	Average
Throughput (t)	7824	29,242	18,142
Environmental temperature (°C)	−7.6	15.36	3.9
Temperature drop (°C)	2.7	11.5	5.2
Pressure drop (MPa)	0.3	1.5	0.82

reducing their amplitude, making it more appropriate for preserving certain relatively tiny amplitude features in the data.

Considering that the neural network training process requires a diversity of data, we choose to use a hard threshold method for processing. The processing flowchart of the wavelet denoising method is shown in Fig. 4.

3.3 Data Set Partitioning. The oil pipeline operation data for three consecutive years (36 months) are selected as the data set, and each piece of data is recorded daily. The training set and the test set are divided at a ratio of 8:2. In view of the fact that different division methods may cause differences in the distribution of sample data and the overall data, the data set was divided using the random sampling method and the stratified sampling method [40], respectively, and the optimal division method was selected to minimize the impact of the sampling error effectively.

Due to the varying throughput of distinct samples, the throughput is used as the sample characteristic for stratified sampling. Figure 5 displays the sample distribution created using stratified sampling and random sampling. The error bar represents the sample's standard deviation.

Stratified sampling improves the model's generalization ability on unseen data by precisely reflecting the properties of the total data, as shown in the figure.

3.4 Data Normalization. Before training the neural network model, it is customary to normalize the data to guarantee that data with different characteristics are on a consistent scale [41], often by mapping the data set to the value range of [0, 1]. The research

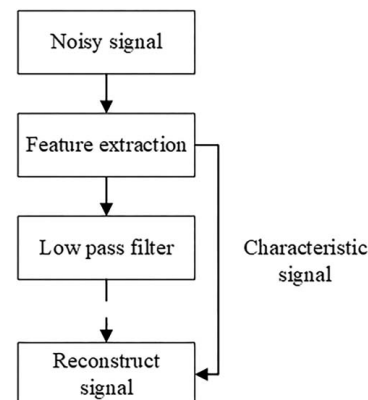


Fig. 4 Adam optimizes the neural network flowchart

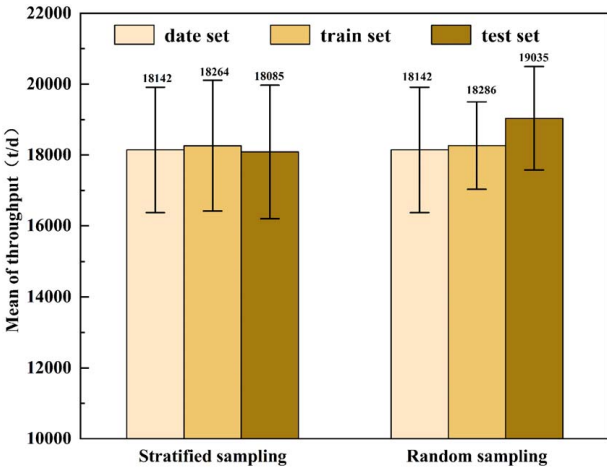
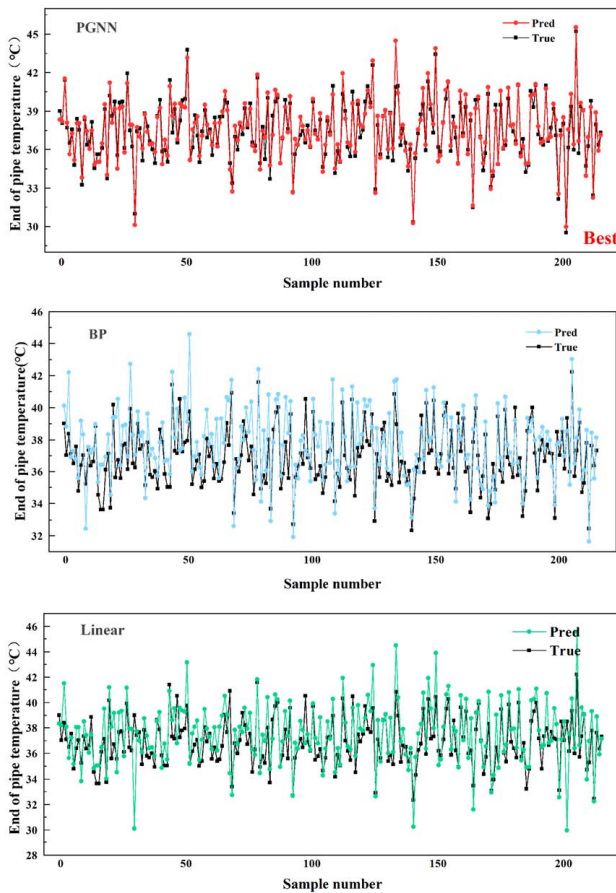


Fig. 5 Stratified sampling and random sampling mean

utilizes linear normalization to standardize features and improve model accuracy through the use of normalized data. The formula for linear normalization is expressed as follows:

$$x' = \frac{x - A_{\min}}{A_{\max} - A_{\min}} \quad (12)$$

In the formula, x' is the normalized data; A_{\max} is the maximum value in data set A; and A_{\min} is the minimum value.



4 Results and Analysis

4.1 Model Evaluation Criteria. Selecting suitable evaluation metrics is crucial for developing top-notch machine learning models. Prediction accuracy and training time are typically regarded as the most crucial variables among all evaluation indicators. Due to the limited training data for the pipeline temperature drop and pressure drop prediction model and the permitted training time, we will prioritize assessing the prediction accuracy index. The evaluation indicators in this paper for determining if the model calculation results can direct on-site production are the root mean square error (RMSE) and the coefficient of determination (R^2) [18]. The calculating formula is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (14)$$

In the formula, N is the number of samples; \hat{y}_i is the predicted value; y_i is the true value; and \bar{y} is the mean.

A reduced RMSE indicates a higher degree of accuracy in the computation result. A higher value of R^2 indicates a stronger correlation between the variables, resulting in a more accurate fitting effect. A commonly held belief is that a model with a goodness of fit of more than 0.8 is considered to be quite high.

4.2 Model Prediction Accuracy. To assess the prediction performance of several models, LR, BPNN, and PGNN are employed to predict the same dataset. The Adam algorithm is utilized to

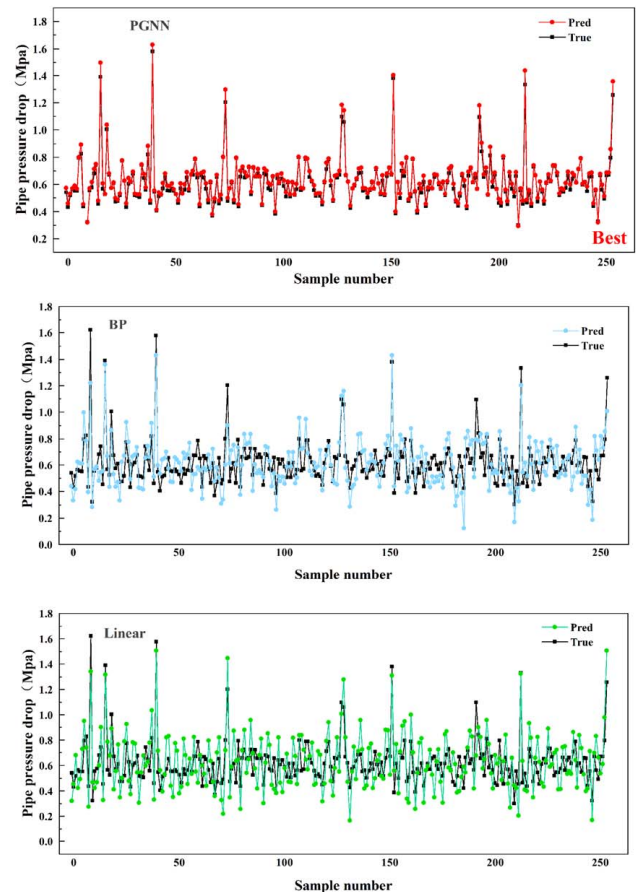


Fig. 6 The predicted value of temperature and pressure drop at the pipeline endpoint

Table 3 Different model performance indicators (temperature)

Model	RMSE (°C)	R^2
PGNN	0.066	0.987
BP	0.184	0.841
LR	0.203	0.797

Table 4 Different model performance indicators (pressure)

Model	RMSE (MPa)	R^2
PGNN	0.071	0.946
BP	0.195	0.832
LR	0.227	0.780

optimize the models' parameters. Figure 6 displays the forecast accuracy of each model, while Tables 3 and 4 show the performance metrics.

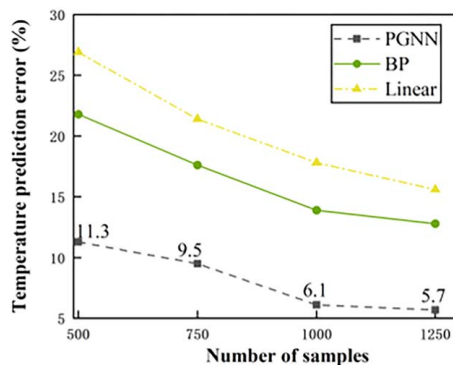
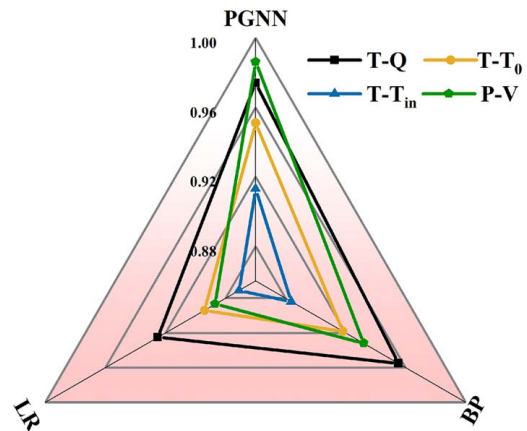
Predictions were tested on datasets of varying sample sizes to thoroughly assess the model's prediction accuracy. Figure 7 displays the average prediction error for each model.

A comparison of the predicted and actual temperature and pressure values for each model in Fig. 6 shows that the values predicted by the PGNN model are closer to the actual test set data. In addition, the PGNN model has the smallest RMSE and the largest R^2 in its prediction performance, both of which are better than the other two models.

From Fig. 7, it can be seen that the PGNN model can also achieve good prediction results with small sample data and minimize the average prediction error. Hence, the PGNN model is superior in forecasting the temperature and pressure of oil pipelines.

4.3 Physical Consistency. The decrease in temperature in the pipeline is inversely proportional to the transmission volume (Q) and the ambient temperature (T_0) and directly proportional to the beginning point temperature (T_{in}), as shown in Eqs. (1)–(4). The pressure loss in the pipeline increases in direct proportion to the flowrate (V). Hence, prediction outcomes that adhere to physical consistency must align with this principle. The Pearson correlation coefficient is a statistical measure used to quantify the level of correlation between two variables. The following formula determines the value:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (15)$$

**Fig. 7 Model prediction accuracy for different sample sizes****Fig. 8 Pearson coefficient of predicted results and related variables****Table 5 Model performance metrics with different loss functions added (temperature)**

Physical intermediate variables	RMSE (°C)	R^2
K	0.066	0.987
None	0.097	0.811

In order to reflect whether the predicted value of the model conforms to the physical consistency relationship, the specific calculation results of the absolute value of the Pearson correlation coefficient are shown in Fig. 8. It can be seen that the predicted result of the PGNN model has a Pearson coefficient closer to one, which is more consistent with the physical consistency requirements.

4.4 The Impact of Physical Intermediate Variables. To assess the impact of physical intermediate variables on the model, various loss functions are employed to forecast temperature and pressure. The detailed outcomes are displayed in Tables 5 and 6. The table demonstrates that incorporating physical intermediate variables into the loss function notably enhances the predictive accuracy of the model.

5 Results and Discussion

The PGNN model is established to predict the temperature and pressure of the oil pipeline, and the input data are selected from

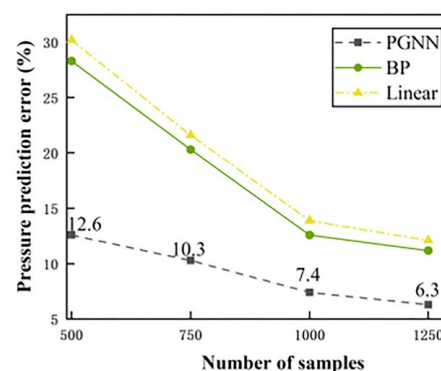


Table 6 Model performance metrics for adding different loss functions (pressure)

Physical intermediate variables	RMSE (°C)	R^2
λ	0.071	0.946
None	0.113	0.803

the actual production data of a long-distance pipeline. The results are shown in Fig. 5 by comparing the effects of the stratified sampling method and the random sampling method in the delineation of the data set. It can be clearly seen that the mean and standard deviation of the training set and test set of the stratified sampling method are closer to the actual data and better reflect the characteristics of the overall data, which lays the foundation for the subsequent training and validation of the model.

In the examination of the model prediction accuracy, the model evaluation indexes are chosen as the RMSE and the R^2 . Linear regression, backpropagation neural network, and PGNN models are used to predict the same dataset. The prediction accuracies are shown in Fig. 6, and the performance indexes are shown in Tables 3 and 4. As can be seen from Fig. 6, the values predicted by the PGNN model are closer to the actual test set data, and because of the introduction of physical intermediate variables in the model, the model is able to predict parameters with a wide range of transmission variations more accurately. The data in Tables 3 and 4 show that the PGNN model has the smallest predictive performance index RMSE for temperature and pressure, 0.066 and 0.071, respectively, and the largest R^2 , 0.987 and 0.946, respectively, which indicates that the model is able to accurately predict the parameters, proving the effectiveness of the PGNN model.

In order to explore the effect of different sample sizes on the prediction performance of the models, the average prediction errors of the three models under different sample sizes were compared. As shown in Fig. 7, the prediction error of the PGNN model is much smaller than that of the BP neural network model and the linear regression model, and good prediction results can be realized under small sample data.

Figure 8 evaluates the relationship between model predictions and physical consistency using the Pearson coefficient, and the Pearson coefficient of the prediction results of the PGNN model is close to one, indicating that the prediction results are highly consistent with the physical consistency requirements.

In order to test the degree of influence of physical intermediate variables on the prediction performance of the PGNN model, the differences in prediction performance metrics of the PGNN model using different loss functions were compared, as shown in Tables 5 and 6. From the tables, it can be seen that the temperature drop and prediction model RMSE performance indexes were reduced by 0.031 and 0.042, respectively, with the addition of physical intermediate variables. Additionally, the R^2 performance indexes are improved by 0.176 and 0.143, respectively. These changes indicate that the introduction of physical intermediate variables enhances the prediction performance of the model.

6 Conclusion

This research suggests a PGNN-based model to predict temperature and pressure in oil pipelines located in cold regions. The Adam algorithm is utilized for model parameter optimization. By incorporating physical intermediate variables as neurons in PGNN and including the prediction error of these variables in the loss function, the mechanism becomes part of the prediction model, enhancing the model's interpretability.

Three prediction models, linear regression, backpropagation neural network, and PGNN, are applied to an oil pipeline to compare the predictive effectiveness of the models. The PGNN model exhibited the highest prediction accuracy, enhanced the

model's generalization capability, and demonstrated the minimum average prediction error. The PGNN model can produce superior prediction outcomes even with limited sample sizes.

The Pearson correlation coefficient between the PGNN prediction results and variables influencing energy usage is close to one, indicating that the prediction findings align well with the requirements of physical consistency.

Comparing various loss functions on prediction accuracy reveals that incorporating physical intermediate variables into the loss function enhances the model's predictive performance.

Funding Data

- The Natural Science Foundation of Heilongjiang Province (Grant No. LH2022E025).

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

- [1] Chen, C., Li, C., Reniers, G., and Yang, F., 2021, "Safety and Security of Oil and Gas Pipeline Transportation: A Systematic Analysis of Research Trends and Future Needs Using WoS," *J. Clean. Prod.*, **279**, p. 123583.
- [2] Wang, Z., Wang, H., Zhu, C., Rui, Z., and Liu, Y., 2020, "A Novel Method for Characterizing the Aggregation of Wax Crystals and an Improvement in Wax Deposition Modeling," *ASME J. Energy Resour. Technol.*, **142**(10), p. 103003.
- [3] Liu, D., Liu, Y., Lai, N., Ji, Y., and Hu, T., 2022, "A New Method for Calculating the Inflection Point Temperature of Heavy-Oil Rheology Transforming From Non-Newton Fluid Into the Newton Fluid," *ASME J. Energy Resour. Technol.*, **144**(8), p. 083001.
- [4] Gong, J., Kang, Q., Wu, H., Li, X., Shi, B., and Song, S., 2023, "Application and Prospects of Multi-phase Pipeline Simulation Technology in Empowering the Intelligent Oil and Gas Fields," *J. Pipeline Sci. Eng.*, **3**(3), p. 100127.
- [5] Xu, L., Hou, L., Zhu, Z., Li, Y., Liu, J., Lei, T., and Wu, X., 2021, "Mid-Term Prediction of Electrical Energy Consumption for Crude Oil Pipelines Using a Hybrid Algorithm of Support Vector Machine and Genetic Algorithm," *Energy*, **222**, p. 119955.
- [6] Duan, J., Wang, W., Zhang, Y., Liu, H., Lin, B., and Gong, J., 2012, "Calculation on Inner Wall Temperature in Oil-Gas Pipe Flow," *J. Cent. South Univ.*, **19**(7), pp. 1932–1937.
- [7] Zhou, M., Zhang, Y., and Jin, S., 2015, "Dynamic Optimization of Heated Oil Pipeline Operation Using PSO-DE Algorithm," *Measurement*, **59**, pp. 344–351.
- [8] Vandurangi, S. K., Lemma, T. A., Mujtaba, S. M., and Pedapati, S. R., 2021, "Determination and Analysis of Leak Estimation Parameters in Two-Phase Flow Pipelines Using OLGA Multiphase Software," *Sustain. Comput.: Inform. Syst.*, **31**, p. 100564.
- [9] Yi, X., Liping, Z., and Tingting, Z., 2014, "Research of Emergency Venting Time in Natural Gas Pipeline Accidents With SPS," *Procedia Eng.*, **84**, pp. 927–932.
- [10] Longzhu, G., Hao, C., and Tengli, Z., 2012, "Predicting the Formation of Gas Hydrate by PIPE-SIM," 2012 Fourth International Conference on Computational and Information Sciences, Chongqing, China, Aug. 17–19, pp. 1167–1170.
- [11] Mannerling, F., Bhat, C. R., Shankar, V., and Abdel-Aty, M., 2020, "Big Data, Traditional Data and the Tradeoffs Between Prediction and Causality in Highway-Safety Analysis," *Anal. Methods Accid. Res.*, **25**, p. 100113.
- [12] Panwar, B., Dhuriya, G., Johri, P., Singh Yadav, S., and Gaur, N., 2021, "Stock Market Prediction Using Linear Regression and SVM," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, Mar. 4–5, pp. 629–631.
- [13] Alabdulrazzaq, H., Alenezi, M. N., Rawajfih, Y., Alghannam, B. A., Al-Hassan, A. A., and Al-Anzi, F. S., 2021, "On the Accuracy of ARIMA Based Prediction of COVID-19 Spread," *Results Phys.*, **27**, p. 104509.
- [14] Muñoz, J., and Felicísimo, Á. M., 2004, "Comparison of Statistical Methods Commonly Used in Predictive Modelling," *J. Veg. Sci.*, **15**(2), pp. 285–292.
- [15] Jiang, P., Liu, Z., Niu, X., and Zhang, L., 2021, "A Combined Forecasting System Based on Statistical Method, Artificial Neural Networks, and Deep Learning Methods for Short-Term Wind Speed Forecasting," *Energy*, **217**, p. 119361.
- [16] Yarveicy, H., Ghiasi, M. M., and Mohammadi, A. H., 2018, "Determination of the Gas Hydrate Formation Limits to Isenthalpic Joule–Thomson Expansions," *Chem. Eng. Res. Des.*, **132**, pp. 208–214.

- [17] Yarveicy, H., Ghiasi, M. M., and Mohammadi, A. H., 2018, "Performance Evaluation of the Machine Learning Approaches in Modeling of CO₂ Equilibrium Absorption in Piperazine Aqueous Solution," *J. Mol. Liq.*, **255**, pp. 375–383.
- [18] Yarveicy, H., and Ghiasi, M. M., 2017, "Modeling of Gas Hydrate Phase Equilibria: Extremely Randomized Trees and LSSVM Approaches," *J. Mol. Liq.*, **243**, pp. 533–541.
- [19] Sun, Y., Ma, L., and Morris, J., 2009, "A Practical Approach for Reliability Prediction of Pipeline Systems," *Eur. J. Oper. Res.*, **198**(1), pp. 210–214.
- [20] Su, H., Zio, E., Zhang, J., Yang, Z., Li, X., and Zhang, Z., 2018, "A Systematic Hybrid Method for Real-Time Prediction of System Conditions in Natural Gas Pipeline Networks," *J. Nat. Gas Sci. Eng.*, **57**, pp. 31–44.
- [21] Peng, S., Zhang, Z., Liu, E., Liu, W., and Qiao, W., 2021, "A New Hybrid Algorithm Model for Prediction of Internal Corrosion Rate of Multiphase Pipeline," *J. Nat. Gas Sci. Eng.*, **85**, p. 103716.
- [22] Shadloo, M. S., Rahmat, A., Karimpour, A., and Wongwises, S., 2020, "Estimation of Pressure Drop of Two-Phase Flow in Horizontal Long Pipes Using Artificial Neural Networks," *ASME J. Energy Resour. Technol.*, **142**(11), p. 112110.
- [23] Zhang, Y., Wang, B., and Huang, X., 2020, "Online Optimization of Heated-Oil Pipeline Operation Based on Neural Network System Identification," *J. Pipeline Syst. Eng. Pract.*, **11**(1), p. 04019040.
- [24] Li, G., Sheng, Y., Jin, H., Ma, W., Qi, J., Wen, Z., Zhang, B., Mu, Y., and Bi, G., 2010, "Forecasting the Oil Temperatures Along the Proposed China–Russia Crude Oil Pipeline Using Quasi 3-D Transient Heat Conduction Model," *Cold Reg. Sci. Technol.*, **64**(3), pp. 235–242.
- [25] Wei, L., Zhang, Y., Ji, L., Ye, L., Zhu, X., and Fu, J., 2022, "Pressure Drop Prediction of Crude Oil Pipeline Based on PSO-BP Neural Network," *Energies*, **15**(16), p. 5880.
- [26] Zhang, F., Zhang, Y., Li, J., Zhu, H., Zhang, J., and Qi, Y., 2023, "Pressure Drop Prediction Model of the Gas-Liquid Stratified Flow Development Section in the Horizontal Pipeline," *ASME J. Energy Resour. Technol.*, **145**(11), p. 113101.
- [27] Jin, X.-B., Robert Jeremiah, R. J., Su, T.-L., Bai, Y.-T., and Kong, J.-L., 2021, "The New Trend of State Estimation: From Model-Driven to Hybrid-Driven Methods," *Sensors*, **21**(6), p. 2085.
- [28] Zheng, J., Du, J., Liang, Y., Wang, C., Liao, Q., and Zhang, H., 2021, "Deeppipe: Theory-Guided LSTM Method for Monitoring Pressure After Multi-product Pipeline Shutdown," *Process Saf. Environ. Prot.*, **155**, pp. 518–531.
- [29] Zhang, X., Hou, L., Zhu, Z., Yang, M., Hu, Z., and Liu, J., 2024, "Prediction of Liquid Holdup and Pressure Drop in Gas-Liquid Two-Phase Flow Based on Integrating Mechanism Analysis and Data Mining," *Geoenergy Sci. Eng.*, **239**, p. 212971.
- [30] Zhang, X., Hou, L., Liu, J., Yang, K., Chai, C., Li, Y., and He, S., 2022, "Energy Consumption Prediction for Crude Oil Pipelines Based on Integrating Mechanism Analysis and Data Mining," *Energy*, **254**, p. 124382.
- [31] Yuan, Z., Chen, L., Shao, W., Zuo, Z., Zhang, W., and Liu, G., 2021, "A Robust Hybrid Predictive Model of Mixed Oil Length With Deep Integration of Mechanism and Data," *J. Pipe. Sci. Eng.*, **1**(4), pp. 459–467.
- [32] Yan, L., Zhang, H., Zhang, D., Shang, Z., Xu, H., and Qiang, G., 2024, "Separate-Layer Injection Scheme Optimization Based on Integrated Injection Information With Artificial Neural Network and Residual Network," *ASME J. Energy Resour. Technol.*, pp. 1–31.
- [33] Pazikadin, A. R., Rifai, D., Ali, K., Malik, M. Z., Abdalla, A. N., and Faraj, M. A., 2020, "Solar Irradiance Measurement Instrumentation and Power Solar Generation Forecasting Based on Artificial Neural Networks (ANN): A Review of Five Years Research Trend," *Sci. Total Environ.*, **715**, p. 136848.
- [34] Abdolrasol, M. G. M., Hussain, S. M. S., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R., Ali, J. A., Mekhilef, S., and Milad, A., 2021, "Artificial Neural Networks Based Optimization Techniques: A Review," *Electronics*, **10**(21), p. 2689.
- [35] Sleiti, A. K., Al-Ammari, W. A., Vesely, L., and Kapat, J. S., 2022, "Carbon Dioxide Transport Pipeline Systems: Overview of Technical Characteristics, Safety, Integrity and Cost, and Potential Application of Digital Twin," *ASME J. Energy Resour. Technol.*, **144**(9), p. 092106.
- [36] Rengasamy, D., Jafari, M., Rothwell, B., Chen, X., and Figueredo, G. P., 2020, "Deep Learning With Dynamically Weighted Loss Function for Sensor-Based Prognostics and Health Management," *Sensors*, **20**(3), p. 723.
- [37] Kingma, D. P., and Ba, J., 2017, ADAM: A Method for Stochastic Optimization.
- [38] Agrawal, R., Malik, A., Samuel, R., and Saxena, A., 2021, "Real-Time Prediction of Litho-Facies From Drilling Data Using an Artificial Neural Network: A Comparative Field Data Study With Optimizing Algorithms," *ASME J. Energy Resour. Technol.*, **144**(4), p. 043003.
- [39] Halidou, A., Mohamadou, Y., Ari, A. A. A., and Zacko, E. J. G., 2023, "Review of Wavelet Denoising Algorithms," *Multim. Tools Appl.*, **82**(27), pp. 41539–41569.
- [40] Mahmud, M. S., Huang, J. Z., Salloum, S., Emara, T. Z., and Sadatdiynov, K., 2020, "A Survey of Data Partitioning and Sampling Methods to Support Big Data Analysis," *Big Data Min. Anal.*, **3**(2), pp. 85–101.
- [41] Sola, J., and Sevilla, J., 1997, "Importance of Input Data Normalization for the Application of Neural Networks to Complex Industrial Problems," *IEEE Trans. Nucl. Sci.*, **44**(3), pp. 1464–1468.