processes

MDPI

Article

Transformer–Gate Recurrent Unit-Based Hourly Purified Natural Gas Prediction Algorithm

Chang Su [1], Jingcai Huang [1], Shasha Dong [1], Yuqi He [1], Ji Li [2], Luyao Hu [1], Xiao Liu [3] and Yong Liao [4,*]

[1] Chongqing Gas Mine, PetroChina Southwest Oil & Gasfield Company, Chongqing 401120, China
[2] Production and Operation Department, PetroChina Southwest Oil & Gasfield Company, Chongqing 610017, China
[3] Chongqing Jinyuyun Energy Technology Company Limited, Chongqing 400050, China
[4] School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China
* Correspondence: liaoy@cqu.edu.cn

Abstract: With the rapid development of industrial automation and intelligence, the consumption of resources and the environmental impact of production processes cannot today be ignored. Today, natural gas, as a commonly used energy source, produces significantly lower emissions of carbon dioxide, sulphur dioxide, and nitrogen oxides from combustion than coal and oil, and can be further purified to remove the small amount of impurities it contains, such as sulphur compounds. Therefore, purified natural gas (hereinafter referred to as purified gas), as a clean energy source, plays an important role in realising sustainable development. At the same time, It becomes more and more important to dispatch purified gas resources reasonably and accurately, and the paramount factor is that the load of purified gas needs to be predicted accurately. Therefore, this paper proposes a Transformer–GRU-based hourly prediction model for purified gas. The model uses the Transformer model for data fusion and feature extraction, and then combines the time series processing capability of the Gate Recurrent Unit (GRU) model to capture long-term dependencies and short-term dynamic changes in time series data. In this paper, the purified gas load data of Chongqing Municipality in 2020 was first preprocessed, and then divided into daily and hourly load datasets according to the measurement step. Meanwhile, considering the influence of temperature factor, the experimental dataset is subdivided according to whether it includes temperature data or not, and then the Transformer–GRU model was built for prediction, respectively. The results show that, compared with the Dual-Stage Attention-Based Recurrent Neural Network (DA-RNN) and the Transformer and GRU models alone, the Transformer–GRU model exhibits good performance in terms of the coefficient of determination, the average absolute percentage error, and mean square error, which can well meet the requirement of hourly prediction accuracy and has greater application value.

Keywords: purified natural gas; transformer; GRU; predictive modelling; load forecasting

check for updates

Academic Editor: Jun-Qiang Wang

Received: 5 December 2024
Revised: 24 December 2024
Accepted: 30 December 2024
Published: 4 January 2025

Citation: Su, C.; Huang, J.; Dong, S.; He, Y.; Li, J.; Hu, L.; Liu, X.; Liao, Y. Transformer–Gate Recurrent Unit-Based Hourly Purified Natural Gas Prediction Algorithm. *Processes* **2025**, *13*, 116. https://doi.org/10.3390/pr13010116

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

With the acceleration of industrialisation and the increase in energy consumption, the problem of environmental pollution is becoming more and more serious, especially atmospheric pollution, which has a great impact on human health and the ecosystem. In order to solve the problems of environmental pollution and energy consumption, people have tried to process traditional natural gas on the basis of its further purification, in order

to remove a small amount of natural gas contained in the sulphide and other impurities; the purified natural gas can be referred to as purified gas. Purified gas is a kind of natural gas that has been processed and desulphurised by a purification device to reach a certain quality index, and at the same time some impurities, such as hydrogen sulphide, carbon dioxide, and other acid gases, have also been removed from it. As a key link in industrial production for the reduction of harmful gas emissions, it is important for environmental protection to make accurate estimations and predictions of its usage data so as to carry out reasonable and accurate scheduling for purified gas [1].

Traditional purified gas prediction methods rely on empirical models and statistical analyses, which often have limitations when dealing with complex, non-linear industrial data. In recent years, machine learning techniques, with their powerful feature learning capabilities and ability to handle complex data, have demonstrated excellent performance in a number of fields, offering new possibilities for purge gas forecasting. The use of machine learning for purified gas as well as natural gas load prediction has become a popular method; for example, reference [2] proposed a natural gas load based on a Convolution Neural Network (CNN) and bidirectional Long Short-Term Memory (LSTM). A natural gas load forecasting model based on LSTM was proposed in reference [3], a natural gas load forecasting model using Empirical Mode Decomposition (EMD) combined with a time series model was proposed in reference [4], and a natural gas load forecasting model based on the Sparrow Search Algorithm (SSA) was proposed in reference [5]. A Sparrow Search Algorithm (SSA)-optimised Recurrent Neural Network (RNN) natural gas load forecasting model was proposed in reference [6], and reference [7] proposed a natural gas load forecasting model based on a Back-Propagation Neural Network (BPN). Back-Propagation Neural Network (BPNN) algorithm-based natural gas load forecasting model was proposed in reference [8], resulting in a natural gas load forecasting model based on a Genetic Algorithm (GA)-improved BPNN algorithm; reference [9] proposed a seasonal time series analysis model based on seasonal time series analysis considering the seasonal change pattern of the natural gas load data; and reference [10] proposed a seasonal time series analysis model based on SSA-LSTM. A natural gas load forecasting model based on a seasonal time series analysis algorithm was proposed in reference [10], which firstly decomposes the seasonal trend of natural gas data, and then establishes different neural network models for the decomposed data to make forecasts. And a short-term natural gas load forecasting model based on a Deep Belief Network (DBN) was proposed in reference [11]. In reference [12], a short-term gas prediction model based on a Time Convolution Network (TCN) combined with a two-way GRU algorithm was proposed, while in reference [13] a gas prediction model based on a Wavelet Neural Network (WNN) was proposed. However, the literature outlined above still has some limitations and drawbacks. For example, when the time sequence prediction algorithm Autoregressive Integrated Moving Average Model (ARIMA) is used to predict the purified natural gas load, it is not sensitive enough to stochastic variations and its predictive performance is more limited. This type of method is usually applied to time series with a trend and no seasonal component, which makes this method less accurate due to seasonal trends in purified natural gas load. Also, regarding the literature mentioned above, it all proposes to predict daily or even yearly natural gas load data, so it cannot predict hourly natural gas load data precisely. For hourly data prediction, compared to daily or even yearly data prediction, the measured data can reduce the influence of data variance and measured error. Simultaneously, the established model is more real-time so that it is better able to meet the requirements of real-time scheduling and distribution for most industries.

Transformer, nowadays one of the more popular deep learning models, which was mainly proposed by Vaswani et al. in 2017 [14], has achieved remarkable results in the

field of natural language processing, such as machine translation, text categorisation, sentiment analysis, etc., as well as in tasks such as computer vision and speech recognition, due to its excellent parallel processing capability and long-distance dependency capture ability. As for the Gate Recurrent Unit (GRU) model, it was proposed as an improved version of CNN by Cho et al. in 2014 [15]. It effectively balances short-term and long-term memory mainly through the gating mechanism, reduces model parameters, alleviates the gradient vanishing problem, and demonstrates powerful modelling capabilities in the fields of natural language processing and time series analysis such as stock price prediction, weather prediction, and so forth.

In summary, in terms of the above algorithms that have been used for natural gas load forecasting, they all use days as the forecasting unit, with a low granularity of prediction, and at the same time, their accuracy of prediction is low, and the robustness and applicability of these models are also poor. Therefore, on the basis of the above studies, this paper proposes an hourly prediction algorithm for purified natural gas based on Transformer–GRU. The main contributions and innovations of this paper are as follows:

(1) Real purified gas dataset. The datasets used in this paper are all from the real data for Chongqing Municipality in 2020–2022, and the datasets are divided into hourly load data with hourly measurement steps and daily load data with day measurement steps, as well as a higher granularity of prediction.
(2) The fusion of data samples with multiple features to better characterise real needs. In this paper, the effect of temperature on purified gas load forecasting is considered, and historical load data combined with temperature data are jointly used as input data for model training, while historical load data are also used separately as a control in order to train the model to analyse the effect of temperature on the forecasting effect.
(3) The experimental comparison verifies that the proposed model has higher prediction accuracy. Compared with the mainstream DA-RNN [16] and the separate Transformer and GRU prediction models, the Transformer–GRU-based prediction model constructed in this paper is better in terms of prediction accuracy for purified gas load data; in terms of the coefficient of determination (R-squared, R^2) index, the proposed Transformer–GRU model is better than the DA-RNN model by about 9%, the Transformer model alone by about 4%, and the GRU model alone by about 15%; the Transformer–GRU model is better than the DA-RNN model by about 9%, the Transformer model alone by about 4%, and the GRU model alone by about 15% in terms of the mean square error (MSE). The Transformer–GRU model improves by about 38% on the DA-RNN model, by about 25% on the Transformer model alone, and by about 47% on the GRU model alone. In terms of the Mean Absolute Percentage Error (MAPE) metrics, the proposed Transformer–GRU model improves by about 80% on the DA-RNN model, by about 59% compared to the Transformer model alone, and by about 87% compared to the GRU model alone.

2. Problem Description and Analysis

Natural gas can contain varying amounts of heavier hydrocarbons, as well as non-hydrocarbon gases such as water vapour, sulphides (e.g., hydrogen sulphide), carbon dioxide, nitrogen, and helium. Natural gas is transported by pipelines, and natural gas generally contains saturated water vapour. When the temperature of the medium around the pipeline is lower than the temperature of the gas, the water vapour will condense into a liquid, or even freeze or form a hydrate, which will seriously block the valves or pipelines. As for the acidic gases contained in natural gas, such as carbon dioxide and hydrogen sulphide, they are even more harmful. The presence of water will turn these gases into acids, aggravating the corrosion of the pipe wall and reducing the service life

of the pipeline. When natural gas is used as a fuel, it can endanger people's lives. When used as a chemical raw material, these acidic gases will also poison the catalyst, reduce the catalytic effect, or even nulify the catalytic effect, affecting product quality. Therefore, crude natural gas needs to be purified to meet the requirements of life and production.

Furthermore, after dehydration and desulphurisation, natural gas still cannot be used directly as fuel gas or chemical raw material, but the mercaptans and heavier hydrocarbons in the natural gas must be separated according to certain standards and requirements, mainly by liquefying and recovering the hydrocarbons above C3. In this way, not only can we avoid many unreliable factors caused by the condensate produced during the transmission of natural gas, but we can also purify the natural gas, eliminating many factors dangerous for users in the process of use. Meanwhile, after separating the components above C3, the combustion calorific value of natural gas will be greatly improved, while the components above C3 are made into qualified liquefied gas and light oil, and better economic benefits can also be achieved.

In industry, the whole process of purifying gas, from the initial input of sulphur-containing natural gas to the final regulation and dispatch of the purified gas according to the quantity used and demanded by the user, is very complex and is roughly shown in Figure 1.

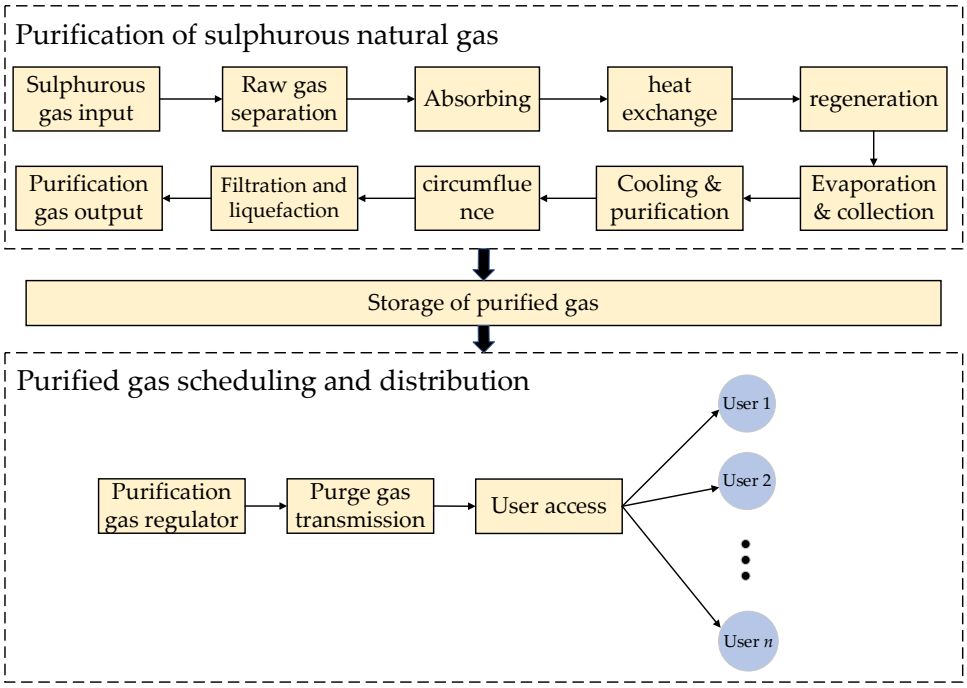


Figure 1. Purification gas treatment and scheduling process.

Sulphur-containing natural gas is first imported, and then feed gas separation is performed to remove any solid particles and liquid droplets that may be present. The separated sulphur-containing natural gas enters the absorption tower, where it is in counter-current contact with the lean solution (the regenerated desulphurisation solution). In the absorption tower, the alcohol amine (a commonly used desulphurising agent) reacts chemically with the hydrogen sulphide, carbon dioxide, and organic sulphur in the natural gas to form the rich liquor (a solution containing dissolved acidic gases). The rich liquor from the bottom of the absorption tower and the lean liquor from the top of the regeneration tower are then exchanged in a rich–rich liquor heat exchanger to recover heat and preheat the rich liquor to improve system efficiency. The preheated rich solution enters the regeneration tower, where it is heated (using a heat source) and depressurised to release dissolved acid

gases from the solution, forming flash vapour. The regenerated solution (lean solution) is recycled back to the absorption tower for further use. The flash vapour released from the top of the regeneration tower enters the flash tank where the acid gases are further separated and collected. The separated liquid then enters the solution cooler where it is cooled and recycled back to the absorption tower. The natural gas (purge gas) that has been treated in the absorption tower then enters the purge gas separator, where possible residual liquid droplets are removed before it is further cooled and purified by passing through the condenser at the top of the tower. After being separated in the purge gas separator, the purified gas enters the return collection tank, ready for subsequent delivery or use. At the same time, the liquid that may be produced during the purification process will be filtered through filters to ensure cleanliness, and then liquefied to obtain the final purified gas. The general flow of its purification process in industry is shown in Figure 2. The stored purified gas needs to be regulated according to specific user needs before being dispatched for distribution, then transported through special natural pipelines, and finally delivered to each user for their use through access points connected to each user's speciality.

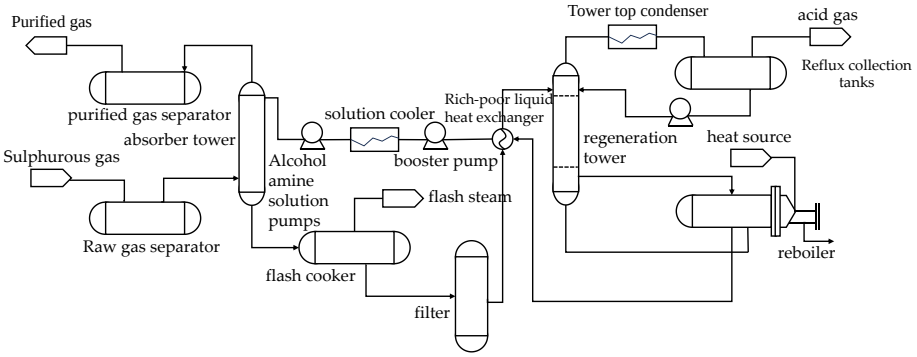


Figure 2. Natural gas purification flow chart.

Therefore, accurate prediction of users' purified gas demand as the key to purified gas pressure regulation and transport scheduling can well avoid the waste of purified gas, greatly improve the utilisation rate of purified gas, and meet well the needs of energy saving and environmental protection.

3. Transformer–GRU-Based Prediction Model

3.1. Data Preprocessing

3.1.1. Source of Data

The accurate selection and construction of data is the key to influencing whether the model can accurately predict purge gas load data. It has been shown that historical load data and temperature data have a greater impact on purified gas load prediction [17]. Therefore, this paper uses two sets of real purified gas load value data from different areas in Chongqing. One set is the daily load data from 1 July 2020 to 31 August 2022, which contain four features; namely, date, load value, and the minimum and maximum temperature of each day, with a total of 762 data points, and part of the dataset is shown in Table 1.

Table 1. Daily load data from 1 July 2020 to 31 August 2022.

Date	Load (L/h)	Highest Daily Temperature (°C)	Lowest Daily Temprature (°C)
2020/7/1	19,034,634	27	23
2020/7/2	20,748,201	28	23
2020/7/3	20,378,724	33	25
2020/7/4	21,311,935	27	24
...	...	...	...
2022/7/30	24,926,544	39	28
2022/7/31	24,881,768	39	29
2022/8/1	25,868,872	39	30

The other set is the hourly load data from 0:00 on 1 December 2020 to 24:00 on 31 July 2021 for a place in Chongqing, which contain one feature of date, hourly load value, and hourly average temperature, with a total of 5832 data points, and part of the dataset is shown in Table 2.

Table 2. Hourly load data from 0:00 on 1 December 2020 to 24:00 on 31 July 2021.

Time	Load (L/h)	Hourly Average Temperature (°C)
2021/12/1 0:00	172,193	7.18
2021/12/1 1:00	172,193	8.38
2021/12/1 2:00	172,193	10.54
2021/12/1 3:00	172,193	12.25
...	...	...
2022/7/31 21:00	152,070	25.65
2022/7/31 22:00	148,171	25.42
2022/7/31 23:00	133,366	25.59

It is clear from the analysis that natural gas load data have cyclical variability, i.e., they will show daily periodicity and cyclicity. The load values of seven consecutive days from 6 December 2021 (Monday) to 12 December 2021 (Sunday) in the hourly load data in Table 2 are taken to show the daily cyclicity of the load data, and the load changes are shown in Figure 3. It can be seen that the gas load has obvious peak and valley characteristics in the 24 h of a day, showing the trend of more gas use during the day and less gas use at night, and the load value changes of each day have similarity, and the figure presents a very obvious daily cyclical change in the natural gas load. Then, the load data of three consecutive weeks from 3 January to 23 January 2022 in Table 2, the hourly load data, are taken to show the periodicity of the load data, and its load change is shown in Figure 4, which shows that the weekly load change trend is very similar, showing obvious periodicity.

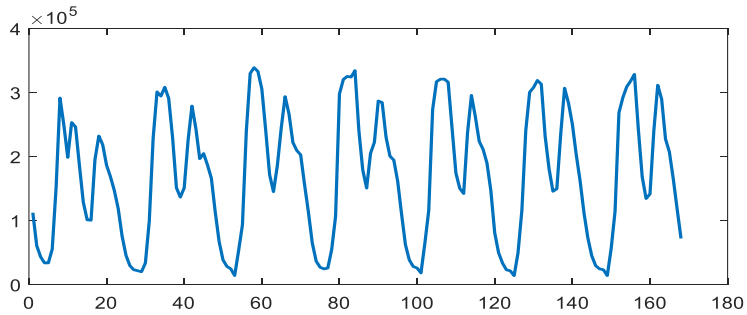
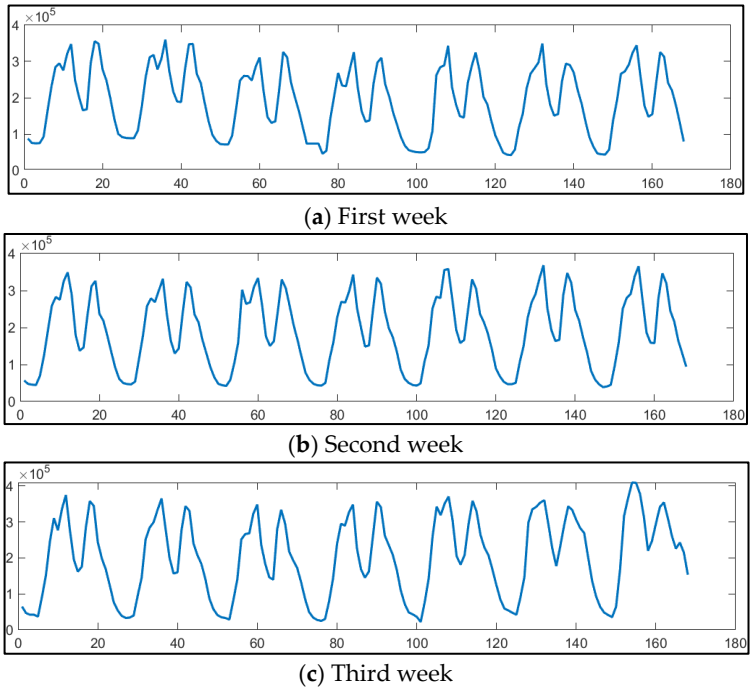


Figure 3. Load changes for seven consecutive days.



(**a**) First week

(**b**) Second week

(**c**) Third week

Figure 4. Load changes for three consecutive weeks.

However, the collected purge gas data are affected by the external environment, resulting in some anomalies in some data, while some data are affected by some special circumstances resulting in missing data, which cannot be used as input data to be directly applied to the Transformer–GRU model established in this paper. Therefore, it is firstly necessary to carry out data preprocessing on the collected purge gas load data, which are mainly divided into three parts, which are divided into the processing of missing values, the denoising of the sliding window averaging method, and the normalisation process.

3.1.2. Missing Value Handling

For missing data, we chose to fill in the data using one linear interpolation. The main principle is as follows: assuming that two points on the known dataset are $A(x_0, y_0)$ and $B(x_1, y_1)$, and then using a straight line $p(x)$ through the points A and B to represent them, we can obtain the expression of the straight line as:

$$p(x) = \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1 \tag{1}$$

Thus, the data can be populated by introducing specific values for the missing data based on the horizontal coordinates of the missing data and the expression p(x).

3.1.3. Sliding Window Averaging Denoising

After using the purge gas data as the input series for the prediction model, the sliding window averaging method is used to average the data in the sample after summing neighbouring data consecutively according to a fixed window radius as shown in the following equation:

$$P_t = \frac{\sum_{i=1}^{n} (x_{t-i} + x_{t+i}) + x_t}{2n + 1} \tag{2}$$

where $\boldsymbol{P_t}$ denotes the denoising result at moment $\boldsymbol{t}$, $\boldsymbol{x_{t-1}}$ denotes the observation at moment $\boldsymbol{t-1}$, and $\boldsymbol{n}$ represents the sliding window radius. The size of the sliding window radius may have an impact on the prediction accuracy, and too much noise and too many anomalies may be captured when the window radius is too large. Therefore, the length of

the window radius is set to 5 and the moving step is 1 to ensure that the front windows always overlap while filtering out suitable noise points.

3.1.4. Data Normalisation

Since the input purified gas data are of different orders of magnitude, in order to unify the orders of magnitude and adapt them to the neural network training process, and reduce the impact of the data scale on the results, this paper adopts the zero-mean normalisation method to preprocess the data input into the model. The mean value of the data obtained after processing is 0 and the variance is 1, which can indirectly avoid the influence of outliers and extreme values. The principle of the processing is as follows:

$$x = \frac{x_0 - \mu}{\sigma} \tag{3}$$

where $\boldsymbol{x}$ is the processed data, μ is the sample mean, and $\boldsymbol{\sigma}$ is the sample standard deviation.

3.2. Transformer–GRU-Based Prediction Model for Purified Natural Gas

Transformer is an encoder–decoder model that relies on the self-attention mechanism, and the structure is shown in Figure 5, which mainly consists of a four-layer network, with the first layer being the input data processing layer, which mainly consists of data embedding and location coding; the second layer is the coding layer, which performs feature extraction on the data through the referential-attention mechanism, which is based on the following principles:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\frac{\mathbf{Q}\mathbf{K}^{\mathrm{T}}}{\sqrt{\mathrm{d_k}}}\mathbf{V} \tag{4}$$

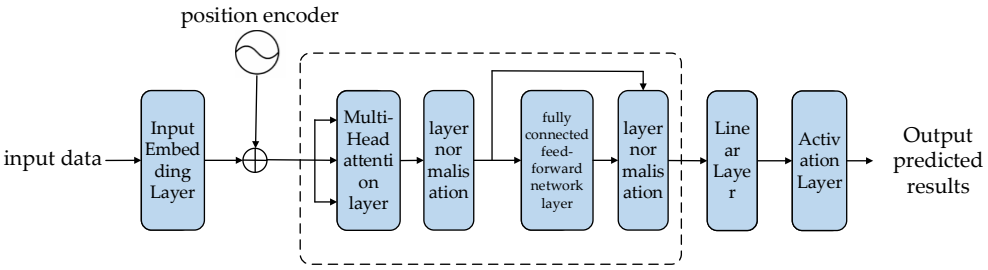


Figure 5. Transformer model structure.

Each encoder block includes a multi-head self-attention sublayer and a feed-forward fully connected neural network, in which the original $\boldsymbol{Q}$, $\boldsymbol{K}$, $\boldsymbol{V}$ are projected into low dimensions through linear layers by the multi-head attention sublayer, and then the similarity functions of different patterns are obtained by matching with the attention mechanism of m times dot product, and finally the output vectors are merged and projected back to make the model extract the global features. At the same time, residual addition and normalisation are also performed at the time of connection, to speed up the training speed and improve the stability of training. The final output can be expressed as:

$$\text{FFN}(\mathrm{x}) = \max(0, \mathrm{x}\mathrm{W}_1 + \mathrm{b}_1)\mathrm{W}_2 + \mathrm{b}_2 \tag{5}$$

The third layer is the decoding layer, which has roughly the same structure as the coding layer, and the fourth layer is the output layer, which mainly consists of one linear transformation and one activation function output.

GRU is a simplified LSTM algorithm, the main principle of which is to replace the original LSTM's forgetting, input, and output gates with reset and update gates, resulting in a simpler structure and faster training, as well as better model performance and higher computational efficiency. Its general structure is shown in Figure 6.

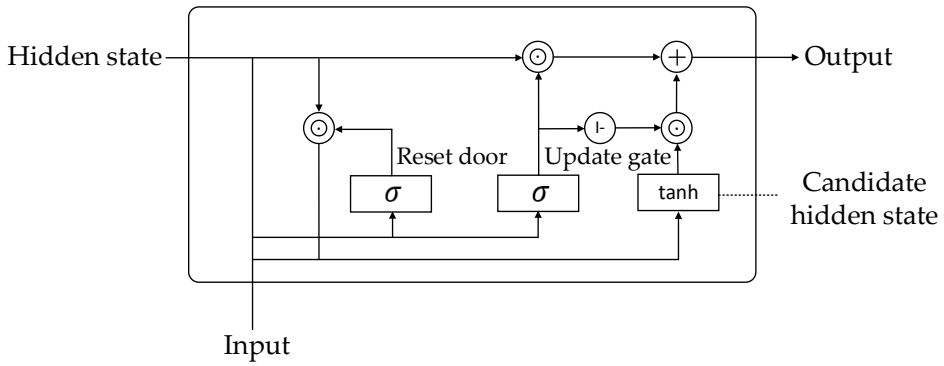


Figure 6. GRU model structure.

Therefore, for its reset gate, the main expression is:

$$r_t = \sigma(W_r * [h_{t-1}, x_t]) \tag{6}$$

It also includes a layer of candidate hidden states, which is based on the following equation:

$$\tilde{h_t} = tanh(W * [r_t * h_{t-1}, x_t]) \tag{7}$$

For its update gate, the main expression is:

$$Z_t = \sigma(W_z * [h_{t-1}, x_t]) \tag{8}$$

where σ denotes the activation function, and commonly used activation functions include sigmoid, ReLu, and tanh. Its specific updating principle is:

$$h_t = (1 - Z_t) * h_{t-1} + Z_t * \tilde{h_t} \tag{9}$$

Combining the respective advantages of the Transformer model and the GRU model, this paper establishes the Transformer–GRU model to predict the usage data of purified gas. In this model, the input data processing layer and the coding layer of the Transformer model are mainly used to process and code the input purified gas data, and then perform data fusion and feature extraction through the encoder, and then the output data features are passed as inputs to the GRU network for prediction. The GRU network reads the long- and short-term dependencies based on the input data features, so as to make accurate predictions for future time periods. For the two dataset types, the model is used to predict them separately, in which the specific construction and application offline training process is shown in Figure 7, and its online prediction process is shown in Figure 8.

3.3. Model Setup Parameters

All the specific parameter settings for the Transformer–GRU model are shown in Table 3.

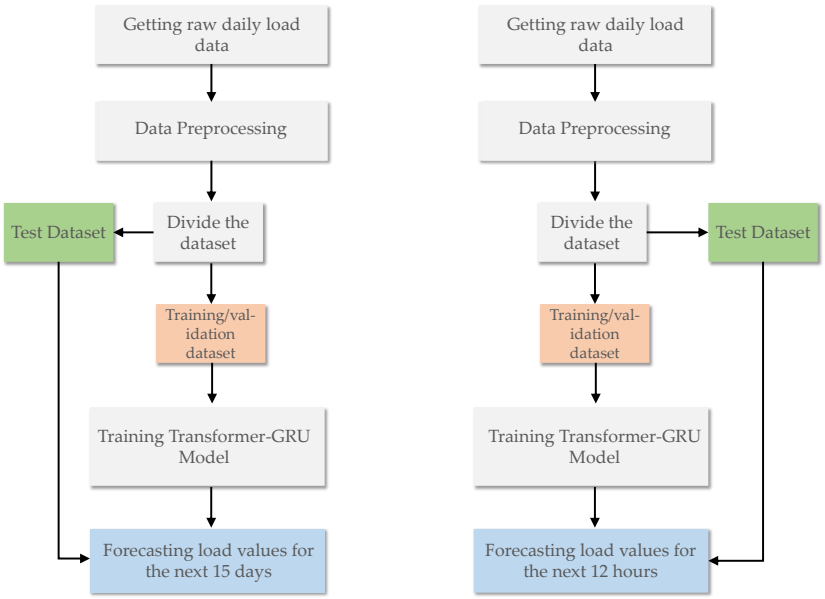


Figure 7. Flowchart of Transformer–GRU model offline training process.



Figure 8. Flowchart of Transformer–GRU model online prediction process.

Table 3. Transformer–GRU model parameter settings.

Parameter	Value
Number of encoders	2
Number of long attention spans	4
Type of activation function	ReLU
Discard rate	0.1
Initial learning rate	0.0001
Maximum number of training iterations	200
Model dimension	64
Fully connected network dimension	64
Optimiser	Adam
Number of attention factors	5
Training set ratio	0.8
Single batch size	32

3.4. Indicators for Model Evaluation

In order to assess the accuracy and validity of the model predictions, three evaluation metrics are introduced; namely, the coefficient of determination (R^2), the mean square error (MSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE), whose corresponding formulas are shown below:

$$R^2 = 1 - \frac{\sum_{i=0}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n}(y_i - \overline{y})^2} \tag{10}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{12}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{13}$$

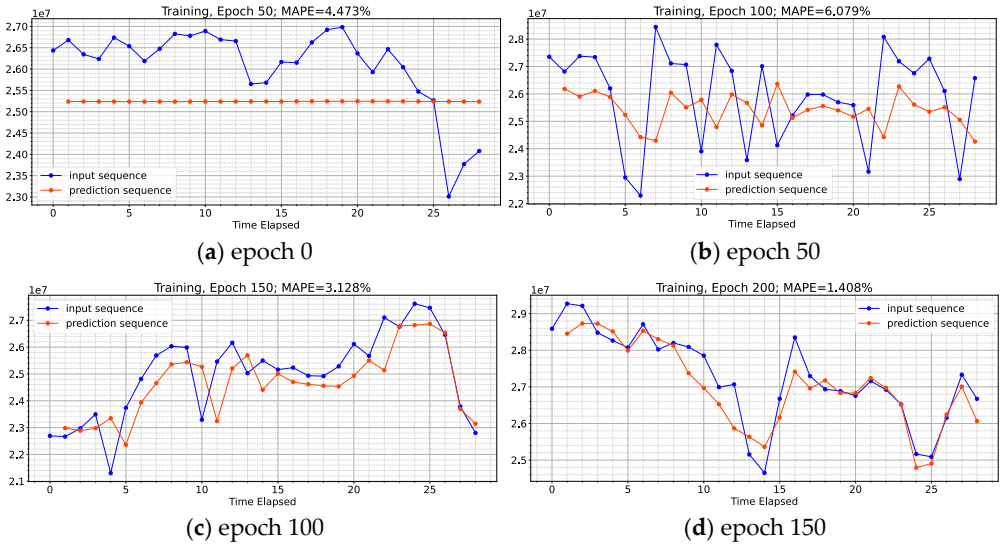
where $\hat{y}_i$ is the predicted value, y_i is the actual value, $\overline{y}$ is the mean value of the sample, and n is the number of samples. Also, for the prediction model used, the expression for the prediction accuracy can be approximated as:

$$accuracy = 1 - MAPE \tag{14}$$

4. Simulation Result and Analysis

4.1. Daily Load Forecast Results

We selected the prediction results with iterations 0, 50, 100, and 150 for analysis and the results obtained are shown in Figure 9.



(**a**) epoch 0 (**b**) epoch 50

(**c**) epoch 100 (**d**) epoch 150

Figure 9. Daily Load Forecast Effectiveness.

It is known that the prediction is best after iteration 150. Therefore, we use the final training model to predict the daily load data, and the results obtained are shown in Figure 10.

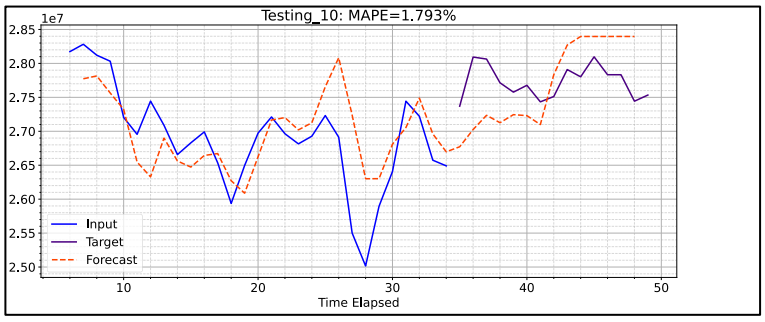
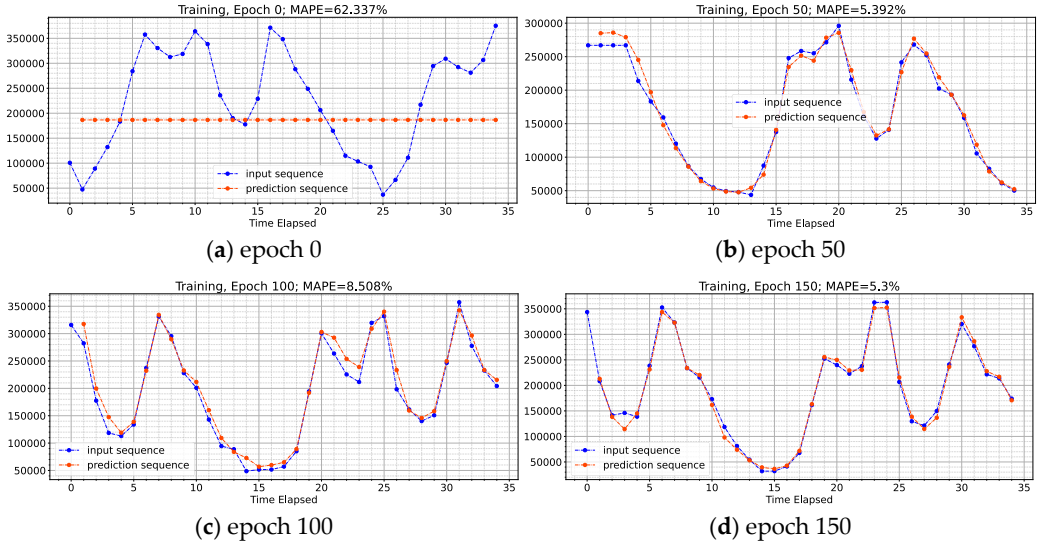


Figure 10. Daily load model forecast results.

The predicted MAPE is 1.793%, so its predictive accuracy can be calculated as 98.207%. It can be known that its prediction effect is good.

4.2. Hourly Load Forecast Results

We first disregarded the effect of temperature on the results and used only historical purge gas load data for the predictions. The predictions with iterations of 0, 50, 100, and 150 were selected for analysis and the results obtained are shown in Figure 11.



(**a**) epoch 0 (**b**) epoch 50

(**c**) epoch 100 (**d**) epoch 150

Figure 11. Hourly load forecast effect diagram.

It is known that the prediction is best after 150 iterations. Therefore, we use the final training model to predict the daily load data, and the results obtained are shown in Figure 12.

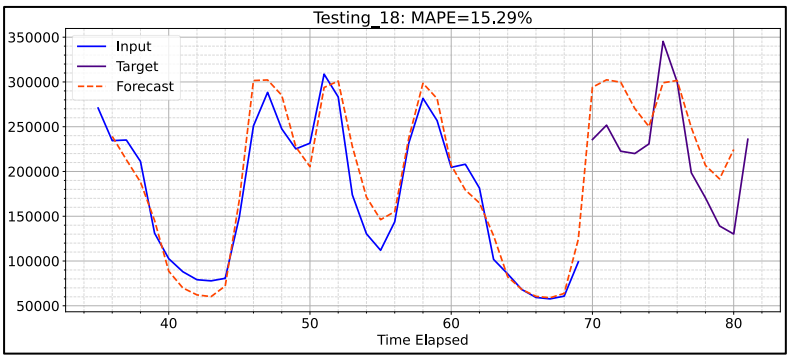


Figure 12. Modelled forecasts of hourly loads (excluding air temperature).

The predicted MAPE is 15.29%, so its predictive accuracy can be calculated as 1-MAPE = 84.71%. At the same time, we added the temperature factor, and the new prediction results obtained are shown in Figure 13.

The predicted MAPE is 8.725% and hence its predictive accuracy can be calculated as 91.275%. Next, we compare the three predictions obtained and the results of the comparison are shown in Table 4.

Table 4. Comparison of the prediction effect of different datasets.

Projected Targets	Data Characteristic	Model	MAPE	Accuracy
Daily load forecast	Load+Temperaure	Transformer–GRU	1.793%	98.217%
Hourly load forecast	Load+Temperaure	Transformer–GRU	8.725%	91.275%
Hourly load forecast	Load	Transformer–GRU	15.290%	84.710%

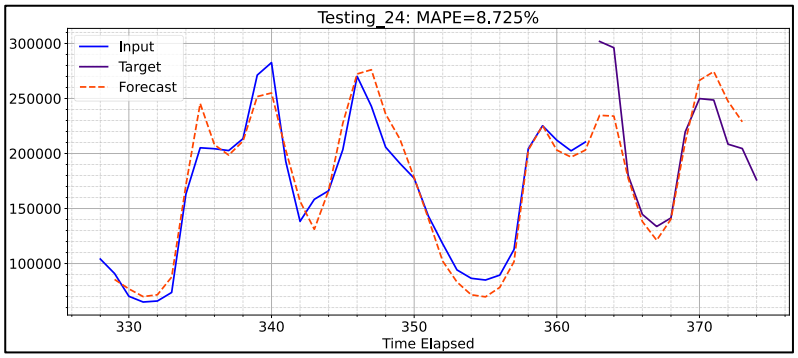


Figure 13. Modelled forecasts of hourly loads (including air temperature).

Therefore, it can be seen from the comparison that the daily load data of the purified gas data are selected for prediction with higher accuracy and better prediction, and for the hourly load data, there is a significant improvement in the prediction performance of the model after the introduction of the air temperature data as an input. Meanwhile, as a comparison, we used the DA-RNN prediction model as well as the separate Transformer model and GRU model to predict the hourly load data containing air temperature data, and the obtained prediction results are shown in Table 5.

Table 5. Comparison of model prediction effects.

Model	R^2	MSE	MAPE	Cost Time (s)
Transformer–GRU	0.90	14,816.19	8.725%	12.4
DA-RNN	0.82	23,858.46	13.587%	9.2
Transformer	0.86	19,246.59	10.384%	9.8
GRU	0.76	28,362.76	18.829%	4.6

Compared to the DA-RNN model, the Transformer—GRU model is improved by about 9%, and in terms of MSE by about 38%, and in terms of MAPE by about 80%; compared to the Transformer model alone, the Transformer–GRU model improves on it in terms of the coefficient of determination by about 4%, MSE by about 25%, and MAPE by about 59%; compared to the GRU model alone, the Transformer–GRU model improves on it by about 15% in terms of the coefficient of determination, by about 47% in terms of MSE, and by about 87% in terms of MAPE. Therefore, compared with DA-RNN and the separate Transformer and GRU prediction models, the Transformer–GRU model developed in this paper is more accurate in predicting purified gas load data, has a better prediction effect, and can be well applied in the prediction of purified gas load data.

In addition, for the computational cost, the computational time cost is measured and shown in Table 5. It is obvious that the Transformer–GRU model proposed costs more time compared to the other prediction models mentioned. For example, compared to DA-RNN model, the Transformer–GRU model costs 3.2 s more, and compared to the Transformer model, the Transformer–GRU model costs 2.6 s more. However, considering the improved prediction performance, the Transformer–GRU model still has a practical advantage. Also, with the proposed Transformer–GRU model, offline training is possible without the need for real-time online training, thus the training time cost is not the most important influence factor.

5. Conclusions and Future Work

5.1. Conclusions

This paper firstly describes the natural gas purification and storage scheduling process, and at the same time points out the necessity of purified natural gas load forecasting. Secondly, it introduces the type of dataset selected in this paper, the construction process of the Transformer–GRU prediction model, as well as the specific model parameters and evaluation indexes. Then, we compare and analyse the prediction results after applying the model to two purified gas datasets, and conclude that the prediction effect is better when using purified gas daily load data and air temperature data for prediction. Finally, this paper compares its preidiction performance with the mainstream model containing DA-RNN, Transformer, and GRU prediction models, and finds that the Transformer–GRU model developed in this paper has a significant advantage in the prediction of purified gas load.

In fact, the Transformer–GRU model proposed above can be applied not only to simplify natural gas load forecasting, but also for wind power and other energy sources with obvious time series characteristics. Thus, it can facilitate subsequent energy scheduling and supply, benefiting many energy sectors and homogeneous industries.

5.2. Future Work

In the next step, we will consider the influence of work days and non-work days on the data, and further divide the sample data into two categories, one category containing only the purified data obtained for work days and another category only containing the purified data obtained for non-work days. For the online prediction part, if the real-time purified natural gas load data are obtained on work days, they will be predicted by the Transformer–GRU model trained only for work days. Also, the optimisation of the Transformer–GRU model developed in this paper can be considered to further improve its prediction accuracy, such as by using a bidirectional GRU model and so forth.

Author Contributions: Conceptualization, Y.L. and C.S.; methodology, S.D. and Y.L.; software, X.L., J.H. and Y.H.; validation, Y.L.; formal analysis, J.L.; investigation, X.L., C.S. and S.D.; data curation, L.H.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L.; visualisation, Y.L.; supervision, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific Research and Technology Development Project of PetroChina Southwest Oil & Gasfield Company—Research on Simulation and Intelligent Dispatch Technology of Multi-gas Source and Multi-Working Conditions Pipeline Network System in Key Areas (2024D105-02-07).

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: The authors declare that this study received funding from PetroChina Southwest Oil & Gasfield Company. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

1. Ayodele, F.O.; Mustapa, S.I.; Ayodele, B.V.; Mohammad, N. An overview of economic analysis and environmental impacts of natural gas conversion technologies. *Sustainability* **2020**, *12*, 10148. [CrossRef]
2. Zhu, J.S. Research on CNN-BiLSTM Natural Gas Load Forecasting Model Based on Attention Mechanism. Mater's Thesis, Xi'an University of Architecture and Technology, Xi'an, China, 2023.
3. Zhou, N. Research on the Application of Multiple Machine Learning Algorithms in Natural Gas Consumption Load Forecasting. Mater's Thesis, Xi'an University of Architecture and Technology, Xi'an, China, 2022.

4. Fan, S.H. Research and Application of Natural Gas Load Forecasting Based on Combined Time Series Models. Mater's Thesis, Xi'an Petroleum University, Xi'an, China, 2021.
5. Liu, T. Research on Natural Gas Load Forecasting Model Based on Machine Learning. Mater's Thesis, Xi'an University of Architecture and Technology, Xi'an, China, 2022.
6. Fu, Z.H.; Huang, C.; Liu, Y.F.; Tian, W.C.; Qiao, W.B. Study of short-term natural gas load forecasting model based on WT and SSA-LSTM. *Contemp. Chem.* **2023**, *52*, 1710–1716.
7. Hu, W.; Tao, Z.; Guo, D.; Pan, Z. Natural gas prediction model based on wavelet transform and BP neural network. In Proceedings of the 33rd Youth Academic Annual Conference of Chinese Association of Automation(YAC), Nanjing, China, 18–20 May 2018; IEEE: New York, NY, USA, 2023; pp. 952–955.
8. Tang, Y. Natural Gas Load Forecasting Based on Improved Genetic Algorithm and BP Neural Network. In Proceedings of the IEEE 5th International Conference on Knowledge Innovation and Invention(ICKII), Aichi, Japan, 16–18 August 2024; IEEE: New York, NY, USA, 2022; pp. 164–168.
9. Xiang, X.; Shen, J.; Yang, K.; Zhang, G.; Qian, J.; Zhu, C. Daily natural gas load forecasting based on sequence autocorrelation. In Proceedings of the 37th Youth Academic Annual Conference of Chinese Association of Automation(YAC), Beijing, China, 19–20 November 2022; IEEE: New York, NY, USA, 2022; pp. 1452–1459.
10. Liu, Z.; Tong, J.; Ou, Y.; Zheng, Y.; Lan, R.; Zhang, Y. Short-term Gas Load Prediction with STL and Hybrid Neural Network. In Proceedings of the China Automation Congress, Xiamen, China, 25–27 November 2022; IEEE: New York, NY, USA, 2022; pp. 1386–1391.
11. Huan, J.; Hong, H.; Pan, X.; Sui, Y.; Zhang, X.; Jiang, X.; Wang, C. Short-term load forecasting of integrated energy systems based on deep learning. In Proceedings of the 5th Asia Conference on Power and Electrical Engineering, Chengdu, China, 4–7 June 2020; IEEE: New York, NY, USA, 2020; pp. 16–20.
12. Xu, D.; Zhu, D. Short-term gas load forecast based on tcn-bigru. In Proceedings of the China Automation Congress, Beijing, China, 22–24 October 2021; IEEE: New York, NY, USA, 2021; pp. 7978–7983.
13. Fu, Q.; Han, B. Study on application of wavelet neural network (WNN) model in city gas load forecasting. In Proceedings of the IEEE 4th International Conference on Cloud Computing and Big Data Analysis, Chengdu, China, 12–15 April 2019; IEEE: New York, NY, USA, 2019; pp. 317–320.
14. Vaswani, A. Attention Is All You Need. Advances in Neural Information Processing Systems. *arXiv* **2017**. [CrossRef]
15. Chung, J.; Gulcehre, C.; Cho, K.H.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
16. Qin, Y.; Song, D.; Chen, H.; Cheng, W.; Jiang, G.; Cottrell, G. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv* **2017**, arXiv:1704.02971.
17. Yang, H.M.; Pan, Z.S.; Bai, W. An overview of time series forecasting methods. *Comput. Sci.* **2019**, *46*, 21–28.

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.