

Article Medium- and Long-Term Load Forecasting for Power Plants Based on Causal Inference and Informer

Kaiyu Yang ^{1,2} and Fanhuai Shi ^{1,*}



- ² Shanghai Turbine Works Co., Ltd., Shanghai 201100, China
- * Correspondence: fhshi@tongji.edu.cn

Abstract: Accurate forecasting of power plant loads is critical for maintaining a stable power supply, minimizing grid fluctuations, and enhancing power market trading mechanisms. However, the data on power plant generation load (hereinafter abbreviated as load) are non-stationary. The focus of existing load forecasting methods has been on continuously improving the ability to capture the dependent coupling between outputs and inputs, while research on external factors, which are the causes of non-stationary load data, has been neglected. The identification of the causal relationship between external variables and load is a significant factor in accurately predicting load. In the present study, the causal effects of various external variables on load were identified and then quantitatively calculated using various methods. Based on the improved Informer model, a long-time series forecasting model, a hybrid forecasting method was proposed called causal inference-improved Informer (hereinafter abbreviated as Causal-Informer). In the present study, the mutual information method was used to remove insignificant external variables. Subsequently, external factors such as GDP, holidays, ambient temperature, wind speed, power plant maintenance status, and rainfall were selected as input features of the proposed forecasting model. Finally, the proposed Causal-Informer method was evaluated using the historical load of a power plant in East China. Compared with four popular forecasting models, measurements on Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) for the proposed method were reduced by 89.8 million kwh-672.3 million kwh, 56.8 million kwh-637.9 million kwh, and 5.1-25.4%. The proposed method achieved the most accurate and stable results. The MAPE reached 10.4% and 24.8% in 30 time steps ahead and 90 time steps ahead of forecasts, respectively.

Keywords: multi-step load forecasting; causal inference; Informer

1. Introduction

1.1. Background of the Present Study

With the development of the economy and the improvement of people's quality of life, a stable and reliable power supply becomes increasingly important to society. The full implementation of power market reform has led to the establishment of a power market trading system and intensified efforts to strengthen power market trading. Using Jiangsu Province in China as an example, according to an official document [1], Jiangsu Province has issued trading rules [2] for electricity, which allow power plants to participate in the electricity market and independently negotiate power tariffs with other market players. Therefore, it is essential for power plants to study generation load forecasts in order to develop more accurate trading strategies. For example, based on generation load forecasts, power plants can adjust their bidding strategies and develop operational strategies such as the purchase of raw materials and staffing in advance, which can reduce the production costs of power generation companies and increase their profits.

Based on the length of the forecast, power load forecasts can be classified into the following categories: long-term series forecasts, medium- and long-term series forecasts,



Citation: Yang, K.; Shi, F. Mediumand Long-Term Load Forecasting for Power Plants Based on Causal Inference and Informer. *Appl. Sci.* 2023, *13*, 7696. https://doi.org/ 10.3390/app13137696

Academic Editor: Juan P. Torreglosa

Received: 24 May 2023 Revised: 18 June 2023 Accepted: 26 June 2023 Published: 29 June 2023



Copyright: © 2023 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/).



medium-term forecasts, and short-term forecasts [3]. Power plants frequently engage in market trading and set purchase targets for a minimum duration of one month. Thus, in the present study, load for at least one month was predicted, which belongs to medium-and long-term load forecasting. The objective of the present study is to propose an accurate medium- and long-term load forecasting method to assist power plants in making informed decisions regarding the development of strategies.

Load forecasting involves integrating historical data with other relevant factors to predict load values for a specific period in the future. Load forecasting is a challenging task that is influenced by external factors such as holidays, seasons, and weather, making it non-linear and non-stationary in its typical characteristics. There are four main categories of factors which can have an impact on load: economic factors, time factors, natural factors, and uncertainties [4].

The rapid growth of the market and the government's macroeconomic regulation directly impact load. In economically developed regions, the presence of a significant number of industrial units and employees leads to higher electricity consumption compared to economically underdeveloped regions, thereby affecting load. Additionally, the industrial structure also impacts electricity consumption, and for regions with a high proportion of manufacturing industries, their electricity consumption is also greater. The time factor refers to the longer time scales in which load trends usually show cyclical and regular changes, such as holidays and seasonal variations. Obviously, load prediction is affected by natural factors such as ambient temperature, wind speed, and humidity. Moreover, its relevance is more obvious in certain extreme natural environments, such as cold winters and hot summers, which have a more significant impact on load variations. Uncertainties are unpredictable future events that affect the normal operation of the power system.

The performance of proposed method was evaluated in comparison with popular medium- and long-term load forecasting methods. The innovations and key contributions of the present study are as follows:

- Causal inference was introduced into the time series prediction scenario to enhance the interpretability and robustness of the model. To the best of my knowledge, this is the first time that causal learning has been introduced in the field of medium- and long-term load forecasting to improve forecasting.
- The Informer algorithm was improved to make it more suitable for predicting nonstationary data in real scenarios.
- A novel prediction model, Causal–Informer, was proposed and compared with other popular forecasting approaches.

The remaining sections are in the following order: The preliminary work is introduced in the second section. Then, the proposed method is introduced in detail in the third section. The overall procedure and experiment analysis are described in the fourth section. The method was discussed with those in the popular art. Additionally, extensive experiments were conducted to show the effectiveness and efficiency of the present method.

1.2. Existing Research Gap

Early power load forecasting tasks relied on the expertise and mathematical skills of practitioners [5], which had advantages in terms of model interpretability. However, in the current era of big data, such methods perform poorly in dealing with large and complex volumes of data. Data-driven forecasting techniques are gaining prominence due to the rising advancement of machine learning methods, including Seasonal Auto-Regressive Integrated Moving Average (SARIMA) [6], Support Vector Machine (SVM) [7], and iterative ResBlocks [8].

The aforementioned methods are currently mainly applied to short-term forecasting, relying on the ability of machine learning algorithms to fit input–output variables, which often yields good results. Despite such results, when the described methods are directly applied to medium- and long-term power plant generation load (hereinafter abbreviated as load) forecasting, the results may not be satisfactory. To solve the problem of medium-

and long-term load forecasting, some scholars have proposed advanced methods. In one study [9], the load data were decoupled and a novel method was used to convert the tabular data into image data to finally achieve medium- and long-term load forecasting for distribution transformers. Variables such as outdoor temperature, humidity, and wind speed were used as inputs to achieve multi-time step forecasting of heating load based on Informer [10]. An improved Long Short-Term Memory (LSTM) model has also been proposed [11] and the effectiveness of the method was verified on a residential load dataset. Additionally, several researchers have expressed the belief that using a hybrid forecasting method based on the optimal combination of multiple forecasting models can help improve forecasting accuracy. A hybrid model based on Prophet and LSTM has also been proposed [12]. Through experiments, the hybrid approach was demonstrated to be able to improve the forecasting accuracy of power load.

The current literature on medium- and long-term load forecasting is sparse, primarily due to the following difficulties:

(1) Non-stationary data are typically more challenging to predict than stationary data. Load is a nonlinear variable, and external factors have a considerable impact on it [13]. Moreover, since load is affected by various external factors, such as ambient temperature, holidays, plant maintenance status, GDP, and other factors, it can be considered non-stationary data, that is, data of which the statistical properties and distribution patterns are constantly changing over time [14].

(2) Medium- and long-term forecasting is more difficult compared to short-term forecasting, and errors in forecasting tend to increase as the time step for forecasting grows longer.

(3) The correlation between various variables and load lacks qualitative analysis. The degree of influence of various variables on load is difficult to calculate quantitatively.

The aforementioned problems constrain the performance of medium- and long-term load forecasting methods. The focus of existing timing prediction methods has been on continuously improving the model's ability to couple the correlation between output and input, while research on the causes of the coupling relationship has been neglected. In the present study, Informer was used, a novel algorithm in the field of long timeseries forecasting that has shown good results on open source datasets but lacks practical applications in generation load-forecasting scenarios. Several improvements were made to Informer to adapt it to non-stationary data in real-world scenarios. In addition, the mutual information method was adopted to remove load-independent external variables. Further, the effect of different external variables on load was quantitatively investigated using causal inference as the primary method. The causal inference model was combined with the improved Informer model to obtain a hybrid model (Causal–Informer) with high interpretability and robustness.

2. Preliminaries

Since load is highly dependent on external factors such as climate and economy, the influence of external factors should be considered for accurate forecasting. However, the multiple covariance and deeper causal relationships between external factors are ignored in traditional methods. In the present study, the causal relationship between external factors and load was explored using causal inference techniques, which not only increased the performance of the model, but also enhanced its interpretability and robustness. Causal inference has already found numerous applications in fields such as medicine, sociology, and computer science [15–19], but owing to a shortage of data, causal inference, which is an emerging research area, is not yet widely used in energy forecasting. Nevertheless, it holds great potential in this field.

Despite often being used interchangeably, causality and correlation have different meanings. Correlation is a general relationship that refers to the degree of mutual influence between different variables, and this relationship is not concerned with the directionality of the influence relationship between different variables. Causality is typically divided into two components: cause and effect. The presence of a cause results in the occurrence of an effect, and the existence of an effect is, to some degree, attributable to the cause [20]. Correlational inference and causal inference differ significantly in their research focus. The latter focuses on the response of the outcome variable when the causal variable is changed and quantifies the magnitude of the response [21]. There are two common frameworks for causal analysis that are based on observed data: the potential outcome model and the structural causal model. Among the two, the design of potential outcome models is inspired by the concept of randomized controlled trials and potential outcomes in statistics. Unlike the potential outcome model algorithm, which focuses on estimating the effect of a single variable, the structural causal model aims to identify causal relationships among multiple variables. It proposes a causal function model by examining the features of the data distribution between variables resulting from causal mechanisms. The regression model is used to learn the causal functions given the data of the variables. In the present study, a structural causal model based on directed acyclic graphs (DAGs) [22] is used to construct the causal effect assessment model. Therefore, based on the PC algorithm [23] and industry experience with external variables and load, a causal diagram was first constructed, as illustrated in Figure 1. The diagram includes load and the external variables that affect load. The start of the arrow represents the 'cause' and the pointing end of the arrow represents the 'outcome'.



Figure 1. Cause diagram of external factors.

The graphical model allows for the visualization of causality and can be used to obtain joint distributions from statistical analysis of observed data. However, the exact causal relationships cannot be obtained from the graphical model alone. Several methods, that is, the intervention, need to be used to change the distribution of the existing data and thus confirm causality. Pearl [22] creatively proposed do-calculus to express the intervention. The variables to be examined are denoted as X and Y. The other covariates are combined into a vector $Z = (Z_1, Z_2, Z_3, ...)$. For simplicity, X is assumed to be a binary variable and can only take the value of 1 or 0. A set of data can be denoted as $D = (X^{(i)}, Y^{(i)}, Z^{(i)})$. P(Y = y | X = x) and P(Y = y | do(X = x)) have different meanings, with the former representing the probability distribution of Y = y conditional on X = x in the original data and the latter representing the probability distribution of Y = y when X = x is reached by intervention. Due to the presence of confounding variables, which are the variables that affect both Xassignment and Y, $P(Y = y | X = x) \neq P(Y = y | do(X = x))$ is usually true [22]. To examine the causal effect of each X on Y, it is essential to eliminate the influence of all confounding variables on the causal model. The effect of confounding variables can be eliminated by means of backdoor adjustment [21] and frontdoor adjustment [21]. For a pair of ordering variables (X,Y), backdoor adjustment can be performed using the set Z of confounding variables if the set Z satisfies certain conditions. These conditions are as follows: there are no descendants of X in the set Z, and all backdoor paths, that is, all paths between X and Y starting with the arrow of Z pointing to X, are blocked by the confounding

variables *Z*. With the backdoor adjustment, the causal effect can be calculated using the following formula:

$$P(Y|do(X)) = \sum_{z} [P(Y|X, Z = z)P(Z = z)]$$
(1)

If all of the factors on the blocking path are not observable, then the backdoor path cannot be computed. In this case, frontdoor adjustment can be used to confirm causality. Two conditions must be satisfied for frontdoor adjustment: first, all paths from *X* to *Y* have element *z*; second, all paths from element *z* to *Y* are single paths. In this case, P(Y | do(X)) can be computed using the set *Z* of *z* to complete the computation. By making the set *Z* of confusing variables on the front door path conditional, Equation (2) can be used to convert a formula with do-calculus into one without do-calculus.

$$P(Y|do(X)) = \sum_{z} \left[P(Z=z|X) \sum_{x} P(Y|X, Z=z) P(X=x) \right]$$
(2)

By intervening in the causality graph using the frontdoor adjustment and the backdoor adjustment, causal relationships can be derived from observed data and distributions alone without do-calculus. After determining the causal relationships between each node on the constructed structural causal diagram, historical data is used to assess specific causal effects and thus calculate the impact of each external variable on load.

3. Methods

As described in Section 1, load is influenced by various variables, rendering it nonlinear and non-stationary in its typical characteristics. The causal effects of different variables on load had to be considered, so as to allow for accurate prediction of future load. Causality is a type of correlation relationship. In the present study, the correlation relationships between external variables and load were first screened by the removal of low-information factors, as described in Section 3.1. Informer-based time series forecasting was implemented for medium- and long-term load forecasting. In addition, several improvements were made to the Informer model to enhance its performance and adapt it better to non-stationary data forecasting scenarios, which is presented in Section 3.2. The causal inference method, described in detail in Section 3.3, was used to determine the causal effects of different variables on load. Given the important role of external factors, the time series forecasting module was combined with the causal inference module in order to obtain accurate forecasting results. The method is explained in Section 3.4. The complete flow of the proposed method is shown in Figure 2.

3.1. Removing Unimportant Factors by Mutual Information

Mutual Information (MI) is a widely used method to measure the correlation between two variables. A higher value of MI represents a higher correlation between two variables. MI can effectively measure the correlation between variables with nonlinear relationships, which has been extensively adopted for feature selection.

Using $\mu_{X,Y}$ to refer to the joint probability density function of two random variables, the marginal density functions of these two random variables can be represented respectively as:

$$\mu_{X}(x) = \int \mu_{X,Y}(x,y) dy; \\ \mu_{Y}(y) = \int \mu_{X,Y}(x,y) dx$$
(3)

Then, H(y) is the information entropy of variable y, which describes the uncertainty of y. The entropy of y is proportional to its uncertainty. The uncertainty of Y can be obtained as:

$$H(Y) = -\int \mu_Y(y) log \mu_Y(y) dy$$
(4)





The measure of uncertainty of variable *Y* obtained after observing another variable *X* is called its conditional entropy, denoted as H(Y|X). Specifically, H(Y|X) can be defined as follows:

$$H(Y|X) = -\int \mu_X(x) \int \mu_Y(y|X=x) log\mu_Y(y|X=x) dydx$$
(5)

Then, the MI between *X* and *Y* is calculated as:

$$M(X,Y) = H(Y) - H(Y|X)$$
(6)

The correlation of various external variables with load was examined using the MI method. The results are presented in the form of a horizontal bar graph in Figure 3.



Figure 3. Chart of MI values.

The degree of correlation between external variables and load is proportional to the value of the MI, which was used in the present study to obtain the correlation ranking. After confirming with energy professionals, the correlation ranking of the external variables in Figure 3 was found to be generally consistent with empirical judgments. On the basis of the correlation ranking and empirical judgment, the proportion was excluded from the input to the model. As shown in Figure 3, temperature and servicing were the most relevant to load. Servicing is a binary variable, wherein 1 indicates that there are units in the plant under maintenance, while 0 indicates that all units in the plant are operating normally. Temperature is a continuous variable, and the causal effect of temperature on load was investigated subsequently. The remaining external variables were transformed into binary variables by setting certain conditions in order to calculate their causal effects. The specific transformation conditions are presented in Section 4.2.1.

3.2. Improved Informer Model for Non-Stationary Data

As one of the most popular artificial intelligence models, Transformer [24] has been widely used in numerous fields. However, the model is not directly applicable to the field of time series prediction due to high memory consumption, high time complexity, and the structural constraints of encoder–decoder. In order to overcome the described problems, the Informer model employs an attention mechanism that reduces the time complexity (denoted as 'O') and memory consumption to $O(L_k lnL_k)$. The dominant attention is highlighted in self-attention extraction by halving the cascading layer inputs, thereby efficiently handling the ultra-long input sequences.

The attention mechanism used in the Informer model is referred to as ProbSparse self-attention. Following the formulation in a previous study [25], the original *i*-th query's attention (denoted as 'A') is defined as a kernel smoother in a probability form:

$$\mathcal{A}(q_i, \mathbf{K}, \mathbf{V}) = \sum_j \frac{k(q_i, k_j)}{\sum_l k(q_i, k_l)} v_j = \mathbb{E}_{\mathcal{P}(k_j | q_i)} [v_j]$$
(7)

where $p(\mathbf{k}_j | \mathbf{q}_i) = \frac{k(\mathbf{q}_i, \mathbf{k}_j)}{\sum_i k(\mathbf{q}_i, \mathbf{k}_j)}$ and $k(\mathbf{q}_i, \mathbf{k}_j)$ are used to select the asymmetric exponential kernel $\frac{q_i k_j^T}{\sqrt{d}}$. Using Kullback–Leibler (denoted as '*KL*') divergence to measure the 'likeness' of attention probability $p(\mathbf{k}_j | \mathbf{q}_i)$ and uniform distribution $q(\mathbf{k}_j | \mathbf{q}_i) = \frac{1}{L_K}$, the following can be obtained:

$$KL(q||p) = \ln \sum_{l=1}^{L_k} \exp\left(\frac{q_i k_l^T}{\sqrt{d}}\right) - \frac{1}{L_k} \sum_{j=1}^{L_k} \frac{q_i k_j^T}{\sqrt{d}} - \ln L_k$$
(8)

Dropping the constant in (8), the *i*-th query's sparsity measurement (denoted as 'M') is defined as:

$$M(\boldsymbol{q}_i, \boldsymbol{k}) = \ln \sum_{l=1}^{L_k} \exp\left(\frac{\boldsymbol{q}_i \boldsymbol{k}_l^T}{\sqrt{d}}\right) - \frac{1}{L_k} \sum_{j=1}^{L_k} \frac{\boldsymbol{q}_i \boldsymbol{k}_j^T}{\sqrt{d}}$$
(9)

The distribution of the attention probability of the dominant query will be different from the uniform distribution. The number of the dominant query is $u = c*lnL_k$. By allowing each key to pay attention only to the top u queries, the ProbSparse Self-attention is defined as follows:

$$\mathcal{A}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Softmax\left(\frac{\mathbf{Q}\mathbf{K}^{T}}{\sqrt{d}}\right)\mathbf{V}$$
(10)

Finally, the computational complexity of attention is reduced in each layer to $O(L_k ln L_Q)$ by sampling the $L_k ln L_k$ dot product pairs. Details of the specific derivation of the formula are provided in prior research [26].

The structure of Informer is shown in Figure 4:



Figure 4. The Structure of Informer.

As shown in Figure 4, after input embedding and position encoding, the data were sent to the encoder for valid information extraction and later to the decoder to complete the prediction. Considering the difficulty of dynamic decoding output for fast prediction, Informer uses one forward procedure to calculate the prediction results. For this purpose, a certain length of sequence from the input is needed as 'Start token' [27].

The original Informer model arrives at long time series prediction but does not consider the effect of non-stationary data in real scenarios. In fact, load is affected by various factors and shows significant variations. For instance, thermal power plants need to be shut down first when the electricity demand decreases. The original Informer model chooses the Mean Squared Error (MSE) loss function, which is not suitable for load forecasting of power plants that have large deviations in data distribution under extreme operating conditions, such as shutdown conditions. The MSE loss function is typically tailored to fit the distribution pattern of data and tends to penalize outliers, resulting in a biased prediction towards them. To address such issue, the SmoothL1Loss function [28] was employed in the present study instead of the original MSE loss function. The SmoothL1Loss function is less sensitive to outliers, leading to a more balanced predictive outcome. When the predicted value is close to the actual value, the SmoothL1Loss is similar to the MSE loss function to speed up the model convergence. The formula of SmoothL1Loss is as follows:

$$\log(x_i, y_i) = \begin{cases} \frac{1}{2} (x_i - y_i)^2, & if |x_i - y_i| < \delta \\ |x_i - y_i| - \frac{1}{2}, & otherwise \end{cases}$$
(11)

where δ is an artificially set hyperparameter. In addition, given the unsteady nature of load, better results may be obtained by regarding data fluctuation as a feature in model construction. In the present study, the residuals of load were constructed as a feature with the following formula:

$$Y_{residuals} = Y_t - Y_{t-1} \tag{12}$$

The effects of Informer with improvements on non-stationary data (Informer_nos) and original Informer were tested and analyzed as described in Section 4.

3.3. Causal Effect Estimation Model of External Factors

Based on the screening results of the mutual information method obtained in Section 3.1, the causal effect of ambient temperature on load was the focus of the present study. Other external variables were treated as binary variables for rough causal effect calculation. First, the causal effect of external variables other than ambient temperature was calculated using propensity score matching (hereafter referred to as PSM) [29]. All external variables except ambient temperature were converted to binary variables using carefully designed rules. Therefore, the treatment of the variables was also binary, and the treatment could only take the value of 0 or 1. Then, *Y* was used to refer to the outcome of the causal effect, that is, load. Additionally, feature *X* consists of each external factor, where X_i is sample *i*'s feature vector. The causal effect of treatment *t* on *Y* was the focus of the present study, specifically the average treatment effect (hereinafter abbreviated as ATE).

$$V_{ATE} = E[V_{ITE}] = E[Y_1] - E[Y_0]$$
(13)

The V_{ITE} mentioned in the above equation refers to the individual treatment effect (ITE), which is defined as $V^{i}_{ITE} = Y^{i}_{1} - Y^{i}_{0}$.

For PSM, each sample is first divided into two groups according to the true value of treatment t, which is a binary variable. PSM assumes that t is a random variable drawn from a particular distribution, that is, $t_i \sim P(t_i | x_i)$. In each group, the probability of receiving a certain t is calculated for each instance based on given x. This probability is called the propensity score. Finally, the samples with the closest propensity scores

in different groups are matched. Taking sample *i* as an example, sample *j* s.t argmin $dist(i,j) = |P(t | X_i) - P(t | X_j)|$ is found. The ATE that is obtained from PSM is as follows:

$$\hat{\tau} = \frac{1}{n} \left[\sum_{i:t_i=1} (y_i - y_j) + \sum_{i:t_i=0} (y_j - y_i) \right]$$
(14)

As one of the external variables with the most significant impact on load, the continuous treatment effect of temperature is computed with an interval of 1 °C. The Double Machine Learning (hereafter referred to as DML) [30] approach is applied to achieve unbiased estimation of the effect of temperature on load. The random forest algorithm [31] is applied to fit the load Y and the temperature T separately. The residuals are obtained by removing the corresponding random forest algorithm fit results from the load Y and temperature T, respectively.

$$\hat{Y} = \hat{Y} - \hat{Y}(X) \tag{15}$$

$$\tilde{T} = T - \hat{T}(X) \tag{16}$$

where \tilde{Y} and \tilde{T} are the fitted residuals. After obtaining the residuals Y and T, the causal effect can be obtained by modeling a regression equation. The specific equation is $\tilde{Y} = \theta(X)\tilde{T}$. The residuals \tilde{Y} and \tilde{T} are the quantities of variation and are related by $\theta(X)$. $\theta(X)$ describes the association between \tilde{Y} and \tilde{T} , which is the causal effect ATE. Because $\theta(X)$ is an equation related to x, it can be solved by the method of least squares.

3.4. Load Forecasting Model for Power Plants

As shown in Figure 2, the external factors were first screened by the MI method to obtain the factors related to load, which was fully described in Section 3.1. Subsequently, the load series with the specified forecast length was obtained using the improved Informer presented in Section 3.2. Then, the causal inference model introduced in Section 3.3 was used: first, a causal diagram was constructed; second, causality was identified by the back door criterion and the front door criterion; and then, causal effects were calculated via different methods. The causal effects were calculated to determine the contribution or inhibition of different external variables to load and thus adjust the prediction results. Historical data were used for the causal effect estimation in the present study. The final predicted load L_n is:

$$L_n = A_n * \frac{1919 + \sum_{i=1}^6 ATE_i}{1919}$$
(17)

where *n* represents the specific prediction time length; A_n represents the prediction result of the improved Informer model; *I* represents the external factor. ATE_i is the causal effect of external variable *i* in the forecast period, and 1919 is the historical single-day average value of load. Notably, in a real scenario, the external variables used to calculate the causal effects are predicted by the Informer model, including temperature, rainfall, wind speed, and GDP. Other external variables, such as workday and servicing, can be obtained by inference. To reduce prediction error, external variables other than temperature were designated as binary variables for the rough causal effect calculation.

4. Overall Flow and Discussion of the Experiment

In the present study, the daily load data were used in combination with the economic and environmental data of Jiangsu Province to construct the dataset. The load data were obtained from a thermal power plant in Jiangsu Province, China, for the period from 1 July 2019 to 28 September 2022, with a time granularity of days, as shown in Figure 5. A total of 1096 data points were used to train, test, and validate the model from 1 July 2019 to 30 June 2022. The training, validation, and test set ratio was 7:2:1. Model performance was evaluated using data from 1 July to 28 September 2022 for prediction. The external variable



data for the corresponding period were obtained from the official website of the National Bureau of Statistics of China [32] and the public meteorological information website [33].

Figure 5. The actual generation load curve.

For the external variable data, the following factors were considered: GDP of Jiangsu Province, work days in China, thermal power share in Jiangsu Province (hereafter referred to as proportion), ambient temperature, rainfall records, wind speed, and the maintenance status of the power plant from which the load data were derived.

4.1. Experiment Preparation

In the present study, Causal–Informer was used for power load forecasting. Data were collected from a power plant located in Jiangsu Province, China. The dataset contained a total of 1186 time steps with a minimum time interval of days. The input features and labels of the model are shown in Table 1. Except for the load residuals and date, the remaining input features were also used for the causal effects calculation.

	x1	Date (Use in Informer only)
	x2	Workday
	x3	Servicing
Input	x4	GDP
-	x5	Rainfall
	x6	Wind Speed
	x7	Temperature
	x8	Load residuals (Use in Informer only)
Output	x9 (Target)	Load

Table 1. Inputs and output of the model.

4.2. Overall Flow of the Experiment

The overall flow of the experiment is shown in Figure 6.

- Data collection: The data included daily load data from a power plant and data on external factors. The datasets provided a range from 1 July 2019 to 28 September 2022.
- Data preprocessing: The load and external factor data were obtained from various sources, and their time stamps had to be aligned when integrated. For data such as temperature, wind speed, and rainfall, the minimum time interval for processing was on a daily basis, and the specific value recorded was the maximum value per day.
- Feature engineering: The external factors were filtered by means of the mutual information method to obtain six variables, namely ambient temperature, workday, servicing, GDP, wind speed, and rainfall.
- Data splitting: In the present study, the training, validation, and test set ratio was 7:2:1.

- Model training: The present experiments involved comparing Causal–Informer with Informer, LSTM, Prophet, and ARIMA to determine the optimal model.
- Model evaluation: MAPE, MAE, and RMSE were adopted as the evaluation metrics. MAPE is a percentile of error that removes the effect of data scaling. MAE is the factual error between prediction and actual data, which is dependent only on the volume of the data. RMSE ensures that each item is positive and distinguishable.



Figure 6. Overall flow of the experiment.

$$MAPE = \frac{100\%}{N} \sum_{n=1}^{N} |\frac{P_n - P_n^{\sim}}{P_n}|$$
(18)

$$MAE = \frac{1}{N} \sum_{n=1}^{N} |P_n - P_n^{\sim}|$$
(19)

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (P_n - P_n^{\sim})^2}$$
(20)

 P_n is the actual load value and P_n^{\sim} is the corresponding result of the model prediction. *N* in the above equation denotes the number of time steps predicted by the model at a time (hereinafter abbreviated as pred_length).

 Results analysis: Visualization of forecast results. Model performance can be more visually observable for analysis.

4.2.1. Experiment 1: Causal Effect Estimation

External factors other than temperature were considered binary variables in the present study, following the analysis in Section 3.3. The specific setting ranges for the binary variables are shown in Table 2. The causal effect of temperature was calculated by treating temperature as a continuous variable with 1 °C intervals. Depending on the temperature in the region where the power plant was located, the causal effect of the temperature from minus 1 °C to 39 °C on the load was analyzed.

Table 2. Binary variables design rules.

External Factors	Conditions for Binary Variables Set to 1
Workday	National working days
Servicing	Planned unit shutdown maintenance
Rainfall	Rainfall > 10 mm
GDP	Single-day GDP (converted from quarterly GDP) > RMB 30,954 million
Wind Speed	Wind speed $> 6 \text{ m/s}$

After data preprocessing, the causal effects of various external variables on load were calculated using the DML and PSM algorithms. The specific calculation results are presented in Figures 7 and 8, respectively. An observation can be made that the causal effect on load increased rapidly when the temperature increased above 30 °C. When the temperature increased above 36 °C, the plant was close to the upper limit of generation capacity, resulting in a decrease in causal effect. Other binary variables in addition to temperature had a significant impact on causal effect. The power plant in the present study had two generating units, and the units were shut down for servicing by rotating them during periods of low power consumption to extend their service life. The plant's single-day generation load was reduced by an average of 987 million kwh when a unit was under servicing. In addition, there was a significant causal effect by workdays and GDP on electricity generation. The causal effect of rainfall was the weakest in relative terms.

4.2.2. Experiment 2: Model Performance Comparisons

In Experiment 2, Causal–Informer was analyzed in comparison to four other popular time series forecasting models under equivalent conditions. As shown in Table 3, Table 4 and Figure 9, by forecasting 30 time steps and 90 time steps, the performance of all models was summarized.



Figure 7. The causal effect of various temperatures on load.



Figure 8. Causal effects of various external factors on load.

Table 3. Evaluation metrics for all models in the same inp	out data (Pred_length = 30 d).
--	--------------------------------

	Arima	Prophet	LSTM	Informer	Causal–Informer
MAE	593.36 million kwh	888.29 million kwh	630.8 million kwh	525.24 million kwh	250.31 million kwh
RMSE	649.01 million kwh	1004.2 million kwh	727.19 million kwh	624.57 million kwh	331.96 million kwh
MAPE	22.4%	35.8%	24.5%	19.7%	10.4%

Table 4. Evaluation metrics for all models in the same input data (Pred_length = 90 d).

	Arima	Prophet	LSTM	Informer	Causal–Informer
MAE	641.54 million kwh	717.71 million kwh	487.34 million kwh	507.09 million kwh	430.59 million kwh
RMSE	709.99 million kwh	822.83 million kwh	638.51 million kwh	625.75 million kwh	535.94 million kwh
MAPE	31.4%	35.8%	29.9%	31.2%	24.8%



(a)



(**b**)

Figure 9. Results of Forecasting. (a) Pred_length = 30 d. (b) Pred_length = 90 d.

Model performance is indicated by the model error values. When pred_length = 30 d in advance, Causal–Informer obtained the smallest values among all error evaluation metrics, and the Informer model obtained the second smallest values. When pred_length = 90 d in advance, Causal–Informer obtained the smallest values among all the error assessment metrics. Among the second smallest values of the three error evaluation metrics, LSTM had two and Informer had one.

As can be seen from Figure 9, the performance of the different models varied considerably. Such findings could be attributed first to the difficulty of predicting load as non-stationary data and second to the difference in the predictive ability of different models. The forecast data were the plant's actual load from 1 July 2022 over the next 30 or 90 days, respectively. In early August 2022, power generation spiked due to a significant increase in electricity demand caused by continuous high temperatures. By the end of August, load decreased rapidly as temperatures dropped and power was dispatched. Beginning on 18 September 2022, one of the power plant's generating units underwent maintenance, resulting in a reduction in the plant's generating capacity. The data for the aforementioned time period were significantly different from historical data. Therefore, there were prediction difficulties. A method with robustness had to be proposed.

LSTM, a widely used time series prediction algorithm, performed well on the dataset. The prediction curves follow the trend of the actual data, but there are slight discrepancies between the predicted values and the actual values. The Arima model performed poorly on this dataset. Despite debugging with various tuning methods, the model could not obtain satisfactory results. Prophet performed better when forecasting 90 days compared to forecasting 30 days. Specifically, Prophet was not able to capture the data trend during the first 25 days, but showed better performance in fitting the data trend after that period. Informer's prediction results at 30 time steps are more consistent with the real data. However, it performed poorly at 90 time steps and was weak at predicting trends. Causal–Informer successfully predicted the trend of the actual data without the effect of time lag. The predictions are close to the real data at multiple time points. A more prominent advantage is the accurate prediction of multiple turning points, which could be a good reference for power plant generation planning. The disadvantage is that there will still be a gap between the predicted value and the actual value at some prediction points. In general, most models are better at predicting 30 time steps than 90 time steps.

4.2.3. Experiment 3: Evaluation of Causal–Informer Performance

In the present study, experiments with ablation were also performed to verify the robustness of the proposed method. The proposed innovation points were found to be effective and could improve model performance, as shown by the results of the ablation experiments.

The results of the ablation experiments are shown in Table 5, Figures 10 and 11. Better results were obtained for all innovations. Specifically, when Informer and Informer + ATE were compared, both accurately captured the trend of the data in the 30 time steps ahead forecast; Informer + ATE produced results closer to the real data than Informer. In the 90 time steps ahead forecast, Informer + ATE captured the trend of the data well and solved the problem of long-term forecasting. When Informer and Informer_nos were compared, Informer_nos produced prediction results closer to the true value than Informer in the prediction at 30 time steps ahead, proving that attempts such as replacing the loss function can produce positive results; for the prediction at 90 time steps ahead, both had the problem of not capturing the trend of the data. The combination (that is, Causal–Informer) of Causal Effect Calculation (ATE) and Informer_nos produced optimal results. The combination produced optimal results because it correctly handled the problem of difficult-to-predict nonstationary data and the problem of difficult-to-predict trends in long time series, thus making the proposed method robust.

Table 5. Ablation study.						
Pred_Length	30 d			90 d		
Evaluation Metrics	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Informer	525.24 million kwh	624.57 million kwh	19.7%	507.09 million kwh	625.75 million kwh	31.2%
Informer + ATE	430.59 million kwh	518.03 million kwh	16.4%	395.81 million kwh	560.81 million kwh	25.3%
Informer_nos Causal-Informer	315.47 million kwh 250.31 million kwh	380.29 million kwh 331.96 million kwh	12.5% 10.4%	512.03 million kwh 430.59 million kwh	593.71 million kwh 535.94 million kwh	28.1% 24.8%

Informer + ATE combines the causal effect of external factors. See Section 3.4 for details of the calculation, replacing A_n with the original informer's prediction. Informer_nos represents an improved Informer model for non-stationary data prediction, as described in Section 3.2.



Figure 10. Ablation experiments (Pred_length = 30 d). (a) Ablation Experiments about ATE. (b) Ablation Experiments about Informer_nos.



Figure 11. Ablation experiments (Pred_length = 90 d). (a) Ablation Experiments about ATE. (b) Ablation Experiments about Informer_nos.

5. Conclusions

In the present study, the causal effect calculation and the time series forecasting model were combined to establish a multi-factor-influenced load forecasting model for power plants. In the present study, external variables that correlate with load were identified, and the causal effect of different variables on load were quantified. A robust and interpretable method for medium- and long-term load forecasting was proposed. The performance of five time series models was compared using three different error evaluation metrics, and the advancement of the proposed Causal-Informer model was demonstrated. Ablation experiments were also performed to verify the improvement in model performance from all the innovations. The proposed method achieved the most accurate and stable results. The MAE, RMSE, and MAPE indicators of the proposed method reached 250.31 million kwh, 331.96 million kwh, and 10.4%, respectively, in 30 time step forecasting. In 90 time step forecasting, the MAE, RMSE, and MAPE indexes reached 430.59 million kwh, 535.94 million kwh, and 24.8%, respectively. Based on the experimental results, the following conclusions could be drawn:

- 1. Causal–Informer performs well in multi-step forecasting and has strong performance potential.
- 2. The incorporation of causal inference within the time series prediction scenario is beneficial in improving the model's interpretability and robustness.
- 3. The improvements to the proposed Informer model have some positive effects on predicting non-stationary data.

Author Contributions: Methodology, K.Y.; data curation, K.Y.; software, K.Y.; draft writing, K.Y.; Conceptualization, F.S.; project administration, F.S.; supervision, F.S.; draft review, F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data of this study are available from the corresponding author upon request (fhshi@tongji.edu.cn).

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. The CPC Central Committee and the State Council. *A Number of Views on Further Deepening the Reform of the Power System;* The CPC Central Committee and the State Council: Beijing, China, 2015.
- 2. Jiangsu Provincial Development and Reform Commission. *Jiangsu Medium and Long-Term Trading Rules for Electricity;* Jiangsu Provincial Development and Reform Commission: Nanjing, China, 2021.
- 3. Kang, C.; Xia, Q.; Zhang, B. Review of Power System Load Forecasting and its Development. *Autom. Electr. Power Syst.* 2004, 28, 1–11.
- Chen, X. Research on Medium and Long Term Power Load Forecasting Method Considering DSM and Intelligent Power Use. Master's Thesis, Southeast University, Dhaka, Bangladesh, 2018.
- 5. Deng, J. Gray System Theory Tutorial; Huazhong University of Science and Technology Press: Wuhan, China, 1990.
- Fang, T.; Lahdelma, R. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Appl. Energy* 2016, 179, 544–552. [CrossRef]
- 7. Aisyah, S.; Simaremare, A.A.; Adytia, D.; Aditya, I.A.; Alamsyah, A. Exploratory Weather Data Analysis for Electricity Load Forecasting Using SVM and GRNN, Case Study in Bali, Indonesia. *Energies* **2022**, *15*, 3566. [CrossRef]
- Hong, Y.; Zhou, Y.; Li, Q.; Xu, W.; Zheng, X. A Deep Learning Method for Short-Term Residential Load Forecasting in Smart Grid. IEEE Access 2020, 8, 55785–55797. [CrossRef]
- 9. Zhang, D.; Guan, W.; Yang, J.; Yu, H.; Xiao, W.; Yu, T. Medium—And Long-Term Load Forecasting Method for Group Objects Based on the Image Representation Learning. *Front. Energy Res.* **2021**, *9*, 739993. [CrossRef]
- 10. Gong, M.; Zhao, Y.; Sun, J.; Han, C.; Sun, G.; Yan, B. Load forecasting of district heating system based on Informer. *Energy* **2022**, 253, 124179. [CrossRef]
- Li, C.; Dong, Z.; Ding, L.; Petersen, H.; Qiu, Z.; Chen, G.; Prasad, D. Interpretable Memristive LSTM Network Design for Probabilistic Residential Load Forecasting. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2022, 69, 2297–2310. [CrossRef]
- 12. Shohan, J.A.; Faruque, O.; Foo, S.Y. Forecasting of Electric Load Using a Hybrid LSTM-Neural Prophet Model. *Energies* **2022**, *15*, 2158. [CrossRef]
- Mo, R.; Chen, P.; Li, X.; Li, Y.; Guo, M. Coordination Analysis and Prediction of Energy Distribution and Economic Development Based on Gray Theory. J. Electr. Power 2018, 33, 86–92. [CrossRef]
- 14. Hyndman, R.J.; Athanasopoulos, G. Forecasting: Principles and Practice; OTexts: Melbourne, Australia, 2018.
- 15. Imai, K.; Ratkovic, M. Covariate balancing propensity score. J. R. Stat. Soc. Ser. B-Stat. Methodol. 2014, 76, 243–263. [CrossRef]
- Lawlor, D.A.; Harbord, R.M.; Sterne, J.A.C.; Timpson, N.; Smith, G.D. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* 2008, 27, 1133–1163. [CrossRef] [PubMed]
- 17. Hedström, P.; Ylikoski, P. Causal Mechanisms in the Social Sciences. Annu. Rev. Social. 2010, 36, 49–67. [CrossRef]

- Niu, Y.; Tang, K.; Zhang, H.; Lu, Z.; Hua, X.S.; Wen, J.R. Counterfactual VQA: A Cause-Effect Look at Language Bias. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021. [CrossRef]
- Yao, L.; Chu, Z.; Li, S.; Li, Y.; Gao, J.; Zhang, A. A Survey on Causal Inference. ACM Trans. Knowl. Discov. Data 2021, 15, 1–46. [CrossRef]
- 20. Pearl, J. Models, Reasoning and Inference; Cambridge University Press: Cambridge, UK, 2000; p. 19.
- 21. Pearl, J. Causal inference in statistics: An overview. *Stat. Surv.* **2009**, *3*, 96–146. [CrossRef]
- 22. Pearl, J. Causality: Models, Reasoning, and Inference, 2nd ed.; Cambridge University Press: New York, NY, USA, 2009.
- 23. Kalisch, M.; Buhlmann, P. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *J. Mach. Learn. Res.* 2007, *8*, 613–636.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
- Tsai, Y.H.H.; Bai, S.; Yamada, M.; Morency, L.P.; Salakhutdinov, R. Transformer Dissection: An Unified Understanding for Transformer's Attention via the Lens of Kernel. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Hong Kong, China, 3–7 November 2019.
- Zhou, H.; Zhang, S.H.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond effificient transformer for long sequence time-series forecasting. *arXiv* 2020, arXiv:2012.07436. [CrossRef]
- Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert:Pre-training of deep bidirectional transformers for language understanding. arXiv 2018, arXiv:1810.04805.
- 28. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015. [CrossRef]
- 29. Rosenbaum, P.R. Matched Sampling for Causal Effects: The Central Role of the Propensity Score in Observational Studies for Causal Effects; Oxford University Press: Oxford, UK, 2006.
- Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; Robins, J. Double/debiased machine learning for treatment and structural parameters. *Econ. J.* 2018, 21, C1–C68. [CrossRef]
- 31. Ho, T.K. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 1998, 20, 832–844.
- 32. National Bureau of Statistics of China. Available online: www.stats.gov.cn (accessed on 19 April 2023).
- 33. Public Meteorological Information Website. Available online: www.rp5.ru (accessed on 19 April 2023).

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.