



## Article

# Hourly Power Consumption Forecasting Using RobustSTL and TCN

Chih-Hsueh Lin<sup>1</sup>, Ulin Nuha<sup>1,\*</sup>, Guang-Zhi Lin<sup>1,2</sup> and Tsair-Fwu Lee<sup>1,3,\*</sup>

<sup>1</sup> Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan; cslin@nkust.edu.tw (C.-H.L.); i110152103@nkust.edu.tw (G.-Z.L.)

<sup>2</sup> Department of Tactical Control Air Traffic Control & Meteorology, Air Force Institute of Technology, Kaohsiung 82047, Taiwan

<sup>3</sup> PhD Program in Biomedical Engineering, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

\* Correspondence: i110152118@nkust.edu.tw (U.N.); tflee@nkust.edu.tw (T.-F.L.)

**Abstract:** Power consumption forecasting is a crucial need for power management to achieve sustainable energy. The power demand is increasing over time, while the forecasting of power consumption possesses challenges with nonlinearity patterns and various noise in the datasets. To this end, this paper proposes the RobustSTL and temporal convolutional network (TCN) model to forecast hourly power consumption. Through the RobustSTL, instead of standard STL, this decomposition method can extract time series data despite containing dynamic patterns, various noise, and burstiness. The trend, seasonality, and remainder components obtained from the decomposition operation can enhance prediction accuracy by providing significant information from the dataset. These components are then used as input for the TCN model applying deep learning for forecasting. TCN employing dilated causal convolutions and residual blocks to extract long-term data patterns outperforms recurrent networks in time series forecasting studies. To assess the proposed model, this paper conducts a comparison experiment between the proposed model and counterpart models. The result shows that the proposed model can grasp the rules of historical time series data related to hourly power consumption. Our proposed model overcomes the counterpart schemes in MAPE, MAE, and RMSE metrics. Additionally, the proposed model obtains the best results in precision, recall, and F1-score values. The result also indicates that the predicted data can fit the pattern of the actual data.

**Keywords:** forecasting; power consumption; RobustSTL; TCN; deep learning



**Citation:** Lin, C.-H.; Nuha, U.; Lin, G.-Z.; Lee, T.-F. Hourly Power Consumption Forecasting Using RobustSTL and TCN. *Appl. Sci.* **2022**, *12*, 4331. <https://doi.org/10.3390/app12094331>

Academic Editors: Ivan A. Parinov, Shun-Hsyung Chang and Hung-Yu Wang

Received: 15 March 2022

Accepted: 21 April 2022

Published: 25 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

The issue of world climate change is a concern for various countries in the world. NOAA's 2020 Annual Climate Report states that the average rate of temperature increase of the world's surface has increased twice every ten years since 1981, compared to every ten years since 1880. This temperature increase can impact the increased demand for power consumption by about 2.7% for every 1 °C increase [1]. Therefore, power management possesses a crucial role in planning sustainable power supply policies. The first essential treatment is to forecast the power consumption of an area. The forecasting of power consumption can bring significant information to help estimate power demand for planning in the future. In addition, such forecasting of power consumption assists the powerhouse in formulating suitable and sustainable operations for electric power systems. A properly conducted forecasting can ensure that the supply will be able to meet the future demand for electrical power [2].

The work of power consumption forecasting can be categorized into two schemes, i.e., the traditional algorithm such as the autoregressive integrated moving average (ARIMA) model, and machine learning such as artificial neural networks (ANN). There are many studies performed that focus on power consumption forecasting at various scales. A comparison study of power consumption forecasting was performed during Korea's highest

peak demands using various models involving traditional, machine learning, and hybrid models. The study result shows that machine learning models, such as long short-term memory (LSTM), and hybrid models, such as seasonal auto-regressive integrated moving average with exogenous factors (SARIMAX) and LSTM gain greater results over the traditional model [3]. In [4], the authors proposed a hybrid model for the month-ahead hourly power demand forecasting for medium industrial consumers. This model employed a non-linear autoregressive network with exogenous inputs (NARX) and artificial neural networks (ANN) to forecast daily power demand using timestamp datasets as exogenous variables and the output of NARX as LSTM-ANN input. Subsequently, a singular spectrum analysis and linear recurrent formula (SSA-LRF) model was proposed for Indonesian power forecasting [5]. This model was also combined with fuzzy or ANN frameworks to address nonlinearity of data. The result indicates that SSA-LRF-ANN is the accurate prediction model for such a related task.

However, the existing studies in power forecasting encounter issues affecting forecasting accuracy. The main challenges are non-regularity distribution and various noise [6]. Several studies have utilized the STL method to decompose the time series data of power demands into seasonal, trend, and remainder components [6–8]. Afterward, these components are fed to the forecasting model as input. Several approaches have gained accurate results for power consumption forecasting with dynamic patterns. Nevertheless, they still need to be enhanced regarding the capture of abrupt changes, burstiness, and high levels of noise in future demands. Therefore, we propose a hybrid model of the RobustSTL and a temporal convolutional network (TCN) to forecast single time series data for hourly power consumption with reliability from interferences such as dynamic patterns, burstiness, and outliers. To obtain a comprehensive historical data understanding, the proposed model first decomposes the raw data using the RobustSTL model, instead of the standard STL. RobustSTL not only captures components such as STL results, but also identifies the residual spike and white noise in the remainder component [9]. These components then become input for the TCN. Through TCN, the proposed model can capture and learn the features of each component's information, although it has long dependencies [10,11].

Based on the analysis above, this study focuses on building hourly power consumption forecasting using a combining model of RobustSTL and TCN. The major contributions of this study can be summarized below:

1. We propose a forecasting model for single time series data regarding hourly power consumption utilizing RobustSTL and TCN;
2. The study's key contribution is the hybrid model of RobustSTL and TCN as the forecasting model;
3. The proposed model can capture and understand the time series data despite containing dynamic patterns and burstiness;
4. The experimental stage was performed based on real hourly power consumption and validated with the existing forecasting models.

This first section organizationally is followed by Section 2 discussing the materials and methods of this study. Section 3 delivers the experimental results and a discussion of hourly power consumption forecasting. Afterward, Section 4 presents the conclusion of this study.

## 2. Materials and Methods

Figure 1 shows the general architecture of the proposed model in this study. Time series data regarding hourly power consumption is first decomposed using the RobustSTL. This decomposition generates three components, i.e., trend, seasonal, and remainder components. These components are simultaneously carried as input for the TCN model. The outputs of the TCN model are flattened with the fully connected layer. We can then obtain the predicted values of power consumption. Through this operation, the model can obtain a better understanding of data patterns and spatial features to estimate the future demand of hourly power consumption.

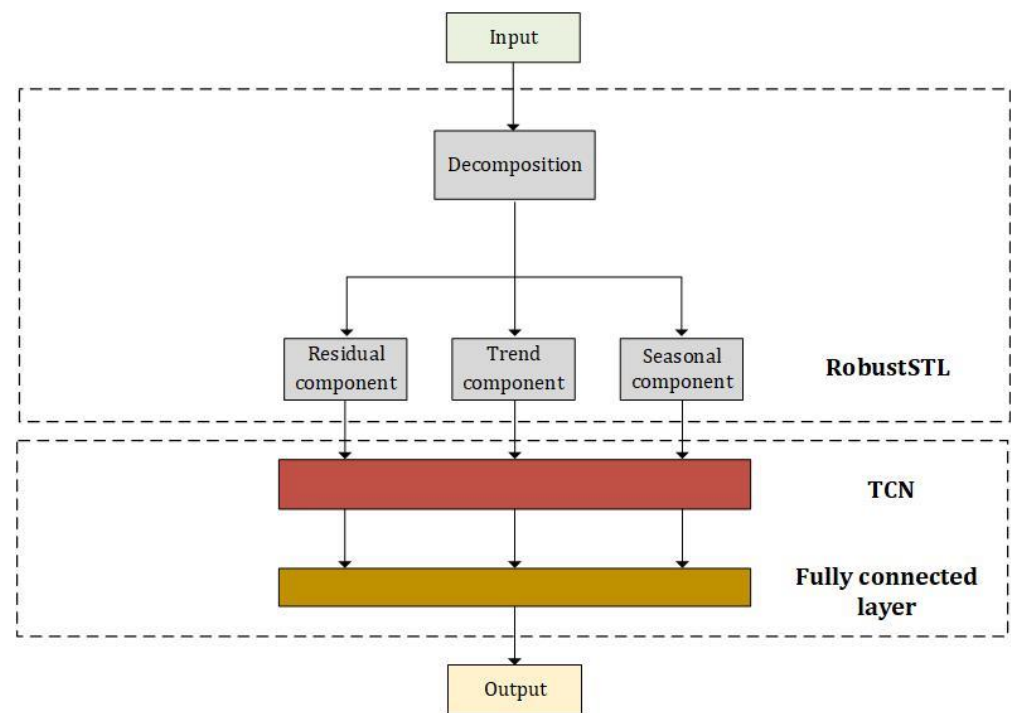


Figure 1. The general architecture of the proposed method.

2.1. RobustSTL

In this study, we selected RobustSTL proposed by Wen et al. [9], instead of standard STL, for the decomposition method. STL (Seasonal-Trend decomposition using LOESS) is a decomposition method utilizing LOESS (locally estimated scatterplot smoothing) to extract estimates of seasonality, trend, and residual components [12]. RobustSTL decomposition has reliability with abrupt changes and various noise compared to the existing methods. In RobustSTL, a residual or remainder component can be further extracted into spike and noise. This decomposition employs the least absolute deviations (LAD) loss with sparse regularizations to extract the trend component, non-local seasonal filtering to decompose the seasonal component, and bilateral filtering to denoise the data. The algorithm summary of RobustSTL to decompose a series data  $y_t$  into trend  $\tau_t$ , seasonality  $s_t$ , and remainder  $r_t$  is shown in Algorithm 1 [9].

Algorithm 1. RobustSTL method summary

**Input:**  $y_t$ , parameter configuration

**Output:**  $\tau_t, s_t, r_t$

**Step 1:** Denoise the time series data using bilateral filtering,  $y'_t = \sum_{j \in J} w_j^t y_j$

**Step 2:** Calculate the relative trend,  $\bar{\tau}_t^r = \begin{cases} 0, & t = 1 \\ \sum_{i=2}^t \nabla \hat{\tau}_i, & t \geq 2 \end{cases}$ ,  $y''_t = y'_t - \bar{\tau}_t^r$

**Step 3:** Calculate the seasonality using non-local seasonal filtering,  $\bar{s}_t = \sum_{(t',j) \in \phi} w_{(t',j)}^t y''_j$

**Step 4:** Adjust the trend, seasonality, and remainder components

$$\tau_t = \bar{\tau}_t^r + \tau_1, s_t = \bar{s}_t - \tau_1, r_t = y_t - s_t - \tau_t$$

where  $w_j^t$  and  $J$  in Step 2 are the filter weight and the filter window. Additionally,  $\nabla \hat{\tau}_i$  is the optimized first-order difference of the trend component and  $\phi$  is neighborhood parameter.

2.2. TCN

The convolutional neural network (CNN) has outstanding results for tasks in two dimensions, i.e., image processing. It is adopted in one dimension for the time series task. However, the main drawback of the basic architecture of CNN is that it is difficult to gain crucial information in long-term time series data. Thus, Bai et al., proposed TCN as an

enhancement of convolutional and recurrent networks due to being able to memorize sequence data with long relationships accurately [13]. TCN is capable of memorizing more historical data than a standard sequence of neural networks by widening the receptive field size [14]. A TCN model consists of residual blocks including dilated causal convolutional neural networks, residual connection, and others as shown in Figure 2.

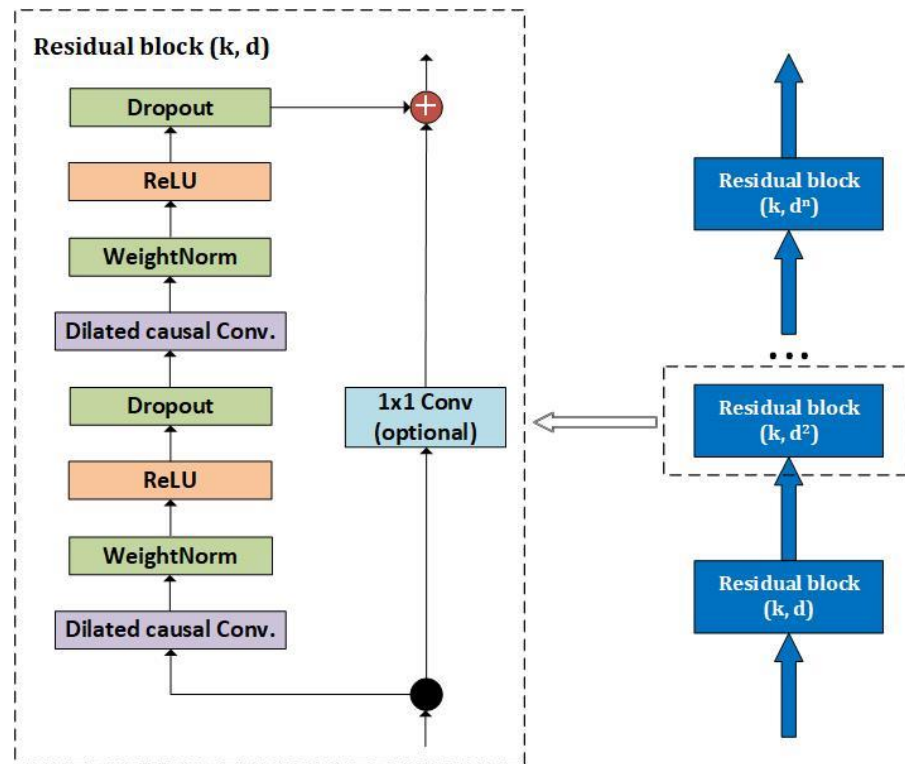


Figure 2. Residual block architecture in TCN model.

A residual block in the TCN model contains some basic layers, i.e., dilated causal convolutions, weight normalization, the rectified linear unit (ReLU) as activation function, and spatial dropout as regularization. Here, the causal convolution term refers to a convolution process obtaining output  $t$  by convolving elements from inputs at step  $t$  and the previous periods. Let us assume to predict output  $y_t$ , we need elements  $x_0, \dots, x_t$  from the previous layer as input.

Afterward, dilated causal convolution is proposed to address the drawback of long-term dependency in causal convolutional neural networks. The dilated causal convolution exponentially enlarges the receptive fields to memorize the long historical information without adding other parameters [15]. The dilated causal convolution steps up the filter in input sequences and skips a certain point with a fixed stride. This convolution has tasks identical to the pooling concept but the input and output channel width in this convolution remains the same size. Figure 3 shows the dilated causal convolution process with dilation factors  $d$  of 1, 2, and 4 and filter size of  $k = 3$ . The blue and red colors refer to the original time series, while the yellow color denotes the padding. This dilated causal convolution can be formulated with Equation (1).

$$F(x) = \sum_{l=0}^{k-1} h(l)f(x - d.l), \tag{1}$$

where  $f(*)$  denotes the input of the dilated causal convolution and  $h(l)$  refers to the filter length  $k$ . Additionally,  $d$  refers to the dilation factor,  $x-d.l$  denotes the past direction of element  $x$ , and  $F(*)$  is the output of this process.

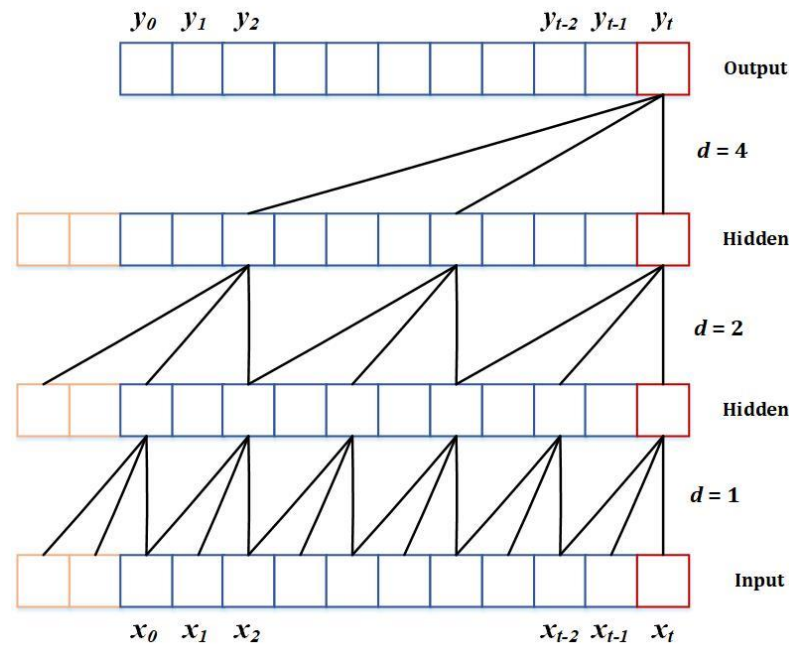


Figure 3. Dilated causal convolution process.

Weight normalization is applied in the TCN model after the dilated causal convolution process to speed up and make the training process robust. ReLU activation function placed after weight normalization can enhance the sparsity of TCN and avoid strong dependency among parameters [16]. Afterward, the spatial dropout is added to avoid overfitting the network model. In addition,  $1 \times 1$  convolution operation, as shown in Figure 2, is incorporated as the residual connection to adjust the channel width of the input and output layer which may have different sizes. Specifically, a residual connection is integrated to the output of TCN to prevent gradient attenuation. The input  $x$  is combined with the output of TCN formulated in Equation (2):

$$O = Activation(x + F(x)). \tag{2}$$

Subsequently, the output of the residual block from multiple layers is considered for the input of the next residual block.

### 2.3. Evaluation Metrics

We introduce several evaluation metrics, i.e., mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE), to evaluate the proposed model’s performance in the experimental stage as shown in Equations (3)–(5) below:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{x_t - \bar{x}_t}{x_t} \right|, \tag{3}$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |x_t - \bar{x}_t|, \tag{4}$$

$$RMSE = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}_t)^2, \tag{5}$$

where  $x_t$  and  $\bar{x}_t$  are actual values and predicted values, while  $n$  is the number of observed data.

We also propose a confusion matrix as the evaluation metric as shown in Figure 4, but we initially cluster the actual values and the predicted values using the Fuzzy c-means (FCM) algorithm. The FCM clustering algorithm assembles a set of data into  $m$  clusters where each data point in the dataset corresponds to each cluster at a certain degree [17]. The higher degree of a data point to a specific cluster denotes the corresponding cluster of

that data point. Here, we take (9) as the number of clusters for the actual values and the predicted values.

		Predicted data				
		Cluster <sub>1</sub>	...	Cluster <sub>j</sub>	...	Cluster <sub>k</sub>
Actual data	Cluster <sub>1</sub>	Z <sub>11</sub>	...	Z <sub>1j</sub>	...	Z <sub>1k</sub>
	⋮	⋮	...	...	...	⋮
	Cluster <sub>j</sub>	Z <sub>j1</sub>	...	Z <sub>jj</sub>	...	Z <sub>jk</sub>
	⋮	⋮	...	...	...	⋮
Cluster <sub>k</sub>	Z <sub>k1</sub>	...	Z <sub>kj</sub>	...	Z <sub>kk</sub>	

Figure 4. Confusion matrix table.

After clustering the actual data and predicted data, we evaluate them utilizing a confusion matrix. The confusion matrix is shown in a table to describe and display the performance of a classification task. The dimension of the confusion matrix is two, where one dimension refers to the actual data and another dimension refers to the predicted data [18]. Through this method, we can evaluate the proposed model using *precision*, *recall*, and *F1-score*. *Precision* denotes accuracy as the ratio of the correctly classified cluster to the total data predicted, while *recall* refers to the ratio of the correctly classified cluster to the total number of data actually belonging to the related cluster, and the *F1-score* represents the equilibrium of the precision and recall. The values of *precision*, *recall*, and *F1-score* can be calculated using Equations (6)–(8):

$$precision = \frac{TP}{TP + FP}, \tag{6}$$

$$recall = \frac{TP}{TP + FN}, \tag{7}$$

$$F1score = \frac{2 \times precision \times recall}{precision + recall}, \tag{8}$$

where *TP* (true positive) is all the diagonal values in the confusion matrix corresponding to each cluster. *FP* (false positive) refers to the sum of values of the corresponding column besides *TP* values of a cluster, and *FN* (false negative) denotes the sum of the values of the corresponding rows besides *TP* values of a cluster.

### 3. Results and Discussion

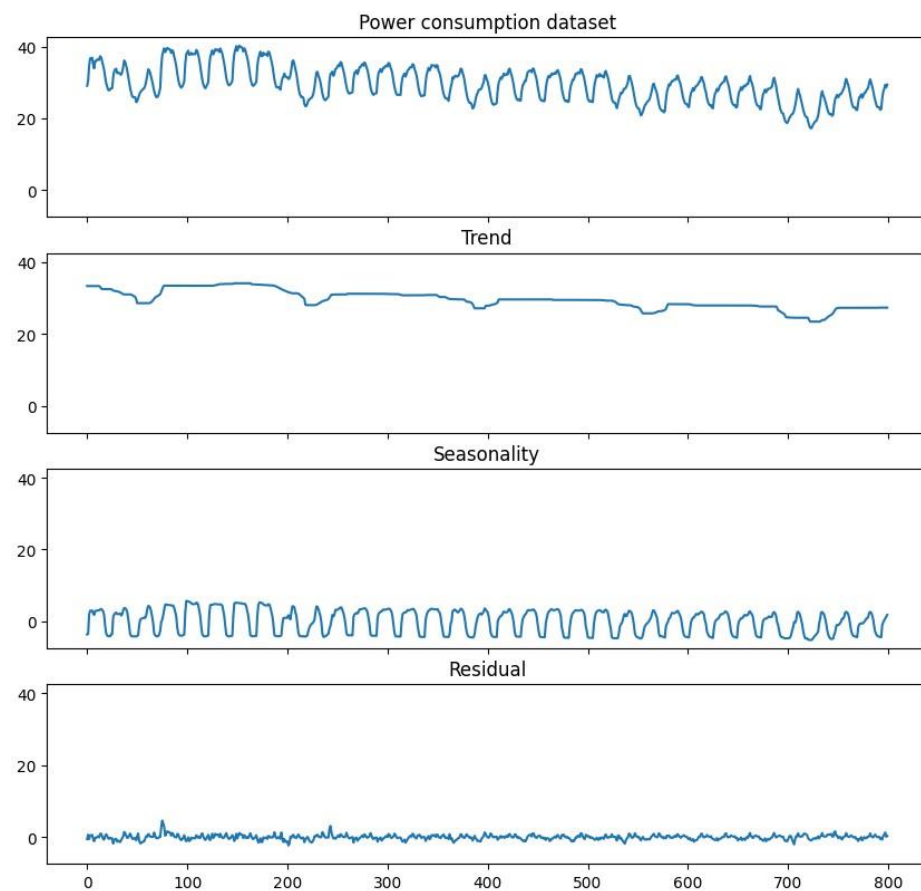
#### 3.1. Data Preparation

The dataset in this study is single time series data for the hourly power consumption of Turkey obtained in Kaggle [19]. This data represents the overall consumption of electrical power in the country. We took the dataset from November 2019 until April 2020 to conduct the experiment. The hourly power consumption in this study was taken for 24 h each day for 6 months. We divided the dataset into 80% for the training process and 20% for the testing process. Before we conducted the experiments, we set the configurations of the RobustSTL and TCN models. We set the length of the seasonality period in the RobustSTL to 24 h, because the hourly power demand for 24 h repeats somewhat in a similar pattern. For the parameter setting in the TCN model, we take only one residual block for the TCN model. Since the input sequence does not have high dimensionality, we set dilation factor  $d = 1, 2, 3$ , and 4, kernel size of  $k = 3$ , and the filter number of 64. In our proposed model, the power consumption  $y_t$  at hour  $t$  can be forecast by referencing the data sequence of previous data such  $y_{t-1}$ ,  $y_{t-2}$ , and  $y_{t-3}$ . Furthermore, we compare the proposed model

and counterpart models. MAPE, MAE, RMSE, and the confusion matrix are utilized as evaluation metrics.

### 3.2. Experimental Results

The time series data input for hourly power consumption is initially decomposed by the RobustSTL method to obtain three components. The decomposition results are shown in Figure 5. The figure consists of four parts, i.e., original dataset of power consumption, trend component, seasonality component, and residual component. The x-axis denotes the period of the data, and the y-axis denotes the power consumption value in GWh (Gigawatt Hour). These trend, seasonality, and residual components become the input data for the TCN.

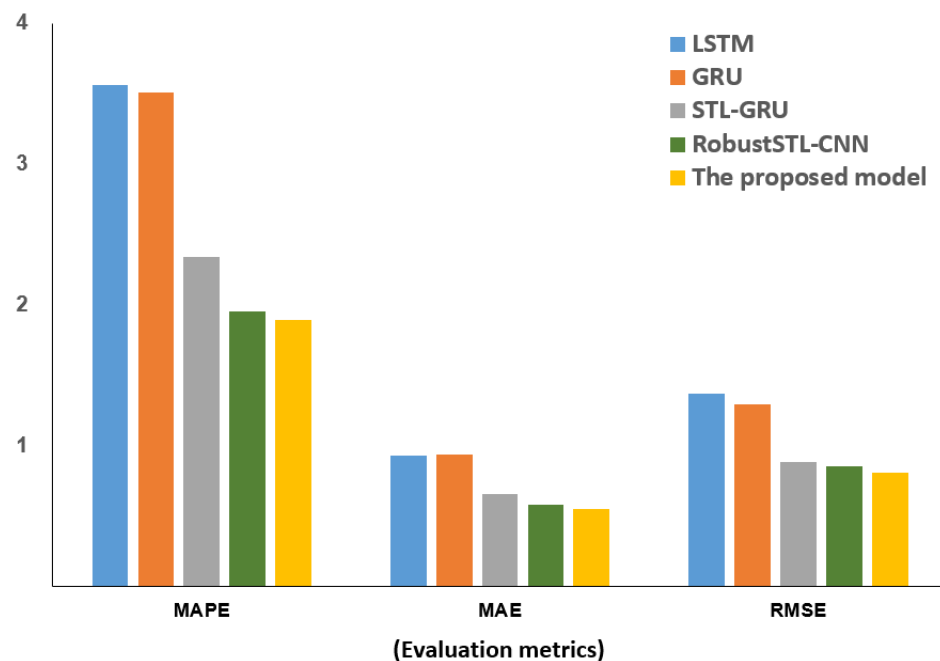


**Figure 5.** The decomposition results.

Then, Table 1 and Figure 6 show the comparison results of hourly power consumption prediction between the proposed model and the counterpart models in MAPE, MAE, and RMSE metrics.

**Table 1.** The results of comparison between the proposed model and counterpart models.

Model	MAPE (%)	MAE	RMSE
LSTM	3.56	0.93	1.37
GRU	3.51	0.94	1.29
STL-GRU	2.34	0.66	0.88
RobustSTL-CNN	1.95	0.58	0.85
The proposed model	1.89	0.55	0.81



**Figure 6.** Evaluation results of the proposed model and counterpart models.

In the comparison, we present LSTM and GRU (gated recurrent units) as the single model, and STL-GRU and RobustSTL-CNN as the hybrid model. In [20], the authors proposed LSTM as a machine learning approach in time series stock data forecasting. The proposed approach, LSTM, outperforms the traditional algorithms in empirical results. GRU was proven to achieve satisfactory accuracy in primary energy consumption forecasting over other methods [21]. A combination of recurrent neural networks and decomposition methods utilizing GRU and STL in short-term power consumption forecasting was proposed [7]. This model obtains good prediction accuracy capturing both local and global information of the data. A hybrid model of CNN and STL, instead of RobustSTL, introduced in [22] also is adopted here.

Based on Table 1 above, the proposed model has the lowest value of MAPE (1.95), while the other models mostly have a value of more than 2.00. The proposed model also has the lowest value of MAE and RMSE (0.58 and 0.85), while the other models gain values more than the proposed model values in these metrics. The lowest values in MAPE, MAE, and RMSE metrics indicate the best results because they have lower errors between the predicted and actual values. We can see in Table 1 that models with the hybrid approach outperform the single models in MAPE, MAE, and RMSE metrics. In the single models, GRU is better than LSTM in MAPE and RMSE metrics. LSTM only outperforms GRU in the MAE metric, by a small margin of 0.1. The best model after the proposed model is RobustSTL-CNN. The difference between the proposed model and RobustSTL-CNN in MAPE, MAE, and RMSE metrics is 0.06%, 0.03, and 0.04. Although the difference values are not huge, they are crucial for evaluation of the big dataset taken in this study.

The single model of LSTM and GRU basically obtains a good result. However, the hybrid model utilizing decomposition can increase forecasting accuracy, especially in the proposed model. Figure 7 shows the comparison between actual values and predicted values using the proposed model to see the real illustration. The predicted values overall can follow the actual values containing the dynamic pattern and burstiness. The dynamic pattern instance can be seen in the time series data after 21 March 2020. The maximum values of power consumption on 17–21 March 2020 each 24 h always exceeded 37 GWh, but the maximum values in the next days did not reach 37 GWh. The burstiness instance is shown in the red box in Figure 7. There are some points having higher values than common values over a short period. However, the proposed model can overcome the issues above through the predicted values having similarities with the real values.



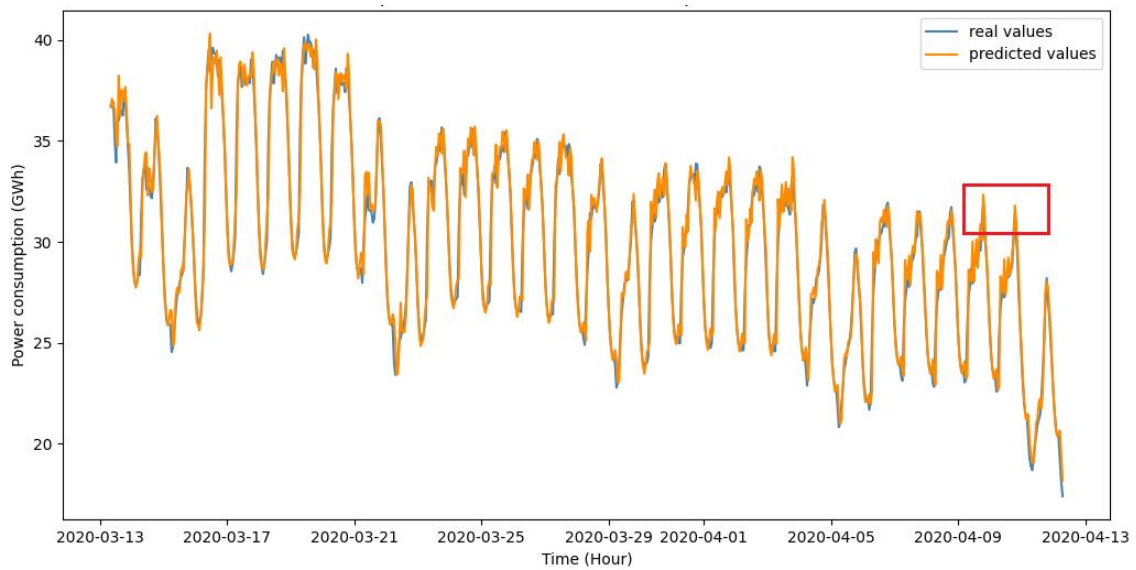


Figure 7. The comparison between the actual values and the predicted values.

Moreover, we also present the evaluation using another perspective, i.e., the classification method. After we obtained the clustering result between the actual data and predicted data, we grouped them into nine clusters. The confusion matrix is used to evaluate the experimental result, and the confusion matrix of the proposed model’s prediction results in each instance is shown in Figure 8. We can see that the total of TP values is 619, or 86% of all the data. We then take Cluster 1 as an instance where TP, TN, FP, and FN are 238, 455, 11, and 16. There are 16 points that should be classified in Cluster 1, but it is classified otherwise. It denotes only 6% of the points that have the incorrect class. For Cluster 5, the TP, TN, FP, and FN values are 70, 698, 26, and 66. There are 26 points that should be classified in Cluster 5, but it is classified otherwise. A total of 27% of the points have an incorrect class in Cluster 5. Since we obtained all the confusion matrices of the counterpart models, we could calculate precision, recall, and F1-score. These values are presented in Table 2. Through this table, we can compare the evaluation results between the models.

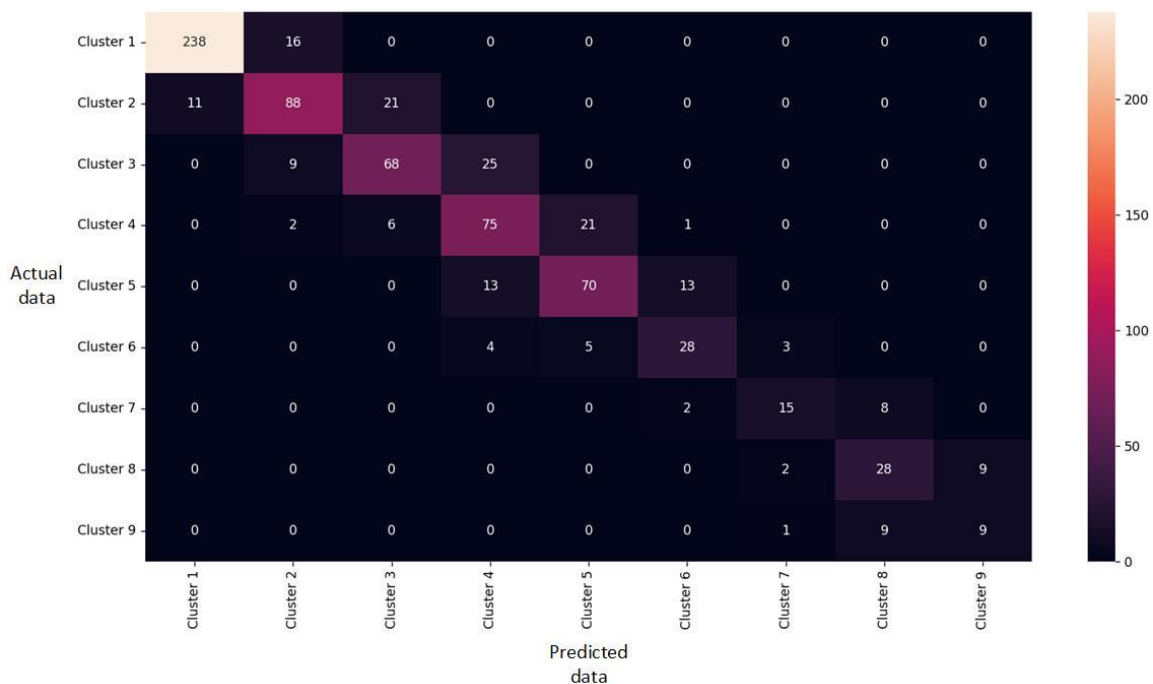


Figure 8. The confusion matrix of the proposed model.

**Table 2.** Precision, recall, and F1-score of the proposed model and counterpart models.

Model	Precision	Recall	F1-Score
LSTM	0.65	0.64	0.64
GRU	0.61	0.63	0.62
STL-GRU	0.60	0.61	0.60
RobustSTL-CNN	0.63	0.62	0.62
The proposed model	0.70	0.70	0.70

In these evaluation metrics, the best result is achieved by a model with the highest values of precision, recall, and F1-score. Based on Table 2, The proposed model remains the best model. It achieves the highest values with precision, recall, and F1-scores of 0.70. The counterpart models mostly have values for precision, recall, and F1-score metrics below 0.70. The proposed model is followed by the LSTM model as the best model in the classification evaluation of clustering data between the actual values and the predicted values. The gap values between the LSTM and the proposed model's precision, recall, and F1-score metrics are 0.05, 0.06, and 0.06. It also shows that only the proposed model reaches these metric values above 0.70. Furthermore, these highest values denote the reliability of the proposed method compared to counterpart models.

#### 4. Conclusions

Power consumption forecasting plays a vital role in describing future demands on power management to help plan sustainable policies. The main issues in hourly power consumption forecasting are the dynamic and intricate patterns. In addition, some burstiness and noise arises in the traffic. The decomposition operation can enhance the understanding of the time series relationship. The hybrid model of RobustSTL decomposition and the TCN model as the means of forecasting indicates excellent results compared to counterpart models. The proposed model obtains optimal accuracy with the lowest value of MAPE, MAE, and RMSE in hourly power consumption forecasting which outperforms the other models. The proposed model also achieves the best values in precision, recall, and F1-score metrics. Moreover, the proposed model can learn the noise and burstiness of the time series dataset to understand the data pattern.

**Author Contributions:** Conceptualization, C.-H.L.; methodology, U.N.; software, C.-H.L.; validation, C.-H.L.; formal analysis, U.N.; resources, C.-H.L.; writing—original draft preparation, U.N.; writing—review and editing, U.N., G.-Z.L. and C.-H.L.; visualization, U.N. and G.-Z.L.; supervision, T.-F.L.; project administration, C.-H.L.; funding acquisition, C.-H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially supported by the Ministry of Science and Technology, Taiwan, under MOST- 110-2218-E-006-014-MBK and MOST- 110-2221-E-992-005-MY2.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in Kaggle (<https://www.kaggle.com/datasets/hgultekin/hourly-power-consumption-of-turkey-20162020>, accessed on 1 November 2021).

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

#### References

1. Zheng, S.; Huang, G.; Zhou, X.; Zhu, X. Climate-change impacts on electricity demands at a metropolitan scale: A case study of Guangzhou, China. *App. Energy* **2020**, *261*, 114295. [[CrossRef](#)]
2. Albuquerque, P.C.; Cajueiro, D.O.; Rossi, M.D.C. Machine learning models for forecasting power electricity consumption using a high dimensional dataset. *Expert Sys. App.* **2022**, *187*, 115917. [[CrossRef](#)]
3. Lee, J.; Cho, Y. National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model. *Energy* **2022**, *239*, 122366. [[CrossRef](#)]

4. Petrosanu, D.M. Designing, developing and validating a forecasting method for the month ahead hourly electricity consumption in the case of medium industrial consumers. *Processes* **2019**, *7*, 310. [CrossRef]
5. Sulandari, W.; Subanar; Lee, M.H.; Rodrigues, P.C. Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks. *Energy* **2020**, *190*, 116408. [CrossRef]
6. Fan, M.; Hu, Y.; Zhang, X.; Yin, H.; Yang, Q.; Fan, L. Short-term load forecasting for distribution network using decomposition with ensemble prediction. In Proceedings of the Chinese Automation Congress (CAC), Hangzhou, China, 22–24 November 2019.
7. Tian, Y.J.; Wen, M.; Li, J.G. A short-term electricity forecasting scheme based on combined GRU model with STL decomposition. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Xi'an, China, 3–6 November 2020.
8. Mendez-Jimenez, I.; Cardenas-Montes, M. Time series decomposition for improving the forecasting performance of convolutional neural networks. In Proceedings of the Advances in Artificial Intelligence, Granada, Spain, 23–26 October 2018.
9. Wen, Q.; Gao, J.; Song, X.; Sun, L.; Xu, H.; Zhu, S. RobustSTL: A robust seasonal-trend decomposition algorithm for long time series. In Proceedings of the 33th AAAI Conference on Artificial Intelligence (AAAI-19), Honolulu, HI, USA, 27 January–1 February 2019.
10. Jiang, F.; Zhang, C.; Sun, S.; Sun, J. Forecasting hourly PM based on deep temporal convolutional neural network and decomposition method. *App. Soft Comp.* **2021**, *113*, 107988. [CrossRef]
11. Wang, Z.; Tian, J.; Fang, H.; Chen, L.; Qin, J. LightLog: A lightweight temporal convolutional network for log anomaly detection on the edge. *Comp. Net.* **2022**, *203*, 108616. [CrossRef]
12. Cleveland, R.B.; Cleveland, W.S.; McRae, J.E.; Terpenning, I. STL: A seasonal-trend decomposition procedure based on Loess. *J. Off. Stat.* **1990**, *6*, 3–73.
13. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271.
14. Yuan, X.; Qi, S.; Wang, Y.; Wang, K.; Yang, C.; Ye, L. Quality variable prediction for nonlinear dynamic industrial processes based on temporal convolutional networks. *IEEE Sens. J.* **2021**, *21*, 20493–20503. [CrossRef]
15. Yin, L.; Xie, J. Multi-feature-scale fusion temporal convolution networks for metal temperature forecasting of ultra-supercritical coal-fired power plant reheater tubes. *Energy* **2022**, *238*, 121657. [CrossRef]
16. Li, J.; Wu, Y.; Li, Y.; Xiang, J.; Zheng, B. The temperature prediction of hydro-generating units based on temporal convolutional network and recurrent neural network. In Proceedings of the 40th Chinese Control Conference (CCC), Shanghai, China, 26–28 July 2021.
17. Katarya, R.; Verma, O.P. Effectual recommendations using artificial algae algorithm and fuzzy c-mean. *Swarm Evol. Comput.* **2017**, *36*, 52–61. [CrossRef]
18. Deng, X.; Liu, Q.; Deng, Y.; Mahdevan, S. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Inf. Sci.* **2016**, *340–341*, 250–261. [CrossRef]
19. Hourly Power Consumption of Turkey (2016–2020). Available online: <https://www.kaggle.com/datasets/hgultekin/hourly-power-consumption-of-turkey-20162020> (accessed on 1 November 2021).
20. –Namina, S.S.; Tavakoli, N.; Namin, A.S. A Comparison of ARIMA and LSTM in Forecasting Time Series. In Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018.
21. Liu, B.; Fu, C.; Bielefield, A.; Liu, Y.Q. Forecasting of Chinese Primary Energy Consumption in 2021 with GRU Artificial Neural Network. *Energies* **2017**, *10*, 1453. [CrossRef]
22. Yang, S.; Deng, Z.; Li, X.; Zheng, C.; Xi, L.; Zhuang, J.; Zhang, Z.; Zhang, Z. A novel hybrid model based on STL decomposition and one-dimensional convolutional neural networks with positional encoding for significant wave height forecast. *Renew. Energy* **2021**, *173*, 531–543. [CrossRef]