

Total electricity consumption forecasting based on temperature composite index and mixed-frequency models

LI Xuerong^{1,2}, SHANG Wei^{1,2,*1}, ZHANG Xun¹, SHAN Baoguo³, WANG Xiang³

(1. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China;

2. MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at UCAS;

3. State Grid Energy Research Institute CO., LTD, Beijing 102209, China)

Abstract The total electricity consumption (TEC) can accurately reflect the operation of the national economy, and the forecasting of the TEC can help predict the economic development trend, as well as provide insights for the formulation of macro policies. Nowadays, high-frequency and massive multi-source data provide a new way to predict the TEC. In this paper, a “seasonal-cumulative temperature index” is constructed based on high-frequency temperature data, and a mixed-frequency prediction model based on multi-source big data (Mixed Data Sampling with Monthly Temperature and Daily Temperature index, MIDAS-MT-DT) is proposed. Experimental results show that the MIDAS-MT-DT model achieves higher prediction accuracy, and the “seasonal-cumulative temperature index” can improve prediction accuracy.

Key words Total electricity consumption; seasonal effect; temperature big data; high-frequency big data; mixed-frequency prediction model

1 Introduction

Since electric power is closely related to industrial production, business activities and residents' living, electricity data could generally reflect the operation condition of the national economy. Electricity statistics are of great value to be explored, which can help the government to formulate macro-control policies and promote governance capacity to look forward the economic or social development.

* Corresponding author: Wei Shang, *E-mail address*: shangwei@amss.ac.cn (W. Shang).

This work was supported by the science and technology project of State Grid Corporation of China (Project Code: 1400-202157207A-0-0-00); the National Natural Science Foundation of China [grant numbers 72273137].

Among the statistical indicators of electricity, total electricity consumption (TEC) is one of the most comprehensive and basic indicators to reflect the electricity consumption situation of a country or region. TEC is generally defined as the total electricity consumption of the primary, secondary and tertiary industries of the country or region, including industrial electricity, agricultural electricity, commercial electricity, residential electricity, public facilities electricity, etc. The important value of TEC lies in that it could reflect the operation condition of the national economy. Accurate prediction of TEC can help track the trend of economic development and provide insights for macro policymaking.

However, the prediction of TEC is a difficult task, and there are few studies in related fields. TEC includes various sectors of electricity consumption with different patterns. Therefore, it is difficult to distinguish the complex factors influencing each other during forecasting, which adds uncertainty to the prediction results. With the development of big data, high-frequency big datasets that can reflect the micro behavior of electricity consumption provide a new idea for the prediction of TEC. At present, the existing researches in related fields mostly focus on the prediction of electricity load [1-3], but there are still few models that can effectively predict TEC by multi-source big datasets.

In this paper, a mixed-frequency prediction method based on temperature composite index and mixed-data sampling (MIDAS) model, MIDAS-MT-DT model, is proposed and applied to TEC prediction, which significantly improves the prediction accuracy compared with the benchmark models. Based on analyzing the electricity consumption behavior in different seasons, this paper constructs the “seasonal-cumulative temperature index”, which can more accurately reflect the electricity consumption behavior affected by temperature. In addition, a high-frequency daily TEC indicator is also introduced into the model to capture other factors except for temperature. In order to simultaneously utilize the above two kinds of high-frequency big data, we propose a mixed-frequency prediction model (MIDAS-MT-DT) for TEC

based on the “season-cumulative temperature index”, and select TEC of Fujian province, China as the sample for empirical research. Through a series of comparative experiments with benchmark models, it is verified that the MIDAS-MT-DT model has higher prediction accuracy, and the “seasonal-cumulative temperature index” has the ability to improve the prediction accuracy. The robustness and superiority of the proposed framework are further verified by comparing it with more benchmark models and multiple time windows.

The main contributions of this paper are in the following aspects: first, we put forward a new perspective of constructing a temperature composite index to predict electricity data. Most previous studies have selected a few specific months (summer or winter) and used temperature data to predict local sample intervals [4-6]. The temperature index constructed in this paper involves all seasons in a unified analytical framework, which is more compatible and helpful to reduce the application cost of the actual system. Moreover, we extend the traditional MIDAS model by incorporating multi-frequency exogenous variables, thus improving the prediction ability of the original model.

The remaining contents of this paper are arranged as follows: Section 2 summarizes the existing literature on electricity consumption prediction and the mixed-frequency model; Section 3 presents the construction of a “seasonal-cumulative temperature index”; Section 4 introduces the mixed frequency TEC forecasting model based on temperature index. Section 5 compares the forecasting results of the models. Section 6 is a summary and outlook.

2 Literature review

In terms of electricity consumption analysis and prediction, there are many forecasting methods proposed by scholars worldwide, which can be roughly divided into classical forecasting methods, traditional forecasting methods and modern

intelligent forecasting methods. Among them, the classical prediction method includes the elastic coefficient method, the calculation of the capacity of the expansion of the industry, etc. The data frequency of traditional prediction methods is mainly annual and monthly. The commonly used models include time series models [1], regression models [2] gray prediction models [3-4], etc. With the development of data processing ability, modern intelligent models have been widely used in electricity consumption forecasting with monthly and daily basis data. Scholars have employed various models including the neural network prediction method [5], support vector machine [6], chaos theory prediction method [7], also include other combination forecast methods, etc.

In recent years, big data technology has gradually been applied in the research of electricity data prediction [8-10]. The model of [11] found that for every 1-degree increase in temperature, peak electricity consumption would increase by 0.45% to 4.6%. Also using deep learning models, Bedi and Toshniwal (2020) propose a deep learning based hybrid approach that firstly implements Variational Mode Decomposition (VMD) and Autoencoder models to extract meaningful sub-signals/features from the data [12]. Ayub et al. (2020) applied the GRU-CNN model to predict the daily electricity consumption of the ISO-NE data set, which improved the prediction accuracy by 7% compared with the SOTA benchmark model [13]. Cui et al. (2023) propose a deep learning framework with a COVID-19 adjustment for electricity demand forecasting [14]. In the study of [15], the adaptive WT (AWT)-long short-term memory (LSTM) is integrated into a hybrid approach for predicting electricity consumption.

Mixed-frequency models have initially been applied to the field of meteorology, and the basic principle is to explore the information contained in the high-frequency data and predict the future before the official release of relevant statistical data. The MIDAS model is a widely used mixed-frequency model, proposed by Ghysels et al. (2004) [16]. Subsequently, many scholars have proposed extended forms of the MIDAS model, such as the MS-MIDAS model [17] and the co-integration MIDAS model [18].

More recently, the MF-VAR model has been applied to estimating the combined endogenous variables [19].

Due to the demand for in-time forecasting in many industries, mixed-frequency models have been applied to wind power forecasting, rail transit passenger flow forecasting, macroeconomic forecasting, financial market forecasting and many other fields. For example, some scholars applied multi-task learning and ensemble decomposition methods to forecast wind power [20-22]. The in-time forecasting of traffic ridership by Yao Enjian et al. (2018) and Bao Lei (2017) have greatly improved the emergency response ability of the traffic system in emergencies [23, 24]. Currently, mixed-frequency models are also widely used in macroeconomic and financial markets [25-27]. For example, Zhang Wei et al. (2020) [28] and Ghysel and Sinko (2011) [29] respectively forecast Gross Domestic Product (GDP) and financial market volatility.

In summary, the existing literature applies various intelligent algorithms and forecasts electricity data using historical datasets. Many studies have applied meteorological big data or remote sensing big data and other natural environment data. Most of the existing models use temperature data in specific months to forecast local sample intervals but have not considered the multi-source high-frequency big data and other predictive information, therefore the forecasting accuracy is expected to be further improved.

3 Data and variables

This section presents the data collection and preprocessing, as well as the construction of the “seasonal-cumulative temperature index”. Figure 1 presents the methodology framework of the MIDAS-MT-DT model for TEC forecasting. The framework includes the following steps: 1) Collect the daily temperature data, daily TEC data, monthly temperature data and monthly TEC data in the historical data; 2) The daily temperature index is obtained by cumulative transformation and seasonal transformation. The monthly temperature index is obtained by seasonal transformation.

3) The monthly temperature index and the lagged variable of the monthly TEC are taken as the low-frequency forecasting variables, and the daily temperature index and daily TEC are taken as the high-frequency forecasting variables. The low-frequency and high-frequency variables are used to predict the monthly TEC. The details of the above framework are presented in section 3 and 4.

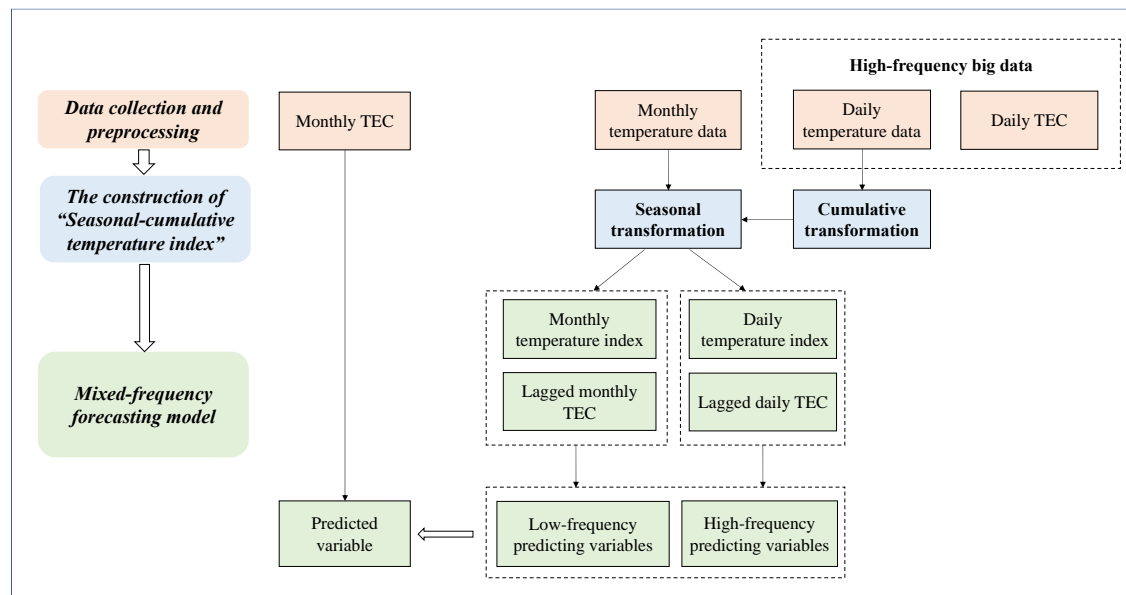


Figure 1. The framework of the MIDAS-MT-DT model

3.1 Data collection and preprocessing

In our empirical study, the monthly TEC of Fujian Province is selected as the predicted variable. The sample period is from January 2017 to November 2020, and the data source is the Wind database. The high-frequency big data used in the prediction model includes 1) the daily TEC of Fujian Province, which is provided by State Grid Energy Research Institute Co., LTD.; 2) The temperature data, which includes Xiamen, Putian, Fuzhou, Nanping, Quanzhou, Ningde, Longyan, Sanming and Zhangzhou of Fujian Province, and the data source is Wind database. The average of daily maximum temperature on 9 cities of Fujian is taken as the original daily temperature data of the temperature index ($T_{i,m}$ in eq. (1)); the average of the monthly average temperature on 9 cities of Fujian is taken as the original monthly temperature data of the temperature index (T_m in eq. (3)). In the out-of-sample forecasting, we first use ARMA model to predict temperature index on testing periods, and then the predicted values of temperature index are inputted into our forecasting models. The sample period of the

above high-frequency data is from January 1, 2017, to November 30, 2020.

3.2 Seasonal-cumulative temperature index

Combined with seasonal changes, it can be analyzed that the electricity consumption behavior has the following two characteristics: (1) Seasonal effect: when the temperature is higher than the comfortable temperature, industrial production, commercial and residential sectors need to use air conditioning to cool down. At the same time, because Fujian province is in the southern region of China when the temperature is lower than the comfortable temperature in winter, it also needs to use air conditioning for heating. In summary, summer temperature should be positively correlated with electricity consumption, while winter temperature should be negatively correlated with electricity consumption. (2) Cumulative effect: the behavior of electricity consumption has a certain inertia. The behavior of using air conditioning in the first few days tends to continue for a short time, so the temperature of the first day has a certain impact on the next few days.

Based on the two characteristics, the daily temperature data is transformed through cumulative transformation and seasonal transformation. The formula of cumulative transformation is:

$$C_{T_{i,m}} = \begin{cases} \left(1 - \sum_{j=1}^4 e^{-j}\right) * T_{i,m} + \sum_{j=1}^4 e^{-j} * T_{i-j,m}, & i \geq 5 \\ \left(1 - \sum_{j=1}^4 e^{-j}\right) * T_{i,m} + \sum_{j=1}^i e^{-j} * T_{i-j+1,m} + \sum_{j=i}^{4-i} e^{-(4-j)} * T_{N-i+1,m-1}, & i = 1,2,3,4 \end{cases} \quad (1)$$

where $C_{T_{i,m}}$ is the daily temperature index of the i day of the m month transformed by cumulative effect; $T_{i,m}$ is the original daily temperature data of the i day of the m month; N indicates the total number of days in the $m-1$ month. j represents the number of days before i . In our study, we assume the cumulative effect of electricity consumption behavior lasts 5 days, thus $j \in \{1,2,3,4\}$; e^{-j} represents the influence coefficient of the temperature of the previous j day, which decreases by the trend of natural logarithm over time.

After that, the daily temperature data is transformed by seasonal effect transformation, wherein the formula is:

$$SC_{T_{i,m}} = \begin{cases} C_{T_{i,m}}, & m \in \{5,6,7,8,9\}, i = 1,2, \dots, N \\ L - C_{T_{i,m}}, & m \in \{10,11,12,1,2,3,4\}, i = 1,2, \dots, N \end{cases} \quad (2)$$

where, $SC_{T_{i,m}}$ is the daily temperature index of the i day of the m month after cumulative effect transformation and seasonal effect transformation, and N is the total number of days of the m month. L is the displacement length to ensure that the temperature index after seasonal transformation remains continuous. In this paper, we estimate the value of L by computing the average of $dist(-C_{T_{i,4}}, C_{T_{i,5}}) + dist(C_{T_{i,9}}, -C_{T_{i,10}})$ of each year in the sample period, where $dist(a, b)$ represents the distance between a and b , that is, $dist(a, b) = |a - b|$.

Furthermore, the monthly temperature data is transformed by seasonal effect to obtain the monthly temperature index, in which the formula of seasonal effect transformation is:

$$S_{T_m} = \begin{cases} T_m, & m \in \{5,6,7,8,9\} \\ L - T_m, & m \in \{10,11,12,1,2,3,4\} \end{cases} \quad (3)$$

where S_{T_m} is the monthly temperature index of the m month after seasonal effect transformation; T_m is the original monthly temperature data of the m month; L is the same as defined in eq. (4).

4 Methodology

In this section, the MIDAS-MT-DT mixed-frequency TEC forecasting model is presented, as well as the single-frequency TEC forecasting models used for benchmark models. After that, we present our experimental design of comparative experiments.

4.1 Mixed-frequency forecasting model based on multi-source big data

In order to comprehensively utilize high-frequency temperature data and high-frequency electricity consumption data, the MIDAS model of Ghysels et al. (2004) is extended in this paper, and a mixed-frequency prediction model of TEC (MIDAS-MT-DT) based on “season-cumulative temperature index” is proposed. The formula of the model is as follows:

$$\begin{aligned}
 Y_{M,t+1} = & \mu + \sum_{j=0}^{p_Y-1} \mu_{j+1} Y_{t-j} + \sum_{j=0}^{p_X-1} \beta_{j+1} S_{-}T_{t-j} \\
 & + \beta \sum_{j=0}^{p_X-1} \sum_{i=0}^{N-1} w_{N-i+j*N}(\boldsymbol{\theta}) Y_{D,N-i,t-j} + \delta \sum_{j=0}^{p_X-1} \sum_{i=0}^{N-1} w_{N-i+j*N}(\boldsymbol{\theta}) SC_{-}T_{N-i,t-j} + u_{t+1}
 \end{aligned} \tag{4}$$

where, $Y_{M,t+1}$ is the monthly TEC of the t month, $S_{-}T_t$ is the monthly temperature index of the t month after seasonal transformation, $Y_{D,N,t}$ is the daily TEC of the N day of the t month, $SC_{-}T_{N,t}$ is the daily temperature index of the N day of the t month after cumulative transformation and seasonal transformation. $w_i(\boldsymbol{\theta})$ is a high-frequency variable lag weighting polynomials of MIDAS model, $\boldsymbol{\theta}$ is the estimated parameter polynomial, and $\sum_{j=0}^{p_X-1} \sum_{i=0}^{N-1} w_{N-i+j*N}(\boldsymbol{\theta}) = 1$; μ , μ_{j+1} , β , β_{j+1} , δ are the parameters to be estimated by the model, p_X and p_Y are the optimal lag period of the model selected by AIC criterion, i and j are the integers in the summative operator, and u_{t+1} is the random error of the model. The parameters are estimated by Non-linear Least Squares (NLS).

In the comparison experiment, we employ Almon and Beta lag weight function of MIDAS model, which are defined as

$$w_i^{Almon}(\theta_1, \theta_2) = \frac{e^{\theta_1 i + \theta_2 i^2}}{\sum_{i=1}^N e^{\theta_1 i + \theta_2 i^2}} \tag{5}$$

and

$$w_i^{Beta}(\theta_1, \theta_2, \theta_3) = \frac{x_i^{\theta_1-1} (1-x_i)^{\theta_2-1}}{\sum_{i=1}^N x_i^{\theta_1-1} (1-x_i)^{\theta_2-1}} + \theta_3 \tag{6}$$

where $x_i = (i - 1)/(N - 1)$.

4.2 Single-frequency forecasting models

In order to verify the predictive power of the MIDAS-MT-DT model and the “seasonal-cumulative temperature index”, several comparative experiments of single-frequency forecasting models are conducted. First, daily and monthly Autoregressive Moving Average (ARMA) models are used to compare the prediction accuracy with and without the temperature index. Specifically, the monthly ARMA model is:

$$Y_{M,t+1} = \mu + \sum_{j=0}^{p_Y-1} \mu_{j+1} Y_{M,t-j} + \sum_{j=0}^{p_X-1} \beta_{j+1} S_{-}T_{t-j} + \sum_{j=0}^{p_Z-1} \delta_{j+1} u_{t+1} \tag{7}$$

where $Y_{M,t}$ is the monthly TEC in month t , S_T is the monthly temperature index of the t month after seasonal transformation, and u_{t+1} is an independent and identically distributed random variable, representing the model error.

The daily ARMA model is:

$$Y_{D,t+1} = \mu + \sum_{j=0}^{p_Y-1} \mu_{j+1} Y_{D,t-j} + \sum_{j=0}^{p_X-1} \beta_{j+1} SC_T_{t-j} + \sum_{j=0}^{p_Z-1} \delta_{j+1} u_{t+1} \quad (8)$$

where, $Y_{D,t}$ is the daily TEC on the day t , and SC_T is the daily high-frequency temperature index on the day t after seasonal and cumulative transformation.

4.3 Experimental design

The prediction accuracy of MIDAS models with daily or monthly temperature index and without temperature index are compared respectively. Furthermore, mixed-frequency models are compared with several single-frequency models. In addition to ARMA models, some intelligent models such as Support Vector Regression (SVR) and Random Forest (RF) model are selected as the benchmark models. To sum up, the model specifications and parameters of the comparison experiments are shown in Table 1.

Table 1. Model specifications and parameters

Model label	Model specifications	Model Parameters
ARMA-M	Single-frequency monthly ARMA model, with lagged variables and without temperature index	In eq. (7), $\beta_j = 0$
ARMA-D	Single-frequency daily ARMA model, with lagged variables and without temperature index	In eq. (8), $\beta_j = 0$
ARMA-MT	Monthly ARMA model with monthly temperature index and lagged variables	In eq. (7), $\beta_j \neq 0$
ARMA-DT	Daily ARMA model with daily temperature index and lagged variables	In eq. (8), $\beta_j \neq 0$
SVR-M	Single-frequency monthly SVR model, with lagged variables and without temperature index	
SVR-D	Single-frequency daily SVR model, with lagged variables and without temperature index	Default values of Python Scikit-
SVR-MT	Monthly SVR model with monthly temperature index and lagged variables	learn library
SVR-DT	Daily SVR model with daily temperature index and lagged variables	

RF-M	Single-frequency monthly RF model, with lagged variables and without temperature index	
RF-D	Single-frequency daily RF model, with lagged variables and without temperature index	
RF-MT	Monthly RF model with monthly temperature index and lagged variables	
RF-DT	Daily RF model with daily temperature index and lagged variables	
MIDAS	Mixed-frequency benchmark model, without temperature index	In eq. (4), $\beta_j = 0, \delta = 0$
MIDAS-MT	MIDAS model with monthly temperature index and its lagged variables	In eq. (4), $\beta_j \neq 0, \delta = 0$
MIDAS-DT	MIDAS model with daily temperature index	In eq. (4), $\beta_j = 0, \delta \neq 0$
MIDAS-MT-DT	MIDAS model with monthly temperature index and its lagged variables, as well as the daily temperature index	In eq. (4), $\beta_j \neq 0, \delta \neq 0$

5 Empirical results

In this section, the forecasting performances of the MIDAS-MT-DT model and the “seasonal-cumulative temperature index” are illustrated through a series of comparative experimental results. First, the description of the “seasonal-cumulative temperature index” are presented, and then the prediction results of the single-frequency models and the mixing-frequency models are compared respectively.

5.1 Data description and correlation analysis

According to the construction method in Section 3.2, the description of the “season-cumulative temperature index” is shown in Figure 2. According to the results in the figure, the temperature index constructed in this paper maintains a general trend of positive correlation with Fujian TEC, which indicates that it may improve the forecasting performance of temperature data.

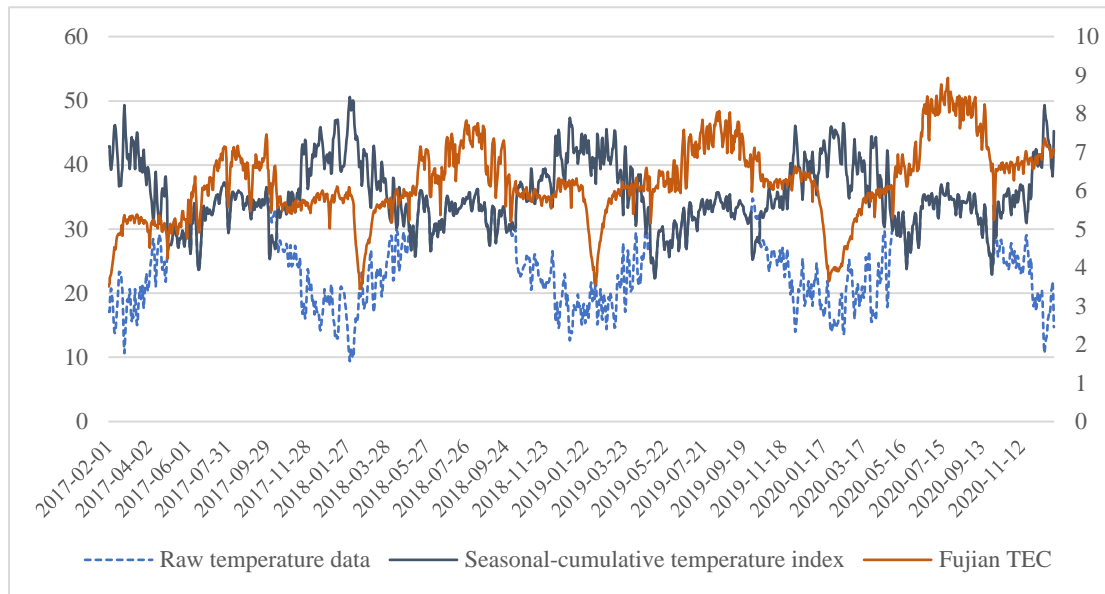


Figure 2. The description of “season-cumulative temperature index”

Table 2 shows the descriptive statistics and correlation analysis results of the raw temperature data, “seasonal-cumulative temperature index” and daily Fujian TEC. The results show that the correlation coefficient between the “seasonal-cumulative temperature index” and Fujian TEC is 0.6540, while the correlation coefficient between the raw temperature data and Fujian TEC is only -0.3060. This indicates that the temperature index construction method in this paper can effectively improve the correlation with the predicted variables.

Table 2. Descriptive statistics and correlation analysis

	Raw temperature data	Seasonal-cumulative temperature index	Fujian TEC
Mean	26.1478	35.5638	6.2101
Maximum	9.4058	50.5942	3.4423
Minimum	37.2591	22.3403	8.9278
Std. Dev.	6.5261	5.1441	0.9815
Skewness	-1.0306	0.2943	0.2392
Kurtosis	-0.2506	2.6980	-0.0345
Pearson correlation coefficient with Fujian TEC	-0.3060	0.6540	1

5.2 Single-frequency forecasting models

To verify the predictive ability of the “seasonal-cumulative temperature index”, daily and monthly models are used to compare the prediction accuracy with and without the temperature index, respectively. The prediction results are shown in Table 3. In the table, column 1 represents the testing period, and the corresponding training period is from the beginning of the sample to the previous month of the testing period. The prediction accuracies are calculated by the following formulas:

$$Accuracy (ACC) = 1 - \frac{1}{N} \sum_{t=1}^N \left| \frac{\hat{x}_t - x_t}{x_t} \right| \quad (7)$$

$$Root Mean Squared Error (RMSE) = \sqrt{\frac{1}{N} \sum_{t=1}^N \left(\frac{\hat{x}_t - x_t}{x_t} \right)^2} \quad (8)$$

where the \hat{x}_t and x_t represent the predicted value and the real value of the forecast model, respectively.

Table 3. Prediction results of single-frequency models

<i>Panel A: Monthly frequency models</i>							
Testing period		ARMA-M	ARMA-MT	SVR-M	SVR-MT	RF-M	RF-MT
2020.1-	ACC	63.42%	83.20%	76.84%	92.67%	77.60%	92.85%
2020.3	RMSE	17.7056	16.0453	23.9215	7.2015	24.1843	7.9507
2020.4-	ACC	62.89%	70.12%	87.11%	93.11%	86.54%	91.71%
2020.6	RMSE	54.6691	53.8693	27.9021	14.8238	20.8601	15.2980
2020.7-	ACC	62.43%	82.31%	87.48%	89.13%	81.16%	86.97%
2020.9	RMSE	58.1760	33.7527	23.9188	21.8095	31.4153	25.0699
2020.10-	ACC	69.75%	75.52%	74.02%	89.64%	73.95%	85.67%
2020.11	RMSE	48.1106	33.5277	39.9282	16.1358	34.1574	21.0209
<i>Panel B: Daily frequency models</i>							
Testing period		ARMA-D	ARMA-DT	SVR-D	SVR-DT	RF-D	RF-DT
2020.8.1-	ACC	61.10%	97.72%	82.92%	95.49%	85.16%	94.98%
2020.8.31	RMSE	3.1863	0.2353	1.6579	0.4335	1.4492	0.4886
2020.9.1-	ACC	67.51%	89.07%	84.03%	95.76%	87.21%	88.73%
2020.9.30	RMSE	2.5071	0.9195	1.4930	0.3945	1.2188	0.9456
2020.10.1-	ACC	76.72%	96.39%	74.48%	95.24%	82.61%	93.19%
2020.10.31	RMSE	1.5694	0.3934	1.9787	0.3897	1.3973	0.5174
2020.11.1-	ACC	74.17%	98.23%	84.62%	95.22%	79.36%	94.30%

2020.11.30	RMSE	1.7447	0.1587	1.3304	0.3790	1.6610	0.4516
------------	------	--------	---------------	--------	---------------	--------	---------------

Note: The bold numbers in the table indicate models with improved predictive accuracy compared to the benchmark model.

According to the results in the table, in all testing periods, regardless on the daily or the monthly basis, the prediction accuracies of forecasting models added with the temperature index are significantly improved compared with the benchmark models. Intelligent models such as SVR and RF perform much better than ARMA models in the cases of the monthly models. On the daily basis, intelligent models and ARMA models are competitive. Among them, the highest prediction accuracy has reached 98.23%. The results in Table 3 show that the temperature index constructed in this paper can significantly improve the forecasting ability of the benchmark models by accurately reflecting the influence of electricity consumption behavior on TEC.

5.3 Mixed-frequency forecasting models

In order to verify the prediction ability of the MIDAS-MT-DT model, the prediction accuracy of the MIDAS model with daily, and monthly temperature index and without temperature index are compared respectively. The prediction results are shown in Table 4. The first column in the table represents the testing period of the model, and the corresponding training period is from the beginning of the sample period to the previous month of the testing period. The prediction accuracies of the corresponding models are calculated by eq. (7-8).

Table 4. Prediction results of mixed-frequency models

Panel A: Total Electricity Consumption of Fujian									
Testing period		Almon-MIDAS				Beta-MIDAS			
		MIDAS	MIDAS-MT	MIDAS-DT	MIDAS-MT-DT	MIDAS	MIDAS-MT	MIDAS-DT	MIDAS-MT-DT
2020.1-2020.3	ACC	83.21%	86.15%	90.86%	86.15%	80.68%	94.30%	81.87%	96.93%
	RMSE	16.6583	21.0858	10.4433	12.4566	18.8926	5.0649	18.6801	3.4150
2020.4-2020.6	ACC	96.54%	93.35%	94.00%	98.39%	84.34%	91.62%	90.94%	94.72%
	RMSE	4.8878	7.2637	9.0697	2.3340	23.8644	14.2741	14.4594	7.3290
2020.7-2020.9	ACC	94.45%	94.91%	94.30%	96.29%	93.64%	91.13%	87.83%	96.73%
	RMSE	9.0774	7.5643	7.9832	6.1133	11.5742	17.2138	20.4678	5.1104
2020.10-2020.11	ACC	94.23%	95.62%	89.58%	95.18%	90.66%	93.09%	93.38%	98.02%
	RMSE	5.3534	4.2126	12.2389	3.5138	13.0001	12.4452	8.8826	3.7987

Panel B: Residential Electricity Consumption of Fujian									
		Almon-MIDAS				Beta -MIDAS			
Testing period		MIDAS	MIDAS-MT	MIDAS-DT	MIDAS-MT-DT	MIDAS	MIDAS-MT	MIDAS-DT	MIDAS-MT-DT
2020.1-2020.3	ACC	91.51%	94.84%	90.57%	92.69%	80.89%	85.07%	84.96%	92.33%
	RMSE	3.4390	2.1221	3.8354	4.1736	16.4449	14.7521	15.5679	9.0528
2020.4-2020.6	ACC	88.79%	87.20%	90.74%	94.96%	88.86%	92.04%	83.15%	95.59%
	RMSE	5.4743	5.6525	4.7849	2.2653	20.1651	11.3245	26.7171	7.5793
2020.7-2020.9	ACC	88.39%	94.68%	86.43%	94.59%	82.40%	91.79%	92.87%	94.95%
	RMSE	7.8735	4.4300	9.9840	3.1155	26.8130	11.6818	10.3671	11.6454
2020.10-2020.11	ACC	87.59%	92.30%	95.49%	93.74%	86.68%	85.01%	90.50%	92.23%
	RMSE	5.0707	3.5147	1.7105	2.5531	21.6887	26.0945	13.6509	13.8566

Panel C: The Tertiary Industry Electricity Consumption of Fujian									
		Almon-MIDAS				Beta -MIDAS			
Testing period		MIDAS	MIDAS-MT	MIDAS-DT	MIDAS-MT-DT	MIDAS	MIDAS-MT	MIDAS-DT	MIDAS-MT-DT
2020.1-2020.3	ACC	82.26%	91.45%	81.06%	94.32%	84.00%	83.56%	78.24%	93.40%
	RMSE	4.8074	2.7748	4.4253	1.6667	4.1694	4.7109	4.7102	2.5181
2020.4-2020.6	ACC	91.48%	93.18%	85.19%	94.57%	84.77%	81.86%	89.79%	91.51%
	RMSE	3.0102	1.8038	4.0636	1.4146	5.1318	5.9643	3.3482	3.0738
2020.7-2020.9	ACC	90.77%	89.33%	93.59%	97.24%	77.35%	90.60%	82.25%	86.34%
	RMSE	5.3767	5.5239	3.2686	1.3215	11.3417	4.8776	7.9887	6.7196
2020.10-2020.11	ACC	95.12%	90.61%	90.73%	97.25%	83.82%	90.98%	79.14%	91.15%
	RMSE	2.0203	3.9438	3.2524	0.9801	5.93041	3.0239	7.1812	3.2041

Note: The bold numbers in the table indicate models with improved predictive accuracy compared to the benchmark MIDAS model.

According to the results in the table, in all testing periods, regardless of whether the daily temperature index or monthly temperature index is added, the prediction accuracies of the models are significantly improved compared with the benchmark models. Among them, the highest prediction accuracy has reached 98.39%. In addition, the MIDAS-MT model and MIDAS-MT-DT model have obtained higher prediction accuracies than the benchmark models in most of the four testing periods. However, only one of the MIDAS-DT models achieves higher prediction accuracy. This indicates that the monthly temperature index has a better predictive ability than the daily temperature index. By comparing different lag weighting polynomials of MIDAS models, we also find that Almon-MIDAS models perform better than Beta-MIDAS on most of the data samples.

To verify the robustness of our model on more datasets, we further apply our models to forecast Residential Electricity Consumption (REC) of Fujian and The Tertiary Industry Electricity Consumption of Fujian. The results in Panel B and C of

Table 4 show that MIDAS-MT-DT model achieves better performances on all testing periods of multiple datasets. We have also observed similar results as those in Panel A that, the monthly temperature index has a stronger ability to improve the prediction accuracy in the mixing-frequency model. Overall, our results demonstrate that the MIDAS-MT-DT model proposed in this paper significantly improves the prediction accuracy of the TEC by incorporating high-frequency temperature data and high-frequency TEC data, and this advantage is not easily affected by the randomness of the data set with good robustness.

6 Conclusions and future research

The total electricity consumption (TEC) reflects the operation of the national economy. Accurate prediction of the TEC is of great significance for the country to look forward the economic development and formulate macro-control policies. The high-frequency and massive multi-source data provides a new idea for the prediction of TEC. Based on the analysis of electricity consumption behavior in different seasons, this study constructs a “seasonal-cumulative temperature index” considering the inertia of seasons and electricity consumption behavior, which can reflect the electricity consumption behavior affected by temperature. In addition, high-frequency daily data of the TEC is also incorporated to supplement the electricity consumption behavior affected by other factors. Based on the above two high-frequency datasets, this study proposes a mixed-frequency prediction model (MIDAS-MT-DT) for the TEC based on the “seasonal-cumulative temperature index” and mixed-frequency models.

According to the empirical results, the temperature index constructed in this paper is able to significantly improve the forecasting ability of the benchmark model by reflecting the electricity consumption behavior affected by temperature and other factors. By incorporating high-frequency temperature data and daily TEC data, the MIDAS-MT-DT model proposed in this paper captures the intricate factors of electricity consumption behavior, thus significantly improves the prediction accuracy of the TEC, and the highest accuracy has reached 98.39%. Through comparative experiments, we find that the monthly temperature index has a stronger ability to improve the prediction accuracy in the mixing-frequency model. The experiments of

multiple testing periods and predicted datasets further verify the robustness of the MIDAS-MT-DT model.

In terms of the limitations of this paper, future research directions include: In addition to the big data of temperature, other microscopic data that can reflect the behavior of electricity consumption can be further collected, and more exogenous variables can be introduced into the prediction model. Examples include remote sensing data and internet data. In terms of the prediction model, the fusion technology of multi-source electricity big data, machine learning and deep learning models can be explored to further improve the prediction accuracy.

Author Contributions

Xuerong Li has collected and processed experimental data, presented experiment results and drafted the original version of the manuscript; Wei Shang has guided the overall direction of this research, and provided advice for comparison experiments. Xun Zhang provided advice for motivation and conclusions presented in Section 1. Baoguo Shan and Xiang Wang have funded the research, provided the proposal of the research, and contributed part of the experimental data. All the authors have made meaningful and valuable contributions in revising and proofreading the resulting manuscript.

References

- [1] A. Hussain, M. Rahman & J. A. Memon. Forecasting electricity consumption in Pakistan: The way forward. *Energy Policy* 90 (2016), 73-80. doi: 10.1016/j.enpol.2015.11.028.
- [2] Y. Feng & S. M. Ryan. Day-ahead hourly electricity load modeling by functional regression. *Applied Energy*, 170 (2016), 455-465. doi: 10.1016/j.apenergy.2016.02.118.

- [3] W. Xu, R. Gu, Y. Liu & Y. Dai. Forecasting energy consumption using a new GM–ARMA model based on HP filter: The case of Guangdong province of China. *Economic Modelling*, 45 (2015), 127-135. doi: 10.1016/j.econmod.2014.11.011.
- [4] L. Zeng, C. Liu & W. Z. Wu. A novel discrete GM (2, 1) model with a polynomial term for forecasting electricity consumption. *Electric Power Systems Research*, 214 (2023), 108926. doi: 10.1016/j.epsr.2022.108926.
- [5] N. Kunwar, K. Yash & R. Kumar. Area-load based pricing in DSM through ANN and heuristic scheduling. *Smart Grid*, 4 (2013), 1275-1281. doi: 10.1109/TSG.2013.2262059.
- [6] Q. Cheng, Y. Yan, S. Liu, C. Yang, H. Chaoui & M. Alzaed. Particle filter-based electricity load prediction for grid-connected microgrid day-ahead scheduling. *Energies*, 13 (2020), 6489. doi: 10.3390/en13246489.
- [7] Z. J. Liu, H. M. Yang & M. Y. Lai. Electricity price forecasting model based on chaos theory. In: *Proceedings of 2005 International Power Engineering Conference*, 2005, pp. 1-9. doi: 10.1109/IPEC.2005.206950.
- [8] X. Guo, Q. Zhao, D. Zheng, Y. Ning & Y. Gao. A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price. *Energy Reports*, 6 (2020), 1046-1053. doi: 10.1016/j.egy.2020.11.078.
- [9] P. Jiang, Y. Nie, J. Wang & X. Huang. Multivariable short-term electricity price forecasting using artificial intelligence and multi-input multi-output scheme. *Energy Economics*, 117 (2023), 106471. doi: 10.1016/j.eneco.2022.106471.
- [10] Y. Jiang, T. Gao, Y. Dai, R. Si, J. Hao, J. Zhang & D. W. Gao. Very short-term residential load forecasting based on deep-autoformer. *Applied Energy*, 2022, 328: 120120. doi: 10.1016/j.apenergy.2022.120120.
- [11] M. Santamouris, C. Cartalis, A. Synnefa, D. Kolokotsa. On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—A review. *Energy & Buildings*, 98 (2015), 119-124. doi: 10.1016/j.enbuild.2014.09.052.
- [12] J. Bedi & D. Toshniwal. Energy load time-series forecast using decomposition and autoencoder integrated memory network. *Applied Soft Computing*, 93 (2020), 106390. doi: 10.1016/j.asoc.2020.106390.
- [13] N. Ayub, M. Irfan, M. Awais, U. Ali, T. Ali, M. Hamdi, A. Alghamdi & F. Muhammad. Big data analytics for short and medium term electricity load forecasting using AI techniques ensembler. *Energies*, 13 (2020), 5193. doi: 10.3390/en13195193.

- [14] Z. Cui, J. Wu, W. Lian & Y. G. Wang. A novel deep learning framework with a COVID-19 adjustment for electricity demand forecasting. *Energy Reports*, 9 (2023), 1887-1895. doi: 10.1016/j.egy.2023.01.019.
- [15] A. Saranj & M. Zolfaghari. The electricity consumption forecast: Adopting a hybrid approach by deep learning and ARIMAX-GARCH models. *Energy Reports*, 8 (2022), 7657-7679. doi: 10.1016/j.egy.2022.06.007.
- [16] E. Ghysels, P. Santa-Clara P & R. Valkanov. The MIDAS touch: Mixed data sampling regression models. UC Los Angeles: Finance. (2004). Available at: <http://escholarship.org/uc/item/9mf223rs>.
- [17] P. Guérin & M. Marcellino. Markov-switching MIDAS model. *Journal of Business & Economic Statistics*, 31 (2013), 45-56. doi: 10.1080/07350015.2012.727721.
- [18] J. I. Miller. Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. *Journal of Financial Econometrics*, 12 (2014), 584-614. doi: 10.1093/jjfinec/nbt010.
- [19] R. Kikuchi, T. Misaka, S. Obayashi, H. Inokuchi, H. Oikawa & A. Misumi. Nowcasting algorithm for wind fields using ensemble forecasting and aircraft flight data. *Meteorological Applications*, 25 (2018), 365-375. doi: 10.1002/met.1704.
- [20] A. Dupré, P. Drobinski P, J. Badosa, C. Briard & P. Tankov. The economic value of wind energy nowcasting. *Energies*, 13 (2020), 5266. doi: 10.3390/en13205266.
- [21] I. Kutiev, P. Muhtarov, B. Andonov & R. Warnant. Hybrid model for nowcasting and forecasting the K index. *Journal of Atmospheric and Solar-Terrestrial Physics*, 71 (2009), 589-596. doi: 10.1016/j.jastp.2009.01.005.
- [22] Y. Wei & M. C. Chen. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C Emerging Technologies*, 21 (2012), 148-162. doi: 10.1016/j.trc.2011.06.009.
- [23] M. Ni, Q. He & J. Gao. Forecasting the subway passenger flow under event occurrences with social media. *IEEE Transactions on Intelligent Transportation Systems*, 18 (2017), 1623-1632. doi: 10.1109/TITS.2016.2611644.
- [24] V. Kuzin, M. Marcellino & C. Schumacher. MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, 27 (2011), 529-542. doi: 10.1016/j.ijforecast.2010.02.006.
- [25] E. Andreou, E. Ghysels, A & Kourtellos. Should macroeconomic forecasters use daily financial data and how?. *Journal of Business & Economic Statistics*, 31 (2013), 240-251. doi: 10.1080/07350015.2013.767199.
- [26] F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7 (2009), 174-196. doi: 10.1093/jjfinec/nbp001.

- [27] H. Bahcivan & C. C. Karahan. High frequency correlation dynamics and day-of-the-week effect: A score-driven approach in an emerging market stock exchange. *International Review of Financial Analysis*, 80 (2022), 102008. doi: 10.1016/j.irfa.2021.102008.
- [28] A. Richardson, T. Mulder & T. Vehbi. Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International journal of forecasting*, 37 (2021), 941-948. doi: 10.1016/j.ijforecast.2020.10.005.
- [29] E. Ghysels & A. Sinko. Volatility forecasting and microstructure noise. *Journal of Econometrics*, 160 (2011), 257-271. doi: 10.1016/j.jeconom.2010.03.035.

Author Biography



Dr. Xuerong Li, assistant professor of Academy of Mathematics and Systems Science, Chinese Academy of Sciences. She received the B.E. degree from Renmin University of China and the Ph.D. degree from University of Chinese Academy of Sciences in 2013 and 2019 respectively. Her research interests are in the area of economic forecasting and machine learning.



Prof. Wei Shang, associate professor of Academy of Mathematics and Systems Science, Chinese Academy of Sciences, chief engineer of Center for Forecasting Science, Chinese Academy of Sciences. She holds a PhD from Harbin Institute of Technology. Her research focuses on macro-economic monitoring and early warning, Internet Data Mining, e-business and information technology and development. She published her researches in major journals such as *Decision Support Systems*, *Electronic Research and Applications*, and *Online Information Review*. She is principle investigator of three regular project of National Science Foundation of China, as well as an ICT4D project of Information Development Research Center (Canada).



Prof. Xun Zhang, associate professor of Academy of Mathematics and Systems Science, Chinese Academy of Sciences. She received the B.E. degree from Renmin University of China and the Ph.D. degree from Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2004 and 2009 respectively. Her research interests are in the area of macroeconomic forecasting and energy economics.



Shan Baoguo is the vice president of State Grid Energy Research Institute and a professor-level senior engineer. He received his bachelor's degree and master's degree from North China Electric Power University in 1993 and 1997, respectively. His research interests are energy power analysis and forecasting, power demand side management.



Dr. Wang Xiang is a senior economist of State Grid Energy Research Institute. He received his bachelor's degree from Jilin University in 2011 and his Ph.D.'s degree from Nankai University in 2014. His research interests are macroeconomic and power market analysis and forecasting.