Improving forecasting accuracy of daily energy consumption of office building using time series analysis based on wavelet transform decomposition

Chengkuan Fang¹, Yuan Gao¹ and Yingjun Ruan²

¹ Postgraduate Student, Tongji University, China
² Associate Professor, Tongji University, China
E-mail: ruanyj@tongji.edu.cn

Abstract. In order to improve the operation, detection and diagnosis of district energy systems, it is necessary to develop energy demand prediction models. Several models for energy prediction have been proposed, including machine learning methods and time series analysis methods. Data-driven machine learning methods fail to achieve the expected accuracy due to the lack of measurement data and the uncertainty of weather forecasts, additionally it is not easy to obtain complete and long-term weather data sets of building as input data in China. In this case, a WT-ARIMA prediction model that combines wavelet transform and time series analysis without meteorological parameters can be a better choice. The predicted performance of the commonly used time series model, WT-ARIMA model and LSTM model was tested based on the energy consumption data for one year. The results show that the model proposed in this paper has a 20% accuracy improvement over the ARIMA model and can reduce data requirement with good forecasting accuracy compared with LSTM-h.

1. Introduction
Growing energy demand has caused worldwide concern about energy efficiency and building energy efficiency plays a vital role in global sustainability. On the one hand, it is because of the high energy density of buildings and has a large potential for energy saving. In general, 80-90% of the total building life cycle energy consumption is used during building operations, where approximately 20% can be saved by applying appropriate controls or timely fault detection on management systems [1]. On the other hand, automation systems and energy management systems are increasingly used in buildings, and their building operations information is available. This abundant building-level data contributes to data-driven analysis of building load. At the same time, the rapid development of analytical tools increases the effectiveness of data-driven models.

Building energy forecasting is the foundation of many building energy management tasks. Most existing load forecasting models use non-DL (deep learning) techniques ranging from the simple multiple linear regression (MLR), autoregressive integral moving average (ARIMA), support vector regression (SVR) to the complex artificial neural network (ANN) [2]. A hybrid prediction model that combines the capabilities of these non-DL technologies is also discussed. In addition to load forecasting, these technologies are also widely used to predict solar photovoltaic power generation, wind power generation and electricity prices.

Researchers now focus on developing new predictive models based on existing models and adapted to the actual engineering goals [3]. A crude oil price prediction model using prior knowledge based on long short-term memory (LSTM) deep learning method has been proposed. Considering the...
equivalence of each historical data, a very creative algorithm is proposed, that is data transmission with prior knowledge, providing more usability data expansion methods (three kind of data type) [4]. Additionally, a weather classification model based on generating confrontation networks and convolutional networks has been established for short-term photovoltaic power generation forecasting. The enhancement of the solar irradiance data set is achieved by the proposed model. By using an augmented data set, classification performance can be improved to achieve more accurate weather status recognition [5]. Deep learning and traditional time series techniques are both used to perform building-level load forecasting. The model has ability to handle high levels of uncertainty in building loads. The multi-step method of convolutional neural networks provides the highest accuracy and high computational efficiency. Based on the RNN model, the LSTM model with non-fixed self-loop weights has a strong processing power for chronological data. Its special and local structure is that each ordinary node in the hidden layer is replaced by a memory unit. LSTM and its variants with long-term memory advantages have increasingly strong fitting capabilities [2].

The idea of many prediction methods is to construct a predictive model of demand by combining weather data with historical data. However, the first problem with this prediction method is that there may be a time delay between the input parameters and the predicted parameter [6]. In the case where the time interval of the prediction result is long, the target curve may not be well fitted, and the machine learning method may learn the law from the data, but the result is not satisfactory. Secondly, in order to make full use of data modeling, it is necessary to measure outdoor temperature, humidity, wind speed, radiation and other indicators in the target building. Indoor temperature and unit operating parameters may also be required. But in practice, detailed and complete data is usually not available. What's more, the uncertainty of the meteorological data on which the predictions are based may make detailed modeling less valuable and difficult to find the cause of the error. In this case, a model with fewer parameters can provide predictive results as good as a complex model, taking less time and computational cost.

The model proposed in this paper assumes that the energy demand of a building has a certain period, which can be fully described as different parts. In the second section, the method of splitting different period data is researched. The effective decomposition can be used to better identify the energy consumption requirements at different time resolutions and this method is easier to learn and predict by the prediction algorithm. The third section will introduce the basis and the number of decompositions. The prediction algorithm is described in Section 4 and gives some results. Finally, Section 5 gives some conclusions.

2. The decomposition method
Using classical time series decomposition can overcome many of the limitations of basic data analysis techniques to identify more specific operational characteristics of a building and point to opportunities for energy savings. Classical time series decomposition techniques have been used in a variety of industries, including economics, weather, energy, and other time-varying data, but not much in the field of architecture. This section will describe the systematic methodologies in detail. The FFT of the time-frequency analysis of the pattern is described in the second subsection. The third section introduces the method of wavelet transform (WT) decomposition. Finally, the WT-ARIMA model used in the article is summarized.

2.1. Seasonal-Trend decomposition procedure based on Loess (STL)
STL is a common algorithm in time series decomposition [7]. Based on LOESS, the data Yv at a certain moment is decomposed into a trend component, a seasonal component and a remainder component:

\[ Y_v = T_v + S_v + R_v \quad v = 1, \ldots, N \quad (1) \]

The STL is divided into an inner loop and an outer loop, wherein the inner loop mainly performs trend fitting and calculation of periodic components. The sample points at the same position in each cycle form a subseries. The inner loop is mainly divided into the following six steps:
Step 1: Detrending, subtracting the trend component of the previous round of results, $Y_v - T^{(k)}_v$.

Step 2: Cycle-subseries smoothing, use LOESS to regress each sub-sequence, and extend each cycle forward and backward; the smoothing result constitutes the temporary seasonal series, $C^{(k+1)}_v$, $v = -n(p) + 1, \cdots, N + n(p)$.

Step 3: Low-Pass Filtering of the periodic subsequence $C^{(k+1)}_v$. The resulting sequence of the previous step is sequentially subjected to a moving average, and then LOESS is returned to obtain a result sequence, $L^{(k+1)}_v$, $v = 1, \cdots, N$.

Step 4: Detrending of Smoothed Cycle-subseries, $S^{(k+1)}_v = C^{(k+1)}_v - L^{(k+1)}_v$.

Step 5: Subtracting the periodic component, $Y_v - S^{(k+1)}_v$.

Step 6: Trend Smoothing, LOESS regression for the sequence after the removal period, to get the trend component, $T^{(k+1)}_v$.

The outer loop is mainly used to adjust the robustness weight.

2.2. Fast Fourier transformation (FFT)

FFT is a fast algorithm of discrete Fourier transform. It uses periodicity and symmetry to improve the discrete Fourier transform (DFT) algorithm, which greatly reduces the amount of operation [8]. FFT can transform the sequence data into the frequency domain. Some data are difficult to find features in the time domain, but if they are transformed into the frequency domain, the periodic performance in the frequency domain is more concentrated, which is good for data processing and easy to find features.

The Fourier transform of the aperiodic continuous time series $x(t)$ can be expressed as:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt$$

(2)

The above formula calculates the continuous spectrum of the time series $X(t)$. However, what is available in an actual building system is the discrete sample value $X(nT)$ of the continuous time series $X(t)$. It is therefore necessary to use the discrete signal $X(nT)$ to calculate the spectrum of the signal $X(t)$.

The DFT of the finite-length discrete signal $X(k)$ is defined as:

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{kn}$$

(3)

$$k = 0, 1, \cdots, N - 1; W_N = e^{-j2\pi/N}$$

Each item is divided into two parts according to the parity of the subscript. The principle of the FFT algorithm is to achieve large-scale transformation through many small and easier transformations, which reduces the computational requirements and improves the speed of operation.

2.3. Wavelet transform (WT)

The wavelet transform replaces the basis of the Fourier transform—changing the infinitely long trigonometric base to a finitely long decaying wavelet base. Wavelet transform is also a signal decomposition idea: the signal is decomposed into a superposition of frequency band signals. Therefore, the performance of wavelet transform is closely related to the selection of wavelet basis. Different wavelet functions are suitable for different target sequence data.

The wavelet transform is defined as:

$$Wf(a, b) = \int f(t)\psi^*_a(u)(t)dt$$

(4)

The kernel function of the transformation is:
A set of wavelet basis functions are generated from basic wavelets by scale factors and translation factors. The kernel function is a wavelet function that is generated by the wavelet generating function and depends on the parameter \((a, b)\). Where: \(a\) is called a scale factor and \(b\) is a translation factor.

The scale corresponds to the frequency (inverse ratio) and the amount of translation corresponds to time. The scaling factor affects different scaling factors to generate different frequency components. The translation factor enables wavelets to implement traversal analysis along the time axis of the sequence data. This method will not only capture the frequency, but also the time. The wavelet transform decomposes the data into approximate components and detail components. The method can observe the data at any scale if the scale of the wavelet function is appropriate.

2.4. Overview of the WT-ARIMA approach

In this paper, the proposed WT-ARIMA method is shown in Figure 1.

![Figure 1. Flow chart of the WT-ARIMA method.](image)

The forecasting process can be summarized as follows:

Step1: Collect and pre-process daily energy consumption data, hourly data.

Step2: Data decomposition. First, the data is subjected to frequency domain transformation, and the frequency domain map is observed to obtain its main period. Second, the data is decomposed to several components by the WT.

Step3: Perform an ARIMA model for each component of the energy consumption.

Step4: Combine the output of the converted prediction result \(P, P = \sum_{i=1}^{k} P_i\).

3. Modelling the time series component

3.1. Data set

The data used in this study came from the Qingdao Building Digital Detection Platform. This time, the hourly energy consumption record of the entire building was used, and the time range was from January 1, 2014 to December 31, 2014.

The research object is building L. As an office building with additional centralized heating (unrecorded energy consumption) in winter, its daily energy consumption data record is shown in Figure 2. It can be seen from the time distribution of energy consumption that the data is nonlinear and non-stationary, with obvious periodic characteristics (years), but the periodic characteristics in days
are not significant, and there is no obvious data trend. The average, maximum, and quantile of the data can be seen from Table 1. The data was divided into two groups: a train set of 334 days (11 months of historical data) and a test set of 31 days (1 month of forecast data).

### Table 1. The describe of data set of building L.  

| Building | Total energy (kWh/day) |
|----------|------------------------|
| L        | Mean: 5774.632, Std: 1473.834, Min: 3673.000, 25%: 4638.000, 50%: 5421.200, 75%: 6209.999, Max: 10198.899 |

![Figure 2. The daily load of building L.](image)

#### 3.2. Evaluation Criteria

To evaluate the performance of the prediction method, three criteria, mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) were used. These standards can be expressed as follows:

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |Q_i - P_i| 
\]  

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|Q_i - P_i|}{Q_i} \times 100\% 
\]  

\[
RMSE = \left( \frac{1}{n} \sum_{i=1}^{n} (Q_i - P_i)^2 \right)^{1/2} 
\]

Q represents the observed energy consumption value and P represents the predicted energy consumption value. These three standards show the closeness of the prediction results from the true values from different aspects, the smaller the better.

#### 4. Results and discussion

##### 4.1. Data preprocessing result

It can be seen in Figure 3 that the Fourier transform results for the hourly energy consumption of building L. The result after the FFT is a complex number of N points. The horizontal abscissa of each point corresponds to the frequency point. The modulus value of the point is the amplitude characteristic (energy consumption) at that frequency value. There are many frequency points with peaks appearing. Considering that when the frequency is converted to the period, the adjacent frequency is not much different. This calculation takes the larger and concentrated point to calculate period of the peak in the FFT result. According to the FFT results, it can be found that the original data has 3 distinct periods, T1=365 days, amplitude is 45, T2=7 days, amplitude is 32, T3=1 day, amplitude is 70.

In this study, the WT method is used, in which the decomposition order is 2nd order, and the wavelet set used for decomposition is Daubechies extremal phase wavelets (db4). The order of the
decomposition corresponds to the number of periods in the FFT result. Figure 4 shows the decomposed wavelet coefficients. Wavelet decomposition can analyze data at different scales (different periods).

4.2. Forecast energy consumption by WT-ARIMA
Difference disposal is applied to raw data according to the identified period, and the model optimal parameters are selected using autocorrelation function (ACF) and partial autocorrelation function (PACF). Table 2 identifies the input and output structure of the data set. The observed values and fitted values of the model for the building L are shown in Figure 5.
Figure 5. Model fitting result by WT-ARIMA of the building L (train: blue, fit: red).

Table 2. Data requirements for WT-ARIMA models.

| Case     | Data set  | Input | Process | Output |
|----------|-----------|-------|---------|--------|
|          | Decomposition part | Number of parameters |         |        |
| Building L | A2        | 334, p=3, q=2 | 8       | 31     |
|          | D2        | 334, p=4, q=0 | 8       | 31     |
|          | D1        | 334, p=4, q=1 | 16      | 31     |

The effect of using different wavelet forms on the accuracy of the prediction results is shown in Table 3, where the Daubechies series wavelets have the best accuracy.

Table 3. Prediction results of different wavelet.

| Wavelet | MAE (kWh) | MAPE (%) | RMSE (kWh) |
|---------|-----------|----------|------------|
| db4     | 521.853   | 9.814    | 684.630    |
| db5     | 586.189   | 10.063   | 724.306    |
| coif4   | 587.623   | 10.549   | 719.039    |
| sym4    | 672.912   | 11.606   | 807.229    |
| dme4    | 783.727   | 13.692   | 906.532    |

The overall final prediction results are shown in Figure 5. The predicted value and the observed value have a good fitting effect. The predicted overall trend is almost consistent with the observed values, and the local troughs and peaks are well predicted compared with the traditional ARIMA model, indicating that the feature recognition effect after wavelet analysis is significant, as shown in Figure 6. This model improved the ARIMA model with 26.29%, 25.98% and 23.55% reductions in the MAE, MAPE and RMSE, respectively. Simple prediction model proposed in this paper has been proved to have good predictive ability. At a certain stationary stage, the WT-ARIMA model is consistent with the observed data, but the error will become larger and larger over time which is a limitation ARIMA model (only predict short-term goal). The result is shown in Table 3.

4.3. Result comparison

The choice of the prediction model has a significant impact on the prediction results, and the requirements of the parameters are different for different models. To demonstrate the availability of a simple model, the model is compared with other methods, including the general time series prediction model Holt's Winter model, SARIMA which take seasonal trend into consideration by STL method,
and machine learning methods Long Short-Term Memory (LSTM). The model parameters are shown in Table 4.

Table 4. Comparison model and its structural parameters.

| Model     | Structure                                                                 |
|-----------|---------------------------------------------------------------------------|
| Holt's Winter | number of input: 334 × 1, number of variables: 3                           |
| SARIMA    | number of input: 334 × 1, number of variables: 3, seasonal_order: 4       |
| LSTM      | 334, Input layer: 8, output layer: 1, hidden neurons: 300, Total params: 371101 |

The input parameters of the time series method only include date data and total energy consumption. In addition to this, machine learning methods also require weather parameters, including mean temperature, max temperature, min temperature, mean humidity, max humidity, min humidity, dew point temperature, wind speed in this case.

Figure 6. The prediction results for the different models.

The results of the model basically follow the trend changes observed in the data set. However, the model structure is different, and the fitting effect is quite different. The Holt's Winter model is not sensitive to trend changes, and the overall forecast results vary smoothly. The linear change in the trend is unrecognizable, and only periodic changes are predicted (same data repetition). The prediction results for data with large fluctuations are not ideal and are easy to deviate. The trend changes identified by the SARIMA model are limited to local data, and the trend is greatly affected by the data of the previous period. It is not possible to identify large periodic changes, and the increase or decrease in the trend will result in a serious deviation of the data. Compared to other models, the LSTM model requires the most complex parameters and the longest training time. The LSTM model can learn complex nonlinear relationships from features, but this requires more historical data to train. In this example, when the training data does not cover the predicted month (the effect of heating and holidays in the target time period is not learned), the result is not satisfactory. The LSTM model with holidays and period information (LSTM-h) can achieve better accuracy but it takes cost to label the data. The MAE, MAPE and RMSE of the different model are shown in Table 5.
Table 5. Prediction results of different model.

| Model      | MAE (kWh) | MAPE (%) | RMSE (kWh) |
|------------|-----------|----------|------------|
| WT-ARIMA   | 521.853   | 9.814    | 684.630    |
| Holt’s Winter | 531.17   | 9.924    | 686.606    |
| SARIMA     | 786.982   | 13.750   | 942.988    |
| LSTM       | 912.598   | 15.352   | 964.148    |
| LSTM-h     | 482.792   | 8.250    | 541.156    |

5. Conclusion
Building energy consumption prediction is a very important task, and the accuracy of the forecast and the workload are very important for the evaluation of the model. This paper proposes a time-series prediction model that artificially decompose energy consumption data into a WT-ARIMA model. The prediction results and parameter requirements indicate that the model can achieve better prediction performance without increasing the requirements of the data and without using complex models.

During the data decomposition phase, the daily time series is split into different parts, improving the ability to express information. During the training phase, the model can better identify the characteristics of the time series data. By decomposing and reintegration, the characteristics of the sequence are clearly described, the model has better nonlinear fitting ability. Compared with the basic ARIMA model, this model improves the prediction accuracy by 20%. The model structure is not significantly complicated, the input parameters of the model are not increased, and the workload is almost the same as the original ARIMA model. The evaluation index of the model is better, and the predictive ability of the model can be accepted. Both this model and the well-trained LSTM-h model can achieve good prediction accuracy, but the WT-ARIMA model can significantly reduce the demand for data and computing time.

References
[1] Fan C, Wang J, Gang W and Li S 2019 Assessment of deep recurrent neural network-based strategies for short-term building energy predictions Applied Energy 236 700-10
[2] Cai M, Pipattanasomporn M and Rahman S 2019 Day-ahead building-level load forecasts using deep learning vs traditional time-series techniques Applied Energy 236 1078-88
[3] Wang Y et al 2018 Determination of polybrominated diphenyl ethers and novel brominated flame retardants in human serum by gas chromatography-atmospheric pressure chemical ionization-tandem mass spectrometry. J Chromatogr B Analyst Technol Biomed Life Sci 1099 64-72
[4] Cen Z, Wang J 2019 Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer Energy 169 160-71
[5] Wang F et al 2019 Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting Energy Conversion and Management 181 443-62
[6] Guo Y et al 2018 Machine learning-based thermal response time ahead energy demand prediction for building heating systems Applied Energy 221 16-27
[7] Pickering EM, Hossain MA, French RH and Abramson AR 2018 Building electricity consumption: Data analytics of building operations with classical time series decomposition and case based subsetting Energy and Buildings 177 184-96
[8] Li C, Tao Y, Ao W, Yang S and Bai Y 2018 Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition Energy 165 1220-7

Acknowledgements
This research has been supported by the national key R & D project (No.2016YF0700104-05)