System Identification Term Project

2102531 Semester 1/64 Year 2021

Electrical load forecasting by SARIMAX model

Phing Lim ID 6372084021
Advisor: Jitkomut Songsiri
Department of Electrical Engineering, Faculty of Engineering
Chulalongkorn University

Contents

1 Introduction

The energy crisis is the biggest problem in the world today. There are many ideas on how to find new renewable energy sources to provide a sustainable energy supply... Another interesting idea is energy management. In this idea, demand side management is a popular technique to use electrical energy according to the demand of important consumers. However, for this it is also necessary to understand the performance of the energy consumption of the load in the grid or in the building. Predicting the energy load can help to make a good decision to improve the efficiency of energy management technique.

According to [HAG+21], both statistical and time series approaches as well as machine learning and other artificial intelligence approaches are the most popular prediction models seen in many previous works. Starting from the simplest linear regression model [LWvS18], multi-linear regression (MLR) [MMC19] through autoregressive integrated moving average (ARIMA) [FL16] [NMC+18], support vector regression (SVR) [CCB13] [JZM+16] to the most advanced artificial neural networks (ANN) [MMC19]. Since many models exist, many authors propose to compare different statistical and machine learning methods, such as MLR, SVR and ANN, which were studied by the authors for different household aggregations [HWVA13]. The result shows that SVR performs best at the aggregate level (by 32 households), while MLR performs best at the household level. Some hybrid modeling approaches are also mentioned in the review paper [HAG+21]. In this part, it is mentioned that ARIMA and ANN models are common components in model combinations. For example, in [BSL14], the authors use a hybrid model that uses ARIMA and ANN to produce point forecasts for total energy and peak load at the low-voltage transformer level for 128 residential customers. They conclude that ANN is more likely to account for small fluctuations, while ARIMAX is better suited for modeling large peak loads. This was the reason for the authors to propose a hybrid model where the load was first forecast using ANNs. When the forecasted demand was above a threshold, ARIMAX was used as the final forecasting model..

In [CPR19], the author mentions that ARIMAX is better at capturing temporal dependence compared to MLR and has better interpretive ability compared to SVR or ANN. This shows the advantage of ARIMAX model compared to other time series models. Since electric load prediction is a non-stationary problem ARIMA model is the best candidates. We know that ARIMA is an extension of ARMA models for non-stationary time series, which can be made stationary by taking a difference (of a certain order) from the original time series. Moreover, as can be seen in [CPR19] and [EF18], the outdoor temperature profile is integrated as an input variable, which has a dominant impact on the electric load, rather than the other variables. These reasons lead to the ARIMAX model being superior to another model.

Project description The aim of this project is to obtain a good time series model for hourly prediction of electrical load for each floor of Chamchuri5 building. The SARIMAX (Seasonal Auto Regressive Integral Moving Average with Exogenous Input) model is used in our work to represent the model prediction. The exogenous input here is outdoor temperature. The data for the electricity consumption of each floor come from CUBEM, as you can see in [PCS+20] while the outdoor temperature data is got from webist of `hptt://meteostate.net` in daily.

2 System modelling

In this project, a seasonal ARIMA model with exogenous input (outdoor temperature) is used as the model for load forecasting for each floor of the building. Therefore, the model is presented here as a general equation for i^{th} floor, $i = 1, ..., 7$ (the number of floors in Chamchuri5 building). We know that SARIMAX is given by the time series patterns within the series and captures the linear covariance between the target variable and the exogenous variables.

ARIMAX $(p, d, q) \times (P, D, Q)S$ represents the standard seasonal ARIMAX model, where $p =$ non-seasonal autoregressive (AR) order, $d =$ non-seasonal differencing, $q =$ non-seasonal moving average (MA) order, $P =$ seasonal AR order, $D =$ seasonal differencing, $Q =$ seasonal MA order, $S =$ time span of the repeating seasonal pattern.This pattern we will discuss more in the methodology part by seeing the time-series graph of our data.

In case of our model, let $y_i(t)$ as the i^{th} floor load value at time t, $e_i(t)$ is white noise of floor i^{th}, and $x(t)$ denotes as outdoor temperature covariate at time t. The ARIMAX model of i^{th} floor can be mathematically expressed by equation below where $\nabla_{iS}y_i(t) = y_i(t) - y_i(t-S)$ and $\nabla_i y_i(t) = y_i(t) - y_i(t-1)$ represent differencing operations. While L is known as the backshit operator.

$$A_i(L^S)a_i(L)\nabla_{iS}^D\nabla_i^d y_i(t) = B_i x(t) + C_i(L^S)c_i(L)e_i(t), \quad i = 1, 2, ..., 7$$

where:

$$\begin{aligned} A_i(L^S) &= 1 - A_{i1}L^S - ... - A_{iP}L^{PS} \\ a_i(L) &= 1 - a_{i1}L - ... - a_{iP}L^p \\ C_i(L^S) &= 1 + C_{i1}L^S + ... + C_{iQ}L^{QS} \\ c_i(L) &= 1 + c_{i1}L + ... + c_{iq}L^q \end{aligned}$$

3 Methodology

According to [HAG+21], the most important for developing an appropriate forecasting model is to select the correct input features. As mentioned in the model chosen, the exogenous variable is the outdoor temperature. In this part, we will discuss more the way to obtain it through an iterative process of plotting, interpreting, and testing.

Figure 3 presents the load-temperature scatter plot of all floors for the period July 01, 2018, until December 31, 2019. The figure shows that the peak of load demand of each floor exists while the temperature is high (in the summer season). This demonstrates that electricity is used for cooling in summer. Moreover, base on the correlation between load consumption of each floor and outdoor temperature are not than 0.34. The linear effect from temperature can be used as exogenous input correctly to our electricity model. However, we should also denote that some of zero ac load see in second floor until the seventh floor scatter graph which indicated that electrical usage in the build each floor depending on the week day and weekend or holiday. We also can see the effect of weekly seasonality by see in the figure 2.

As mentioned in the model section, SARIMAX model is proposed to be forecast model in our project. From the paper [FL16] state the general scheme for ARIMA model includes the following four steps starting from the model structure identification, The second step is to identify the order of the ARMA model by Autocorrelation function (ACF) and partial autocorrelation function (PACF). The parameters of the model are estimated by a maximum likelihood (ML) function. Next step is testing on the estimated model residuals to find the goodness of fit. In the last step, The estimated model can be used to conduct forecast which is obvious that forecasts are less accurate as the forecasting horizon gets larger.

When evaluating which tentative model best fits the data, the Akakike Information Criterion (AIC) and Bayesian Information Criterion are a measure to compare them. The AIC can reward models for a good fit and penalize model for complexity while BIC is related to the AIC but has a larger penalty term then in the AIC. Both AIC and BIC apply a likelihood function to select the best fitted model. They stand for a trade-off between 'fit' measured by log likelihood value and 'parsimony' as measured by the number of free parameters. The target is to choose the model orders that result is in minimum values of AIC and BIC. The most appropriate model for the data set of each floor are found as ARIMAX$(3, 1, 2)$ model seasonally Integrated with Seasonal AR(12) and MA(6), ARIMAX$(3, 1, 2)$ model seasonally Integrated with Seasonal AR(7), ARIMAX$(3, 1, 0)$ model seasonally Integrated with Seasonal AR(7), ARIMAX$(2, 1, 1)$ model seasonally Integrated, ARIMAX$(3, 1, 1)$ model seasonally Integrated, ARIMAX$(2, 1, 0)$ model seasonally Integrated,and ARIMAX$(4, 1, 0)$ model seasonally Integrated, respectively from first to top floor. The estimated significant parameters are illustrated in Table 1.

In this study, we compare the accuracy of each models by out-of-sample test which means the data used in model fitting are different from those used in forecasting evaluation. For our case, the period from July 1, 2018 to July 21, 2019 is used for estimation purposes, and the data from August 1, 2019

Table 1: SARIMAX model estimation results of each floor

Variable	Value	t-Statistic
1st Floor		
Constant	1.0699	-1.1874
AR{1}	0.73811	8.6956
AR{2}	-0.33345	-4.8357
AR{3}	-0.35191	-6.1017
SAR{6}	-0.28035	-5.1566
SAR{12}	-0.11218	-1.9794
MA{1}	-1.2603	-15.0261
MA{2}	0.65779	8.3865
SMA{6}	-0.94654	-40.8751
Beta(TempOut)	0.034944	1.1483
AIC	2909.8355	
BIC	2953.0023	
2nd Floor		
Constant	-0.12515	-0.19714
AR{1}	-0.71558	-15.7629
AR{2}	0.24502	3.9467
AR{3}	-0.0394	-0.71914
SAR{7}	-0.48235	-14.9103
MA{2}	-1	-75.5377
Beta(TempOut)	0.0040809	0.19086
AIC	2659.1407	
BIC	2690.5987	
3rd Floor		
Constant	1.6566	0.099526
AR{1}	-0.4795	-5.649
AR{2}	-0.42935	-5.6573
AR{3}	-0.27578	-3.8988
SAR{7}	0.55415	7.6595
Beta(TempOut)	-0.055981	-0.10485
AIC	1053.422	
BIC	1074.0135	

Variable	Value	t-Statistic
4th Floor		
Constant	0.24343	1.1687
AR{1}	0.035362	0.88603
AR{2}	-0.072149	-1.2021
MA{1}	-1	-47.2507
Beta(TempOut)	-0.0081339	-1.1513
AIC	2447.7043	
BIC	2470.6248	
5th Floor		
Constant	0.36722	1.4611
AR{1}	0.056385	1.3992
AR{2}	-0.095576	-1.2472
AR{3}	0.078836	1.9443
MA{1}	-1	-103.5754
Beta(TempOut)	-0.012247	-1.4445
AIC	2769.0419	
BIC	2796.6964	
6th Floor		
Constant	-11.0799	-1.7741
AR{1}	-0.24516	-6.3187
AR{2}	-0.10077	-1.9697
Beta(TempOut)	0.3757	1.7696
AIC	2760.3117	
BIC	2780.0909	
7th Floor		
Constant	-4.5209	-0.363
AR{1}	-0.66046	-14.1837
AR{2}	-0.63307	-11.5899
AR{3}	-0.3664	-7.6187
AR{4}	-0.22095	-4.9785
Beta(TempOut)	0.15185	0.36776
AIC	3113.0607	
BIC	3140.7152	

are left for forecast evaluation for all floor except the 7th floor that estimate set is chosen from January 1, 2019 to May 31, 2019 and other is used for forecast evaluation.

Serveral measurement statistic can be used to examine the forecast accuracy of different models. Rot mean squared error ($RMSE$), mean absolute percentage error ($MAPE$), and Thil's inequality coefficient (TIC) are use very often to evaluate the performance of the forecasting model. The above mentioned statistical quantities are computed as below:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_t - y_t)^2} \tag{1}$$

$$MAPE = \left(\frac{1}{N}\sum_{t=1}^{N}\left(\frac{|\hat{y}_t - y_t|}{y_t}\right)\right).100\% \tag{2}$$

$$TIC = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{y}_t - y_t)^2}}{\sqrt{\frac{1}{N}\sum_{t=1}^{N} y_t^2} + \sqrt{\frac{1}{N}\sum_{t=1}^{N} \hat{y}_t^2}} \tag{3}$$

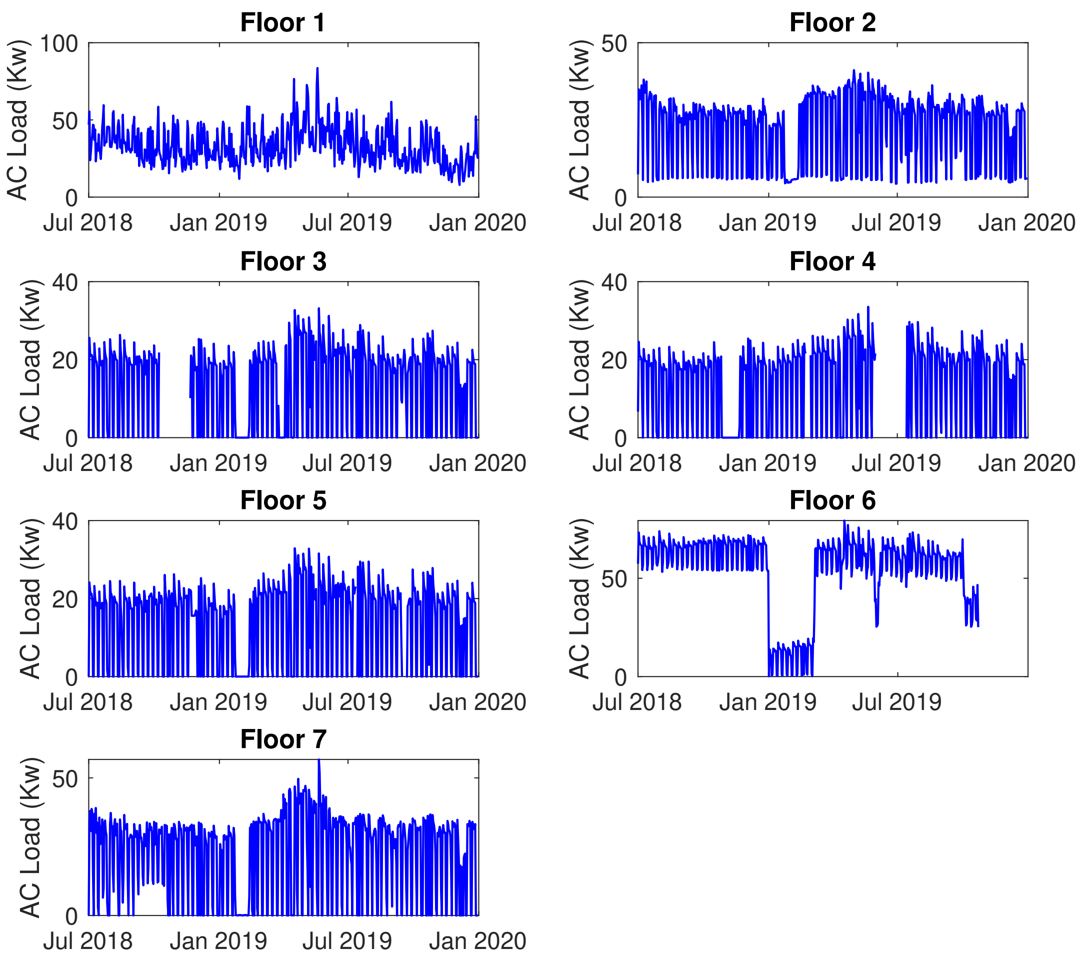


Figure 1: Daily electrical load plot of each floor

RMSE statistics depend on the scale of the scale of the variables. In such cases smaller errors indicate the better forecasting accuracy. MAPE and TIC are insensitive to the scale of the variables. Similarly, smaller MAPE and TIC indicate a better forecasting performance. TIC yields a number of values ranging from 0 to 1, where zero indicates a perfect fit of the forecasted values to the actual.

4 Data description

As previously stated, historical load consumption data for this project was derived from the CUBEMs data of the Chamuri5 building (see in [PCS+20]). The data was provided in a csv file that contained the electricity consumption of AC loads in (Kw) for each level (from $1st$ to $7th$ floor) for each minute. The data was converted to daily as shown in figure 1 because the goal of this project is to estimate load demand a day ahead of time.

From July 1, 2018 to December 31, 2019, the data for each level is shown in 1. According to the graph, the electrical load consumption in each level begins to rise in April, reaches a high in May, and then begins to decline before July. This is show the annual seasonality of data relate to season (weather effect) and to economic and social events during different months (e.g., public and national holidays). Figure 2 show weekly seasonality. The weekly pattern reflects the variation in load on weekdays versus weekends.

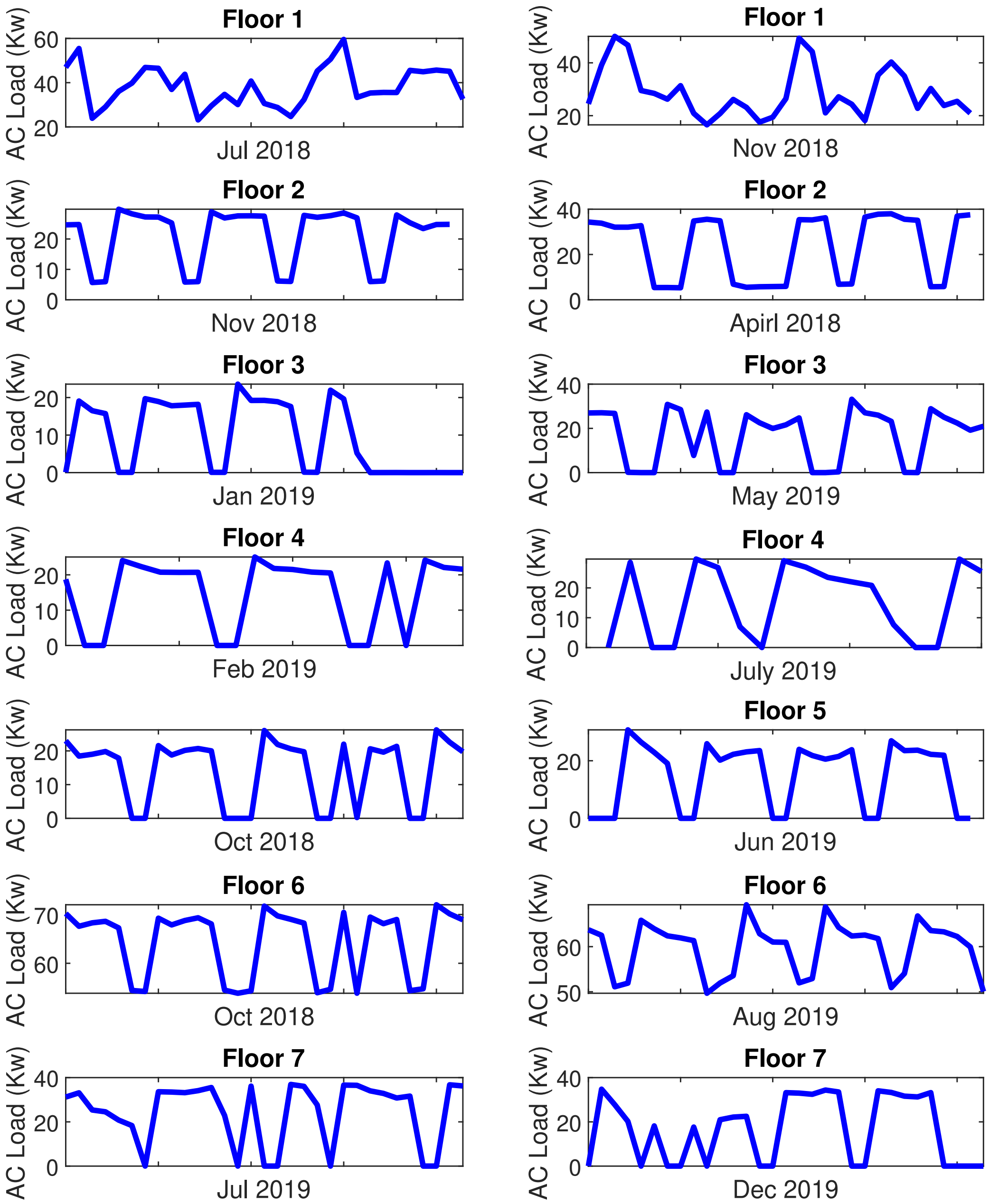


Figure 2: Daily electrical load plot of each floor for random Month

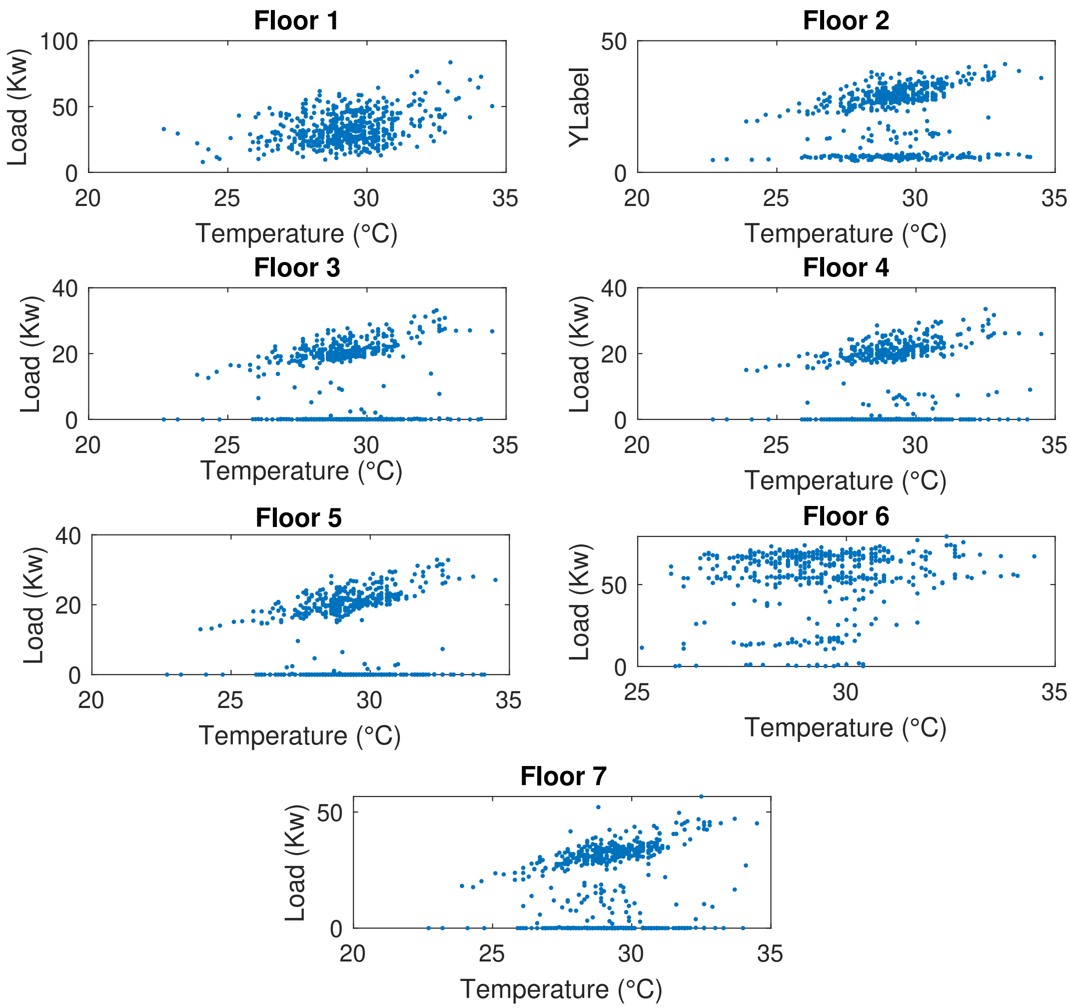


Figure 3: Daily electrical load plot of each floor

5 Experimental results

References

[BSL14] Christopher J Bennett, Rodney A Stewart, and Jun Wei Lu. Forecasting low voltage distribution network demand profiles using a pattern recognition based expert system. *Energy*, 67:200–212, 2014.

[CCB13] Ervin Ceperic, Vladimir Ceperic, and Adrijan Baric. A strategy for short-term load forecasting by support vector regression machines. *IEEE Transactions on Power Systems*, 28(4):4356–4364, 2013.

[CPR19] Mengmeng Cai, Manisa Pipattanasomporn, and Saifur Rahman. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Applied energy*, 236:1078–1088, 2019.

[EF18] Niematallah Elamin and Mototsugu Fukushige. Modeling and forecasting hourly electricity demand by sarimax with interactions. *Energy*, 165:257–268, 2018.

[FL16] Tingting Fang and Risto Lahdelma. Evaluation of a multiple linear regression model and sarima model in forecasting heat demand for district heating system. *Applied energy*, 179:544–552, 2016.

[HAG+21] Stephen Haben, Siddharth Arora, Georgios Giasemidis, Marcus Voss, and Danica Vukadinovic Greetham. Review of low-voltage load forecasting: Methods, applications, and recommendations. *arXiv preprint arXiv:2106.00006*, 2021.

[HWVA13] Samuel Humeau, Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer. Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households. In *2013 Sustainable internet and ICT for sustainability (SustainIT)*, pages 1–6. IEEE, 2013.

[JZM+16] Huaiguang Jiang, Yingchen Zhang, Eduard Muljadi, Jun Jason Zhang, and David Wenzhong Gao. A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization. *IEEE Transactions on Smart Grid*, 9(4):3341–3350, 2016.

[LWvS18] GBMA Litjens, E Worrell, and WGJHM van Sark. Assessment of forecasting methods on performance of photovoltaic-battery systems. *Applied Energy*, 221:358–373, 2018.

[MMC19] Nick MacMackin, Lindsay Miller, and Rupp Carriveau. Modeling and disaggregating hourly effects of weather on sectoral electricity demand. *Energy*, 188:115956, 2019.

[NMC+18] Gde Dharma Nugraha, Ardiansyah Musa, Jaiyoung Cho, Kishik Park, and Deokjai Choi. Lambda-based data processing architecture for two-level load forecasting in residential buildings. *Energies*, 11(4):772, 2018.

[PCS+20] Manisa Pipattanasomporn, Gopal Chitalia, Jitkomut Songsiri, Chaodit Aswakul, Wanchalerm Pora, Surapong Suwankawin, Kulyos Audomvongseree, and Naebboon Hoonchareon. Cu-bems, smart building electricity consumption and indoor environmental sensor datasets. *Scientific Data*, 7(1):1–14, 2020.

6 Appendices

6.1 Appendix A

Autocorrelation function (ACF) and Partial Autocorrelation function (PACF):

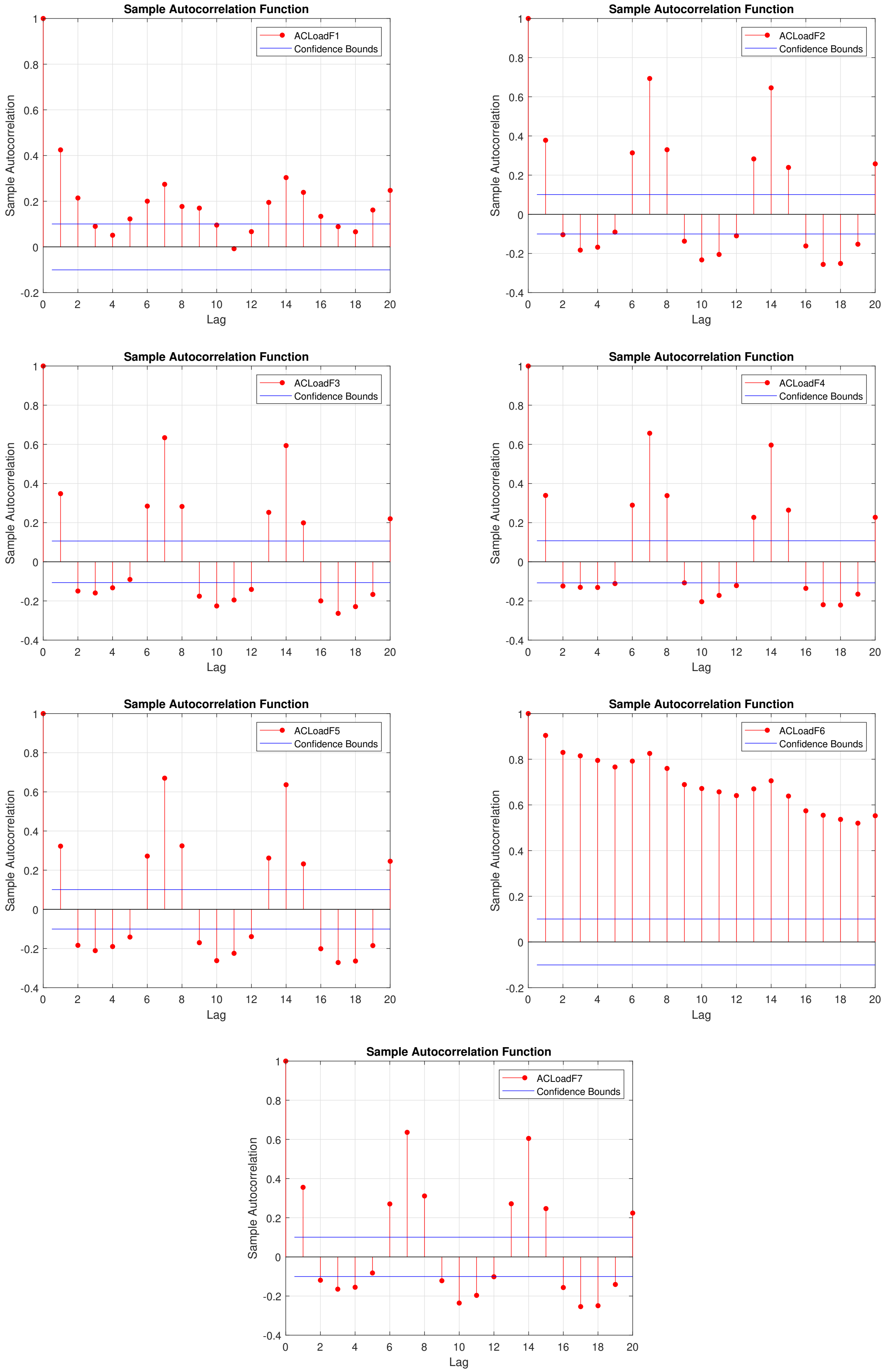


Figure 4: Autocorrelation function (ACF) of each floor

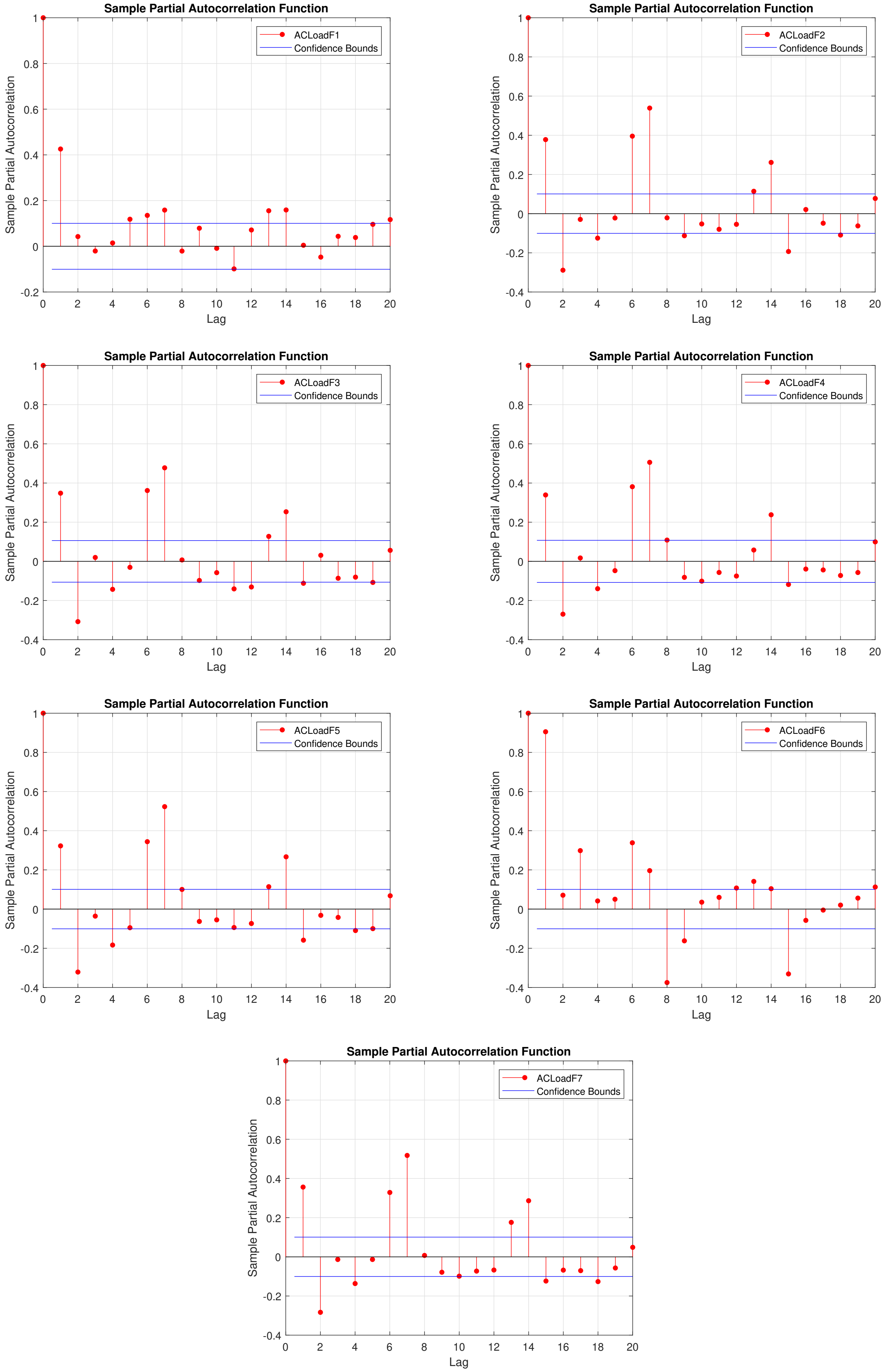
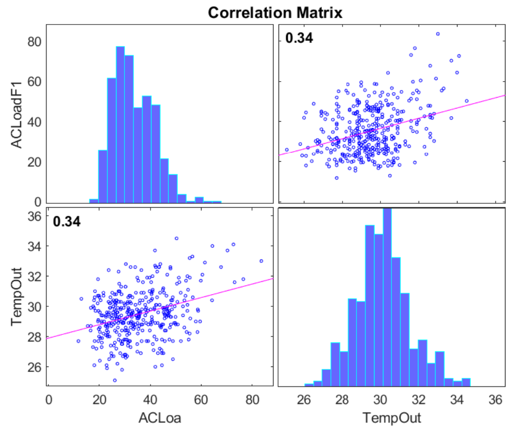
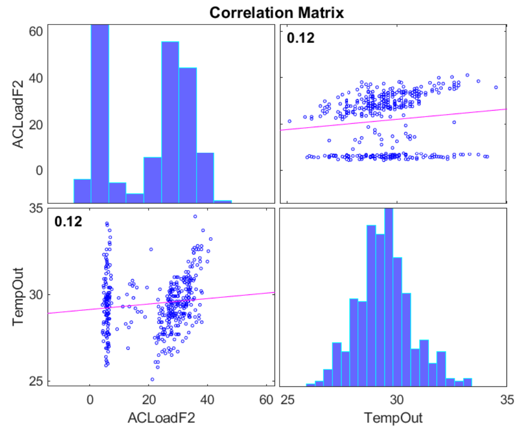


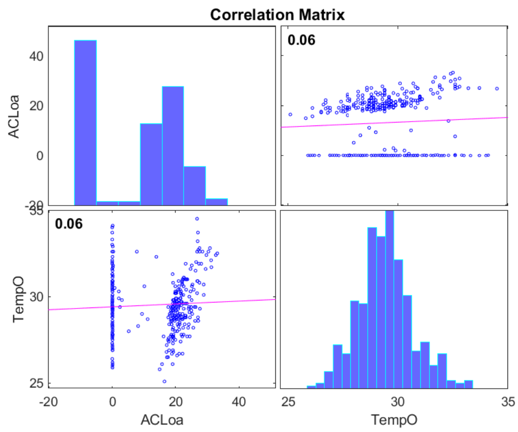
Figure 5: Partial Autocorrelation function (PACF) of each floor

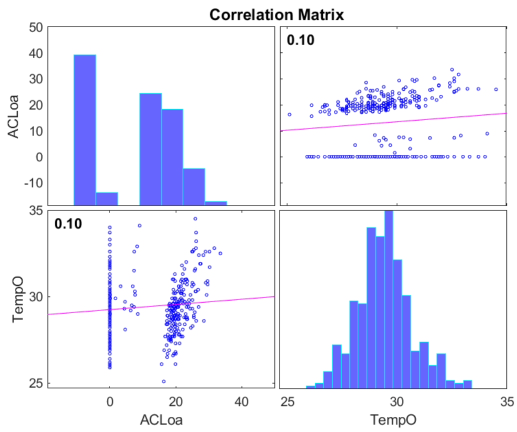
6.2 Appendix B

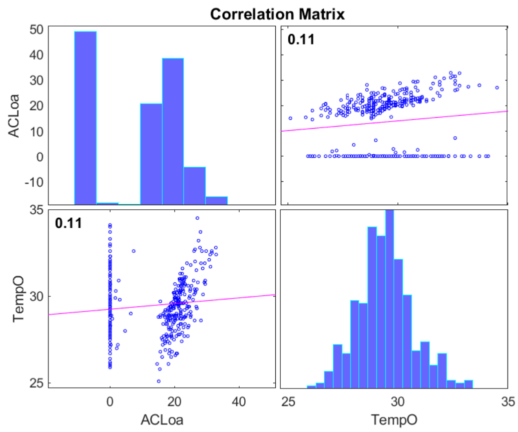
Matrix Correlation of AC load each floor and Outdoor Temperature:

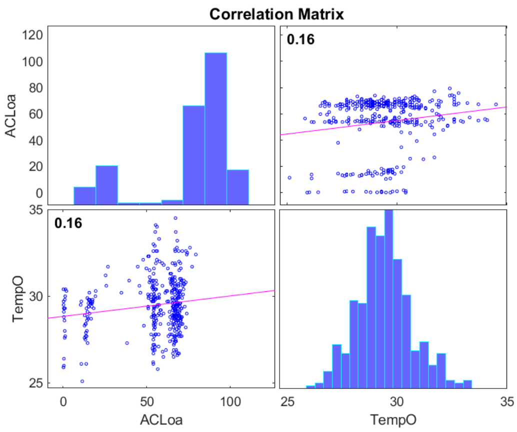

Correlation Matrix
ACLoadF1
TempOut
0.34
0.34
ACLoa
TempOut

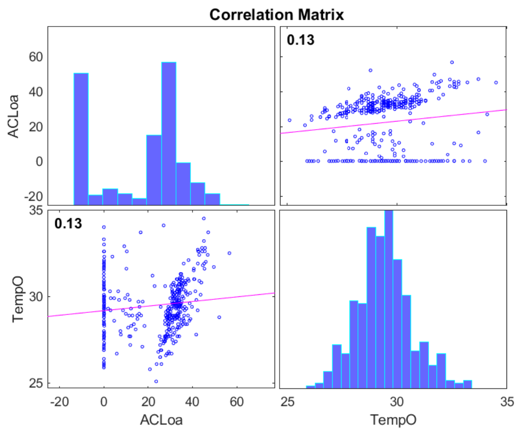

Correlation Matrix
ACLoadF2
TempOut
0.12
0.12
ACLoadF2
TempOut


Correlation Matrix
ACLoa
TempO
0.06
0.06
ACLoa
TempO


Correlation Matrix
ACLoa
TempO
0.10
0.10
ACLoa
TempO


Correlation Matrix
ACLoa
TempO
0.11
0.11
ACLoa
TempO


Correlation Matrix
ACLoa
TempO
0.16
0.16
ACLoa
TempO


Correlation Matrix
ACLoa
TempO
0.13
0.13
ACLoa
TempO