**Prediction of building energy consumption**

[Energy AI 김소희]

1. **Prediction models for forecasting building energy consumption** [1]

Prediction of building energy consumption is crucial for improved decision making towards reducing energy consumption and CO2 emissions, because it can assist in evaluating different building design alternatives and building operation strategies (in terms of their energy efficiency) and improving demand and supply management. However, building energy consumption prediction remains to be a challenging task due to the variety of factors that affect the consumption such as the physical properties of the building, the installed equipment, the outdoor weather conditions, and the energy-use behavior of the building occupants.

There are several popular methods used for forecasting building energy consumption. It can be categorized into two main approaches, engineering (physical) modelling approach and data-driven approach. Physical models (also known as white-box models) rely on thermodynamic rules for detailed energy modelling and analysis. Examples of building energy simulation software that utilize physical models include EnergyPlus, eQuest, and Ecotect. These types of software calculate building energy consumption based on detailed building and environmental parameters such as building construction details; operation schedules; HVAC design information; and climate, sky, and solar/shading information [4]. However, some of such detailed data may not be available to the users at the time of simulation. Failure to provide accurate input can result in poor prediction performance

On the other hand, data-driven building energy consumption prediction modelling does not perform such energy analysis or require such detailed data about the simulated building, and instead learns from historical/available data for prediction. They correlate the energy consumption or energy index with the influencing variables. Data-driven prediction has gained a lot of research attention in recent years, despite its possible limitations. Developing a data-driven model, typically, consists of four primary steps: data collection, data preprocessing, model training, and model testing. SVM, ANN, decision trees, and other statistical algorithms are the most commonly-used supervised machine learning algorithms for model training.

Despite the importance of data-driven approaches, data-driven energy consumption prediction has two main limitations. First, data-driven prediction models may not perform well outside of their training range. For example, a model that was trained by learning from a limited dataset (e.g., data collected from a small set of buildings) may not perform well outside of the training data (e.g., different types of buildings in terms of physical properties, operation strategies, weather conditions, occupant behavior, etc.). Second, data-driven prediction models are black-box models – their internals are not known. A black-box model may provide sufficient prediction accuracy but may be limited in providing a detailed understanding of the different parameters and its behavior in terms of energy consumption.

Hybrid or grey-box modelling approaches, on the other hand, offer a combination of physical and data-driven prediction models, thereby leveraging the advantages and minimizing the disadvantages of both approaches. In grey-box models, some internal parameter and equations are physically interpretable. Grey-box models may also show better performance compared to black-box and white-box models.

So, I will discuss about 2 paper which is each related to data-driven model (black-box model) [2] and a hybrid model (grey-box model) [3]. Chae et al. [2] proposes a short-term building energy usage forecasting model based on an Artificial Neural Network (ANN) model with Bayesian regularization algorithm and investigates how the network design parameters such as time delay, number of hidden neurons, and training data effect on the model capability and generality. Dong et al. [3] developed a hybrid model, which couples a data-driven model and a thermal network model, for predicting the total and non-AC energy consumptions of residential buildings and compared its prediction performance to ANN-, SVM-, LSSVM-, Gaussian mixture model (GMM), Gaussian process regression (GPR)-based models.

1. **Artificial neural network model for forecasting sub-hourly electricity usage in Commercial buildings** [2]
   1. **Introduction**

Forecasting electricity load, especially for commercial and industrial buildings has become one of the important topics recently, to be able to better manage energy usage. Although the forecast time horizon can range from minutes to years, the short-term load forecast (STLF), especially for a period shorter than a day, has been more of an interest in the perspective of buildings because the utility prices may change by seasonality, time-of-use in on/off peak period, and contract demand.

Previous studies on STLF for the sub-hourly electricity consumption of buildings are limited. Escriva et al. [4] proposed STLF model using ANNs, but this model requires an entire whole year’s dataset, and the performance may not be stable when the energy consumption pattern has large daily or annual variability. Therefore, it is useful to explore a model that can perform well under more general setting, in particular, not requiring a large amount of input data for forecasting electricity usage of buildings. For this need, they developed a short-term load forecasting model using data mining and machine learning technique while assuming limited availability of data. In particular, they investigate ANNs model to predict the energy consumption of a commercial building complex.

* 1. **Methodological approach**

 All data set for this study was obtained from a building management system (BMS) of a commercial office building complex, and the data are periodically pulled into a relational database (IBM DB2TM). The site consists of three office buildings in urban area, each of which has different number of floors; five in building 1 (BLDG1), four in building 2 (BLDG2), and two in building 3 (BLDG3). A total floor area of 15,224 m2 spreads over typical office area, small laboratories, cafeteria, parking garage, and small gymnasium (Fig. 1). Although the buildings are separated, they are all managed by one utility billing system.

Figure 1 Typical floor plans of the case study building complex.

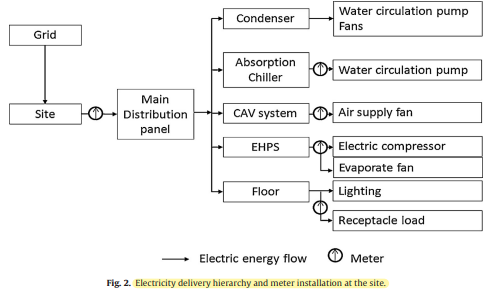
Two absorption chiller systems provide chilled water in summer and hot water in winter for a constant air volume (CAV) system for each floor of BLDG1 and fan coil units (FCU) of perimeter zones of each floor of BLDG2. All three buildings have electric heat pump (EHP) systems with multi-indoor units. EHP systems are supplementary system incorporated with CAV systems for BLDG1 but operate as the main air-conditioning system for BLDG2 and BLDG3. BMS system monitors operational conditions of both primary/secondary system and EHPs. The system also controls all secondary system operation, whilst EHPs are locally controlled by occupants’ indoor thermal demand. For the electricity usage monitoring, one main electric meter and several sub-meters are installed as illustrated in Fig. 2. The main meter measures electricity usage, both the instantaneous power in kW with a minute interval and aggregated electricity usage at every 15 minutes in kWh.

Figure 2 Electricity delivery hierarchy and meter installation at the site.

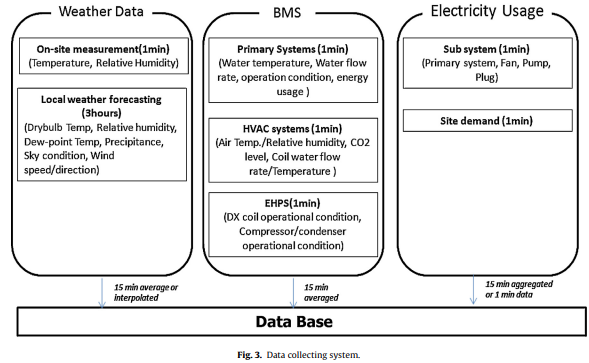
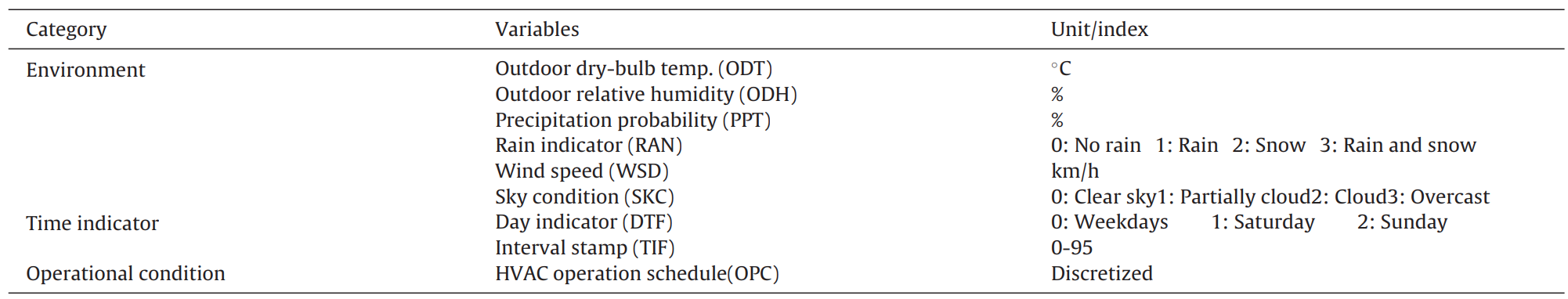
The relational database management system (RDBMS) has been used to collect and store environmental variables, BMS data, and electricity usage as illustrated in Figure 3. Although the data set has over 1000 data points with 20 different measurement types, some variables may not eventually be used for electricity usage forecast. Thus, after taking the data availability into account, the predictors are divided into three categories: environmental data, time indicator, and operational condition, as illustrated in Table 2.

Figure 3 Data collecting system.

Table 1 Potential predictor variables.



The data include outdoor air temperature, relative humidity, wind speed and direction, sky condition, and precipitation(강수량) type in every 3-hour interval for the next 72 hours. The forecast data is updated eight times a day.

The potential predictors are nine independent variables. However, those variables influence the electricity usage of the building in a different way. If some variables in the input data set are irrelevant to the output, it decreases model accuracy, stability, and effectiveness. Therefore, it is necessary to pre-screen the variables by identifying the important variables from the input data set.

Random forests algorithm [5] was used to assess the importance of variables by measuring the candidate parameters in terms of their impacts on the response of prediction. The permutation importance and node impurity were used to select variables in this study. According to permutation importance and node impurity, the operational condition was one of the most important factors. It shows that the operational condition of the secondary system is useful to capture the actual activity in the building such as occupancy and electricity consumptions of lighting systems and receptacles.

Five variables are ranked highly in their importance; the operational condition, time indicator, day type, outdoor dry-bulb temperature, and outdoor relative humidity. These five variables are selected as input attributes, together with previous electricity usages.

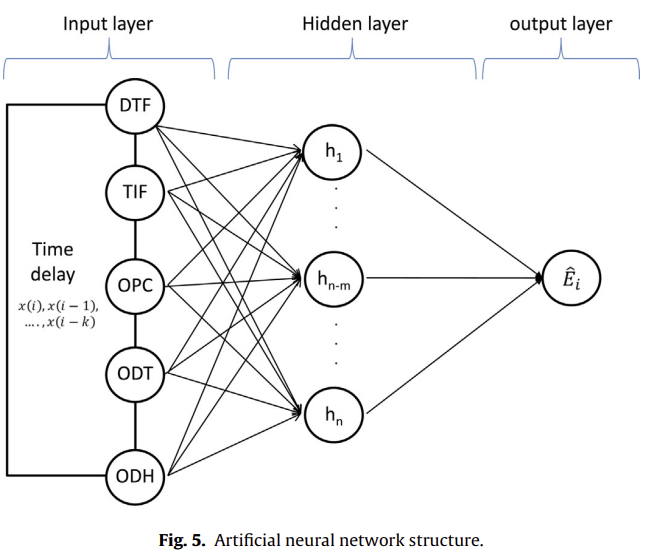
Although many machine-learning algorithms are available, the choice of the specific method to use is not trivial and depends strongly on the specific application and data availability and type. They examined nine different machine-learning algorithms and chose the one that performed best, the ANN. The ANN model in this study has a conventional multi-layered feedforward network using a backpropagation algorithm as illustrated in Figure 5. A Bayesian regularized neural network model with Levenberg–Marquart(LM) backpropagation algorithm is employed for the training process to improve the generalization of model. In this approach, the objective function includes both the conventional error function and the weight decay components or penalty term. The weights and biases in the model are assumed to be random variables with Gaussian distribution and the regularization parameters in the objective function can be optimized by using Bayesian rules.

Figure 4 Artificial neural network structure.

A test and validation procedure of the ANN model was conducted under several predictor conditions and data implementations with time delays during the training period. The 15-minute interval data set and highly ranked five predictor variables were collected from July 1st to July 31st, 2012. By removing data from one day, which has sensor and meter malfunction, 30 days, of which 22 are weekdays and 8 are weekend days, were used. Total of 2880 data points for six input variables, including the electric usage, were used to train the model. The network model selected the training input data set in a random, while three weekdays with a new data set (August 1–3, 2012), due to the weather forecasting time scope, were used to evaluate the out-of-sample testing.

* 1. **Results and discussions**

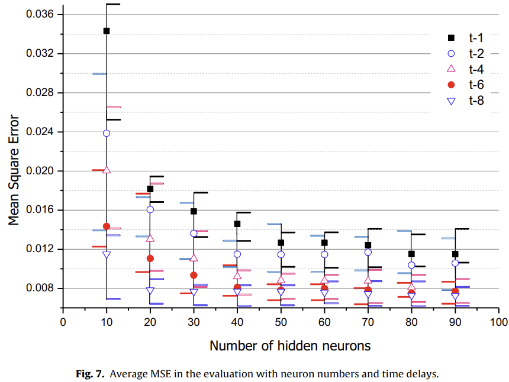
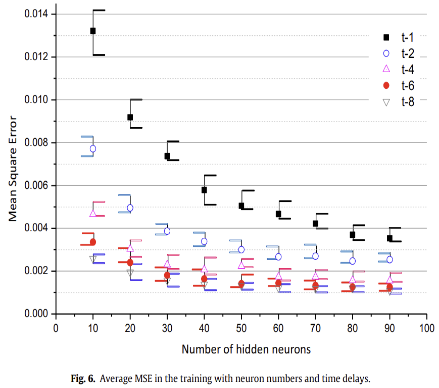


Figure 5 Average MSE in the training(left) and evaluation(right) with neuron numbers and time delays.

Figure 5 shows average and min/max error range of MSE between the actual energy consumption and the model, presented values in the training and evaluation procedures. The total neuron number in the hidden layer is varied from 10 to 90. The input-feedback time delay (i-k) is also parameterized, for the time step k = 1, 2, 4, 6, 8 (up to 2 hours). Although the training algorithm has the regularization function, the network requires a relatively large number of neurons and time delays to provide a stable performance. Considering the model complexity and computation time, it is reasonable to have 50 neurons in the hidden layer and have t-6 time delay for input variables and feedback for the network model in this study.

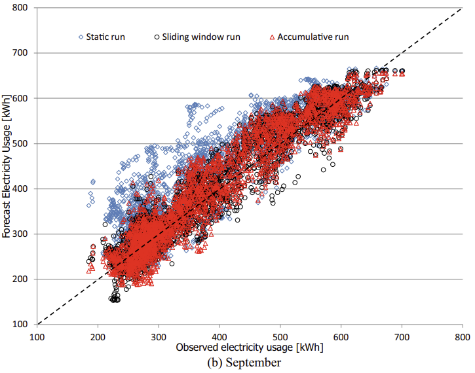
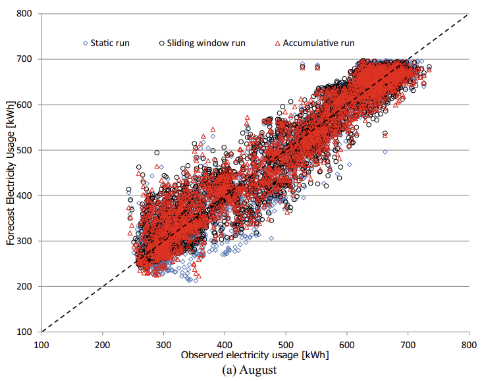


Figure 6 Comparisons of actual observed and forecasted electricity usage.

To verify the forecasting performance of the developed model, two months of data, August and September in year 2012, were used. Three training methods were considered: static, accumulative, and sliding windows. For the static training, the model is trained using four weeks of data in July, and it forecasts for August and September without retraining the model with newly available data. Accumulative training, an adaptive training method, uses accumulated data set from the first day of July to the day before the target day and retrained on a daily basis. The sliding windows method uses a fixed training data window size (four weeks) and the window is shifted by a day by removing the first day of the old training set and adding the new measurements into the data set. The networks are also retrained daily with the new training data

테이블이(가) 표시된 사진

자동 생성된 설명As illustrated in Figure 6, the overall prediction performance of each training type is similar to each other in August. The coefficient of determination (R-square) values is 0.904, 0.912, and 0.902 for each method. The result of September shows that accumulative and sliding windows are similar to August, but static training was not good as in August(R-square = 0.798).

Table 2 15 min prediction results with each training type.

Table 2 presents the statistical summary of CV(RMSE) and its standard deviation() on daily level. All three run types of training set have a similar performance in August. In September, CV(RMSE) and standard deviation of the adaptive training, sliding window and accumulative, are 9% and 2% in weekdays but the average daily forecast error (11.6%) and variance (5.03) by the static training is higher than for the other training types. The static training often overestimated the electricity consumption in weekdays. It may indicate that the energy consumption pattern of the site has 테이블이(가) 표시된 사진

자동 생성된 설명changed by the outdoor weather condition.

Table 3 Daily peak prediction results with each training type.

In addition to the daily electricity usage profile, the daily peak demand forecasting accuracy is one of the important factors in model evaluation. Second table summarizes the daily peak forecasting performance on weekdays and weekends for the two months. As similar to the daily prediction performance, the adaptive training method, accumulative and sliding window, provides more accurate results with smaller variance than the static training in September. With the regard of the uncertainty of actual electricity consumption in the short-term intervals, the day-ahead peak demand forecasting performance by the developed model is reasonable enough to make a management plan for the daily electricity consumption and peak electricity demand in the building.

* 1. **Conclusion**

Many studies have been done on model development for short-term electricity load forecast for electricity supply and demand, and have reported successful results in practical tests over hourly resolution. However, the sub-hourly electricity usage forecast is still a challenging problem due to the complexity of usage pattern and the highly noisy input data.

In this study, they demonstrated a new approach using a feature extraction and an artificial neural network to make a-day-ahead forecast of the electricity usage profile for a commercial building complex in a high temporal resolution of 15 minutes. The implementation results of the developed model for two months illustrate that the daily error with 15-minute resolution forecasting is stable around averaged CV(RMSE) of 10% for weekdays (coefficient of variation) and the model predicts the daily peak demand of weekdays within averaged APE of 5% for the period. It implies that the model can provide a day-ahead electricity usage profile with sub-hourly intervals and daily peak electricity consumption with a reasonable accuracy.

A good predictive model of energy consumption in buildings is useful not only for accurately forecasting future energy consumption but also in developing a good model predictive control (MPC) method that can reduce energy costs in buildings. As a future research, real-time electricity forecasting model with a smart meter which can be adaptive to weather and building operation changes for both the total meter and sub-meter level, and an anomaly detection model in short-term electricity usage pattern will be investigated.

1. **A hybrid model approach for forecasting future residential electricity consumption** [3]
   1. **Introduction**

“Data-Driven” models rely more on the data, assuming there is already a mathematical relationship between inputs (e.g. weather parameters) and outputs (e.g. total building energy consumptions). To further categorize the “Data-Driven” approach, there are “black-box” and “grey-box” models. The “black-box” model is purely machine learning driven and derived by measured data. Previous studies on energy consumption forecasting using “black-box” models are limited in one type of methods such as neural networks. The “Grey-box” model is built upon physical relationships, while the parameters of the physical model are usually unknown or uncertain. The measured data is used to identify those parameters and model tuning for better accuracy. The “Grey-box” models have been applied widely to estimate commercial building heating and cooling load, not many on the residential building sector.

The research gap between commercial and residential building comes from two main sources, which are (a) the lack of hourly or more granular data, and (b) the growing percentage of miscellaneous electrical load (MEL) in residential building, respectively. While MEL is approximately 15–25% of a typical home’s energy use, it is projected to increase by 52% by 2040. Unfortunately, for residential building, none of previous models can work as good as, because MEL is not weather related, and occupies a large portion of total building energy consumption.

In this paper, they try to close this gap by developing a new hybrid modeling approach through integrating “physical” and “data-driven” models to forecast hour ahead and 24 h ahead residential electrical load.

* 1. **Current state of the art**

Data-driven models using machine learning algorithms are explored a lot by researchers for building energy consumptions forecast. Common methods of machine learning are ANN(artificial neural networks), SVM(support vector machines), LS-SVM(least square optimization), and Gaussian family models, such as GPR(Gaussian Process Regression) and GMM(Gaussian Mixture Model).

Physical(Forward) models use laws of thermodynamics. There are two main parts of a model: building thermal zone envelopes and air conditioning modeling. For simplicity, most models integrate all physical phenomena (convection, conduction, and radiation) to one layer, and all layers (walls, floor, roof, etc.) to a thermal envelope zone. Thus, the envelope of the thermal zone in a building can be constructed as thermal resistance-capacity (RC) networks. In energy forecast applications, physical model suffers from a lack of building information.

Furthermore, many tests involved in data-driven methods are only conducted on total building energy consumption forecast. In contrast, physical models are limited to thermal usage prediction such as air conditioning. For onsite buildings, they are restricted either by the availability of building information or by measurements such as internal load profile. This research proposes a hybrid approach, making use of physical and data-driven models to forecast both total and air-conditioning (AC) energy consumptions. Improvement of accuracy can be analyzed by comparing pure data-driven model results. This method combines the advantage of the black box approach using data-driven methods. It overcomes the limits in forward model by adaptation of internal heat gain with data-driven non-air conditioning forecast.

* 1. **Methodology and approach**

Their primary objectives are to assess whether hybrid modeling approach for energy consumption forecasting is feasible and comparable to traditional data-driven models. Five data-driven models (ANN, SVR, LS-SVM, GPR, and GMM) are first built for forecasting the total and non-AC energy consumptions. One method, LS-SVM, is further selected to hybridize with a forward model to produce new total energy consumption prediction Results are compared to previous pure data-driven forecasts using our residential data collected in San Antonio.

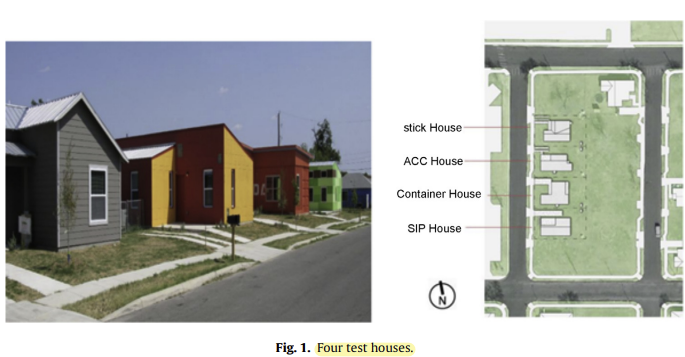
The residential data in our research was collected from four residential houses in San Antonio as shown in Figure 7. Those four samples are single family dwelling around 110 meter squared each. Houses are named according to construction materials: SIP (Structure Insulated Panel), ACC (Autoclaved Aerated Concrete), Container, and Stick. For material information, calibrated thermal resistances and capacities are provided by manufacturers. Energy consumptions are monitored at 5 min intervals for all the rooms including kitchen, bathroom, living and bedroom areas. Sub-metered data such as plugs, lighting, water heater and AC are also available. Outdoor air temperature and global horizontal solar radiation are collected from a weather station near the position. Solar radiations on the surface of each house at the same period are measured from sensors installed on site. Hour ahead and 24-h ahead models are evaluated using performance metrics: mean absolute percentage error (MAPE), and coefficient of variation (CV) of root mean square error. The smaller the MAPE to be calculated, the better the forecast is. Also, the smaller CV became, the more similar dispersions are between the forecast and the true value.

Figure 7 Four test houses.

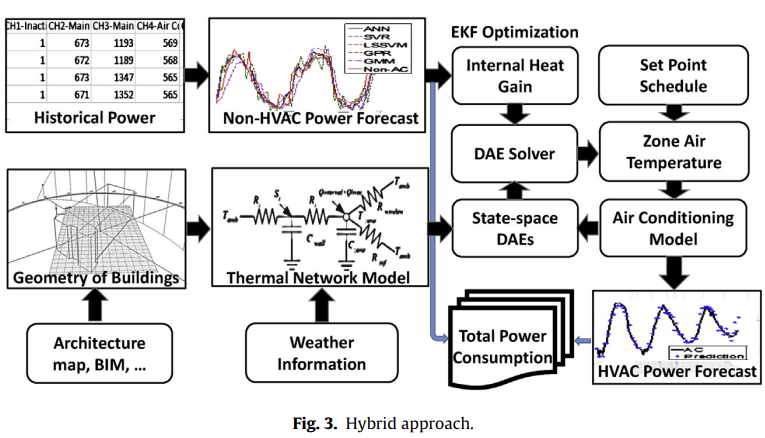
As shown in Figure 8, there are two major steps in the hybrid modeling approach. First, we forecast non-AC electricity consumption from historical non-AC consumption information. As internal heat gain is highly correlated with non-AC electricity consumption, it can be directly estimated from the non-AC prediction using heat convection and conduction. Secondly, predicted weather together with the internal heat gain forecast are inputted to the thermal network differential algebraic equations (DAEs)to simulate zone temperature. Then, the calculated zone temperature change is updated to an AC regression model with a set point schedule. Afterward, the AC cooling power consumption is further predicted. By summing up the AC and non-AC predictions, we have a total electricity consumption.

Figure 8 Hybrid approach.

For the non-AC forecast, the target is the non-AC energy consumption standing for plug load, lighting, water heater etc. The hour ahead input set is used to forecast the non-AC load at hour. For the hour ahead forecast case, input shows that the features need previous 1–5 h historical load information. The 24-h ahead input set is used to forecast the next 24-h load. For one future time load of the next 24-h window, the required input needs not only previous 24–31 h historical load but also moving averages.

* 1. **Results and discussions**

The experimental results of forecasting are organized in the following order: AC, non-AC, and total building energy consumption. For evaluation of the hybrid approach, 1 week from 2013.9.17 to 2013.9.24 is tested for all the residential houses. One week up to 1 month of non-AC power consumption data before the tested period is used for training. For all approaches, both hour ahead and 24-h ahead forecasting windows are presented and compared to the similar studies within the literature.

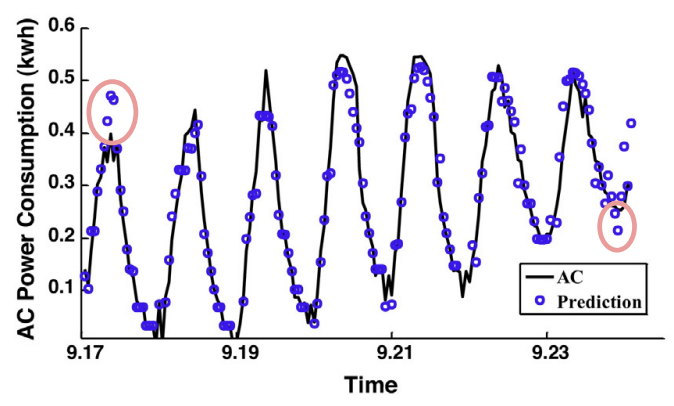
As the first step of hybrid modeling approach, we need to forecast the AC air conditioning energy usage. MAPEs are like 7.03%, 8.29%, 9.15%, and 8.03% for SIP, ACC, Container, Stick houses, respectively. An example of SIP house is shown in Figure 9. Most errors are beyond the baseline are due to the insensitivity of the model to AC cooling peak or transitional stages. With an R squared value of 0.9629, the forecasts show fairly accurate predictions on AC power consumption.

Figure 9 AC power consumption forecasting of SIP house.

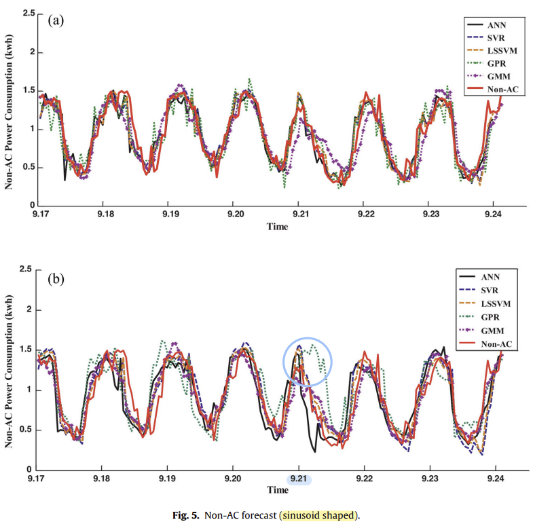
For the non-AC energy forecasting, they conducted a comparative analysis of five data-driven methods (ANN, SVR, LSSSVM, GPM and GMM). They present two types of common energy consumption patterns observed: sinusoid shaped and irregular shaped. For the sinusoid patterns, traditional methods such as ANN, and SVR have good performances with MAPE below 10%. Other methods also show reliable performances for most of the days. For the irregular patterns, it can be seen that some methods (MAPE around 11% for ANN) are still able to account for the larger variation that occurs, but forecasting performance decreased. Results of 24-h ahead forecasts are significantly different than hour ahead cases. They cannot keep the same performance for 24-hour ahead cases. The comparative energy consumption plots in Figure 10. It clearly illustrates that predictions are either over fitting (e.g. Sept. 21st) or insensitive at certain periods of tested houses.

Figure 10 Non-AC forecast (sinusoid shaped).

테이블이(가) 표시된 사진

자동 생성된 설명The reduced performance for 24-h ahead forecasts is due to model limitations and input constraints. Constraints on input feature selections are not obvious in hour ahead cases since the training features include previous 1-5 hour. However, for 24-hour ahead, models learned the patterns only from yesterday. Thus, the successfully prediction of longer forecast window is based on one key assumption: daily usages are similar without large deviation, which may not be the case for all days.

Table 4 Hour(left) and 24-h(right) ahead total load forecast.

Finally, for the total building energy consumption, the values of MAPE and CVs are provided in Tables 4 for hour ahead and 24-h ahead, respectively. In terms of CVs, the best method for hour ahead electricity consumption forecasting is the hybrid model. Meanwhile, in terms of MAPE, hybrid model, ANN, SVR, and LSSVM are not significantly different. However, hybrid approach results are improved.

A comparison of the 24 h ahead forecasting results find similar patterns. The hybrid model tends to have lower MAPE and CV values. However, the forecasting error compared to hour ahead is statistically distinct. 24-hour forecast’s MAPE and CV are higher than hour ahead forecast. This implies that temporal variance (hour or 24 h ahead) in the forecasting window has a significant impact on the predictive capability of both data-driven and hybrid models. Models trained with historical information with more recent load information has higher chances to characterize the 텍스트이(가) 표시된 사진

자동 생성된 설명electricity consumption behavior at an hourly interval.

Table 5 Hour ahead forecast performance of recent studies.

In the residential building energy research sector, forecasting is difficult due to the large variation of energy patterns from stochastic occupancy behavior. Three previous research studies for hour ahead total energy consumption forecasting of residential buildings are presented in Table 5. The forecasting error is considered reasonable if the model returns MAPE values between 9% and 30%, and CV values between 10% and 38% for hour ahead. We can conclude our hybrid approach is slightly better than the other existing models.

For the hybrid approach, the improvements can be traced back to the decomposition of total building energy consumption. By isolating certain appliances such as air conditioning, electricity usages are filtered to a sub-meter level. Another capability of the hybrid approach is the simultaneous prediction on both total energy consumption and sub-meter appliances, such as AC. This will be a valuable feature for the application of HVAC optimization and advance control strategies such as model predictive control.

* 1. **Conclusions**

This paper aims to develop and demonstrate an innovative hybrid modeling approach for residential building energy consumption forecasting. The proposed modeling approach was validated through 1 month measured data from four residential buildings. The final data analysis shows that hybrid modeling approach is slightly better for the hour ahead forecasting in terms of CV. For the 24-h ahead forecasting, all results are similar.

The study presented here uses 5 min interval data to predict AC electricity usage. However, such data is not typically available to homeowners and utility companies. In the future, we will explore the model performance by using hourly data. Further study is also required to understand and improve 24h ahead forecasting. Accurate residential building forecasting in the energy consumption is critical to improve energy efficiency. Fully use of monitored data to understand and explore not only total energy usage but also appliances such as air conditioning will help promote demand response programs and develop energy benchmark models.

**References**

[1] Amasyali, Kadir, and Nora M. El-Gohary. "A review of data-driven building energy consumption prediction studies." *Renewable and Sustainable Energy Reviews* 81 (2018): 1192-1205.

[2] Chae, Young Tae, et al. "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings." *Energy and Buildings* 111 (2016): 184-194.

[3] Dong, Bing, et al. "A hybrid model approach for forecasting future residential electricity consumption." *Energy and Buildings* 117 (2016): 341-351.

[4] Escrivá-Escrivá, Guillermo, et al. "New artificial neural network prediction method for electrical consumption forecasting based on building end-uses." *Energy and Buildings* 43.11 (2011): 3112-3119.

[5] Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.