An inquiry into time-series forecasting methods to quantify on-peak electricity demand savings through direct load control

Burak Gunay[1], Brodie Hobson[1], Andre Markus[1], Farzeen Rizvi[2]
[1]Carleton University, Ottawa, Canada
[2]National Research Council, Ottawa, Canada

Abstract

Direct load control (DLC) for HVAC systems are automated temporary adjustments to the sequences of operation that aim to reduce on-peak electricity demand. DLC algorithms represent a major cost-saving opportunity, particularly for large commercial buildings with automation and control infrastructure to shift and reduce HVAC loads on demand. While DLC can become a mainstream business for energy services companies, methods that can reliably estimate the impact of DLC algorithms on electricity demand during peak events are needed for their use in performance contracts. Simply put, we need methods that can answer the question: What would happen to the electricity demand if a DLC algorithm was not implemented? This paper compares six forecasting methods in terms of their ability to predict the baseline electricity profiles during historical peak periods in Ontario. To this end, an EnergyPlus model of an office building in Ottawa, Canada, was used to generate hourly cooling electricity use data with and without DLC. The results indicate that a motif-learning and rule extraction model outperformed four other SARIMA configurations and Meta's Prophet, and achieved an average demand reduction prediction error of less than 30%.

Introduction

Electrification has emerged as the main pillar of decarbonization. Adoption of electric vehicles and heat pumps for space heating in Canada is expected to double our electricity demand by 2050 (Canada, 2023). It is not cost-effective to address this supply-demand mismatch by simply adding more generation capacity, as much of the energy infrastructure would be underutilized outside a handful of extremely hot or cold hours every year. Demand-side management (DSM) solutions will play a major role in encouraging consumers to change their electricity use behaviour at peak periods.

HVAC systems represent a major grid asset for DSM. Firstly, HVAC systems account for over 50% of the coincident peaks. Secondly, given the inherent thermal mass of buildings, they are suitable for storing electricity as thermal energy for short periods and shifting demand outside peak periods. Lastly, the proliferation of building management systems (BMS) makes automated demand-minimizing algorithms possible.

Specifically, DLCs, typically implemented as a preheat-to-setback sequence in winter peaking jurisdictions and a precool-to-setup sequence in summer peaking jurisdictions, are promising DSM solutions. Particularly, in large commercial buildings, dynamic peak pricing places ample economic incentives to invest in technologies that enable automated load-shedding capabilities. This often means smart supervisory building systems that communicate with the grid operator or a third-party peak forecasting service and interact with a modern BMS via an application programming interface (API). For example, large commercial and institutional buildings in Ontario, Canada, are subject to mandatory demand charges, calculated based on a top-five coincident peak hour pricing scheme known as global adjustment. These demand charges represent up to 80% of their annual electricity bills without a DSM solution. Given the availability of such financial incentives, DLC is a prime candidate to become part of performance contracts, where a DSM solution vendor offers a technology to reduce a building owner's demand charges, and, in turn, receive a portion of the financial benefits. Yet, approaches to forecast short-term electricity demand are required to estimate demand savings.

Literature review

Generating short-term energy demand forecasts often calls for time-series modelling techniques. Autoregressive moving average (ARMA) models form the basis of many of the forecasting approaches employed in the literature (Deb et al., 2017). These models attempt to extrapolate energy use patterns by looking at the recent history of measurements and forecasting errors.

Given that energy use time-series in buildings is non-stationary, detrending is a requirement. While differencing is the generic detrending approach for autoregressive integrated moving average (ARIMA) model development, detrending with a traditional regression model was an alternative used in the literature. For example, Gunay et al. (2019) detrended a central heating plant's energy demand through a change point model (CPM) and then trained an ARMA model for the CPM's residuals. Similarly, Fang and Lahdelma (2016) employed a multiple linear regression model prior to using time-series modelling for a central heating plant.

Building energy time-series also exhibits daily and weekly seasonalities, which underline the need for autoregressive features extending to daily and weekly lags. Seasonal autoregressive integrated moving average (SARIMA) models enable the incorporation of such seasonalities. SARIMA models can be reinforced with external variables such as outdoor temperature, humidity, solar radiation. SARIMA models with external regressors are commonly referred to as SARIMAX, whereby the letter X indicates the presence of exogenous regressors. Determining seasonal and non-seasonal autoregressive, moving average, and differencing orders as well as selecting appropriate exogenous regressors can be a tedious task, particularly with hourly and subhourly resolutions required for DSM applications. While there are algorithms available in programming environments like Python and R to automate the model development process, the model selection and re-training can be computationally expensive. To this end, Blázquez-García et al. (2020), when employing SARIMA to forecast the energy demand of an elevator, put forward a genetic algorithm-based approach to speed up the model training processes.

Another limitation of SARIMA models is that they can only capture single periodic seasonalities, whereas a building's energy demand may exhibit multi-modal seasonalities. Algorithms such as Meta Prophet (Taylor and Letham, 2018) were developed to address the deficiencies of the SARIMA framework with respect to computational efficiency and modularity. Chaturvedi et al. (2022) compared Meta Prophet with SARIMA to forecast monthly energy demand in India, and found its performance to be superior. However, its performance in building-level electricity use forecasting at hourly and subhourly resolutions has not been assessed.

Machine learning models incorporating autoregressive features were also employed in forecasting short-term electricity demand. Various recurrent neural network formulations such as long short-term memory (LSTM) (Dubey et al., 2021; Chaturvedi et al., 2022; Li et al., 2022) and nonlinear autoregressive exogenous model (NARX) (Koschwitz et al., 2018) achieved performance levels exceeding SARIMA. Notably, training recurrent neural network models, unlike other linear forecasting models, requires large amounts of data, which may be a limiting factor in the early stages of deployment. Clustering and rule extraction were also coupled to identify common daily energy use clusters and their occurrence patterns (Ashouri et al., 2018).

The performance metrics typically used in assessing the accuracy of forecasting models include mean absolute percentage error (MAPE), adapted MAPE, Theil's inequality coefficient, root mean squared error (RMSE), and mean absolute error (MAE) (Blázquez-García et al., 2020; Fang and Lahdelma, 2016). Other variables in assessments include time-resolution of forecasts (e.g., hourly, daily, weekly) and the forecasting horizons. The performance of forecasting approaches has not been assessed in terms of their ability to quantify the effectiveness of DSM solutions. The ideal forecasting approach needs to quantify DSM effects accurately at hourly resolutions over forecasting horizons typical for DLC applications. However, there is a lack of published research that quantifies baseline models' predictive performance in DSM applications. Granderson et al. (2023) offers one of the only studies that compared existing baseline models suitable for demand flexibility applications. However, because their data emerge from real buildings, their approach lacked the ground truth necessary to quantify the impact of modelling errors on the measurement and verification of demand savings. To this end, building performance simulation- (BPS-) based data represent an untapped opportunity to generate the ground truth data needed. Thus, the objective of this paper is to compare forecasting techniques in terms of their ability to quantify DSM impacts.

Methodology

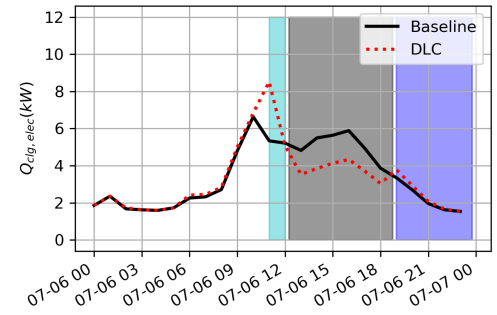
The methodology involves generation of synthetic cooling electricity use data through the BPS tool EnergyPlus. The BPS model is simulated with and without DLC sequences. The simulation without DLC represented the baseline data record and was used to train the forecasting models. The difference between DLC and non-DLC electricity demand represents the ground truth in peak reduction. This ground truth acts as a reference to assess the forecasting models' accuracy.

Synthetic cooling electricity use data generation

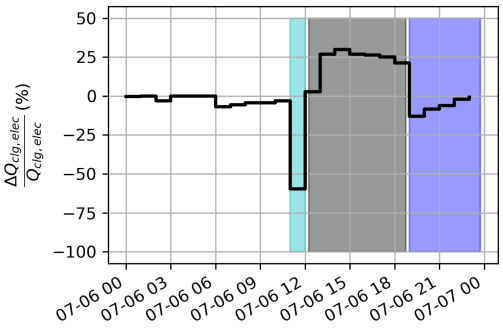
The BPS model used in this study is the top floor of a 662 m^2 academic office building located in Ottawa, Canada (ASHRAE Climate Zone 6A). All model details were extracted from the approved design details for construction. The model's HVAC system is comprised of a 26 zone variable air volume (VAV) air handling unit

Table 1: Parameters of the DLC sequences and the peak reductions during the setup periods.

Date	DLC parameters		Reduction during setup period (W/m^2)		
	Precooling	Setup	Minimum	Average	Maximum
05-Jul	-	12 pm - 9 pm	0.54	1.47	2.74
06-Jul	11 am - 12 pm	12 pm - 7 pm	1.07	2.03	2.77
11-Jul	1 pm - 5 pm	5 pm - 6 pm	1.89	2.15	2.65
26-Jul	1 pm - 4 pm	4 pm - 7 pm	1.03	2.55	4.78
27-Jul	1 pm - 4 pm	4 pm - 7 pm	1.19	1.57	2.54
28-Jul	11 am - 3 pm	3pm - 6 pm	0.88	2.34	3.42
05-Sep	10 am - 12 pm	12 pm - 9 pm	0.46	3.43	6.91



(a)



(b)

Figure 1: (a) The cooling electricity demand and (b) percent variation from the baseline with and without the DLC sequence described in Table 1 for July 6.

(AHU) system without reheat. The system's total airflow capacity is 5147 L/s. The zones are equipped with hydronic baseboard heaters. The outdoor air damper minimum position setpoint is defined as 43%.

A supply air temperature (SAT) setpoint reset based on outdoor air temperature (OAT) is utilized (i.e., an 18.3°C SAT below 4.5°C OAT, and a 12.8°C SAT above 12.8°C OAT, with the SATs linearly interpolated for intermediate OATs). The AHU is equipped with heating, cooling, and humidification coils and does not have heat recovery capabilities. The AHU operates daily from 6 am until 7 pm except for holidays; during operating hours, zone air temperature setpoints of 22°C and 23.5°C are used in the heating and cooling seasons, respectively. Holidays and afterhours use setback and setup mode temperature thresholds of 15°C and 30°C for the heating and cooling seasons, respectively. The AHU is allowed to cycle afterhours if more than 20% of zones have a temperature above the setup threshold or below the setbacks threshold. The VAV terminals and baseboards are hard-sized in accordance with the design specifications.

Hourly measured occupancy data representing presence/absence records of real individuals are randomly assigned to each zone. This is intended to reflect the influence of stochastic disturbances created by occupancy on the electricity demand time series. The lighting power density for each zone is set to 8 W/m^2; lighting behaviour is simplified such that artificial lighting was turned on upon occupants' arrivals in each zone and switched off 30 min after occupants' departures. Plug-in equipment electricity use is set to 100 W/person when zones were occupied. During after-hours on weekdays and weekends/holidays, this is reduced to 75 W/person and 50 W/person, respectively.

The building has a rectangular layout. On two sides, it is attached to similarly conditioned spaces. Thus, these surfaces are treated as adiabatic. The exterior walls have a window-to-wall ratio of approximately 60%. Infiltration through exterior walls is defined as 0.2 L/sm^2 when the AHU is not operating and reduced to 0.05 L/sm^2 when the AHU is operating to account for the effects of positive pressurization. The thermal transmittance of the walls, windows, and roof are 0.16, 0.9, and 0.15 W/m^2°C, respectively. The solar heat gain coefficient of the windows is 0.56.

The model is simulated with actual meteorological year data extracted from NASA's Power API from May 2023 to October 2023. Through the EMS application of EnergyPlus, seven plausible DLC sequences are injected during periods when Ontario's electricity demand exceeded 22,000 MW in 2023. The DLC sequences are implemented as precooling and setup sequences where

precooling and setup durations are modified. During precooling periods, the cooling temperature setpoint is reduced to 21.5°C; and during setup periods, the temperature setpoint is increased to 25.5°C. Table 1 provides a summary of the characteristics of the seven DLC sequences tested in this study. The sequences sustained an average demand reduction of 1.47-3.43 W/m^2, normalized by floor area. The demand reductions were notably higher with DLC sequences involving longer precooling and shorter setup periods. Demand reductions were higher at the beginning of the setup (2.54-6.91 W/m^2) and lower at the end of the setup periods (0.46-1.86 W/m^2). In Ontario, maintaining a 1 W reduction during the highest five peak periods translates to $0.50 to $0.80 of savings, based on historical global adjustment fees.

Figure 1 presents the baseline cooling electricity use and the modified cooling electricity use with the DLC sequence on July 6 listed in Table 1. The figure also enables the isolation of the impact that the DLC sequences had on the electricity demand. The three shaded regions (from left to right, respectively) illustrate the precooling, setup, and post-DLC behaviour. The precooling period is manifested with a sharp increase in the demand. During the setup period, the demand drops abruptly relative to the baseline levels, but as more zones exceed the setup temperature threshold of 25.5°C, the difference between the baseline and DLC trends diminishes. The post-DLC period is manifested with a sharp rebound of demand. The increased demand during precooling and post-DLC periods highlights the importance of accurately detecting the coincident peak periods. It is therefore critical to sustain a continuous two-way communication with the grid operator. As the demand begins to differ from the grid operator's original forecasts, it may be necessary to terminate precooling prematurely or extend the DLC period. While this is outside the scope of this paper, future work is needed to address these issues.

Figure 2 illustrates the weekly cooling electricity demand profile. Notably, the shaded region indicating the 10^{th} and the 90^{th} percentiles represent a wide range. Figure 3 presents the time-series decomposition of the cooling electricity use data. The decomposition results point to a significantly changing trend, indicating non-stationarity, and weekly seasonality, where weekends have smaller additive seasonal components than the weekdays. Also, the autocorrelation tailing off gradually before increasing again at a 24-h lag confirms the daily seasonality, and significant partial autocorrelation at lags one and two indicates a strong autoregressive nature of the trend up to a 2-hour lag (see Figure 4).

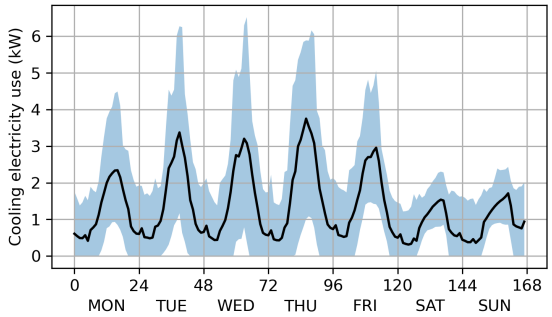


Figure 2: Weekly cooling electricity demand profile.

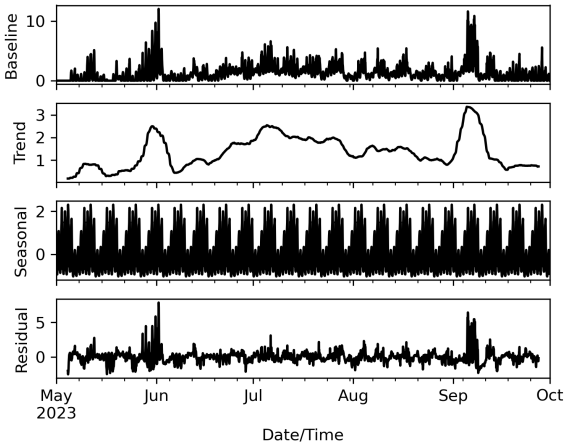
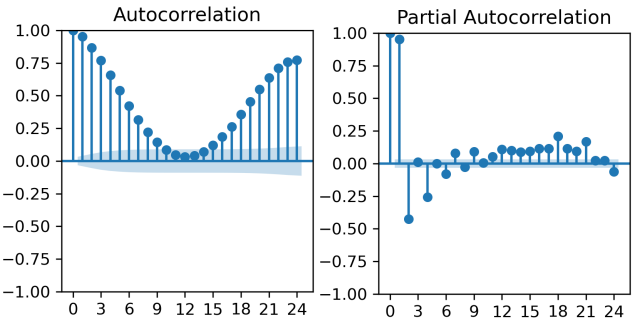


Figure 3: Additive time-series decomposition for the cooling electricity demand.

Forecasting algorithms

Six different forecasting approaches are compared in this study: 1) SARIMA, 2) SARIMAX, 3) uCPM-SARIMA, 4) mCPM-SARIMA, 5) Prophet, and 6) Motif-learning and rule extraction.

The non-seasonal differencing order of the SARIMA model d is identified through an augmented Dickey–Fuller test unit root test of stationarity. The seasonal model order D is determined based on an Osborn-Chui-Smith-Birchenhall (OCSB) unit root test for seasonality. The seasonal and non-seasonal autoregressive orders (P and p) and seasonal and non-seasonal moving average orders (Q and q) are determined through an iterative search minimizing the Akaike information criterion (AIC). The seasonal lag (m) is set as 24-h. The permissible differencing orders d and D are defined as zero, one, or, two. Likewise, p, q, P, and Q are allowed to take values between zero, one, or two. Limiting the model order was necessary to avoid overfitting and to keep the computational burden of the model selection process acceptable. The model selection is executed by using pmdarima v.2.0.4.



(a) *(b)*

Figure 4: (a) Autocorrelation function (acf) and (b) partial autocorrelation function (pcf).

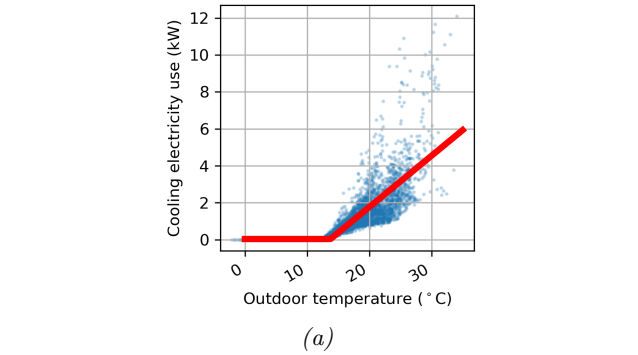
SARIMAX model was identical to that of the SARIMA model with the exception that the SARIMAX model also inputs the outdoor temperature.

uCPM-SARIMA model combines a univariate CPM and a SARIMA model. The univariate CPM has the following functional form:

$$q_{clg} = (T_{oa} - x_1)^+ \cdot x_2 + x_3 \quad (1)$$

where q_{clg} (kW) is the cooling electricity use, $T_{oa}{}^{\circ}C$ is outdoor temperature, and x is the unknown regression parameters. The superscript + indicates that the parenthetical term $(T_{oa} - x_1)$ can only take a positive value. If outdoor temperature T_{oa} is less than the change point temperature x_1, the term $(T_{oa} - x_1)$ becomes zero. The parameters x are estimated through nonlinear least squares method as implemented in scipy's *curve_fit* function. Figure 5(a) presents the univariate CPM trained with the dataset. A SARIMA model is then trained with the CPM's residuals $\hat{q}_{clg} - q_{clg}$, where $\hat{q}_{clg}$ (kW) is the cooling electricity use predicted by the CPM. Essentially, the CPM detrends q_{clg} and the SARIMA model forecasts the CPM's residuals $\hat{q}_{clg}$. The structure of this model follows the methodology described in Gunay et al. (2019). As shown in Figure 5(b-c), the residuals still have daily seasonality and an autoregressive behaviour for the first few lags.

mCPM-SARIMA model is identical to uCPM-SARIMA with the exception that mCPM-SARIMA employs a multivariate CPM instead of a univariate CPM. The multivariate CPM inputs the outdoor temperature and a binary operating status indicator as regressors. In particular, the building is expected to require higher space heating and cooling loads during operating hours, mainly due to ventilation (i.e., the need to condition the outdoor air in the supply air). In order to eliminate the burden of entering the operating schedule from the model user, the model automatically identifies the operating status following the procedure described in



(a) *(b)* *(c)*

Figure 5: (a) Univariate CPM defined in Eqn. 1, and (b) acf and (c) pcf of the residuals of the CPM.

Markus et al. (2021). The multivariate CPM has the following functional form:

$$q_{clg} = \begin{cases} (T_{oa} - x_1)^+ \cdot x_2 + x_3, & \text{if } b = on \\ (T_{oa} - x_4)^+ \cdot x_5 + x_6, & \text{otherwise.} \end{cases} \quad (2)$$

where b is the binary operating schedule. It is formulated as follows:

$$b = \begin{cases} (t_{day} > x_7) \wedge (t_{day} < x_8) & \text{if } weekday \\ (t_{day} > x_9) \wedge (t_{day} < x_{10}) & \text{otherwise.} \end{cases} \quad (3)$$

where x_7 and x_9 represent the start time of the daily operating schedule on weekdays and weekends, respectively. Similarly, x_8 and x_{10} represent the end time of the daily operating schedule on weekdays and weekends, respectively. The regression parameters $x_{1\,to\,10}$ are identified through a metaheuristic search algorithm as implemented in scipy's *differential_evolution* function. Figure 6(a) presents the multivariate CPM trained with the current dataset, where the red line represents the workhours and the black line represents afterhours electricity use behaviour. With the mCPM-SARIMA model, a SARIMA model is trained with the residuals of this multivariate CPM. Again, as shown in Figure 6(b-c), the residuals still have daily seasonality and autoregressive behaviour for the first few lags.

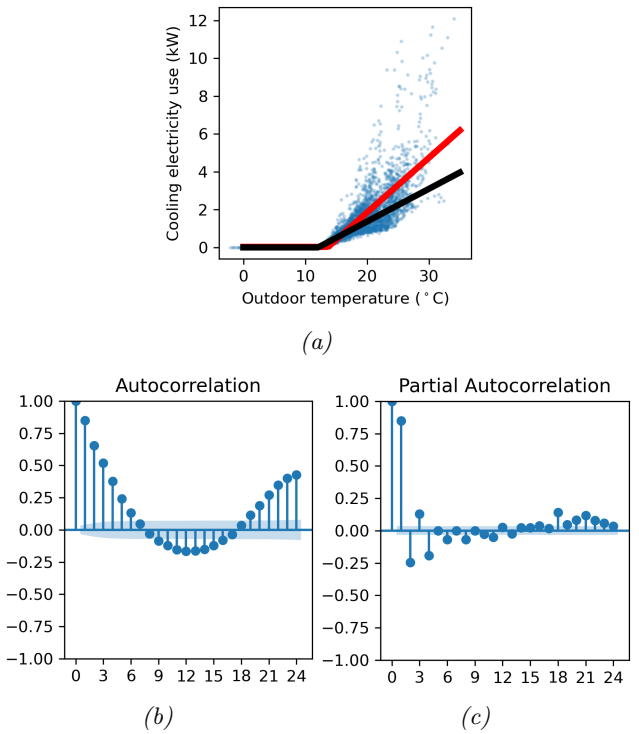


Figure 6: (a) Multivariate CPM defined in Eqn. 2, and (b) acf and (c) pcf of the residuals of the CPM.

The fifth forecasting approach is Meta's Prophet. Prophet adopts a piecewise linear or logistic growth curve trend upon change points automatically detected in trends. The Prophet model is defined with both weekly and daily seasonalities where the seasonalities are modelled as multiplicative. The model inputs outdoor temperature as an exogenous variable. These settings were identified through a trial-and-error process by manually inspecting the model performance.

Motif-learning and rule extraction method follows Ashouri et al. (2018)'s approach. First, daily cooling electricity demand profiles are clustered into up to ten distinct day types via k-means algorithm using their Euclidean distances. To this end, sklearn's cluster package was used. Figure 7 presents the ten day types identified through clustering. The daily profiles represent the median values for each hour of the day for each cluster. The day types are then classified based on two features: the day of week as a categorical variable and the mean daily outdoor temperature. To this end, sklearn's tree package was used. The classification model offered concise rules that effectively state: when the daily outdoor temperature is between a and b and the day of week is c, expect a daily cooling electricity profile per clustered day-type i.

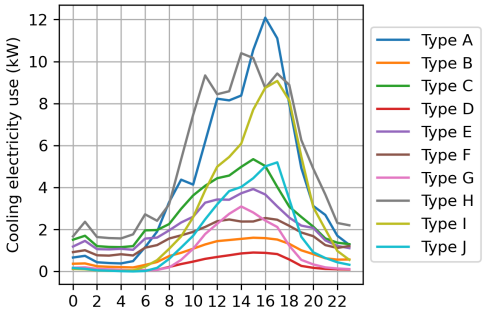


Figure 7: Centre-of-cluster daily cooling electricity demand profiles representing ten day types.

Results and discussion

The models need to generate a forecast of the baseline electricity demand from the beginning of precooling until the end of setup. For the DLC sequences tested in this study, this represents a 5- to 11-hour forecasting horizon. Figure 8 presents forecasts generated by the six models for one of the DLC sequences from 11 am to 7 pm on July 6. Visual inspection confirms that the forecasts by all six approaches were reasonably close to the ground truth baseline, and all were able to closely follow the timing and magnitude of the local maxima and minima during the setup period.

While the DLC event on July 6 was appropriately captured by all models, the same cannot be said for the other DLC events (see Table 2). For example, the error in predicting the average demand reduction during the setup period on September 5 was greater than 3.4 W/m^2 for all models except motif-learning and rule extraction. None of the six models were able to predict the average demand reduction during the setup period on July 26 at less than 30% error. Of the seven DLC events, SARIMA and Prophet failed to accomplish less than 30% error in predicting the average demand reduction during setup three times; the uCPM-SARIMA and mCPM-SARIMA failed four times, while the SARIMAX and motif-learning and rule extraction failed twice. The MAPE in predicting the average demand reduction during setup was less than 30% only with the motif-learning and rule extraction model. Notably, the accuracy of the forecasts was higher in predicting the maximum demand reduction during setup than the minimum demand reduction during setup. This is because the accuracy of forecasts degrades over time and the fact that maximum demand reduction occurs at the beginning of the setup period and the minimum demand reduction occurs at the end of the setup period.

It is important to note that none of the tested forecasting models offer a sufficiently adequate solution to accu-

Table 2: Predicted and actual average, minimum and maximum demand reductions sustained during the six setup periods. The figure reports the error statistics MAPE and MAE for the average, maximum, and minimum.

	05-Jul	06-Jul	11-Jul	26-Jul	27-Jul	28-Jul	05-Sep	MAE (W/m^2)	MAPE (%)
	Average reduction during setup (W/m^2)								
Ground truth	1.47	2.03	2.15	2.55	1.57	2.34	3.43	-	-
SARIMA	1.03	1.21	2.05	0.80	1.13	2.29	-0.27	2.51	47
SARIMAX	1.52	1.66	1.93	0.68	1.24	1.93	0.05	2.40	43
uCPM-SARIMA	1.31	1.75	1.37	0.26	1.60	1.46	-0.48	2.65	54
mCPM-SARIMA	1.43	1.62	1.47	0.36	1.76	1.39	-0.83	2.66	56
Prophet	1.13	1.48	2.71	0.93	1.82	1.59	-3.87	2.88	73
Motif-learning	1.15	2.99	1.95	1.47	1.20	2.72	2.56	1.74	27
	Maximum reduction during setup (W/m^2)							**MAE (W/m^2)**	**MAPE (%)**
Ground truth	2.74	2.77	2.65	4.78	2.54	3.42	6.91	-	-
SARIMA	1.55	1.99	2.05	1.20	2.30	2.55	3.27	1.56	42
SARIMAX	2.89	2.41	1.93	1.03	2.48	2.26	3.81	1.33	36
uCPM-SARIMA	3.23	2.70	1.37	0.58	2.91	1.90	2.76	1.73	47
mCPM-SARIMA	3.46	2.88	1.47	0.68	3.13	1.83	2.46	1.82	49
Prophet	2.56	2.60	2.71	1.36	2.78	1.90	0.47	1.72	47
Motif-learning	2.02	3.72	1.95	2.97	2.03	3.88	5.39	0.95	26
	Minimum reduction during setup (W/m^2)							**MAE (W/m^2)**	**MAPE (%)**
Ground truth	0.54	1.07	1.89	1.03	1.19	0.88	0.46	-	-
SARIMA	0.16	-0.22	2.05	0.25	-0.15	1.92	-4.13	3.71	100
SARIMAX	0.15	0.33	1.93	0.25	-0.18	1.50	-3.54	3.62	98
uCPM-SARIMA	-0.85	0.27	1.37	-0.16	0.02	0.92	-3.99	4.03	109
mCPM-SARIMA	-0.81	0.09	1.47	-0.10	0.12	0.87	-4.19	4.05	110
Prophet	0.17	0.56	2.71	0.26	0.57	1.18	-7.89	4.05	110
Motif-learning	-0.39	1.95	1.95	0.34	0.03	1.59	0.50	2.84	77

rately quantify demand reductions of a DSM strategy. Each of these methods, in its own right, is capable of generating load forecasts with close resemblance to the baseline demand profile. In fact, if the forecasting errors were normalized by the cooling electricity demand, the MAPE would have been 10% or less for almost all DLC events and models. However, because the goal is to isolate the impact of a DLC strategy (typically not influencing the baseline profile more than 25%), even the slightest forecasting error can have a dramatic impact on the methods' ability to quantify demand reductions. This underlines the need for research dedicated for the direct estimation of demand savings.

An important consideration in designing forecasting models for demand savings is their need for data. Training forecasting models requires at least several months' worth of data, which means that a functional model will not be available for new buildings and existing buildings without historical data. Simpler models (e.g., merely looking at the past several days and hours) may be used temporarily until a full-fledged forecasting model can be reliably trained.

Conclusion

This study compared six forecasting models from the literature in terms of their ability to quantify the impact of a DLC sequence. To this end, two synthetically generated cooling electricity use data records were used, one with and one without DLC sequences. Different DLC sequences were injected on seven different days in 2023, when Ontario's coincident peak electricity demand exceeded 22,000 MW. An EnergyPlus model of an academic office building was used with the DLC sequences implemented in EnergyPlus' EMS application. The results indicate a $\sim$ 25% average cooling demand reduction during setup periods - which corresponds to an average electricity demand reduction of 1.47 W/m^2 to 3.43 W/m^2 (normalized by floor area).

The forecasting models tested include a SARIMA and SARIMAX model, two SARIMA models detrended with a univariate CPM and a multivariate CPM, Meta's Prophet model, and a motif-learning and rule extraction model. The motif-learning and rule extraction model was the only one capable of sustaining a MAPE of less than 30% in predicting the average demand re-

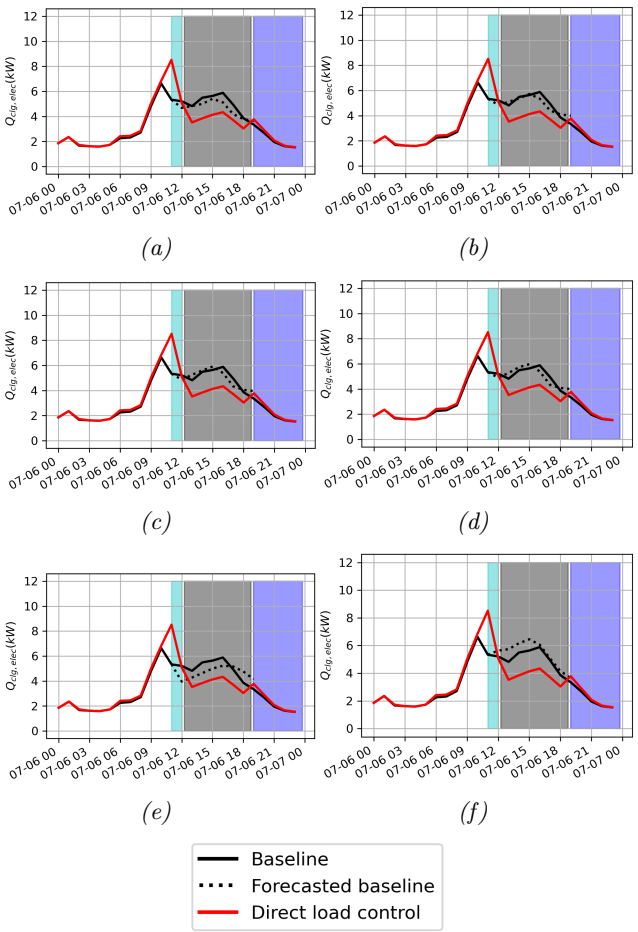


Figure 8: The ground truth and the forecasts by the (a) SARIMA, (b) SARIMAX, (c) uCPM-SARIMA, (d) mCPM-SARIMA, (e) Prophet, and (f) Motif-learning and rule extraction models on July 6.

duction during setup periods. The study also highlighted the need for new models to quantify demand savings during DLC events, as none of the existing models were consistently accurate on all seven test periods.

Acknowledgment

This research is funded by a research contract with the National Research Council Canada (contract no: 1012783).

References

Ashouri, A., Y. Hu, G. R. Newsham, and W. Shen (2018). Energy performance based anomaly detection in non-residential buildings using symbolic aggregate approximation. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, pp. 1400–1405. IEEE.

Blázquez-García, A., A. Conde, A. Milo, R. Sánchez, and I. Barrio (2020). Short-term office building elevator energy consumption forecast using sarima. *Journal of Building Performance Simulation 13*(1), 69–78.

Canada (2023). *Budget 2023*. Department of Finance Canada.

Chaturvedi, S., E. Rajasekar, S. Natarajan, and N. McCullen (2022). A comparative assessment of sarima, lstm rnn and fb prophet models to forecast total and peak monthly energy demand for india. *Energy Policy 168*, 113097.

Deb, C., F. Zhang, J. Yang, S. E. Lee, and K. W. Shah (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews 74*, 902–924.

Dubey, A. K., A. Kumar, V. García-Díaz, A. K. Sharma, and K. Kanhaiya (2021). Study and analysis of sarima and lstm in forecasting time series data. *Sustainable Energy Technologies and Assessments 47*, 101474.

Fang, T. and R. Lahdelma (2016). Evaluation of a multiple linear regression model and sarima model in forecasting heat demand for district heating system. *Applied energy 179*, 544–552.

Granderson, J., S. Fernandes, E. Crowe, M. Sharma, D. Jump, and D. Johnson (2023). Accuracy of hourly energy predictions for demand flexibility applications. *Energy and Buildings*, 113297.

Gunay, B., A. Ashouri, and W. Shen (2019). Load forecasting and equipment sequencing in a central heating and cooling plant: A case study. *ASHRAE Transactions 125*, 513–523.

Koschwitz, D., J. Frisch, and C. Van Treeck (2018). Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and narx recurrent neural network: A comparative study on district scale. *Energy 165*, 134–142.

Li, Y., Z. Tong, S. Tong, and D. Westerdahl (2022). A data-driven interval forecasting model for building energy prediction using attention-based lstm and fuzzy information granulation. *Sustainable Cities and Society 76*, 103481.

Markus, A. A., B. W. Hobson, H. B. Gunay, and S. Bucking (2021). A framework for a multi-source, data-driven building energy management toolkit. *Energy and Buildings 250*, 111255.

Taylor, S. J. and B. Letham (2018). Forecasting at scale. *The American Statistician 72*(1), 37–45.