Les Cahiers du GERAD ISSN: 0711–2440

Local cold load pick-up estimation using time-stamped measurements

S. Bajic, F. Bouffard, H. Mishalska, G. Joós

G–2021–68

October 2021

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2021
– Bibliothèque et Archives Canada, 2021

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2021
– Library and Archives Canada, 2021

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Local cold load pick-up estimation using time-stamped measurements

Sanja Bajic [a, b]

François Bouffard [a, b]

Hannah Michalska [b]

Géza Joós [b]

[a] *GERAD, Montréal (Qc), Canada, H3T 1J4*

[b] *Department of Electrical and Computer Engineering, McGill University, Montréal (Qc), Canada, H3A 0E9*

sanja.dzeletovic@mail.mcgill.ca
francois.bouffard@mcgill.ca
hannah.michalska@mcgill.ca
geza.joos@mcgill.ca

October 2021
Les Cahiers du GERAD
G–2021–68

Abstract : Thermostatically-controlled loads have a significant impact on electricity demand after service is restored following an outage, a phenomenon known as cold load pick-up. Widespread deployment of smart metering devices is also opening up new opportunities for data-driven load modelling and prediction. In this paper, we propose an architecture for local estimation of cold load pick-up using time-stamped local load measurements. The proposed approach uses ARIMA modelling for short-term foregone energy consumption prediction during an outage. Predictions are made on a hourly basis to estimate the energy to potentially recover after outages lasting up to several hours. Moreover, to account for changing customer behavior and weather, the model order is adjusted dynamically. Simulation results based on actual smart meter measurements are presented for 50 residential customers over the duration of one year. These results are then validated using physical modelling of residential loads and match well ARIMA-based predictions.

Keywords: Cold load pick-up, load forecasting, power distribution systems

Résumé : Les charges thermostatiques, telles que le chauffage électrique de l'eau et des bâtiments, ont un impact significatif sur la demande en électricité sur une artère de distribution lorsque le service est rétabli suite à une panne. Également, le déploiement à grande échelle de compteurs communicants ouvrent la porte à une multitude d'occasions de modélisation et de prévision de la charge guidé par les données. Dans cet article, nous proposons une architecture visant l'estimation locale du processus de reprise en charge d'un client à partir de ses données temporelles de consommation. L'approche préconise l'utilisation d'une modélisation de type ARIMA afin d'estimer la consommation perdue lors d'une panne. Les prévisions sont établies à un intervalle régulier afin d'estimer l'ńergie ȑecouvrer à la suite d'une gamme de durées de pannes allant jusqu'à quelles heures. De plus, afin de prendre en compte l'influence du changement dans les conditions météorologiques et le comportement du client, on y intègre une mise à jour dynamique du modèle de reprise en charge. Des résultats expérimentaux basés sur des données de compteur rélles réparties sur une période d'un an sont présentés pour un échantillon de 50 clients. Ces résultats sont validés sur des modl̀es basés sur les caractéristiques physiques de charges résidentielles et montrent que l'approche guidée par les données est hautement précise.

Mots clés : Reprise en charge, prévision de la demande, réseaux de distribution d'électricité

Acknowledgements: This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, Ottawa, ON.

1 Introduction

Utilities, whose customers have large thermostatically controlled loads (TCL), face significant challenges due to sharp load increases when service is restored following extended outages. This is especially important in regions where and when winter temperatures are low (e.g. $-40°$C), as in Quebec, Canada [1]. Similar comments apply during heatwaves for power systems with significant air conditioning loading.

The sudden and significant increase in the load after service restoration, referred to as *cold load pick-up* (CLPU), is principally attributed to loss of load diversity. CLPU is habitually accounted for in planning and operation procedures; in fact, it is one of the top determining factors of high-to-medium voltage transformer sizing and their overload capability [2, 3]. Utilities are actively working on the improvement of strategies to reduce CLPU demand and in that way reduce the peak loading and thus potentially postpone investments.

To deploy approaches like the one proposed in [3], accurate and single-customer CLPU estimations are needed. This paper proposes a methodology for estimating CLPU of a single residential dwelling with significant proportion TCLs based on smart meter (time-stamped) measurements. The emphasis is placed on obtaining best estimation performance for consumers with electrical heating, for which CLPU during low-temperature periods represents an additional challenge for the distribution utility both in operations and planning.

The first step of the proposed approach is based on the estimation of future energy consumption for the duration of the potential outage using ARIMA. The ARIMA parameters are updated hourly to adapt to changing conditions that may affect demand and eventual post-outage CLPU. Model order selection is performed for the most recent data sample based on the analysis of data stationarity, values of the auto-correlation function (ACF) and partial auto-correlation function (PACF).

The remainder of this paper is organized as follows. In Section II, we present a literature review of CLPU phenomena and short term energy consumption prediction. Next, we present the architecture of local CLPU estimation in Section III. In Section IV, we develop the mathematical formulation for the ARIMA model and model selection. Section V consists of the verification of CLPU ARIMA model selection efficiency on actual smart meter data and validation of CLPU duration assessment against simulated cases based on explicit load models implemented in GridLab-D, an industry-grade distribution system simulation software. The conclusion of the work follows in Section VI.

2 Literature review

Previous work on the subject of CLPU has focused on physically-based load modeling and post-restoration load response. Physically-based CLPU models of heating and cooling loads were first developed in [4]. In [5], a multi-state physical model is employed to evaluate the magnitude and duration of CLPU. Physically-based models for computing electric loads on residential feeders are presented in [2]. A feeder CLPU curve for an aggregated load consisting of single-family homes with electrical space and water heating is computed as a piecewise linear function of outdoor temperature by the authors of [6]. An aggregate model for a heterogeneous population of TCLs is derived in [7] to capture accurately the collective load behavior under demand response. All of the above models were developed for aggregated loads, and thus cannot be readily applied for single residential CLPU estimation. Despite the high accuracy of physically-based load modelling, they are inherently difficult to build and maintain because of their significant and changing model calibration requirements.

Hence, measurement-based approaches show better potential for CLPU modelling. The CLPU model of aggregate residential loads developed in [8] is based on field measurements and a load model. In that article, the authors introduce the CLPU factor that represents the ratio between the rated power and the expected power during the restoration and quantifies the energy to be recovered which depends on the duration of an outage. A general model for determining the maximum restorable load was developed in [9]. Here, CLPU is estimated using only *practical guidelines* - the rules defined based on the practical experience for accounting for CLPU peak and duration in the service restoration procedure. Practical guidelines for CLPU estimation have routinely been applied in many utilities, as they proved to be sufficiently robust and effective provided that the guideline assumptions are satisfied. The drawback of this approach, however, is that it does not apply to the prediction of a single load CLPU peak and duration. The aforementioned represents a limitation in novel approaches for CLPU mitigation strategies, as the assumptions are built based on the aggregated load response.

Recently, there has been an increase in deployment, and consequently, the utilization of timestamped customer load measurements. The authors of [10] developed a data-driven framework for estimation of CLPU response of aggregated load. The timestamped customer load measurements are used to enhance the results of a least square support vector machine. In [11], timestamped customer load measurements along with outdoor temperature measurements are used to estimate energy to be recovered after an outage. Compared to the temperature-based linear regression approaches, auto-regression has shown better performance in terms of mean squared error (MSE).

One of the determining factors of CLPU duration is the amount of energy not served to a customer during an outage. Different approaches have been proposed in the literature for short-term load and energy consumption prediction. These include artificial neural networks (ANN), time series analysis and multiple linear regression. In [12], the four most competitive methods (ANN, ARMA, Hold-Winters Taylor (HWT) exponential smoothing and singular value decomposition (SVD) based exponential smoothing) were applied to predict short term load demand. The aim was to establish an univariate forecasting method that can potentially be used for generating real-time online forecasts in an automated framework. In [13], exponential smoothing methods for short-term load forecasting outperformed weather based approaches in a time window of five hours. Paper [14] presented the method for short-term load forecast based on auto-regressive integrated moving average (ARIMA) in combination with input in a form of the knowledge of a network operator with past experience regarding how an area of the network behaves. Here, ARIMA has shown good performance in estimating load peaks using time series of measured load. The vast majority of the literature on the topic of short-term load forecasting, including [12, 13, 14], refer to forecasting the expected electricity demand at aggregated levels. Since consumption recorded from individual smart meters exhibits much higher volatility, the methods that are proven as effective for aggregated load do not necessarily dominate in the area of single load consumption estimation and prediction.

The massive smart meter deployment over the past decade has provided the industry with large amounts of data which are highly granular, both temporally and spatially [15]. A comparison of the four most common approaches was performed in [16]: ARMA, SVM and two approaches based on neural networks: nonlinear auto-regressive recurrent neural networks, and long-term short-memory networks. Here, ARMA was outperformed by the other approaches. However, this conclusion is questionable since an ARMA model assumes stationary time series data, which is generally not the case for timestamped household load measurements. An ARIMA model would be a more suitable choice since it can be applied to non-stationary timeseries data. A detailed forecast model based on the four approaches (multiple linear regression (MLR), regression trees, support vector regression (SVR) and ANN) is proposed in [17]. Here, a large number of features is introduced, such as temperature, humidity, precipitation rate, wind chill, average load values for several intervals and others and compartmental forecasting models are generated for every hour which introduces an additional complexity to the model.

ARIMA models have several advantages that make it suitable for short term energy consumption prediction:

1. As a statistical forecasting approach, ARIMA has shown better properties in terms of accuracy and computational performance than ANN approaches [18, 19, 20].

2. ARIMA has shown good forecasting results in univariate modelling [12]. Univariate modelling of electrical energy consumption is more convenient for local and independent execution since it does not require additional data apart from timestamped load measurements available locally.
3. ARIMA does not assume knowledge of any underlying model or relationships, which makes it a good choice for predicting models with uncertainty [21].

3 Local CLPU assessment architecture

In this section, we present the architecture of the proposed framework for local CLPU estimation. The method is designed for local execution at the service delivery point, fed by the most recent local smart meter data. The outputs of the algorithm are predictions of the CLPU peak and the duration of CLPU events following a range of outage durations.

3.1 Single load CLPU

Due to the significant difference between a CLPU curve at the feeder and at the single dwelling level, we should approach differently the estimation of CLPU for this two cases. Figure 1 shows a typical feeder load during a CLPU event. On the other hand, Figure 2 illustrates CLPU of a single residential load for the same outage as in Figure 1. By inspection, we can see that feeder aggregated CLPU has exponential decay, while the single customer CLPU is discrete (i.e., all or nothing) in nature, due to the limited number of devices that can influence undiversified power consumption.

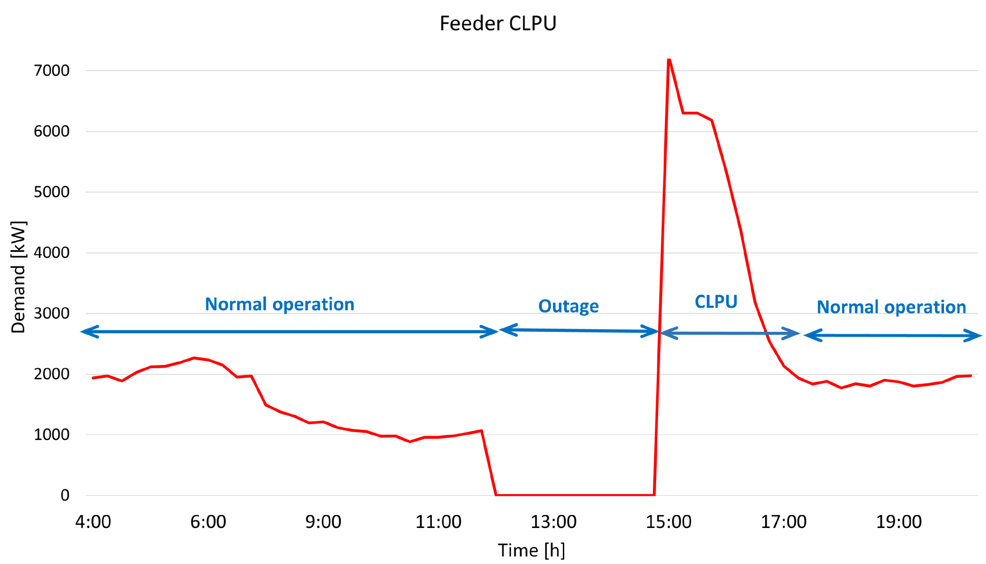

Figure 1: Feeder CLPU

For a single dwelling, CLPU characterization boils down to two decoupled problems:

1. Estimation of peak demand during CLPU,
2. Estimation of energy not served during the outage, which is proportional to the time required to regain normal operation.

Due to the discrete nature of residential load CLPU, the duration of the CLPU peak demand varies with the energy not delivered during the outage. The peak power, on the other hand, is limited by the maximum power of the dwelling's major TCLs. Consequently, in this paper, we address the CLPU prediction for a single customer by first estimating the energy not delivered during the outage (lasting

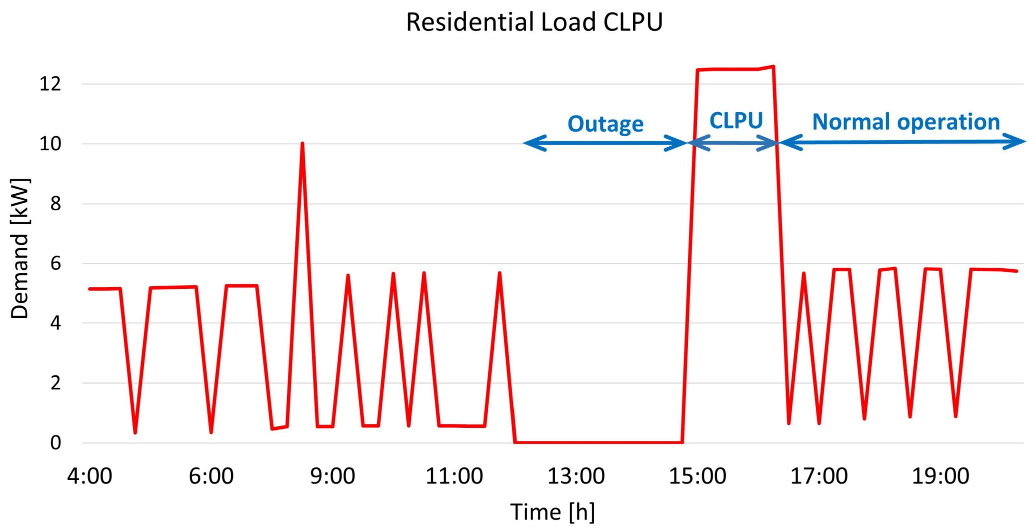

Figure 2: Single dwelling CLPU

from t_0 to τ_1) and the peak power, P_{CLPU}, at which this energy is consumed in the post-outage period (lasting from τ_1 to τ_2) (Figure 3). In this Figure 3, the continuous curve represents the energy consumption in normal operating condition had no outage actually occurred.

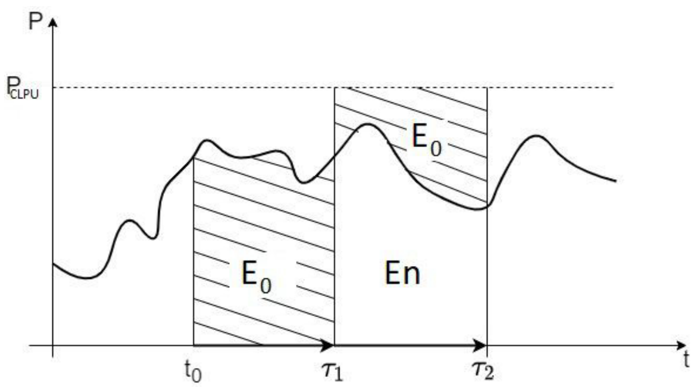

Figure 3: CLPU estimation using foregone energy

If an outage starts at t_0, the CLPU duration $T_C = \tau_2 - \tau_1$ depends on the outage duration time, $\tau_1 - t_0$ because of the need to recover foregone energy lost during the outage—which maps to an actual temperature change inside the dwelling:

$$E_C(t_0, \tau_1, \tau_2) = E_o(t_0, \tau_1) + E_n(\tau_1, \tau_2) \tag{1}$$

where E_C is the net CLPU energy, E_o is the foregone energy during the outage and E_n represents the no-outage "normal" energy consumption (i.e., as if there had been no outage) during the CLPU period. Using the anticipated energy consumption in normal operating conditions, E_o, and the amount of energy to recover E_C, we can estimate the time needed for return to non-CLPU operation, by estimating the time to consume the energy E_C at maximum power P_{CLPU}.

3.2 CLPU peak

The maximum power consumption of a single load is limited by the contract with the utility and the rating of household protection equipment. Daily peak power values depend on the sum of the ratings of the appliances likely to be used simultaneously. The dominant part of the load in the low-temperature periods consists of electrical heating units.[1] Both the temperature dependency of the heating load and household appliances' power consumption influence the daily maximum load power. Hence, the power in the post-outage periods of a single load, due to the load power limitations, can be approximated by the daily peak power. Furthermore, we can calculate maximum daily power by applying an autoregressive model to close historical daily maximum power measurements:

$$P_{CLPU}(t) = \hat{P}_{CLPU}(t) + a_t = \sum_{i=1}^{N} \varphi_i P_{peak}(t-i) + a_t \tag{2}$$

Here, the peak load in CLPU conditions $P_{CLPU}(t)$ at time t is estimated by $\hat{P}_{CLPU}(t)$ using an autoregressive model of peak power values P_{peak} recorded in the N previous days, with coefficients φ_i and residual error a_t. This approach is justifiable since during multi-day cold snaps daily peak power values tend to converge.

3.3 CLPU duration

In this paper, we define the CLPU duration of a single customer as the interval of time during which a customer has continuous energy consumption at a high power value, as shown in Figure 2. The CLPU duration T_C after $\tau_1 - t_0$ long outage is estimated using the prevailing CLPU peak power estimate $\hat{P}_{CLPU}(t_0)$ and $E_o(t_0, \tau_1)$, which is also predicted by an ARIMA model addressed formally in the next section

$$\hat{T}_C(t_0, \tau_1) = \frac{\hat{E}_o(t_0, \tau_1)}{\hat{P}_{CLPU}(t_0)} \tag{3}$$

4 Implementation

Residential power consumption is a stochastic process, and, therefore, it should be modelled as such. We show next how future consumption (in kWh) of a dwelling can be easily and accurately estimated in real time using an ARIMA model. On a rolling horizon basis, E_C and P_{CLPU} are estimated at the current time t_0 for a range of outage durations $\tau_1 - t_0$.

4.1 Architecture overview

The proposed solution architecture is shown in Figure 4. The architecture consists of the smart meter at a residential home with an additional calculation module that performs CLPU peak and duration estimation, available on-demand to the distribution system operator. The model order selection algorithm, from which the ARIMA model is obtained, i.e. the values of the ARIMA model order parameters p, d and q, is performed continuously using fresh time series data. Using that same data, a model of specified order is trained for estimation of the forgone energy during a potential outage. The model order selection procedure is re-triggered on a weekly basis or in the case of the divergence of the maximum likelihood estimation (MLE) optimization algorithm used for estimation of ARIMA model parameters. The resulting CLPU peak and CLPU duration estimates are intended to facilitate service restoration by providing impact assessments of specific customer reconnection sequences, for instance.

4.2 ARIMA model formulation

A general ARIMA model can be expressed through the dependency of the current value from its past values and past error values

$$\phi_p(B) W_t = \theta_q(B) a_t \tag{4}$$

[1]For the remaining parts of the paper, we will only refer to heating loads to be the leading cause of CLPU in cold weather. The converse is applicable during heat waves for cooling loads.

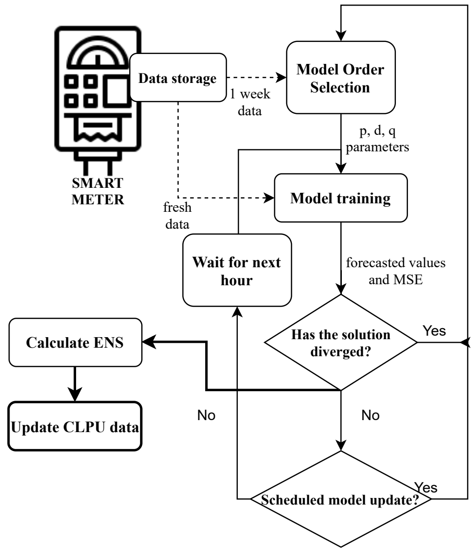

Figure 4: CLPU assessment algorithm overview

Here, a_t represents the residual model error at time t, while the stationary series W_t can be expressed through measurements of the variable X_t and the lag operator B

$$W_t = (1 - B)^d X_t \tag{5}$$

assuming that the input data series X_t is such that it can be reduced to stationarity by differencing it a finite number of times. The backshift operator B on the index of the time series is such that $BX_t = X_{t-1}$, $B^2X_t = X_{t-2}$ and so on

$$B^p X_t = X_{t-p} \tag{6}$$

In its extended form, an ARIMA model consists of the auto-regressive and moving average coefficients associated with the corresponding order of backshift operator B

$$\begin{gathered} \left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right) W_t = \\ \theta_0 + \left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right) a_t \end{gathered} \tag{7}$$

where the values of $\phi_1, \phi_2, \ldots, \phi_p$ and $\theta_0, \theta_1, \ldots, \theta_q$ represent the coefficients of the model that are estimated from the data. Hence, an ARIMA(p, d, q) model employs $p+q+1$ unknown parameters that are estimated from the data, where:

- p is the number of auto-regressive terms,
- d is the number of differences needed to make W_t stationary, and
- q is the number of lagged forecast errors in the prediction equation

An ARIMA model formulation including a potential control input, such as outdoors temperature here is presented in Appendix A. Lastly, as operating conditions evolve over days, weeks and seasons, the ARIMA model order should be re-estimated using fresh data from the smart meter. Considering the cyclicality of electricity demand, we argue that weekly model order re-estimation should adequate performance considering the computational effort required.

4.3 ARIMA model selection

Energy consumption time series are notoriously non-stationary—one can think of how the mean and the variance of the energy consumed varies from season to season. ARIMA models are tailored to address such a lack of stationarity in the mean sense, when compared to simpler AR and ARMA models, which cannot since they can only model purely stationary processes.

The major challenge when using ARIMA models is to select an adequate model order, namely: the number of auto-regressive terms (p), the number of moving average terms (q), and the order of differentiation of the time series (d).

The first parameter which has to be set is the order of differencing, d, which is determined from data stationarity. As mentioned previously, intuitively energy consumption time series are seasonal and do not display constant means over periods of time of a few hours. Therefore, they are non-stationary. In fact, one can apply standard unit root tests like the Augmented Dickey-Fuller (ADF) test to establish stationarity or lack thereof [22]. For processes displaying time-variable mean, it is common to apply differencing to remove its seasonal component. The order of differencing, d, sets the number of times the data has had earlier values subtracted from them. Hence, in determining d, one has to increase it from zero up to the point where the resulting differenced time series passes the ADF test [22].

The next step in the preliminary identification is to determine the values of p and q in the now mean-sense stationary time series based on the estimated time series autocorrelation function (ACF) and its partial autocorrelation function (PACF). From these, one can evaluate which combination of p and q can result in a minimumprediction error. The mathematical definitions of ACF and PACF are provided in Appendix B. For a given auto-regressive process order (p), an analysis of PACF is performed, while for the moving average process order (q), the ACF is calculated [23]. Based on the number of significant lags in ACF and PACF, the maximum values of lags for AR and MA processes are set.

The simplest approach in determining optimal values for the model order is to perform a full grid search over p and q. The most suitable model order is selected based on some of the standard criteria, such as mean squared error (MSE), root mean squared error (RMSE), Akaike information criterion (AIC) and others. However, this is a computationally expensive process. To overcome this challenge, a reduced grid search is developed in this paper. The reduced grid search is an approach that aims to reduce the search space of the ARIMA model order using the statistical properties of the data sample, such as ACF, PACF and unit root test results. Figure 5 shows the overall algorithm for model order selection.

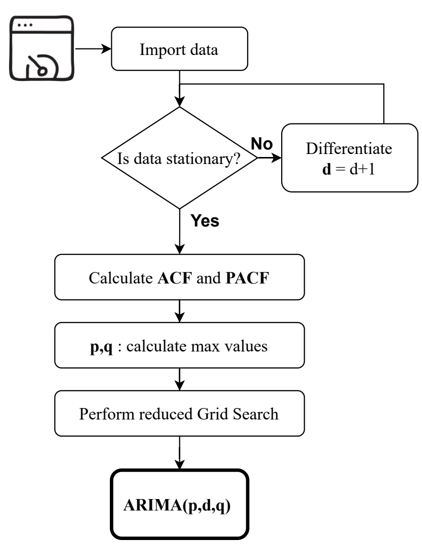

Figure 5: ARIMA model selection algorithm

Based on the number of lags of ACF and PACF functions that have values larger than a cut-off values (as defined in Appendix B), also called significant lags, maximum model orders p_{max} and q_{max}

are determined as a minimum between the calculated number of significant lags of PACF and ACF (n_{pacf} and n_{acf}) obtained from the criteria laid out in Appendix B and exogenous limiting values of the auto-regressive and moving average model orders ($p_{\lim}$ and $q_{\lim}$)

$$p_{\max} = \min(n_{pacf}, p_{\lim}) \tag{8}$$

$$q_{\max} = \min(n_{acf}, q_{\lim}) \tag{9}$$

The optimal combination of d, $p \in \{0, 1, \ldots, p_{\max}\}$, and $q \in \{0, 1, \ldots, q_{\max}\}$ parameters is selected based on the minimum value of MSE for a maximum expected outage duration

$$(p, q)_{opt} = \arg\min_{p,q} \mathrm{MSE} \tag{10}$$

5 Results and discussion

In this section, we illustrate how CLPU characteristics of individual dwellings are determined from local smart meter data. Here, we assess CLPU prediction abilities for outages of up to 12 hours in one-hour time steps.

5.1 Data set description

In this paper, actual real-world time-stamped dwelling-level measurements are used for verification of the proposed approach. The measured load data for residential homes were obtained from the UMassTraceRepository [24]. They are part of the *Smart* Home Data Set* which contains data for 114 single-family apartments for the period from 2014 to 2016 measured at 15-minute time intervals. Specifically, we have randomly selected 50 apartments from the 'Apartment' dataset.

5.2 System configuration

All calculations were implemented in Python, using approaches from [25], and ran on a computer with the following configuration: Intel i7 8700 CPU 3.19-GHz processor with 16.0 GB of RAM and a 64-bit operating system.

5.3 ARIMA model selection – Performance results

The execution performance of the proposed ARIMA model order selection algorithm is presented in this section. Figure 6 shows the value of the MSE for every residential customer for a full and reduced grid search, compared to random walk results. A random walk is defined as a process where the current value of a variable is composed of the past value plus an error term defined as a white noise (a normal variable with zero mean and variance one). The values are independent from one another and sorted in decreasing order of MSE for reduced grid search.

We can see that the reduced grid search approach has a higher error per customer than the full grid search, which is expected due to the reduced search space. However, we can also observe that the performance of the reduced grid search for the majority of the loads is close to the accuracy achieved by a full grid search. In Table 1, the statistics of an increase in values for reduced grid search compared to the full grid search are presented. It can be seen that for 50 apartments analyzed, the median mean square error increased by 22.54% from the full grid search approach when going with the reduced grid search.

The execution time comparison is presented in Figure 7. The reduced grid search has significantly outperformed the full grid search for all of the dwellings, which makes the reduced grid search more suitable for local, real-time execution.

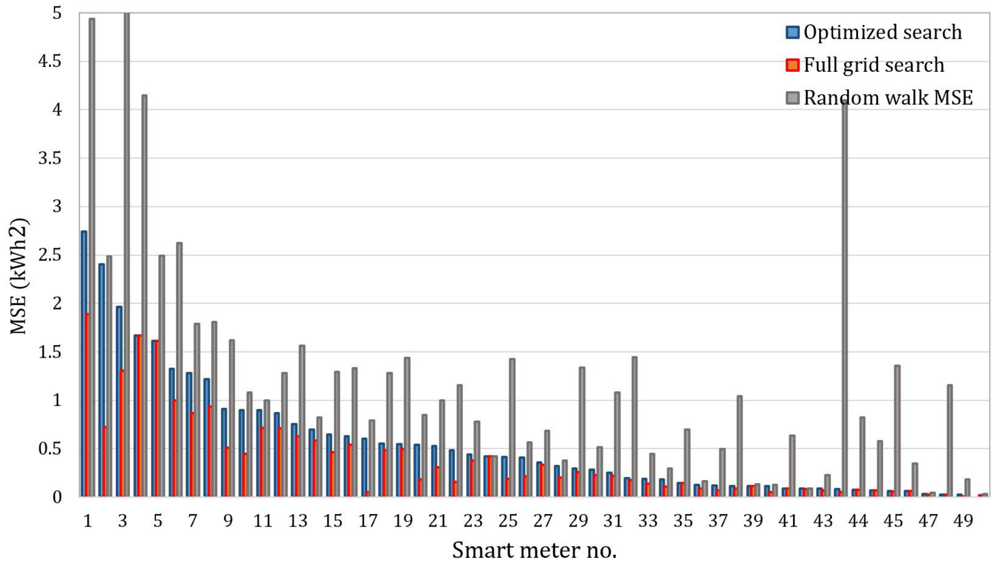

Figure 6: MSE for 50 apartments - random walk, reduced and full grid search

Table 1: MSE change between reduced grid search compared to full grid search (%)

Min	25%	50%	75%	Max	Average
0	9.84	22.54	70.25	1100	82.15

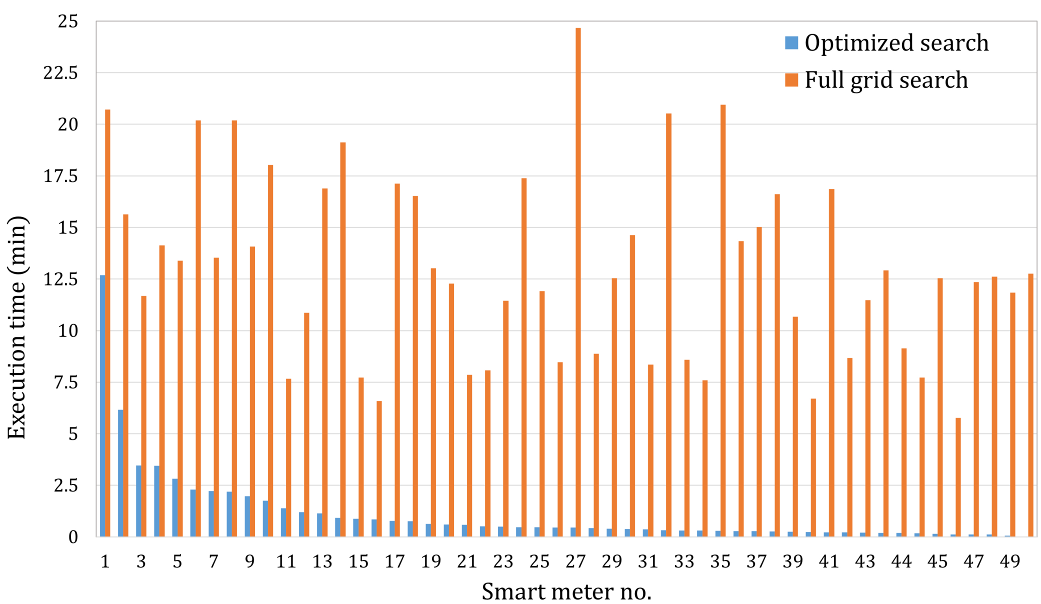

Figure 7: Execution time comparison for reduced and full grid search

5.4 Rolling window validation

To verify the robustness of the developed approach when exposed to raw and variable smart meter data, a rolling window forecast is executed for one-year worth's of data for the 50 apartments. By rolling window forecast, we consider the continuous execution of the proposed algorithm for one full year, where model parameters are updated hourly using fresh (unseen) data. With regularly updated model parameters, the CLPU model is predicted for the outages of up to 12 hours and MSE is recorded. Considering that MSE increases with outage duration, the 12-hour MSE is taken as the worst case. In each execution, the last week's most recent measurements are used. Considering the weekly periodicity of residential demand, the ARIMA model order (p, d and q) is revised on a weekly basis or in the case of the divergence of the MLE algorithm used to calculate parameters value. A weekly model update approach is compared to the one-year rolling window of ARIMA algorithm using the same, initially estimated, model order for the whole year. The average cumulative error comparison is shown in Figure 8. We can observe that the regular update of model order improves the accuracy significantly. The approach we propose can be executed independently for any unknown data set, without the need for user intervention. This is one of the most significant properties for enabling local smart meter CLPU estimation.

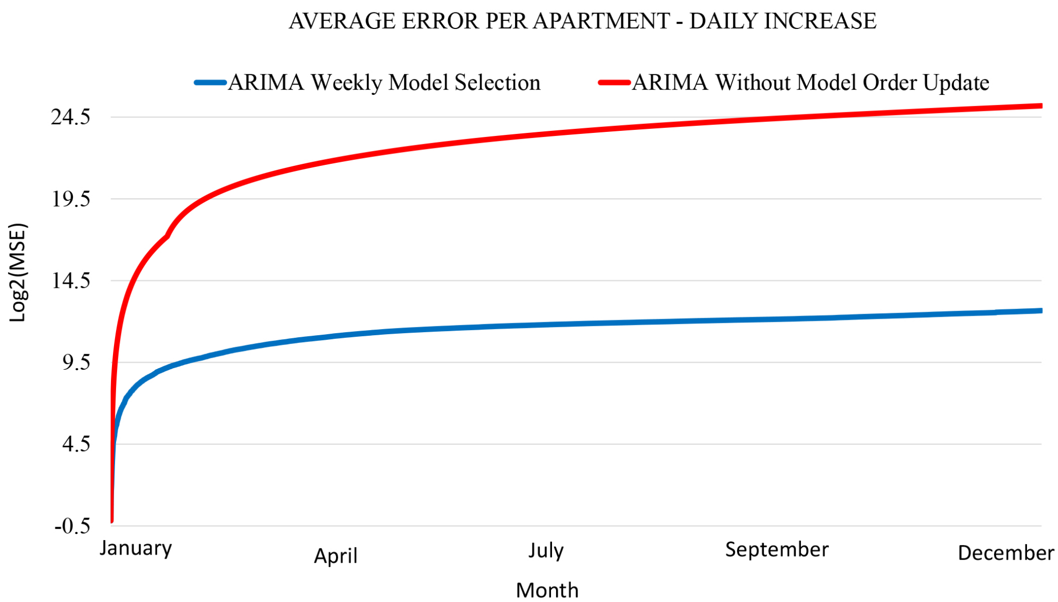

Figure 8: Comparison of ARIMA model performance using weekly model order update and ARIMA model performance with no model order update

The weekly model update shows significant improvement in terms of accuracy. Since the proposed model order selection has good performance in terms of execution time, the proposed approach can be run locally.

5.5 Comparison of performance for linear system with and without input control variable

The accuracy of CLPU estimation performance with added lagged temperature values as regressors as seen in (18) of Appendix A is compared against univariate ARIMA. For this purpose, two months of rolling window validation is performed for 15 sets of smart meter data. The results are presented in Figure 9 showing daily cumulative error.

The estimation with temperature as a control input has shown unsatisfactory performance in terms of convergence of MLE algorithm and execution time. The utilization of temperature values improves only slightly the accuracy of CLPU forecast. Moreover, the inclusion of p lagged temperature values increases significantly the complexity of the model. In addition, the procedure for ARIMA order selection would often diverge, even with an increased number of iterations. For example, a two-month rolling window model selection without the extra control input executes in less than an hour on average, while the ARIMA model order selection with lagged temperature values can take up to 12 hours. This

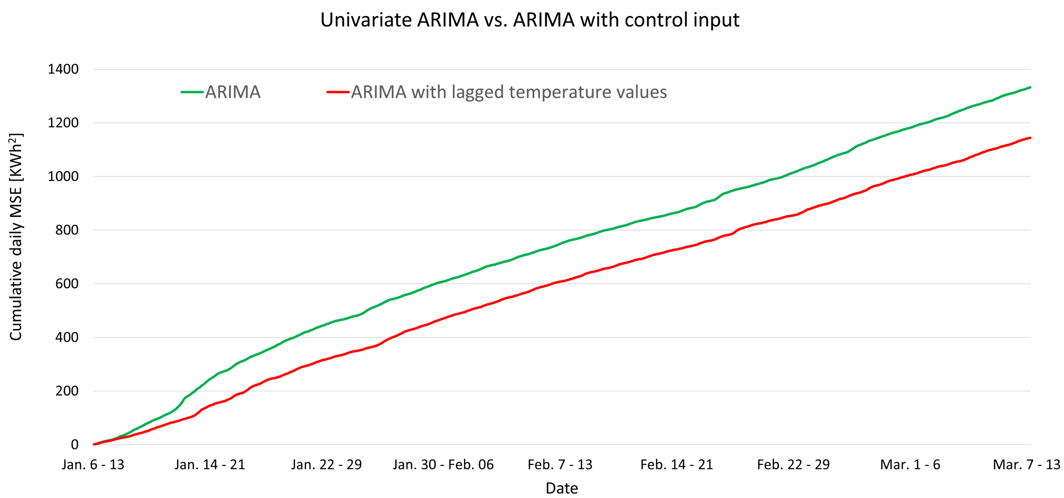

Figure 9: Rolling window comparison of the ARIMA with and without lagged temeprature values

is due to the fact that TCL energy consumption and lagged temperature values are essentially co-linear, which leads to singularity of the normal equations used to determine the weights (ϕ_i and θ_i) of the ARIMA model.

Figure 10 shows an example of calculation of the CLPU peak and duration using timestamped load measurements. Here, the outage starts at 9:00 AM, and we consider outage durations of one, two and three hours and the demand in non-outaged operating conditions.

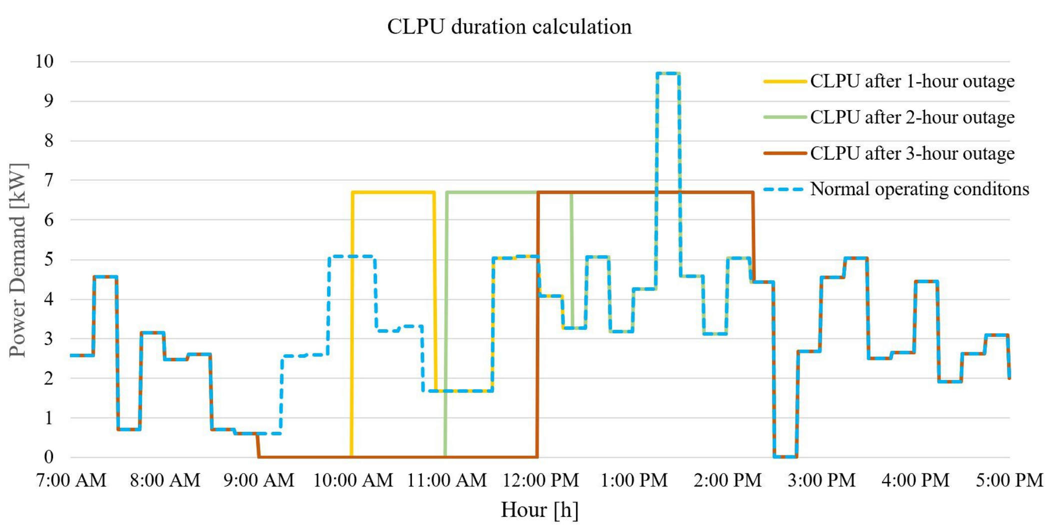

Figure 10: CLPU duration estimation results

5.6 CLPU duration verification against GridLab-D

The proposed CLPU estimation approach of a single customer is verified using GridLab-D [26], a distribution network modeling software, which can implement physical models of TCLs, as presented in Appendix C. An outage range of one to 10 hours is simulated for a customer with electrical space and water heating using temperature data for the Boston, MA area in January. Energy consumption measurements are simulated at 15-minute intervals for one week. These are then used to train the

ARIMA CLPU model. The peak power at which post-outage energy is consumed is calculated using auto-regression of the maximum power values for the previous seven days. The estimated peak value by our approach is 16.5 kW, while the average maximum power calculated by GridLab-D is 17.9 kW.

Figure 11 shows the CLPU duration calculated using the proposed CLPU estimation approach with 25%, 50% and 75% confidence interval (CI) of P_{CLPU} estimation, in comparison to GridLab-D CLPU duration. We can observe that the GridLad-D calculations closely match results using our proposal for short outage durations (about up to six hours), and later they stay within 50% CI of our estimator.

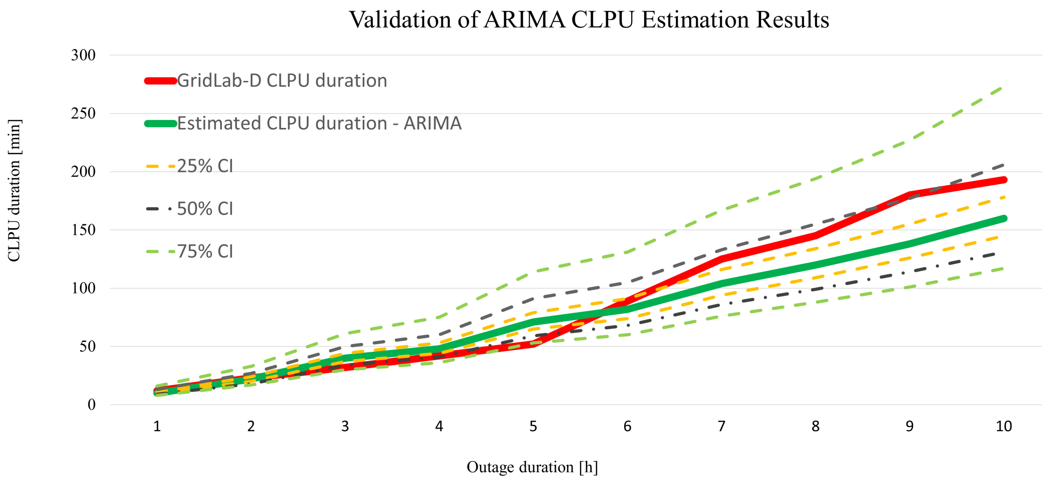

Figure 11: Estimated CLPU duration vs GridLab-D for residential home

6 Conclusion

In this paper, we have proposed an approach to estimate locally the CLPU power and duration of a single customer using its smart meter data. The duration of CLPU is estimated based on the foregone energy during the outage. The energy not served is estimated using an ARIMA approach, where only timestamped load measurements are used. The ARIMA model order is determined using a reduced grid search. We have shown that the reduced grid search results in shorter execution time of model selection while maintaining the reasonable accuracy. Since the proposed framework is designated for local execution, the execution time and the data storage are very important. We have observed that using temperature values significantly increases the execution time and model complexity, and, therefore, should not be used.

This work is foundational to performing adequate CLPU management during distribution system restoration in cold (and also in hot) climates. With simple and low-computation customer-per-customer short-term energy and peak power predictions, the proposed solution can be implemented along with customer-driven service restoration, where single-customer CLPU values can play a role in restoration time optimization and post-outage feeder overload prevention.

A ARIMA with lagged temperature values as an input

The ARIMA model can be extended to include an external control variable. In the context of this paper, lagged temperature values can be used as regressors with the objective of improving model performance considering how TCL-based demand is temperature-dependent. In this case, the ARIMA model is extended with $p+1$ input variables, of which p are lagged temperature values and the current temperature at time t

$$X_t = a_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i a_{t-i} - \sum_{i=0}^{p} \beta_i u_{t-i} \tag{A.1}$$

In the event where the input data X_t is not stationary, the data series should be differenced as in (5) to obtain a stationary W_t, and thus

$$\phi_p(B) W_t = \theta_q(B) a_t - \beta_p(B) u_t \tag{A.2}$$

where u_t represents the temperature at time t.

B Analysis of the ACF and PACF for the given data sample

After a time series has been stationarized by differencing, the next step in fitting an ARIMA model is to determine how many AR or MA terms are needed to correct any autocorrelation that remains in the differenced series. By looking at the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the differenced series, one can tentatively identify the numbers of AR and/or MA terms needed.

For an observed series $X_1, X_2, \ldots, X_T$, we denote the sample mean by $\bar{X}$. The sample lag-h autocorrelation is given by

$$\hat{\rho}_h = \frac{\sum_{t=h+1}^{T} \left(X_t - \bar{X}\right)\left(X_{t-h} - \bar{X}\right)}{\sum_{t=1}^{T} \left(X_t - \bar{X}\right)^2} \tag{A.3}$$

and the lag-h PACF is

$$\hat{\psi}_h = \mathrm{Cov}(X_t, X_{t-h} | X_{t-1}, \ldots, X_{t-h+1}) \cdot [\mathrm{Var}(X_t | X_{t-1}, \ldots, X_{t-h+1}) \cdot \mathrm{Var}(X_{t-h} | X_{t-1}, \ldots, X_{t-h+1})]^{-1/2}$$

The cut-off values, from which the limit for the model order is set, are calculated from confidence intervals for the ACF and the PACF determined from the training data sample. The MA order q is determined from the ACF function confidence intervals (CI) for the q^{th} lag

$$CI(q) = \pm 1.96 \sqrt{\frac{1}{T}\left(1 + 2\sum_{i=1}^{q} \hat{\rho}_i^2\right)} \tag{A.4}$$

where the 1.96 factor is an approximate value of the quantile of the normal distribution for typical values of 95% confidence interval. If the ACF of the time series is less than than $CI(q+1)$, then the order of the MA portion of the model should be at most q.

On the other hand, the AR order p is determined from the PACF function. Confidence intervals for PACF are independent of p

$$CI(p) = \pm 1.96 / \sqrt{T} \tag{A.5}$$

If the PACF of the time series is less than than CI for some $p+1$, then the order of the AR portion of the model should be at most p.

C GridLab-D load model

Residential load in GridLab-D is modeled through the "house" class that represents a thermodynamical load with integrated heating and cooling models. The load is modelled as a combination of a ZIP (constant impedance-current-power) load, heat pump and water heater. In GridLab-D, the thermal flow of a residential home is calculated using the Equivalent Thermal Parameter (ETP) model [26], as shown in Figure A.1. Given at a given time t, the differential equations for the indoor air temperature T_A and the building mass temperature T_M are

$$0 = Q_A - U_A\,(T_A - T_O) - H_M\,(T_A - T_M) - C_A \frac{d}{dt} T_A \tag{A.6}$$

$$0 = Q_M - H_M\,(T_M - T_A) - C_M \frac{d}{dt} T_M \tag{A.7}$$

Here, the constant values, that depend on the physical characteristics of the house, are: T_O is the outdoor air temperature, Q_A is the heat added to the indoor air, Q_M is the heat added to the building mass, U_A is the building envelope conductivity to the indoor air, C_A is the heat capacity of the indoor air, H_M is the building mass conductivity to the indoor air, and C_M is the heat capacity of the building mass.

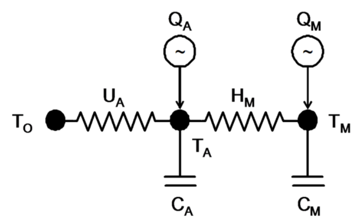

Figure A.1: Equivalent Thermal Parameter model

References

[1] S. Dzeletovic, F. Bouffard, and G. Joos, Reduction of CLPU overload through active load management, 2019 CIGRE Canada Conference, 2019.

[2] S. Lefebvre and C. Desbiens, Residential load modeling for predicting distribution transformer load behavior, feeder load and cold load pickup, International Journal of Electrical Power and Energy Systems, 24(4):285–293, 2002.

[3] S. Dzeletovic, F. Bouffard, and G. Joos, Service restoration with active cold load pick-up management, in 2019 IEEE Power Energy Society General Meeting (PESGM), Aug 2019, pp. 1–5.

[4] S. Ihara and F. C. Schweppe, Physically based modeling of cold load pickup, IEEE Transactions on Power Apparatus and Systems, PAS-100(9):4142–4150, Sep. 1981.

[5] K. P. Schneider, E. Sortomme, S. S. Venkata, M. T. Miller, and L. Ponder, Evaluating the magnitude and duration of cold load pick-up on residential distribution feeders using multi-state load models, IEEE Transactions on Power Systems, 31(5):3765–3774, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7776868/

[6] J. Aubin, R. Bergeron, and R. Morin, Distribution transformer overloading capability under cold-load pickup conditions, IEEE Transactions on Power Delivery, 5(4):1883–1891, Oct 1990.

[7] W. Zhang, K. Kalsi, J. Fuller, M. Elizondo, and D. Chassin, Aggregate model for heterogeneous thermostatically controlled loads with demand response, in 2012 IEEE Power and Energy Society General Meeting, July 2012, pp. 1–8.

[8] E. Agneholm and J. Daalder, Cold load pick-up of residential load, IEE Proceedings - Generation, Transmission and Distribution, 147(1):44–50, Jan 2000.

[9] H. Qu and Y. Liu, General model for determining maximum restorable load, in 2012 IEEE Power and Energy Society General Meeting, July 2012, pp. 1–6.

[10] F. Bu, K. Dehghanpour, Z. Wang, and Y. Yuan, A data-driven framework for assessing cold load pick-up demand in service restoration, IEEE Transactions on Power Systems, 34(6):4739–4750, Nov 2019.

[11] S. Dzeletovic, F. Bouffard, H. Michalska, and G. Joos, Regression-based wintertime energy consumption prediction for cold load pick-up management, in 2020 IEEE Power Energy Society General Meeting (PESGM), Aug 2020, pp. 1–5.

[12] S. Arora and J. W. Taylor, Short-term forecasting of anomalous load using rule-based triple seasonal methods, IEEE Transactions on Power Systems,28(3):3235–3242, Aug 2013.

[13] J. W. Taylor, Short-term load forecasting with exponentially weighted methods, IEEE Transactions on Power Systems, 27(1):458–464, 2012.

[14] N. Amjady, Short-term hourly load forecasting using time-series modeling with peak load estimation capability, IEEE Transactions on power systems, 16(3):498–505, 2001.

[15] T. Hong and S. Fan, Probabilistic electric load forecasting: A tutorial review, International Journal of Forecasting, 32(3):914–938, 2016.

[16] R. Bonetto and M. Rossi, Machine learning approaches to energy consumption forecasting in households, arXiv preprint arXiv:1706.09648, 2017.

[17] P. Lusis, K. R. Khalilpour, L. Andrew, and A. Liebman, Short-term residential load forecasting: Impact of calendar effects and forecast granularity, Applied Energy, 205:654–669, 2017.

[18] C. Nichiforov, I. Stamatescu, I. Făgărăşan, and G. Stamatescu, Energy consumption forecasting using arima and neural network models, in 2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE), 2017, pp. 1–4.

[19] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, Statistical and machine learning forecasting methods: Concerns and ways forward, PloS one, 13(3):e0194889, 2018.

[20] A. Meyler, G. Kenny, and T. Quinn, Forecasting irish inflation using arima models, 1998.

[21] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, Comparison of arima and artificial neural networks models for stock price prediction, Journal of Applied Mathematics, vol. 2014, 2014.

[22] J. S. Armstrong, Principles of forecasting: a handbook for researchers and practitioners. Springer Science & Business Media, 2001, vol. 30.

[23] A. Nielsen, Practical Time Series Analysis: Prediction with Statistics and Machine Learning. O'Reilly Media, Inc., 2019.

[24] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, Smart*: An open data set and tools for enabling research in sustainable homes.

[25] S. Seabold and J. Perktold, statsmodels: Econometric and statistical modeling with python, in 9th Python in Science Conference, 2010.

[26] D. P. Chassin and J. C. Fuller, Etp closed form solution, http://gridlab-d.shoutwiki.com/wiki/ETP_closed_form_solution, 2013.