USING SOCIOECONOMIC DATA TO PREDICT MULTI-FAMILY RESIDENTIAL ELECTRICITY CONSUMPTION

By

Grace E. Wackerman

Senior Thesis in Electrical Engineering

University of Illinois at Urbana-Champaign

Advisor: Ashlynn S. Stillwell

May 2020

Abstract

Electricity supply responds to changes in demand, and changing populations alter energy demands for an area. This project characterizes how different population compositions affect electricity consumption using Commonwealth Edison (ComEd) anonymized meter-level data, which show the electricity usage at 30-minute intervals in 2016 for the whole service area, sorted by zip code. The following tasks were completed:

- Compare multi-family residences with different population densities and median incomes in Chicago.
- Characterize different electricity profiles for different zip codes using mean electricity usage for an average day in each month for each zip code.
- Predict multi-family electricity consumption as a function of zip-code-level socioeconomic predictor variables using linear regression.

This analysis shows that median age of home, mean commute time, percent of multi-family housing units, median age of population, and percent female are statistically significant predictors of multi-family residential electricity consumption. Daily and monthly electricity profiles also vary notably across zip codes in Chicago. These results can inform electricity providers regarding how forecasted changes in population will likely affect the electricity demand of a particular area.

Subject Keywords: smart meters; socioeconomic; electricity usage; multi-family housing

Acknowledgments

This work was supported by the Department of Civil and Environmental Engineering and the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. Commonwealth Edison electricity data were obtained from the Environmental Defense Fund, under a data sharing agreement with the University of Illinois.

Contents

1. Introduction

Variances in time and weather have been studied extensively to predict household electricity consumption, but research is much more limited on how social and economic factors affect electricity consumption in multi-family residential households. With the widespread installation of smart meters, there are more data available across a larger geographic area, supporting the study of geography and demographic effects on electricity consumption.

This work examined Commonwealth Edison (ComEd) smart meter data from Chicago in conjunction with data collected from the U.S. Census for 2016, comparing differences across zip codes to reveal electricity consumption trends for multi-family housing units related to socioeconomic traits. The scope was restricted to zip codes in the City of Chicago because of the uniformity in local governance as well as the wide social variation of neighborhoods and zip codes within the study area.

This research analyzes socioeconomic characteristics of zip codes and how they can be used to predict electricity consumption for multi-family homes in Chicago. The research goals were to (1) visualize and compare electricity demand profiles for multi-family units, and (2) predict electricity consumption in multi-family housing as a function of socioeconomic data.

2. Literature Review

2.1 Smart Meter Data Analysis

Previous studies have leveraged smart electricity meters to better understand predictors of customer consumption and acceptance of smart meter infrastructure. When using smart meter data to examine climate, occupant, and housing impacts on residential electricity consumption, Kavousian et al. [27] found that the number of occupants, high consumption appliances, weather, location, floor area, number of refrigerators, entertainment devices, and pet ownership had significant correlation with the electricity consumption profiles of the households. However, income, home ownership, and building age had no significant correlation [27]. Zip code variation explains up to 46% of electricity consumption variability [27]. A study of the willingness-to-pay for smart meters among household customers in Germany found that trust in data protection, intention to change usage behaviors, and usefulness of consumption feedback were the most important factors, in that order [16].

2.2 Factors Related to Residential Electricity Consumption

The residential sector, specifically the multi-family housing sector, is a relatively under-studied aspect of building energy consumption. When reviewing data-driven energy consumption prediction studies related to buildings, Amasyali and El-Gohary found that while 81% of the reviewed research efforts focused on developing energy consumption prediction models for commercial and/or educational buildings, only 19% focused on residential buildings [4]. Energy-efficient features are adopted much less regularly in rented units than in owned housing units, with rates of adoption ranging from 5.3% to 21.6%, in a study of 10 U.S. cities [23]. Chicago has a comparably high rate of energy efficiency adoption in rental units, but the associated percent increase in rent for units with energy efficiency measures was also among the highest in the study [23]. Kiefer and Krentz found that while multi-family water use (which is closely related to electricity consumption) varies with respect to water pricing and income, it varies less than in single-family homes, and climate differences affect single-family use patterns more than multi-family use patterns [28]. Residential electricity consumption patterns also show seasonality, with weather effects more pronounced in the winter than in the summer [20].

2.3 Socioeconomic Predictors of Electricity Consumption

Various socioeconomic factors affect residential electricity consumption, with statistical significance depending on context. A study of 189 Dutch households found that energy use was related to sociodemographic variables, while changes in energy use were often related to psychological variables [1]. A later study of Dutch households found that households with children or elderly tended to consume more energy than other households [8]. Similarly, electricity load profiles in Europe showed strong dependence on household size, net income, age of reference person, and employment status [21]. Electricity consumption has also been studied on a U.S. zip code level to determine the impact of large-scale electric vehicle adoption [3]. Elnakat and Gomez found that there was an 80% higher per capita energy consumption in female-dominated households compared to male-dominated households, with twice the natural gas consumption in the former [12]. When studying the socioeconomic, demographic, and gendered influences on a household's energy consumption at the zip code level in San Antonio, Texas, Elnakat et al. found higher energy use to be associated with zip codes that were female-dominant, with a median age over 40, and with higher levels of income and education [13]. The study also found that renters tended to use less energy than home owners [13]. Similarly, Karatasou et al. found that households with more occupants, living in older and less insulated buildings, with greater floor area and electric water heating were more likely to be high energy consumers [26].

While data available on a zip code level can be helpful for analysis, there are some potential issues. Zip codes can include very different geographic ranges depending on whether the area is rural or urban, and may also consist of two discontinuous areas. Information grouped by census block can be better for analysis of demographic and socioeconomic data [18]. Socioeconomic factors have been studied for their effect on CO_2 emissions in Iran, and were evaluated using a multivariate statistical analysis [19]. However, Harris and Liu found that income did not have a statistically significant effect on U.S. residential electricity consumption for the period 1969-1990 [20].

2.4 Electricity Consumption Prediction Models

Several quantitative approaches have been used to describe electricity consumption variability with location and time. Cluster analysis, multivariate linear regression, or support vector machines combined with classification based on the surveys can provide a more detailed analysis of household load profiles [11, 17, 22, 29, 38]. Relationships between income and energy consumption have been studied using regression [2], and clustering techniques have been used to group customers for electrical load pattern analysis [10, 36]. In one study, a hierarchical clustering method was used to detect resident profiles, finding statistical significance for months of the year, working versus weekend days, hours of the day, temperature, and baseline energy consumption [2]. A study of Danish households' hourly electricity consumption predicted that the peak electricity consumption for households will likely increase significantly for workdays in January 2030 [2]. A comparison study of neural network, conditional demand analysis, and engineering model approaches determined that while all three methods can be used, a neural network model was best for end-use energy consumption modeling [6]. Appliance, lighting, and cooling energy consumption reached a minimum and stayed constant as household income decreased, while space heating electricity consumption increased linearly, further increasing with additional people in the household [6]. Another end-use simulation/forecasting model combined load data with survey results to estimate the total residential load curve in New South Wales, Australia, including considerations for monthly and daily variations as well as weather dependencies [7]. Parti and Parti similarly obtained detailed household level data for 5,286 households in San Diego County via a mail questionnaire, using the results to determine a conditional demand framework to disaggregate household demand into 16 appliance categories [32].

Traditional methods for energy demand forecasting for demand side management include time series, regression, and econometric modeling, while soft techniques such as fuzzy logic, genetic algorithms, and neural networks have also been used [34]. Upgrade options for reducing residential energy consumption in Canada were evaluated using an end-use electricity consumption model, which found that upgrading appliances would lead to significant savings [14]. Different statistical analysis methods have been used to estimate residential end-use load curves, such as conditional demand analysis, seemingly unrelated regressions, and the random coefficient model [15], while others use artificial intelligence for load forecast models and research gaps [33]. Despite these advances in modeling electricity demands and load patterns, the multi-family residential housing sector remains as an understudied aspect of residential electricity consumption. This work aims to fill that knowledge gap for the multi-family residential housing sector in Chicago, Illinois, analyzing both electricity load profiles and socioeconomic factors related to electricity consumption.

3. Methodology

3.1 Electricity Consumption Across Zip Codes

Commonwealth Edison (ComEd) has installed smart electricity meters throughout its service area in northern Illinois. Through a data sharing agreement with the Environmental Defense Fund, this work leveraged 30-minute resolution anonymized smart meter data for 2016, within the ComEd service area. Each smart meter and its corresponding data were reported by zip code, including a unique account identifier and the delivery service type, designating multi-family or single-family residential and electric or non-electric space heating.

With over 500 zip codes served by ComEd, this work focused on zip codes in the City of Chicago as a large, diverse city with a range of housing options and distinct neighborhoods with reported demographic information. Multi-family households are much less studied in terms of electricity consumption than single-family homes. With an increasing global population, as well as an increase in urbanization of city centers, the understanding of load profiles of high density and multi-family housing becomes even more essential.

The analysis of ComEd smart meter data for multi-family residential electricity consumption used the following approach, implemented in the open-source Python programming language:

1. Extract smart meter data for multi-family units in Chicago zip codes from a collection of comma separated value files with electricity consumption data organized by month and by zip code.
2. Average the daily consumption profiles in each month to create a multi-family residential electricity consumption profile for an average day in each month.

3. Create a box plot for each month and zip code pair containing half-hourly average electricity consumption for a typical day in the month for each of the multi-family smart meters within the zip code for that month.
4. For each month, plot the average half-hourly consumption for the typical day in that month, comparing the averages from each zip code in the same data visualization.
5. Record the daily average multi-family residential electricity consumption profile for each zip code and month pair to be used for linear regression analysis.

3.2 Socioeconomic Data for each Zip Code

To study the real-world impact of demographic and socioeconomic variables on electricity consumption, it is necessary to have robust data about both electricity consumption and demographic variables for the same geographic location. With anonymized smart meters, it is impossible to determine the characteristics of the households of each individual smart meter. However, the American Community Survey 5-Year Estimate collects demographic information on a zip code level [35]. With the smart meters labeled by their respective zip codes, it is possible to analyze the smart meter data in conjunction with collected demographic and socioeconomic data to estimate correlations and create a predictive model. The following data from the American Community Survey 5-Year Estimates [35] were collected for each zip code in Chicago:

- **Demographics**
    - Population
    - Median age
    - Percent 65 and over
    - Percent female

  - Households with children under 6
  - Households with children 6 to 17
- **Income**
  - Median household income
  - Mean household income
  - Percent under poverty line
  - Unemployment percent
- **Education**
  - Percent with a high school degree
  - Percent with a bachelor's degree
- **Housing and Location**
  - Median age of home
  - Percent of multi-family housing units
  - Percent occupancy
  - Total housing units
  - Mean commute time

These socioeconomic data were selected as possible predictors based on previous literature [6, 8, 12, 17, 21, 27].

3.3 Annual Electricity Consumption Prediction Model

With the developed multi-family residential electricity profiles and the corresponding socioeconomic indicators for each zip code, a regression model was created to evaluate socioeconomic indicators as possible predictors of multi-family residential electricity consumption. Since socioeconomic data were available on an annual timestep, a multi-family residential electricity profile for each zip code for an average day in the year was used with linear regression to reveal the best socioeconomic indicators of electricity consumption in multi-family homes. The best-fit model was determined through ordinary least squares (OLS) regression using linear modeling in the open-source R statistical computing software, using the form listed in Equation 1.

$$\mathrm{Y}_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \cdots \beta_i X_i \tag{1}$$

for multi-family residential electricity consumption Y_i and socioeconomic predictor variables X_i. To verify the accuracy of the model, the model OLS assumptions of normality and constant variance of residuals, and independence of predictor variables and residuals were evaluated. The normality and constant variance of the residuals were quantified using the Shapiro-Wilk normality test and Tukey test, respectively. Independence of predictor variables was determined by evaluating multicollinearity and the Durbin-Watson test for autocorrelation. Diagnostics were performed on the statistical significance of coefficient estimates and the model form, including estimation of the goodness-of-fit R^2 statistic.

4. Results

4.1 Electricity Profile Graphical Analysis

Through visual analysis of monthly average daily electricity consumption for each zip code, high variances were observed between months within each zip code. High variances also exist between zip codes within each month. Figures 1(a) and 1(b) illustrate the daily averages across zip codes by month, showing variance both temporally and spatially.

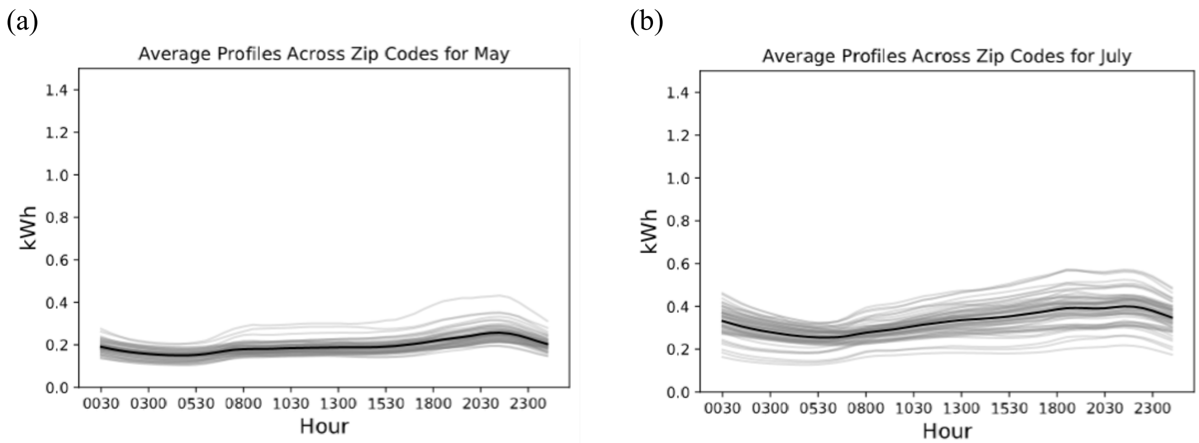

Figure 1. Multi-family residential electricity load varies with time of day across the analyzed zip codes for (a) May, and (b) July. The average profiles are represented as gray lines for each zip code, with the black line reflecting the average across zip codes.

During summer months, there was higher electricity consumption compared to other months. Winter months also had high electricity consumption, while spring and fall months had the lowest electricity consumption. Multi-family residential electricity load profiles for all months are shown in the Appendix for all zip codes analyzed, shown in Figure 2. Many of the daily electricity profiles had peaks in the evening, while earlier in the day electricity consumption was lower. Figures 3(a) and 3(b) illustrate the differences between zip codes with differing median age. The locations of the zip codes in Figures 3(a) and 3(b) are shaded in orange in Figure 2.

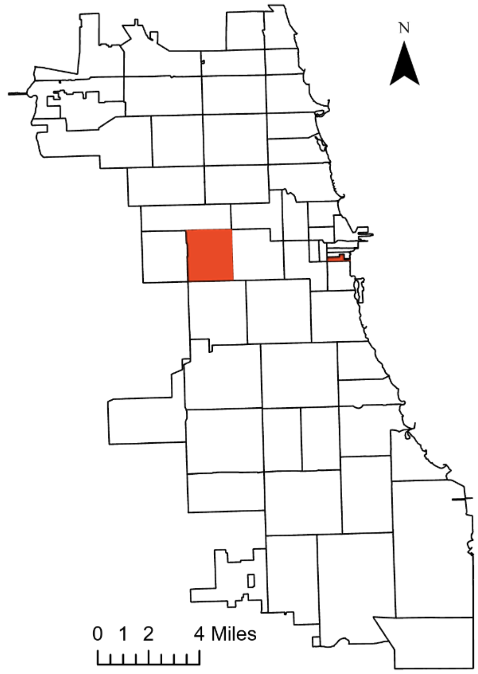

Figure 2. Zip codes (n = 56) in the City of Chicago were included in the analysis based on high amounts of multi-family residential housing. Garfield Park and The Loop neighborhoods, depicted in orange, represented differing load profiles, as shown in Figures 3(a) and 3(b).

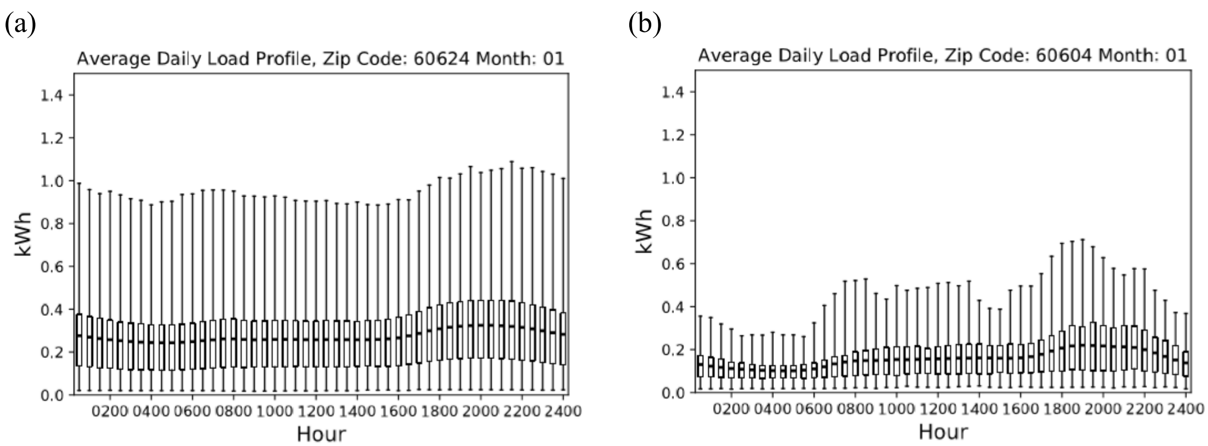

Figure 3. Daily load profiles vary for multi-family residential housing across zip codes, shown for January 2016, in (a) Garfield Park, and (b) The Loop.

Garfield Park had a much lower median age than The Loop, at 28.7 years compared to 44.6 years, highlighting a potential impact of socioeconomic indicators on multi-family residential electricity consumption, shown in Figure 3 for these zip codes. A multiple linear regression model was created to further investigate the ability of socioeconomic indicators to explain variation in multi-family residential electricity consumption in Chicago.

4.2 Multiple Linear Regression Analysis

Using an ordinary least squares modeling approach, a best-fit multiple linear regression model was created following the form of Equation 3.1, using the socioeconomic data for Chicago zip codes as indicators. The OLS approach used backwards stepwise regression to create a best-fit model of statistically-significant predictor variables. The coefficient estimates shown in Table 1 reflect the multiple linear regression model, rounded to two significant figures, explaining approximately 41% of the variability in multi-family residential electricity consumption in Chicago. Table 1 also includes measures of the statistical significance of each coefficient and the model form.

Table 1. Socioeconomic indicators were statistically significant predictors of multi-family residential electricity consumption for Chicago zip codes, explaining approximately 41% of the variability.

Factor	Coefficient	Estimate	Standard Error	*t*-Value	Pr(>|*t*|)
Constant	β_0	18	3.6	5.2	4.4e-6
Median Age of Home	β_1	0.029	0.012	2.3	0.024
Percent Multi-Family	β_2	-1.9	0.86	-2.2	0.031
Percent Female	β_3	0.14	0.071	1.9	0.059
Median Age	β_4	-0.28	0.046	-6.1	1.4e-7
Mean Commute Time	β_5	-0.14	0.051	-2.8	0.0082
Multiple R^2 = 0.47; Adjusted R^2 = 0.41; *F*-statistic = 8.8 (*p*-value = 4.2e-6)					

OLS assumptions of normality and constant variance of residuals, and independence of residuals and predictor variables were evaluated statistically. The normality of the residuals was measured using the Shapiro-Wilk normality test, which is based on hypothesis tests:

H_0: Residuals are normally distributed.

H_A: Residuals are not normally distributed.

The Shapiro-Wilk test provided a value of 0.97, with a p-value of 0.26, with the decision of fail to reject the null hypothesis, confirming the assumption of normally distributed residuals. Constant variance of the residuals was quantified using the Tukey test, based on the following hypotheses:

H_0: The quadratic term in a residual trend line is zero (e.g., $ax^2 + bx + c; a = 0$).

H_A: The quadratic term in a residual trend line is not zero (e.g., $a \neq 0$).

Based on the Tukey test statistic of 0.11, with a p-value of 0.91, the hypothesis test decision is to fail to reject the null hypothesis, confirming the assumption of constant variance across residuals, shown graphically in Figure 4.

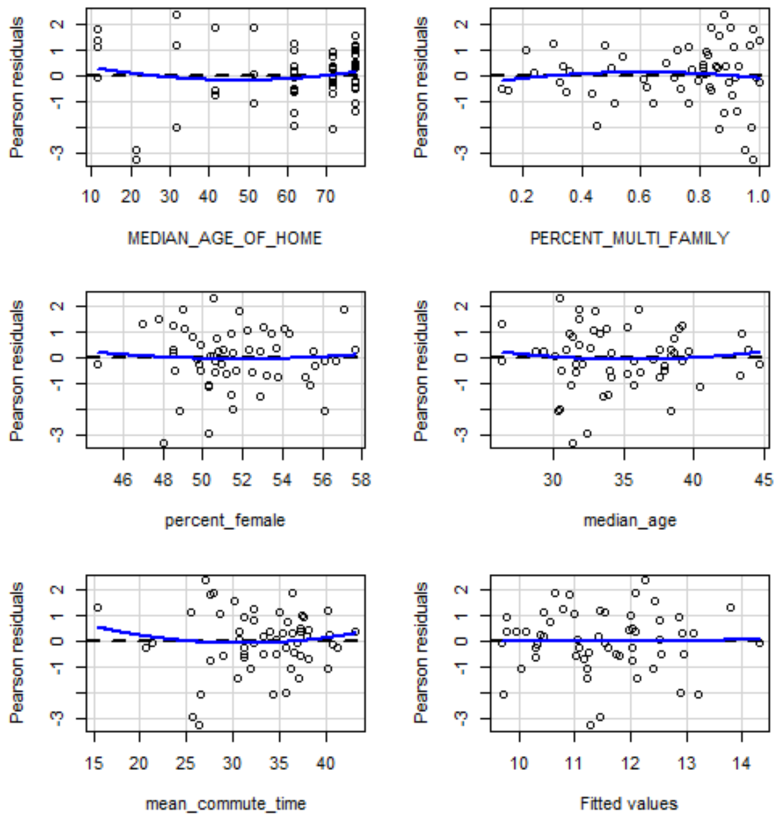

Figure 4. Pearson residuals of the multiple linear regression model satisfy the ordinary least squares regression assumptions of normality, constant variance, and independence.

Independence assumptions were evaluated for both the residuals and the socioeconomic predictor variables. To assess independence of the residuals, shown in Figure 4, the Durbin-Watson statistic was calculated to quantify autocorrelation, based on the following hypotheses:

H_0: Autocorrelation is zero.

H_A: Autocorrelation is not zero.

For the calculated Durbin-Watson statistic of 1.95, with a p-value of 0.75, the hypothesis test decision is to fail to reject the null hypothesis, confirming the assumption of independence of the residuals. Independence of the socioeconomic predictor variables was determined by estimating variance inflation factors (*VIF*) as a measure of multicollinearity. *VIF* > 10 is an indication of correlation between predictor variables, violating assumptions of independence. Table 2 summarizes the *VIF* values for the socioeconomic indicators, confirming independence among the predictor variables.

Table 2. The socioeconomic indicators did not exhibit multicollinearity as all *VIF* values were less than 10.

Predictor Variable	Variance Inflation Factor
Median Age of Home	2.32
Percent Multi-Family	1.69
Percent Female	1.40
Median Age	1.43
Mean Commute Time	2.97

With OLS assumptions verified, the best-fit model of annual multi-family residential electricity consumption (Y_i), in kilowatt-hours, was the model shown in Equation 2:

$$\begin{aligned} Y_i = 18 &+ 0.029(\text{Median age of home}) - 1.9(\text{Percent multi-family}) \\ &+ 0.14(\text{Percent female}) - 0.28(\text{Median age}) \\ &- 0.14(\text{Mean commute time}) \end{aligned} \tag{2}$$

Based on these results, higher multi-family residential electricity consumption is associated with older homes, lower percent multi-family housing, higher percent female populations, younger median age, and shorter mean commute time, based on data at a zip code level.

Equation 2, however, only explains about 41% of the variability in multi-family residential electricity consumption across zip codes. Many other factors, including personal behaviors and seasonal effects, influence residential electricity consumption, as reported in literature [2, 7, 20].

5. Conclusion

Data from Commonwealth Edison (ComEd) smart electricity meters in Chicago were analyzed in conjunction with data collected from the U.S. Census for 2016, to compare differences in multi-family residential electricity consumption across zip codes. A statistically significant multiple linear regression model was created to predict annual electricity consumption in the multi-family housing sector using socioeconomic characteristics as indicator variables. Even for the restricted scope of zip codes in the City of Chicago, sufficient social variation existed across zip codes and multi-family electricity consumption to visualize differences in daily load profiles and create a statistically-relevant model.

The results show that there are differences in electricity load profiles across zip codes and those differences are present with time of day and season of the year. There are also many drivers of multi-family residential electricity consumption in Chicago. This research demonstrates that socioeconomic characteristics of zip codes can be used to predict electricity consumption for multi-family homes in Chicago using a multiple linear regression model. This model was found to be robust, with the following socioeconomic variables as statistically-significant predictors: median age of home, percent of multi-family housing units, percent female, median age of population, and mean commute time.

While using zip code-level socioeconomic data can give an estimate of the demographics of the data sample in aggregate, these data do not reflect individual households as the electricity meter data are anonymized. Socioeconomic information is also estimated on an annual scale, such that there are limitations in using U.S. Census data to understand sub-annual variation in multi-family residential electricity consumption. Smart meter data are limited to electricity consumption such that additional research is needed to examine both electricity and natural gas consumption to gain a more holistic perspective on home energy consumption.

References

[1] Abrahamse, W., & Steg, L. (2009). How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? *Journal of Economic Psychology*, *30*(5), 711–720. doi: 10.1016/j.joep.2009.05.006

[2] Abreu, P. H., Silva, D. C., Amaro, H., & Magalhães, R. (2016). Identification of Residential Energy Consumption Behaviors. *Journal of Energy Engineering*, *142*(4), 04016005. doi: 10.1061/(asce)ey.1943-7897.0000340

[3] Ahkamiraad, A., & Wang, Y. (2018). An Agent-Based Model for Zip-Code Level Diffusion of Electric Vehicles and Electricity Consumption in New York City. *Energies*, *11*(3), 640. doi: 10.3390/en11030640

[4] Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, *81*, 1192–1205. doi: 10.1016/j.rser.2017.04.095

[5] Andersen, F. M., Baldini, M., Hansen, L. G., & Jensen, C. L. (2017). Households' hourly electricity consumption and peak demand in Denmark. *Applied Energy*, *208*, 607–619. doi: 10.1016/j.apenergy.2017.09.094

[6] Aydinalp-Koksal, M., & Ugursal, V. I. (2008). Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. *Applied Energy*, *85*(4), 271–296. doi: 10.1016/j.apenergy.2006.09.012

[7] Bartels, R., Fiebig, D. G., Garben, M., & Lumsdaine, R. (1992). An end-use electricity load simulation model. *Utilities Policy*, *2*(1), 71–82. doi: 10.1016/0957-1787(92)90055-n

[8] Brounen, D., Kok, N., & Quigley, J. M. (2012). Residential energy use and conservation: Economics and demographics. *European Economic Review*, *56*(5), 931–945. doi: 10.1016/j.euroecorev.2012.02.007

[9] Census profile: Chicago, IL. (n.d.). Retrieved from https://censusreporter.org/profiles/16000US1714000-chicago-il/.

[10] Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, *42*(1), 68–80. doi: 10.1016/j.energy.2011.12.031

[11] Do, H., & Cetin, K. S. (2018). Residential Building Energy Consumption: A Review of Energy Data Availability, Characteristics, and Energy Performance Prediction Methods. *Current Sustainable/Renewable Energy Reports*, *5*(1), 76–85. doi: 10.1007/s40518-018-0099-3

[12] Elnakat, A., & Gomez, J. D. (2015). Energy engenderment: An industrialized perspective assessing the importance of engaging women in residential energy consumption management. *Energy Policy*, *82*, 166–177. doi: 10.1016/j.enpol.2015.03.014

[13] Elnakat, A., Gomez, J. D., & Booth, N. (2016). A zip code study of socioeconomic, demographic, and household gendered influence on the residential energy sector. *Energy Reports*, *2*, 21–27. doi: 10.1016/j.egyr.2016.01.003

[14] Farahbakhsh, H., Ugursal, V. I., & Fung, A. S. (1998). A residential end-use energy consumption model for Canada. *International Journal of Energy Research*, *22*(13), 1133–1143. doi: 10.1002/(sici)1099-114x(19981025)22:13<1133::aid-er434>3.0.co;2-e

[15] Fiebig, D. G., Bartels, R., & Aigner, D. J. (1991). A random coefficient approach to the estimation of residential end-use load profiles. *Journal of Econometrics*, *50*(3), 297–327. doi: 10.1016/0304-4076(91)90023-7

[16] Gerpott, T. J., & Paukert, M. (2013). Determinants of willingness to pay for smart meters: An empirical analysis of household customers in Germany. *Energy Policy*, *61*, 483–495. doi: 10.1016/j.enpol.2013.06.012

[17] Gouveia, J. P., & Seixas, J. (2016). Unraveling electricity consumption profiles in households through clusters: Combining smart meters and door-to-door surveys. *Energy and Buildings*, *116*, 666–676. doi: 10.1016/j.enbuild.2016.01.043

[18] Grubesic, T. H. (2008). Zip codes and spatial analysis: Problems and prospects. *Socio-Economic Planning Sciences*, *42*(2), 129–149. doi: 10.1016/j.seps.2006.09.001

[19] Hajilary, N., Shahi, A., & Rezakazemi, M. (2018). Evaluation of socio-economic factors on CO_2 emissions in Iran: Factorial design and multivariable methods. *Journal of Cleaner Production*, *189*, 108–115. doi: 10.1016/j.jclepro.2018.04.067

[20] Harris, J. L., & Liu, L.-M. (1993). Dynamic structural analysis and forecasting of residential electricity consumption. *International Journal of Forecasting*, *9*(4), 437–455. doi: 10.1016/0169-2070(93)90072-u

[21] Hayn, M., Bertsch, V., & Fichtner, W. (2014). Electricity load profiles in Europe: The importance of household segmentation. *Energy Research & Social Science*, *3*, 30–45. doi: 10.1016/j.erss.2014.07.002

[22] Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., & Modi, V. (2012). Spatial distribution of urban building energy consumption by end use. *Energy and Buildings*, *45*, 141–151. doi: 10.1016/j.enbuild.2011.10.061

[23] Im, J., Seo, Y., Cetin, K. S., & Singh, J. (2017). Energy efficiency in U.S. residential rental housing: Adoption rates and impact on rent. *Applied Energy*, *205*, 1021–1033. doi: 10.1016/j.apenergy.2017.08.047

[24] Ingraham, C. (2019, April 26). Map: The age of housing in every zip code in the U.S. Retrieved from https://www.washingtonpost.com/news/wonk/wp/2014/06/24/map-the-age-of-housing-in-every-zip-code-in-the-u-s/.

[25] Jones, R. V., Fuertes, A., & Lomas, K. (2015). The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings (Version 1). figshare. Retrieved from https://hdl.handle.net/2134/16696

[26] Karatasou, S., Laskari, M., & Santamouris, M. (2018). Determinants of high electricity use and high energy consumption for space and water heating in European social housing: Socio-demographic and building characteristics. *Energy and Buildings*, *170*, 107–114. doi: 10.1016/j.enbuild.2018.04.019

[27] Kavousian, A., Rajagopal, R., & Fischer, M. (2013). Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate,

building characteristics, appliance stock, and occupants' behavior. *Energy*, *55*, 184–194. doi: 10.1016/j.energy.2013.03.086

[28] Kiefer, J. C., & Krentz, L. R. (2018). Water Use in the Multi-Family Housing Sector. *Water Resource Foundation*.

[29] McLoughlin, F., Duffy, A., & Conlon, M. (2015). A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, *141*, 190–199. doi: 10.1016/j.apenergy.2014.12.039

[30] Measuring Socioeconomic Status and Subjective Social Status. Retrieved from https://www.apa.org/pi/ses/resources/class/measuring-status.

[31] Narayan, P. K., & Smyth, R. (2005). Electricity consumption, employment and real income in Australia evidence from multivariate Granger causality tests. *Energy Policy*, *33*(9), 1109–1116. doi: 10.1016/j.enpol.2003.11.010

[32] Parti, M., & Parti, C. (1980). The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector. *The Bell Journal of Economics*, *11*(1), 309. doi: 10.2307/3003415

[33] Raza, M. Q., & Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, *50*, 1352–1372. doi: 10.1016/j.rser.2015.04.065

[34] Suganthi, L., & Samuel, A. A. (2012). Energy models for demand forecasting—A review. *Renewable and Sustainable Energy Reviews*, *16*(2), 1223–1240. doi: 10.1016/j.rser.2011.08.014

[35] US Census Bureau. (n.d.). Census.gov. Retrieved from https://www.census.gov/.

[36] Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., & Zhao, Y. (2015). Load profiling and its application to demand response: A review. *Tsinghua Science and Technology*, *20*(2), 117–129. doi: 10.1109/tst.2015.7085625

[37] Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. doi: 10.1080/01621459.1963.10500845

[38] Yildiz, B., Bilbao, J., Dore, J., & Sproul, A. (2017). Recent advances in the analysis of residential electricity consumption and applications of smart meter data. *Applied Energy*, *208*, 402–427. doi: 10.1016/j.apenergy.2017.10.014

Appendix: Generated Graphs and Data

A1. Electricity Daily Load Profiles by Zip Code

Zip Codes with Complete Data for 2016

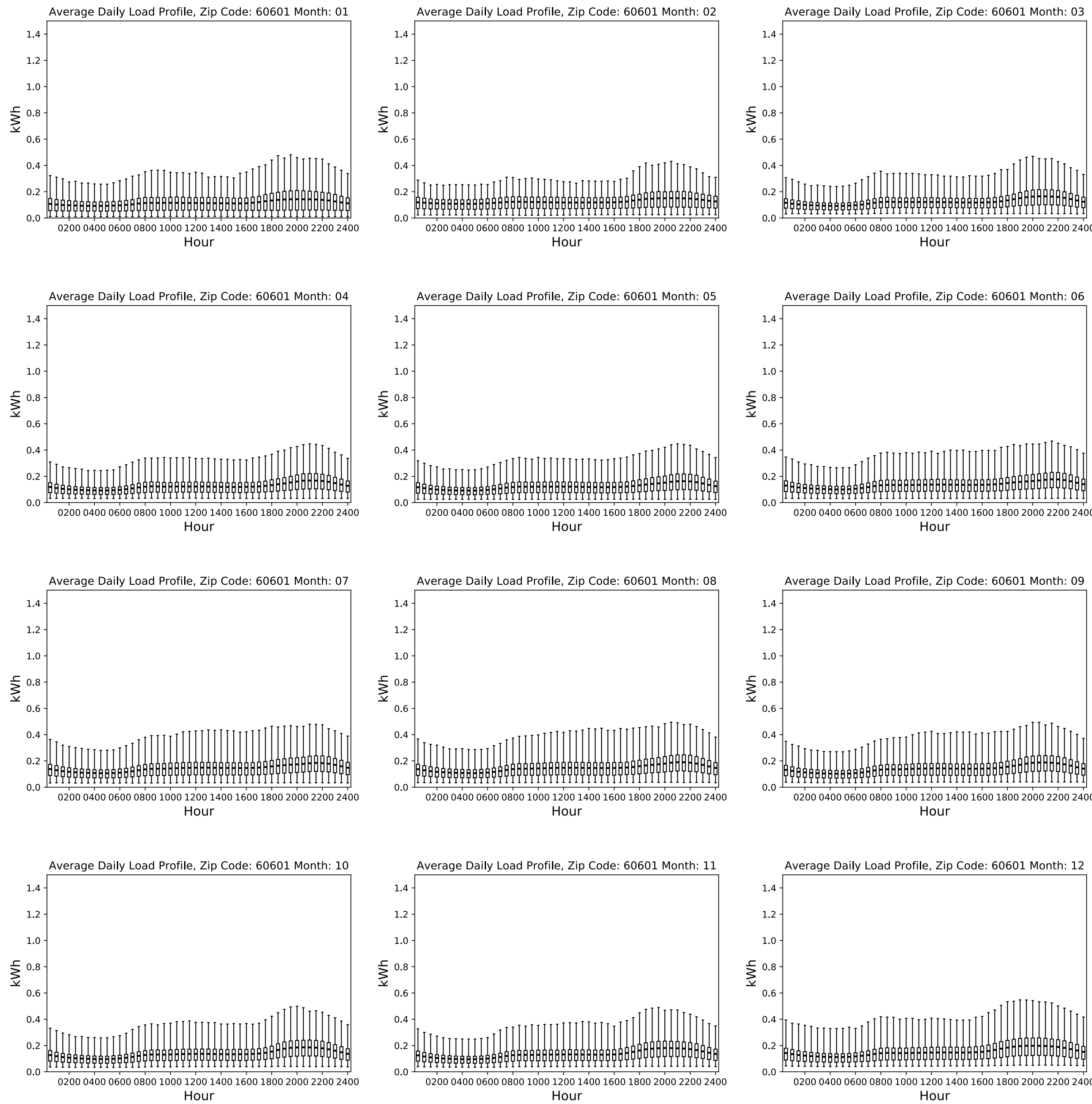

Figure A1. Daily load profiles for multi-family residential electricity consumption by month: 60601.

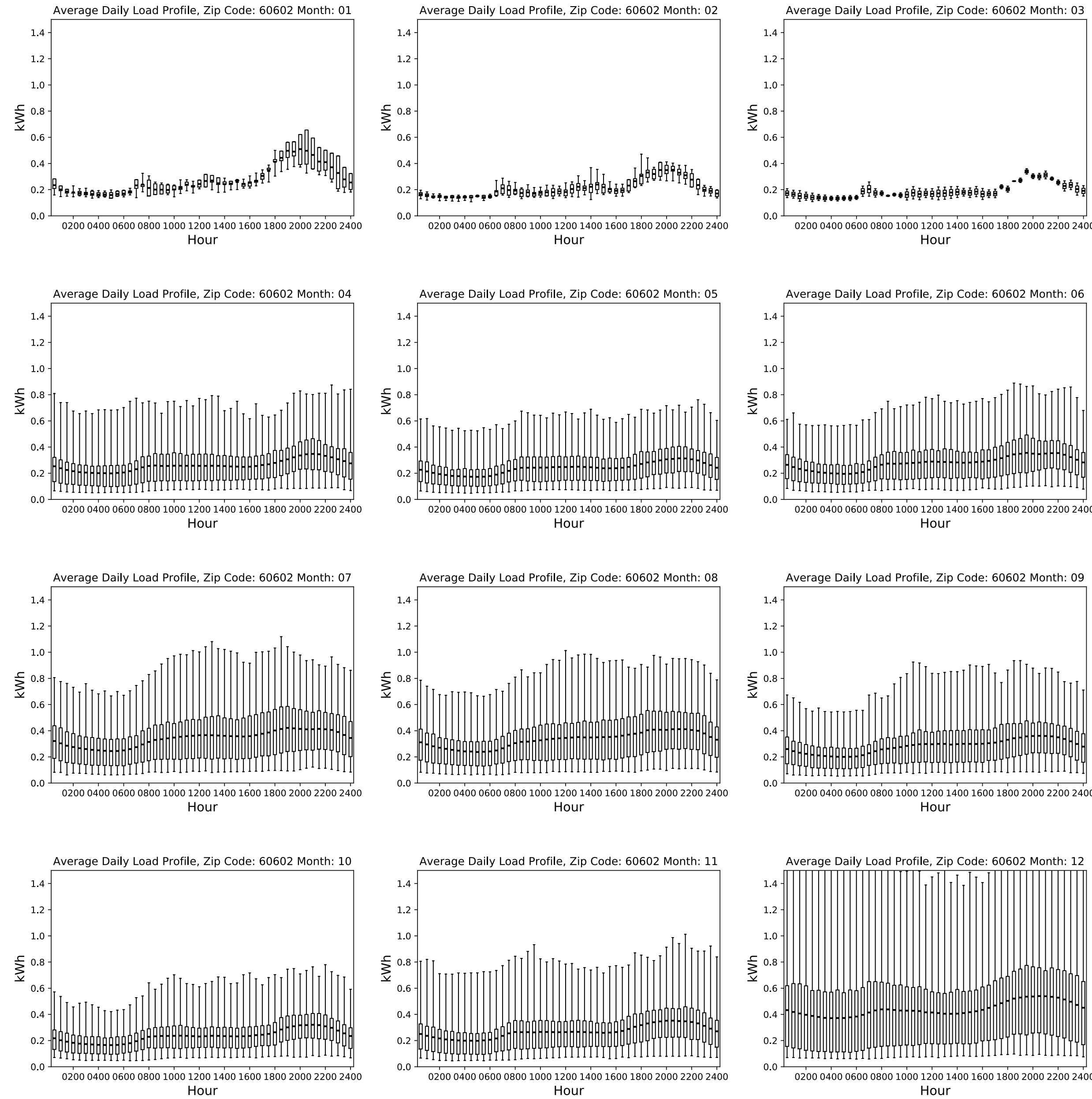

Figure A2. Daily load profiles for multi-family residential electricity consumption by month: 60602.

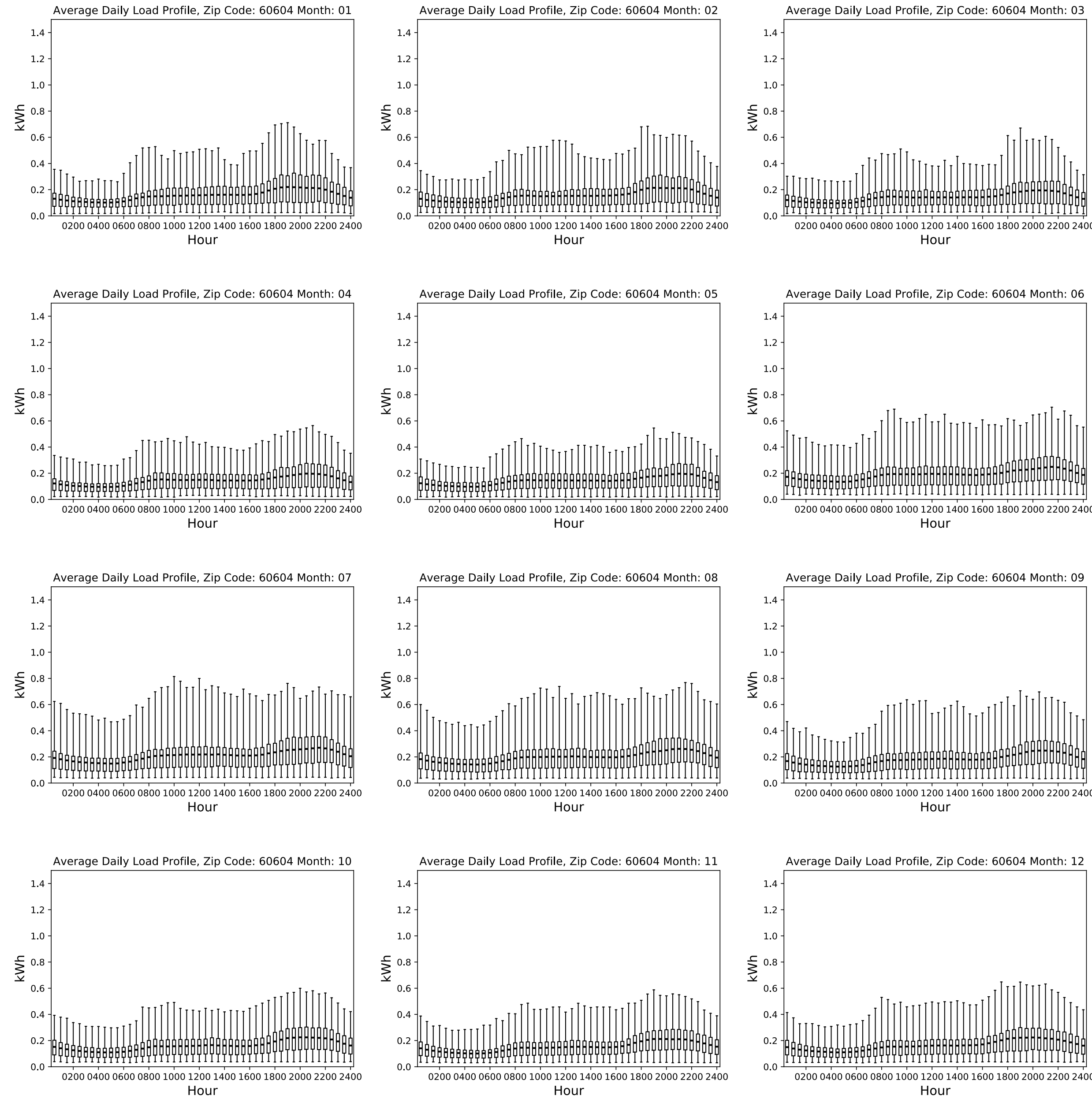

Figure A3. Daily load profiles for multi-family residential electricity consumption by month: 60604.

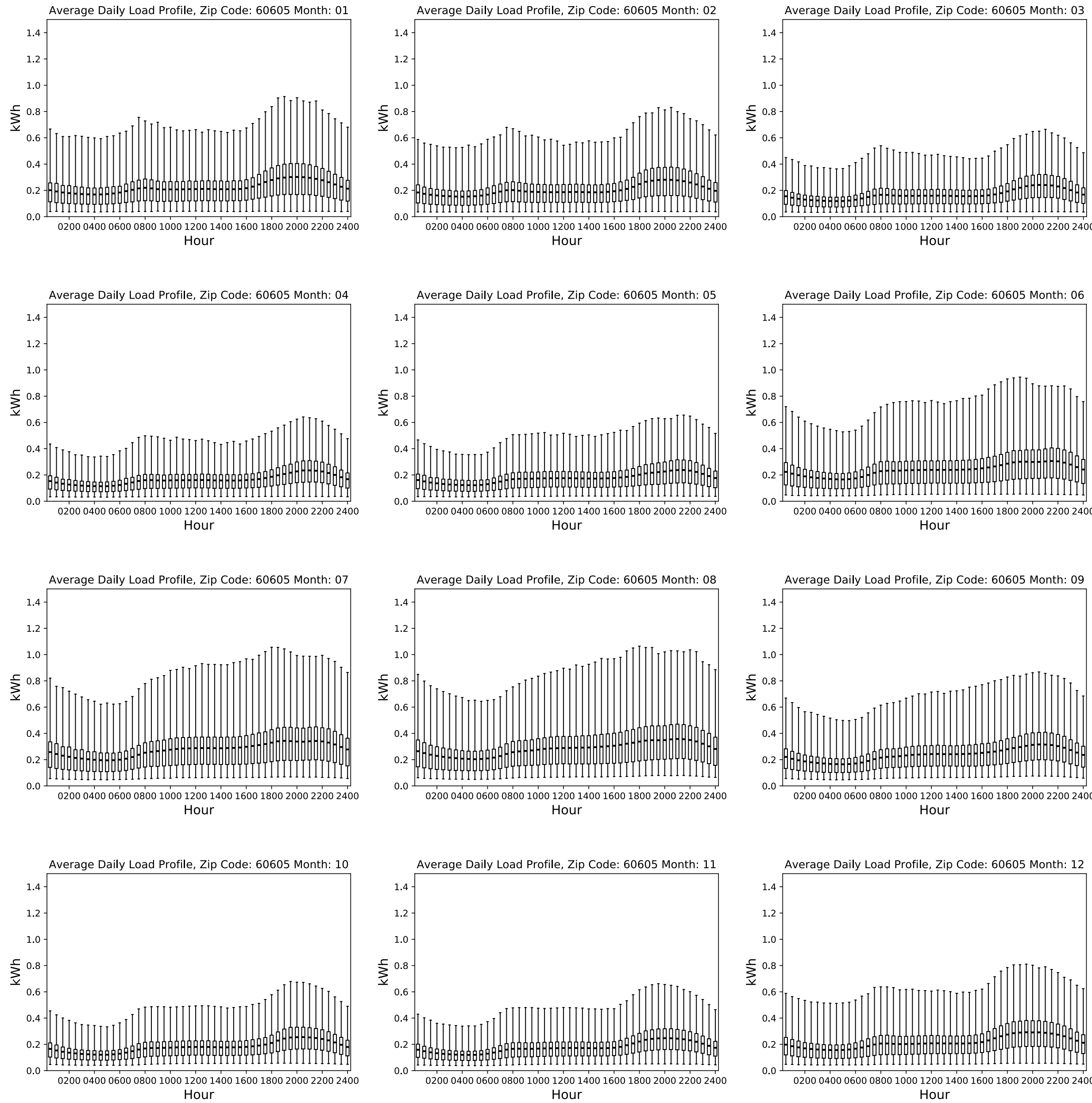

Figure A4. Daily load profiles for multi-family residential electricity consumption by month: 60605.

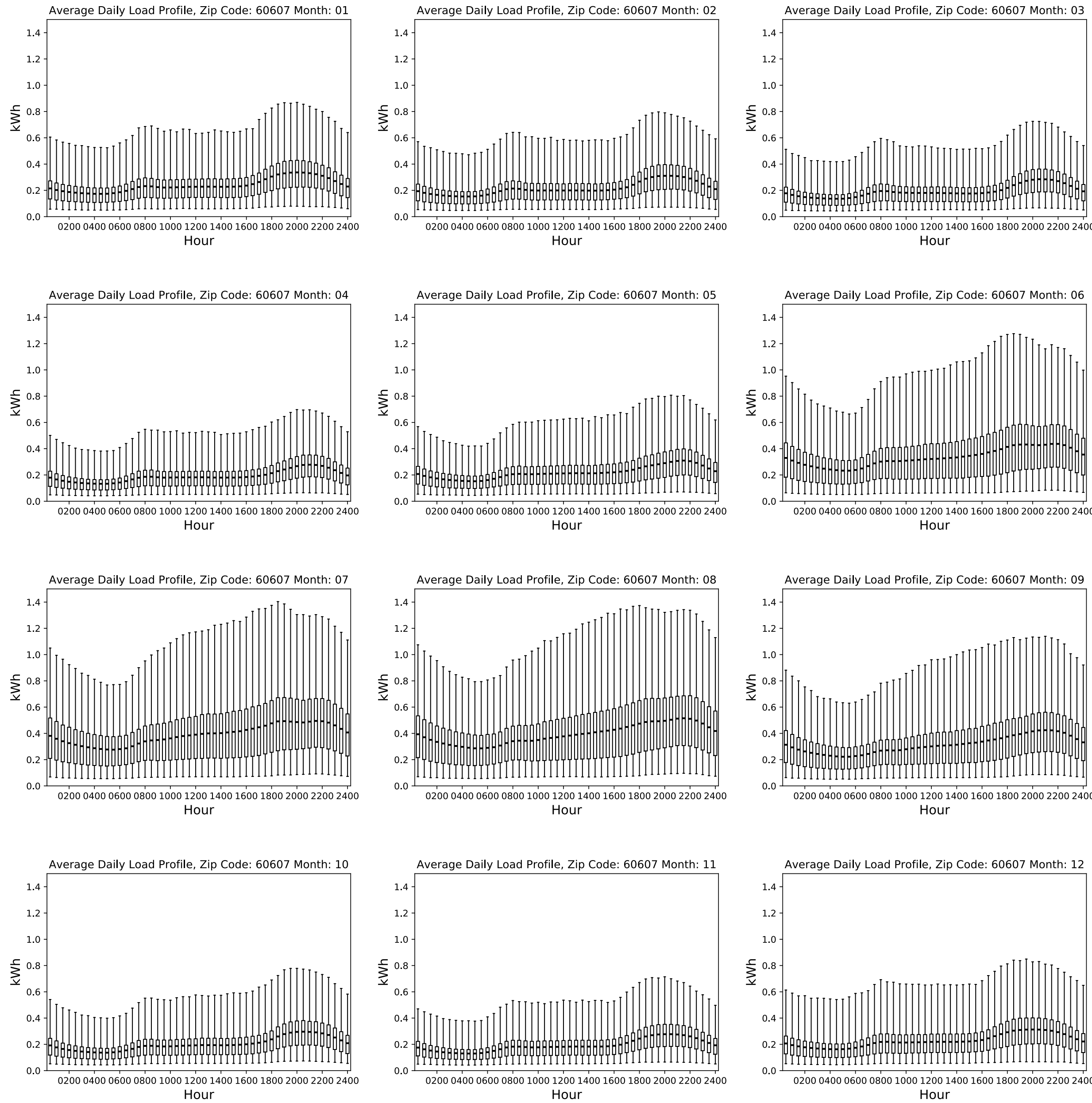

Figure A5. Daily load profiles for multi-family residential electricity consumption by month: 60607.

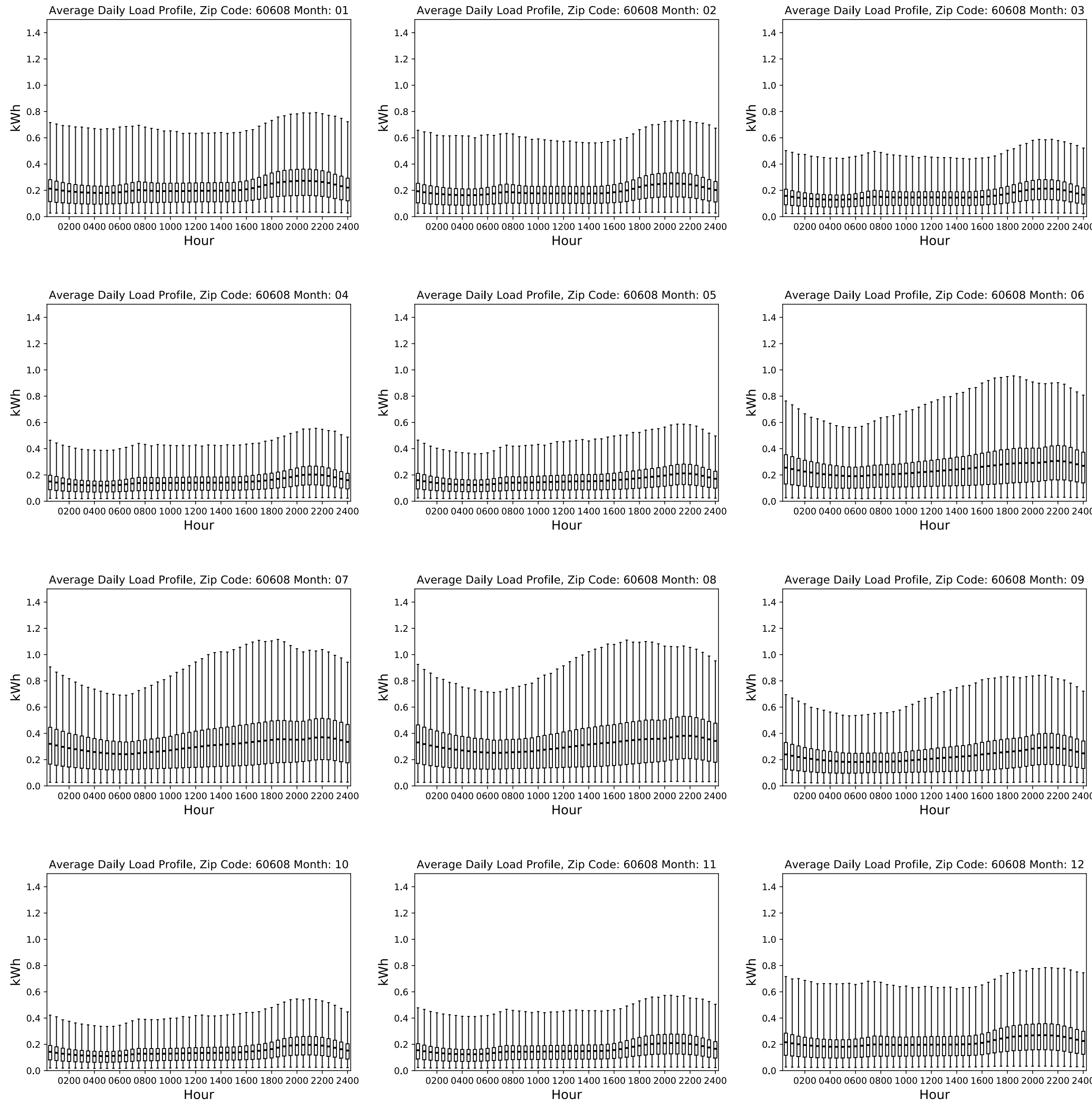

Figure A6. Daily load profiles for multi-family residential electricity consumption by month: 60608.

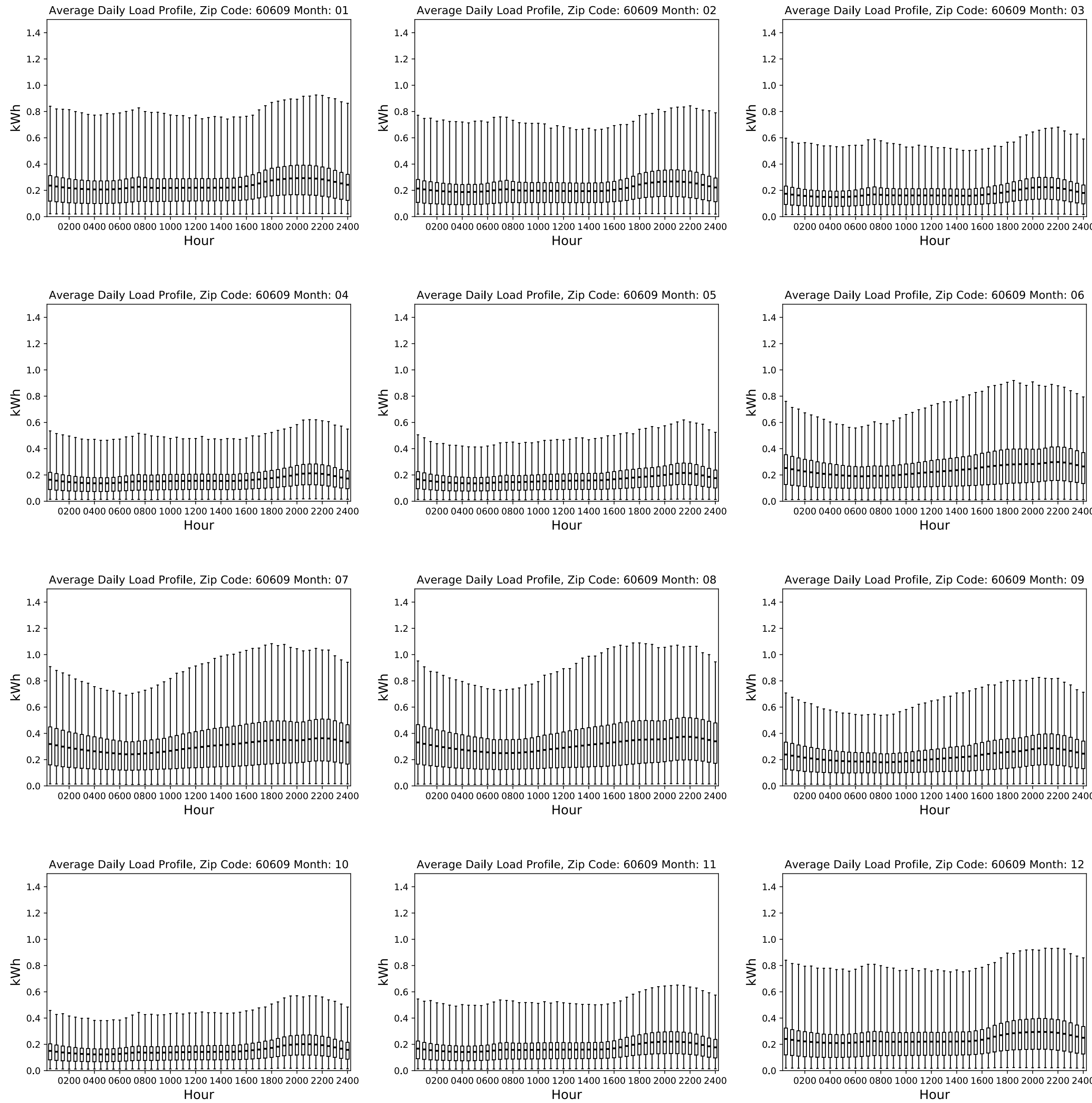

Figure A7. Daily load profiles for multi-family residential electricity consumption by month: 60609.

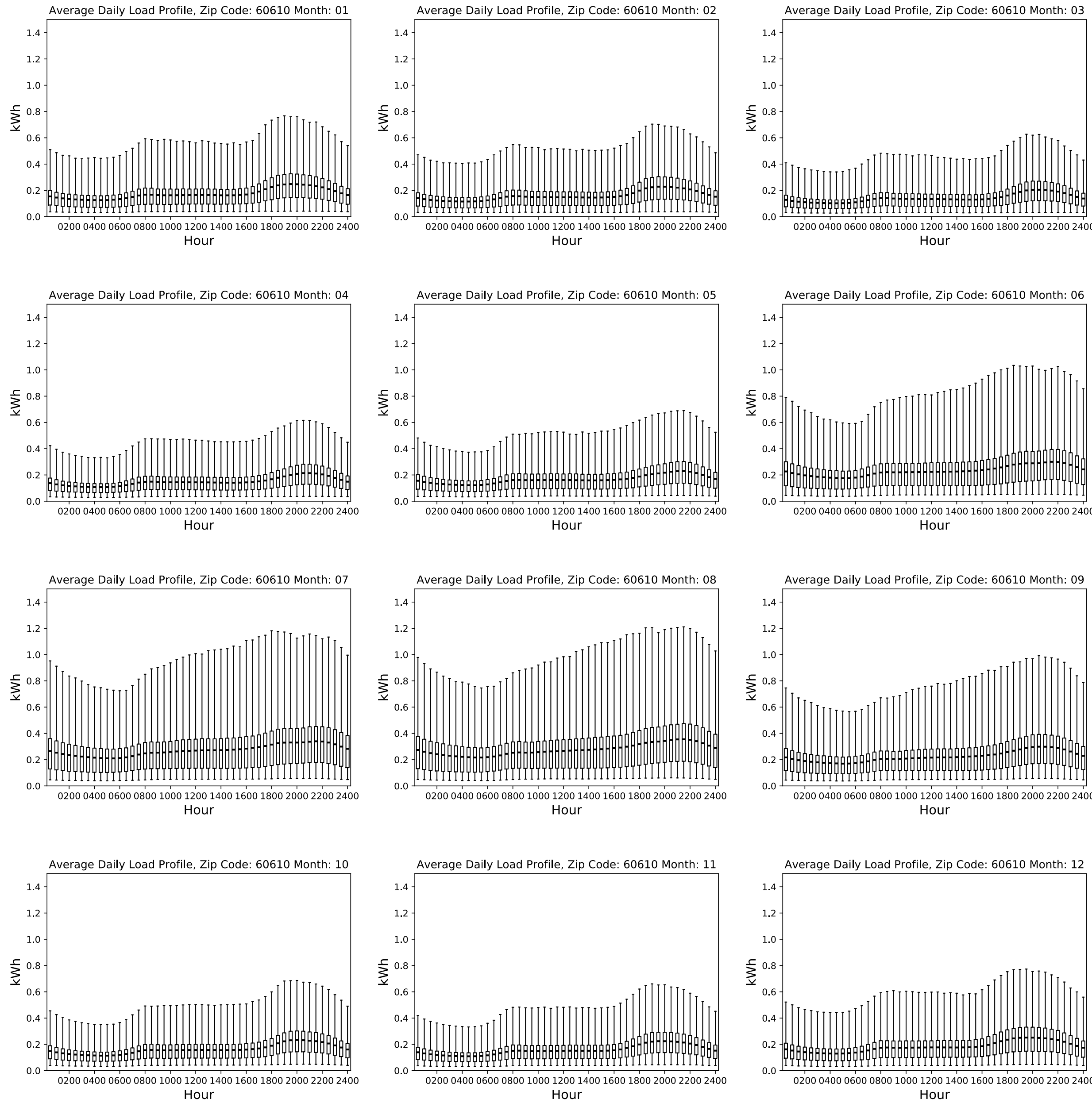

Figure A8. Daily load profiles for multi-family residential electricity consumption by month: 60610.

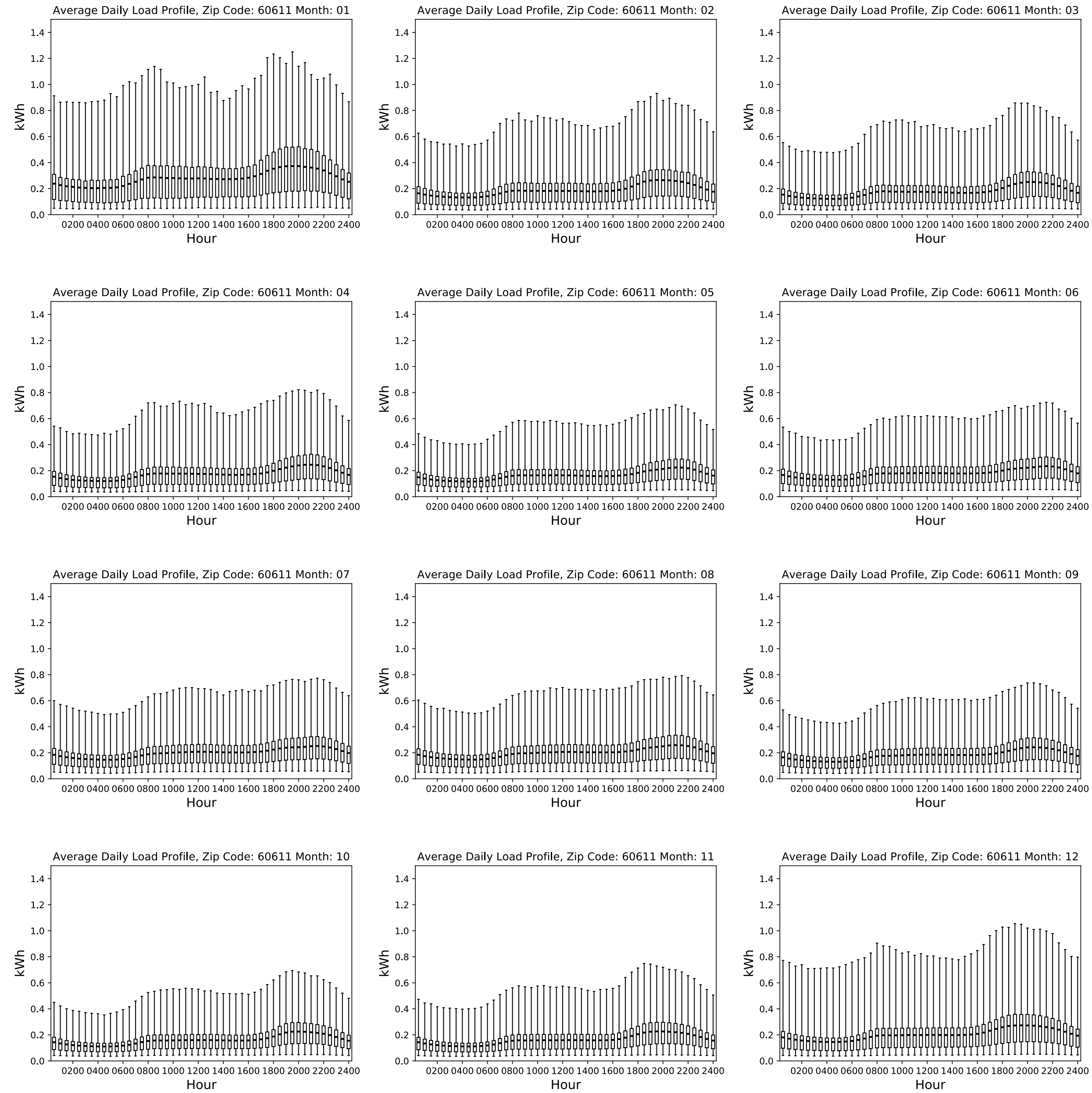

Figure A9. Daily load profiles for multi-family residential electricity consumption by month: 60611.

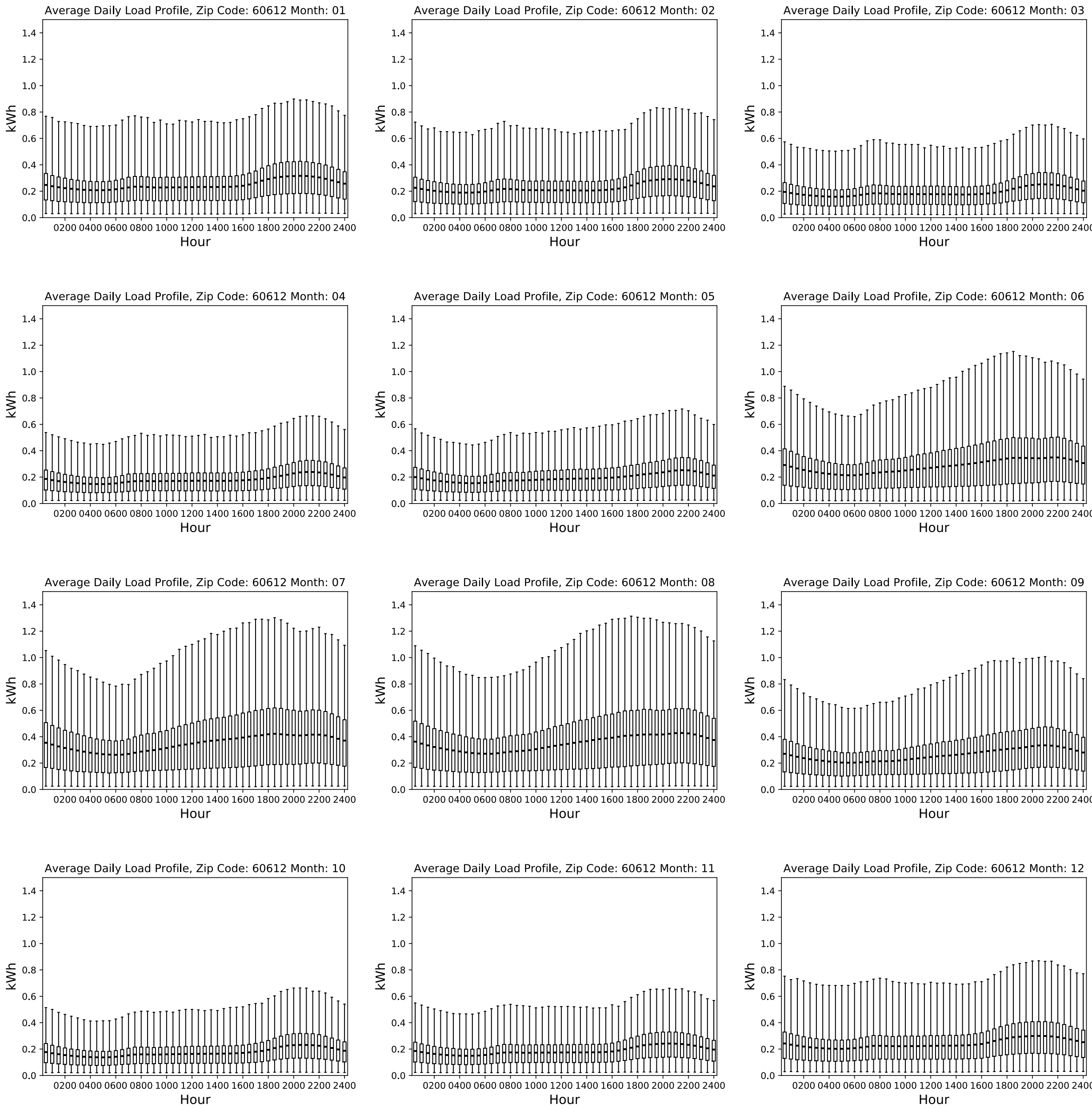

Figure A10. Daily load profiles for multi-family residential electricity consumption by month: 60612.

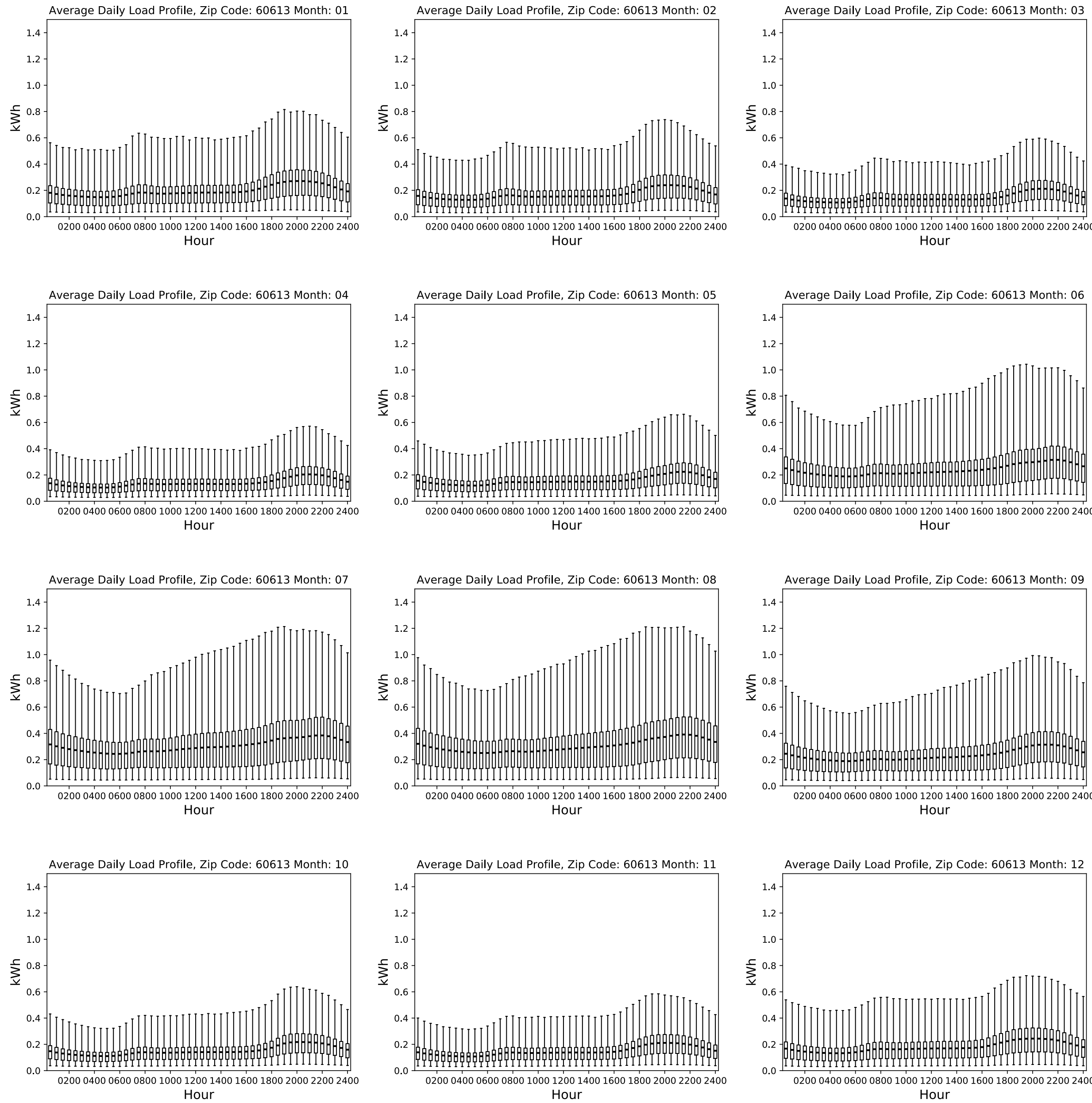

Figure A11. Daily load profiles for multi-family residential electricity consumption by month: 60613.

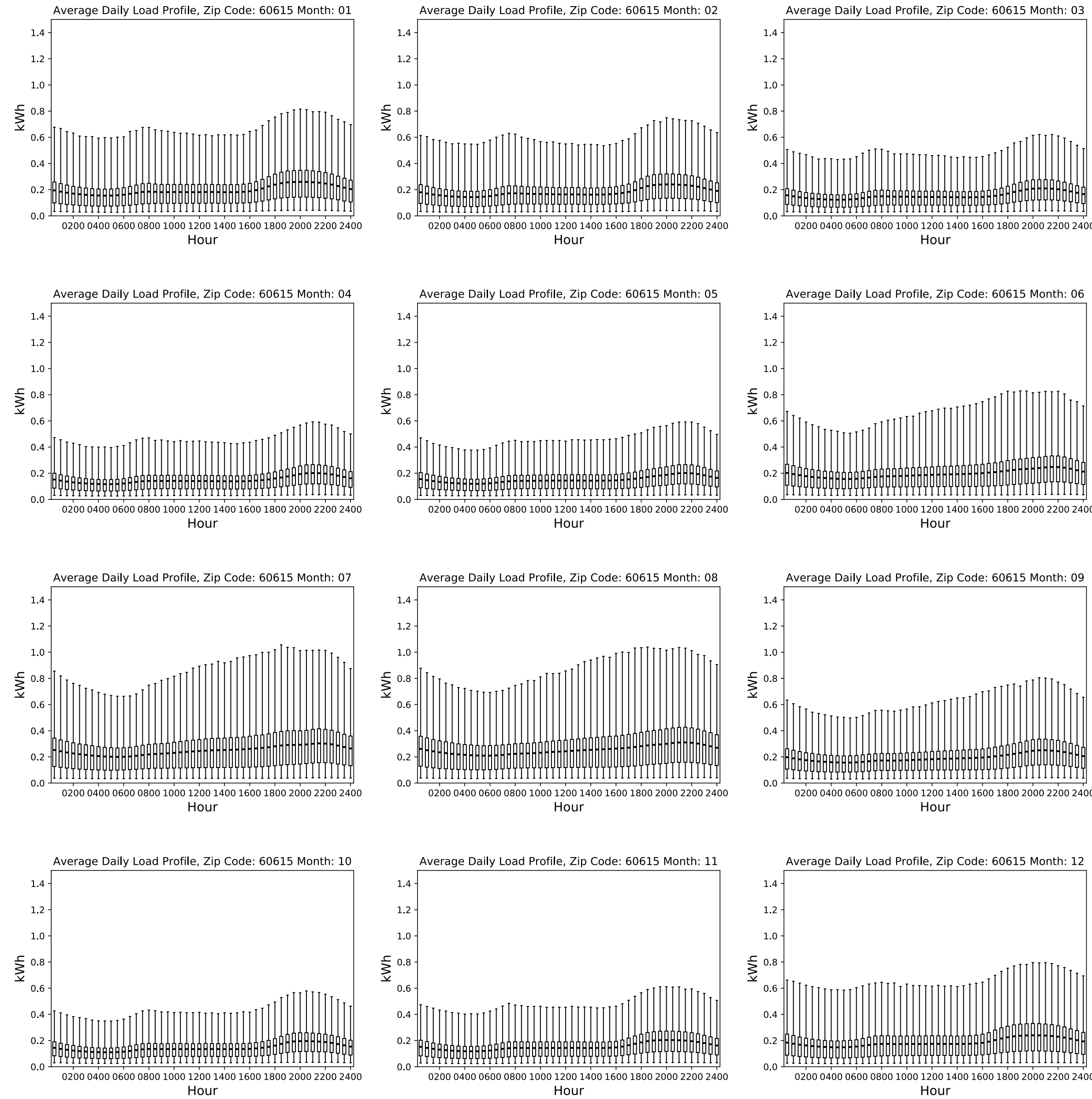

Figure A12. Daily load profiles for multi-family residential electricity consumption by month: 60615.

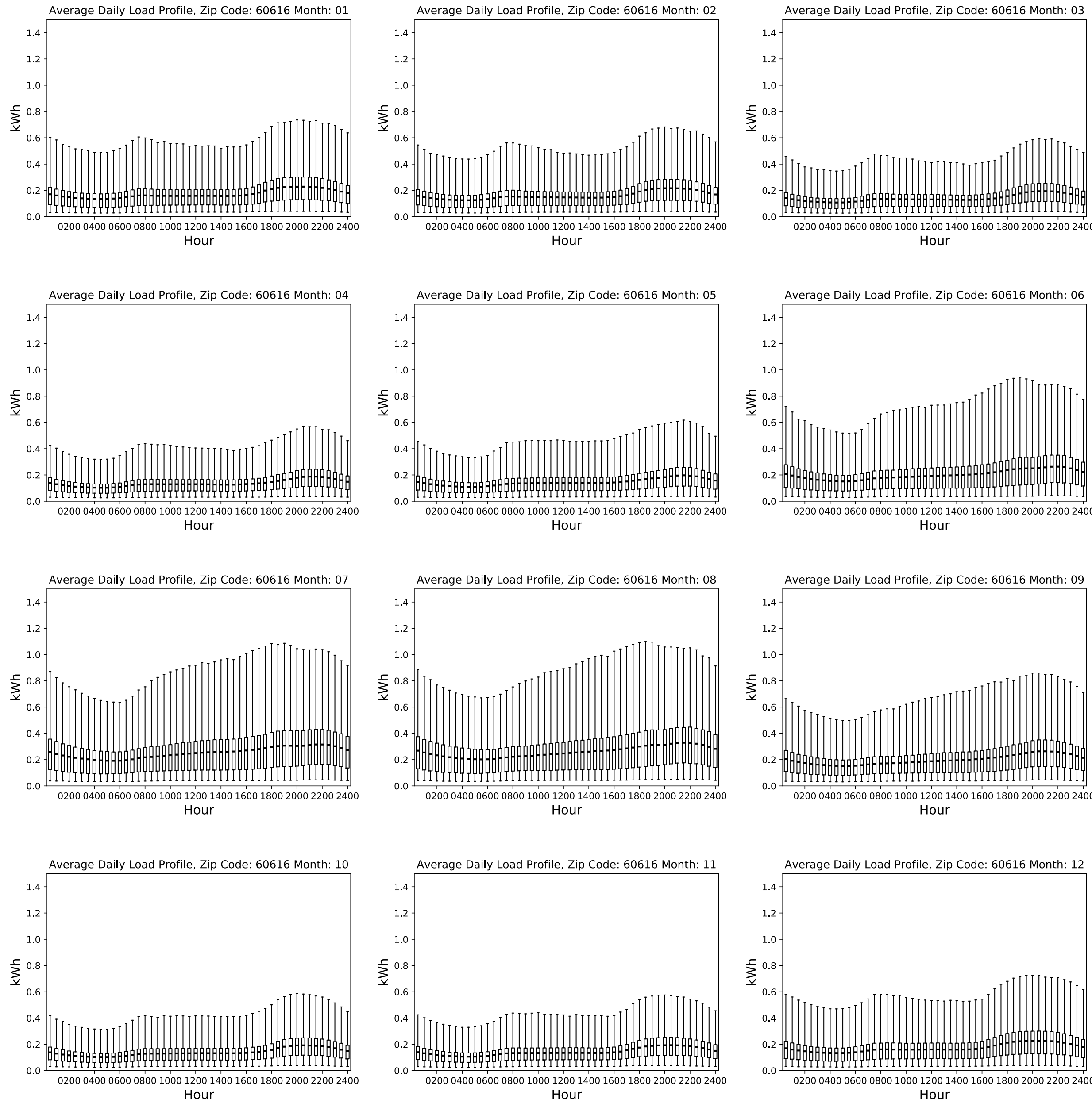

Figure A13. Daily load profiles for multi-family residential electricity consumption by month: 60616.

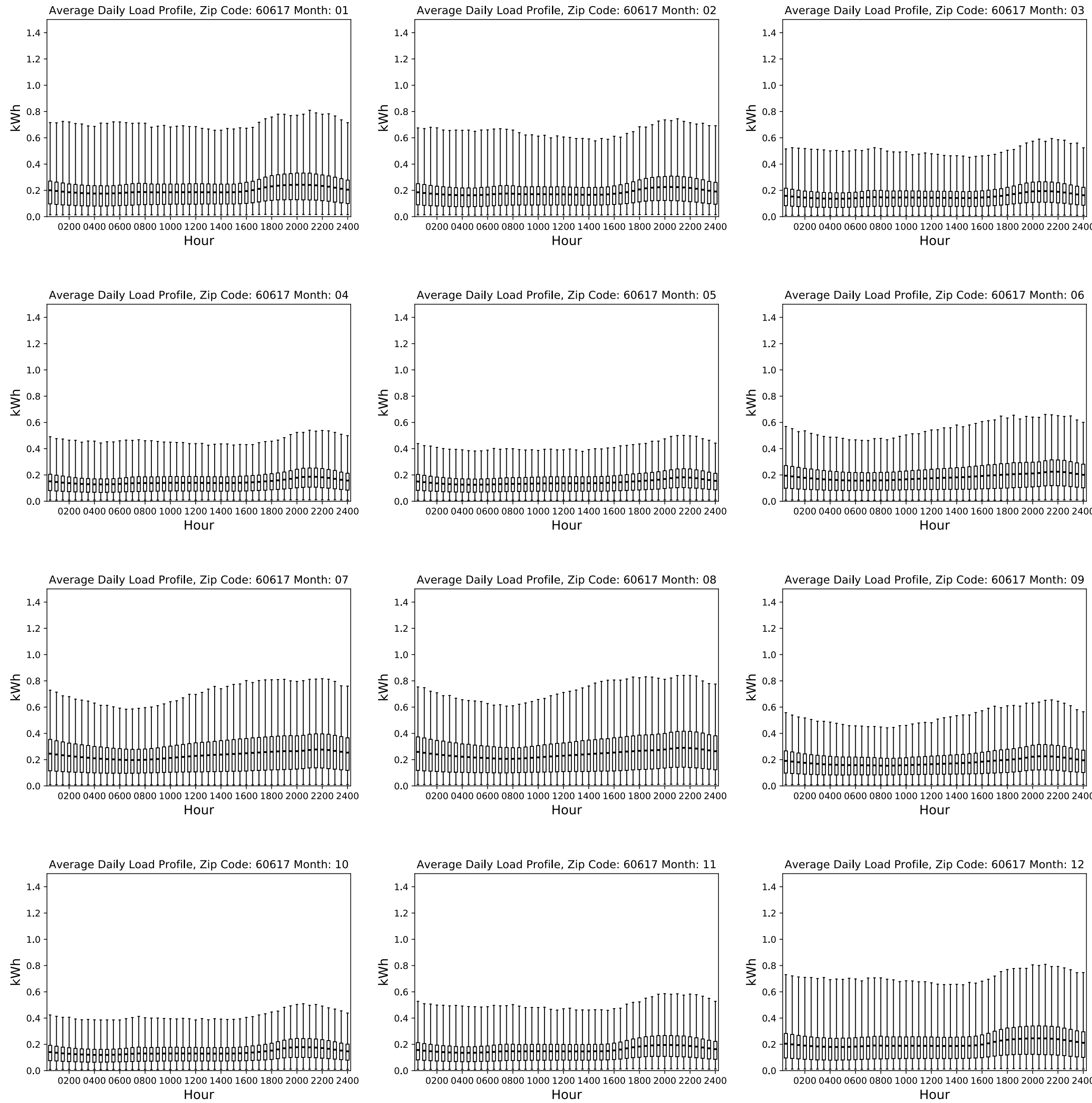

Figure A14. Daily load profiles for multi-family residential electricity consumption by month: 60617.

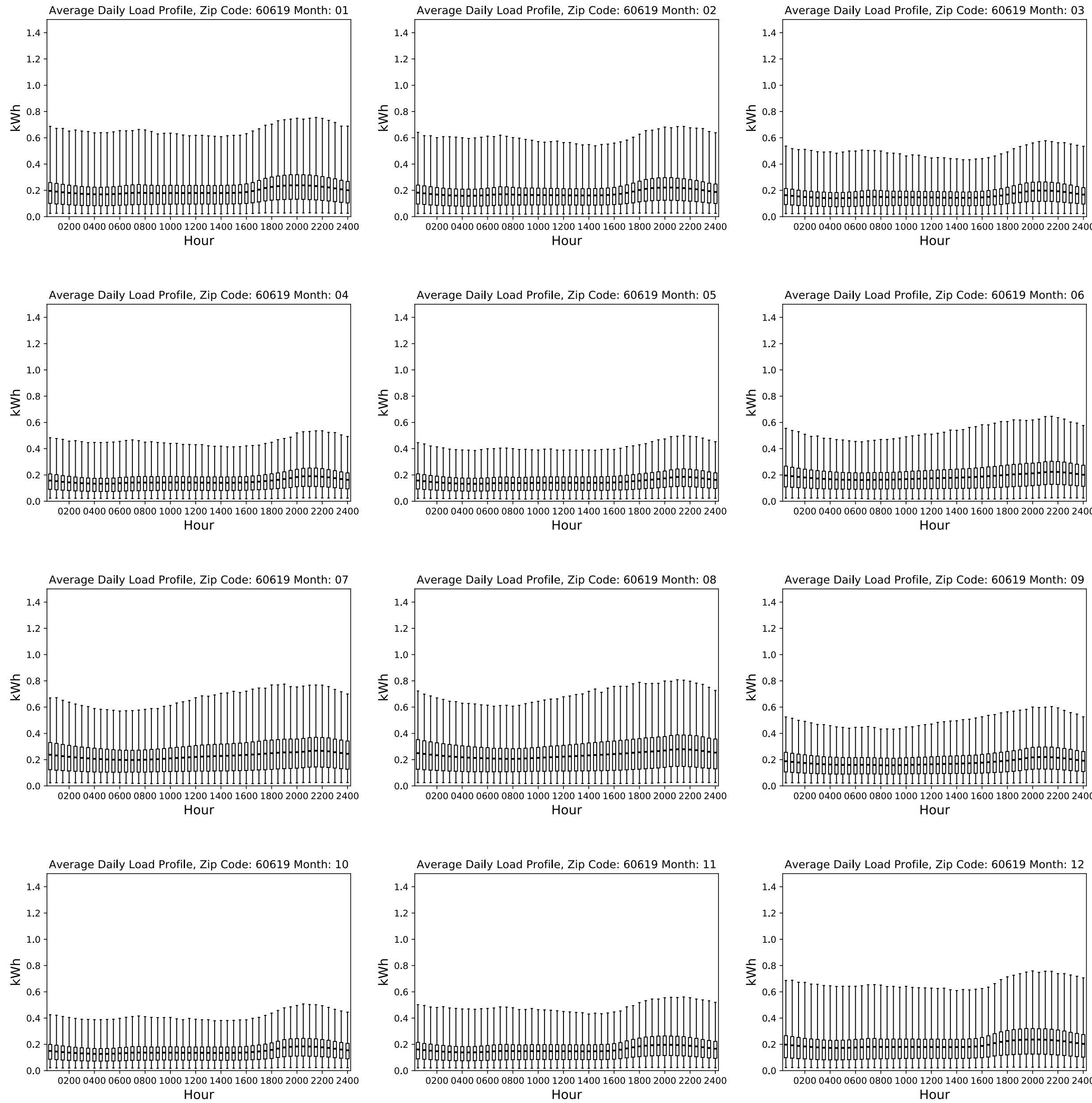

Figure A15. Daily load profiles for multi-family residential electricity consumption by month: 60619.

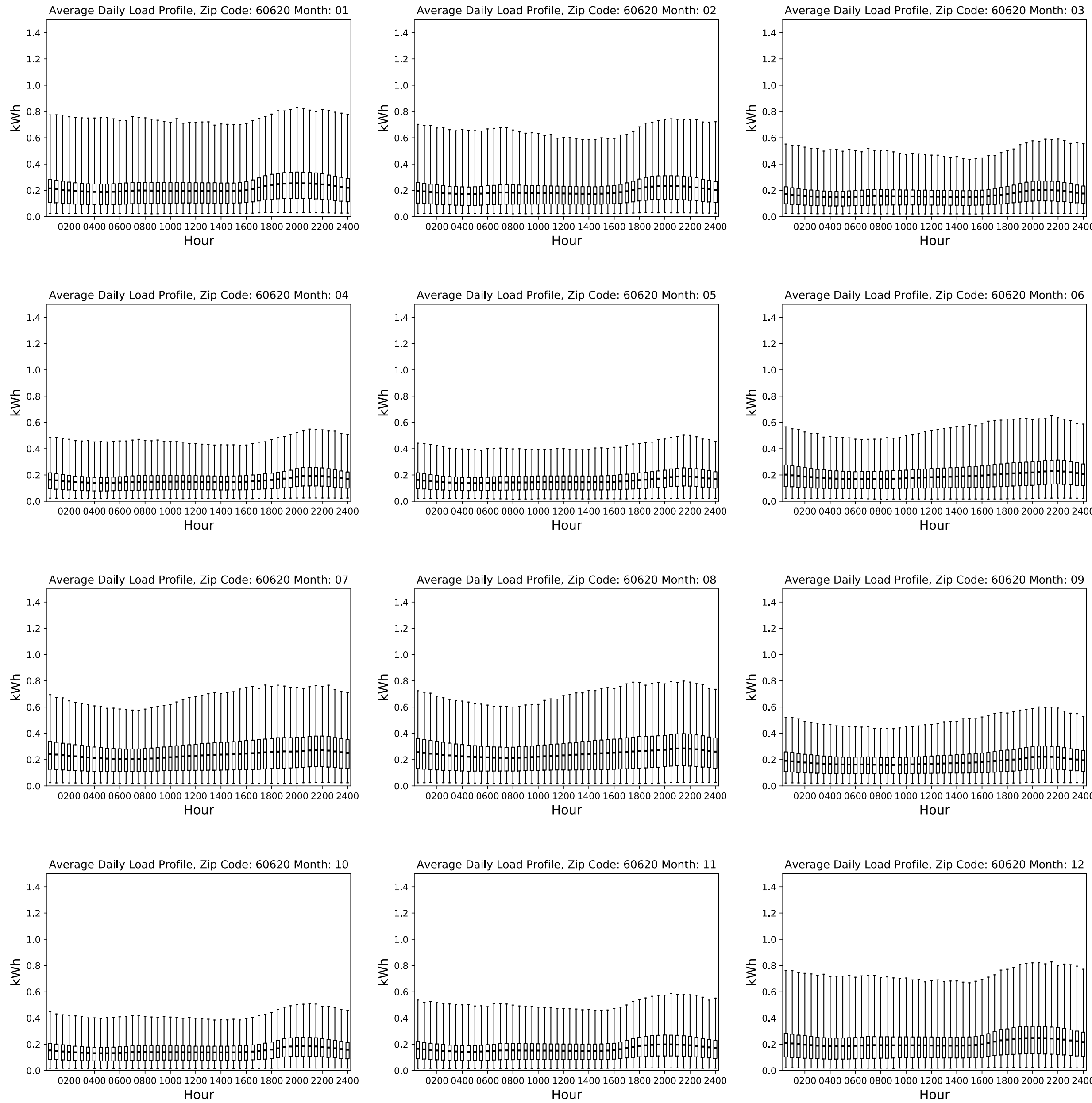

Figure A16. Daily load profiles for multi-family residential electricity consumption by month: 60620.

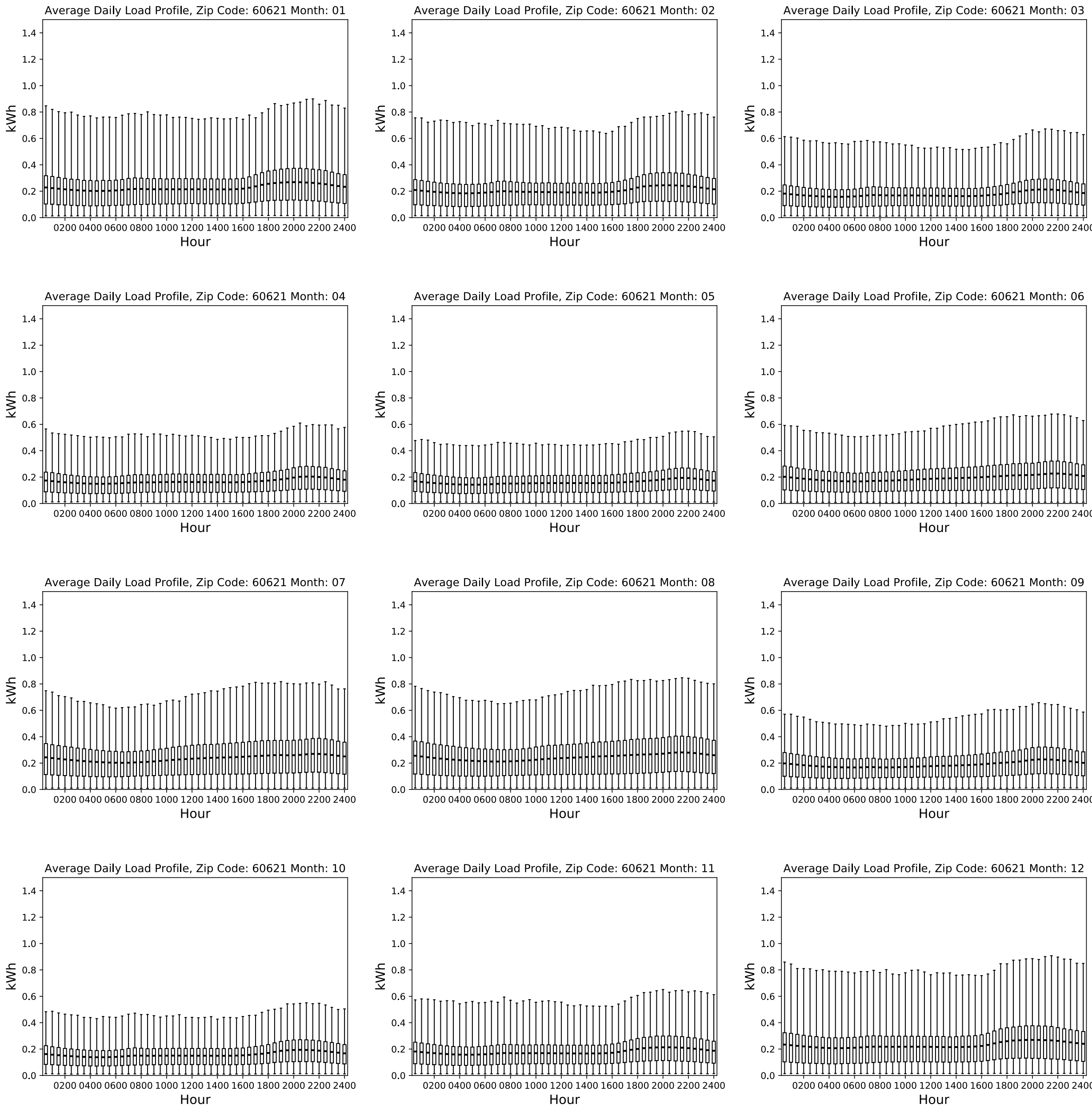

Figure A17. Daily load profiles for multi-family residential electricity consumption by month: 60621.

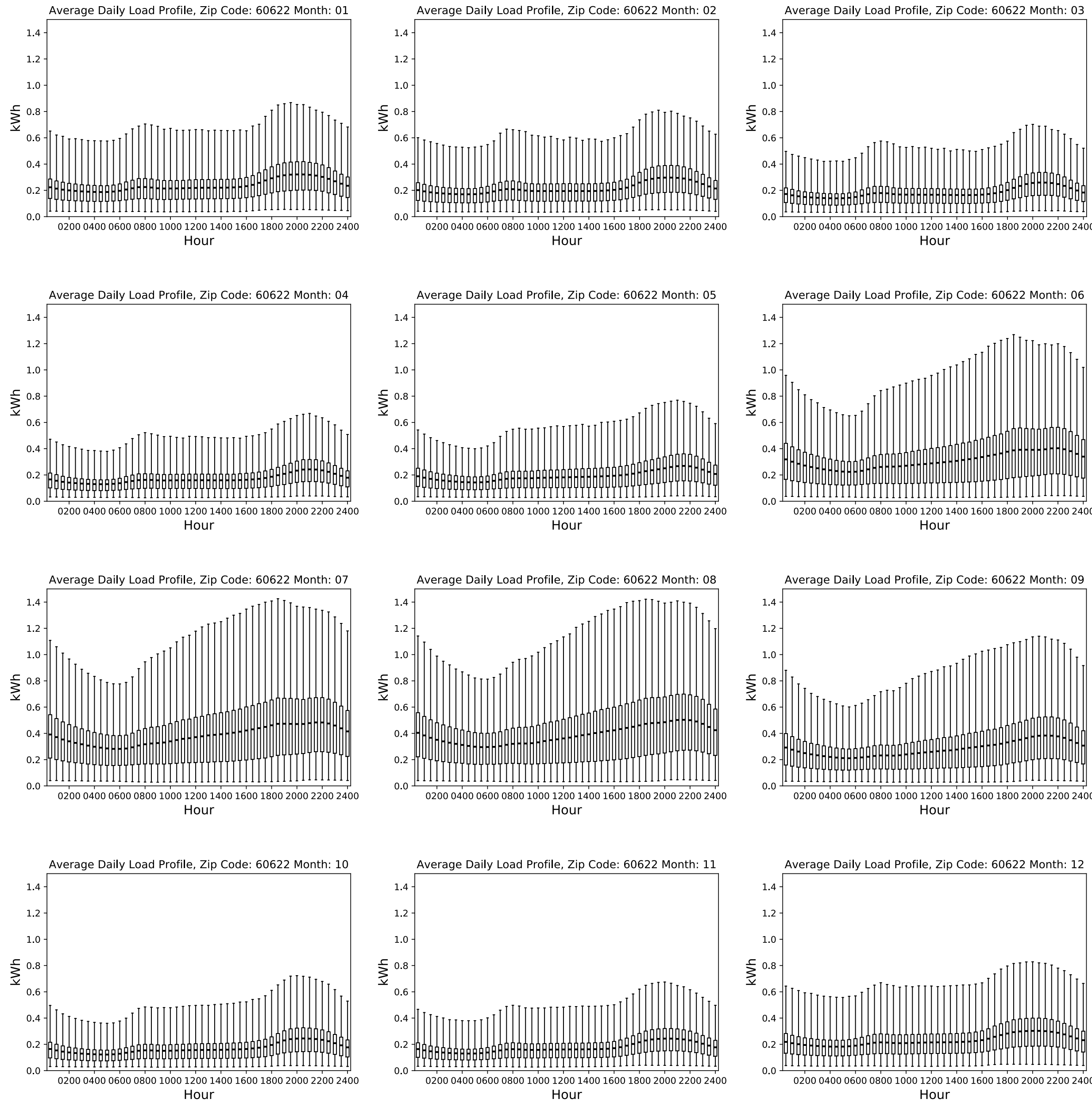

Figure A18. Daily load profiles for multi-family residential electricity consumption by month: 60622.

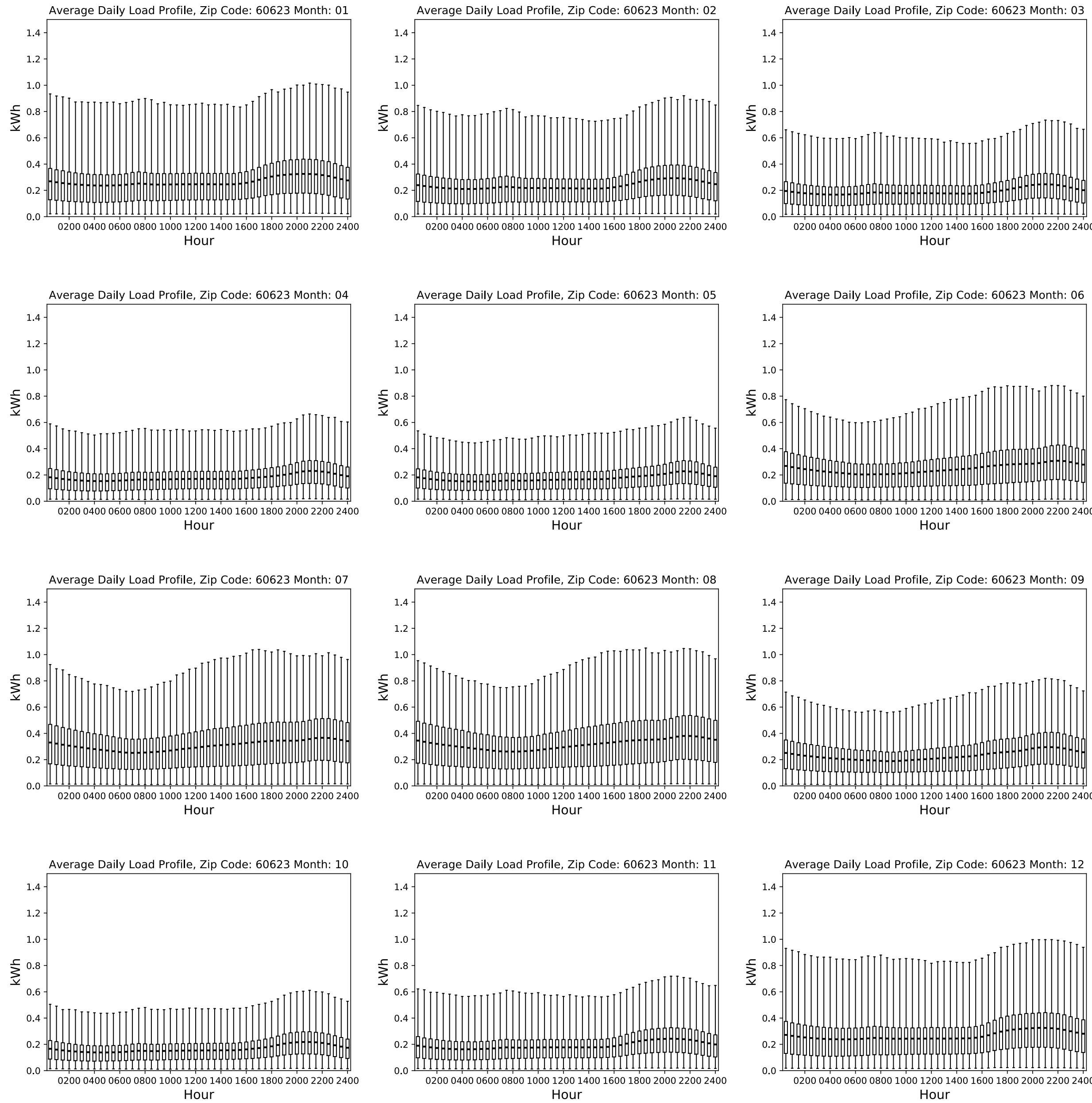

Figure A19. Daily load profiles for multi-family residential electricity consumption by month: 60623.

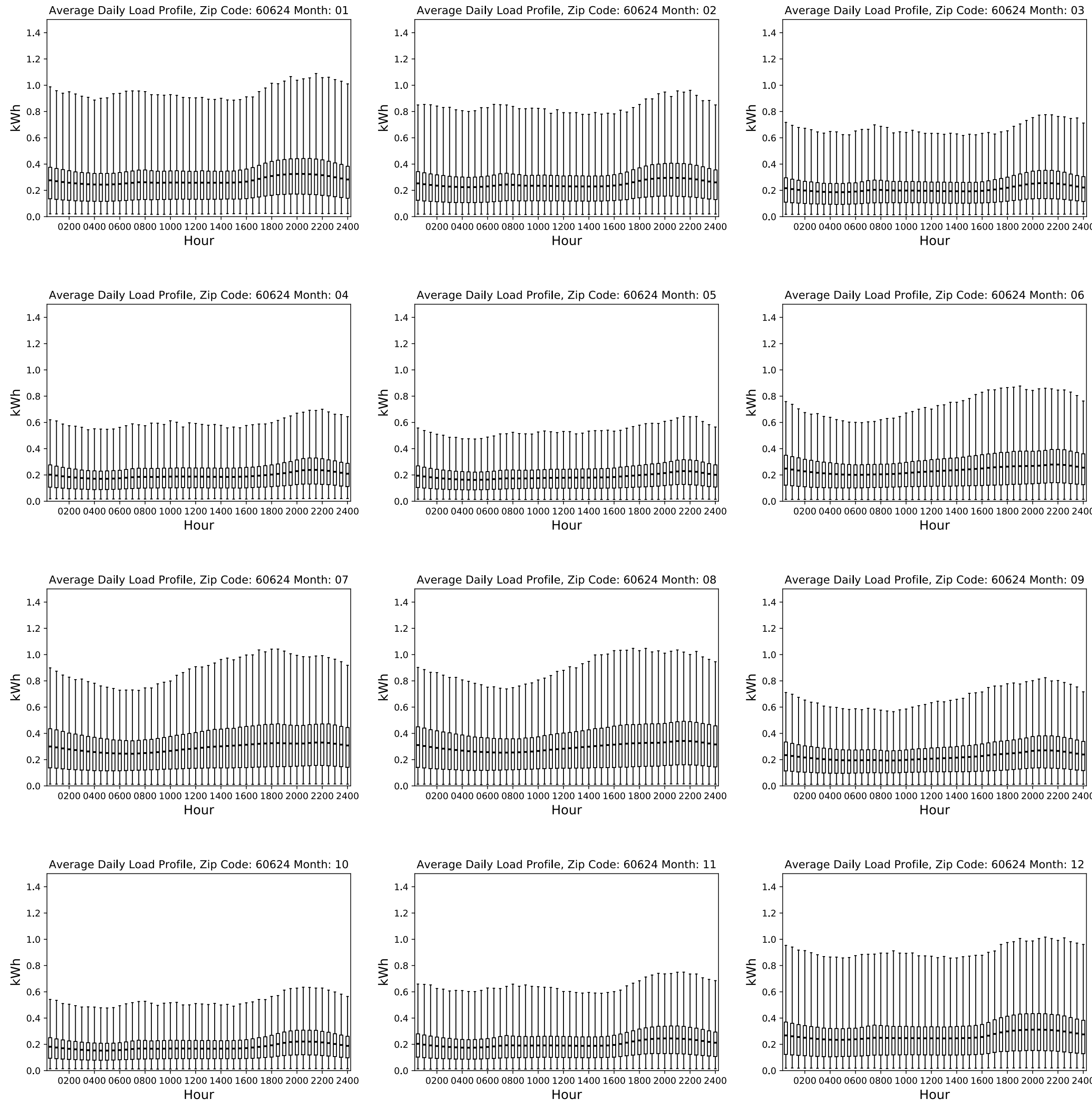

Figure A20. Daily load profiles for multi-family residential electricity consumption by month: 60624.

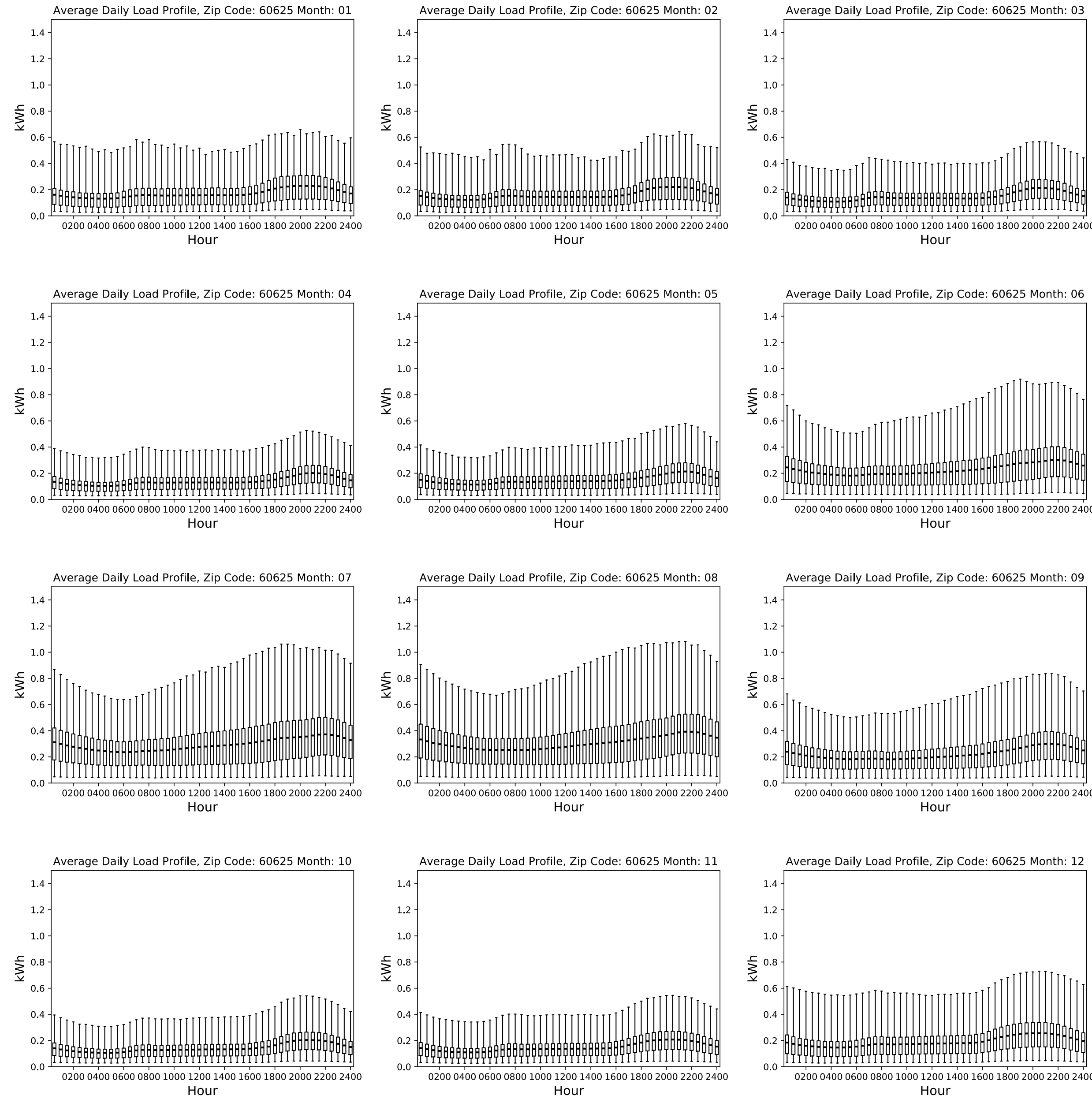

Figure A21. Daily load profiles for multi-family residential electricity consumption by month: 60625.

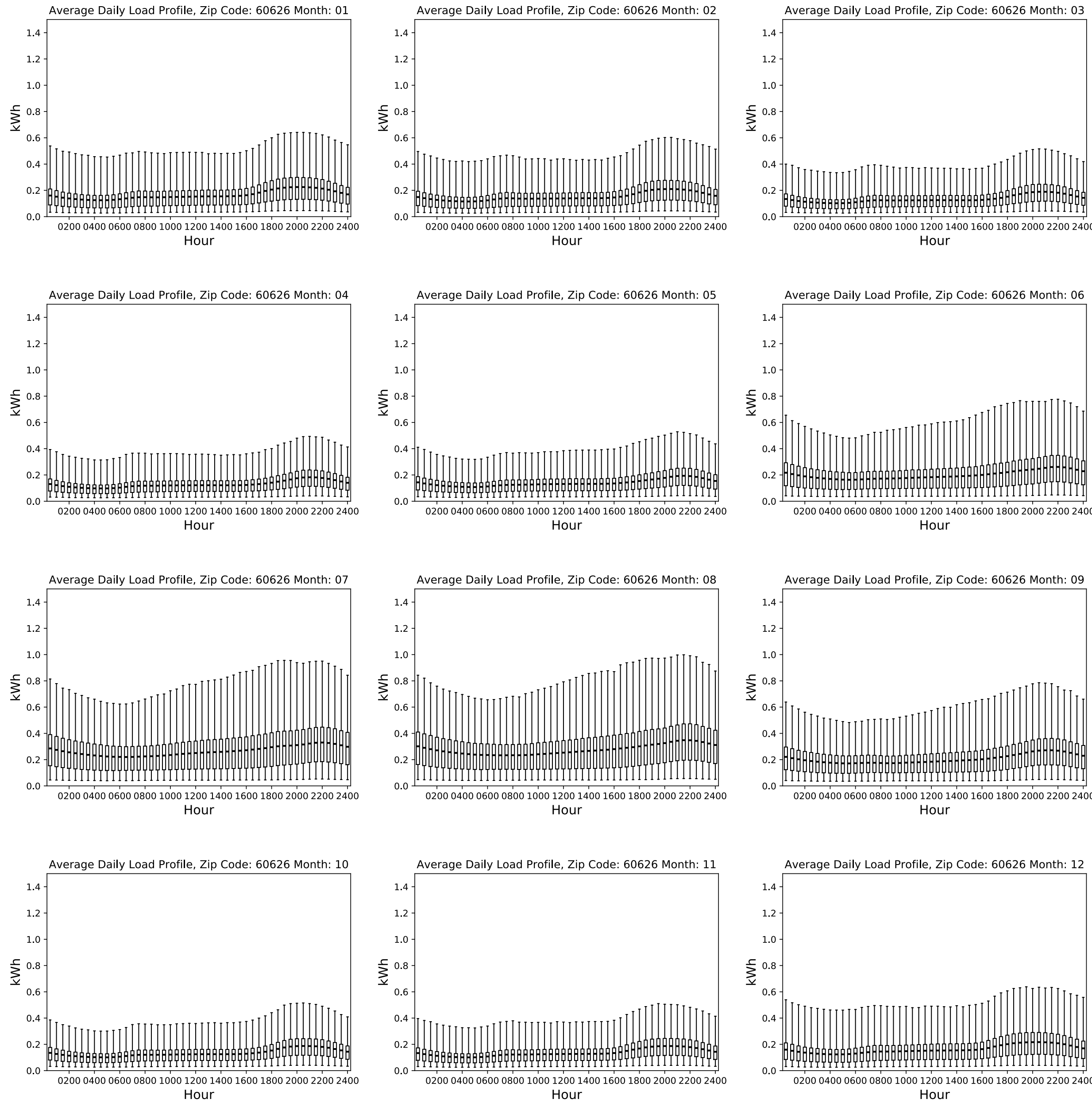

Figure A22. Daily load profiles for multi-family residential electricity consumption by month: 60626.

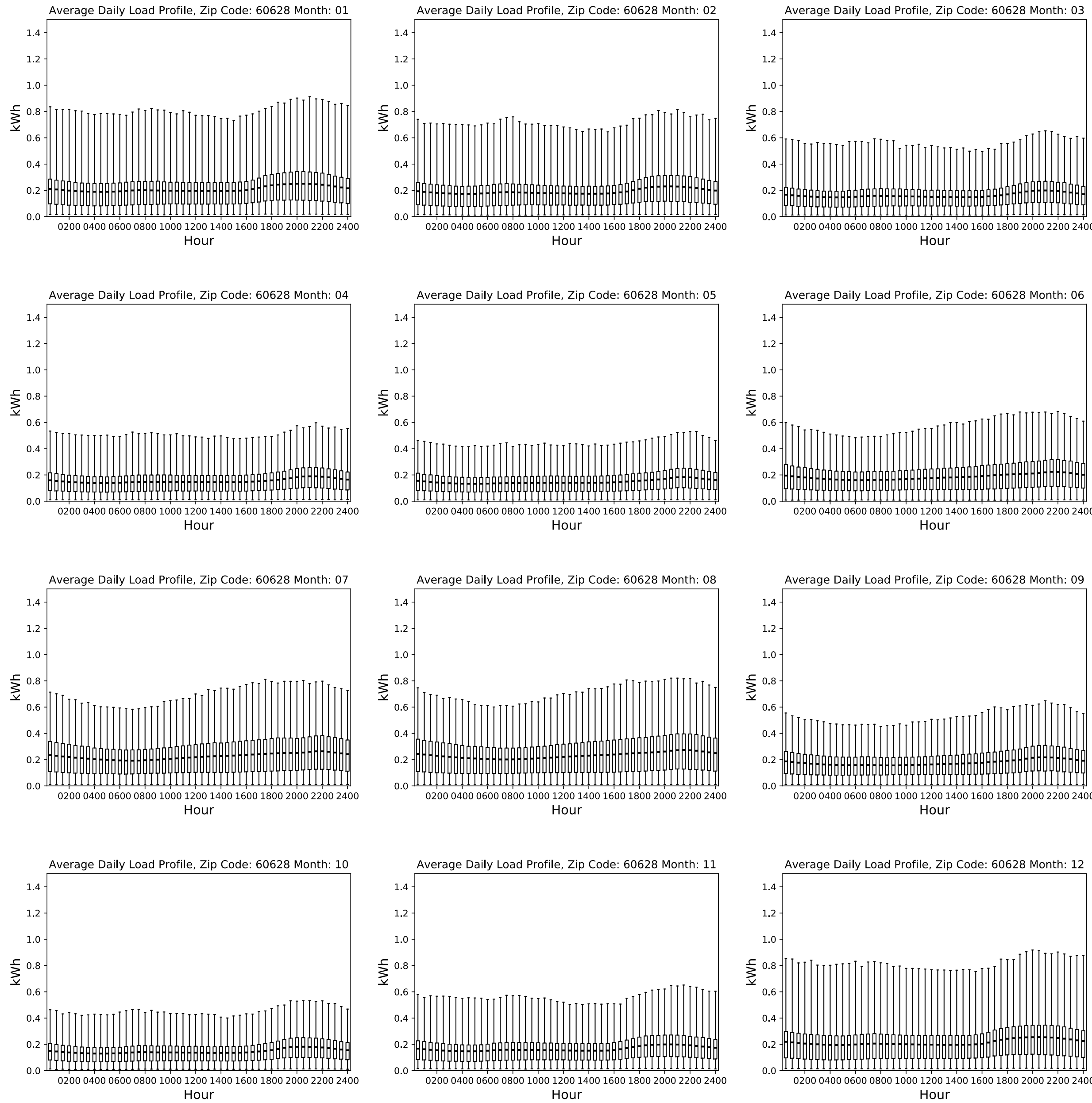

Figure A23. Daily load profiles for multi-family residential electricity consumption by month: 60628.

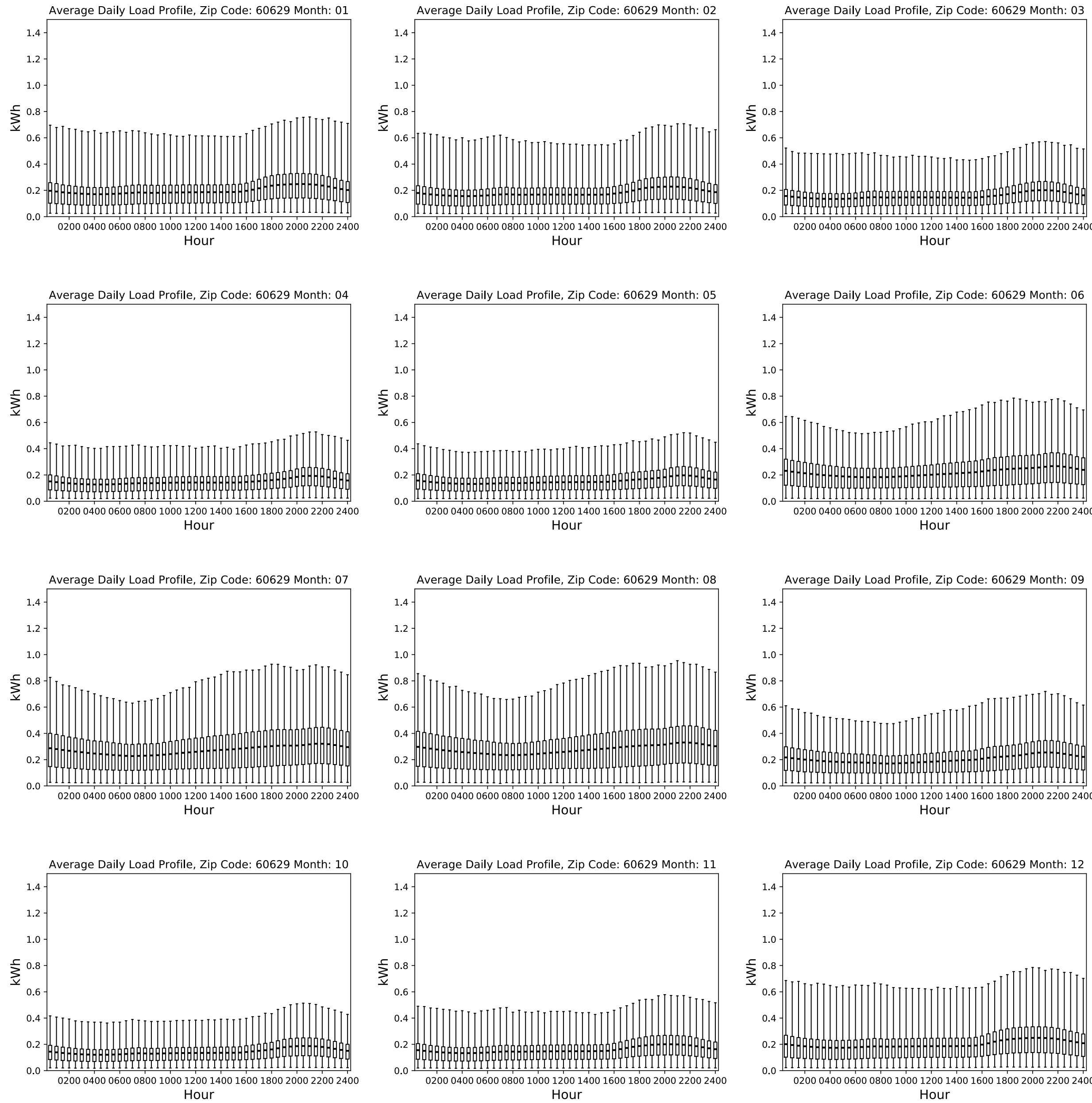

Figure A24. Daily load profiles for multi-family residential electricity consumption by month: 60629.

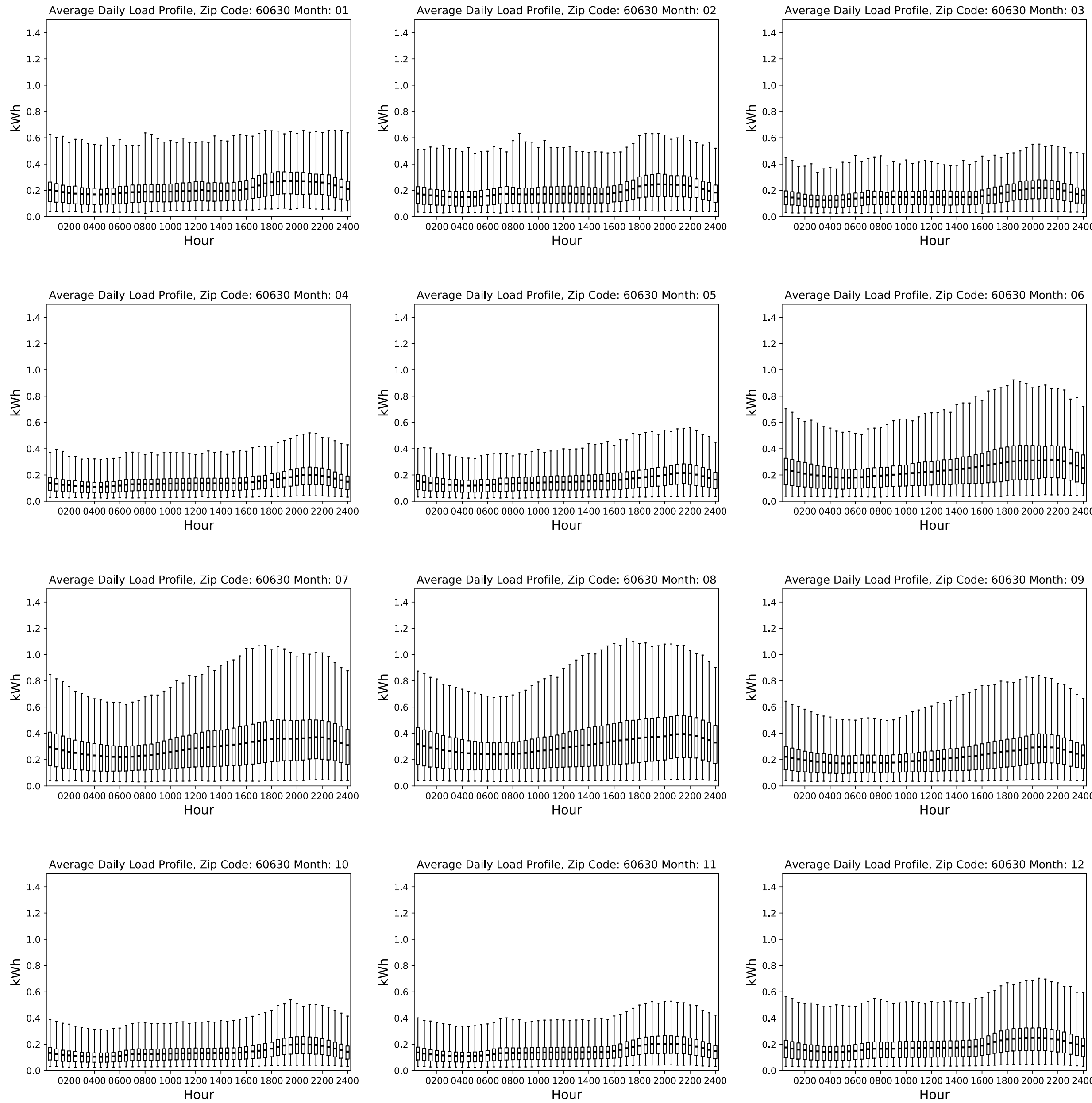

Figure A25. Daily load profiles for multi-family residential electricity consumption by month: 60630.

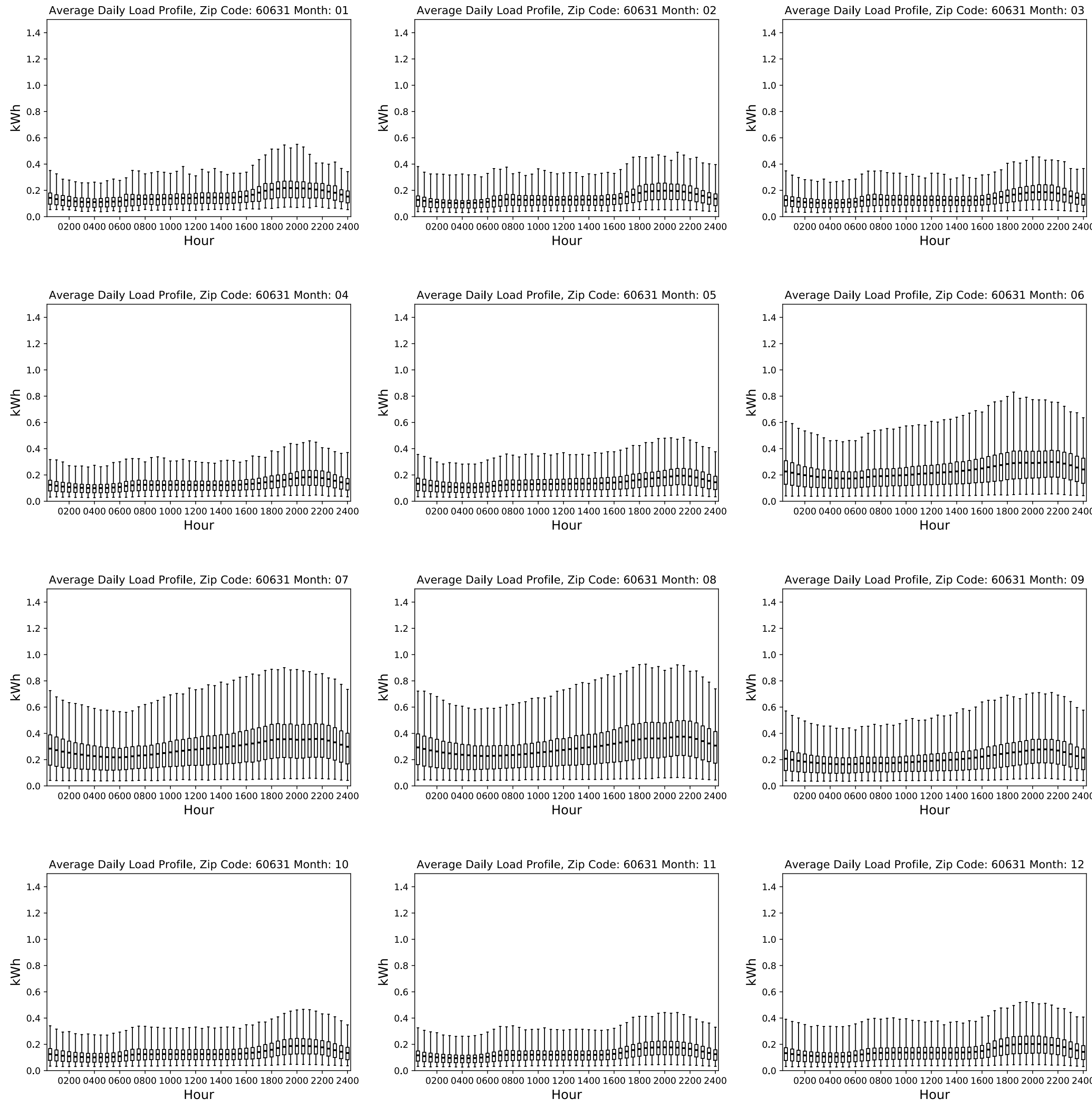

Figure A26. Daily load profiles for multi-family residential electricity consumption by month: 60631.

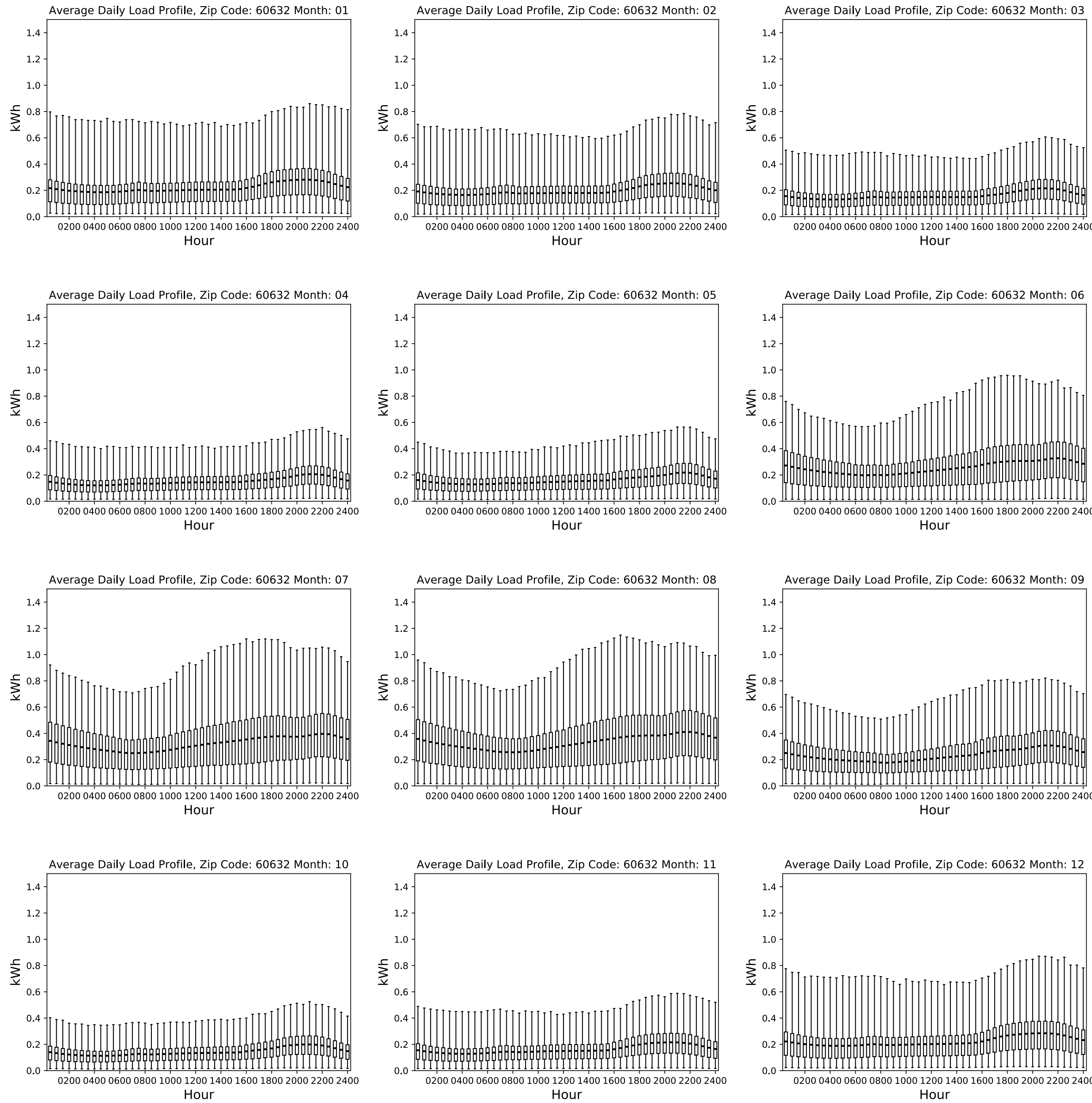

Figure A27. Daily load profiles for multi-family residential electricity consumption by month: 60632.

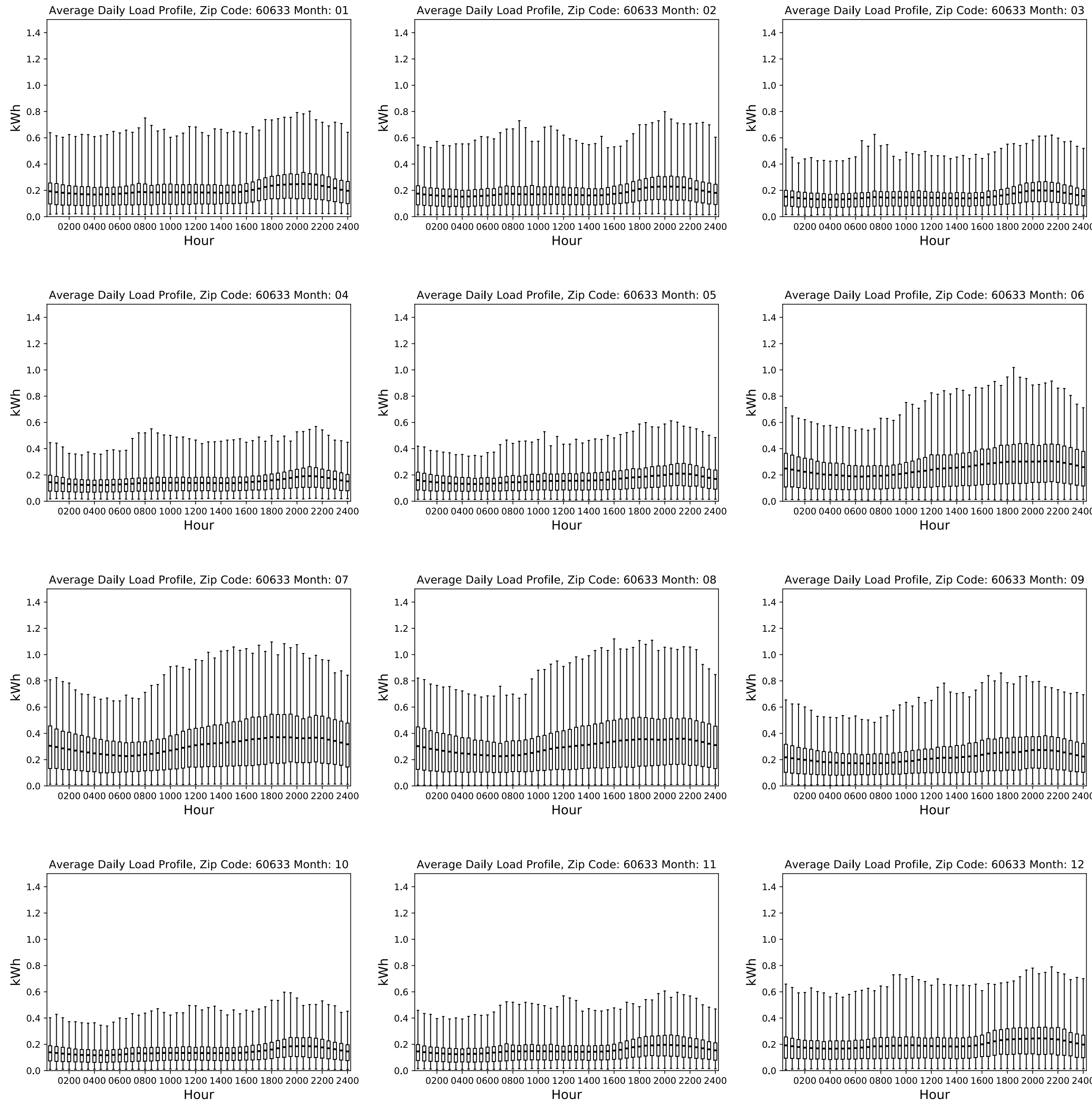

Figure A28. Daily load profiles for multi-family residential electricity consumption by month: 60633.

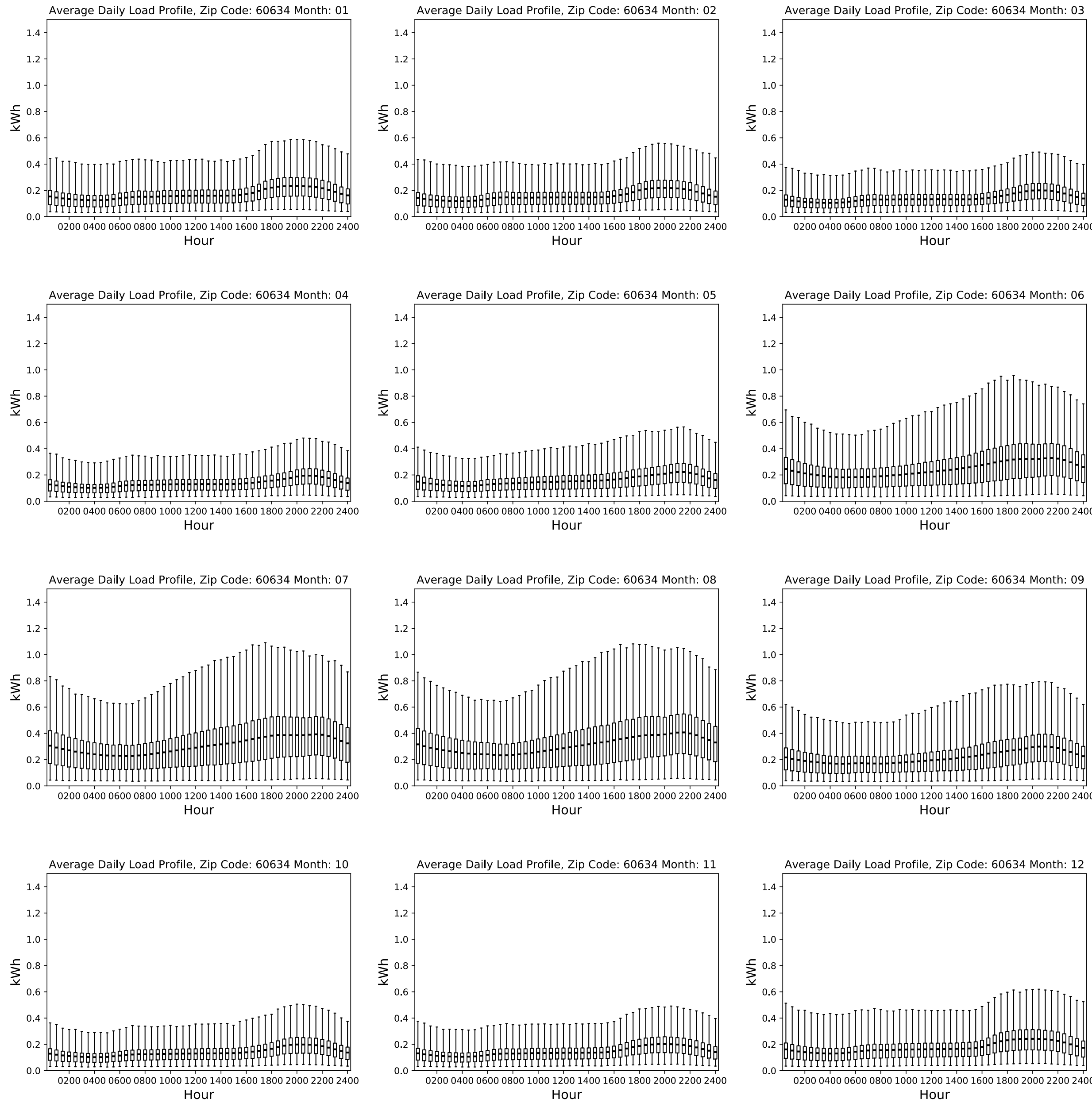

Figure A29. Daily load profiles for multi-family residential electricity consumption by month: 60634.

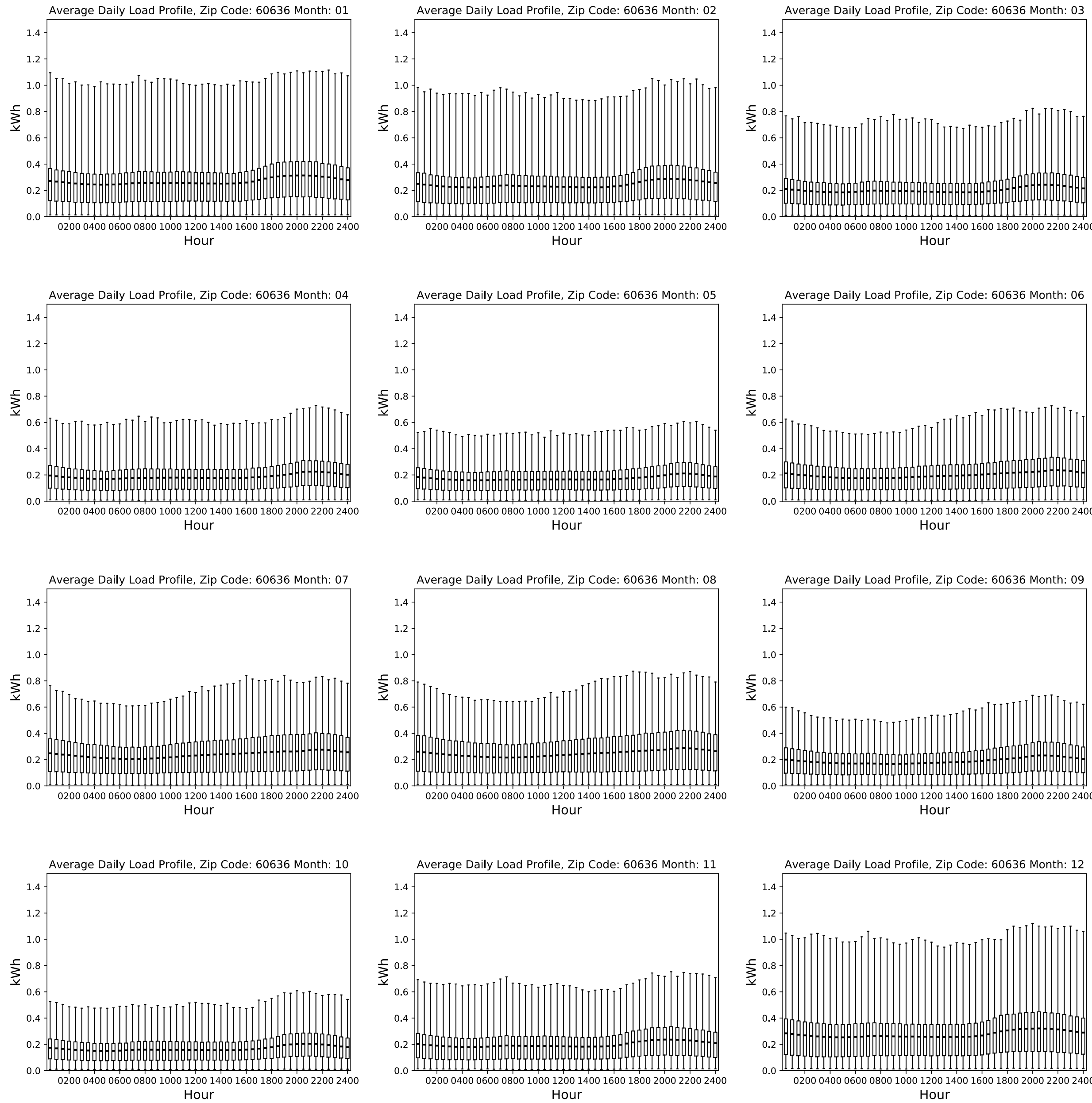

Figure A30. Daily load profiles for multi-family residential electricity consumption by month: 60636.

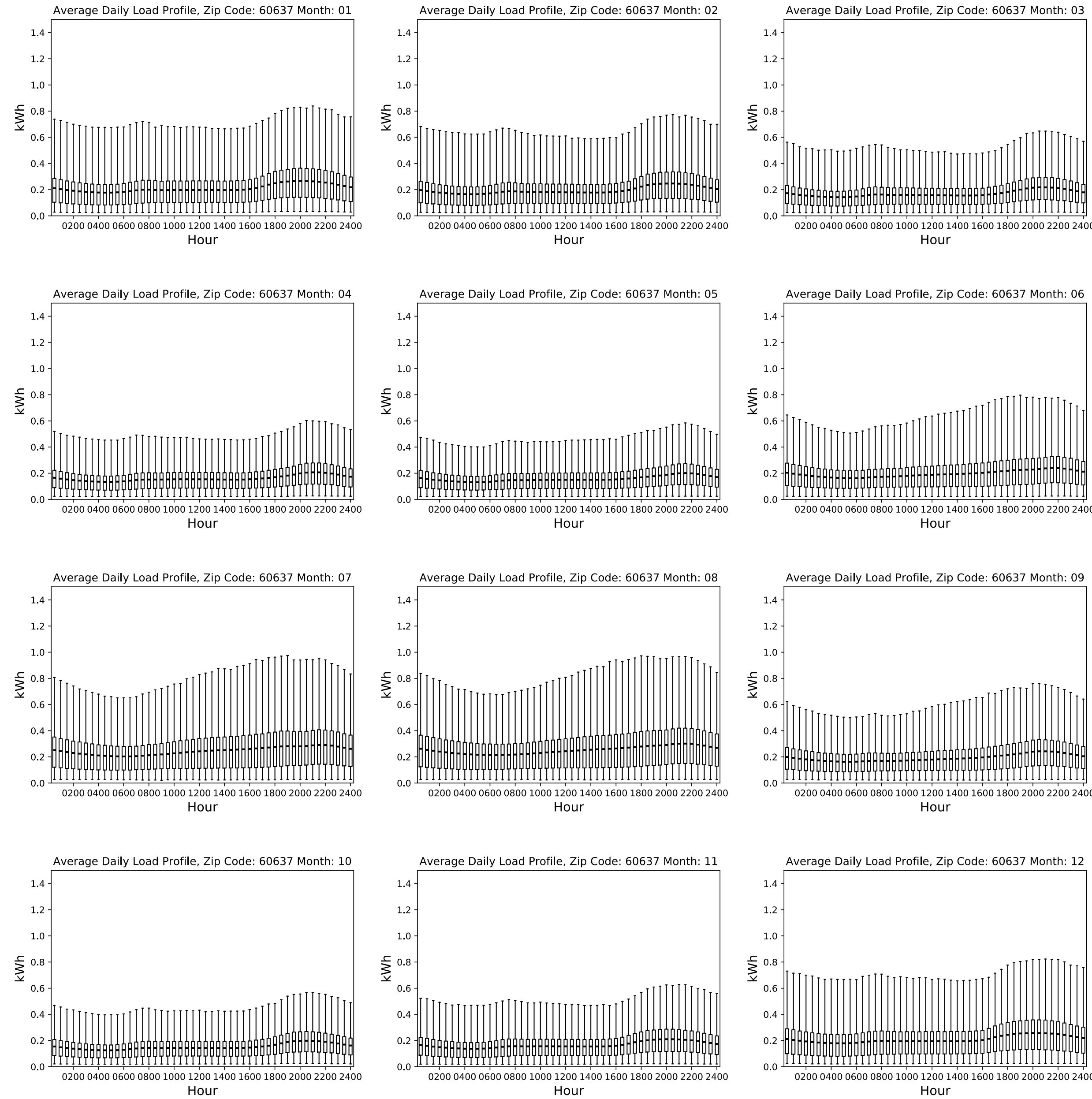

Figure A31. Daily load profiles for multi-family residential electricity consumption by month: 60637.

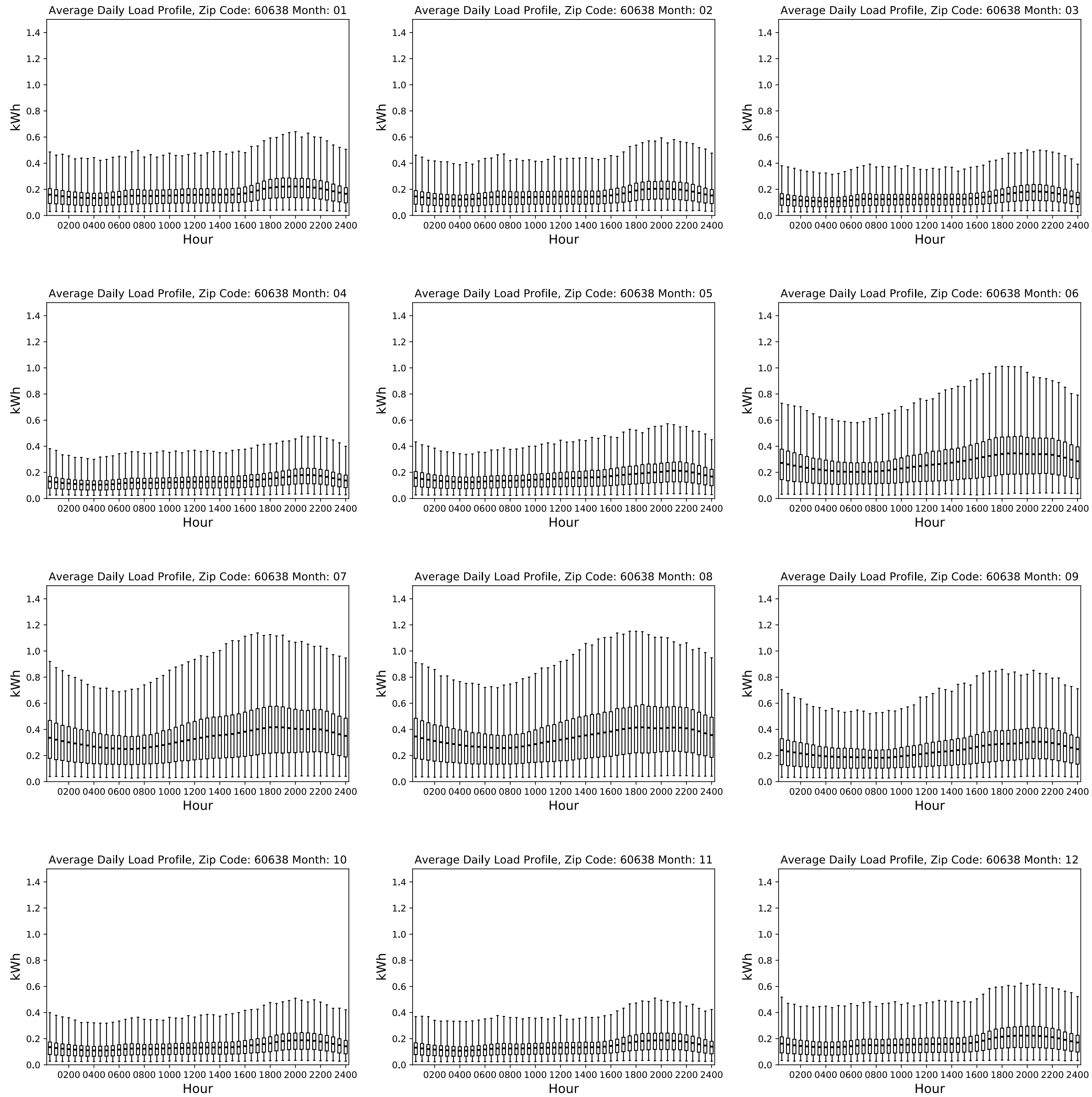

Figure A32. Daily load profiles for multi-family residential electricity consumption by month: 60638.

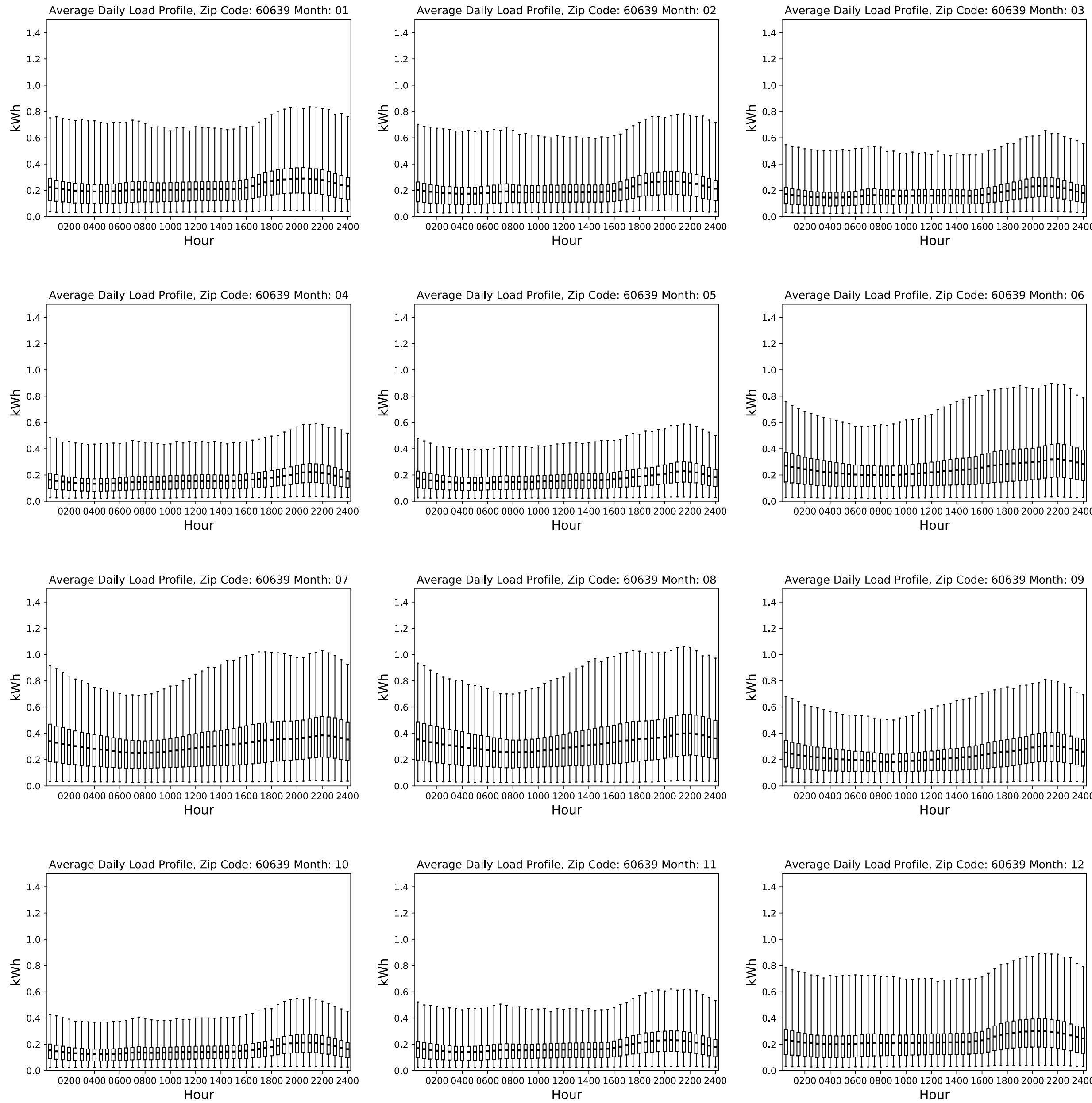

Figure A33. Daily load profiles for multi-family residential electricity consumption by month: 60639.

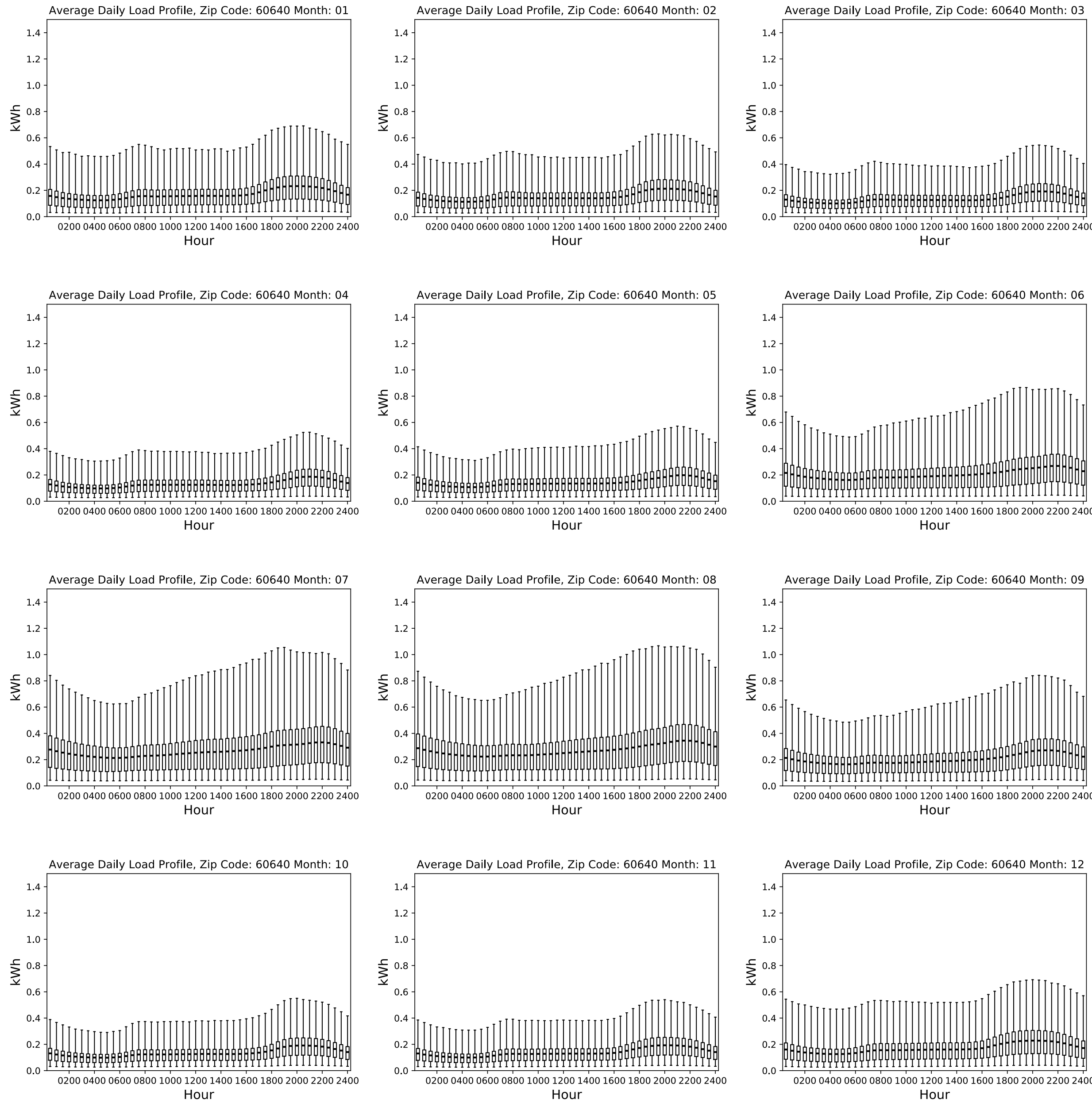

Figure A34. Daily load profiles for multi-family residential electricity consumption by month: 60640.

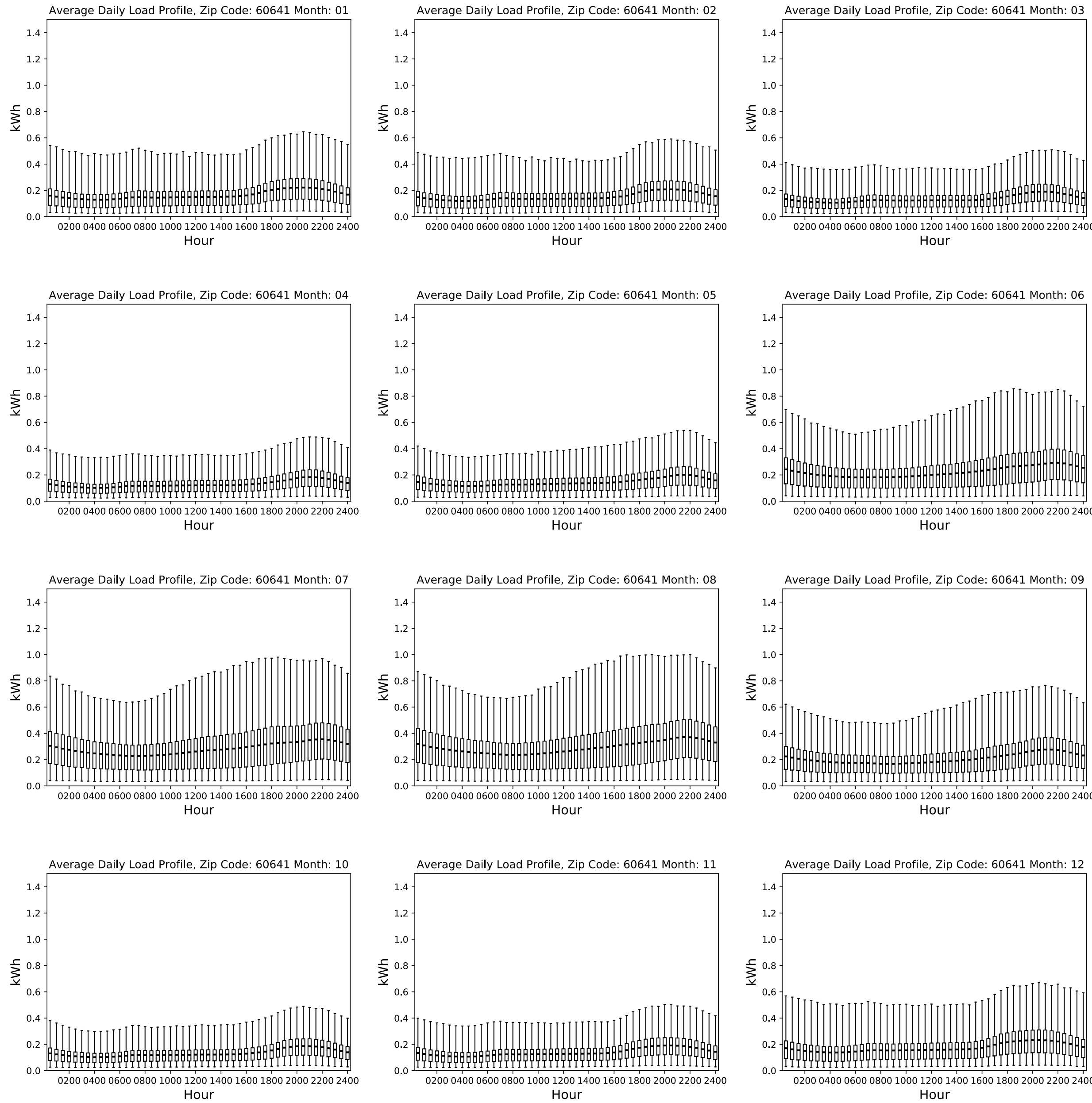

Figure A35. Daily load profiles for multi-family residential electricity consumption by month: 60641.

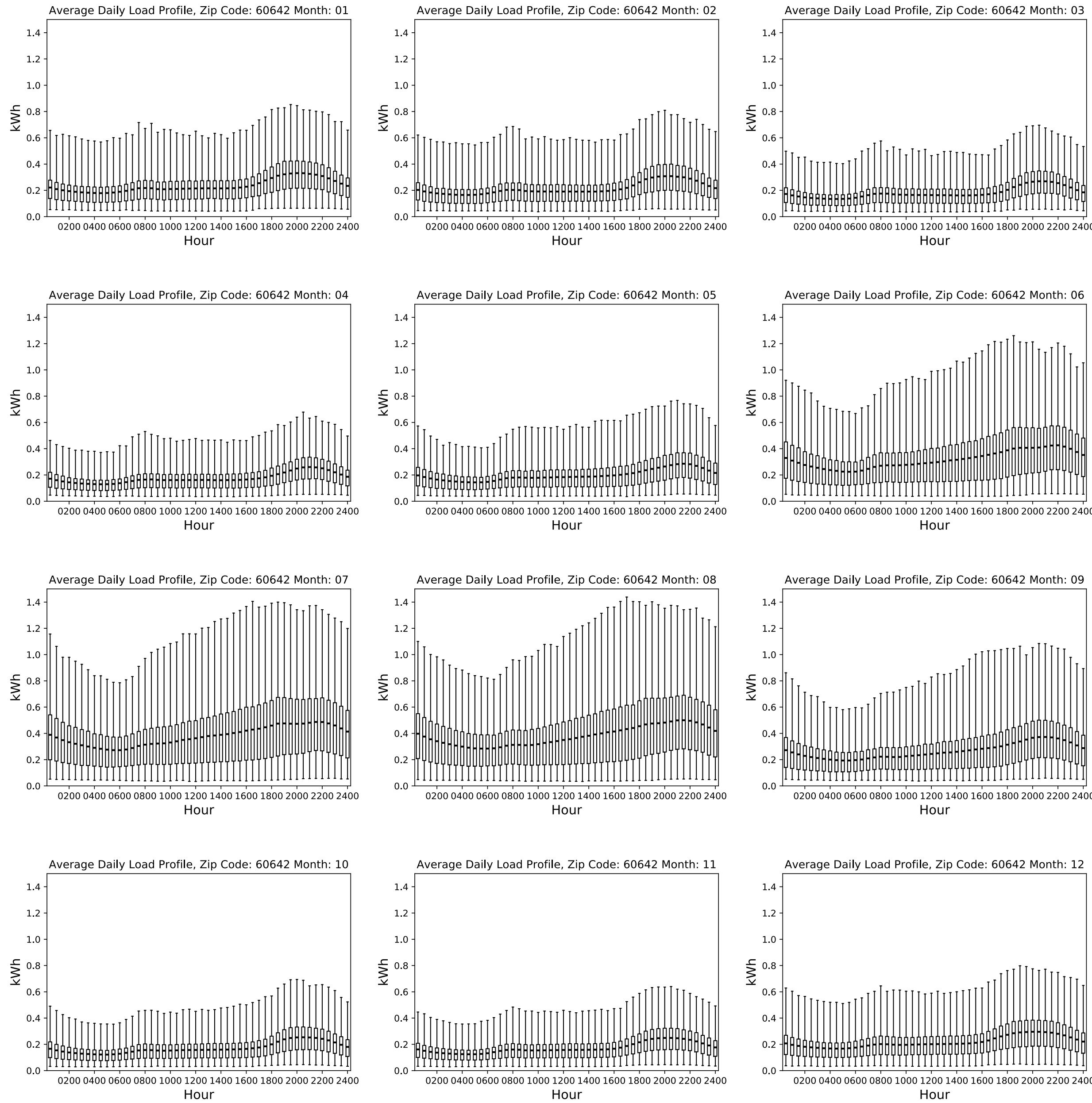

Figure A36. Daily load profiles for multi-family residential electricity consumption by month: 60642.

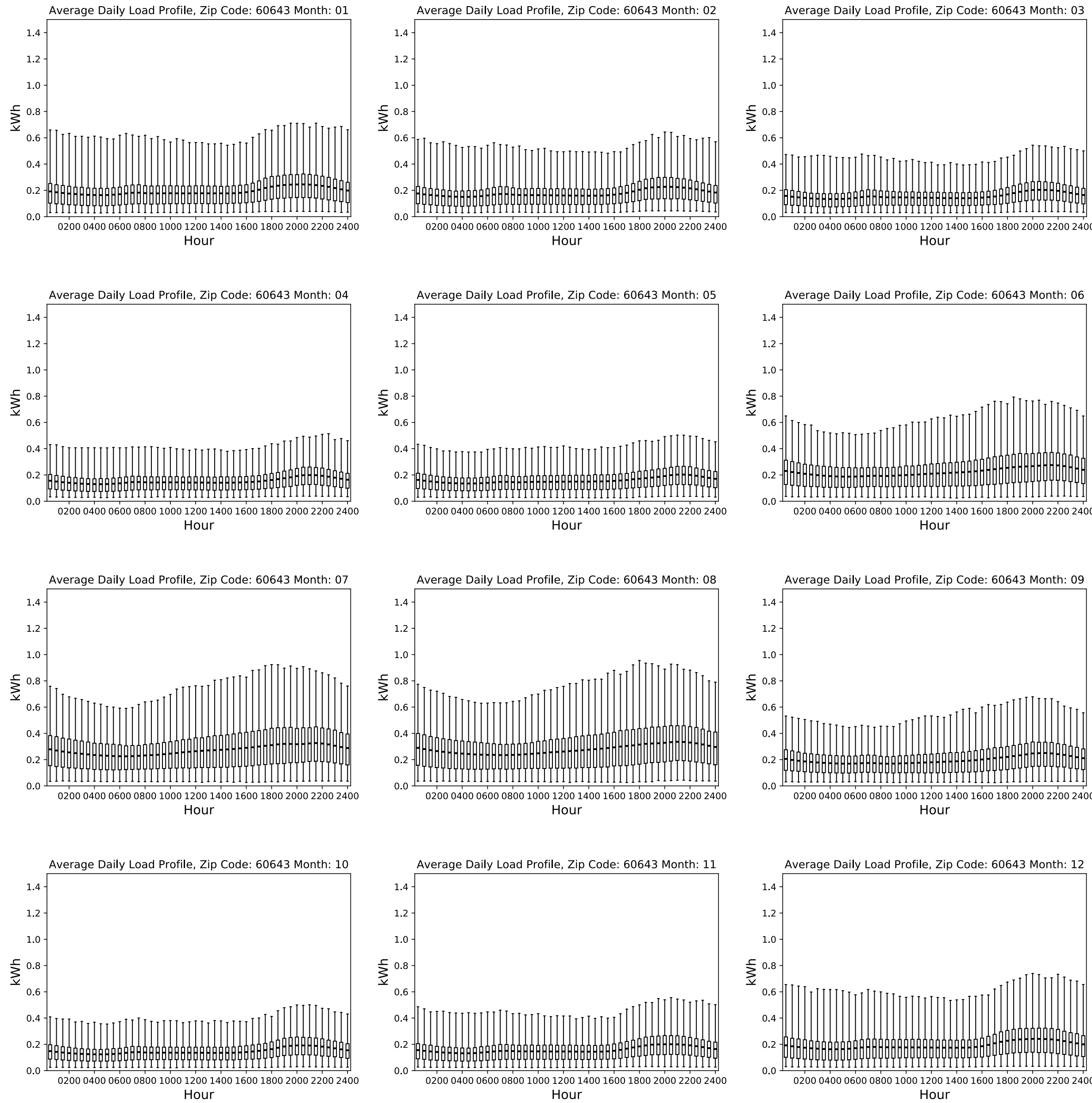

Figure A37. Daily load profiles for multi-family residential electricity consumption by month: 60643.

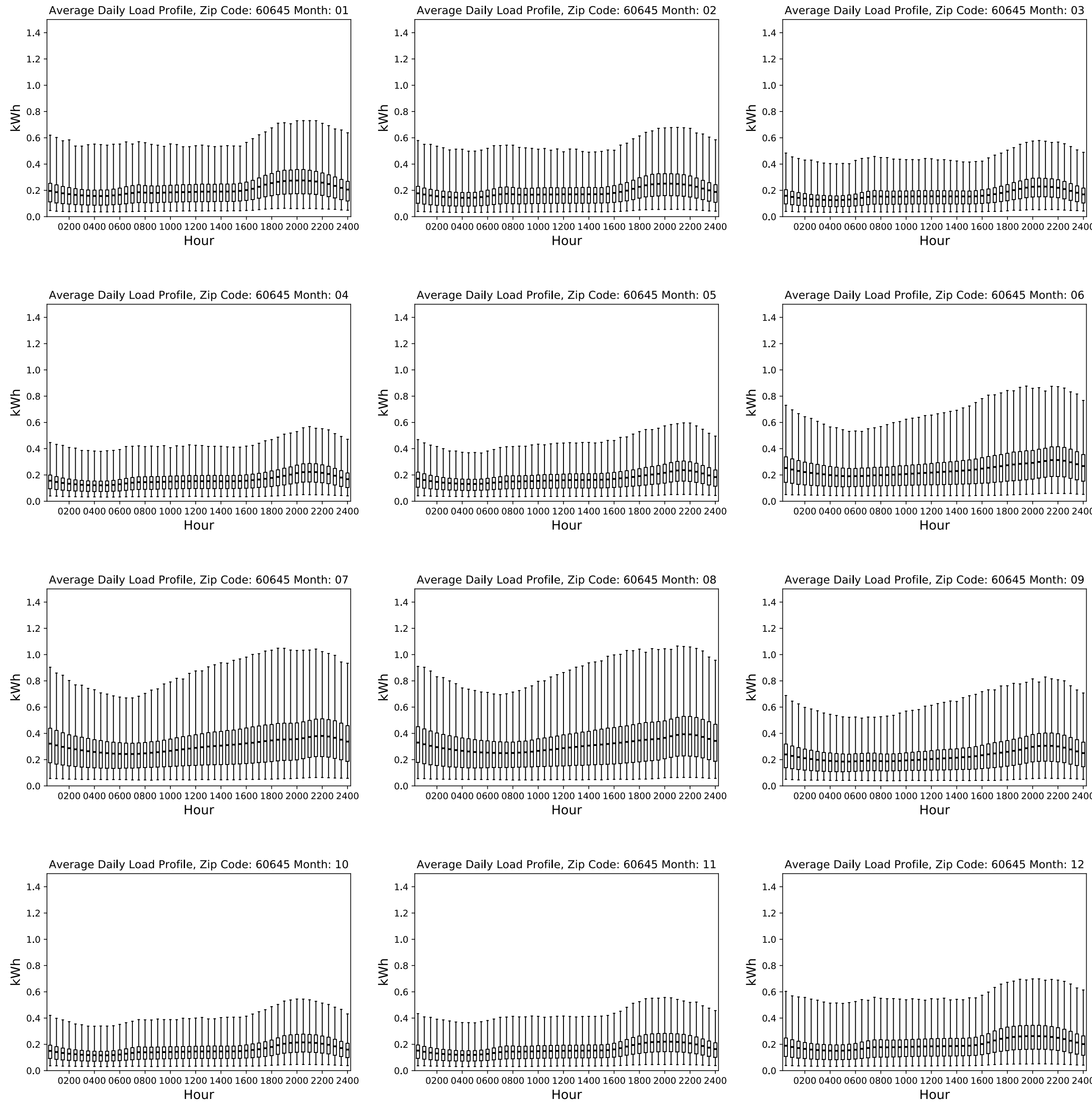

Figure A38. Daily load profiles for multi-family residential electricity consumption by month: 60645.

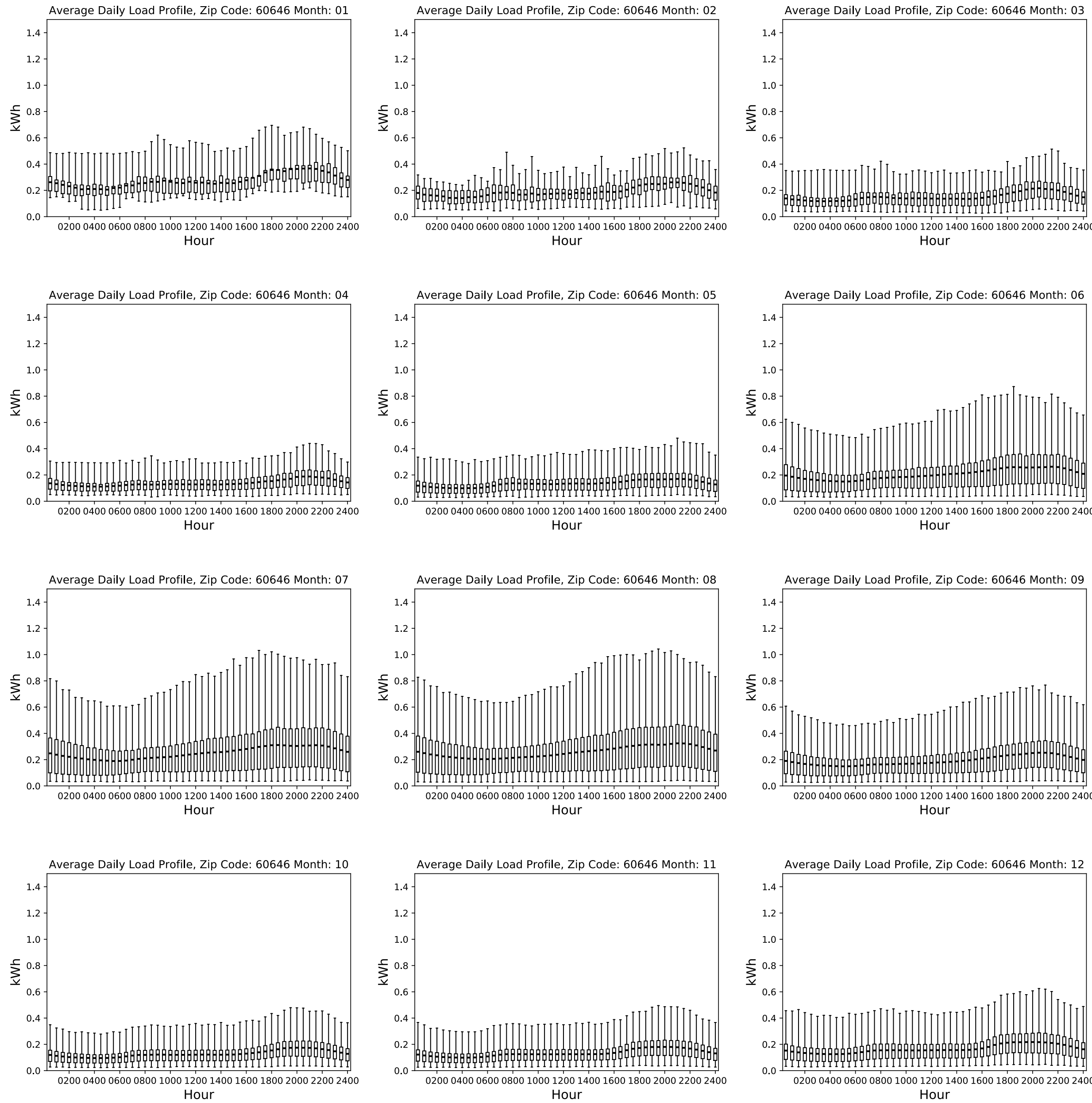

Figure A39. Daily load profiles for multi-family residential electricity consumption by month: 60646.

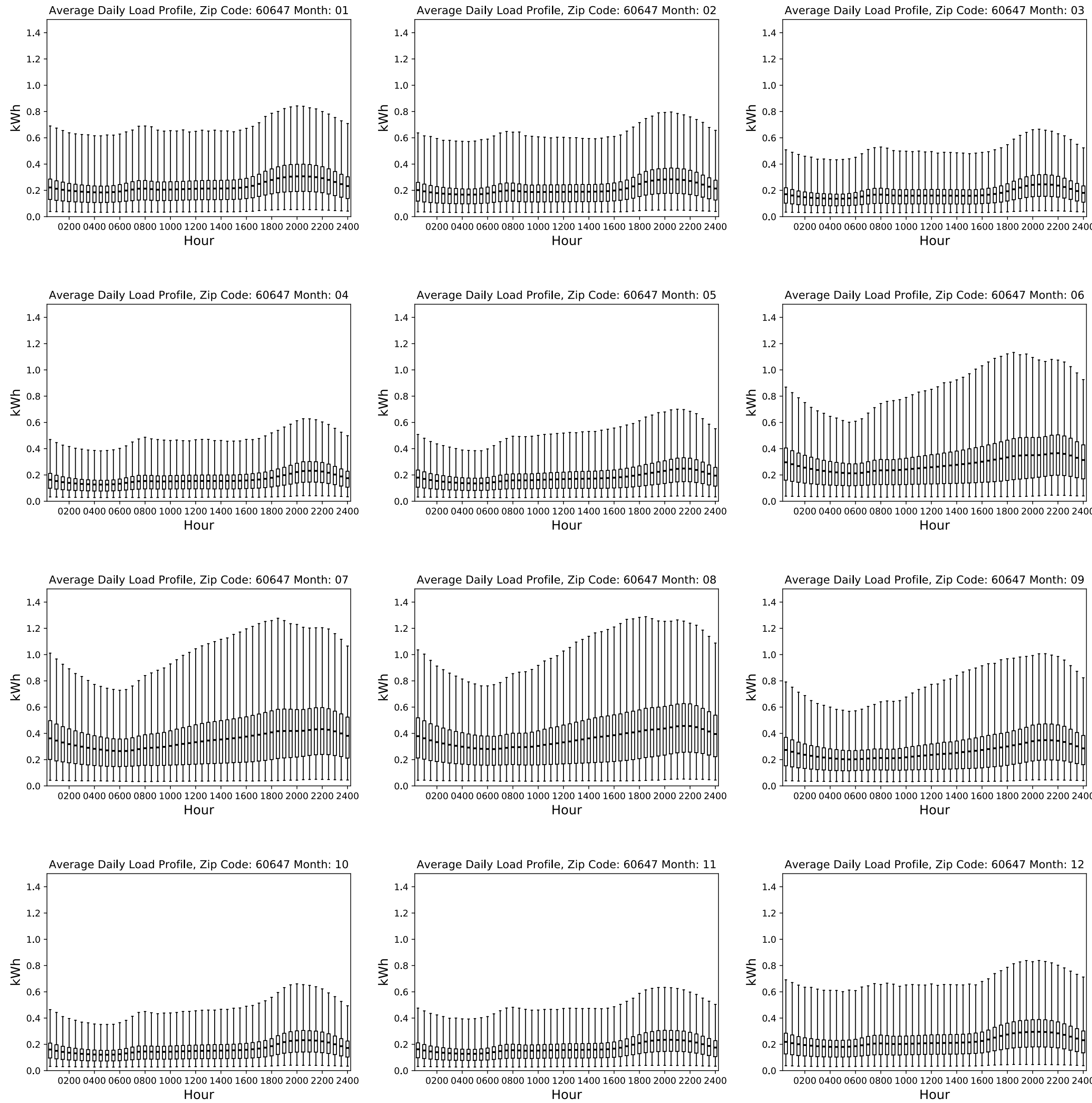

Figure A40. Daily load profiles for multi-family residential electricity consumption by month: 60647.

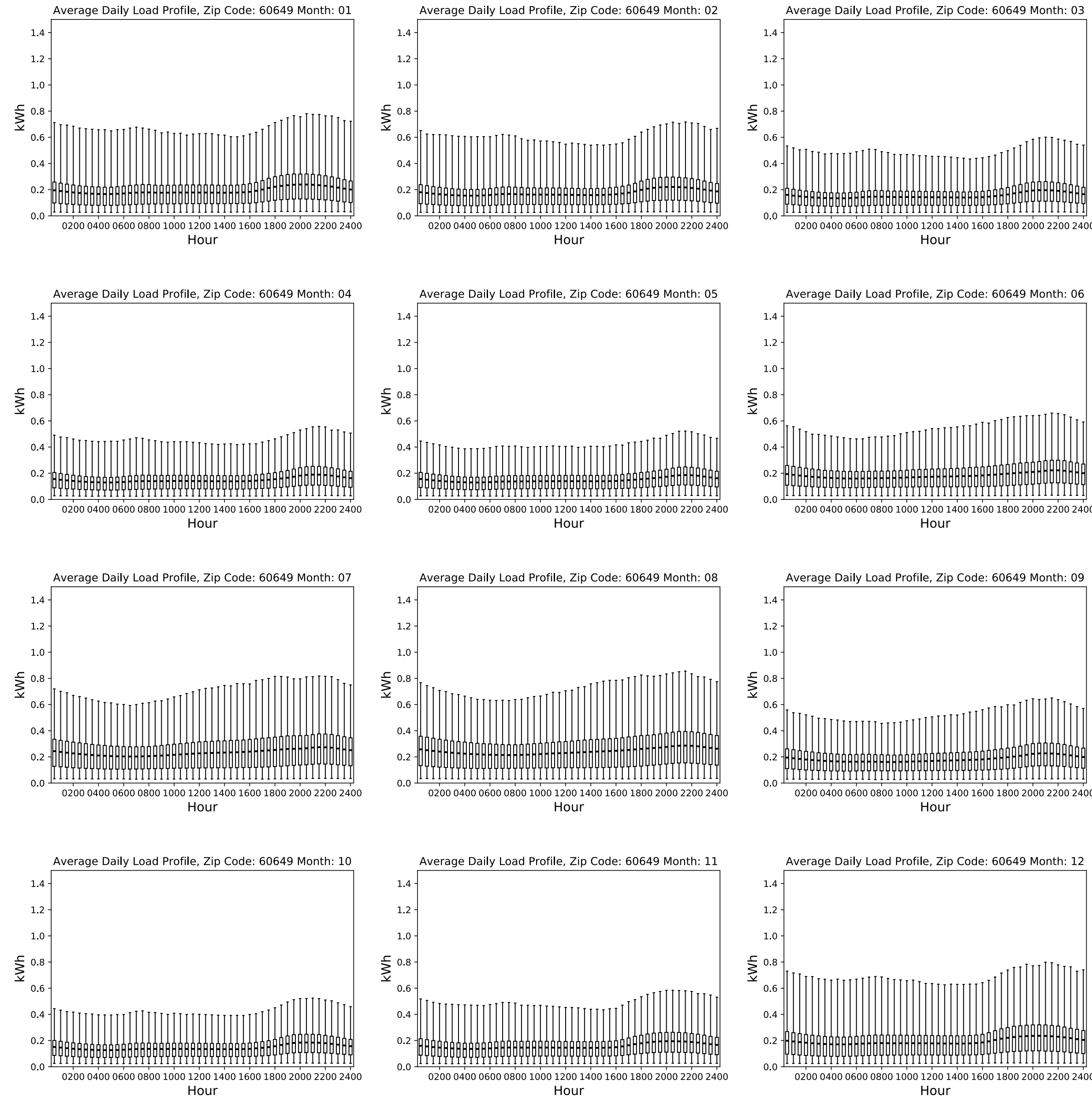

Figure A41. Daily load profiles for multi-family residential electricity consumption by month: 60649.

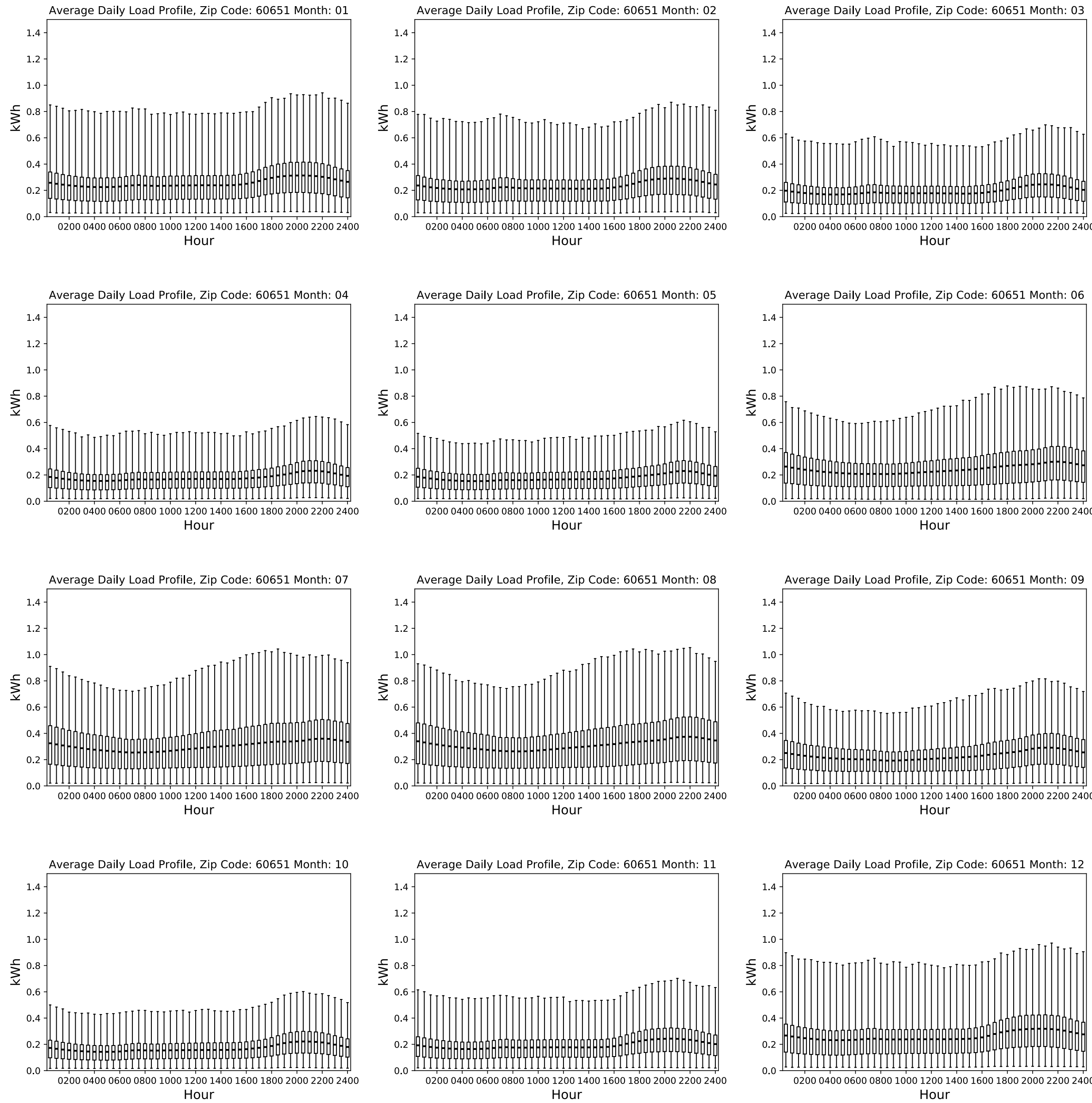

Figure A42. Daily load profiles for multi-family residential electricity consumption by month: 60651.

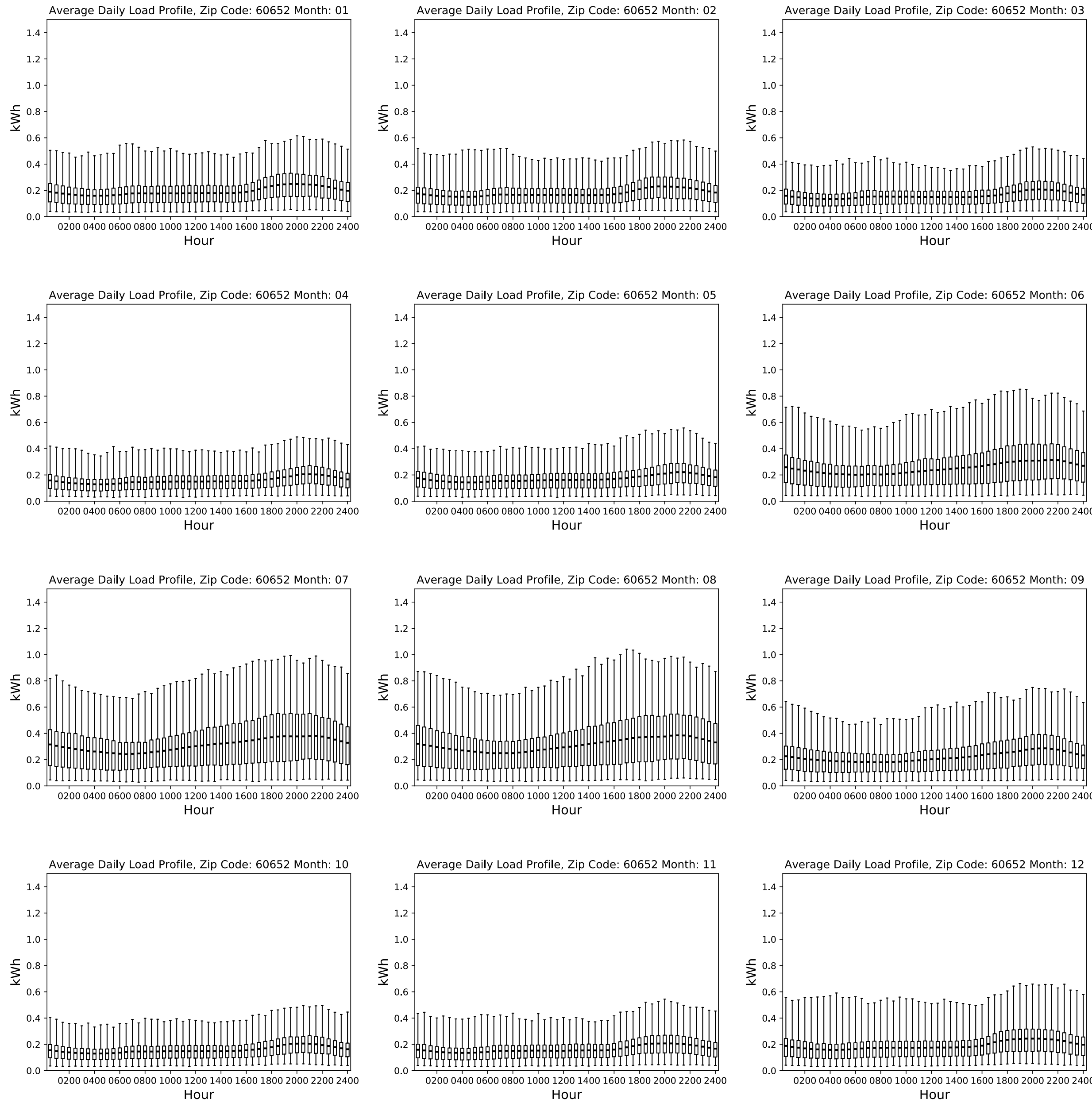

Figure A43. Daily load profiles for multi-family residential electricity consumption by month: 60652.

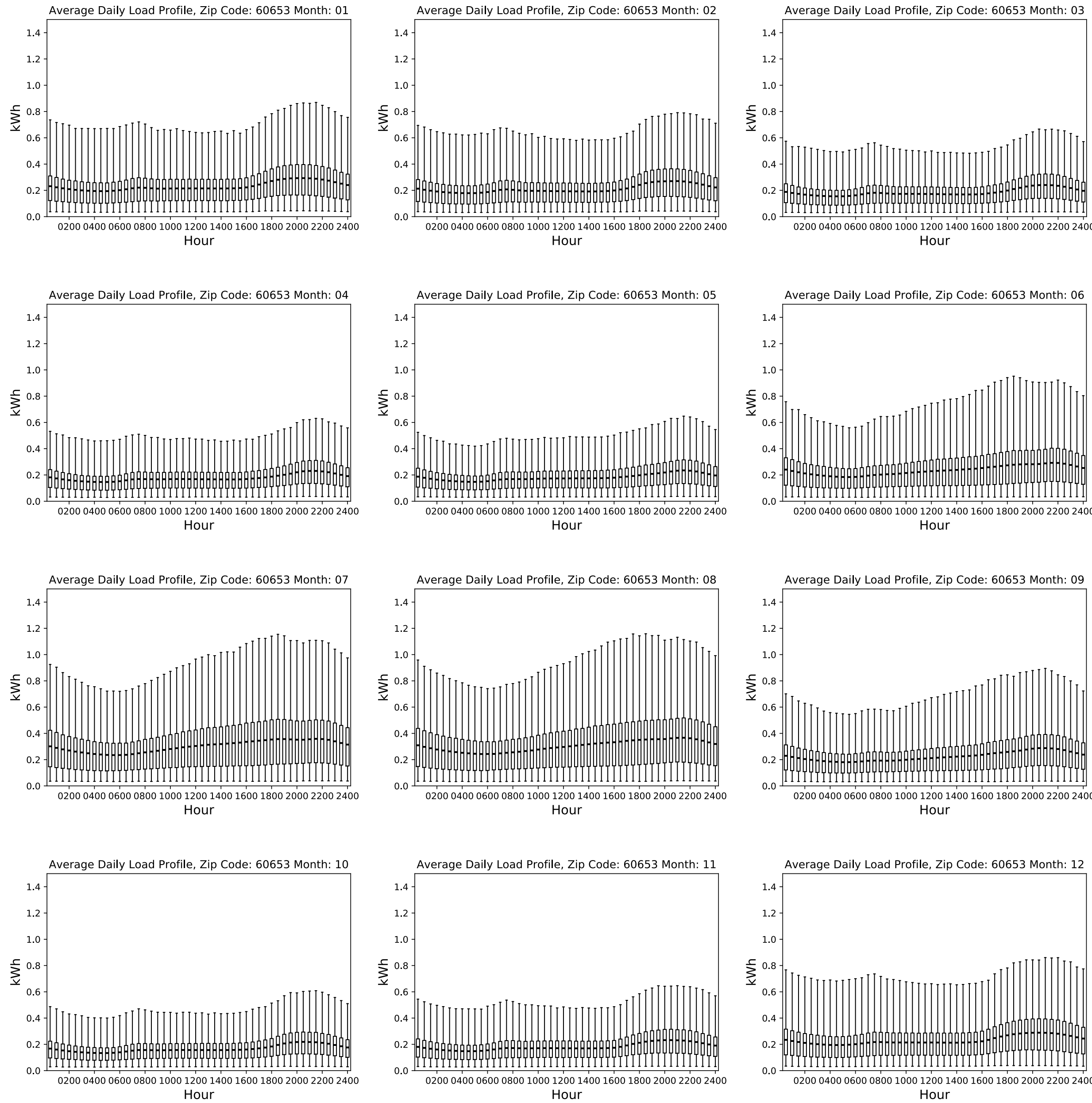

Figure A44. Daily load profiles for multi-family residential electricity consumption by month: 60653.

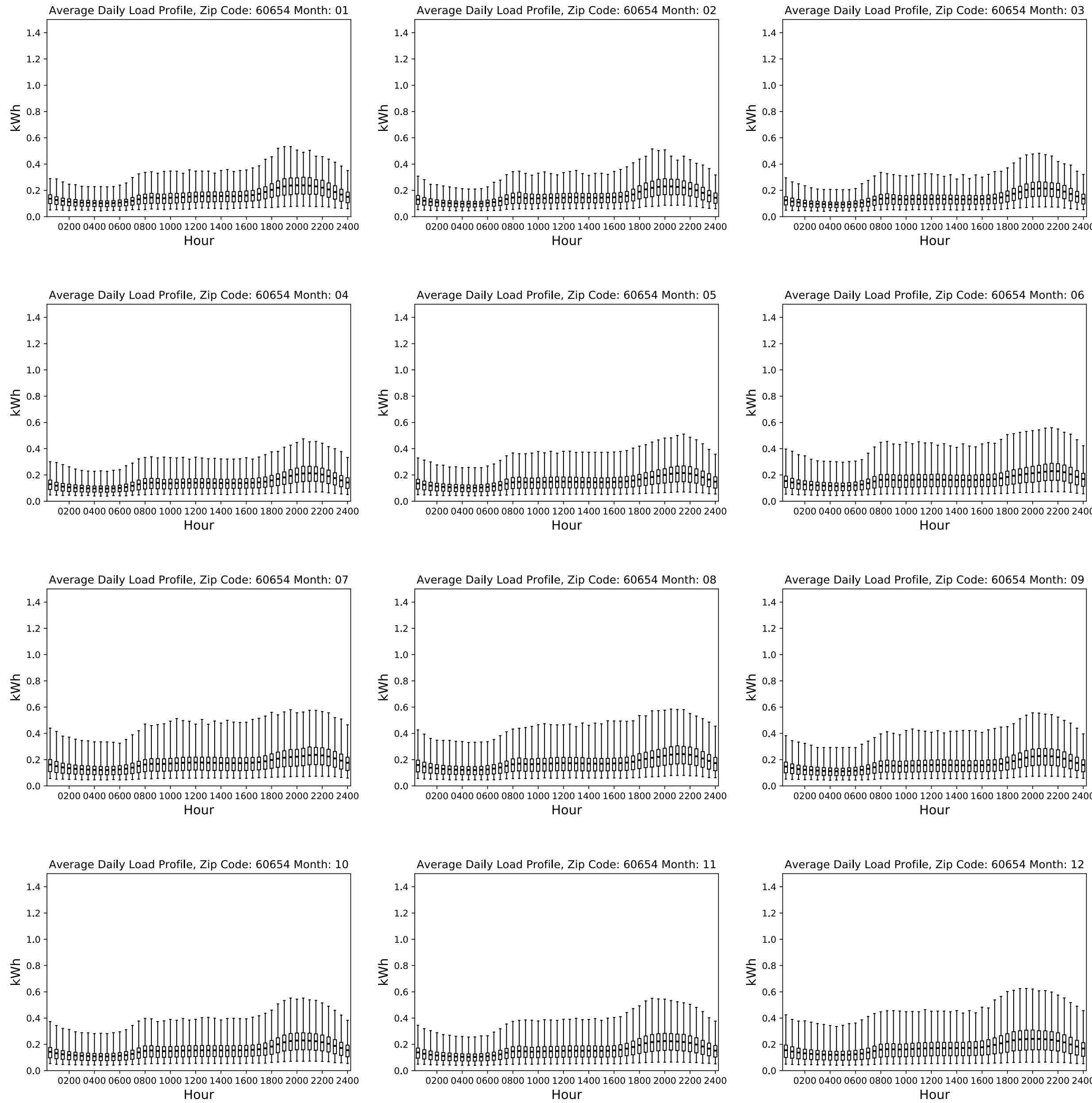

Figure A45. Daily load profiles for multi-family residential electricity consumption by month: 60654.

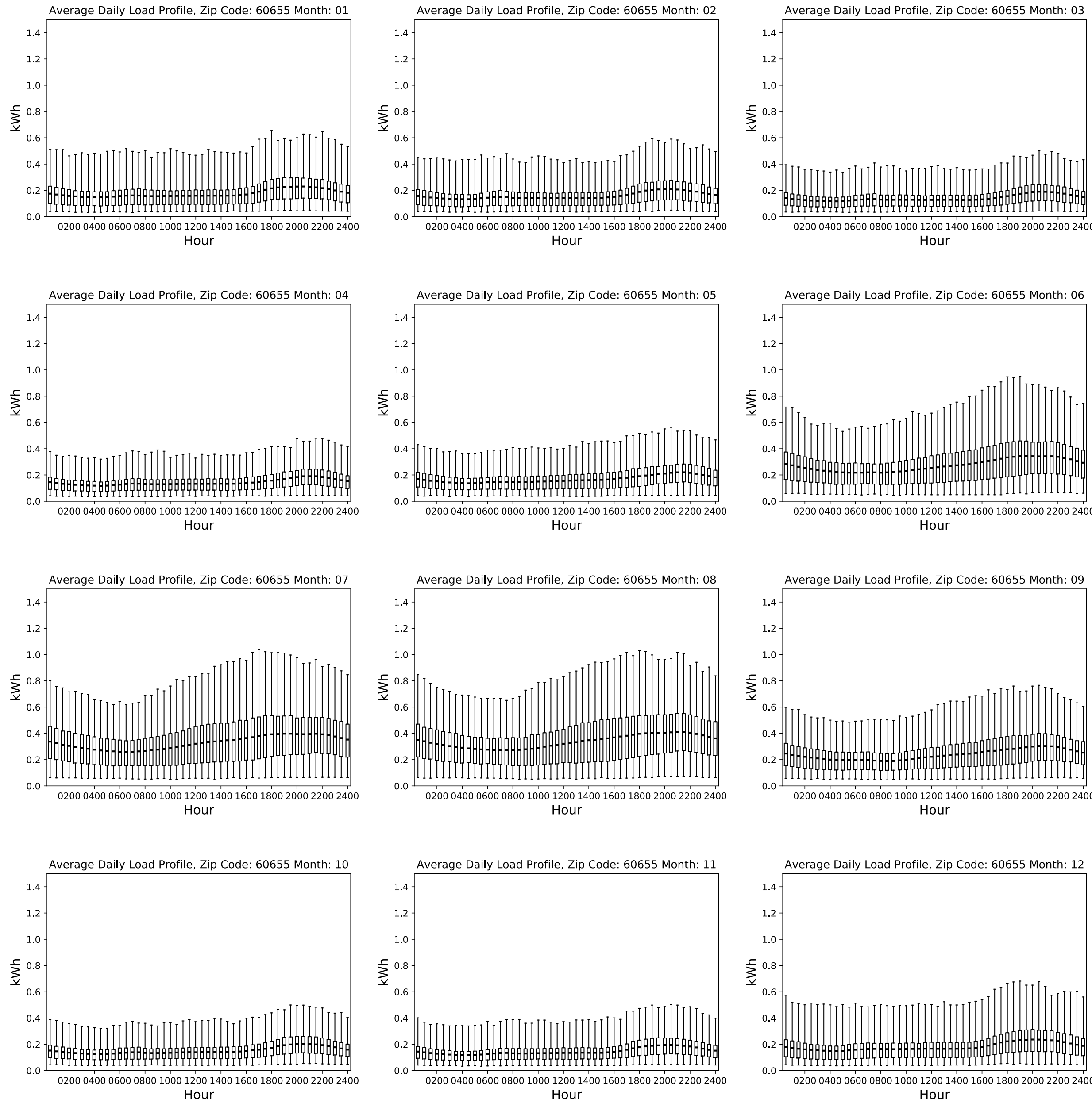

Figure A46. Daily load profiles for multi-family residential electricity consumption by month: 60655.

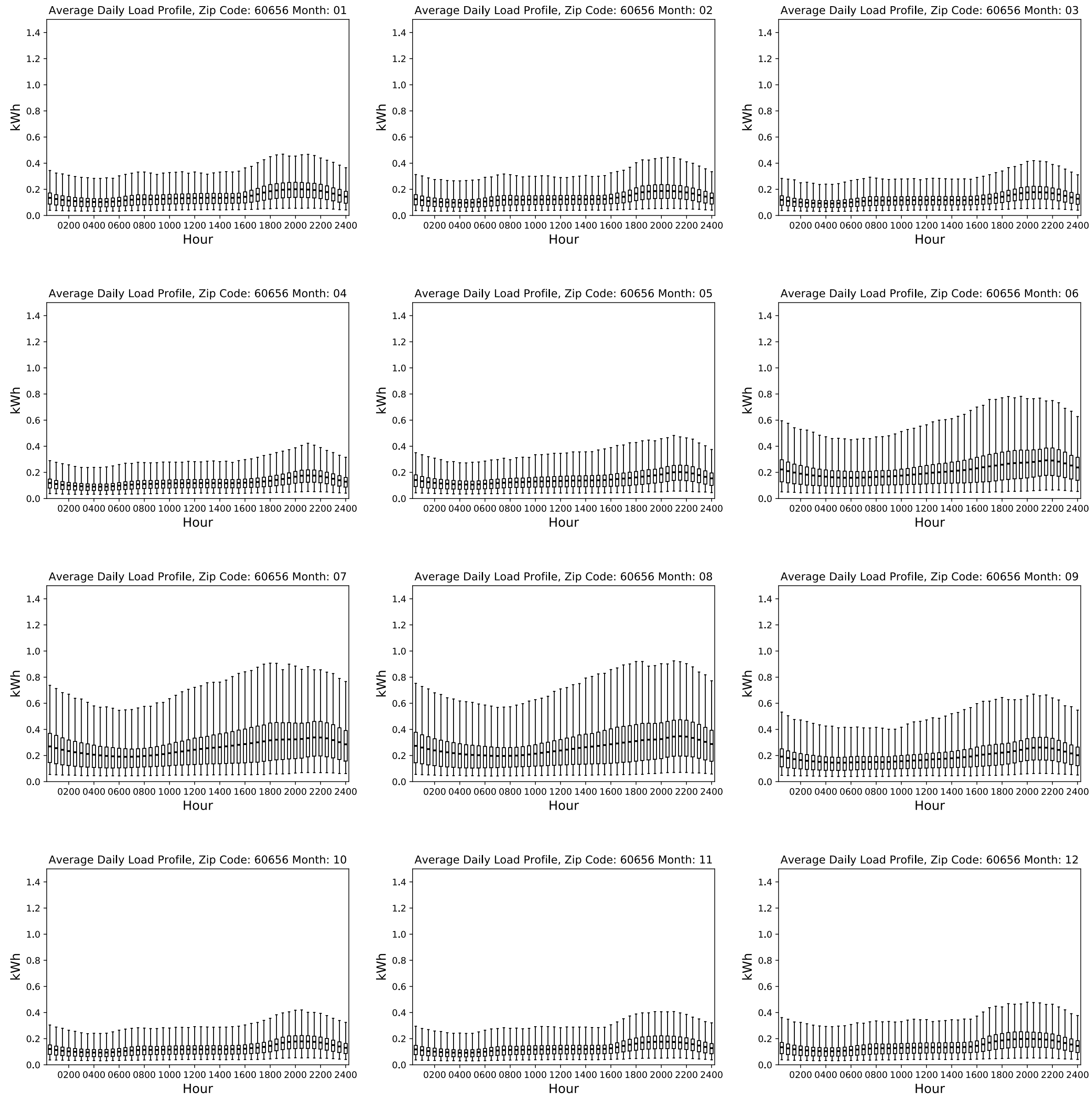

Figure A47. Daily load profiles for multi-family residential electricity consumption by month: 60656.

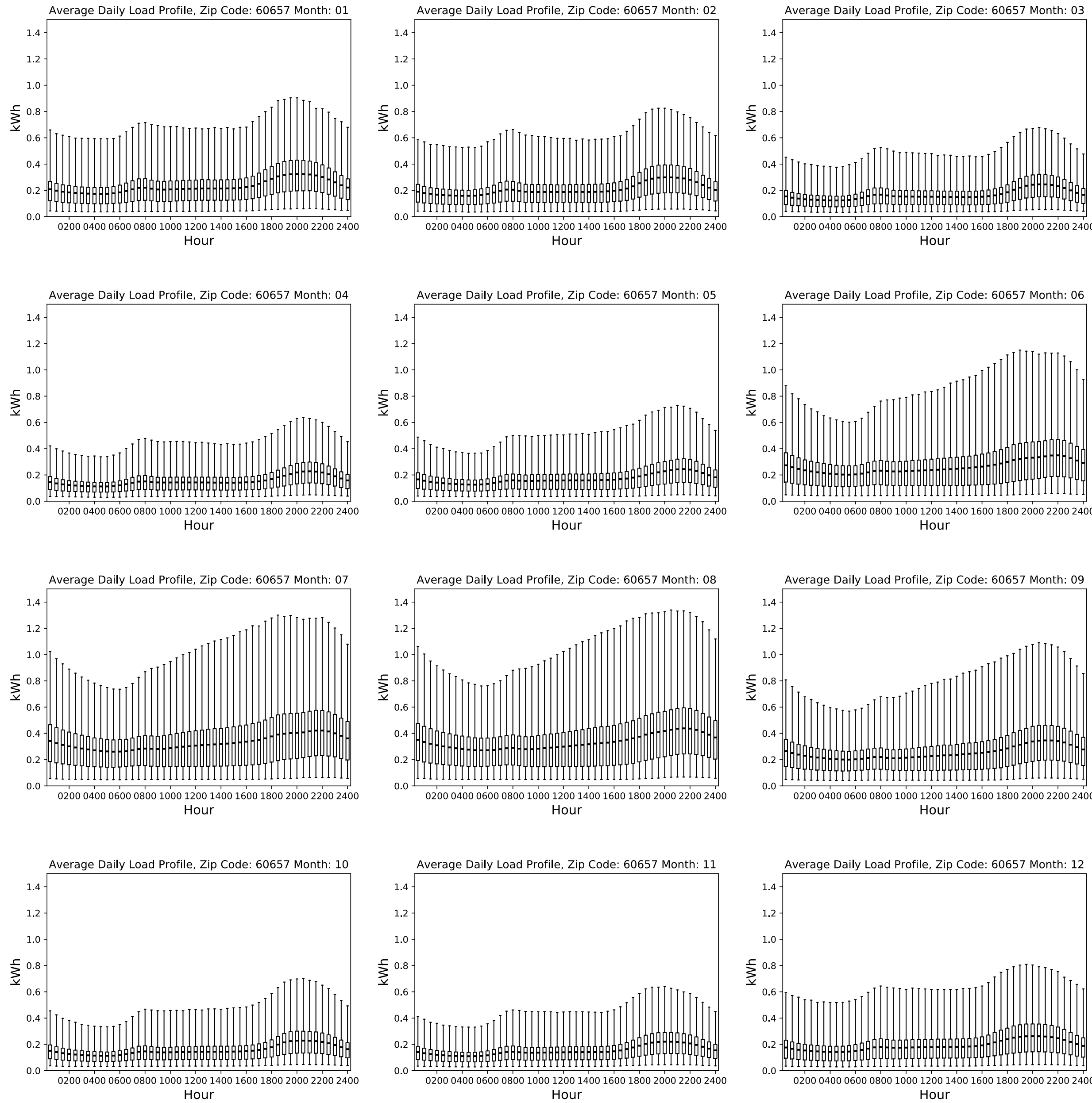

Figure A48. Daily load profiles for multi-family residential electricity consumption by month: 60657.

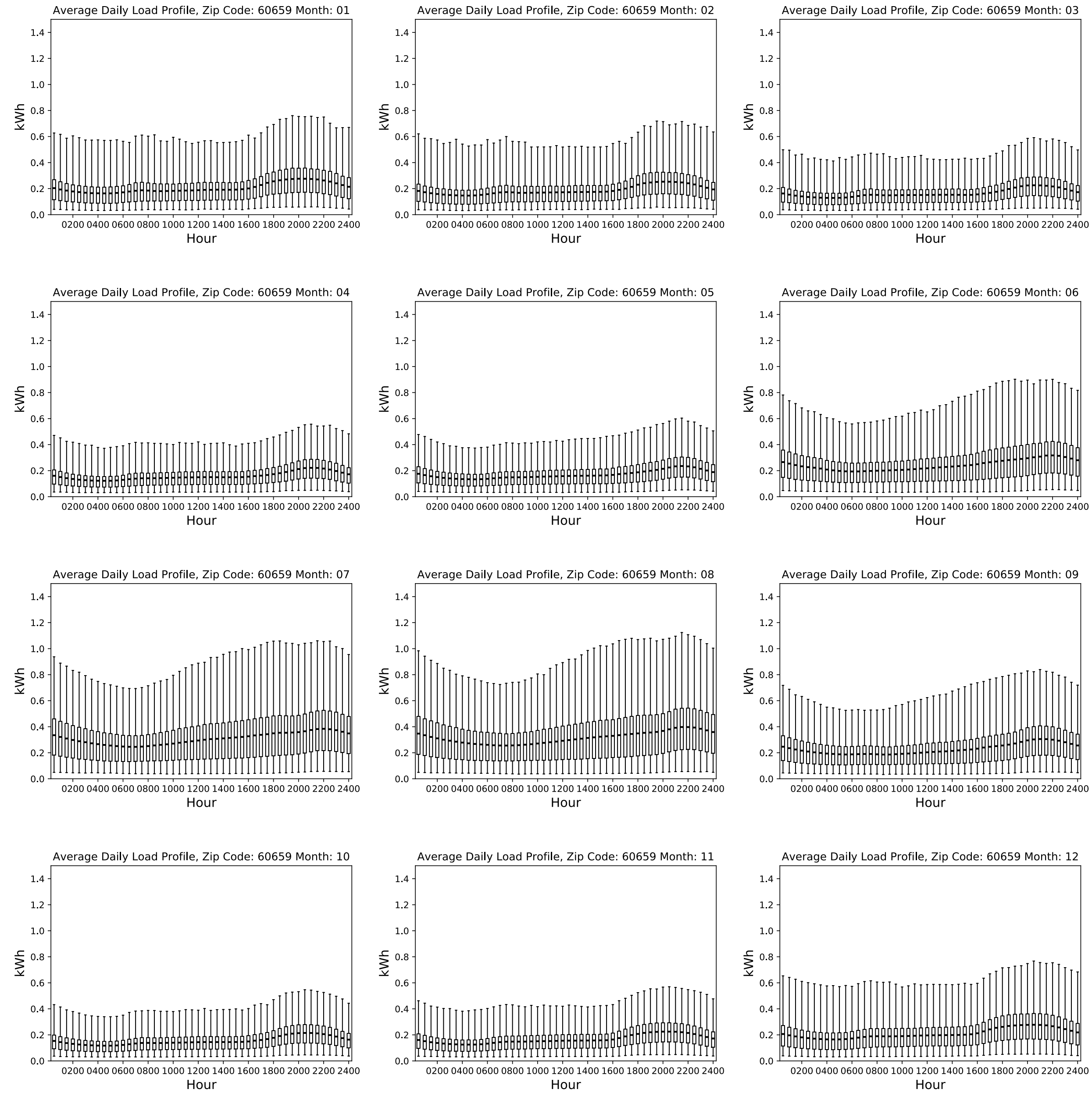

Figure A49. Daily load profiles for multi-family residential electricity consumption by month: 60659.

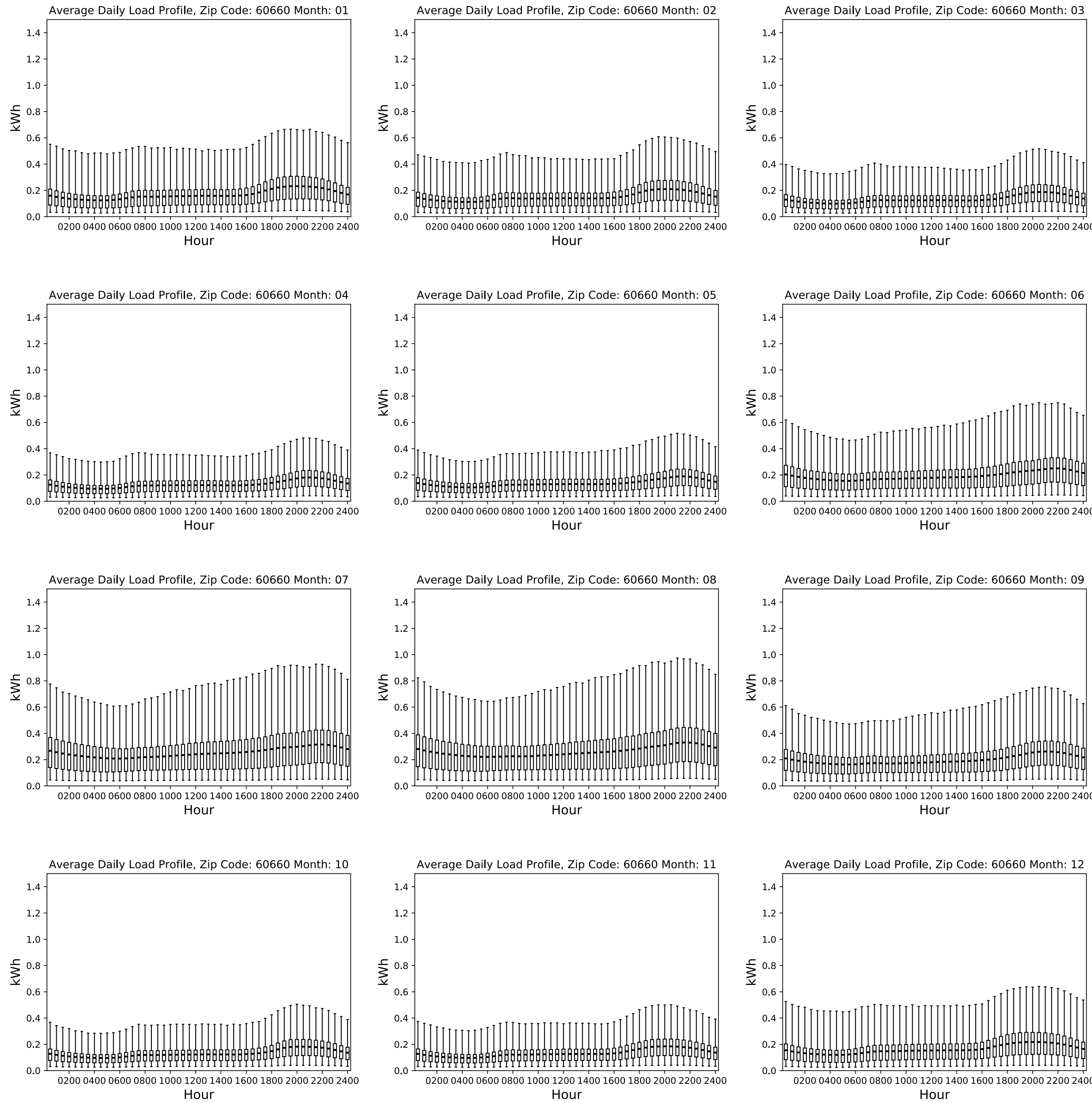

Figure A50. Daily load profiles for multi-family residential electricity consumption by month: 60660.

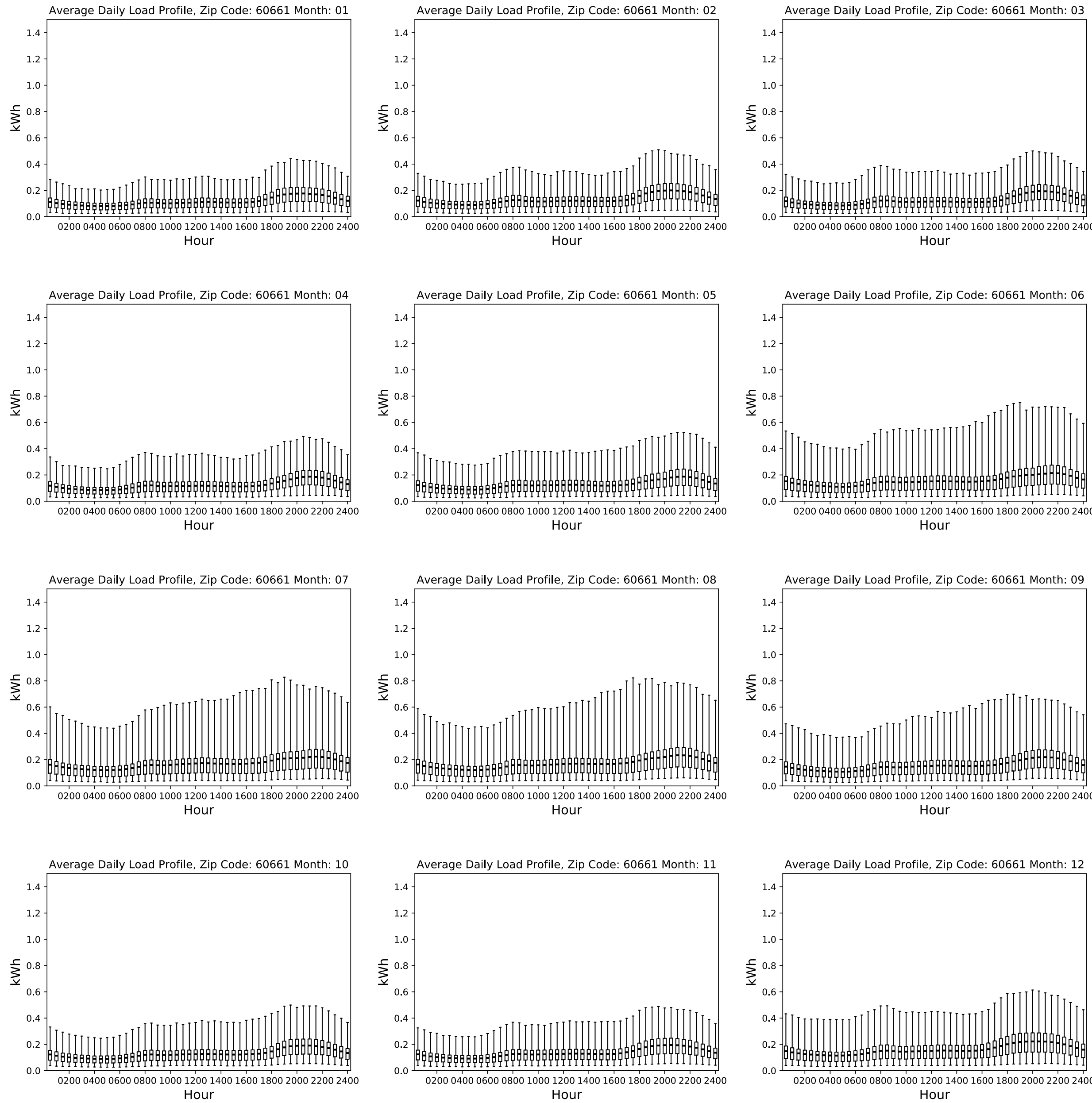

Figure A51. Daily load profiles for multi-family residential electricity consumption by month: 60661.

Zip Codes with Incomplete Data for 2016

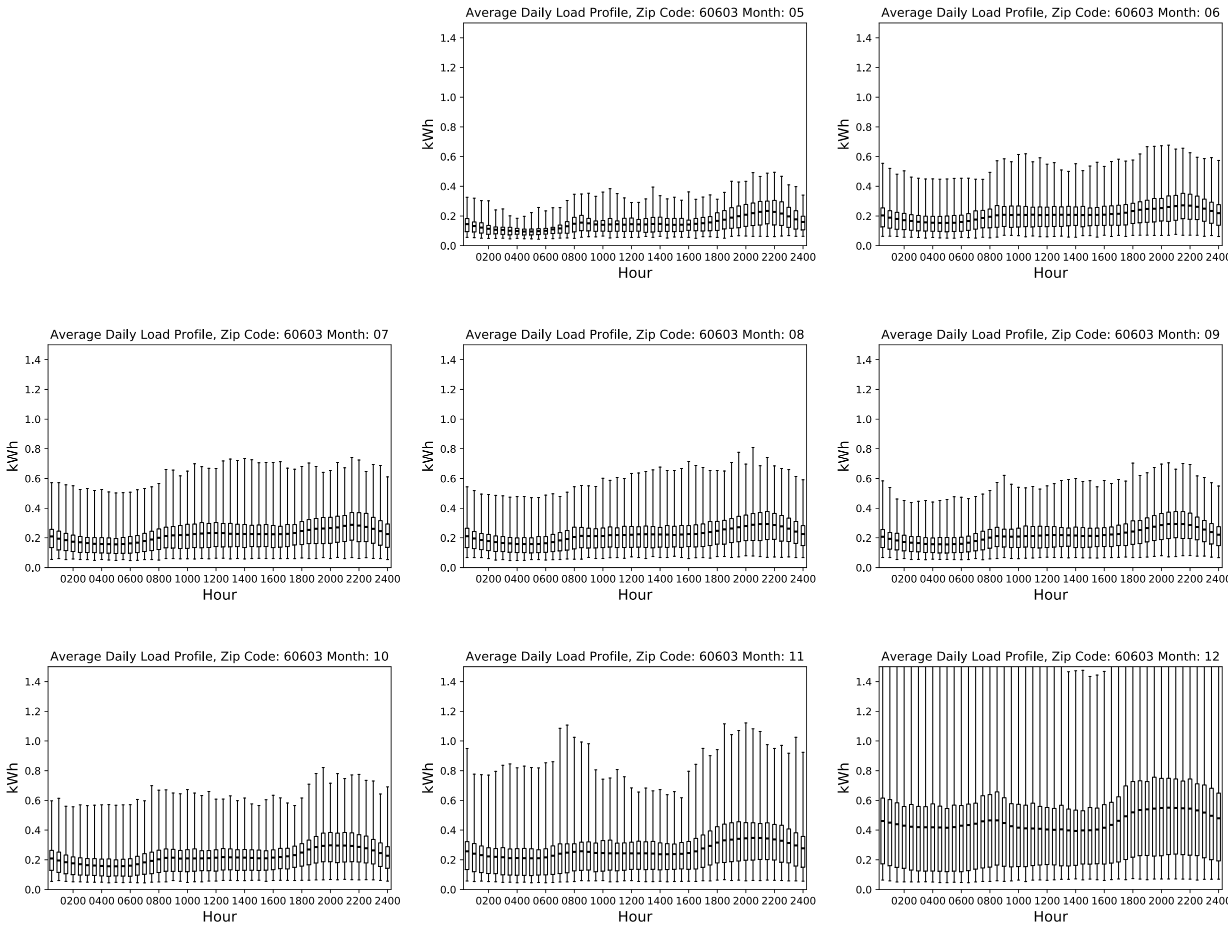

Figure A52. Daily load profiles for multi-family residential electricity consumption by month: 60603.

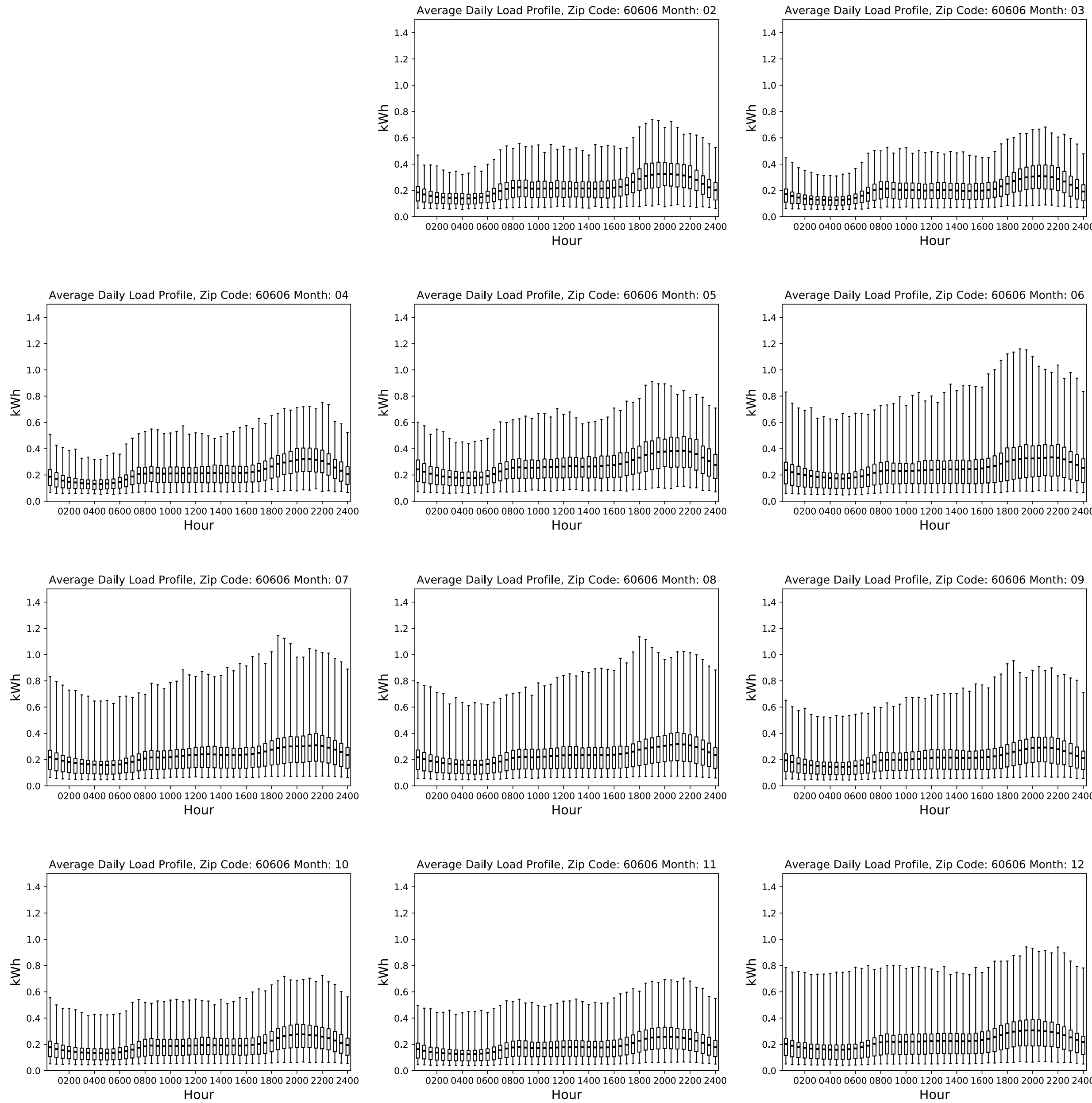

Figure A53. Daily load profiles for multi-family residential electricity consumption by month: 60606.

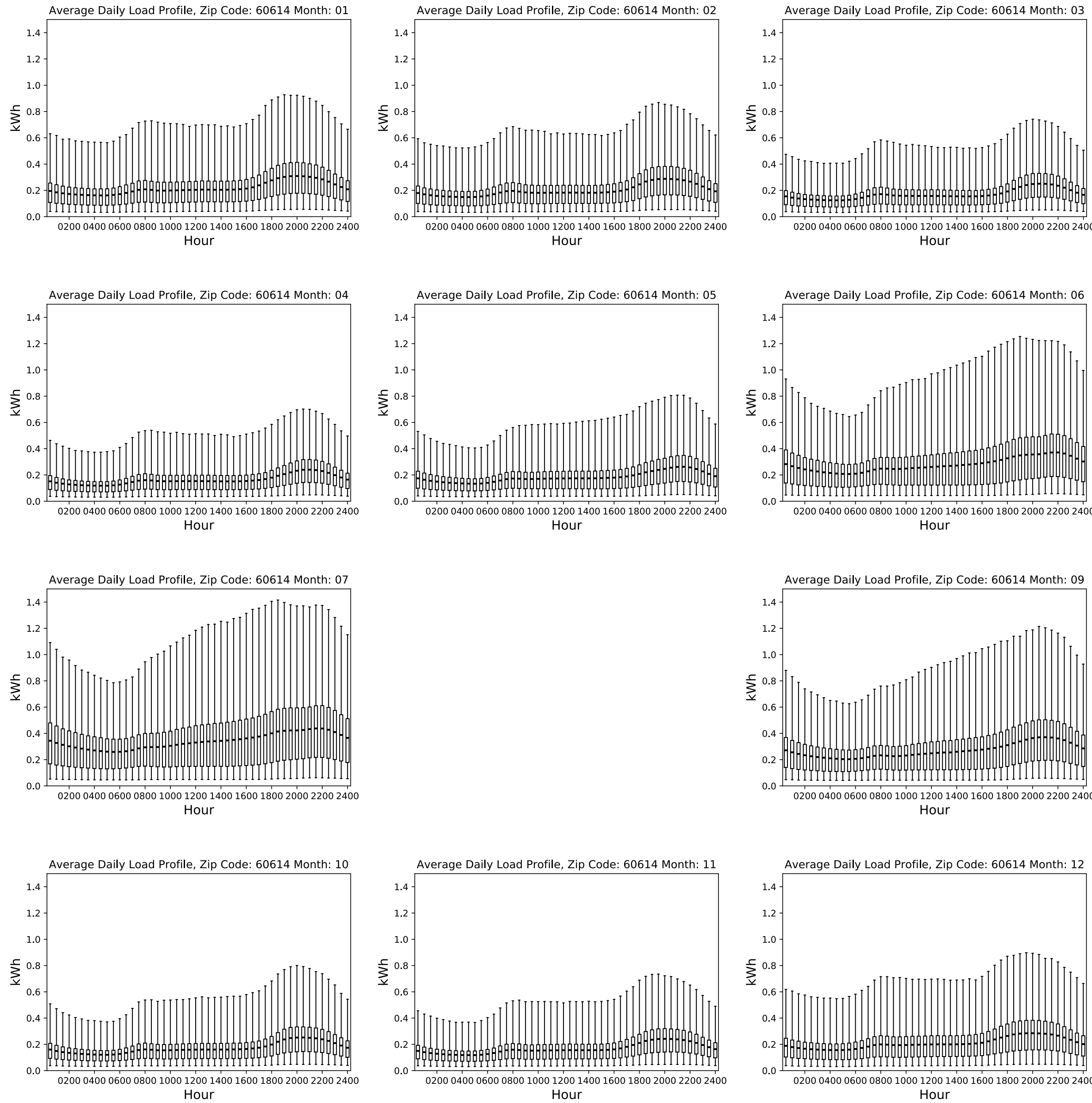

Figure A54. Daily load profiles for multi-family residential electricity consumption by month: 60614.

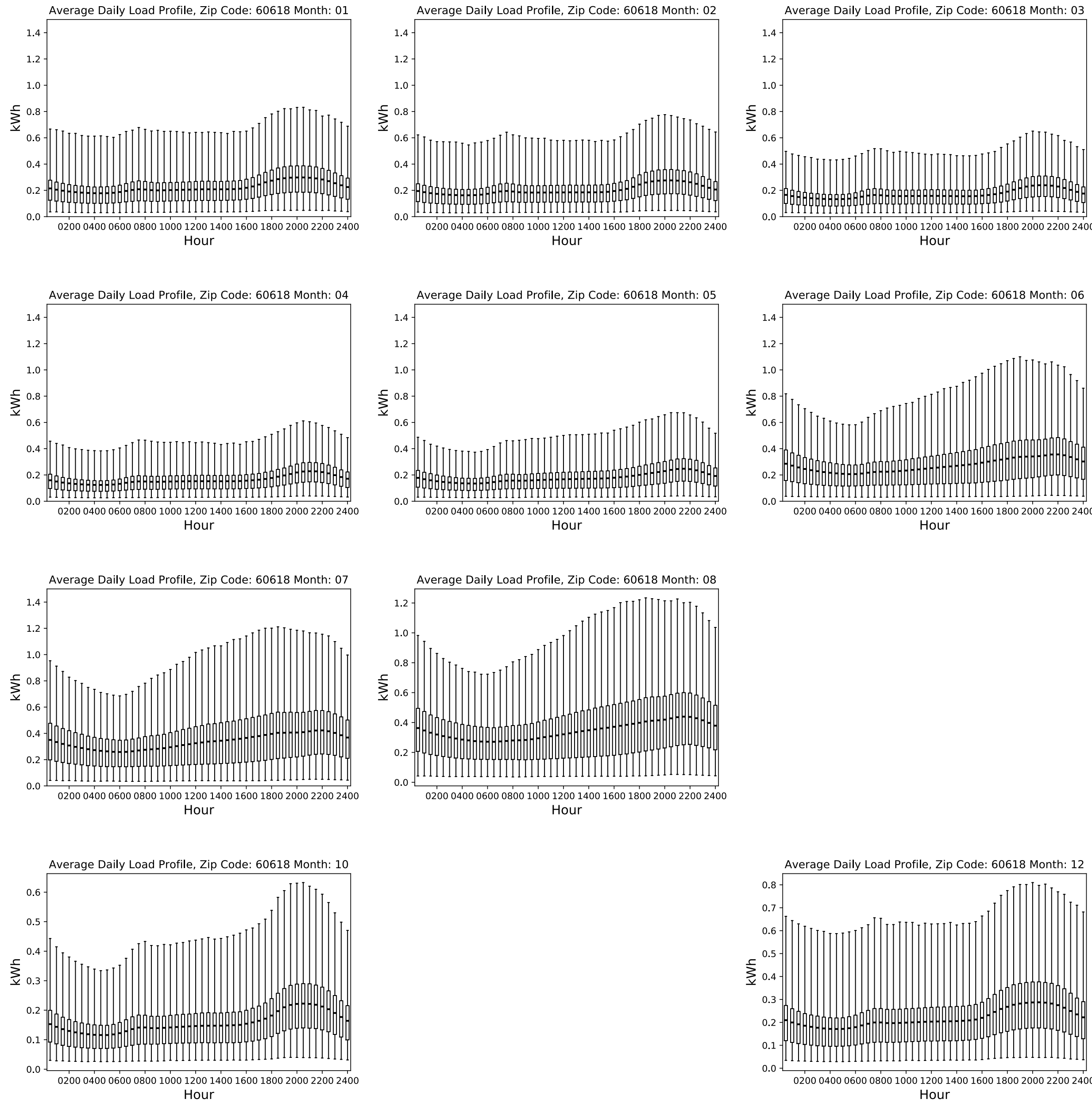

Figure A55. Daily load profiles for multi-family residential electricity consumption by month: 60618.

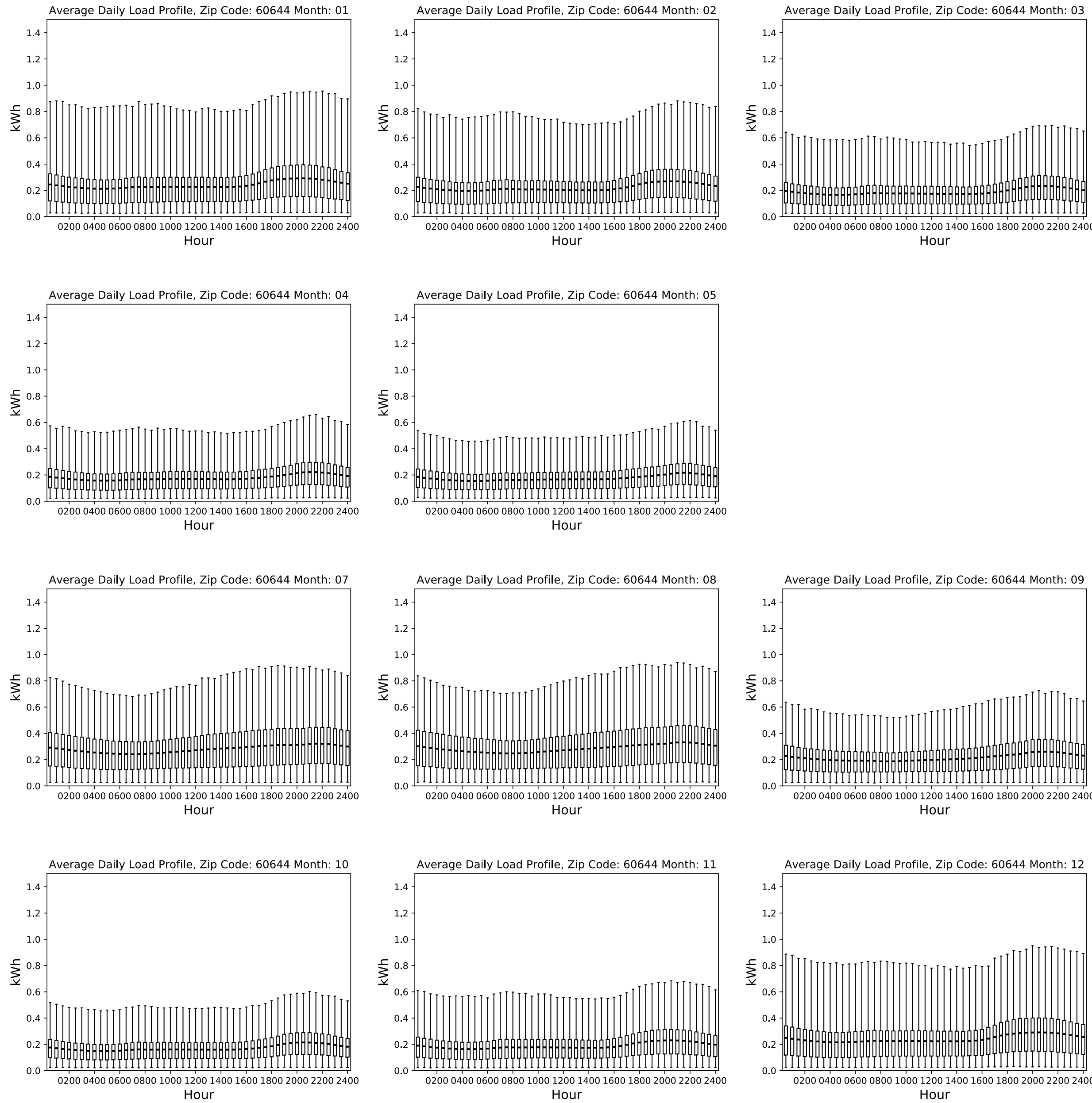

Figure A56. Daily load profiles for multi-family residential electricity consumption by month: 60644.

A2. Electricity Profiles Across Zip Codes

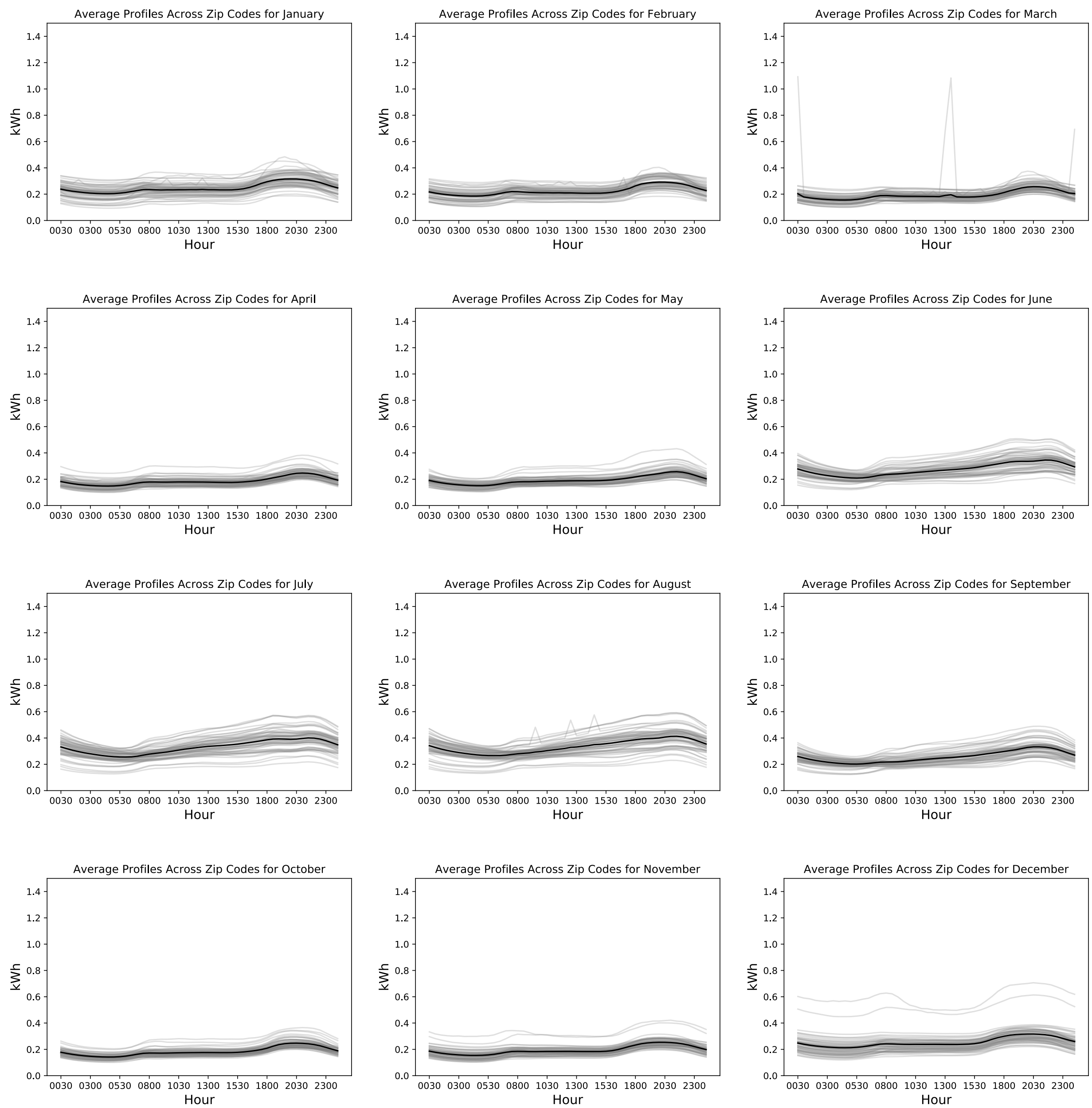

Figure A57. Average daily load profiles for multi-family residential electricity consumption by month, for all analyzed Chicago zip codes.

A3. Socioeconomic Data

The relevant 2016 multi-family residential electricity demand and socioeconomic data for reproducing the multiple linear regression model are available online at https://stillwell.cee.illinois.edu/data/.