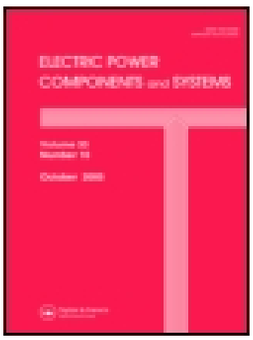

Taylor & Francis
Taylor & Francis Group

Electric Power Components and Systems

ISSN: 1532-5008 (Print) 1532-5016 (Online) Journal homepage: http://www.tandfonline.com/loi/uemp20

Time-Series Modeling of Aggregated Electric Vehicle Charging Station Load

Henry M. Louie

To cite this article: Henry M. Louie (2017): Time-Series Modeling of Aggregated Electric Vehicle Charging Station Load, Electric Power Components and Systems, DOI: 10.1080/15325008.2017.1336583

To link to this article: https://doi.org/10.1080/15325008.2017.1336583

Published online: 21 Dec 2017.

Electric Power Components and Systems, 0(0):1–14, 2017

ISSN: 1532-5008 print / 1532-5016 online
DOI: 10.1080/15325008.2017.1336583

Time-Series Modeling of Aggregated Electric Vehicle Charging Station Load

Henry M. Louie
Department of Electrical and Computer Engineering, Seattle University, Seattle, WA, USA

CONTENTS

Abstract—The widespread proliferation of Electric Vehicles (EVs) can have a transformative effect on the electric power system. The power and energy consumed by EVs when charging is substantial, which has consequences on power system operation and planning. This paper identifies, evaluates, and proposes time-series seasonal autoregressive integrated moving average (ARIMA) models of aggregated EV charging station load. The modeling is based on 2 years of time-stamped aggregate power consumption from over 2400 charging stations in Washington State and San Diego, California. The different data sets allow the influence of time-of-use pricing on the time-series models to be explored. Weekday, weekend, and near-term and long-term models are developed and analyzed. The best performing near-term weekday models are $(2, 0, 0) \times (0, 1, 1)_{24} \times (1, 0, 0)_{120}$ for Washington State and $(2, 0, 0) \times (1, 1, 0)_{24} \times (0, 0, 1)_{48}$ for San Diego. Applications of the seasonal ARIMA models to aggregate EV charging station load forecasting and creation of synthetic time-series at different penetration levels are discussed.

Keywords: ARIMA, charging stations, electric vehicles, forecasting, time-of-use pricing, time series.

Received 20 July 2016; accepted 3 May 2017

Address correspondence to Henry M. Louie, Department of Electrical and Computer Engineering, Seattle University, Seattle, 98122, USA. E-mail: louieh@ seattleu.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uemp.

1. INTRODUCTION

The International Energy Agency estimated that over one million electric vehicles (EV)—highway-capable plug hybrid electric vehicles (PHEVs) and battery electric vehicles (BEVs)—were added to the world's roadways in 2015 [1]. Although this represents a small share of the worldwide vehicle fleet, the growth in EV sales despite depressed petroleum prices is notable. The potential shift in vehicle energy supply from petroleum to electricity at a macro scale has been investigated in several contexts [2–5] in recent years. These studies point to the technical plausibility of EV penetration levels above 70%.

In addition to EV integration studies, contemporary research has investigated control algorithms for EV charging, including the use of EV batteries in demand response [6–8]. A key input to many of this research is a time-series model of individual or aggregate EV charging load. Researchers have had to rely on bottom-up physics-based models [4], surveys of driving behavior [9–11], or use of limited empirical data sets [12, 13]. Other studies use larger data sets, but do not provide

sufficient details to replicate the time series. In previous work by the author, statistical characteristics and probabilistic models of aggregate EV charging load based on measured data were presented. Data from over 2400 charging stations in two regions in the United States were considered, one with time-of-use pricing and one without. The present article builds from this same data set by developing and analyzing time-series models of aggregate EV charging load.

Time-series models of aggregate EV charging station load has several applications. The models themselves provide insight into the characteristics of aggregate EV charging station load. Researchers can use the model to synthesize time series for use in high-temporal resolution integration studies and evaluation of control strategies. Lastly, time-series models can be used in forecasting EV charging station load in the short term and long term.

There are various approaches to developing models of general time series, including autoregressive integrated moving average models (ARIMA), Markov chain-based models, autoregressive conditional heteroscedasticity (ARCH) models, neural network models, and wavelet models [14, 15]. Recent research has applied various machine-learning and big data techniques such as pattern sequence-based forecasting, *k*-nearest neighbor, and support vector regression to EV load forecasting [16–18]. While several of these techniques are promising, the research considers specific EV charging outlets or requires large, diverse data sets of traffic and weather data, which might be challenging to obtain.

In this article, the ARIMA modeling approach is used. ARIMA models are conceptually simple, portable and, in the case of aggregate EV charging load, strike a reasonable balance between parsimony and accuracy. The extant literature is scant of ARIMA-based EV charging load models. One recent paper [19] developed ARIMA models for load forecasting, but instead of being based on actual load data, models of factors such as expected driving distance and charging time were used.

The remainder of this article is arranged as follows. The data sets considered in this work are described in Section 2. Section 3 explores the time-series characteristics of the data. The time-series models are developed and analyzed in Section 4. Applications of the time-series models are demonstrated in Section 5. Conclusions and future outlook of this work are provided in Section 6.

2. DESCRIPTION OF DATA SETS

The EV charging station data sets considered in this work were provided by ECOtality [20]. The data sets correspond to public and private EV charging stations San Diego, California, and portions of Washington State, namely the greater Olympia and Seattle areas, as shown in the Appendix. The data sets cover from 1 January 2011 to 31 January 2013.

The data sets consist of the aggregated real power load of the charging stations in each region. The raw data are 15-min average values, which were subsequently converted into hourly averages. Since aggregated charging station load is considered, spatial diversity and other characteristics such as charger type and station accessibility and location are implicitly accounted for. It is emphasized that the models developed in this article are for aggregate—not individual—charging station load.

The number of charging stations in each region is an important consideration in the time-series models as it affects the stationarity of the time series. In the Washington State data set, there were approximately 270 charging stations in June 2011, rising to 650 in January 2012 and 1200 in January 2013. The numbers for San Diego are similar: approximately 330 in June 2011, 600 in January 2012, and 1250 in January 2013 [21]. It is not possible to accurately relate the number of charging stations with the number of electric vehicles—which surely increased during the period. However, from a practical perspective, utilities are more likely to know the number of charging stations than EVs in their territory, provided special permitting is needed to install a charging station. Almost all the chargers were Blink brand [22] AC Level 2 chargers (240 VAC, 30 A), with the exception being one public DC Fast Charger in San Diego (200–450 VDC, up to 200 A) [21, 23]. Although reactive power was not included in the data set, EV charging in general is done at high power factor, for example the Blink DC Fast Charger power factor is in excess of 0.90 [24]. This provides a basis for converting the real power time-series models developed in this paper to a reactive power time-series model.

An important difference in the data sets is that time-of-use (TOU) electricity rates are used in San Diego. The lowest rates ("super off-peak") occur between 0:00 and 5:00, whereas the highest rates ("on-peak") occur between 12:00 and 18:00 [25]. This scheme provides an economic incentive to charge EVs between 0:00 and 5:00.

3. TIME-SERIES CHARACTERISTICS

The first step in developing the time-series models is to characterize the EV charging station load through qualitative inspection. Figure 1 shows a representative 2-week period for the Washington State and San Diego data sets. Within each data set, there is a pronounced diurnal trend. A salient difference between the two plots is the sharp peaks in the San Diego data set—a result of TOU pricing, which tends to compress

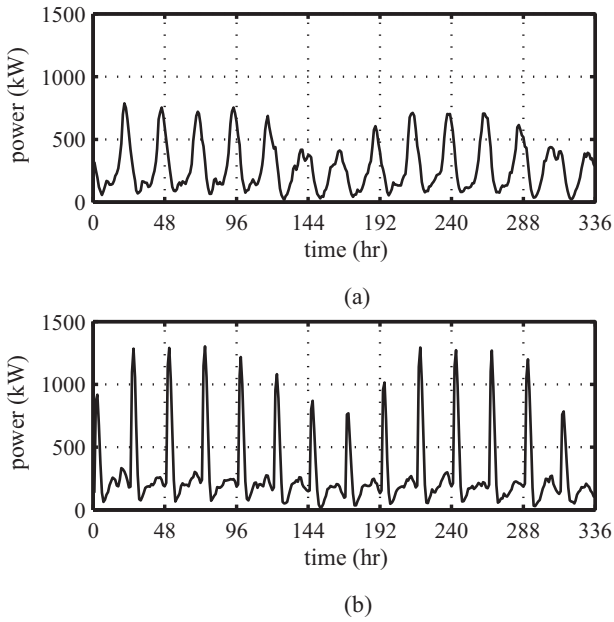

FIGURE 1. Representative 2-week time-series plot of aggregate charging station load. (a) Washington State. (b) San Diego.

the load into the interval between 0:00 and 5:00 (the super off-peak hours). It is also evident that the weekends, shown as hours 120 to 168 and again from 318 to 366 in Figure 1, exhibit a different pattern than the weekdays. Although not shown in Figure 1, the pattern of charging station load on holidays also differs from that on weekdays and is similar to that of weekends.

Figure 2 shows the charging station load for Washington State and San Diego across the entire data set. A 24-hr moving average is shown so that the trend is evident. In each region, the load is low for approximately the first quarter of 2011. Starting around May 2011, the load increases through approximately November 2012. Thereafter, the load stabilizes with a slight decrease at the end of the data set. As the load grows, so does the variability. The increase in mean load and its variance suggests nonstationarity of the data, which must be accounted for in the time-series model.

The autocorrelation (ACF) and partial autocorrelation (PACF) functions of the Washington State data set with holidays and weekends censored are shown in Figure 3. Figure 4 shows the same plots for the San Diego data set. The dashed horizontal lines in these plots are the 95% null confidence intervals for the correlation coefficients. The diurnal pattern is apparent in the ACF, peaking at 24-hr intervals. The weekend plots exhibit similar characteristics but are not shown for the sake of concision.

4. TIME-SERIES MODELING

4.1. Seasonal ARIMA Models

This article seeks to develop ARIMA models of the aggregate EV charging station load. An ARMA model of a stochastic linear time series with response z_t and innovation ϵ can be expressed as [14, 15]:

$$z_t = c + \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \tag{1}$$

where c is a constant term, ϕ and θ are coefficients, and p and q are the orders of the autoregressive and moving-average elements of the model. ARIMA models include differencing. Let L be the lag operator where

$$L^i z_t = z_{t-i} \tag{2}$$

The lag operator polynomials are

$$\phi(L) = 1 - \phi_1 L - \phi_2^2 L^2 - \cdots - \phi_p^p L^p \tag{3}$$

$$\theta(L) = 1 + \theta_1 L + \theta_2^2 L^2 + \cdots + \theta_q^q L^q \tag{4}$$

$$\Phi(L) = 1 - \Phi_{p_1} L^{p_1} - \Phi_{p_2} L^{p_2} - \cdots - \Phi_{p_s} L^{p_s} \tag{5}$$

$$\Theta(L) = 1 + \Theta_{q_1} L^{q_1} + \Theta_{q_2} L^{q_2} + \cdots + \Theta_{q_s} L^{q_s} \tag{6}$$

where Φ and Θ are seasonal coefficients, and p_s and q_s are the orders of the seasonal autoregressive and moving average polynomials, respectively. The general seasonal ARIMA

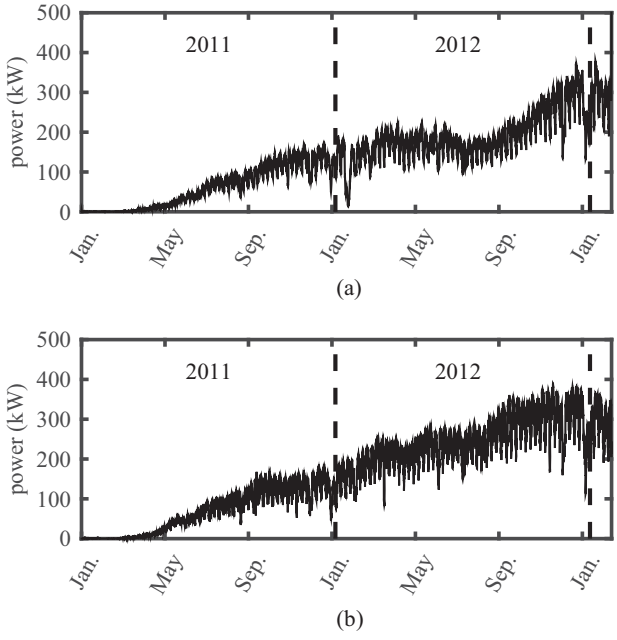

FIGURE 2. Time-series plot of aggregate charging station load spanning the entire data set. (a) Washington State. (b) San Diego.

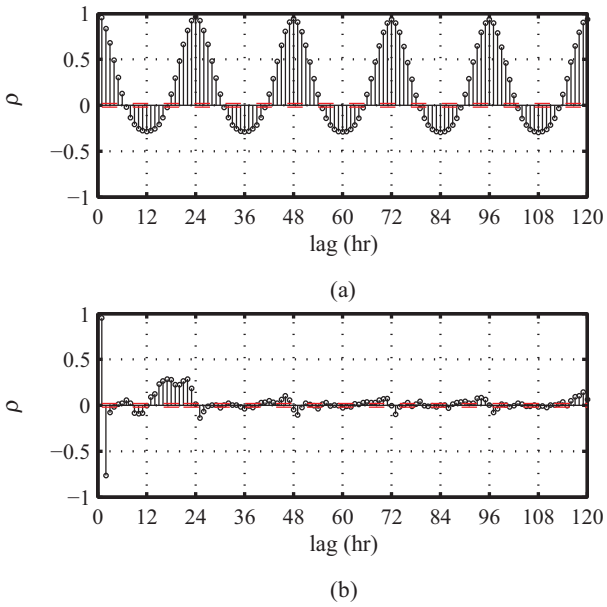

FIGURE 3. Time-series characteristics of Washington State weekday charging station load. (a) Autocorrelation function. (b) Partial autocorrelation function.

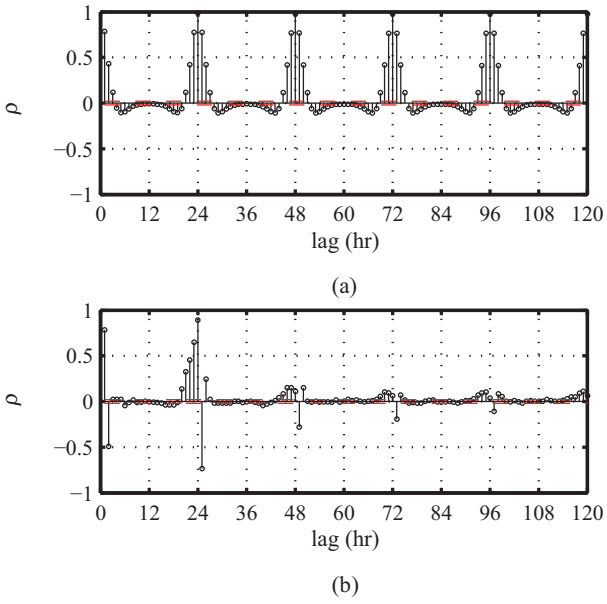

FIGURE 4. Time-series characteristics of San Diego weekday charging station load. (a) Autocorrelation function. (b) Partial autocorrelation function.

model is then

$$\phi(L)(1-L)^{D}\Phi(L)(1-L^{s})^{D_s}z_t = c + \theta(L)\Theta(L)\epsilon_t \qquad (7)$$

where D and D_s are the degrees of integration and seasonal integration, respectively.

4.2. Data Preparation

The qualitative characterization performed in Section 3 guides the time-series modeling approach. Given their differences, separate models are developed for Washington State and San Diego. The data sets for each region are subdivided into weekday and weekend data sets. Because weekends have been censored from the weekday data sets, the time-series of the data for Friday is adjacent to the data for Monday. This introduces discontinuities in the data set, but nonetheless reasonably well-fitting models are found. The few holidays that fall on weekdays are also censored from the weekday time-series.

From each subset of data, the first 140 calendar days have been censored due to the low charging station load during this period. In addition, the last 14 calendar days have also been censored as they will be used later to demonstrate the forecast capability of the models. Therefore, the censored weekday data set contains 418 days of hourly values and the censored weekend data set contains 168 days of hourly data for both Washington State and for San Diego.

In addition to developing time-series models for the censored data set, models are developed for the most recent 60 weekdays and 24 weekend days. It should be expected that these near-term models will perform better in terms of model diagnostics than the models developed for the entire censored data set because the underlying process is less variable. In other words, the model of the censored data covers a span of over 1 year, and must account for, for example, differences in driving behavior between summer and winter and Day Light Savings. Additionally, the number of EV charging stations in each region varies considerably over this time span. This change in the underlying time-series process should temper the expectations of the performance of a long-term model.

The apparent non-stationarity in the data suggests that differencing and/or seasonal differencing be applied. The diurnal characteristics in the data further suggest that the models include seasonal differencing with a lag of 24 hr. It was found that seasonal differencing with a lag of 24 hr alone is sufficient in achieving the dual goals of removing the trend of increasing mean load and eliminating the diurnal pattern. Let z_t be the original time series of aggregate EV charging station load. The resulting time series w_t after seasonal differencing is

$$w_t = L^{24}z_t = z_t - z_{t-24} \qquad (8)$$

where L^{24} is the difference operator with a lag of 24 hr. A representative 2-week weekday time-series of w_t is provided in Figure 5. When compared with Figure 1, the effect of the differencing is notable as the diurnal trend is largely removed.

A useful tool to investigate whether or not the differenced time series is stationary in a weak sense is the variogram [14].

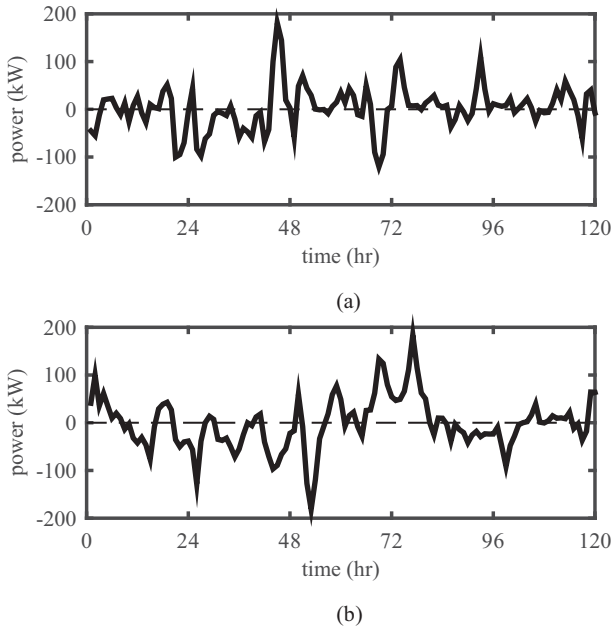

FIGURE 5. Representative 5-day time series of w_t. (a) Washington State. (b) San Diego.

Briefly, a variogram compares the variance of a time-series at various lags to that of the variance when the lag is one time unit, in this case 1 hr. Mathematically

$$G_k = \frac{V\{w_{t+k} - w_t\}}{V\{w_{t+1} - w_t\}}, \quad k = 1, 2, \ldots \tag{9}$$

where G_k is variogram value at lag k and $V\{\cdot\}$ is the variance operator of the set $\{\cdot\}$. For a weakly stationary process, as the lag increases, G_k will approach an asymptote [14]. The variograms for each data set are shown in the top plots of Figures 6–9. For each data set, the variogram of w_t stabilizes, suggesting that w_t is weakly stationary. The model identification is then concerned with w_t of Washington State and San Diego—or to describe it differently, the model of z_t will feature seasonal integration with a lag of 24 (SI(24)).

4.3. Model Identification

The ACF and PACF of w_t are provided in the middle and bottom plots of Figures 6–9 for Washington State and San Diego weekdays and weekends, respectively. The ACF and PACFs exhibit similar qualities among the different regions and data sets. The seasonal differencing attenuated the 24-hr oscillations in the ACF for all cases. An inspection of the ACFs and PACFs in Figures 6 and 9 suggests that several possible seasonal ARIMA models should be entertained. The ACFs of w_t exhibit a damped sine wave beginning at the first lag. The PACFs appear to cut off at a lag of two, suggesting that the models have a second-order autoregressive (AR(2)) component. However, the PACF at the lag of two is small so a model with AR(1) component should also be considered. The ACFs an PACFs also show oscillations and spikes at and near multiples of 24. In the ACF of the weekday time series, the peaks at and around 120 hr are more pronounced than the weekends. This is sensible because the nominal period of the weekday is 5 days (120 hr). Similarly, the weekend ACFs have more notable peaks around 48 hr. It is also notable that the amplitudes of the peaks in the San Diego weekend ACF are generally more pronounced than in the other data sets.

It is difficult to discern a single specific model whose components result in the observed ACF and PACF signatures. However, it is reasonable to entertain models that contain seasonal AR (SAR) and/or seasonal moving average (SMA) components with lags at multiples of 24.

With the inspection of the ACF and PACF not leading to a distinct seasonal ARIMA model, several were considered for each data set. A total of 144 models were considered for each data set containing selected combinations of autoregressive:

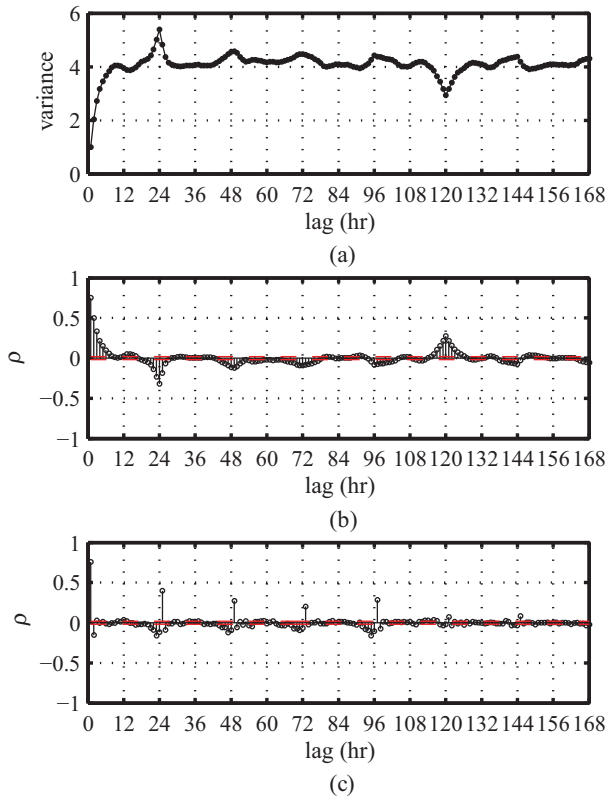

FIGURE 6. Time-series characteristics of Washington State weekday load after 24-hr seasonal differencing. (a) Variogram. (b) Autocorrelation function. (c) Partial autocorrelation function.

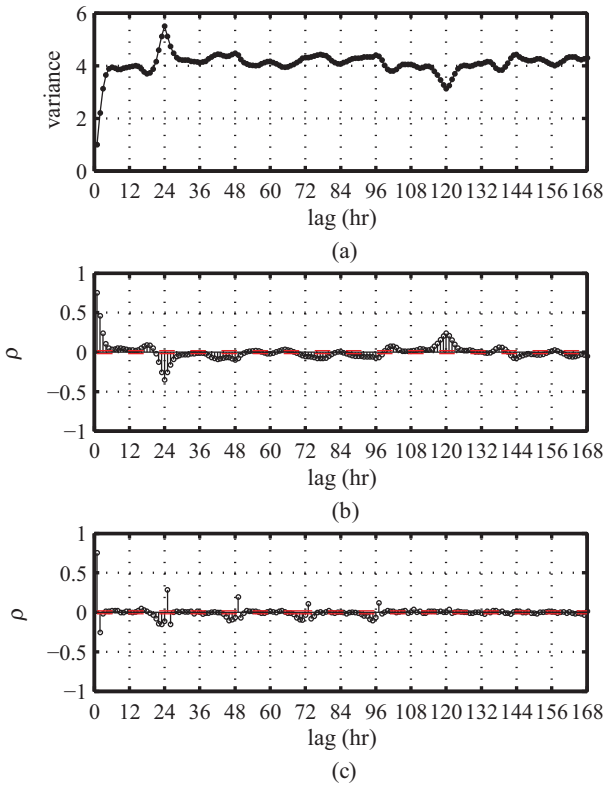

FIGURE 7. Time-series characteristics of San Diego weekday load after 24-hr seasonal differencing. (a) Variogram. (b) Autocorrelation function. (c) Partial autocorrelation function.

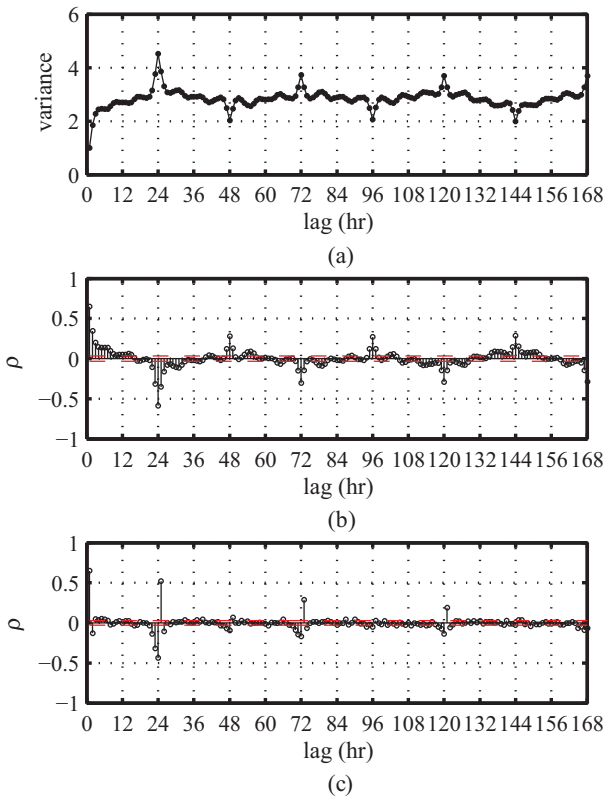

FIGURE 8. Time-series characteristics of Washington State weekend load after 24-hr seasonal differencing. (a) Variogram. (b) Autocorrelation function. (c) Partial autocorrelation function.

AR(0), AR(1), AR(2); moving average: MA(0), MA(1); seasonal autoregressive: SAR(0), SAR(24), SAR(120); seasonal moving average: SMA(0), SMA(24), SMA(48), SMA(120); and seasonal integration SI(24). The parameters of each model were fit using a Kalman filter and the maximum likelihood method in MATLAB (The Mathworks, Natick, Massachusetts, USA) [15].

Considered first are the near-term models for the censored weekday data sets. The long-term models are discussed in the next section. Recall that the near-term models are developed from the most recent 60 days of the censored data set.

The Bayesian Information Criterion (BIC) and mean squared error (MSE) of the top-ranking models were similar within each data set, suggesting that multiple models are reasonably appropriate. The top two models for each weekday data set based on the BIC are provided in Table 1, with estimated coefficients in Table 2. The difference in BIC between the weekday and weekend data sets is largely attributed to the different lengths of days considered.

The models for the weekday data sets differ from one another. The Washington State model is of the form: $(2, 0, 0) \times (0, 1, 1)_{24} \times (1, 0, 0)_{120}$, whereas for San Diego, it is: $(2, 0, 0) \times (1, 1, 0)_{24} \times (0, 0, 1)_{48}$. When the Washington State weekday model is applied to San Diego—after re-estimating the coefficient values—the resulting BIC is 13385 with MSE of 606 kW^2. Similarly, when the San Diego model is applied to the Washington State data, the BIC is 13727 with MSE of 823 kW^2. The models perform reasonably well across both data sets, with limited decreases in their accuracy.

Model	BIC	MSE (kW^2)
WA State (weekday)	13446	808.4
San Diego (weekday)	13150	610.5
WA State (weekend)	5316	727.3
San Diego (weekend)	5264	641.2

TABLE 1. Near-term model estimation summary.

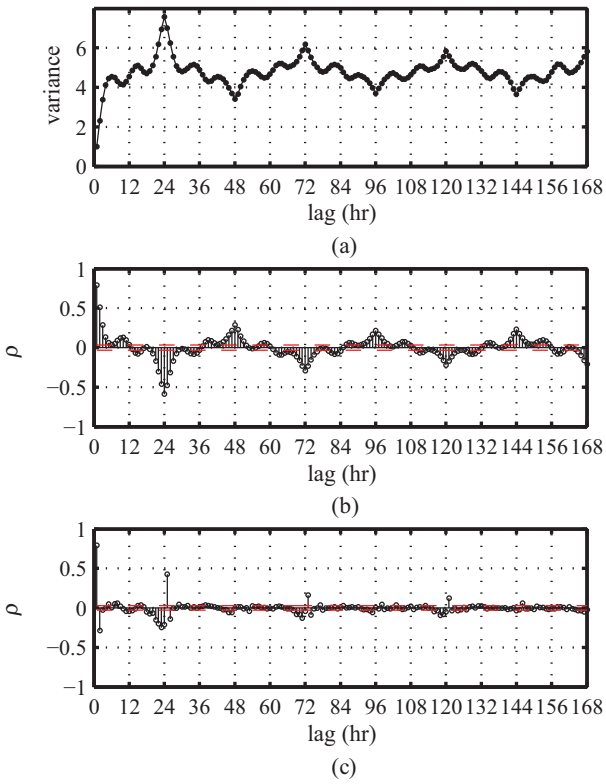

FIGURE 9. Time-series characteristics of San Diego weekend load after 24-hr seasonal differencing. (a) Variogram. (b) Autocorrelation function. (c) Partial autocorrelation function.

	Model term	Coefficient	t	p-value
Weekday WA State	AR 1	0.828	42.68	1.000
	AR 2	− 0.132	− 6.69	0.000
	SAR 120	0.149	7.35	1.000
	SMA 24	− 0.869	− 79.37	0.000
Weekday San Diego	AR 1	1.011	49.74	1.000
	AR 2	− 0.298	− 14.15	0.000
	SAR 24	− 0.800	− 52.62	0.000
	SMA 48	− 0.697	− 38.60	0.000
Weekend WA State	AR 1	0.897	32.92	1.000
	AR 2	− 0.149	− 5.25	0.000
	SAR 24	− 0.976	− 188.67	0.000
	SMA 24	0.135	6.15	1.000
	SMA 48	− 0.733	− 30.81	0.000
Weekend San Diego	AR 1	0.991	32.82	1.000
	AR 2	− 0.257	− 8.18	0.000
	SAR 24	− 0.966	− 128.88	0.000
	SMA 24	0.092	4.11	1.000
	SMA 48	− 0.658	− 29.68	0.000

TABLE 2. Near-term model coefficient estimation summary.

It is notable that the weekend models contain the same elements: $(2, 0, 0) \times (1, 1, 1)_{24} \times (0, 0, 1)_{48}$. The model contains a moving average component at a lag of 48 hr, which as discussed previously, is likely associated with the weekends being 2 days long.

4.4. Model Diagnostics

There are several diagnostic checks that should be performed on time-series models, as outlined in Figure 10. To begin, the coefficients of the models are tested to see if they could reasonably be omitted. The t statistic and associated p-values of the estimated model coefficients are provided in Table 2. The p-values associated with the coefficients are greater than 0.975 or less than 0.025, which suggests that each component of each model is not superfluous at a 95% level of significance.

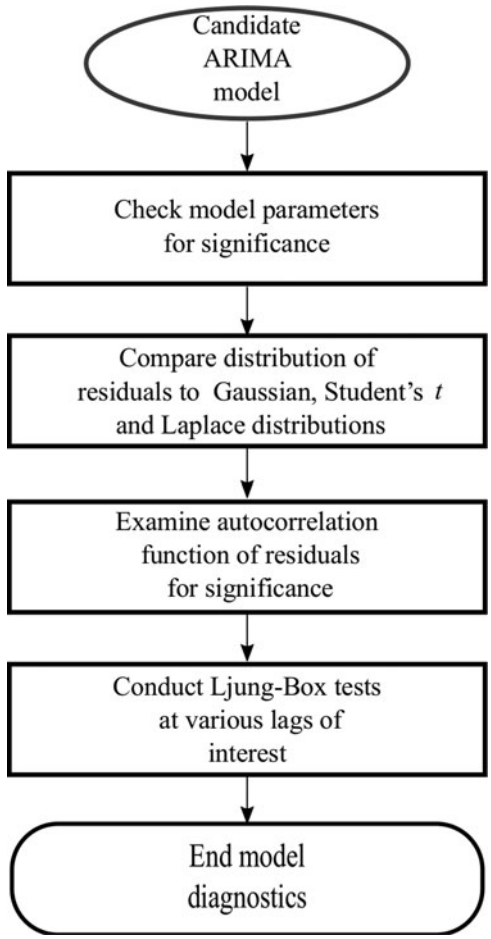

FIGURE 10. Flowchart of model diagnostic evaluation approach.

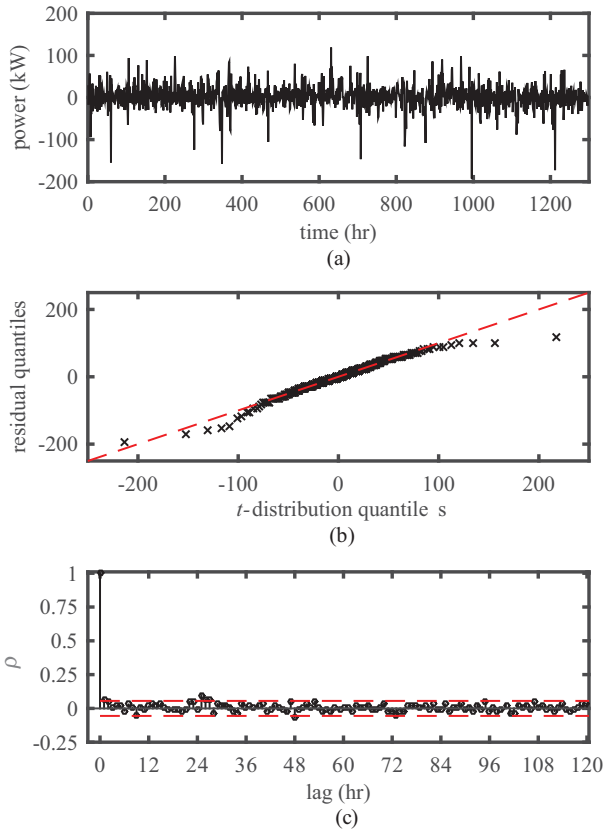

FIGURE 11. Residual analysis of Washington State near-term weekday model. (a) Residual time series. (b) Residual quantile-quantile plot. (c) Residual autocorrelation.

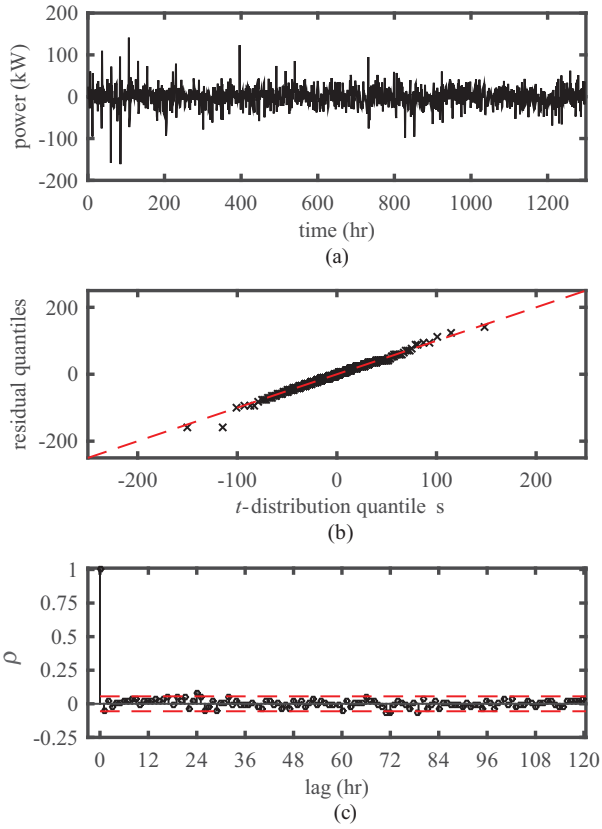

FIGURE 12. Residual analysis of San Diego near-term weekday model. (a) Residual time series. (b) Residual quantile-quantile plot. (c) Residual autocorrelation.

Next, the residuals of the models are tested in three ways: plotting their time-series, inspecting their distribution, and examining their autocorrelation. The residual $\hat{r}_t$ time series of a model is

$$\hat{r}_t = z_t - \hat{z}_t \tag{10}$$

where $\hat{z}_t$ is the model output time series. Note that if the model is exact, $\hat{r}_t$ is equal to ϵ_t, the innovation in the process. The top plots in Figures 11–14 are the time series of the residuals. The residuals of a well-fitting model are expected to have a zero-mean with an unchanging variance. Indeed, the residuals all appear to randomly fluctuate around zero, with no noticeable trend or periodicity.

Not easily recognizable from a time-series plot is the distribution of the residuals. A general assumption in time-series model formulation is the assumption of zero-mean Gaussian-distributed innovations. In an exact model, the residuals should have the same statistical characteristics of the innovations. It is possible, however, for the innovations to follow a distribution other than Gaussian. Therefore, the fit of several distributions including Gaussian, Student's t, and Laplace were tested using a χ^2 goodness-of-fit test. In each case, the residuals were found to fit Student's t distribution at a level of significance (95%) rather than Gaussian. This implies that extreme values are somewhat more likely to be present in the residuals than would be expected in a Gaussian distribution. The fit degrees-of-freedom parameter for each model is provided in Table 3.

Note that the estimated coefficients in Table 2, and the residuals in this section, reflect that the residuals are distributed as Student's t and not Gaussian as usually assumed when a maximum likelihood method is applied. The middle plot in Figures 11–14 is the quantile-quantile plot of the residuals compared to the individually fit Student's t distribution.

Model	Degrees of freedom
WA State (weekday)	3.31
San Diego (weekday)	4.13
WA State (weekend)	2.97
San Diego (weekend)	2.93

TABLE 3. Near-term model residual distribution coefficients.

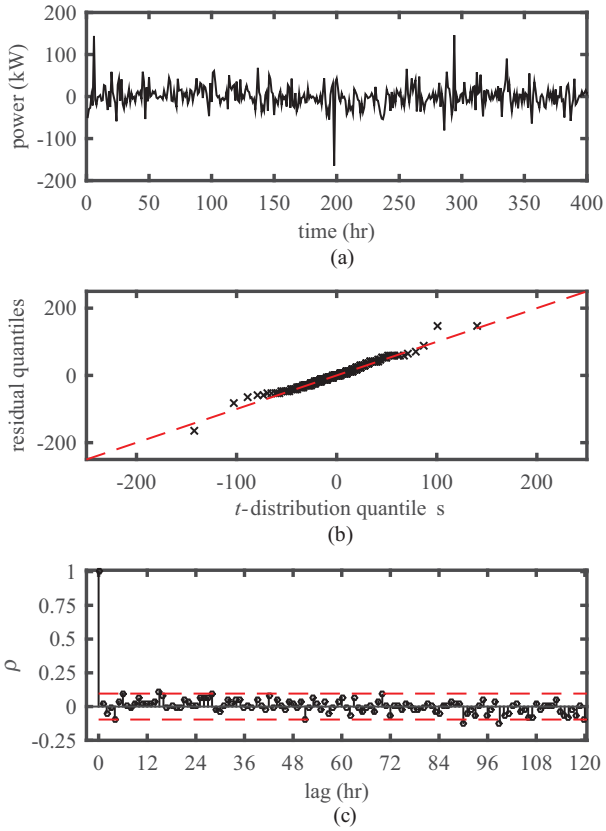

FIGURE 13. Residual analysis of Washington State near-term weekend model. (a) Residual time series. (b) Residual quantile-quantile plot. (c) Residual autocorrelation.

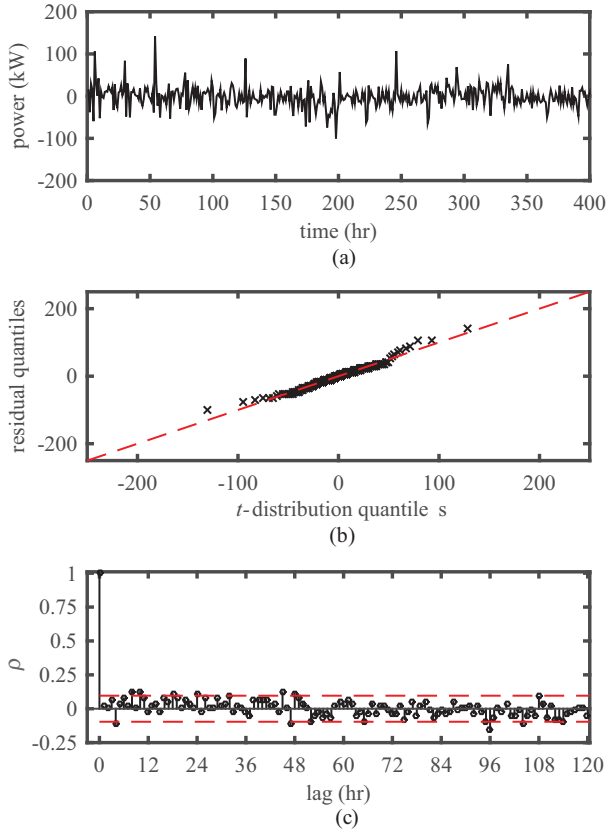

FIGURE 14. Residual analysis of San Diego near-term weekend model. (a) Residual time series. (b) Residual quantile-quantile plot. (c) Residual autocorrelation.

That the residuals, marked as "x" in the figures, generally lie along the diagonal dashed line indicate that the Student's t distribution is well fitting, with a few extreme exceptions.

The autocorrelations of the residuals are shown in the bottom plots of Figures 11–14. It is expected that the residuals are not autocorrelated. Inspection of the plots show that autocorrelation at almost all lags is below the level of statistical significance. To provide more insight, Ljung-Box tests were conducted [14]. This test is based on the notion that the residuals in a time series should not be correlated over time. Lags of 12, 24, 48, 120, and 168 hr were considered with resulting p-value of the test statistic shown in Table 4. Tests whose p-values are greater than 0.05 resulted in the null hypothesis of independence being accepted—the desired outcome for a better performing model. These values are indicated in bold in Table 4. Note that the critical values for the statistics are different for different lags. Three of the four models pass the Ljung-Box tests at nearly all lags. However, the San Diego weekend model did not pass at any lag, suggesting that the model did not capture all the autocorrelation in the time series. Nonetheless, the model performs well in the other diagnostic measures.

In summary, the developed seasonal ARIMA models balance parsimony with accuracy. The best-performing models have similar formats, and indeed the coefficients themselves are reasonably close to each other.

4.5. Long-Term Model

Considered next are the long-term models developed from the entire censored data set. The model parameter evaluation and estimation procedure is identical to that of the near-term

	Lag				
Model	12	24	48	120	168
WA State (weekday)	0.048	**0.164**	0.004	**0.289**	**0.475**
San Diego (weekday)	**0.214**	0.021	**0.056**	**0.263**	**0.252**
WA State (weekend)	**0.109**	**0.211**	**0.400**	**0.128**	**0.403**
San Diego (weekend)	0.000	0.000	0.000	0.000	0.000

TABLE 4. Ljung-box residual test (weekdays).

Model	BIC	MSE (kW^2)
WA State (weekday)	90561	399.3
San Diego (weekday)	93209	516.6
WA State (weekend)	36107	377.3
San Diego (weekend)	38246	634.7

TABLE 5. Long-term model estimation summary.

models discussed previously. The BIC and MSE values are provided in Table 5 with coefficients in Table 6. Note that the BIC and MSE values for the long-term models are not directly comparable to those of the near-term models. The number of samples is different, which affects the BIC. Further, the MSE values are larger in the near-term model in part due to the larger average load values during the period considered.

As with the near-term models, all the coefficients in the long-term models are statistically significant. However, when the Ljung-Box tests were conducted, none of the models passed at the lags considered. The distributions of the residuals were not found to be Student's t. Despite the apparent autocorrelation in the residuals, the long-term models do have the benefit that all are of the same form: $(2, 0, 0) \times (1, 1, 1)_{24} \times (0, 0, 1)_{48}$.

It is worthwhile to briefly discuss the logarithmic transformation of the data prior to seasonal differencing. Models based on this approach were developed and evaluated. However, they performed worse than the previously described models in terms of MSE when converted back to the original units.

	Model term	Coefficient	t	p-value
Weekday Wash. State	AR 1	0.976	151.94	1.000
	AR 2	− 0.172	− 25.21	0.000
	SAR 24	0.221	42.71	1.000
	SMA 24	− 0.865	− 126.11	0.000
	SMA 48	− 0.057	− 8.43	0.000
Weekday San Diego	AR 1	0.935	225.57	1.000
	AR 2	− 0.242	− 39.74	0.000
	SAR 24	0.120	30.39	1.000
	SMA 24	− 0.602	− 207.71	0.000
	SMA 48	− 0.118	− 36.49	0.000
Weekend Wash. State	AR 1	0.901	73.78	1.000
	AR 2	− 0.128	− 10.06	0.000
	SAR 24	− 1.000	− 967.17	0.000
	SMA 24	0.111	13.35	1.000
	SMA 48	− 0.812	− 116.12	0.000
Weekend San Diego	AR 1	0.988	156.44	1.000
	AR 2	− 0.268	− 31.73	0.000
	SAR 24	− 1.000	− 1113.70	0.000
	SMA 24	0.378	85.87	1.000
	SMA 48	− 0.556	− 113.11	0.000

TABLE 6. Long-term model coefficient estimation summary.

5. APPLICATIONS

Two important applications of time-series models are discussed hereafter. In the first, the application and performance of the models for forecasting aggregate EV charging station load across various timescales are demonstrated. The performance is compared against benchmark persistence forecasts. The second application shows how synthetic time-series data can be constructed based on a target number of EV charging stations.

5.1. Forecasting

The developed models are useful in forecasting future values of the time series. As an example of this application, the near-term models for San Diego are applied using 1-, 2-, and 24-hr forecast horizons for weekdays and weekends. The example considers the last 14 calendar days of data, which were censored from the data set for this purpose. The weekday forecast therefore was applied to 10 days and the weekend forecast applied to 4 days.

The forecasted and actual values are shown in Figures 15 and 16 along with 95% prediction intervals for the 1-hr-ahead forecast. The first 2 days of the weekdays and weekend forecasts are shown. The prediction intervals are based on the finding that the innovations in the process can be modeled by an appropriately fit Student's t-distribution, as discussed earlier, rather than the usual Gaussian distribution. The figures show that the time series forecasted by the model closely tracks the actual values and is generally bound within the prediction intervals.

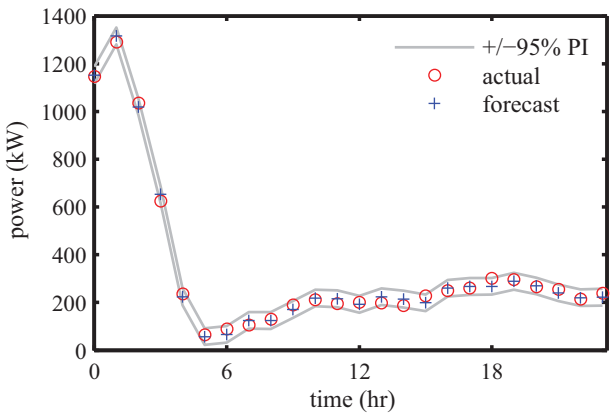

FIGURE 15. Forecast of San Diego weekdays using the near-term model with 1-hr forecast horizon.

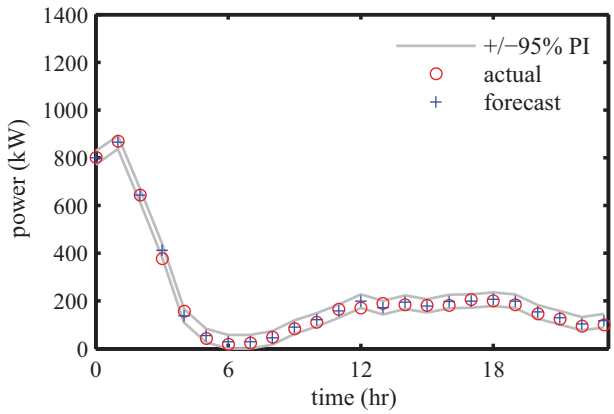

FIGURE 16. Forecast of San Diego weekends using the near-term model with 1-hr forecast horizon.

The MSE and mean absolute percentage error (MAPE) of the forecasts considering all the forecasted days are provided in Table 7. For comparison purposes, the accuracy of two other methods is considered. The first method is a persistence forecasts with 1-hr $(1, 0, 0)$, 2-hr $(1, 0, 0)_2$, and 24-hr $(1, 0, 0)_{24}$ forecast horizons. The second method is the modified pattern sequence-based forecast (MPSF) algorithm.

The MPSF algorithm applied is detailed in [16, 17]. It is based on the pattern of sequence-based forecast algorithm described in [26]. In brief, the MPSF uses a clustering algorithm to optimally group the EV load profile of previous days of aggregate EV load. Each cluster has an associated template load profile. The clustering is based on similarity of the load profiles within each cluster and the dissimilarity between clusters. Each cluster is assigned a unique label, and the sequence of daily load profiles is replaced with the unique label corresponding to its cluster. Therefore, each day is represented by a single label, instead of 24-hr of load data. For example, the weekday near-term model is based on 60 days. The MPSF replaces the 1440 hourly values with 60 labels where each label corresponds to the cluster that a particular day belongs to. Together the labels form a sequence and the algorithm predicts the next label by examining the sequence for similar sub-sequences. For example, using alphabetical labels, assume an arbitrary 5-day period within the 60 days is represented by the sequence "$\ldots, A, C, D, D, F, \ldots$". If the most recent 3 days were "C, D, D", the algorithm would forecast the next day as the template load profile corresponding to cluster F. Further details of the MPSF are described in [16, 17].

Horizon	Model (MSE)	Persistence (MSE)	MPSF (MSE)	Model (MAPE)	Persistence (MAPE)	MPSF (MAPE)
	San Diego (weekday)					
1 hr	511	53317	—	9.3	37.4	—
2 hr	897	143330	—	13.3	95.1	—
24 hr	1296	16469	1686	15.5	16.6	17.9
Horizon	Model (MSE)	Persistence (MSE)	MPSF (MSE)	Model (MAPE)	Persistence (MAPE)	MPSF (MAPE)
	San Diego (weekend)					
1 hr	340	25573	—	9.6	35.6	—
2 hr	752	64128	—	13.9	87.1	—
24 hr	1147	3765	2283	17.3	39.7	22.3

TABLE 7. Forecast accuracy.

The MPSF algorithm is selected as shown in [17] to have greater accuracy than k-nearest neighbor and support vector regression when used to forecast load from individual EV charging outlets. To be consistent with the existing literature on this method, it is only applied to the 24-hr forecast horizon.

The resulting error of the forecasts is provided in Table 7. In all cases, the near-term model outperforms the benchmark persistence and MPSF models. Due to the strong diurnal pattern, the errors associated with the 24-hr forecast horizon are less than the errors of the 2-hr horizon. The models are superior in particular in the one and 2 hr-ahead forecasts, partly due to the sharp rise and fall in load that occurs during the start and end of the super-off-peak pricing period, which the persistence forecasts are unable to anticipate. That the near-term model has greater accuracy than the MPSF is somewhat surprising. In [16], the opposite was shown. However, it must be noted that there are differences between this research and that in [16]. Most significant is that the present research considers aggregate EV charging station load, with the number of stations changing over time, whereas [16] considers a static number of specific EV charging outlets. The number of days of data used in developing the MPSF models was also generally greater in [16], which could improve the performance of the MPSF. Lastly, the ARIMA model in [16] was developed using a different approach, which did not seek to optimize according to AIC or BIC, as was done in this research. With these differences noted, the forecast accuracy of the seasonal ARIMA near-term model developed in this research is at least comparable to existing methods.

5.2. Synthetic Time Series

A useful application of time-series models is the creation of synthetic time series for research and analysis purposes. The seasonal ARIMA models developed can be applied to a wide

range of scenarios. The steps to creating synthetic aggregate EV charging station time series are as follows.

First, the seasonal ARIMA model is selected. For systems with TOU pricing, it is recommended to use the San Diego near-term model, otherwise use the near-term Washington State model as developed in Section 4. Next, because the models have a seasonal component, a prototypical load profile is required. In previous work by the author, average load profile values for aggregate EV charging station load are given for systems with and without TOU pricing. Depending on the desired penetration of EV charging stations, the load profile may need to be adjusted. It has been shown that for systems with more than approximately 600 charging stations, the prototypical load profile can be linearly scaled. The scaled prototypical load profile serves as the pre-response time series upon with the seasonal ARIMA model operates. Note that when the model features seasonal autoregressive terms greater than 24, the prototypical load profile will need to be repeated until the required amount of pre-response data is obtained.

The selected seasonal ARIMA model is applied to the data, and at each time step of the synthesized data, a random variable representing the innovation (residual) is added. The random variable should emulate the statistical characteristics and distribution of the residuals, as discussed in Section 4, but may be scaled to increase or decrease its effect. Limits on the output values may need to be applied, for example, to prevent negative load. The described procedure is applied separately for weekdays and weekends. The prototypical load profiles for weekends are provided in the Appendix of this article; profiles for weekdays are found in previously published work by the author.

As an example, assume that 1 week of synthesized time series of aggregate EV charging station load is desired. The system does not have TOU pricing, and the number of charging stations is 1800. From previously published work by the author, the prototypical load profile of Washington State is based on 1200 charging stations, so that the scaling factor is $\alpha = 1800/1200 = 1.5$. The prototypical mean load profile of Washington State for weekdays and weekends is multiplied by α. The models are separately run for the weekdays and weekends, and then "spliced" together to form one continuous time series, as shown in Figure 17. The synthetic time series approximates the shape of the load profile of Washington State (see Figure 1). However, in the synthetic case, the number of charging stations has increased by 50%, so that the load is commensurately higher. Also note that the change in shape of the time series during the weekend—the last 2 days synthesized—is representative of the decreased load during weekends.

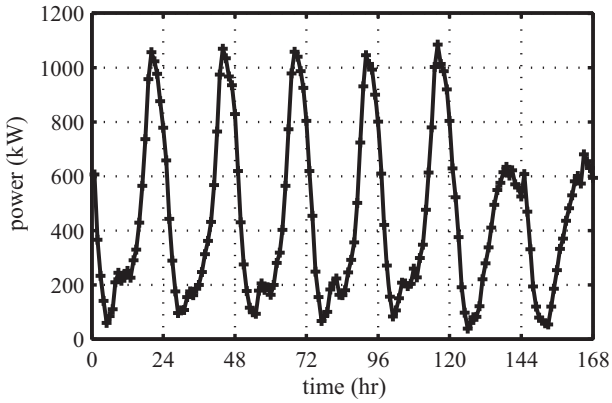

FIGURE 17. Synthetic 7-day time series of a system with 1800 EV charging stations without time of use pricing.

6. CONCLUSION AND FUTURE OUTLOOK

This article developed, analyzed, and applied time-series models of electric vehicle charging station load based on approximately 2 years of data in two different regions. Inspection showed that the time-series data were not stationary, as the load characteristics evolved as the number of charging stations increased. Additionally, the load for weekends was distinct from weekdays. Seasonal ARIMA models were developed for the regions, separated by weekend, weekday, and near- and long-term portions of the data sets. The near-term models, which considered the last 84 days were rigorously analyzed and passed a number of statistical tests. It was shown that their residuals follow Student's t distribution, rather than the Gaussian. The long-term models were decidedly less accurate, reflecting underlying nonstationarity and the general challenges of modeling lengthy data sets.

In general, the models exhibited a balance between parsimony and accuracy, with just four or five parameters required. A challenge, however, with the described approach in the time series synthesis application is the need for a prototypical load profile. Although these data are becoming more available, more are needed to model other applications. The underlying requirement or assumption of stationarity requires that, for example, a utility employing custom seasonal ARIMA model would have to potentially update the model over time as the number of EV stations or driving patterns change. These disadvantages should be weighed against the discussed advantages when deciding whether use the discussed seasonal ARIMA models.

Potential future work includes applying other time-series modeling techniques such as ARCH or Markov-based models to aggregate EV charging station load.

ACKNOWLEDGMENT

The author thanks ECOtality.

REFERENCES

[1] International Energy Agency, "Global EV outlook 2016: Beyond one million electric cars," June 2016, available at: https://www.iea.org/publications/freepublications/publication/Global_EV_Outlook_2016.pdf (Accessed Jul. 2016).

[2] Kintner-Meyer, M., Schneider, K., and Pratt, R. "Impacts assessment of plug-in hybrid vehicles on electric utilities and regional U.S. power grids part 1: Technical analysis," November 2007.

[3] Meliopoulos, S., Meisel, J., Cokkinides, G., and Overbye, T. "Power system level impacts of plug-in hybrid vehicles," October 2009. available at: http://www.pserc.wisc.edu/documents/publications/reports/ (Accessed Oct 2013).

[4] Dyke, K. J., Schofield, N., and Barnes, M. "The impact of transport electrification on electrical networks," *IEEE Trans. Ind. Electron.*, Vol. 57, No. 12, pp. 3917–3926, December 2010.

[5] Ungar, E., and Fell, K. "Plug in, turn on, and load up," *IEEE Power Energy Mag.*, Vol. 8, No. 3, pp. 30–35, May/June 2010.

[6] Shao, S., Pipattanasomporn, M., and Rahman, S. "Grid integration of electric vehicles and demand response with customer choice," *IEEE Trans. Smart Grid*, Vol. 3, No. 1, pp. 543–550, March 2012.

[7] Jin, C., Tang, J., and Ghosh, P. "Optimizing electric vehicle charging with energy storage in the electricity market," *IEEE Trans. Smart Grid*, Vol. 1, No. 4, pp. 311–320, March 2013.

[8] Masoum, A. S., Deilami, S., Moses, P., Masoum, M., and Abu-Siada, A. "Smart load management of plug-in electric vehicles in distribution and residential networks with charging stations for peak shaving and loss minimisation considering voltage regulation," *IET Gener. Transm. Distrib.*, Vol. 5, No. 8, pp. 877–888, 2011.

[9] Darabi, Z., and Ferdowsi, M. "Aggregated impact of plug-in hybrid electric vehicles on electricity demand profile," *IEEE Trans. Sust. Energy*, Vol. 2, No. 4, pp. 501–508, October 2011.

[10] Lojowska, A., Kurowicka, D., Papaefthymiou, G., and van der Sluis, L. "Stochastic modeling of power demand due to EVs using copula," *IEEE Trans. Power. Syst.*, Vol. 27, No. 4, pp. 1960–1968, November 2012.

[11] Liu, Z., and Wu, Q. "EV charging analysis based on the national travel surveys of the nordic area," *Proceedings of the IEEE PES General Meeting*, National Harbor, MD, July 2014.

[12] Gong, Q., Midlam-Mohler, S., Marano, V., and Rizzoni, G. "Study of PEV charging on residential distribution transformer life," *IEEE Trans. Smart Grid*, Vol. 3, No. 1, pp. 404–412, March 2012.

[13] Alizadeh, M., Scaglione, A., Davies, J., and Kurani, K. S. "A scalable stochastic model for the electricity demand of electric and plug-in hybrid vehicles," *IEEE Trans. Smart Grid*, Vol. 5, No. 2, pp. 848–860, March 2014.

[14] Bisgaard, S., and Kulahci, M. *Time Series Analysis and Forecasting by Example*. Hoboken, NJ: John Wiley & Sons, 2011.

[15] Chatfield, C. *The Anaysis of Time Series: An Introduction*, 6th ed. Boca Raton, FL: Chapman & Hall/CRC, 2004.

[16] Majidpour, M., Qiu, C., Chu, P., Gadh, R., and Pota, H. "Modified pattern sequence-based forecasting for electric vehicle charging stations," *Proceedings of the IEEE International Conference on Smart Grid Communications*, Venice, Italy, November 2014.

[17] Majidpour, M., Qiu, C., Chu, P., Pota, H., and Gadh, R. "Forecasting the EV charging load based on customer profile or station measurement?," *Appl. Energy*, Vol. 163, pp. 134–141, February 2016.

[18] Arias, M., and Bae, S. "Electric vehicle charging demand forecasting model based on big data technologies," *Appl. Energy*, Vol. 183, pp. 327–339, December 2016.

[19] Amini, M., Kargarian, A., and Karabasoglu, O. "ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation," *Electric Power Systems Research*, Vol. 140, pp. 378–390, November 2016.

[20] The EV Project, "Overview," 2013, available at: http://www.theevproject.com/overview.php (Accessed May 2013).

[21] The EV Project, "EV project documents," 2013, available at: http://www.theevproject.com/documents.php (Accessed May 2013).

[22] Blink Network, "Blink," 2016, available at: https://www.blinknetwork.com/ (Accessed April 2017).

[23] Electric Transportation Engineering Company, "Electric vehicle charging infrastructure deployment guidelines for the central Puget Sound area," May 2010, available at: http://www.theevproject.com/documents.php (Accessed May 2013).

[24] Blink Network, "Blink DC fast charger," 2016, available at: http://prod.blinknetwork.com/brochures.html (Accessed April 2017).

[25] San Diego Gas and Electric, "EV rates," 2013, available at: http://www.sdge.com/clean-energy/ev-rates (Accessed March 2013).

[26] Martínez-Álvarez, F., Troncoso, A., Riquelme, J., and Aguilar-Ruiz, J. "LBF: A labeled-based forecasting algorithm and its application to electricity price time series," *Proceedings of the IEEE International Conference on Data Mining*, Pisa, Italy, December 2008.

APPENDIX

The prototypical (mean) load profile for Washington State and San Diego Quarter 4 are provided in Table A.1.

Time period	WA state (kW)	San Diego (kW)
0:00–1:00	421.24	826.40
1:00–2:00	326.95	884.29
2:00–3:00	217.84	652.77
3:00–4:00	129.39	399.00
4:00–5:00	74.07	165.50
5:00–6:00	46.33	60.95
6:00–7:00	37.33	52.80
7:00–8:00	47.66	64.53
8:00–9:00	65.75	87.23
9:00–10:00	93.27	118.01
10:00–11:00	124.97	151.49
11:00–12:00	168.84	182.77
12:00–13:00	220.01	201.08
13:00–14:00	268.82	213.89
14:00–15:00	306.99	209.19
15:00–16:00	343.26	208.58
16:00–17:00	378.88	216.78
17:00–18:00	398.90	230.14
18:00–19:00	393.17	232.29
19:00–20:00	365.40	219.30
20:00–21:00	350.38	196.44
21:00–22:00	350.96	166.81
22:00–23:00	366.25	147.08
23:00–24:00	362.97	170.69

TABLE A.1. Washington state and San Diego weekend prototypical load profiles.

The geographic boundaries of the San Diego and Washington State areas in which the charging stations are installed is provided in Figure A.1.

(a)

(b)

FIGURE A.1. Geographic boundaries of the charging stations as of August 2012.

BIOGRAPHIE

Henry M. Louie is an associate professor in the Department of Electrical and Computer Engineering at Seattle University. He received his B.S.E.E. from Kettering University, his MS from the University of Illinois, and his PhD from the University of Washington. His research interests include renewable energy resource modeling and humanitarian engineering.