Statistical approach and multiplicative models for electric vehicles charging behaviour patterns

D Y Koeva

1 Department of Electrical Engineering and Electronics, Technical University of Gabrovo, 4 H Dimitar, Gabrovo 5300, Bulgaria

E-mail: dkoeva@abv.bg

Abstract. Since the charging processes of electric vehicles are stochastic and time-dependent, the paper views an approach based on a statistical analysis of real data on electricity consumption at charging station connection points. Other types of data (geographical, public sites, distance between individual charging stations, etc.) are also taken into account when making the analysis. Multiplicative models are the most suitable for studying and forecasting time series with pronounced cyclicity and seasonality. Their application allows us to consider the correlation of the load in the consuming nodes with regional features, climatic factors and seasonality. The method and approach discussed in this paper make possible the processing of a large amount of data and the detection of load cyclicity in the load schedule of electricity facilities. The results of the model will identify the requested charging power in a developing charging infrastructure.

1. Introduction

The ever-increasing number of electric vehicles and the consequent major changes in transport sector that we are witnessing today pose enormous challenges to the electricity distribution network and the charging infrastructure connected to it. A careful stakeholder analysis is needed for the proper expansion of the charging infrastructure: from the point of view of electricity suppliers (appropriate place to provide the necessary charging power without peak overloads of electricity consumption at the point of connection of the charging station); from the point of view of owners of electric vehicles (utilization of the structure for maximum accessibility and services); from the point of view of government and local government structures. Nowadays the remarkable evolution in the field of mobility is related to digitalization, autonomous driving, shared mobility and dynamically evolving electric mobility.

The reduction of CO2 emissions, atmospheric pollution and noise and the controlled dependence on oil are at the focus of this sustainable development. But how do we evaluate the participation of electric vehicles (EVs) in the implementation of each of these important goals? It seems that they are quite ambitious and the current 5.3 million (according to the International Energy Agency) electric vehicles will not cover them. Therefore, there is an urgent need to improve charging technology and build relevant charging station infrastructure. The charging process of electric vehicles is stochastic and time-dependent and the most common practice is to connect charging stations to an existing electricity supply network. The energy system operators are responsible for managing energy flows and constantly monitoring the balancing energy at the connection points of large consumers with peak loads, such as charging stations (CS). There are two main reasons why they do this: the load in the local distribution networks and the required transmission network are strongly increasing with all the negative
consequences of this. The energy market regulation is a constant corrective. In Bulgaria, the development of the charging station infrastructure (CSI) turns out to be a simple function of the number of registered electric vehicles, without any research or analysis done in advance. There is no profile that prevails at the moment – business customers, as well as private and municipal installations are approximately equal. It is important to expand the network, first at public places and then at company parking lots. Business desire depends on the number of installed CSs and they are the most difficult thing in the field of electric mobility since more preparation is required. The secondary market for electric vehicles is developing slowly, mostly because of the undeveloped and incomplete CSI. There are several web sites providing information about charging stations in Bulgaria, as well as worldwide. However, there is not a real platform for charging infrastructure management with various options for planning, booking, payment and management, as well as for attracting other participants in this process. In conclusion, the main steps towards low-carbon economy and more sustainable transport with the increasing use of EVs are careful forecast analysis of appropriate multiplicative models for electricity consumption and subsequent integration of renewable energy sources in the charging infrastructure.

2. Background

The structure of the energy distribution network to which charging stations are connected is interdependent and therefore a careful preventive analysis is recommended in order to avoid peak loads. Some factors that determine the behavior during charging are time-dependent, others are geographical and local, but overall the charging process is defined as stochastic. The results of an optimized multiplicative model for electricity consumption at the connection point of the charging station will be an indication of the requested power and location of the expanding charging infrastructure. This is particularly true for countries like Bulgaria, where the rate of use of EVs is described as low, „exploratory“, and charging infrastructure as slowly developing and chaotically positioned.

The uncontrolled charging combined with the strongly increasing trend of EVs cause major problems in the distribution networks at some places. The solution is „smart charging management algorithm strategy“ based on multiplicative forecasting models. In [1, 2] it is shown how the power transformers in PS overload when CS are connected. In [1, 2, 3], stochastic models based on the behavior of EVs’ owners during charging are offered. Selected models are used to evaluate the impact of CS on the distribution network, as well as a tool to compare different types of EVs and the consumption of residential sites and other objects nearby. Another stochastic model is used in [4, 5] and it is based on the consumption of households supplied by the urban distribution network. Since 2016 the analyses of potential risks of charging according to geographical region, travel time between work and home, average mileage, charging time, etc. have started. This is how the so-called „charging profile of a typical EV battery“ is built. Nowadays, according to [5], it is evident that studying travel patterns and charging behavior is beneficial.

The first approach involves the collection and processing of data for the purpose of forecast analysis of the required charging power.

The second approach is based on building a model of the entire EVs lot and using the data as an indicator of the required power provided by the charging infrastructure at certain points and times, [6]. Obviously, the forecast analysis of the energy consumed by the charging stations is extremely important when it comes to evaluating the load on the distribution networks and determining the exact location and power when the CSI is expanding.

3. General statement and approach

The most common models for identifying discrete objects and time-dependent processes are:

3.1. Autoregressive model – AR (Auto Regressive)
It is considered as the simplest and described as follows:
\[ A(z).y(t) = e(t) \]  
(1)

\[ A(z) = 1 + a_1.z^{-1} + a_2.z^{-2} + \cdots + a_n.z^{-n_a} \]  
(2)

### 3.2. ARX model – (Auto Regressive with eXternal input)

In generalized and expanded form is given with the expressions:

\[ A(z).y(t) = B(z).u(t) + e(t) \]  
(3)

\[ y(t) + a_1.y(t - 1) + \cdots + a_{n_u}y(t - n) = b_1.u(t) + b_2.u(t - 1) + \cdots + b_{n_b}.u(t - m) + e(t) \]  
(4)

\[ B(z) = b_1 + b_2.z^{-1} + \cdots + b_{n_b}.z^{-n_{b+1}} \]  
(5)

Optimization with ARX models is most effective when applying a polynomial approximation. The forecasts made by ARX models have a high degree of accuracy, especially if the approximate polynomials are of a higher order. Distortions (disturbances) as part of the dynamic system are a disadvantage. These models are widely used because of their main advantage: it is not necessary to know the physical characteristics of the system, which greatly reduces engineering work.

### 3.3. ARMAX model – (Auto Regressive – Moving Average with eXternal input)

In this case, the object of study (electricity consumption) is time dependent. The models used for forecasting are non-deterministic as it is assumed that they allow us to determine the probability of future values in a certain range. Such models are called stochastic and they play an important role. They are divided into stationary and non-stationary. Stationary models are based on the assumption that the process remains in equilibrium around a relatively constant average level. But often in industry, trade, economy and so on, there are time series that are best described as non-stationary, which means that they do not have a natural average level. In this case an exponentially weighted moving average value is used. The stochastic model, where forecasting by exponentially weighted moving average is optimal, belongs to the class of non-stationary processes, also known as autoregressive processes - with integrated moving average or ARIMA (Auto Regressive Integrated Moving Average) models.

\[ A(z).y(t) = B(z).u(t - n_k) + C(z).e(t) \]  
(6)

where \( n_k \) is a value that characterizes the delay.

\[ C(z) = 1 + c_1.z^{-1} + c_2.z^{-2} + \cdots + c_{n_c}.z^{-n_c} \]  
(7)

To achieve greater flexibility in the construction and operation of the model in terms of the observed time series, sometimes it is more appropriate for the models to be combined in one autoregression and moving average model. The combined model is described by the expression:

\[ \tilde{z}_t = \phi_1.\tilde{z}_{t-1} + \cdots + \phi_p.\tilde{z}_{t-p} + a_t - \theta_1.a_{t-1} - \cdots - \theta_q.a_{t-q} \]  
(8)

where \( p+q+2 \) are unknown parameters: \( \mu, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, a_{t}^2 \), which are determined by the observations.

It is important to know that the combined models have the following distinguishing features: stationarity, reversibility and reciprocity. Mixed autoregression – the moving average model is always preferred because of its economy. In general, the process is expressed by the equation:

\[ \tilde{z}_t = \phi_1.\tilde{z}_{t-1} + \cdots + \phi_p.\tilde{z}_{t-p} + a_t - \theta_1.a_{t-1} - \cdots - \theta_q.a_{t-q} \]  
(9)
\( (1 - \phi_1B - \phi_2B^2 - \cdots - \phi_pB^p)\tilde{z}_t = (1 - \theta_1B - \theta_2B^2 - \cdots - \theta_qB^q)\alpha_t \) \hspace{1cm} (10)

\( \phi(B)\tilde{z}_t = \theta(B)\alpha_t \) \hspace{1cm} (11)

where: \( \phi(B) \) and \( \theta(B) \) are polynomials of \( B \) of degree \( p \) and \( q \).

Many time series in industry, trade, economy are non-stationary and do not fluctuate around a relatively fixed average value. Along with that, the process fluctuates around a different average level and, taking into account the differences in levels, a similar behavior of the order is observed. Such an order can be represented by a generalized autoregression operator \( \varphi(B) \) where one or more zeros of the polynomial \( \varphi(B) \) are equal to 1. When similar features of the time series are repeated after the \( s \) interval (they are called reference intervals), it is said that the time series has periodicity \( s \). The most common mistake in the analysis of time series is mixing the procedures for building the forecasting model. Therefore, a common approach to make the analysis is to divide the order into three components: trend, seasonal component, and random component. The multiplicative models are suitable for researching and forecasting time series with pronounced seasonal periodicity. If the time series is \( \{Z_t, t = 1, 2, 3, \ldots, k\} \), a SARIMA model (Seasonal ARIMA model) is generated for it from row \((p, d, q) \times (P, D, Q)\) with average value \( \mu \), as follows \([6, 7, 8, 9]\):

\[\phi_p\Phi_p(B^S)\cdot(1 - B)^d\cdot(1 - B^S)^D\cdot(Z_t - \mu) = \theta_q(B)\Theta_q(B^S)\cdot\alpha_t\] \hspace{1cm} (12)

where: \( \phi(B) \) is a polynomial of order \( p \), \( p \) is the order of non-seasonal autoregression; \( \theta(B) \) - polynomial of order \( q \), \( q \) - the order of the non-seasonal component of the moving average; \( \Phi(B^S)\) - polynomial of order \( P \), \( P \) - order of seasonal autoregression; \( \theta(B^S) \) - polynomial of order \( Q \), \( Q \) - the order of the seasonal component of the moving average; \( d \) - the number of differences; \( D \) - the number of seasonal differences; \( S \) - the length of the season.

The development and analysis of the model is performed in the following sequence:

**Descriptive Methods** – the data are visualized and the frequency of seasonal occurrence is established, if it is not precisely determined in advance;

**Seasonal Decomposition** – the data are divided into four components: trend \((T_t)\), cyclical component \((C_t)\), seasonal component \((S_t)\), and random component \((R_t)\). The data are presented as a product of the individual components \(Y_t = T_t\cdot C_t\cdot S_t\cdot R_t\).

**User-Specified Model** – the most appropriate models are selected;

**Automatic Forecasting** – automated model selection is performed with the best approximation to the source data.

Estimated Autocorrelations, Partial Autocorrelations and Tests for Randomness of residuals – the autocorrelation function, the partial autocorrelation function and a randomness test of the difference between data and forecast are calculated, i.e. the residual error. The type and values of the functions are used to diagnose the model.

**4. Results and discussion**

The object of monitoring and analysis is the consumption of three publicly available charging stations (CS 1, CS 2 and CS 3), connected to the city distribution network. Considering the problems described in sections 1 and 2, we use the data collected for the energy profiles of the three power substations (PS 1, PS 2, PS 3) supplying these terminals, Figure 1. The data include active and reactive electricity measured in 15 minutes for three typical months: January, May and July, (2976 values for each month), 2020. The energy profile of the terminals to which the CSs are connected is monitored at the same time, Figure 2, Figure 3, Figure 4. The energy profiles represent active and reactive energy for one typical working day of each month, a total of 100 measured values for 24 hours (x-axis). The analysis of the application results and the optimized multiplicative model will be an indication of the correlation of the load with climatic factors, seasonality, electricity tariff zones and EVs charging behavior.
**Figure 1.** Power substations PS and CSs location: PS1 and CS1 (a), PS2 and CS2 (b), PS3 and CS3 (c).

**Figure 2.** CS 1 energy profile for January (a), May (b) and July (c), 2020.

**Figure 3.** CS 2 energy profile for January (a), May (b) and July (c), 2020.

**Figure 4.** CS 3 energy profile for January, May and July 2020.
Table 1. Initial data processing.

|       | Unique Values | Minimum | Maximum |
|-------|---------------|---------|---------|
| CS1   | 57            | 12      | 146.25  |
| CS2   | 116           | 10.2    | 98.8    |
| CS3   | 136           | 0       | 79.8    |

Table 2. ARIMA Model Summary

| Parameter | Estimate  | Standard Error | t        | P-value |
|-----------|-----------|----------------|----------|---------|
| AR (1)    | 0.885454  | 0.0192577      | 45.9793  | 0       |
| MA (1)    | -0.18089  | 0.0397865      | -4.54652 | 5E-6    |
| SMA (1)   | 0.646419  | 0.0289428      | 22.3343  | 0       |

In a three-coordinate system, the data from January (x-axis), May (y-axis) and July (z-axis) are compared, the so-called data distribution density, Figure 5 a). Each monthly energy profile is checked for random data (Normal Probability plot with 95% limits). Box and Whisker plot helps to find datacenter, median and interquartile range, as it is done for May, figure 6. Because of the processing of 288 data on the consumption of CS for each month considered, non-repeating, minimum and maximum values for each of them are established. Table 1. The density of distribution is presented on figure 5 b). A completely adequate description of the seasonal time series using a multiplicative model is impossible to obtain. The adequacy of the model gives an idea about how accurate (as close as possible to the specific data) and reliable it is (in which cases it gives good results).

![Figure 5](image)

**Figure 5.** The density of distribution.

![Figure 6](image)

**Figure 6.** Data randomness check.
The adequacy analysis is performed after calculating the residual error once the model has been started. An autocorrelation checks of the residual errors using the method of least squares is made. In addition, a check of a periodogram that shows the probabilistic distribution of the residual error is made. However, both checks are considered relatively rough and do not provide an accurate picture of the adequacy of the model. Selected from a total of seven models, the ARIMA \((1,0,1)\times(0,1,1)_7\) model for forecasting the daily profile of electricity consumption gives very good results, Table 2. According to the analysis of the coefficients of the autocorrelation and the private autocorrelation functions of the residue, only between two private autocorrelation functions out of 24 been calculated a correlation is found. The forecast can be more precise if the relationship between the residual of the data after the forecast analysis and the predicted values is used, Figure 7:

\[
R = e^{[0.282929+0.980795\ln(F)]}
\]  

where: \(R\) – Residual; \(F\) – Forecast.

![Figure 7. Finding a relation between predicted values and residual data.](image)

5. Conclusion
The disadvantage of the model is the difficulty in predicting values with a low frequency of occurrence, but in all cases, the quality of the data used for the model is of great importance. It is necessary to have a long period of observations and thus to form a database of values of the observed parameters and their change. The electricity consumed in January differs from the trend and cyclicity of those in May and July. The reason is compliance with all restrictions and lockdown due to COVID-19. These measures affect households and industry, and therefore the whole economy. Therefore, another statistical approach is required to study their impact.

6. References
[1] Brady J, O’Mahony M 2016 Modelling charging profiles of electric vehicles based on real-world electric vehicle charging data Sustainable Cities and Society vol 26 p 203-216 doi: 10.1-16/j.scs.2016.06.014
[2] Huang Y 2019 Day-Ahead Optimal Control of PEV Battery Storage Devices Taking Into Account the Voltage Regulation of the Residential Power Grid IEEE Transactions on Power Systems vol 3, no 6, doi: 10.1109/TPWRS.2019.2917009
[3] Muratori M, Morgan M J, Serra E, Rizzoni G 2013 Highly resolved modeling of personal transportation energy consumption in the United States *Energy* vol 58 pp 168-177

[4] McKenna E, Thomson M 2016 High resolution stochastic integrated thermal-electrical domestic demand model *Applied Energy* vol 165 pp 445-61

[5] Arghira N, Hawarah L, Ploix S, Jacomino M 2012 Prediction of appliances energy use in smart homes *Energy* vol 48, no 1 pp 128-134

[6] Pevec D, Babic J, Kayser M, Carvalho A, Ghiassi-Farrokhfal Y, Podobnic V 2018 A data-driven statistical approach for extending electric vehicle charging infrastructure *Energy Research* vol. 42 pp 3102-20 https://doi.org/10.1002/er.3978

[7] Dong Y, Wang J, Jiang H, Wu J 2011 Short-term electricity price forecast based on the improved hybrid model, *Energy Conversion and Management* vol 52 pp 2987-2995 doi: 10.1016/j.enconman.2011.04.020

[8] Wang Sh, Feng J, Liu G 2011 Application of seasonal time series model in the precipitation forecast, *Mathematical and Computer Modelling* vol 58, pp 677-683 doi: 10.1016/j.mcm.2011.10.034

[9] Box G E P, Jenkins G M, Reinsel G, Ljung G 2015 *Time Series Analysis: Forecasting and Control* 5th Edition, Holden-Day San Francisco ISBN: 978-1-118-67502-1

**Acknowledgements**

The implementation of this paper is thanks to the support from the University Project 2104E “Electric drives for electric vehicles and industrial applications, electrical components and systems – practical and mathematical model studies in terms of energy and economic efficiency”, financed by the Ministry of Education and Science of the Republic of Bulgaria.