Revised manuscript (clean version)

Decoupling Based Monthly Net Electricity Consumption Prediction Model Considering High Penetration of Distributed Solar PV Systems

Xin Chen[a], Fei Xu[b], Guixiong He[c], Zhenghui Li[a], Fei Wang[a, d, e,*], Kangping Li[f], João P.S. Catalão[g]

[a] *Department of Electrical Engineering, North China Electric Power University*, Baoding 071003, China
[b] *State Key Lab of Power Systems, Department of Electrical Engineering, Tsinghua University*, Beijing 100084, China
[c] *China Electric Power Research Institute, Beijing*, China 100192, China
[d] *State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University)*, Beijing 102206, China
[e] *Hebei Key Laboratory of Distributed Energy Storage and Microgrid (North China Electric Power University)*, Baoding 071003, China
[f] *College of Smart Energy, Shanghai Jiao Tong University, Shanghai 200240, China*
[g] *Faculty of Engineering, University of Porto and INESC TEC*, Porto 4200-465, Portugal

Abstract

The large-scale introduction of distributed photovoltaic (DPV) increases the need for retailers to consider and quantify the differences in monthly electricity consumption of customers to maximize their interests in trading in the forward electricity market. For customers with DPV, retailers need to predict net electricity consumption (NEC), which is actual electricity consumption (AEC) minus DPV generation. However, the DPV is behind the meter and DPV generation data is invisible to retailers. Therefore, the issue of how to distinguish the transition of customers from no DPV to with DPV and their DPV installation information needs to be addressed. To better capture the additions of DPV timely under high penetration of DPV, a decoupling-based monthly NEC prediction model considering the DPV installation update is proposed. Firstly, the features are extracted from the hourly NEC data of known customers with DPV to distinguish other customers whether installing DPV. Secondly, an online update framework of DPV installation evaluated by two validations is proposed. Thirdly, based on the difference in the electricity consumption series before and after the installation of DPV, the NEC is decoupled into AEC and DPV generation. Finally, the monthly DPV generation prediction results are subtracted from the monthly AEC prediction results to obtain the final monthly NEC results. Different scenarios of DPV penetration are set in case studies to test the performance between the proposed model and other direct models. The results indicate the superiority of the proposed method under high penetration of DPV.

Keywords: Retailers; Forward Electricity Market; Distributed Photovoltaic Penetration; Monthly Net Electricity Consumption Prediction; Online Update Framework

Nomenclature

Abbreviations

DPV	Distributed Photovoltaic
AEC	Actual electricity consumption
NEC	Net electricity consumption
SVR	Support vector regression
ARIMA	Auto-regressive integrated moving average

Symbols

T	Set of timeslot
Ω	Set of customers
t	Indices of timeslot
g, i, j	Indices of the hour, month and year
k	Indices of prediction error
$\boldsymbol{x}$, $\boldsymbol{y}$	Input/output vector of SVR
u	Binary variable indicating the identification of customers
w, b	Parameters of SVR
p, d, q	Partial correlation order, differential order, and auto-correlation order of ARIMA
B, C	Value of the Bayesian/Akaike information validation
H_g, M_i, Y_j	The g th hour, i th month and j th year
$s_{cl,g}$	The slope of the cut line at H_g
$r_{tp,g}$	Degree of concavity at H_g
$r_{s,g}$	The average rate of change of the slope
ε_k	Prediction error of the prediction of P_{Load,H_g,M_i,Y_j} at k th time
ε_{av}	Average prediction error
$P_{Net,t}$, $P^0_{Net,t}$	NEC of each customer whose DPV information is unknown/known
$P_{Load,t}$, $P^0_{Load,t}$	AEC of each customer whose DPV information is unknown/known
$P_{DPV,t}$, $P^0_{DPV,t}$	DPV generation of each customer whose DPV information is unknown/ known
$\hat{y}_{Load,H_g,M_i,Y_j}$	The real value of AEC of P_{Load,H_g,M_i,Y_j}
y_{Load,H_g,M_i,Y_j}	Prediction result of AEC of P_{Load,H_g,M_i,Y_j}
y_{DPV,H_g,M_i,Y_j}	Prediction result of DPV generation of P_{Load,H_g,M_i,Y_j}
$P_{Net,\,fluctuation}$	Fluctuation of NEC
$P_{Net,trend}$	Trend of NEC
n_{cor}	Correct identification number of non-solar customers converted to solar customers
n_{tol}	The real number of all non-solar customers to be identified that are converted to solar customers
a_{update}	Accuracy of the DPV installation update check validations
y_{ac_t}, y_{pre_t}	Real/predicted value of NEC

1. Introduction

In the past decades, the traditional monopoly utility form of supplying electricity to customers has revealed a series of problems such as negativity and inefficiency, and the deregulation of the electricity industry in various countries has introduced market competition, which has activated the electricity market and improved economic efficiency [1-2]. Under the retail competition model, retailers participate in the wholesale market on behalf of small customers, and then sell electricity to small customers in the retail market, profiting from the price difference. Retailers need to predict customers' electricity consumption in the coming month before purchasing electricity in advance in the forward electricity market [3]. The error of prediction affects the final profit, and the imbalance of electricity is the source of the trading risk. Therefore, an accurate monthly electricity consumption prediction is beneficial in helping retailers plan the amount of electricity to be purchased in advance, which concerns their interests.

The monthly electricity consumption prediction is confronting new obstacles in the current environment, where the proportion of distributed energy generation is increasing. The large-scale installation on the demand side of the power system, represented by distributed photovoltaic (DPV), allows customers to generate their electricity. The characteristics of customers' electricity consumption curves during the daytime are altered by DPV generation. The electricity consumption that needs to be bought by retailers is considered net electricity consumption (NEC), which refers to the AEC minus the distributed energy generation. This paper focuses on DPV, and the NEC in this paper refers to the AEC minus the DPV generation. The DPV penetration is the ratio between DPV generation and the total power generation in a given period [4].

Therefore, the stochastic nature of DPV generation overlaid with the uncertainty of customer load brings new characteristics to monthly NEC. This brings new questions for retailers to consider when predicting monthly NEC for customers: 1) Is the original method of predicting monthly electricity consumption considering DPV still applicable? 2) The installation of DPV by customers is a random event, so DPV for households in a certain area is constantly changing over time. As DPV penetration increases, how can retailers get first-hand information about DPV installations of customers to guide monthly NEC prediction? To address the above issues, it is necessary to figure out novel methods of monthly NEC considering the increasing DPV penetration [5].

The impact of DPV penetration on monthly NEC characteristics varies with different levels. The study of monthly NEC has significant implications for grid dispatch, demand-side management, and forward electricity market trading. The NEC prediction ideas can be classified into two categories: the direct idea and the indirect idea. The direct idea, whose prediction target is the NEC, involves using the DPV generation data monitored by the additional metering equipment installed, but this will significantly increase the cost due to the large number of DPV. The indirect idea is decomposing the NEC into AEC and DPV generation by mining the relevant information included in various data sources with adequate temporal and spatial granularity [6-9], with the prediction targets being the AEC and DPV generation produced after NEC decomposition. Deep learning algorithms [10] and probabilistic statistics [11] are popular among scholars studying net load prediction. In [12], the proposed ANN-based short-term residential net load forecasting method is validated in 75 single-family composition micro-neighborhood cases in the Netherlands.

Considering the uncertainty of human behavior and distributed energy generation, probabilistic algorithms have received attention recently [13]. In [14], a new hybrid probabilistic scheme approach is proposed to model the uncertainties associated with electrical and thermal load demand and renewable distributed output. Reference [15] provides a detailed description of the application of probabilistic prediction methods in solar irradiance prediction and load prediction. In [16], it finds that increasing the proportion of DPV output power in the net load of residential customers can improve the clarity and reliability of probability prediction in spring and winter. The photovoltaic generation prediction method proposed in [17] aiming for invisible photovoltaic sites is different from [18], which belongs to unsupervised learning. Therefore, the model training does not require historical data of the sites, and the proposed method applies to high penetration.

In addition, with the support of advanced metering facilities, it is easy to collect a full range of electricity data [19]. Reference [20] designs the experiments and verifies that the data obtained from smart meters have advantages in mining residential electricity consumption features and that the extracted features interpret residential electricity consumption features better than those using the standard load profile methods. For the huge amount of data collected from smart meters, reference [21] proposes a clustering algorithm that reduces dimensionality while preserving load characteristics.

Despite the developments in NEC prediction, there are still issues that need to be addressed and improved. First, they fail to account for non-solar customers' DPV installation information changes, resulting in a portion of DPV generation not being included in the NEC calculation. This portion is commonly referred to as the AEC prediction error, as it can cause variances in the extraction of customer load characteristics, affecting the NEC prediction accuracy. Second, it is insufficient to research the NEC characteristics based only on the NEC curve under variable weather conditions. As an example, the possible scenarios behind a decrease in the net load curve during daytime hours compared to neighboring days are 1) slightly better weather conditions and a minor increase in DPV generation, 2) a combination of lower actual customer electricity consumption and smaller DPV generation, and 3) no DPV installation (NEC equals AEC), where the AEC decreases. The first two cases may also be mistaken for no newly installed DPV due to the similarity to the AEC curve.

The main contributions of this paper are as follows:

(1) A novel monthly NEC prediction method that utilizes decoupling technology is proposed to sort out the factors affecting NEC and the degree of influence (For example, weather factors have an impact on both AEC and DPV generation, but the impact mechanism is different) so that the model can be more targeted.

(2) An online update framework of customer DPV installation information is proposed to address the incomplete, non-reciprocal, and severely delayed DPV

installation statistics held by grid dispatch operators or retailers and improve the performance of the monthly NEC prediction model further.

(3) The validations are proposed to evaluate the performance of the online update framework. In particular, based on the differences in fluctuation between AEC and DPV generation, it is judged whether it is the error of the AEC prediction or the installation of DPV.

The rest of the paper is shown below: The problem statement and overall framework will be introduced in Section II. Then, the specific process of the proposed method will be presented in Section III. Next, the data set, case study, and performance evaluation will be proposed in Section IV. Finally, the conclusion and future work will be written in Section V.

2. Problem statement and overall framework

2.1 Problem statement

Since the NEC is equal to the AEC minus DPV generation and the DPV penetration is variable, it is necessary to study the effect of DPV generation on the characteristics of the AEC. The typical daily curve of NEC is called the duck curve, which is famous for its "belly" appearance during the daytime [22]. How monthly NEC characteristics vary with DPV penetration needs to be known. In addition to the numerical difference, there is a difference in the slope change of the monthly AEC and monthly NEC curves for an individual residential customer, as shown in Fig. 1. The numerical and shape difference between monthly AEC and monthly NEC is expanded as the DPV penetration increases for a cluster of 300 customers, as shown in Fig. 2. How the NEC curves differ from AEC depending on the rate and trend of the change in DPV penetration for a cluster of 300 customers in Fig. 3. When the DPV penetration increases to a particular level, the NEC curve differs significantly from the AEC curve.

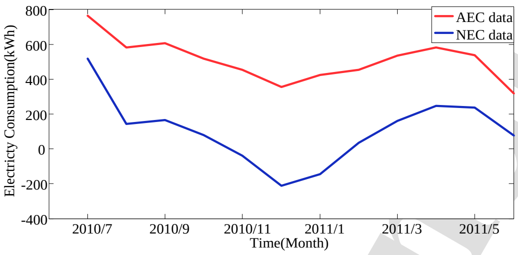

Fig. 1.The comparison of monthly AEC and monthly NEC for a customer

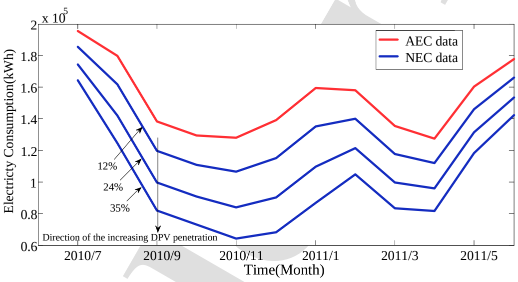

Fig. 2.The comparison of monthly AEC and monthly NEC for 300 customers under different fixed DPV penetration

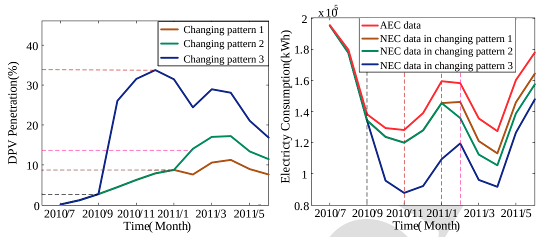

Fig. 3.The comparison of monthly AEC and monthly NEC for 300 customers under different changing patterns of DPV penetration

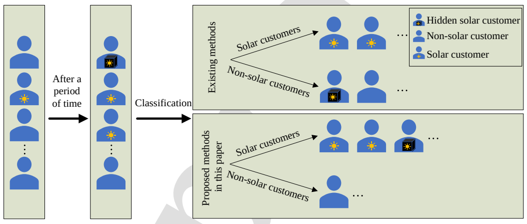

Fig. 4.The process of customer classification

Driven by incentive or subsidy policies, residents are enthusiastic about installing DPV. Therefore, there is a possibility of untimely reporting of customer DPV installation information, incomplete statistics, or non-authorized private installation. It is not conducive to the monthly NEC prediction results. After some time, new solar customers may add, requiring the retailers to update the solar customer information. It is necessary to identify the characteristics that reflect the transformation of customer status (including non-solar customers and solar customers), as shown in Fig. 4. Let P_{Load}, P_{DPV} and P_{Net} respectively denote AEC, DPV generation, and NEC of each customer. After dividing all customers into solar customers and non-solar customers, monthly AEC is predicted for non-solar customers and monthly NEC is predicted for solar customers. The relationship among AEC, DPV generation and NEC is shown as follows:

$$\sum_{l\in\Omega} P_{Net_l} = \sum_{l\in\Omega}\left(P_{Load_l} - u_l P_{DPV_l}\right) \tag{1}$$

where u_l indicates whether the l th customer has installed DPV and 0-1 variable. Ω is the set of customers.

2.2 Benefits for Decoupling of Net Electricity Consumption

The benefits of decoupling the NEC will be discussed in two aspects. From the perspective of qualitative analysis, it can be seen in (1) that monthly AEC plays a positive growth role on NEC, while DPV generation plays a decreasing role on NEC. Under the joint effect, AEC plays a positive growth role on NEC, while DPV generation plays a decreasing role on NEC. When NEC eventually exhibits a descend trend due to the combined influence of AEC and DPV generation, the causes could be 1) AEC decreases, 2) DPV generation grows, and 3) AEC decreases more than DPV generation. For machine learning, if the NEC is directly used as input, the selected parameters may not reflect the inherent characteristics of the NEC change pattern. The ambiguity in learning the patterns behind the changes in NEC could easily further lead to large prediction errors when the changes in DPV penetration are more complex.

From the perspective of influencing factors, since changes in either AEC or DPV generation directly affect NEC, the factors that affect AEC and DPV generation also affect NEC. Reference [23] found that the problem of predicting DPV power generation and AEC can be converted into predicting solar irradiance and ambient temperature through the derivation of mathematical equations. The mechanisms of ambient temperature on DPV power generation and actual power consumption are different. The high temperature of DPV panels is not favorable for the operation of power electronics [24]. Besides, the summer and winter are the peak periods of actual residential demand for electricity [25]. Therefore, modeling AEC and DPV generation separately after decoupling the NEC can make the mapping relationship between input and output variables clearer and the model more relevant.

2.3 Overall framework

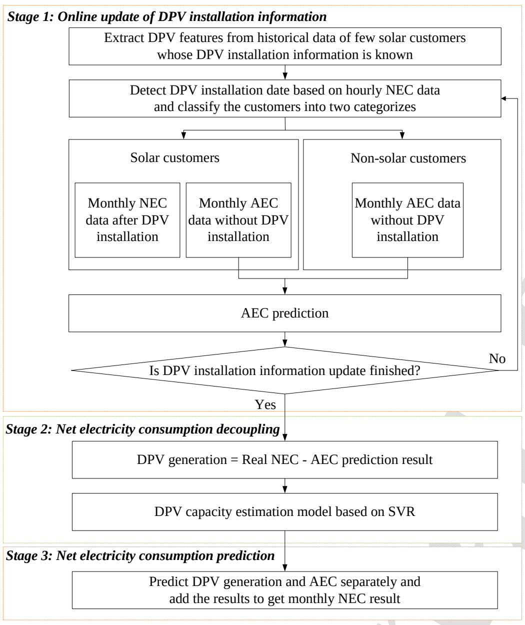

Fig. 5.Data updates and prediction framework

Fig. 5 illustrates the overall framework of the monthly NEC prediction method with an online DPV installation information update. It concludes with three stages on the whole and the detailed processes of each stage are introduced in section III. Stage I: *Online Update of DPV Installation Information*. To identify the solar customers and pave the way for NEC decoupling, this stage is divided into two steps. First, customers are classified into solar and non-solar customers by the extracted characteristics characterizing DPV installations. Then, an online DPV installation check is performed for non-solar customers, as shown in Fig. 6. Stage II: *NEC Decoupling*. NEC is decoupled into AEC and DPV generation. The DPV capacity information is afterward obtained from the DPV generation as the input. Stage III: *Monthly NEC Prediction*. The monthly NEC is obtained by separately predicting the monthly AEC and the monthly DPV generation.

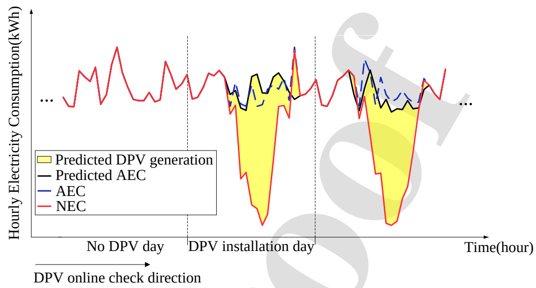

Fig. 6.Online update check diagram

3. Proposed methodology

3.1 Feature Exaction and Classification

The difference between NEC and AEC comes from the amount of DPV generation. Typical features to distinguish the DPV installation are extracted using hourly electricity consumption data from a small number of customers who are known to be solar customers. The fine-grained data helps study the monthly electricity consumption [26]. The monthly NEC curves smooth out fluctuations compared with the daily electricity consumption curve. It is easier and more obvious to extract the DPV features from the daily electricity consumption curve.

During the DPV generation period, the actual electricity consumption curve is abated by DPV generation to the extent that it first increases to a peak and then decreases. Based on this, it is mathematically described from different perspectives, as mentioned in [27]. When installing DPV, the magnitude of the slope will decrease during the period of sunrise. The magnitude and rate of change of the slope in the morning are considered features. The calculations are shown below:

$$s_{cl,g} = \left(P_{Net,H_{g+\Delta g},M_i,Y_j} - P_{Net,H_g,M_i,Y_j}\right) / \left(H_{g+\Delta g} - H_g\right) \quad (2)$$

$$r_{s,g} = \left(s_{cl,g+\Delta g} - s_{cl,g}\right) / \left(H_{g+\Delta g} - H_g\right) \quad (3)$$

where $s_{cl,g}$ is the slope of the cut line at H_g. $r_{s,g}$ is the average rate of change of the slope during H_g and $H_{g+\Delta g}$. T denote time and consist of the hour H_g, month M_i, and year Y_j, namely $T = \{H_g, M_i, Y_j\}, t \in T$. Among them, *g*, *i*, and *j* represent the number of hours *H*, month *M* and year *Y* respectively.

The irradiation intensity at noon is the greatest during the day, so the magnitude of NEC decreases the most during this time. The degree of concavity at noon is calculated as follows:

$$r_{tp,g} = \left(P_{Net,H_{g_2},M_i,Y_j} - P_{Net,H_{g_1},M_i,Y_j}\right) / \left(H_{g_2} - H_{g_1}\right) - P_{Net,H_g,M_i,Y_j}, g \in [g_1, g_2] \quad (4)$$

where $r_{tp,g}$ represents the degree of concavity during H_{g_1} and H_{g_2} . P_t denote the electricity consumption of each customer whose DPV information is unknown and $P_t = \left\{ P_{Net,t}, P_{Load,t}, P_{DPV,t} \right\}$. Let P_t^0 be the electricity consumption of each customer whose DPV information is known and define $P_t^0 = \left\{ P_{Net,t}^0, P_{Load,t}^0, P_{DPV,t}^0 \right\}$.

Customers are classified into solar customers and non-solar customers according to the change of features before and after the installation of DPV. For known solar customers, the features show a significant difference before and after the installation of DPV. The changes before and after the installation of DPV for the same feature are recorded as references. For other customers, these features of the daily electricity consumption series are calculated in turn. By determining whether the changes of each feature of the two adjacent days match the references, the customers are classified into two categories. If yes, the customer is classified as a solar customer. If not, the customer is currently a non-solar customer.

3.2 DPV Installation Update Check

Let Ω_1 denote the number of solar customers after classification, $l_1 \in \Omega_1$; Ω_2 denote the number of non-solar customers after classification, $l_2 \in \Omega_2$. To capture the new DPV installations by non-solar customers, the hourly electricity consumption series data of non-solar customers are predicted. The hourly AEC data of each non-solar customer is predicted per day based on an artificial neuron network [28] and the prediction errors per time are recorded. The error between the prediction results and the actual electricity consumption data collected from the smart meters is calculated. The non-solar customer whose hourly AEC prediction error is higher than the average of the former error ε_{av} is suspected of installing DPV newly and the equations are as follows:

$$\varepsilon_{av} = \frac{1}{m}\sum_{k=1}^{m}\varepsilon_k \tag{5}$$

$$\left| \hat{y}_{Load,H_g,M_i,Y_j} - y_{Load,H_g,M_i,Y_j} \right| > \varepsilon_{av} \tag{6}$$

where ε_k is the error of the prediction of P_{Load,H_g,M_i,Y_j} and k is the time of the prediction. $\hat{y}_{Load,H_g,M_i,Y_j}$ is the real AEC of P_{Load,H_g,M_i,Y_j} and y_{Load,H_g,M_i,Y_j} is the predicted result of AEC P_{Load,H_g,M_i,Y_j} .

Further verification is required to determine whether the error is caused by the DPV installation and to exclude the cause of the prediction method. The time-series data contains trend and fluctuation information [29]. The random fluctuations of the AEC and NEC sequences during the daytime DPV generation period are different. Based on this, the analysis compares the stochastic fluctuations of the adjacent two-day electricity consumption series. The stochastic fluctuations $P_{Net,fluctuation}$ in the AEC series are obtained by removing the trend $P_{Net,trend}$ from the actual consumption series, as shown in the following:

$$P_{Net,fluctuation} = P_{Net,t} - P_{Net,trend} \tag{7}$$

Wavelet packet analysis divides the time-frequency plane more carefully, and it has a higher resolution of the high-frequency part of the signal than the binary wavelet [30]. Wavelet package decomposition is performed on the electricity consumption series to extract the high-frequency components. It is determined which segment of high-frequency information is distinguishable between AEC and NEC by comparing the decomposed hourly AEC and NEC data before and after the DPV installation for known solar customers, which is then used to determine whether a non-solar customer becomes a solar customer. As a result, if the above two validations are met, the prediction error is determined to be due to the addition of DPV, implying that this non-solar customer becomes a solar customer.

3.3 Net Electricity Consumption Decoupling

The difference between AEC and NEC is DPV generation. Before installing DPV, the curves of hourly AEC and the hourly NEC are overlapped. According to the hourly AEC prediction of the day before the date of installation of DPV, the amount of DPV generation is obtained by using the hourly AEC on the day of installation of DPV minus the hourly AEC prediction results, as shown in the following equation:

$$y_{Load,M_i,Y_j} = \sum_{q=1}^{k_2} y_{Net,H_q,M_i,Y_j} + \sum_{q=k_1}^{k_2} y_{DPV,H_q,M_i,Y_j} \tag{8}$$

where k_2 is the total number of the hours in DPV installation month M_i *and* k_1 is the number of the hours after installing DPV in the month M_i .

The predicted DPV generation after decoupling is shown as:

$$y_{DPV,H_q,M_i,Y_j} = y_{Load,H_q,M_i,Y_j} - y_{Net,H_q,M_i,Y_j} \tag{9}$$

3.4 DPV Capacity Estimation

The customer may install the DPV on any day of the month, so the smart meter collects the NEC data after the installation of the DPV. To restore the monthly AEC data, the actual NEC data is used to add the predicted DPV generation in the days after the installation of DPV. In light of the support vector regression (SVR) model's benefits in handling small sample data and excellent generalization [31], the model is based on SVR. Let $\boldsymbol{x}$ be the vector of the inputs of SVR, namely $\boldsymbol{x} = \left[x_{wc}, x_{te}, x_g \right]$. It includes weather conditions x_{wc} (such as sunny, rainy, and cloudy), temperature x_{te} , and DPV generation x_g . Let $\boldsymbol{y}$ denote the DPV capacity information of a few solar customers whose DPV information is known. A small number of the known solar customer hourly data is set as the training set, and the rest of the solar customer hourly data is set as the test set. Targeted training samples are as follows:

$$D = \left\{ (x_1, y_1), (x_2, y_2), (x_3, y_3), \cdots, (x_n, y_n) \right\} \tag{10}$$

$$f(x_i) = w^T x_i + b \tag{11}$$

where $f(x)$ is a function expected to be obtained through learning, and w and b are the parameters to be determined.

3.5 Monthly Net Electricity Consumption Prediction

To make the output of the prediction model better follow the actual results, the seasonal trend is removed. Differential operations can smooth a class of non-smooth sequences (i.e., sequences with a trend). The essence of the differential operator is to extract determination information by self-regression:

$$\nabla^d x_t = (1-B)^d x_t = \sum_{i=0}^{d} (-1)^i C_d^i x_{t-1} \tag{12}$$

where d is the differential order. B is the value of the Bayesian information criterion and C is the value of the Akaike information criterion (AIC). AIC criterion for determining the order of the model: S is the total number of unknown parameters of the model. $\hat{\sigma}^2$ is the estimate of the variance of the series. N is the length of the series.

$$AIC(S) = \ln \hat{\sigma}^2 + \frac{2S}{N} \tag{13}$$

Then the monthly AEC and DPV generation are predicted separately based on the auto-regressive integrated moving average (ARIMA) of order (*p*, *d*, *q*):

$$\nabla^d X(t) = \sum_{i=1}^{p} \phi_i \nabla^d X(t-i) + a_t - \sum_{j=1}^{q} \theta_j a_{t-j} \tag{14}$$

where ∇^d is the *d*th difference operator, a_t is white noise with variance σ_a^2, ϕ_i and θ_j are model parameters [32]. The AEC and DPV generation obtained by NEC decomposition have seasonal cycle variations to different degrees. A differential operation with a step size of the cycle length for a series with a fixed cycle can usually extract the cycle information better. The ARIMA model parameters *p* and *q* are determined by the trailing and truncated tails of the auto-correlation and partial correlation coefficients respectively.

3.6 Evaluation Criteria

To quantify the accuracy of the DPV installation update check validations, the accuracy a_{update} is calculated by the following:

$$a_{update} = n_{cor} / n_{tol} \tag{15}$$

where n_{cor} is the correct identification number of non-solar customers converted to solar customers. n_{tol} is the real number of all non-solar customers to be identified that are converted to solar customers.

To quantify the accuracy of the proposed method, the error is evaluated using the mean absolute percentage error (MAPE) and root mean squared error (RMSE):

$$MAPE = \frac{1}{m} \sum_{t=1}^{m} \left(\left| y_{pre_t} - y_{ac_t} \right| / y_{ac_t} \right) \tag{16}$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{t=1}^{m} \left| y_{pre_t} - y_{ac_t} \right|^2} \tag{17}$$

where y_{pre_t} is the predicted value of NEC and y_{ac_t} is the real value of NEC. m is the number of the sequence y_{pre_t} and y_{ac_t}.

4. Case Study

4.1 Dataset and Parameter Settings

The dataset used in our work is from an Australian grid named Ausgrid [33]. The framework proposed in this paper is for a real-life situation where DPV penetration is constantly changing. Based on a half-hourly dataset of 300 known solar households over three years, including hourly DPV generation and hourly AEC, DPV installation dates and DPV generation have been randomly selected to be added to the non-solar customers' AEC series. Thus, DPV penetration is increasing by adding solar customers. Three scenarios have been set up as shown in Fig. 7. As the penetration of DPV changes, it is obvious that the difference between monthly NEC and actual monthly electricity consumption becomes larger. Scenario 1: three hundred customers without DPV for three years; Scenario 2: One hundred customers without DPV installation and two hundred solar customers, of which two hundred solar customers' DPV installation dates are randomly determined over three years, reaching 27% penetration by the end of the third year; Scenario 3: three hundred solar customers by the end of the third year, reaching a penetration of 39%. The DPV installation dates for three hundred solar customers have been randomly set over the three years.

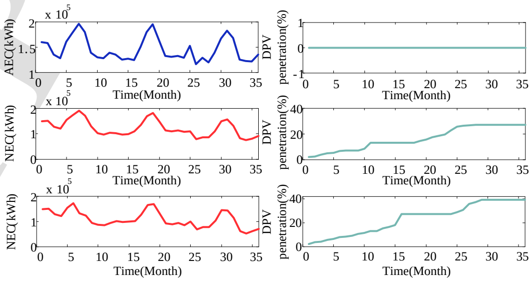

Fig. 7. Three scenarios of DPV penetration setting

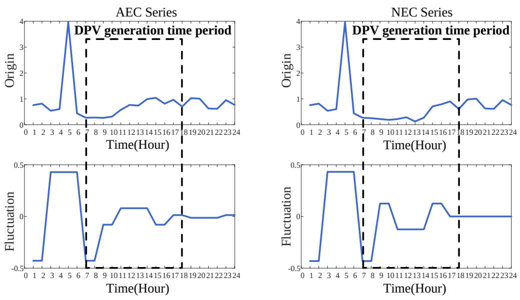

Fig. 8.Comparison of the fluctuation of electricity consumption sequence before and after the installation of DPV by one customer

4.2 Results and Analysis

Fig. 8 illustrates the comparison of the fluctuation of electricity consumption sequence before and after the installation of DPV by one customer. Although the AEC curve and the NEC curve are very similar due to the small output of DPV, the characteristics of the AEC series are found to be changed during the noon hour after the installation of DPV by observing the fluctuation of the two curves. As seen in Fig. 9, adding the validation of random fluctuations in the electricity consumption series has a positive effect on improving the accuracy of DPV installation updates. This is because when the weather is bad or the DPV capacity is small, it is somewhat strained to determine whether to install DPV only by the average error of AEC prediction.

The monthly prediction results of NEC, AEC and DPV generation based on the method proposed in this paper under scenario 3 are respectively shown in Fig. 10, Fig. 11 and Fig. 12. It can be seen that the proposed method overall performs well. Above all, the performance evaluation of the proposed method and direct NEC prediction based on ARIMA under different scenarios is shown in Table 1. When in the absence of DPV installations, the direct prediction method is slightly better than the method proposed in this paper, because the cumulative error of decomposition and reorganization is larger in this case. However, when the simulated scenario is that the DPV penetration rate changes over time, the method proposed in this paper has the advantage of detecting the installation of DPV and updating the database of solar customers, making the monthly NEC prediction by the decomposition better than the direct prediction. This advantage becomes obvious as the DPV penetration increases.

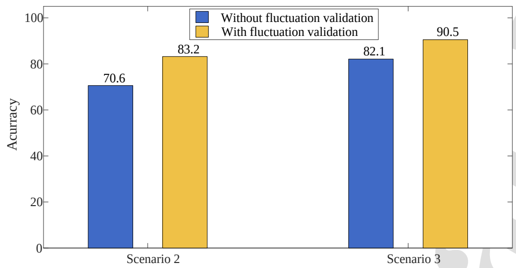

Fig. 9. Comparison of the accuracy of the DPV installation update check with and without fluctuation validation in scenario 2 and scenario 3

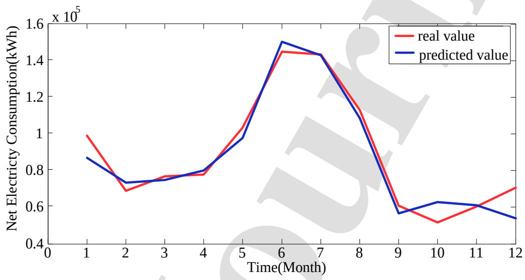

Fig. 10. Monthly NEC prediction results under scenario 3

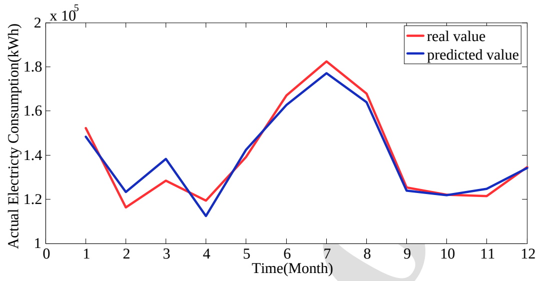

Fig. 11. Monthly AEC prediction results under scenario 3

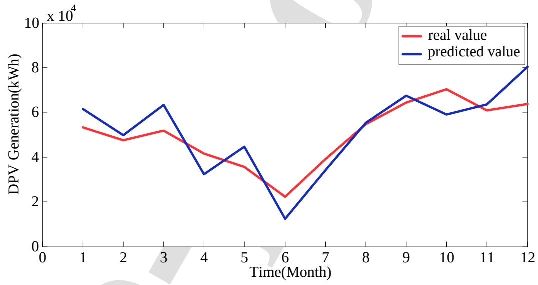

Fig. 12. Monthly DPV generation prediction results under scenario 3

Table 1 Performance evaluation of the proposed method and ARIMA under different DPV penetration

Scenario	Method	MAPE	RMSE (kW)
1	Proposed Method	0.1174	14372.59
	ARIMA	0.0983	14461.18
2	Proposed Method	0.0927	11892.62
	ARIMA	0.1094	13574.31
3	Proposed Method	0.0825	10876.52
	ARIMA	0.1126	13892.45

5. Conclusion

Some of the factors affecting DPV and AEC are the same but have different mechanisms of action, such as weather. There are also different influencing factors, such as calendar effects for AEC and national policies for DPV. Considering the different patterns of variation of DPV and AEC under the action of multiple influencing factors, the idea of decoupled prediction is proposed. To cope with the difficulty of uncertainty added to the monthly electricity consumption prediction by the random additions of DPV, a monthly electricity consumption prediction method under the framework of online update of DPV information is proposed. Given the continuous updating of the DPV database, the process of change in DPV penetration is simulated in this paper, making the prediction of DPV generation more realistic. The validation that the fluctuations of monthly AEC series are injected new characteristics of DPV generation is used to determine the former non-solar customers have installed DPV, which ensures that the newly installed DPV generation is subtracted. The proposed monthly NEC prediction model taking into account DPV installation information update outperforms the direct model based on ARIMA under high DPV penetration.

Though the DPV installation check for each non-solar customer in turn is easier to find DPV additions, it also increases the time of prediction. Future work could improve this point and extend the proposed method to monthly NEC prediction considering the demand response to verify its applicability [34-36]. Furthermore, the integration of various distributed renewable energy generation [37-40] and a massive amount of data collected from advanced metering infrastructure will bring more challenges.

References

[1] Lee B. L., Wilson C., Simshauser P., et al., Deregulation, efficiency and policy determination: An analysis of Australia's electricity distribution sector. Energy Economics 2021; 98: 105210.

[2] Ajayi V., Weyman-Jones T., State-level electricity generation efficiency: Do restructuring and regulatory institutions matter in the US. Energy Economics 2021; 104: 105650.

[3] Carrion M., Conejo A. J., Arroyo J. M., Forward Contracting and Selling Price Determination for a Retailer. IEEE Trans. Power Syst. 2007; 22(4): 2105-2114.

[4] B. Kroposki, et al., Achieving a 100% Renewable Grid: Operating Electric Power Systems with Extremely High Levels of Variable Renewable Energy. IEEE Power and Energy Magazine 2017; 15(2):61-73.

[5] Chen X., Li Z., Wang F., et al., Monthly Net Electricity Consumption Prediction Under High Penetration of Distributed Photovoltaic System, 2021 International Conference on Smart Energy Systems and Technologies (SEST), 2021, pp. 1-6.

[6] Sossan F., Nespoli L., Medici V., et al., Unsupervised disaggregation of photovoltaic production from composite power flow measurements of heterogeneous prosumers. IEEE Trans. Ind. Inform. 2018; 14(9): 3904–3913.

[7] Bu F., Dehghanpour K., Yuan Y., et al., A Data-Driven Game-Theoretic Approach for Behind-the-Meter PV Generation Disaggregation. IEEE Trans. Power Syst. 2020; 35(4): 3133-3144.

[8] Van der Meer D.W., Munkhammar J., Widén J., Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals. Sol. Energy 2018; 171: 397-413.

[9] Stainsby W., Zimmerle D., Duggan G. P., A method to estimate residential PV generation from net-metered load data and system install date. Appl. Energy 2020; 267: 114895.

[10] Sun M., Zhang T., Wang Y., et al., Using Bayesian Deep Learning to Capture Uncertainty for Residential Net Load Forecasting. IEEE Trans. Power Syst. 2020; 35(1): 188-201.

[11] Van der Meer D.W., Shapero M., Svensson A., et al., Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes. Appl. Energy 2018; 213: 195-207.

[12] Kobylinski P., Wierzbowski M., Piotrowski K., High-resolution net load forecasting for micro-neighbourhoods with high penetration of renewable energy sources. International Journal of Electrical Power & Energy Systems 2020; 117: 105635.

[13] Nouha D., Julien E., Stéphane T., et al., A survey of modelling and smart management tools for power grids with prolific distributed generation. Sustainable Energy, Grids and Networks 2020; 21: 100284.

[14] Nibir B. R., Debapriya D., Optimal allocation of active and reactive power of dispatchable distributed generators in a droop controlled islanded microgrid considering renewable generation and load demand uncertainties. Sustainable Energy, Grids and Networks 2021; 27: 100482.

[15] Van der Meer D.W., Widén J., Munkhammar J., Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew. Sustain. Energy Rev. 2018; 81(1): 1484-1512.

[16] Van der Meer D.W., Munkhammar J., Widén J., Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals. Sol. Energy 2018; 171: 397-413.

[17] Shaker H., Zareipour H., Wood D., Estimating Power Generation of Invisible Solar Sites Using Publicly Available Data. IEEE Trans. on Smart Grid 2016; 7(5): 2456-2465.

[18] Shaker H., Zareipour H., Wood D., A Data-Driven Approach for Estimating the Power Generation of Invisible Solar Sites. IEEE Trans. on Smart Grid 2016; 7(5): 2466-2476.

[19] Karthick T., Charles Raja S., Jeslin Drusila Nesamalar J., et al., Design of IoT based smart compact energy meter for monitoring and controlling the usage of energy and power quality issues with demand side management for a commercial building. Sustainable Energy, Grids and Networks 2021; 26: 100454.

[20] Stegner C., Glaß O., Beikircher T., Comparing smart metered, residential power demand with standard load profiles. Sustainable Energy, Grids and Networks 2019; 20: 100248.

[21] Choksi K. A., Jain S., Pindoriya N. M., Feature based clustering technique for investigation of domestic load profiles and probabilistic variation assessment: Smart meter dataset. Sustainable Energy, Grids and Networks 2020; 22: 100346.

[22] California ISO. What the duck curve tells us about managing a green grid. California ISO, Shaping a Renewed Future, 2012.

[23] Wang Y., Zhang N., et al. Data-Driven Probabilistic Net Load Forecasting with High Penetration of Behind-the-Meter PV. IEEE Trans. on Power Syst. 2018; 33(3): 3255-3264.

[24] Mallal Y., Sharma D. K., Bahir L. E., et al., Temperature prediction-based realistic performance analysis of various electrical configurations of solar PV panels. Sol. Energy 2021; 228: 612-624.

[25] Cao G., Wu L., Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. Energy 2016; 115(1): 734-745.

[26] Hong T., Wilson J., Xie J., Long Term Probabilistic Load Forecasting and Normalization With Hourly Information. IEEE Trans. Smart Grid 2014; 5(1): 456-462.

[27] Wang F., Li K., Wang X., et al., A Distributed PV System Capacity Estimation Approach Based on Support Vector Machine with Customer Net Load Curve Features. Energies 2018; 11(7): 1750.

[28] Chau T. K., Yu S. S., Fernando T., et al., A Load-Forecasting-Based Adaptive Parameter Optimization Strategy of STATCOM Using ANNs for Enhancement of LFOD in Power Systems. IEEE Trans. on Ind. Inform. 2018; 14(6): 2463-2472.

[29] Gonzalez-Romera E., Jaramillo-Moran M. A., Carmona-Fernandez D., Monthly Electric Energy Demand Forecasting Based on Trend Extraction. IEEE Trans. on Power Syst. 2006; 21(4): 1946-1953.

[30] The MathWorks, Wavelet Packets: Decomposing the Details, [Online]. Available: https://ww2.mathworks.cn/help/wavelet/ug/wavelet-packets-decomposing-the-details.html. [Accessed Mar.18, 2021].

[31] K. Li, F. Wang, Z. Mi, M. Fotuhi-Firuzabad, N. Duić, T. Wang, Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation, Applied Energy, 2019; 253: 113595.

[32] Box G.E.P., Jenkins G.M., Time Series Analysis. Forecasting and Control. San Francisco: Holden-Day, 1976.

[33] Ausgrid, Solar home electricity data, [Online] Available: https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data. [Accessed Mar.18, 2021].

[34] K. Li, J. Yan, L. Hu, F. Wang, and N. Zhang, Two-stage Decoupled Estimation Approach of Aggregated Baseline Load under High Penetration of Behind-the-Meter PV System. IEEE Trans. Smart Grid 2021; 12(6): 4876-4885.

[35] F. Wang, X. Lu, X. Chang, X. Cao, S. Yan, K. Li, N. Duić, M. Shafie-Khah, and J.P.S. Catalao, Household profile identification for behavioral demand response: A semi-supervised learning approach using smart meter data. Energy 2022; 238, Part. B: 121728.

[36] F. Wang, K. Li, C. Liu, Z. Mi, M. Shafie-Khah, and J. P. S. Catalao. Synchronous pattern matching principle-based residential demand response baseline estimation: Mechanism analysis and approach description. IEEE Trans. Smart Grid 2018; 9(6): 6972–6985.

[37] Worighi I., Maach A., Hafid A., et al., Integrating renewable energy in smart grid system: Architecture, virtualization and analysis. Sustainable Energy, Grids and Networks 2019; 18: 100226.

[38] X. Lu, K. Li, H. Xu, F. Wang, Z. Zhou, and Y. Zhang, Fundamentals and Business Model for Resource Aggregator of Demand Response in Electricity Markets, Energy, 2020, 204: 117885.

[39] Q. Chen, F. Wang, B.M. Hodge, J. Zhang, Z. Li, M. Shafie-Khah, and J.P.S. Catalao, Dynamic price vector formation model-based automatic demand response strategy for PV-assisted EV charging stations, IEEE Trans. Smart Grid, 2017, 8(6): 2903-2915.

[40] Fathabad A. M., Cheng J., Pan K., Qiu F., Data-Driven Planning for Renewable Distributed Generation Integration. IEEE Trans. on Power Syst. 2020; 35(6): 4357-4368.