A SPATIO-TEMPORAL FORECASTING METHOD FOR DISPERSED PV PLANTS

Maria Malvoni1*, Nikos Hatziargyriou1

1 School of Electrical and Computer Engineering, National Technical University of Athens, Greece.
*maria.malvoni@gmail.com

Keywords: PHOTOVOLTAIC POWER FORECASTING, PREDICTIVE MODELS, CORRELATION, WAVELET TRANSFORM, SUPPORT VECTOR MACHINES.

Abstract—The study proposes a novel approach based on the three-dimensional wavelet transform (3D-DWT) and the Least Square Support Vector Machines (LS-SVM) to forecast the power output of spread PV systems. The proposed forecasting method can capture the correlations of distributed PV installations providing more accurate predictions. The output power of different PV plants in Rhodes, Greece is applied to forecast the power generation of each PV system at 1, 12 and 24 hours-ahead time horizons. The accuracy is also expressed by the error metrics.

1. Introduction

The grid-connected solar PV systems require more flexibility of the electric grid. The knowledge in advance of PV solar generation is an important issue for the system operators for the balancing of load and generation in the grid operations successfully. Accurate forecasts can provide an effective support in power system management, leading to higher integration of the variable PV solar plants [1]. Regardless of the large advances achieved in the developing of PV generation forecasting models, some issues need to address. The generation prediction for individual PV plant remains the most used approach. Furthermore, the recent smart grid enables to exchange of generation data from the distributed PV installations, supporting the forecasts task by exploring of potential correlations of the spread PV systems and by developing the spatio-temporal forecast approaches [2]. However, few studies present forecasting methods based on spatio-temporal analysis of distributed solar PV plants. Asadi et al. [3] combined a time series decomposition and fuzzy clustering approach to investigate spatial and temporal patterns of power generation. The corresponding main features were used as input of neural network to get spatial-temporal forecasts, outperforming the results with respect to the state-of-the-art neural network models. A collection of different sensors network is investigated to analyze their spatio-temporal dependencies in order to improve PV generation forecasts accuracy up 20% in comparison with common forecasting methods [4]. The geographical correlation can help to identify a suitable sensors network for neighboring PV plants by exploring different spatial-time resolutions [5]. In the smart grid of Évora, Portugal the output power of residential PV systems is predicted at 6 hours ahead by applying a spatial–temporal forecasting method based on the autoregressive method with exogenous input. Such approach improves the forecast accuracy up to 10% in comparison to an regressive model [6]. The spatial and temporal correlations of different solar generation plants are implemented in a probabilistic forecasting method based on the Gaussian Conditional Random Fields (GCRF). The forecasts reach more accurate results than the common models such as the persistent model and the autoregressive model [7]. The power forecasts for PV plants in the South of Italy were performed by a Compressive Spatio-Temporal Forecasting (CSTF) algorithm, using data from different meteorological stations [8]. The PV generation forecasts for several PV rooftop were performed by gradient boosted regression tree (GBRT) model in using weather predictions at multiple sites across Japan [9]. However, the machines learning methods have already demonstrated high potential for PV power forecasting [10]-[11]. In this context Artificial Neural Networks and Support Vector Machines play an important role. As well as the data mining techniques enable to handle the big data from PV generation systems [12][13].

In such contest, the current study aims to contribute in the development of spatio-temporal forecasting models and to get more accurate generation forecasts for individual PV power plants by a novel spatial-temporal forecasting model. The wavelet transform is applied to three-dimensional input dataset of collected PV power output to explore the correlations in space and time domain. The Least Square Support Vector Machines (LS-SVM) is then applied to predict the PV power generation of individual PV installation for different hours ahead. The spatio-temporal forecasting model was originally published in [14]. Furthermore, the accuracy assessment of the generation forecasts of individual PV plants at 24 hours-ahead was discussed. The present work provides an extension of results shown in [14]. Here the power generation predictions are performed at 1 and 12 hours-ahead time horizons and compared with the performance originally reported in [14] for 24 hours ahead. The paper is organized as follows: methodology is shown in Section II. Section III presents the error indexes for the forecast accuracy assessment. Results of a case study are discussed in Section IV, Section V reports the conclusions.
2. Methodology

In this section the method to predict the PV generation of different and dispersed PV installations using the 3D data is shown. First, we introduce the Least Squares Support Vector Machines and the Wavelet Transformation.

2.1 Least Square Support Vector Machine

The Least Squares Support Vector Machines (LS-SVM) is a statistical learning technique widely applied for nonlinear questions, through primal-dual representation of the input variables [15]. For a given set \( S_N = \{x_k, y_k\}_{k=1}^N \) of \( N \) elements, with \( x_k \in \mathbb{R}^d \) the \( k \)-th input variable and \( y_k \in \mathbb{R} \) the \( k \)-th output variable, a nonlinear estimation \( \hat{y} \) can be given by:

\[
\hat{y} = w^T \varphi(x) + b \quad (1)
\]

with \( \varphi: \mathbb{R}^d \rightarrow \mathbb{R} \) unknown function, \( w \in \mathbb{R} \) the weight vector and \( b \in \mathbb{R} \) the bias. The optimization problem minimizes the cost function \( J(w, e) \) as follows:

\[
\min_{w,e} J(w, e) = \frac{1}{2}w^T w + \frac{1}{2} \sum_{k} e_k^2 \quad (2)
\]

where \( e_k \) is an artificial variable and \( y \) the regularization factor. An estimation of \( y \) can be given by:

\[
\hat{y} = \sum_{k=1}^N \alpha_k K(x_k, x_k) + b \quad (3)
\]

with \([\alpha_1, ..., \alpha_N]^T\) the variables vector in dual space and \( K(x_k, z) \) the kernel matrix defined as:

\[
K(x_k, z) = \varphi^T(x_k) \varphi(z) \quad k = 1 \ldots N \quad (4)
\]

The Radial Basis Function (RBF) is a common Kernel function adopted, defined as:

\[
K(x_k, x_j) = \exp \left( -\frac{\|x_k - x_j\|^2}{\sigma^2} \right) \quad k, j = 1 \ldots N \quad (5)
\]

with a \( \sigma \) tuning parameter.

2.1 Wavelet Transformation

The wavelet decomposition shifts and translates a signal \( x(t) \) as follows [16]:

\[
X_{\psi}(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \psi \left( \frac{t-b}{a} \right) x(t) dt \quad (6)
\]

with \( \psi(t) \) the mother wavelet, the parameters \( a \) and \( b \) represent the translation parameter and shifted parameter in time domain respectively. The Discrete Wavelet Transform (DWT) shifts and scales the mother wavelet, to get \( n \) representations of the original signal \( x(t) \), as follows:

\[
x(t) = \sum_{n=1}^{n} c_{1,n} \psi_{1,n}(t) \quad (7)
\]

with \( \psi_{1,n}(t) \) is the father function and \( c_{1,n} \) are the wavelet coefficients at level \( n \).

The Fast Wavelet Transform (FWT) [16] implements the DWT by the approximation and detail components that can be further decomposed into approximation and detail until a given level. The Inverse Discrete Wavelet Transform (IDWT) enables the algorithm to reconstruct the original signal.

This section presents the method to predict the PV generation of different and sparse PV installations using the 3D dataset as input. Given \( p_n(t) \) the hourly data related to the output power of individual PV plant with \( n = 1, 2 \ldots N \) where \( N \) the number of PV systems, \( i \) is the \( i \)-th observation (\( i = 1, 2 \ldots T \)), \( T \) are the total observations, so the approach build a \( N \) by \( N \) matrix corresponding to the geographical coordinates for a number \( N \) of PV installation. In order to have a three-dimensional data input, it organizes the \( N \) time series \( p_n(t) \) in a \( N \)-by-\( N \)-by-\( T \) matrix in which the third coordinate represents the hourly temporal step. Successively, it implements the discrete wavelet transform of the three-dimensional data matrix and uses the corresponding detail coefficients as data input to train the LS-SVM and to predict the wavelet decomposition’s coefficients at different head time horizons. Finally, it applies the 3D-IDWT to obtain the power output for each single PV site at each time horizon. Fig. 1 illustrates the adopted approach. Further details related to model implementation are clearly presented in [14].

![Flow chart of the proposed forecasting approach.](image)

The forecast accuracy assessment of the spatial-temporal forecasting approach is performed by adopting of the normalized error \( e(i) \), the Normalized Mean Bias Error (NMBE) and the Normalized Mean Absolute Error (NMAE) defined as:

\[
e(i) = \frac{\bar{x}(i) - x(i)}{\max x(i)} \quad (8)
\]

\[
NMBE = \frac{1}{M} \sum_{i=1}^{M} e(i) * 100 \quad (9)
\]

\[
NMAE = \frac{1}{M} \sum_{i=1}^{M} |e(i)| * 100 \quad (10)
\]

where \( \bar{x}(i) \) represents the PV power forecast achieved by the proposed method, \( X(i) \) is the actual value of the PV power at \( i \)-th time instant, \( M \) is the size of the testing dataset.

3. Results

The proposed method is investigated considering nine PV systems located Rhodes, Greece. The ground-mounted PV systems have a nominal power of about 1 MW. Fig. 2 shows the position of PV systems over the Greek island. Additional specifications can be found in [14]. The model is implemented to forecast the PV power at 1, 12 and 24 hours ahead, and the accuracy is also assessed. The hourly power generation data of each PV system related to whole years 2014, 2015 and 2016 was derived by PVGIS [17]. Noted the geographical coordinates, the power output data of each PV plant for years 2014 and 2015 was organized in a 3D matrix. The power generation data of the whole year 2016, for a total of 8760 samples, was considered to test the forecasting model and to evaluate its performance.
The Wavelet Toolbox™ [18] and the LS-SVMlab Toolbox [19] in Matlab® environment were used to perform the output power forecasts for individual PV plants. The 3D-WDT was carried out at level 3 with the wavelet ‘sym4’ and the LS-SVM was tested to forecast the output power at 1, 12 and 24 hours horizon. The errors analysis of the nine PV plants is performed using NMBE and NMAE.

NMBE and NMAE for 1, 12 and 24 hours ahead forecasts of each PV plant are depicted in Fig. 3. The NMBE varies between -0.16% for PV2 and 0.25% for PV5 respectively at 12 hours ahead. Considering that NMBE provides the forecast error on average, the proposed model overestimates the power generation for PV1, PV5 and PV8, which show positive errors for each time horizon with highest errors for the PV5 plant. The power output related to the remaining installations is underestimated for 1, 12 and 24 hours ahead. This means that the forecasting model does not change its trend in over-forecasting and under-forecasting at different time horizons.

The NMAE is in the range between 0.04% for PV7 at 24 hours ahead and 1.66% for PV5 at 12 hours ahead. It is clear that the 12 hours-ahead forecasts exhibit higher NMAEs than 1 and 24 hours ahead. Indeed, the solar irradiance is a 24 hours periodic signal. Considering that the PV power generation maintains the same trend of the solar irradiance signal in the absence of faults of the PV system, for a given time instant during a day, it is easier to predict 1 and 24 hours ahead, than 12 hours ahead due to the solar irradiance periodicity. Furthermore, the worst error is for the PV5 plants and the best values is for the PV7 installation. In fact, Fig.4 shows the normalized errors \( e(i) \) for the PV5 and PV7 systems at different time horizons related to the first week of the year 2016. It can be seen that the blue curves are always above the red curves that demonstrates higher errors in the predicted power for PV5 than PV7 for each hour ahead.

The normalized error distributions plotted in Fig. 5 for each system highlight the shift of the histograms related the occurrences of the hourly errors for each plant for three-time horizons. It is clearly noticeable that the error distributions at 12 hours are low with long tails, different than in the case of 1 and 24 hours, where the histograms are narrower and with notable peaks. This implies that negative and positive errors have the similar probability to occur. Furthermore, one can possible identify the shift left and shift right of the distribution indicating that it is more probable to underestimate and to overestimate the power generation respectively.

Therefore, considering that the histograms for the PV1, PV5, PV6, PV8 and PV9 are shifted right, their power is generally overestimated. Unlike for the PV2, PV3, PV4 and PV7 plants of which the histograms are shifted left, the output power is in average underestimated at both 1 and 24 hours.
The correlation coefficients of the bias errors at different time horizons are plotted in Fig.6. Correlation values higher than 0.7 are put in evidence. So, it can be seen that in the case of the 12 hours ahead forecasts, the errors are strictly correlated. The forecast errors for PV2, PV3 and PV4 show a correlation among them at 1 and 24 hours ahead. Furthermore, the bias of PV5 and PV8 plants are correlated in 1 hour and 1 day predictions.

A geographical disposition of NMAE is plotted in Fig. 7. At 1 and 24 hours ahead the furthest away PV systems present the highest absolute errors, meanwhile the installations located in the internal area of investigation show lower error, demonstrating that they can benefit from the neighboring plants to obtain improved forecasts.

4. Conclusions

This paper presents a forecasting model for PV output power based on the 3D wavelet decomposition and the LS-SVM using PV power data of different PV installations. The three-dimensional wavelet transform allows to integrate spatial information, improving the forecast accuracy. Power generation data of nine PV plants installed in the Rodhes (Greece) is used to investigate the performance at 1, 12 and 24 hours-ahead. Bias error varies between -0.16% and 0.25% at 12 hours ahead. Absolute error is in the range between 0.04% at 24 hours ahead and 1.66% at 12 hours ahead. Results highlight the importance to consider the neighboring PV plants in order to improve the forecast performance.
5. Acknowledgment

The research presented was funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 799835.

6. References

[1] Raza, M.Q., Mithulananthan, N. Ekanayake, C., ‘On recent advances in PV output power forecast’, Solar Energy, 2016, 136, pp. 125-144.

[2] A. Tascikaraoglu, ‘Evaluation of spatio-temporal forecasting methods in various smart city applications’, Renewable and Sustainable Energy Reviews, 2018, 82 (1), pp 424-435.

[3] R. Asadi, A.C. Regan, ‘A spatio-temporal decomposition based deep neural network for time series forecasting, Applied Soft Computing, 2020, 87, 105963.

[4] X. G. Agoua, R. Girard and G. Kariniotakis, "Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production," IEEE Transactions on Sustainable Energy, 2018, 9, (2), pp. 538-546.

[5] Amaro e Silva, R., C. Brito, M. ‘Impact of network layout and time resolution on spatio-temporal solar forecasting’, Solar Energy, 2018, 163, pp. 329-337.

[6] R. J. Bessa, A. Trindade and V. Miranda, ‘Spatial-Temporal Solar Power Forecasting for Smart Grids’, IEEE Transactions on Industrial Informatics, 2015, 11 (1), pp. 232-241.

[7] B. Zhang, P. Dehghanian and M. Kezunovic, ‘Spatial-temporal solar power forecast through use of Gaussian Conditional Random Fields’, IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, 2016, pp. 1-5.

[8] A. Tascikaraoglu et al., ‘Compressive Spatio-Temporal Forecasting of Meteoerological Quantities and Photovoltaic Power’, IEEE Transactions on Sustainable Energy, 2016, 7(3), pp. 1295-1305.

[9] Persson, C., Bacher, P., Shiga, T., Madsen, H. ‘Multi-site solar power forecasting using gradient boosted regression trees’, Solar Energy, 2017, 150, pp. 423-436.

[10] S. Sobri, S. Koohi-Kamali, N. A. Rahim, ‘Solar photovoltaic generation forecasting methods: A review’, Energy Conversion and Management, 2018, 156, pp. 459-497.

[11] Malvoni M.; Chaib Y.,’ Machine Learning Based Approaches for Modeling the Output Power of Photovoltaic Array in Real Outdoor Conditions’, Electronics, 2020, 9, 315.

[12] M. Malvoni, M. G. De Giorgi, P. M. Congedo, ‘Forecasting of PV Power Generation using weather input data preprocessing techniques’, Energy Procedia, 2017, 126, pp 651-658.

[13] Gabriel de Freitas Viscondi, Solange N. Alves-Souza, ‘A Systematic Literature Review on big data for solar photovoltaic electricity generation forecasting’, Sustainable Energy Technologies and Assessments, 2019, 31, pp. 54-63.

[14] Malvoni, M., Hatziaigiou, N., One-day ahead PV power forecasts using 3D Wavelet Decomposition, 2019, SEST 2019 - 2nd International Conference on Smart Energy Systems and Technologies, art. no. 8849007. DOI: 10.1109/SEST.2019.8849007

[15] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, ‘Least Squares Support Vector Machines’, World Scientific, 2002.

[16] S. Mallat, ‘A wavelet tour of signal processing’. Academia, 1999.

[17] http://re.jrc.ec.europa.eu/pvgis/ accessed 06 June 2020

[18] https://it.mathworks.com/products/wavelet.html accessed 06 June 2020

[19] http://www.esat.kuleuven.be/sista/lssvmlab/ accessed 06 June 2020