PHOTOVOLTAIC FORECAST BASED ON HYBRID PCA-LSSVM USING DIMENSIONALITY REDUCTED DATA

M. Malvoni, M.G. De Giorgi, P.M. Congedo,*

Dipartimento di Ingegneria dell'Innovazione, Università del Salento, via per Monteroni, 73100 LECCE, Italy

*Corresponding Author: mariagrazia.degiorgi@unisalento.it

ABSTRACT

The power forecasting plays a significant role in the electrical systems. Furthermore the high-dimensional data reduction without losing essential information represents an important advantage in the forecasting models. Low computational costs and short execution time together with high predicted performance are the main goals to be reached in the development of a prediction method. In this paper a hybrid method based on an active selection of the support vectors, using the quadratic Renyi entropy criteria in combination with the principal component analysis (PCA), is shown to dimensionally reduce the training data in the forecasting models. The reduced data have been used to implement the Least Squares Support Vector Machines (LS-SVM) in order to predict the photovoltaic (PV) power in the day-ahead time horizon. The model has been validated using historical data of a PV system in the Mediterranean climate. Additionally the weather variations have been taken into account to evaluate the outcome of the sunny and cloudy condition in the PV forecasting models. The proposed technique gives fulfill results. A training data size same as 30% original dimension allows to improve the forecasting accuracy and reduces the computational time of 70% respect to an implementation without dimensionality reduction data.

Highlights

- A dimensionally reduction of the time series data, based on the quadratic Renyi entropy criteria and in combination with PCA is shown.
- LS-SVM models are applied to predict the PV output power in a one-day head frame.
- Photovoltaic forecasting model is performed using the historical PV power data.
- The predicted method takes in account the weather fluctuation, as sunny and cloudy condition.
- The accuracy and executive time of the hybrid method has been investigated.

Keywords

Photovoltaic forecast, Least squares support vector machines, Principal component analysis, Dimensionality reduction, Quadratic Renyi entropy.

1. INTRODUCTION

Nowadays the implementation of models for the prediction of the output of a renewable power plant is a popular research topic [1-2]. The generation systems prediction, as well the load forecasting, is essential into the electricity network. The predictions are the basis for an efficient scheduling and dispatch of the energy in the electric grid. Accurate forecasting methods reduce the operating costs and enhance reliability associated with the integration of renewable systems into the existing electricity grid.

In the literature intelligent forecasting methods as artificial neural networks (ANNs) have been widely used in developing wind and solar forecast models [3-6]. However, ANNs require a complex training process. In the recent years Kernel based estimation techniques, such as Support Vector Machines (SVMs) and Least Squares Support Vector Machines (LS-SVM), have been also applied as powerful nonlinear regression methods suitable for renewable power forecasts [7-9]. SVMs are more resistant to the over-fitting problem with respect to ANNs and permit to achieve high generalization performance in solving forecasting problems of various time series. LS-SVM is simpler and computationally less expensive, even if presents the same advantages of the ANNs and SVMs models [10].

In [8] ANN and LS-SVM were compared. Authors underlined that LS-SVM based models outperform ANNs in the prediction of the photovoltaic power.

Usually the PV power time series show important seasonal patterns (yearly, weekly, intra-daily patterns) that need to be taken into account in the modeling strategy. Furthermore the model is trained on the times series data, but when the available dataset is large, the learning model becomes more expensive in terms of time and computational resources.

The PV power is affected by weather and topographic factors, such as temperature, irradiance, humidity, which lead to large PV power variations. Furthermore the measured data could contain random fluctuations given by measurement errors and random factors. The meteorological fluctuations influence the prediction results with different impact on the forecasting accuracy if the weather is sunny or cloudy [8]. Previous works [8, 11 - 12] applied the wavelet transform to reduce the noise contained in the data to be used in the forecasts. The weather data are well correlated, this permits the use of historical time series of these data to implement forecasting models. Nevertheless the redundant information can lead difficult modeling if the historical data, used as the model input, are a highly correlated. The principal component analysis (PCA) is one of traditional technique, extensively applied, to eliminate redundant information and improve the accuracy of the forecast of renewable power [13-15]. The PCA permits to extract essential features and to reduce high-dimensional data into low-dimensional ones, which serve as inputs for the forecasting methods, as neural network or support vector machine, with a reduction of the CPU time. Some researchers used dimensionally reduction techniques in the support vectors for short term load forecasting [16] or wind prediction [17-18]. Despite its advantages, this technique has been rarely applied in the field of photovoltaic power prediction.

The main goals are to build a forecasting model that takes into account meteorological conditions together with the seasonal patterns of the PV power and to investigate the model accuracy when a dimensionally reduction is applied to the training dataset. The historical time series which consist of hourly values of the PV power, the module temperature, the ambient temperature and the plane of array irradiance, recorded for a period of about 1 year, are used to learn the model. In order to take into account the weather variation, as sunny and cloudy conditions, additional parameters are introduced in the learning set. This leads to an increase of the data set size. Furthermore, an efficient method to dimensionally reduce the learning data set is proposed based on the active selection of the recorded hourly samples in accordance to the quadratic Renyi entropy criteria and the implementation of the PCA to eliminate redundant information. The reduction technique is tested to predict the PV output power in a one-day head frame of a system located in a Mediterranean climate implementing the LS-SVM. A detailed analysis is carried out to evaluate the implemented models performance. The executive time and the computational complexity of the hybrid method are investigated.

This research will allow to improve the forecasting models accuracy of the PV power production in view of the computational resources needed to implement them.

This paper is organized as follows: the section 2 describes the theory regarding the LS-SVM, the quadratic Renyi entropy, the PCA decomposition and the proposed procedure. In the section 3 the dataset and problem statement are presented. The metric to evaluate the performance of model are illustrated in section 4. The results of the method to dimensionality reduce are discussed in section 5. Finally Section 6 states the conclusions of the paper.

2. THEORY AND METHODS

This section explains the theory that is the base of the proposed model in the present work and how it is applied to forecast PV power output in the proposed method.

2.1. Least Square Support Vector Machine

The Least-squares support vector machines allows to build a nonlinear representation of the original inputs using positive-definite kernel functions, based on a primal-dual formulation.

Given a training set of N data points, $\mathcal{D}_N = \{x_k, y_k\}_{k=1}^{N}$where $x_k \in \mathbb{R}^d$ is the k-th input data and $y_k \in \mathbb{R}$ is the k-th output data, a regression model can be constructed using $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}^s$ unknown function map of the input space:

$$y_k = w\varphi(x_k) + b \quad k = 1 \dots N \qquad (1)$$

where $w \in \mathbb{R}^s$ is the weight vector and $b \in \mathbb{R}$ is the bias term. The regression equation is transformed into an optimization problem with constraint that means to minimize a cost function:

$$\text{mix}_{w,e}\, \mathcal{J}\,(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_k^N e_k{}^2 \quad k = 1 \dots N \qquad (2)$$

where e_k is an artificial variable and γ is the regularization factor. In order to solve this optimization problem, the Lagrange function is defined as:

$$L(w, b, e, \alpha) = \mathcal{J}(w, e) - \sum_{k=1}^{N} \alpha_k \{y_k[w\varphi(x_k) + b] - 1 + e_k\} \quad k = 1 \dots N \qquad (3)$$

where $\alpha_k \in \mathbb{R}$ are the Lagrange multipliers. The optimal conditions are:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 & \rightarrow \mathrm{w} = \sum_{j=1}^{N} \alpha_j\, \varphi(x_j) \\ \frac{\partial L}{\partial b} = 0 & \rightarrow \sum_{k=1}^{N} \alpha_k = 0 \\ \frac{\partial L}{\partial e_j} = 0 & \rightarrow \alpha_j = e_j \gamma \quad j = 1 \dots N \\ \frac{\partial L}{\partial \alpha_j} = 0 & \rightarrow y_j = w\varphi(x_j) + b + e_j \end{cases} \qquad (4)$$

Applying the Mercer's theorem [19]:

$$\varphi^T(x_k)\varphi(x_j) = \mathrm{K}(x_k, x_j) \quad k, j = 1 \dots N \qquad (5)$$

The Radial Basis Function (RBF) is introduced as kernel function K and defined as:

$$\mathrm{K}(x_k, x_j) = \exp\left(-\frac{\|x_k - x_j\|_2^2}{\sigma^2}\right) \qquad (6)$$

where σ is a tuning parameter. So, the solution in matrix notation is:

$$\begin{bmatrix} \Omega + \frac{1}{\gamma} I & 1 \\ I^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (7)$$

where $\Omega_{kj} = K(x_k, x_j)$ is the kernel matrix and $\alpha = [\alpha_1, \ldots\ldots \alpha_N]^T$ are the variables in dual space. So, the approximate model of y is expressed in following form:

$$y(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \quad (8)$$

So calculating α and b, it is possible obtain an estimation of y in dual representation.

2.2. Approximation of the feature map for LS-SVM dual representation

The solution of Eq.(8) permits to determinate α and b, it is need to solve a system of size N x N that is unfeasible when the number N is more large. Furthermore it is advantageous to have a finite-dimensional approximation of the feature map $\hat{\varphi}: \mathbb{R}^d \rightarrow \mathbb{R}^M,\ M \ll N$. So, an expression for the approximation of the feature map, based on the Nyström approximation [20-22], is given as follows:

$$\hat{\varphi} = \frac{M}{\sqrt{\lambda_{i,M}}} \sum_{m=1}^{M} u_{mi,M} K(x, x_m) \quad (9)$$

where $\lambda_{i,M}$ is the i-th eigenvalue and the $u_{mi,M}$ is the *m*-th element of the *i*-th eigenvector for a M x M kernel matrix with i=1,…M. Hence, an M-approximation of y can be obtained replacing the expression for the approximation of the feature map given by Eq. (9) in Eq.(1):

$$\hat{y}_m = w\hat{\varphi}(x_m) + b \quad m = 1 \ldots M \quad (10)$$

The Eq. (10) requests a subsample of size *M* of the original training set.

2.3. Active selection of the support vector based on quadratic Renyi entropy criteria

The subsample selection is a very important step of the algorithm because it influences the model accuracy. The selection criteria can use a random selection or an active selection. A popular active selection method, based on the density distribution of the sample, consists to maximize the quadratic Renyi entropy [23 - 24].

Given a training subset of M data points with different probabilities $\{p_1, \ldots p_M\}$, the Renyi entropy of order α ($\alpha \geq 0$, $\alpha \neq 1$) is defines as [25]:

$$H_\alpha(x) = \frac{1}{1-\alpha}\log(\sum_{m=1}^{M} p_m{}^{\alpha}) \qquad (11)$$

It measures the degree of the set randomization. More random samples involve more large entropy [26 - 27]. A case of particular interest is α=2, noted as quadratic Renyi entropy.

If the probability density function is a Gaussian function, the quadratic Renyi's entropy can be calculated as follows [20, 28]:

$$H_2(X) = -\log\left(\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} K\left(x_i, x_j\right)\right) \qquad (12)$$

The quadratic Renyi entropy can be used as criteria to select a subsample of dataset based on the high value of the entropy [29]. The active selection algorithm for a training subset is shown in Table 1.

2.4. Principal Component Analysis

The Principal Components Analysis (PCA) is a statistical technique that allows to reduce the size of the original data set, taking in to account the uncorrelated variables and removing the redundant information. The PCA technique is based on the research of the most significant variations of the variables [30]. The covariance method is one of approach to the computing PCA [31].

Given a dataset $\mathcal{S}$ of dimensions Q × M where M is the observations number of Q variables, it is possible to reduce a new subset, represented by L variables, L< Q. A matrix B of dimension M × Q is defined as:

$$B = \mathcal{S} - h\,u^T \qquad (13)$$

where $u[j] = \frac{1}{M}\sum_{i=1}^{M}\mathcal{S}[i,j]$ and $h[i] = 1$ with j=1…Q and i=1…M. The Q × Q covariance matrix C of matrix B is:

$$C = \frac{1}{M-1} B^T B \qquad (14)$$

The D diagonal matrix Q x Q of eigenvalues of C

$$D[k,l] = \lambda_k \text{ for } k = l \qquad D[k,l] = 0 \text{ for } k \neq l \qquad (15)$$

where λ_k is the kth eigenvalues of the covariance matrix C. The Q x Q eigenvectors matrix V that diagonalizes the covariance matrix C is:

$$V^{-1}CV = D \qquad (16)$$

Furthermore V contains Q column vectors, each of length Q, which represent the Q eigenvectors of the covariance matrix C. It is need sort the columns of the matrix V in order of decreasing eigenvalue of the matrix D. The columns of Q × L matrix W are the first L columns of V as:

$$W[k,l] = V[k,l] \qquad (17)$$

with k=1,..Q and l=1,…L where L<Q.

Hence the dataset $\mathcal{S}$ can be transformed into a new M-by-L space by:

$$\bar{\mathcal{S}} = BW \qquad (18)$$

In the PCA it is necessary to choose an appropriate number of the components in order to not worsen the accuracy of the predicting model [32]. Generally, the first principal component defines the highest percentage of the variance of the samples and the second one refers to the next highest variance.

2.5. PV forecasting method with dimensionally reduction of training dataset

The dimensionality reduction preprocessing method, as explained previously, can be applied to the training dataset for the PV output power prediction by LS-SVM

The original training dataset is a N-by-P matrix, where N rows represent the observations of P monitored parameters at every time instant. The dimensionally reduction of the training set consists to select M rows (M<N) and R columns (R<P) starting from the original training matrix without loss the information.

The value M that represents the number of the support vectors is freely chosen by the user. The maximum value of M that can be used depends on the computational resources and the features of the dataset.

Fixed M, the quadratic Renyi entropy criteria is used to select a subsample of dataset. So a new data matrix is obtained of dimension M x P, whose rows represent the perfect vector (PVT). Subsequently in order to reduce further the size of each support vector, the PCA has been performed, eliminating the principal components that contribute less than 2% to the total variation in the data set. The identification procedure of the PVT in combination with the decomposition through the principal components (PVTPC technique) produces a new training dataset of reduced dimension M x R that is used for the training of the LS-SVM in the prediction the PV power (Fig. 1).

In the present study the RBF has been chosen as the kernel function. During the training of LS-SVM an optimal selection of the regulation constant of the cost function γ and the tuning parameters σ of the kernel function has been performed using a 10-fold cross-validation procedure, maximizing the performance of each implementation. Table 2 shows the main steps of the proposed approach to forecast PV power using the LS-SVM with the dimensionally reduction PVTPC technique of training data.

3. DATASET AND PROBLEM STATEMENT

The collected data are related to the PV systems, sited in the campus of the University of Salento, in Monteroni di Lecce (LE), Puglia. More details of the PV systems are reported in [33 - 36]. The $\mathcal{A}_F$ time series dataset consists of F hourly values of the main parameters recorded for a period of about 1 year, from March 2012 until March 2013 (9.192 hourly records). In particular the collected data are:

- the PV power, P (kW)
- the module temperature, T_m (°C)
- the ambient temperature, T_a(°C)
- the irradiance on plane of module with tilt 3, I_3 (W/m^2)
- the irradiance on plain of module with tilt 15, I_{15} (W/m^2).

For each hour i of the observed records, the normalized PV power value has been calculated as follows:

$$\hat{P}(i) = \frac{P(i)}{Max_{i=1}^{F}(P(i))} \quad (19)$$

Fig.2 shows the normalized hourly PV power for one week of four months (May, August, November and February), observing the power changes within the time of the day and the months. Furthermore an irregular trend of PV power is observed in several hours of some days.

To better take into account this behavior, the clear sky irradiance model was used. When the solar radiation passes through the atmosphere, it is subjected to phenomena of scatting and absorption by air molecules, water, and dust. Hence the solar radiation on the surface of the earth is lower than in the extraterrestrial radiation. In the clear sky model the atmosphere is neglected and the solar radiation on the earth' surface is equal to the solar radiation outside the earth [37].

In this paper a PVGIS (Photovoltaic Geographical Information System) tool, development by Joint Research Center – Institute for Energy and Transport (JRC IET) and available at webpage [38], is used to evaluate the hourly clear sky irradiance on tilted plane 3° and 15° at the Latitude 40°21' - Longitude 18°11'.

The difference between the measured (I_k) and clear sky (I_{sky}) irradiance is estimated as follows:

$$G_k(i) = \frac{I_{sky,k}(i) - I_k(i)}{I_0} \qquad i=1....F \qquad (20)$$

where I_0=1.000W/m^2 is the solar radiance at Standard Test Condition and k is equal to the inclination of modules (3° or 15°).

Considering the tilted of PV modules equal to 15°, the measured irradiance data I_k and the difference G_k are normalized as follows:

$$I_{15,n}(i) = \frac{I_{15}(i)}{Max_{i=1}^{F}(I_{15}(i))} \qquad I_{sky,15,n}(i) = \frac{I_{sky,15}(i)}{Max_{i=1}^{F}(I_{sky,15}(i))} \qquad G_{k,n}(i) = \frac{G_k(i)}{Max_{i=1}^{F}(G_k(i))} \qquad (21)$$

Fig. 3 shows $I_{15,n}$, $I_{sky,15,n}$ and $G_{15,n}$ on one week for each considered months. As expected, the $I_{sky,15,n}$ is always higher than $I_{15,n}$ because of neglecting the effects of the atmosphere. It is evident the seasonal variability in irradiance, as expected, the normalized measured irradiance in spring (May) is higher than in winter (February).

Furthermore the irradiance behavior is well related to the weather conditions and permits to identify the sunny days when the trend of $I_{15,n}$ is quite regular, therefore $G_{15,n}$ is low, while cloudy days correspond to fluctuation of $I_{15,n}$ and peaks of $G_{15,n}$.

Hence it's evident that the irradiance has an hourly variation, seasonal dependence and its behavior is related to the different weather conditions, consequently the PV power is influenced by the effects of the sunny or cloudy days and the seasons.

For this reason an efficient prediction technique of the PV output power in a short time should take into account the different weather conditions. This leads to increase the training dataset dimension with high computational and temporal resource requirements. A possible solution to optimize the resource consists in the dimensionally reduction of the original data without loss information. Hence, a main aim of this study is the implantation of a suitable method to dimensionality reduce the training data sets.

4. PERFORMANCE EVALUATION

A code in Matlab® software, with the LS-SVMlab Toolbox [39], was developed using a PC equipped with a Intel(R) Xeon(R) CPU E5-1650 3.20 GHz CPU and 8-GB RAM for showing the applicability of the proposed method.

The performance of the proposed algorithm is evaluated by the Normalized Mean Absolute Error (NMAE) that is defined as:

$$NMAE(h) = \left(\frac{1}{N_{TS}} \sum_{i=1}^{N_{TS}} |e_h(i)|\right) \qquad (22)$$

where i is the generic hour of the predicted data from 1 to the number of the testing sample N_{TS}; h is hourly prediction horizon and the $e_h(i)$ is the normalized error defines as follows:

$$e_h(i) = \widehat{T}_h(i) - \widehat{P}_h(i) \qquad (23)$$

where

$$\widehat{P}_h(i) = \frac{\overline{P}(i)}{\mathrm{Max}_{i=1}^{N_{TS}}(P_h(i))} \qquad (24)$$

is the normalized predicted PV power at generic hourly instant i for the time horizon h;

$$\widehat{T}_h(i) = \frac{T_h(i)}{\mathrm{Max}_{i=1}^{N_{TS}}(T_h(i))} \qquad (25)$$

is the normalized PV power value used as testing data at hour i for time horizon h. The target at the h prediction time horizon has been defined as the sum of the next h values at hour i of the power values as:

$$T_h(i) = \sum_{i+1}^{i+h} P(k) \qquad \forall\, h = 1, 2, 3\ldots.24 \qquad (26)$$

5. RESULTS AND DISCUSSION

A preliminary analysis of the impact of the weather conditions on the PV power prediction has been carried out taking into account different input vectors to learn the LS-SVM which has been performed to predict the PV power in the day-ahead time frame.

A sample of 4 months (May, August, November and February), approximately the same as the 65% of records (N=6.336 hourly observations) has been used to train and the 35% remaining of data (N_{TS}=2.856) to test the model. The value of M has been chosen equal to M_1=192, M_2=672 and M_3=1920 which represent two, seven and twenty days for each season.

5.1. Definition of the input vectors

Three different input vectors (IV) of the available historical values of exogenous variables have been defined in order to take in account the dynamic variations of PV power at different weather conditions. The first IV (IV1) is composed by the collected hourly data as previously described. In the second one (IV2), in order to follow the weather fluctuations, the difference G_k between the measured and clear-sky model irradiance value (Eq. 21) has been replaced to the measured irradiance values of IV1. A third input vector (IV3) considers the IV2 and in addition the hourly instant and the season at which the data have been measured. The use of IV3 allows to track the daily and seasonal cycle of the PV power. The “season” variable has been considered by introducing a numerical vector {1, 2, 3, 4} to identify the four seasons: spring, summer, autumn and winter (Fig.4).

5.2. Forecasting model results

Firstly, the LS-SVM without dimensionally reduction of the training data has been applied using the three input vectors in order to evaluate the NMAE and to compare the performance. Fig. 5 plots NMAE in the day-ahead time frame. When the time horizon increases, the NMAE increases more quickly in cases of IV1 and IV2 than IV3. It's clear that the highest NMAE is obtained for the training based on IV1 and that the IV3 better performs the LS-SVM for all investigated look-ahead horizons with NMAE in the range [2%; 12%].

Fig.6 and Fig.7 show the normalized forecasted PV power (Eq.24), implementing the three different IV for 1h and 12h time horizons. The comparison with the normalized target (Eq. 25) is plotted for the second week of each month of the test. A good agreement between the target and forecasted values of all models is evident at 1h. Instead at 12h, the models have a tendency to underestimate the PV power during the hot months and overestimate during the cold month and in the cloudy days.

To better understand this behavior, in Fig.8 and Fig.9 the normalized error, as defined by Eq.23, is plotted together with the normalized value $G_{15,n}$ that allows to track the weather variations. In general the normalized error is in the range [-0.2; 0.2], so it's lower at 1 hour than 12h ahead prediction time. Substantial differences for the three models are not evident, they mainly follow the same trend. In particular, focusing on 1 hour prediction horizon, high values $G_{15,n}$ lead to high values of the error e_h that assumes positive polarity when the $G_{15,n}$ increases and negative polarity if the $G_{15,n}$ decreases. At this horizon, the error appears more dependent by the parameter $G_{15,n}$. Instead at 12h prediction horizon high errors occur even if the $G_{15,n}$ assumes a value near to zero. In particular in sunny days ($G_{15,n} \approx 0$), the IV1 and IV2 underestimate the PV power during the hot months and overestimate in the autumn and winter. Differently, the IV3 seems to be less affected by the parameter $G_{15,n}$.

Furthermore, the IV3 is able to track the fluctuation of PV power at different weather conditions with the better performance. Hence, the IV3 has been chosen to implement the PVTPC technique.

The three dimensionality reduced datasets have been used to train the LS-SVM in order to forecast the PV power at five time horizons (1h, 3h, 6h, 12h and 24h).

The predicted PV power by each subsample M is compared with the target value for the five analyzed horizons (Fig. 10 and Fig. 11). At 1h time horizon the forecasted values of all subsamples are quite in agreement with the target. An evident overestimation is evident in the autumn and winter and at the cloudy days of the spring and summer.

To evaluate the forecasting performance for each subsample, the NMAE has been calculated and plotted in Fig. 12. The results show that the LS-SVM improves its accuracy with larger values of M. The use of M_1 leads to lower performance with respect to the M_2 and M_3 at 1h time horizon. When the time horizon increases, the difference between M_1 and M_2 becomes less noticeable. Significant improvements are obtained in case of M_3.

In Fig. 13 the error statistical distributions have been also plotted at five time horizons. For all subsamples the trend is quite similar. For each training subset M, the highest values of the probability are given when the errors assume values in the range [-30%; -20%]. Hence the forecasted PV power is generally underestimated by the proposed strategy. Better results are obtained with M_3. When the time horizon increases, the probability decreases. For long time horizons, the histograms are flat. It means that the models have equal tendency to overestimate or underestimate.

Finally, the NMAE of the proposed technique PVTPC with M_3 is plotted in Fig.14 and compared with the performance of the ANN and LS-SVM with Wavelet Decomposition (WD), which were analyzed in [8]. It's evident that the novel approach presents the best results with a substantial reduction of the prediction error for each time horizon.

Additionally, the CPU time has been estimated and compared for implemented models at 24h prediction time horizon as reported in Table 3. In the analysis of the executive time, the method based on PCA-LSSVM without the dimensionally reduction of the IV3 (LS-SVM PCA IV3) and the PVT method without the PCA decomposition for three subsamples (LS-SVM PVT IV3 – M_1; LS-SVM PVT IV3 – M_2; LS-SVM PVT IV3 – M_3) have been taken into account. As expected, the model simulation with IV3 requests more time because the number of input variables is greater than IV1 and IV2. In LS-SVM PCA IV3, the PCA introduces not significant reduction of the executive time. The LS-SVM PVT allows to test the models with a computation time reduced of the 70% with respect to the baseline case with all the observations. The LS-SVM PVTPC and LS-SVM PVT are different of few seconds. It means that the reduction of the computational time is mainly given by the PVT method rather than by PCA, with final results very satisfactory. Computational complexity of the proposed algorithm is analyzed. With reference to the step 2 of the Algorithm 2 (Table 2), the entropy value (Eq.12) is given by the sum of all elements of kernel matrix that has a computational cost of M^2 [29, 40]. In Algorithm 1 (Table 1) for each time a new sample vectors is considerate thought the active selection therefore the new entropy value has to be compute that replaces the previous one and then can be removed from the memory [29]. At the step 3 the PCA takes $O(M^2Q)$ time to compute the covariance matrix and $O(M^3)$ time to eigendecompose it [41].The kernel matrix is again computed using the decomposed set $\bar{S}_M$ with a computational cost of M^2 at the step 4. The calculating the Nyström approximation with eigenvalue decomposition of the kernel matrix requires $O(M^3 + M^2N)$ at the step 5 [29, 42]. The total complexity of the proposed method is then given by the sum of each complexities $(2 + Q + N) M^2 + NM^3$. Considering that the number of sample N is fixed and the number variables Q<<N and, the resulting complexity depends on M approximately as $NM^2 + NM^3$ and the algorithm is $O(M^4)$.

6. CONCLUSION

In this work the LS-SVM model has been applied to predict the photovoltaic power at different time horizons up to the day-ahead time. An analysis of the external variables has been carried out to define the parameters that more influence the training of the model. Furthermore in the definition of the input vector of the forecasting model, in addition to the historical power output data, it is necessary to take into account the measured irradiance, related to a clear-sky model in order to include the meteorological variations in terms of sunny and cloudy weather. Hence, three different input vectors have been identified and a sample set of N= 9192 observations has been implemented in the LS-SVM. The results show that an appropriated choice of the external variables in the training dataset leads to significant improvements, with a decrease of the forecasting error up to 10% at long prediction time horizon.

The PVTPC dimensionally reduction technique has been also applied. Three subsamples of support vectors with different dimensions have been selected based on the quadratic Renyi entropy criteria and reduced again with the classical principal component analysis. The results show that a number of support vectors M equal to 30% of the dimension of original training set leads to a decrease of the NMAE of 4% in the prediction at 24 hours with respect to cases in which M is smaller. Also it is observed a reduction of the computational time of 70% respect to the case in which all the original observations have been used for the training.

The efficient selection of the support vectors depends on the computational resources. Even if the computational complexity of the proposed technique is higher than that of the simple method, it allows to improve the prediction performance. This aspect is particularly important for the online implementation in the electric market. Forecasting models with high performance permit to reach the greatest revenue with lower costs for unbalancing penalty.

In future studies, techniques of dimensionally reduction will be developed to avoid the redundancy of the input data that represent the same external variables.

REFERENCE

[1] De Giorgi, M.G., Ficarella, A., Tarantino, M. Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods. Energy (2011), 36(7): 3968-3978

[2] De Giorgi MG, Ficarella A, Tarantino M.. Error analysis of short term wind power prediction models. Applied Energy (2011), 88: 1298-1311

[3] A. Mellit, S.A. Kalogirou, Artificial intelligence techniques for photovoltaic applications: a review Progr. Energy Combust. Sci. (2008), 34(5):574–632

[4] M. Alexiadis Short-term forecasting of wind speed and related electrical power. Solar Energy (1998), 63 (1), 61–68

[5] A.S.S. Dorvlo, J.A. Jervase, A.A. Lawati Solar radiation estimation using artificial neural networks. Applied Energy (2002), 71:307–319

[6] Skittides, C., Früh, W., Wind forecasting using Principal Component Analysis, Renewable Energy (2014), 69:365-374

[7] De Giorgi MG, Campilongo S, Ficarella A, Congedo PM.Comparison between wind power prediction models based on wavelet decomposition with leastsquares support vector machine (LS-SVM) and artificial neural network (ANN). Energies (2014), 7(8):5251–72

[8] De Giorgi, M.G., Congedo, P.M., Malvoni, M., Laforgia, D., Error analysis of hybrid photovoltaic power forecasting models: A case study of mediterranean climate. Energy Conversion and Management (2015), 100:117-130 DOI: 10.1016/j.enconman.2015.04.078

[9] Gala, Y., Fernández, A., Díaz, J., R. Dorronsoro, J., Hybrid machine learning forecasting of solar radiation values, Neurocomputing, http://dx.doi.org/10.1016/j.neucom.2015.02.078.

[10] J.A.K. Suykens, T. Van Gestel, J. Debrebanter Least squares support vector machines World Scientific Publishing Co., Singapore (2002)

[11] Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell Jul (1989), 11(7):674–93.

[12] Shayeghi H, Ghasem A. Day-ahead electricity prices forecasting by a modified CGSA technique and hybrid WT in LSSVM based scheme. Energy Conversion Management (2013);74:482–91.

[13] C. Paoli, C. Voyant, M. Muselli, M. Nivet Forecasting of preprocessed daily solar radiation time series using neural networks Sol. Energy, (2010), 84 (12): 2146–2160

[14] Skittides, C., Früh, W.G., Wind forecasting using Principal Component Analysis, Renewable Energy, (69): 365-374, http://dx.doi.org/10.1016/j.renene.2014.03.068.

[15] Da Silva Fonseca Junior, J.G, Oozeki, T., Ohtake, H., Shimose, K., Takashima, T., Ogimoto, K., Regional forecasts and smoothing effect of photovoltaic power generation in Japan: An approach with principal component analysis, Renewable Energy, (68): 403-413, http://dx.doi.org/10.1016/j.renene.2014.02.018.

[16] Espinoza, M., AK Suykens, J, De Moor, J. "Load forecasting using fixed-size least squares support vector machines."Computational Intelligence and Bioinspired Systems. Springer Berlin Heidelberg, (2005), 1018-1026.

[17] Oliver,k., Treiber, N.A. Sonnenschein, M. Wind power ramp event prediction with support vector machines. Hybrid Artificial Intelligence Systems. Springer International Publishing, (2014), 37-48.

[18] Kong, X., Liu, X., Shi, R., Lee, K.Y. Wind speed prediction using reduced support vector machines with feature selection, Neurocomputing, (2015) 169: 449-456 http://dx.doi.org/10.1016/j.neucom.2014.09.090.

[19] Vapnik V (1998) Statistical learning theory. Wiley, New Yorkut. Management Sci., vol. 3, no. 2, pp. 113–129, April 2006.

[20] Girolami, M., Orthogonal series density estimation and the kernel eigenvalue problem, Neural Computing (1998), 10(6):1455–1480

[21] Williams C., Seeger, M. Using the Nystrom method to speed up kernel machines. Proc. NIPS (2000), 13:682–688

[22] De Brabanter, K., De Brabanter, J., Suykens, J. A., & De Moor, B. Optimized fixed-size kernel models for large data sets. Computational Statistics & Data Analysis, (2010), 54(6), 1484-1504

[23] Espinoza, M., Suykens, J.A.K., De Moor, B. Least squares support vector machines and primal space estimation, Proceedings of the IEEE 42nd conference on decision and control (2003), 5716–5721

[24] Weigend AS, Gershenfeld NA (1994) Time series predictions: forecasting the future and understanding the past. Addison-Wesley, Reading

[25] Principe, J. C. (2010).Information theoretic learning: Renyi's entropy and kernel perspectives. Springer Science & Business Media.

[26] Renyi, A. On measures of entropy and information. Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability, University of California Press, Berkeley, CA, 1(1961), 547-561

[27] Renyi, A. Introduction a la theorie de linformation. Calcul des probabilites. Dunod, Paris, (1966)

[28] Thomas, A. (2008). Sprayed concrete lined tunnels. CRC Press.

[29] K. De Brabanter, J. De Brabanter, J.A.K. Suykens, B. De Moor, Optimized fixed-size kernel models for large data sets, Computational Statistics & Data Analysis (2010), 54(6): 1484-1504, http://dx.doi.org/10.1016/j.csda.2010.01.024.

[30] Samarasinghe, S. (2006). Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. CRC Press.

[31] Engineering Statistics Handbook Section 6.5.5.2, Retrieved 19 January 2015.

[32] I.T Jolliffe Principal Component Analysis (2nd Edition)Springer, New York (2002)

[33] Congedo, P.M., Malvoni, M., Mele, M., De Giorgi, M.G., (2013). Performance measurements of monocrystalline silicon PV modules in South-eastern Italy. Energy Conversion and Management, 68, 1-10 DOI: 10.1016/j.enconman.2012.12.017

[34] De Giorgi, M.G., Congedo, P.M., Malvoni, M., (2014). Photovoltaic power forecasting using statistical methods: Impact of weather data. IET Science, Measurement and Technology, 8(3), 90-97 DOI: 10.3182/20140824-6-ZA-1003.01184

[35] De Giorgi, M.G., Congedo, P.M., Malvoni, M., Tarantino, M., 2013. Short-term power forecasting by statistical methods for photovoltaic plants in south Italy. In: 4th IMEKO TC19 Symposium on Environmental Instrumentation and Measurements: Protection Environment, Climate Changes and Pollution Control, 171-175

[36] Donateo, T., Congedo, P.M., Malvoni, M., Ingrosso, F., Laforgia, D., Ciancarelli, F. An integrated tool to monitor renewable energy flows and optimize the recharge of a fleet of plug-in electric vehicles in the campus of the University of Salento: Preliminary results (2014) IFAC Proceedings Volumes (IFAC-PapersOnline), 19, pp. 7861-7866. DOI: 10.3182/20140824-6-ZA-1003.01184

[37] Duffie JA, Beckman WA. Solar engineering of thermal processes. 4th ed; 2013. doi: http://dx.doi.org/10.1002/9781118671603.

[38] http://re.jrc.ec.europa.eu/pvgis/apps4/pvest.php

[39] http://www.esat.kuleuven.be/sista/lssvmlab/

[40] Bottou, L., & Lin, C. J. (2007). Support vector machine solvers. Large scale kernel machines, 301-320.

[41] Günter, S., Schraudolph, N.N., Vishwanathan, S.V.N. (2007). Fast iterative kernel principal component analysis. Journal of Machine Learning Research, 8, 1893-1918.

[42] Chapelle, O. (2007). Training a support vector machine in the primal. Neural Computation, 19(5), 1155-1178.