



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on Support Vector Machines (SVM) model to forecast photovoltaic power



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

Department of Engineering for Innovation, University of Salento, 73100 Lecce, Italy

ARTICLE INFO

Article history:

Received 30 June 2016

Received in revised form

21 July 2016

Accepted 10 August 2016

Available online 18 August 2016

Keywords:

Least Squares Support Vector Machines (LS-SVM)

Forecast photovoltaic

ABSTRACT

The data concern the photovoltaic (PV) power, forecasted by a hybrid model that considers weather variations and applies a technique to reduce the input data size, as presented in the paper entitled “Photovoltaic forecast based on hybrid pca-lssvm using dimensionality reduced data” (M. Malvoni, M.G. De Giorgi, P.M. Congedo, 2015) [1]. The quadratic Renyi entropy criteria together with the principal component analysis (PCA) are applied to the Least Squares Support Vector Machines (LS-SVM) to predict the PV power in the day-ahead time frame. The data here shared represent the proposed approach results. Hourly PV power predictions for 1,3,6,12, 24 ahead hours and for different data reduction sizes are provided in [Supplementary material](#).

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Engineer, Neurosciences, Physics.
More specific subject area	Renewable resource, Solar energy, Photovoltaic, Forecasting methods
Type of data	Table, Excel file
How data was acquired	Data were obtained implementing the proposed model in [1] through MATLAB code.
Data format	Processed

* Corresponding author.

E-mail address: paolo.congedo@unisalento.it (P.M. Congedo).<http://dx.doi.org/10.1016/j.dib.2016.08.024>2352-3409/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Experimental factors	Sensors for measurements of solar irradiance on plane of array, module temperature and ambient temperature are connected to a Supervisory Control And Data Acquisition (SCADA) system. Thought protocols Modbus, the measured parameters are available in the private web site. Input data are treated to referred them at same time step.
Experimental features	The pre-processed data of outdoor measurements and PV output power represent the input set to implement the forecasting models. LS-SVM in combination with technique to resize the training data is implemented to predict the PV power in ahead day time frame.
Data source location	Lecce, Italy (40° 19'32"16 N, 18° 5'52"44 E)
Data accessibility	Data are within this article

Value of the data

- The provided data may be useful to compare the PV system performance in different climatic conditions.
 - The PV power predictions performed through LS-SVM, in combination with a technique to dimensionally reduce the training data, may be used to compare the prediction methods accuracy.
 - The forecasted PV power can be applied in the planning, operation and management of power systems in view of the future development of smart grids.
 - The prediction of PV power can be used to optimize the electric vehicles integration and to manage the recharging status [2].
-

1. Data

The hourly PV power predicted by LS-SVM, combined with a technique to reduce the size of the training data, is provided for different input data dimensions in [Supplementary file](#) (online). For each size, hourly PV power predictions are referred at 1,3,6,12, 24 ahead hours.

2. Experimental design, materials and methods

2.1. Experimental data

The shared data refer to a grid connected 960 kWp PV system sited in Lecce, Southern Italy (40°19'32"16N, 18°5'52"44E), to supply the electricity to the utilities of the University of Salento Campus. The 3000 monocrystalline silicon PV modules are installed on shelters as car parking roof. The PV modules present the same azimuth angle of 10°, but they are tilted of two different angle, 3° and 15° respectively. More technical details may be found in [3]. An integrated web-connected system is available to measure and monitor the weather parameters and the PV output power. The PV field is equipped of a pyranometer to measure of solar radiation on the module's plane every 1 minute. Temperature probes provide measurements of the ambient temperature and the module temperature every 10 min. PV output power is sampled every 1 min. The data as monitored are collected in web site with private access.

2.2. Data pre-processing

For more accurate predictions, the PV energy forecasting models request the knowledge of meteo parameters [4,5]. Therefore the weather data, referred to one year of measurements from 06/03/2012 to 30/03/2013, was considered for a total number of 9.192 hourly samples. In order to align them to the same sample step of 1 h, a data pre-processing was performed. For this aim, hourly average values of temperatures, solar irradiance and PV power were computed, as defined in [6].

Table 1
Implementation of LS-SVM model for different size of input data set.

	Description
LS-SVM 0	LS-SVM implemented without technique to dimensionally reduce the training data;
LS-SVM M1	LS-SVM implemented with technique to dimensionally reduce the training data to M1 = 192 records;
LS-SVM M2	LS-SVM implemented with technique to dimensionally reduce the training data to M2 = 672 records;
LS-SVM M3	LS-SVM implemented with technique to dimensionally reduce the training data to M3 = 1920 records;

The same meteo data have already been applied by the authors in [7] to investigate a hybrid statistical models based on LS-SVM with Wavelet Decomposition (WD) and in [8] to propose a novel hybrid algorithm GLSSVM (Group Least Square Support Vector Machine), based on the combination of the LS-SVM and a unexplored neural network known as GMDH (Group Method of Data Handling) implementing different strategies (Direct, Recursive and DirRec) for multi-step ahead forecast.

2.3. Data processing method

In [1] a hybrid method based on an active selection of the support vectors, using the quadratic Renyi entropy criteria in combination with the PCA is proposed to forecast the PV power output at different ahead hours.

In order to implement the forecasting methods, the input dataset of the following parameters was considered:

- S , variable to represent the season in which the data was measured;
- H , hourly instant to refer the data;
- T_a , ambient temperature (°C);
- T_{mod} , module temperature (°C);
- $G_{3,n}$ and $G_{15,n}$ solar irradiance difference at tilt 3° and 15° respectively, defined as the difference between the solar irradiance on tilted plane and the related hourly clear sky irradiance for each hour i , and computed as follow:

$$G_k(i) = \frac{I_{sky,k}(i) - I_k(i)}{I_0} \quad i = 1, 2 \dots 9.192 \text{ and } k = 3^\circ, 15^\circ \tag{1}$$

where I_k is the measured irradiance for each plane of array, $I_{sky,k}$ is the corresponded solar irradiance in clear sky condition and $I_0 = 1.000 \text{ W/m}^2$ is the solar radiance at Standard Test Condition. Difference time series $G_k(i)$ was normalized as follows

$$G_{k,n}(i) = \frac{G_k(i)}{\text{Max}_{i=1}^{9.192}(G_k(i))} \quad i = 1, 2 \dots 9.192 \text{ and } k = 3^\circ, 15^\circ \tag{2}$$

- T_h , target of PV output power at the h prediction time horizon given as the cumulated PV power in h consecutive hours and scaled to at its maximum value, defined by:

$$T_h(i) = \frac{\sum_{k=i}^{i+h} P(k)}{\text{Max}_{k,P(i)}} \quad i = 1, 2 \dots 9.192 \text{ and } h = 1, 3, 6, 12, 24 \tag{3}$$

where $P(i)$ is the hourly average values PV power as computed during the data pre-processing.

The hourly mean ambient temperature, module temperature, solar irradiance measured on two tilted planes (I_3 and I_{15}) and the hourly mean PV power are provided in [6]. The time series of $G_{k,n}$ and

T_h were computed. The input dataset was divided in two subsets. The samples referred to May, August, November and February (6336 about the 35% of the total records) were applied to test and the remain months (2856, about 65% of records) to train them.

The outcomes of the forecasting models represent the cumulated and normalized power, predicted for each time horizon $\hat{T}_h(i)$. The output data referred to the train set and the test set are shared in [Supplementary material](#), File 1. The time series $\hat{T}_h(i)$ is provided at five time horizons, implementing the proposed methodology for four cases, as summarized in [Table 1](#). More details are provided in [1].

Transparency document. Supplementary material

Transparency data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.08.024>.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.08.024>.

References

- [1] M. Malvoni, M.G. De Giorgi, P.M. Congedo, Photovoltaic forecast based on hybrid pca-lssvm using dimensionality reduced data, *Spec. Issue Recent Adv. Support Vector Mach.* (2015), <http://dx.doi.org/10.1016/j.neucom.2016.01.104>, <<http://www.sciencedirect.com/science/article/pii/S0925231216305677>>.
- [2] T. Donato, P.M. Congedo, M. Malvoni, F. Ingrosso, D. Laforgia, F. Ciancarelli, 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, in: *The IFAC Proceedings Volumes (IFAC-PapersOnline)*, 19, 2014, pp. 7861–7866. <http://dx.doi.org/10.3182/20140824-6-ZA-1003.01184>.
- [3] P.M. Congedo, M. Malvoni, M. Mele, M.G. De Giorgi, Performance measurements of monocrystalline silicon PV modules in South-eastern Italy, *Energy Convers. Manag.* 68 (2013) 1–10. <http://dx.doi.org/10.1016/j.enconman.2012.12.017>.
- [4] M.G. De Giorgi, P.M. Congedo, M. Malvoni, Photovoltaic power forecasting using statistical methods: Impact of weather data, *Sci., Meas. Technol.* 8 (3) (2014) 90–97. <http://dx.doi.org/10.1049/4-smt.2013.0135>.
- [5] M.G. De Giorgi, P.M. Congedo, M. Malvoni, M. Tarantino, Short-term power forecasting by statistical methods for photovoltaic plants in south Italy, in: *The 4th 5 TC19 Symposium on Environmental Instrumentation and Measurements: Protection Environment, Climate Changes and Pollution Control*, 2013, pp. 171–175.
- [6] M. Malvoni, M.G. De Giorgi, P.M. Congedo, Data on photovoltaic power forecasting models for Mediterranean climate, *Data Brief.* 7 (2016) 1639–1642. <http://dx.doi.org/10.1016/j.dib.2016.04.063>.
- [7] M.G. De Giorgi, P.M. Congedo, M. Malvoni, D. Laforgia, Error analysis of hybrid photovoltaic power forecasting models: a case study of Mediterranean climate, *Energy Convers. Manag.* 100 (2015) 117–130.
- [8] M.G. De Giorgi, M. Malvoni, P.M. Congedo, Comparison of strategies for multi-step ahead photovoltaic power forecasting models based on hybrid group method of data handling networks and least square support vector machine, *Energy* 107 (2016) 360–373. <http://dx.doi.org/10.1016/j.energy.2016.04.020>.