Renewable Energy 80 (2015) 197–204

Contents lists available at ScienceDirect

Renewable Energy

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

Short-term predictability of photovoltaic production over Italy

Matteo De Felice [a, *], Marcello Petitta [a, b], Paolo M. Ruti [a]

[a] Casaccia R.C., ENEA Energy and Environment Modelling Technical Unit, Rome, Italy
[b] Institute for Applied Remote Sensing, EURAC, Viale Druso 1, Bolzano/Bozen, Italy

ARTICLE INFO

Article history:
Received 29 September 2014
Accepted 6 February 2015
Available online

Keywords:
Photovoltaic system
Solar power forecasting
Renewable energy modelling
Solar irradiance

ABSTRACT

Photovoltaic (PV) power production increased drastically in Europe throughout the last years. Since about the 6% of electricity in Italy comes from PV, an accurate and reliable forecasting of production would be needed for an efficient management of the power grid. We investigate the possibility to forecast daily PV electricity production up to ten days without using on-site measurements of meteorological variables. Our study uses a PV production dataset of 65 Italian sites and it is divided in two parts: first, an assessment of the predictability of meteorological variables using weather forecasts; second, an analysis of predicting solar power production through data-driven modelling. We calibrate Support Vector Machine (SVM) models using available observations and then we apply the same models on the weather forecasts variables to predict daily PV power production. As expected, cloud cover variability strongly affects solar power production, we observe that while during summer the forecast error is under the 10% (slightly lower in south Italy), during winter it is abundantly above the 20%.

1. Introduction

Europe is experiencing a growing penetration of photovoltaic (PV) production, in particular Italy that in 2012 had almost 480 000 PV plants (16.4 GW of total installed power) [6], 44% more than in 2011. Modelling of daily electricity generation of PV power systems can be useful for an effective management and balancing of a power grid, supporting real-time operations especially in countries with a considerable amount of solar energy potential. Forecasting the expected PV power production could in fact help to deal with its intermittency, mainly due to weather conditions. Moreover, short-term forecasting information can also be valuable for electric market operators.

PV plant production can be modelled in two ways: with a mathematical model or through a data-driven approach, the latter often called black-box modelling. A mathematical model consists of a set of equations describing the physical behaviour of the photovoltaic module, a good example can be found in Bellini et al. [2], Sandrolini et al. [16] and Massi Pavan et al. [11]. While this approach can be considered physics-aware, the data-driven approach tries to reproduce the behaviour of the system "just" modelling the relationship between observed inputs (e.g. meteorological conditions) and outputs (e.g. power output). Both the approaches have their pros and cons, the former can be more accurate but in addition to weather variables (incoming solar radiation, air temperature, wind speed, etc.) it needs solar panel characteristics (technology, area, orientation, etc.) and their evolution during time (e.g. due to degradation). Conversely, the black-box approach does not require information about the typology of PV panel but it needs long time-series of input and output variables to calibrate a reliable model. From a system identification perspective, the two approaches can be defined respectively white box and black box model modelling. If we go beyond this dichotomous point of view, we may consider a tradeoff between these two approaches, commonly called a "grey box" model (see the paper by Ljung [9] for an interesting introduction on system identification modelling). Since the aim of this paper is an analysis of PV power production forecast without using on-site measurements, we limit our approach to the black-box modelling in order to focus on the uncertainty associated to meteorological data. Furthermore, the black-box choice is also due to the absence of detailed information about solar panel characteristics and site location. However, we plan to explore a more complex modelling approach in a future paper.

* Corresponding author.
E-mail address: matteo.defelice@enea.it (M. De Felice).

http://dx.doi.org/10.1016/j.renene.2015.02.010
0960-1481/

In this work, we use a Support Vector Machine (SVM, briefly introduced later in Sec. 4) to perform the prediction of daily production using both solar radiation and temperature information.

SVMs have been already used for similar applications, Zeng and Qiao [22] tested a SVM-based approach using data from three different sites outperforming both autoregressive and neural network-based models; Bouzerdoum et al. [3] proposed a hybrid SARIMA-SVM approach which performed better than both the single models in predicting hourly power output of a small PV plant; Shi et al. [19] applied a SVM-based approach using weather forecast data on a PV station in China. More in general, black-box methods are common for forecasting applications related to solar power and solar radiation (e.g., see Pedro and Coimbra [14]).

Our work is based on daily power production data of 65 grid-connected PV systems in Italy. For each plant a SVM model has been built and tested with the best available observations of solar radiation and air temperature. Finally, the same SVM models are used for forecasting power production considering as inputs data provided by numerical weather forecasts.

In the next section, we introduce and describe weather and production data that will be used for modelling and forecasting parts, respectively presented in Sections 4 and 5. For a better comprehension of the forecasting results, we also analyze the predictability of solar radiation and temperature provided by weather forecasts in Sections 3.1 and 3.2. The final section provides a summary and conclusions.

2. Data

In this work a data-driven approach has been chosen, mainly due to the unavailability of detailed data about power plants and local weather measurements. The effectiveness of a data-driven approach, as the name suggests, strongly relies in the appropriateness and quality of input/output data. Input data are here two meteorological variables: solar radiation and air temperature, while the output variable is the electricity production. Solar radiation is converted into electricity by photovoltaic modules and for this reason the choice of surface incoming solar radiation as model input is obvious. Air temperature is also a relevant variable: solar panels efficiency depends by module temperature. Depending by the technology used, above a certain threshold (generally about 25 °C) the panel efficiency begins to drop. For an improved modelling of the module temperature the cooling effect of the wind also should be taken into account (as described in Schwingshackl et al. [18]) and its inclusion will be object of future work.

2.1. Meteorological data

Solar radiation data used in this paper are based on the algorithm described by Mueller et al. [13] and obtained from the Satellite Application Facility on Climate Monitoring (CM-SAF) [17], part of EUMETSAT's SAF Network. Considered variable is the surface incoming shortwave (SIS) radiation on the Meteosat (MSG) full disk. In Fig. 1a is visible the average daily solar radiation and its coefficient of variation (Fig. 1b), i.e. the ratio between standard deviation and average.

For air temperature, we instead consider the E-OBS gridded dataset [8], a land-only high-resolution temperature dataset obtained interpolating on a 0.25° regular grid the available meteorological stations (4200 stations at the latest release made available in October 2013).

Weather forecast of solar radiation and temperature data are provided by the ECMWF Integrated Forecasting System (IFS) which runs twice per day with a resolution of 16 km.

In Table 1 are summarized all the data sources used in this paper.

2.2. Electricity production data

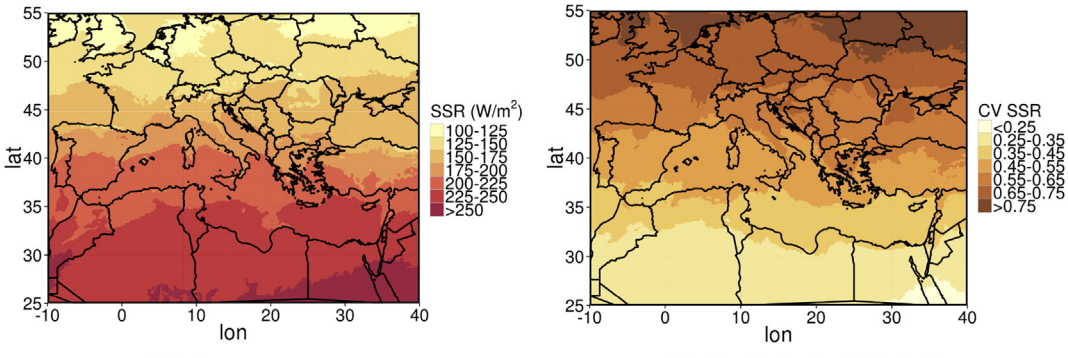
In this work we consider 65 different PV power plants located in different Italian regions. We divide the plants in two groups: North and South. In the first group (North) we have all the PV plants above the 44° 50′ latitude, 34 PV plants with a total of 127 MW of installed capacity. Remaining plants are in the other group (South), 31 PV plants with a total of 288 MW.

For each plant we have a time-series of daily power production of variable length, between 18 and 24 months (550–731 daily samples).

3. Daily predictability of meteorological data

In this section we analyze the capability of the ECMWF numerical weather prediction (NWP) model to forecasts the two main predictors for solar power production: solar radiation and air temperature. Both the meteorological variables are provided by the ECMWF global forecast model, which data is available on 0.25° grid up to ten days in advance with a time step of 3 h.

An assessment on the forecasting skills of ECMWF model can be found in Richardson et al. [15]. Other studies on the use of solar radiation forecasts can be found in Lorenz et al. [10] and Mathiesen & Kleissl [12].



(a) Average (b) Coefficient of Variation

Fig. 1. Surface Solar Radiation statistics for the years 2011–2012 from CM-SAF satellite observations. (a) Annual average of surface solar radiation (b) Coefficient of variation defined as the ratio between standard deviation and average. It measures the variability of the solar radiation, we can observe as the Northern Europe shows a higher variability (generally a CV > 0.55 above the 45° of latitude) and lower average solar radiation than the Southern part of the continent.

Table 1
Summary of weather data sets used in this work.

	Observed	Forecast
2-m temperature	E-OBS (~25 km)	ECMWF IFS (~16 km)
Downward solar radiation	CM-SAF (~15 km)	ECMWF IFS (~16 km)

3.1. Solar radiation

ECMWF operational deterministic forecasts are issued every day and it provides hourly estimation of several variables up to ten days. We use the surface solar radiation downwards variable, i.e. the incident shortwave radiation accumulated over the day.

We compare the forecasts with the CM-SAF satellite data for the years 2011–2012. The two data sets have different resolutions: CM-SAF data are on a sinusoidal grid of 15 × 15 km while the weather forecasts are on a Gaussian reduced grid of about 16 km of resolution. To make them comparable, we applied a bilinear interpolation on the satellite data to match the NWP resolution.

In Fig. 2 we observe the spatial correlation between forecasts and satellite data on the entire domain as a function of the lead time of forecast. We define spatial correlation the Pearson correlation coefficient computed on the spatial dimensions (instead of time as it is usually done) and then averaged on all the time steps. Given two variables $A_{x,y,t}$ and $B_{x,y,t}$ where t is the time and x, y the coordinates on a grid of size $M \times N$ the spatial correlation is defined as:

$$\rho^s = \frac{1}{T}\sum_{k=1}^{k=T} \frac{\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(A_{i,j,k}-\overline{A}_k\right)\left(B_{i,j,k}-\overline{B}_k\right)}{\sqrt{\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(A_{i,j,k}-\overline{A}_k\right)^2}\sqrt{\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(B_{i,j,k}-\overline{B}_k\right)^2}} \tag{1}$$

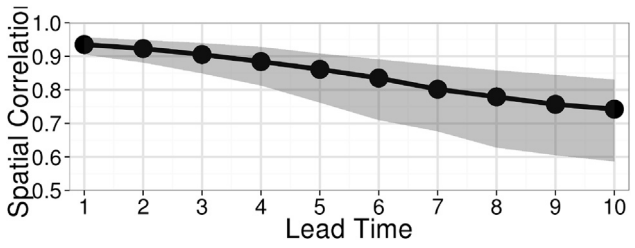
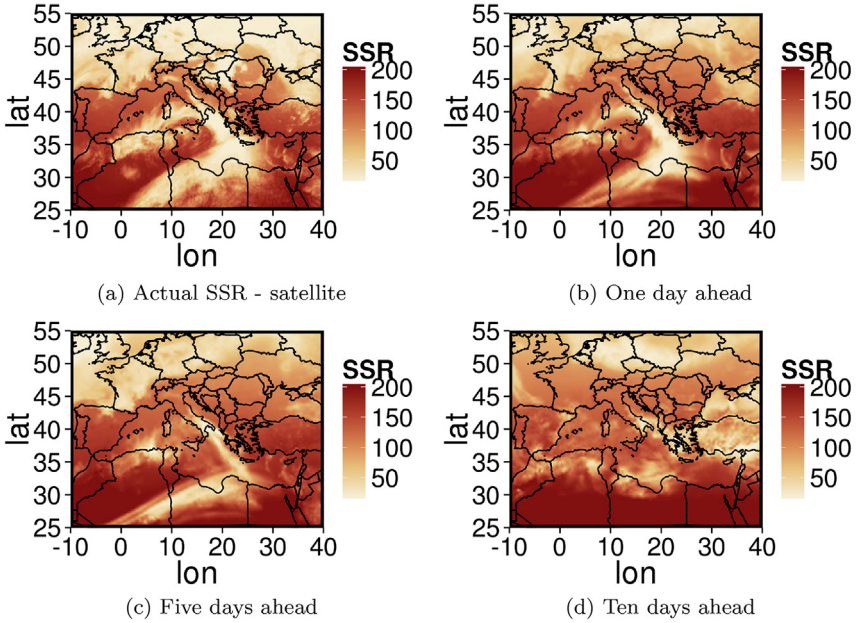


Fig. 2. Average spatial correlation for the period 2011–2012 on the entire domain between operational forecasts and satellite measurements of solar radiation. Shaded area represents the interquartile range (IQR) for each lead time. We observe an average decrease of correlation of 2.5% and an increment of IQR of 20% for each lead time.

The two sample means $\overline{A}_k$ and $\overline{B}_k$ are equivalent to the average of all the points (i.e. all the pairs x, y) for the time step k.

An example of the forecast data is presented in Fig. 3 where we show predicted solar radiation of a specific day for three different lead times: one, five and ten days. Correlating the CM-SAF observations with the three above mentioned forecasts we obtain respectively a ρ^s of 0.93 (Fig. 3b), 0.90 (Fig. 3c) and 0.67 (Fig. 3d).

Moreover, solar radiation exhibits a clear seasonal cycle and for this reason absolute error measures (e.g. RMSE) might not be sufficient to describe the performance of the models used in predicting it. For example, a RMSE of 20 W/m^2 can be a small fraction of the



(a) Actual SSR - satellite (b) One day ahead

(c) Five days ahead (d) Ten days ahead

Fig. 3. Example for a specific day (2/2/2011) of solar radiation forecasts provided by ECMWF operational forecasts with one, five and ten days of lead time. The spatial correlations of the shown forecasts with the observations are respectively 0.93, 0.90 and 0.67.

total incoming solar radiation in Summer but a relevant portion in Winter. For this reason, for a complete and meaningful description we use two different error measures: an absolute and a percentage one.

As absolute error measure we select the RMSE (Root Mean Square Error), defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{t=n} (\hat{y}_t - y_t)^2}{n}} \tag{2}$$

where y_t is the observed value and $\hat{y}_t$ the estimation at time t.

As a percentage error measure, we choose the Median Absolute Percentage Error (MdAPE) defined as:

$$\mathrm{MdAPE} = \mathrm{median}(|100(\hat{y}_t - y_t)/y_t|) \tag{3}$$

We preferred this error measure over the more common MAPE (Mean Absolute Percentage Error) because the former is less sensitive to outliers (see Armstrong & Collopy [1] for an interesting discussion on error measures).

Fig. 4 illustrates the MdAPE and the RMSE of the predicted solar radiation with respect to the latitude for three lead times (1, 5, 10 days, the other lead times have been omitted for sake of clarity). Looking at the RMSE (bottom part of the figure), the difference among the three lead times is more evident, with the prediction with one day of lead time having an error nearly constant at all the latitudes. Obviously, given that the average solar radiation is lower at higher latitudes (see Fig. 1a), the percentage error shows a steeper trend correspondingly. However, it is evident how the prediction error is related to the lead time, with one day the average MdAPE on the entire domain (30–50° latitude) is 8.25%, with five days is 11.59% and at ten days is 17.04%. Respectively, the average RMSE is instead 29.19 W/m^2, 43.62 W/m^2 and 58.06 W/m^2.

The decrease of the forecast performance at high latitudes is due to the higher weather variability, as can be also depicted in Fig. 1b.

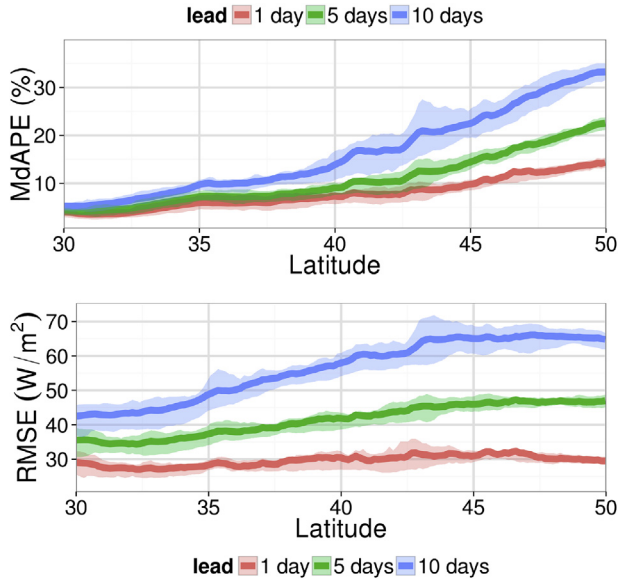


Fig. 4. Error (MdAPE upper panel, RMSE lower panel) on solar radiation forecast versus latitude over Europe with selected lead times (one, five, and ten days). Shaded area represents the interquartile range (IQR). The range 30°–35° is related to the North Africa and East Mediterranean where the solar radiation variability is low, in this case in fact the errors for the three lead times are close to each other. Instead the range 40° and 45° includes the majority of the European mountain areas (Alps, Pyrenees, Carpathians, Balkans), in fact we observe an large forecast error variability (i.e. high IQR).

According to the North/South classification proposed in Section 2.2, Fig. 5 shows the density plot of solar radiation provided by CM-SAF and by the forecast at one, five and ten days of lead time. Looking at the density plot for the North Italy (Fig. 5a), we can quickly see the difference among the three lead times in describing the two peaks, especially for the minor one. Observing the density comparison for the South Italy (Fig. 5b) we instead see how the three lead times show a similar distribution. It can be also seen that for the South Italy the forecasts tend to underestimate the highest peak.

3.2. Air temperature

As for the downwards solar radiation, we analyze the predictability of air temperature provided by ECWMF deterministic forecasts by comparing it with the observations. As stated in Section 2.1, we used as observation the E-OBS dataset for the years 2011–2012. In order to have both the data sets with the same spatial grid, we applied the same interpolation procedure described in Section 4.

Fig. 6 shows descriptive statistics of observed temperature over Italy. The coefficient of variation (Fig. 6b) clearly follows Italian orography, with the higher variability of temperature mostly in the mountain areas. The density plot of observed and predicted temperature (Fig. 7) shows a higher correspondence of forecasts with respect to the similar plot for solar radiation in Fig. 5.

4. Modelling PV production using satellite data

To perform a forecast of the solar power production we first need to find an accurate relationship between daily meteorological variables (here solar radiation and temperature) and power

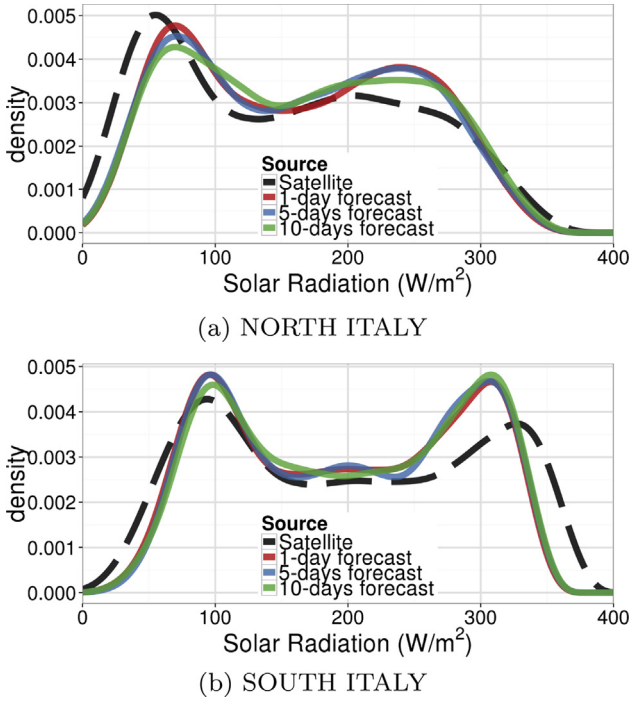
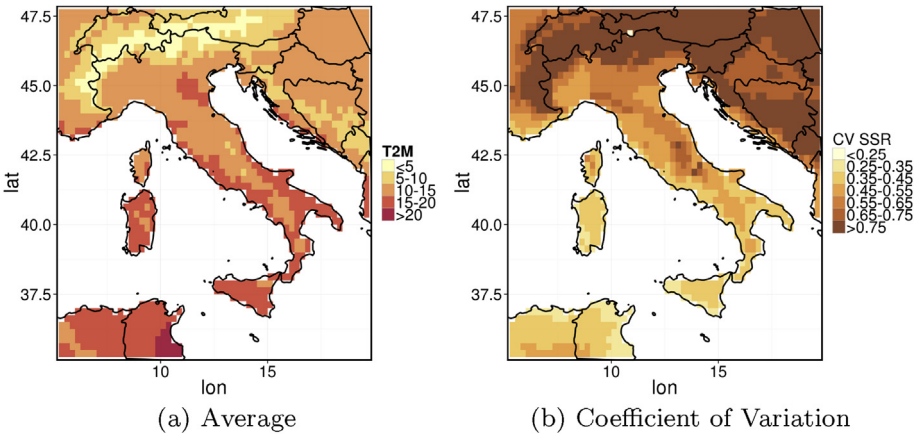


Fig. 5. Comparison of Gaussian kernel density estimation of the observed solar radiation with the predictions at three lead times (one, five and ten days). We can see that the weather forecasts tend to overestimate the "winter" (left one) peak in the North of Italy and to underestimate the "summer" peak (right one) in the South part. A cosine kernel has been used and the bandwidth has been selected using the Silverman rule of thumb [20] adjusted with a factor 1.5.



(a) Average (b) Coefficient of Variation

Fig. 6. Air Temperature statistics for the years 2011–2012 from E-OBS dataset. Both the statistics highlight clearly the Italian coastal areas (higher average temperature and lower variability) and the mountain areas (Alps and Apennines with lower temperature and higher variability).

production. We need to find a set of functions f_i (one for each PV plant) with the following form:

$$\hat{y} = f_i(\mathrm{SSR}, \mathrm{T}) \tag{4}$$

with $\hat{y}$ the predicted power output, SSR and T respectively the surface solar radiation and the air temperature available for the i-th PV plant. These functions aims to model the relationship between the meteorological variables and the electricity produced, trying to minimize the error between observed and estimated values. A black-box approach will focus at the same time on the minimization of the modelling error and on the maximization of the generalization, i.e. the capability of giving consistent outputs with new observed inputs. Given the absence of on-site measurements, here we consider as inputs the bilinear interpolation among the four nearest grid points of solar radiation and temperature data.

Although the photovoltaic process is non-linear, it is a good practice to start with the simplest model for the f function, a linear regression model with the following form:

$$\hat{y} = a_1 \mathrm{SSR} + a_2 \mathrm{T} + a_3 \tag{5}$$

Minimizing the error through Ordinary Least Squares, we obtain an average MdAPE of 12.4% on cross-validation on all the PV plants. A k-fold (with $k = 10$) cross-validation procedure here is used: as first step we divide the available dataset in k subsamples of equal size, and then for k times the chosen model is calibrated using $k - 1$ subsets and then tested on the remaining one. At the end of the k steps, the cross-validation error is given as the average of all the k obtained errors.

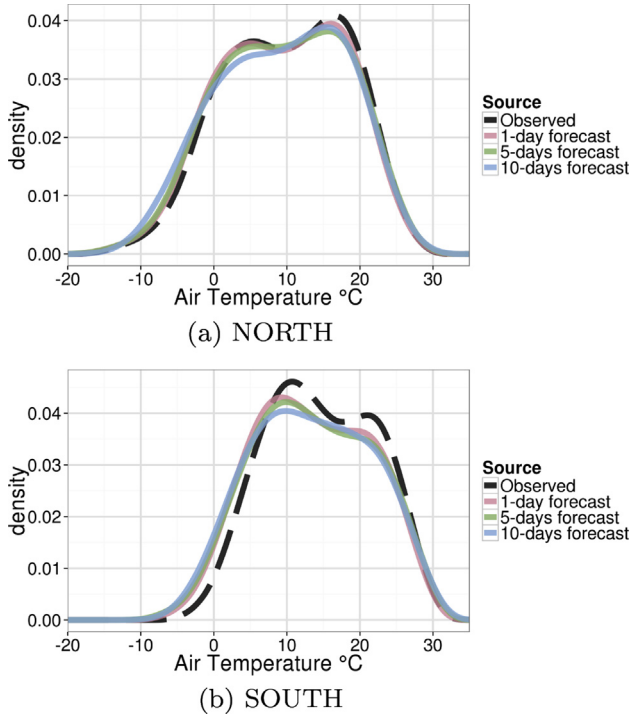
Here the average parameters with the associated standard deviation of the 65 linear models: $a_1 = 0.13 \pm 0.28$, $a_2 = -0.22 \pm 0.40$, $a_3 = 5.12 \pm 9.50$.

In order to take into account the non-linearity of the PV physical processes, we use a Support Vector Machine (SVM), a well-established non-linear approach.

SVMs were developed by Cortes & Vapnik [4,21] for binary classification and then extended to regression problems (Support Vector Regression). The idea behind the support vector-based methods is to use a non-linear mapping Φ (kernel function) to project the data into a higher dimensional space where solving the classification/regression task is easier than in the original space.

In our case, we use a Support Vector Regression method called ε-SVR [5], which tries to find a function $f(x) = \langle w, \Phi(x) \rangle + b$ that has at most ε deviation from the target values. The input vector x is mapped with a non-linear function into a higher dimensional space where the regression is performed. The kernel function here used is a Gaussian kernel $K(\mathbf{u}, \mathbf{v}) = \exp(-\gamma \left|\left| \mathbf{u} - \mathbf{v} \right|\right|^2)$. There are several possible kernel functions (the functions need to satisfy the Mercer's theorem, see the book by Haykin [7] for an in-depth description of the SVM theory) but their optimal choice is generally problem-dependent. We selected the Gaussian kernel among other common typologies (e.g. linear, sigmoid) after some preliminary tests.

A ε-SVR model has three parameters: the regularization parameter C, the ε value, and the width of the kernel γ. The parameter C can be considered the tradeoff between the model complexity and the empirical risk (i.e. the average loss of the estimator): a large value of C implies that the model designer has high



(a) NORTH

(b) SOUTH

Fig. 7. Comparison of kernel density estimation of the observed temperature with the predictions at three lead times (one, five and ten days). A cosine kernel has been used and the bandwidth has been selected using the Silverman rule of thumb [20] adjusted with a factor 1.5.

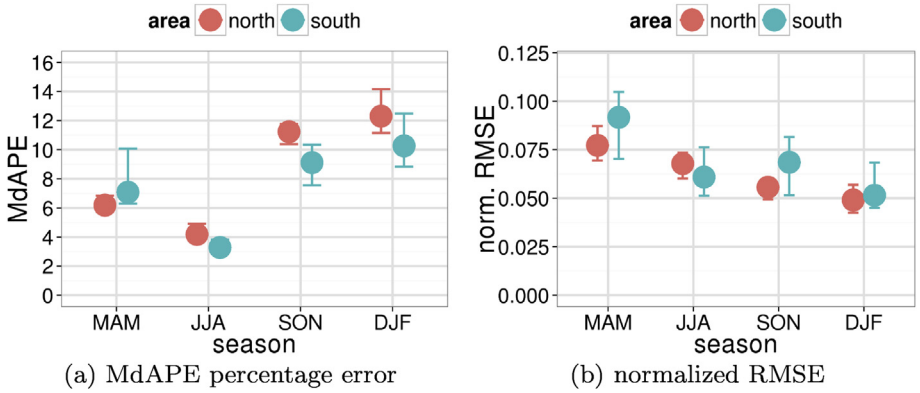


Fig. 8. Cross-validation modelling errors for SVM using observed meteorological variables (satellite solar radiation and E-OBS temperature). Error bars represent the interquartile range (IQR). Looking at the percentage error, the proposed model is able to model the power production better in the South Italy than in the North, due to the lower weather variability, except during Spring. The normalized RMSE is instead influenced by the larger errors that can happen in the South of Italy because of the larger average solar radiation (see Fig. 1a).

confidence in the quality of the training data, on the other hand a small value is needed when the available data is noisy (e.g. to avoid overfitting). The ε parameter represents instead the "width" of the ε-insensitive zone, in the ε-SVR approach the loss function is different from zero only when the error is larger than ε, in other words this parameter denotes how much error you are willing to allow per each training data sample. Finally, the γ parameter is the width of the Gaussian kernel as described before.

For each PV plant we choose the optimal parameters of the SVR model applying a grid search among 75 combinations of $C \in [10^{-2}, 10^{2}]$, $\varepsilon \in [10^{-2}, 1])$ and $\gamma \in [2^{-2}, 2^{2}]$. After the parameters' selection, as for the linear models, we compute the cross-validation error. We obtain an average MdAPE of 7.6%, about the 40% lower than in the linear case. This improvement was already expected, given the highest modelling power due to the inherent non-linearity of SVR with respect to linear regression.

Aggregating the PV plants by North and South (see Sec. 2.2) we obtain the modelling errors (both the MdAPE and the RMSE) divided by season (see Fig. 8). We observe how the percentage error is lowest during summer for entire Italy, and, except for spring, we get for South Italy lower errors at all the seasons. On the other hand, on the right (Fig. 8b) we see that the normalized RMSE (i.e. RMSE divided by the maximum PV plant power output) is lowest during winter, which is the period with the lowest incoming radiation and thus PV production during the entire year.

5. Short-term forecast of solar power production

In this section we assess the forecasting skill using the SVM models created in the previous section and driven by NWP variables instead of observations.

As summarized in Table 1 and explained in Section 2.1, we use the meteorological data coming from the ECMWF operational forecasts. Following the same approach of the modelling part, for each PV plant we apply bilinear interpolation of the nearest four grid points as input variables for each PV plant.

For each day of lead time Fig. 9a shows the MdAPE error of the power production. Similarly, Fig. 9b depicts the correlation between predicted and observed output. The minimum error is with one day of lead time (10–12%) and it grows steadily up to 15–20% with ten days of lead time. In all the cases the prediction of the PV plants in South of Italy is more accurate than in the North. We observe that the interquartile range also increases with the lead time, highlighting the higher uncertainty due to the weather forecasts at increasing lead times. Looking at the correlation, one day of lead time for both the cases is in the range 0.7–0.8 while at ten days it drastically decreases below 0.1.

The error analysis can be improved grouping the errors by season, as in Fig. 10 where we display both the MdAPE and the normalized RMSE (i.e. the RMSE divided by the maximum PV power output). In this figure is evident the different magnitude of

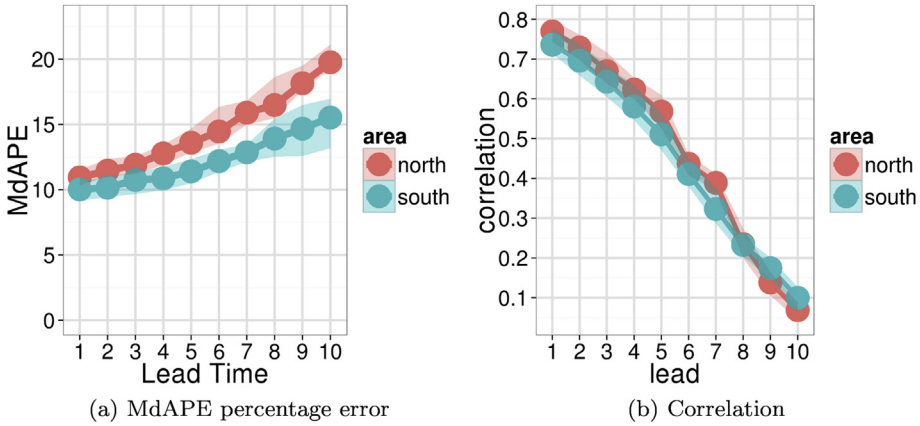


Fig. 9. PV power production forecast for SVM using predicted meteorological data. Shaded area represents interquartile range (IQR). The percentage error and the correlation between observed and predicted power production are shown for each lead time of the weather forecasts.

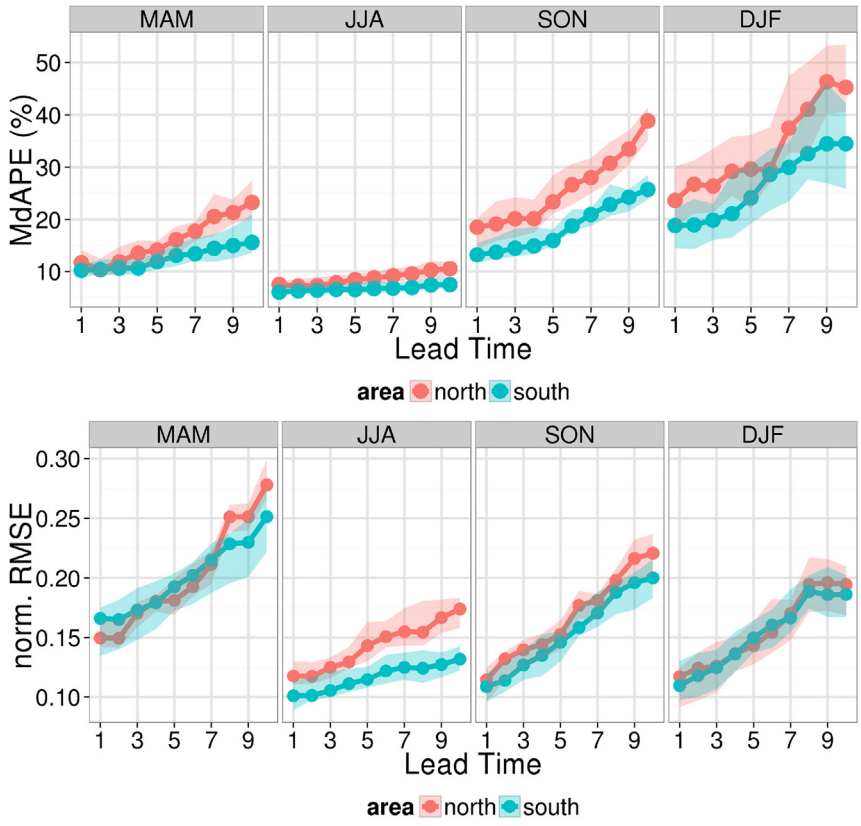


Fig. 10. Prediction error (median percentage absolute error and normalized RMSE) for SVM using forecasted weather data by season. Shaded area represents interquartile range (IQR). The evident error differences among the seasons (especially between summer and the other season) is due to the weather variability and then to the capability of the weather forecast models to predict effectively the meteorological predictors used as inputs for the SVM.

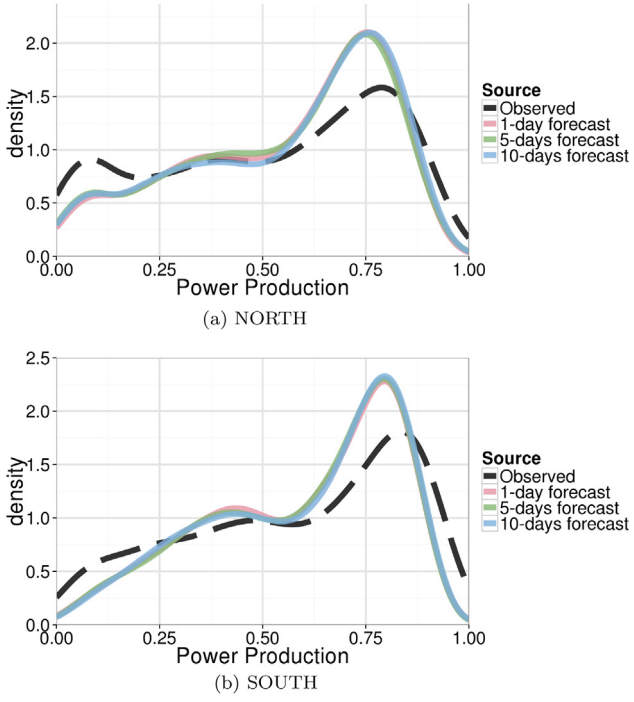


Fig. 11. Comparison of SVM estimation of the normalized solar power production with the predictions at three lead times (one, five and ten days). The SVM model tends to underestimate the power production in both the geographical domains and basically the power distributions of the three lead times are hardly distinguishable. A cosine kernel has been used and the bandwidth has been selected using the Silverman rule of thumb [20] adjusted with a factor 1.5.

errors between summer, where it is common to have clear sky in most of the country, and winter, when the MdAPE is about the 50%. Observing the normalized RMSE, the difference between North and South is less pronounced except for summer, where the two error curves are well distinguishable. It is worth remembering that the average incoming solar radiation (see Fig. 1a) is different between North (125–175 W/m^2) and South (175–225 W/m^2) of Italy, this means that the same absolute error can lead to different percentage errors as it has been discussed before and it is shown in Fig. 10.

Finally, Fig. 11 highlights how in both cases, North and South, the prediction densities of the three lead times are similar, indicating a general tendency to underestimate high yields.

6. Conclusions

In this paper, we have shown an assessment about the short-term predictability of photovoltaic daily power production over Italy without the use of meteorological on-site PV plant measurements. We have performed a detailed analysis of the accuracy of solar radiation and temperature predicted by NWP models to evaluate the associated uncertainty.

Through Support Vector Machine methodology, we have analysed the modelling error of power production using solar radiation from satellite as well as temperature observations from weather stations. Then, with NWP forecasts as inputs on the same models, we have compared the prediction error for lead times between one and ten days.

The results can be outlined as follows:

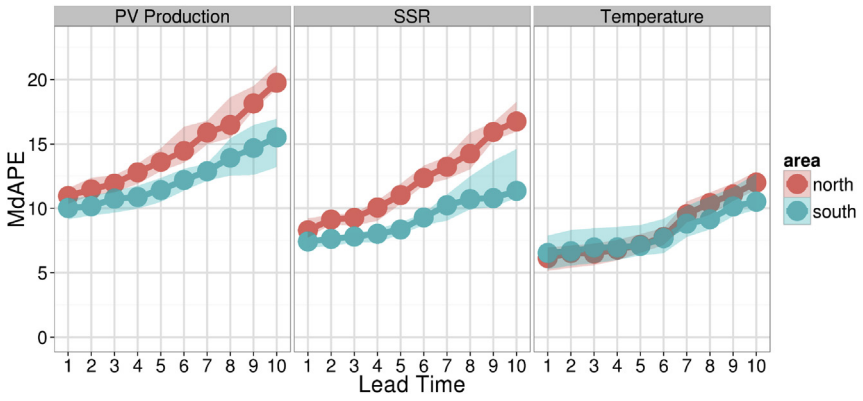


Fig. 12. Prediction errors (median percentage absolute error). Left panel: error for power production with SVM using predicted meteorological data. Center: error between observed solar radiation and NWP prediction. Right: error between observed temperature and NWP prediction. Shaded area represents interquartile range (IQR).

1. Given the absence of meteorological measurements on PV plants, we have used remote sensing and ground-based data obtaining an average cross-validation percentage error (MdAPE) of 12.4% using a linear model and 7.6% a SVM on the interpolated PV plant location.
2. Solar power production obtained by SVM modelling on Italy was found to be more accurate during summer than in the rest of the year: the percentage error is below the 5% when we use observed meteorological data as predictors and below the 12% when we use forecasted predictors on the entire prediction range. The normalized RMSE is below 0.08 and 0.18 respectively.
3. The prediction results for the PV plants in the South Italy were comfortably better than those in the North, mainly due to the lower weather variability in the southern part of the country.

Uncertainty due to the absence of information related to local phenomena (e.g. orography, shading effects, etc.) becomes certainly critical in predicting PV power production, especially for the higher lead times.

We have analysed the nature of our results' uncertainty and it can be seen as the combination of three concurrent sources: (i) Modelling limitations of the SVM methodology (ii) Errors in the observations used to calibrate the models (iii) Weather forecast accuracy (as discussed in Sections 3.1 and 3.2).

The error propagation of the NWP forecasts on the solar power production can be estimated observing the differences between the modelling (Fig. 8) and prediction (Figs. 9 and 10) errors.

Concluding, Fig. 12 summarizes this uncertainty propagation showing the relationship between the PV production error (the same as in Fig. 9a) and the forecast error of the used meteorological predictors (solar radiation and temperature).

These results demonstrate on the one side the potentiality in using black-box approach in spite of the absence of on-site measurements; on the other side, the crucial importance of estimating the magnitude and the nature of uncertainties in forecasting electricity production.

Finally, we believe that an in-deep analysis of the uncertainty is the key factor for a reliable management of renewable and conventional sources in power grids.

Acknowledgements

We thank TERNA for providing photovoltaic data. EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) intermediate products were used by permission of Deutscher Wetterdiens. We also acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu).

References

[1] Armstrong JS, Collopy F. Error measures for generalizing about forecasting methods: empirical comparisons. Int J Forecast 1992;8(1):69–80.
[2] Bellini A, Bifaretti S, Iacovone V, Cornaro C. Simplified model of a photovoltaic module. In: Applied electronics, 2009. AE 2009. IEEE; 2009. p. 47–51.
[3] Bouzerdoum M, Mellit A, Massi Pavan A. A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. Sol Energy 2013;98:226–35.
[4] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.
[5] Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support vector regression machines. Adv Neural Info Process Syst 1997;9:155–61.
[6] GSE. Rapporto statistico 2012 Solare fotovoltaico (Italian). May 2013. http://www.gse.it/it/Statistiche/RapportiStatistici/Pagine/default.aspx.
[7] Haykin SS. Neural networks and learning machines, vol. 3. Pearson Education Upper Saddle River; 2009.
[8] Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M. A european daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J Geophys Res Atmos 2008;113(D20).
[9] Ljung Lennart. Perspectives on system identification. Annu Rev Control 2010;34(1):1–12.
[10] Lorenz E, Hurka J, Heinemann D, Beyer HG. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. Sel Top Appl Earth Observ Remote Sens IEEE J 2009;2(1):2–10.
[11] Massi Pavan A, Mellit A, De Pieri D, Lughi V. A study on the mismatch effect due to the use of different photovoltaic modules classes in large-scale solar parks. Prog Photovolt Res Appl 2014;22(3):332–45.
[12] Mathiesen P, Kleissl J. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Sol Energy 2011;85(5):967–77.
[13] Mueller RW, Matsoukas C, Gratzki A, Behr HD, Hollmann R. The CM-SAF operational scheme for the satellite based retrieval of solar surface irradianceA LUT based eigenvector hybrid approach. Remote Sens Environ 2009;113(5):1012–24.
[14] Pedro HTC, Coimbra CFM. Assessment of forecasting techniques for solar power production with no exogenous inputs. Sol Energy 2012;86(7):2017–28.
[15] Richardson DS, Bidlot J, Ferranti L, Haiden T, Hewson T, Janousek M, et al. Evaluation of ECMWF forecasts, including 2012–2013 upgrades. Technical Memorandum 710, ECMWF. November 2013.
[16] Sandrolini L, Artioli M, Reggiani U. Numerical method for the extraction of photovoltaic module double-diode model parameters through cluster analysis. Appl Energy 2010;87(2):442–51.
[17] Schulz J, Albert P, Behr H-D, Caprion D, Deneke H, Dewitte S, et al. Operational climate monitoring from space: the EUMETSAT satellite application facility on climate monitoring (CM-SAF). Atmos Chem Phys 2009;9(5):1687–709.
[18] Schwingshackl C, Petitta M, Wagner JE, Belluardo G, Moser D, Castelli M, et al. Wind effect on PV module temperature: analysis of different techniques for an accurate estimation. Energy Procedia 2013;40:77–86.
[19] Shi J, Lee W-J, Liu Y, Yang Y, Wang P. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. Indus Appl IEEE Trans 2012;48(3):1064–9.
[20] Silverman BW. Density estimation for statistics and data analysis, vol. 26. CRC Press; 1986.
[21] Vapnik V. The nature of statistical learning theory. Springer; 2000.
[22] Zeng J, Qiao W. Short-term solar power prediction using a support vector machine. Renew Energy 2013;52:118–27.