SOLAR RADIATION FORECAST USING NEURAL NETWORKS FOR THE PREDICTION OF GRID CONNECTED PV PLANTS ENERGY PRODUCTION (DSP PROJECT)

C. Cornaro*#, F. Bucci*, M. Pierro*, F. Del Frate°, S. Peronaci°, A. Taravat°
*Department of Enterprise Engineering, °Department of Civil Engineering and Computer Science Engineering, #CHOSE
University of Rome "Tor Vergata"
Via del Politecnico, 1 00133 Rome, ITALY
e-mail: cornaro@uniroma2.it, fabio.delfrate@disp.uniroma2.it, frabucci@gmail.com, simone.peronaci@hotmail.it, marco.pierro@gmail.com, art23130@gmail.com

ABSTRACT: The work presented in this paper is part of a project aimed to develop a prototype device (DSP) able to forecast with a day in advance the energy produced by PV plants. The energy forecast is required by the National Authority for the electricity in order to control the high instabilities of the electric grid induced by unpredictable energy sources such as photovoltaic. In the paper several models to forecast the hourly solar irradiance with a day in advance using Artificial Neural Network (ANN) techniques are described. Statistical (ST) models that use only local measured data and Hybrid model (HY) that also use Numerical Weather Prediction (NWP) data are tested for the University of Rome "Tor Vergata" site. The performance of ST, NWP and HY models, together with the Persistence model (PM), are compared. The ST models and the NWP model exhibit similar results improving the performance of the PM of around 20%. Nevertheless different sources of forecast errors between ST and NWP models are identified. The Hybrid models give the better performance, improving the forecast of approximately 39% with respect to the Persistence model.
Keywords: solar radiation, grid stability, photovoltaic, forecast, neural networks

1 INTRODUCTION

With the growing of PV installed power, the forecast of the solar energy production becomes more and more important. In particular the 24/72 hours horizon forecast is essential for transmission scheduling and day ahead energy market. In Italy the PV energy production during the 2012 reached the 16.7 GWh providing an average of 7% of yearly electrical consumption with monthly pick of 9%. Thus Italy, as many other European countries, is starting to deal with some criticisms of the integration of the PV plants into the National Grid. Recently the national authority for the electricity delivered actions with the objective to regulate imbalances of the grid induced by electricity inputs from the unpredictable energy sources, such as solar energy. One action will foresee penalties for producers that will not correctly forecast the electricity produced by the PV plants with a day in advance. This work is part of a project aimed to develop a prototype device (DSP) able to forecast with a day in advance the energy produced by PV plants.

The techniques to forecast the solar radiation or PV productions on the 24/72 hours horizon can be divided in three mains groups [1]:

1) Numerical Weather Prediction models (NWP)
2) Statistical models (ST)
3) Hybrid models (HY)

The Numerical Weather Prediction models [2] are essentially based on the numerical integration of coupled differential equations that describe the dynamic of the atmosphere and radiation transport mechanisms.

The main advantage of these forecasting methods is that they are based on deterministic physical models. On the other hand, the main problem, in addition to the non-linearity of the used equations, is the spatial resolution of the integration grid (from 100 km^2 to few km^2) that is too wide with respect to the PV plants size. Inside the grid cell the cloud cover and aerosol are homogeneously fixed at their average values thus great errors could be induced both in the amount and in the time of the forecasted irradiance on the PV site. Besides many NWP models have a temporal output interval greater than one hour while, as in this case, the hourly irradiance forecast is required. Just to cite an example, Perez et al. [3] presented an extensive validation of short and medium term solar radiation forecast for various sites in the US.

The Statistical models are based on methods to reconstruct the relations between the hourly irradiance and past meteorological parameters (cloud ratio, air temperature, relative humidity, pressure etc.) or past irradiance observations. The most used models for the one day horizon irradiance forecast are Artificial Neural Networks (ANN). With this method the forecast could be achieved by fast simple algorithms that use only local meteorological measurements [4, 5, 6] and statistical feature parameters [7]. Thus spatial and temporal resolution problems are overcome. On the other hand these methods are not able to provide a good forecast in unstable weather conditions since in these cases the correlation between the irradiance and the precursors rapidly falls down.

The Hybrid models combine both NWP and ST models [8, 9]. The first one is used for the forecast while the second to correct the site effects through local measurements. The ST models are essentially used to down scale the irradiance forecast.

It has to be noted that some authors in the literature use ANN and Hybrid models to directly forecast the PV plant production [9, 10, 11].

In this article all these methods were tested to forecast the solar irradiance at the University of Rome "Tor Vergata" site with a horizon of 24 hours. The NWP data were provided by the European Centre for Medium-Range Weather Forecast (ECMWF) [12]. The results obtained by seven ST models based on different kind of Neural Networks Algorithms that use only in situ measurements, are reported. Finally the forecast obtained by four different Hybrid models that use NWP data and local measurements as input of ANN are analyzed. It appears that, for the present site, the local approach with ST models could provide a comparable hourly irradiance forecast than NWP model. On the contrary the cumulative daily irradiation is better predicted with the NWP data. It has been confirmed that the deterministic and statistic approaches have different sources of errors in their forecasting outputs. Each method could reduce the error of the other and vice versa. Thus all the Hybrid models provide similar and better results.

It should be remarked that the majority of the works in the forecasting literature provide the predictions of the global horizontal irradiance while to forecast the PV energy production the irradiance on the plane of arrays (POA) is required. Thus the transposition factor from the horizontal plane to POA could introduce an additional error in the forecast. For this reason, with the exception of NWP data, all the presented models directly provide the forecast of the POA irradiance.

2 DATA DESCRIPTION

2.1 Local Experimental Data and pre-processing

The local experimental data used as input and to train and test the models come from the ESTER outdoor Laboratory - University of Rome “Tor Vergata” [13]. The global irradiance used for the forecast is the one measured on the plane of PV modules exposed at ESTER lab from January 2009 to the end of October 2010. In particular the global and diffuse horizontal irradiance, the global POA irradiance, air temperature and the energy produced by a c-Si module (Kyocera KC 125) were measured each minute during the considered period. The data were filtered removing the not physically consistent measurements due to monitoring problems. For each day data reconstruction by linear interpolation was applied if no more than 60 consecutive missing samples were encountered, otherwise the whole day was removed from the data set. After this operation the hourly and daily data were calculated.

The data reconstruction was introduced to overcome the Data Monitoring System problems that bring to underestimation of the produced energy.

2.2 NWP data

The NWP data are the output of the ERA-INTERIM model that is the latest global atmospheric reanalysis software produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) [12]. In particular they consist of a 24 hours horizon forecast of the horizontal global solar irradiation. The spatial resolution is 3 km^2. The temporal output interval is 3 hours thus, for each day, the data contain 8 forecast of the cumulative integral of horizontal global irradiance starting from midnight UTC. According to the local measurements the considered period is between the 1 January 2009 and 31 October 2010.

3 MODELS INPUT PARAMETERS

Since the final aim of the work is to forecast the PV energy production, all the inputs of the reported forecast models should be variables that are commonly measured by a medium-large size PV plant monitoring systems.

For each day, the hourly solar irradiance features ($Gpoa_1$... $Gpoa_{24}$), to be forecasted, have been characterized by the following five parameters:

- OD= Ordinal day
- Hh and Hpoa= daily global irradiation on horizontal and POA [kWh/m^2 day];
- CR= Cloud Ratio [-];
- NMHV= Normalized Maximum Hourly Variation of the solar irradiance;
- NADV=Normalized Absolute Daily Variation of the solar radiation between the day (t) and the day (t-1).

The OD takes into account for the yearly variations of sunrise and sunset hours.

The Hh and Hpoa are the integral of hourly solar irradiance and depict both the yearly solar energy variation and meteorological features of the day. Besides the Hpoa takes into account orientation and tilt of the PV plane.

The Cloud Ratio is defined as the ratio between the horizontal diffuse (Hsh) and global (Hh) daily irradiation:

$$\mathrm{CR} = \frac{\mathrm{Hsh}}{\mathrm{Hh}}\ [-]$$

This parameter is strictly related to the stochastic meteorological conditions. For CR<0.4 the day could be considered clear while CR>0.4 indicates overcast days. It requires specific measurements that are not always available at the PV plant, but for practical applications it could be replaced by the clearness index Kt [6].

The Normalized Maximum Hourly Variation of the solar irradiance is defined as:

$$NMHV = \frac{max_{h}\left\{\sqrt{\frac{\sum_{min=1}^{60}(G_h^{min}-fit(G_h^{min}))^2}{60}}\right\}}{< G_h^{min} >_{\mathrm{day}}}\ [-]$$

Where:

G_h^{min} = irradiance at minute min of the hour h $[\frac{W}{m^2}]$;

$fit(G_h^{min})$ = linear fit of G_{min}^{h}($hourly\ trend$) $[\frac{W}{m^2}]$;

$< G_h^{min} >_{\mathrm{day}}$= daily average irradiance $[\frac{W}{m^2}]$.

This statistical parameter is the maximum fluctuation of the measured irradiance around the hourly trend with respect to the daily average irradiance and it is used to describe the daily variability of the irradiance. In clear sky days it could be near to zero while in overcast days it could reach the value of seven (fluctuations are 7 times the mean daily irradiance). Variable days show NMHV>0.4.

The Normalized Absolute Daily Variation of the solar irradiation is defined as:

$$\mathrm{NADV} = |NDV| = \left|\frac{(H(t)-H(t-1))}{(H(t)+H(t-1))/2}\right|\ [-]$$

Where:

$H(t)\ and\ H(t-1)$ irradiations at days: $(t)$and $(t-1)$

It is used as an indicator of the weather persistence of one day with respect to the other. NADV<0.4 has been considered as indicator of stable weather condition with respect to the past day. Figure 1 shows the Probability Density Function (PDF) of NDV of the horizontal irradiation Hh of five years measurements (2008-2012) at the ESTER location. It can be observed that small variations of weather conditions between two consecutive days (low NADV) are much more probable than fast weather changing (high NADV). In general the mean value and the variance of PDF changes from year to year and from site to site but the shape of the PDF remains the same. Thus the weather tends to be persistent.

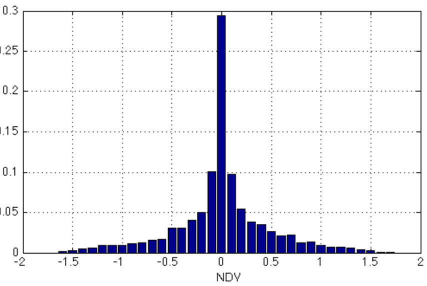

Figure 1: Probability Density Function of NDV

Also for this reason, to evaluate a forecast model performance the persistence model is commonly used (see section 3.2.1). Moreover fast weather perturbations (high NADV) could be forecasted only using NWP models since the correlations between the input parameters at lag0 (day t) and the forecasted irradiance at lag1 (day t+1) are very low. Nevertheless the ST models could show good performance for the majority of the days that have relatively persistence conditions (low NADV).

Since all the five variables measured at lag0 are correlated with the daily irradiation at lag1 (see Table I) they were used as ST models inputs. Another meteorological parameter that is usually measured by all the PV plant monitoring systems is the ambient temperature (Ta). Since it is strongly correlated to the daily irradiation (see Table I), it has also was used as input variable.

Table I: Pearson correlations of the inputs variables at lag0 and Hh at lag1 in the year 2009.

	Hpoa (lag0)	Ta (lag0)	CR (lag0)	NMHV (lag0)	NADV (lag0)
Hpoa (lag1)	49%	51%	40%	19%	30%

4. FORECAST MODELS DESCRIPTION

In this section the general features of all the used forecast models are discussed. In Table II all the technical specifications of each model are reported.

4.1 Persistence model

Since the weather tends to be persistent it is possible to define a trivial model:

$$[Gpoa_1(t+1), \dots\ Gpoa_{24}(t+1)] = \\ = [Gpoa_1(t), \dots\ Gpoa_{24}(t)]$$

A forecasting model should have better performance than the Persistence Model (PM). Moreover the performance improvement of a model with respect the PM is a parameter much less yearly and site dependent.

4.2 NWP model

The NWP data contains the one day ahead forecast of the cumulative irradiation each 3 hours, thus to obtain the hourly irradiance, a linear interpolation over the average 3 hours irradiance has been performed:

$$[Gh_1(t+1), \dots\ Gh_{24}(t+1)] = \\ = Linear\ Interpolation\ of\ ECMWF\ ERA\ INTERIM\ Outputs$$

4.3 Statistical and Hybrid models

To develop the ST and Hybrid models the MultiLayer Perceptron Neural Network (MLPNN) has been used.

The MLPNN architecture, reported in Figure 2, uses meteorological parameters to forecast the one day ahead hourly irradiance:

$$[Gpoa_1(t+1), \dots\ Gpoa_{24}(t+1)] = \\ = f(\text{meteorological parameters})$$

The inputs meteorological parameters could come only from past local measurements (in the case of ST models) or also from NWP forecast data (in the case of Hybrid models).

In this work the performance of 8 different MLPNN used to develop ST and HY models, are reported. These ANNs were developed with two software: MatLab and Java NNS to verify the reliability of the techniques. Similar performance have been found and the results of the most performing models are reported.

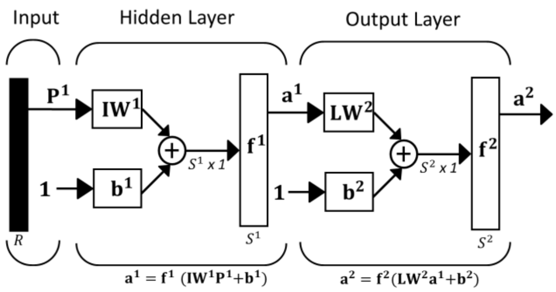

Figure 2: MLP architecture

To define the best numbers of neurons and optimize the Network the following procedure was used:

- 284 days selected from the 01/11/2009-31/10/2010 with the condition that three consecutive days data exist were used for training and validation. The 80% of these data were randomly sorted for training the MLP and the 20% for validation.
- To select the best number of neurons of the hidden layer (best S^1) for each dimension S^1=[1,5,10,15,20,25,30] the NN was trained almost 20 times. Then the architecture that exhibits the minimum MSE on the validation set was chosen.
- To optimize the model (best IW^1 and LW), the selected NN architecture was trained almost 50 times, then the MLP that presents the minimum MSE on the validation set was selected.
- To test the best model the data of 270 days selected from the 1/01/2009-31/12/2009 were used with the condition that three consecutive days data should exist.

For the NN developed with MatLab tool the training algorithm is the Lervemberg-Marquardt (LMA), while for the one developed with Java NNS the training Algorithm is the Batch Backpropagation (BBP). For each training operations the convergence process is stopped when the MSE on the validation set reaches its minimum.

Table II: technical specifications of the networks models

TYPE	NAME	DESCRIPTION
Persistence	PM	[$Gpoa_1$(t+1)... $Gpoa_{24}$(t+1)]=[$Gpoa_1$(t)... $Gpoa_{24}$(t)]
NWP	NWP	[Gh_1(t+1)... Gh_{24}(t+1)]=Linear Interpolation of ECMWF ERA-INTERIM Output
ST MODELS with Feed Forward Multi Layer Perceptron	1MLP	[$Gpoa_1$(t+1)... $Gpoa_{24}$(t+1)]=f1(OD(t), Hpoa(t), Ta(t),CR(t)) JavaNNS: R=6, S^1=30 f^1 = tansigmoid, S^2=24 f^2 = purelinear, BBP
	2MLP	[$Gpoa_1$(t+1)...$Gpoa_{24}$(t+1)]=f2(OD(t),Hpoa(t),Ta(t),CR(t), NMHV(t),NADV(t)) JavaNNS: R=6, S^1=30 f^1 = tansigmoid, S^2=24 f^2 = purelinear, BBP
	3MLP	4 seasonal Neural Network in parallel: [$Gpoa_1$(t+1)...$Gpoa_{24}$(t+1)]=f3(OD(t),Hpoa(t),Ta(t),CR(t),NMHV(t),NADV(t)) JavaNNS: R=6, S^1=30 f^1 = tansigmoid, S^2=24 f^2 = purelinear, LMA
	4MLP2Net: *MLP4.1*	2 Neural Network in series: [Hpoa(t+1)]=f4.1(OD(t-1), Hpoa(t-1), Ta(t-1),CR(t-1),OD(t), Hpoa(t), Ta(t),CR(t)) MatLab: R=8, S^1=10 f^1 = tansigmoid, S^2=1 f^2 = purelinear, LMA
	MLP4.2	[$Gpoa_1$(t+1)...$Gpoa_{24}$(t+1)]=f4.2(OD(t+1), Hpoa(t+1), Ta(t),CR(t),NHV(t),NDV(t)) MatLab: R=6, S^1=10 f^1 = tansigmoid, S^2=1 f^2 = purelinear, LMA
HYBRID MODELS with Numerical Weather Prediction and Multi Layer Perceptron Feed Forward	8NWPMLP	[$Gpoa_1$(t+1)...$Gpoa_{24}$(t+1)]=f8(OD(t), Hpoa(t), Ta(t),CR(t), OD(t+1), Hh_nwp(t+1)) MatLab: R=6, S^1=10 f^1 = tansigmoid, S^2=15 f^2 = purelinear, LMA
	9NWPMLP	[$Gpoa_1$(t+1)... $Gpoa_{24}$(t+1)]=f9(OD(t+1), [Gh_nwp_1(t+1)... Gh_nwp_{24}(t+1)]) MatLab: R=25, S^1=20 f^1 = tansigmoid, S^2=15 f^2 = purelinear, LMA
	10NWPMLP	[$Gpoa_1$(t+1)... $Gpoa_{24}$(t+1)]=f10(OD(t), Hpoa(t), Ta(t),CR(t),OD(t+1), [Gh_nwp_1(t+1)... Gh_nwp_{24}(t+1)]) MatLab: R=25, S^1=20 f^1 = tansigmoid, S^2=15 f^2 = purelinear, LMA
	11NWPMLP2Net: NWPMLP11.1	2 Neural Network in series: [Hpoa(t+1)]=f41(OD(t-1), Hpoa(t-1), Ta(t-1),CR(t-1),OD(t), Hpoa(t), Ta(t),CR(t), OD(t+1), Hh_nwp(t+1)) MatLab: R=10, S^1=10 f^1 = tansigmoid, S^2=1 f^2 = purelinear, LMA
	MLP4.2	[$Gpoa_1$(t+1)... $Gpoa_{24}$(t+1)]=f42(OD(t+1), Hpoa(t+1)) MatLab: R=6, S^1=10 f^1 = tansigmoid, S^2=1 f^2 = purelinear, LMA

Notes: LMA=Levemberg-Marquardt Algorithm, BR=Bayesian Regulation, BBP=Batch Back Propagation, OD=Ordinal Date (ISO 8601), H=daily irradiance [kWh/m^2 day], CR=cloud ratio, Ta=mean daily temperature, NMHV=Normalized Maximum Hour Variation, NADV=Normalized Absolute Day Variation.

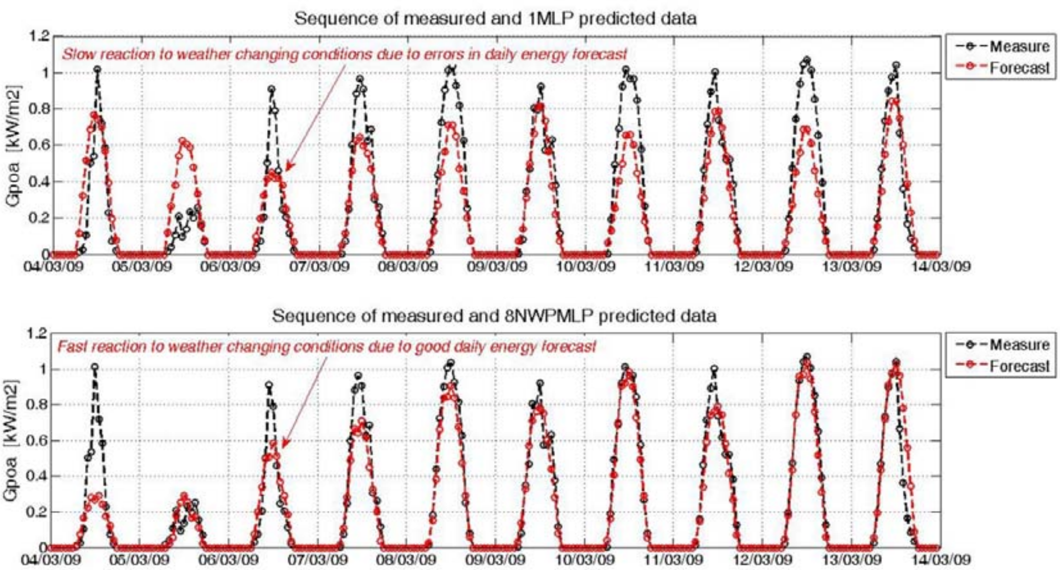

Figure 3: example of sequence of measured and forecasted data

Table III: models performance main results

Forecast variable	NAME	test days	KS	Corr	RMSE [W/m^2]	NRMSE %	I_{RMSE} %	NMAE %	D_{NMAE} %
Gh	PM	270	0	0.81	166	38.4	1	22.7	1
	NWP	270	0.08	0.89	128	29.6	**23**	22.7	**0**
Gpoa	PM	270	0	0.72	236	50.6	1	29.6	1
	ST models with local data								
	1MLP	270	0.13	0.81	188	40.2	**20.6**	29.1	**0.5**
	2MLP	270	0.09	0.79	197	42.3	17.1	28.4	1.2
	3MLP	270	0.05	0.72	231	49.6	2.35	32.9	-3.3
	4MLP2Net	270	0.09	0.81	187	40.1	21.3	27.4	2.2
	Hybrid models with local and NWP data								
	8NWPMLP	270	0.06	0.89	145	31	**38.7**	20.2	**9.4**
	9NWPMLP	270	0.06	0.89	145	31.2	38.4	20.9	8.7
	10NWPMLP	270	0.05	0.89	147	31.5	37.9	20.7	8.9
	11NWPMLP2Net	270	0.07	0.884	149	31.9	37.2	20.8	8.8

5 RESULTS AND DISCUSSION

5.1 Statistical Performance Indicators

The main used statistical performance indicators are the following:

1. Kolmogorov-Smirnov test

   $$\text{KS} = sup_G\left|CDF^m(G) - CDF^f(G)\right|$$

   CDF= cumulative distribution function

2. Pearson Correlation index

   $$\text{Corr} = \frac{\sum_{i=1}^{n}(G_i^m - \overline{G^m})\left(G_i^f - \overline{G^f}\right)}{\sqrt{\sum_{i=1}^{n}(G_i^m - \overline{G^m})^2 \sum_{j=1}^{n}(G_j^f - \overline{G^f})^2}}$$

3. Root Mean Square Error and Normalized RMSE

   $$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(G_i^m - G_i^f)^2}{n}}\ [\text{kW/m}^2]$$

   $$NRMSE = 100\left(\frac{\text{RMSE}}{\overline{G^m}}\right)\ [\%]$$

4. Daily Cumulative Absolute Error

   $$CAE = \sum_{i=1}^{24}\left|G_i^m - G_i^f\right|\ [\text{kWh/m}^2\ \text{day}]$$

5. Mean Absolute Error and Normalized MAE

   $$MAE = \frac{\sum_{i=1}^{n}\left|G_i^m - G_i^f\right|}{n}\ [\text{kW/m}^2]$$

   $$\text{NMAE}=100\left(\frac{\sum_{i=1}^{n}\left|G_i^m - G_i^f\right|}{n\,\overline{G^m}}\right)[\%]$$

6. Improvement

   $$I_{RMSE}=100\left(\frac{RMSE(PM)-RMSE(model)}{RMSE(PM)}\right)[\%]$$

   $$D_{NMAE}=NMAE(PM) - NMAE(model)\ [\%]$$

Where

G_i^m= measured hourly irradiance [kW/m^2]

G_i^f = forecast hourly irradiance [kW/m^2]

All the statistical indexes are calculated considering only the day time hourly irradiance (G > 20 W/m^2), indeed including the nighttime values all the average calculations would be underestimated.

It should also be pointed out that the NMAE is exactly the measure of the unbalanced energy with respect to total PV electricity fed into the grid. Even if, in the literature, the most used indicator is the RMSE, also the NMAE evaluation is very important. In particular for the Italian Law the unbalanced PV energy could be a cost for the producers (if NMAE > 15%) or a gain (if NMAE < 15%). Thus, for the Italian case, the daily or annual NMAE target is 15%.

5.2 Models Performance analysis

The main results of all the forecast models are summarized in Table III.

First of all, it should be underlined that the horizontal global irradiance (Gh) is more persistent than the POA (Gpoa). Indeed, if the plane of the array is set to maximize the energy production, in clear sky days the Gh assumes much lower values then the Gpoa (mainly due to the direct component) while in cloudy days Gh and Gpoa are very similar. Thus, in instable weather conditions, the Gh forecast errors of the persistence model (PM) is much lower with respect Gpoa forecast. This brings to a lower NRMSE and NMAE of the horizontal irradiance PM prediction.

The NWP model shows a D_{NMAE} almost equal to zero while the I_{RMSE} is 23%.

From figures 4a and 4b, it can be observed that the forecast of the daily irradiation (Hh) is much better than the prediction of the hourly irradiance (Gh). This considerations indicate that the main source of errors of the NWP model is in the hourly values predictions, probably induced by site effects and low output time resolution (3 hours interval output data).

The NWP model shows comparable performance to the one found in [3] for the site of Bondville (which has a RMSE persistence similar to the Rome site). On the whole, this model provides just a small improvement of the hourly irradiance forecast.

It has to be remarked that the NWP provides the forecast of the horizontal irradiance while the other models predict the POA irradiance, so that the results are not directly comparable. Moreover, it should be pointed out that in the NWP model performance the "transposition factors" error is not taken into account. However, since the PM performance is calculated both for the horizontal and the POA irradiance the improvement of the models ($I_{RMSE,}$ D_{NMAE}) can be compared.

The performance of four ST models developed with MLPNN architecture were analyzed.

The model 2MLP uses as input, all the daily variables described in section 3, while the model 1MLP uses only the parameters OD, Hh, CR and Ta that have the maximum correlation with daily irradiance at day (t+1)

(Table I). The model 1MLP shows the best performance proving the right minimum choice of the input variables and confirming the result obtained in [5]. The model 3MLP explores a seasonal approach, thus for each season a MLPNN was developed and used in parallel. Even if this approach could potentially brings to good results, one season is insufficient to train and validate each NN. Thus, in this case, the 3MLP model shows worst performance with respect to the others.

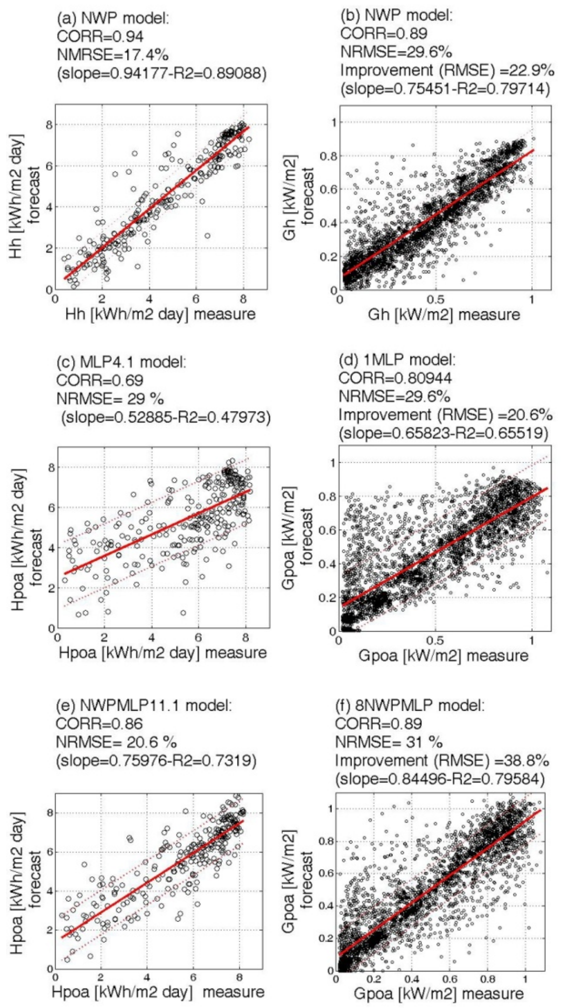

Figure 4: correlations between measured and forecasted data from different models

It has been observed that the correlation between the measured daily irradiation (Hpoa) at lag0 and lag1 is greater than the one between the hourly irradiance (Gpoa). Thus the forecast problem has been split in two steps, using two NN in series: one (MLP4.1) to forecast the Hpoa(t+1) and one (MLP4.2) to reconstruct the Gpoa from the forecasted daily irradiation (Hpoa(t+1)). The model 4MLP2Net summarizes the results of this 2 steps approach.

The MLP4.2 NN that predicts the hourly irradiance from the same day irradiation (shape irradiance forecast model), shows very good performance (*Corr=0.95, NRMSE=20% and NMAE=13%* see Figure 5).

Nevertheless the first step MLP4.1 is not enough well performing model (*Corr=0.69, NRME=29% and NMAE=23%*, see Figure 4c). Besides for the MLP4.1 model an over fitting trends in the training phase was observed, thus even if the 4MLP2Net model shows the best performance, this model has been considered less reliable than the others. Probably this NN model could be improved using more than one year data for training and validation.

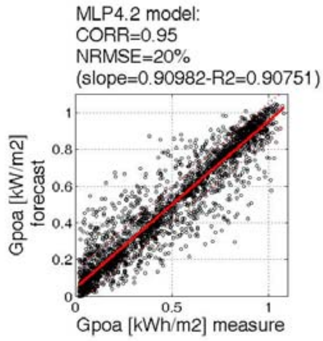

Figure 5: correlations between measured data and MPL4.2 model forecasted data (hourly forecasting from the same day irradiation)

These ST models exhibit similar performance to the one realize by the NWP. They realize an improvement around 20% in terms of RMSE but do not achieve any gains in terms of MAE (imbalanced energy measure). However, in this case, a different source of errors should be pointed out. From figures 4c and 4d, it appears that the MLP NN forecast of the daily irradiation is not better than the irradiance prediction (with a lower correlation of 0.7). On the other hand, Figure 5 shows that the MLP NN is able to provide a very good forecast of the hourly irradiance form the same day irradiation. Thus these ST models mainly fail in the daily irradiation forecast showing a slow reaction to weather changing conditions (see Figure 3). On the whole, they exhibit a forecast behavior comparable to the PM model, just providing a better performance during variable periods and worst performance in the more stable weather days.

Finally it should be remarked that there is not a general standard for ST NN model performance evaluation: different statistical indicators and test time interval are used or different variables are forecasted (horizontal, POA irradiance, PV power). Thus the one year results (270 days spread over one year) presented in this article, are not easily comparable with the ones reported in the literature [5, 6, 7, 9, 10]. Moreover a systematic study on the performance improvement dependence from the year and site should be done, thus the presented results could not be generalized.

Four Hybrid models based on MLPNN architecture were developed and studied. These models use NWP and local data as input and are trained with the site measured hourly POA irradiance. The 8NWPMLP model uses as input both the local measured meteorological parameters: OD, Hpoa, Ta, CR at lag0 and OD, and the NWP forecasted daily irradiation: Hh_nwp at lag1. The 9NWPMLP model uses directly the NWP predicted hourly horizontal irradiance: Gh_nwp_h(t+1) (*with h=1, 24*), while 10NWPMLP uses both the meteorological parameters and the forecasted NWP irradiance. Finally the model 11NWPMLP uses the 2 step approach: NWPMLP11.1 predict the daily POA irradiation Hpoa(t+1) from the meteo parameters measured at day (t) and (t-1) and Hh_nwp at (t+1) while the described MLP4.2 shape forecast model reconstructs the Gpoa(t+1) from the forecasted daily irradiation. All the models present an improvement over 37%. Since the obtained results are all very similar, the increasing performance of the Hybrid models does not depend neither to local meteo parameters (input of 8NWPMLP NN) nor to NWP predicted hourly irradiance (input of 9NWPMLP NN). On the other hand, from figures 4c and 4e a high improvement appears in the daily POA irradiation forecast between the MLP4.1 NN (that use only local daily parameters) and NWPMLP11.1 NN (that use also the NWP predicted horizontal irradiation).

Thus the main contribution of the NWP data to the Hybrid models performance is related to the good prediction of the daily horizontal irradiation (see Figure 4a), while the contributions of the ANN models is related to best hourly prediction (see Figure 5) and to the irradiance transposition. Hybrid models correct NWP hourly forecast and transpose the predicted horizontal irradiance on the POA. From figures 4a and 4e the transposition error can be evaluated: the NWP horizontal daily irradiation forecast shows a Corr=0.93 and an NRMSE= 17.4% while the POA daily irradiation forecast exhibits a Corr=0.86 and a NRMSE= 20.6%. Thus the irradiance transposition brings a loss of 7 points in correlation and 6 points in NRMSE.

Figure 3 shows the fast reaction of this model to the weather changing conditions due to the NWP irradiation forecast.

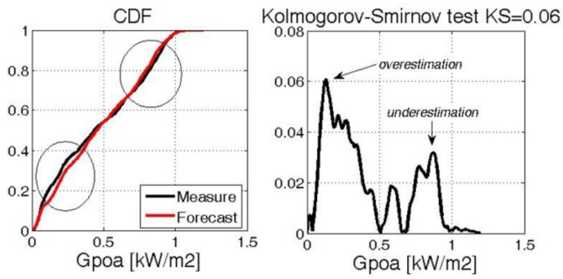

Figure 6: 8NWPMLP model Kolmogorov-Smirnov test

On the whole, these Hybrid models bring to high improvements: I_{RMSE} up to 37% and D_{NMAE} around 9% (see Table III). From the reporting site a NMAE of 20.2% has been realized, very near to the target of 15%.

All the reported models exhibit an underestimation of the forecasted value at high irradiance level and an overestimation at low irradiance level. This could be see also from the measured and forecasted Cumulative Distribution Function (CDF) curves (see for example Figure 6).

Figure 7 reports the monthly NMAE of the PM, NWP, 1MLP and 8NMPMLP models. All the models exhibit a lower performance with respect to the PM in the month of July and August while 1MLP model also in February and May. Besides, from this figure, the small improvement of the NWP and ST model can be observed.

From the seasonal point of view, the NWP fails in summer while the ST model also in spring time (see Table IV).

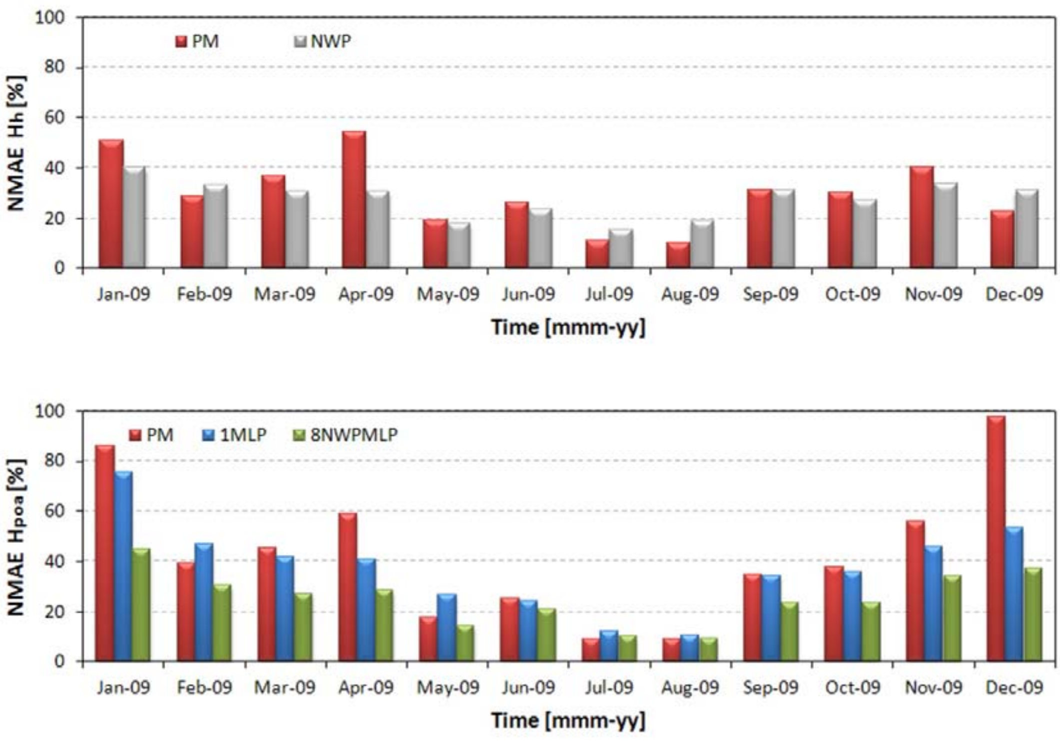

Figure 7: monthly performance of different forecast model

Table IV: seasonal performance (in bold the worst performance with respect to PM)

	NMAE Hh [%]		NMAE Hpoa [%]		
Season	**NWP**	**PM**	**1MLP**	**8NWPMLP**	**PM**
Winter	31.7	34.5	50.6	30.6	50.2
Spring	21.3	25.6	**28.5**	19.3	25.5
Summer	**19.8**	14.7	**15.1**	12.8	14.0
Autumn	28.0	30.5	38.5	26.2	45.5

Finally from Table V the performance difference between sunny and cloudy days can be observed. It should be noted that the NRMSE and Corr between sunny and cloudy days for the NWP forecast is less than two points while for the other models is greater than four points. This depends on the best daily irradiance forecast and on the more persistent features of the horizontal irradiance.

Table V: sunny and cloudy days performance

Model	NRMSE(%)		Corr	
	Sunny	Cloudy	Sunny	Cloudy
NWP	29	30.3	0.9	0.88
1MLP	37.1	44.5	0.83	0.78
8NWPMLP	28	34.9	0.91	0.87

5 CONCLUSIONS

Using Artificial Neural Network algorithms, several different POA irradiance forecast models on 24 hours horizon were developed. The performance of 4 Statistical models (ST) and 4 Hybrid models have been evaluated and discussed. The ST models use only site measured meteorological parameters as inputs while the Hybrid models also input NWP data. The improvement with respect to the Persistence model (PM) has been compared with the one obtained for the horizontal irradiance forecast using only the NWP model.

The reference site is the University of Rome “Tor Vergata” and the reference year for the models test is 2009.

It appears that the used NWP model and the ST models produce similar results with an improvement around 20% in terms of RMSE and no improvement in terms of NMAE. These performance results from different sources of forecast errors. The NWP model provides a very good forecast of the daily irradiation but fails in the hourly irradiance prediction. This error is probably due to site effect (low spatial resolution: 3 km^2) and to the low output temporal resolution (3 hours time interval of forecasted irradiation instead of 1 hour). On the other hand, the ST models are able to provide a good hourly forecast but could not well predict the daily irradiance in instable weather conditions.

The Hybrid models correct NWP hourly forecast taking into account the site effects and transpose the predicted horizontal irradiance on the POA. Thus they show the best performance increasing the improvement over 37% in terms of RMSE and 9% in terms of NMAE. The annual imbalanced energy measure (NMAE) of these model is around 20%, very near to the Italian threshold of 15%.

6 ACKNOWLEDGEMENTS

The DSP project is funded by the 2010 MISE-ICE-CRUI agreement between the Italian Ministry of the Economic Development (MISE), the Italian Trade Promotion Agency (ICE) and the Conference of the Italian Deans (CRUI). M. Pierro, F. Bucci and A. Taravat are supported in the framework of the same agreement. Thanks to Tecno El srl and Etexia srl that also funded the DSP project.

ECMWF ERA-40 (ECMWF Re-analysis) data used in this study have been provided by ECMWF.

The ESTER facility is part of the Centre for Hybrid and Organic Solar Energy (CHOSE).

7 REFERENCES

[1] H. M. Diagne, M. David, P. Lauret, J. Boland. Proceedings World Renewable Energy Forum WREF 2012, Denver, USA, May 2012.

[2] S.C. Muller, J. Remund. Proceedings 25th EUPVSEC Conference (2010) 4629.

[3] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., D. Rennè, T. E. Hoff. Solar Energy 84 (2010) 2161.

[4] L. Martin, L.F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, M. Cony. Solar Energy 84 (2010) 1772.

[5] A. Mellit, A. Massi Pavan. Solar Energy 84 (2010) 807.

[6] C. Voyant, P. Randimivololona, M.L. Nivet, C. Poli, M. Muselli. Meteorological Applications (2013) DOI: 10.002/met.1387.

[7] F. Wang, Z. Mi, S. Su, H. Zhao. Energies 5 (2012) 1355.

[8] R. A. Guarnieri, E.B. Pereira, S.C. Chou. Proceedings 8th ICSHMO (2006) 1777.

[9] C. Chen, S. Duan, T. Cai, B. Liu. Solar Energy 85 (2011) 2856.

[10] H. T.C. Pedro, C. F.M. Coimbra. Solar Energy 86 (2012) 2017.

[11] E. Lorenz, D. Heinermann, H. Wickramarathne, H. Beyer, S. Bofinger. Proceedings 20th EUPVSEC (2007).

[12] D. P. Dee et al. Q. J. R. Meteorol. Soc. (2011) 137 553–597.

[13] A. Spena, C. Cornaro, S. Serafini. Proceedings of the 33rd IEEE Photovoltaic Specialists Conference, (2008).