DOI: http://dx.doi.org/10.1109/EESMS.2014.6923259

Multi-point Solar Prediction through Feed-forward Neural Networks

Stefano Ferrari and Claudio Leani and Vincenzo Piuri
Department of Computer Science
Università degli Studi di Milano, Milan, Italy
Email: {stefano.ferrari,vincenzo.piuri}@unimi.it, claudioleani@gmail.com

***Abstract*—In this paper we present a study on the feasibility of the prediction of the solar radiation on a location giving the meteorological measurement in surrounding locations on a mesoscale system scale. The data from four public stations run by the Lombardy regional agency for environmental protection (ARPA) have been used as dataset for training a neural network in order to predict with one-hour lag the global radiation in one station using the data from the other three stations. The results have been compared with other two models: the first makes use of only the data from the station to be predicted, while the second exploits all the available information considering all the four stations as input sources. The dataset has been formed using data from the ARPA stations in Milano, Crema, Osio sotto, e Cassano d'Adda, considering the years 2002–2007.**

I. Introduction

Among the renewable sources, the solar energy plays a fundamental role, but has the drawback of its non-constant nature. Although its daily and yearly seasonality can be easily modeled, the actual quantity of the radiation that reaches the ground depends on meteorological factors. Proper forecasts of the radiation availability are required for almost all the processing phases, from the planning to the production, with different timescales (e.g., for planning where to place the site, long-term climatic forecasts are needed, while the daily management of a plant requires short term forecasts) [1].

The predictability of the solar radiation at a given site depends by large scale meteorological phenomenon as well local geographic features such as orography, presence of woods, urban induced micro-climatic phenomena.

Several techniques can be used to forecast the solar radiation. In literature approaches that use autoregressive models (e.g., ARIMA [2][3]), statistical models (e.g., Markovian models [4]), neural networks (e.g., feed-forward [5][6] or recurrent models [7]), Support Vector Machines (e.g., [8][9]) can be found. Besides the predictive paradigm used to model the phenomenon, these approaches differ also for the input data (number of previous samples, measured features), the prediction horizon, and the used performance figures. Moreover, it should be stressed that the direct comparison between approaches challenged on similar but different tasks and different data is not easy nor fair, since the variability of the solar radiation phenomenon with respect to both the geographical position and the year.

The use of temporal series can allow to capture both the seasonal and the local trends. However, the measurement of the meteorological parameter is usually expensive and requires trained personnel.

In this paper, we explored the feasibility of an indirect approach, which exploits the data collected by public station for producing the forecast for which the solar radiation measurement are not available. In particular, we investigated the possibility of predicting the solar radiation on one site given the measurements at three surrounding sites.

The paper is organized as follows: in Section II, the paradigm used for the prediction, the feed-forward neural network, is introduced; in Section III the elements that characterize the experiments (i.e., the dataset, the indices used to measure the performance of the predictors, the information used for the prediction, and the models we challenged) are described; in Sections IV and V the results of the experiments are respectively reported and discussed.

II. Prediction models

A. Feed-forward Neural Networks

The feed-forward neural networks (FNN) [10][11][12] are computational models that enjoy the universal approximation property, which made them a suitable tool to approximate continuous functions. They are composed of processing units (called neurons) organized in layers. Each neuron computes its output as a function of a linear combination of the output of the neurons of the previous layer (this function is often called transfer function or activation function). The information, hence, flows only from the input layer to the output layer. It can be proved that a network with one hidden layer (i.e., a layer between the input and the output layers) enjoys the universal approximation property. A FNN is characterized by the number of neurons of the hidden layer, the activation function, Ψ, (usually sigmoidal), and by the learning algorithm used

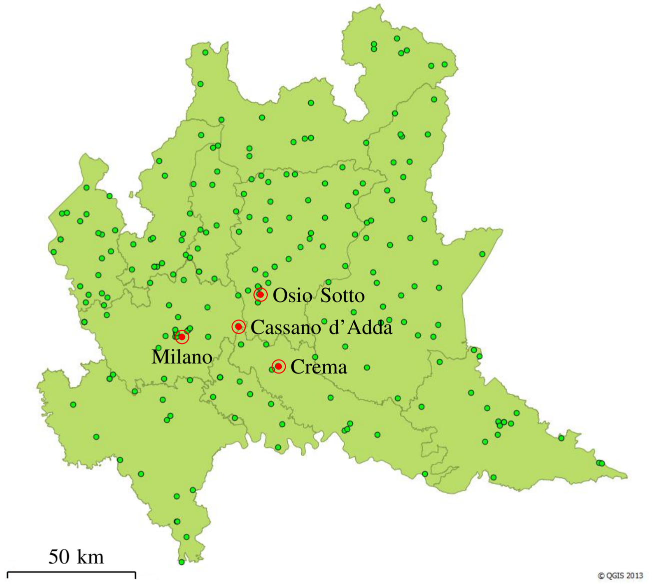

Fig. 1. The ARPA stations. The green circles identify the location of the stations in the Lombardy area. Among them, the stations involved in our experiments are depicted using concentric red circles and the name of the city where they are deployed.

(usually gradient descent based, such as Marquardt algorithm [12]).

More formally, the output, $f_{\text{FNN}}(\cdot)$, of a single layer FNN is

$$f_{\text{FNN}}(x) = \beta_0 + \sum_{j=1}^{L} \beta_j \Psi(\gamma_j^T \cdot x) \tag{1}$$

where L is the number of units of the hidden layer, the β_j is the weight of each neuron (β_0 is a bias term), the γ_j represents the weight vector of the linear combination of input for the j-th neuron.

The function Ψ, which can be chosen among different functions, is often the hyperbolic tangent

$$\Psi(z) = \frac{1 - \exp(-2z)}{1 + \exp(-2z)}. \tag{2}$$

III. Experiments

A. Datasets

The use of data from public stations allows to relieve the application from the costs of the instrumentations and their maintenance. ARPA Lombardia [13] is the Lombardy regional agency for the environmental protection which provides meteorological data archive of the measurements done from the nearly 250 stations distributed in the regional area (see Fig. 1). The ARPA dataset includes the hourly average of several measured quantities, such as: global radiation (GR), air temperature (AT), rainfall (RF), atmospheric pressure (AP), wind speed (WS), and relative humidity (RH). We identified four stations suitable for our experiments, namely:

- Milano (via Juvara);
- Crema (via XI febbraio);
- Osio Sotto;
- Cassano d'Adda.

The position of the considered station is reported in Fig. 1.

The selection considered both the geographic location and the data availability. In fact, in order to distribute the relative importance of each input source for the prediction, the geometrical configuration of the selected station should be such that one station (the one that will be predicted) is almost regularly surrounded by the others (those that will provide the information for the prediction). Besides, in order to improve the available information, the missing data for the main instruments must be restrained as much as possible, with the availability of several years (in order to capture the yearly seasonality). The distance between the stations has been also considered: if it is too short, the prediction would be a trivial task, due to the locality of the phenomenon; if it is too large, the prediction would be impossible, due to the low correlation between the measurement of the different stations. The distance between the surrounding stations and the central one ranges 15–23 km, which correspond to a mesoscale system size. Since the feed-forward neural network does not need the time series of the data, the presence of missing data does not prevent the use of the dataset for training the model, but only implies the use of a smaller example dataset, because the missing data are simply discarded. We used the data from the years 2002 and 2003 for the training, the 2004 for the feature selection, and 2005 for the testing.

Since the present work concerns the solar radiation, only the daylight measurements have been considered. In particular, only the data in the time range 07:00–18:00 have been used for the experiments.

B. Performance indices

To quantitatively evaluate the performance of the challenged models, several indices have been used, namely, the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), and the Relative Weighted Error ($\widetilde{\text{ER}}$). In the following, the formal description of the above mentioned indices is provided, where y_i refers to the i-th measured value, $\hat{y}_i$ is the i-th predicted value, and n is the number of examples used in the performance evaluation.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{3}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{4}$$

$$\widetilde{\text{ER}} = \sum_{i=1}^{n} \frac{|y_i|}{\sum_{i=1}^{n} |y_i|} \cdot \frac{|\hat{y}_i - y_i|}{|y_i|} = \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{\sum_{i=1}^{n} |y_i|} \tag{5}$$

Both RMSE and MAE weight equally each value considered in the prediction. The main difference is that the RMSE is more sensitive to the outliers; hence, their difference is proportional to the variance of the error distribution. Instead,

$\widetilde{\text{ER}}$ weights more the contribution of those elements where the real value is (relatively) high. In particular, it measures the relative error weighted with the relative importance of the real value. This performance index can be more meaningful for an application where the ability to predict the peaks of the energy production is more valuable than the reliability of the prediction on the average.

C. Persistence

The persistence is a naïve predictor that can be used to provide a baseline forecast. It can be formalized as a predictor which realizes the identity function defined on the output domain; since it produces the output (the prediction) always equal to the input (the observable), it performs well only for trivial prediction, where the phenomenon to be modeled is almost constant. Hence, it is obviously inappropriate for prediction of interesting real cases, and any other model is supposed to perform better than the persistence model.

D. Feature selection

In order to find the best set of input features, a preliminary experiment has been run. Several neural models have been challenged with a different set of input features and different network size. The models have been used to predict the global radiation (GR) with one hour lag for the Cassano d'Adda station. In particular, in addition to the time of the prediction, the following sets:

- $x =$ [RH, GR]
- $x =$ [AT, RH, GR]
- $x =$ [AT, RU, GR, AP]
- $x =$ [RF, AT, WS, RH, GR]
- $x =$ [RF, AT, WS, RH, GR, AP]

and the following network sizes have been challenged:

$$L = \{15, 20, 25, 30, 35, 40, 45, 50\} \tag{6}$$

for a total of 40 different configurations.

For each configuration, five trials have been run, in order to average the random effects of the training procedure. The models have been trained using the 2002–2003 data and tested on the 2004 data. The average RMSE for each configuration, E, and the related standard deviation, σ, have been used to select the best configuration (the one which minimized $E+\sigma$). The input features of the best model (which resulted to be $x =$ [AT, RH, GR]) has been then considered as the input set for the models challenged in the main experiment: all the models considered in the following will make use of the air temperature, the relative humidity, and the global radiation measurements at the input stations measured at a given time and will predict the global radiation value at the Cassano d'Adda station at one hour ahead.

E. Prediction models challenged

In order to evaluate the capability to predict the solar radiation in an arbitrary location, we challenged different models on the 1-hour prediction of the global radiation in the Cassano d'Adda station. In particular, the prediction based on the three surrounding stations (Milano, Crema, and Osio Sotto) has been confronted to a prediction based on the Cassano d'Adda measurements only and to the prediction obtained using the measurement from all the four stations. These three models will be named after the acronyms of the used locations: MiCrOs, Ca, MiCrOsCa, respectively. The rational of considering also the Ca and the MiCrOsCa models is to allow the proper evaluation of the predictive capability of the proposed model, MiCrOs. In fact, the Ca model provides a baseline for comparing the ability of the paradigm to provide a prediction using the geographically closest station (the Cassano d'Adda station itself). The MiCrOsCa model, instead, provides a hint of the paradigm's capability to exploit all the available information.

For each of the models, networks of different size, L have been trained. In particular, the same number of neurons for the hidden layer as in the feature selection experiment have been tried (as listed in (6)). For each set-up, a pool of five networks have been trained, in order to mitigate the randomness of the learning procedure. This allows to obtain a more robust estimation of the performance of each configuration as the winner set-up can be obtained as the one of the model that minimize the average test error.

The models have been trained using the data of the years 2002 and 2003, while they have been tested on the year 2005 data. They have been fed with the input feature set resulting from the procedure described in III-D, as reported in the Results Section (see Section IV-A), and trained for predicting the global radiation with one hour lag.

The best networks for the MiCrOs, Ca, MiCrOsCa prediction have been then validated with the year 2007 data.

IV. Results

A. Feature selection

Table IV-A reports the results obtained with the best model for each feature set on the year 2004 data of the Cassano d'Adda station, as described in III-D. For each set-up, the performance has been evaluated as the sum of the average and the standard deviation of the RMSE, $E + \sigma_E$, on the five trials. The model fed with air temperature, relative humidity, and global radiation ([AT, RH, GR]) shown the best prediction capability, achieving an error of 65.77 W/m^2. This feature set has been adopted for the main experiment.

For comparison purposes, also the persistence performance computed on the global radiation data only is reported in

TABLE I
FEATURE SELECTION. COMPARISON OF THE PERFORMANCE OF THE FIVE INPUT SETS CHALLENGED (IN W/M^2).

Model	E	σ_E	$E + \sigma_E$
[RH, GR]	65.73	0.14	65.87
[AT, RH, GR]	**65.46**	**0.31**	**65.77**
[AT, RU, GR, AP]	65.42	0.37	65.79
[RF, AT, WS, RH, GR]	66.12	0.27	66.39
[RF, AT, WS, RH, GR, AP]	66.03	0.44	66.47
Persistence [GR]			113.8

TABLE II
MODEL CA PERFORMANCE (IN W/M^2).

L	E	σ_E	$E+\sigma_E$	$\widetilde{\text{ER}}$	$\sigma_{\widetilde{\text{ER}}}$	$\widetilde{\text{ER}}+\sigma_{\widetilde{\text{ER}}}$	E_g
15	60.6	21.8	82.4	0.166	0.0647	0.231	19.03
20	60.9	21.2	**82.2**	0.167	0.0622	0.230	18.86
25	60.9	21.5	82.3	0.167	0.0668	0.234	19.57
30	62.0	21.2	83.2	0.166	0.0673	0.233	19.65
35	61.8	21.2	83.0	0.169	0.0672	0.236	19.87
40	61.1	21.4	82.4	0.165	0.0571	**0.222**	**18.27**
45	63.1	20.6	83.7	0.174	0.0626	0.236	19.77
50	62.5	20.5	83.0	0.172	0.0649	0.237	19.65
Persistence			115.8			0.293	34.0

TABLE III
MODEL MICROS PERFORMANCE (IN W/M^2).

L	E	σ_E	$E+\sigma_E$	$\widetilde{\text{ER}}$	$\sigma_{\widetilde{\text{ER}}}$	$\widetilde{\text{ER}}+\sigma_{\widetilde{\text{ER}}}$	E_g
15	69.9	21.3	91.2	0.198	0.0726	0.270	24.6
20	68.1	22.1	90.2	0.191	0.0708	**0.262**	**23.6**
25	69.0	21.3	90.3	0.199	0.0751	0.274	24.8
30	71.3	18.9	90.2	0.212	0.0934	0.305	27.5
35	71.3	18.4	89.7	0.218	0.0876	0.306	27.4
40	70.8	21.9	92.7	0.205	0.0874	0.292	27.1
45	72.6	22.6	95.2	0.218	0.0905	0.308	29.3
50	71.8	16.2	**88.0**	0.200	0.0824	0.282	24.8
Persistence			111.8			0.301	33.6

Table IV-A.

B. Prediction of solar radiation

Table II reports the performance achieved by the trained Ca models. For every network size, the average RMSE, E, its standard deviation, σ_E and their sum is reported in the columns from the second to the fourth. In the next columns, the weighted relative error, its standard deviation, and their sum is reported. Since the models that perform well in the sunny days may achieve a better E, while those that perform better with low values of solar radiation may achieve a lower ER, in the last column, a new indices has been introduced to summarize both the behavior: the global error, E_g, which is defined as:

$$E_g = (E + \sigma_E) \cdot (\widetilde{\text{ER}} + \sigma_{\widetilde{\text{ER}}}) \tag{7}$$

The rationale on the use of E_g is that a model performing reasonably on both the indices can also achieve a low value of E_g.

TABLE IV
MODEL MICROSCA PERFORMANCE (IN W/M^2).

L	E	σ_E	$E+\sigma_E$	$\widetilde{\text{ER}}$	$\sigma_{\widetilde{\text{ER}}}$	$\widetilde{\text{ER}}+\sigma_{\widetilde{\text{ER}}}$	E_g
15	58.0	25.8	83.8	0.184	0.0718	0.255	21.4
20	58.9	27.2	86.1	0.197	0.0816	0.279	24.0
25	56.1	27.6	83.7	0.177	0.0734	**0.251**	**21.0**
30	56.7	26.1	**82.8**	0.191	0.0841	0.275	22.8
35	60.8	25.6	86.4	0.193	0.0717	0.265	22.9
40	59.2	26.6	85.8	0.202	0.0809	0.283	24.2
45	62.5	27.2	89.7	0.221	0.0931	0.314	28.2
50	62.0	27.2	89.1	0.216	0.0835	0.299	26.6
Persistence			110.6			0.291	32.1

TABLE V
COMPARISON OF THE AVERAGE ERRORS OF THE THREE MODELS EVALUATED ON THE 2007 DATA.

Model	Error	$\hat{y}$	σ	$\hat{y}+\sigma$
Ca	RMSE	61.05	21.39	82.44
	MAE	39.66	12.83	52.49
	$\widetilde{\text{ER}}$	0.18	0.08	0.26
	E_g			21.4
MiCrOs	RMSE	66.75	21.40	88.15
	MAE	44.82	13.28	58.10
	$\widetilde{\text{ER}}$	0.19	0.07	0.26
	E_g			22.9
MiCrOsCa	RMSE	60.86	22.02	82.88
	MAE	40.54	13.26	53.79
	$\widetilde{\text{ER}}$	0.18	0.07	0.25
	E_g			20.7

Similarly, Tables III and IV report the performance achieved by the models MiCrOs and MiCrOsCa, respectively. All the Tables report also the persistence performance computed on the year 2005 data. While for the Ca model case the persistence has been computed as described in Section III-C, for the other models the indices have been computed considering as input the average of the GR input data.

The best models selected using the E_g index for the Ca, MiCrOsCa, and MiCrOs type of predictor have been tested on the year 2007 data. The performance indices described in Section III-B have been computed and are reported in Table V.

V. DISCUSSION

From Tables II–IV several considerations common to all the models can be done. The persistence model, as described in Section III-C, is often used as baseline for the predictors of phenomena that evolves in time. For all the considered models, the persistence computed using the global radiation at the input station is always larger than the corresponding indices obtained by the trained predictors. In particular, the lowest values for the RMSE, the $\widetilde{\text{ER}}$, and the E_g are respectively the 71.0%, 75.8%, and 53.7% of the persistence for the networks of the Ca model, while these figures are 78.7%, 87.0%, and 70.2% for the MiCrOs and 74.9%, 86.3%, and 65.4% for the MiCrOsCa model. These values confirm that all the models have the capability of modelling the considered solar radiation phenomenon. However, for all the models, the networks that have the lowest value of RMSE are typically different from those that minimize the $\widetilde{\text{ER}}$. Introducing the E_g index help to take a decision in the best network topology. It should be noted, however, the indices of the different networks are quite similar. Hence, the selection of the model can be biased by the specific data used for the selection of the best network.

A similar consideration can be done for the comparison of the MiCrOs model with the Ca, MiCrOsCa models. For this task, we challenged the networks selected by the E_g index on the year 2007 data (Table V). All the indices are quite similar among the models, although the MiCrOs performance is always poorer than the Ca, MiCrOsCa models. Noticeably, the Ca model is slightly better than MiCrOsCa in terms of RMSE and MAE, while it is slightly worsen in terms of $\widetilde{\mathrm{ER}}$, and E_g. Although no clear conclusions can be drawn, from Table V it seems that the use of the three surrounding stations can provide a reliable prediction, with a loss of performance of the 9.59% with respect to the case when also the information relative to the central station are available.

The selection of the features to be used in the experiment has been realized challenging different feature sets on the one hour lag prediction from one single station (the Cassano d'Adda station). Although the results in Table IV-A identify as the best performing the [AT, RH, GR] set, it should be noticed that the achieved performance is not very different by the performance of the other sets. In particular, using the persistence as baseline, they score in the range 0.578–0.584 of that index. Hence, there is not a clear evidence that the [AT, RH, GR] set performs better than the other feature sets. The procedure for the feature selection made use of only the Cassano d'Adda information, and hence it could bias the performance at the advantage of the Ca model. This procedure has been chosen to avoid to add more complexity in a preliminary study, with the working hypothesis that features well performing for a single station prediction would be suitable also for the multi-station case. In the light of the results, this decision can be questioned, and new experiments can be required.

In order to be usable for the prediction of a general place, the latitude and longitude of the site to be predicted should be added as input for the network. Besides, the prediction can enjoy also of features that can help in the estimation of the atmospheric condition (and hence of the transparency to the solar radiation) such as gases concentration in the atmosphere. However, the increment in the number of the input features should be carefully considered, since it can cause a decrease in the performance due to the well known effect called "curse of dimensionality". As the dataset size will increase, a larger number of features could be effectively used.

The experiments have been run using five networks to average the effect of the randomness of the training procedure. This fact can also be exploited to obtain the confidence interval of the prediction.

VI. Conclusion

In this paper, the use of single layer feed-forward neural networks for the solar radiation prediction has been studied. In particular, the prediction used the data from three stations to predict the level of solar radiation in a site placed in the region delimited by the considered stations.

The experiments have been run using data from public stations. This solution allows to obtain reliable prediction avoiding the efforts required to maintain a measurement stations at an efficient level.

The experiments shown that although the prediction is degraded with respect to the cases when the local information are available, an appreciable prediction can be obtained.

Future directions of the research will consider new feature sets and different network topologies.

References

[1] V. Kostylev and A. Pavlovski, "Solar power forecasting performance — towards industry standards," in *1st International Workshop on the Integration of Solar Power into Power Systems*, Aarhus, Denmark, Oct. 2011.

[2] D. Yang, P. Jirutitijaroen, and W. M. Walsh, "Hourly solar irradiance time series forecasting using cloud cover index," *Solar Energy*, vol. 86, no. 12, pp. 3531–3543, 2012.

[3] S. Ferrari, M. Lazzaroni, V. Piuri, L. Cristaldi, and M. Faifer, "Statistical models approach for solar radiation prediction," in *Instrumentation and Measurement Technology Conference (I2MTC), 2013 IEEE International*, May 2013, pp. 1734–1739.

[4] M. Muselli, P. Poggi, G. Notton, and A. Louche, "First order Markov chain model for generating synthetic "typical days" series of global irradiation in order to design photovoltaic stand alone systems," *Energy Conversion and Management*, vol. 42, no. 6, pp. 675–687, 2001.

[5] M. Rizwan, M. Jamil, and D. Kothari, "Generalized neural network approach for global solar energy estimation in India," *Sustainable Energy, IEEE Transactions on*, vol. 3, no. 3, pp. 576–584, July 2012.

[6] S. Ferrari, M. Lazzaroni, V. Piuri, A. Salman, L. Cristaldi, and M. Faifer, "Computational intelligence models for solar radiation prediction," in *Instrumentation and Measurement Technology Conference (I2MTC), 2013 IEEE International*, May 2013, pp. 757–762.

[7] G. Capizzi, C. Napoli, and F. Bonanno, "Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 23, no. 11, pp. 1805–1815, Nov 2012.

[8] J. Zeng and W. Qiao, "Short-term solar power prediction using a support vector machine," *Renewable Energy*, vol. 52, no. 0, pp. 118–127, 2013.

[9] F. Bellocchio, S. Ferrari, M. Lazzaroni, L. Cristaldi, M. Rossi, T. Poli, and R. Paolini, "Illuminance prediction through SVM regression," in *Environmental Energy and Structural Monitoring Systems (EESMS), 2011 IEEE Workshop on*, Sep. 2011, pp. 1–5.

[10] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[11] R. Lippmann, "An introduction to computing with neural nets," *ASSP Magazine, IEEE*, vol. 4, no. 2, pp. 4–22, 1987.

[12] M. Hagan and M.-B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *Neural Networks, IEEE Transactions on*, vol. 5, no. 6, pp. 989–993, 1994.

[13] ARPA Lombardia. [Online]. Available: http://ita.arpalombardia.it/