© 2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Illuminance Prediction through Statistical Models

S. Ferrari, A. Fina, M. Lazzaroni, and V. Piuri
Università degli Studi di Milano
Milan, Italy
Email: {stefano.ferrari, massimo.lazzaroni, vincenzo.piuri}@unimi.it
alberto.fina@studenti.unimi.it

L. Cristaldi, M. Faifer, and T. Poli
Politecnico di Milano
Milan, Italy
Email: {loredana.cristaldi, marco.faifer, tiziana.poli}@polimi.it

***Abstract*—A reliable forecast of renewable energies production, like solar radiation, is required for planning, managing, and operating power grids. Besides, the short-term prediction of the climatic conditions is very useful for many other purposes (e.g., for Climate Sensitive Buildings). Data for the prediction can be produced by several sources (satellite and ground images, numerical weather predictions, ground measurement stations) with different resolution in time and space. However, the unsteadiness of the weather phenomena and the variability of the climate make the prediction a difficult task, although the data collected in the past can be used to capture the daily and seasonal variability. In this paper, several autoregressive models (namely, AR, ARMA, and ARIMA) are challenged on a two-year ground solar illuminance dataset measured in Milan, and the results are compared with those of simple predictor and results in literature.**

I. Introduction

The exploitation of renewable energies is an important aspect of the actual and future energy policies. Among the renewable energy sources, the photovoltaic (PV) technology is one of the most actively studied, since it allows to obtain electric energy from solar radiation [1]. Although the efficiency of the nowadays technology is questionable, the appealing of this way to produce energy lies in its cost-effectiveness and low environmental impact.

The main weakness of this renewable energy source is that its availability cannot be controlled. In fact, the availability of the solar radiation depends on different factors: geographic position, local climate, and weather are the most important. Among these, the position and the climate influence onto the solar radiation can be easily stated from astronomical and statistical data, but the weather is characterized by a high variability and depends on many physical factors. Besides, the PV energy harvesting is operated by many small plants, geographically spread out on a large area. The energy production is often consumed for facing the needs of nearby buildings, but the excess flows into the power grid. The unpredictability of the amount of energy available is a problem for the grid operator, that has to integrate the energy required by the users with traditional non-renewable energy sources. Hence, a trustable solar radiation forecasting is a powerful tool for an efficient grid management.

Depending on their use, the forecasts are required with a different horizon and with a different granularity. According to [2], the forecasts required by the activity related to the grid management can be partitioned in three categories (intra-hour, hour ahead, and day ahead), while those related to planning and assets optimization can be categorized in medium and long-term (monthly and yearly forecasts, respectively).

Since the main factor for solar radiation availability is the local weather, approaches based on weather forecast have been widely used in literature. These are based on data obtained from satellite observations and ground stations. Two main aspects should be taken into account, namely, geographic and time availability. Since direct measurement of solar radiation is not available for all the sites of interest, data from nearby sites have to be used to obtain the desired data. Although several approximation techniques and algorithms are largely available in literature, in this case the local geographical features of the site has to be carefully considered. For instance, large variations in elevation or in surface type [3] can produce micro-climate conditions very different from those of the available sites. Besides, the sampling rate of the available measurement have to be related to the granularity of the forecast.

The solar radiation prediction can be based on data produced by several data sources. They are characterized by the type of data they produce, as well the space-time granularity they provide:

- Numerical Weather Prediction (NWP) models are global models that provide forecast at 1 to 3 hours granularity, with a few days horizon, but with a few square kilometers spatial resolution. They are used to produce predictions on physical variable that can be related to the solar radiation.
- Satellite-base forecast are very effective in capturing the cloudiness of a large geographical area. Since the cloudiness is the main factor that affects the solar radiation, satellite imagery can provide critical information for the PV production forecast. The spatial granularity of the prediction is limited to one square kilometer, although they can provide better one to six hours ahead predictions than NWP-based approaches. However, the long refresh time of the satellite images, as well as the problem of estimating the cloud altitudes and their projected shadows, limits the accuracy of the approaches of this type.

© 2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

- All-sky imagers allow to capture the image of the hemispherical sky around the ground based observation station and provide near real-time prediction of the cloud position in a region nearby the station site. A sequence of images, acquired at a high sampling rate, can be processed with cloud-detecting algorithms and used to predict the clouds movement and evolution. These approach can be very effective for short term (minutes) and geographically local predictions.
- Ground measurements are very useful since they allow a direct measurement of the solar radiation. Their data can be used both for producing a local prediction and for assessing the performance of the models used for the prediction. Ground measurement are usually available at a high rate, although the data are valid only in a limited neighborhood of the site. Besides, the measurement accuracy can be affected by the presence of dirt or birds.

Several forecasting approaches have been used in literature. Among these, the most effective in producing hour-ahead predictions are based on empirical regression, neural networks [4] and time-series models (e.g., ARMA, ARIMA) [5][6].

In this paper, a two years hourly dataset will be used to model the time series of the global horizontal illuminance is considered. The dataset has been collected by the MeteoLab [3][7] between October 2005 and October 2007. In a previous work [8], a Support Vector Machine (SVM) model has been used for forecasting this dataset through regression. In the present work, the illuminance values are considered as a time series and autoregressive models are used to perform a forecast. The prediction operated by the autoregressive models will be compared with a naïve predictor, the persistence model, a simple predictive model, namely the k-Nearest Neighbor (k-NN) model, and the SVM model from [8].

II. Time-series Models

A time series is composed of a sequence of observation x_t sampled by a sequence of random variables X_t. Usually, the ordering value is related to the time and the observation are related to a phenomenon that varies with the time. A practical assumption is that the observations are taken in equally spaced instants.

A. Autoregressive Models

An autoregressive model describes the values of a particular time series in terms of its past values [9]. In particular, the value of X_t is modeled as a combination of a part that is determined by the past values of the series and a part determined by an unpredictable event that happens at the time t (innovation). More formally, given a time series $\{X_t\}$, its autoregressive representation of order p, often denoted by AR(p), is:

$$X_t = \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{t-k} + \varepsilon_t \tag{1}$$

where α_0 is a constant and the innovation ε, is assumed to be white noise ($E(\varepsilon) = 0$, $E(\varepsilon^2) = \sigma^2$) and $\{\varepsilon_t\}$ are supposed to be normal i.i.d. variables.

A moving average model describes the time series values in terms of linear combination of (unobserved) innovation values. A moving average representation of order q, often denoted by MA(q), of the time series $\{X_t\}$ is:

$$X_t = \mu + \sum_{h=1}^{q} \beta_h \varepsilon_{t-h} + \varepsilon_t \tag{2}$$

The autoregressive and moving average models can be combined in the autoregressive moving average model (ARMA). An ARMA representation of autoregressive order p and moving average order q, ARMA(p, q) is formally described as:

$$X_t = \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{t-k} + \sum_{h=1}^{q} \beta_h \varepsilon_{t-h} + \varepsilon_t \tag{3}$$

When the time series is sampled from a stationary process, it can be represented by the above mentioned models. However, when the time series shows a trend or a seasonality, a more advanced class of models, namely the autoregressive integrated moving average models (ARIMA), have to be used. The ARIMA model take into consideration also the difference series (i.e., the series resulting by computing the difference of time lagged series). In particular, the notation ARIMA(p, d, q) is commonly used for indicating the ARIMA model with p, d, and q order of respectively autoregression, differencing, and moving average. In order to formalize this model, the backward shift operator, B, have to be introduced:

$$X_{t-1} = B\, X_t \tag{4}$$

This allows to express X_{t-k} as $B^k X_t$. The ARIMA(p, d, q) representation of the time series $\{X_t\}$ is:

$$\left(1 - \sum_{k=1}^{p} \alpha_k B^k\right)(1-B)^d X_t = \left(1 + \sum_{h=1}^{q} \beta_h B^h\right)\varepsilon_t \tag{5}$$

B. Persistence

In order to assess the performance of model in the short-term prediction of a time series, the persistence model is often used. It is a naïve predictor that assumes that the next value of the time series, x_t will be equal to the last known, x_{t-1}. It is obviously inappropriate for long-term prediction of time-series of interest in real cases, but it can be used as a baseline forecast: it is supposed that any other model will perform better than the persistence model.

C. k-Nearest Neighbor Interpolator

The k-Nearest Neighbor (k-NN) model is a instance-based or lazy learning paradigm used both for function approximation and classification [10]. It is used to predict the value of a function, f, in unknown points, given a sampling of the function itself (training data), $\{(x_i, y_i) \,|\, y_i = f(x_i)\}$. For an unknown point, x, the value of $f(x)$ is estimated from the value of its k nearest neighbors, for a given k, using a suitable voting scheme or an average. The most simple scheme, often used in classification, estimates $f(x)$ as the most common output value among its neighbors, while in function

approximation the average output value is often used. More complex schemes, such as the use of weighted averaging, or a sophisticated norm for computing the distance can be used as well. The k-NN can be used in time series prediction using some previously observed values for composing the input vectors. For instance, when using a two-dimensional feature space, the training dataset will be composed by triples of the form (x_{t-2}, x_{t-1}, x_t), where will be assumed that $x_t = f(x_{t-2}, x_{t-1})$.

III. Experiments

The dataset used in the experiments described in the present paper has been collected by the MeteoLab [3][7] between October 2005 and October 2007. MeteoLab measures:

- air temperature;
- relative humidity;
- global horizontal irradiance;
- diffuse horizontal irradiance;
- global horizontal illuminance.

The station samples the data every ten minutes, but the dataset used here considers only their hourly average.

The illuminance varies both on daily and seasonal basis. The surface reported in Fig. 1 shows this behavior. It has been obtained by averaging the illuminance samples measured in the same hour of the same day of the year. Although there is a clear trend, the variability of the illuminance (which depends also by fast changing meteorological phenomena) makes the resulting surface very rough.

Figure 2, instead shows the relation between the illumination acquired at two successive hours. In particular, in Fig. 2a the distribution of the points along the identity line supports the use of the persistence predictor. However, the maximum of the prediction error of the persistence can be considerably high: in fact, it can be estimated as the length of the vertical section of the cloud of points, which is at least 40 long. The histogram in Fig. 2b resembles a mixture of two normal distributions with the same mean. This is due to the fact that in the early and the late daylight hours (not to mention the hours around midday), the illuminance is almost the same (especially in the winter). Hence, the consecutive samples acquired in those period of time are quite similar, while the other moments of the day show a larger variability.

A. Dataset Pre-Processing

For this work, we focused only on the global horizontal illuminance (i.e., the fraction of the solar radiation that can be actually perceived by the human eyes). Since time series models requires that all the values are equally time spaced, the few values that are missing are interpolated using a simple rule that exploits the daily seasonality of the solar radiation. For each missing value, x_t, the set $\{x_{t-1}, x_{t+1}, x_{t-24}, x_{t+24}\}$, i.e., the set composed of the illuminance one hour before and ahead, and one day before and ahead are considered. The missing value is then replaced with the average of the collected values. Since the missing data are few, the selected set has a meaningful number of elements even though some of the selected elements are missing too.

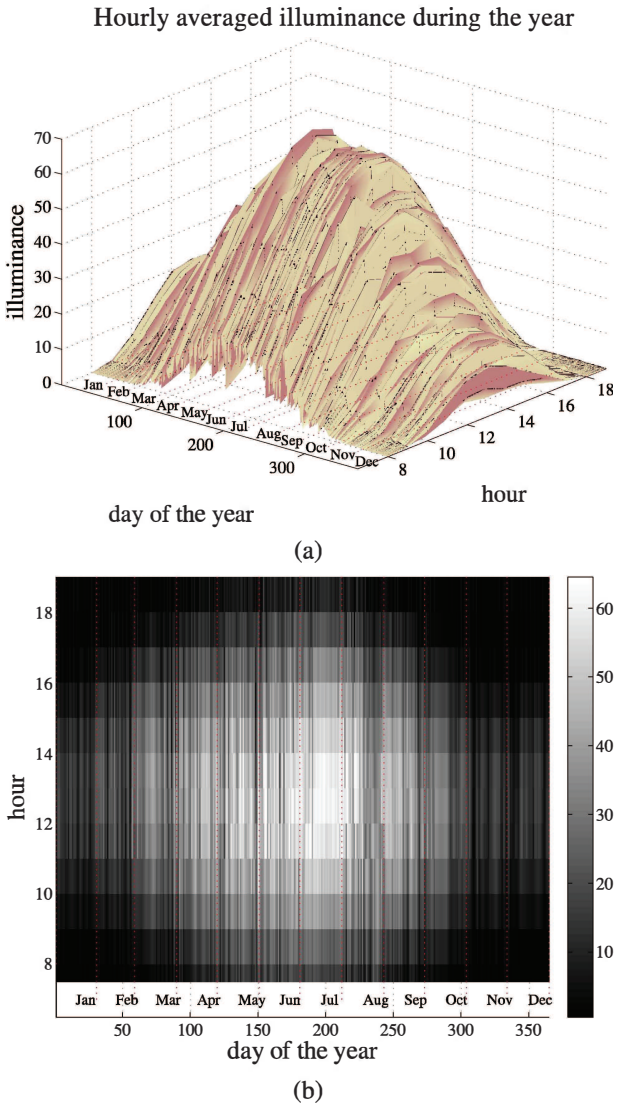
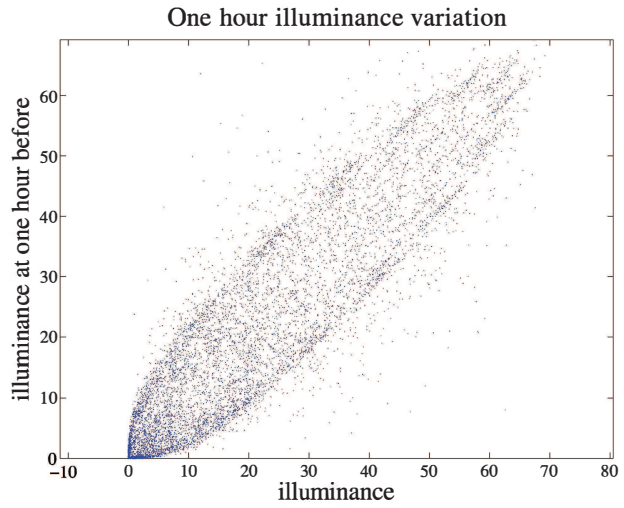


Fig. 1. The illuminance samples belonging to the same day of the year and to the same hour have been averaged and plotted as a surface. It is evident that there is a trend, but with an high variability which results in a non-smooth surface.

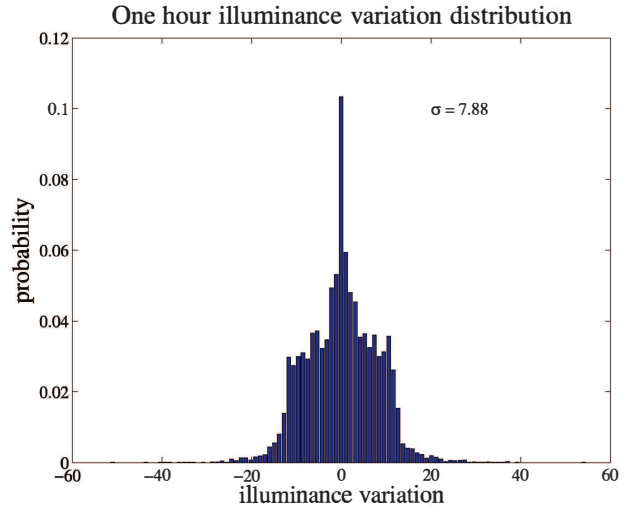
The resulting dataset is composed of 18096 samples. Since the dataset covers a period of time of two years, and the autoregressive models require a training set composed of consecutive data, the first year has been used as training set. In this way, the yearly variability have a chance of being captured by the models. The data belonging to the second year has been randomly partitioned in the validation and training set. Hence, training, validation, and testing set are composed of, respectively, 9048, 4524, and 4524 samples.

B. Performance Evaluation

For the evaluation of the performances, only the daylight hours data ([8, 19]) has been considered. Besides, since the il-



(a)



(b)

Fig. 2. The persistence predictor makes direct use of the illuminance value measured one hour before. In panel (a), the relationship between the two illuminance measurement made at distance of one hour one from the other is shown. Although the samples populate the region around the identity line, an evident dispersion is shown. In panel (b), the estimated probability density function of the variation (which standard deviation is 7.88).

luminance cannot be negative, all the negative values predicted by the models are remapped to zero.

The prediction error has been measured by means of the average of the absolute error achieved on the testing set data:

$$\mathrm{Err}(f) = E(|x_t - f(x_t)|) \tag{6}$$

where $f(x)_t$ is the value for x_t predicted by the model f.

Another performance statistics used in solar radiation prediction is the mean relative error:

$$\mathrm{Rel}(f) = E\left(\frac{|x_t - f(x_t)|}{|x_t|}\right) \tag{7}$$

C. Prediction through k-NN Models

The performance of a k-NN predictor depends on several hyperparameters. Since it does not requires other training process than just storing the training values, all the hyperparameters of a k-NN predictor operate in the prediction stage. In particular, the behavior of the k-NN predictor is ruled by:

- k: the number of neighbors;
- D: the number of dimension of the input space; it corresponds to the number of previous values used for the prediction;
- the weighting scheme: the law to assign the weights for the weighted averaging prediction;
- the norm of the input space.

The following values for the hyperparameters has been challenged:

$$k \in [1, 30] \tag{8}$$

$$D \in [1, 5] \tag{9}$$

Three weighting schemes have been tried: equal weight, weight proportional to the inverse of the neighborhood rank, and weight proportional to the inverse of the distance. Only the Euclidean norm has been used to compute the distance in the input space.

For the sake of comparison, the rules for generating the training, validation and test set will be the same used for the autoregressive models, described in III-A.

D. Prediction through Autoregressive Models

In order to train an autoregressive predictor, the hyperparameters that regulate the optimization procedure (i.e., the autoregression order, p, the moving average order, q, and the differencing order, d), have to be set to the proper value. Several values have been tried and their effectiveness have to be assessed through cross validation. In particular, the AR models have been challenged with $p \in \{1, \ldots, 63\}$; the ARMA models have been challenged with the combination of p and q for $p \in \{1, \ldots, 25\}$ and $q \in \{1, \ldots, 25\}$; and the ARIMA model have been challenged with the combination of the following values of p, d, and q:

$$p \in \{1, \ldots, 20\} \tag{10}$$

$$d \in \{1, \ldots, 3\} \tag{11}$$

$$q \in \{1, \ldots, 20\} \tag{12}$$

Since the training of the ARMA and ARIMA models requires consecutive training data, for avoiding of considering two separated periods of time for evaluating the validation and training error (which involves the risk of biased estimation due to the seasonality of the phenomenon under study), the prediction on the data not used to train the predictor has been carried out first and then the predicted period has been sampled for obtaining the validation and testing data.

TABLE I
TEST ERROR ACHIEVED BY THE PREDICTORS.

Predictor	Err(f)
Persistence	6.09
k-NN	3.17
AR	2.84
ARMA	2.86
ARIMA	2.85
SVM	2.89
SVM [8]	2.34

IV. RESULTS

The persistence and k-NN predictors, described in Section II, have been coded in Matlab, while for the autoregressive models (AR, ARMA, and ARIMA) their implementation in R have been used. Their performances have been evaluated using the prediction error, Err(f), described in (6). Since the persistence predictor configuration does not need any hyperparameters, the whole dataset described in Section III-A has been used to assess its performances. Instead, the training of the k-NN and the autoregressive models are regulated by a pool of hyperparameters. Hence, the training set has been used to estimate the model's parameters for each combination of the hyperparameters, then the validation dataset has been used to identify the best model (i.e., the one that achieved the lowest prediction error on the validation dataset) and the prediction error of that model on the testing set has been used to measure the performance of the class of the predictors.

The experiments have been carried out on a PC equipped with an Intel Core 2 Quad CPU at 2.5 GHz and 4 GB of RAM.

As reported in Table I, the persistence predictor has achieved an error $\text{Err}(f_{\text{P}}) = 6.09$, while the k-NN achieved an error $\text{Err}(f_{k\text{-NN}}) = 3.17$, for $D = 4$, $k = 18$, and using the inverted distance weighting scheme.

The AR model that scores the lower validation error has been trained using $p = 54$ and achieved $\text{Err}(f_{\text{AR}}) = 2.84$; the best ARMA model has been trained using $p = 24$ and $q = 19$, achieving $\text{Err}(f_{\text{ARMA}}) = 2.86$; the best ARIMA model, trained using $p = 16$, $d = 2$, and $q = 7$, achieved a testing error $\text{Err}(f_{\text{ARIMA}}) = 2.85$.

For the sake of comparison, the prediction error obtained using Support Vector Machines (SVM) as predictor in [8] is also reported. Although based on datasets belonging to the same source, the experiments are loosely comparable. In fact, it worth noting that the datasets used differ both for randomization and composition. In particular, in the present work the data used for training and for assessment (validation and testing) belong to two different time intervals (one year for training, the following year for assessment), while in [8] the dataset has been randomly equally partitioned without any consideration for the time span. Besides, also the information used are different, since in [8] the input variables were: the day of the year, the hour, the illuminance of the previous hour, and the average illuminance of the previous day and the previous week. Here, instead, only the illuminance values have been used.

In order to provide a proper comparison, some experiments have been run also with the SVM (using the LibCVM Toolkit [11]) on the datasets used for the present paper. The best performing model, as reported in Table I, achieved an error of $\text{Err}(f_{\text{SVM}}) = 2.89$; it is composed of 5325 support vectors with a Gaussian kernel ($\sigma = 13.8$), makes use of four previous values of the illuminance as input and has been trained using the following values for the hyperparameters: $\epsilon = 0.1$ and $C = 10$. This figure is more comparable to the error achieved with the autoregressive models (and, on the other hand, it can be seen as a measure of the influence of the dataset partitioning on the prediction ability).

From the comparison of the results reported in Table I, all the autoregressive models clearly outperform the persistence and the k-NN predictors and are slightly better than the SVM predictor. However, it worth noting that both the k-NN and the SVM make use of only four previous illuminance samples, while the the AR, ARMA, and ARIMA predictors base their prediction on respectively 54, 24, and 16 previous samples. On the other hand, the SVM is composed of 5325 support vectors, while the k-NN predictor stores the 9048 training samples: their prediction requires a high computational cost.

Figure 3 shows the distribution of the prediction error of AR, ARMA and ARIMA models. They are very similar and, in particular, the peaks in the error in Figs 3a–c are in the same position, which means that the corresponding values in the dataset are quite departed from the usual illuminance pattern.

While Fig. 3 shows the substantial equivalence of the predictors, Figs. 4 and 5, shows the robustness of the optimal pool of hyperparameters selected through cross validation. In these figures, the testing error for all the combinations of the hyperparameters are reported as circles.

In Fig. 4a, the performance of the AR models for each value of p is represented. It can be noted that the testing error decreases step-wise as the number of previous values used for the estimate increases, and, in particular, that local minima can be found after around multiples of 24 hours. This is intuitively explained by the 24 hours seasonality of the dataset. Although the error can possibly decrease for larger values of p, the search has been stopped because the long time consumed by the computation of the model parameters for a large values of p. Besides, for large values of p, numerical errors can arise in the optimization routine, which results in failure in the estimate of the model parameters. In fact, it can be noted that in the graph some testing errors are missing (e.g., for $p = 52$).

In Fig. 4b–c, the performance of the ARMA models with respect to the values of p and q are represented. The solid line connects the error for each value of p and q when, respectively, $q = 19$ and $p = 24$, i.e., the behavior of the error when one hyperparameter is fixed to the optimal value. It can be noticed that, although the the variance of the error tends to decrease with the increase of both p and q, the testing error do not exhibit a stable trend. This means that the solution is not very robust: for a different randomization of the validation and the testing sets, the best model could be characterized by different

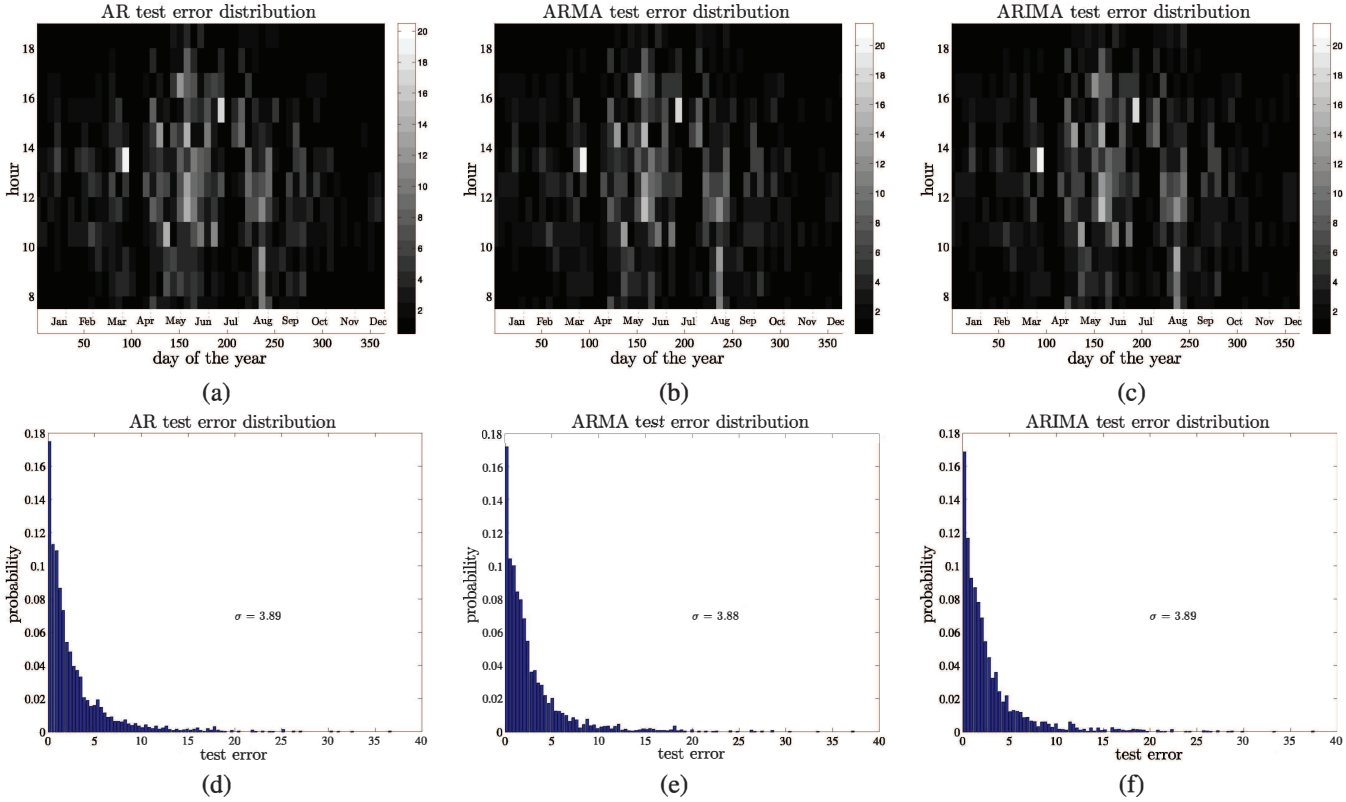


Fig. 3. Test error distribution. In panels (a)–(c), the test error produced by respectively the AR, the ARMA, and the ARIMA predictor are reported with respect to the day of the year and the hour. Since the test set does not include all the possible time combinations, the error have been reported averaging those of seven consecutive days. In panel (d)–(f), the estimated probability density function of the test error of the autoregressive models (which have standard deviations respectively of 3.89, 3.88, and 3.89). It can be noticed that both the distribution and the histogram of the considered predictors are very similar.

values of p and q. In particular, the error tends to decrease when p increases. Like for the AR model, larger values of p could be considered, but at large computational cost and numerical instability of the model parameters estimate.

In Fig. 5a–c, the performance of the ARIMA models with respect to the values of p, d, and q are represented. In each graph, the solid line represents the error as a function of the considered hyperparameter, when the other two are fixed at their optimal value. Although the error variance appears to be larger than in the ARMA model, the testing error becomes closer to the minimum for a large number of combinations of the parameters. For the p hyperparameter, it is clear from the graph that for value of p greater than 14, the performance of the predictor tends to improve (circles are more dense in the region close to the minimum). Larger values of p has not been considered in the experiment because the increase in computational time and the numerical instability of the optimization procedure. The differencing parameter, d, that characterize ARIMA with respect to the other models here considered, shows a slight improvement in the prediction perform ace when it is assumed to be equal to two instead of one. This can be an effect of the randomization and the best value can change for different choices of the assessment datasets.

V. Conclusions

In this paper, several autoregressive models (namely the AR, ARMA and ARIMA models) has been challenged with a problem of time series prediction.

All the models have been able to reach a similar testing error, with also a similar distribution. The analysis of the behavior of the error with respect to the hyperparameters of the models shows the tendency of increasing the performance as the numbers of previous values of the illuminance considered in the prediction approach the seasonality of the time series. However, since the illuminance has a seasonality of 24 hours, the number of parameters of the model can give rise to numerical error in the estimation procedure. Besides, a large number of parameters can slow the convergence of the estimation procedure.

Since a common approach in the literature is to prefer the simpler model, that is the model with less parameters, the ARIMA model can be considered the best fitting to the time series studied in this work.

Future works can take into consideration also the other information available in the original dataset, such as the day of the year, or the time of the prediction. This should provide an improvement in the performance.

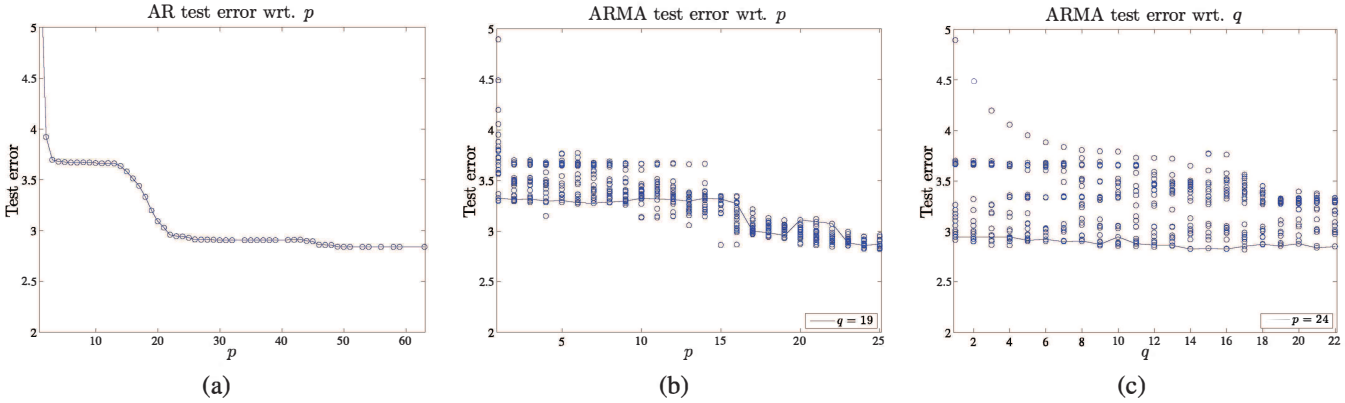


Fig. 4. AR, (a), and ARMA, (b)–(c), test error wrt. the hyperparameters.

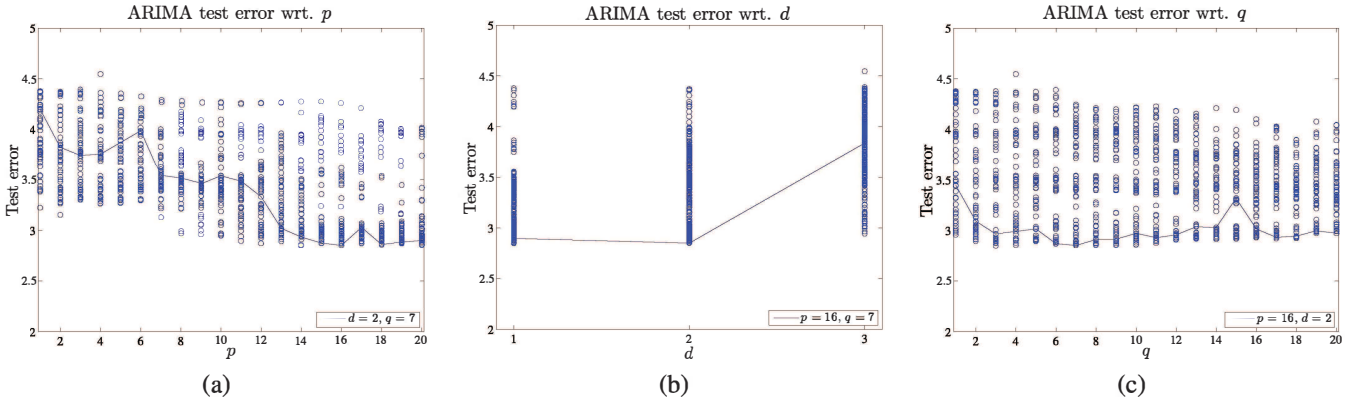


Fig. 5. ARIMA test error wrt. the hyperparameters.

REFERENCES

[1] M. Catelani, L. Ciani, L. Cristaldi, M. Faifer, M. Lazzaroni, and P. Rinaldi, "FMECA technique on photovoltaic module," in *IEEE International Instrumentation And Measurement Technology Conference (I2MTC 2011)*, May 2011, pp. 1717–1722.

[2] 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.

[3] T. Poli, L. P. Gattoni, D. Zappalà, and R. Gottardi, "Daylight measurement in Milan," in *Proc. of PLEA2006, Conf. on Passive and Low Energy Architecture*, 2006.

[4] A. Mellit, M. Benghanem, and S. Kalogirou, "An adaptive wavelet-network model for forecasting daily total solar-radiation," *Applied Energy*, vol. 83, no. 7, pp. 705–722, 2006.

[5] R. Perdomo, E. Banguero, and G. Gordillo, "Statistical modeling for global solar radiation forecasting in Bogotà," in *Photovoltaic Specialists Conference (PVSC), 2010 35th IEEE*, jun 2010, pp. 002 374–002 379.

[6] M. H. T.A. Raji, A.O. Boyo, "Analysis of global solar radiation data as time series data for some selected cities of western part of Nigeria," *International Journal of Advanced Renewable Energy Research*, vol. 1, no. 1, pp. 14–19, 2012.

[7] T. Poli, L. P. Gattoni, D. Zappalà, and R. Gottardi, "Daylight measurement in Milan," in *Clever Design, Affordable Comforta Challenge for Low Energy Architecture and Urban Planning*. Geneve - CH: Raphael Compagnon & Peter Haefeli and Willi Weber, 6 2006, pp. 429–433.

[8] 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.

[9] G. E. P. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*. Holden-Day, Incorporated, 1990.

[10] T. Cover and P. Hart, "Nearest neighbor pattern classification," *Information Theory, IEEE Trans. on*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[11] I. Tsang, J. Kwok, and P.-M. Cheung, "Core Vector Machines: Fast SVM training on very large data sets," *Journal of Machine Learning Research*, vol. 6, pp. 363–392, 2005.