Application of Fuzzy Logic to Forecast Hourly Solar Irradiation

Remus Boata 1, 2, *, Marius Paulescu 2

1 Astronomical Institute of the Romanian Academy, Timisoara Astronomical Observatory, A. Sever Sq. 1, 300210, Timisoara, Romania
2 Faculty of Physics, West University of Timisoara, V. Parvan 4, 300223, Timisoara, Romania; E-Mail: marius@physics.uvt.ro (M. P.)

* Author to whom correspondence should be addressed; E-Mail: rboata@physics.uvt.ro (R. B.); Tel.: +40 256 275092.

Received: 14 January 2014 / Accepted: 8 March 2014 / Published: 14 March 2014

Abstract: Accurate forecasting of solar resource is a key issue for a successful integration of the solar power plants into the grid. This paper proposes the fuzzy logic approach as alternative to classical statistics, aiming to forecast hourly global solar irradiation. Hourly clearness index, defined as the ratio of hourly global solar irradiation at the ground and at the top of the atmosphere, is the quantity directly processed by the fuzzy algorithms. It takes into account all random meteorological influences, being a measure of the atmospheric transparency. Thus, clearness index isolates the stochastic component of the solar irradiation data series. Four new autoregressive-fuzzy models are studied in this paper. The models are mainly differentiated by the number of the input variables and attributes. The results show that a model which includes seasonal terms is the most performant. The model can be applied in sites where measurement of hourly global solar irradiation is currently performed, only a re-estimation of the parameters being necessary. Overall results demonstrate that the fuzzy models have the strength to translate the information enclosed in past measurements into an actual prediction with acceptable accuracy.

Keywords: Solar irradiation; Forecasting; Fuzzy logic.

1. Introduction

The weight of solar electricity in the energy mix experienced an impressive augment in the last decade, and this trend is expected to continue. The European photovoltaic (PV) cumulative installed capacity has progressed from less than 1 GW in 2003 to over 30 GW in 2010 and 70 GW in 2012. Regarding the energy mix, at the end of 2012 PV covers 2.6% of the electricity demand and 5.2% of
the peak electricity demand in Europe [1]. To date, there are two challenges standing against the growing share of PV systems in the energy mix. The first challenge refers to the price of the electricity which is still high in comparison to that of power plants based on fossil fuels. Many efforts are spent all over the world to reduce the costs associated to solar electricity production. The second challenge stems from the intrinsic nature of solar energy which is stochastic fluctuating in time owing to irregular weather pattern. Therefore, in order to reduce the costs of integrating the solar plants in the existing power grid, forecasting the energy generated by the solar plants is a key issue. Currently, there are several ongoing research projects in Europe, testing different procedures for accurately forecast the PV output power. For example, COST Action ES1002 “Weather Intelligence for Renewable Energies” has the main objective to improve the forecast of the output power of solar and wind power plants [2].

There are two methods currently used to perform the forecast: (1) Statistical modeling of the PV plant output power time series (e.g. [3]) and (2) Modeling the output power, using as entries forecasts of radiometric quantities (solar irradiance/irradiation) and weather parameters (air temperature, cloud cover) (e.g. [4]).

The accuracy of forecasting the output power of a solar plant is closely related to the accuracy of forecasting of solar resource. Depending on the time horizon, different forecasting methods are considered [5]. In nowcasting (horizon time less than 3 hours) the forecasts are based on the extrapolations of real-time measurements [6-9]. The most popular methods are: Autoregressive Integrated Moving Average (ARIMA) and artificial neural network. A concise but outstanding review of these methods can be read in [10].

This paper is focused on the practice of hourly solar irradiation forecasting by using the fuzzy logic approach.

Fuzzy logic has been introduced in 1965 by Lotfy A. Zadeh [11]. Basically it replace the binary 0/1 logic with a multi-valued logic. Because the fuzzy approach is quite different from classical approaches, a short introduction is given in Appendix A. Additional references and examples can be found in many books (e.g. [12]).

Although fuzzy logic has been adopted as a standard method in various applications [13] its use is still incipient in modeling solar radiation at the Earth surface. There are several works dealing with the estimation of the solar irradiation via fuzzy approach [14-15]. Up to now, only a few attempts related to solar resource forecasting were published. In Ref. [16], our group reported a fuzzy model for forecasting daily global solar irradiation. Here we report an extension of this study by dropping the forecasting time horizon to one hour. Four auto-regressive fuzzy models constructed in an innovative manner are proposed and assessed.

The structure of the paper is as follows. Next section is devoted to the database description. In Sec. 3 the models construction is illustrated and their prediction accuracy is discussed. The main conclusions are gathered in Sec. 4.

2. Database

Measurements performed in the Romanian town of Timisoara (45°46′N, 21°25′E) are used in this study. Timisoara is placed at 85 m asl on the southeast edge of the Pannonian plain. Timisoara is characterized by a warm temperate climate, fully humid (Köppen climate classification Cfb - based on
the digital Köppen-Geiger world map on climate classification [17]) with warm summer, typical for the Pannonia Basin.

Global and diffuse solar irradiance recorded on the Solar Platform of the West University of Timisoara are used here [18]. Measurements are performed all day long at equal time intervals of 15 seconds. DeltaOHM LP PYRA 02 first class pyranometers which fully comply with ISO 9060 standards and meet the requirements defined by the World Meteorological Organization are employed. The sensors are integrated into an acquisition data system based on National Instruments PXI.

Hourly clearness index \( k_t \) is the quantity directly processed by the proposed fuzzy algorithm. This choice is motivated by the fact that \( k_t \) isolates the stochastic component of the solar irradiation [19]. In the present study, time series of hourly \( k_t \) was derived from solar irradiance measurements following the definition:

\[
\begin{align*}
  k_t &= \frac{H}{H_{ext}} \\
  \text{Eq. (1)}
\end{align*}
\]

In Eq. (1) \( H \) and \( H_{ext} \) denote the hourly global solar irradiation on the ground and at the top of the atmosphere, respectively. \( H \) was calculated by summing up the 240 solar irradiance measurements in every hour while \( H_{ext} \) was calculated by integrating the extraterrestrial radiation \( G_{ext} \) between sunrise and sunset times:

\[
H_{ext} = \frac{12}{\pi} \int_{\omega_1}^{\omega_2} G_{ext}(\omega) d\omega \quad \text{Eq. (2)}
\]

In Eq. (2) \( \omega \) is the hour angle and \( \omega_1 \) and \( \omega_2 \) delimit the hour interval. The factor \( 12/\pi \) assigns the unit: if \( G_{ext} \) is in W/m², then \( H_{ext} \) is in Wh/m². For a given day, the extraterrestrial radiation was calculated as function of the elevation angle \( h \): \( G_{ext}(\omega) = G_{sc} \epsilon \sin h(\omega) \), where \( G_{sc} = 1366.1 \) W/m² is the solar constant, \( \epsilon \) is the eccentricity correction factor that can be calculated with the equation of Spencer [20].

Another quantity used in this study is the relative sunshine \( \sigma \). It was evaluated following the pyranometer method. At first, the sunshine number \( \xi_t \), a Boolean quantity stating if the sun shining or not [21, 22], was calculated using the World Meteorological Organization sunshine criterion [23]. According to this, it is considered that the “sun is shining” at time \( t \) if direct solar irradiance exceeds 120 W/m², i.e.:

\[
\begin{align*}
  \xi_t &= \begin{cases} 
    1 & \text{if } \left( G_t - G_{d,t} \right) / \sin h > 120 \text{ W/m}^2 \\
    0 & \text{otherwise} 
  \end{cases} \\
  \text{Eq. (3)}
\end{align*}
\]

where \( G_t \) and \( G_{d,t} \) denote the global and diffuse solar irradiance at the moment \( t \) and \( h \) represents the sun’s elevation angle.

Then, the relative sunshine in a given time interval \( \Delta t \) was simply computed as the average values of \( \xi_t \) over \( \Delta t \):

\[
\sigma(\Delta t) = \bar{\xi}(\Delta t) \quad \text{Eq. (4)}
\]
The series of hourly relative sunshine was derived from measurements assuming in Eq. (4) \( \Delta t = 1 \) hour. In the following we denote \( \sigma = \sigma(\Delta t = 1 \text{hour}) \).

The database used in this study consists of 17520 values for each variable, \( k_t \) and \( \sigma \), corresponding to all hourly intervals of 2009 and 2010 (including zero night values). The data from 2009 were used to build the models while data from 2010 were used to test the models.

3. Results and Discussions

In this section the structure of the fuzzy models is presented in detail and their accuracy is assessed.

3.1. Models

Four autoregressive-fuzzy models are studied in this paper. The models are mainly differentiated by the number of the input variables and the number of attributes. Model #1 has only one variable at the input \( k_{t-1} \) (measured at time \( t-1 \)) characterized by 3 attributes. Model #2 extends the number of inputs increasing the order of auto-regressive terms to two, \( k_{t-1}, k_{t-2} \). Following the findings from [16] the variable \( k_{t-1} \) preserves the three attributes while the variable \( k_{t-2} \) is characterized by two attributes. The model #3 adds to model #1 an exogenous input, namely relative sunshine. The model #4 includes a seasonality term, \( k_{t-24} \) adjacent to \( k_{t-1} \).

The models #1, #2 and #3 were applied to the daily series (DS) of hourly \( k_t \). This series was obtained by the following procedure. First, for each day a sub-series of \( k_t \) has been prepared. Only hourly measurements associated to sun elevation angle greater than 5° were considered. Next, the DS series was build by stacking the sub-series prepared for each day in part. Since model #4 includes a seasonal term, it was applied to the whole day-night series (DNS) of \( k_t \) (i.e. including zero night values). The results presented in Sec. 3.3 shows that the model #4 is the most accurate. Thus, more details on the structure and the operation of this model are presented in the following.

Figure 1. The membership functions of the variables: (a) \( k_{t-1} \); (b) \( k_{t-24} \) and (c) \( k_t \).
3.2. Seasonal autoregressive fuzzy model

Model #4 is a seasonal autoregressive fuzzy model with two input variables \(kt_{t-1}, kt_{t-24}\) and one output variable \(kt_t\). The membership functions are specified in Fig. 1. The geometry of all membership functions was choosing triangular and always the peak of a triangle matches with the corresponding extremities of the adjacent membership functions. The membership functions of all variables, either in or out, reads:

\[
m_{t-v} = \begin{cases} 
\max \left(0, \frac{kt_{t-v} - a_i}{b_i - a_i}\right) & \text{if } kt_{t-v} < b_i \\
\max \left(0, 1 - \frac{kt_{t-v} - b_i}{c_i - b_i}\right) & \text{otherwise}
\end{cases}
\]  

(5)

where \(v = 0, 1\) or 24 specify the variable and the index \(i\) counts the attributes of a given variable. The parameters \(a_i, b_i\) and \(c_i\) have the meaning as is illustrated in Fig. 1a on the M1 attribute. The membership functions of the attributes S24 and H24 are saturated toward zero and infinite, respectively.

<table>
<thead>
<tr>
<th>(kt_{t-1})</th>
<th>(kt_{t-24})</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>M1</td>
</tr>
<tr>
<td>S24</td>
<td>S</td>
</tr>
<tr>
<td>H1</td>
<td>M</td>
</tr>
<tr>
<td>S24</td>
<td>M</td>
</tr>
<tr>
<td>H24</td>
<td>H</td>
</tr>
</tbody>
</table>

Table 1. Matrix of the system rule base of the model #4.

The mapping of the input to the output of the system, materialized in the rules-base, is listed in Table 1, as a matrix. There are 6 rules, each rule being an implication in sense of Eqs. (A3a,b).

The models #1, #2 and #3 have a similar structure with the model #4.

3.3. Models performance assessment

The models performance has been assessed with three statistical indicators: relative root mean squared error \((rRMSE)\), relative mean bias error \((rMBE)\) and relative mean absolute error \((rMAE)\), defined as follows

\[
rRMSE = \sqrt{\frac{\sum_{i=1}^{N} (kt_i^{estimated} - k_{i}^{measured})^2}{\sum_{i=1}^{N} k_{i}^{measured}}} \]  

(6)

\[
rMBE = \frac{\sum_{i=1}^{N} (kt_i^{estimated} - k_{i}^{measured})}{\sum_{i=1}^{N} k_{i}^{measured}} \]  

(7)

\[
rMAE = \frac{\sum_{i=1}^{N} |kt_i^{estimated} - k_{i}^{measured}|}{\sum_{i=1}^{N} k_{i}^{measured}} \]  

(8)

where \(N\) is the number of measurements taken into account.
Table 2. Statistical indicators of the models accuracy in the fitting period.

<table>
<thead>
<tr>
<th>Model</th>
<th>Fitting period</th>
<th>Testing period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rMBE</td>
<td>rMAE</td>
</tr>
<tr>
<td>#1</td>
<td>0.023</td>
<td>0.202</td>
</tr>
<tr>
<td>#2</td>
<td>0.023</td>
<td>0.202</td>
</tr>
<tr>
<td>#3</td>
<td>0.022</td>
<td>0.202</td>
</tr>
<tr>
<td>#4</td>
<td>0.013</td>
<td>0.182</td>
</tr>
<tr>
<td>ARIMA(1,0,1)×(1,0,1)24</td>
<td>-0.021</td>
<td>0.184</td>
</tr>
<tr>
<td>Persistence</td>
<td>-0.060</td>
<td>0.221</td>
</tr>
</tbody>
</table>

**Fitting period.** Table 2 shows the models ability to fit the data. Only forecasts for the daylight time were considered. It can be seen that model #4 is the best fit to the data, with an improvement in rRMSE over persistence of 25.2%. The persistence model assumes that the conditions at the time of the forecast will not change. The ability of the fuzzy model to trace the measured time series is well illustrated in Fig. 2, where the measured and the forecasted kt series with model #4 in 10 days (16 to 25 June 2009) are plotted. The DNS series was also modeled by a seasonal autoregressive integrated moving average sARIMA model (see Appendix B for an introduction), as the first competitor. The model ARIMA(1,0,1)×(1,0,1)24 was identified as the best, with rRMSE = 0.254 and rMBE = -0.021. It shows that there are no notable differences between the performance of the fuzzy logic model and the performance of the sARIMA model. However, a slightly advantage for the fuzzy model can be noted.

**Figure 2.** Measured and forecasted clearness index with model #4 in ten days of the fitting period (16 to 25 June 2009).

Testing period. All the models were tested against data measured in 2010. This is in line with rather common practice that a forecasting model is often developed with data from only one location. Then, only the model parameters are re-estimated when it is applied to another location.

The models performance in the testing period is also presented in Table 2. Again, model #4 registered the best performance with rRMSE = 0.283. This is a reasonable value taking into account that the statistical indicators were calculated over the entire year. The improvement in rRMSE over persistence is of 24.3%. A smaller improvement in rRMSE of only 4% is found compared to sARIMA model.

The ultimate goal of this work is the forecasting of hourly solar irradiation. Reikard [7] reported a comparison of different models (including ARIMA, UCM - unobserved components model, NN – neural networks and hybrid models) for forecasting mean hourly solar irradiance at different horizons
of time, against data measured at six stations. The results were assessed in terms of mean absolute percentage errors ($MAPE \equiv \frac{1}{N} \sum_{i=1}^{N} |\frac{k_{t_i}^{estimated} - k_{t_i}^{measured}}{k_{t_i}^{measured}}|$). For one hour forecast horizon, $MAPE$ reported in [7] falls between 19.6% and 75.4%. Comparing the forecasted solar irradiation time series generated by the model #4 with measurements we found $MAPE = 29.4\%$ in the fitting period and $MAPE = 37.0\%$ in the testing period. Therefore, our results are in good agreement with the results from [7].

**Table 3.** Statistical indicators of accuracy of the model #4 in each month of the year 2010.

<table>
<thead>
<tr>
<th>Month</th>
<th>Clearness index</th>
<th>Solar irradiation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rMBE</td>
<td>rMAE</td>
</tr>
<tr>
<td>January</td>
<td>0.089</td>
<td>0.2461</td>
</tr>
<tr>
<td>February</td>
<td>0.071</td>
<td>0.2357</td>
</tr>
<tr>
<td>March</td>
<td>0.007</td>
<td>0.2234</td>
</tr>
<tr>
<td>April</td>
<td>0.011</td>
<td>0.2114</td>
</tr>
<tr>
<td>May</td>
<td>0.058</td>
<td>0.2488</td>
</tr>
<tr>
<td>June</td>
<td>0.025</td>
<td>0.1963</td>
</tr>
<tr>
<td>July</td>
<td>0.036</td>
<td>0.1874</td>
</tr>
<tr>
<td>August</td>
<td>-0.009</td>
<td>0.1613</td>
</tr>
<tr>
<td>September</td>
<td>0.022</td>
<td>0.2042</td>
</tr>
<tr>
<td>October</td>
<td>-0.009</td>
<td>0.255</td>
</tr>
<tr>
<td>November</td>
<td>0.094</td>
<td>0.2119</td>
</tr>
<tr>
<td>December</td>
<td>0.189</td>
<td>0.3228</td>
</tr>
</tbody>
</table>

When the fuzzy models were tested monthly, the models accuracy was better in summer months than in winter months. As example, monthly statistical indicators of accuracy for model #4 are presented in Table 3. It can be seen that $rRMSE$ decreases from 32.4% in January to 19.6% in August and increases again to 41.6% in December, which indicates a seasonal dependence of the model accuracy. This is in accord with the results reported in Ref. [9], which demonstrated that the accuracy of the autoregressive models for sunshine number is primarily linked to the stability of the state of the sky, which, however, depends on the season.

**4. Conclusions**

In this paper four auto-regressive fuzzy models for forecasting hourly global solar irradiation are assessed. Being a measure of the stochastic component of solar irradiation, the hourly clearness index is the effective forecasted quantity. A seasonal fuzzy model which includes two auto-regressive terms of order 1 and 24 was found as the most performing. The proposed model exploits a very simple rules-base matrix. Since the data series used to build the model can be considered coming from an arbitrary environment, the procedure is general and it can be applied in any place where hourly global solar irradiation is currently measured, only a re-estimation of the parameters being necessary. The detailed
presentation of the model is intended to help the potential users to devise an appropriate fuzzy model for their own requirements.

The comparison with the traditional ARIMA model shows that the fuzzy logic approach is a competitive alternative for accurate forecasting short-term solar irradiation. Allowing intermediate values between the two binary options 0 and 1, the fuzzy sets theory can provide mathematics with the ability to capture uncertainties associated with natural phenomena. Thus, fuzzy logic may be regarded as an extension of the binary logic, which is successful in many applications, like computer science, but may lack the flexibility needed in other applications, like solar irradiation forecasting.

Further efforts will be devoted to the integration of some meteorological parameters in the fuzzy algorithm (related to the state of the sky, atmospheric pressure), aiming to increase the model accuracy.

Appendix

A. Fuzzy logic

The basis of Zadeh’s logic theory is the fuzzy set concept, which is defined as follows. If \( X \) is a collection of objects, the associated fuzzy set \( A \) is defined as:

\[
A \equiv \{ (x, m_A(x)) : x \in X \}
\]  

(A1)

where \( m_A(x) \) is the membership function showing the degree of affiliation of the element \( x \) to the fuzzy set \( A \). A physical variable is named linguistic variable and its values are linguistic values, called attributes, expressed by words. The membership function indicates the level of confidence with which an attribute characterizes a certain element \( x \in X \). The number of attributes of a linguistic variable and the shape of the membership functions depend on the application, being specified in a heuristic way.

In fuzzy sets theory, these operations are defined through the membership functions, e.g.:

Intersection: \( m_{A\cap B} = \min (m_A(x), m_B(x)), \quad \forall x \in X \)  

(A2a)

Reunion: \( m_{A\cup B} = \max (m_A(x), m_B(x)), \quad \forall x \in X \)  

(A2b)

Generally, a fuzzy logic model is a functional relation between two multidimensional spaces. The associative rules between different input and output fuzzy sets are expressed in the form:

IF \( \text{(premises)} \) THEN \( \text{(conclusions)} \)  

(A3a)

Every premise or conclusion consists of an expression like:

(\( \text{variable} \) IS \( \text{attribute} \))  

(A3b)

Thus, the information is carried out from the input to the output of a fuzzy model in three steps:

1. **Fuzzification.** A coding process in which each numerical input of a linguistic variable is changed into the membership function values of its attributes.
(2) **Inference.** A process consisting of two steps: (i) The computation of the degree in which a rule is fulfilled by the intersection of individual premises (Eq. A2a). (ii) Sometimes some rules drive to the same attribute of an output linguistic variable. For finding the confidence level of this conclusion, the individual degrees of fulfilling the rules driving to this conclusion are joined (Eq. A2b).

(3) **Defuzzification.** It is a decoding operation of the conclusion resulted from the inference process, on the purpose of providing the output crisp value. There are many defuzzification methods (see e.g. [12]); in this work, the following equation was applied:

\[
y_{\text{crisp}} = \sum_i c_i \int m_{y_i}(x)dx / \sum_i \int m_{y_i}(x)dx
\]

(A4)

where \( c_i \) is value of the variable \( x \) where \( m_{y_i}(x) \) reaches its peak and the integral \( \int m_{y_i}(x)dx \) represents the area delimited by \( m_{y_i}(x) \).

**B. Seasonal ARIMA models**

A seasonal ARIMA model ARIMA\((p,d,q)\times(P,D,Q)s\) includes the following elements: (1) \( AR(p) \) autoregressive part of order \( p \); (2) \( MA(q) \) moving average part of order \( q \); (3) \( I(d) \) non-seasonal differencing of order \( d \); (4) \( AR(P) \) seasonal autoregressive part of order \( P \); (5) \( MA(Q) \) seasonal moving average term of order \( Q \); (6) \( I_s(D) \) seasonal differencing of order \( D \). \( s \) is the period of seasonal pattern (in this case \( s = 24 \) hours). The general equation of ARIMA\((p,d,q)\times(P,D,Q)s\) model is:

\[
\left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right) \left(1 - \beta_1 B - \beta_2 B^2 - \ldots - \beta_q B^q\right) \left(1 - B^d\right)^{\delta} \left(1 - B^s\right)^D z_t = c + \\
\left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_P B^P\right) \left(1 - \psi_1 B - \psi_2 B^2 - \ldots - \psi_Q B^Q\right) \epsilon_t
\]

(B1)

where \( B \) is the backshift operator:

\[
Bz_t = z_{t-1}
\]

(B2)

and \( \epsilon_t \) is the estimated shock at time \( t \) (white noise).

The statistical software Statgraphics was used to fit an ARIMA\((p,d,q)\times(P,D,Q)s\) model on the DNS data series. The model ARIMA\((1,0,1)\times(1,0,1)24\) was selected; The model equation is:

\[
\tilde{z}_t = \phi_1 z_{t-1} + \beta_1 z_{t-24} - \beta_1 \phi_1 z_{t-25} - \theta_1 \epsilon_{t-1} - \psi_1 \epsilon_{t-24} + \psi_1 \theta_1 \epsilon_{t-25} + \epsilon_t
\]

(B3)

More details on the sARIMA models can be found in e.g. Ch. 6 from [24].

**Acknowledgments**

This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS – UEFISCDI, project number PN-II-ID-PCE-2011-3-0089A.
Conflicts of Interest

The authors declare no conflict of interest.

References and Notes


© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).