

SEASONAL MODELING OF HOURLY SOLAR IRRADIATION SERIES

M. PAULESCU¹, N. POP², N. STEFU¹, E. PAULESCU¹, R. BOATA³, D. CALINOIU²

¹West University of Timisoara, Physics Department, V. Parvan 4, 300223 Timisoara, Romania

²Politehnica University Timisoara, Fundamentals of Physics for Engineers Department, V. Parvan 2, 300223 Timisoara, Romania, E-mail: nicolina.pop@upt.ro

³Astronomical Institute of the Romanian Academy, Timisoara Astronomical Observatory, Axente Sever Square 1, 300210 Timisoara, Romania

Received November 30, 2016

Abstract. Increasing usage of solar energy and the integration of photovoltaic power plants into the grid demand accurate knowledge and forecasting of the solar energy resources. This paper reports on the seasonal autoregressive integrated moving average model (sARIMA) applied to the hourly solar irradiation data series. The standard sARIMA model, a new adaptive sARIMA model and a model based on multi-valued logic are compared in terms of forecasting accuracy. A simple procedure of including relevant information into sARIMA model is proposed, by expressing the model coefficients as linear functions of external predictors, such as various radiometric and meteorological quantities and quantifiers for the state of the sky. Global and diffuse solar irradiance measured on the Solar Platform of the West University of Timisoara in 2009 and 2010 are used for testing. The results show that the models performances depend on the stability of the radiative regime, and that only minimal improvement is obtained by using the adaptive method.

Key words: Solar irradiation, forecasting, ARIMA modeling.

1. INTRODUCTION

The main challenge for the power grid operators is to synchronize at every moment the electricity production with the demand. The equilibrium is constantly broken by the fluctuation of the demand. Nowadays this equilibrium is furthermore threatened by the increasing penetration of the renewable energy sources, such as wind and solar, whose inherent variability may induce significant energy fluctuation into the grid. Thus, forecasting the energy production of the wind and solar plants has become a crucial task for enabling operators to take control actions to balance the power grid.

The accuracy of the forecasting of the PV plant output power is mainly dictated by the accuracy of the forecasting of the solar energy resources [2]. The forecasting of the solar irradiance is currently done by statistical methods, among which the *Autoregressive Integrated Moving Average* (ARIMA) model [6] is the most popular. A major limitation standing against increasing the forecasting accuracy of solar energy is associated with the persistence property, *i.e.* the general tendency of the statistical models to extrapolate the current state in the future. To avoid this limitation different methods have been recently proposed, *e.g.* by increasing the relationship complexity [11] or based on artificial intelligence [8]. Only partial success was achieved, as shown in the quoted papers.

In this paper, we study the opportunity of using the seasonal ARIMA model (sARIMA) for modeling the hourly solar irradiation series. The standard sARIMA model and a new adaptive sARIMA model are compared to each other, and to a model based on multi-valued logic. The simplest way to produce a forecast, namely the persistence model, is assumed as reference.

2. DATABASE

Measurements performed in the Romanian town of Timisoara ($45^{\circ}46'N$, $21^{\circ}25'E$) 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 (Köppen climate classification Cfb), fully humid with warm summer, typical for the Pannonia Basin [7].

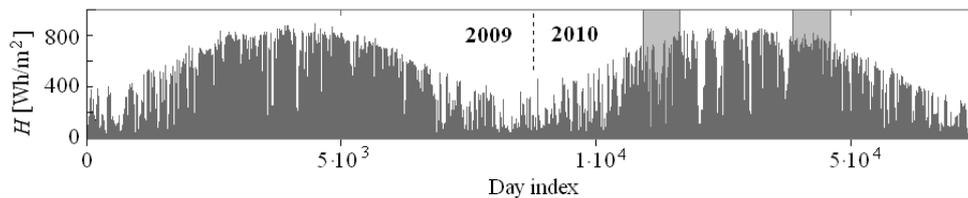


Fig. 1 – Hourly global solar irradiation during 2009–2010.

Global and diffuse solar irradiance recorded on the Solar Platform of the West University of Timisoara are used [10]. Measurements are performed all day long at equal time intervals of 15 seconds. DeltaOHM LP PYRA02 first class pyranometers, which fully comply with ISO 9060 standards and meet the requirements defined by the World Meteorological Organization, are employed. The database used in this study consists of 17520 values for each variable, corresponding to all the hourly intervals of years 2009 and 2010 (including zero night values). The data from 2009 (further denoted HF series) were used to build

the models while data from April and August 2010 were used to test the models (Fig. 1).

3. MODELS DESCRIPTION

Four different models are tested in this paper: (1) Persistence, (2) *Seasonal Autoregressive Integrated Moving Average* – sARIMA [6], (3) Autoregressive fuzzy logic [3] and (4) Adaptive ARIMA model (new model developed within this work). The four models, very different by nature and complexity, were implemented in order to highlight their potential advantages and disadvantages. These models are summarized in the following. Section 4.2 presents the proposed model in detail.

3.1. PERSISTENCE

The persistence method is the simplest way to produce a forecast. It assumes that the state of a system at a given moment $t-1$, given by the observation z_{t-1} , will not change in the short future, at a next moment t , *i.e.*:

$$z_t = z_{t-1}. \quad (1)$$

The extent to which a model improves the forecasting accuracy over the persistence, may be regarded a measure of the model quality.

3.2. sARIMA MODEL

An ARIMA model expresses the observation z_t at the time t as a linear function of previous observations, a current error term and a linear combination of previous error terms. A seasonal ARIMA model is usually denoted $ARIMA(p,d,q) \times (P,D,Q)_s$ and contains the following terms: (1) $AR(p)$ – the non-seasonal autoregressive term of order p ; (2) $I(d)$ – the non-seasonal differencing of order d ; (3) $MA(q)$ – the non-seasonal moving average term of order q ; (4) $AR_s(P)$ – the seasonal autoregressive term of order P ; (5) $I(D)$ – the seasonal differencing of order D ; (6) $MA_s(Q)$ – the seasonal moving average term of order Q . The general equation of the sARIMA model is written:

$$\underbrace{(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)}_{AR(p)} \underbrace{(1 - \beta_1 B^s - \beta_2 B^{2s} - \dots - \beta_P B^{Ps})}_{AR_s(P)} \underbrace{(1 - B)^d}_{I(d)} \underbrace{(1 - B^s)^D}_{I(D)} z_t = c + \underbrace{(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)}_{MA(q)} \underbrace{(1 - \psi_1 B^s - \psi_2 B^{2s} - \dots - \psi_Q B^{Qs})}_{MA_s(Q)} a_t \quad (2)$$

where B is the backshift operator defined as:

$$Bz_t = z_{t-1} \quad (3)$$

a_t is a random error at time t usually assumed with normal distribution, zero mean and standard deviation σ_a (white noise). Sometimes an adjustment constant c is included in Eq. (2). The coefficients ϕ , θ , β and ψ as well as the standard deviation of the white noise are obtained using the maximum likelihood method [6].

While Eq. (2) looks formidable, the ARIMA models used in practice are much simpler cases. The equation of a particular model can be easily deduced from Eqs. (2–3). For example, the equation of the most referred model in this paper, sARIMA(1,0,0)×(0,1,1)₂₄, is:

$$z_t = z_{t-24} + \phi_1(z_{t-1} - z_{t-25}) - \psi_1 a_{t-24} + a_t. \quad (4)$$

Basically, Eq. (4) represents the starting point of the model proposed in this paper.

3.3. AUTOREGRESSIVE FUZZY LOGIC

Basically, fuzzy logic [13] replaces the binary 0/1 logic with a multi-valued logic. A brief introduction can be read in [12]. As an alternative to classical statistics, a seasonal autoregressive fuzzy model reported in [3] is also tested in this paper. The model includes two auto-regressive terms of order 1 and 24. 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. Clearness index takes into account all random meteorological influences, thus isolating the stochastic component of the solar irradiation data series. The comparison with the traditional sARIMA performed in [4] shows that the fuzzy logic approach is a competitive alternative for accurate forecasting short-term solar irradiation.

3.4. ADAPTIVE sARIMA MODEL

This is a new model proposed by this paper. It was inspired by the results reported by our group in [9] on the accuracy of the forecasting with the random walk model, applied to seasonally adjusted hourly global solar irradiation time series. In that paper we demonstrated that an adaptive procedure, in which the seasonal indices are re-estimated every day, may lead to an amazing improvement in rRMSE of over 100% in comparison to the standard procedure (in which the seasonal indices are estimated only once, and then used for all the future forecastings).

The model proposed here is developed based on two hypotheses: (1) the information from the recent past (encapsulated in the model coefficients) may

allow for the subtle characterization of the time series dynamics and (2) the information from the far past may alter the model ability to handle the actual dynamics of the time series. Therefore, two facets of the sARIMA model applied to the hourly solar irradiation time series are explored: (1) the possibility to improve the overall model accuracy by estimating its coefficients in a window sliding in time and (2) the possibility to relate the sARIMA model coefficients with exogenous predictors which may encapsulate relevant information about the series dynamics. Both approaches are extensively discussed in Sec. 4.2.

4. RESULTS AND DISCUSSIONS

4.1. THE STANDARD sARIMA MODEL

Table 1 compares the results of fitting different sARIMA models to HF data series (see [1] for details on the fitting procedure). The highest order of any term in Eq. (2) (*i.e.* p , P q or Q) was limited to 2. As expected, the more complex a model, the better its performance. However, the performance of different sARIMA models did not differ appreciably. Between the model $ARIMA(2,1,1) \times (2,0,2)_{24}$ ranked first and the model $ARIMA(1,0,0) \times (0,1,1)_{24}$ there is a difference of 0.57% in terms of *RMSE* and 0.12% in terms of *Akaike Information Criterion* (AIC). Following the parsimony principle the model with the lowest number of coefficients $ARIMA(1,0,0) \times (0,1,1)_{24}$ was preferred. The model is given by Eq. (4), with the coefficients $\phi_1 = 0.804$ and $\psi_1 = 0.909$.

Another motivation for choosing the $ARIMA(1,0,0) \times (0,1,1)_{24}$ model is presented in Fig. 2. The procedure used to obtain the results presented in this figure is explained in the following. Different sARIMA models were fitted on sets of 720 data (corresponding to a *window* of 30 days) and tested against the next 120 data (corresponding to 5 days). This data *window* was then slid forward on the database, 24 hourly data at a time, over the entire HF series, and the fitting process was performed each time. The first sample in the window from 1 January to 30 January was indexed to 1, the sample from 2 January to 5 February was indexed by 2, and so on. The last sample from 27 November to 26 December was indexed by 331. Again, the highest order of any term in Eq. (2) was assumed equal to maxim 2. The models fitted in a given window were ranked according to AIC. Figure 2 shows the sARIMA models hierarchy according to the frequency of occupying the first position (lowest AIC) in all windows. The model $ARIMA(1,0,0) \times (0,1,1)_{24}$ was ranked on first position the most often.

Table 1
Statistical indicators of fitting different sARIMA models on HF series

Model	<i>MBE</i> [Wh/m ²]	<i>RMSE</i> [Wh/m ²]	<i>AIC</i>
ARIMA(2,1,1)×(2,0,2)24	−0.09346	50.78	7.8569
ARIMA(2,1,1)×(2,0,1)24	−0.16798	50.80	7.8571
ARIMA(2,1,1)×(1,1,1)24	−0.06124	50.81	7.8574
ARIMA(1,0,0)×(0,1,1)24	−0.04913	51.07	7.8666

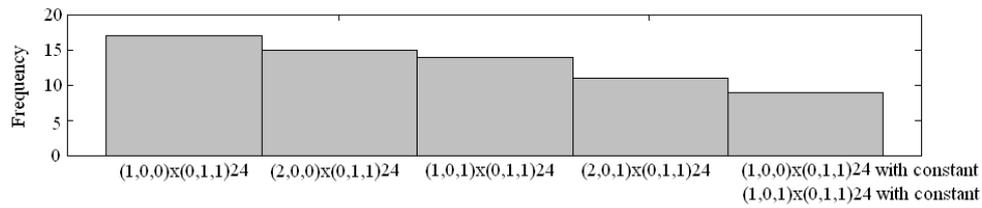


Fig. 2 – The frequency each model was ranked first when fitted to a 720 data window from HF series sliding in time.

4.2. ADAPTIVE sARIMA MODEL

Coefficients dynamic. In this section the dependence of the ARIMA(1,0,0)×(0,1,1)24 coefficients on the properties of the sample used for their fitting is studied. For this, the coefficients ϕ and ψ were estimated by fitting the model in a *sliding window* over HF series as explained in Sec. 4.1. Figure 3 shows the dependence of the coefficients ϕ and ψ on the sample index $i = 1 \dots 331$. The mean value of the coefficients ϕ_i ($\bar{\phi}_i = 0.800$) equals with good approximation the value of the coefficient ϕ obtained by fitting the model over the entire HF series, $\phi = 0.804$. It can be noticed, although, that ϕ_i displays significant variation around the mean estimated value $\bar{\phi}_i = 0.800$ (min = 0.660, max = 0.806 and stdev = 0.059). A seasonal behavior can also be observed, the values of ϕ_i during summertime being smaller than the average, while the values during winter are larger than the average. Coefficient ψ presents a different behavior, a small variation around the estimated average $\bar{\psi} = 0.909$ (min = 0.777, max = 0.977 and stdev = 0.033). Therefore, the seasonal term will be approximated to a constant in the following, $\psi_i = \psi$, and only the dynamics of the first order autoregressive term ϕ_i will be analyzed.

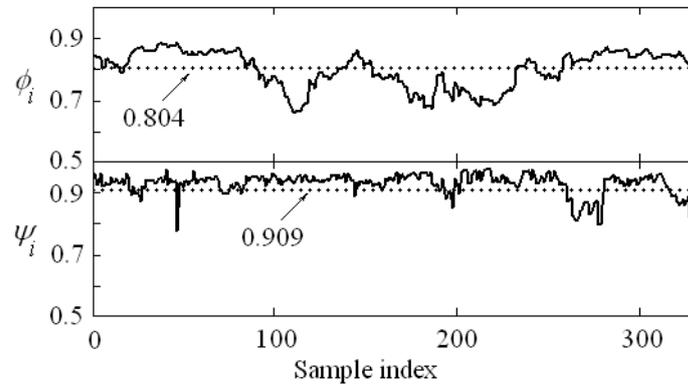


Fig. 3 – Dependence of the ARIMA(1,0,0)×(0,1,1)24 coefficients on the sample index i . The dotted lines indicate the values of the coefficients fitted on the entire HF series.

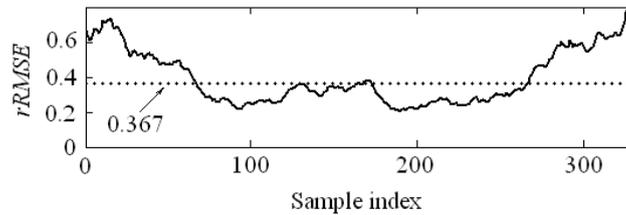


Fig. 4 – Accuracy of fitting the ARIMA(1,0,0) × (0,1,1)24 model to the data from the sliding window. The dotted line indicates the level of RMSE for the ARIMA(1,0,0) × (0,1,1)24 model fitted on the entire HF series.

A seasonal behavior similar to that of coefficient Φ_i can be noticed in the accuracy of the fit with the ARIMA(1,0,0) × (0,1,1)24 model on the data within the sliding window (Fig. 5). The accuracy is much better in summertime ($0.2 < rRMSE < 0.35$) than in wintertime ($0.35 < rRMSE$). The shape of $rRMSE$ in Fig. 5 is no surprise, although, because there are many consecutive sunny stable days during summer, while the days during winter are mostly cloudy. Cloud transparency is perhaps one of the most difficult radiometric quantities to forecast.

4.3. SENSITIVITY ANALYSIS

The observations in the previous section lead to the idea of introducing some exogenous predictors into the model, in order to increase the model performance. The classical procedure would be to combine the sARIMA model with a linear regression model, obtaining aso-called generalized additive model [5].

A simpler procedure of including relevant information into sARIMA model is proposed in this paper, namely by expressing the model coefficients as linear

functions of external predictors. By using the example $ARIMA(1,0,0) \times (0,1,1)_{24}$, this section is dedicated to the study of the sensitivity of coefficient Φ_i to different radiometric and meteorological quantities, and some quantifiers for the state of the sky. The general equation considered for Φ_i is:

$$\phi_i = \beta_0 + \sum_{k=1}^N \beta_k X_k, \quad (5)$$

where β_k , $k = 0 \dots N$ are coefficients to be estimated, X_k , $k = 1 \dots N$ are the predictors and N is the number of predictors.

The predictors considered here are the quantities listed below, as averages of the daily values over the time length of the sliding window (30 days in this case):

h – sun elevation angle computed at the middle of the day, \bar{H} – global solar irradiation, \bar{H}_d – diffuse solar irradiation, \bar{H}_b – beam solar irradiation, \bar{T} – diurnal average of the air temperature, \bar{T}_{24} – average of the air temperature over 24 hours, T_{min} minimum air temperature, T_{max} – maximum air temperature, \bar{H}_u – diurnal average of the relative humidity, \bar{H}_{24} – average of the relative humidity over 24 hours, \bar{v} – diurnal average of the wind speed, \bar{v}_{24} average of the wind speed over 24 hours, σ – relative sunshine, $\bar{\zeta}$ – average of the sunshine stability number, $\Delta\bar{t}$ – average of daily amplitude of air temperature.

First, the procedure for selecting the appropriate predictors of the regression model for ϕ was designed. The procedure considers all possible linear regressions involving different combinations of maximum $N = 10$ predictors and compares the results based on the adjusted R-Squared. A *parsimonious* model is desired, *i.e.*, a model including as few variables as possible, without altering the predictive capability of the model. Figure 5 shows the equations with the highest adjusted R-Squared values. The best adjusted R-squared increases noticeably until $N = 4$. Table 2 summarizes the fitted models, sorted in decreasing order of the adjusted R-squared. Two models were selected for further analysis: model #1 and model #6. The equations of these models are:

Model #1:

$$\begin{aligned} \phi = & 0.497354 + 0.0236182h + 0.00973719\bar{H} - 0.00998174\bar{H}_d - \\ & - 0.0099149\bar{H}_b + 0.213843\bar{T} - 0.15584\bar{T}_{24} - 0.0681037T_{max} + \\ & + 0.0377789\bar{v}_{24} + 0.320917\sigma - 8.44435\bar{\zeta} \end{aligned} \quad (6)$$

The R-Squared indicates that model #1 explains 86.46% of the variability of the ϕ coefficients series.

Model #6:

$$\begin{aligned} \phi = & 0.681393 + 0.00709257h - 0.00011235\bar{H} - 0.164455\bar{v} + \\ & + 0.273817\bar{v}_{24} + 0.231262\sigma \end{aligned} \quad (7)$$

The R-Squared indicates that model #6 explains 78.7% of the variability of the ϕ coefficients series.

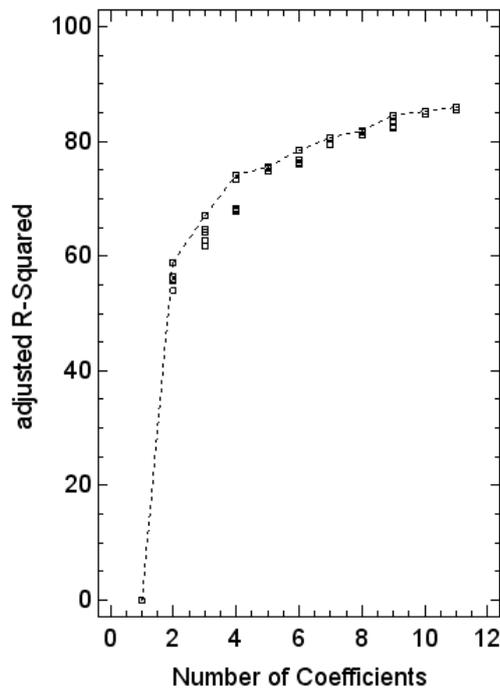


Fig. 5 – Adjusted R-squared as function of coefficients number for different linear regression of ϕ .

Table 2

Models with the highest adjusted R-Squared

Model	N	Dependent variables														R-Squared	Adjusted R-squared	
		h	\bar{H}	\bar{H}_d	\bar{H}_b	\bar{T}	\bar{T}_{24}	T_{min}	T_{max}	\bar{H}	\bar{H}_{24}	\bar{v}	\bar{v}_{24}	σ	ζ			$\Delta\bar{T}$
#1	10	x	x	x	x	x	x	x	x				x	x	x		86.462	86.039
#2	9	x	x		x	x	x	x				x	x	x	x		85.556	85.151
#3	8	x	x		x	x	x	x				x	x	x			84.981	84.608
#4	7	x	x		x	x	x	x				x	x				82.187	81.801
#5	6	x	x		x	x	x	x						x			81.041	80.690
#6	5	x	x									x	x	x			78.725	78.397

Table 3

Statistical indicators of the accuracy of the models tested in this study

Model	Data set	April				August			
		<i>rRMSE</i>	<i>rMBE</i>	<i>MAPE</i> *	<i>IMP</i> [%]	<i>rRMSE</i>	<i>rMBE</i>	<i>MAPE</i>	<i>IMP</i> [%]
Persistence	D&N	0.582	0.000	–	–	0.458	0.000	–	–
	D	0.391	–0.026	0.489	–	0.314	–0.015	0.476	–
<u>Adaptive sARIMA</u>									
Sliding window	D&N	0.495	–0.035	–	14.9	0.277	0.005	–	39.5
	D	0.340	–0.043	0.448	13.0	0.197	0.001	0.234	37.2
<i>N</i> = 5	D&N	0.493	–0.031	–	15.2	0.272	0.007	–	40.6
	D	0.338	–0.040	0.450	13.5	0.203	0.006	0.263	35.3
<i>N</i> = 10	D&N	0.496	–0.025	–	14.7	0.271	0.004	–	40.8
	D	0.340	–0.034	0.459	13.1	0.203	0.003	0.265	35.3
ARIMA(1,0,0) × (0,1,1) ₂₄	D&N	0.494	–0.029	–	15.1	0.277	0.005	–	39.5
	D	0.339	–0.039	0.451	13.2	0.197	0.001	0.234	37.2
Fuzzy autoregressive	D&N	0.454	–0.028	–	21.9	0.281	–0.034	–	38.6
	D	0.270	–0.009	0.432	30.9	0.185	–0.020	0.258	41.0

4.4. EVALUATION OF THE MODELS ACCURACY

The different models studied in this paper were tested against data measured in two months of 2010: April and August. As Fig. 1 shows, these months are characterized by different solar radiative regimes: in April the solar radiative regime is rather unstable, while in August the solar radiative regime is mostly stable. Two series of data are considered in each of the two months: (1) the full series including the zero night values, denoted D&N series and (2) the series of daily data, denoted D, obtaining by excluding the zero night values and considering only the hourly values with the sun elevation angle greater than ten degrees.

The models accuracy was evaluated using three statistical indicators: normalized root mean square error (*rRMSE*), normalized mean bias error (*rMBE*) and mean absolute percentage error (*MAPE*). The performances of the models were compared in terms of *IMP* – improvement in *RMSE*. The results of the models testing are gathered in Table 3. First, major differences between the results of the tests performed on the two sets of data can be seen. The models perform better on the *D* set ($0.185 < rRMSE < 0.391$) than on the D&N set ($0.271 < rRMSE < 0.582$). A second observation is that, generally, the models perform better in stable periods ($0.185 < rRMSE < 0.458$ in August) than in unstable periods ($0.270 < rRMSE < 0.582$). By adding exogenous predictors to the adaptive sARIMA(1,0,0) × (0,1,1)₂₄ the forecasting accuracy remains almost unchanged, regardless of the stability of the solar radiative regime.

The best performing model is the fuzzy autoregressive model, with the highest improvement in *RMSE* over the persistence model. In three of the four testing cases, the fuzzy model ranks first: (1) April, dataset D&N, *IMP* = 21.9%; (2) April, dataset D, *IMP* = 30.9%; (3) August, dataset D, *IMP* = 41.0%. Only in

one case, the adaptive sARIMA takes the lead: August, dataset D&N $IMP = 40.8\%$. Figure 6 compares the solar irradiation values forecasted with this model with the measured ones in the first 15 days of August 2010. There is a good similarity between the sequential features of the measured time series and the forecasted time series.

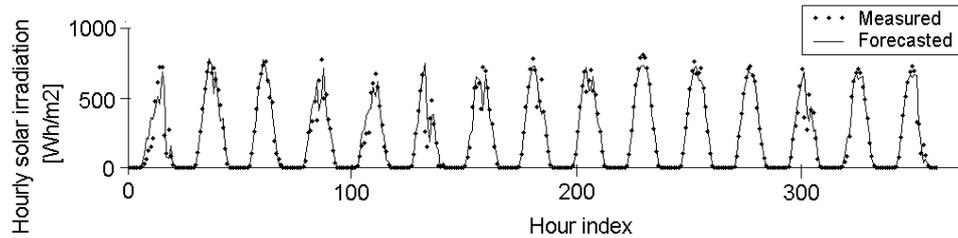


Fig. 6 – Measured and forecasted hourly solar irradiation in the first 15 days of August 2010.

The sliding window effect is very small. No substantial improvement in the accuracy of the prediction is noted. This is an unexpected result, taking into account that high improvement was obtained in forecasting the hourly solar irradiation using non-seasonal models, by using a sliding window for the seasonal decomposition [9]. A possible explanation may stem from the intrinsic nature of the sARIMA models. The seasonality always connects the forecasts with the very recent past. However, further investigations are required.

5. CONCLUSIONS

The paper focuses on forecasting the hourly solar irradiation series using different seasonal ARIMA models. The study was conducted on data measured on the Solar Platform of the West University of Timisoara, Romania. The performance of the standard seasonal ARIMA model, $sARIMA(1,0,0) \times (0,1,1)_{24}$, was compared with the performance of the proposed sARIMA model. Overall results show that the proposed adaptive procedure for fitting the sARIMA parameters (in which the model parameters are refitted daily) adds a minimal benefit to the classical procedure (in which the model parameters are fitted only once on the entire database). An autoregressive fuzzy model has proved better performances. It was shown that the fuzzy logic modeling is a promising approach for accurate forecasting of short-term solar irradiation. The models performances depend on the stability of the radiative regime (increasing with decreasing instability). Further studies are required to enhance the forecasting accuracy when large fluctuations are present in the hourly global solar irradiation time series.

REFERENCES

1. V. Badescu, M. Paulescu, *Meteorology and Atmospheric Phys.* **112**, 139-154 (2011).
2. P. Bacher, H. Madsen, H. A. Nielsen, *Solar Energy* **83**, 1772-1783 (2009).
3. R. Boata, M. Paulescu, Proc. of the 1st International e-Conference on Energies, 14-31 March 2014: <http://sciforum.net/conference/ece> (2014).
4. R. Boata, N. Pop, *Romanian Journal of Physics* **60**, 3-4, 593-602 (2015).
5. M. Brabec, M. Paulescu, V. Badescu, *Solar Energy* **101**, 272-282 (2014).
6. GEP Box, G. M. Jenkins, *Time series analysis. Forecasting and Control*, Holden-Day, San Francisco, 1970.
7. M. Kottek, J. Grieser, C. Beck, B. Rudolf, F. Rubel, *Meteorologische Zeitschrift* **15**, 259-263 (2006).
8. P. Lauret, C. Voyant, T. Soubdhan, M. David, P. Poggi, *Solar Energy* **112**, 446-457 (2015).
9. A. Pacurar, N. Stefu, O. Mares, E. Paulescu, D. Calinoiu, N. Pop, R. Boata, P. Gravila, M. Paulescu, *Journal of Renewable and Sustainable Energy* **5**, Article Number: 063140 (2013).
10. Solar Platform, Solar Platform of the West University of Timisoara, Romania: <http://solar.physics.uvt.ro/srms> (2016).
11. T. Soubdhan, J. Ndong, H. Ould-Baba, M.T. Do, *Solar Energy* **131**, 246-259 (2016).
12. E. Tulcan-Paulescu, M. Paulescu, *Theoretical and Applied Climatology* **91**, 181-192 (2008).
13. L.A. Zadeh, *Information and Control* **8**, 338-353 (1965).