Article type: Conf.-Conference paper (CASAMANSUN2018)

ARMA model for short-term forecasting of solar potential: application to a horizontal surface on Dakar site

A. Mbaye[1], M.L. Ndiaye[1], D.M. Ndione[2], M. Sylla[1], M.C. Aidara[1], M. Diaw[2,1], V. Traoré[2], A. Ndiaye[1] and P.A.S. Ndiaye[1]

[1]International Center for Training and Research in Solar Energy /ESP/UCAD, BP 5085, Dakar-Fann, Senegal.
[2] Hydraulic and Fluids Mechanics Laboratory (HFML), Department of Physics, Faculty of Sciences and technology, Cheikh Anta Diop University (UCAD), BP 5005, Dakar-Fann, Senegal.

Corresponding author : amy1.mbaye@ucad.edu.sn

Abstract: This paper presents a model for short-term forecasting of solar potential on a horizontal surface. This study is carried out in to the context of valuing of energy production from photovoltaic solar sources in the Sahelian zone. In this study, Autoregressive Moving Average (ARMA) process is applied to predict global solar potential upon 24 hours ahead. The ARMA (p, q) is based on finding optimum parameters p and q to better fit considered variable (sunshine). Data used for the model calibrating are measured at the station of Ecole Supérieure Polytechnique of Dakar. Records are hourly and range from October 2016 to September 2017. The choice of this model is justified by its robustness and its applicability on several scales through the world. Simulation is done using the RStudio software. The Akaike information criterion shows that ARMA (29, 0) gives the best representation of the data. We then applied a white noise test to validate the process. It confirms that the noise is of white type with zero mean, variance of 1.252 and P-value of about 26% for a significant level of 5%. Verification of the model is done by analyzing some statistical performance criteria such the RMSE =0.629 (root mean squared error), the R^2 = 0.963 (Coefficient of determination), the MAE=0.528 (Mean Absolut Error) and the MBE=0.012 (Mean Bias Error). Statistics criteria show that the ARMA (29,0) is reliable;then, can help to improve planning of photovoltaic solar power plants production in the Sahelian zone.

Introduction

Senegal has significant potential for enhancing solar resource (PV and Thermal) with 3000 hours of sunshine per year and an average of global irradiation about 5.8 kWh / m2 / day. Solar photovoltaic market is growing in the world in general particularly in Africa. Currently there is a proliferation of photovoltaic solar power plants in Senegal. Integration of these intermittent sources on the SENELEC electricity system causes many problems on planning and maintaining the consumption/production balance. Such difficulties are stochastic behaviour of the renewable energy sources related. To overcome such difficulties, forecasting can be considered. Thus, prediction of solar potential can be a substantial contribution in renewable energy production. It allows to evaluate the intermittent energies rate on the distribution network, in order to mitigate unbalances risks. Indeed, forecasting production of intermittent energies allows anticipating on the availability of the production sources and facilitating the network management. It is inspired by the stochastic behaviour of the atmospheric situations that are deterministic factor for the solar potential the earth surface. Considering above uncertainty sources, operators in solar power plants fields need appropriate tools to predict the energy production and availability.

In the literature, stochastic prediction of the solar potential on horizontal surface can be done either at long-term scale [1, 2] or short-term scale [4, 3]. Thus, this study has been done to contribute to the search for reliable prediction methodologies of photovoltaic power. Four categories of models exist in the literature [5, 6, 4]. Among these models, we have linear, nonlinear models, probabilistic models and models of naive types. An Autoregressive Moving Average (linear model) process is used to forecast solar potential at Dakar. We have chosen to work with the model Autoregressive Moving average because of its robustness and applicability on several sites in the world on different forecast horizons [7, 28, 9, 10, 11]. ARMA model is based on researching optimal fit parameters p and q for better representation of considered time series (solar radiation). Our contribution in this field of study focuses on implementating reliable solar potential prediction model that able to predict the SENELEC's capacity of production for 24 hours forecast horizon and can be used as a decision support tool for planning and managing photovoltaic solar power plants in Senegal and surrounding. Section2 describes the methodology and tools used to reach our target, section 3 presents results of issues form forecasting model and interpretation. The last section is devoted to the conclusion and perspectives.

Materials and Methods

1. Presentation of study area

Dakar is located in the extreme west of the Cape Verde peninsula, at the edge of the Atlantic Sea, between latitude 14° 75 N and longitude -17.33 ° W with an average altitude of about 16m (Figure1). It contains some meteorological stations including one of measuring of the International Center for Training and Research in Solar Energy (CIFRES) of the Polytechnic High School of Cheikh Anta Diop University (UCAD) of Dakar. This station measures global, direct and diffuse solar radiation, ambient temperature, relative humidity, rainfall, wind speed and wind direction. Dakar covers an area of 550 km². Its climate is of tropical type characterized by alternation of two seasons: a very long and dry season on one hand (9 months) and a short and rainy season on the other (3months). The dry season is characterized by the predominance of the Harmattan dry wind that is neither hot nor cold; the rainy season is characterized by the hot and moist Monsoon wind. Temperatures vary according to the seasons. A low range of diurnal temperature caused by the maritime wind exchanges composed of a steady wet and fresh wind blowing especially in the dry season explaining the leniency of temperatures at Dakar in comparison to that inside the country. Dakar has a relatively mild climate, due to prerogatives of its geographical position and oceanic influences determining special characteristics of this region. Rainfall in this area usually occurs between July and October. The rainy month of this area is August. Annual rainfall is estimated about 500 mm. Within variation of the air moisture average depends on the temperature and hygrometric characteristics of the one. Average of the relative humidity rate in the site is close to 95%, especially during the rainy season between August and September. Lower values of relative humidity are observed during the dry season in which the time evaporation remains significant.

In this work, we used solar radiation data from CIFRES station. The data series runs from October 1st, 2016 to September 30th, 2017 and measurements are made every one hours. A Campbell Scientific model CR800 datalogger is used.
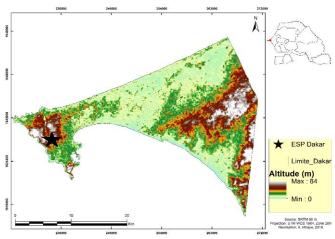

Figure1: Location of Dakar in Senegal area

2. Autoregressive moving average process (ARMA)

In the literature, some authors [9, 28] argue that autoregressive moving average process (ARMA) seems adequate for fitting solar radiation time series. Moreover, the ARMA process can be used to forecast time series. Running such processes is based on observations in order to determine adequate stochastic model that fit the studied variable. In time series analysis, Wold proposed in 1938 a decomposition of time series into components referred to as deterministic component (trend and periodic component) and stochastic one [18]. Deterministic component can be computed through known physical laws and steady mathematical schemes. The stochastic component can't be exactly determined by known physical and mathematical lows due to its random feature, then are modelled through stochastic processes. Stochastic processes have also been referred to as Yule processes (infinite linear combination process that cannot be expected). In 1954, Wold defined from a random process a linear model composed of an autoregressive process noticed by AR and a moving average one noticed by MA [16]. Combination of both processes gives the ARMA process. An AR process fits a time series through a linear function of its past values while the MA process smooth fluctuations around a mean of a time series. In addition, the trend component is represented by some weighted mean based on the past values of the time series. Elsewhere, the residual is composed of strictly random part of the raw time series noticed by ε_i.

ARMA process is characterized by following equation:

$$x_t = \sum_{i=1}^{p} \varphi_i x_{t-1} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \qquad (1)$$

φ_i : parameter of Autoregressive process
θ_i : parameter of the moving average process ($\theta_0 = 1$)
ε_i a residual.

3. Application of the ARMA process (p, q) on the studied site.

3.1 Dickey-Fuller test

To check for non-stationarity of recorded data, a general function noticed by "adf.test" has been used. The function is applied using the package "tseries" of the R software. Among tests, one can cite the classical Dickey-Fuller test, the augmented Dickey-Fuller test and the Phillips-Perron test. For above tests, the null hypothesis is non-stationarity of the time series [15]. In this study, the augmented Dickey-Fuller test is used to verify non-stationarity of the solar radiation.

3.2 Validation of the ARMA process (p, q)

In the determination p and q orders of the ARMA (p, q) process, the autocorrelation and partial autocorrelation functions can be used. Through above functions, we can select several possible orders p and q. The AIC criterion is used to select the optimal order of p and q. Validation of the process includes an examination of the estimated coefficients and by

analysing the residue, (the estimated residues must follow a white noise process). Indeed, analysis of residual is based on the autocorrelation, Box, and Pierre tests [12, 19]. This test applied to the residual to validate whether residual is "white noise" (zero mean, fixe and uncorrelated variance).We use the Box.test function of the R software to perform this test. In practice, this function gives a p-value for the applied test. The null hypothesis H_0 of white noise is accepted, when the p-value is greater than 5% and is rejected otherwise.

4. Information Criterion

We study the specific criteria of the stochastic models such the Akaike's AIC (An information criterion) criterion and the Bayesian one (BIC) to check for optimum orders of the model. Among these criteria, we chose to work with the Akaike criterion. Indeed, this criterion of Akaike or AIC is based on researching optimum parameters p and q [16]. The best of the ARMA models (p, q) is the model that minimizes the statistic determined in priori.

This criterion AIC is described follow the eq. (2):

$$AIC(p,q) = \log \hat{\sigma}^2 + \frac{2(p+q)}{T} \qquad (2)$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=0}^{N} (\varepsilon_t - \bar{\varepsilon}_t)^2$$

$$\bar{\varepsilon}_t = 0$$

5. Model Performance Criteria

In forecasting purposes, several methods are available for verifying the performances and the reliability of the used model [17, 22, 14]. In this paper, the root mean squared error (RMSE), the mean absolute error (MAE), the mean bias error (MBE) and the coefficient of determination (R^2) are used to evaluate issues of the model.

The mean root error (RMSE) representing the standard deviation between predictions and observations is commonly used. The RMSE is given by the following equation.

$$RMSE = (\frac{1}{N} \sum_{t=1}^{N} (P_i - \hat{P}_i)^2)^{\frac{1}{2}} \qquad (3)$$

The Mean Absolute Error (MAE) measures the difference between predictions and observations and is given by eq. (4)

$$MAE = \frac{1}{N} \sum_{1=1}^{N} \left| P - \hat{P} \right| \qquad (4)$$

Mean Bias Error (MBE) is the estimation of algebraic difference between forecasted time series and observed one. A positive value of MBE means that predicted values overestimates expectations, while a negative value characterize an underestimation. The MBE value is calculated using the following equation.

$$MBE = \frac{1}{N} \sum_{t=1}^{N} (P_i - \hat{P}_i) \qquad (5)$$

The coefficient of determination (R²) indicates how the regression curve adjusts the data or how variation in time of forecasts fits observations. R² is calculated using eq. (6)

$$R^2 = 1 - \frac{\sum\limits_{i}^{n}(P_i - \hat{P}_i)^2}{\sum\limits_{i}^{n}(P_i)^2} \qquad (6)$$

Results

6. Autocorrelation partial autocorrelation tests

Fig.2 and Fig.3 present respectively the plots of the partial autocorrelation and autocorrelation functions of the solar potential recorded at the Dakar site. This AFC curve shows oscillatory behaviour and the deterministic feature of the solar potential. Indeed, the bars of the AFC curve go beyond the confidence interval, so the data are random. The AFC curve maintains periodic (deterministic) behaviour of a time series if it exists. In addition, the augmented Dickey-Fuller test has been carried out for this purpose through the R software. Results show that data are random with a p-value of 0.01 for a significant level of 5% (thus greater than 5%). The null hypothesis for this test is the non-stationarity of the time series.

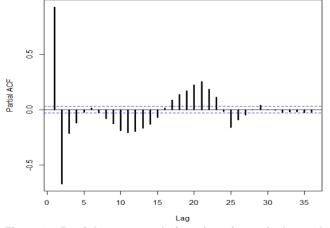**Figure2: Autocorrelation function of the solar potential.**

Figure3: Partial autocorrelation function of the solar potential.

7. Calibration of the ARMA model

To determine the order of the ARMA process, the AIC information criterion has been used. According to the test, the retained orders are of 29 for p and 0 for q with a variance equal to 1.252. Then, we find that an ARMA process (29, 0) fits better the data on Dakar site according to the AIC information criterion. Residual is of zero mean (white noise) of which statistical characteristics are presented in Table1.

8. Noise validation

Statistic characteristic of the noise from the ARMA (29, 0) in the calibrating period (October 2016 to September 2017) are presented in Table 1. Results show residual is white noise, then retained orders can be used to predict the solar potential for 24 hours forecast horizon on the Dakar site. In addition, the Box-Pierce test has been applied to the data in order to confirm randomness of the residual. Performed test has a P-value of about 0.26 at significant level of 5%. Then, the null hypothesis of randomness is accepted.

Table1: Statistical characteristic of the residual.

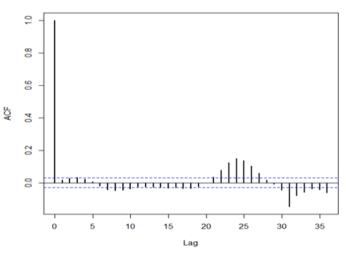| Site	Residual	Mean	Var	P-value
CIFRES/ESP	ε	0	1.252	0.26

Figure4: Autocorrelation function of the residual from applied ARMA (29, 0) at the Dakar site.

9. Study of the forecast quality

Table 2 shows performances of the ARMA (29, 0) for 24 hours forecast horizon. The Coefficient of Variation between observations and forecasts is about 0.963 close to 1. This means that the forecasts are well representative of the observations. The Mean Bias Error (MBE) used to predict solar potential at 24 hours time horizon is about 0.012. The positive value of the MBE means that the model overestimate solar potential in the study area. The Root Mean Squared Error (RMSE) is equal to 0.629, accordingly the model is acceptable because of reduce RMSE. In

a same way, the MEA of which value is about 0.528 confirms the reliability of the model. Globally, analysis of the performances criteria confirms the reliability of the ARMA (29, 0). Then, the model can be used to predict the solar radiation for 24 hours forecast horizon one the Dakar site.

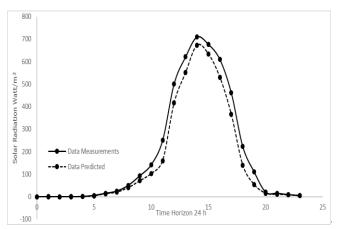Table 2: Performances of the Prediction Model ARMA (29, 0).

RMSE	MBE	MAE	R²
0.629	0.012	0.528	0.963

Figure5: Evolution of measured and predicted data of the study area.

Discussion

According to the tests carried out, the selected orders are p = 29 and q = 0. It is found that an ARMA (29, 0**)** reproduces the data of the Dakar site better according to the Akaike information criterion (AIC). The Box-Pierce test was applied to check the randomness of the noise. It confirms that the noise is white with zero mean, variance of 1.252 and P-value of about 26% above the 5% threshold. The model accuracy is studied by calculation of the RMSE = 0.629 (root mean squared error), the R² = 0.963 (Coefficient of determination), the MAE = 0.528 (mean absolute error) and the MBE = 0.012 (mean bias error). These Performance criteria confirm well the reliability of the ARMA (29, 0) in 24 hours ahead forecasting of the solar potential of the site of Dakar.
Other similar work has already been carried out as part of the prediction of solar potential [23, 24, 7, 25, 27].
In A. Mbaye and al., (2018), discrete Kalman filter model is applied at the Dakar site (Senegal); results show very good performance (RMSE = 0.048 and R ² = 0.99) for 20 minute forecast horizon.
M. Willy and al., (2018) applied a multi-layer perception (PMC) model at the Gandon (Senegal) site. The results also show very good performance (RMSE = 0.03 and R ² = 0.99) for 10 minute forecast horizon.

In C. Voyant and al., (2013), used an ARMA(1,0) to predict the global solar radiation at several Mediterranean sites with the best mean annual of RMSE standardized equal to 19.4% on the site of Ajaccio and 16% on Marseille.
In M. Y. Sulaiman and al., (1997), applied the Box-Jenkins approach to daily solar radiation from four different locations in Malaysia. They removed the deterministic component using Fourier analysis and modelled the stochastic component modeled using ARMA model. For this case of study, the residuals are better described by ARMA (2, 0).
In As'ad. Mohamad, (2012**)** works toward finding the best model for predicting peak of electricity demand for up to seven days in Australlia considering time series upon one year. Four Autoregressive Integrated Moving Average (ARIMA) models have been compared for. Results show that, the ARIMA model calibrated using the data for December 2010-May 2011 is the best (RMSE = 369.678) in comparison to the ARIMA model calibrated from September 2010 to May 2011 (RMSE = 370.735).
In this paper, we confirm that the ARMA (p, q) model is applicable on different spatio-temporal scales. In addition, the results almost show statistical similarity between studies. Globally, one can retain that the ARMA (p, q) can give very good performance and can be applied on various regions with different climates around the world. This study can help to improve the planning of photovoltaic solar power plants production in the Sahelian zone and in the sub region.

Conclusion

This study focuses on short-term solar potential prediction on a horizontal surface at Dakar. An ARMA (p, q) model was applied to the data measured on Dakar (Senegal) site for 24 hours forecast horizon solar potential prediction purpose. ARMA model is performed on the basis on finding optimum parameters p and q that better fit data (sunshine). Non-stationarity of the time series is first checked applying the Dickey-Fuller test. It shows that data are random with a p-value of about 0.01% for significance level of 5%. The Akaike information criterion was used to determine the orders of the ARMA process. Retained parameters are p=29 and q=0.The Box-Pierce test was used to test the residual. It found that residual series are white noise for significance level of 5% with zero mean, variance of 1.252 and a p-value of 26%. Verification of the model is done by analyzing some statistical performance criteria such the RMSE = 0.629 (root mean squared error), the R ² = 0.963 (Coefficient of determination), the MAE = 0.528 (mean absolut error) and the MBE = 0.012 (mean bias error). Statistics criteria show that the ARMA (29, 0) is reliable. This model can be used as a decision support tool for planning and managing photovoltaic solar power plants in Senegal and surrounding. We projected to apply same ARMA model in other tropical and intertropical areas in comparison purpose.

Complementary informations on authors:

Amy Mbaye: e-mail and Research Gate: amy1.mbaye@ucad.edu.sn ; ORCID and Linkedln profiles: amymbaye86@yahoo.fr.

Didier Maria Ndione: e-mail, Research Gate, ORCID and Linkedln profiles: didiermaria.ndione@ucad.edu.sn.

REFERENCES

1. A. K. Yadav and S. S. Chandel, Solar radiation prediction using Artificial Neural Network techniques, Sustain Energy, vol.33, p 772 (2014).

2. M. Ikhsan, A. Purwadi, N. Hariyanto, N. Heryana and Y. Haroen, Study of Renewable Energy Sources Capacity and Loading Using Data Logger for Sizing of Solar-wind Hybrid Power System, Procedia Technol, vol.11, p 1048 (2013).

3. H. Madsen, H. A. Nielsen and T. S. Nielsen, A tool for predicting the wind power production of offshore wind plants, Proceedings of the Copenhagen Offshore Wind Conference & Exhibition (2005).

4. A. Mellit, A. Pavan and V. Lughi, Short-term forecasting of power production in a largescale photovoltaic plant, Solar Energy, vol.105, p 401 (2014).

5. P. Brockwell and D. RA, Time series: theory and methods, Springer series in statistics, second edition (2006).
6. J. D. Hamilton, Times series analysis, one of Princeton University Press's Notable Centenary Titles, Hard over (1994).

7. C. Voyant, C. Paoli, M. Muselli and M. L. Nivet, Multi-horizon solar radiation forecasting for Mediterranean locations using time series models, Renew Subtain Energy, vol.28, p 4452 (2013).

8. A. Zaharim, A. M. Razali, T. P. Gim and K. Sopian, Time Series Analysis of solar radiation data in the Tropics, European Journal of Scientific Research, Vol.25, p 672 (2009).

9. J. Boland, Time series and statistical modelling of solar radiation, Recent Advances in Solar Radiation Modelling, Viorel Badescu (Ed.), Springer Verlag, p 283 (2008).

10. M. David, F. Ramahatana, P. J. Trombe and P. Lauret, Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models, Solar Energy, Elsevier, vol.133, p 55 (2016).

11. C. Benard, E. Boileau and B. Guerrier, Modeling of global solar irradiation using ARMA processes: application to low-time (hourly) prediction, with a view to establishing optimal control in the habitat, Revue Phys,vol.20,p 845 (1985).

12. G. E. P. Box and David A. Pierce, Distribution of residual autocorrelations in autoregressive integrated moving average time series models, Journal of the American Statistical Association, vol.65, p 1509 (1970).

13. C. Join, M. Fliess, C. Voyant and F. Chaxel, Solar energy production: Short-term forecasting and risk management, IFAC Papers Online, Vol.49 p 686 (2016).

14. C. Voyant, Prediction of time series of global solar radiation and photovoltaic energy production from artificial neural networks, Thesis, university of Corsican-Pascal Paoli doctoral school environment and society,Chapter2, p 87 (2011).

15. R. Davidson and J. G. Mackinnon, Report that the Phillips Perron test performs worse in finite samples than the augmented Dickey–Fuller test, Journal of Economic Dynamics and Control, vol.12, p 297 (2004).

16. T. Fang and R. Lahdelma, Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system, Applied Energy, vol.179, p 544 (2016).

17. H. Jiang, Y. Dong and L. Xiao, A multi-stage intelligent approach based on an ensemble of two-way interaction model for forecasting the global horizontal radiation of India, Energy Conversion and Management, vol.137, p 142 (2017).

18. H. Wold, A study in the Analysis of Stationary Time Series, J. N. Journal of the Royal Statistical Society, Vol. 102, p 295 (1939).
19. G. M. Ljung and G. E. P. Box, on a measure of lack of fit in time series models, Biometrika vol.65, p 297 (1978).

20. M. Lazzaroni, S. Ferrari, V. Piuri, A. S. Salman, L. Cristaldi and M. Faifer, Models for Solar Radiation Prediction based on

Different Measurement Sites, Elsevier, measurement, vol.11, p 037 (2015).

21. S. Monjoly, R. Calif, M. Andre and T. Soubdhan, Prediction tools for wind power generation: Application to optimize coupling to electricity distribution networks, University of the West Indies and Guyana December, (2013).

22. T. Soubdhan, J. Ndong, H. O. Baba and M. ThangDo, A robust forecasting framework based on the Kalman filtering approach with a twofold parameter tuning procedure: Application to solar and photovoltaic prediction, Solar Energy, vol.131, p 246 (2016).

23. A. Mbaye, J. Ndong, M. L. NDiaye, M. Sylla, M. C. Aidara, M. Diaw, M. F. NDiaye, P. A. Ndiaye and A. Ndiaye, Kalman filter model as a tool for short-term forecasting of solar potential: case of the Dakar site, EDP Science, vol.57, p 2267 (2018).

24. W. M. Nkounga, M. F. Ndiaye, M. L. Ndiaye, O. Cisse, M. Bop and A. Sioutas, Short-term forecasting for solar irradiation based on the multi-layer neural network with the Levenberg-Marquardt algorithm and meteorological data: application to the Gandon site in Senegal, 7TH International Conference on Renewable Energy Research and Applications, IEEE, p 869 (2018).

25. Y. M. Sulaiman, O. W. M. Hliang and M. A. Wahab, Modeling of solar radiation in time series Domain, Renewable Energy, vol.11, p 97 (1997).

26. M. A. Alghoul, M. Y. Sulaiman, B. Z. Azami and M. A. Wahab, Modeling of solar radiation in time series Domain, International Energy Journal, vol.7, p 261 (2006).

27. M. As'ad, Finding the Best ARIMA Model to Forecast Daily Peak Electricity Demand, Proceedings of the Fifth Annual Applied Statistics Education and Research Collaboration (ASEARC) Conference Papers, University of Wollongong Australia (2012).