INTERNATIONAL JOURNAL of RENEWABLE ENERGY RESEARCH
Belhaj I. and Fathi E.O., Vol.9, No.1, March, 2009

A Benchmark of Statistical Models for Forecasting Monthly Direct Normal Irradiation (DNI) for the Region of Ouarzazate Morocco

Ismail Belhaj*‡, Omkaltoume El Fatni*

*Physics Department, Faculty of Sciences Rabat, LPHE-Modelling and Simulations, 4 Avenue Ibn Battouta B.P. 1014 RP, Rabat, Morocco

(ibelhaj13@gmail.com, omfatni@gmail.com)

‡ Corresponding Author; Ismail Belhaj, 4 Avenue Ibn Battouta B.P. 1014 RP, Rabat, Morocco, Tel: +212 671 694 767,

ibelhaj13@gmail.com

Received: 11.11.2018 Accepted:21.12.2018

Abstract- Direct Normal Irradiation (DNI) is the main component of solar thermal (ST) systems and concentrated solar power systems (CSP and CPV). The volatility of this component due its renewable resource nature can cause fluctuations that affect the electrical grid, hence the importance of forecasting it. In this work, the forecasting of monthly average Direct Normal Irradiation (DNI) is explored. A benchmark of statistical forecasting models is used to select the best statistical forecasting model. Seventeen models are evaluated and compared, namely Trend models, Box Jenks models and Exponential Smoothing models. The satellite data used are of the period extended from 1994 to 2012 for the region of Ouarzazate in Morocco. The daytime period is from 07h30 till 17h30. The design region of the models is from 1994 to 2009 divided into an estimation period and a validation period. Data from 2010 to 2012 are used as forecasting years. Several error metrics are used for performance evaluation and comparison. The results indicate that a seasonal ARIMA model outperforms the other statistical models with a MPE=-0.490442% in the validation period and MPE=-5.1907% in the forecasting region.

Keywords Forecasting, Time Series, Direct Normal Irradiation (DNI), Smoothing, Solar radiation, Concentrated Solar Power (CSP), Ouarzazate.

1. Introduction

Solar radiation has received much attention in recent years due to its use in harvesting different forms of energy such as electricity and heat. Forecasting Direct Normal Irradiation (DNI) has become very important recently as the DNI is the main component of concentrated solar power (CSP) and concentrated photovoltaic (CPV) systems. The uncertain nature of renewable energy resources such as solar energy produces fluctuations that affect the electrical grid which lead in some cases to blackouts hence the necessity to reduce the cost of imbalance in the grid using forecasting [1-4], which proved to be a better way to increase flexibility in the grid compared to other solutions such as increasing storage capabilities [5], forecasting the energy demand [6, 7] and prediction of electricity energy consumption [8]. Numerous studies were conducted on intra-hourly and hourly DNI forecasting [9-15] useful for operations and load-following , however not many researches were conducted on ~~to~~ benchmark studies of monthly average solar radiation forecasting especially the component of DNI which is important for the task of planning in CSP solar sites. This gap in this particular forecasting time horizon for DNI and its use in the industry make it an interesting subject to tackle.

There is a growing concern on long term solar forecasting in the latest years. A recent research involved long-term GHI forecasting (monthly and seasonal) for regional sites in Queensland, Australia where 3 methods were compared, namely artificial neural networks (ANN), multiple linear regression (MLR) and ARIMA [16]. The results showed that the ANN outperformed the other

methods in monthly and seasonal forecasting with an overall average RMSE=1.23MJ m-2 day-1. Another study revealed the results of the investigation on models for forecasting monthly mean daily global solar radiation from in-situ measurements in the tropical climate of India [17]. The results showed that Hargreaves and Samani model (HSM) performed satisfactory both on Hyderabad and Vishakhapatnam regions with a mean percentage error MPE=-5.39% and -20.24% respectively. Long term DNI forecasting, in this case using monthly average data, is very important in planning, management and decision making for engineers, policy makers and energy experts for cases such as load pricing and energy management in areas where solar thermal plants are installed or intended for implantation. That being said, the choice of a reliable predictive model is crucial for making the best forecasts hence the importance of conducting a study of benchmarking forecasting models for electricity balancing market prices [18] or the renewable resource itself, in this case DNI, which is subject of this manuscript.

The purpose of the present work is to investigate a time series based benchmark of forecasting methods on monthly average DNI using satellite data. The Moroccan region of Ouarzazate in the South West Mediterranean basin area is the case study due to the availability of data and the importance of this region in the Moroccan Solar Plan (MSP), Africa and the Mediterranean region as it shelters Africa's biggest solar plant "NOOR" and the closest one to Spain which can be later considered for energy exportation. Data considered lay from the period 1994 to 2012 using 07h30 to 17h30 daytime hours. Several metric measures were performed, such as the Akaike Information Criterion (AIC) and root mean squared error (RMSE), to assess the forecasting skills and select the best predictive model among the seventeen models being tested. The rest of the paper is organized as follows: section 2 presents the data pre-processing procedures; Section 3describes the proposed methods tested for forecasting while section 4 reveals the findings in details. Finally, section 5 reports the conclusions made and discusses the results with an insight of the next future coming works.

Nomenclature

Acronyms	
AIC	Akaike Information Criterion
AR	Autoregressive
ARMA	Autoregressive moving average
ARIMA	Autoregressive integrated moving average
MA	Moving average
MAE	Mean Absolute error
MAPE	Mean absolute percentage error
ME	Mean error
MPE	Mean percentage error
MSE	Mean square error
RMSE	Root mean squared error
SARIMA	Seasonal autoregressive integrated moving average
Greek symbols	
α	A smoothing constant used to estimate the level
β	A smoothing constant used to estimate the slope
γ	A smoothing constant used to estimate the seasonality
ε_t	White noise
ε_{t-j}	Lagged errors used as predictors in the $MA(q)$ and $ARMA(p,q)$ models
φ_i	Coefficients of the lagged observations in the $AR(p)$ and $ARMA(p,q)$ models
θ_j	Coefficients of the lagged errors in the $MA(q)$ and $ARMA(p,q)$ models
Roman symbols	
B	Lag operator also called the "Backshift"
C	Constant
$\hat{a},\hat{b}$ And $\hat{c}$	Estimates of the parameters a,b,c of the trend models using the OLS method
d	Order of the non-seasonal differencing
D	Order of the seasonal differencing
$F_t(k)$	The forecast for the horizon $t+k$ at time t
I_t	The seasonal estimate of Y_t
k	The forecast horizon or the number of steps in the future
L	Likelihood of the data

p	Order of the non-seasonal AR term
P (SAR)	Order of the seasonal autoregressive term
q term	Order of the non-seasonal moving MA
Q (SMA)	Order of the seasonal moving average term
s	The number of seasons
S_t	The estimate of the local Level
S_t'	The singly-smoothed series obtained by applying simple exponential smoothing to series Y at time t
S_t''	The doubly-smoothed series obtained by applying simple exponential smoothing
S_t'''	The triple smoothed series obtained by applying simple exponential smoothing
T_t	The estimate of the local Trend
Y_t / y_t	A given time series
$\overline{Y}$ time t	The average of the data up to and including
y_{t-i} the $AR(p)$ and $ARMA(p,q)$ models	Lagged observations used as predictors in

2. Pre-processing Data

The region of Ouarzazate located in the latitude 31°.00'83" and longitude -06°.86'27" in south Morocco is a strategic region for the Moroccan Solar Plan (MSP) that aims to produce 2000 MW of electricity from renewable energy resources by 2020 to reduce Morocco's energy dependency with a goal of 42% of electricity from renewable resources by 2020 and 52% by 2030 not to mention the reduction of CO2. The Direct Normal Irradiation (DNI) used is the hourly data accumulated from METEOSAT MSG and MFG satellite data (EUMETSAT) and from atmospheric data (ECMWF and NOAA) by SOLARGIS method for the region of Ouarzazate. The spatial resolution for solar radiation data is 250m□250m and for meteorological data is 1000m□1000m. The entire hourly database used represents 19 years of data starting from 1st January 1994 to 31st December 2012. The period between 07:30 to 17:30 is considered as it represents the daytime solar duration. The data were pre-processed by excluding the 29th February day from each year to avoid the leap years and homogenize the data. The monthly average DNI data were then computed and divided into two regions (partitions): The design region where the model is trained, tested and validate; in this case the design period is from 1994 to 2009 divided into an estimation period from January 1994 to February 2005 which constitutes 70% of the data used in the design period, and a validation period from March 2005 to December 2009 which constitutes 30% of the data; and the forecast region where the model is used to generate forecasts blindly, which is chosen to be three years from January 2010 to December 2012.

3. Methods

The present work focuses on bencharking several statistical forecasting models using monthly average DNI time series. In this section, three meain branches of models are introduced, namely Trend models, Box-Jenkins models (ARMA and ARIMA models) and exponential smoothing models.

3.1. Trend Models

Trend models assume that future forecasts are fitting various types of regression models based only on time as an independent variable with random fluctuations. Trend odels weight all data equally and are fit by least squares, resulting in estimates of up to 3 coefficients *a, b* and *c* denoted $\hat{a}$, $\hat{b}$ and $\hat{c}$ respectively. In this work, The Mean, Linear Trend, Quadratic Trend, Exponential Trend, and S-Curve models are explored.

Mean Model: $$F_t(k) = \overline{Y} \quad (1)$$

Here is the average of the data up to and including time *t*

Linear Trend: $$F_t(k) = \hat{a} + \hat{b}(t+k)) \quad (2)$$

Quadratic Trend: $$F_t(k) = \hat{a} + \hat{b}(t+k) + \hat{c}(t+k)^2 \quad (3)$$

Exponential Trend: $$F_t(k) = \exp(\hat{a} + \hat{b}(t+k)) \quad (4)$$

S-Curve: $$F_t(k) = \exp(\hat{a} + \hat{b}/(t+k)) \quad (5)$$

In the five models, $F_t(k)$ denotes the forecast for time $t+k$ at time t.

3.2. ARMA Models

(ARMA) is a group of models that handle stationary time series. The ARMA models are namely:

- The Autoregressive model (AR)
- The Moving Average model (MA)
- The Autoregressive Moving Average model (ARMA)

The *AR(p)* model:

The (AR) model is a multiple regression model with lagged observations as y_{t-i} predictors:

$$y_t = C + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t = C + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t \quad (6)$$

ε_t is a white noise, φ_i are the coefficients of the lagged observations, C is a constant and p is the autoregressive order.

The *MA(q)* model:

The (MA) model is a multiple regression model with lagged errors ε_{t-q} as predictors:

$$y_t = C + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t = C + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (7)$$

ε_t is a white noise, θ_j are the coefficients of the lagged errors ε_{t-j}, C is a constant and q is the moving average order.

The *ARMA(p,q)* model:

The (ARMA) model is a multiple regression model that combines both lagged observations with lagged errors as predictors:

$$y_t = C + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t = C + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (8)$$

3.3. ARIMA models

In general, two types of ARIMA models can be distinguished:

- Nonseasonal ARIMA models
- Seasonal ARIMA models

3.3.1. Nonseasonal ARIMA models

The (ARIMA) models, also known as Box-Jenkins models are time series models. They are the generalized form of ARMA models that take into consideration stationary and non stationary by adding a differencing parameter. The ARIMA models can be written as *ARIMA(p,d,q)* where p is the autoregressive (AR) order which is the number of lags, d the degree of first differencing involved and q the oving average (MA) order.

The general equation of an ARIMA model is given by:

$$\left(1 - \sum_{i=1}^{p} \varphi_i B^i\right)(1-B)^d y_t = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right)\varepsilon_t \quad (9)$$

B is the lag operator called the "backshift" operator such that $B^k y_t = y_{t-k}$. In special situations, where p, d or q is equal to zero the following cases arise:

Case 1: when $d=0$ the ARIMA model turns into an *ARMA(p,q)* model

Case 2: when $d=0$ and $q=0$ the ARIMA model is reduced to *AR(p)* model

Case 3: when $d=0$ and $p=0$ the ARIMA model is reduced to *MA(q)* model

For further readings about the ARIMA models, the references [19-22] are suggested.

3.3.2. Seasonal ARIMA models (SARIMA)

The seasonal ARIMA models, also called Multiplicative ARIMA models are ARIMA models with components of seasonality variables (SAR, SMA and the differencing of seasonality). The multiplicative Seasonal ARIMA (SARIMA) model is denoted by $ARIMA(p,d,q)\square(P,D,Q)_s$, where s is the number of seasons, p is the order of the non-seasonal autoregressive term, d is the order of non-seasonal differencing term, q is the order of the non-seasonal moving average term, P is the order of the seasonal autoregressive term (SAR), D is the order of seasonal differencing and Q is the order of the seasonal moving average term (SMA). The SARIMA model is described with following equation [23]:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D x_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (10)$$

B is the lag operator, ε_t are the residuals, $\phi_p(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ is a polynomial in B of degree p, $\theta_q(B) = 1 + \sum_{i=1}^{q} \theta_i B^i$ is a polynomial in B of degree q, $\Phi_P(B^s) = 1 - \sum_{i=1}^{P} \Phi_i B^{is}$ is a polynomial in B^s of degree P, and $\Theta_Q(B^s) = 1 + \sum_{i=1}^{Q} \Theta_i B^{is}$ is a polynomial in B^s of degree Q. It is worth mentioning that seasonal differencing $\Delta_s y_t = (1-B^s)y_t = y_t - y_{t-s}$ removes seasonality in the same way that ordinary differencing $\Delta y_t = y_t - y_{t-1}$ removes a polynomial trend.

3.4. Exponential Smoothing models

Smoothing is a technique based on averaging over multiple periods in order to minimize the noise using an exponential function [24-25]. The use of the exponential functions is intended to assign weight exponentially to the corresponding observations in a decreasing order. The different smoothing models are called "smoothers". To generate the forecasts, up to three passes of an exponential smoother are made:

Pass 1 $$S_t^{'} = \alpha Y_t + (1-\alpha) S_{t-1}^{'} \quad (11)$$

α is the smoothing constant such that 0<α<1 and $S_t^{'}$ is the singly smoothed series obtained by applying simple exponential smoothing to series Y at time t.

Pass 2: $$S_t^{''} = \alpha S_t^{'} + (1-\alpha) S_{t-1}^{''} \quad (12)$$

$S_t^{''}$ is the doubly-smoothed series obtained by applying simple exponential smoothing using the same α to series $S_t^{'}$ at time *t*

Pass 3 $$S_t^{'''} = S_t^{''} + (1-\alpha)S_{t-1}^{'''} \quad (13)$$

$S_t^{'''}$ is the triple smoothed series obtained by applying simple exponential smoothing using the same α to series $S_t^{''}$ at time *t*. The initial values at time *t=0* are determined by back-forecasting by smoothing first the time series backwards and then using the back-forecasts to initialize the forward smoothing.

3.4.1. Simple exponential smoothing

The simple exponential smoothing estimates trend similar to the Mean Trend model [26]:

$$F_t(k) = S_t^{'} \quad (14)$$

3.4.2. Brown's linear exponential smoothing

Brown's linear exponential smoothing estimates trend similar to the linear trend model [27]:

$$F_t(k) = 2S_t^{'} - S_t^{''} + k\frac{\alpha}{1-\alpha}(S_t^{'} - S_t^{''}) \quad (15)$$

3.4.3. Quadratic exponential smoothing

The Quadratic exponential smoothing is described by the following formula [28]:

$$F_t(k) = 3S_t^{'} - 3S_t^{''} + S_t^{'''} + k\frac{\alpha}{2(1-\alpha)^2}((6-5\alpha)S_t^{'} - (10-8\alpha)S_t^{''} + (4-3\alpha)S_t^{'''}) + k^2\frac{\alpha^2}{2(1-\alpha)^2}(S_t^{'} - 2S_t^{''} + S_t^{'''}) \quad (16)$$

k are the steps ahead in the future.

3.4.4. Holt's linear exponential smoothing

Holt's Linear Exponential smoothing is similar to Brown's Linear Exponential Smoothing in that it generates forecasts that follow a linear trend. However, Holt's procedure uses smoothing constants α (alpha) and β (beta), one to estimate the level of the series at time t and a second to estimate the slope [29]. The procedure is as follows:

Smooth the data to estimate the level using

$$S_t = \alpha Y_t + (1-\alpha)(S_{t-1} + T_{t-1}) \quad (17)$$

Smooth the first smooth to estimate the slope using

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1} \quad (18)$$

Calculate the forecasts using

$$F_t(k) = S_t + kT_t \quad (19)$$

S_t is the estimate of the local level and T_t is the estimate of the local trend.

3.4.5. Winter's exponential smoothing

All of the forecasting methods described above handle the seasonality by first seasonally adjusting the data, then applying the forecasting model, and then putting back the seasonality. Winter's exponential Smoothing procedure handles the seasonality directly by estimating seasonality at the same time that it estimates the level and trend. In extends Holt's procedure by adding an additional parameter gamma γ to use in a third smoother [30]. The procedure is as follows:

Estimate the seasonality by smoothing the ratio of the data to the estimated level at time *t* using:

$$I_t = \gamma\frac{Y_t}{S_t} + (1-\gamma)I_{t-s} \quad (20)$$

I_t is the seasonal estimate of Y_t

Estimate the level of the series by smoothing the data divide by the estimated seasonality using

$$S_t = \alpha\frac{Y_t}{I_{t-s}} + (1-\alpha)(S_{t-1} + T_{t-1}) \quad (21)$$

Estimate the slope of the series using

$$T_t = \beta(S_{t-1} + T_{t-1}) + (1-\beta)T_{t-1} \quad (22)$$

Calculate the forecasts using

$$F_t(k) = (S_t + kT_t)I_{t-s+k} \quad (23)$$

Here $S_t + kT_t$ is the extrapolation of level and trend from period *t* and I_{t-s+k} is the most recent estimate of the seasonal index for k^{th} period in the future. It is worth mentionning that in all the exponential smoothing techniques used, the parameters α, β and γ are heuristic parameters between 0 and 1 both excluded.

The metrics used to assess the accuracy of the forecasting methods are the (RMSE), (ME), (MPE), (MAE), (MAPE) and (AIC).

The Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} \quad (24)$$

The RMSE is a measure of average squared deviation of forecasted values which penalizes large individual errors occurred during the forecasting. Here, e_t is the error at a time *t* compute by subtracting the forecasted values from the actual data and *n* is the number of samples.

The Mean Error (ME)

$$ME = \frac{1}{n}\sum_{t=1}^{n} e_t \quad (25)$$

The Mean Percentage Error (MPE)

$$MPE = \frac{1}{n}\sum_{t=1}^{n} \frac{e_t}{y_t} \times 100 \quad (26)$$

The MPE is a measure of overall bias error or systematic error. It is used as a percentage error.

The Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \quad (27)$$

The MAE shows the magnitude of overall error, occurred due to forecasting. It does not penalize extreme forecast errors.

The Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{e_t}{y_t}\right| \times 100 \quad (28)$$

The MAPE represents the percentage of average absolute error occurred. Here, n is the number of observations, e_t is the error at a time t compute by subtracting the forecasted values from the actual data and y_t are the actual data at time t. Each of the RMSE, MAE, MAPE, ME and MPE, is based on the one-ahead forecast errors, which are the differences between the data value at time t and the forecast of that value made at time t-1. The first three statistics measure magnitude of the errors while the last two statistics measure bias. A better model will give a small value. Details about the importance of use of these metrics can be found in [31].

The Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is a criterion formulated by the Japanese Statistician Hirotugu Akaike in the early 1970s [32]. AIC is an information criterion used for model selection described by the following equation:

$$\mathrm{AIC} = -2\log(L) + 2(p+q+k) \quad (29)$$

L is the likelihood of the data, p is the autoregressive order and q is moving average order and k represents the intercept of the ARIMA model. For a model to be selected, it has to be the one with the lowest AIC. AIC discourages over-fitting and rewards goodness of fit using the likelihood function L.

4. Results and Discussion

In this section, the results are presented and discussed. In the present work, the MSE is used as a minimization function to select the forecasting model. Models with and without constant were tested. Table 1 presents the optimized best models found after test.

Table 1. The different 17 predictive models found after test

N° of Models	Models found after test
(M1)	Random walk
(M2)	Random walk with drift = 0,0226256
(M3)	Constant mean = 605,114
(M4)	Linear trend = 602,979+ 0,03164 t
(M5)	Quadratic trend = 585,098 + 0,820477 t + -0,00584323 t^2
(M6)	Exponential trend = exp(6,39709 + 0,0000495657 t)
(M7)	S-curve trend = exp(6,40309 + -0,0648205 /t)
(M8)	Simple moving average of 2 terms
(M9)	Simple exponential smoothing with alpha = 0,0083
(M10)	Brown's linear exp. smoothing with alpha = 0,0031
(M11)	Holt's linear exp. smoothing with alpha = 0,2372 and beta = 0,0188
(M12)	Brown's quadratic exp. smoothing with alpha = 0,0065
(M13)	Winter's exp. smoothing with alpha = 0,2222, beta = 0,0219, gamma = 0,1083
(M14)	ARIMA(1,0,0)x(2,0,1)12 with constant
(M15)	ARIMA(0,0,1)x(2,0,2)12 with constant
(M16)	ARIMA(2,1,1)x(2,0,2)12
(M17)	ARIMA(2,0,0)x(2,0,1)12 with constant

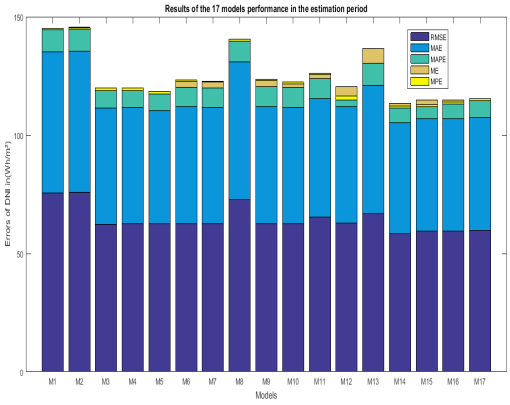

Fig. 1. Results of the 17 models' performance in the estimation period.

In this case, the models were estimated from the first 134 data values and the last 58 data values were withheld for validation. Figure 1 and Figure 2 reveals the results of the models found based on their performance at the estimation and validation periods respectively.

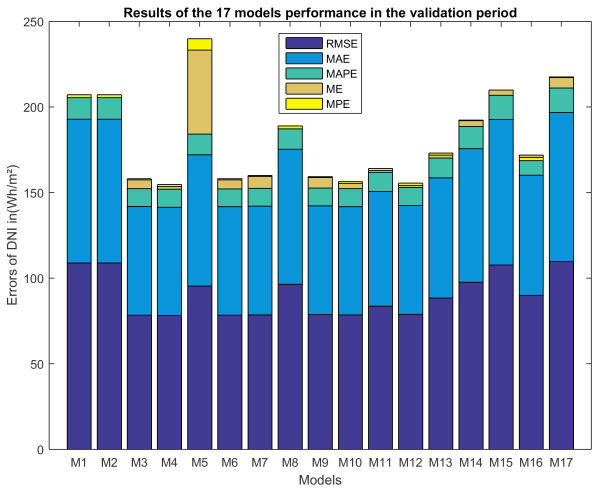

Fig. 2. Results of the 17 models' performance in the validation period.

The selection of the best model is based on all the performance metrics with a focus on the RMSE and MPE to measure the magnitude of the errors and the error bias respectively. AIC is also used to rank the models based on their quality. In the estimation period, the model with the lowest value of the RMSE, MPE and AIC (with RMSE= 58.4456, MPE=-0.945163 and AIC=8.2082) is model M14 which is a mixed seasonal autoregressive integrated moving average model *ARIMA(1,0,0)□(2,0,1)12* with constant. This model assumes that the best forecast for future data is given by a parametric model relating the most recent data value to previous noise.

While comparing the out-of-sample error statistics (validation period) to the in-sample statistics (estimation period), the noticeable rise of error and the change of performance of some models suggests the presence of non-randomness in the residuals. For this reason, five diagnostic tests for residual randomness and stationarity were performed namely, the test for excessive runs up and down (RUNS), the test for excessive runs above and below median (RUNM), Box-Pierce test for excessive autocorrelation (AUTO), the test for difference in mean 1st half to 2nd half (MEAN) and the test for difference in variance 1st half to 2nd half (VAR). Table 2 represents the results of the five tests.

Table 2. Table of the five diagnostic tests

Model	*RMSE*	*RUNS*	*RUNM*	*AUTO*	*MEAN*	*VAR*
(M1)	75,5007	OK	*	***	OK	OK
(M2)	75,8121	OK	*	***	OK	OK
(M3)	62,4089	OK	OK	***	OK	OK
(M4)	62,5942	OK	OK	***	OK	OK
(M5)	62,4422	OK	OK	***	OK	OK
(M6)	62,6673	OK	OK	***	OK	OK
(M7)	62,5437	OK	OK	***	OK	OK
(M8)	72,8012	OK	OK	***	OK	OK
(M9)	62,5533	OK	OK	***	OK	OK
(M10)	62,4598	OK	OK	***	OK	OK
(M11)	65,4576	OK	OK	***	OK	OK
(M12)	62,8633	OK	*	***	OK	OK
(M13)	66,7386	OK	OK	*	OK	OK
(M14)	58,4456	OK	OK	OK	OK	OK
(M15)	59,363	OK	OK	OK	OK	OK
(M16)	59,4894	OK	OK	OK	OK	OK
(M17)	59,6277	OK	OK	OK	OK	OK

In the diagnostic tests for residual randomness and stationarity tests, ideally an OK or * symbol, indicate none statistically significant violations of model assumptions. If many flags appear (** or *** symbols) meaning significant (p-value between 0,001 and 0,01) or highly significant (with a p-value less or equal to 0,001) probability of non-randomness in the residual plots. The p-value is the probability used in testing a null hypothesis to either reject it or failing doing that (the term "accepting" a hypothesis does not exist in statistics as the logic is that not enough evidence is found to reject that hypothesis) In general, the smaller and more random are the errors, the better. From Table 2, the model that passed all the tests and respected the models assumptions with the lowest error is the model M14, which is *ARIMA(1,0,0)□(2,0,1)12* with constant.

Figure 3 shows the time sequence of the data used which cover 192 time periods with a monthly seasonality (s=12) and how it is explained by the fitting model. It can be clearly seen that the fitting model passes through most of the points of the actual data with some exceptions in the first four years.

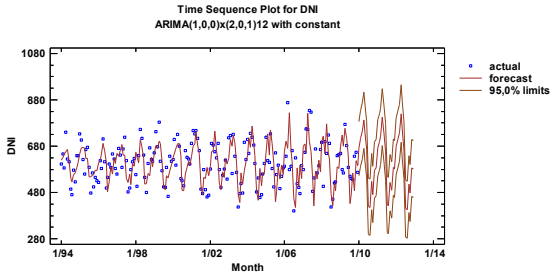

Fig. 3. Time sequence plot for DNI (Wh/m²) using the fitting model ARIMA(1,0,0)□(2,0,1)12 with constant.

Figure 4 presents the forecasts generated for the years 2010, 2011 and 2012 with their 95% confidence interval.

It shows the plot of the forecasted values generated from the selected model versus the actual DNI data. The plot shows a very good following seasonal behaviour from the model and describes in a very good manner the dynamic within the data with differences in the peaks representing the errors.

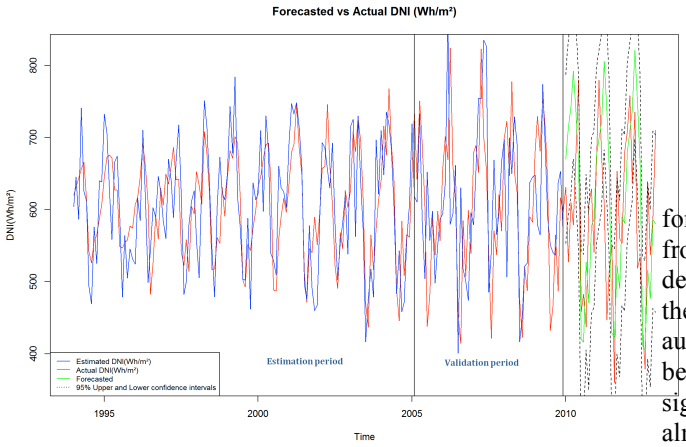

Fig. 4. Plot Forecasted versus Actual DNI in (Wh/m²) in the period (estimation and validation) and the forecast period (2010-2012).

Table 3 presents the summary of the model found. It is worth mentioning that each value of DNI was adjusted using a multiplicative seasonal adjustment before the model was fit. Terms with p-values less 0,05 are statistically significantly different from zero at the 95,0% confidence level. The p-value for the constant term, AR(1), MA(2), and SMA(1) terms is less than 0,05, so they are significantly different from 0. The estimated standard deviation of the input white noise equals 59,294. For more details about the time series analysis, see Box, Jenkins and Reinsel (1994) [33].

Table 3. SARIMA Model Summary

Parameter	*Estimate*	*Stnd. Error*	*T*	*P-value*
AR(1)	0,278525	0,0860106	3,23826	0,001529
SAR(1)	0,898977	0,107209	8,38529	0,000000
SAR(2)	0,189864	0,09973	1,90378	0,059166
SMA(1)	0,985683	0,0720349	13,6834	0,000000
Mean	603,211	18,1462	33,2417	0,000000
Constant	-38,6639			

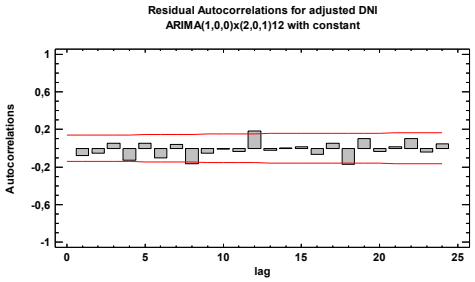

Fig. 5. Residual Autocorrelations for adjusted ARIMA(1,0,0)□(2,0,1)12 with constant.

It is also useful to examine the autocorrelations of the residuals. The residual autocorrelation at lag *k* measures the strength of the correlation between residuals *k* time periods apart. The residual lag *k* autocorrelation is calculated from

$$r_k = \frac{\sum_{t=1}^{n-k}(e_t - \bar{e})(e_{t+k} - \bar{e})}{\sum_{t=1}^{n}(e_t - \bar{e})^2} \quad (30)$$

Where n is the sample size and e_t is one period ahead forecasting errors calculated from subtracting the forecast from the observed data value (i.e. $e_t = Y_t - F_{t-1}(1)$). If a model describes all of the dynamic structure in a time series, then the residuals should be random and all of their autocorrelations should be insignificant. Bars extending beyond the upper or lower limit correspond to statistically significant autocorrelations. From Figure 5 it can be seen that almost no bars appear to be extending from the confidence interval which suggests that the model explains well the dynamic structure in the data. Table 4 presents the results of the model performance in the forecast period where it can be noticed that an increase in all the error metrics has occurred which is normal since most of the times the forecasting models do not perform well in the forecast period yet a MPE of -5,1907% and a RMSE equal to 115,8791Wh/m² still represent an acceptable forecasting accuracy of DNI monthly average data.

Model	*RMSE*	*MAE*	*MAPE*	*ME*	*MPE*
ARIMA(1,0,0)x(2,0,1)12	115,8791	95,7763	16,9438	-22,2736	-5,1907

Table 4. Table of error metrics for the forecast period

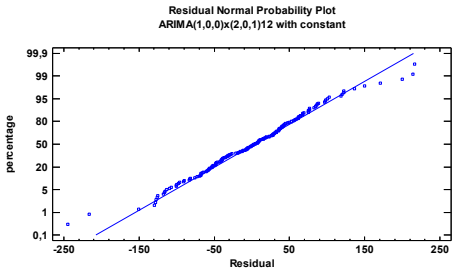

Fig. 6. Residual Normal Probability plot ARIMA(1,0,0)□(2,0,1)12 with constant.

5. Conclusion

The aimof this study is to forecast monthly average DNI using the different statistical methods to assess their forecasting accuracy The results show the performance evaluation of different models for long term forecasting horizon of (DNI) in the region of Ouarzazate. The ARIMA(1,0,0)x(2,0,1)12 with constant was selected after assessment, evaluation and comparison with seventeen other models. The selected seasonal model was used to generate 36 time period forecasts representing the three years (2010, 2011 and 2012)use for forecasting. The model proved to

generate forecasts with a good accuracy represented by MPE= -0.9452% in the estimation period, MPE= -0.490442% in the validation period and MPE= -5.1907% in the forecast region. Although the model reveals a difficulty in including most of the outliers which is penalized by the marginalization of the RMSE, it captures very well the seasonality present in the data. Future works will involve using artificial intelligence methods such as neural networks [34-38] and classification techniques to test their data driven concept for better forecasting skills.

References

[1] E. Cogliani, "The role of the Direct Normal Irradiance (DNI): Forecasting in the Operation of Solar Concentrating Plants", Energy Procedia, https://doi.org/10.1016/j.egypro.2014.03.170, vol. 49, pp.1612-1621, 2014.

[2] N. Kumar, U. K. Sinha, S. P. Sharma, Y. K. Nayak, "Prediction of Daily Global Solar Using Neural Networks With Improved Gain Factors and RBF Networks", International Journal of Renewable Energy Research, vol. 7, pp. 1235-1244, 2017.

[3] M. Demirtas, M. Yesilbudak, S. Sagiroglu, I. Colak, "Prediction of solar radiation using meteorological data", 2012 International Conference on Renewable Energy Research and Applications (ICRERA), Nagasaki Japan, 11-14 November 2012.

[4] T. Koyasu, K. Yukita, K. Ichiyanagi, M. Minowa, M. Yoda, "Forecasting Variation of Solar Radiation and Movement of Cloud by Sky Image Data", 5th International Conference on Renewable Energy Research and Applications (ICRERA), Birmingham UK, pp. 401-406, 20-23 November 2016.

[5] A. Kaur, L. Nonnenmacher, H.T.C. Pedro, C.F.M. Coimbra, "Benefits of solar forecasting for energy imbalance markets", Renewable Energy, https://doi.org/10.1016/j.renene.2015.09.011, vol. 86, pp. 819-830, 2016.

[6] I. Ghalehkhondabi, E. Ardjmand, G.R. Weckman, W.A. Young, "An overview of energy demand forecasting methods published in 2005-2015", Energy Systems, https://doi.org/10.1007/s12667-016-0203-y, vol. 8, pp. 411-447, 2017.

[7] L. Suganthi, A.A. Samuel, "Energy models for demand forecasting-A Review", Renewable and Sustainable Energy Reviews, https://doi.org/10.1016/j.rser.2011.08.014, vol. 16, pp. 1223-1240, 2012.

[8] F. Gürbüz, C. Öztürk, P. Pardalos, "Prediction of electricity energy consumption of Turkey via artificial bee colony: a case study", Energy Systems, https://doi.org/10.1007/s12667-013-0079-z, vol. 4, pp. 289-300, 2013.

[9] Y. Chu, M. Li, H.T.C. Pedro, C.F.M. Coimbra, "Real-time prediction intervals for intra-hour DNI forecasts", Renewable Energy, https://doi.org/10.1016/j.renene.2015.04.022, vol. 83, pp. 234-244, 2015.

[10] Y. Chu, M. Li, C.F.M. Coimbra, "Sun-tracking imaging system for intra-hour DNI forecasts", Renewable Energy, https://doi.org/10.1016/j.renene.2016.05.041, vol. 96, pp. 792-799, 2016.

[11] Y. Chu, C.F.M. Coimbra, "Short-term probabilistic forecasts for Direct Normal Irradiance", Renewable Energy, https://doi.org/10.1016/j.renene.2016.09.012, vol. 101, pp. 526-536, 2017.

[12] R. Marquez, C.F.M. Coimbra, "Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database", Solar Energy, https://doi.org/10.1016/j.solener.2011.01.007, vol. 85, pp. 746-756, 2011.

[13] C.W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, B. Washom, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed", Solar Energy, https://doi.org/10.1016/j.solener.2011.08.025, vol. 85, pp. 2881-2893, 2011.

[14] S. Quesada-Ruiz, Y. Chu, J. Tovar-Pescador, H.T.C. Pedro, C.F.M. Coimbra, "Cloud-tracking methodology for intra-hour DNI forecasting", Solar Energy, https://doi.org/10.1016/j.solener.2014.01.030, vol. 102, pp. 267-275, 2014.

[15] V. Bone, J. Pidgeon, M. Kearney, A. Veeraragavan, "Intra-hour direct normal irradiance forecasting through adaptive clear-sky modeling and cloud tracking", Solar Energy, https://doi.org/10.1016/j.solener.2017.10.037, vol. 159, pp. 852-867, 2018.

[16] R.C. Deo, M. Şahin, "Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland", Renewable and Sustainable Energy Reviews, https://doi.org/10.1016/j.rser.2017.01.114, vol. 72, pp. 828-848, 2017.

[17] D.V. Siva Krishna Rao K, M. Premalatha, C. Naveen, "Models for forecasting monthly mean daily global solar radiation from in-situ measurements: Application in Tropical Climate, India", Urban Climate, https://doi.org/10.1016/j.uclim.2017.11.004, vol. 24, pp. 921-939, 2017.

[18] G. Klæboe, A.L. Eriksrud, S.E. Fleten, "Benchmarking time series based forecasting models for electricity balancing market prices", Energy Systems, https://doi.org/10.1007/s12667-013-0103-3, vol. 6, pp. 43-61, 2015.

[19] A. Wold, Analysis of stationary time series, Uppsala: Almquist and Wicksell, 1938.

[20] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time series analysis: forecasting and control, 4th ed, Wiley, 2008.

[21] R.H. Inman, H.T.C. Pedro, C.F.M. Coimbra, "Solar forecasting methods for renewable energy integration", Progress in Energy and Combustion Science, https//doi.org/101016/j.pecs.2013.06.002, vol. 39, pp. 535-576, 2013.

[22] M. Diagne, M. David, P. Lauret, J. Boland, N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids", Renewable and Sustainable Energy Reviews, https://doi.org/10.1016/j.rser.2013.06.042, vol. 27, pp. 65-76, 2013.

[23] R.H. Shumway, D.S. Stoffer, Time series analysis and its applications: with R examples, 3rd ed, Springer texts in statistics Springer New York, 2011.

[24] G. Schmueli, C. Peter, I. Yahav, R.P. Nitin, K.C. Lichtendahl Jr, Data Mining for Business Analytics-Concepts, Techniques and Applications in R, Wiley United States, 2018.

[25] G. Schmueli, K.C. Lichtendahl Jr, Practical Time Series Forecasting with R: A Hands-On Guide, 2nd ed, Axelrod Schnall Publishers, 2016.

[26] https://people.duke.edu/~rnau/411avg.htm#SES (Last checked 15th November 2018)

[27] https://people.duke.edu/~rnau/411avg.htm#BrownLES (Last checked 11th November 2018)

[28] cdn2.hubspot.net/hubfs/402067/PDFs/Exponential_Smoothing_Statlet.pdf (Last checked 15th May 2018)

[29] https://people.duke.edu/~rnau/411avg.htm#HoltLES (Last checked 11th November 2018)

[30] https://people.duke.edu/~rnau/Slides_on_forecasting_with_inflation_seasonal_adjustment_and_Winters_model--Robert%20Nau.pdf (Last checked 11th November 2018)

[31] R. Adhikari, R. K. Agrawal, "An Introductory Study on Time Series and Forecasting", LAP Lambert Academic Publishing, Germany, https://arxiv.org/abs/1302.6613v1, 2013.

[32] A. Hirotugu, "A New Look at the Statistical Model Identification", IEEE Transactions on automatic control, DOI: 10.1109/TAC.1974.1100705, vol. 19 issue 6, pp.716-723, 1974.

[33] G. E. P. Box, G. M. Jenkins, G.C. Reinsel, Time Series Analysis: Forecasting and Control, 3rd ed, Prentice Hall: Englewood Cliff New Jersey, 1994.

[34] E. Jahani, S.M. Sajed Sadati, M. Yousefzadeh, "Assessment of solar data estimation models for four cities in Iran", physica status solidi C, DOI: 10.1002/pssc.201510105, vol 12, No 9-11, pp. 1272-1275, 2015.

[35] A. H. Assi, M. H. Al-Shamisi, H. A. N. Hejase, "Prediction of Global Solar Radiation in UAE Using Artificial Neural Networks", 2013 International Conference on Renewable Energy Research and Applications (ICRERA), Madrid Spain, pp. 196-200, 20-23 October 2013.

[36] K. Gairaa, F. Chellali, S. Benkaciali, "Daily Global Solar Radiation Forecasting Over a Desert Area Using NAR Neural Networks Comparison with Conventional Methods", 4th International Conference on Renewable Energy Research and Applications (ICRERA), Palermo Italy, pp. 567-571, 22-25 November 2015.

[37] A. Alzahrani, P. Shamsi, M. Ferdowsi, C. Dagli, "Solar Irradiance Forecasting Using Deep Recurrent Neural Networks", 6th International Conference on Renewable Energy Research and Applications (ICRERA), San Diego USA, pp. 988-994, 5-8 November 2017.

[38] R. Al-Hajj, A. Assi, M. M. Fouad, "A Predictive Evaluation of Global Solar Radiation using Recurrent Neural Models and Weather Data", 6th International Conference on Renewable Energy Research and Applications (ICRERA), San Diego USA, pp. 195-199, 5-8 November 2017.