Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 1, March 2020, pp.455-463

Statistical model of forecasting indicators in one area of sustainable development

Shatha Abdullah Mohammed[1], Obay Abdullah Mohammed[2], Mohammed Ghanim Fathi [3]
[1] Deptartment of Software, University of Mosul
[2] Ministry of Industry and Minerals, State Company for the Manufacture of Medicines and Medical Supplies Samarra, Nineveh Pharmaceuticals Factories
[3] University of Mosul

ABSTRACT

This study is based on the seventh goal of sustainable goals development and under the title of "clean and affordable energy". As a result of the continuing increase in the number of the world's population, the dramatic rise in energy demand, especially in the production of the renewable energy worldwide, was observed.

The Arab region has the highest levels of solar radiation. Many Arab countries are still taking slow steps to develop their renewable energy capacity, which has led many Arab countries today to seek to meet this demand through renewable energy solutions and improved energy efficiency.

The field of study presented in this research includes a proposal for the statistical model to analyze the prediction in the field of solar energy, which is one of the most important areas of renewable energy, by providing a comparison between the time series model and the simple regression model to predict data which has been obtained in the field of solar energy.

Keywords: Sustainable Development (SD), Clean Energy(CE), Forecasting, Time Series Model (TSM), Simple Regression Model(SRM)

Corresponding Author:
Shatha Abdullah Mohammed
Deptartment of Software
University of Mosul, Iraq
Email: shathaabdullah@uomosul.edu.iq

1. Introduction

Sustainable development is the progress that meets the needs of the present days without compromising the ability of future generations to meet their needs. Equality and equity between generations are one of the factors required for sustainable development. Sustainable development is the ideal and effective use of all environmental resources, social life and economy for the distant future with a focus on a better life of high value to every member of society now and in the future [6].

The field of development from abstract environmental thinking to sustainable development includes the inevitable linkage of the environment to economic and social transformations. This is in order to find alternatives and improve the efficiency of economic growth to use less natural resources and reduce consumption. These goals include the initiative (WEHAB) which is an abbreviation of the words (water, energy, health Agriculture and Biological Diversity), as well as the 2002 meeting of the Council of Arab Ministers Responsible for the Environment, presented by the Sustainable Development Initiative [10]

The challenge for the international community now is how to achieve economic development and social welfare with minimal consumption of natural resources and minimal pollution and environmental damage, which is the essence of sustainable development [12].

The usual figures show a change in one direction only without showing the related points between society, economy, social life, and the environment if the society consists of three separate parts (Economy, Social life, and Environment). At the ninth session of the United Nations Commission on Sustainable Development (CD9) on 5/2001, it stated that energy is the most important for achieving sustainable development and identified 5 key issues related to [13,19]:

1 -Increasing access to advanced energy services.
2 -Improve energy efficiency and consumption.
3 -Developing the uses of renewable energy resources.
4-Developing cleaner technologies for fossil fuels.
5 -Energy in transport

The sustainable development is based on environmental, economic, social, and technological elements. The fourth element is the tool to achieve a rapid transformation in the technological base of industrial communities to new technology which is cleaner, more efficient and able to reduce environmental pollution and create a technological transformation in industrializing developing countries while avoiding Repeating development mistakes and avoiding environmental pollution caused by industrialized countries, so as to reconcile development goals with the rules imposed by the environment [21].

1.1. Research problem

One of the elements of sustainable development represented by renewable energy has been adopted through the analysis and investigation of the indicators, based on both simple regression models and time series models first, and then the trade-offs between them. This was done through the adoption of solar energy generation and its positive effects on the environment, which is the main pillar of man. Population communities are the fourth element of sustainable development through the identification of renewable energies and their wealth of positive effects on the environment. The solar energy and its stages and requirements and the way of its production, so we must identify the positive effects, especially the economic energy production through the presentation of expenses or imports by employers in this area.

1.2. Solar cell systems and components

Solar, photovoltaic or photovoltaic device A, is a device that converts solar energy directly into electrical energy utilizing the photovoltaic effect. The origin of the term “photovoltaic” from Greek (ō (phōs)) means light and the name Volta, an Italian physicist, Volt is a unit of electric impulse and thus the term 'photovoltaic' has been in English since 1849. The photovoltaic effect was first recognized in 1839 by the French physicist Becquerel. However, the first photovoltaic cell was built in 1883 by Charles Fritz. Concentrations of solar cells (solar modules) are used to capture energy from sunlight to convert it into electricity when multiple modules are grouped together, where the priority is to install a portable polar tracking system. These photovoltaic cells are installed as a single unit that is directed on a single surface, called a power board. Examples of the use of solar energy are discussed in [2,3,4,5] and [22].
Photovoltaic is the field of technology and research related to the practical application of electricity production from light but nevertheless is often used specifically to refer to the generation of electricity from sunlight. Cells are described as photovoltaic cells, although the light source is not the sun like lamp light or artificial light. Photovoltaic is used to detect light or other electromagnetic radiation near a visible light array, such as infrared detection, or measuring the intensity of light [7,8,9,18].

The essential components as depicted in Figures 1-4 are:

1- Solar Panels
2- Charger Controllers
3- Batteries
4- Power Inverters

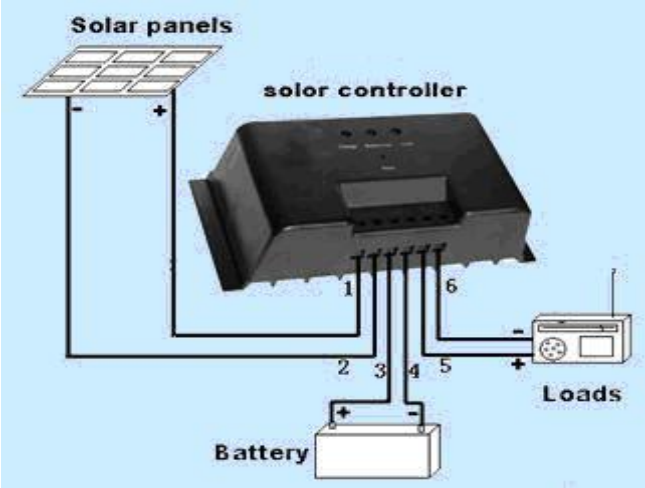

Figure 1. Components of the solar cell system

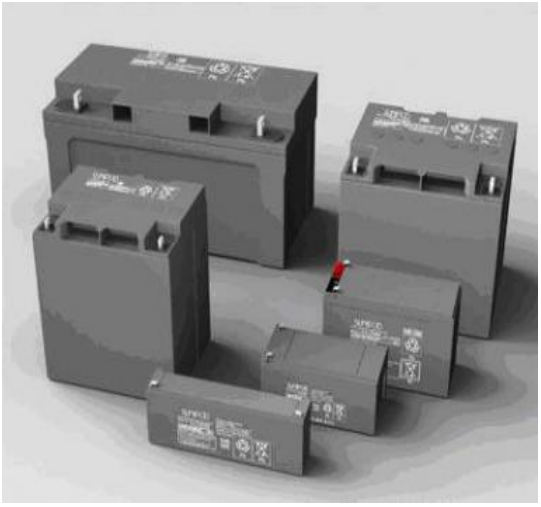

Figure 2. Batteries

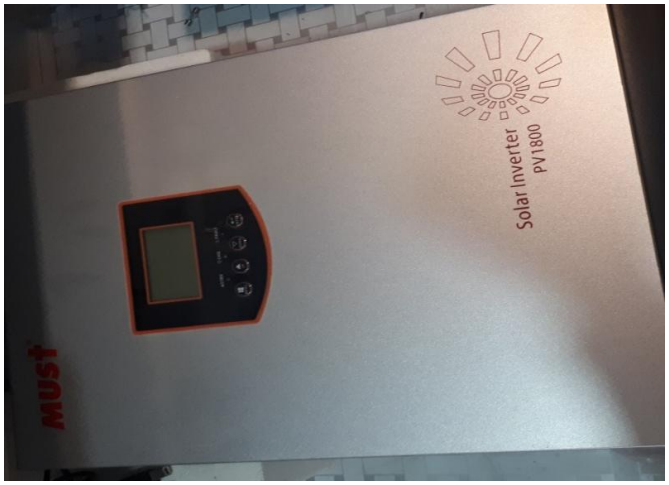

Figure 3. Power Inverters

Figure 4. Solar Panels

The initial cost of purchasing the system is fairly high, which includes the price of panels, inverters, batteries, wires, and the installation structure.

This cost is an obstacle to low-income people who currently have access to electricity subsidized by the government and do not need a very expensive solar system. So, it was necessary to give some clarification to show the amount of cost and profit to employers in this area, which is the profit and expenses when establishing solar power generation systems [14,15].

2. Research method

The using methodology is a very important that build's the models for Analyzing, Data Evaluation, Forecasting & Controlling depending on classic statistic methods (simple regression and time series model) that describe the behavior of statistic phenomenon, the methodology works by simulating the data to reach a robust optimum model. Those represent the statistic phenomenon and we can use the model in any time and states [1,11,16,17,20].

This work depends on the received power to build a robust model for forecasting, analyzing and controlling in the renewable energy, which represented by the solar energy.

The sales volume has been approved in addition to the expenses by the employers, when establishing the solar systems, for a period of 24 months (two years), if we want to predict the volume of sales for subsequent periods, we have two cases or two possibilities so that the volume of sales can be considered a function of the number of expenses, cost, as a time series and can be analyzed by time series in the interpretation of the phenomena.

In an attempt to find the best model, the differentiation criteria will be applied, as in the following steps:

1. Building a simple regression model and perform various statistical tests. Then use the prediction model for the next period.

2. Build a time series model of sales and conduct various statistical tests and then use the model in the prediction of the next period and then determine the scope of the predicted value as we measure the accuracy of predictive results of the model.

3. We conduct different criteria to differentiate between the two models through:

• Statistical tests to judge the model in the prediction, which depends on the field of the predicted value.

• The trade-off between the two models according to the Thiel rule.

• The trade-off between the two models according to the predictive accuracy indicators of the output.

Table 1. Total costs and sales

	Time		Cost	Sales
First Time(FT)	FT-1	1	912.059	34549.7
	FT-2	2	1381.6	33532.8
	FT-3	3	1644.79	34612.3
	FT-4	4	1232.64	26154.7
	FT-5	5	1181.6	47553.3
	FT-6	6	874.708	30397.45
	FT-7	7	275.02	18763.75
	FT-8	8	703.781	31494.35
	FT-9	9	1411.49	32872.35
	FT-10	10	1696.42	47558
	FT-11	11	1728.03	33571.1
	FT-12	12	1912.02	45940.35
Second Time(ST)	ST-1	13	1500.61	54609.2
	ST-2	14	1704.48	45035.7
	ST-3	15	2134.55	52692
	ST-4	16	1786.17	53964.65
	ST-5	17	2249.81	56573.15
	ST-6	18	2353.87	49437.45
	ST-7	19	1443.68	41857.55
	ST-8	20	710.331	31224.05
	ST-9	21	1946.57	41842.1
	ST-10	22	1681.06	39889.8
	ST-11	23	1495.45	39648.35
	ST-12	24	1734.68	39590.8

2.1. The simple regression model

From the sales and expense diagram, the set of coordinate points shown in a cloud of points are close to each other, indicating a relationship between the two variables. So, the correlation coefficient must be calculated to determine how strong the relationship is between the sales and the expenses (costs), see Figure 5.

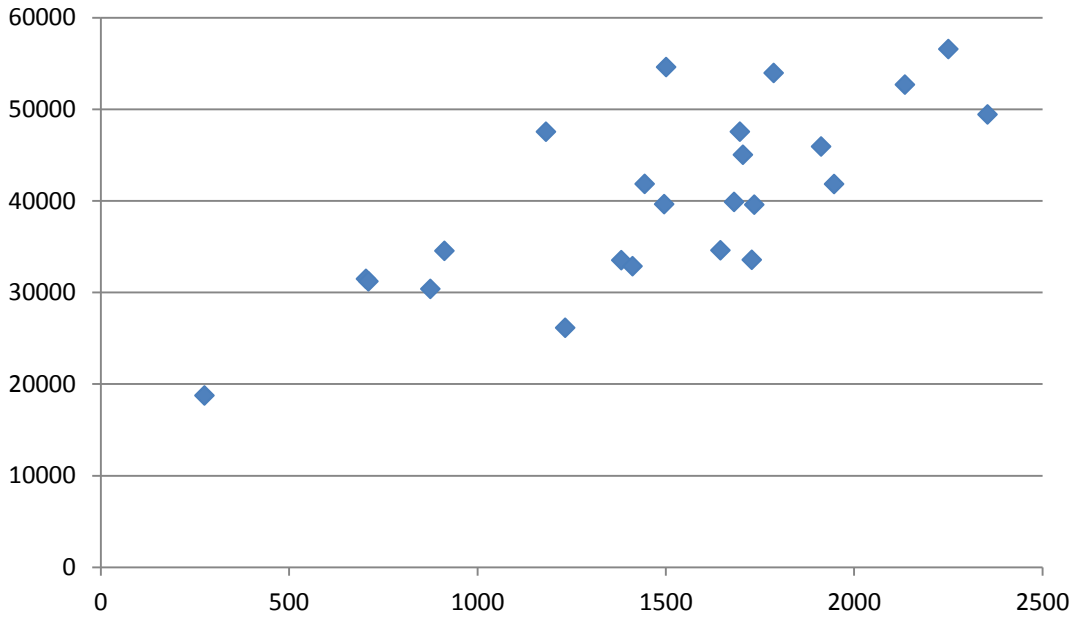

Figure 5. Sales according to cost expenses

$$r = \sqrt{\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}}\ ,\quad r = 0{,}758$$

Which indicates that the relationship between sales and expenses is a direct relationship (i.e. increasing)

Where n<30, therefore $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \rightarrow t = 5.44, \alpha = 5\%, df = 22$

$$\rightarrow t_{table} = 1.717.$$

Also, the descent between the variables is Y_i,

$$Y_i = a + bX_i + e_i$$

Where $a\ and\ b$ are :

$$b = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i)^2\, n\bar{X}^2},\quad a = (\bar{Y} - b\bar{X}) \rightarrow a = 18712{,}22, b = 144{,}07$$

$$Y_i = 18712{,}22 + 144{,}07X_i \qquad (1)$$

Equation (1) represents the Simple Regression.

2.2. Time series model

The time series diagram depicted in Figure 6, represents the number of sales where the presence of a general trend vehicle and the random vehicle is shown. Time series is based on aggregate, multiplier, and mixed varieties.

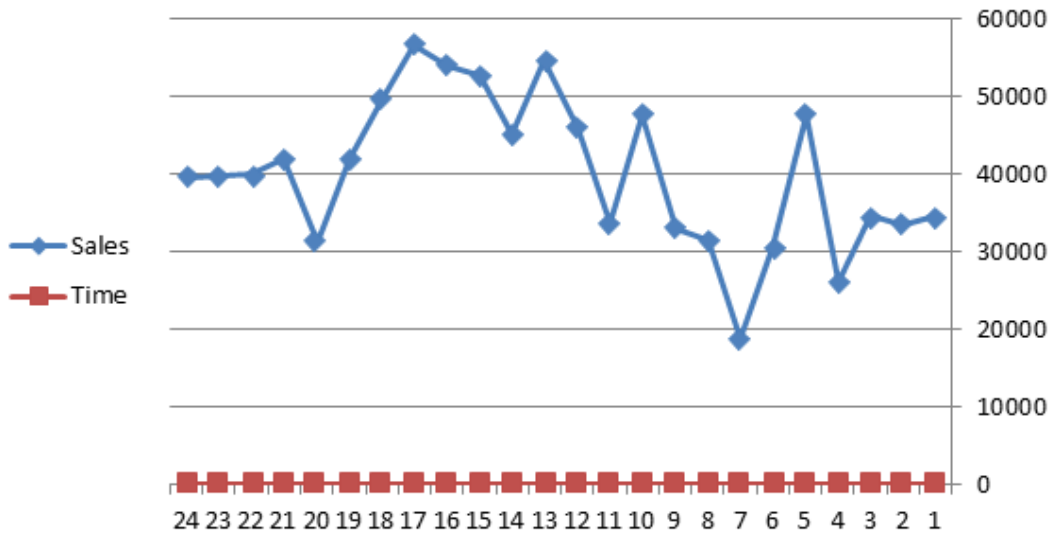

Figure 6. Sales according to time

Where $a\ and\ b\ as$:

$$b = \frac{\frac{(n1)}{2}\left[\sum(Y_{ij}) - (n2)\bar{\bar{Y}}\right]}{n(n3)},\quad a = \left(\bar{\bar{Y}}\right) - b(n4)$$

$$\rightarrow a = 29077.2,\quad b = 898.4$$

$$Y_i = 29077.2 + 898.4(j + m(i-1)) + K \qquad (2)$$

$$K = [K1 \quad K2 \quad K3 \quad K4]'$$

Where $K, n1, n2, n3$:

$$K1 = [9213.45 \quad 3019.85 \quad 6489.35]$$

$$K2 = [1998.47 \quad 13103.62 \quad 59.45\ ]$$

$$K3 = [-10445.75 \quad -10295.6 \quad -5195.97\ ]$$

$$K4 = [272.3 \quad -7740.27 \quad -2482.82\ ]$$

Equation (1) represents the time series model.

2.3. The trade-off between the two models

Before judging which models are better at predicting sales volumes, the results of statistical tests and conciliation quality tests for each model should be addressed. The results are summarized in Table 2.

Table 2. Summary of the results of statistical tests and the quality of conciliation

Test	Simple regression model	Time series model
The correlation coefficient	0.78	0.618
Significance of correlation coefficient	Correlation coefficient is significance(isn`t stochastic)	Correlation coefficient is significance(isn`t stochastic)
Statistical tests of parameters: standard error tests : For: 1- fixed amount a,	The presence of statistical significance	The presence of statistical significance

Test	Simple regression model	Time series model
2- slop b t-student tests of parameters: standard error tests: for: 1- fixed amount a, 2- slop b	The presence of statistical significance	The presence of statistical significance
Conciliation quality tests: R2	Middle	High
Conciliation quality tests: By distribution F	statistical significance of the model as a whole	statistical significance of the model as a whole

From the observation of Table 2, the statistical criteria show the statistical significance of the simple regression model and the time series model, while there is some inconsistency in the conciliation quality tests. For sales, the quality of reconciliation is fairly high with the statistical significance of the model. So, you can't prefer one model over the other.

Therefore, more precise criteria are used in the results to determine the preference for any model. In this step, the following criteria are used to differentiate between the two models as follows.

2.3.1. According to the judgment of predictions (areas of the prediction)

The predicted value according to the simple regression model is 46085.52 units sold, and the simple regression model shows a range of predicted values with a probability of 95% is: $\hat{Y} \epsilon$[59853.62;32317.41], and the extent of this area is 27536.2 sales.

The predicted value according to the sales time series model shows that the time series model is: 60750.65, it also shows the scope for the predicted value with a probability of 95% is: $\hat{Y} \epsilon$[73599.57;47901.72]. The range of this area is 25698 for sales.

2.3.2. Thaeil-rule

The mean error estimation square for a simple regression model is 42305240.59. The average estimation error box for the time series model is 32594351.8.

Through this criterion (Thiel's trade-off rule), it can be judged that a simple sales regression model is better than a sales time series model in predicting the sales.

2.3.3. Using the comparison between the results of the measurement accuracy

Table 3 summarizes the four indicators of criteria for measurement accuracy results for both models.

Table 3. The summarization of measurement accuracy for the models

The scalars		Simple regression model	Time series model
Mean Absolute Error (MAE)	$\left(\sum \lvert e_i \rvert\right)/n$	5035.25	4572.29
Mean Squared Error (MSE)	$\left(\sum {e_i}^2\right)/n$	38779803.89	29878155.83
Mean Absolute Percentage Error (MAPE)	$\left(\sum \frac{\lvert e_i \rvert}{Y_i}\right)/n$	13.11%	12.43%
Mean Percentage Error (MPE)	$\left(\sum \frac{{e_i}^2}{Y_i}\right)/n$	972.13%	804.44%

All indicators of accuracy of the predictive results require that the model with a lower value for each indicator to be the best model (in terms of measurement accuracy). The table shows that the sales time series model has the lowest values for the four indicators and conversely in the simple regression model. Therefore, the time series model is better than the simple regression model in predicting the sales volume, according to the indicators of measurement accuracy.

So, all three differentiation criteria refer to the preference of the sales time series model, which interprets sales according to the time estimate of sales over a simple regression model. It interprets sales as an affiliate of expenses, then the sales time series model is better than the simple regression model to predict the sales volume, by choosing sales volume forecasting series The approved model estimates the total sales volume (A1, A2, A3) and its value (60750.65) through which sales can be distributed (A1, A2, A3) in the two years as follows:

	First-year	Second-year	Total	Mean
A1:	9194.15	24507.35	33701.5	1685075
A2:	17665.05	17284.35	34949.4	17474.7
A3:	7690.5	12817.5	20508	10254
Total:	34549.7	054609.2	989158.9	44579.45

A1: 22963,61, A2: 23813,64, A3: 13973,63

So the selling of A1, A2, A3 is expected as in the above values, as these estimates are found respectively. By using the both two models according to the previous steps, by subjecting the two models to the different criteria for differentiation, it is possible to determine which of the two models is most appropriate to predict the volume of sales without the other model. The differentiation process must be repeated periodically and at the beginning of each session.

3. Conclusions

In this research, a simple regression model and a time series model of sales from solar power generation systems were used. Due to the importance of forecasting sales volumes, the various estimates have made by the employers adopted these functions. The forecasts have a clear impact on the performance and spread of this technique.

Through the good construction of the model used in the sales volume forecasting process, it is possible to obtain good predictions about the volume of sales and the resulting good estimates and make efficient decisions.

This research revolves around the preference of the model used in predicting sales volume. It gives the best predictions and the best interpretation of sales, besides the adoption of the basic premise that the preference (or efficiency) of the model used in predicting sales based on a set of statistical considerations and not on the nature of the model itself. In order to answer the questions of the research problem and confirm its hypotheses, the study was presented in three main axes, namely:

1. Clarifying various concepts related to forecasting in general and forecasting sales volume in particular.
2. Analytical study of each model by studying the elements of each model and its most important factors.
3. Determining the various criteria used in the comparison between the two models, how they are calculated and how they are used.

References

[1] A. David et al," Distribution of the Estimators for Autoregressive Time Series With a Unit Root", Journal of the American Statistical Association, vol. 74, no. 366, United states, 1979.

[2] A.J., Marszal, P., et al. Zero Energy Buildings,' A Review of definitions and calculation methodologies, Energy and Buildings', vol.43, no. 4, pp. 971-979, 2011.

[3] Dajiang Yang & Huiming Yin, Energy Conversion Efficiency of a Novel Hybrid Solar System for Photovoltaic, Thermoelectric, and Heat Utilization, IEEE Transactions on Energy Conversion, vol. 26 ,no. 2, pp.662 – 670, 2011.
[4] M.R.AlRashidi, et al "New Estimation Approach for Determining the I–V Characteristics of Solar Cells",vol. 85, no. 7, pp.1543-1550, 2011.
[5] Z.Omar, et al "Solar Collector Systems: Part I – Fundamentals, Design Considerations and Current Technologies", Concentrated photovoltaic thermal (CPVT), vol. 50, pp.1500-1565, 2015.
[6] A. Starr, "Naming the Enemy: Anti-Corporate Movements Confront Globalization", Sydney: Pluto Press, 2000.
[7] E. Maldonando, "Implementing the Energy Performance of Building Directive (EPBD)". ISBN 978-92-9202-090-3 EA-30-11-026-ENC. Brussels, April 2011.
[8] H. Dokka, et al. "Energi Effektivisering i by Gninger", Oslo: SINTEF Report in Norwegian, 2009.
[9] J. Andrews et al, "Energy Science – Principles, Technologies and Impacts", Oxford: Oxford University Press, 2013.
[10] United Nations Development Programme (UNDP), Human Development Report, 1995, New York: Oxford University Press, 1995.
[11] V. Zarnowitz, "*Theory, History, Indicators, and Forecasting*", Chicago, University of Chicago Press, 1992.
[12] WCED (World Commission on Environment and Development), Our Common Future, Oxford: Oxford University Press, 1987.
[13] P. Wenz. Environmental Justice. New York: State University of New York Press, 1988.
[14] E.Mlecnik, et al. "Web Platforms Integrating Supply and Demand for Energy Renovations. Paper presented", at the 4th Nordic Passive House Conference in October 2011. ISBN 978-951-758-535-4, 2011.
[15] S. Tamrakar1, et al, "Feasibility Study for Utilization of Solar Energy in the Arctic Areas", IOP Conference Series: Materials Science and Engineering, Vol.700, 2nd Conference of COTech & OGTech 2019) 27–29, Stavanger, Norway, IOP Publishing Ltd, November 2019.
[16] B. S.Everitt, Torsten Hothorn, A handbook of statistical analysis using R,2nd edition, Taylor & Francis Group LLC, USA, 2010.
[17] R.J. Hyndman, and G. Athanasopoulos, Forecasting: principles and practice, 2nd edition, Australia, 2018.
[18] A. R. Jha, "Solar Cell Technology and Applications", CRC Press, New York, Auerbach Publication, 2009.
[19] R. L. Thomas, *Modern econometrics: an introduction*. Addison-Wesley Longman, 1997.

[20] I. Pardoe, "Applied regression modeling", 2nd edition, John Wiley & Sons Inc, New Jersey, 2012
[21] V. Shiva, Staying Alive: Women, Ecology, and Development. London: Zed Books, 1988.
[21] A. Khechekhouche, N. Elsharif, I. Kermerchou, and A. Sadoun, "Construction and performance evaluation of a conventional solar distiller", *Heritage and Sustainable Development*, vol. 1, no. 2, pp. 72-77, Dec. 2019.